Biologicals xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Biologicals

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

Use of a new RNA next generation sequencing approach for the specific detection of virus infection in cells

Audrey Brussel[a], Kerstin Brack[b], Erika Muth[c], Rudolf Zirwes[c], Justine Cheval[c], Charles Hebert[c], Jean-Marie Charpin[c], Alice Marinaci[b], Benoit Flan[a], Horst Ruppach[b], Pascale Beurdeley[c], Marc Eloit[c,d,e,*]

[a] LFB, Courtaboeuf, France
[b] Charles River Laboratories Germany GmbH, Erkrath, Germany
[c] PathoQuest, Paris, France
[d] National Veterinary School of Alfort, Paris-Est University, Maisons-Alfort, France
[e] Pathogen Discovery Laboratory, Biology of Infection Unit, Institut Pasteur, Paris, France

ARTICLE INFO

Keywords:
Virus testing
Next generation sequencing
Metagenomics
Biologicals
Viral safety
Adventitious virus

ABSTRACT

The utilization of the current combination of *in vitro*, *in vivo* and PCR assays for the identification of adventitious viruses in production cells has a limited range of detection. While Next Generation Sequencing (NGS) has a broader breadth of detection, it is unable to differentiate sequences from replicating viruses versus background inert sequences. In order to improve NGS specificity, we have designed a new NGS approach which targets subsets of viral RNAs only synthesized during cell infection. In order to evaluate the performance of this approach for detecting low levels of adventitious viruses, we selected two difficult virus/cell systems. This included B95-8 cells persistently infected by *Human herpesvirus 4* (HHV-4) and serially diluted into HHV-4 negative Ramos cells and Madin-Darby bovine kidney cells with an early infection produced via a low dose of *Bovine viral diarrhea virus*. We demonstrated that the sensitivity of our RNA NGS approach was equivalent to targeted PCR with an increased specificity for the detection of viral infection. We were also able to identify a previously undetected *Murine Leukemia Virus* contaminant in Ramos cells. Based on these results, we conclude that this new RNA NGS approach is suitable for conducting viral safety evaluations of cells.

1. Introduction

Insuring viral safety during the development and manufacture of biological and medicinal products is a critical quality parameter. All animal-derived materials used for the manufacture of biological products need to be free from adventitious viruses. Currently *in vivo* and *in vitro* assays are utilized for the detection of viruses in biological products. This includes the testing of starting materials such as cell substrates, virus seeds, and cell culture intermediates or ancillary materials such as porcine trypsin and animal serum.

Currently utilized tests were developed with known viruses discovered via isolation in cell culture or in animals. This results in a false assumption that all potential viral contaminants can be identified by *in vitro* and *in vivo* assays. With the advent of the use of molecular biology assays like PCR, the discovery of viruses is now primarily based on the detection of viral nucleic acids. This includes the identification of wild-type strains of human viral pathogens which cannot be isolated and identified by current *in vitro* methods. Recent studies have demonstrated the viral detection and sensitivity limitations associated with current *in vivo* assays and recommended *in vitro* 9 CFR testing method [1,2]. This is especially important for biological products manufactured using bovine or porcine raw materials. While electron microscopy is used as an alternative method to screen for viral contaminants in biologic products since it is suitable for evaluating the presence of virus-like particles, it lacks sensitivity and is operator dependent. The above lead to both *in vitro* assays and electron microscopy frequently failing to clearly identify contaminants and the requirement for subsequent analyses.

Although targeted PCR is a useful testing alternative, it is limited to only detecting and identifying viruses known or suspected to be potential contaminants in the material being analyzed. When animal origin materials are used for the production of biologicals, the number of PCR tests is typically limited to a small list of animal viruses with known pathogenicity in the original host species, however, these may

* Corresponding author. PathoQuest, 11 rue Watt, 75013, Paris, France.
E-mail address: marc.eloit@pathoquest.com (M. Eloit).

https://doi.org/10.1016/j.biologicals.2019.03.008
Received 21 December 2018; Received in revised form 15 March 2019; Accepted 31 March 2019

Abbreviations

(−)ssRNA	negative sense single stranded RNA
(+)ssRNA	positive sense single stranded RNA
ATCC	American Type Culture Collection
BVDV-2	*Bovine Viral diarrhea virus* type 2
BVDV-3	*Bovine Viral diarrhea virus* type 3
CFR	code of federal regulations
d.p.i.	days post infection
dsRNA	double stranded RNA
ECACC	European Collection of Authenticated Cell Cultures
FCS	Fetal Calf Serum
HHV-4	*Human Herpes Virus type 4* (Epstein-Barr virus)
MEM	Minimum Essential Medium
MDBK	Madin-Darby Bovine Kidney Cells
MuLV	*Murine leukemia virus*
NGS	Next Generation Sequencing
PBS	Phosphate Buffered Saline
ssRNA	single stranded RNA
TPA	12 o-tetradecanoylphorbol-13-acetate
RNA-Seq	RNA sequencing

have questionable relevance in humans. Even when using human material, the standard list of PCR tests utilized do not encompass all known human viruses. Targeted PCR also cannot typically detect unknown viruses and frequently fails to detect variants. The above explains why cases in which adventitious viruses are found in biological products or substrates used for production still occur [3,4]. Newer technologies capable of detecting and identifying a wider range of potential adventitious viral contaminants, both known and unknown, are becoming increasingly utilized due to these limitations in adventitious agent testing.

Unlike PCR testing approaches which use prior knowledge of the target of interest, NGS does not require prior assumptions regarding what virus should be included as a part of the testing. NGS is able to detect all types of nucleic acid sequences in a sample, therefore theoretically enables the detection of all known viruses as well as viruses that have yet to be discovered which show some identity for essential viral proteins, like polymerases. This difference is reinforced by the major role NGS plays in the discovery of novel viruses [5]. NGS technology has gained significant attention in the field of biologics since 2010 when an academic laboratory found a porcine circovirus 1 (PCV1) contamination in a licensed pediatric vaccine [3]. Since then, NGS has become increasingly appealing as a supplemental test and is now considered as an alternative method for adventitious virus testing during the production of biologics [6–9].

