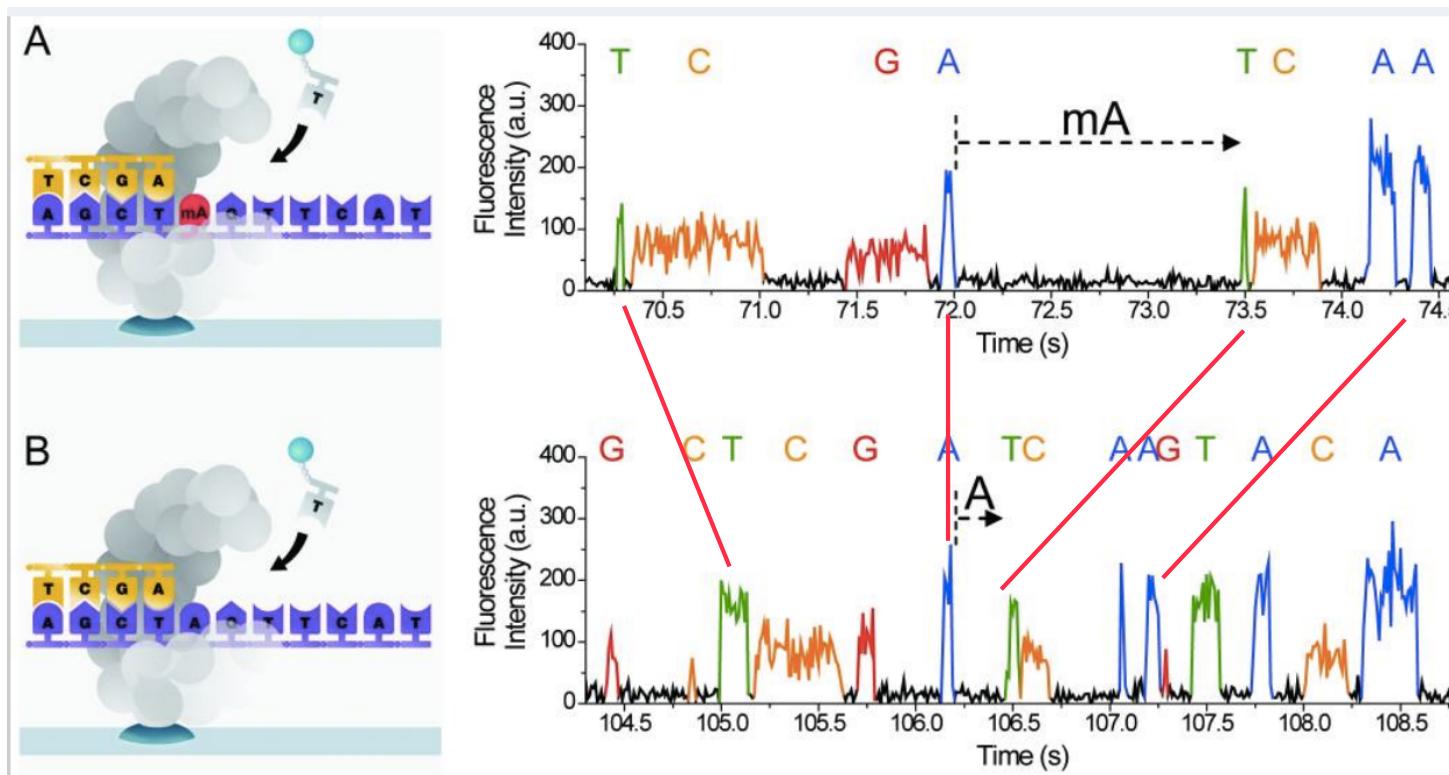


DETECTION OF MODIFIED BASES



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

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

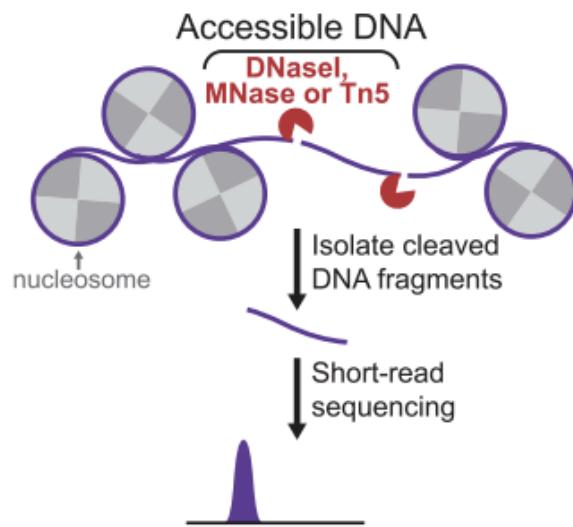
Feng et al. *PLOS Comput Biol* (2013)

Detection of m6A with CCS

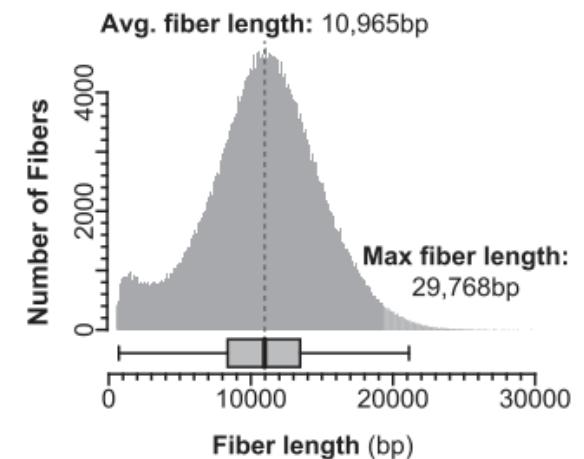
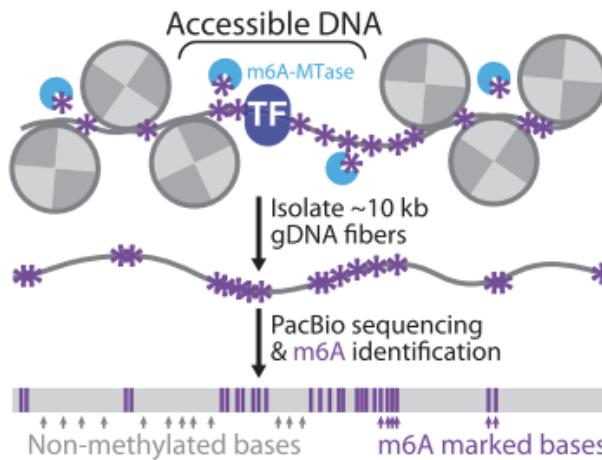
Single-molecule regulatory architectures captured by chromatin fiber sequencing
Sternbach et al. Science (2020)

DnaseI-seq.

Cleavage-based assay:

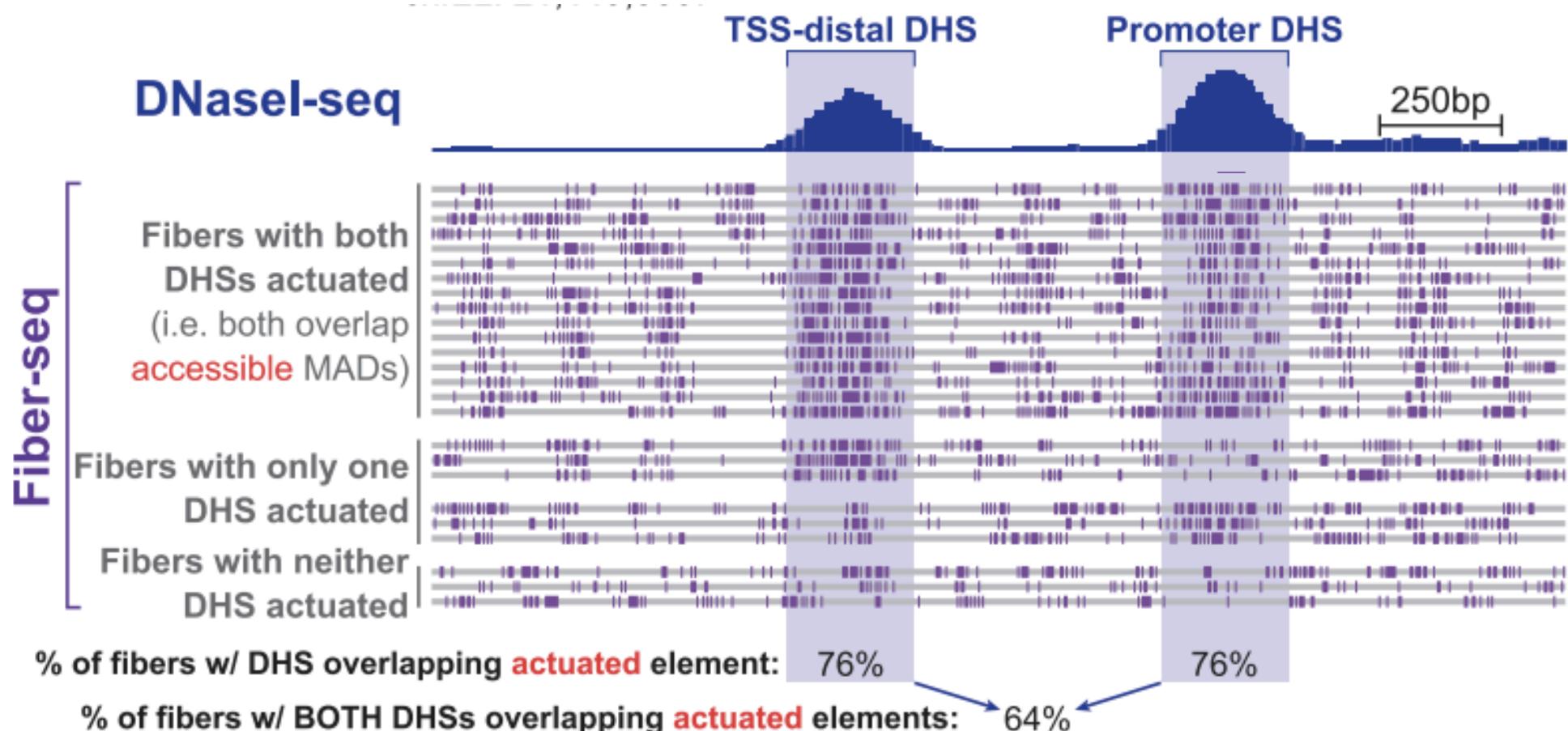


Fiber-seq.



Detection of m6A with CCS

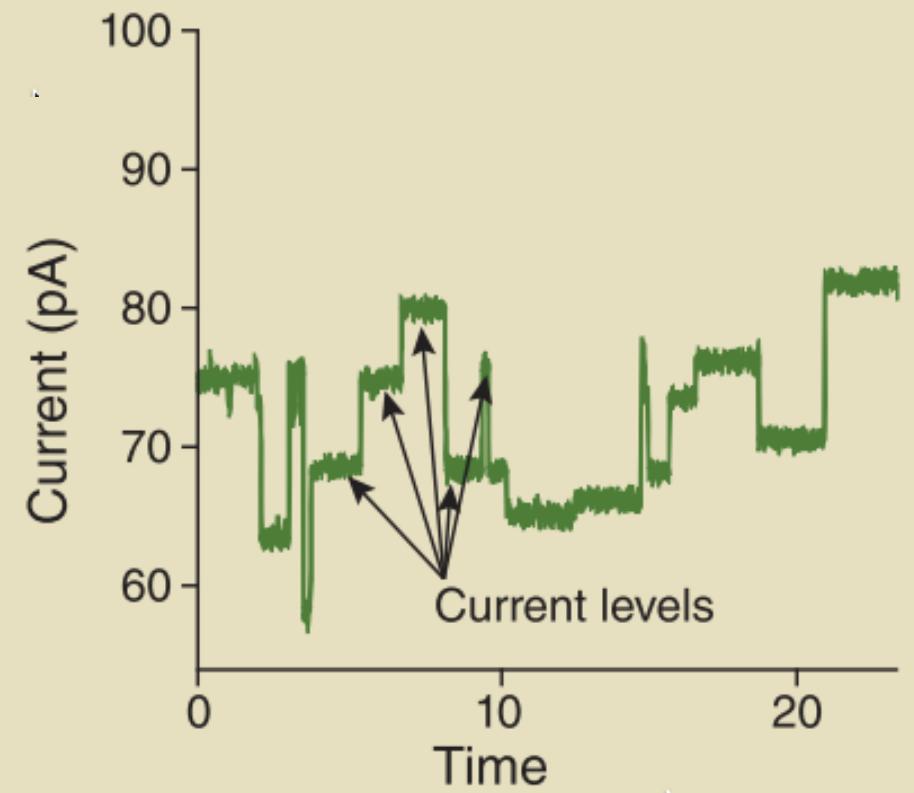
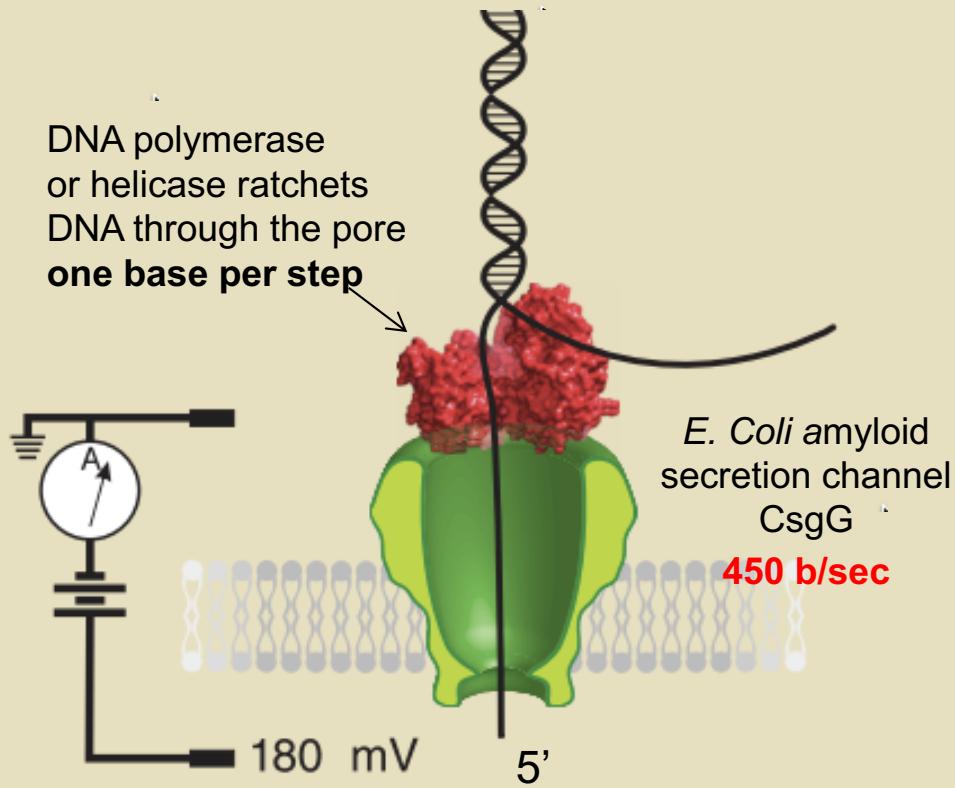
Single-molecule regulatory architectures captured by chromatin fiber sequencing
Sternbach et al. Science (2020)



