Just published on 08 April 2024

Zhang *et al. Genome Biology* (2024) 25:90
https://doi.org/10.1186/s13059-024-03235-5

Genome Biology

SHORT REPORT **Open Access**

scifi-ATAC-seq: massive-scale single-cell chromatin accessibility sequencing using combinatorial fluidic indexing

Xuan Zhang[1], Alexandre P. Marand[1,2], Haidong Yan[1] and Robert J. Schmitz[1*]

GeneCore EMBL

Sequencing today:
an end point or the new beginning?…

Vladimír Beneš

EMBL Course: “scATAC-seq: attacking open chromatin in single cells”

Heidelberg, 08-12 April 2024

GeneCore EMBL

47 years ago…

Proc. Natl. Acad. Sci. USA
Vol. 74, No. 2, pp. 560–564, February 1977
Biochemistry

977

Biochemistry

A new method for sequencing DNA

(DNA chemistry/dimethyl sulfate cleavage/hydrazine/piperidine)

ALLAN M. MAXAM AND WALTER GILBERT

Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, Massachusetts 02138

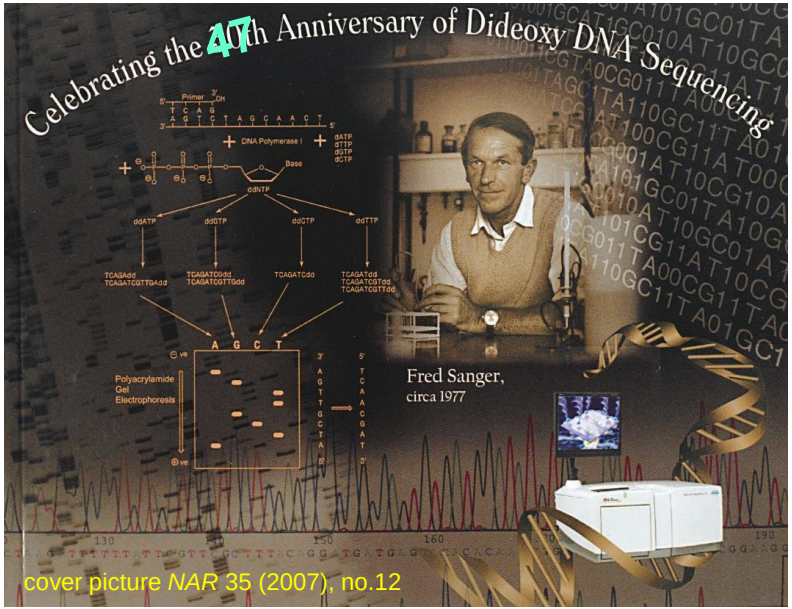
Contributed by Walter Gilbert, December 9, 1976

ABSTRACT DNA can be sequenced by a chemical procedure that breaks a terminally labeled DNA molecule partially at each repetition of a base. The lengths of the labeled fragments then identify the positions of that base. We describe reactions that cleave DNA preferentially at guanines, at adenines, at cytosines and thymines equally, and at cytosines alone. When the products of these four reactions are resolved by size, by electrophoresis on a polyacrylamide gel, the DNA sequence can be read from the pattern of radioactive bands. The technique will permit sequencing of at least 100 bases from the point of labeling.

Fred Sanger, Walter Gilbert & Paul Berg:
Nobel prize in chemistry 1980

GeneCore EMBL

Nucleic-acids sequencing – state of the art



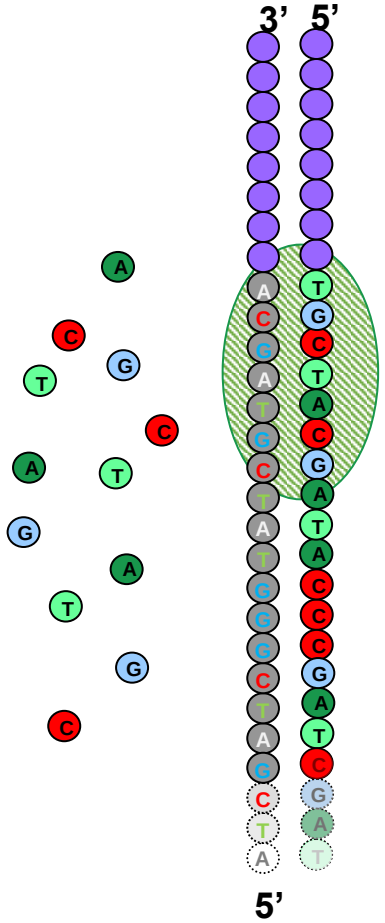
Sanger sequencing (1977 to date)

2024

- Sequencing by synthesis (SBS)
  - Sequencing by avidity (DNA pol still needed)
  - Sequencing with DNA nanoballs (DNA pol still needed)
  - Sequencing by binding (DNA pol still needed)
- Sequencing by threading through the nanopore
- Sequencing by cascade pyrophospohorolysis
- Sequencing by TEM
- Sequencing by base specific modification & cleavage

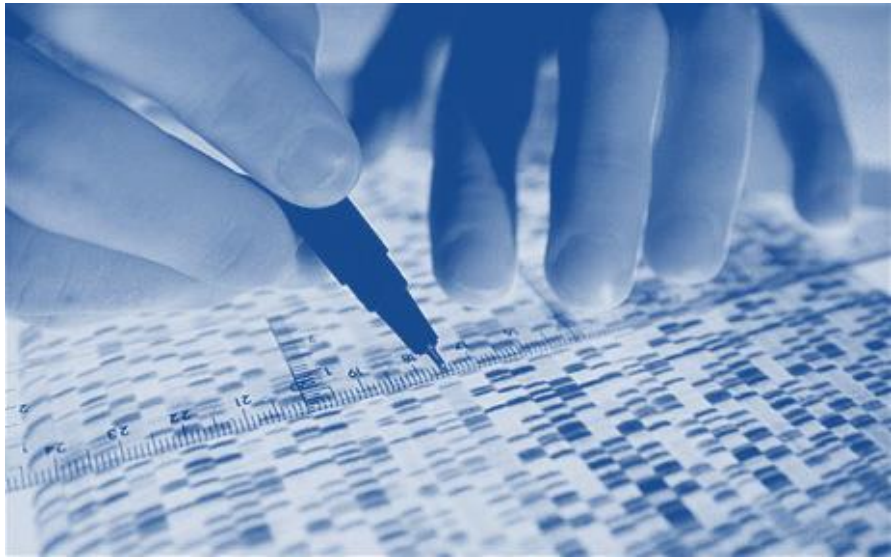
…

Sequencing by *de novo* synthesis of the ‘daughter’ strand



GeneCore EMBL

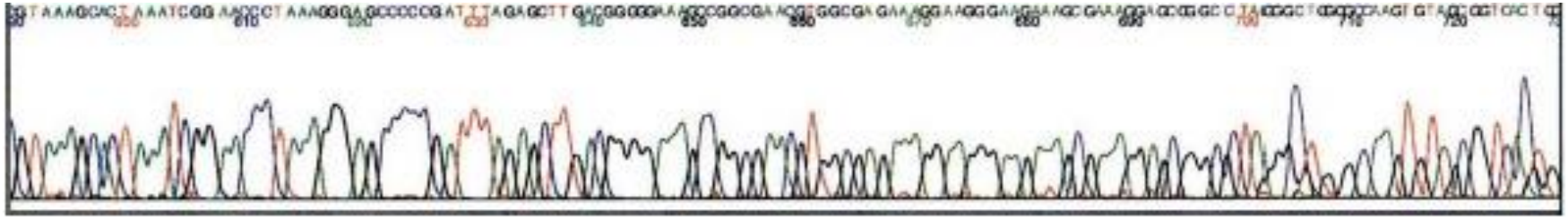
First generation sequencing



GeneCore EMBL

Developments in technology and biochemistry towards higher ‘everything’

