UNIVERSITA' DEGLI STUDI DI TORINO

Ph.D. School "Health and Life Sciences"

Ph.D. Programme "Complex Systems for Life Sciences"

XXXIII Cycle

DEVELOPMENT OF *IN VITRO* ANALYTICAL METHODS FOR BIOSAFETY ASSESSMENT TO SUPPORT THE PRODUCTION OF BIOTECHNOLOGICAL DRUGS

Thesis' author:

Dott. Giorgia Piersigilli

Supervisors:

Simone Olgiati, Ph.D.

Prof. Michele De Bortoli

Ph.D. Programme Co-ordinator :

Prof. Enzo Medico

ACADEMIC YEARS: 2017-2020

Sommario

1 INTRODUCTION

1.1 BIOPHARMACEUTICAL MOLECULES MANUFACTURING

1.1.1 MANUFACTURING PROCESS AND GMP

Biologicals are defined by the International Conference on Harmonization (ICH) Q5D guideline [1] as "*any product prepared from cells cultivated from cell banks with the exception of microbial metabolites such as, for example, antibiotics, amino acids, carbohydrates, and other low molecular weight substances*". Most of biotechnological products approved by regulatory authorities are produced using recombinant DNA technology in various expression systems. In recent years, there has been a rapid growth in the biotechnology industry which is reflected in the availability of more than one hundred biotechnological molecules in several therapeutic areas [2]. Mammalian expression systems are the most common platforms for the manufacturing of biopharmaceuticals as these cell lines produce high quantity of complex proteins with human-like post translational modifications [3].

The entire process for manufacturing of a biotechnological product is divided into two major steps: the upstream and the downstream processing. The upstream processing involves the generation of the cell lines used for recombinant protein production and their cultivation in bioreactors, while the downstream processing includes all production steps from cell harvest to the purification of the final product [4]. The manufacturing process starts with the generation of the Master Cell Bank (MCB), which is usually prepared under defined culture conditions from an initial clone selected for its optimal growth and protein production rate. A stable transfection, with plasmid DNA presenting the sequence of interest, is used to create a recombinant cell line from the selected host cell. Transfected cells are then cultivated multiple times and screened to evaluate cell growth, product titer and quality [5]. These parameters are evaluated to choose the best clone that will be cryopreserved and expanded in order to obtain the MCB and the derived Working Cell Bank (WCB). Prior to be used in the production process WCB productive capacity are tested in laboratory scale. After that,

WCB are cultivated in a large bioreactor to produce the recombinant protein by a fermentation process.

At the end of the upstream process, the bulk harvest, consisting in cells, media (including cell byproducts, such as sugars, proteins, and amino acids), and the protein of interest, is submitted to a purification step performed to remove any impurities and contaminants before the final batch release (Figure 1).

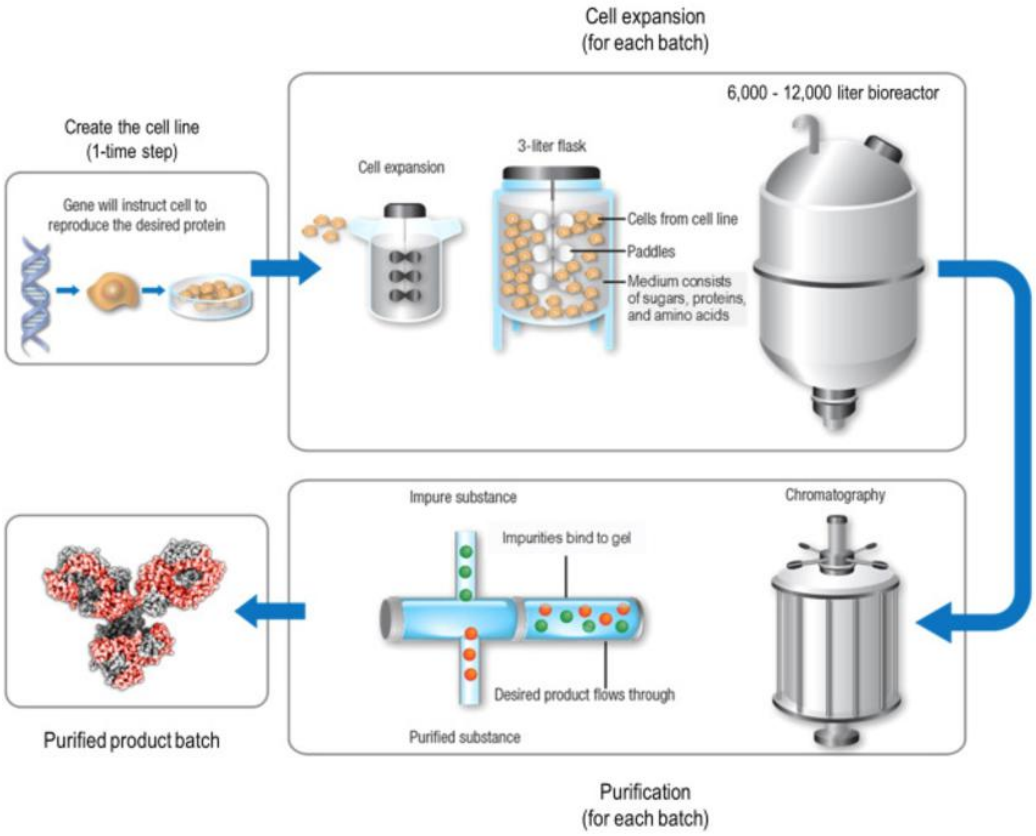


Figure 1. Biopharmaceuticals manufacturing process [6] Manufacturing process for biologics involves three main steps: 1) development of a genetically modified cell line able to express the protein of interest, 2) Cell expansion and scale-up of cell bank from shaker flask culture to bioreactor, 3)protein isolation and purification

Biologicals are processed under strictly controlled conditions required to guarantee the production of a potent and safe product, and in order to avoid the introduction of environmental contaminants [7].

Regulatory authorities require that all biopharmaceuticals must be manufactured according to the current Good Manufacturing Practice (cGMP) guidelines. These

guidelines were redacted by the World Health Organization (WHO) [8] and they are currently adopted by national regulatory bodies as the European Medicines Agency (EMA) or the Food and Drug Administration (FDA). However, The manufacturing processes are constantly under improvement to increase productivity, efficiency, and product quality/safety [9]. Therefore, the GMP system is not static, and instead it adapts and evolves with new technologies and new challenges in biosafety that gradually emerge. Moreover, to harmonize the process development and manufacturing of biotechnologicals, regulatory authorities of different country follow common guidelines released by ICH [10].

1.1.2 ADVENTITIOUS VIRUS SAFETY OF BIOPHARMACEUTICALS

A critical aspect in the development of biotechnological products is the biosafety evaluation of materials and intermediates used in the manufacturing process. This requires a number of stringent tests to demonstrate that a product is free of contamination and that it is safe for patients. For this reason, the 1998 International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use defined measures to guarantee viral safety through testing of cell banks, raw materials and bulk harvest, integrated with downstream purification activities for viral clearance [11]. These guidelines are specified in the ICH Topic Q5A (R1) [12]; this document states that *"viral contamination can be reasonably assured only by the application of a virus testing program and assessment of virus removal and inactivation achieved by the manufacturing process"*

Viruses that may have been accidentally introduced into the manufacturing process of a biotechnological molecules are defined as adventitious viruses [13]. These viruses can come from several sources, including contaminated biological reagents, cells and medium handled or from virus used for the induction of gene expression. In addition to adventitious viral contaminants many cells may have endogenous retrovirus or latent virus [12]. These contaminants could have severe consequences for patient's health.

Assays that are required for the detection of viruses on cell banks and bulk harvest are routinely performed to guarantee the safety of cell banks and of the final products. All assays performed are fully validated according to ICH Q2 [14]. Virus detection assays include broad screen *in vitro* assays (cell and nucleic acid-based assays), tests for Retroviruses and *in vivo* assays (antibody production tests and viral inoculation in embryonated eggs). Table 1 outlines examples for these assays. In vitro tests are performed using detector cell lines to reveal an extensive variety of adventitious viral agents able to induce cytopathic or hemadsorbing effects. Other viruses that cannot be detected using the traditional cell-based methods are identified with Real-time polymerase chain reaction (RT-PCR) assays. For the detection of retroviruses, a battery of different tests is used, including cell-based assays, reverse transcriptase and

electron microscopy (EM). In vitro assays and retroviruses tests are beyond the focus of this thesis.

To evaluate the presence of viruses that cannot grow in cell culture, in vivo assays that involve inoculation of test article into Embryonated chicken eggs and specific pathogens free (SPF) animals are used. Between the in vivo tests that use SPF animals there are the Antibody production tests, including the: Mouse antibody production (MAP) test, Hamster antibody production (HAP) test and Rat antibody production (RAP) test. These assays are immunologically-based procedures for the detection of adventitious viral agents in cell lines used to produce biologics. All these tests include appropriate positive controls to ensure their ability to detect contaminants. Although the use of animal studies is an important tool for human safety assessment, the objections to *in vivo* biosafety studies on both ethical and scientific grounds have found strength with the availability of new alternative approaches not requiring the use of animals. The development of alternative tests applies the 3Rs principles (replacement, reduction and refinement). The 3Rs were first described in 1959 by Russel and Burch in *The Principles of Humane Experimental Techique* [15] and their purpose is to improve the animal welfare and promote the development of new method that can improve the sensitivity and accuracy of traditional viral safety tests. In this context companies are moving towards the replacement of *in vivo* tests using alternative methods. Attractive alternatives and valid support to traditional biosafety tests are represented by molecular assay such as, next-generation sequencing (NGS) and droplet digital polymerase chain reaction (ddPCR).

Table 1. Summary of the assays used for the detection of endogenous and adventitious viruses. Readapted from [16]

ADVENTITIOUS VIRUS DETECTION
In-vitro assay • MRC-5/Vero/CHO or other detector cell lines *In-vivo* assay • adult & suckling mice • embryonated eggs
DETECTION OF RETROVIRUSES
Transmission electron microscopy of cells Cell based assay • S+L- Assay for infectious Xenotropic retrovirus • Mus dunni co-cultivation assay for infectious Murine Retrovirus Reverse transcriptase assay
SPECIES SPECIFIC VIRUS DETECTION
Hamster and Mouse Antibody Production Assay RT-PCR

1.2 NEXT GENERATION SEQUENCING TECHNOLOGY

Next Generation sequencing (NGS) is a high-throughput technique able of sequencing multiple DNA molecules in parallel. The success of NGS technology is due to its capacity to sequencing millions of DNA fragments at the same time, generating large amount of data within relatively short time [17]. Another advantage of this technology is the high sensitivity which allows the detection of rare DNA molecules (such as viral contaminants) and therefore making possible the study of single nucleotide variants at low frequencies [18]. Moreover the development of NGS has reduced sequencing cost enabling the widespread use of this technology [19]. In the last decades, several NGS-based applications have been developed, including metagenomic [20], transcriptomic [21] and epigenomic [22].

Among the several NGS platforms currently available, Illumina is the most widely adopted technology. In this study, we used this technology for the biosafety assessment of biotechnological products. Illumina's sequencers use a sequencing-by-synthesis method producing short sequences (also called reads).

The Illumina's workflow involves four major steps: library preparation, cluster amplification, sequencing and data analysis. NGS libraries are prepared by random fragmentation of the sample (DNA, cDNA or amplicon) followed by the adapter ligation at 5' and 3' positions (Figure 2A). Illumina adapters are oligonucleotide sequences composed by two binding regions which bind their complementary oligos on the flowcell and index that acts as a "barcode sequence" for each read. After that through a polymerase chain reaction (PCR) the fragments are amplified. The prepared libraries are loaded into a flow cell coated with oligos complementary to the library adapters. In the NGS sequencer, through a process called Bridge Amplification, each fragment is amplified forming distinct clonal clusters [23]. In particular, single-molecule clusters are generated in the range of millions to billions in each channel of the flowcell. Each cluster will contain multiple identical copies of the same library fragment, this process is required to boost the fluorescent signal used to read the sequence nucleotide (Figure 2B). The instrument can then analyze the sequence information of all the clusters simultaneously. For the sequencing phase, Illumina used the sequencing by synthesis (SBS) technology (Illumina). This technology employees one-channel, two-channel and four-channel methods for the detection of single nucleotide. The four-channel SBS uses 4 types (A/T/C/G) of reversible dye terminators, each nucleotide is tagged with different fluorophores and blocked at 3' position [25]. The sequencing process occurs in multiple cycles, each one reading a single nucleotide. Each cycle includes multiple steps. First, a fluorescently-label nucleotide is added to the growing nucleic acid chain based on the sequence of the template. Second, the clusters are excited by a laser source in order to record the addition of the nucleotide. This fluorescence is detected by a camera that takes a picture of the flow cell after each synthesis. Last, the 3' blocking group and fluorophore are cleaved to allow the incorporation of the next fluorescence nucleotide by DNA polymerase (Figure 2C). Instead of using a dye for each base in two-channel SBS are used 2 fluorescent dyes, and in one-channel SBS is used only one dye [26].

Illumina NGS systems support Single-End (or Single Read) and Paired-End (PE) sequencing. In the Single-End sequencing the instrument reads a fragment from only one end, generating a single sequence for each DNA molecule. In the PE sequencing,

instead, both ends of the fragment are sequenced, generating two sequences for each molecule. At the end of the sequencing of the forward strands, the newly synthesized reverse strands are regenerated by bridge amplification. The forward strands are removed, leaving attached on the flow cell only the newly synthesized reverse strands to be sequenced to produce paired end sequence data. PE sequencing allows the production of high-quality reads alignment because the sequences in pair can span longer distances, increasing the accuracy of the alignment (Illumina).

The NGS data analysis process can be divided into three major steps: primary, secondary, and tertiary data analysis. During the primary analysis a base-calling algorithm converts digital images to FASTQ files, providing sequences and associated quality score to each read. In the secondary analysis, short reads are aligned against a reference sequence (reads mapping) or used to build longer sequences (d*e-novo* assembly) (Fig. 2D). Lastly, collected data are interpreted through bioinformatics, integrating public database information or other sample-related information.