Next Generation Sequencing

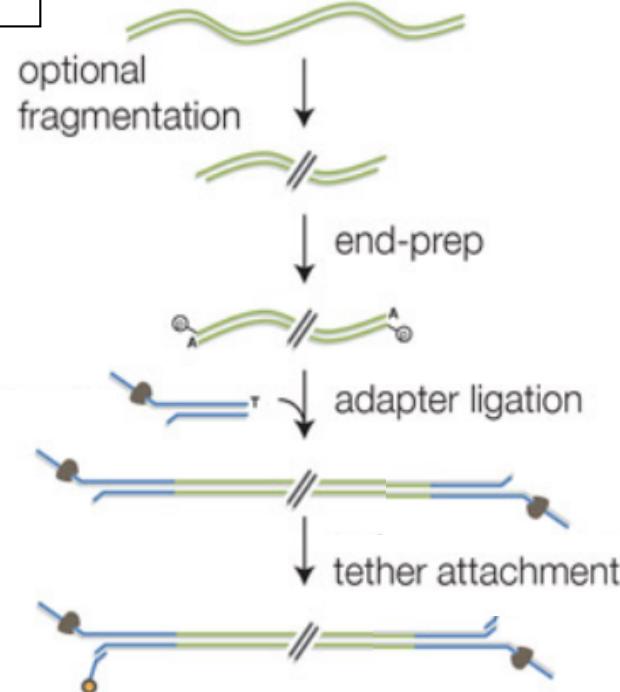


BASIC CONCEPTS

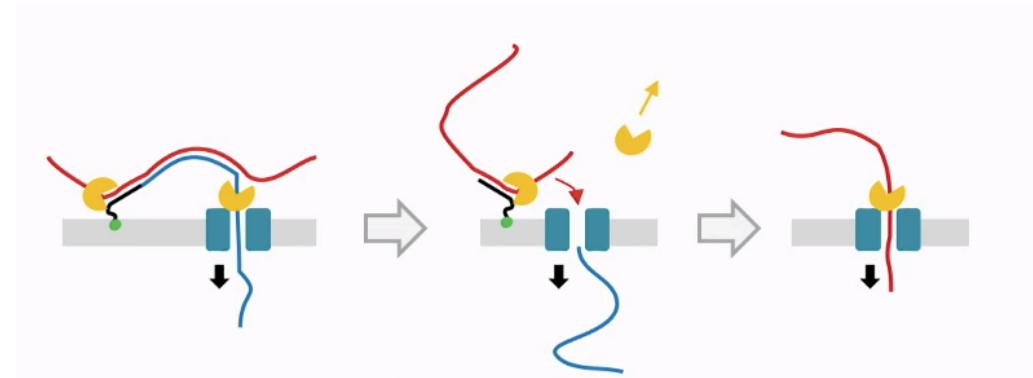
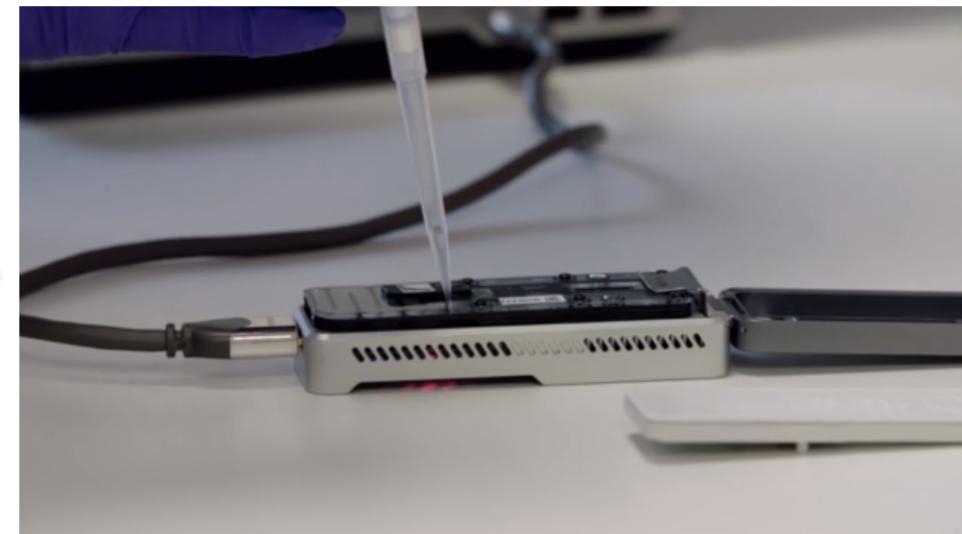


SEQUENCING PROCESS

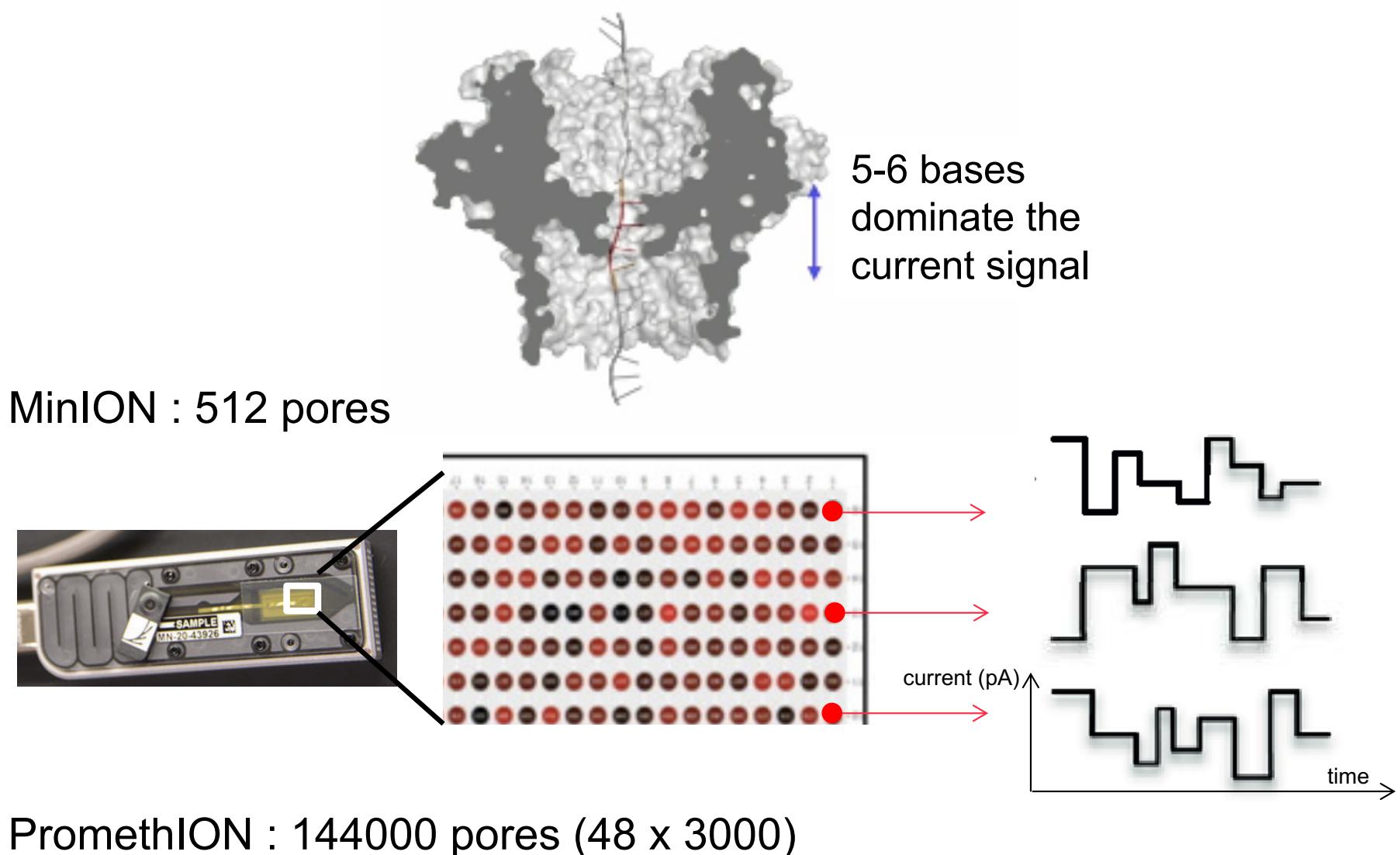
1D² Library
(2017)



