

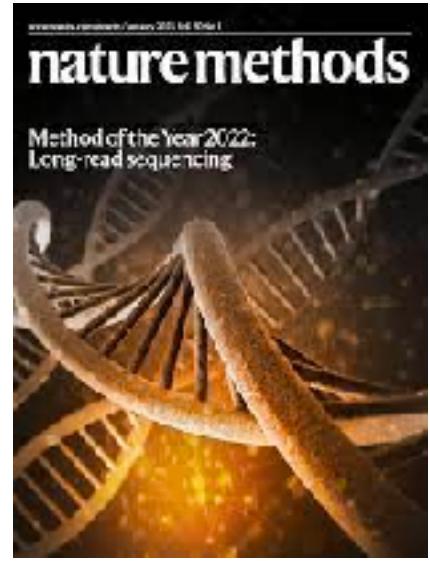
# 生物資訊在長讀長定序技術中所扮演的關鍵角色

從裡到外的改進 - 以Nanopore定序技術為例

邱家軍 (米夏糯 or Michael Nostalgie)

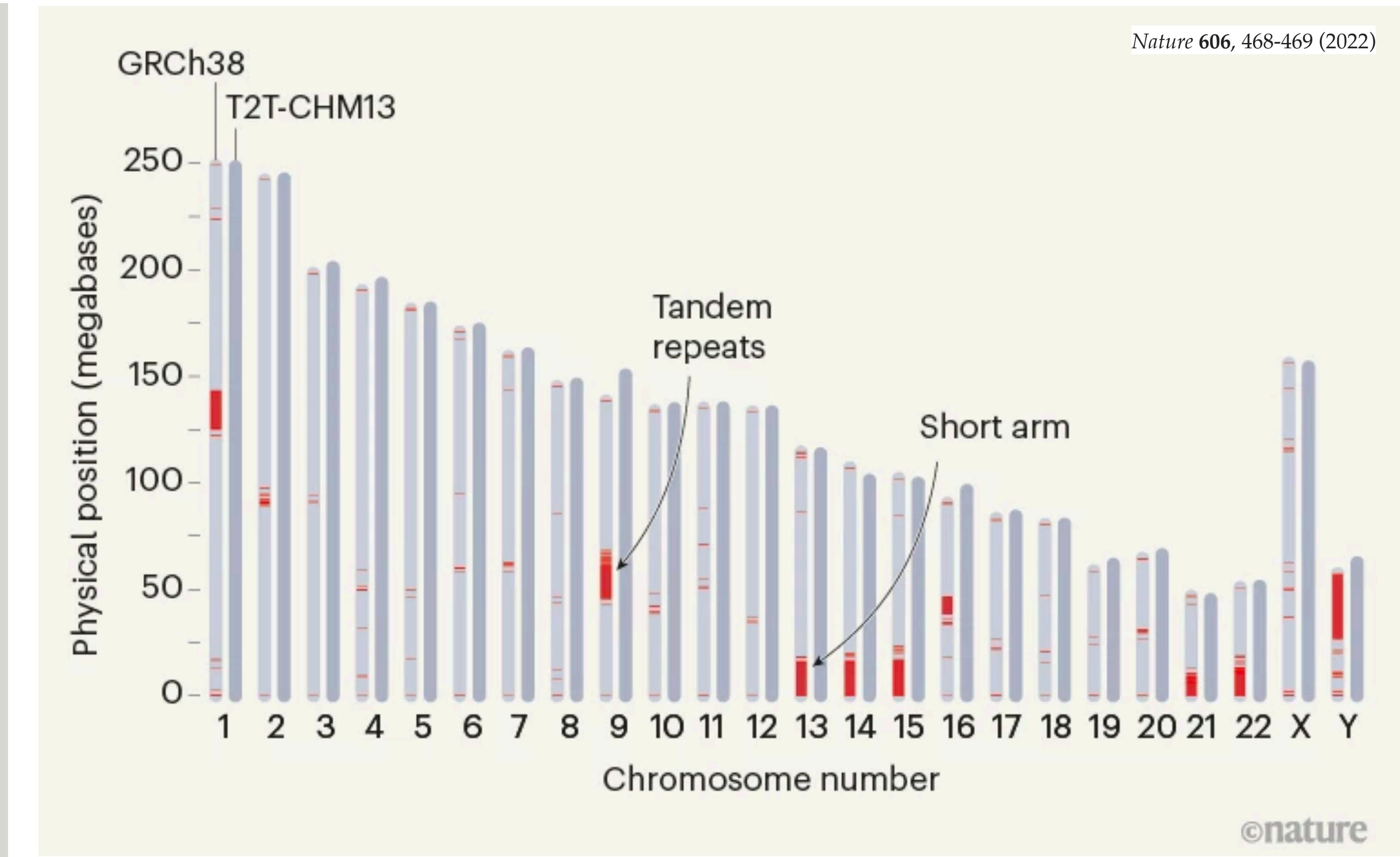
# 本日演講內容

- 長讀長定序技術的重要性
- 長讀長定序技術 (Long-read sequencing, LRS) 簡介及其背後的運作原理
  - PacBio
  - Nanopore
  - Others
- 目前LRS技術的發展狀況、應用場景及其背後用到的生物資訊理論方法
  - State-of-the-art improvements & applications of LRS
  - Development of bioinformatic methods behind these technologies
- 未來LRS的發展及可能的應用所帶來的挑戰，生物資訊端如何因應？

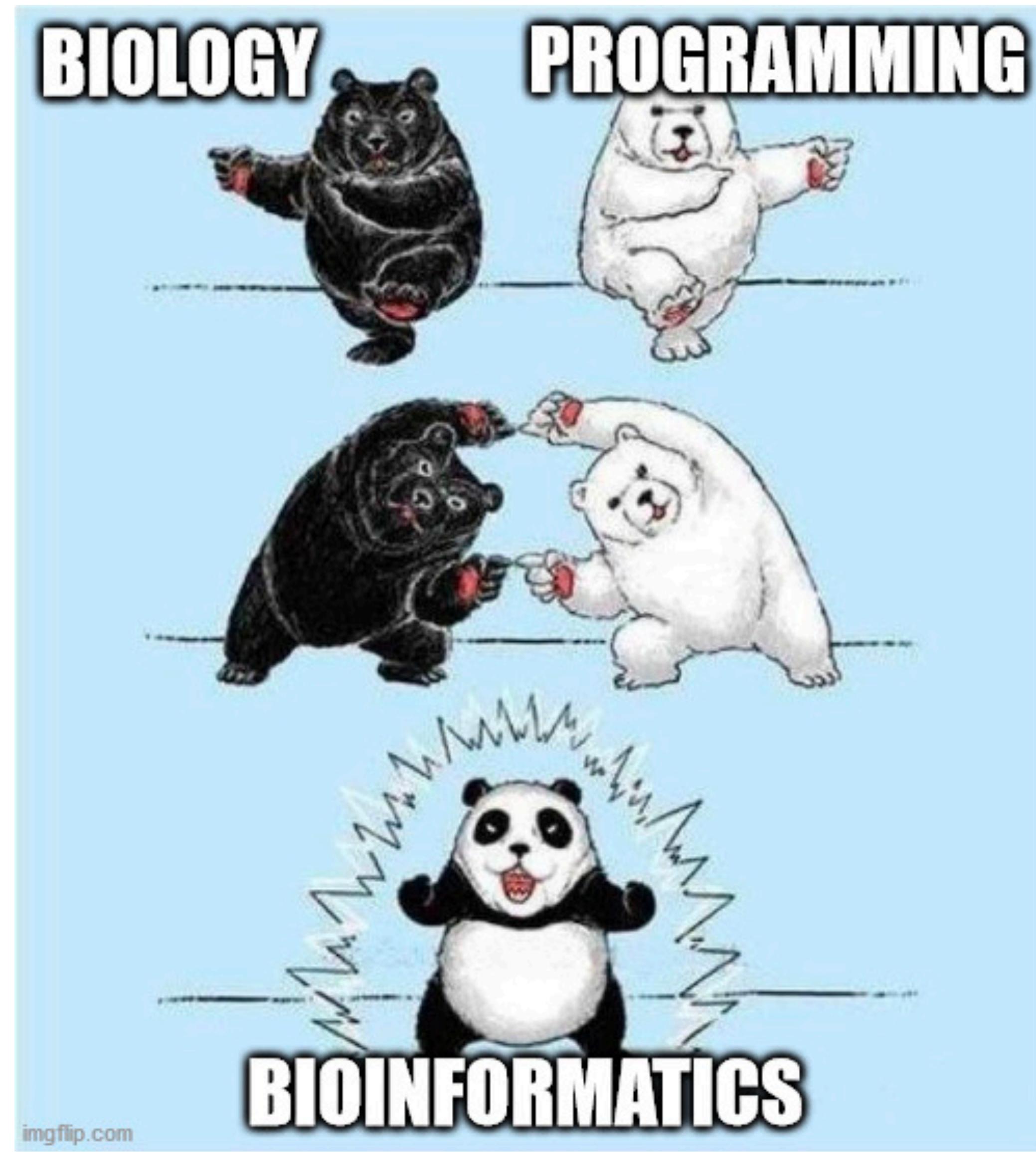


# Method of the Year 2022

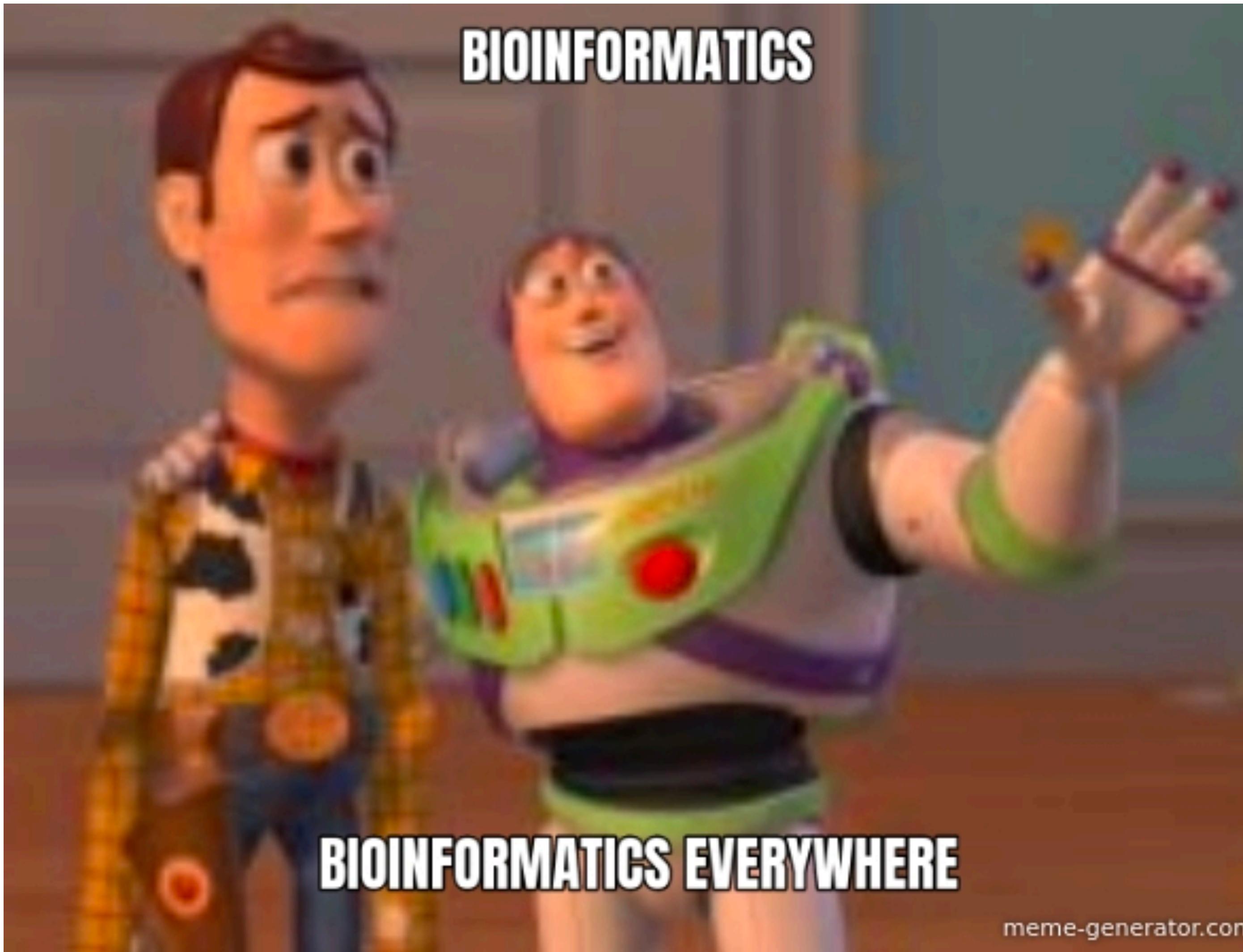
## Why is Long-Read Sequencing?



# Why Bioinformatics? (1/3)

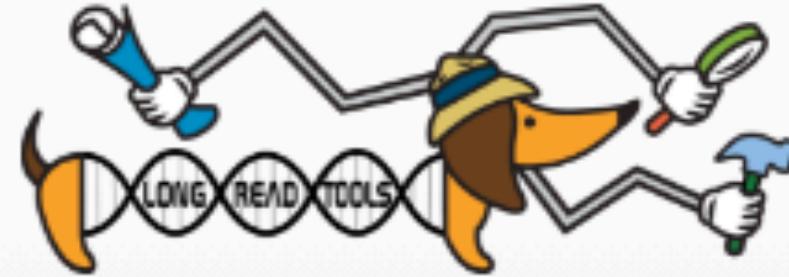


# Why Bioinformatics? (2/3)



- File processing / parsing
- Algorithms
- Tools
- More insights
- Tuning experiment protocols