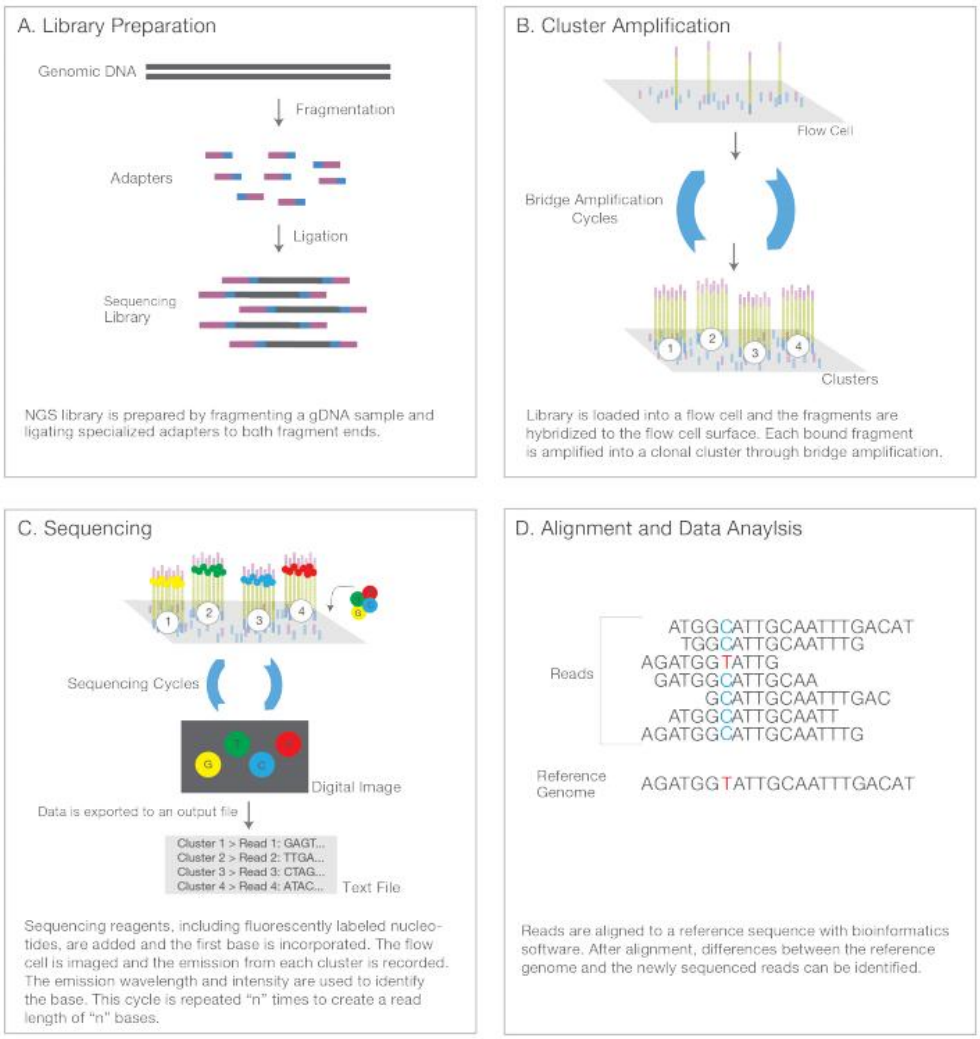


Figure 2. Next generation sequencing overview (Illumina) Illumina NGS includes four steps: (A) library preparation, (B) cluster generation,(C) sequencing, and (D) alignment and data analysis.

1.3 DROPLET DIGITAL PCR

Droplet digital polymerase chain reaction (ddPCR) is an advanced technology that offers the advantage of an absolute quantification of target DNA copies per input sample without the need of a standard curve. This technology allows more precise and reproducible data versus quantitative PCR (qPCR), especially in the presence of PCR inhibitors [28].

Digital PCR is used in many areas of biology, including microbial research and management of infectious disease for detecting pathogen [29,30]. Moreover, it can be a useful tool for the detection and quantification of rare sequences and for oncology research to detect circulating tumor cells or circulating tumor DNA in blood. Recently, ddPCR was applied in the biopharmaceutical industry, for the detection of host residual DNA [31] or impurities in the purified drug [32].

The ddPCR system is an implementation of digital PCR based on water-oil emulsion technology created by Bio-Rad Laboratories Inc. (Hercules, CA), commercially available since 2011. This technology uses surfactant chemistries and microfluidic circuits to divide a 20 µL reaction volume of reagents and sample into 20,000 nanoliter-sized droplets [33]. As a result of the partitioning step, some droplets will contain one or more copies of the target molecule and some will not contain any copies of the target. Each droplet represents a small bioreactor in which a PCR amplification is carried out and the contained DNA is amplified using target-specific primers and fluorescent probes. Following PCR, an automated droplet reader that works like a flow-cytometer is used to analyze the sample. The ddPCR workflow is illustrated in Figure 3. The system can use either hydrolysis probe or fluorescent DNA-binding dyes. In our work, the hydrolysis probe approach was used; this assay includes a probe with a sequence specific labelled with a fluorescent reporter at 5' and a quencher at the 3' end in addition to the sequence-specific primers. During the amplification phase the 5' to 3'exonuclease activity of the polymerase degrades the probe allowing the fluorophore release and the subsequent emission of fluorescence. The reader is tuned to the excitation/emission wavelengths of the reporter dye employed on the hydrolysis probes (often FAM or HEX). In this way, the reader can analyze each droplet and

determine whether the droplet is positive or negative for each fluorophore. Positive droplets contain at least one copy of the target molecule and show an increased fluorescence as compared to negative droplets. Ultimately, the reader measures the fraction of positive droplet in the set of droplets analyzed. The discrimination between negative and positive droplets is obtained by applying a threshold on the fluorescence amplitude in the software program used for data acquisition and analysis.

Figure 3. Schematic representation of the ddPCR workflow: Firstly, droplets of PCR Mix with primers, probe, and DNA sample are made in an oil suspension. Secondly, a PCR amplification is performed. Lastly, fluorescence is measured in each droplet by flow cytometry. Adapted from Biorad [34]

The collected data are analyzed using Poisson statistics in order to calculate the target concentration in the sample. This distribution was first introduced by Simeon Denis Poisson in 1837 and it expresses the probability of a given number of events occurring independently in a fixed interval of time and/or space when the average rate of occurrence is known. The formula used for Poisson distribution is:

$$-ln(1-p) = \text{Target copies per droplet}$$

Where p is the fraction of positive droplets

An optimized ratio between the number of positive droplets and the total number of droplets is at the base of an accurate Poisson analysis. A higher total number of droplets corresponds to a higher precision of Poisson based counting providing an accurate estimation from very low target copy levels to high copy levels [35]

2 SCOPE OF THE THESIS

The goal of this thesis is the development of methods based on molecular innovative technologies, such as ddPCR and NGS, to support or replace standard safety testing methods.

As previously stated, all viral safety tests should include a positive control. The first two experimental chapters of the thesis are therefore focused on the development of methods for the characterization of viral stocks. These methods are also important to gain enough information on the stocks that are used during the validation of methods aim at the detection of viral contaminants.

In the third chapter, a new viral titration method based on ddPCR is presented. This technology is used to perform an absolute nucleic acid quantification of the viral stocks used as internal controls. The section describes the design of specific assays used to perform ddPCR and the optimization steps followed to develop this new protocol. Moreover, the viral load of viruses obtained with ddPCR is compared with the titer obtained by the cell-based assays.

In the fourth chapter, the development and the validation of an NGS based method to qualify the viral stock identity and exclusivity is described. NGS is used in order to characterize the viral stocks, verifying the virus species and exclude the presence of other contaminant viruses. The method required optimization of the workflow for viral library preparation and the development of a dedicated bioinformatic pipeline for data analysis. The robustness and Limit of Detection of the method was evaluated.

In the fifth chapter, we present an alternative method based on NGS to replace two *in vivo* methods: the Mouse antibody production (MAP) and the Hamster antibody production (HAP) tests. NGS is used in order to verify the presence of species-specific viruses present in rodent cell lines. The optimization of the method includes the setup of the experimental workflow and the creation of a bioinformatic pipeline for virus detection. During the setup of the method, spiking studies were performed to evaluate the Limit of Detection (LOD) of the method.

3 QUANTIFICATION OF VIRAL LOAD USING ddPCR

3.1 INTRODUCTION

Viral safety tests are routinely used for the biosafety assessment of cell banks and biotechnological molecules. These tests should include appropriate controls to guarantee adequate sensitivity and specificity. In order to perform viral safety tests, positive controls viruses need to be propagated and prepared stocks have to be subsequently characterized by the evaluation of the identity and titer. The focus of this chapter is the viral titration step.

Currently, different methods are routinely used for viral titration depending on the type of virus analyzed. The most used assays are the plaque assay [36], the focus formation assay (FFA) and the 50% tissue culture infective dose ($TCID_{50}$) [37]. All these assays are performed to evaluate the infectious titer of viruses which can induced cytopathic effect (CPE). In the first two assays, each plaque/focus corresponds to an infectious unit and the titer is usually expressed as Plaque forming unit per ml (PFU/ml) or focus forming units per ml (FFU/ml). The last one instead is an end point dilution assay used to measure the load of virus needed to kill the 50% of infected cells. A mathematical method is used to calculate the $TCID_{50}$, the recommended one is the Spearman-Kaerber [38]. *Bryan et al*. observed that there was a correlation between PFU and $TCID_{50}$, in particular 1 $TCID_{50}$ corresponds to 0.69 PFU [39].

Even if those assays are considered the "gold standard" for viral titration, they have some limits. Firstly, they require appropriate cell substrates because each virus has a specific host cell and some viruses does not produce any cytopathic effect. Secondly, using cell-based assays we cannot estimate the absolute number of viral genome copies in a sample. As this information is useful for the molecular assays used for the detection of viral contaminants in quality control (QC) viral safety tests ddPCR could represent an innovative tool to support these activities. The ddPCR has many advantages, including, a high sensitivity, high reproducibility, and the possibility to perform an absolute viral genome quantification without need of standard curve [40]. Moreover, once the method has been set (thermal profile, input sample concentration

and specificity are defined) it is fast, semi-automated and easy to perform. Conversely, cell-based assays are laborious and time-consuming delivering results in an average of a week.

The aim of this study is to develop and optimize several ddPCR assays for the absolute quantification of viral stocks prepared and used as positive control for viral safety tests. For this purpose, 5 different viruses were propagated and titrated. In particular, two types of viral titration assays were performed to evaluate virus infectivity and nucleic acids molecules. Infectious particles were enumerated by end-point infectivity assay, while the detection of nucleic acids was performed by ddPCR. In order to obtain the absolute quantification of the viral stocks, we followed two experimental phases. Firstly, viruses were sequenced in order to design three different set of primers and probe specific for each virus. After that, we optimized each ddPCR experimental protocol by the assessment of thermal profiles, the definition of the amount of input material used as input in the ddPCR reaction, and the evaluation of the assay specificity. Finally, once concluded the optimization experiments, we selected the best assay for each virus, and we used it to perform the absolute viral genome quantification of one viral propagation for each virus type.

3.2 MATERIAL AND METHODS

3.2.1 HOST CELLS AND VIRUSES

Each of the five viruses was propagated infecting its specific host cell line and both were purchased from American type culture collection (ATCC). Five different viruses were used: Mouse adenovirus type 1 strain FL (MAV-1), Minute virus of mice (MVM), Reovirus 3 strain Abney (REO3), Theiler's murine encephalomyelitis virus (TMEV) and Murine pneumonia virus (PVM).

MAV-1 ATCC VR-550 is a non enveloped, with double-stranded (ds) linear DNA virus[41]. This virus was propagated into the BALB/3T3 clones A31 (ATCC CCL-163), a mouse embryonic fibroblast cell line.

MVM [42] ATCC VR-1346 was propagated into the A9 (APRT and HPRT negative derivative of strain l) (ATCC CCL-1.4), another murine fibroblast cell line. The virus contains a single-stranded (ss) DNA genome.

REO3 ATCC VR-232 is a nonenveloped virus that contains a genome of ten segments of dsRNA. The host cells used for the propagation are the LLC-MK2 Derivative (ATCC CCL-7.1), a monkey (*Macaca mulatta*) epithelial cell lines.

TMEV [43] ATCC VR-995 was propagated into the BHK-21 [C-13] (ATCC CCL-10), hamster (*Mesocricetus aureus)* fibroblast cell lines and it's genome is a ssRNA of positive polarity.

PVM ATCC VR-1819 is an enveloped, negative sense, ss RNA viruses [44]. It was propagated into the BHK-21 [C-13] (ATCC CCL-10), hamster (*Mesocricetus aureus*) fibroblast cell lines.

3.2.2 VIRUS PROPAGATION

Cells were cultivated in media without antibiotic for 2 weeks until they reached semi-confluence. The day of the infection the cell growth medium from the cell culture flasks was removed and cultured cells were inoculated with the virus and resuspended with specific media. For virus propagation the FBS concentration was reduced at 2%, in all prepared media. When ~90% cytopathic effect was reached (at day 3 to 10 post infection [p.i.]), the culture supernatant was harvested, centrifugated at 600 × *g* and

filtered through 0.45 µm filters. The clarified culture supernatant was divided into aliquots and stored at −80°C to be used as positive controls in the in vitro experiments.

3.2.3 VIRAL TITRATION BY TCID50

To perform viral titration, permissive cells ($1,5x10^5$ cells/1 ml) were seeded in each well of a 96-wells plate and incubated at 37°C in 5% $C0_2$ and humid atmosphere for 24 hours. The day of the infection eight ten-fold serial dilutions of virus were prepared. Subsequently, 5 replicates of 25 µl of each dilution were plated. The same number of replicates was used for the negative control. Cells were incubated at 37°C since the CPE induced by inoculated virus could be observed. Then cells were washed and fixed 10 min with absolute methanol. A crystal violet staining was performed to microscopically observe the cytopathic effect

3.2.4 VIRAL GENOME EXTRACTION AND ILLUMINA SAMPLE PREPARATION

The workflow used for the sample preparation is described in the paragraphs 4.2.3 and 4.2.4. Briefly, nucleic acids were extracted and retro-transcribed to have the double strand complementary DNA (dscDNA) required for the library preparation.

Library preparation was carried out by means of Nextera XT DNA Library protocol (Illumina).

3.2.5 ddPCR ASSAY DESIGN

Quering public viral genome database the regions without mutations were identified for each virus. In those regions we designed our set of primers and probe. In particular, three different sets were designed for each virus.

3.2.6 DROPLET DIGITAL PCR

The QX200 AutoDG Droplet Digital PCR System (Biorad) was used in this study. We used 2 different kits: ddPCR Supermix for Probes (No dUTP) (Biorad) for DNA viruses and the One-Step RT-ddPCR Advanced Kit for Probes (Biorad) for RNA viruses. Both kits follow the same protocols, with the exception that the One-Step reverse transcription and Droplet Digital PCR (RT-ddPCR) Advanced Kit for Probes starts directly from total

RNA combining in a single-reaction the reverse transcription and the ddPCR. PCR Mastermixes were prepared in a final volume of 24µl. The Mixture used for DNA viruses contained: 2 µl of diluted DNA, 1x Supermix for Probes (No dUTP), 900 nm primers and 250 nm probe. The Mixture used to analyze RNA viruses contained: 2 µl of diluted RNA, 1x Supermix, Reverse Trancriptase 20U/ µl, DTT 15 mM, 900 nm primers, and 250 nm probe. In all cases extracted viral genomes were diluted to be within the range of detection of the instrument.

Droplets were generated using Automated Droplet Generator and fluorescence signals were measured using a QX200 Droplet Reader. Data were analyzed by QuantaSoft Regulatory Edition Software (Ver. 1.7) and thresholds were manually set for each sample. Raw fluorescence data were processed by the software to produce a copies/µL value for each well tested by counting positive and negative droplets. The test was considered valid if the number of droplets generated was ten thousand or above.

3.2.7 THERMAL GRADIENT OPTIMIZATION

The ddPCR reaction conditions were optimized for each designed assay. For DNA viruses we set only the optimal annealing temperature for primers and probes, while for the RNA viruses the RT reaction temperature has to be set as well in order to enhance the specificity and efficiency of the reaction and ensure the full activation of the enzyme used for cDNA synthesis. For DNA viruses the optimization experiments consisted in the assessment of the optimal Annealing/Extension temperature, performed using a C1000 Touch thermal cycler (Biorad) with a thermal gradient between 55°C and 65°C.

The thermal protocol consisted in five stages: enzyme activation at 95°C for 10 min, 40 cycles of denaturation at 94°C for 30s, than annealing/extension with a thermal gradient between 55-65°C for 1 min (temperature ramp -2°C/s) and, finally enzyme deactivation at 98°C for 10 min and storage at 4°C.

For RNA viruses the optimization experiments consisted in the assessment of the optimal temperature for the reverse transcription before the establishment of the Annealing/Extension optimal temperature. The thermocycling protocol included six main stages: reverse transcription with thermal gradient between 42-50°C for 60 min

(temperature ramp -2°C/s) , enzyme activation at 95°C for 10 min, 40 cycles of denaturation at 95°C for 30s, annealing/extension with a temperature between 55-65°C for 1 min and, finally enzyme deactivation at 98°C for 10 min and storage at 4°C. All samples were run in 8 replicates with no-template controls (NTCs) and noRT negative controls were run for RNA viruses.

