

Sequencing technology

Bio5488

Ting Wang

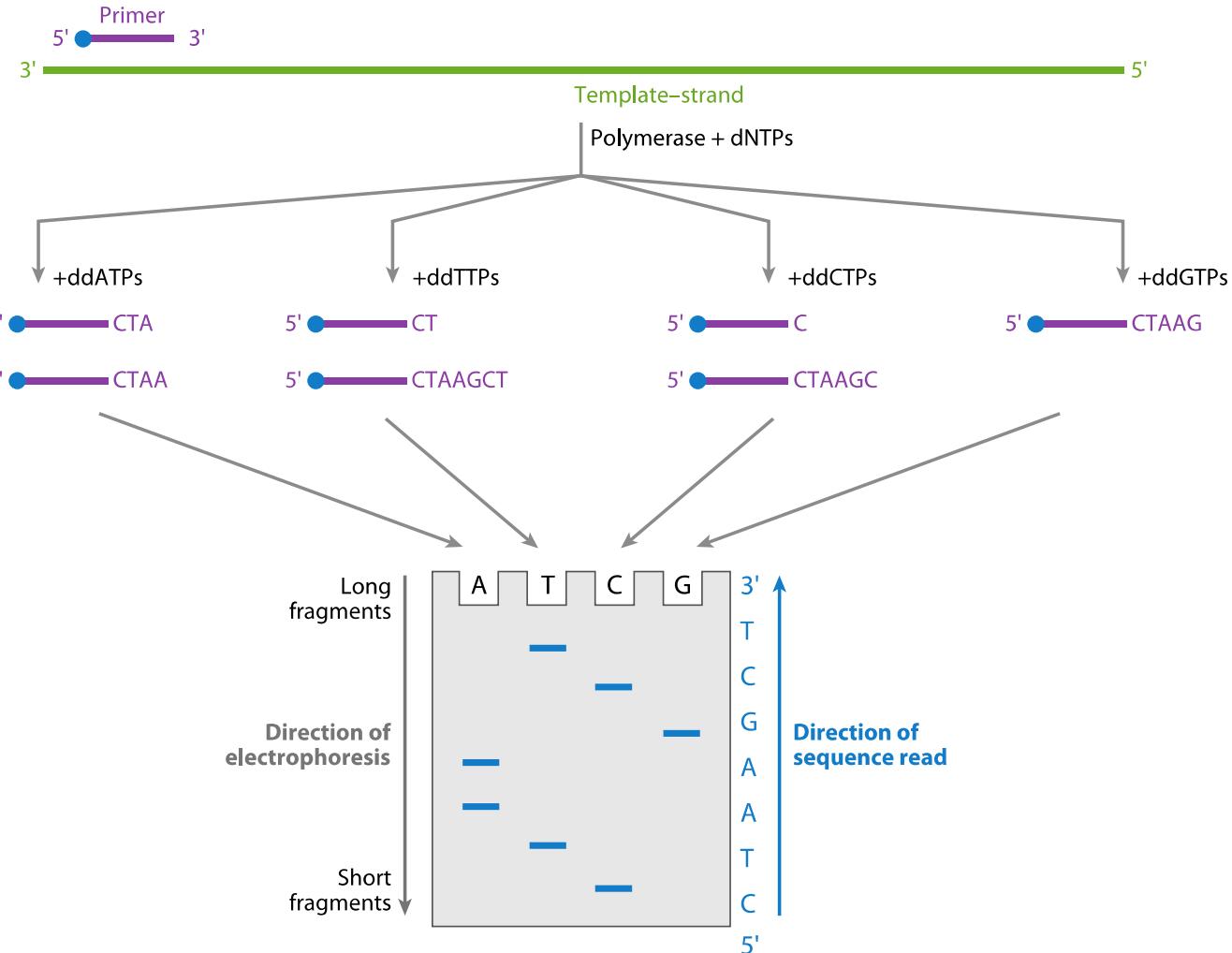
1/21/26

Outline

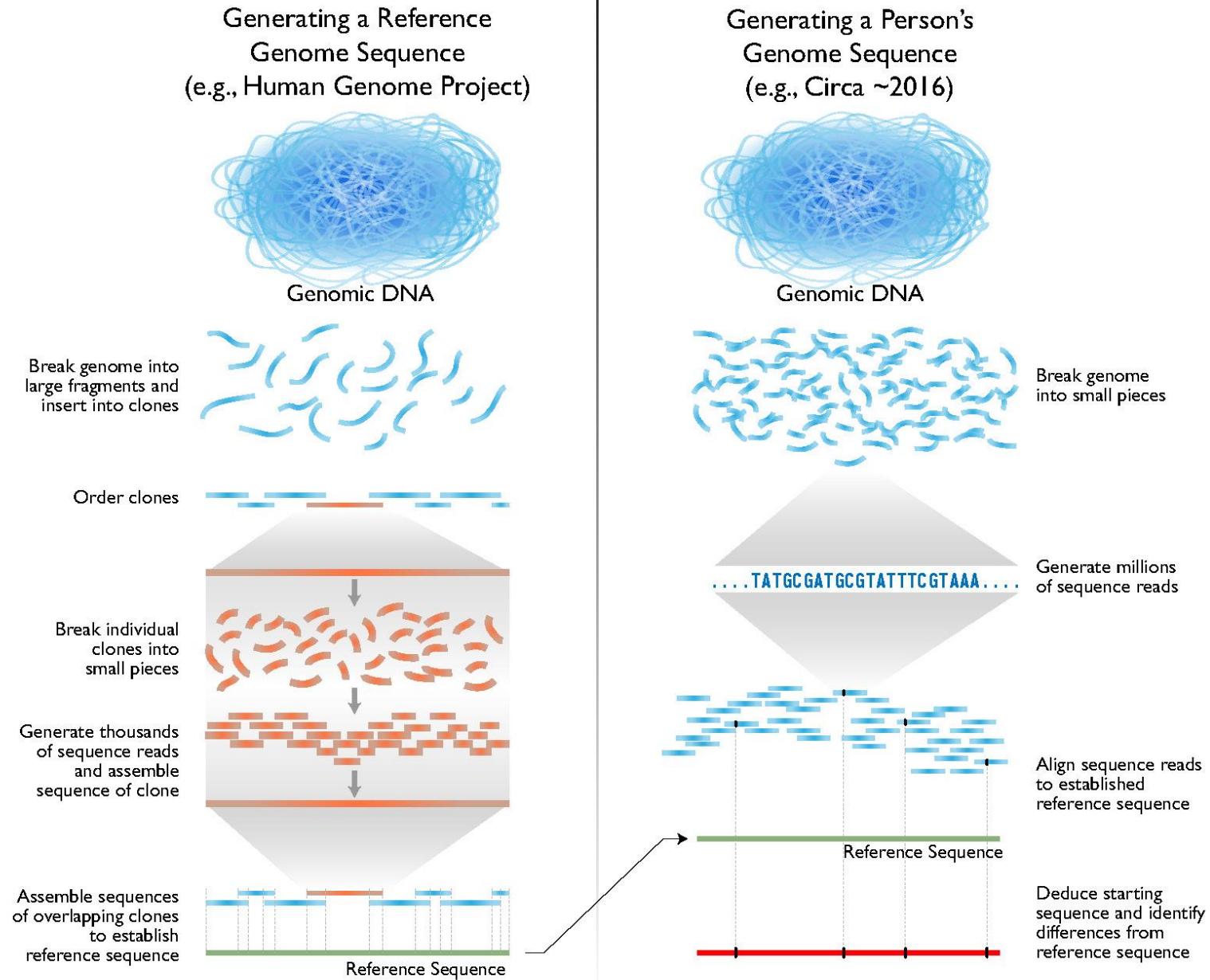
- First generation
 - Sanger sequencing
 - (Maxam and Gilbert's chemical cleavage technique)
- Second generation
 - Illumina
 - (SOLID)
 - (Ion Torrent)
 - (AVITI, Ultima, etc.)
- Third generation
 - PacBio
 - Oxford Nanopore
 - Illumina long-read (not really)
- Some math basics
- Sequence a genome
 - Assembly
- Re-sequence a genome
 - Variations
- Readout of functional genomics
 - YourName-seq

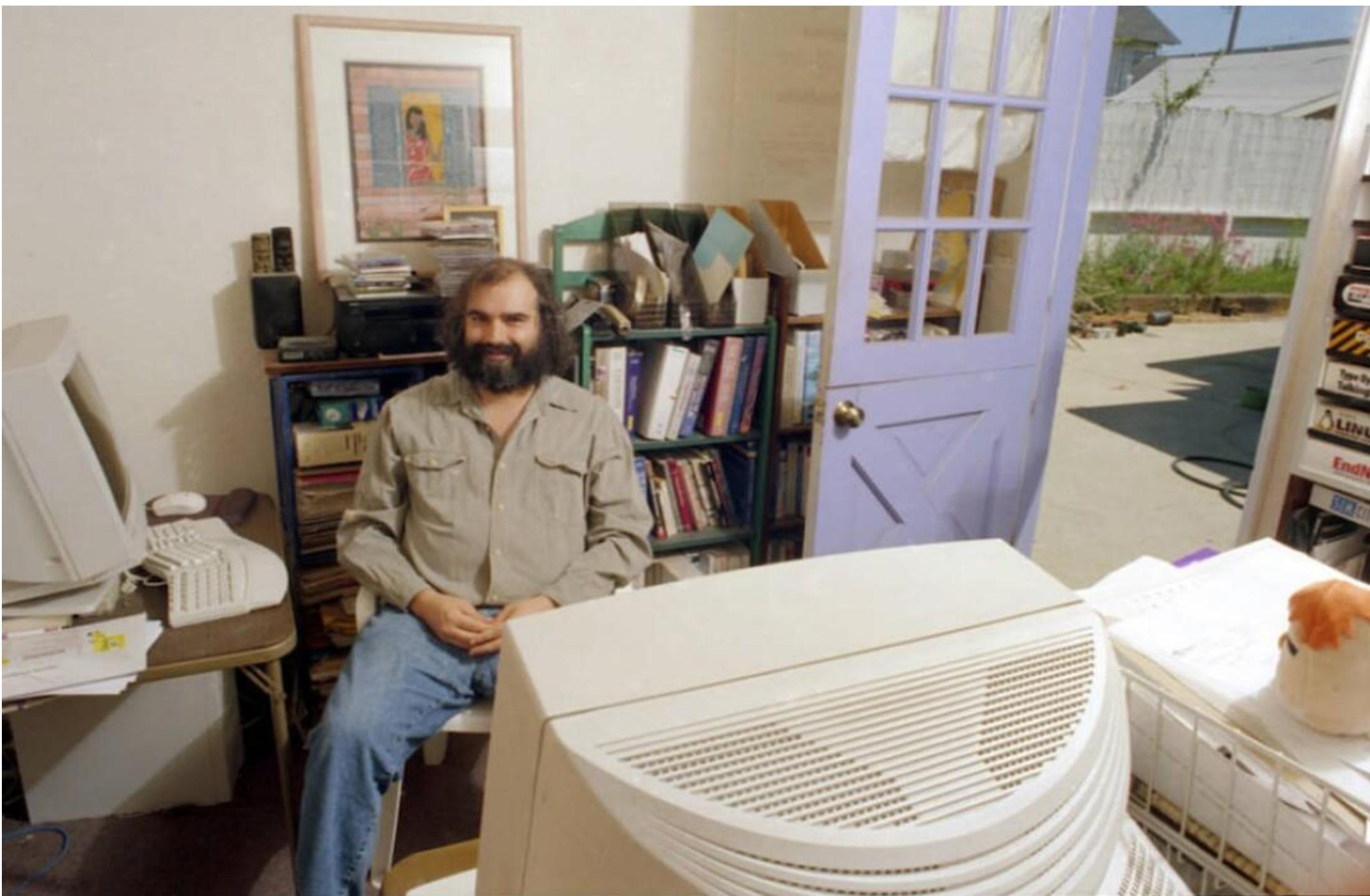
Sanger sequencing (1977)

- Dideoxy chain termination
- Dideoxynucleotides (ddNTPs) lack the 3' hydroxyl group that is required for extension of DNA chains, and therefore cannot form a bond with the 5' phosphate of the next dNTP
- Radiolabeled dideoxy nucleotides get randomly incorporated as the strand extends, halting further progression
- Four parallel reactions containing each individual ddNTP base and running the results on four lanes of a polyacrylamide gel
- Fluorescent labeling replaces radiolabeling



