

Introduction to Nanopore sequencing

Thomas Haverkamp

Norwegian Veterinary Institute

thomas.haverkamp@vetinst.no



Veterinærinstituttet
Norwegian Veterinary Institute

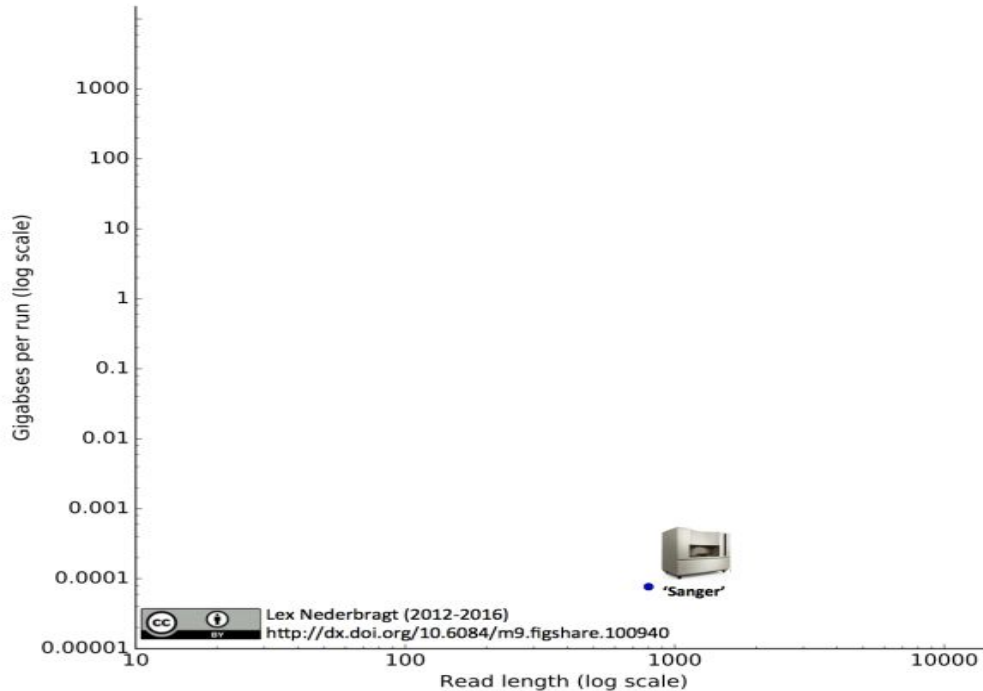


Outline

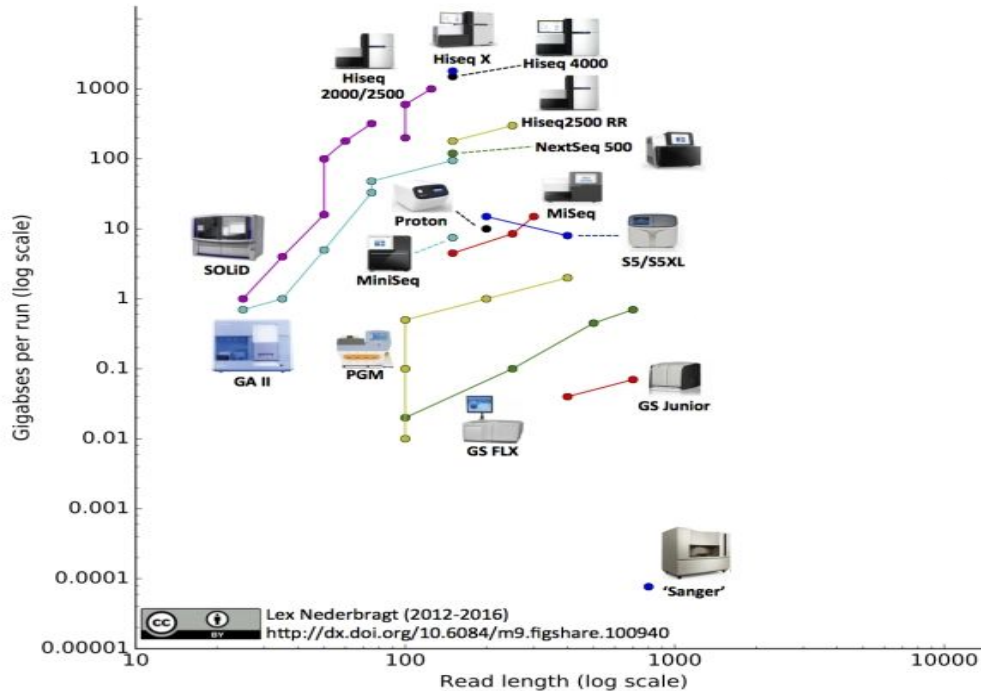
- The nanopore sequencing method
- Software applications for Nanopore
 - Genome assembly
 - Amplicon Sequencing
- A small NGS comparison

Rapid development in instrumentation

- Drastic increase in both
- Read length
 - Amount of sequence / run



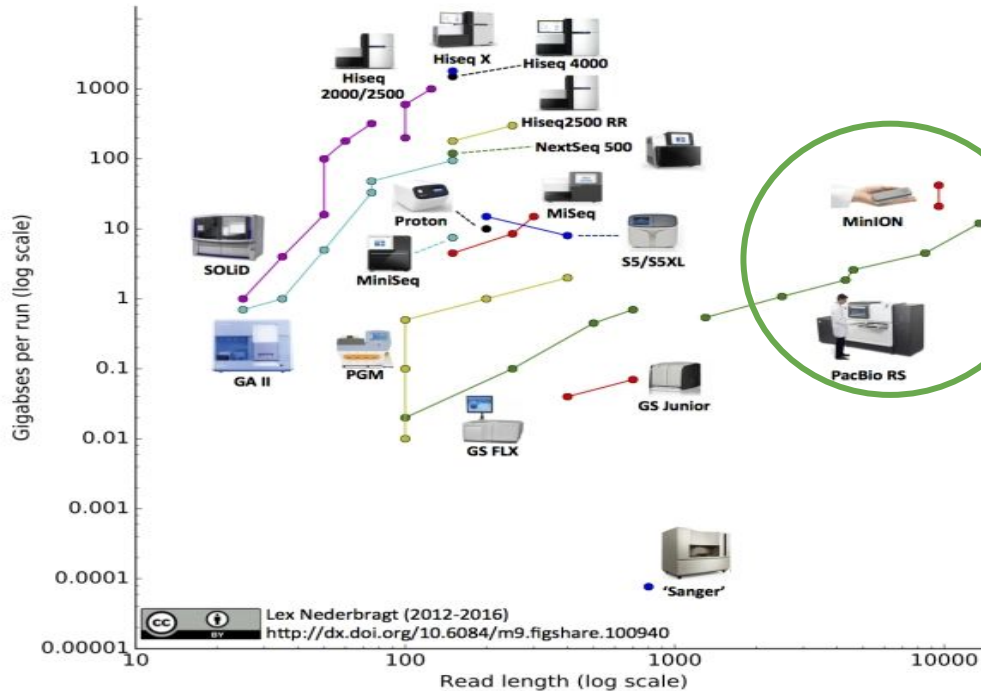
Rapid development in instrumentation



Drastic increase in both

- Read length
- Amount of sequence / run
- Single molecule sequencer
- Long read sequencers

Rapid development in instrumentation



Drastic increase in both

- Read length
- Amount of sequence / run

- Single molecule sequencer
- Long read sequencers

Oxford Nanopore sequencers



Machine	MinION	GridION 5X	Promethion P2S / P2i	promethION 24 / 48
Flowcells	1	1 to 5	1 or 2	1 to 24 / 48
Data output	10-50 Gb	75-175 Gb	580 Gb	13.3 Tb
Pores per flowcell	800	5 X 800	> 5000	>5000
Application	Field / lab based	Sequencing service	Sequencing service	Sequencing service



The nanopore flowcells



Flongle flowcell - 2.8 Gb



Normal flowcell - 50 Gb



Promethion flowcell - 290 Gb

minION & GridION machines

Promethions



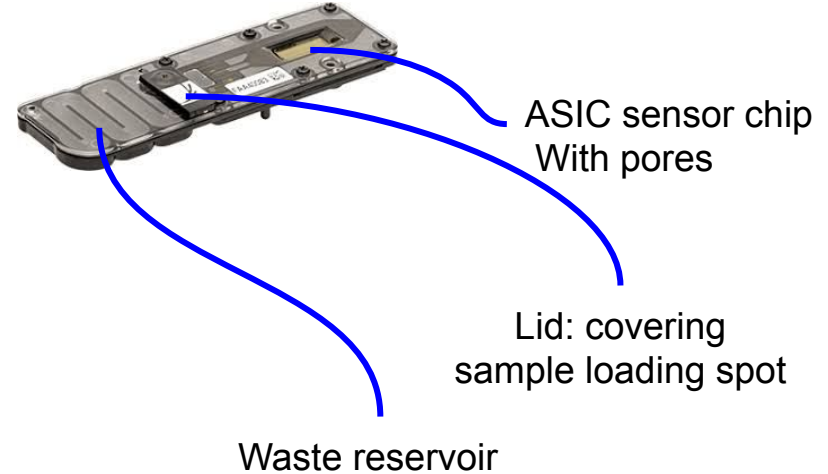
Veterinærinstituttet
Norwegian Veterinary Institute