We selected the optimal primers/probe annealing temperature and RT temperature for each set of primers and probes. The optimal annealing/RT temperature was the one resulting in the greatest difference in fluorescence amplitude between the positive and negative droplets avoiding non-specific amplification. Figure 4 represents a one-dimensional (1-D) plot of ddPCR thermal gradient optimization experiment.

3.2.8 ASSAY EVALUATION

The optimized assays were tested for their linearity and specificity. To evaluate the linearity of the method different concentration of nucleic acids were analyzed. Starting from extracted nucleic acids, ten-fold serial dilutions for each virus were performed. Considering that the starting nucleic acids concentration of each viral stock is variable, different dilutions of the extracted input material were analyzed in order to stay in the dynamic range of ddPCR. The dilutions tested for each virus are listed below:

1. REO3: 10^{-3} -10^{-4}-10^{-5}
2. TMEV: 10^{-4} -10^{-5} -10^{-6}
3. PVM: 10^{-1} -10^{-2} -10^{-3}
4. MVM: 10^{-5} -10^{-6} -10^{-7}
5. MAV: 10^{-3} -10^{-4} -10^{-5}

To test the specificity of our assays, it was verified that no amplification occurred on the host cell genome in the absence of virus infection. In particular, the RNA was extracted from host cells used to propagate RNA viruses, while DNA was extracted from host cells used to propagate DNA viruses. After that, a ddPCR was performed on DNA or RNA.

At the end of the optimization phase we selected the best assay for each virus considering the specificity and the linearity. The selected assays were used to assess the absolute quantification of the genome copies number for each virus. The absolute quantification of the genome copies number was performed testing 3 vials of each virus in order to have a more representative quantification of the whole propagation. For each sample, eight technical replicates were tested. We observed that ddPCR represents a highly precise method to measure viral titer.

Therefore, viral abundance was calculated for each virus using the following formula:

Aliquots concentration (copies/µl) =

Average copy concentration/µL multiplication factor* Dilution Factor*

Where:

Average copy concentration/µl = $\frac{mean\ total\ copies\ per\ 20\mu l\ obtained\ from\ technical\ replicates}{2\mu l\ diluted\ sample}$

multiplication factor = $\frac{\mu l\ of\ input\ sample}{\mu l\ elution\ volume}$

3.3 RESULTS

3.3.1 TISSUE CULTURE INFECTIVE DOSE 50% TITRATION

The amount of infectious particles was expressed in PFU/ml. The titer was measured at different time points post-infection, as the replication kinetic is different for each virus analyzed. The titers obtained for each virus are shown in Table 3. No morphological changes were observed in cells infected with PVM even at long post-infection times. This observation is in contrast to the findings of Compans and colleagues, who observed that in BHK-21 cells, the CPE began about 48 hours post-infection, but the viral cytopathogenicity may be low in these cells as reported by *Miyata and coworkers* [45] [46].

Table 3. Summary of PFU/ml values obtained by end-point dilution assays

Virus	Cell Culture	Titer (PFU/ml)[a]	Days p.i.
MAV-1	BALB/3T3	8.9×10^5	8
REO-3	LLC-MK2	1.4×10^5	7
MVM	A9	8.8×10^4	6
TMEV	BHK-21	1.8×10^4	3
PVM	BHK-21	n.a.	n.a.

[a]PFU/ml values were obtained by the following equation : $TCID_{50}$/ml x 0,69 [39]

3.3.2 ANNEALING AND RT TEMPERATURE

A thermal gradient experiments (Figure 4) was used to optimize annealing and RT temperatures of the tested sets of primers and probe(Data not shown).

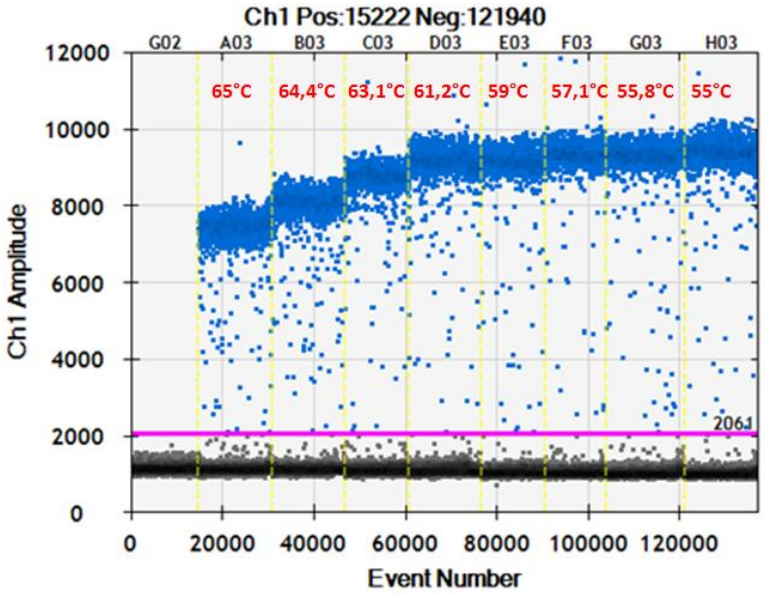


Figure 4. Annealing temperature gradient experiments. A gradient PCR ranging from 55°C to 65°C is represented in the 1-D Plot. The pink line is the threshold. Under the threshold there are negative droplets (grey) while over the threshold there are positive droplets (blue). Eight reactions are performed with the same amount of nucleic acid. All reactions are divided by the vertical yellow line. In the first column no amplification occurred because there is the NTC. The best separation of the droplets occurred at 57.1 °C (column F03).

3.3.3 DROPLET DIGITAL PCR ASSAY SPECIFICITY

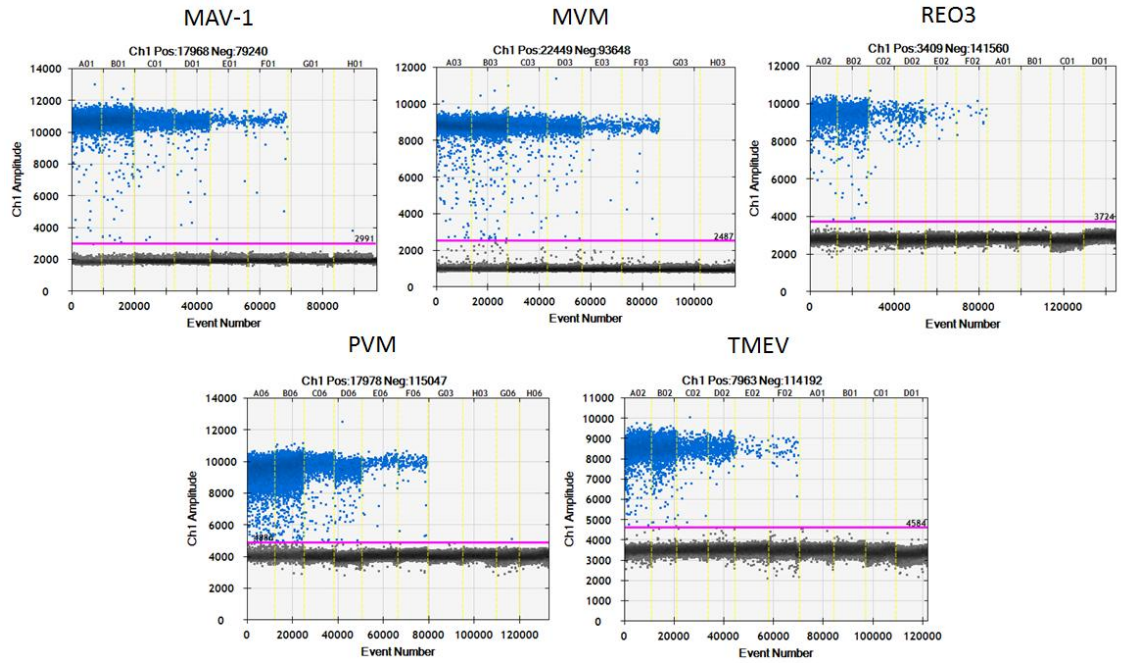
Using the Nucleotide Basic Local Alignment Search Tool (BLAST) we performed the *in-silico* testing of the primers and probes. Additionally, we checked the specificity of the ddPCR assays comparing results obtained by target nucleic acids and cells nucleic acids. The ddPCR assays provided negative results from non-target samples indicating no cross reaction. The obtained results confirmed the specificity of primers and probes for viruses.

3.3.4 DROPLET DIGITAL PCR ASSAY LINEARITY

The linearity of the assay was assessed by quantification of three ten-fold serial dilutions of extracted genome for each virus. Two technical replicates were tested for each sample. The serial dilutions of nucleic acids were plotted against the concentration (copies/µl) measured by ddPCR. Concentrations were calculated using QuantaSoft Software and represent the measurement of concentrations in merged wells for each sample. In all cases there was a good linearity between the target input amounts and quantified values with the coefficient of determination (R^2) ranging between 0,9 and 1. Linear regressions of the ddPCR assays on viral genomes are shown

in panel B (Figure 5). Only data of the selected primers/probes sets are reported. Panel A (Figure 5) represents a one-dimensional (1-D) plot of ddPCR reactions with serially diluted targets.

A



B

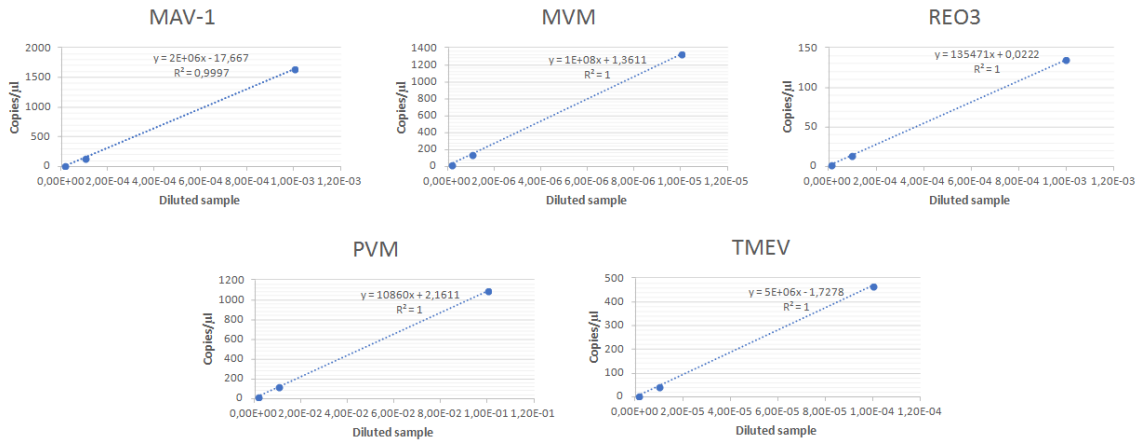


Figure 5. A) 1-Dplot of ddPCR assays. Serial dilutions of target nucleic acids, six ddPCR reactions with three serially diluted targets are divided by the vertical yellow line. The right most columns contain negative controls. **B) Correlation between diluted samples and copies/µl obtained from ddPCR assays**. The equation represents the linear regression between both data. Each data point represents the mean of two replicates of one experiment.

We used this analysis to determine the optimal dilution of input DNA or RNA for the assays. Indeed, in order to obtain the best results, the input sample amount has to be adjusted to stay within the dynamic range of the system because high viral nucleic acids concentration generally corresponds to a bad separation between positive and negative droplets with high rain effect.

3.3.5 DROPLET DIGITAL PCR TITRATION

For each vial analyzed, eight technical replicates were tested. Results obtained from the analysis showed that virus titers were variable inside the propagations (Figure 6), even if in all cases the same order of magnitude is measured (Figure 7).

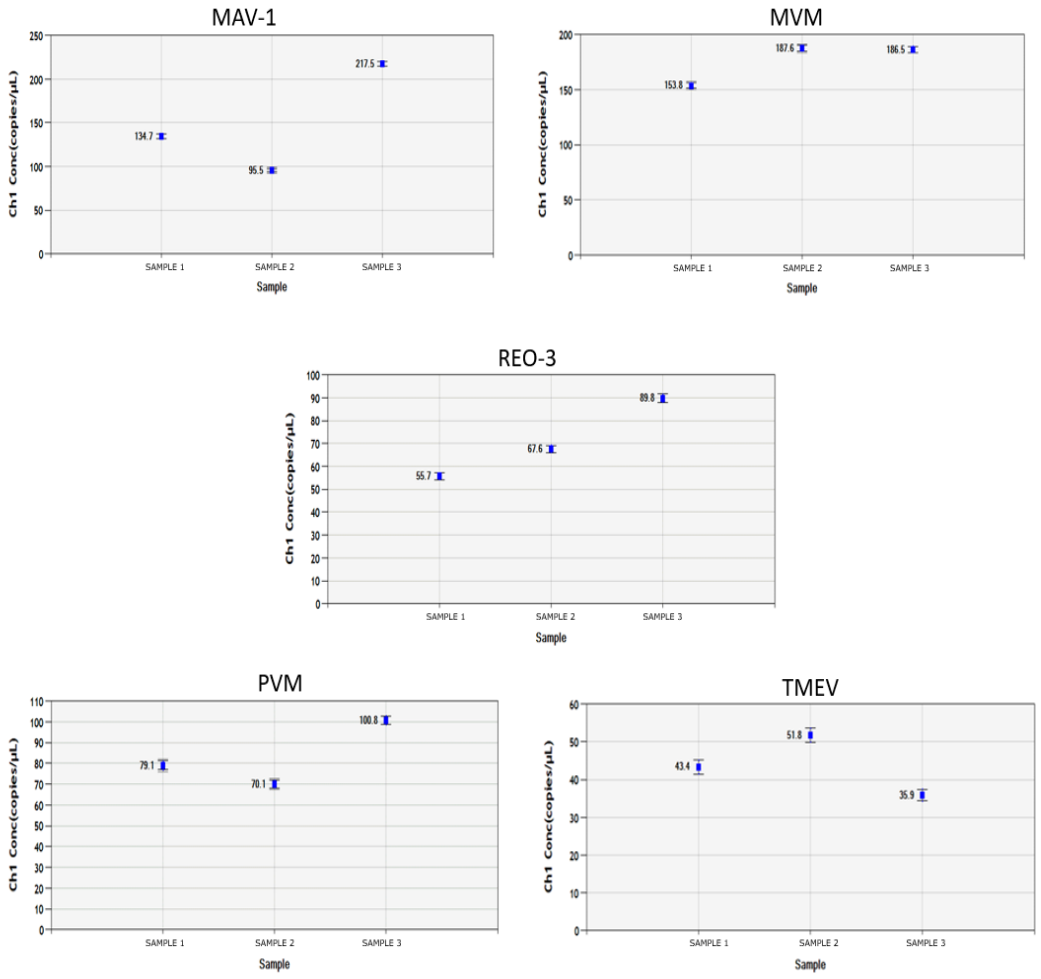


Figure 6. Viral genomes quantification of the 5 virus propagation. Three vials of each propagation are analysed (replicates=8). Concentrations are provided in units of copies per microliter. The inner error bars indicate the Poisson 95% confidence interval (CI) and the outer error bars show the total 95% CI of replicates.