Human Genome Sequencing





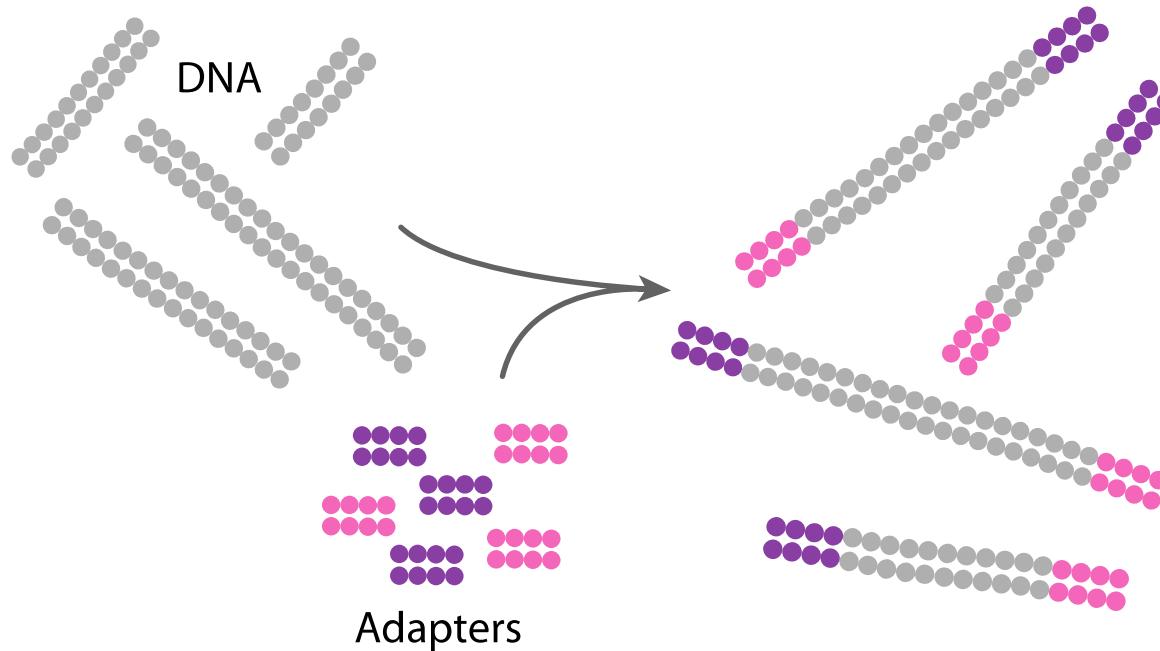
Next-gen sequencing

- Integration
 - Things get lost in translation
- Parallelization
 - If I can get two cents from everyone on earth...
- Miniaturization
 - Small is the new big.
- At what cost?
 - There are free donuts, but you have to take the exam.

Illumina: Sequencing by synthesis (SBS)

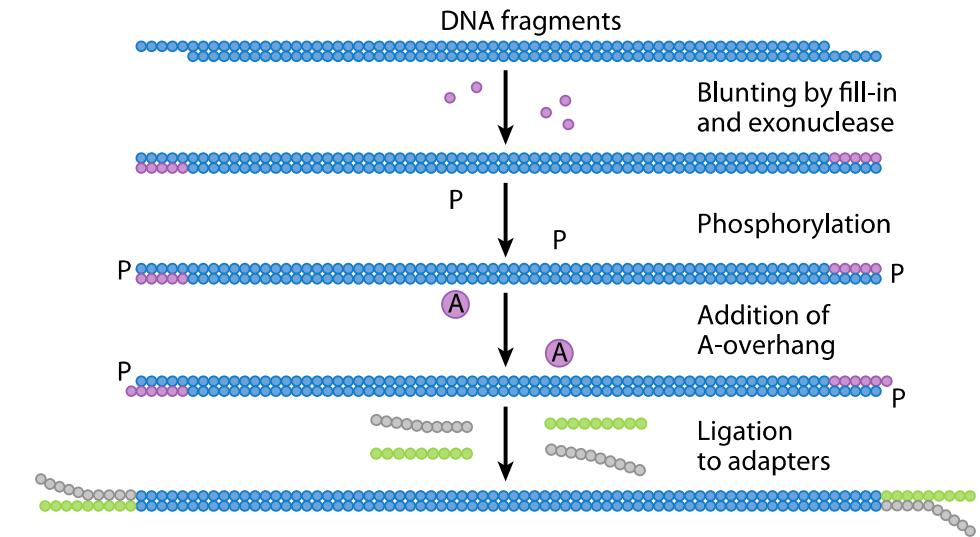
Making a library, adding adapters

a

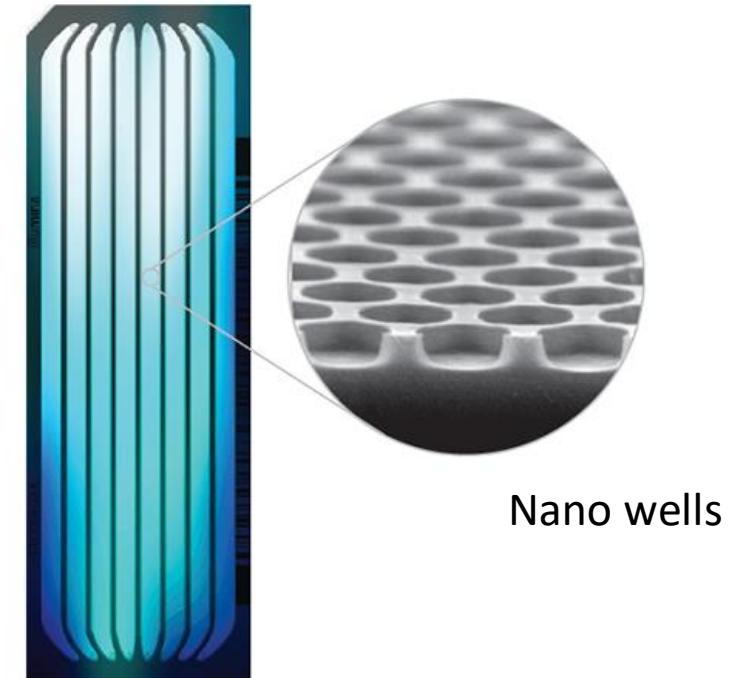
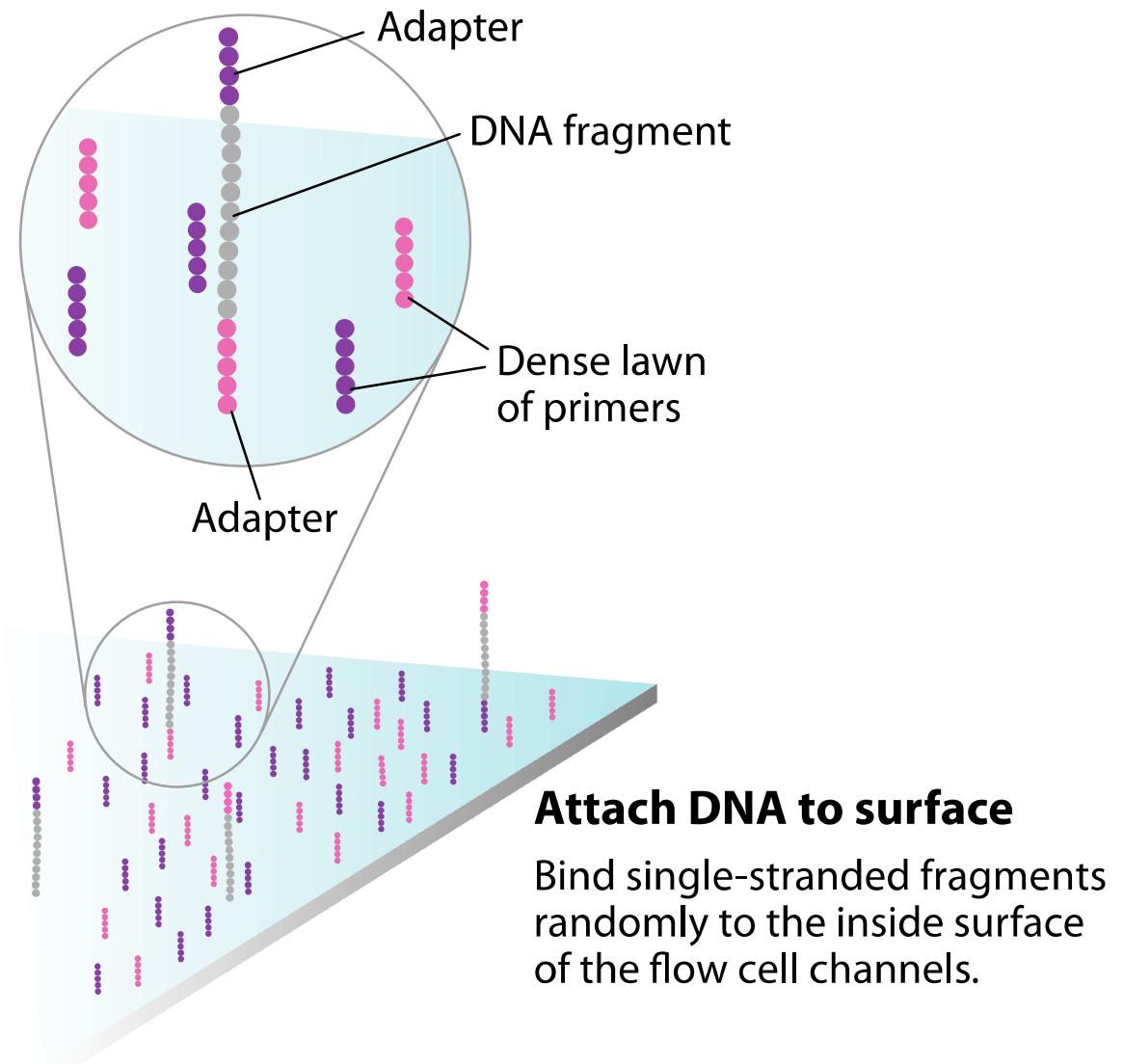


Prepare genomic DNA sample

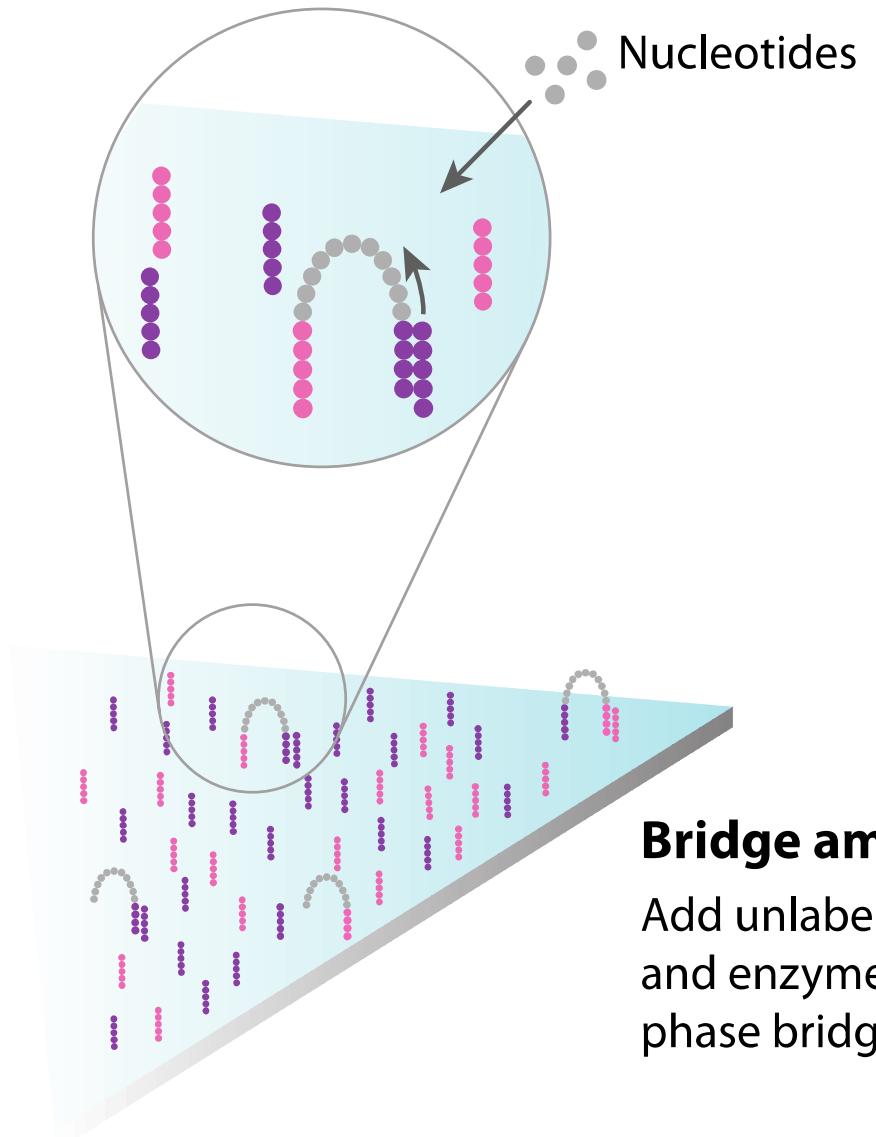
Randomly fragment genomic DNA and ligate adapters to both ends of the fragments.



