Biologicals 69 (2021) 70–75

Contents lists available at ScienceDirect

Biologicals

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

Short paper

Metagenomic characterization of parental and production CHO cell lines for detection of adventitious viruses

Katarina Bačnik [a,b,1], Denis Kutnjak [a,*,1], Barbara Jerič Kokelj [a], Nika Tuta [c], Tan Lončar [c], Matjaž Vogelsang [c], Maja Ravnikar [a,d]

[a] National Institute of Biology, Department of Biotechnology and Systems Biology, Večna Pot 111, 1000, Ljubljana, Slovenia
[b] Jožef Stefan International Postgraduate School, Jamova cesta 39, 1000, Ljubljana, Slovenia
[c] Novartis Technical Research & Development, Biologics Technical Development, Lek Pharmaceuticals d.d., Kolodvorska 27, 1234, Mengeš, Slovenia
[d] University of Nova Gorica, Vipavska Cesta, 5000, Nova Gorica, Slovenia

ARTICLE INFO

Keywords:
Adventitious viruses
CHO cells
High-throughput sequencing
Metagenomics
Parental cell lines

ABSTRACT

Viral contamination is a major concern for biological products. Therefore, virus testing of raw materials and cells is essential for the safety of the final product. We used high-throughput sequencing to detect viral-like sequences in selected CHO cell lines. Our aim was to test various approaches of sample preparation, to establish a pipeline for metagenomic analysis and to characterize standard viral metagenome of production and parental CHO cell lines. The comparison of the metagenomics composition of the differently prepared samples showed that among four tested approaches sequencing of ribosomal RNA depleted total RNA is the most promising approach. The metagenomics investigation of one production and three parental CHO cell lines of diverse origin did not indicate the presence of adventitious viral agents in the investigated samples. The study revealed an expected background of virus-like nucleic acids in the samples, which originate from remains of expression vectors, endogenized viral elements and residuals of bacteriophages.

1. Introduction

Chinese hamster ovary (CHO) cells are one of the main cell lines used in the production of recombinant therapeutics [1] and have the highest occurrence of reported virus contaminations compared with other mammalian cell lines [2]. The utilization of the current combination of in vitro, in vivo, and PCR assays for the identification of adventitious viruses has a limited range of detection. High-throughput sequencing (HTS) on the other hand is able to detect all types of nucleic acid sequences in a sample [3] and has gained significant attention in the field of biologicals since a porcine circovirus 1 contamination has been detected in a licensed pediatric vaccine [4] and a novel rhabdovirus was discovered in the Sf9 insect cell line [5]. Since then, HTS has been utilized as a supplemental test and is now considered as an alternative method for adventitious virus testing [6–9]. The aim of this study was to test various approaches of sample preparation, to establish a pipeline for metagenomic analysis and to characterize standard viral metagenome of selected CHO cell lines. This study serves as a baseline for further investigations of CHO cell lines for adventitious viruses using HTS.

2. Materials and methods

2.1. CHO cell lines

Four different CHO cell lines (T-CHO, P-CHO-1, P-CHO-2, P-CHO-3) were used in the study. P-CHO-1, P-CHO-2 (engineered) and P-CHO-3 are parental cell lines originating from different lineages, including CHO-K1, CHO-DUKXB1 and CHO-DG44. As previously described [10], the CHO-K1 and CHO-DG44 (dihydrofolate reductase (DHFR) deficient) cell lineages were generated from the spontaneously immortalized cells from Chinese hamster ovaries by single-cell cloning and gamma radiation, respectively. The DHFR-deficient CHO-DUKXB1 lineage was generated from CHO-K1 cells by chemical mutagenesis. The test (T-CHO) cell line is a monoclonal antibody producing cell line derived from the engineered P-CHO-2 parental cell line.