# Why Bioinformatics? (3/3)



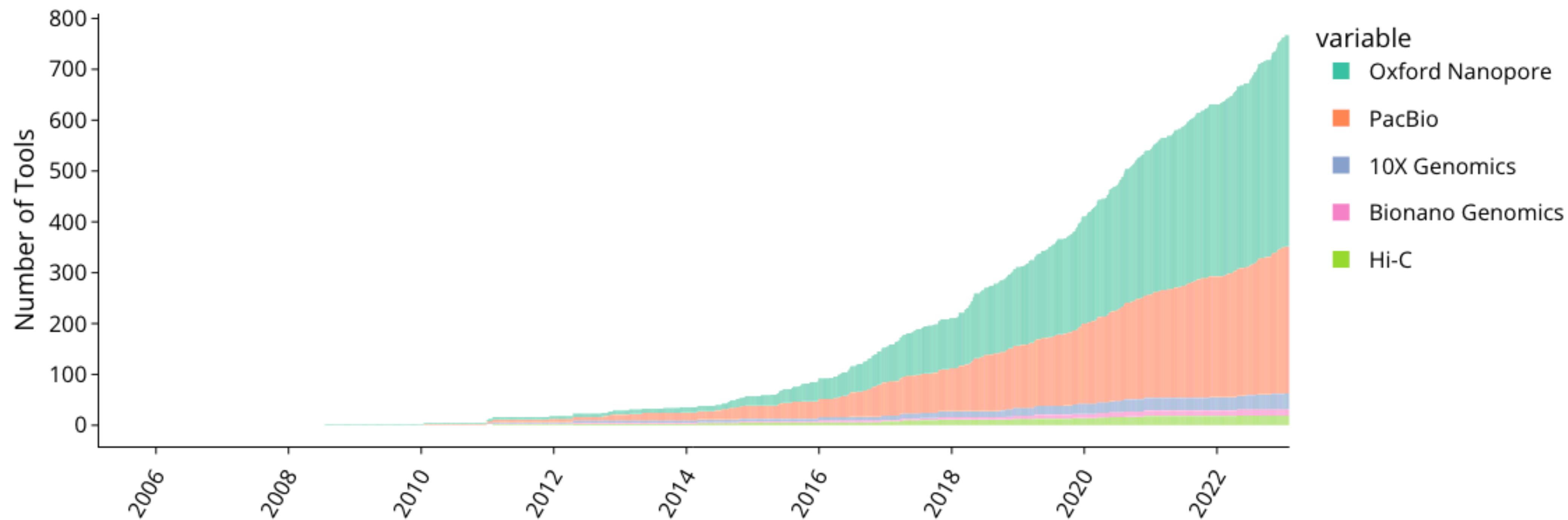
Search tools by full or partial name



HOME TABLE TOOLS TUTORIALS BENCHMARKS STATISTICS SUBMIT UPDATES FAQS CONTACT

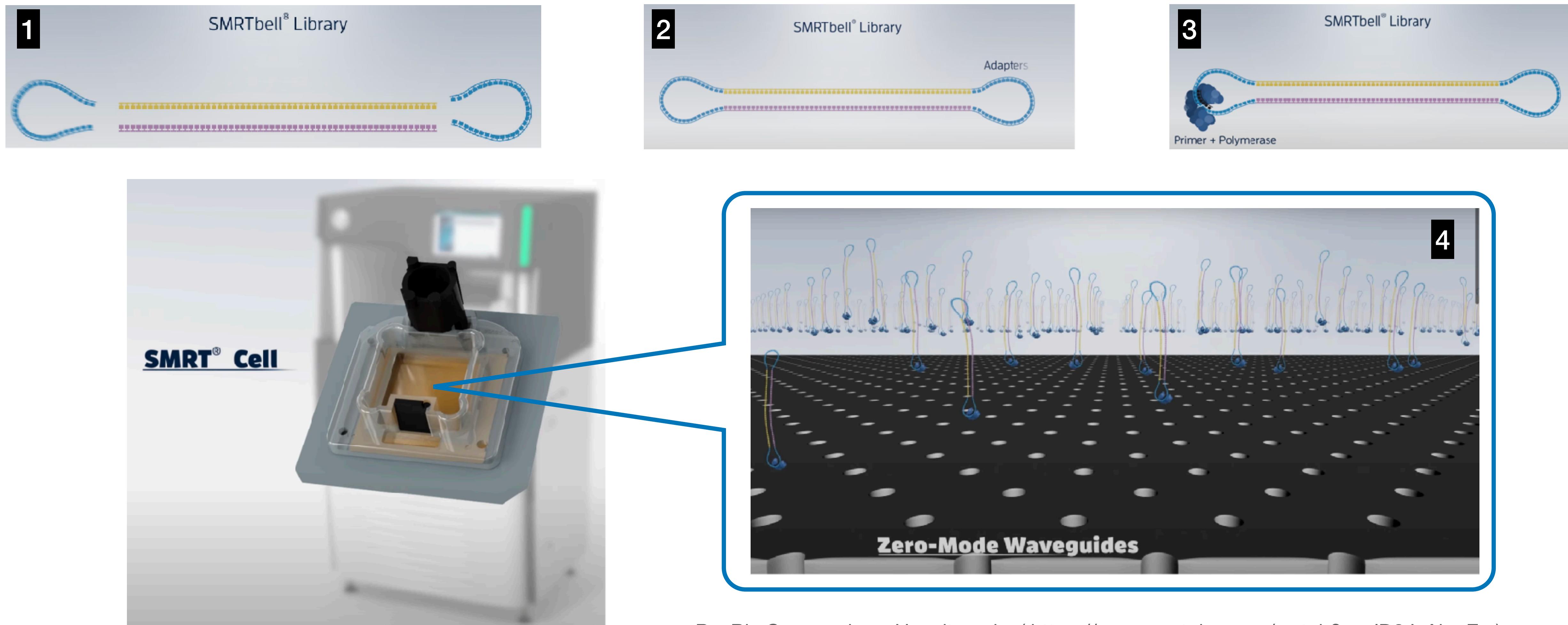


## Tools Across Time



# Long-read sequencing technologies

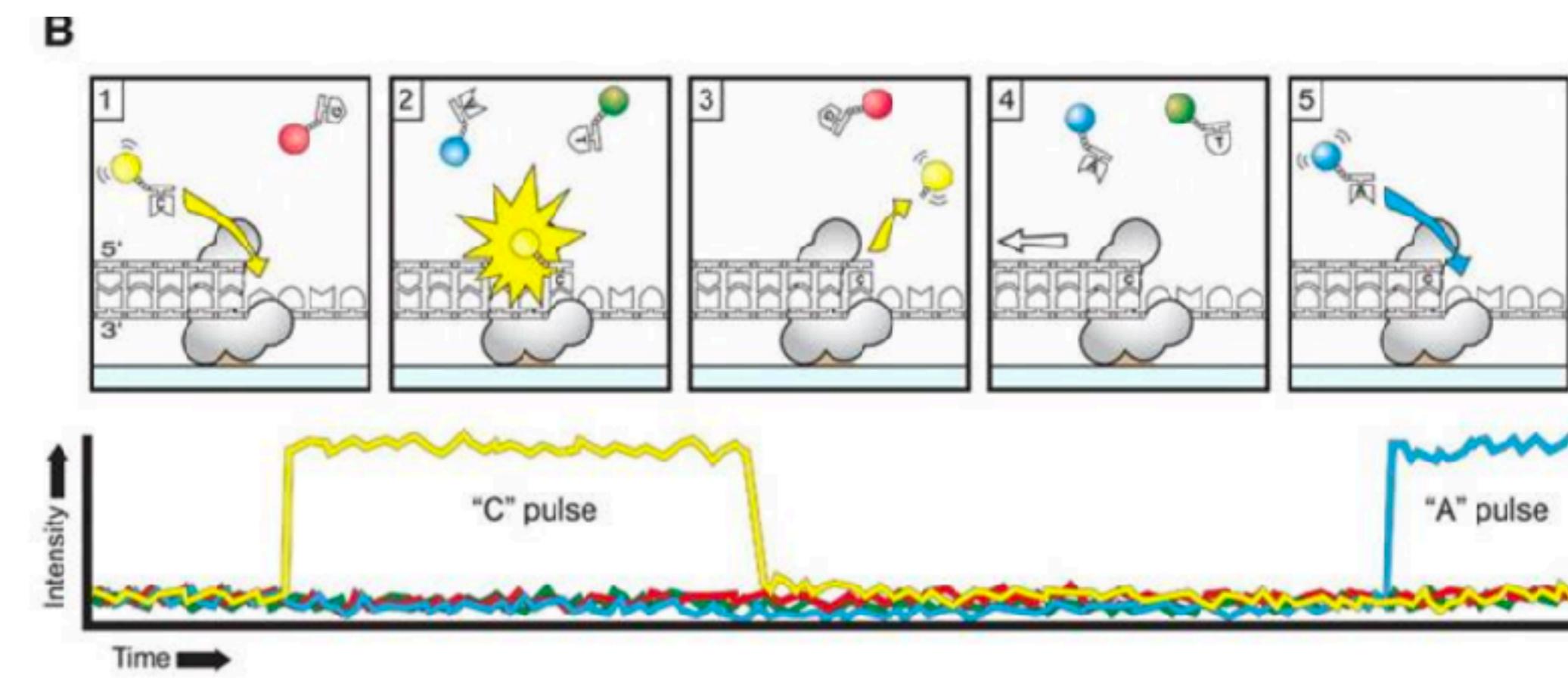
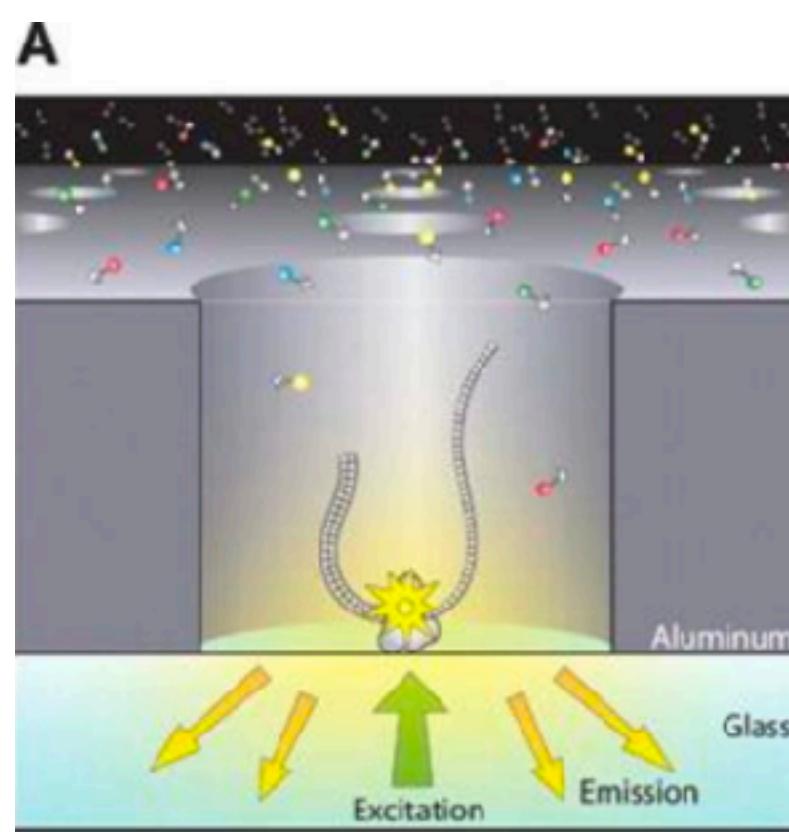
## Sequencing Principle of PacBio (1/2)



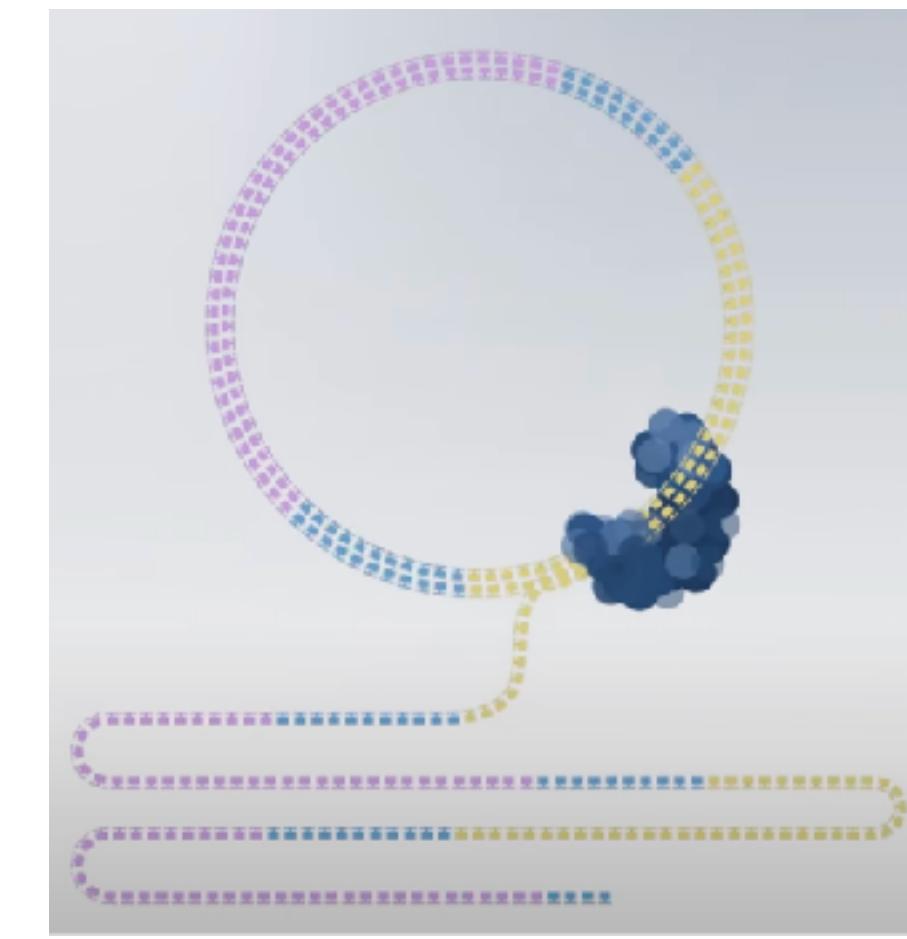
# Long-read sequencing technologies

## Sequencing Principle of PacBio (2/2)

Circular consensus sequencing (CCS)  
=> HiFi reads (>99% accuracy)



[Genomics Proteomics Bioinformatics](#). 2015 Oct; 13(5): 278–289

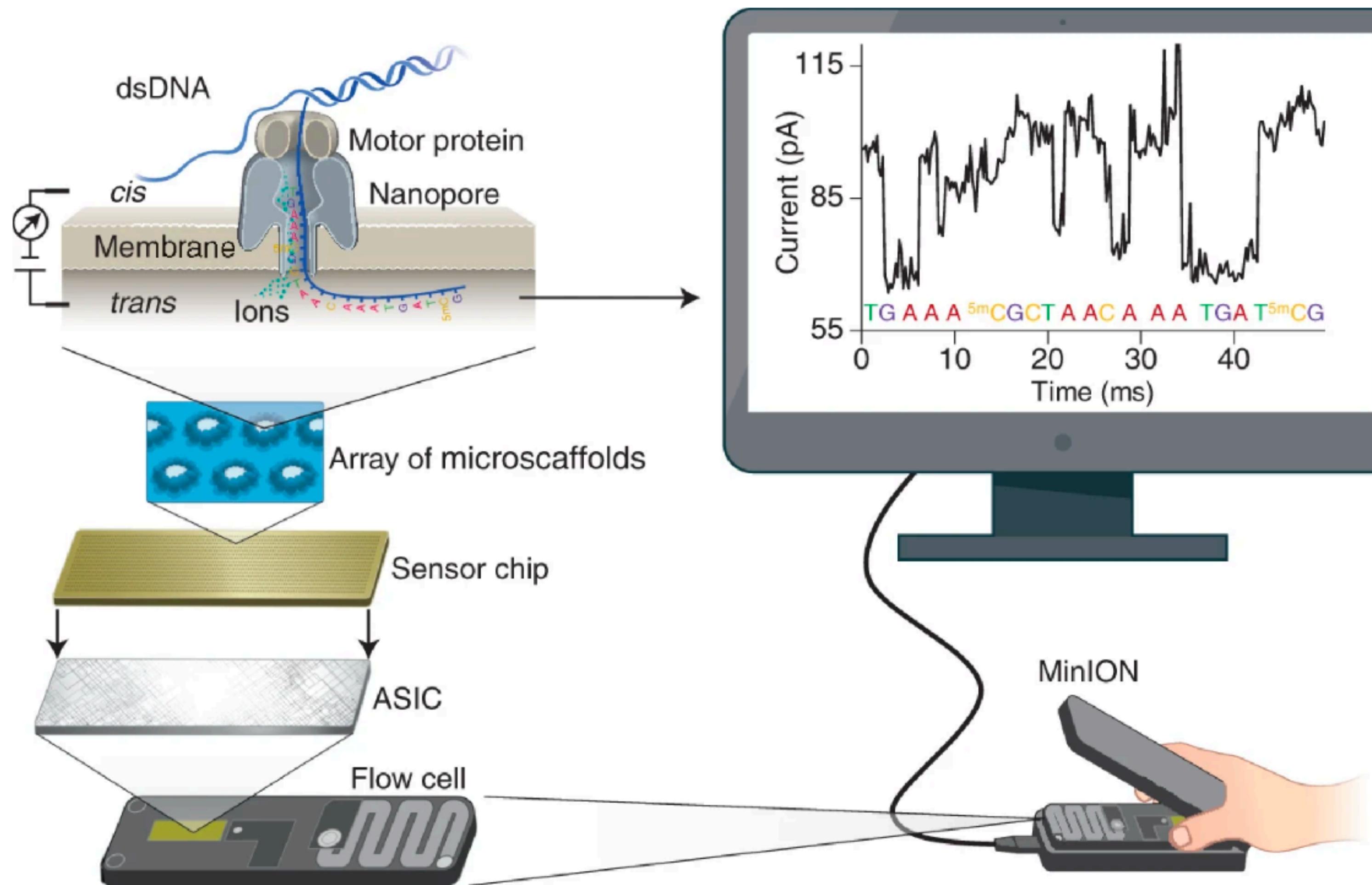


Continuous long-read (CLR) sequencing  
=> Long reads (>50% reads with length > 50kb)