A major concern restricting the use of NGS in testing of biologics is its potential to detect viral sequences that are not related to infectious viruses. Carryover of nucleic acids from inactivated viruses introduced through reagents, water or media can lead to false positive results. These false positive results could potentially trigger unnecessary, lengthy and resource-consuming investigations to evaluate the biological relevance of the findings in order to rule out the presence of replicating viruses. In the present paper we demonstrate the ability to use NGS to test cell substrates for viral infection and differentiate hits from non-infectious viral sequences by using agnostic transcriptome analysis via the targeting a subset of viral RNAs which are only synthesized during cell infection.

To test our approach, we compared NGS and RT-PCR data from early time points after acute infection of cells with BVDV-2 (*Flaviviridae*, positive ssRNA virus) to data obtained during cultivation with non-infectious gamma irradiated BVDV-2. A non-cytopathogenic BVDV-type 2 strain was selected since it mimics cases where BVDV-2 or other viruses potentially contained in serum is gamma-irradiated before being used in cell culture, raising issues to distinguish viral nucleic acids from those coming from a true cell infection. The second model consisted of the detection of HHV-4 (*Herpesviridae*, linear dsDNA) in persistently infected cells serially diluted in HHV-4 non-infected cells to mimic low level of cell infection.

2. Material and methods

2.1. Cells and viruses

We utilized three different cell lines for the present study. The first was B95-8 cells (ECACC 85011419), a cotton-top tamarin monkey (*Saguinus oedipus*) peripheral blood lymphocyte B-cell line transformed with HHV-4, a dsDNA virus. In most B95-8 cells approximately 20 copies per cell of HHV-4 genomic DNA remain latent, expressing only a limited number of HHV-4 gene products (nuclear proteins, membrane proteins, small RNAs) [10]. HHV-4 genomic DNA molecules are also only replicated once during the S phase of the cell cycle. Only a fraction of the cells is productively infected and releases HHV-4. In the presence of TPA or under certain conditions of culture, a productive viral cycle is induced in the cell population [11,12]. Virus production was therefore further increased in some test samples by using TPA (Sigma, Germany) induction at a 2 ng/ml final concentration. The cells were then cultured at 37 °C in a humidified atmosphere with 5% CO_2 in RPMI 1640 medium supplemented with 10% (v/v) FCS, heat inactivated, (ThermoFisher, Cat. No. 10500, Lot No. 08F1457K). We also utilized Ramos cells (ECACC 85030802), a human B lymphocyte cell line derived from a Burkitt's lymphoma that was negative for HHV-4. These cells were cultured at 37 °C in a humidified atmosphere with 5% CO_2 in RPMI 1640 medium supplemented with 10% (v/v) heat-inactivated FCS. The third cell line we used was a MDBK (NBL-1) bovine kidney cell line purchased from CLS Cell lines Service GmbH (Eppelheim, Germany). These cells were grown at 37 °C in a humidified atmosphere with 5% CO_2 using MEM, supplemented with non-essential amino acids and 10% (v/v) gamma-irradiated FCS (GE Healthcare/HyClone, Cat. # SV30160.03IR, Lot No. RAE35928). Accutase (Innovative Cell Technology, San Diego, USA) was used to detach the cells for each subculture.

The BVDV-2, non-cytopathic strain Munich 2 used for the study was obtained from the Friedrich-Loeffler Institute, Greifswald-Insel Riems (Germany) (Cat. No. RVB-0241). A virus stock was produced using MDBK cells in the presence of 10% gamma-irradiated FCS. The titer of the virus stock was determined on MDBK cells using indirect immunofluorescence staining (see section 2.4). For additional experiments, BVDV-2 was inactivated by gamma-irradiation (40 kGy ± 10%).

2.2. BVDV-2 infection of MDBK cells

MDBK cells were seeded into 175 cm^2 cell culture flasks. One day after seeding, the cells were inoculated with MEM supplemented by 10% FCS containing either infectious (2 $\log_{10}$ $TCID_{50}$/ml medium) or gamma-irradiated, non-infectious BVDV-2 (4.64 $\log_{10}$ $TCID_{50}$/ml medium before irradiation) with 15 ml of the inoculum transferred to each flask. Negative control cells were inoculated with the same volume of cell culture medium. After an incubation period of 2 h ± 15 min, the inoculum was removed from cultures infected with infectious BVDV-2. The cells were then washed once with Dulbecco's PBS and fresh cell

culture medium was added. No medium exchange was performed for cultures inoculated with gamma-irradiated BVDV-2 or the negative control. The cultures were maintained for up to six days. Samples were taken directly after the incubation period (day 0) and subsequently after 2, 4, and 6 d.p.i. with the cell culture supernatant completely removed and stored at −80 °C. The cells were then washed once with Dulbecco's PBS prior to detachment with Accutase. The cells were counted, and pellets of 2×10^6 cells/pellet were obtained by centrifugation at 500 to 550 × *g* for 3 min. The supernatant was then discarded, and the cell pellets were stored at −80 °C. Two cell pellets were prepared for each culture. All samples were then processed using BVDV-2 RT-PCR. Additionally, one sample of gamma-irradiated FCS used for cultivation of MDBK cells and a sample of gamma-irradiated non-infectious BVDV-2 were prepared.

2.3. BVDV-2 detection by indirect immunofluorescence staining

MDBK cells were seeded into 96-well microtiter plates one day prior to inoculation. Serial 3-fold dilutions of the BVDV-2 virus stock in MDBK cell culture medium were done on the day of infection. One hundred µl of virus dilutions were added to MDBK cells in eight wells each. The cultures were further incubated for 4 days at 37 °C and 5% CO2. Cells were washed once with Dulbecco' PBS and then fixed by adding methanol (−20 °C) and incubated for 10 min at room temperature. Cells were washed twice with Dulbecco's PBS and the cells were incubated with BVDV Direct FA Conjugate (VMRD, Inc., USA) for 30 min at 37 °C. The antibody solution was removed, and the cells were washed with FA Rinse Buffer (VMRD, Inc., USA). A caprine Anti-Porcine IgG FITC conjugate (VMRD, Inc., USA) was added to the cells. After an incubation period of 30 min at 37 °C the antibody solution was removed, and the cells were washed with FA Rinse Buffer. Dulbecco's PBS was added to cells and the immunostaining was evaluated by fluorescence microscopy at a 40-200x magnification with a Zeiss Axiovert 40 CFL (Zeiss, Jena, Germany) microscope with a 450–490 nm filter. BVDV-positive cells were identified by green fluorescence in the cytoplasm. The $TCID_{50}$ titer was calculated using the Spearman and Karber formula. The titer of the virus stock was the mean $TCID_{50}$ titer of three titrations.