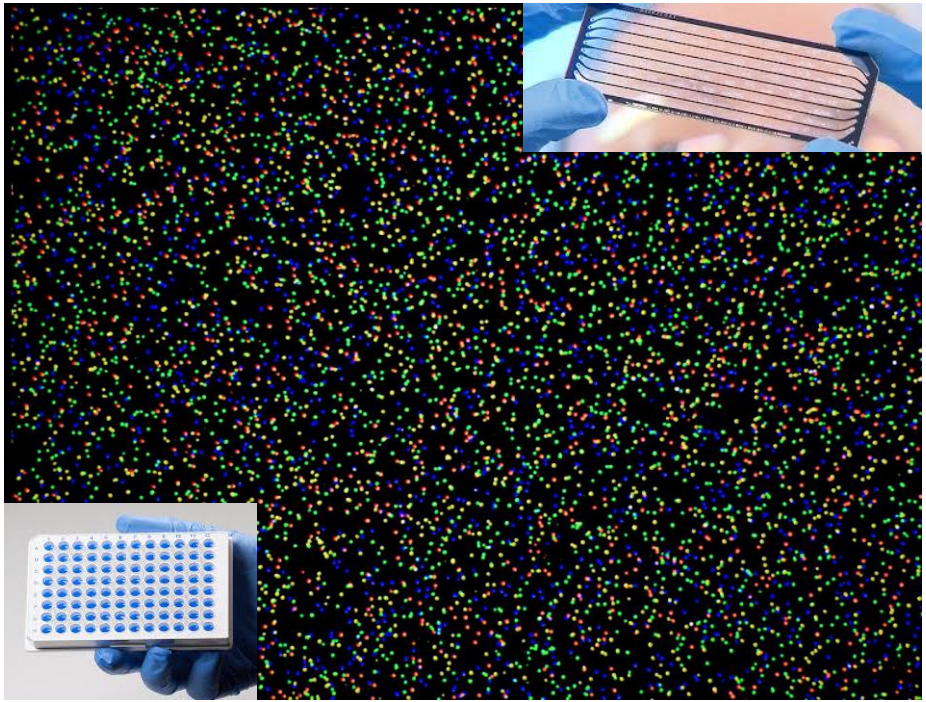
- Radioactivity gave way to the fluorescence detection
- Introduction of DNA polymerases non-discriminating between dNTPs and ddNTPs
- Fluorescently labeled primers gave way to fluorescently labeled terminators
- Slab polyacrylamide gels gave way to capillaries, *first generation sequencing starts*
- Introduction of thermostable DNA polymerases
- Linear single-pass primer extension evolved to the cycling protocol
- Optimization of sample preparation methods and reaction conditions
- **Significant reduction of the DNA input amount** (from micrograms to nanograms!)
- **Increased accuracy and read-length**



GeneCore EMBL

The second, “next” generation sequencing

Reading optical signals as bases and their analogs are added in the *de novo* DNA synthesis steps

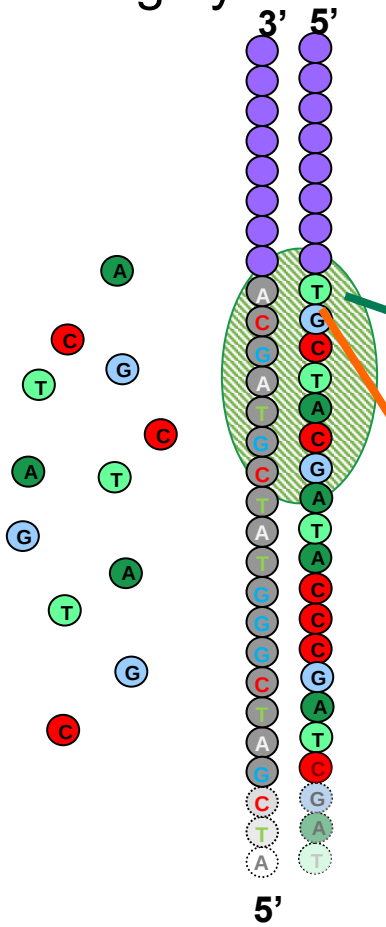


NGS *modi operandi*

- Short-/long-read technologies
- Clonal amplification still required for short-read technology (clusters, DNB, polonies) => only an indirect detection of modified bases possible
- Sequencing by *de novo* synthesis:
  - ‘static’ (Illumina & co.) *vs.* ‘dynamic’ real-time detection (PacBio)
- Detection: mostly fluorescence but strong efforts to improve signal-to-noise ratio
- Sequencing by threading through the nanopore: ‘dynamic’ real-time detection through sensing of the change of electric current through the pore
- Error model: ‘static’ mode: mostly substitutions

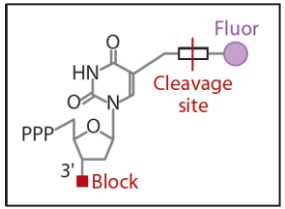
  ‘dynamic’ mode: mostly insertions/deletions
- Limitations of Poisson distribution
- Speed: nanopore by far the fastest => real-time applications, *e.g.*, FJ Müller & H Kretzmer (MPI Molecular Genetics, Berlin): Real-time cancer classification with nanopore sequencing, https://www.youtube.com/watch?v=DmDZEyvkmOo

