Revolutionizing Vaccine Safety with Next-Generation Sequencing for Adventitious Agent Detection

November 23-24, 2023
DCVMN Regulatory Working Group

Dr. Sameer Naik
Asst. Gen. Manager-Quality Control
Serum Institute of India Pvt. Ltd.

Outline of the talk

- Introduction
- Overview of Current Challenges in NGS
- Regulatory guidance
- Strategy used for adventitious agent detection at SII
- Closing remarks

Introduction

- Next-generation sequencing (NGS) is a dynamic and evolving alternative for adventitious agents/ virus detection due to its sensitivity, speed, reproducibility, and unambiguous nature.
- NGS-based methods are 3Rs (replace, reduce, and refine) compliant and can be used at different stages of manufacturing as non-hypothesis-derived testing of adventitious viruses.
- Presently, all NGS platforms operate on two strategies a) short-read technology (e.g. IonTorrent, Illumina, MGI, Genapsis, etc.) and b) long-read technologies by Oxford nanopore and Pac Bio.
- These technologies generate sequences of nucleic acids from the sample that can be used for the detection of adventitious viruses (known and unknown)
- The most commonly used bioinformatics databases used for sequence analysis are, The National Center for Biotechnology (NCBI) and the Reference Viral database (RVDB) developed by Dr. Khan at CBER.

Overview of challenges in NGS

Challenge	Sample selection, processing, and preparation of 'quality' nucleic acid	• Virus standard • Internal controls • Validation of assays	• Bioinformatics platform • Validation of bioinformatics platform	Follow-up strategies • Confirmation of the 'hit' • Designing decision tree
Solution	• Sample matrix • Check the integrity of nucleic acid, quality, and quantity • Size distribution	• WHO international reference standard for adventitious agent detection by NGS • Define work flow	• Curated viral database, removal of host sequences, selection of appropriate program • Reference virus database (RVDB) • Cloud technology	• Full-length viral genome sequences • Infectivity assays • Risk assessment

Regulatory guidance

European Pharmacopoeia

- Ph. Eur. Chapters 2.6.16 and 5.2.14 advocate the application of NGS methods for adventitious agent testing
- A new chapter of 5.2.14 elaborates substitution of *in vivo* methods by *in vitro* methods such as NGS
- Chapter no. 2.6.41 of Ph. Eur. describes the HTS methods and guidelines for their validation

World Health Organization

- WHO TRS 978 annex-3
- WHO TRS 993, annex-2
- WHO/BS/2020.2394

ICH

- ICH Q5A (R2)

US FDA

- US Food and Drug Administration. 2012. FDA briefing document: vaccines and related biological products advisory committee meeting: cell lines derived from human tumors for vaccine
- AVDTIG: Collaborative efforts between FDA-CBER and Industry, supported by PDA

