3105
June 12, 2022

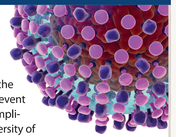
Evaluating Unbiased RNA Sequencing for Biothreat and Emerging Disease Detection

Signature Science, LLC
Austin TX, 78759

Amanda N. Scholes • Nicolette C. Keplinger • Matthew B. Scholz • Krista L. Ternus • F. Curtis Hewitt • Anthony D. Kappell*

+1 512.533.2092
akappell@signaturescience.com

ABSTRACT



RNA viruses pose a significant global biothreat. Beyond the current coronavirus pandemic, filoviruses (e.g., ebolaviruses), alphaviruses (e.g., *Venezuelan Equine Encephalitis Virus* (VEEV)), and flaviviruses (e.g., *Zika virus),* among others, cause recurring outbreaks in the Americas and around the world. Unfortunately, the detection of viral pathogens that are associated with a biothreat event or that could cause novel emerging infectious disease (EID) is complicated by three factors: 1) the high mutation rates and genetic diversity of RNA viruses; 2) the common requirement for reverse transcription prior to DNA-based characterization; and 3) the need for cell culture to propagate such viruses. Unbiased sequencing has the potential to overcome these factors to detect any pathogen present in a sample. Nucleic acid sequencing approaches continue to advance in speed and accuracy with lower associated costs. This trend is highlighted by nanopore sequencing, which combines real time sequence analysis with relatively inexpensive, disposable sequencing reagents. We evaluated and optimized a suite of standardized protocols using the Oxford Nanopore Technologies' (ONT) MinION sequencing device to rapidly and accurately detect RNA viruses categorized as biothreats by CDC. Four workflows were evaluated, including the input of synthesized double stranded cDNA into a cDNA Sequencing Kit or Rapid Sequencing Kit, Direct RNA Sequencing Kit, and Direct cDNA Sequencing Kit. We focused our study to *SARS-CoV-2* in remnant viral transport media and VEEV in human blood plasma using contrived clinical samples. These viruses represent wider classes of RNA viruses that should be detectable using similar sequencing approaches. The evaluation compared performance data including the sensitivity, limits of detection, and associated costs of the best unbiased sequencing method to current targeted sequencing methods. The improvement of performance by ribosomal RNA depletion and the applicability of sample multiplexing to decrease costs and increase scalability was also evaluated.

INTRODUCTION

RNA viruses pose a significant global biothreat. Beyond the current coronavirus pandemic, filoviruses (e.g., ebolaviruses), alphaviruses (e.g., Venezuelan Equine Encephalitis Virus (VEEV)), and flaviviruses (e.g., Zika virus), among others, cause recurring outbreaks in the Americas and around the world.

Unbiased, metagenomic sequence-based approaches could lead to early detection of both previously characterized and novel threats RNA virus threats.[1] It 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] Metagenomics for pathogen detection in public health can overcome current challenges in traditional methods. Techniques that require serial testing against a list of suspected pathogens or culturing can lead to delayed actionable results especially for slow-growing pathogens, such as *Mycobacterium tuberculosis,* while metagenomic approaches comprise a single test and are increasing in speed. While performing multiple tests can very expensive, a single test with a metagenomic approach can be performed and associated costs are declining making them more economically justifiable. These trends of increasing speeds and reduced cost is highlighted by nanopore sequencing, which can combine real time sequence analysis with relatively inexpensive, disposable sequencing reagents.

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] Rapid workflows for cDNA synthesis and direct RNA sequencing have been established as a foundation for unbiased viral sequencing, but significant work is needed to evaluate and validate the best performing methods to enable the implementation of these promising new tools in public health labs.