Sequencing by '*de novo*' synthesis



Cycle 1: Add sequencing reagents

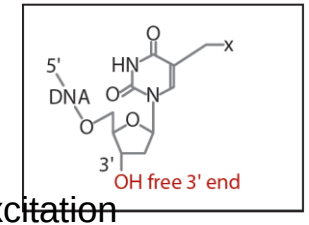
First base incorporated



Emission

Detect signal

Cleave off the block (terminator) and dye

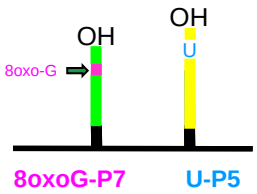


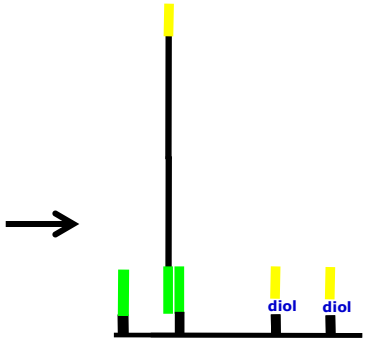
Excitation

Cycle 2-n: Add sequencing reagents and repeat

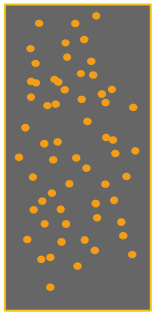
GeneCore EMBL

Cluster generation: fragment binding

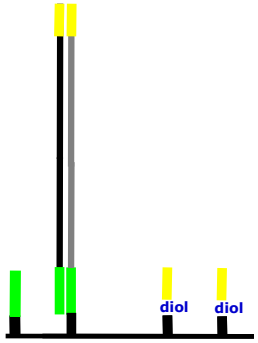




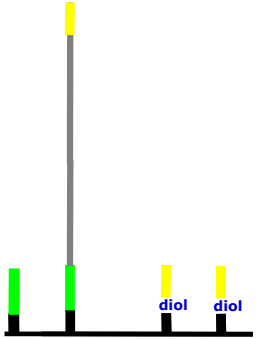
Template hybridization



Single molecules bound to a flow cell in a random pattern



Initial extension

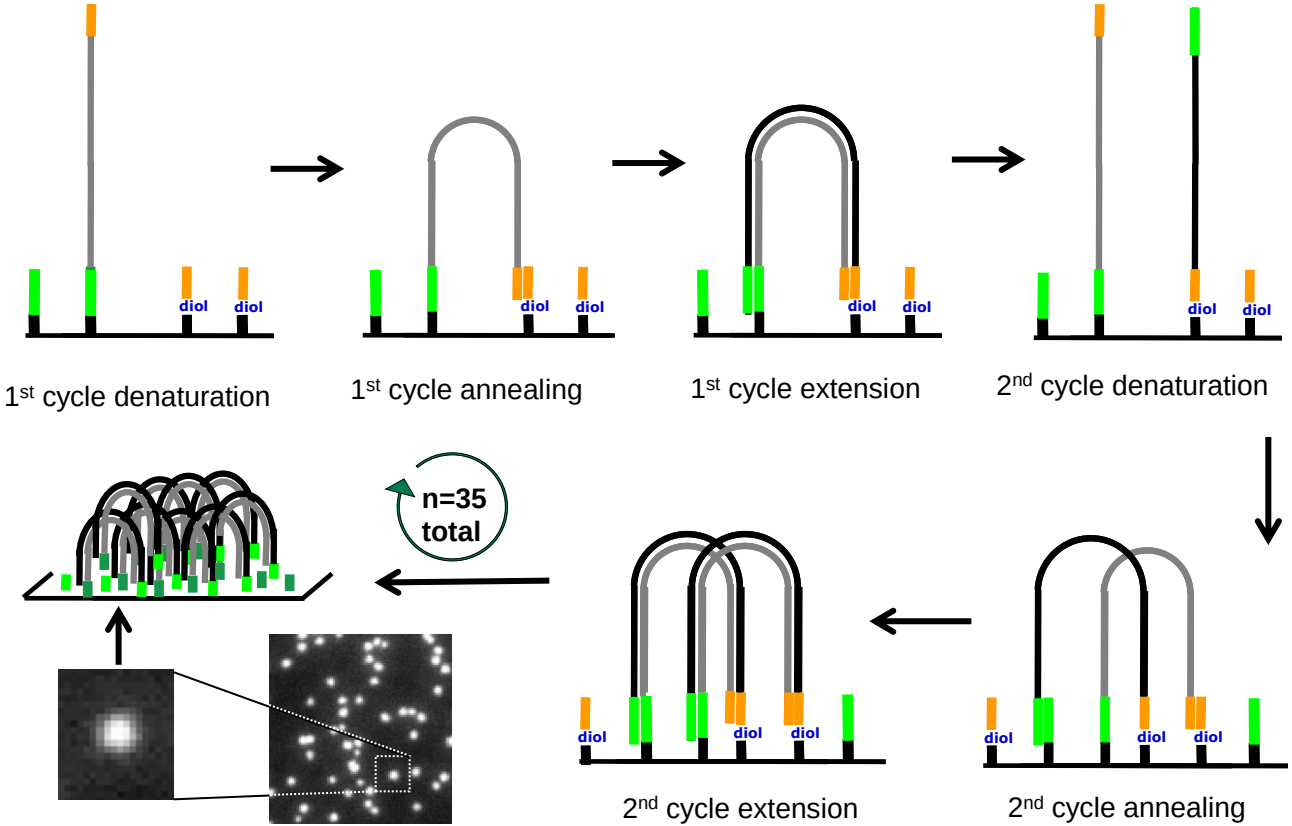


Denaturation

Grafted flowcell
Single-/paired-end
Uracil specific excision reagent
Formamidopyrimidine glycosylase (fpg)

