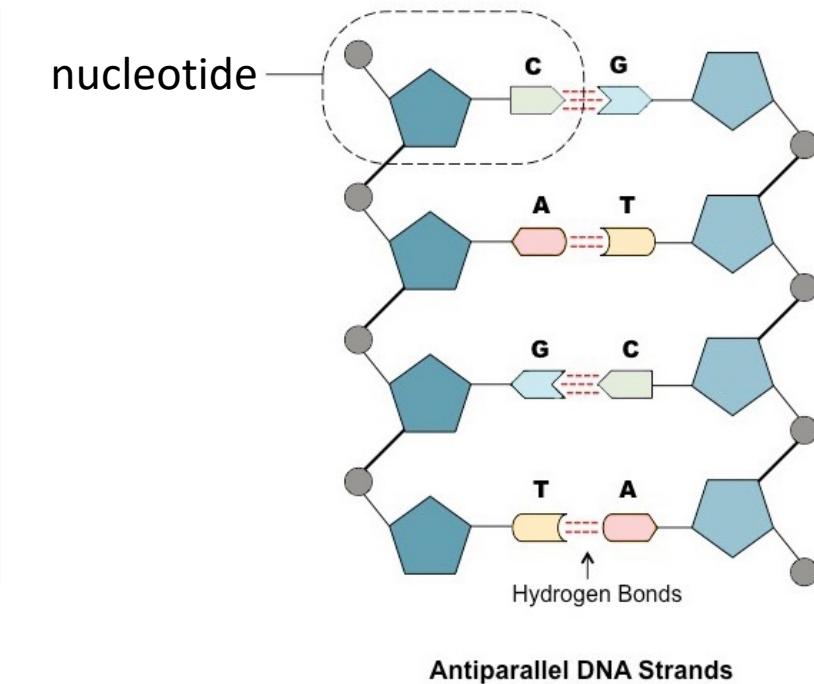
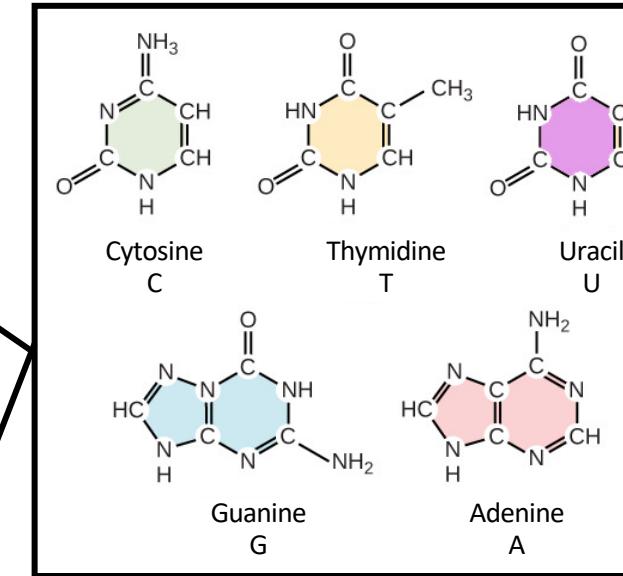
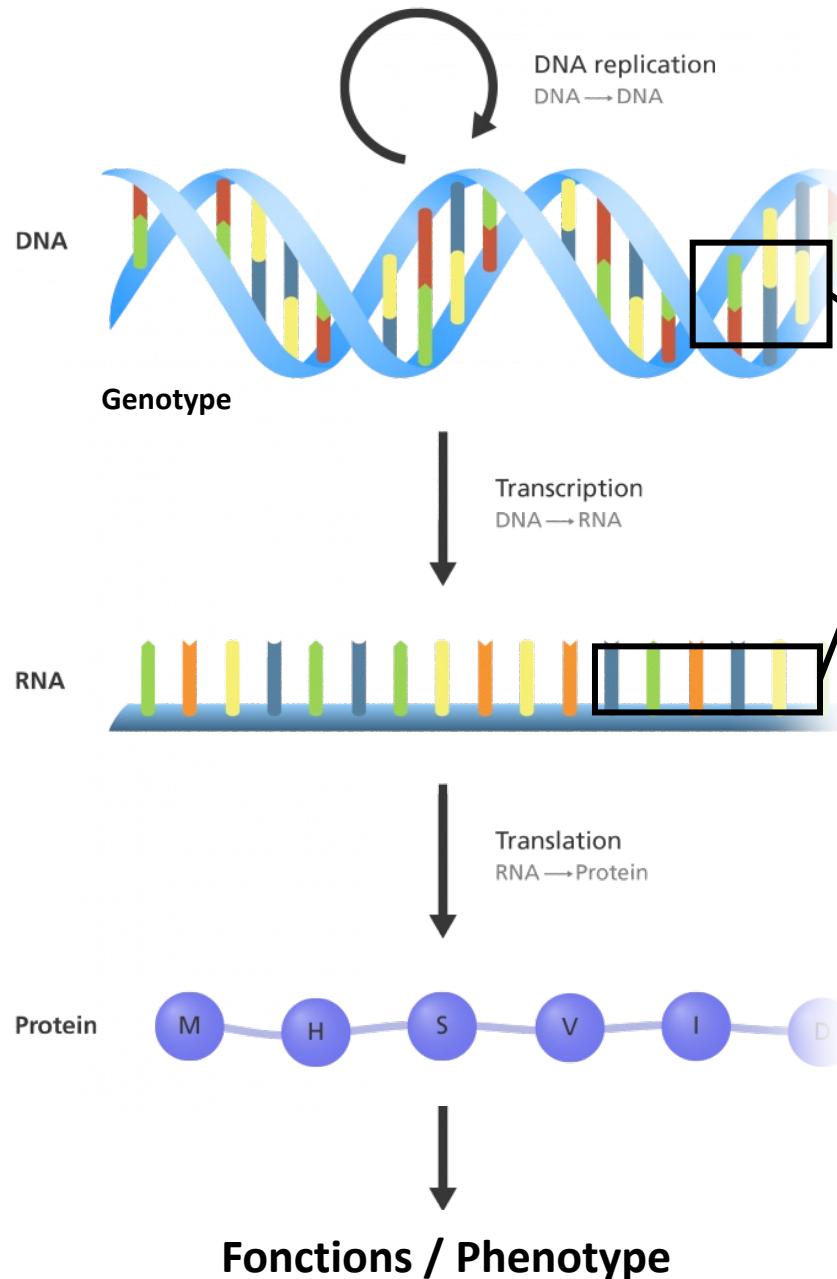


