CPHM 308
June 16, 2024

Evaluation of Agnostic RNA Sequencing for Biothreat and Emerging Infectious Disease Detection

512.533.2092
akappell@
signaturescience.com

Signature Science, LLC: Kathleen Q. Schulte, Elizabeth A. Scheuermann, Nicolette C. Keplinger, Matthew B. Scholz, Leah W. Allen, Amanda N. Scholes, Taylor A. Wolt, Viviana M. June, Cole J. Schulte, Krista L. Ternus, F. Curtis Hewitt, and Anthony D. Kappell*

INTRODUCTION

Background

- RNA viruses pose a significant global biothreat.
  - Beyond the 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.[1]
  - 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.[2]
  - 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 real-time sequencing and detection.
  - Nanopore sequencing for public health threats is well established and has been previously demonstrated.[3–6]
  - Targeted nanopore sequencing for viral detection in public health labs has been successful as part of the COVID-19 pandemic response.[7]

Study

- We have established a 6 clinical sample multiplex (8 total with controls) library preparation workflow for the untargeted, unbiased sequencing of RNA viruses from Viral Transport Media (VTM) and Human Plasma
- This study evaluated the final library preparation workflow using contrived clinicals samples and patient remnant samples.

Ebola Virus

MATERIALS AND METHODS

Contrived Samples: Virus free pooled remnant VTM or plasma spiked with relevant viruses

Remnant Samples: Remnant VTM samples for respiratory viruses

Internal Control: MS2 Phage added to each sample at extraction

Process Control: Green-fluorescent protein mRNA

Library Preparation Workflow

1 Viral RNA Extraction
1 – 1.5 hr.

2 Double-stranded cDNA Synthesis
3 – 3.5 hr. • Includes AMPure Bead Cleanup

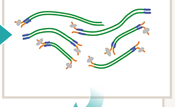
3 Whole Genome Amplification
4 – 4.5 hr. • Includes AMPure Bead Cleanup and Qubit Quantification

4 T7 Endonuclease Digestion
1.2 –1.5 hr. • Includes Custom Bead Cleanup and Qubit Quantification

5 DNA Repair
1.2 – 1.5 hr. • Includes AMPure Bead Cleanup

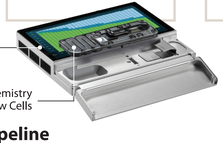
6 Barcode Ligation
1 – 1.2 hr. • Includes AMPure Bead Cleanup and Pooling of Library

7 Adapter Ligation
1.5 – 2 hr. • Includes AMPure Bead Cleanup and Loading of Library



MinION Mk1c Onboard Fast Basecalling

R9.4.1 Chemistry Type Flow Cells



Sequenced Using a MinION (Mk1B or Mk1C)
24 hour run time

SigSciDx Bioinformatic Pipeline

Use EPI2ME Labs Framework and Jupyter Notebooks for Execution within a Docker Container → **Run Quality Control** (i.e. Read Quality Read Length Distribution No. of Reads per Barcode) → **Host (Human) Sequence Removal** → **GFP and MS2 Phage Alignment Statistics (Evaluate Controls)**

SeqScreen's Taxonomical Inference Module[8,9]

barcode04 | 5 entries per page

Barcode	Species	Taxonomy ID	Source Database	Assembly Level	No of Reads	Coverage	Depth	Coverage-2nd	Depth-2nd	ANI	Combined	Presence/Absence
barcode04	Human coronavirus 229E	11137	NaN	Complete Genome	1315	0.987297	37.635671	0.987151	37.615021	0.929547	95.17	Present
barcode04	Respiratory syncytial virus	12814	NaN	Complete Genome	516	0.966493	32.283340	0.966493	32.283340	0.976022	95.48	Present
barcode04	Escherichia phage MS2	12022	NaN	Complete Genome	15437	0.959933	1865.009048	0.959933	1864.934910	0.997351	95.87	Present

