Long Read Sequencing Lecture and Lab

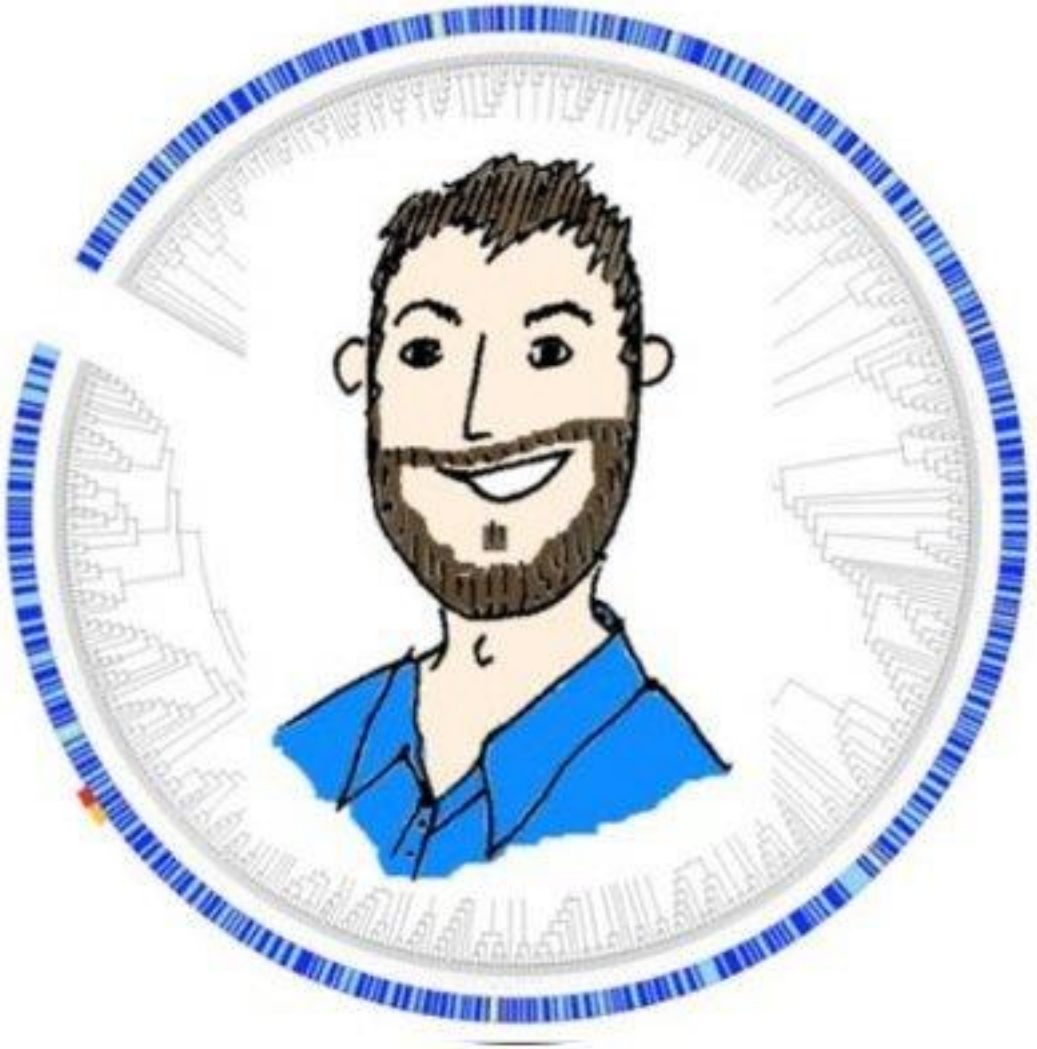
AGRO5431 – Feb 5 2024

Andy Read
USDA-ARS Wheat and Oat Geneticist
@hashtag_read
Andrew.read@usda.gov
office – 522 Borlaug Hall

Warm up question:

Group 1 name one sequencing technology, the next group can either share the read-length from that technology, or name a different sequencing technology

Introduction

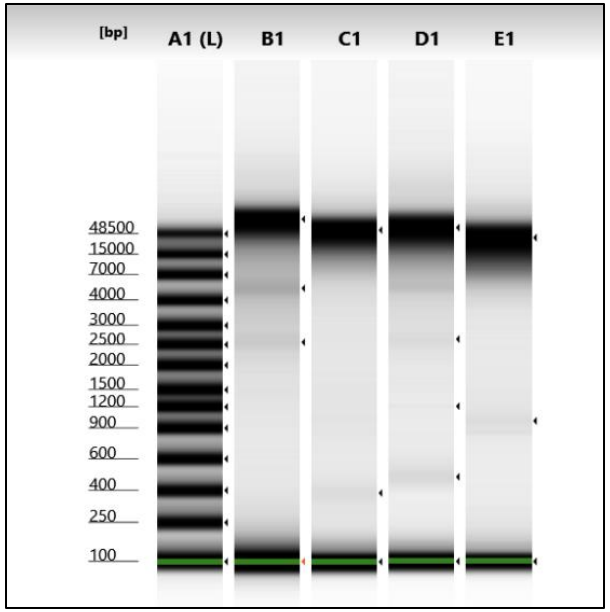


Goals for today

1. Become familiar with **DNA requirements, library prep options, and flowcell loading**
2. Be able to **troubleshoot** Nanopore (a little bit)
3. **Basecall** raw Nanopore data
4. **Align** reads to a reference genome
5. Think about where **un-mappable reads** are coming from

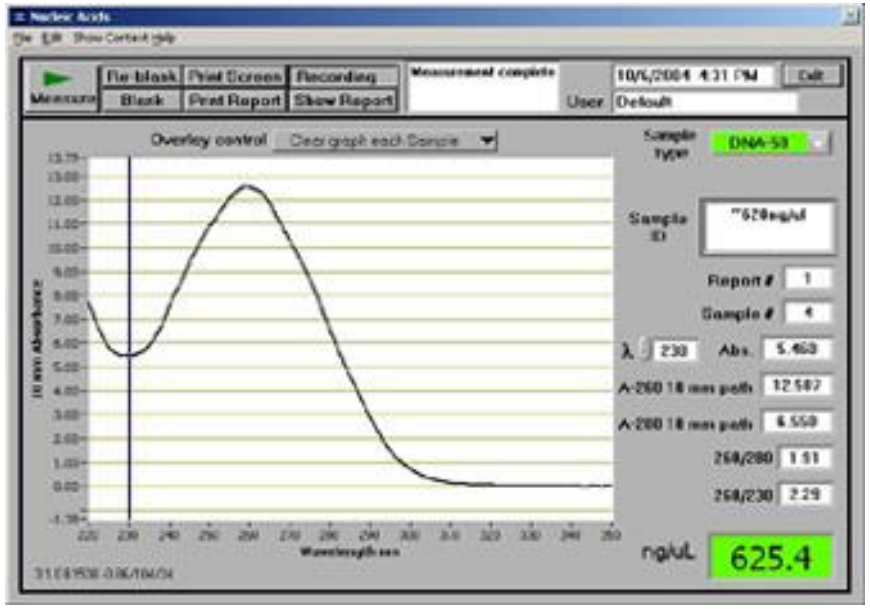
It starts with purified nucleic acid (DNA or RNA)

How can you tell if you've isolated high molecular weight DNA?



Agarose gel
Size selection
Tape Station/Bioanalyzer

Is my DNA pure enough?

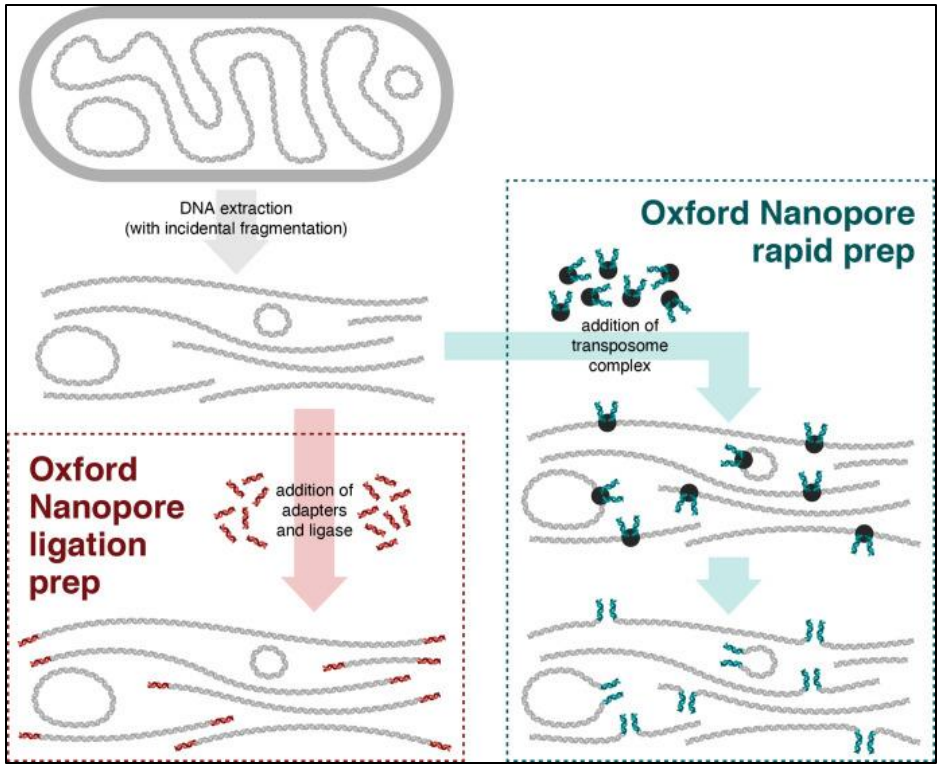


What are optimal 260/280 and 260/230 ratios?
What do low ratios indicate?
Can I trust the Nanodrop quantification?

