CNRS UPMC
Station Biologique Roscoff
aviesan
alliance nationale pour les sciences de la vie et de la santé

LONG READS

"Chasing perfection"

Claude THERMES

PLATEFORME DE SÉQUENÇAGE I2BC

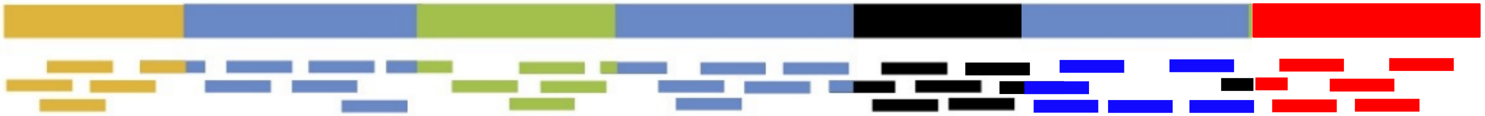
INSTITUT DE BIOLOGIE INTÉGRATIVE DE LA CELLULE

GIF-SUR-YVETTE

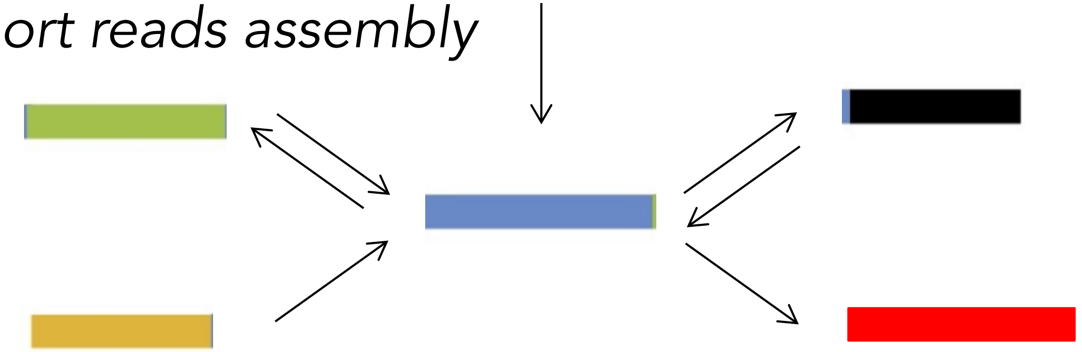
10ème ÉCOLE DE BIOINFORMATIQUE EBAII - 23/11/2021

LONG-READS VERSUS SHORT-READS

Assembly of DNA fragments with repeated sequences

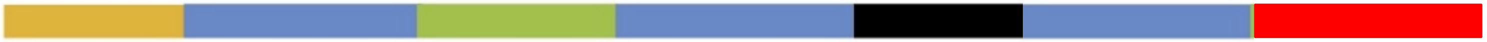


NGS short reads assembly



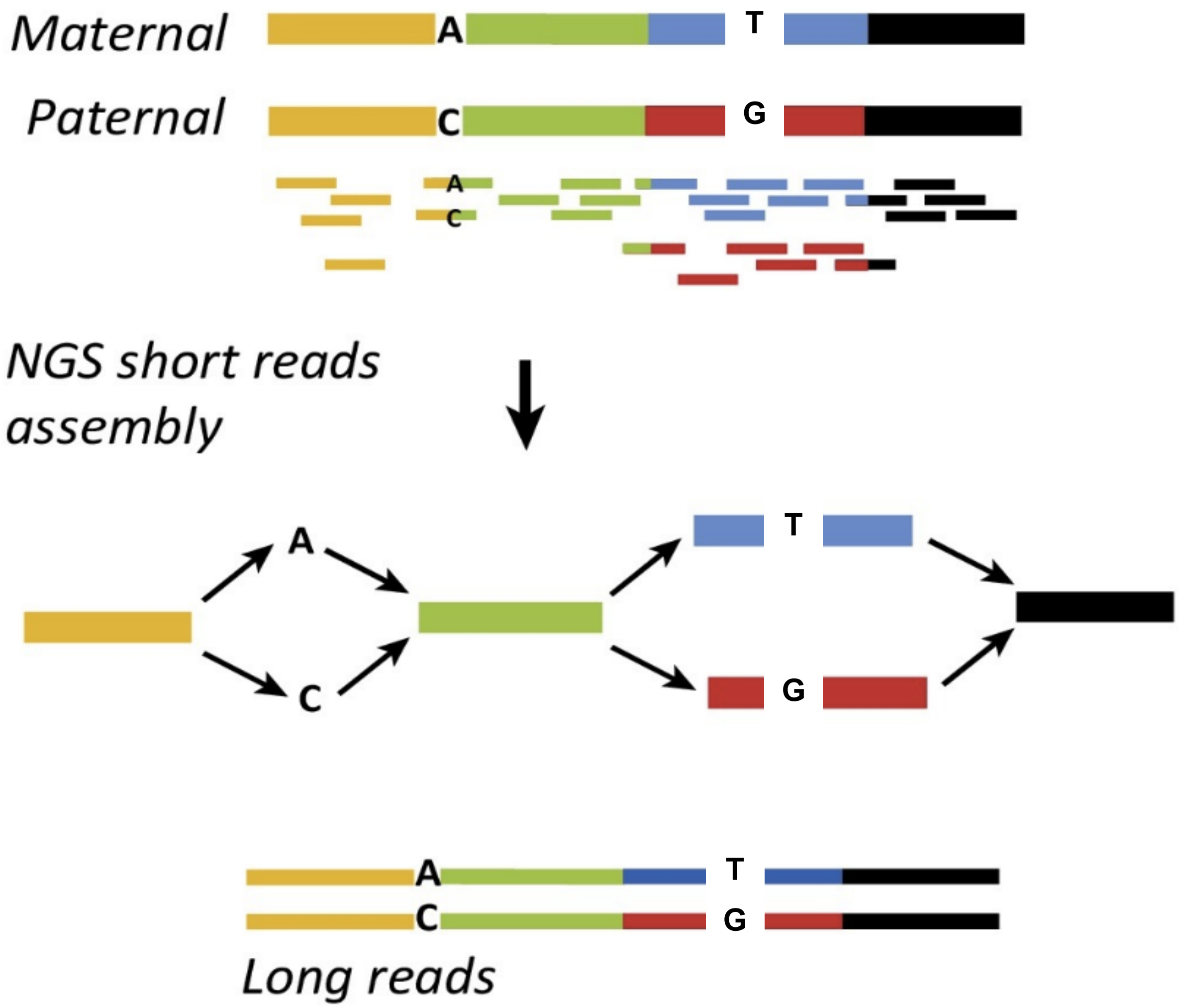
Several contigs ⟶ incomplete assembly, underestimation of repeats

Long reads assembly



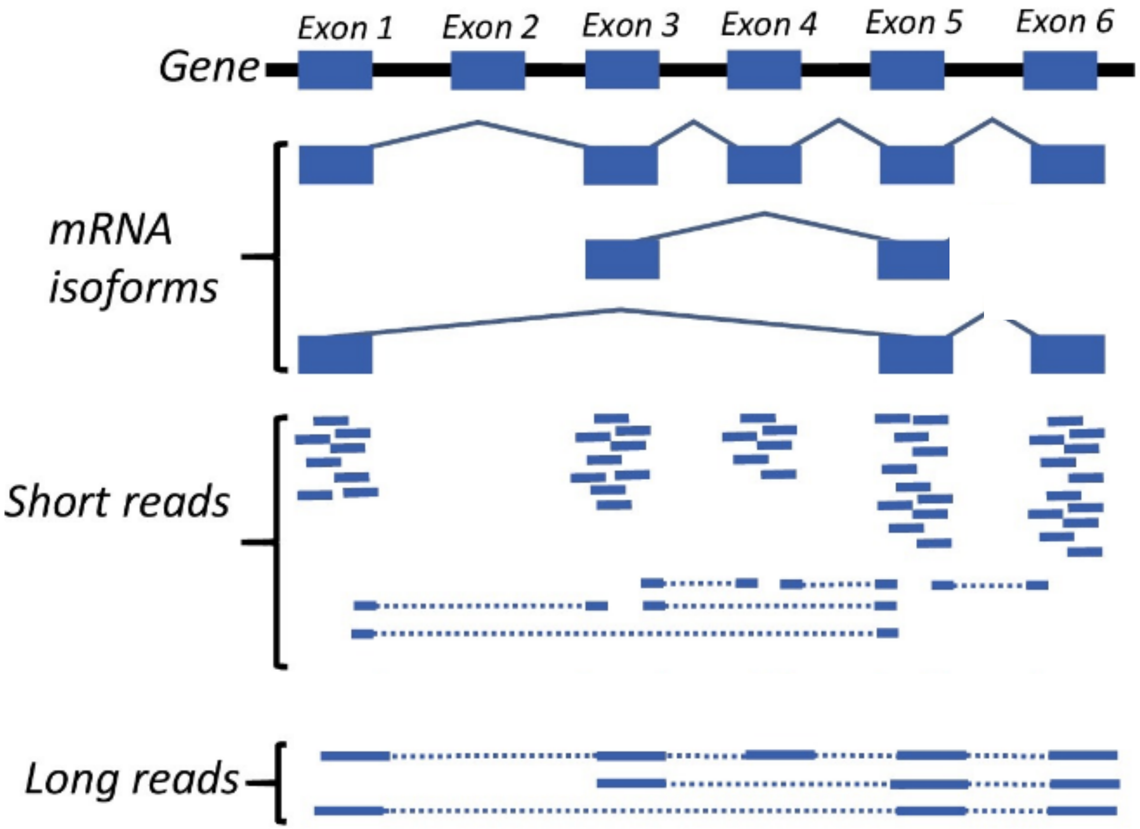
LONG-READS VERSUS SHORT-READS

Haplotype phasing



LONG-READS VERSUS SHORT-READS

Detection of splicing isoforms



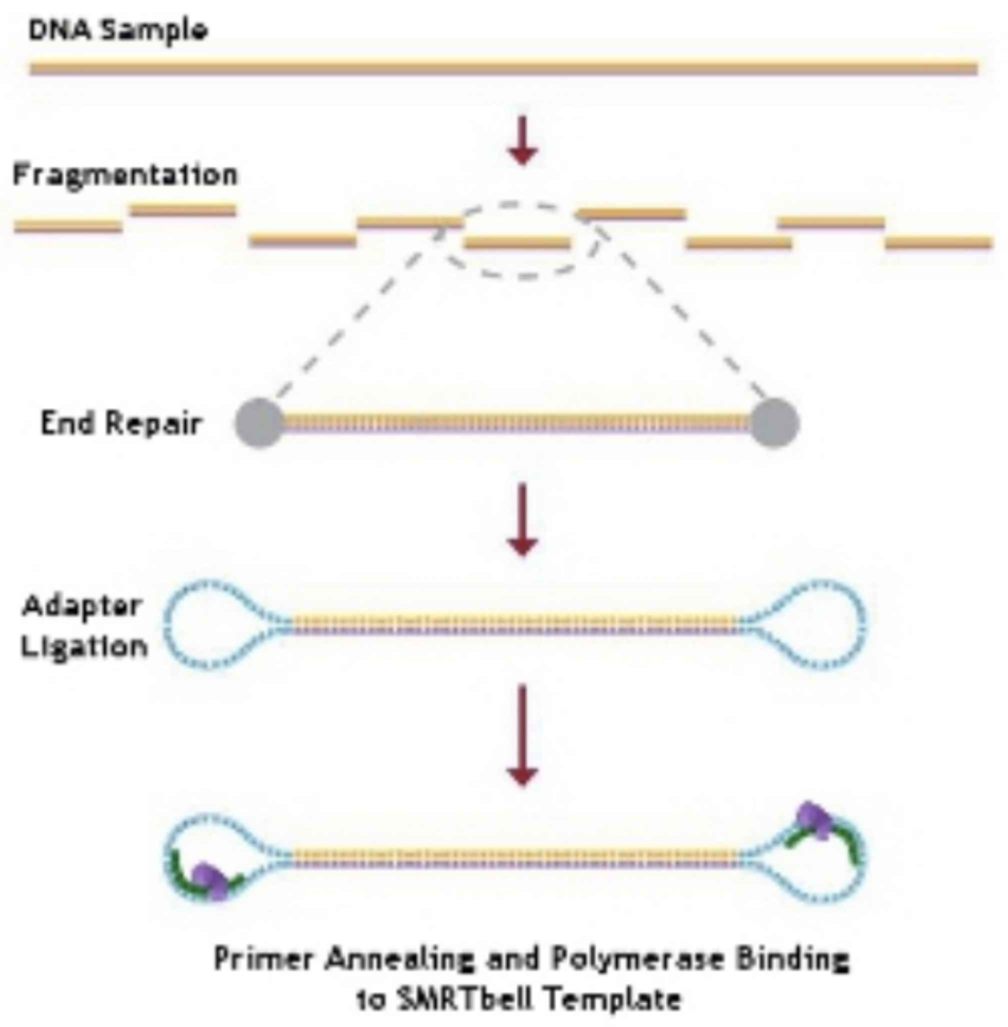
The 3rd generation winning technologies

Sequel - Pacific Biosciences
Single molecules
Up to 150 kbp long
Error rate ≈ 10-15 % - CCS: <1%

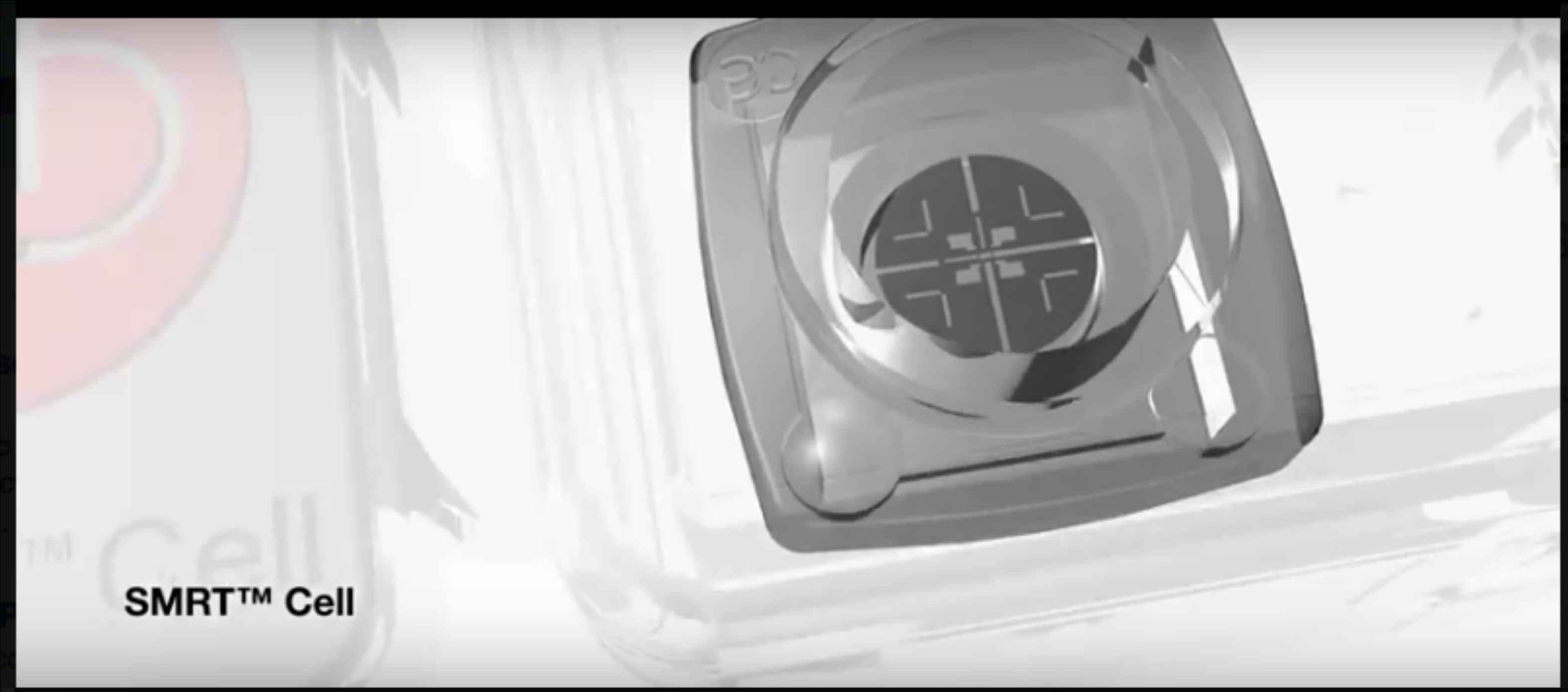
MinION - Oxford Nanopore
Single molecules
Up to 1 Mbp long
Error rate ≈ 10-15 %
Compensated by coverage

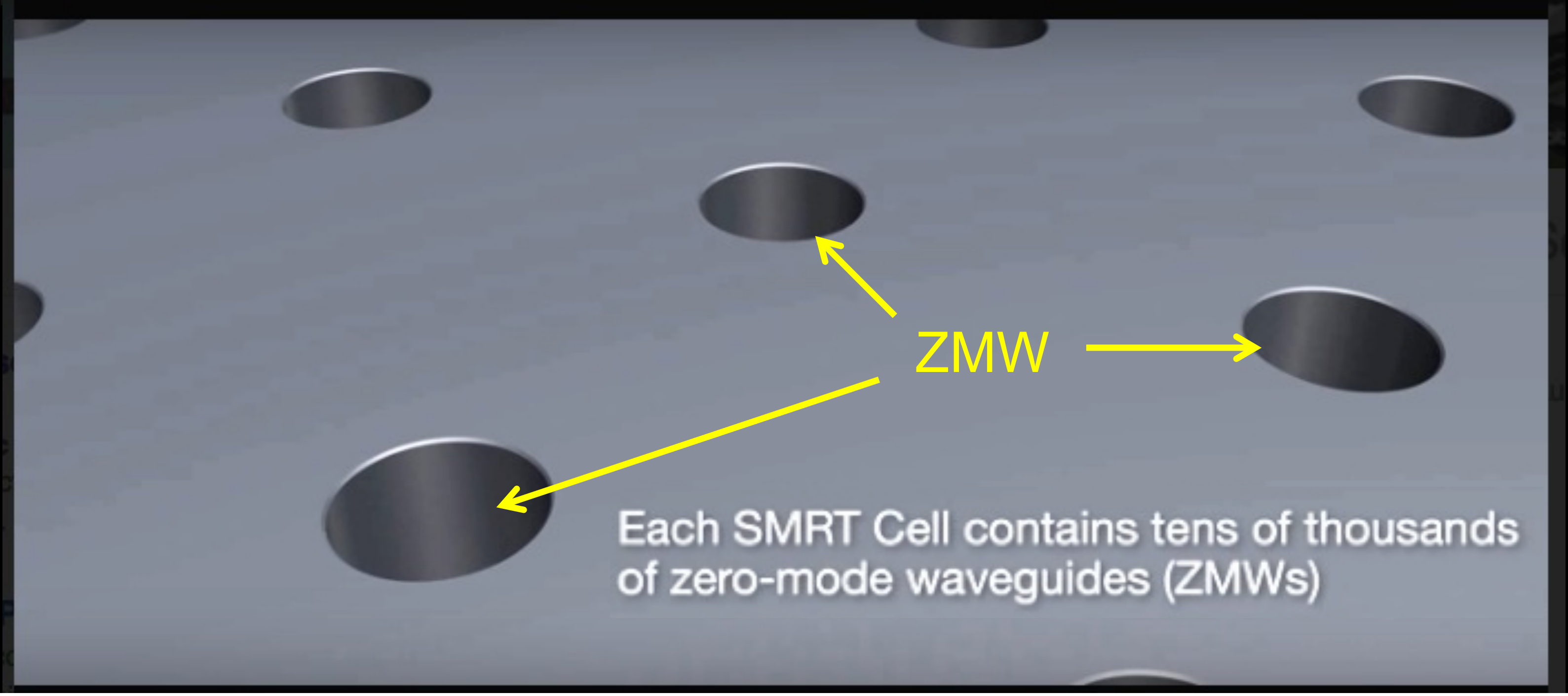
PacBio : Single Molecule Real Time (SMRT) sequencing

PacBio DNA-seq library


DNA Sample
Fragmentation
End Repair
Adapter Ligation
Primer Annealing and Polymerase Binding to SMRTbell Template

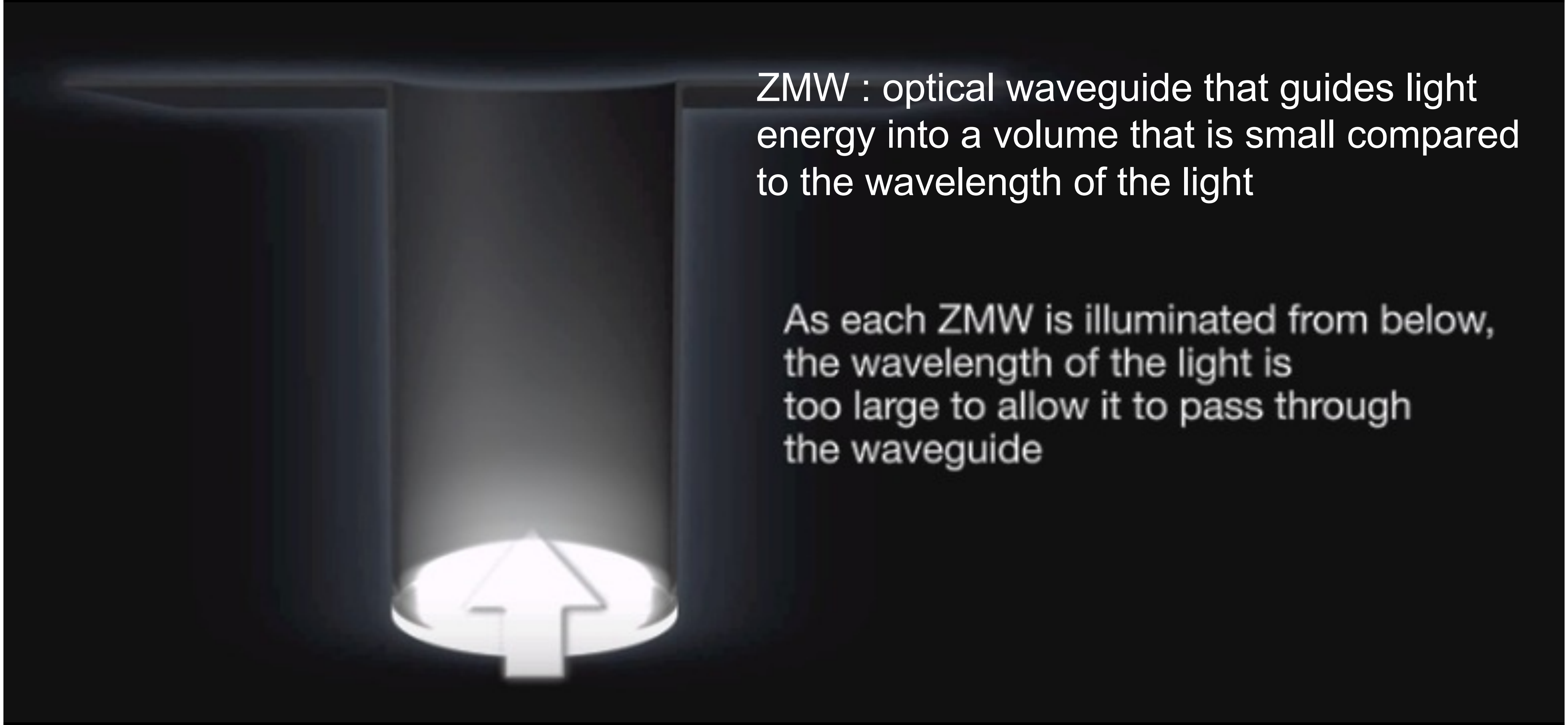
PACIFIC BIOSCIENCES





Each SMRT Cell contains tens of thousands of zero-mode waveguides (ZMWs)

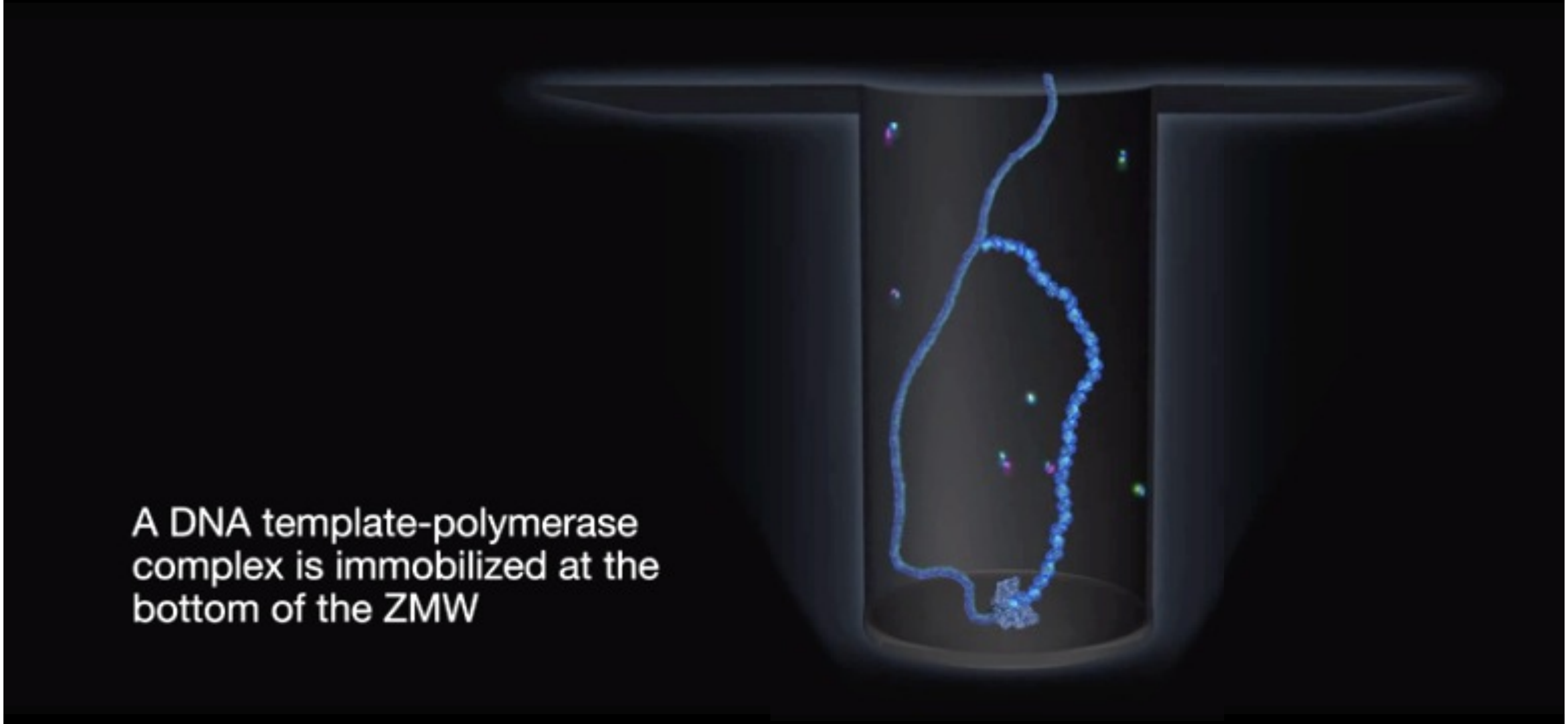
PACIFIC BIOSCIENCES



ZMW : optical waveguide that guides light energy into a volume that is small compared to the wavelength of the light