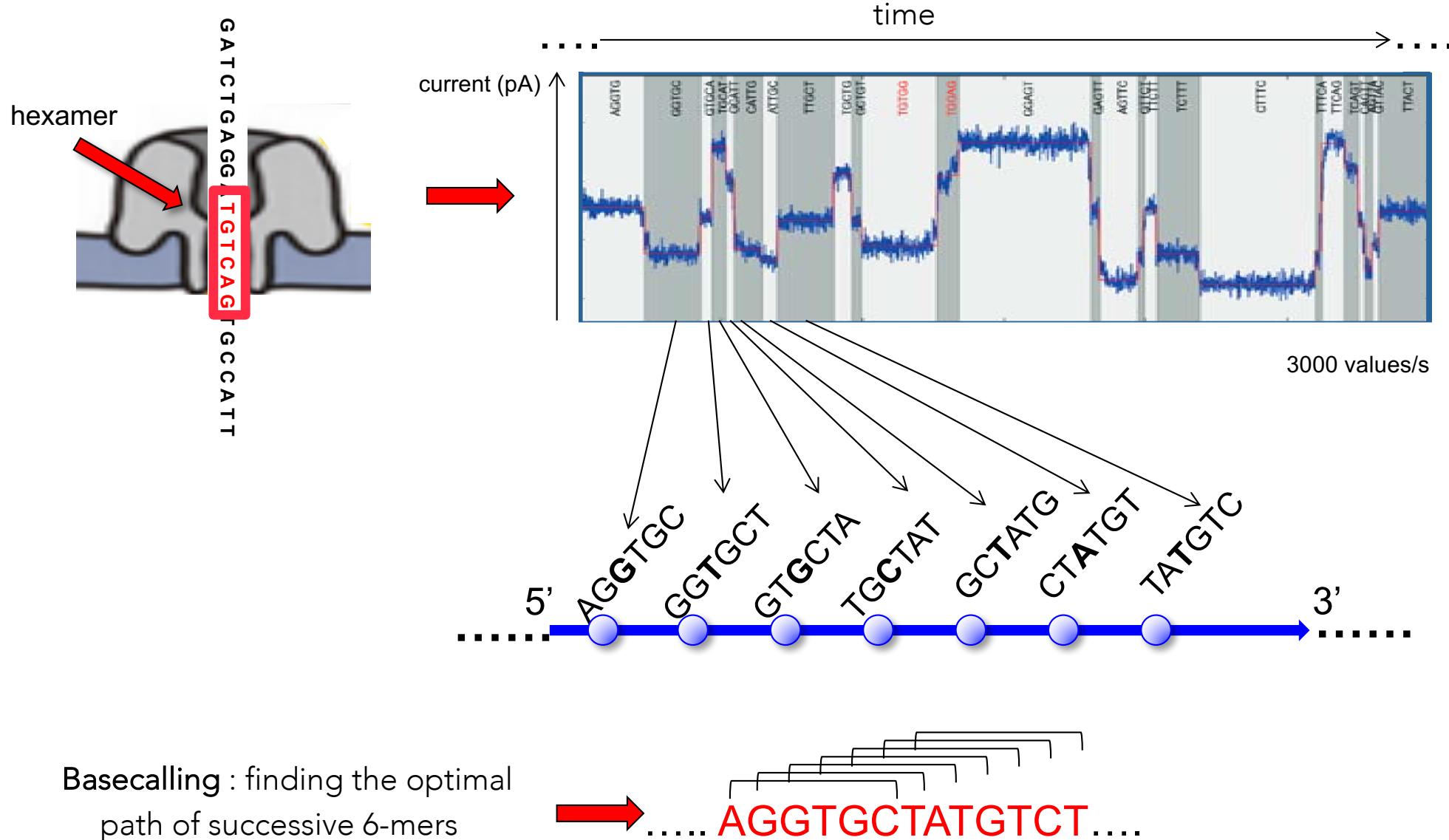
SEQUENCING



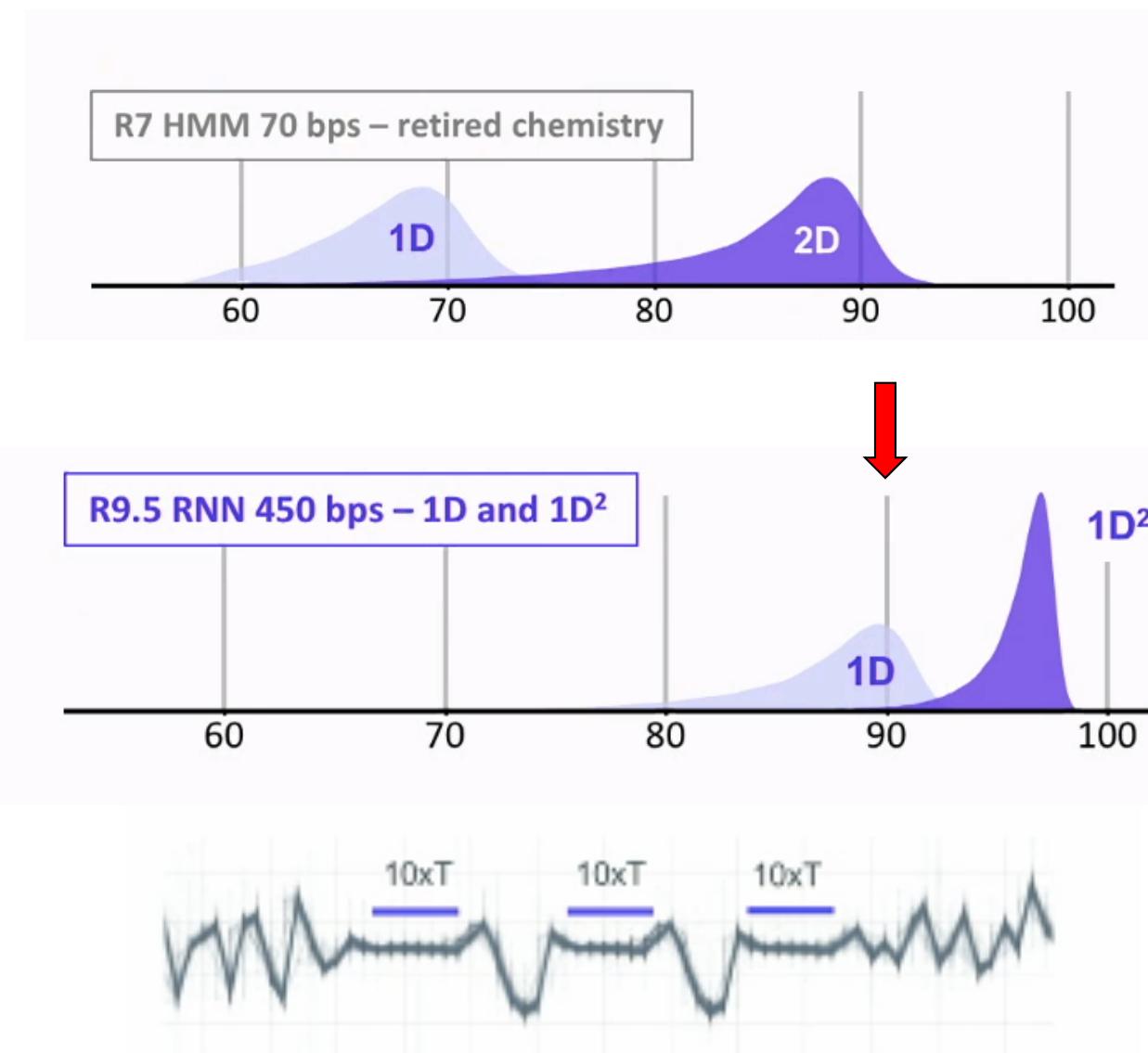
SEQUENCING PROCESS : MinION FLOW CELL



BASECALLING



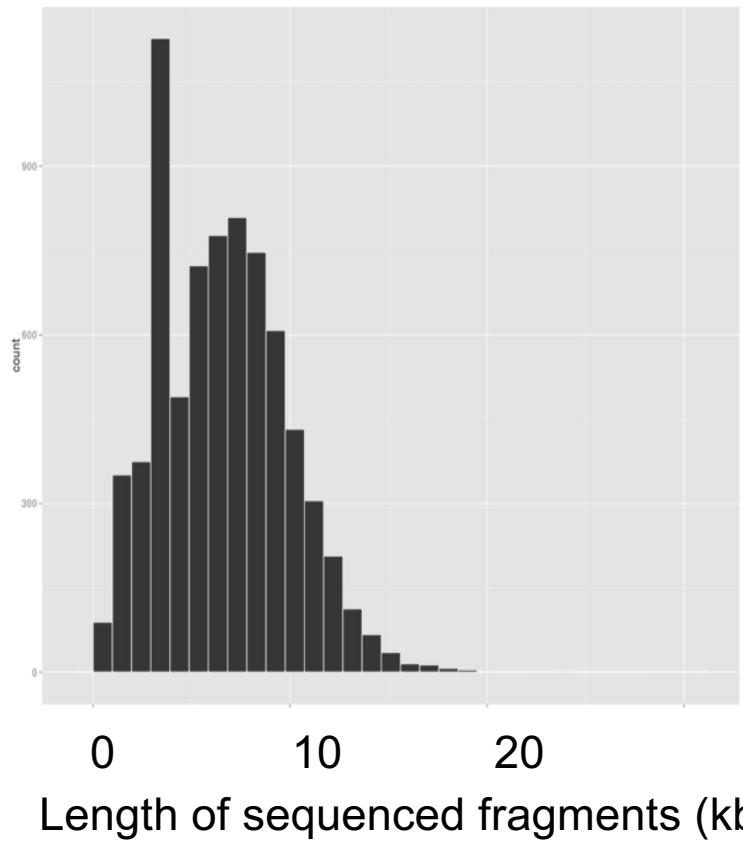
QUALITY



Homopolymers difficult to sequence

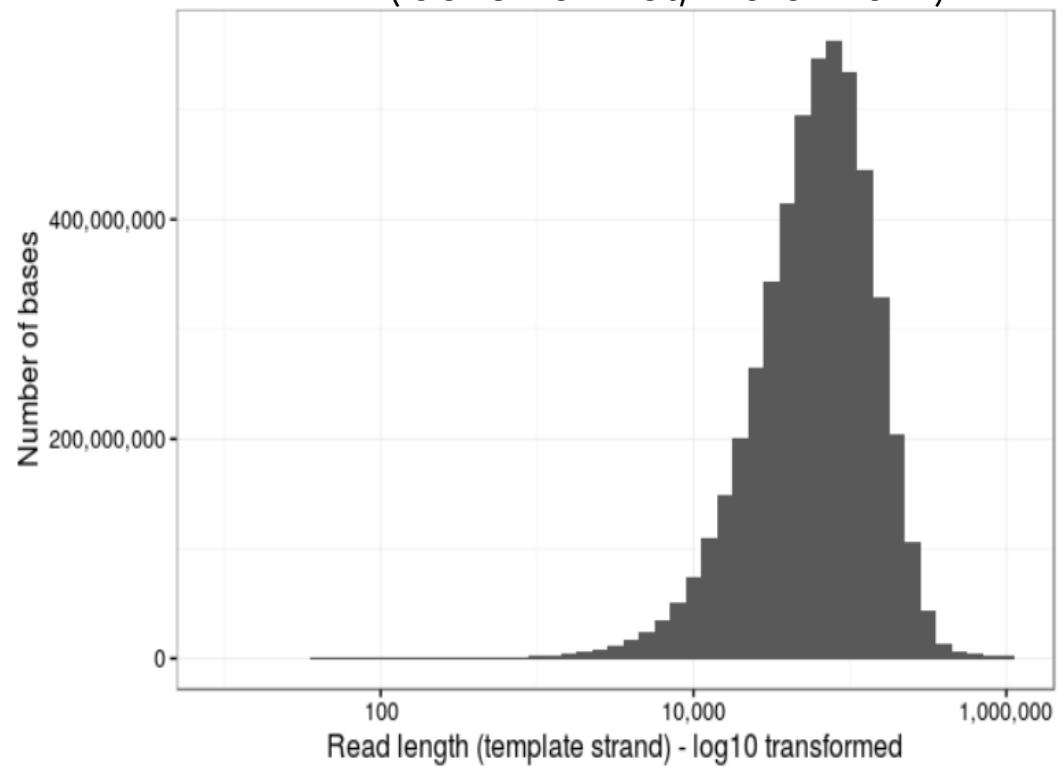
SIZE OF SEQUENCED DNA FRAGMENTS

Typical profile of fragment size



(Risse et al. *GigaScience*, 2015)

