Genomics workshop

Sequencing technologies

Neeman Mohibullah, Ph.D.

Director, Integrated Genomics Operation

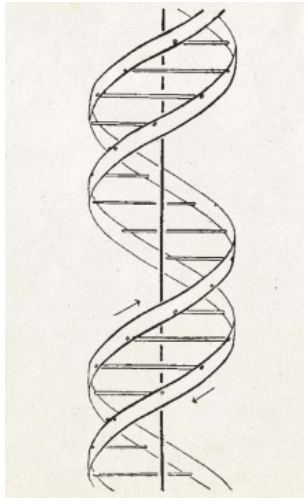


Memorial Sloan Kettering Cancer Center

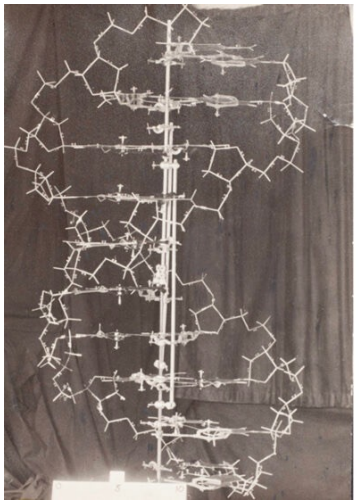


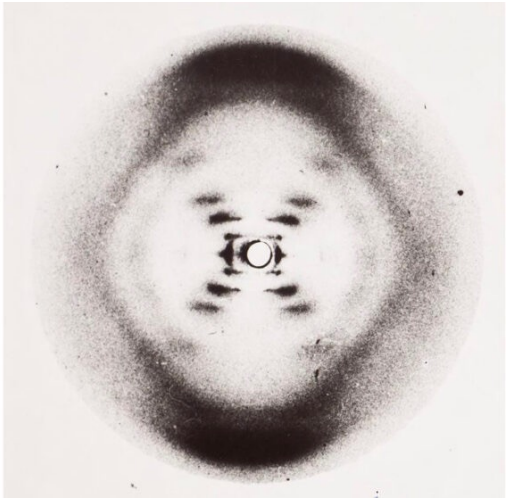
MSK Confidential – Do not distribute

DNA structure (1953)



by Odile Crick in Nature 171 (1953)





Franklin, R.E. and Gosling, R. Nature 171 (1953):740

Focus on composition of DNA molecules rather than order

RNA-seq precedes DNA seq

The analyses in Table I are consistent with the formulas for the purified RNAs.[3]

Alanine RNA: pGp[(Ap)$_{10}$(Cp)$_{28}$(Gp)$_{28}$(Up)$_{17}$(P$_s$Up)$_3$]A
Valine RNA: pGp[(Ap)$_{16}$(Cp)$_{22}$(Gp)$_{24}$(Up)$_{16}$(P$_s$Up)$_4$]A
Tyrosine RNA: pGp[(Ap)$_{18}$(Cp)$_{22}$(Gp)$_{23}$(Up)$_{14}$(P$_s$Up)$_4$]A

As indicated in Table I, the partition coefficients of the valine and tyrosine RNAs in the

TABLE I

NUCLEOTIDE COMPOSITIONS OF PURIFIED ALANINE-, VALINE- AND TYROSINE-ACCEPTOR RIBONUCLEIC ACIDS OF YEAST[a]

Nucleotide	RNA, Mole %: Alanine	Valine	Tyrosine
Ap[b]	12.1	19.1	21.7
Cp	29.9	27.5	26.7
Gp	33.7	29.5	30.2
Up	20.8	19.0	17.1
PsUp	3.7	4.9	4.4
Partition coefficient[c]	0.12	0.21	4.2

[a] Alkaline hydrolysates of 1 mg. of the RNAs were chromatographed on Dowex 1 columns 0.2 × 15 cm. according to the procedure of W. E. Cohn and E. Volkin, *Nature*, **167**, 483 (1951). The values given are averages of three or more determinations, and are not corrected for the terminal nucleoside and nucleoside diphosphates. [b] The abbreviations used are: Ap, adenylic acid; Cp, cytidylic acid; Gp, guanylic acid; Up, uridylic acid; and PsUp, pseudouridylic acid. [c] Partition coefficients of the purified RNAs in a countercurrent distribution solvent system composed of phosphate buffer, formamide and 2-propanol (see ref. 1).

Holley et al *J. Am. Chem. Soc.* (1961): 4861

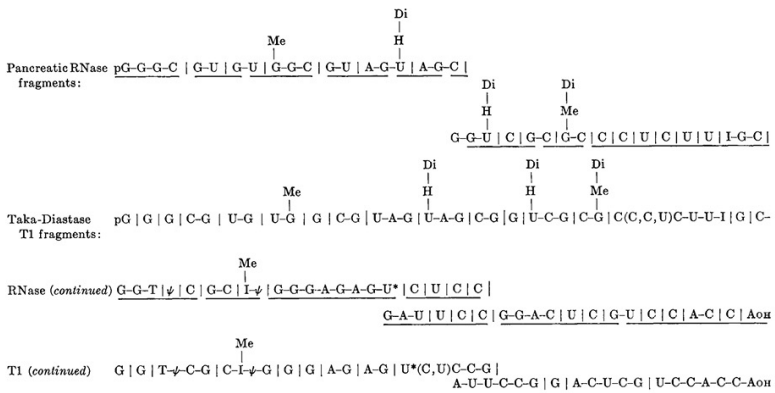
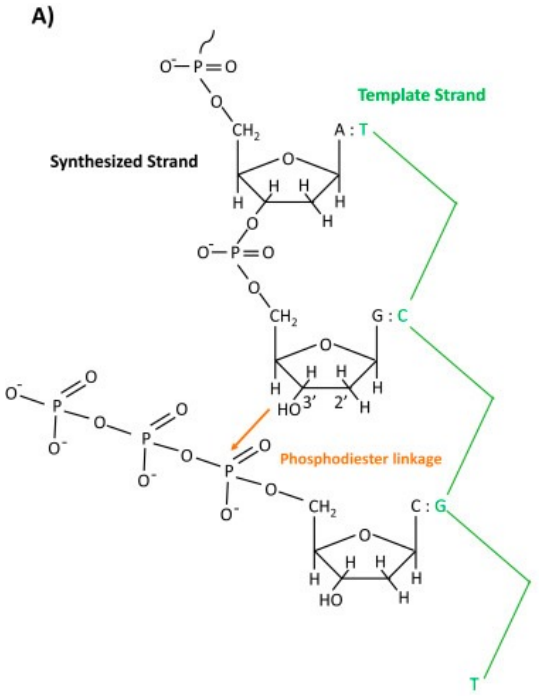
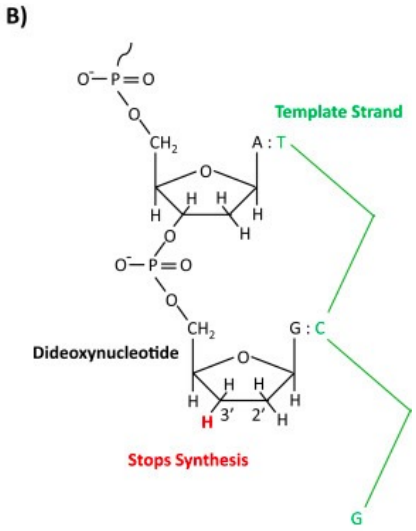


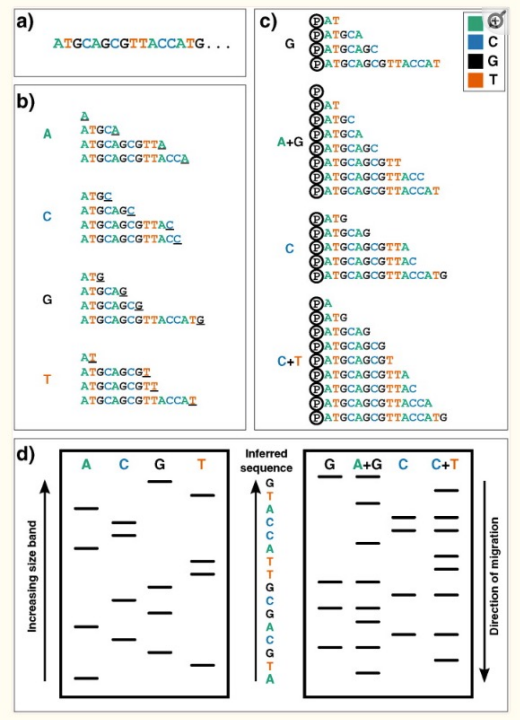
FIG. 7. One of many possible arrangements of the pancreatic RNase and RNase T1 digest fragments that shows the overlaps between the two digests. The RNA molecule is accounted for by the 16 oligonucleotide sequences indicated by the *solid lines*. Only the positions of the two terminal sequences are known. *Vertical lines* indicate the position of enzymatic attack. The *asterisk* indicates that the uridine may be partially substituted by DiHU.

Holley et al *J. Biol. Chem.* (1965): 2122

Advent of PAGE, Maxam and Gilbert and Sanger







Sanger

Maxam-Gilbert

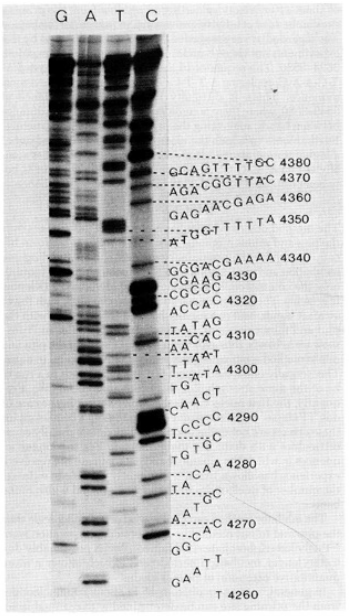
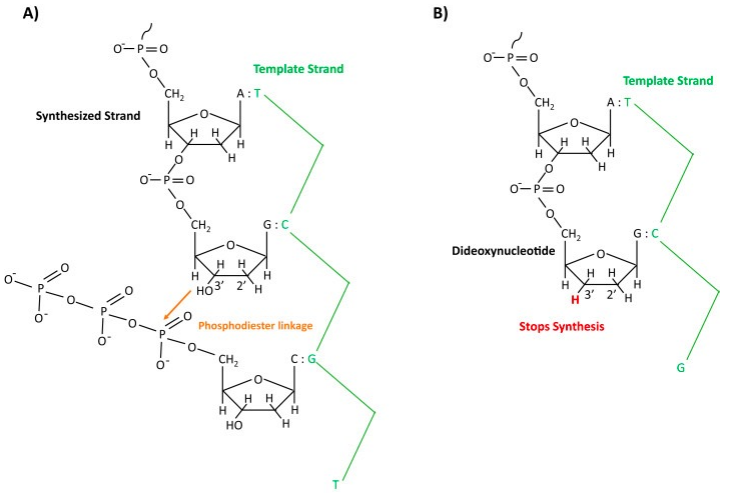
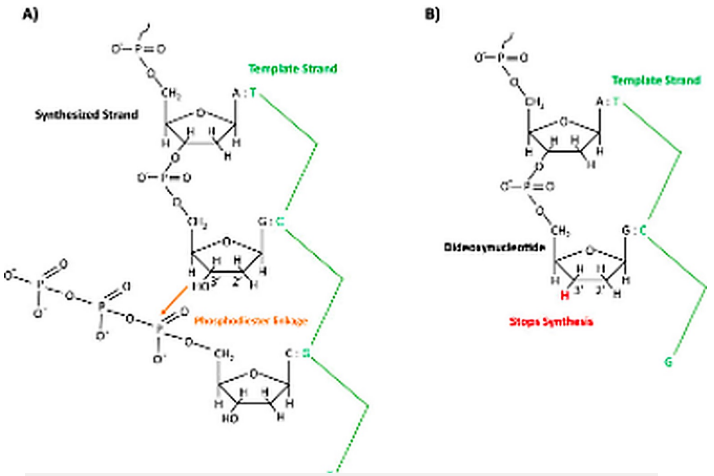