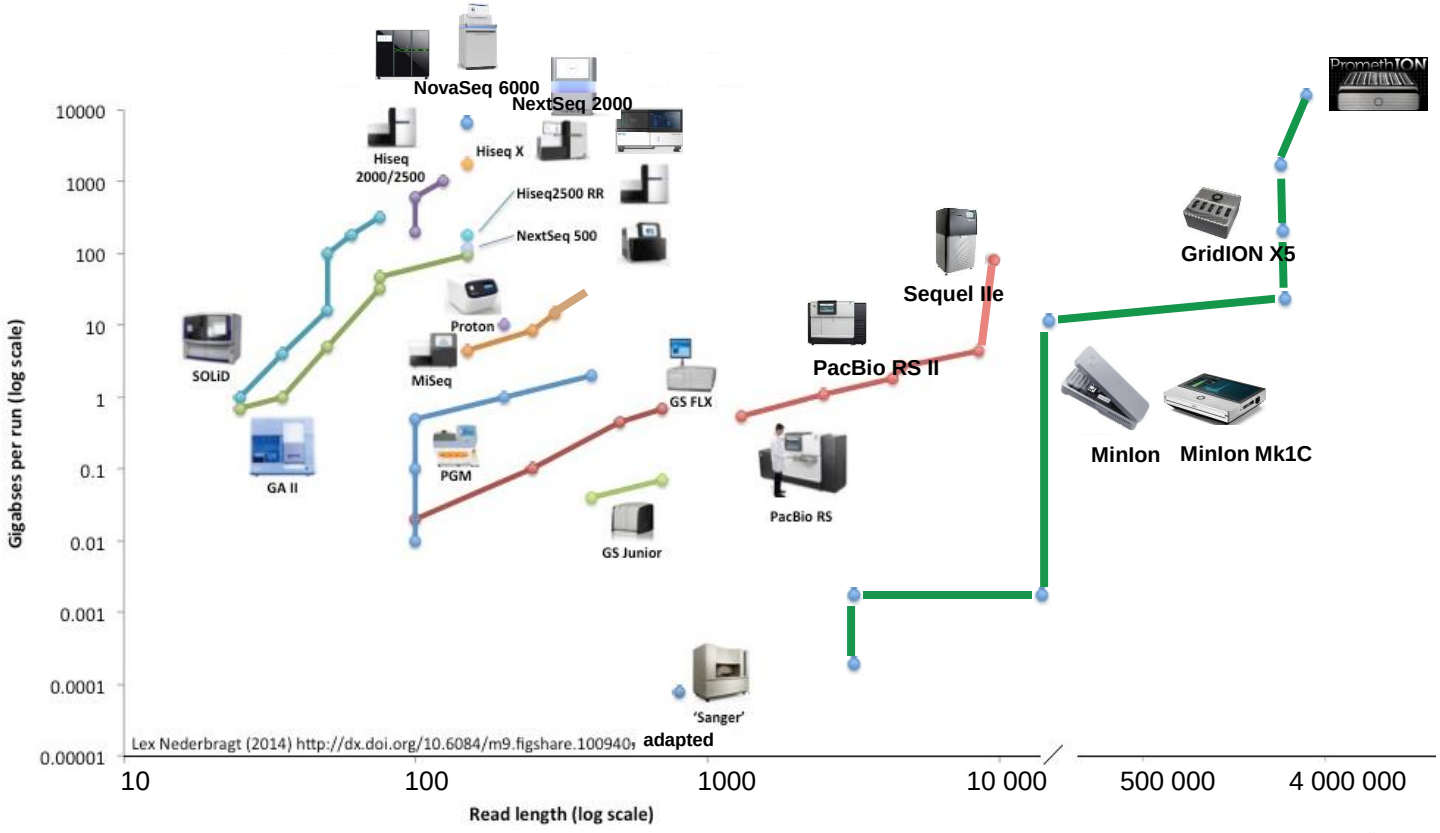
GeneCore EMBL

Cluster generation: amplification



GeneCore EMBL

Sequencing technology (r)evolution



GeneCore EMBL

Short-read NGS technology has become interesting again…

Element Biosciences

AVITI, Avidity

SINGULAR GENOMICS

G4

PACBIO

Onso, SBB

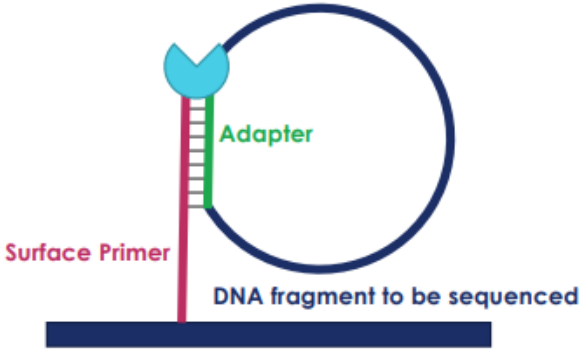
MGI

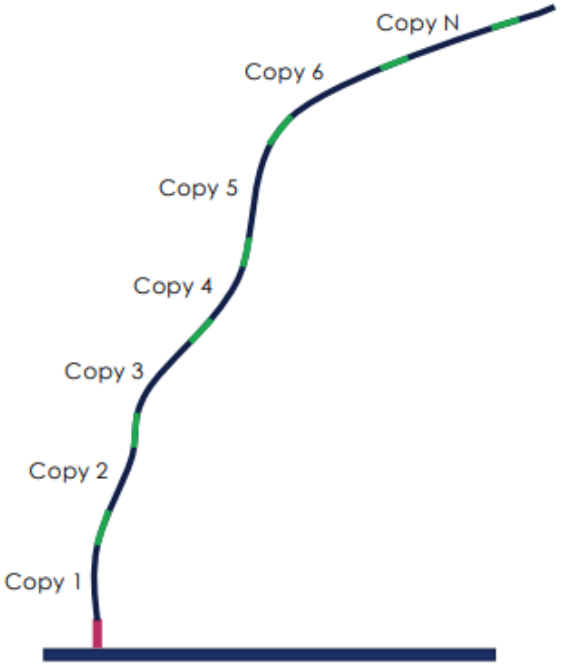
ULTIMA GENOMICS

Sequencers are not to scale.

GeneCore EMBL

AVITI – rolling circle amplification on the surface: polony generation



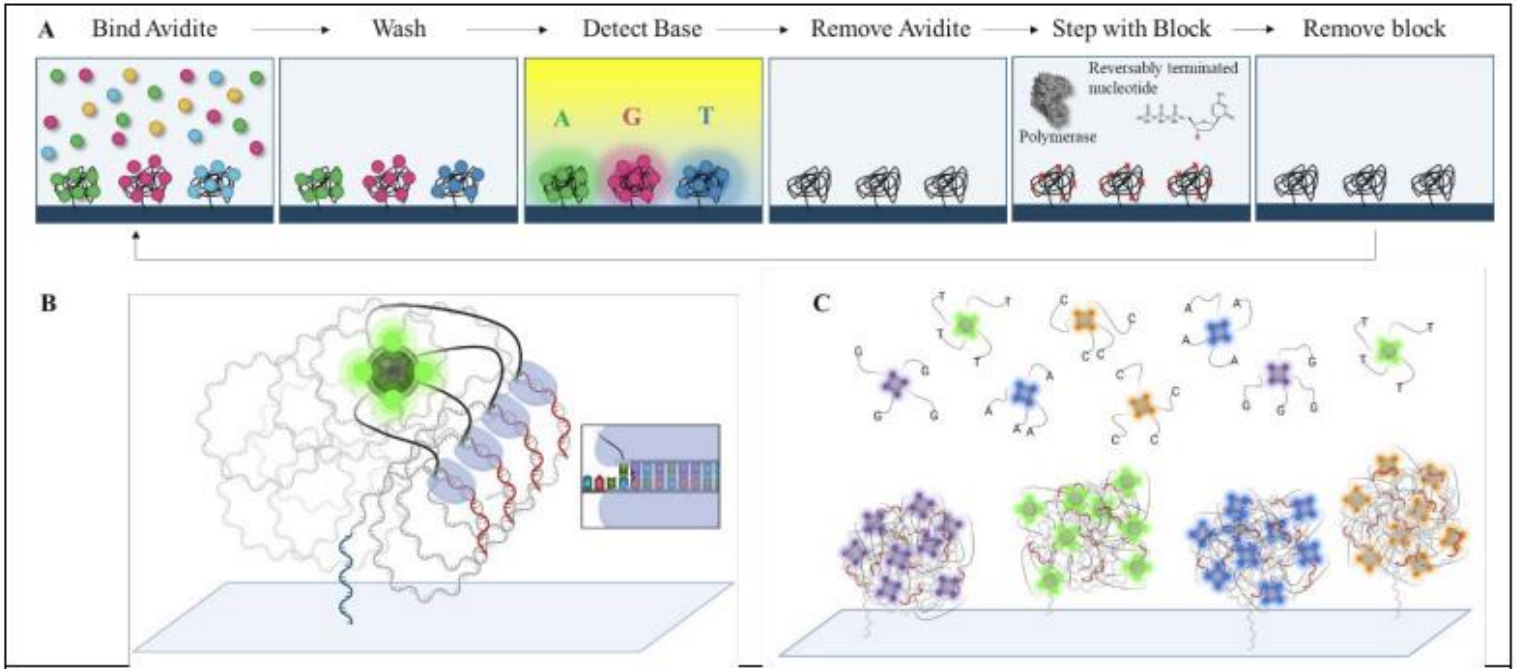


10^3 copies of substrate DNA in a close proximity: a polony

Polony: polymerase colony by clonal amplification of a single DNA molecule

GeneCore EMBL

Sequencing by avidity

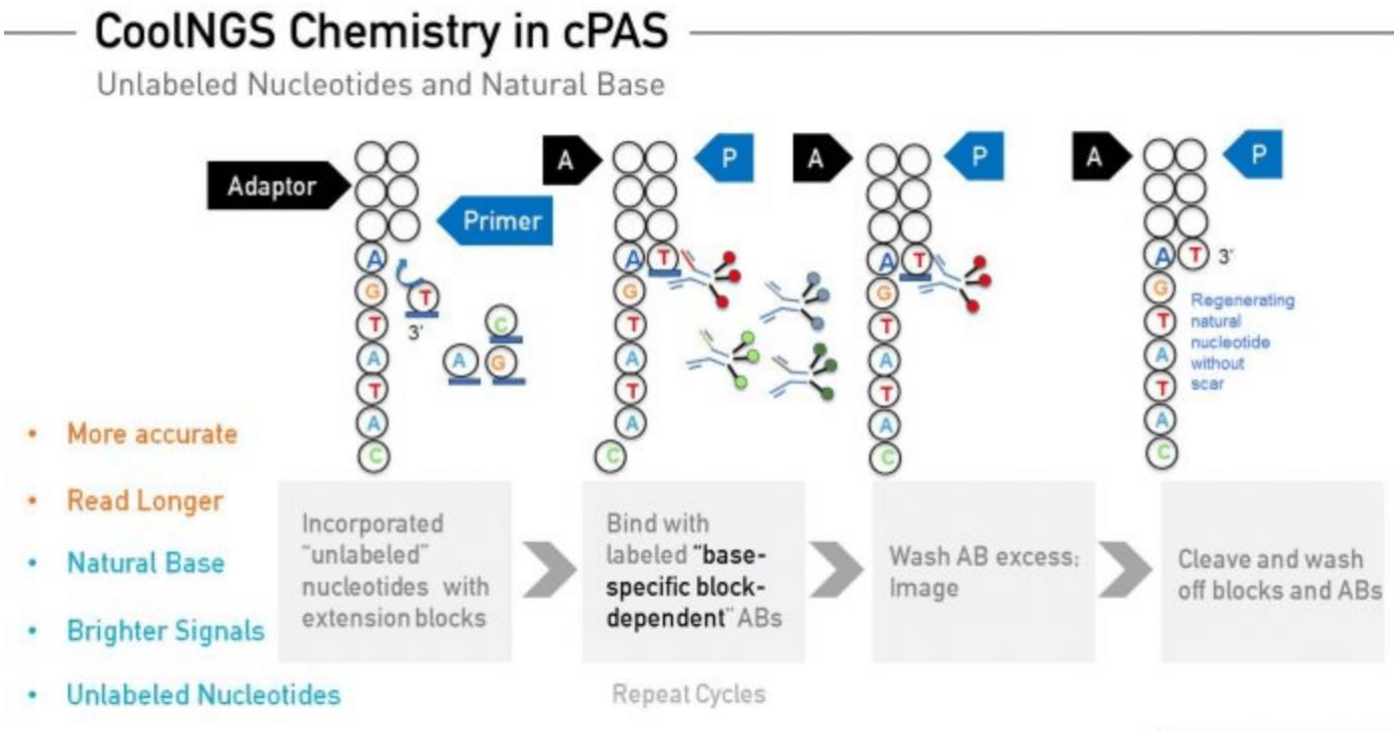


Avidity: the accumulated strength of multiple affinities of individual noncovalent binding interactions

