epiXact®: Same-day transmission analysis of nosocomial transmission using nanopore whole genome sequencing

Mohamad Sater[1], Remy Schwab[1],Tim Farrell[1], Ian C. Herriott[1], and Miriam Huntley[1]
[1] Day Zero Diagnostics, Inc. (Boston, MA)

Introduction

Healthcare associated infections (HAIs) are a major contributor to patient morbidity and mortality worldwide. HAIs are increasingly important due to the rise of multidrug resistant pathogens which can lead to deadly nosocomial outbreaks. Current methods for investigating transmissions are slow, costly, or have poor detection resolution. A rapid, cost-effective and high-resolution method to identify transmission events is imperative to guide infection control. Whole genome sequencing of infecting pathogens paired with a single nucleotide polymorphism (SNP) analysis can provide high-resolution clonality determination, yet these methods typically have long turnaround times. Here we examined the utility of the Oxford Nanopore Technologies (ONT) platform, a rapid sequencing technology, for whole genome sequencing based transmission analysis. We introduce a SNP calling pipeline customized for ONT data, which exhibit higher sequencing error rates and can therefore be challenging for transmission analysis. The pipeline leverages the latest basecalling tools as well as a suite of custom variant calling and filtering algorithms to achieve highest accuracy in clonality calls compared to short-read based sequencing. We also capitalize on ONT long-reads by assembling outbreak-specific genomes in order to overcome the need for an external reference genome. We demonstrate the utility of ONT for HAI investigation, establishing the potential to transform healthcare epidemiology with same-day high-resolution transmission determination.

epiXact®, DZD commercial Illumina-based HAI sequencing and analysis service, is used by multiple partnering hospitals to investigate a wide variety of suspected outbreaks (in both clinical and laboratory settings). When infection control (IC) suspects transmission, cultured bacterial samples are sent to DZD; epiXact genomic relatedness analysis identifies sometimes clonal transmission, in other cases not. Results are reported back in 2-3 days allowing IC to use the definitive genomic evidence to inform decisions regarding ward cleaning, staff screening, and equipment contamination.

Comparison of whole genome sequencing technologies for HAI

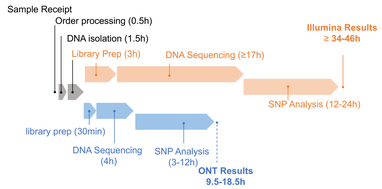
Whole genome sequencing (WGS) is a well-established, high-resolution method for measuring pathogen relatedness to better understand infectious disease transmission.
ONT sequencing offers many advantages with faster speed and lower costs over short-read technologies that make this an attractive platform for a commercial HAI investigation service that is rapid, scalable and cost effective. However, the lower single base accuracy has challenged the utility of ONT data for SNP discrimination and outbreak investigations which depend on accurate quantification of genomic relatedness.

	ONT MinION	Illumina iSeq100 (fastest turn around Illumina product)
Single base accuracy	95% (1D)	>99%
Mean read length	1000-50,000bp	300bp
Typical data output	0-15Gbp (as needed)	1.2Gbp
Library preparation time	30 min	180 min
Data acquisition	Real-time (<300Mbp/hour)	Post-sequencing (1.2 Gp after 17 hours)
Capital investment	$1,000	$20,000
Cost per run*	$33 per 1Gbp	$540 per 1Gbp
Samples per run	1-50**	5-6 depending on genome size

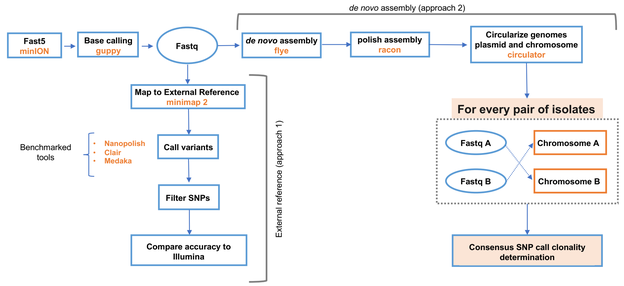
**Estimated, ONT based on flow cell bulk discount*
***Requires empirical testing*
Data adapted from Rang et al. (2018)

Turn-around time of Day Zero Diagnostic (DZD) HAI pipelines.

DZD standard WGS HAI service (epiXact®) uses Illumina short-read sequencing for clonality determination, and results are typically delivered in ≥34-46 hours (orange pipeline).
Our ONT based WGS approach (**epiXact-ONT,** blue pipeline) allows us to reduce this time to 9.5-18.5 hours, **enabling a same-day HAI service** including sequencing, bioinformatic analysis and reporting.



ONT bioinformatic analysis pipeline

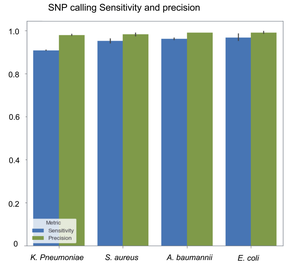


Establishing SNP calling accuracy (Approach 1): We finetuned our ONT SNP calling pipeline to maximize accuracy as compared to Illumina (short read gold standard). We benchmarked several variant calling tools, and developed a custom SNP filtering step that removes erroneously called SNPs based on observed error profiles. We compared the performance of the tools to previous HAI projects performed at Day Zero using epiXact® with Illumina WGS data mapped to an external reference genome to select the most accurate ONT SNP calling pipeline.
Standalone reference free ONT pipeline (Approach 2): To capitalize on ONT long-read data we apply the SNP calling pipeline developed in Approach 1 to *de novo* assemblies of all samples produced from ONT WGS, to reproduce clonality calls obtained using external references.

