1

Evaluation of RNA extraction and Illumina NGS library preparation methods to detect viral RNA from different sample matrices

Sayma Afroj[1], Binsheng Gong[2], Aaron Scholl[1], Bingjie Li[1], Joshua Xu[2] and Sandip De[1]*

[1]Tumor Vaccine & Biotechnology Branch, Division of Cellular Therapy 2, Office of Cell Therapy and Human Tissues, Office of Therapeutic Products, US Food and Drug Administration, 10903 New Hampshire Avenue, Silver Spring, MD, 20993; [2]Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, Food and Drug Administration 3900 NCTR Rd Jefferson, AR 72079.

Introduction

1. Human cells, tissues, and cellular and tissue base products (HCT/Ps) are required to comply with the donor eligibility requirements as per 21 CFR part 1271 and applicable guidance documents to prevent the introduction, transmission, and spread of communicable diseases by HCT/P's.
2. To ensure the safety, purity, and quality of both investigational and licensed biological products, adventitious agents testing, and clearance are mandated at various manufacturing stages (Title 21 CFR 610.1, 610.13, 21 CFR 312.23(a)(7)(i) and (iv).
3. Next-Generation Sequencing (NGS), or high-throughput sequencing, is a technology capable of massively parallel sequencing of nucleic acid sequences. Hence, this sequencing technique holds potential applications for comprehensive virus detection in biologics.
4. A critical step for the virus detection from cell and tissue high-throughput sequencing is efficient extraction of nucleic acids from adventitious agents and next-generation sequencing library preparation. Another critical step in detection of adventitious agents is to identify the reads from extraneous agents using bioinformatics.
5. This project seeks to evaluate the RNA extraction methods and next generation sequencing library preparation methods to detect RNA from adventitious agents from diverse sample matrices. In addition, we are aiming to evaluate and develop a bioinformatics workflow for efficient detection of these agents.

Objectives

1. Evaluate different RNA extraction methods including several commercial kits for viral RNA extraction and Identify best RNA extraction method from different sample matrices.
2. Test different Illumina library preparation methods for efficient amplification of RNA viruses from extracted RNA.
3. Create an enhanced Illumina library preparation method surpassing existing technologies.
4. Evaluate existing bio-informatic work-flows and identify the most efficient method for detection of viral sequences from NGS data.
5. Develop reference materials for qualification and validation of NGS assays for detection of adventitious agents.

Materials and Methods (M&M)

Sample matrices tested

Design 1: viruses present outside the cells.
A) We spiked virus reference panel (MSA 2008) at 1 genome copy/cell into U937 cells.
B) We spiked Zika virus (ZIKV) at 1 MOI into U937 cells.

Design 2: viruses present within the cells.
We used U937 cells persistently infected with Zika virus (ZIKV-MR766). Used 100000 cells for RNA extraction.

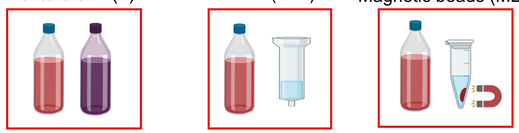
Design 3: ZIKV infected cells spiked into buffy coat matrix. Either 100 or 100000 cells were mixed into 120 ul of buffy coat.

Figure 1. Three different sample matrices were used, U937, U937 cells persistently infected with Zika virus, Zika virus infected cells spiked into buffy coat.

M&M continued

Viral RNA extraction methods tested

1. TRIzol + Chloroform (T)
2. TRIzol + Silica column (TPL)
3. Lysis buffer + Magnetic beads (MB)



4. Kit for RNA extraction from eukaryotic cells (PL)
5. Kit for viral RNA extraction from supernatant/liquid samples (Q)

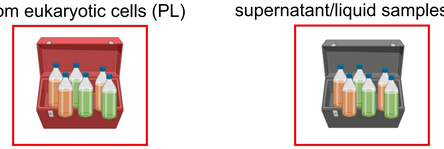


Figure 2. RNA from spiked sample matrices were extracted using five methods following manufacturer's protocols.

