

Nanopore Sequencing Workshop

Data Science Core

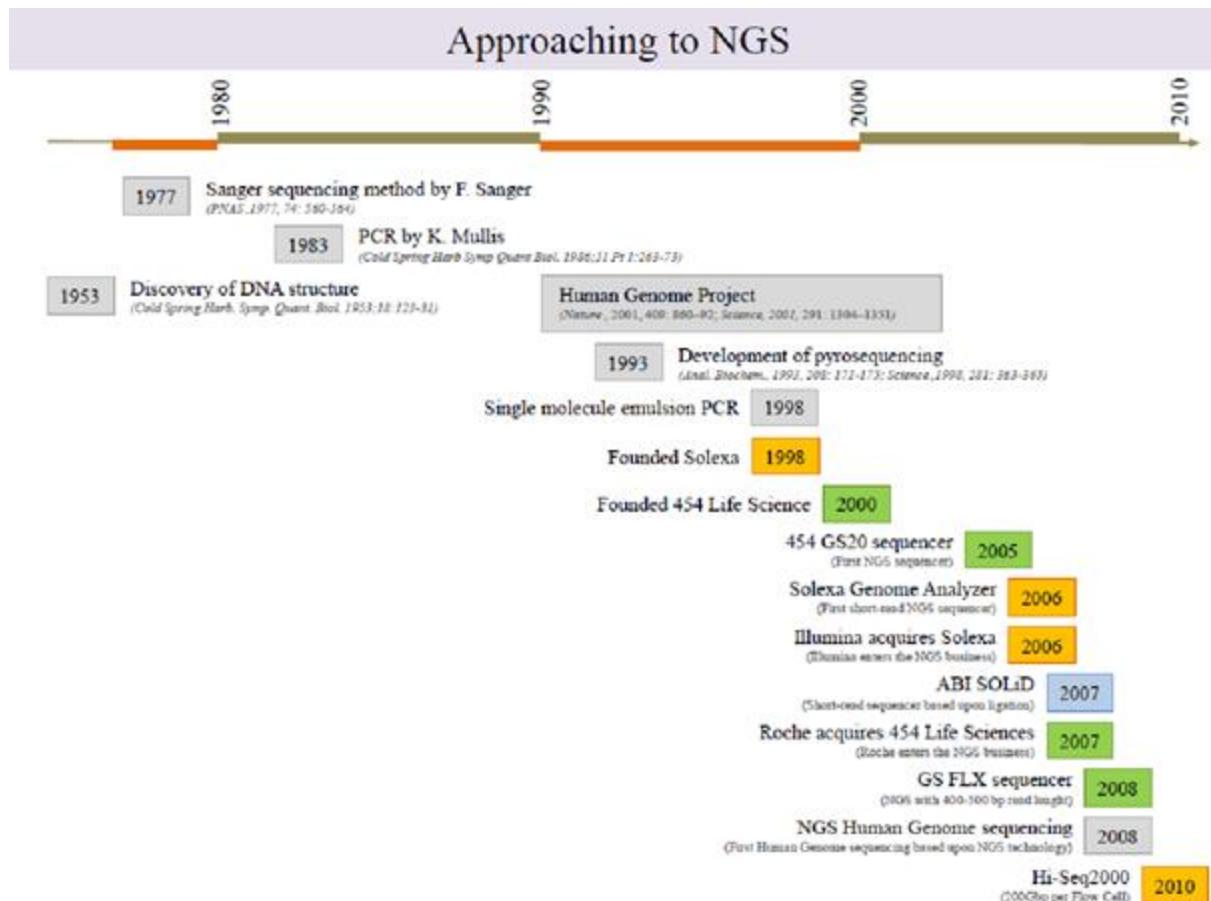
Alaska INBRE NIH IDeA (P20GM103395)





Introduction to Nanopore Sequencing

History of DNA sequencing

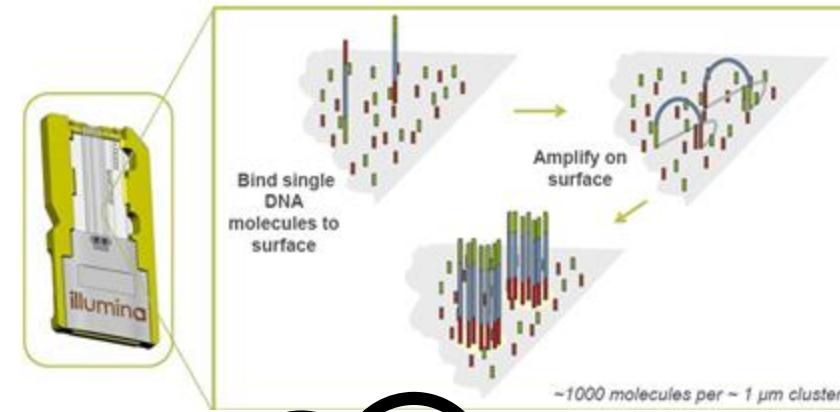


Or for the video check out:
<https://www.youtube.com/watch?v=s9UbA7VylSQ>

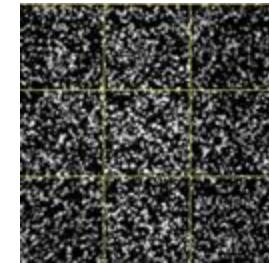
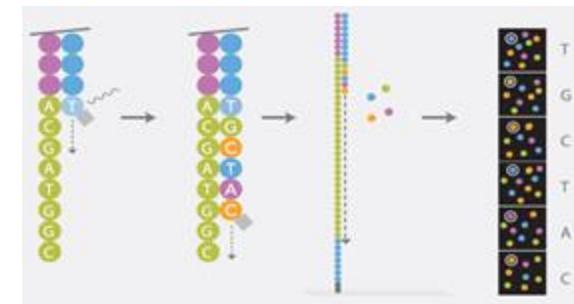
From Slideshare presentation of Cosentino Cristian
<http://www.slideshare.net/cosentia/high-throughput-sequencing>