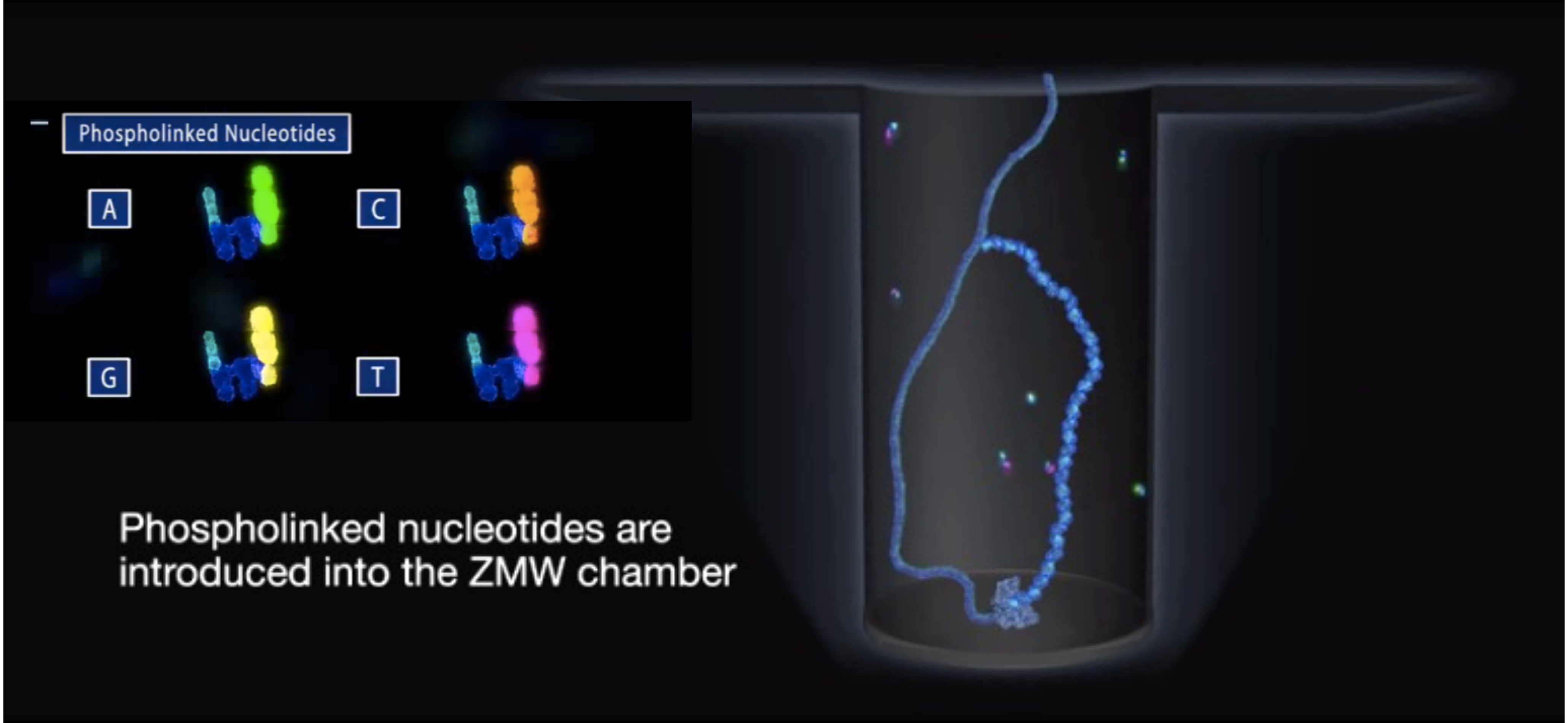
As each ZMW is illuminated from below, the wavelength of the light is too large to allow it to pass through the waveguide

PACIFIC BIOSCIENCES



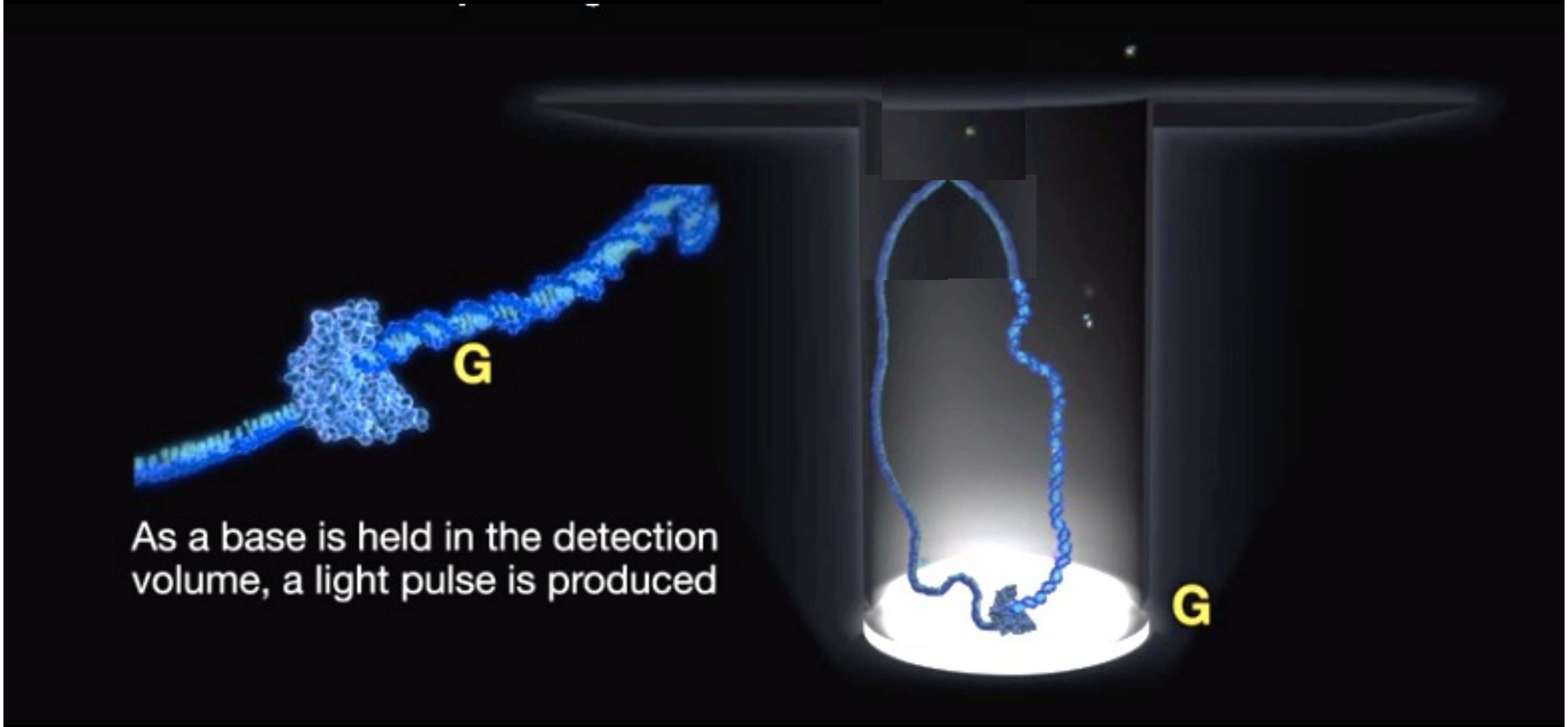
A DNA template-polymerase complex is immobilized at the bottom of the ZMW

PACIFIC BIOSCIENCES



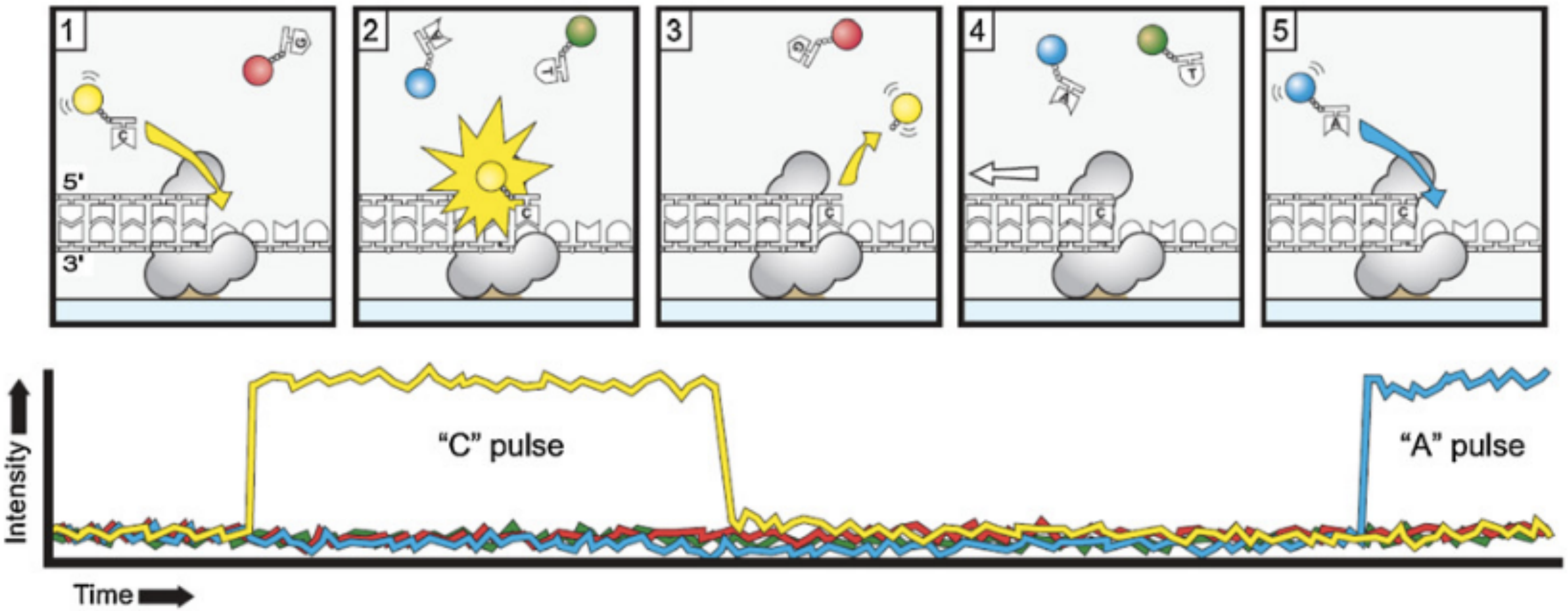
Phospholinked nucleotides are introduced into the ZMW chamber

PACIFIC BIOSCIENCES



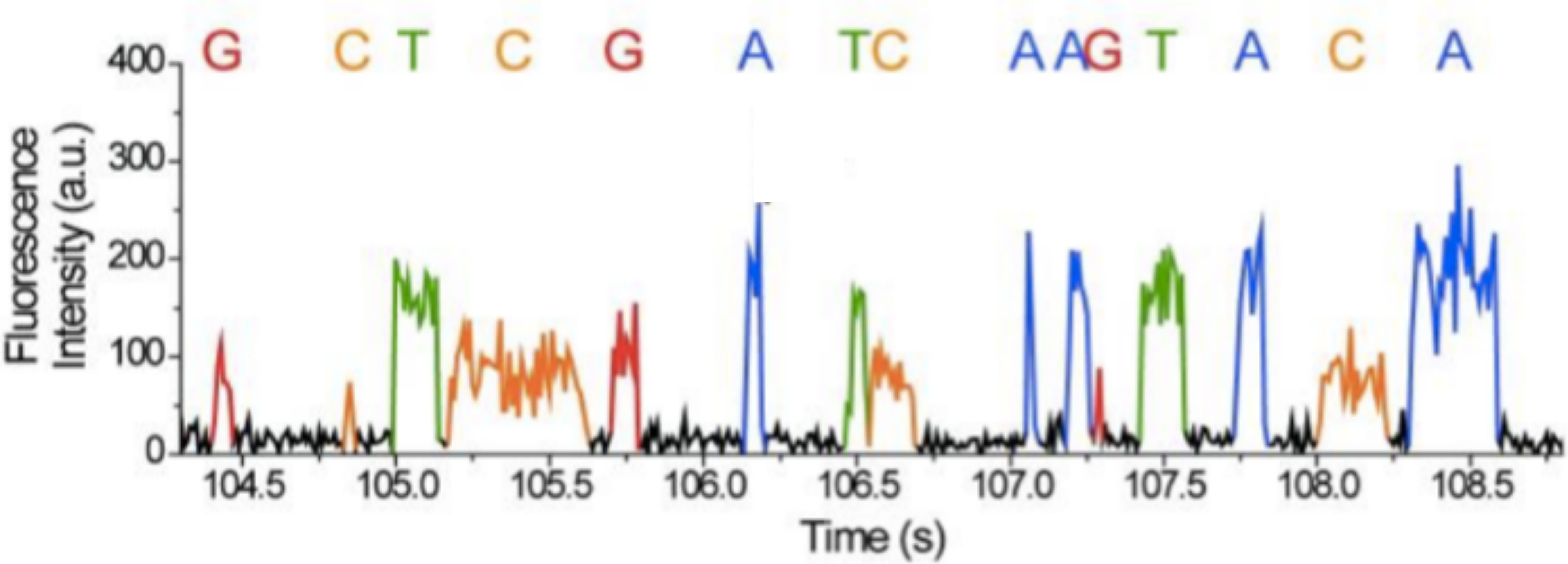
As a base is held in the detection volume, a light pulse is produced

PACIFIC BIOSCIENCES

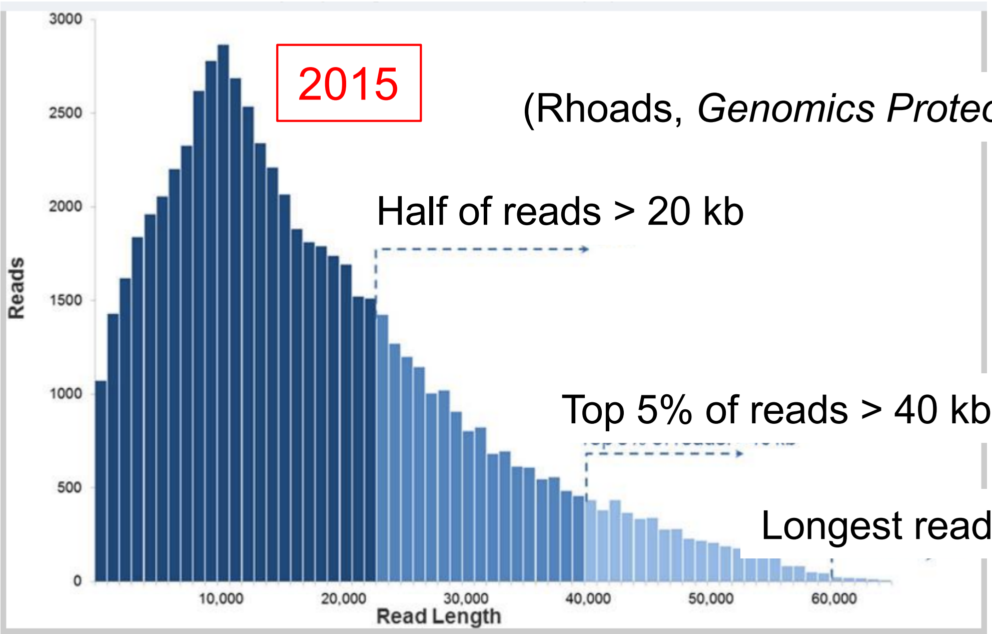


Eid, J., et al. Science (2009)

PACIFIC BIOSCIENCES



Length of PacBio reads



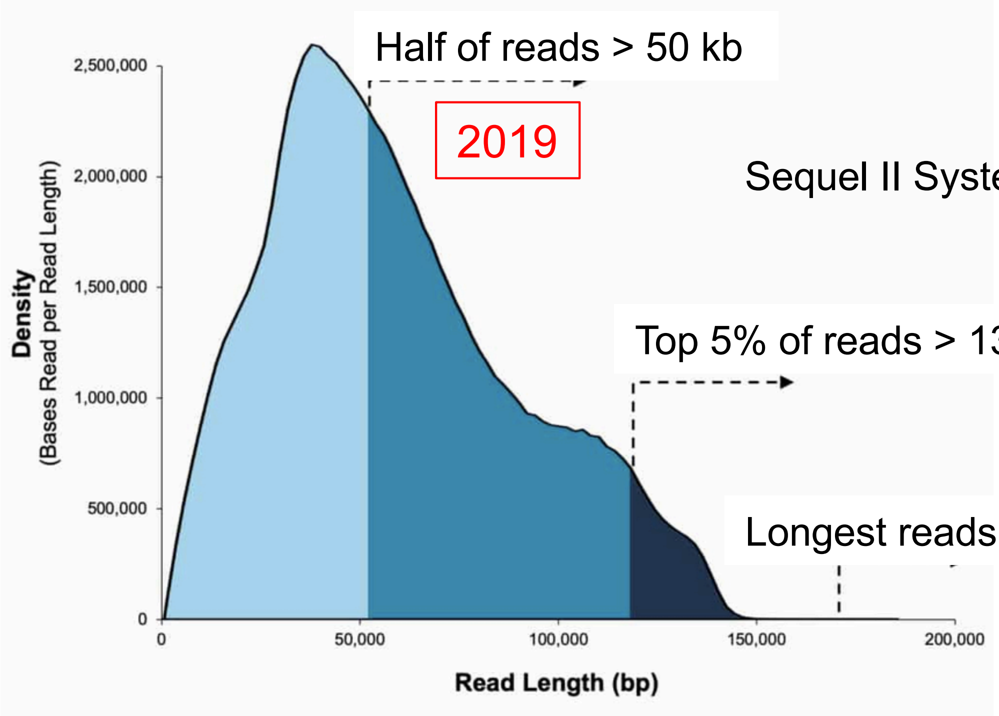
2015

(Rhoads, *Genomics Proteomics Bioinformatics*, 2015)

Half of reads > 20 kb

Top 5% of reads > 40 kb

Longest reads > 40 kb



Half of reads > 50 kb

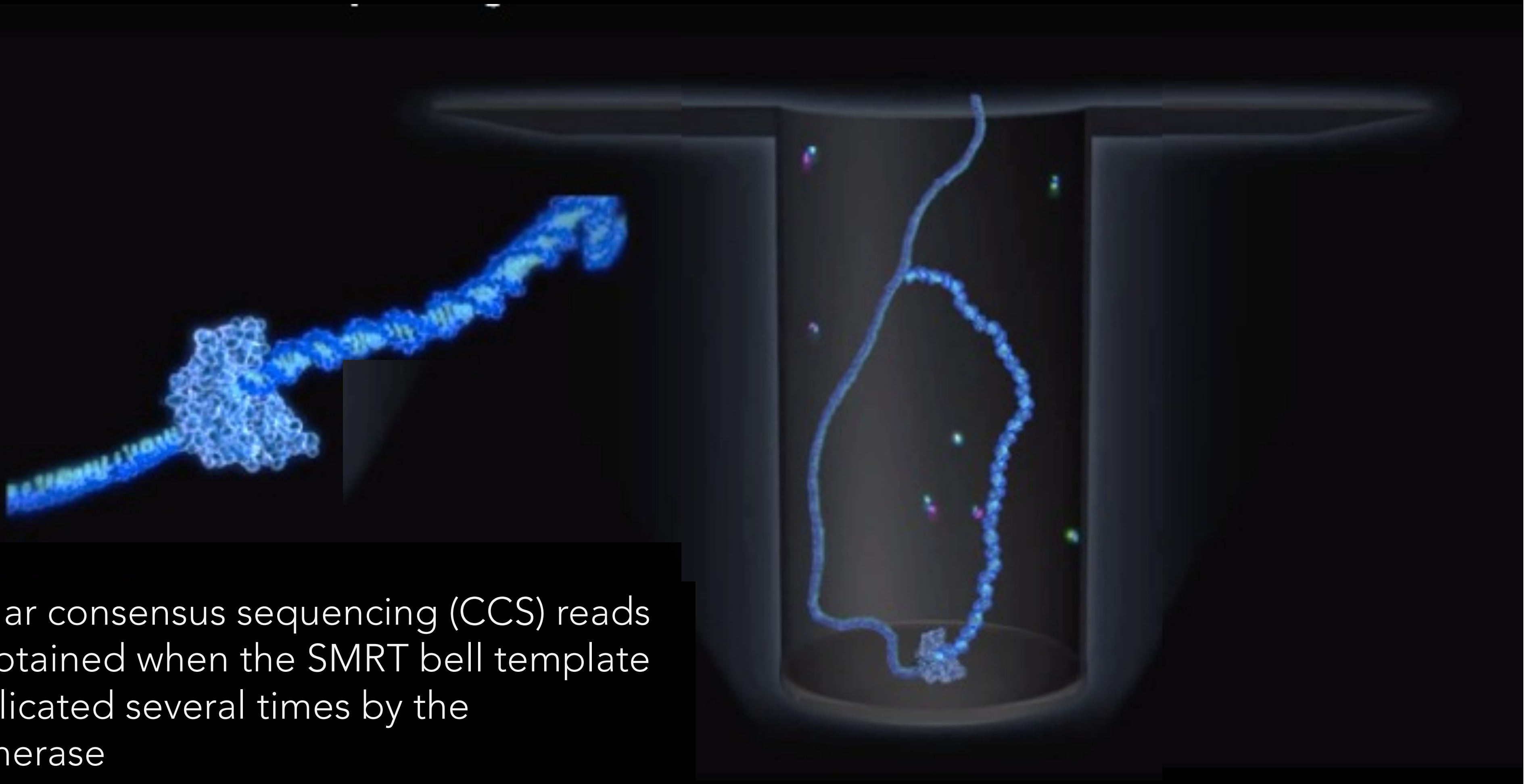
2019

Sequel II System ; 2.0 Chemistry

Top 5% of reads > 135 kb

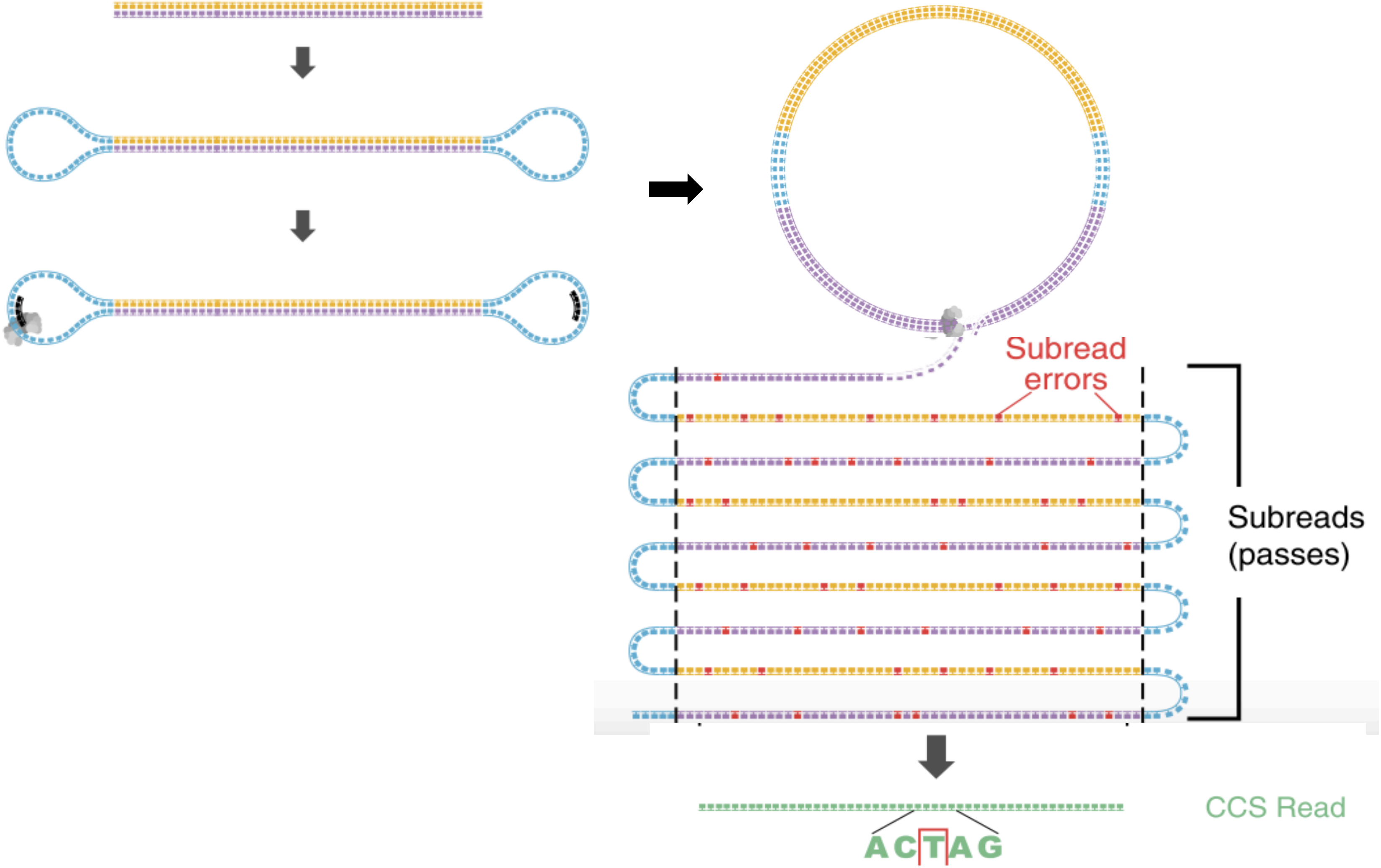
Longest reads > 175 kb

Improvement with new chemistry : Circular Consensus Sequence (CCS)