The minION flowcell

Specifications:

- 512 pores (Guaranteed)
- Needs to be stored at 2-8 °C
- Pores deteriorate over time - Fresh is best
- Longest single read of a single molecule sequence
'Record' > 4 Mbp
- 'Happy' at about 25-50 kb
- Up to 450 bases per second / sampling rate 4000 kHz
- May give a near 'realtime sequencing' data for up to 48 hrs
- Current capacity up to 48 hrs/15-40gb



The minION nanopore sequencer



minION sequencer & flowcell

Working with the minION



WHAT THE COMPANY SAYS IT LOOKS LIKE



WHAT YOUR PI THINKS IT LOOKS LIKE



HOW THE ACTUAL DATA LOOKS LIKE



HOW REALITY LOOKS LIKE

The nanopore sensor chip

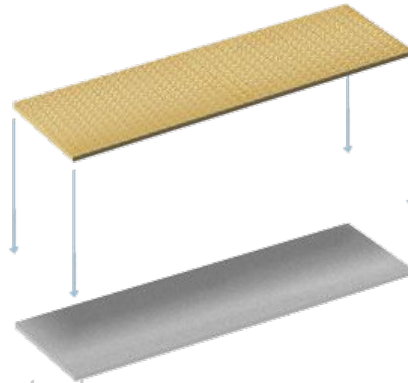


Nanopore A protein nanopore is set in an electrically-resistant polymer membrane.



Array of microcaffolds

Each microcaffold supports a membrane and embedded nanopore.



Sensor chip

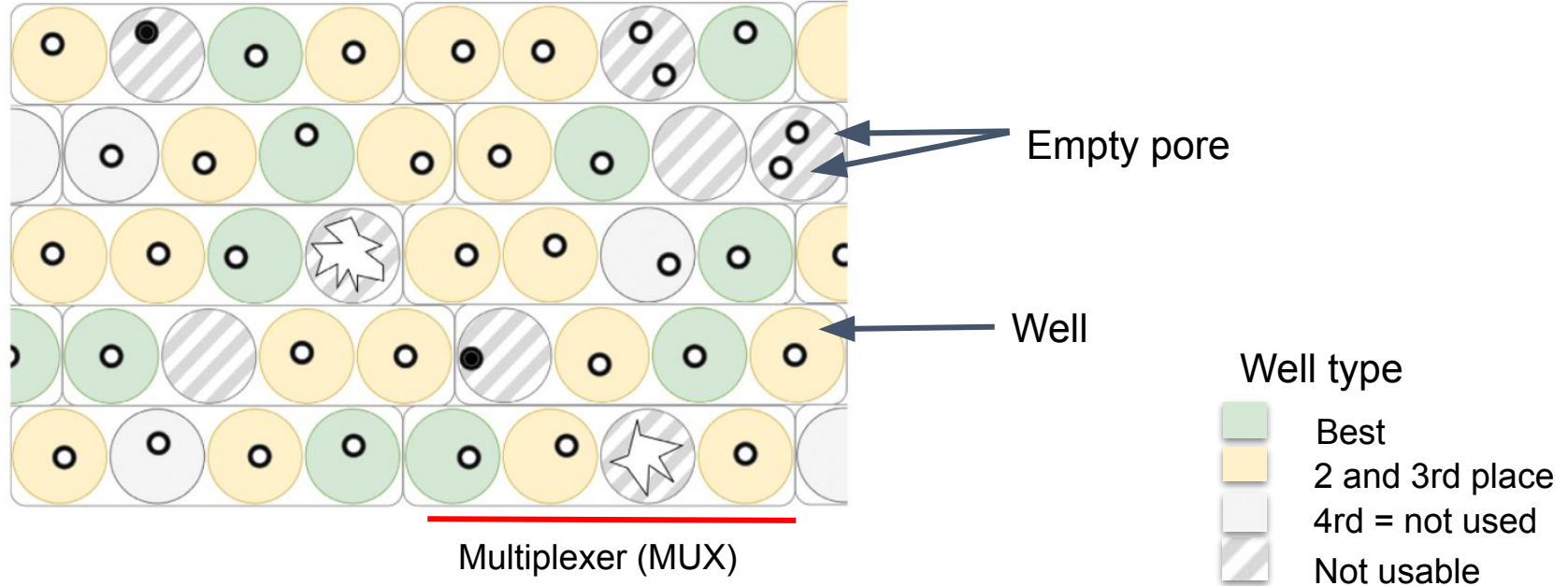
Each microcaffold corresponds to its own electrode that is connected to a channel in the sensor array chip.

ASIC Application-Specific Integrated Circuit

Each nanopore channel is controlled and measured individually by the bespoke ASIC.