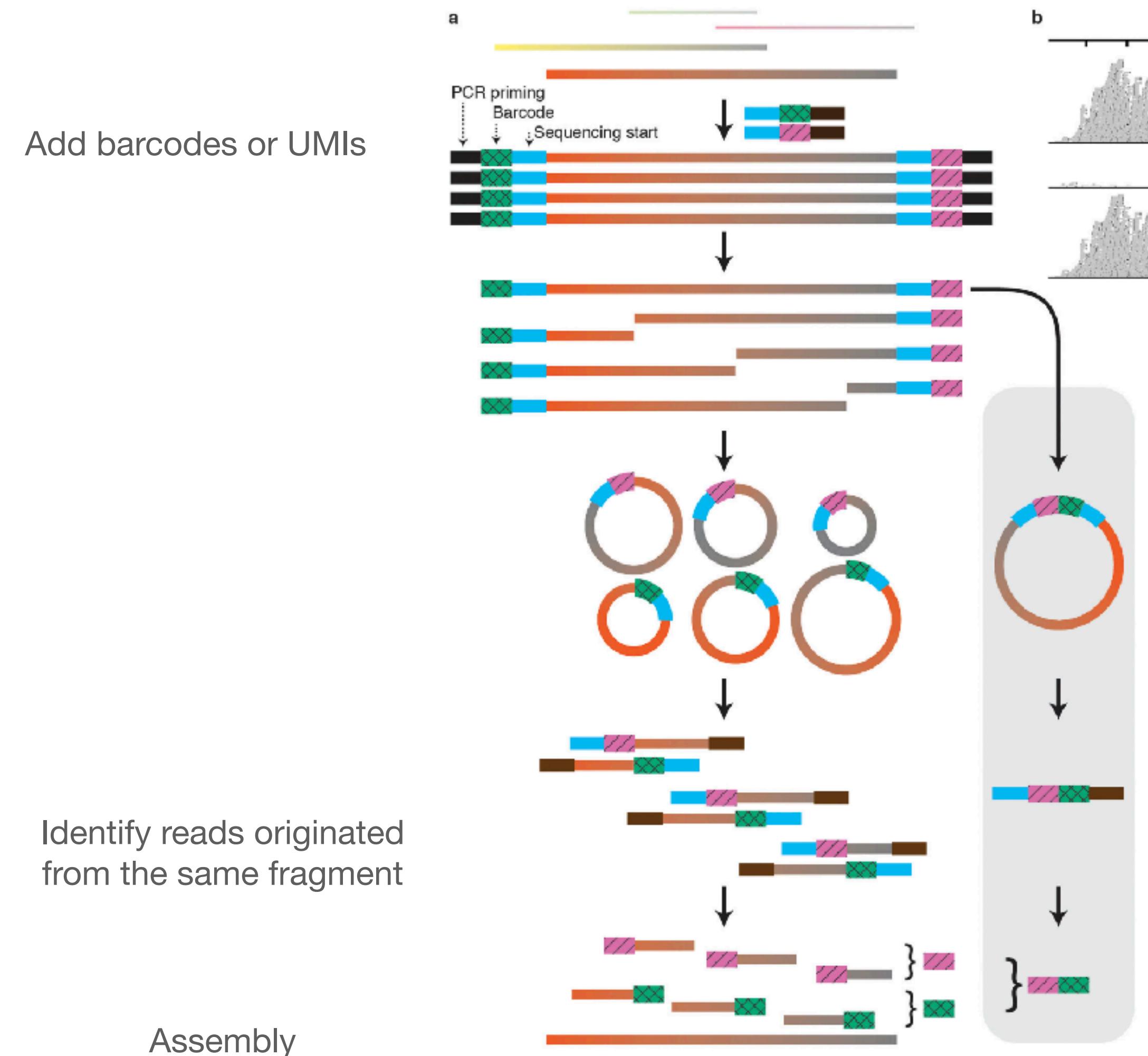
# Long-read sequencing technologies

## Sequencing Principle of Oxford Nanopore Technology



# Long-read sequencing technologies

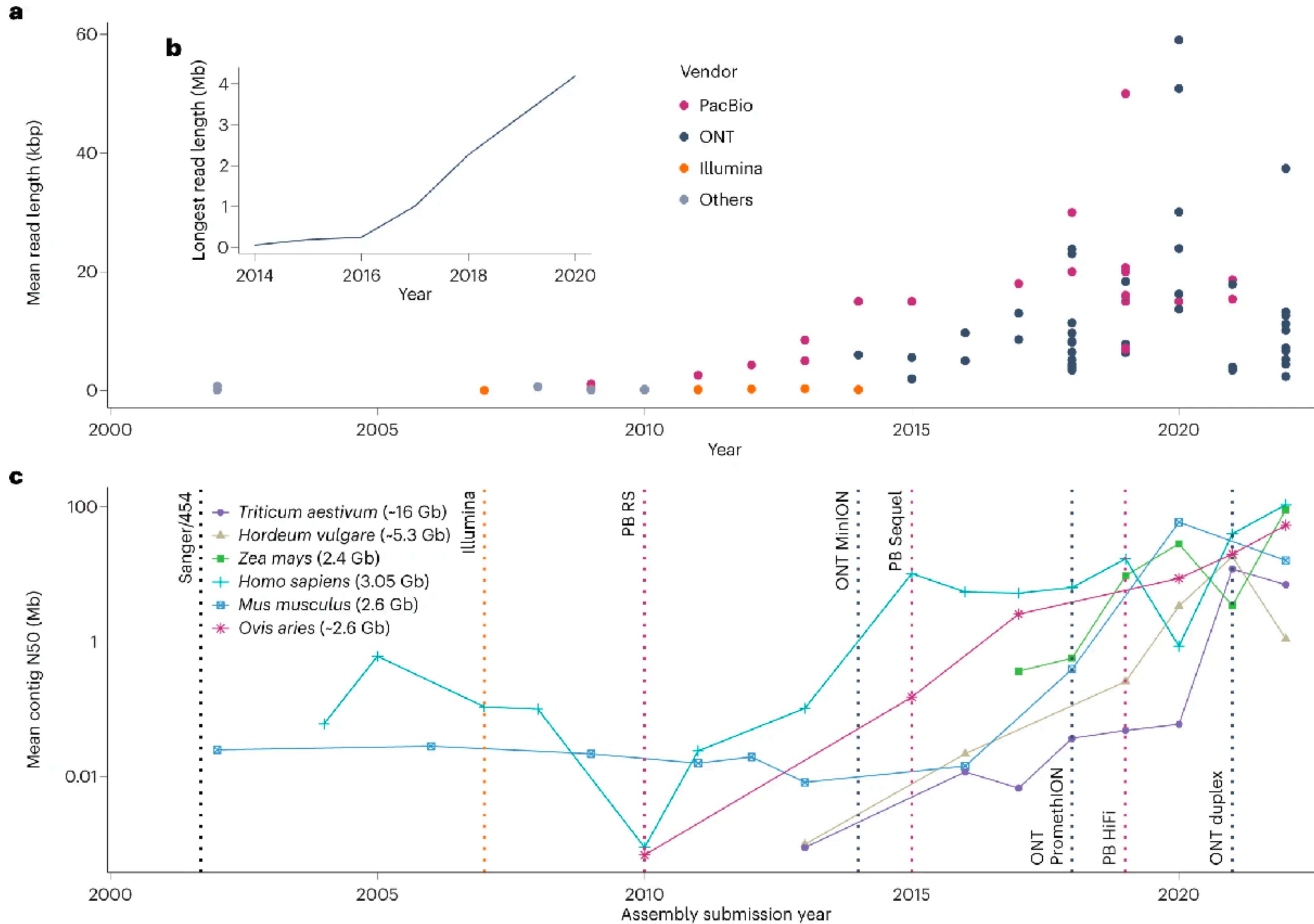
## Sequencing Principle of other technologies



- Based-on NGS
- Synthetic long-read sequencing

Element Bioscience, Ultima Genomics, MGI, Illumina etc.

# State-of-the-art LRS technologies (1/3)



# State-of-the-art LRS technologies (2/3)

- 25M ZMW per SMRT Cell, 4 SMRT Cells in parallel
- Generate 360Gb of HiFi reads per day
- \$1000 30x human whole genome sequencing with HiFi reads
- Nvidia GPU accelerated
- Include Google DeepConsensus
- Methylation detection
- Support short-read sequencing



PacBio

Credit: PacBio website

Nanopore

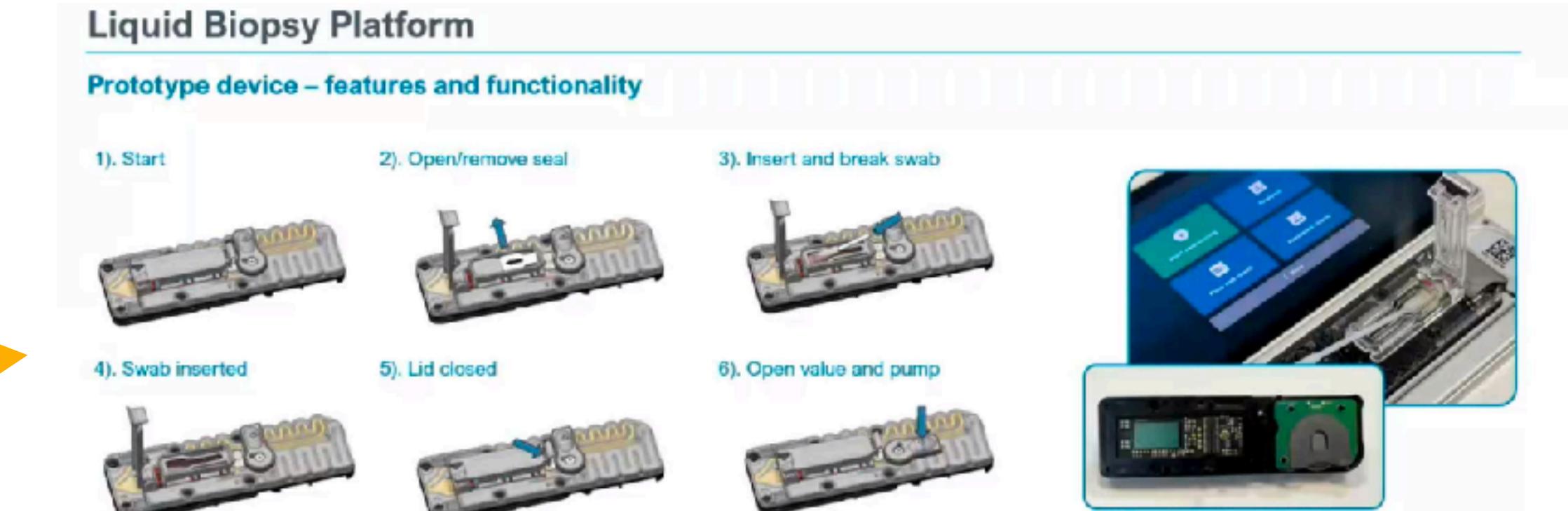


- Q20+ platform: V14 kit + R10.4 flow cell, > 237Gb throughput per PromethION FC
- Length of the longest read ~ 4Mb with Ultra-long kit
- Length of the shortest read ~ 20bp with short fragment mode
- Direct RNA-sequencing
- Nvidia GPU / Apple Silicon accelerated

Credit: ONT website

# State-of-the-art LRS technologies (3/3)

- More throughput generated
- Longer sequencing read
- Higher accuracy
- Detect epigenetic status simultaneously
- Support short-read sequencing
- Deep learning based algorithms
- Sample prep. free (?)



- Raw sample input via swab receptacle
- Integrated semi-permeable layer separating sample input chamber and sequencing chamber
- An additional set of electrodes for electrophoresis
- Battery powered circuit to generate electric field between chambers (in development)
- User actuated pumping mechanism (in design stage)

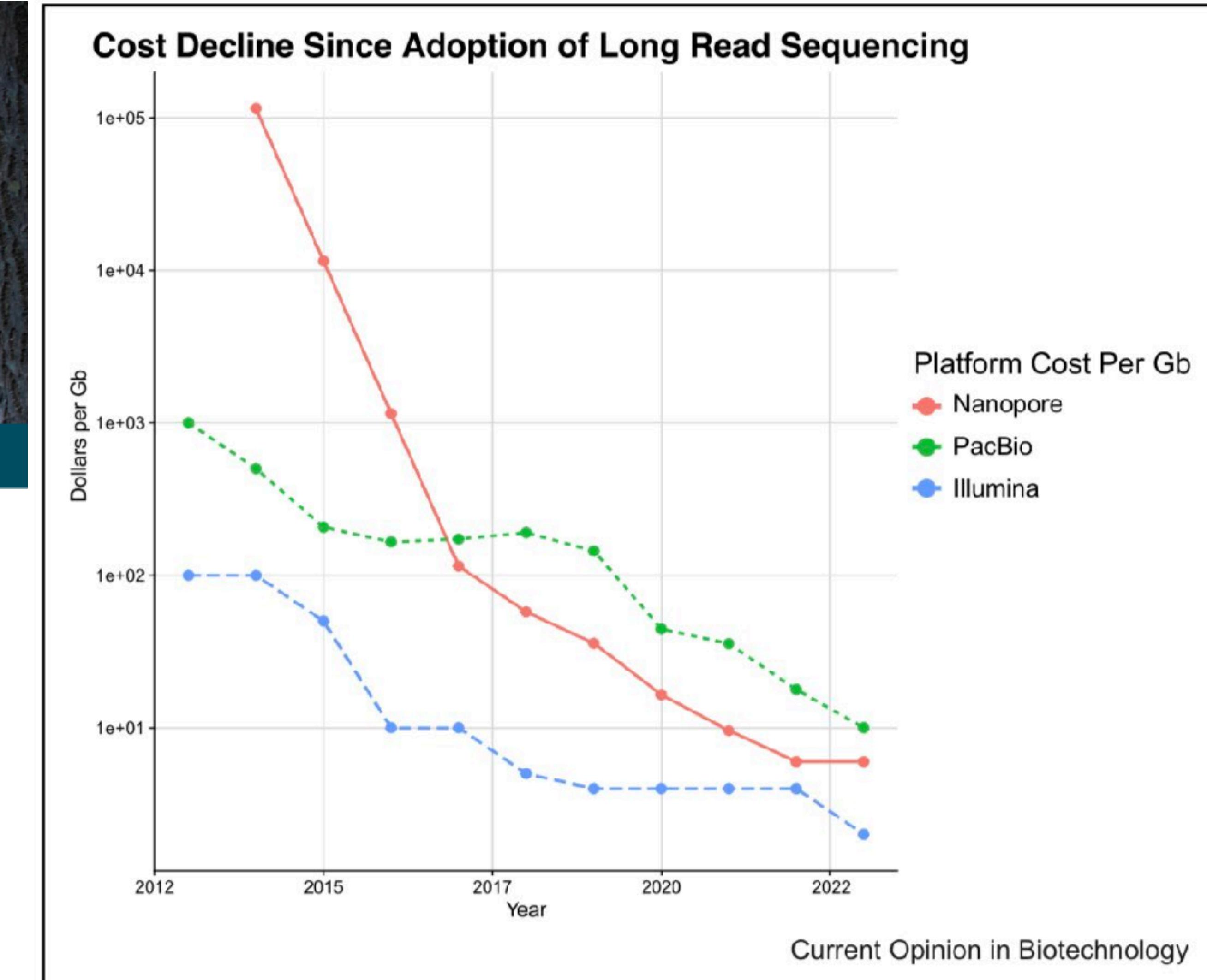
# Applications of LRS in Genomics

## Genome assembly (1/3)

<https://nanoporetech.com/oo/organisms>

The screenshot shows the OrgOne website interface. At the top left is the OrgOne logo. In the center is a large image of a golden snub-nosed monkey hanging from a tree. To the right of the monkey, the text "Golden Snub-nosed Monkey" and its scientific name "Rhinopithecus roxellana" are displayed, along with the status "ENDANGERED". Below the main image is a search bar with the placeholder "Search for a species". Underneath the search bar are five category buttons: "All (8637)" (highlighted in green), "Fungi (28)", "Plants (4880)", "Animals (3020)", and "Chromista (4)". The word "OrgOne" is centered below the category buttons. At the bottom of the page is a grid of nine smaller images showing various people and their dogs.