You have good DNA – time to order some sequencing supplies

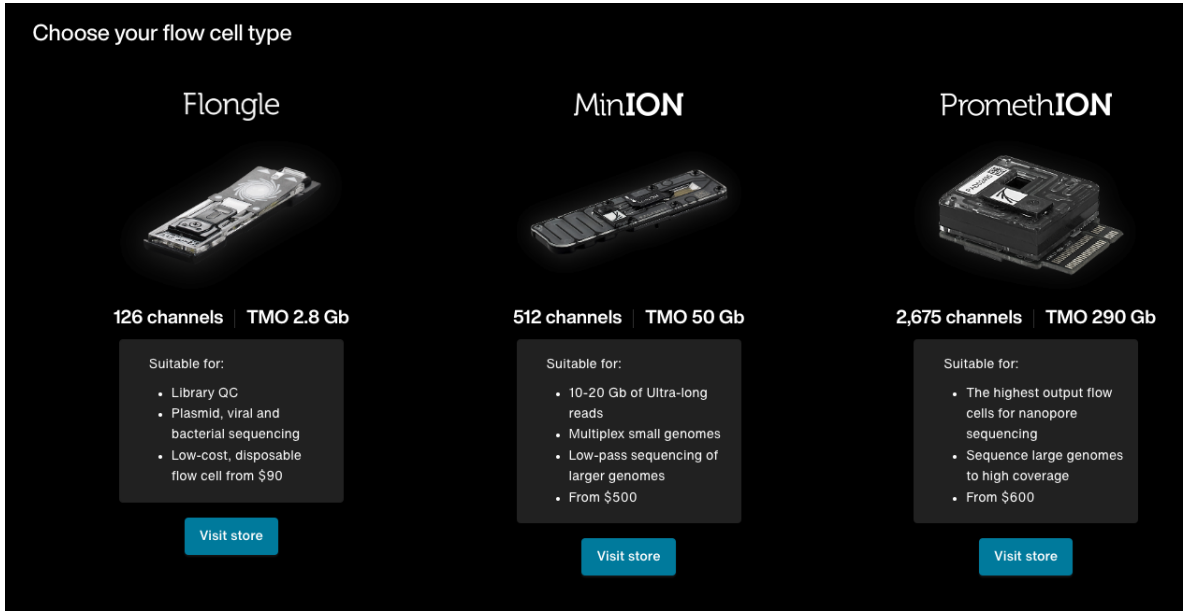
https://nanoporetech.com/products/kits

Library prep kit



Microb Genom. 2021 Aug;7(8):000631.
doi:10.1099/mgen.0.000631.

Flowcell/Hardware



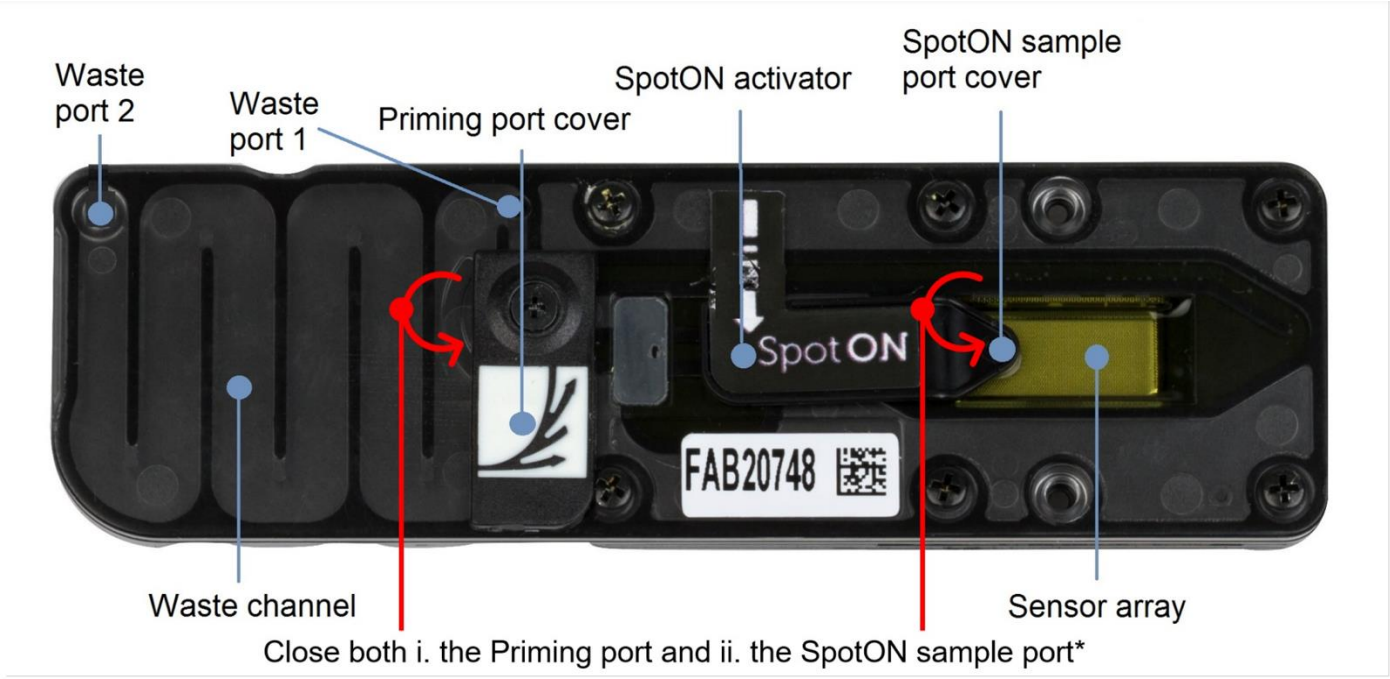
MinION | MinION Mk1C | MinION Mk1D | GridION | P2 Solo | P2 | PromethION 24 | PromethION 48

MinION and Flongle Flow Cell compatible

PromethION Flow Cell compatible

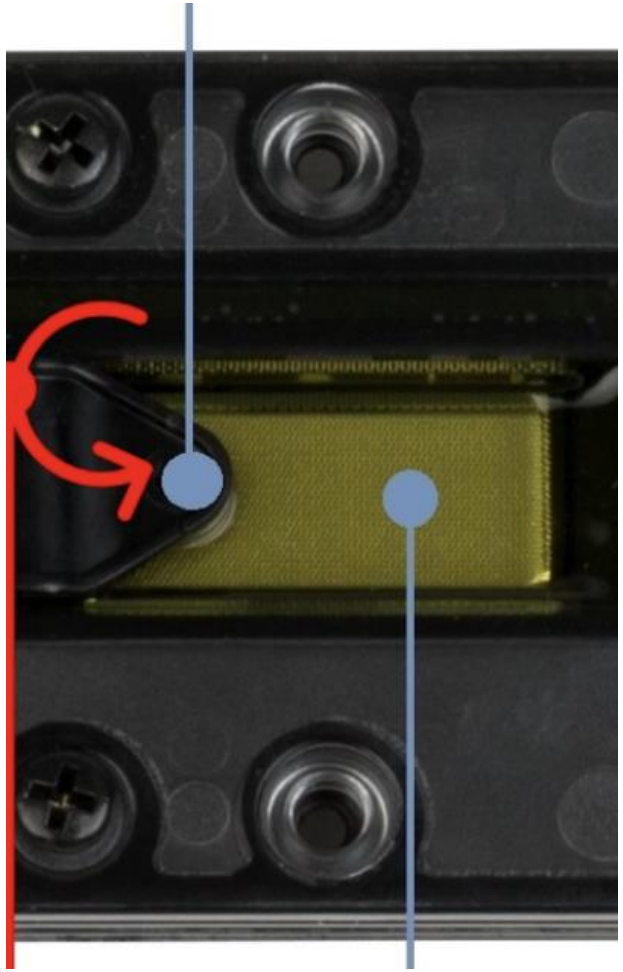
Configuration	Platform			Techniques	Tech specifications			
Number of flow cells per device	1	1	1	5	2	2	24	48
Maximum number of channels per flow cell	512	512	512	512	2,675	2,675	2,675	2,675
Run time	72 Hours	72 Hours	72 Hours	72 Hours	72 Hours	72 Hours	72 Hours	72 Hours
Device TMO†	50 Gb	50 Gb	50 Gb	250 Gb	580 Gb	580 Gb	~7 Tb	~14 Tb
Maximum number of flow cells per year*	104	104	104	520	208	208	2,596	4,992
Offer sequencing as a service	No	No	No	Yes	Yes	Yes	Yes	Yes

