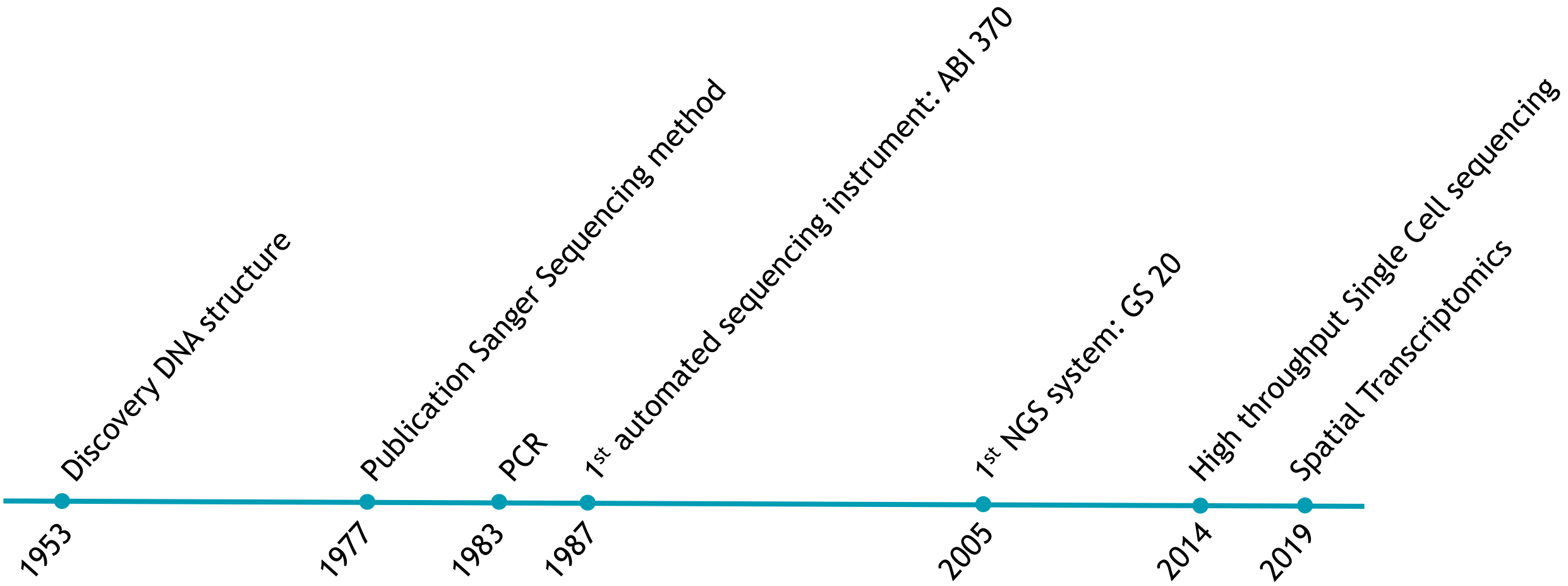
Amsterdam UMC
Universitair Medische Centra

Next Generation Sequencing

Core Facility Genomics

Marja Jakobs

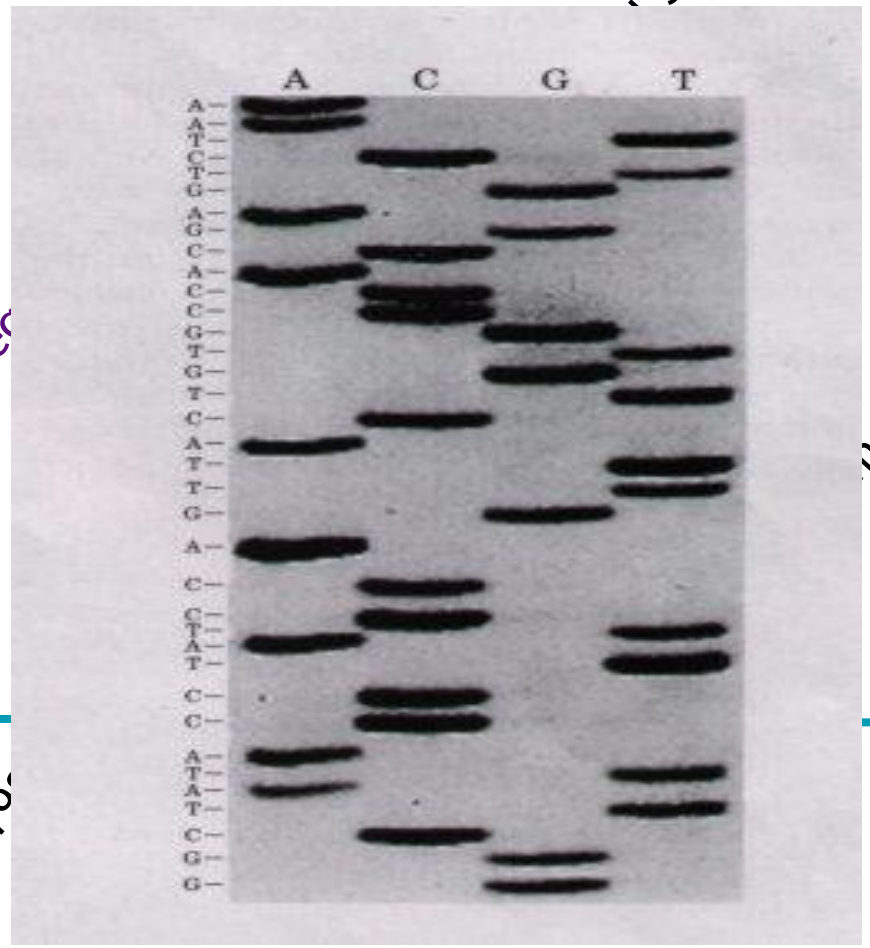

1953
Discovery DNA structure
1977
Publication Sanger Sequencing method
1983
PCR
1987
1st automated sequencing instrument: ABI 370
2005
1st NGS system: GS 20
2014
High throughput Single Cell sequencing
2019
Spatial Transcriptomics

1953 Discovery DNA structure

1977 Publication Sanger Sec

1983 PCR

370

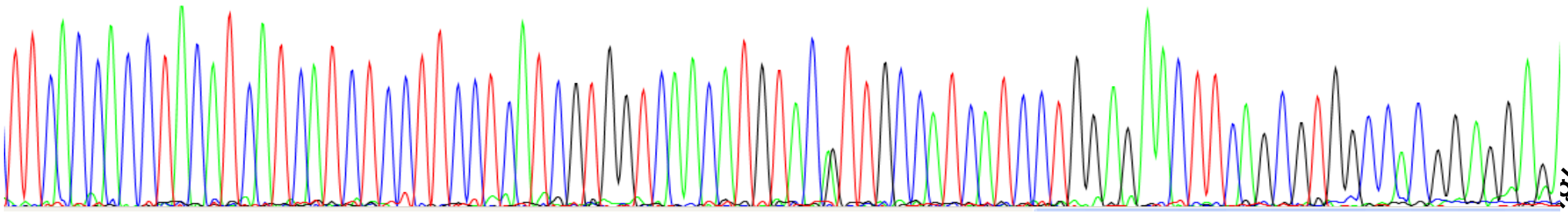


Dideoxynucleotides and
radioactive labeled fragments

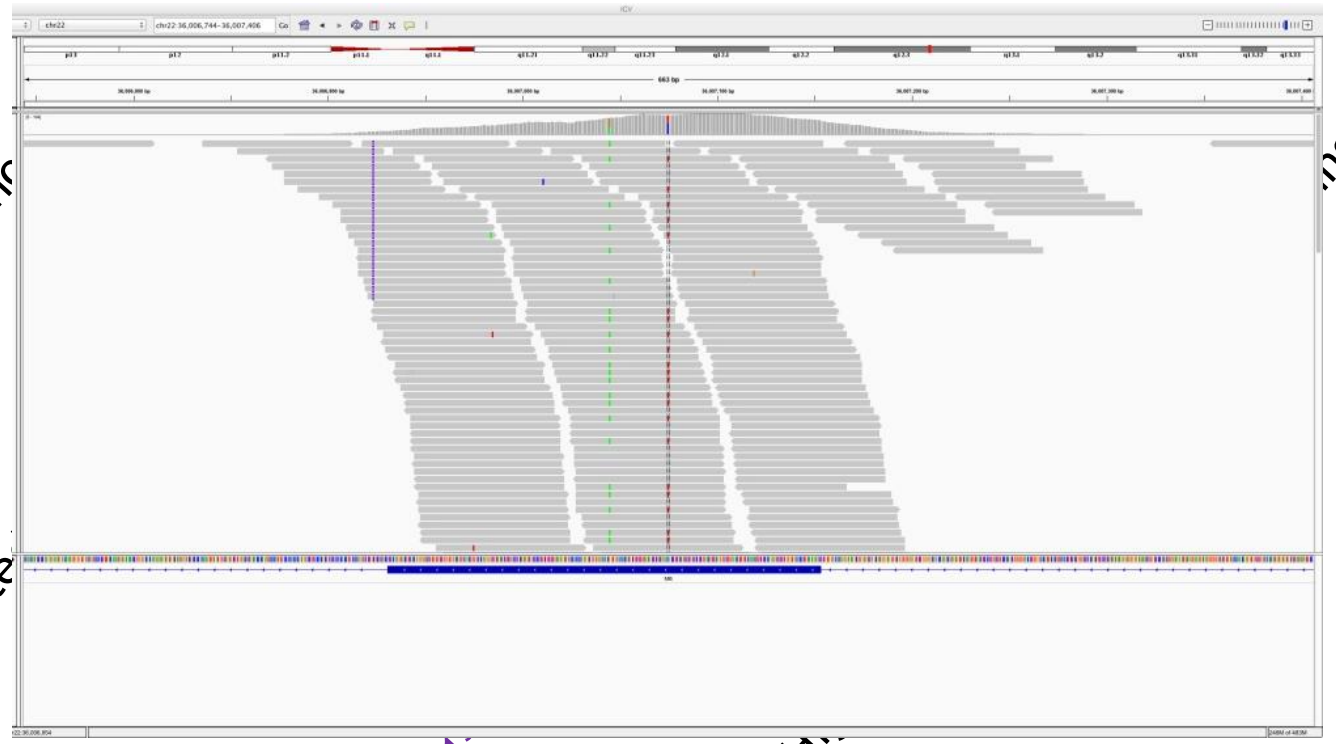
n: GS 20

2014 High throughput Single Cell sequencing

2019 Spatial Transcriptomics



1996 ABI 310: Capillary sequencer using fluorescent labelled ddNTPs



1953 Discovery DNA structure

1977 Publication Sanger Sequencing meth

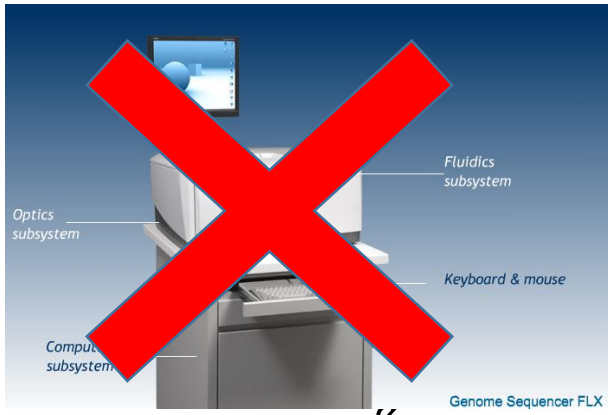
1983 PCR

1987 1st automate

2005

2014

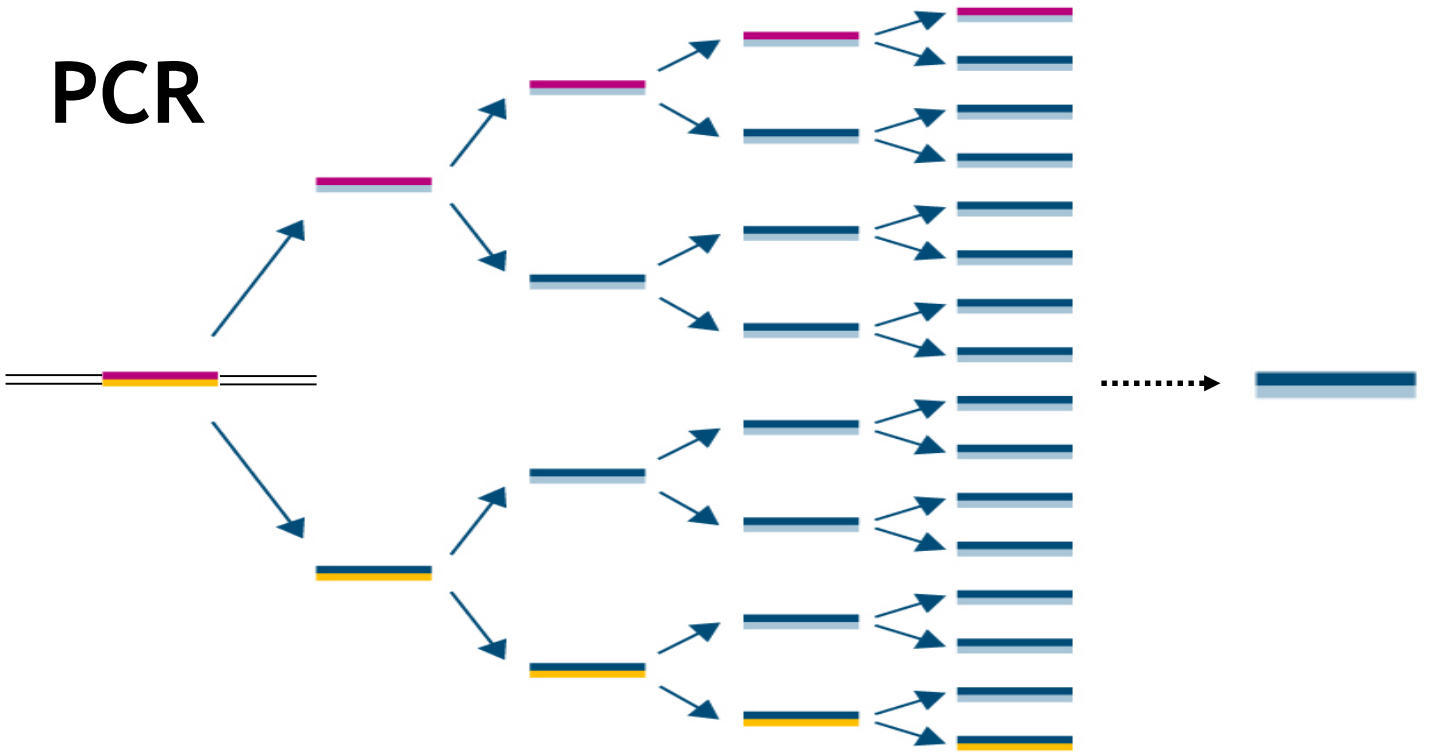
2019


Fluidics subsystem
Optics subsystem
Keyboard & mouse
Genome Sequencer FLX


Ion S5
ion torrent

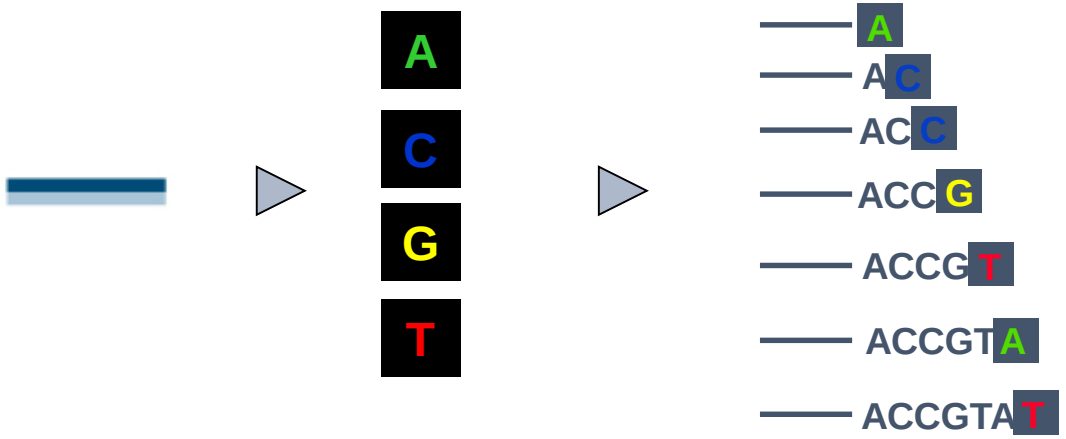

NovaSeq 6000
illumina

Conventional Sanger sequencing



Sanger Sequencing

DNA polymerase, dNTP's
and dye-labeled dideoxy-
nucleotides (terminators)



PCR product extension products

Next Generation Sequencing

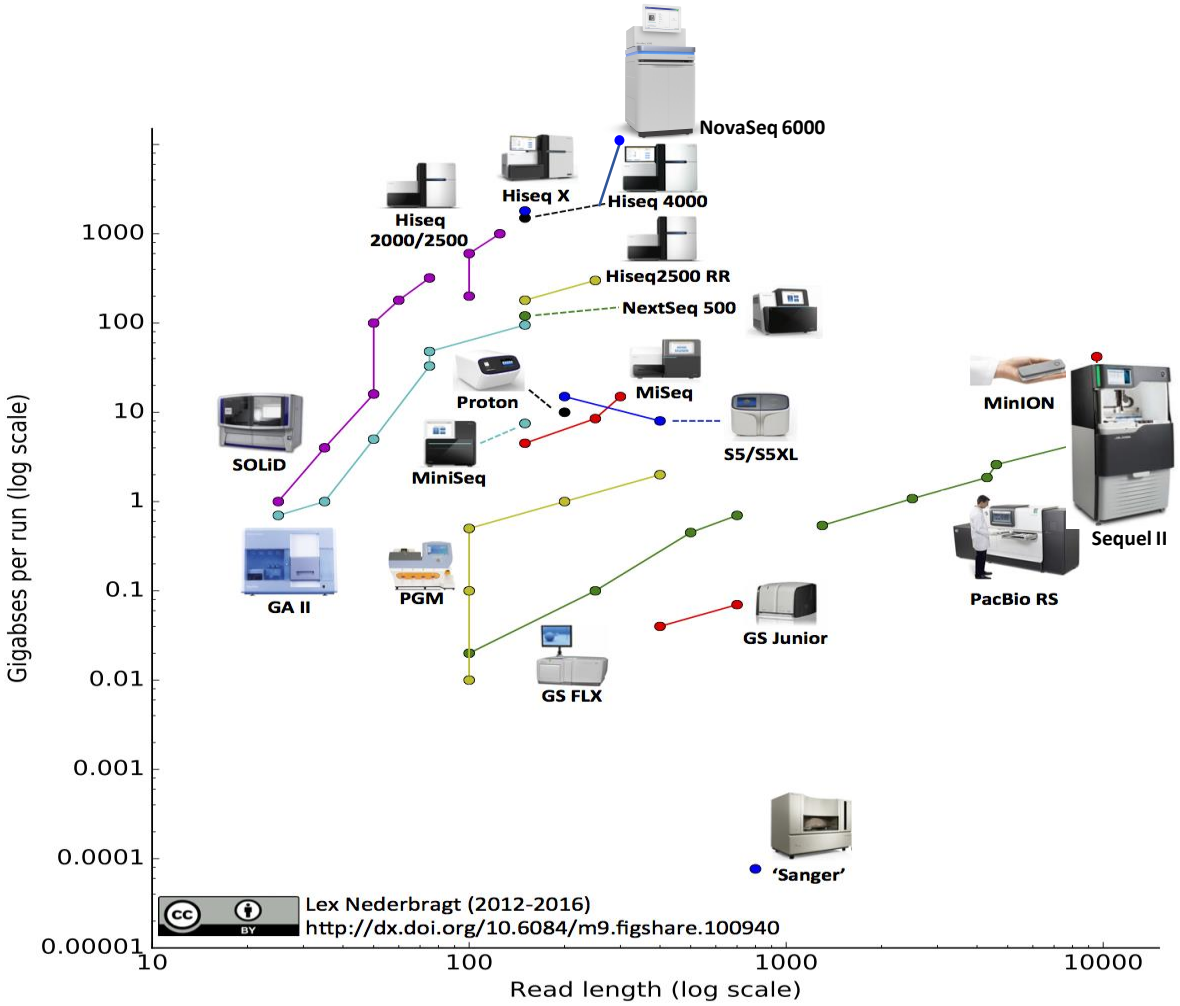
Mass Parallel Sequencing of unique DNA molecules

Main NGS platforms:

- PacBio
- Oxford Nanopore
- Illumina
- Ion Torrent

Throughput vs readlenght

Sequencing platforms



Differences in Sequencing Strategies

CONVENTIONAL		NGS
one sample	⟷	Pool of molecules
one tube	⟷	one reaction vessel
one reaction	⟷	many reactions
one result	⟷	many results

Several NGS strategies

Single molecule vs clonal amplification

- Sequencing by synthesis
- Sequencing with terminators

Two approaches

Single molecule

- SMRT sequencing

  Detection of incorporation of dNTPs in a single molecule (PacBio)

- Nanopore sequencing

  Pulling a lineair (c)DNA / RNA strand through a nanopore (NanoPore)

Clonal amplification

- Amplify a single molecule to obtain a pool of identical molecules

  - Emulsion PCR (IonTorrent)
  - Polony formation (Illumina)

Single vs Amplified

Single strand

- No amplification, no bias
- No copying erros
- Long reads
- Modifications stay intact

- Very low signal
- High error rate (~14%)

Clonal Amplification

- Strong signals
- PCR errors are averaged
- High throughput

- PCR bias/errors
- Short reads
- Loose modifications

Next Generation Sequencing

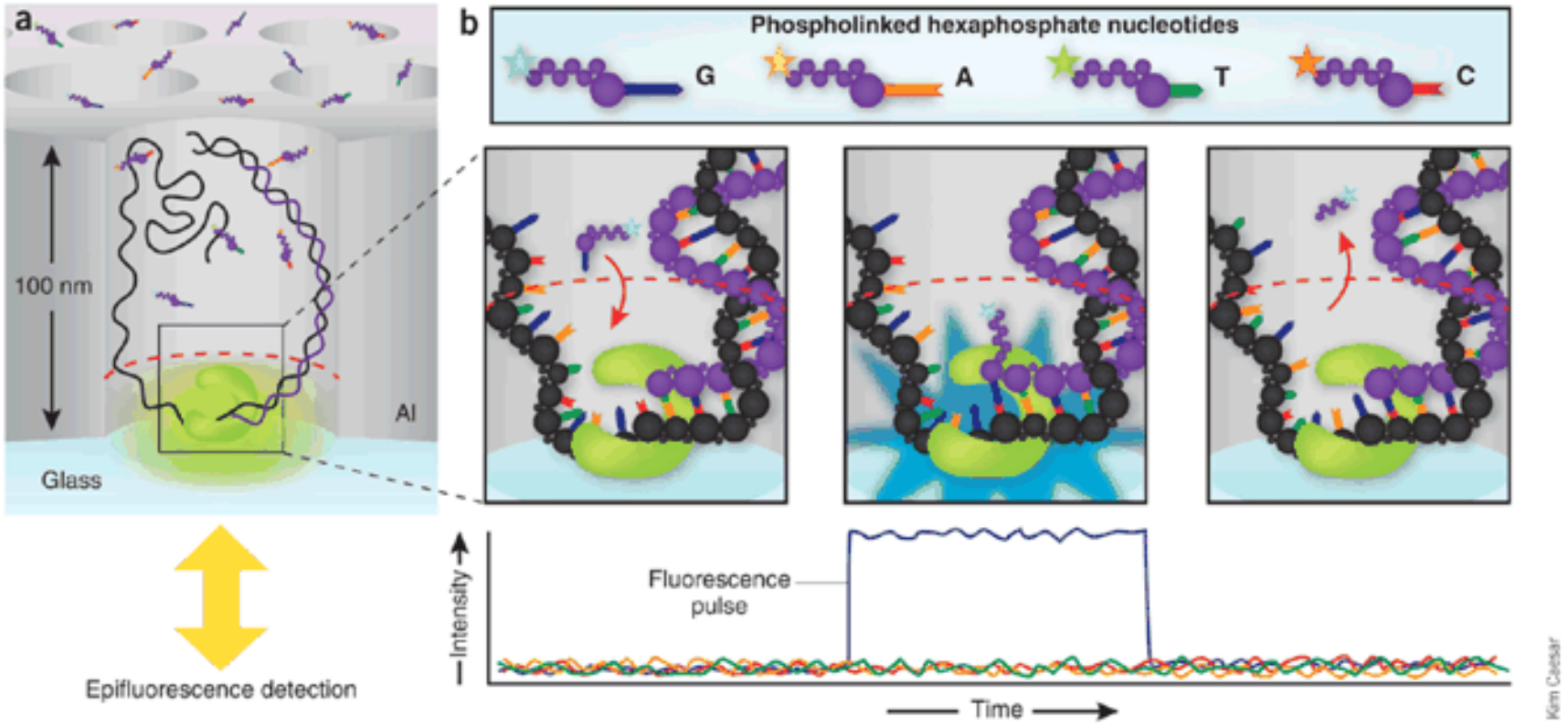
Mass Parallel Sequencing of unique DNA molecules

Main NGS platforms:

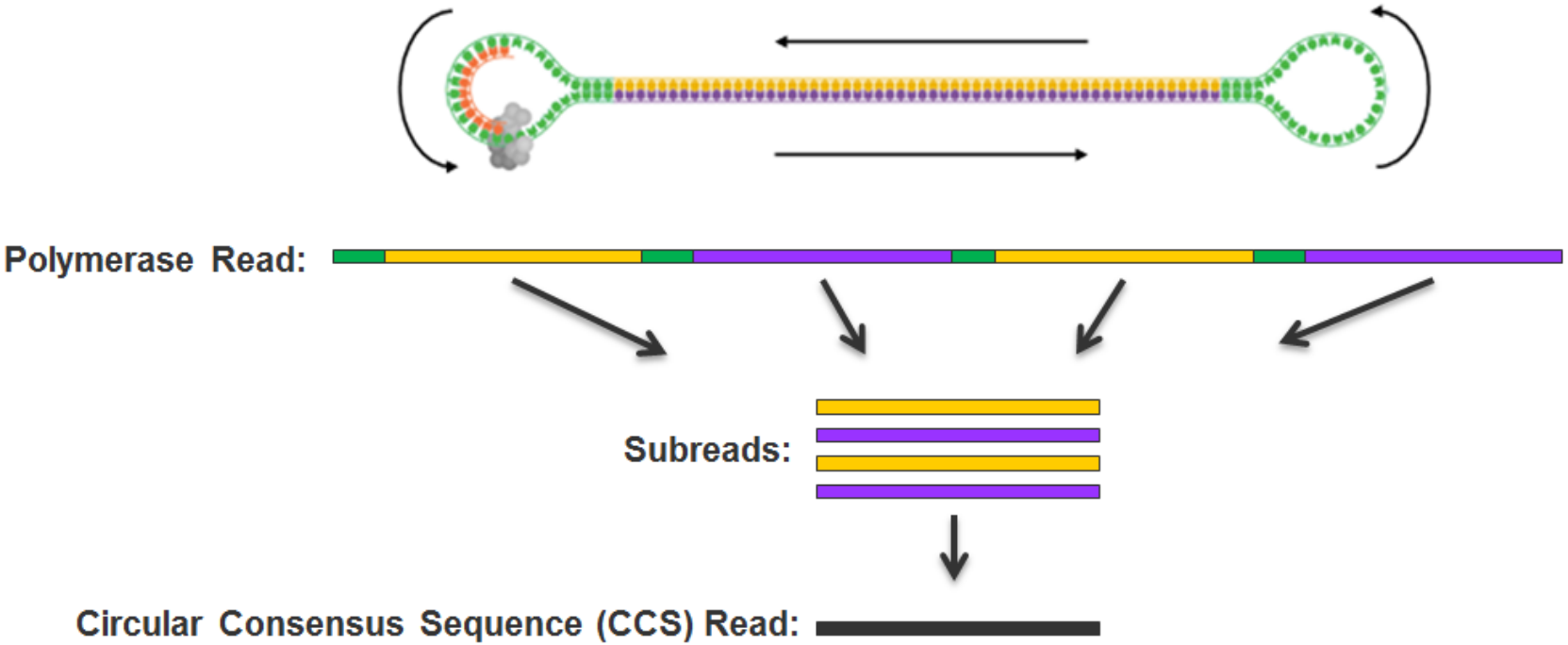
- PacBio
- Oxford Nanopore
- Illumina
- Ion Torrent

Single molecule, labeld dNTP's

Single Molecule Real Time sequencing
(PacBio)



Single Molecule Real Time sequencing
(PacBio)