The flowcell layout



A minION flow cell has 2048 wells → 512 pores sequenced in parallel

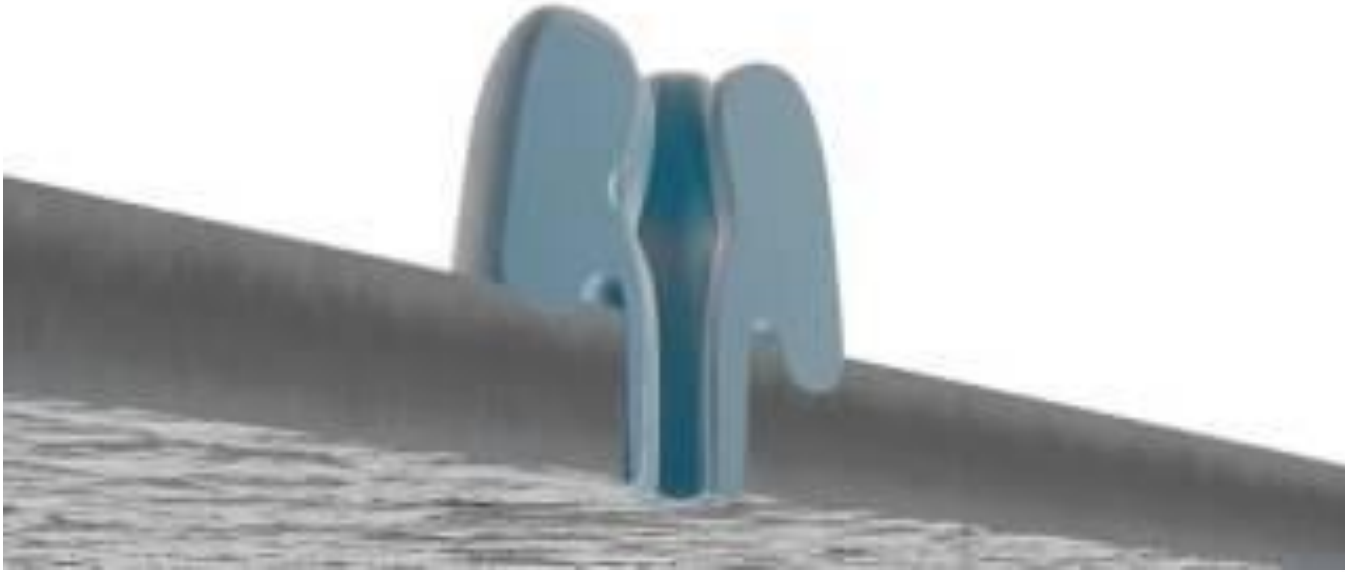


Veterinærinstituttet
Norwegian Veterinary Institute

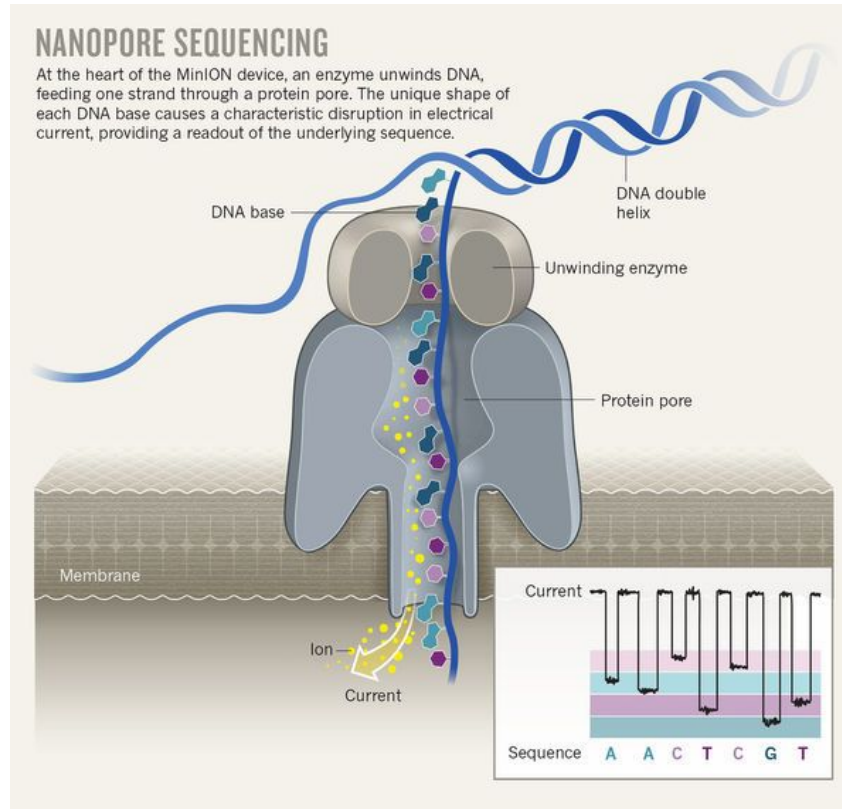
www.nanoporetech.com



Nanopore sequencing explained



Nanopore sequencing explained



DNA

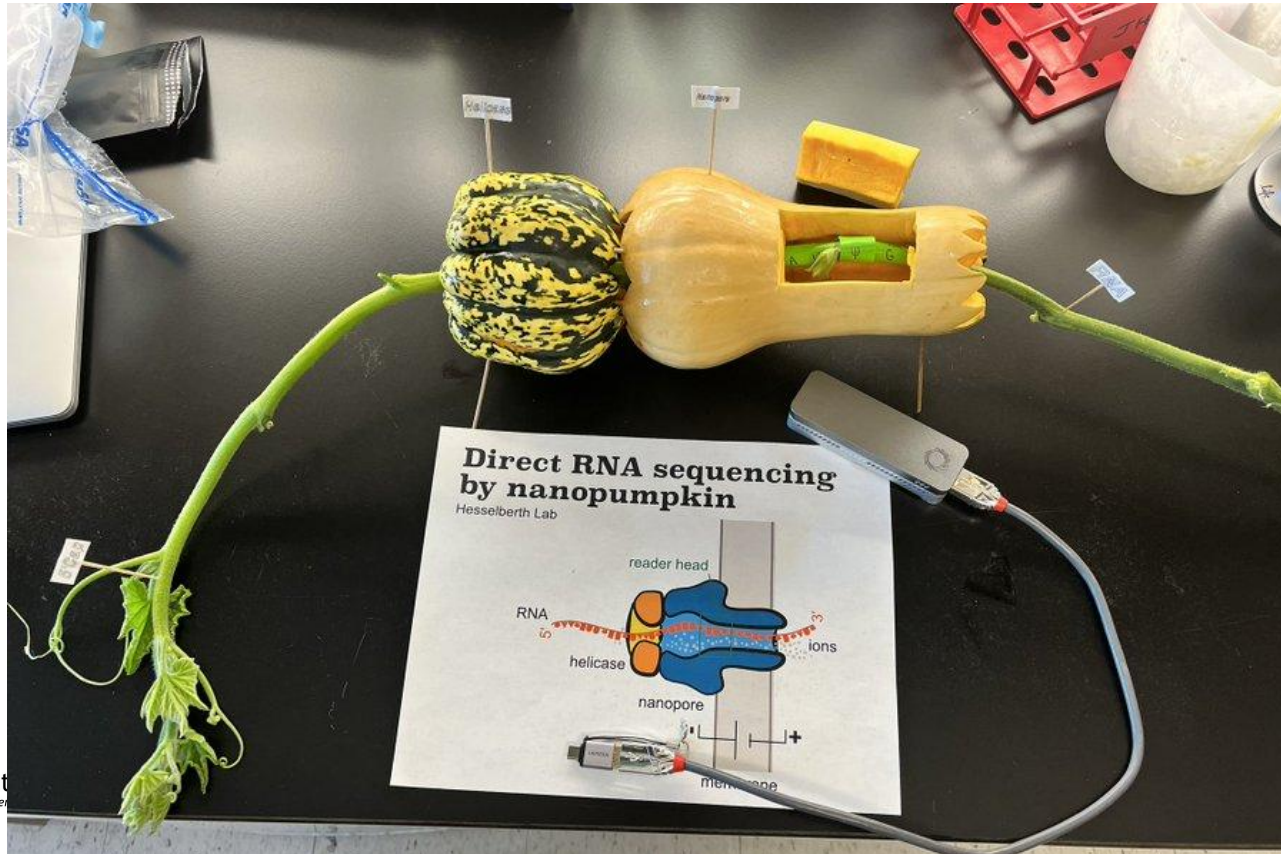


Veterinærinstituttet
Norwegian Veterinary Institute

www.nanoporetech.com



Nanopore sequencing explained



RNA

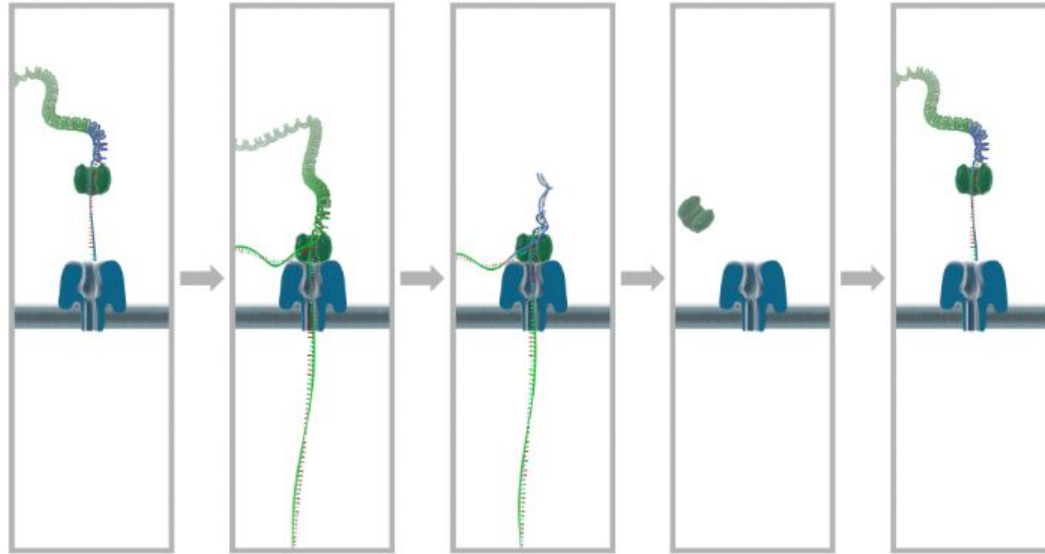
by
Jay Hesselberth



Veterinærinstituttet
Norwegian Veterinary Institute

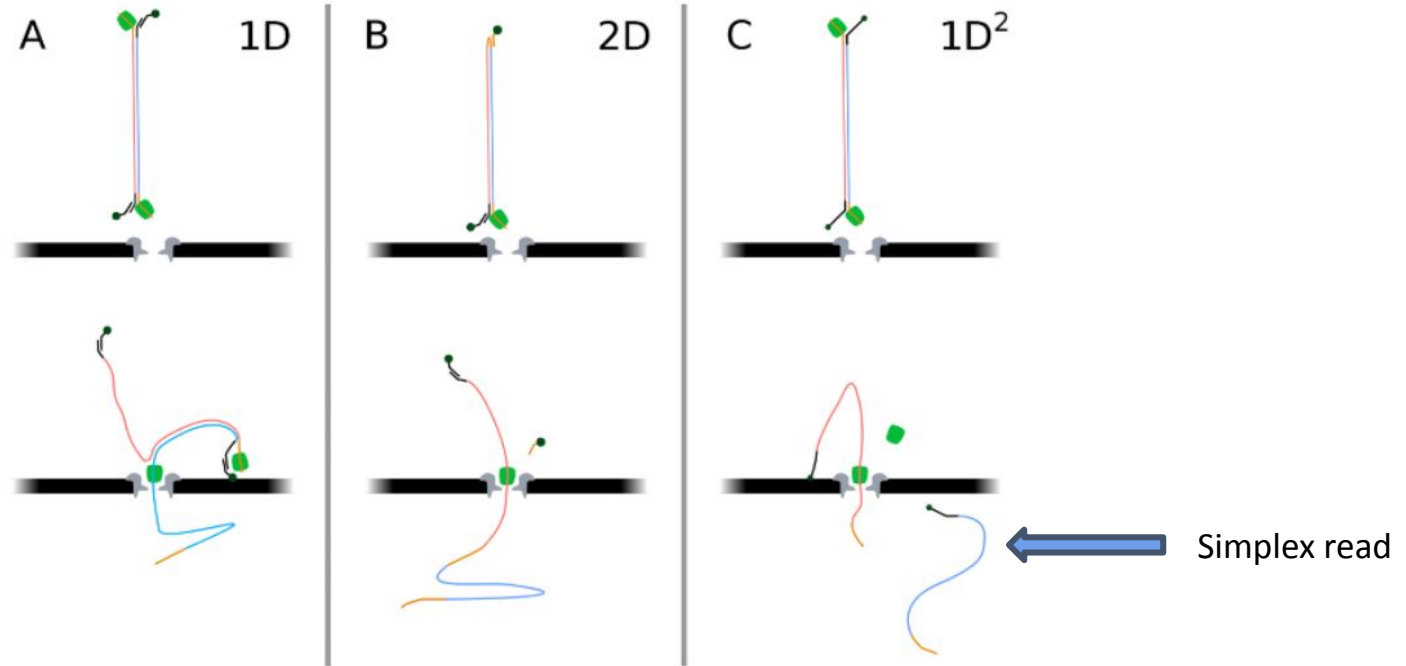


Nanopore sequencing



- The electric potential over the membrane pulls the DNA toward the nanopore.
- The motor protein regulates the speed of sequencing ($\approx 400 \text{ bases s}^{-1}$).
- Current changes are measured when a base is pulled through the pore.

1D vs 2D sequencing



Note: 2D sequencing is no longer available. 1D² is now the standard.



Veterinærinstituttet
Norwegian Veterinary Institute