Let’s take a break and check out a flow cell

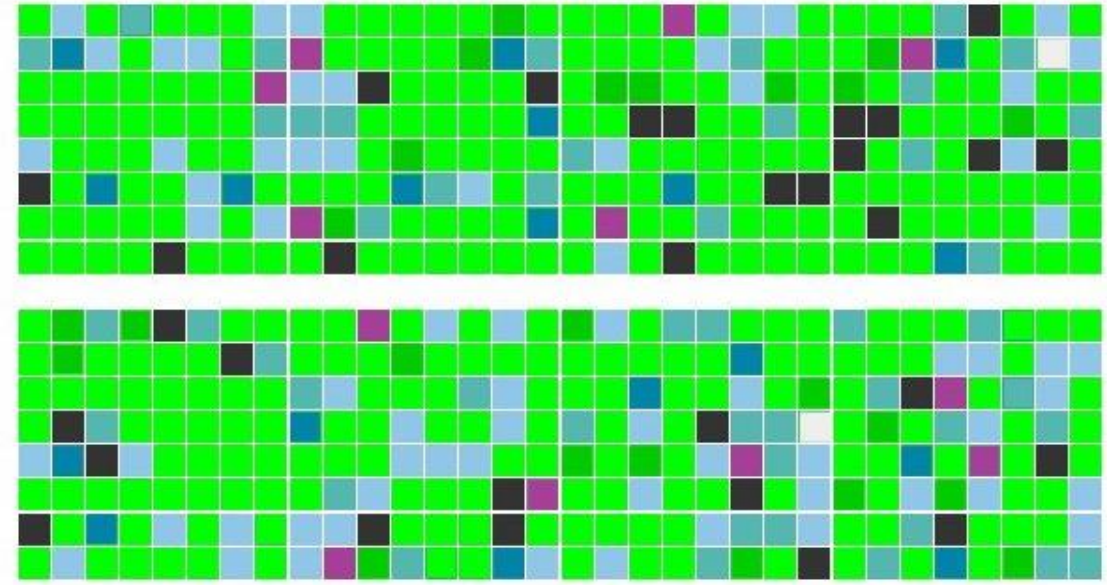


*Both ports are shown in a closed position

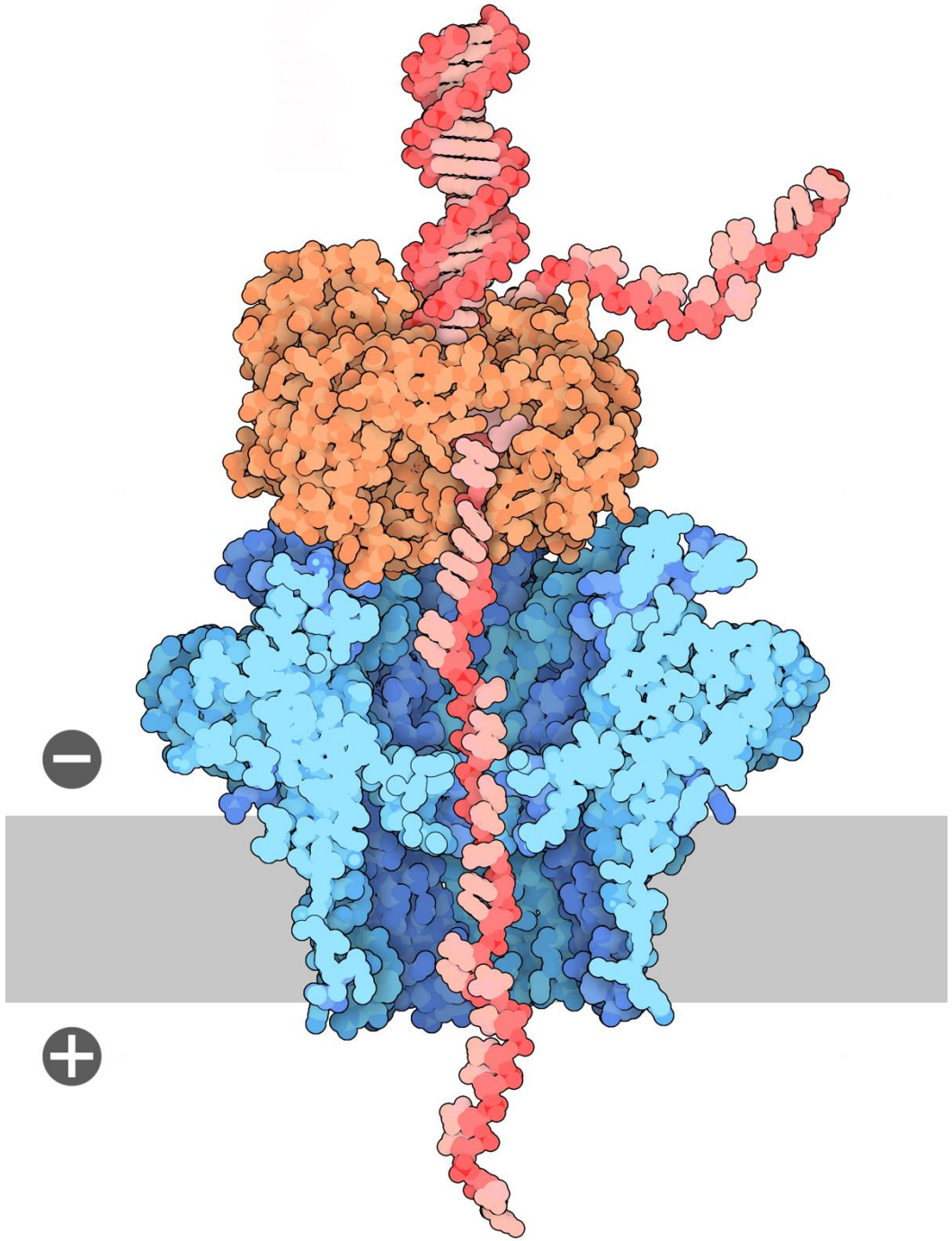
What's really going on in there?



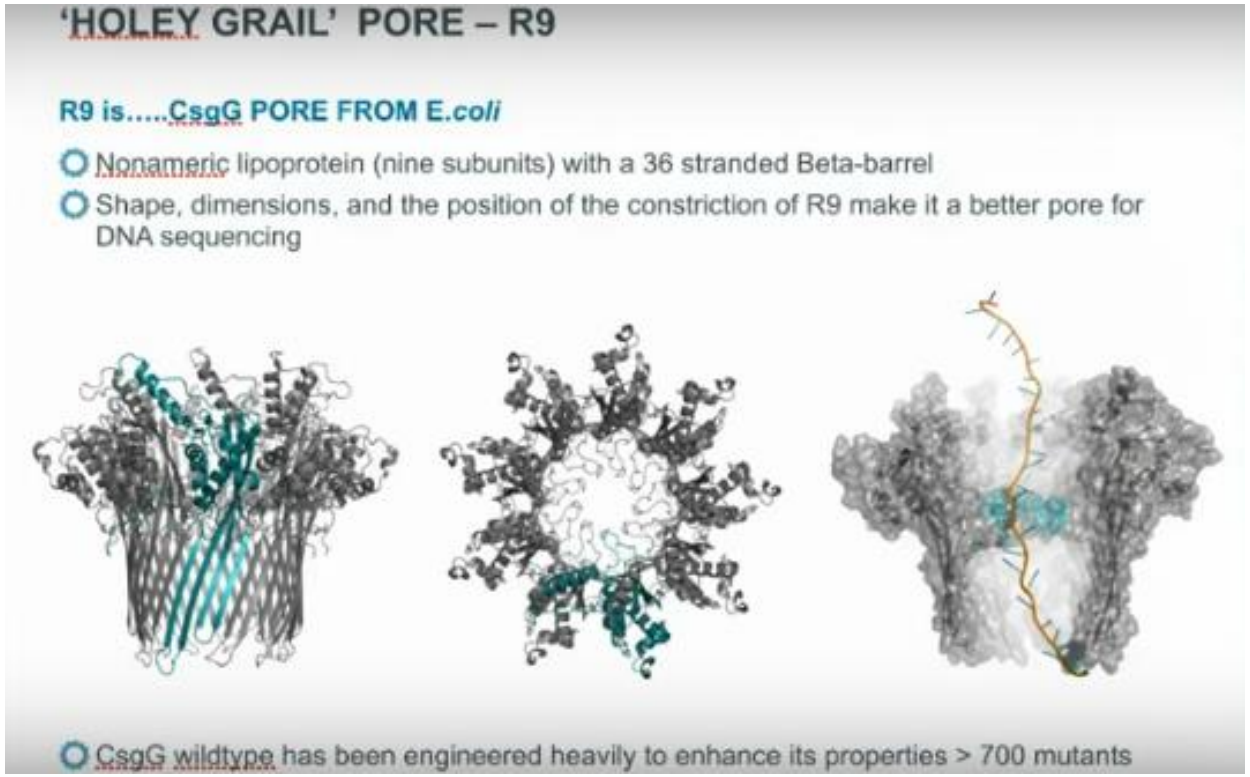
- 512 channels – up to 4 nanopores at each channel
- 2,048 is the theoretical maximum number of pores, but only 1 pore per channel can be sequencing at a time



