# **Developing Unbiased RNA Sequencing for Biothreat and Emerging Disease Detection**

## Signature Science, LLC: Kathleen Schulte, Amanda N. Scholz, Leah W. Allen, Matthew C. Murray, Krista L. Ternus, F. Curtis Hewitt, and Anthony D. Kappell\*

## INTRODUCTION

309

June 18, 2023

#### Background

- RNA viruses pose a significant global biothreat.
- Beyond the recent coronavirus pandemic, filoviruses, alphaviruses, and flaviviruses, among others, cause recurring outbreaks in the Americas and around the world.
- Untargeted, metagenomic sequence-based approaches that are unbiased are a potential solution in detecting novel viruses.
- Untargeted sequence-based approaches could lead to early detection of both previously characterized and novel threats RNA virus threats.<sup>1</sup>
- Untargeted sequence-based approaches has the potential to serve as a hypothesis-free, single, and universal assay for diagnostics of known and novel infectious disease and Emerging Infectious Diseases (EIDs) directly from samples.<sup>2</sup>
- Single agnostic test reduces the need for serial testing against a list of suspected pathogens or culturing which can lead to delayed actionable results.
- Nanopore sequencing has the promise of allowing near realtime sequencing and detection.
- Nanopore sequencing for public health threats is well established and has been previously demonstrated.<sup>3–6</sup>
- Targeted nanopore sequencing for viral detection in public health labs has been successful as part of the COVID-19 pandemic response.<sup>7</sup>

#### Study

- We have previously established a 6-plex (8 total) library preparation workflow for the untargeted, unbiased sequencing of RNA viruses from Viral Transport Media (VTM) and Human Plasma (ASM Microbe 2022)
- This study expands the evaluation of the final library preparation workflow. We also addressed the outstanding question of barcode binning with barcodes on both ends (both barcodes) is better than single end (single barcodes)



## **MATERIALS AND METHODS**

**Contrived Samples:** Generated using pooled remnant VTM or plasma not associated with an infection and spiked with a pool of matri related viruses. These were then serial dilute into relevant matrix.

Remnant Samples (Real-world): Were obtained for related viruses used in the contrived samples

**Controls:** MS2 Phage was added to each sample at extraction as internal standard including in a phosphate-buffered saline bla sample. Green-fluorescent protein mRNA was used as a positive control for processes upstream of the extraction without MS2 Pha

## Library Preparation Workflow



Ebola Virus

## **Contrived Samples within Clinical Matrices**

| LT1   |               |          |  |   |                                |                        |                                     | Human Plasma                   |                               |
|-------|---------------|----------|--|---|--------------------------------|------------------------|-------------------------------------|--------------------------------|-------------------------------|
| rix   |               | Media    | Viral Transport Media                                    |   |                                | Human Plasma           |                                     |                                |                               |
| ted   |               | Virus    | Influenza A<br>Virus (A/<br>Wisconsin/15/<br>2009 (H3N2) | Human<br>Respiratory<br>Syncytial Virus<br>(A 1998/3-2) | Human<br>Coronaviurs<br>(229E) | Zika Virus<br>(MR 766) | Hepatitis A<br>Virus<br>(HM175/18f) | Yellow<br>Fever Virus<br>(17D) | Chikungunya<br>Virus (181/25) |
|       |               | BEI Cat# | NR-42007   | NR-28529  | NR-52726                       | NR-50065               | NR-137                              | NR-116                         | NR-56523                      |
|       |               | ID       | InfA   | RSV   | 229E                           | Zika                   | HepA                                | YF                             | Chik                          |
|       |               | А        | 7.10   | 5.60  | 3.85                           | 7.35                   | 6.85                                | 6.60                           | 6.35                          |
|       |               | В        | 6.10   | 4.60  | 2.85                           | 6.35                   | 5.85                                | 5.60                           | 5.35                          |
|       | Load Levels   | С        | 5.10   | 3.60  | 1.85                           | 5.35                   | 4.85                                | 4.60                           | 4.35                          |
| olank | LOG10 (TCID50 | D        | 4.10   | 2.60  | 0.85                           | 4.35                   | 3.85                                | 3.60                           | 3.35                          |
|       | per mL)       | E        | 3.10   | 1.60  | -0.15                          | 3.35                   | 2.85                                | 2.60                           | 2.35                          |
| S     |               | F        | 2.10   | 0.60  | -1.15                          | 2.35                   | 1.85                                | 1.60                           | 1.35                          |
| nage. |               | G        | NA   | NA  | NA                             | 1.35                   | 0.85                                | 0.60                           | 0.35                          |
|       |               |          |  |   |                                |                        |                                     |                                |                               |

