Pathogen detection in human blood and plasma using long read sequencing

Sean Smith[1], Moussa Kourout[2], Scott Espich[2], Carolyn Fisher[2], Irina Tiper[2], Anjan Purkayastha[3], Luis Santana-Quintero[1], Robert Duncan[2]

1 – Office of Biostatistics and Pharmacovigilance (OBPV), Center for Biologics Evaluation and Research (CBER), Food and Drug Administration (FDA), Silver Spring, MD, USA

2 - Division of Emerging and Transfusion Transmitted Diseases, Office of Blood Research and Review (OBRR), Center for Biologics Evaluation and Research (CBER), Food and Drug Administration (FDA), Silver Spring, MD, USA

3 – Openbox Bio, Vienna, VA, USA

Abstract

Evaluated potential use of nanopore sequencing for multiple pathogen blood testing.

Introduction: Pathogen detection in the blood supply may involve multiple tests. Utilizing multiplexed sets of primers, we developed a long-read sequencing pipeline to identify multiple pathogens simultaneously and compared results with a microarray test.

Methods: Multiplexed sets of primers were used to amplify nucleic acids extracted from viral, bacterial, and protozoan pathogen-spiked blood samples. Nucleotide sequences were determined using an Oxford Nanopore Minion sequencer and aligned to a reference of the targeted pathogen sequences. These results were compared to an approach that hybridized the amplified nucleic acids to a microarray as part of the Blood Borne Pathogen Resequencing Micorarray Expanded (BBP-RMAv.2) platform. Human blood or plasma samples were spiked with one of 6 viruses, 2 bacteria, and 5 protozoans.

Results: Nanopore sequencing correctly identified most samples, however, the BBP-RMAv.2 platform had a higher accuracy. Similarity of targeted regions with human and other species presented challenges limiting the sensitivity and specificity of the nanopore sequencing pipeline.

Conclusion and/or Implications: Nanopore sequencing showed potential for use as part of a multiple pathogen blood test. Results suggest additional improvements are needed to reach performance levels of alternative methods such as microarrays. Areas of further development for potential improvement in performance include selection of target regions, alignment filtering, and optimization of aligner parameters.

Introduction

Viral, bacterial, and protozoan pathogens may be present in blood donations and can occur at low concentrations. Ensuring blood supply safety often includes multiple methodologies such as antibody testing, nucleic acid testing, and donor screening. Our goal was to develop a sensitive single-donor test capable of detecting a wide range of pathogens reducing the need for multiple testing. Next generation sequencing (NGS) is an agnostic approach with the potential to detect multiple pathogens using one platform. Utilizing the long-read sequencer, Oxford Nanopore MinION, may provide an opportunity to rapidly detect pathogens with high sensitivity and accuracy.

Disclaimer

This poster reflects the views of the authors and should not be construed to represent FDA's views or policies.

Materials and methods

A full description of materials and methods can be found in Kourout, 2024. Briefly, pathogen genome regions varying from 45 to 876 nucleotides were selected for targeted sequencing. Primers were designed for PCR amplification of the targeted regions. The functionality of the primers was tested under multiplex PCR conditions. Viral copies and bacterial and protozoan cells were spiked into plasma (virus) or blood (bacteria, protozoa) at 100-100,000 / mL. Samples were processed through extraction, reverse transcription, multiplex PCR reactions, pooling, and cleanup, and the Oxford Nanopore Amplicons-by-Ligation kit (SQK-LSK112) (Oxford Nanopore Technologies, 2023) or ligation sequencing amplicons-native barcoding kit (SQKNBD112.24) (Oxford Nanopore Technologies, 2021) was used for library prep from about 300 ng of DNA of the pooled PCR product following manufacturer's instructions. Sequencing data was produced, and quality filtered using a MinION Mk1b with a SpotON Flow Cell, version R9 (FLO-MIN106D, Oxford Nanopore Technologies, Oxford, United Kingdom) and MinKNOW software, version 22.03.5 (Oxford Nanopore Technologies, Oxford, United Kingdom). Sequence files were transferred to the High-Performance Integrated Virtual Environment (HIVE) for storage and data analysis. Using Minimap2 (Li, 2018), reads aligned (mapping quality ≥20) to a unique reference file including targeted pathogen regions, human 18S rRNA decoy regions, and negative controls were reported. The unique reference file was created from 37 tile sequences for pathogens of interest, 2 tile sequences for controls, and a human 18S rRNA sequence.