De Lannoy et al. - DOI: 10.12688/f1000research.12012.1



Simplex vs Duplex reads

Simplex reads: Single strand of a DNA molecule

Duplex reads: two strand of the “same” DNA molecule

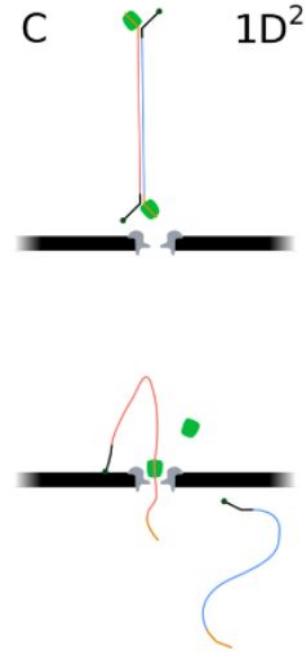
Algorithm:

Sequencing of two single strands

—> at the same pore

—> with little time between them

—> Create a consensus sequence of both reads, with higher quality scores.

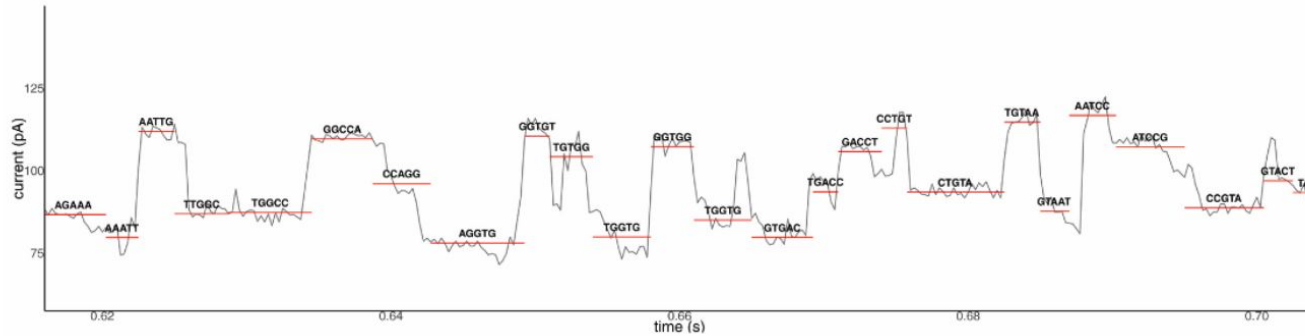


About 5 -10 % of all single reads can be used for this.

Veterinærinstituttet
Norwegian Veterinary Institute

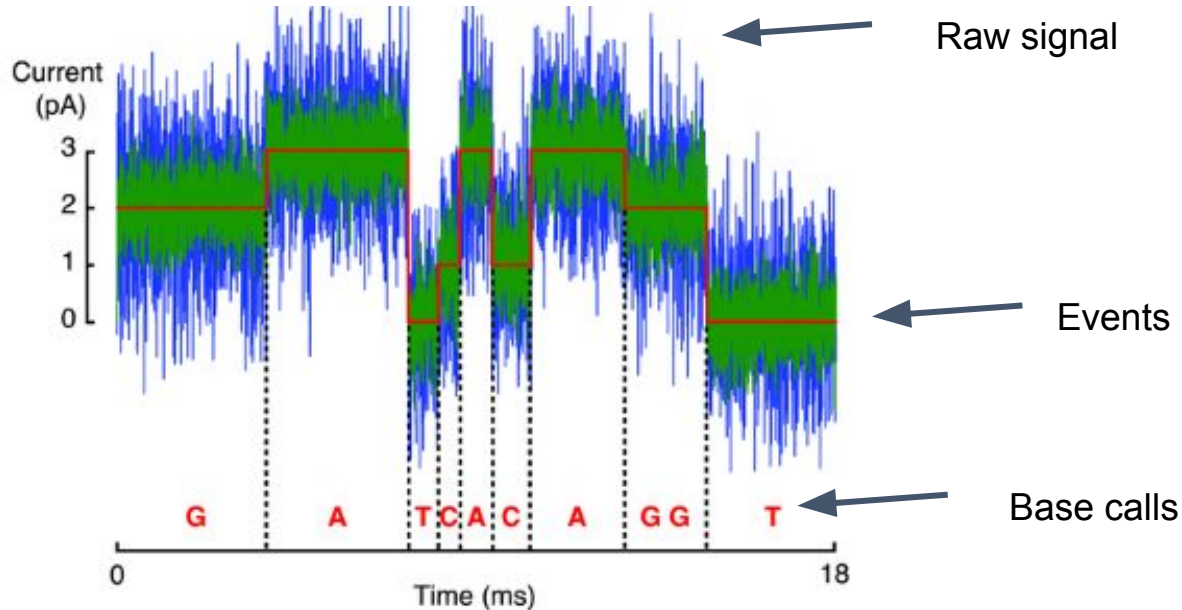


Nanopore basecalling



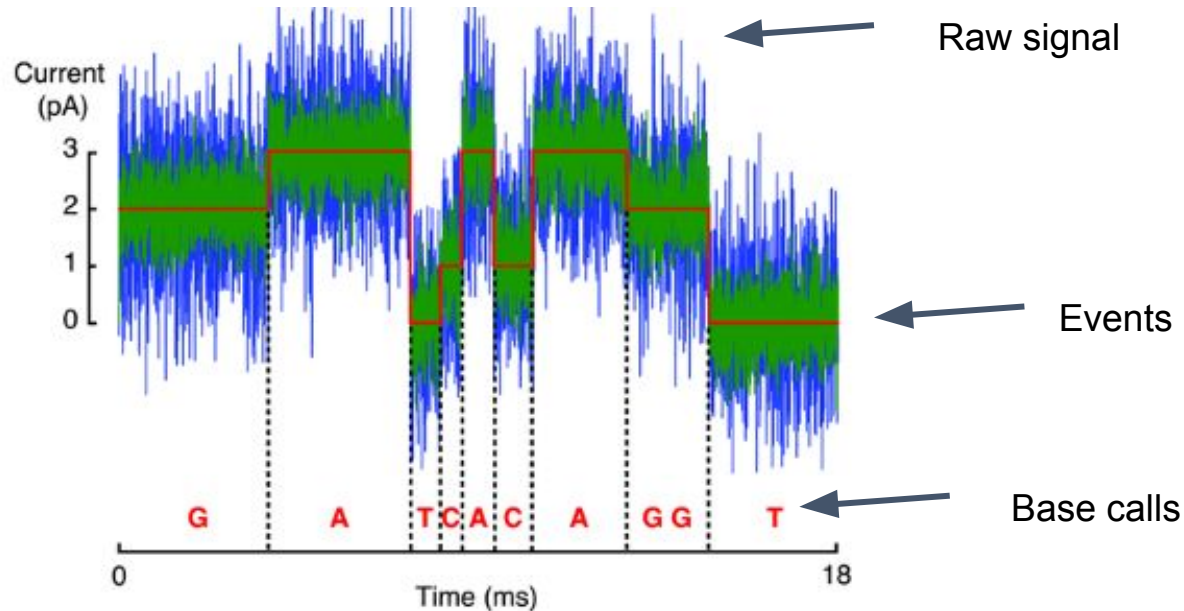
- The length of the passage (pore) determines the signal
- The assumption was that 5 bases fitted in the pore.
- Newer basecaller software dropped assumption and derive basecalls directly from the signal

Variation in basecalling



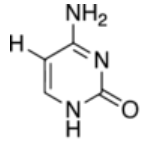
- Translocation time through the pore time is variable
- Depending on the surrounding sequence
- Basecallers need advanced algorithms to deal with this “noisy data”.

