UPPSALA UNIVERSITET

Introduction to Next Generation Sequencing

Olga Vinnere Pettersson, Uppsala University

Scientific Lead of Planetary Biology Strategic Area, SciLifeLab

ERGA Vice-chair

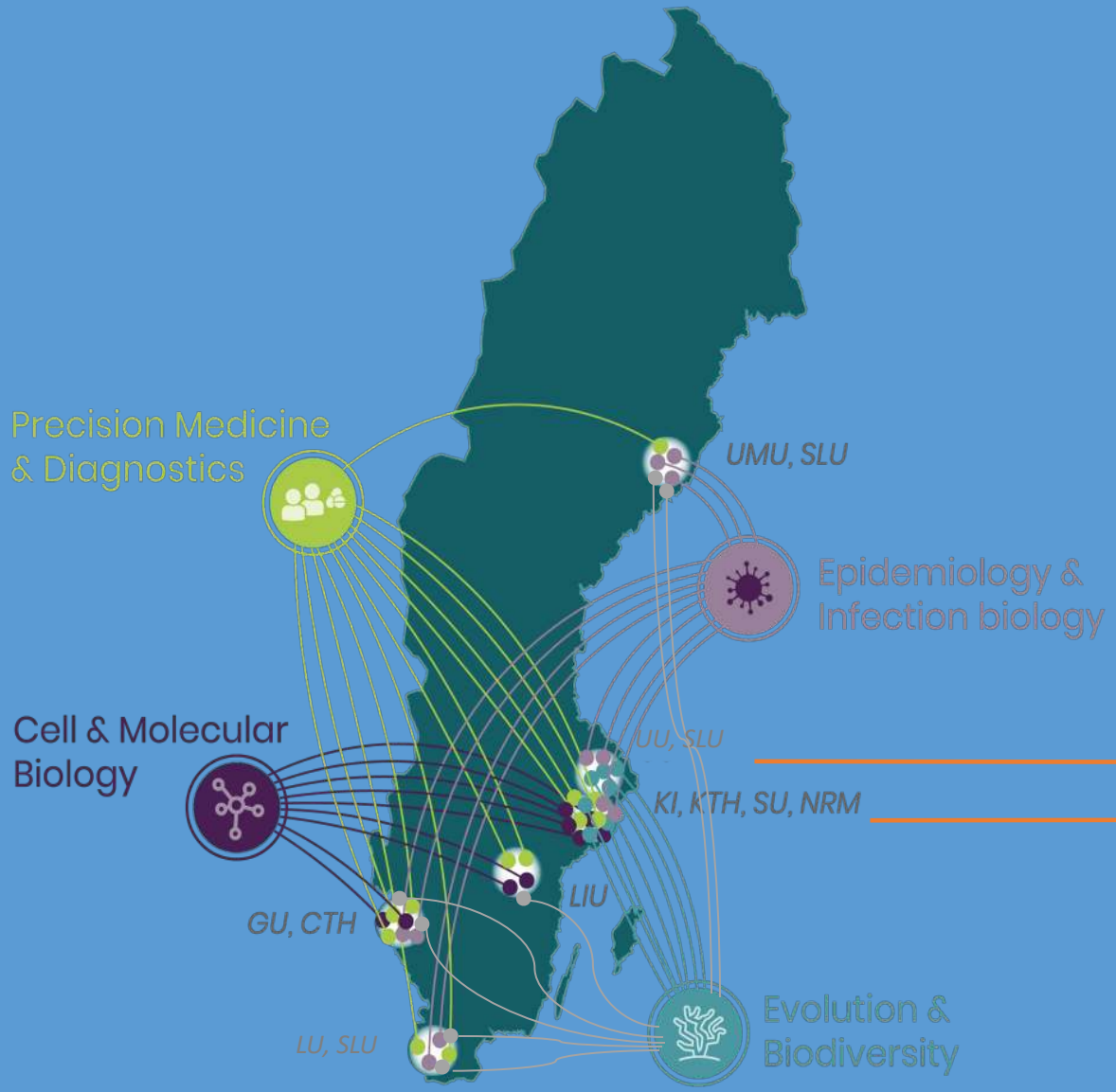
EVOMICS,
Český Krumlov
2026-01-12

Outline

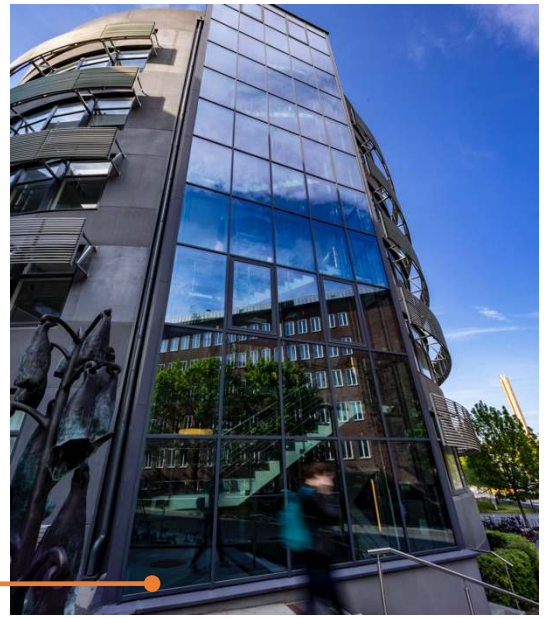
- Brief history of Genomics
- What is sequencing?
- Glossary
- How do sequencing technologies work?
  - Short reads: Illumina, Ion, MGI, Element Biosciences (Aviti), Roche
  - Long reads: PacBio and ONT

Where am I from: SciLifeLab

SciLifeLab



Sweden's national center for molecular biosciences

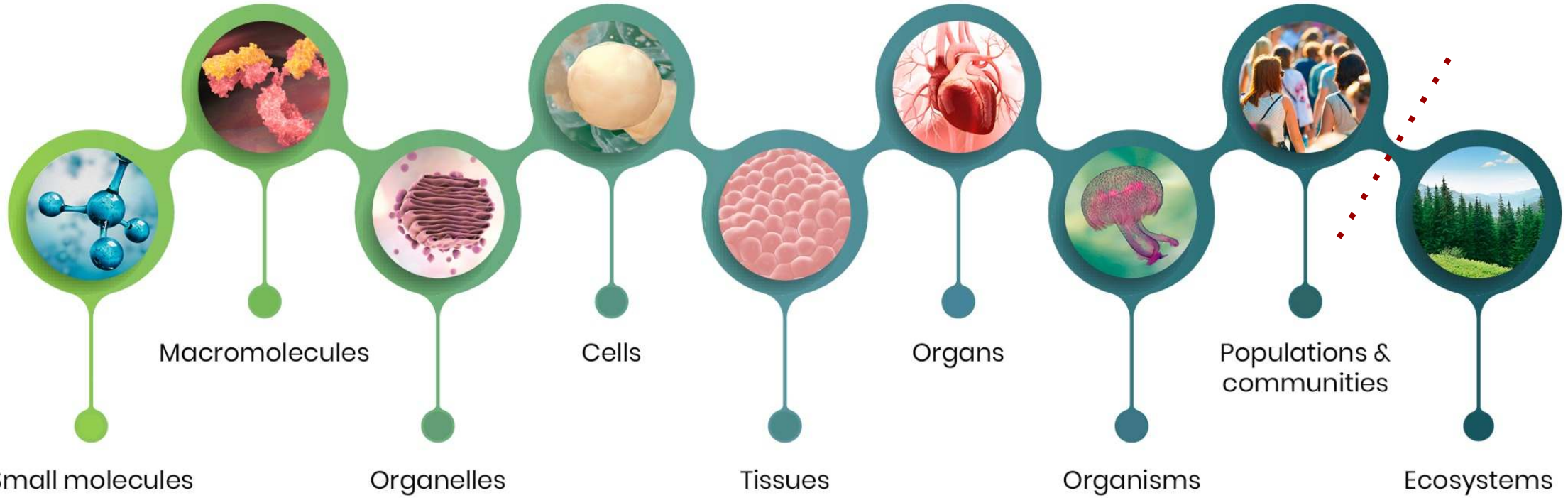


Connecting strong research environments

Enabling research across the full spectrum of life science

SciLifeLab infrastructure technologies:

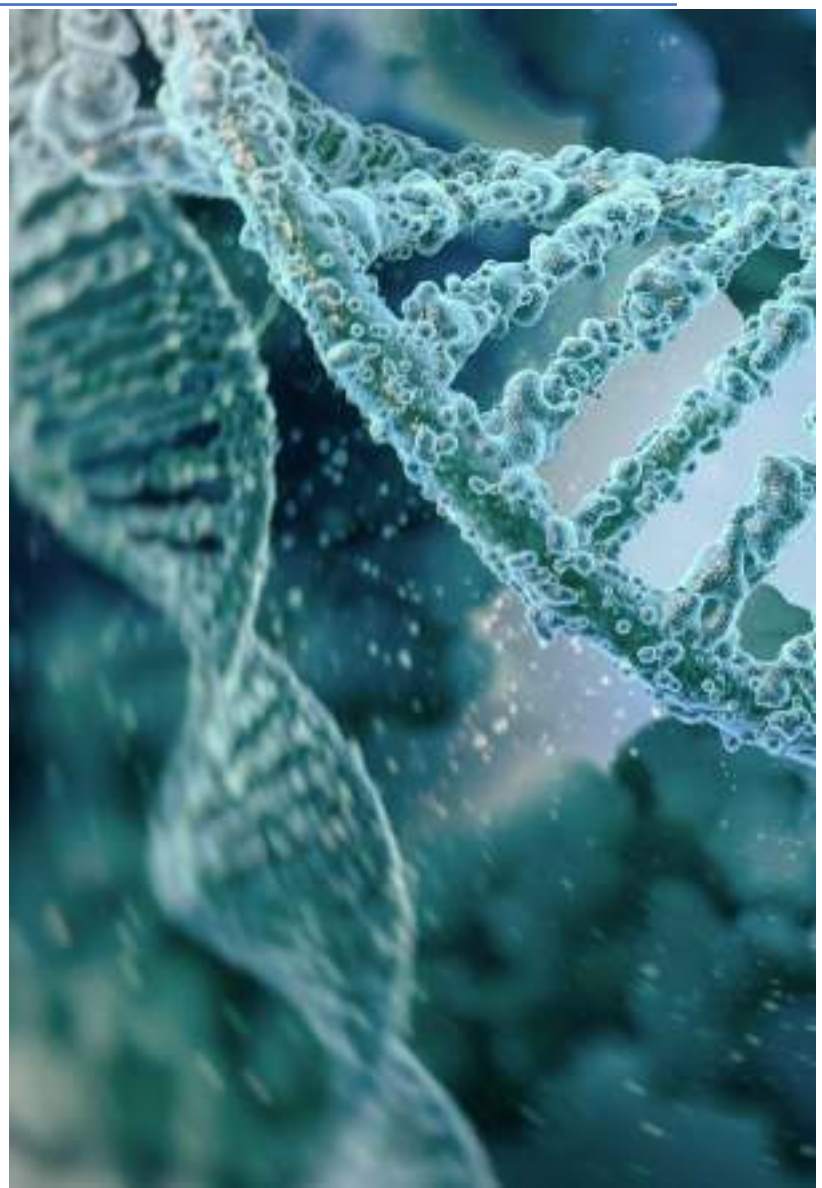
- Can be used to study the molecular aspects of life **ranging from the atomic scale up to entire ecosystems**
- Are applicable across a **large spectrum of disciplines and research fields** in life science
- Are **available to all academic researchers in Sweden** on equal terms
- Are available to **healthcare and industry** all over the country, as well as international users



SciLifeLab Genomics Platform

- Consists of:
  - **National Genomics Infrastructure** – 3rd largest NGS core in Europe
  - Ancient DNA facility

- Advanced user support (even without previous NGS experience):
  - ✓ Project design
  - ✓ Choice of sequencing technology and methodology
  - ✓ Sample requirements and experimental design
  - ✓ DNA and RNA extraction for reference genome sequencing
  - ✓ Ancient DNA extraction and library construction
  - ✓ Sequencing
  - ✓ Primary data analysis



Brief history of Genomics and NGS

It all begun in late 19th century…

Johannes Friedrich Miescher
(13 August 1844 – 26 August 1895)

Die Spermatozoen einiger Wirbelthiere.

Ein Beitrag zur Histochemie*)

von

F. Miescher.

Hiezu Tafel I.

Bekanntlich wird in Basel der Fang des Rheinlachses (Salmo Salar) ziemlich lebhaft betrieben. Während der Laichzeit, im November, kann man zuweilen diese stattlichen Fische in grosser Zahl auf dem Markte sehen. Die reifen Geschlechtsprodukte dieser Thiere sind dabei als Abfall in beträchtlicher Menge zu erhalten. Die grosse Anstalt für künstliche Fischzucht in dem benachbarten Hüningen bezieht ihren ganzen Bedarf an Lachseiern, im Betrage von mehreren Millionen jährlich, von Herrn Friedrich Glaser, dem Besitzer der bedeutendsten hiesigen Fischhandlung.

Besonders verlockend ist hier für den Physiologen die Gelegenheit zur Gewinnung von Sperma. Von der rahmigen Flüssigkeit, die man als „Lachsmilch" bezeichnet, habe ich zuweilen mit Erlaubniss der Verkäufer fast einen Schoppen auf einmal
als blendend weisse Crême
lebenden Fisch; bei todten
Galle, Harn oder Blut mit

Der Samen der Fisch
vor andern werthvoll. Keine
ihre Produkte dem Sekret

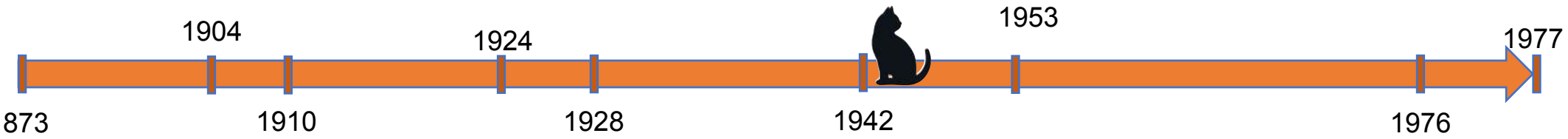
*) Nach Vorträgen, gehalte
ber 1873.

brechung der Spermatozoenköpfe erkennbar sind, ist auch das Protamin nachzuweisen.

Das Nuclein.

Der Rückstand nach Extraction mit Salzsäure zeigt unter dem Mikroskop noch Hülle und Inhalt und gibt die Millon'sche Reaction. In Kochsalzlösung quillt er nicht mehr, dagegen etwas in destillirtem Wasser.

From Miescher to Sanger



Miescher
Nuclein discovery
1873

Kossel
Nobel Prize
for discovery of
A, T and C
1910

(G discovered by von Liebig, in 1844)

Sutton and Boveri
Chromosome
theory of heredity
1904

Koltzoff
Prediction of a
”giant hereditary molecule
of 2 mirror strands”
1924

Griffith
DNA carried
genetic information
(theory)
1928

Avery
DNA carried
genetic information
(experimental)
1942

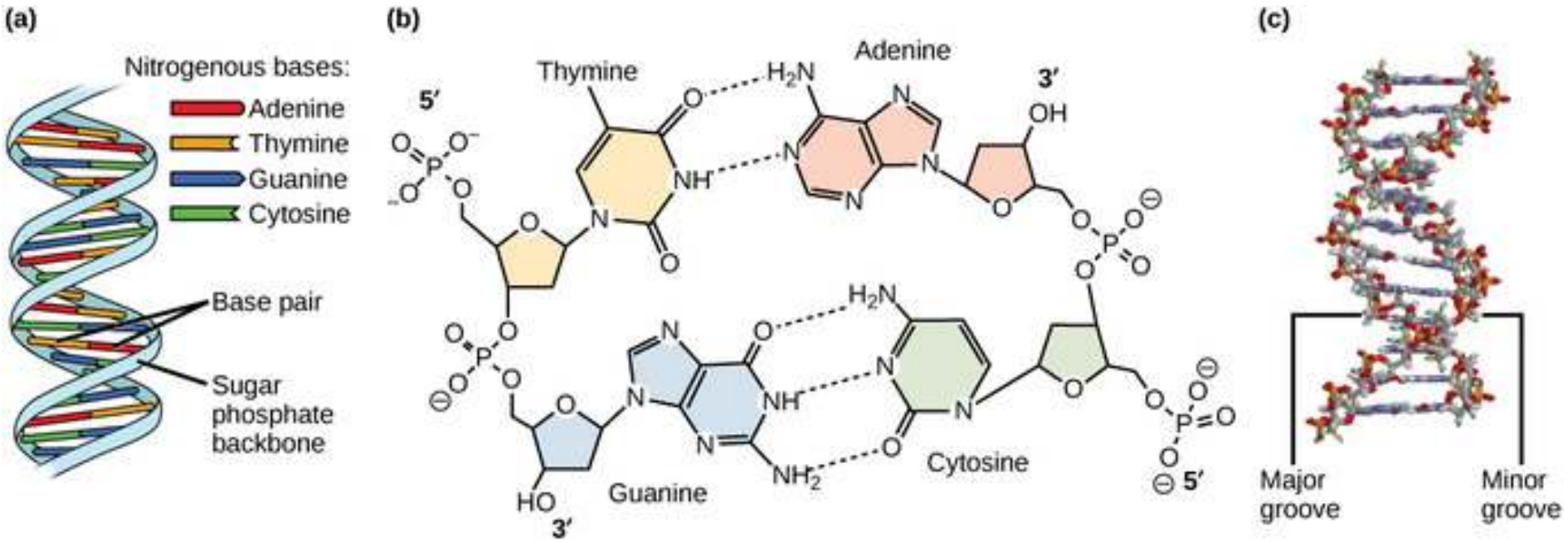
Watson, Crick, Franklin
Double helix DNA model
1953