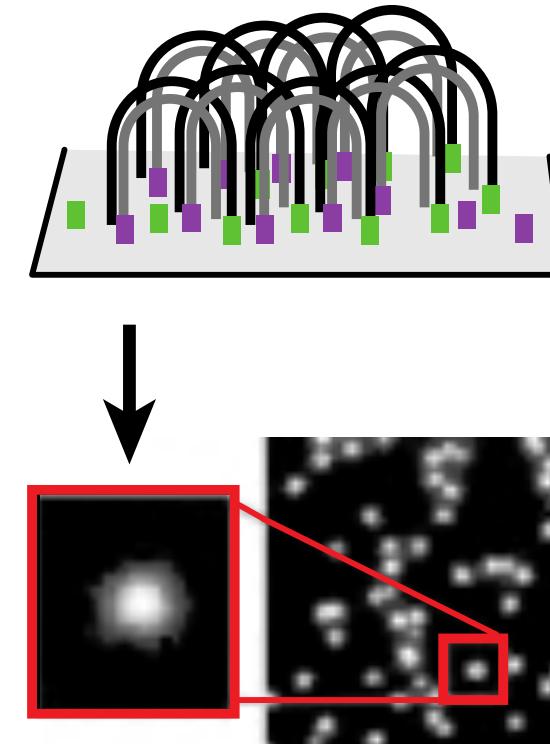
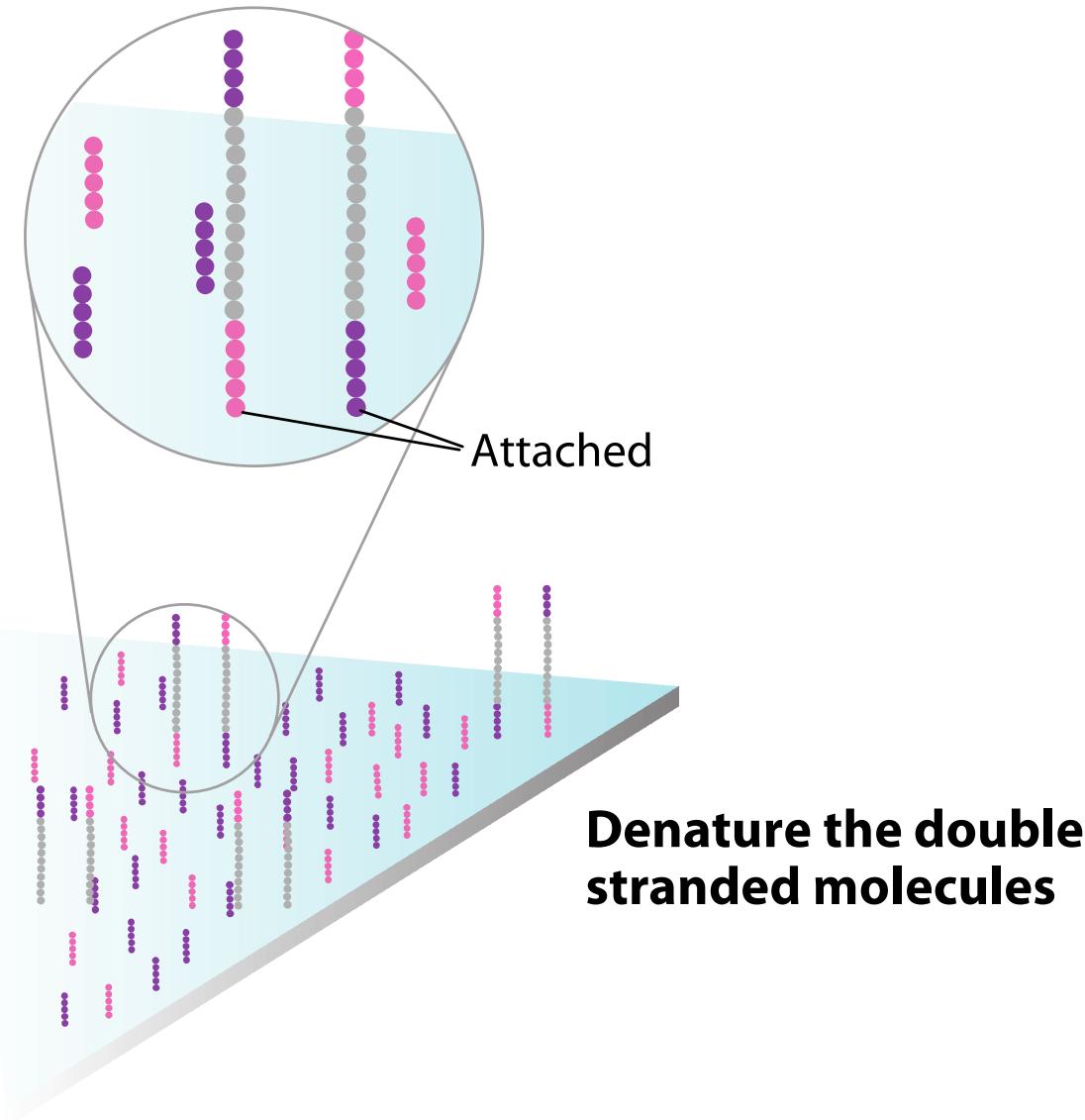
Illumina: Making of a clone/colony



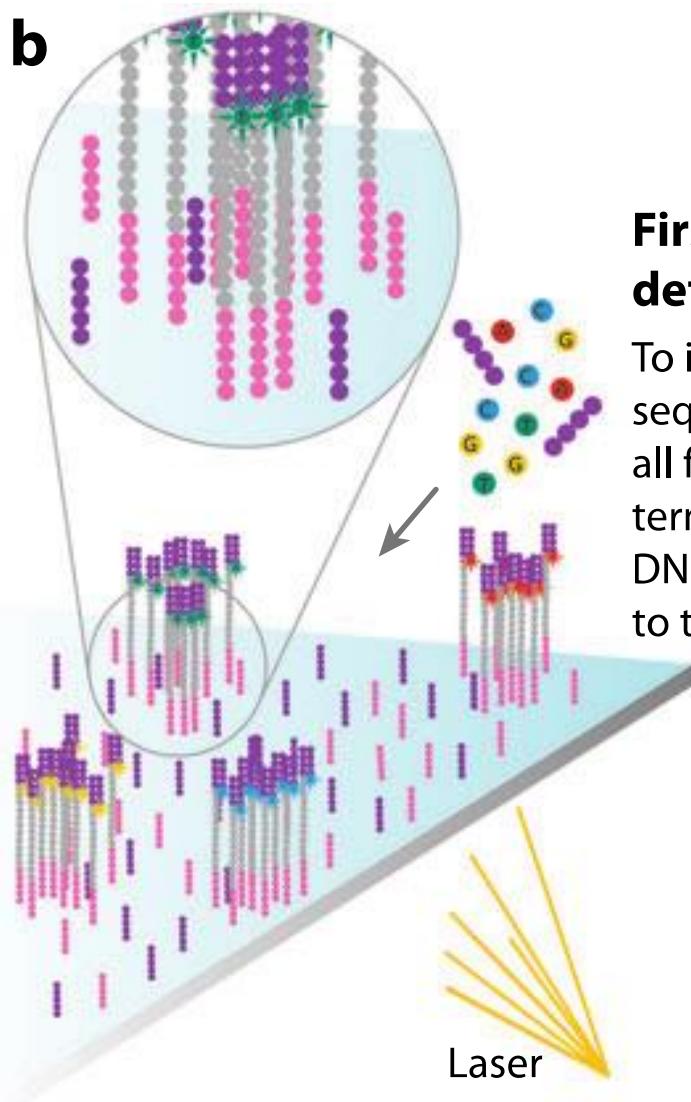
Illumina: Bridge amplification



Illumina: Sequencing by synthesis (SBS)

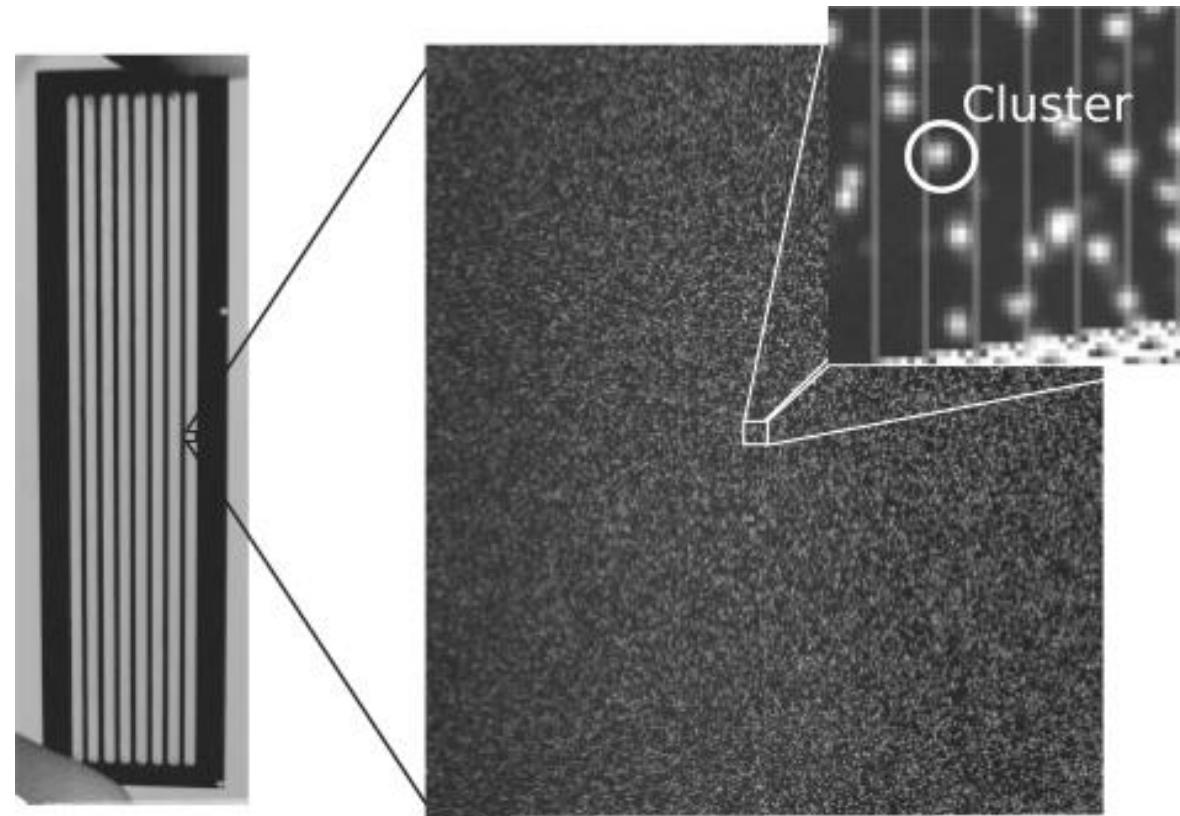


Illumina: Sequencing by synthesis (SBS)



First chemistry cycle: determine first base

To initiate the first sequencing cycle, add all four labeled reversible terminators, primers, and DNA polymerase enzyme to the flow cell.