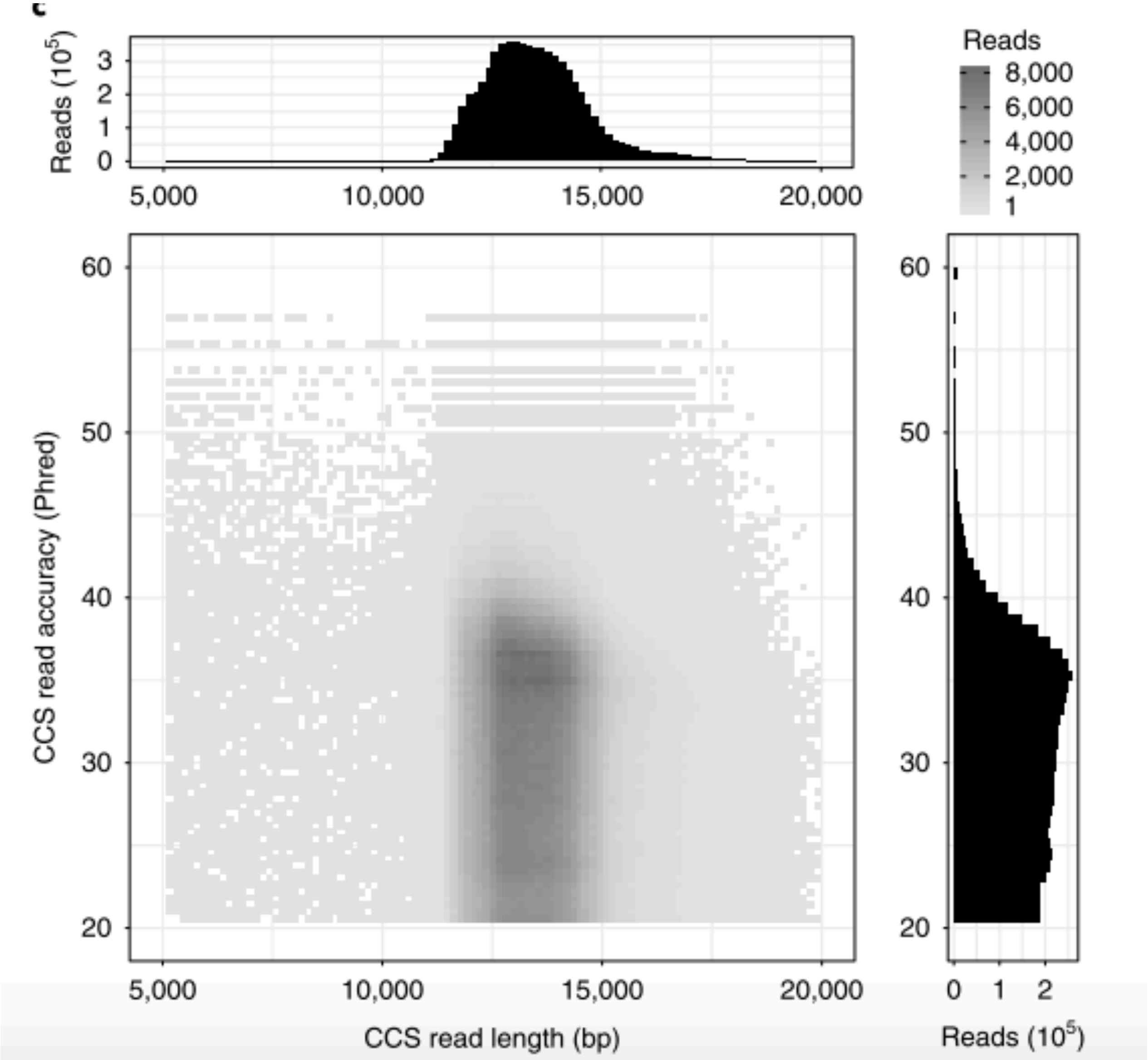
Circular consensus sequencing (CCS) reads are obtained when the SMRT bell template is replicated several times by the polymerase

Circular Consensus Sequences (CCS): HIFI READS



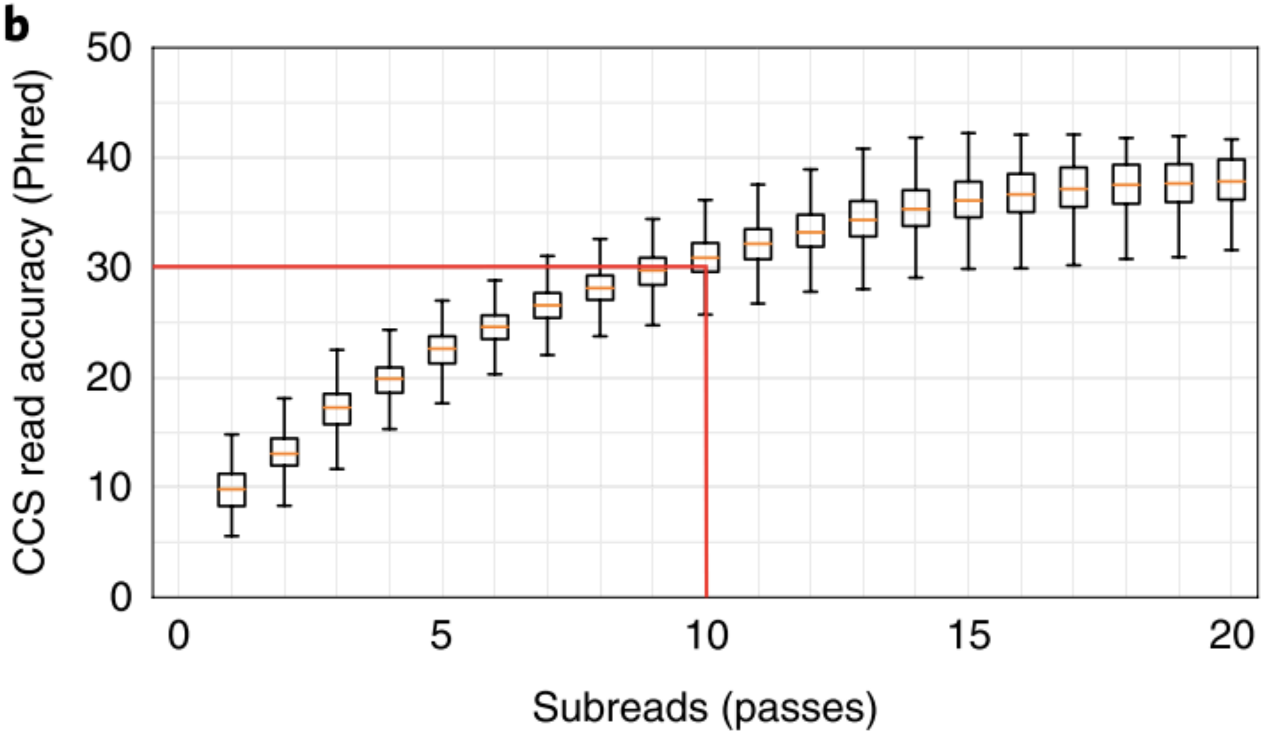
GENOME ASSEMBLY WITH CCS

Circular consensus assembly of a human genome
Wenger et al. *Nat. Biotechnol.* oct. 2019



GENOME ASSEMBLY WITH CCS

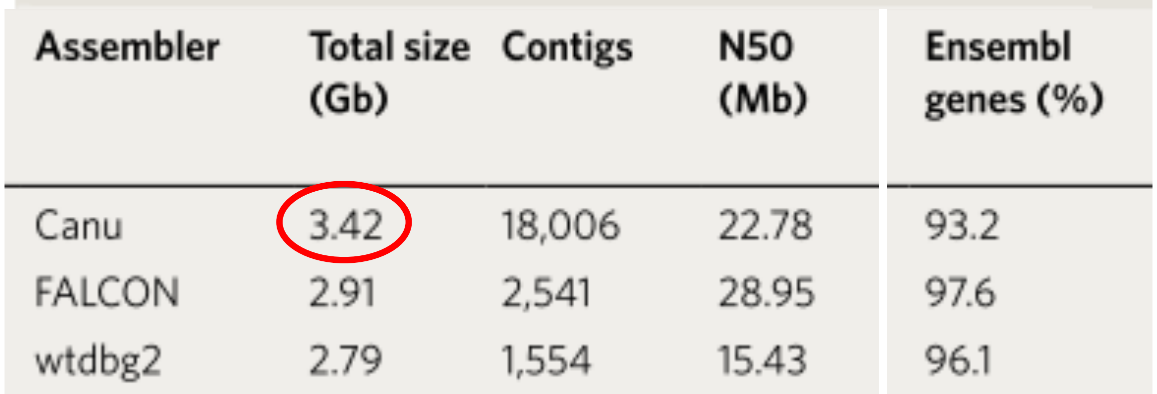
Circular consensus assembly of a human genome
Wenger et al. *Nat. Biotechnol.* (2019)



1. GENOME ASSEMBLY WITH CCS

Circular consensus assembly of a human genome
Wenger et al. *Nat. Biotechnol.* (2019)

CCS reads alone : high quality contiguous genome : concordance of 99.997%



Assembler	Total size (Gb)	Contigs	N50 (Mb)	Ensembl genes (%)
Canu	3.42	18,006	22.78	93.2
FALCON	2.91	2,541	28.95	97.6
wtdbg2	2.79	1,554	15.43	96.1

Canu assembly

- genome size > expected haploid genome because it resolves some heterozygous alleles into separate contigs

Majority of CCS read discordances

- 3.4% mismatches → 1 mismatch every 13,048 bp
- 4.6% indels in non homopolymers. → 1 non-homopolymer indel every 9,669 bp
- 92.0% indels in homopolymers → 1 homopolymer indel every 477 bp

Comparison with NovaSeq

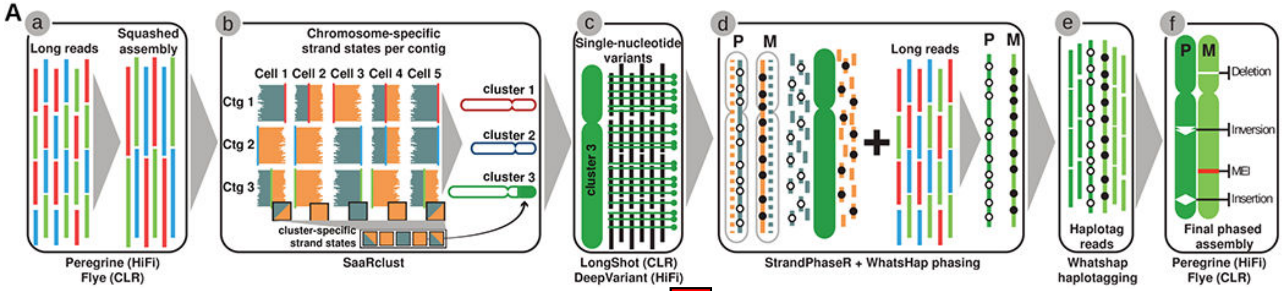
- CCS mismatch rate is 17× lower than reads from NovaSeq
- CCS indel rate is 181× higher than reads from NovaSeq

2. GENOME ASSEMBLY WITH CCS

Haplotype-resolved diverse human genomes and integrated analysis of structural variation
Ebert et al. *Science* April 2021

New methodology to produce fully phased diploid genome assemblies that combines :
- long-read PacBio
- Strand-seq Illumina

Methodology
1. generation of a non-haplotype-resolved clustered assembly
2. clustering of assembled contigs into "chromosome" clusters based on Strand-seq Illumina
3. calling of single-nucleotide variants (SNVs) relative to the clustered assembly
4. chromosome-wide phasing
5. tagging of input long reads by haplotype
6. phased genome assembly based on haplotagged long reads



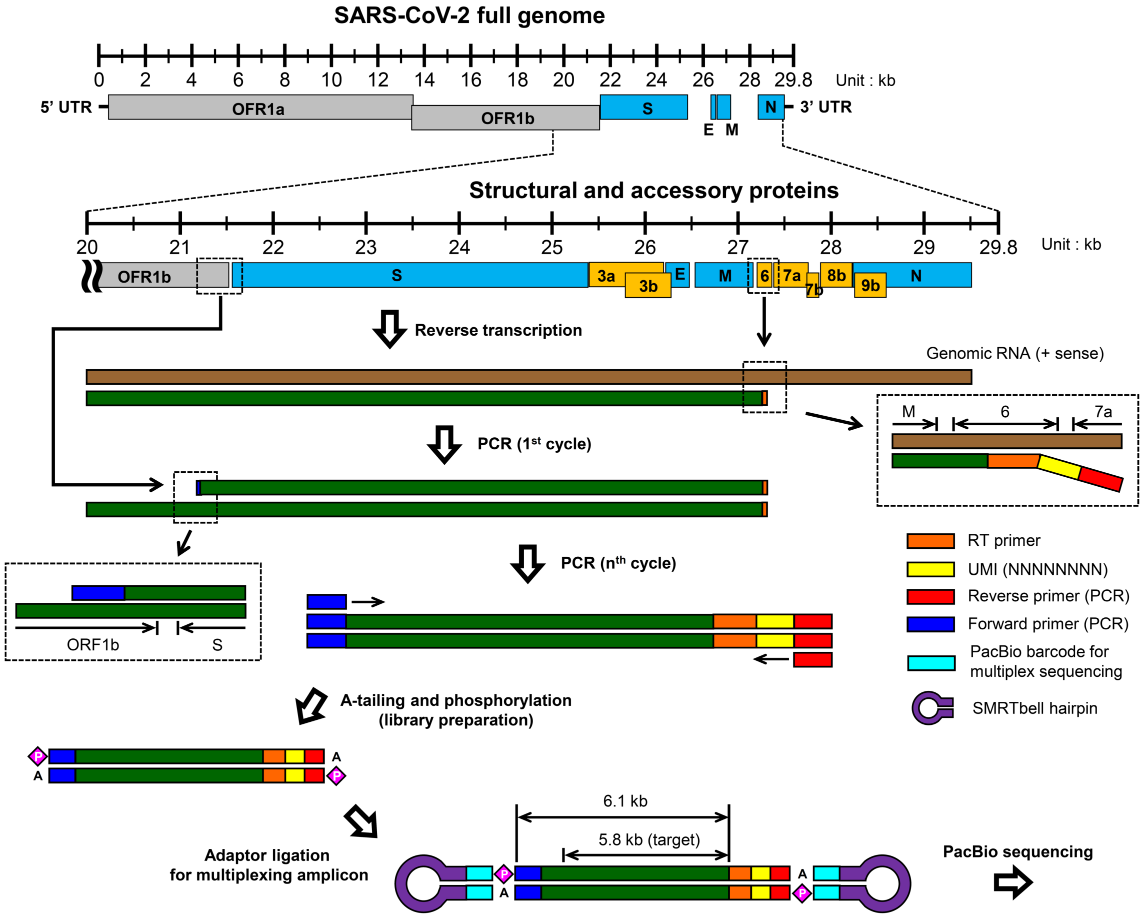
64 ASSEMBLED HAPLOTYPES FROM 32 DIVERSE HUMAN GENOMES

- Comparison of these 32 Highly contiguous phased haplotype assemblies allows identification of :
  - 107,590 structural variants of which 68% not discovered by short-read sequencing
  - By contrast, analysis of 2,504 short-read sequenced genomes (1000GP) reported 69,000 SVs

3. GENOME ASSEMBLY WITH CCS

High-throughput, single-copy sequencing reveals SARS-CoV-2 spike variants coincident with mounting humoral immunity during acute COVID-19
Ko S.H. et al. *PLOS Pathogens* 2021

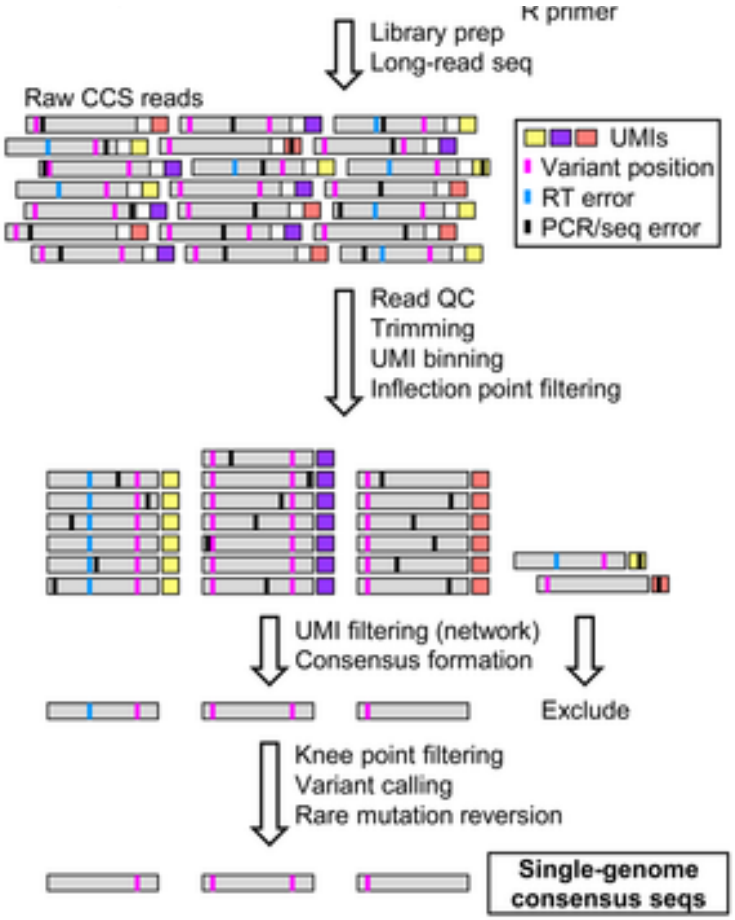
Study of intra-individual evolution of SARS-CoV-2 : standard sequencing yields single consensus sequence for each sample, rather than multiple sequences representing virus quasispecies diversity.

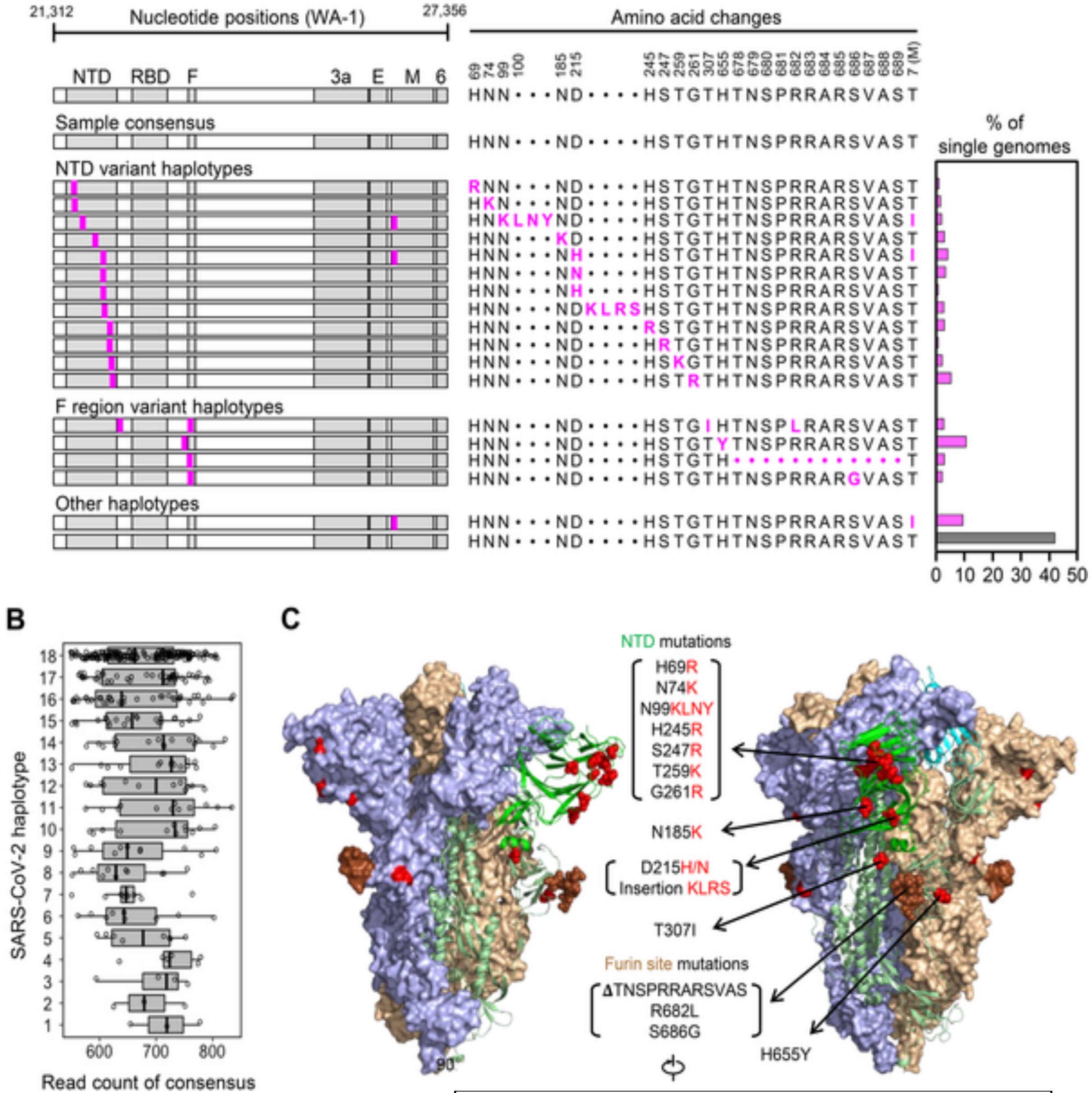


3. GENOME ASSEMBLY WITH CCS

Each sequence corresponds to a single viral genome

Analysis of CCS reads

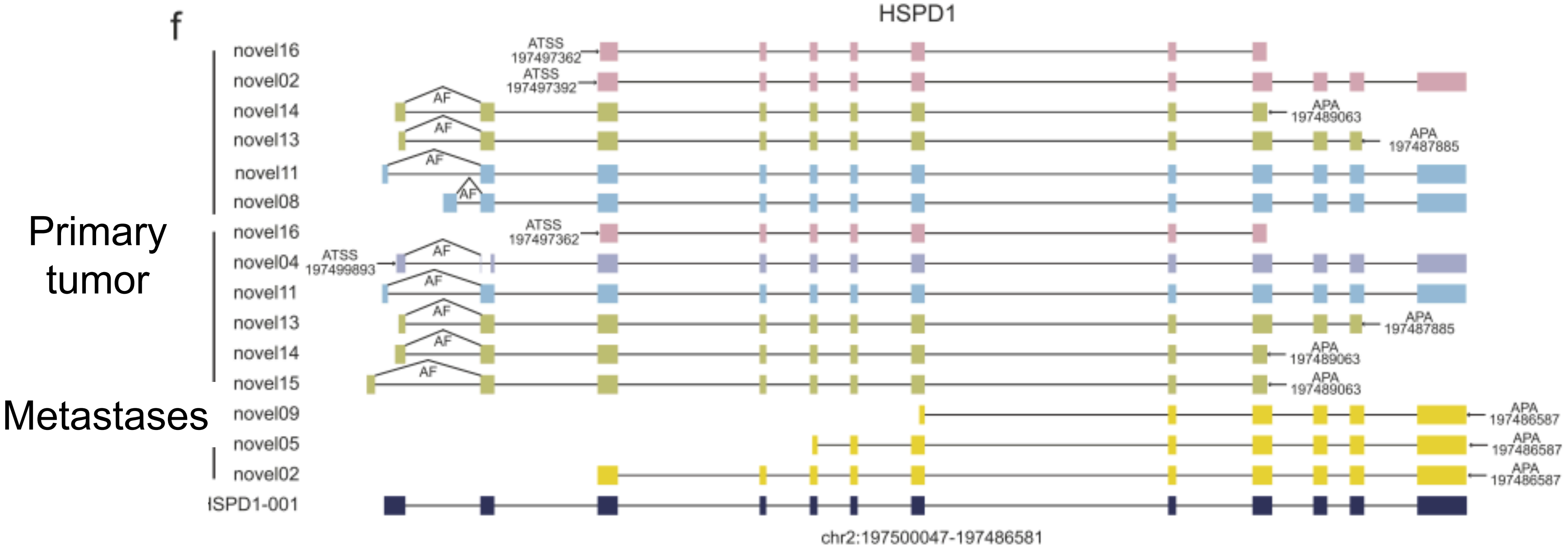




Ko S.H. et al. *PLOS Pathogens* 2021

cDNA SEQUENCING

Hybrid full-length transcriptome in metastatic ovarian cancer
Jing et al. *Oncogene* 2019



Long-read full-length transcriptome analysis :

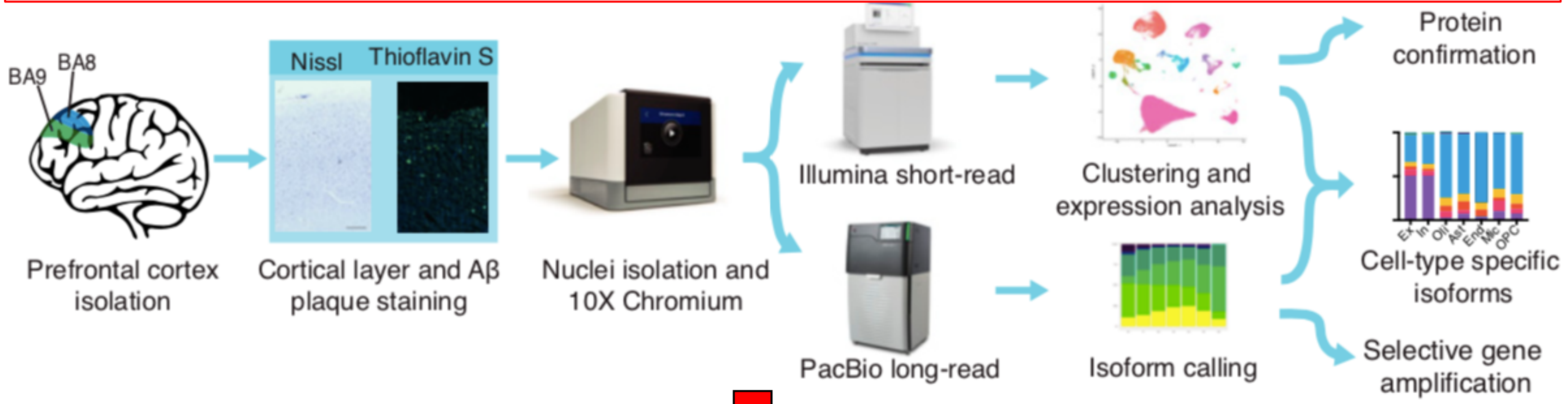
- improves molecular diagnostic

SINGLE CELL PacBio cDNA SEQUENCING

Altered cell and RNA isoform diversity in aging Down syndrome brains
Palmer et al. *PNAS* Aug. 2021