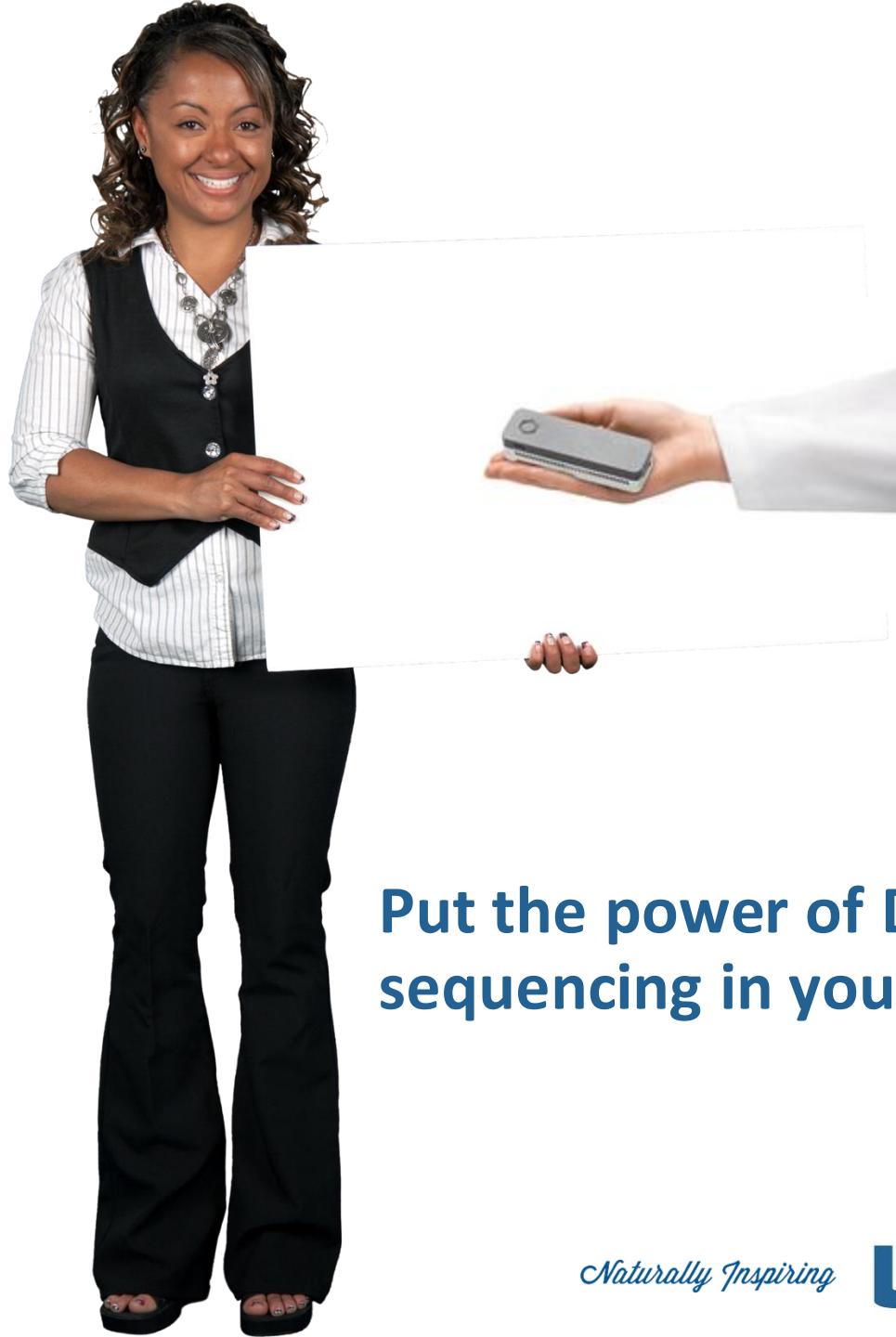
MiSeq at UAF IAB Genomics Core



Ask me for more information



Naturally Inspiring



**Put the power of DNA
sequencing in your hand**

Naturally Inspiring

UAF UNIVERSITY OF
ALASKA FAIRBANKS

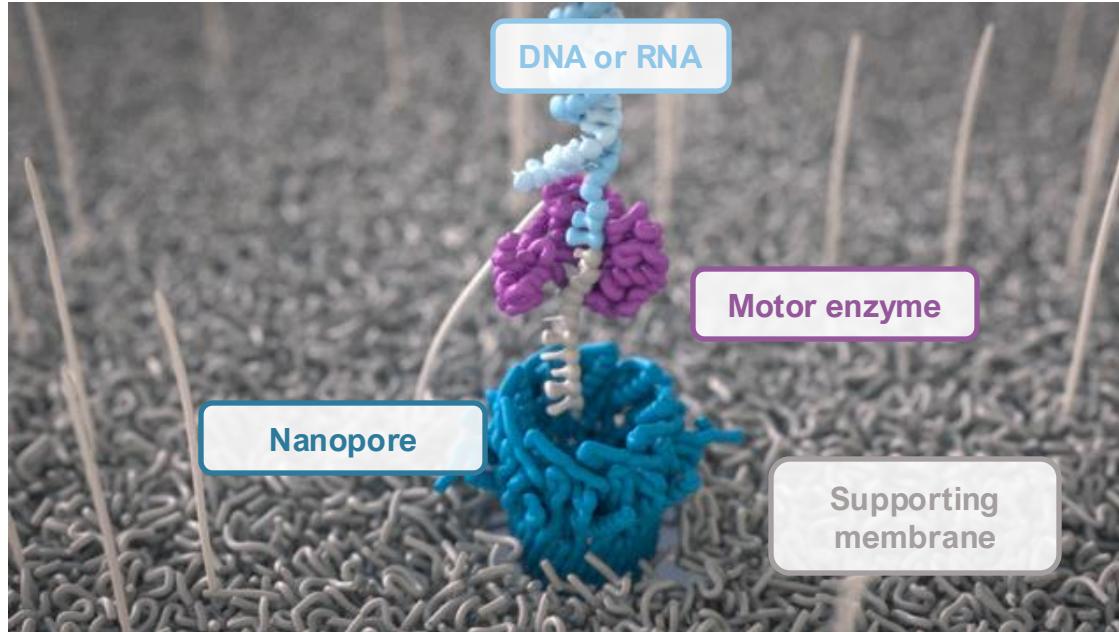


Agenda

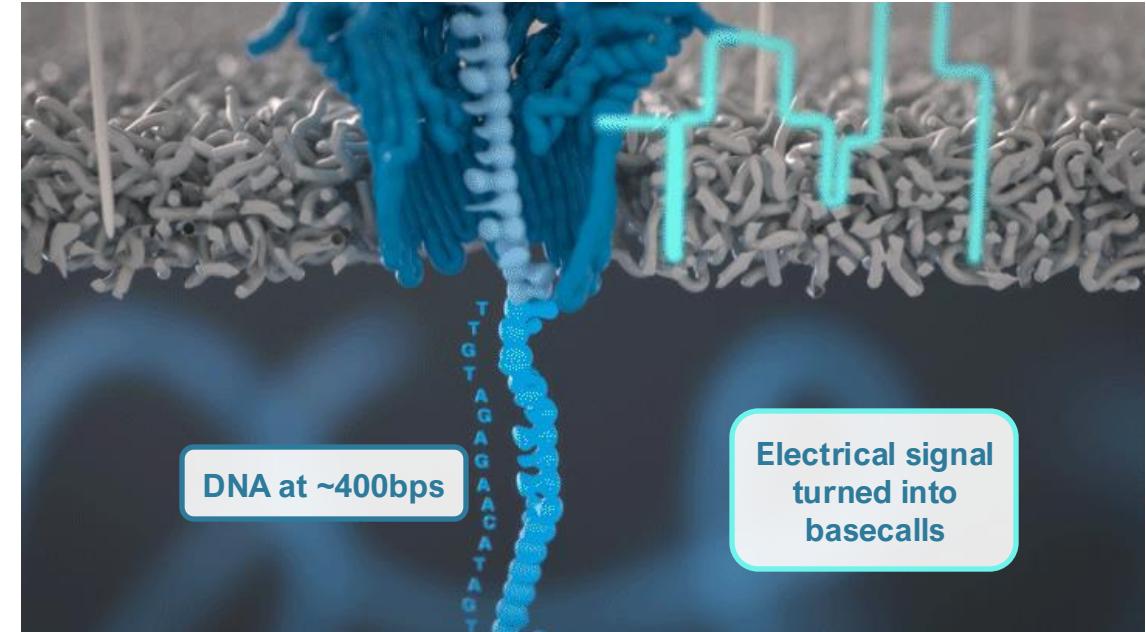
- Overview of Nanopore Sequencing
- Basecalling – from signal to sequence
- Library Preparation
- Flow cell – hands on
- More with Nanopores!

How nanopore sequencing works

A DNA / RNA strand is passed through a nanopore



An electrical signal is interpreted into sequence data



The advantages of nanopore sequencing

Real-time
Analysis

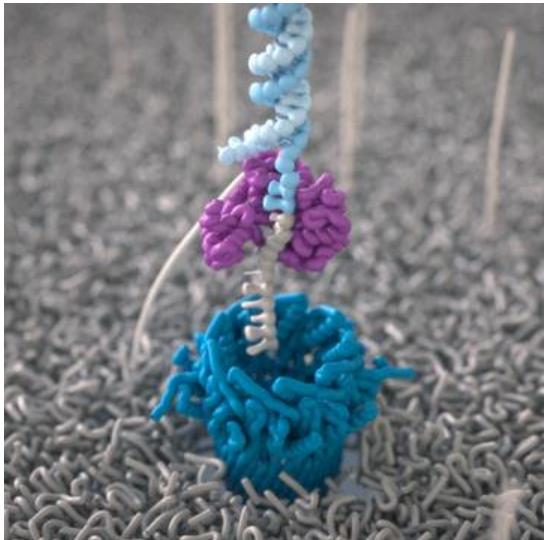
PCR free, no
amplification bias

Modified base
detection

Read length-
agnostic

Direct sequencing
of DNA / RNA

How nanopore sequencing works

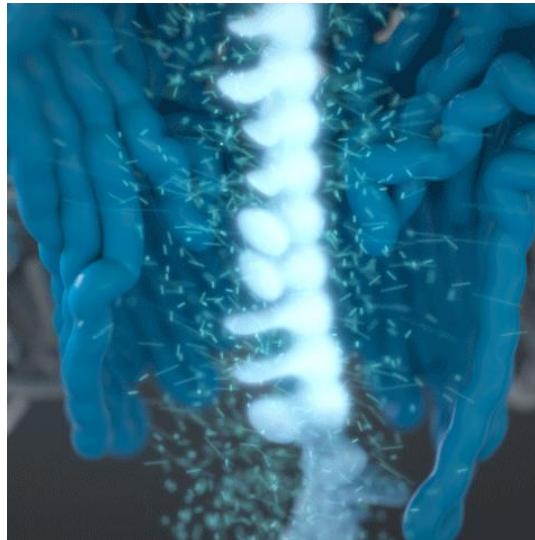


1.

A motor feeds DNA
through a nanopore

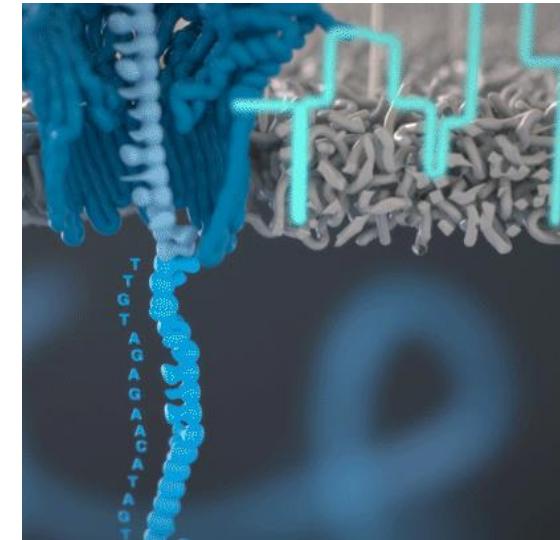


Watch the full video on how nanopore sequencing works



2.

The DNA blocks the flow
of current through the
pore

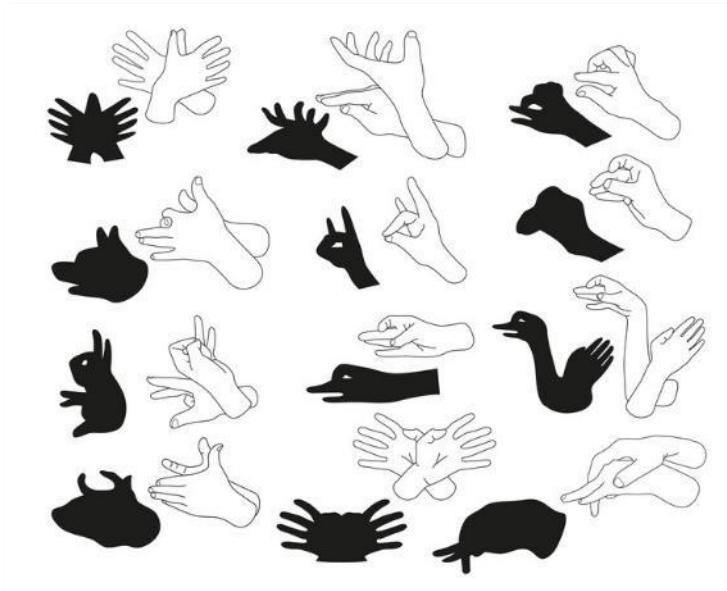


3.

The changes in current
are decoded into the DNA
sequence – this is called
basecalling

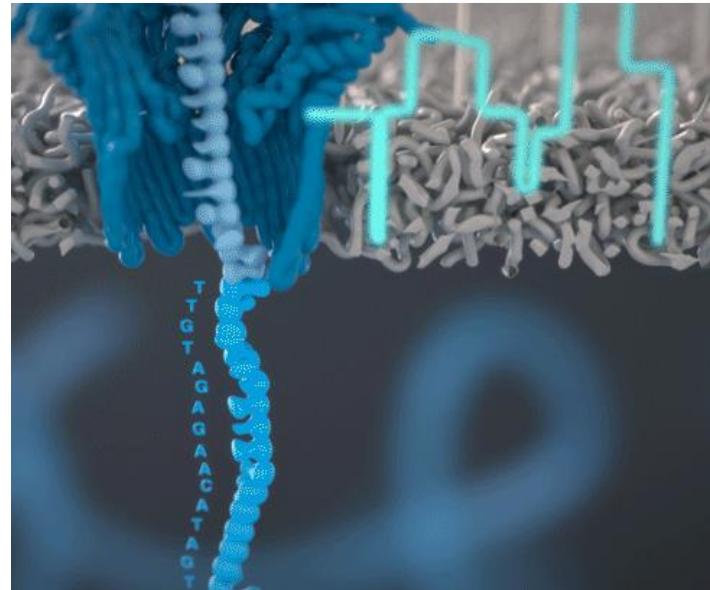
How does basecalling of nanopore data work?

The shadow puppet analogy



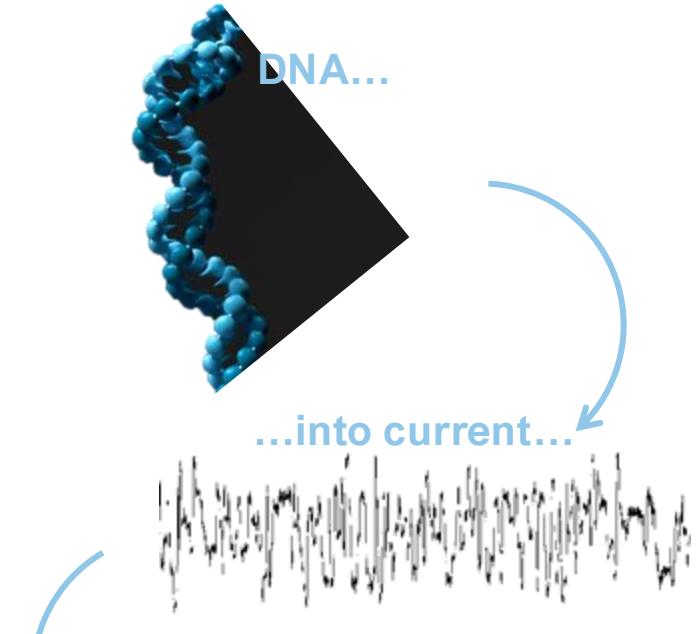
The hands block the light

Because the hands are a certain shape, it creates a shadow we recognise



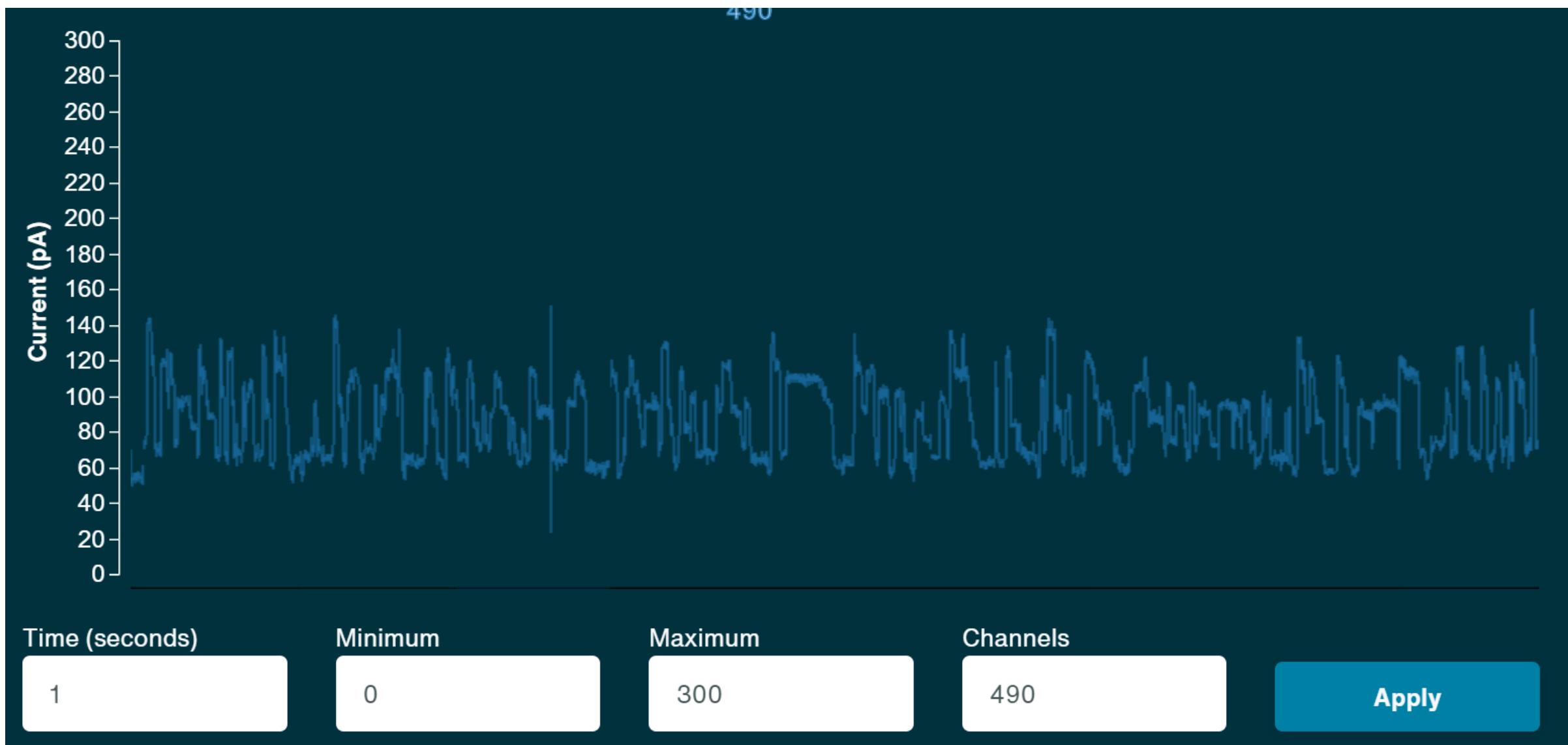
Each base of DNA blocks the current

Because the bases are a certain shape and size, it creates a signal we can recognise