2.4. BVDV-2 detection by RT-PCR

A quantitative RT-PCR was used to detect BVDV-2 RNA in cell pellets for the negative control at 6 d.p.i., MDBK cells inoculated with infectious BVDV-2 and MDBK cells inoculated with gamma-irradiated non-infectious BVDV-2 (0, 2, 6 d.p.i.). The analyses were performed by the Institute of Virology, Center for Infectious Diseases at the University of Veterinary Medicine Hannover (Hannover, Germany). SYBR® Green based PCR was used to detect a fragment within the 5′ non-translated region (5′NTR) of Pestiviruses using Quanti Tect SYBR® Green RT-PCR Kit (Qiagen, Germany). *In vitro* transcribed GFP plasmids were added to the samples prior to sample preparation in order to check the quality of RNA extraction using GFP specific primers. Total RNA was extracted using RNeasy (Qiagen, Hilden, Germany). RNA concentration was determined using NanoDrop. Typically, 850 ng of RNA were used per reaction for each analysis. The q(RT)-PCR was shown to specifically detect BVDV-1 and -2, *Border disease virus*, *Classical swine fever virus* and Hobi-like viruses (also known as BVDV-3) [13].

2.5. Low-dose infection of MDBK cells and comparison with MDBK culture in the presence of high dose of inactivated BVDV-2

Our objective was to assess the capability of NGS to differentiate a true viral cell infection from carry-over of non-infectious viral sequences and to evaluate the capacity of NGS to detect a low level of infected cells in comparison with a well-established RT-PCR assay. We prepared samples from MDBK cells cultured up to 6 days in the presence of a high dose of BVDV-2 inactivated by gamma-irradiation (initial titer of 4.64 $\log_{10}$ $TCID_{50}$/ml) to reproduce a medium containing gamma-irradiated FCS contaminated with BVDV-2. A previously published risk analysis estimated the BVDV titer in batches of bovine fetal serum submitted to gamma-irradiation could be approximately 4.7 $\log_{10}$ $TCID_{50}$/ml [14]. We also prepared samples from a kinetic of infection (day 0 to day 6) of MDBK cells with a low dose of infectious BVDV-2 (2 $\log_{10}$ $TCID_{50}$/ml) in order to mimic cells infected just before sampling by residual replication-competent BVDV from poorly inactivated FCS. We tested all samples by BVDV-2 RT-PCR and results were used to select samples that were informative enough to be tested by NGS.

2.6. Detection of HHV-4 in B95-8 persistently infected cells

In order to assess the performance of RNA-Seq NGS to detect a low level of persistently infected cells and to compare this with an established specific PCR used for HHV-4 detection, HHV-4 infected, non-induced, B95-8 cells were serially diluted 10-fold up to $1{:}10^8$ into an HHV-4-negative human lymphoid cell line (Ramos). These cells were then seeded into 175 cm^2 cell culture flasks at a cell density of 3×10^5 cells/ml. Ramos cells were seeded into 175 cm^2 flasks at a cell density of 2×10^5 cells/ml. One culture of B95-8 cells was treated with TPA and served as a positive control. The cells were maintained under standard cell culture conditions for 48 h.

The cells were counted using a haemocytometer and the cell densities were adjusted with cell culture medium to prepare B95-8 cells serially diluted $1{:}10^3$ to $1{:}10^8$ with Ramos cells with 2×10^6 cells. Ramos cells, B95-8 cells, TPA treated B95-8 cells and TPA treated B95-8 cells mixed $1{:}10^4$ with Ramos cells were included as controls. The cells of all samples were pelleted by centrifugation and the supernatant was discarded. Cell pellets were stored at −80 °C. B95-8 cells cultured with TPA, undiluted or diluted to $1{:}10^4$ in Ramos cells were used as positive controls. For each sample, independent DNA and RNA extractions was performed. HHV-4 DNA genome copy number was determined by PCR and RNA was submitted to RNA-Seq.

2.7. HHV-4 detection by PCR

HHV-4 PCR was carried out using the RealStar® EBV PCR Kit 1.0 (Altona Diagnostics GmbH, Hamburg, Germany) by the Medizinisches Versorgungszentrum (Dr. Eberhard & Partner, Dortmund, Germany).

2.8. Infectivity assay for Murine leukemia viruses

Ramos cells were tested for the presence of infectious xenotropic Murine leukemia viruses (MuLV) by an S^+L^- assay using the feline PG-4 cell line. PG-4 cells grown in 25 cm^2 cell culture flasks were inoculated with the supernatant of Ramos cells one day after seeding. The medium of the PG-4 cells was removed and the cells were inoculated with 1 ml sample per flask. Each sample was supplemented with polybrene (Sigma, Germany) up to a final concentration of 10 µg/ml. The xenotropic MuLV strain pNFS Th-1 (ATCC VR-1447) was used as a control virus. Positive controls with 1.5 and 15 $TCID_{50}$/mL were included as well as interference controls consisting of the sample spiked with the two doses of the control virus. The negative control was inoculated with 1 ml cell culture medium containing polybrene. The inoculated cultures were incubated for 1.5–2 h in culture conditions. The inoculum was then removed, and the cultures were further incubated using standard cell culture conditions. All cultures were passaged once at 4 d.p.i.

2.9. Sample extraction and next generation sequencing

Total RNA was extracted from 2 million cells using the RNEasy mini kit (Qiagen, Hilden, Germany) and on-column DNAse digestion. A

stranded RNA-Seq library was performed from 10 ng of total RNA using the SMARTer Stranded Total RNA-Seq – Pico Input Mammalian (Clontech, Mountain View USA). Using this approach, the enzyme's terminal transferase activity adds a few non-templated nucleotides to the 3′ end of the cDNA when the RT reaches the 5′ end of the RNA fragment. A primer included in the kit then pairs with the non-templated nucleotides, creating an extended template to enable the RT to continue replicating to the end of the oligonucleotide. The resulting cDNA contained sequences derived from the random primer linked to constant sequence. This sequence is then targeted by PCR enabling the incorporation of Illumina adaptors and indexes in an oriented sense defined by the PCR primers. Libraries were then multiplexed by four in equimolar ratio and loaded onto NextSeq High Output kit v2 (Illumina, San Diego, USA) before running for 150 cycles in single reads.