Fiers *et al.*
Complete nucleotide sequence
of bacteriophage MS2 RNA
(3569 nucleotides)
1976

Sanger and Coulson
Nobel Prize for
Chain Termination Sequencing
1977

1944, **Schrödinger**:
Heredity must be carried by a solid, large, stable molecular structure - aperiodic crystal

What is sequencing?



https://figures.boundless-cdn.com

Sequencing is the process of arranging items, steps, or elements in a specific, logical order

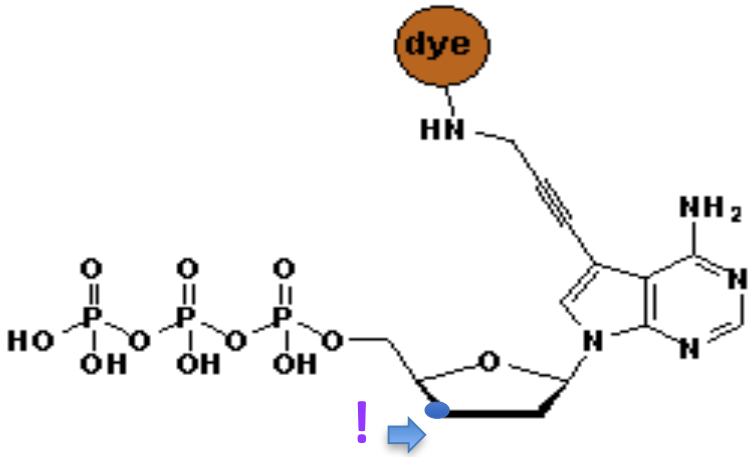
Once upon a time…

- Fredrik Sanger and Alan Coulson

  Chain Termination Sequencing (1977)

  Nobel prize 1980

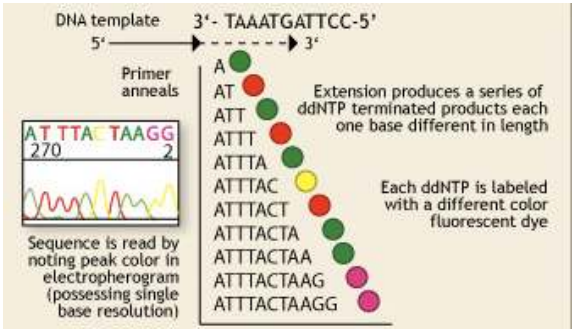
  Principle:



Lack of OH-group at 3’ position of deoxyribose

SYNTHESIS of DNA is randomly **TERMINATED** at different points

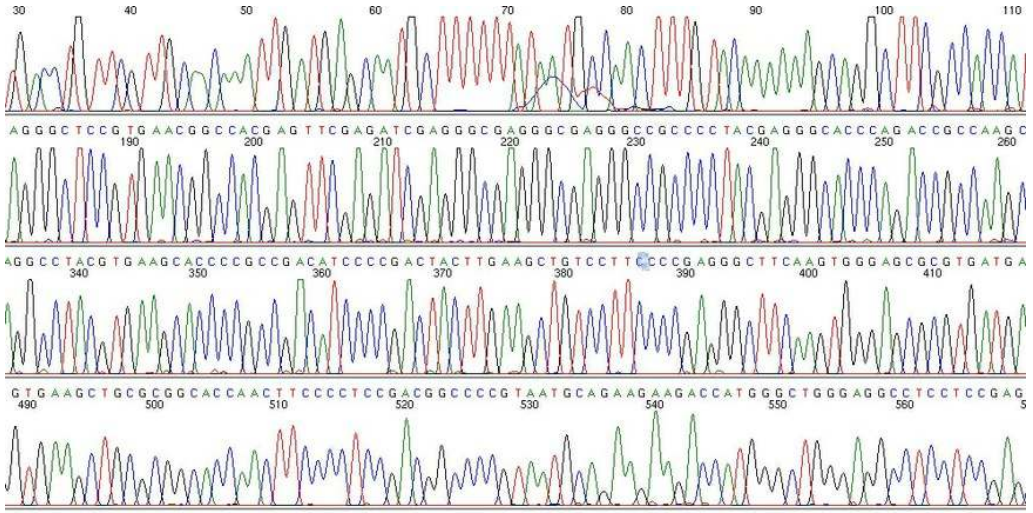
Separation of fragments that are 1 nucleotide different in size

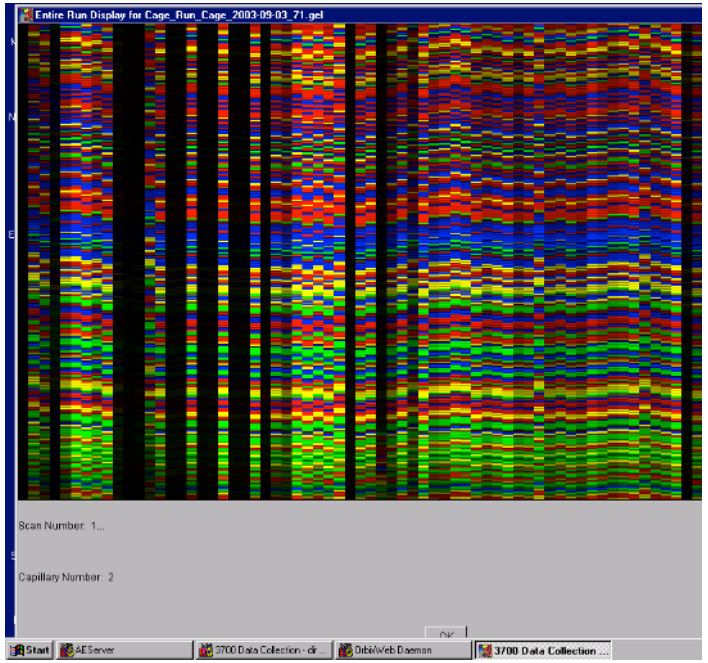


1 molecule sequenced at a time = 1 read

Capillary sequencer: 384 reads per run

Sanger sequencing



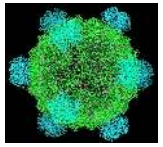



```
>eugene3.02190008
ATGTGGGGAATTGTTTTTTACCACCAACCAATTCTTGCAGCTACTATCATGTCAGATCAAAAGAACAAAA
CACCCCCTCCCAATGTAGAACTTAAATTCCAAATGCAAGCCATGACGAAGATGATGGAAAGAATGAATTC
CGTGATGGGGAATGTGTGTGACAGACTTGAGAAAGTGGGAAAACAAGGTAATGTCAGAACATGTACCCAA
GACGTGAGAAAGGTTGGGGCTGAACCAAAATCAAACAATGGCAGAGGGGCTGAAAGGCCAAGGTGGGCTG
ATTATGCGGATTTTGAGGTGGACGTTGATGATATTGTTGATGGTGGTTTTAAGGATGAGACCATAGGCCA
TCAAAAAGGTTTTCAACACCATAGAAACCGAAGGGATTTCATGTATTTTGACGGGGTGTTATGGCAAAAG
AAAATGAGGATTCAAAAGGAGAGGTGTCAAAGGGAGAGAAATAAAGAGATTGGTGTTCTAAAAAATGAAT
CCAAGAGTCTATACCGTATTCTAGGGGAGATGAAGCAAGAACTTGATGTGTTAATGGCAATAGTCAATGC
CCAACAAGCTTCAGAAAGAGAGGAGAAAATACGCTGCAATGATGATATACGAAATAGAATGGATGCTACT
TTCATCAAAAGTTGGTGGCGACGATTGTTGGCAAGAAAAGAACTCCAAAGGCTTCAAAAAGAGGCTAAGg
aatttggtccttaa
```

Chromatogram file size: c:a 250 kb
FASTA file size: 12 kb

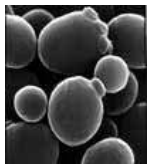
Prerequisite: amplified DNA

At the very beginning of genome sequencing era…

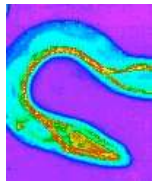


- First DNA genome: virus ϕ X 174 - 5 368 bp (1977)

- First organism: *Haemophilus influenzae* - 1.5 Mb (1995)



- First eukaryote: *Saccharomyces cerevisiae* - 12.4 Mb (1996)



- First multicellular organism: *Cenorhabditis elegans* - 100 MB (1998)

- First plant: *Arabidopsis thaliana* - 157 Mb (2000)

The Yeast Genome project

Life with 6000 Genes

A. GOFFEAU, B. G. BARRELL, H. BUSSEY, R. W. DAVIS, B. DUJON, H. FELDMANN, F. GALIBERT, J. D. HOHEISEL, C. JACQ, [...], AND S. G. OLIVER

& Affiliations

SCIENCE · 25 Oct 1996 · Vol 274, Issue 5287 · pp. 546-567 · DOI: 10.1126/science.274.5287.546

"The genome of the yeast Saccharomyces cerevisiae has been completely sequenced through ***an international effort involving some 600 scientists in Europe, North America, and Japan.*** *It is the largest genome to be completely sequenced so far (a record that we hope will soon be bettered) and is the first complete genome sequence of a eukaryote."*

"New graduate students are already wondering how we all managed in the "dark ages" before the sequence was completed. We must now tackle a much larger challenge, that of elucidating the function of all of the novel genes revealed by that sequence. ***As with the sequencing project itself, functional analysis will require a worldwide effort.*** *In Europe, a new research network called EUROFAN [for European Functional Analysis Network has been established to undertake the systematic analysis of the function of novel yeast genes. Parallel activities are underway in Germany, Canada, and Japan. In the United States, the National Institutes of Health has recently sent out a request for applications for "Large-Scale Functional Analysis of the Yeast Genome." "*

GENETICS

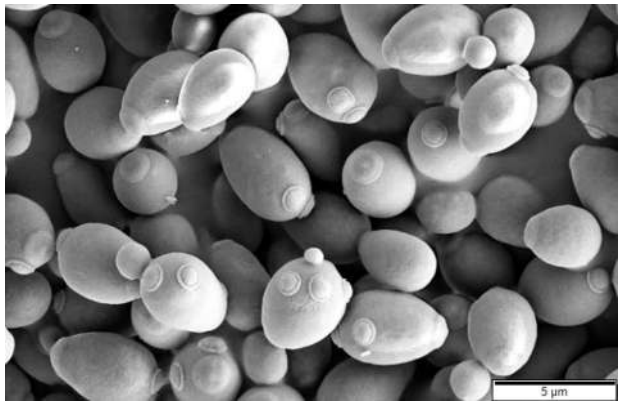
Genetics. 2013 Jun; 194(2): 291–299.
doi: 10.1534/genetics.113.151258

PMCID: PMC3664842
PMID: 23733847

The Modest Beginnings of One Genome Project

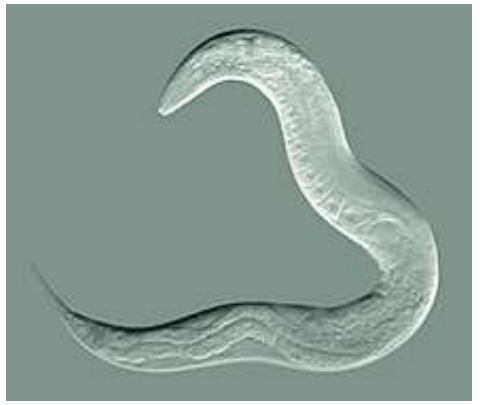
David B. Kaback†

First genomic references

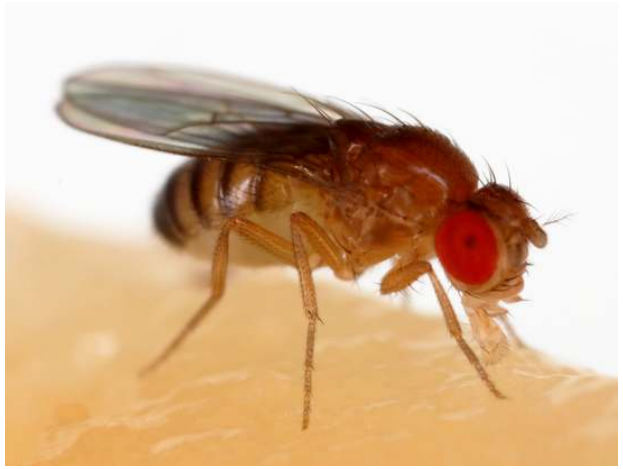


1/3 of genes related to human by homology

Basic cell functions



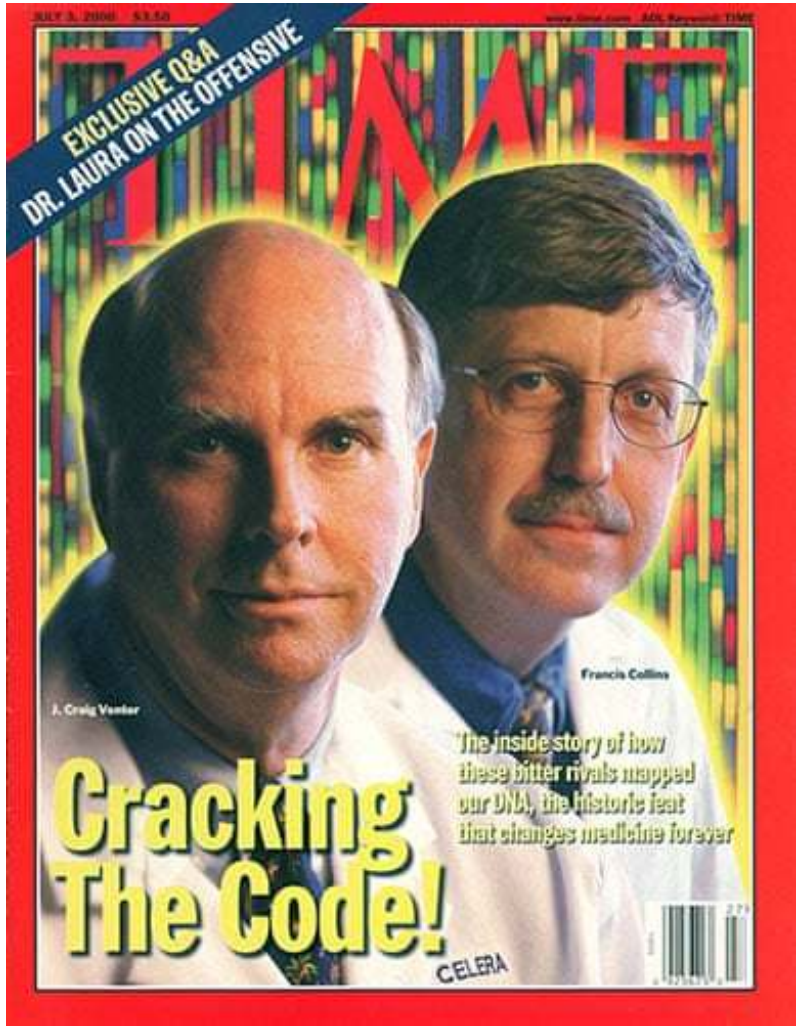
Human disease gene discovery



Phenotypical traits

Plant genome
structure and function

But it was not enough…



GENOME SEQUENCING WORKSHOP

MARCH 3 & 4, 1986

SANTA FE, NEW MEXICO

SPONSOR

DOE

OFFICE OF HEALTH AND

ENVIRONMENTAL RESEARCH

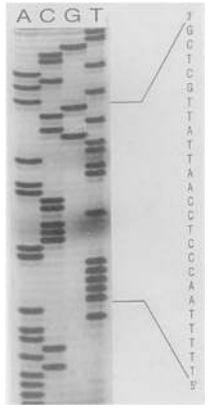
HOST

LIFE SCIENCES DIVISION

LOS ALAMOS NATIONAL LABORATORY

It is thus important that we identify here what real benefits and liabilities might emerge from the contemplated sequencing activity, which would aim at capturing the entire human genome in a period of 10 or 12 years. Do we have the technologies necessary to do this, and do we have the computational power and algorithms needed to integrate and anaylze this data? Will this information provide both clinical and basic benefits of such magnitude to warrant an accelerated effort?