The strategy used for adventitious agent detection at SII

Adventitious agent testing by NGS at SII

Fig.1 Sampling Scheme

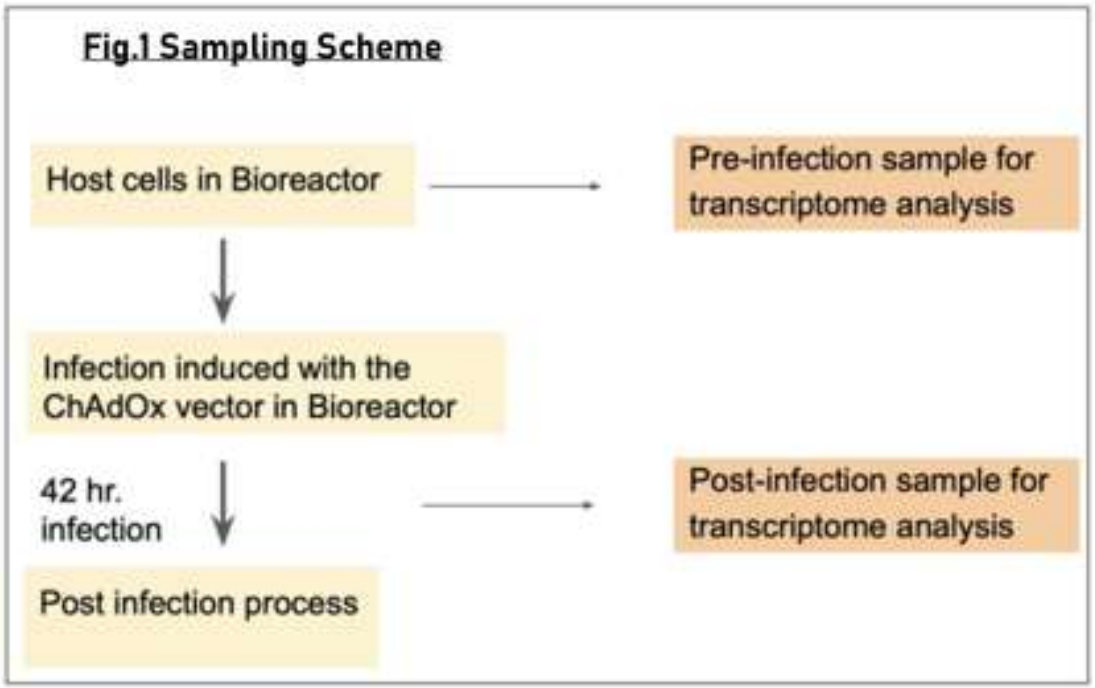
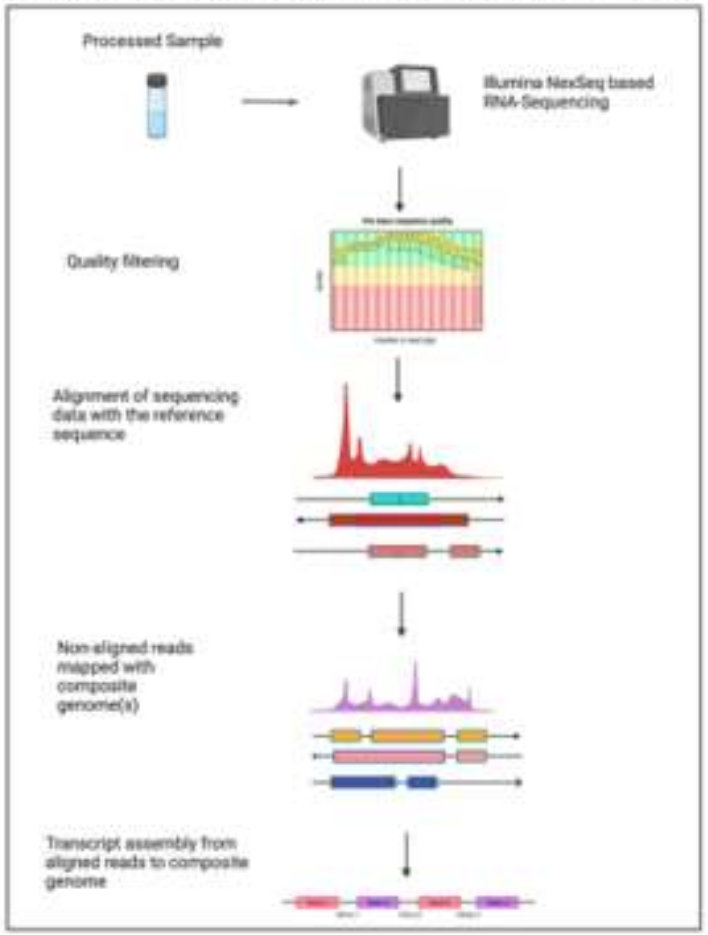


Fig.2 RNA-Seq Analysis Pipeline Schematic



Steps for RNA-Seq analysis

Step # 1

A parallel shotgun sequencing of all RNA species from a single sample was used in the present analysis. The RNA from the source material was extracted, fragmented, and ligated with RNA adapters. These oligomers were subjected to amplification.