2.10. Initial agnostic bioinformatics analysis

Raw data were first sorted to suppress or cut duplicates, low quality reads and homo-polymers (Dupli_Qual_Filter version 1.2.0, PathoQuest, Paris, France). Sequences introduced during Illumina library preparations (adapters, primers) were removed with Skewer version 0.1.114 [15]. Reads aligned to the host genome with the mapping software Burrows-Wheeler Aligner (BWA) [16] (human genome - Reference GRCh37/hg19 for Ramos and B95-8 cells, bovine genome (Reference BosTau7) for MDBK cells) or to bacterial rRNA were discarded. Host genomes were downloaded from the UCSC Genome Browser [17]. The bacterial rRNA database was downloaded from the EMBL-EBI ENA rRNA database (https://rnacentral.org/) followed by in-house sequence cleaning and clustering (RNAdbmaker version 3.0.0, PathoQuest, Paris, France). A *de novo* assembly step was performed with CLC Assembly Cell solution (https://www.qiagenbioinformatics.com/) on the set of filtered reads. Resulting contigs and non-assembled reads (singletons) were aligned (BLAST alignment [18]) using viral and comprehensive databases downloaded during November 2017 from the EMBL-EBI nucleotide sequence database (https://www.ebi.ac.uk/). Blastbankmaker version 2.2.0 (PathoQuest, Paris, France) was developed to remove duplication and low confidence sequences (too short, multiple taxonomies, low-quality associated keywords …). Hits were then aligned on a comprehensive nucleotide database (https://www.ebi.ac.uk/) and were considered for further analysis if their best hit was still a viral taxonomy. In the absence of any viral nucleotide hit and to identify more distant viral hits, contigs were similarly aligned successively on protein viral and comprehensive databases which downloaded during November 2017 from the Uniref100 database [19]. The Uniref100 database is already non-redundant but a taxonomic cleaning process was performed to produce the final databases.

For BVDV-2 identification, all contigs and singletons assigned to the *Pestivirus* genus were reprocessed to validate the BVDV-2 species taxonomic assignation and reject false assignations due to cellular sequences, such as ubiquitin, inserted in genomes of pestiviruses [20]. *Pestivirus* member hits were realigned on the exhaustive nucleotide database and all taxonomies aligned with a high score (more than 85% of best score) were considered. Hits were considered validated if BVDV-2 was listed among these taxonomies in the absence of any non-viral taxonomy.

2.11. Stranded targeted bioinformatics analysis

The targeted and stranded analysis was performed for viruses identified using the agnostic approach as shown in sections 3.1 to 3.3. The reference genomes used as references for mapping were: HHV-4: NC_007605.1, MuLV: AF221065.1, BVDV-2: KT832818.1. This analysis enabled a more stringent mapping alignment of filtered reads. This alignment provided a detailed horizontal genome coverage and depth profile. Local alignments were performed with BWA [16]. The RNA strand information was retained since the sample libraries were performed using the SMARTer Stranded RNA-Seq Kit (see section 2.9).

2.12. Splicing analysis

When MuLV sequences were detected (see results), a search for splicing events was conducted to identify reads specific for the junctions generated during the splicing of *env* gene mature transcripts. In the edited sequence, the splicing donor site (nt 204 gggaggtaagct) is joined to a splicing acceptor site (nt 5502 cttacaggccct) to give the sequence gggaggccct [21]. Comparison with the number of reads mapped to the non-edited sequence from the nascent precursor messenger RNA enabled the quantification of the edition rate. To identify these reads, we utilized a proprietary script (PathoQuest, Paris, France) to search for stringent (no gap) DNA motives.

3. Results

3.1. Performance of NGS and RT-PCR for detection of BVDV-2 infected cells compared to cells put in contact with high dose of inactivated BVDV-2

Samples from MDBK cells cultured in the presence of a high dose of BVDV-2 inactivated by gamma-irradiation were compared to samples from a kinetic of infection of MDBK cells with a low dose of infectious BVDV-2. No viral sequences other than BVDV-2 were detected in any sample using agnostic RNA-Seq. MDBK cells cultured in the absence of BVDV-2 spike showed a very low RT-PCR signal and 2 NGS reads (see

Table 1
RT-PCR and agnostic RNA-Seq results of MDBK cells cultured with inactivated BVDV-2 or infected with low dose BVDV-2.

Sample description	RT-PCR BVDV-2	Next Generation Sequencing			
	copies/reaction (mean)	Total number of reads (per library)	Number of reads post filtration	Number of BVDV-2 reads	Genome coverage (%)
MDBK cells	0.91	166 563 633	9 428 029	2	2.4
Infection with low dose BVDV-2 - day 0	2.16	N/A	N/A	N/A	N/A
day 2	1 591 000	176 004 808	10 262 347	187 710	100
day 4	7 709 500	178 269 826	10 646 679	1 230 533	100
day 6	3 746 500	N/A	N/A	N/A	N/A
MDBK cells	0.91	N/A	N/A	N/A	N/A
Culture with inactivated BVDV-2 - day 0	27 700	163 603 920	9 960 735	499	81.6
day 2	17.25	N/A	N/A	N/A	N/A
day 4	13.04	N/A	N/A	N/A	N/A
day 6	7.11	N/A	N/A	N/A	N/A

N/A: not available.

Table 1) at day 0. In our experience this is a frequent finding for cells previously cultivated in presence of inactivated FCS. Cells cultured with a high dose of gamma-irradiated BVDV-2 gave a higher RT-PCR and NGS signals at day 0 which decreased on the following days when tested by RT-PCR. Following MDBK infection with a low dose of infectious BVDV-2, the initial signal detected by RT-PCR at day 0 was strongly amplified in the following time points tested (day 2, day 4 and day 6) by RT-PCR and was also detected by NGS (only day 2 and day 4 tested). This demonstrated that RT-PCR targeted to the viral genome and agnostic transcriptome analysis similarly detected BVDV-2 at the earlier time point tested after infection. We observed a strong quantitative difference in both techniques between cells infected with small amounts of infectious BVDV and those cultured with high amounts of inactivated BVDV-2. At this stage, results with NGS or with RT-PCR at any individual time point were suitable to conclude if the detected signal accounted for a replicating virus.