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

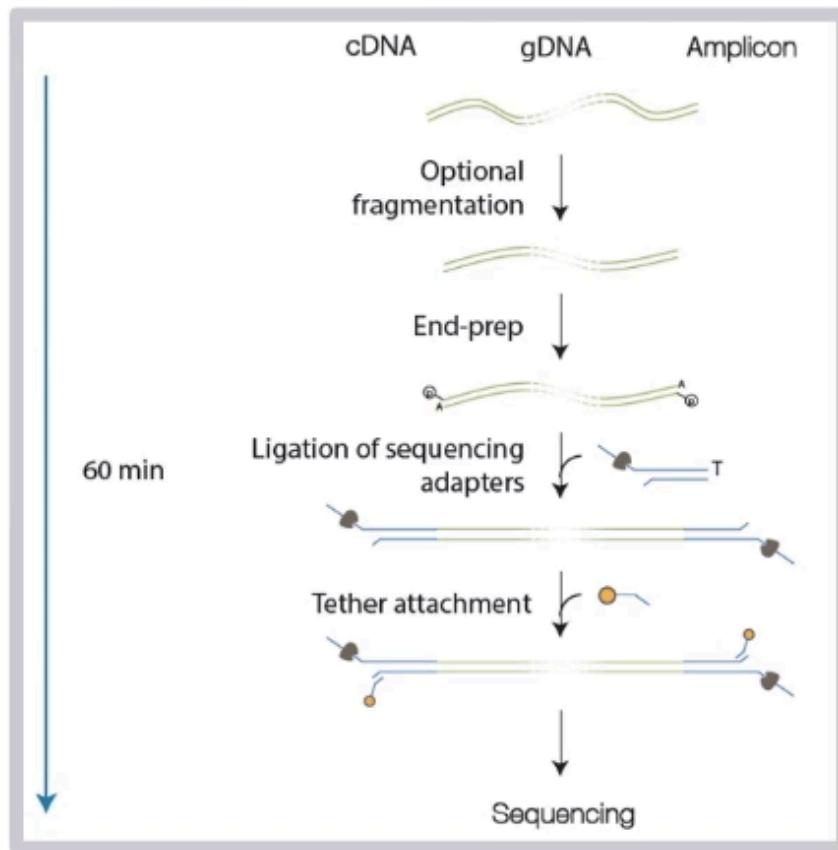


Size of the longest read : 778 kb

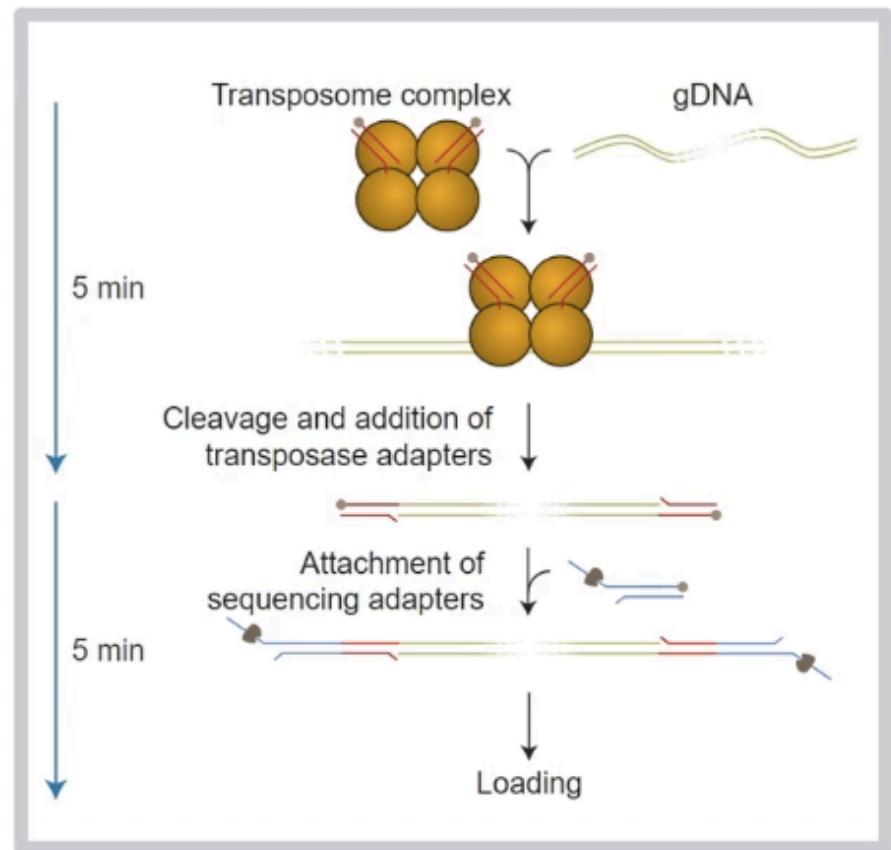
1 contig of the 4.6Mb chromosome of *E. coli*
obtained with just the 7 longest reads

SIZE OF SEQUENCED DNA FRAGMENTS

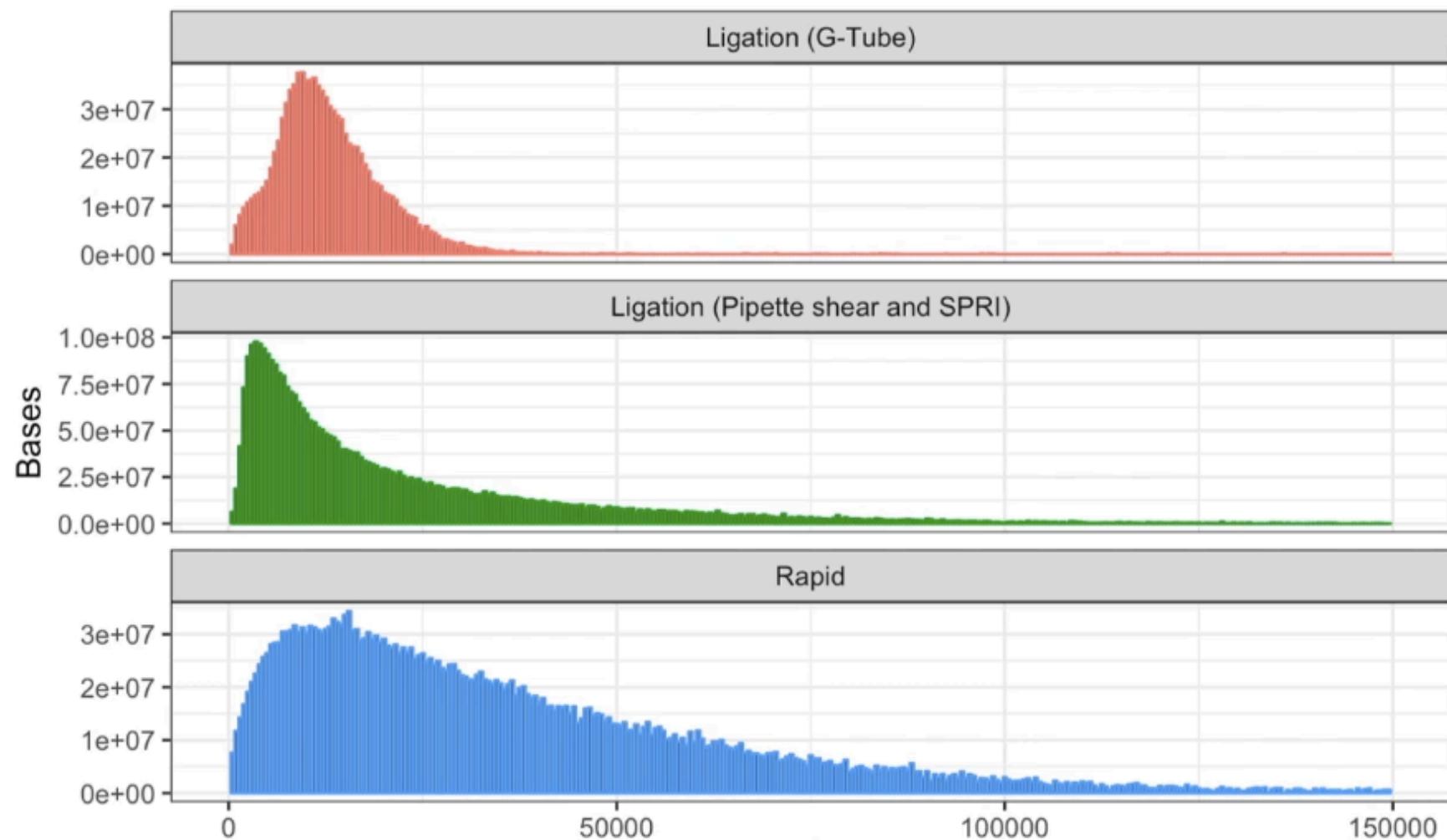
Ligation method



Transposase method



SIZE OF SEQUENCED DNA FRAGMENTS



Josh Quick, Nick Loman

see John Tyson's video (ONT website)

HYBRID GENOME ASSEMBLY : NANOPORE AND ILLUMINA DATA

Acinetobacter baylyi (data from Oxford Nanopore)

Assemblies	Illumina only	Illumina + MinION
Input Coverage	50X	13X
# contigs	20	1
Assembly size (Mb)	3.59	3.62
N90 size (Kb)	326	3 621
NA75 size (Kb)	194	1 002
Genome fraction (%)	99.73	99.997
# misassemblies	4	2
# local misassemblies	3	4
# mismatches per 100 Kb	6.49	3.11
# indels per 100 Kb	0.33	0.14

Jan. 2018

Nanopore sequencing and assembly of a human genome with ultra-long reads

Miten Jain^{1,13}, Sergey Koren^{2,13}, Karen H Miga^{1,13}, Josh Quick^{3,13}, Arthur C Rand^{1,13}, Thomas A Sasani^{4,5,13}, John R Tyson^{6,13}, Andrew D Beggs⁷, Alexander T Dilthey², Ian T Fiddes¹, Sunir Malla⁸, Hannah Marriott⁸, Tom Nieto⁷, Justin O'Grady⁹, Hugh E Olsen¹, Brent S Pedersen^{4,5}, Arang Rhie², Hollian Richardson⁹, Aaron R Quinlan^{4,5,10}, Terrance P Snutch⁶, Louise Tee⁷, Benedict Paten¹, Adam M Phillippy², Jared T Simpson^{11,12}, Nicholas J Loman³ & Matthew Loose⁸

eserved.

Using nanopore reads alone assembly of a human genome :

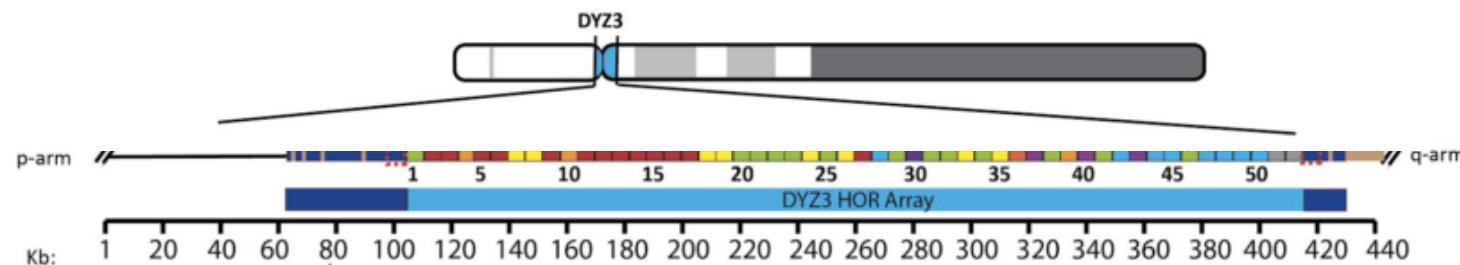
- NG50 contig size of ~6.4 Mb
- covers >85% of the reference
- 99.88% accuracy
- MHC locus on a single contig, phased over its full length
- closure of 12 large (>50 kb) gaps in the reference human genome

ASSEMBLY OF A HUMAN Y CENTROMERE

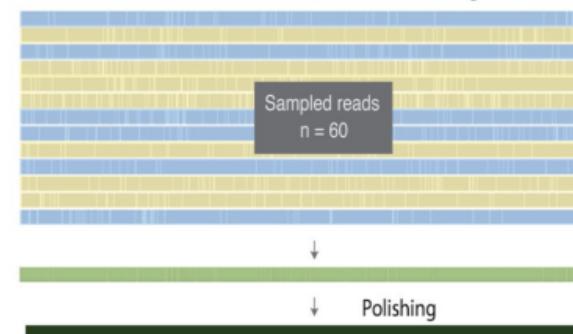
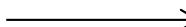
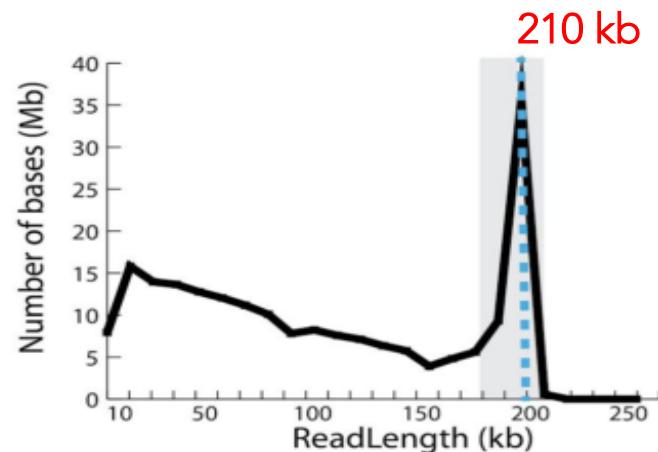
(Jain et al., *bioRxiv*, 2017)

300 kb array of 5.8 kb sequence repeated in an uninterrupted head-to-tail orientation

To date, no technology has been capable of sequencing centromeres due to requirement for extremely high-quality long reads



9 BACs
100kb to 210kb



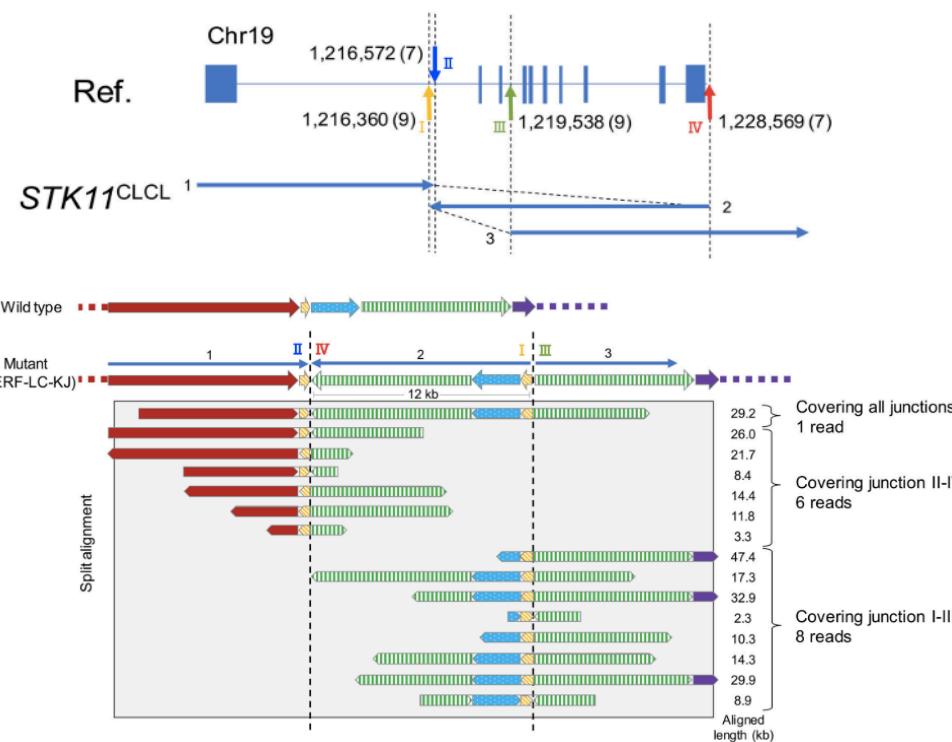
FIRST COMPLETE SEQUENCE OF A HUMAN CENTROMERE

GENOME SEQUENCING

Long-read sequencing for non-small-cell lung cancer genomes
Sakamoto et al. Genome Research, Sept. 2020