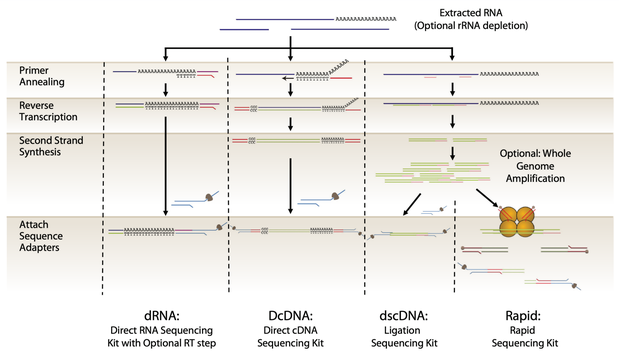
In this study, we evaluate methods using the Oxford Nanopore MinION sequencing device to rapidly and accurately detect RNA viruses in an unbiased manner. This approach capitalizes on the strengths of the sequencing device by generating sequence data for real time analysis to dramatically shorten the time required to sequence each sample, and critically, enabling workflows for unbiased sequencing to detect novel pathogens and EIDs.

MATERIALS and METHODS

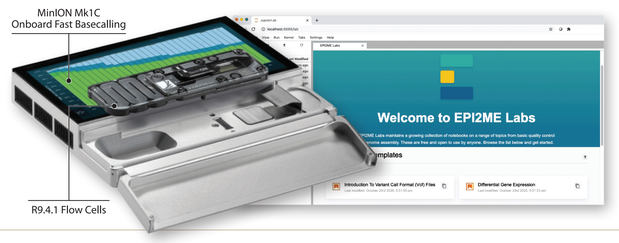
Viruses and Contrived Samples Used for Analysis

Virus	Matrix	Transmission	Genome Features	Relevant Pathogen
SARS-CoV-2	Remnant Viral Transport Medium (VTM)	Primarily Respiratory	+ Strand RNA Genome; Polyadenylated	• Other Betacoronaviruses (SARS-CoV, MERS-CoV) • Other Respiratory Pathogens
Human CoV 229E	VTM	Primarily Respiratory	+ Strand RNA Genome; Polyadenylated	• Other Alphacoronaviruses • Other Respiratory Pathogens
Venezuelan Equine Encephalitis Virus (VEEV)	Human Plamsa	Primarily Arthropod Vectors	+ Strand RNA Genome; Polyadenylated	• Other Alphaviruses (EEEV, WEEV, Sindbis, Chikungunya) • Other RNA arboviruses detected in blood samples (Hantavirus, West Nile Virus, Dengue Virus, Mayaro Virus)

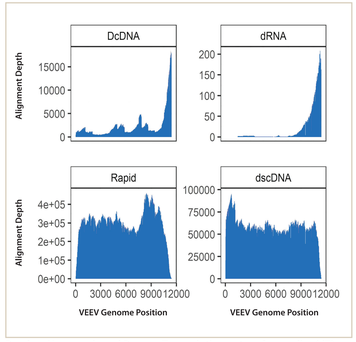
Library Preparation Workflows Examined



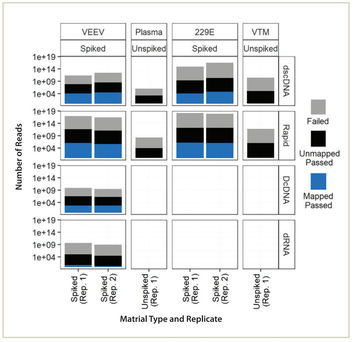
Sequencing and Analysis



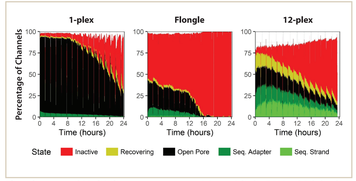
RESULTS



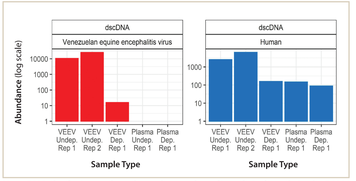
Random Hexamer based Reverse Transcription out performs oligo(dT) methods for coverage and depth.



Random Hexamer based methods also yielded more data for mapping and alignment.



Flongle use and Multiplex was most promising with dscDNA which had lowest utilization, but data loss was too high.



Host RNA depletion reduced number and quality of viral reads in dscDNA and Rapid methods.

