GenEpi-BioTrain – Virtual training 17 – Nanopore sequencing technology

Introduction to Nanopore technology

Chloé Baum, PhD
Biomics, Institut Pasteur
April 09th 2025

Intended Learning Objectives

Specific objectives of this session:

1. Understand the basics of Nanopore sequencing
2. Learn about Nanopore sequencing applications
3. Key considerations to start with Nanopore sequencing

Outline

This session consists of the following elements:

1. Brief introduction to sequencing
2. Overview of the Biomics core facility
3. The basics of Nanopore sequencing
4. Applications of Nanopore sequencing
5. Challenges and future directions
6. Q&A session

Have you worked with any sequencing technologies before?

1. Brief introduction to sequencing

What is DNA/RNA sequencing?

- Sequencing is the process of reading the order of nucleotides (A, T, C, G or U) in a DNA or RNA molecule.

- "Reading the code of life" : it tells us about the genes present and how they might function.

- Over time, technologies have been developped to read DNA, and they continue to evolve!

A quick history of sequencing

Sanger Sequencing (1977): The first method, accurate but slow and low-throughput.

Next-Generation Sequencing (NGS): Introduced in the 2000s; massively parallel, faster and cheaper; but produces short reads (100–300 bp).

Third-Generation / Long-Read Sequencing:

- Technologies like PacBio and Oxford Nanopore allow for sequencing of long DNA molecules (even whole chromosomes).
- Particularly useful for structural variants, genome assembly, and complex regions.

Why sequencing is important

- Understanding genetic diseases, identify mutations (cancer, rare diseases)
- Tracking of infectious diseases, monitor outbreaks (COVID-19 pandemic), bacterial antibiotic resistance…
- Environmental monitoring (biodiversity, contaminants, pathogens)
- Studying Microbiomes in the human body, environment, and other ecosystems
- Forensics
- Non-exhaustive list!

Different types of sequencing projects

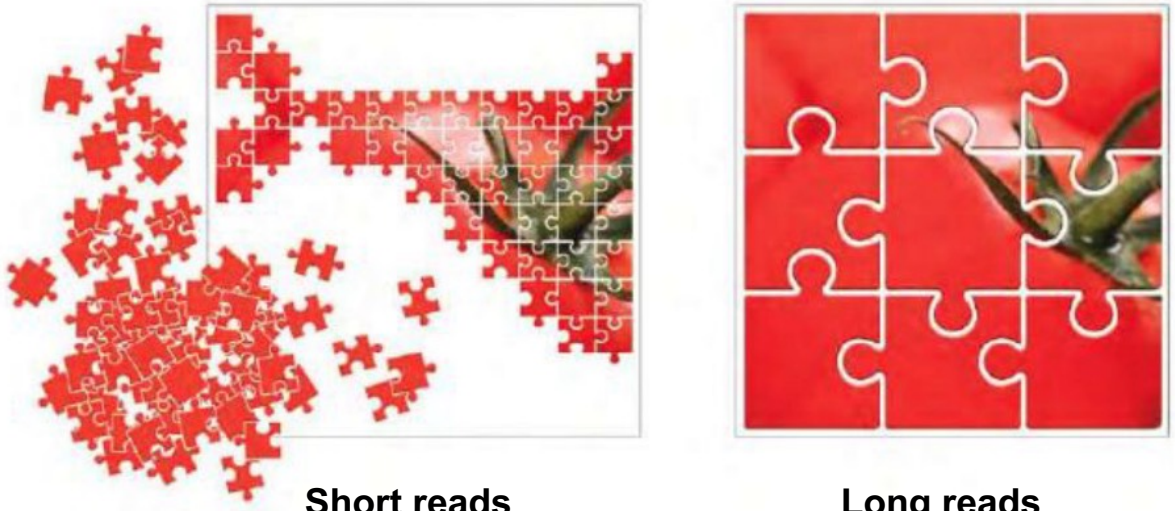
- **Whole Genome Sequencing (WGS)**: Reading the entire genome of an organism
- **RNA sequencing (RNA-seq)**: Capturing gene expression and transcript variants
- **Single-cell sequencing:** Sequencing the genome or transcriptome of individual cells
- **Epigenetic sequencing:** Identifying DNA modifications such as methylation
- **Targeted sequencing**: Focusing on specific genes of interest
- **Metagenomics**: Analyzing mixed DNA from environmental or microbiome samples
- …

Difference between short and long reads

Short-read sequencing (like Illumina)
- 100-300bp
- High accuracy
- Cost effective and high-throuhput
- Challenging for repetitive or complex genomic regions

Long-read sequencing (like Nanopore)
- >10kb (20bp to 4Mb)
- Lower accuracy
- Less cost effective and high-throuhput
- Great for repetitive or complex genomic regions

Example of genome assembly



Short reads
1000 pieces

Long reads
9 pieces

2. Overview of Biomics core facility

Biomics sequencing core facility

3 complementary teams to facilitate scientific discovery through high-throughput sequencing technologies

Head of Biomics:
Marc Monot

Support-Lab

Imène Najjar
Project Manager

Laurence Motreff
Technician

Natacha Hakine
Administrative

Florence Jagorel
Technician

Wet-Lab

HEAD

Yakov Vitrenko
Researcher

Georges Haustant
Technician

Azimdine Habib
Engineer

Elodie Turc
Technician

Laurence Ma
Technician

Chloé Baum
Engineer

Dry-Lab

HEAD

Etienne Kornobis
Engineer

Laure Lemée
Technician

Rania Ouazahrou
Engineer

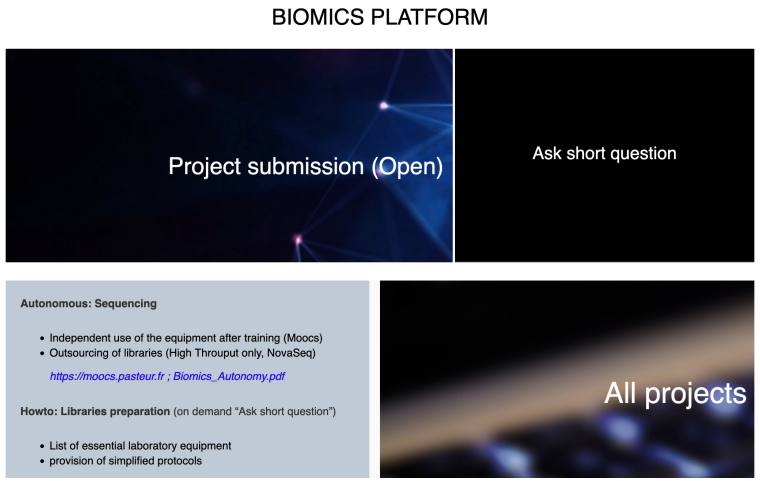
1. **Customized** service per project for research labs (wide range of organisms)

2. **Provide open-access** to equipements and sequencers (training, MOOCs) https://moocs.pasteur.fr/

3. **Development** to keep up with the latest technological advances in NGS

Technological offer at Biomics

- Short-Reads (Illumina/MGI), Long-Reads (PacBio/Nanopore)
- DNA-Seq, RNA-Seq, Single-cell seq, Metagenomics (targeted or shotgun)
- Bioinformatics: NGS data analysis (rna-seq, variant calling, assembly, methylation analysis…)