Nanopore sequencing of **cancer cell lines** (PromethION)

- Maximum length : 0.99 Mb
- N50 : 32 kb
- Average mapped reads : 14 kb



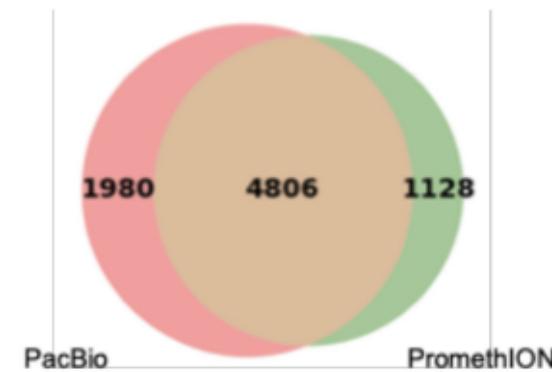
Biological relevance of SV further revealed by :

- epigenome,
- transcriptome,
- protein analyses

Sequencing of clinical tumor samples

- Structural aberrations also found in **clinical lung adenocarcinoma specimens**

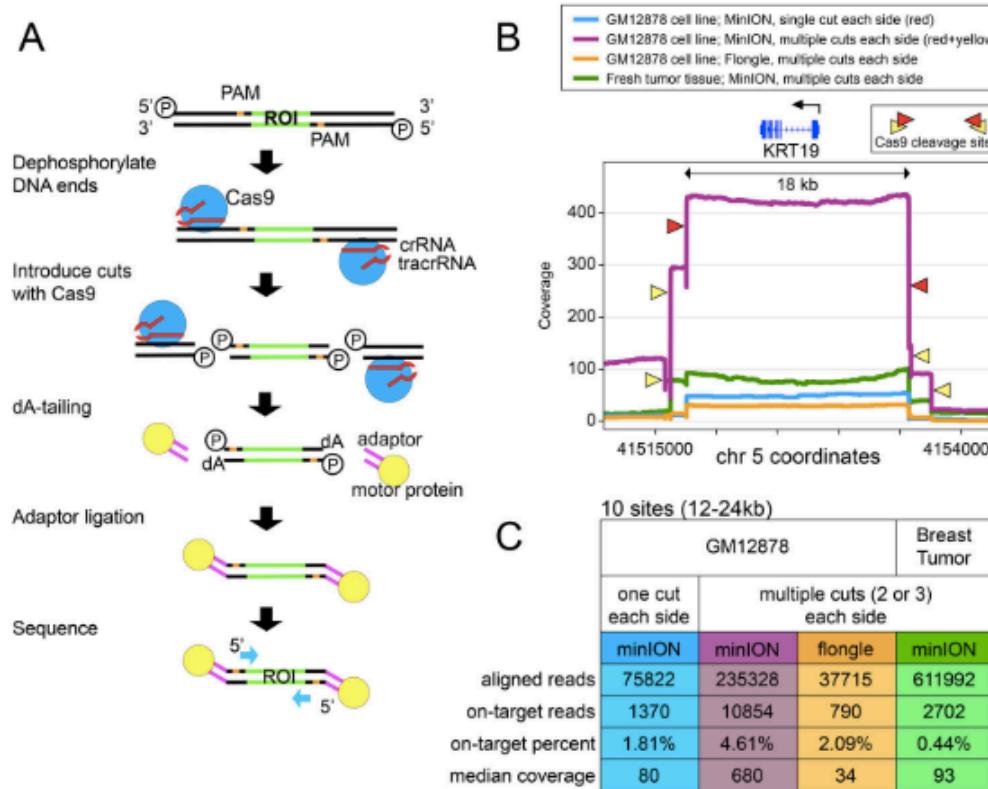
Structural variants : comparison with PacBio sequencing



"These results indicated that neither the PacBio nor the PromethION platform is currently perfect; therefore, they should be used to complement each other."

GENOME SEQUENCING : TARGETED SEQUENCING

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



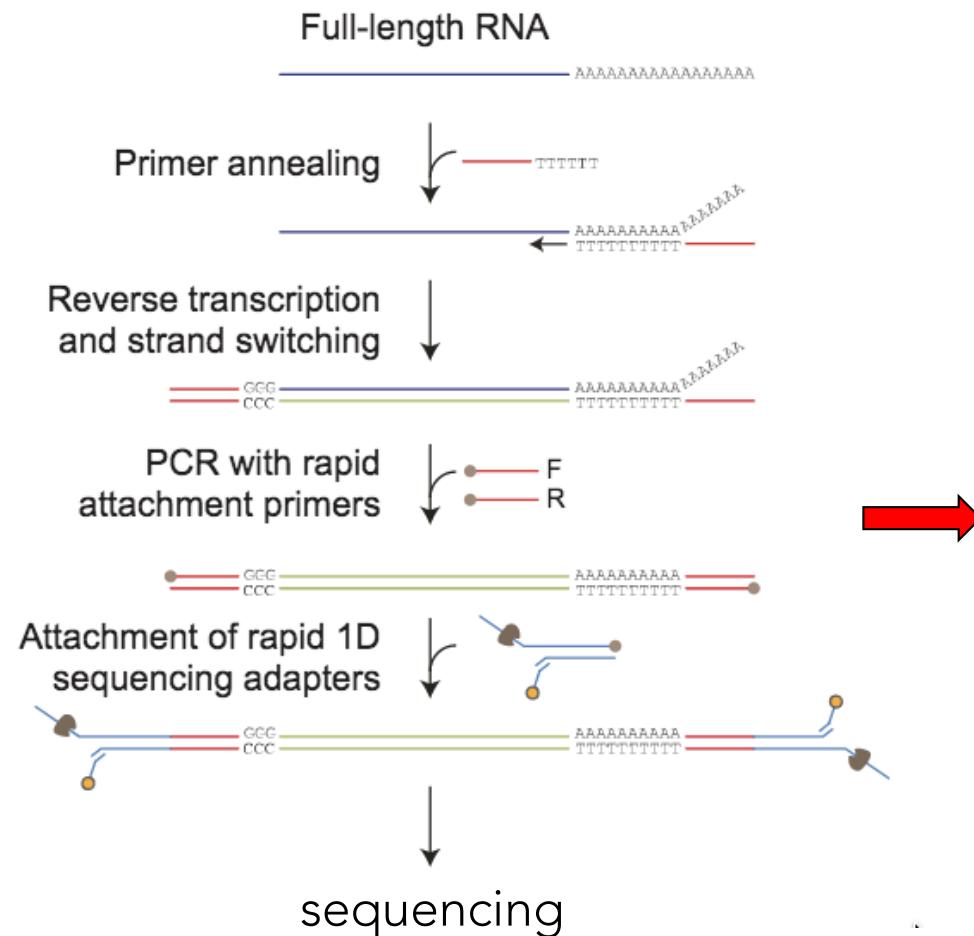
nCATS = nanopore Cas9-targeted sequencing : enrichment strategy using targeted cleavage of DNA to ligate adapters for nanopore

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.

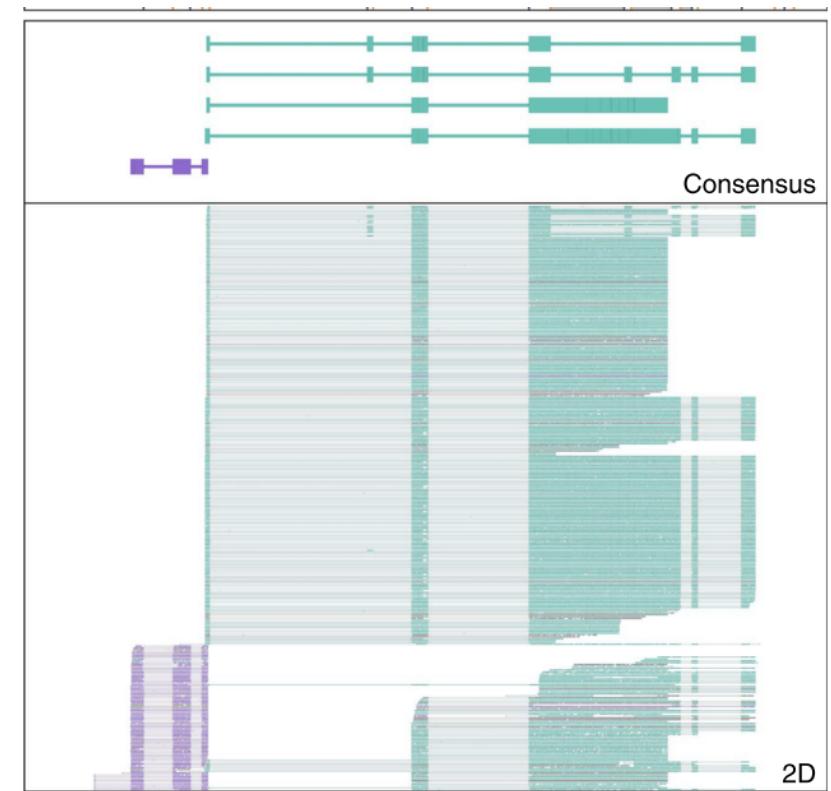
cDNA SEQUENCING

Library preparation



Detection of splice variants in surface receptor of B cells

(Byrne et al. Nat. Comm. 2017)



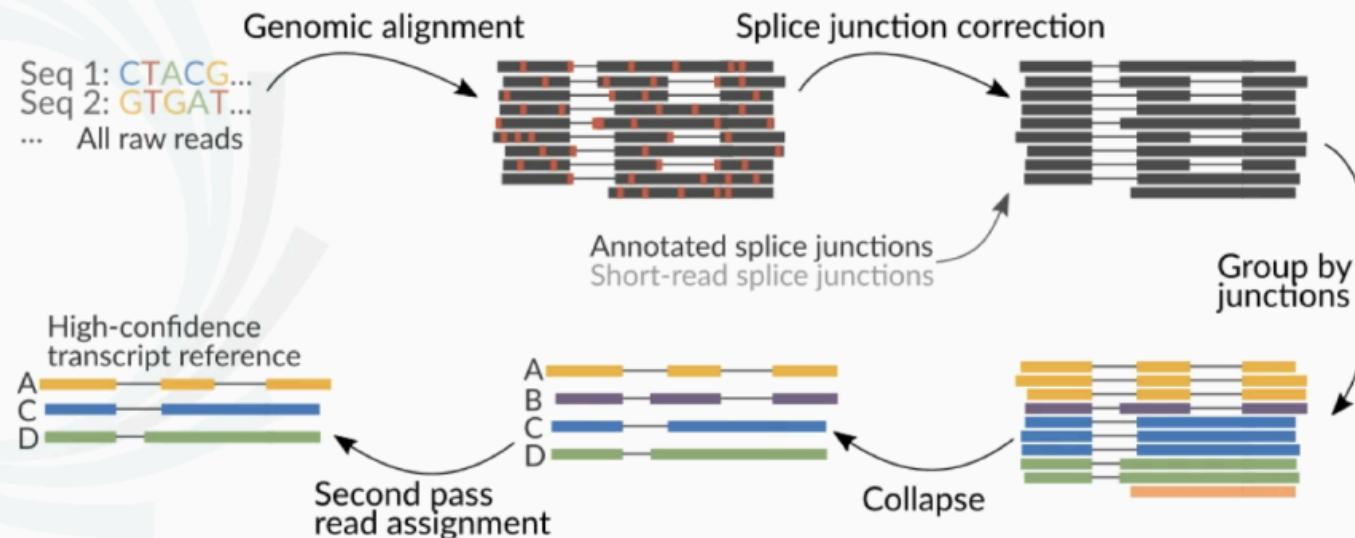
- Splice alignment uneasy due to high (5-10%) error rate
- Reads are frequently truncated from 5' end

