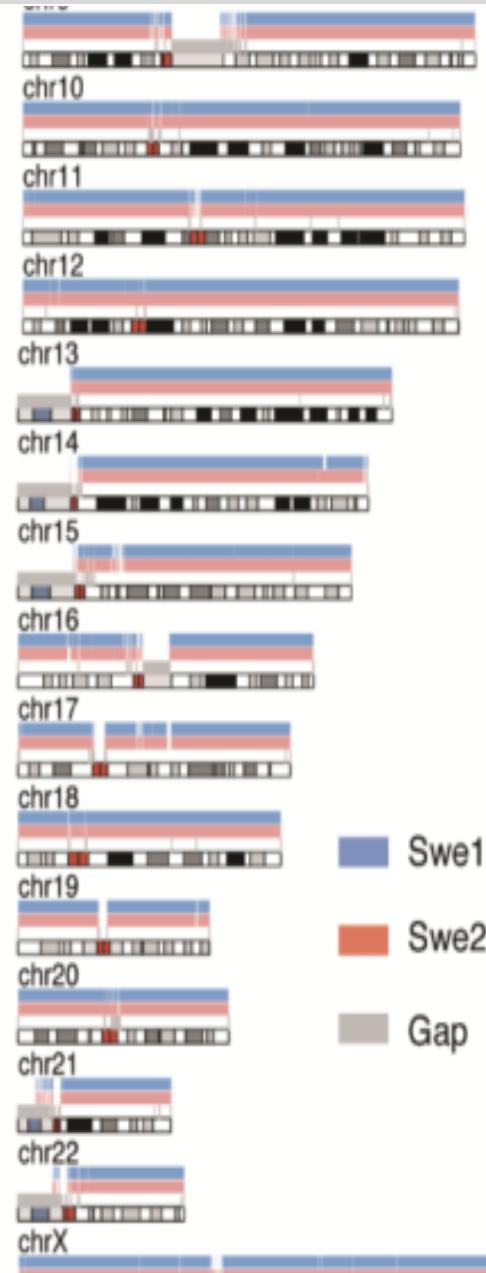


# PacBio GENOME ASSEMBLY

De novo Assembly of Two Swedish Genomes Reveals Missing Segments from the Human Reference (hg38)

Ameur et al. Genes, 2018

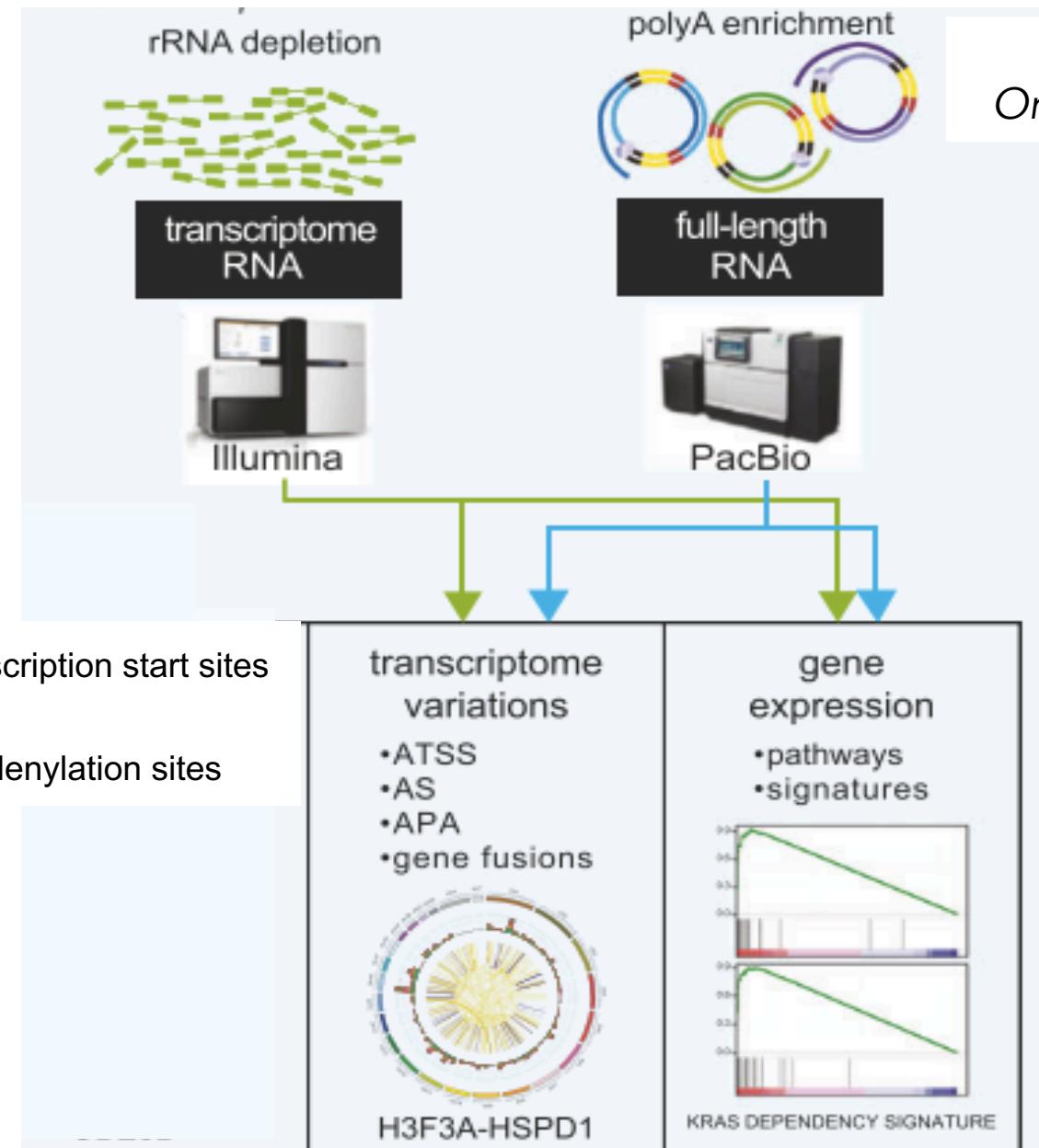
- 10 Mb of the 2 genomes are absent from hg38 reference
- 1 Mb are assigned to chr. Y
- 6 Mb are shared with a Chinese personal genome
- Inclusion of these sequences in GRCh38 genome radically improves alignment and variant calling from short-read data :
- re-analysis :
  - yields > 75,000 putative novel single nucleotide variants (SNVs)
  - removes > 10,000 false positive SNV calls per individual
- It becomes possible to represent specific population groups by assembly of representative genomes from different populations.



# PacBio cDNA SEQUENCING

## Hybrid full-length transcriptome in metastatic ovarian cancer

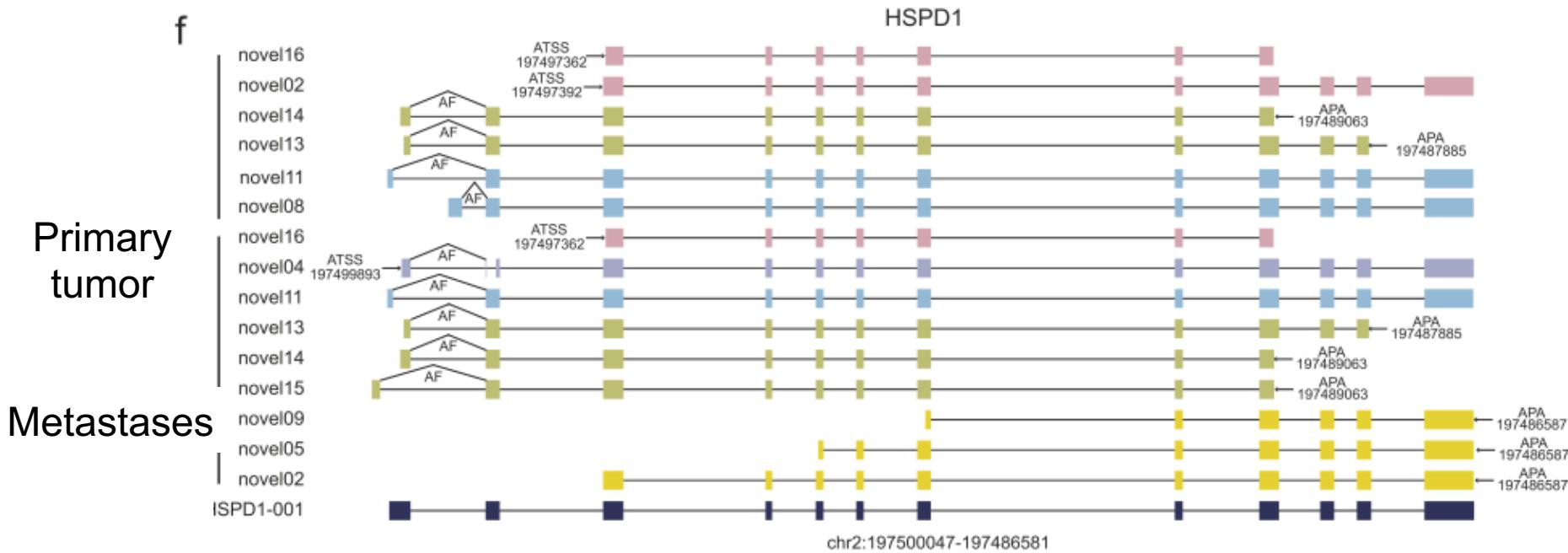
Jing et al.  
Oncogene 2019



# PacBio cDNA SEQUENCING

## Hybrid full-length transcriptome in metastatic ovarian cancer

Jing et al. *Oncogene* 2019



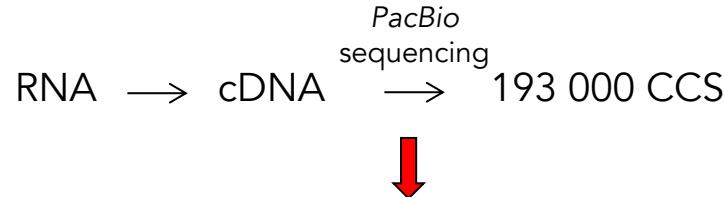
Long-read full-length transcriptome analysis

- improves molecular diagnostic
- reveals novel therapeutic vulnerabilities

# 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

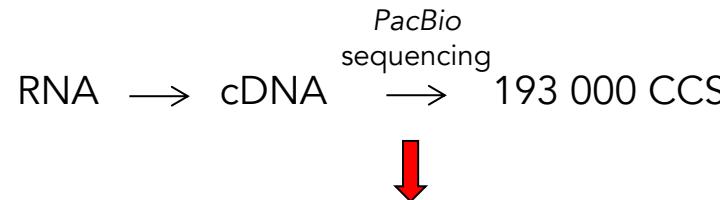


- Increased the total genomic exon length by 1.9 Mb (12.4%)
- 1609 (9.2%) new genes

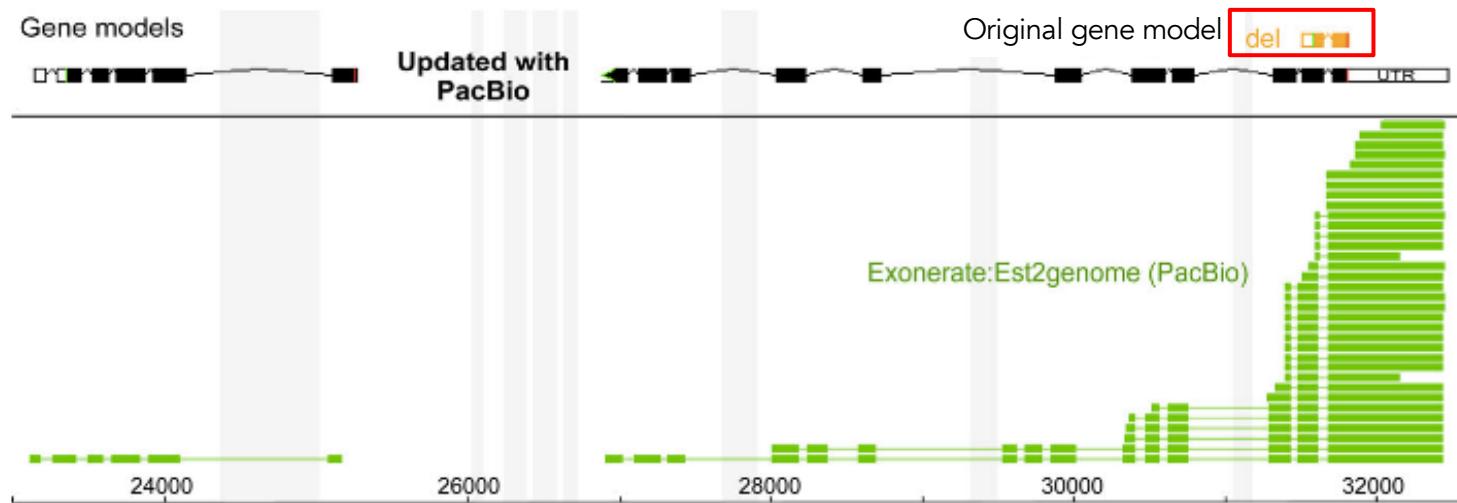
# SEQUENCING cDNA USING CIRCULAR CONSENSUS SEQUENCES

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

Magrini et al. *BMC Genomics*, 2018



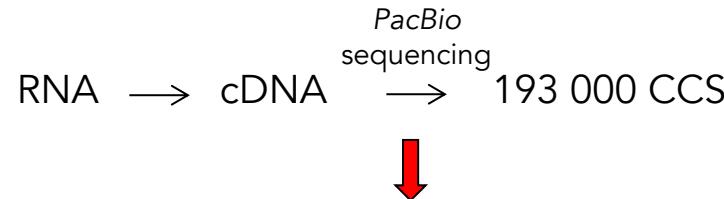
- Increased the total genomic exon length by 1.9 Mb (12.4%)
- 1609 (9.2%) new genes



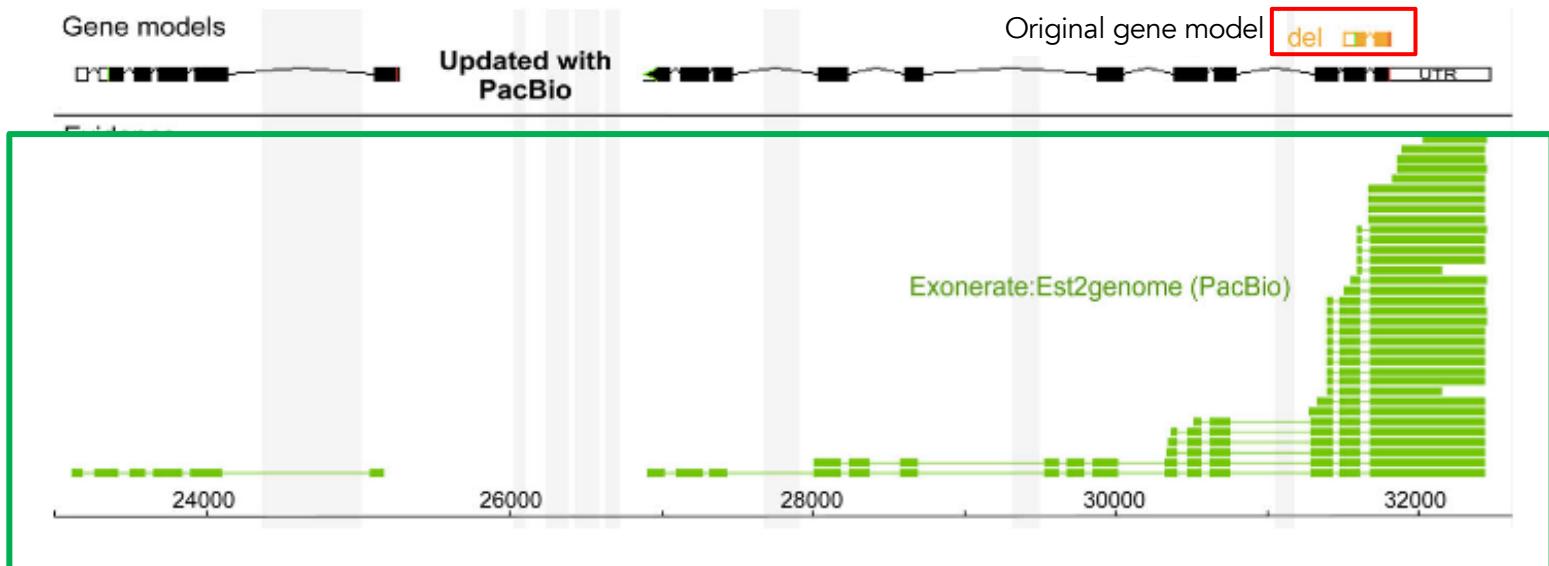
# SEQUENCING cDNA USING CIRCULAR CONSENSUS SEQUENCES

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

Magrini et al. *BMC Genomics*, 2018



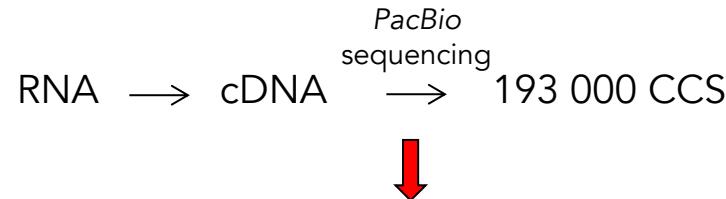
- Increased the total genomic exon length by 1.9 Mb (12.4%)
- 1609 (9.2%) new genes



