Validation of Influenza Sequencing on Illumina and Oxford Nanopore Technologies Platforms

Cherrelle Barnes[1,2], Brian Schwem[2], Shannon Delaney[2], Maria-Magdalene Pugliese[2], Dana Woell[2], Richard Siderits[2] and Thomas Kirn[2]

1. APHL-CDC Public Health Laboratory Fellowship

2. New Jersey Department of Health, Public Health and Environmental Laboratories, Ewing, NJ 08628

Introduction

Surveillance of Influenza A and B viruses is a concerted effort between local/state public health laboratories and the Centers for Disease Control and Prevention, and is incredibly important as its genome is dynamic, resulting in small mutations and major reassortment events. Next-generation sequencing has become a pivotal tool to monitor these changes over time, which in turn provides vital information for public health decisions to be made for the upcoming Influenza season.

To increase the depth of Influenza surveillance in New Jersey, we aimed to adopt the CDC's Multi-Segment Reverse Transcription PCR (MRT-PCR), a one-step protocol for generation and amplification of Influenza cDNA. Furthermore, we validated library preparation and sequencing of Influenza on both the Illumina and Oxford Nanopore Technologies (ONT) platform. The validation studies vetted the accuracy, precision, sensitivity and specificity of MRT-PCR on both platforms.

Methodology

Initially, Influenza specimens are processed using a RT-PCR assay to determine if it is positive or negative for Influenza. The assay will also provide a corresponding cycle threshold (Ct) value. The same specimen were used for these validation studies. Generally, the specimen are used for MRT-PCR and cleaned using AMPure XP beads and ethanol afterwards. Subsequently, the cleaned PCR products move forward though the library preparation and sequencing on each platform. Finally, the sequencing results are analyzed in Terra using the respective workflows for Illumina and ONT.

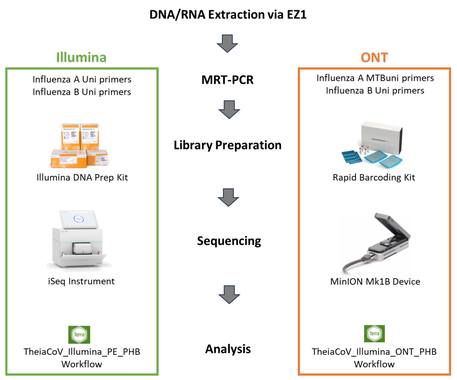


Figure 1. Method for Influenza Sequencing. The schematic shows the overall steps from extractions to analysis. The green and orange box highlights specifics of library preparation, sequencing instrument/device and analysis workflow for the Illumina and ONT platform, respectively.

Accuracy Study

Type/Subtype: The sequencing results were compared to the RT-PCR assay to assess if the correct type and subtype was determined. Illumina and ONT demonstrated 100% accuracy for type/subtype on Influenza samples.

Sample ID	Ref % Cov	Type/Subtype (FluSC2)	Type/Subtype (MRT-PCR SEQ)
INFA_PC	101.02	INF A / H1N1	INF A / H1N1
INFB_PC	102.92	INF B / VIC	INF B / VIC
V22029057	100.98	INF A / H3N2	INF A / H3N2
V22029847	99.16	INF A / H3N2	INF A / H3N2
V22030183	68.66	INF A / H3N2	INF A / H3N2
V22041734	98.67	INF A / H3N2	INF A / H3N2
V22042786	101.02	INF A / H3N2	INF A / H3N2
V22044128	99.66	INF A / H1N1	INF A / H1N1
V22045381	101.02	INF A / H3N2	INF A / H3N2
V22045703	99.69	INF A / H3N2	INF A / H3N2

Figure 2. Example of Influenza Type and Subtype Accuracy. The table provides sample ID and the percent coverage of the reference genome of the sequencing result for 10 of 54 samples sequenced on the Illumina platform. The type and subtype was compared between the RT-PCR assay (Flu-SCR) and sequencing (MRT-PCR SEQ).

Hemagglutinin (HA)/ Neuraminidase (NA) Clade: The sequencing results were compared to CDC's sequencing results to verify if the same clade was called. Illumina showed 100% accuracy, while ONT had 98% accuracy for the HA and NA clade determination.