Illumina: Sequencing by synthesis (SBS)

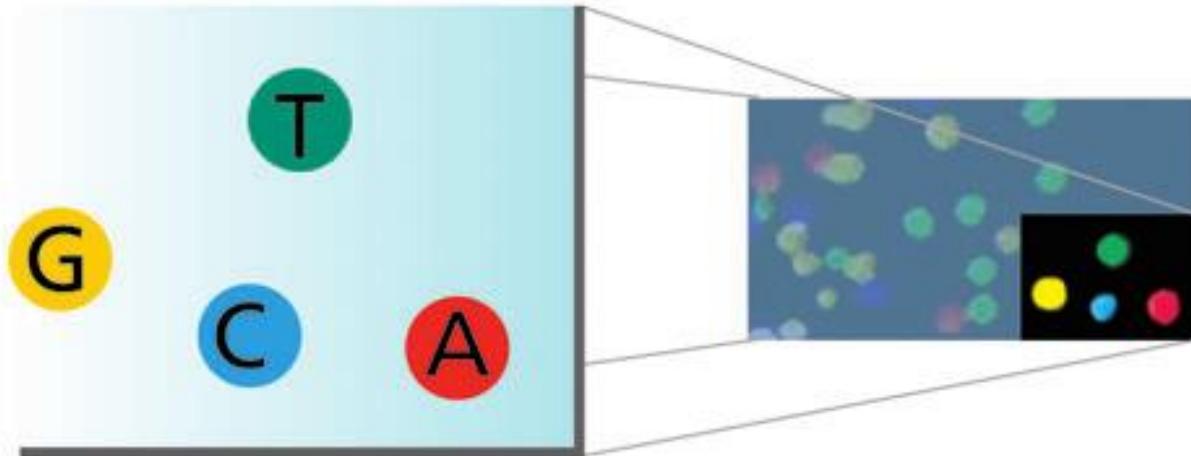
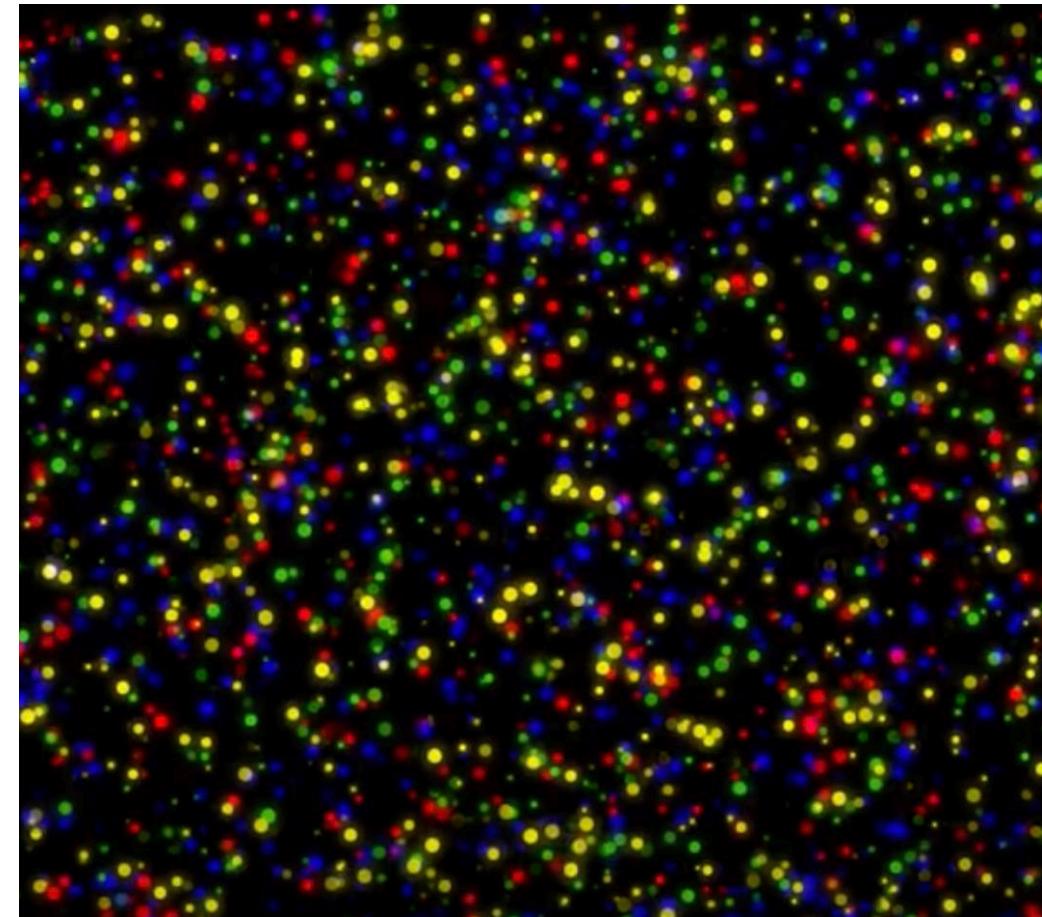


Image of first chemistry cycle

After laser excitation, capture the image of emitted fluorescence from each cluster on the flow cell. Record the identity of the first base for each cluster.

Before initiating the next chemistry cycle

The blocked 3' terminus and the fluorophore from each incorporated base are removed.



Illumina: Sequencing by synthesis (SBS)



Sequence read over multiple chemistry cycles

Repeat cycles of sequencing to determine the sequence of bases in a given fragment a single base at a time.

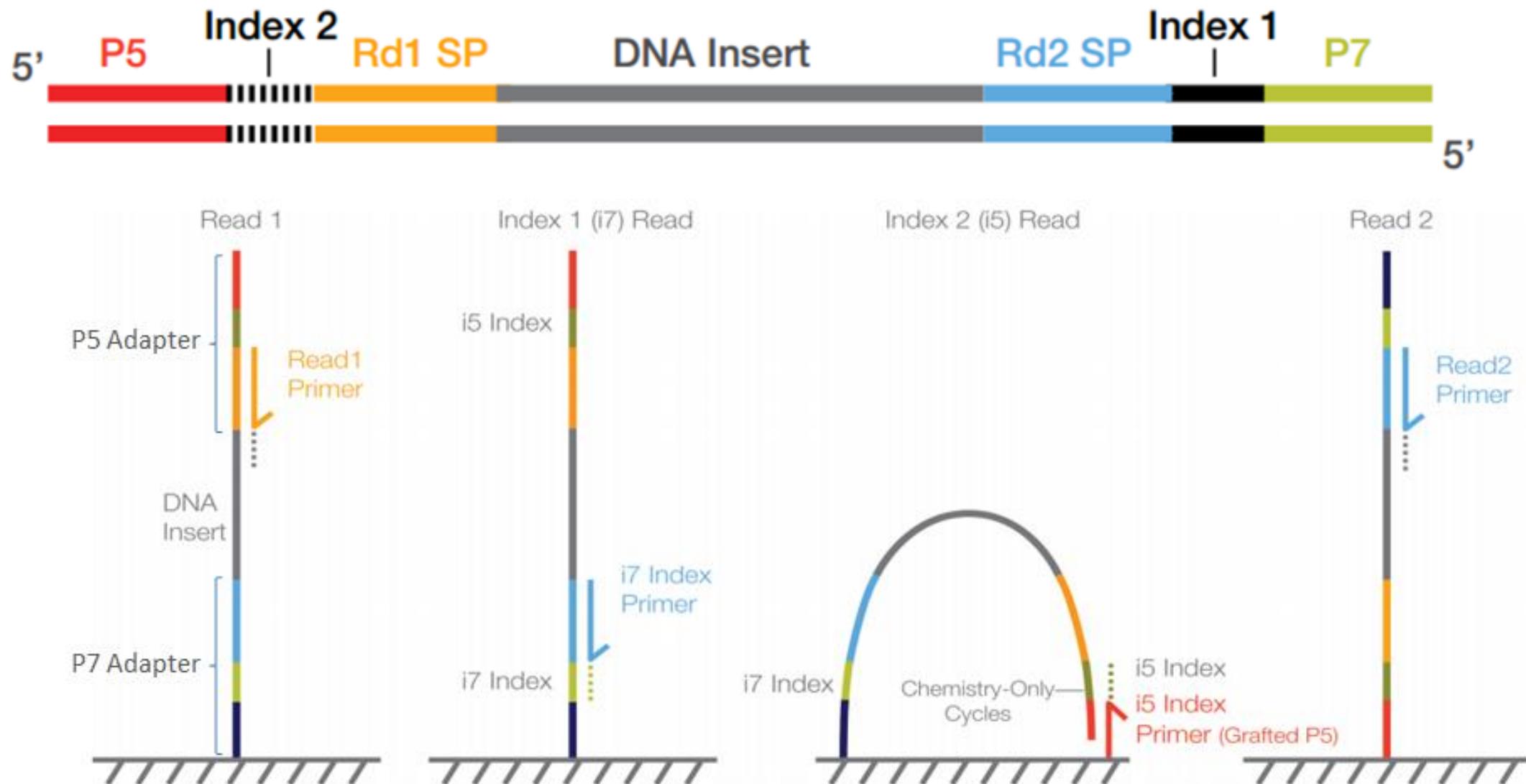
Illumina: Sequencing by synthesis (SBS)

- Smile, you are on camera!
- Y-shaped adapters
- How long can you sequence?
 - PCR
 - The problem of phasing/error accumulation
 - Mate pair libraries (don't confuse with paired end sequencing)
- Paired end sequencing
- Index, barcode
- How to increase colony density?
- How many colors?
- Nanowells and optical duplication

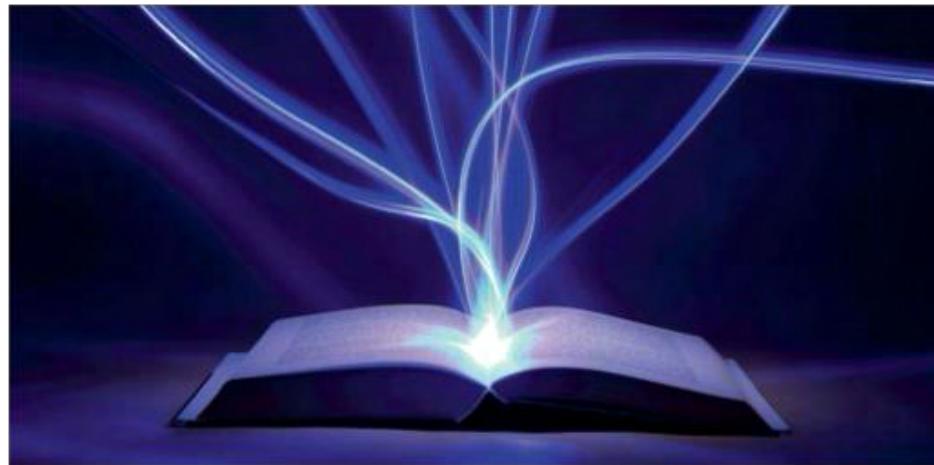
Illumina: Index, barcode, paired-end sequencing



Illumina: Index, barcode, paired-end sequencing



<https://doi.org/10.1038/s41592-022-01730-w>



Long-read sequencing has buoyed projects in small genomics labs and large-scale projects.

METHOD OF THE YEAR: LONG-READ SEQUENCING

To large-scale projects and individual labs, long-read sequencing has delivered new vistas and long wish lists for this technology's future. **By Vivien Marx**

To the delight of scientists across the life sciences, reads, which are the output of sequencing instruments, have been getting longer. Reads might be sequenced DNA or RNA and could one day routinely be entire genomes, transcriptomes and epigenomes at high throughput and accuracy, and maybe even the amino acid sequences of proteins.

Academics have happy tales about how long-read technologies have empowered their genomics projects. A few companies have facilitated this journey, notably Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT). Of late, other firms presenting long-read approaches include Element Biosciences, Illumina and MGI. Ultima Genomics and others have plans in this area.

Long reads have buoyed numerous findings in individual labs. In larger ventures, among the celebrated achievements are those in the Vertebrate Genomes Project (VGP) and the Telomere-to-Telomere Consortium (T2T)¹. A set of papers and news features related to the T2T Consortium can be found as a *Nature* Collection online. During the T2T project, says Adam Phillippy, a researcher at the National Institutes of Health (NIH) National Human Genome Research Institute (NHGRI) who co-leads the T2T Consortium, the longest read he and his colleagues handled had one million base pairs. Long-read sequencing is being used by the Human PanGenome Reference Consortium (HPRC)^{2–4}. The HPRC teams want to assemble the human genome at the T2T level of completion and capture a "better

spectrum of humanity in terms of how they represent allelic diversity," says University of California Santa Cruz researcher Karen Miga, who co-leads the T2T Consortium with Phillippy and is part of the HPRC.

Population-level data from diverse populations are needed, says Heidi Rehm, who, among other appointments, is the chief genomics officer at Massachusetts General Hospital's department of medicine and medical director of the clinical research sequencing platform at the Broad Institute of MIT and Harvard. She and her colleagues have found instances in which Black people received information about risk of a heart condition without sufficient evidence on genetic variants to support it⁵. Population data had been lacking about these variants, and such data are still limited, says Rehm.