An outcome of HUGO – Genomic Revolution



1977

Sanger

P32 labeled gels

1990 HUGO

2000

Sanger

ABI 3000 series

NGS

2005

454

Solexa
ABI SOLiD

2006

PacBio
ONT

2012

2010

IonTorrent

2021 HiFi reads

MGI 华大智造 2023

PacBio
REVIO

2026 WE ARE HERE

Element Biosciences

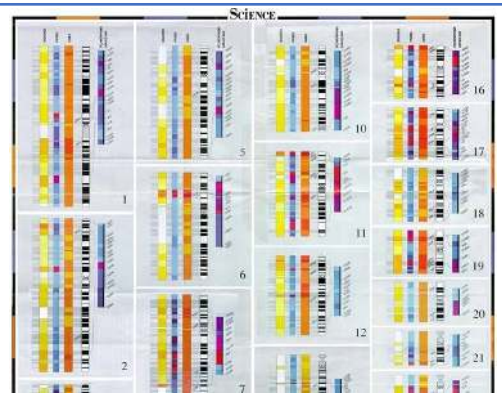
Roche

ULTIMA GENOMICS

Just a comparison

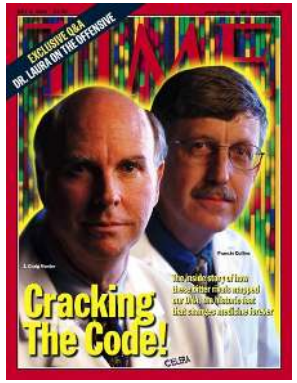
1990 - 2003
HUGO

Sanger traditional
$2.7 bln



2007
Craig Venter's genome

Sanger ABI 3730
$300 mln



2008
Jim Watson's genome

454 FLX
$1 mln

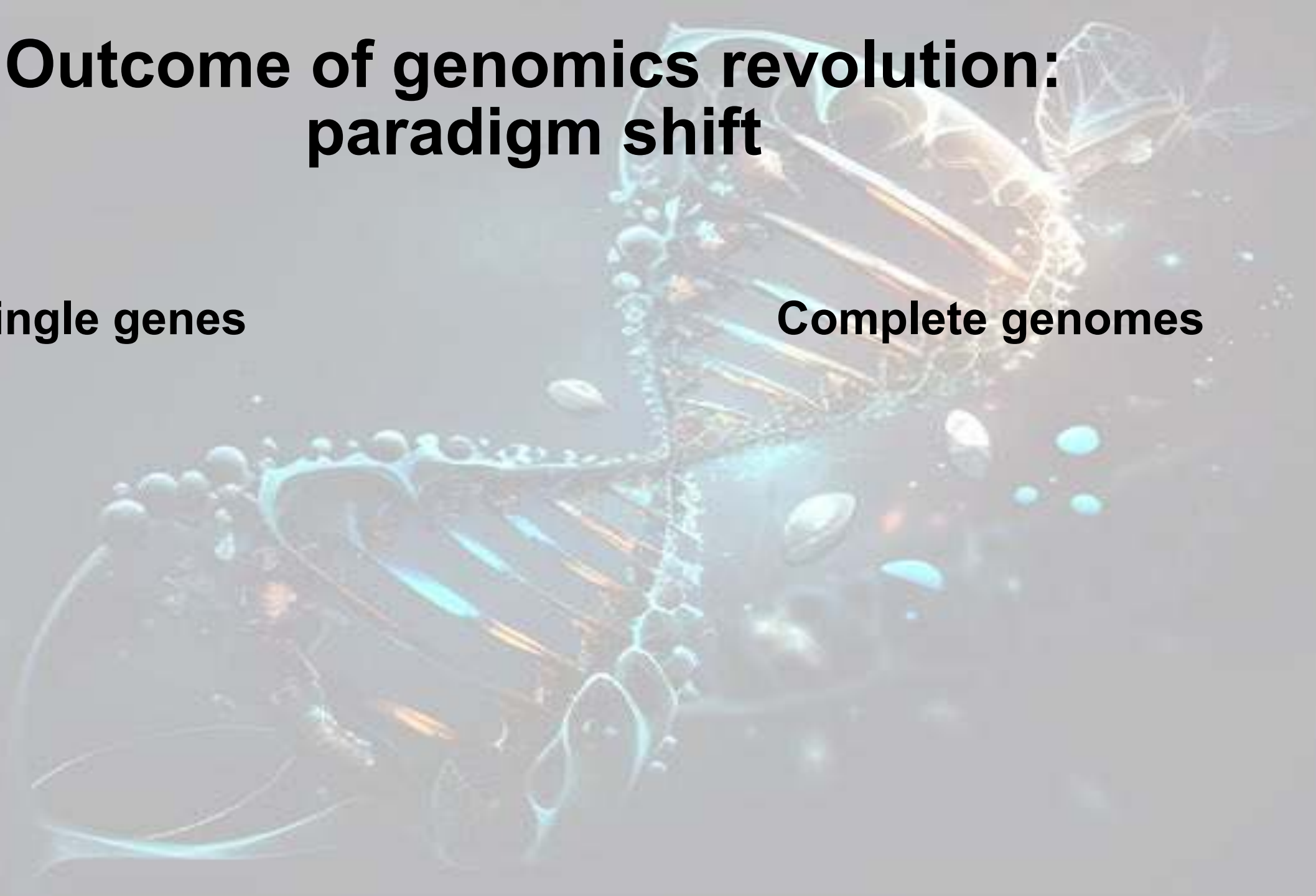
TODAY
any human

$200-600 with short reads
$1-3k with long reads

Outcome of genomics revolution: paradigm shift

Single genes

Complete genomes



Outcome of genomics revolution: paradigm shift

Single genes

Single transcripts

Complete genomes

Whole transcriptomes

Outcome of genomics revolution: paradigm shift

Single genes	Complete genomes
Single transcripts	Whole transcriptomes
Single organisms	**Metagenomes**

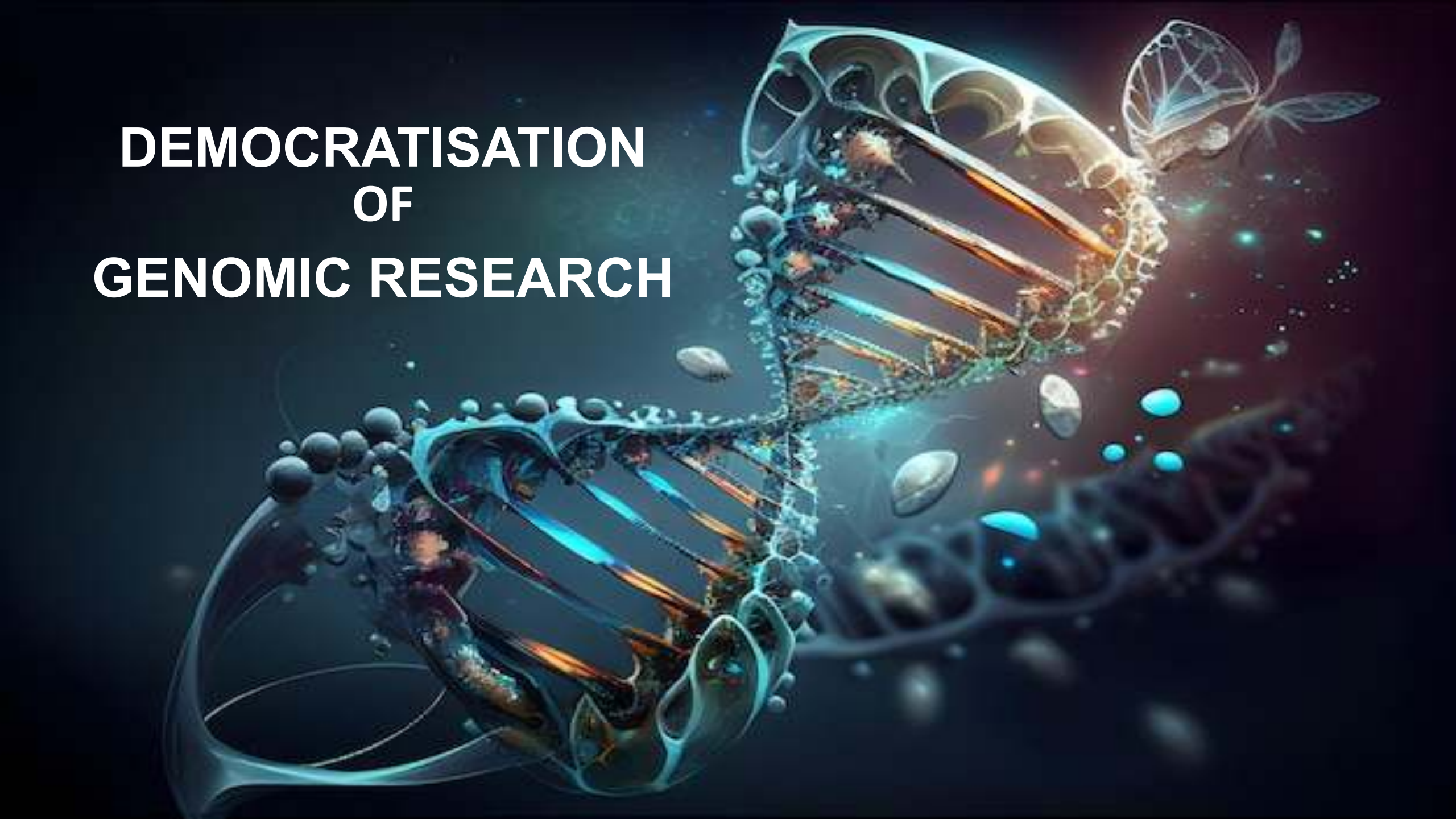
Outcome of genomics revolution: paradigm shift

Single genes	Complete genomes
Single transcripts	Whole transcriptomes
Single organisms	Metagenomes
Model organism	**Any species**

Outcome of genomics revolution: paradigm shift

Single genes	Complete genomes
Single transcripts	Whole transcriptomes
Single organisms	Metagenomes
Model organism	Any species
Available to highly specialized labs	**Available to anyone**

DEMOCRATISATION OF GENOMIC RESEARCH



Sequencing technologies: theory and principles

NGS market overview / Read length

Oxford NANOPORE Technologies

Element Biosciences

ULTIMA GENOMICS

MGI 华大智造

illumina

PACBIO®

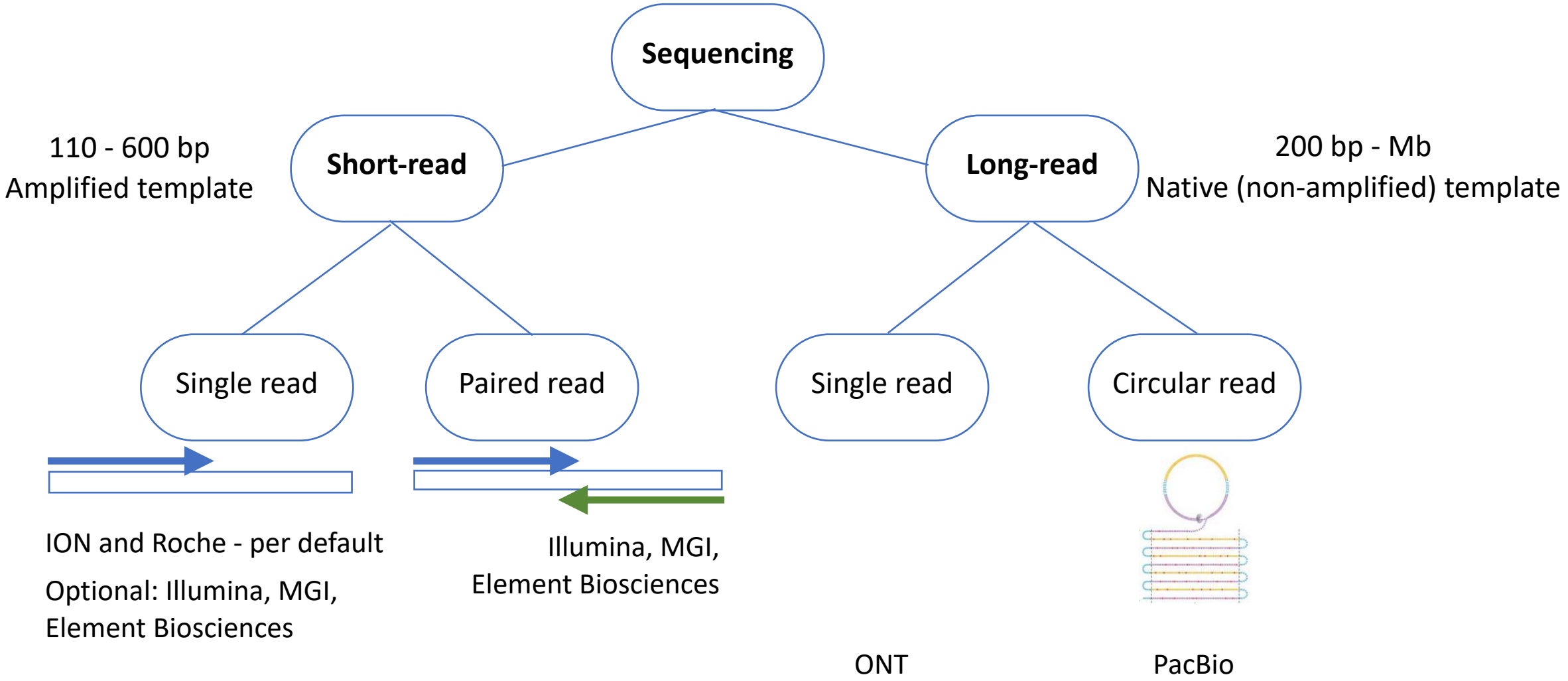
iontorrent by Thermo Fisher Scientific

Roche

1 2 3 4 5 6 7

Before going into details, let's understand the main principles.

Types of sequencing by **read length**



Paired reads (PE) = better mappability

Any NGS starts with a library

Template (aka “insert”)

Adaptor ligation

Tells the reaction where to start

OR:

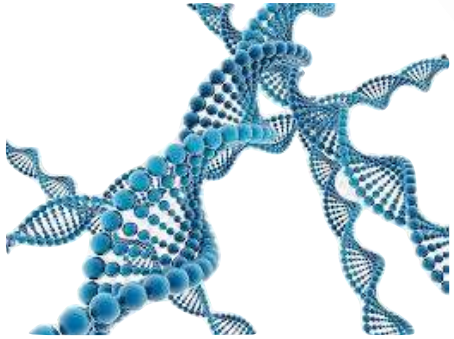
Barcode and adaptor ligation

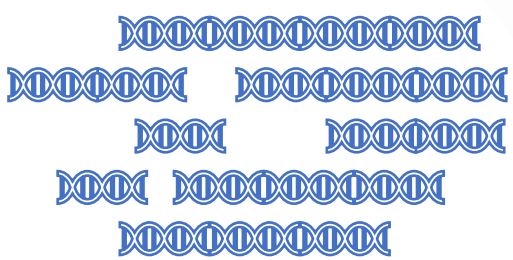
Barcodes different samples for pooling

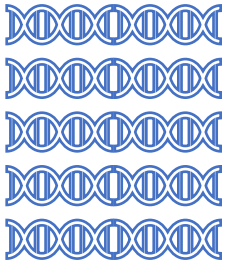
Type of adaptor, primer and insert size are technology- and application-specific

For **all** short-read technologies: libraries have to be amplified

NGS library construction procedure



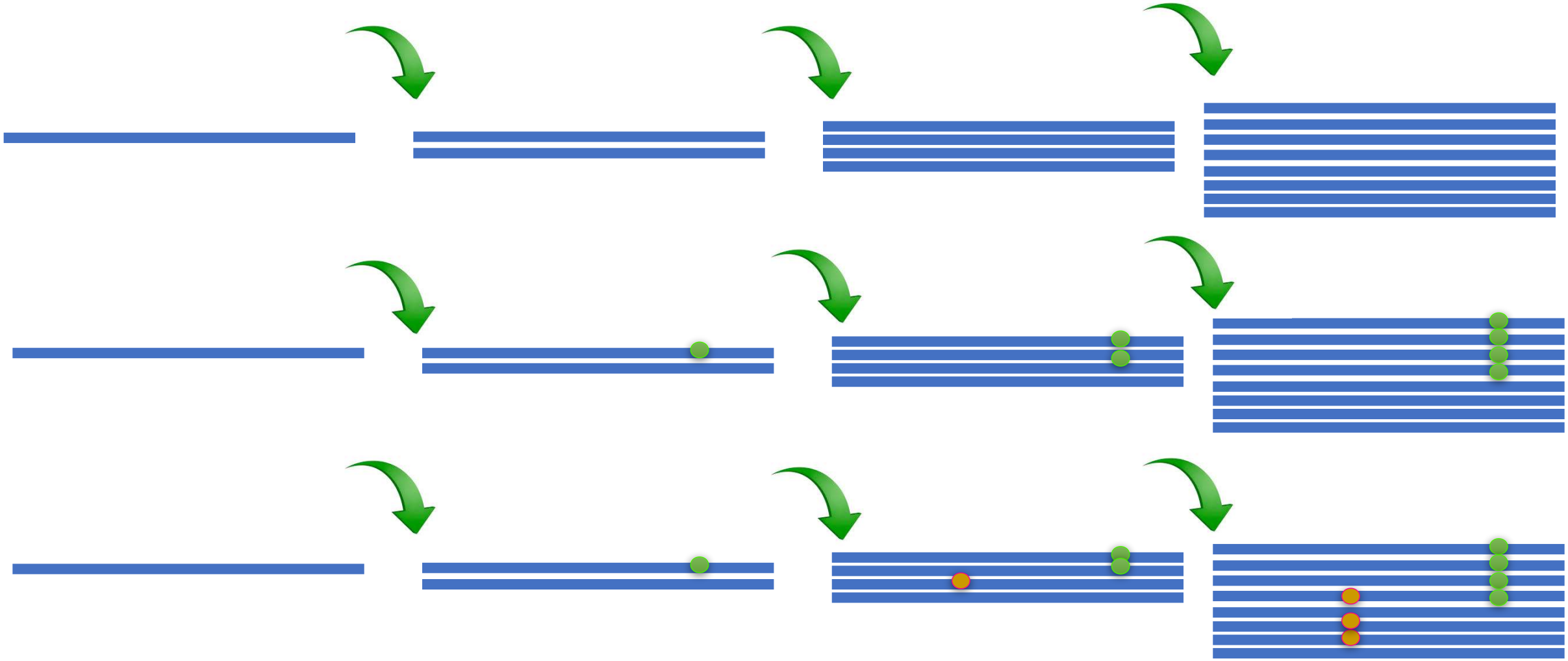




Library construction steps ->	**DNA fragmentation**	**Size selection**	**End repair and adaptor ligation**
Quality control steps ->	Sizing Quantification	Sizing Quantification	Sizing Quantification Purification

Up to 40% of material is lost in each step

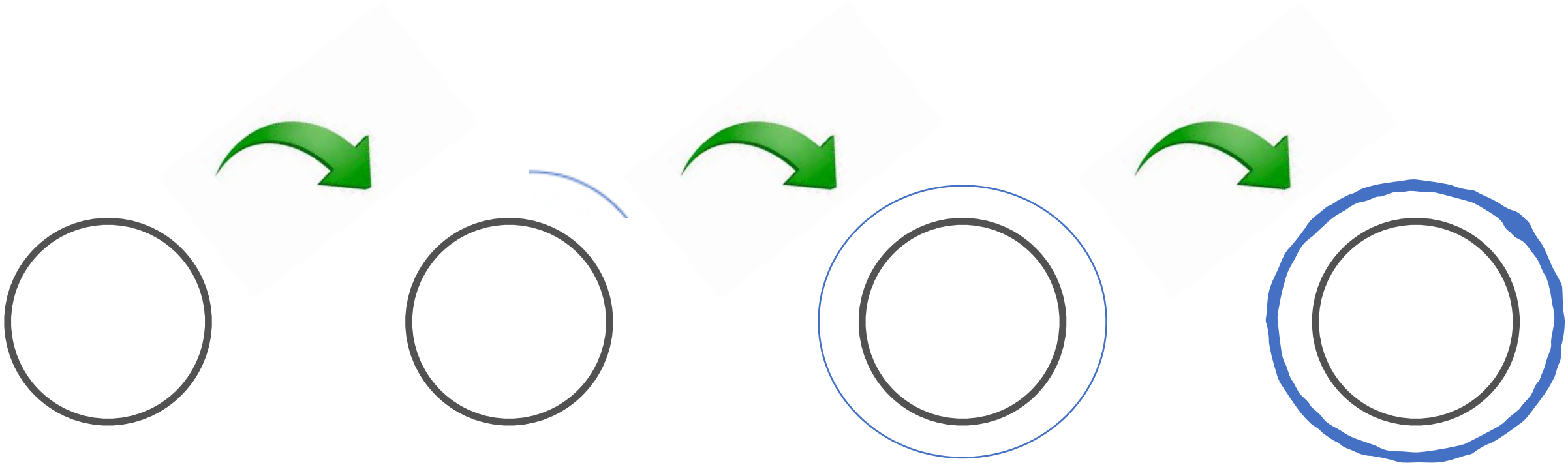
Amplification: PCR (Copy of a copy)



Sanger, ION, Illumina

Is it a SNP or a PCR artifact?