Nanopore sequencing and detection of DNA methylation

A
C
G
C
A
G
C

Mathias Boulanger, PhD
Predoc course – Nov 2022

Gene regulation and DNA sequencing



- ~22k coding genes can code for more than million of proteins
- Multilayer regulations give rise to specific phenotypes
- Scientists try to understand life and disease by retro-engineering:
 - Protein and PTM (proteomic)
 - RNA / Transcript (transcriptomic)
 - DNA / Gene (Genomic / Epigenetic)

Sequencing = determination of nucleotide sequence of a DNA molecule

DNA sequencing technologies

Illumina 



- Gold standard
(2nd generation sequencing)
- Optical readout
- Short reads
- Very high depth
(~1.3 billion reads NextSeq2k)
- PCR amplified fragments

Nanopore

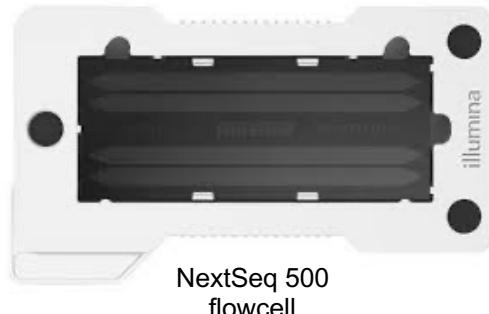


- 3rd generation sequencing
- Electric readout
- Ultra long reads
- Low depth
(~13-15Gb GridION)
- Native DNA/RNA sequencing or PCR amplified fragments

PacBio

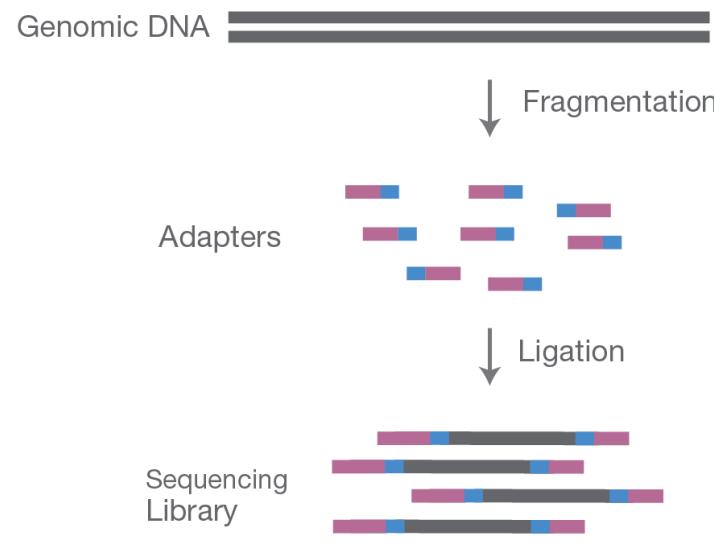


- 3rd generation sequencing
- Optical readout
- Long reads (up to 15kb)
- High depth
(~35Gb Sequel II)
- Native DNA/RNA sequencing or PCR amplified fragments
- HiFi reads