FIG. 2. Autoradiograph from an experiment R4 as primer on the complementary strand of φX174 DNA. Conditions were as in Fig. 1 with the following exceptions used as inhibitor instead of araCTP. After incubation of the solutions at room temperature for 15 min, 1 μl of 0.5 mM dATP and 1 μl of restriction enzyme *Hae* III (4 units/μl) were added and the solutions were incubated at 37° for 10 min. The *Hae* III cuts close to the *Hin*dII site and it was used because it was more readily available. The electrophoresis was on a 12% acrylamide gel at 40 mA for 14 hr. The top 10.5 cm of the gel is not shown.

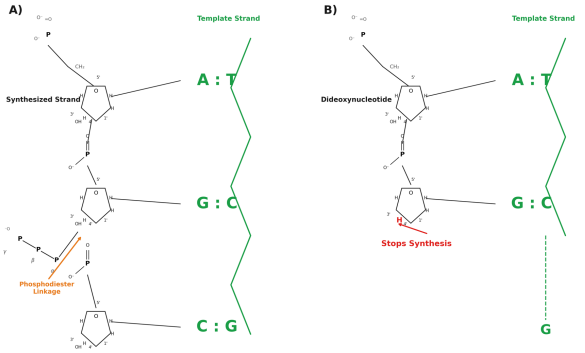
J. Mol. Biol. (1975) 94:441, Genomics (2016) 107(1):1, PNAS 74:560, PNAS 74:5463



can you polish this image to make it easier to read when projected on the screen. i can't read the letters



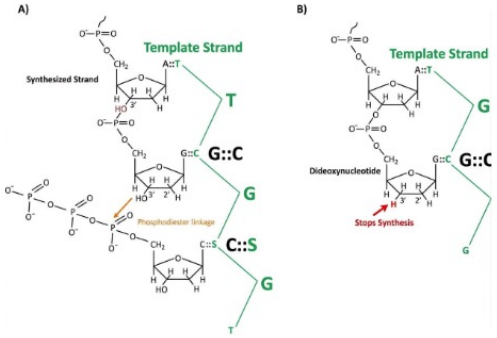
please redraw this diagram focusing on the letters in the template strand and their base-pair partners



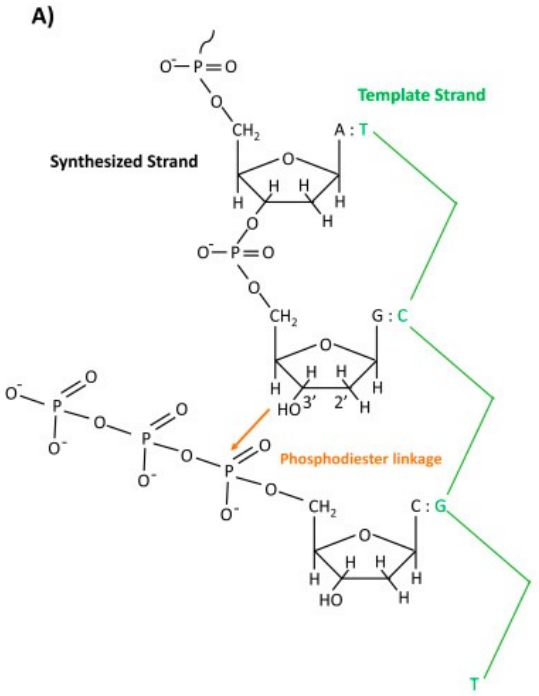
can you increase the font size for the green letters and their partners

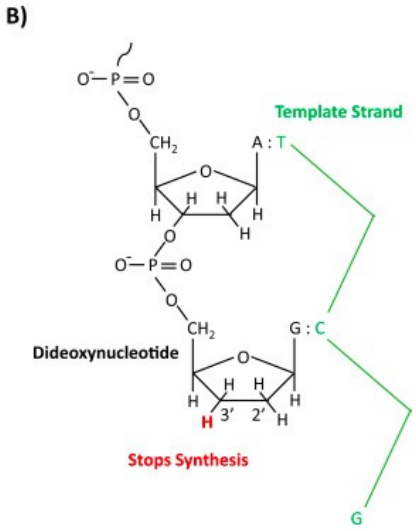
I have updated the diagram, enlarging the font for the green template strand letters (`A` , `T` , `G` , `C`), the overall `Template Strand` title, and their complementary partner bases on the black synthesized strand to make them much easier to read.

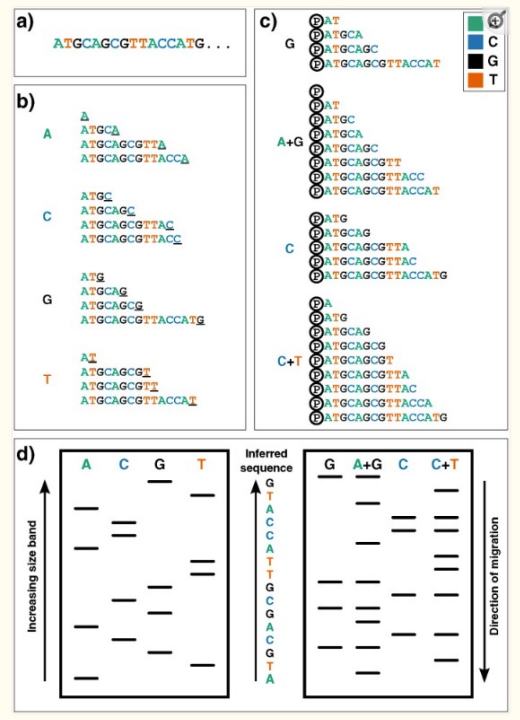
Here's your generated image.



Advent of PAGE, Maxam and Gilbert and Sanger







Sanger

Maxam-Gilbert

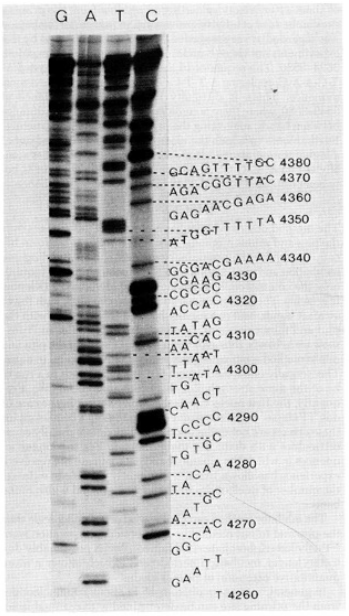
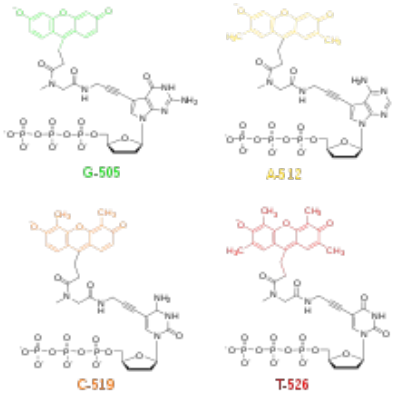


FIG. 2. Autoradiograph from an experiment R4 as primer on the complementary strand of φX174 DNA. Conditions were as in Fig. 1 with the following exceptions used as inhibitor instead of araCTP. After incubation of the solutions at room temperature for 15 min, 1 μl of 0.5 mM dATP and 1 μl of restriction enzyme *Hae* III (4 units/μl) were added and the solutions were incubated at 37° for 10 min. The *Hae* III cuts close to the *Hin*dII site and it was used because it was more readily available. The electrophoresis was on a 12% acrylamide gel at 40 mA for 14 hr. The top 10.5 cm of the gel is not shown.

J. Mol. Biol. (1975) 94:441, Genomics (2016) 107(1):1, PNAS 74:560, PNAS 74:5463

ABI 370 prototype





Abstract

References

eLetters (0)

Large-Scale and Automated DNA Sequence Determination

T. HUNKAPILLER, R. J. KAISER, B. F. KOOP, AND L. HOOD Authors Info & Affiliations

SCIENCE • 4 Oct 1991 • Vol 254, Issue 5028 • pp. 59-67 • DOI: 10.1126/science.1925562

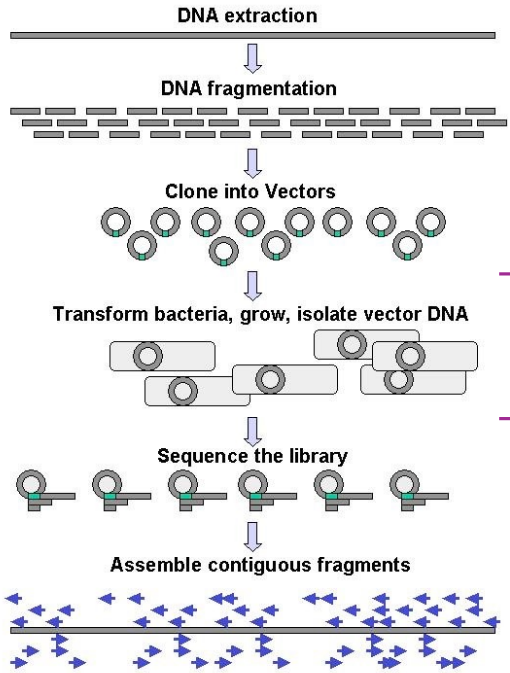
89 4 CHECK ACCESS

Abstract

DNA sequence analysis is a multistage process that includes the preparation of DNA, its fragmentation and base analysis, and the interpretation of the resulting sequence information. New technological advances have led to the automation of certain steps in this process and have raised the possibility of large-scale DNA sequencing efforts in the near future [for example, 1 million base pairs (Mb) per year]. New sequencing methodologies, fully automated instrumentation, and improvements in sequencing-related computational resources may render genome-size sequencing projects (100 Mb or larger) feasible during the next 5 to 10 years.