Amplification: Rolling circle (RCA)
(Copy of the same molecule)



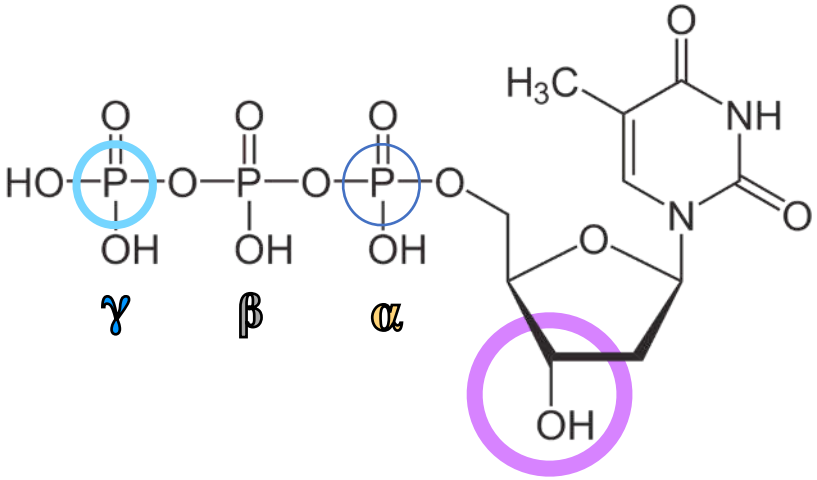
Element Biosciences, MGI

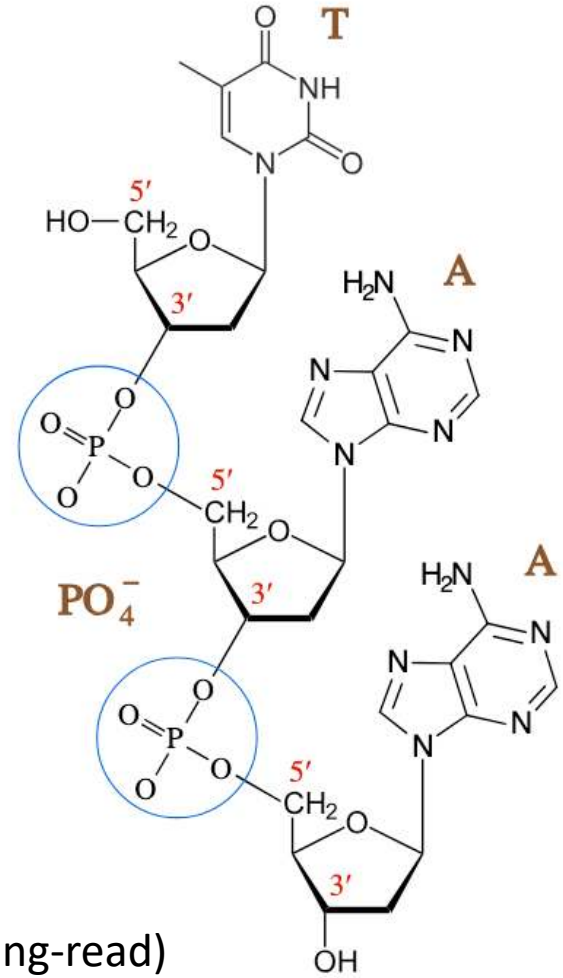
Reduced amplification bias

DNA synthesis - Back to school!

SBS-sequencing by synthesis

Sequencing fuel:
Deoxyribonucleoside triphosphates (dNTP)





DNA polymerization:

Pyrophosphate is released
(and with that - **color signal**)

O of 3’-OH stays put
H is released as a proton, **H+**

Phosphate groups

α Gets attacked to form a bond
γ Where “dye” is attached

3’ position of deoxyribose

Present = reaction happens (long-read)
Blocked = reaction stops (most short-read)

Nucleotide incorporation

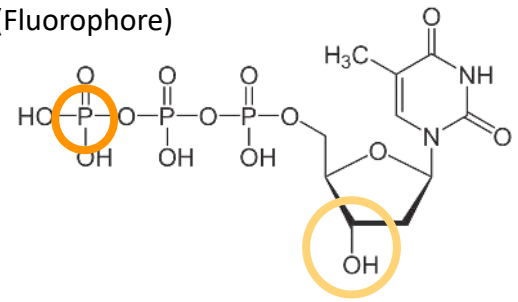
Short reads: 1 at a time (cycle)
Long reads: constant flow

Mode of sequencing: SBS and SBB

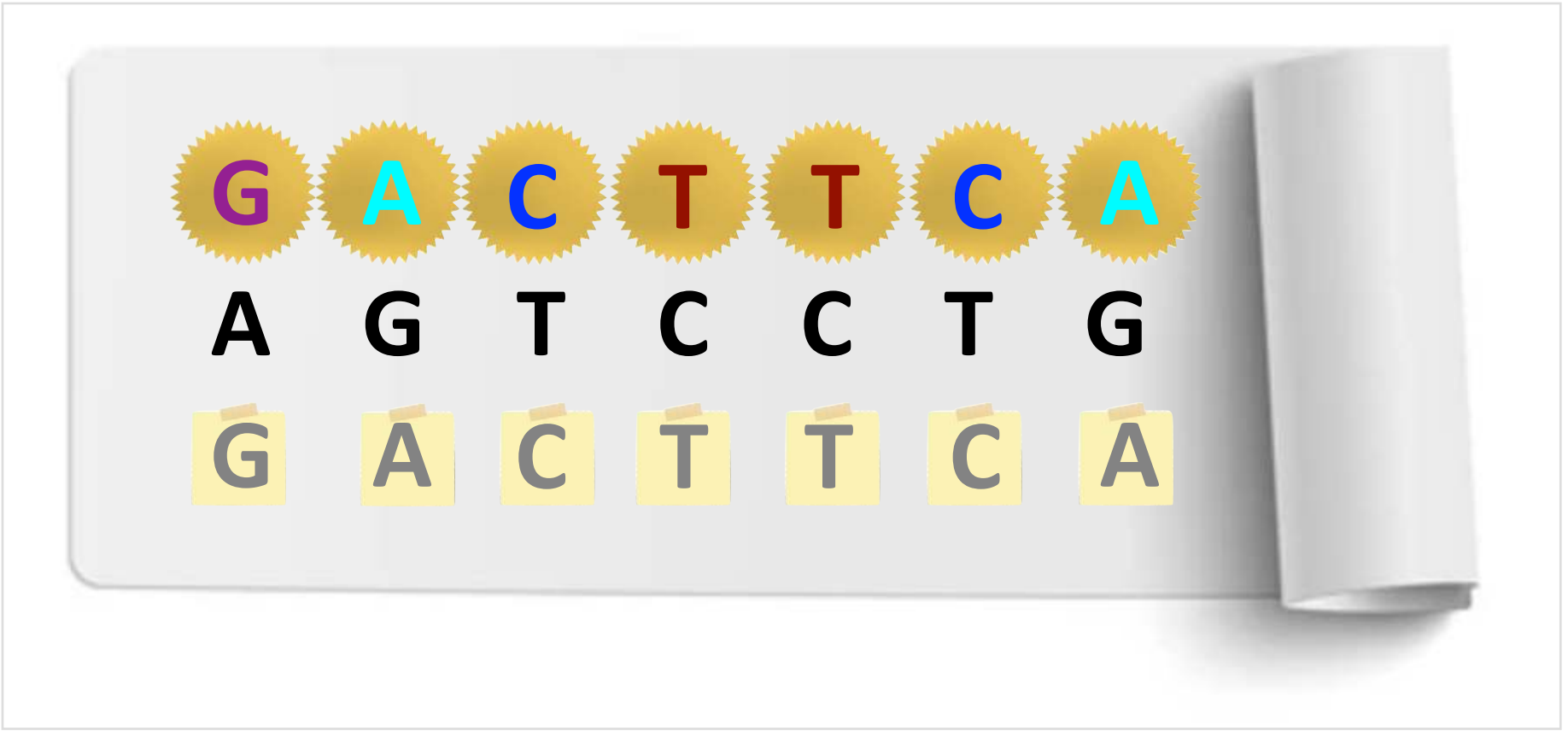
Sequencing by Synthesis

dNTP* is **BOUND**

(Fluorophore)



(Block)



Sequencing by Binding

dNTP* presented, removed, and then a native nucleotide is bound

SBB is gentler than SBS

SBS and SBB comparison

Traits:	SBS	SBB
Core differences	Detection = incorporation Modified nucleotides Errors locked in	Detection -> incorporation Natural nucleotides Errors can reset
Quality drop with cycle	Steep	Shallow
Read accuracy	Lower	Higher
End-of-read noise	Common	Reduced
Phasing errors	Common	Rare

SBS is prone to phasing errors

Q-score

Q (quality) score - measurement of machine confidentiality that the base (A, G, T, C) was correctly identified

Q-Score	Accuracy	Error Rate	Probability of a Wrong Call
Q10	90%	1 in 10	10% chance of error
Q20	99%	1 in 100	1% chance of error
Q30	99.9%	1 in 1,000	0.1% chance of error
Q40	99.99%	1 in 10,000	0.01% chance of error
Q50	99.999%	1 in 100,000	0.001% chance of error

Why it matters? Example:

Looking for a rare mutation (SNP), appears 1 in 5000 DNA molecules
Q30 = 1 error per 1 000 bases -> the somatic mutation will be drowned in the noise of incorporated errors

To correctly call a 1:5000 rare SNP, one needs a technology with Q40 or higher

Types of sequencing errors

Indels: Insertions and deletions - machine sees something that is not there

Insertion
Actual: A T G C
Read: A T **T** G C

Deletion
Actual: A T G C
Read: A T _ C

Substitutions: incorporation of wrong bases

Actual: A T G C
Read: A T **A** C

Homopolymers: stretches of the same nucleotide

Actual: A A A A
Read1: A A A A
Read2: A A A
Read3: A A A A A

All technologies have their own error profiles

Short break and then:

Finally, the technologies…

illumina®

Current leader on the NGS market

SBS

Instrument	Run time	Max output	Max reads/run	Max read length, bp
iSeq	9.5 – 19 hrs	1.2 Gb	4 mln	PE 150
MiniSeq	4-24 hrs	7.5 Gb	25 mln	PE 150
MiSeq	4-55 hours	15 Gb	25 mln	PE 300
NextSeq series	12-48 hours	120-300 Gb	0.4 – 1 bln	PE 150
NovaSeq 6000	13-44 hours	6 Tb	26 bln	PE 250
NovaSeq X Plus	15-18 hrs	16 Tb	52 bln	PE 150

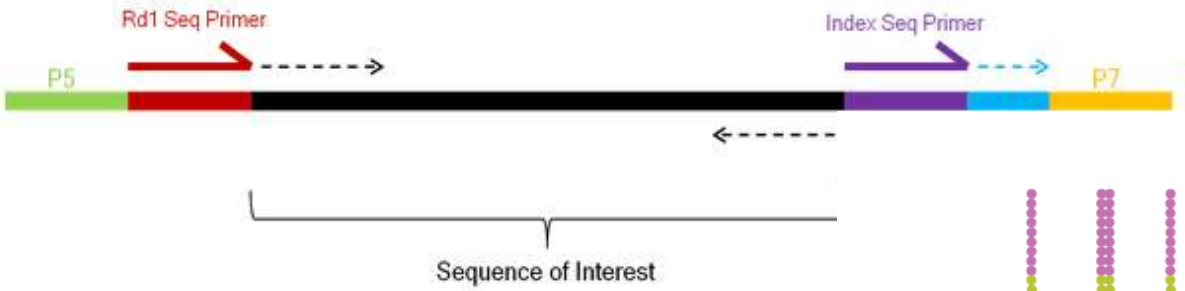
Technology highlight: bridge amplification

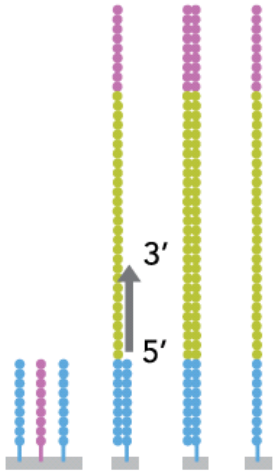
Used for: everything

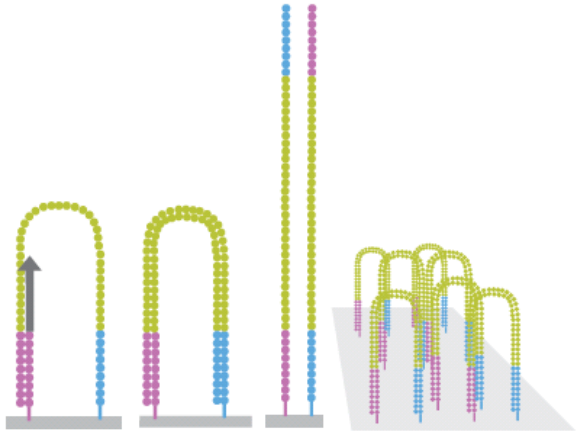
Strength: cheap; robust data analysis pipelines

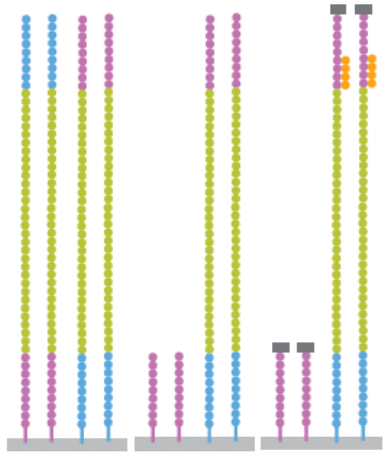
Weakness: more bias than people think, GC-bias