Dedicated website for project submission:
https://biomics.pasteur.fr/

Equipments in open-access for the campus

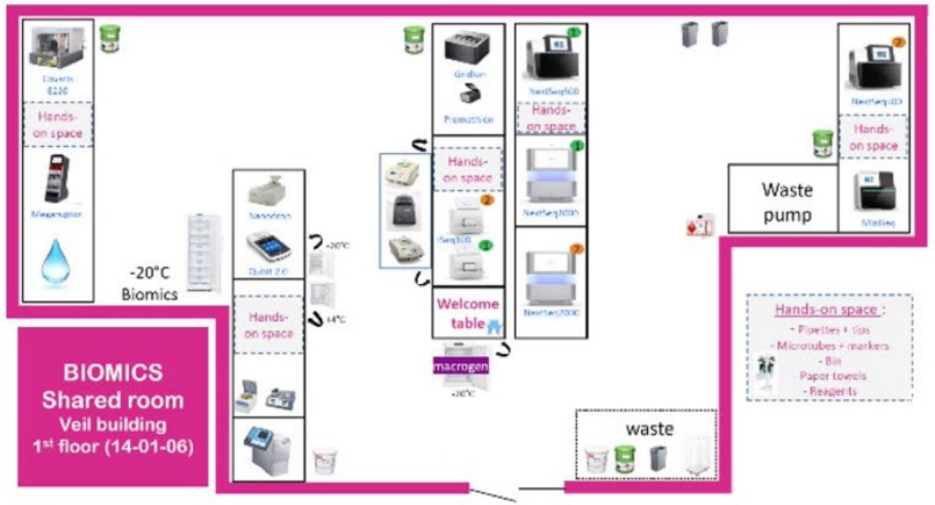
Shared Room

Fragmentation

Nanopore sequencing

Illumina sequencing

Quality Control



Illumina sequencing

Beyond the sequencing projects: capacity building

Objective : Transfer knowledge and help implementing NGS core facilities in the Pasteur network.

PerMediNA
Personalized Medicine in North Africa

MINISTÈRE DE L'EUROPE ET DES AFFAIRES ÉTRANGÈRES
Liberté Égalité Fraternité

2022 - 2024

Algeria, Marocco, Tunisia

AFROSCREEN
Sequencing capacity-building in Africa

AFD
AGENCE FRANÇAISE DE DÉVELOPPEMENT

2021 - 2024

DIPs

Gouvernement Princier
PRINCIPAUTÉ DE MONACO

2018 - 2025

Tunisia

PHIND access

HORIZON 2020

2018 - 2022

Tunisia

Amazed

Guyana, Guadeloupe, New-Caledonia

How familiar are you with Nanopore sequencing?

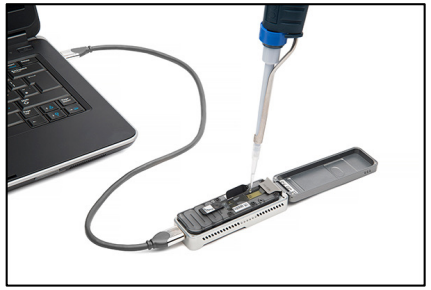
3. The basics of Nanopore sequencing

Nanopore sequencing offers new perspectives

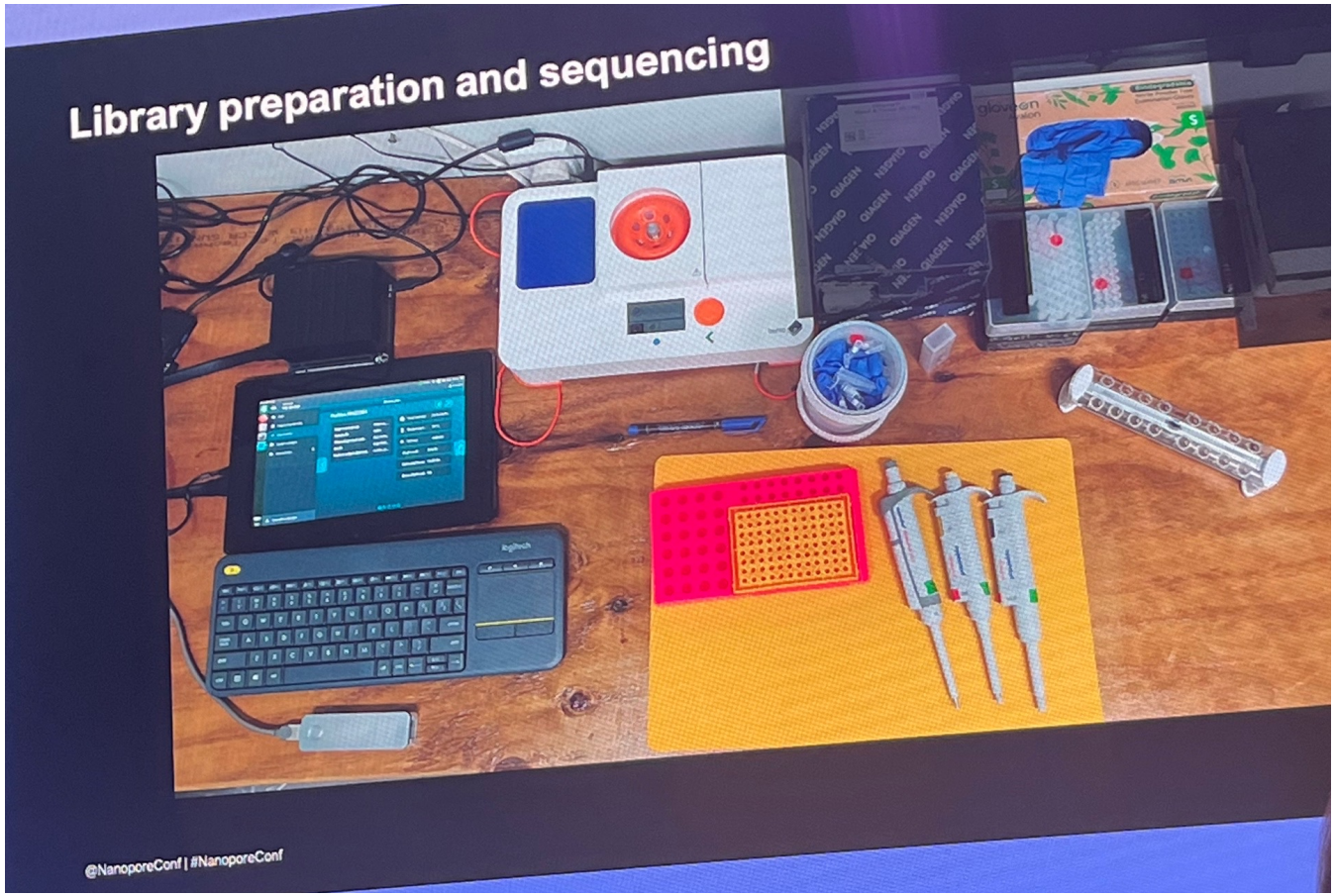
« Sequencing anything, by anyone, anywhere »

Portable, real-time analysis

Zane Libke
Biodiversity research in Ecuadorian tropical andes (Amazonia)