AATCACATCAAGACTGCTTAAAGGAAGCAACTTTGAGTGTTAATATGGTACCCGTGATAA
GGGATGGGGTAGATGAGATGGGCCAGGGCACCCGAGGTCAAGTCATTACCCCTTCATAGGGAGTG
TTCTAGATCATACACCATATCATTGAGCATGACACTATCTGGGTGTCGCCAACCTGGTAG
TCATTATTCATCATGCCTTGAGTAGCTGGTGAGGATATCCCCAGAATGAAAATCTTTTC
ACTGACAGTCATATTGGGGTGCTCTTAAGCTTTCCACTTGGCTGGGTCTGAGGCTCCGGCC
GACCTATTTGTGGGGACACTCGGGGTAGTCGTTGGCTTATGACCCGTAAGTCTCCGGCC
GACCTATTTGTGGGGACACTCGGGGTAGTCGTTGGCTTATGACCCGTAAGTCTCCGGCC
CTCCCGCTACAGAAGATGATAAGCTCCGCAAGCAATTATGAACAACCCAAGGATCGGGATATAA
AACAGAGAACAGGGCTGATTACACTTGTGTTGGTATCCGCTAAATAGCCTCGGGAGCCTTATG
CATACTGCTCCGGGAGCACTCTGGTAACGGTTATGTCCTAGGACATTATGCTTCGGGTAT
GGGCTCTATTGACGATCCTTGGGCCAGAGATGCTGGCACAGGACTAAATTAGAGCAGTCACA
ACTGTAAGGTCCTCACCGCAGCACGGCCAGGGAGACACTGACCCATCAACCTGTAACGGAAC
CTTCCTGATCTGCTGGAGAGATACTACAGTCGGCTTACAGGCCCTTGTGCGTCGCC
CTGCTGATCTGCTGGAGAGATACTACAGTCGGCTTACAGGCCCTTGTGCGTCGCC
GATCTGATCTGCTGGAGAGATACTACAGTCGGCTTACAGGCCCTTGTGCGTCGCC
GAAATGAAGTCTGATGCGACAGAAACTGTCAGCTACCTATCTCTTGTGAGCTTGTGAC
GATTCTGTC

Raw signal from a Read



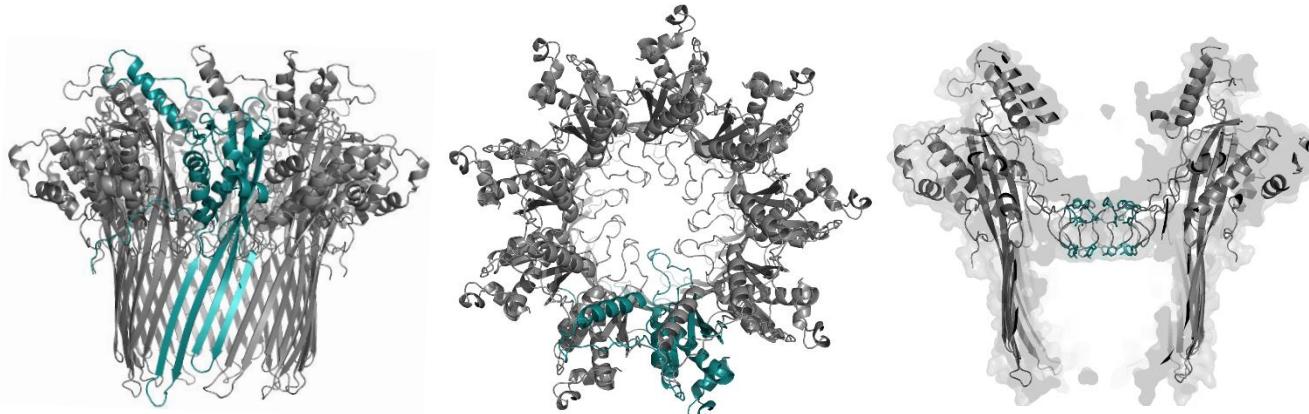
Where do nanopores come from?

They exist in nature but are engineered for nanopore sequencing

This is the CsgG pore

A nine-subunit protein from the outer membrane of *E.coli*

In nature it facilitates excretion of amyloid proteins for biofilms

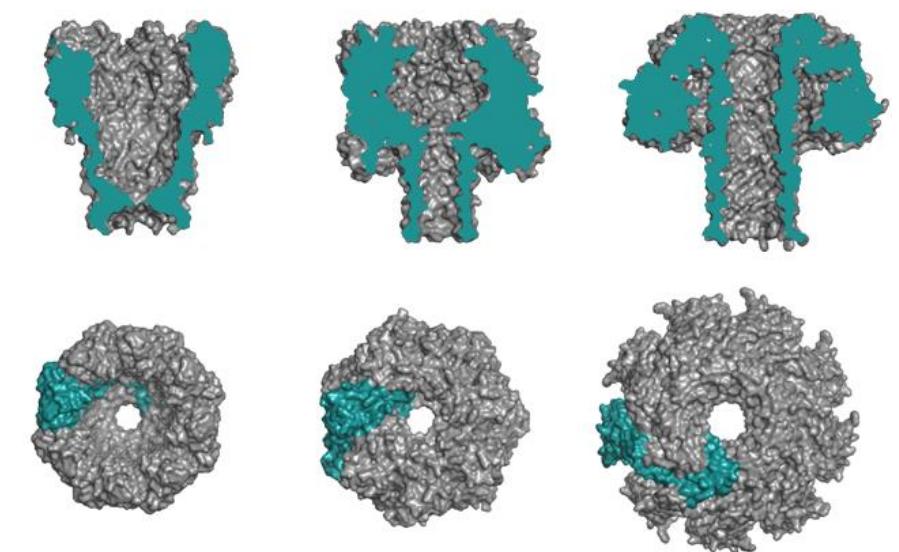


Read more about CsgG

Lots of nanopore families exist

They exhibit different structures

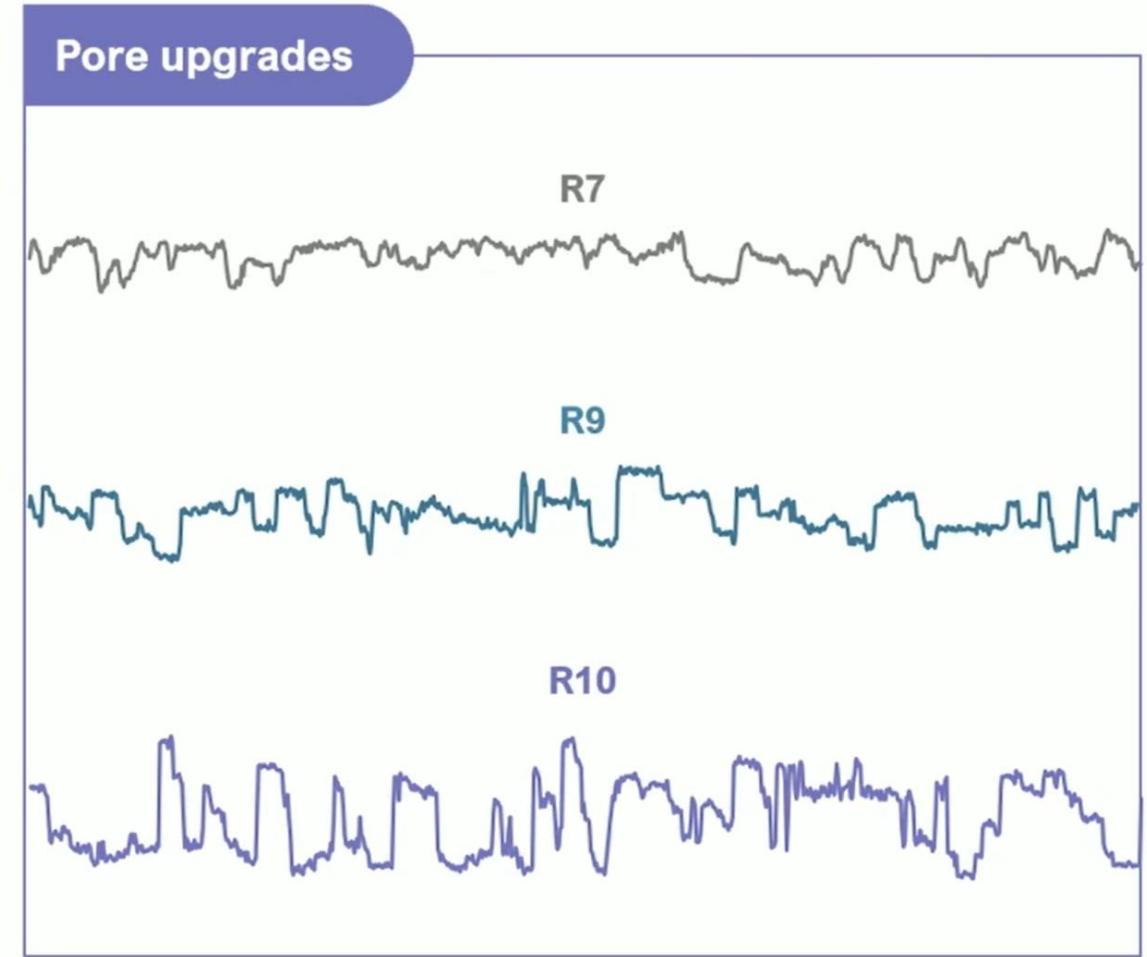
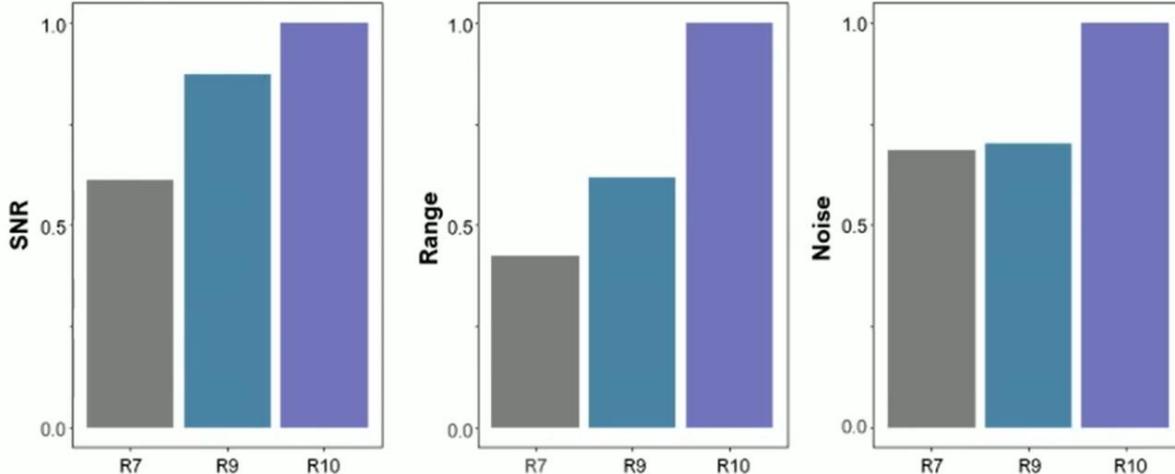
They may be useful for different applications