To increase the specificity of detection for actively replicating viruses, we refined our RNA-Seq approach in order to identify viral RNA molecules produced during cell infection. BVDV-2 is a positive single stranded RNA virus that produces an antigenome intermediate (i.e. negative sense RNA) in cells, as a matrix of transcription and genome replication. Hence, the detection of BVDV-2 negative sense RNA in cells is evidence of ongoing viral replication. Since our read library preparation method comprised a differential tagging system according to the read polarity, we mapped the reads from raw data to the closest BVDV-2 reference genome identified by blast analysis and retrieved information about their polarity (Table 2). While no antigenomic reads could be found in MDBK cells cultured with inactivated BVDV-2, these antigenomic reads covered 48.6% (day 2) to 82.5% (day 4) of the BVDV-2 genome in infected MDBK cells. This differentiation could not be made with standard RT-PCR and therefore would not enable the ability to distinguish positive and negative strands.

This data demonstrated that both RNA-seq and RT-PCR were able to detect viral infection at the first time point tested (day 2). Additionally, the use of stranded RNA-Seq differentiated BVDV-2 infection from viral nucleic acid carry-over.

3.2. Detection of HHV-4 in B95-8 persistently infected cells

HHV-4 was selected as a second challenging virus model because it can persist in a state close to latency in the lymphoid cotton-top tamarin monkey B95-8 cells. Our objective was to test the selective identification of cell infection with this DNA virus through viral RNA transcriptome detection, which is not feasible using PCR targeted to genomic DNA. A secondary objective was to evaluate the performance of persistently infected cells in a background of non-infected cells. The Ramos cell line was selected as diluting cells as it matched the lymphoid B lineage of B95-8, with the objective to be close to the transcriptomic background.

As expected, no HHV-4 sequences were found in the Ramos cell line by PCR or NGS (Table 3). HHV-4 DNA and RNA transcripts of HHV-4 were detected at the same dilution (1 persistently B95-8 cell into 10^5 Ramos cells) respectively by HHV-4 PCR and agnostic RNA-Seq. At this very low ratio of infected to virus-free cells, RNA-Seq covered 2060 nt (0.9%) of the genome, which enabled unambiguous identification of HHV-4. At a ratio of one B95-8 to 10^4 Ramos cells, the horizontal coverage strongly increased (13400 nt, 7.3% of the genome). Furthermore, when cells were induced with TPA the number of HHV-4 DNA cp (PCR) or reads (RNA-Seq) increased by a factor in the range of 7–8 (see dilution $1/10^4$ with or without TPA), thus increasing the detection of productively infected cells. Identification of transcripts from this DNA virus was direct evidence of cell infection. In addition, we were able to identify reads mapping to both strands of HHV-4 DNA with a proportion different from 50% (Table 4). These results exclude the possibility of contaminating DNA being the source of the reads (which should lead to a 1/1 ratio relative to each sense). These results are consistent with the known bidirectional transcription pattern of HHV-4, as both strands are highly transcribed during viral replication [22].

Our RNA-Seq results strongly indicated that the B95-8 cells were infected with HHV-4. RNA-Seq was also able to identify cell infection over a very high background of uninfected Ramos cells. As shown in Table 3, sequences for HHV-4, MuLV and BVDV-3 were found in several samples. We considered the hypothesis of productive BVDV-3 infection of Ramos cells highly unlikely due to the very low number of reads which also precluded significant RNA strand analysis. We also excluded any BVDV-3 infection of the B95-8 cells as there was no correlation between the number of BVDV-3 reads with the dilution of the cells in the range of 10^{-4} to 10^{-6}. We concluded that these reads likely corresponded to carryover of non-infectious BVDV-3 RNA from the FCS used as cell reagent and focused our analysis on MuLV instead.

3.3. Detection of MuLV in Ramos cells

We found 77 309 to 190 007 reads corresponding to MuLV in the Ramos cells and in every tested sample using these cells. The reads showed 99% nt identity with the strain DG-75, a xenotropic MuLV (Table 3). All MuLV open reading frames were intact, encoding full-length proteins without any stop codon. Mapping of the reads onto the DG-75 genome demonstrated that 99.5% of the reads corresponded to the positive strand and covered the whole genome (e.g., corresponding to the genomic RNA and/or to the transcripts). This demonstrated the presence of retroviral RNA and enabled us to differentiate this from the contamination of the RNA library with proviral DNA (which would have given a 1:1 ratio between reads relative to the sense).

Our results were not sufficient to demonstrate MuLV expression since genomic RNA and transcripts have the same polarity. We therefore looked for splicing events specific for mature RNA transcripts produced during cell infection and focused on the *env* gene mRNA (Fig. 1A). Splicing events were detected in 38% of the reads at the *env* splicing site with 62% of the reads corresponding to genomic or nascent RNA. The demonstration of a functional MuLV genome (full length genome coverage with intact ORFs) and direct evidence for active transcription (presence of spliced transcripts) pointed to the likely contamination of cells with an infectious virus. A xenotropic retrovirus infectivity assay was conducted by infection of S^+L^- PG-4 cells as indicator cells with Ramos cell culture supernatant. A strong cytopathic effect was evidenced after one week of incubation (Fig. 1B). Therefore, the retrovirus infectivity assay confirmed productive infection of the Ramos cells by a xenotropic MuLV, as already strongly suggested by NGS data.

Table 2
Targeted stranded BVDV-2 analysis: number of positive and negative BVDV-2 reads and genome horizontal coverage.

Sample description	Number of BVDV-2 reads[a] genomic strand (genome coverage)	Number of BVDV-2 reads[a] antigenomic strand (genome coverage)
MDBK cells	0 (0%)	0 (0%)
Infection with low dose BVDV-2 - day 2	184 355 (99.9%)	366 (48.6%)
day 4	1 196 781 (99.9%)	2427 (82.5%)
Culture with inactivated BVDV-2 - day 0	437 (78.5%)	0 (0%)

[a] Reads mapped on BVDV-2: reference KT832818.1.

Table 3
RT-PCR and NGS agnostic analysis results of Ramos and B95-8 infected HHV-4 cells.