Generation and characterization of persistently infected cells

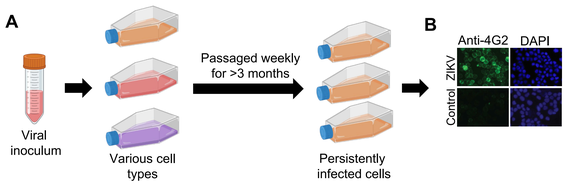


Figure 3. Generation of persistent ZIKV infected U937 cells. A) Persistent ZIKV infections were established by initiating acute infections at MOI=0.01 in U937 cells, allowing growth for over 3 months with weekly passaging. B) Immunofluorescence assay of persistently infected ZIKV cells compared to uninfected control cells. The green color panels (left) represent cells stained with anti-4G2 flavivirus antibody and the blue color panels (right) represent cells stained with DAPI. Images were taken with 20x objective using exposures of 1/10 seconds (anti-4G2) and 1/20 seconds (DAPI). 100% cells are infected with ZIKV, ~15% cells are highly infected.

Bioinformatics workflow used for detecting viral reads from sequencing data

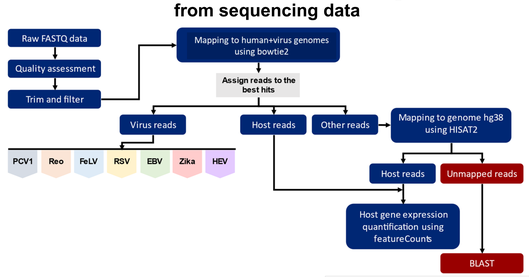


Figure 4. A sequential workflow of the steps from the initial input of raw data to analysis of the processed data is shown. Genomes of viruses used for data analysis is shown and highlighted with different colors..

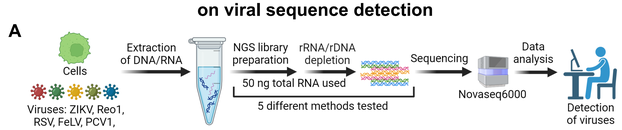
Results

Noticeable differences observed in the efficiency of viral RNA extraction across the methods tested

Viruses		RNA extraction Methods				
		T	TPL	PL	Q	MB
Reference viral panels ATCC-MSA 2008 spiked in U937, design 1A	Orthoreao virus	22.59	28.657	29.122	29.638	30.287
	Influenza virus	26.905	25.434	24.949	25.92	26.089
	RSV	29.879	28.997	26.881	27.831	28.509
	Zika virus	27.828	25.281	24.416	24.981	26.117
Zika viruses present outside the cells (1 MOI), spiked in U937, design 1B		26.833	23.15	26.165	28.137	26.711
Zika Virus present within the cells (10^5 cells),		19.562	18.85	18.358	20.836	19.408
Zika virus infected cells in Buffy Coat matrix		30.511	28.94	31.521	33.405	X

Table 1. CT values from qRT-PCR assays for detecting spiked viruses in U937 cells using various RNA extraction methods. Green CT values indicate the lowest (most optimal) value for the virus tested, blue CT values represent the second-best method, and red CT values denote the third-best method. The virus panel ATCC-MSA 2008 includes four different viruses.

Assessment of different RNA library preparation methods on viral sequence detection



B

Methods tested for NGS library preparation:
Method-1: rRNA depleted- single 6 nucleotide index.
Method-2: no rRNA depletion- single 6 nucleotide index.
Method-3: rRNA depleted- dual 6 nucleotide indexes.
Method-4: isothermal amplification carried out and rRNA depleted- single 8 nucleotide index.
Method-5: isothermal amplification carried out and rRNA depleted- dual 6 nucleotide indexes.