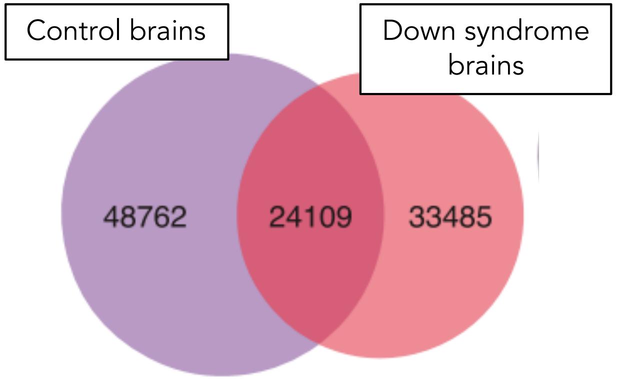
Down syndrome (trisomy 21) :

- single-nucleus long read RNA sequencing
- >170,000 cells from 29 aging DS and control brains

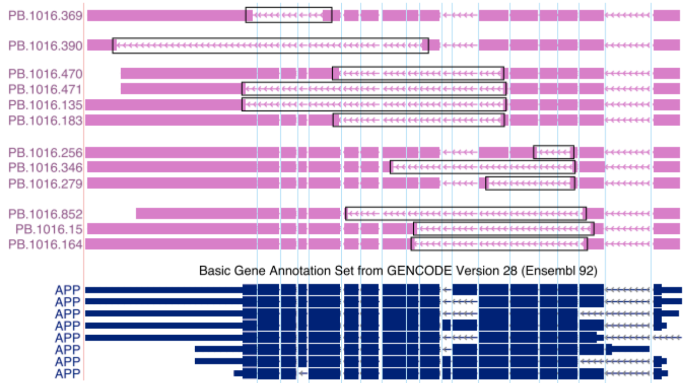

BA9
BA8
Prefrontal cortex isolation
Nissl
Thioflavin S
Cortical layer and Aβ plaque staining
Nuclei isolation and 10X Chromium
Illumina short-read
PacBio long-read
Clustering and expression analysis
Isoform calling
Protein confirmation
Ex
In
Oli
Ast
End
Mic
OPC
Cell-type specific isoforms
Selective gene amplification

New splicing isoforms :

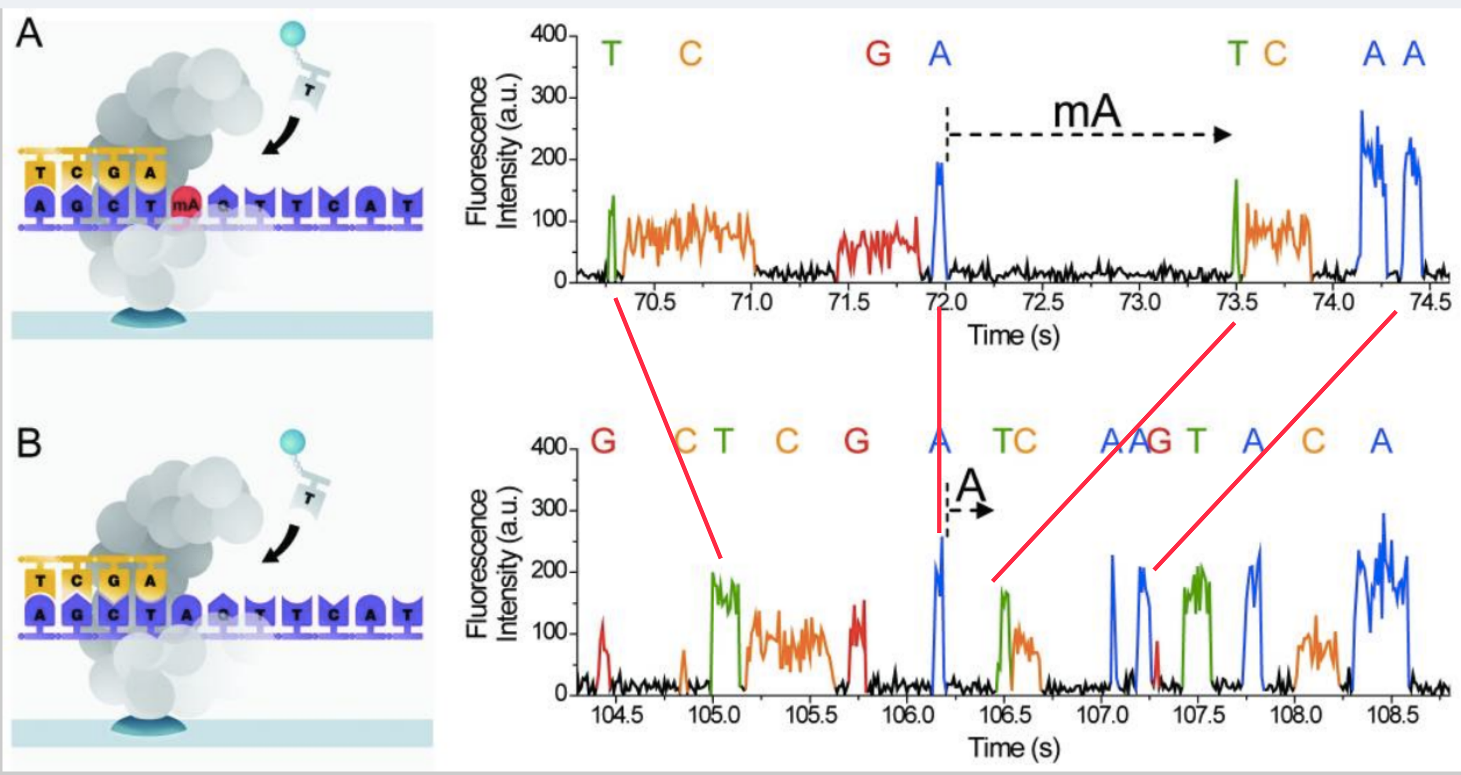
- new splice sites
- novel exon junctions
- entirely new exons
- intron retention



Amyloid precursor protein (Alzheimer's disease gene)



DETECTION OF MODIFIED DNA BASES



from Fusberg et al. *Nature Methods* (2010)

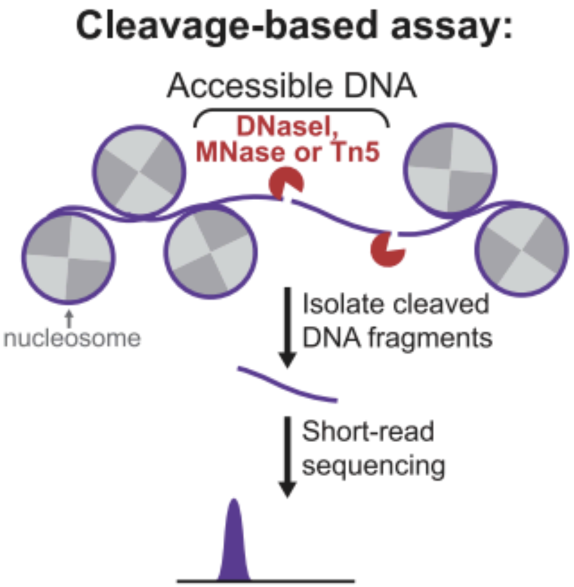
Detection of 5mA with strong influence of sequence contexts : requires high coverage

Feng et al. *PLOS Comput Biol* (2013)

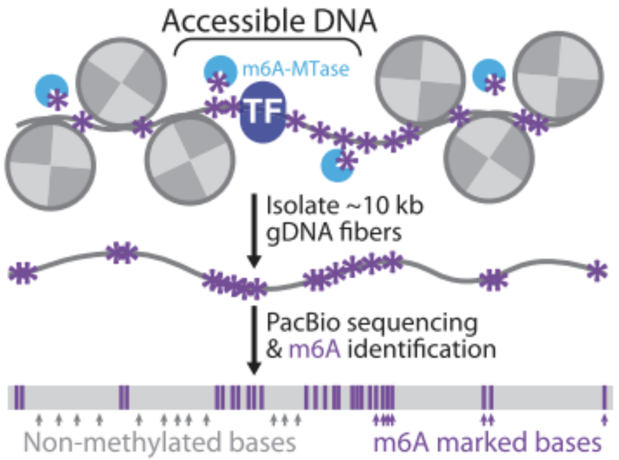
DETECTION OF DNA m6A WITH CCS

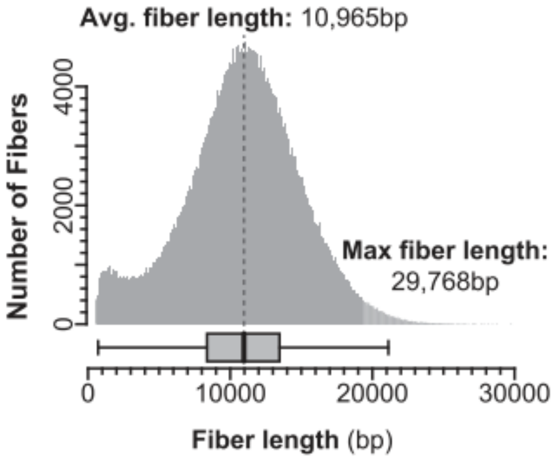
Single-molecule regulatory architectures captured by chromatin fiber sequencing
Stergachis et al. *Science* (2020)

Dnasel-seq.



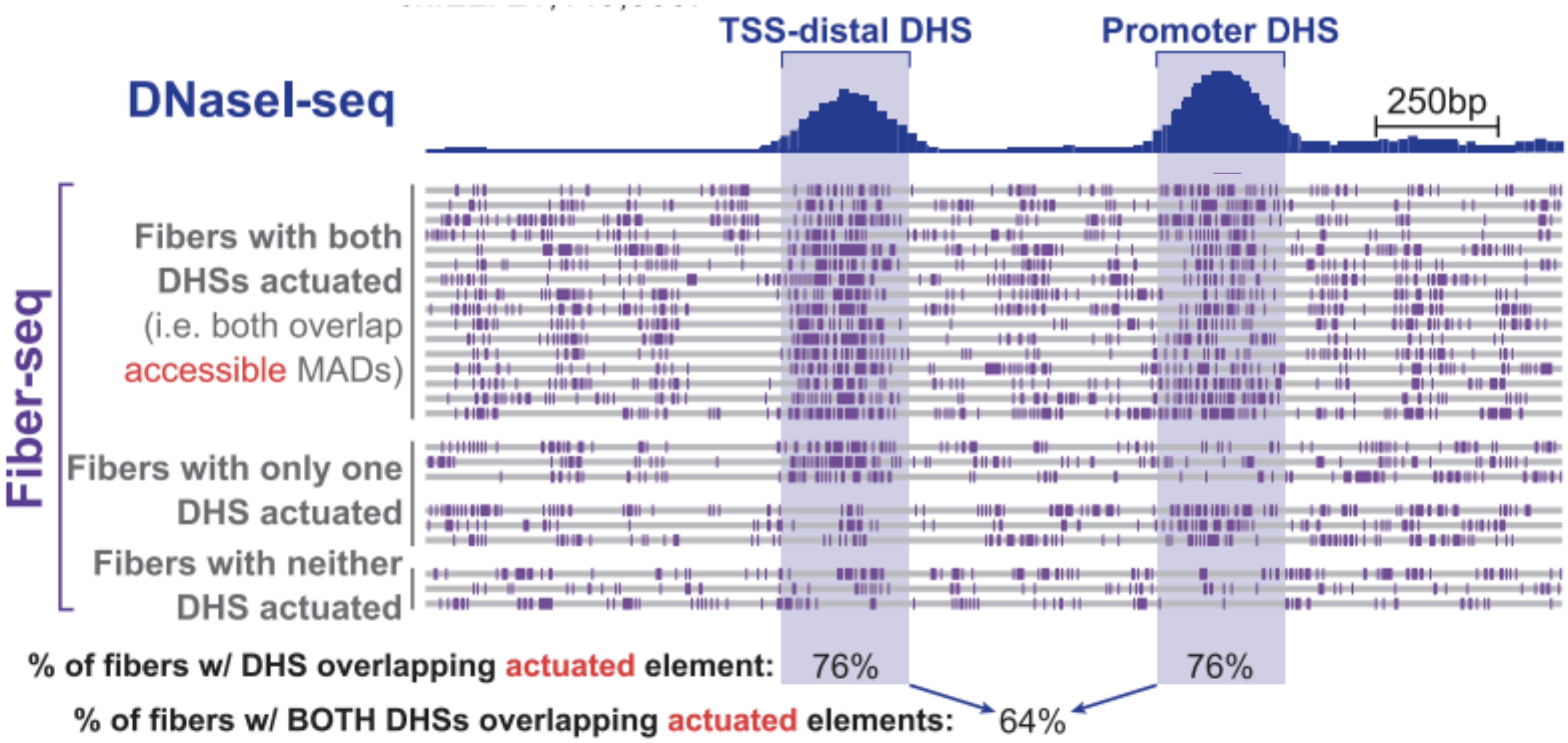
Fiber-seq.


Accessible DNA
m6A-MTase
TF
Isolate ~10 kb gDNA fibers
PacBio sequencing & m6A identification
Non-methylated bases
m6A marked bases



DETECTION OF DNA m6A WITH CCS

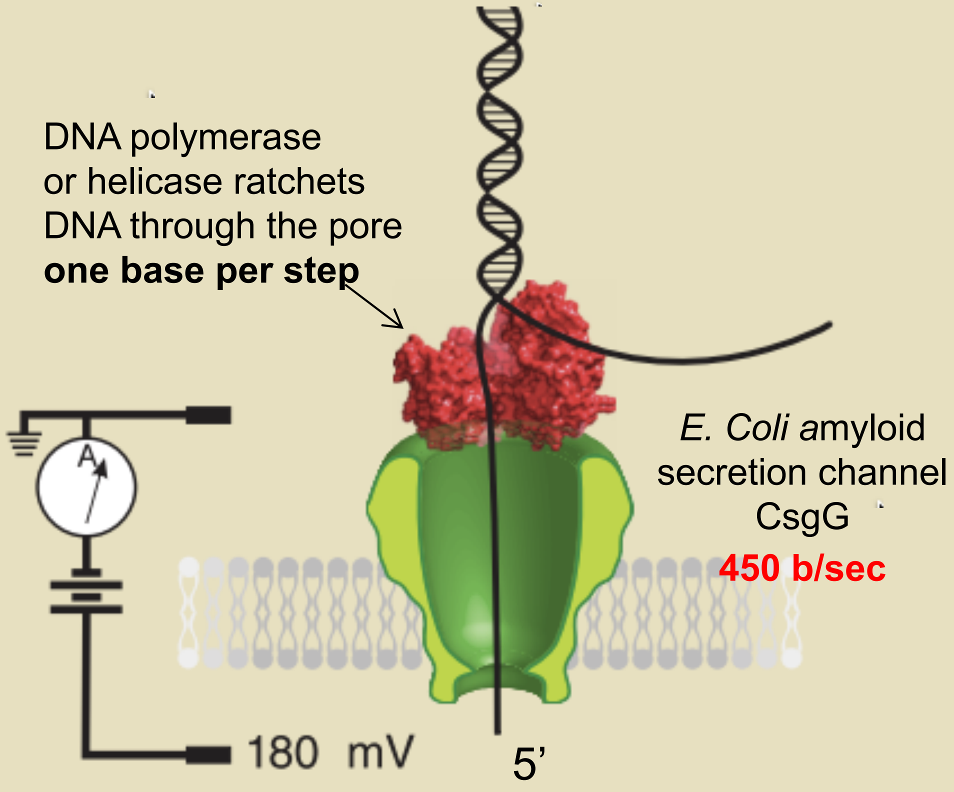
Single-molecule regulatory architectures captured by chromatin fiber sequencing
Stergachis et al. *Science* (2020)



Next Generation Sequencing

Oxford NANOPORE Technologies®

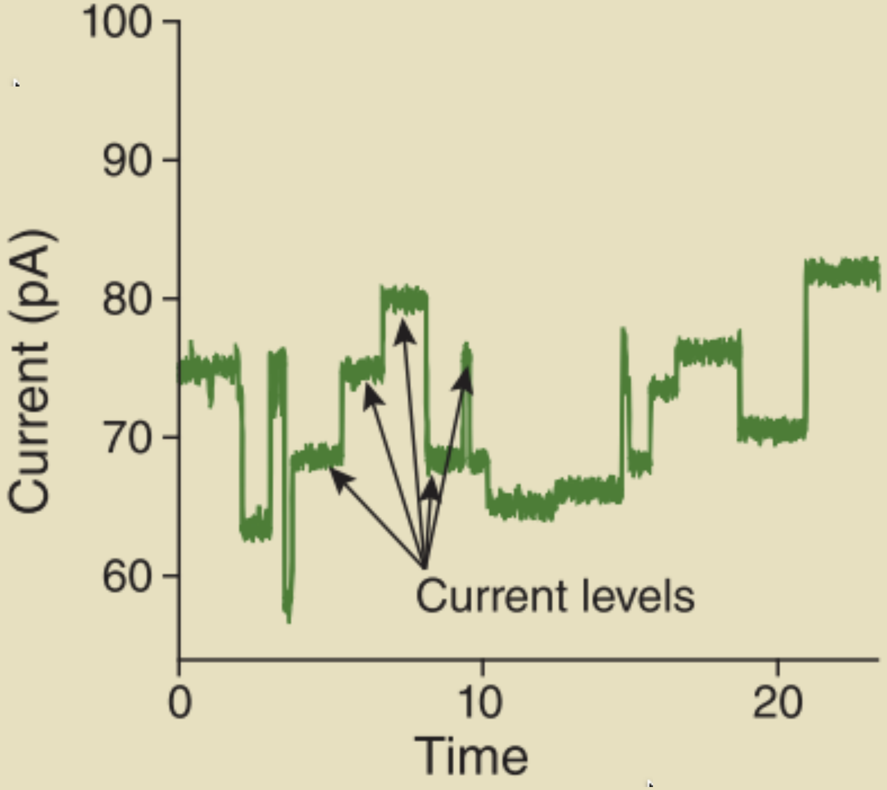
BASIC CONCEPTS



DNA polymerase
or helicase ratchets
DNA through the pore
one base per step

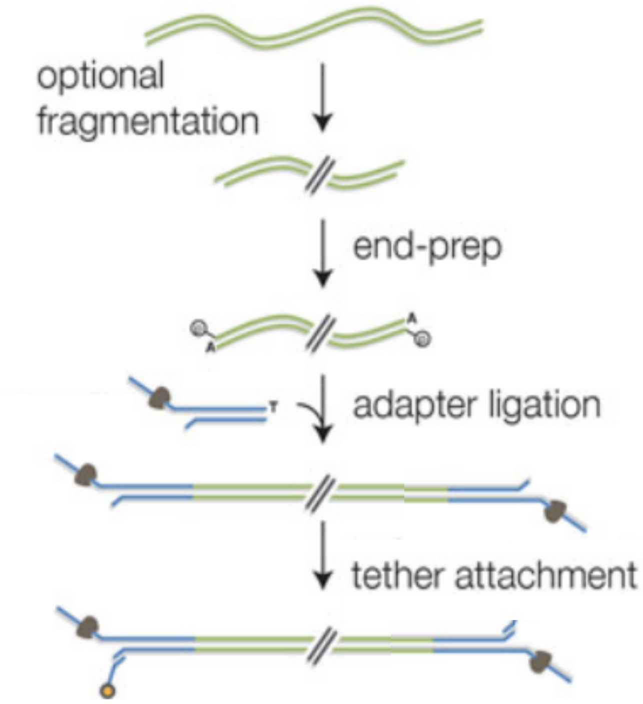
*E. Coli a*myloid
secretion channel
CsgG