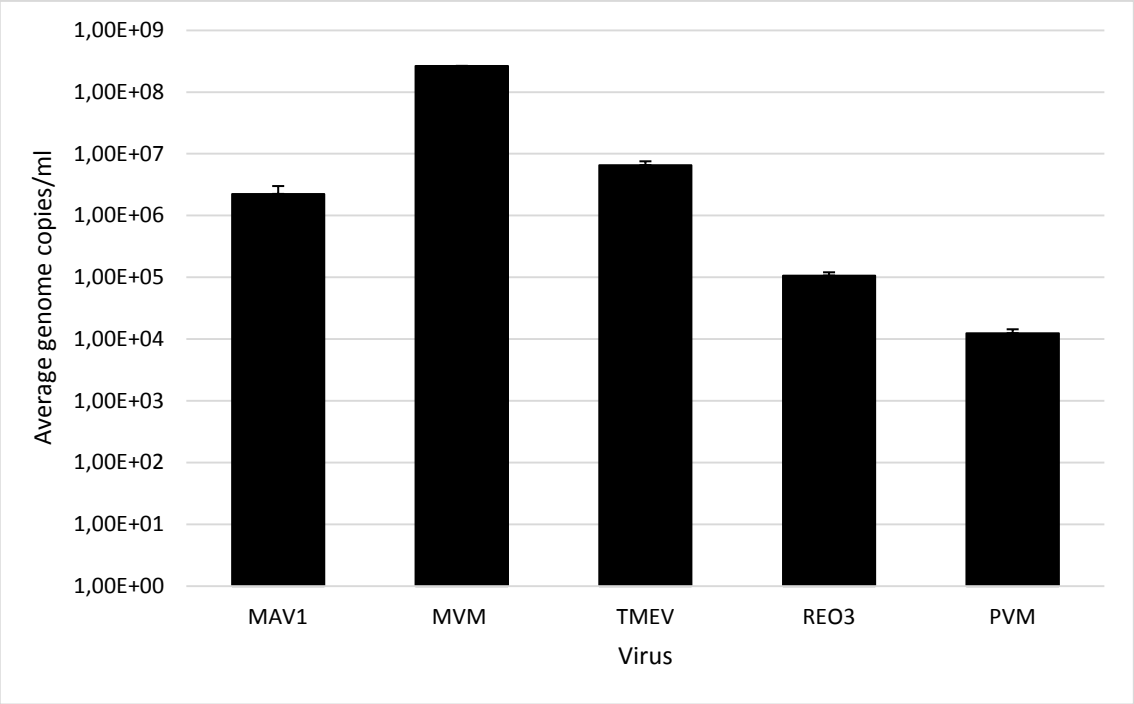


Figure 7. Average of samples genome copies (gc) per mL

3.4 DISCUSSION

The titration of viruses used as positive controls for viral safety tests is routinely conducted using cell-based assays, such as the $TCID_{50}$ method. This method allows the assessment only of the infectious particles that are present in a sample. For the development of molecular assays, such as those described in the next chapters, there is a need to introduce a method able to determine absolute number of viral copies present in a given sample. Therefore, we developed an alternative method based on ddPCR. Several papers in literature have demonstrated that this technology represents a new promising tool for the precise viral load quantification [30,47]. In the present study, the assessment of ddPCR on five different viruses was carried out for the first time in our laboratories.

We compared ddPCR-derived titers with titers obtained by TCID50. We observed that the target copy number determined by ddPCR was from 3-fold to 7-fold higher than the number of PFUs determined by the end point dilution assay. This difference may be due to the presence in the supernatant of infectious and non-infectious virions [48] and to the levels of viral messenger RNA in the sample [49]. Moreover, the two assays did not deliver comparable titers may be due to the efficiency of the extraction and retrotranscription steps [40]. For what concerns PVM, we could not compare results obtained from $TCID_{50}$ and ddPCR, because we did not observe any CPE with the cell-based assay. We hypothesized that it could be attributable to a low number of infectious particles in the harvested supernatant. However, viral genome copies were detected confirming the presence of viruses in the viral stock prepared.

The high specificity of the designed primers and probes, used for ddPCR assay, was demonstrated by the fact that cell's DNA or RNA did not interfere with the quantification of viral genome copies. In this study linearity, precision and dynamic range of ddPCR was evaluated. The assays exhibited in all cases a good linearity (R^2 ranging between 0,9 and 1) over the dynamic range tested in both the RNA and DNA viruses. Moreover, this procedure has a high dynamic range, which reduces the number of dilutions needed for analysis and it does not require tissue cultures. Finally,

although the “gold standard” for viral titration are cell-based assays, these tests takes almost ten days to obtain the final results while the method developed in this study required, after the optimization, only one day. We observed titer variability between vials of the same viral propagation, and we speculate that this difference between samples may be related to the manually performed supernatant collection. Further quantification using more samples for each propagation is needed to evaluate if the difference is statistically significant, because each viral propagation is made of more than two hundred vials.

In conclusion, ddPCR represents a powerful tool to perform viral titration and provide useful information for Quality Control molecular viral safety tests that can be used in association with the viral titration methods required by regulatory authorities.

4 VIRAL SEED STOCK IDENTITY BY NGS

4.1 INTRODUCTION

The production of drug substances by animal cells implies the risk of adventitious virus contamination of the final product. For this reason, cell banks and unprocessed bulks harvest must be tested for the presence of adventitious viral agents or endogenous viruses. A number of assays can be used for the detection of these contaminants and these tests include, as mentioned in paragraph 3.1, positive controls. In this chapter we designed and tested an NGS-based method designed to both confirm the identity of viral seed stocks used as positive controls and exclude cross-contaminations due to other viruses that are propagated in the same laboratories.

NGS allows the identification of viruses by the sequencing of random genome fragments in the test sample. In this way, NGS can potentially detect any virus and can be used for the detection of viruses of any genome type (ssDNA, dsDNA, ssRNA and dsRNA).

In this context, the NGS platform is described as a technology tool able to confirm viral stock identity and detect eventual contaminations among other viruses propagated in our laboratories. The sample composition and complexity can influence the sequencing results, thus the sample preparation needs to be adapted by the introduction of sample pre-treatment methods such as virus enrichment, filtration and/or nuclease treatment. For this purpose, we developed a method based on DNAse and RNAse treatments of viral stocks, followed by nucleic acid extraction, reverse transcription reaction and library preparation by Illumina technology.

In our method, the data obtained from NGS are then used to interrogate a curated internal database containing only the genome sequences of viruses that could potentially contaminate the virus propagation (meaning all the viruses propagated in the same lab).

For the assessment of the method performance we evaluated three characteristics: the Limit of Detection, the specificity and the robustness.

4.2 MATERIALS AND METHODS

4.2.1 VIRUS STOCKS

Viruses used for this study were viral stocks prepared and used for QC routine testing in Merck Ivrea laboratories (Table 4). ssRNA bacteriophage, MS2, was used as positive control for specificity and limit of detection experiments.

Table 4. List of viruses used as positive control in the Laboratory 1 (LAB1) and Laboratory 2 (LAB2), characteristics and host cells. Bacteriophage MS2 is mentioned as is used as positive control in this method.

Unique name	Virus	Size (~kbp)	Genome	Host cells	Laboratory
RVB-574	Suid Herpesvirus 1 (SuHV-1)	142	DsDNA	MDBK (NBL-1)	LAB2
ATCC® VR-93	Human Parainfluenza 3 (HPIV3)	15	ssRNA(-)	LLC-MK2 derivative	LAB2
ATCC® VR-232	Reovirus 3 (REO3)	23.5	DsRNA	LLC-MK2 derivative	LAB2
ATCC® VR-1350	Moloney murine leukemia virus – ecotropic (MLV1350)	8332	ssRNA (+)	SC1	LAB1
ATCC® VR-1450	Hybrid Moloney/Amphotropic nurine leukemia virus (MLV1450)	8041	ssRNA (+)	NIH 3T3	LAB1- LAB2
ATCC® VR-861	Murine leukemia virus (MLV861)	8275	ssRNA (+)	MV1LU	LAB1
ATCC® VR-158	Vesicular stomatitis virus (VSV)	10832	ssRNA (-)	MDBK	LAB1
ATCC® VR-95	Influenza virus H1N1 (INFLA)	13283	ssRNA (-)	EMBRIONATED CHICKEN EGGS	LAB1
ATCC® VR-260	Human Herpes virus (HSV)	147029	DsDNA	VERO	LAB1
ATCC® VR-995	Theiler's murine encephalomyelitis virus (TMEV)	7238	SsRNA	BHK-21	LAB1
ATCC® VR-764	Murine hepatitis virus (MHV)	31314	ssRNA (+)	NCTC clone 1489	LAB1
ATCC® VR-1346	Minute virus of mice (MVM)	5128	SsDNA	A9	LAB1 – LAB2
ATCC® 15597B1	Escerichia coli bacteriophage MS2 (MS2)	3569	ssRNA	E. coli	-

4.2.2 METHOD WORKFLOW

The overall method workflow is summarized in Figure 8 and is detailed in the following paragraphs.

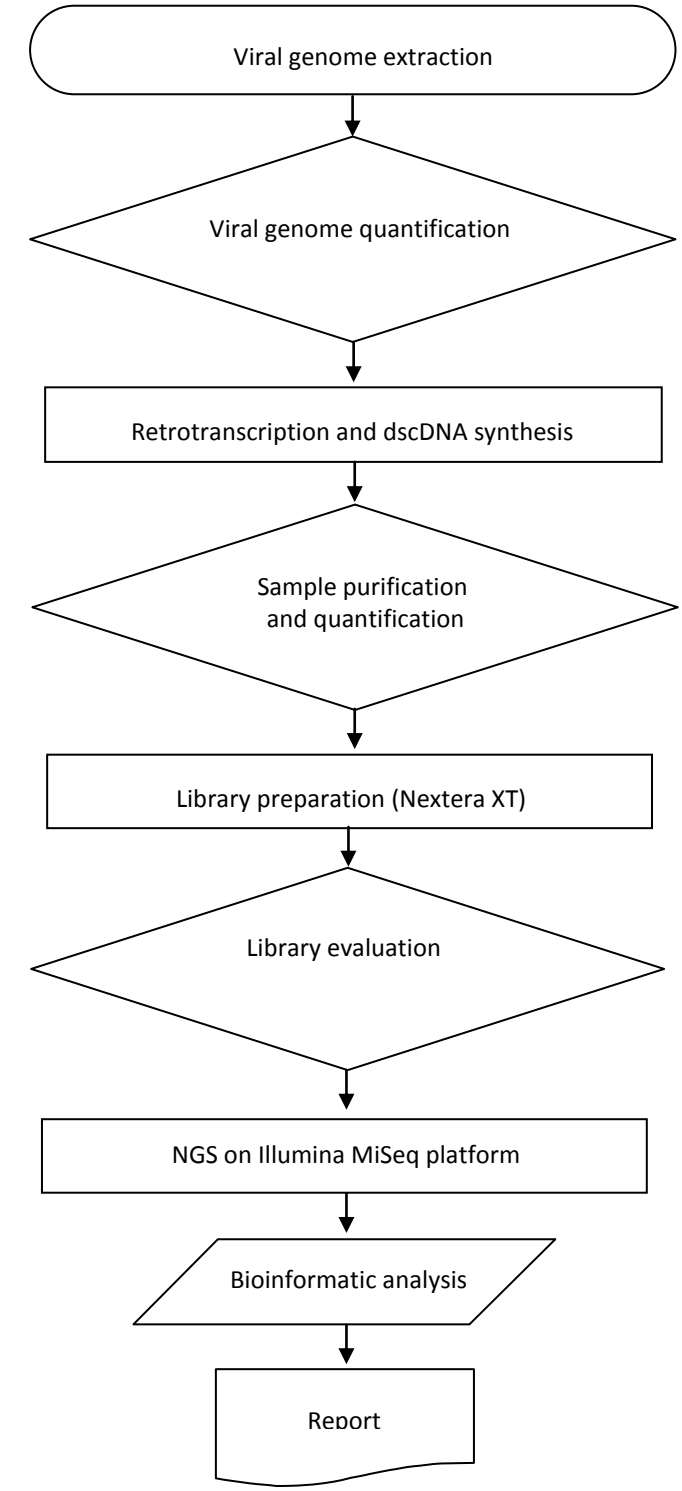


Figure 8. Summary of the method workflow

Start of process | Activity of process | Data | Document | Acceptance criteria

4.2.3 VIRAL GENOME EXTRACTION

Before the extraction, 200µl of viral seed stocks were filtered through 0.45 µm pore-size nylon membranes to remove particulates larger than those of most viruses. Turbo DNAse (Thermo Fisher) and Ambion RNAse I (Thermo Fisher) enzymes were added to the eluates and samples were incubated for 30 min at 37°C, this step improves the recovery of viral nucleic acids eliminating exogenous materials. Viral genomes were then extracted with a QIAmp MinElute Virus Spin Kit according to the manufacturer's instruction. At the end samples were quantified using Qubit® 2.0 Fluorometer. Depending on the genome of the extracted virus different kits were used for the quantification including ssDNA, dsDNA or RNA high sensitivity assay kit. The extracted were used as template for the dscDNA synthesis. Only if the extract concentration was greater than 2 ng/µl we proceed to the next step of the protocol.

Retro-transcription was carried out using an internal protocol. Briefly, the RNA was cleaved into small fragment (~200 nt) by Elute/Prime/Fragment Buffer, used heat and divalent cation. The dscDNA synthesis was performed using the TruSeq RNA Library Preparation Kit. This method consists of a preliminary incubation with random primers and a first strand synthesis using the SuperScript® III Reverse Transcriptase (Thermo Fisher). After that, Second Strand Master Mix is added into the reaction mix to synthesize the second strand. Then, AMPure XP beads (Beckman Coulter) were used to purify the dscDNA. Finally, the dscDNA was quantified by Qubit® 2.0 Fluorometer and then used for the library preparation. The input dscDNA was then used for library preparation only if the obtained concentration was greater than 2 ng/µl.

4.2.4 LIBRARY PREPARATION AND LOADING

Starting from 10 ng of dscDNA, libraries were generated using Nextera XT DNA Library Prep Kit (Illumina), following the manufacturer's instruction. The first step of the Nextera XT provides the DNA fragmentation by bead-linked transposomes and the tagmentation of fragments with sequencing adapters.

After an incubation of 5 minutes at 55°C, the tagmented DNA is amplified by limited-cycle PCR program. Contextually, index adapters are added on both ends of the

fragments. At the end of the protocol samples were cleaned up using the 1X AMPure XP beads (Beckman Coulter) according to the manufacturer's instructions.

Purified libraries were quantified by Qubit 3.0 and sized using the Agilent 2100 Bioanalyzer. Libraries molar calculation was performed using validated spreadsheet. In order to be sequenced, the mean of the fragments size had to be higher than or equal to 200 bp and the molarity had to be greater than or equal to 2 nM. NGS libraries where then diluted to 2nM and prepared for sequencing.

4.2.5 SEQUENCING

Sequencing was performed on an Illumina MiSeq sequencer generating paired-end 2x150 bp reads. For libraries sequencing, both 300-cycle (MiSeq Reagent Kit V2) and 600-cycle (MiSeq Reagent Kit V3) sequencing kits were used. To analyze only one sample the kit V2 was used obtaining up to 30 million of reads. While, when 7 samples were multiplexed the kit V3 was used obtaining up to 7 million of reads for each loaded sample.

Based on the Lander/Waterman equation, using the Kit V3 a coverage of 6000x was expected for the HSV1, that was the virus with the largest genome size listed in our databases. This means that analyzing viruses with smaller genome an increase in the coverage was expected. To guarantee similar throughput between the libraries, similar library sizes have to be analyzed. For this reason, libraries could be pooled and sequenced together only if their fragments length ranged from 300 bp to 600 bp.