Science (1991) 254:59

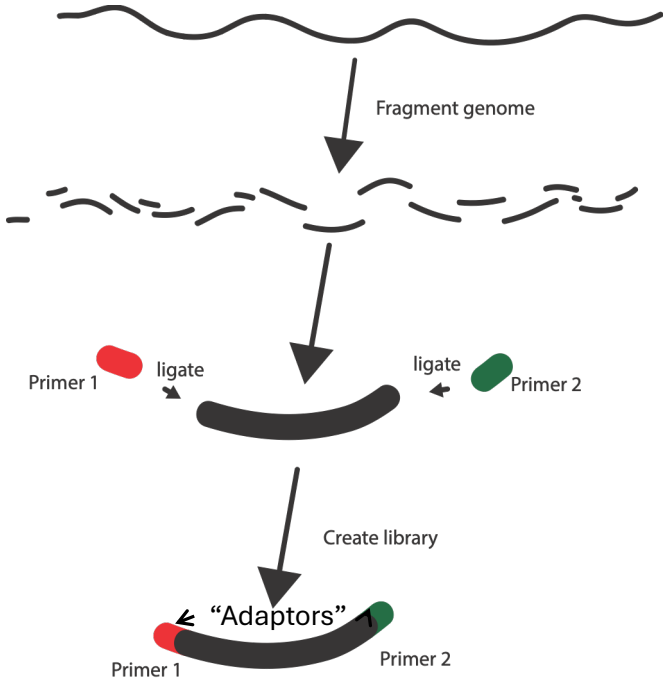
Sequencing the human genome



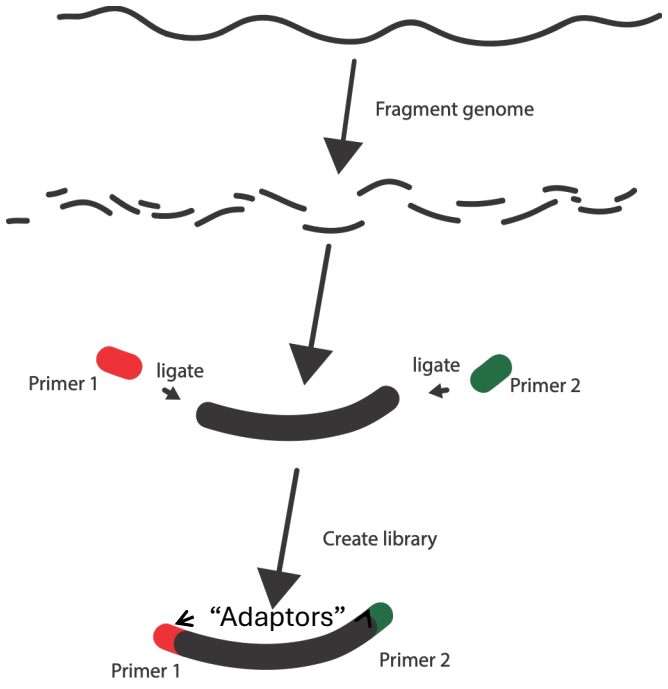
LIBRARY = a collection of genetic material fragments that can be stored and replicated

clonal amplification of fragments

Generating a sequencing library

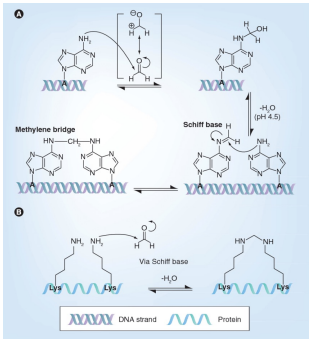


Source material considerations



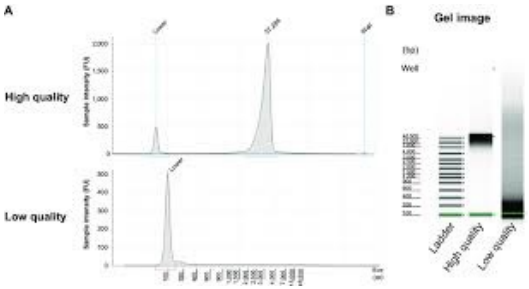
Source of nucleic acid? Quantity available?

Quality?

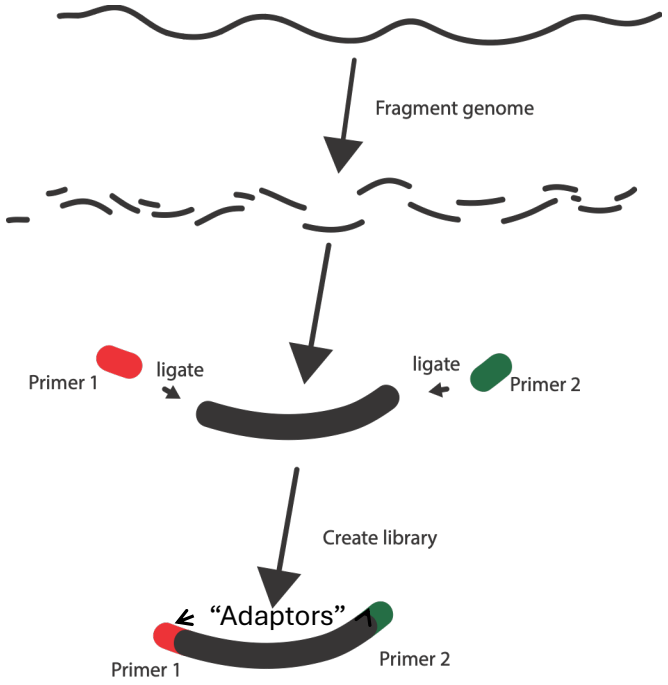


Delrue et al 2012





Ways to fragment DNA

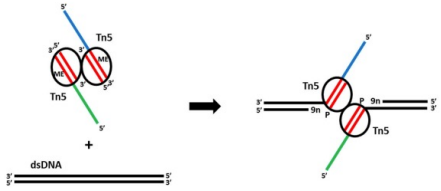


Fragmentation technique?



Sonication

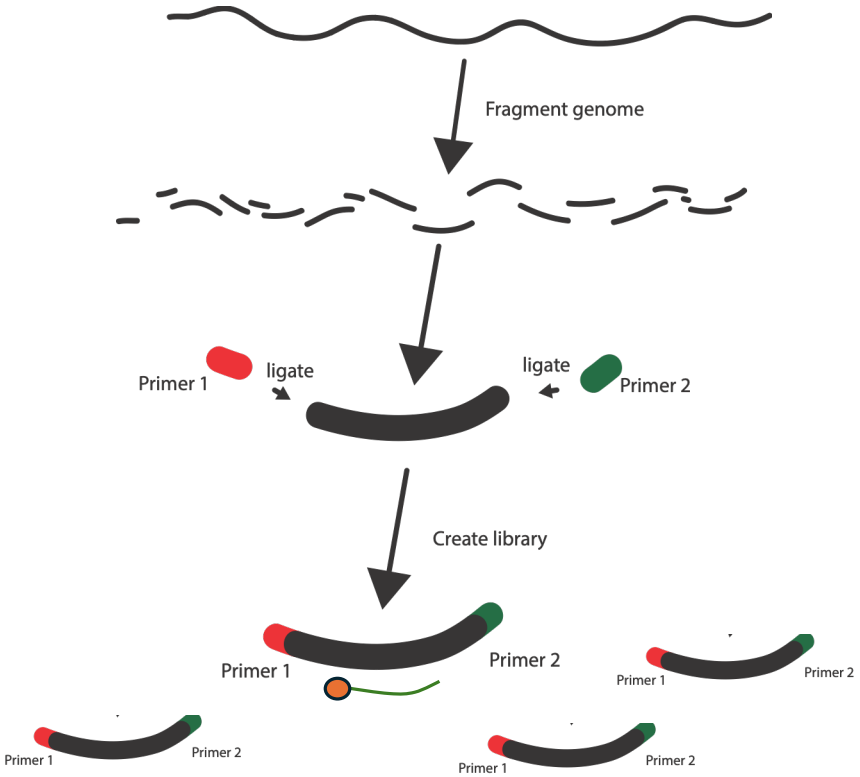
Covaris



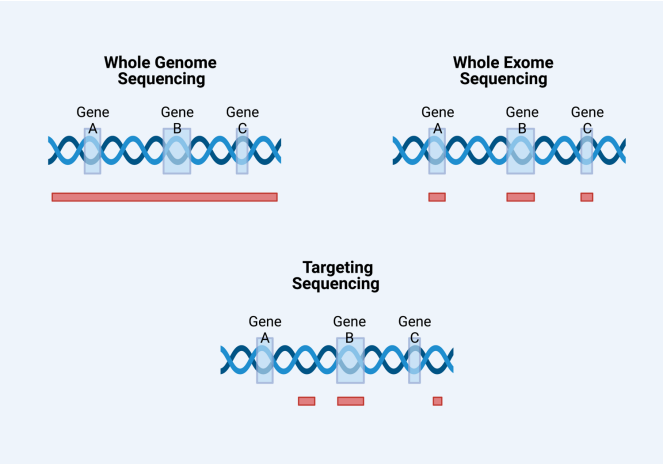
Penkov et al 2023

“Tagmentation”

What is the region of interest

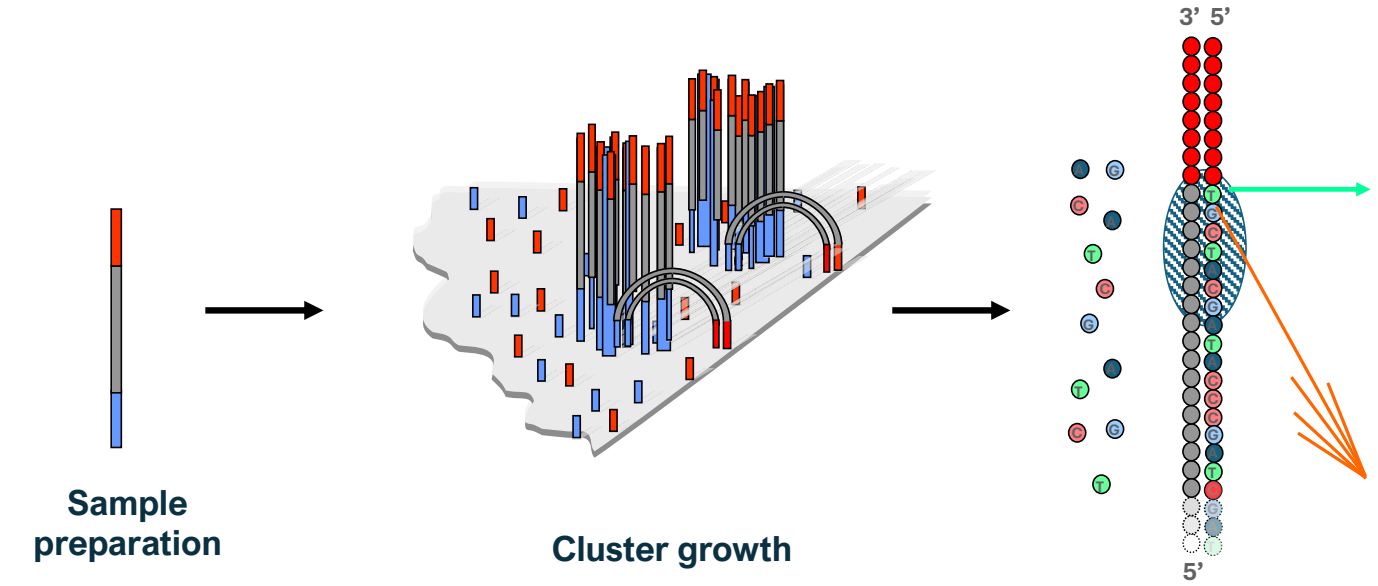


Every single fragment from the cell? Or enrich areas of interest? What's in a genome?