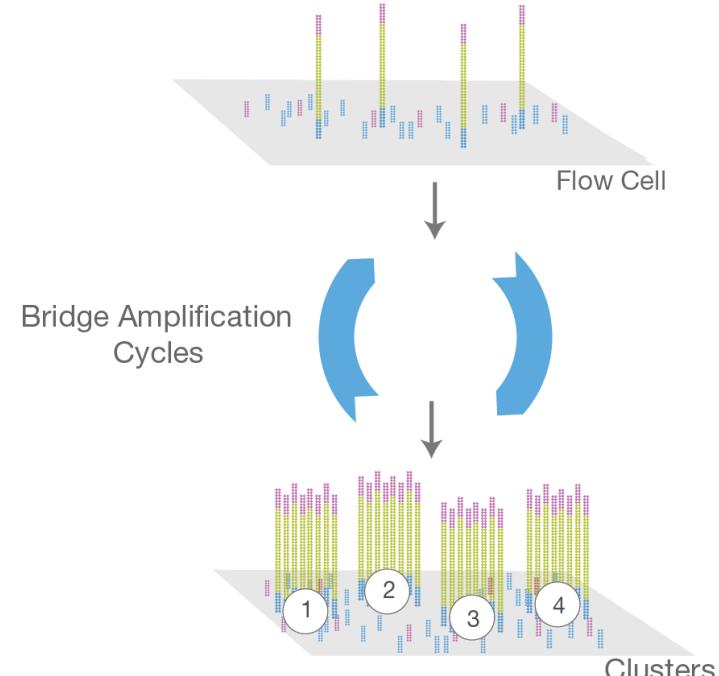


- Sequencing by synthesis
- PCR based amplification
- Image read out by incorporation of fluorescent nucleotides