Sample description	PCR HHV-4	Next Generation Sequencing				
	Mean number of copies/ reaction	Total number of reads (per library)	Number of reads post filtration	Number of HHV-4 reads	Number of MuLV reads	Number of BVDV-3 reads
Ramos (negative control)	negative	164 604 917	6 014 833	0	190 007	0
B95-8/Ramos [$(1/10^3)$]	24 691	N/A	N/A	N/A	N/A	N/A
B95-8/Ramos [$(1/10^4)$]	1604	169 304 831	6 362 936	317	77 309	3
B95-8/Ramos [$(1/10^5)$]	802	168 907 048	6 531 975	31	104 987	0
B95-8/Ramos [$(1/10^6)$]	negative	168 718 801	6 588 113	0	89 540	5
B95-8/Ramos [$(1/10^7)$]	negative	N/A	N/A	N/A	N/A	N/A
B95-8/Ramos [$(1/10^8)$]	negative	N/A	N/A	N/A	N/A	N/A
B95-8 + TPA/Ramos [$(1/10^4)$]	11 234	161 142 284	6 323 478	2657	143 264	2
B95-8 (positive control)	11 049 382	N/A	N/A	N/A	N/A	N/A
B95-8 + TPA (positive control)	60 432 098	N/A	N/A	N/A	N/A	N/A

Table 4
Targeted stranded HHV-4 analysis: number of HHV-4 reads and genome horizontal coverage.

Sample description	Number of HHV-4 reads[a] positive strand (genome coverage)	Number of HHV-4 reads[a] negative strand (genome coverage)
B95-8/Ramos [$(1/10^4)$]	97 (3.6%)	211 (3.7%)
B95-8/Ramos [$(1/10^5)$]	17 (0.7%)	2 (0.2%)
B95-8 + TPA/Ramos [$(1/10^4)$]	591 (17.2%)	2137 (19.6%)

[a] Reads mapped on HHV-4, reference NC_007605.1.

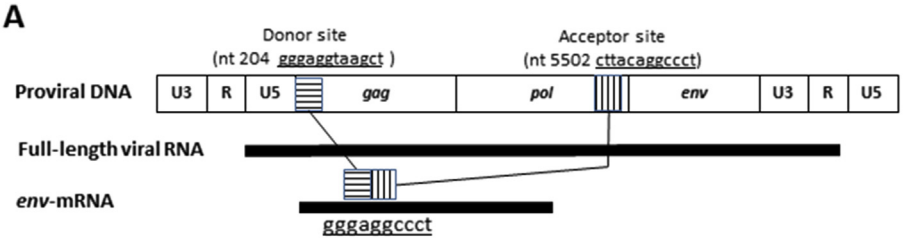
4. Discussion

NGS is a powerful agnostic method which is widely used for virus discovery. Its utility for the testing of biologicals has been demonstrated by the finding of viral contamination of a marketed live vaccine which was not identified via compendial controls for adventitious virus detection [3] and of a production cell substrate which was previously assumed to be virus free [4]. The widespread use of NGS has nevertheless been hampered by the fear of finding nonrelevant hits which correspond to inert viral sequences or non-infectious viruses. Our objective was to develop a NGS strategy which adds specificity to the detection of cell infection in addition to the sensitivity and broad range of detection by NGS.

For the current study, we focused on testing of cell substrates since they are a main target of viral testing when used for the production of biologicals. We reasoned that focusing on nucleic acids that indicate productive or latent cell infection was a more productive approach to search for viral genomes as opposed to deep sequencing all cellular nucleic acids (DNA and RNA). As a result, we elected to selectively sequence viral RNAs produced during viral replication cycles and use these as biomarkers of virus replication and latency [23,24].

Viral RNAs are *per se* indicative of transcription for DNA viruses. Transcripts have been shown to be incorporated into virions of some enveloped dsDNA viruses such as herpesviruses [25,26], poxviruses [27], and mimiviruses [28]. Nevertheless, these RNAs correspond to a subset of genes and at low concentration are identifiable in highly concentrated virus stocks and cannot be mistaken for the wide and wide range of viral transcription products found in infected cells.

For RNA viruses, non-encapsidated RNAs like transcripts and/or intermediates of replication are unlikely to be found in raw materials or



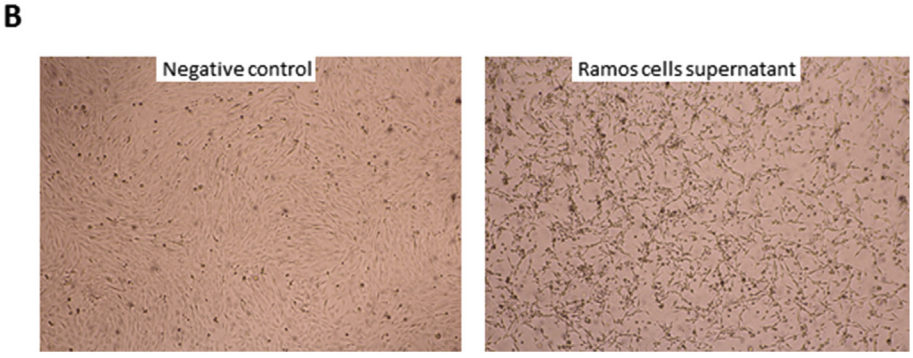


Fig. 1. Follow-up studies of MuLV sequences found in Ramos cells: A) analysis of *env* gene mRNA splicing: the sequences of the donor and acceptor sites together with the resulting sequence of the splice junction of the subgenomic mRNAs, which was identified in the sequencing reads, are shown; B) infectivity assay in PG-4 S^+L^- cells (magnification 40x) showing the appearance of a cytopathic effect for cells put in contact with Ramos cells supernatant.

reagents because of the large presence of RNAses. This enables these to be utilized as markers of infection. It is essential to differentiate these from genomic RNAs which are protected from degradation by capsids, even in inactivated viruses, and can therefore be a source of positive results not linked to viral contamination of molecular tests, including RT-PCR. This is typically the case for bovine viruses like BVDV which are frequently found in gamma-irradiated FCS.

In cells infected by a ssRNA virus, the antigenomic strand can be used as a marker. Nevertheless, for some (−)ssRNA viruses, like for members of the *Paramyxoviridae* and *Rhabdoviridae*, full-length positive strands are also present in the virion [29–31] but the ratio of genomic/antigenomic reads in transcribed versus non-transcribed loci is evidence for viral expression (our unpublished data). Additionally, antigenomes are not packaged in the virion for (+)ssRNA viruses and the antigenomic strand accumulates in low to very low amounts in infected cell. For example, this accumulation is in the range of 0.4%–10% for members of the *Flaviviridae* [32–34] and *Coronaviridae* [35,36]. Ultra-deep sequencing is necessary to identify this low amount of negative sense RNA.