Dog Genome Project



<https://twitter.com/TJesse62/status/1616004729007476736/photo/1>

# Applications of LRS in Genomics

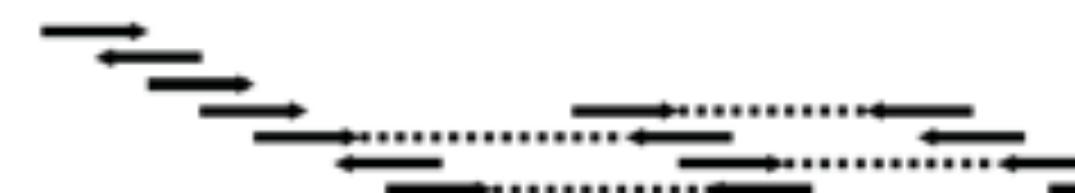
## Genome assembly (2/3)

### OLC-based methods

1. Find all read-read overlaps



2. Layout



3. Consensus



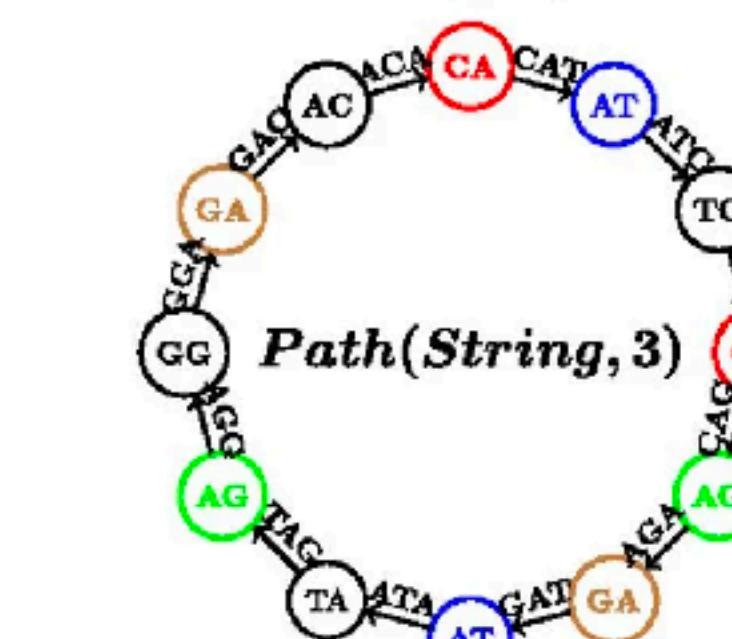
4. Scaffolds



[https://bioinformaticsworkbook.org/dataAnalysis/GenomeAssembly/Intro\\_GenomeAssembly.html#gsc.tab=0](https://bioinformaticsworkbook.org/dataAnalysis/GenomeAssembly/Intro_GenomeAssembly.html#gsc.tab=0)

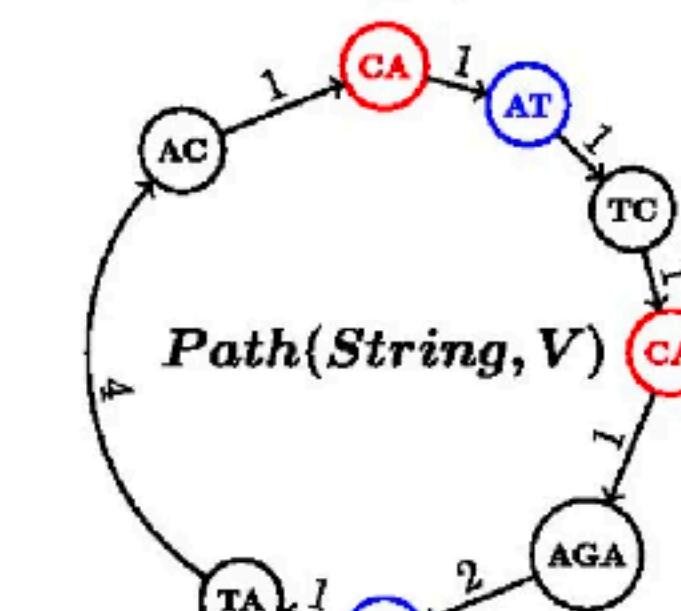
### De Bruin Graph-based methods

#### De Bruin graph

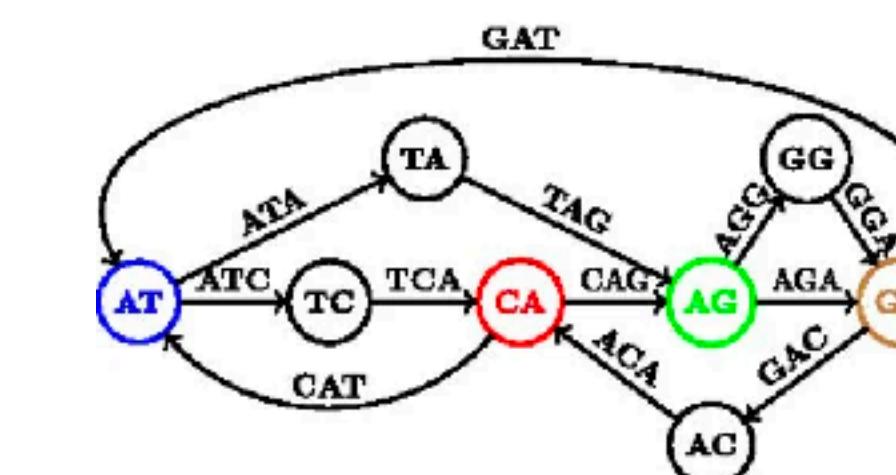
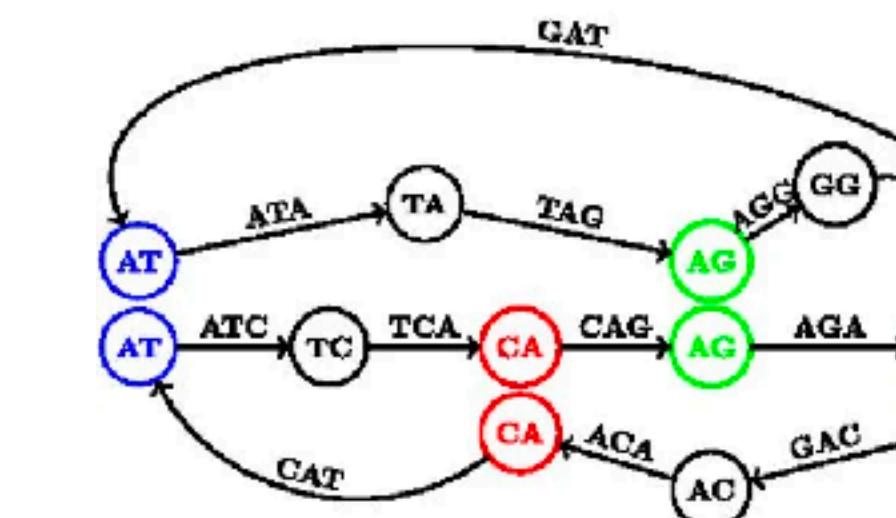


*Path(String, 3)*

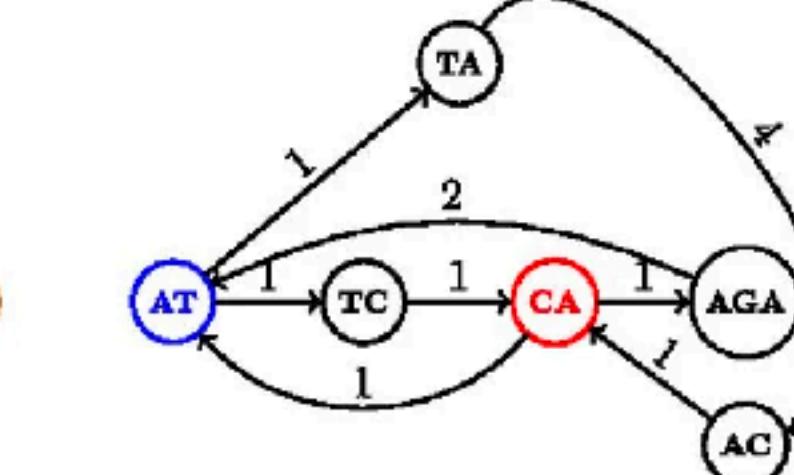
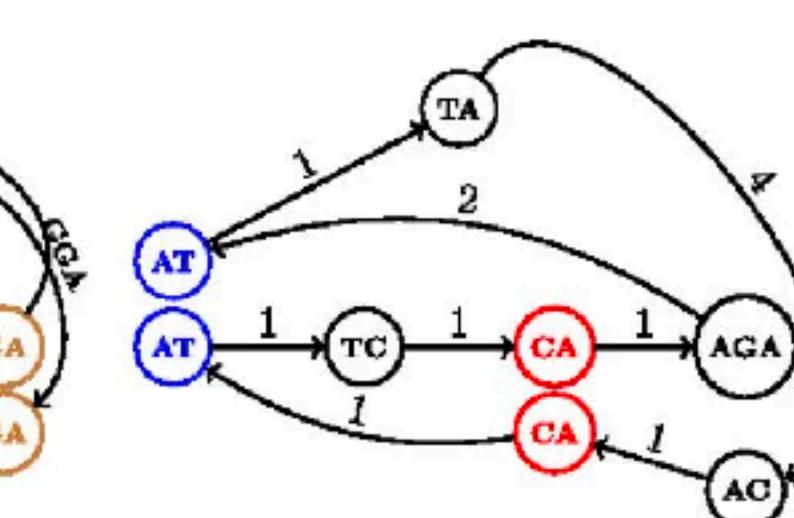
#### A-Bruijn graph



*Path(String, V)*



*DB(String, 3)*

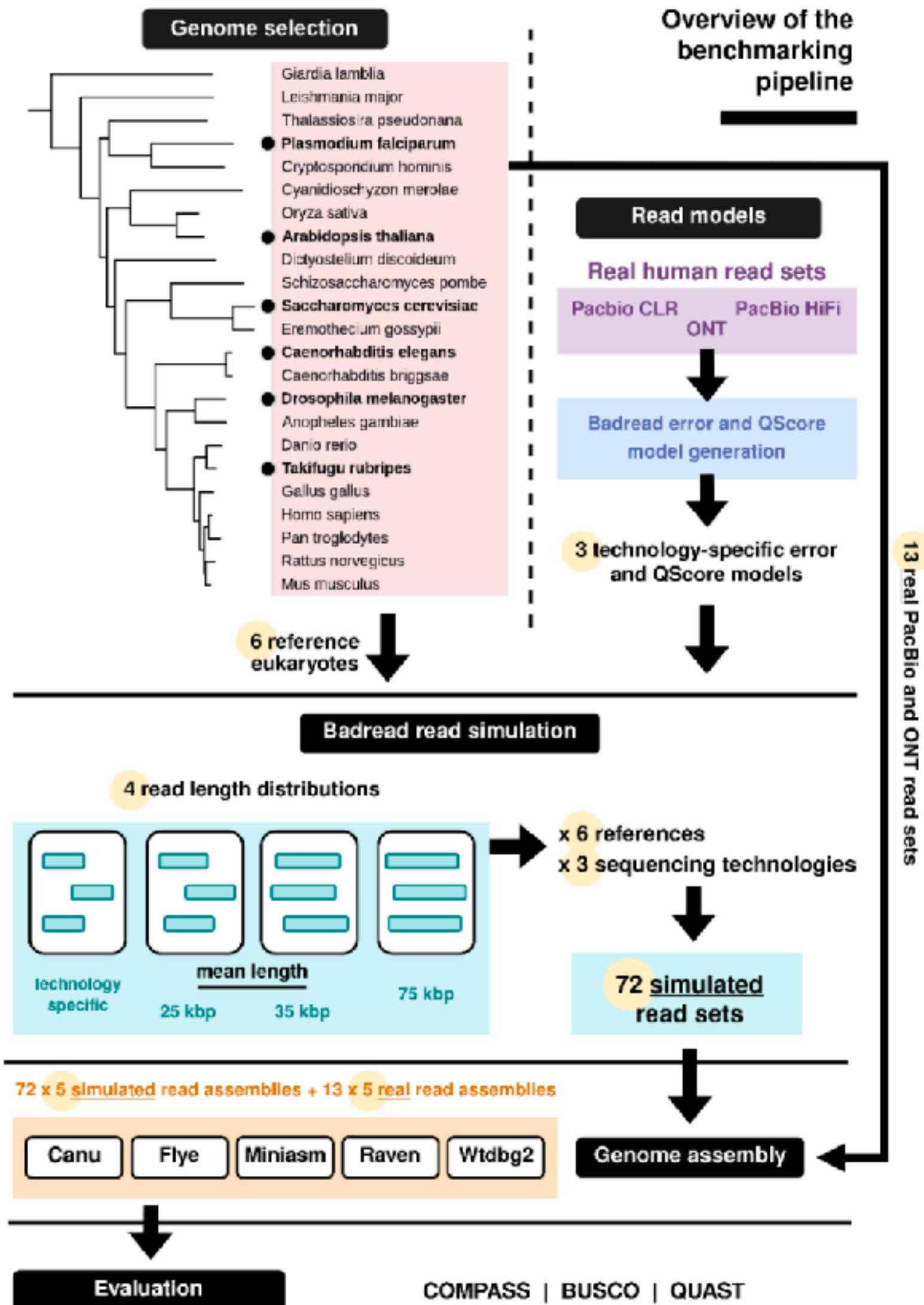


*AB(String, V)*

<https://www.pnas.org/doi/10.1073/pnas.1604560113>

# Applications of LRS in Genomics

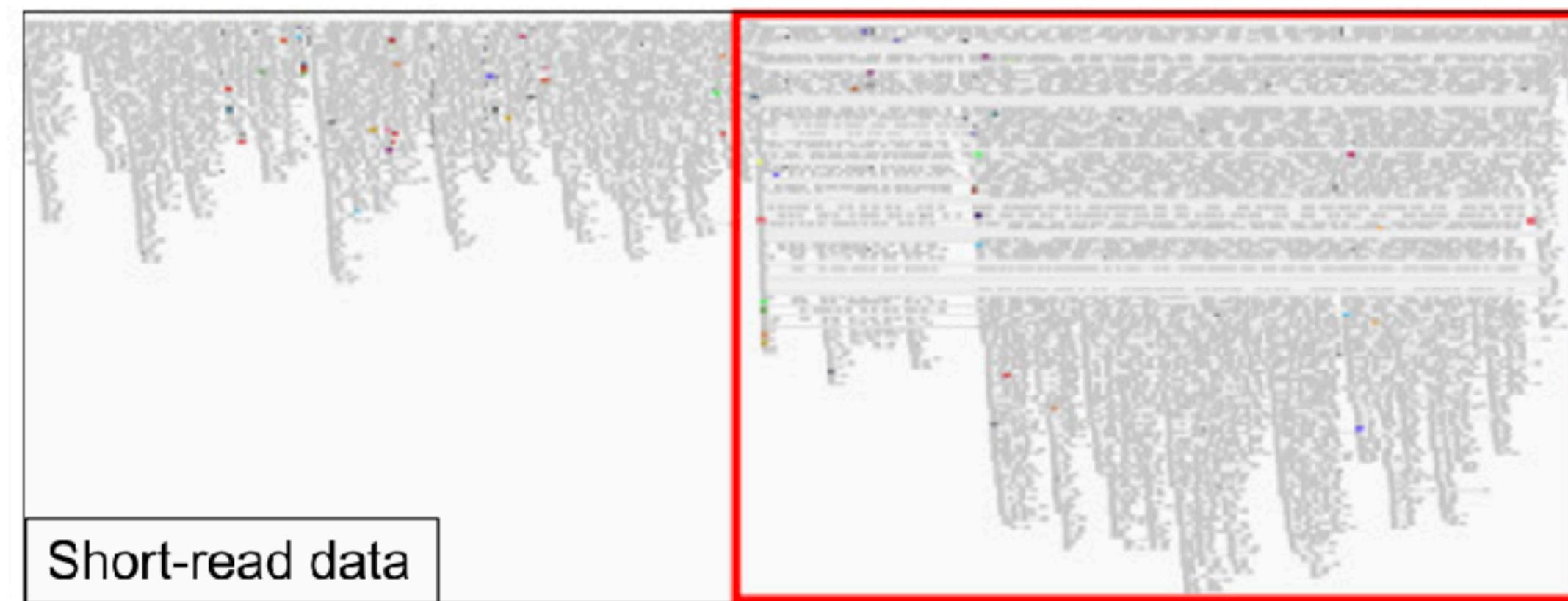
## Genome assembly (3/3)



