

SINGLE CELL TECHNOLOGIES

Single cell transcriptomics allows to study transcriptome heterogeneity, to investigate differences in transcript expression and gene regulation *in individual cells* :

- ❖ Differences in transcript abundance
- ❖ Alternative splicing and differential expression of isoforms

Most widely used device to study single-cell transcriptomics : *Chromium controller (10x Genomics)*

Several applications :

Single cell Gene expression

Measures gene activity on a cell-by-cell basis, characterize cell populations, cell types, ...

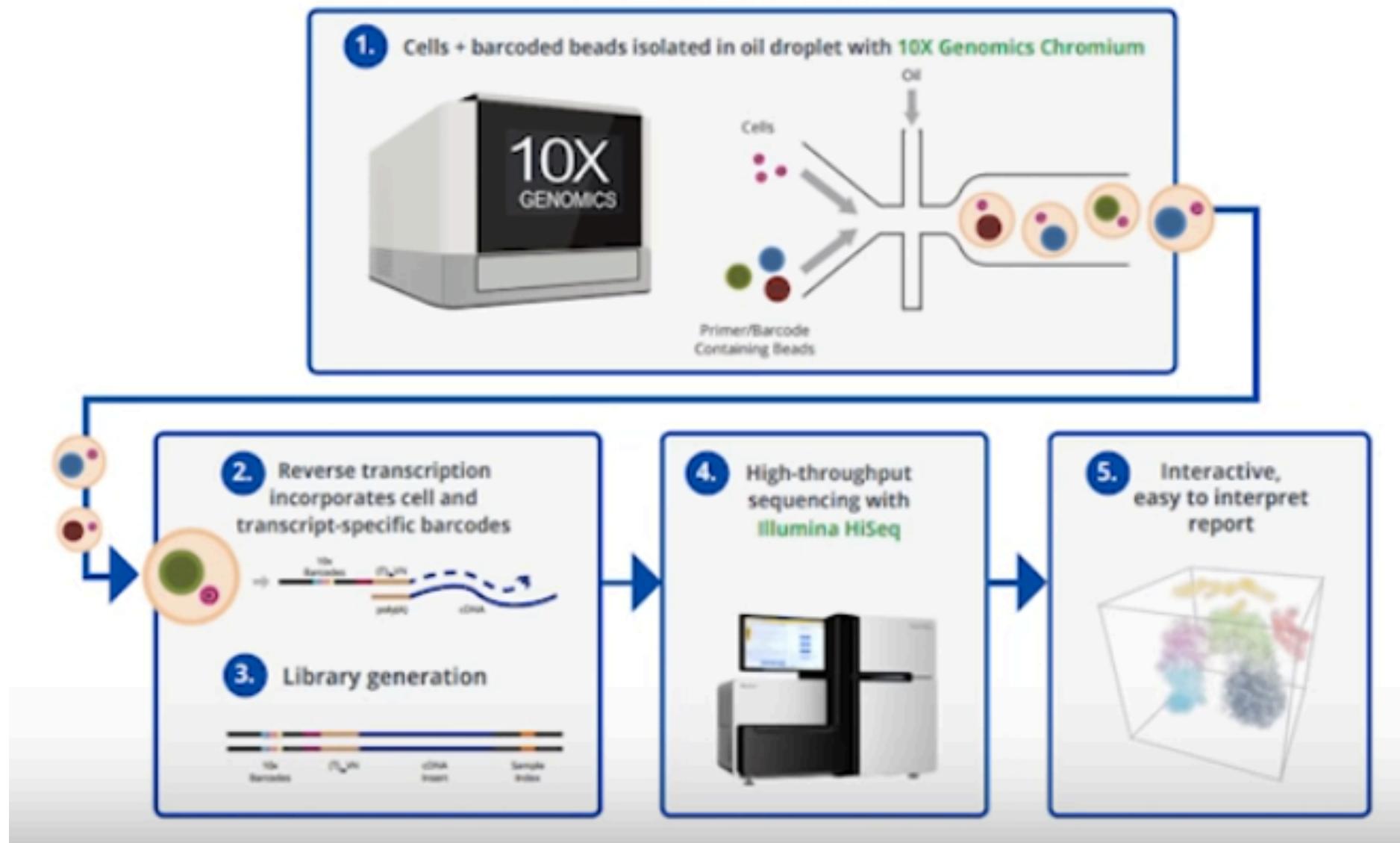
Linked read genomics

Performs diploid de novo assembly, phase haplotypes, genetic variations

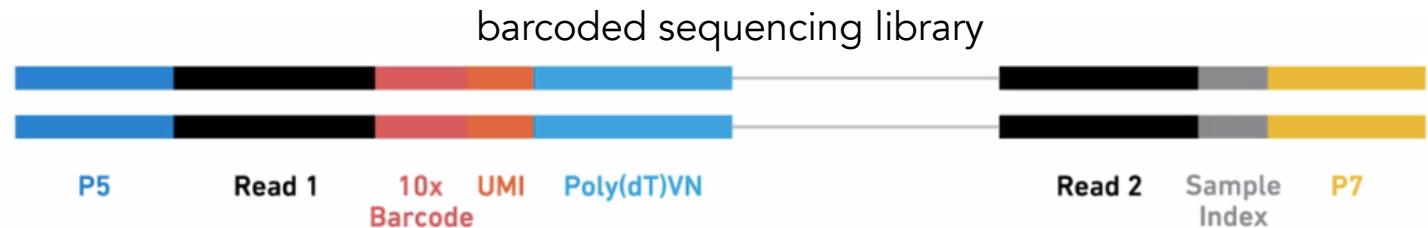
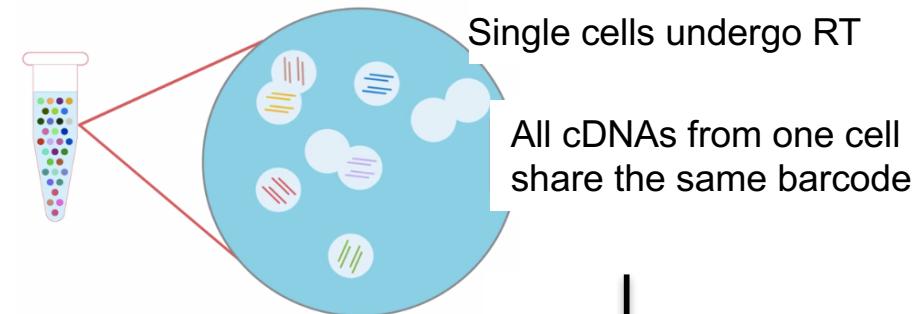
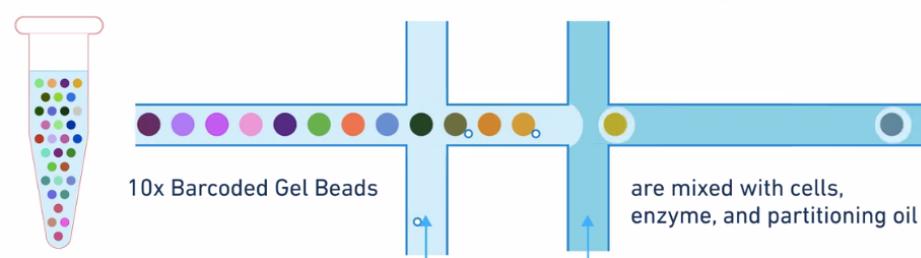
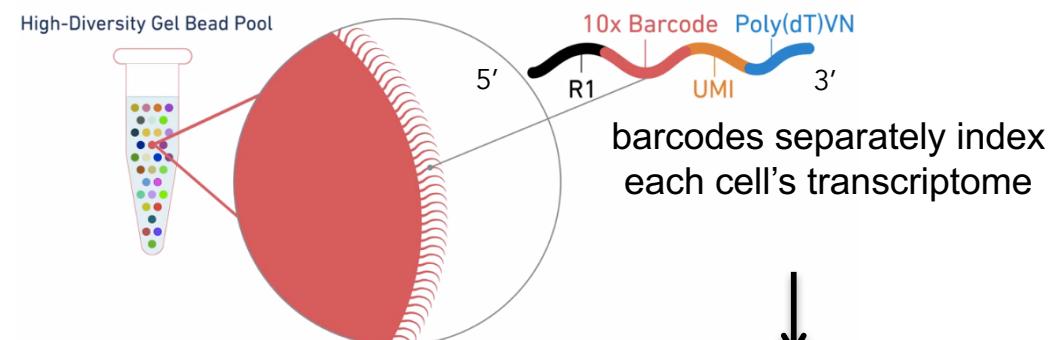
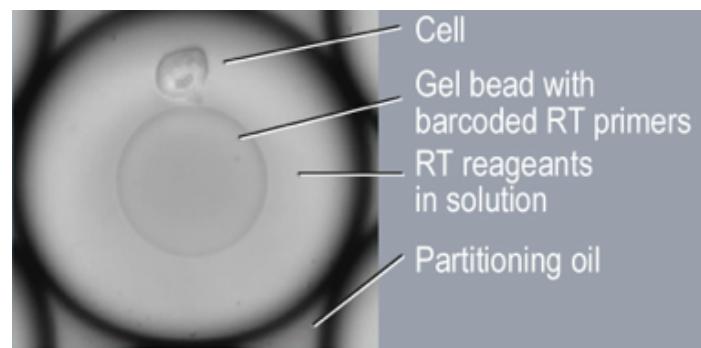
Single cell ATAC