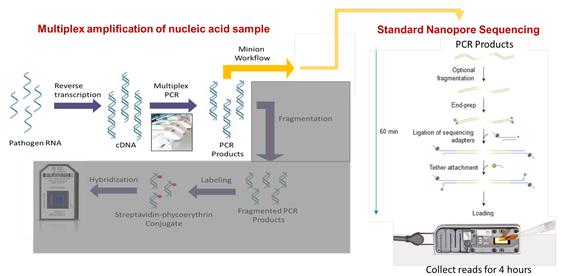

Figure 1. MinION Nanopore Sequencing Workflow

Table 1. Blastn alignment to nt database for *S. epidermidis* 100,000 cells/mL spike-in blood.

Rank	Species	Unique Hits
1	Unaligned	27811
2	MF164260.1 Homo sapiens clone BAC JH15 genomic sequence	3526
3	AY046759.1 Uncultured eukaryote isolate A3_E004 18S ribosomal RNA gene, partial sequence	931
4	XR_004921155.1 PREDICTED: Halichoerus grypus 18S ribosomal RNA (LOC118543604), rRNA	924
5	XM_020154710.1 PREDICTED: Castor canadensis collagen alpha-1(I) chain-like (LOC109679536), mRNA	847
…	…	…
2051	CP034111.1 Staphylococcus epidermidis strain CDC120 chromosome, complete genome	1

Results and discussion

Nanopore sequencing from the spiked-in samples was aligned to the NCBI nt database using Blastn (Altschul, 1990). This produced excessive false positives as shown for a Staphylococcus epidermidis 100,000 cells / mL blood spike-in where the spike-in pathogen was the 2051st most common alignment (Table 1).

To reduce noise, a custom reference fasta file was created containing:

- Targeted pathogen sequences
- Control sequences

Observations after applying this approach with Minimap2 alignment included:

- Greatly reduced false positives
- Misalignments to 18S rRNA regions still limited specificity

Addition of 18S rRNA human sequence with similarity to targeted pathogen 18S rRNA regions reduced misalignments as shown for P. vivax 1,000 cells/mL spike-in blood where T. brucei 18S rRNA misalignments were not reported after addition of the human decoy sequence (Table 2).

In tables 3 and 4, comparison of microarray and nanopore long-read sequencing for 3 negative control samples and 4 viral, 2 bacterial, and 7 protozoan spike-in samples indicated less control false positives for nanopore (0) compared to microarray (2). Pathogen detection for the 13 spiked-in pathogen samples was observed in more samples for the microarray platform (12 true positive, 0 possible positive) compared to nanopore (8 true positive, 2 possible positive). 'Possible positive' indicates uncertainty of pathogen detection due to a low number of reads aligning to the pathogen reference sequence and/or ambiguity due to read alignments to multiple pathogen sequences.

Table 2. Read alignment before and after add 18S rRNA human decoy to custom reference for *P. vivax* 1,000 cells/mL spike-in blood. Excluded regions with no aligned reads.

Region	No human 18s rRNA decoy	With human 18s rRNA decoy
human 18S rRNA	NA	118069
NAC (control)	2	2
P. Vivax 18S rRNA	6	5
T. Brucei 18S rRNA	2	0

Table 3. Summary of microarray and MinION nanopore results for 3 negative control samples and 4 viral, 2 bacterial, and 7 protozoan spike-in samples. TP, true positive; FP, false positive; TN, true negative; FN, false negative; PP, partial positive.