Improving basecalling

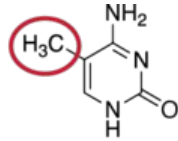


- Addition of Lambda DNA might improve basecalling per run.
- But the software needs to be able to use that information

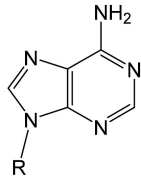
DNA methylation



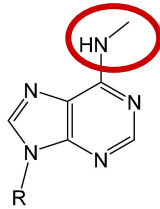
Cytosine



methyated Cytosine



Adenine (A)



N⁶-methyladenine (m⁶A)

Methylated nucleotides.

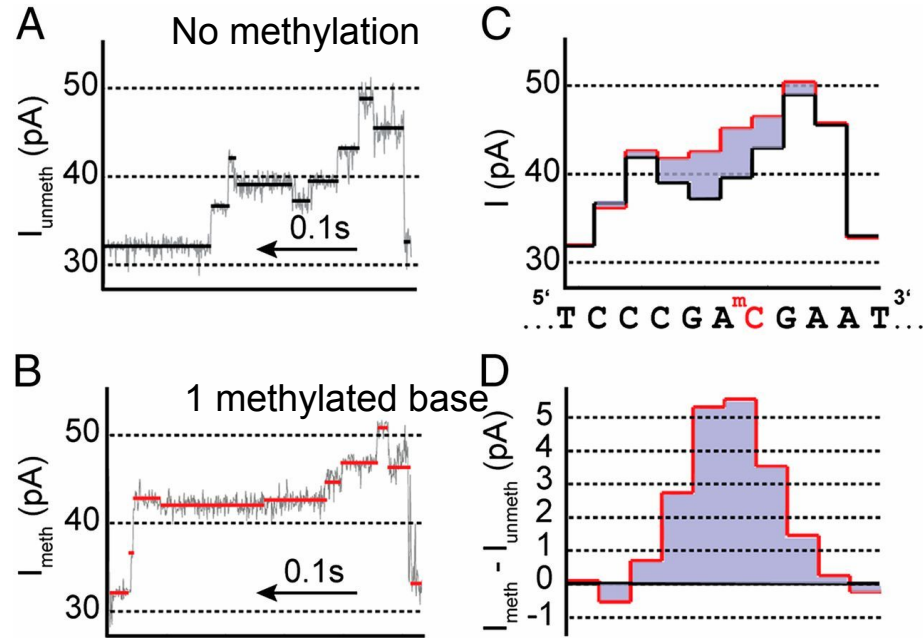
Methylation in Eukaryotes needed for:

- Gene regulation
- Cell differentiation
- Silencing of mobile elements

Methylation in Prokaryotes:

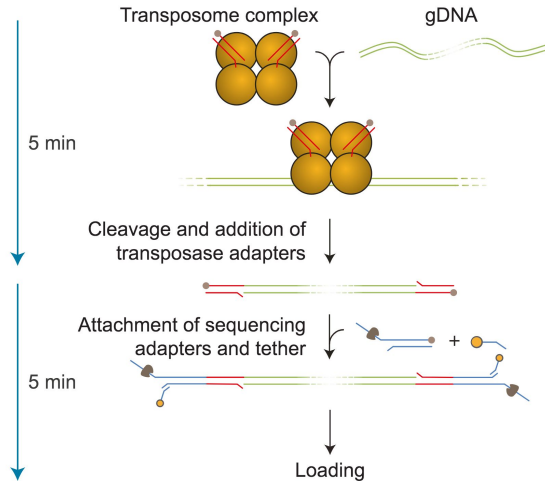
- Silencing of mobile elements
- Phages recognized
- Gene regulation

Detecting methylation



Methylation changes the detected current

Sequencing library preparation - DNA



Rapid Barcoding Kit protocol

- Input: 200ng HMW DNA
- Typical output:
 - 1-2 Gb in 6 hrs
 - 4-8 Gb in 48 hrs
- Enzymatic Shearing of DNA
→ 40-60 % GC required

A very quick library preparation is possible

Sequencing output

Sequencing E.coli K-12 MG1655

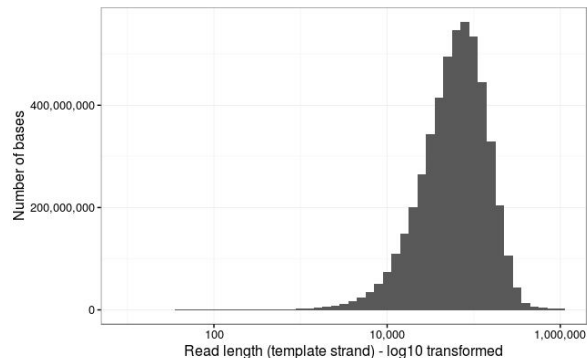
minION output

Total bases: 5.014.576.373 (5Gb)

Number of reads: 150.604

N50: 63.747

Mean length: 33.296,44



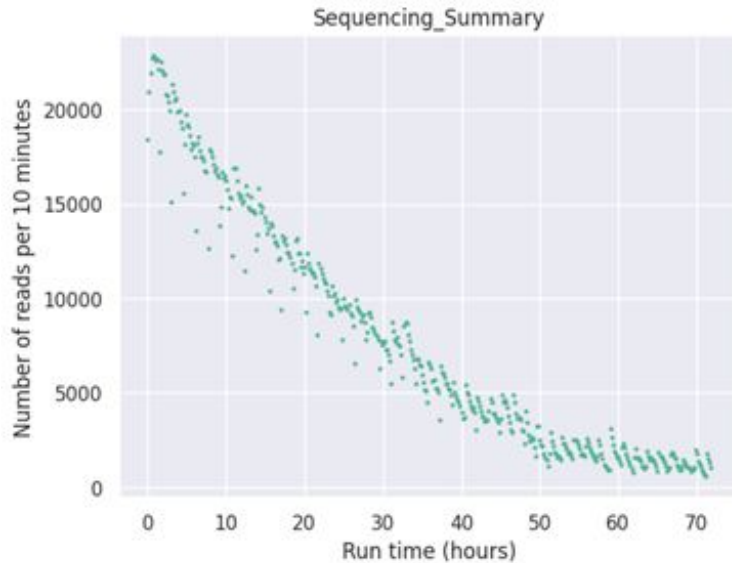
Longest alignable sequence: 2,272,000 bp (2018)

> 4 million bp

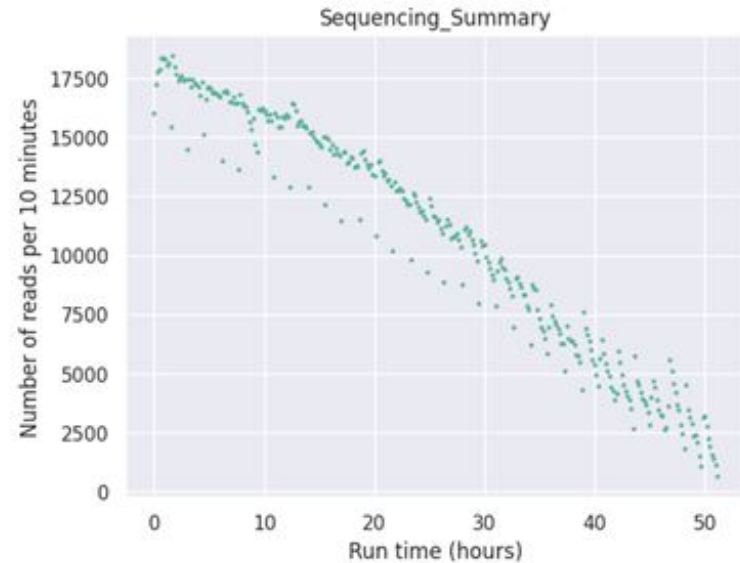
Possible due to very careful phenol / chloroform extractions
with very pure DNA (260/280 $\approx$ 2.0).!!!

