4th Conference on NGS for Adventitious Virus Detection in Biologics for Humans and Animals

## Advanced Virus Detection Technologies Working Group

## Spiking Study 4

# Evaluation of Long-Read Sequencing for Adventitious Virus Detection in a Low Complexity Viral Background

Co-leaders:
Valeria Maria Zanda (Merck), Simone Olgiati (Merck)

*December 4, 2024*

# DISCLAIMERS

- Dr. Valeria Zanda is employed by RBM S.p.a., Italy, an affiliate of Merck.

- The content and views expressed in this presentation are those of the individual experts participating to the AVDTWG Spiking study #4 and should not be construed to represent the views or policy of the organization, regulatory authority, or governmental agency they represent.

- All criteria, tools or ways of working described must be considered as “considerations” only and should not be taken as recommendations, mandatory use or regulatory views.

# Advanced Virus Detection Technologies WG

## Mission

To facilitate the use of advanced technologies (such as NGS) for the detection of adventitious viruses in biologics by providing an informal, scientific forum for knowledge exchange, scientific discussions and collaborations among scientists across different organizations

### 2024

**Subgroup AB**

Sample selection/ preparation/processing
Virus standards and reference materials

**Subgroup C**

Complete and correctly annotated,
virus reference database

**Subgroup DE**

Bioinformatics pipelines analysis
Follow-up strategies to confirm the identity of a "hit"

### Short-read NGS

| | |
|---|---|
| **Spiking study 2A**<br>Started 2017 | MVM in *CHO cell background* |
| **Spiking study 2B**<br>Started 2016 | WHO 5 viruses in high-titer virus background: *viral seed/viral vector* |
| **Spiking study 3**<br>Started 2019 | Transcriptomics: infected cells in uninfected cells background |

### Long-read NGS

| | |
|---|---|
| **Spiking study 4**<br>Started 2020 | Viral seed/viral vector background |
| **Spiking study 5**<br>Started 2023 | Unprocessed bulk |
| **Spiking study 6**<br>Started 2023 | Transcriptomics/Cell bank |

# AVDTWG Collaborative Spiking Study 4

## Introduction

**Purpose**
To evaluate the application of long-read sequencing for adventitious virus detection

**Models**

- Viral stocks spiked in a high-titer viral background (Adenovirus5)
- Viruses*:
  - Epstein-Barr virus (EBV)
  - Mammalian orthoreovirus type 1 (REO)
  - Human respiratory syncytial virus (RSV)
  - Feline leukemia virus (FeLV)
  - Porcine Circovirus 1 (PCV1)

**Members:**
12 different organizations

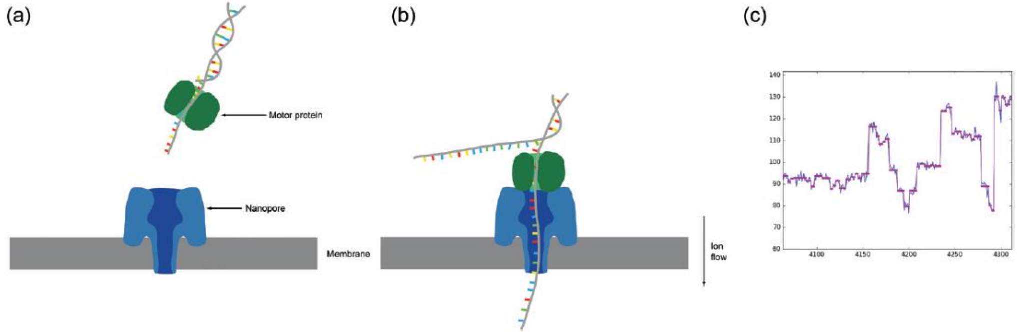
**Timing**
Expected completion:
end-2025