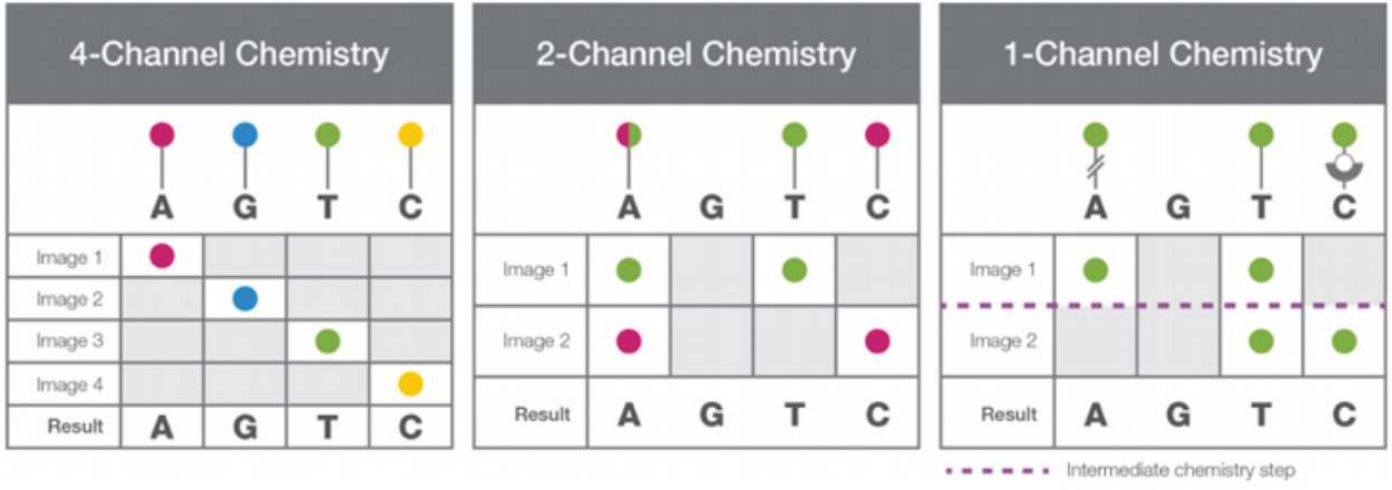
Illumina: bridge amplification



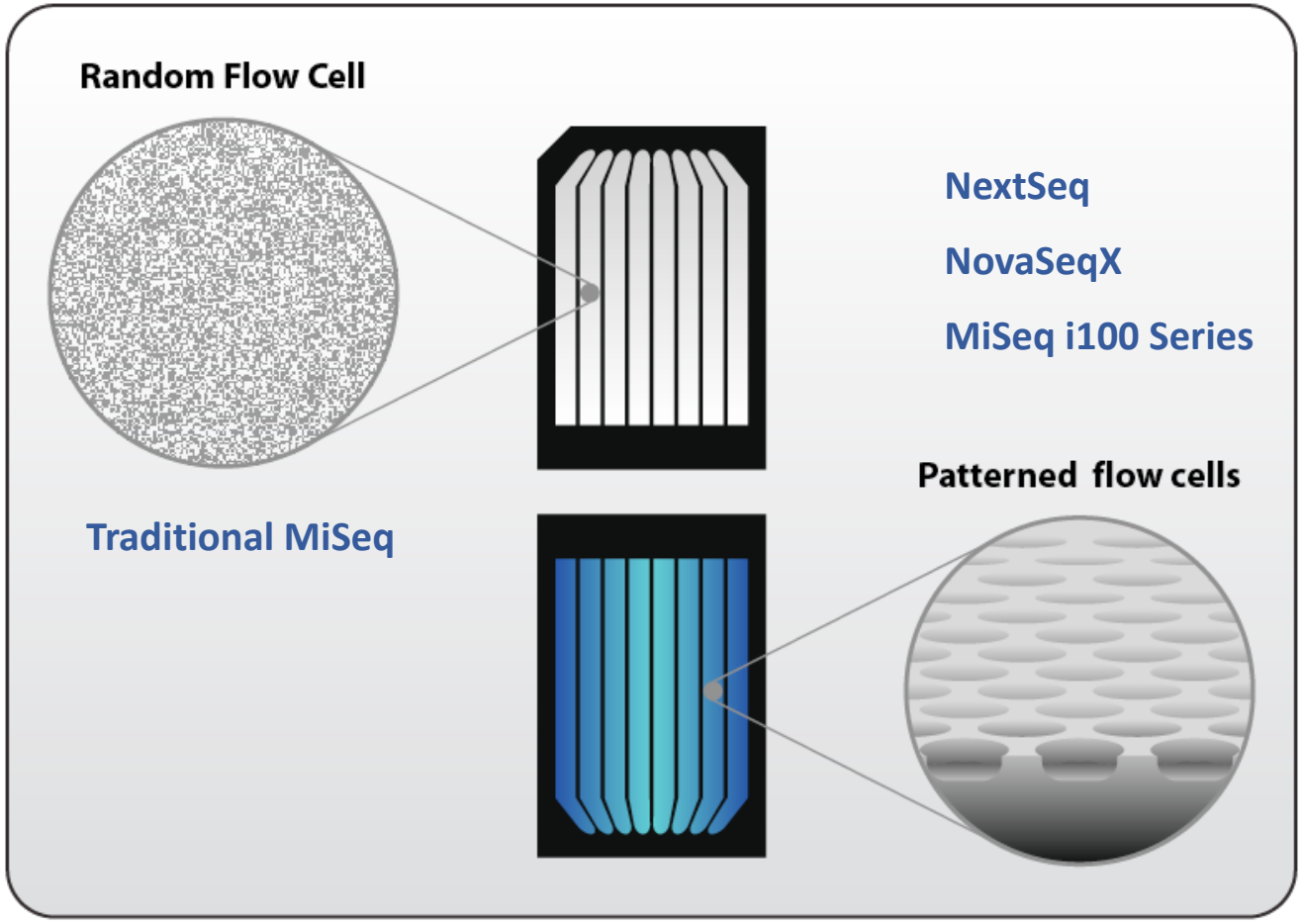


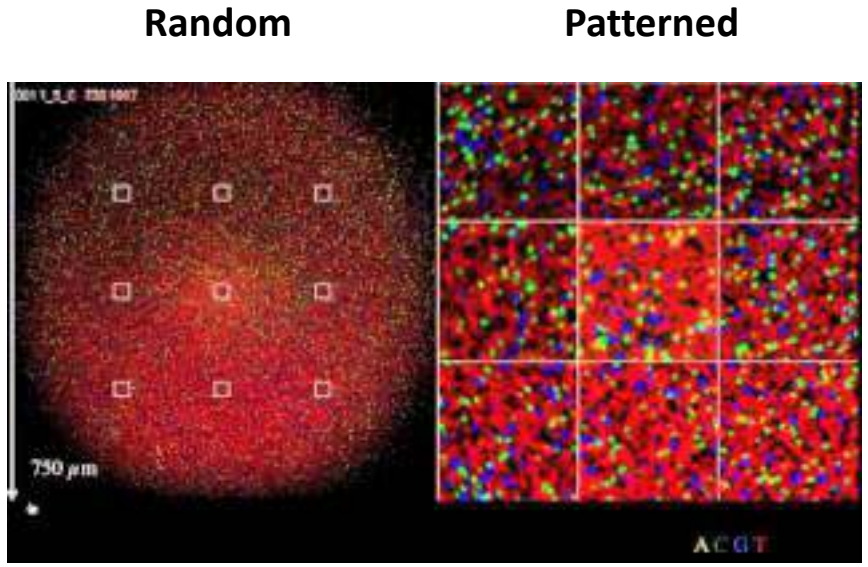






Illumina flow cells: important differences





1 flow cell per run

ION (ThermoFischer)

Instrument	Run time	Max output	Max reads/run	Max read length, bp
Gene Studio S5	3 - 24 hrs	50 Gb	150 mln	SE 600
Genexus	14-24 hrs	20 Gb	60 mln	SE 400

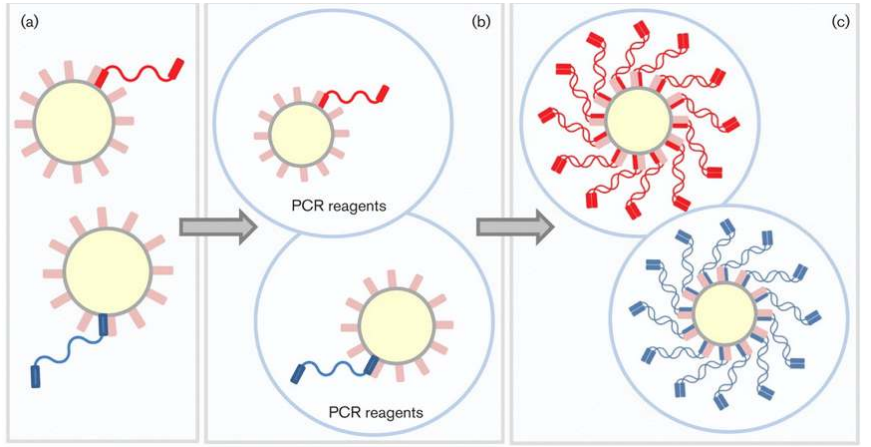
Technology highlight: H^+ ion-sensitive field effect transistors

Used for: gene panels, advantage in clinical setting

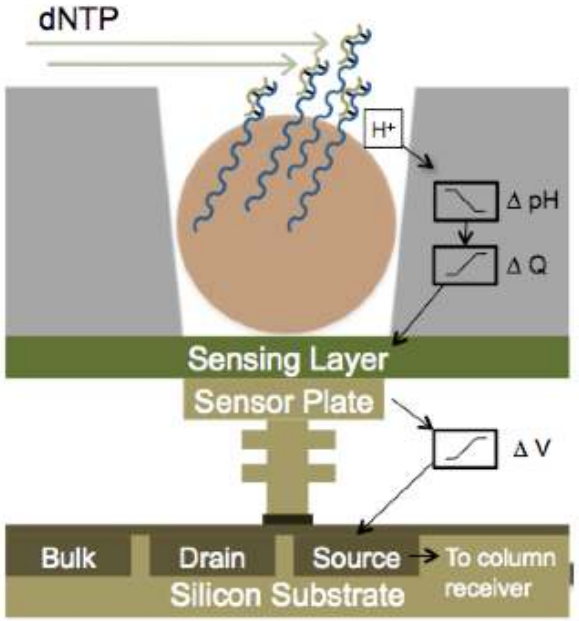
Strength: scalable, very fast turn-around, inbuilt analysis software

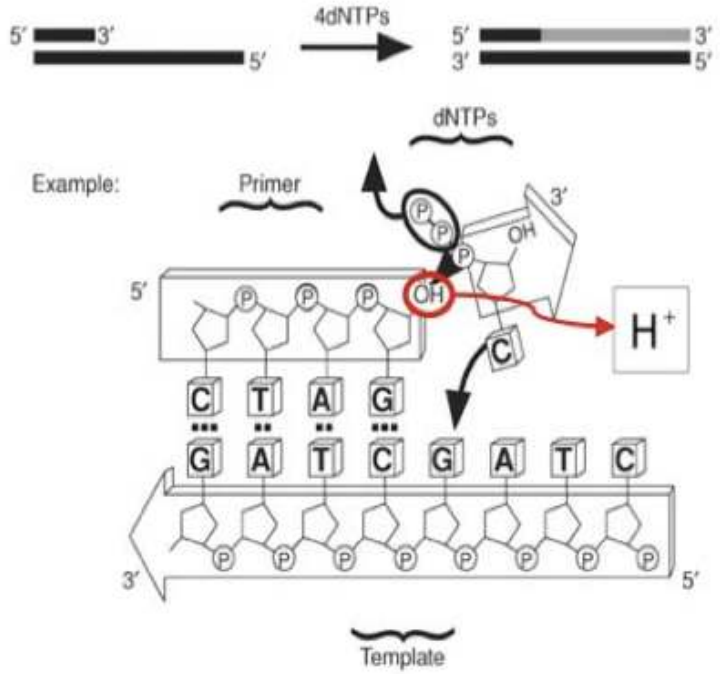
Weakness: limited number of applications, AT-bias, (homopolymers bias)

Ion Torrent: H^+ ion-sensitive field effect transistors



https://doi.org/10.1099/vir.0.043182-0





Ion technology: scalability

1 chip per run

Ultima

SBS

Instrument	Run time	Max output	Max reads/run	Max read length, bp
UG 100	12 - 14 hrs	2.4 Tb	6-8 bln	SE 300
UG 100 Solaris	14 hrs	3.6 Tb	10-12 bln	SE 300
UG Solaris Boost	24 hrs	unspecified*	100 bln/day	SE 300

* based on 100 billion reads/day

Technology highlight: no flow cell, lots of AI

Used for: potentially - everything, still new on the market

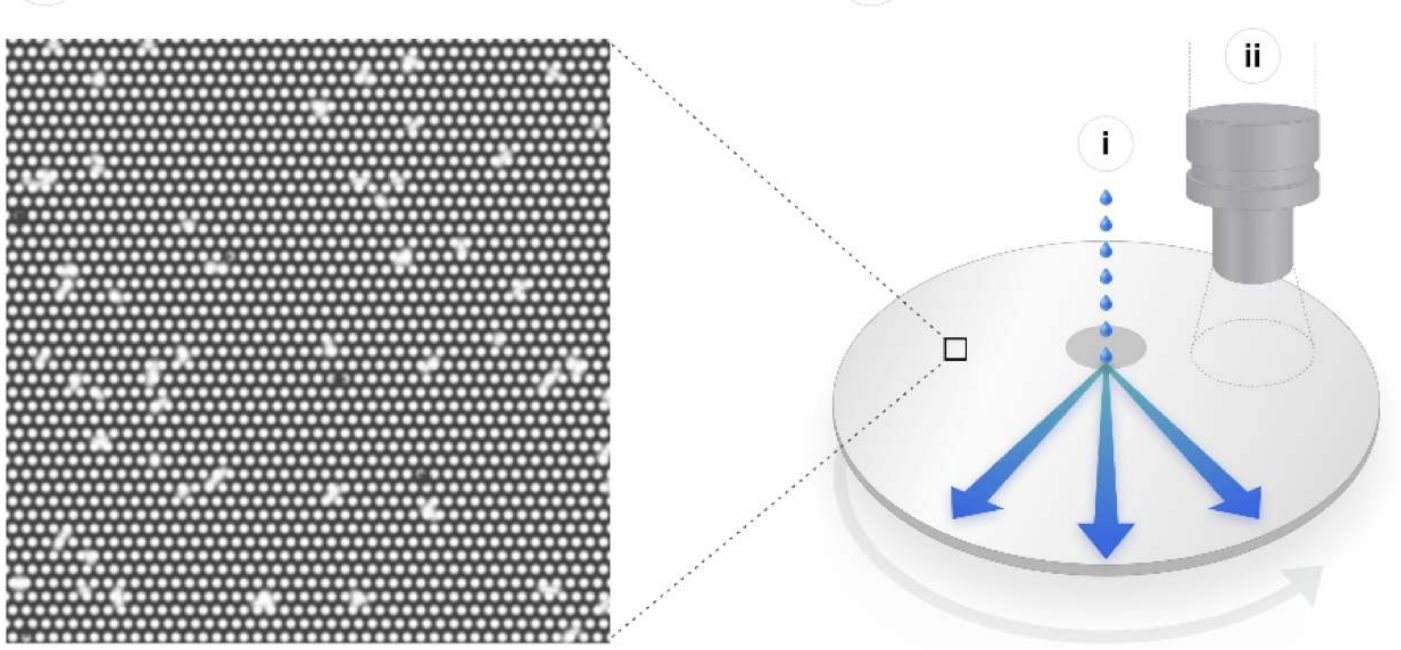
Strength: very cheap but accurate

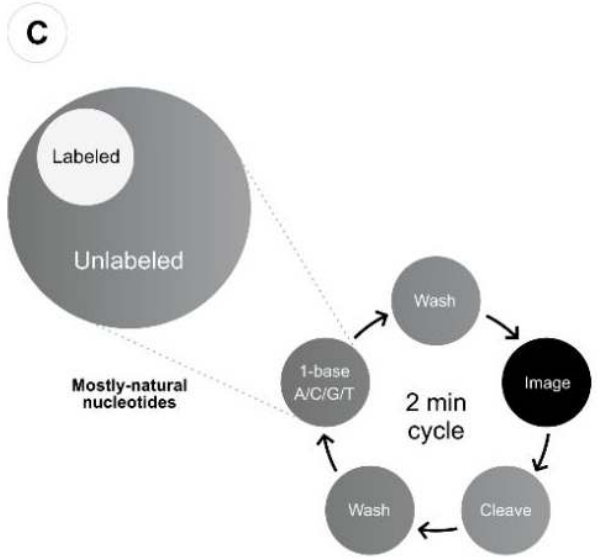
Weakness: Few analysis pipelines available, GC-bias, analysis bias

Ultima technology

No more flow cells, chips and SMRT cells - introduced wafers

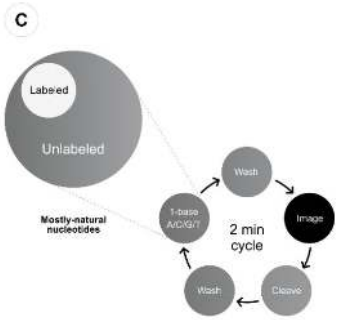
Library: emulsion PCR, similar to ION technology

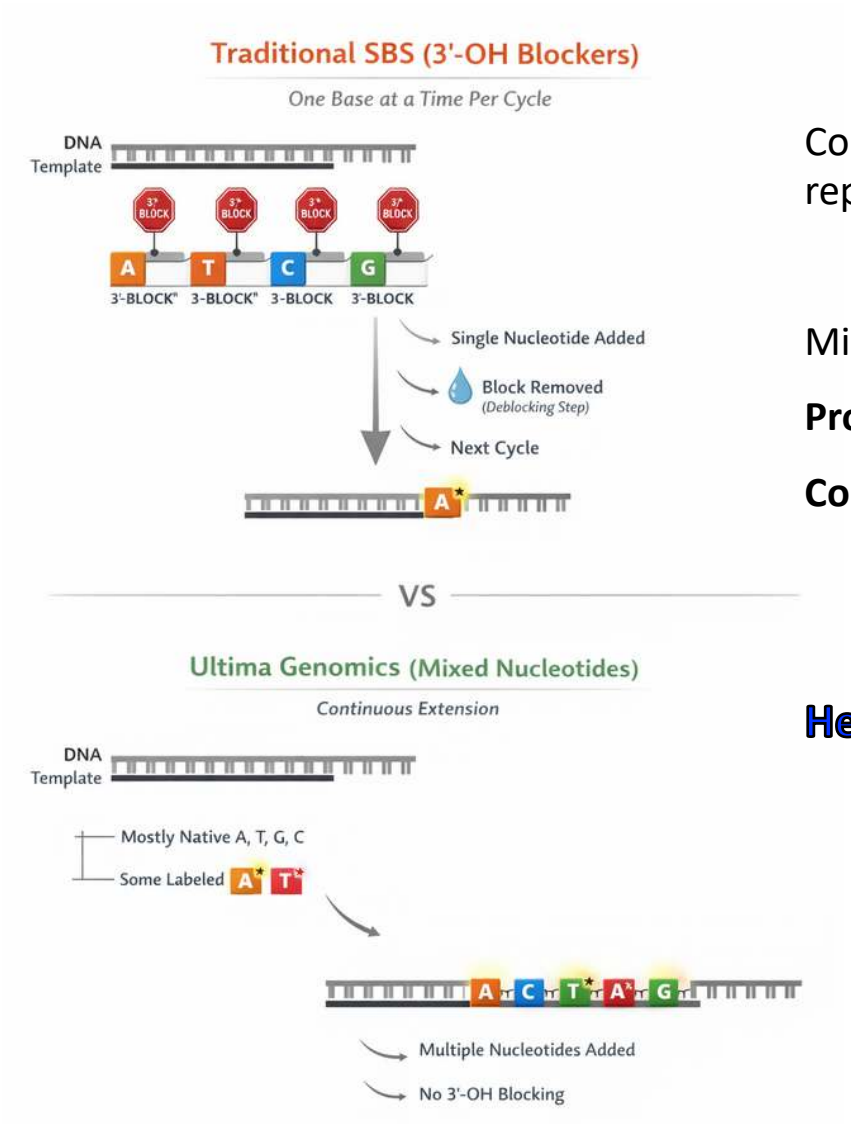




CD-manufacturing precision

Ultima technology





Continuous extension: **higher speed and throughput** (avoiding repetitive chemical de-blocking steps).