Telomeres
Centromeres
rDNA
Introns
Exons
Promoters
Repetitive DNA
"Junk DNA"
Transposable elements

Sequencing by synthesis



Sample preparation

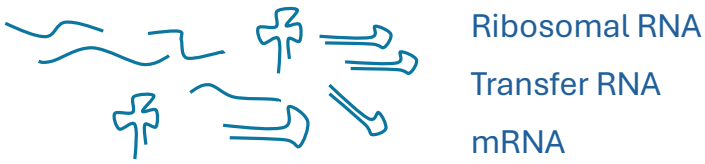
Cluster growth

Sequencing

Image acquisition

Base calling

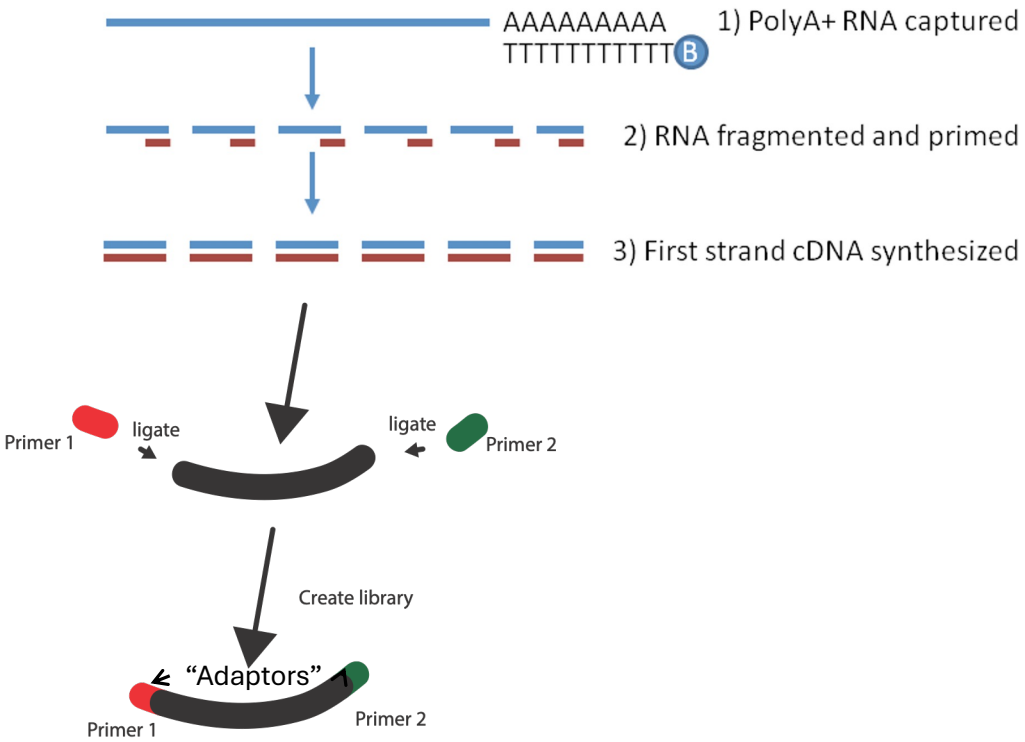
Sequencing RNA



Ribosomal RNA

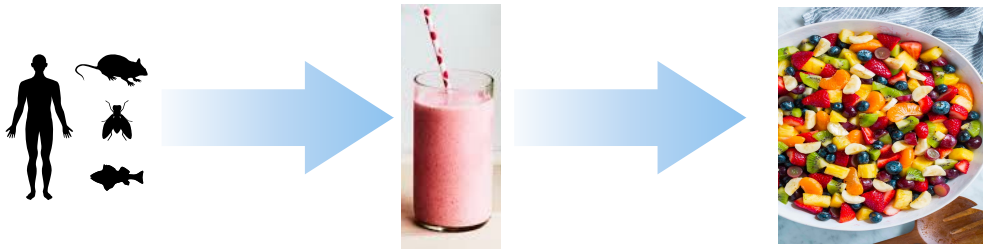
Transfer RNA

mRNA

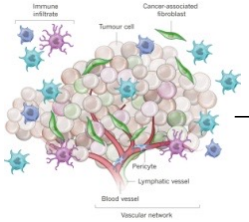




How about single-cell RNA-seq?



Global RNAseq tells a very different story than single cell RNAseq



Total RNA → RNAseq

Gene	Expression level
Gene 1	10,000
Gene 2	1,000
Gene 3	5,000
Gene 4	200*
Gene 5	0

*:limit of detection

Gene	Expression level
Gene 1	0
Gene 2	2000
Gene 3	1000
Gene 4	800
Gene 5	0

Gene	Expression level
Gene 1	0
Gene 2	2000
Gene 3	1000
Gene 4	0
Gene 5	0









Gene	Expression level
Gene 1	60,000
Gene 2	0
Gene 3	1000
Gene 4	200
Gene 5	0

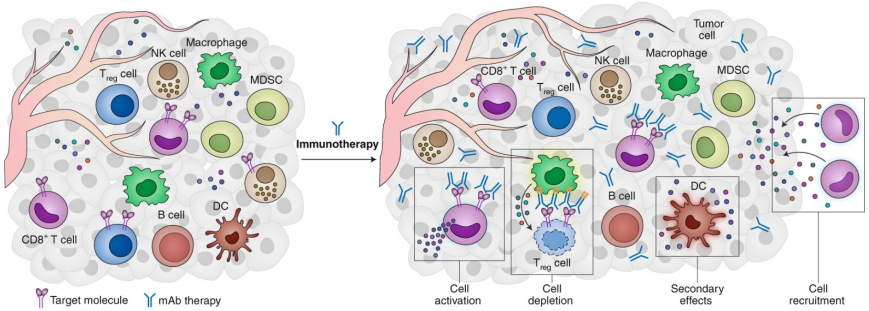
Gene	Expression level
Gene 1	0
Gene 2	1000
Gene 3	1000
Gene 4	0
Gene 5	250

Gene	Expression level
Gene 1	0
Gene 2	0
Gene 3	21 000
Gene 4	0
Gene 5	0

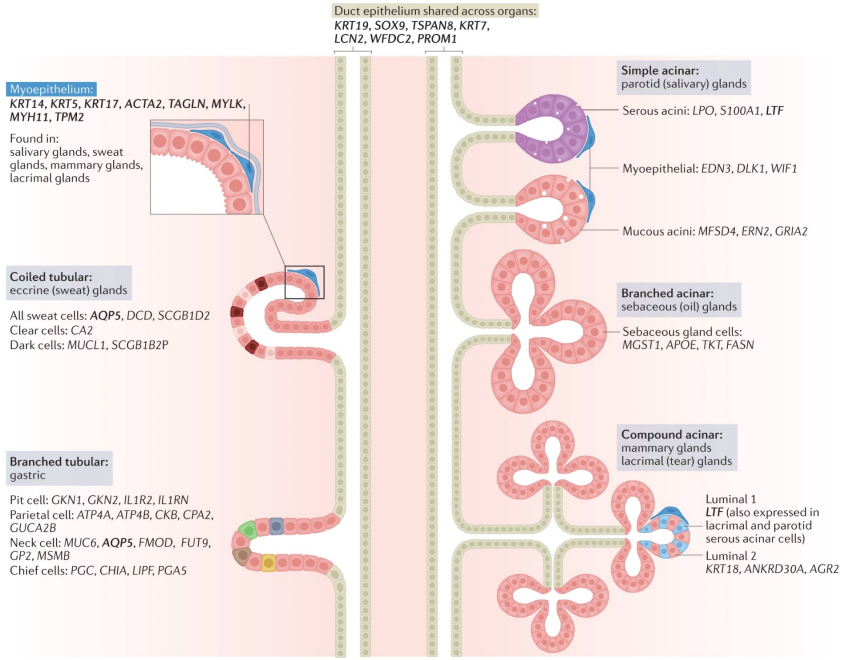
Global RNAseq tells a very different story than single cell RNAseq



Lineage tracing – cells grouped according to similarity, reveal developmental relationships



Understanding disease mechanisms, cellular crosstalk



Cell atlas – revelation of specialized cell types in organs

Nat. Med. **26**, 171–177 (2020)
Nat. Rev. Genet. **23**, 395–410 (2022).
Nat. Rev. Genet. **21**, 410–427 (2020).