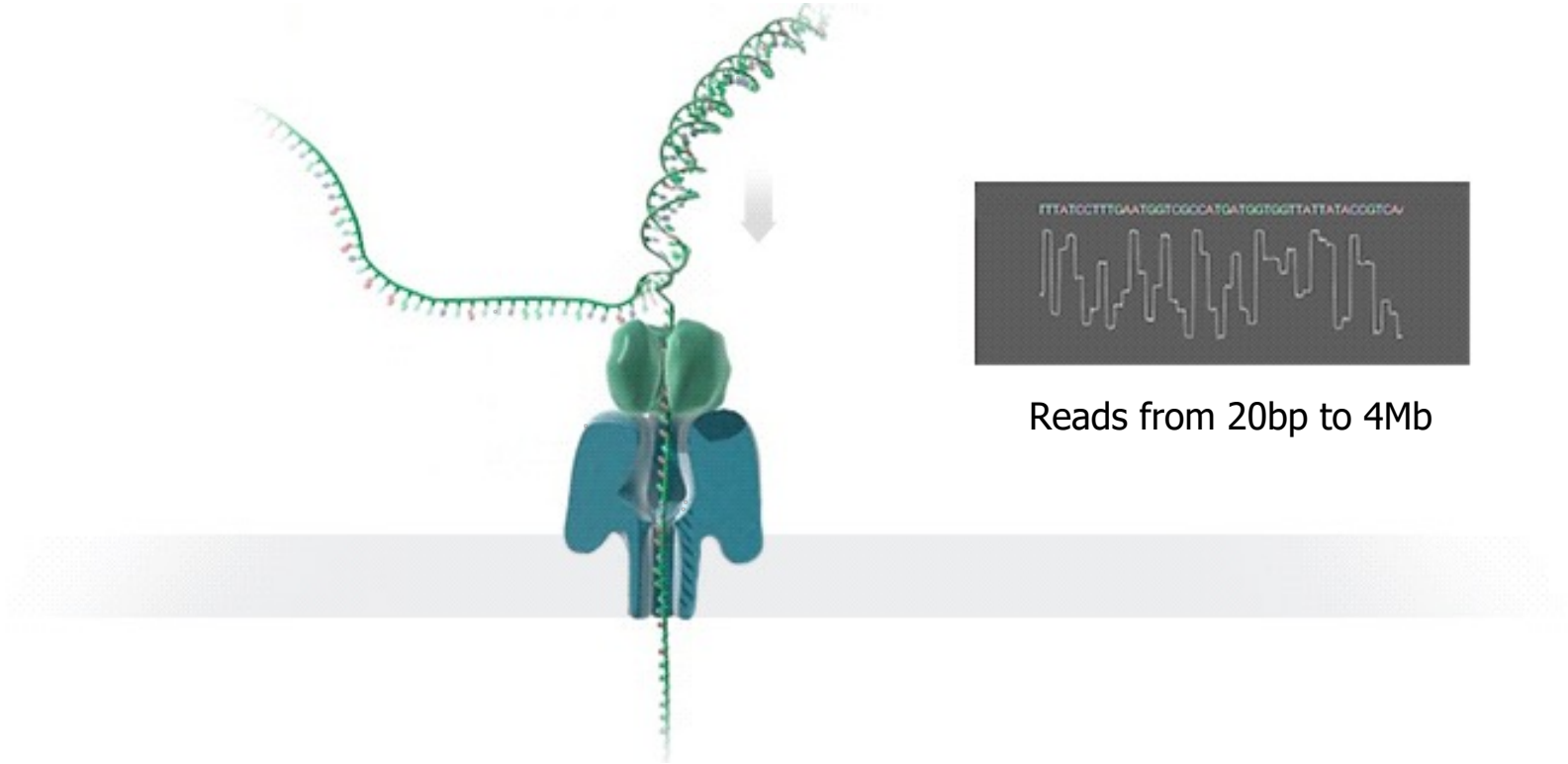
Field-based ONT sequencing



https://bento.bio/devices/

Principle of Nanopore sequencing

Nanopore library structure



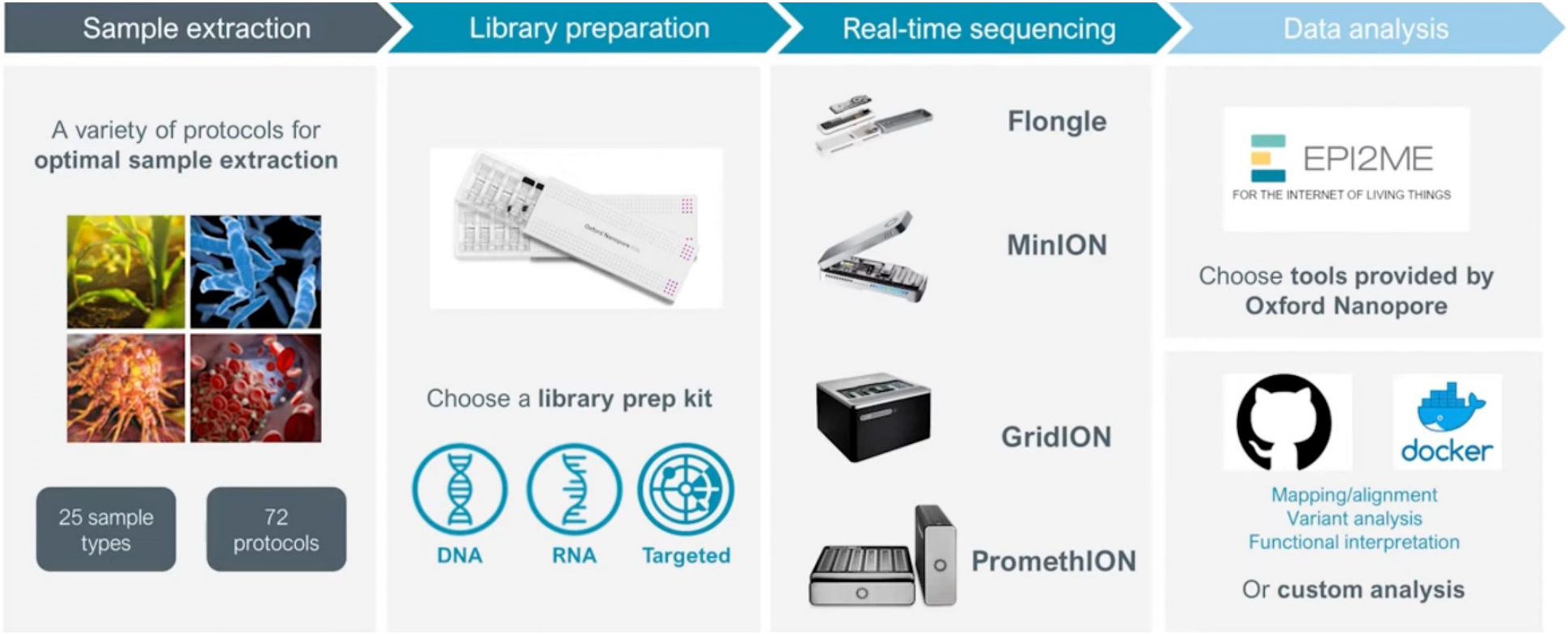
Reads from 20bp to 4Mb

2014 : First Nanopore sequencer (MinION)

1. DNA/RNA molecules pass through nanopores embedded in a membrane
2. DNA/RNA bases cause a characteristic current disruption
3. Changes in electrical current are translated into nucleotide sequence

