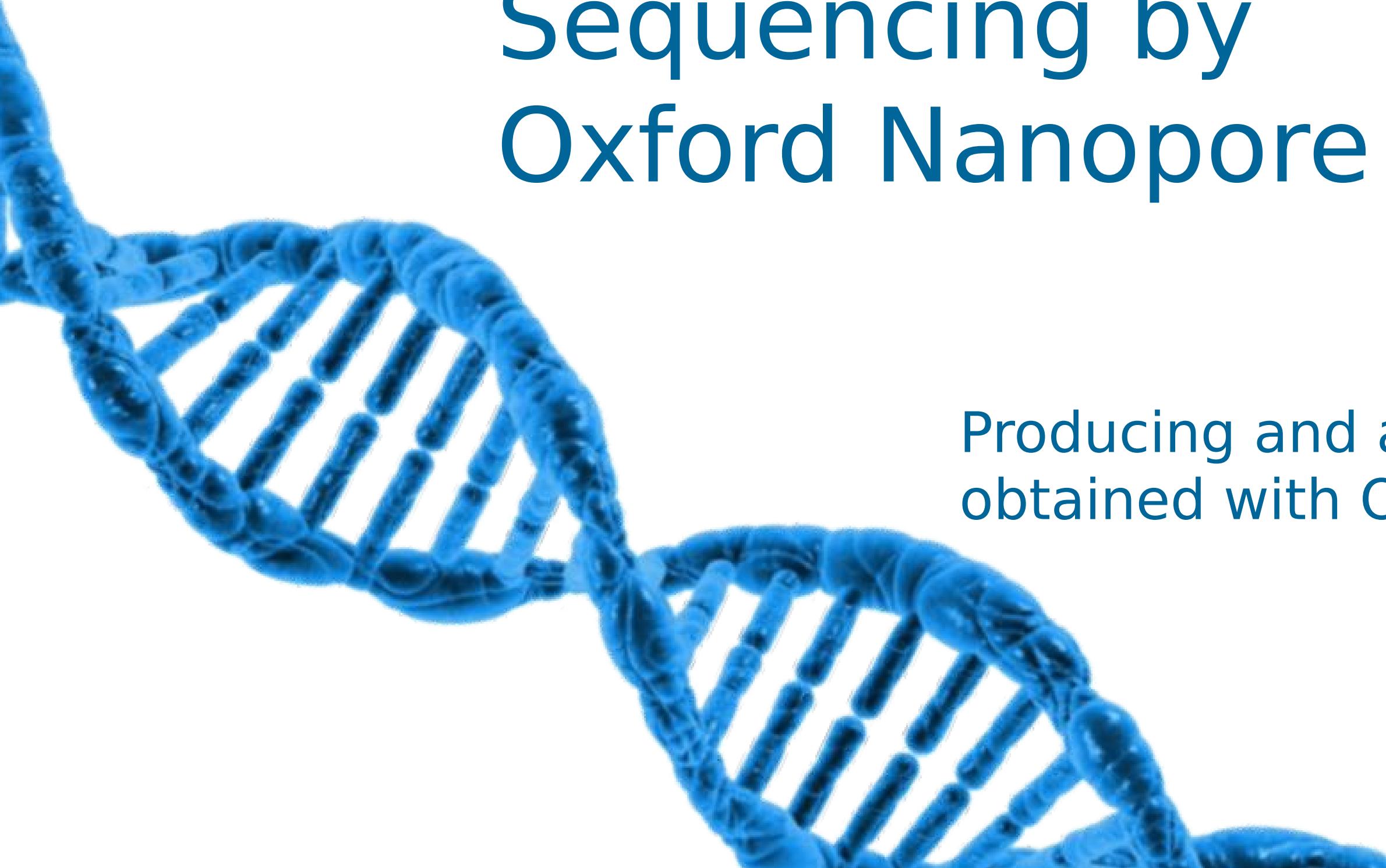


# Sequencing by Oxford Nanopore Technology



Producing and analyzing genomic data  
obtained with ONT

November, 2023

# Timeline QUITO

## Laboratory

- Reminder of sequencing principle

- ✓ Sequencing principles
- ✓ Applications
- ✓ ONT, equipment
- ✓ Lab, equipment
- ✓ extraction protocol explained

Thursday  
9/11

- HMW DNA extraction

- ✓ HMW DNA extraction (laboratory)
- ✓ DNA qualification

- Preparation Librairies

- ✓ Presentation of MK1C, LSK110, NEB, Ampure XP
- ✓ *Preparation of 2 libraries (laboratory)*
- ✓ Launching run

Friday  
10/11

- ✓ *Preparation of librairies by participants (laboratory)*
- ✓ *Loading on blank flowcell (laboratory)*
- ✓ *Priming flowcell (laboratory)*
- ✓ Qualitative analysis

- Data results

- ✓ Preliminary analysis of the sequencing RUN
- ✓ Discussion on graphics
- ✓ Applications

Monday  
13/11

- ✓ Critical parameters for run
- ✓ Sequencing action possibilities
- ✓ Nanoplot
- ✓ Students Talks

## Bioinformatique

- ✓ NGS course and application
- ✓ Basecalling
- ✓ Demultiplexing
- ✓ Polishing/correction
- ✓ Assembly quality

Tuesday  
14/11

- ✓ Read Quality analysis
- ✓ Assembly

Wednesday  
15/11

- ✓ Comparative genomics

- ✓ SV detection through mapping

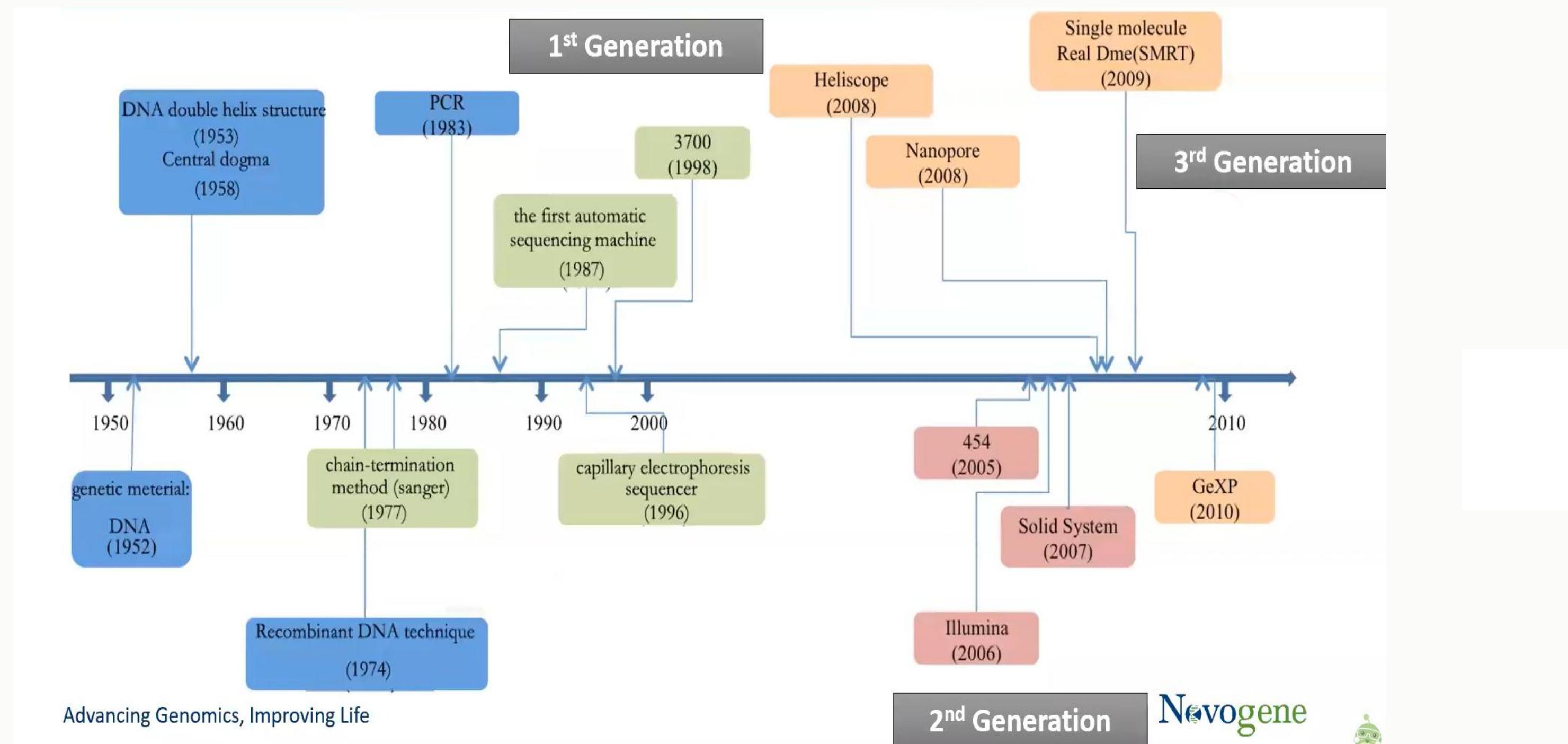
Thursday  
15/11

- ✓ Discussion

## Sequencing principles

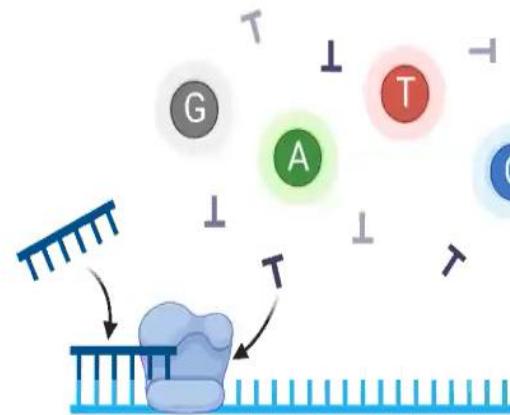


# History of sequencing

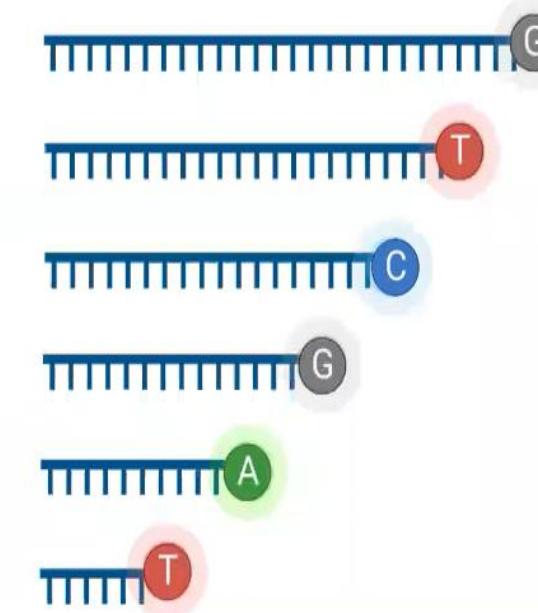


# Methodology of first generation : sanger

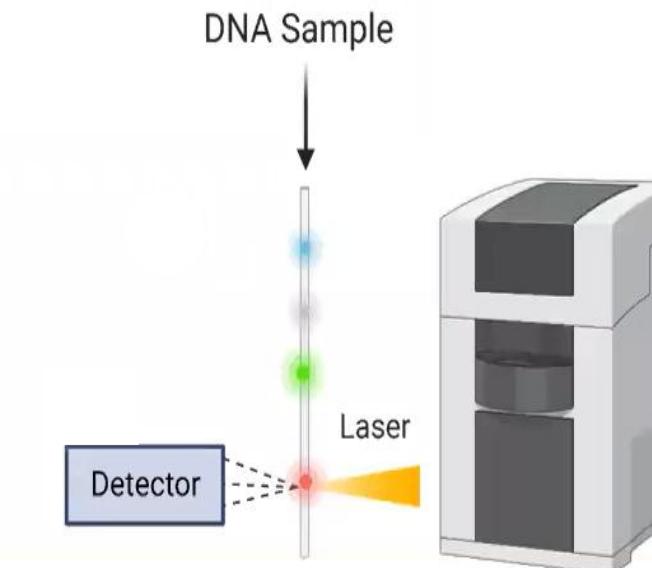
① Primer annealing and chain extension



② Fluorescently labelled DNA sample



③ Capillary gel electrophoresis and fluorescence detection



## • Phase 1

Annealing of specific primer

## • Phase 2

Copy of dna with dNTPs et ddNTPs

## • Phase 3

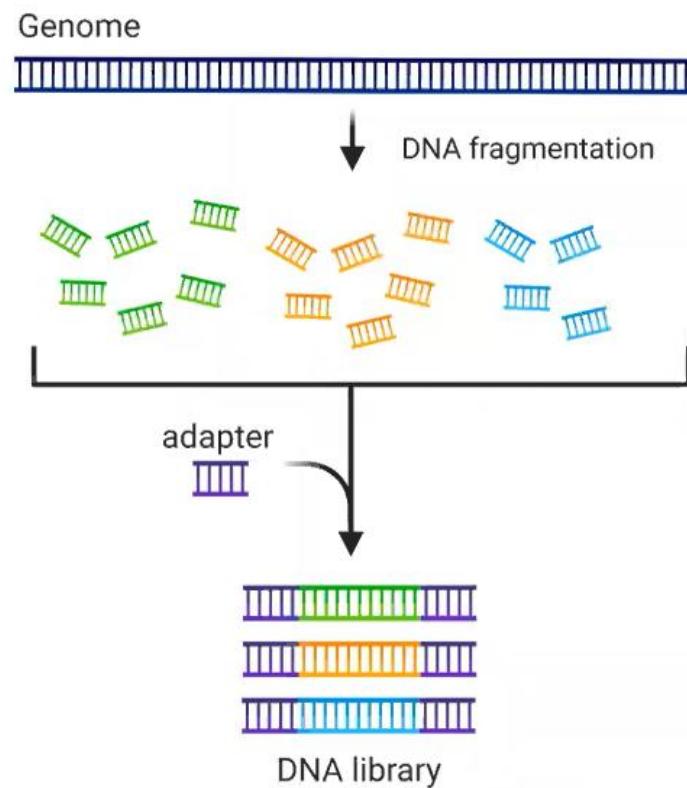
fluorescence reading by capillary sequencer

## • Results

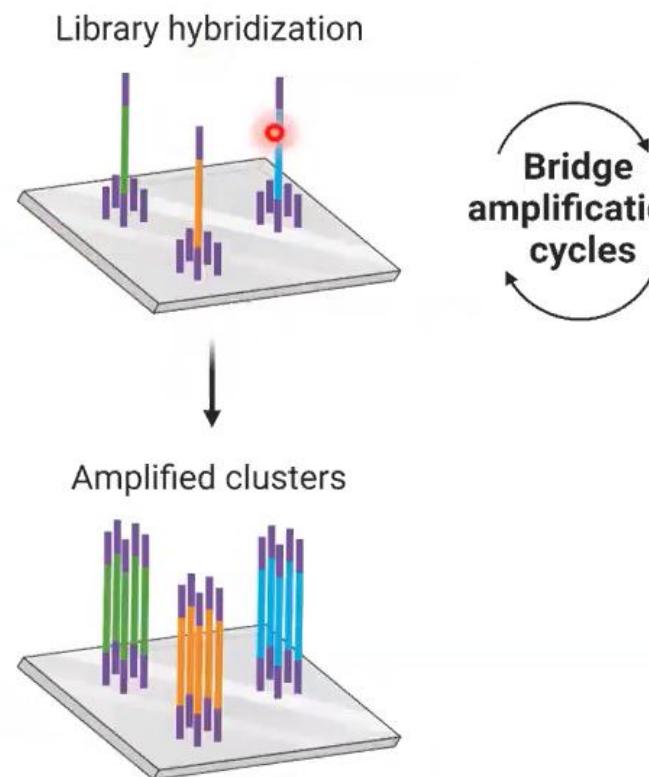
- ✓ High accuracy
- ✓ Simple data analysis
- ✓ Low error rate < 1%
- ✓ Limited length of readings <1kb
- ✓ High Cost per Base

# Methodology of 2nd generation : illumina

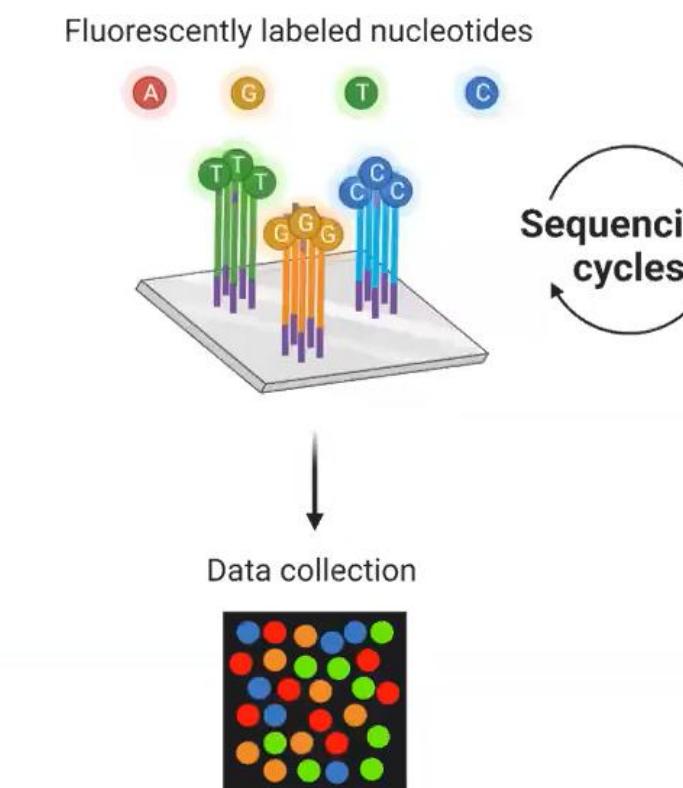
## ① Library preparation



## ② DNA library bridge amplification



## ③ DNA library sequencing

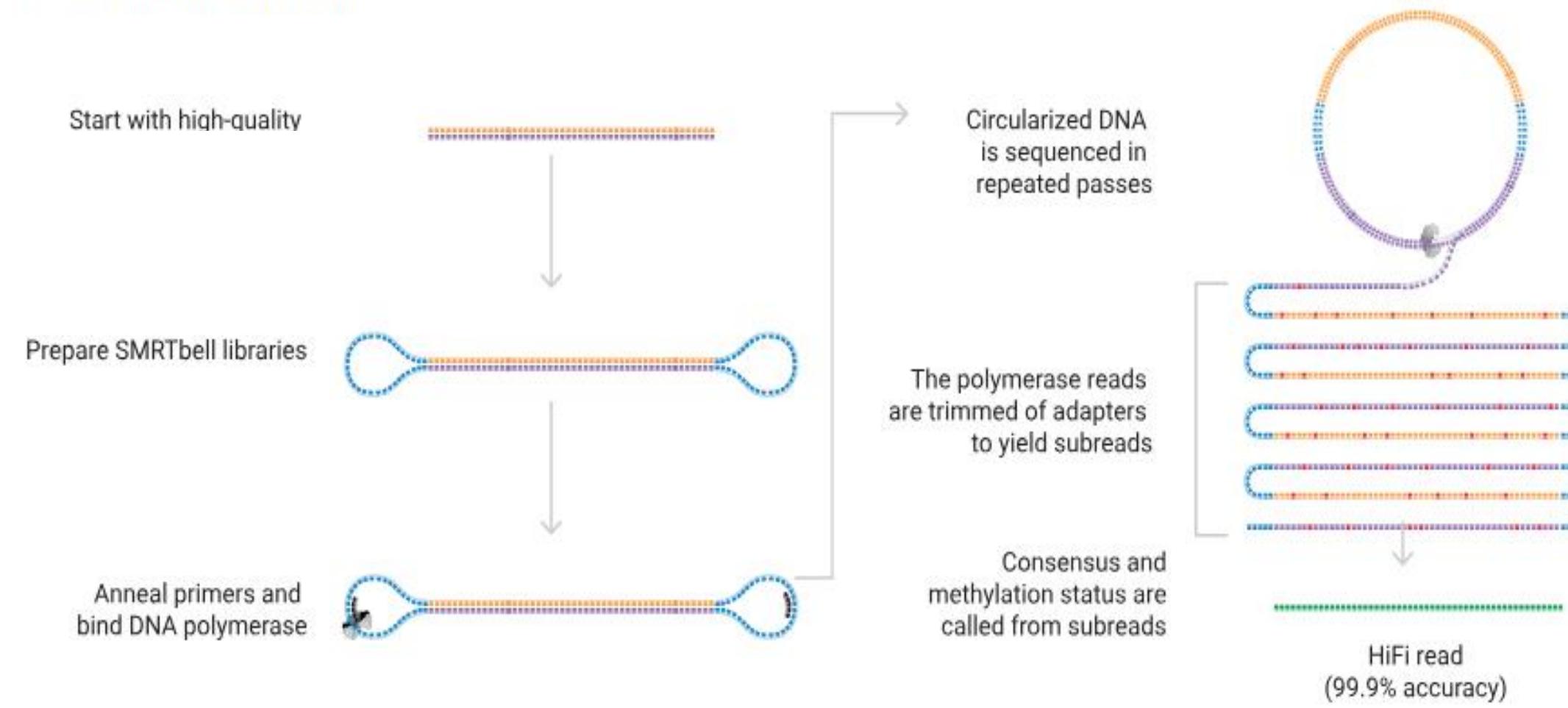


## • Results

- ✓ High accuracy
- ✓ Low error rate < 1%
- ✓ Multiplexing
- ✓ Limited length of readings 150 pb
- ✓ Complex data analysis
- ✓ Cost 9\$/Gb

# Methodology of 3rd generation : PacBio HiFi

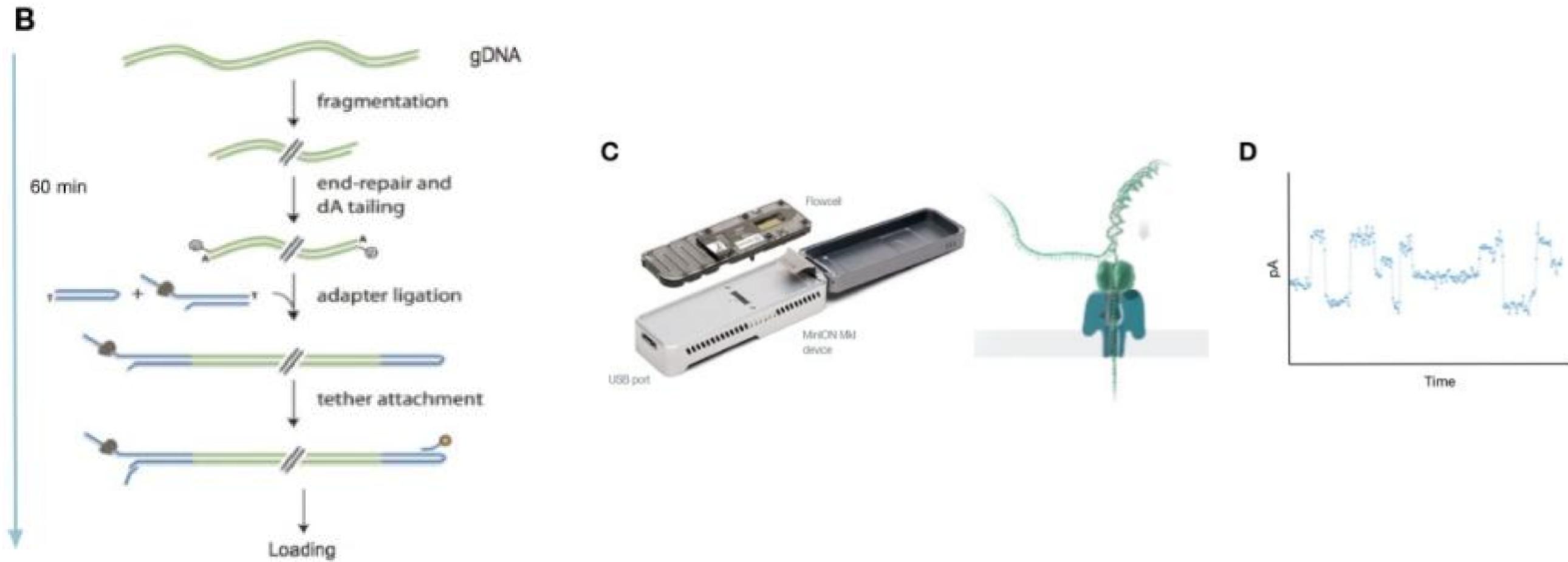
How are HiFi reads generated?