Single cell RNA seq timeline

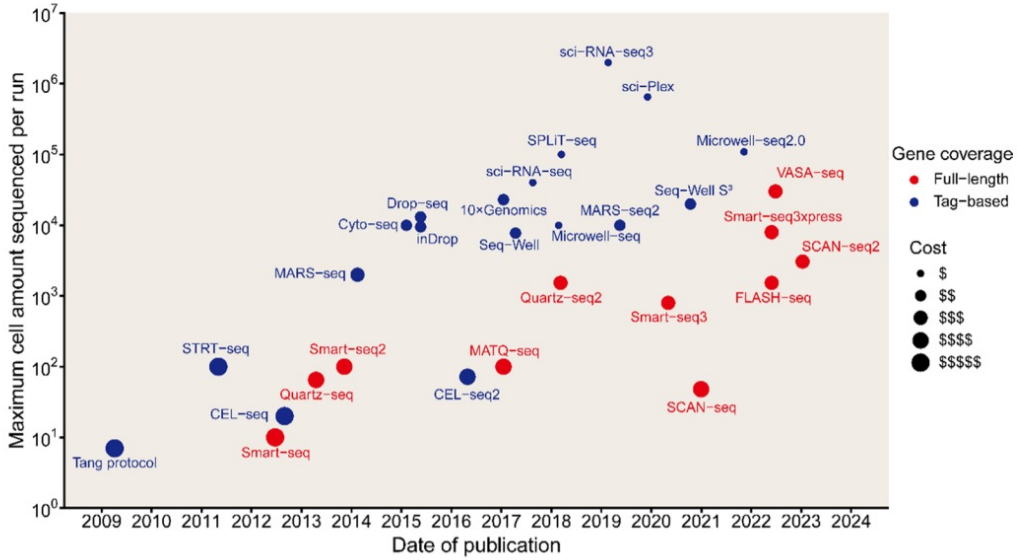
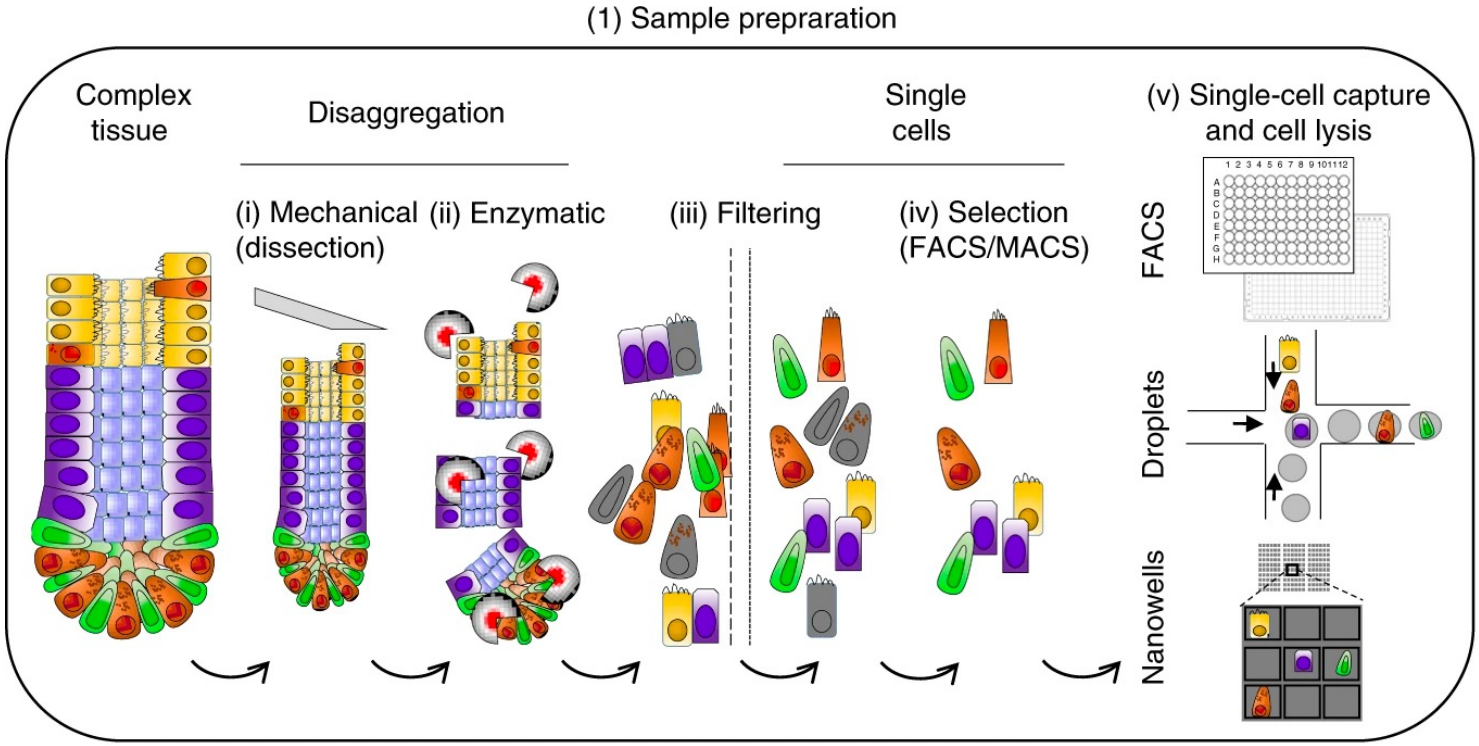


Fig. 2 Timeline and throughput of various scRNA-seq methods. Scatterplot depicts the published date and throughput of sequencing for each technology. The color indicates the different gene coverage. Size indicates the cost per sequenced cell of scRNA-seq methods

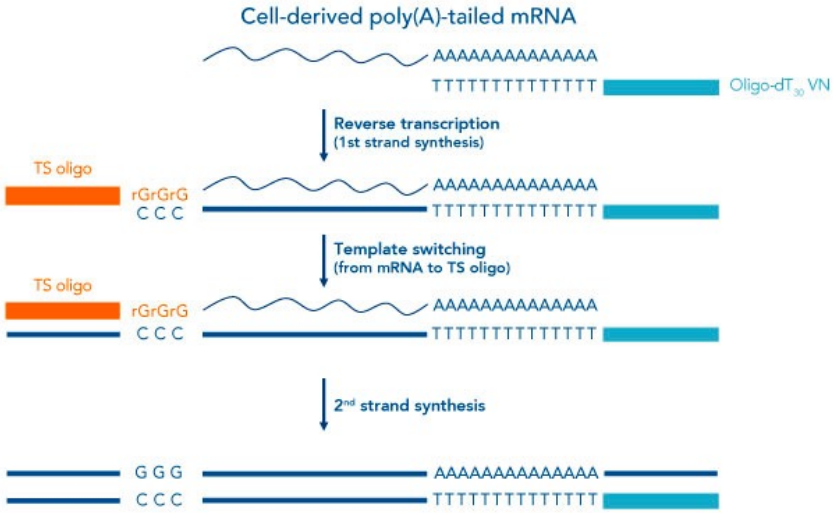
Single cell wet lab workflow

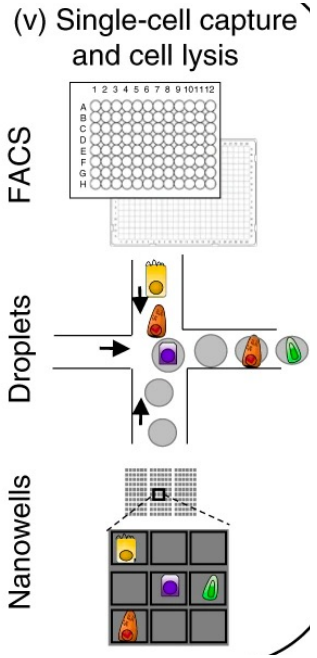


Lafzi et al Nature Prot 2018

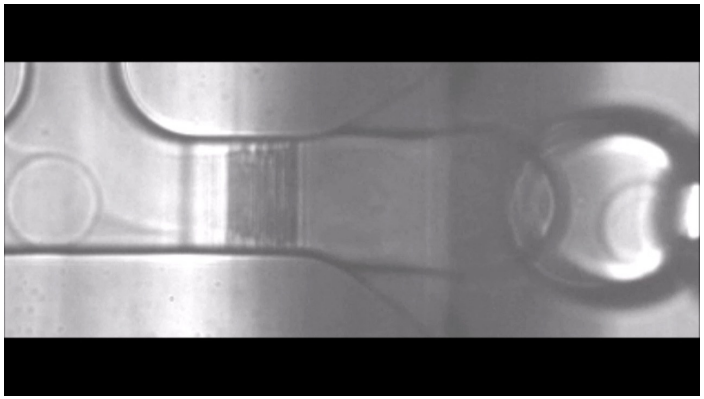
What could go wrong?