**CBER NGS Virus Reagents :* https://www.beiresources.org/Catalog/animalviruses/NR-59622.aspx

# Long-Read Sequencing Technology

## Rationale for the Study

### Nanopore sequencing technology overview



Leggett RM, et al. J Exp Bot. 2017. PMID: 28992056

Evaluate the new technology on sensitivity and specificity for viral detection:

- Direct DNA/RNA sequencing
- Full length genome and circular genome
- Native or amplified DNA/RNA sequencing
- Real time analysis
- Cost effective
- Error rate

## Aim of the study

Evaluate the long-read sequencing performance on viral contaminants detection in high-titer viral background, mimicking the Illumina based study 2B for comparing with short-read sequencing

# AVDTWG Collaborative Spiking Study 4

## Experimental Design & Timeline

### Samples to be analyzed

High-titer viral background: Adenovirus 5 ($10^9$ GC in 200μL)

Two shared spike levels: $10^4$ and $10^6$ GC

### Experimental procedure

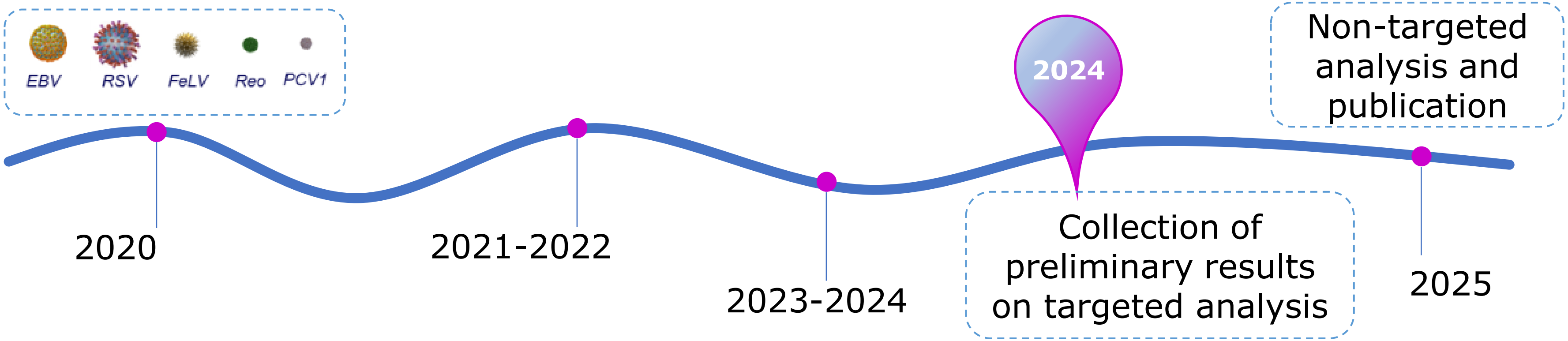
- Common protocol
- Individual protocol
- Both common protocol and individual solution

### Data analysis

- Individual pipeline for targeted and non-targeted analysis

### Data collection & publication

- Common datasheet creation
- Data collection
- Final publication



2020

2021-2022

2023-2024

2024

Collection of preliminary results on targeted analysis

2025

Non-targeted analysis and publication

# AVDTWG Collaborative Spiking Study 4

## Common Protocol

- To explore the new technology and gain experience

**Sample Preparation** → **Extraction** → **Reverse Transcription** → **Library prep & Sequencing** → **Data Analysis**

**Sample Preparation**

Viral stock dilutions

Preparation of spiked samples

**Same for all labs**

**Extraction**

Recover both RNA and DNA viral nucleic acids

**Reverse Transcription**

Obtain double stranded cDNA

**Library prep & Sequencing**

Different instruments

Different chemistry versions

**Data Analysis**

Individual pipelines

- Points considered:

Low quantity of nucleic acids

Preserve long viral fragments

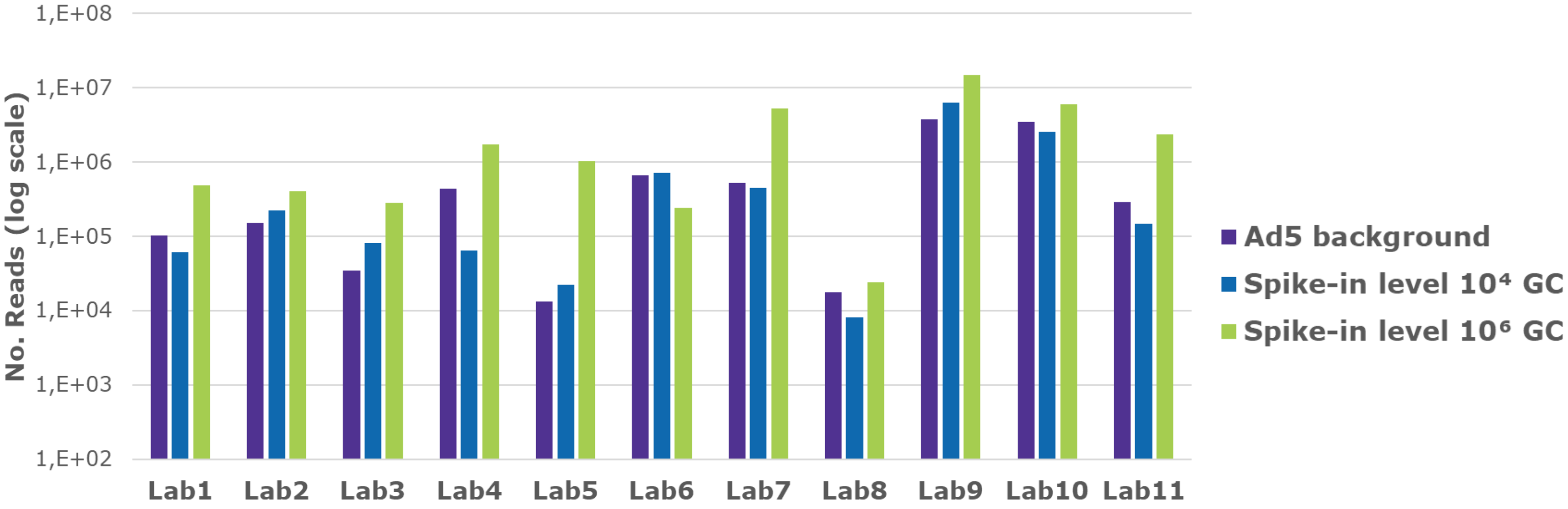
Optimize for higher viral recovery

Chemistry /software update

Pipeline adaptation

# Sequencing results

## Total number of reads

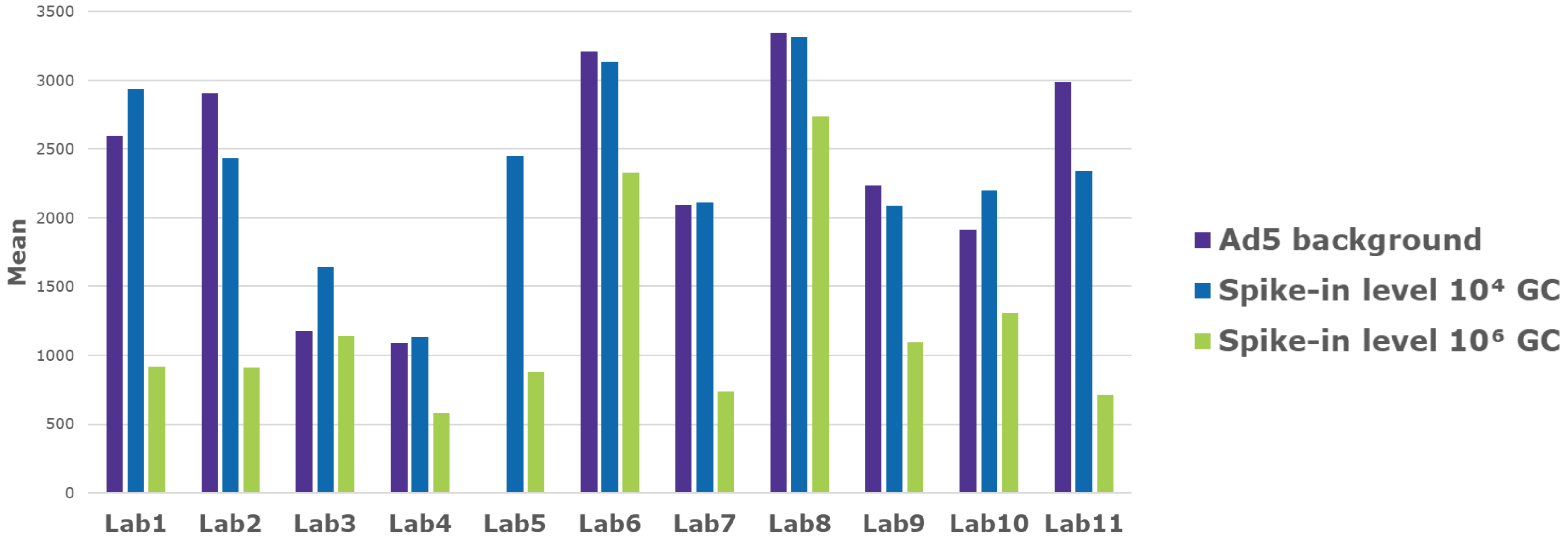


- Range of reads between 8K and 14M
- Substantial variability in the total number of reads was observed between different labs in all the different spike-in levels
- Highest number of reads observed for sample $10^6$ spike-in level for majority of the labs