450 b/sec

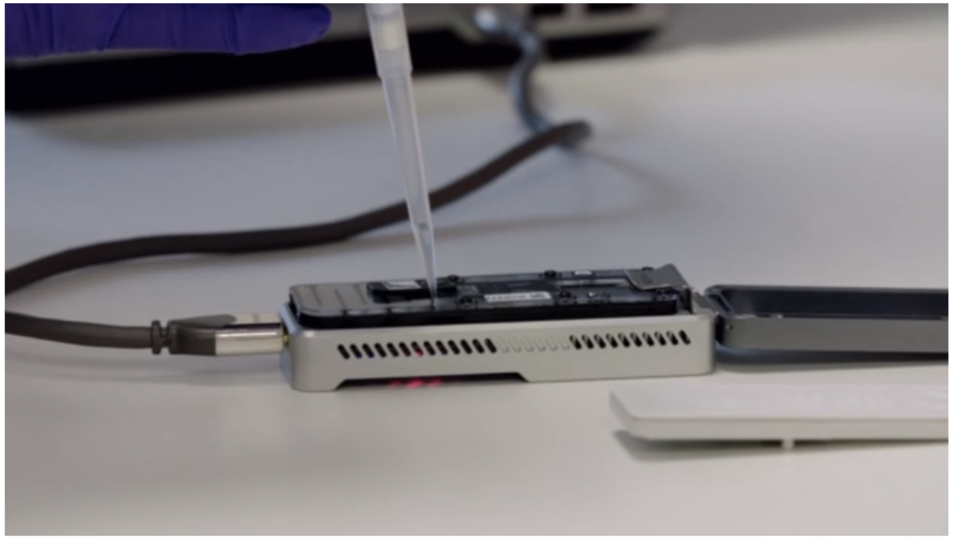


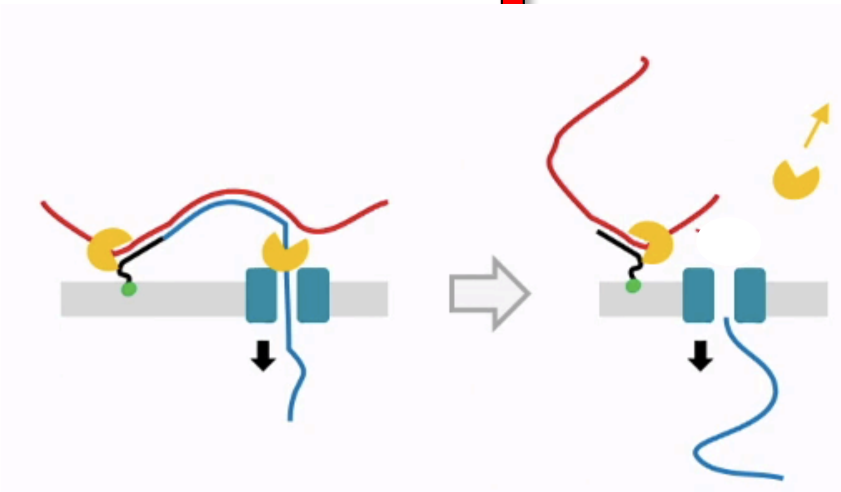
SEQUENCING PROCESS

Library preparation

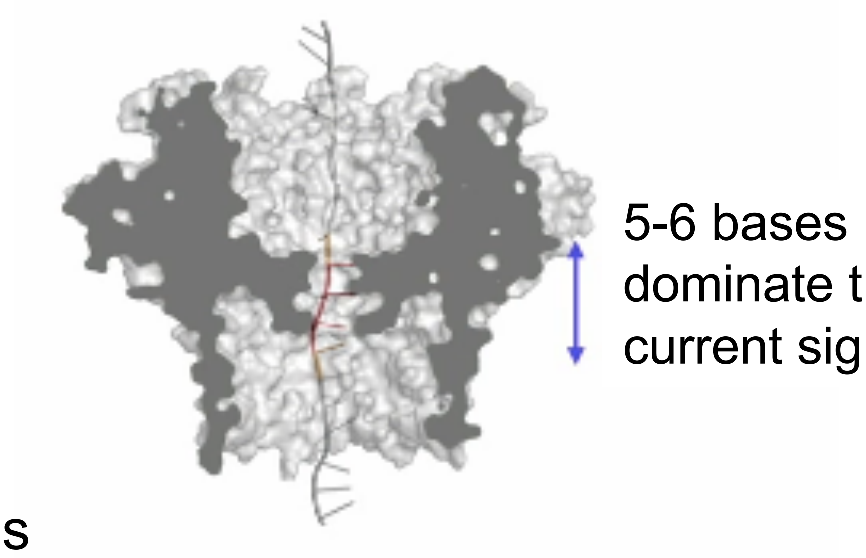


SEQUENCING



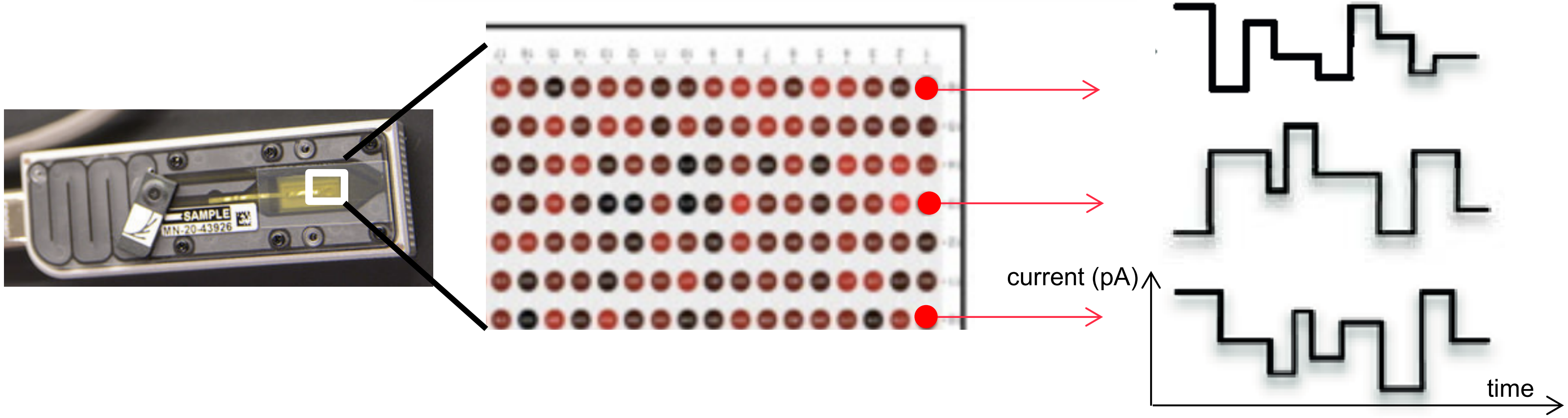


SEQUENCING PROCESS : MinION FLOW CELL



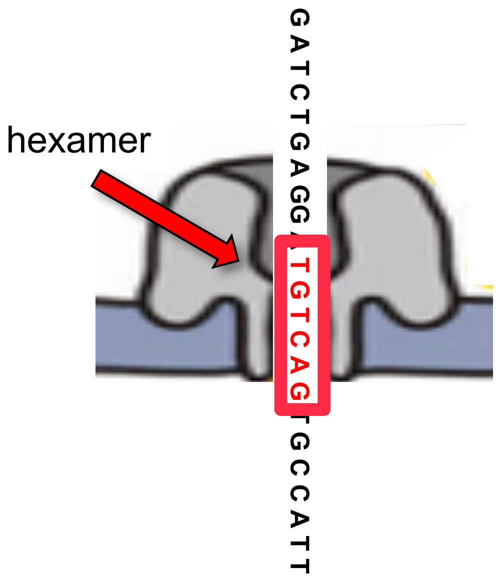
5-6 bases dominate the current signal

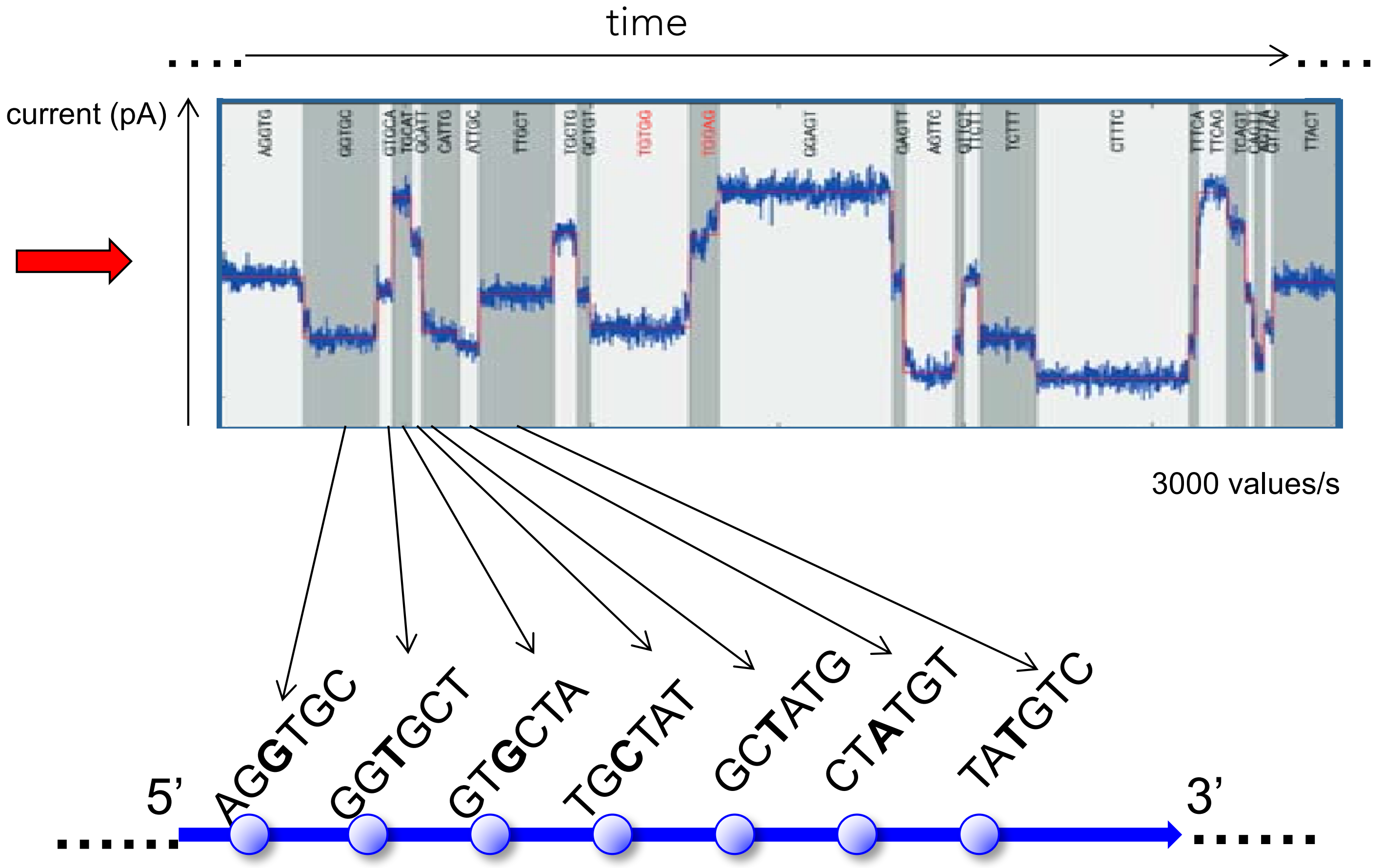
MinION : 512 pores



PromethION : 144000 pores (48 x 3000)

BASE CALLING



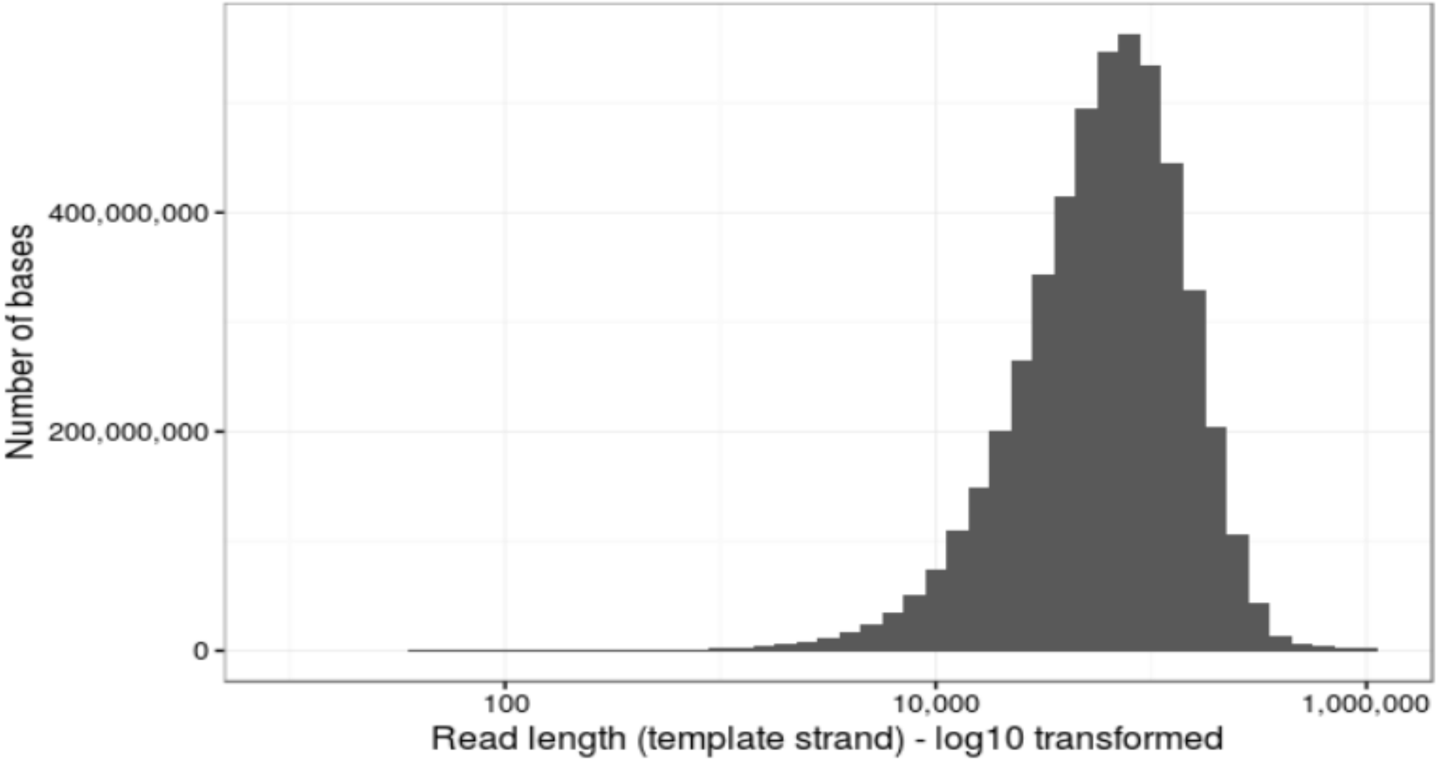


Basecalling : finding the optimal path of successive 6-mers

..... AGGTGCTATGTCT....

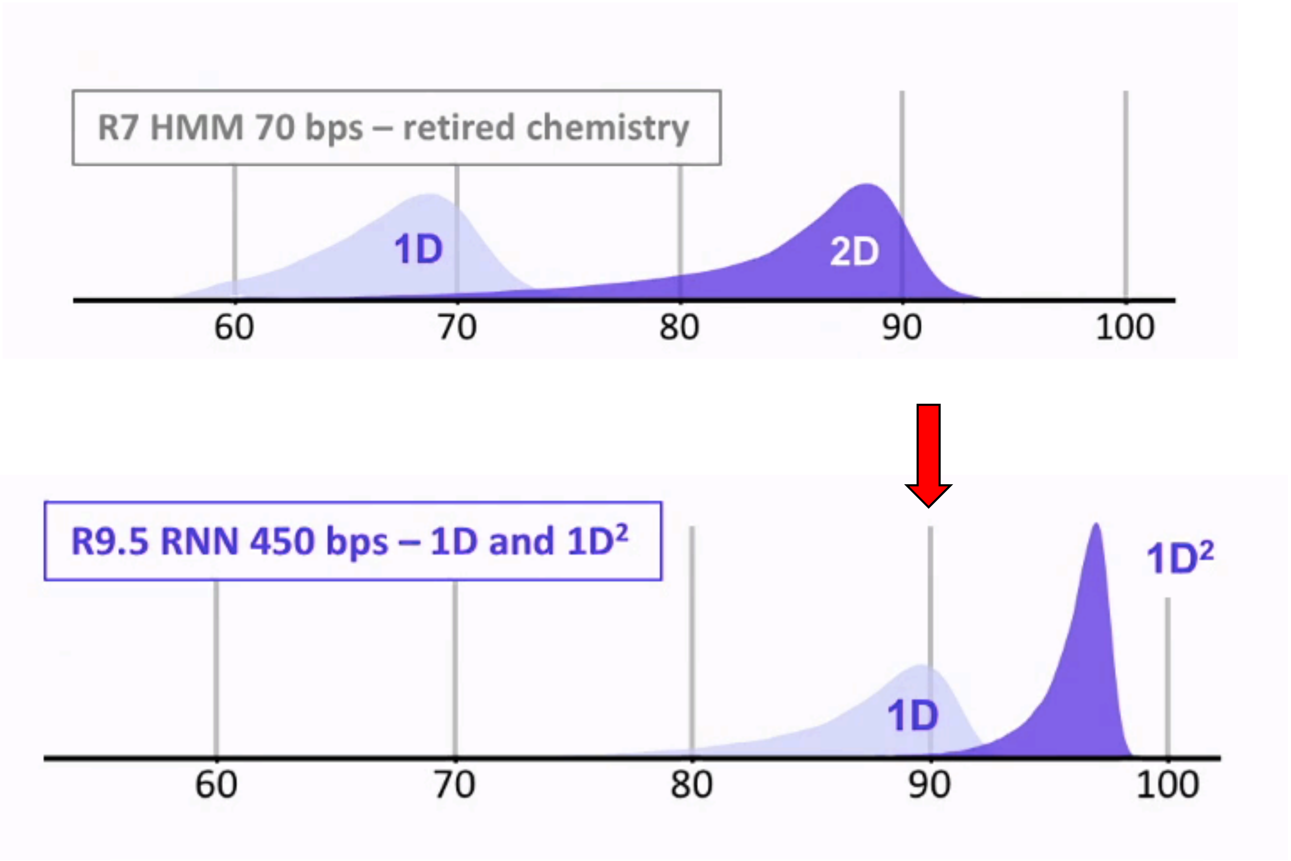
SIZE OF SEQUENCED DNA FRAGMENTS

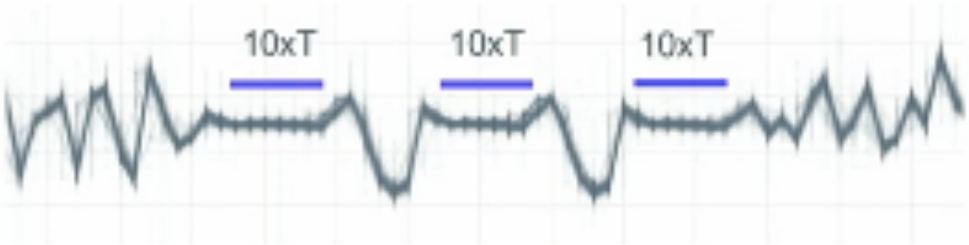
"Ultra long" reads
(lab.loman.net, March 2017)



Size of the longest read : 778 kb

READ QUALITY

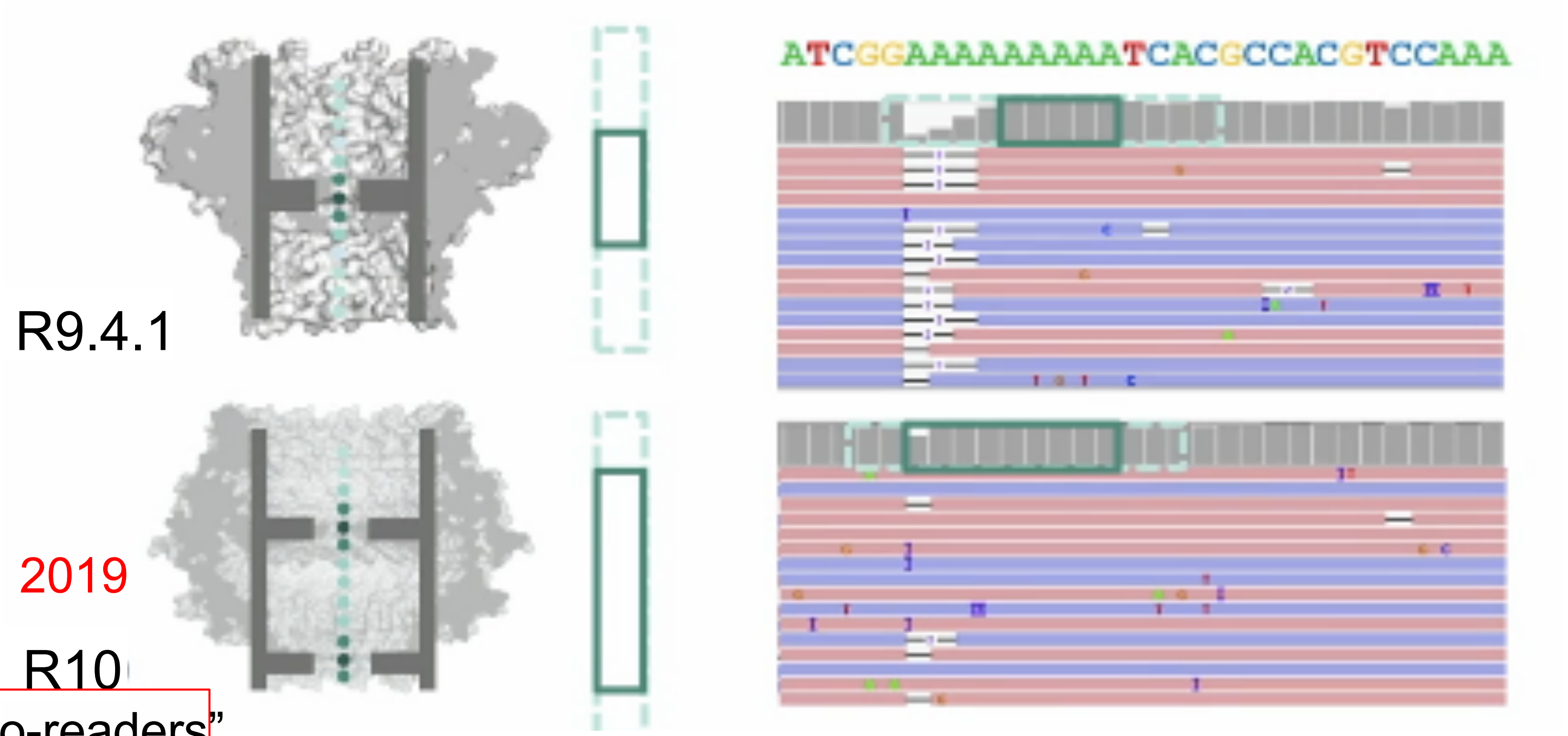




Homopolymers difficult to sequence

Recent improvements: ”Two readers” nanopore

”One-reader” pore has difficulty to read homopolymers

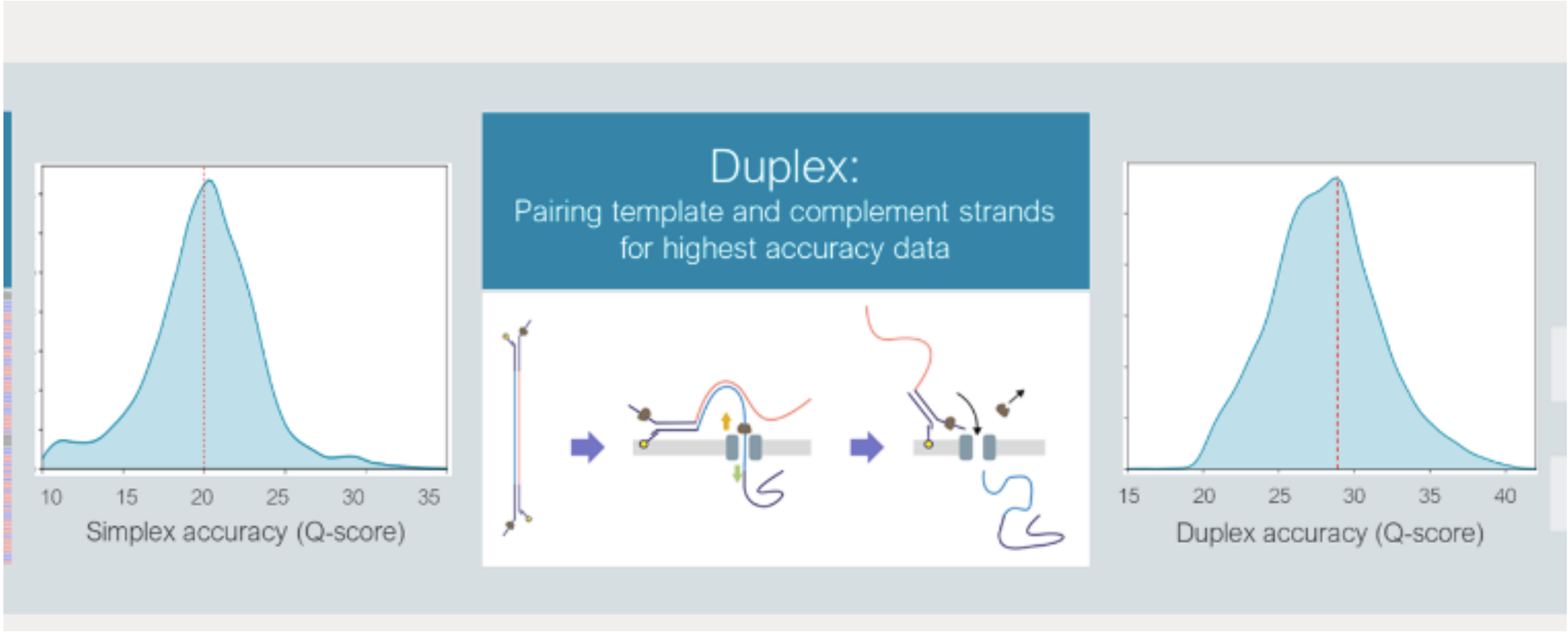


New pore accurately calls homopolymers

- A pore with a longer or multiple “readers” has more bases dominating the signal
- Longer homopolymers are “seen” by the pore and can be decoded with high accuracy

Recent improvements: "Q20+ reads

The Q20+ chemistry enables users to generate raw read sequencing data to an accuracy greater than Q20 (99%+)

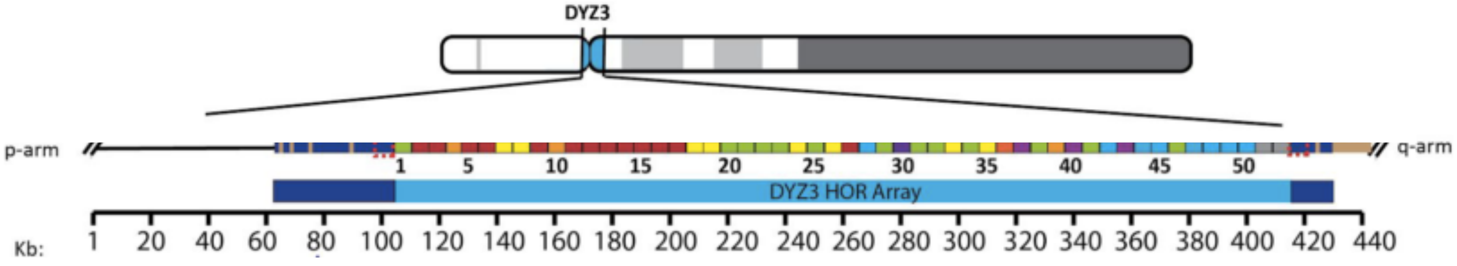


Duplex:

Pairing template and complement strands for highest accuracy data

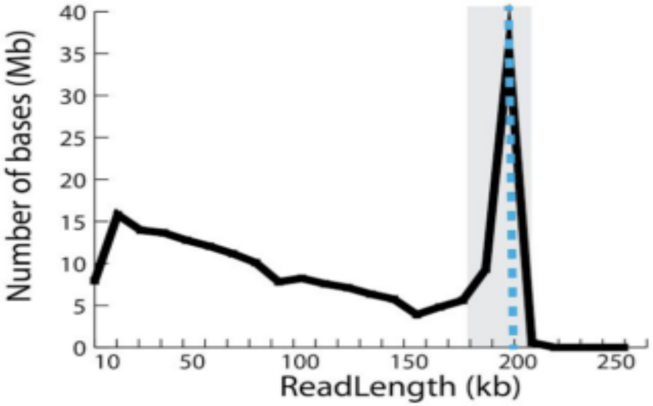
1 - GENOME ASSEMBLY WITH NANOPORE

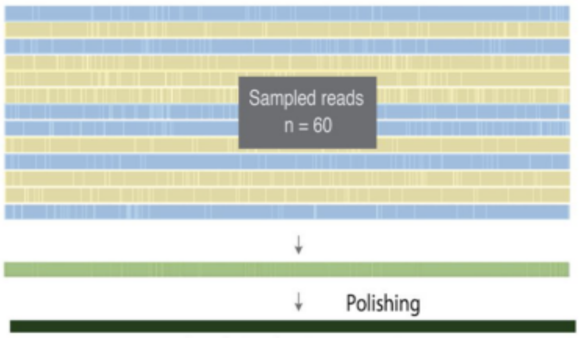
Linear Assembly of a Human Y Centromere using Nanopore Long Reads
Jain et al., *bioRxiv*, 2017