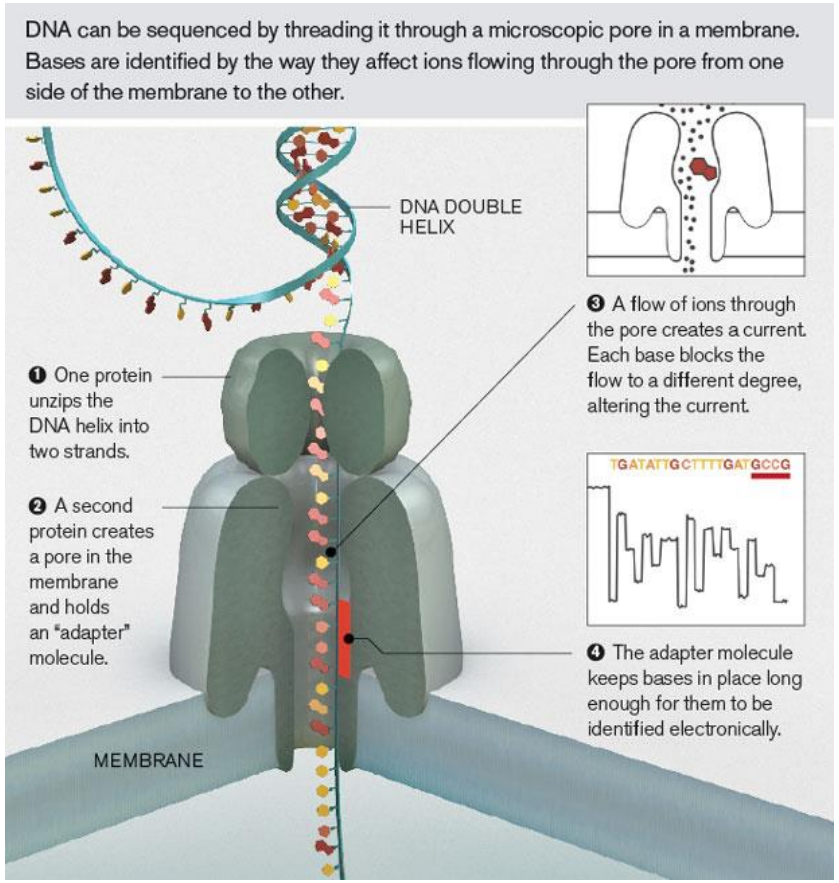
Next Generation Sequencing

Mass Parallel Sequencing of unique DNA molecules

Main NGS platforms:

- PacBio
- Oxford Nanopore — Single molecule
- Illumina
- Ion Torrent

Single Molecule sequencing
(Oxford Nanopore)



Next Generation Sequencing

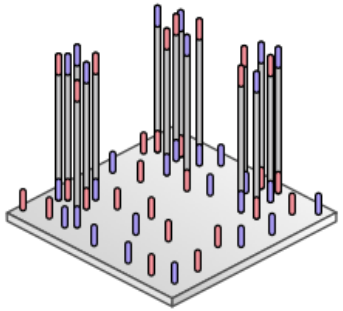
Mass Parallel Sequencing of unique DNA molecules

Main NGS platforms:

- PacBio
- Oxford Nanopore
- Illumina — Clonal amplification, terminator Seq
- Ion Torrent — Clonal amplification, synthesis Seq

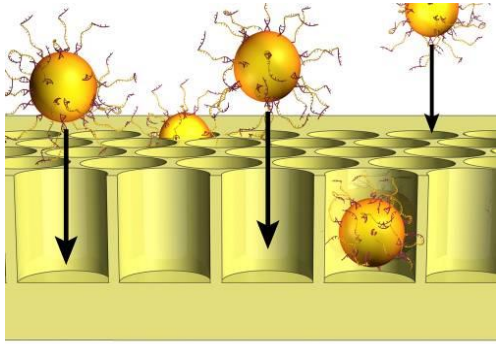
Clonal amplification

- Illumina (polony formation)



Repetition forms clusters
of identical strands

- IonTorrent (emulsion PCR)



Next Generation Sequencing platforms AMC / VUmc

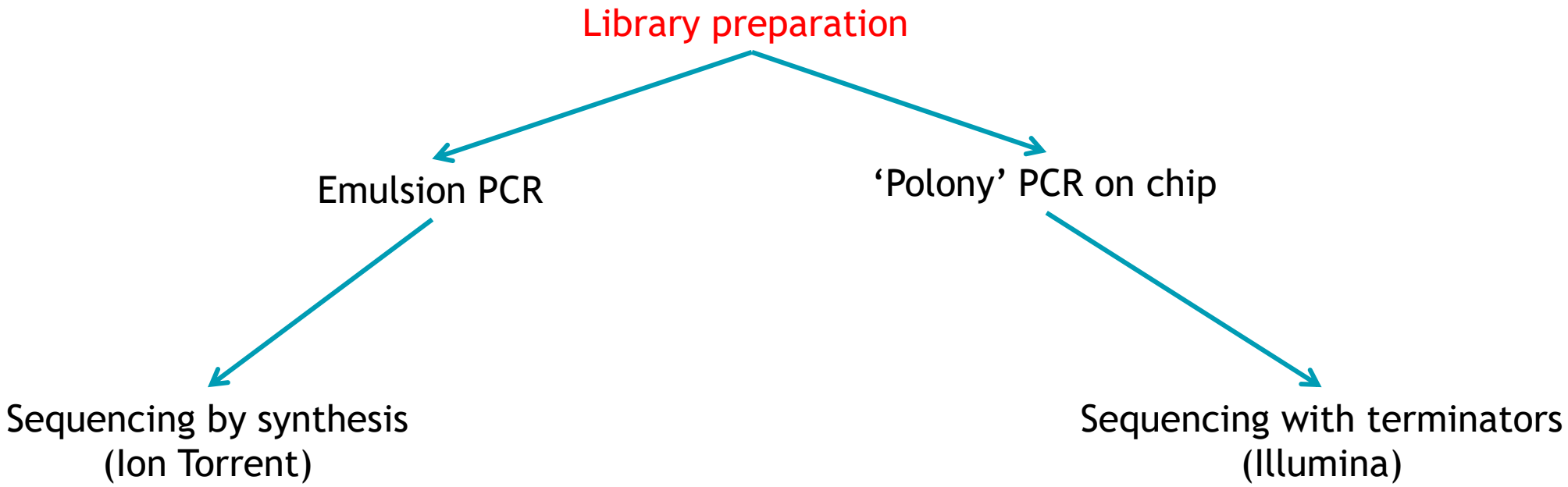
- **PacBio**
  - Sequel II
- **Oxford Nanopore**
  - MinIon
- **Illumina**
  - MiSeq
  - HiSeq 4000
- **Ion Torrent**
  - GeneStudio S5 Prime

High-throughput sequencing

	Sanger ABI
Run time sequencer	96 samples / hour
Through put	5 *10^5 x 500 bp ↓ 250M bp / **year!**
Max sequence length	500 bp
Analysis time	15 min sample