## RESULTS

#### **Binning of Sequences Based on Barcodes on** at Least One end (Single) or Both Ends (Both)

| Mapping to GFP from Different Barcodes |                   |              |                                       |  |  |
|--|-------------------|--------------|---------------------------------------|--|--|
| Sample                                 | Single<br>Barcode | Both Barcode | Reduction of<br>Barcode<br>Cross-talk |  |  |
| GFP                                    | 173,105           | 120,911      | NA                                    |  |  |
| NEG                                    | 1,054             | 39           | 96%                                   |  |  |
| Sample 1                               | 1,130             | 72           | 94%                                   |  |  |
| Sample 2                               | 815               | 37           | 95%                                   |  |  |
| Sample 3                               | 940               | 80           | 91%                                   |  |  |
| Sample 4                               | 1,359             | 136          | 90%                                   |  |  |
| Sample 5                               | 1,534             | 98           | 94%                                   |  |  |
| Sample 6                               | 764               | 70           | 91%                                   |  |  |

#### **Abbreviated Table of Clinical Remnant Samples**

#### Example of Little to No Loss of Detection when usin on Both Ends Setting Compared to Single End Setti

| VTM                               | RSV               |                  | InfA               |                  | 229E               |                  |
|-----------------------------------|-------------------|------------------|--------------------|------------------|--------------------|------------------|
| Contrived<br>Sample Viral<br>Load | Single<br>Barcode | Both<br>Barcodes | Single<br>Barcodes | Both<br>Barcodes | Single<br>Barcodes | Both<br>Barcodes |
| Α                                 | 4/4               | 4/4              | 4/4                | 4/4              | 4/4                | 4/4              |
| В                                 | 8/8               | 8/8              | 8/8                | 7/8              | 8/8                | 8/8              |
| С                                 | 10/10             | 10/10            | 0/10               | 0/10             | 10/10              | 10/10            |
| D                                 | 4/8               | 4/8              | 0/8                | 0/8              | 8/8                | 8/8              |
| E                                 | 0/4               | 0/4              | 0/4                | 0/4              | 0/4                | 0/4              |
| F                                 | 0/2               | 0/2              | 0/2                | 0/2              | 0/2                | 0/2              |

= All Replicates Detected = At Least One Detected = None Detected

#### Limit of Detection from Contrived Samples within Clinical Matrices

| 95% Probability | Both Ba         | arcodes          | Single B        | Probit           |            |
|-----------------|-----------------|------------------|-----------------|------------------|------------|
| Organism        | Logit Analysis* | Probit Analysis* | Logit Analysis* | Probit Analysis* | Difference |
| RSV             | 2.71            | 2.79             | 2.71            | 2.79             |            |
| InfA            | 5.65            | 5.72             | 6.14            | 6.17             | 0.45       |
| 229E            | 0.41            | 0.47             | 0.41            | 0.47             |            |
| Zika            | 3.42            | 3.41             | 3.42            | 3.41             |            |
| Chik            | 2.42            | 2.42             | 2.92            | 2.88             | 0.46       |
| YF              | 3.03            | 3.00             | 3.17            | 3.14             | 0.14       |
| HepA            | 4.52            | 4.50             | 4.52            | 4.50             |            |

\* LOG10 (TCID50 per mL)

| Sample           | Gold Standard       | Organism of Assembly   | SeqScreen Call       | Modified Call | Stat   |  |
|------------------|---------------------|--|----------------------|---------------|--------|--|
| Human            | Genmark EPlex PCR   | Streptococcus anginosus  | Genus Present        | Absent        | FN     |  |
| metapneumovirus  |                     | Magnetospirillum gryphiswaldense MSR-1 v2                        | Genus Present        | Absent        | 1      |  |
|                  |                     | Streptococcus equi subsp. zooepidemicus Genus Present            |                      | Absent        |        |  |
| Parainfluenza IV | Genmark EPlex PCR   | Human parainfluenza virus 4a                                     | Genus Present        | Present       | TP     |  |
| Parainfluenza IV | Genmark EPlex PCR   | Human parainfluenza virus 4a                                     | Absent               | Present       | TP     |  |
| SARS-CoV-2       | Gene Xpert Infinity | SARS coronavirus Tor2  | Absent               | Present       | TP     |  |
| SARS-CoV-2       | Gene Xpert Infinity | SARS coronavirus Tor2  | Genus Present        | Present       | ТР     |  |
|                  |                     | Carnobacterium maltaromaticum                                    | Genus Present Absent |               | 1      |  |
|                  |                     | Corynebacterium rouxii   | Genus Present        | Absent        | 1      |  |
| SARS-CoV-2       | Gene Xpert Infinity | SARS coronavirus Tor2  | Genus Present        | Present       | ТР     |  |
| RSV              | Biofire             | NA   | Absent               | Absent        | FN     |  |
| RSV              | Biofire             | Human orthopneumovirus   | Genus Present        | Present       | TP     |  |
|                  |                     | Streptococcus pneumoniae   | Present              | Absent        |        |  |
| RSV              | Biofire             | Respiratory syncytial virus                                      | Present              | Present       | ТР     |  |
|                  |                     | Human orthopneumovirus   | Absent               | Present       |        |  |
| Influenza A      | Biofire             | Influenza A virus (A/California/07/2009(H1N1))                   | Genus Present        | Present       | ТР     |  |
|                  |                     | Cyclobacteriaceae bacterium                                      | Genus Present        | Absent        | 1      |  |
| Influenza A      | Diasorin Integrated | Influenza A virus (A/California/07/2009(H1N1))                   | Absent               | Present       | ТР     |  |
|                  | Cycler              | Streptococcus anginosus  | Present              | Absent        | 1      |  |
| Influenza A      | Biofire             | NA   | Absent               | Absent        | FN     |  |
| Enterovirus      | Diasorin Integrated | Rhinovirus A   | Absent               | Present       | ТР     |  |
| Enterovirus      | Diasorin Integrated | Rhinovirus A   | Genus Present        | Present       | ТР     |  |
| Enterovirus      | Diasorin Integrated | Rhinovirus A   | Absent               | Present       | ТР     |  |
|                  | Cycler              | Lactobacillus crispatus  | Genus Present        | Absent        |        |  |
|                  |                     | Cutibacterium acnes HL096PA1                                     | Present              | Absent        |        |  |
|                  |                     | Corynebacterium efficiens YS-314                                 | Present              | Absent        |        |  |
| Hepatis A        | DiaSorin Liaison XL | NA   | Absent               | Absent        | FN     |  |
| Hepatis A        | DiaSorin Liaison XL | Salmonella enterica subsp. enterica serovar Typhimurium str. LT2 | Absent               | Present       | FN/ FP |  |
| Hepatis A        | DiaSorin Liaison XL | Bacillus cereus  | Genus Present        | Absent        | FN     |  |
|                  |                     | Acinetobacter baumannii  | Genus Present        | Absent        |        |  |