CHALLENGES OF NANOPORE TRANSCRIPTOME ANALYSIS

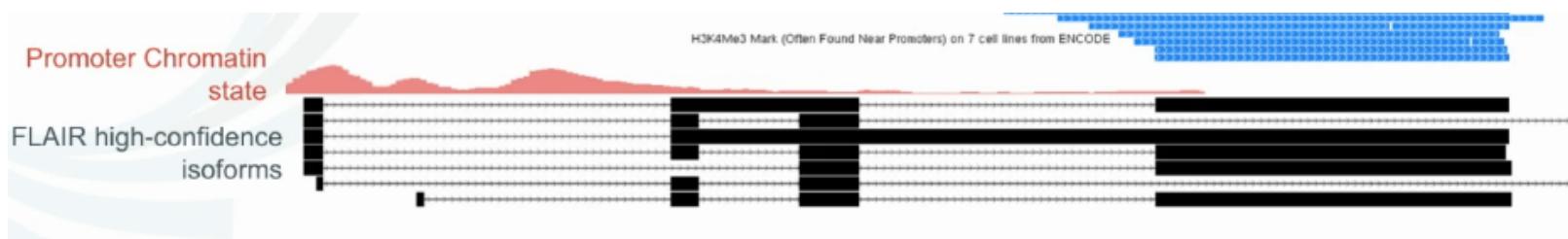
FLAIR : a pipeline for splicing isoform determination

Tang et al. *bioRxiv* 2018

FLAIR CONTAINS TWO ALIGNMENT STEPS TO PRODUCE A HIGH-CONFIDENCE TRANSCRIPT REFERENCE



FLAIR incorporates promoter chromatin states to distinguish 5' truncations from true novel start sites

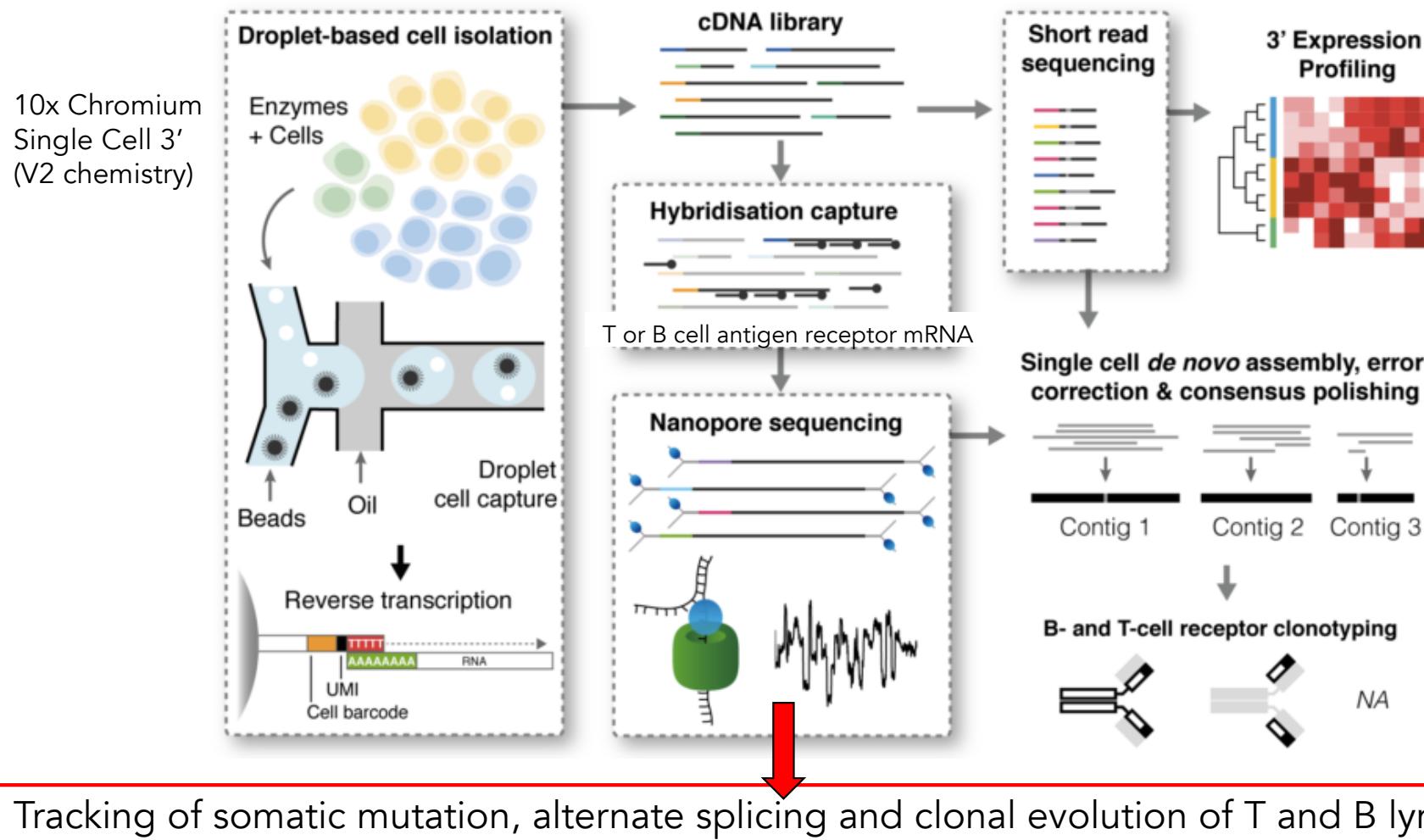


cDNA from chronic lymphocytic leukemia (CLL)

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

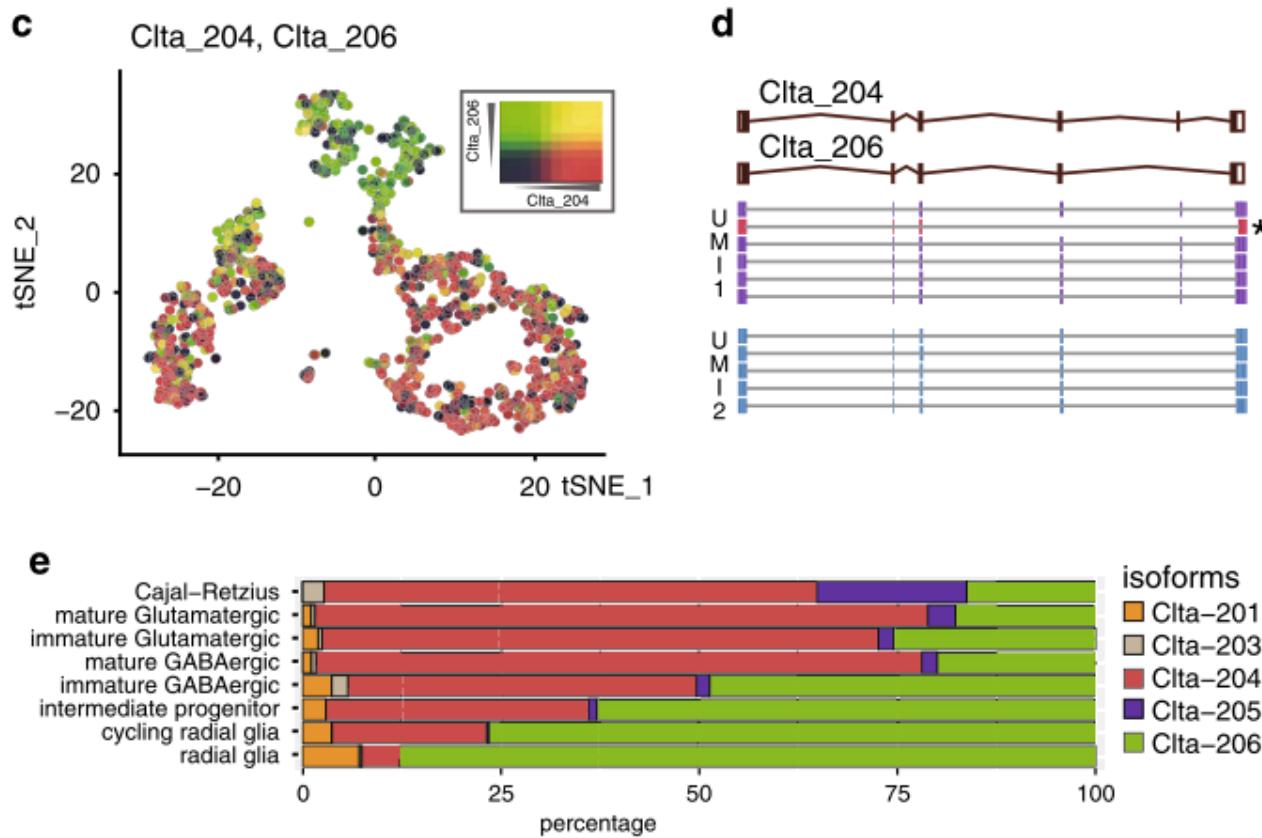


NANOPORE and SINGLE CELL cDNA SEQUENCING

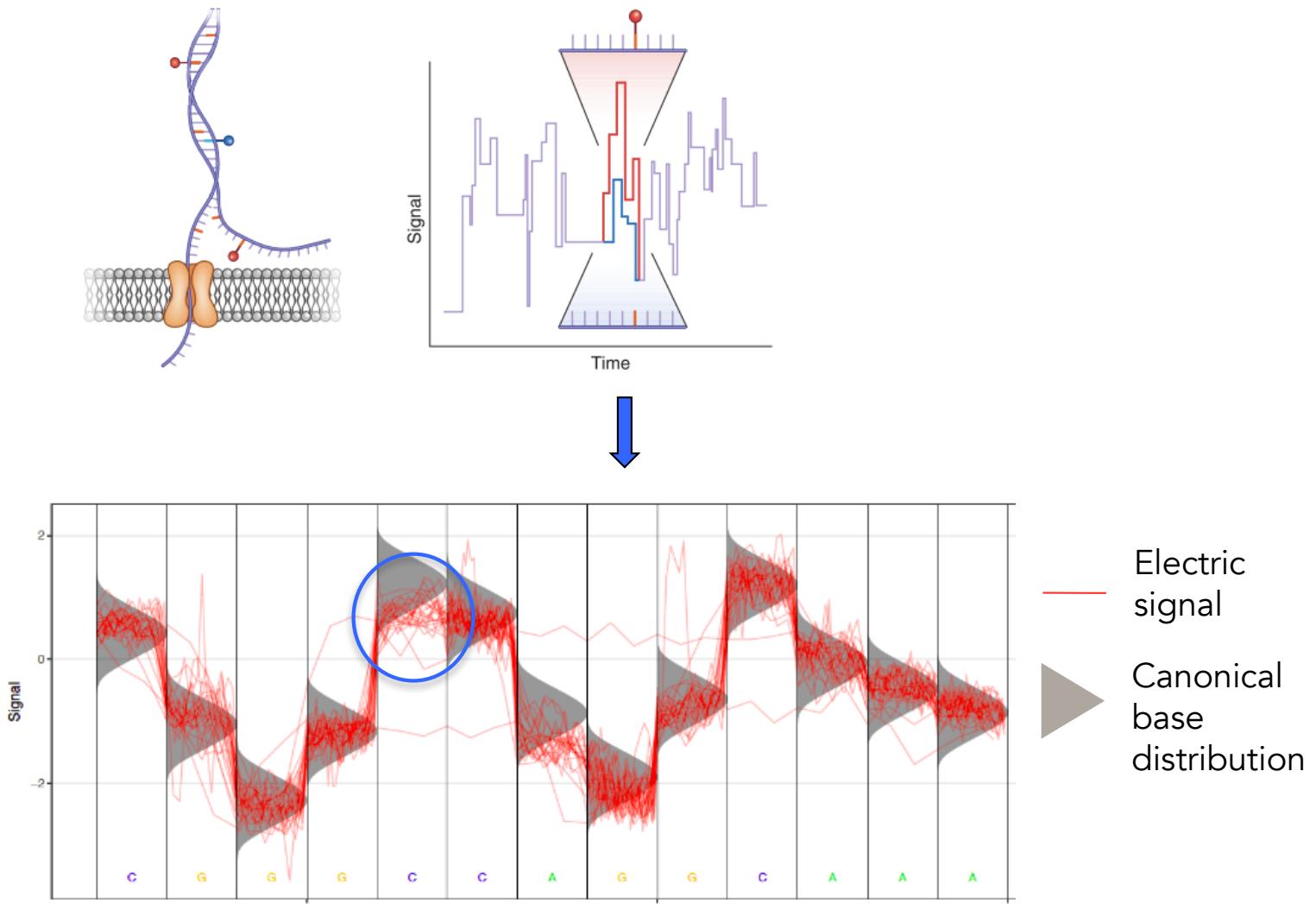
High-throughput targeted long-read single cell sequencing reveals the clonal and transcriptional landscape of lymphocytes
Lebrigand et al., *Nature Communications*, 2020

ScNaUmi-seq : Single-cell Nanopore sequencing with UMIs (10x Genomics Chromium system)

- High accuracy cell BC and UMI assignment
- Analysis of splicing and sequence variation at the single-cell level