Evolution of nanopores

Designed to **improve base resolution** and **resolve low complexity signal**

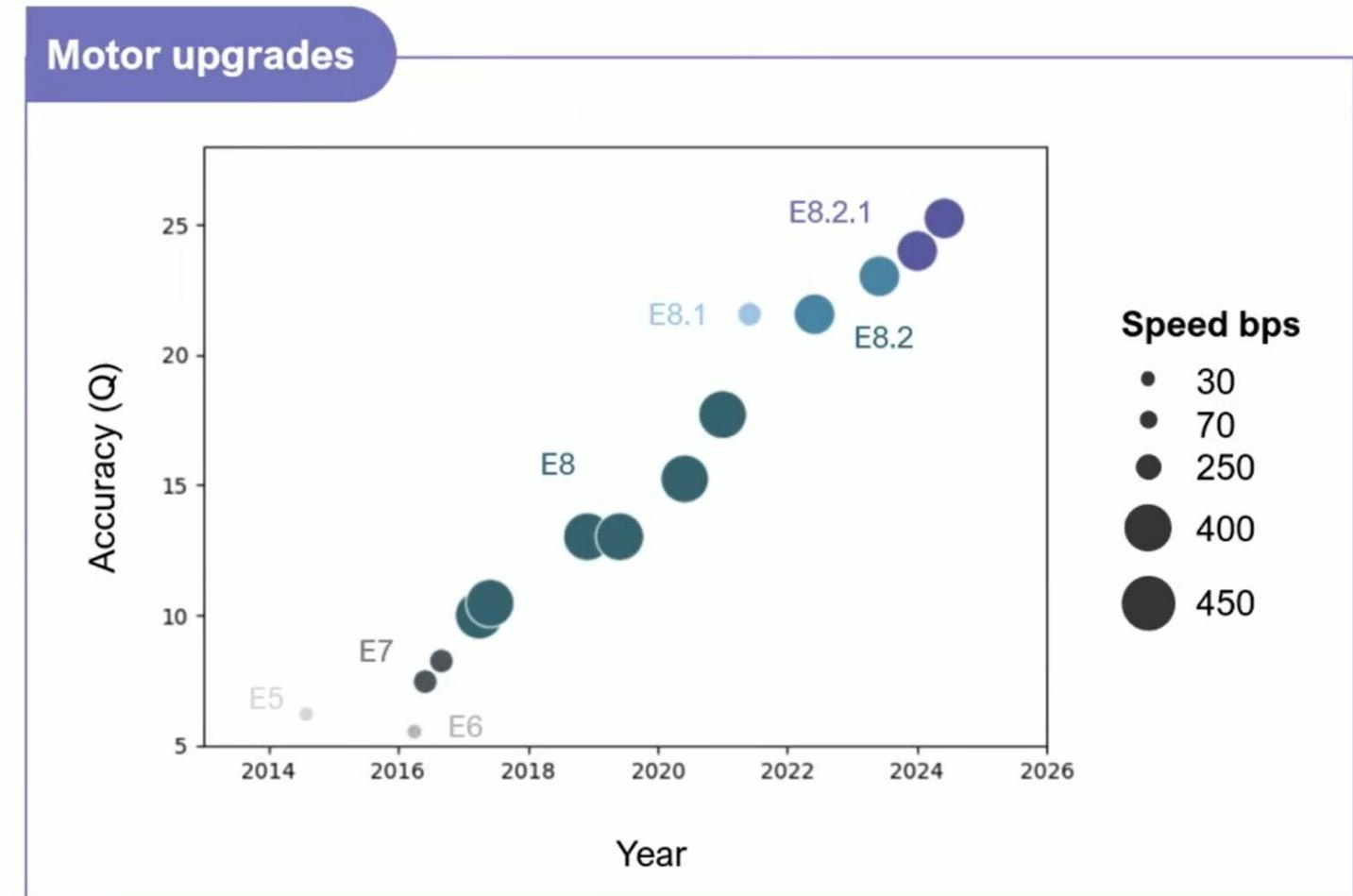
- Significant accuracy gains have come from **upgrading** the class of reader
- Minor revisions have increased **stability, robustness** and **sensitivity**
- Over **11,000 pore mutants** have been tested in-house to date



Evolution of motors

Enzyme engineering

- Motor controls the movement of **DNA / RNA** through the pore
- Motor performance **increased** over the years
 - ✓ Faster motors increase output
 - ✓ Movement consistency improves accuracy
 - ✓ Motor processivity increased read length
 - ✓ Motor stability extends run time
- AlphaFold-like **guided designs**
- **Automated** screening pipelines in place

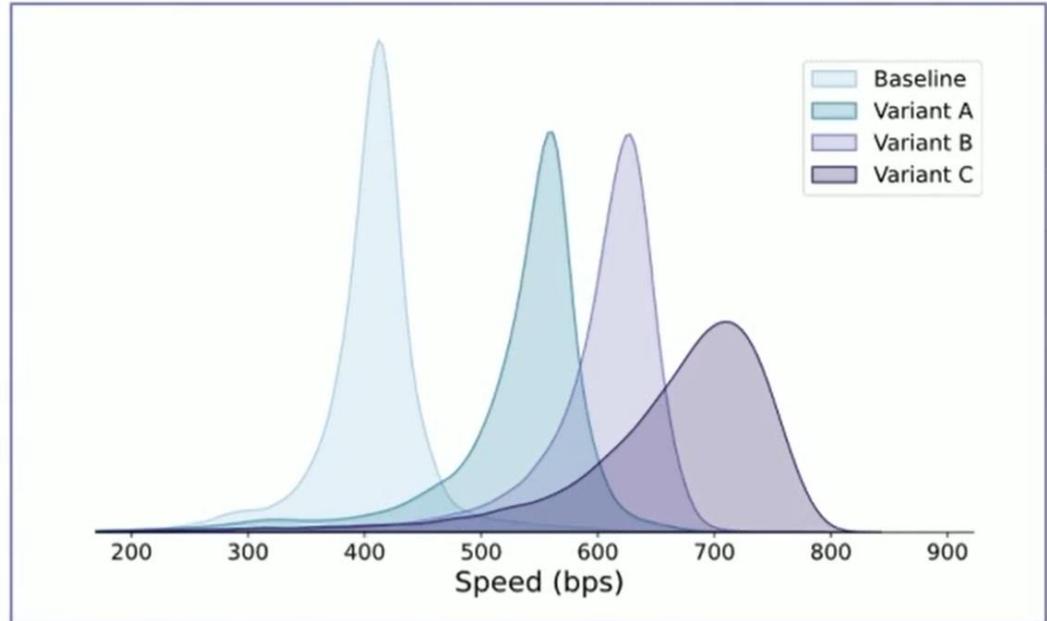
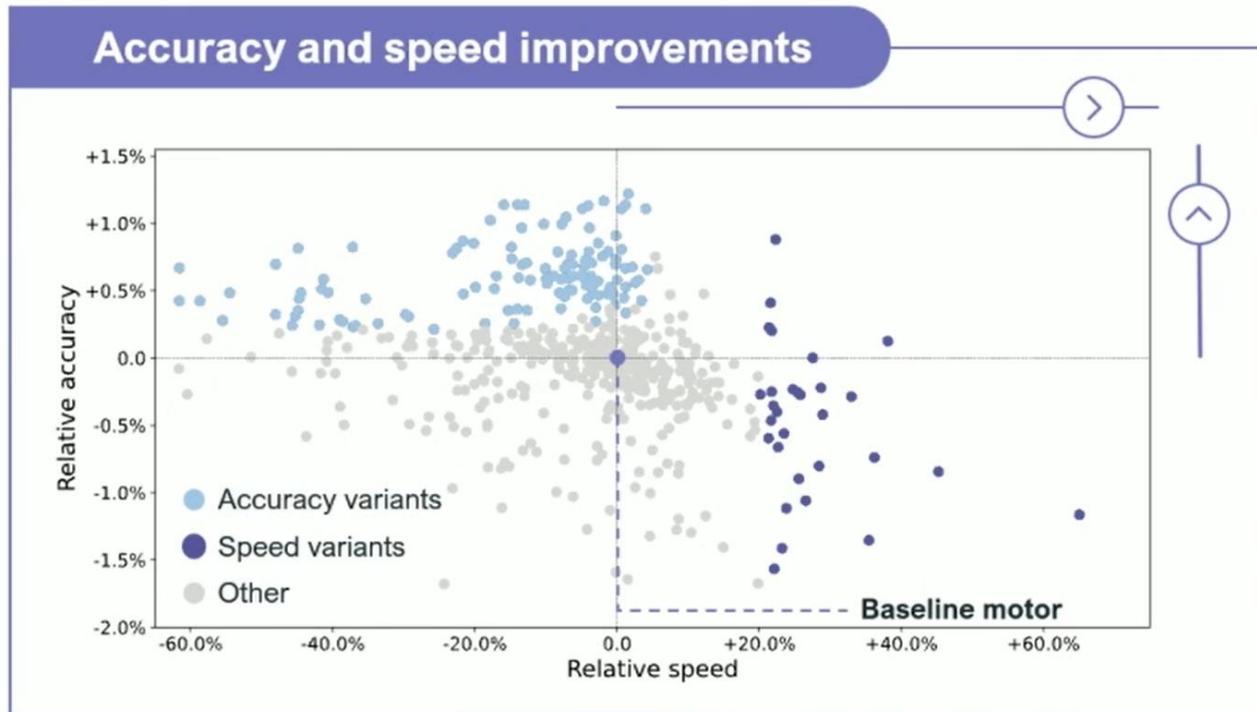


Improving output

High-speed motors

AI guided motor engineering

- Use protein-language, 3D-structure and hybrid models
- *In-silico* ranking and selection of variant libraries
- Enabling targeted exploration of possible mutants and variants



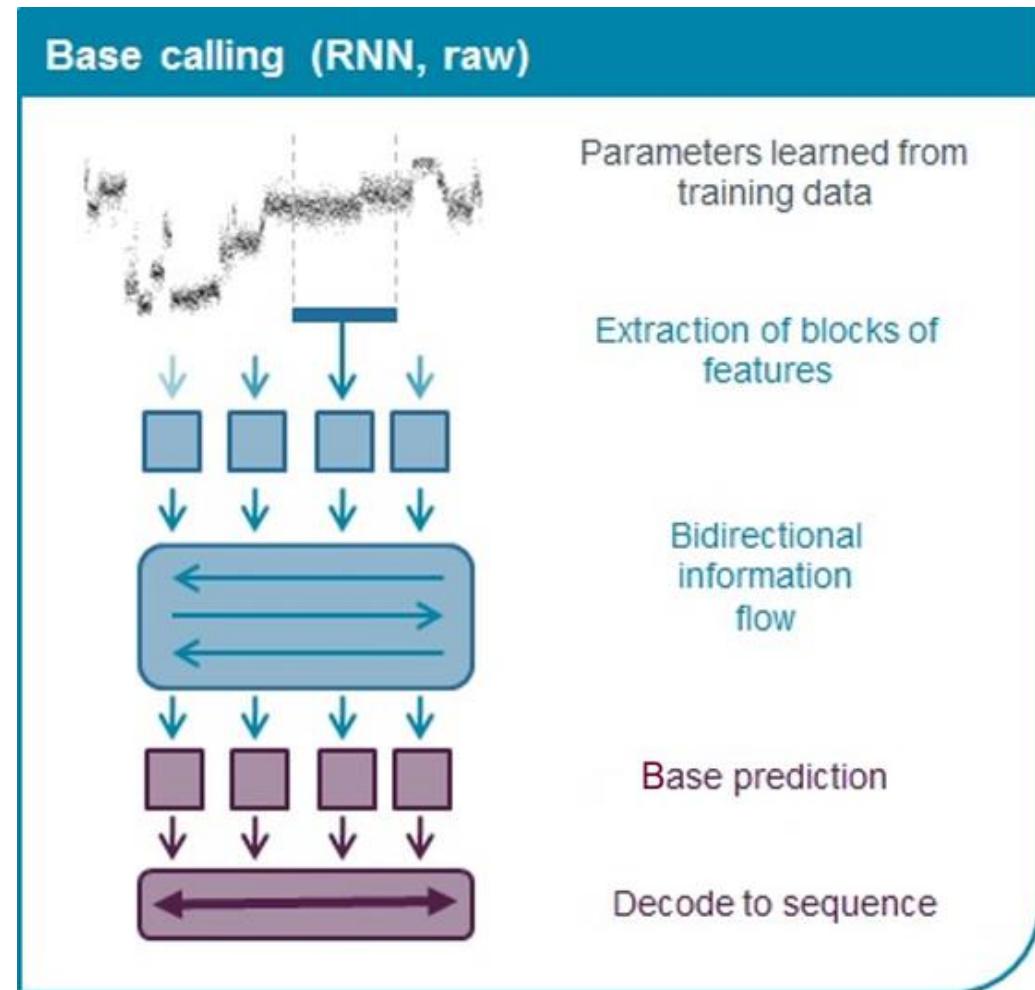
High throughput automated screening pipeline

- Liquid handling robots
 - lysate to sequencing library
- Improved analysis pipeline for:
 - speed, accuracy, stability, blocking, movement



Basecalling with neural networks

- Basecallers use bi-directional Recurrent Neural Networks (RNNs) to convert raw signal data into basecalls.
- Neural networks, inspired by the human brain, are trained computational models excellent at signal processing and pattern recognition.
- RNNs have an internal memory, allowing current computations to leverage information from previous data points.
- Bi-directional RNNs enhance accuracy by considering both preceding and succeeding signal data.