scheme: https://nanoporetech.com/

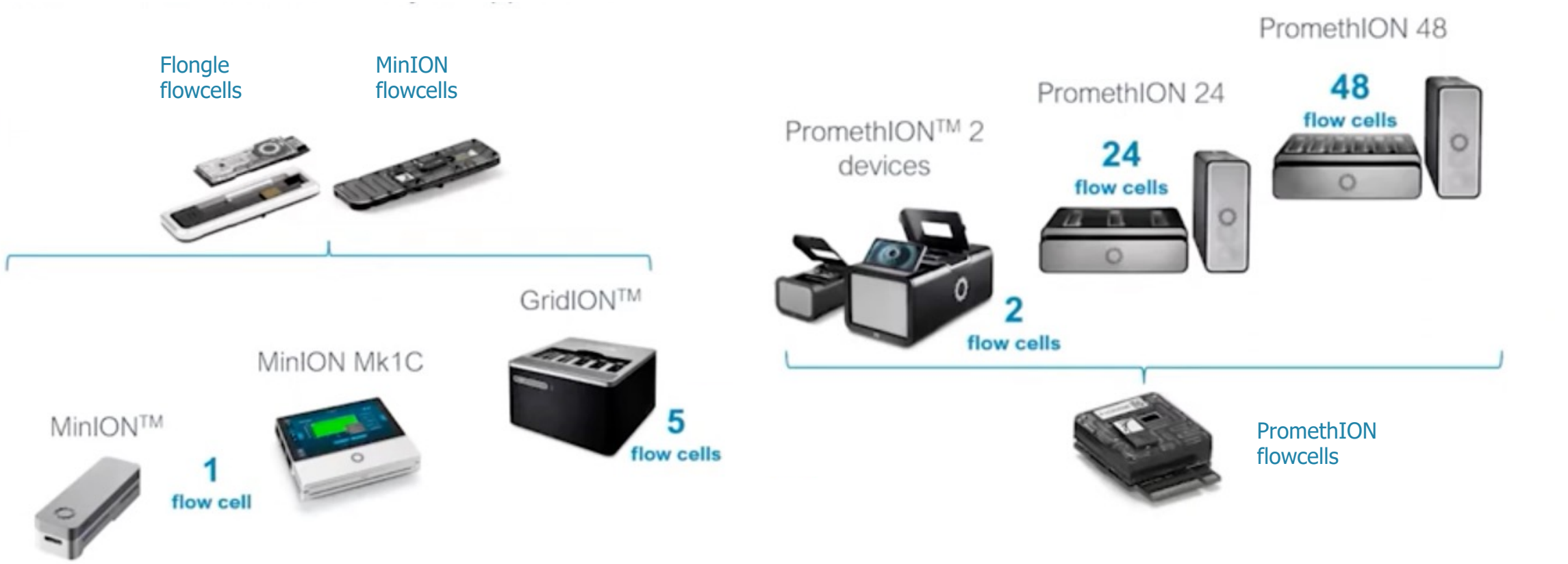
Global Nanopore workflow



- Sample quality is key!
- Flexbility in sample type
- Lots of different protocols!
- Sequence with the flowcell and device that's adapted to your project
- Different options according to level of bioinformatics expertise

scheme: https://nanoporetech.com/

Nanopore sequencing devices



scheme: https://nanoporetech.com/

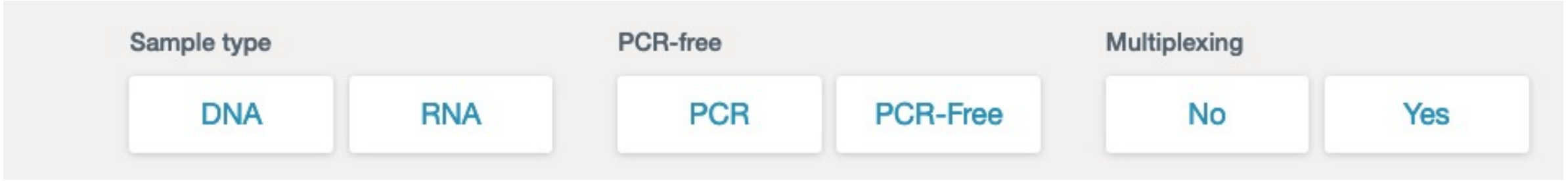
Nanopore flowcells

	Flongle	MinION	PromethION
			
Output max (ONT)	2,8 Gb	50 Gb	250 Gb
Output (average observed)	<1 Gb	10 Gb	100 Gb
Reads	100 k	1-10M	5-100 M
Quality	Q10 – Q30		

Nanopore library preparation methods

Library preparation is the process of converting raw DNA or RNA samples into a format compatible with sequencing. Typically involving optional fragmentation, adapter ligation, and optional amplification.

- Unmatched flexibility in sample type (only platform for direct-RNA sequencing)
- Choice of library prep will depend on the application, sample type, input quantity, fragment length…
- Examples:
  - Low input or rare targets – PCR based method
  - Methylation calling or fragment length beyond the limit of PCR – Ligation based method
  - Low-resource setting, fast results – Transposon based

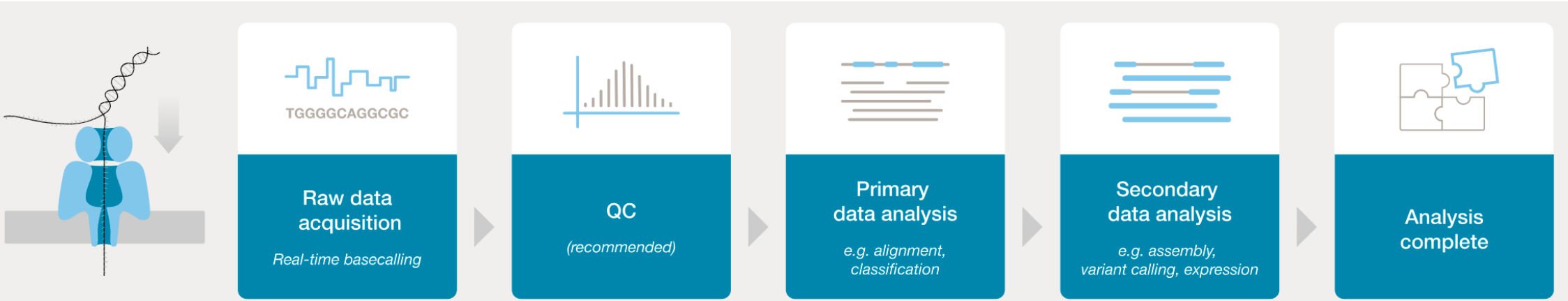


Comparison of the main Nanopore library preparations

Library Prep Type	Best For	Key Features	Pros	Cons
Ligation-based	High accuracy and long reads	Adapters ligated to double-stranded DNA	Preserves base modifications and long fragments	Requires more input DNA, longer prep time
Rapid (Transposase-based)	Quick and simple workflow	Transposase fragments and tags DNA in one step	Fast (1h), minimal hands-on steps	Shorter read lengths, less suitable for modified bases
PCR-based	Low-input or degraded DNA	DNA is amplified before adapter ligation	Works with small amounts of DNA	Removes native modifications, may introduce bias
Direct-RNA	RNA modifications, transcriptomics	Adapters added to RNA, sequenced directly without reverse transcription	Captures native RNA and modifications	Lower throughput, higher input requirements

Overview of data analysis workflow

1. Basecalling raw data (Dorado)
2. Data QC
3. Primary data analysis (removing adapters, mapping to reference)
4. Secondary data analysis (variant calling, taxonomic assignation, genome assembly…)



scheme: https://nanoporetech.com/

Basecalling the raw data : 2 options

1. Live and Local

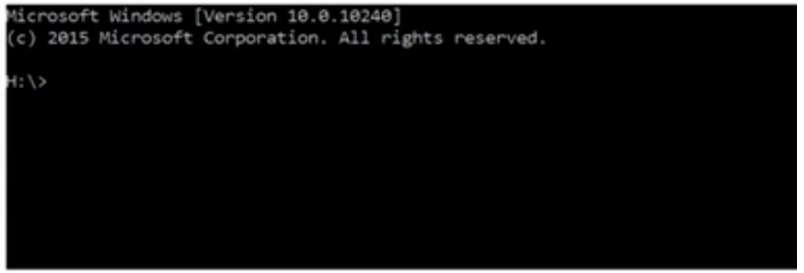
MinKNOW integrated analysis

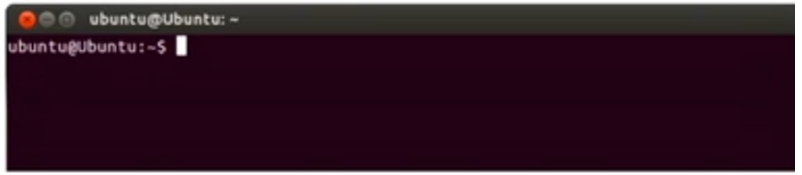
- "Catch-up" strategy for reads skipped during real time processing