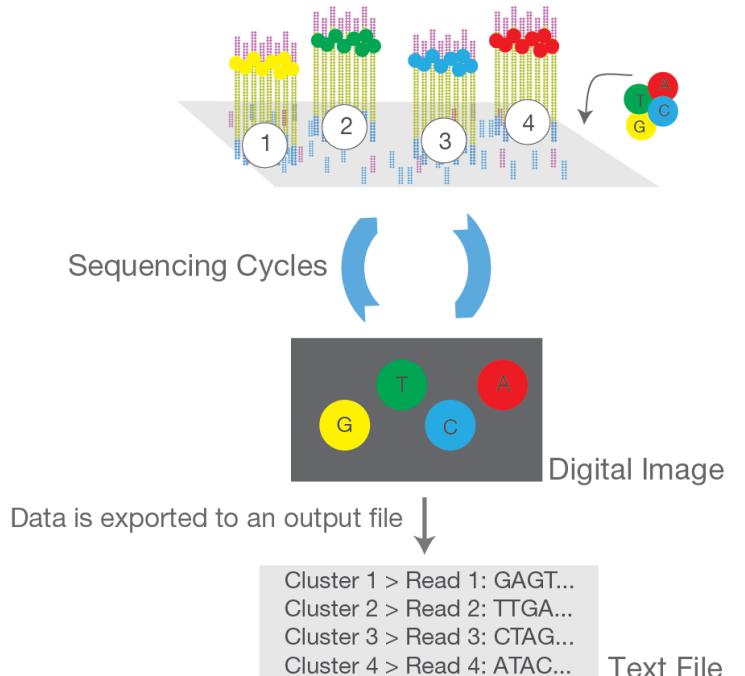
1 – Library preparation



2 – Cluster amplification



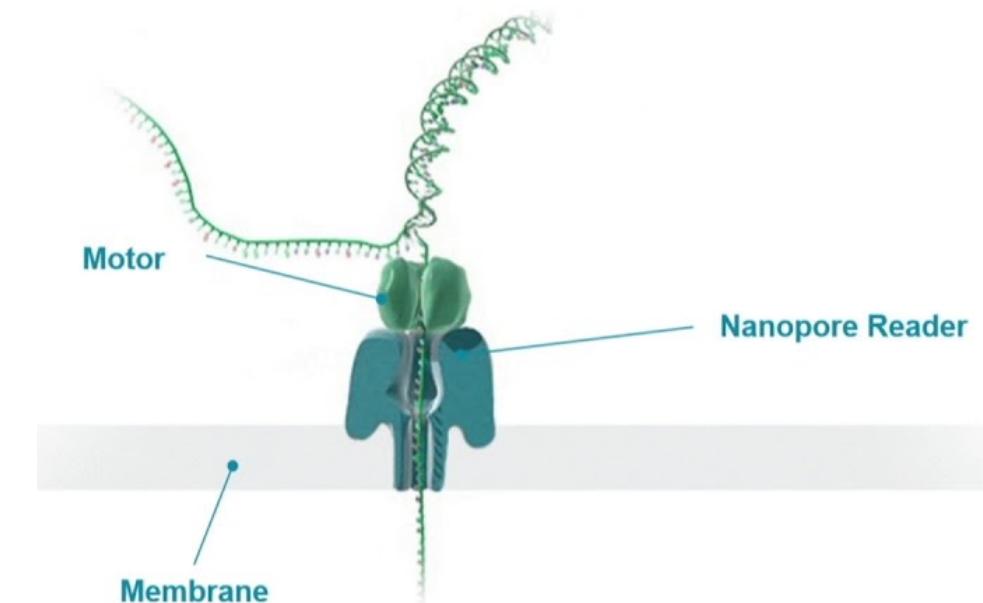
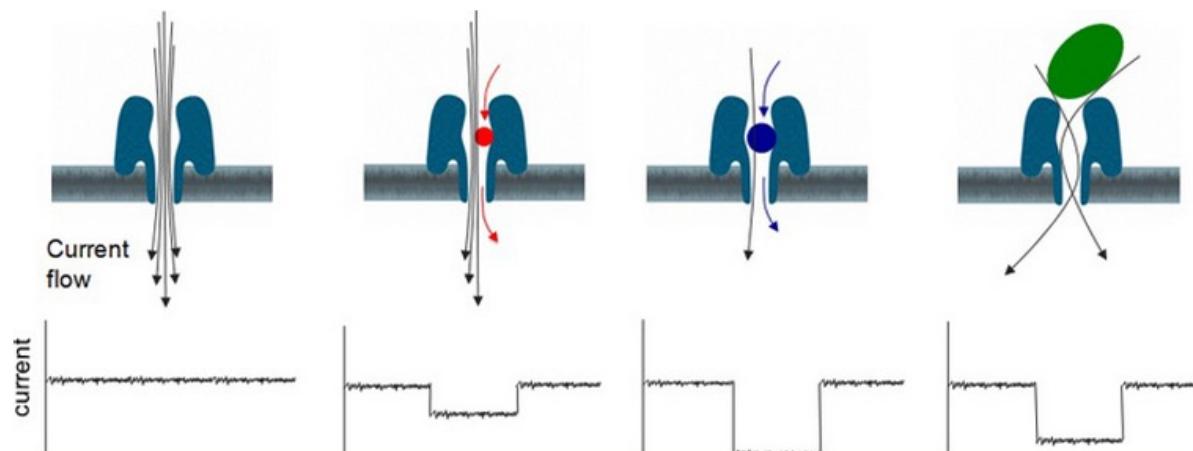
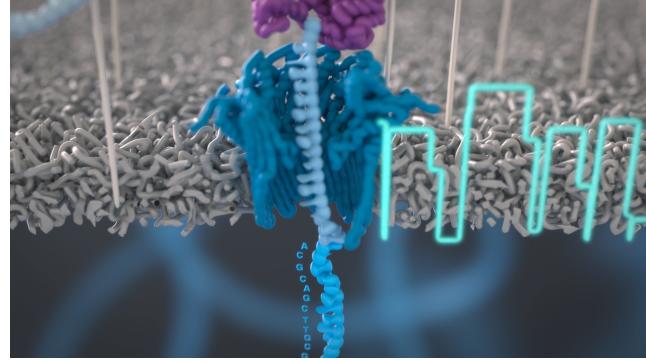
3 – Sequencing





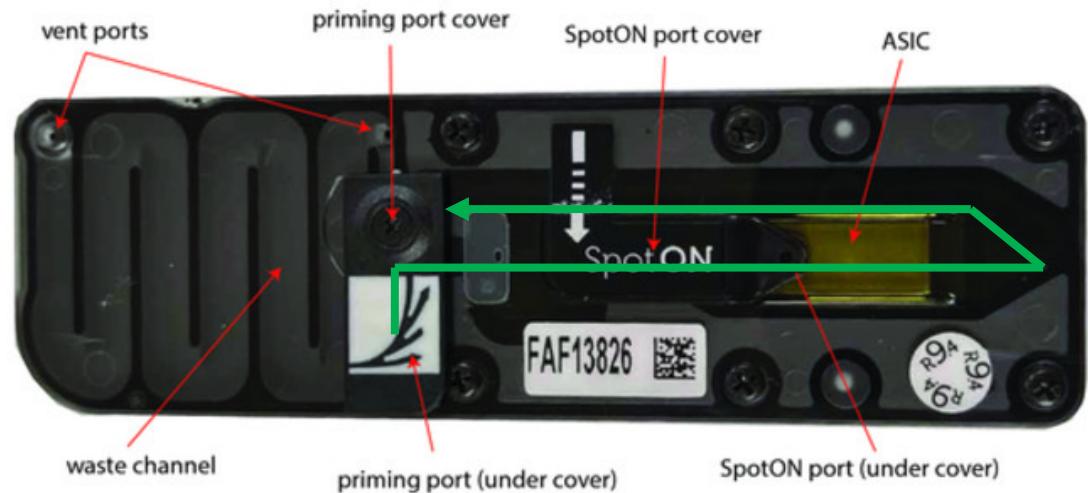
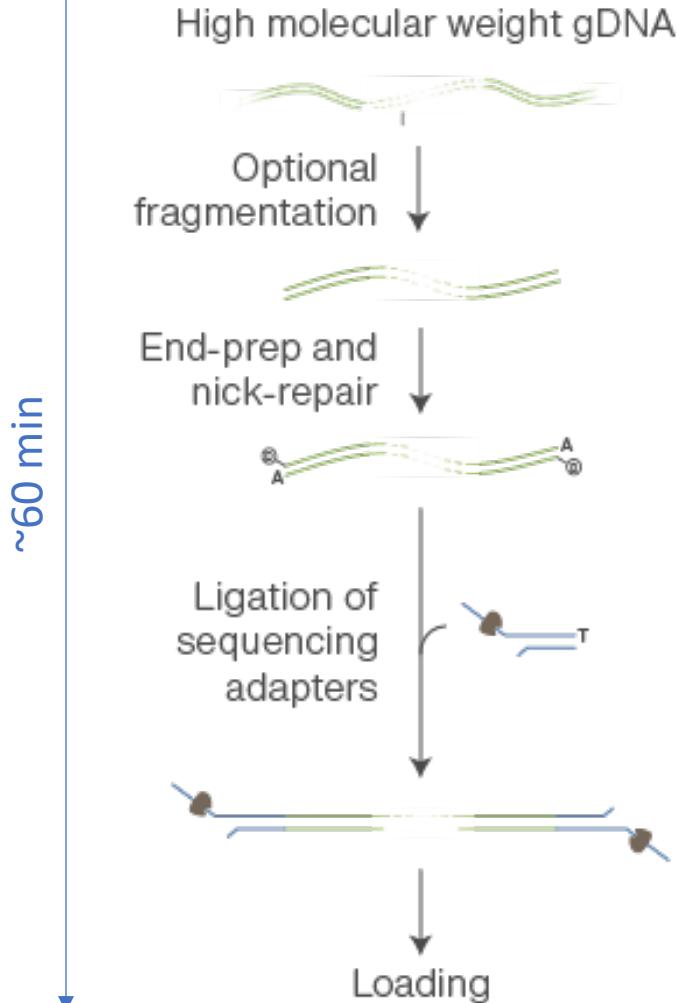
Oxford Nanopore Technology (ONT)