Workflow

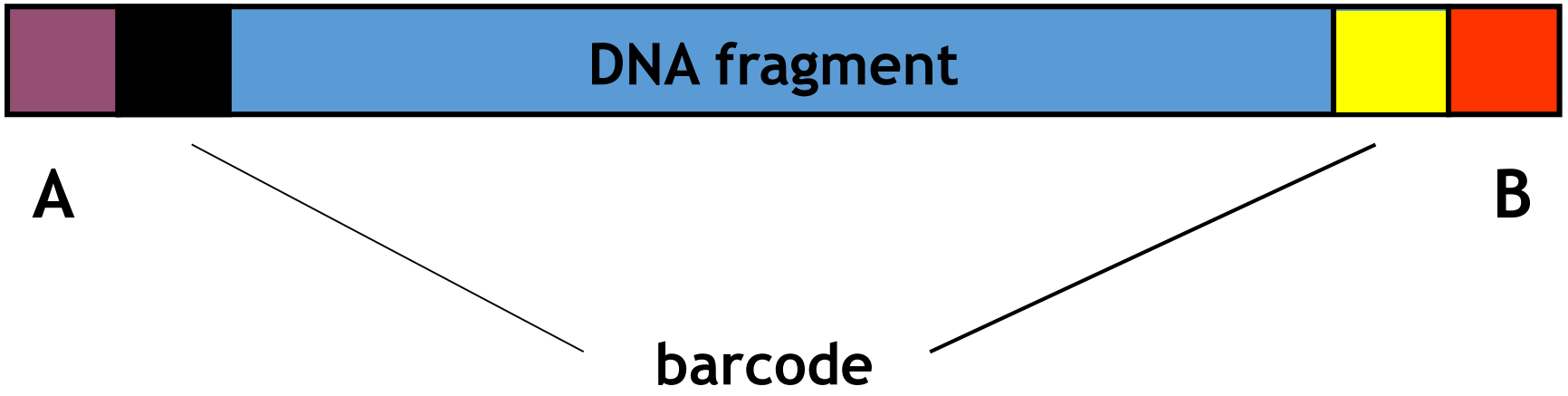
Sequencing clonally amplified DNA



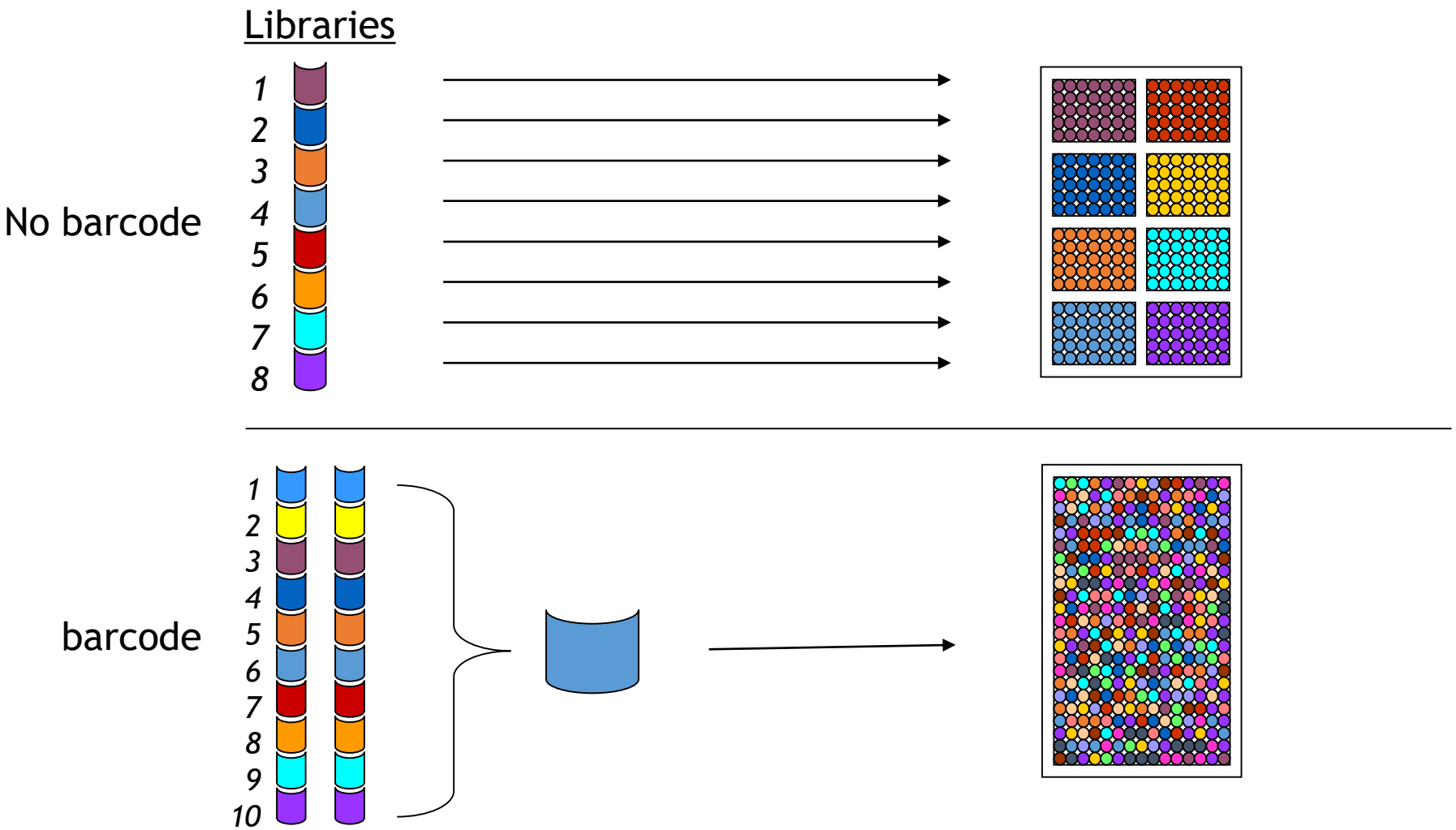
Data Analysis

Library preparation

platform specific adapter ligation / primers to (c)DNA fragments

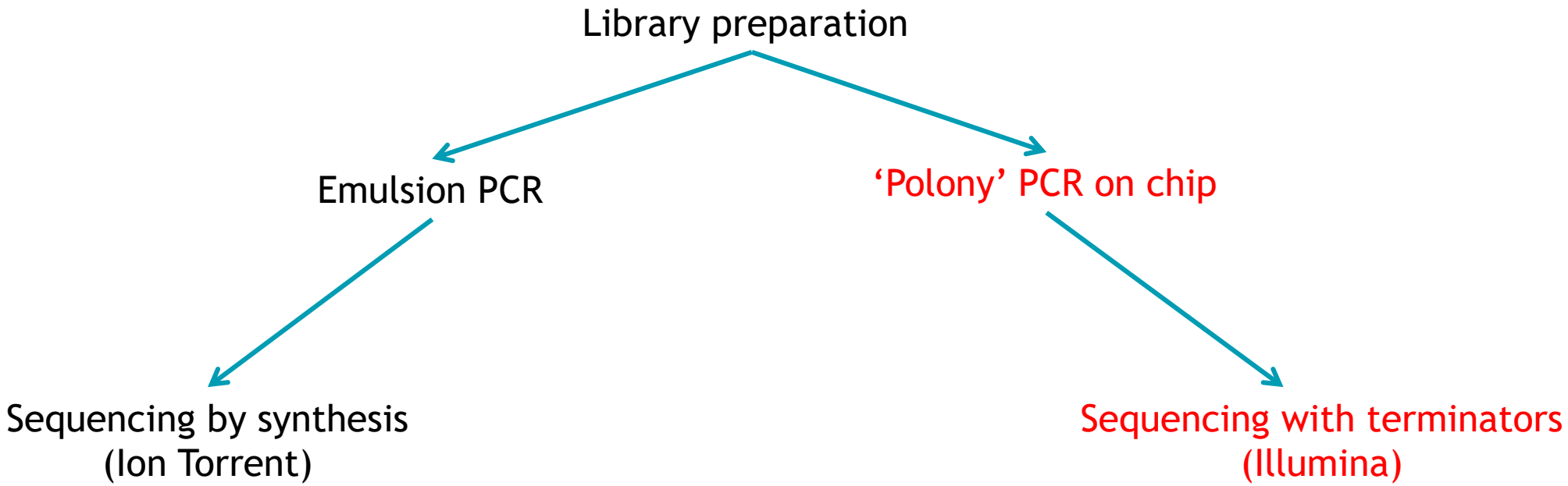


Barcoding > multiplex advantages



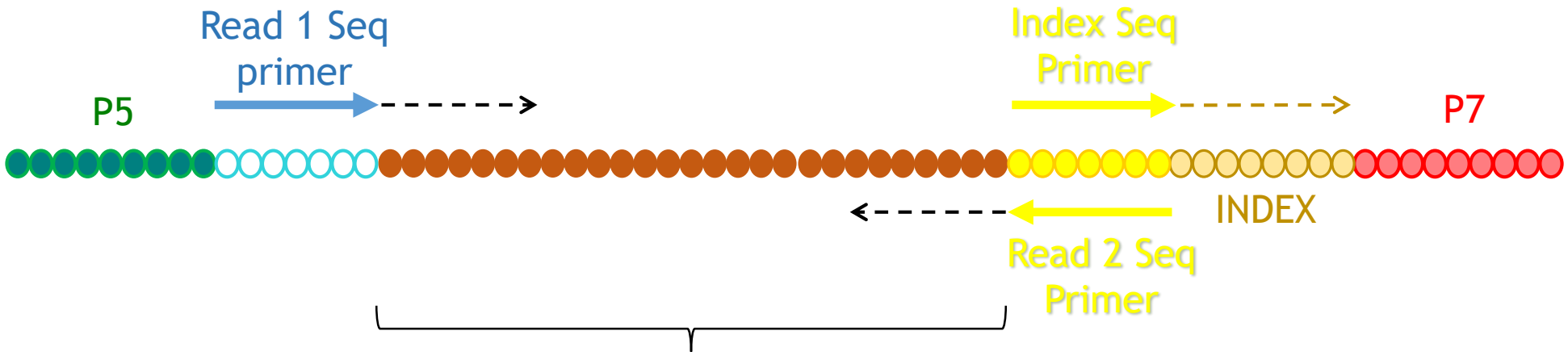
Workflow

Amplified Single Molecule Sequencing

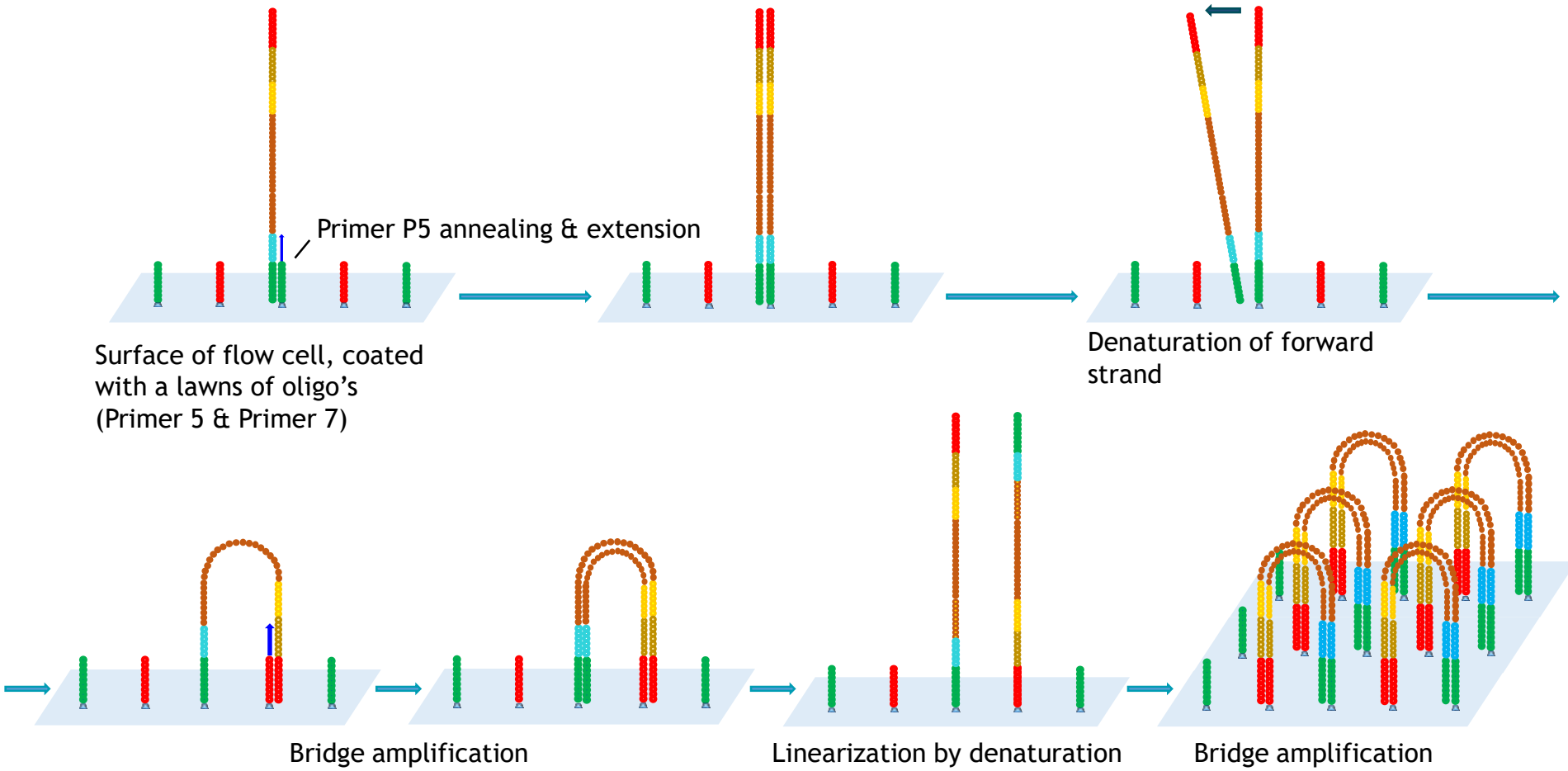


Data Analysis

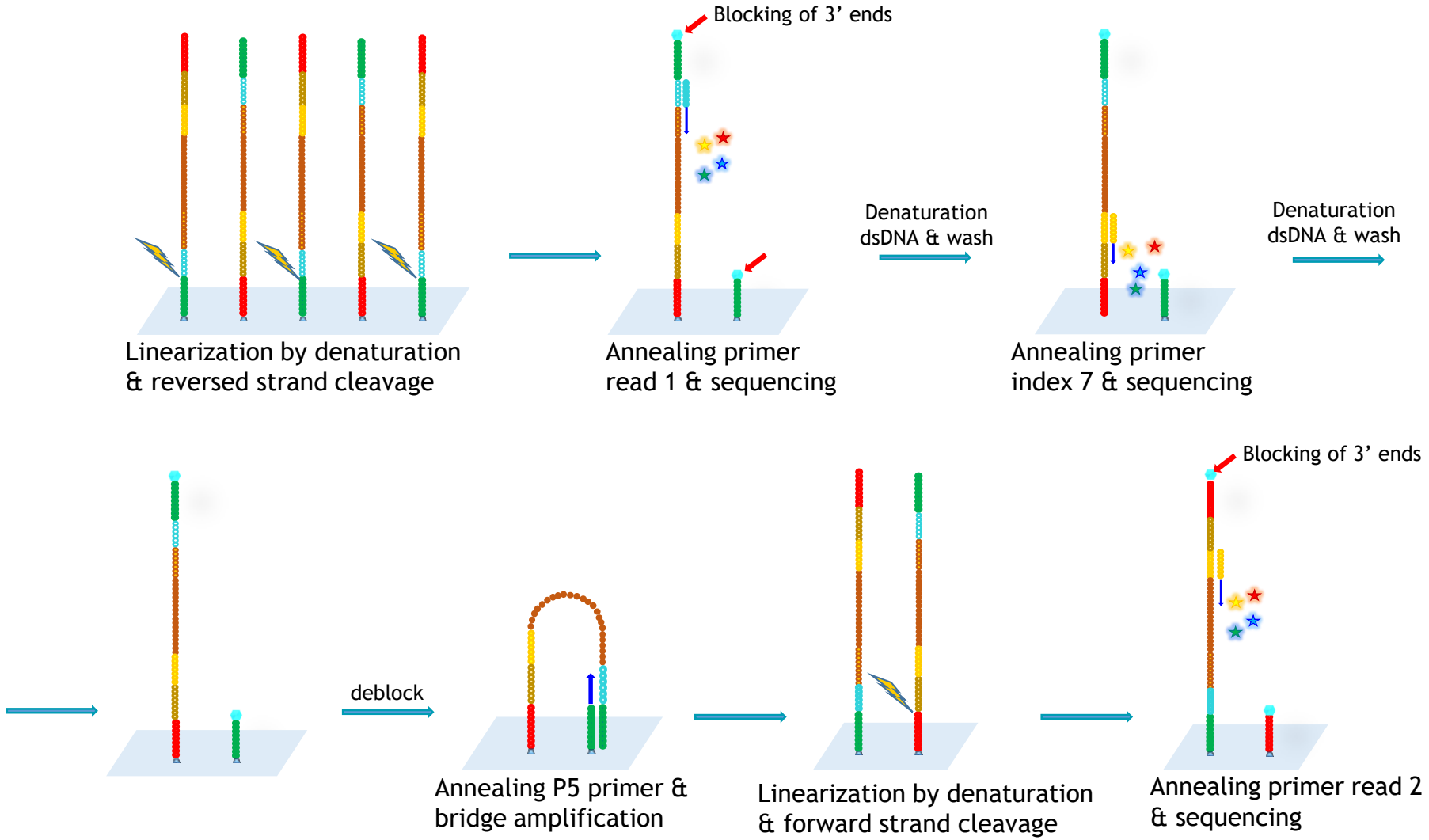
Single Index Illumina Library



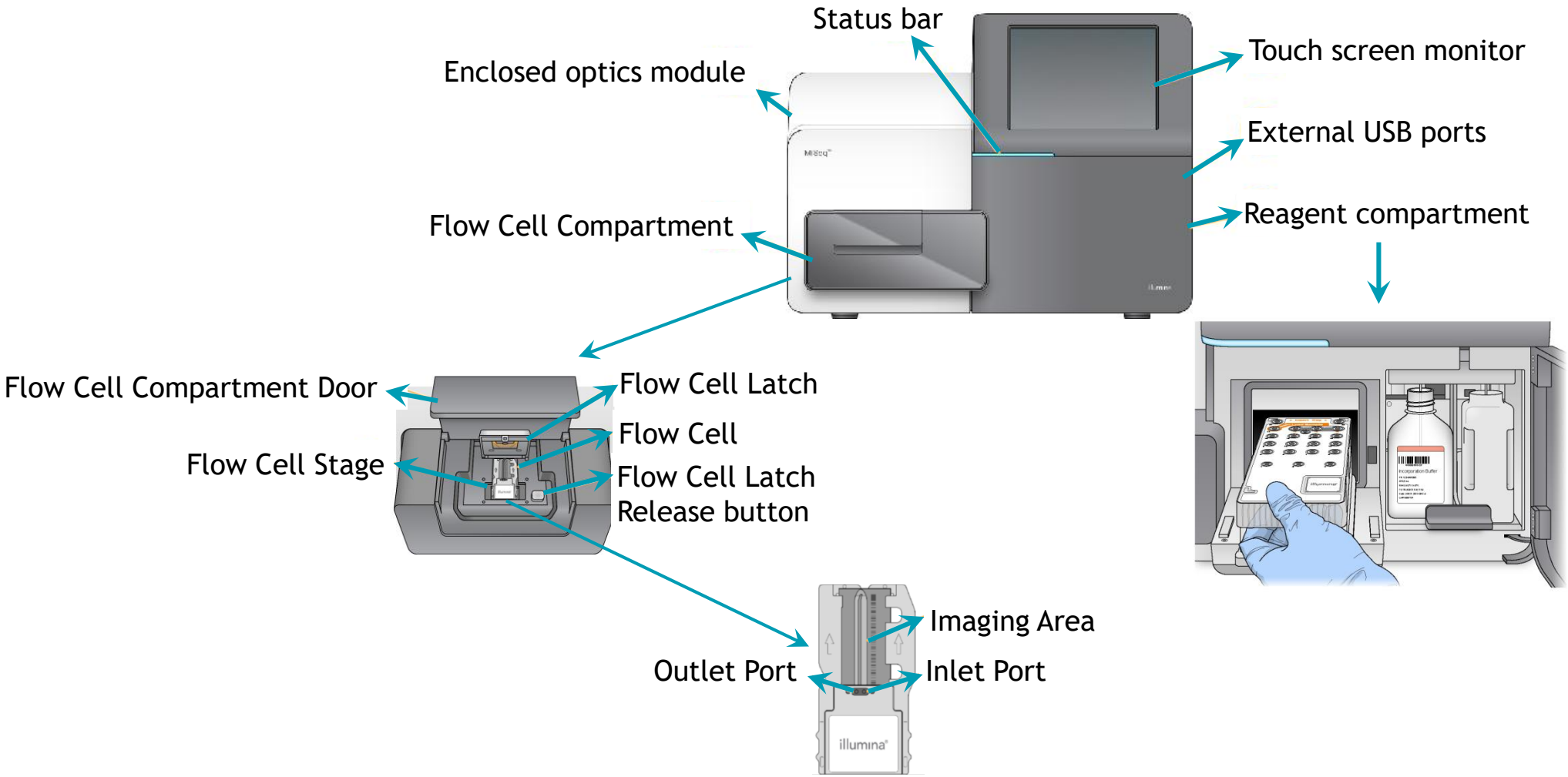
Single index Paired End Sequencing



Single index Paired End Sequencing

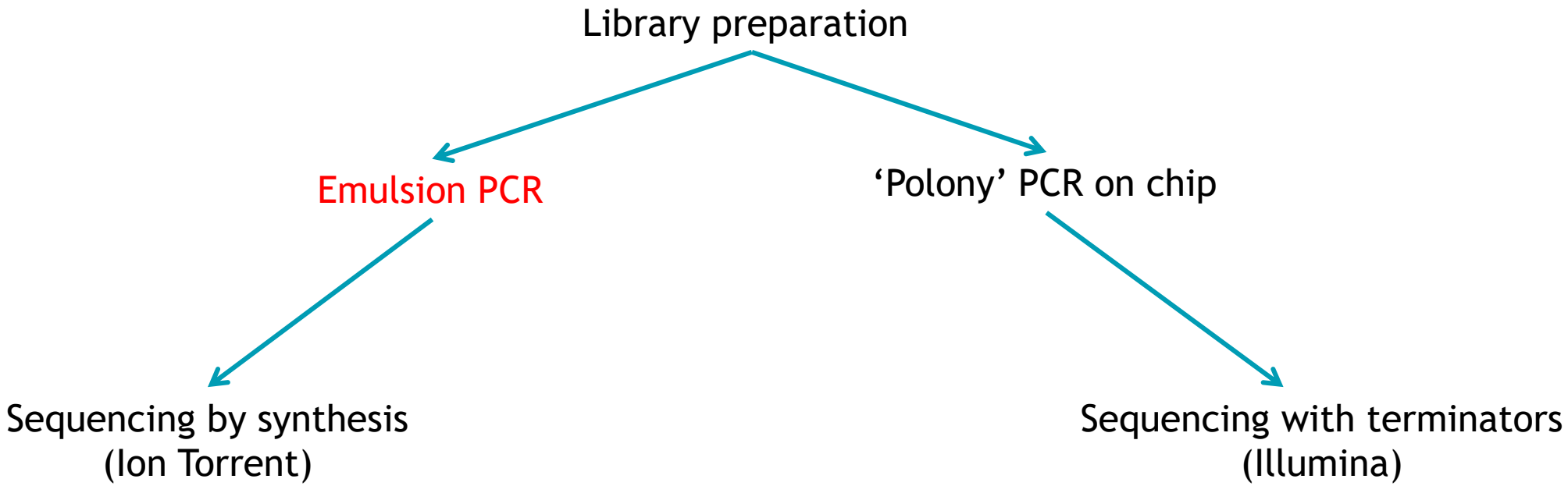


Illumina MiSeq



Workflow

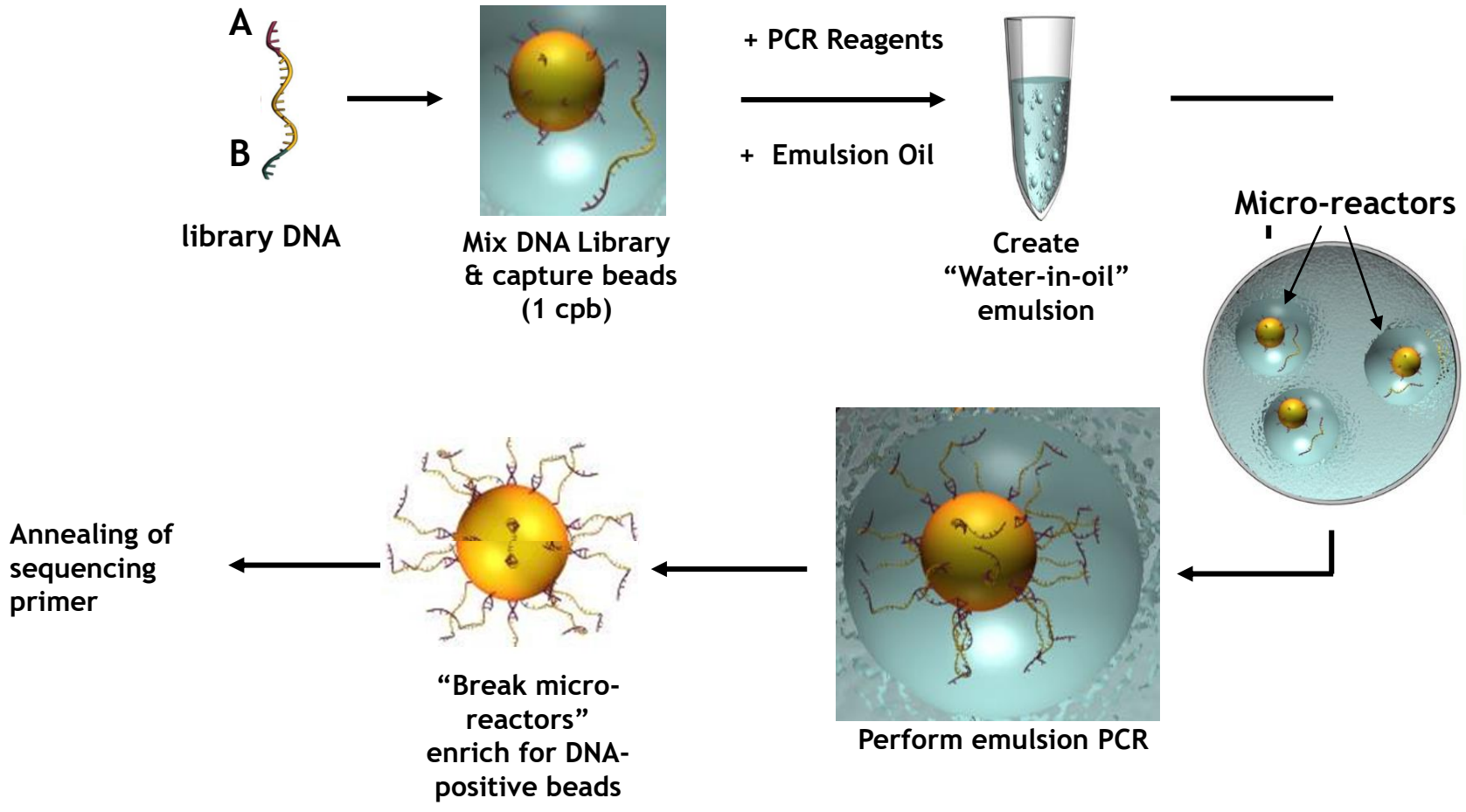
Amplified Single Molecule Sequencing



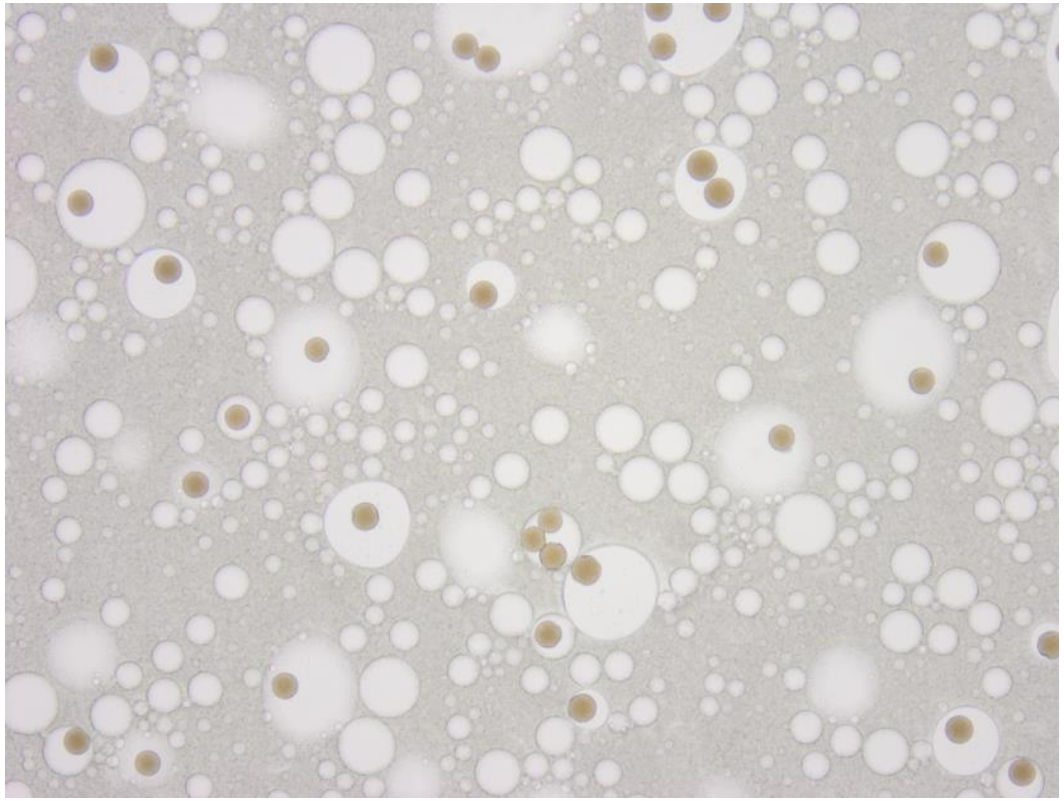
Data Analysis

Emulsion PCR

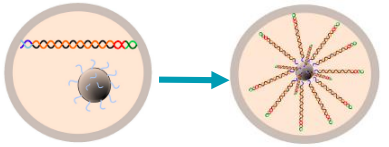
clonal amplification of DNA fragments



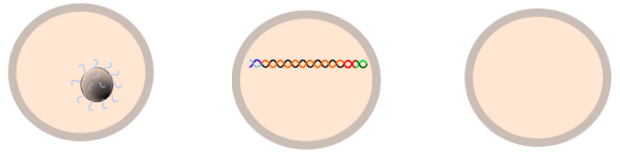
Emulsion PCR



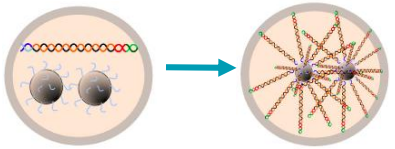
Preferably 1cpb (clonal amplification)

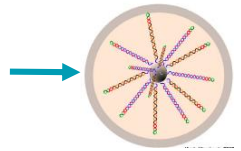


No product



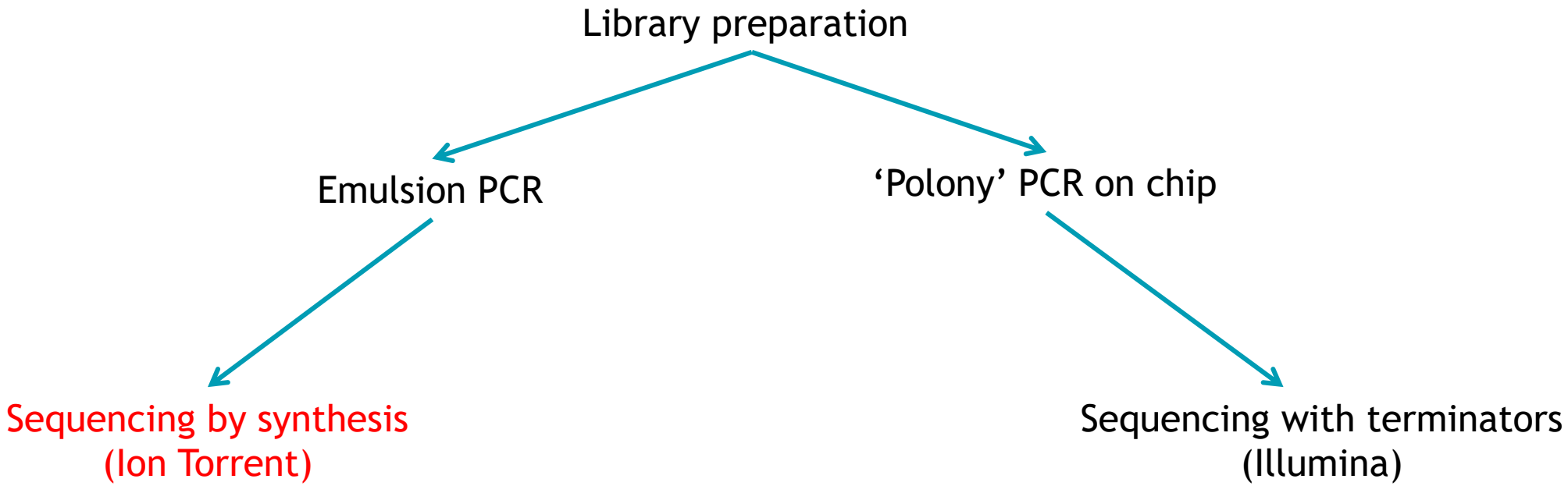
Also possible: • 1 copy, ≥ 2 beads





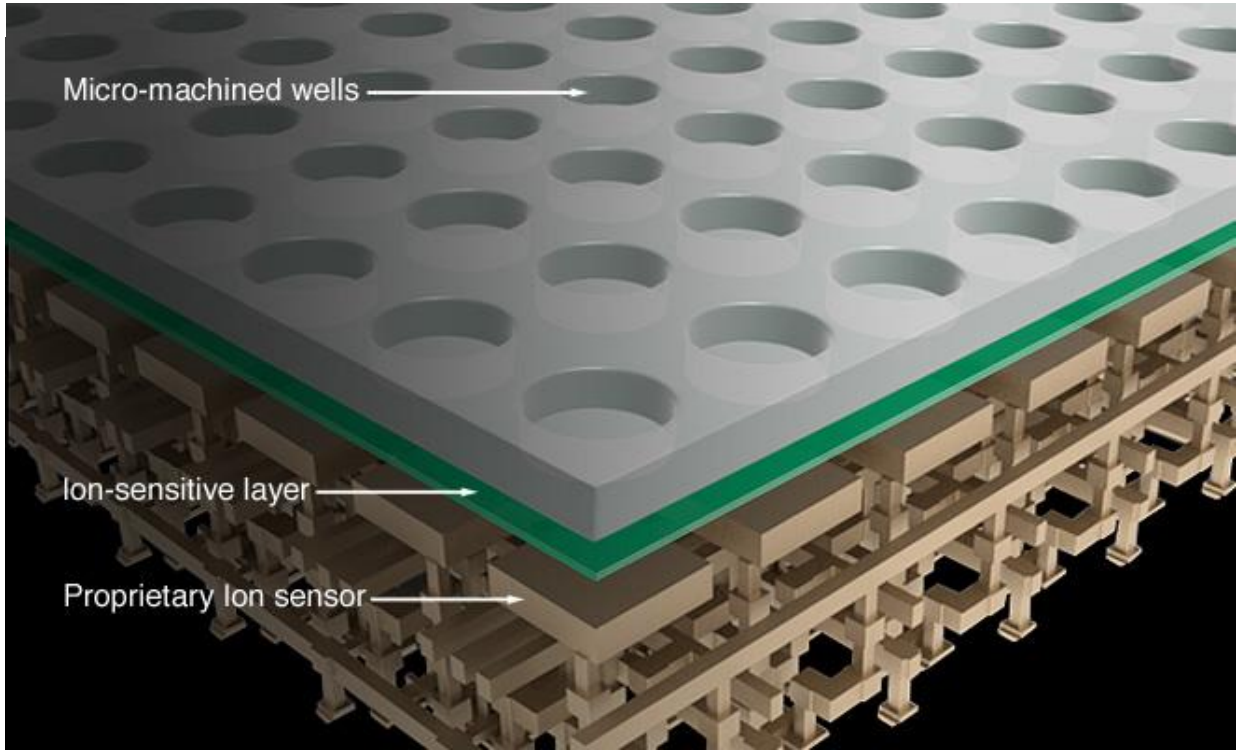
Workflow

Amplified Single Molecule Sequencing



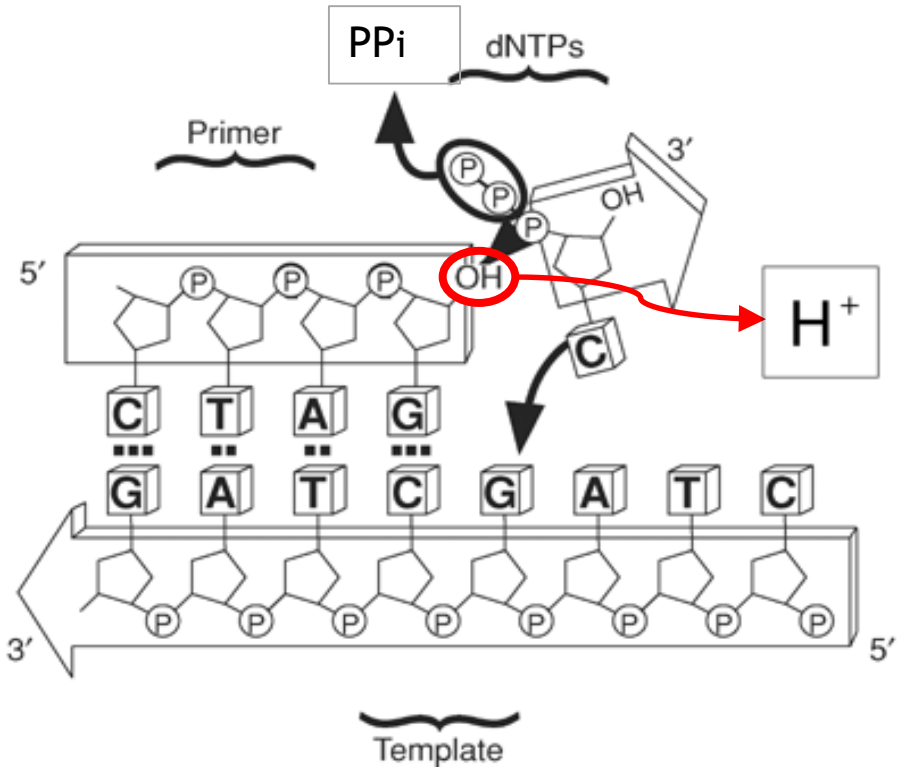
Data Analysis

Semiconductor Sequencing chip

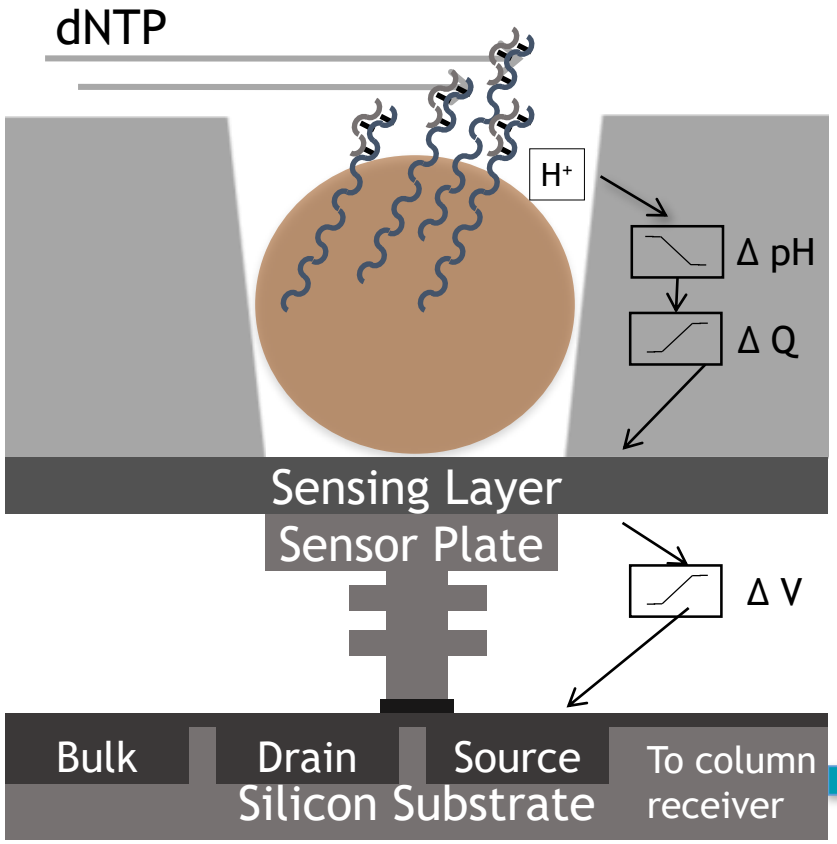


Ion Torrent

Semiconductor sequencing

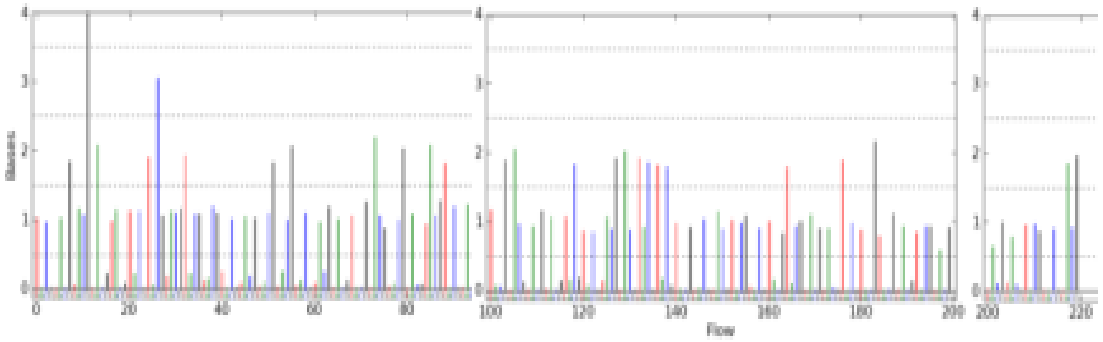


Fast (real time) Direct Detection



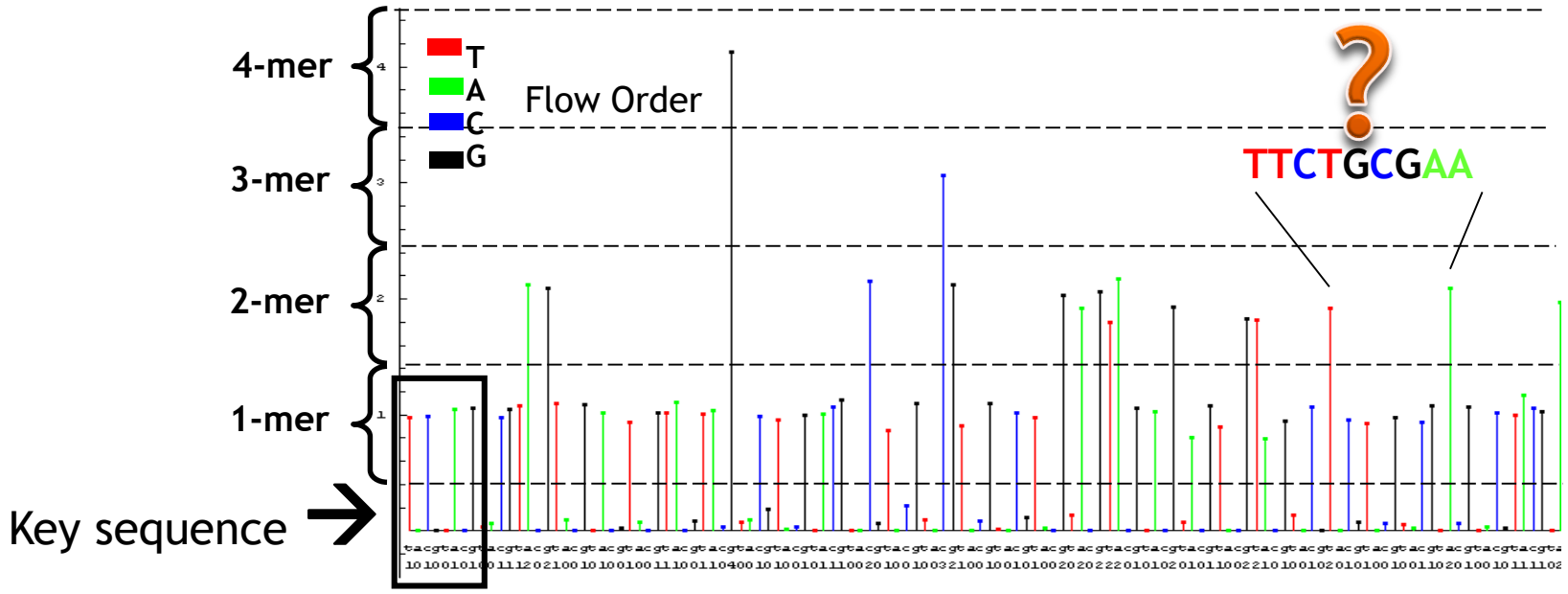
DNA → Ions → Sequence

- Nucleotides flow sequentially over Ion semiconductor chip
- One sensor per well per sequencing reaction
- **Direct** detection of natural DNA extension, **no** camera's
- Millions of sequencing reactions per chip
- Fast detection, fast cycle time, real time detection



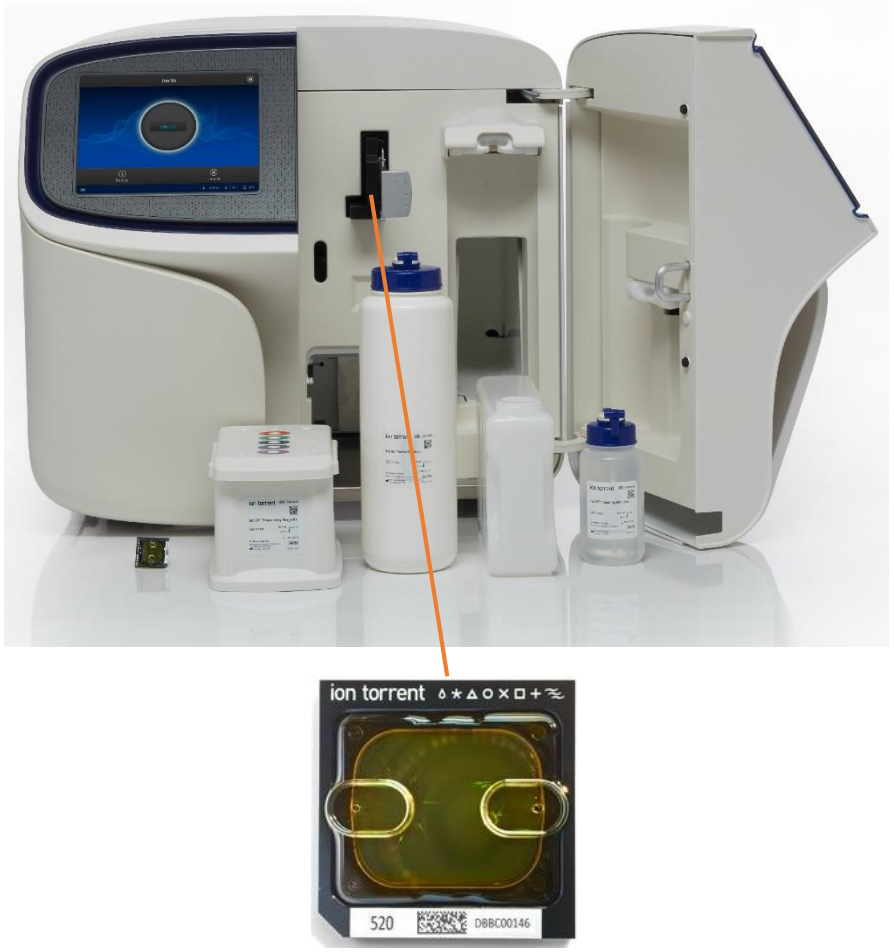
Sequencing Workflow

The signal strength is proportional to the number of nucleotides incorporated



Key: TCAG for signal calibration and normalization

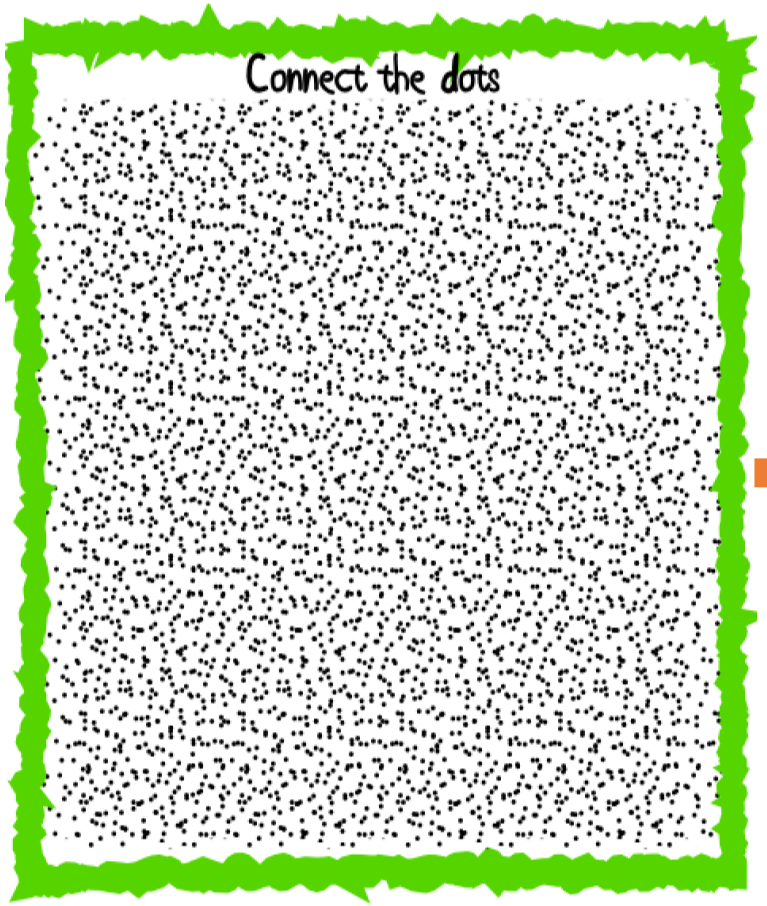
GeneStudio S5™ Sequencer



Summary NGS techniques

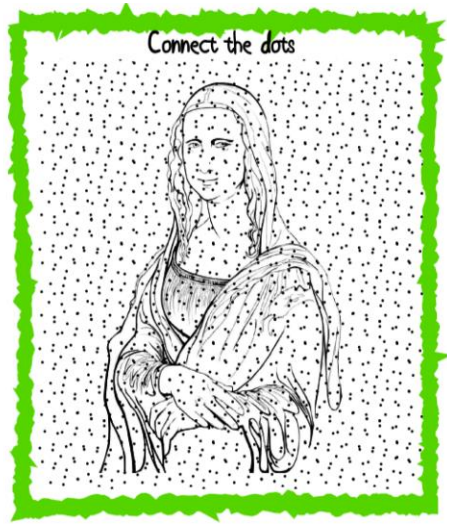
- Single Molecule: PacBio & Oxford Nanopore
  - Sequencing by terminators vs detection by Δ V
- Clonally amplified DNA: Illumina & Ion Torrent
  - Sequencing by terminators vs synthesis