Arslan *et al*., *Nature Biotechnology* (2023), doi: 10.1038/s41587-023-01750-7

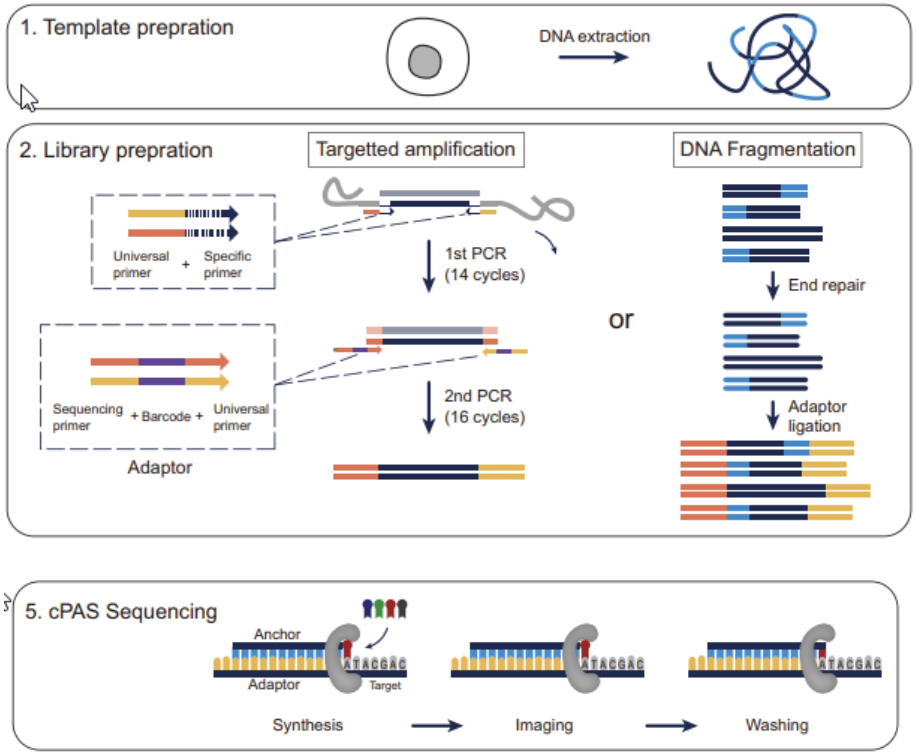
GeneCore EMBL

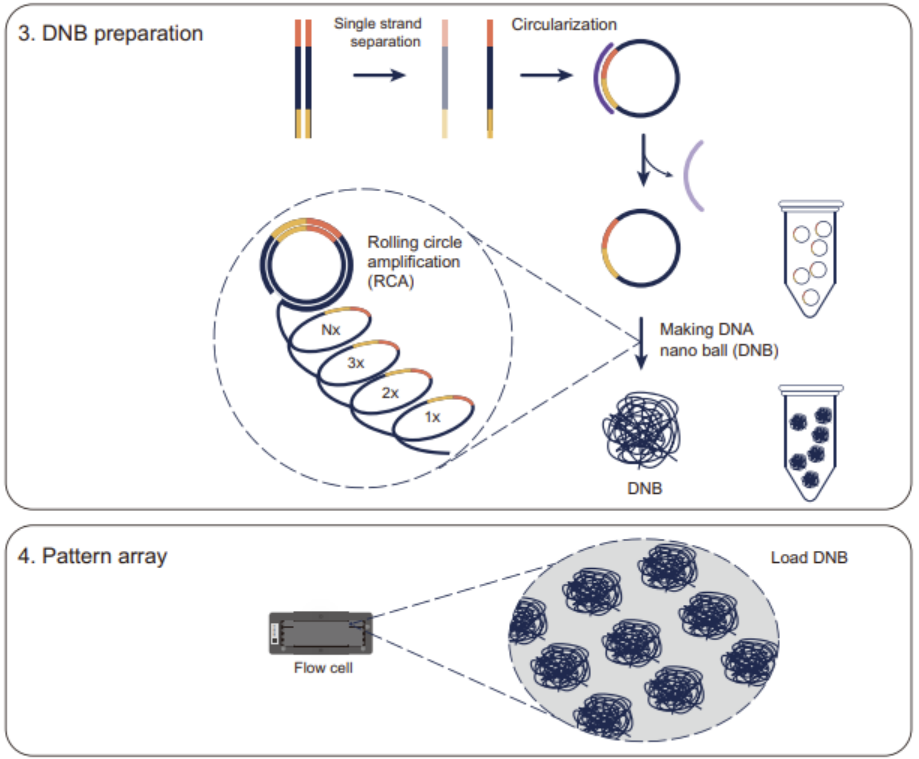
A novel way to detect incorporated bases



cPAS = combinatorial probe anchored synthesis; image courtesy by MGI
https://www.youtube.com/watch?v=0bqIBtMS3xw

DNA nanoballs





Li *et al*. *Intl J Legal Medicine* (2021), doi: 10.1007/s00414-021-02507-0

GeneCore EMBL

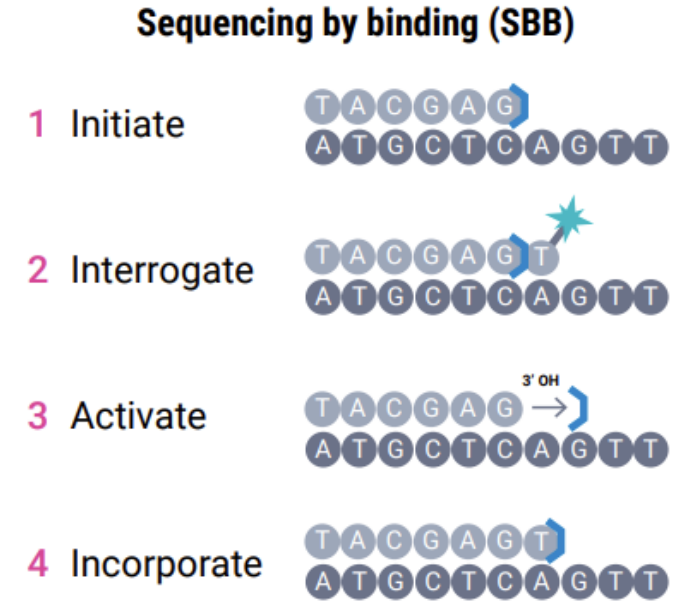


Figure 1. Conceptual diagram of SBB sequencing.

(1) Each cycle initiates with a 3′ reversible blocked nucleotide. (2) Fluorescently labelled nucleotides are then flowed over the flow cell allowing for the appropriate base to bind. Unbound nucleotides are washed away so that the base can be interrogated with reduced background signal. (3) The 3′ end of the nucleotide is activated via removal of the reversible terminator. (4) Native, unlabeled, reversible blocked nucleotides are flowed over the flow cell and the cognate base can incorporate into the growing strand. The process is then repeated for each new sequenced base.