Barcode	01	02	03	04	05	06	07	08	09	10	11	12	unclassified
Virus or Sample	VEEV	VEEV	VEEV	VEEV	VEEV	VEEV	NA	NA	NA	NA	NA	NA	NA
GE in total	8.21E+05	8.21E+05	8.21E+03	8.21E+03	8.21E+01	8.21E+01	0	0	0	0	0	0	0
GE/mL in Sample	1.95E+07	1.95E+07	1.95E+05	1.95E+05	1.95E+03	1.95E+03	0	0	0	0	0	0	0
Replicate	1	2	1	2	1	2	0	0	0	0	0	0	0
Mean Depth	2954.70	3914.28	150.02	137.55	47.41	66.24	0.00	0.00	0.00	0.18	0.00	0.00	1071.17
Coverage (%, minimap2)	99.99	99.97	98.34	98.44	90.33	92.42	0.00	0.00	0.00	2.52	0.00	0.00	99.98
Coverage (%, blastn)	99	99	99	99	94	95	0	0	0	10	0	0	99
Identity of Coverage (%)[1]	96.34	96.37	96.29	96.33	96.94	97.70	0.00	0.00	0.00	97.57	0.00	0.00	96.34
N50	3376	2940	2734	3476	3483	3496	6202	0	3145	4028	4225	1642	3813
Longest Alignment (Kb)	3.19	2.96	1.84	3.15	1.84	2.86	0	0	0	0.29	0	0	2.13
Total Reads for a Barcode	98649	139714	136040	108623	82638	120026	13	0	22	9	15	32	237016
Percent of Total Reads in Run	10.69	15.14	14.74	11.77	8.96	13.01	0.00	0.00	0.00	0.00	0.00	0.00	25.68
Passing Quality Reads (%)	68.51	68.40	69.74	68.21	71.28	68.91	15.38	0.00	13.64	22.22	6.67	9.38	26.84
Mapped Passed Reads to Virus (%)	25.89	26.83	0.96	1.01	0.48	0.51	0.00	0.00	0.00	50.00	0.00	0.00	10.21
Passing bp (%)	66.86	66.20	67.02	65.28	68.15	66.83	22.72	0.00	16.68	23.07	18.54	4.13	31.48

Threshold limit for study was set to >90% identity over >90% coverage of the genome.

Best Method: REPLI-g supplemented dscDNA Method was successfully multiplex (6) and LOD ≤ 1.95E03 Genomic Equivalence of VEEV per mL of Human Plasma Sample.

TABLE NOTES
- REPLI-g increase overall data (read counts and passed bp) which generated greater depth, identity, and coverage of the genome for identification.
- 6-plex of the REPLI-g dscDNA (and Rapid) also showed less barcode cross-talk compared to 12-plex standard dscDNA and Rapid methods.
- Unclassified read not assigned a barcode represents a potential reservoir of additional data.
- Limit of detection will be assessed in single-plex and multiplex in future studies.

CONCLUSIONS

cDNA Synthesis	REPLI-g Amplification	End Repair and Ligation	Loading and Sequencing
Hands-on time: 45 min (6-plex: 1 hr) Incubation time: 2.5 hrs	Hands-on time: 45 min (6-plex: 1 hr) Incubation time: 3 hrs	Hands-on time: 45 min (6-plex: 1 hr) Incubation time: 45 min (6-plex: 1.25 hr)	Hands-on time: 30 min Incubation time: up to 24 hrs

- Best method identified was REPLI-g assisted dscDNA: Double-stranded cDNA synthesis followed by REPLI-g WGA (with T7 endonuclease digestion) as input into the Ligation Sequencing Kit (or Native Barcoding Kit for Multiplex).
- Supports multiplex up to 6 samples with approximately 9 (1-plex) to 11 (6-plex) hours of sample preparation.
- Run time as low as 6 hrs.
- Cost ranges from $750 (1-plex) to $1565 (6-plex).
- Can detect viral genomes extracted as little as 1.95E03 Genomes/mL sample.
- Further study to determine sensitivity and limit of detection is underway.

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

ACKNOWLEDGEMENTS

This work was financially supported by the Centers for Disease Control and Prevention under award number 75D301-21-C12250. The opinions, findings, and conclusions or recommendations expressed in this presentation are those of the author(s).