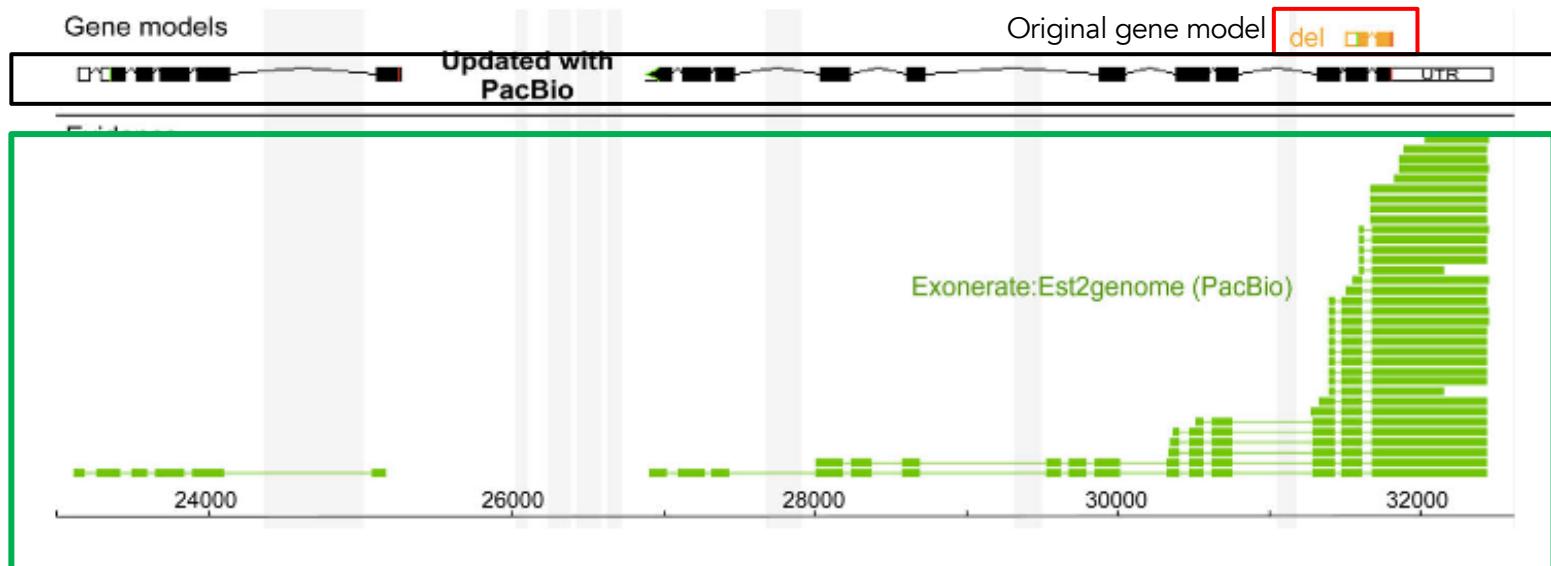
# SEQUENCING cDNA USING CIRCULAR CONSENSUS SEQUENCES

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

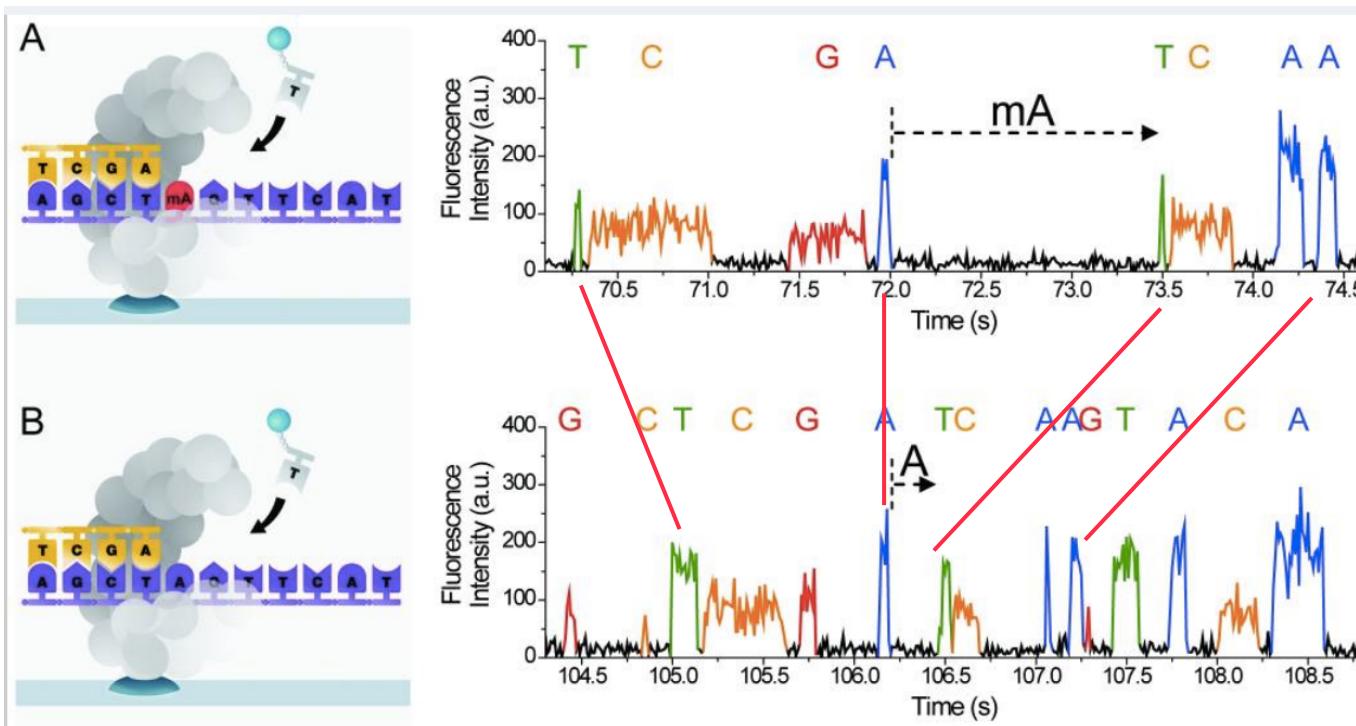
Magrini et al. *BMC Genomics*, 2018



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



# DETECTION OF MODIFIED BASES



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

Detection of 5mA and possible detection of other modified bases : 5mC, 5-hmC, etc.  
but with strong influence of sequence contexts (requires high coverage)

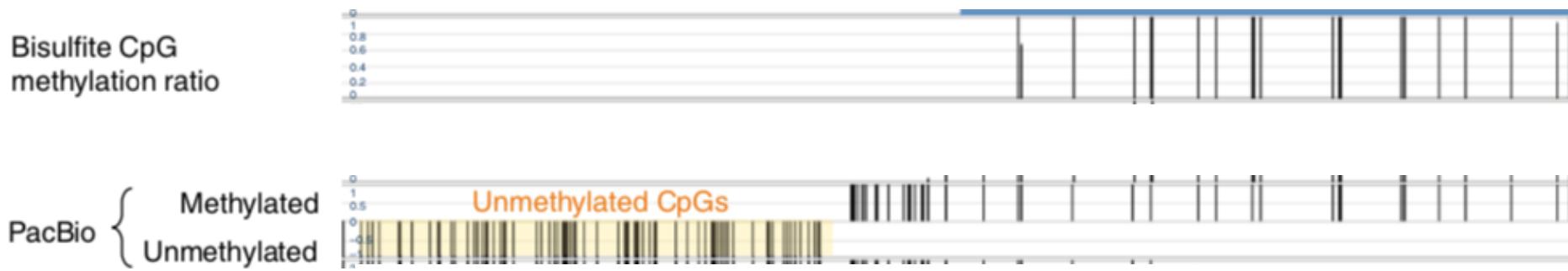
Feng et al. *PLOS Comput Biol* (2013)

# PacBio DETECTION OF MODIFIED BASES

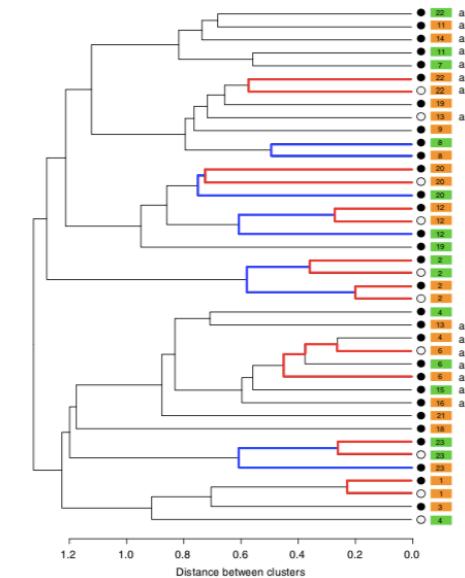
## Centromere evolution and CpG methylation during vertebrate speciation

Ichikawa et al. *Nature Communications* 2009

Long-read genome assembly of three inbred medaka strains that undergo speciation



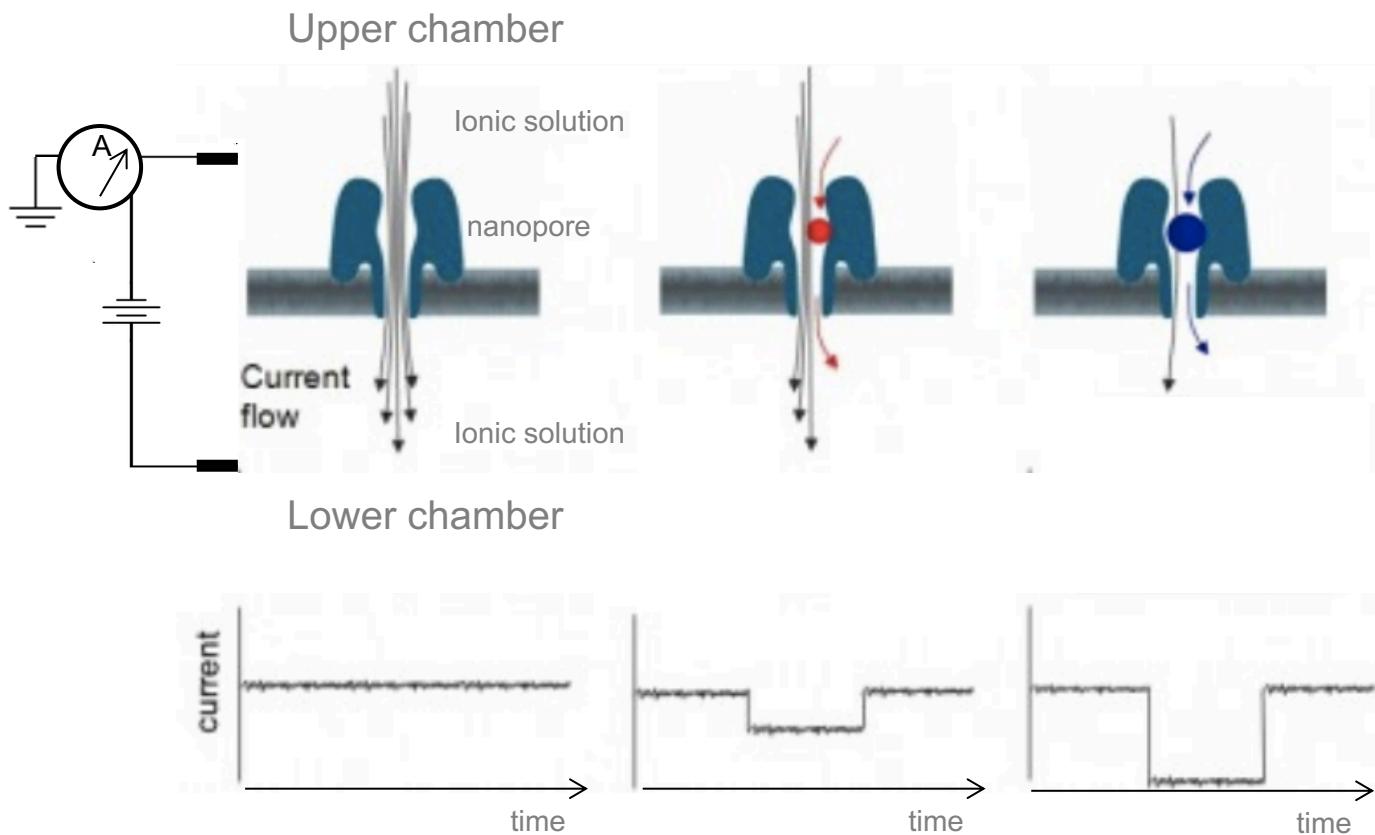
Variation of CpG methylation occurred after the divergence of two medaka strains demonstrating that centromeres accumulated epigenetic diversity during speciation



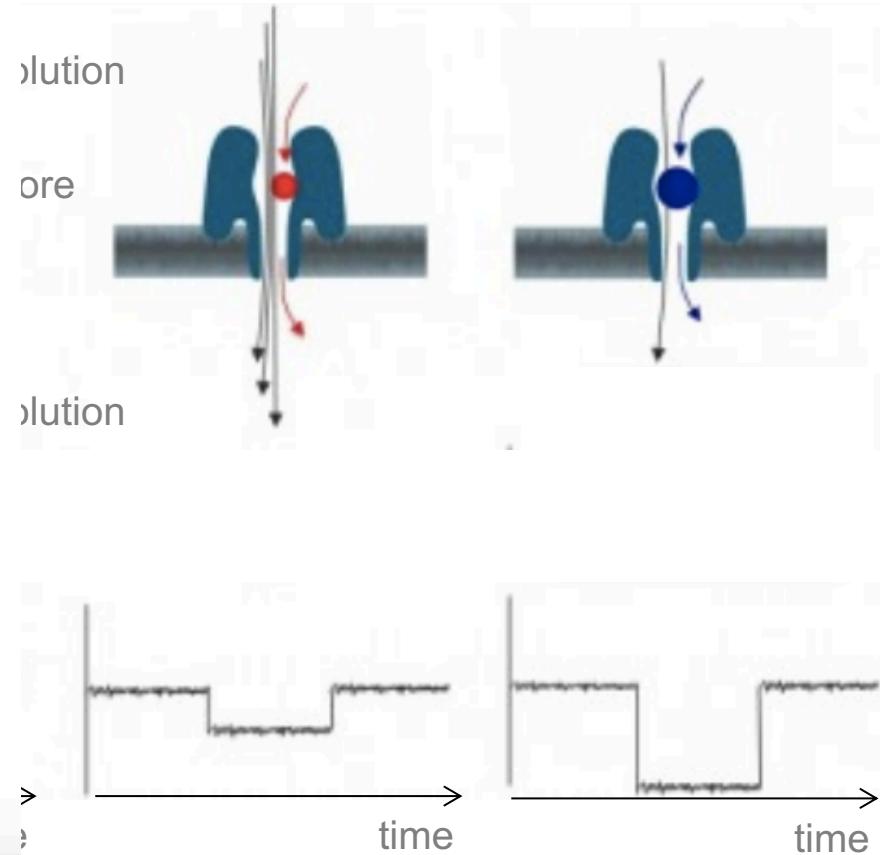
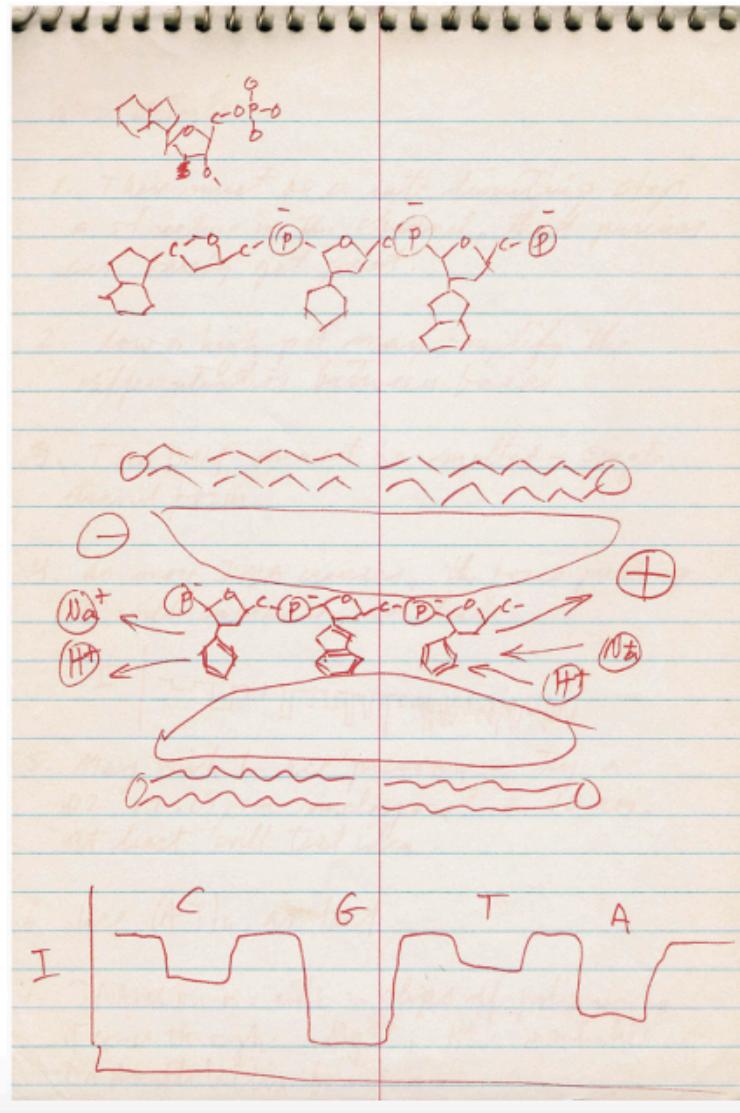
# Next Generation Sequencing



# BASIC CONCEPTS

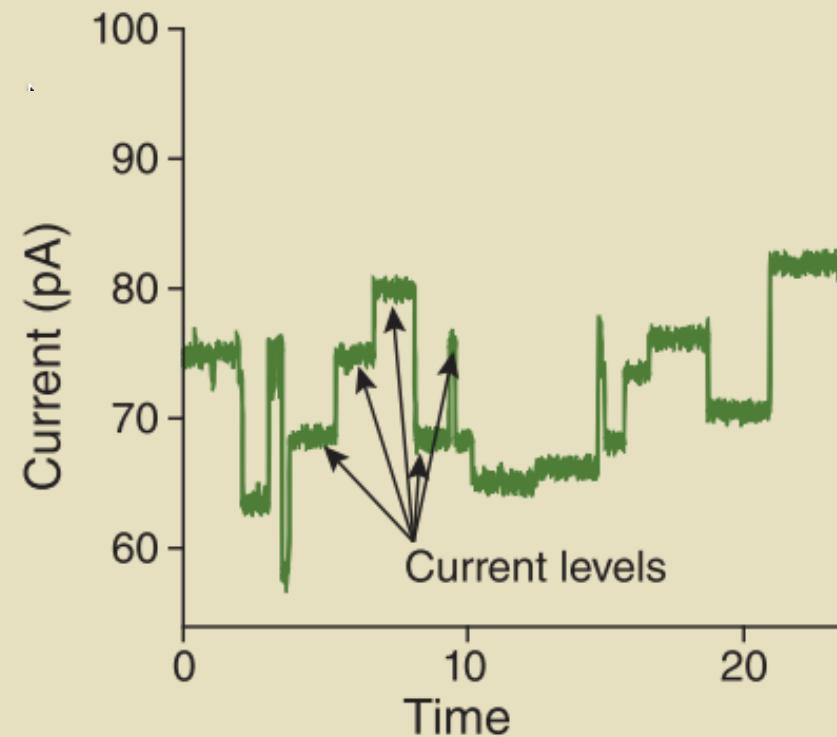
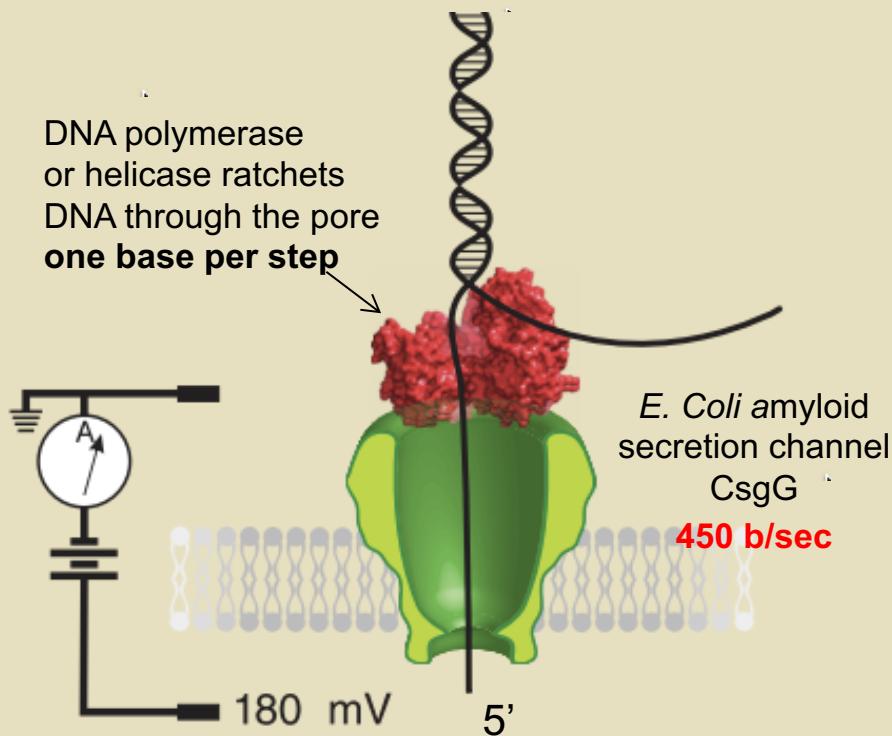