By separating the interrogation and incorporation steps, and optimizing chemistries for each step, SBB chemistry enables high signal-to-noise levels to be achieved. Moreover, the lack of molecular scarring from the use of native nucleotides for incorporation drives industry-leading accuracy.

https://www.pacb.com/wp-content/uploads/SBB-Product-note.pdf

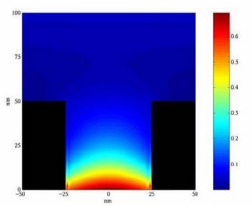
PacBio

GeneCore EMBL

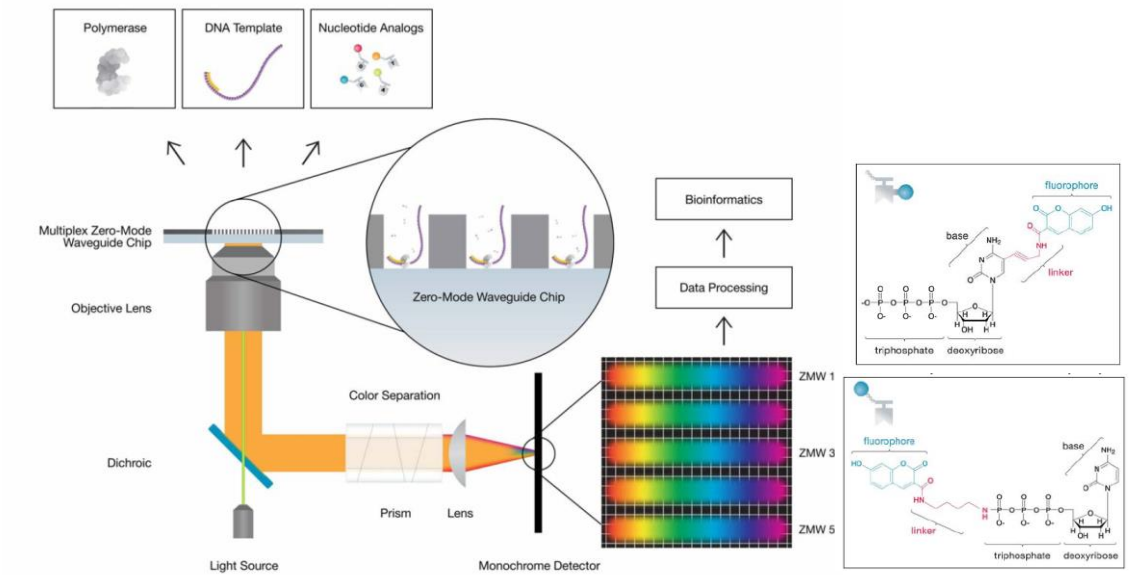
The third generation: long-read/single-molecule sequencing

PacBio Sequel IIe
complete bacterial genomes
detection of modified bases limited
Long reads with a **very** high accuracy

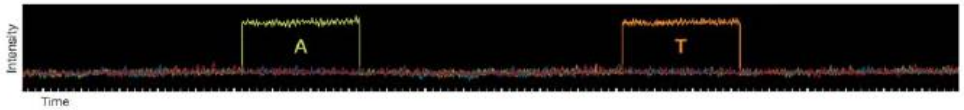
Pacific Biosciences (PacBio)



Detection volume



Modified bases



By courtesy of Pacific Bioscience

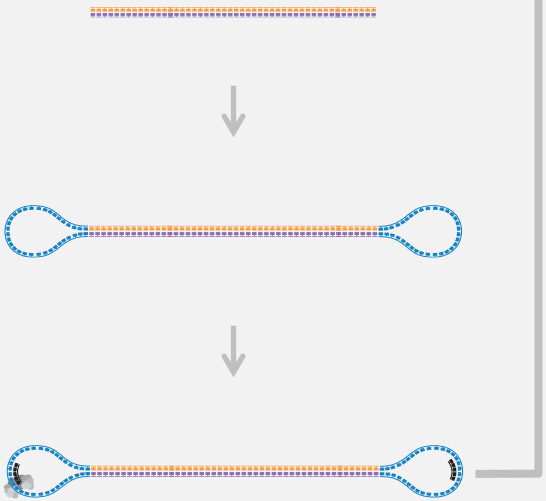
GeneCore EMBL

PacBio: highly accurate long HiFi reads

Start with **high-quality** double-stranded DNA

Ligate SMRTbell adaptors and size select

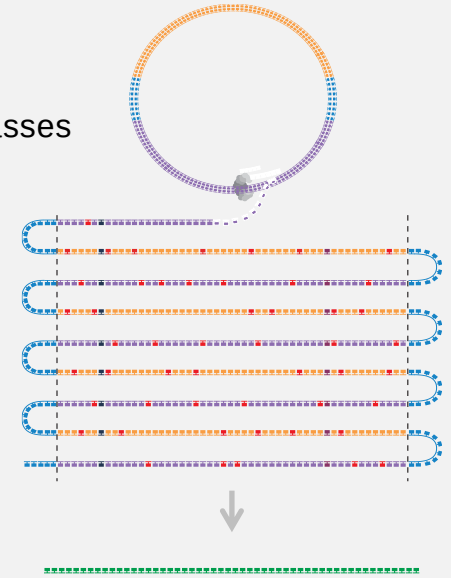
Anneal primers and bind DNA polymerase



Circularized DNA is sequenced in repeated passes

The polymerase reads are trimmed of adapters to yield sub-reads

Consensus is called from sub-reads



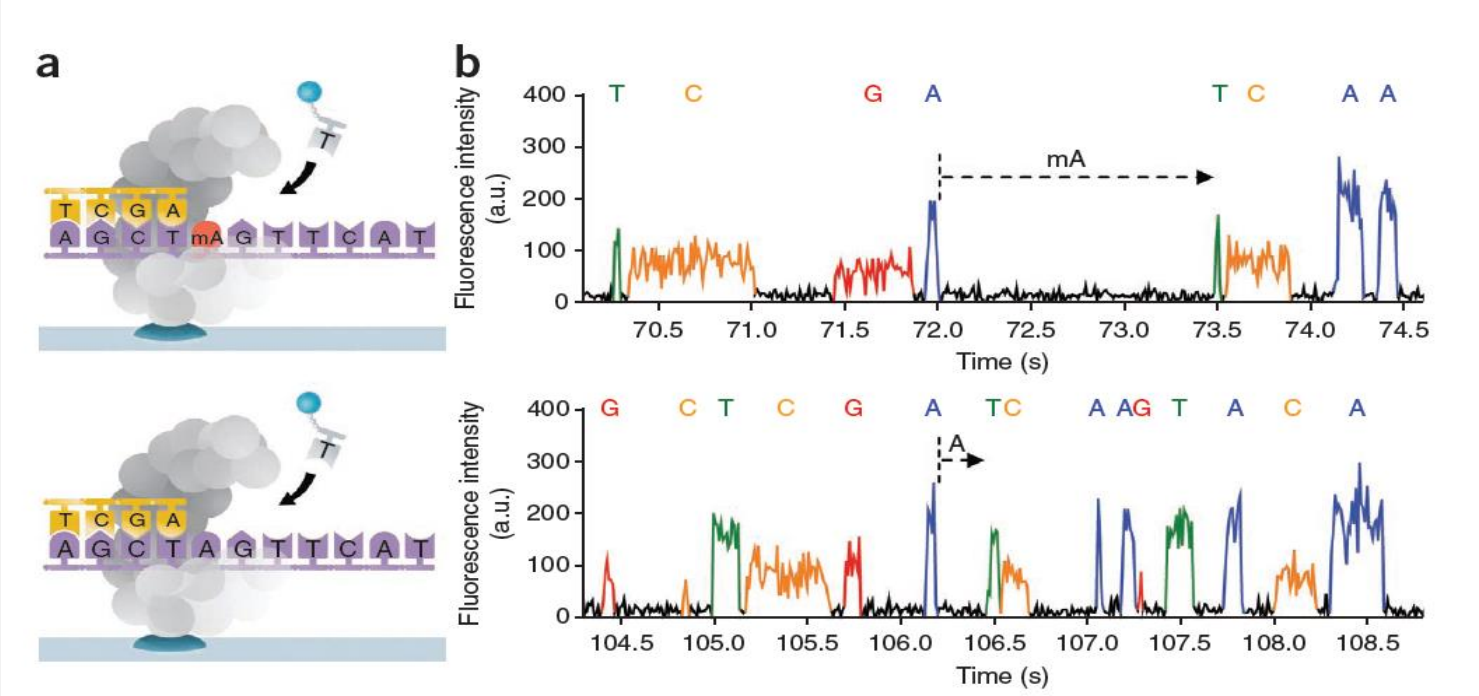
HiFi reads

>99.9% accuracy

Fragments up to 25 kb

GeneCore EMBL

Direct detection of modified bases



Flusberg *et al.*, *Nature Methods* (2010)

