Long Read Sequencing at Integrated Genomics Operation

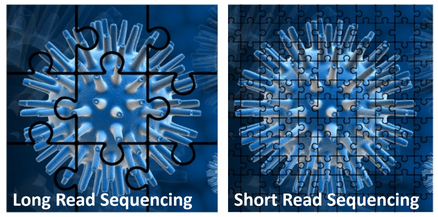
Cassidy Illum, Juan Li, Stephanie Chrysanthou, Neeman Mohibullah

1. INTRODUCTION

As knowledge of the relationship between the human genome and cancer continues to expand, the need for more robust, cost-efficient and cutting-edge technologies increases. Here, we present the interworking's of long read sequencing technology and IGO's experience using it.

Why Long Read Sequencing?

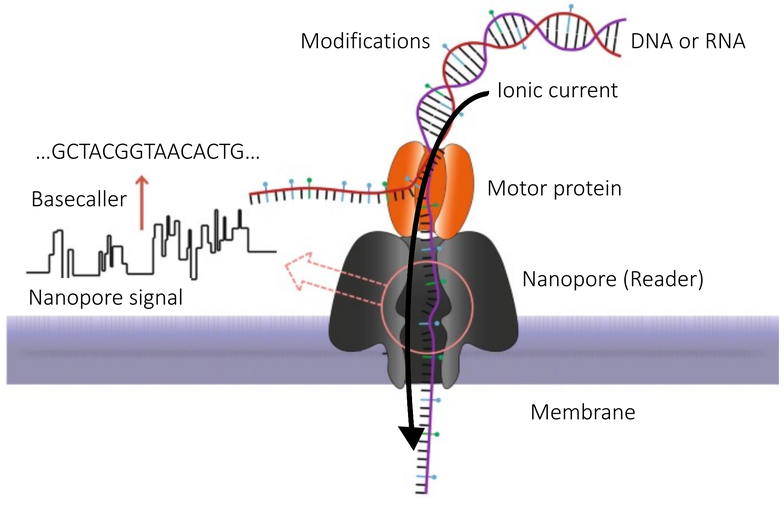
- Oxford Nanopore Technologies (ONT) sequencing is limited only by the length of the DNA/RNA fragment presented to the pore and can therefore span **entire repetitive regions**, resolve **structural variants**, differentiate between different **isoforms,** facilitate **de novo assembly** and generate content rich data, including **methylation**.
- Characterize and quantify full-length transcripts — up to single-cell resolution.
- Explore DNA and RNA modifications and eliminate any PCR bias with direct sequencing of native DNA/RNA.
- Stop sequencing when sufficient depth is obtained with real-time basecalling and analysis.



Try to complete two jigsaws of the same photograph – one with significantly larger pieces than the other. A jigsaw with only 9 pieces to assemble is much easier than one with 900.