2. Local and Offline

Stand alone basecaller

- Data analysis package
- Command line application
- Windows, Linux, OSX (in development)





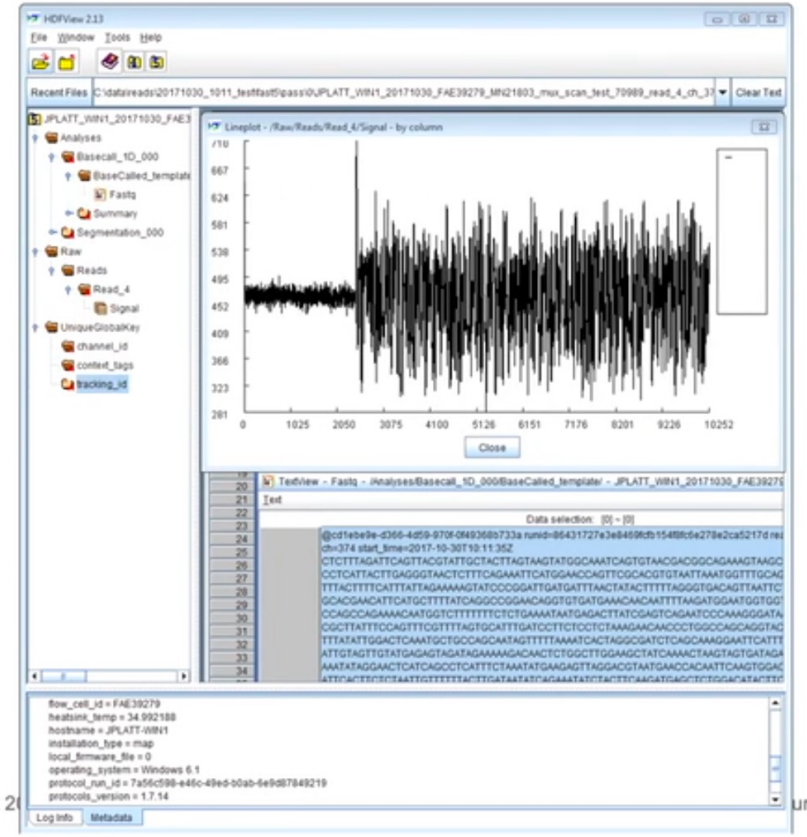
- Require powerful computing resources (GPU), especially if real-time analysis
- Basecalling algorithm has improved a lot in the past years and is constantly evolving!

scheme: https://nanoporetech.com/

Output files from sequencing

Device Control MinKNOW → **Data collection** MinKNOW → **Live basecalling** MinKNOW

pod5 Raw data (electric signal)



fastq Basecalled data with base quality

```
@de4d97b8-c1c5-4d48-83dd-cf35c8b1b262 runid=86431727e3e8469fcfb154f8fc6e278e2ca5217d read=4 ch=443
start_time=2017-10-30T10:11:28Z
TGATGCTTCGTTCAGTTGCGTATTGCTCTGCCAGATACGCCCGCGCAGATGATTACGGCTCACCTGCTGGCTGCCAGCTGCTCAGGTCCACCTGCGCCGAC
CGGTGCCGTGTTCCGGGTAGCCTGTTTCATCAGTAGTTTCAGTGTATGACGTGGGCGTTTATGCGAAACTGGTCTGTGCTGTCAGCGTCATCCTGCGCATC
CGGATGTAAACGGGCGCGGCGGCGGGTTACCCATCACCACCGGGGGCCCGGATGCGGGTGGTTTGCTATAATGGGTGATGATGTCTTTTCCGTCACCCAGCC
ACCGTTACGTTGTATCTAGACAGCAGGCGGACTACGACGGTTCTATCACCCTTTGGGGTGGTTTCCTGTACGAACGGAAGGTAAAGCGCAGACGGTCGATG
TTTGCAGATTGTAATGGGTGCGGGTGATCGGCGTGTCGTATTTCGCTTGTACTTGGCGCGCACCGTCTCAGGCCAGCGGAGTTCAAATCCCTCCAGCGGAT
CTATCCTCTCACCAGCCCGGAACACCTGTGGTACGGATATATTTAGTATTCCCCTCAGTATCCAGCACCAGCGTACTGTGTTCAGCCGCGCTTTTCCGTAT
CCACCGGACCTTCAATCCAGCCTTCGCTGGTGGCATCGATCCACACTCCAGCAGCCGTGGACTTGGAGTTGTCAGCCAGCTGCGCAGGGGTATGCCCTTAC
TGCTTCCTTTACCCATTTCTCCGCTCCACCATTAGCCGAAACCGCAGGCCAGTTCCACATAAAACATTTTGCATCTTGACCGAAATCACCTGGCAACCACC
GTCCCCTTCGTCTGCCGTGCCAGTCTCGCAGAGCCGCGTATTTGCGCATTCCCCATGCAGACAGGCCAGAACATTTGCCTGGGCAGCCATTATCAGTGGGG
AGAAATAGGTGTTCTGCTTACCGTTATCCGTTGTCTGTATACGGGGTTCGGCACCGGTGCCAGCATCCGCCACTACCGGGCGCCATACGGCACCGGAAAAC
AGGATGCCAGTCATACCACCGGCCCCAATGAAAACTGCCCCCATGCTGCAAGGGTGGCTCCGGCGAGTAAAGAATGATCCGGCGTGGCGGCAGCCCAGGAC
AATCTGGAATCGCCACCGGCTGACTGGCGACTCTGGGAACAATGAATTACAGCGCCATCAGGCCAGAGTCTCATGCATAACACCGCCGTTAACCGGACATT
TGGCGTCCGGCAATCCGTACCTGATACCAGCCAATCGCTCAGTTTCACAGCCGACGCCAGGGCTATGTGGCCAGTGCCCGGATGGCTTCAGCCCA
+
##$$$(')-,7A83<<D+5.)+,731++)),(''))-$./6<8+2*3,::;744*322(+(1*210265.7,,)5,'+::8789*/-,+/+17026289<::
4'2?+0<<>;>4/,)6*2$$$'.*56/.3:<%(4',1;@3(C@:3**+.17-,%-05162($(&'%+))).3852-),*,/-19-32=8?<=AC?452647<
4=?>:*2(3,2-6AAAA<</-50*4269%/(-)-..<@,14/-?36:'*??,+,&.;2>>230((,/30+6.3/-,-.1387<>8ED>0:169-298,+.3
(,('-(#&*,0--++*'&--.'(((*+/*4188/3;898=.@0,)8;5>C79D2+-.(0/?4A(*'$$%&,+%&57)6*9,.%')+/2E8A>@/CB96887
8C=7<120-*$(',-4:.((,&*&'(.02.83+3*++&/(5C0+&'&%#%%&'%%*.,)'**296--,,+7:8?*+&3-*+*2*&&(*:=-+:.%.43:,0
10798*((,0(('*$+))/:352*/./0.+,+-4,;+0'.+6C6(//4<D5CD.*53'6&,11669%6+02334;8,*-%5>)))*(+/,+**(&(118::
9839@4C25++-%3,1?(%-/1-1.-./.03,.4),0*-;>((1.1+2'*&,.&,.(7-A/677+)%)/5*//0&.-'%',,+%'&45>6=45';D?)2+-
))(/21))8((-)($+88:32--22;C-39(3-+)+.(+,);E256754-&17C0-*.)/+20+---A2,/2,('&(*15/&1052<9<90-)*(&(+*55
9=2?:+09:;*C;66:/3/4*$/<<78$)&'3'(,(***%,-(),+-(,--)+*&&"%&+65/409(%-,50/1?105A:;2*(/+:433:024'87/+9;
--(1A-,3-</5?61*))''-0+6501=29:D223.*-,-':(346-7.,&&#&%//@<@,*,,'5$(,/+)+&'..+**5&*.,($'*5-(.;/6/3439
&87>86.*&)1*&,-=*&,'//+-.*0-:857/&*(,4BFD><:65,111/-8314<938,-/6-)&*8B(')116:CCB8=?,>0@2'62+-94*$4620
<40;;+9$%'327:'5$,&)#($)&-*,..=884,8899961.*%'-)'(%.5>B?<><095;*+**,+40+%&$$&&*/.),316=(:(0$*(0.8<'*-
+)/-+')'),5<A==8'<+0>*((,,:*3%&%.&,16;<?34>=),$%&#%$#=5048+4:=50+,((+-/5**+,.)(+668)'(%5#$'))&$
@4ea9b632-cbca-455d-9061-88117b879114 runid=86431727e3e8469fcfb154f8fc6e278e2ca5217d read=4 ch=107
start_time=2017-10-30T10:11:30Z
TGTTACTTTAGTTCAGACCATTGCACCATCAGATTATGTTTATTAATTCTTTTTTTTTTTGGAATTTTTTTTTGAGATTTTTTTTTGCGCTAACAGCCTCC
ACACGTTTTGCCCGTACTCGGTCACGAACAAATCCTGATTACTAAACACATTAGCCTCGTTTGTCTATCAATCAGCCTTATTCTAATTAAATAGAACATTC
CCCTTATTGAGGGTAAAGACGCCCAGAATTGAAAACGCGGCCTGTTGTACTTATTCTCGCAGCCAGGAACAAAACATCAGGCAATCCTTGCGTTTGCAATG
```