# Applications of LRS in Genomics

## Detection of structural variants (1/3)

5 non-small cell lung cancer cell lines  
20 lung cancer clinical samples

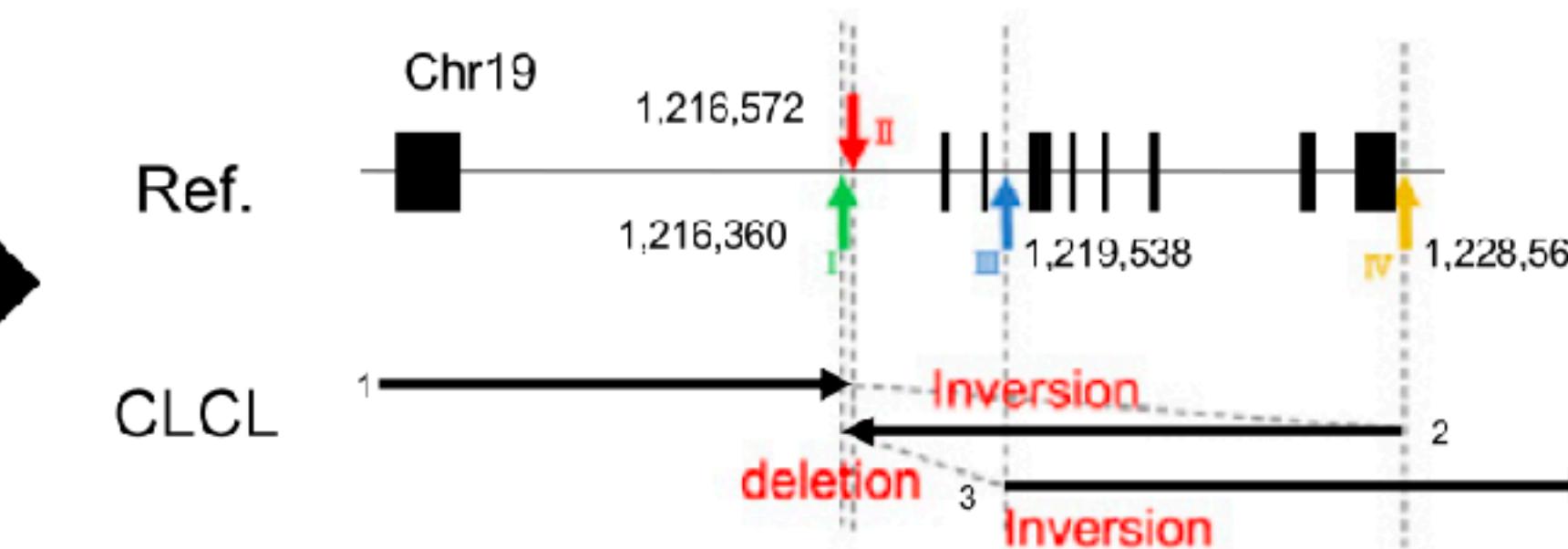


Copy-number aberration



CLCL (Cancerous Local Copy-number Lesions)  
Tandem duplication, inversion, micro deletion

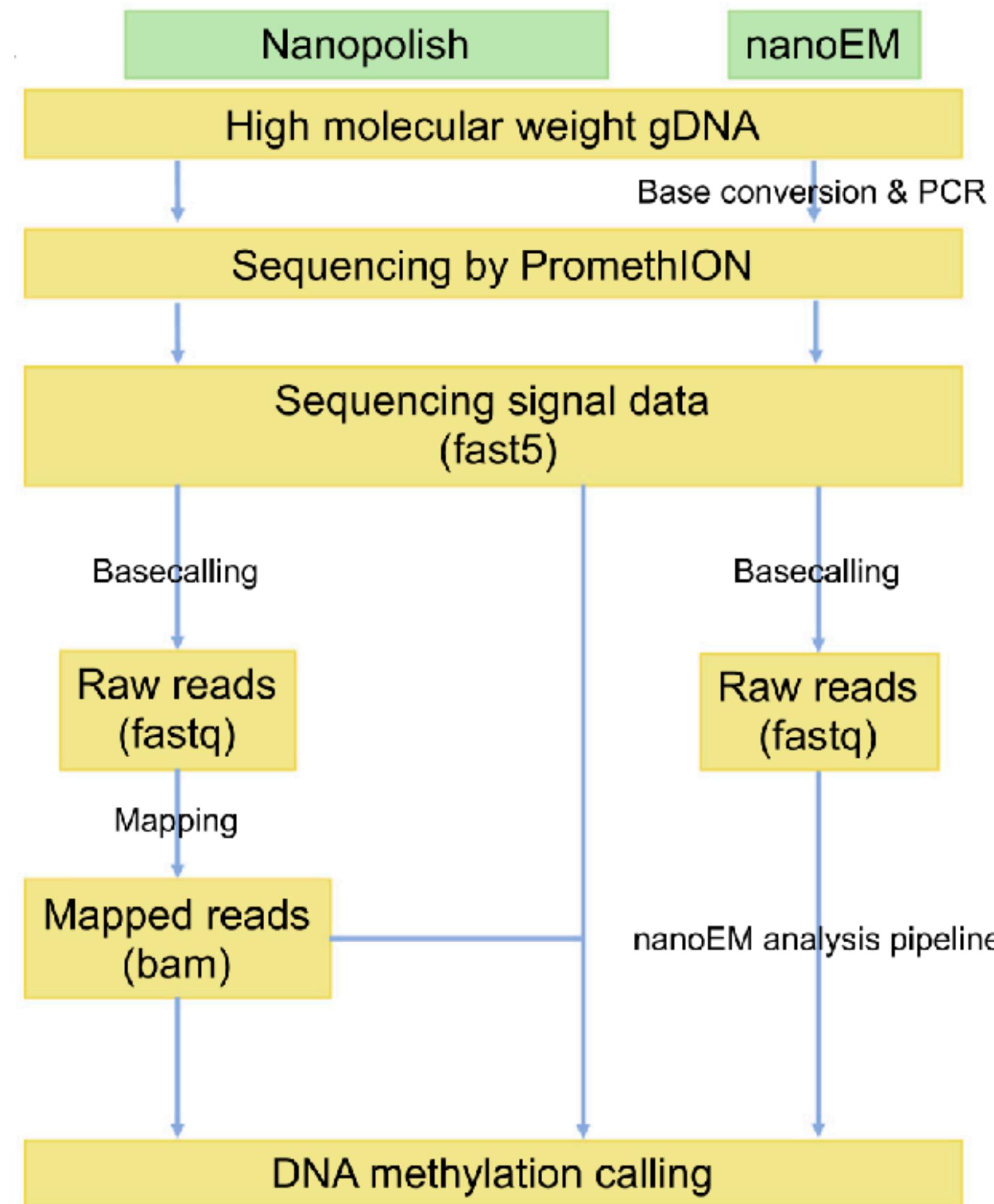
*STK11*



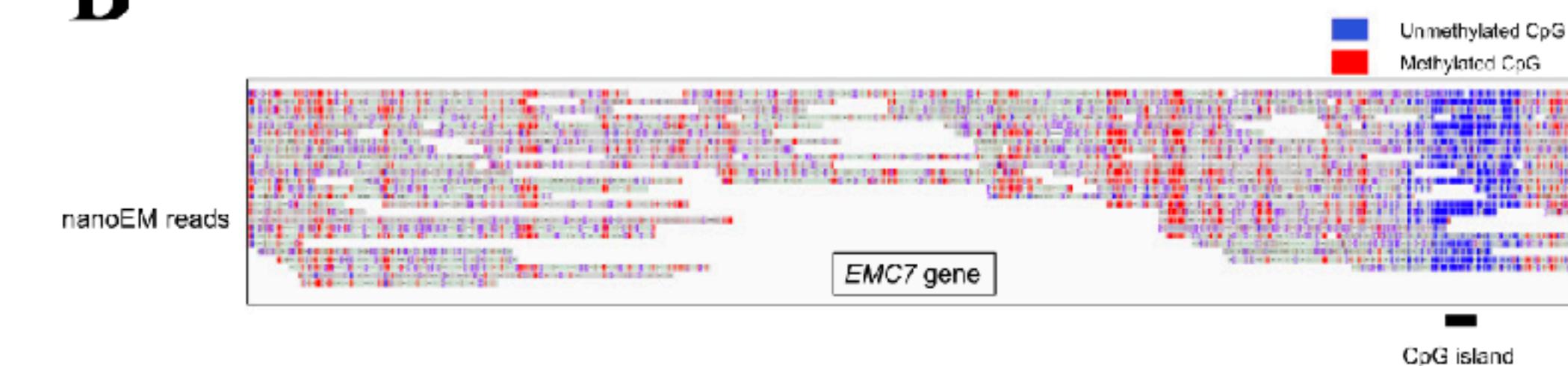
**Complete structure of SVs**

# Applications of LRS in Genomics

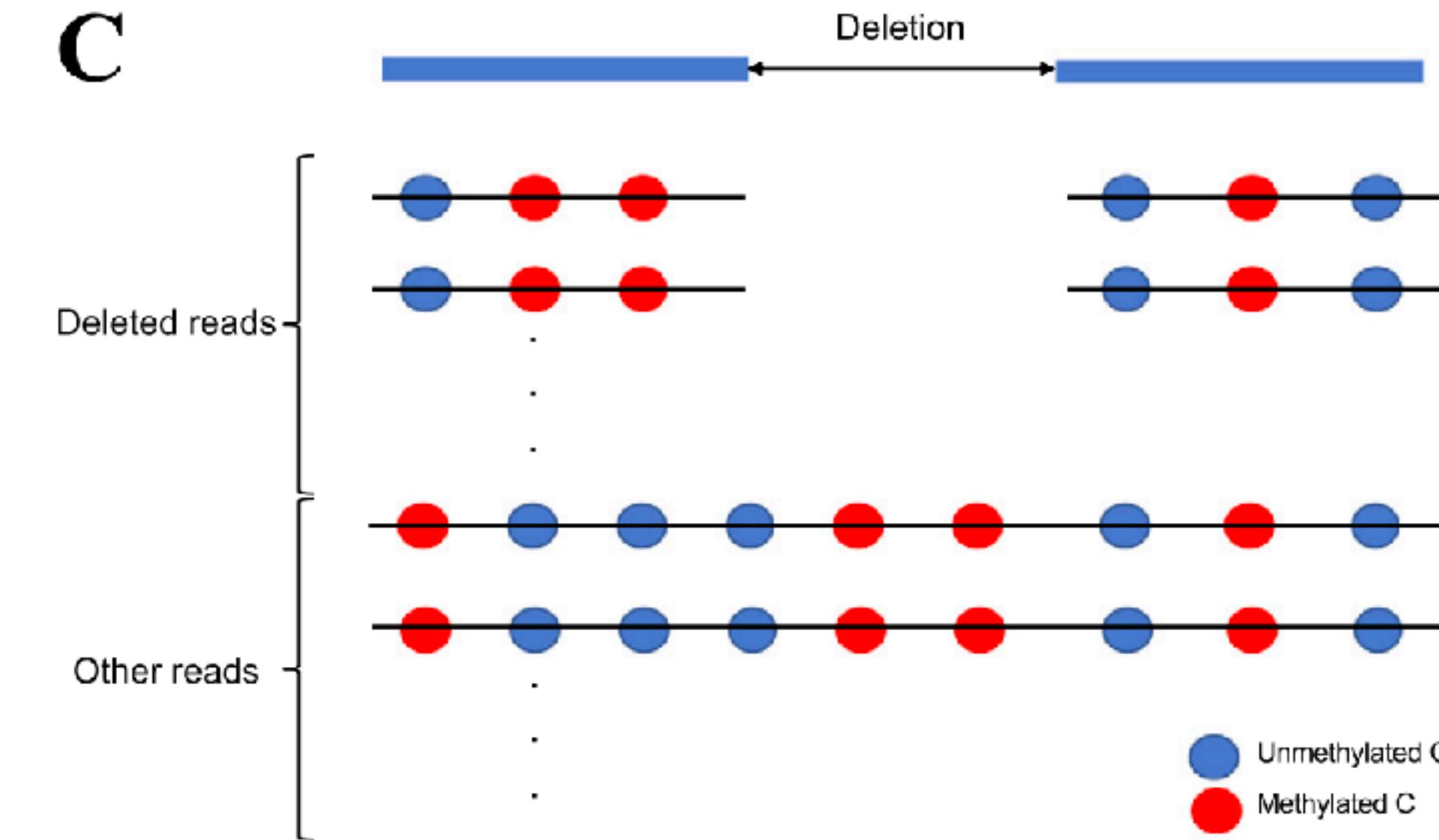
## Detection of structural variants (2/3)