## • Results

- ✓ Long read : 10 kb
- ✓ Low error rate < 1%
- ✓ Multiplexing
- ✓ Complex data analysis
- ✓ Cost 13-100\$/Gb

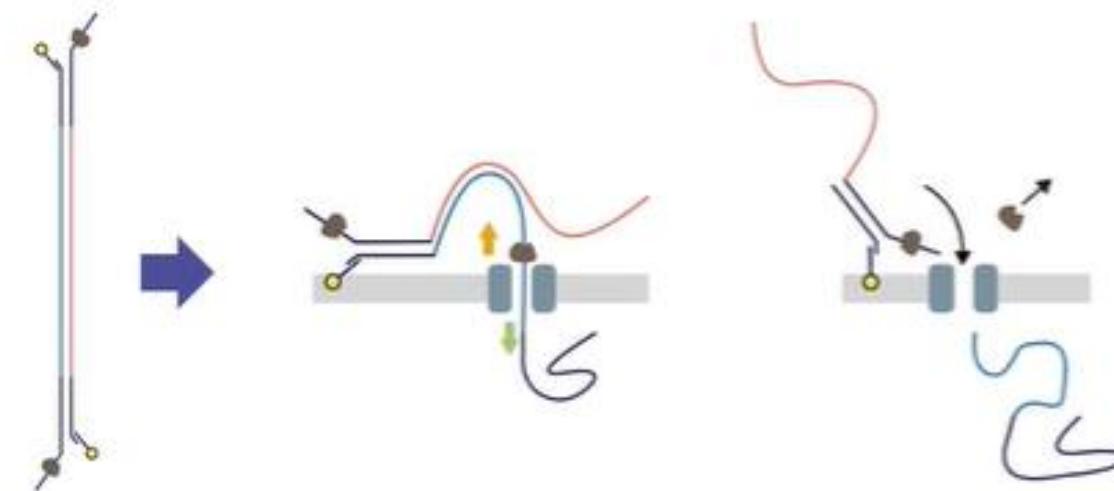
# Methodology of 3rd generation : ONT



## • Results

- ✓ Long read : N50 > 25 kb, max 4Mb
- ✓ Error rate < 3%
- ✓ No matrix amplification
- ✓ Complex data analysis
- ✓ Cost 8-17\$/Gb

# Methodology of 3rd generation : ONT - Q20+



Latest updates to nanopore sequencing achieve:

Flow cell	Kit	Sequencing & basecalling parameters	Sample	Raw read accuracy	Output
R10.4.1	Ligation Sequencing Kit V14	400 bps, 5 kHz, HAC basecalling	Human HG002	99.0% (Q20)	●●●
R10.4.1	Ligation Sequencing Kit V14	400 bps, 5 kHz, SUP basecalling	Human HG002	99.5% (Q23)	●●●
R10.4.1	Ligation Sequencing Kit V14	400 bps, 5 kHz, Duplex basecalling	Human HG002	>99.9% (Q30)	●

## • Results

- ✓ Long read : up to 4Mb
- ✓ Increased sensitivity
- ✓ Medium throughput
- ✓ Complex data analysis
- ✓ Error rate : < 1%

- Short read type  
illumina

Short reads 0.1-0.6 Kb  
DNA fragmentation (short)  
Matrix amplification  
Limited error rate : ~ 0.5%  
High throughput : 800 Gb (3Tb for S4)  
Cost: 9 €/Gb

- Long read type  
ONT

Long reads max 1 Mb  
DNA HMW  
NO Matrix amplification  
ONT error rate : <1 % (kit14)  
Medium throughput 10-300Gb (9Tb for Promethion)  
Cost ONT : 8 - 17 \$/Gb ([available for : 1000 \\$](#))

Human genome :  
2006 : \$20-25 million  
2021 : \$1000

# Applications



# Which applications for DNA sequencing

1

## Whole Genome Sequencing : WGS

Sequencing the entire genome of an organism

2

## Whole Exome sequencing : WES

Sequencing all the protein-coding region of genes in a genome

3

## Target Region Sequencing : TRS

Sequencing a selected set of genes or regions

4

## Metagenomics

Study of microbial communities in their habitats

5

## Epigenetic

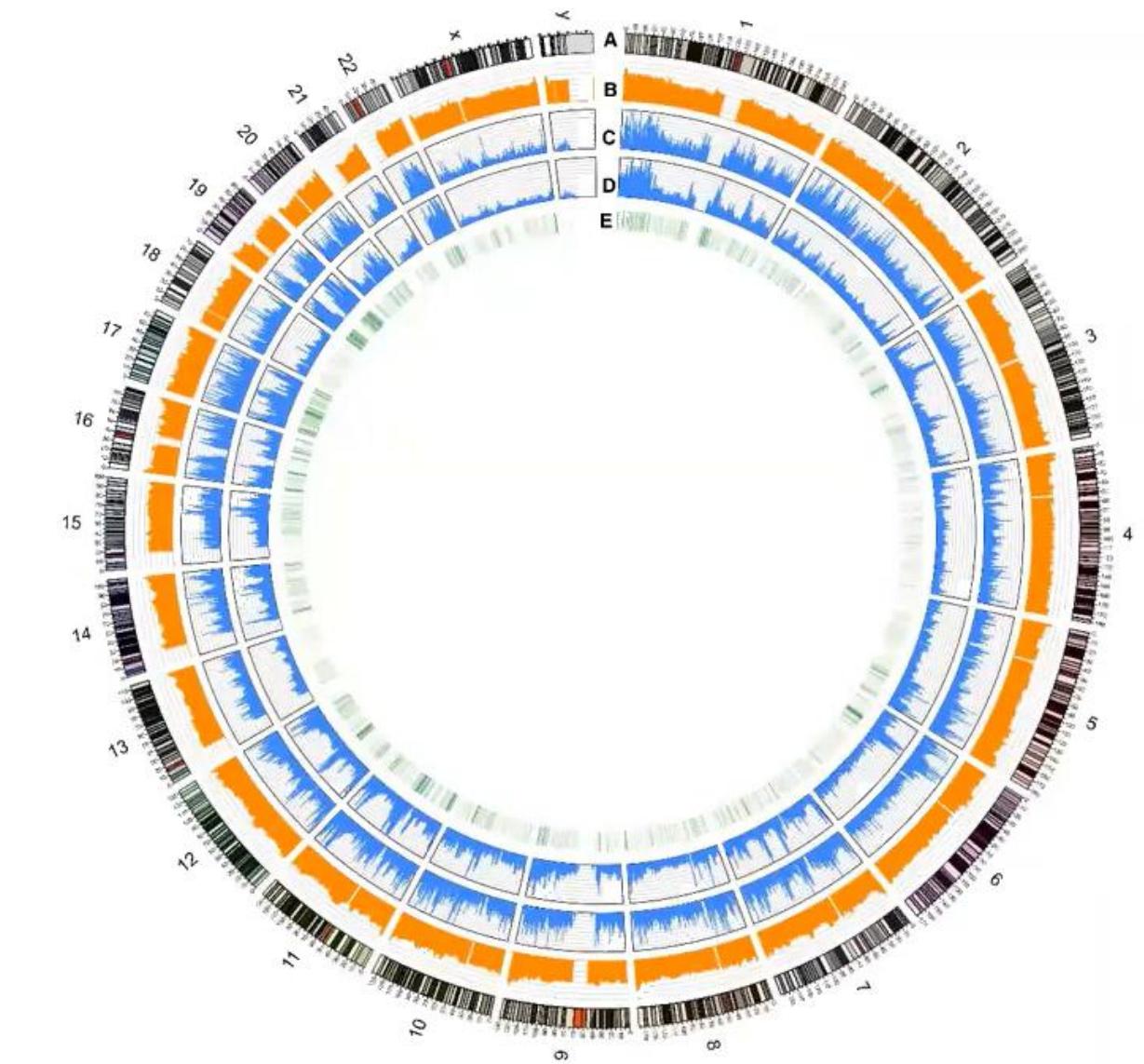
Study of stable phenotypic changes that do not involve alteration in the DNA sequence

# Which applications for DNA sequencing

## 1

### Whole genome sequencing

- **Whole Genome Sequencing (WGS)** – is a technique for sequencing the entire genome of an organism at a single time
- WGS provides a **deep insight** into the DNA sequence of humans, animals, plants, and microbial genomes
- WGS enables the identification and analysis of various **variants** of the genome at a single-base resolution

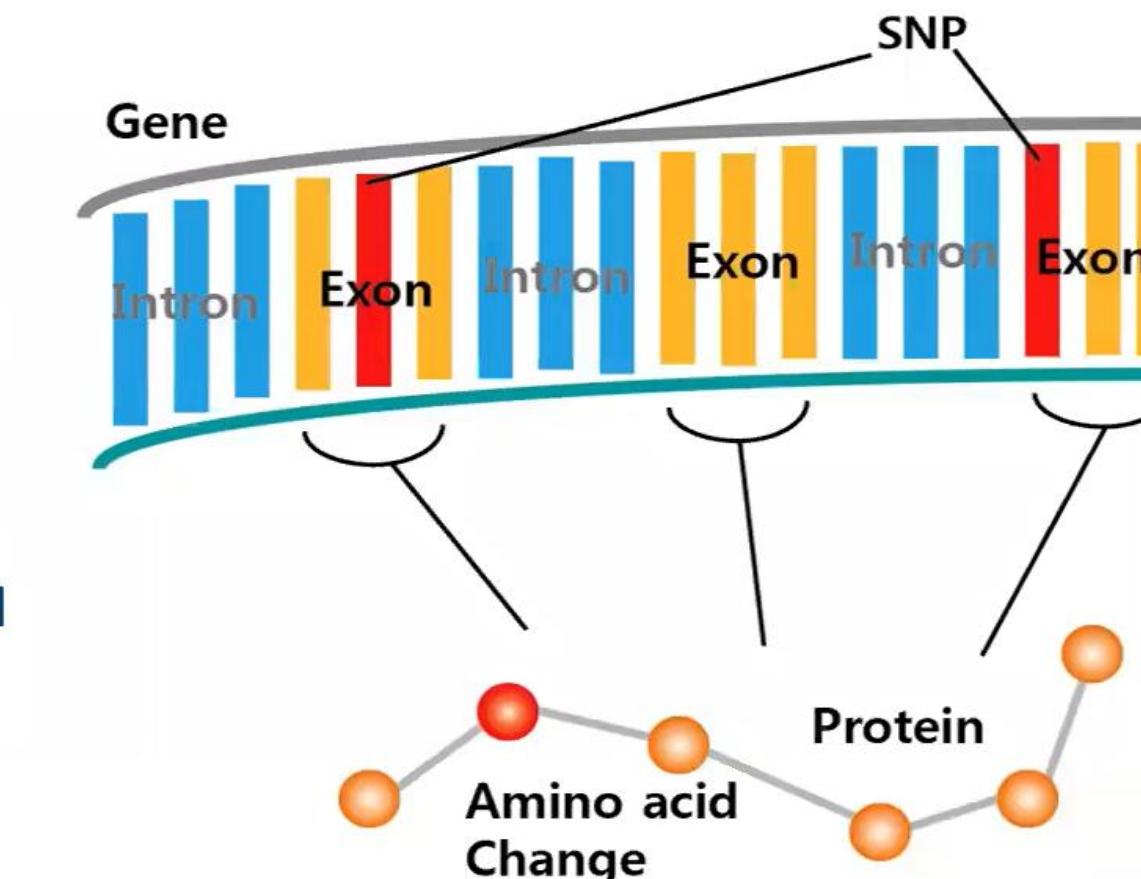


# Which applications for DNA sequencing

## 2

### Whole Exome Sequencing WES

- **Whole Exome Sequencing (WES)** – is a technique for sequencing all of the protein-coding regions of genes in a genome (known as the exome)
- The exome represents around **1%** of the human genome, but contains **~85%** of known disease-related variants
- WES aims is to identify **genetic variants** that alter protein sequences, including those responsible for Mendelian and polygenic diseases, such as Alzheimer's disease
- WES is **cost-effective** compared to WGS, and can be applied in both academic research and clinical diagnostics



## 3

## Target Region sequencing TRS

- **Target Region Sequencing (TRS)** – is a technique for sequencing a **selected set** of genes or regions that might be involved in a specific disease or phenotype
- TRS requires **less sample input** and produces a smaller data set
- Methods: **hybridization capture** (target enrichment) and **amplicon sequencing**

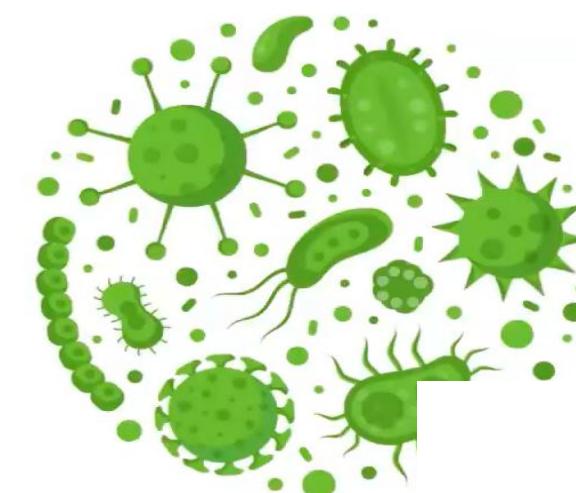


Hybridization capture	Amplicon sequencing
Larger gene content, <b>&gt; 50 genes</b>	Smaller gene content, <b>&lt; 50 genes</b>
More comprehensive, longer hands-on time	More affordable, easier workflow
<ul style="list-style-type: none"><li>• Exome sequencing</li><li>• Genotyping</li><li>• <b>Oncology</b></li><li>• Gene discovery</li></ul>	<ul style="list-style-type: none"><li>• Genotyping by sequencing</li><li>• <b>Detecting CRISPR editing events</b></li><li>• Detecting disease-associated variants</li></ul>

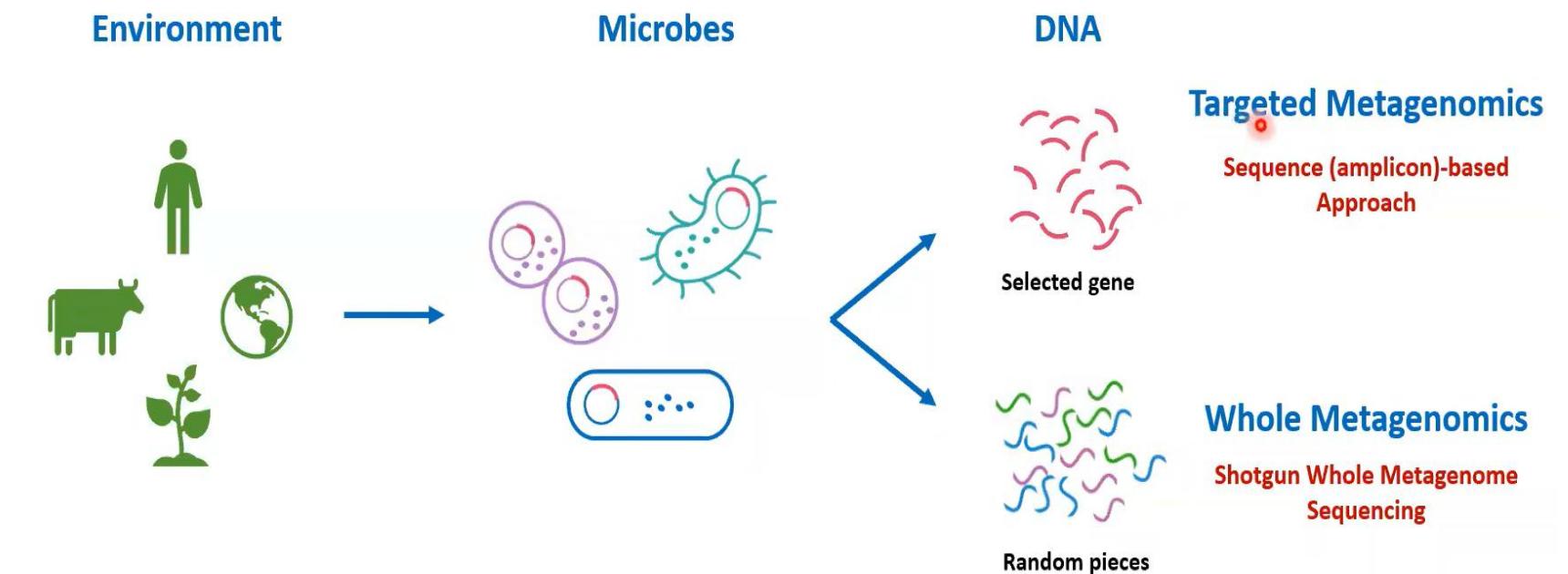
## 4

### Metagenomics

- **Metagenomics** is the study of microbial communities in their original habitats
- Metagenomics provides a **comprehensive insight** into the interactions within microbial communities



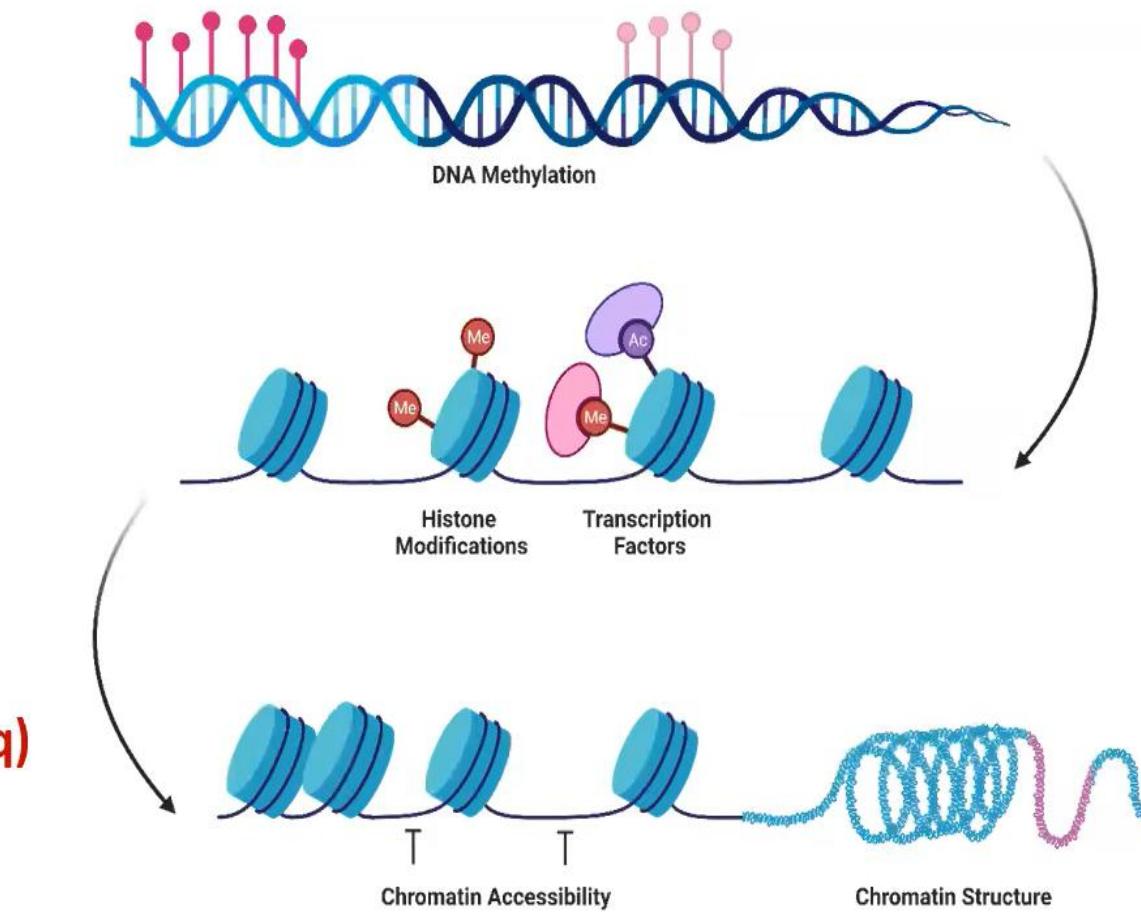
Advancing Genomics, Improving Life



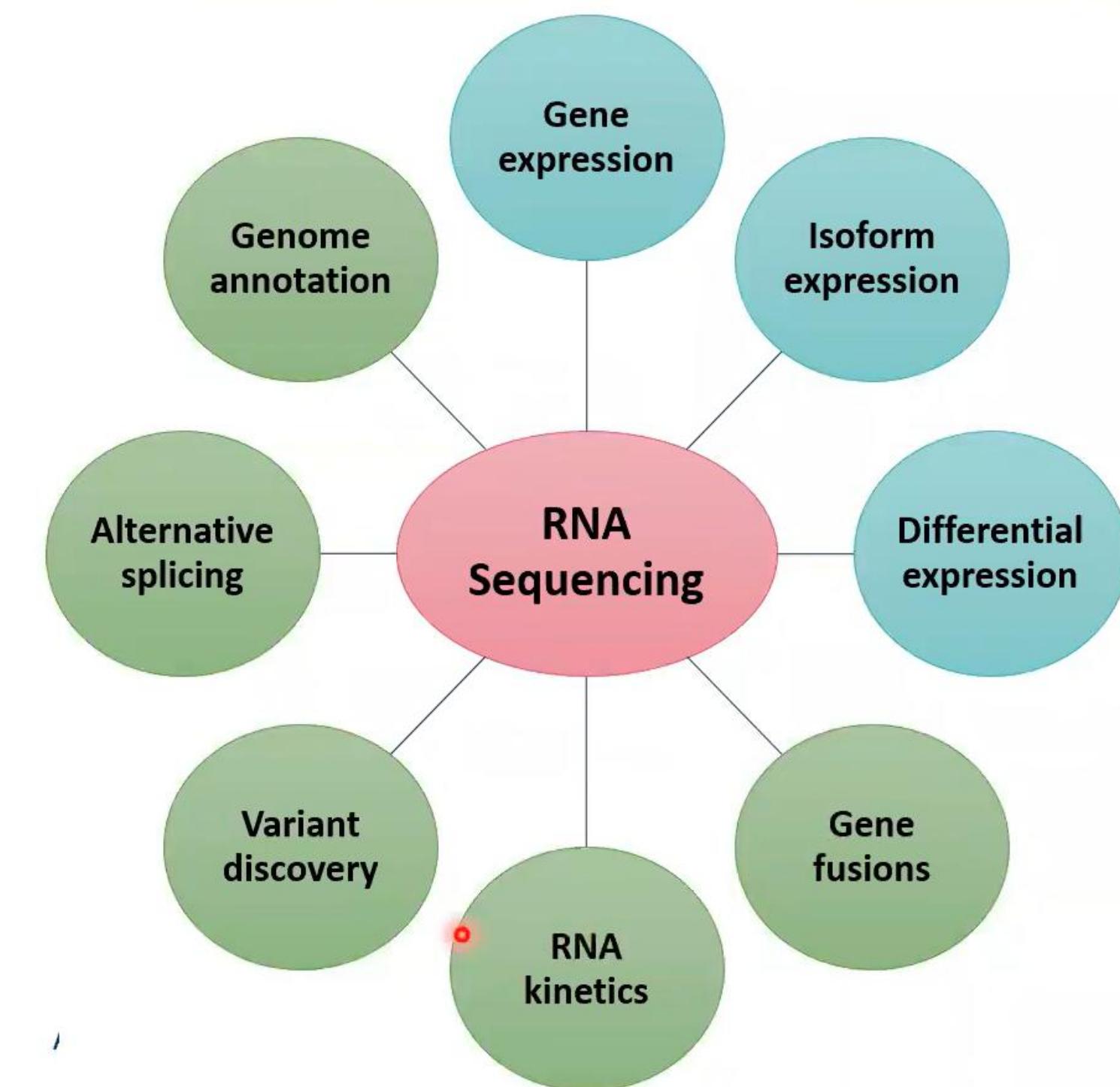
## 5

## Epigenetics

- **Epigenetics** is the study of stable phenotypic changes that do not involve alterations in the DNA sequence
- Mitotically and/or meiotically **heritable**
- **Reversible**
- DNA Methylation:  
**Whole Genome Bisulfite Sequencing (WGBS)**
- Histone modifications & Transcription Factors:  
**Chromatin Immunoprecipitation Sequencing (ChIP-seq)**
- Chromatin Accessibility: **ATAC-seq**
- Chromatin Conformation: **Hi-C**



# Which applications for RNA sequencing



# Oxford Nanopore Technology

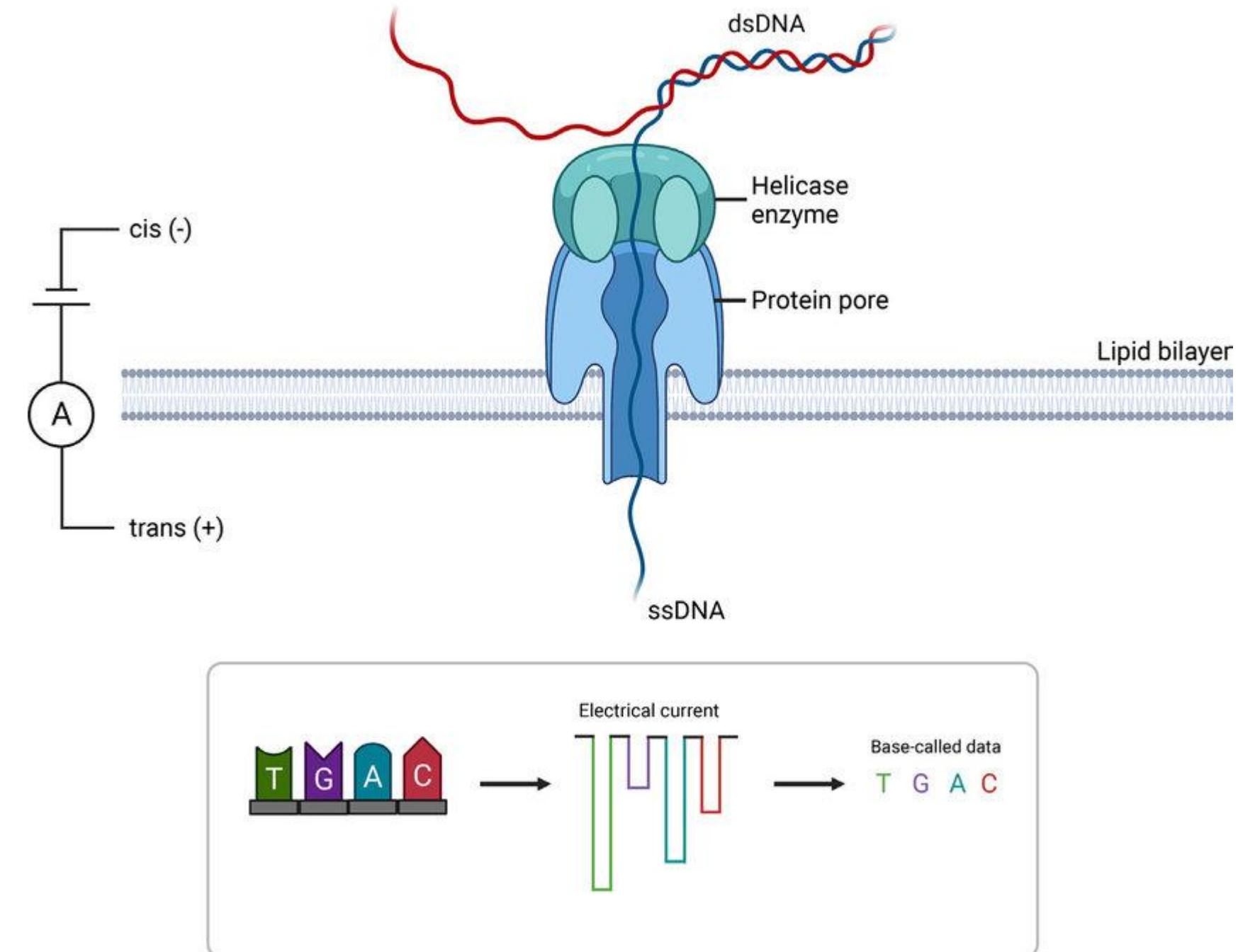


# The ONT sequencing

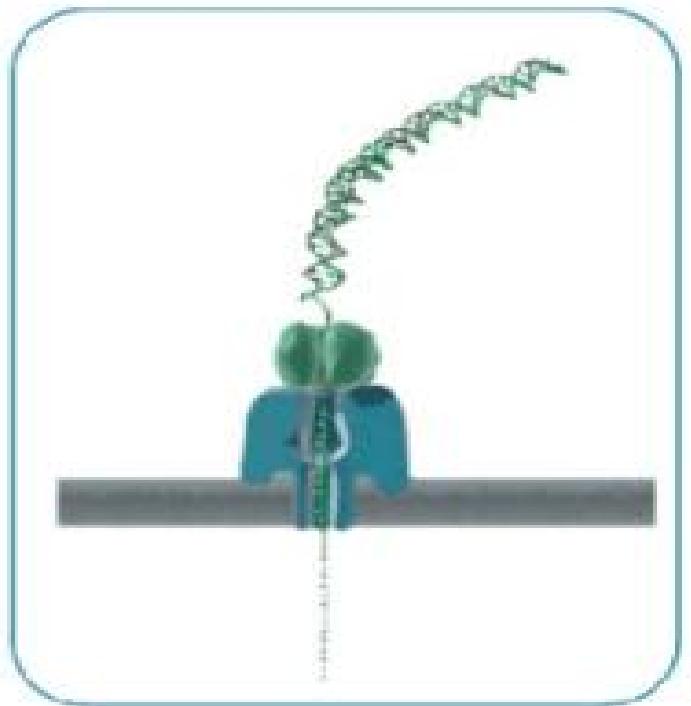
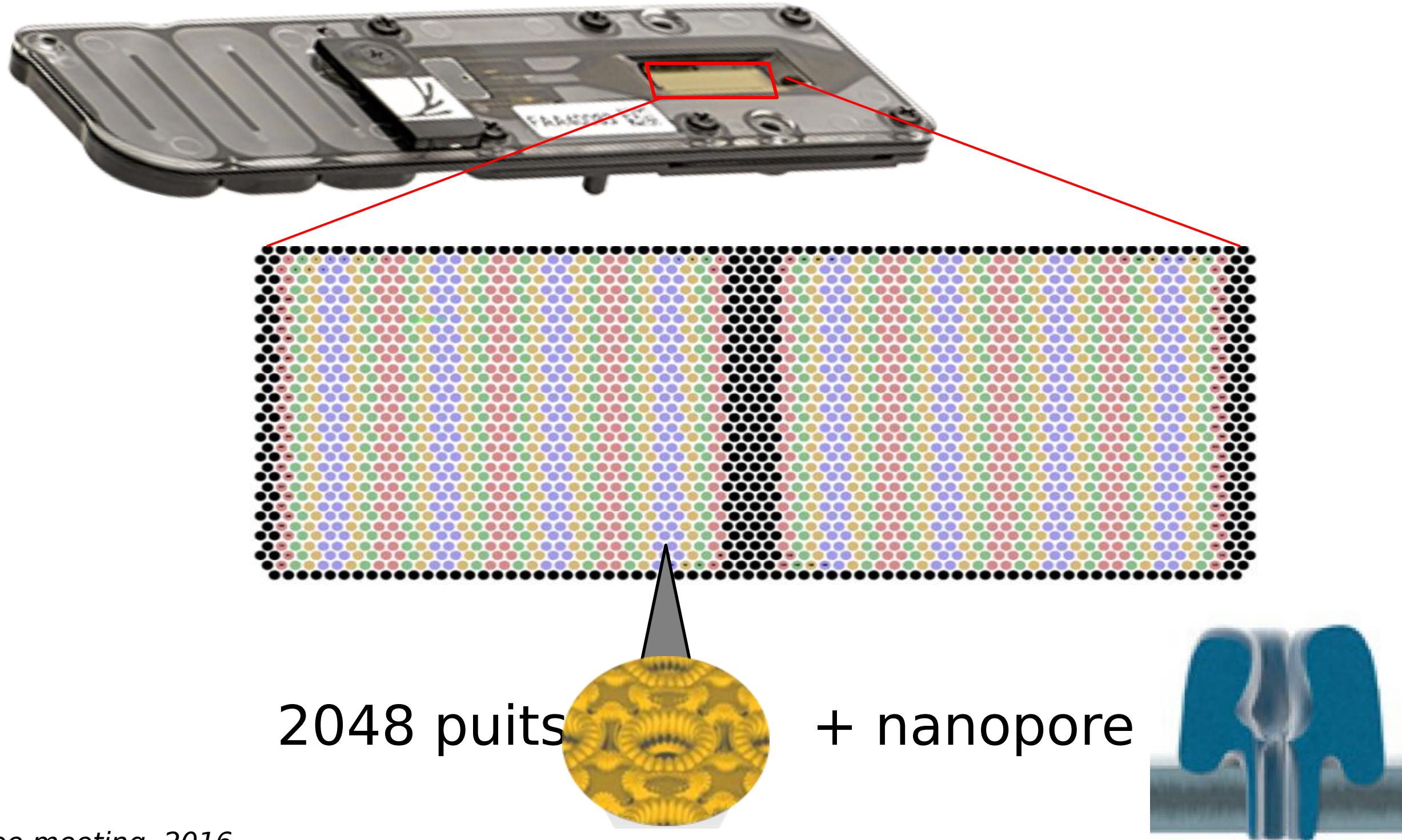