Mix of fluorescent and native dNTPs: **signal events are rare**

Pro: High speed - no need to resolve every single base individually.

Con: weaker fluorescent signal

Hence - machine learning for base calling

MGI

SBS

Instrument	Run time	Max output	Max reads/run	Max read length, bp
DNBSEQ-T7	5.5 - 24 hrs	7 Tb	23 bln	PE 150
DNBSEQ-G400	13-109 hrs	1.4 Tb	3.6 bln	PE 300
DNBSEQ-G99	11 hrs	0.24 Tb	0.2 mln	PE 150
DNBSEQG-50	9-40 hrs	150 Gb	500 mln	PE 150
DNBSEQ-T1+	24 hrs	1.5 Tb	12 bln	PE 150

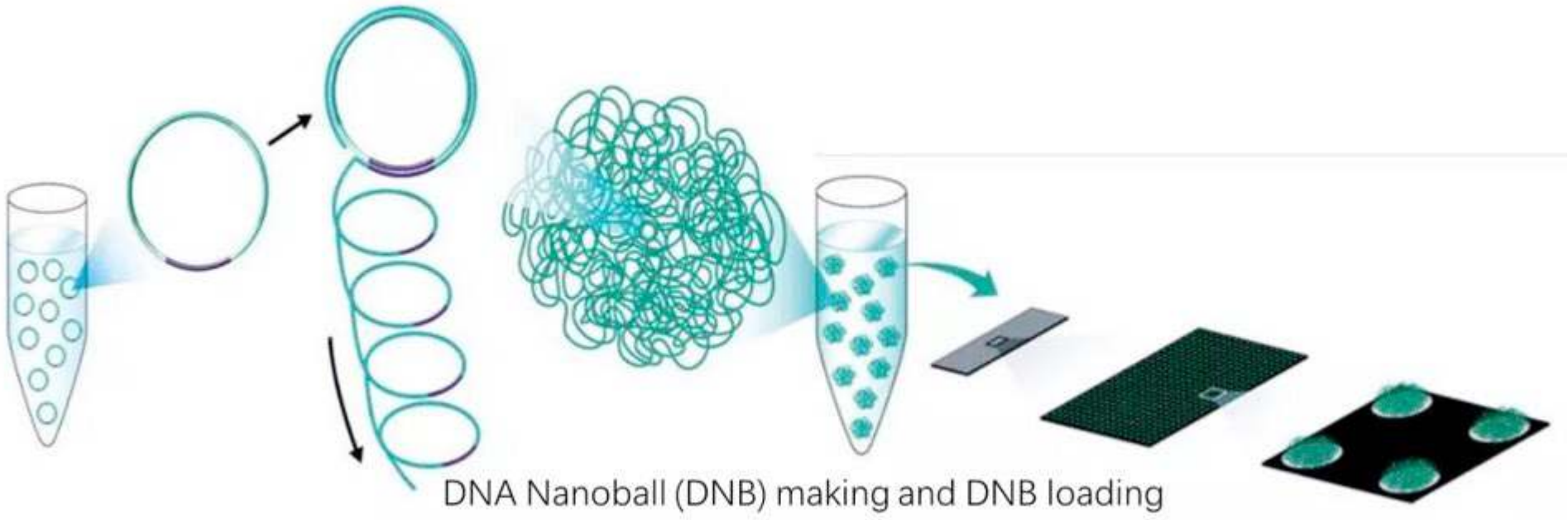
Technology highlight: DNA nanoballs (RCA)

Used for: high-throughput projects

Strength: high accuracy and throughput, coupled with automatization

Weakness: labour-intensive, GC bias, software limitations

MGI: DNA nanoball (DNB) technology



Crucial step: library circularization, followed by rolling-circle amplification.

Result: *negatively charged* nanoballs.

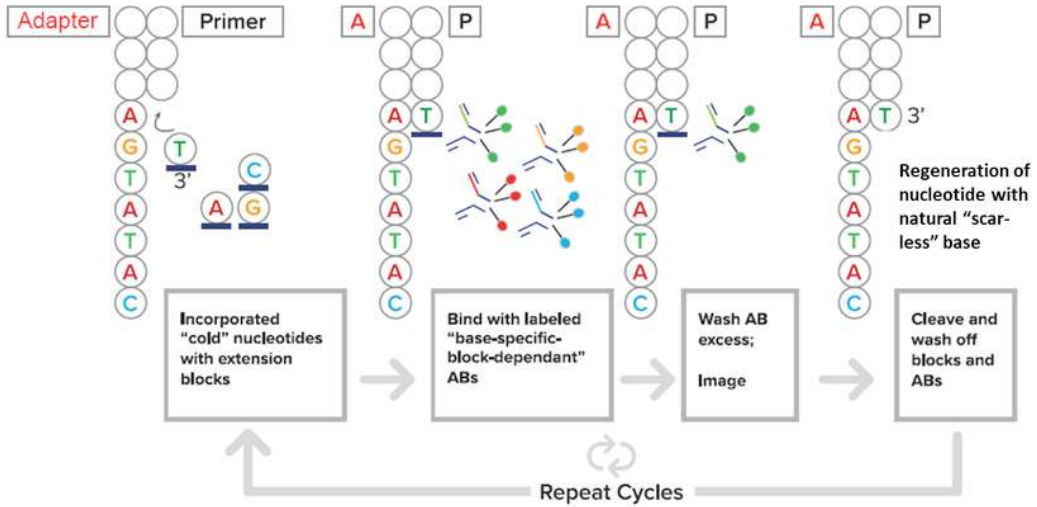
Physical confinement to patterned nanoarrays (nanowell of 200-300 nm):

- No competition for primers
- No runaway amplification
- No bias towards shorter fragments

MGI chemistry: cPAS

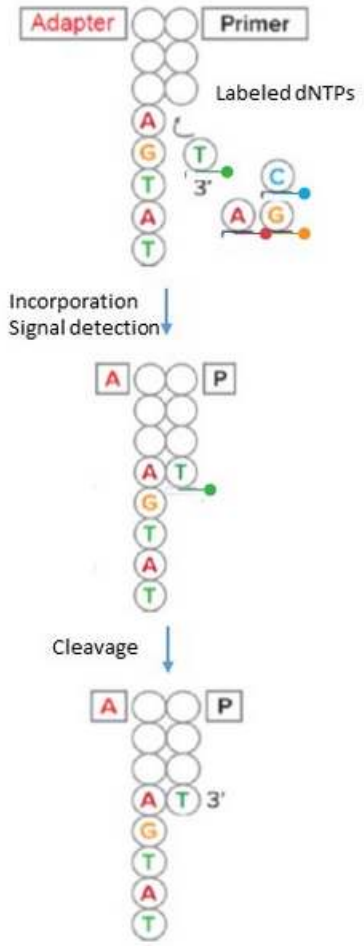
Combinatorial Probe-Anchor Synthesis

Cool



Feature	Cool Mode	Hot Mode
Temperature	Lower	Higher
Cycle time	Longer	Shorter
Accuracy	Higher	Slightly lower
Throughput/day	Lower	Higher
Typical applications	WGS, clinical	RNA-seq, large cohorts
Error profile	Lower substitution rate	Slightly higher substitution rate

Hot



Element Biosciences

SBB

Instrument	Run time /cell	Output /cell	Max reads /cell	Max read length, bp*
AVITI™	24-60 hrs	300 Gb	1 bln	PE 300
AVITI24™	24-60 hrs	450 Gb	1.5 bln	PE 300
AVITI LT™	17-51 hrs	150 Gb	500 mln	PE 300

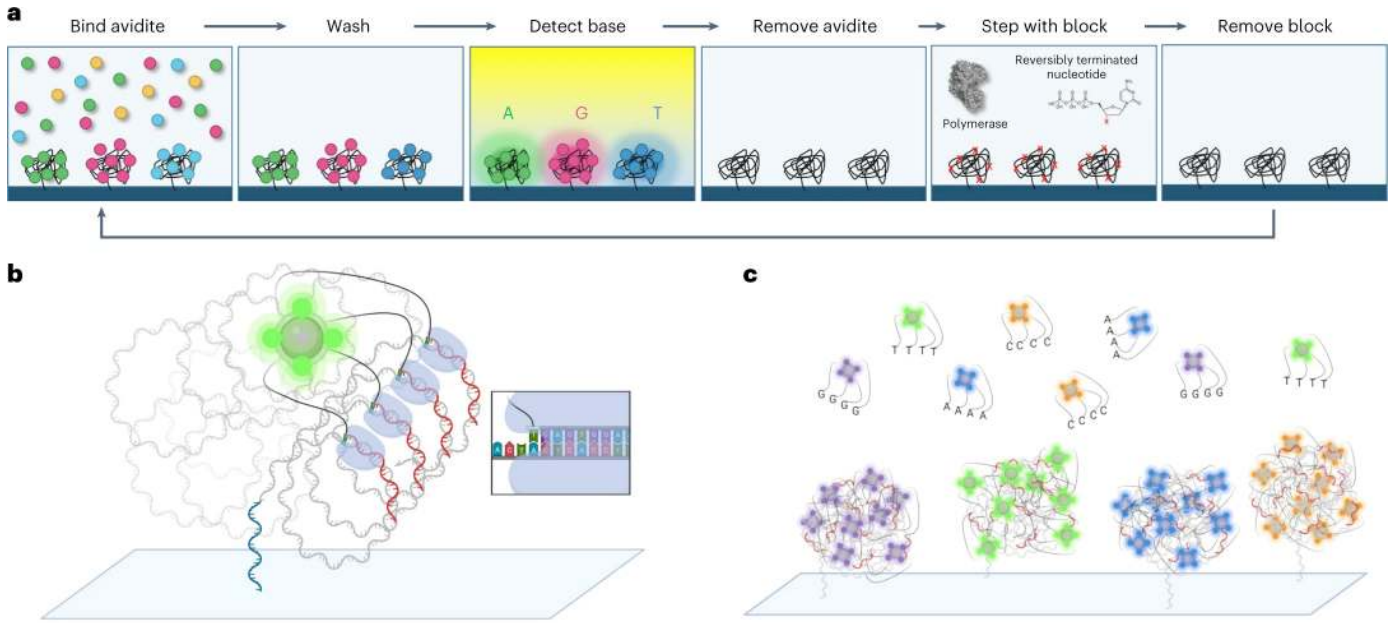
Technology highlight: DNA colonies (polonies), multiomics

Used for: everything + single cell multiomics

Strength: low cost, few artifacts, muti-omics on the same sample, Illumina-library compatible

Weakness: throughput limitations, GC homopolymers

Avidities and polonies



Polony = polymerase colony
Avidite = *multivalent* labelled probe, recognizes the incorporated nucleotide in a polony.
Reusable, reversible, strong signal.

Library is Illumina-compatible (P5 & P7 adaptors)

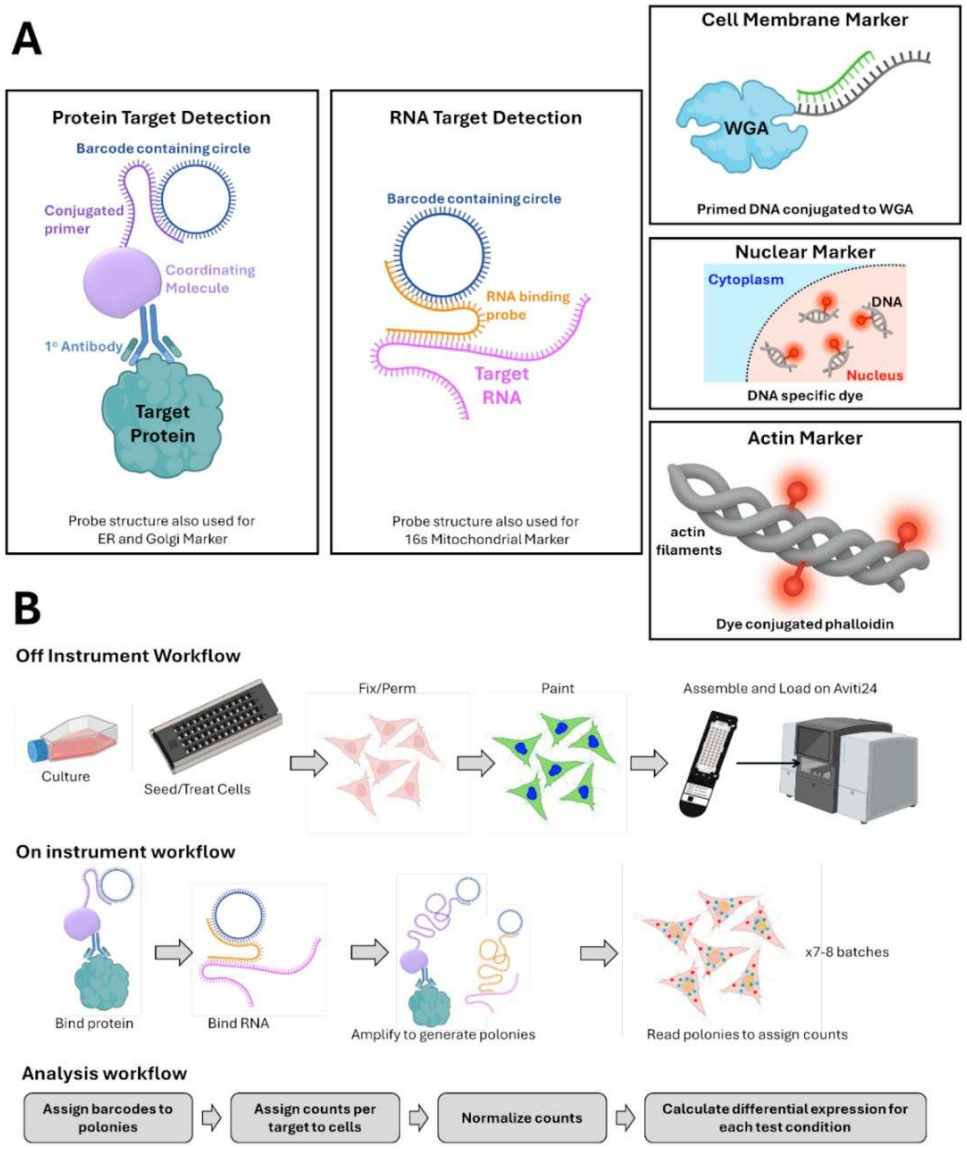
Similar to Illumina: amplification on a slide
Similar to MGI: **RCA**

Patterned flow cell

Decoupled incorporation and detection = low error rates

Think of Velcro

Multiomics on AVITI



Teton™ CytoProfiling

- Simultaneous detection of RNA, proteins and morphology in cells - *in situ* profiling approach
- Provides spatial information

Cells fixed on a slide > target probes hybridized > polonies formed > avidities added > signal detected.

RNASeq: ~350 transcripts with seq-specific probes
Spatial & single cell differential expression

Protein profiling: ~50 markers in the same cell

Cell morphology is scanned

PACBIO®

Instrument	Run time /cell	Output /cell	Max reads /cell	Max read length, bp*
Sequel II	12-30 hr	*30**	*4 M**	20 k, circular consensus HiFi
Revio	12-30 hr	*100-120 Gb**	*7 M**	20 k, circular consensus HiFi
Onso	48 hr	150 Gb	800-1000 M	PE 150

* What company tells you (achievable mainly on fresh human samples)

Technology highlight Revio: Single Molecule Real Time (SMRT) sequencing
Used for, Revio: everything where long reads are needed

Technology highlight Onso: SBB
Used for, Onso: "needle in the stack" applications

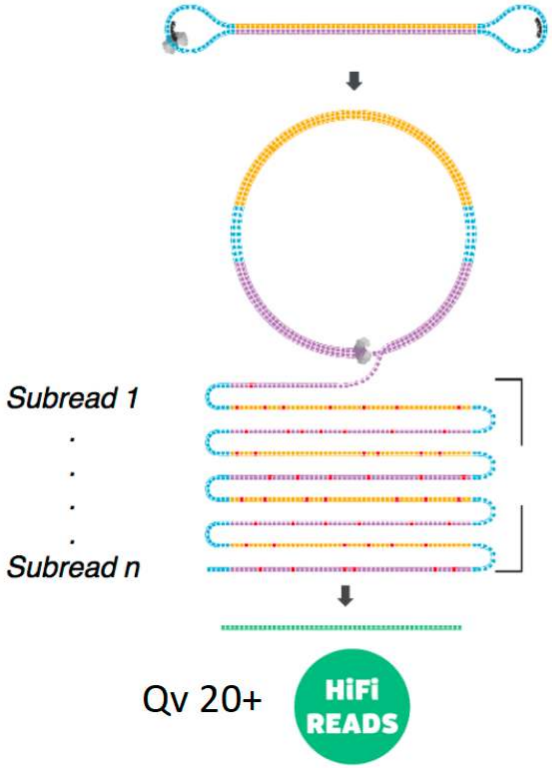
Strength: very high quality of reads
Weakness, Revio: sensitivity to DNA quality, GA bias
Weakness, Onso: higher cost, low throughput, new technology