Behold, the nanopore…



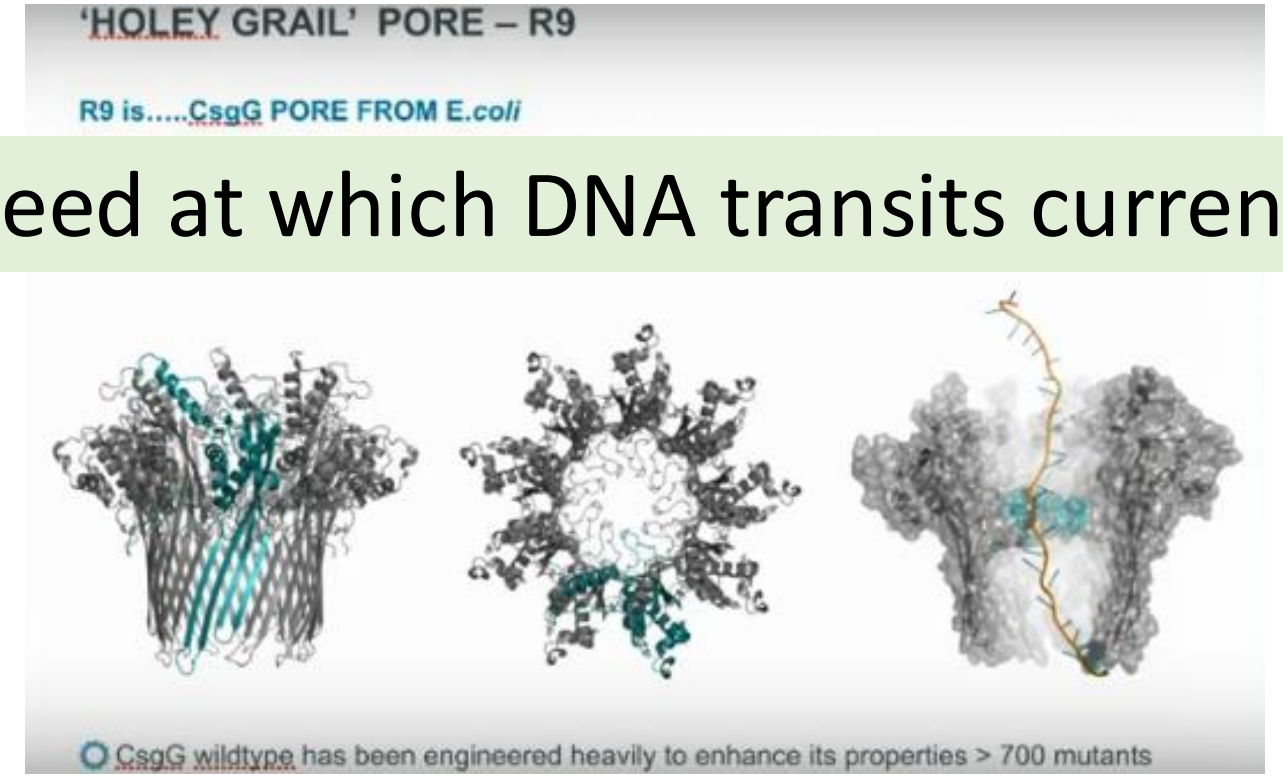
A view of the pore in 2016



Furthermore, fast mode is now going to a reality: with the new kits (going to developers this month and the whole community next month, with R7 phased out shortly thereafter), speeds of 280 bases per second (or about 3-fold higher than the current 80bps). The system is believed to be capable of even faster; Oxford plans to roll out software updates to up the speed to closer to 500 bps (perhaps in steps) as they gain confidence in the new system.

omicsomics.blogspot.com/2016/03/oxfords-no-thanks-ive-already-got-one.html

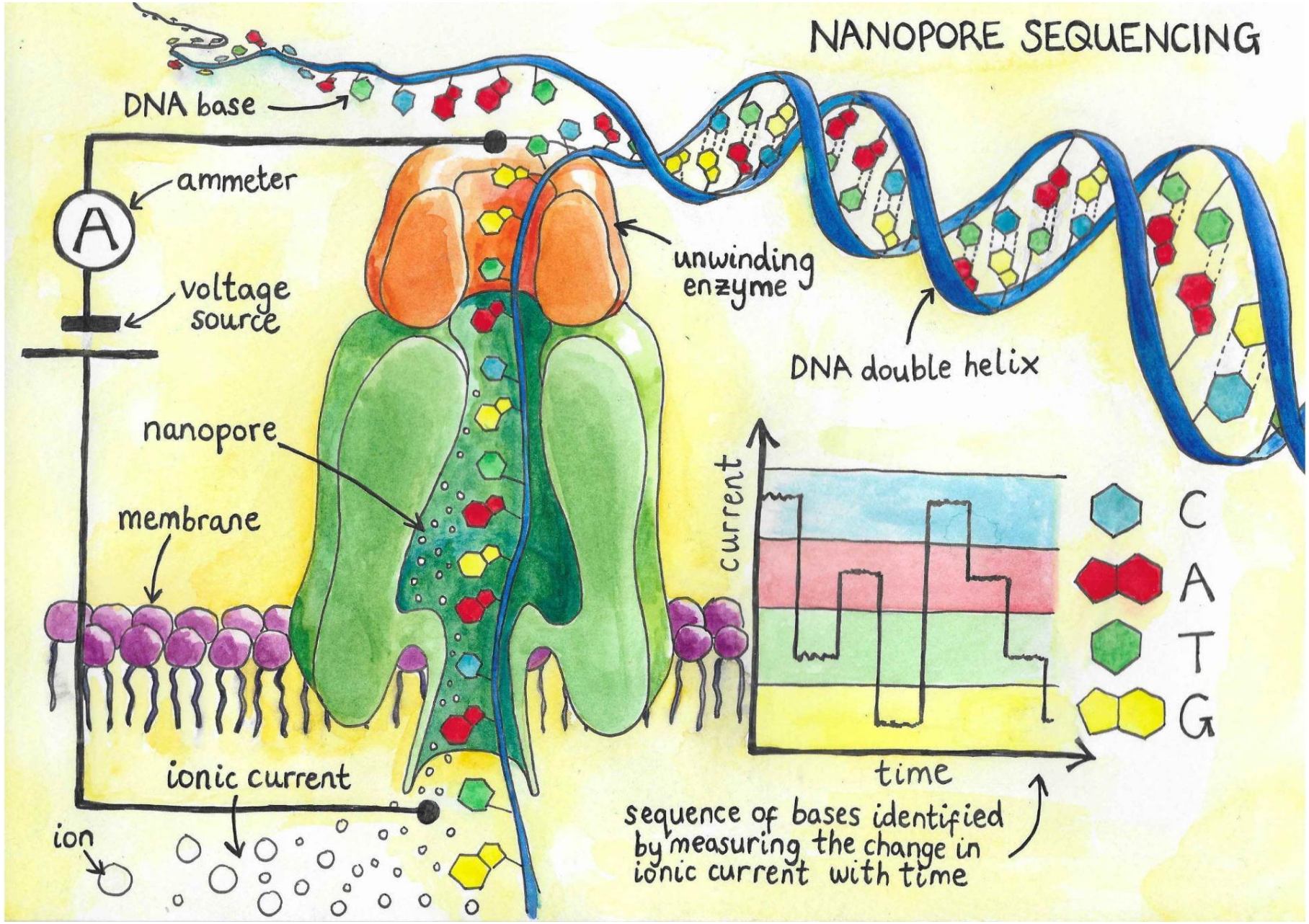
A view of the pore in 2016

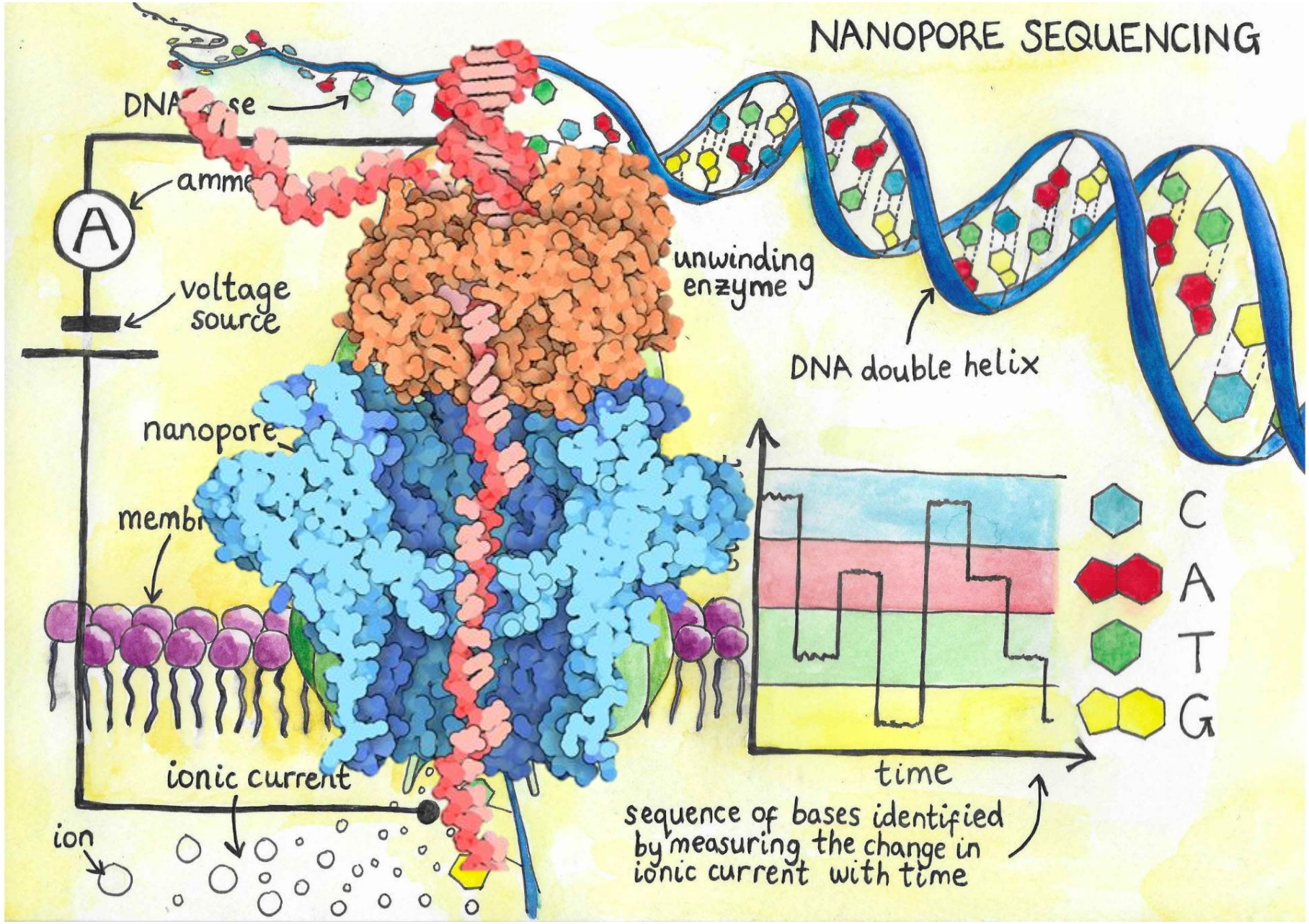


What is the speed at which DNA transits current pore-types?