Sample ID	Ref % Cov	CDC HA Clade	PHEL HA Clade	CDC NA Clade	PHEL NA Clade
INFA_PC	101.02	6B.1A.5a.1	6B.1A.5a.1	B.2	B.2
INFB_PC	102.92	V1A	V1A	A.1	A.1
V22029057	100.98	3C.2a1b.2a.2a.3	3C.2a1b.2a.2a.3	B.3	B.3
V22029847	99.16	3C.2a1b.2a.2a.1	3C.2a1b.2a.2a.1	B.2	B.2
V22030183	68.66	3C.2a1b.2a.2a.1	3C.2a1b.2a.2a.1	B.2	B.2
V22041734	98.67	3C.2a1b.2a.2a.1	3C.2a1b.2a.2a.1	B.2	B.2
V22042786	101.02	3C.2a1b.2a.2a.3	3C.2a1b.2a.2a.3	B.3	B.3
V22044128	99.66	6B.1A.5a.2a	6B.1A.5a.2a	C.2	C.2
V22045381	101.02	3C.2a1b.2a.2b	3C.2a1b.2a.2b	A.2.2.2	A.2.2.2
V22045703	99.69	3C.2a1b.2a.2a.3b	3C.2a1b.2a.2a.3b	B	B

Figure 3. Example of Influenza Clade Accuracy. The table provides sample ID and the percent coverage of the reference genome of the sequencing result for 10 of 54 samples sequenced on the Illumina platform. The hemagglutinin (HA) and neuraminidase (NA) clade determination was compared between the CDC's and NJ PHEL's sequencing results.

Precision Study

A set of 7 or 8 samples were processed by three different operators, on different days and sequenced on two different instruments for the Illumina and ONT platforms. Each sample was sequenced a total of 6 times on each platform to determine if the same type/subtype and clade was called.

Illumina showed 100% precision

User A			
Sample ID	% Coverage	Type/Subtype	HA Clade
V22029847	99.16	INF A / H3N2	3C.2a1b.2a.2a.1
V22042786	101.02	INF A / H3N2	3C.2a1b.2a.2a.3
V23012230	97.3	INF A / H1N1	6B.1A.5a.2a.1
V22044128	99.66	INF A / H1N1	6B.1A.5a.2a
V23009768	89.08	INF B / VIC	V1A.3a.2
INF_A PC	101.02	INF A / H1N1	6B.1A.5a.1
INF_B PC	102.92	INF B / VIC	V1A
User B			
Sample ID	% Coverage	Type/Subtype	HA Clade
V22029847	99.58	INF A / H3N2	3C.2a1b.2a.2a.1
V22042786	99.67	INF A / H3N2	3C.2a1b.2a.2a.3
V23012230	91.19	INF A / H1N1	6B.1A.5a.2a.1
V22044128	101.02	INF A / H1N1	6B.1A.5a.2a
V23009768	62.52	INF B / VIC	V1A.3a.2
INF_A3	101.02	INF A / H1N1	6B.1A.5a.1
INF_B2	105.42	INF B / VIC	V1A
User C			
Sample ID	% Coverage	Type/Subtype	HA Clade
V22029847	100.61	INF A / H3N2	3C.2a1b.2a.2a.1
V22042786	101.02	INF A / H3N2	3C.2a1b.2a.2a.3
V23012230	97.32	INF A / H1N1	6B.1A.5a.2a.1
V22044128	100.89	INF A / H1N1	6B.1A.5a.2a
V23009768	89.88	INF B / VIC	V1A.3a.2
INF_A3	101.02	INF A / H1N1	6B.1A.5a.1
INF_B2	105.42	INF B / VIC	V1A

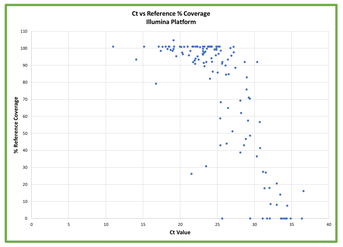
ONT showed 100% precision

User A			
Sample ID	% Coverage	Type/Subtype	HA Clade
V22045381	99.58	INF A / H3N2	3C.2a1b.2a.2b
V22047301	100.18	INF A / H3N2	3C.2a1b.2a.2b
V22048210	100.89	INF A / H3N2	3C.2a1b.2a.2b
V22050106	101.02	INF A / H3N2	3C.2a1b.2a.2b
V22050609	100.96	INF A / H3N2	3C.2a1b.2a.2b
V22052705	101.03	INF A / H3N2	3C.2a1b.2a.2a.1b
INF_A2	98.61	INF A / H3N2	3C.2a1b.2
User B			
Sample ID	% Coverage	Type/Subtype	HA Clade
V22045381	101.03	INF A / H3N2	3C.2a1b.2a.2b
V22047301	101.03	INF A / H3N2	3C.2a1b.2a.2b
V22048210	101.03	INF A / H3N2	3C.2a1b.2a.2b
V22050106	101.02	INF A / H3N2	3C.2a1b.2a.2b
V22050609	101.03	INF A / H3N2	3C.2a1b.2a.2b
V22052705	101.03	INF A / H3N2	3C.2a1b.2a.2a.1b
INF_A2	101.02	INF A / H3N2	3C.2a1b.2
User C			
Sample ID	% Coverage	Type/Subtype	HA Clade
V22045381	101.03	INF A / H3N2	3C.2a1b.2a.2b
V22047301	101.03	INF A / H3N2	3C.2a1b.2a.2b
V22048210	101.03	INF A / H3N2	3C.2a1b.2a.2b
V22050106	101.03	INF A / H3N2	3C.2a1b.2a.2b
V22050609	101.03	INF A / H3N2	3C.2a1b.2a.2b
V22052705	101.03	INF A / H3N2	3C.2a1b.2a.2a.1b
INF_A2	101.03	INF A / H3N2	3C.2a1b.2