PacBio HiFi: SMRT - technology

TWO MODES OF SMRT SEQUENCING

Circular Consensus Sequencing (CCS) Mode

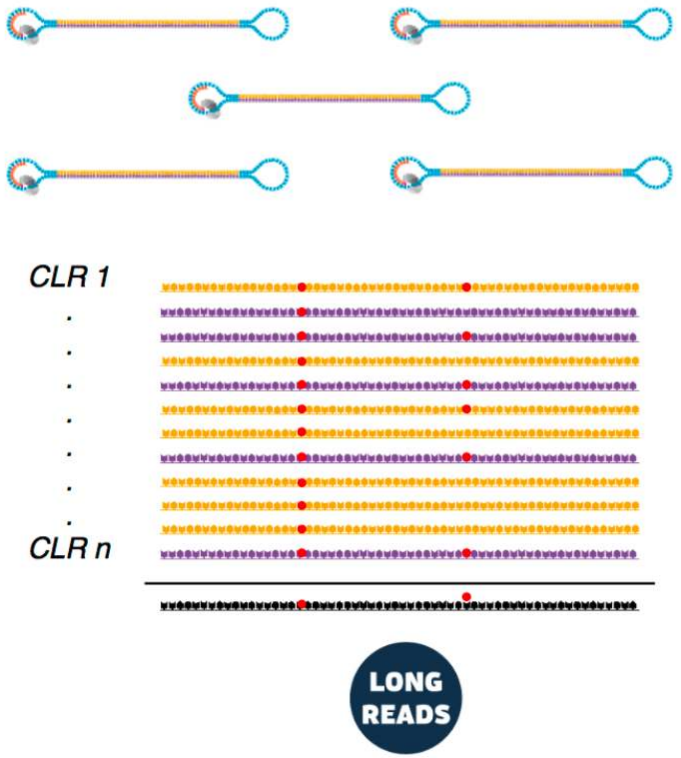
Inserts 10-20 kb



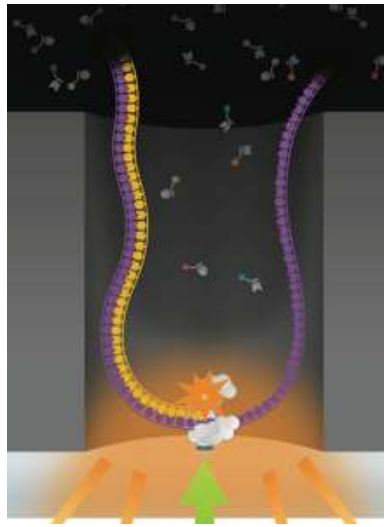
Single-molecule consensus sequence

Continuous Long Read (CLR) Sequencing Mode

Inserts >25 kb, up to 175 kb

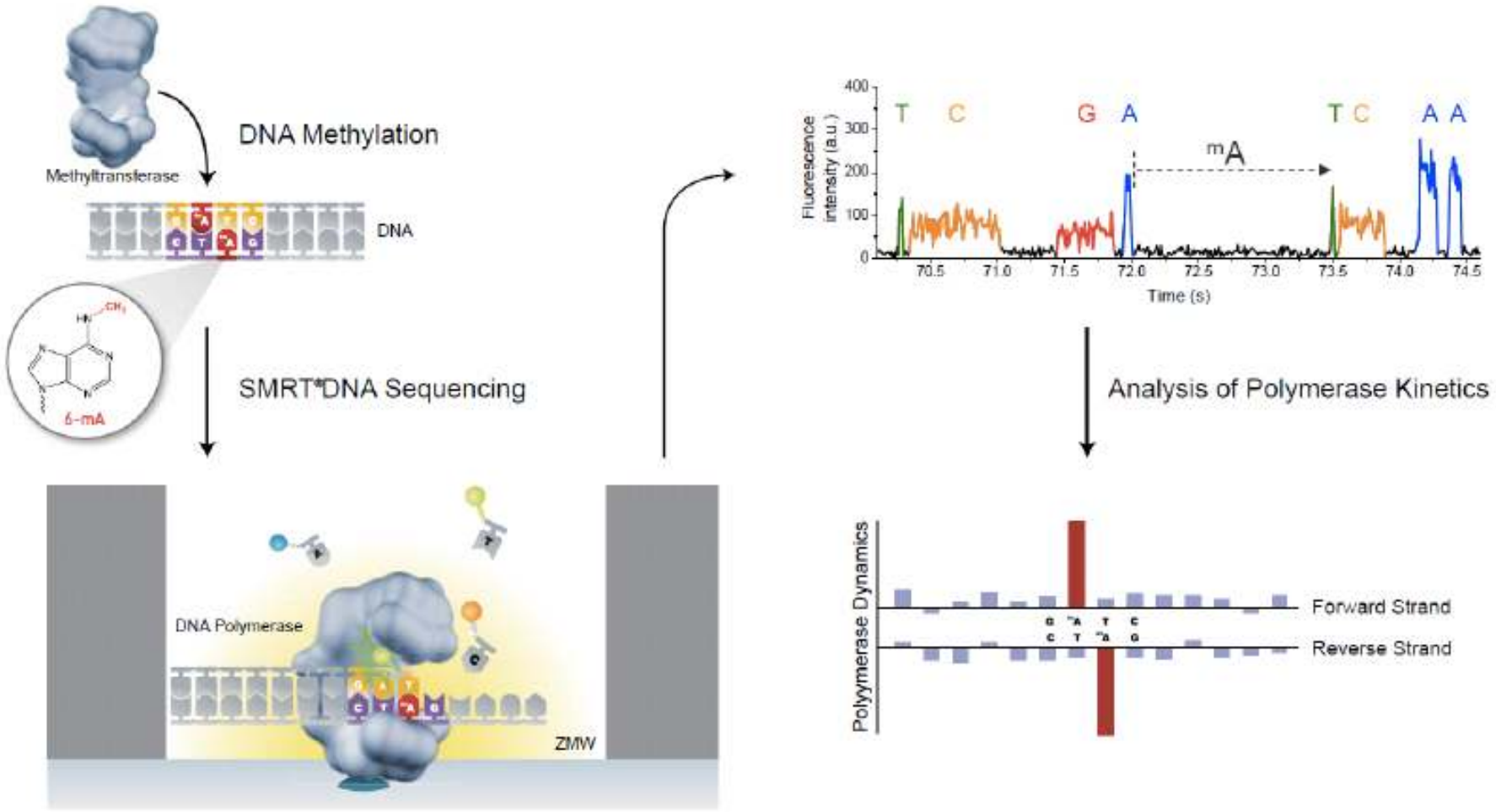


Multi-molecule consensus sequence



HiFi Q-score up to 50+

Base Modification: Discover the Epigenome

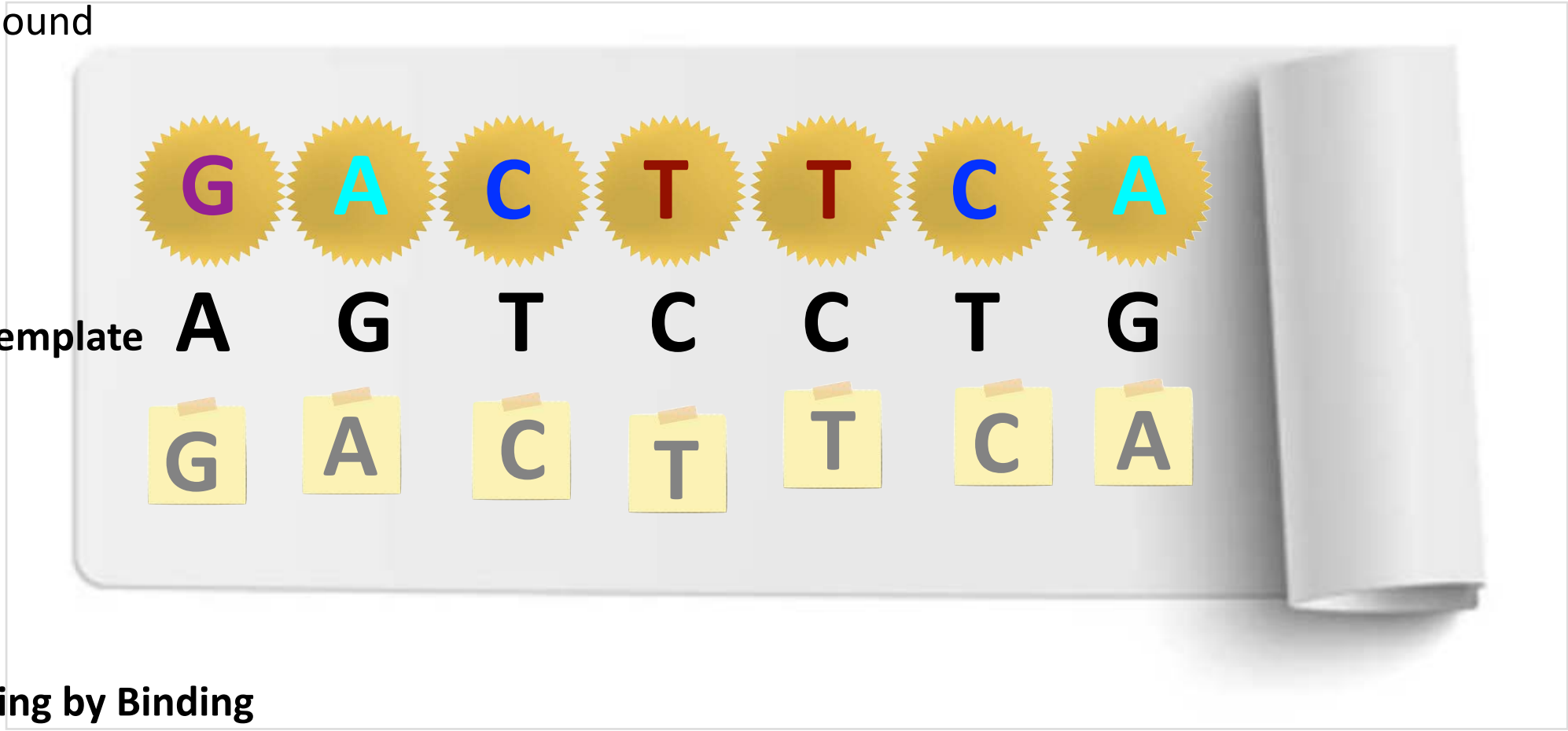


Detect base modifications using the kinetics of the polymerization reaction during normal sequencing

PacBio Onso: SBB reminded

Sequencing by Synthesis

dNTP* bound



Sequencing by Binding

dNTP* presented, then native nucleotide is bound

Comparison Illumina vs Onso

Onso utilizes a proprietary cluster generation technology, non-patterned flow cell

Feature	Sequencing by Synthesis (Illumina)	Sequencing by Binding (Onso)
Detection	Fluorescent labeled nucleotides	Transient binding probes
Incorporation	Modified nucleotides	Native nucleotides
Accuracy	High (Q30–Q40)	Ultra-high (Q40+)
Error type	Phasing errors common	Very low phasing issues
Read type	Short reads	Short but ultra-accurate

Oxford NANOPORE Technologies™

Instrument	Run time /FC	Output / FC	Nr of pores	Max read length
Flongle	16 hrs	1 Gb	126	1 Mb
MinION	24 hrs	2-15 Gb	512	1 Mb
GridION	24 hrs	2-15 Gb	512	1 Mb
PromethION	72 hrs	10 – 150 Gb	3 000	2 Mb

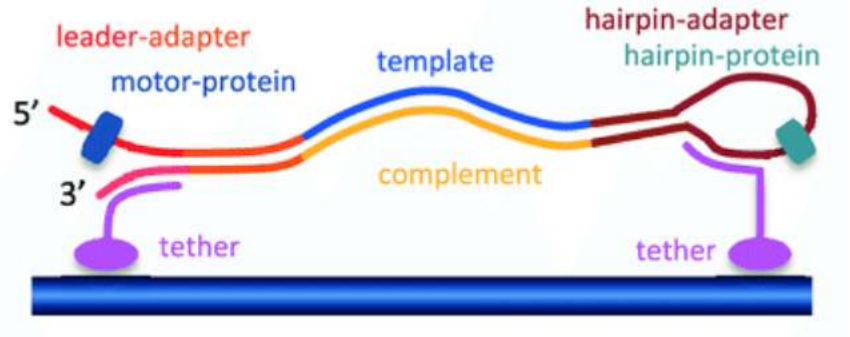
Technology highlight: nanopores, no fluorescence, ultra-long reads, native RNA

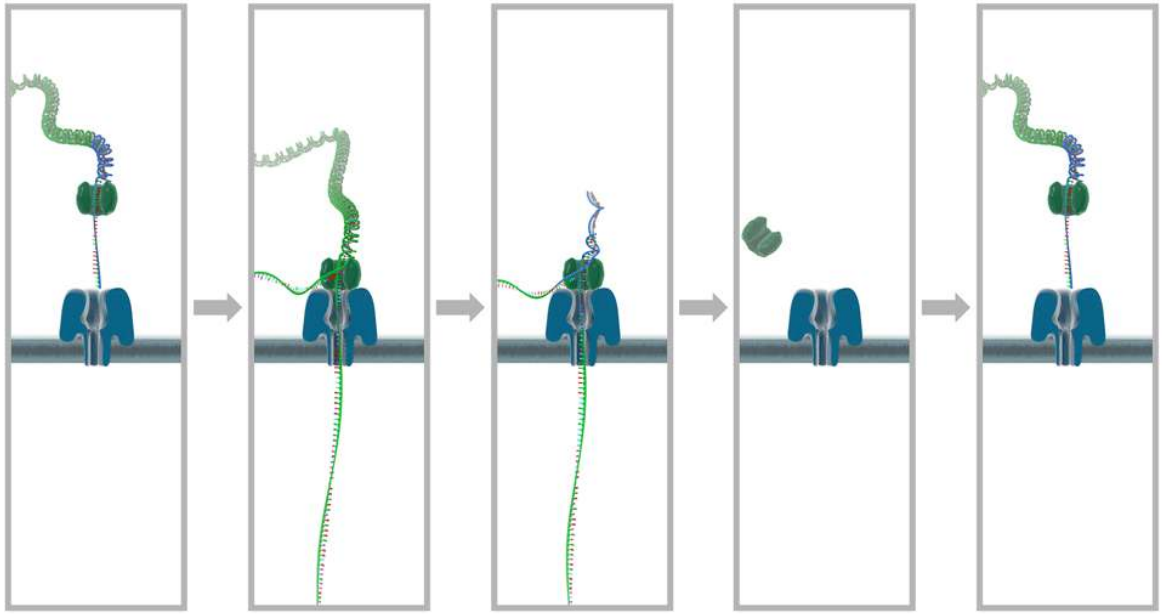
Used for: reference genomes, SV detection

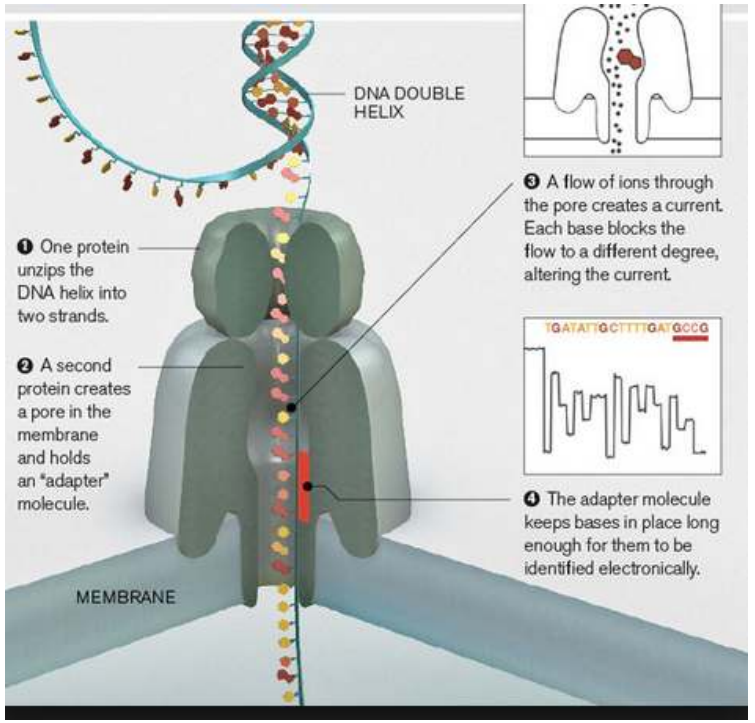
Strength: size and portability

Weakness: sample-sensitive, software limitations, lower Q-score than PacBio

ONT: DNA + Motor + Pore







Q-score 20-30

Main advantages of ONT: SPEED and PORTABILITY

Rapid Confirmation of the Zaire Ebola Virus in the Outbreak of the Equateur Province in the Democratic Republic of Congo: Implications for Public Health Interventions

Placide Mbala-Kingebeni, Christian-Julian Villabona-Arenas, Nicole Vidal, Jacques Likofata, Justus Nsio-Mbeta, Sheila Makiala-Mandanda, Daniel Mukadi, Patrick Mukadi, Charles Kumakamba, Bathe Djokolo ... Show more

Clinical Infectious Diseases, Volume 68, Issue 2, 15 January 2019, Pages 330–333, https://doi.org/10.1093/cid/ciy527

Published: 29 June 2018 **Article history**

ORIGINAL ARTICLE BRIEF REPORT

A Novel Coronavirus from Patients with Pneumonia in China, 2019

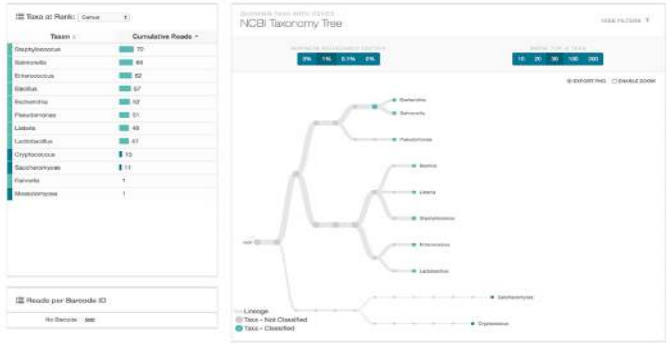
Na Zhu, Ph.D., Dingyu Zhang, M.D., Wenling Wang, Ph.D., Xinwang Li, M.D., Bo Yang, M.S., Jingdong Song, Ph.D., Xiang Zhao, Ph.D., Baoying Huang, Ph.D., Weifeng Shi, Ph.D., Roujian Lu, M.D., Peihua Niu, Ph.D., Faxian Zhan, Ph.D., et al., for the China Novel Coronavirus Investigating and Research Team