→ Raw data can be re-basecalled later with new algorithm… Save both the pod5 and fastq files

scheme: https://nanoporetech.com/

Different analysis approaches

- Different options according to level of bioinformatics expertise
- Lots of tools under development, time needed should not be underestimated

	EPI2ME	Protocol & analysis tutorials	Community developed tools	Custom analysis pipelines
Bioinformatic capability needed	● ○ ○ ○	● ● ○ ○	● ● ● ○	● ● ● ●
How	Use the cloud-based EPI2ME platform for real-time analysis workflows.	Get analysis recommendations and clear tutorials on the use of open-source tools.	Run open-source tools written and developed by the Nanopore Community.	All the data, raw or basecalled, can be used in custom analysis pipelines written by the user for specific applications.
Where	EPI2ME Workflows ⓘ	Community knowledge ⓘ	Community	User defined

scheme: https://nanoporetech.com/

4. Applications of Nanopore sequencing

Human genomics and cancer research

- **Genetic disease research**

Identification of structural variants (SVs), SNVs, large indels
Improved resolution for complex genomic regions (e.g., repeat expansions, telomeres)
Phasing of alleles

- **Cancer genomics**

Detection of mutations, fusion genes, and large-scale rearrangements
Full-length sequencing of oncogenes and tumor suppressors (e.g., TP53, BRCA1)

- **Cancer transcriptomics**

Sequencing of full-length RNA transcripts
Identification of novel transcripts and RNA editing events

- **Epigenetics and methylation**

Direct detection of DNA methylation and other epigenetic modifications
Sequence native DNA, so no need for bisulfite treatment or chemical conversion

Microbiology and infectious diseases

- **Pathogen identification & genomic surveillance**

Rapid, on-site identification of pathogens
Whole genome sequencing of pathogens for outbreak tracing
Real-time monitoring of emerging infectious diseases (COVID-19)

- **Antimicrobial Resistance (AMR) Detection**

Direct identification of resistance genes in clinical and environmental samples

- **Field-based sequencing**

Portable devices (MinION) for direct on-site sequencing
Immediate access to results
Mobile labs

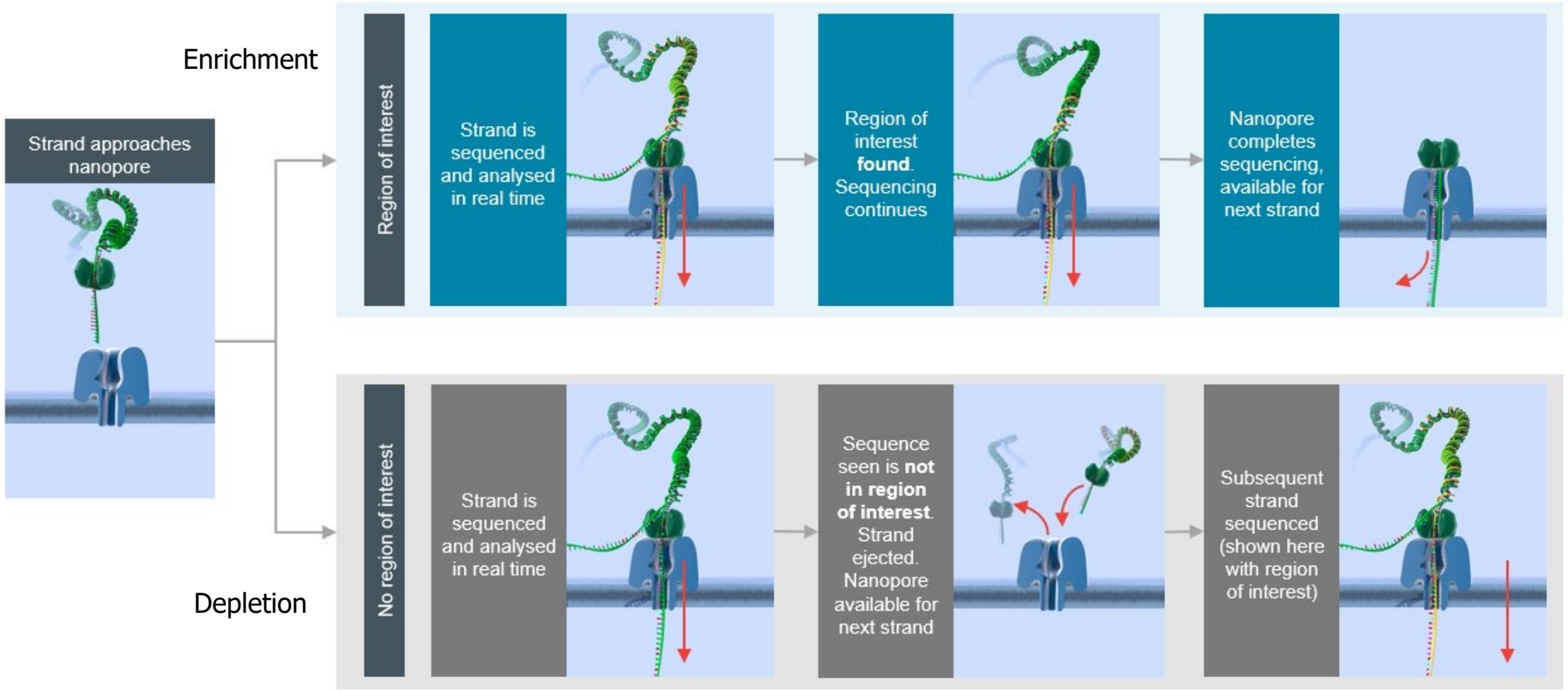
- **Environmental research**

Analyse envirnmental DNA (eDNA) and microbiome samples
Assess biodiversity, animal conservation, monitor ecosystems

Sequencing samples in antarctic

Johnsson SS, *et al.*

Adaptive sampling: targeted sequencing in real-time



- Increases coverage of specific regions
- Detect various pathogens from various samples
- Host depletion (microbiome samples)

scheme: https://nanoporetech.com/

Adaptive sampling: what do you need?