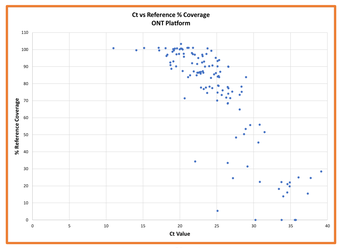
Figure 4. Precision Study Results. The table provides the sample ID and percent coverage of the reference genome for the sequencing results. The type/subtype and hemagglutinin (HA) clade determination for each user on the Illumina (green table) and ONT (orange table) platforms.

Sensitivity Study

Over 90 positive Influenza specimens with Ct values ranging from 10 to 40 were sequenced on both platforms and plotted against the resulting percent reference coverage to determine a cut-off that would pass quality metrics for submission to National Center for Biotechnology Information (NCBI) (≥ 90% coverage; 50x minimum depth).



Illumina cut-off value = 28



ONT cut-off value = 25

Figure 5. Analytical Sensitivity Results. The graph plots the Ct value against the percent reference coverage for sequencing results for the Illumina (top, green) and ONT (bottom, orange) platforms.

Specificity Study

A set of 8 samples that tested negative for Influenza were used for MRT-PCR with Influenza A and Influenza B primer pool. The percent reference coverage was assessed to ensure only Influenza would successfully sequence on both platforms.

Illumina Platform			
Sample ID	Pathogen Detected	% Ref Cov (A)	% Ref Cov (B)
V23017555	None - True Negative	0	0
V23017562	None - True Negative	0	0
V23009007	Rhinovirus/Enterovirus	0	0
	Adenovirus		
V23010342	Adenovirus	0	0
V23010477	Rhinovirus/Enterovirus	0	0
V23005120	Human coronavirus	0	0
INF_A PC	INF A / H1N1	101.02	61.39
INF_B PC	INF B / VIC	44.59	102.92

ONT Platform			
Sample ID	Pathogen Detected	% Ref Cov (A)	% Ref Cov (B)
V23017555	None - True Negative	47.93	49.45
V23017562	None - True Negative	29.98	40.47
V23009007	Rhinovirus/Enterovirus	49.09	37.42
	Adenovirus		
V23010342	Adenovirus	32.71	46.62
V23010477	Rhinovirus/Enterovirus	36.82	53.21
V23005120	Human coronavirus	36.23	29.37
INF_A PC	INF A / H1N1	101.05	45.45
INF_B PC	INF B / VIC	83.58	105.36

Figure 6. Analytical Specificity Results. The table provides the sample ID and pathogen detected for each samples sequenced on the Illumina (top, green table) and ONT (bottom, orange table) platform. MRT-PCR was performed using the Influenza A primers (A) and Influenza B primers (B). The samples that would be considered for submission to NCBI are highlighted in yellow.

Conclusions

Overall, we were able to successfully sequence Influenza A and B viruses on the Illumina and ONT platforms. Both platforms were able to demonstrate >98% accuracy for the type/subtype and clade assignment. Illumina and ONT showed 100% precision and 100% between operators, days and instruments.

Some differences between the platforms were observed in the sensitivity and specificity. Illumina showed that samples with a Ct value up to 28 were sequenced with the required quality metrics for submission to NCBI. The ONT platform showed to be less sensitive with a Ct cut off value of 25, meaning less samples were meeting the required quality metrics. We are interested in optimizing ONT to be as sensitive as Illumina. The specificity study revealed that only Influenza specimens are be successfully sequenced using this protocol. However, the ONT showed some generation of a sequencing result, but does not meet the required quality for submission to NCBI.

The benefit of having two sequencing platforms includes having complementary short and long read for genome resolution to identify unique strains/mutations, the ability to scale up or down sequencing if needed and to flex to a different platform given procurement supply/demand issues. In conclusion, we believe that both platforms are viable for Influenza sequencing in an effort to bolster surveillance of the virus in New Jersey.

References

- Centers for Disease Control & Prevention Laboratory Standard Operating Procedure for:
  - "Multi-Segment Reverse Transcription PCR (MRT-PCR) of Influenza A and B Viruses"
  - "MRT-PCR of Influenza A Viruses for Sequencing"
- Illumina - DNA Prep Reference Guide
- Oxford Nanopore Technologies – Rapid Sequencing gDNA Barcoding Kit V14

Acknowledgements

- APHL-CDC Laboratory Fellowship
- NJ Public Health and Environmental Laboratories
- Dana Woell
- Brian Schwem
- Shannon Delaney
- Maggie Pugliese
- Byeong Jeong
- Molecular Virology Unit