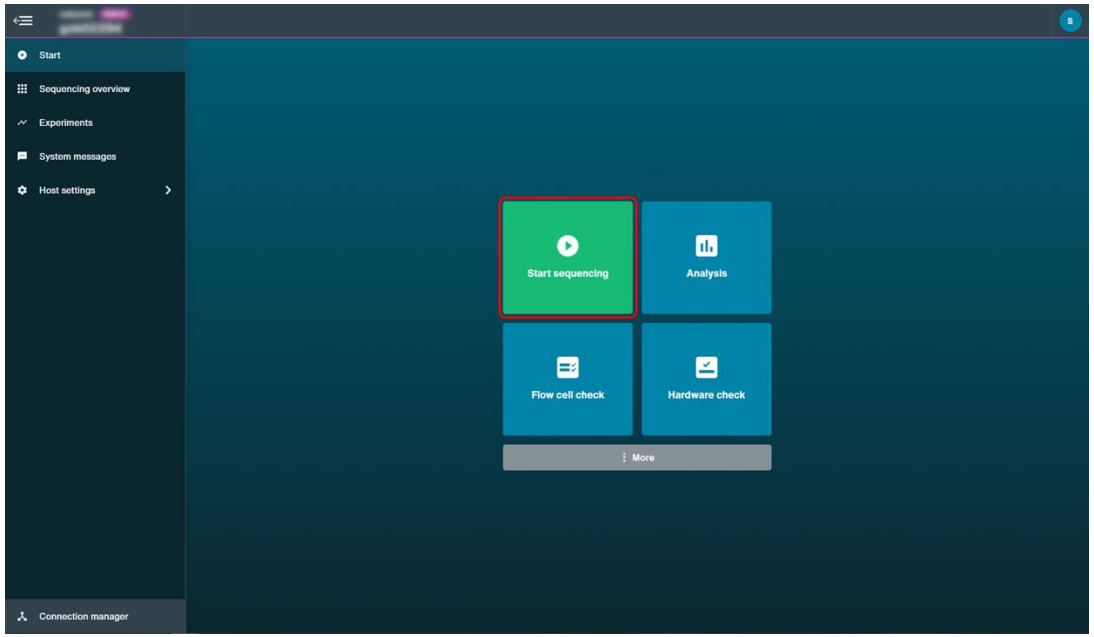
Furthermore, fast mode is now going to a reality: with the new kits (going to developers this month and the whole community next month, with R7 phased out shortly thereafter), speeds of 280 bases per second (or about 3-fold higher than the current 80bps). The system is believed to be capable of even faster; Oxford plans to roll out software updates to up the speed to closer to 500 bps (perhaps in steps) as they gain confidence in the new system.

omicsomics.blogspot.com/2016/03/oxfords-no-thanks-ive-already-got-one.html

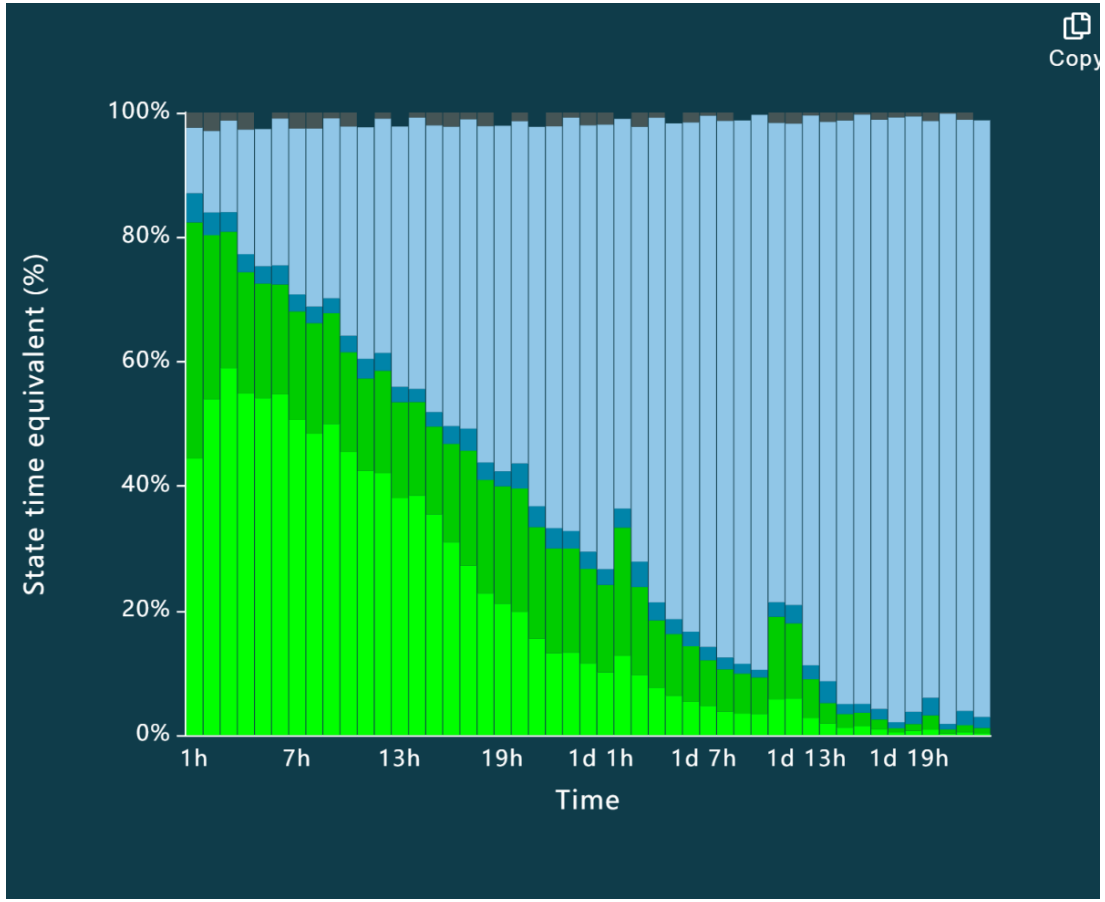

NANOPORE SEQUENCING
DNA base
ammeter
A
voltage source
unwinding enzyme
DNA double helix
nanopore
membrane
current
time
C
A
T
G
ionic current
ion
sequence of bases identified by measuring the change in ionic current with time


NANOPORE SEQUENCING
DNA
amme
A
voltage source
unwinding enzyme
DNA double helix
nanopore
membr
ionic current
ion
time
C
A
T
G
sequence of bases identified by measuring the change in ionic current with time

Starting a Nanopore run



Monitoring flowcell progress

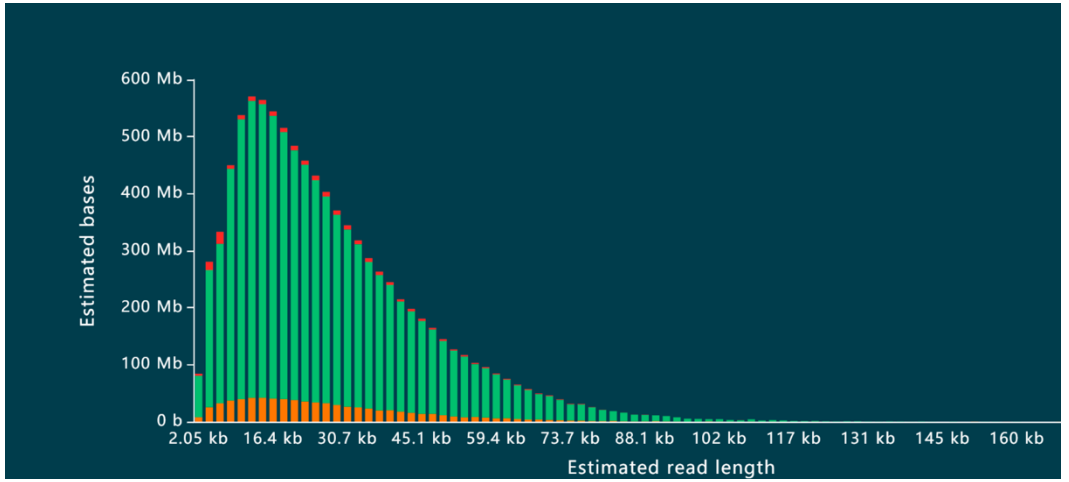


Active pore count reduces over time

What might cause this?