C

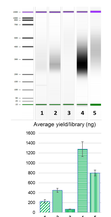


Figure 5. Outline of NGS library preparation methods. A) Virus panels BEI-NR-59622 (contains PCV1, Reo1, FeLV, RSV, EBV), and Zika persistently infected cells were spiked in U937 cells, extracted total RNA and libraries were constructed. The libraries were sequenced in Illumina Novaseq 6000 platform. B) NGS libraries were constructed using different methods, details of the library prep methods are mentioned. C) Top panel- bioanalyzer analysis of NGS libraries prepared, bottom panel- total amount of libraries produced using different methods analyzed using Qubit.

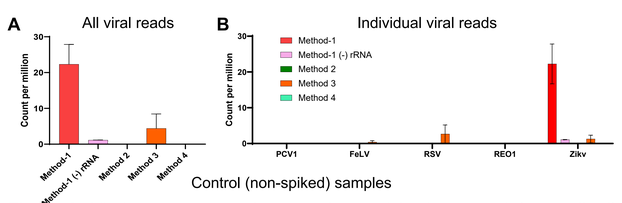


Figure 6. Viral sequence reads were detected in the control (non-spiked) samples due to index hopping. A) Normalized counts per million of viral sequence reads observed for all viruses in different NGS library preparation methods. B) Normalized counts per million of viral sequence reads observed for 5 spiked viruses in NGS library preparation methods.

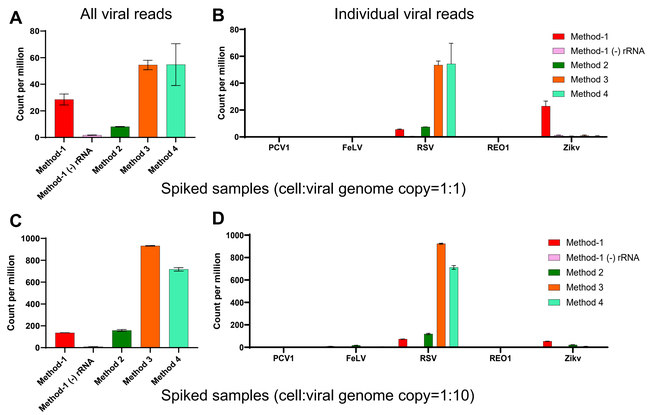


Fig. 7. Identification and quantification of viral sequence reads in spiked samples. A & C) The normalized counts per million of viral sequence reads for all viruses across various NGS library preparation methods. B & D) Show the normalized counts per million for five spiked viruses within these methods. Panels A and B represent sequencing data from samples with a 1:1 ratio of cell to viral genome copy, while Panels C and D correspond to a 1:10 ratio.

Conclusion

1. We detected variability in the efficiency of different RNA extraction methods when isolating viral RNA. The PureLink Mini Kit (PL) from Invitrogen, typically used for RNA extraction from eukaryotic cells, proved to be the most effective for extracting viral RNA from spiked sample matrices.
3. Index hopping was noted in samples or libraries prepared using single indexes of either 6 or 8 nucleotides. This issue was resolved by utilizing dual indexes.
4. FeLV, RSV and ZIKV were detected with all the library preparation methods though we observed higher number of reads for RSV compared to other viruses as reported by other study by Khan et al, 2020, Biologicals.
5. Isothermal amplification method- SPIA showed biased amplification of the RSV, which is not desired for NGS library preparation.

Future direction

1. Create a single tube-based approach for lysis, as well as the extraction and purification of host and viral DNA/RNA. This approach aims to reduce the observed loss of viral nucleic acids typically encountered in standard nucleic acid extraction and purification methods.
2. Determine the optimal RNA extraction and RNA-seq library preparation protocol from our study and establish its detection limit.
3. Evaluate and develop the best bioinformatics pipeline for detection of pathogens and adventitious agents from next generation sequencing data.

Acknowledgements and Funding

We thank CBER FBR core facility for carrying out NGS. This work was supported by the Intramural Research Program of the Center for Biologics Evaluation and Research (CBER), U.S. Food and Drug Administration. This project was also supported in part by Aaron Scholl's appointment to the Research Participation Program at CBER administered by the Oak Ridge Institute for Science and Education through the US Department of Energy and U.S. Food and Drug Administration.