Template-switching

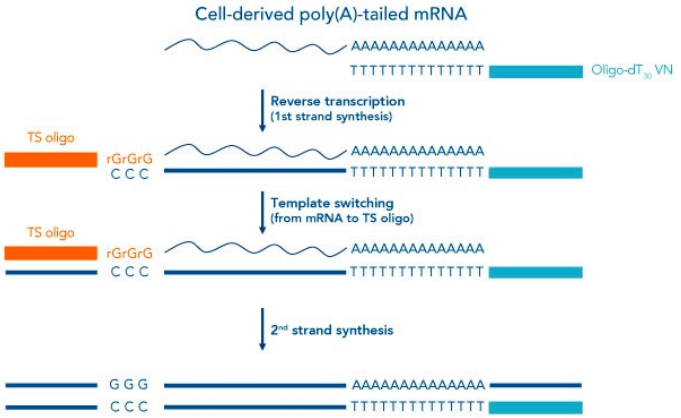


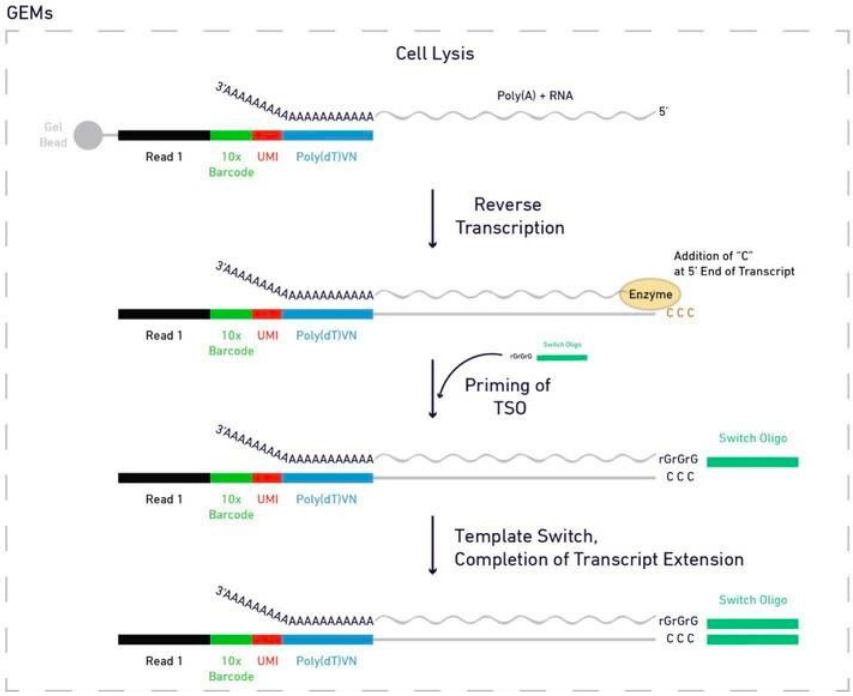


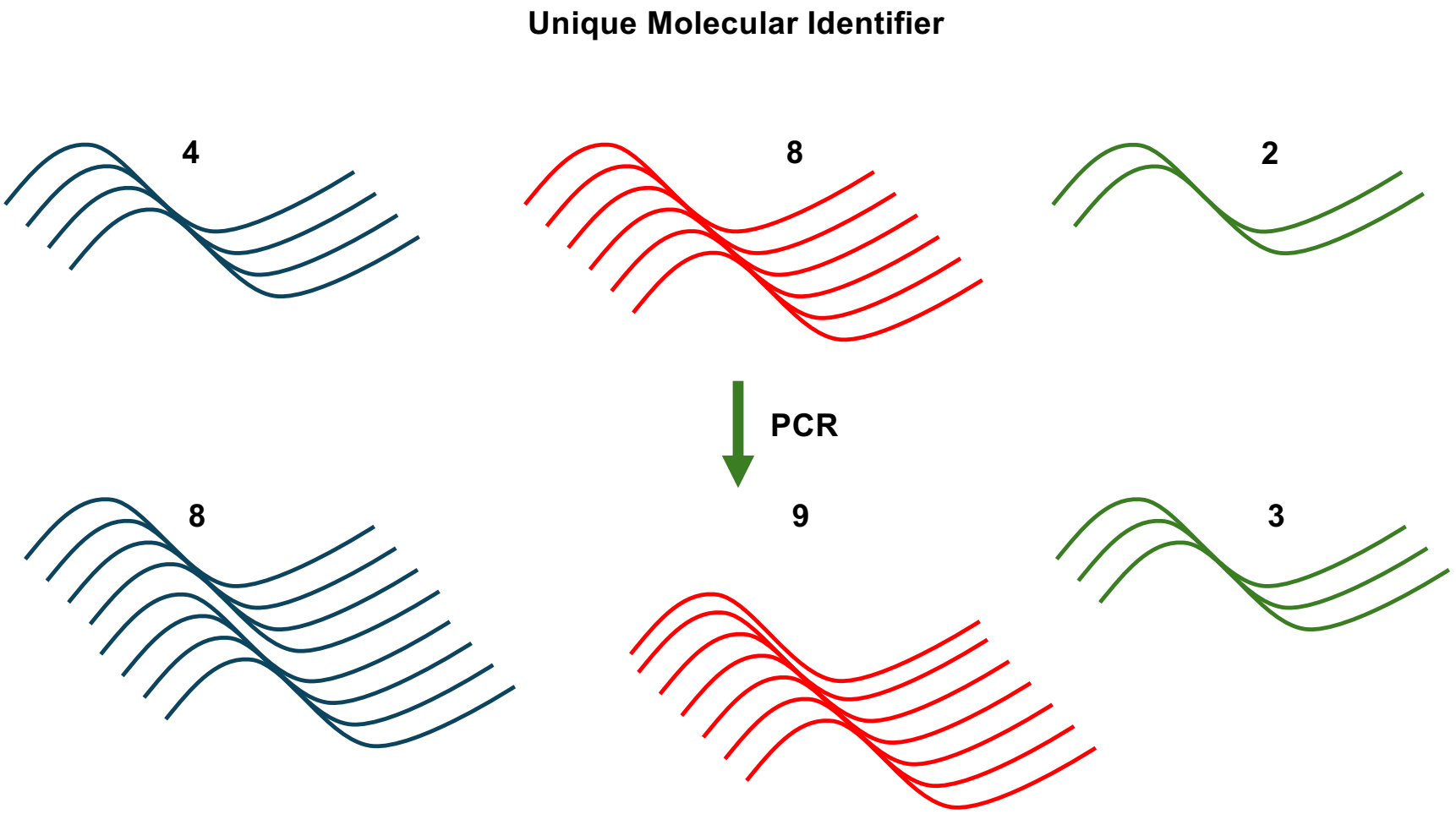
10X genomics

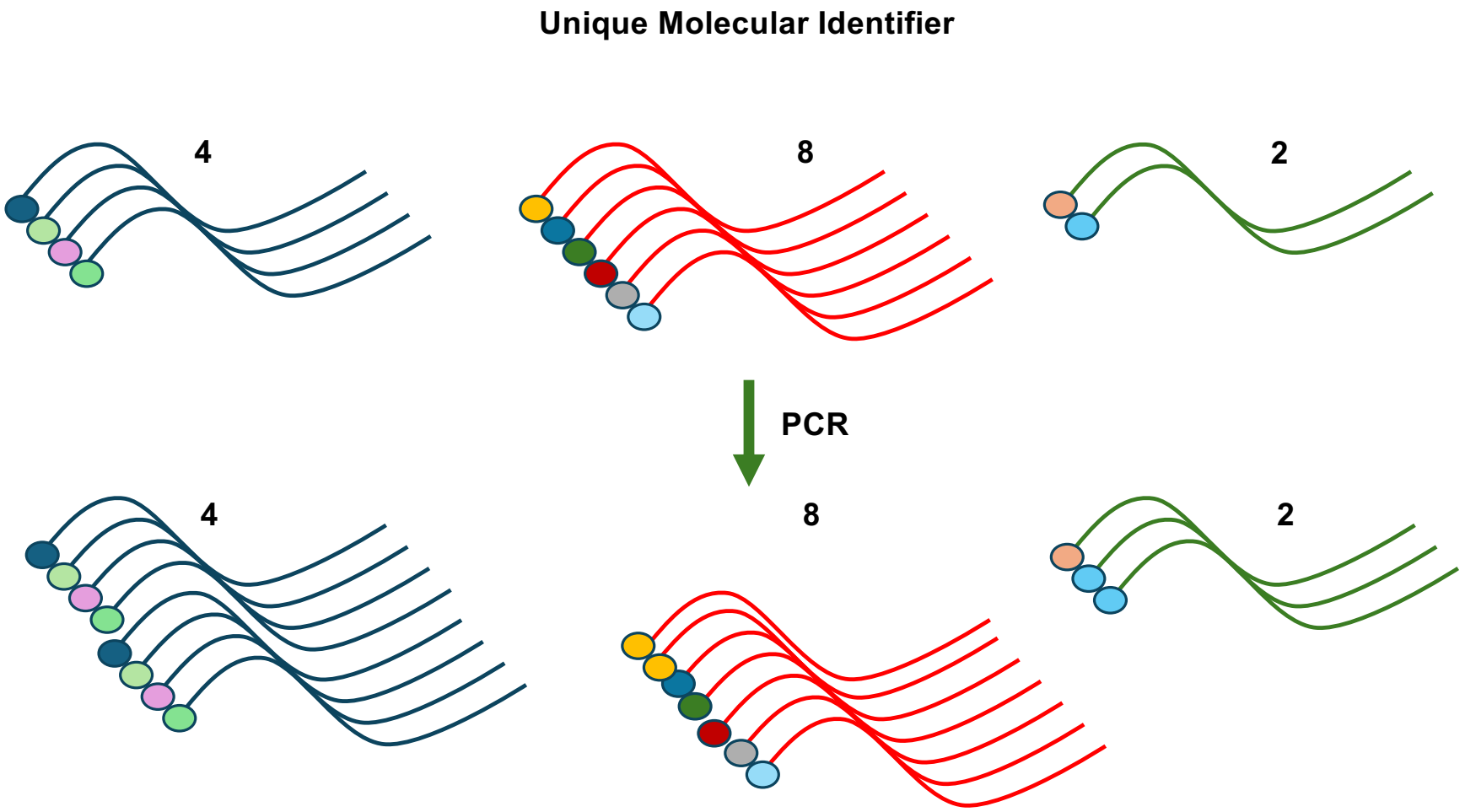


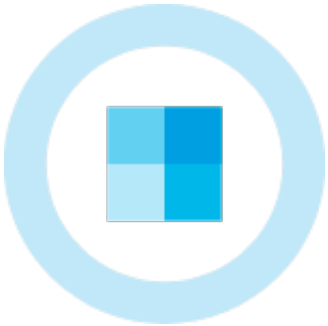
Template-switching with cell barcode





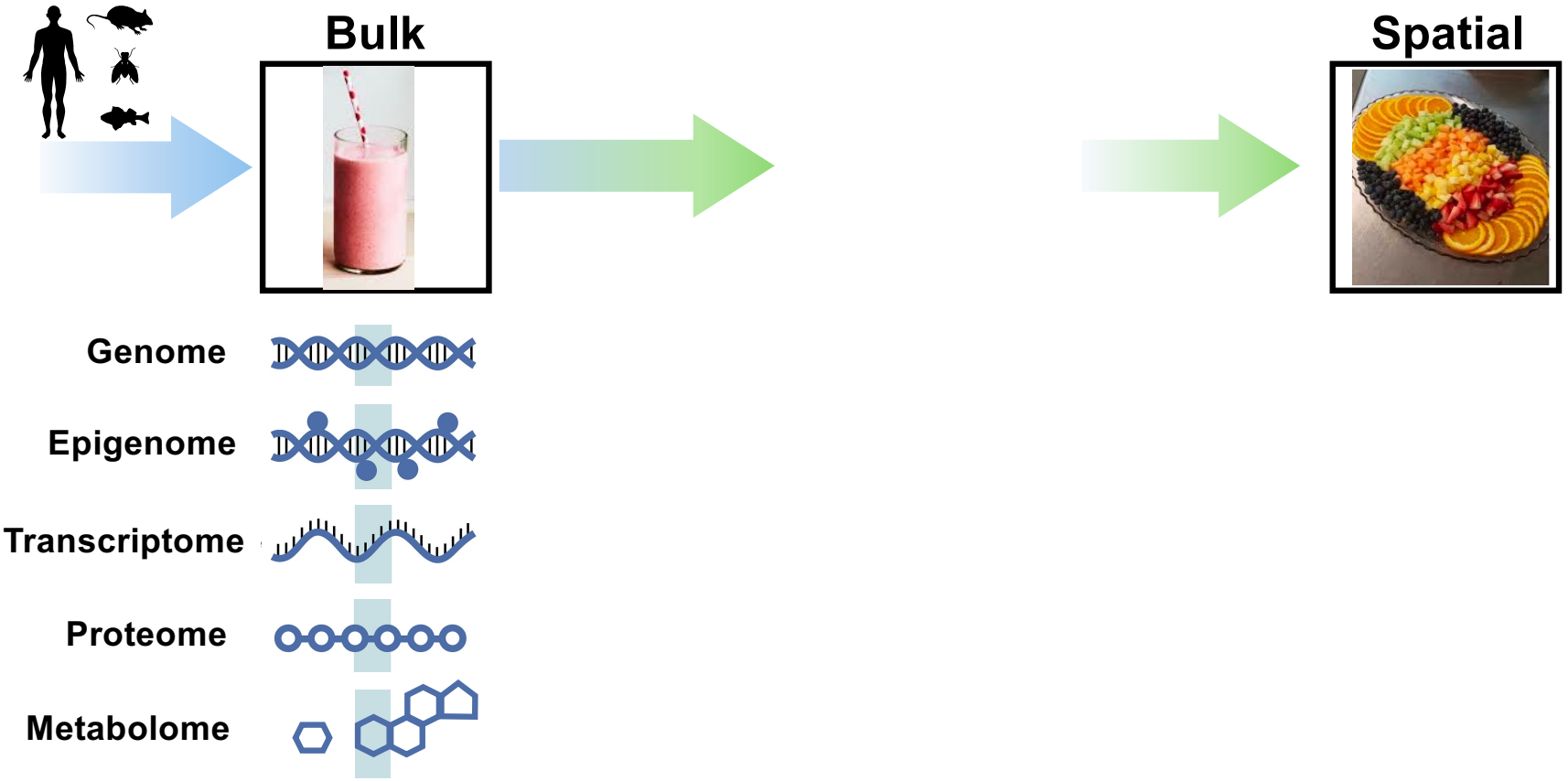

Unique Molecular Identifier
4
8
2
PCR
8
9
3


Unique Molecular Identifier
4
8
2
PCR
4
8
2
counting UMIs instead of transcripts



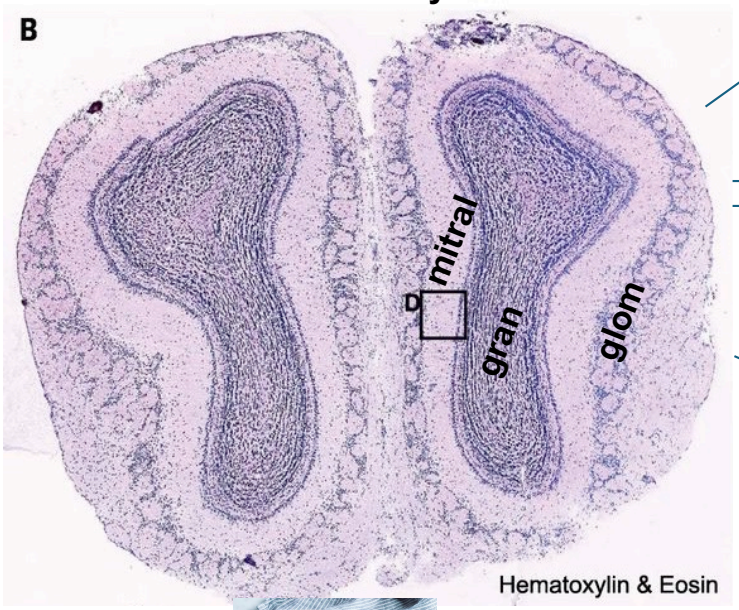
SPATIAL GENE EXPRESSION