Waiting for Lab12 data

# Sequencing results

## Mean read length

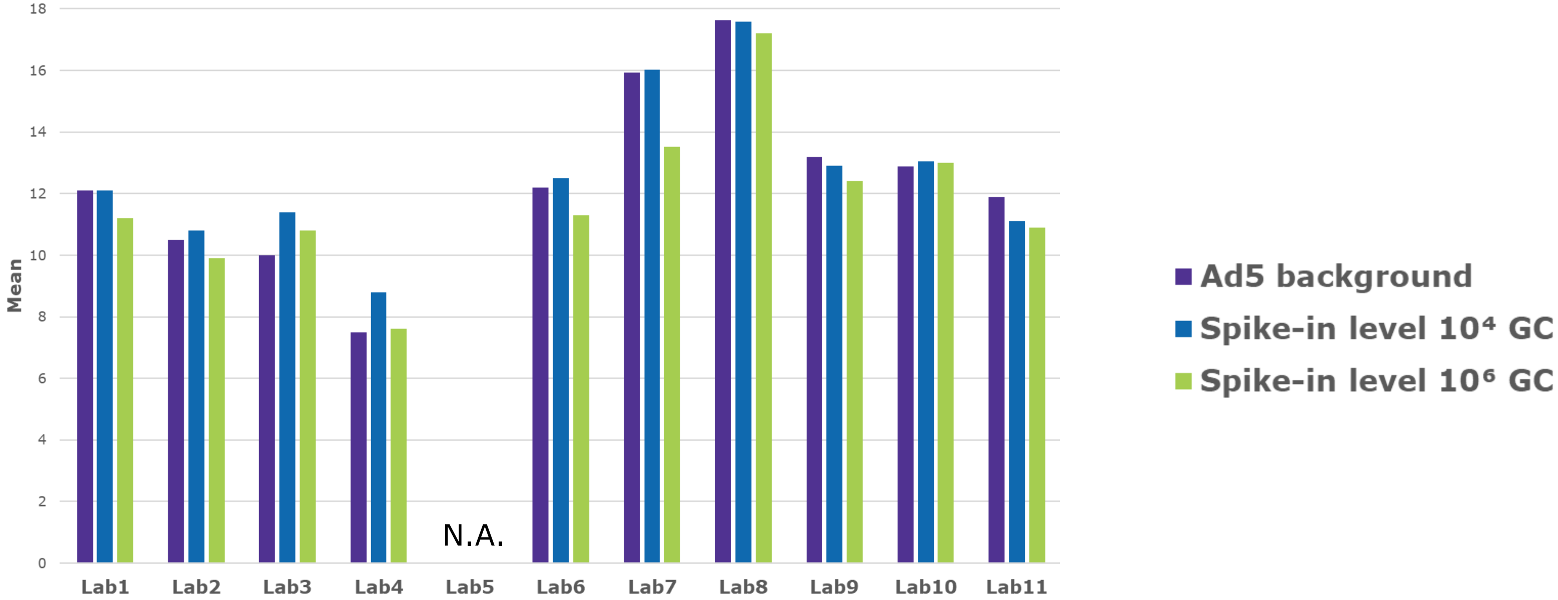


- The mean read length observed between 581 bp and 3,345 bp
- Substantial variability of the mean read length observed between different labs in all the different spike-in levels
- All the labs reached the lowest mean read length for sample $10^6$ spike-in level