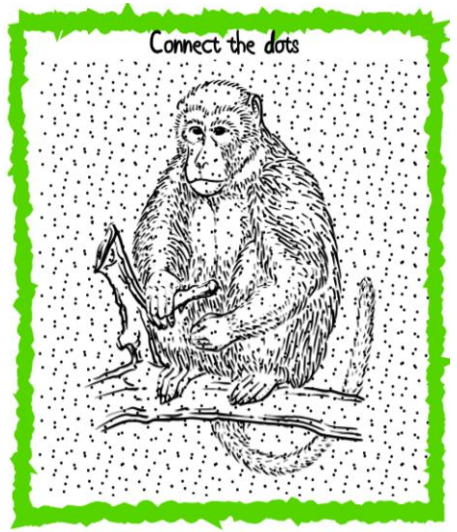
Data analysis



Impossible to assemble / align manually

Same dataset,
different pipelines





Applications Next Generation Sequencing

- Genome sequencing
- Epigenome sequencing
- Metagenome sequencing
- Transcriptome sequencing

- Single Cell sequencing

Applications Next Generation Sequencing

- Genome sequencing
- Epigenome sequencing
- Metagenome sequencing
- Transcriptome sequencing

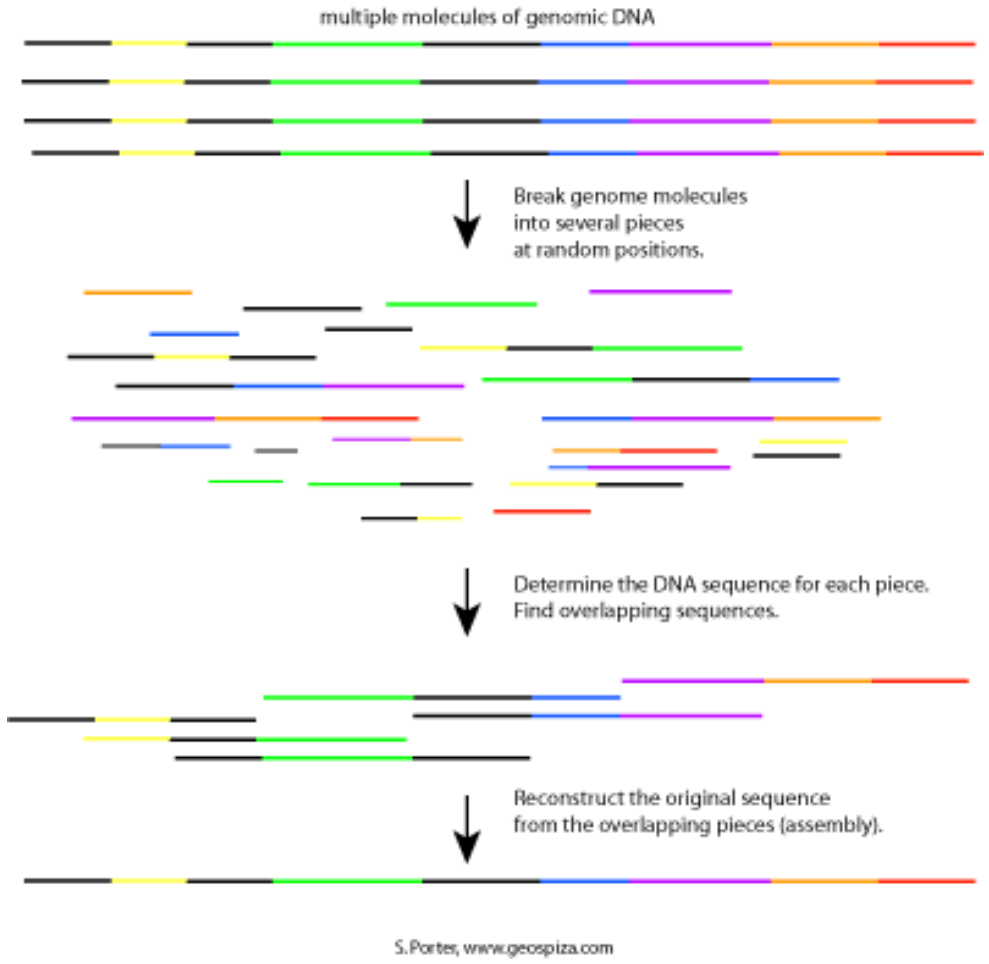
- Single Cell sequencing

Genome sequencing:

- De novo sequencing
- Whole Genome / Exome Sequencing
- Targeted resequencing
  - Amplicons
  - Hybridisation captures

Genome sequencing:

- De novo sequencing
- WGS



Genome sequencing:

- Whole Exome Sequencing
- Targeted resequencing

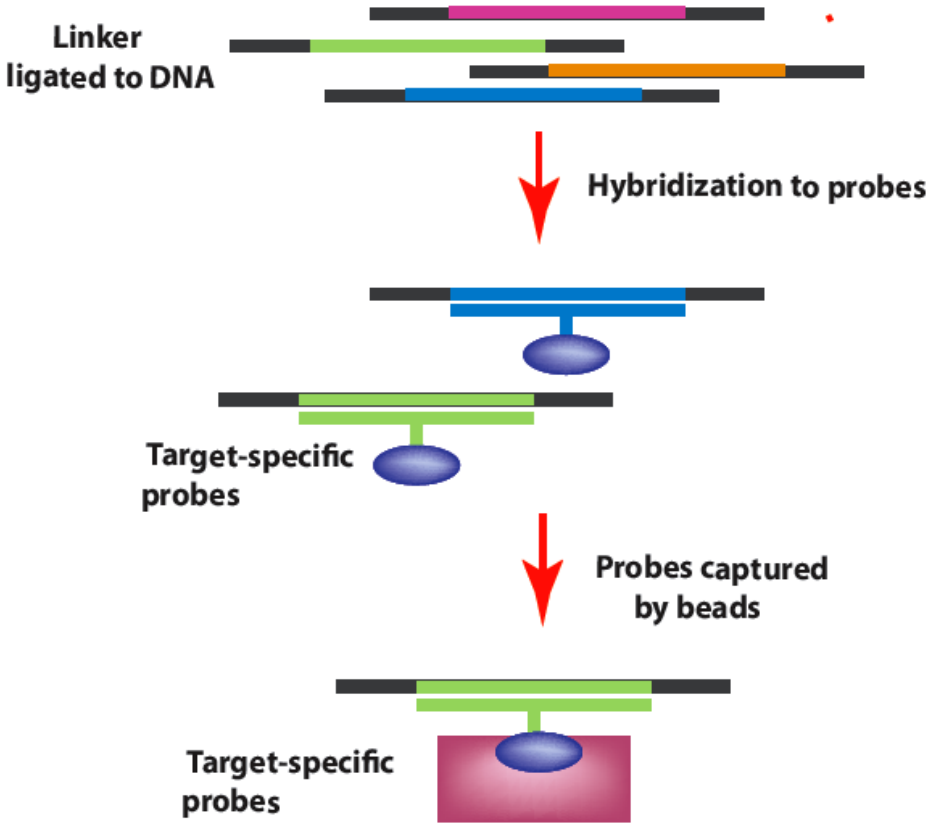
✓ selected gene panels,

✓ mutation analysis (SNPs, low frequent mutations etc),

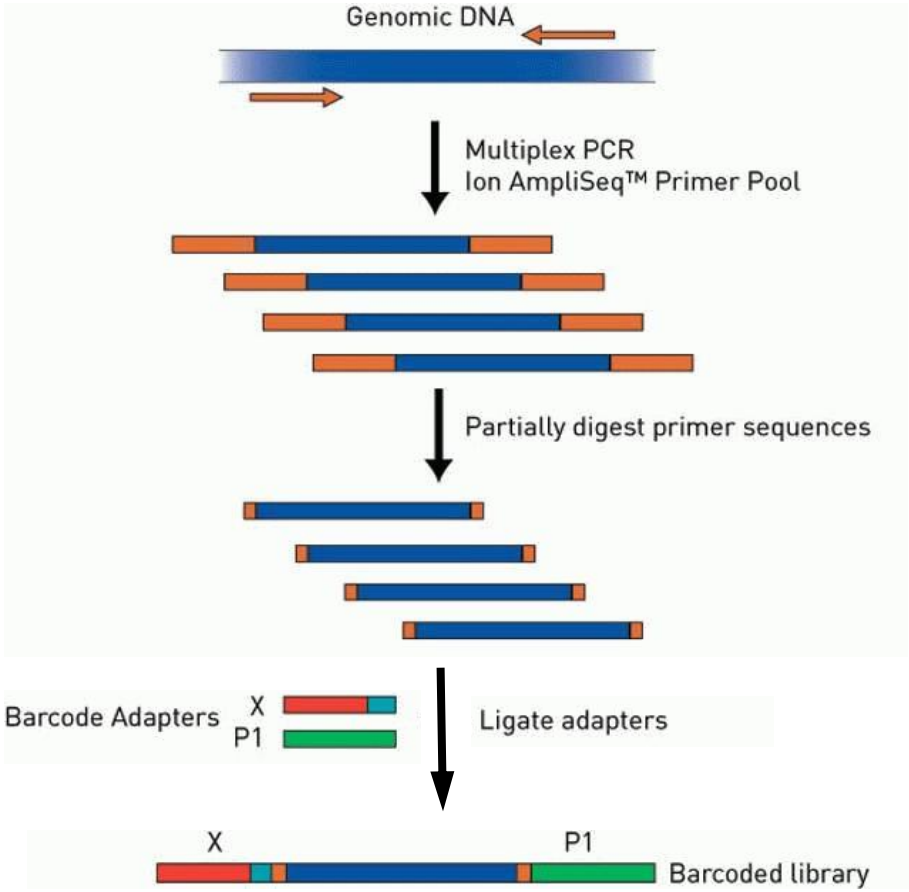
✓ structural variation (deletions, insertions, inversions, CNVs)

✓ mitochondrial sequencing

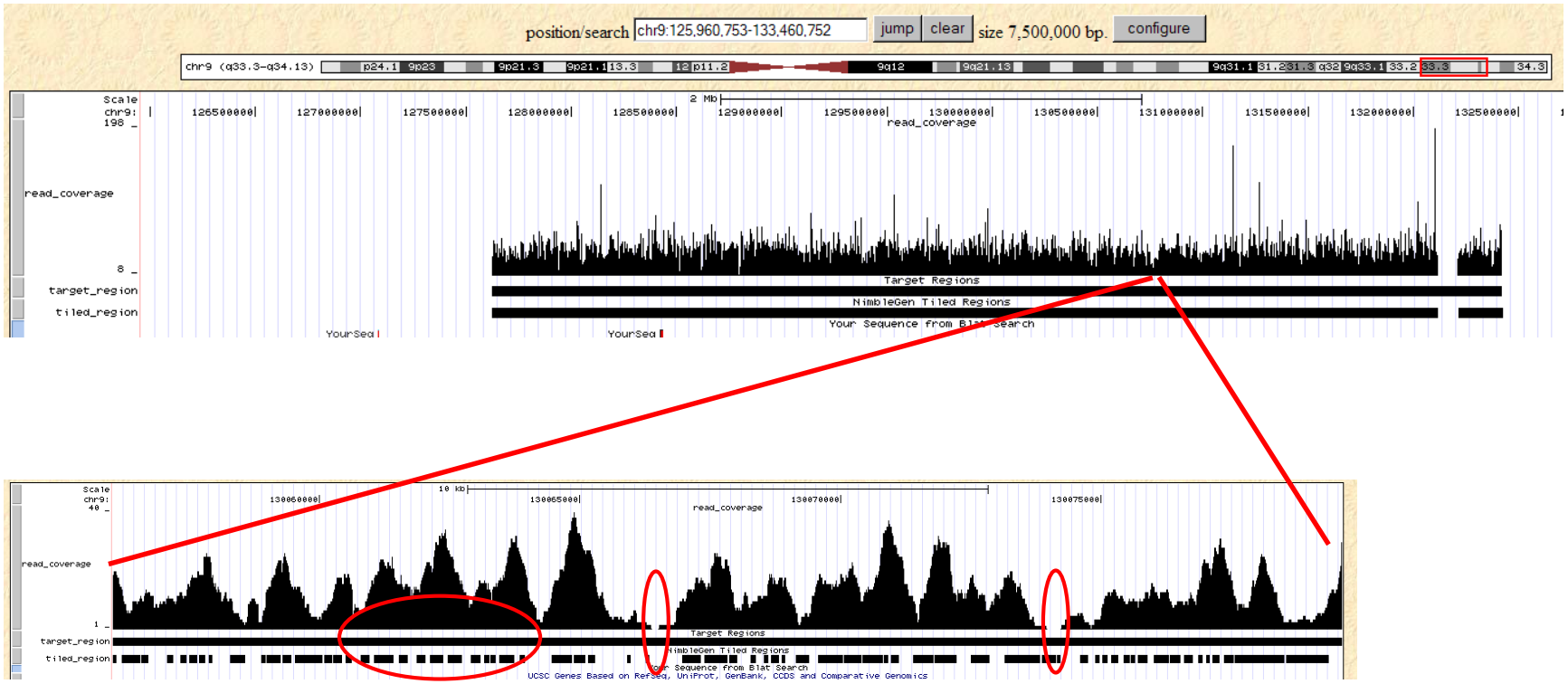
Hybridization target capture



(Multiplex) Amplicon sequencing



Read Coverage: Captures are not uniform



Summary genome sequencing

- De novo sequencing
- Whole Genome
- Exome Sequencing
- Targeted resequencing

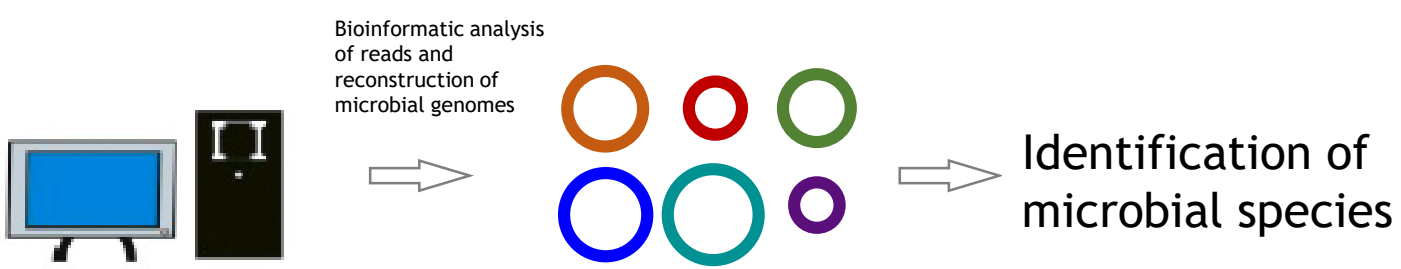
Pause

Applications Next Generation Sequencing

- Genome sequencing
- Epigenome sequencing
- Metagenome sequencing
- Transcriptome sequencing

- Single Cell sequencing

Metagenome Sequencing

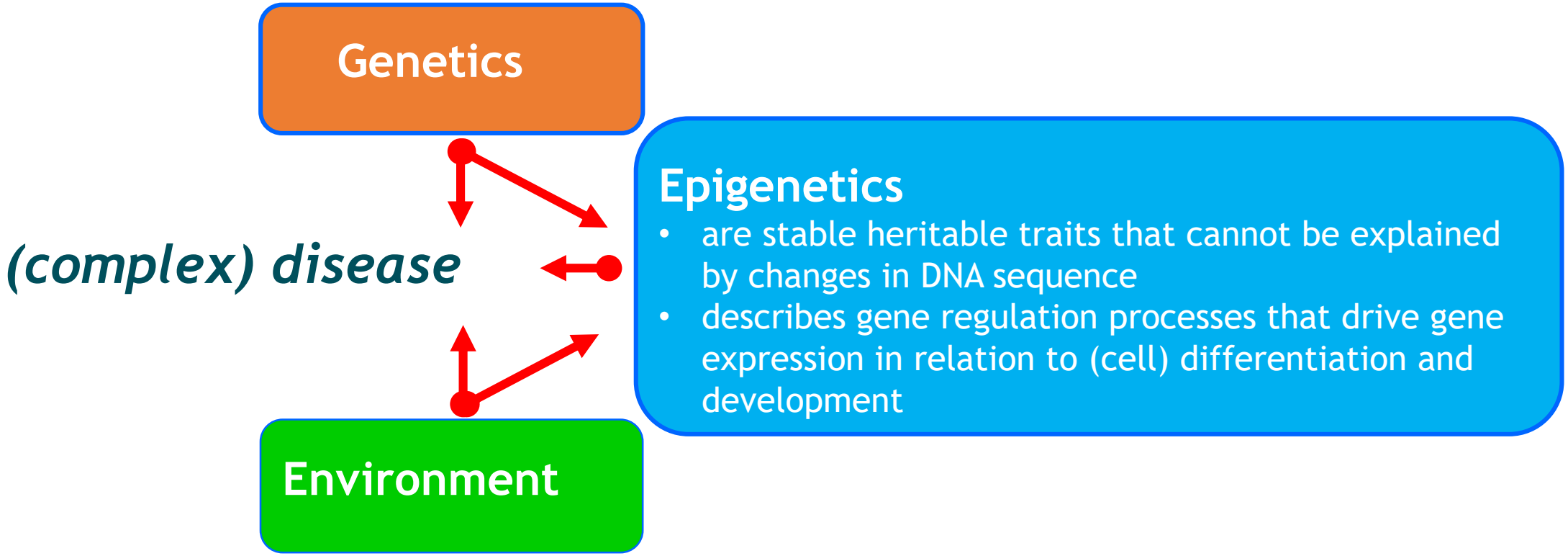


Applications Next Generation Sequencing

- Genome sequencing
- Epigenome sequencing
- Metagenome sequencing
- Transcriptome sequencing

- Single Cell sequencing

Epigenome sequencing



Epigenetics: bridge between "nature" and "nurture"

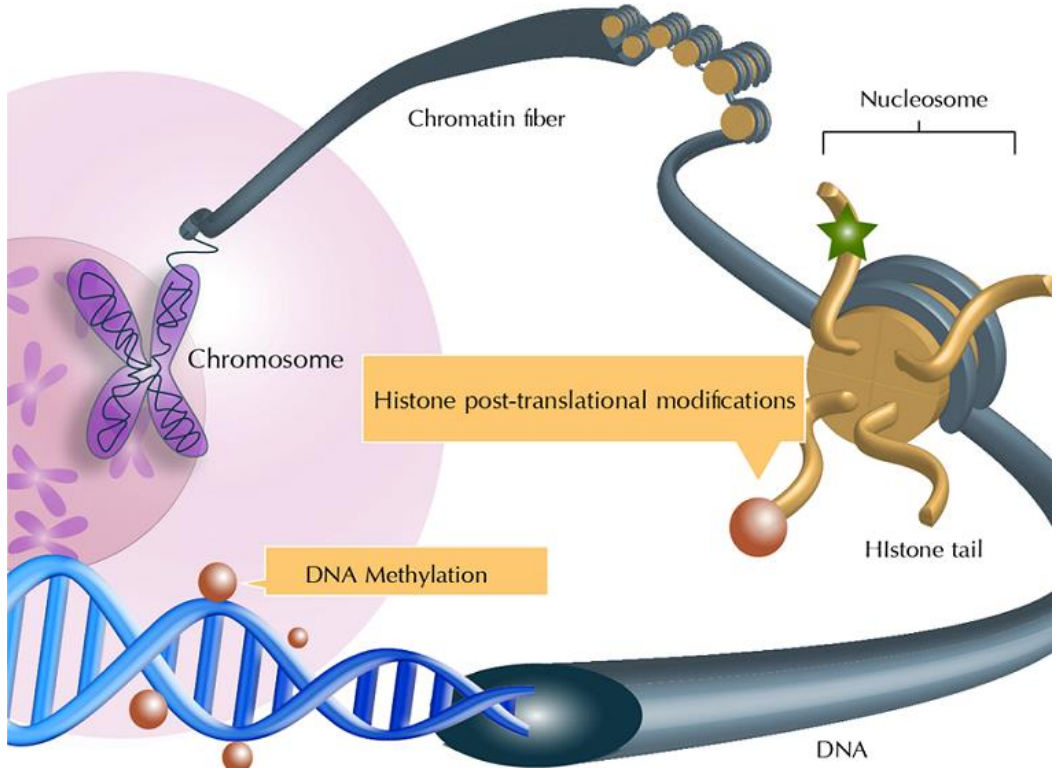
Epigenetics is involved in:

- (in)activation of genes during development
- X-chromosome inactivation in females
- Genomic imprinting (parent specific active alleles)

- (in)activation of genes under influence of the genetics / environment

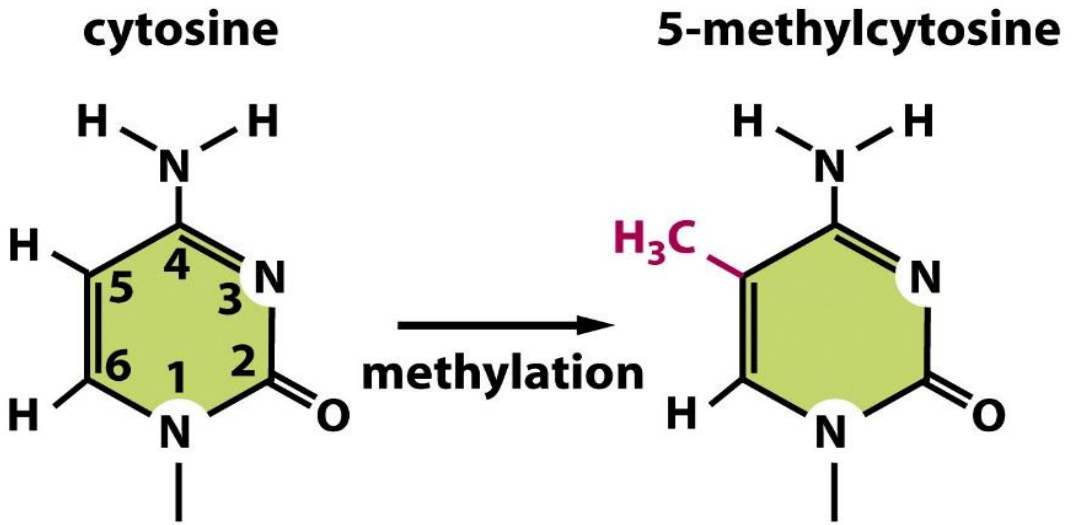
NGS approaches Epigenetics

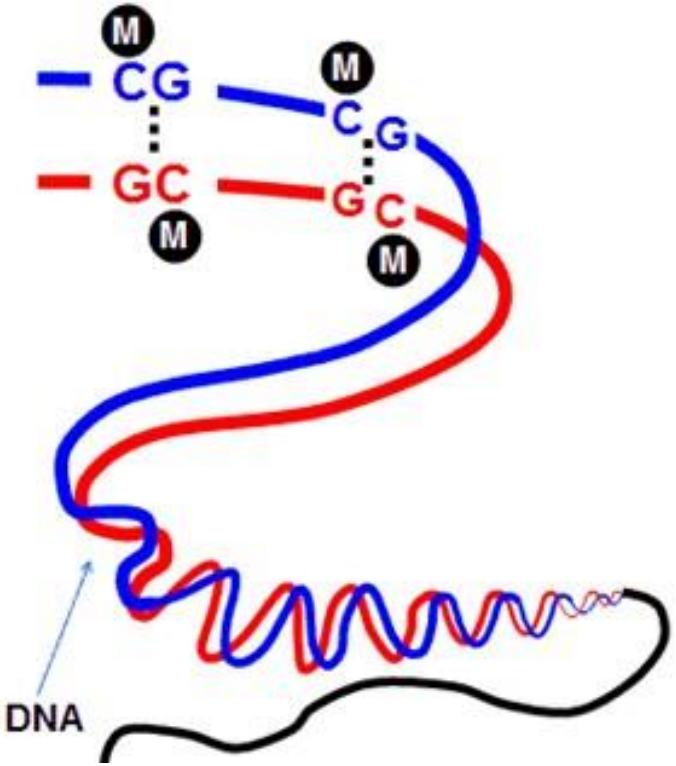
- DNA-methylation
- Histon modifications
- Chromatin remodelling