<https://nanoporetech.com/>

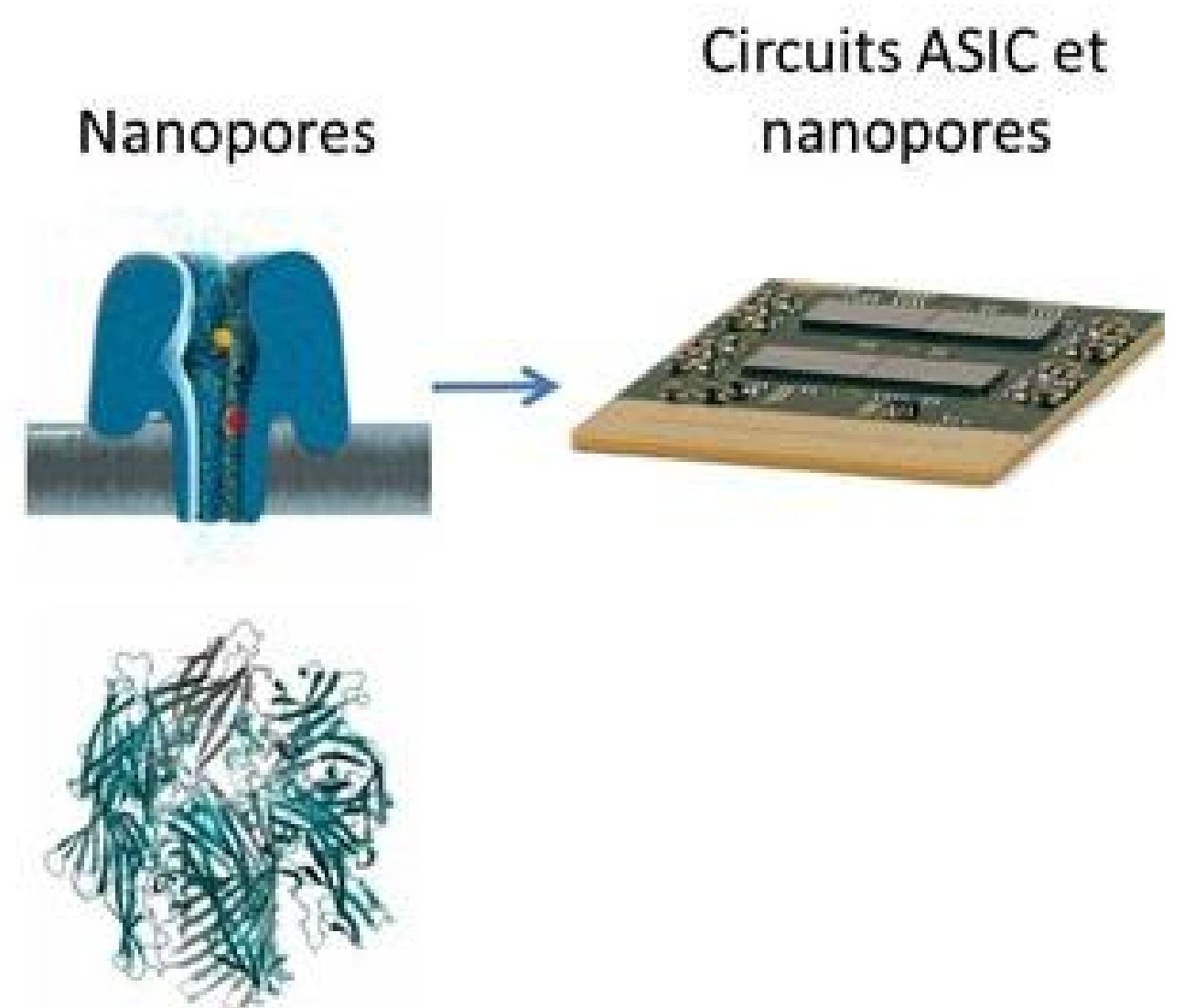
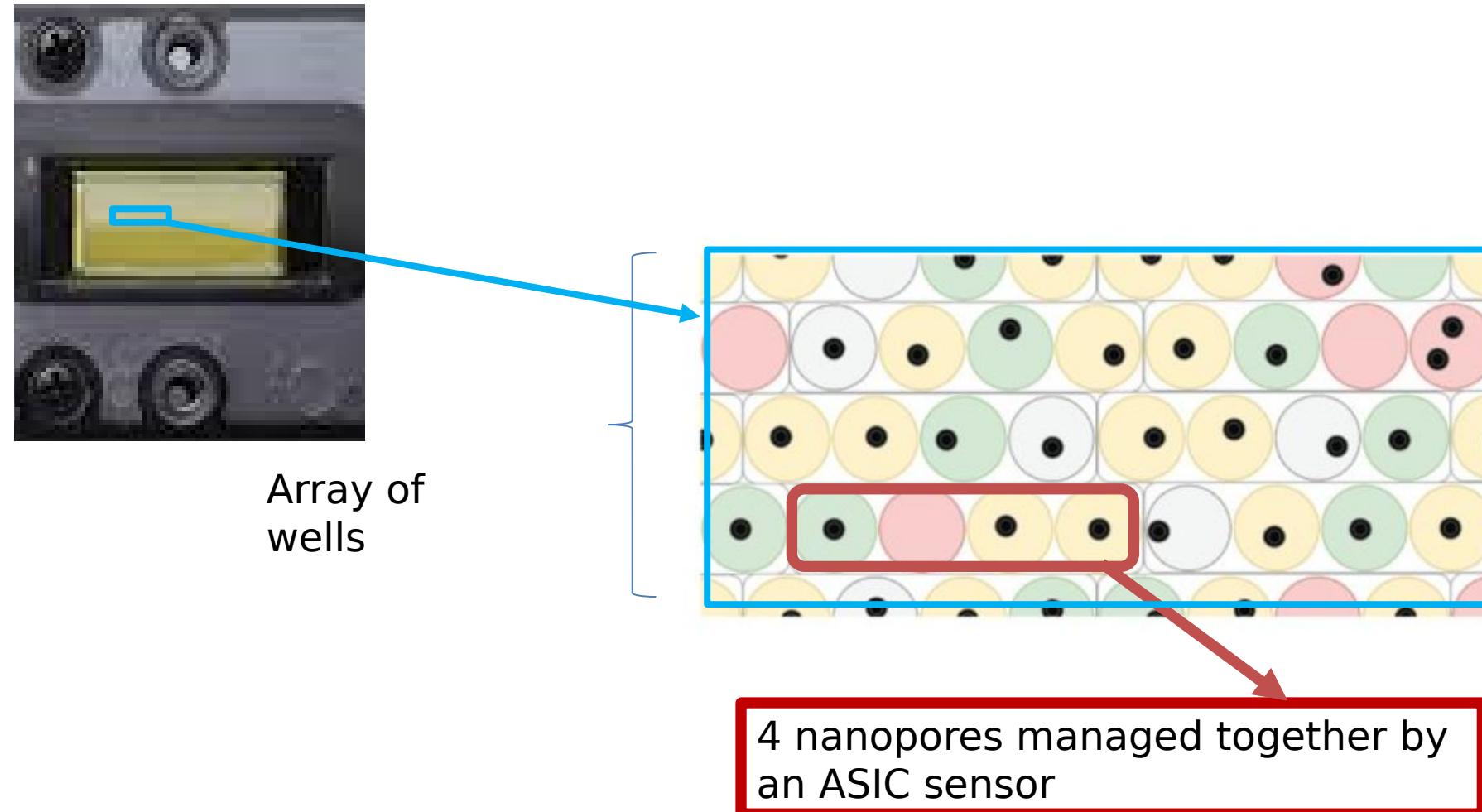
# The operating sequencing



# The flowcell



# The flowcell-ASIC



# The operating sequencing

## Generate new biological insights



### Whole genome sequencing

- De novo assembly and resequencing
- Scaffolding and finishing
- Variant analysis: structural variation, SNVs, phasing, base modifications
- Chromatin conformation



### Targeted sequencing

- Amplicon and PCR-free enrichment
- Real-time targeting with adaptive sampling (see page 15)
- 16S rRNA analysis
- Variant analysis: structural variation, SNVs, phasing, base modifications



### RNA sequencing

- Direct RNA, direct cDNA, and cDNA
- Characterise and quantify full-length transcripts
- Sequence complete viral genomes
- Variant analysis: splice variants, gene fusions, SNVs, base modifications



### Metagenomics

- Real-time, unbiased analysis of mixed samples
- Enhanced species identification using long reads



### Epigenetics

- Base modifications (e.g. methylation)
- Histone modification
- Non-coding RNA activity (e.g. lncRNA)



# The ONT sequencing

## Library preparation kits

Select the library preparation kit that matches your specific experimental needs — your choice of read length (short to ultra-long), turnaround time, input amount, sample multiplexing, modification detection, and output requirements. Find out more and view our complete library prep portfolio at [store.nanoporetech.com](http://store.nanoporetech.com).

	Native DNA			Amplified DNA	
	Ligation	Rapid/Field	Ultra-Long	PCR	Rapid PCR
 Prep time	60 mins	10 mins	90 min + 1x O/N incubation	60 mins + PCR	15 mins + PCR
 Input	1,000 ng dsDNA	From 50 ng HMW gDNA	6M cells / 1 ml blood	100 ng dsDNA	1-5 ng gDNA
 Multiplexing options	Yes	Yes	-	Yes	Yes
 Read length	Equal to fragment length	Random distribution, dependent on input fragment length	N50 >50 kb	Equal to fragment length post-PCR	<2 kb
 PCR required	No	No	No	Yes	Yes
 Product range highlights	Detect modified bases for free. Automatable workflows and XL kits enable production-scale sequencing			Ideal for low input amounts	

# The ONT sequencing

Biology for anyone, anywhere



MinION & GridION



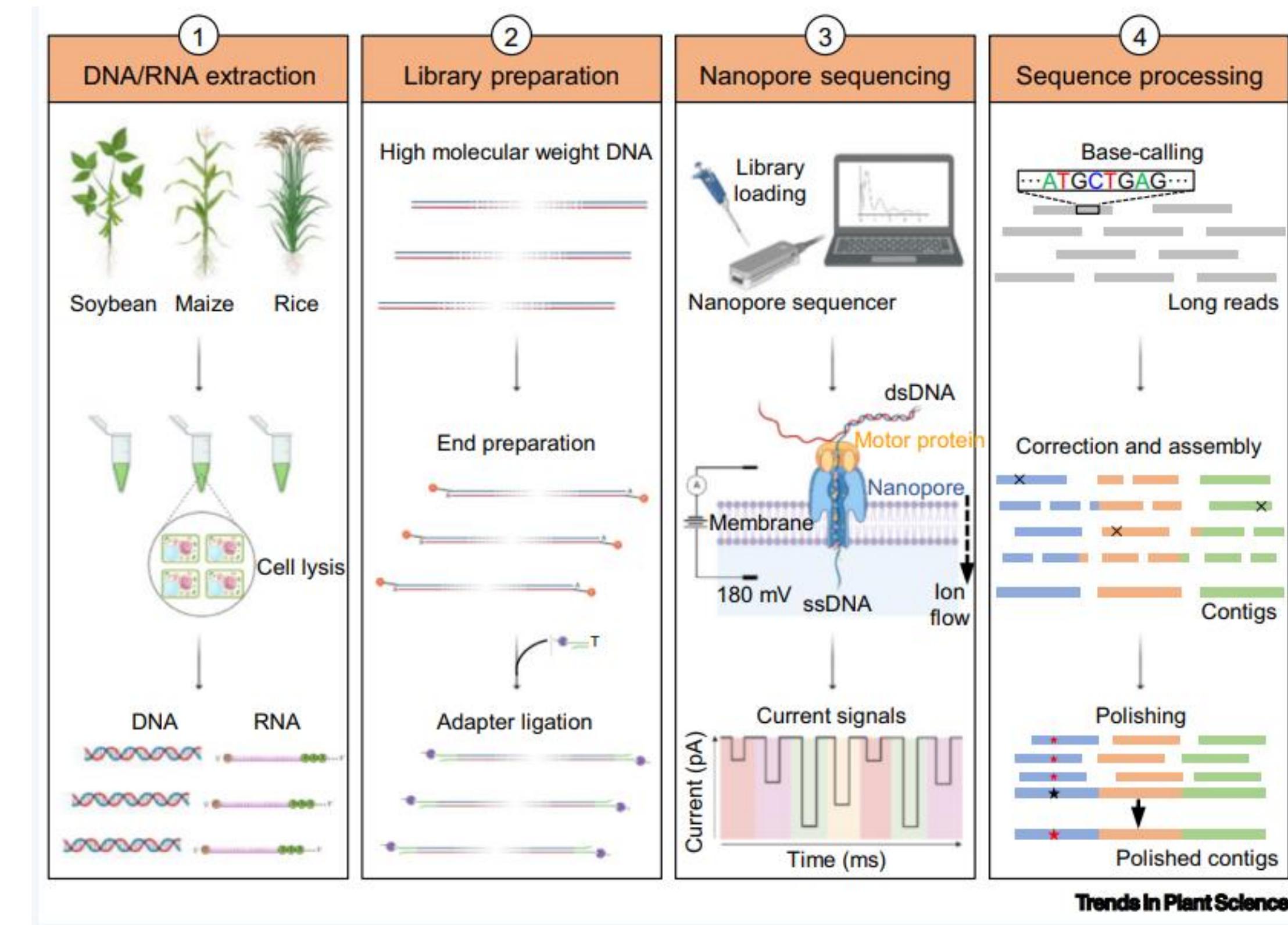
PromethION



Flongle



# Process ONT sequencing

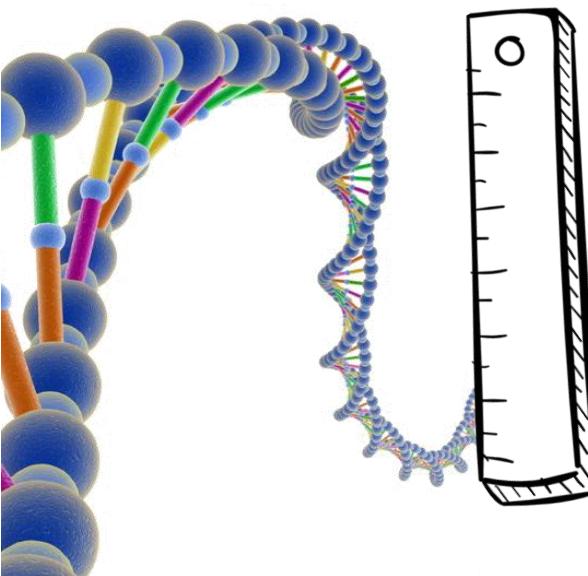


# High molecular weight DNA (HMW DNA) Extraction

Major requirement for ONT sequencing



# High molecular weight DNA: What is it and why is it needed?



From 50 to  
hundreds Kb

Whole Genome Analysis

Identification of Genetic Variants

Study of Repeated and Complex Regions

Mapping of Loci of Interest

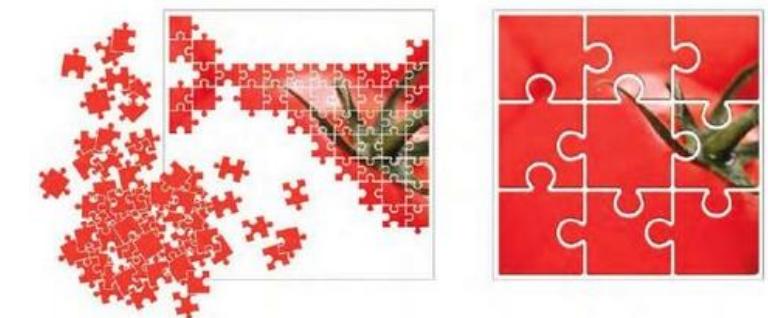
Study of Genomic Duplications

Research on Species Evolution

Epigenetic Mechanisms

...

the higher the molecular weight of the DNA,  
the longer the reads,  
the easier the assembling !



# HMW DNA extraction protocole

## Specificity of plant DNA extraction:

Polysaccharides, proteins, tannins, alkaloids, polyphenols, and other secondary metabolites...  
Selection of tissue, developmental stage...  
Need for customization and adjustment based on certain species

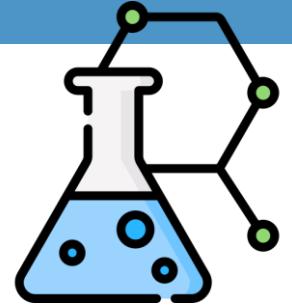


# HMW DNA extraction protocole

General DNA extraction : What are the components and why we use them?

Lysis buffer

Purpose	Chemistry
pH stabilization	Tris
Maintain cellular integrity	NaCl
DNA protection from degradation	EDTA
Help DNA concentration and purity	PEG6000
Reduce disulfide bonds within proteins	DTT
Disruption of bi-lipid membrane	CTAB $\simeq$ MATAB
Lipid and protein removal	Chloroform : Isoamylalcohol (24:1)
Nucleases inactivation	Proteinase K
Enzymatic RNA removal	RNase A
DNA preferential precipitation	Isopropanol, Sodium acetate
Salt removal	70% Ethanol
Re-suspending DNA precipitate	Water, TE 0.1X



# High molecular weight DNA: How to extract it?



## Tips and best practices



Use fresh (or flash frozen), young and healthy tissue



Forget your favorite column



Use wide-bore tips



Avoid mechanical stress



Gentle agitation and homogenisation



Unplug your vortex

# HMW DNA extraction protocole

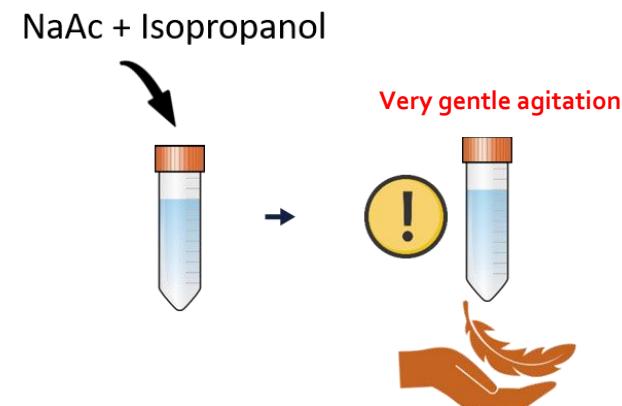
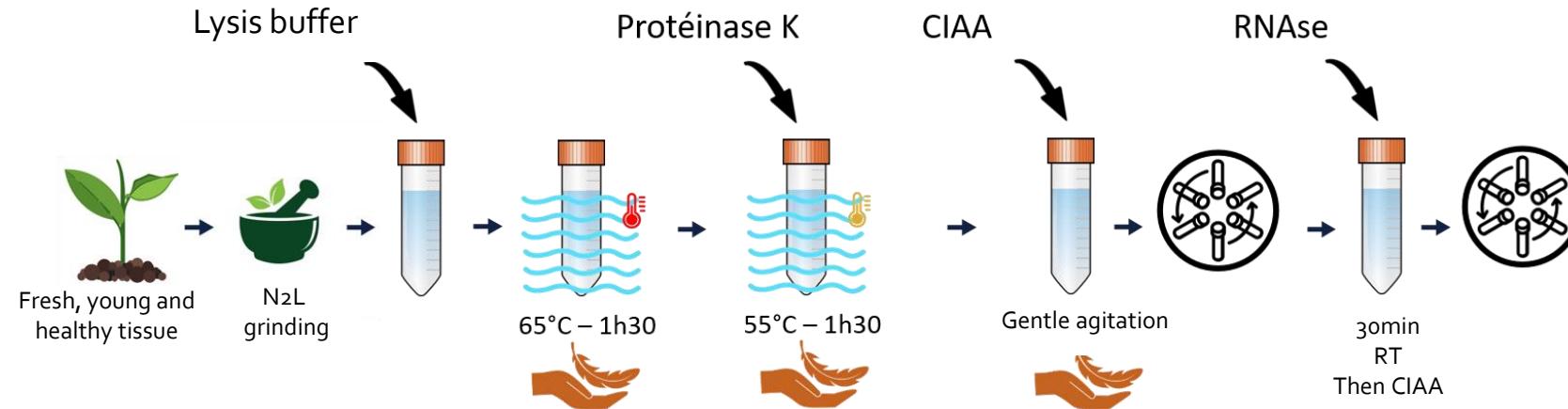


From low cost plant HMW DNA extraction to MinION sequencing

Julien Serret<sup>1</sup>, marie.couderc<sup>1</sup>, Cedric Mariac<sup>1</sup>, Laurencealbar<sup>1</sup>, Francois Sabot<sup>1</sup>

<sup>1</sup>Institut de Recherche pour le Développement

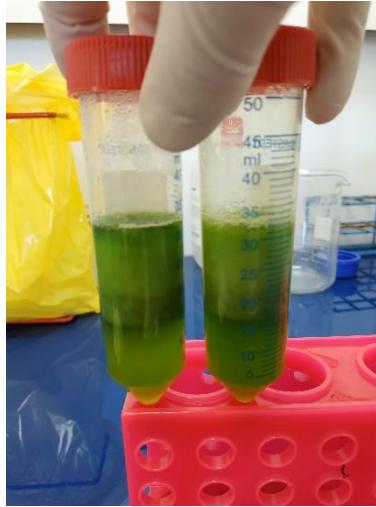
[dx.doi.org/10.17504/protocols.io.bu3vnyn6](https://dx.doi.org/10.17504/protocols.io.bu3vnyn6)



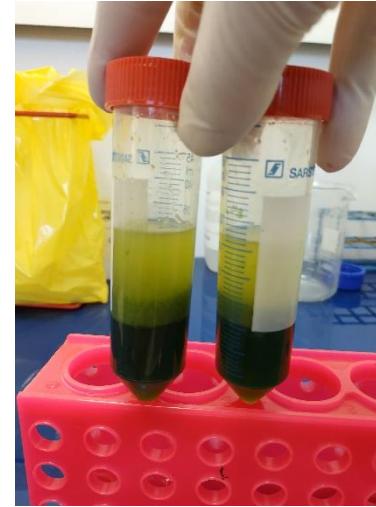
# HMW DNA extraction protocole steps



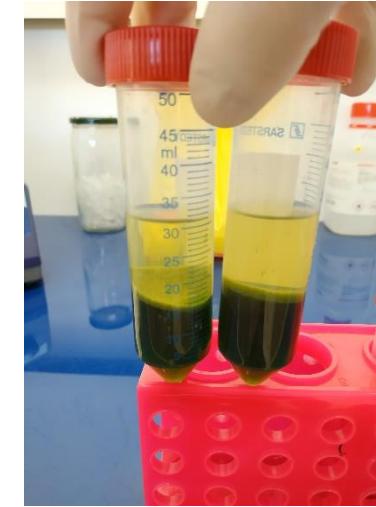
Powder in lysis buffer



CIAA addition



After gentle agitation



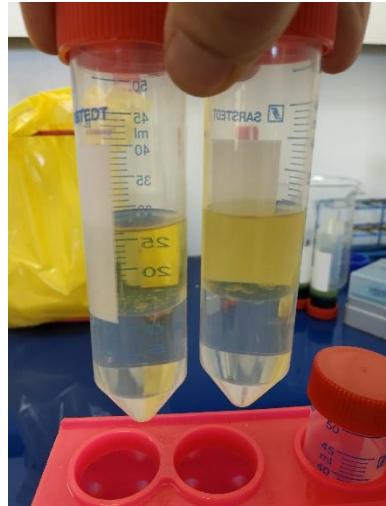
After centrifuge



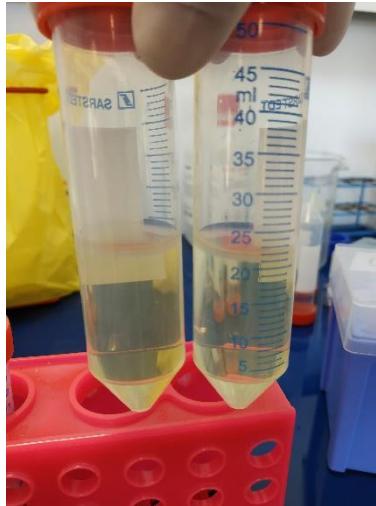
Upper phase recovered and treated with RNase