Transformer models (V5 SUP)

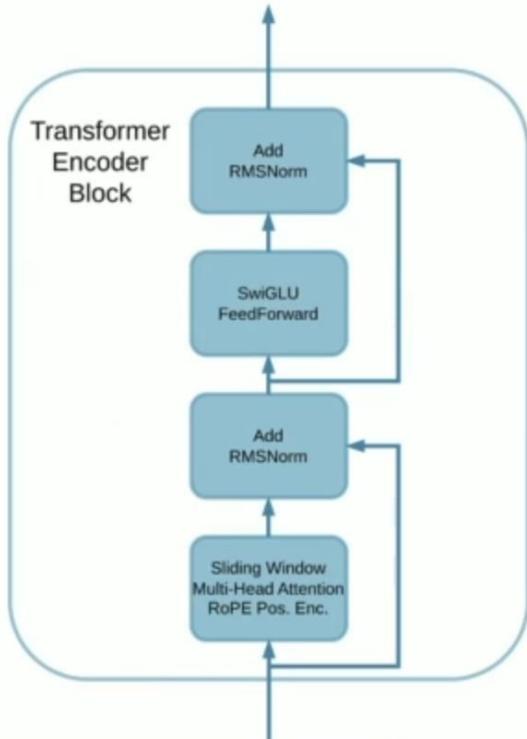


Released in Dorado

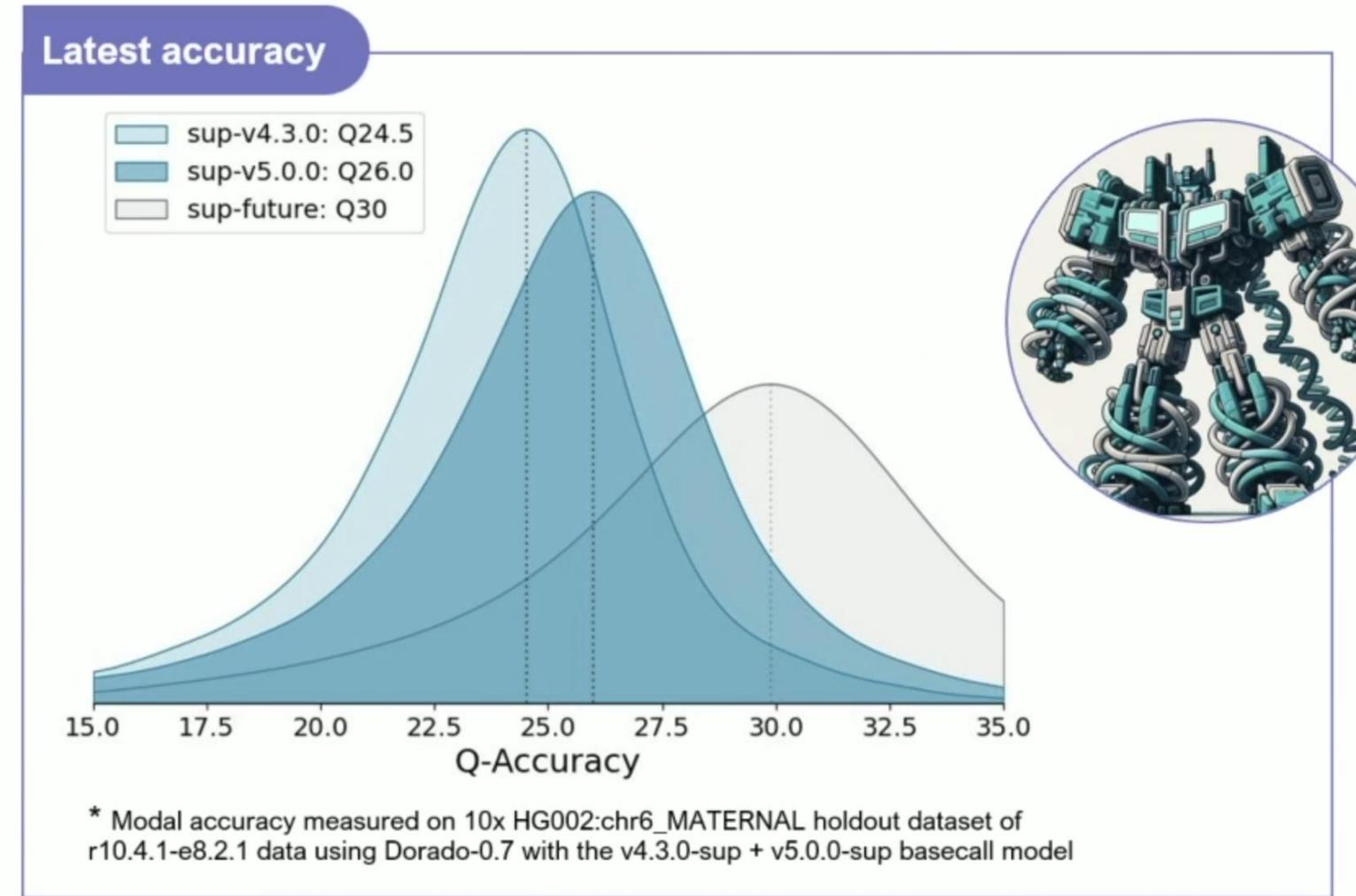
LC
LONDON CALLING
2024

The first **big change** to Oxford Nanopore basecalling model architecture in years

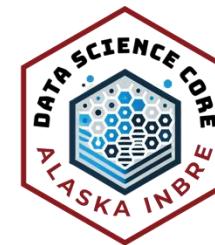
- Model architecture inspired by **Llama 3**
- HAC models remain **LSTM**



Now in Dorado 0.7

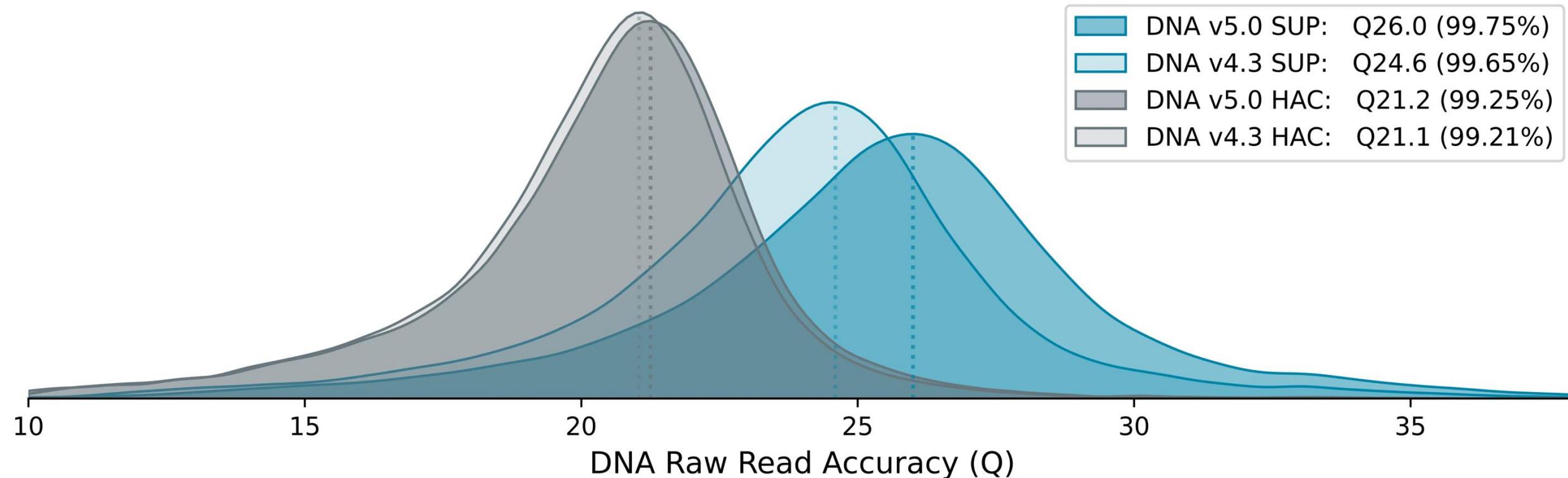


Tuning accuracy for your experimental need



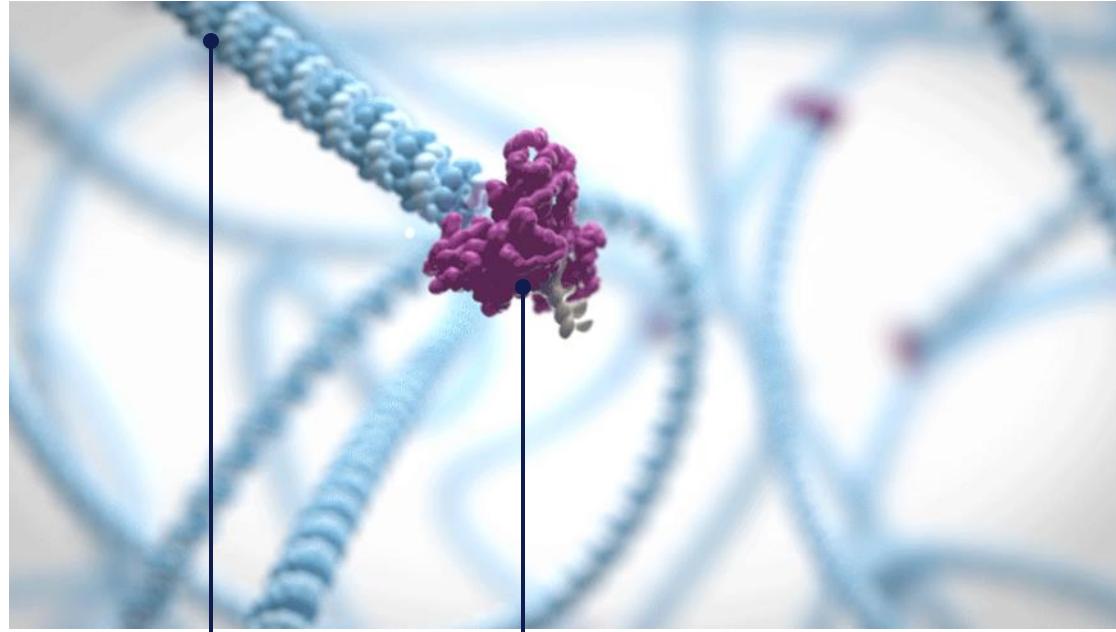
- Fast basecalling: fastest, least computationally intense.
- High accuracy basecalling (HAC): highly accurate, intermediate speed and computational requirement.
- Super accuracy basecalling (SUP): the most accurate and computationally intense.
- Duplex basecalling: is recommended for hemi-methylation investigation, enabling the methylation signature of each DNA strand to be distinguished.