Step # 2

This fragmented 'library' of RNA molecules is subjected to Reverse Transcription followed by PCR amplification.

This mixture of RNA fragments is sequenced using massively parallel di-deoxy sequencing using DNA polymerase and specially modified nucleotides.

Step # 3

The 'hg38' build of the human genome was used as a reference for this analysis. (https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.26/), accession ID: GCF_000001405.26 is used as a 'reference' for the transcript assembly.

Step # 4

Sequence information of the fragments/reads itself is used to parse, assemble, and join the reads into ever-larger contiguous sequences.

Comparing the contiguous sequences with all known transcripts from the Hg38 genome can classify the reads into those originating from the host genome and those that could be non-host in origin.

RNA-Seq analysis process

Quality filter and adapter removal

The read quality of Q30 was used for the fast-filter data. Quality-filtered read pair where either read had a length smaller than 74 were removed from the dataset. The adaptors used post-quality filtered reads were checked using GATK V- 4.0.3.

.

Alignment to the composite pathogen genome

A composite pathogen 'genome' was constructed using the sequences. Total 54 viral genomes including 9 CFR viruses, human, simian viruses & and 13 Mycoplasma complete genomes. The reads that did not align to the hg38 genome were aligned to this composite genome using the hisat2 aligner.

Alignment to the hg38 genome

hisat2 version 2.1 was used to align the paired-end reads, The pairs that aligned concordantly were considered actual alignments to the reference. The reads that did not align concordantly or singly were written to fastq files using the —un and the —un-conc options of the hisat2

Transcript assembly of aligned reads using stringtie

The aligned reads to the viral genome were then used to construct transcripts using the stringtie pipeline. stringtie version 2.2 was used to assemble all possible transcripts using the composite viral genome as a reference. No external GTF file was provided as an annotation

Construction and validation of the spike in data set

Sensitivity

- We constructed a combined pseudo-reference genome with the HEK genome and other genomes.
- We detected a 100% alignment of the spike in reads exclusively with the respective genome sequences they were derived from.
- This indicates that the bioinformatics pipeline is sensitive to the detection of nonhost genome sequences even when the spike in data:host transcriptome data is in a ratio of 1:10000 skewed in the favour of the host

Specificity

- To measure specificity a reference genome that included the entire hg38 human genome sequence but omitted varying numbers of the spike was constructed
- We aligned the entire spiked in RNA-Seq dataset containing all the spike in data to this custom-made reference that contained some but not all of the original sequences of the spike in candidates.
- We observed that read identity denoted successful alignment and complied with the pre-determined criteria. The reads derived from sequences present in the reference demonstrated a valid alignment

Reproducibility

- This analysis was repeated with multiple combinations of viral and bacterial sequences omitted from the reference with very similar results.
- We carried out at least 10 of this analysis.
- Our pipeline would detect the presence of nucleotide sequences even when there is an overwhelming majority of reads that arise from the host genome.

Outcome of RNA-Seq analysis

- No adventitious virus transcripts were detected from production control cells and virus harvest.
- The production cultures and virus harvest were negative for the presence of Mycoplasmas and Mycobacteria
- In parallel, the same samples were tested using classical methods and targeted screening at an MHRA-audited contract research laboratory. Results of this testing indicate that samples were negative for adventitious viruses, mycoplasma, and mycobacteria.