Second CIAA



After centrifuge



NaAc + Isprop precipitation



DNA pellet

# Laboratory

HMW DNA extraction Demo



# High molecular weight DNA (HMW DNA) Qualification



# How to visualize HMW DNA ?

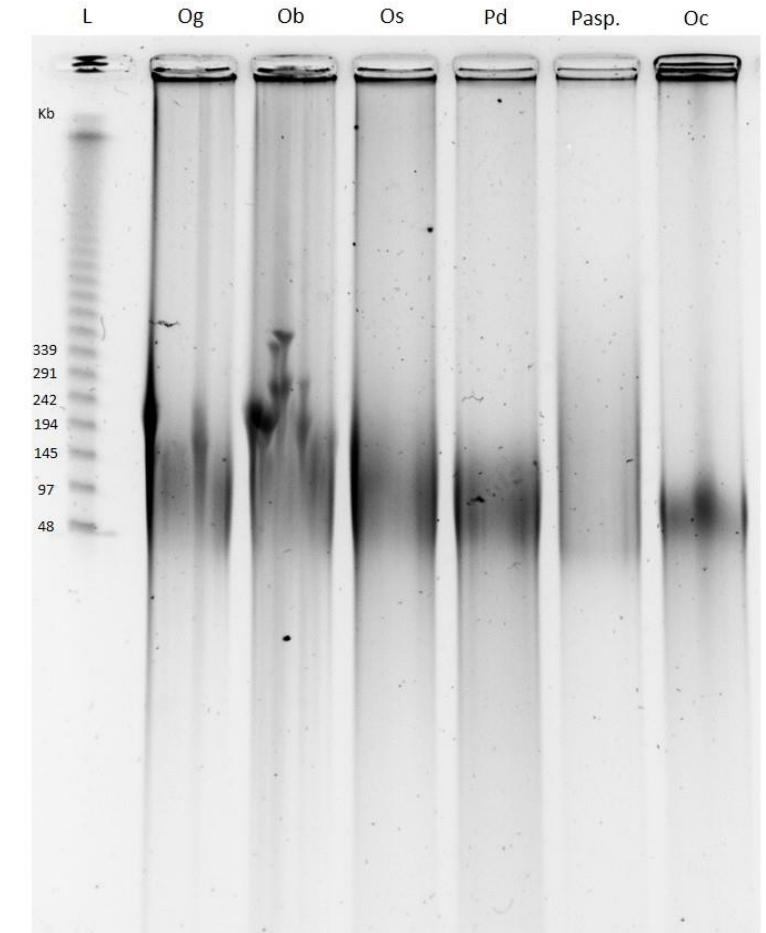
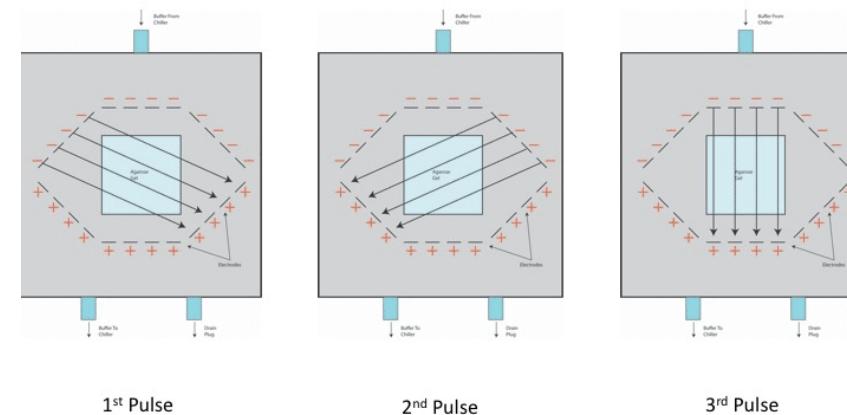
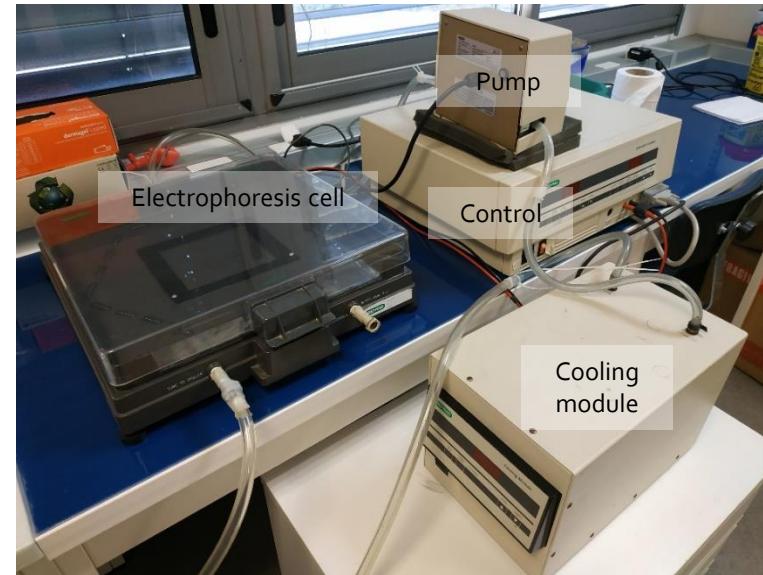
Volume 37, Issue 1, May 1984, Pages 67-75

Article

Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis

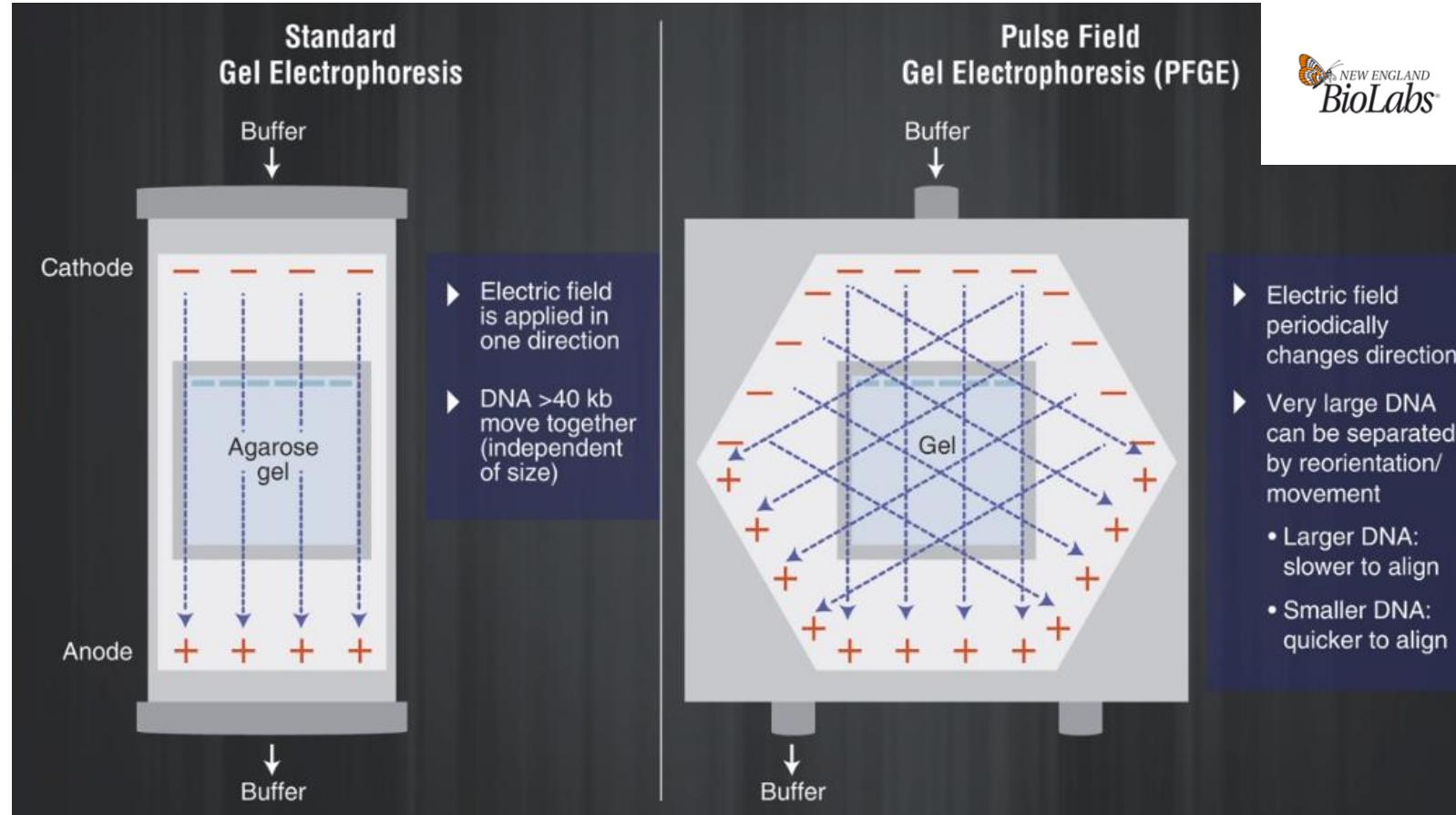
David C. Schwartz, Charles R. Cantor

## Pulsed field gel electrophoresis (PFGE)



1.5µg of DNA loaded on 1% pulsed-field electrophoresis gel  
Lambda PFG Ladder (L), *Oryza glaberrima* (Og), *Oryza barthii* (Ob), *Oryza sativa* (Os), *Phoenix dactylifera* (Pd), *Paspalum* (Pasp.), *Ocotea obtusata* (Oc).

# Standard electrophoresis Vs PFGE

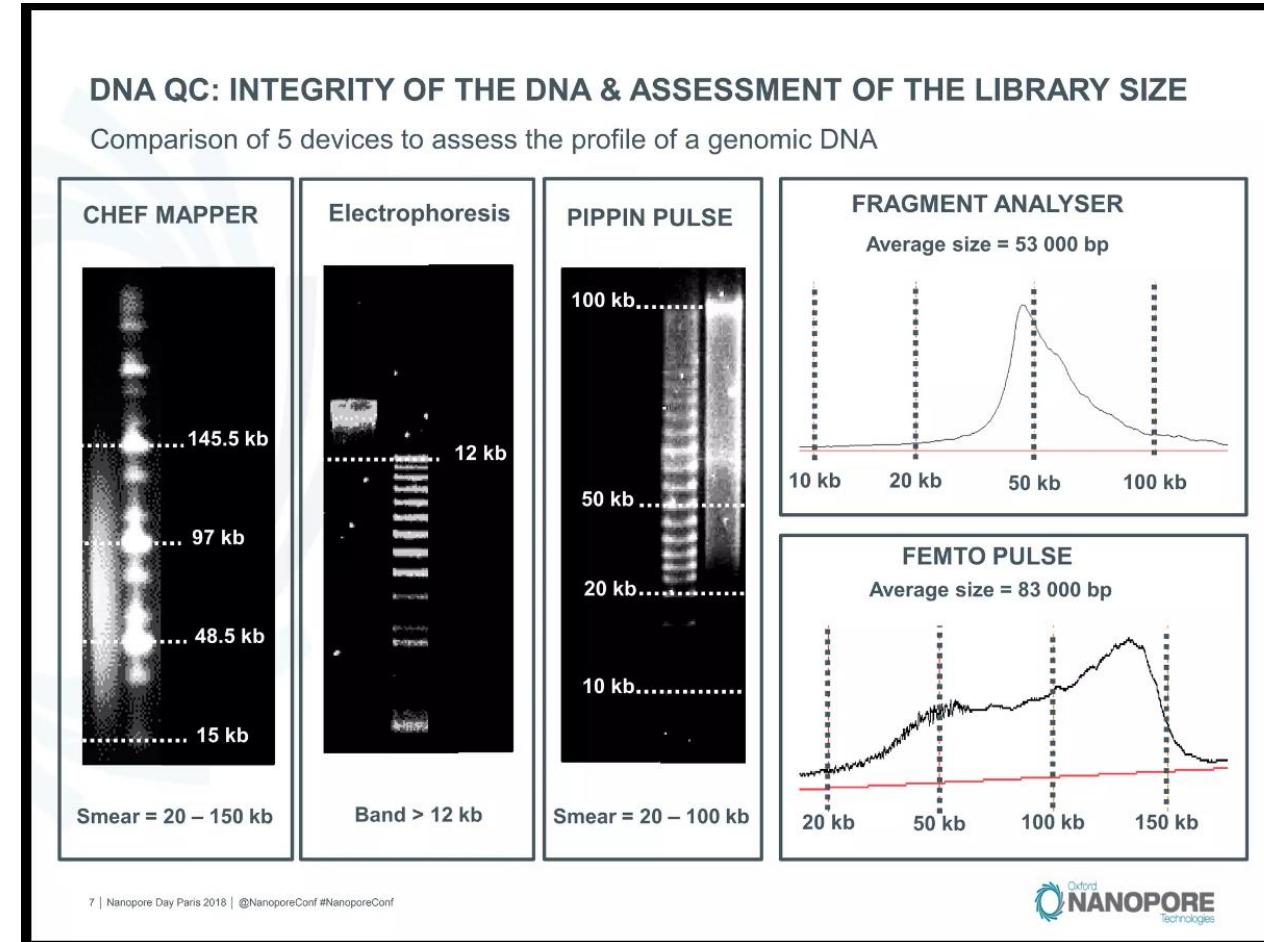


100 pb – 40 Kb



50 Kb – 700 Kb

# How to visualize HMW DNA ?



Purity of the sample is a key parameter !

- *Library prep* : impurities can reduce the efficiency of enzymatic reactions
  - *Sequencing* : pores could be very sensitive to contaminants
- > Ultra pure DNA is required to obtain optimal results

## Concentration measurement

Qubit  
fluorometer



Fluorescence dsDNA specific  
Accurate at low concentrations  
Reduced sample consumption

## Quality evaluation

Nanodrop  
spectrophotometer



Determination of quality ratios



### Nota bene

The concentration readings between NanoDrop & Qubit should be less than 50% different.

Ratio  $A^{260}/A^{280}$  1,8 - 2,0

If  $< 1,8$

Too little DNA compared to other components; organic contaminants (proteins, phenol...) [absorb at 280nm](#)

Ratio  $A^{260}/A^{230}$  2,0 - 2,2

If  $< 2$

Salt contamination, peptides, aromatic compounds, polyphenols, guanidine, thiocyanates... [absorb at 230nm](#)

A blue, textured 3D rendering of a DNA double helix molecule, oriented diagonally across the left side of the slide.

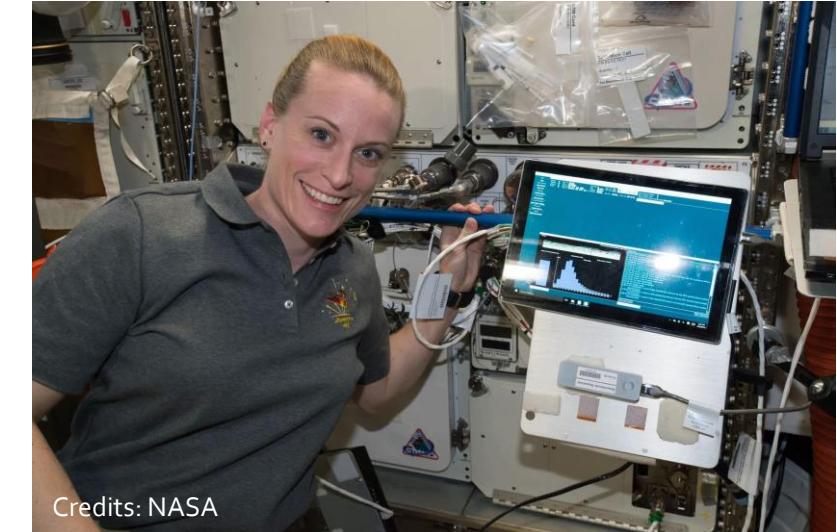
## DNA library preparation and sequencing

# ONT's promises

## Sequence Everything, Everywhere

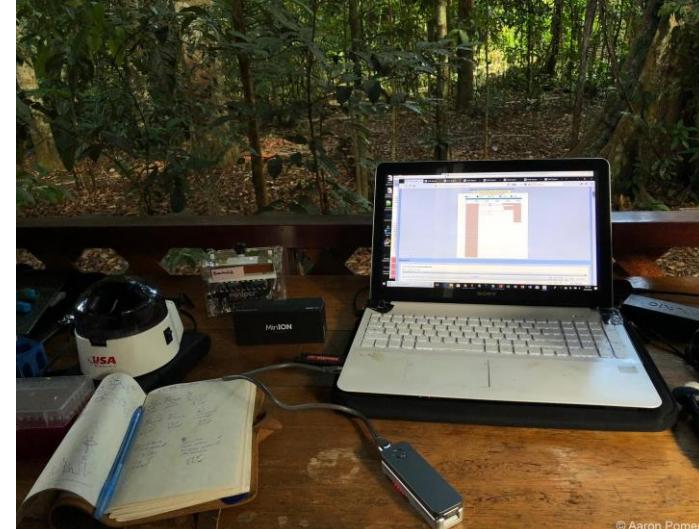


doi: [10.7171/jbt.17-2801-009](https://doi.org/10.7171/jbt.17-2801-009)



Credits: NASA

Oxford  
**NANOPORE**  
Technologies



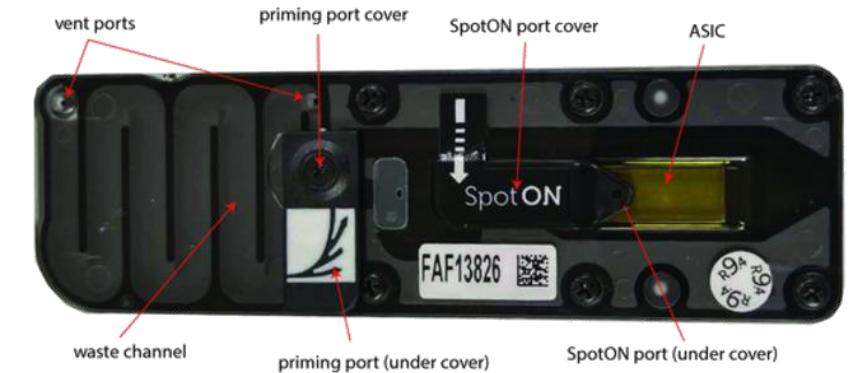
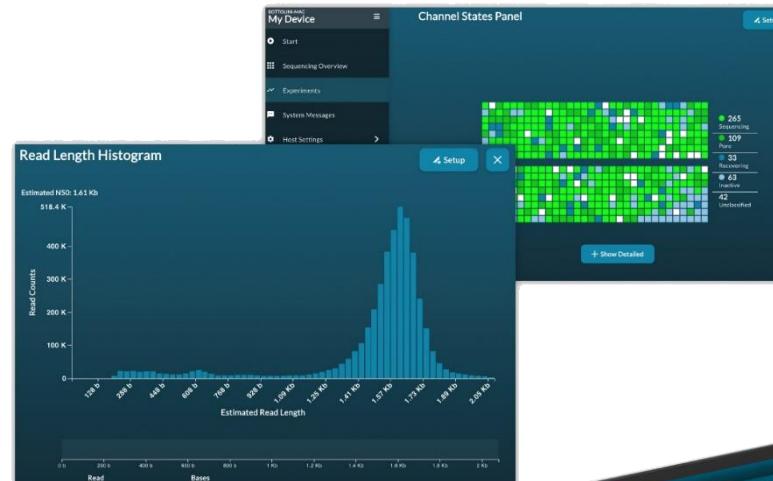
Oxford

# MinION devices

Real-time sequencing monitoring

Visualisation of the size and volume of data produced

Evolution of flow cell health



MinION Flow Cell  
512 channels  
4 pores/channel

# Short read eliminator (SRE) / short fragment eliminator (SFE)



## SRE kit

SKU 102-208-300 (previously SS-100-101-01)

Reagents to complete depletion/size selection of DNA <25 kb

\$150.00

1 item in the cart

Quantity: 1

Add More

Go to Checkout

## Short Fragment Eliminator Expansion

EXP-SFE001



An expansion pack designed for size selection of high molecular weight DNA samples to progressively deplete short fragments of <25 kb.

€95.00

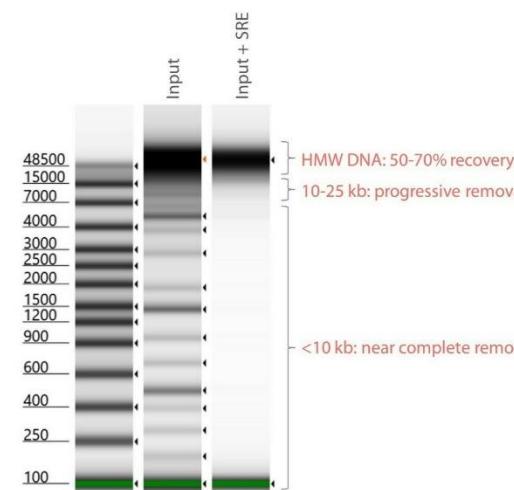
-

1

+

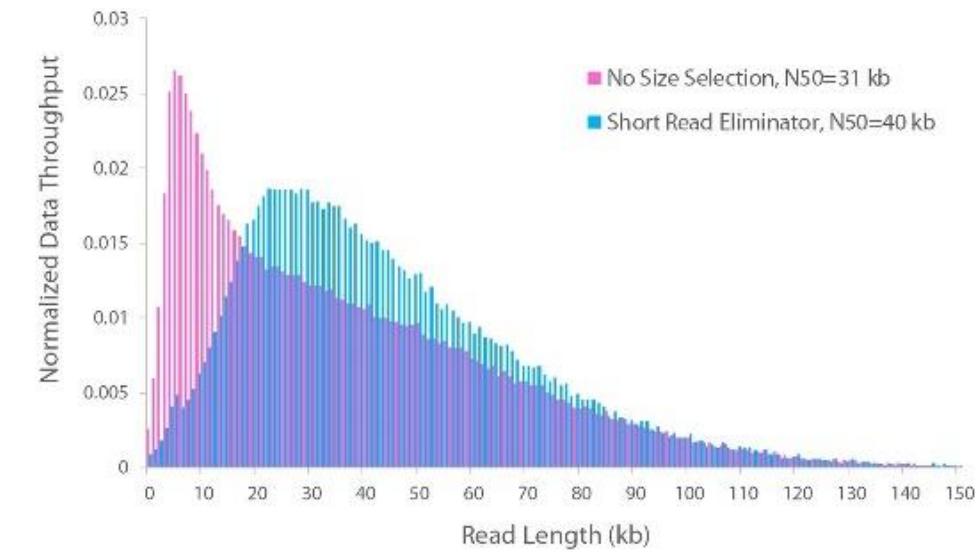
Add to basket

2 Early Access

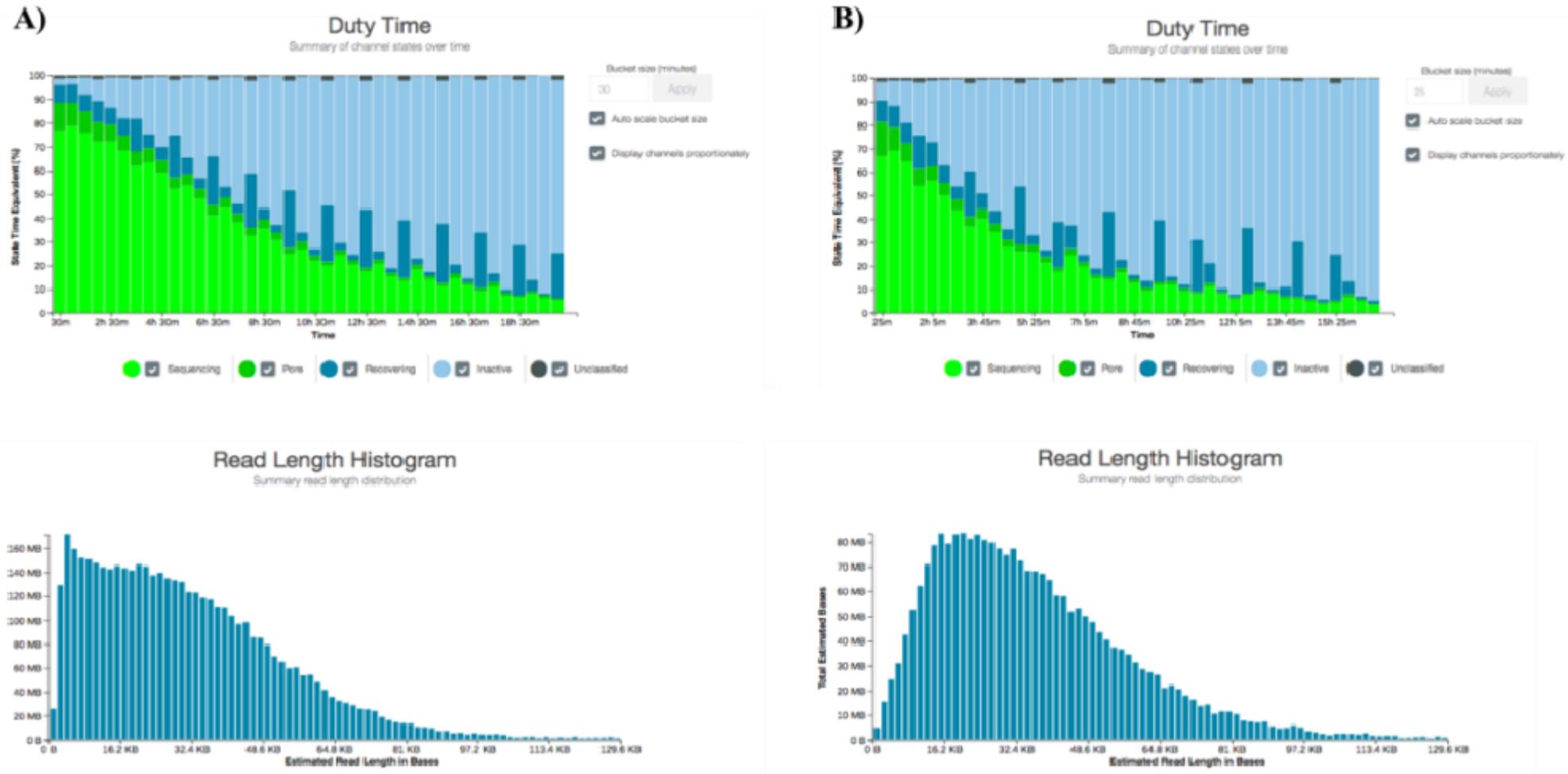


The Short Read Eliminator Kit was used to size select HMW DNA from GM12878 cells that was isolated using the Nanobind CBB Big DNA Kit and spiked with a 1 kb DNA ladder. Short DNA <10 kb is undetectable by agarose gel or CE while DNA from 10-25 kb is progressively removed. Each version of the Short Read Eliminator Kit operates under the sample principle but with different cutoffs suitable for different quality DNA samples.

## Nanobind Extracted Human Blood DNA on ONT GridION



N50 improvement



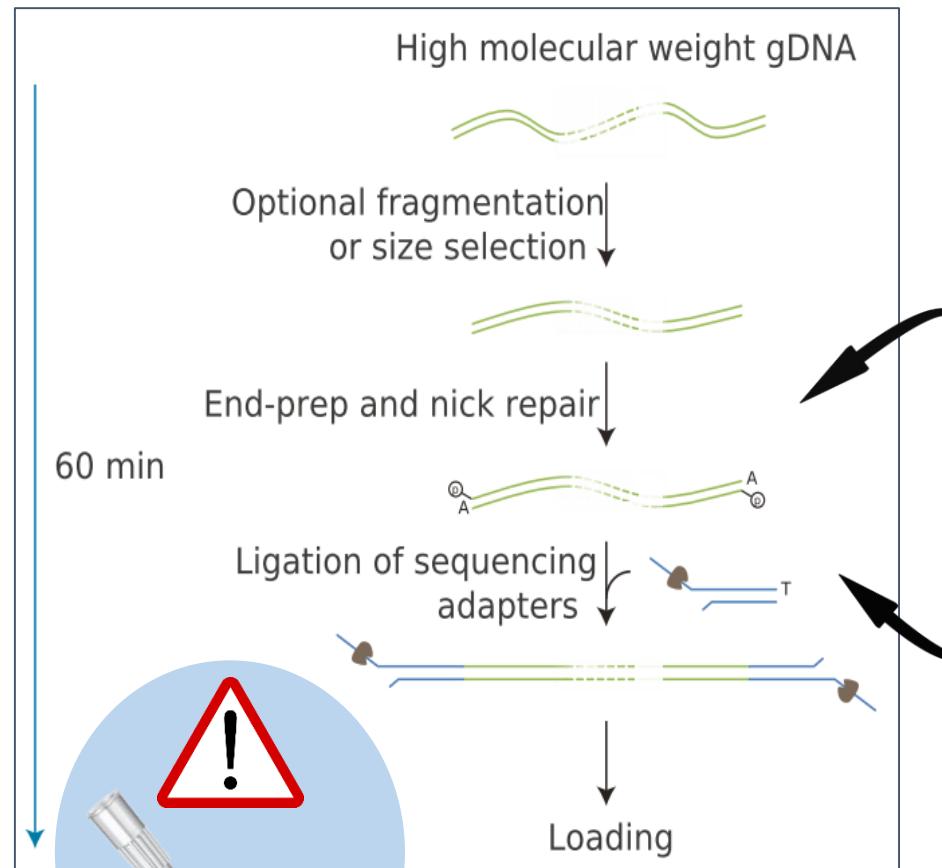
## Circulomics size selection test.