B

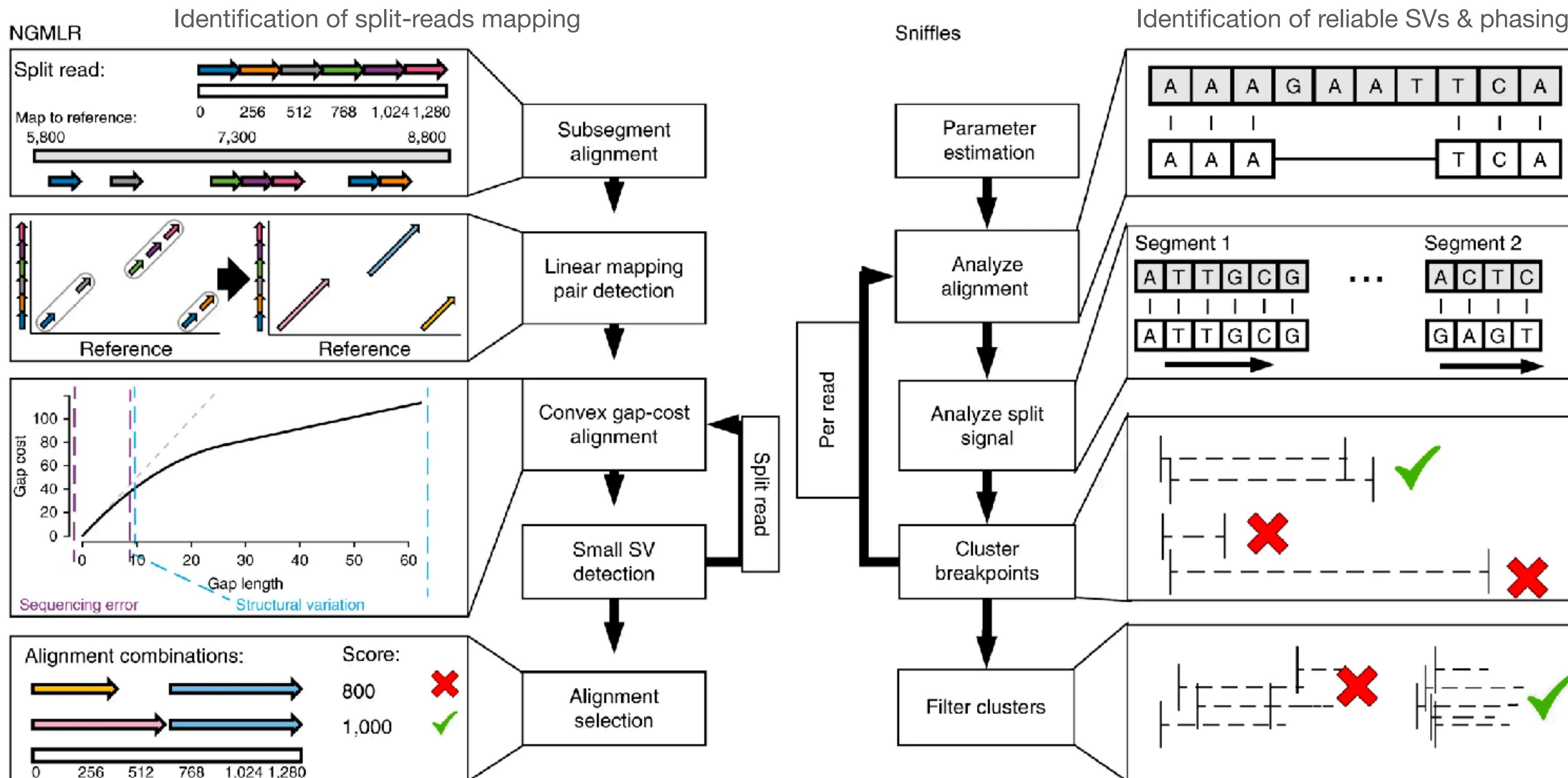


C



# Applications of LRS in Genomics

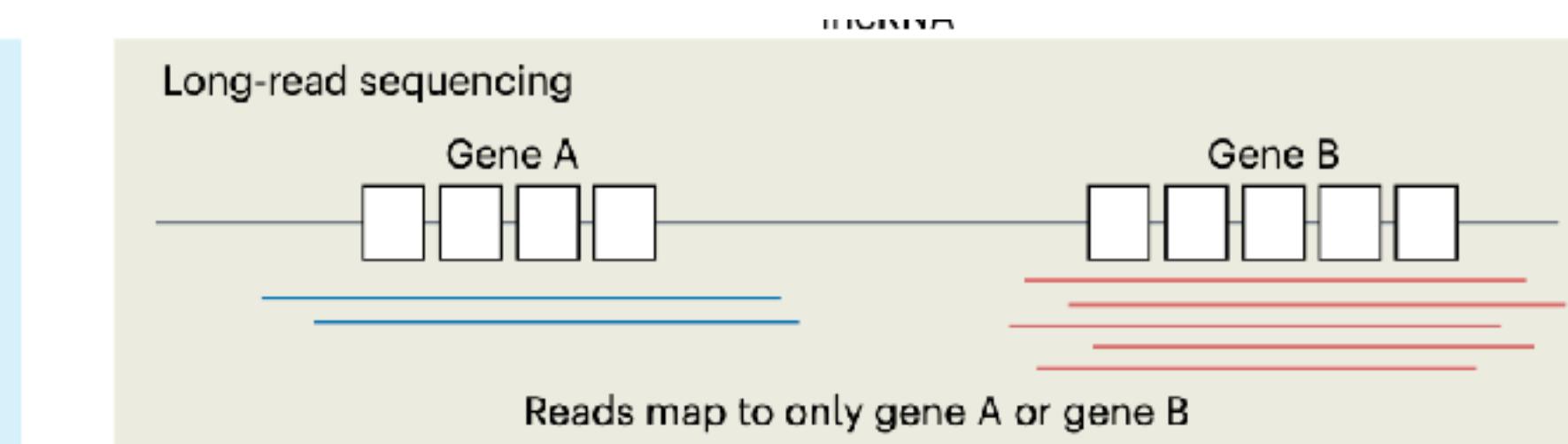
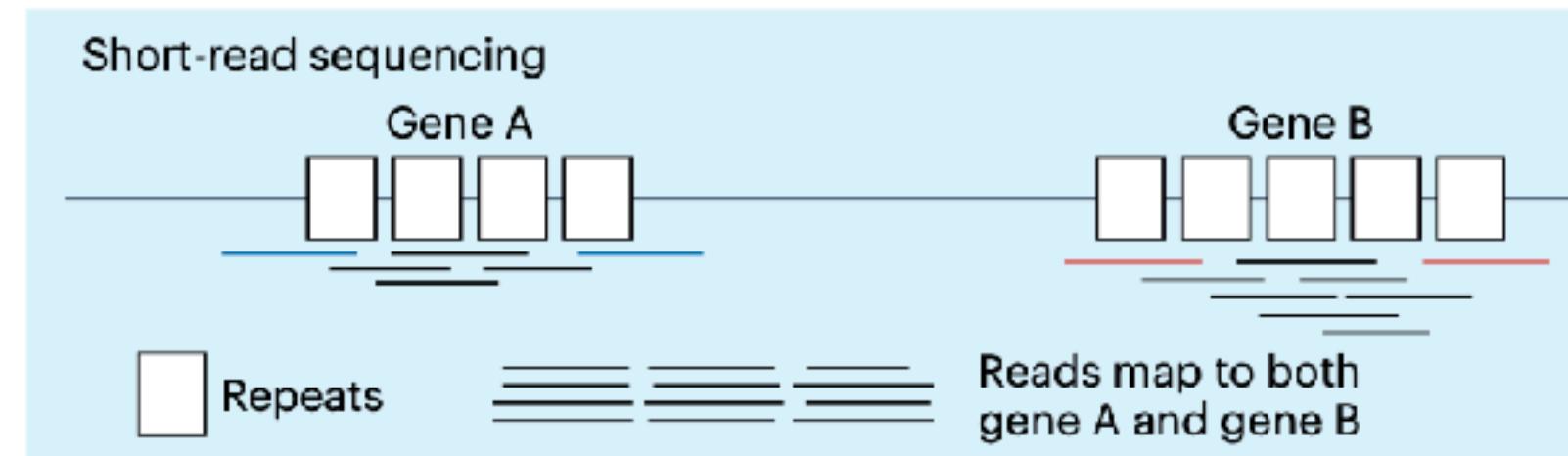
## Detection of structural variants (3/3)



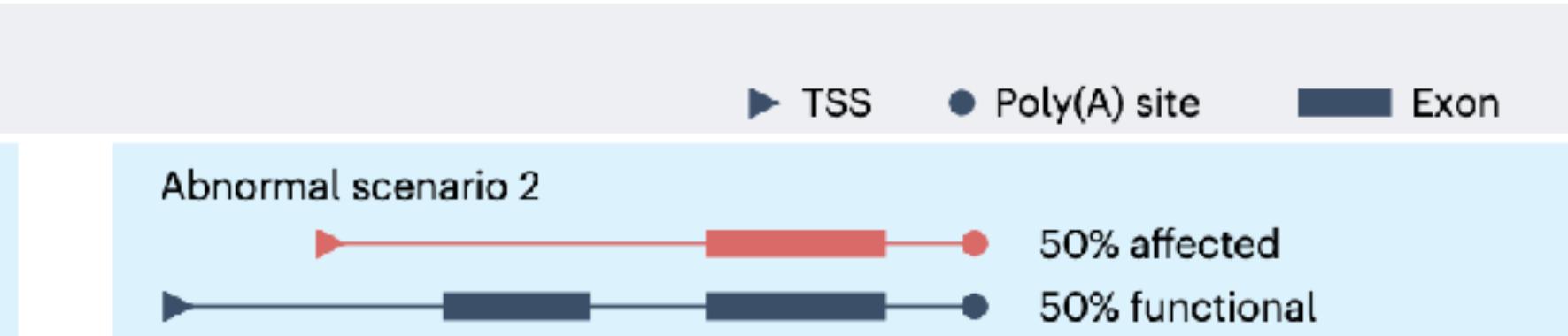
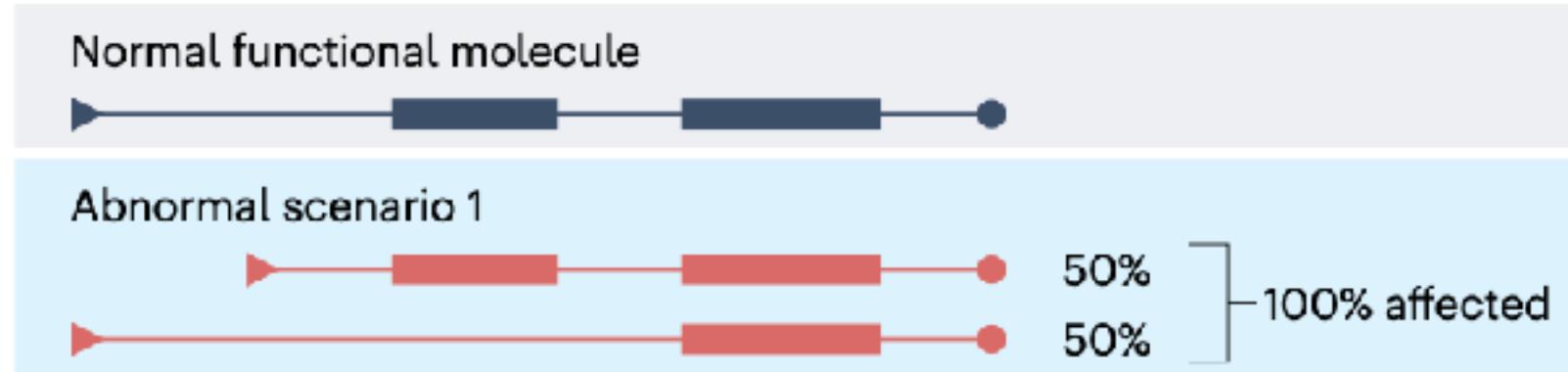
# Applications of LRS in Transcriptomics (1/4)

## Solve the problems that NGS cannot solve

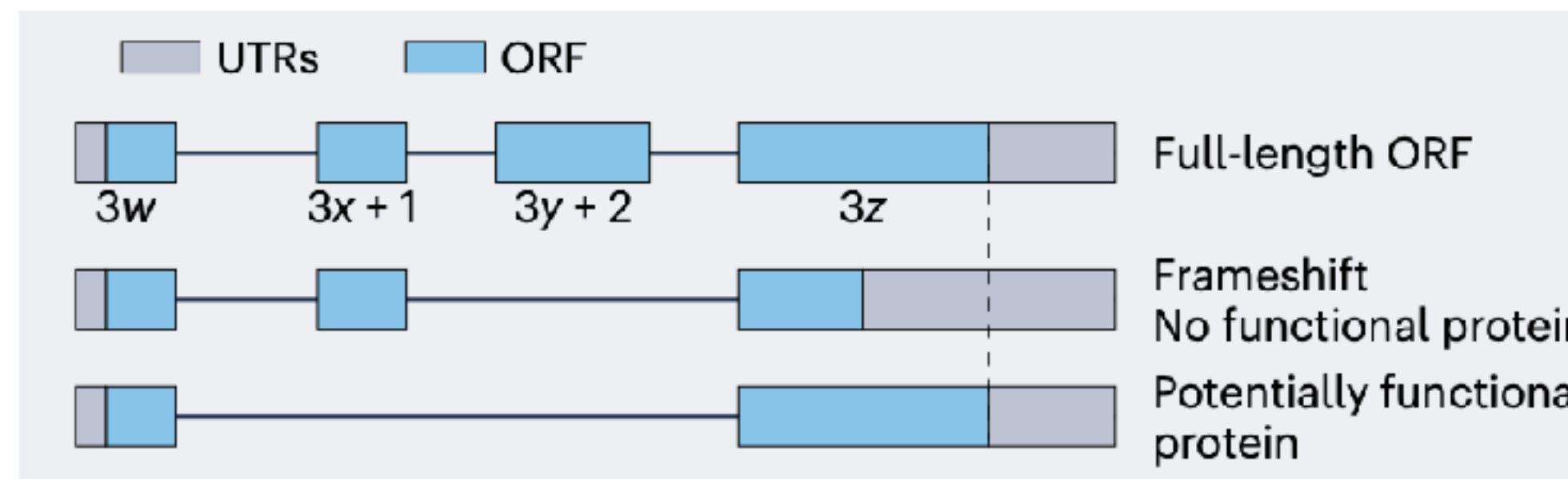
Condition 1



Condition 2



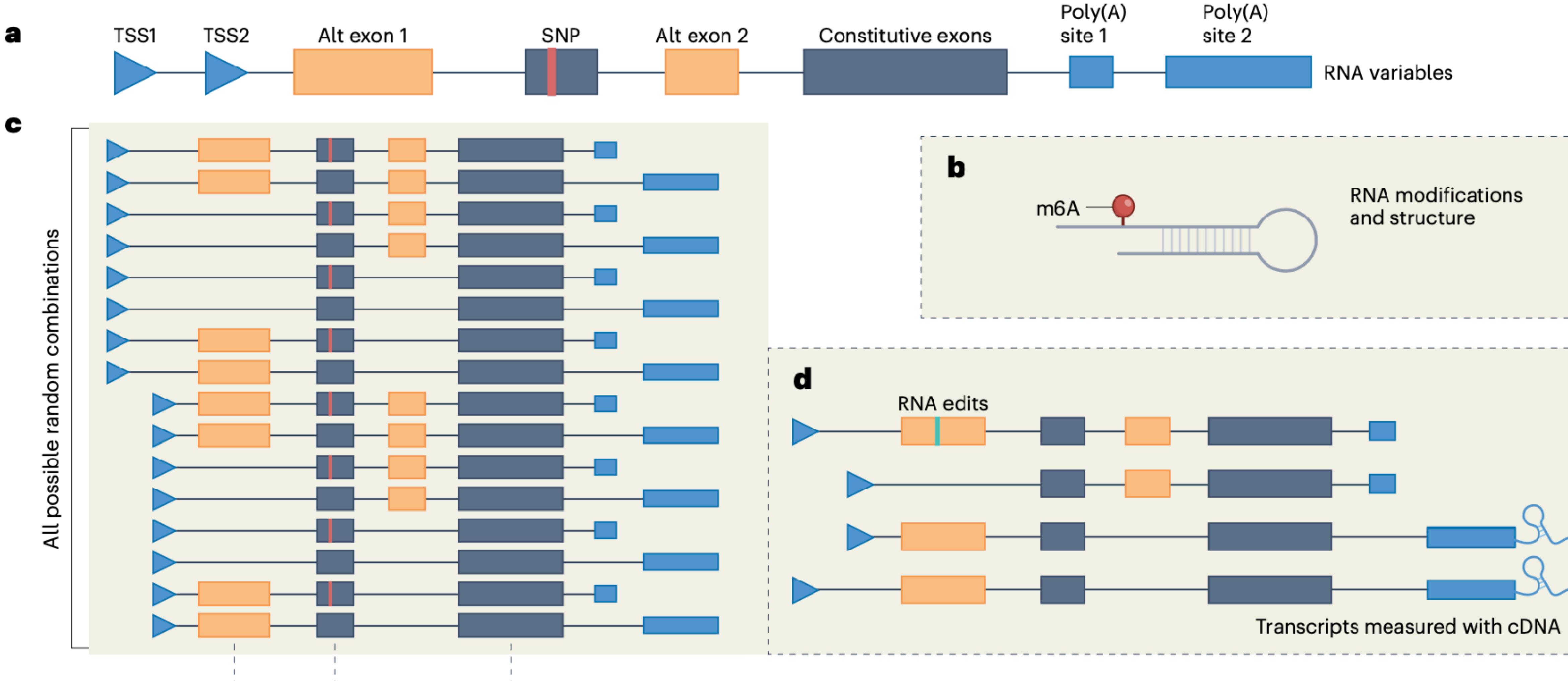
Condition 3



<https://doi.org/10.1038/s41592-022-01715-9>

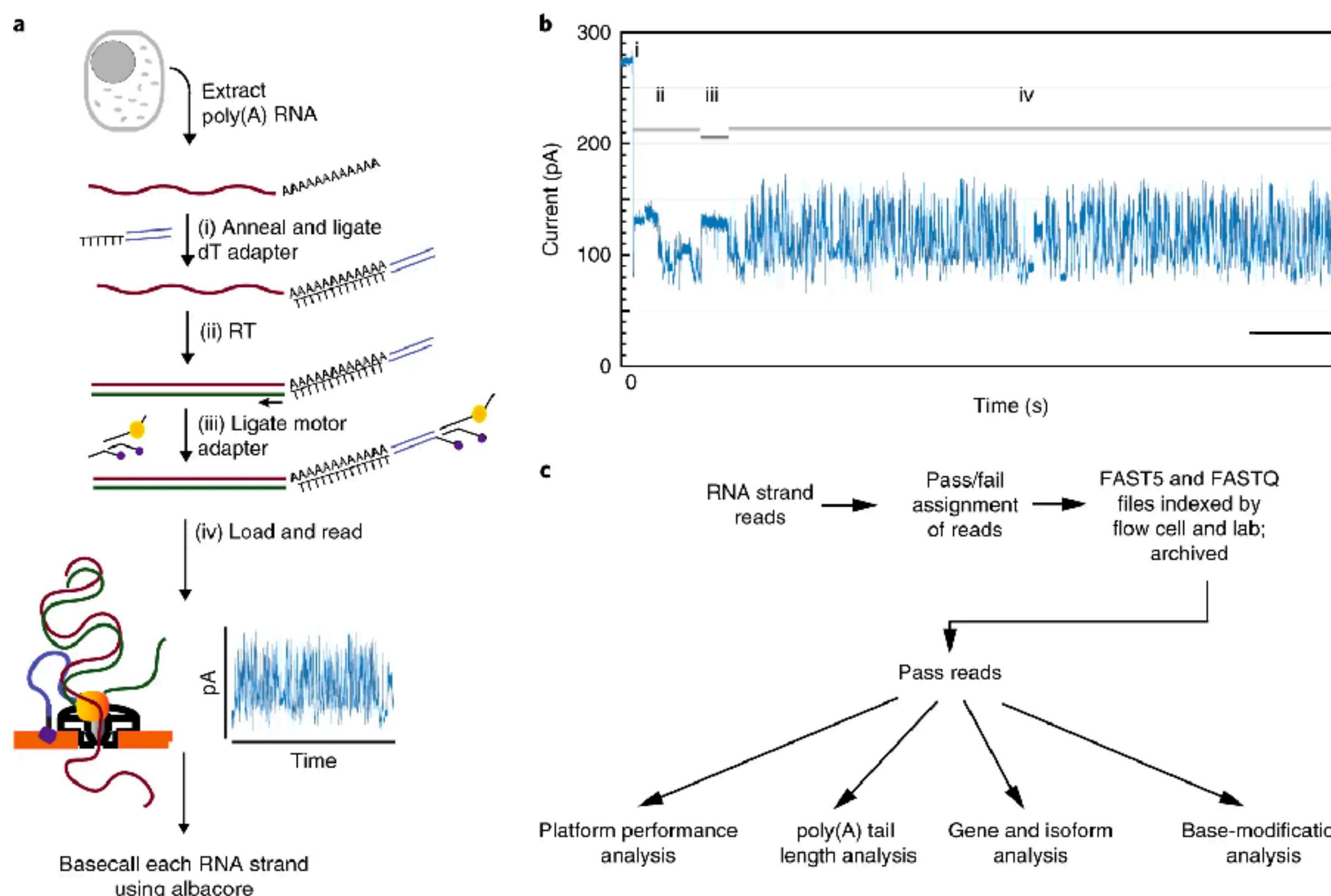
# Applications of LRS in Transcriptomics (2/4)