GeneCore EMBL

The fourth generation: our wish list

- Higher throughput
- Rapid turnaround
- Little starting material
- No complex library preparation
- Label free

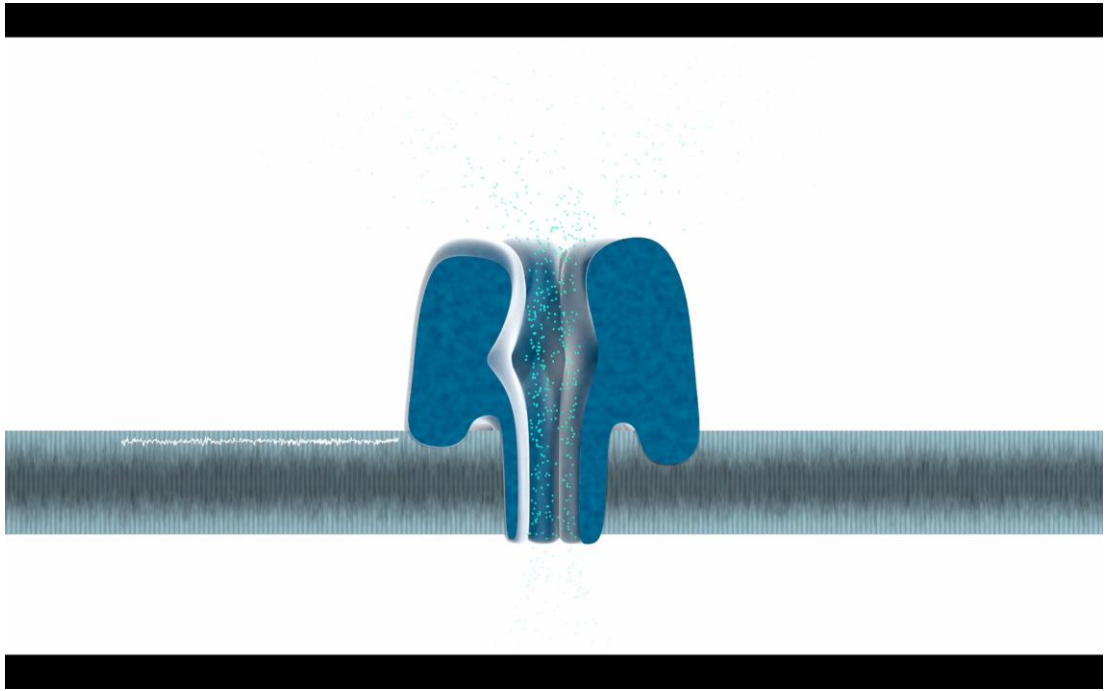
- Higher accuracy
- Longer reads
- Lower cost

Single Molecule + *Continuous measurement* + *Light off*

The fourth generation: nanopore sequencing (still revolutionary)

ONT Grid- & Prometh-ION
Direct RNA sequencing
unrestricted detection of modified bases
Very long reads with a lower accuracy

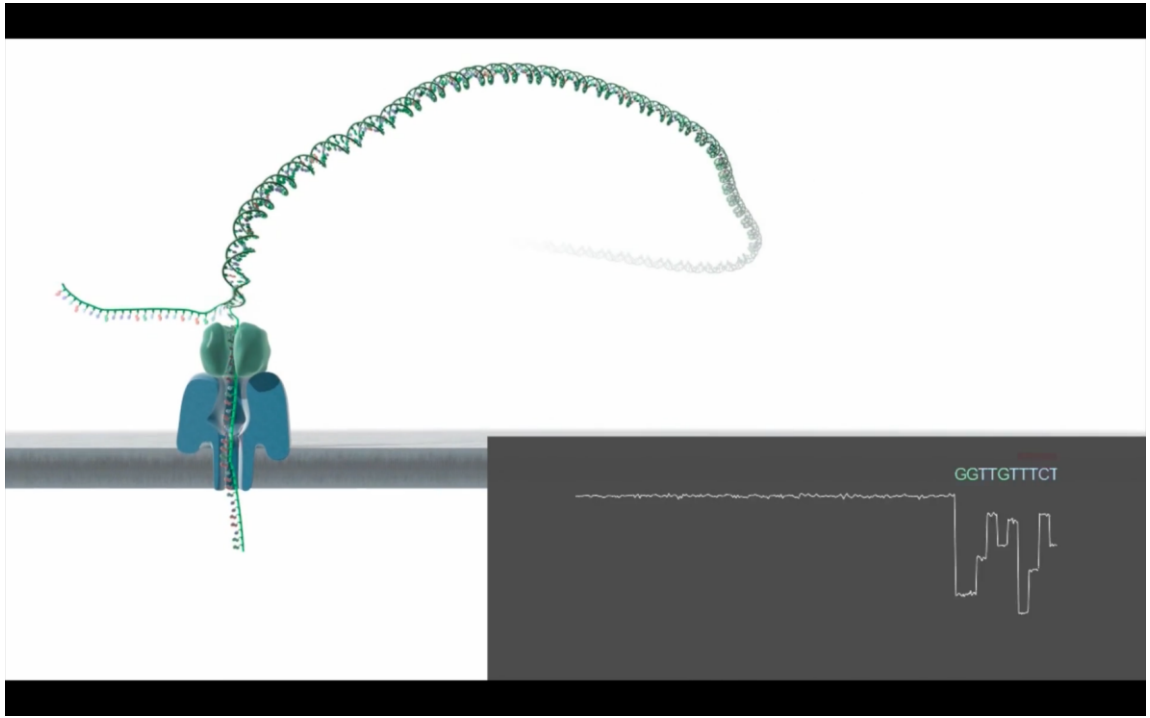
How it works – the pore



Oxford NANOPORE Technologies

GeneCore EMBL

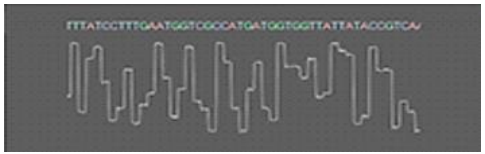
How it works - sequencing



GeneCore EMBL

Nanopore: the state-of-the-art

Algorithms: HMM, RNN etc...