The minION output - reads

run_20220831

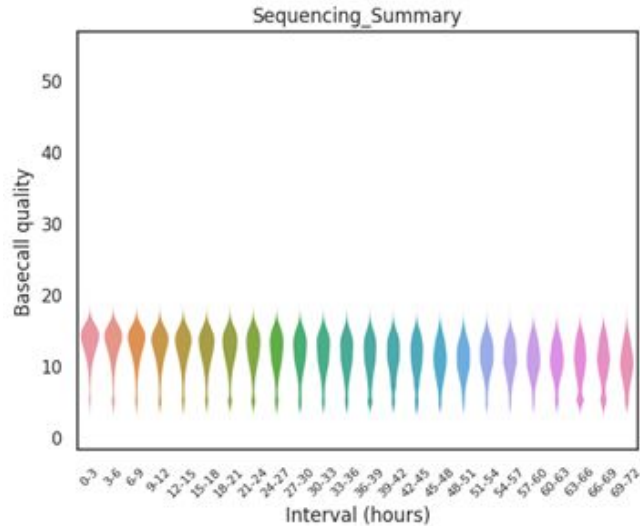


run_20220920



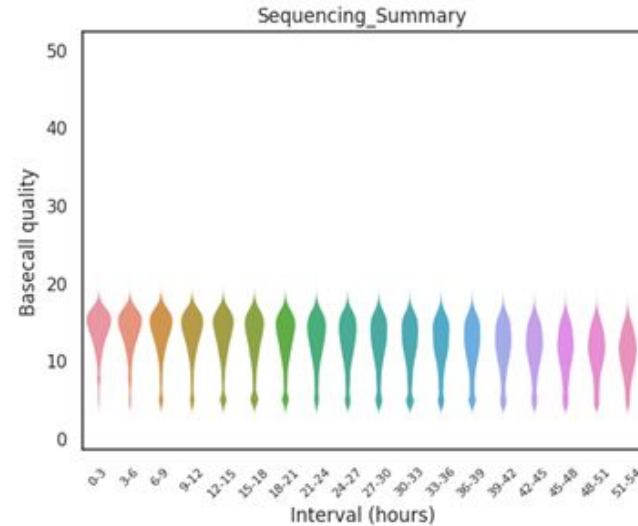
The minION output - quality

run_20220831



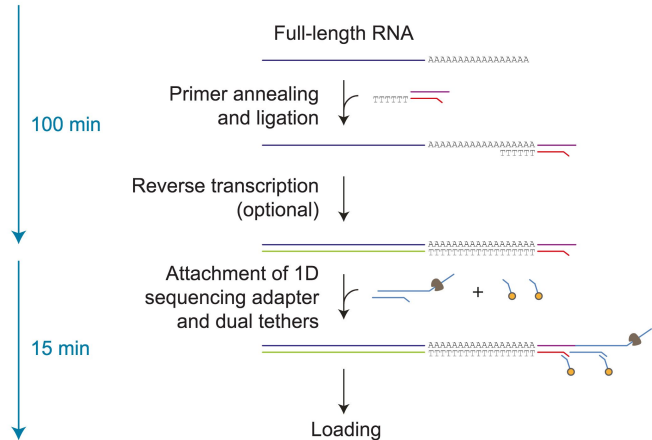
Highest quality
54.2 (696)

run_20220920



Highest quality
49.9 (230)

Sequencing library preparation - RNA



Direct RNA sequencing

- Poly-A tail needed
- Optional reverse transcriptase to make cDNA → improves output
- Input : 500 ng RNA
- Typical output:
 - < 1 Gb in 6 hrs
 - 1-4 Gb in 48 hrs

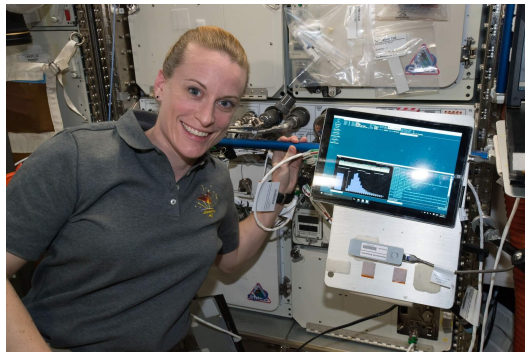
RNA is very easily degraded.

With this “quick” protocol direct sequencing is possible !

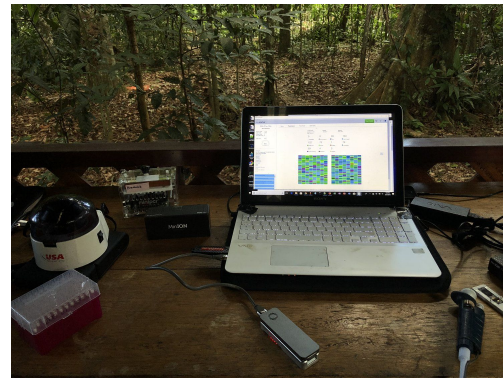
Out of the lab usage...



Antarctica



ISS spacestation



The Jungle

You do need lab equipment to process your samples !!!

MinION applications

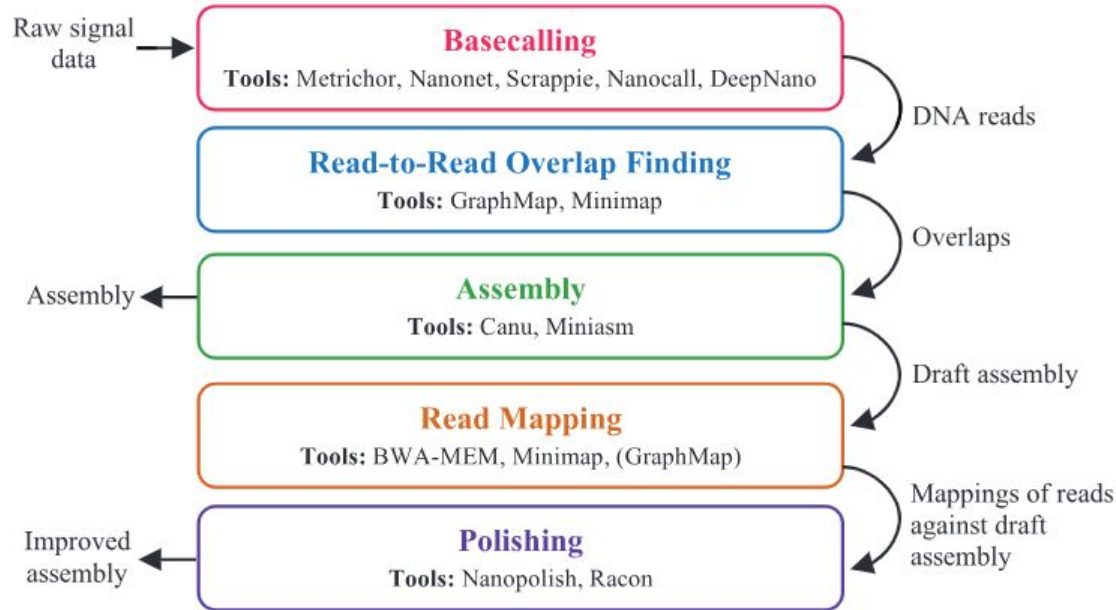
- *De novo* shotgun sequencing (pcr / primer free sequencing)
 - Especially good for repetitive regions
 - Finishing Prokaryote / Eukaryote genomes
 - Detection of structural genome variation (indels)

MinION applications

- Amplicon sequencing
 - Prokaryotes / Eukaryotes: 16S rRNA / 18S rRNA
 - Fungi: ITS-1
 - Animal barcoding: CO1
- Shotgun metagenomics
- Transcriptomics / Direct RNA sequencing
 - Detection of RNA isoforms
- Epigenome (methylation) sequencing
- Direct Protein sequencing !!!



De novo genome assembly



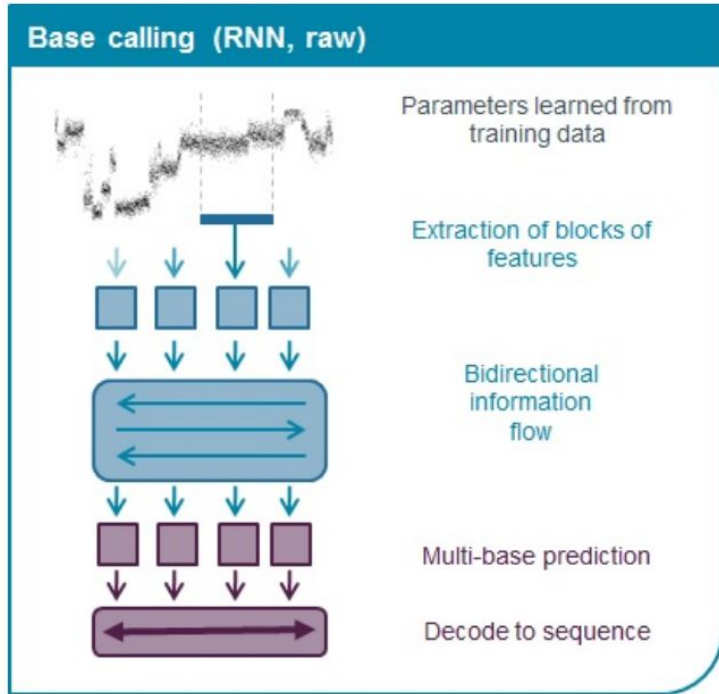
Basecalling software

Many options available:

- Nanopore provides several basecallers
 - MINKnow (Included in the sequencing software)
 - Guppy
 - Bonito
 - **Dorado (current standard)**
- Other groups have also made basecallers for the nanopore machines:
 - Metrichor (In the cloud basecaller, part of minION workflow)
 - Chiron
 - DeepNano
 - etc



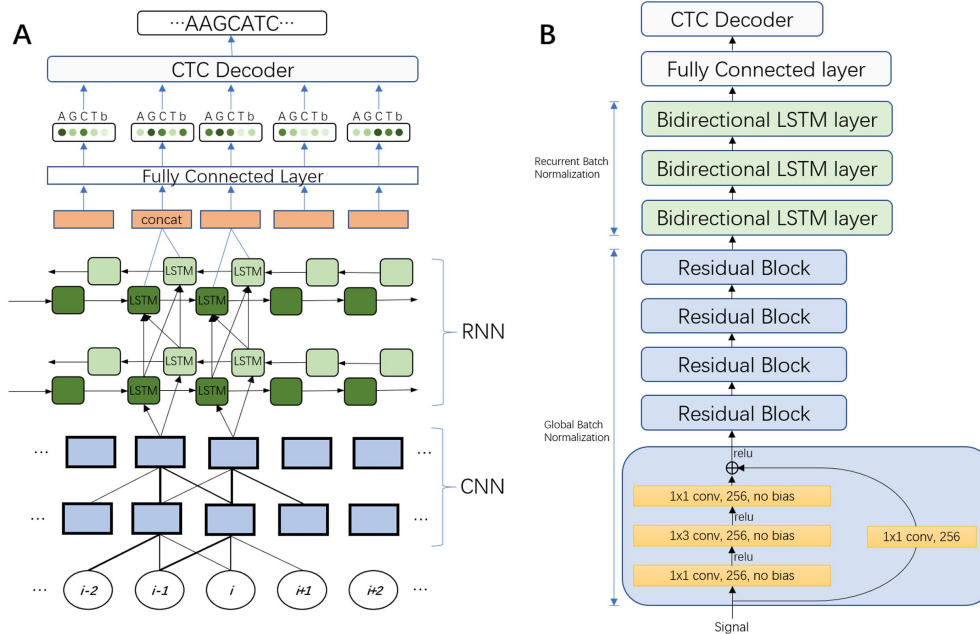
Nanopore basecalling



Original basecallers used Hidden Markov Models

Latest basecallers use Recurrent Neural Network (RRN)

Basecalling software - Chiron



A combined convolutional neural network and a Recurrent Neural Network



Genome assembly

Table 1. Summary of comparisons between long read assemblers. (A) Selected metrics for three benchmarking efforts on MinION reads, including chemistries used in the respective studies. Bold values denote the best score per metric. (B) Short descriptions and reference papers for all assemblers discussed in this paper. †: reads were corrected by Canu prior to assembly.

A	Judge <i>et al.</i> ⁴¹			Istace <i>et al.</i> ⁴⁰			Giordano <i>et al.</i> ³⁹		
	subs/kbase	indels/kbase	N50 (Mbase)	subs/kbase	indels/kbase	N50 (Mbase)	subs/kbase	indels/kbase	N50 (Mbase)
PBcR	1.0	12.2	1.20				0.2	17	0.616
Canu	0.3	7.8	2.80	0.105	10.0	0.610	0.1	17	0.698
SMARTdenovo				0.580	11.1	0.783	0.3	14	0.625
Minimap & miniasm	6.7	18.6	6.60	0.207 [†]	13.5 [†]	0.736 [†]	34	67	0.739
ABruijn				0.130	10.1	0.816	0.1	15	0.769
Chemistry		MAP006			MAP005/MAP006			MAP006/007	
Read type		2D			2D			2D	
Pore		R7.3			R7.3			R7.3/R9	
Basecaller		EPI2ME			EPI2ME			EPI2ME	
Organism		<i>Enterobacter kobei</i>			<i>S. cerevisiae</i>			<i>S. cerevisiae</i>	
B	Description								Ref.
PBcR	Celera OLC assembler adapted for long error-prone reads.								42
Canu	The more accurate successor of PBcR.								43
SMARTdenovo	Fast and reasonably accurate assembler without prior error correction step.								Github
Minimap & miniasm	Fast assembly pipeline without error correction and consensus steps.								44
ABruijn	DBG assembler that fuses unique strings prior to assembly, produces highly contiguous assemblies.								45
TULIP	uses seed extension principle to efficiently assemble large genomes.								25
HINGE	Assesses coverage of low complexity regions prior to assembly and processes them more efficiently.								46