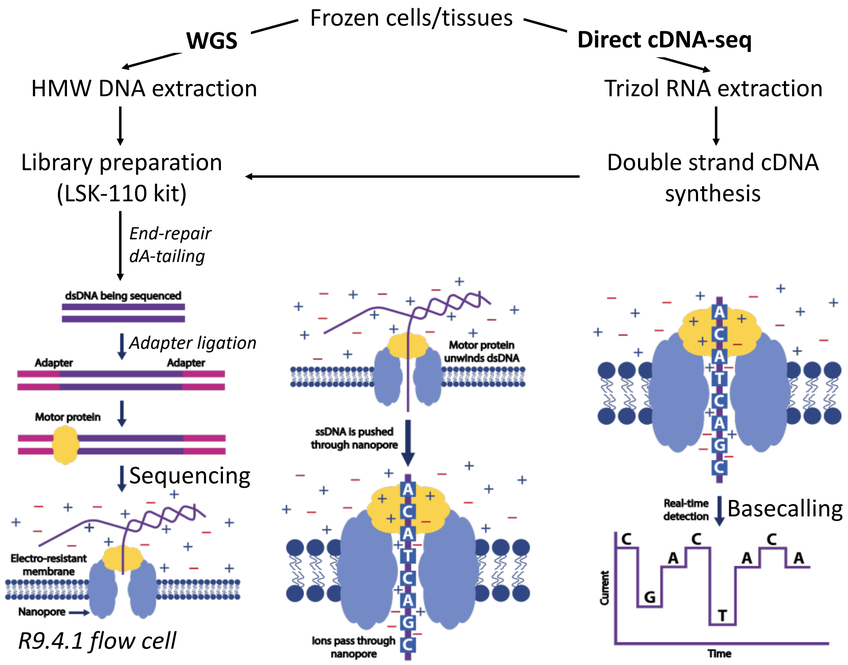
The Nanopore

All ONT sequencing devices use flow cells which contain an array of tiny holes — nanopores — embedded in an electro-resistant membrane. The nanopores are capable of measuring changes in electric current as DNA or RNA passes through them. The library has a motor protein adapter which unzips the double stranded DNA or cDNA, sending a single strand through the nanopore. Each base (modified or unmodified) has a unique signal disruption which is then translated into a base call in real-time.



2. METHODS

The Workflow *(most popular)*



The Instruments

	Flongle (adaptor on MinION)	MinION Mk1C	PromethION 24
Channels per flow cell	126	512	3000
Maximum number of flow cells per device run	1	1	24
Run time	Up to 20 hours	Up to 48 hours	Up to 96 hours
DNA sequencing yield per flow cell *! Yields are dependent on sample and preparation methods.*	0.5 - 2 Gb	5 - 25 Gb	50 - 150 Gb
Suitable applications	• Amplicons • Panels/targeted sequencing • Quality testing • Small sequencing tests	• Small genomes/ exomes • Metagenomics • Targeted sequencing • Whole transcriptomes (direct RNA and cDNA)	• Large genomes • Whole transcriptomes • Multiplexing • Multiomics (simultaneous DNA and RNA sequencing)

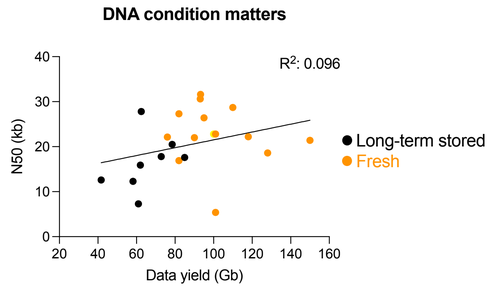
MinION (+ Flongle adapter)

PromethION

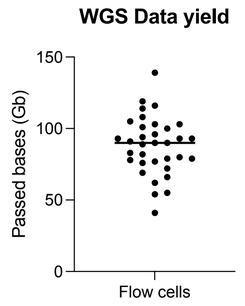
3. RESULTS – PromethION

Protocol	Libraries
WGS	51
WGS-Multiplex	20
Direct cDNA-seq	3
Direct cDNA-seq-Multiplex	6
10x + cDNA-seq	3
Direct RNA-seq	1

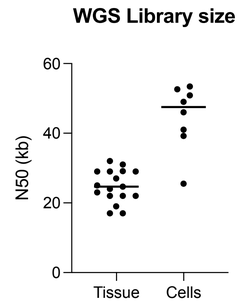
Starting November 2021, IGO has sequenced over 84 samples with 6 different ONT kits.



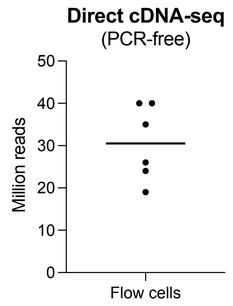
Tests included DNA samples stored in -80°C for >1 year and fresh DNA. Data shows that freshly extracted DNA of high purity yields higher throughput data.



Whole genome sequencing runs yielded on average 90 Gb of data per flow cell, equivalent to **~30X** coverage of the human genome.



The average library size (N50) in kilobases (kb) for WGS with freshly extracted HMW DNA from tissue and cells.



Direct cDNA sequencing yields about **30 million reads** per flow cell.

4. IMPACT

Long-read sequencing solves genome assembly difficulties and improves completeness of genome assemblies compared to traditional NGS systems. This new technology provides new abilities to explore a diverse variety of experimental design options, particularly in cancer research where structural variation is important.

Advancing Technology at IGO

- Upgrade to LSK-114 Ligation kit and R10.4.1 flow cells

This is the latest chemistry with tunable settings, allowing the user to choose between throughput or higher accuracy. In general, it delivers over 99% raw read accuracy and duplex accuracies ~99.9%.

- Adaptive Sampling

Software controlled enrichment of regions/genes of interest for increased sequencing depth.

- Methylation Detection

Remora, ONT's latest methylation detection software, is being integrated into the analysis platform to provide this information in parallel during the sequencing run. Epigenomic data included at no additional cost!