Visualize Gene Expression within Tissue Organization


Bulk
Spatial
Genome
Epigenome
Transcriptome
Proteome
Metabolome

Spatial transcriptomics

Mouse olfactory bulb



Bulk RNAseq

In situ RNAseq
Microdissect and RNAseq

Dissociate tissue and single cell RNAseq

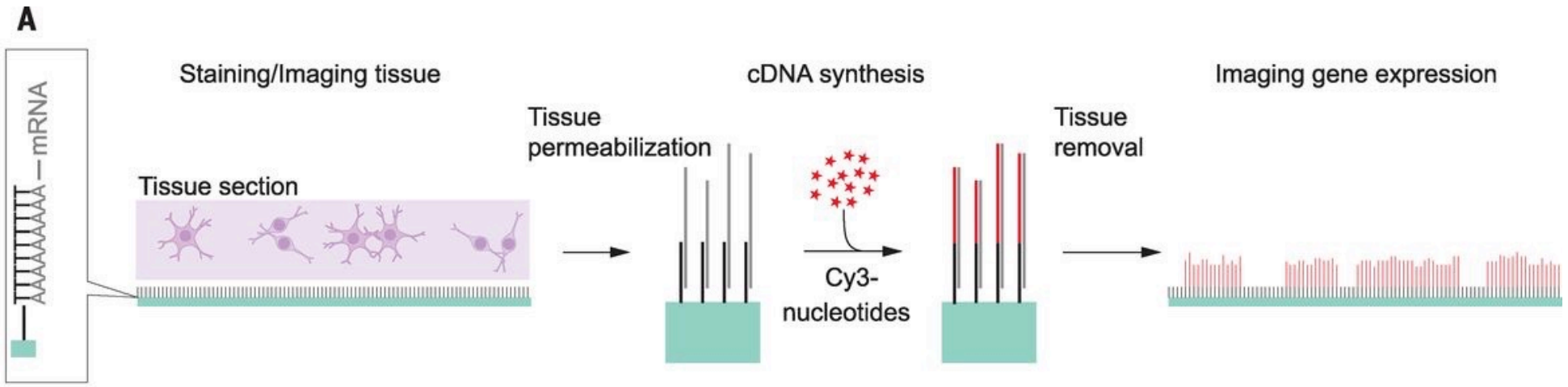




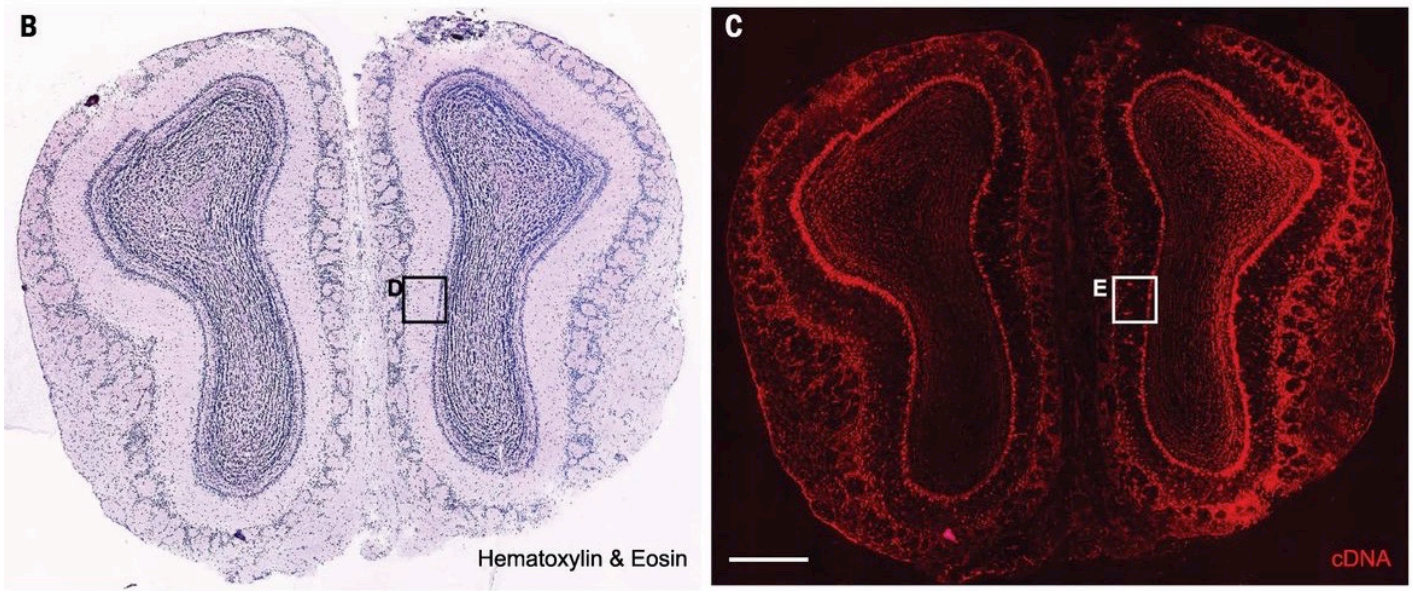


Patrik L. Ståhl et al. Science 2016;353:78-82

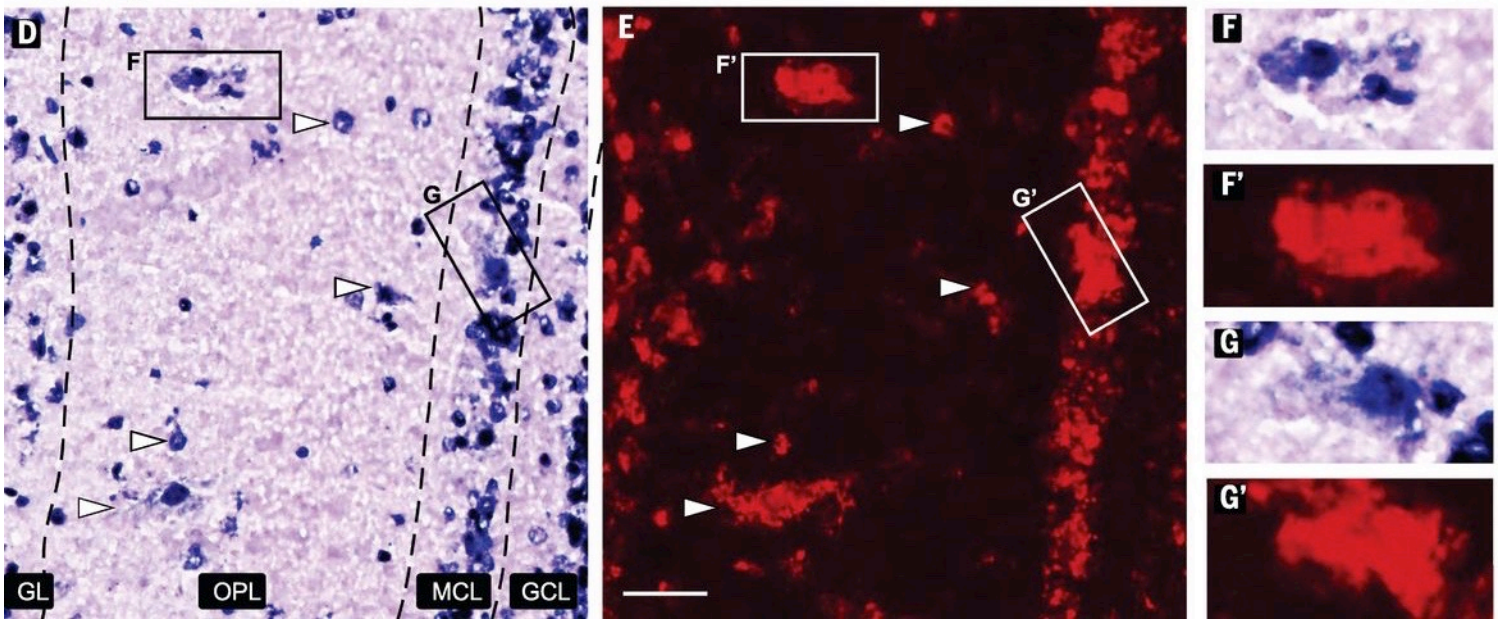
Spatial transcriptomics workflow



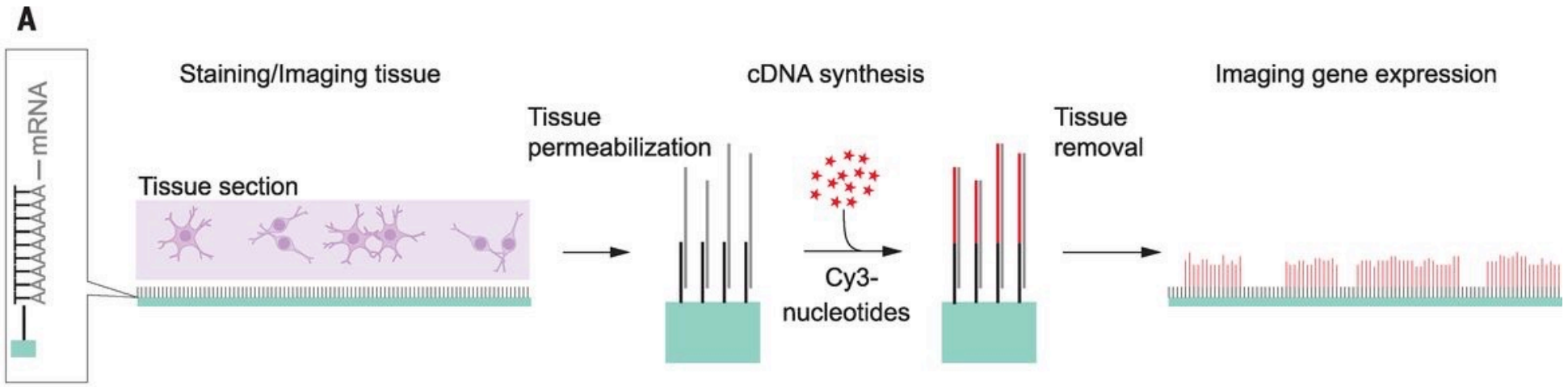
Patrik L. Ståhl et al. Science 2016;353:78-82