# BASIC CONCEPTS

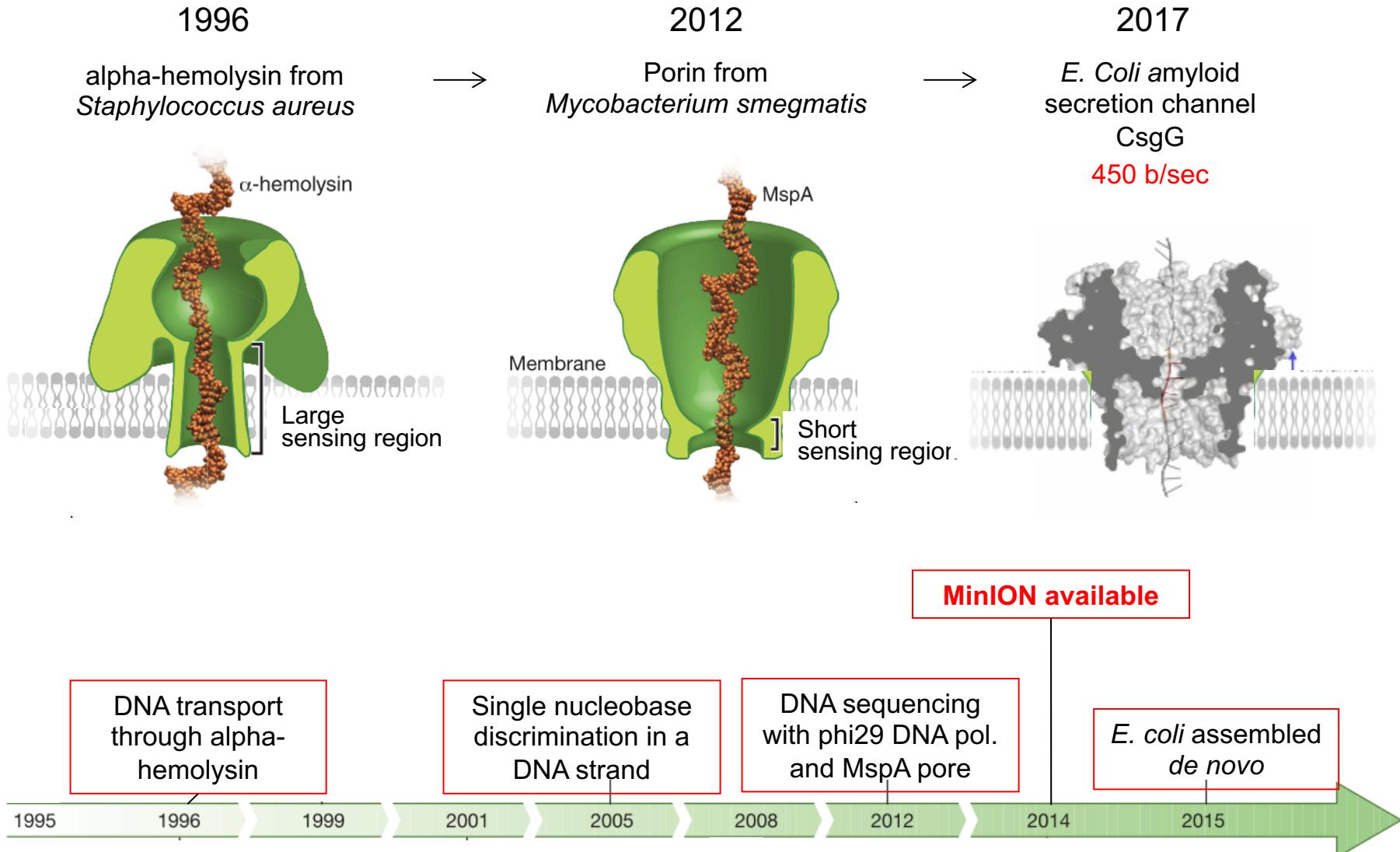


1989 David Dreamer's notebook

## BASIC CONCEPTS

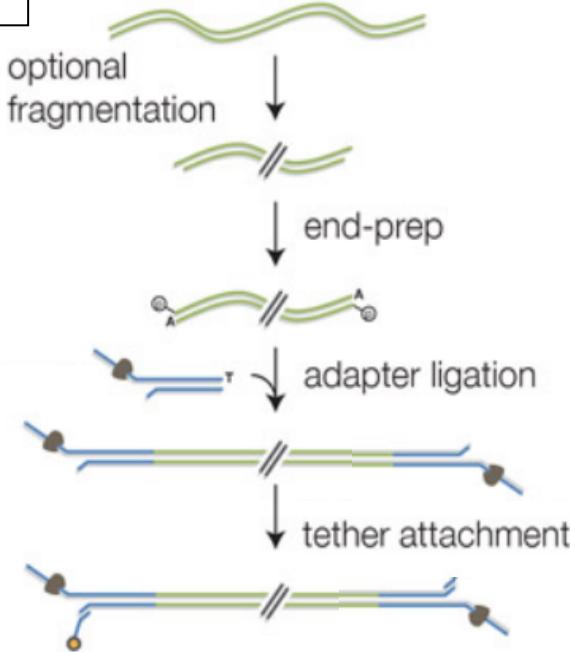


# NANOPORES USED IN THE MinION

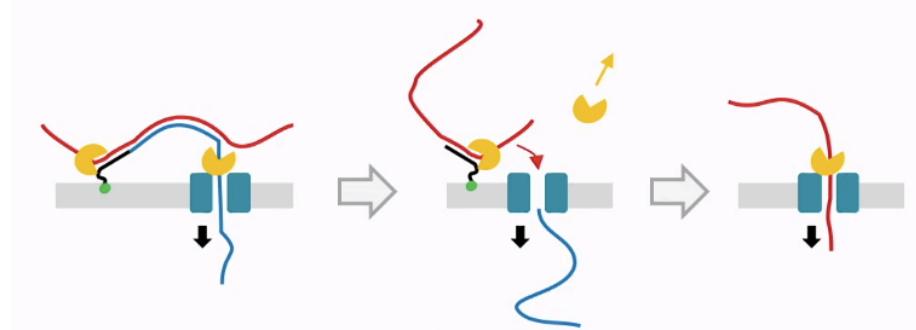
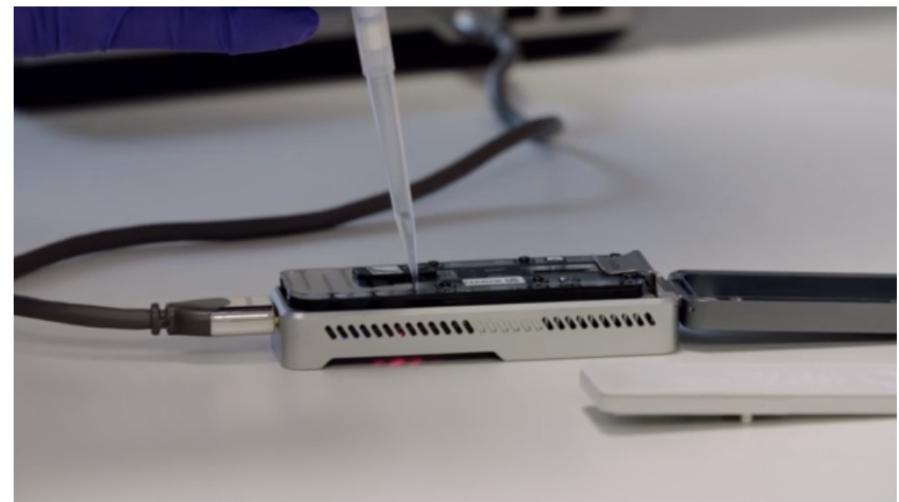


# SEQUENCING PROCESS

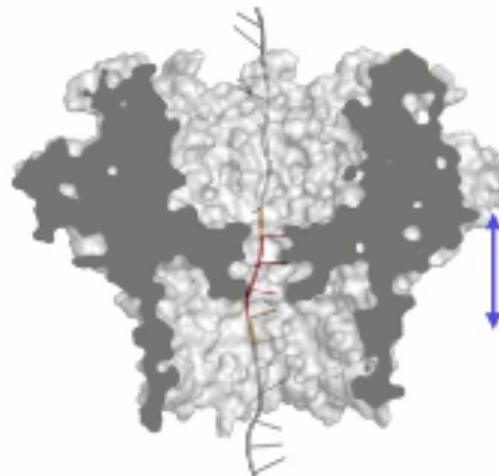
1D<sup>2</sup> Library  
(2017)



SEQUENCING

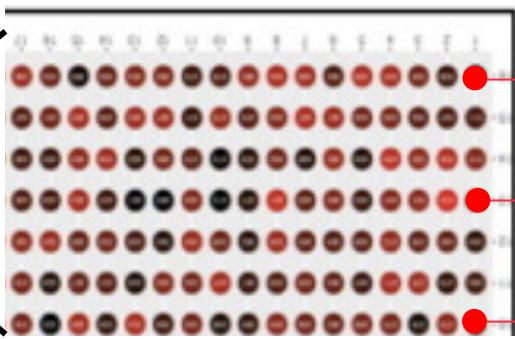


## SEQUENCING PROCESS : MinION FLOW CELL

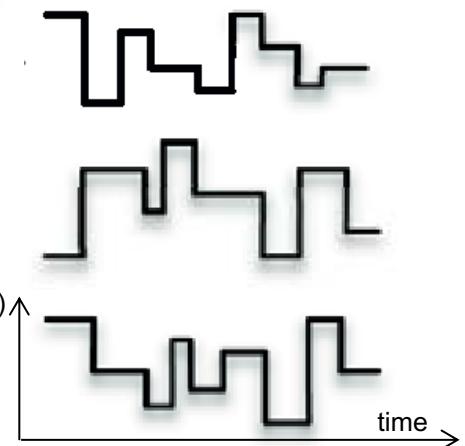


5-6 bases  
dominate the  
current signal

MinION : 512 pores

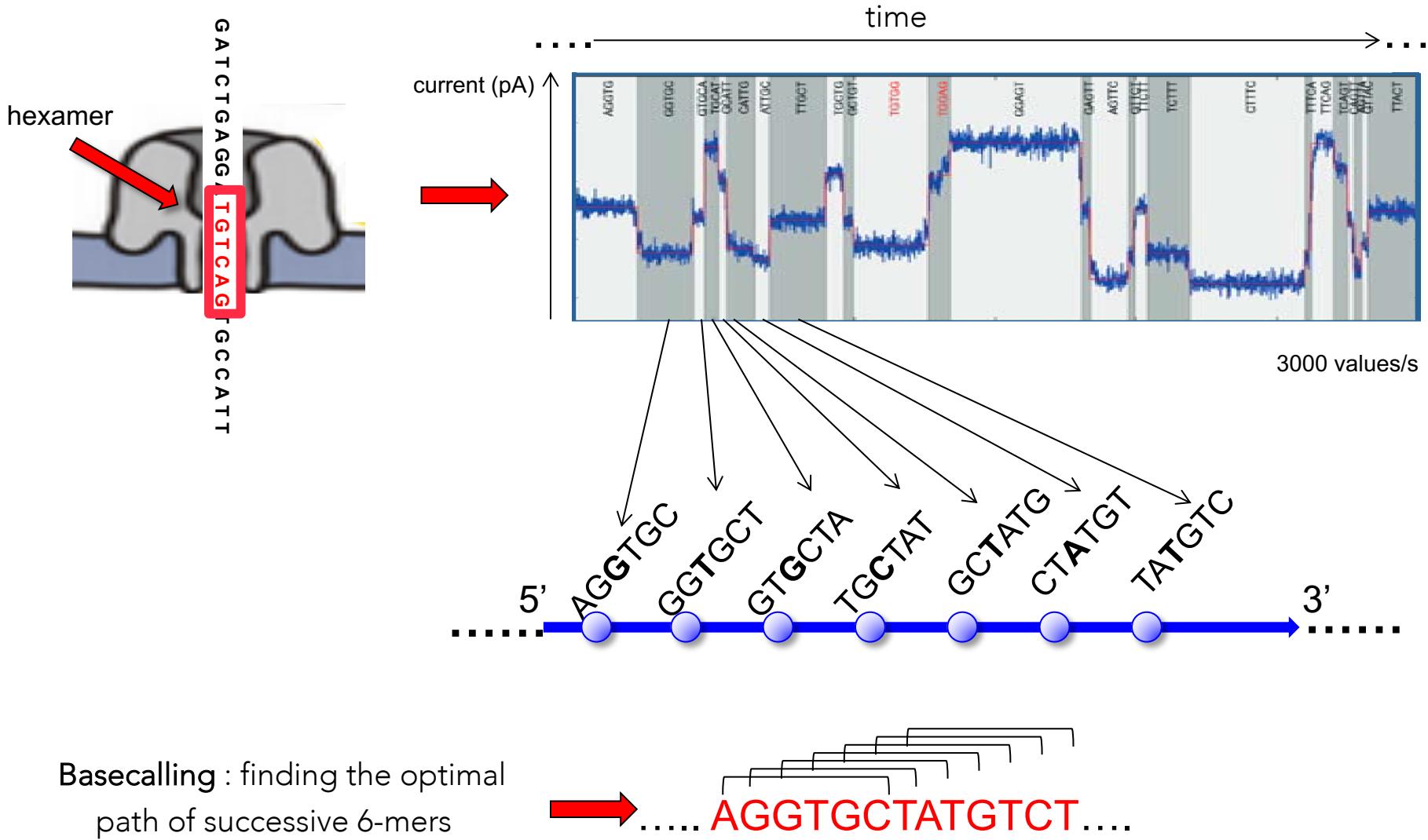


current (pA)

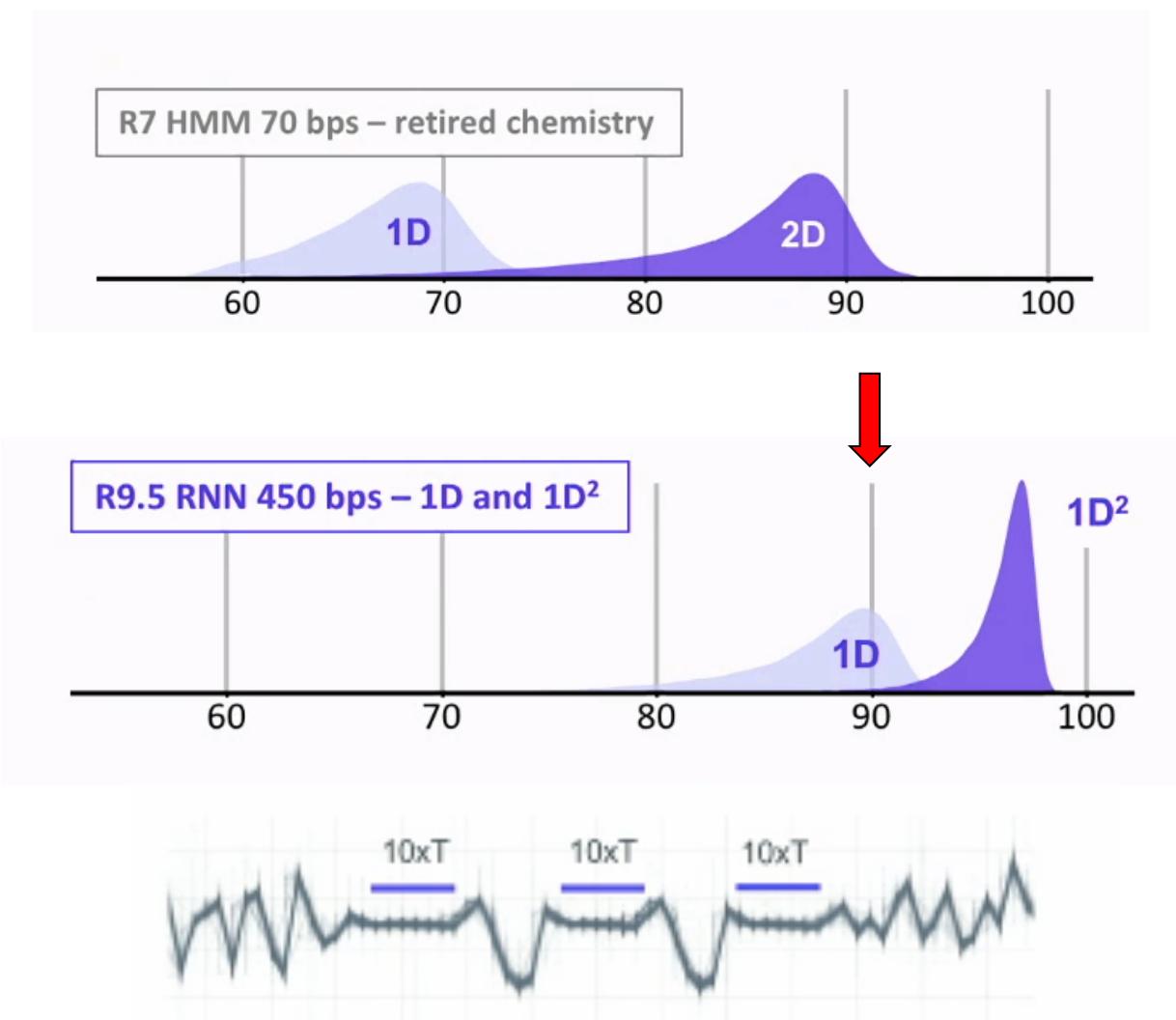


PromethION : 144000 pores (48 x 3000)