9 BACs
100 kb to 210 kb

210 kb





Final high quality consensus BAC sequence

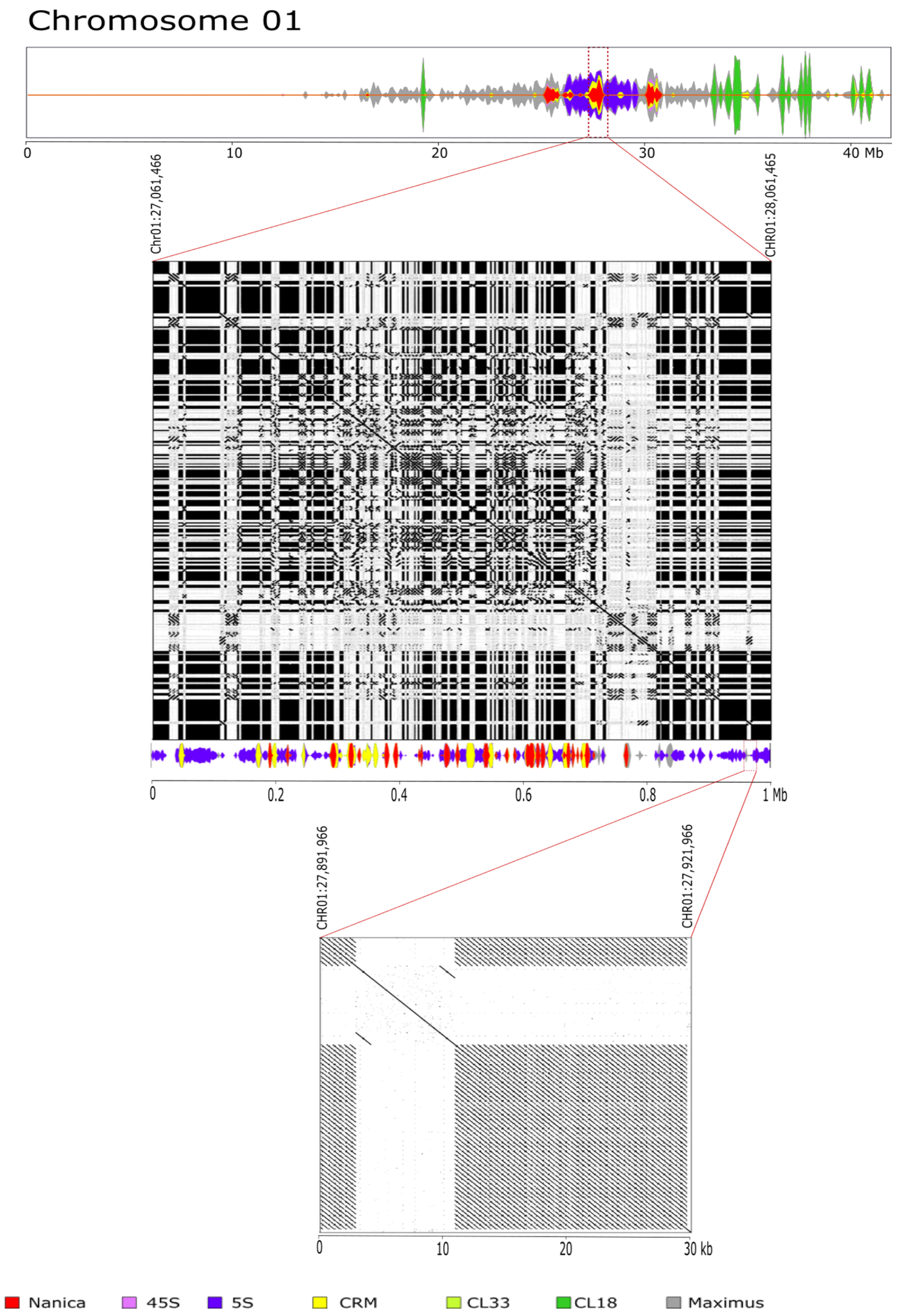
FIRST COMPLETE SEQUENCE OF A HUMAN CENTROMERE

2 - GENOME ASSEMBLY WITH NANOPORE

Telomere-to-telomere gapless chromosomes of banana using nanopore sequencing
Belser et al. *Communications Biology* Sept 2021

- haploid genome :
  - ~500 Mbp,
  - 11 chromosomes:
- 3 samples of reads:
  - 177X of all reads
  - 30X of the longest reads
  - 30X of the **Filtlong** highest-score reads
- assembler: NECAT11,
- 124 contigs polished with:
  - Racon (nanopore reads)
  - Medaka (nanopore reads)
  - Hapo-G (Illumina reads) : incorporates phasing information (Aury & Istace, NAR Apr. 2021)
- Bionano:
  - validate order and orient the contigs:
  - all contigs but 1 in accordance with optical maps
- ➡ 5 chromosomes reconstructed telomere to telomere
- reveal centromeres, clusters of paralogous genes
- Ex. : in previous versions : 130 5S rDNA genes
- New version : 7696 rDNA genes

Fine structure of repeated elements



3 - GENOME ASSEMBLY WITH NANOPORE

Long-read and chromosome-scale assembly of the hexaploid wheat genome
Aury et al., *bioRxiv*, Aug 2021

- First hexaploid wheat genome based on ONT long-reads
- hexaploid genome (15.5 Gb)
- sequencing began in 2005 : International Wheat Genome Sequencing Consortium (IWGSC)
- first sequence in 2018

- This work:
  - ✓ organize contigs in chromosomes using:

    ONT
    - 20 ONT flow cells (2 MinION and 18 PromethION)
    - produced 12M reads representing 1.1 Tb
    - base calling: (i) guppy 2.0 and then guppy 3.6 (High Accuracy)
    - coverage: 63x, N50: 24.6 kb
    - 3.1M reads > 50 kb, coverage: 14x

    Bionano Genomics (BNG) Saphyr
    - direct Label and Stain Chemistry (DLS) with the DLE-1 enzyme
    - total size: 14.9 Gb, N50: 37.5 Mb

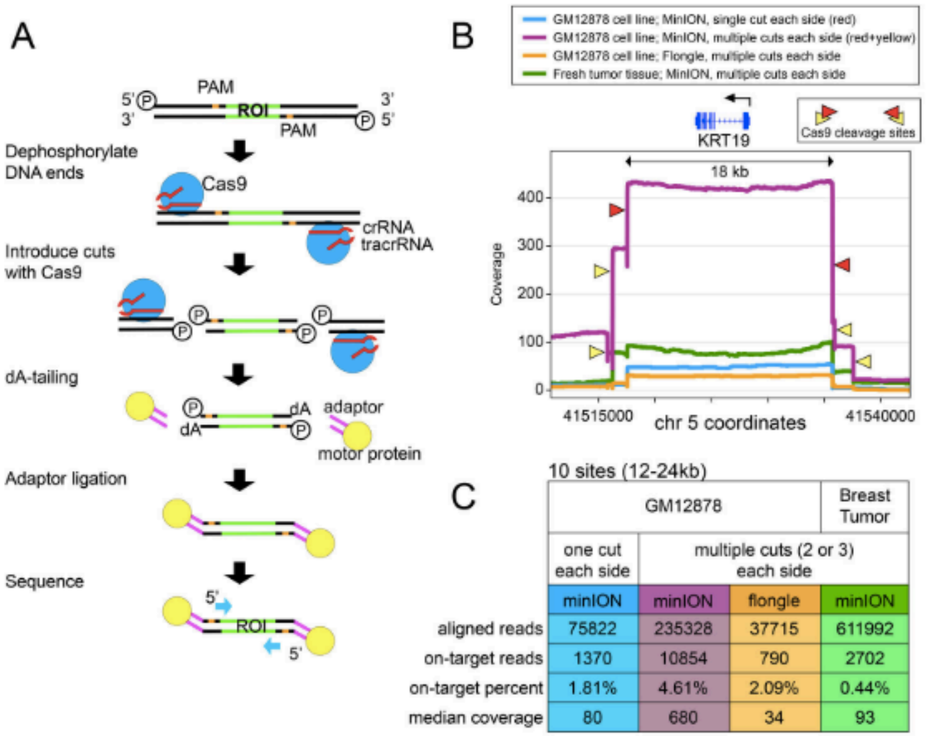
    Hi-C
    - 4 Hi-C libraries, Arima Genomics protocol
    - Illumina sequencing -> 537 Gb, 35x
    - We used a sample of 240 million read pairs (72 Gb, 5x) to build a Hi-C map

⬇

Most contiguous and complete chromosome-scale assembly of a bread wheat genome

GENOME SEQUENCING : TARGETED NANOPORE SEQUENCING

Targeted nanopore sequencing with Cas9-guided adaptor ligation
Gilpatrick et al. *Nature Biotechnology* April 2020



10 sites (12-24kb)

	GM12878			Breast Tumor
	one cut each side	multiple cuts (2 or 3) each side		
	minION	minION	flongle	minION
aligned reads	75822	235328	37715	611992
on-target reads	1370	10854	790	2702
on-target percent	1.81%	4.61%	2.09%	0.44%
median coverage	80	680	34	93

nCATS = nanopore Cas9-targeted sequencing : enrichment strategy using targeted cleavage of DNA to ligate adapters for nanopore

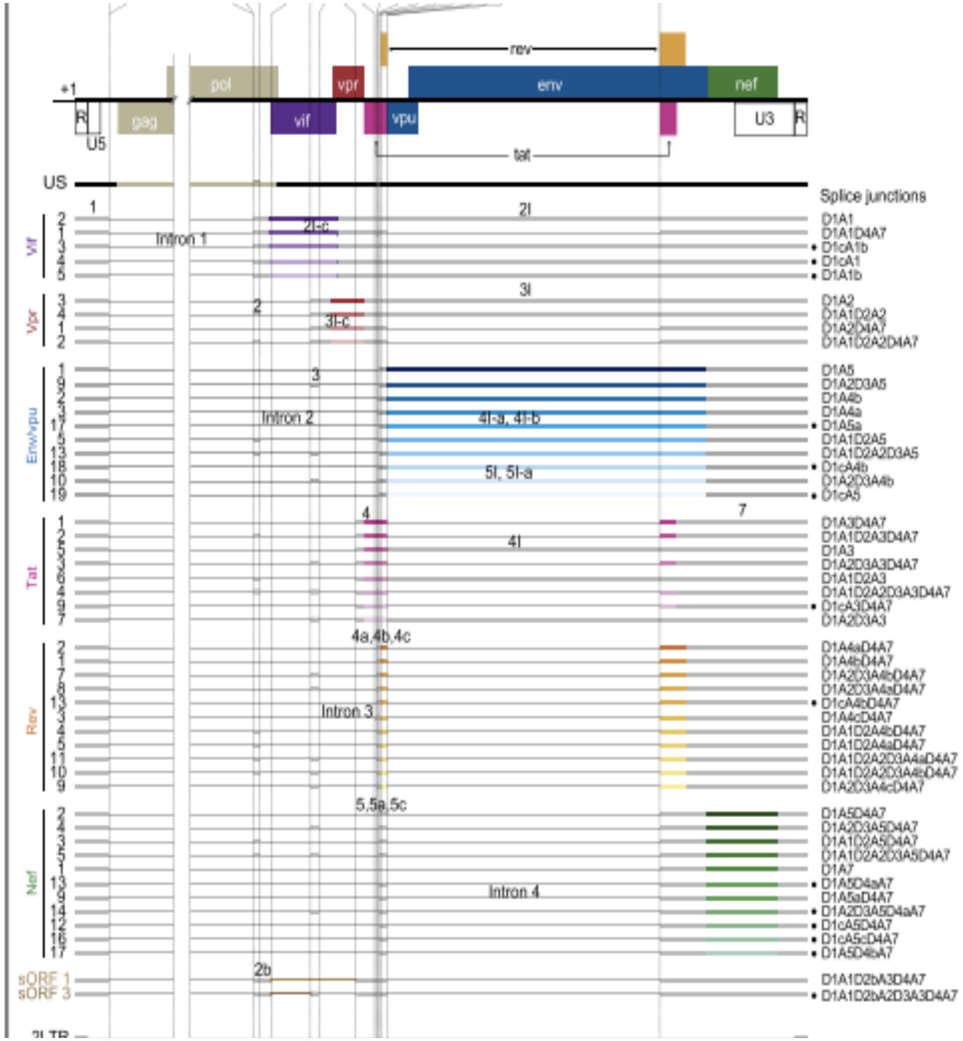
nCATS can simultaneously assess :

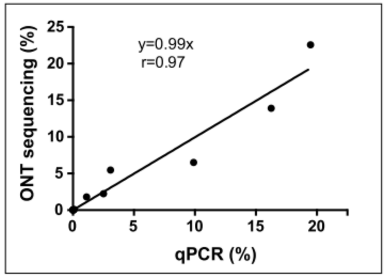
- haplotype-resolved single-nucleotide variants (SNVs)
- structural variations (SVs)
- CpG methylation…
- Best median sequencing coverage : 680 X
- nCATS uses only ~3 µg of genomic DNA + can target a large number of loci in a single reaction.

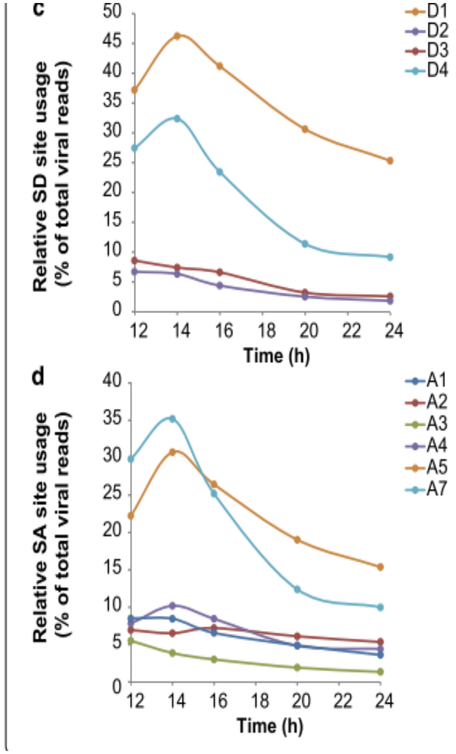
cDNA NANOPORE SEQUENCING

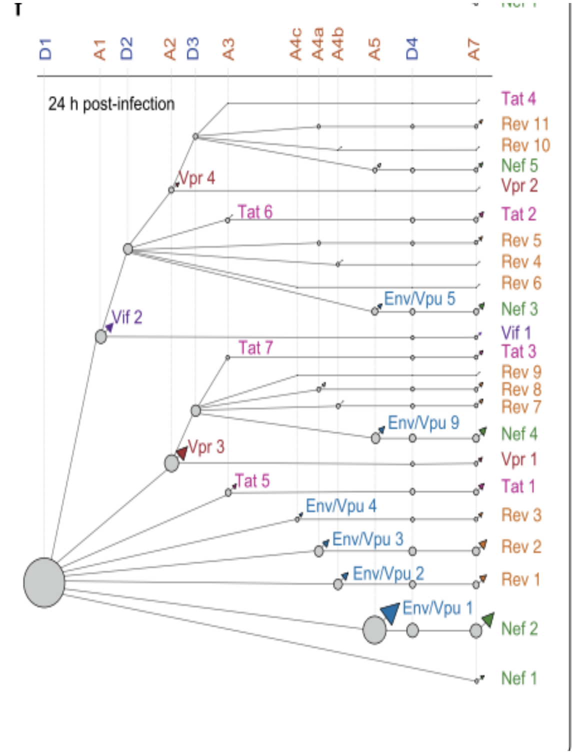
Dynamic nanopore long-read sequencing analysis of HIV-1 splicing events during the early steps of infection
Quang et al. *Retrovorology* 2020

- 53 viral RNA isoforms, including 14 new ones
- Relative levels highly correlated with qPCR
- First dynamic picture of the cascade of events occurring between 12 and 24 h of viral infection
- -> importance of non-coding exons in viral RNA transcriptome regulation







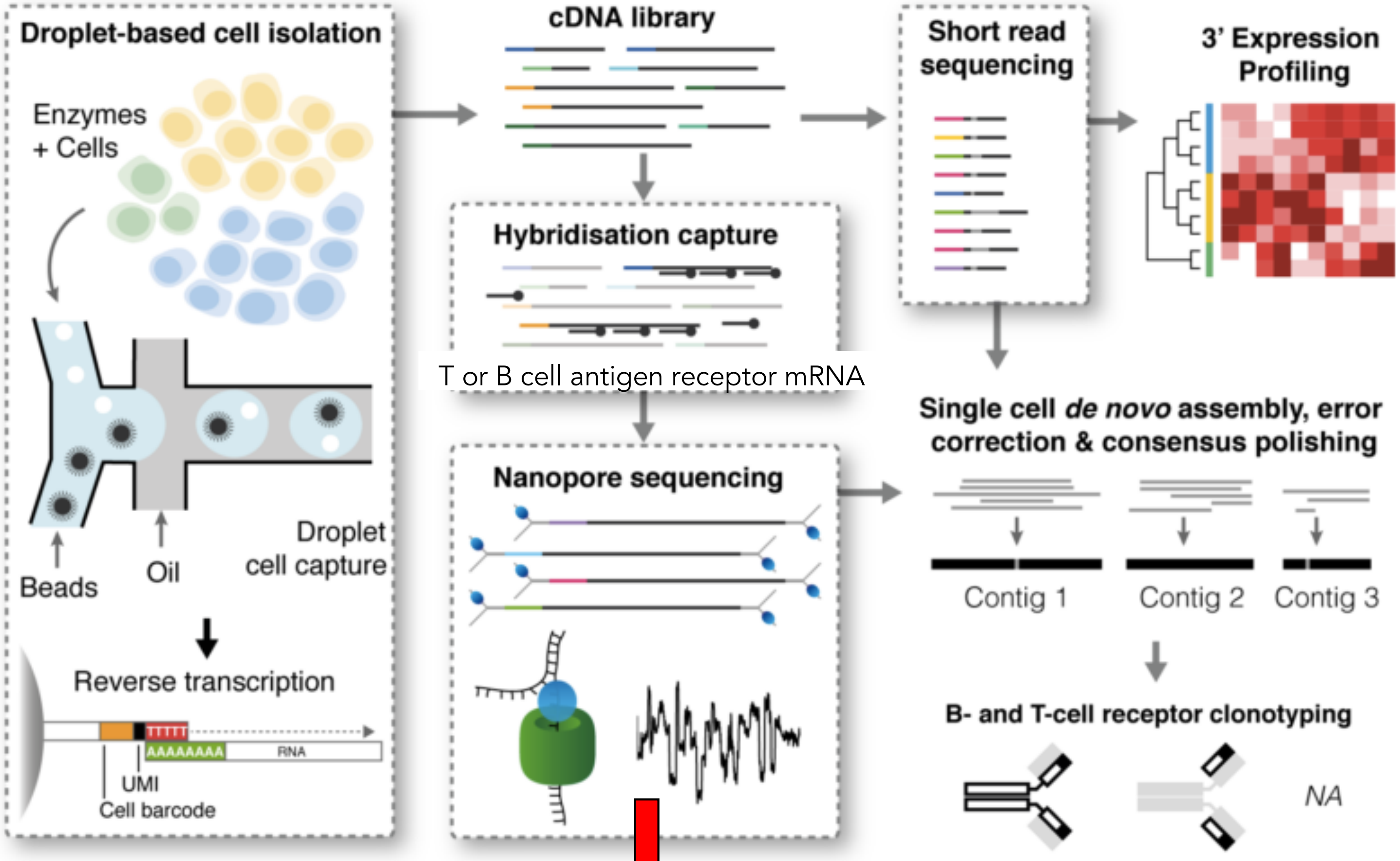


NANOPORE and SINGLE CELL cDNA SEQUENCING

High-throughput targeted long-read single cell sequencing reveals the clonal and transcriptional landscape of lymphocytes
Singh et al., *bioRxiv*, 2018

RAGE-seq (Repertoire And Gene Expression sequencing) : combines targeted long-read sequencing with short-read transcriptome of barcoded single cell libraries

10x Chromium
Single Cell 3'
(V2 chemistry)



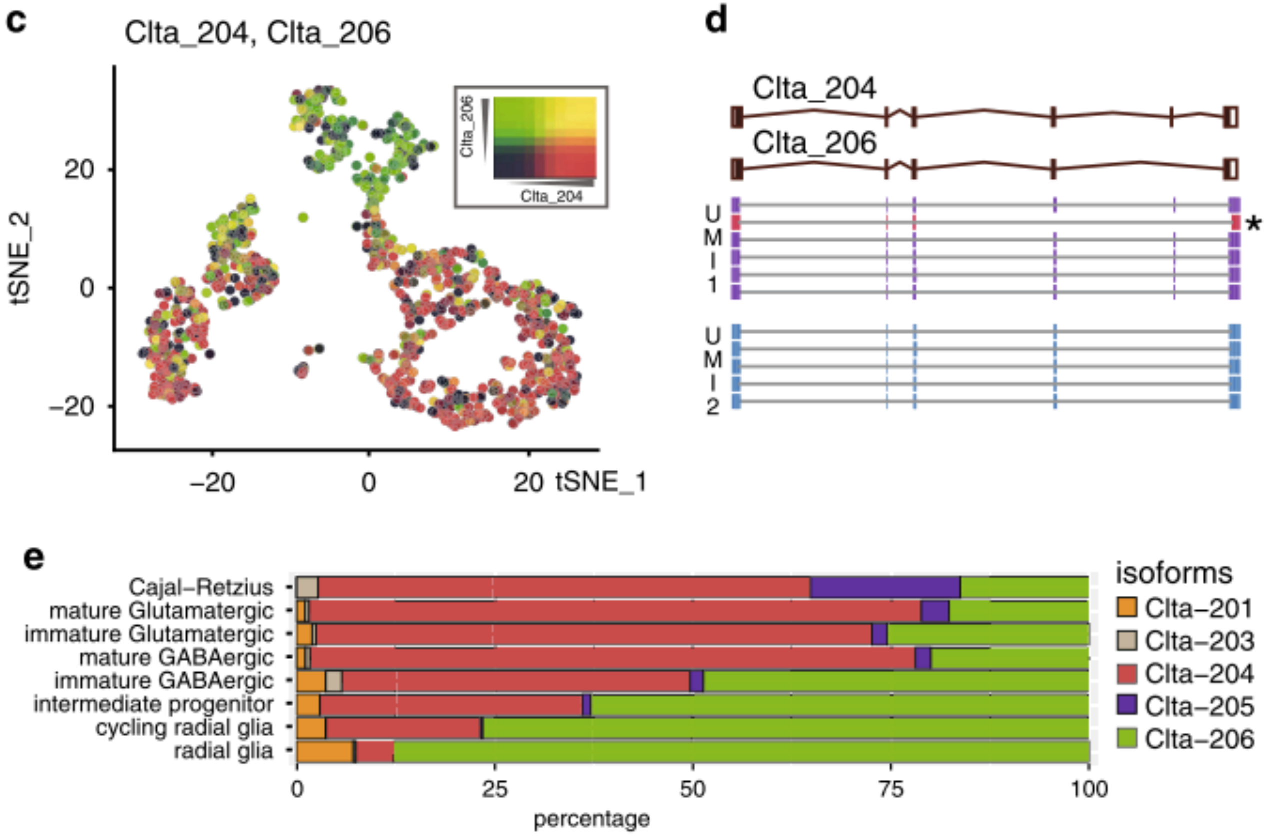
T or B cell antigen receptor mRNA

Tracking of somatic mutation, alternate splicing and clonal evolution of T and B lymphocytes
BUT
Does not correct for PCR biases

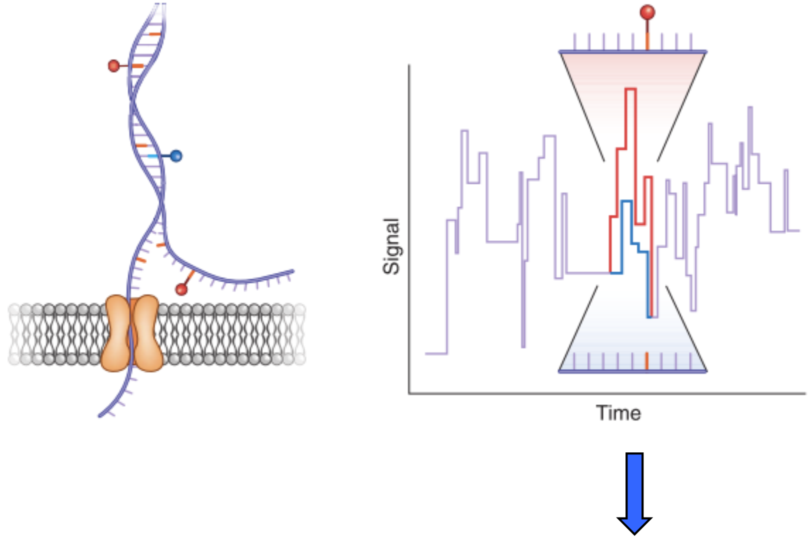
NANOPORE and SINGLE CELL cDNA SEQUENCING

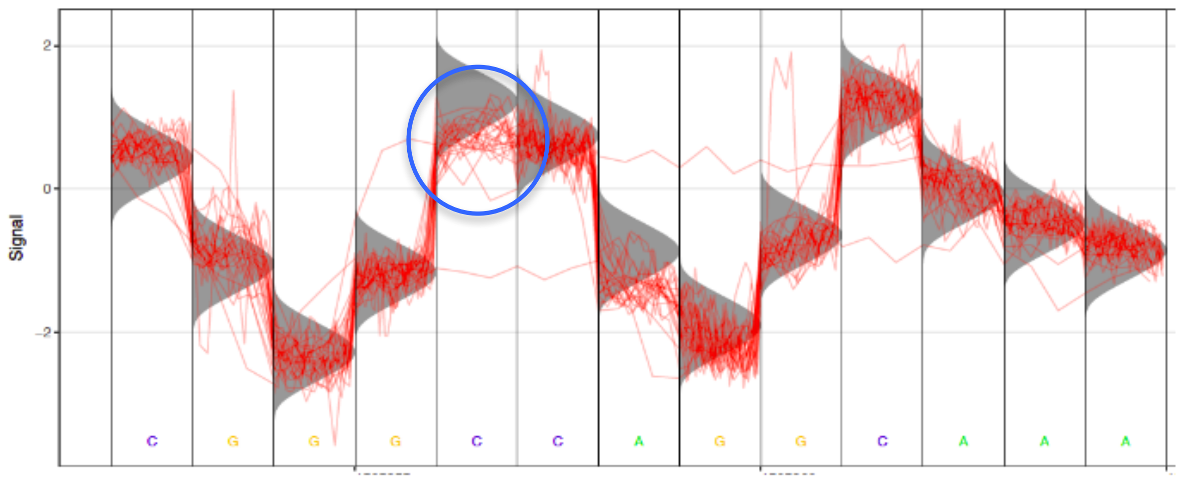
High-throughput targeted long-read single cell sequencing reveals the clonal and transcriptional landscape of lymphocytes
Lebrigand et al., *Nature Communications*, 2020

ScNaUmi-seq : Single-cell Nanopore sequencing with UMIs (10x Genomics)
- High accuracy cellBC and UMI assignment
- Analysis of splicing and sequence variation at the single-cell level