Modified Call is based on accepting ≥0.15 Breadth.Coverage and ≥1 Depth.Coverage as calculated by SeqScreen-Nano • TP: True Positive, FP: False Positive, FN: False Negative

# 512.533.2092 akappell@ signaturescience.com

| ing  | Barcodes |
|------|----------|
| ting | J        |

## CONCLUSIONS

- Best method identified was REPLI-g assisted Double-stranded cDNA synthesis as input into the Native Barcoding Kit and Ligation Sequencing Kit to generate an 8-plex with 6 samples and 2 controls.
- Sequencing library preparation takes between 10 to 16 hours, based on experience of analyst.
- Current limit of detection is greater than targeted PCR assays and near that of immuno-based assays.
- Detection limits appear to be associated with genome size and segmentation of the virus.
- Sensitivity is 0.5714 based on remnant samples with currently no detection of Human metapneumovirus and Hepatis A.
- Precision is 0.8571 including the common detection contaminants of Klebsiella pneumoniae and Escherichia coli O157:H7.
- F1 Score is 0.6857 based on remnant samples.
- We are continuing to sequence remnant samples to determine the limitations and associated statistics with this assay.

## REFERENCES

- 1. N. Sapoval, et al. (2021) Hidden genomic diversity of SARS-CoV-2: implications for qRT-PCR diagnostics and transmission, Genome Res.
- 2. Y. Xu, et al. (2018) Detection of Viral Pathogens With Multiplex Nanopore MinION Sequencing: Be Careful With Cross-Talk, Front. M2017icrobiol. 9. (And Citations Within)
- F.C. Hewitt, et al. () Toward Rapid Sequenced-Based Detection and Characterization of Causative Agents of Bacteremia, BioRxiv. 162735.
- 4. H.P. McLaughlin, et al. (2020) Rapid Nanopore Whole-Genome Sequencing for Anthrax Emergency Preparedness, Emerg. Infect. Dis. 26 (2020) 358–361.
- 5. J.A. Russell, et al. (2018) Unbiased Strain-Typing of Arbovirus Directly from Mosquitoes Using Nanopore Sequencing: A Field-forward Biosurveillance Protocol, Sci. Rep. 8. 5417.
- 6. M.R. Lindberg, et al. (2016) A Comparison and Integration of MiSeq and MinION Platforms for Sequencing Single Source and Mixed Mitochondrial Genomes, PLOS ONE. 11. e0167600.
- CDCgov/SARS-CoV-2\_Sequencing, Centers for Disease Control and Prevention, 2021. https://github.com/CDCgov/SARS-CoV-2\_Sequencing
- 8. S.T. Calus (2018). NanoAmpli-Seq: a workflow for amplicon sequencing for mixed microbial communities on the nanopore sequencing platform. Gigascience, 7(12), giy140. [Workflow contains adapted figure]
- 9. A. Balaji et al. (2022). SeqScreen: accurate and sensitive functional screening of pathogenic sequences via ensemble learning. Genome Biology, 23(1), 133.
- 10. A. Balaji et al. (2023). SeqScreen-Nano: a computational platform for rapid, in-field characterization of previously unseen pathogens. bioRxiv, 2023-02.

## ACKNOWLEDGEMENTS

This research was funded by Centers for Disease Control and Prevention under award numbers 75D30121C12250 and 75D30122C15359. Additionally, the development of SeqScreen and SeqScreen-Nano was funded through IARPA Award No. W911NF-17-2-0089. The view and conclusions contained herein are those of the authors and should not be interpreted as official policy or endorsement of the HHS, CDC, ODNI, IARPA, ARO or the US Government.