RESEARCH ARTICLE | Full Access

Semi-quantitative characterisation of mixed pollen samples using MinION sequencing and Reverse Metagenomics (RevMet)

Ned Peel, Lynn V. Dicks, Matthew D. Clark, Darren Heavens, Lawrence Percival-Alwyn, Chris Cooper, Richard G. Davies, Richard M. Leggett, Douglas W. Yu

First published: 15 July 2019 | https://doi.org/10.1111/2041-210X.13265

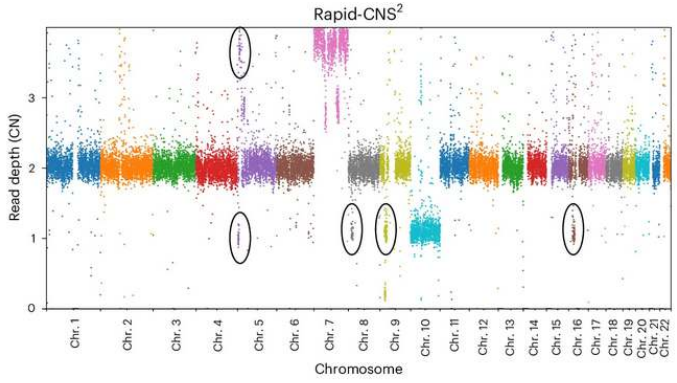


Article | Open access | Published: 11 October 2023

Ultra-fast deep-learned CNS tumour classification during surgery

C. Vermeulen, M. Pagès-Gallego, L. Kester, M. E. G. Kranendonk, P. Wesseling, N. Verburg, P. de Witt Hamer, E. J. Kooi, L. Dankmeijer, J. van der Lugt, K. van Baarsen, E. W. Hoving, B. B. J. Tops & J. de Ridder

Nature **622**, 842–849 (2023) | Cite this article



Roche

Sequencing by expansion

Instrument	Run time /cell	Output /cell	Max reads /cell	Max read length, bp*
AXELIOS	500 mln bp/sec	5 bln/hour	75-90 bln	SE 1 kb

Technology highlight: novel sequencing by expansion, uses nanopores
Used for: massive sample size, high sensitivity, poor quality samples

Strength: highest speed on the market, very high accuracy
Weakness: high running costs, labor-intensive, software limitations

Sequencing by expansion (SBX)

Xpandomer - synthetic molecule
50x longer than the template

Standard NGS library construction
SBX encoding: from template to Xpandomers - proprietary expandable nucleotides (X-NTPs)
Single molecule detection by CMOS sensor array (*Complementary Metal-Oxide-Semiconductor*)

X-A, X-G, X-C, X-T carry a **reporter code** - later will produce a high-signal readout
Translocation control elements - movement through a nanopore
Enhancer - robust synthesis
Acid-cleavable bonds - allow polymer expansion after synthesis

Enzyme: XP synthase + help from polymerase enhancers (PEMs)

No amplification

Xpandomers, CMOS and nanopores

Xpandomers make the electronic signal easy to read

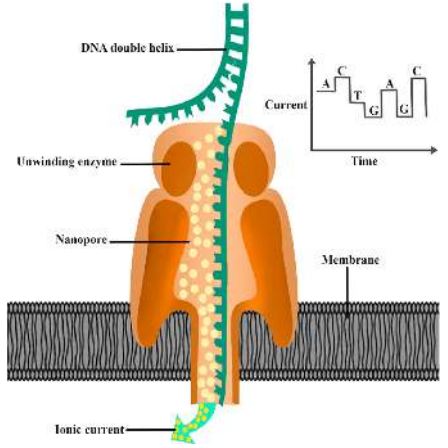
CMOS detect Xpandomer units, not bases
Each Xpandomer = large, base-specific reporter structure
X-A, X-G, X-C and X-T have unique engineered electrical signatures

1 Xpandomer passes the pore at a time (unlike constant stream in ONT)

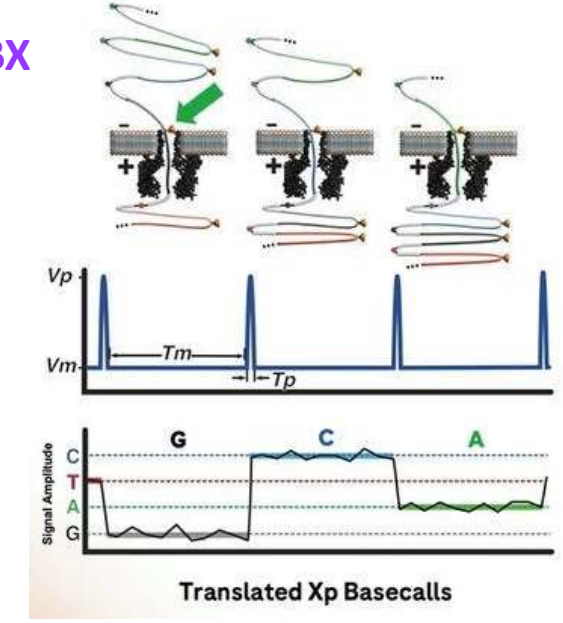
Signal is strong, base-specific and time-separated

Homopolymers: each Xpandomer is read separately, but long stretches is always an issue

ONT



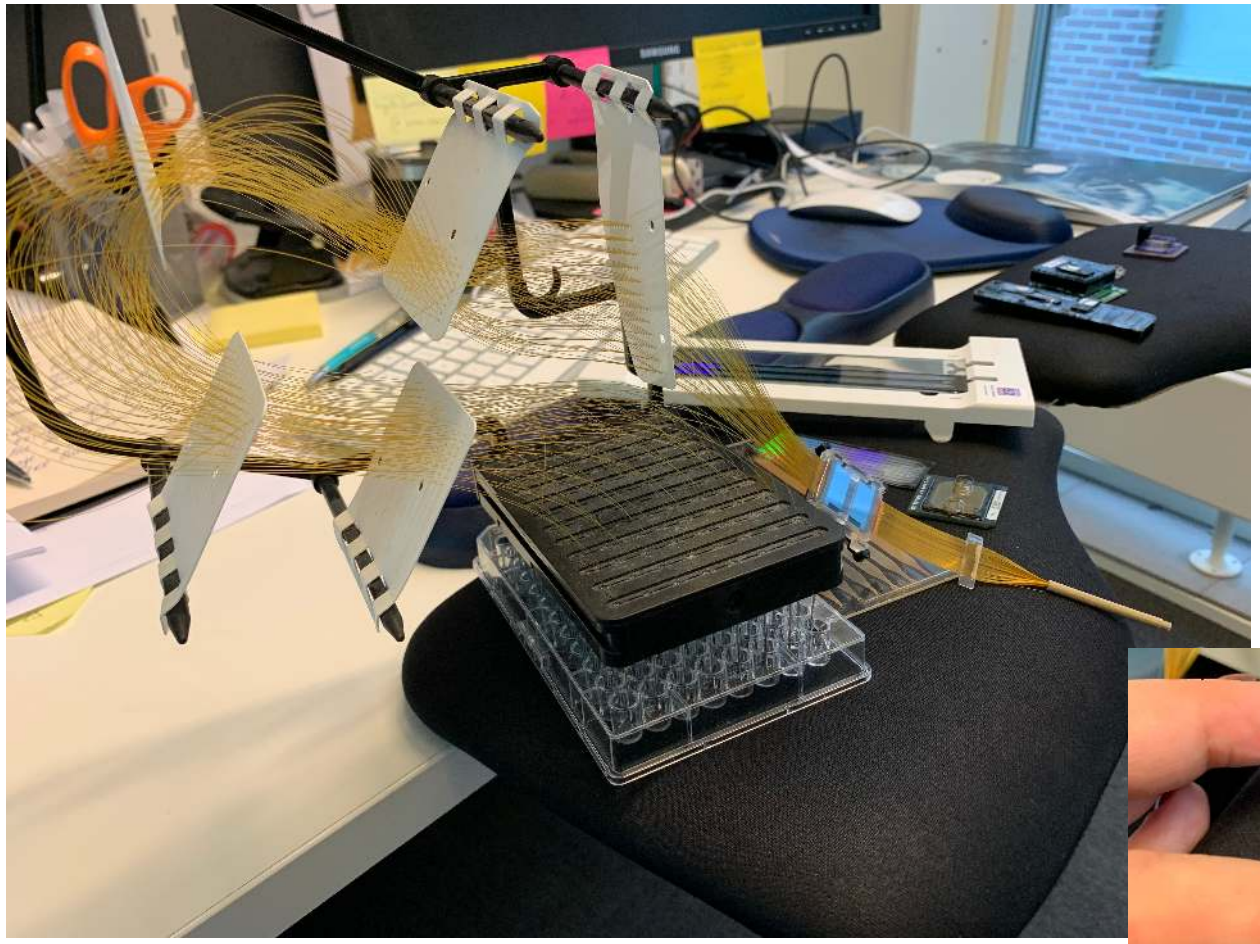
SBX

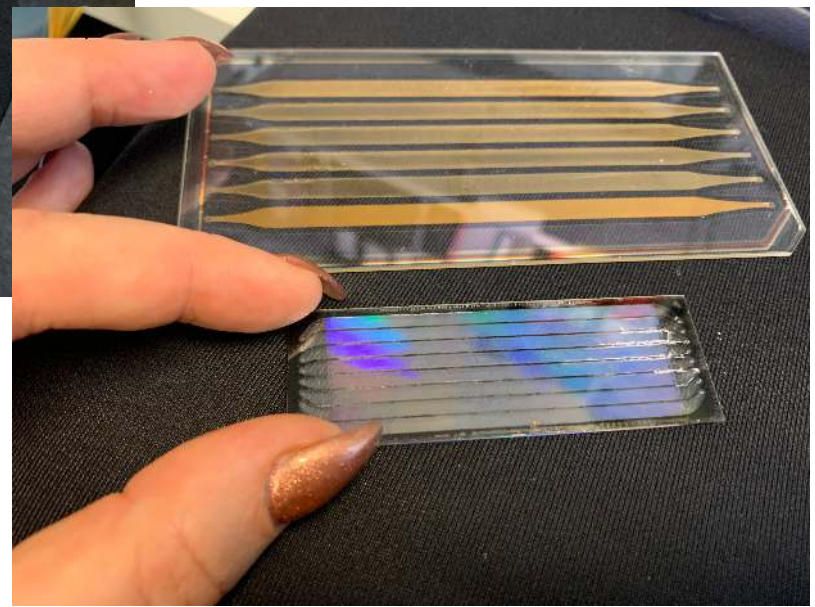


Let’s compare all the short reads

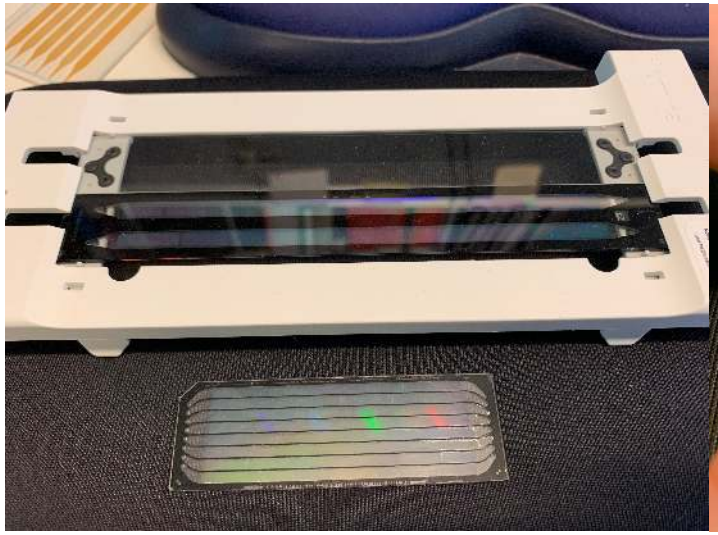
Platform	Sequencing type	Library / amplification	Signal detection	Main advantage	Main disadvantage	Error profile	Q score	Best uses
Illumina	SBS, optical	Bridge PCR	Fluorescence	Accurate, common	Old?	Substitutions	Q30-40	WGS, RNA
Utima	SBS, flow-based	Emulsion PCR	Fluorescence	Throughput, cost	Homopolymers, repeats	Substitutions, homopolymers	Q30-35	High-throughput
Element	SBB with avidities	RCA polonies	Fluorescence	High accuracy	Small ecosystem	Low substitutions	Q40+	Variant calling
MGI	SBS with DNB	RCA nanoballs	Fluorescence	High accuracy	Complex, loading issues	Low substitutions	Q30-35	High-throughput
ION	Semiconductor pH	Emulsion PCR	pH change	Fast, simple Inbuilt analysis	Homopolymers, single-end	Indels in homopolymers	Q30	Gene panels
Roche	SBE	Xpandomers	Electronic	Ultra-fast	Higher cost/base, new tech	New tech	Q40+	Fast, rare variant detection
PacBio ONSO	SBB	?	Fluorescence	High accuracy, superb in homopolymers	Higher cost/base, lower output	?	Q40-50+	Rare variant detection

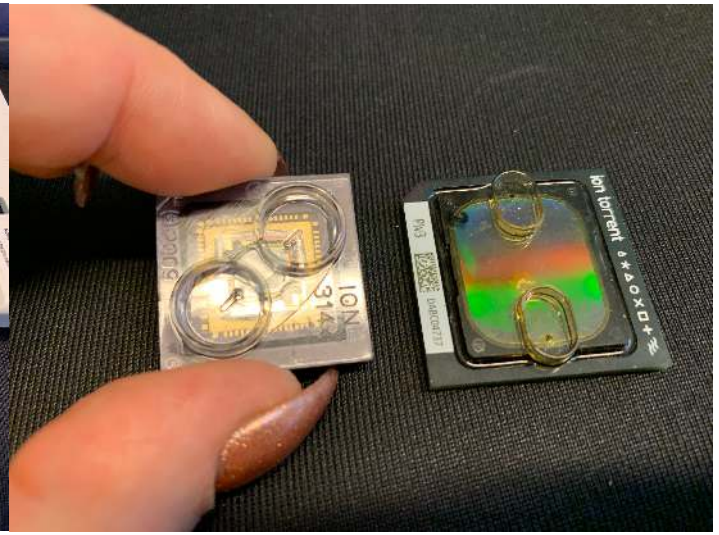
How it looked yesterday

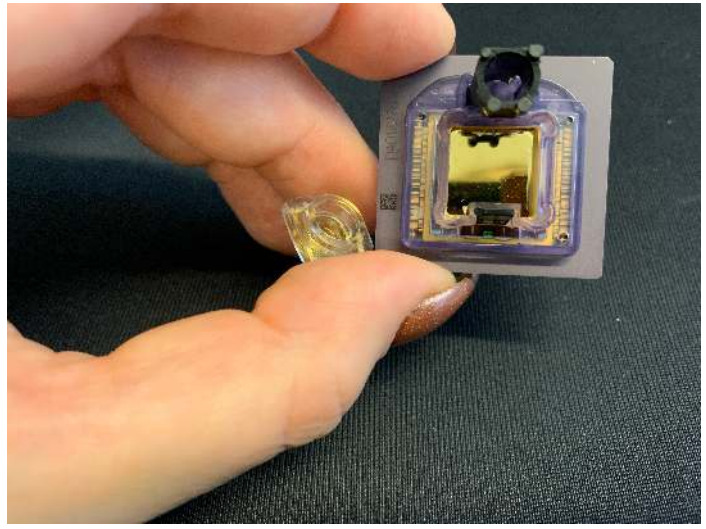




How it looks now: technology progress







- Sequencing technologies
- Informatics
- Biotech & medical applications
- Ecosystem science
- Reference genome sequencing

NGS Technologies and Applications

NGS technologies

Short read NGS

Long-read NGS





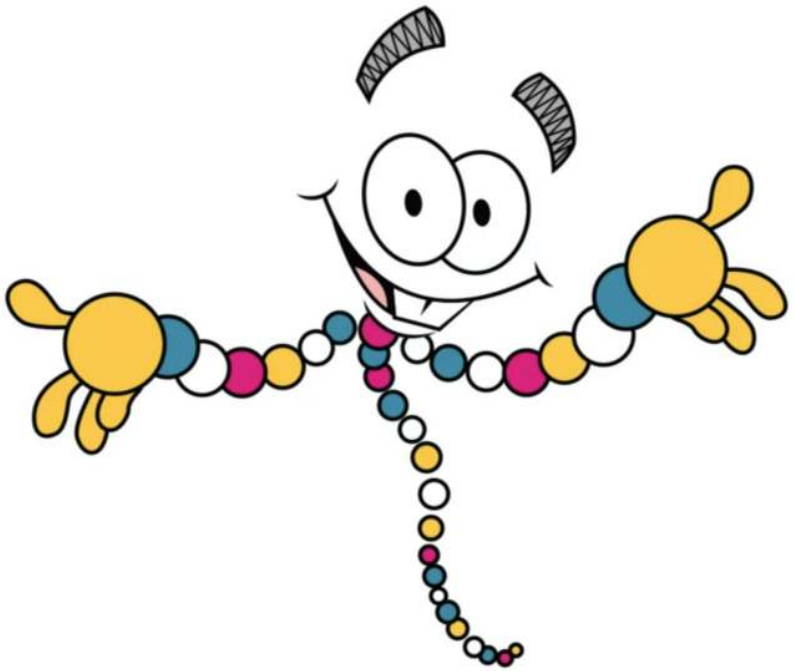



Whole genome re-sequencing
RNA-seq
Targeted re-seq
Panels
Amplicons up to 600 bp

De novo genome sequencing
Whole-transcript sequencing
Structural variant resolving
Targeted re-seq
Amplicons up to 13 kb

Research and development

Thank you!



Swedish Research Council

UPPSALA UNIVERSITET

SciLifeLab

European Commission
Horizon Europe
2021-2027