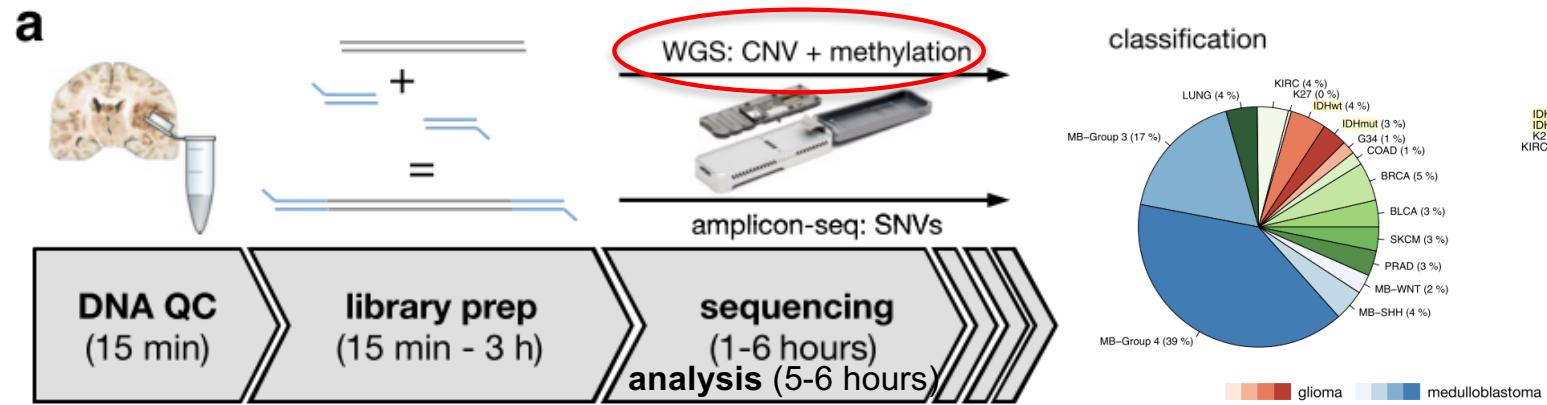


DETECTION OF MODIFIED BASES



DETECTION OF MODIFIED BASES IN CANCER GENOMES

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



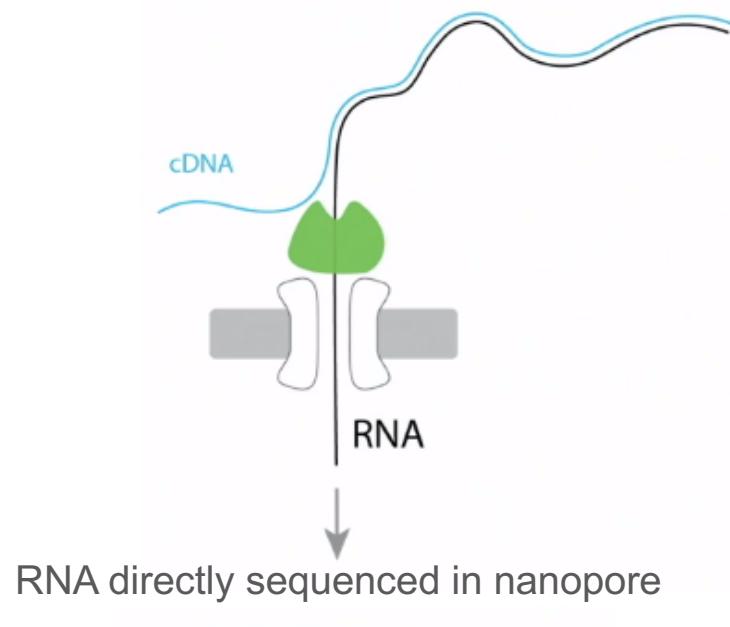
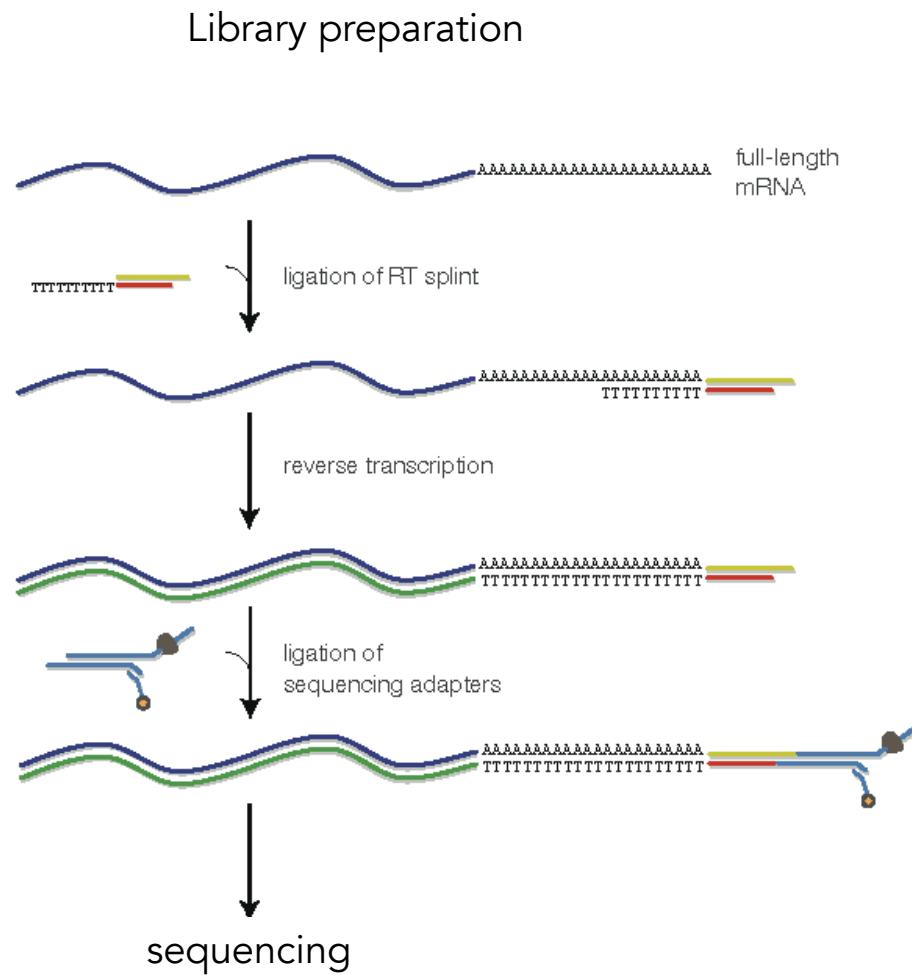
Same-day detection of :

- structural variants
- point mutations
- methylation profiling

Single device with negligible capital cost :

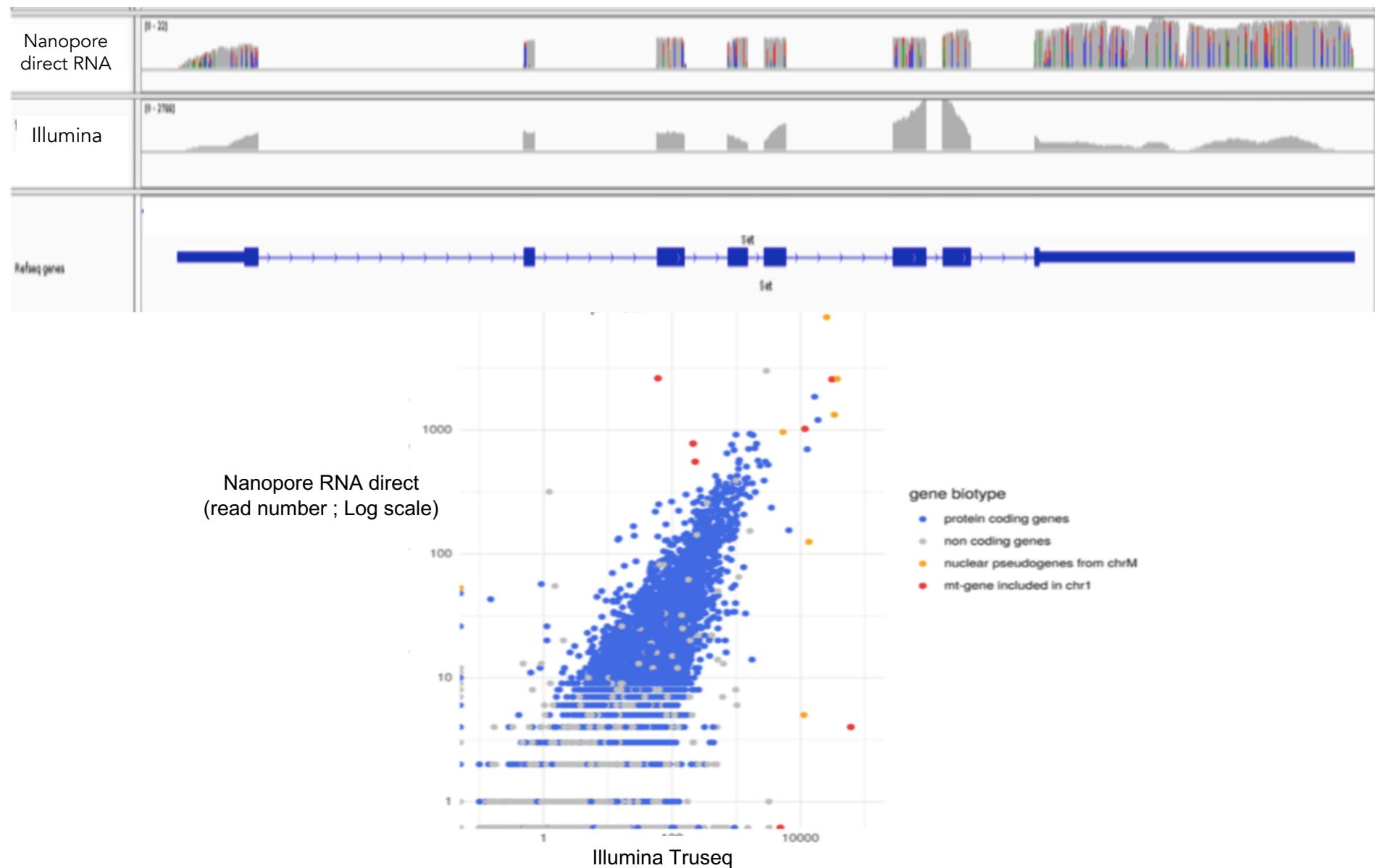
- outperforms hybridization-based and current sequencing technologies
- makes precision medicine possible for every cancer patient

DIRECT RNA SEQUENCING



- No PCR bias
- Quantitative

DIRECT RNA SEQUENCING vs ILLUMINA

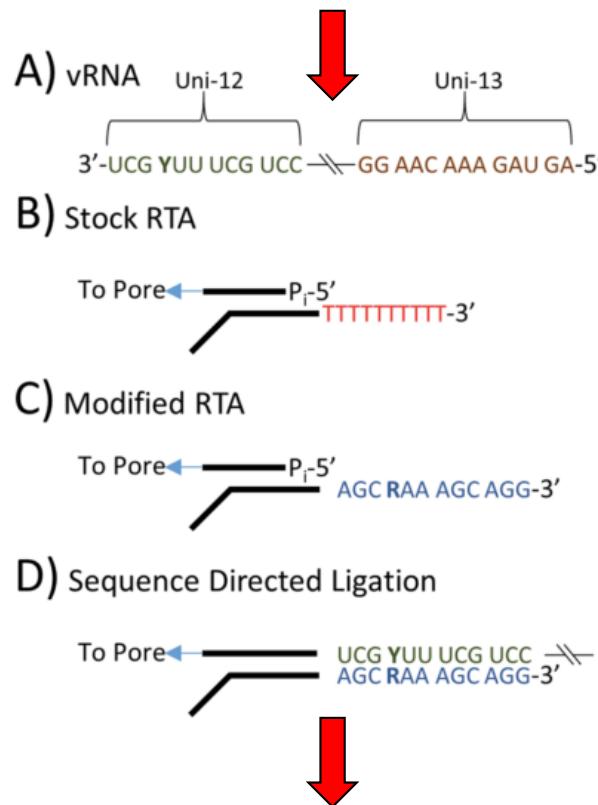


DIRECT RNA SEQUENCING: INFLUENZA VIRUS GENOME

Direct RNA Sequencing of the complete Influenza A Virus Genome
Keller et al. *Scientific Reports*, Sept. 2018

For the first time a complete genome of an RNA virus sequenced in its original form

Influenza A viruses are negative-sense segmented RNA viruses (8 segments)