	All		Viral		Bacterial		Protozoan	
Interpretation	Microarray	Nanopore	Microarray	Nanopore	Microarray	Nanopore	Microarray	Nanopore
TP	12	8	4	3	2	0	6	5
TN	1	3	0	0	0	0	0	0
PP	0	2	0	0	0	2	0	1
FP	2	0	0	0	0	0	0	0
FN	1	3	0	1	0	0	1	1

Table 4. Microarray and MinION nanopore results for 3 negative control samples and 4 viral, 2 bacterial, and 7 protozoan spike-in samples. TP, true positive; FP, false positive; TN, true negative; FN, false negative; PP, possible positive.

Test Sample	Microarray Results (C3 score)	Interpretation	MinION Minimap2(>20) Result (score)	Interpretation
Chikungunya Virus 1,000 copies/ml Plasma	CHIKV (93)		CHIKV (269)	
Dengue Virus Type 2 100 copies/ml Plasma	DENV2 (53)		E. coli (1)	
HIV-1 GrpM 10,000 copies/ml Plasma	HIV1M (86)	TP	HIV (23)	TP
Sindbis Virus 1,000 copies/ml Plasma	Sindbis (100)	TP	Sindbis V. (16)	TP
Negative Control Plasma	Ld (21)	FP	None (0)	TN
Negative Control Plasma	HCV (30.5)	FP	None (0),	TN
Babesia microti 10,000 cells/ml Blood	Bm (92)	TP	*B. microti* (11)	TP
Babesia microti 1,000 cells/ml Blood	(0)	FN	None (0)	FN
Leishmania aethiopica 10,000 cells/ml Blood	*L. aeth* (82)		*L. aeth.* (654)	
Leishmania aethiopica 1,000 cells/ml Blood	*L. aeth.* (86)		*L. aeth.* (104)	
Plasmodium vivax 10,000 cells/ml Blood	PLVI (88)		*P. vivax* (8)	
Plasmodium vivax 1,000 cells/ml Blood	PLVI (75)	TP	*P. vivax* (5)	
Trypanosoma brucei 10,000 cell/ml Blood	*T brucei* (29)		*T. brucei* (3)	PP
Staphylococcus epidermidis 100K/ml Blood	*S. epi* (75)		*S. epi* (1)	
Y. pseudotuberculosis 10,000 cells/ml Blood	*Y. pseudo* (43)	TP	None (0)	FN
Negative Control Whole Blood	None	TN	None (0)	TN

Conclusion

With the use of a custom reference of targeted sequences and human 18S rRNA decoy sequence, nanopore sequencing demonstrated potential for use as part of a multiple pathogen blood test. Initial results suggest additional improvements are needed to increase sensitivity for detecting pathogens that may exist at low levels in blood and plasma. Potential areas of further investigation to improve performance include selection of target regions, alignment filtering, optimization of aligner parameters, and increased sequencing depth.

References

Kourout M, Espich S, Fisher C, Tiper I, Purkayastha A, Smith S, Santana-Quintero L, Duncan R. (2024) Multiplex detection and identification of viral, bacterial, and protozoan pathogens in human blood and plasma using an expanded high-density resequencing microarray platform. Front Mol Biosci. 11:1419213. doi: 10.3389/fmolb.2024.1419213. PMID: 38966129; PMCID: PMC11222771.

Oxford_Nanopore_Technologies (2021). Ligation sequencing amplicons – native barcoding (SQK-NBD112.24). Oxford, United Kingdom: ONT website. Available at: https://community.nanoporetech.com/docs/prepare/library_prep_protocols/nativebarcoding-kit-24-with-amplicons-sqk-nbd112-24/v/nba_9135_v112_revl_01dec2021.

Oxford_Nanopore_Technologies (2023). Ligation sequencing amplicons (SQKLSK112). Oxford, United Kingdom. Available at: https://community.nanoporetech.com/docs/prepare/library_prep_protocols/amplicons-by-ligation-sqk-lsk112/v/acde_9142_v112_revi_01dec2021.

Li, H. (2018).Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34 (18), 3094–3100. doi:10.1093/bioinformatics/bty191

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990) Basic local alignment search tool. J. Mol. Biol. 215:403-410..