## BASECALLING



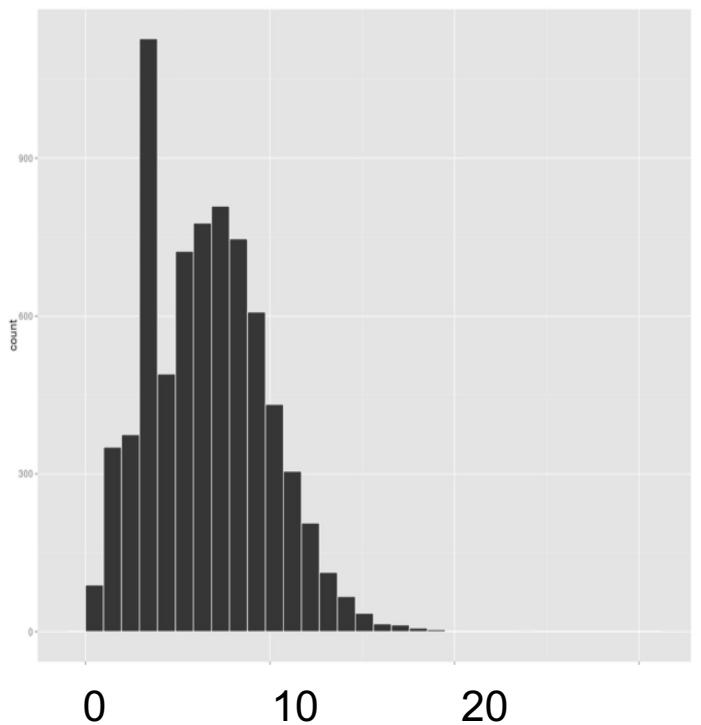
# QUALITY



Homopolymers difficult to sequence

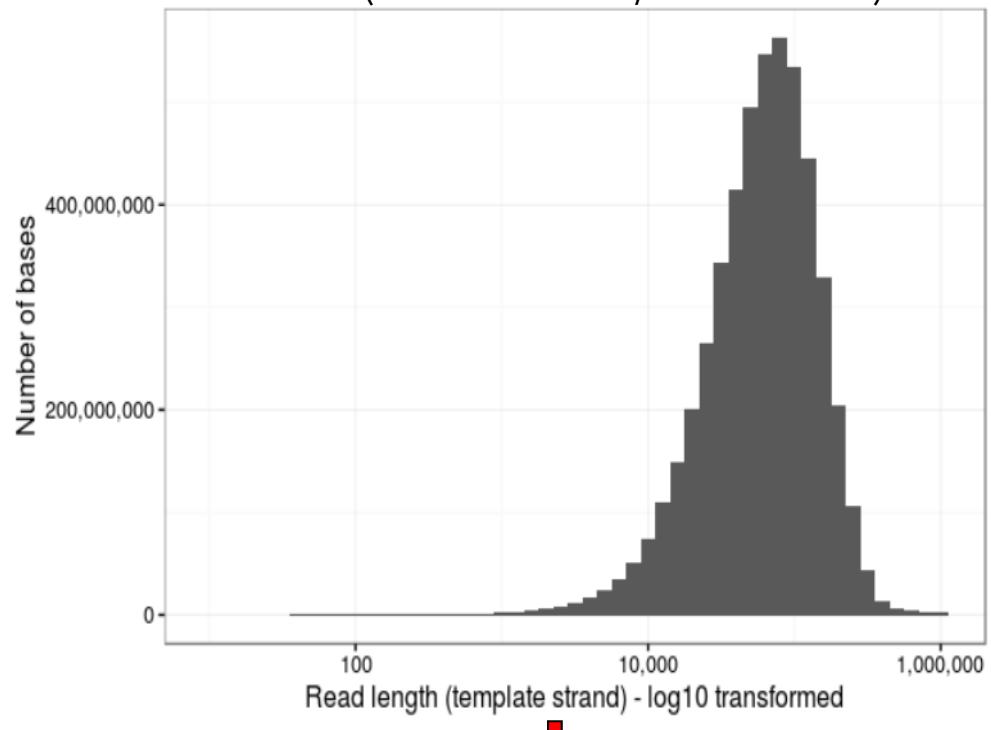
# SIZE OF SEQUENCED DNA FRAGMENTS

Typical profile of fragment size



(Risse et al. *GigaScience*, 2015)

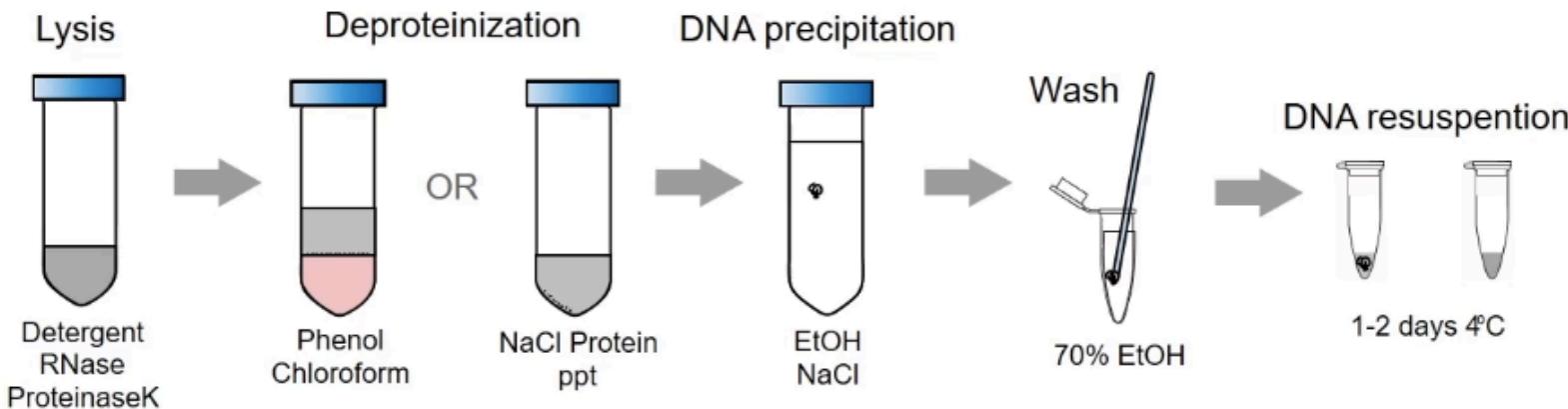
"Ultra long" reads  
(lab.loman.net, March 2017)



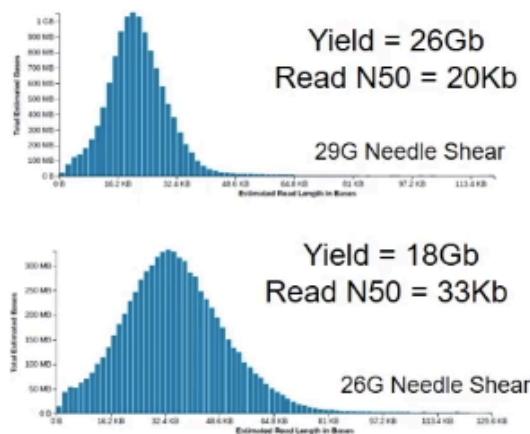
Size of the longest read : 778 kb

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

# SIZE OF SEQUENCED DNA FRAGMENTS

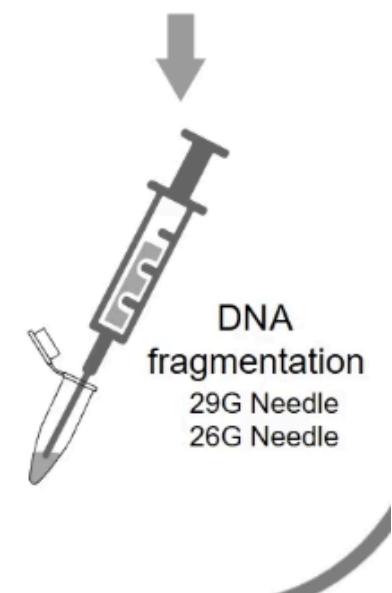


## Read Length Distributions



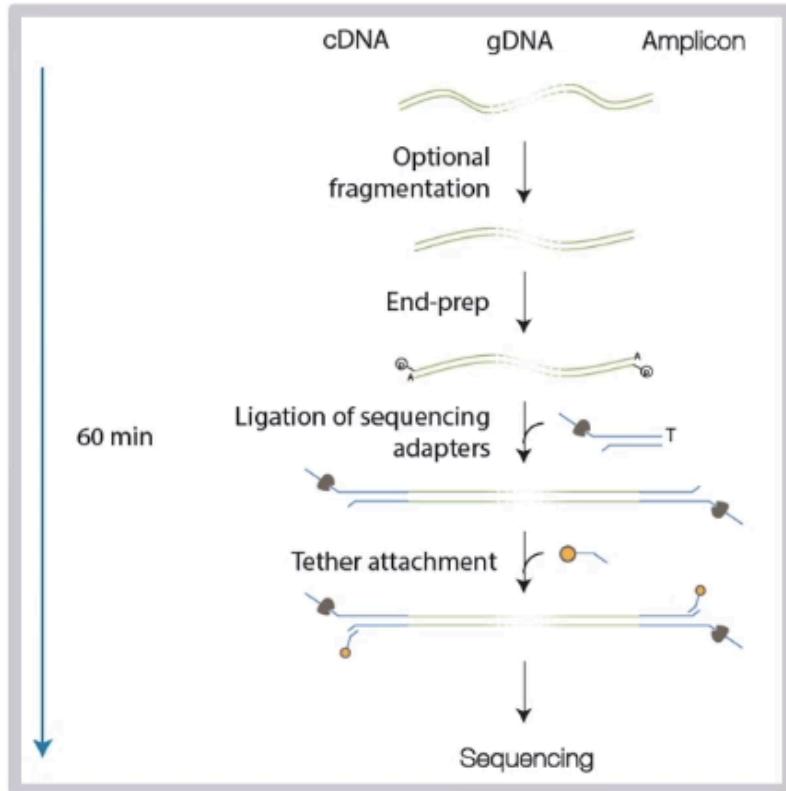
## Nanopore Library

Modified SQK-LSK109  
Ligation Prep

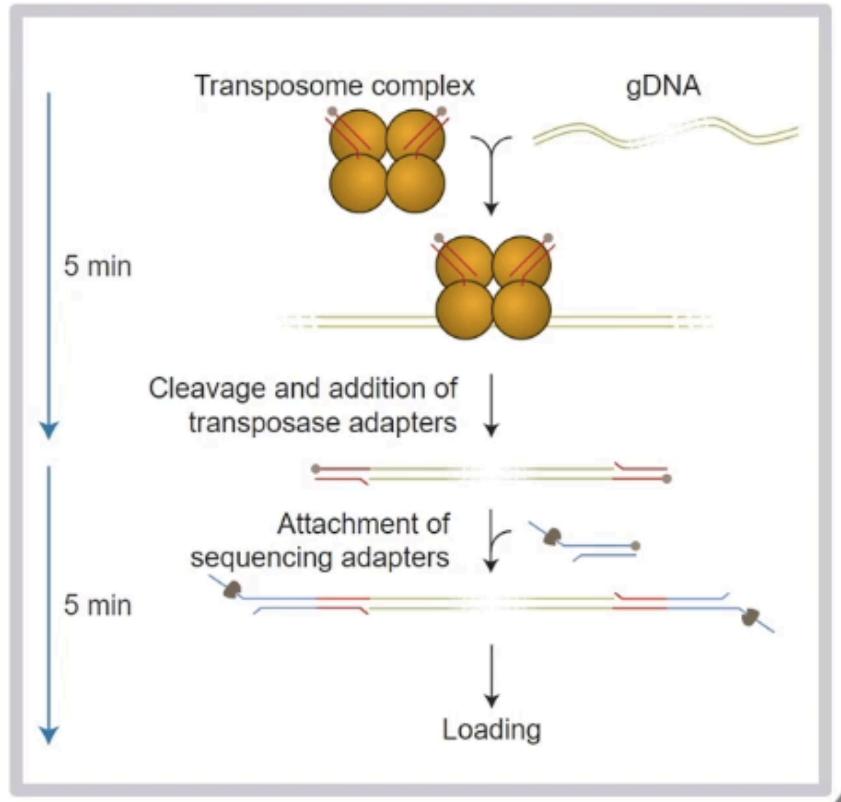


# SIZE OF SEQUENCED DNA FRAGMENTS

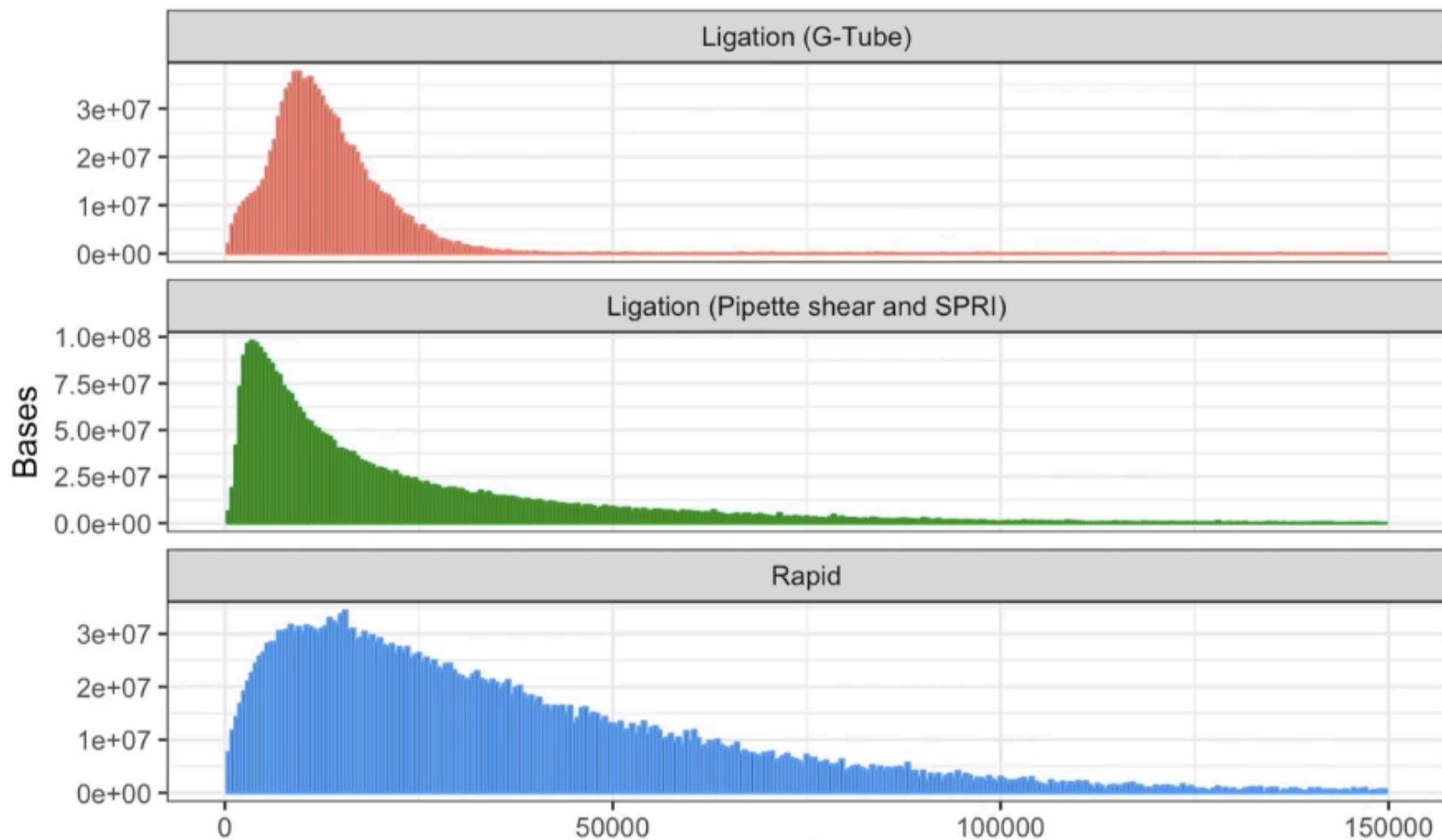
Ligation method



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

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

Miten Jain<sup>1,13</sup>, Sergey Koren<sup>2,13</sup>, Karen H Miga<sup>1,13</sup>, Josh Quick<sup>3,13</sup>, Arthur C Rand<sup>1,13</sup>, Thomas A Sasani<sup>4,5,13</sup>, John R Tyson<sup>6,13</sup>, Andrew D Beggs<sup>7</sup>, Alexander T Dilthey<sup>2</sup>, Ian T Fiddes<sup>1</sup>, Sunir Malla<sup>8</sup>, Hannah Marriott<sup>8</sup>, Tom Nieto<sup>7</sup>, Justin O'Grady<sup>9</sup>, Hugh E Olsen<sup>1</sup>, Brent S Pedersen<sup>4,5</sup>, Arang Rhie<sup>2</sup>, Hollian Richardson<sup>9</sup>, Aaron R Quinlan<sup>4,5,10</sup>, Terrance P Snutch<sup>6</sup>, Louise Tee<sup>7</sup>, Benedict Paten<sup>1</sup>, Adam M Phillippy<sup>2</sup>, Jared T Simpson<sup>11,12</sup>, Nicholas J Loman<sup>3</sup> & Matthew Loose<sup>8</sup>

eserved.