- Sequencing based on electric readout
- Pores are transmembrane protein channels (~1nm diameter) modified from biological proteins used for the cell macromolecule and ions exchanges (eg. A-Hemolysin, MspA...)
- Array: membrane (e.g. lipid bilayer) or synthetic material (e.g. silicon nitride or graphene)
- The pore is blocked due to passage of a DNA molecule. Therefore, the current flow is also altered depending of the nucleotides inside the pore.
- Bases are different at the atomic scale and then produce different ionic current torrents when translocated through the pore.

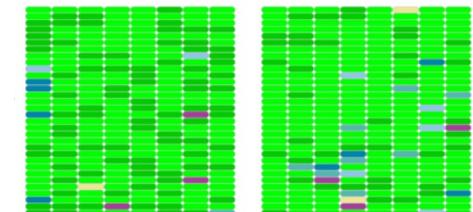
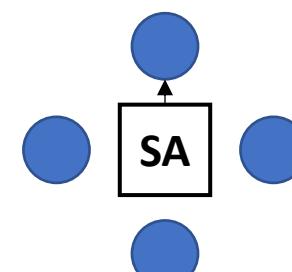


ONT library prep and flowcell

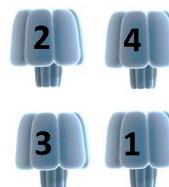
Library preparation



Array of 2048 wells with nanopore embedded in polymer membrane. Underneath is an Application Specific Integrated Circuit (ASIC) which contains 512 signal amplifiers (SA).

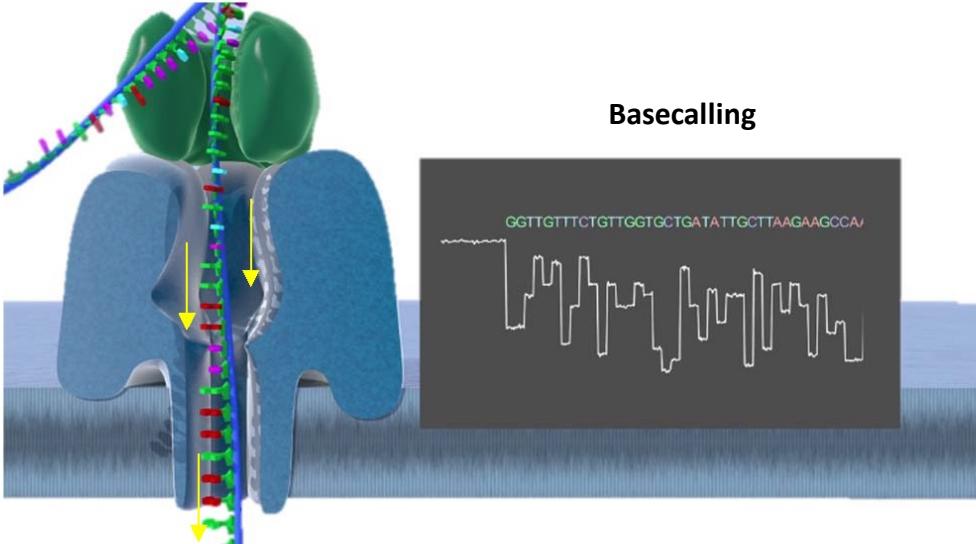


Every 1,5hrs (normal settings), the pore in a well can be saturated (exhausted) and the SA can switch to a new well