- No need for additional library prep (ligation kit recommended) → direclty upload regions of interest when setting up the run in MinKNOW software (.bed and .fasta files)
- Need a known reference sequence
- Powerful computing ressources (GPUs) needed
- Target must be 1-10% of the genome/sample
- Sufficient fragment length (5kb-10kb is optimal)
- Enrichment of ~5-10 fold when using adaptive sampling
- Does not work for direct-RNA sequencing (yet)

5. Challenges and future directions

Advantages and challenges

Advantages

- Read length
- Portable
- Real-time sequencing
- Sequence native DNA/RNA (only technology for direct-RNA seq !)
- Direct epigenetic modifications detection
- No amplification needed
- Easy to handle

Challenges

- Accuracy
- Need high quality DNA and high quantity
- Variability in flowcells, pore numbers, stability
- No standardized protocol, need to adapt
- Data analysis can be tricky
- Need to constantly keep updated as the technology evolves fast

Upcoming developments

- Enhanced accuracy and basecalling (machine learning, AI)
- Multiplexing for direct-RNA
- Expanded list of modifications that can be detected
- Increased throughput
- Direct protein sequencing

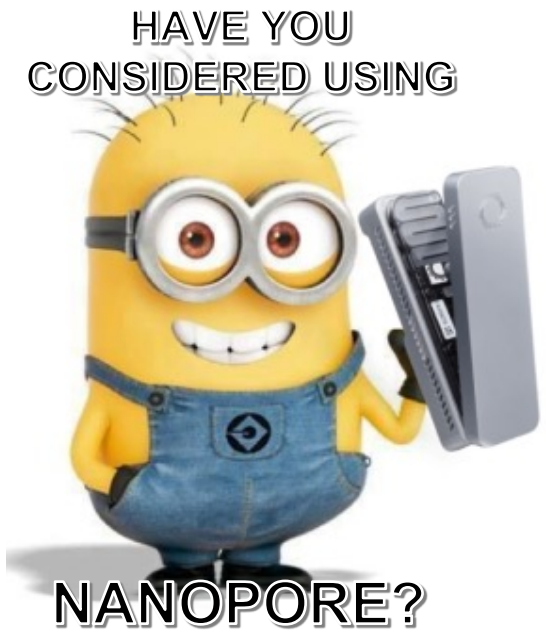
Direct sequencing of protein amino acid sequences without chemical modification or labeling

Challenges such as protein folding and pore blockage

- Annual Nanopore conference is in May… Stay tuned !

LONDON CALLING 2025

Summary: a short guide to get started with Nanopore



- Is Nanopore the right technology to answer my question?

- Which extraction protocol, library prep should I use?

Choice depends on the application, sample type, input quantity, fragment length…

- What throughput do I need? (Flongle, MinION, PromethION)

- Analysis: Are there tools or end-to-end pipelines already developed that I could use?

If not, do I have the minimum bioinformatic skills required? (command line, linux terminal, cluster use)

Nanopore is constantly evolving and subject to change. Keep up to date!

Do not hesitate to use the Nanopore Community

https://community.nanoporetech.com/

To go further

Nanopore community
https://nanoporetech.com/community
You can explore to find everything related to protocols, tech updates, general information…
You can also ask a question to the nanopore community users using the forum.

How Nanopore sequencing works
https://www.youtube.com/watch?v=RcP85JHLmnI&list=PLxpxXZj-IgXyR0it4QjoBETC5JeEanW-o&index=1

Nanopore masterclasses 2023
https://www.youtube.com/playlist?list=PLxpxXZj-IgXx-avVuZI3F-6nnJlkm-kV6

Technology videos
https://www.youtube.com/playlist?list=PLxpxXZj-IgXyR0it4QjoBETC5JeEanW-o

Adaptive sampling
https://www.youtube.com/watch?v=fwHreHs9FHg

References

https://biomics.pasteur.fr/

https://nanoporetech.com/

https://community.nanoporetech.com/

- Wang Y, *et al*. Nanopore sequencing technology, bioinformatics and applications. *Nat Biotechnol.* 2021
- Zheng P, *et al*. Nanopore sequencing technology and its applications. *MedComm* (2020). 2023
- Johnson SS, *et al*. Real-Time DNA Sequencing in the Antarctic Dry Valleys Using the Oxford Nanopore Sequencer. *J Biomol Tech.* 2017
- Wrenn DC, Drown DM. Nanopore adaptive sampling enriches for antimicrobial resistance genes in microbial communities. *GigaByte*. 2023
- Lu C, Bonini A, Viel JH, Maglia G. Toward single-molecule protein sequencing using nanopores. *Nat Biotechnol*. 2025

Wet-lab

Marc Monot
Yakov Vitrenko
Laurence Ma
Elodie Turc
Azimdine Habib
Georges Michel Haustant

Dry-lab

Etienne Kornobis
Rania Ouazahrou
Laure Lemée

Support-lab

Laurence Motreff
Imène Najjar
Natacha Haikine
Florence Jagorel

Feel free to reach out!

chloe.baum@pasteur.fr

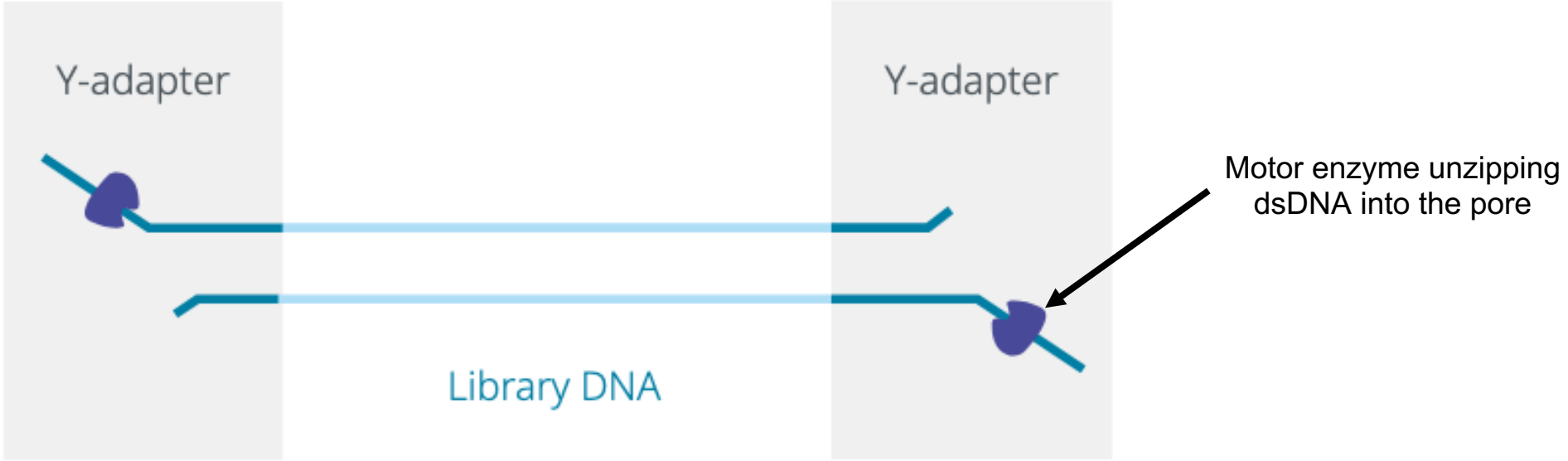
Acknowledgements

The creation of this training material was commissioned by ECDC to Institut Pasteur with the direct involvement of <name of speaker(s)>