Using nanopore reads alone assembly of a human genome :

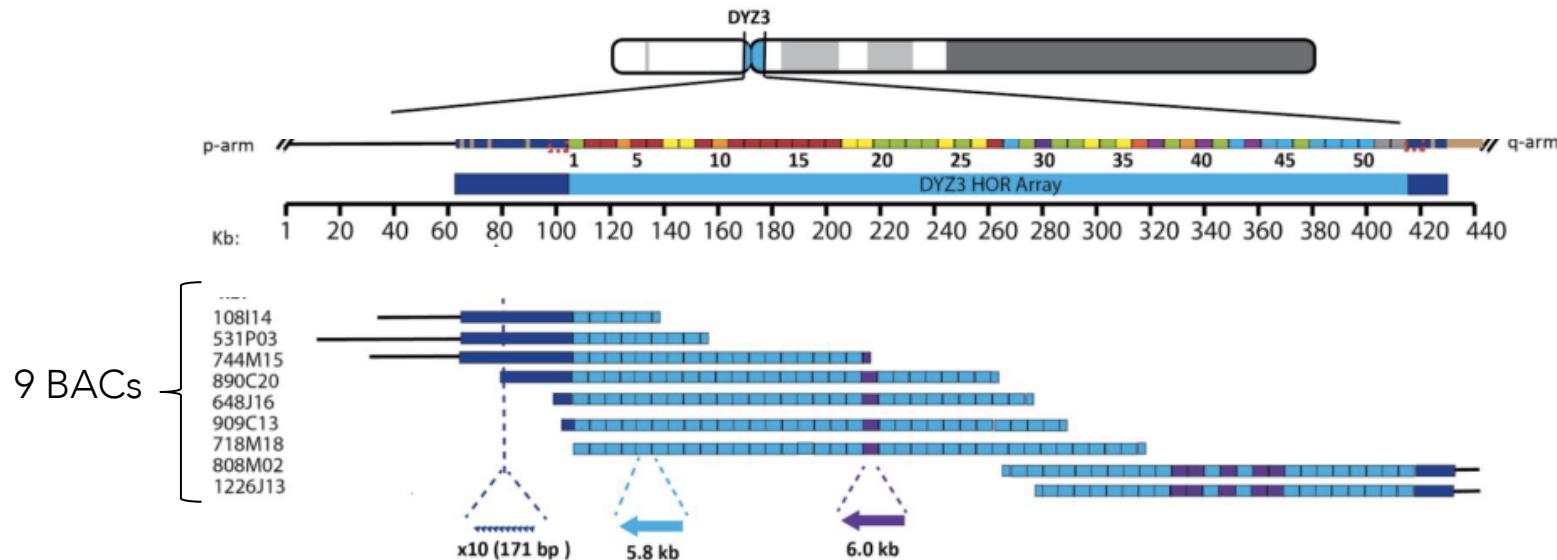
- NG50 contig size of ~6.4 Mb
- covers >85% of the reference
- 99.88% accuracy
- MHC locus on a single contig, phased over its full length
- closure of 12 large (>50 kb) gaps in the reference human genome

# 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**

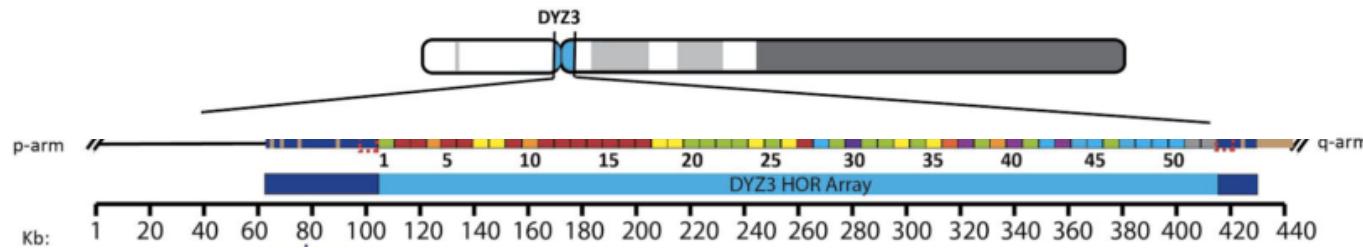


# ASSEMBLY OF A HUMAN Y CENTROMERE

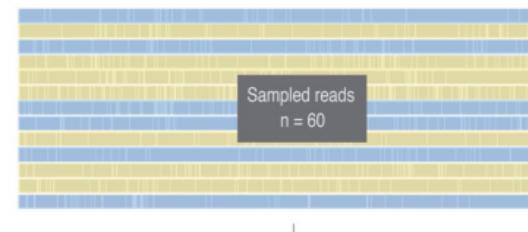
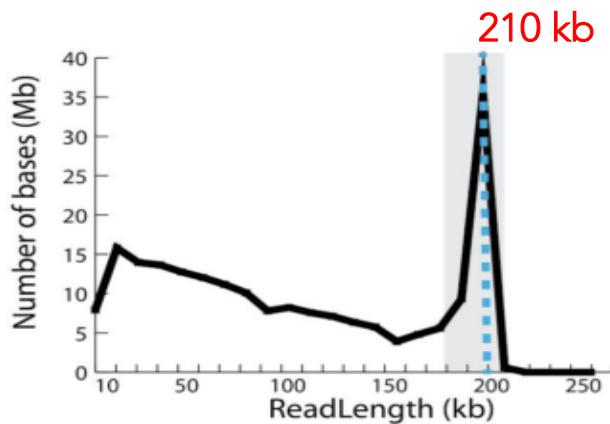
(Jain et al., bioRxiv, 2017)

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



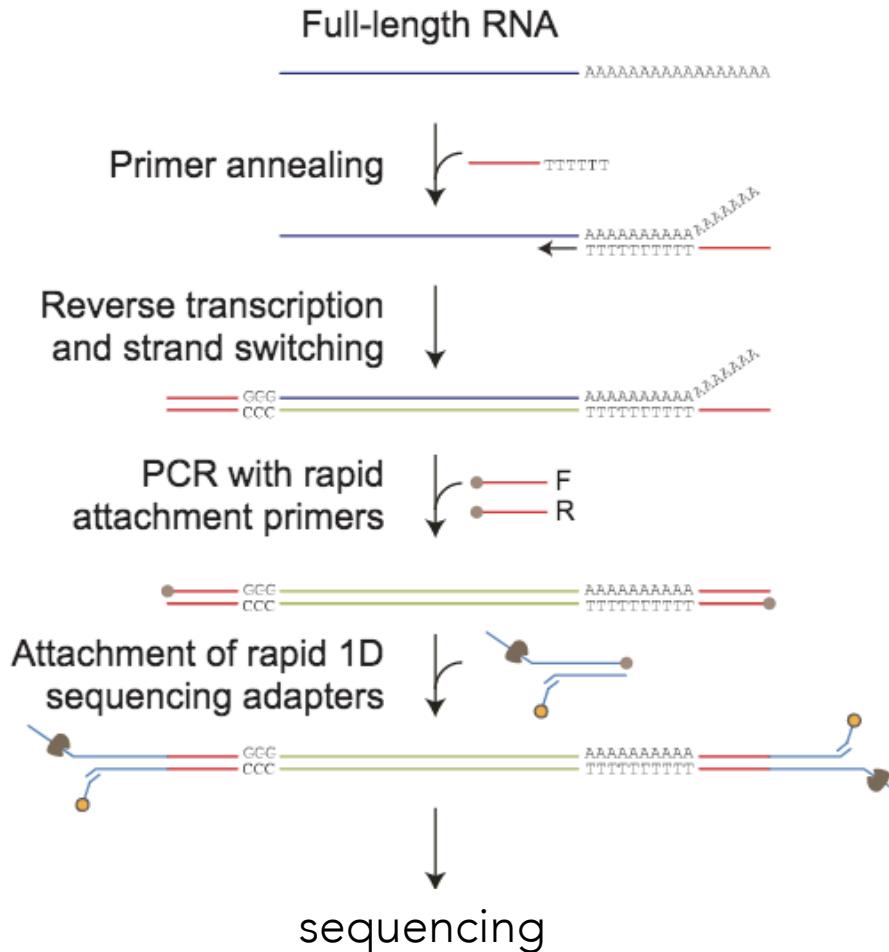
Final high quality consensus BAC sequence



FIRST COMPLETE SEQUENCE OF A HUMAN CENTROMERE

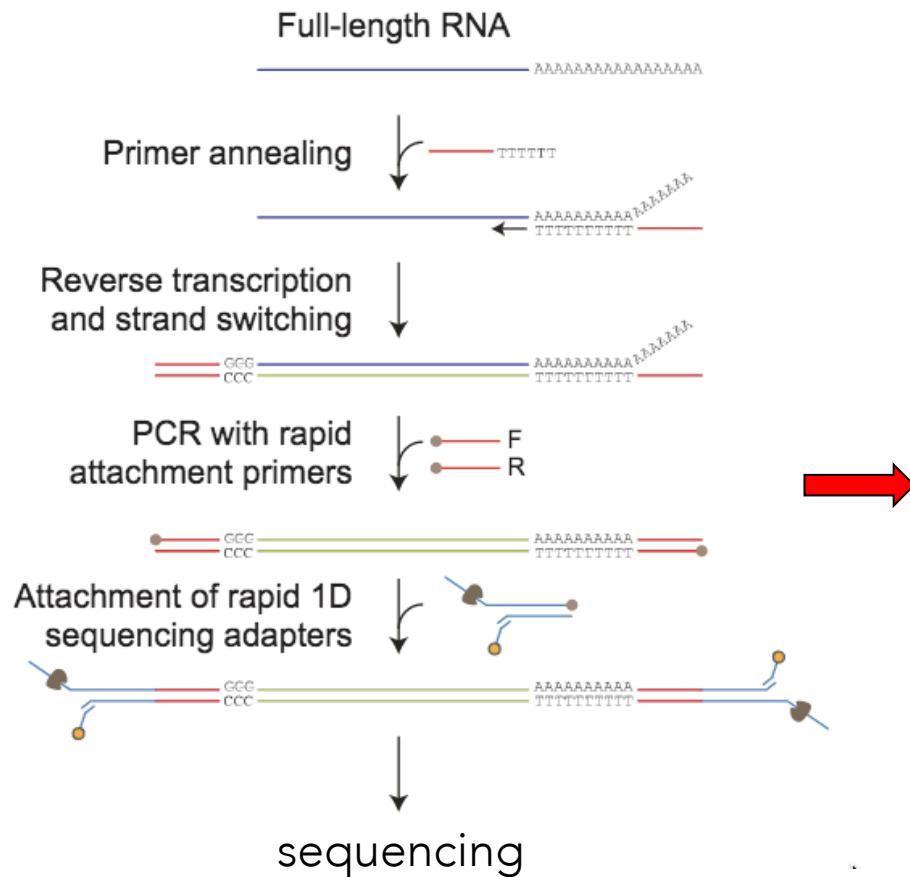
# cDNA SEQUENCING

## Library preparation

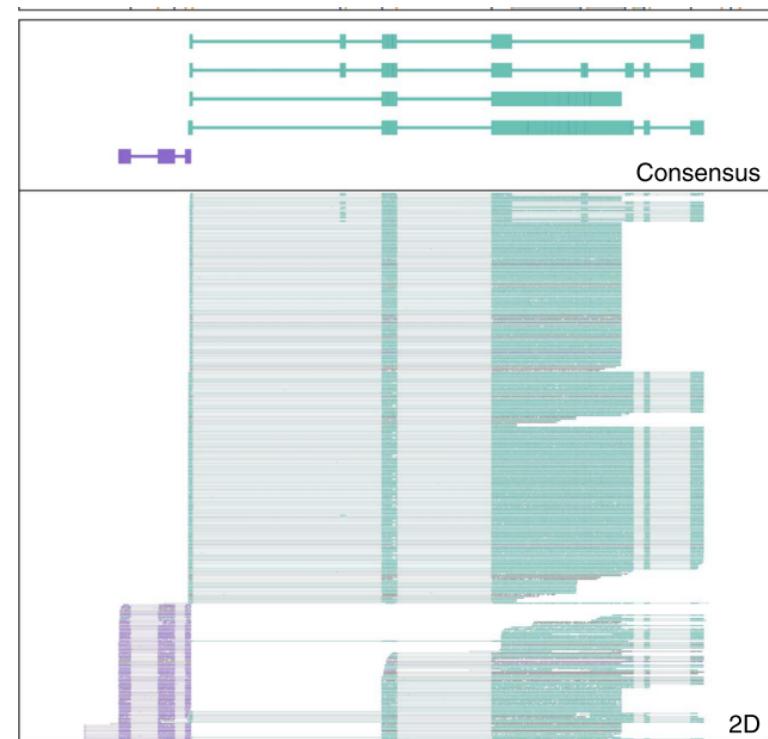


# cDNA SEQUENCING

## Library preparation



Detection of splice variants in surface receptor of B cells  
(Byrne et al. Nat. Comm. 2017)



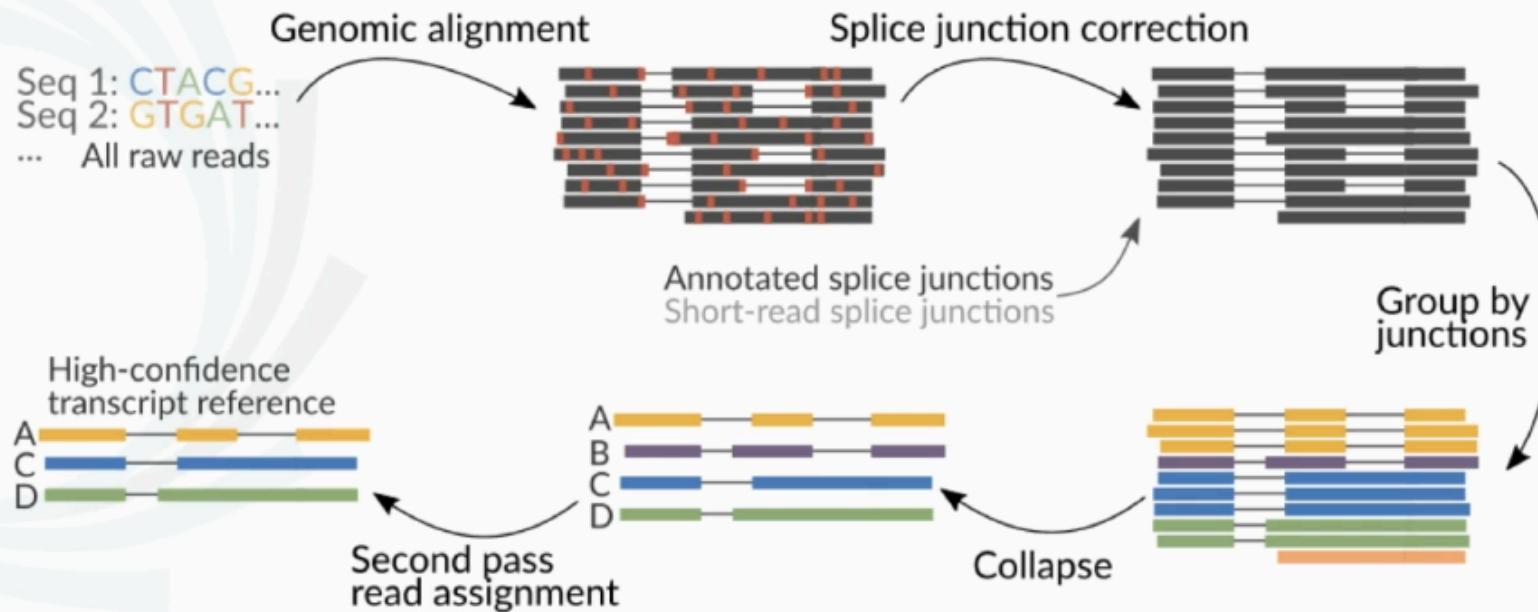
- Splice alignment difficult due to high (5-10%) error rate
- Reads are frequently truncated from 5' end

# CHALLENGES OF NANOPORE TRANSCRIPTOME ANALYSIS

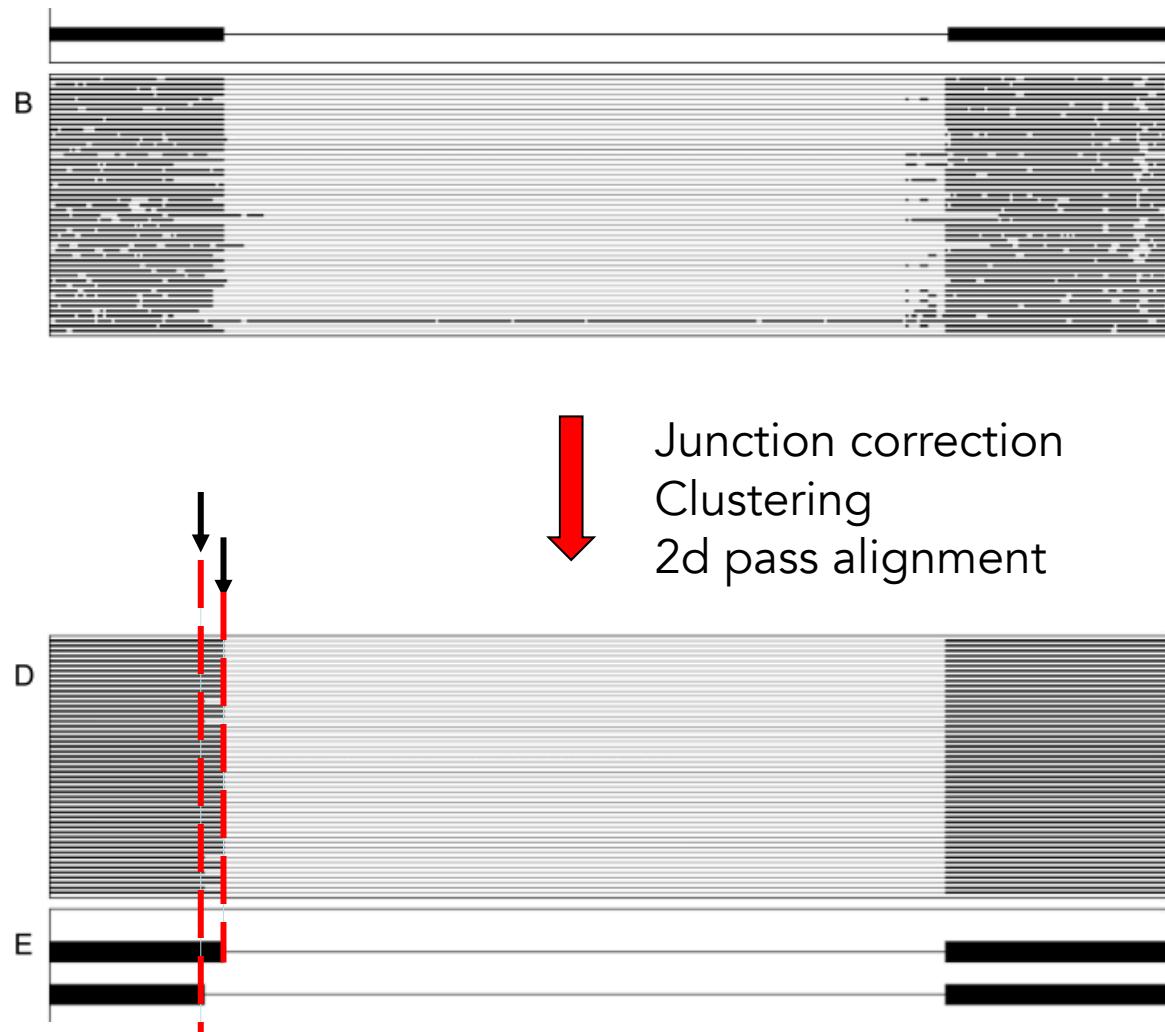
## FLAIR : a pipeline for splicing isoform determination