CHERIE LARIZZI/STOCK PHOTOGRAPHY

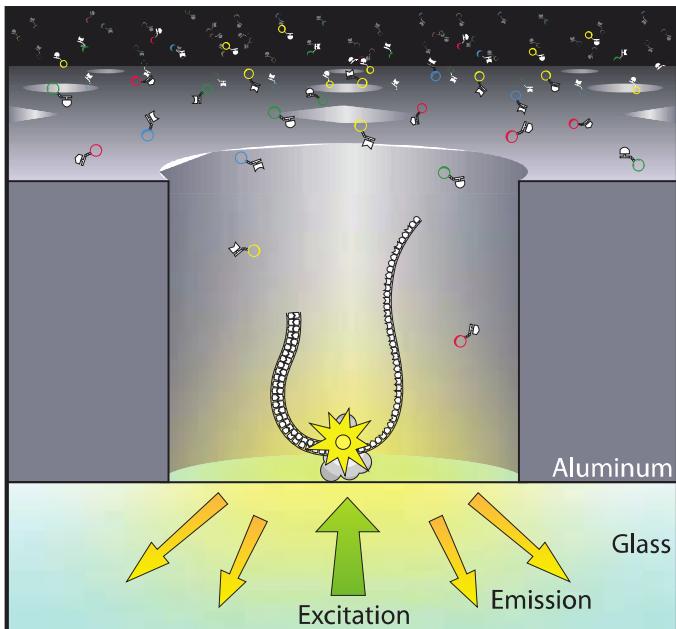
Pacific Biosciences: Single Molecule, Real-Time (SMRT) Sequencing

Real-Time DNA Sequencing from Single Polymerase Molecules

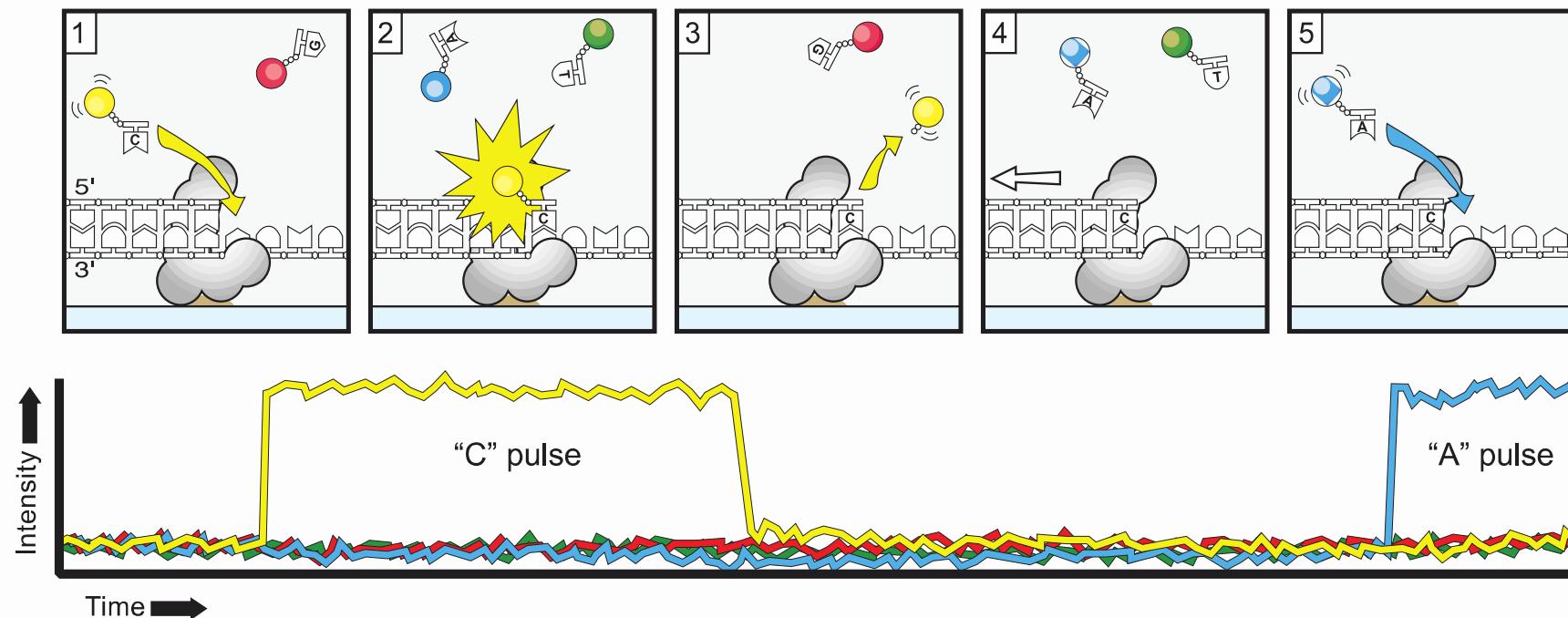
John Eid,* Adrian Fehr,* Jeremy Gray,* Khai Luong,* John Lyle,* Geoff Otto,* Paul Peluso,* David Rank,* Primo Baybayan, Brad Bettman, Arkadiusz Bibillo, Keith Bjornson, Bidhan Chaudhuri, Frederick Christians, Ronald Cicero, Sonya Clark, Ravindra Dalal, Alex deWinter, John Dixon, Mathieu Foquet, Alfred Gaertner, Paul Hardenbol, Cheryl Heiner, Kevin Hester, David Holden, Gregory Kearns, Xiangxu Kong, Ronald Kuse, Yves Lacroix, Steven Lin, Paul Lundquist, Congcong Ma, Patrick Marks, Mark Maxham, Devon Murphy, Insil Park, Thang Pham, Michael Phillips, Joy Roy, Robert Sebra, Gene Shen, Jon Sorenson, Austin Tomaney, Kevin Travers, Mark Trulson, John Vieceli, Jeffrey Wegener, Dawn Wu, Alicia Yang, Denis Zaccarin, Peter Zhao, Frank Zhong, Jonas Korlach,† Stephen Turner†