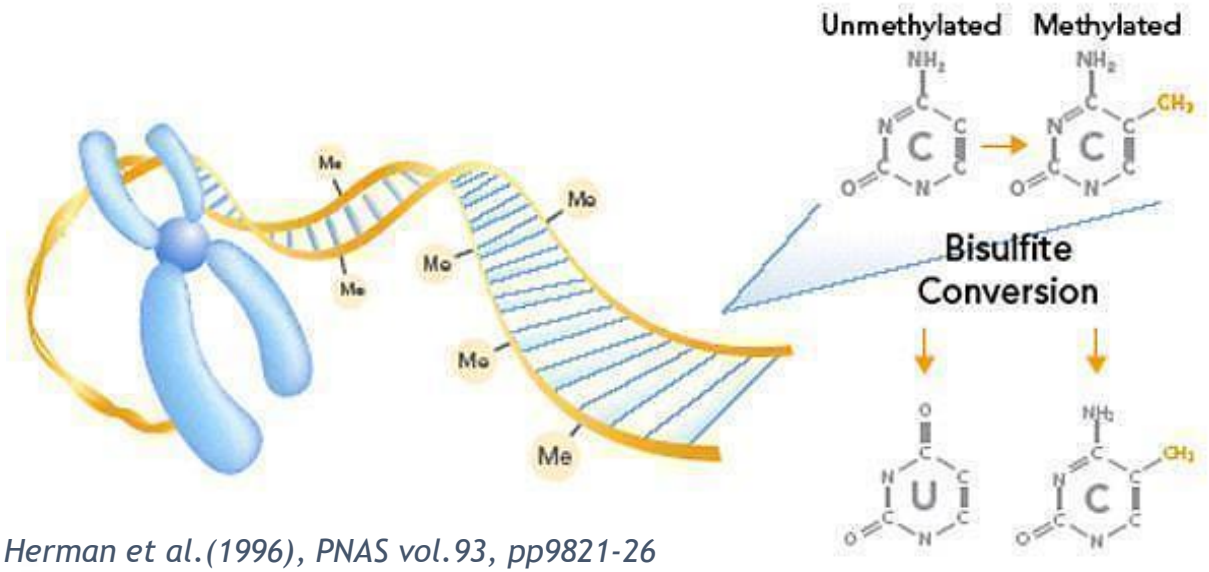
• DNA-methylation

- Histon modifications
- Chromatin remodelling

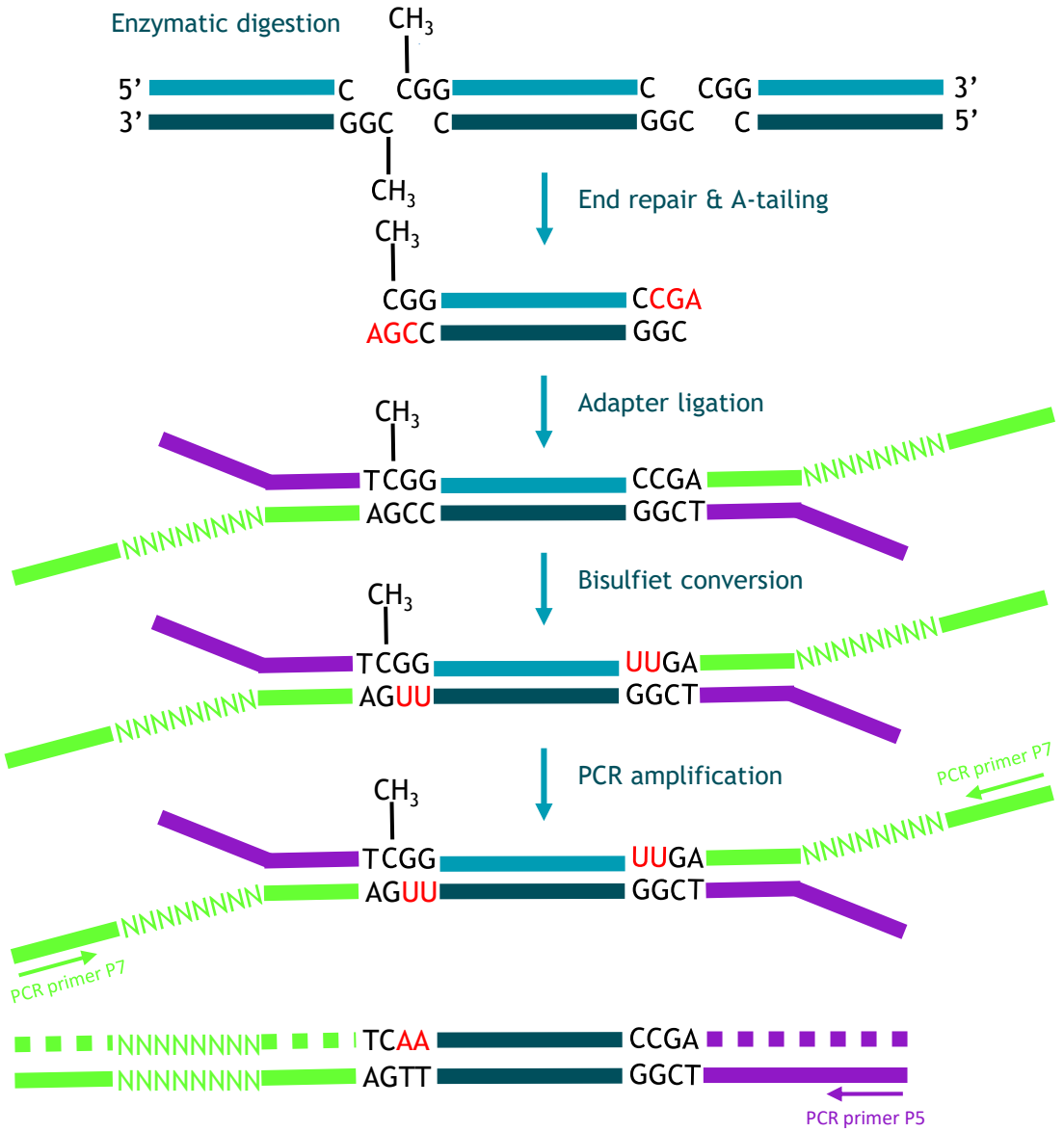




Bisulfite Conversion of DNA



Herman et al.(1996), PNAS vol.93, pp9821-26

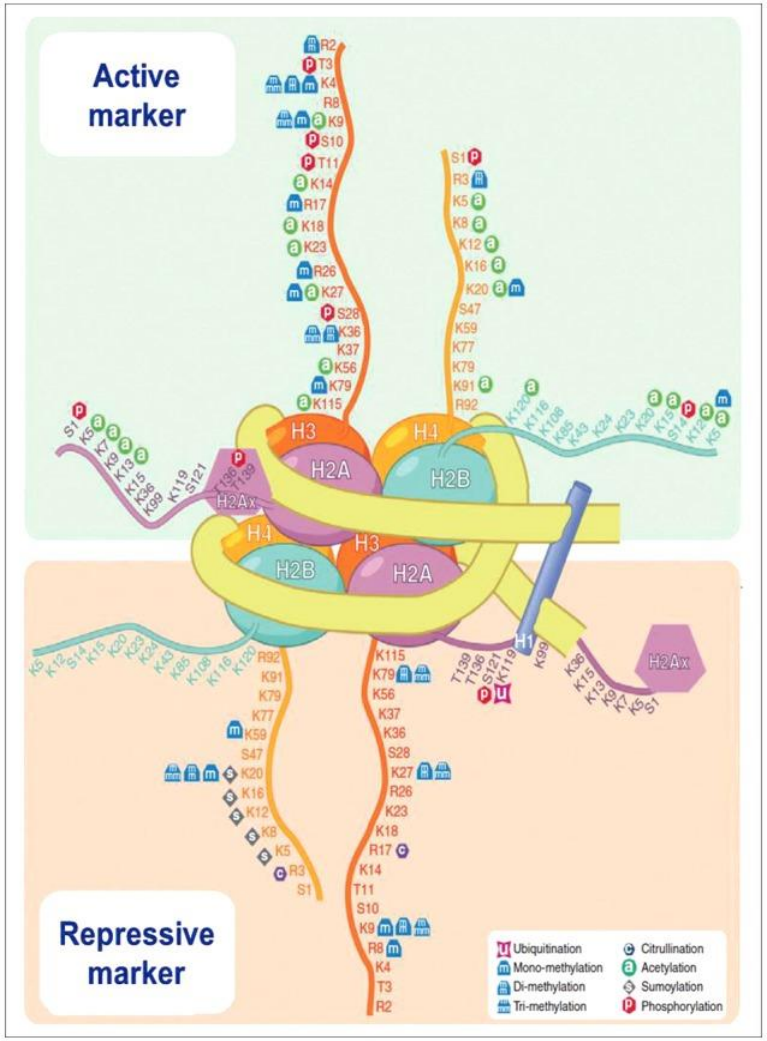


- DNA-methylation
- **Histon modifications**
- Chromatin remodelling

Most common Histone modifications:

- **Acetylation**
- **Phosphorylation**
- **Methylation**
- **Ubiquitylation**

ChIP seq



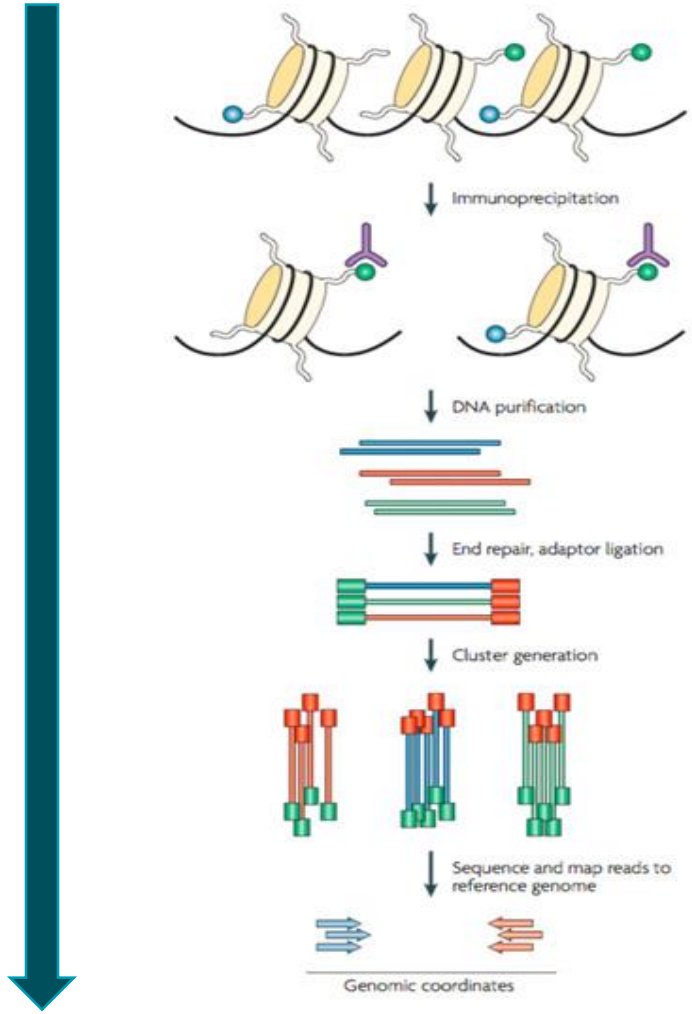
Chromatin Immuno Precipitation (ChIP) Sequencing

Crosslinking

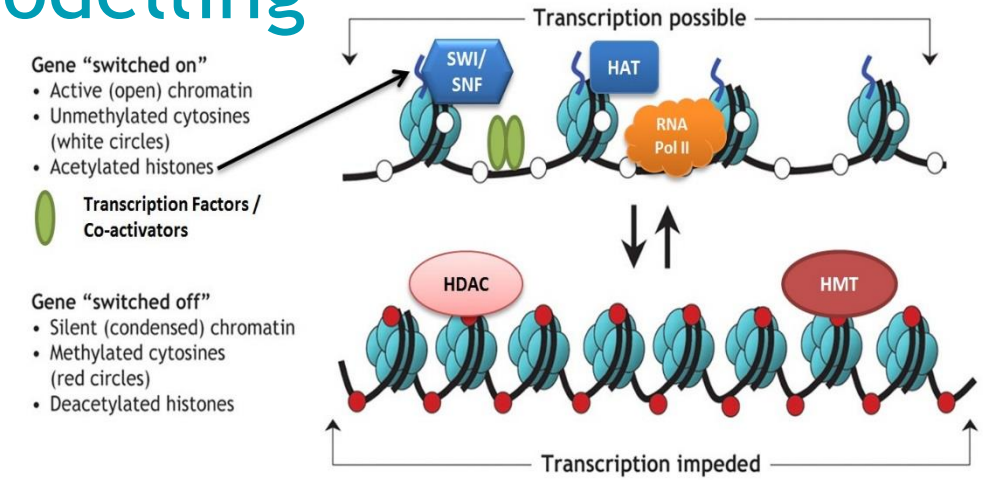
Antibody against e.g. H3k27Ac

Sequencing

Mapping



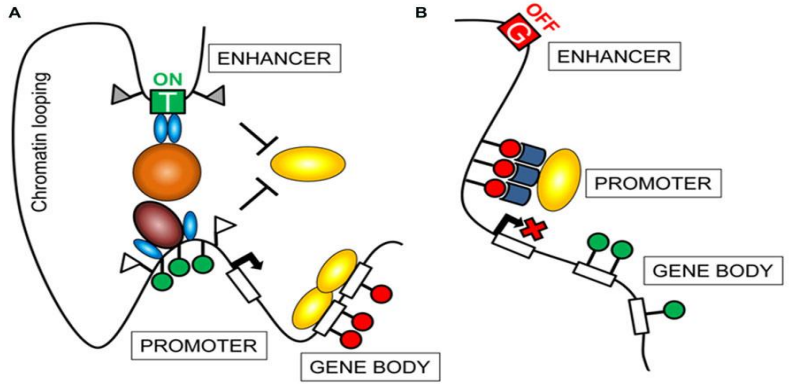
- DNA-methylation
- Histon modifications
- **Chromatin remodelling**



"Open"

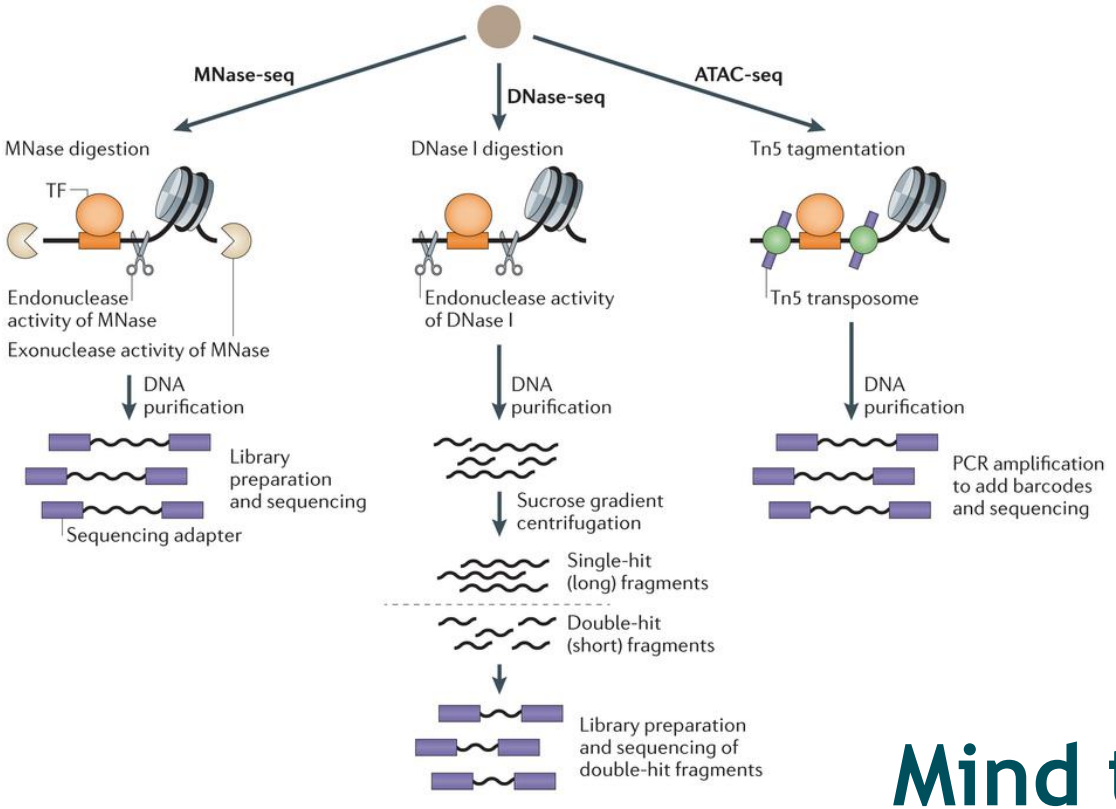
"Closed"

Chromatin Accessibility

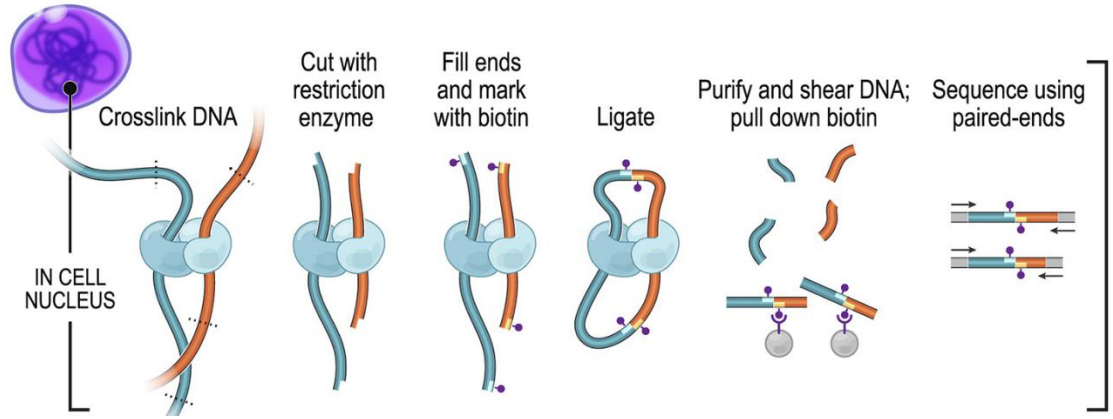


Chromatin Interaction

Chromatin Accessibility



Chromatin interactions: Hi-C



Mind the tissue but cell cycle as well !!

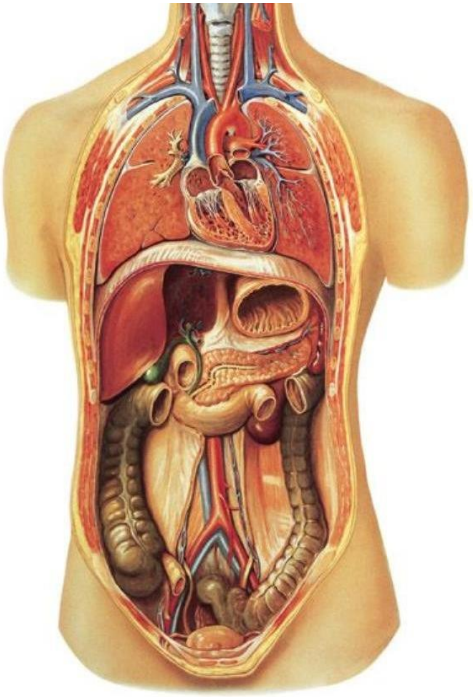
Summary Epigenetics approaches and techniques

- DNA-methylation
- Histon modifications
- Chromatin remodelling

Important confounding factors / bias in Epigenetic studies

Biological:
- Age
- Gender
- Tissue (cell mixture)

Technical:
- Batch
- Passage (cell culture)
- Medium (cell culture)

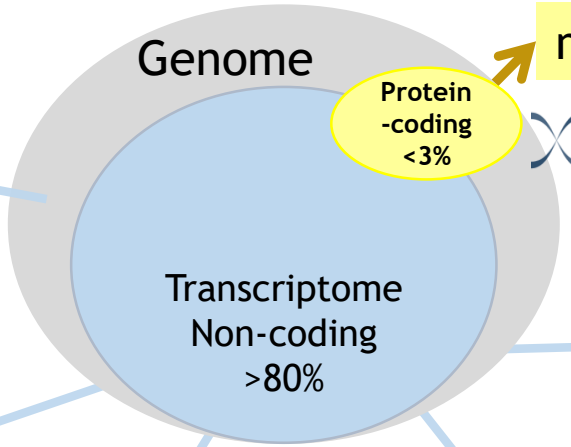


Applications Next Generation Sequencing

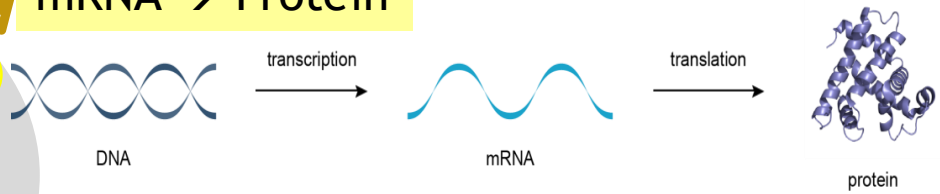
- Genome sequencing
- Epigenome sequencing
- Metagenome sequencing
- Transcriptome sequencing

- Single Cell sequencing

Transcriptome sequencing



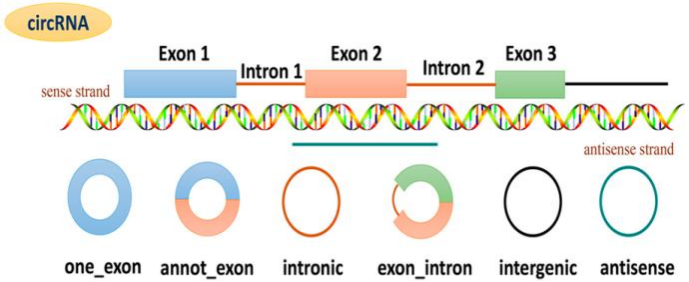
mRNA → Protein



Repetitive DNA

Interspread repeats (SINEs,LINEs)
Processed pseudogenes
Simple sequence repeats
Segmental duplication
Blocks of tandem repeats

Circular RNA (circRNAs) (100- >4000 nt)



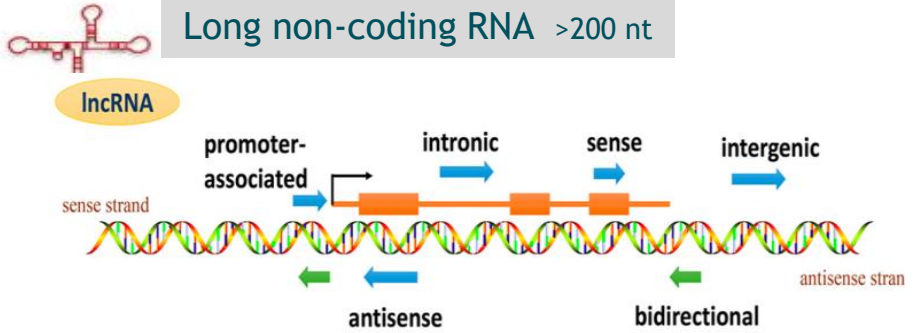
Transcriptional RNA

tRNA (70-100nt) → *Translation*
rRNA (120-4700nt) → *Translation*

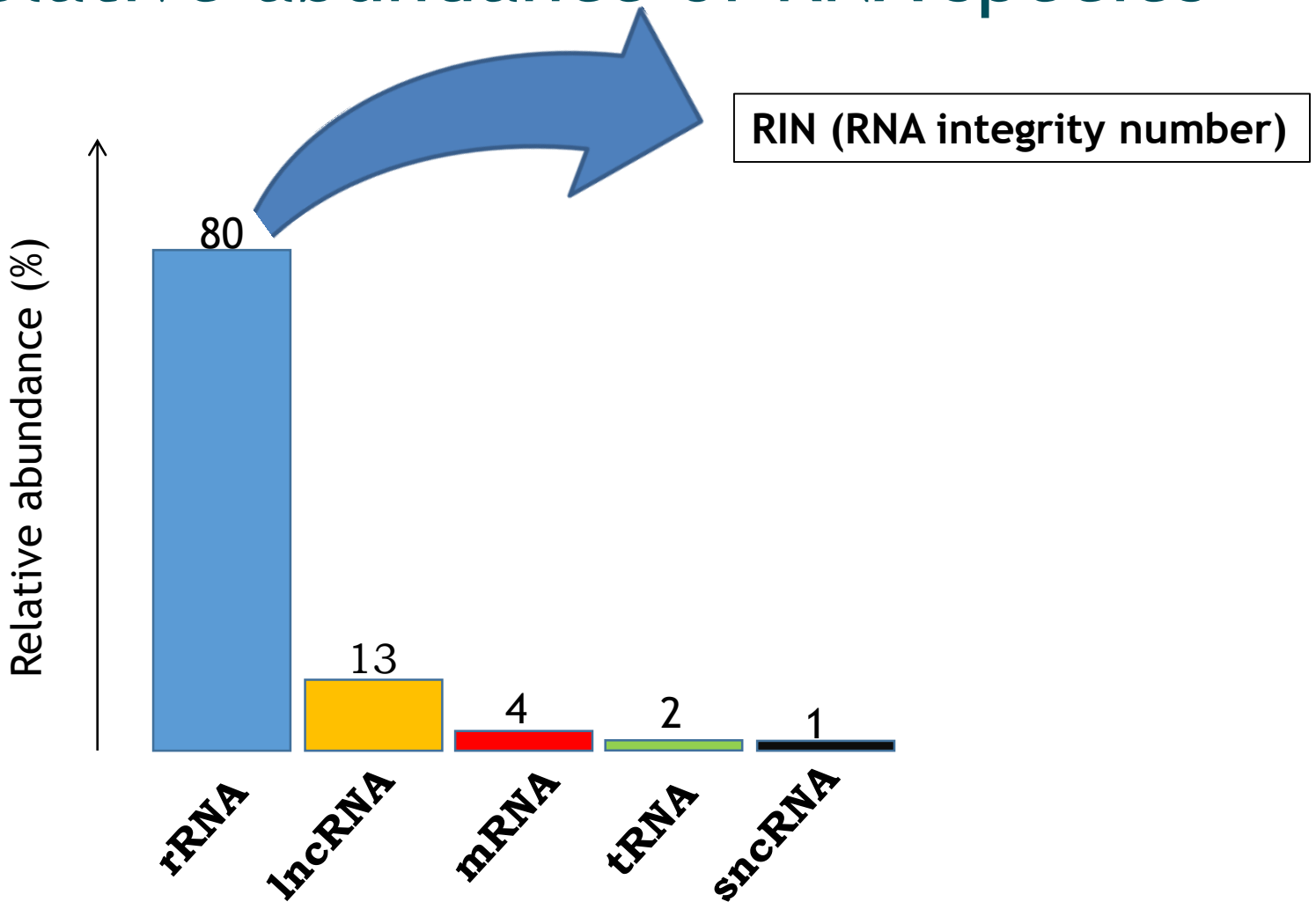
Small non-coding RNA <300 nt

miRNA (18-24 nt) → *Post-transcriptional gene regulation*
piRNA (26-31 nt → *Germ line transposon silencing*
snRNA (¬150nt) → pre-mRNA splicing
snoRNA (60-300 nt) → *RNA modification, rRNA processing*
scaRNA → *RNA modification*
Y RNA → *DNA replication, small RNA maturation*
vRNA (88-98nt)
siRNA *(21-23nt)*