DETECTION OF MODIFIED DNA BASES



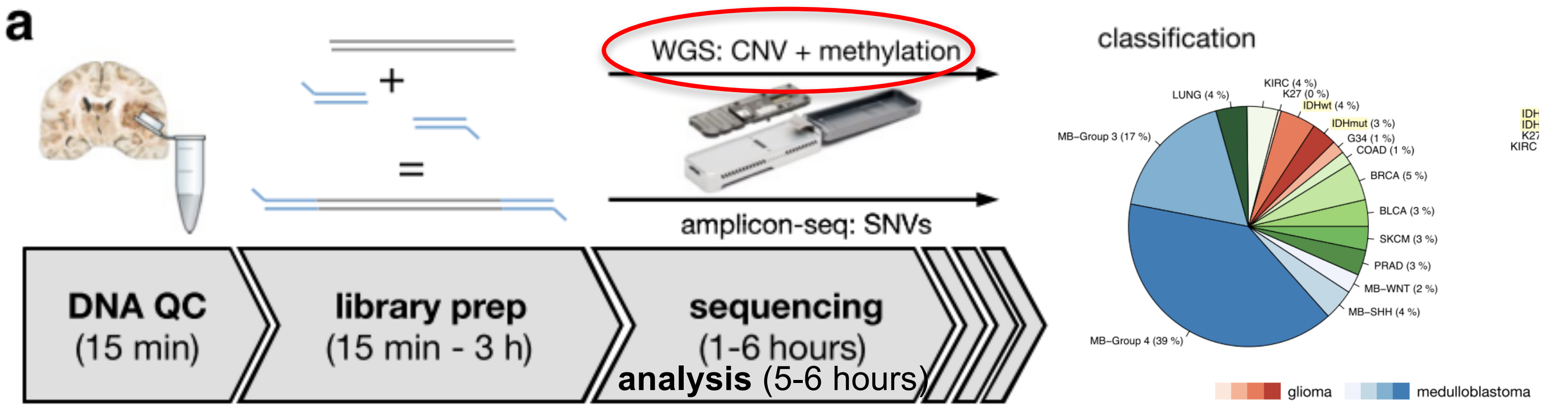


— Electric signal

▶ Canonical base distribution

DETECTION OF MODIFIED DNA BASES : 5mCpG in CANCER GENOMES

Same-day genomic and epigenomic diagnosis of brain tumors (gliomas, medulloblastomas) with nanopore sequencing
Euskirchen et al., *Acta Neuropathol.* (2017)



Same-day detection of :

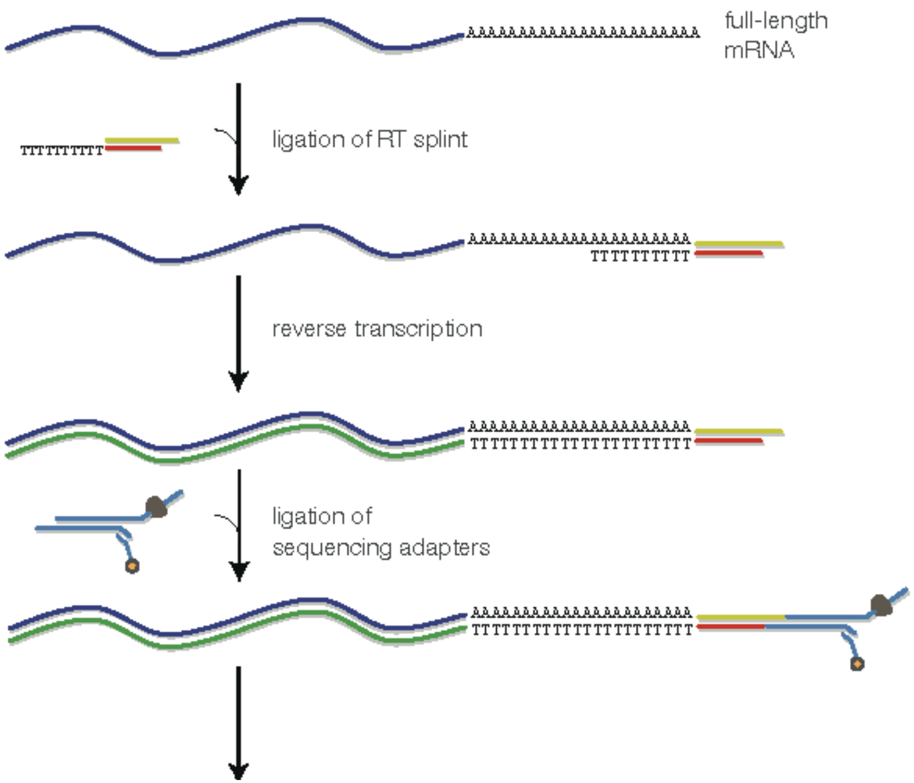
- structural variants
- point mutations
- CpG methylation profiling

Single device with negligible capital cost :

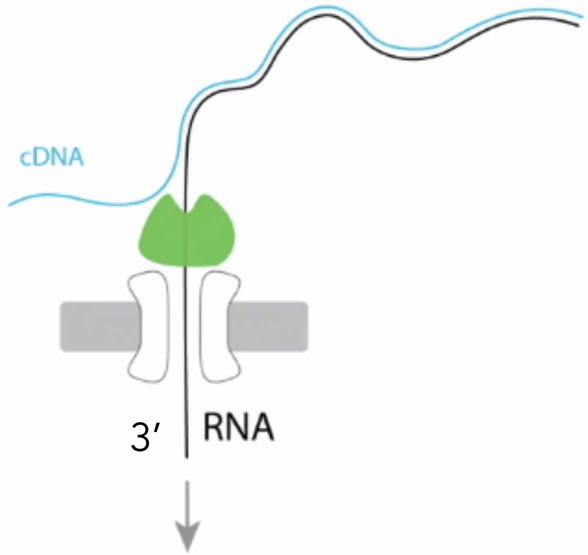
- outperforms hybridization-based and current sequencing technologies
- makes precision medicine possible for every cancer patient

DIRECT RNA SEQUENCING

Library preparation



sequencing

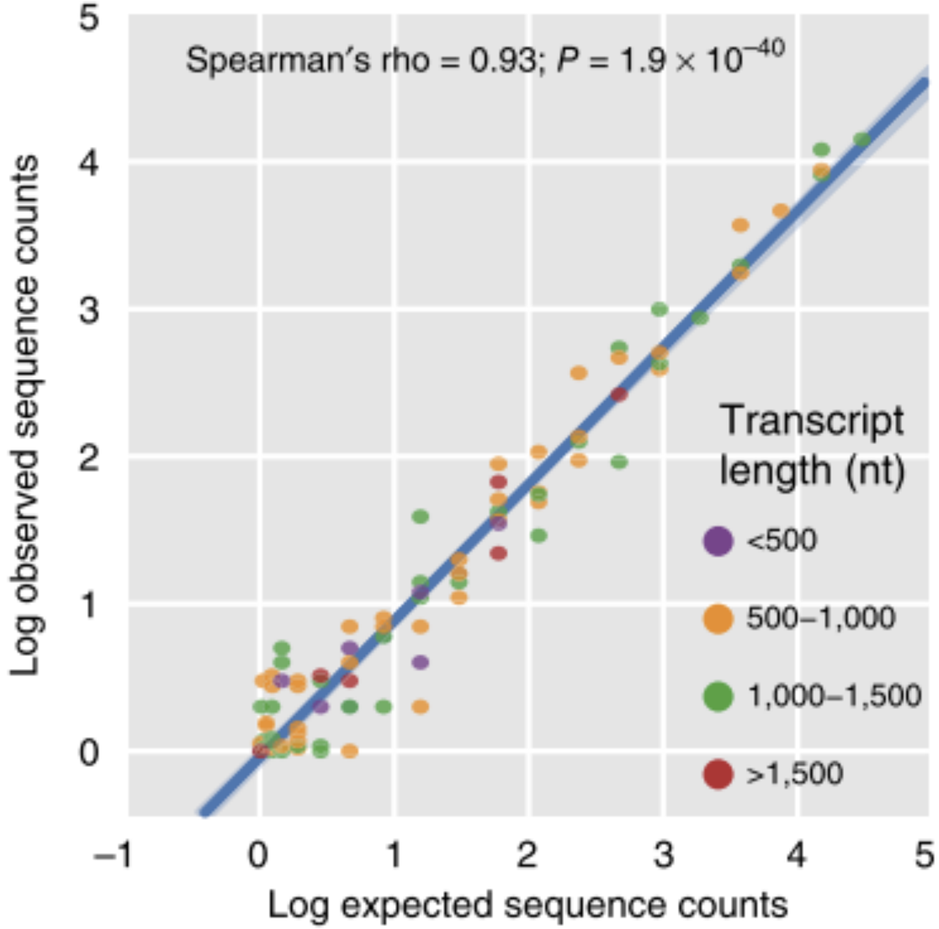


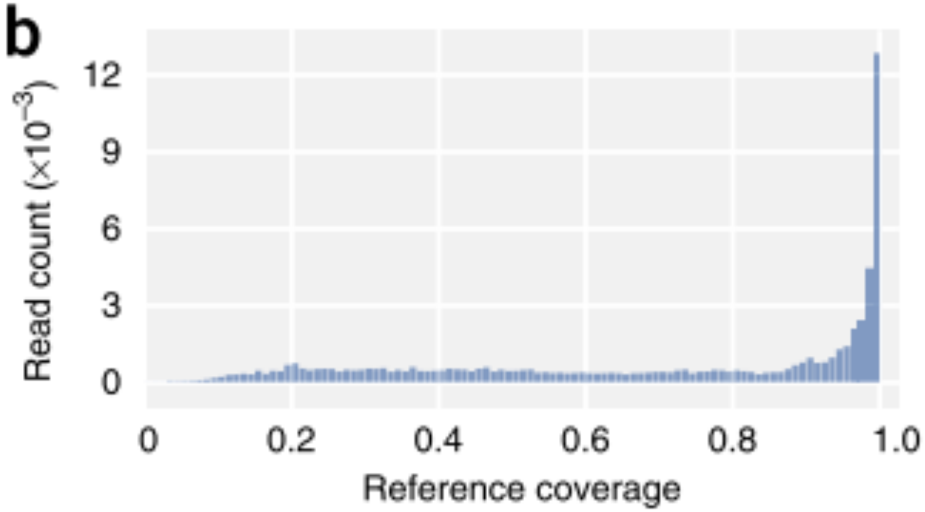
RNA directly sequenced in nanopore

- No PCR bias
- Quantitative

DIRECT RNA SEQUENCING : CONTROLS

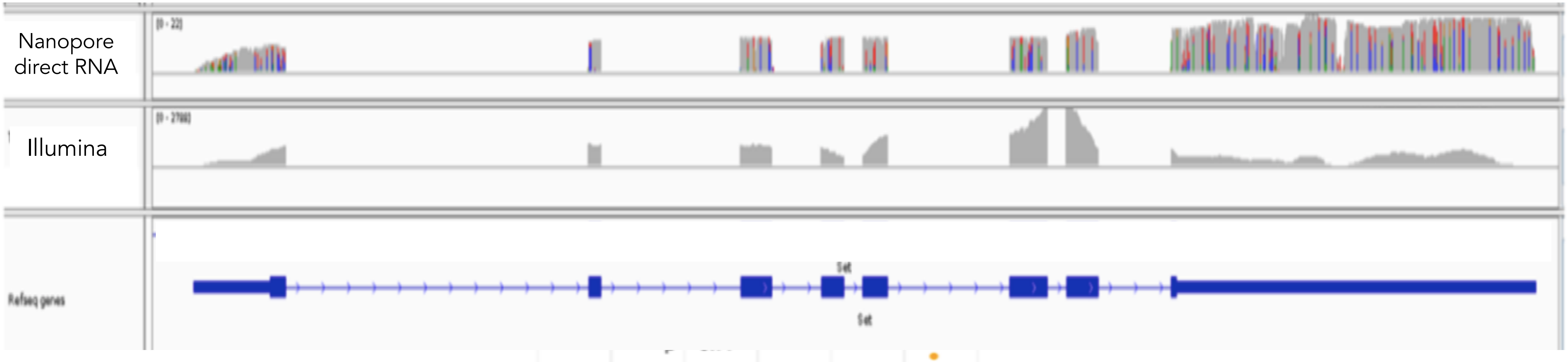
RNA spike in

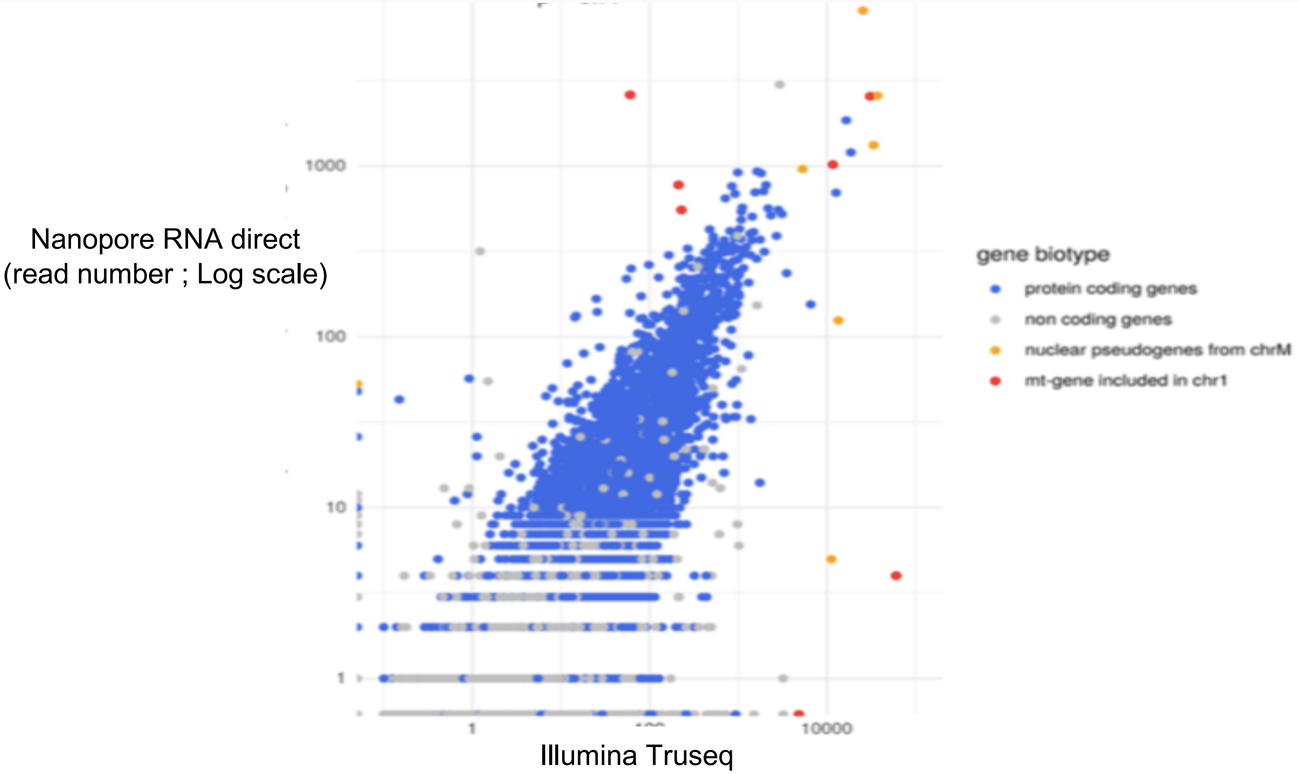




Garalde et al. *Nat. Methods* 2018

DIRECT RNA SEQUENCING vs ILLUMINA

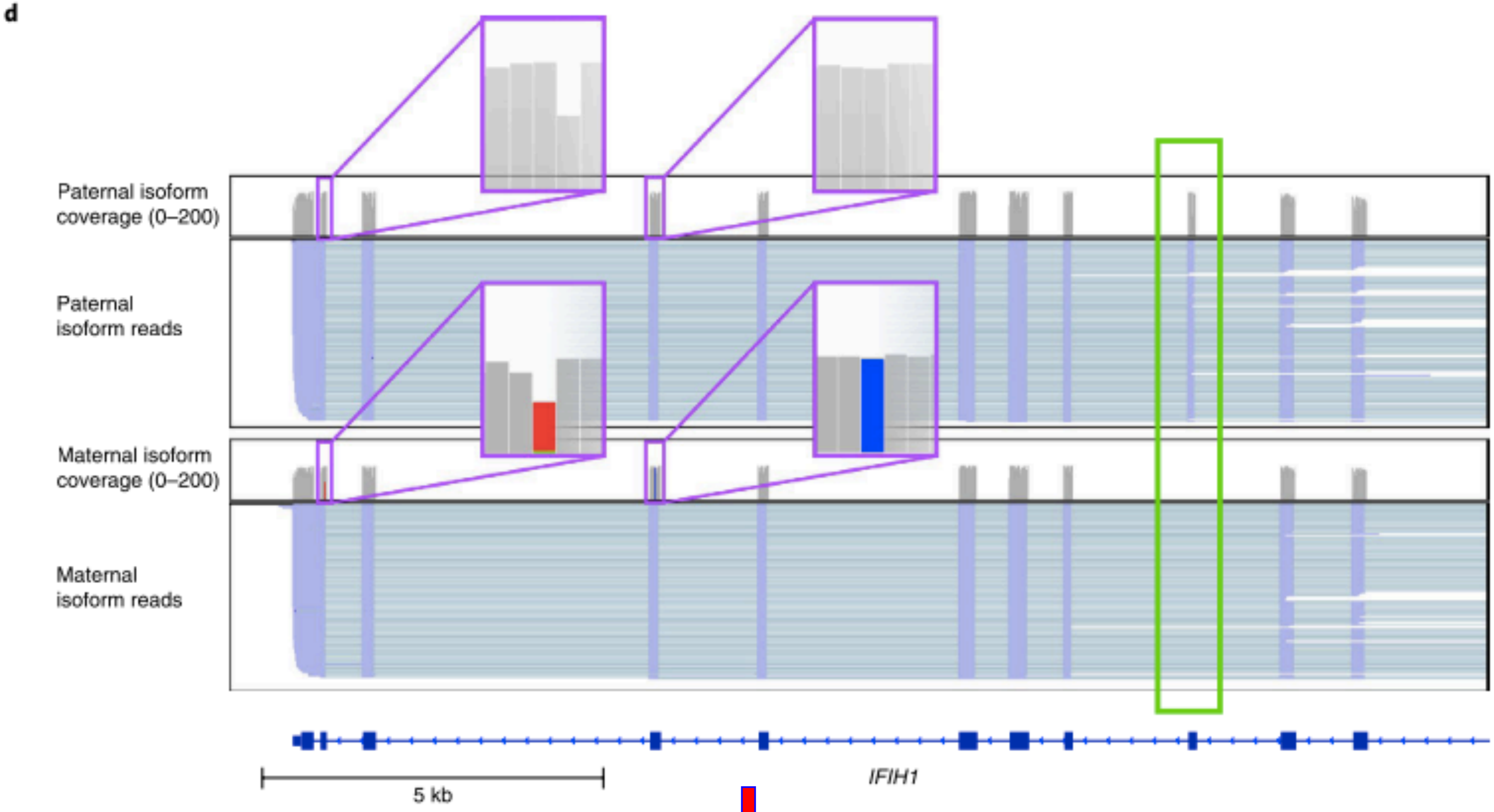




Sessegolo et al. *Sci. Reports* 2019

DIRECT RNA SEQUENCING: TRANSCRIPT HAPLOTYPE

Nanopore native RNA sequencing of a human transcriptome
Workman et al. *Nat. Methods* (2019)



34 genes with discordant allele specificity in two isoforms

DIRECT RNA SEQUENCING: DETECTION OF MODIFIED RNA

RNA modifications (> 150) play important roles in regulating RNA fate :

- RNA folding and structure
- base pairing
- recruitment of RNA-binding proteins
- can be dynamic and reversible

In mRNAs (translation, stability, splicing..)

- *6mA* most abundant and better characterized
- *pseudoU*

Also found in ncRNAs

- microRNAs (miRNAs)
- long non-coding RNAs (lncRNAs)
- circular RNAs (circRNAs)

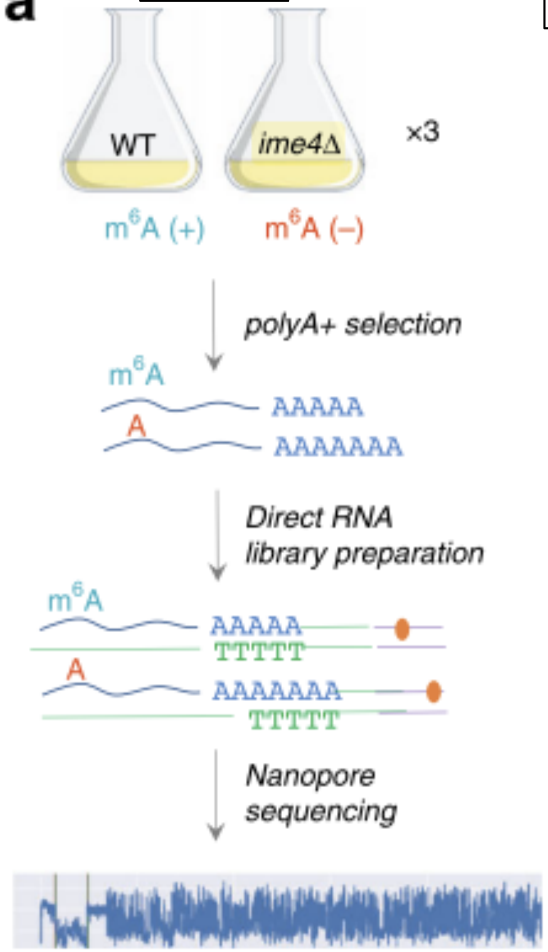
Viral RNAs contain high levels of modifications (modulate virus cycle)

- HIV RNA rich in :
    - *6mA*
    - *5mC*
    - *2'O-methyl*

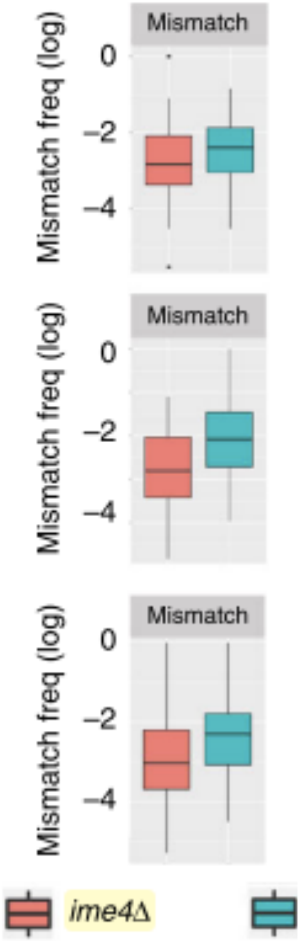
DIRECT RNA SEQUENCING: DETECTION OF m6A

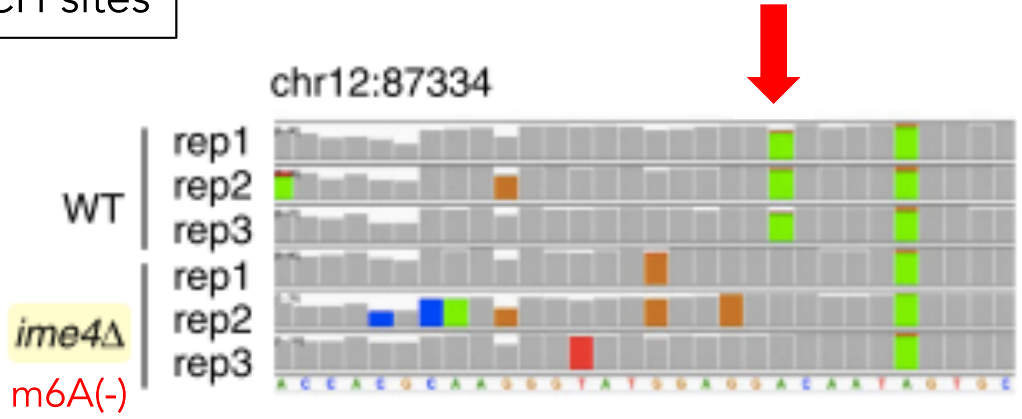
Accurate detection of m6A RNA modifications in native RNA sequences
Liu et al. *Nat. Comm.* 2019

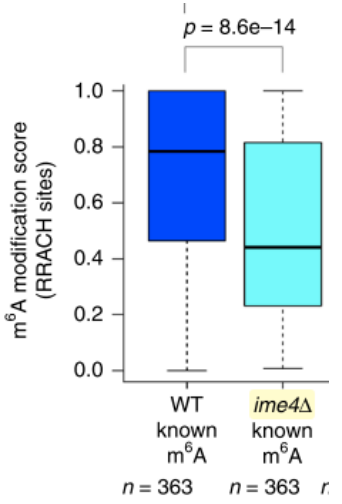
yeast

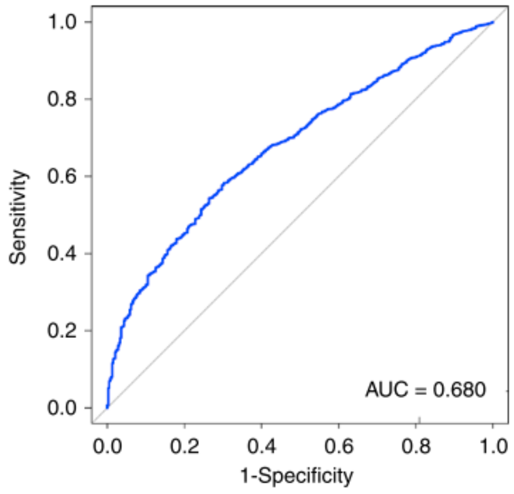


m6A-modified RRACH sites









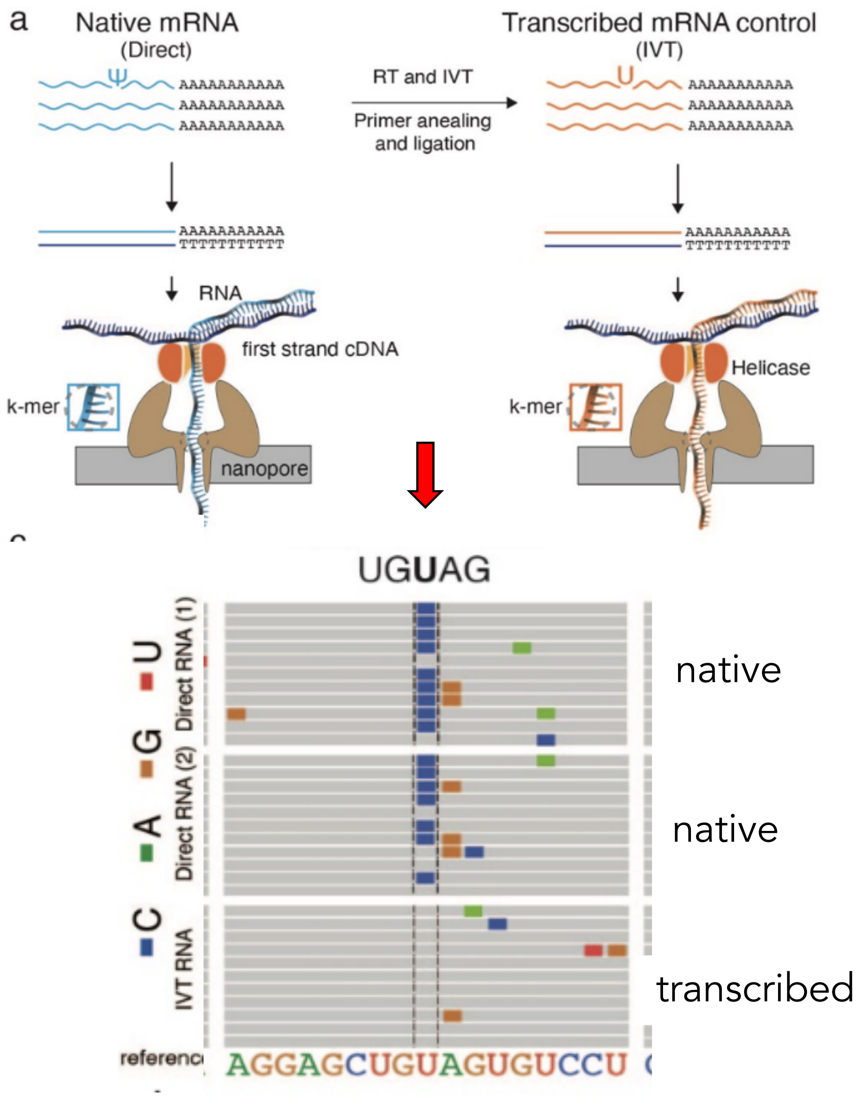
DIRECT RNA SEQUENCING : DETECTION OF pseudo-U