2.2. Sample preparation and nucleic acids isolation

In the first batch of analyses, we compared different sample

* Corresponding author.
E-mail address: denis.kutnjak@nib.si (D. Kutnjak).
[1] Both authors contributed equally to this paper.

https://doi.org/10.1016/j.biologicals.2020.11.001
Received 25 August 2020; Received in revised form 26 October 2020; Accepted 9 November 2020
Available online 24 November 2020

preparation approaches using T-CHO cell line. Total (tot)RNA extracted from cells and nucleic acids isolated from either nontreated, ultracentrifuged or filtrated growth media were used as inputs for HTS (detailed in Fig. 1 and Supp. materials and methods). Additional two negative controls: processed growth medium and fresh growth medium were treated in the same way and stored for possible troubleshooting and additional confirmation analyses.

In the second batch of analyses, three parental cell lines were processed using the selected approach (totRNA sequencing of CHO cells) based on the results of the first batch of the analyses. First, cells were separated from growth medium by centrifugation (5 min, 4 °C, 400×*g*) and were further used for nucleic acids isolation (P-CHO-1, P-CHO-2, P-CHO-3).

For all analyses, RNA from cell samples was isolated using RNeasy Plus Mini kit (Qiagen) (batch 1 and batch 2). Sample of nuclease free water spiked with luciferase was used as a negative control of nucleic acid isolation from parental cell lines (batch 2). For simultaneous purification of DNA and RNA from growth media samples (batch 1) QIAamp MinElute Virus Spin Kit (Qiagen) was used.

2.3. Sequencing and bioinformatics analysis

For all samples (Table 1), depletion of ribosomal RNA using Ribo-Zero Gold rRNA Removal Kit (Illumina) and reverse transcription of RNA were performed. For first batch of production cell line samples sequencing libraries were prepared using Illumina Nextera DNA Library Prep Kit and sequenced by Illumina MiSeq (2 × 250 bp). For parental cell lines samples Illumina TruSeq Stranded total RNA library preparation kit was used and libraries were sequenced using Illumina HiSeq 2500 (2 × 150 bp). For first batch of samples less data was generated compared to sequencing of parental cell lines, where we increased the sequencing depth to get a better insight into the presence of virus-like nucleic acids.

The pipeline used for data analysis of all samples is presented in Fig. 1 and described in details in supplementary materials and methods. Additionally, reads from parental cell line samples were used for de novo assembly using software SPAdes [11] and were further classified (Supp. materials and methods). The presence of retroviral-like reads corresponding to the sequence of known transcriptionally active retrovirus sequences from CHO genomes were investigated by mapping of reads from parental cell lines to endogenous retroviral sequences (Supp.