- A) Run without size selection: regular pore-loss, mean read length leaning towards 15 kbp and 4.97 Gbp called.
- B) Run with Circulomics size selection: increased pore-loss, mean read length leaning towards 20 kbp, and 2.36 Gbp called.

# ONT library construction



Exemple of kit SQK-LSK110 (Ligation sequencing kit)



**Input DNA: 9µg**

**Repair enzymes**

FFPE DNA Repair Mix

Ultra™ II End Repair/dA-Tailing



[AMPure XP magnetic beads purification](#)

**Ligation enzyme**

Quick T4 DNA Ligase



[AMPure XP magnetic beads purification](#)

**Load 1-3µg of library**

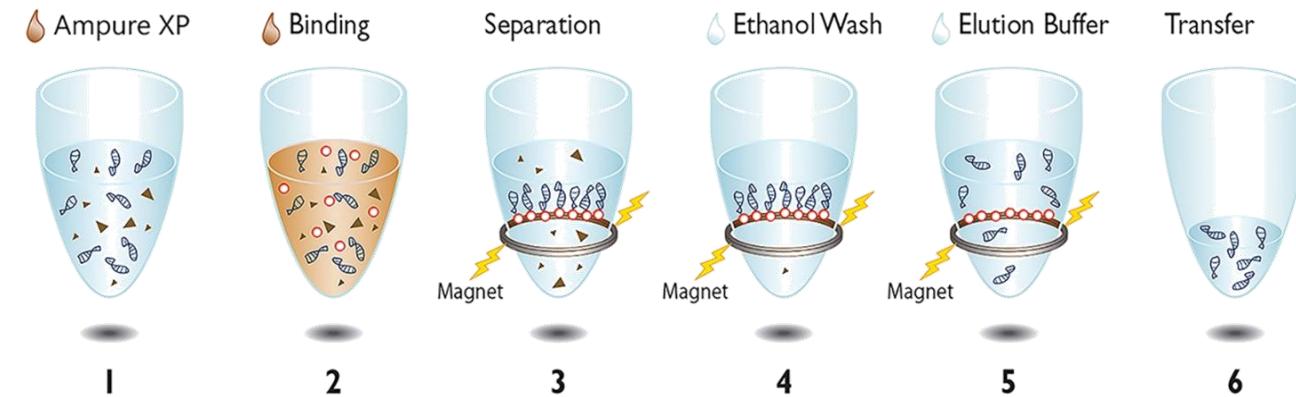
**Nick fill:** damaged double-stranded DNA reparation

**End Repair:** A nucleotide addition at each end of the double-stranded DNA, enabling ligation of adapters

**Ligase:** attaches Y-shaped adapters with a motor protein to DNA strands

# Different stages of purification

AMPure XP magnetic beads (Beckman) for cleaning  
Removal of enzymes and other compounds  
Buffer change between reactions  
DNA concentration

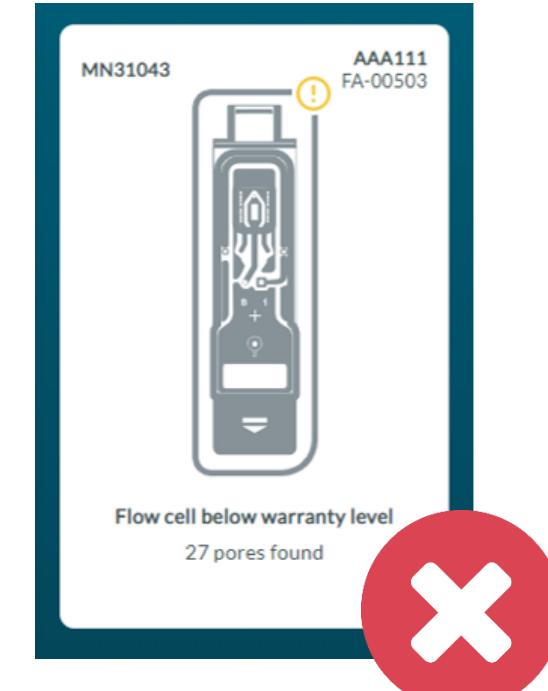
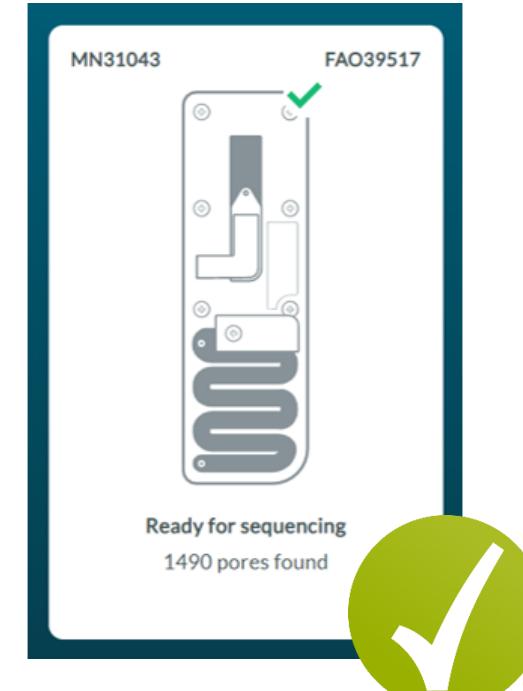
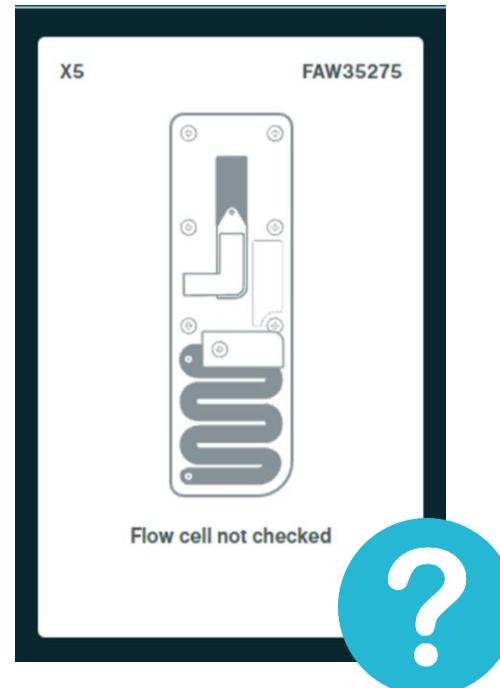


# Flowcell checking

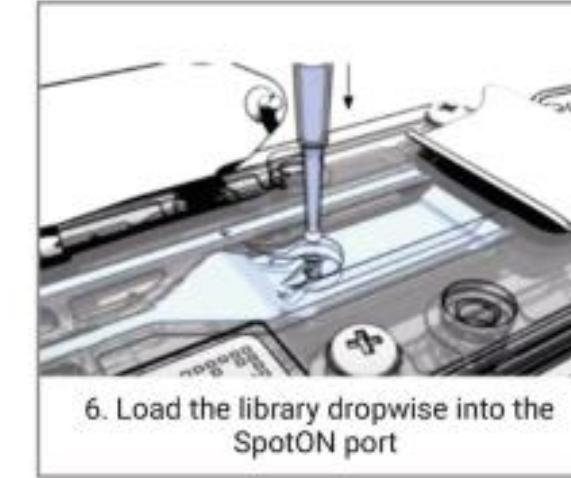
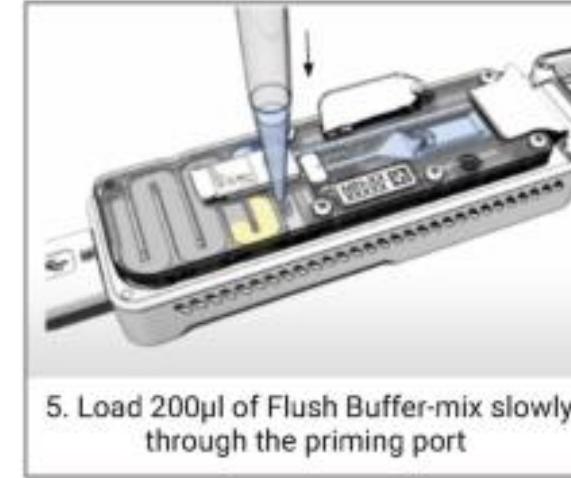
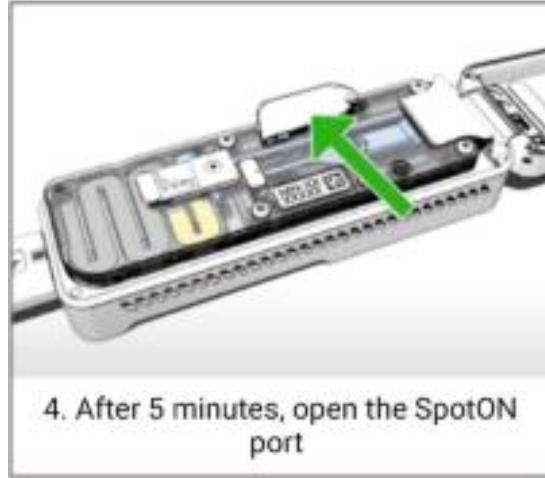
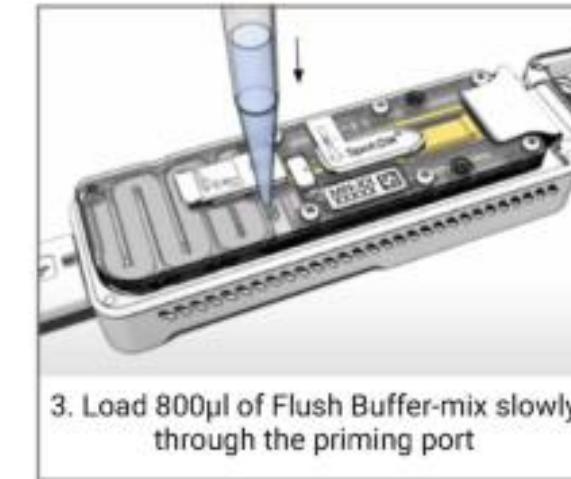
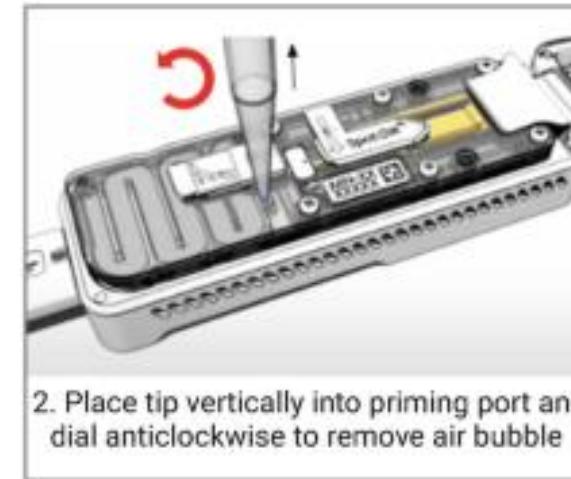
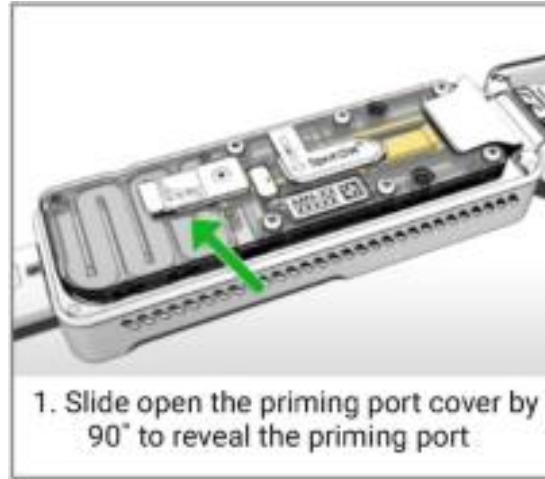
MinION Flow Cell  
512 channels  
4 pores/channel  
  
2048 pores

Flowcell age  
Storage conditions  
Conception  
Other...

Flowcell check :  
number of pores  
actually available



# How to prepare the MinION flowcell and load the library



## Some key figures

Starter Pack	€950.00
1x MinION Sequencing Device MIN-101B	
1x Control Expansion Kit EXP-CTL001	
1x Flow Cell Wash Kit EXP-WSH004	
1x Flow Cell (R10.4.1) FLO-MIN114	



**Flow Cell (R9.4.1)**  
FLO-MIN106D



From 455€ (pack 96 FC) to  
855€ (pack 1 FC)

**Ligation Sequencing Kit**  
SQK-LSK110



**6 reactions**

A versatile sequencing kit optimised for output, and long reads, and processing singleplex samples. We recommend using our latest version [Ligation Sequencing Kit V14 \(SQK-LSK114\)](#).

This uses Kit 10 chemistry.

This product is scheduled to be discontinued. The recommended replacement is [SQK-LSK114](#).

€570.00



AMPure XP SPRI Reagent, 5 mL



384€



€ 980,00

**NEBNext Companion Module**  
for Oxford Nanopore Ligation Seq  
Cond.: 24 rxns

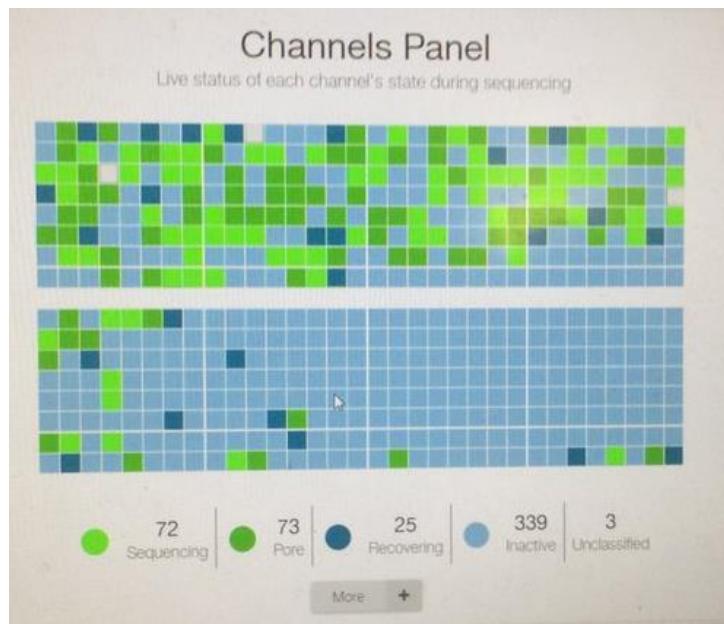
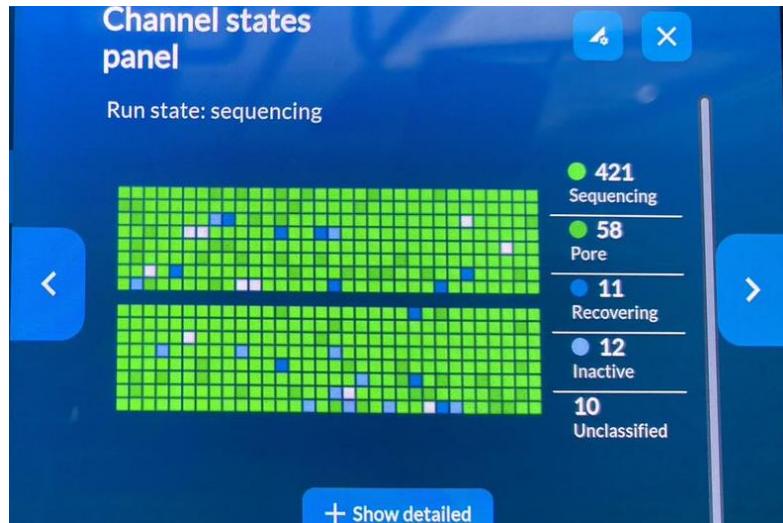


Flowcells are single-use

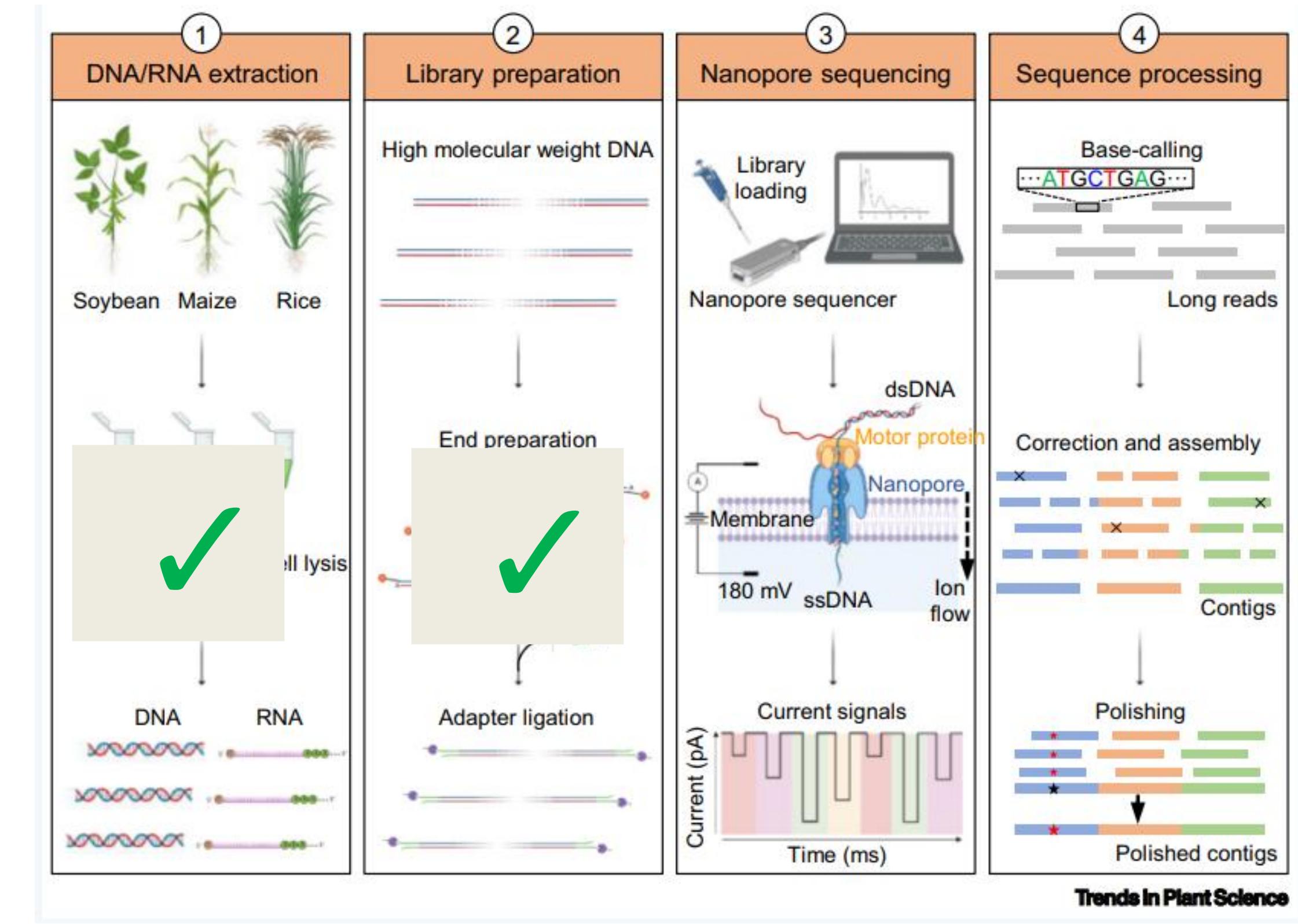
Multiple (2 or 3) loads are possible with the same genotype (or different if barcoding kit)