Tang et al. *bioRxiv* 2018

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

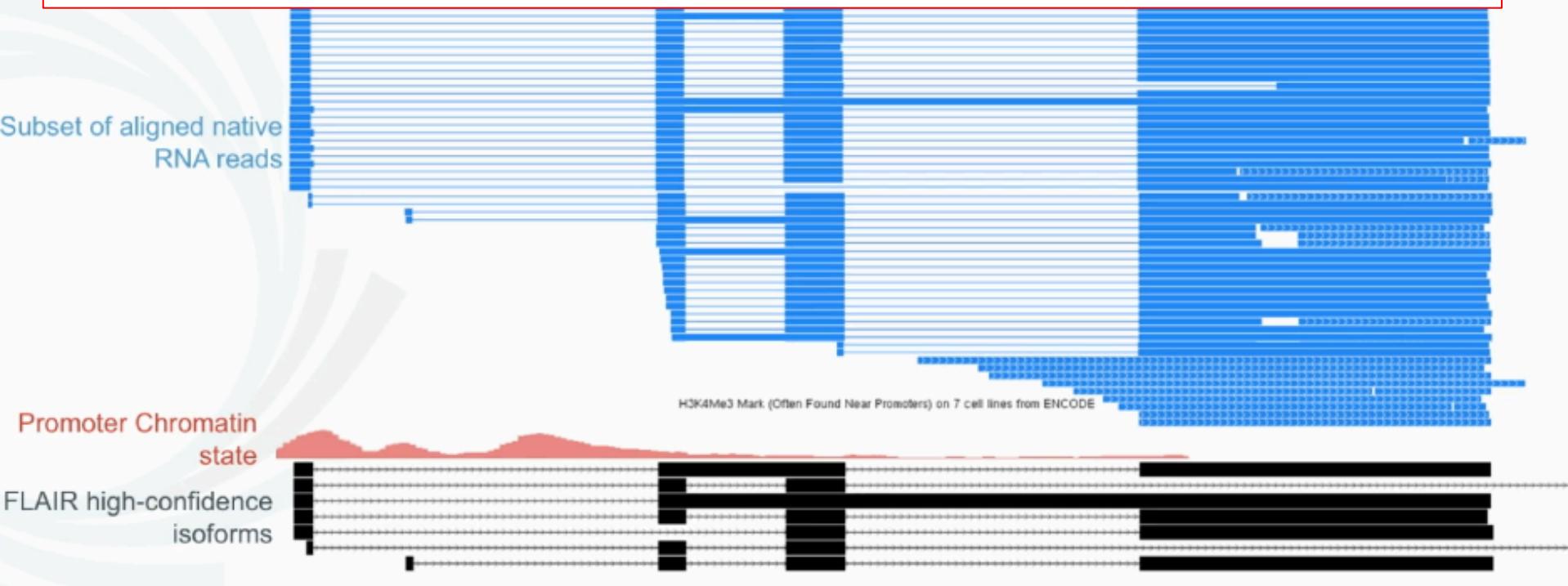


# CHALLENGES OF NANOPORE TRANSCRIPTOME ANALYSIS



# CHALLENGES OF NANOPORE TRANSCRIPTOME ANALYSIS

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

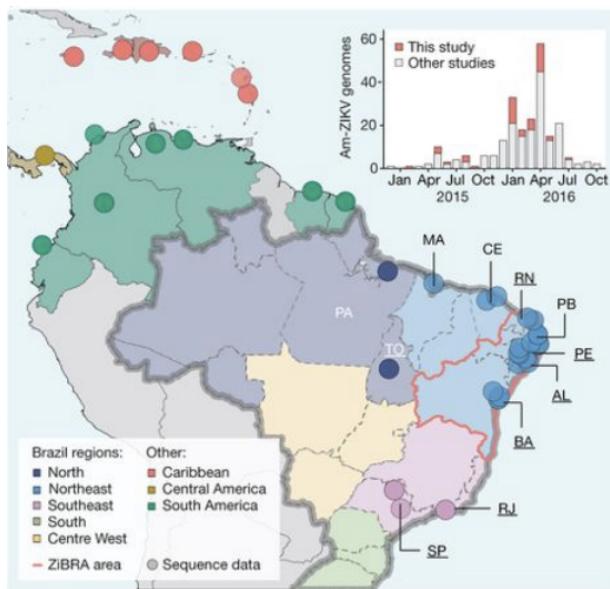


cDNA from chronic lymphocytic leukemia (CLL)

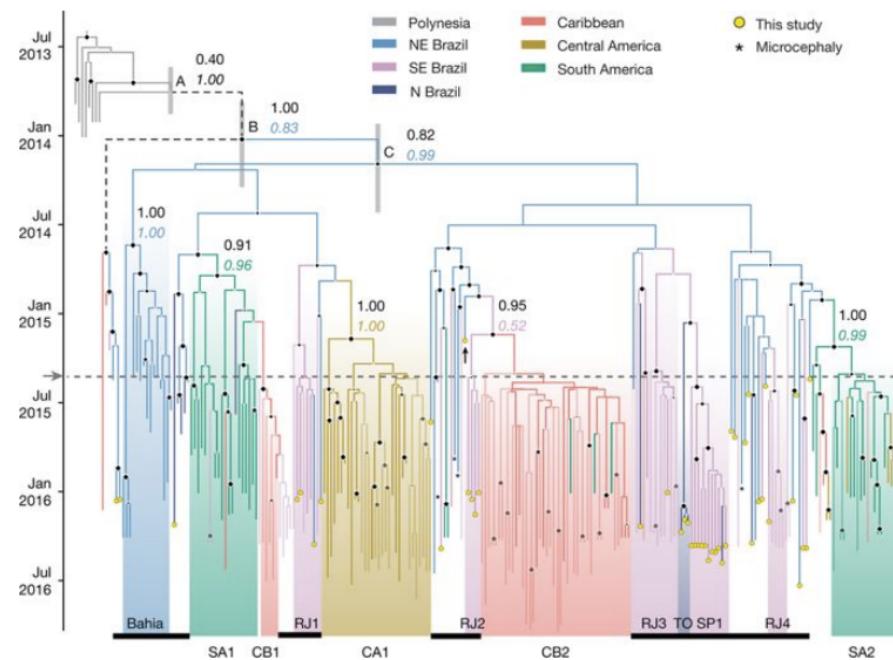
THE ZIBRA PROJECT : Establishment and cryptic transmission of Zika virus in Brazil and the Americas  
 Mobile genomics laboratory that travelled through northeast Brazil during June 2016.  
 The ZibRA laboratory screened 1,330 samples (blood) from patients in 82 municipalities across 5 federal states  
 The MinION protocol does not require an Internet connection for analysis, making it suitable for field applications

Viral RNA genome : + sense, 10 kb

**Viral consensus sequences can be achieved in 1-2 days.**



**Figure 3: Phylogeography of ZIKV in the Americas.**



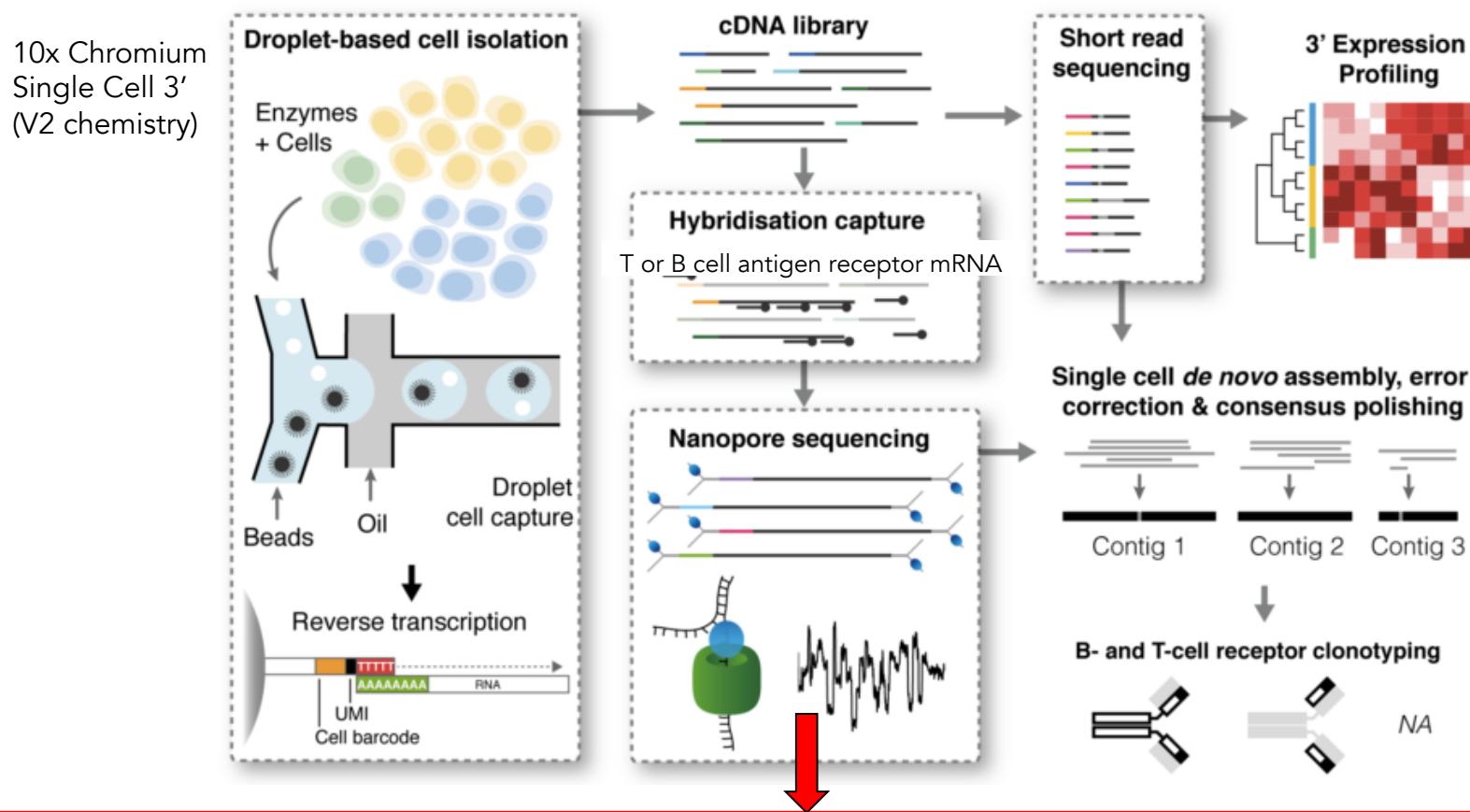
ZIKV was present in northeast Brazil by February 2014 and is likely to have disseminated from there, nationally and internationally, before the first detection of ZIKV in the Americas.

# COUPLING NANOPORE and SINGLE CELL cDNA SEQUENCING

High-throughput targeted long-read single cell sequencing reveals the clonal and transcriptional landscape of lymphocytes

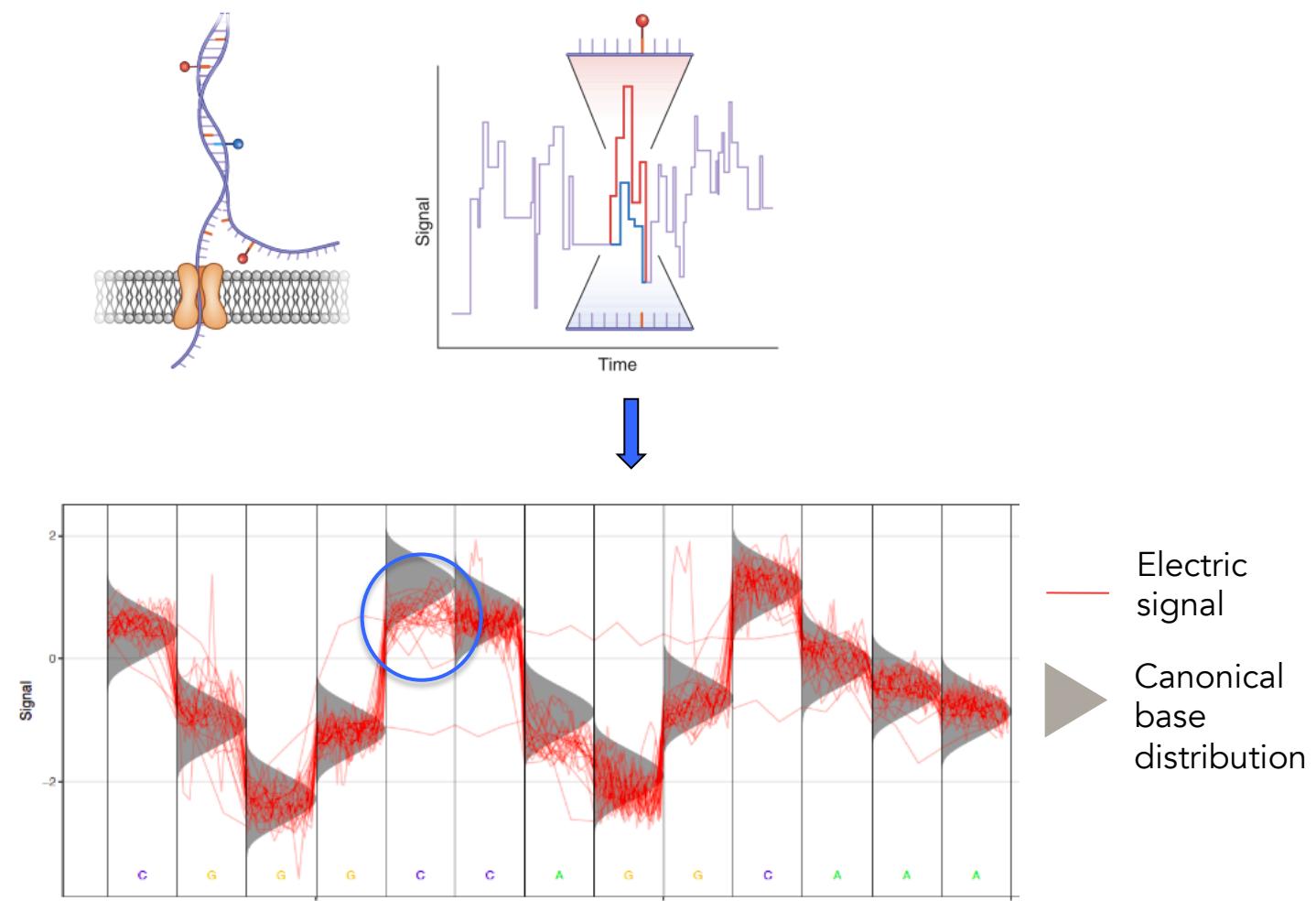
Singh et al., *bioRxiv*, 2018

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



Tracking of somatic mutation, alternate splicing and clonal evolution of T and B lymphocytes

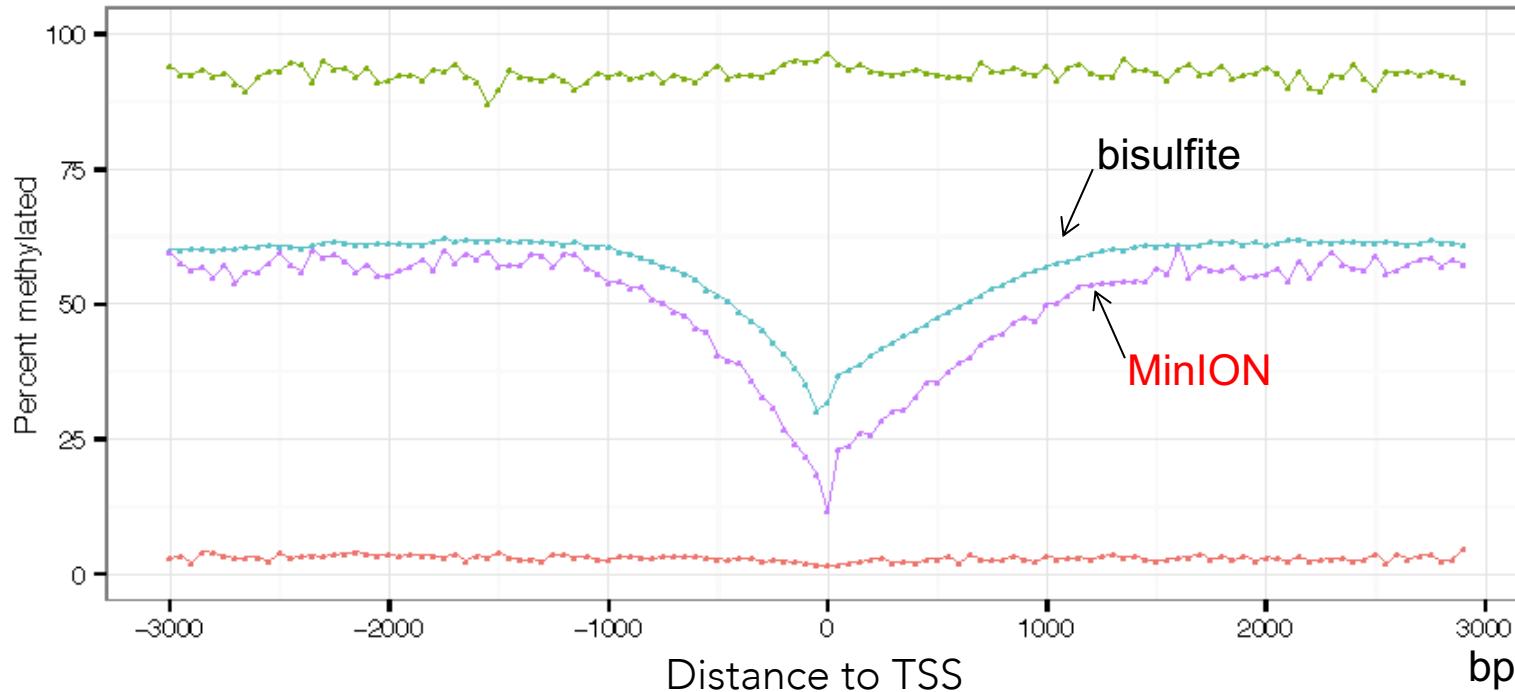
# DETECTION OF MODIFIED BASES USING NANOPORE



# DETECTION OF MODIFIED BASES USING NANOPORE

modified bases → modified currents

Detection of methylated cytosine (m5C) in CpG sites in the human genome



Simpson et al., *Nat. Methods*, 2017

Softwares available:

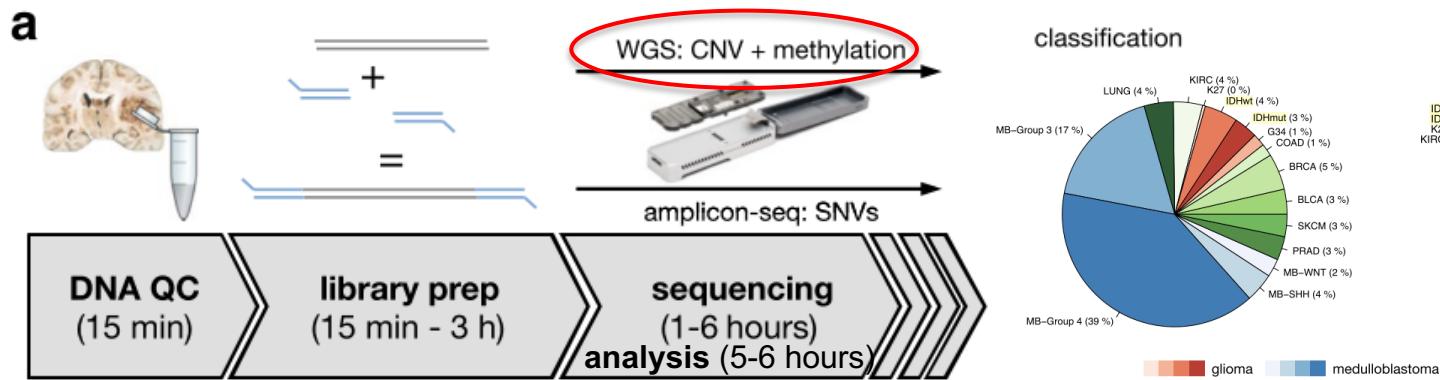
Simpson et al., 2017, *Nat. Methods*

Rand et al., 2017, *Nat. Methods*

# DETECTION OF MODIFIED BASES USING NANOPORE

Same-day genomic and epigenomic diagnosis of brain tumors (gliomas, medulloblastomas) using real-time nanopore sequencing

Euskirchen et al., *Acta Neuropathol.* (2017)

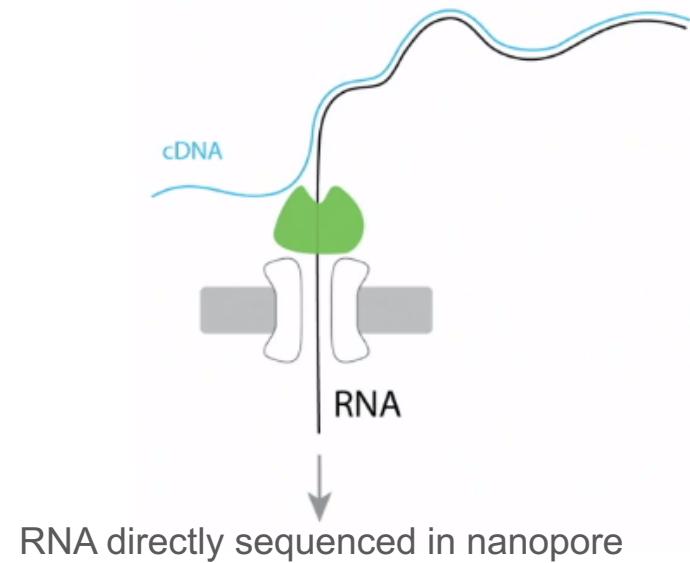
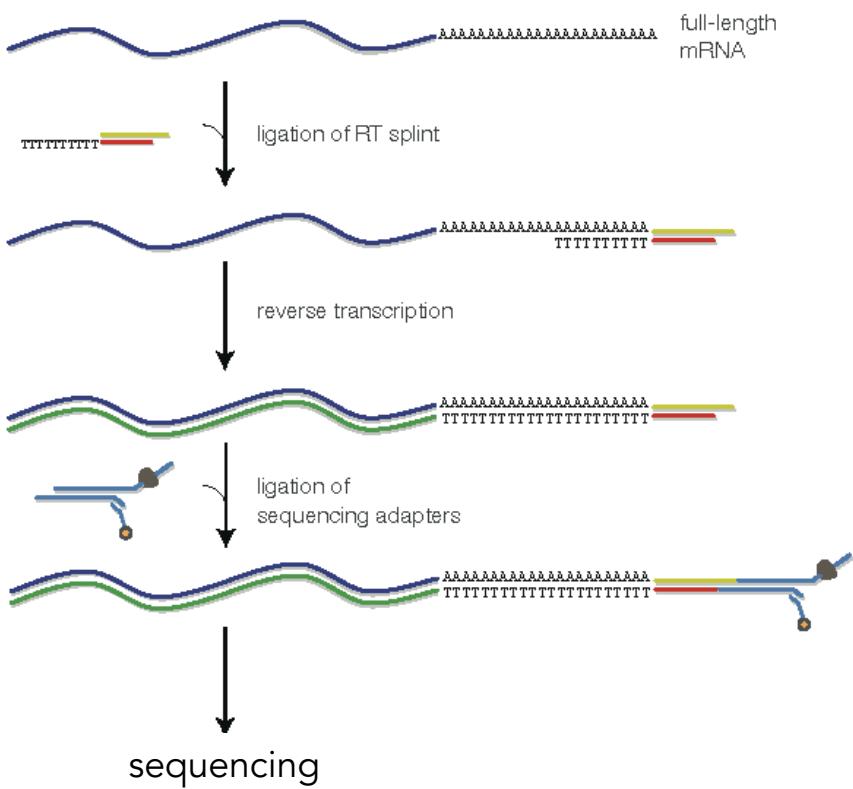


Nanopore sequencing allows same-day detection of structural variants, point mutations, and methylation profiling using a single device with negligible capital cost.

It outperforms hybridization-based and current sequencing technologies aiming to make precision medicine possible for every cancer patient.

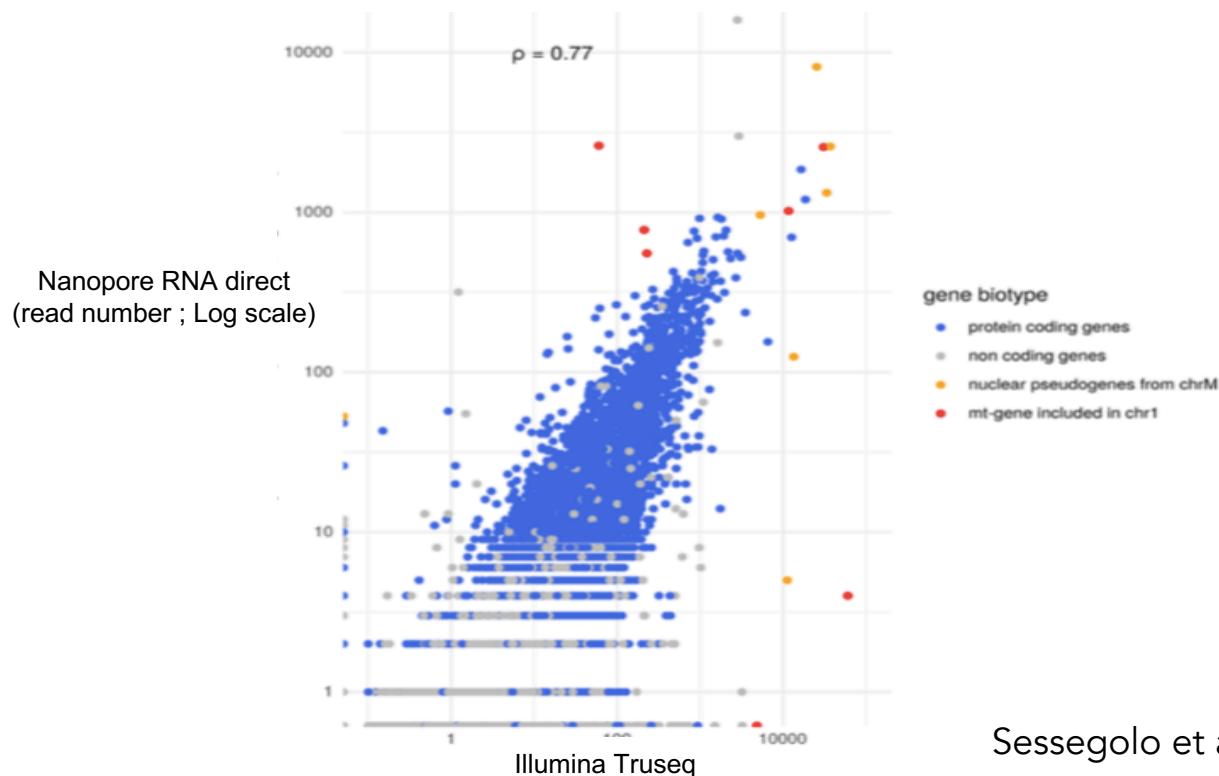
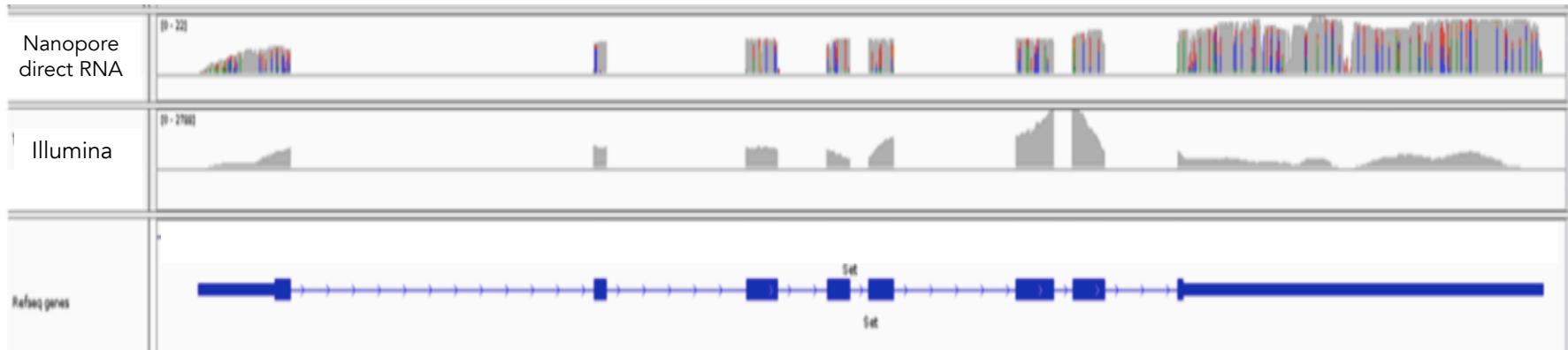
# DIRECT RNA SEQUENCING

## Library preparation

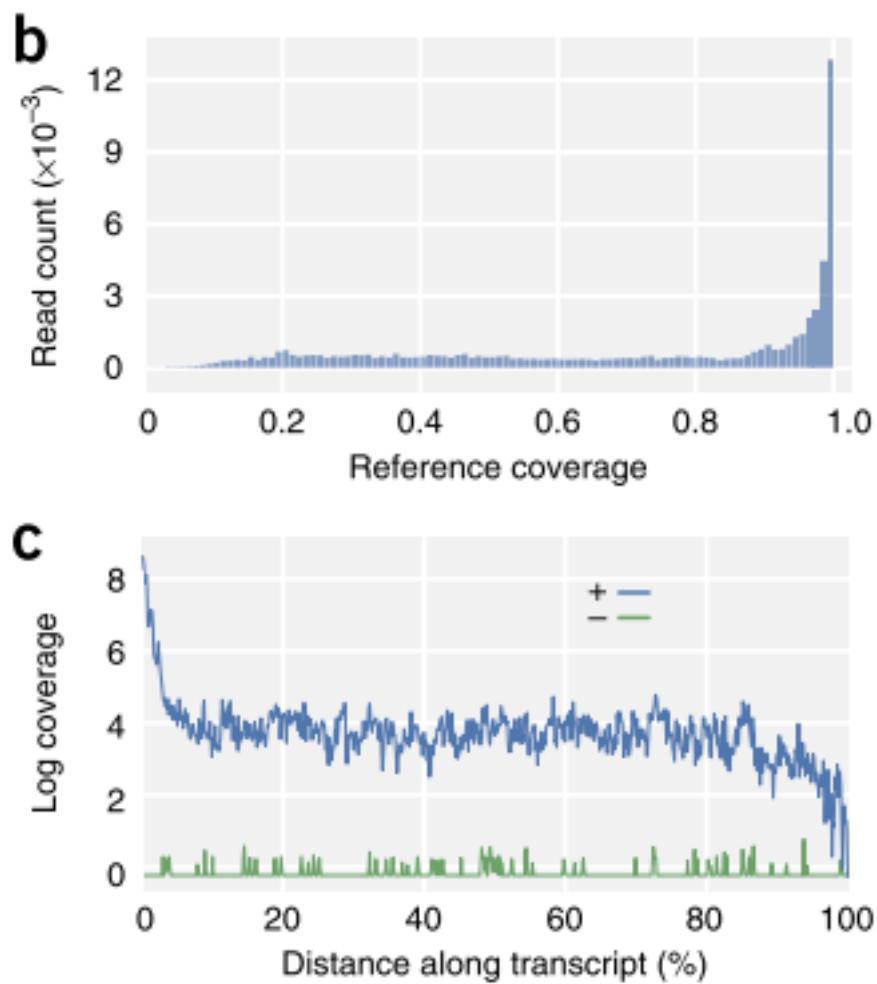
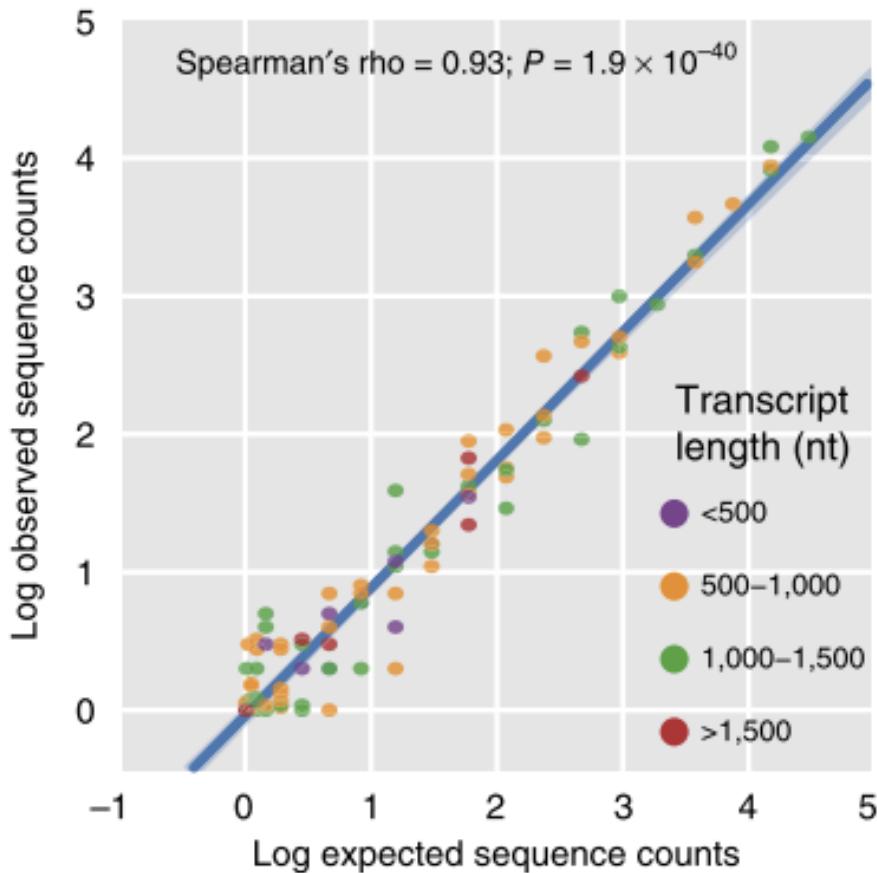


- No PCR bias
- Quantitative

# ILLUMINA cDNA vs NANOPORE DIRECT RNA



# DIRECT RNA SEQUENCING / SPIKE-IN CONTROLS

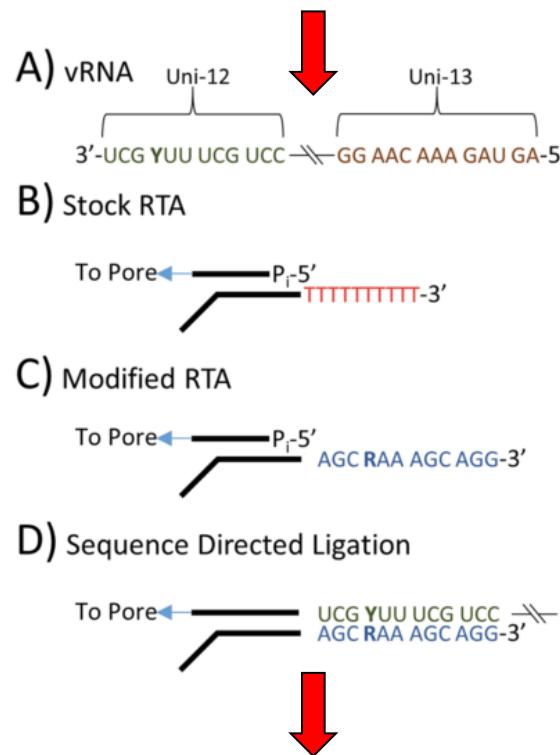


# DIRECT RNA SEQUENCING

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)