Engine protein
highly processive helicase E8

Nanopore reader: R9, R10
E. coli lipoprotein CsgG

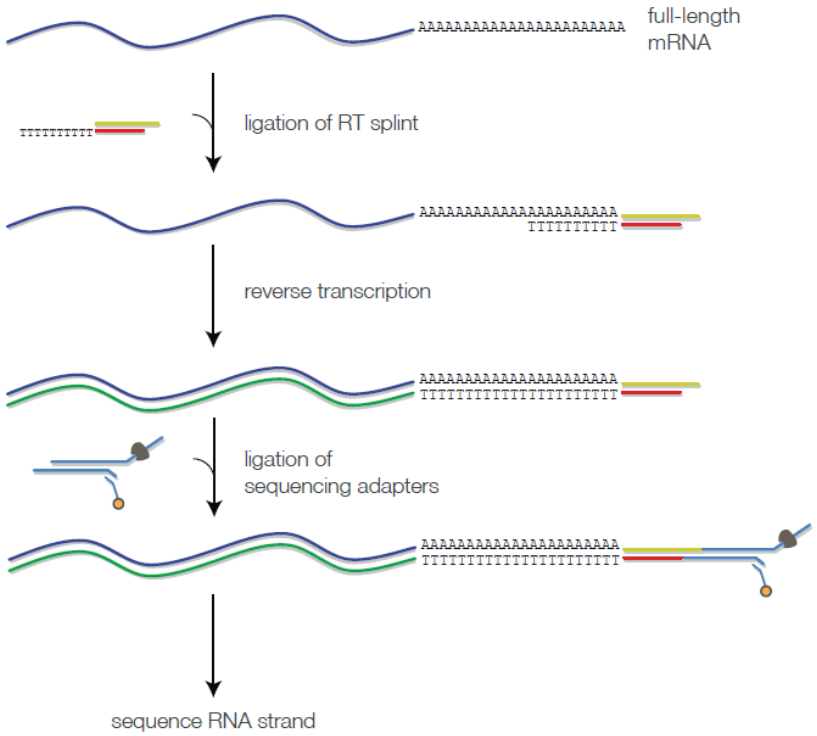
Running conditions:
Salt, ‘fuel’, script, temperature...

Membrane
M10, synthetic, not lipid bi-layer

Animation & pictures on this and 2 preceding slides by courtesy of ONT

GeneCore EMBL

Direct RNA sequencing with nanopore



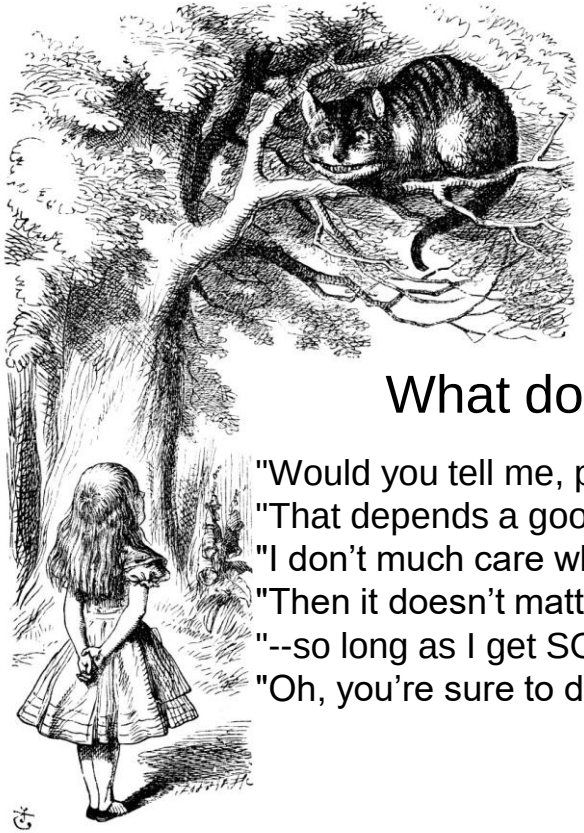
Garalde *et al*, *Nature Methods* (2018), doi: 10.1038/nmeth.4577

GeneCore EMBL

NGS features

- Unprecedented discovery power
- Hypothesis-forming potential
- Heading towards unbiased results
- Sensitivity & specificity
- Almost whole-genome, -transcriptome, -methylome, -epigenome *etc*… view

Importance of an ‘experimental’ design



What do I want to achieve?

"Would you tell me, please, which way I ought to go from here?"
"That depends a good deal on where you want to get to," said the Cat.
"I don’t much care where--" said Alice.
"Then it doesn’t matter which way you go," said the Cat.
"--so long as I get SOMEWHERE," Alice added as an explanation.
"Oh, you’re sure to do that," said the Cat, "if you only walk long enough.”

Lewis Carroll’s Alice in Wonderland

GeneCore EMBL

Experimental design/decisions, decisions…

Points to consider

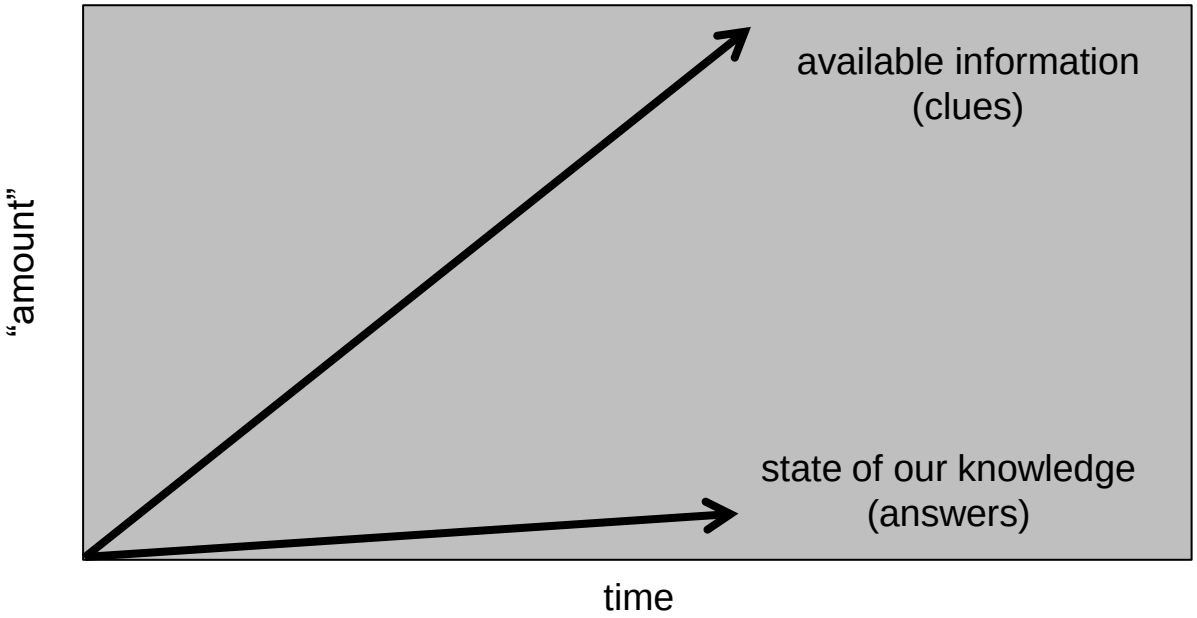
- Background information on the studied system (*e.g.*, genome size)
- Sample choice and its homogeneity, contamination
- Complexity of a “sample”, its repertoire, incl. base composition
- Proper sampling
- Biological replication, randomization, controls
- Sample harvesting, DNA/RNA isolation, library preparation
- Read-type, read-length, read-depth/coverage
- Data QC, analysis, corroboration & validation
- *What comes next?*

GeneCore EMBL

"Data don't make any sense,
we will have to resort to statistics."

biologist *vs.* bioinformatician
https://www.youtube.com/watch?v=Hz1fyhVOjr4

Caution required



We are drowning in information and starving for knowledge.
R. D. Roger

GeneCore team 2024

Hilal Özgür
NGS, PacBio

Mireia Osuna Lopez
NGS, scRNAseq, PacBio

Daphne Welter
NGS, nanopore

Jessica Sutter
NGS, scRNAseq

Ferris Jung
NGS, Automation

Laura Villacorta
NGS, scRNAseq, nanopore

Tobias Rausch
Bioinformatics
Korbel group

Jonathan Landry
Bioinformatics
Leaving in VI. 2024

Jan Provaznik
Bioinformatics, LIMS

Viviane Heins
NGS, Olink
GSK-Cellzome

Claudio Asencio
NGS, methods
GSK-Cellzome

GeneCore EMBL

Thank you!

https://www.genecore.embl.de
solexateam@embl.de

Photo: Konstantin Milberger