epiXact-ONT performance using real HAI cases

For this analysis 20 bacterial isolates from 5 HAI investigations were selected to represent different nosocomial species as well as to showcase performance on both clonal transmission and non-clonal cases. The samples are obtained from a mix of patient and environmental samples, e.g. multiple patients in NICU with MRSA of which 2/3 had clonal samples. Additional details on one of the cases, *A. baumannii*, can be found in (Shenoy *et al* 2020).

epiXact-ONT pipeline (Approach 1, against an external reference genome) achieved greater than 90% SNP-calling sensitivity and precision, **yielding 100% accuracy of clonality classification** compared to Illumina-based results across four common HAI species.



Accurate clonality calling achieved using ONT *de novo* assemblies

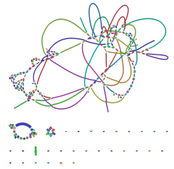
The availability of an external reference genome that is high quality as well as closely related to the outbreak strain is critical for accurate SNP calling[2]. However, for most species it can be very challenging to find a reference genome that is a good match to the outbreak strain. A low quality or distant reference genome will reduce the genomic resolution by limiting the analysis to the core-genome thereby missing large parts of the accessory genome. Furthermore, SNP calling accuracy drops significantly when mapping to a highly divergent reference[2,3].
Short-read WGS limits the use of a *de novo* assembly genome from an outbreak strain as a mapping reference. Long-read WGS, conversely, suffers from higher rates of sequencing errors which can lower SNP calling accuracy. Given the results of "Approach 1" overcoming ONT higher error rates, we demonstrate the validity of our SNP analysis pipeline using assembled genomes (Approach 2). Applied to the outbreak cases across 4 common HAI species with varying genome sizes and compositions, we demonstrate **100% clonality call concordance** between external reference short-read mapping and ONT long-read assembly and mapping. The epiXact-ONT pipeline capitalizes on long-read WGS to maximize SNP calling accuracy and resolution by using an outbreak specific reference genomes.

				Illumina reads	ONT reads
Sample 1	**Sample2**	**Species**	**Clonality**	**External reference**	**ONT assembly**
E-1	E-2	*E. coli*	Distant	97181	112962
E-1	E-3		Distant	97207	112954
E-2	E-3		Clonal	8	19
S-h1-1	S-h1-2	*S. aureus*	Distant	39671	45551
S-h1-1	S-h1-3		Distant	39665	45551
S-h1-2	S-h1-3		Clonal	2	0
S-h2-2	S-h2-3		Clonal	5	0
S-h2-2	S-h2-4		Clonal	3	2
S-h2-2	S-h2-5		Clonal	4	3
S-h2-2	S-h2-6		Distant	29387	54710
S-h2-3	S-h2-4		Clonal	9	3
S-h2-3	S-h2-5		Clonal	8	3
S-h2-3	S-h2-6		Distant	29413	54709
S-h2-4	S-h2-5		Clonal	5	7
S-h2-4	S-h2-6		Distant	29372	54716
S-h2-5	S-h2-6		Distant	29366	54710
K-1	K-2	*K. pneumoniae*	Clonal	5	2
A-2	A-3	*A. baumannii*	Clonal	18	26
A-2	A-4		Clonal	15	3
A-2	A-5		Clonal	12	30
A-2	A-6		Clonal	11	1
A-3	A-4		Clonal	20	27
A-3	A-5		Clonal	14	2
A-3	A-6		Clonal	11	10
A-4	A-5		Clonal	14	29
A-4	A-6		Clonal	6	1
A-5	A-6		Clonal	12	8

Results of SNP calling and clonality determination using ONT *de novo* assembled outbreak strain as shown in Approach 2 versus external reference / Illumina gold standard. Orange highlighted rows indicate clonal cases.

A. Baumanni long-read assembly

A. baumanni short-read assembly



A graphical representation of the same genome assembly produced by ONT long-reads (left) compared to the assembly produced using Illumina short-reads (right). Our long-reads assembly pipeline produced more contiguous sequences (often fully circular as above case), with >99% per-base accuracy.
Our pipeline leverages ONT long-read WGS to produce high quality *de novo* assemblies that can be used as outbreak specific references for mapping and SNP calling, omitting the need for an external reference genome.

Conclusions

- epiXact-ONT achieves high SNP calling sensitivity and precision
- We show the validity of our SNP analysis pipeline using ONT assembled genomes from each outbreak allowing a reference free, truly species agnostic, and outbreak specific method for accurate clonality determination
- Day Zero Diagnostics demonstrates the utility of epiXact-ONT for ultra rapid HAI investigations paving the way for real-time HAI transmission detection and prospective outbreak warning system

References:
1. Rang, F. J., *et al. Genome Biol* **19**, 90 (2018).
2. Shenoy, E. S. *et al. Infect. Control Hosp. Epidemiol.* 1–8 (2020)
3. Bush, S. J. *et al. GigaScience* **9**, (2020).
4. Olson, N. D. *et al.. Frontiers in Genetics* **6**, 235 (2015).

Contact information:
Mohamad Sater, PhD mohamad@dayzerodiagnostics.com
epiXact Team epiXact@dayzerodiagnostics.com
URL https://www.dayzerodiagnostics.com/hai-services/

DAYZERO DIAGNOSTICS