Solve the complexity of non-random pairing among RNA variables

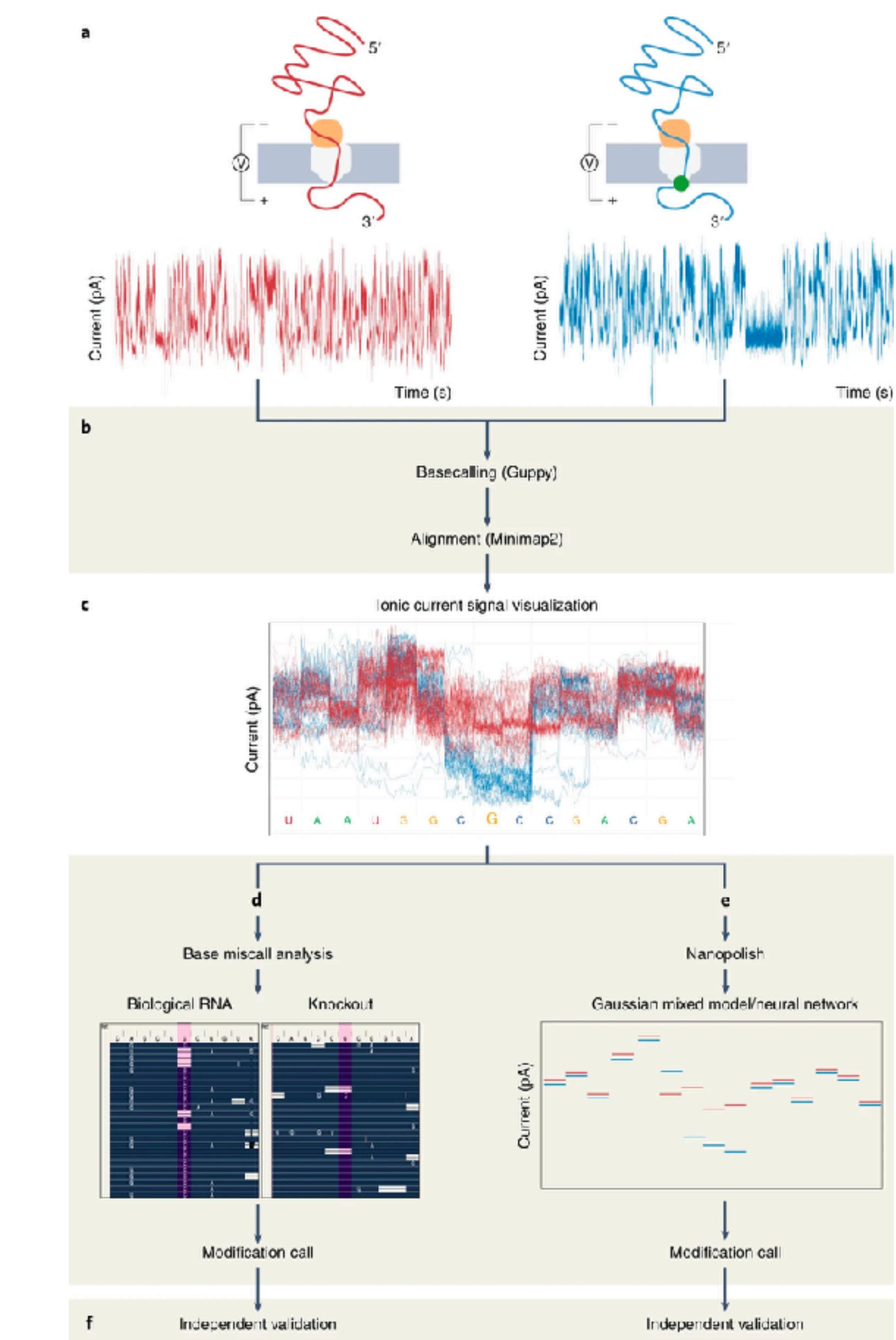


# Applications of LRS in Transcriptomics (3/4)

## Direct RNA sequencing



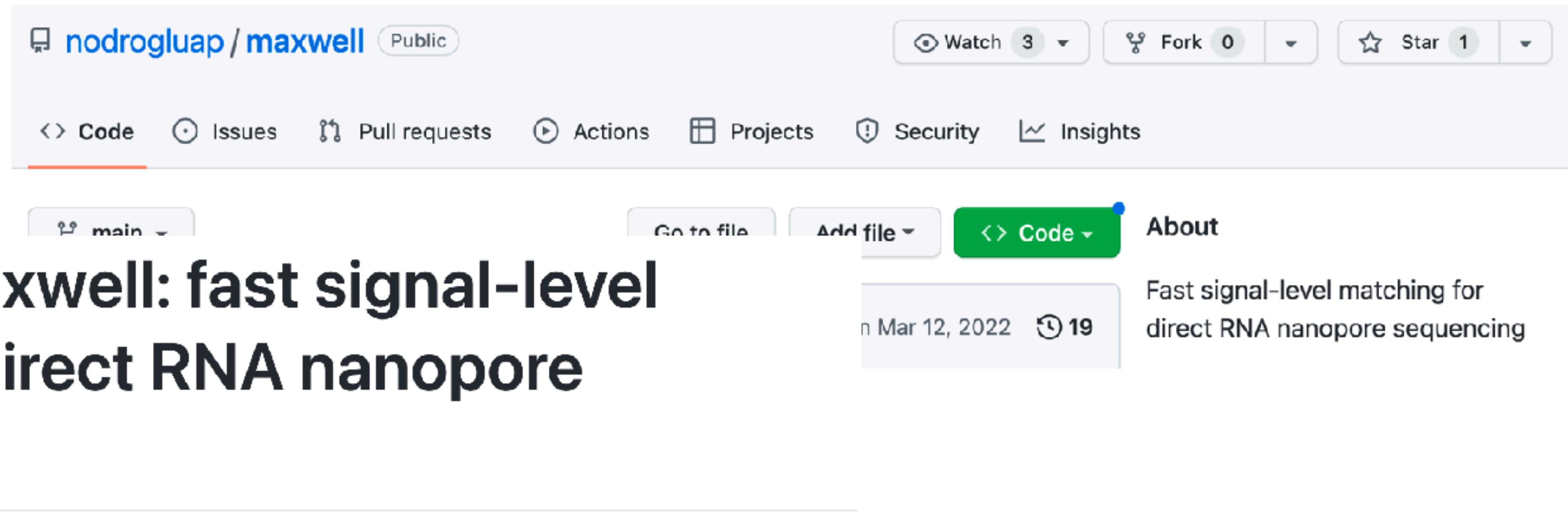
*Nature Methods* volume 16, pages 1297–1305 (2019)



*Nature Methods* volume 19, pages 1160–1164 (2022)

# Applications of LRS in Transcriptomics (4/4)

## Direct RNA sequencing + Read-Until



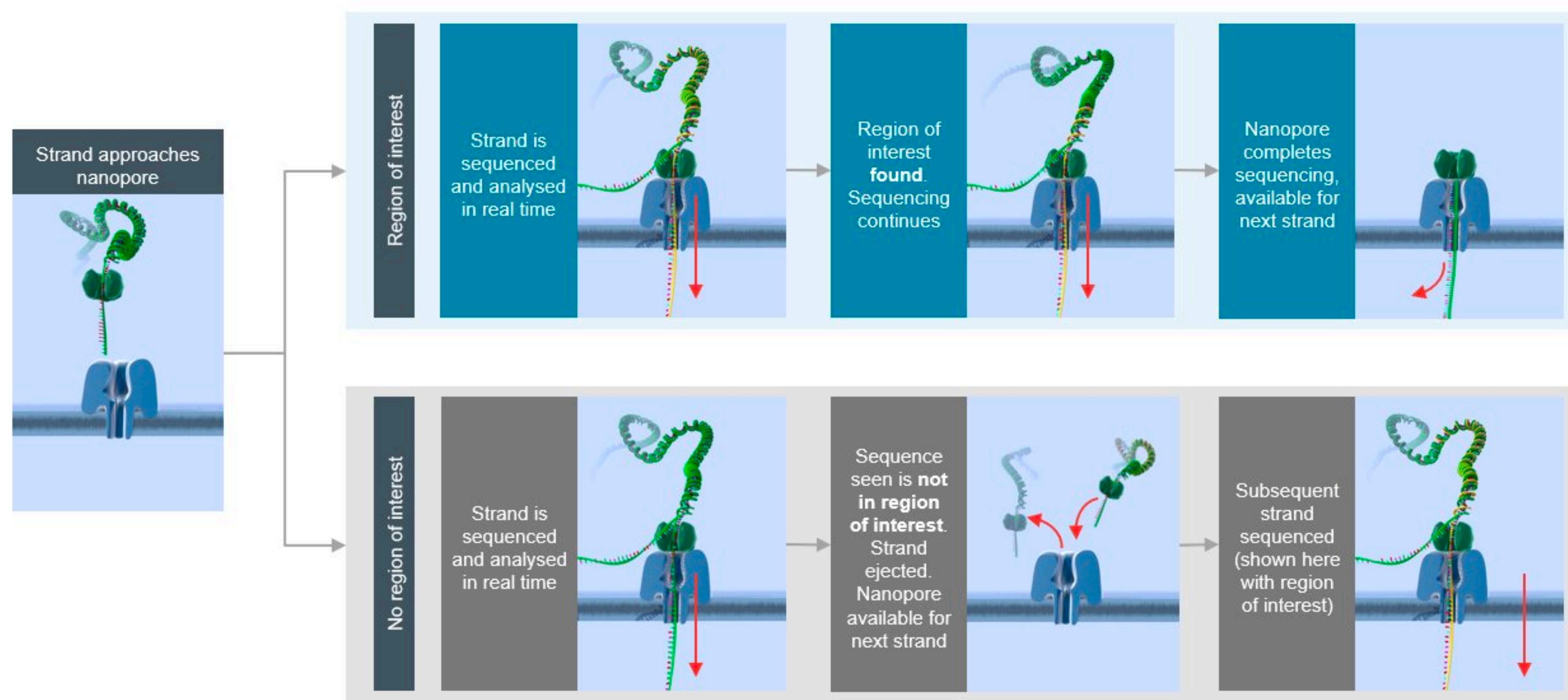
The screenshot shows a GitHub repository page for 'noderogluap/maxwell'. The repository is public and has 3 watches, 0 forks, and 1 star. The navigation bar includes links for Code, Issues, Pull requests, Actions, Projects, Security, and Insights. Below the navigation bar, there's a dropdown for the main branch, a 'Go to file' button, an 'Add file' button, and a 'Code' button which is highlighted in green. A timestamp indicates the page was last updated on Mar 12, 2022, at 19:19. The main content area features a large title: 'Magenta & Maxwell: fast signal-level matching for direct RNA nanopore sequencing'. To the right of the title is a brief description: 'Fast signal-level matching for direct RNA nanopore sequencing'.

Magenta is novel GPU code for very fast Dynamic Time Warp (DTW) matching of a query signal (e.g. a nanopore direct RNA raw signal) against a much larger reference signal database (e.g. a reference human transcriptome).

Maxwell is a standalone executable that wraps the Magenta code and the MinKNOW API in C++ to perform real time signal-level streaming DTW alignment of Oxford Nanopore Technologies (ONT) device data, so that include or exclude sequencing criteria can be applied to the biological sample using the ONT Read-Until capability. Existing approaches using the ONT Read-Until capability for DNA nanopore signal include [ReadFish](#) and [UNCALLED](#). Maxwell is intended for the trickier task of accurately match the noisier direct RNA signal.