4.2.6 DATABASES

Three different databases were created, retrieving viral genome sequences from NCBI data bank (Table 5).

The DB1 is used to analyze viruses propagated in LAB1. The DB2 is used for viruses propagated in LAB2 and, finally, the DB3 is used for the positive control.

Moreover, a host cells genome dataset, comprising nine different reference sequences obtained by NCBI, is used by the pipeline to filter and subtract host-derived sequences (Table 6).

Table 5. List of databases used by the pipeline

Database	Viral reference sequence name
DB3	E. COLI BACTERIOPHAGE_MS2_(MS2)
DB1	MINUTE_VIRUS_OF_MICE_(MVM) MURINE_HEPATITIS_VIRUS_(MHV) THEILERS_MURINE_ENCEPHALOMYELITIS_VIRUS_(TMEV) VESICULAR_STOMATITIS_VIRUS_(VSV) *HUMAN_HERPESVIRUS_1_(HSV1)* *INFLUENZA_A_VIRUS_(INFLA)* *MURINE_LEUKEMIA_VIRUS_AMPHO_(Mo-A-MuLV_MLV1450)* *MURINE_LEUKEMIA_VIRUS_XENO_(MLV861)* *MURINE_LEUKEMIA_VIRUS_(MLV1350)*
DB2	MINUTE_VIRUS_OF_MICE_(MVM) MURINE_LEUKEMIA_VIRUS_AMPHO_(Mo-A-MuLV_MLV1450) *HUMAN_PARAINFLUENZA-3_(HPIV3)* *MAMMALIAN_ORTHOREOVIRUS_3_(REO3)* *SUID_HERPESVIRUS_(SuHV-1)*

Table 6. List of host reference sequences used for host subtraction

Host species	NCBI Reference sequenze
Mus musculus	GCF_000001635.26
Mesocricetus auratus	GCF_000349665.1
Macaca mulatta	GCF_003339765.1
Chlorocebus sabaeus	GCF_000409795.2
Gallus gallus	GCF_000002315.6
Neovison vison	GCA_900108605.1
Bos Taurus	GCF_002263795.1
Sus scrofa	GCF_000003025.6
Escherichia Coli	NC_007779.1

4.2.7 DATA ANALYSIS

The raw sequencing data produced by MiSeq instrument were analyzed with a custom pipeline. This pipeline allows to confirm the identity of viruses propagated internally to our laboratories employed as positive control for quality control tests. Moreover, it permits the identification of potential cross-contamination between different viral stocks. As mentioned (paragraph 4.2.6), three viral databases obtained from NCBI public data bank, were interrogated.

The bioinformatic pipeline aligns all generated reads with its similar sequences present inside the database. After the exclusion of the reads that aligned with the host genome dataset (Table 6), the remaining ones were aligned against the specific viral database. The presence of virus or viruses inside the sample was evaluated considering the viral sequences coverage rate. In particular:

- % coverage describes the fraction of reference that is covered by reads
- % coverage 10x indicate the fraction of bases that reach at least 10X sequence coverage
- Mapped reads refer to the number of reads that align on the reference genome

Only the viruses that exceed the defined cut-off values were considered as positive: *% coverage > 90*, *% coverage 10X > 80* and *number of mapped reads > 1000*.

4.2.8 SPECIFICITY

For the assessment of the method performance were evaluated the specificity and the robustness of the method. The term specificity refers to the ability of the test to correctly identify the analytical subject of the test and discriminate it from all other possible elements present within the sample. To identify the specificity of a method, it is necessary to demonstrate that the analytical procedure is able to discriminate between similar elements that could be simultaneously present in the same sample. The method specificity is intended as the ability to correctly identify the analysed virus and exclude contamination among other viruses propagated in the same laboratory. For this purpose, 13 different viruses were analyzed and bacteriophage MS2 was used as control. All viruses and MS2 tested, were processed following the workflow described in Figure 8. The method passed our specificity test if the pipeline identifies for each viral propagation only the virus that was present in the analyzed viral stock. To confirm viral identity the reads obtained from each virus were aligned against the specific Viral Database (Table 5).

4.2.9 ROBUSTNESS

To evaluate the performance of the method during its development was evaluated the robustness. Robustness is the term used to identify the capability of an analytical

method to not undergo variations after analytical conditions adjustment. For this purpose, a series of operational variations are introduced to verify that the method is not susceptible to those operational changes.
The robustness analysis was evaluated through the design of a risk assessment that had the scope to determine which parameters were to be considered critical for the analytical method and had to be tested.

Contextually to the method robustness analysis its risk assessment was determined by calculating the risk index of each experimental step.

In particular, three test phases were considered:

1. Nucleic acids extraction phase
2. Library preparation phase
3. Sample preparation and sequencing phase

For each test phase an overall of 9 analytical steps were identified for the risk assessment.

Three score were assigned to each analytical step, to evaluate its probability, severity and detectability and determine which of them could impact on the analytical method.

In particular severity describes the impact that a problem related to any phase of the test could have on its results; detectability is defined as the possibility to detect any problem within a specific tested phase; probability represents the possibility and the frequency to which the problem is likely to occur. By multiplying the values of the three attributes assigned to each experimental parameter, the risk priority number index (RPNi) was calculated. The RPNi can be defined as low, medium, high or very high. The steps that were rated at medium or high and very high risk had to be involved in the evaluation of the robustness of the method.
These factors were not tested individually but were included in an experimental matrix design (Design of Experiment - DoE) using the JMP software V. 12.0. After entering the number of factors that have to be evaluated, JMP determines the minimum number, type and order of tests to be performed, in order to achieve a statistically significant

analysis. The experimental tests were conducted using the approach defined in Table 7.

Table 7. Design of experiment. Nine Samples were analysed under different experimental conditions. The experimental conditions tested were: incubation timing with LB and protease(1st), incubation timing with the NTB (2nd) and incubation timing with the NaOH (3rd). The Pattern indicates if the incubation times used change in comparison with the standard conditions. Shorter times than standard conditions are indicated with (-) , while higher times with (+). The Sample 4 was tested using standard conditions.

Sample	Pattern	1st experimental condition	2nd experimental condition	3rd experimental condition
1	−−+	13’	5’	7’
2	+−+	17’	5’	7’
3	+++	17’	7’	7’
4	0	15’	6’	5’
5	+−−	17’	5’	3’
6	−+−	13’	7’	3’
7	++−	17’	7’	3’
8	−−−	13’	5’	3’
9	−++	13’	7’	7’

4.2.10 LIMIT OF DETECTION

The Limit of Detection (LOD) is described as the lowest concentration of analyte that can be detected by an analytical procedure. We established that the LOD corresponds to the minimum amount of each virus, intended as viral genome copies per microliter, detected by the method within the analysed viral stock.

The propagated viruses belong to different families, as they present different features such as genome length, genome type (DNA or RNA; ss or ds) and dimension; considering all these aspects, different viruses were selected to evaluate the LOD.

In particular, 4 viruses were selected, in order to cover all different genome types, belonging to different families, characterized by different genome size and different virion dimension.

To perform LOD tests, the MLV1450 was used as matrix, and spiking studies were performed using 3 different viruses: TMEV, MVM and REO3.

In particular, the matrix was spiked with 7 different concentration of the selected viruses:

- MIX 5M : $5x10^7$ viral genome copies
- MIX 1M : $1x10^6$ viral genome copies
- MIX 500K : $5x10^5$ viral genome copies
- MIX 100K : $1x10^5$ viral genome copies
- MIX 50K : $5x10^4$ viral genome copies
- MIX 10K : $1x10^4$ viral genome copies
- MIX 1K : $1x10^3$ viral genome copies

For each MIX 3 biological replicates were prepared.

For all samples, the VSQ workflow was followed up to the nucleic acid extraction step. After that, only one replicate for each spiked sample was retrotranscribed and sequenced in order to define the LOD value. Once identified the LOD, libraries corresponding to the LOD values and those above and below the defined LOD were retested, using the 2 prepared biological replicates.

4.3 RESULTS

4.3.1 SPECIFICITY RESULTS

NGS detected the specific virus of the viral stock analyzed (Table 8) demonstrating that the cut-offs that were set (%coverage, %coverage 10 and mapped reads) guarantee the specificity of the results for each virus. In the case of SuHV1, MLV1350 and HSV, the number of total reads obtained was lower as compared to the reads obtained for the other viruses: this is due to the sequencing with a 300-cycle V2 flow cell instead of a 600-cycle V3 flow cell. The host cell reads were higher in some samples (MLV-1450 and MLV-1350), thus reducing the number of reads mapping against the viral Database. However, this did not impact the identification of the virus as the number of mapping reads and the coverage was above the cut-off values of the method.

Table 8. Overview of the specificity result. The parameters considered as cut-off were % coverage, % coverage 10 and mapped reads.

Database	Virus	Genome length	% Coverage	% Coverage 10	Coverage depth	Total reads	Mapped reads	% reads w/o host
DB3	MS2	3569	99		307013	7.25+06	7.03+06	97
DB1	MLV-1350	8332	96	89	9971	8.27E+05	2.98E+05	55
	MLV-1450	8041	96	93	31750	5.61E+06	5.20E+05	31
	MLV-861	8275	100	100	103049	6.91E+06	4.70E+06	82
	VSV	10832	100	100	61529	5.70E+06	3.79E+06	85
	INFLA	13283	100	100	44978	5.80E+06	3.14E+06	78
	HSV1	147029	100	98	485	5.42E+05	3.23E+05	68
	TMEV	7238	100	100	137122	6.67E+06	5.26E+06	79
	MHV	31314	100	100	20019	6.79E+06	3.27E+06	78
	MVM	5128	100	100	251121	8.82E+06	8.29E+06	96
DB2	SUHV1	143423	100	98	560	6.00E+05	4.67E+05	87
	HPIV3	15409	100	100	21682	2.75E+06	1.91E+06	85
	REO3	23570	101	99	9754	3.89E+06	9.09E+05	58
	MLV-1450	8041	98	94	47499	4.49E+06	1.83E+06	73
	MVM	5128	100	100	228680	8.08E+06	7.59E+06	97

4.3.2 ROBUSTNESS RESULTS

Based upon the the risk assessment, three critical parameters were identified and evaluated:

- INCUBATION WITH LYSIS BUFFER AND PROTEASE DURING THE EXTRACTION PHASE: an improper incubation timing between sample and lysis buffer and protease enzyme could lead to an uncompleted sample lysis making necessary step repetition. Moreover, the incubation period is monitored only by the operator and a human error could potentially occur.

To test the robustness for this parameter, three vials of bacteriophage MS2 were extracted at different conditions. The time of treatment with lysis buffer was differently set for each sample. In particular, the time points tested were: 13, 15 and 17 minutes. Different incubation periods with LB did not affect the nucleic acid extraction yield, and actually the extracts concentration always exceeded our acceptance criteria of 2ng/µl. The extracts concentration was quantified with the Qubit® 2.0 fluorimeter using the dsDNA HS Assay Kit, ssDNA HS Assay Kit and the RNA HS Assay Kit depending on the viral genome type.

-INCUBATION WITH THE NEUTRALIZE TAGMENT BUFFER (NTB) DURING THE LIBRARY PREPARATION PHASE: an incorrect blocking of the dscDNA tagmentation could lead to problems related to libraries preparation. Also, in this case a human error can occur during the timing monitoring.

To test the robustness for this parameter, the dscDNAs with a concentration that exceeded 2ng/µl were used to prepare libraries. After the tagmentation step the Neutralize Tagment buffer was added to stop the reaction. Different periods of treatment with NT were tested: 5, 6 and, 7 minutes. The final assessment of the produced libraries was compliant with the defined acceptance criteria:

-Average fragment size greater than or equal to 200bp

-Average concentration greater than or equal to 2nM

Libraries concentration was estimated with the Qubit® 3.0 fluorimeter applying the dsDNA HS Assay Kit while the libraries average size was analyzed with the Bioanalyzer

using the Agilent DNA HS kit. Also, in this case the alteration of the incubation timing between samples and NT did not affect the library preparation.

-INCUBATION TIMING WITH NaOH DURING THE PRE-SEQUENCING PHASE: an insufficient time of incubation between libraries and NaOH, could impair DNA denaturation and result in lower cluster density. This has a significant impact on run performance, specifically on data quality and total data output.

To test the robustness for this parameter, prepared libraries were denatured with NaOH for different periods: 3, 5 and 7 minutes. Nine different libraries were sequenced in three separated MiSeq runs. Seven libraries were pooled together in equimolar amounts and then denatured in order to be sequenced using a V3 flow cell. The remaining 2 libraries were analysed in two V2 flow cells. The denaturation timing did not affect the results, as the parameters of the sequencing run were compliant with the acceptance criteria.

4.3.3 LIMIT OF DETECTION RESULTS

The libraries were prepared from MLV1450 spiked with different concentration of 3 viruses and were used to establish the limit of detection of the method. Only one replicate for each mix was sequenced for the first analysis in order to define the LOD that was at 1 million reads as at this concentration the *% coverage*, *% coverage 10X* and *number of mapped reads* obtained for each virus (TMEV,MVM, REO3 and MLV1450) exceeded the defined cut-off values (Data not shown). After the initial LOD estimation, mixes corresponding to the LOD values and those above and below the settled LOD (i.e. MIX500K, MIX1M, and MIX5M) were sequenced, using the 2 prepared biological replicates. The defined cut-off values were reached for all viruses (Figure 9) when the MIX1M were analyzed, confirming that 1 million reads were the LOD of the method.

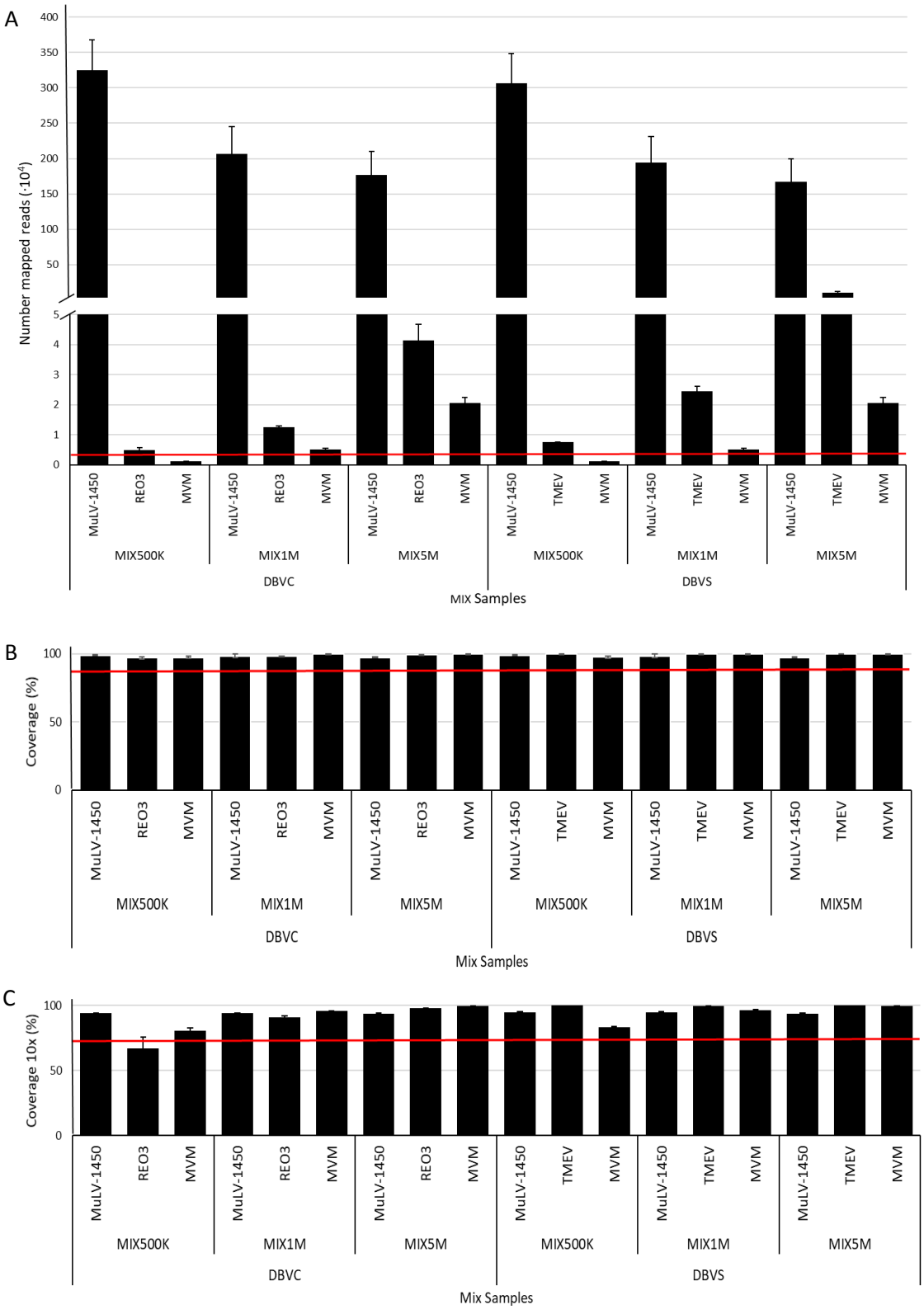


Figure 9. The histograms show the number of mapped reads (Panel A), the coverage (Panel B) and the Coverage 10X (Panel C) on the reference sequence of every virus of the analyzed MIXes. For each parameters the cut off is indicated by the red line. All error bars indicate mean ± SE of three technical replicates.