Max flowcell yield obtained at IRD 25Gb

# Quelques exemples de status d'une flow cell



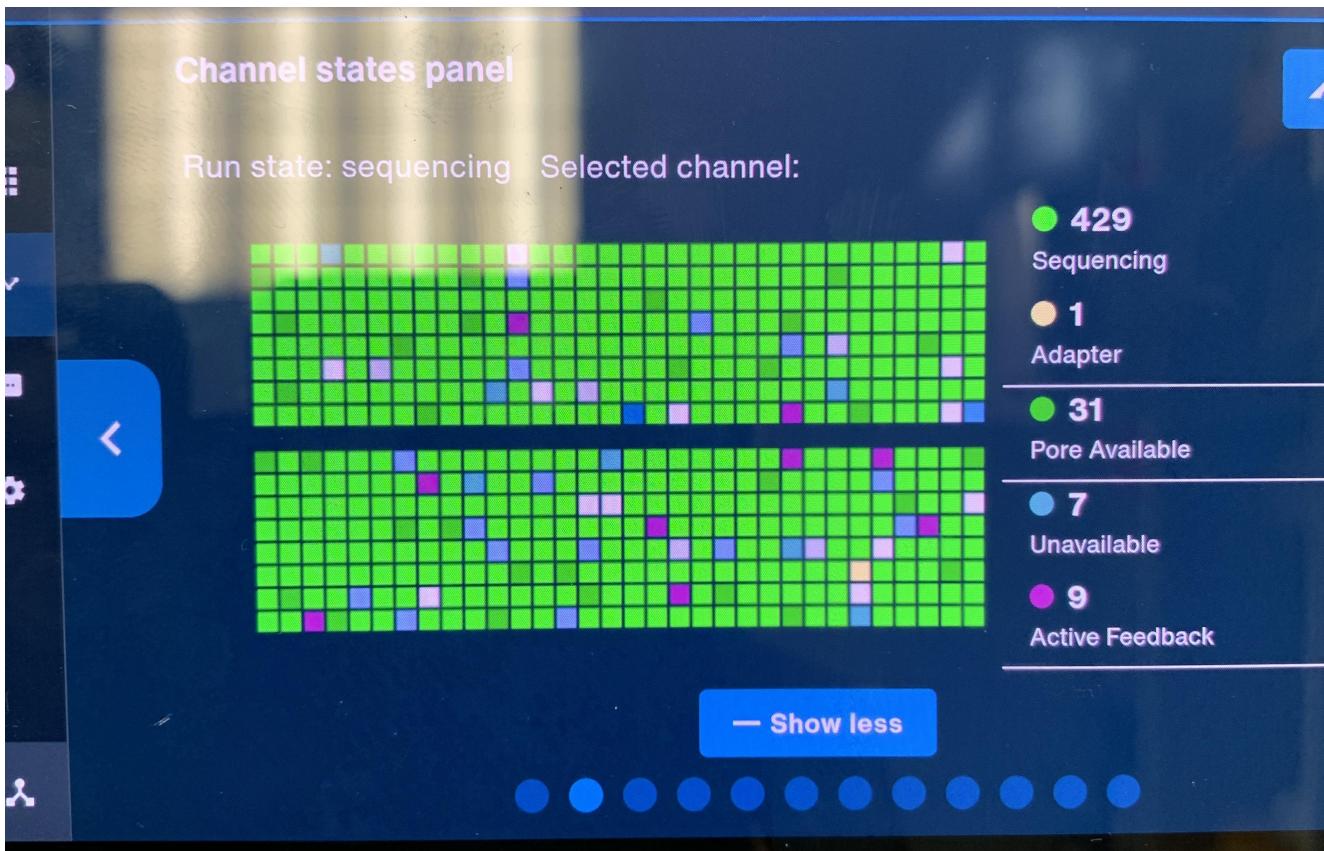
# The ONT sequencing



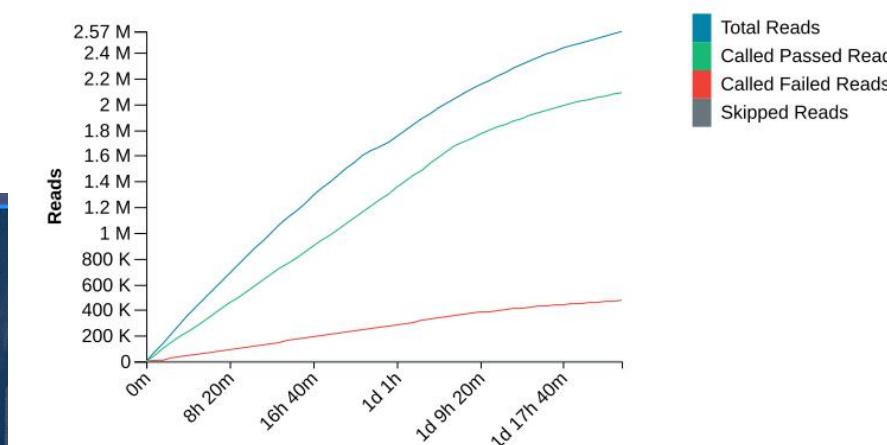


Qualitative analysis of sequencing runs in progress

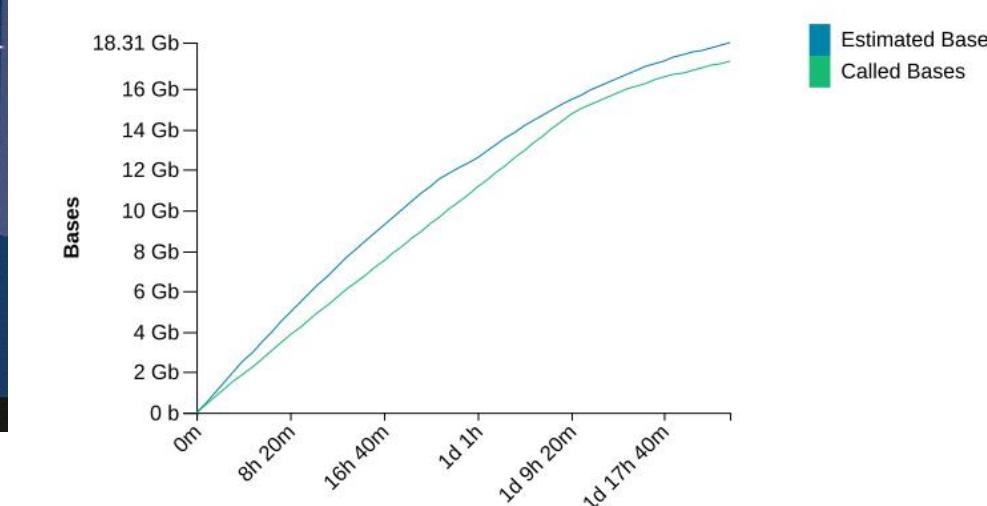
# Sequencer management



Cumulative Output Reads

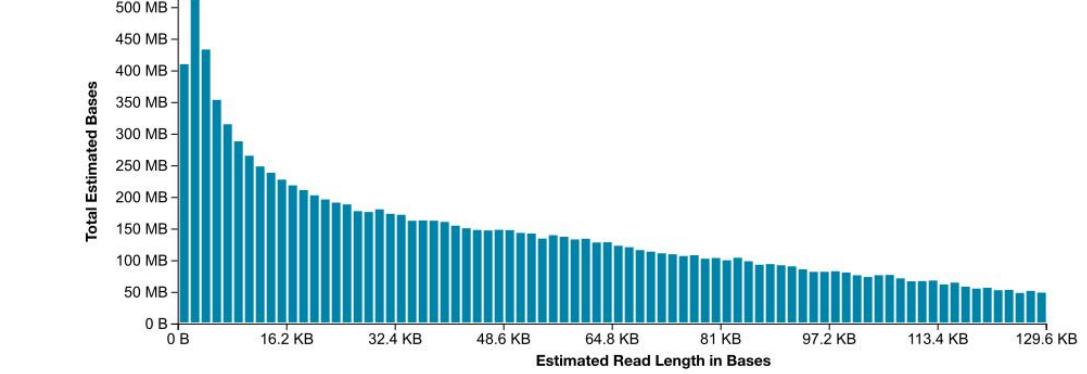


Cumulative Output Bases

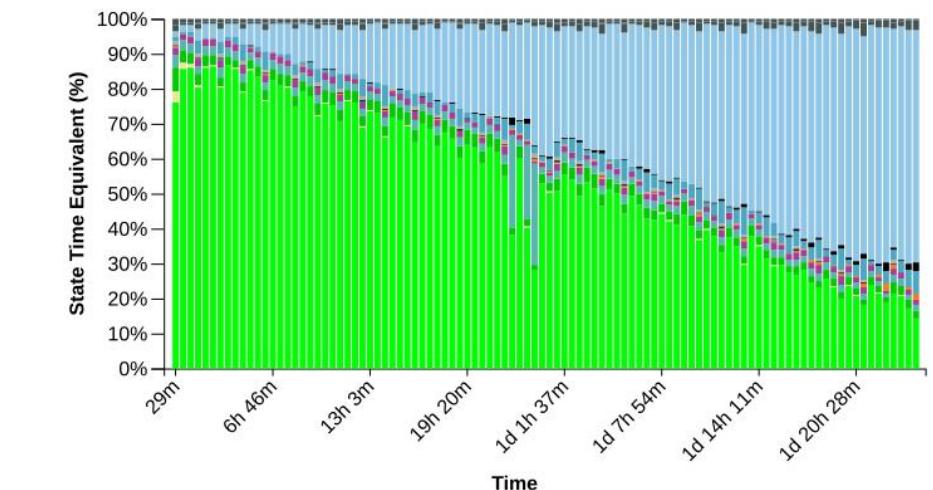


Read Length Histogram

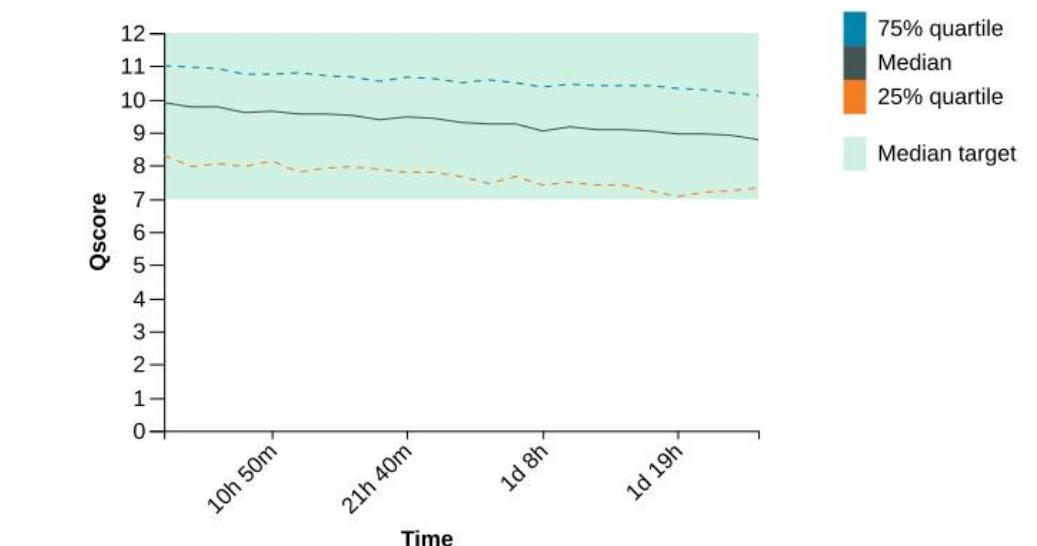
Summary read length distribution



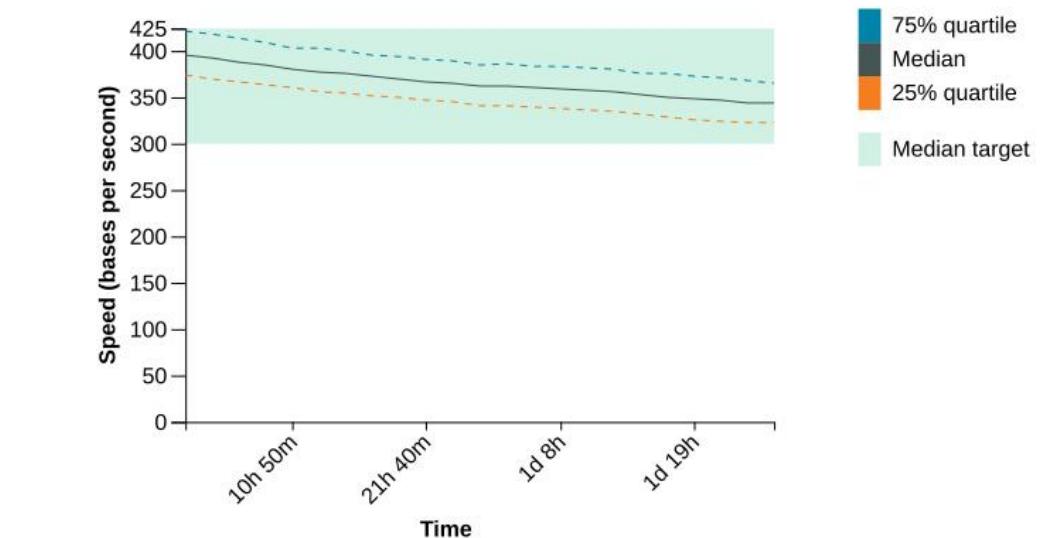
Duty time Categorised



QScore

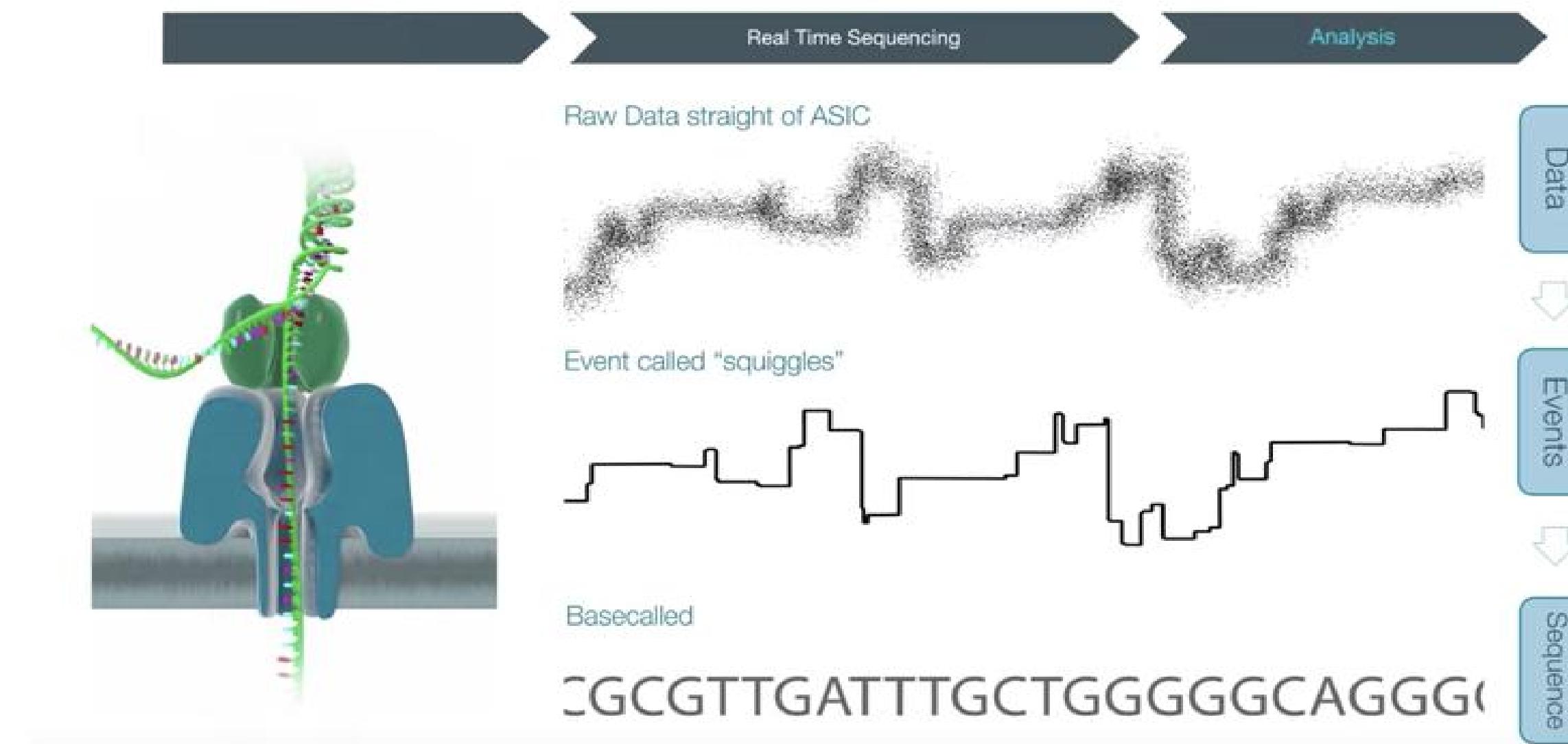


Translocation Speed



# Data monitoring through fast basecalling

This is the process of translating raw electrical signal data from an ON sequencer to DNA sequence. Basecalling is a critical step in the analysis workflow as poor basecalling makes poor sequence data.



## Phred Scores Q

reflects the probability of error in the associated sequence

$Q=10 \rightarrow$  1 chance in 10 that the base call is incorrect  
 $\rightarrow$  90% base call accuracy

$Q=20 \rightarrow$  1 chance in 100 that the base call is incorrect  
 $\rightarrow$  99% base call accuracy

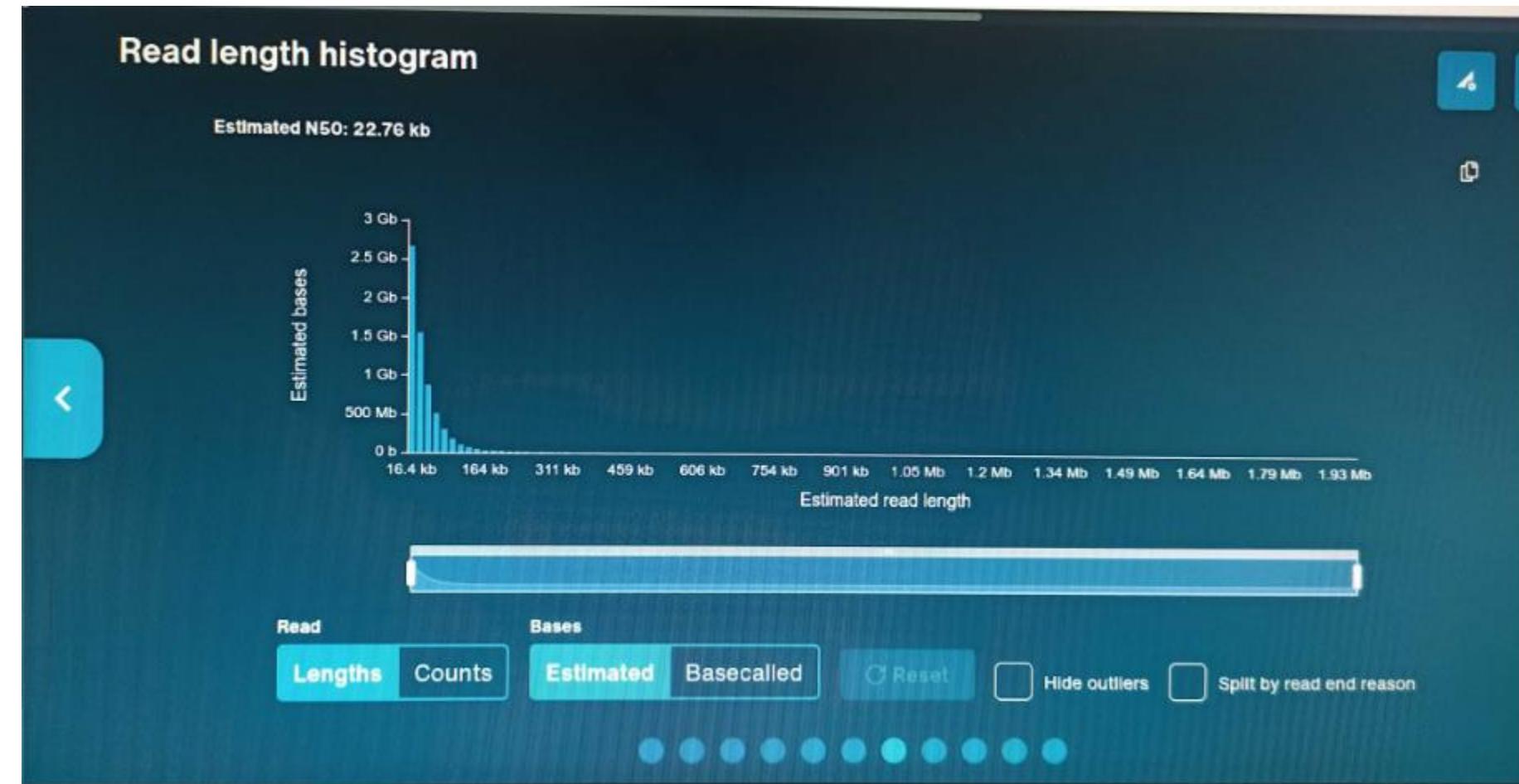
<https://labs.epi2me.io/quality-scores/>

$$\text{Taux d'erreur} = 10^{-\frac{Q}{10}}$$

Dans cette formule,  $Q$  est le score de qualité, donc pour le Q20 :

$$\text{Taux d'erreur} = 10^{-\frac{20}{10}} = 10^{-2} = 0.01$$

## N50

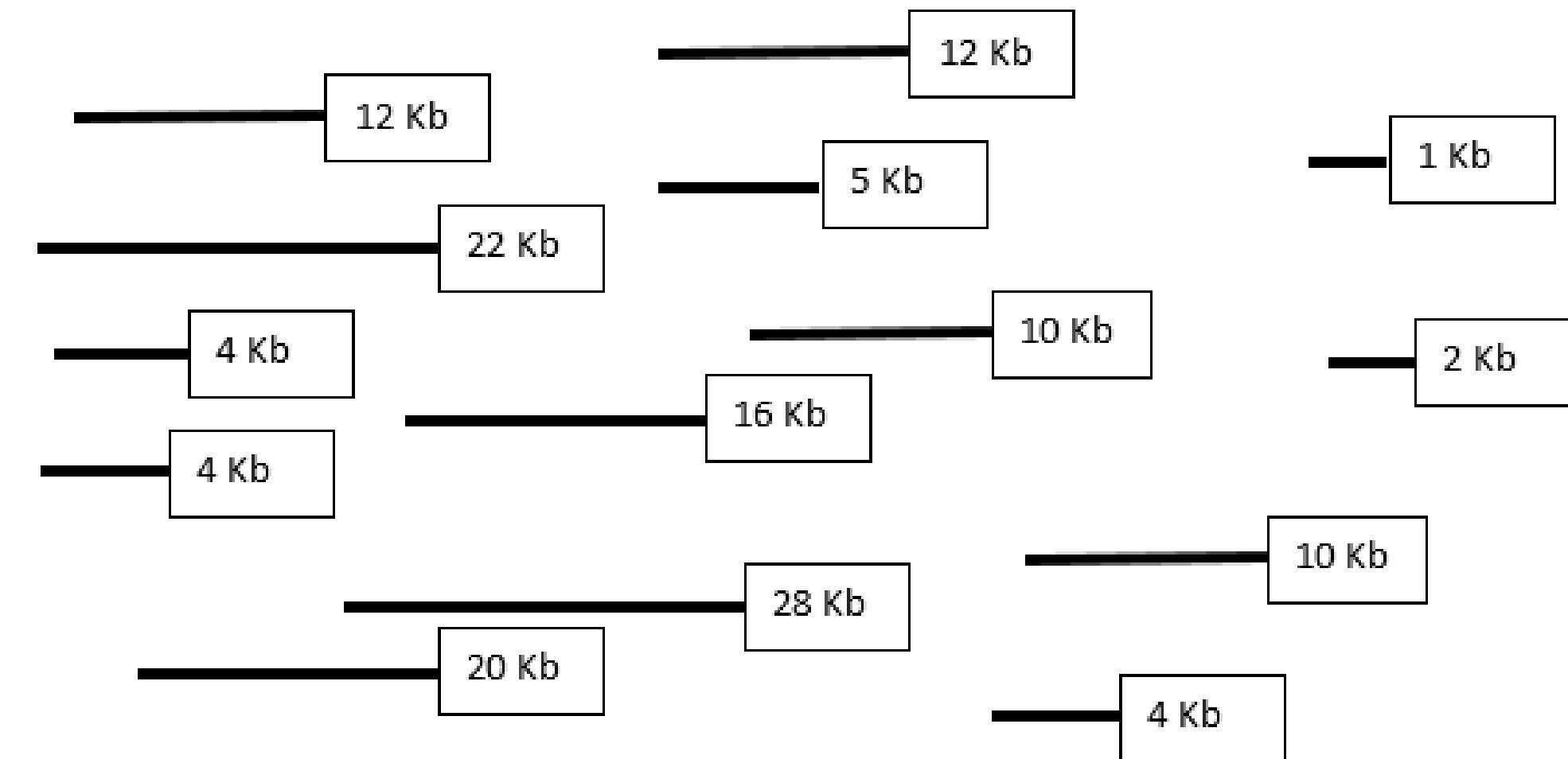


$N50 = 20 \text{ kb}$

Half of the sequenced bases  
are in reads that are at least  
20 kb in size

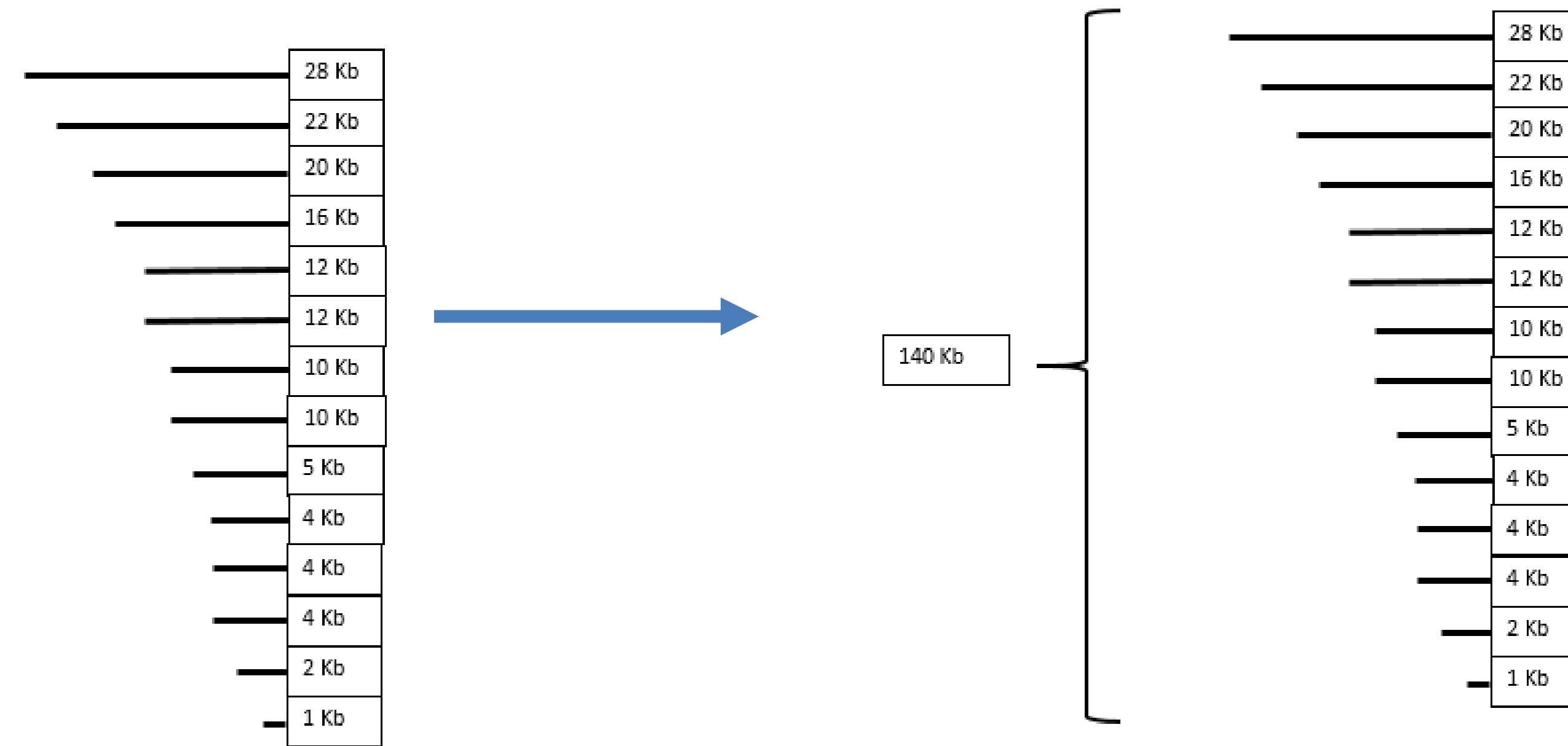
$$N50 = 20 \text{ kb}$$

Half of the sequenced bases are in reads that are at least 20 kb in size



$$N50 = 20 \text{ kb}$$

Half of the sequenced bases are in reads that are at least 20 kb in size



N50 = 20 kb

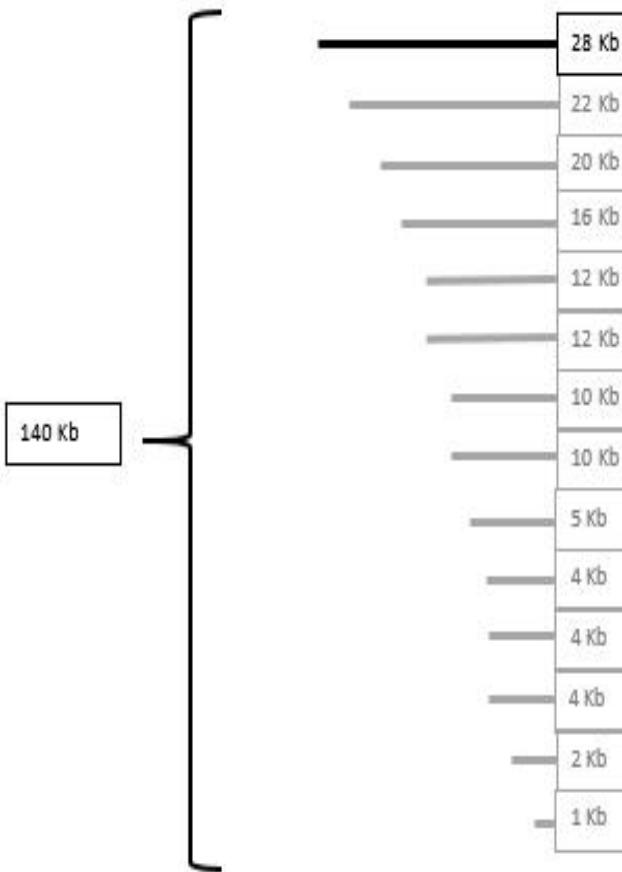
$$140 \text{ Kb} / 2 = 70 \text{ Kb}$$

Which is the read size to reach at least half of the sequenced reads i.e. 70kb?