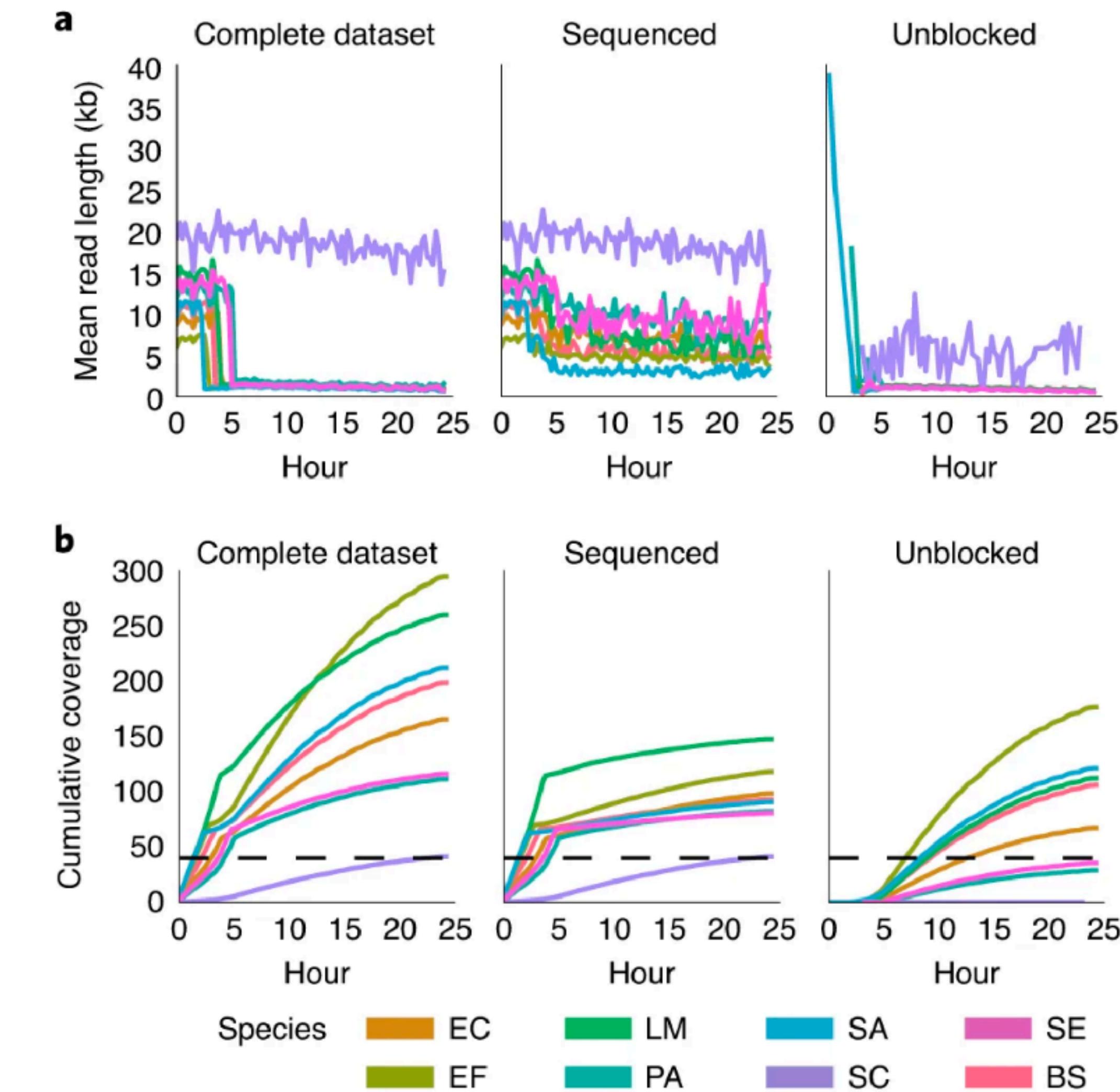
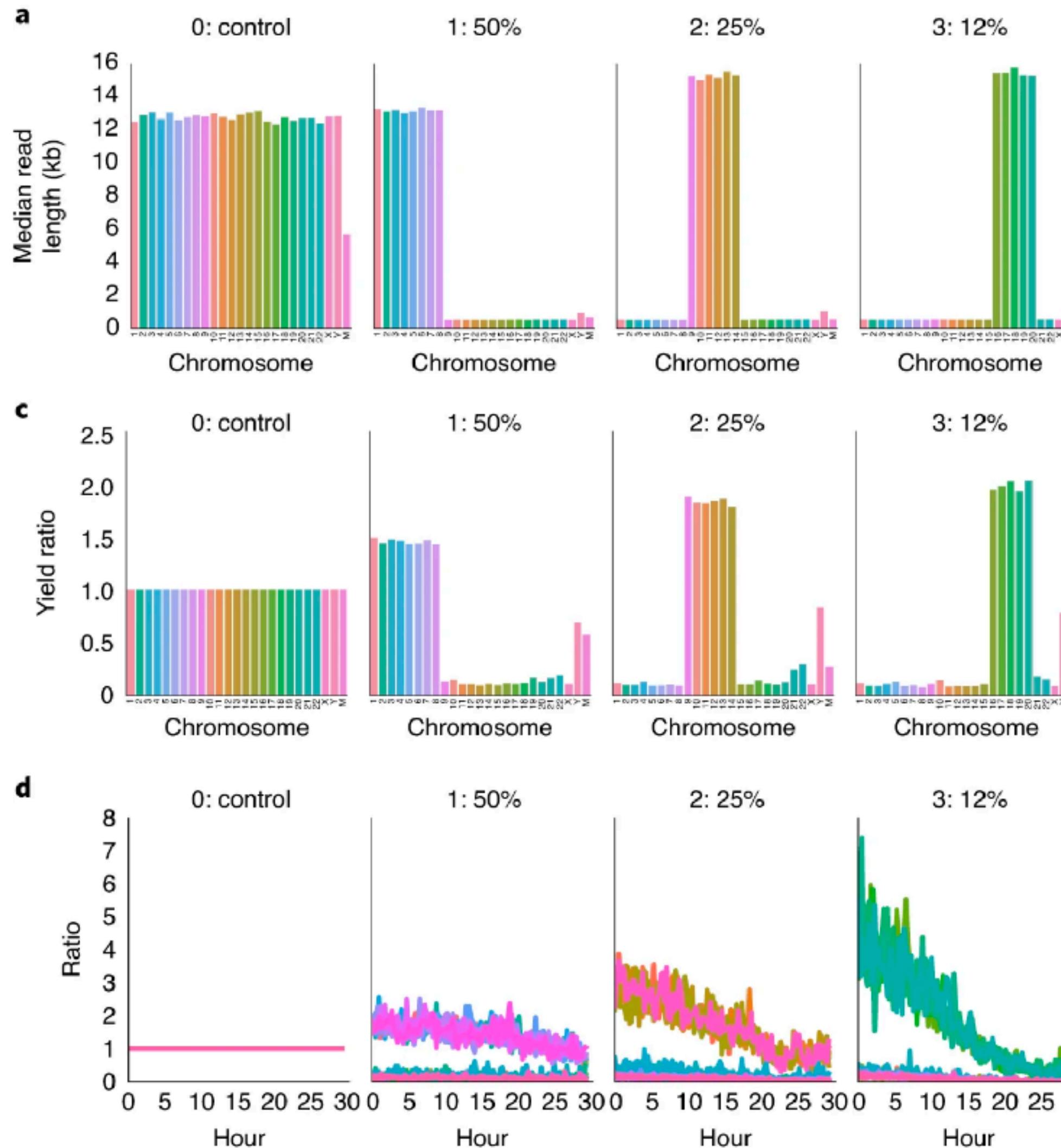
# Applications of LRS in sample enrichment

## Adaptive sampling (Read Until)



# Applications of LRS in sample enrichment

## Applications



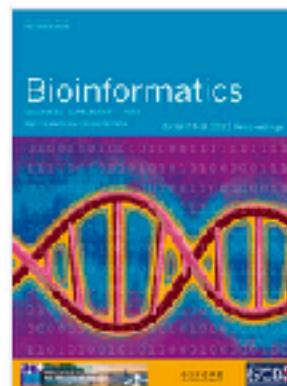
# Applications of LRS in sample enrichment

Novel methods: improve adaptive sampling or power its applications

## Bioinformatics

Issues   Advance articles   Submit ▾   Alerts   About ▾

Bioinformatics



Volume 38, Issue  
Supplement\_1  
July 2022

JOURNAL ARTICLE

### ReadBouncer: precise and scalable adaptive sampling for nanopore sequencing

Jens-Uwe Ulrich , Ahmad Lutfi, Kilian Rutzen, Bernhard Y Renard 

*Bioinformatics*, Volume 38, Issue Supplement\_1, July 2022, Pages i153–i160, <https://doi.org/10.1093/bioinformatics/btac223>

Published: 27 June 2022

## BMC Bioinformatics

Home   About   Articles   Submission Guidelines   Join The Board

Research | Open Access | Published: 28 January 2023

### DeepSelectNet: deep neural network based selective sequencing for oxford nanopore sequencing

Anjana Senanayake , Hasindu Gamaarachchi, Damayanthi Herath & Roshan Ragel

*BMC Bioinformatics* 24, Article number: 31 (2023) | [Cite this article](#)

668 Accesses | 5 Altmetric | [Metrics](#)

bioRxiv posts many COVID19-related papers. A reminder: they have not been formally peer-reviewed and should not guide health-related behavior or be reported in the press as conclusive.

#### New Results

 [Follow this preprint](#)

### RawHash: Enabling Fast and Accurate Real-Time Analysis of Raw Nanopore Signals for Large Genomes

 Can Firtina,  Nika Mansouri Ghiasi,  Joel Lindegger,  Gagandeep Singh,  Meryem Banu Cavlak,  Haiyu Mao,  Onur Mutlu

doi: <https://doi.org/10.1101/2023.01.22.525080>

## nature biotechnology

Explore content ▾   About the journal ▾   Publish with us ▾

[nature](#) > [nature biotechnology](#) > [articles](#) > [article](#)

Article | Open Access | Published: 02 January 2023

### Dynamic, adaptive sampling during nanopore sequencing using Bayesian experimental design

Lukas Weilguny, Nicola De Maio, Rory Munro, Charlotte Manser, Ewan Birney, Matthew Loose & Nick Goldman 

[Nature Biotechnology](#) (2023) | [Cite this article](#)

# Applications of LRS in clinical use

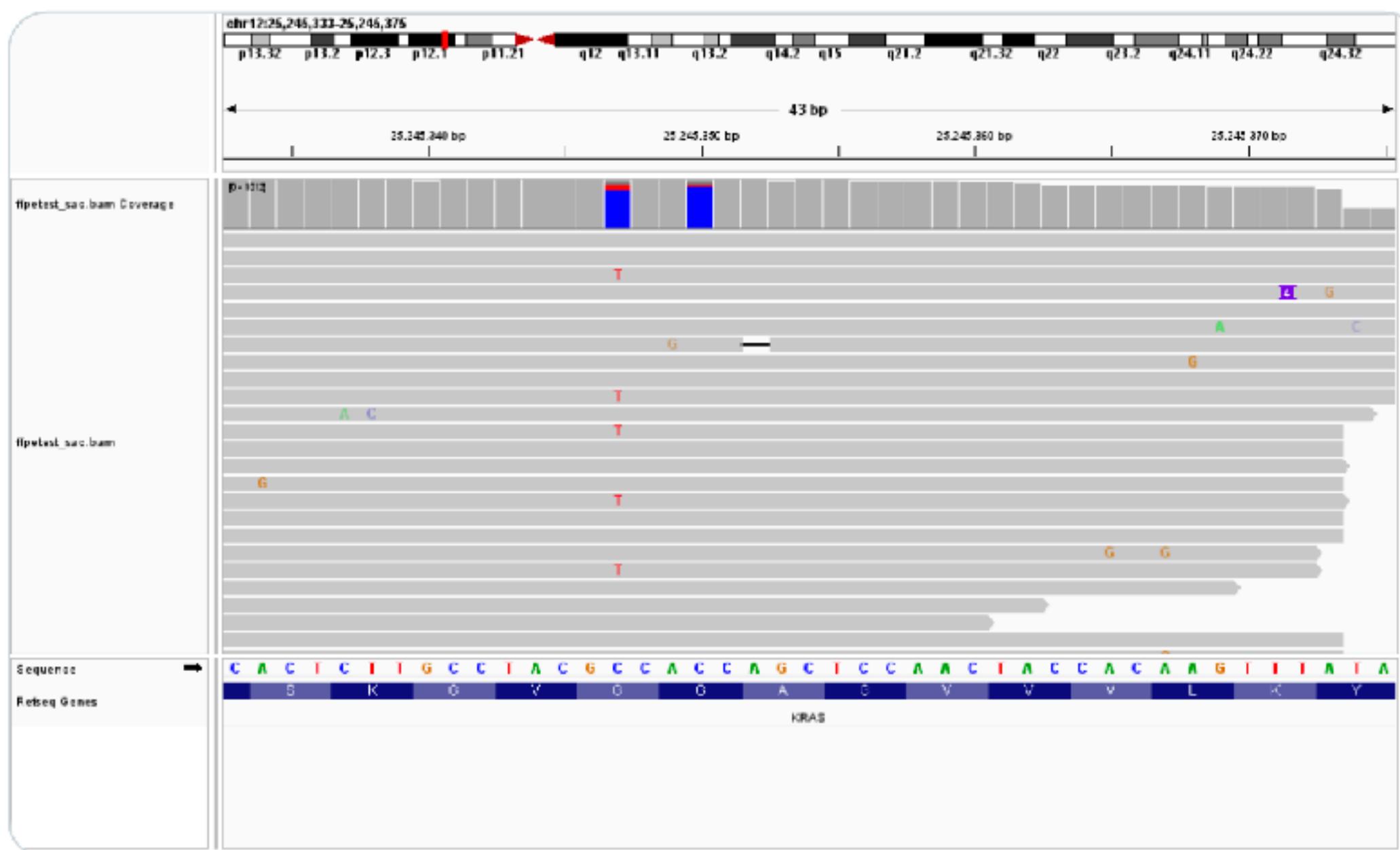


Andrew Beggs  
@adbeggs

回覆給 @coregenomics @nanopore 和其他 9 人

Funny you should ask - here is todays run from the Horizon FFPE Oncospan control showing a G13D (VAF 0.15) and G12D (VAF 0.06) in KRAS.. off a FFPE sample on a @nanopore Flongle R10.4.1 with super accuracy calling

翻譯推文



...



frontiers | Frontiers in Cellular and Infection Microbiology

TYPE Original Research  
PUBLISHED 23 January 2023  
DOI 10.3389/fcimb.2023.1064317

Check for updates

## OPEN ACCESS

### EDITED BY

Xin Zhou,  
Stanford University, United States

### REVIEWED BY

Jiyuan Hu,  
New York University, United States  
Matt Loose,  
University of Nottingham, United Kingdom

### \*CORRESPONDENCE

Huiying Liu  
✉ liuhuiying1982@sina.com  
Peng Li  
✉ jiekenlee@126.com  
Hongbin Song  
✉ hongbinsong@263.net

<sup>†</sup>These authors have contributed equally to  
this work.

## Application of nanopore adaptive sequencing in pathogen detection of a patient with *Chlamydia psittaci* infection

Yanfeng Lin<sup>1,2†</sup>, Yan Dai<sup>3†</sup>, Shuang Zhang<sup>1,4†</sup>, Hao Guo<sup>3†</sup>, Lang Yang<sup>2</sup>, Jinhui Li<sup>2</sup>, Kaiying Wang<sup>2</sup>, Ming Ni<sup>1,4</sup>, Zongqian Hu<sup>1,5</sup>, Leili Jia<sup>2</sup>, Huiying Liu<sup>6\*</sup>, Peng Li<sup>2\*</sup> and Hongbin Song<sup>1,2\*</sup>

<sup>1</sup>Academy of Military Medical Sciences, Academy of Military Sciences, Beijing, China, <sup>2</sup>Chinese PLA Center for Disease Control and Prevention, Beijing, China, <sup>3</sup>State Key Laboratory of Translational Medicine and Innovative Drug Development, Jiangsu Simcere Diagnostics Co., Ltd., Nanjing, China,

<sup>4</sup>Institute of Health Service and Transfusion Medicine, Beijing, China, <sup>5</sup>Beijing Institute of Radiation Medicine, Beijing, China, <sup>6</sup>College of Pulmonary & Critical Care Medicine, 8th Medical Center, Chinese PLA General Hospital, Beijing, China

# Applications of LRS in the future

- Monitor various RNA variables (such as isoforms, splicing patterns, TSS and poly-A site choice) in single-cell level or spatial manner simultaneously.
- Play a key role in the development and approval of RNA-based therapies (such as novel mRNA vaccines) by facilitation of identifying where modified nucleosides are incorporated in individual RNA molecules.
- Sequence one genome per species on Earth (including microbes).

# What's the next of LRS?

- All technologies in one-box
- Lower sequencing cost
- Higher read accuracy
- Fewer time to assemble a T2T genome



# Advances in LRS brings new challenges to bioinformatics

- New file format / algorithms to lower requirement of massive storage, to reduce the time in file transferring, and to increase the efficiency of file processing.
- More ML/DL-based algorithms for solving complex genome structure or identifying transcriptional regulatory mechanisms are needed.
- Higher computational resources demanded.
- How to lower CO<sub>2</sub> emission?



**OPEN**  
**Fast nanopore sequencing data analysis with SLOW5**

Hasindu Gamaarachchi<sup>1,2</sup>, Hiruna Samarakoon<sup>1,2</sup>, Sasha P. Jenner<sup>1</sup>, James M. Ferguson<sup>1</sup>, Timothy G. Amos<sup>1</sup>, Jillian M. Hammond<sup>1</sup>, Hassaan Saadat<sup>2</sup>, Martin A. Smith<sup>3,4</sup>, Sri Parameswaran<sup>2</sup> and Ira W. Deveson<sup>1,5</sup>



Introducing POD5 - [@nanopore](#)'s new serialisation format backed by [@ApacheArrow](#).

- smaller file sizes.
- hugely improved read/write performance.
- no GIL.

#NanoporeConf  
翻譯推文

nanoporetech/pod5-file-format

Pod5: a high performance file format for nanopore reads.



# Summary

- Basics of principles of mainly LRS technologies and their state-of-the-art development were introduced.
- Applications of LRS in genomics & transcriptomics and bioinformatics were also introduced.
- Adaptive sampling & signal-based analysis would be the next hot-spot.
- Faster, more accurate, and cheaper will be the main trend in the development of LRS. One-box is a plus.
- Faster, more accurate, and cheaper are also the challenges to bioinformatics.

# Acknowledgement



Dr. Meng-Han Tsai

Dr. Shu-Chen Kuo  
Dr. Ying-Chi Huang

Dr. Chien-Feng Li

Dr. Wei-Sheng Wu

# Thank you for your attention!

- Email: [chiachun.chiu@gmail.com](mailto:chiachun.chiu@gmail.com)
- Facebook: @godkin1211 or Michael Chiu
- Twitter: @chiachunchiu
- Mastodon (genomic.social): @michaelchiu
- Discord: nostalgie1211#2691



<https://www.facebook.com/groups/446434039038963>

<https://genomic.social/>



<https://discord.gg/BjSHNdh>