4.4 DISCUSSION

The use of well-characterized viruses as positive control in viral safety assays guarantees that the methods are suitable for its intended purpose. In the current study, our goal was to develop a process to characterize positive control virus seed stocks used in quality control tests. We successfully established a workflow enabling the identification of both DNA and RNA viruses and detect virus seed stock contamination in a single assay. In order to identify the analyzed viruses a pipeline was developed.

We defined a protocol that maximizes removal of host nucleic acids while not affecting the numbers of viral sequences. This enrichment step is based on the use of DNAse and RNAse treatment [50] and included the use of a protease that allowed the removal of the nucleases before the viral nucleic acids release. Notably, the percentage of reads that mapped against host are less than 50% for all the viral stocks except for one (MLV-1450) (Table 9), demonstrating that the process was effective at reducing non-viral sequences.

To monitor the efficiency of the protocol, the MS2 bacteriophage was used as control as described [51]. The MS2 was selected as control because this virus is simple to produce in high titer, its genome is well characterized, and it does not contain mammalian cells sequences. Moreover, an RNA virus was required to control the reverse transcription reaction.

The method described in this work showed to be accurate and robust. All 14 viruses analyzed were correctly identified with a percentage of genome coverage that exceeded 95% even if variations in method parameters were introduced. The high coverage over a full genome makes virus identification accurate, therefore providing accurate and reliable results. Moreover, deliberate variations to the method parameters introduced at critical steps, allowed us to demonstrate that the method is robust. Despite the changes, indeed, the obtained results felt within the ranges defined during validation tests.

The LOD for a panel of 14 model viruses was assessed. Based on the results, this method is capable of identifying all viral stocks analyzed with a sensitivity of 10^6 genome copies in 200 µl corresponding to 5000 genome copies per µl.

We developed a method based on the same protocol for the analysis of both RNA and DNA viral stocks. The commercial kits selected during method development, was demonstrated to release robust results. Moreover, the developed method allowing the achievement of highly specific and sensitive results.

5 MAP/HAP METHOD REPLACEMENT BY NGS

5.1 INTRODUCTION

To guarantee the absence of viral contaminants in the manufacturing process, Adventitious virus testing is performed on unprocessed bulk harvest, as well as master and working cell banks, according to ICH Q5A [12]. As mentioned in paragraph 1.1.2, several analytical methods are used including *in vivo* adventitious virus assays. The use of animals in scientific procedures, including biosafety testing, is regulated by international and national law. On September 2010 the EU adopted the Directive 2010/63/EU on the protection of animals used for scientific purpose, and is based on the application of the 3Rs principles (replacement, reduction and refinement). Regulatory authorities are increasingly recognizing the potential of NGS to detect adventitious viral agents. In particular, they recognize its potential and encourage its use to substitute or supplement *in vivo* testing,

For this reason, in this chapter we have developed an NGS based method aiming at the replacement of two in vivo assays: the Mouse antibody production test (MAP) and Hamster antibody production test (HAP). MAP and HAP are immunologically based procedures used to detect and identify species-specific viruses present in rodent cell lines. Viruses that are searched by these methods are listed in Table 9.

We here describe our NGS-based method for the replacement of MAP and HAP, from the bench to sequence analysis. In order to evaluate the applicability of NGS for the replacement of MAP and HAP in vivo assays, we performed virus spiking experiments on cell banks. The method workflow consists of these different steps: nucleic acid extraction, reverse transcription reaction, library preparation by Illumina technology and sequencing on the Illumina platform. The data obtained from NGS are then used to interrogate a curated internal database containing only the genome sequences of viruses that are listed in table 9 (meaning all the viruses detected with the *in vivo* assays). Finally, it was evaluated the sensitivity of the NGS for adventitious virus detection.

Table 9. Virus detected in MAP and HAP assays.

MAP	HAP
Mouse Cytomegalovirus (MCMV)	Lymphocytic Choriomeningitis
Mouse Hepatitis Virus (MHV)[2]	Pneumonia Virus of Mice (PVM)[2,3]
Mouse Rotavirus (EDIM)[2,3]	Reovirus Type 3 (Reo3)[1,3]
Pneumonia Virus of Mice (PVM)[2,3]	Sendai Virus[1,3]
Polyoma Virus [2]	SV5
Reovirus Type 3 (Reo3)[1,3]	
Sendai Virus[1,3]	
Thymic Virus [2]	
Minute Virus of Mice [2,3]	
Mouse Adenovirus [2,3]	
Ectromelia Virus[2,3]	
K Virus [2]	
Lactic Dehydrogenase Virus (LDM)	
Lymphocytic Choriomeningitis Virus (LCM)	
Mouse Encephalomyelitis Toolan Virus (Theilers, GDVII)[2]	
Hantaan Virus[1,3]	

[1] Viruses for which there is evidence of capacity for infecting humans or primates.
[2] Viruses for which there is no evidence of capacity for infecting humans.
[3] Virus capable of replicating in vitro in cells of human or primate origin.

Note. Adapted from ICH Q5A(R1) [12]

5.2 MATERIALS AND METHODS

5.2.1 EXPERIMENTAL DESIGN AND SETUP

The overall method workflow is summarized in Figure 9 and is detailed in the following paragraphs.

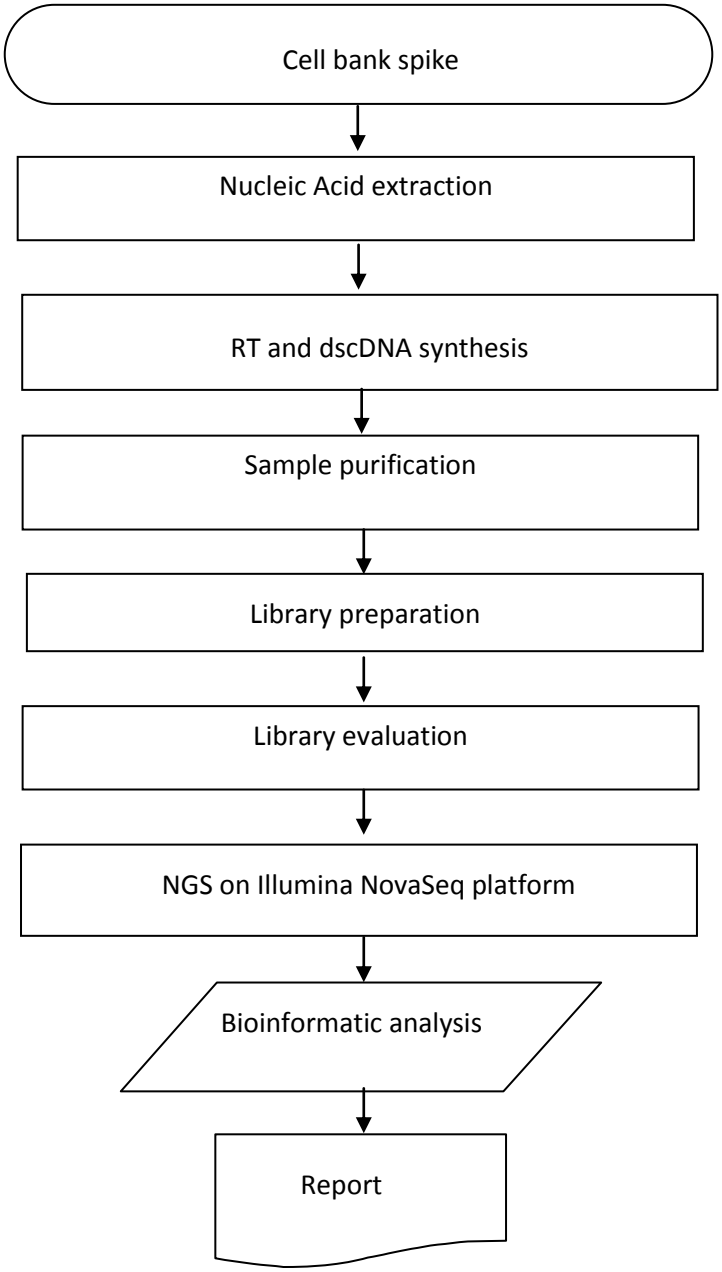


Figure 9. Summary of the method workflow

Start of process | Activity of process | Data | Document

5.2.2 SPIKING EXPERIMENT

The viruses used in the spiking study were selected from the MAP and HAP virus panel on the basis of physical and biochemical properties, to represent different types of virus families. In the spiking study were included the RNA viruses Reovirus 3 strain Abney (REO3), Theiler's murine encephalomyelitis virus (TMEV) and pneumonia virus of mice (PVM) as well as the DNA viruses Mouse adenovirus type 1 strain FL (MAV-1) and Minute virus of mice (MVM). The stocks were subjected to viral titration by ddPCR (Paragraph 3.3.5). All five viruses were included in the MAP panel, while only two viruses (REO3 and PVM) were included in the HAP panel.

In the spiking study were included two representative recombinant Chinese hamster ovary cell lines (CB1 and CB2) and two representative recombinant murine cell lines (CB3 and CB4) which are used at Merck in the manufacture of recombinant proteins. The cells were grown and propagated under defined conditions until pellet preparation. For each pellet, cells were dissociated with trypsin and $1x10^6$ cells were placed in each tube, and centrifuged for 10 min at 1000 rpm, supernatant media was removed, and the pellets were stored at −80°C.

For each in-house propagated cell lines $1x10^6$ cells were spiked with different concentrations of the selected viruses. In particular, Hamster recombinant cells were spiked only with REO3 and PVM, while the murine recombinant cell lines were spiked with all the five viruses selected. Three replicate experiments were performed for spiking at both 10^6 and 10^5 viral genome copies as it was hypothesized that the LOD for the assay may be close to this spiking level. In each spiked sample was added the enterobacteria phage MS2 as internal control.

5.2.3 VIRAL GENOME EXTRACTION AND dscDNA SYNTHESIS

Viral genomes were then extracted with a QIAmp MinElute Virus Spin Kit (Qiagen) according to the manufacturer's instruction.

Retro-transcription was carried out using two different kit the TruSeq RNA Library Preparation Kit (Illumina) and the SuperScript Double-Stranded cDNA Synthesis Kit (Termo Fisher). In the first case, the RNA was cleaved into small fragment (~200 nt) by Elute/Prime/Fragment Buffer, used heat and divalent cation. Fragment were incubated

with random primers and the first strand of dscDNA was performed using the SuperScript III Reverse Transcriptase (RT) (Thermo Fisher). After that, Second Strand Master Mix is added into the reaction mix to synthesize the second strand. Then, AMPure XP beads (Beckman Coulter) were used to purify the dscDNA. The second kit used random hexamer primer (Thermo Fisher) and the Superscipt II RT for the dscDNA synthesis. Then, QIAquick PCR purification Kit (Qiagen) was used to purify the dscDNA. Finally, the dscDNA obtained by use of two commercial kits was quantified by Qubit® 2.0 Fluorometer and then used for the library preparation.

5.2.4 LIBRARY PREPARATION AND SEQUENCING

Starting from 500 ng of dscDNA, libraries were generated using Nextera Flex DNA Library Prep Kit (Illumina), following the manufacturer's instruction. The first step of the Nextera Flex provides the DNA tagmentation by transposomes that simultaneously fragment and tag the input DNA with adapters. Once the adapters are ligated, the tagmented DNA is amplified by limited-cycle PCR program. Contextually, index adapters are added on both ends of the fragments. The clean-up purification step was performed with the 1X AMPure XP beads (Beckman Coulter) according to the manufacturer's instructions.

Purified libraries were quantified by Qubit 3.0 and sized using the Agilent 2100 Bioanalyzer. Libraries molar calculation was performed using validated spreadsheet.

Sequencing was performed on an Illumina NovaSeq 6000 system generating paired-end 2x150 bp reads. For libraries sequencing, both S1 and S2 flow cells were used. Sequence data were converted from the .bcl to FASTQ formats using the Illumina Bcl2Fastq2 (ver. 2.19.1.403) software.

5.2.5 VIRAL DATABASE

A viral database was created containing viral genome reference sequences of all the viruses included in the MAP and HAP assays (Table 9). The reference sequences were retrieved from NCBI data bank.

5.2.6 DATA ANALYSIS

Reads were trimmed of low quality bases, adapter sequences using Trimmomatic (v. 0.36). Then, reads were mapped with BWA (Burrows-Wheeler Aligner) (v. 0.7.15) to the curated Viral Database and to the sequence of the plasmid used for the production of recombinant proteins. The alignment to reference plasmid sequence, which is specific for each cell line analysed, removes known plasmid sequences. In this step the pipeline calculates the percentage of “breadth coverage" as the ratio between the number of viral genome bases covered by the reads and the total length of the genome. However, this coverage does not indicate if the entire genomic sequence is uniformly covered. For this reason, the pipeline also calculates the "% 1kb bin coverage mask". In this phase, the pipeline, using a Browser Extensible Data (BED) file, here called Mask file, to subtract from the alignment the reads which showing high similarity with host cell genome sequences. To do this, each viral reference genome in the database was divided into bins of 100 base pairs (bp) overlapped by at least 50% of the length. Subsequently, the pipeline counted for each genome the number of positive and negative bins founded, considering as positive the bins that have at least one reads aligned to the reference viral sequence and as negative the bins that did not have aligned reads. Subsequently the pipeline uses 1000 bp bins which are counted as positive if they contain positive 100 bp bins and negative otherwise. Finally, the pipeline calculates the% 1kb bin coverage mask as the ratio between the number of 1000 bp positive bins and the number of 1000 bp total bins.

Only the viruses that exceed the defined cut-off values were considered as positive: *% breadth coverage > 50 %,* % 1kb bin coverage mask *> 80%* and *number of mapped reads > 10*.

5.3 RESULTS

To evaluate the performance of both, developed experimental workflow and bioinformatic pipeline, artificially viral spiked sample were analyzed. Obtained data demonstrated the ability and specificity of the pipeline to detect all the viruses included in the MAP and HAP panels.