Measures epigenetics by detecting open chromatin regions

Project Workflow

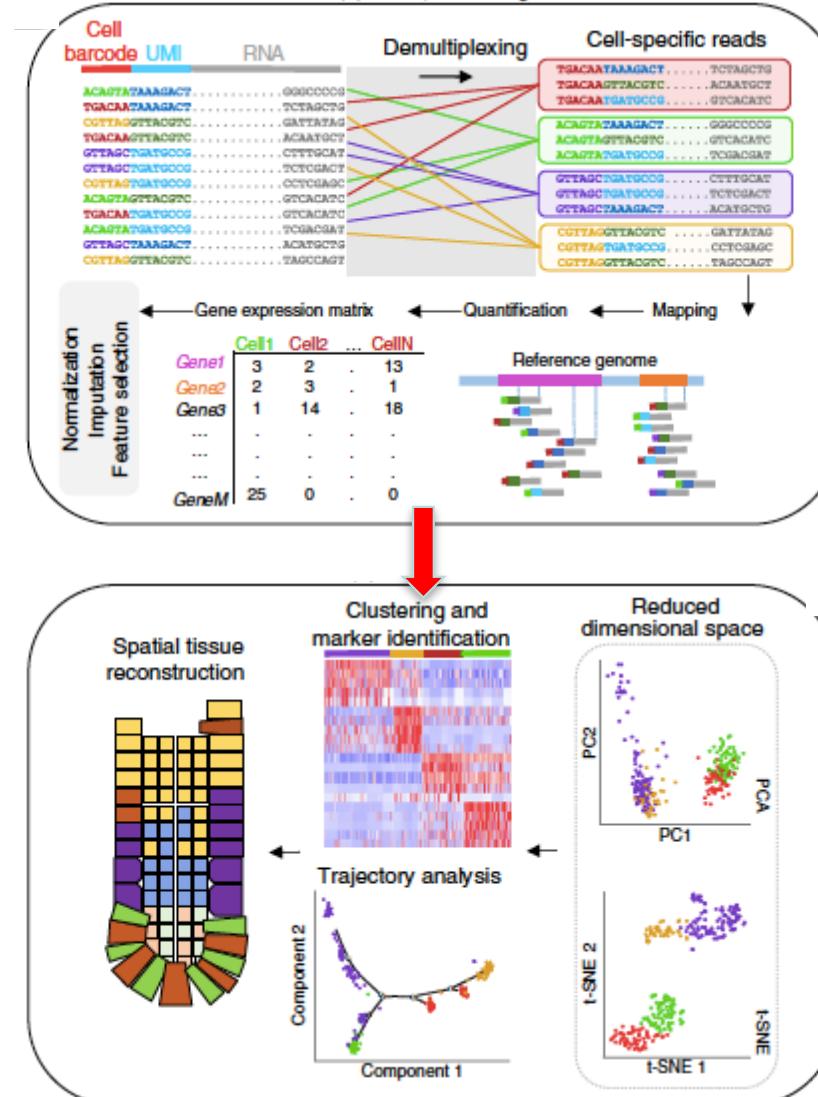


CHROMIUM SINGLE-CELL RNA SEQUENCING



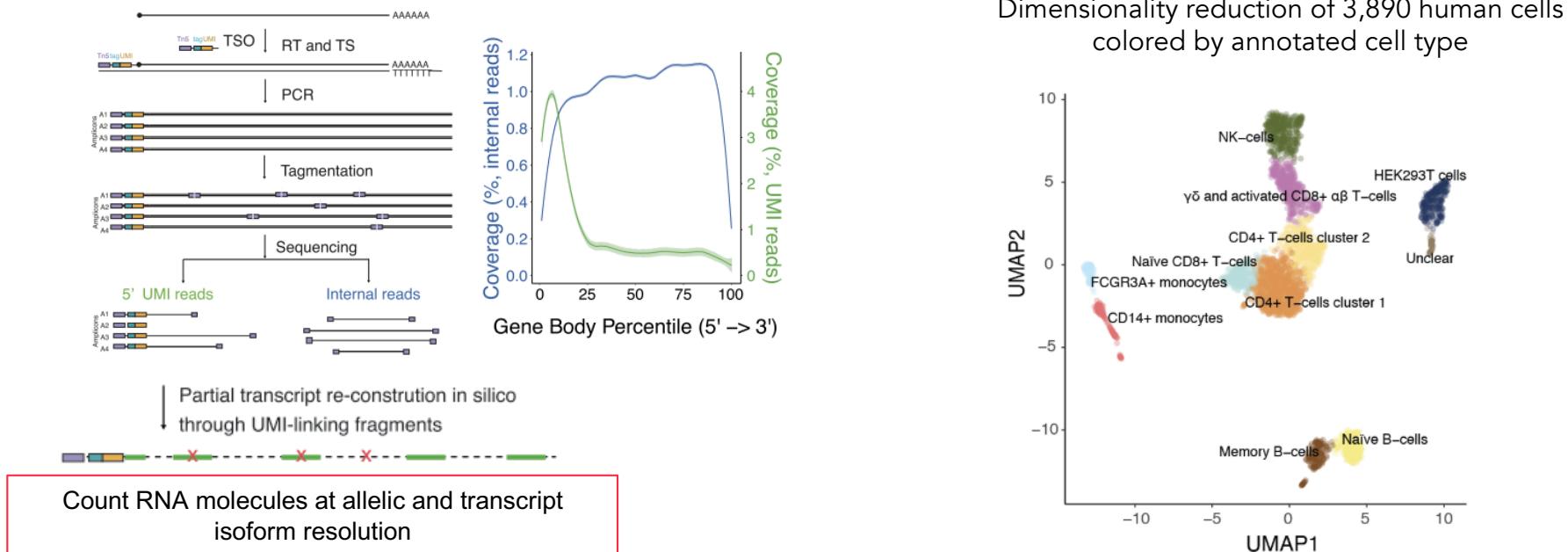
CHROMIUM SINGLE-CELL RNA SEQUENCING

DATA ANALYSES



Recent advances : Illumina sequencing with Smart-seq3

Single-cell RNA counting at allele- and isoform-resolution using Smart-seq3
Hagemann-Jensenn et al. *Nature Biotechnology*, 2020



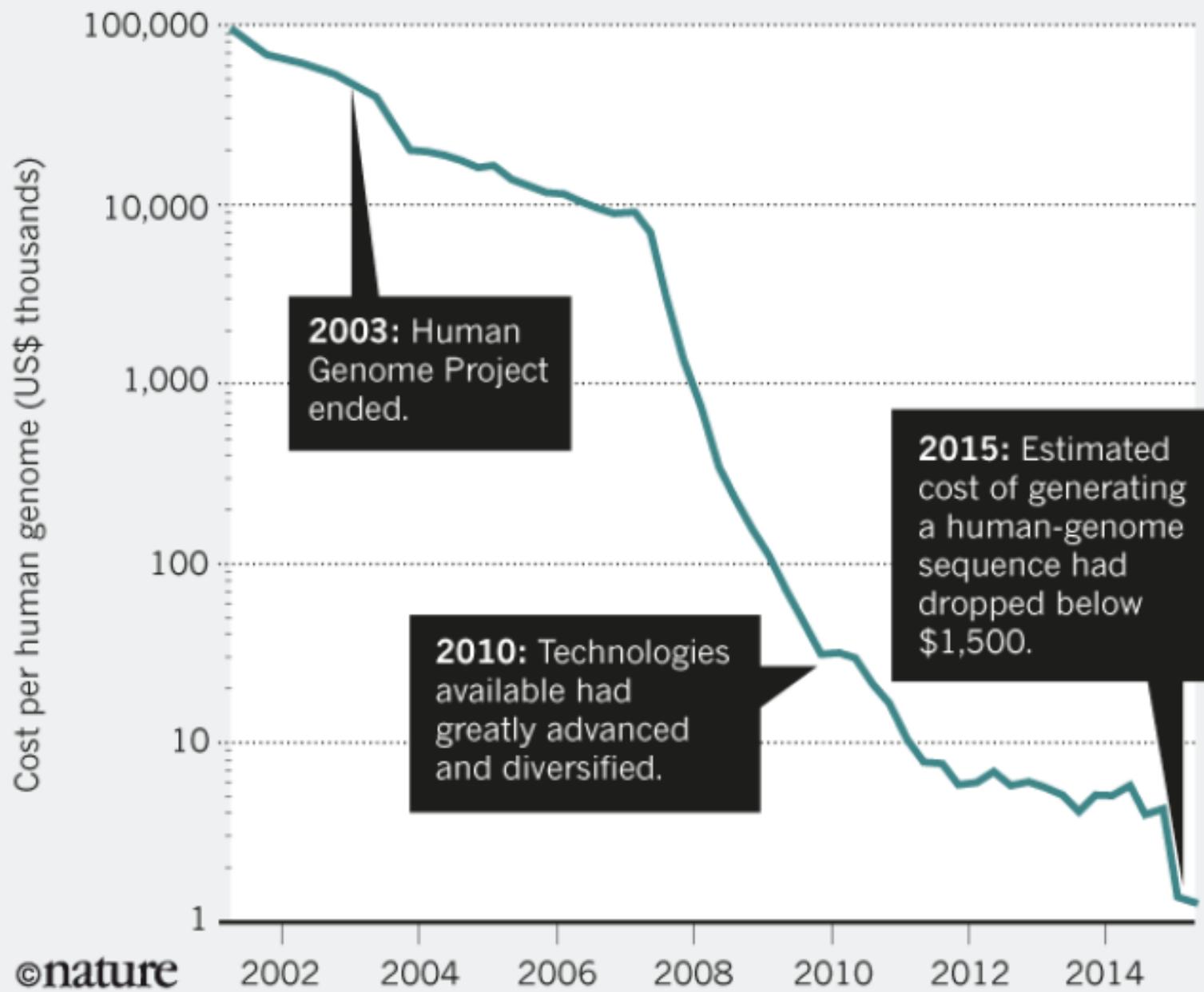
Smart-seq3 :

- full-length transcriptome coverage
- 5' UMI RNA → in silico reconstruction of thousands of RNA molecules/cell
- 60% assignments to allelic origin
- 30–50% assignments to specific isoforms

Smart-seq3 greatly increased sensitivity compared to Smart-seq2 detecting thousands more transcripts/cell

BETTER, CHEAPER, FASTER

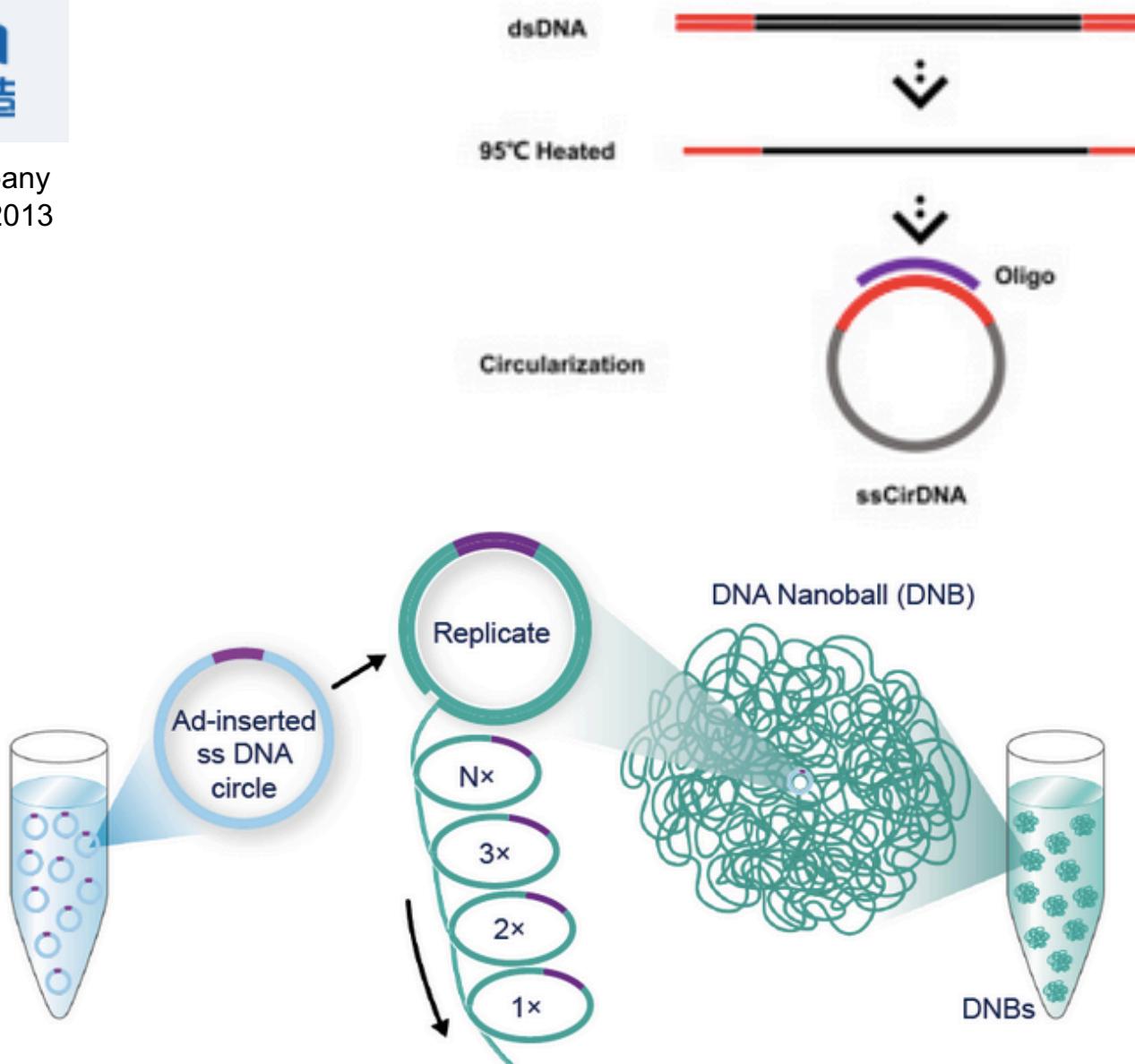
The cost of DNA sequencing has dropped dramatically over the past decade, enabling many more applications.