Nanopore raw reads now achieve 99.75% (Q26) accuracy with the latest Dorado basecalling models (v5)





Preparing DNA for nanopore sequencing



DNA strand
Adapters with motor
protein added to the
end

Motor protein
Moves the DNA
through the pore

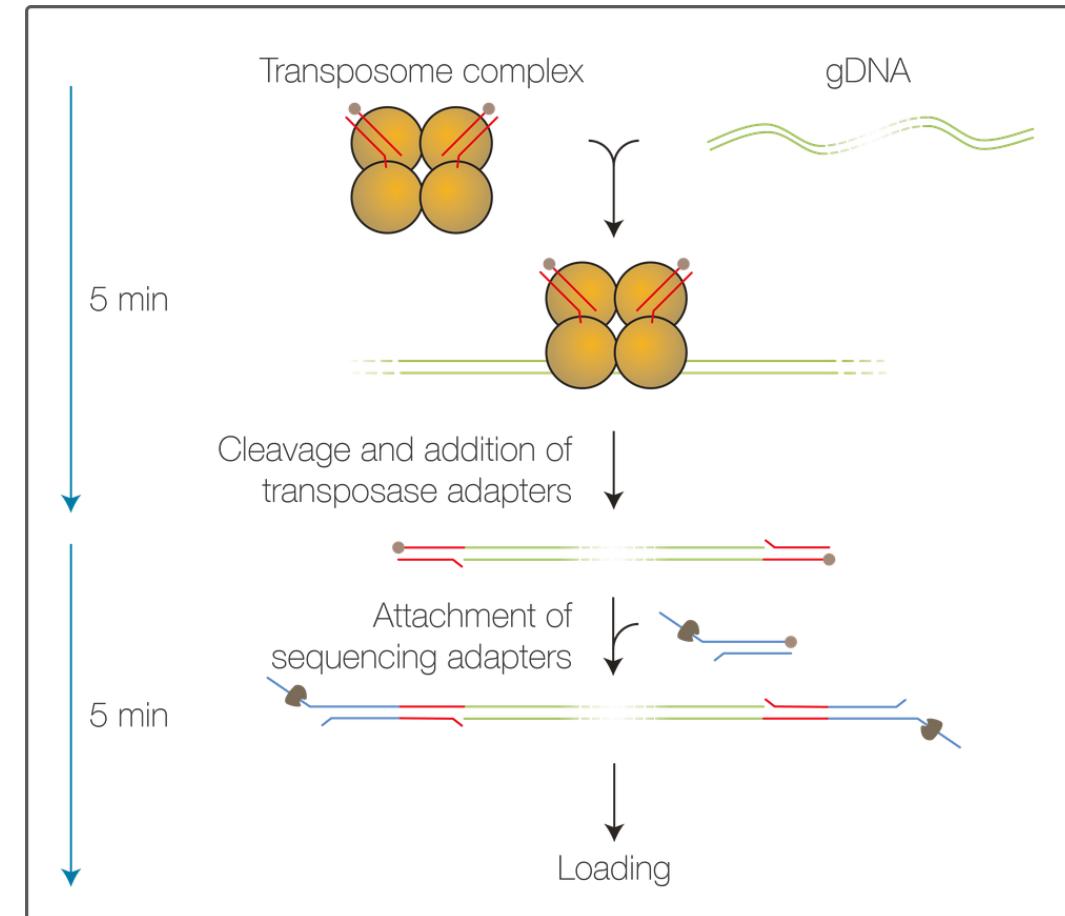


Comparison of Sequencing kit options

	Ligation Sequencing	Rapid Sequencing
Optimized for	Output/Yield	Speed
Preparation time	60 mins	10 mins
Input	~1000 ng gDNA	~200 ng gDNA
Fragmentation	Optional	Transposase-based
Amplification	No	No
Barcode options	24, 96	24, 96
Typical output	***	**
Adaptive sampling	Yes	Yes
Methylation included	Yes	Yes

Rapid Sequencing Library Preparation

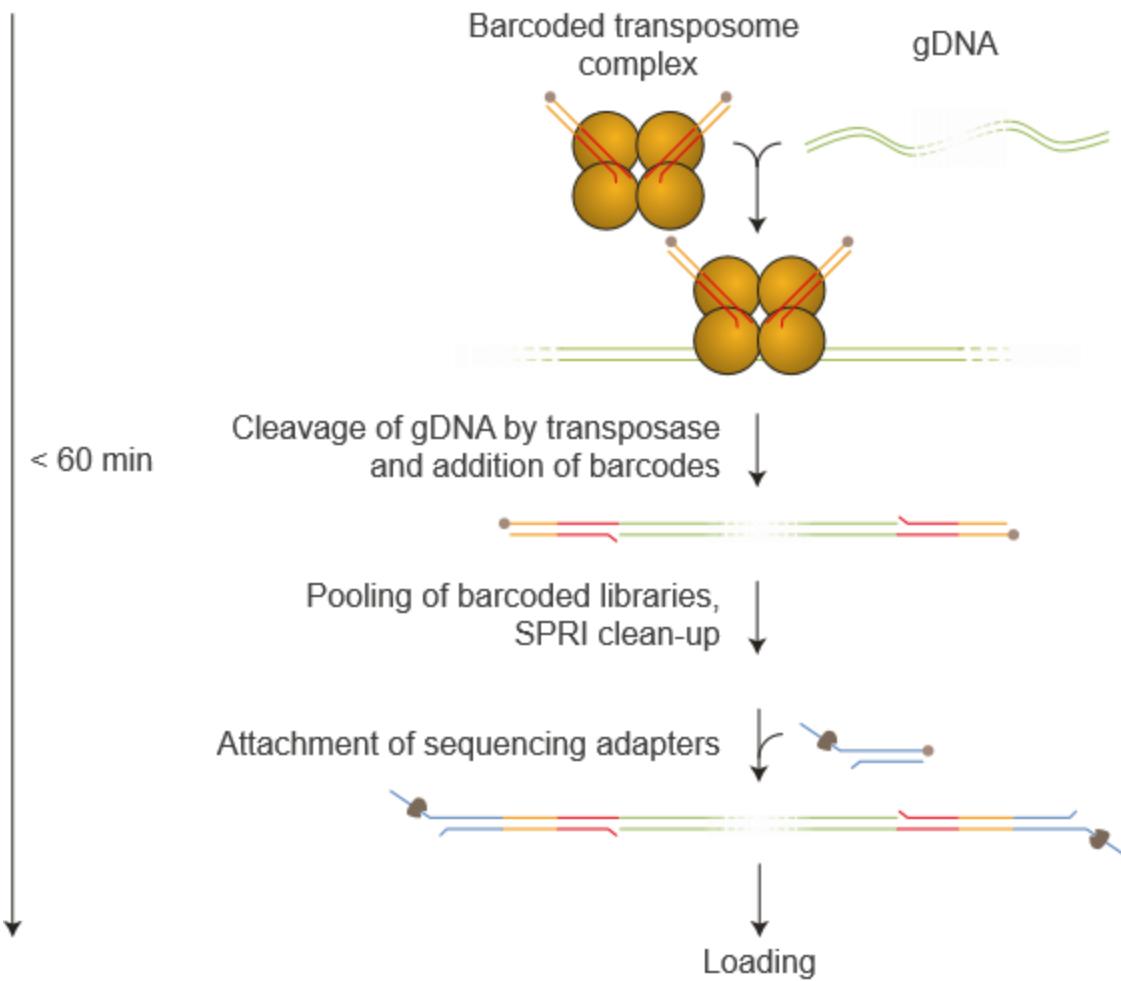
Library preparation step	Process	Time
Fragmentation	Tagment your DNA using the Fragmentation Mix	5 mins
Adapter attachment	Attach sequencing adapters to the DNA ends	5 mins
Priming and loading the flow cell	Prime the flow cell and load the prepared library for sequencing	5 mins





Rapid Barcoding Sequencing Library Preparation

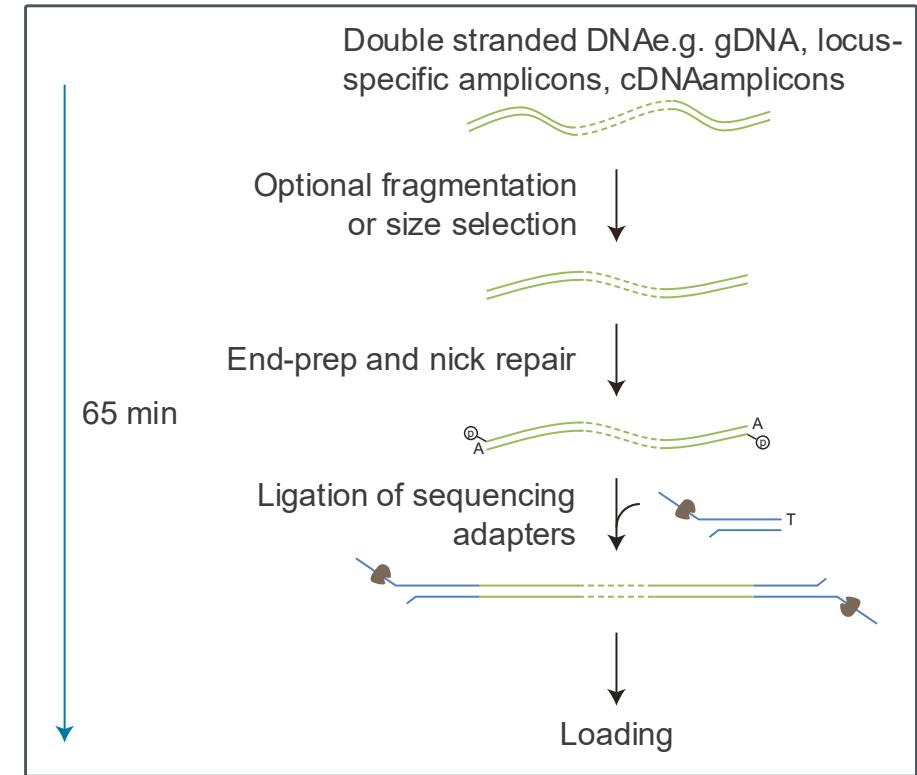
Library preparation step	Process	Time
DNA barcoding	Fragmentation of the DNA using the Rapid Barcoding Kit V14	15 mins
Sample pooling and clean-up	Pooling of barcoded libraries and SPRI Bead clean-up	25 mins
Adapter ligation	Attach the sequencing adapters to the DNA ends	5 mins
Priming and loading the flow cell	Prime the flow cell and load the prepared library for sequencing	5 mins





Ligation Sequencing Library Preparation

Library preparation step	Process	Time
DNA repair and end-prep	Repair the DNA and prepare the DNA ends for adapter attachment	35 mins
Adapter ligation and clean-up	Attach the sequencing adapters to the DNA ends	20 mins
Priming and loading the flow cell	Prime the flow cell and load the prepared library for sequencing	10 mins