5.3.1 SPECIFICITY

Data obtained from different run sequencing, of the 4 cell banks spiked with one million of viral genome copies per ml, showed that the specificity of the bioinformatic pipeline is guaranteed if are used the following cutoffs:

- *% breadth coverage > 50 %*
- % 1kb bin coverage mask *> 80%*
- *number of mapped reads > 10*.

Viruses used for the spiking studies were the only viruses detected with a % breadth of coverage, % coverage bins and a number of mapped reads over the defined cutoffs for both, the MAP and HAP panels (Table 10).

Table 10. Overview of the specificity results for the HAP panel (Table 10.A) and the MAP panel (Table 10.B)**.** The parameters considered as cut-off were % Breadth of coverage, % Coverage bins and Virus mapped reads.

Table 10.A

HAP PANEL					
Cell Bank	**Virus**	**Genome length**	**Mapped reads**	**% Breadth of coverage**	**% Coverage bins**
CB1[a]	Enterobacteria phage MS2*	3569	1790	97.8	100.0
	Reo 3 **	23570	307	69.7	90.1
	PVM**	14885	152	53.6	89.2
	LCM	10610	3	0.8	5.7
	MCMV	230301	336	0.6	1.1
	Mouse Encephalomyelitis Toolan Virus	8120	12	0.4	0.0
	MHV	31357	6	0.1	0.0
CB2[b]	Enterobacteria phage MS2*	3569	1598	97.3	100.0
	Reo 3 **	23570	696	90.4	98.9
	PVM**	14886	344	76.4	96.3
	LCM	10628	1	0.3	5.7
	MCMV	230301	19	0.2	0.5
	MHV	31357	9	0.1	0.0
	Ectromelia virus	209771	1	0.0	0.3

[a] Total reads $9x10^8$
Database mapped reads $1,5x10^5$

[b] Total reads $1x10^9$
Database mapped reads $1,7x10^5$

* Process control virus
**Viruses used in spike-in studies

Table 10.B

MAP PANEL					
Cell Bank	**Virus**	**Genome length**	**Mapped reads**	**% Breadth of coverage**	**% Coverage bins**
CB3[c]	Mouse adenovirus **	30944	17304	100.0	100.0
	Mouse Encephalomyelitis Toolan Virus **	8098	2060	99.2	100.0
	Minute virus of mice**	5149	228	95.2	100.0
	Enterobacteria phage MS2*	3569	2776	94.9	100.0
	PVM**	14886	507	80.6	99.0
	Reo 3 **	23570	485	75.5	92.9
	LCMV	10610	2	0.8	5.7
	MHV	31357	9	0.1	0.0
	MCMV	230301	11	0.0	0.0
CB4[d]	Mouse adenovirus **	30944	28152	100.0	100.0
	Minute virus of mice**	5149	546	99.0	100.0
	Mouse Encephalomyelitis Toolan Virus **	8100	5255	98.0	100.0
	Enterobacteria phage MS2*	3569	7602	97.6	100.0
	PVM **	14887	1397	95.8	100.0
	Reo 3**	23570	1361	90.8	99.6
	MHV	31357	14	0.1	0.0
	MCMV	230301	9	0.0	0.0
	Ectromelia virus	209771	3	0.0	0.3

[c] Total reads $9x10^8$
Database mapped reads $4{,}4x10^5$

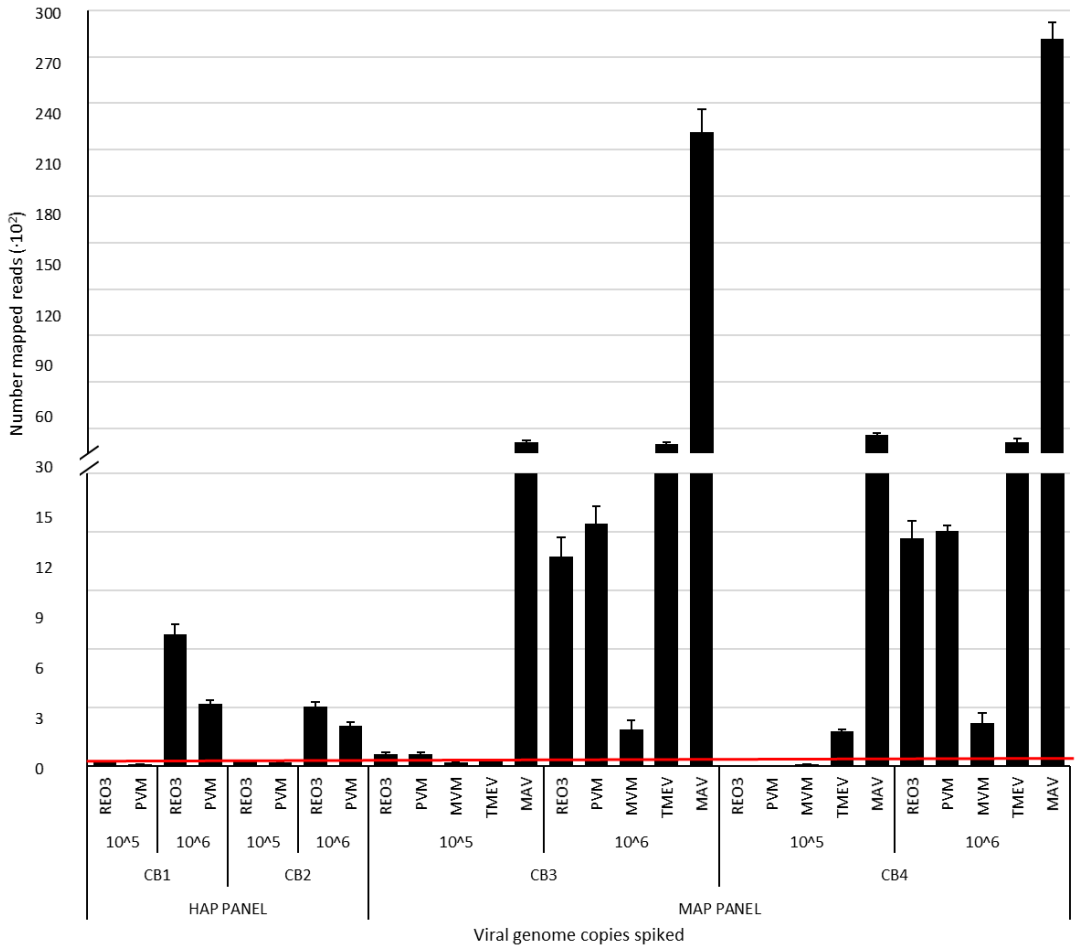
[d] Total reads $1x10^9$
Database mapped reads $7{,}2x10^5$

* Process control virus
**Viruses used in spike-in studies

5.3.2 ASSESSMENT OF LOD

The LOD was estimated using two concentrations of viral spikes, 10^5 and 10^6 viral genome copies per ml. To eliminate the endogenous virus sequences, 4 cell banks were analyzed to create a background of the endogenous retroviral sequences. Running the analysis with the defined cutoffs NGS detects all viruses at 10^6 viral genome copies per ml in each cell line spiked. The number of mapped reads, the breadth of coverage and the coverage bins for all viruses were over the values defined for each cutoff (Figure 10). At 10^5 viral genome copies per ml, instead, only MAV exceed all the defined cut-off value.

A



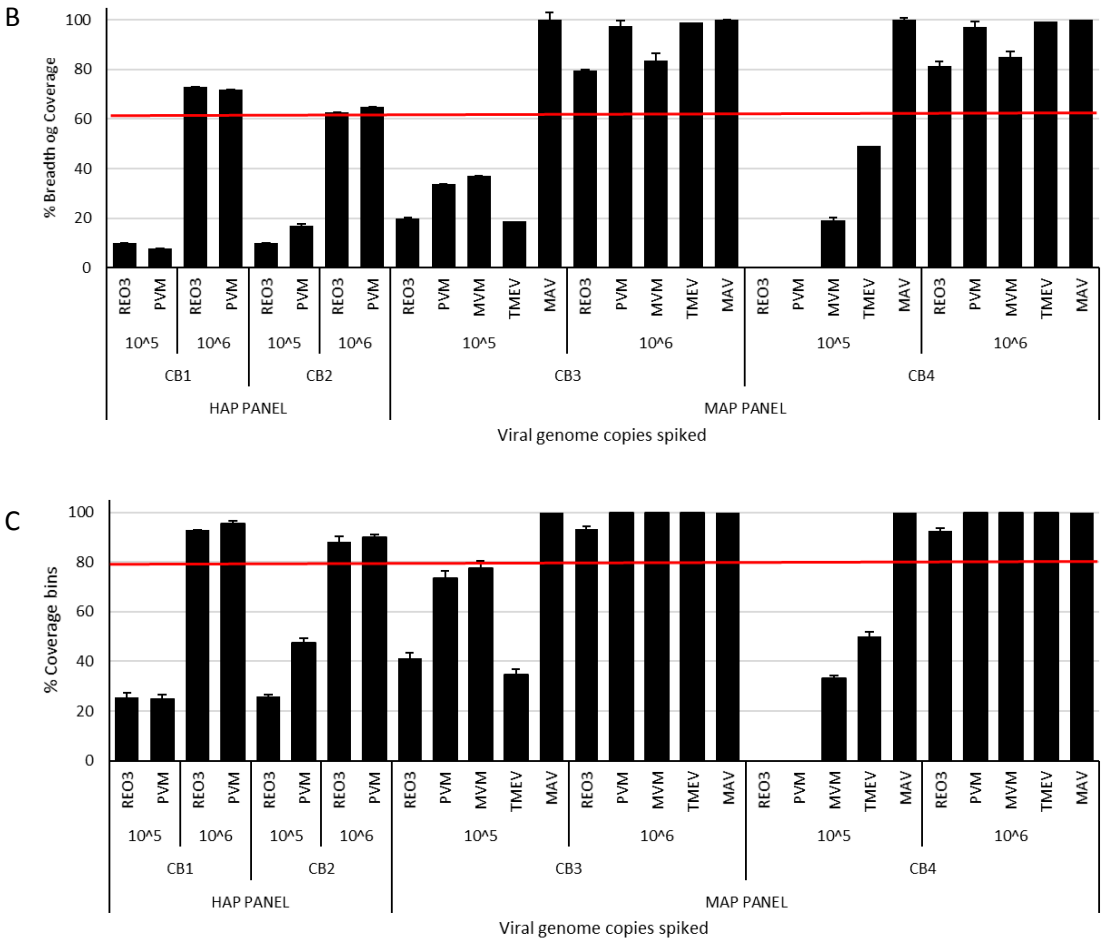


Figure 10. The histograms show the number of mapped reads (Panel A), the percentage breadth of coverage (Panel B) and the percentage coverage bins (Panel C) of each virus used in the spiking studies. For each parameters the cut off is indicated by the red line. All error bars indicate mean ± SE of three technical replicates.

5.3.3 COMPARISON RT

Data derived by the sequencing of libraries, obtained from sample retrotranscribed with two different reverse transcription kit, were compared (Figure 11). The number of mapped reads was always over the defined cutoff value. The variability between the number of mapped reads was in most cases not statistically significant as well as the % of breadth coverage and the % of coverage bins (data not shown). However, there is a significant difference in the number of reads obtained from the RNA viruses in some cases.

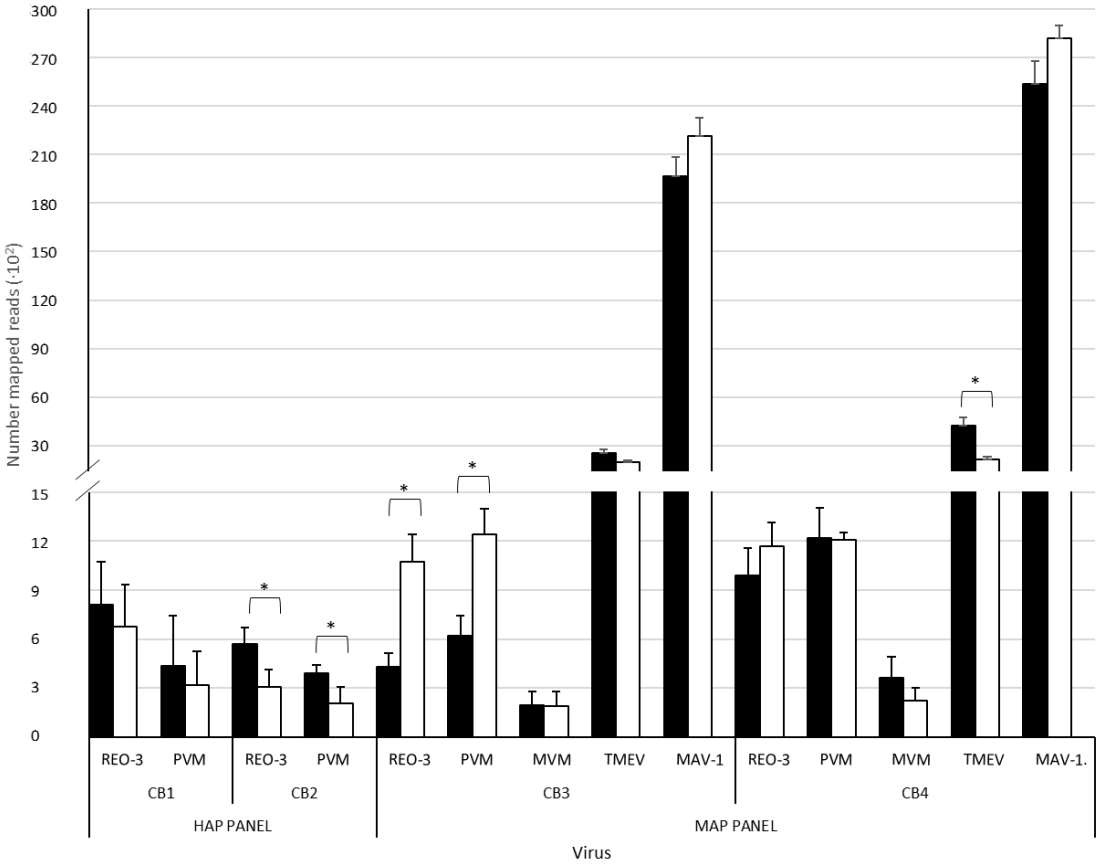


Figure 11. Comparison of number mapped reads obtained from sample retrotranscribed with two different reverse transcription kit. All error bars indicate mean ± SE of three technical replicates. Black bars indicate the TruSeq RNA Library Preparation Kit (ILLUMINA), while grey bars indicate the SuperScript Double-Stranded cDNA Synthesis Kit (Termofisher). * p<0,05

5.4 DISCUSSION

Regulatory authorities require that biotechnological products derived from animal cell lines are demonstrated to be free of adventitious viral agents before their use for therapeutic purposes. Adventitious virus testing of biological materials has typically included in vivo tests, including the mouse and hamster antibody production tests.

In the current study, our ultimate goal was to develop an *in vitro* test to replace these *in vivo* assays.

We developed a workflow in which, combined DNA and RNA viruses were correctly identified following nucleic acid extraction, double-stranded cDNA synthesis, library preparation using the Nextera Flex Kit and Illumina sequencing on NovaSeq 6000 platform. In order to identify the analyzed viruses, a bioinformatic pipeline was developed and a curated viral database created.

To eliminate the endogenous virus sequences, 4 cell banks were analyzed to create a background of the endogenous retroviral sequences. Moreover, the sensitivity of the method was assessed performing spike-in studies with different classes of viruses (ss/ds DNA or RNA). Based on the obtained results, a limit of detection of $1x10^6$ viral genome copies per ml (1 viral genome copies per cell) was identified as the value at which the method is able to correctly identify the spiked-in viruses.