Conversely, the positive strand for dsRNA viruses is found in much higher quantities in infected cells. Additionally, retroviruses are transcribed from a proviral DNA and full length transcripts are not differentiable from genomic RNA. Nevertheless, subgenomic RNAs used as mRNAs are generated by splicing events [37] and junctions between donor and acceptor sites can be detected to demonstrate expression of viral genes as we have reported here.

To demonstrate our approach we selected two types of viruses [dsDNA and (+)ssRNA] under conditions of a very low level of cell infection (e.g., highly diluted persistently infected HHV-4 cells or recent infection at a very low multiplicity of infection with BVDV-2). The results were compared to those of genomic RT-PCR or PCR assays which are routinely used for testing cell lines. The analytical sensitivity of RNA-Seq was identical to PCR when based on the lowest detectable concentration for HHV-4 (one HHV-4-persistently infected cell in 10^5 non-infected cells) and on the first time point detected (2 d.p.i.) for BVDV infection. The efficacy of detection, as measured by the percentage of positively identified infected samples, depends not only on the analytical sensitivity, but also on the range of viruses that can be detected. This is indeed an advantage of NGS over PCR since we have demonstrated this enabled the detection of the unexpected expression of a functional clone of MuLV in the human lymphoid Ramos cell line we utilized for our study. To increase the specificity of our approach, specifically the percentage of virus-free samples identified as negative, we compared our RNA-Seq results with the known transcriptome pattern of the suspected virus and were able to identify the likelihood of viral infection. Interestingly, for our unexpected finding of MuLV in Ramos cells, the analysis of the splicing events together with the demonstration of intact MuLV open reading frames strongly suggested retroviral expression, which we confirmed by a positive infectivity assay. This example illustrates the potential benefits of agnostic testing followed by in depth analysis of the transcriptome on the same dataset. Several human lymphoid cell lines have been shown to be contaminated with MuLV, most likely a result of initial grafting of the cells into mice [38]. While some Ramos cell lines were found MuLV-free (DSMZ Ref 603 [38], ATCC [38]), Ramos cells from other sources were found MulV-positive [39], strongly suggesting a contamination after the establisment of the cell line. The Ramos cells we utilized for the present study were amplified from cells obtained from the ECACC.

Our new NGS approach and the results we report are based on the use of two separate methods being used in conjunction with each other. The first method is the use of transcriptome instead of whole nucleic acid analysis, and the second is the use of a RNA-Seq method that maintains the link between the sequence and the mother strand. The information obtained using our approach is not accessible using RT-PCR since random priming for the reverse transcription typically targets only the genomic strand and results in the loss of the link to the mother strand.

Despite the strengths that are associated with our method, there are some potential drawbacks. For (+)ssRNA viruses, antigenomes are present in low amount and require ultra-deep sequencing to be identified. As mentioned above, full-length antigenomes are also present in capsids for some (−)ssRNA viruses resulting in the need for additional targeted bioinformatic analysis to differentiate these from subgenomic transcripts which are a potential sign of infection. Also, purified virus stocks must be used for validation in order to avoid the presence of cell-free RNA in virus stocks: according to good practices, virus stocks are often frozen immediately after harvest, which preserves cell-free RNAs. These cell-free RNAs could be mistaken for biomarkers of replication present in the spiked cell line subject to testing.

We have demonstrated in the present study that NGS can be successfully designed to specifically detect infected cells with a sensitivity similar to that of (RT-)PCR. Based on our results, we believe this new NGS approach is a valuable alternative testing option which can be utilized during the development and production of biotherapeutics. We also envision its potential use for the testing of new cell substrates, cell banks, cell culture intermediates, and cell therapies and for investigating suspected contamination of biologics. NGS analysis can be conducted following Good Laboratory Practices, which specifically require prior validation to guarantee raw data integrity and archiving [40]. It is also possible as a further step to conduct analysis in a GMP environment by implementing required validation methods relying on the guidelines of the pharmaceutical industry.

Conflicts of interest

KB, RZ, AM, HR are members of Charles River Laboratories. EM, JC, PB, JMC are members of PathoQuest. EM, JC, ME holds shares of PathoQuest.

Funding

This work was supported by LFB.

Role of the funding source

The sponsor of the study was LFB. Members of its staff (AB, BF) participated in the design of the study, in the writing of the manuscript, and in the decision to its submission for publication.

Acknowledgments

The authors would like to thank Lydia Zimmermanns and Florian Steffens from Charles River Laboratories Germany for their dedicated technical assistance.

References

[1] 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. https://doi.org/10.1016/j.vaccine.2014.02.021.

[2] Marcus-Sekura C, Richardson JC, Harston RK, Sane N, Sheets RL. Evaluation of the human host range of bovine and porcine viruses that may contaminate bovine serum and porcine trypsin used in the manufacture of biological products. Biologicals 2011;39:359–69. https://doi.org/10.1016/j.biologicals.2011.08.003.

[3] 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. https://doi.org/10.1128/JVI.02690-09.

[4] 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. https://doi.org/10.1128/JVI.00780-14.

[5] Chiu CY. Viral pathogen discovery. Curr Opin Microbiol 2013;16:468–78. https://doi.org/10.1016/j.mib.2013.05.001.

[6] European Pharmacopoeia. General chapters, 5.2.14, Substitution of in vivo method (s) by in vitro method(s) for the quality control of vaccines. 2018.