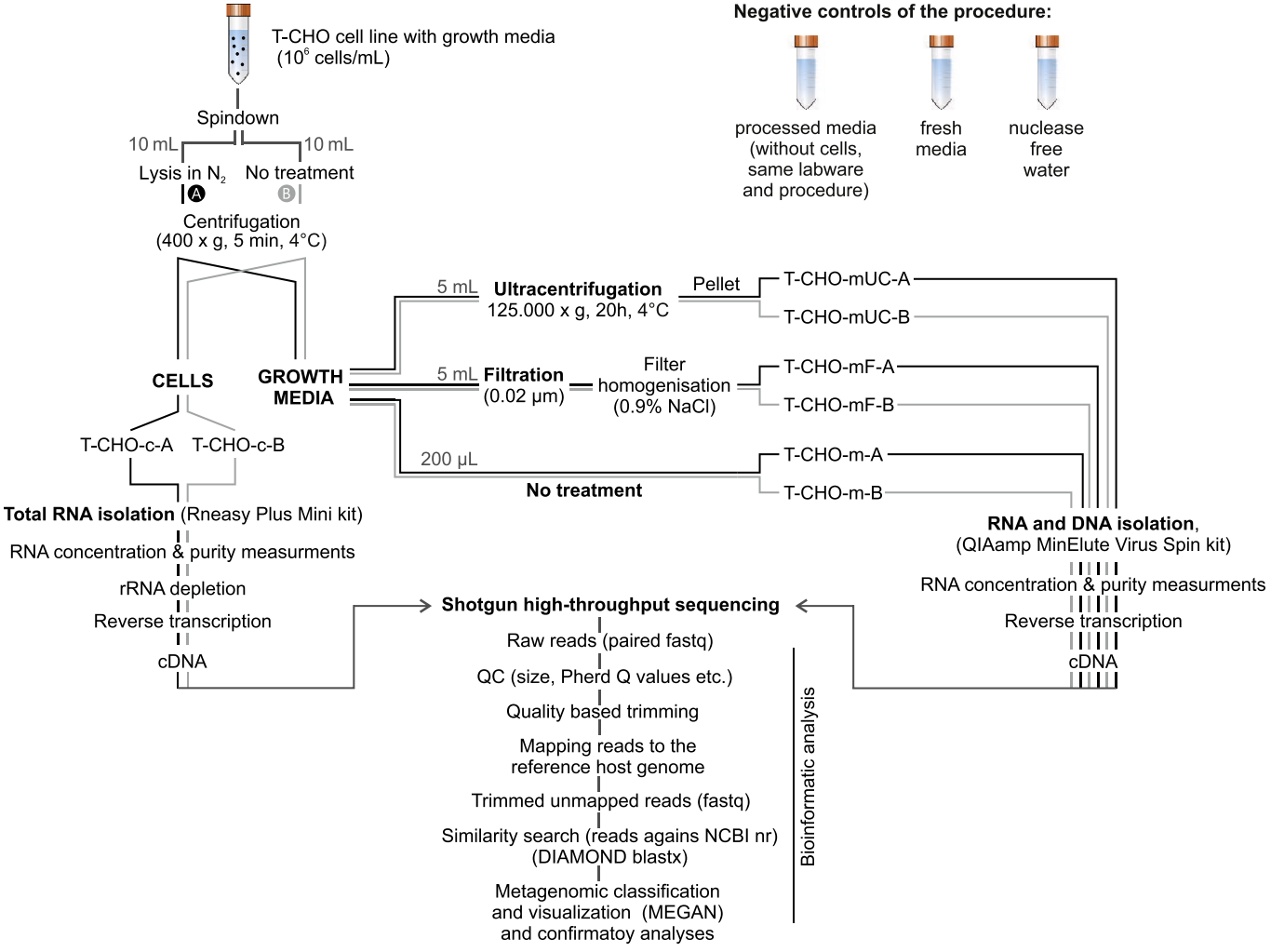

Fig. 1. Detailed schematic representation of sample preparation workflow and bioinformatics pipeline. Two aliquots of cell samples were first briefly centrifuged. In order to lyse the cells, one of the aliquots was sequentially frozen three times in liquid nitrogen (samples originating from this aliquot are marked with A). The other aliquot was left untreated (samples originating from this aliquot are marked with B). Both samples were further centrifuged to separate cells (T–CHO–*c*-A, T–CHO–*c*-B) from growth media. Total RNA was isolated from pelleted cells (T–CHO–*c*-A, T–CHO–c-B). Aliquots of growth media samples were stored for nucleic acids isolation (T–CHO–m-A, T–CHO–m-B). The rest of the growth media was either ultracentrifuged or filtrated through inorganic filter. After ultracentrifugation, supernatant was removed from the tubes and pellets were used for nucleic acid isolation (T–CHO–mUC-A, T–CHO–mUC-B). After sample filtration, the filter was transferred in a sterile Petri dish and fragmented into small fragments, which were then fragmented and used for the nucleic acid isolation (T–CHO–mF-A, T–CHO–mF-B).

Table 1
Summary of reads' mapping in CLC Genomics Workbench and reads' classification using DIAMOND for different samples tested in this study.