For comparison, Charlebois and colleagues [52] tested the detection of 22 viruses by a spike-in study into a cellular matrix, establishing an LOD of $1x10^4$ viral genome copies (per mL?), being 2-Log lower than the LOD we show in our study. Nonetheless, our method can detect, with an $1x10^6$ viral genome copies LOD, the presence of MVM and Reo3 viruses that are instead non-detectable in Charlebois's study. The detection of MVM and Reo3 is of great importance as these are included in the list of viruses required in the ICH guidelines for MAP and HAP assays. For this reason, we conclude that the LOD of our method is the optimal one to guarantee the detection of all viruses required by regulatory authorities.

During the setup of the method, we tested different reverse transcriptase, but they did not change the result in most cases, except for RNA viruses. There is a significant difference in the number of reads obtained from the RNA viruses in some cases. Indeed, both Reo3 and PVM, showing a highest number of mapped reads with the

ILLUMINA RT kit rather than with INVITROGEN RT kit in the CB2, while in the CB3 the results were opposite. Moreover, it is not possible to define if one kit is better than the other one because we did not observe always the same result for the same virus. These differences are probably due to the efficiency of the retro-transcription step, or to the presence of single-stranded RNA, including both messenger RNA and cellular ribosomal RNA. Considering that with both kits the number of mapped reads, the % of coverage and the % coverage bin were over the defined cutoffs both will be used on the routine workload.

The experimental entire workflow takes approximately one week to be completed, reducing dramatically the turnaround time if compared with the *in vivo* assays.

This work highlights the potential of the NGS to be used in concert with *in-vivo* viral safety assay, however, additional experiments are needed to determine the robustness of our NGS method before it can be routinely applied to detect the presence of adventitious viral agents in our samples.

6 GENERAL DISCUSSIONS AND CONCLUSIONS

The goal of this PhD project is the development of methods based on molecular innovative technologies, such as ddPCR and NGS, to support or replace standard safety testing methods.

Two preparatory methods were developed for quantification and characterization of viral seed stocks used as positive controls in viral safety tests. At the same time, we used the information produced on viral seed stocks to develop an NGS based method aiming the replacement of two in vivo assays, the MAP and HAP tests.

First, our focus was on the development of a new viral titration method based on ddPCR to support traditional methods providing further information on the viral stocks produced. The traditional methods for the viral titration used in Merck were based on cell-based assays. With these methods we cannot estimate the absolute number of viral genome copies, a useful information for the establishment of the limit of detection afforded by viral safety NGS based methods. The proposed viral titration method provides absolute numbers of DNA or RNA viral genome copies in a sample and can be easily adapted to different types of viruses. Moreover, the absence of cells culture makes the method labor and time saving compared to the traditional methods. Indeed, the procedure is semi-automated and can be completed in one working day, rather than one week. The choice of use ddPCR, alternatively to quantitative real-time PCR (qPCR), was due to the evaluation of several aspects. Firstly, ddPCR does not required a standard curve, secondly ddPCR is more accurate as well as reproducible than qPCR and finally ddPCR is less affected by the PCR inhibitors [53]. In conclusion we report the implementation of a new method for the accurate quantification of viral seed stocks.

For the characterization of viral seed stocks, we proposed an NGS-based method to characterize the viral stocks, verify the virus species, and exclude the presence of other viruses propagated in the same laboratories. The method required the optimization of the experimental workflow for viral library preparation and the development of a dedicated bioinformatic pipeline for data analysis. We developed an efficient and easy-

to-use process allowing the characterization of both DNA and RNA viruses using the same workflow. The method was found to be specific and robust, with an LOD of $1x10^6$ viral genome copies in 200 µl. The method is currently used to characterize viruses used in the quality control routine test.

Finally, as there is a growing attention to decrease the use of animal testing wherever possible in accordance to the 3Rs principle (Refinement, Reduction, Replacement)[15], we developed an in vitro assay to replace two in vivo methods. Alternative testing methods have many advantages over traditional animal tests. Firstly, i*n-vitro tests* are less variable and more sensitive, secondly, they usually take less time and money to be completed and, finally, they do not use animals. In particular, we worked on the setup of the experimental conditions required to validate this method, allowing its use in a GMP environment. In this study, after the development of the experimental workflow, we used artificially spiked recombinant cell banks to evaluate the lowest genome copy number of known viruses that could be detected using high throughput sequencing based on Illumina platforms. The MAP/HAP NGS based assay showed several advantages respect to the in vivo assays. Firstly, the selection of animals used can affect the sensitivity of in vivo assays, because some viruses induce different levels of antibody response with do not require animal infection. Secondly, NGS provides a fast turnaround time of the results respect to in vivo assay. Indeed, turnaround time can be reduced from 4-6 weeks to 1-2 weeks. Finally, this method is in line with the 3R principles eliminating the use of animals.
Of note, the developed method does not give information about the viruses' viability. However, the presence, in sequenced samples, of a high number of reads that maps to a reference sequence with a high percentage of coverage means that a contamination was occurred. This results in an alarm of suspected contamination even if we are not able to define the viruses' viability.

In conclusion, new technologies could be considered not only supplementary method of traditional biosafety testing, but also an alternative testing strategy, especially NGS. The recommended assays are generally able to detect the presence of adventitious

viral agents, however in some cases in the past they failed. For these reason more sensitive assays based on next generation sequencing technologies has been introduced. Even if its use is not yet mature to be used in a GMP regulated environment, NGS is able to meet their requirements. A continuous dialogue between regulatory authorities and industry is needful to implement the introduction of this technology for quality and safety testing.

7 REFERENCES

1. EMEA. *ICH Topic Q 5 D Quality of Biotechnological Products: Derivation and Characterisation of Cell Substrates Used for Production of Biotechnological/Biological Products Step 5 NOTE FOR GUIDANCE ON QUALITY OF BIOTECHNOLOGICAL PRODUCTS: DERIVATION AND CHARACTER*. http://www.emea.eu.int (1998).

2. Covic, A. *et al.* Biosimilars and biopharmaceuticals: what the nephrologists need to know--a position paper by the ERA-EDTA Council. *Nephrol. Dial. Transplant* **23**, 3731–7 (2008).

3. Dumont, J., Euwart, D., Mei, B., Estes, S. & Kshirsagar, R. Human cell lines for biopharmaceutical manufacturing: history, status, and future perspectives. *Critical Reviews in Biotechnology* vol. 36 1110–1122 (2016).

4. Jungbauer, A. Continuous downstream processing of biopharmaceuticals. *Trends in Biotechnology* vol. 31 479–492 (2013).

5. Li, F., Vijayasankaran, N., Shen, A., Kiss, R. & Amanullah, A. Cell culture processes for monoclonal antibody production. *mAbs* vol. 2 466–479 (2010).

6. AL-Sabbagh, A., Olech, E., McClellan, J. E. & Kirchhoff, C. F. Development of biosimilars. *Semin. Arthritis Rheum.* **45**, S11–S18 (2016).

7. Morrow, T. & Felcone, L. H. Defining the difference: What Makes Biologics Unique. *Biotechnol. Healthc.* **1**, 24–9 (2004).

8. Lowe, C. R., Lowe, A. R. & Gupta, G. New developments in affinity chromatography with potential application in the production of biopharmaceuticals. *Journal of Biochemical and Biophysical Methods* vol. 49 561–574 (2001).

9. Haryadi, R. *et al.* Optimization of heavy chain and light chain signal peptides for high level expression of therapeutic antibodies in CHO cells. *PLoS One* **10**, e0116878 (2015).

10. Mutti, M. & Corsini, L. Robust Approaches for the Production of Active Ingredient and Drug Product for Human Phage Therapy. *Front. Microbiol.* **10**, 2289 (2019).

11. Wisher, M. Biosafety and product release testing issues relevant to replication-competent oncolytic viruses. *Cancer Gene Ther.* **9**, 1056–61 (2002).

12. EMEA. ICH Topic Q 5 A (R1) Quality of Biotechnological Products: Viral Safety Evaluation of Biotechnology Products Derived from Cell Lines of Human or Animal Origin. *Wiley Encycl. Clin. Trials* 1–29 (2008) doi:10.1002/9780471462422.eoct456.

13. WHO. Recommendations for the evaluation of animal cell cultures as substrates for the manufacture of biological medicinal products and for the characterization of cell banks Table of contents. (2010).

14. EMEA. *ICH Topic Q 2 (R1) Validation of Analytical Procedures: Text and Methodology Step 5 NOTE FOR GUIDANCE ON VALIDATION OF ANALYTICAL PROCEDURES: TEXT AND METHODOLOGY*. http://www.emea.eu.int (1995).

15. W.M.S. Russell and R.L. Burch. The Principles of Humane Experimental Technique.

https://caat.jhsph.edu/principles/the-principles-of-humane-experimental-technique (1959).

16. Lovatt, A. A., Scientific, P. H. D. & Vitrology, S. G. S. Cell Substrate Characterisation in Vitro-Based Virus Detection Methods. (2012).

17. König, K. *et al.* Implementation of amplicon parallel sequencing leads to improvement of diagnosis and therapy of lung cancer patients. *J. Thorac. Oncol.* **10**, 1049–1057 (2015).

18. Rivas, M. A. *et al.* Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nat. Genet.* **43**, 1066–1073 (2011).

19. Pareek, C. S., Smoczynski, R. & Tretyn, A. Sequencing technologies and genome sequencing. *Journal of Applied Genetics* vol. 52 413–435 (2011).

20. Prifti, E. & Zucker, J.-D. *The new science of metagenomics and the challenges of its use in both developed and developing countries*. (2013).

21. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics* vol. 10 57–63 (2009).

22. Sarda, S. & Hannenhalli, S. Next-Generation Sequencing and Epigenomics Research: A Hammer in Search of Nails. *Genomics Inform.* **12**, 2 (2014).

23. Kawashima, E., Farinelli, L. & Mayer, P. WO1998044151A1 - Method of nucleic acid amplification - Google Patents. *Patent* https://patents.google.com/patent/WO1998044151A1/en (1998).

24. Illumina. *An introduction to Next-Generation Sequencing Technology*. www.illumina.com/technology/next-generation-sequencing.html.

25. Chen, F. *et al.* The History and Advances of Reversible Terminators Used in New Generations of Sequencing Technology. *Genomics, Proteomics and Bioinformatics* vol. 11 34–40 (2013).

26. Reuter, J. A., Spacek, D. V. & Snyder, M. P. High-Throughput Sequencing Technologies. *Mol. Cell* **58**, 586–597 (2015).

27. Illumina. Paired-End vs. Single-Read Sequencing Technology. https://www.illumina.com/science/technology/next-generation-sequencing/plan-experiments/paired-end-vs-single-read.html.

28. Rački, N., Dreo, T., Gutierrez-Aguirre, I., Blejec, A. & Ravnikar, M. Reverse transcriptase droplet digital PCR shows high resilience to PCR inhibitors from plant, soil and water samples. *Plant Methods* **10**, 42 (2014).

29. Pholwat, S., Stroup, S., Foongladda, S. & Houpt, E. Digital PCR to Detect and Quantify Heteroresistance in Drug Resistant Mycobacterium tuberculosis. *PLoS One* **8**, e57238 (2013).

30. Strain, M. C. *et al.* Highly Precise Measurement of HIV DNA by Droplet Digital PCR. *PLoS One* **8**, (2013).

31. Anderson, J. & Hussain, M. A Direct Droplet Digital PCR Method for E. coli Host Residual

DNA Quantification. *Pharmacol. & Pharm.* **09**, 117–123 (2018).

32. Hussain, M., Fantuzzo, R., Mercorelli, S. & Cullen, C. A direct droplet digital PCR method for quantification of residual DNA in protein drugs produced in yeast cells. *J. Pharm. Biomed. Anal.* **123**, 128–131 (2016).

33. Hindson, B. J. *et al.* High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal. Chem.* **83**, 8604–8610 (2011).

34. Biorad. h*ttps://www.bio-rad.com/it-it/product/qx200-droplet-digital-pcr-system?ID=MPOQQE4VY* https://www.bio-rad.com/it-it/product/qx200-droplet-digital-pcr-system?ID=MPOQQE4VY.

35. Pinheiro, L. B. *et al.* Evaluation of a droplet digital polymerase chain reaction format for DNA copy number quantification. *Anal. Chem.* **84**, 1003–1011 (2012).

36. Dulbecco, R. & Vogt, M. Plaque formation and isolation of pure lines with poliomyelitis viruses. *J. Exp. Med.* **99**, 167–182 (1954).

37. Reed, L. J. & Muench, H. A SIMPLE METHOD OF ESTIMATING FIFTY PER CENT ENDPOINTS12. *Am. J. Epidemiol.* **27**, 493–497 (1938).

38. Finney, D. J. Statistical method in biological assay. *Stat. method Biol. assay.* (1952).

39. Bryan, W. R. INTERPRETATION OF HOST RESPONSE IN QUANTITATIVE STUDIES ON ANIMAL VIRUSES. *Ann. N. Y. Acad. Sci.* **69**, 698–728 (1957).

40. Hindson, C. M. *et al.* Absolute quantification by droplet digital PCR versus analog real-time PCR. *Nat. Methods* **10**, 1003–1005 (2013).

41. Hartley, J. W. & Rowe, W. P. A new mouse virus apparently related to the adenovirus group. *Virology* vol. 11 645–647 (1960).

42. Crawford, L. V. A minute virus of mice. *Virology* **29**, 605–612 (1966).

43. Theiler, M. Spontaneous encephalomyelitis of mice, a new virus disease. *J. Exp. Med.* **65**, 705–720 (1937).

44. Easton, A. J., Domachowske, J. B. & Rosenberg, H. F. Animal Pneumoviruses: Molecular Genetics and Pathogenesis. *Clinical Microbiology Reviews* vol. 17 390–412 (2004).

45. Compans, R. W., Harter, D. H. & Choppin, P. W. Studies on pneumonia virus of mice (PVM) in cell culture. II. Structure and morphogenesis of the virus particle. *J. Exp. Med.* **126**, 267–276 (1967).

46. Miyata, H. *et al.* New Isolates of Pneumonia Virus of Mice (PVM) from Japanese Rat Colonies and their Characterization. *Exp. Anim.* **44**, 95–104 (1995).

47. Sanmiguel, J., Gao, G. & Vandenberghe, L. H. Quantitative and digital droplet-based AAV genome titration. in *Methods in Molecular Biology* vol. 1950 51–83 (Humana Press Inc., 2019).

48. Thompson, K. A. S. & Yin, J. Population dynamics of an RNA virus and its defective interfering particles in passage cultures. *Virol. J.* **7**, 257 (2010).

49. Fleige, S. & Pfaffl, M. W. RNA integrity and the effect on the real-time qRT-PCR

performance. *Molecular Aspects of Medicine* vol. 27 126–139 (2006).

50. Hall, R. J. *et al.* Evaluation of rapid and simple techniques for the enrichment of viruses prior to metagenomic virus discovery. *J. Virol. Methods* **195**, 194–204 (2014).

51. Dreier, J., Störmer, M. & Kleesiek, K. Use of bacteriophage MS2 as an internal control in viral reverse transcription-PCR assays. *J. Clin. Microbiol.* **43**, 4551–4557 (2005).

52. Charlebois, R. L. *et al.* Sensitivity and breadth of detection of high-throughput sequencing for adventitious virus detection. *npj Vaccines* **5**, 1–8 (2020).

53. Dobnik, D. *et al.* Accurate quantification and characterization of adeno-associated viral vectors. *Front. Microbiol.* **10**, (2019).