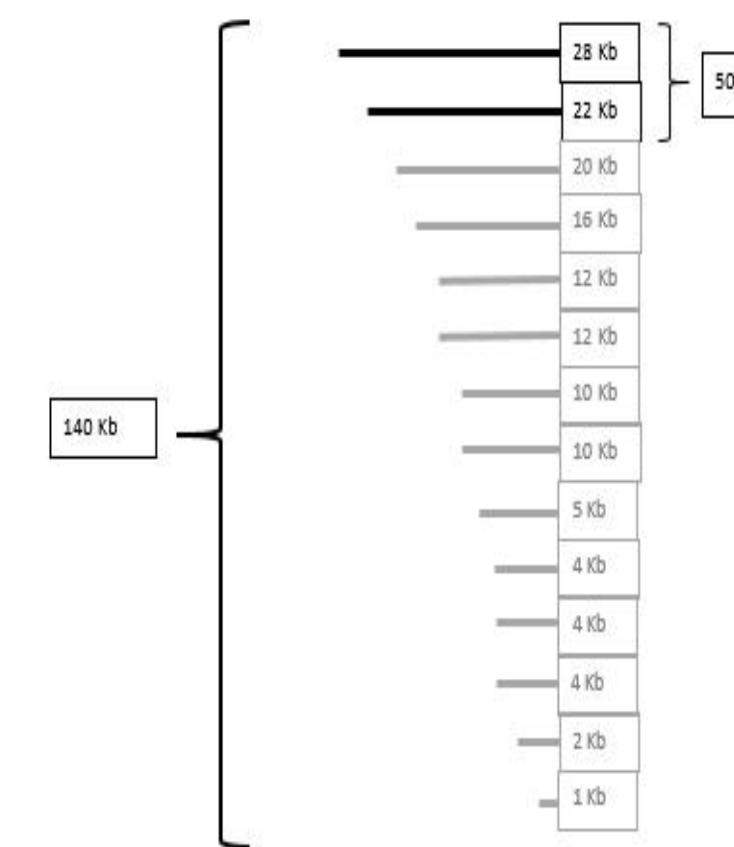
# Data monitoring through fast basecalling

$$140 \text{ kb} / 2 = 70 \text{ kb}$$

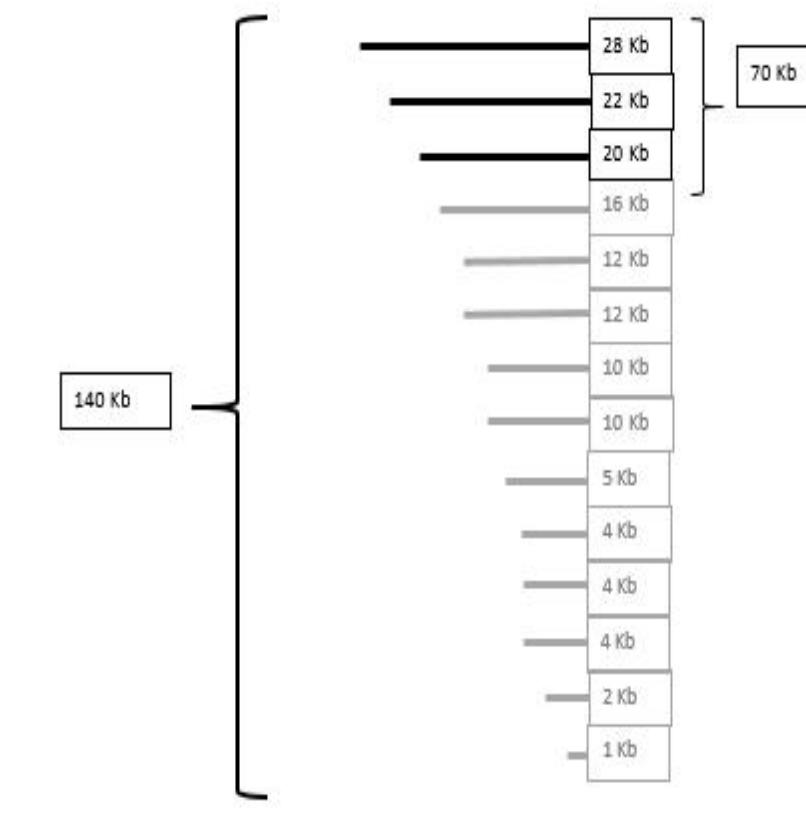
$28 < 70$



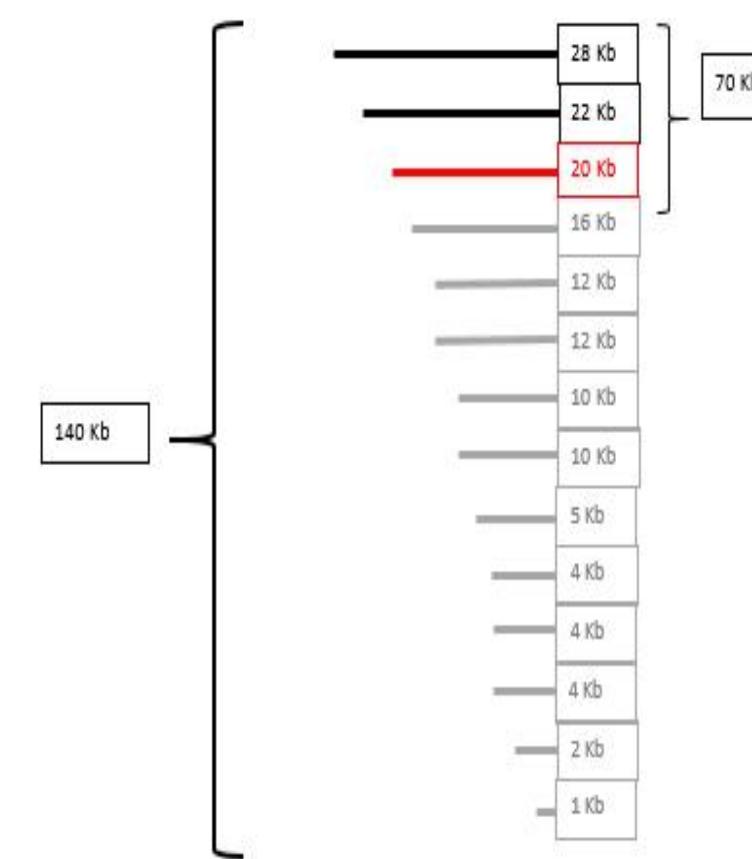
$50 < 70$



$70 \geq 70$

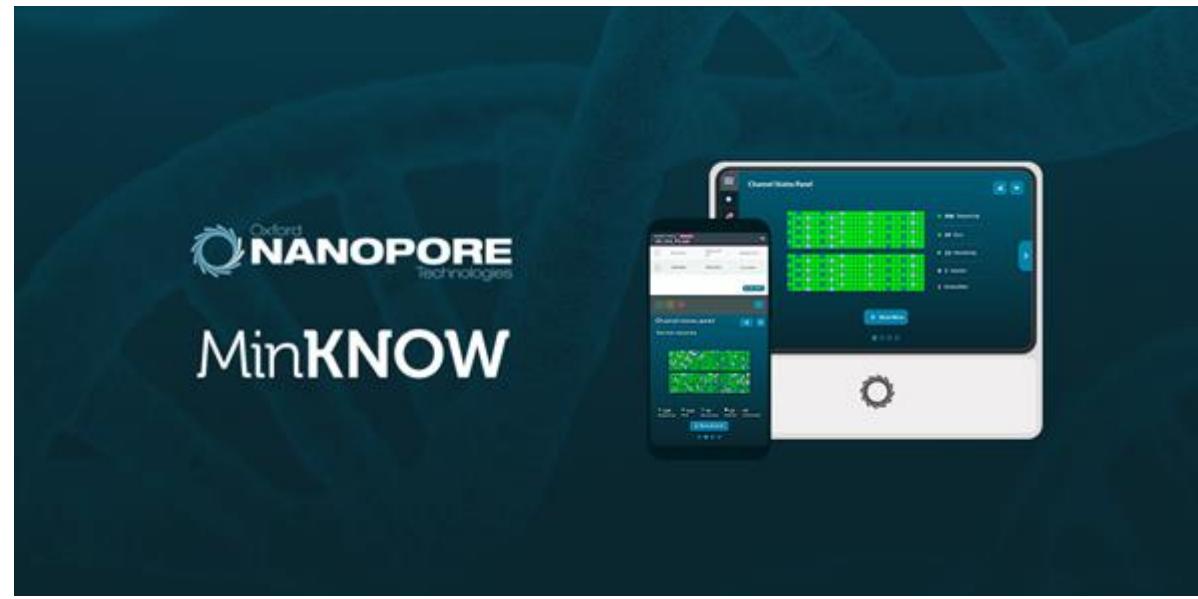


$N50 = 20 \text{ kb}$



What can be done during sequencing?





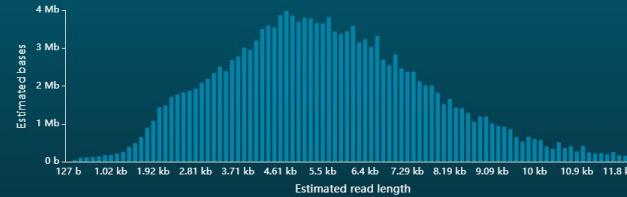
MinKNOW software to view the sequencing

Flow cell status, size and quality of sequences obtained, data volume, sequencing speed, etc...

# Examples

Read length histogram

Estimated N50: 5.35 kb

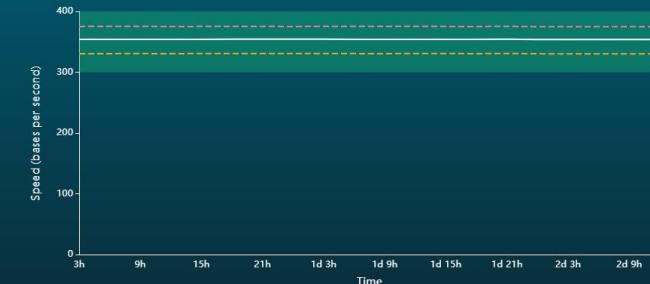


Read Lengths Counts Estimated Basecalled

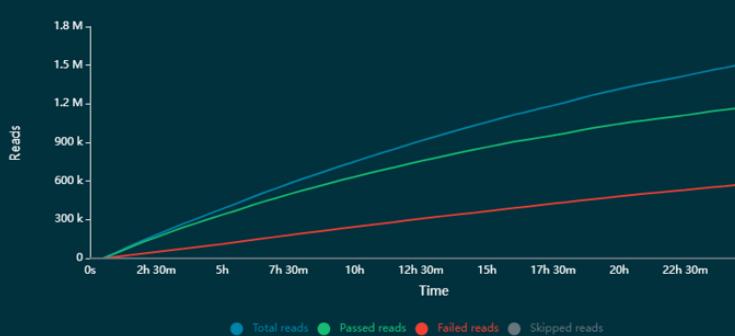
Reset  Hide outliers  Split by read end reason

Translocation speed

75% quartile Median 25% quartile



Cumulative output



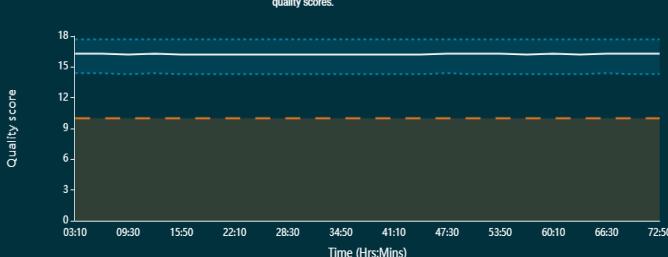
Reads Bases

Quality score

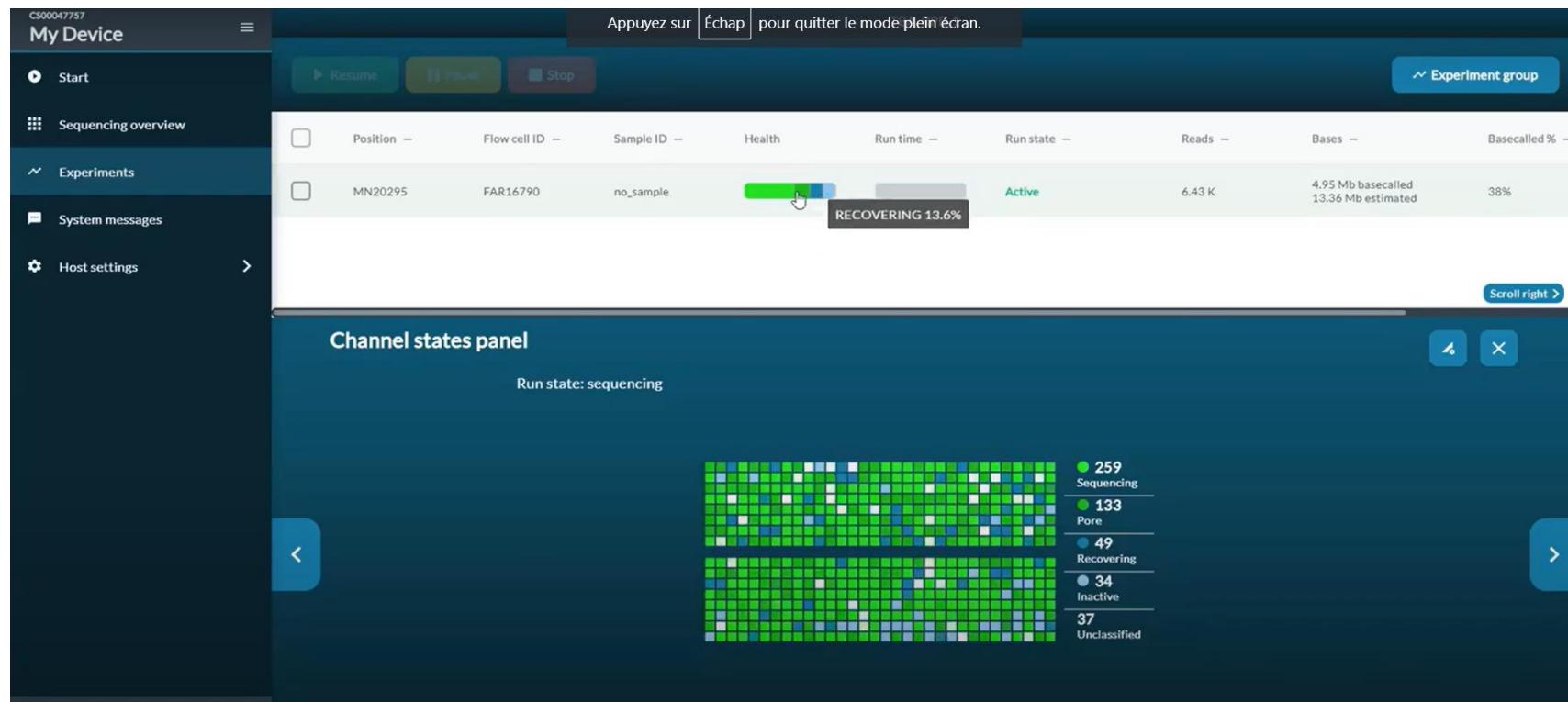
Median —  
The 50th percentile of quality scores.

Interquartile range □  
The spread between the 25th and 75th percentiles of quality scores.

Min. quality score ---  
Minimum quality score for a read to pass.



Mode Median

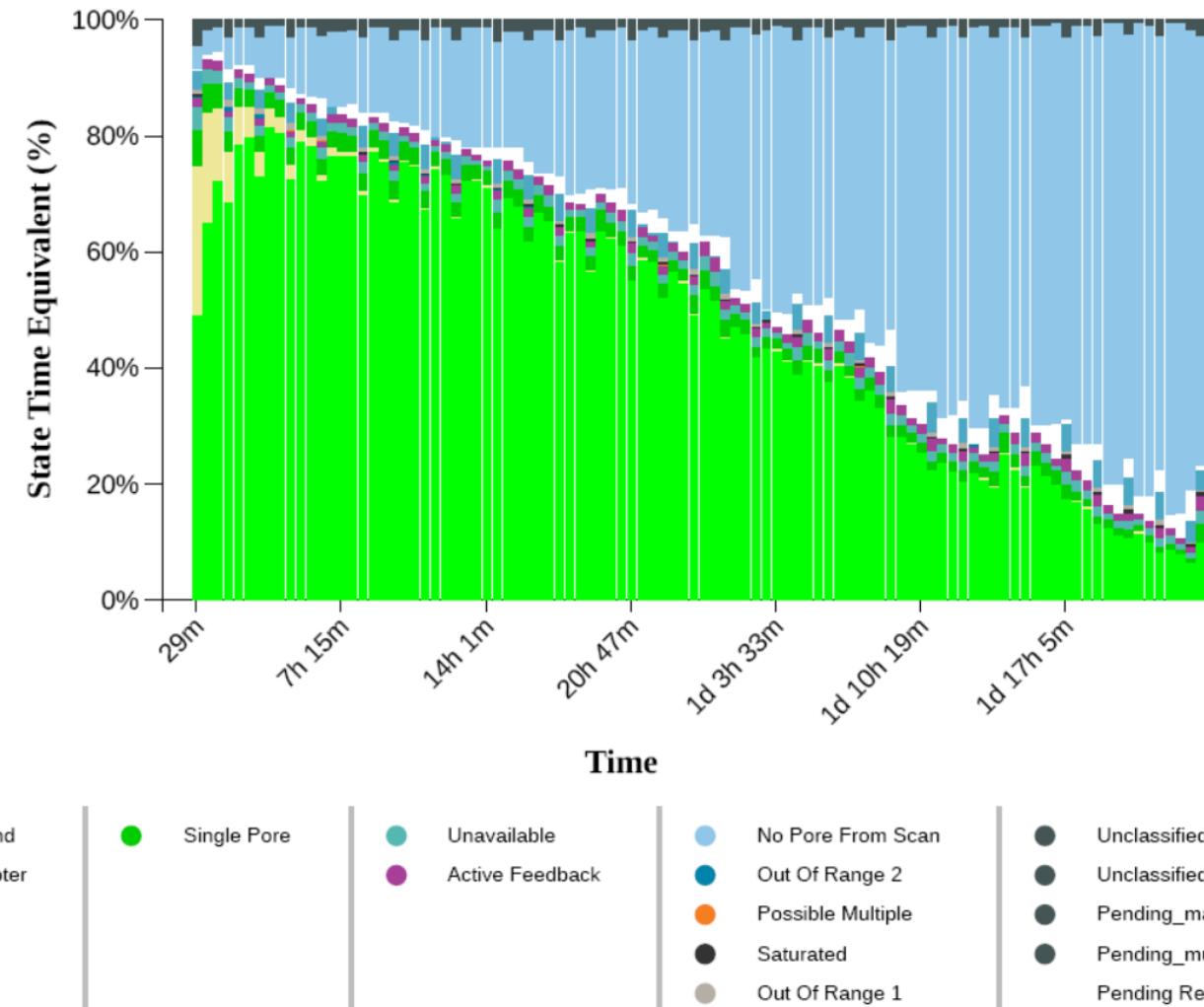


Flow cell health in real time -> number of pores active, waiting, inactive, etc...

Depends on :

- The flow cell (age, storage conditions, integrity, etc.)
- The initial gDNA (quality/purity of extraction, fragment size)
- The construction of the library (DNA fragmentation, loss during purification, assay reliability)

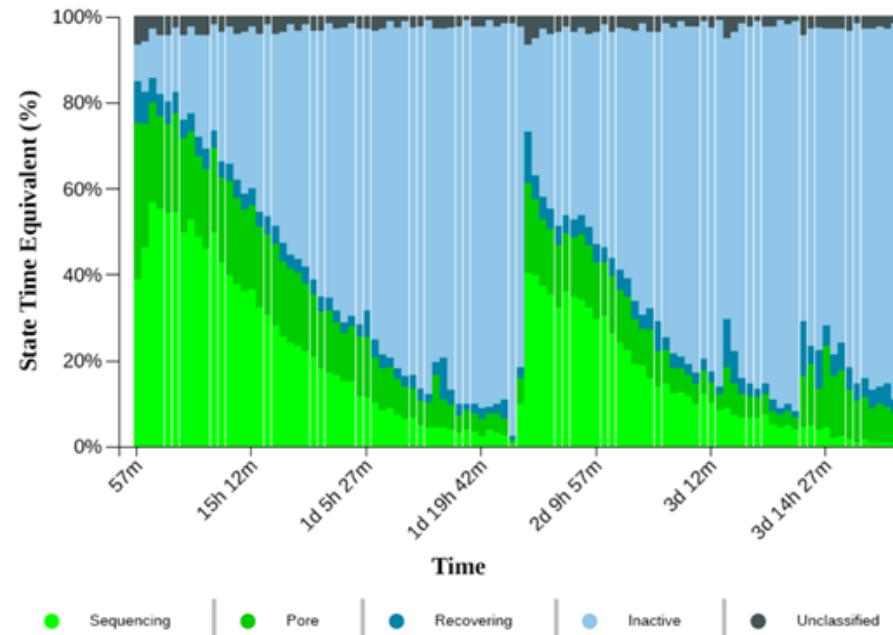
### Duty time Categorised



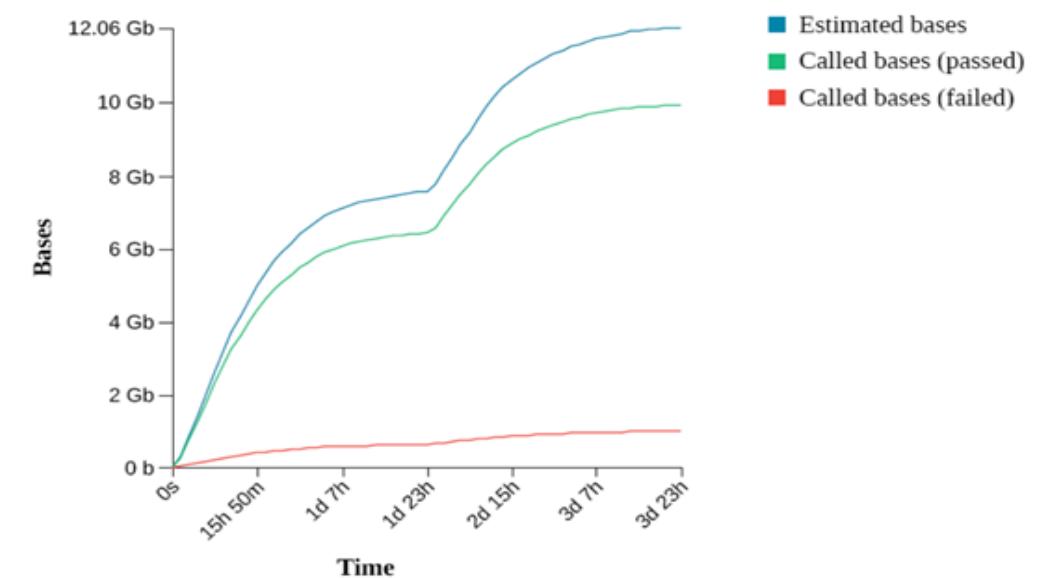
# Optimisation of flow cell yield by making two loads of the same library separated by a DNase wash



Duty Time Grouped

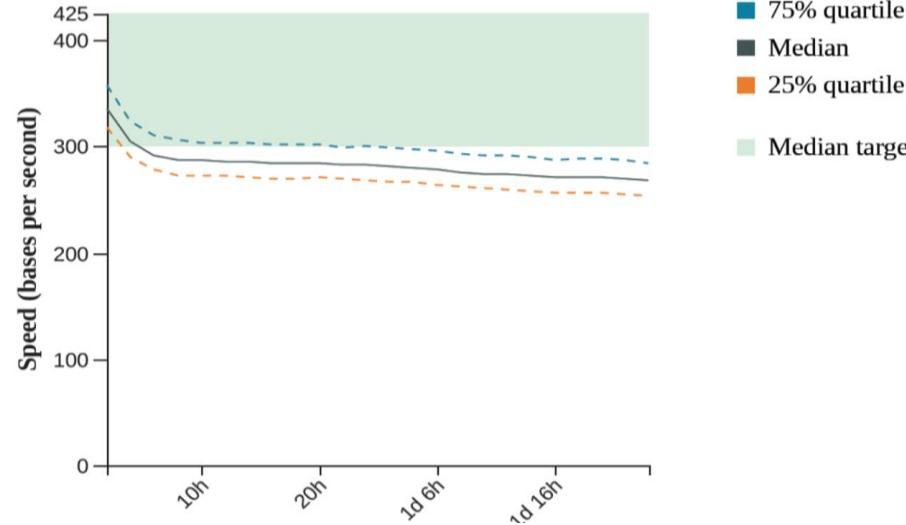


Cumulative Output Bases

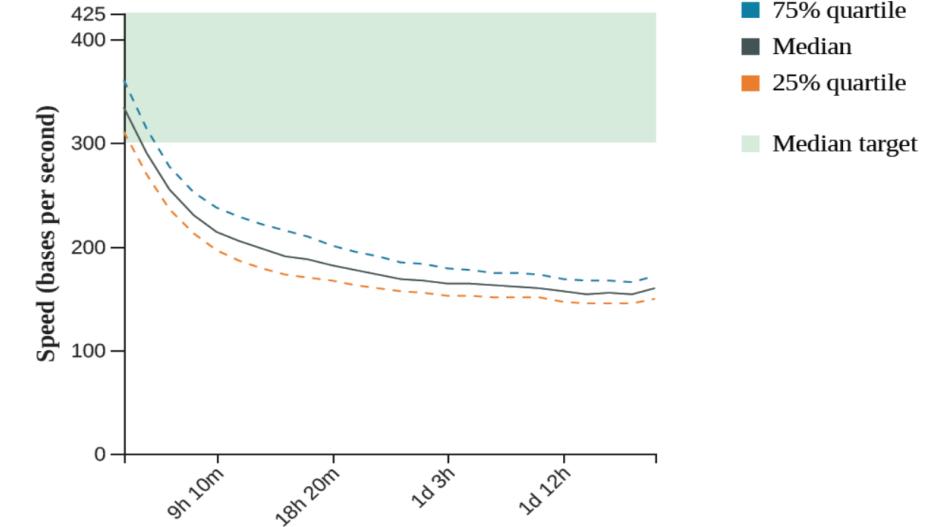


# Loss of translocation speed during sequencing

**Translocation Speed**



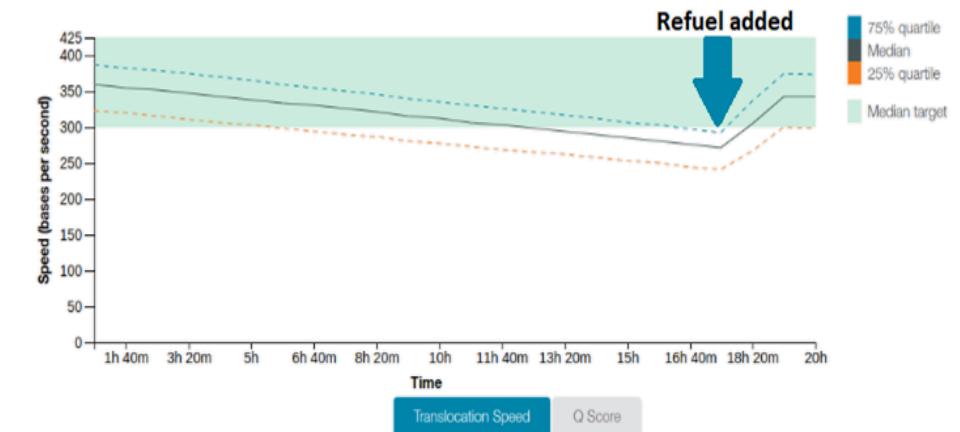
**Translocation Speed**



Loss specific to each run, depending on the flow cell (age, integrity), its occupation, etc...

The addition of Flushing Buffer (FB) during the run compensates for the loss of translocation speed

**Translocation Speed**



# Applications at IRD



François

# Population Genomics/Genomics

---

- Gene discovery/GWAs

# Population Genomics/Genomics

---

- Gene discovery/GWAs
- Species Definition

# Population Genomics/Genomics

---

- Gene discovery/GWAs
- Species Definition
- Subspecies/specific subgroup definition

# Population Genomics/Genomics

---

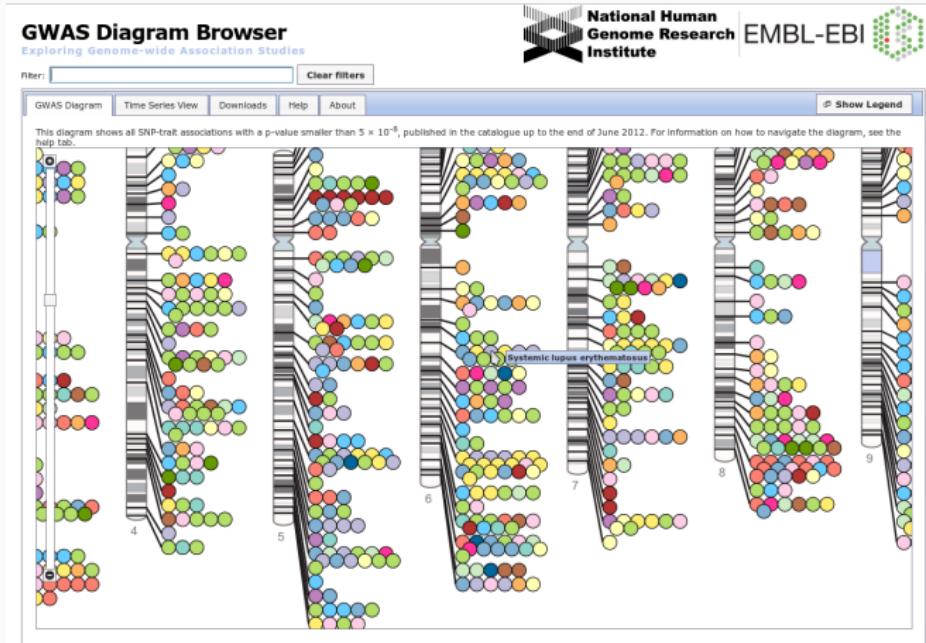
- Gene discovery/GWAs
- Species Definition
- Subspecies/specific subgroup definition
- Global genotyping (for breeding in agriculture e.g.)