Sample designation	Description	Freezing	Processing	Isolation type	Total N of trimmed reads	Mapped to *C. griseus* genome/luciferase (CLC)	Percentage of unmapped assigned reads (DIAMOND)	Percentage of unmapped reads assigned to Eukaryota (DIAMOND)	Percentage of unmapped reads not assigned to Eukaryota (DIAMOND)	Number of unmapped reads assigned to Viruses (DIAMOND)
T–CHO–c-A	CHO cells	YES	–	RNA	2.912.756	55.05	73.72	68.96	4.76	292
T–CHO–c-B	CHO cells	NO	–	RNA	2.454.601	52.82	77.04	71.98	5.06	137
T–CHO–m-A	medium (cells)	YES	–	DNA + RNA	2.673.285	77.02	19.72	13.84	5.88	164
T–CHO–m-B	medium (cells)	NO	–	DNA + RNA	1.714.020	73.51	20.24	13.31	6.92	103
T–CHO–mF-A	medium (cells)	YES	filtration	DNA + RNA	2.507.973	84.78	17.07	11.76	5.31	46
T–CHO–mF-B	medium (cells)	NO	filtration	DNA + RNA	3.036.897	64.18	22.59	12.79	9.80	214
T–CHO–mUC-A	medium (cells)	YES	ultracent.	DNA + RNA	2.089.904	84.68	13.55	11.70	1.85	53
T–CHO–mUC-B	medium (cells)	NO	ultracent.	DNA + RNA	2.377.070	76.39	12.76	11.15	1.61	52
P-CHO-1	CHO cells	NO	–	RNA	181.580.192	67.61	70.02	65.73	4.29	150
P-CHO-2	CHO cells	NO	–	RNA	193.718.096	69.36	68.19	63.59	4.60	2629
P-CHO-3	CHO cells	NO	–	RNA	169.045.514	68.31	68.37	63.74	4.63	214
NKI-luc	NKI and luc	NO	–	RNA	179.049.102	96.23	42.59	0.19	42.40	72

materials and methods).

Reads classified as viral in Diamond/Megan analysis were extracted and further analyzed to possibly confirm/reject that they correspond to viral genomes. Those viral-like reads were then imported into CLC Genomics Workbench (Qiagen) and compared against the latest NCBI nt and nr databases using blastn and blastx, respectively. Reads showing significant similarity only to viral taxa in blastn or blastx analysis were further investigated to determine if they might originate from CHO genomic DNA. They were compared for similarity against selected known CHO short read archive (SRA) datasets (Supp. Table 1) using blastn and against reference viral database (RVDB) U-RVDBv19.0 and U-RVDBV18.0 using blastn and blastx, respectively.

3. Results and discussion

3.1. Comparison of sample preparation approaches and standard viral metagenome of selected CHO cell line

Comparing the number of viral-like reads and detection of different viral taxonomic groups can give some insight about which of the tested sample preparation approaches would be more appropriate for discovery of adventitious viral agents. When comparing different sample preparation approaches, the differences are not striking, however, highest number of viral-like reads (292) and detected viral taxonomic groups (Fig. 2) was found in sample of isolated totRNA from cells. Analyzing ribosomal RNA-depleted totRNA from cells required the smallest amount of handling and therefore smaller chance of the contaminants introduction among tested approaches, additionally supporting our choice of this approach for subsequent analyses (for parental cell lines).

Majority of reads classified as viral in Diamond/Megan analysis corresponded to retroviral sequences (Fig. 2), which are known to be endogenized in CHO genomes. A substantial amount of reads were classified as human betaherpes virus 5 (family *Herpesviridae*), and some to polyomaviruses, both originating from the expression vectors used to manipulate the cell line. Some samples contained remains of bacterial and bacteriophage reads, which are expected to be present as residuals from chemicals and labware used in the procedure. Few reads were misclassified as viral, however further analysis showed they correspond to parts of bacterial or CHO genomes. Finally, very few reads were showing similarities to filoviruses or parvoviruses and were analyzed in detail to better determine their possible origin.