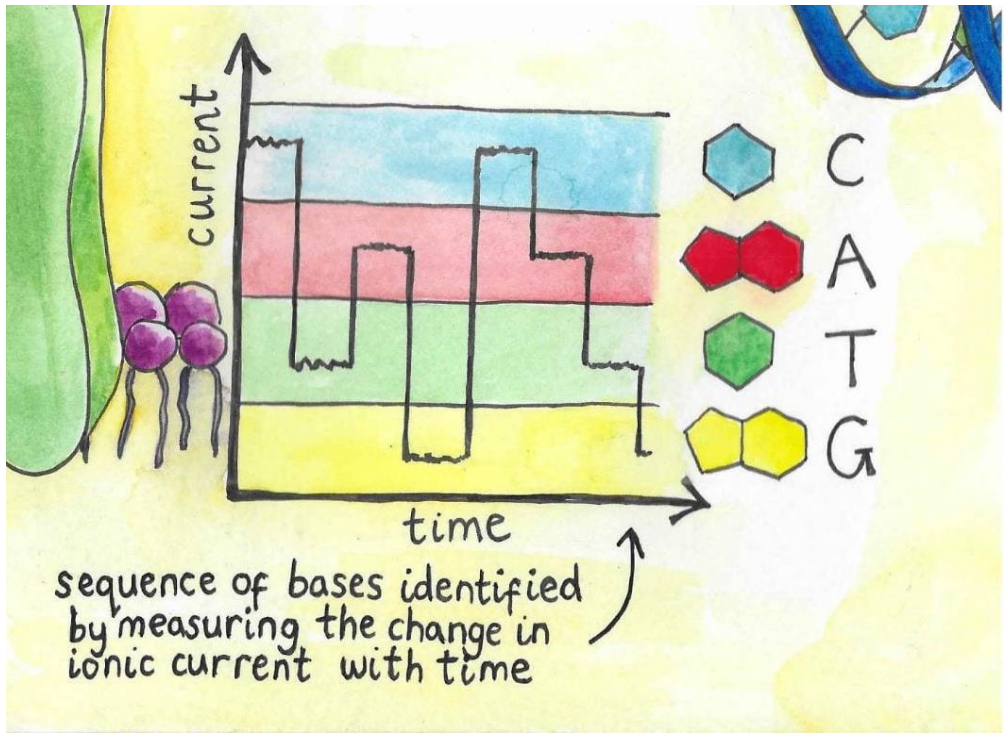
Why are there some increases in active pore count? (at 1d 13h for example)

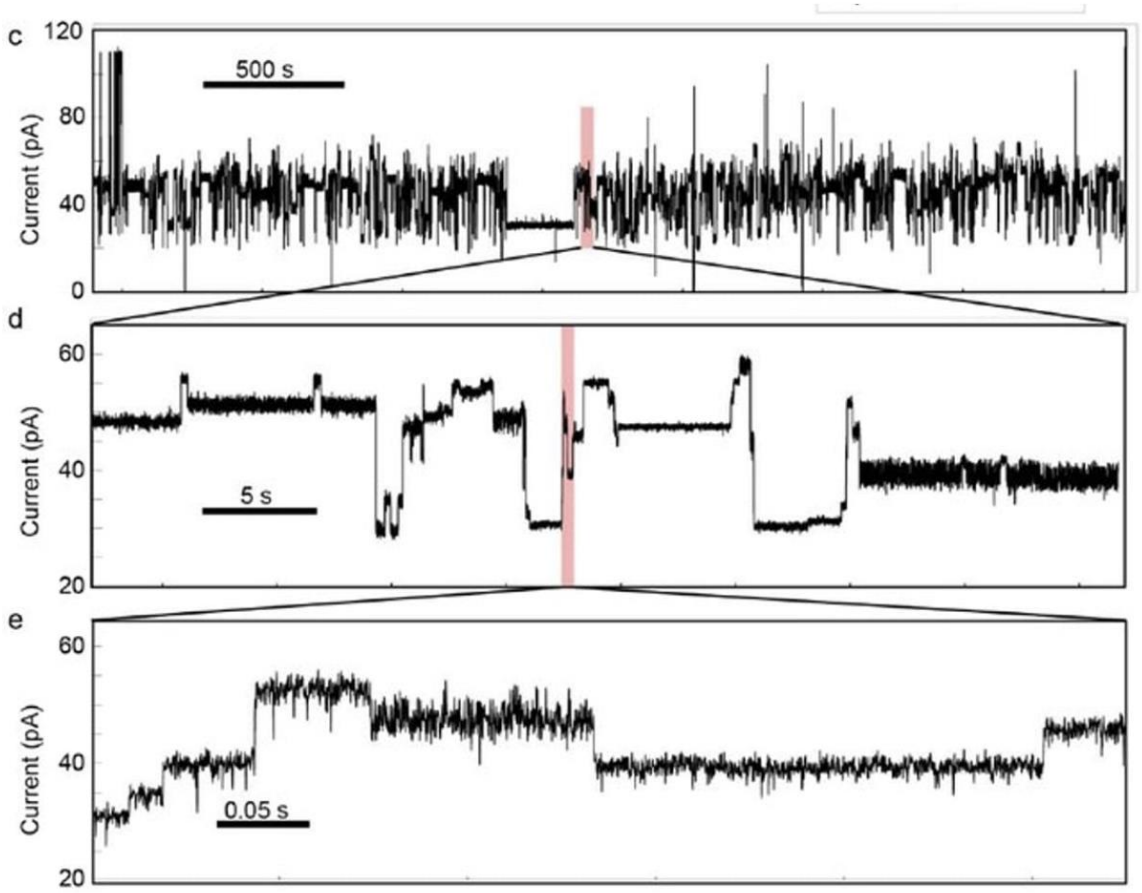
Read length distribution



Note: **Estimated** read length

What's with the "squiggle"?





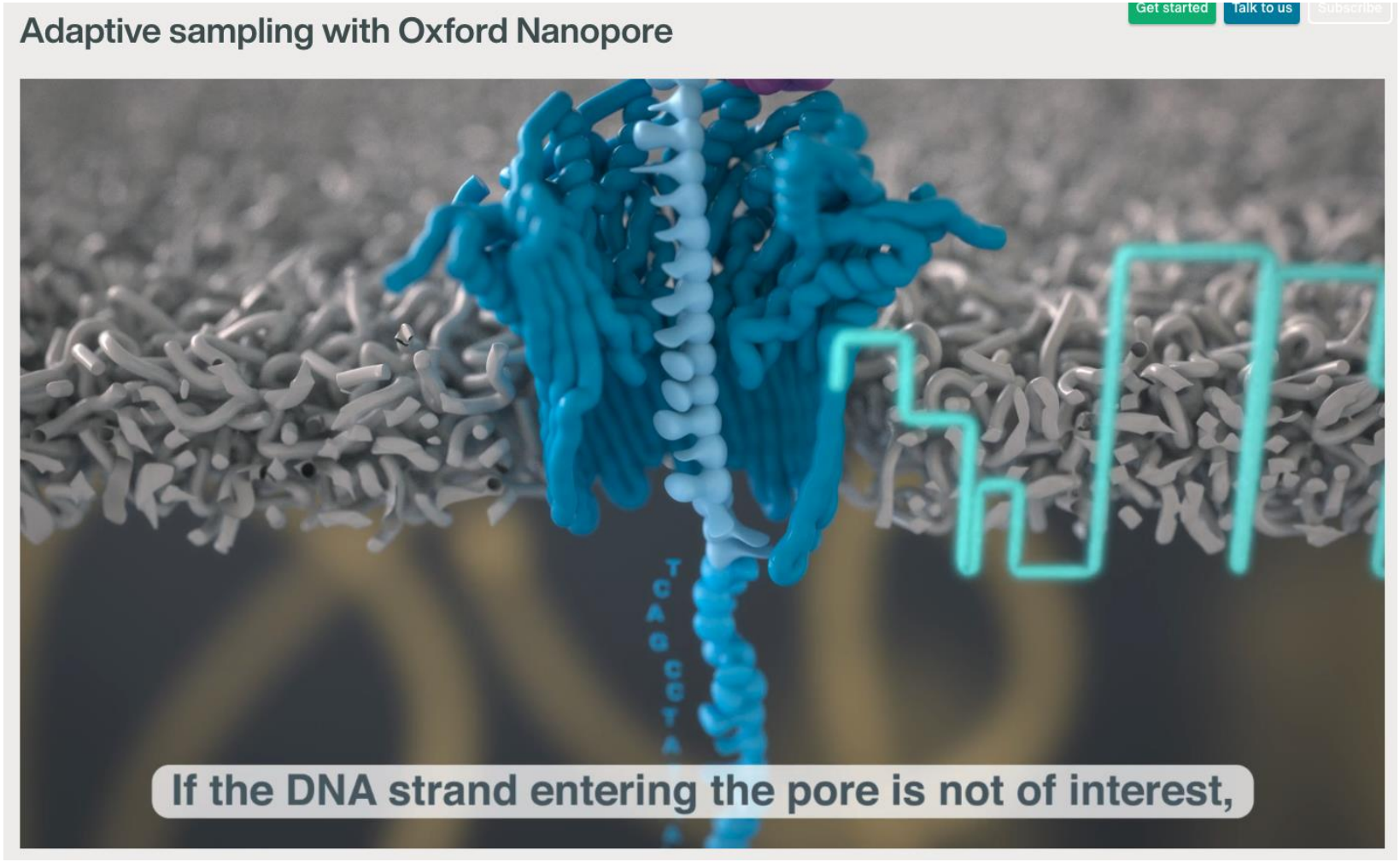
Converting the squiggle to accurate base-calls is challenging!
This is an area where Nanopore has made huge improvements (more data = better algorithms)
The raw data is stored in .fast5 file format

Let’s take a shot at base-calling some raw data!