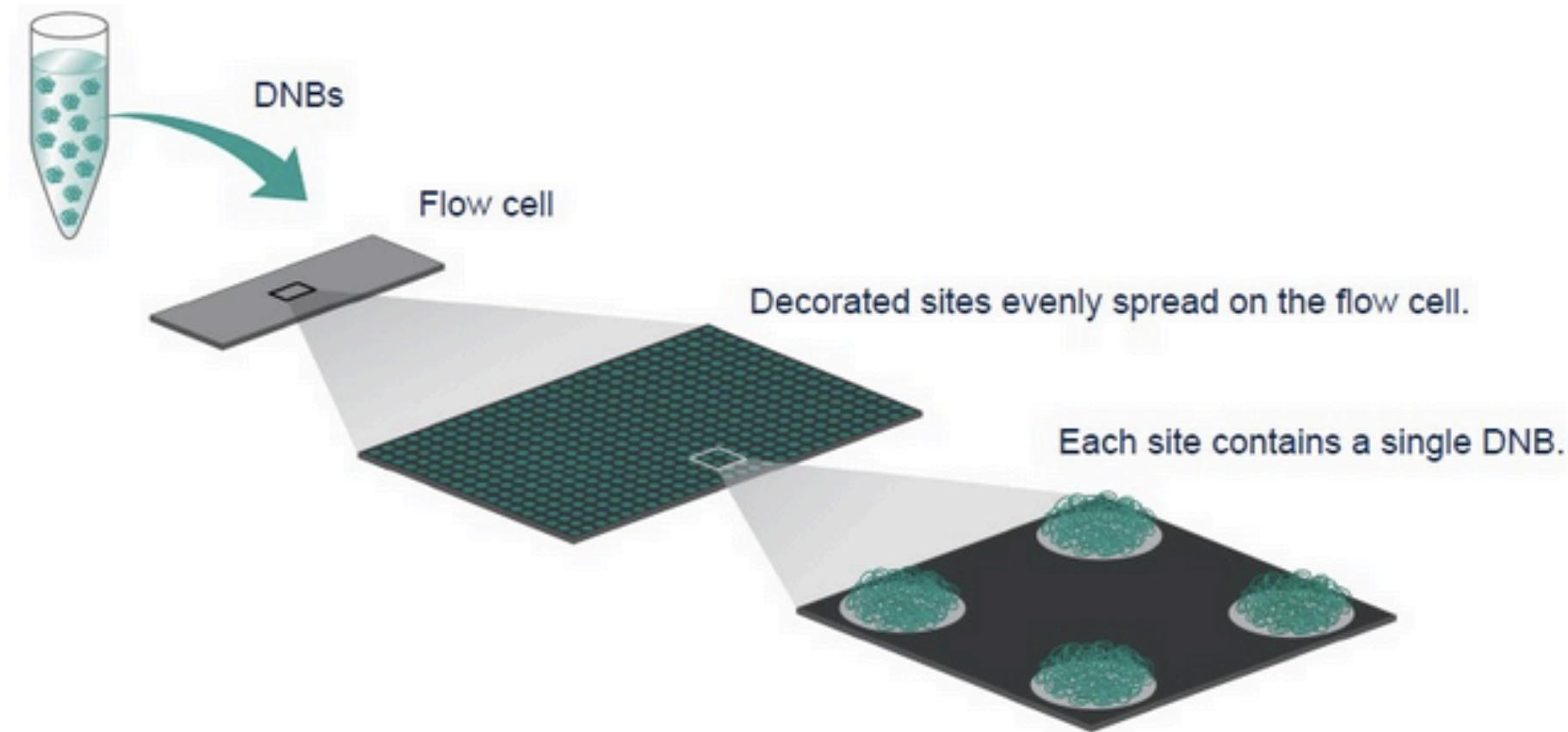
NEW SEQUENCING TECHNOLOGY



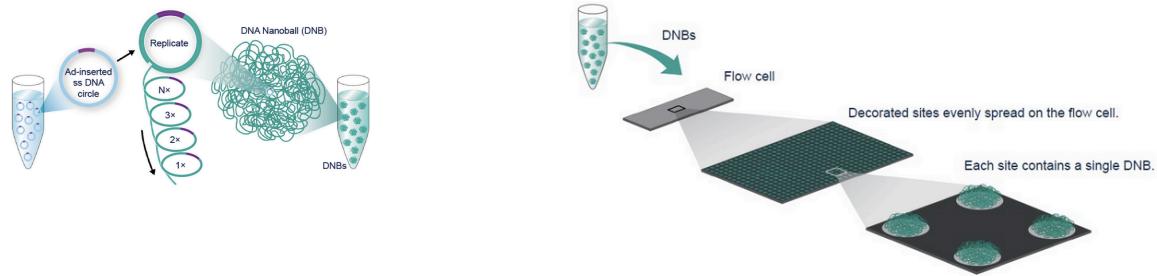
acquisition of U.S. company
Complete Genomics in 2013



NEW SEQUENCING TECHNOLOGY



— NEW SEQUENCING TECHNOLOGY —



Tests préliminaires (Genoscope, CNRGH)

- Qualité : Q30 moyen élevé (même en 2x200)
- Taux d'erreur MGI < taux d'erreur NovaSeq
- % reads mappés MGI > % reads mappés Illumina

Mais :

- Runs plus longs (+ lavage 6h)
- Preparation of nanoballs « délicate »
- Régions riches A/T et G/C moins couvertes

	G400	NovaSeq S1	NovaSeq S4
Run time	66h	24h	48h
Output	500 Gb	500 Gb	1600-2000 Gb
Cost/Gb (€)	3.8 - 4.5	9.9 - 12.4	6.6 - 8.3