Waiting for Lab5 and Lab12 data

# Sequencing results

## Mean read quality

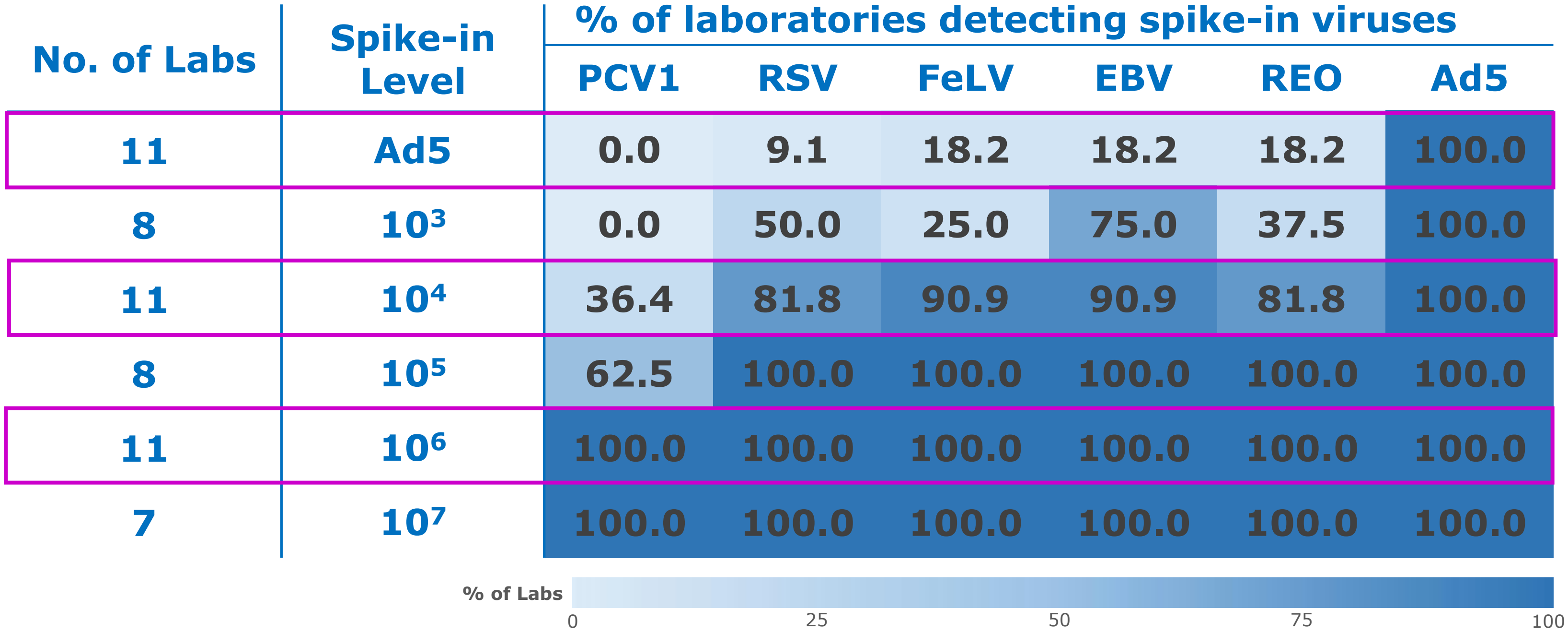


- Mean read quality between Q7 and Q17
- Each lab obtained consistent q score among the different runs
- Variability observed between the quality score obtained by the different labs

Waiting for Lab12 data

# Detection of Spike-in Viruses

## Targeted Analysis Results



| No. of Labs | Spike-in Level | % of laboratories detecting spike-in viruses | | | | | |
|---|---|---|---|---|---|---|---|
| | | PCV1 | RSV | FeLV | EBV | REO | Ad5 |
| 11 | Ad5 | 0.0 | 9.1 | 18.2 | 18.2 | 18.2 | 100.0 |
| 8 | $10^3$ | 0.0 | 50.0 | 25.0 | 75.0 | 37.5 | 100.0 |
| 11 | $10^4$ | 36.4 | 81.8 | 90.9 | 90.9 | 81.8 | 100.0 |
| 8 | $10^5$ | 62.5 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| 11 | $10^6$ | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| 7 | $10^7$ | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |

All viruses detected at $10^6$ GC by all labs and at $10^5$ GC by the majority of labs

Waiting for Lab12 data

# Detection of Spike-in Viruses

## Targeted Analysis Results

### Number of Mapping Reads Detected

| Spike-in Level | PCV1 Min-Max | RSV Min-Max | FeLV Min-Max | EBV Min-Max | REO Min-Max |
|---|---|---|---|---|---|
| **Ad5** | 0 | 0-1 | 0-1 | 0-12 | 0-5 |
| **$10^3$** | 0 | 0-3 | 0-2 | 0-18 | 0-4 |
| **$10^4$** | 0-26 | 0-357 | 0-277 | 0-3,459 | 0-2,217 |
| **$10^5$** | 0-72 | 1-1,910 | 9-808 | 5-1,888 | 3-876 |