https://github.umn.edu/agro5431-2023/Lab3_NanoporeLongReads

Nanopore is always innovating

Adaptive Sequencing – Enrich for sequences you're interested in

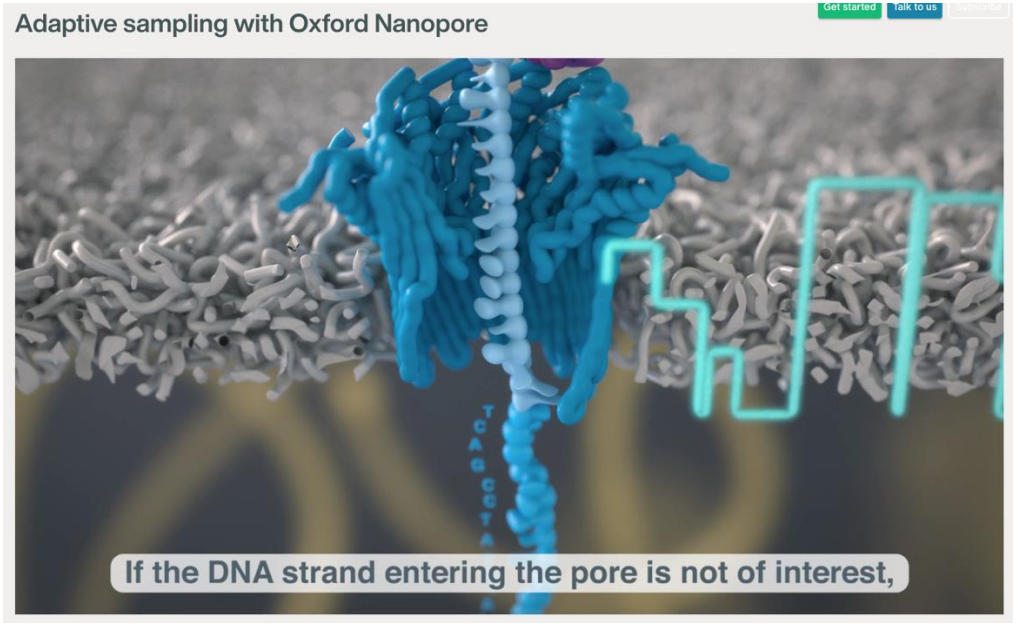


https://nanoporetech.com/resource-centre/adaptive-sampling-oxford-nanopore

Adaptive Sequencing – Enrich for sequences you're interested in

What is actually happening?

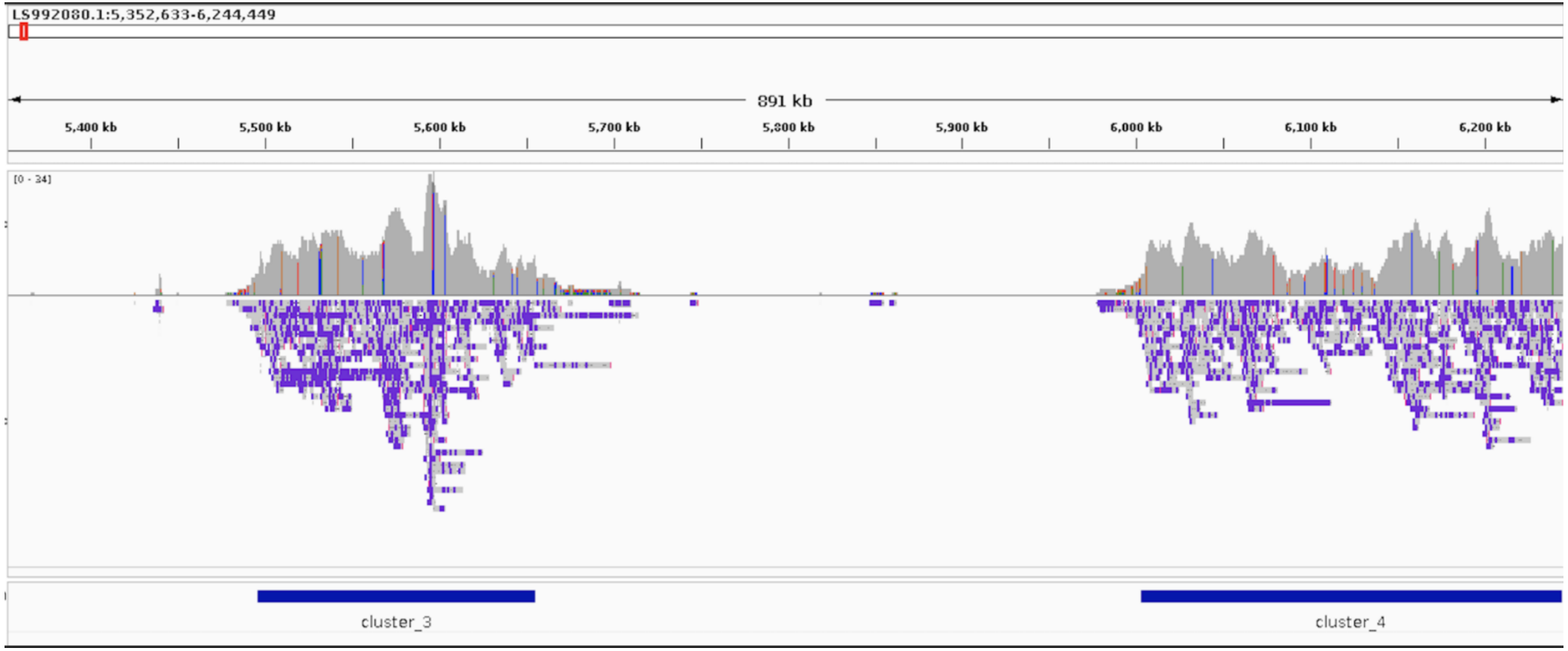
1. ssDNA passes through the pore
2. basecalling occurs at the same time – this requires a GPU
3. basecalled read is mapped to reference genome
4. if it maps to a region of interest, it continues, if not the current is reversed, kicking out the read



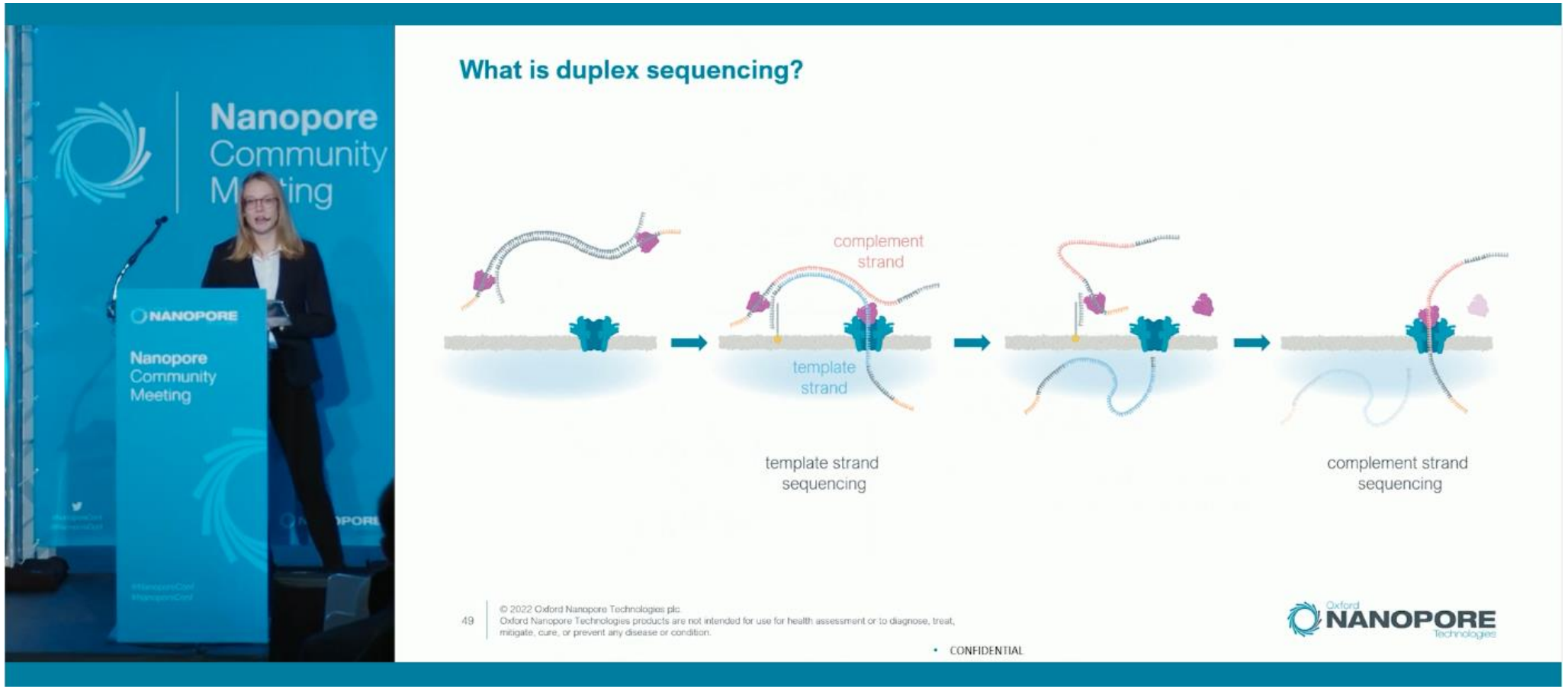
When would this be useful?