8 reads remained classified as protoparvovirus sequences after the additional blastx and blastn analysis in media samples (T–CHO–m-A and T–CHO–m-B). Sequences could be further assembled, resulting in two contigs that were most similar to sequences of rat parvovirus (82% identity at the nucleic acid level and 74% at the amino acid level) and canine parvovirus (74% identity on nucleic acid level and 71% on amino acid level), confirmed also by blasting against RVDB. Blastn analysis did not show significant similarity to the reads in CHO SRA archives. There have been several reports of endogenization of parvoviral sequences into animal genomes [12], including rodent genomes [13]. Moreover, there have been several reports on contamination of nucleic acid extraction columns with parvoviral nucleic acids [14]. Therefore, it seems highly unlikely that such reads correspond to an adventitious viral agent replicating in the CHO cells, due to the extremely low number of observed reads (corresponding to less than 12% of parvoviral genome length) and low similarity of these reads to known pathogenic viruses.

2 reads were classified as filoviral. Blastn analysis showed that their sequences were highly similar to *C. griseus* sequence (100% sequence identity), annotated as magnesium transporter, but was identified as filoviral using blastx (77% sequence identity). We could find a similar sequence in other existing CHO SRA datasets. We were unable to find a report about CHO genome integrated filoviral sequences, however, integrated filoviral sequences have been found in other rodent species