[7] European Pharmacopoeia General chapters, 5.2.3. Cell substrates for production of vaccines for human use. 2018.
[8] WHO expert committee on biological standardization: sixty-eighth report. Geneva: World Health Organization; 2018. (WHO Technical Report Series, No. 1011). 2018.
[9] Pharmacopea European. Genneral chapters, 2.6.16. Tests for extraneous agents in viral vaccines for human use. 2018.
[10] Tsurumi T, Fujita M, Kudoh A. Latent and lytic Epstein-Barr virus replication strategies. Rev Med Virol 2005;15:3–15. https://doi.org/10.1002/rmv.441.
[11] Miller G, El-Guindy A, Countryman J, Ye J, Gradoville L. Lytic cycle switches of oncogenic human gammaherpesviruses. Adv Cancer Res 2007;97:81–109. https://doi.org/10.1016/S0065-230X(06)97004-3.
[12] Shaw JE, Petit RG, Leung K. Growth of B95-8 cells and expression of Epstein-Barr virus lytic phase in serum-free medium. J Virol 1987;61:4033–7.
[13] Becher P, Orlich M, Thiel H-J. Complete genomic sequence of border disease virus, a Pestivirus from sheep. J Virol 1998;72:5165–73.
[14] Scientific Committee on Animal Health and Animal Welfare. Virus inactivation in bovine blood and blood products. European Commission; 2000. Sanco/C3/AH/R21/2000 https://ec.europa.eu/food/sites/food/files/safety/docs/sci-com_scah_out50_en.pdf.
[15] Jiang H, Lei R, Ding S-W, Zhu S. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. BMC Bioinf 2014;15:182. https://doi.org/10.1186/1471-2105-15-182.
[16] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinforma Oxf Engl 2009;25:1754–60. https://doi.org/10.1093/bioinformatics/btp324.
[17] Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. Genome Res 2002;12:996–1006. https://doi.org/10.1101/gr.229102.
[18] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215:403–10. https://doi.org/10.1016/S0022-2836(05)80360-2.
[19] Suzek BE, Wang Y, Huang H, McGarvey PB, Wu CH. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. Bioinformatics 2015;31:926–32. https://doi.org/10.1093/bioinformatics/btu739.
[20] Meyers G, Tautz N, Dubovi EJ, Thiel HJ. Viral cytopathogenicity correlated with integration of ubiquitin-coding sequences. Virology 1991;180:602–16.
[21] Raisch KP, Pizzato M, Sun HY, Takeuchi Y, Cashdollar LW, Grossberg SE. Molecular cloning, complete sequence, and biological characterization of a xenotropic murine leukemia virus constitutively released from the human B-lymphoblastoid cell line DG-75. Virology 2003;308:83–91.
[22] O'Grady T, Cao S, Strong MJ, Concha M, Wang X, Splinter Bondurant S, et al. Global bidirectional transcription of the Epstein-Barr virus genome during reactivation. J Virol 2014;88:1604–16. https://doi.org/10.1128/JVI.02989-13.
[23] Zhang X, Ma X, Jing S, Zhang H, Zhang Y. Non-coding RNAs and retroviruses. Retrovirology 2018;15:20. https://doi.org/10.1186/s12977-018-0403-8.
[24] Knipe DM, Raja P, Lee J. Viral gene products actively promote latent infection by epigenetic silencing mechanisms. Curr Opin Virol 2017;23:68–74. https://doi.org/10.1016/j.coviro.2017.03.010.
[25] Terhune SS, Schröer J, Shenk T. RNAs are packaged into human cytomegalovirus virions in proportion to their intracellular concentration. J Virol 2004;78:10390–8. https://doi.org/10.1128/JVI.78.19.10390-10398.2004.
[26] Sciortino MT, Suzuki M, Taddeo B, Roizman B. RNAs extracted from herpes simplex virus 1 virions: apparent selectivity of viral but not cellular RNAs packaged in virions. J Virol 2001;75:8105–16.
[27] Grossegesse M, Doellinger J, Haldemann B, Schaade L, Nitsche A. A next-generation sequencing approach uncovers viral transcripts incorporated in poxvirus virions. Viruses 2017;9. https://doi.org/10.3390/v9100296.
[28] Raoult D, Audic S, Robert C, Abergel C, Renesto P, Ogata H, et al. The 1.2-megabase genome sequence of Mimivirus. Science 2004;306:1344–50. https://doi.org/10.1126/science.1101485.
[29] Plumet S, Duprex WP, Gerlier D. Dynamics of viral RNA synthesis during measles virus infection. J Virol 2005;79:6900–8. https://doi.org/10.1128/JVI.79.11.6900-6908.2005.
[30] Kolakofsky D, Bruschi A. Antigenomes in Sendai virions and Sendai virus-infected cells. Virology 1975;66:185–91.
[31] Finke S, Conzelmann KK. Ambisense gene expression from recombinant rabies virus: random packaging of positive- and negative-strand ribonucleoprotein complexes into rabies virions. J Virol 1997;71:7281–8.
[32] Fan Y-H, Nadar M, Chen C-C, Weng C-C, Lin Y-T, Chang R-Y. Small noncoding RNA modulates Japanese encephalitis virus replication and translation in trans. Virol J 2011;8:492. https://doi.org/10.1186/1743-422X-8-492.
[33] Raquin V, Lambrechts L. Dengue virus replicates and accumulates in Aedes aegypti salivary glands. Virology 2017;507:75–81. https://doi.org/10.1016/j.virol.2017.04.009.
[34] Kreil TR, Zimmermann K, Burger I, Attakpah E, Mannhalter JW, Eibl MM. Detection of tick-borne encephalitis virus by sample transfer, plaque assay and strand-specific reverse transcriptase polymerase chain reaction: what do we detect? J Virol Methods 1997;68:1–8.
[35] Perlman S, Ries D, Bolger E, Chang LJ, Stoltzfus CM. MHV nucleocapsid synthesis in the presence of cycloheximide and accumulation of negative strand MHV RNA. Virus Res 1986;6:261–72.
[36] Sethna PB, Hung SL, Brian DA. Coronavirus subgenomic minus-strand RNAs and the potential for mRNA replicons. Proc Natl Acad Sci U S A 1989;86:5626–30.
[37] Leblanc J, Weil J, Beemon K. Posttranscriptional regulation of retroviral gene expression: primary RNA transcripts play three roles as pre-mRNA, mRNA, and genomic RNA vol. 4. Wiley Interdiscip Rev RNA; 2013. p. 567–80. https://doi.org/10.1002/wrna.1179.
[38] Uphoff CC, Lange S, Denkmann SA, Garritsen HSP, Drexler HG. Prevalence and characterization of murine leukemia virus contamination in human cell lines. PLoS One 2015;10:e0125622. https://doi.org/10.1371/journal.pone.0125622.
[39] Lin Z, Puetter A, Coco J, Xu G, Strong MJ, Wang X, et al. Detection of murine leukemia virus in the Epstein-Barr virus-positive human B-cell line JY, using a computational RNA-Seq-based exogenous agent detection pipeline. PARSES. J Virol 2012;86:2970–7. https://doi.org/10.1128/JVI.06717-11.
[40] OECD. Digital security risk management for economic and social prosperity. OECD Recommendation and Companion Document; 2015.