| **$10^6$** | 1-1,102 | 12-21,116 | 37-19,978 | 24-100,429 | 1-19,563 |
| **$10^7$** | 2-539 | 69-21,222 | 669-12,591 | 108-30,847 | 108-182,008 |

# Conclusions

- All viruses were detected in the targeted analysis at $10^6$ viral GC
- Long-read technology is able to detect circular viruses
- All but PCV1 virus observed at $10^4$ viral GC in the majority of the labs
- High spiked levels displayed higher number of total reads with lower mean length

# Next Step

- Aggregation of non-targeted analysis results
- Performance detection among individual and common protocol will be evaluated once data package will be completed

# Lessons Learned

- Technology is **rapidly evolving** with frequent updates of chemistry and software – to be fixed for results comparison
- Workflow:
  - Evaluate impact of **starting material** – quantity/quality
  - Adapt the protocol to **preserve long nucleic acids**
- Sequencing:
  - Batch-to-batch **flow cell variability**
  - **Quality criteria** for base calling: impact on sensitivity to be further investigated
- Analysis:
  - Presence of **full-length reads** longer than viral genome size, possible chemistry artefacts
  - **Bioinformatic pipeline** should be adapted to work on long reads

# Acknowledgments

## All Spiking Study 4 Members

- FDA
  - CBER OVRR
  - CBER OBRR
- Genentech
- GSK
- Kobe University
- Merck
- Massachusetts Institute of Technology (MIT)
- NIIMBL/University of Delaware
- PathoQuest
- Regeneron
- Sanofi
- ViruSure