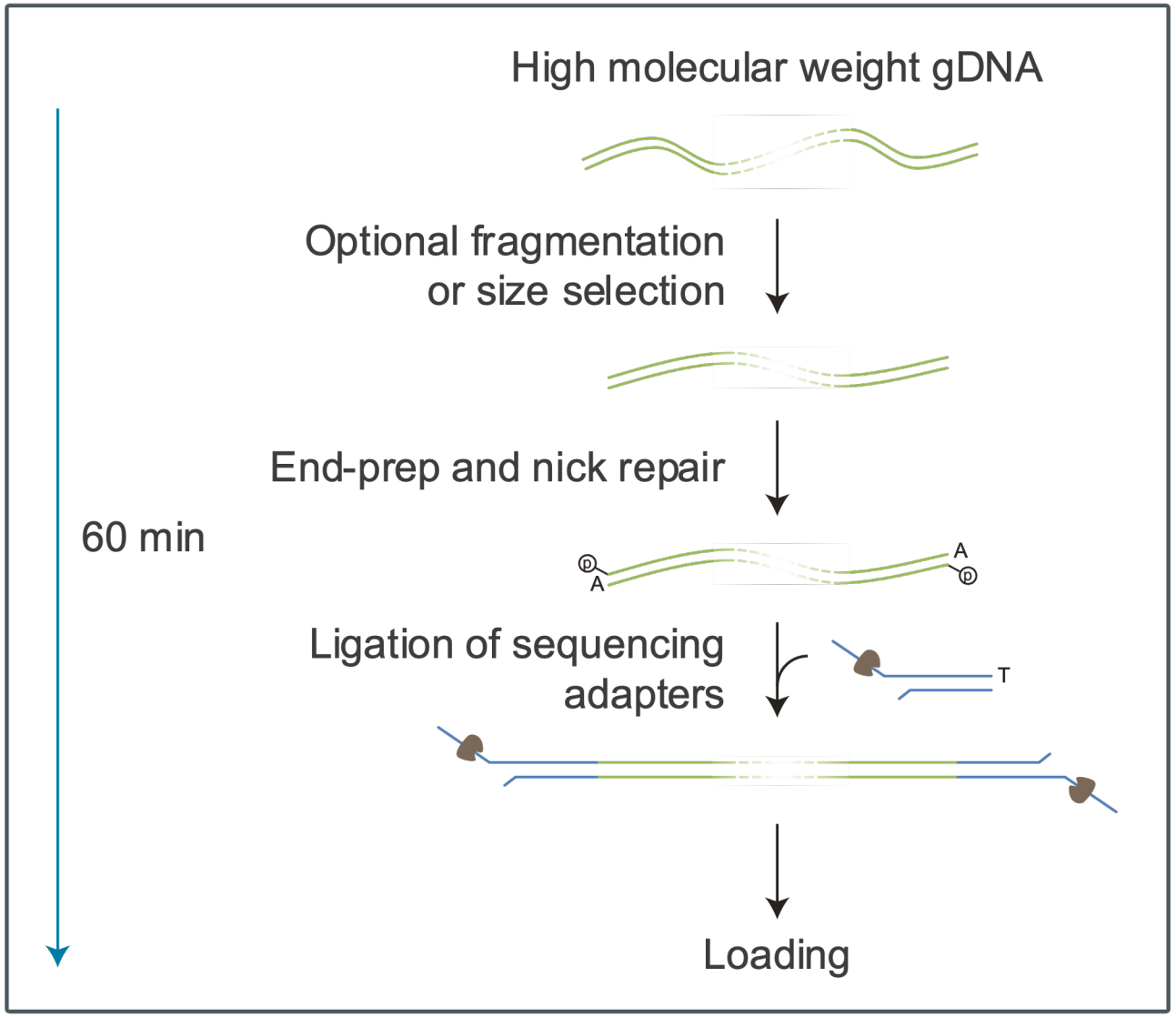
6. Q&A session

Appendix

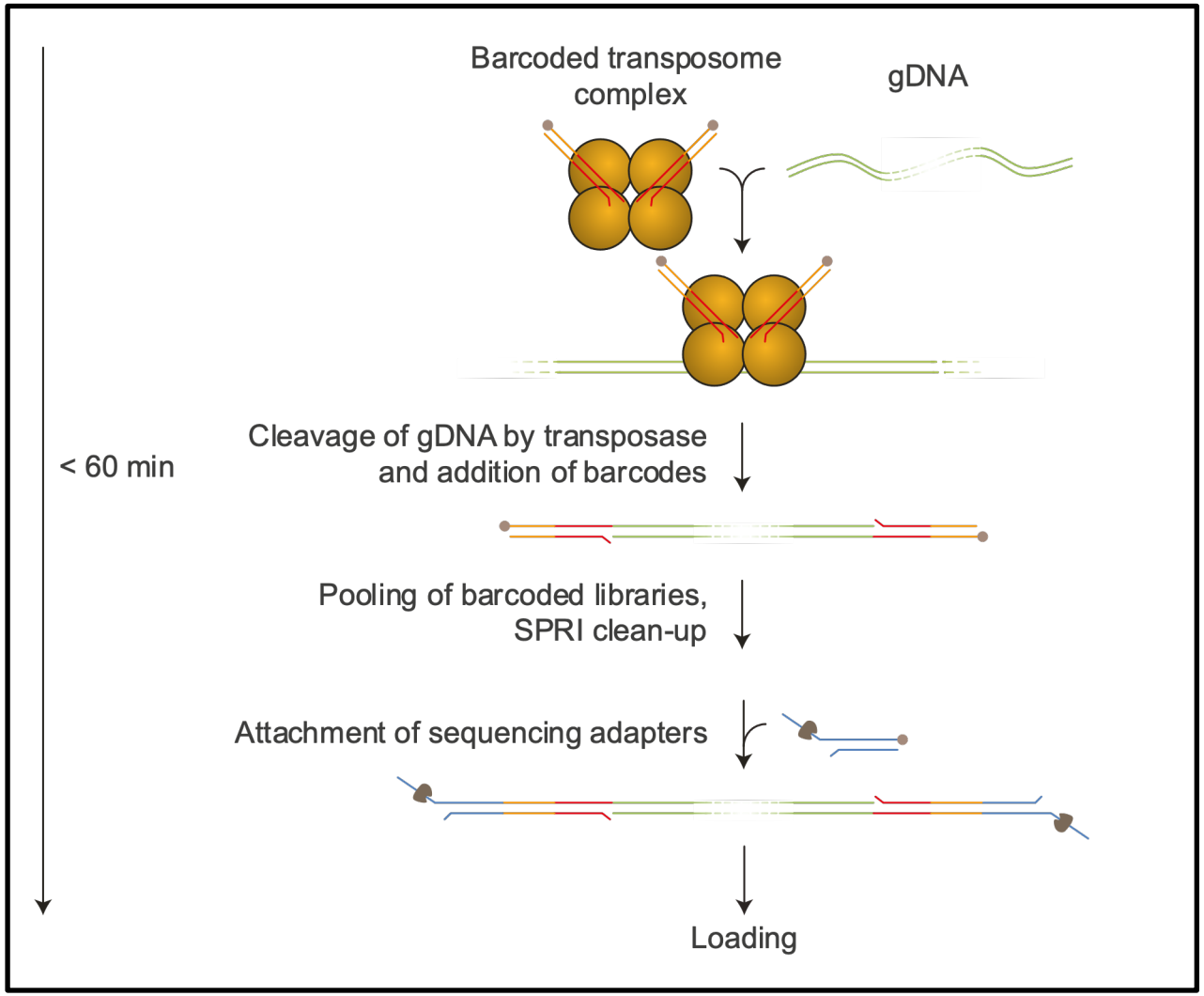
Nanopore library structure



Ligation-based library prep



Transposon-based library prep (rapid kit)



Direct-RNA library prep