4 pores per channel

ONT sequencing pros and cons

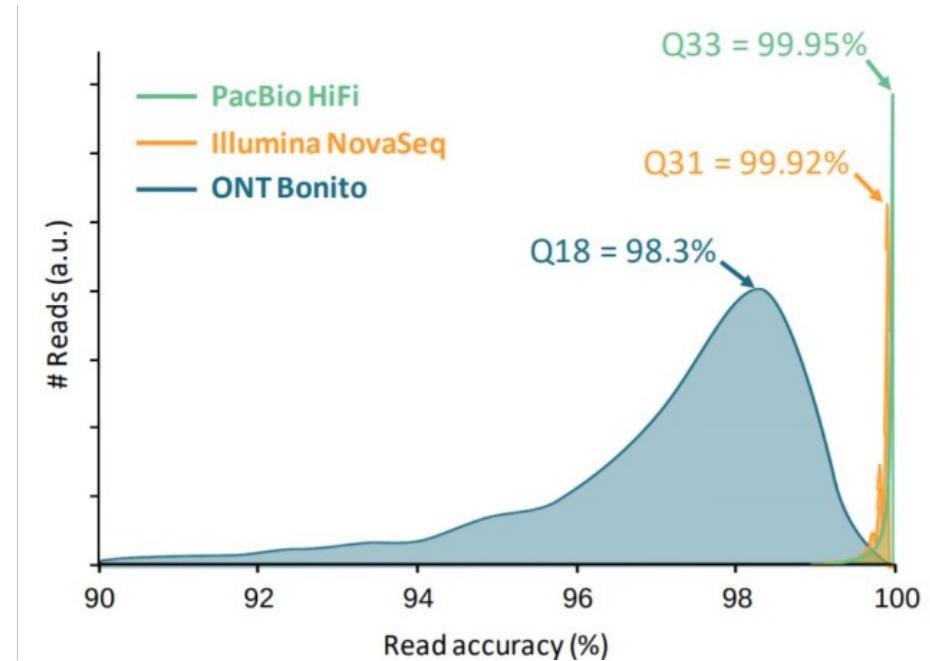


PROS

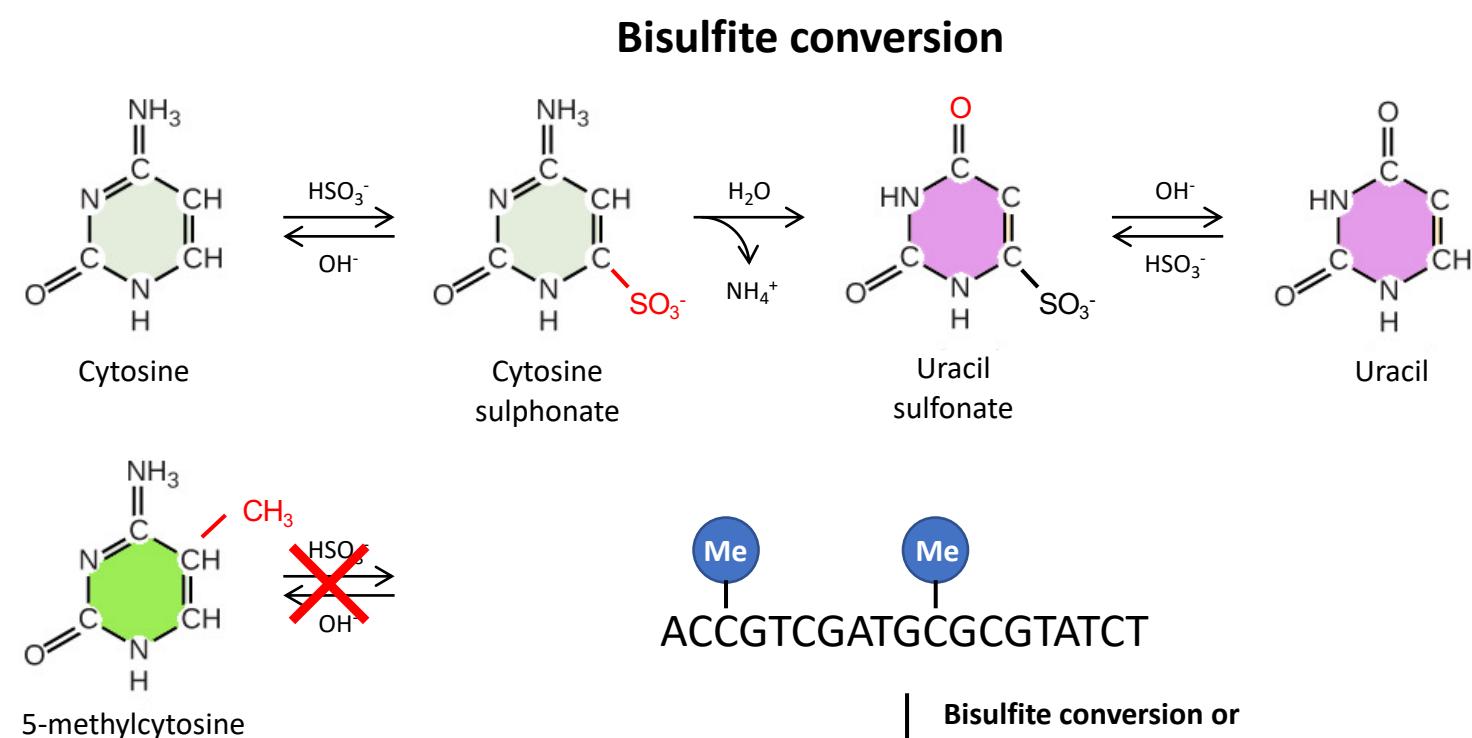
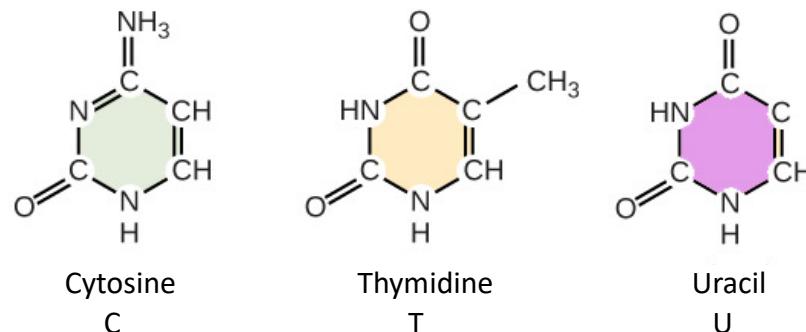
- No strand amplification required.
- No bias due to sequencing amplification.
- Low cost
- Read size: limited only by preparation (ultra long)
- Large-scale structural variation can be detected at lower depth of coverage.
- Enable long-range haplotyping.
- No need for expensive and time-consuming mate pair library construction.