Hands on with MinIONs



The Oxford Nanopore MinION

Consumable flow cell
Contains sensing chemistry, nanopores, and electronics

Sample added to flow cell here

Sensor chip
With multiple nanopores

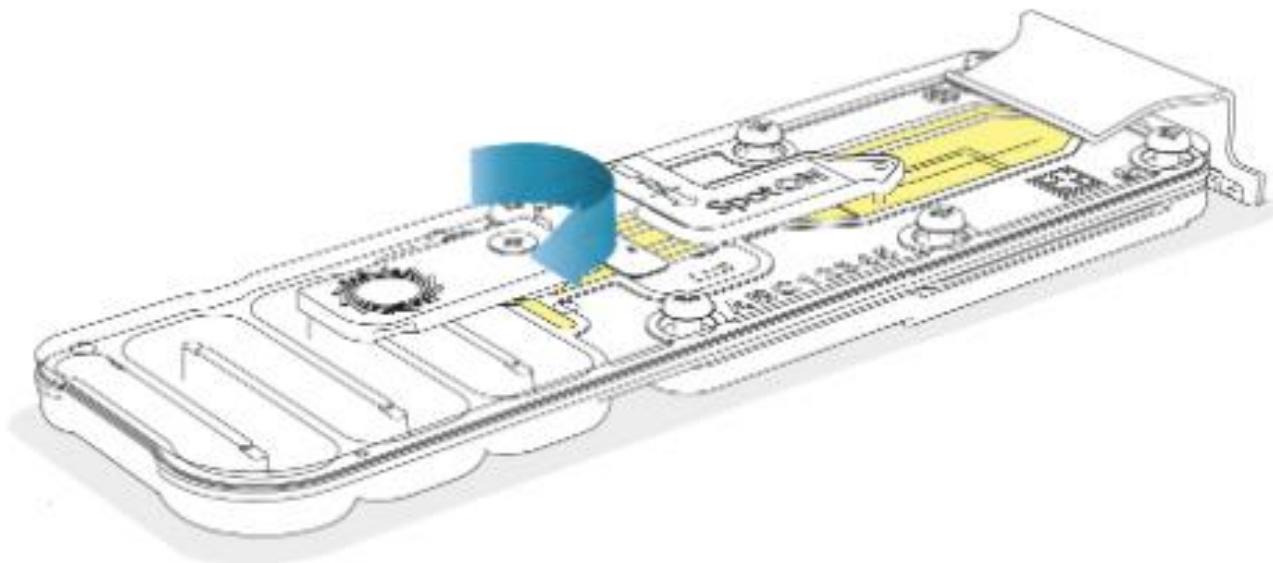


USB
Powers device and passes data to PC

MinION
sequencing device

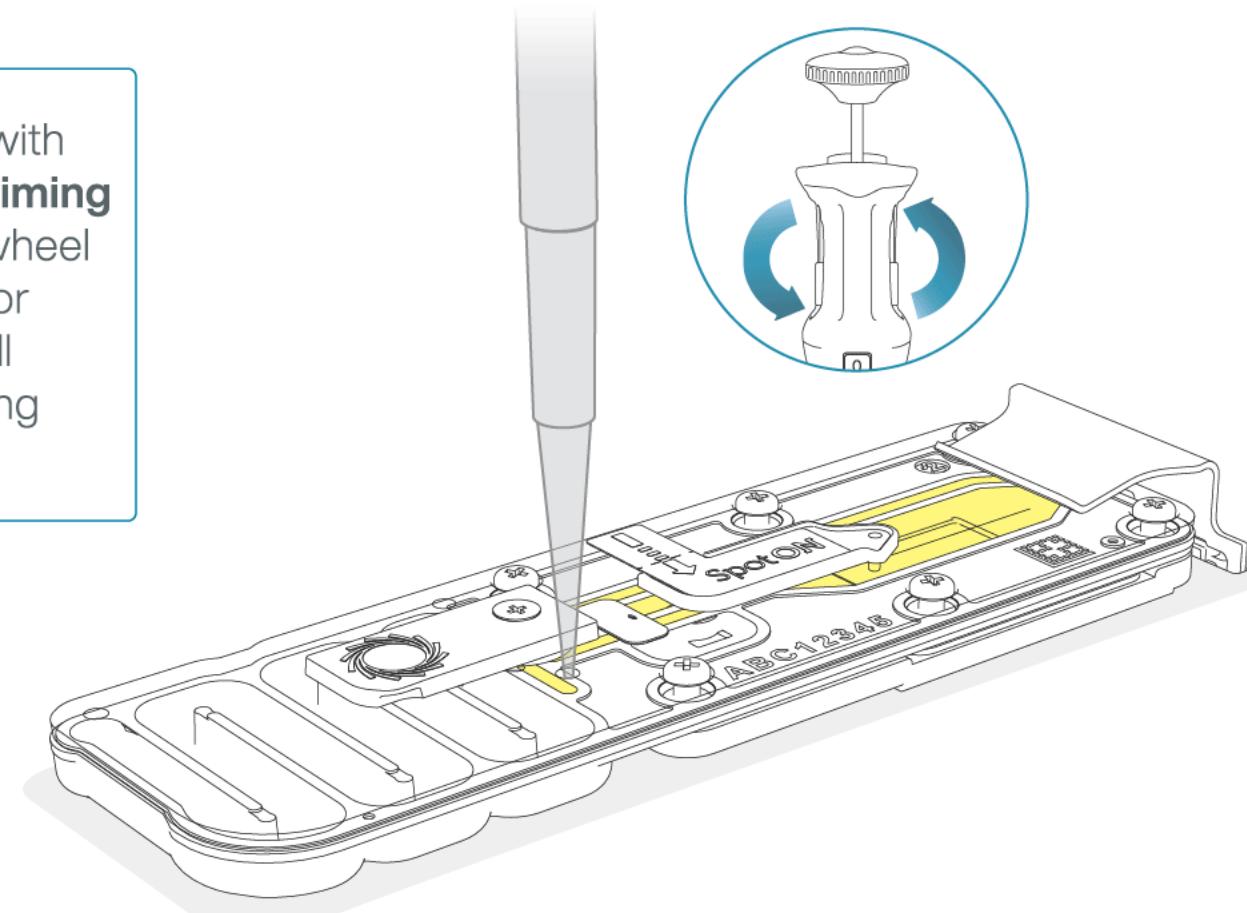


Priming Flow Cell



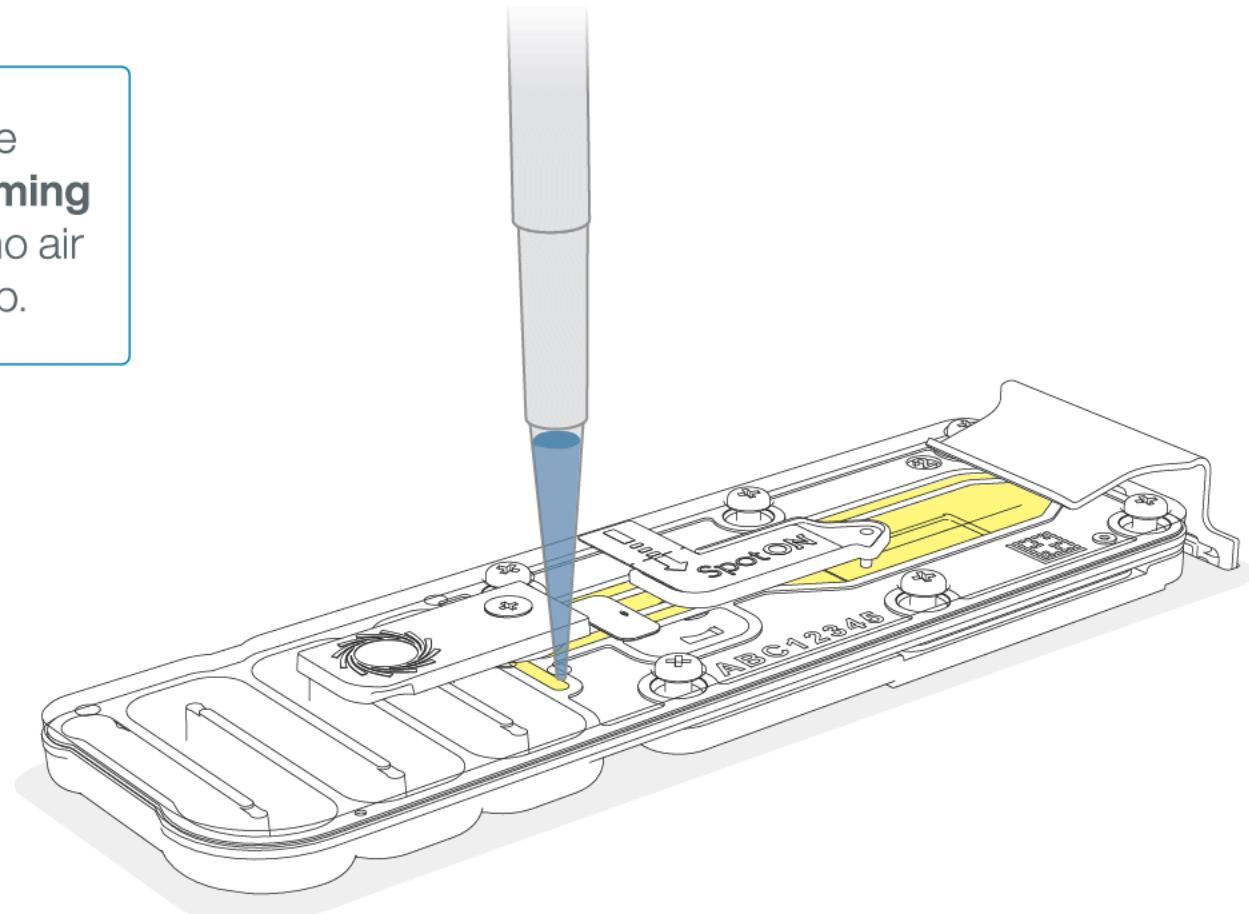
3

Insert a P1000 pipette with an empty tip into the **Priming port**. Turn the pipette wheel to draw back 20-30 μ l or until you can see a small volume of buffer entering the pipette tip.



4

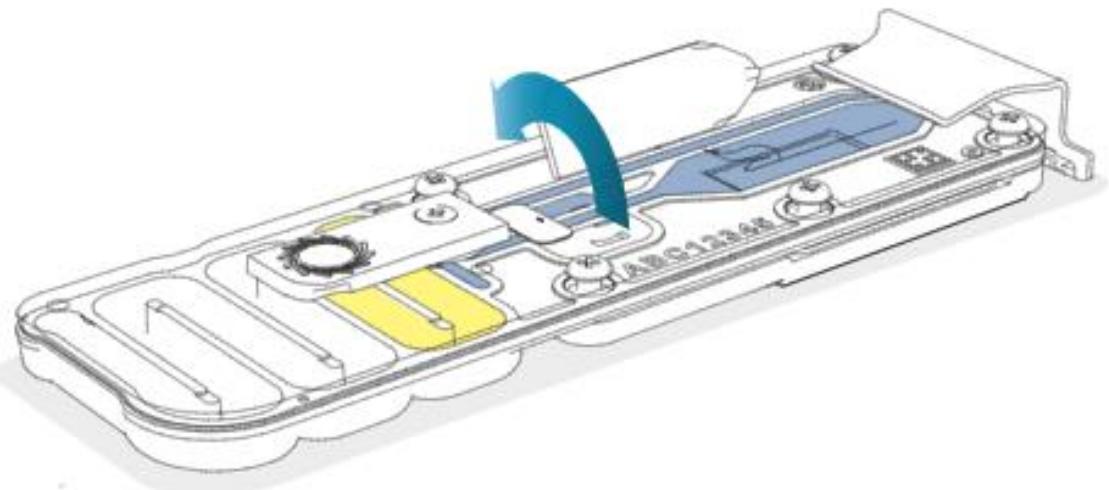
Slowly load 800 μ l of the priming mix into the **Priming port**. Ensure there are no air bubbles in the pipette tip.



Wait 5 minutes before proceeding to the next step.

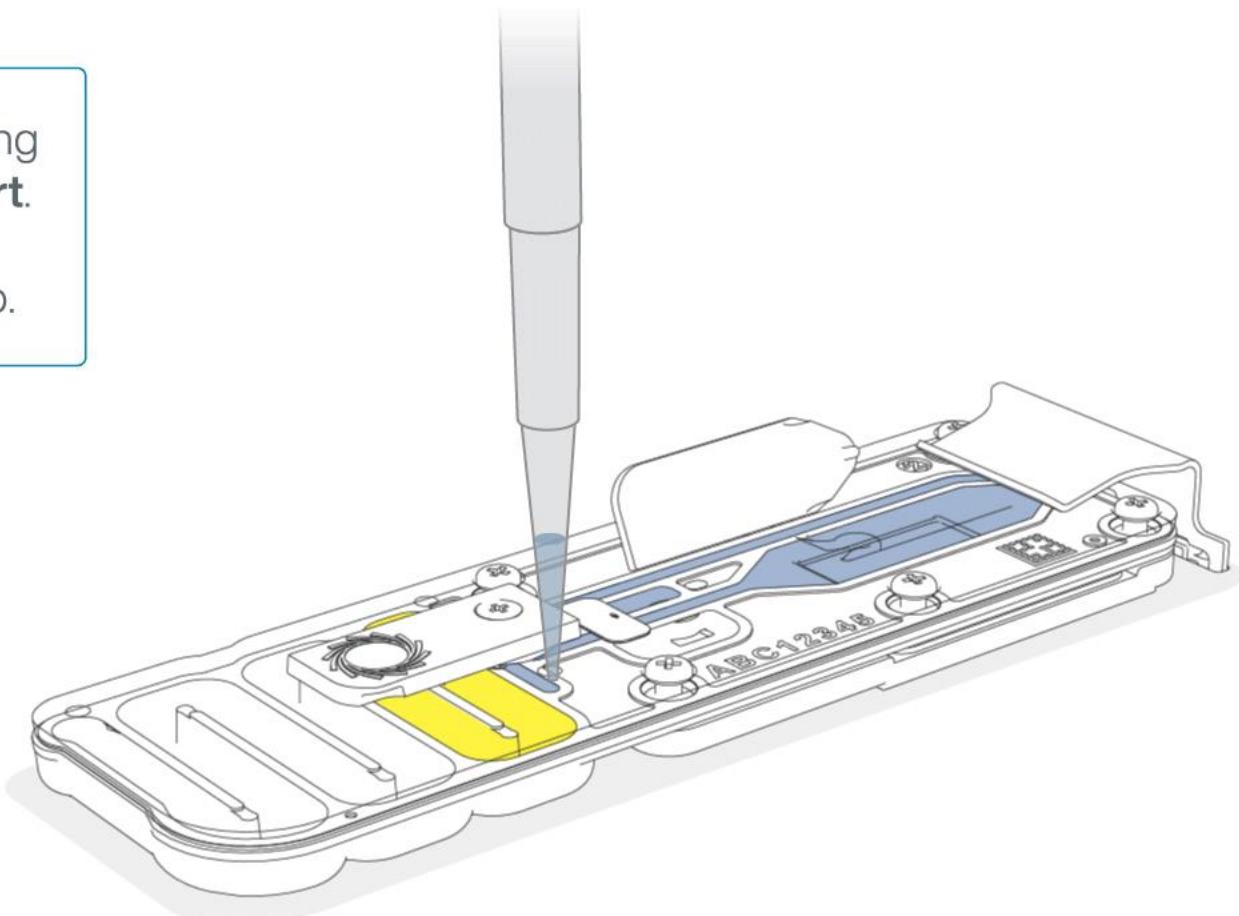
Prepare and load final library

Reagent	Volume per flow cell
Sequencing Buffer (SB)	37.5 μ l
Library Beads (LIB) mixed immediately before use	25.5 μ l
DNA library	12 μ l
Total	75 μl