Canu assembler

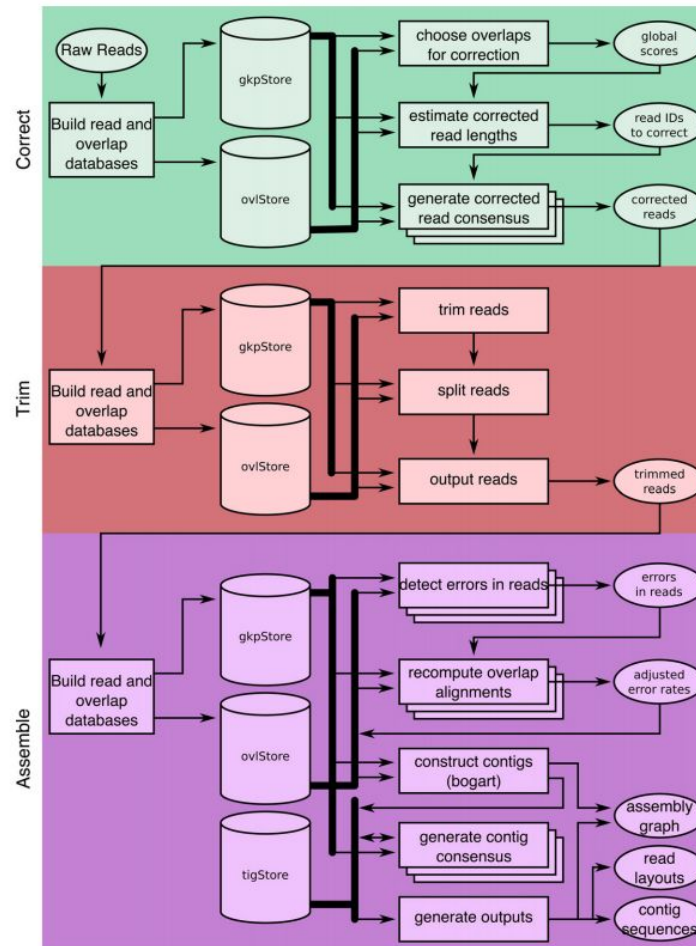
Canu Assembly pipeline

1. Error correction
2. Trimming
3. Assembly

gkpStore: reads database

ovlStora: overlap database

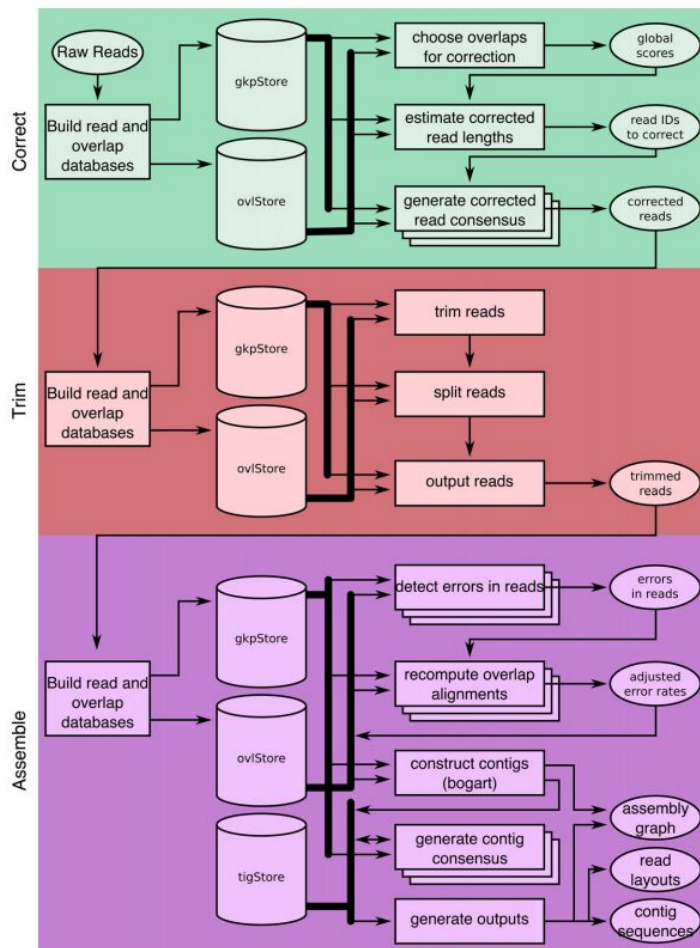
tigStore: contigs database



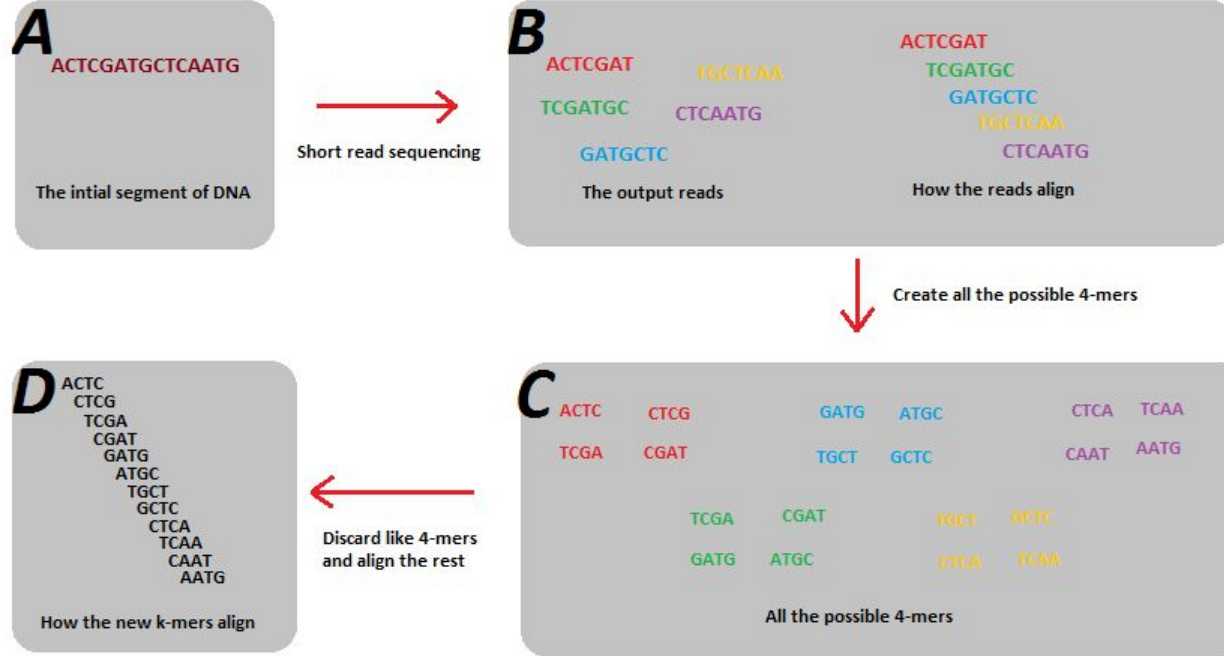
Error correction

1. Reads split into **kmers**
2. Kmers used to identify overlap
3. Correct reads using overlap

Corrected reads are trimmed



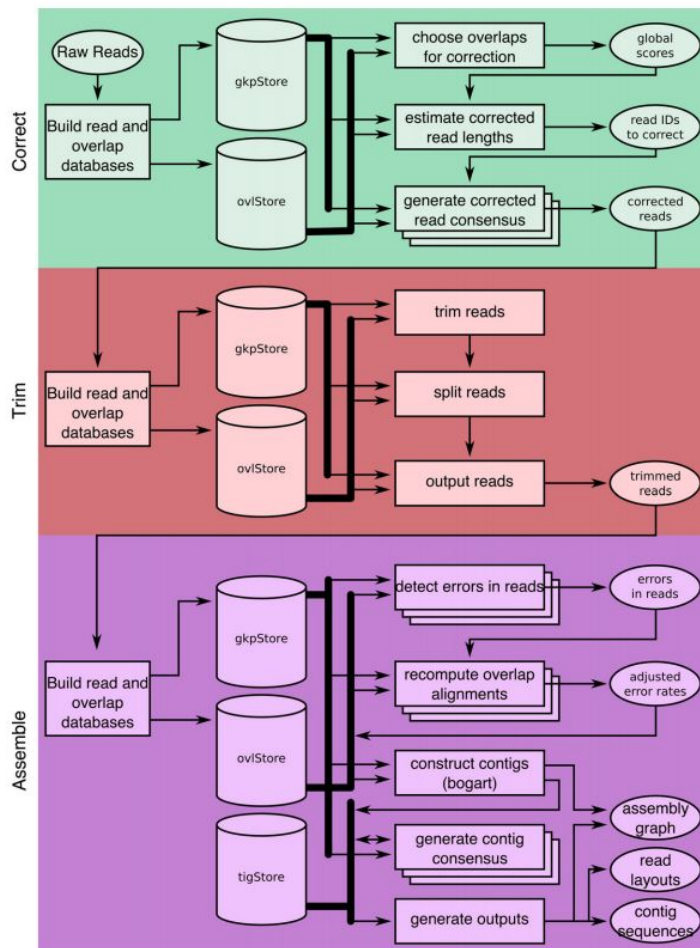
Kmers



Error correction

1. Reads split into kmers
2. Kmers used to identify overlap
3. Correct reads using overlap

Corrected reads are trimmed



Canu assembly *E. coli* genome

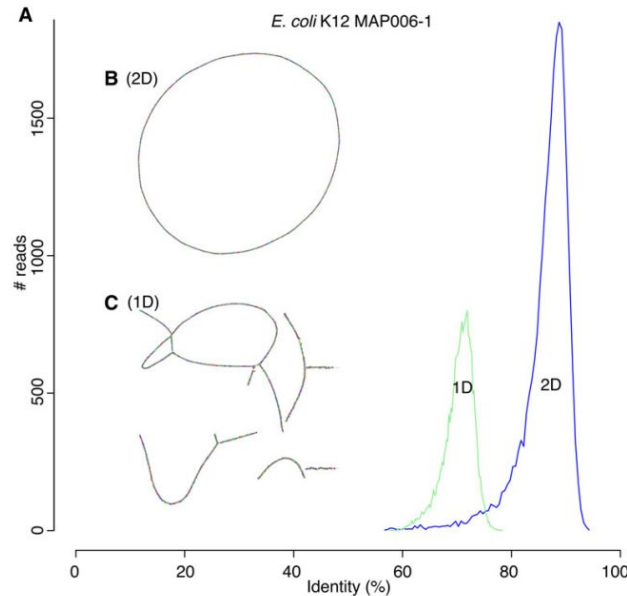


Figure 5. Canu can assemble both 1D and 2D Nanopore *Escherichia coli* reads. (A) A comparison of error rates for 1D and 2D read error rates versus the reference. Template 1D and 2D reads from the MAP006-1 *E. coli* data set were aligned independently to compute an identity for all reads with an alignment >90% of their length (95% of the 2D reads and 86% of the 1D reads had an alignment >90% of their length). The 2D sequences averaged 86% identity, and the 1D reads averaged 70% identity. (B) Bandage plot of the Canu BOG for the 2D data. The genome is in a single circle representing the full chromosome. (C) The corresponding plot for 1D data. While highly continuous, there are multiple components due to missed overlaps and unresolved repeats (due to the higher sequencing error rate).

Polishing

Nanopolish: Improve consensus sequence of assemblies

Options:

- Predict methylated bases
- detect SNPs and indels with respect to a reference genome
- calculate an improved consensus sequence for a draft genome assembly
- align signal-level events to k-mers of a reference genome
 - Align raw sequence data to deal with homopolymers and other hard to analyse sequences



A short comparison

	Illumina	PacBio	minION
Output (Gb)	7.5 – 6000	5-8	10-50
Reads (million)	25 – 20-000	0.15 - 1	≈ 0.15
Read length	150 – 300 bp	0 - 70 Kbp	0 - 800 Kbp
Pros	• Many reads • High quality • Tolerant for poor input material	• Long reads • Improve genome assemblies	• High mobility • Long reads • Improve genome assemblies
Cons	• Fragmented genome assemblies	• High quality input needed • expensive	• High quality input needed • Flowcell has limited shelf life

Experimental design important to decide which platform to use.



The End

A few papers:

The long reads ahead: *de novo* genome assembly using the MinION

- <https://f1000research.com/articles/6-1083/v2>

Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis

- <https://doi.org/10.1186/s13073-015-0220-9>

Nanopore sequencing technology, bioinformatics and applications (2021)

- <https://pubmed.ncbi.nlm.nih.gov/34750572/>

A website with a big list of software for handling Nanopore data in a wide variety of applications:

- <https://github.com/goekelab/awesome-nanopore>



The End

Contact details:

Thomas Haverkamp

Thomas.haverkamp@vetinst.no

bluesky: @thomieh73.bsky.social



Veterinærinstituttet
Norwegian Veterinary Institute