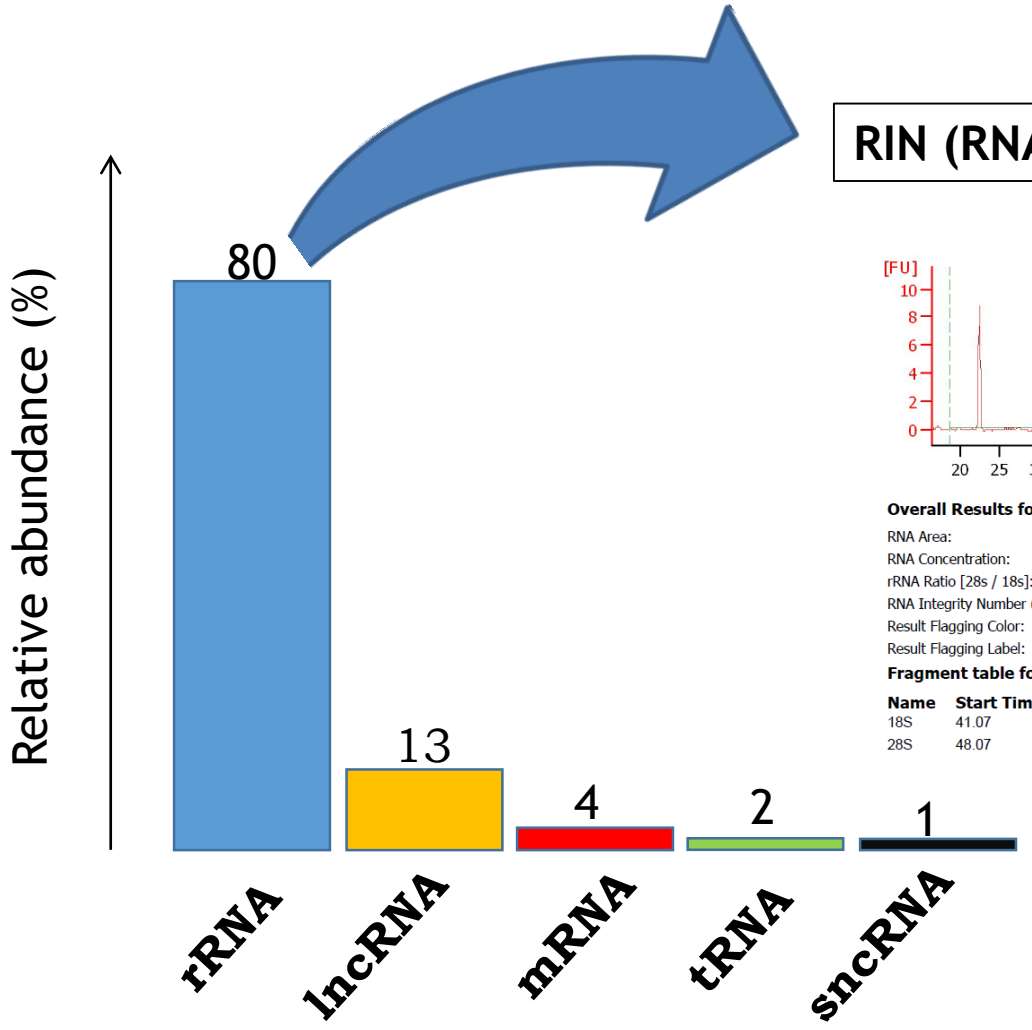
Long non-coding RNA >200 nt



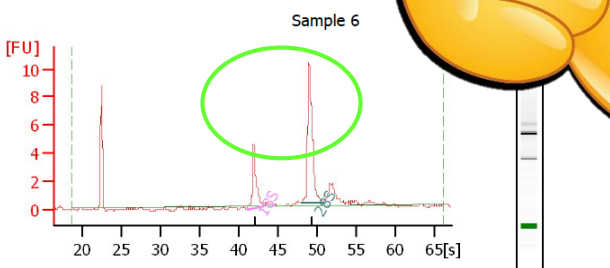
Relative abundance of RNA species



Relative abundance of RNA species



RIN (RNA integrity

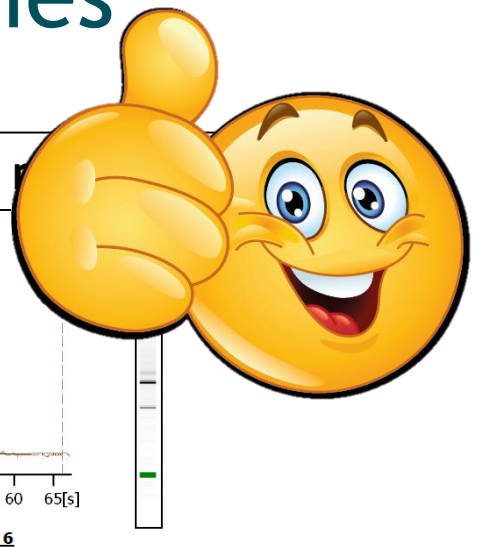


Overall Results for sample 6 : **Sample 6**

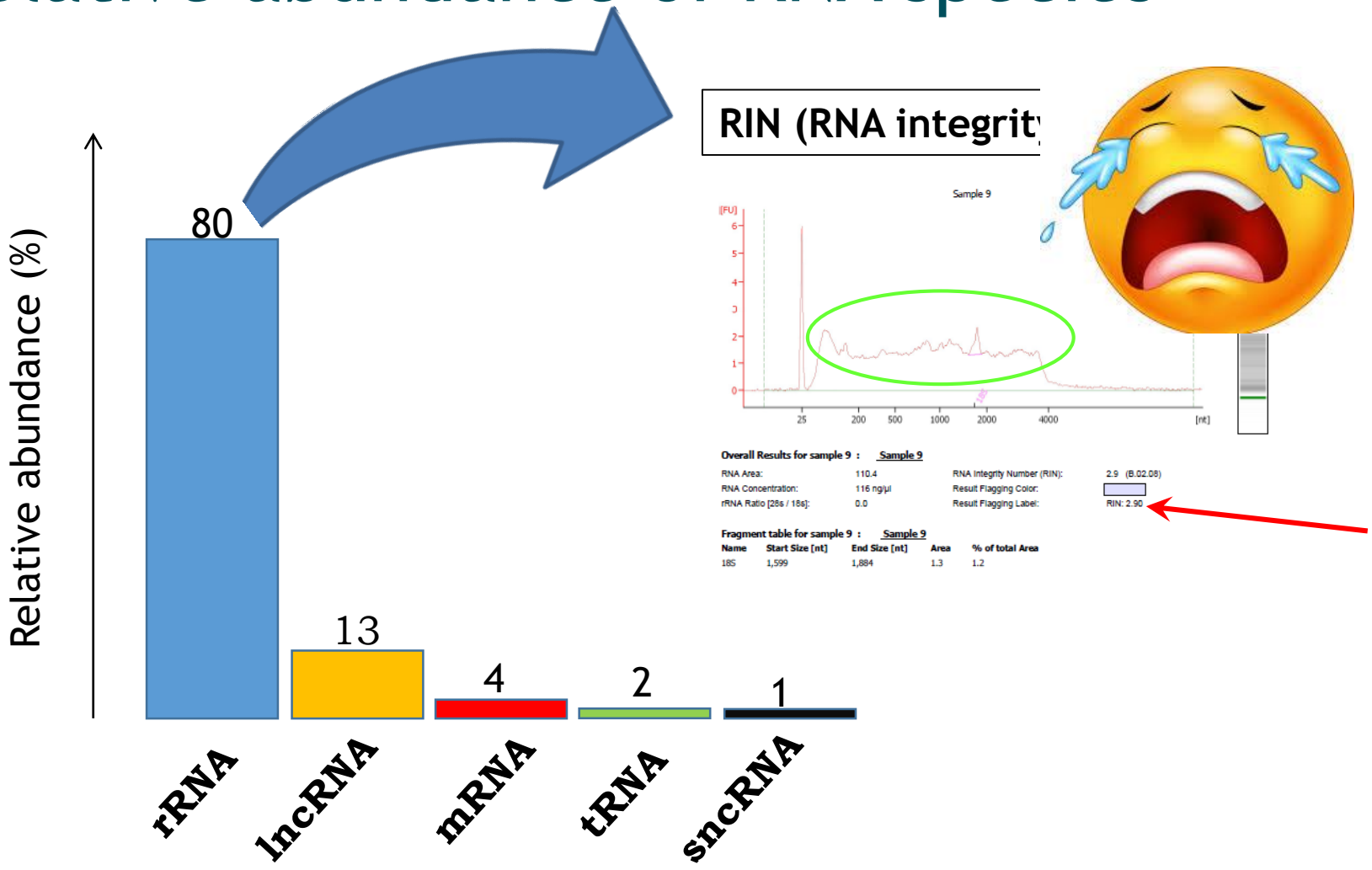
RNA Area:	29.4
RNA Concentration:	25 ng/µl
rRNA Ratio [28s / 18s]:	2.5
RNA Integrity Number (RIN):	10 (B.02.08)
Result Flagging Color:	
Result Flagging Label:	RIN:10

Fragment table for sample 6 : **Sample 6**

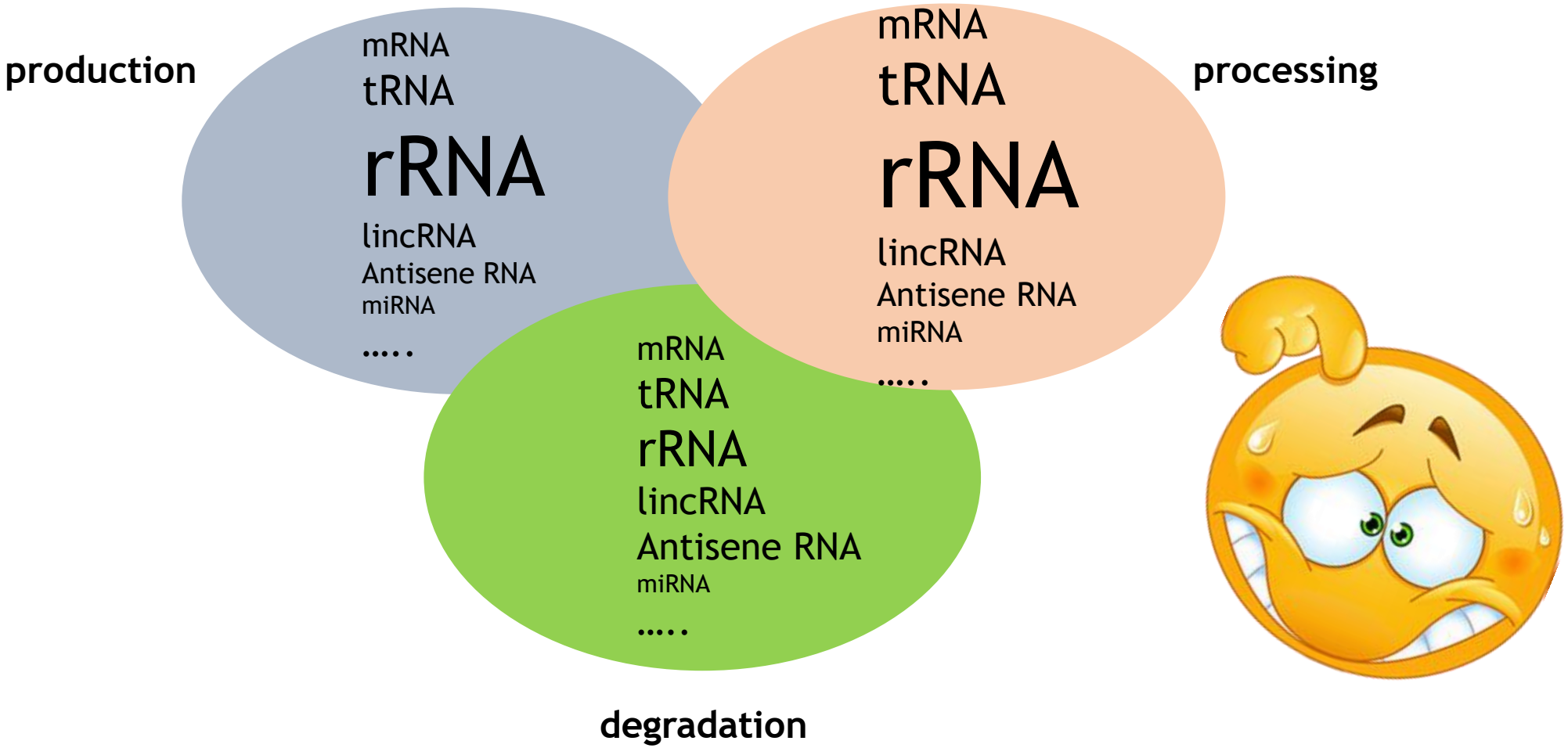
Name	Start Time [s]	End Time [s]	Area	% of total Area
18S	41.07	43.32	5.9	20.2
28S	48.07	50.85	14.6	49.6



Relative abundance of RNA species



What kind of RNA have I isolated?



Approaches for transcriptome sequencing

Gene expression	Target poly(A) mRNAs (enrich or selectively amplify).
Alternative splicing	Target exon/intron boundaries by either doing long read sequencing (>300 bp) or paired end read sequencing (≥ 2 × 100).
miRNA (small RNAs)	Target short reads using size selection purification because miRNAs are in the 18–23 bp range.
Non-coding RNA	Directional RNA sequencing is critical (strand-specific)
Anti-sense RNA	Consider combining mRNA expression with directional RNA sequencing
Single cell RNA	Critical challenge is the technical noise created by amplification.

Preparing RNA for next generation sequencing

The core steps in preparing RNA (or DNA) for NGS analysis are:

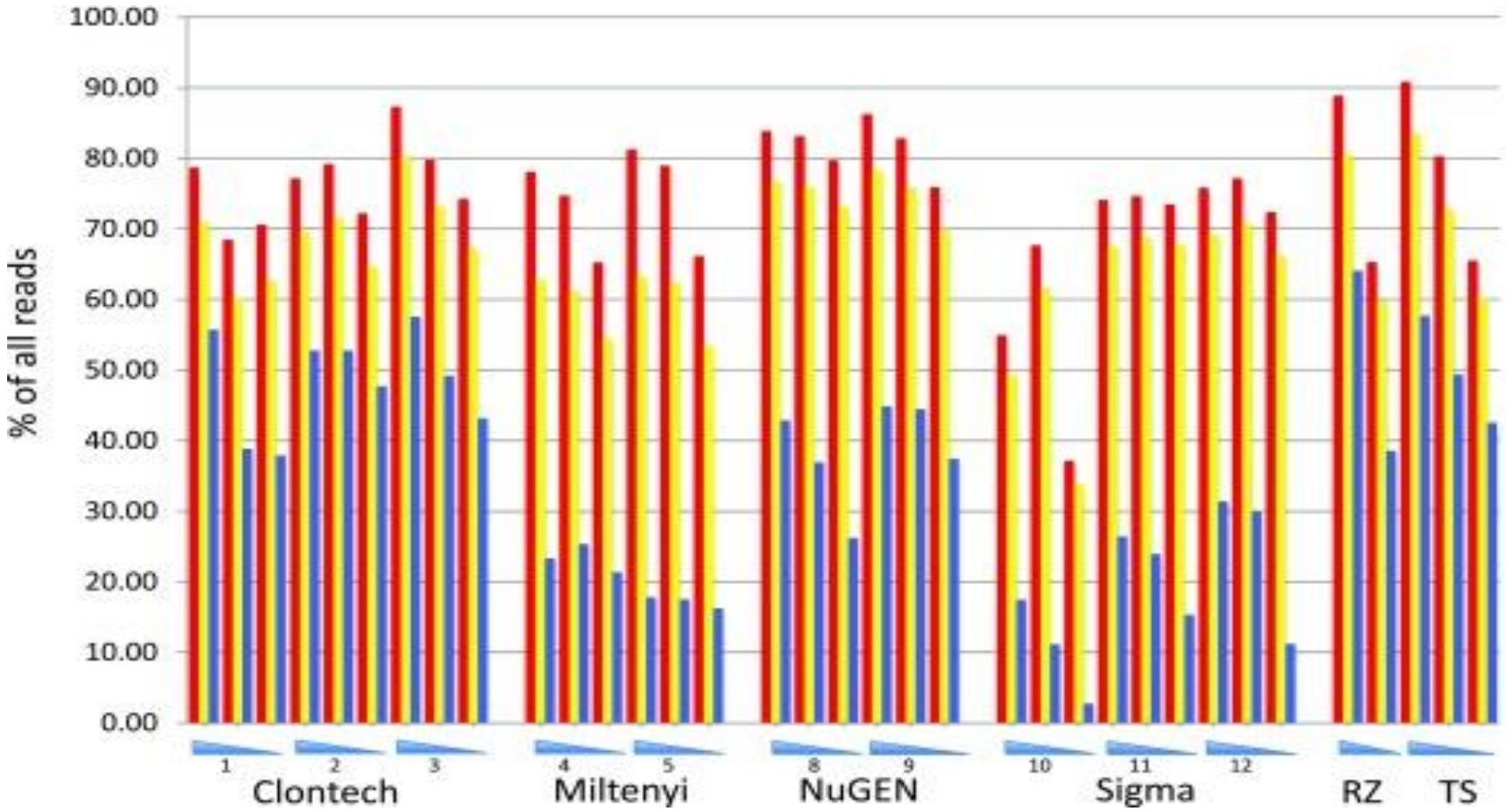
- RNA isolation: kit used, maintaining integrity and complexity !
- converting target to double-stranded DNA !!
- fragmenting and/or sizing the target sequences to a desired length
- attaching oligonucleotide adapters to the ends of target fragments
- quantitating the final library product for s

PCR and [illegible]plification bias

Direct RNA seq with Oxford Nanopore for full length transcripts

Dependency of recovered reads on library prep protocol

RNA concentrations were 5ng, 500pg and 50pg



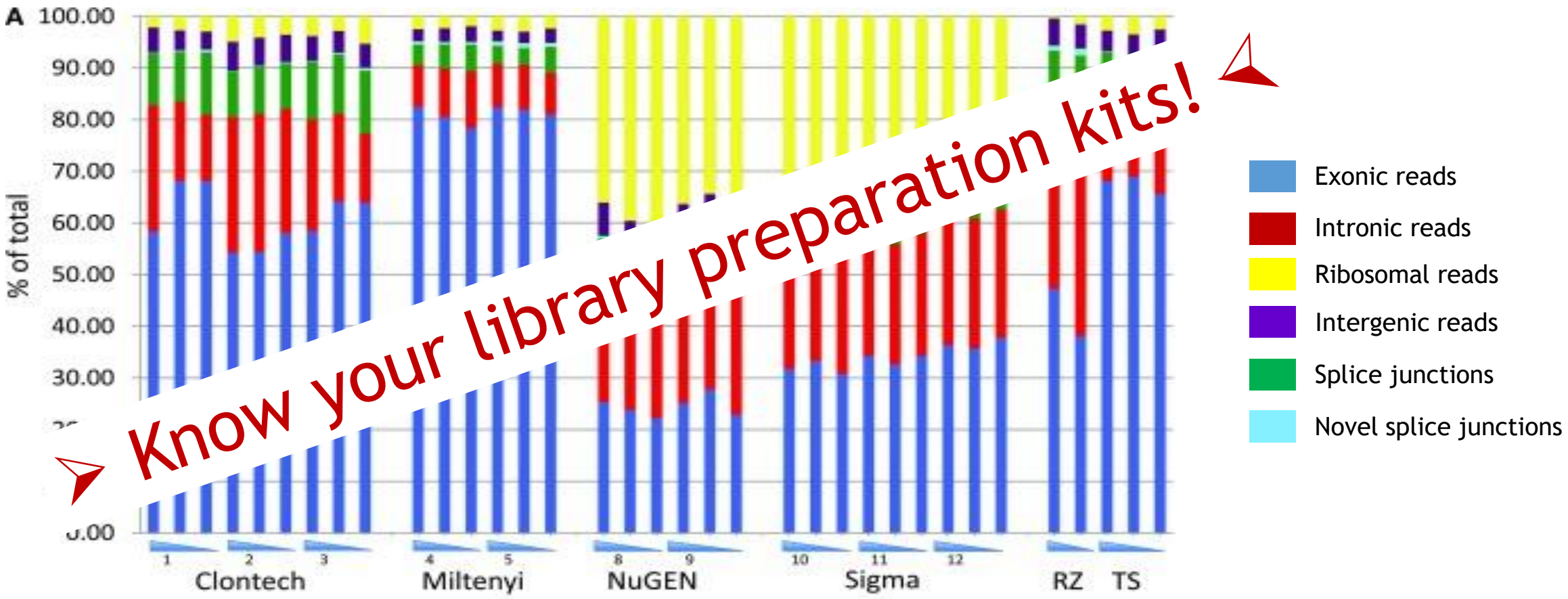
All aligning reads

All aligning reads mapping to unique genomic positions

Putative PCR duplicates

Shanker S etal. J Biomol Tech. 2015

Dependency of recovered RNA biotypes on library prep protocol



Know your library preparation kits!

Shanker S etal. J Biomol Tech. 2015

Summary transcriptome sequencing

Which
- RNA,
- tissue
- isolation kit

} quality and quantity

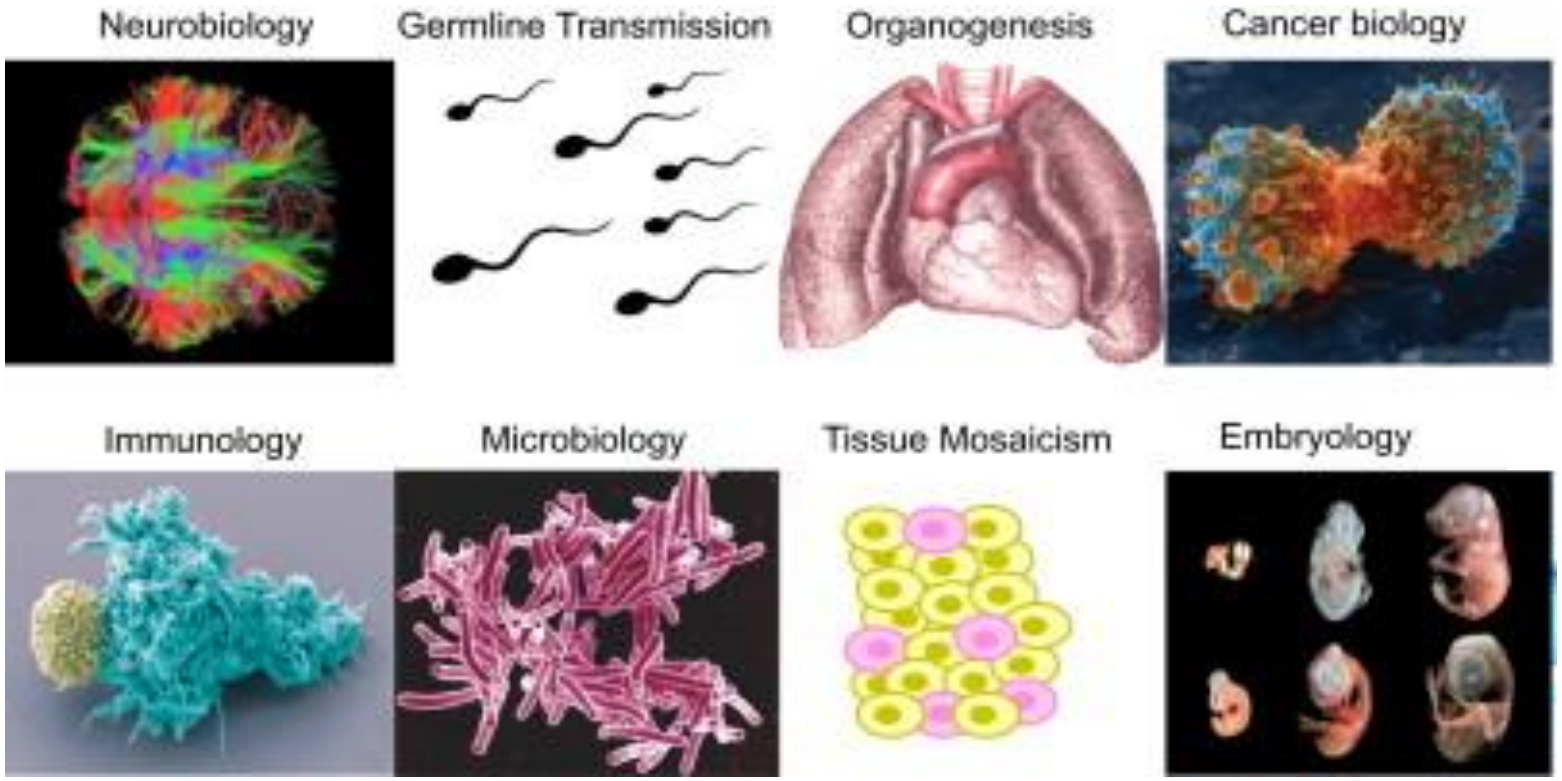
- library prep kit
- sequencing platform

Applications Next Generation Sequencing

- Genome sequencing
- Epigenome sequencing
- Metagenome sequencing
- Transcriptome sequencing

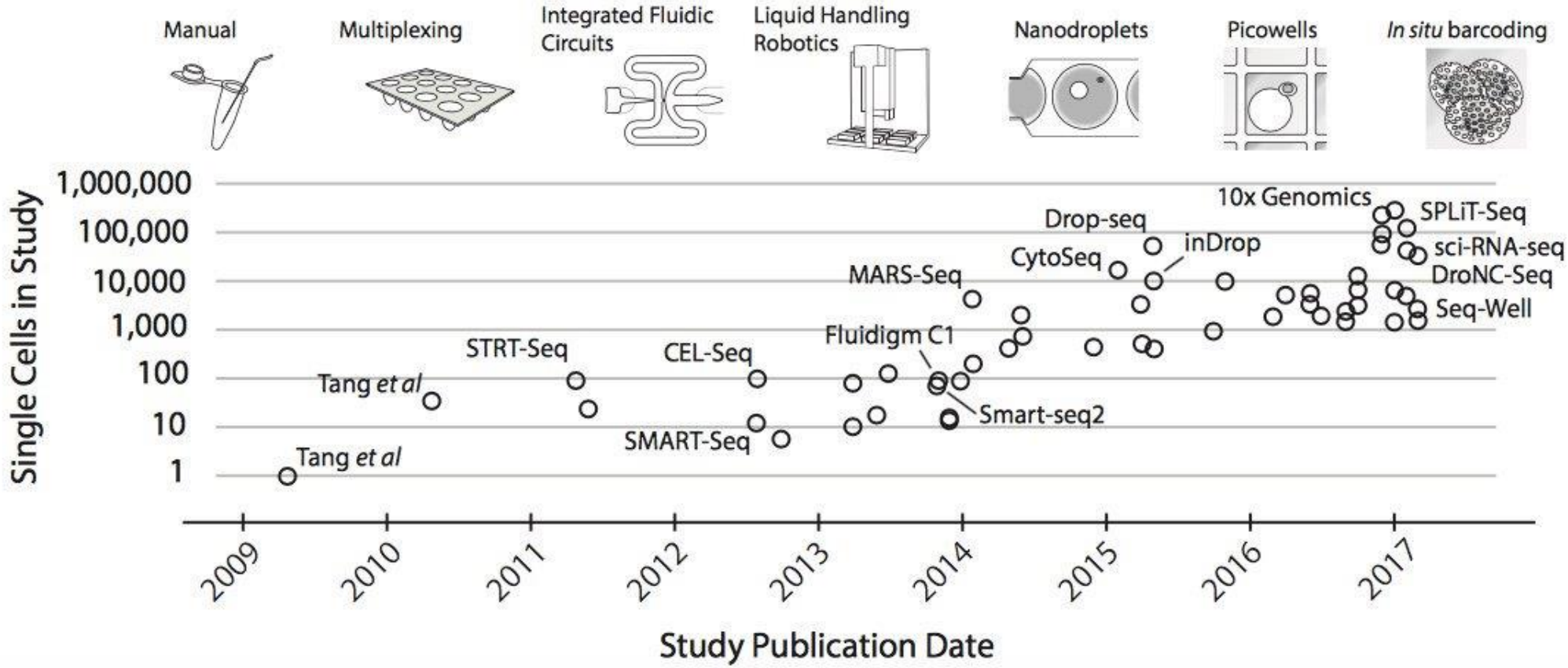
- Single Cell sequencing

Single Cell biology impacts most areas of research



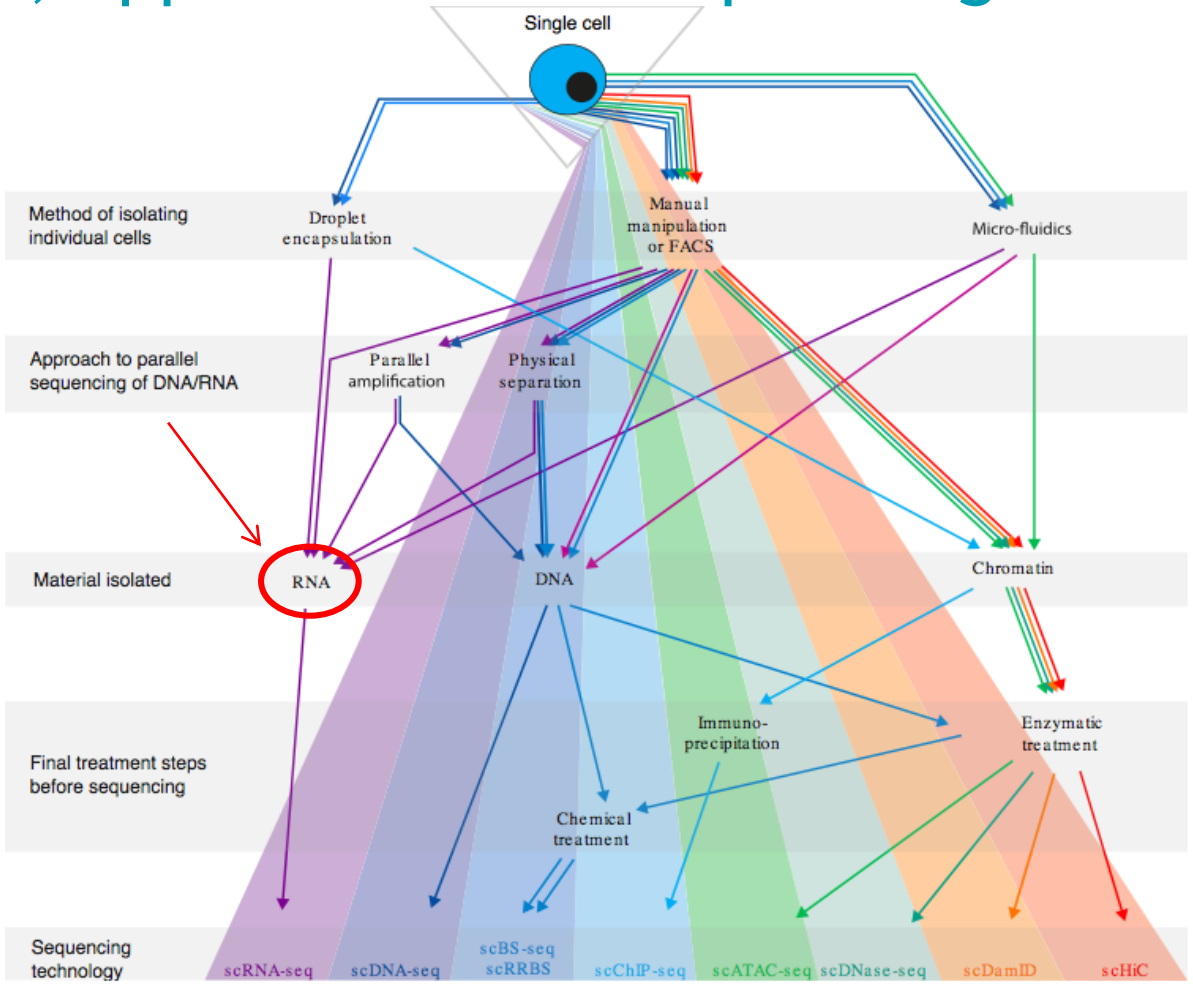
Yong Wang, Nicholas E. Navin
Advances and Applications of Single-Cell Sequencing Technologies
Mol Cell, Volume 58, Issue 4, 2015, 598-609

Scaling Single Cell Transcriptomics



V Svensson, R Vento-Tormo, SA Teichmann
https://arxiv.org/abs/1704.01379

Single Cell Isolation, approaches and sequencing



Technical possibilities in Single Cell Seq

Single Cell Genomics

Copy Number Variation

Single Cell Transcriptomics

- Gene Expression Profiling
- [illegible]
- Gene Immune Profiling (& Cell Surface Protein (cell hashing))
- Gene Immune Profiling (& Antigen Specificity)

Single Cell Epigenomics

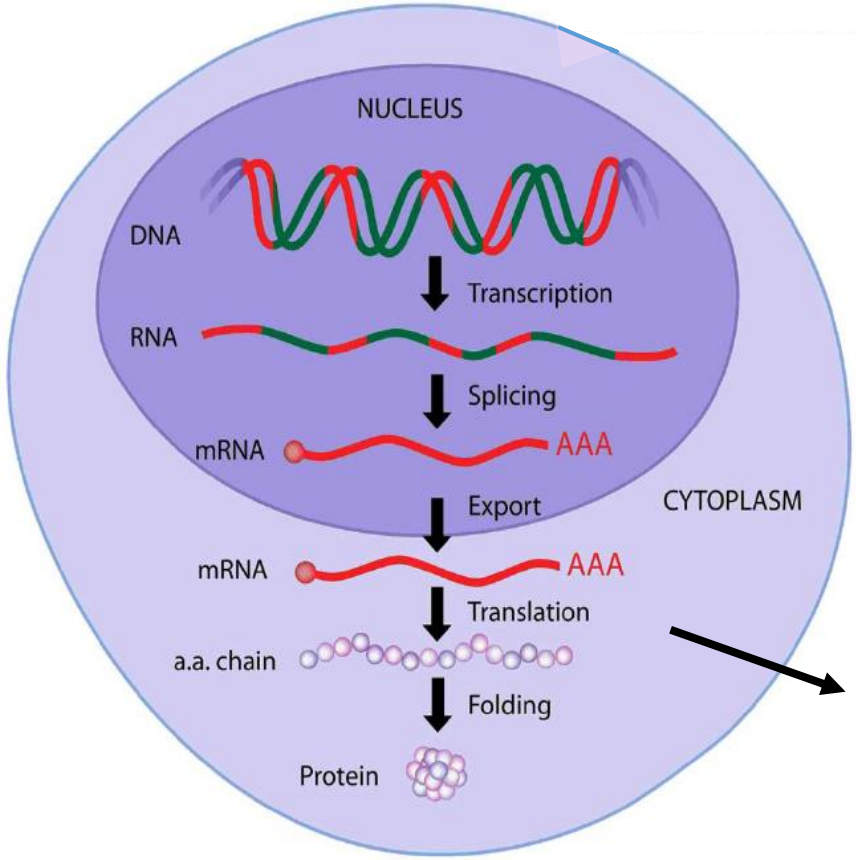
Chromatin Accessibility (ATAC)

Spatial Transcriptomics

Spatial Gene Expression Profiling

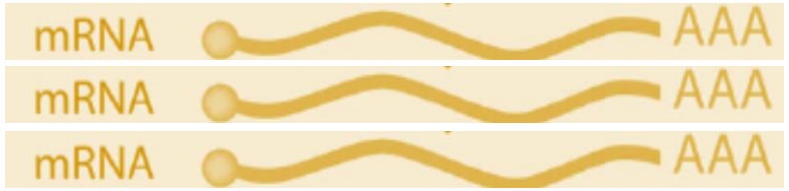
Why Single Cell (mRNA) sequencing?