Genome integrity using NGS

- We performed High Throughput Next Generation Sequences (HT-NGS) shotgun sequencing of purified viral DNA. This NGS data was then used to assemble the viral genome *in silico* and identify the antigenic sequence present. The assembled DNA sequences will then be assembled into complete scaffolds and then a single contiguous genome.
- The complete genome sequence will also be analyzed against the parent sequence provided to ensure that the pre-GMP seed, Master seed virus, working seed virus and the virus samples at the end of production are identical to that of the parent sequence.
- A high-coverage sequencing experiment was carried out to ensure the capture of all variants that could be extant within the viral sample. Combined with extremely stringent alignment parameters during genome assembly and alignment, this approach enables the detection of all variants within the sample.
- This dataset of extremely high coverage can also be aligned with the genomes of any potential pathogenic strains of viruses to detect the presence of those contaminants in the sample

Genome integrity using NGS

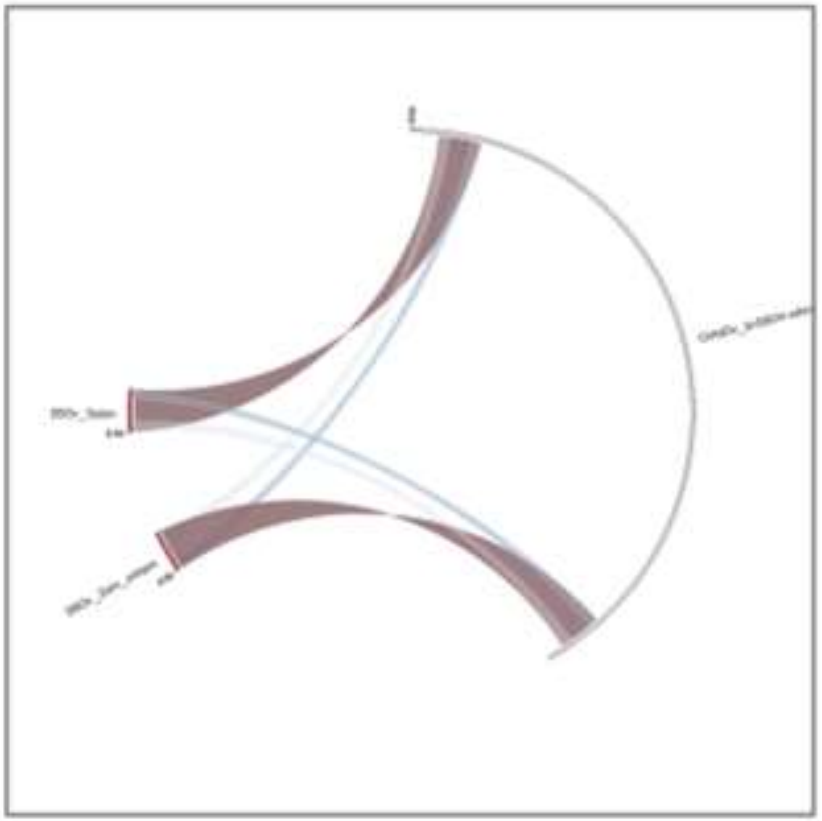


Fig. 1. Identification of insert and its genetic stability

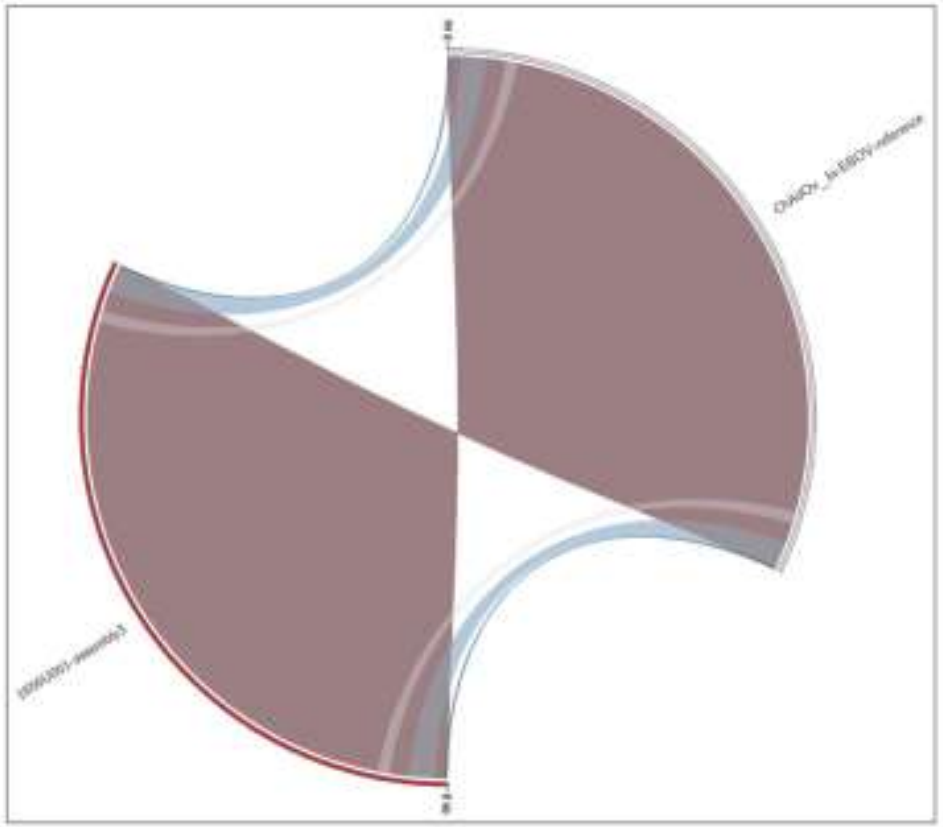


Fig. 2. Genome integrity and stability

Figures are derived with mummer2circos

The Outcome of the study

- Each sample was sequenced at a minimal depth of ~ 10000X
- Each sample was 100% identical to the reference ChAdOx-1-bi EBOV viral vector reference genome.
- There are no rearrangements in any genomes analyzed as compared to the ChAdOx-1-bi EBOV reference genome.
- The antigen sequences for the Sudan and Zaire antigens are identical to the reference sequence at the nucleotide level.