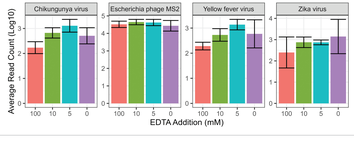
RESULTS

Performance Characteristics for the Agnostic RNA Sequencing Assay

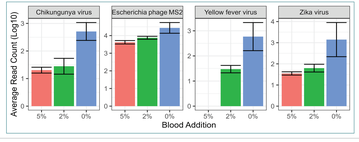
Performance Metric	Method	Result
Limit of detection (LoD)	▪ Qualitative detection of RNA virus dilution replicates by probit analysis	**Pathogen → LoD (Estimated PFU[1])** Influenza A Virus → 26,915 $TCID_{50}$ /mL (18,840 PFU/mL) Human Respiratory Syncytial Virus → 214 $TCID_{50}$ /mL (150 PFU/mL) Human Coronavirus 229E → 1 $TCID_{50}$ /mL (1 PFU/ mL) Zika Virus → 550 $TCID_{50}$ /mL (385 PFU/ mL) Hepatitis A Virus → 1318 $TCID_{50}$ /mL (923 PFU/ mL) Yellow Fever Virus → 68 $TCID_{50}$ /mL (47 PFU/ mL) Chikungunya Virus → 115 $TCID_{50}$ /mL (80 PFU/ mL)
Precision	▪ Qualitative detection over 2 to 7 contrived sample runs of each organism (inter-assay)[2] ▪ Qualitative detection of duplicate contrived samples on the same run (intra-assay)[2] ▪ Qualitative detection over 21 remnant sample (duplicate or triplicate) runs (inter-assay)	▪ 100% concordance ▪ 100% concordance ▪ 100% concordance
Interference	▪ Quantitative read count of viruses with spiked blood (5%, 2%) ▪ Quantitative read count of viruses with spiked EDTA (100 mM, 10 mM, 5 mM) ▪ Quantitative read count of viruses with spiked bacteria (Micrococcus luteus or *Staphylococcus epidermidis*)	▪ Decrease of number of reads by 1.3-Logs to 1.5-Logs mapped to viruses ▪ Decrease of number of reads by 0.6-Logs and 0.1-Logs with addition of 100 mM and 10 mM EDTA, respectively ▪ No significant changes in number of reads mapped to RNA viruses
Accuracy	▪ 18 of single-detection remnant sample runs, results comparisons (3 for each organism).	**Pathogen → Sensitivity → Specificity** Human metapneumovirus → 33.3%[3] → 100% Parainfluenza IV → 33.3% → 100% SARS-CoV-2 → 100% → 100% RSV → 100% → 100% Influenza A → 100% → 100% Enterovirus → 100% → 100%

[1] Estimated PFU calculated by multiplying the $TCID_{50}$ /mL by 0.7 • [2]Nearest higher loading to that of the LOD (within 1-Log) • [3]Includes two of the three samples that did not pass internal control QC

High Concentration of EDTA Negatively Impacted Read Count



High Concentration of Blood Negatively Impacted Read Count



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-sample multiplex library preparation with 6 samples and 2 controls.
- Supports multiplex up to 6 samples with approximately 9 (1-plex) to 11 (6-plex) hours of hands-on sample preparation (no automation).
- Current limit of detection is higher than targeted PCR assays and near that of immuno-based assays.
- Detection limits appear to be associated with genome size and complexity of the genome (e.g. segmentation) of the virus.
- Sensitivity is 100% down to 33% - Discrepancy testing was not performed to confirm remnant sample original result .
- Specificity was at 100%.
- Results are in Preprint: medRxiv 2024.03.26.24304688 and original workflow development: medRxiv 2024.03.26.24304686

FUTURE DIRECTIONS

- Transitioning to R10 chemistry.
  - Increase in read quality and decrease in error is expected.
  - New R10 Kit is compatible with workflow.
- Reduce time from sample to prepared library.
  - WGA produces excess amounts of DNA, therfore incubation time can be reduced.
  - Reverse-transcription incubation time can be reduced and/or a faster enzyme kit could be used.
- Re-examine the use of the Rapid Barcoding Kit.
  - Will need to revisit barcode cross-talk.
- Reduce overall processing and time for a single-sample library preparation and automate that process (Currently funded).
- Scale the automated single-sample library preparation to be capable of processing multiple samples (6 to 12) for a single sequencing run.

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. Microbiol. 9. (And Citations Within)
3. F.C. Hewitt, et al. (2017) 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.
7. 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 utilizing the taxonomical inference module 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 HGS, CDC, ODNI, IARPA, ARO or the US Government.