Protein expression is proportional to gene transcription

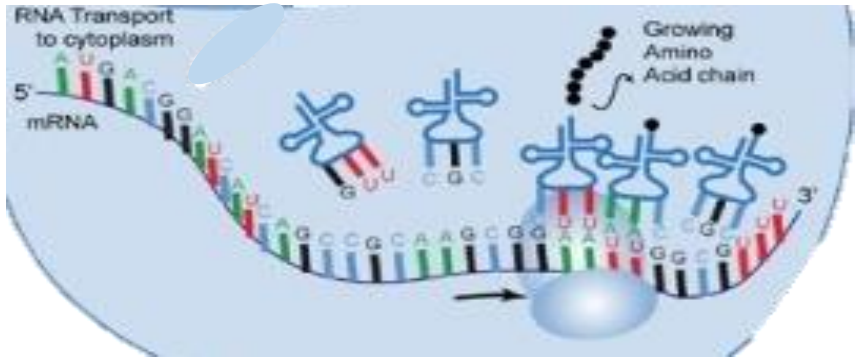


Gen A

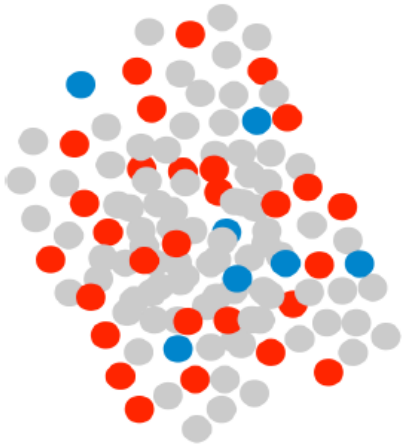
Gen B



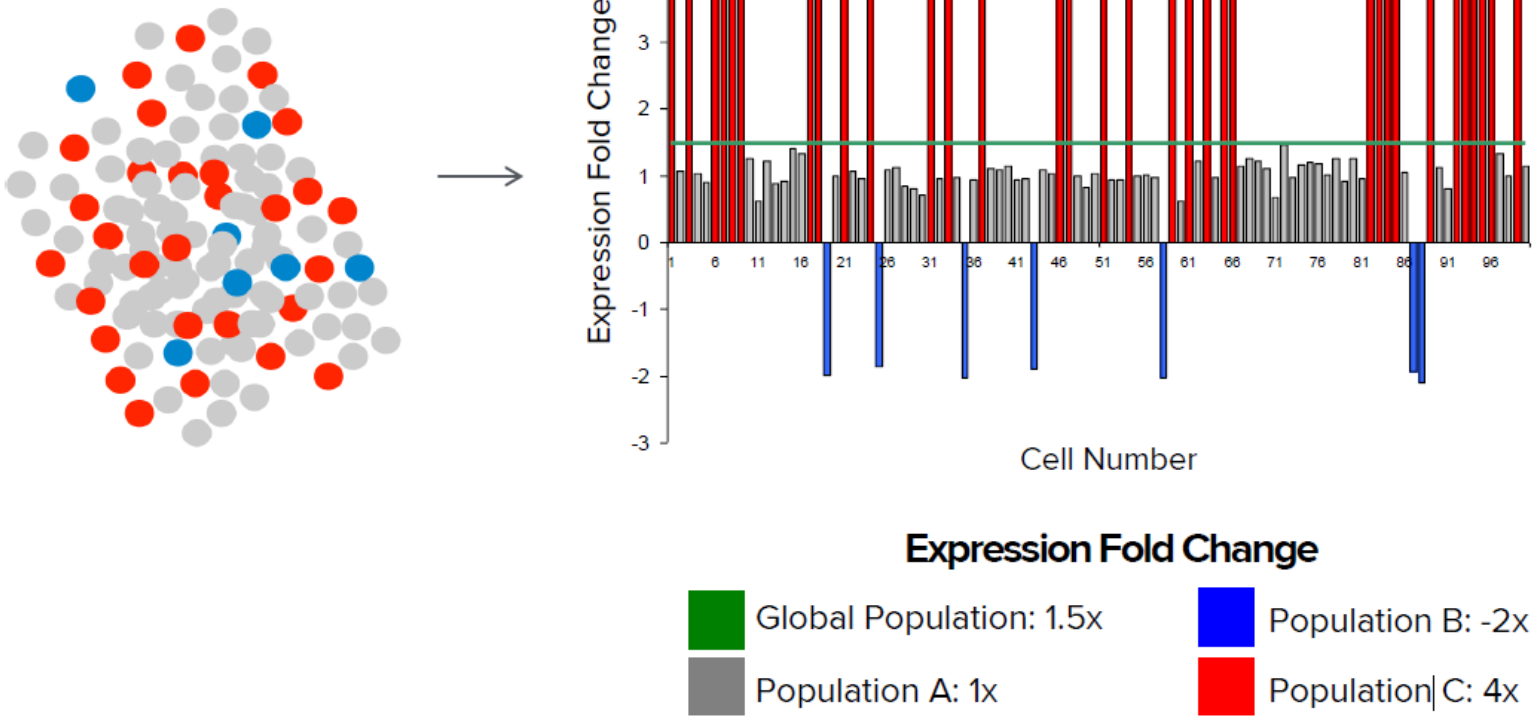
Gen C



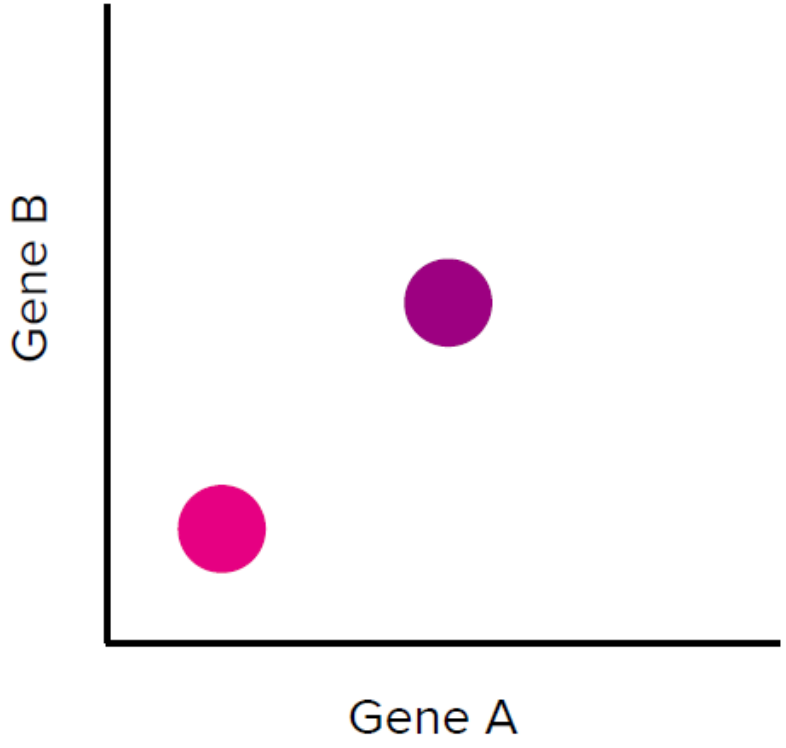
Individual cells behave differently from the average of many cells



Individual cells behave differently from the average of many cells



Pooled cell data can be misleading

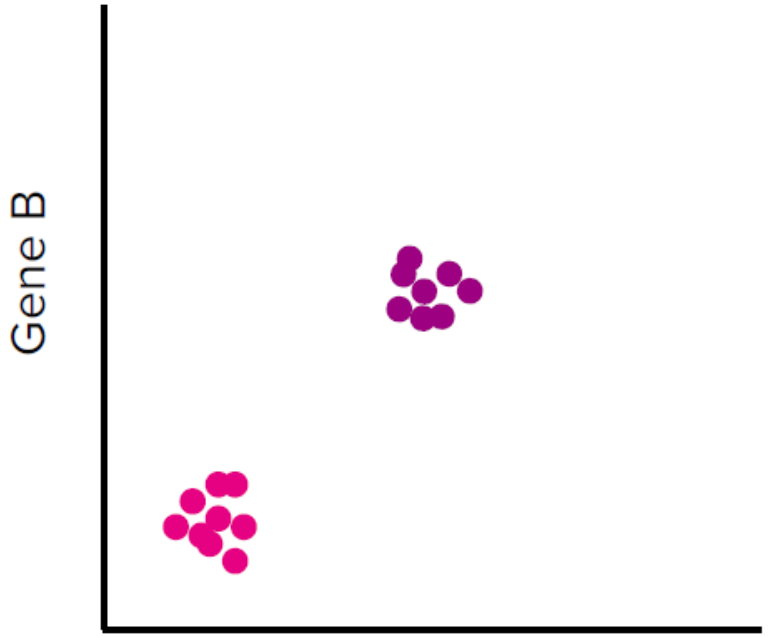


A, B fully correlated

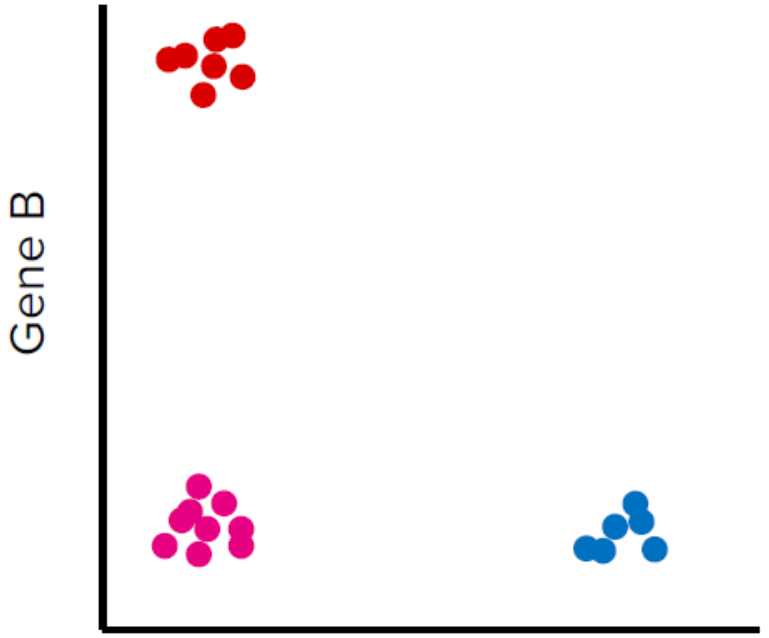
Pooled cell data can be misleading

A, B fully correlated

Pooled cell data can be misleading

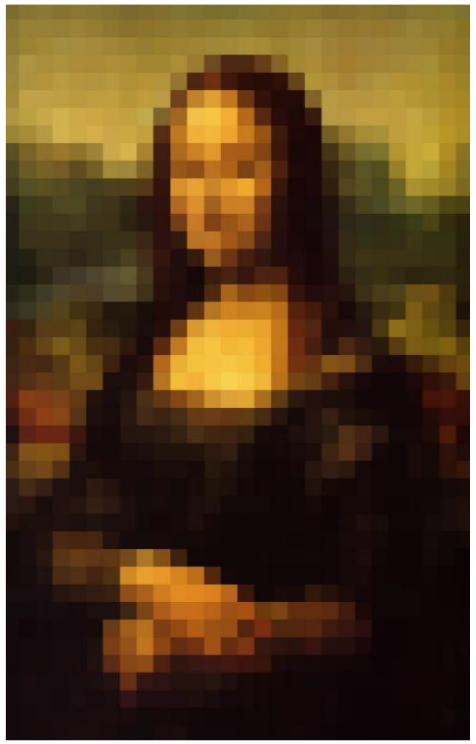


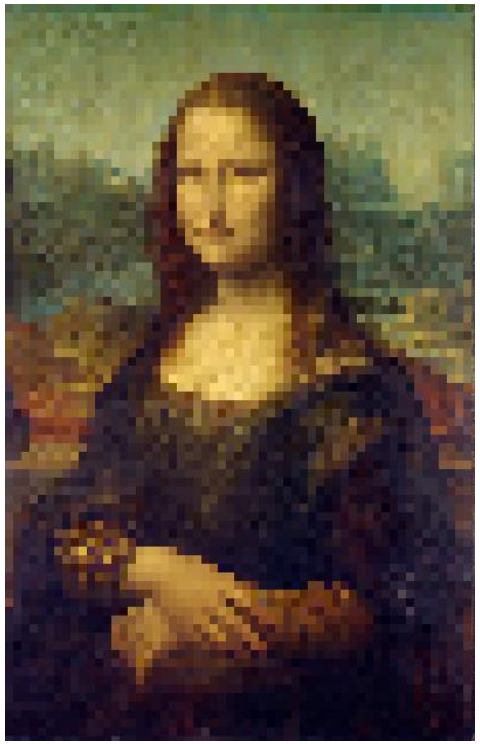
A, B fully correlated



A, B ***anticorrelated***

Single Cell Seq adds level of detail





Technical requirements for Single Cell sequencing

- **Single cell suspension at high enough concentration**
  - No aggregates / clumping
  - No doublets
    - FACS
    - Dnase / trypsin treatment
    - Cell strainer

- **As little manipulation time as possible**
  - Viability cells
  - Avoiding pertubation of transcriptomal profiles in RNA expression

- **Nanoliter protocols**
  - technical and cost efficient

Single Cell Capture
Amsterdam UMC

- FACS
- Chromium Controller, 10X Genomics

Table 1 Brief overview of scRNA-seq approaches

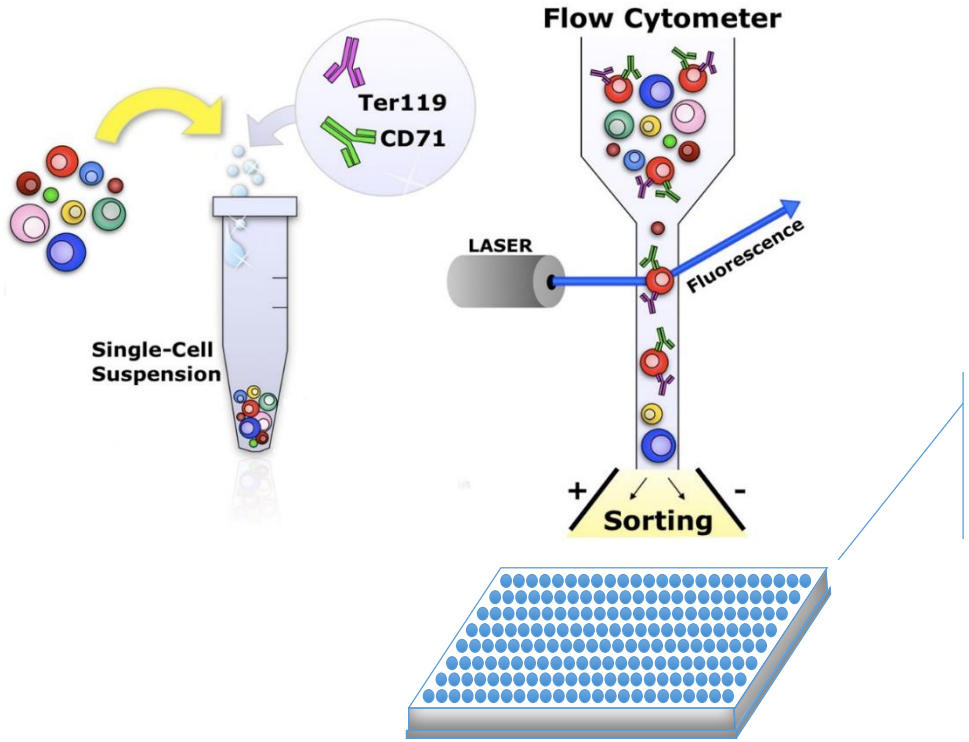
	FACS				microfluidics					
Protocol example	SMART-seq2	MATQ-seq	MARS-seq	CEL-seq	C1 (SMARTer)	DROP-seq	InDrop	Chromium	SEQ-well	SPLIT-seq
Transcript data	Full lenght	Full lenght	3'end counting	3'end counting	Full lenght	3'end counting	3'end counting	3'end counting	3'end counting	3'end counting
Platform	Plate-based	Plate-based	Plate-based	Plate-based	Microfluidics	Droplet	Droplet	Droplet	Nanowell array	Plate-based
Throughput (♯cells)	10^2-10^3	10^2-10^3	10^2-10^3	10^2-10^3	10^2-10^3	10^3-10^4	10^3-10^4	10^3-10^4	10^3-10^4	10^3-10^5
Typical read depth (per cell)	106	106	10^4-10^5	10^4-10^5	106	10^4-10^5	10^4-10^5	10^4-10^5	10^4-10^5	104
Reaction volume	microliter	microliter	microliter	nanoliter	nanoliter	nanoliter	nanoliter	nanoliter	nanoliter	microliter

Haque et al. Genome Medicine (2017) 9:75

Single Cell Capture
Amsterdam UMC

- FACS
- Chromium Controller, 10X Genomics

FACS



384 plate filled with lysis buffer

Reversed Transcription → cDNA → Library preparation → Sequencing

Single Cell Capture
Amsterdam UMC

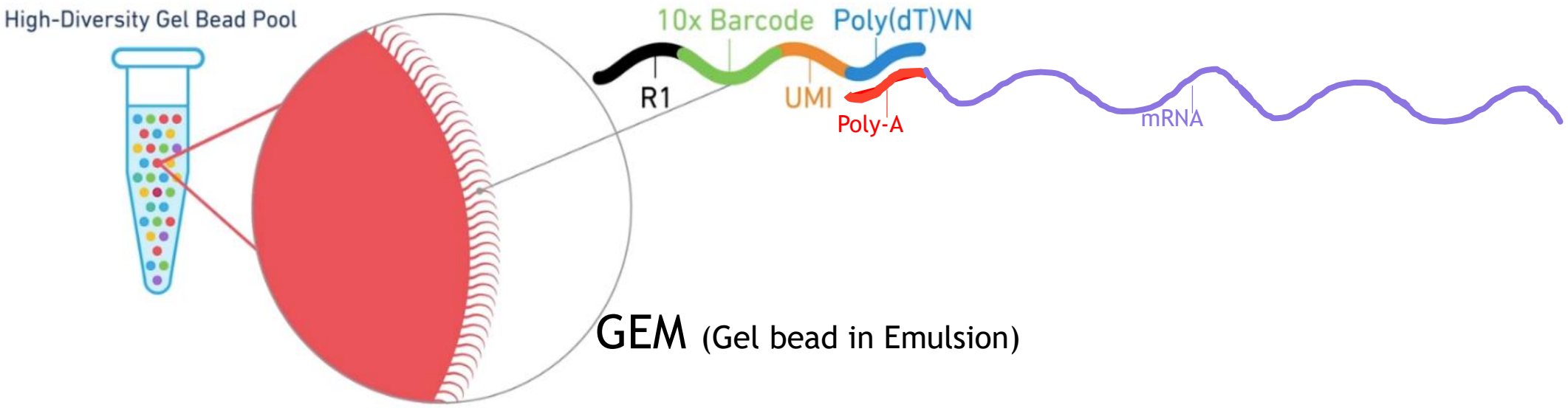
- FACS
- Chromium Controller, 10X Genomics

Amsterdam UMC
Universitair Medische Centra

Chromium Single Cell Solution

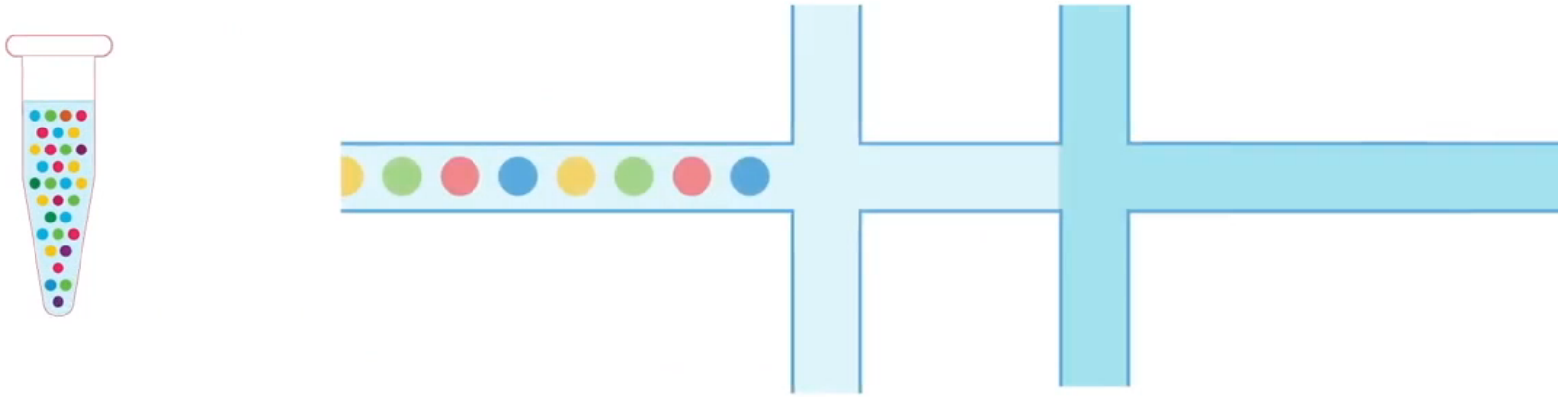
10x Genomics

10X Microfluidics droplet encapsulation

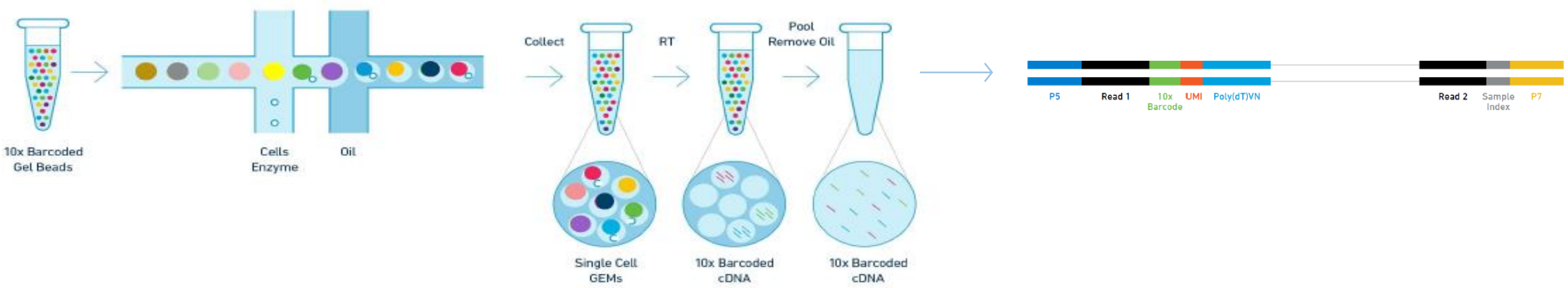


GEM (Gel bead in Emulsion)

10x GemCode™ Technology for Single Cell Partitioning

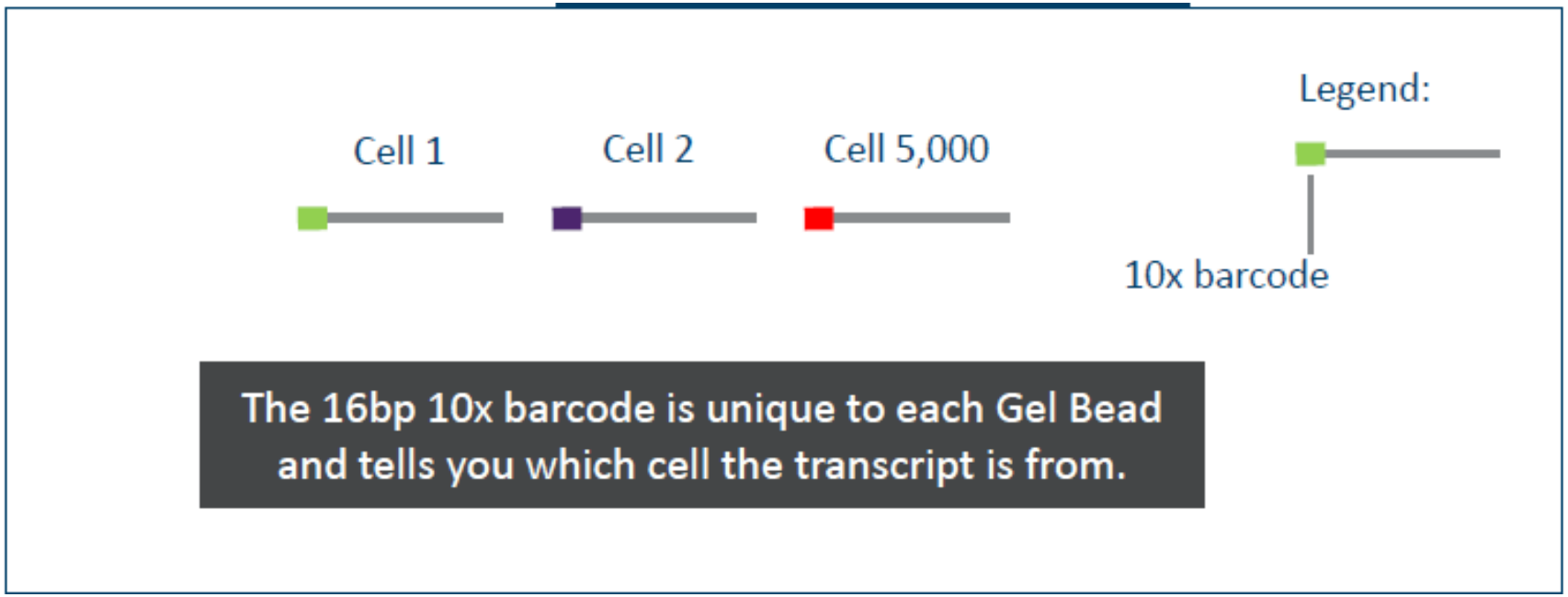


10X Microfluidics droplet encapsulation



10X 3'-expression library

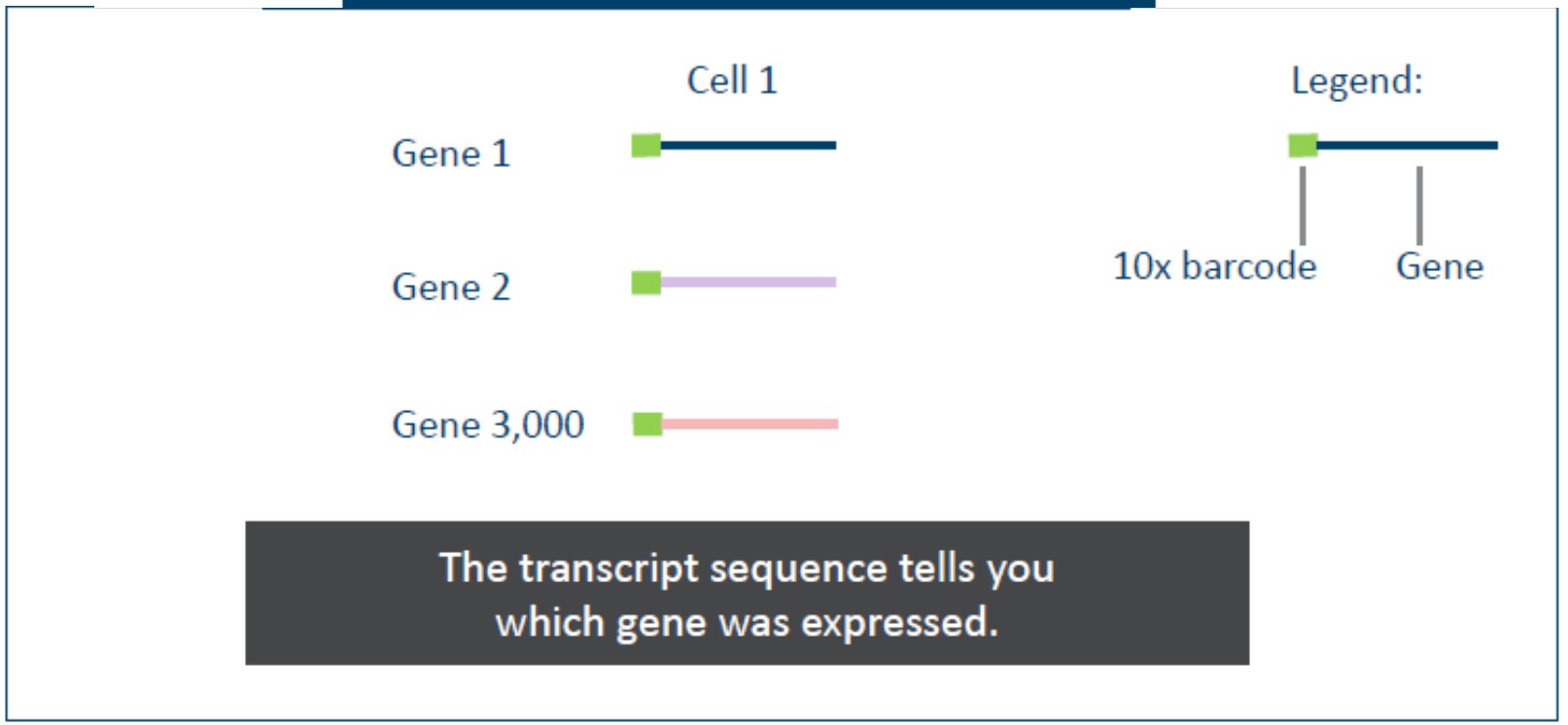
What cell am I from?