PART 2

3rd GENERATION SEQUENCING

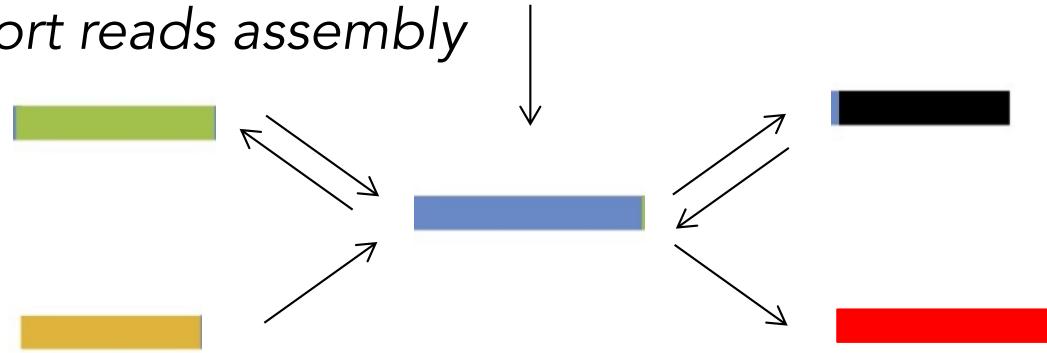
LONG READS

LONG-READS VERSUS SHORT-READS

Assembly of DNA fragments with repeated sequences



NGS short reads assembly

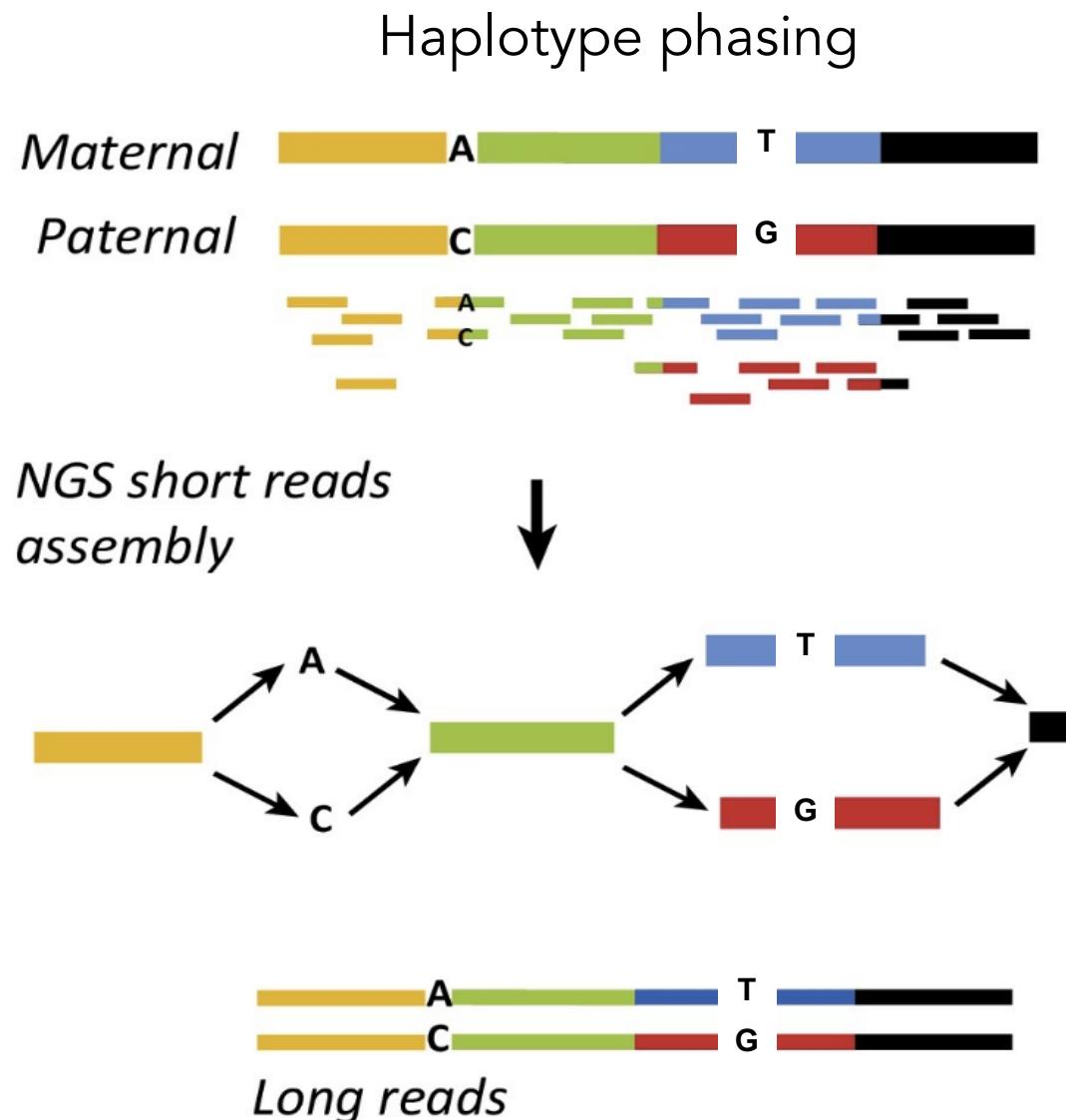


Several contigs → incomplete assembly, underestimation of repeats

Long reads assembly

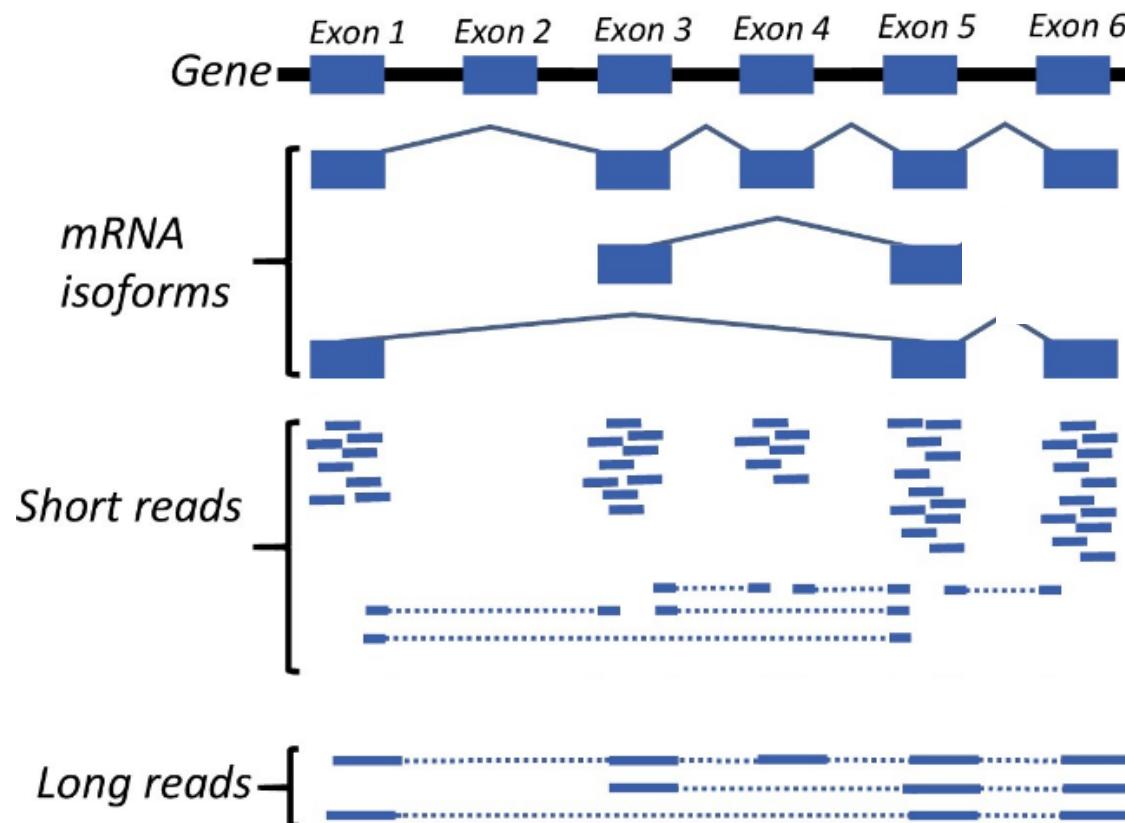


LONG-READS VERSUS SHORT-READS



LONG-READS VERSUS SHORT-READS

Detection of splicing isoforms



The 3rd generation winning technologies



Sequel - Pacific Biosciences

Single molecules

Up to 80,000 bp long

Error rate ≈ 10-15 % - CCS: <1%

Compensated by coverage



MinION - Oxford Nanopore

Single molecules

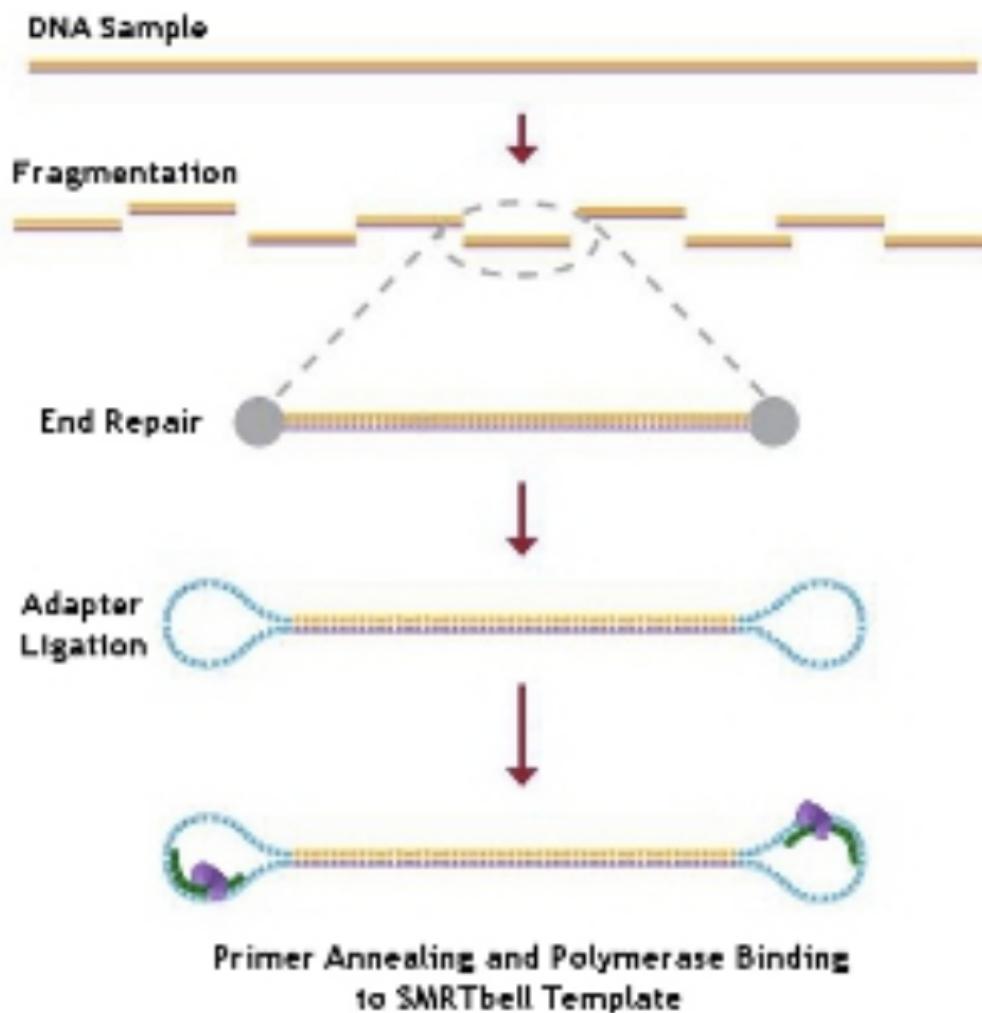
> 200 000 bp long

Error rate ≈ 10-15 %

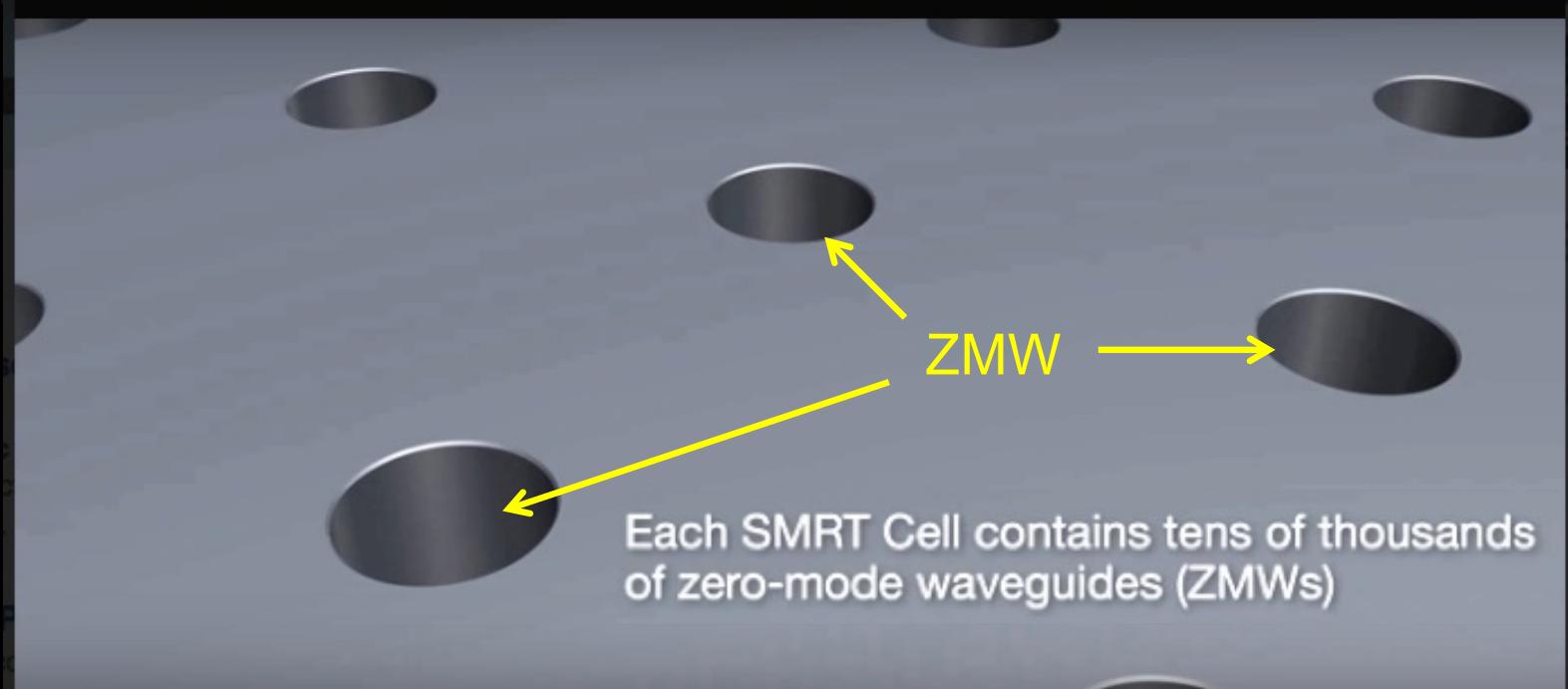
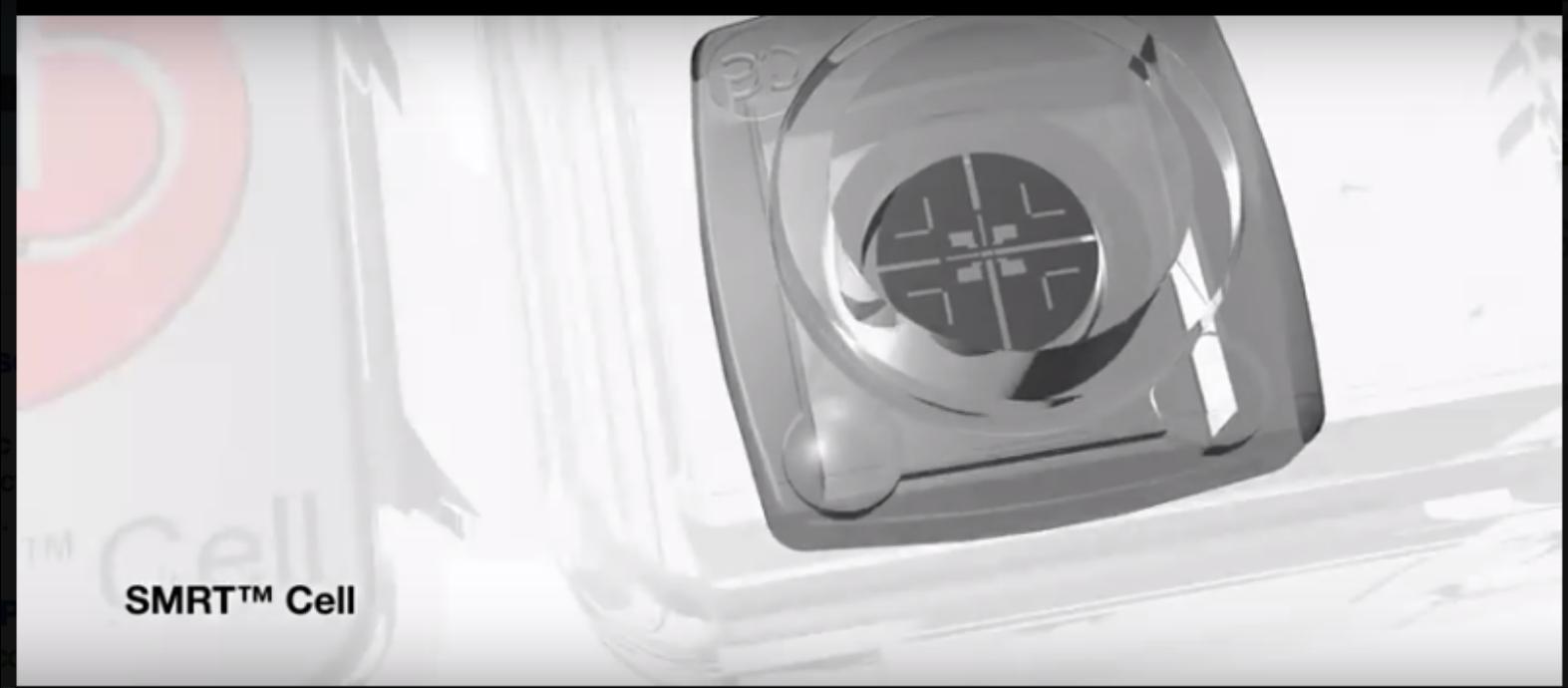
Compensated by coverage