# Population Genomics/Genomics

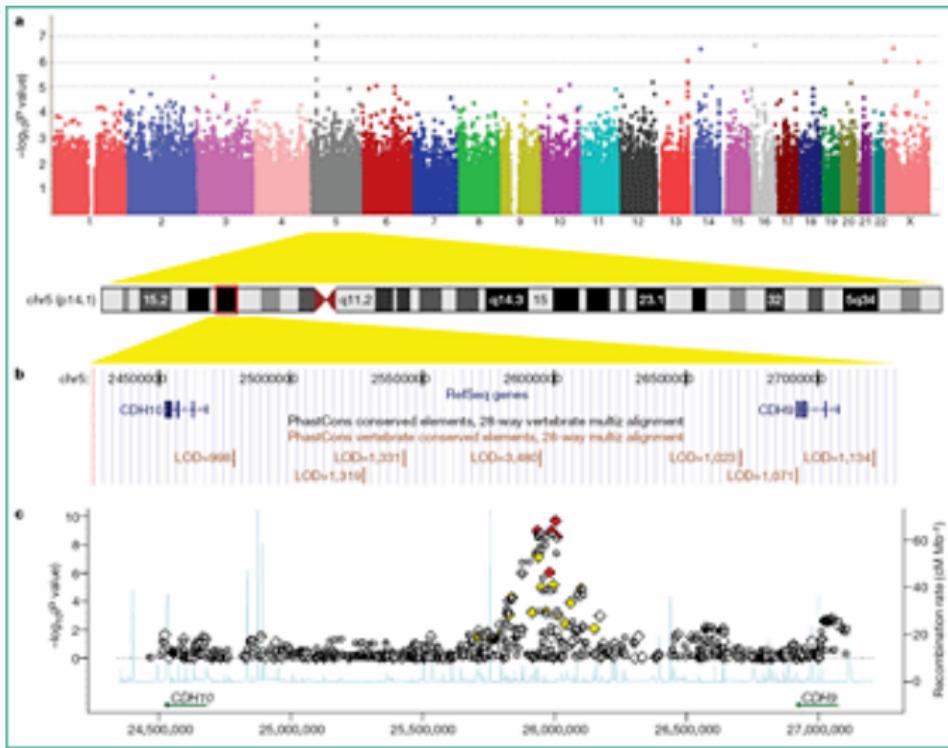
---

- Gene discovery/GWAs
- Species Definition
- Subspecies/specific subgroup definition
- Global genotyping (for breeding in agriculture e.g.)
- Genomic Ecology (Transposable elements, etc...)

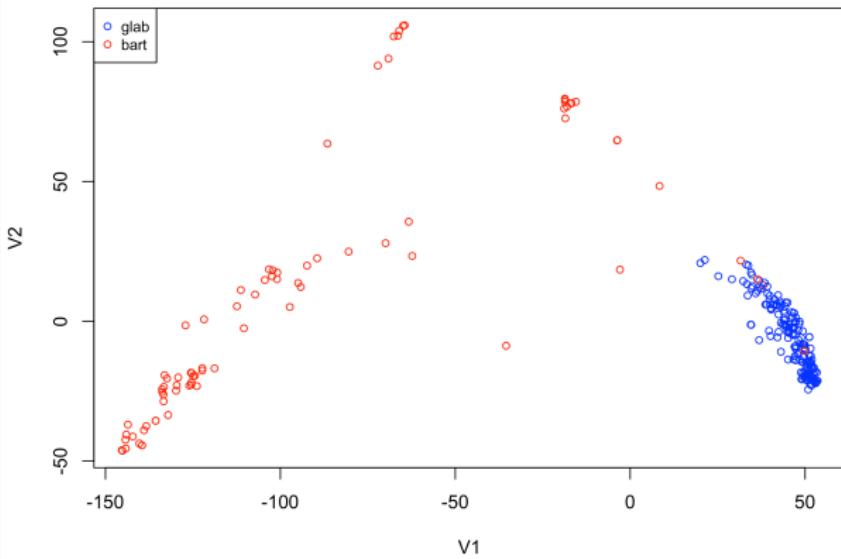
# Example in GWAs & Population Genomics



# Example in GWAs & Population Genomics

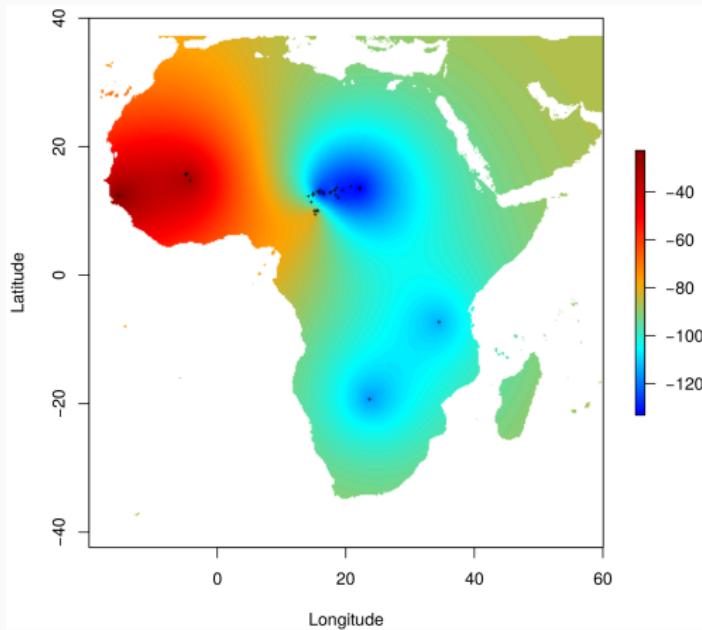


# Example in Global Genotyping & Population Genomics



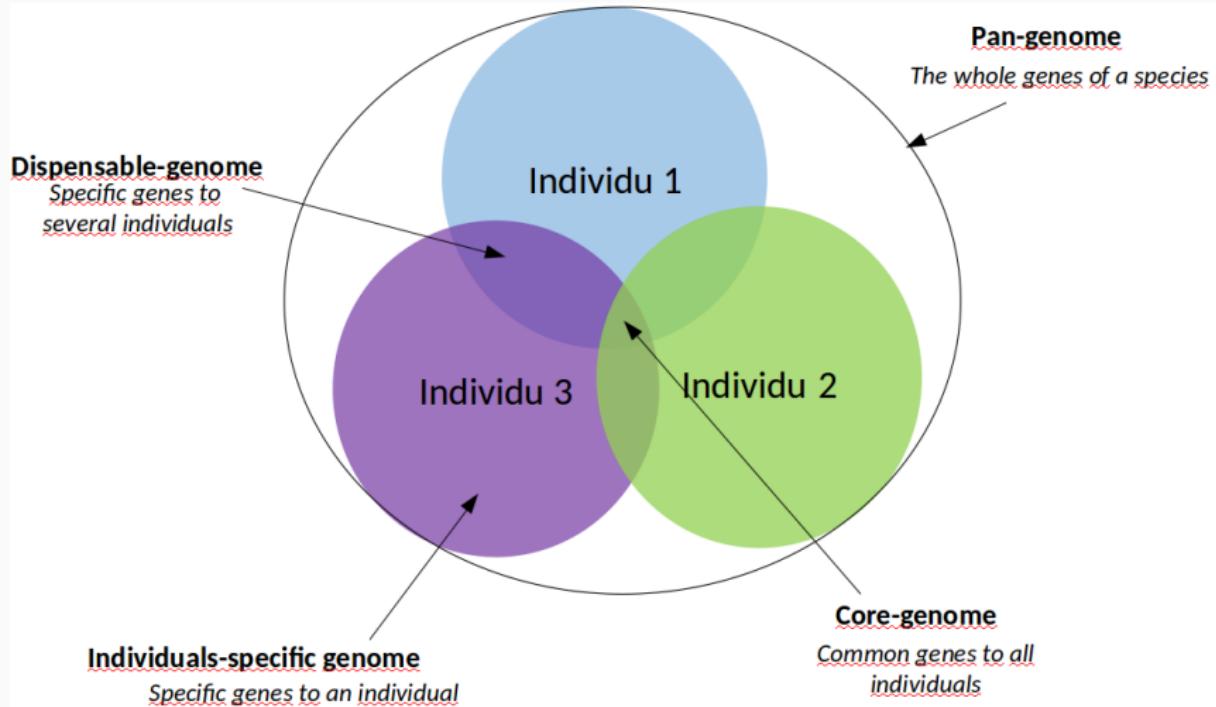
From Cubry et al, 2018

# Example in Global Genotyping & Population Genomics



From Cubry et al, 2018

# Pangenomic



From C. Monat

# Transcriptomics

---

- Level of expression in different conditions or in different individuals

# Transcriptomics

---

- Level of expression in different conditions or in different individuals
- Variation in sequences

# Transcriptomics

---

- Level of expression in different conditions or in different individuals
- Variation in sequences
- Variation of splicing

# Transcriptomics

---

- Level of expression in different conditions or in different individuals
- Variation in sequences
- Variation of splicing
- Variation of editing

# Transcriptomics

---

- Level of expression in different conditions or in different individuals
- Variation in sequences
- Variation of splicing
- Variation of editing
- Detection of putative coding/active sequence

# smallRNA Transcriptomics

---

- Level of expression in different conditions or in different individuals

# smallRNA Transcriptomics

---

- Level of expression in different conditions or in different individuals
- Variation in sequences

# smallRNA Transcriptomics

---

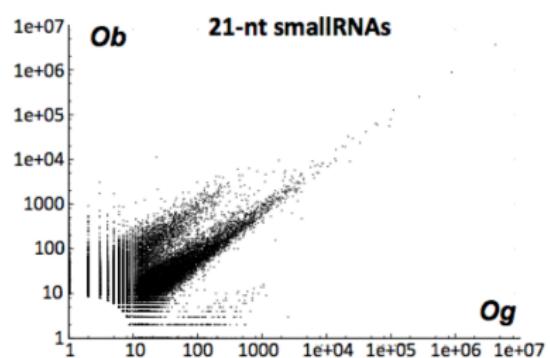
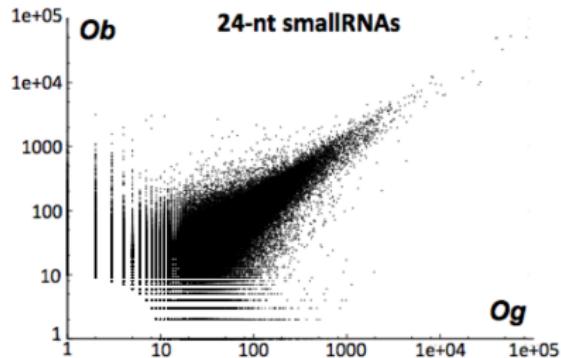
- Level of expression in different conditions or in different individuals
- Variation in sequences
- Variation in specific forms

# smallRNA Transcriptomics

---

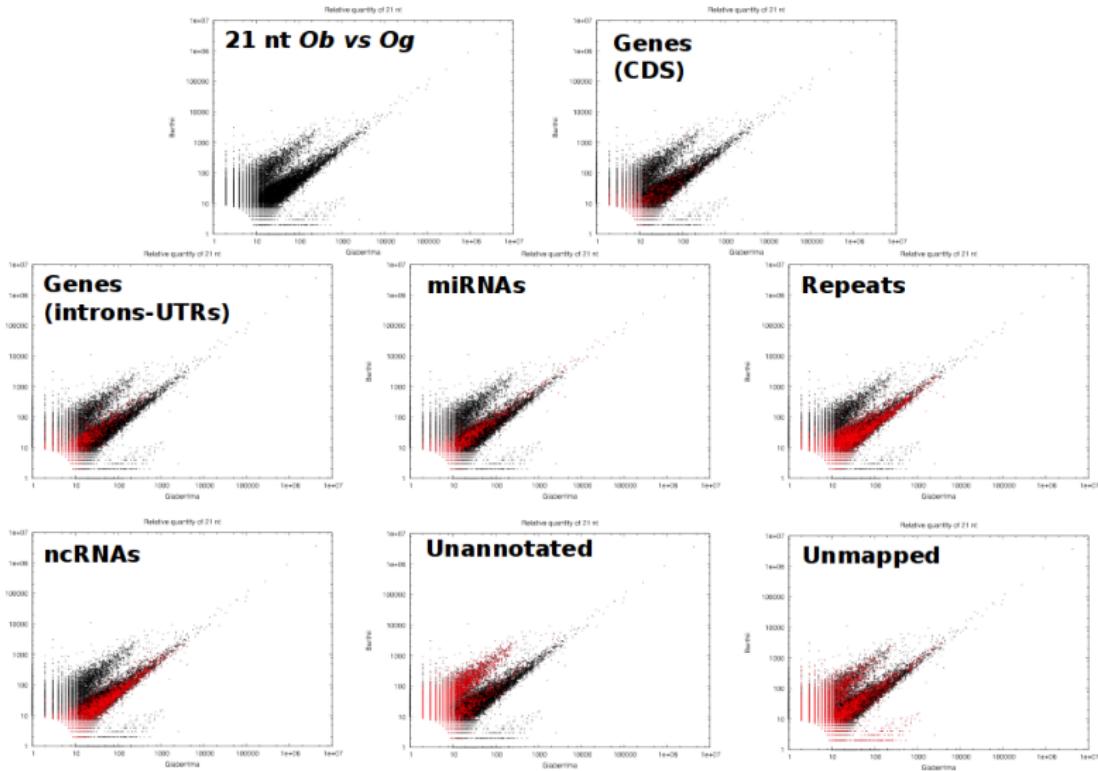
- Level of expression in different conditions or in different individuals
- Variation in sequences
- Variation in specific forms
- Detection of new forms

# Example in smallRNA Transcriptomics

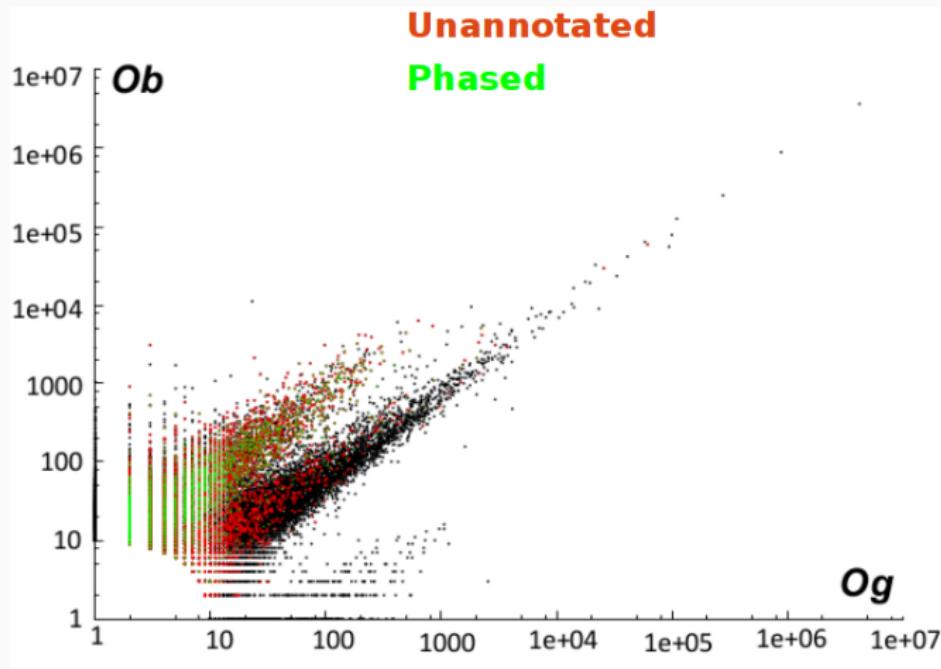


From Ta et al, 2015

# Example in smallRNA Transcriptomics



## Example in smallRNA Transcriptomics



From Ta et al, 2015

## Diagnostic/Pathology

---

- Pre-diagnostic (Genetic illness, putative resistance)

## Diagnostic/Pathology

---

- Pre-diagnostic (Genetic illness, putative resistance)
- Tumor sequencing

## Diagnostic/Pathology

---

- Pre-diagnostic (Genetic illness, putative resistance)
- Tumor sequencing
- Viral sequencing

## Diagnostic/Pathology

---

- Pre-diagnostic (Genetic illness, putative resistance)
- Tumor sequencing
- Viral sequencing
- Risk Assessment

# Diagnostic/Pathology

---

- Pre-diagnostic (Genetic illness, putative resistance)
- Tumor sequencing
- Viral sequencing
- Risk Assessment
- Epidemiological Studies

# Metagenomics at large

## THE METAGENOMICS PROCESS



Extract all DNA from  
microbial community in  
sampled environment

### DETERMINE WHAT THE GENES ARE

#### (Sequence-based metagenomics)

- Identify genes and metabolic pathways
- Compare to other communities
- and more...

### DETERMINE WHAT THE GENES DO

#### (Function-based metagenomics)

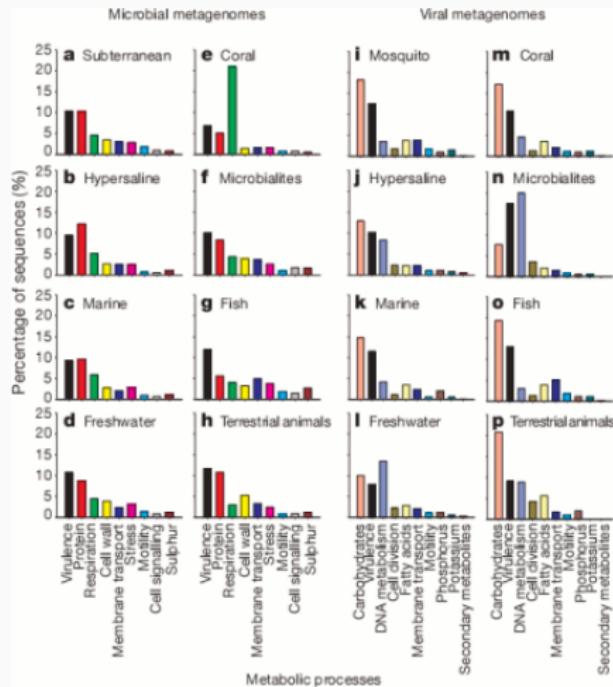
- Screen to identify functions of interest, such as vitamin or antibiotic production
- Find the genes that code for functions of interest
- and more...

# Large Metagenomic assays



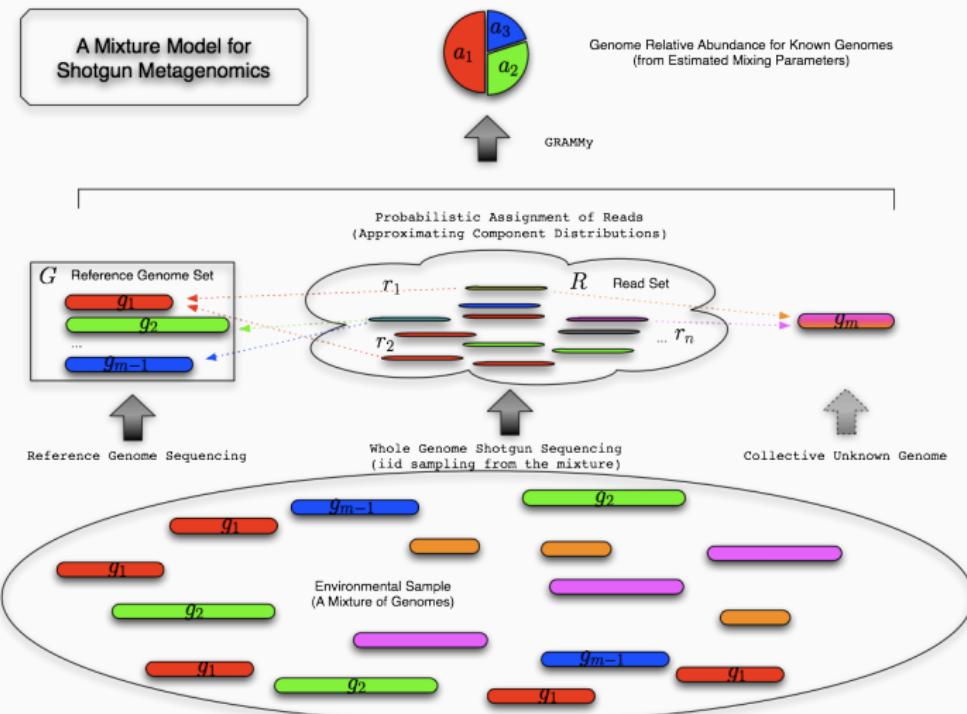
From Tara Ocean website

# Functional Metagenomics



From Dinsdale et al, 2008

# Barcoding



# Large Projects

The image shows a screenshot of the 1000 Genomes Project website, featuring two versions of the homepage side-by-side.

**Left Version (Older Layout):**

- Title:** 1000 Genomes  
A Deep Catalog of Human Genetic Variation
- Navigation:** Home, About, Data, Analysis, Participants, Contact, Bio
- Section:** LATEST ANNOUNCEMENTS
- Headline:** WEDNESDAY FEBRUARY 16, 2011  
**February 2011 Data Up**  
**Full Project Indel Release**
- Text:** Indels calls from [Dindel](#). These calls genome project. This release is ba
- Link:** Data access links: [EBI](#) / [NCBI](#)

**Right Version (Newer Layout):**

- Title:** 1001 Genomes  
A Catalog of *Arabidopsis thaliana* Genetic Variation
- Navigation:** Home, Collaborators, Accessions, Tools, Software, Data Center, Gallery, About, Help desk
- Section:** Welcome to the 1001 Genomes Project
- Header Buttons:** Database & Species lists, News, Events, Publications, Participants, For G10K Organizers (restricted)
- Search Bar:** Search:  Go
- Background Image:** A blue-toned image of DNA helixes and various animal species.
- Text:** GENOME 10K®  
Unveiling animal diversity
- Section:** Genome 10K Project
- Text:** To understand how complex animal life evolved through changes in DNA and use this knowledge to become better stewards of the planet.
- Text:** April 2009—The Genome 10K project aims to assemble a genomic zoo—a collection of DNA sequences representing the genomes of 10,000 vertebrate species, approximately one for every vertebrate genus. The trajectory of cost reduction in DNA sequencing suggests that this project will be feasible within a few
- Call-to-Action:** Join us  
Become a G10K affiliate
- Section:** Genome assembly

# Possibilities in the next 5-10 years (From a presentation in 2013)

---

- Real-time Transcriptomics

# Possibilities in the next 5-10 years (From a presentation in 2013)

---

- Real-time Transcriptomics
- Single-Cell Genomics -> DONE in 2014

## Possibilities in the next 5-10 years (From a presentation in 2013)

---

- Real-time Transcriptomics
- Single-Cell Genomics -> DONE in 2014
- Single-Cells Transcriptomics (and smallRNA) -> DONE in 2015

# Possibilities in the next 5-10 years (From a presentation in 2013)

---

- Real-time Transcriptomics
- Single-Cell Genomics -> DONE in 2014
- Single-Cells Transcriptomics (and smallRNA) -> DONE in 2015
- Personal Genomics medicine (ethical problems...) -> Available

# Possibilities in the next 5-10 years (From a presentation in 2013)

---

- Real-time Transcriptomics
- Single-Cell Genomics -> DONE in 2014
- Single-Cells Transcriptomics (and smallRNA) -> DONE in 2015
- Personal Genomics medicine (ethical problems...) -> Available
- And any new ideas you will have...

# This is the end...

---

- NGS technologies change the way of abording Biology

# This is the end...

---

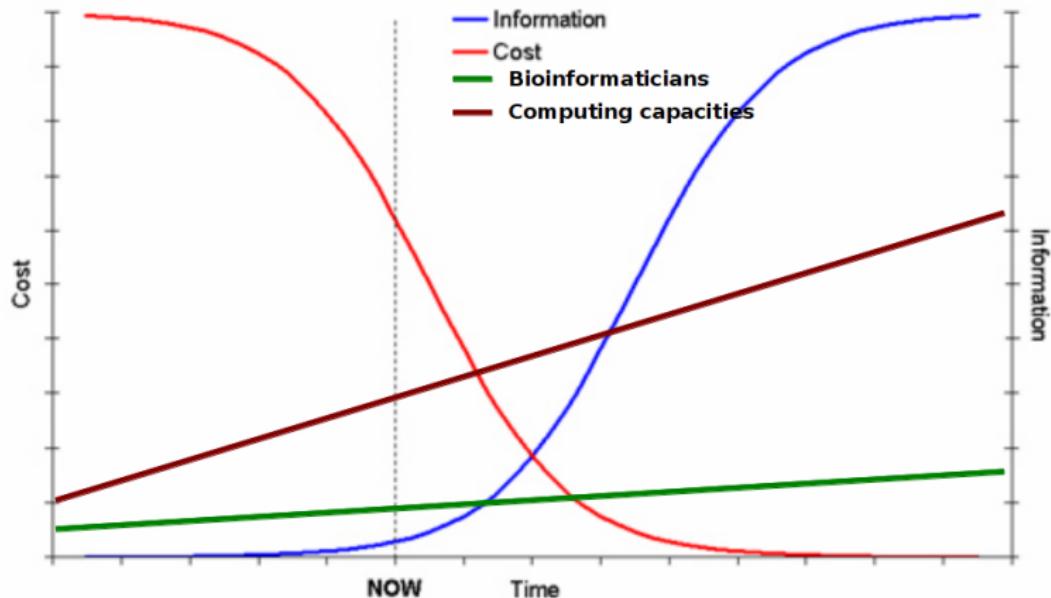
- NGS technologies change the way of abording Biology
- A lot of Possibilities, a lot of limits

# This is the end...

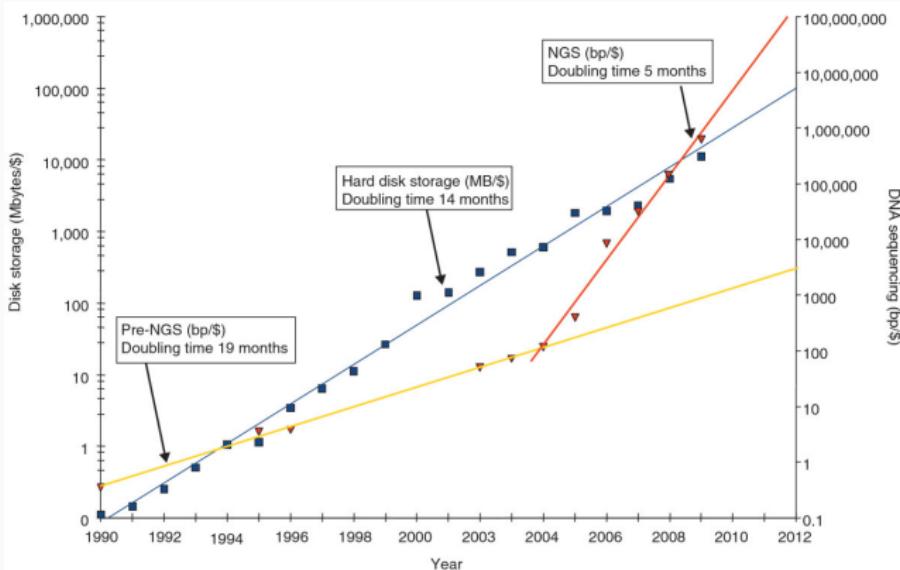
---

- NGS technologies change the way of abording Biology
- A lot of Possibilities, a lot of limits
- The main limit is no more Sequence, but Sample acquisition and Data treatment

## Keep in mind!

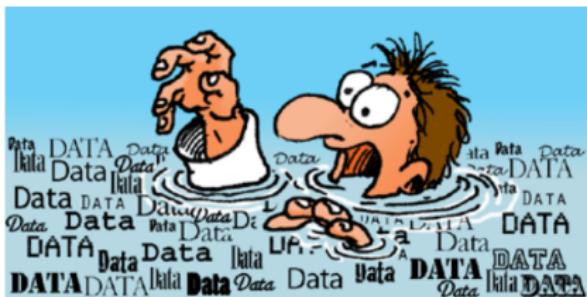


# ...From Data Rarity to Data Deluge



From L. Stein, 2010

# Be Careful to data drowning!



# Thanks for your attention!



Pedagogic material used for these teaching is available under the Creative Common Licence CC-BY-NC-SA 4.0 International - No Commercial Use - Same sharing conditions:

<http://creativecommons.org/licenses/by-nc-sa/4.0/>