The 16bp 10x barcode is unique to each Gel Bead and tells you which cell the transcript is from.

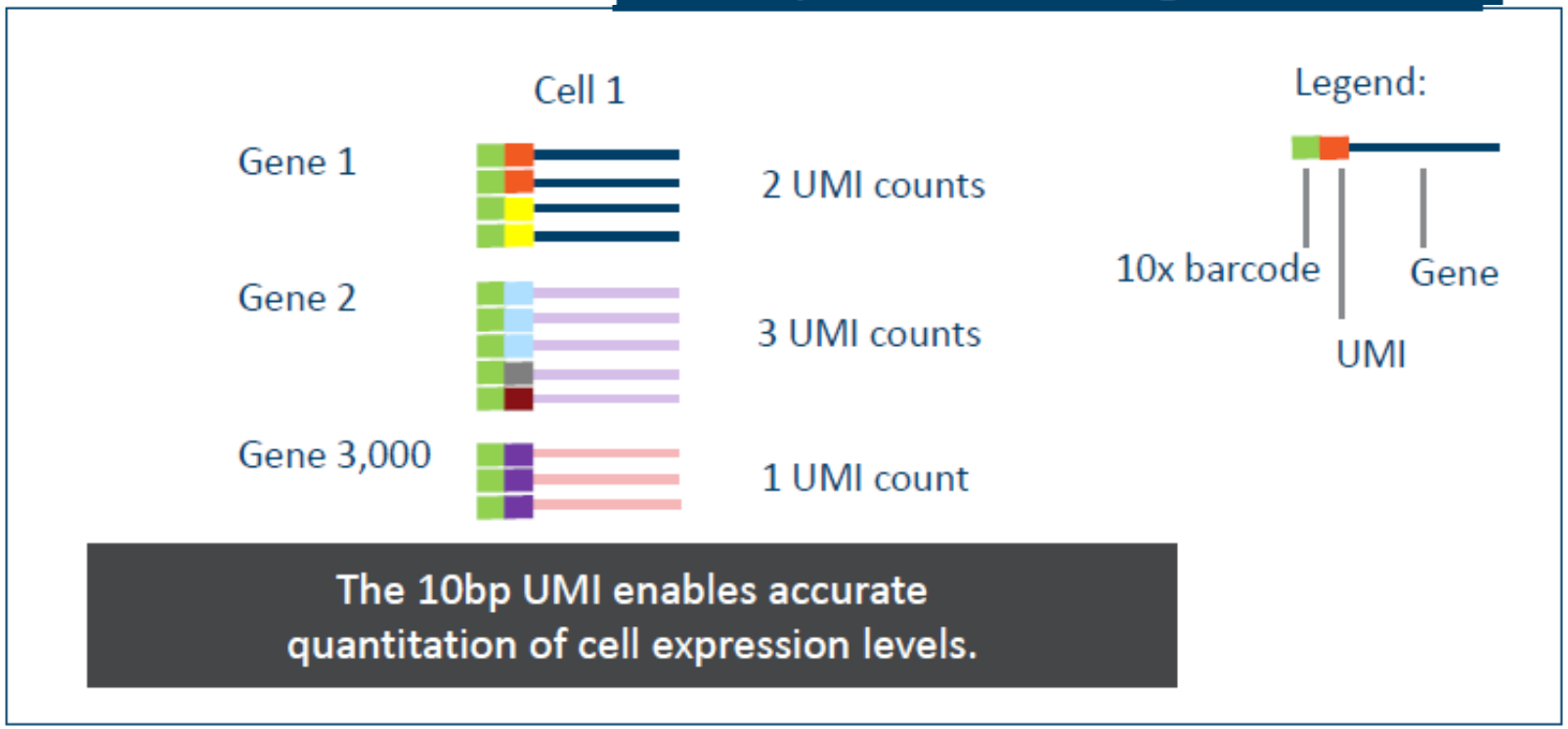
10X 3'-expression library

What gene was expressed?



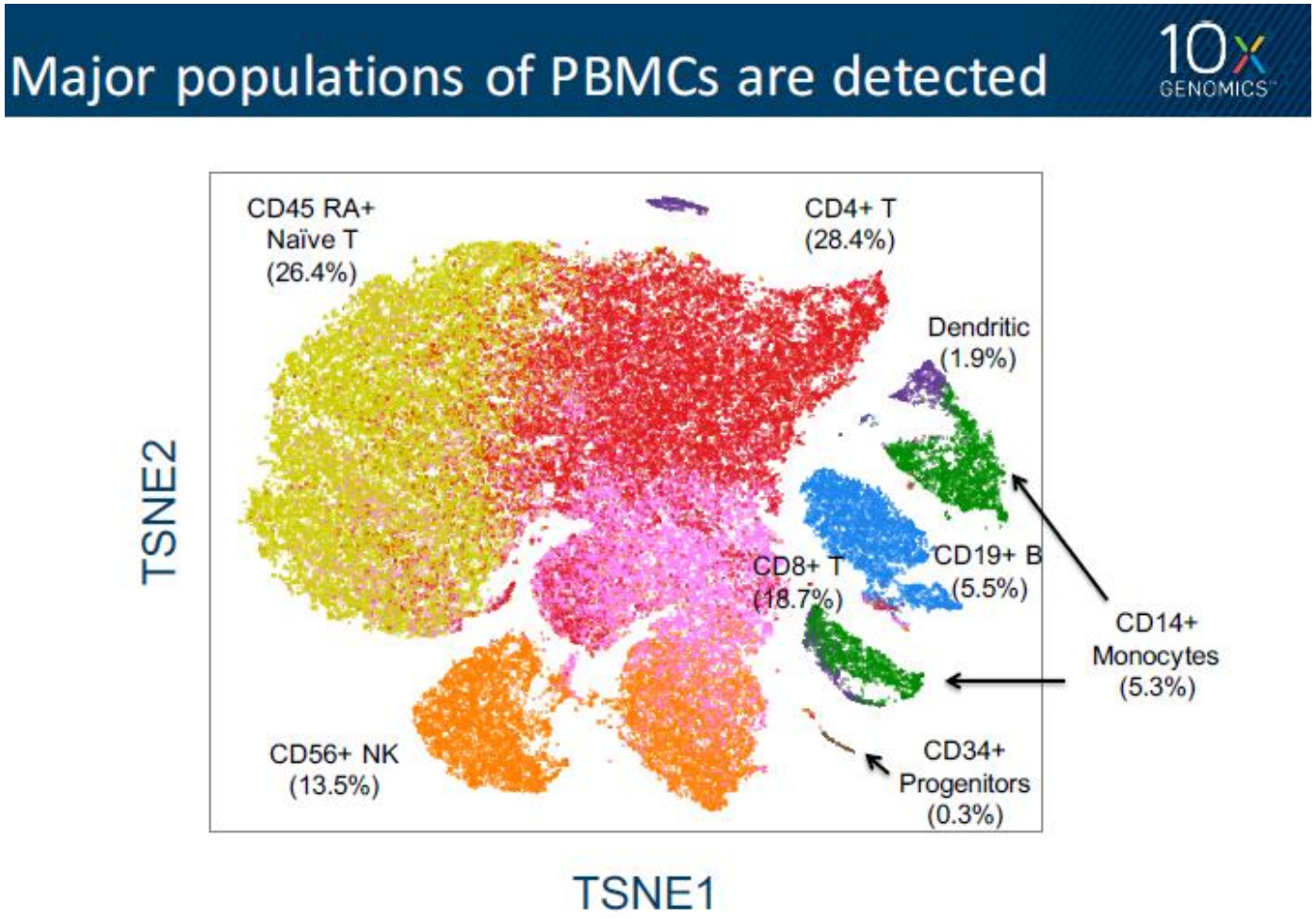
10X 3'-expression library

How many copies of the transcript are present in the single cell?



The 10bp UMI enables accurate quantitation of cell expression levels.

Cell Ranger


Major populations of PBMCs are detected
10X GENOMICS
CD45 RA+
Naïve T
(26.4%)
CD4+ T
(28.4%)
Dendritic
(1.9%)
CD8+ T
(18.7%)
CD19+ B
(5.5%)
CD14+
Monocytes
(5.3%)
CD56+ NK
(13.5%)
CD34+
Progenitors
(0.3%)
TSNE2
TSNE1

Single Cell data analysis

Top differentially expressed genes per cell cluster
(UMI counts/cell)

Tissues

Gene name	Cluster 1	Cluster 2	Cluster 3
Abc1	28.52	0.03	14.7
Xyz2	4.56	8.33	30.85
Fgh3	8.94	17.44	1.27

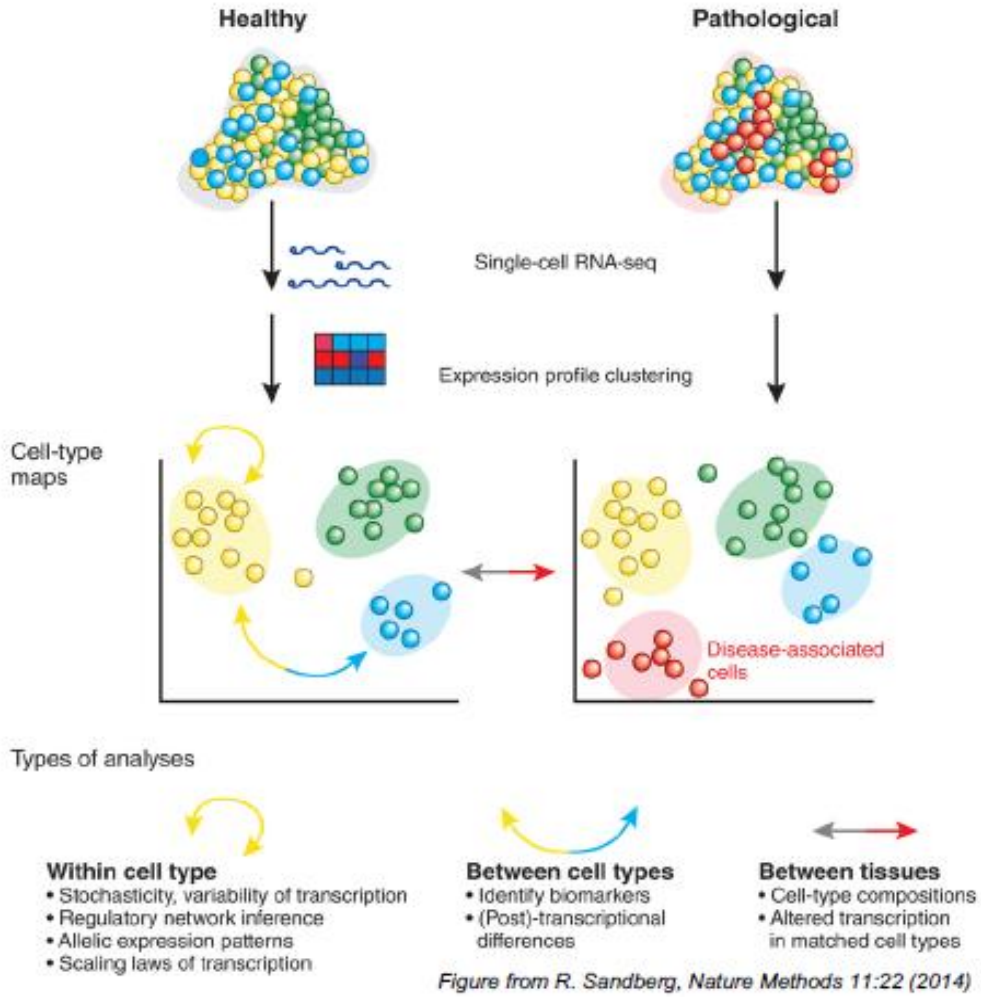


Figure from R. Sandberg, Nature Methods 11:22 (2014)

Technical possibilities in Single Cell Seq

Single Cell Genomics

Copy Number Variation

Single Cell Transcriptomics

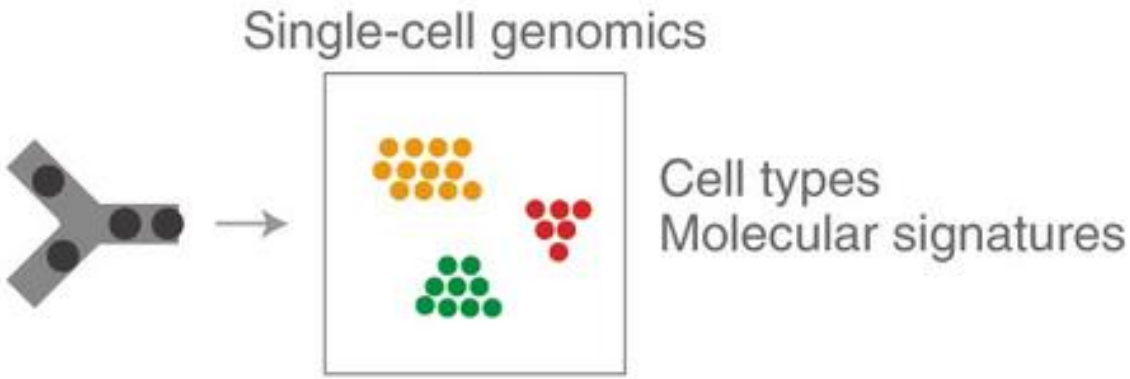
- Gene Expression Profiling (& Cell Surface Protein (cell hashing))
- Gene Expression CRISPR
- Gene Immune Profiling (& Cell Surface Protein (cell hashing))
- Gene Immune Profiling (& Antigen Specificity)

Single Cell Epigenomics

Chromatin Accessibility (ATAC)

Spatial Transcriptomics

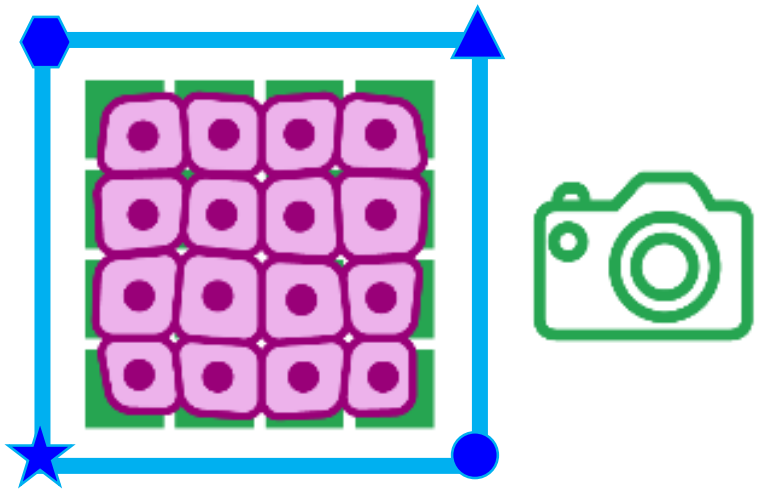
Spatial Gene Expression Profiling


Single-cell genomics
Cell types
Molecular signatures

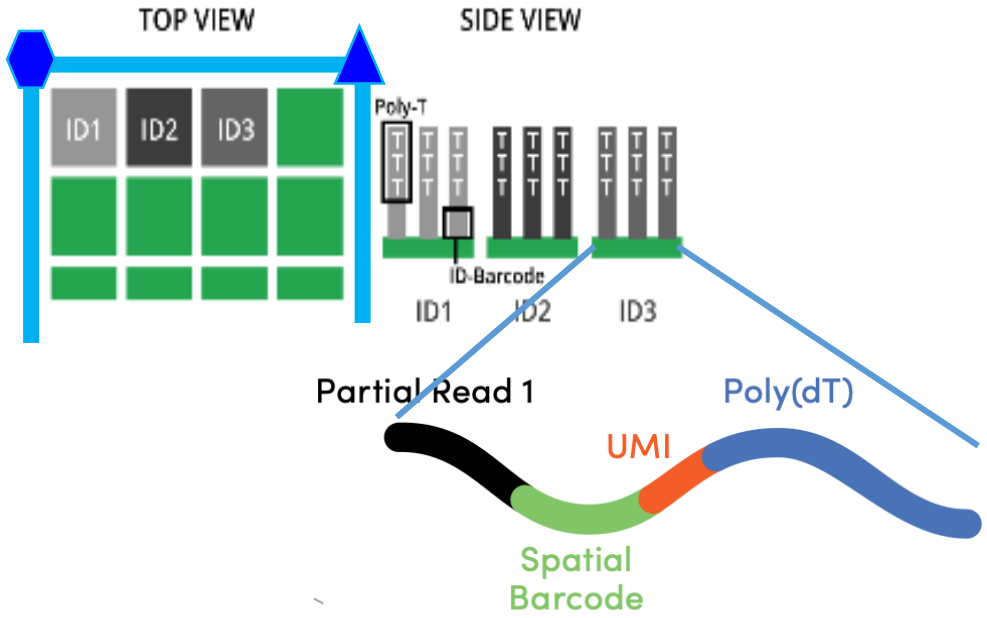
Visium: Spatial Transcriptomics

Section of a fresh frozen tissue on a capture area on a slide.

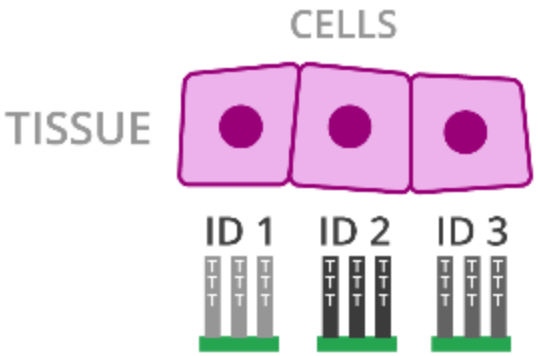
Fixed, stained and imaged together with a detectable frame



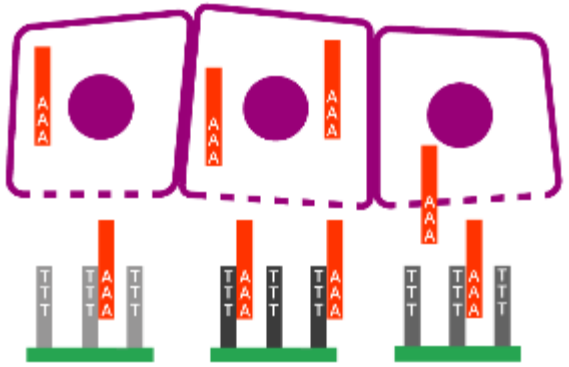
Capture area, coated with an array of barcoded probe spots.



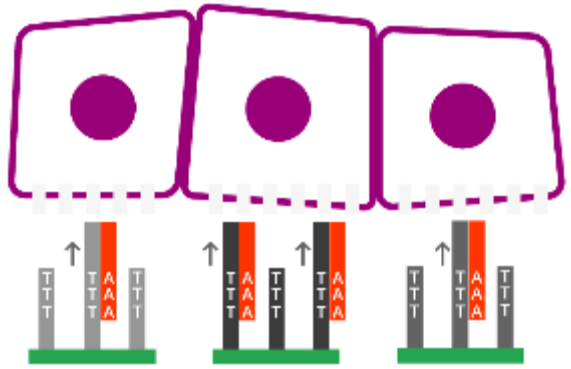
Fixed and stained tissue section on capture area



Permeabilisation of cells, RNA diffuses out of the cells and bind to the probes

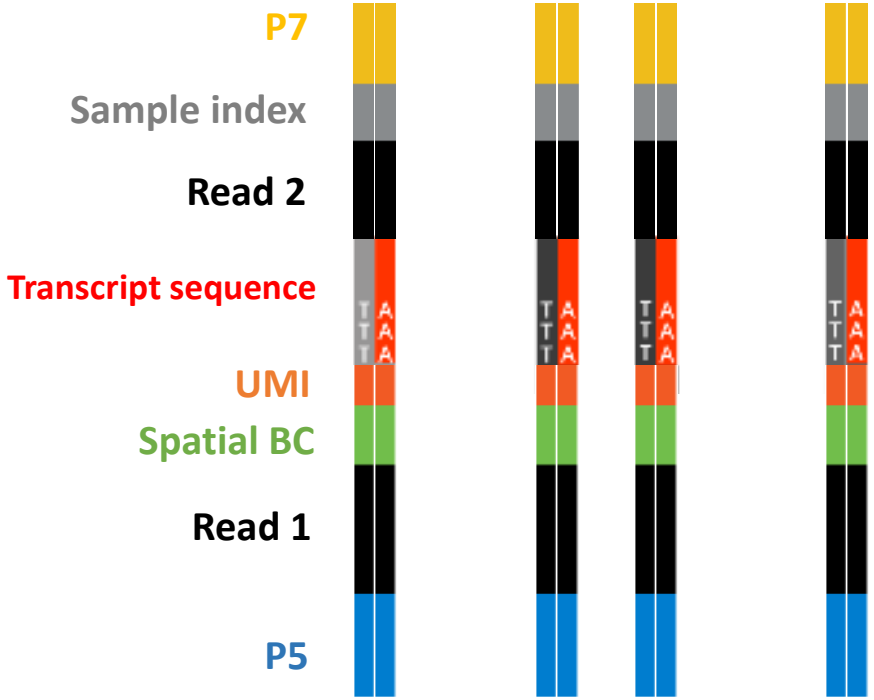


cDNA synthesis of captured RNA

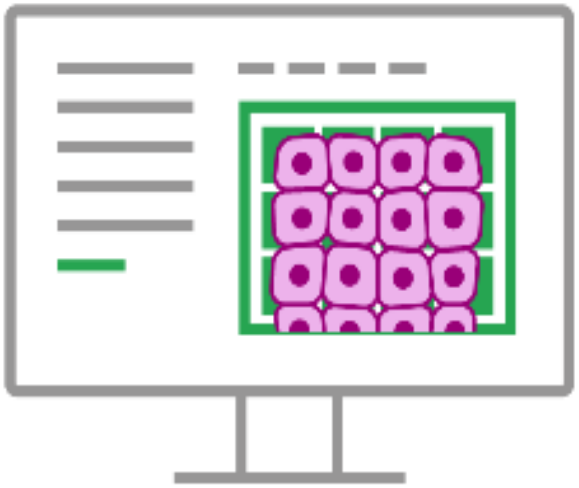


Amsterdam UMC
Universitair Medische Centra

Library prep of the ds cDNA



Sequencing and data analysis of gene expression and histology



Visium: Spatial Transcriptomics example

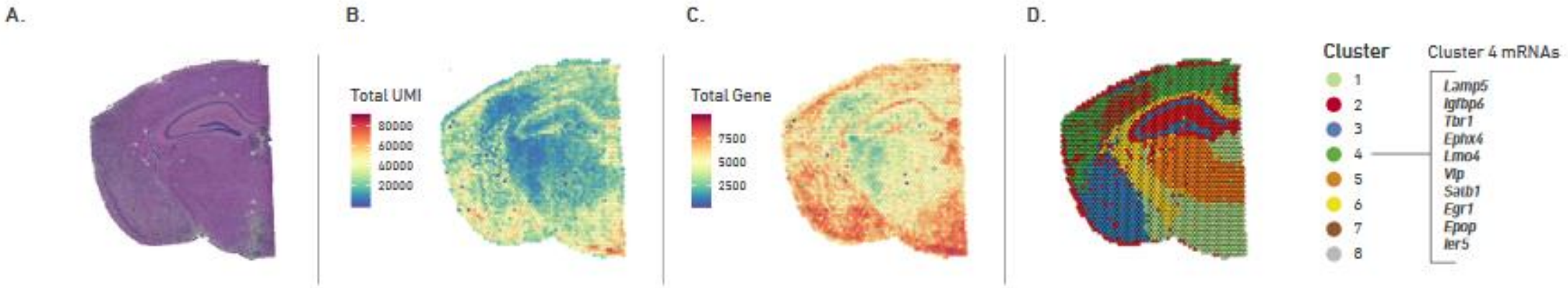


Figure 2. Spatially-resolved clustering and expression in the mouse brain. A. A coronal mouse brain section was H&E stained, imaged, then processed through the Visium Spatial Gene Expression workflow. Shown are image overlays containing data for UMI counts (B), total gene count (C), and spatially naïve clustering based on total differentially expressed genes (D). The top genes that are more highly expressed in cluster 4 (green) than any other cluster are shown to the far right.

Other challenges in NGS

- Bioinformatics
  - Data analysis
  - Tracking and tracing
  - Interpretation
- Data storage
  - New developments in analysis software
- Logistics
  - pre- post PCR laboratories
  - Data and sample tracking

Amsterdam UMC
Universitair Medische Centra

Acknowledgements

Core Facility Genomics

Lourens Bordewijk
Linda Koster
Patrick Ruizendaal
Suzan Kenter
Alex Postma
Peter Henneman
Elisabeth Lodder

Ingrid Bakker
Ted Bradley
Lyra Eken
Sebastiaan de Vriend
Daoud Sie

Clinical Genetics

Izabela Krzyzewska
Andrew Li Yim

Andrea Venema
Adri Mul

Olaf Mook
Martin Haagmans

Alex Postma
Peter Henneman
Erik Sistermans
Marcel Mannens

KEBB

Barbera van Schaik
Aldo Jongejan
Perry Moerland
Antoine van Kampen

Contact info CFG:

Location VUmc
Visiting address: MF-J379 Telephone: 020-4448351

Location AMC
Visiting address: K2-213 Telephone: 020-5663846

E-mail: cfg@amsterdamumc.nl
Website: http://cfg-amsterdamumc.nl