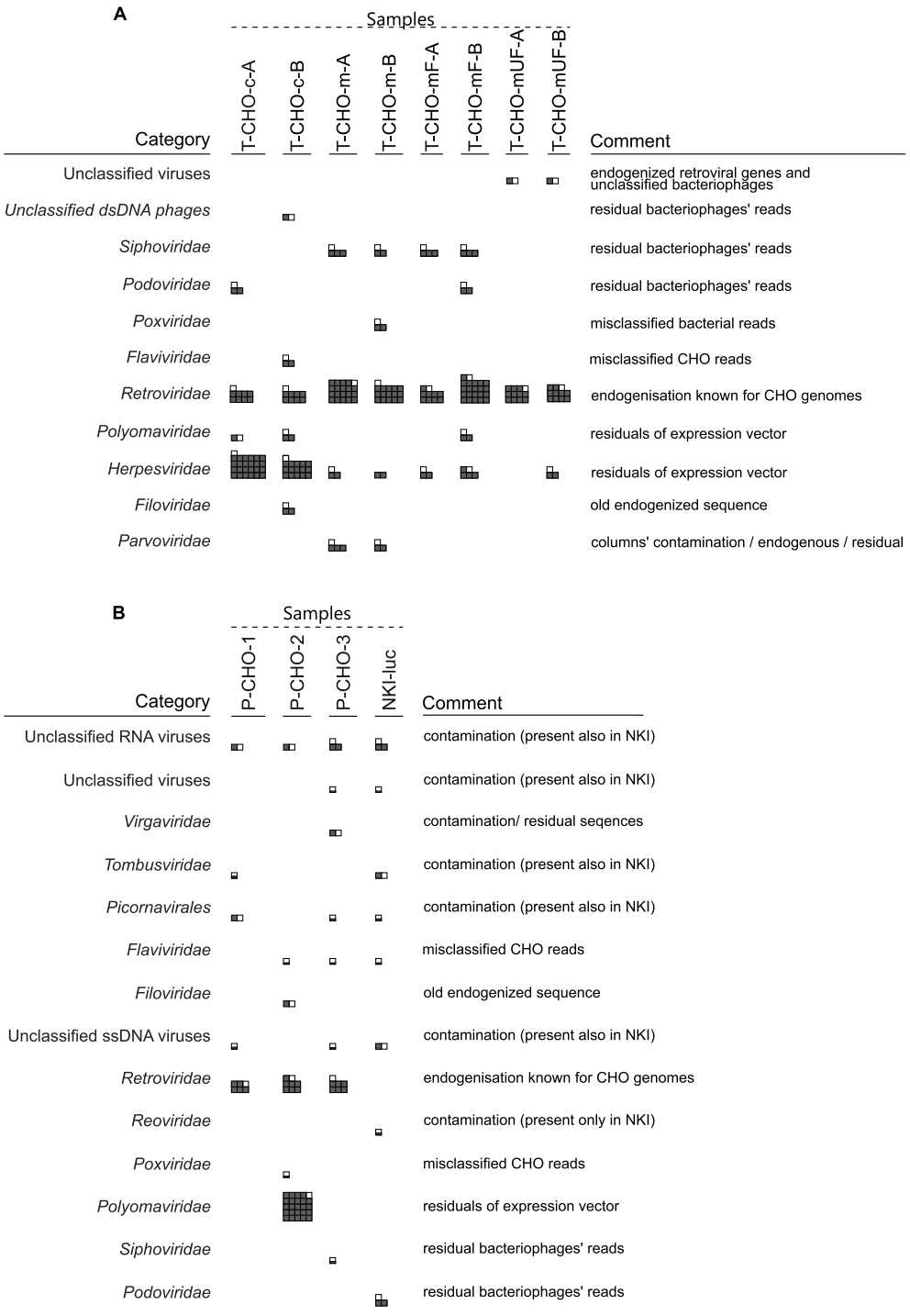

Fig. 2. Overview of the virus-like reads across the tested production cell line samples in first batch of analysis (a) and parental cell line samples with negative control (b). Reads were classified as viral by Diamond and Megan. Squares represent the abundance of the reads classified as a specific category in a specific sample (one square represents 4 reads); the read counts are square root transformed for better representation. The comment column shows final classification or interpretation of reads origin.

[15]. Thus, we suggest that such reads likely correspond to an ancient integrated filovirus sequence in the CHO genome, which likely obtained new function through the evolution of *C. griseus* genome.

3.2. Metagenomic analysis for the detection of adeventitous viruses in selected parental CHO cell lines

We used totRNA sequencing of cells to determine the metagenomic baseline of selected CHO parental cell lines. We obtained a high number of sequencing reads for analyzed samples (Table 1), including a negative control, which enabled us to further filter the results. Most of the reads classified as viral corresponded to retroviral sequences that did not map to the reference host genome using selected parameters. However, additional blastn and blastx similarity searches showed that all detected retroviral-like sequences had high similarities with CHO genomes and thus probably represent endogenized retroviral elements. Additional assembly of sequencing data (trimmed reads) from parental cell lines resulted in 13 contigs (78-1076 nt) classified as retroviral using Diamond/Megan analysis (Supplementary Table 2). Blastn and blastx similarity searches against nt/nr databases showed that all contigs have significant alignments to host transcripts or transcripts or genomes of rodent species (Supplementary Table 2). Reads from parental CHO cell lines mapped to the two previously reported transcriptionally active retroviral sequences [16,17]. However, for only one of the two

sequences (ETC109F) and only in P-CHO-2, complete length of the sequence was covered by the reads (Supp. Fig. 1).

In addition, 4 reads were classified as filoviral in P-CHO-2 sample. Reads were assembled into 1 contig (166 bp) showing similarity with *C. griseus* genome using blastn, and similarity to filoviruses using blastx (54% sequence identity), which we already observed in the previous analysis of T-CHO cell line. We could find similar sequences in other existing CHO SRA datasets. The P-CHO-2 parental cell line has been used to prepare the T-CHO cell line analyzed in previous experiment, which explains the detection of similar sequences, with exception of herpesvirus sequences that originate from the expression vectors used to manipulate the parental cell line. Also in P-CHO-2 sample, we detected a relatively high number of reads classifying as polyomaviruses corresponding to expression vector sequence, supporting the fact that P-CHO-2 is an engineered parental cell line. Reads classifying as polyomaviruses (expression vector sequence) were not detected in other parental cell lines, explaining the lower numbers of reads assigned to viruses in other two analyzed parental cell lines.