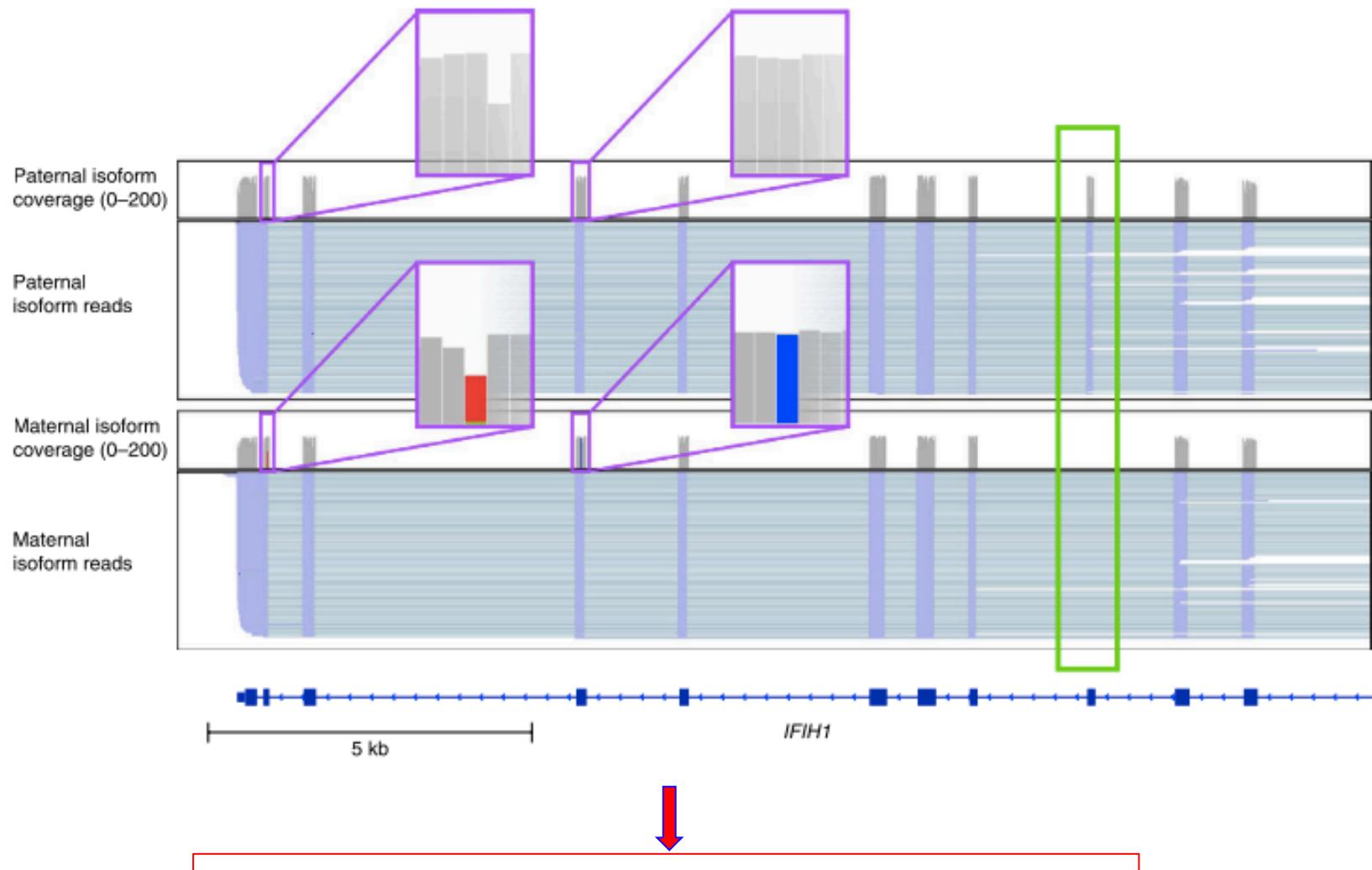
sequencing of complete genome with 100% nucleotide coverage, 99% consensus identity

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

# ALLELIC SPECIFIC ISOFORM ANALYSIS

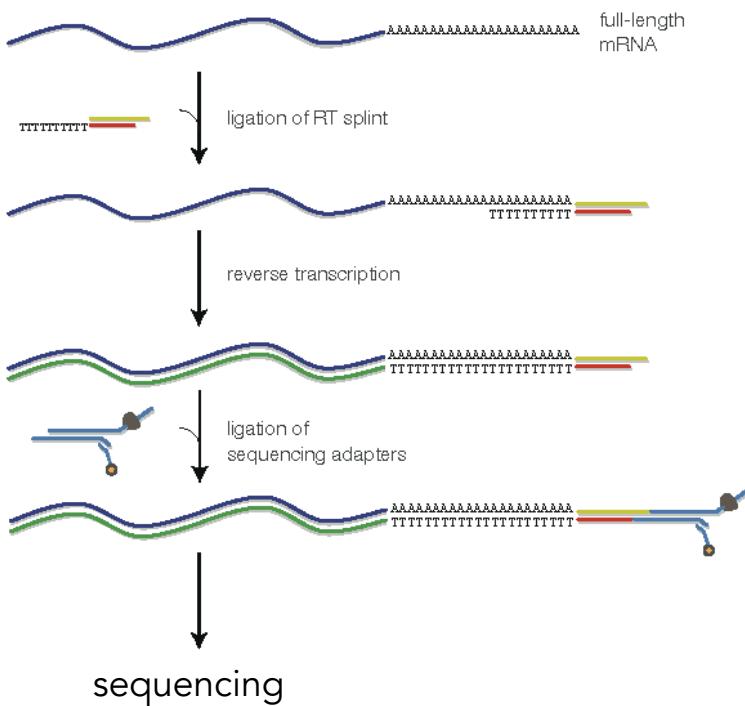
Nanopore native RNA sequencing of a human poly(A) transcriptome

d

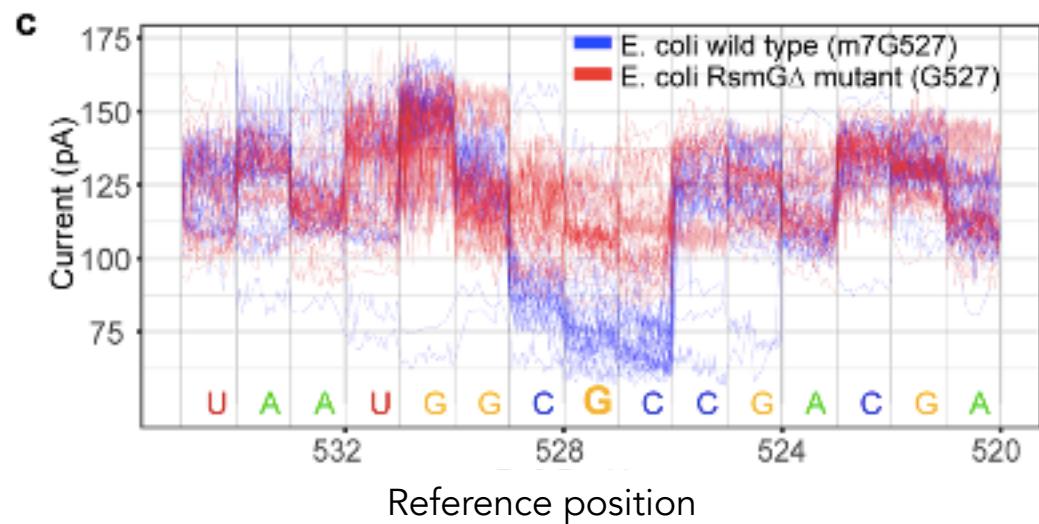


# DIRECT RNA SEQUENCING : DETECTION OF MODIFIED BASES

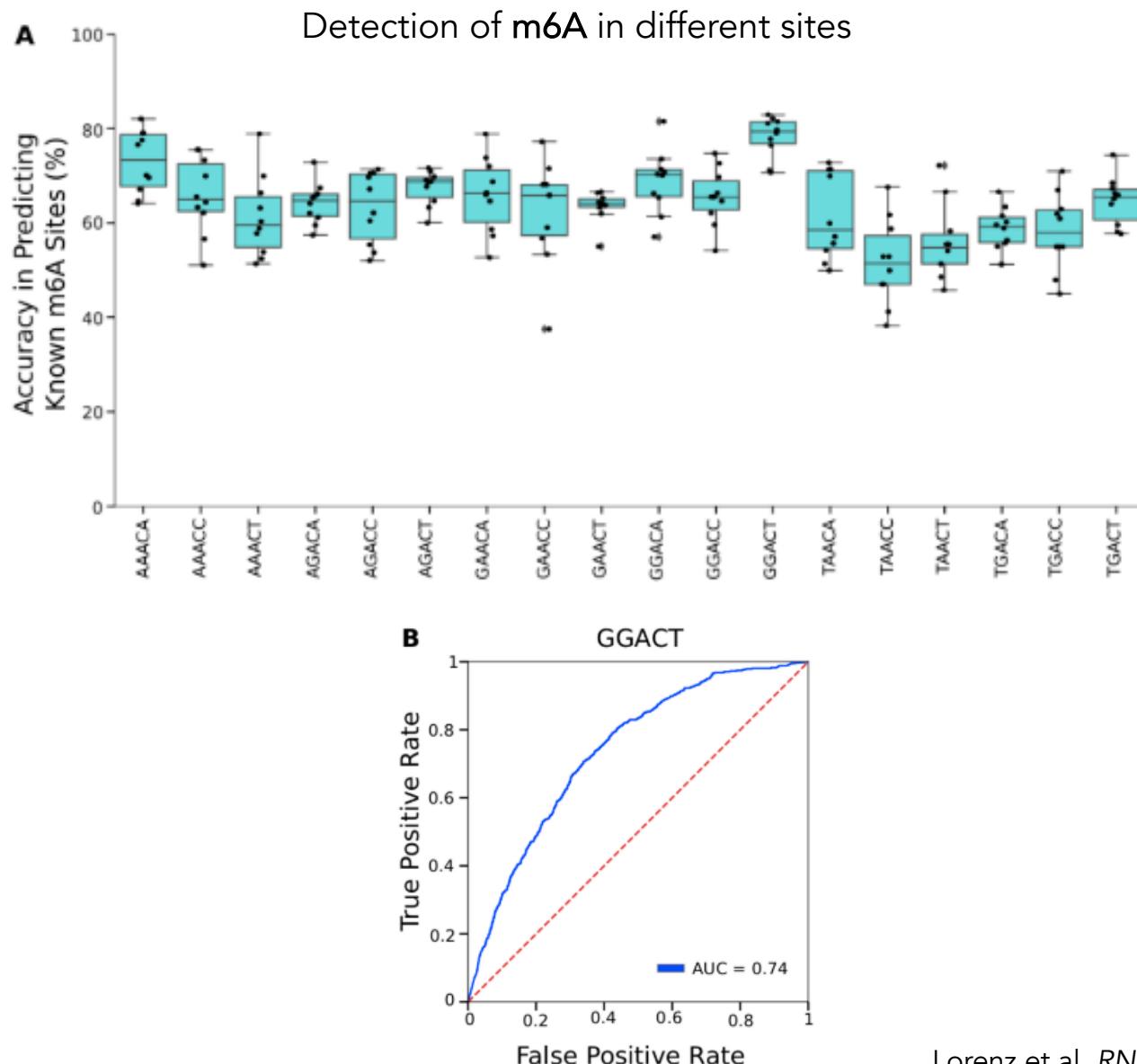
## Library preparation



## Detection of m6A modification in RNA molecules



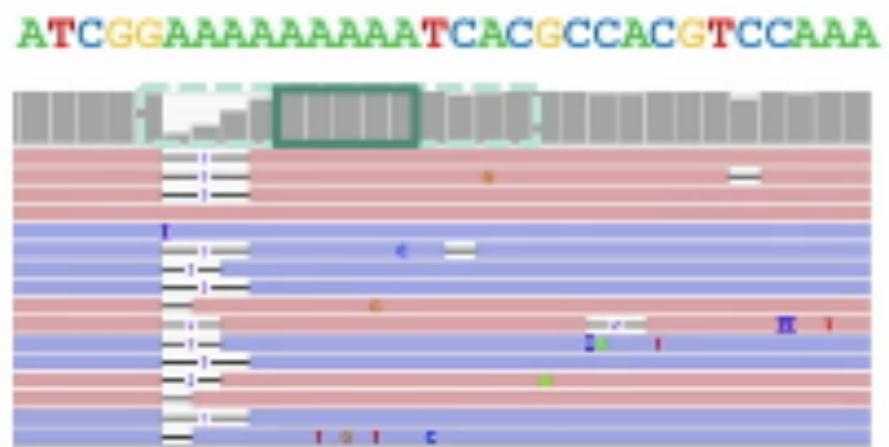
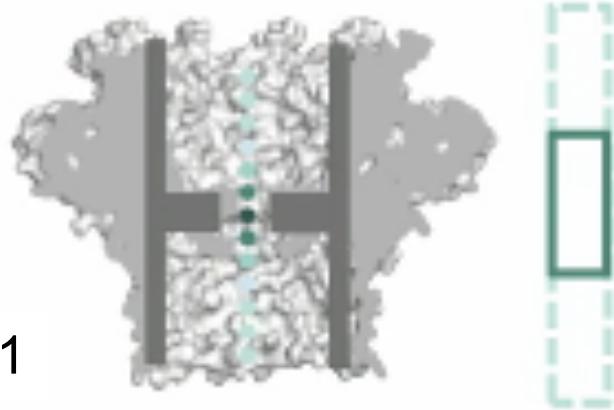
# DIRECT RNA SEQUENCING : DETECTION OF MODIFIED BASES



## RECENT IMPROVEMENTS

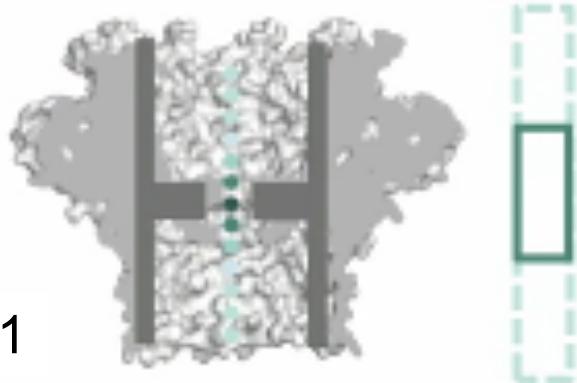
“one-reader” pore has difficulty to read homopolymers

R9.4.1



# RECENT IMPROVEMENTS

R9.4.1



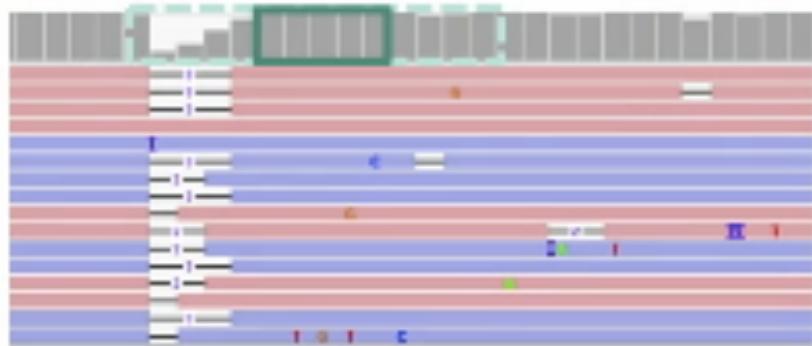
2019

R10  
“two-readers”

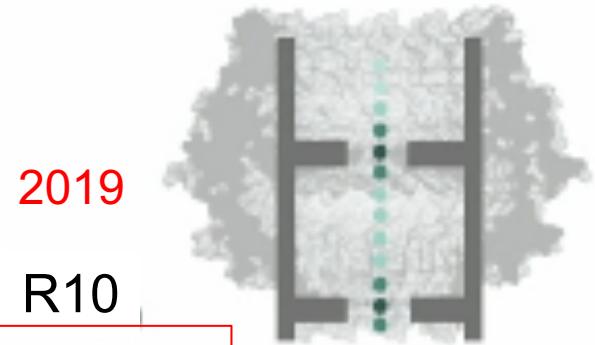
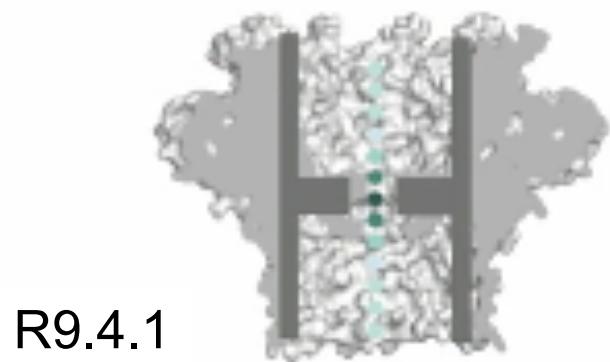
## New pore accurately calls homopolymers

- 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

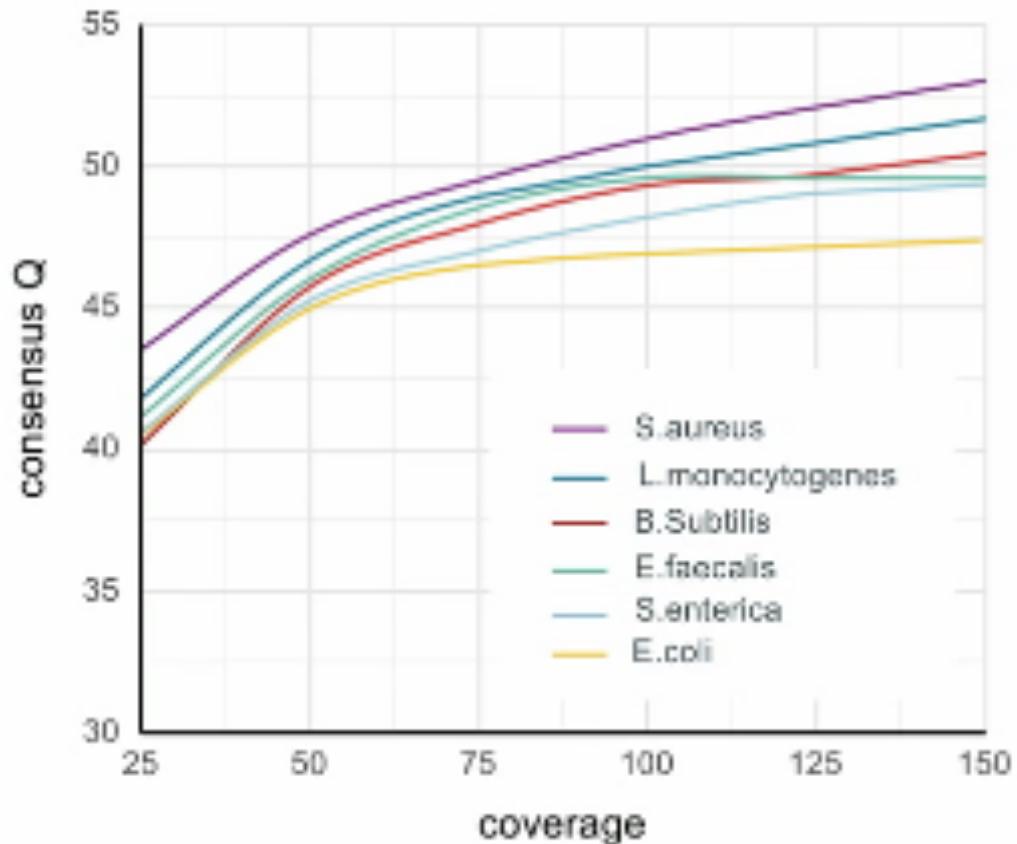
ATCGGAAAAAAATCACGCCACGTCCAAA



## RECENT IMPROVEMENTS



Consensus accuracy (R10 flow cell)



“two-readers” flow cell can reach Q>50

# Conclusions

Very fast-evolving technologies – Strong competition

## *PacBio*

- Maximum read length : 200 kb
- Error rate compensated by highly accurate circular consensus sequencing (CCS) reads
- Sequencing of cDNAs (resolution of alternative splicing)
- Detection of modified DNA with context effects (preferentially 6mA)

## *Nanopore*

- Very light sequencing system
- Very long reads : maximum length  $\gg$  200 kb
- Problems with homopolymers : solution with “two-readers” pore
- Sequencing of cDNAs (resolution of alternative splicing)
- Detection of modified DNA with context effects (preferentially 5mC)
- Direct sequencing of RNA
- Direct detection of modified RNA (6mA)