CONS

- Low throughput (compared to Illumina)
GridION ~13-15Gb
- IT Resources (storage and server) reaching easily To of data
- Lower read accuracy than other sequencing technology



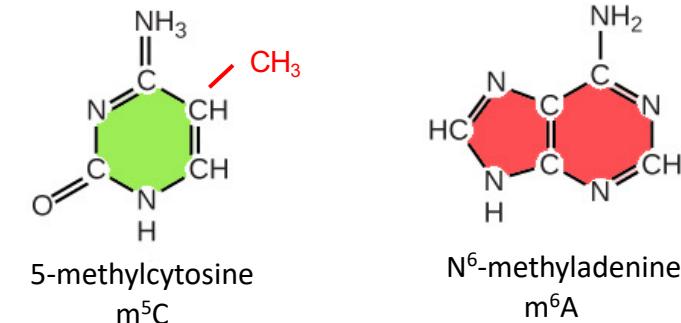
Detection of DNA methylation



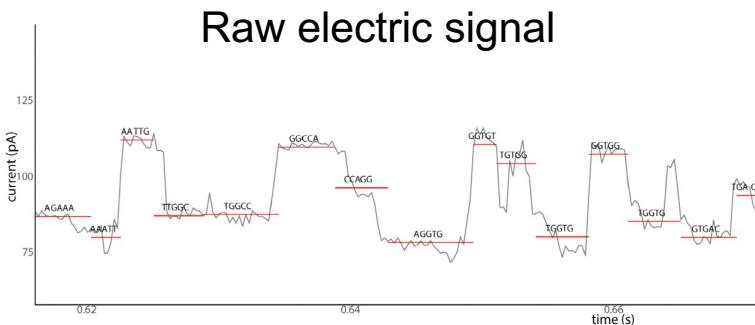
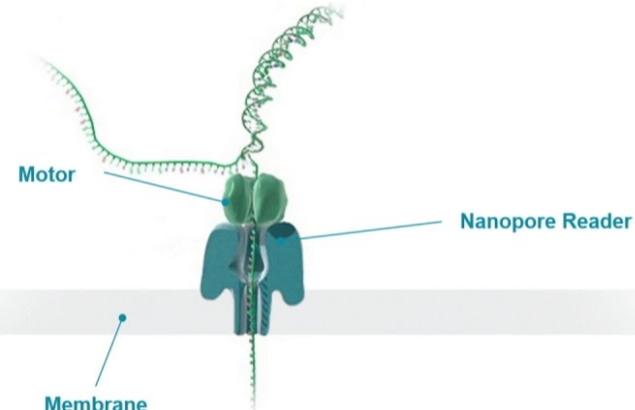
ACCGTCGATGCGCGTATCT
↓
Bisulfite conversion or Enzymatic conversion