Pacific Biosciences: Single Molecule, Real-Time (SMRT) Sequencing

A

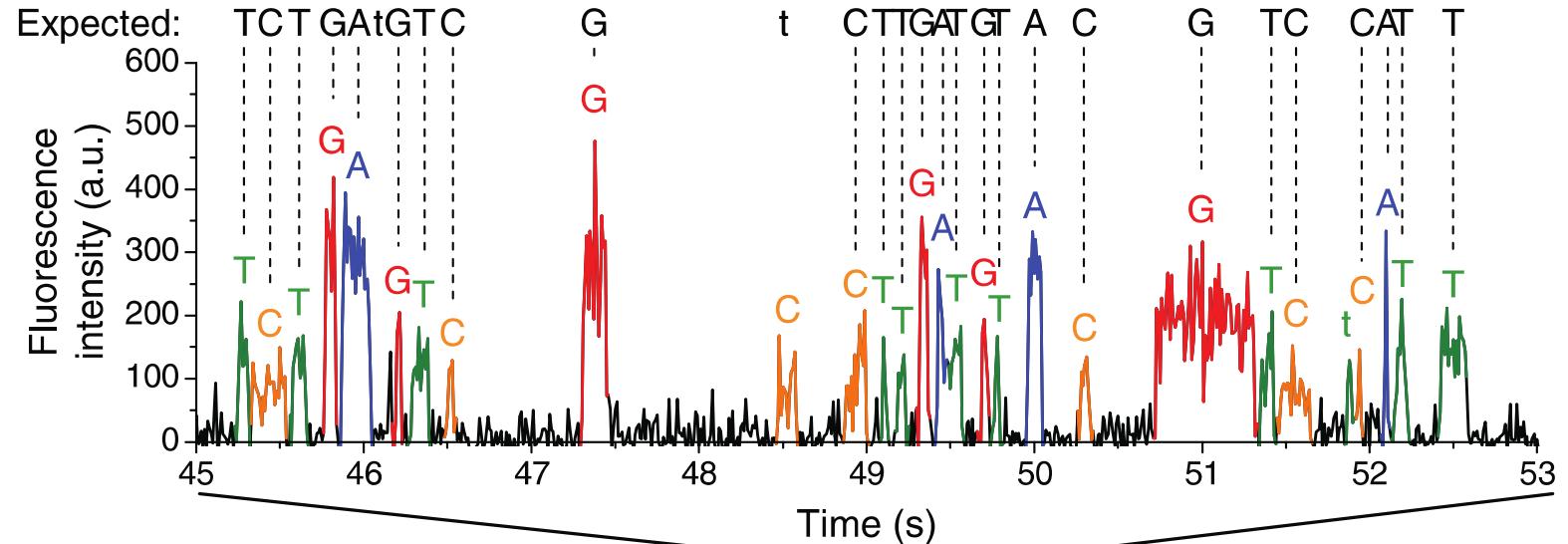


B

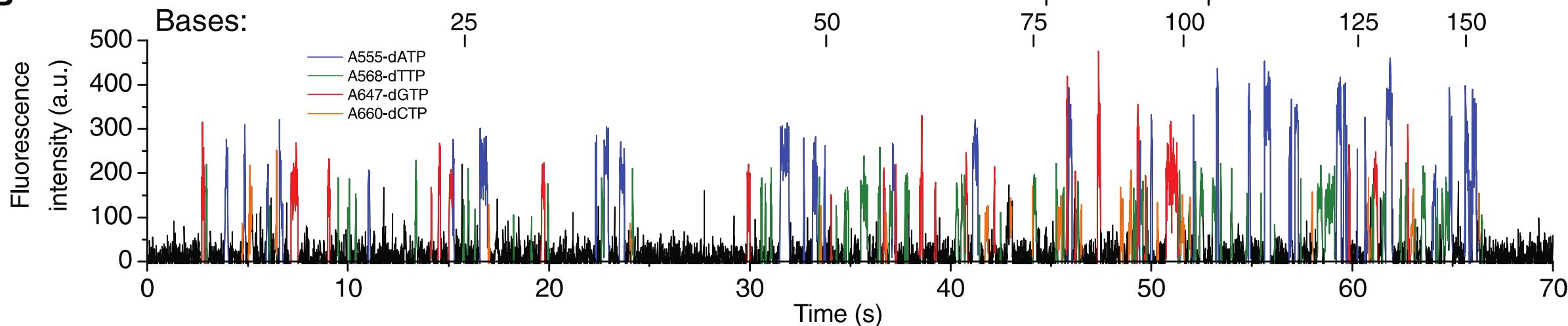


Pacific Biosciences: Single Molecule, Real-Time (SMRT) Sequencing

A



B



Pacific Biosciences: HiFi reads

Start with high-quality double stranded DNA



Ligate SMRTbell adapters and size select



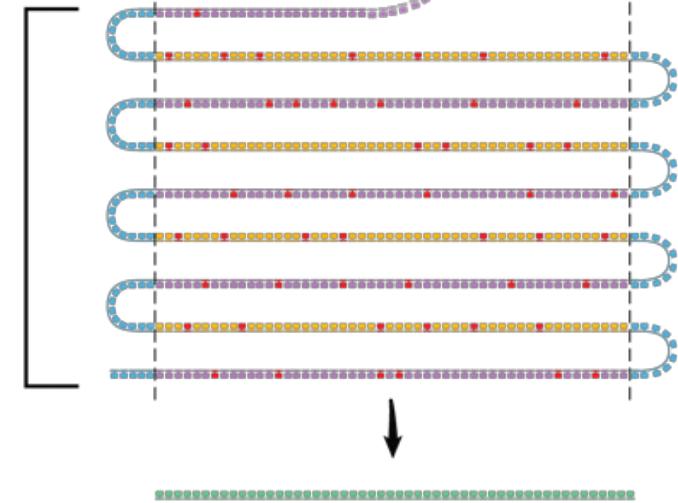
Anneal primers and bind DNA polymerase



Circularized DNA is sequenced in repeated passes



The polymerase reads are trimmed of adapters to yield subreads

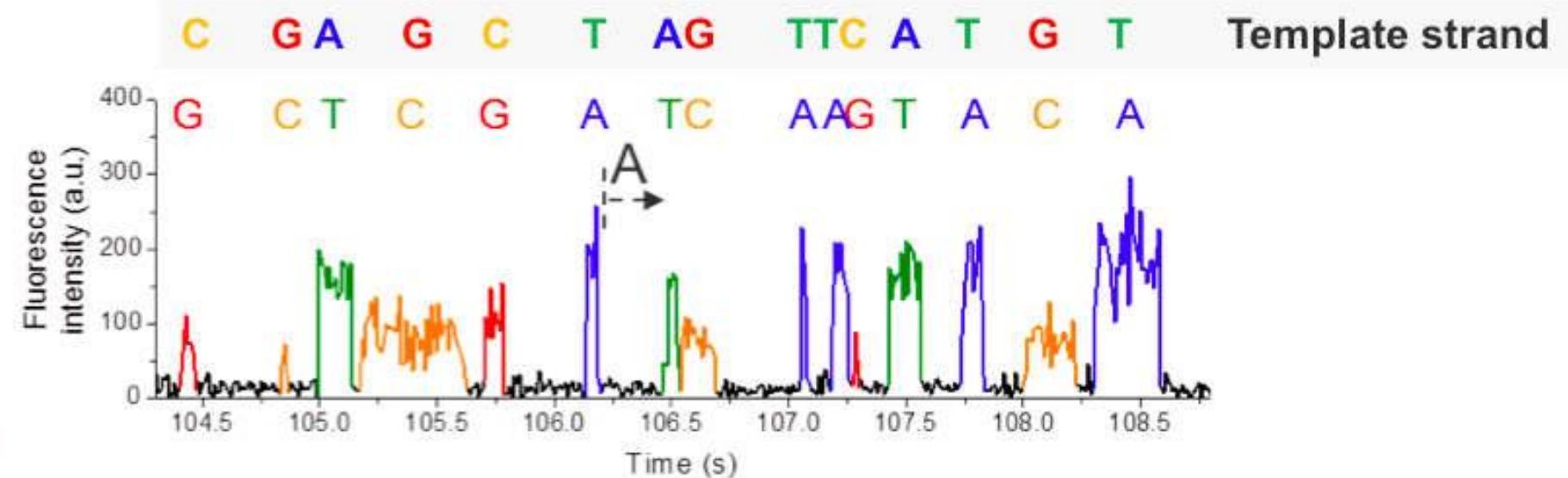
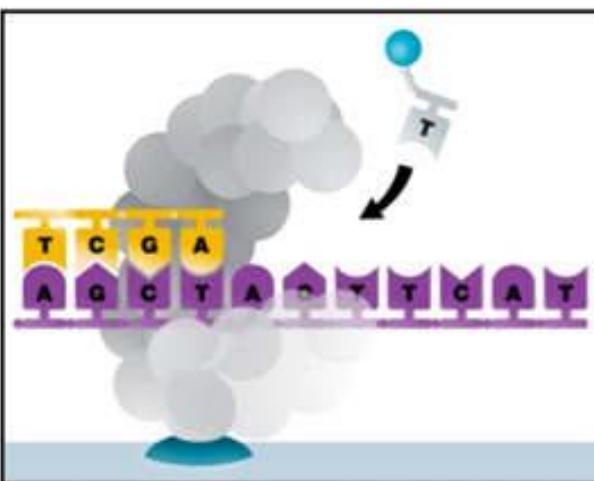
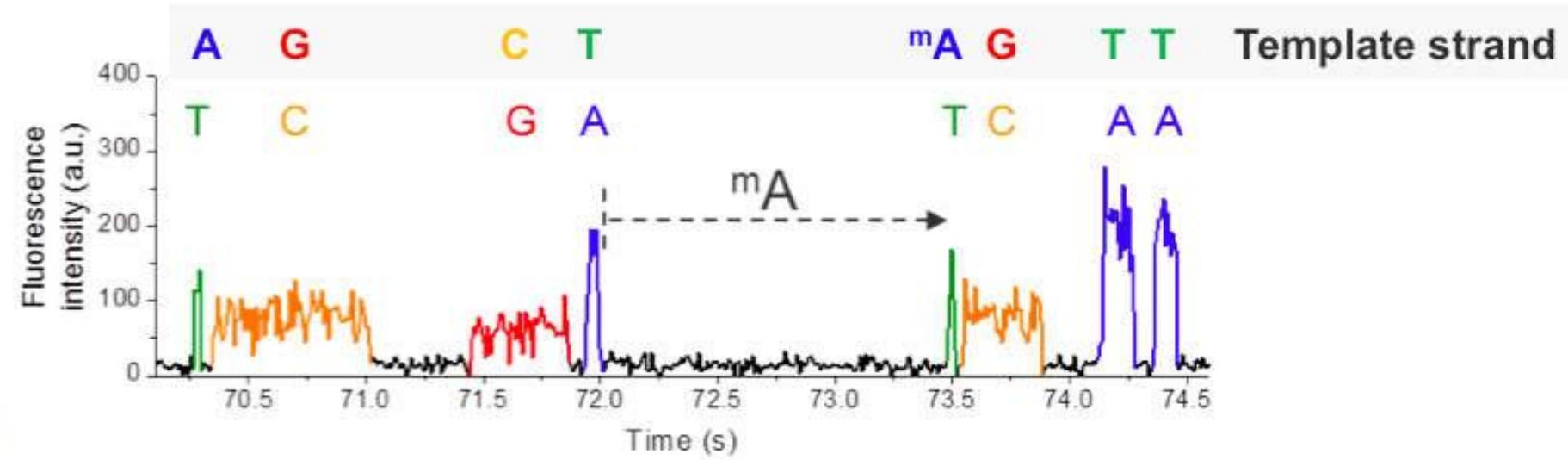
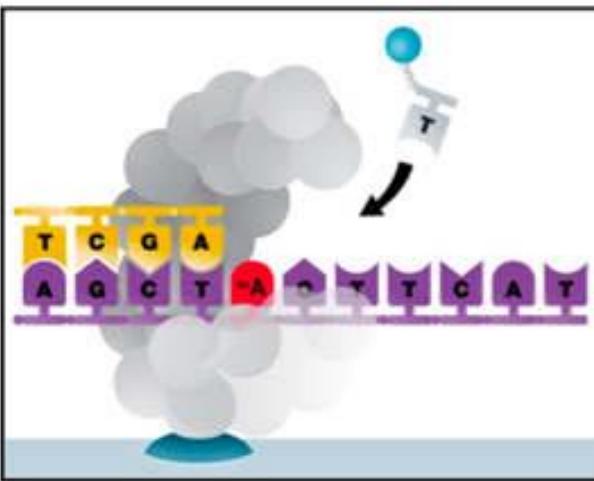


Consensus is called from subreads

HiFi READ

>99.9% accuracy

Pacific Biosciences: Detecting modified bases



Pacific Biosciences: Detecting modified bases

Genome-wide detection of cytosine methylation by single molecule real-time sequencing

O. Y. Olivia Tse^{a,b,1} , Peiyong Jiang^{a,b,1} , Suk Hang Cheng^{a,b,1}, Wenlei Peng^{a,b}, Huimin Shang^{a,b}, John Wong^c , Stephen L. Chan^{d,e} , Liona C. Y. Poon^f, Tak Y. Leung^f, K. C. Allen Chan^{a,b,e}, Rossa W. K. Chiu^{a,b} , and Y. M. Dennis Lo^{a,b,e,2} 

^aLi Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong Special Administrative Region, China;

^bDepartment of Chemical Pathology, Prince of Wales Hospital, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong Special Administrative Region, China; ^cDepartment of Surgery, Prince of Wales Hospital, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong Special Administrative Region, China; ^dDepartment of Clinical Oncology, Prince of Wales Hospital, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong Special Administrative Region, China; ^eState Key Laboratory of Translational Oncology, Prince of Wales Hospital, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong Special Administrative Region, China; and ^fDepartment of Obstetrics and Gynaecology, Prince of Wales Hospital, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong Special Administrative Region, China

Contributed by Y. M. Dennis Lo, December 9, 2020 (sent for review September 25, 2020; reviewed by Shankar Balasubramanian and Andrew P. Feinberg)

Oxford Nanopore: Sequencing by a nanopore

Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 13770–13773, November 1996
Biophysics

Characterization of individual polynucleotide molecules using a membrane channel

JOHN J. KASIANOWICZ*, ERIC BRANDIN†, DANIEL BRANTON†‡, AND DAVID W. DEAMER§

*Biotechnology Division, National Institute of Science and Technology, 222/A353, Gaithersburg, MD 20899; †Department of Molecular and Cellular Biology, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138; and §Department of Chemistry and Biochemistry, University of California, Santa Cruz, CA 95064

Contributed by Daniel Branton, September 5, 1996

Biophysical Journal Volume 77 December 1999 3227–3233

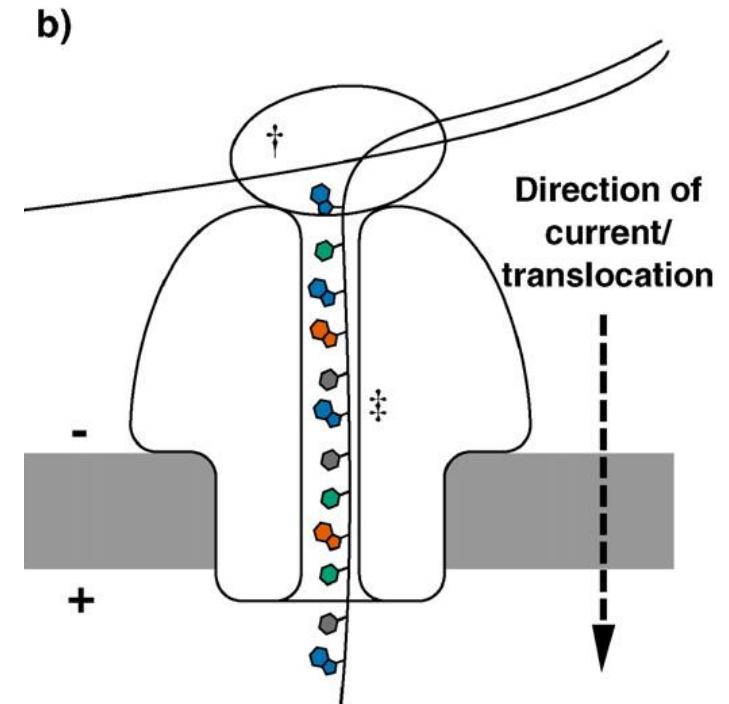
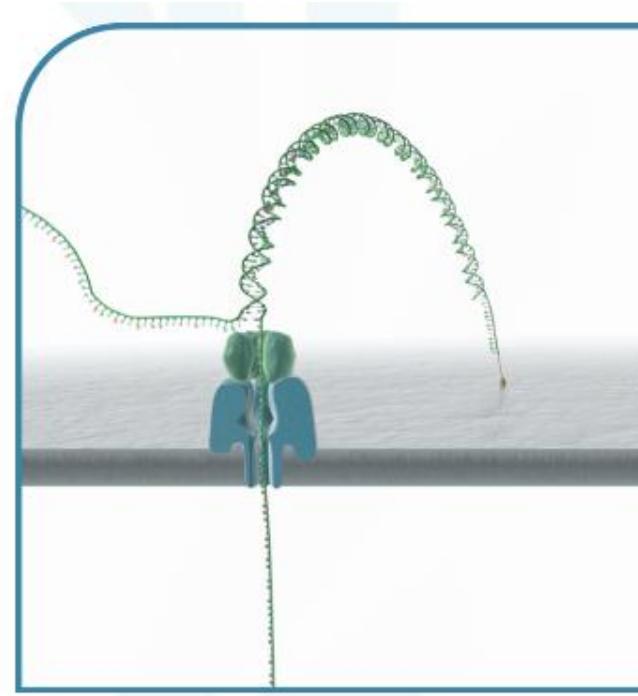
3227

Microsecond Time-Scale Discrimination Among Polycytidylc Acid, Polyadenylic Acid, and Polyuridylic Acid as Homopolymers or as Segments Within Single RNA Molecules