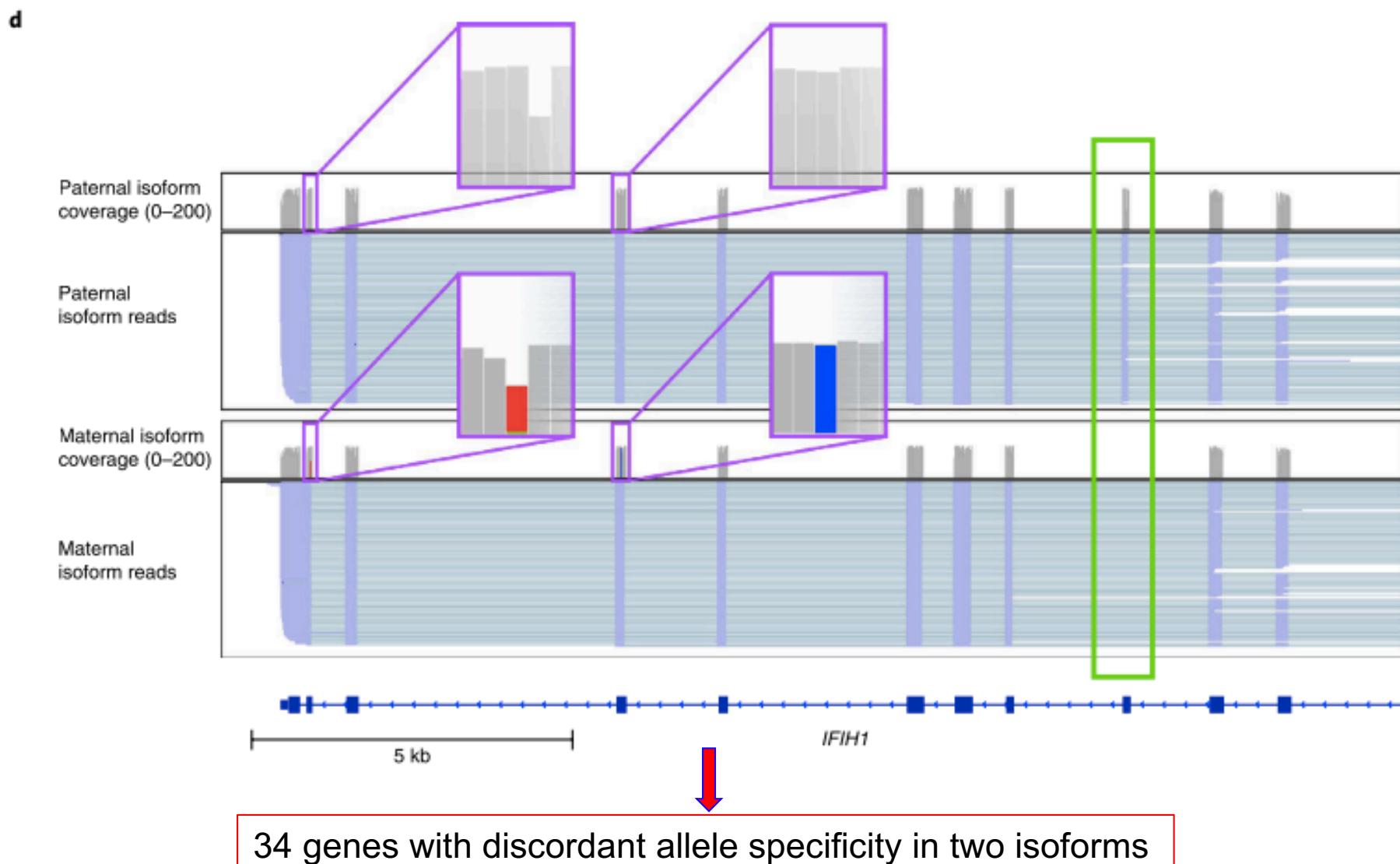


sequencing of complete genome with 100% nucleotide coverage, 99% consensus identity

Potential to identify and quantify splice variants, base modifications
not practically measurable with current methods

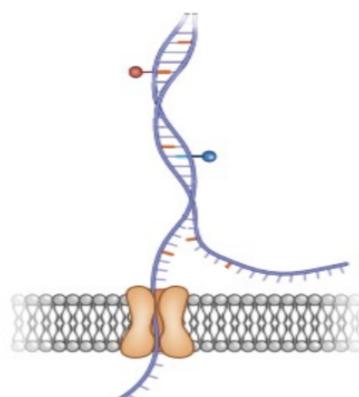
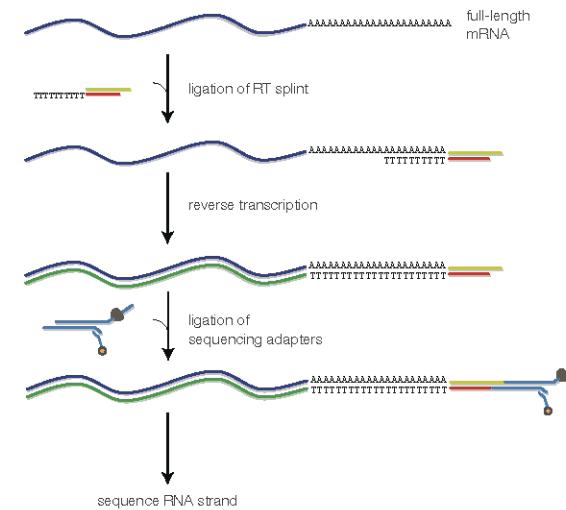
DIRECT RNA SEQUENCING: TRANSCRIPT HAPLOTYPE

Nanopore native RNA sequencing of a human transcriptome

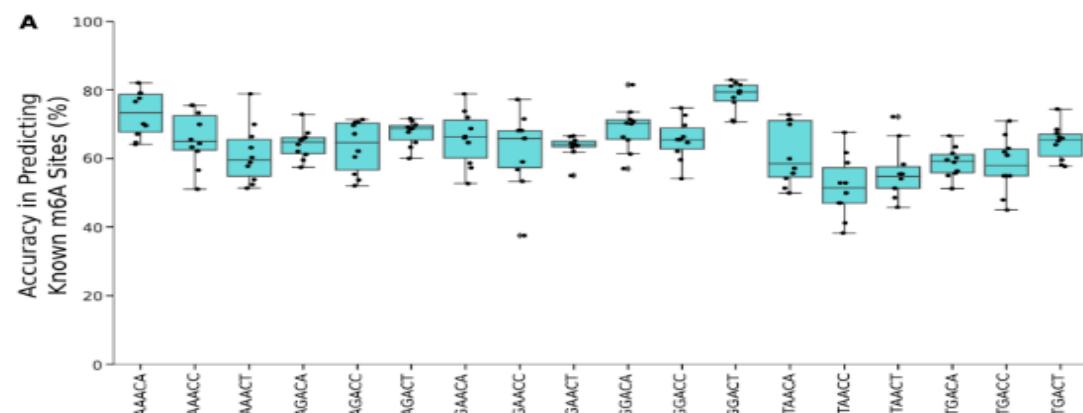


— DIRECT RNA SEQUENCING: DETECTION OF m6A

Library preparation



Lorenz et al. *RNA* 2019



Detection of m6A with Nanopolish :

Different detection efficiency in different sites: 45% to 82%

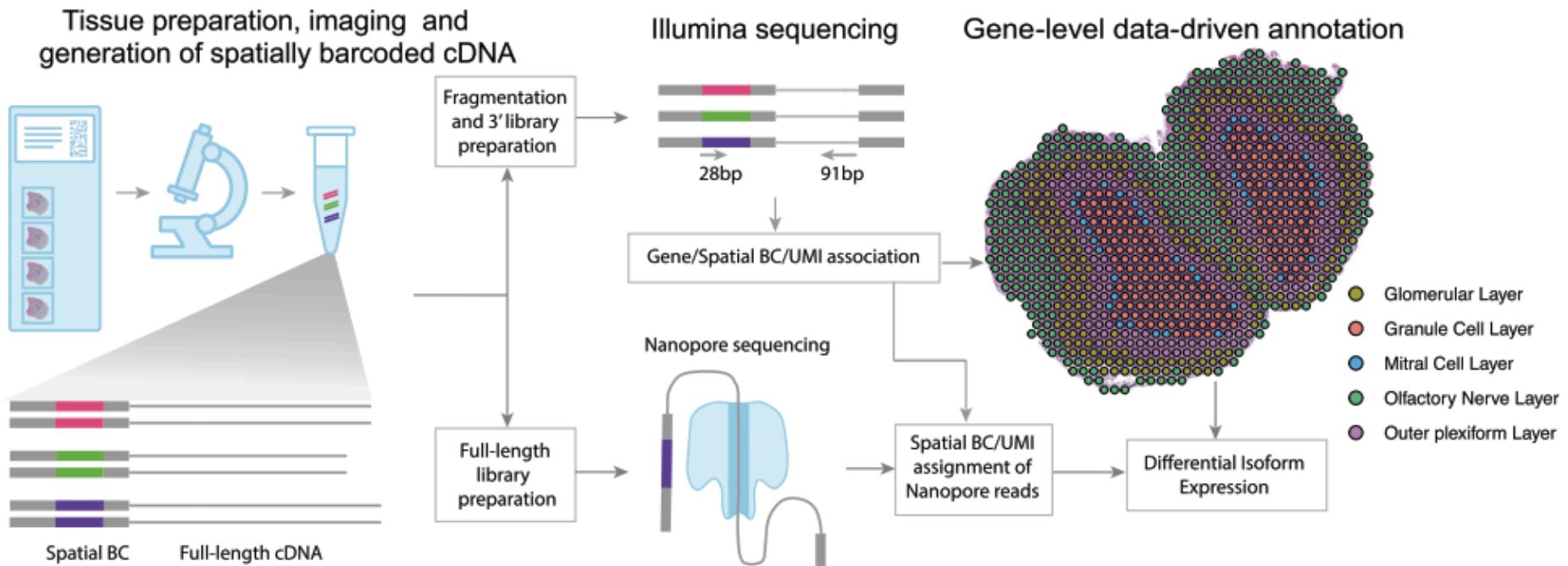
Context dependent detection efficiency

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)



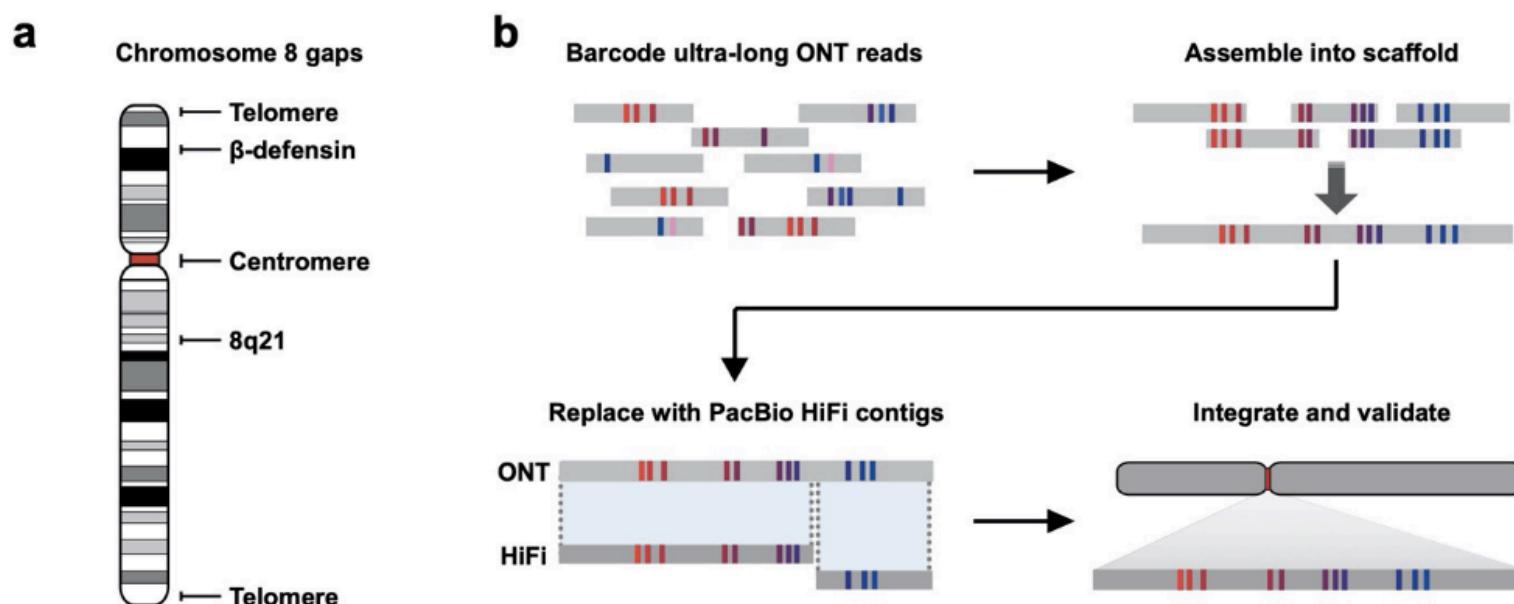
The structure, function, and evolution of a complete human chromosome 8

Logsdon et al., *bioRxiv*, Sept 2020

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..
- 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%.



???



PacBio



Nanopore



Illumina