- The gene sequences for the genes E1 and E3 are absent in all the samples analyzed.

Concluding remarks

1

All SII vaccine products, at different stages of manufacturing, are tested for the presence of adventitious viruses using classical and targeted nucleic acid amplification methods

2

The use of NGS for adventitious agent testing was started during COViD pandemic. All COViD vaccines were tested for the presence of adventitious agents using the Illumina platform of NGS.

3

Recently SII has established a state-of-art molecular biology laboratory and has purchased the Oxford nanopore for the implementation of NGS for adventitious agent testing

4

The SII is an active participant in the development, collaboration, and implementation of the 3Rs. The SII has successfully implemented the in vitro assays for different vaccines as a replacement for animal assays.

5

There are at least two assays of the 3Rs that are currently under development.

6

We, at SII, are willing to participate in any of the collaborative studies for the use of NGS in adventitious agent testing.


thank you
danke
謝謝
ngiyabonga
teşekkür ederim
gracias
merci
obrigado
bedankt
dziękuję
спасибо
Баярлалаа
рахмат
dank je
mochchakkeram
go raibh maith agat
sukriya
kop khun krap
terima kasih
grazie
arigatō
takk
dakujem
мерси
хвала
tapadh leat
ευχαριστώ
감사합니다
ありがとう
hvala
sagolun
taiku
blagodaram
mersi
tack
spas
maruuru
dankie
kiitos
enkosi
rahmat
tanemirt
dhanyavadagalu
shukriya
asante
mamnun
djiere dieuf
matondo
misaotra
paldies
grazzi
gracies
sulpay
dekuji
sobodi
mesi
didi madobe
najis tuke
kam sah hamnida
xiexie
diolch
trugarez
manana
faafetai lava
vinaka
welalin
barka
nuhun