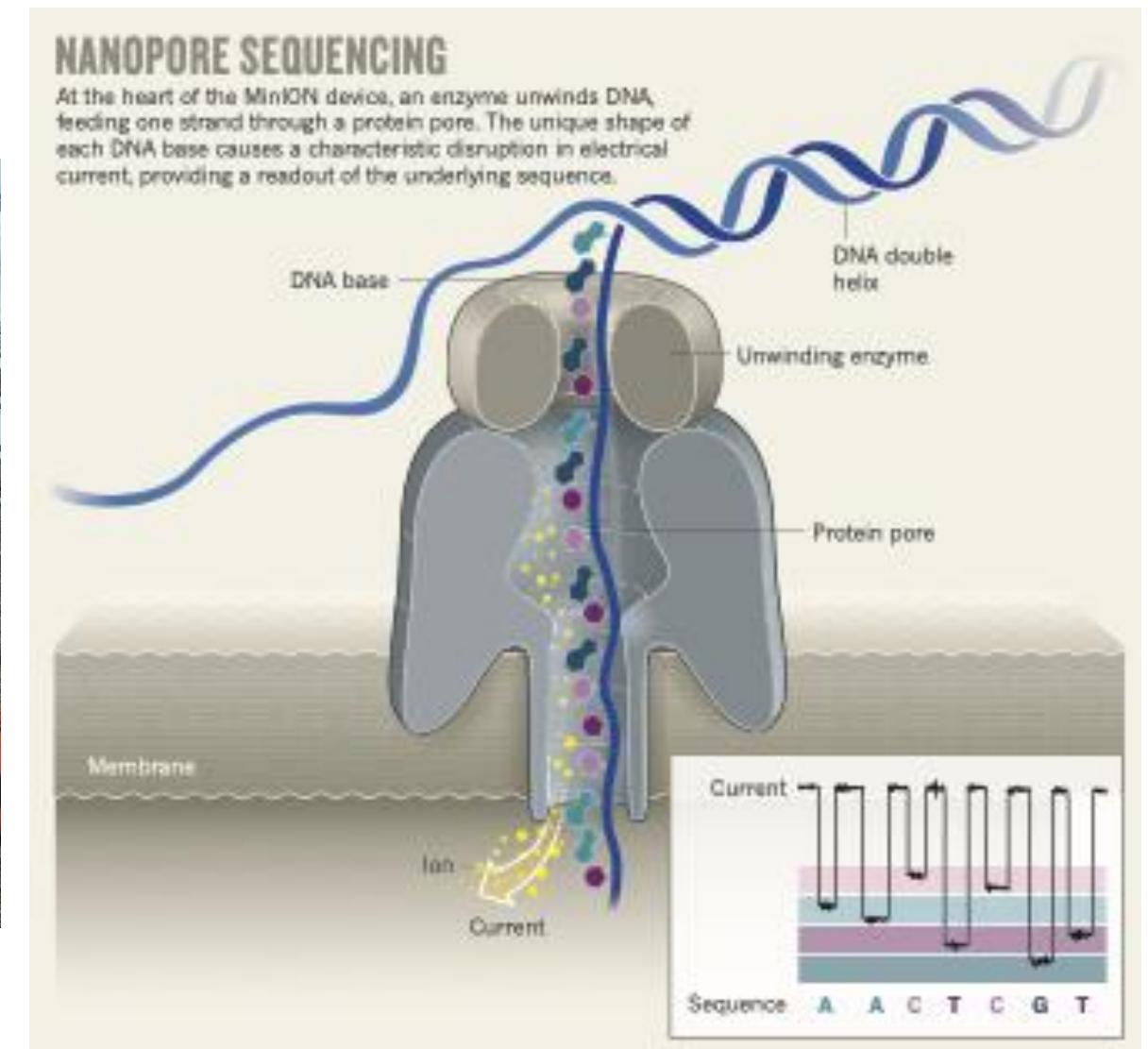
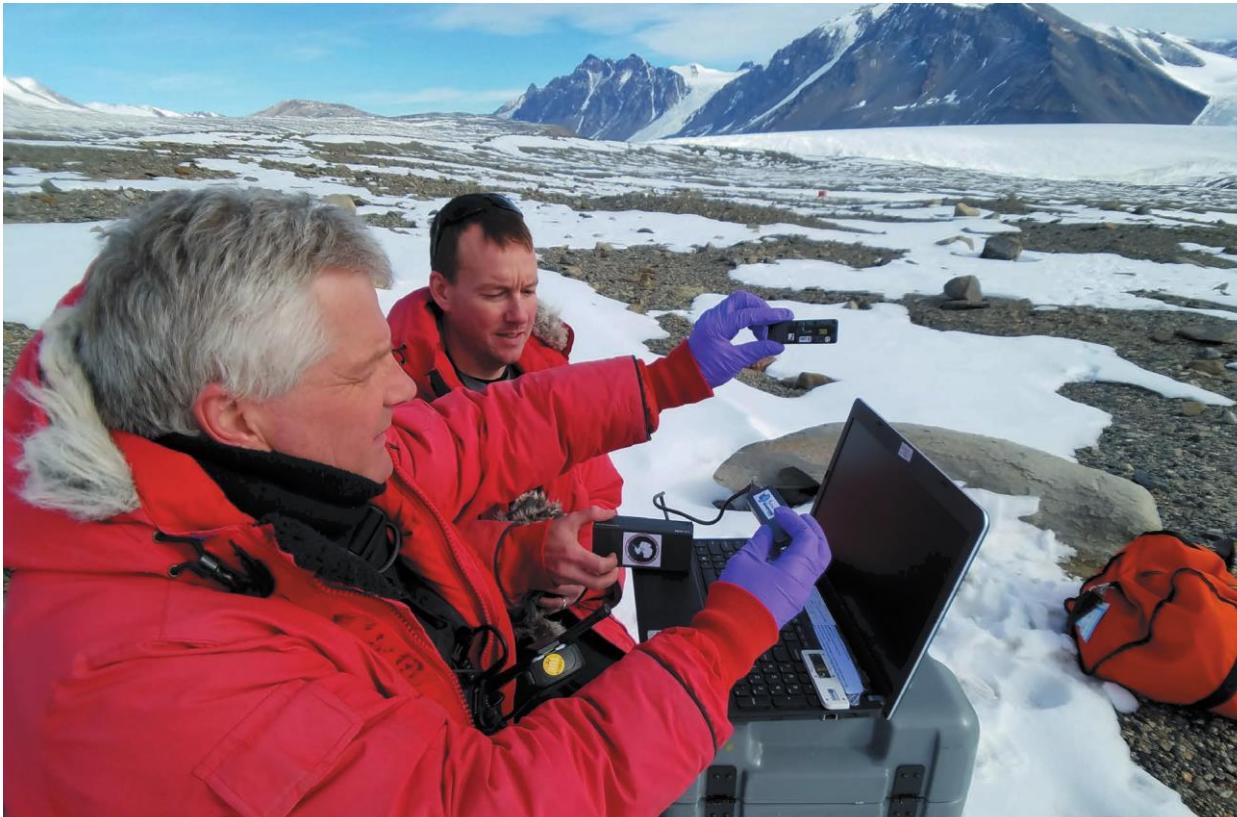
Mark Akeson,* Daniel Branton,† John J. Kasianowicz,§ Eric Brandin,† and David W. Deamer*

*Department of Chemistry & Biochemistry, University of California, Santa Cruz, CA 95064; †Department of Molecular & Cellular Biology, Harvard University, Cambridge, MA 02138; and §Biotechnology Division, National Institute of Standards & Technology, Gaithersburg, MD 20899

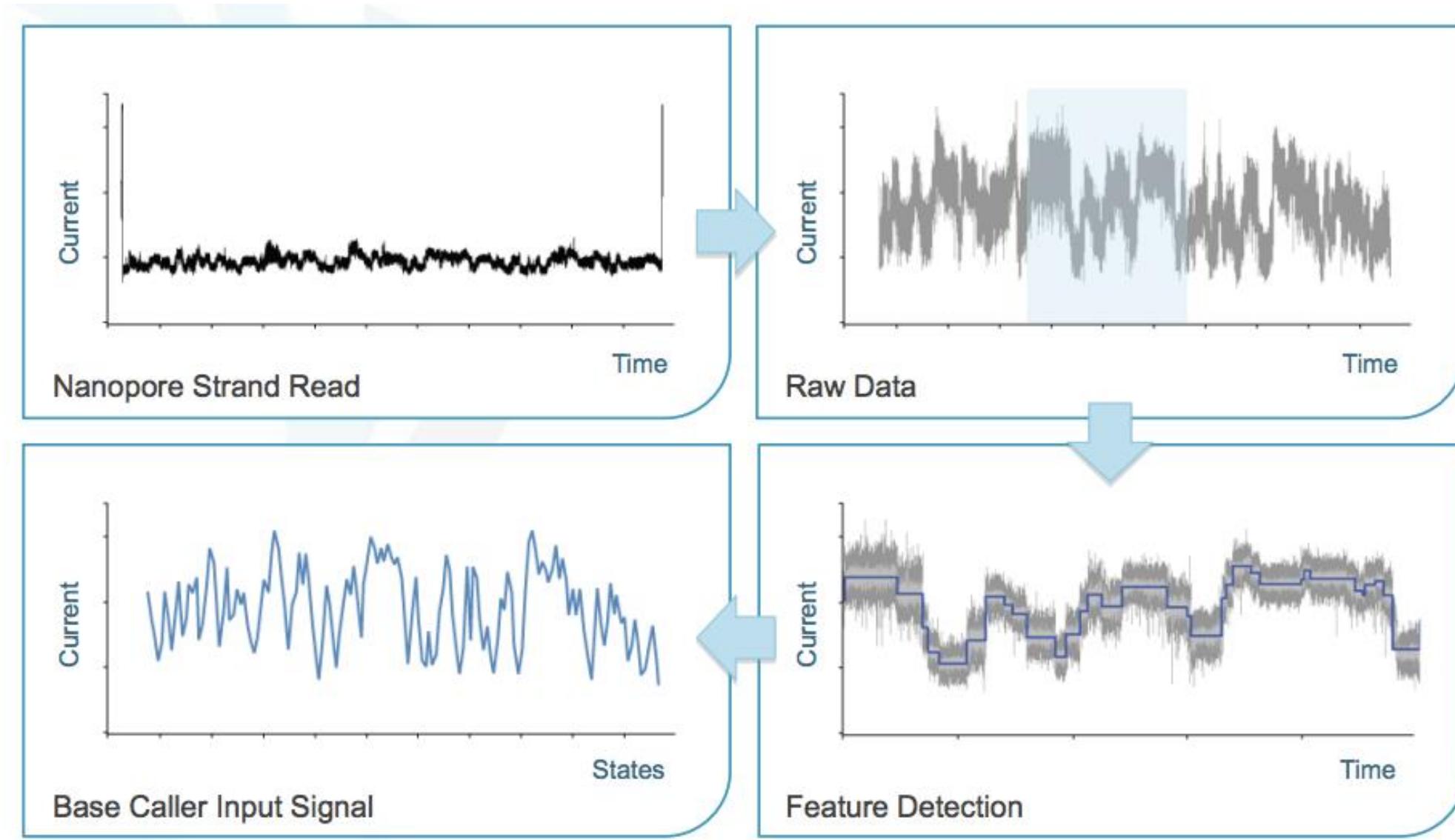
Oxford Nanopore: Sequencing by a nanopore



Oxford Nanopore: Sequencing by a nanopore

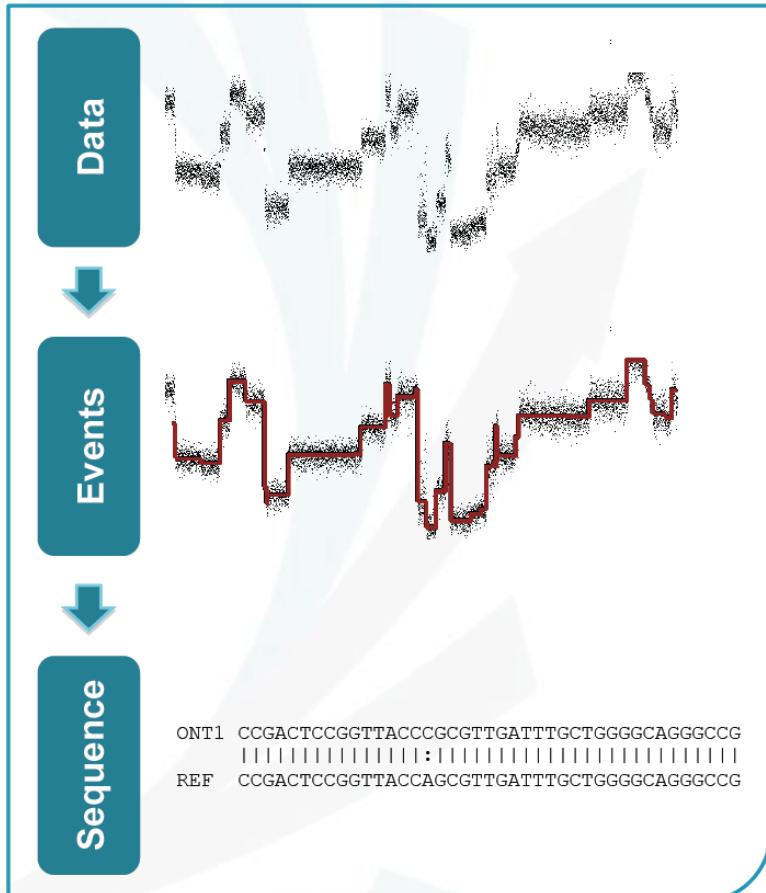


Oxford Nanopore: Sequencing by a nanopore

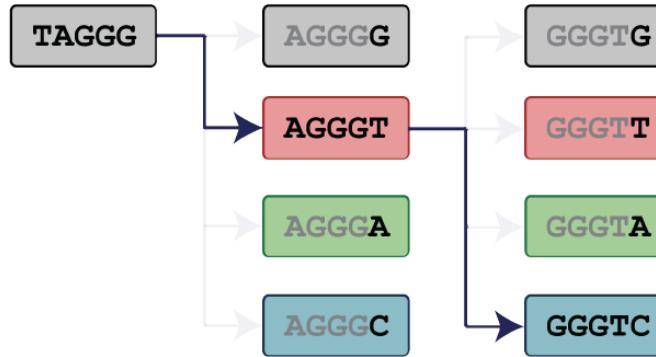


Oxford Nanopore: Sequencing by a nanopore

Data Workflow – (5-mer example)