Detection of pseudouridine modifications and type I/II hypermodifications in human mRNAs using direct long-read sequencing.
Tavakoli et al. *bioRxiv* Nov. 2021

Detection of pseudo-U sites

- U-to-C base-calling errors occur at pseudouridines
- benchmarked against sites previously identified

- Pipeline for direct identification, quantification, and detection of pseudouridine modifications and
- Controls :
  - 1000mer synthetic RNA with single pseudouridine in center position
  - U-to-C occurs at the site of pseudouridylation
- Discovery of human mRNAs with up to 7 unique sites of pseudouridine modification



native

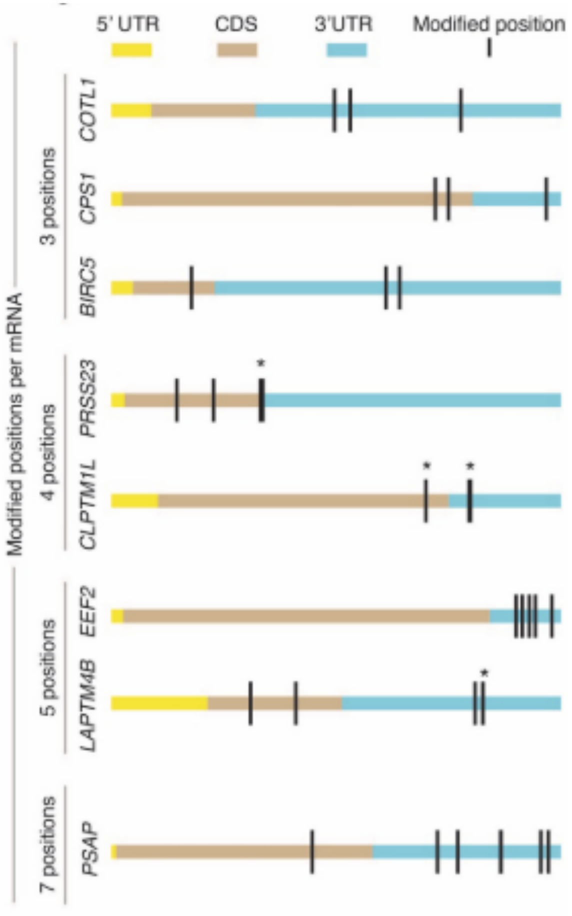
native

transcribed

DIRECT RNA SEQUENCING : DETECTION OF pseudo U

Detection of pseudouridine modifications and type I/II hypermodifications in human mRNAs using direct long-read sequencing.
Tavakoli et al. *bioRxiv* Nov. 2021

Pseudouridinylated human mRNAs :
104 at 2 positions
27 at 3 positions
4 at 4 positions
5 at 5 positions
1 at 6 positions
1 at 7 positions

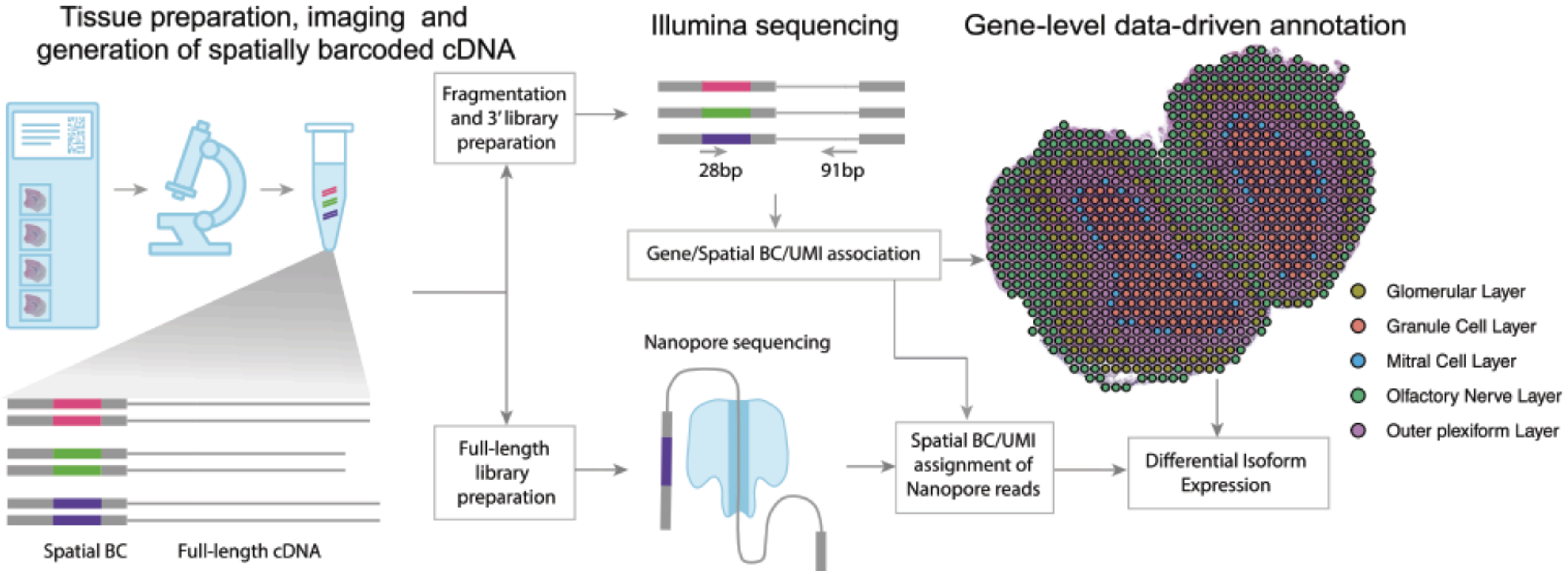


Recent advances : Nanopore and 10x Genomics Visium

The spatial landscape of gene expression isoforms in tissue sections
Lebrigand et al., *bioRxiv*, 2020

Spatial Isoform Transcriptomics (SiT) : Genome-wide approach to explore and discover in a tissue context :

- Isoform expression (bi-allelic expression)
- Sequence heterogeneity (SNP expression)



Tissue preparation, imaging and generation of spatially barcoded cDNA
Spatial BC
Full-length cDNA
Fragmentation and 3' library preparation
Full-length library preparation
Illumina sequencing
28bp
91bp
Gene/Spatial BC/UMI association
Nanopore sequencing
Spatial BC/UMI assignment of Nanopore reads
Gene-level data-driven annotation
Glomerular Layer
Granule Cell Layer
Mitral Cell Layer
Olfactory Nerve Layer
Outer plexiform Layer
Differential Isoform Expression

Genome assembly : Nanopore + PacBio:

Genome assembly : Nanopore + PacBio

- 2001: Celera Genomics and the International Human Genome Sequencing Consortium published their initial drafts of the human genome
- But, due to technological limitations, many other complex regions were left unfinished or incorrectly assembled for over 20 years
- ➡ *8% of the genome*
- T2T assembly : largest addition of new content to the human genome in the past 20 years

Main publications

1 - The structure, function, and evolution of a complete human chromosome 8.
Logsdon et al., *Nature*, May 2021

2 - The complete sequence of a human genome.
Nurk et al., *bioRxiv* May 2021

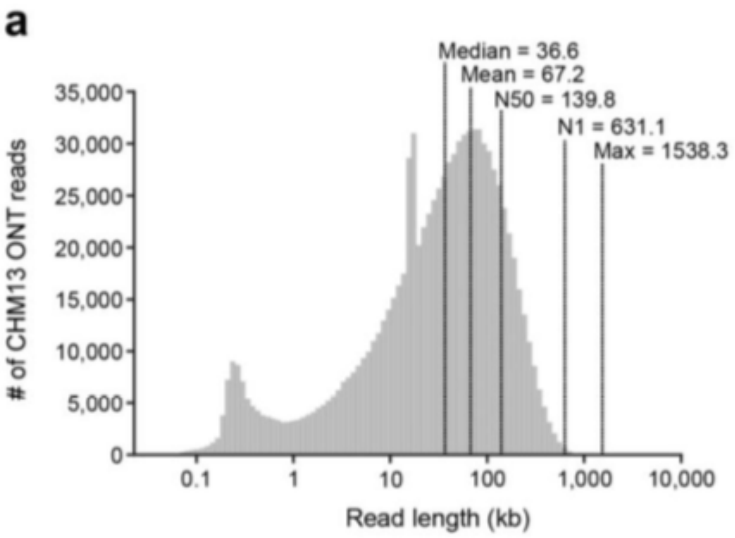
3 - Chasing perfection: validation and polishing strategies for telomere-to-telomere genome assemblies.
Cartney et al., *bioRxiv* July 2021

Genome assembly : Nanopore + PacBio

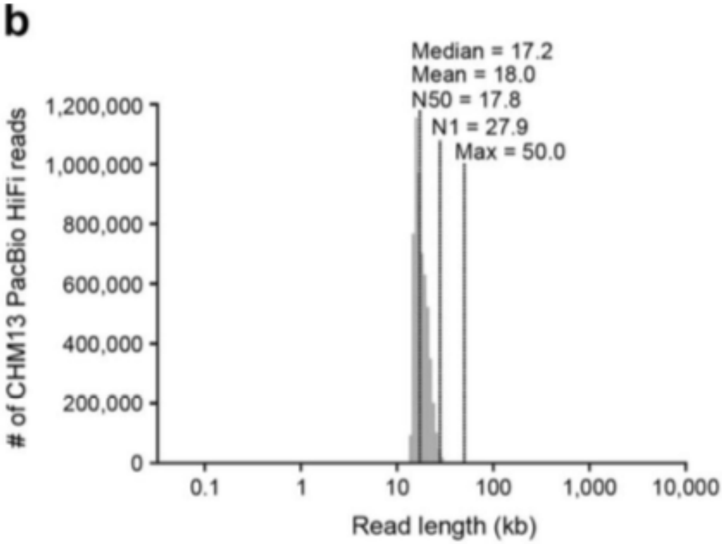
1 - The structure, function, and evolution of a complete human chromosome 8
Logsdon et al., *Nature*, May 2021

- Cell line : "complete hydatidiform mole" (CHM) derived from abnormal form of pregnancy
- Almost completely homozygous and therefore easier to assemble than heterozygous diploid genomes
- 20-fold sequence coverage of ONT ultra-long reads
- 32.4-fold coverage of PacBio HiFi

ONT ultra-long reads



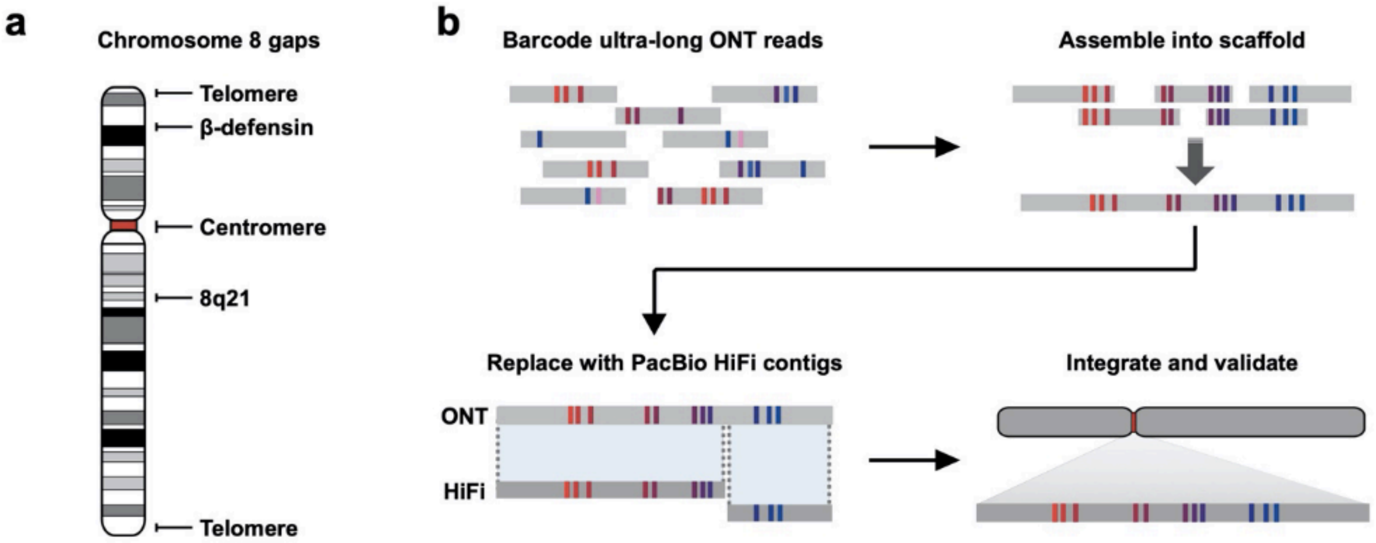
PacBio HiFi reads



Genome assembly : Nanopore + PacBio

1 - The structure, function, and evolution of a complete human chromosome 8
Logsdon et al., *Nature*, May 2021

- Barcoded Ultra-long Nanopore reads assembled into a scaffold
- Regions within the scaffold with high sequence identity with PacBio HiFi contigs are replaced, thereby improving the base accuracy to >99.99%.



- First complete linear assembly of a human autosomal chromosome.
- It resolves the sequence of five previously long-standing gaps :
  - 2.08 Mbp centromeric α-satellite array
  - 644 kbp defensin copy number polymorphism
  - 863 kbp variable number tandem repeat at chromosome 8q21.2 (neocentromere)
  - etc..

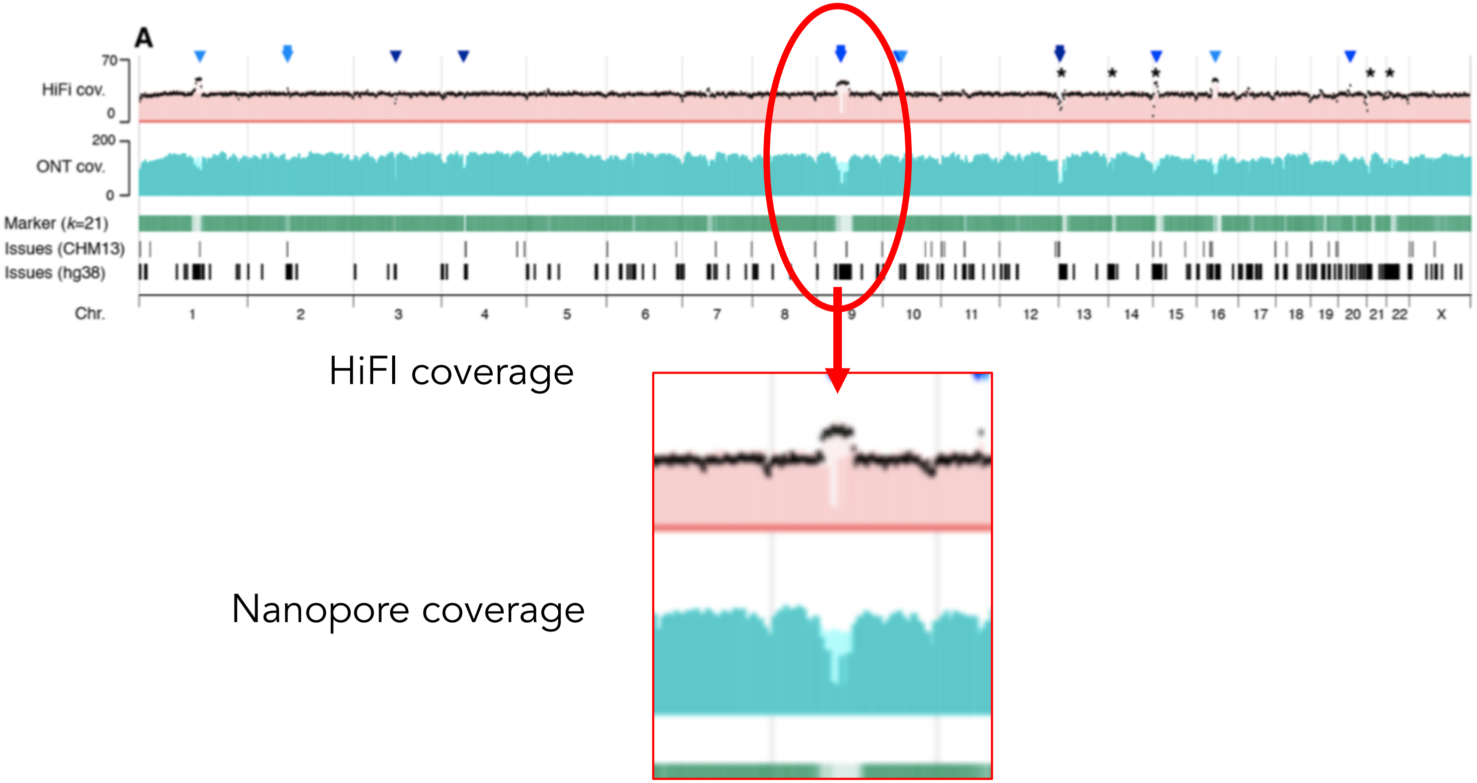
Genome assembly : Nanopore + PacBio

2 - The complete sequence of a human genome
Nurk et al. *bioRxiv* May 2021

SEQUENCING

Data were obtained with a “complete hydatidiform mole” (CHM13) cell line:

- 30× PacBio circular consensus sequencing (HiFi)
- 120× Oxford Nanopore ultra-long read sequencing (ONT)
- 100× Illumina PCR-Free sequencing
- 70× Illumina / Arima Genomics Hi-C (Hi-C)
- BioNano optical maps (*11*)
- Strand-seq



HiFI coverage

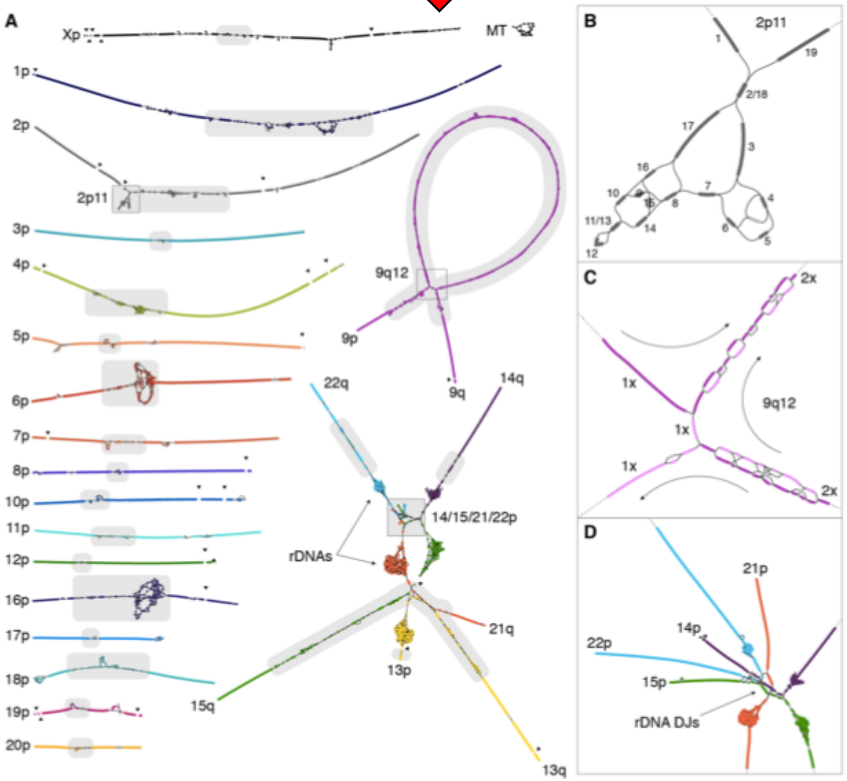
Nanopore coverage

Genome assembly : Nanopore + PacBio

2 - The complete sequence of a human genome
Nurk et al. *bioRxiv* May 2021

ASSEMBLY

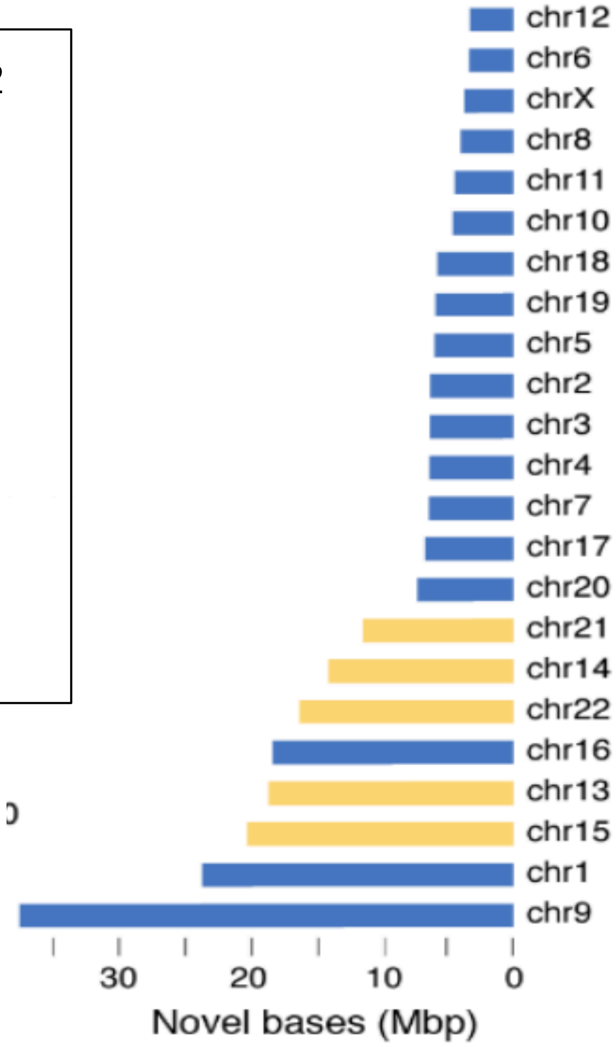
- HiFi-based string graph constructed using a purpose-built method that combines components from
  - HiCanu
  - Miniasm
  - specialized graph processing



Genome assembly : Nanopore + PacBio

2 - The complete sequence of a human genome
Nurk et al. *bioRxiv* May 2021

- 8% of the genome completed by this T2T assembly :including all 22 autosomes plus Chromosome X :
  - ➢ Corrects numerous errors
  - ➢ Introduces 200 million bp of novel sequence
  - ➢ Identifies 2,226 paralogous gene copies, 115 of predicted as protein coding
  - ➢ all centromeric regions
  - ➢ entire short arms (p-arms) of 5 acrocentric chromosomes : 13, 14, 15, 21, 22



Genome assembly : Nanopore + PacBio

3 - Chasing perfection: validation and polishing strategies for telomere-to-telomere genome assemblies
Cartney et al. *bioRxiv* July 2021

Recent Telomere-to-Telomere (T2T) human genome assembly

- this assembly has evidence of small errors and structural misassemblies
- polishing strategy :
  - ✓ Make corrections in large repeats without over-correction
  - ✓ Ultimately fixing 51% of errors and improving the assembly QV to 73.9
  - ✓ show sequencing biases in PacBio HiFi and ONT reads that cause errors that can be correcte

- **1,457 corrections :**
  - ✓ **replacing a total of 12,234,603 bp with 10,152,653 bp**
  - ✓ **ultimately leading to the first complete human genome ever assembled**

Summary

PacBio

- Maximum read length : 200 kb
- CCS sequencing (HiFI reads) :
  - Very low error rate, better genome assembly
  - Sequencing of cDNAs (resolution of alternative splicing)
  - Detection of modified DNA (6mA >> 5mC)
  - cDNA :
    - RNA-seq
    - Efficient for splicing isoforms detection

Nanopore

- Very light sequencing system
- Very long reads : maximum length >> 200 kb
- Detection of modified DNA (5mC >> 6mA)
- Direct sequencing of RNA :
  - Direct RNA sequencing :
    - RNA-seq
    - splicing isoforms detection
    - Detection of modified RNA (6mA, pseudo U)

Conclusion :
Whereas ultra-long nanopore sequencing excels at spanning long, identical repeats, HiFi sequencing excels at differentiating subtly diverged repeat copies or haplotypes
For large genomes, using these technologies simultaneously will likely improve the assembly

Remark

✓ Haplotype-resolved diverse human genomes and integrated analysis of structural variation
Ebert et al. *Science* April 2021

- 65 authors, 29 affiliations : 18 USA, 5 Germany, 2 Spain, 3 China, 1 UK

✓ The complete sequence of a human genome.
Nurk et al. *bioRxiv* May 2021

- 98 authors, 51 affiliations : 39 USA (68 authors), 5 Russia, 2 Germany, 2 UK, 1 Switzerland, 1 Croatia, 1 India

✓ Chasing perfection: validation and polishing strategies for telomere-to-telomere genome assemblies
Cartney et al. *bioRxiv* July 2021

- 20 authors, 14 affiliations : 10 USA, 1 Russia, 1 UK, 1 India, 1 Croatia