Few reads of unclassified ssDNA viruses and unclassified RNA viruses were detected both in samples and in the negative control. They were probably introduced during the extraction or sequencing and were thus regarded as contamination. Sequences classified as unclassified ssDNA viruses showed similarity with widely distributed [18] CRESS (circular Rep-encoding single-stranded DNA) viruses and sequences classified as unclassified RNA viruses were similar to sequences of members from *Picornavirales* order. Viral-like reads only present in negative control were probably introduced together with the synthetic luciferase RNA, which was spiked in the negative control.

6 reads in P-CHO-3 sample were classified as plant infecting tobamoviruses from *Virgaviridae* family (blastn – 94–96% sequence identity, blastx – 100% sequence identity) matching two different parts of tobamovirus genomes. Reads classifying as tobamoviruses are potentially representing a very low input contamination. The sequences belong to plant infecting viruses and are not of concern for human health.

To summarize, in all analyzed CHO cell line samples (T-CHO, P-CHO-1, P-CHO-2, P-CHO-3) we detected known endogenized retroviral sequences and residuals of bacteriophages. Residuals of expression vectors with viral sequences and filoviral-like reads were detected in engineered parental (P-CHO-2) and in corresponding derived production (T-CHO) cell line. Reads classifying as parvoviral were detected only in T-CHO cell line media samples (T–CHO–m-A, T–CHO–m-B). In parental cell line samples, which were investigated with higher sequencing depth, more contaminating sequences were detected; however, they were also detected in negative control, except from plant infecting tobamoviral reads only present in P-CHO-3.

4. Conclusions

The metagenomics investigation of one production and three parental CHO cell lines of diverse origin did not indicate the presence of adventitious viral agents in the investigated samples. The study revealed an expected background of virus-like nucleic acids in the samples, which originate from remains of expression vectors, endogenized viral elements and residuals of bacteriophages. Obtained results together with established bioinformatic pipeline focused on viral-like reads serve as a baseline for future investigations of CHO cell lines for adventitious viruses using HTS. The results including the overlapping virus-like sequences between parental and corresponding production cell line and the contaminations present in negative control are showing the importance of knowing the range and source of background signal in studies of adventitious virus detection. Regulatory authorities are increasingly recognizing the potential of HTS for the detection of a broad range of viruses. Therefore, HTS methodologies can complement, supplement or even replace some of the conventional adventitious virus detection assays (Khan et al., 2020). The method described in this study is currently a research tool, which has a potential to supplement other currently used adventitious virus detection methods. Since HTS is constantly evolving, there is a need for standardization and validation of the methodology, including the technical and bioinformatics steps.

Declaration of competing interest

Matjaž Vogelsang and Nika Tuta are employed by Lek Pharmaceuticals d.d. that prepared the analyzed production CHO cell line. The statement is made in the interest of full disclosure and not because the authors consider this a conflict of interest.

Acknowledgements

We would like to thank Tinkara Tinta from National Institute of Biology for providing Whatman® Anodisc inorganic filter membranes and Kristina Gruden and Špela Baebler for discussion of experiments plans.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.biologicals.2020.11.001.

Funding source

The study has been performed in the frame of SPS BioPharm.Si project (OP20.00363), funded by Ministry of Education, Science and Sport of Republic of Slovenia and European regional fund. This work was supported by Slovenian research agency (P4-0407 and young researchers' grant awarded to K.B.).

Data statement

The raw sequencing data obtained in this study are confidential.

References

[1] Grillberger L, Kreil TR, Nasr S, Reiter M. Emerging trends in plasma-free manufacturing of recombinant protein therapeutics expressed in mammalian cells. Biotechnol J 2009;4:186–201. https://doi.org/10.1002/biot.200800241.

[2] Barone PW, Wiebe ME, Leung JC, Hussein ITM, Keumurian FJ, Bouressa J, et al. Viral contamination in biologic manufacture and implications for emerging therapies. Nat Biotechnol 2020;1–10. https://doi.org/10.1038/s41587-020-0507-2.