AUCGTUGATGCGUGTATUT

↓
Illumina sequencing



Methylation calling - Nanopolish



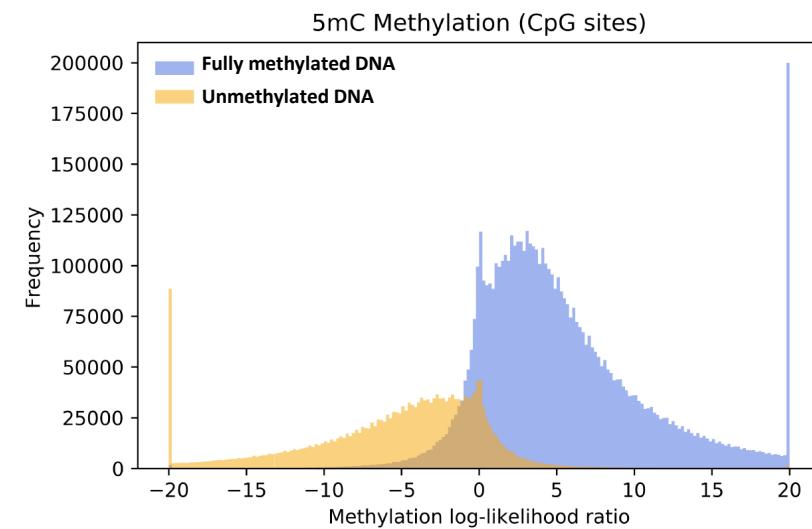
Basecalled Sequence

AGAATTGTTGCCGGCCAGGTGTTGGTGGTGACC

Hidden-Markov-Model

Kmer	Current	Current_sd
TACGAA	102.28	1.54414
TAMGAA	102.30	1.93632
TACGAC	97.0313	2.24722
TAMGAC	97.9398	1.89519
TACGAG	100.581	2.47220
....

Likelihood methylation Ratio

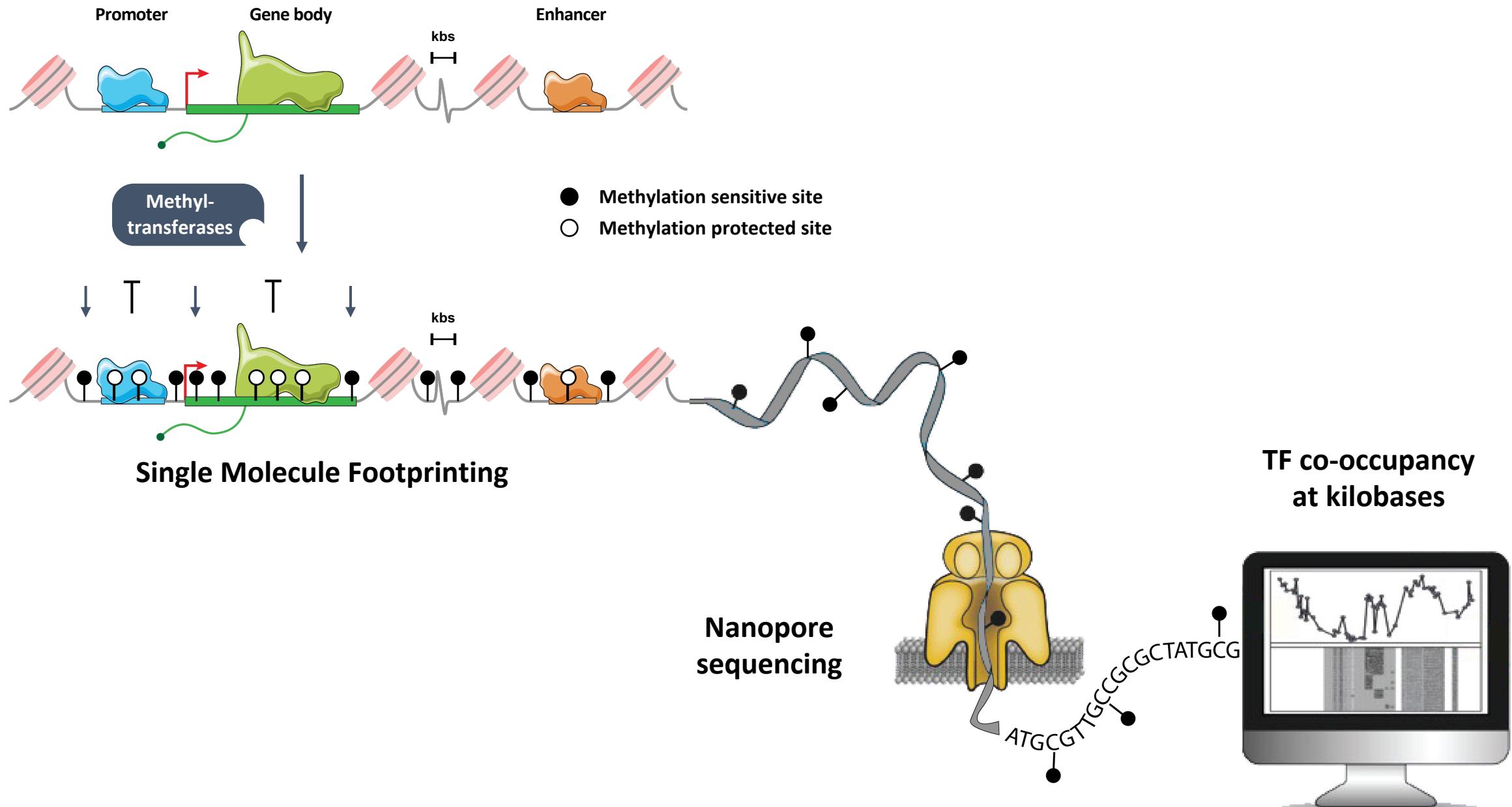


LLR < 0 = Unmethylated C

LLR > 0 = methylated C

-2 < LLR < 2 = underterminated methylation

Single Molecule Footprinting - ONT



Single Molecule Footprinting - ONT