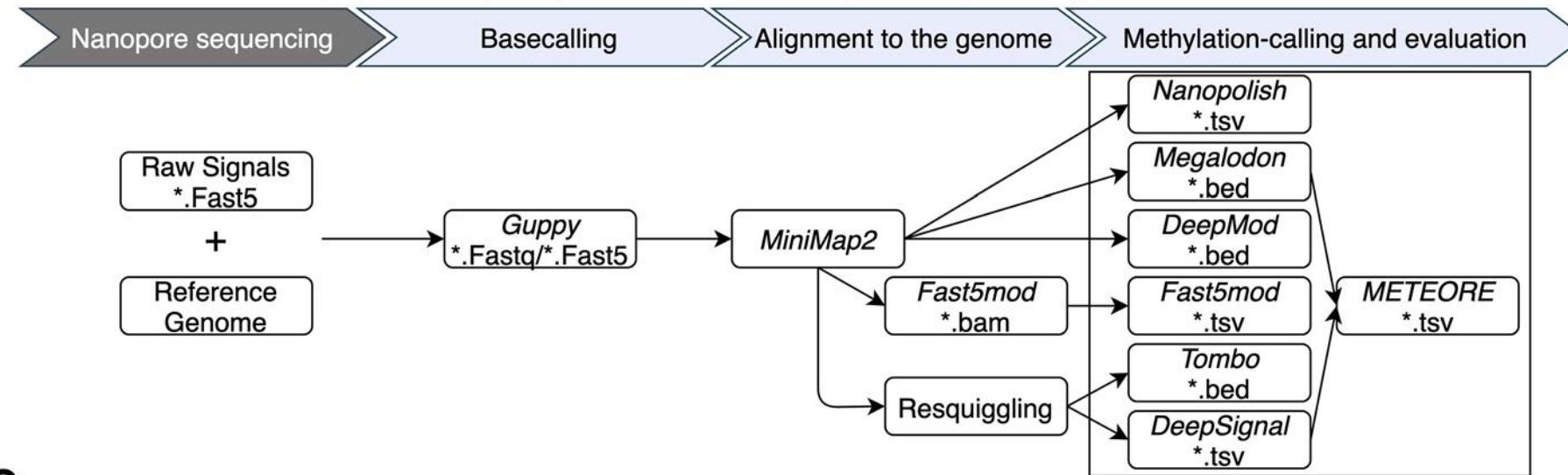
- Hidden Markov model
- Only four options per transition
- Pore type = distinct kmer length



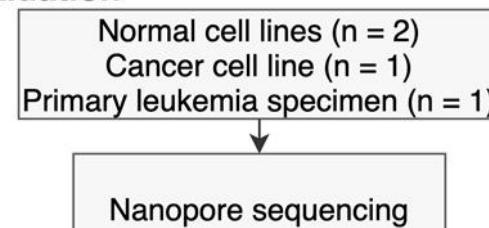
- Form probabilistic path through measured states currents and transitions
 - e.g. Viterbi algorithm

Oxford Nanopore: Detecting modified base

D Workflow for 5-methylcytosine (5mC) detection for nanopore sequencing



C Per-read and per-site performance evaluation



Liu et al. *Genome Biology* (2021) 22:295
<https://doi.org/10.1186/s13059-021-02510-z>

Genome Biology

RESEARCH

Open Access

DNA methylation-calling tools for Oxford Nanopore sequencing: a survey and human epigenome-wide evaluation

Yang Liu^{1†}, Wojciech Rosikiewicz^{1,2†}, Ziwei Pan^{1,3†}, Nathaniel Jillette¹, Ping Wang¹, Aziz Taghbalout¹, Jonathan Foo^{4,5}, Christopher Mason^{4,5,6,7}, Martin Carroll⁸, Albert Cheng^{1,3} and Sheng Li^{1,3,9,10*} 



Oxford Nanopore: Sequencing by a nanopore

- No more expensive camera!
- You can carry the sequencer in your pocket
- But how to increase accuracy?
 - ... two pores!
- Adaptive sampling
- Detecting modified bases directly



☰ Menu

[Stanford Medicine](#) / [News](#) / [Fastest genome sequencing](#)

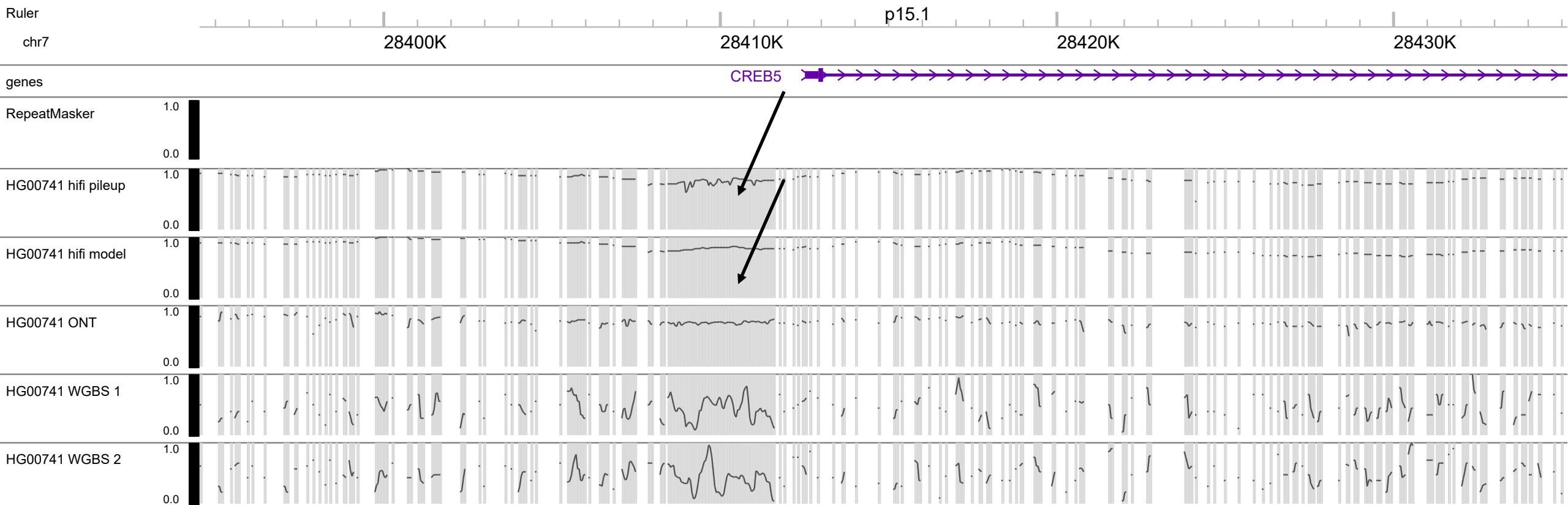
Fastest DNA sequencing technique helps undiagnosed patients find answers in mere hours

A research effort led by Stanford scientists set the first Guinness World Record for the fastest DNA sequencing technique, which was used to sequence a human genome in just 5 hours and 2 minutes.

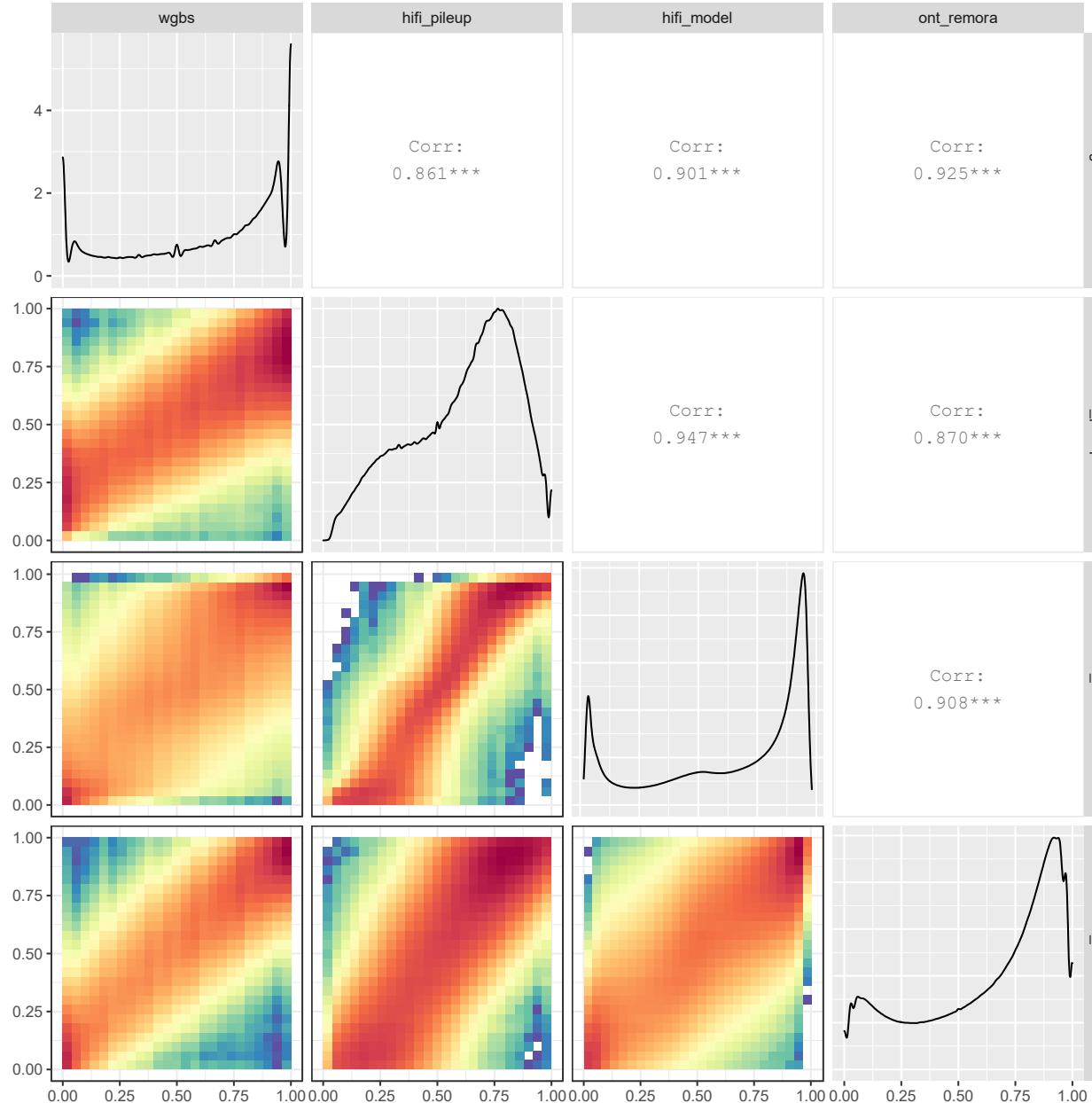
January 12, 2022 - By Hanae Armitage



Benchmarking DNA methylation calls



Benchmarking DNA methylation calls



- Both Pacbio and ONT methylation calling results are correlated well with the golden standard WGBS result
- The HiFi model apparently improves the methylation percentage estimate

Metrics per HiFi read (1x)

- **Precision:** 86%
- **Recall:** 85%
- **Accuracy:** 85%

An era of fast paradigm shifting

- First generation
 - Sanger sequencing
 - (Maxam and Gilbert's chemical cleavage technique)
- Second generation
 - Illumina
 - (SOLID)
 - (Ion Torrent)
 - (AVITI, Ultima, etc.)
- Third generation
 - PacBio
 - Oxford Nanopore
 - Illumina long-read (not really)
- Some math basics
- Sequence a genome
 - Assembly
- Re-sequence a genome
 - Variations
- Readout of functional genomics
 - YourName-seq