Adaptive Sequencing – Enrich for sequences you're interested in

Some data we generated from the VERY LARGE wheat genome

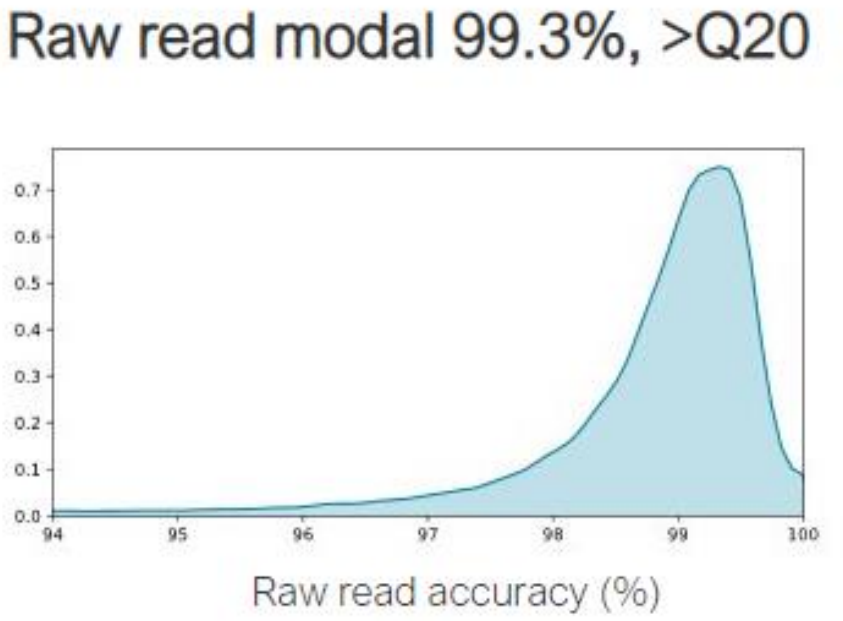


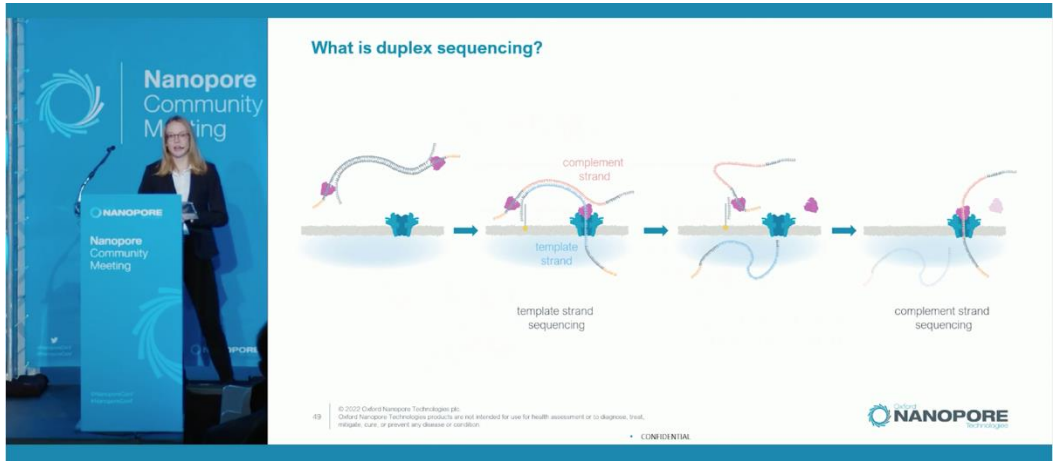
Improving accuracy with duplex reads



https://nanoporetech.com/resource-centre/video/ncm22/advances-in-duplex-basecalling

Improving accuracy with duplex reads

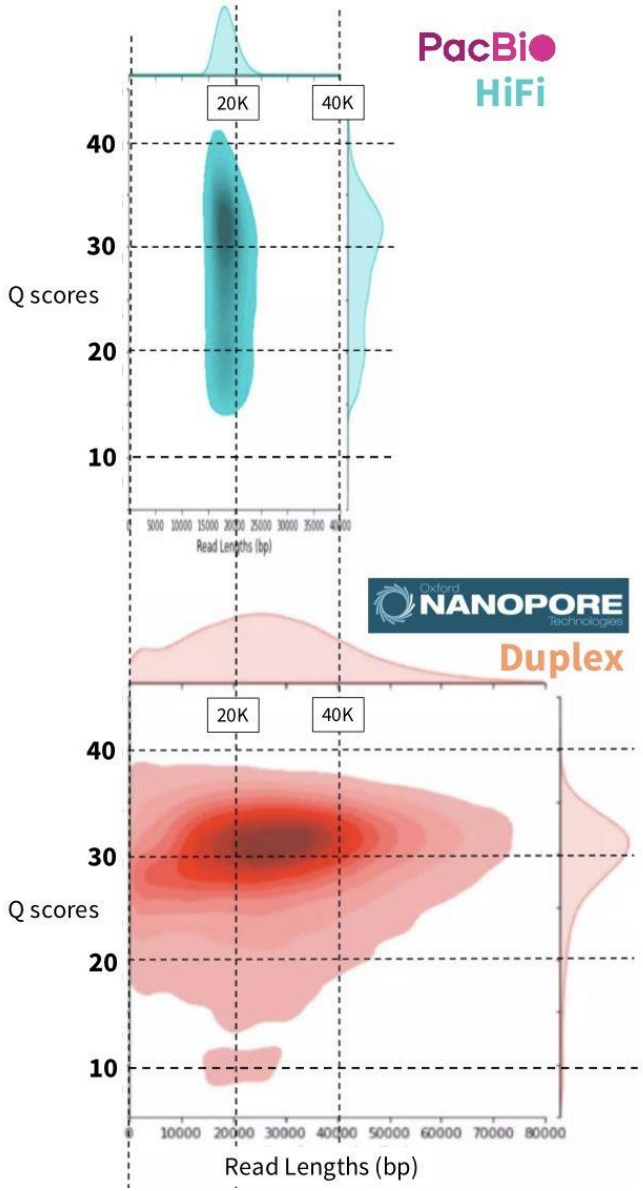




Using both strands to basecall improves accuracy,
But not all reads will be duplex in a given run

https://nanoporetech.com/resource-centre/video/ncm22/advances-in-duplex-basecalling

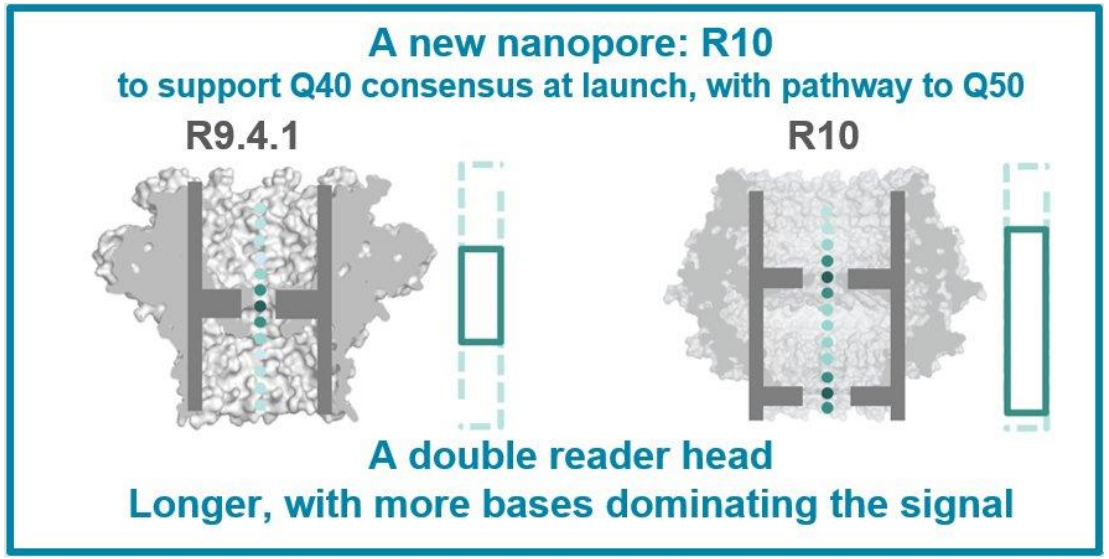
Improving accuracy with duplex reads:
The sequencing wars…

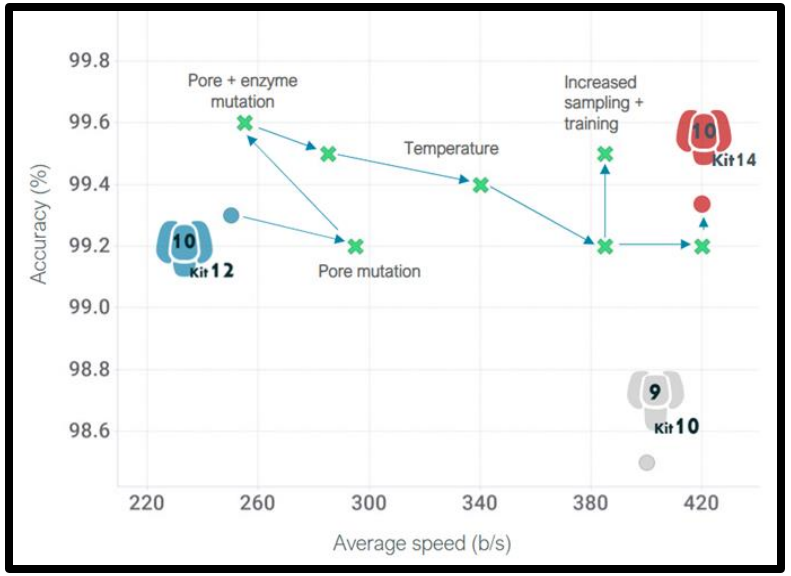


NOTE: this is from Twitter, not a peer-reviewed publication

What is this plot communicating?

R10 has entered the chat





The R10 flowcell is the new standard
Changes to the pore topography allow better base-calling
The user can choose high-accuracy or high throughput modes

How to keep up with the latest Nanopore news?

Clive Brown – CTO - @the-taybor.bsky.social

Alexander Wittenberg – KeyGene - @awngs.bsky.social

Official account - @nanoporetech.com

Sissel Juul – Genomic Applications - @sisseljuul.bsky.social

Thank you all so much!
Who has questions?

Homework on Canvas

read0094@umn.edu