PacBio : Single Molecule Real Time (SMRT) sequencing

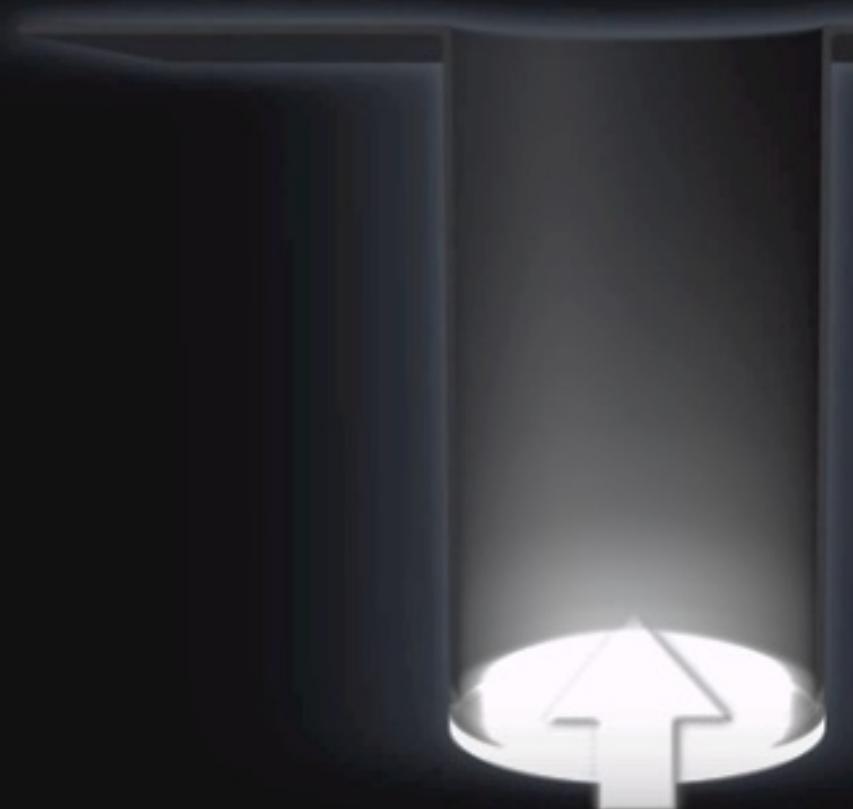
PacBio DNA-seq library



PACIFIC BIOSCIENCES



PACIFIC BIOSCIENCES

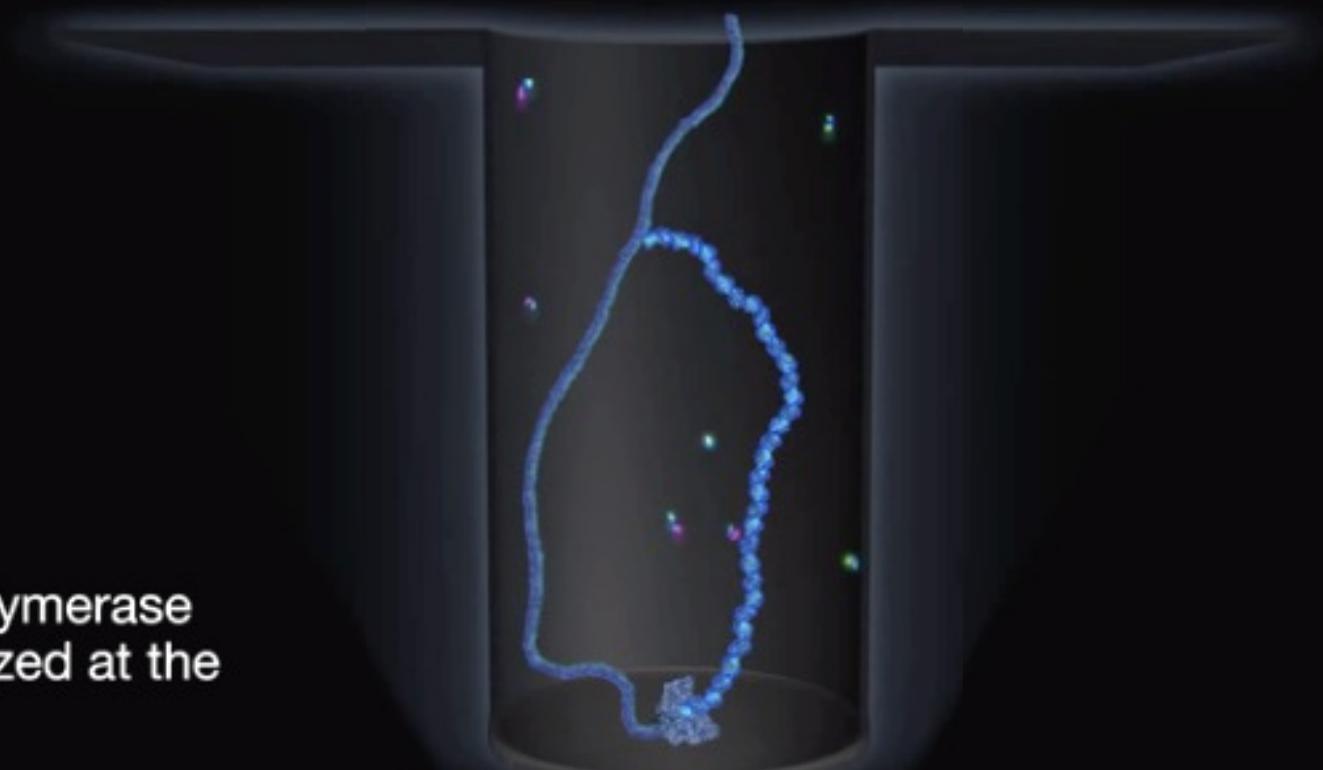


ZMW : optical waveguide that guides light energy into a volume that is small compared to the wavelength of the light

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

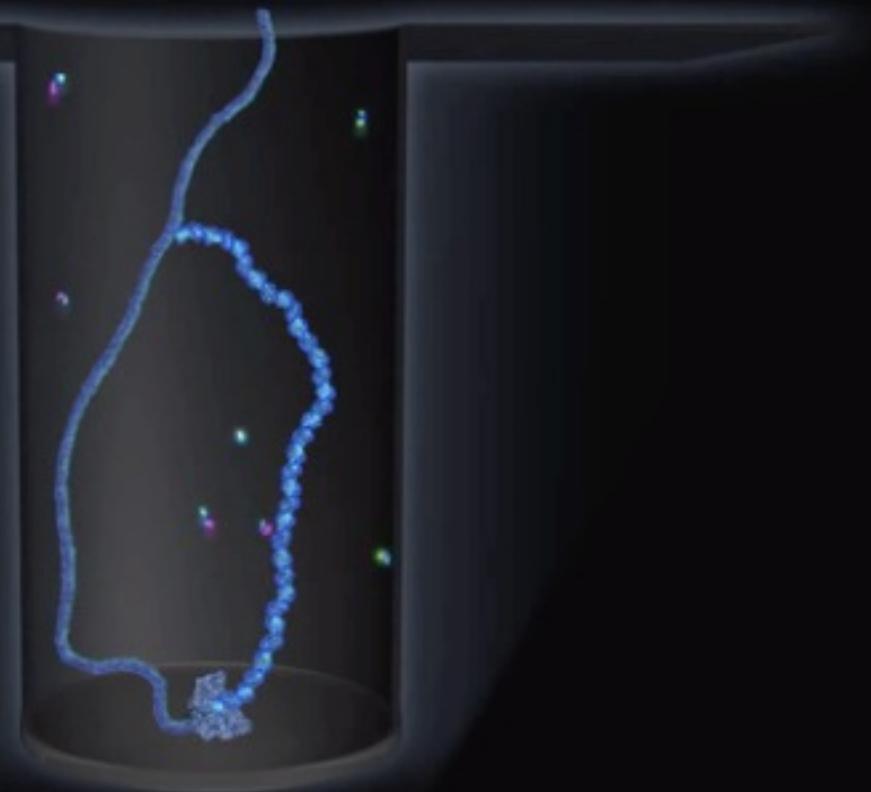
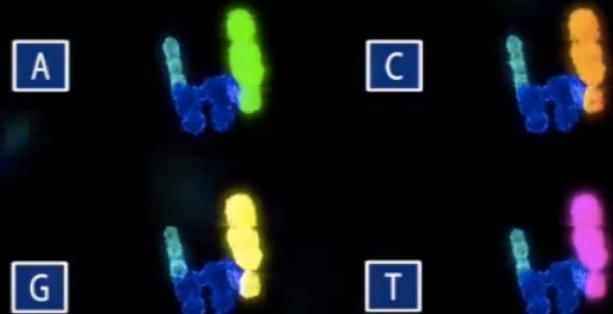
PACIFIC BIOSCIENCES

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



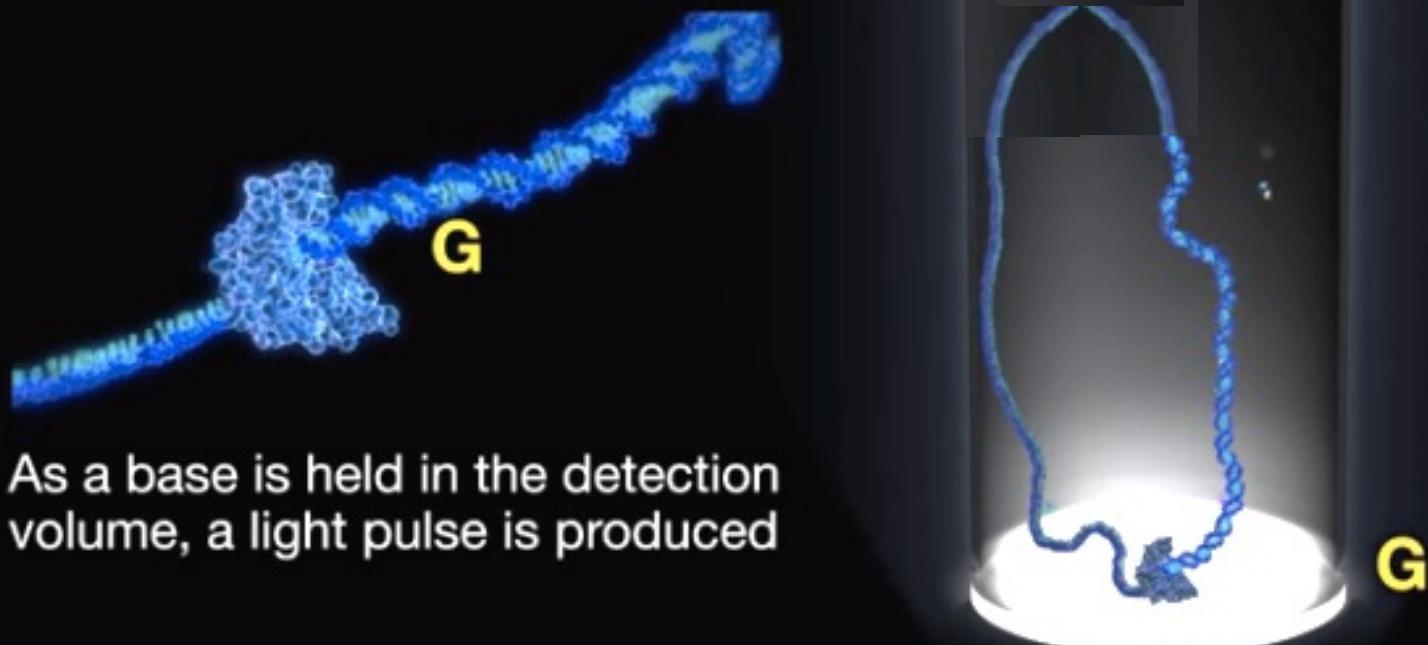
PACIFIC BIOSCIENCES

- Phospholinked Nucleotides



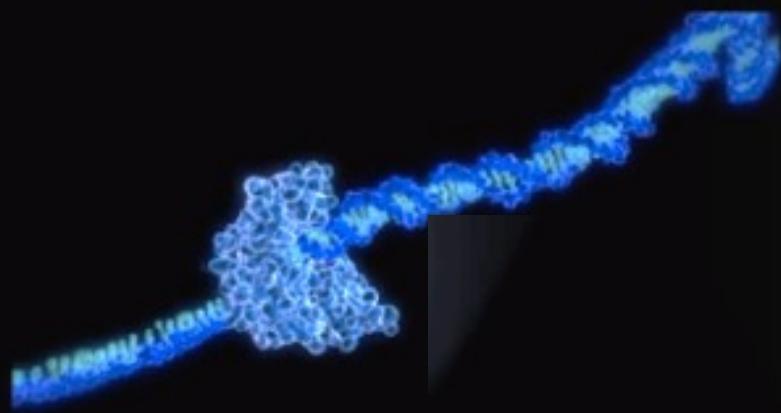
Phospholinked nucleotides are introduced into the ZMW chamber