[3] Lambert C, Braxton C, Charlebois RL, Deyati A, Duncan P, Neve F La, et al. Considerations for Optimization of High-Throughput Sequencing Bioinformatics Pipelines for Virus Detection. 2018. https://doi.org/10.3390/v10100528.

[4] Victoria JG, Wang C, Jones MS, Jaing C, McLoughlin K, Gardner S, et al. Viral nucleic acids in live-attenuated vaccines: detection of minority variants and an adventitious virus. J Virol 2010;84:6033–40. https://doi.org/10.1128/jvi.02690-09.

[5] Ma H, Galvin TA, Glasner DR, Shaheduzzaman S, Khan AS. Identification of a novel rhabdovirus in spodoptera frugiperda cell lines. J Virol 2014;88:6576–85. https://doi.org/10.1128/jvi.00780-14.

[6] Khan AS, Benetti L, Blumel J, Deforce D, Egan WM, Knezevic I, et al. Report of the international conference on next generation sequencing for adventitious virus detection in biologicals. Biologicals 2018;55:1–16. https://doi.org/10.1016/j.biologicals.2018.08.002.

[7] Khan AS, Benetti L, Blumel J, Deforce D, Egan WM, Knezevic I, et al. Report of the second international conference on next generation sequencing for adventitious virus detection in biologics for humans and animals. Biologicals 2020;55:1–16. https://doi.org/10.1016/j.biologicals.2018.08.002.

[8] McClenahan SD, Krause PR. Towards dynamic monitoring of cell cultures using high throughput sequencing. Vaccine 2019;37:1001–5. https://doi.org/10.1016/j.vaccine.2018.12.019.

[9] Brussel A, Brack K, Muth E, Zirwes R, Cheval J, Hebert C, et al. Use of a new RNA next generation sequencing approach for the specific detection of virus infection in cells. Biologicals 2019;59:29–36. https://doi.org/10.1016/j.biologicals.2019.03.008.

[10] Wurm FM, Hacker D. First CHO genome. Nat Biotechnol 2011;29:718–20. https://doi.org/10.1038/nbt.1943.

[11] Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 2012;19:455–77. https://doi.org/10.1089/cmb.2012.0021.

[12] François S, Filloux D, Roumagnac P, Bigot D, Gayral P, Martin DP, et al. Discovery of parvovirus-related sequences in an unexpected broad range of animals. Sci Rep 2016;6:1–13. https://doi.org/10.1038/srep30880.
[13] Arriagada G, Gifford RJ. Parvovirus-derived endogenous viral elements in two south American rodent genomes. J Virol 2014;88:12158–62. https://doi.org/10.1128/jvi.01173-14.
[14] Naccache SN, Greninger AL, Lee D, Coffey LL, Phan T, Rein-Weston A, et al. The perils of pathogen discovery: origin of a novel parvovirus-like hybrid genome traced to nucleic acid extraction Spin columns. J Virol 2013;87:11966–77. https://doi.org/10.1128/jvi.02323-13.
[15] Taylor DJ, Leach RW, Bruenn J. Filoviruses are ancient and integrated into mammalian genomes. BMC Evol Biol 2010;10. https://doi.org/10.1186/1471-2148-10-193.
[16] Duroy PO, Bosshard S, Schmid-Siegert E, Neuenschwander S, Arib G, Lemercier P, et al. Characterization and mutagenesis of Chinese hamster ovary cells endogenous retroviruses to inactivate viral particle release. Biotechnol Bioeng 2020;117:466–85. https://doi.org/10.1002/bit.27200.
[17] Lie YS, Penuel EM, Low MA, Nguyen TP, Mangahas JO, Anderson KP. Chinese hamster ovary cells contain transcriptionally active full-length type C proviruses. J Virol 1994;68:7840–9.
[18] Kazlauskas D, Koonin EV, Krupovic M. Multiple origins of prokaryotic and eukaryotic single-stranded DNA viruses from bacterial and archaeal plasmids. Nat Commun 2019:1–12. https://doi.org/10.1038/s41467-019-11433-0.