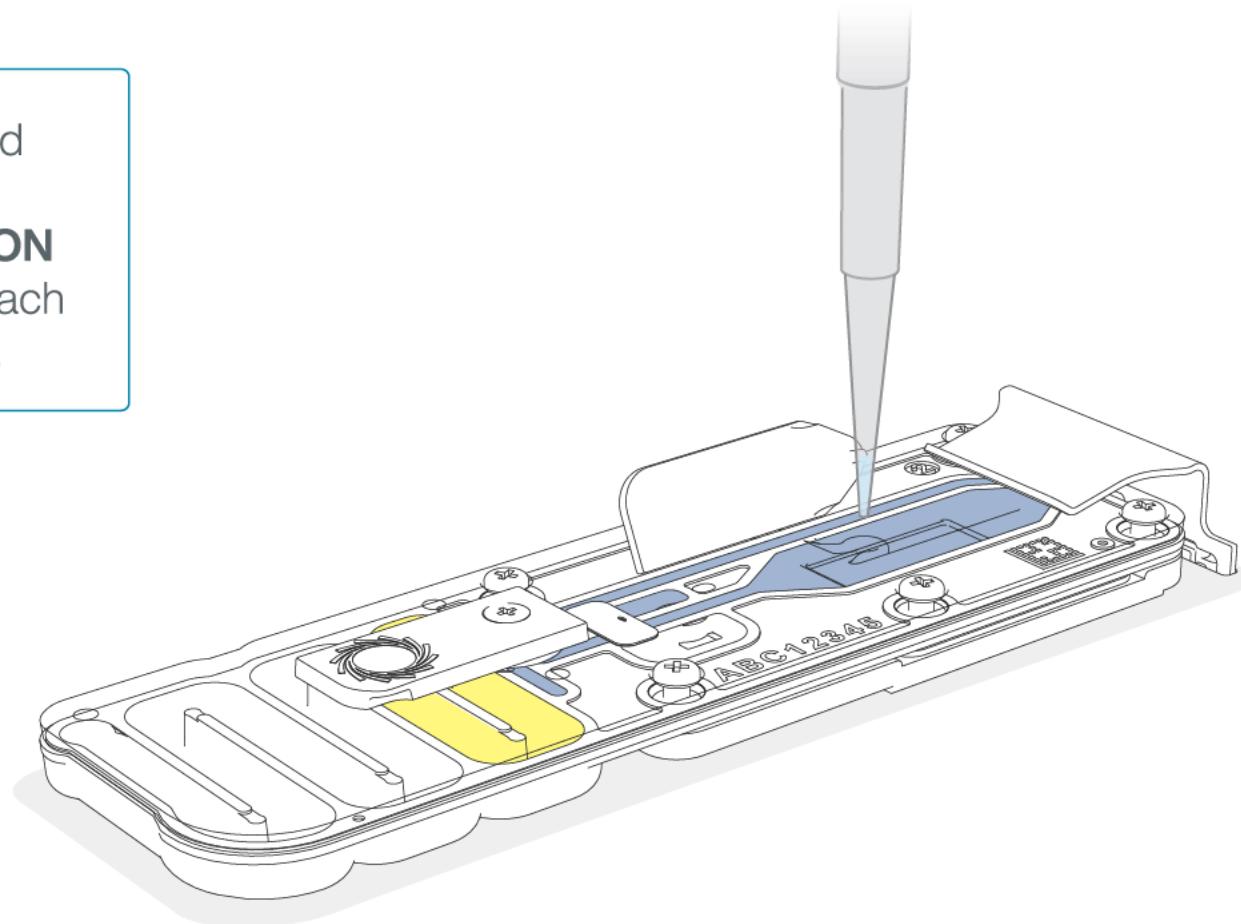
6

Load 200 μ l of the priming mix into the **Priming Port**. Ensure there are no air bubbles in the pipette tip.



7

Pipette mix the prepared library and load 75 μ l dropwise into the **SpotON** sample port, ensuring each drop flows into the port.

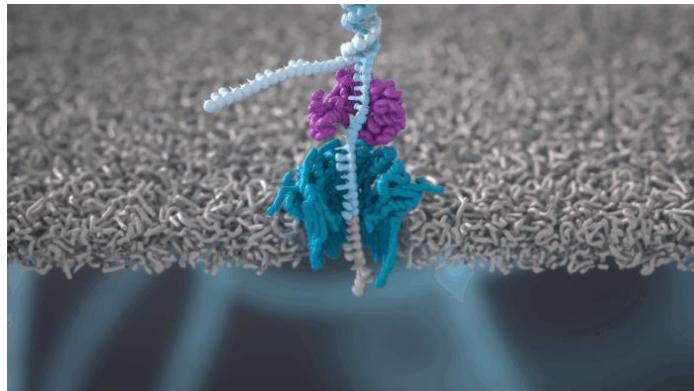




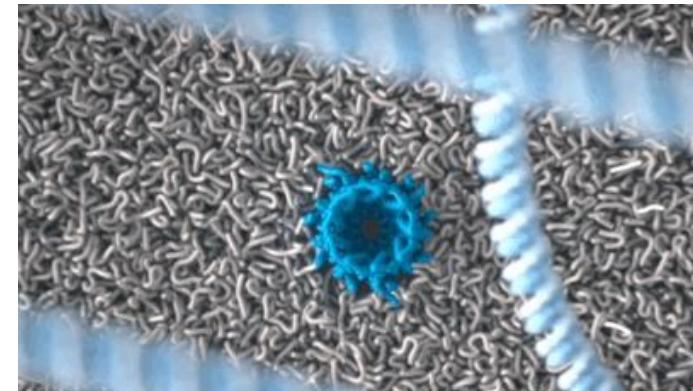
More with Nanopores!

The nanopore sequencing platform can be easily scaled

One nanopore per well



Multiple wells in parallel



Arranged in fixed flow cell sizes

PromethION

2,675
nanopores



MinION

512
nanopores



Flongle

126
nanopores



The benefits of nanopore sequencing

Sequence the original DNA, not a copy

Avoid PCR bias

Modifications are included

Can sequence the very short (20 bases) to the very long (4 million bases)



Ask every question you have of a genome

Identify every type of variant – SNP, INDEL, Methylation, SVs

Determine the parent of origin for variants (phasing)

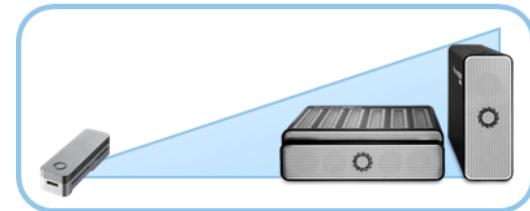
Easily carry out assemblies with long reads



It scales to you

MinION and Flongle sequencing for smaller projects

PromethION sequencing for large genomes and high-throughput projects



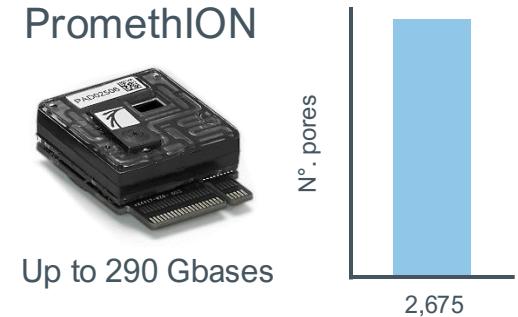
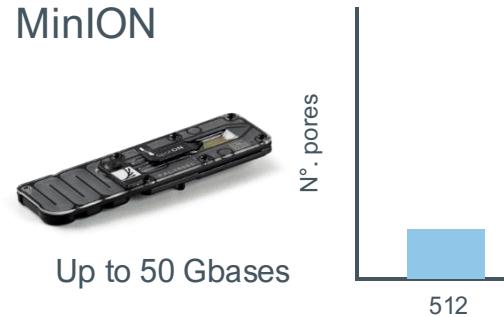
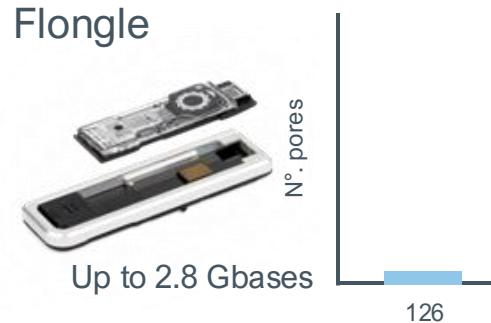
Real-time data

Make decisions as the data appears, not at the end

Carry out targeted sequencing with no extra sample prep



Flow cell types suited for different studies



Library QC
Plasmid, viral and bacterial sequencing

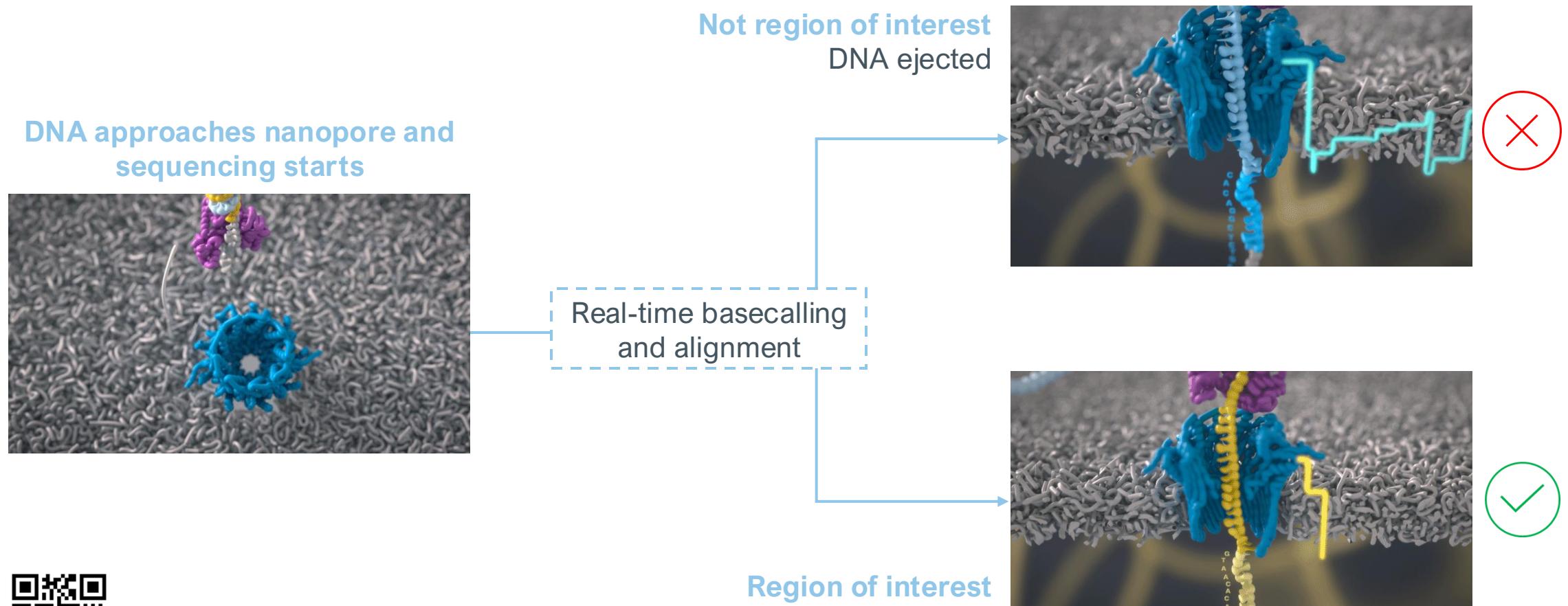
Multiplex small genomes
Low-pass sequencing of larger genomes

Generate hundreds of Gigabases of data
Sequence large genomes to high coverage

All flow cells use exactly the same sequencing libraries and underlying technology

Real-time data used for targeted sequencing

This is called Adaptive Sampling



Watch more about Adaptive Sampling