PACIFIC BIOSCIENCES



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

PACIFIC BIOSCIENCES

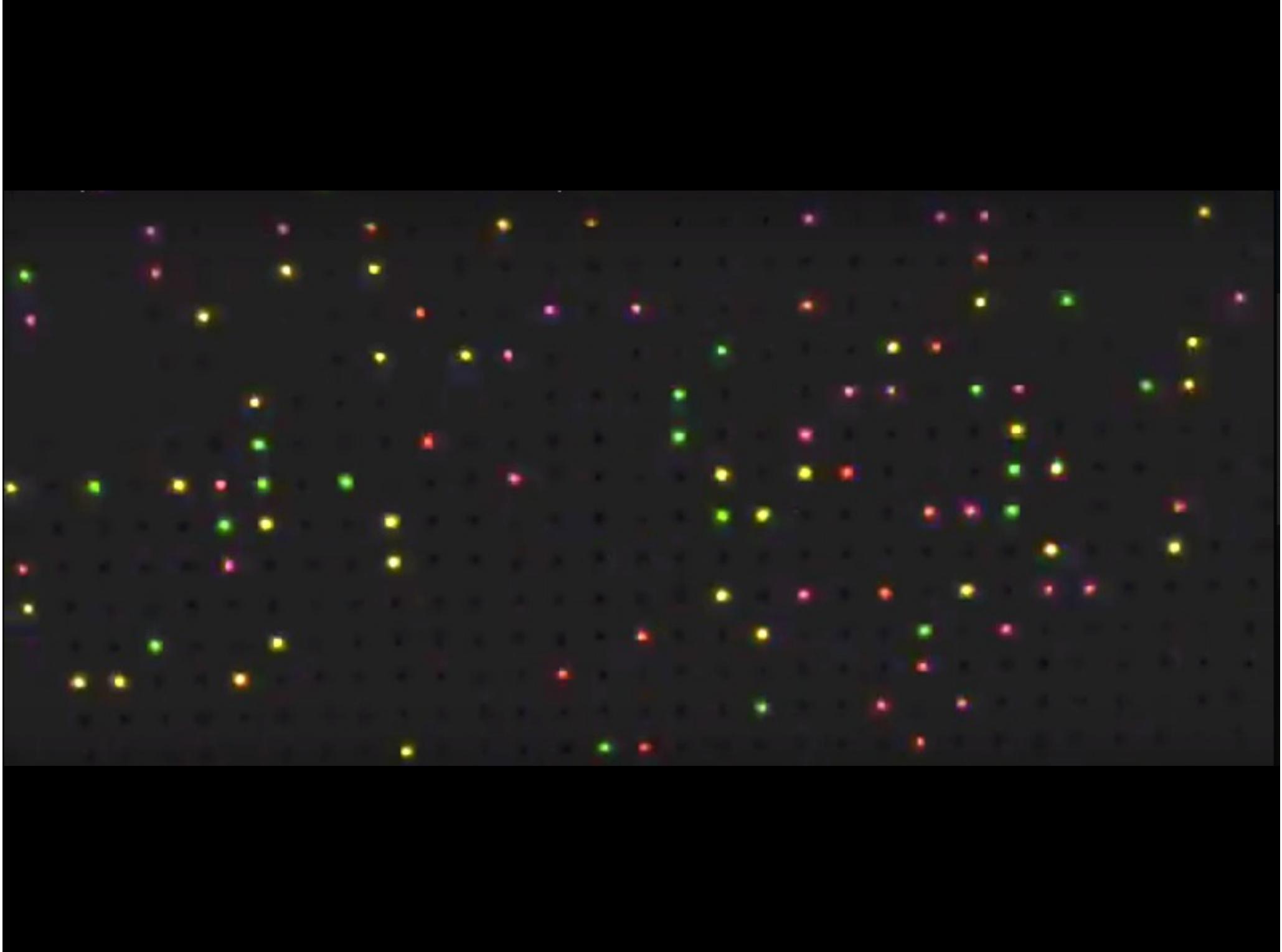


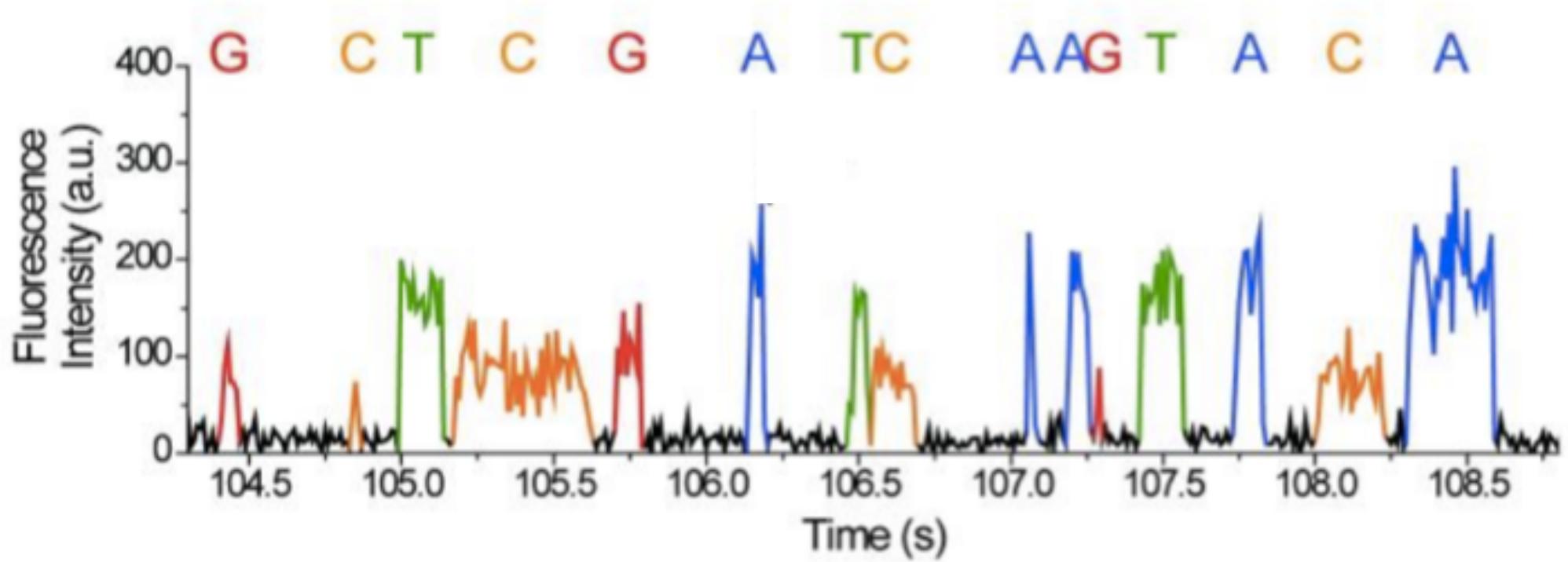
After incorporation the phosphate chain is cleaved, releasing the attached fluorophore



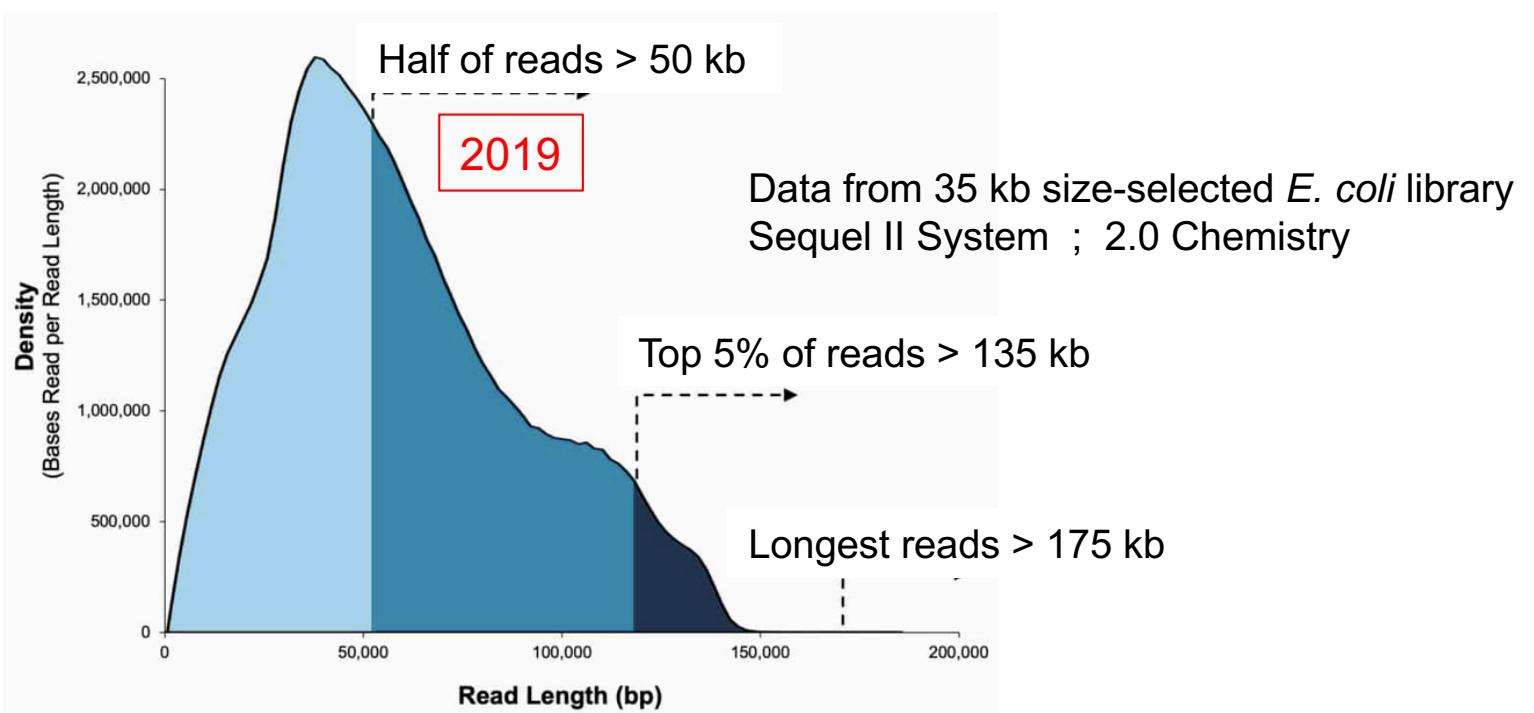
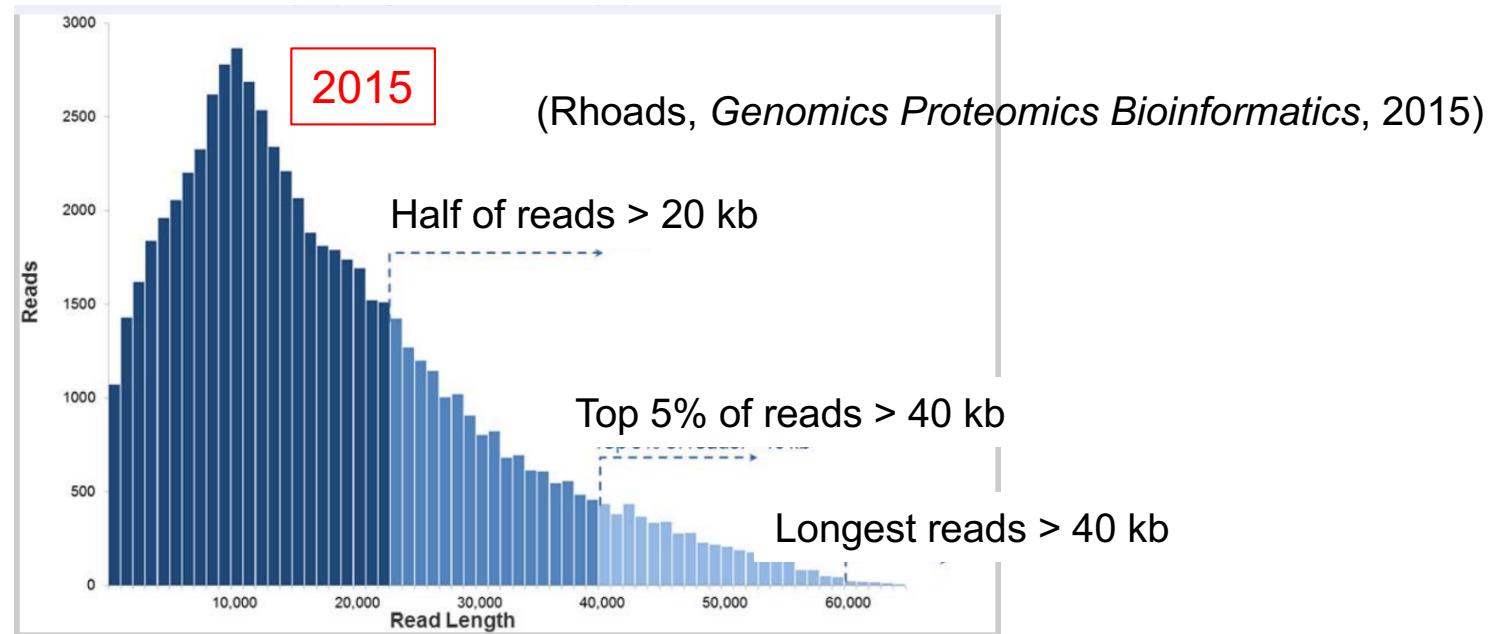
PACIFIC BIOSCIENCES



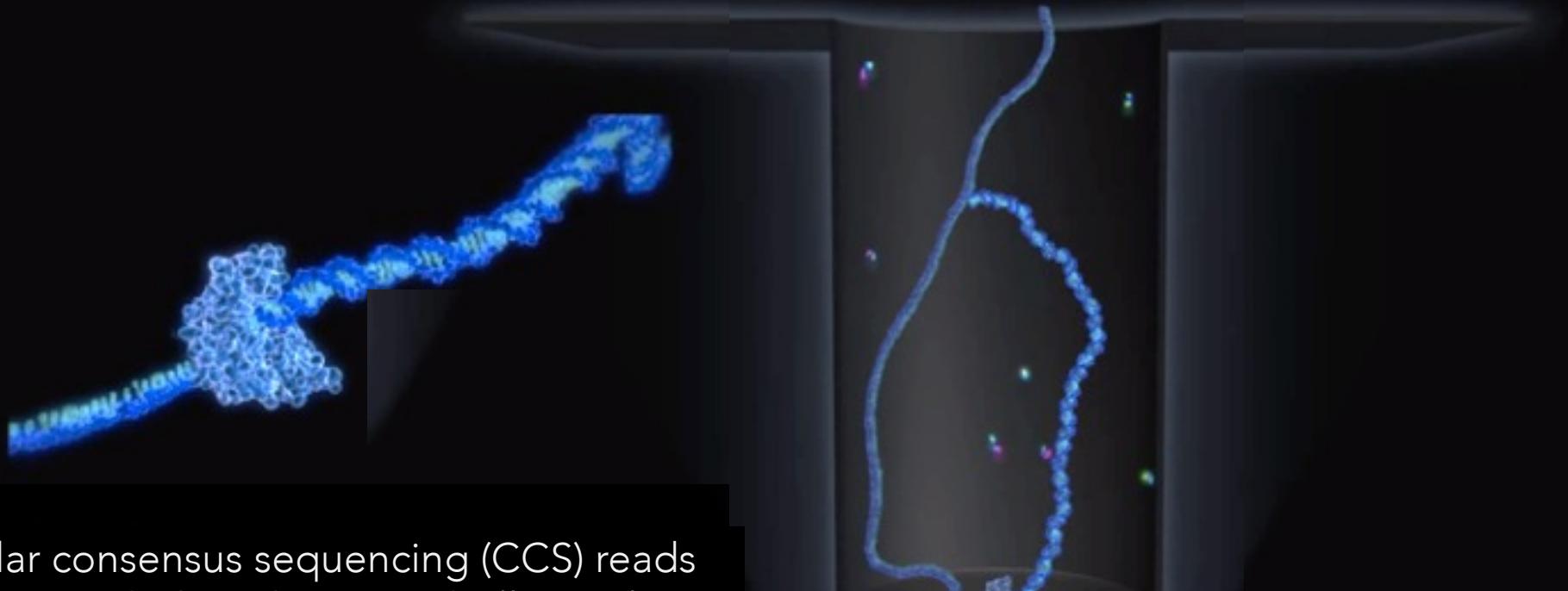




Length of PacBio reads



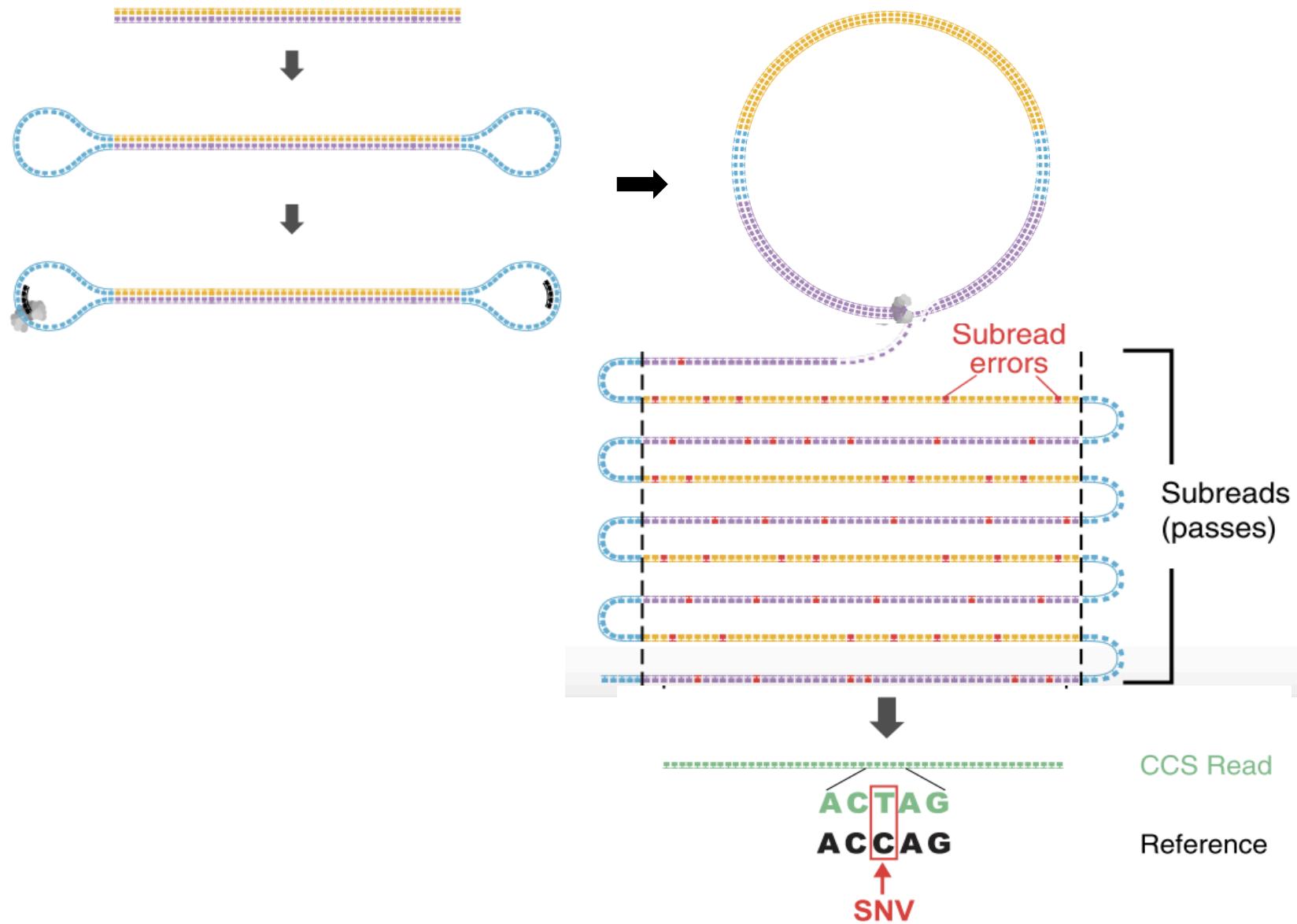
RECENT IMPROVEMENT WITH NEW CHEMISTRY



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

This allows a highly accurate sequencing by correction of random errors

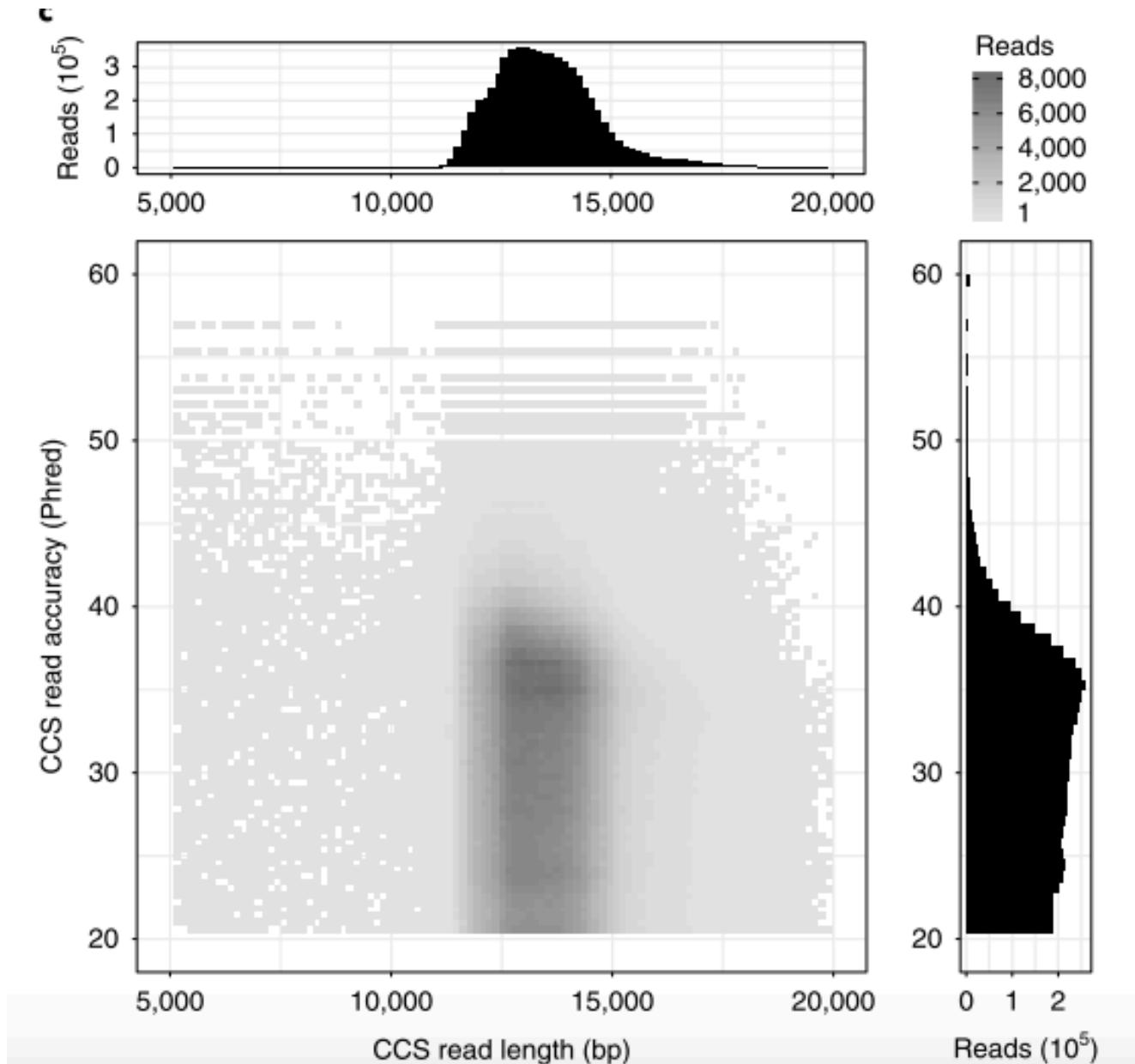
Circular Consensus Sequences (CCS): HIFI READS



RECENT IMPROVEMENT: GENOME ASSEMBLY WITH CCS

Circular consensus assembly of a human genome

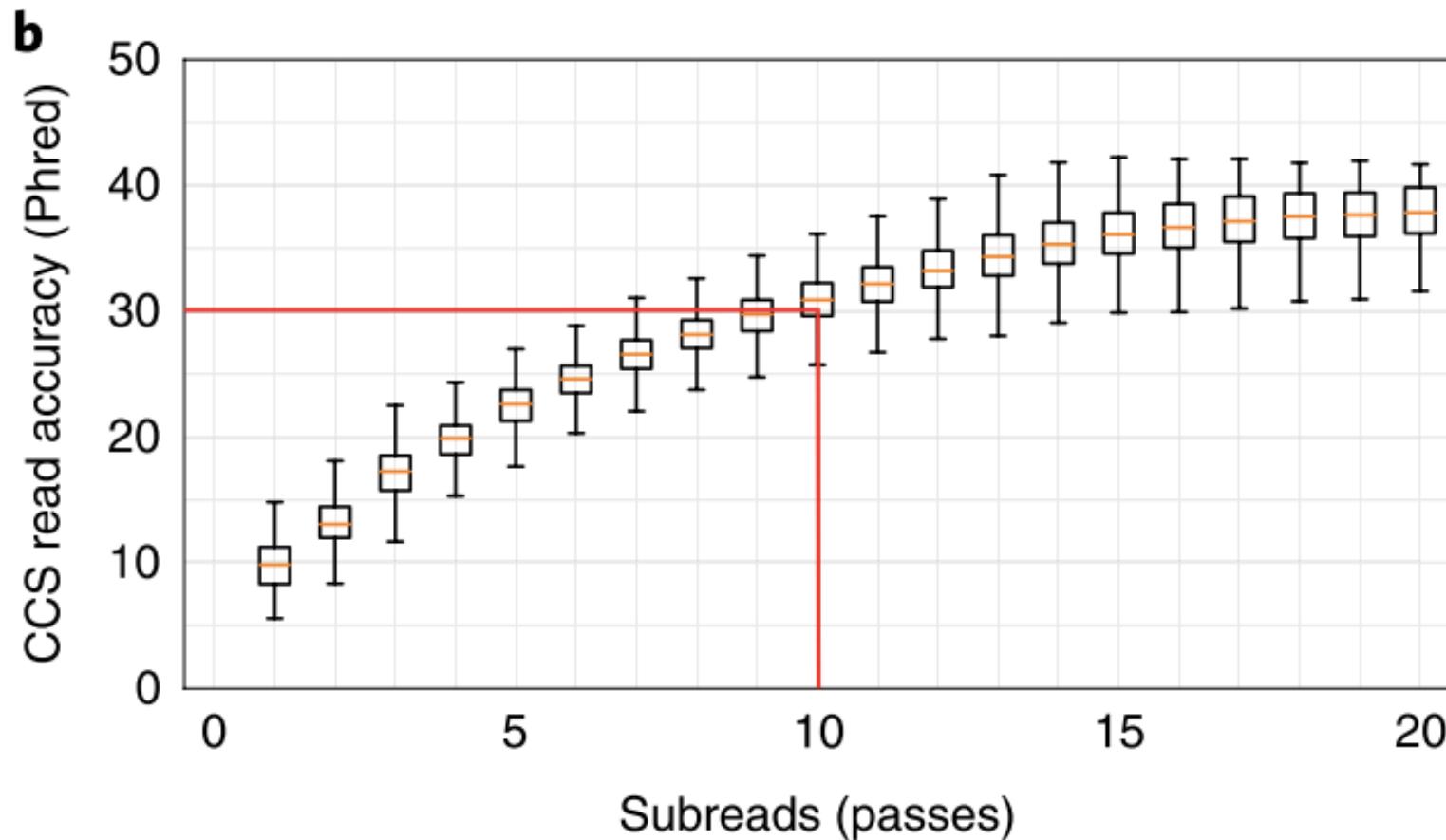
Wenger et al. *Nat. Biotechnol.* oct. 2019



RECENT IMPROVEMENT: 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.* oct. 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

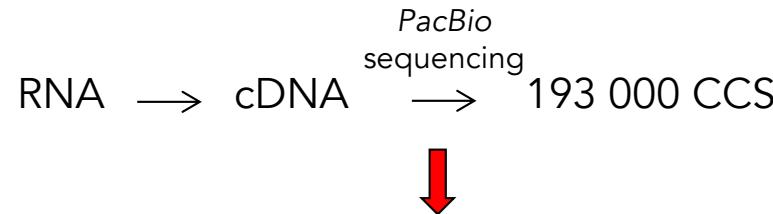
Comparison with NovaSeq

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

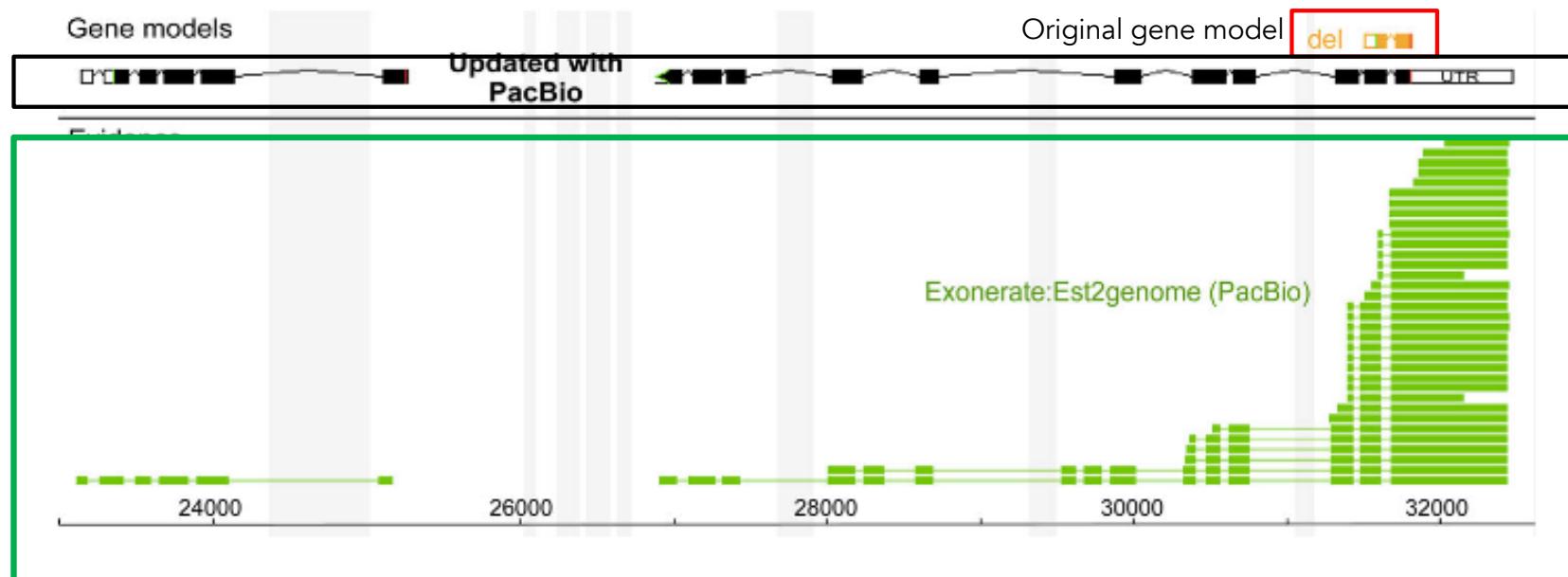
SEQUENCING cDNA USING CIRCULAR CONSENSUS SEQUENCES

Genome annotation of the parasitic hookworm *Ancylostoma ceylanicum*
using single molecule mRNA sequencing

Magrini et al. *BMC Genomics*, 2018



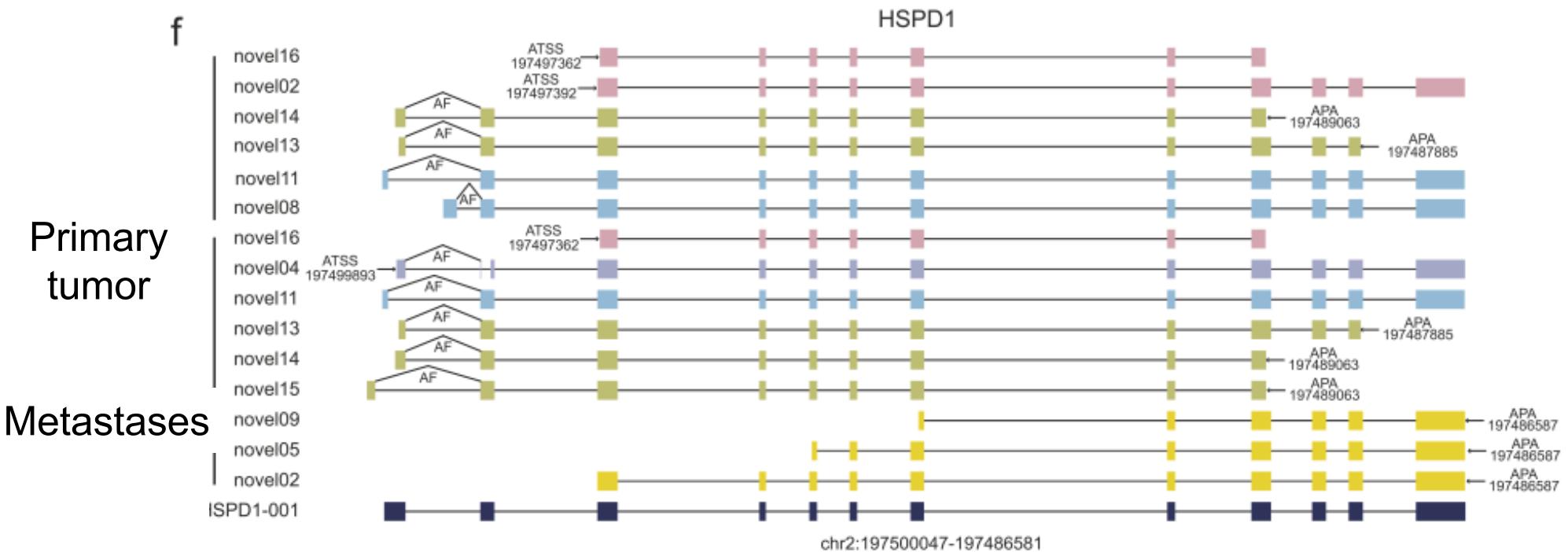
- Increased the total genomic exon length by 1.9 Mb (12.4%)
- 1609 (9.2%) new genes



PacBio cDNA SEQUENCING

Hybrid full-length transcriptome in metastatic ovarian cancer

Jing et al. *Oncogene* 2019



Long-read full-length transcriptome analysis

- improves molecular diagnostic
- reveals novel therapeutic vulnerabilities