Patrik L. Ståhl et al. Science 2016;353:78-82

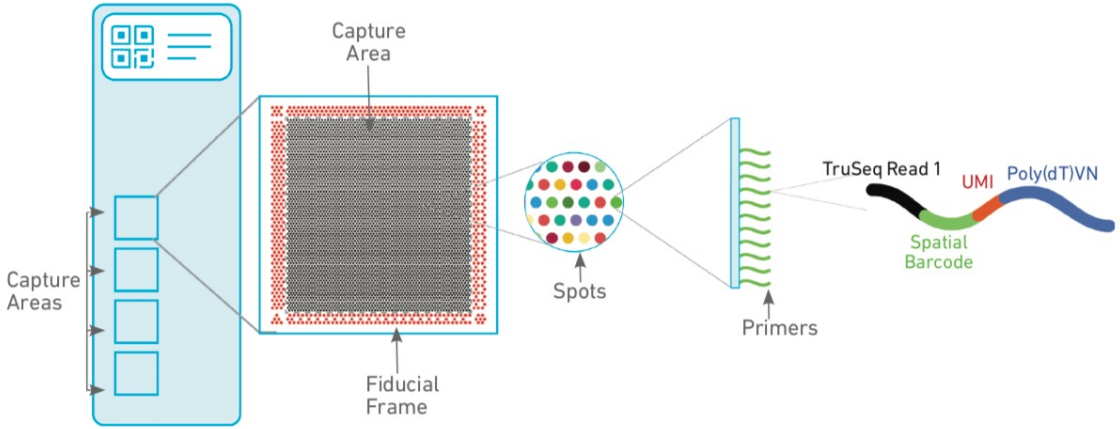


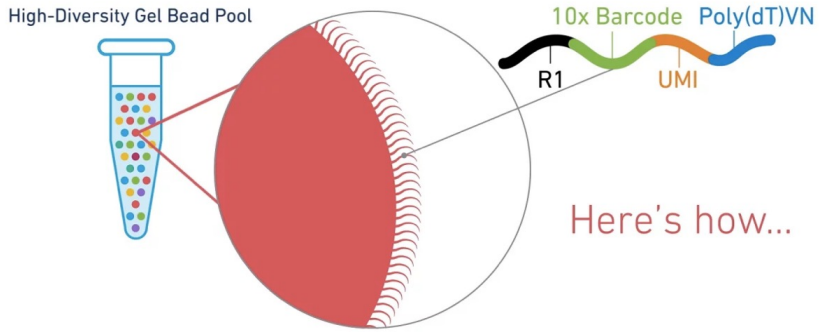
Patrik L. Ståhl et al. Science 2016;353:78-82



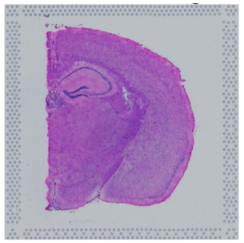
Patrik L. Ståhl et al. Science 2016;353:78-82

10x genomics Visium

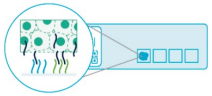


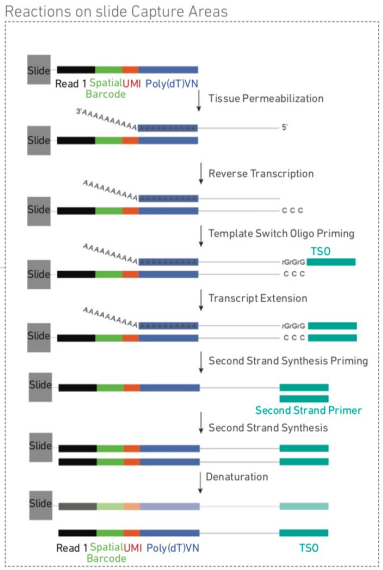


Place tissue on slide
H&E stain

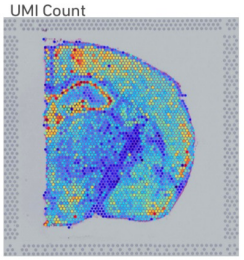


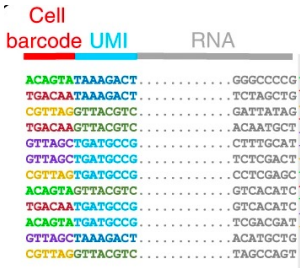
Permeabilize and capture
H&E stain



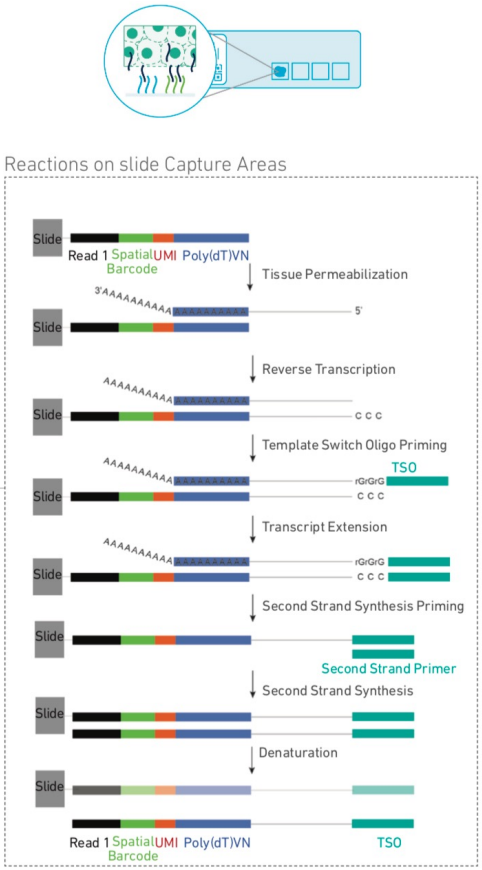


Prepare libraries
Sequence



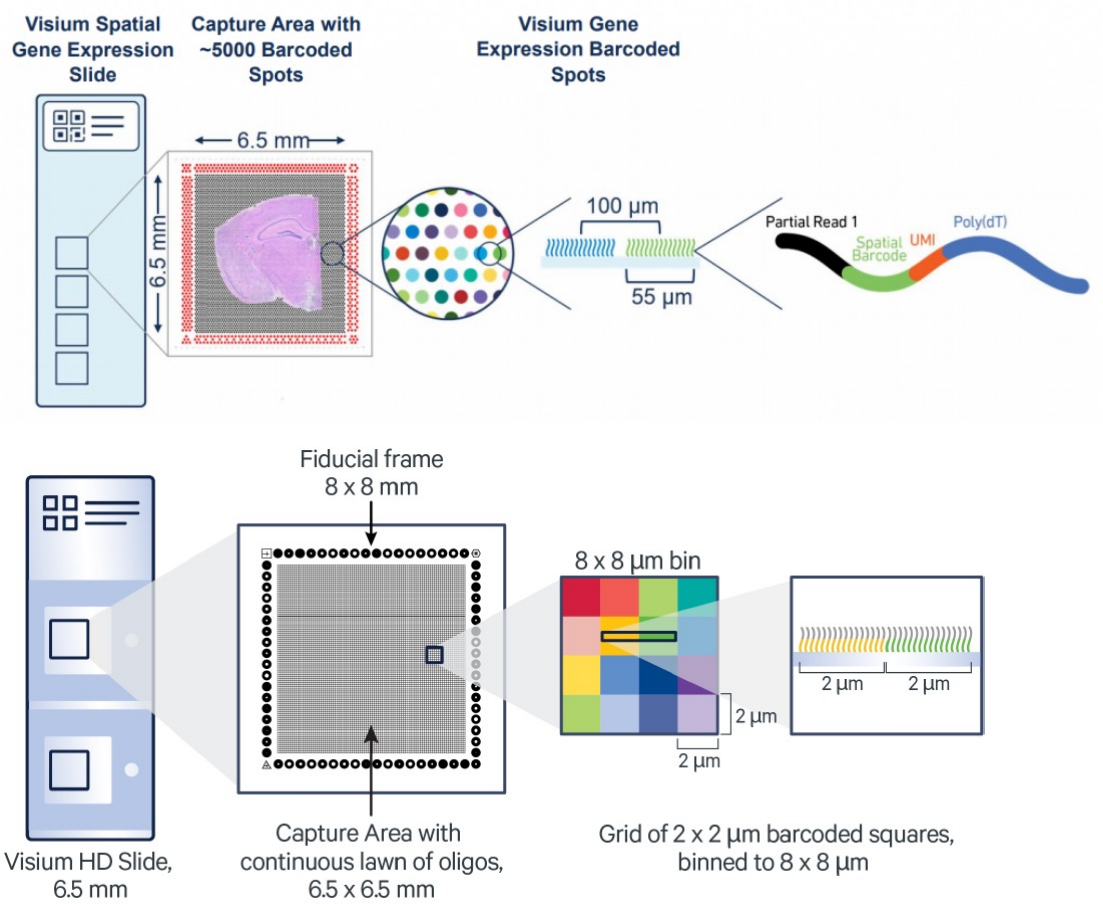


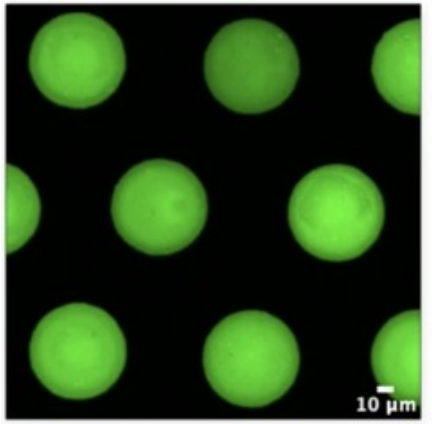
How to make this compatible for FFPE?

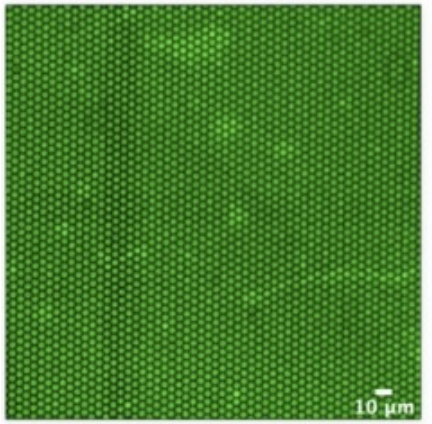




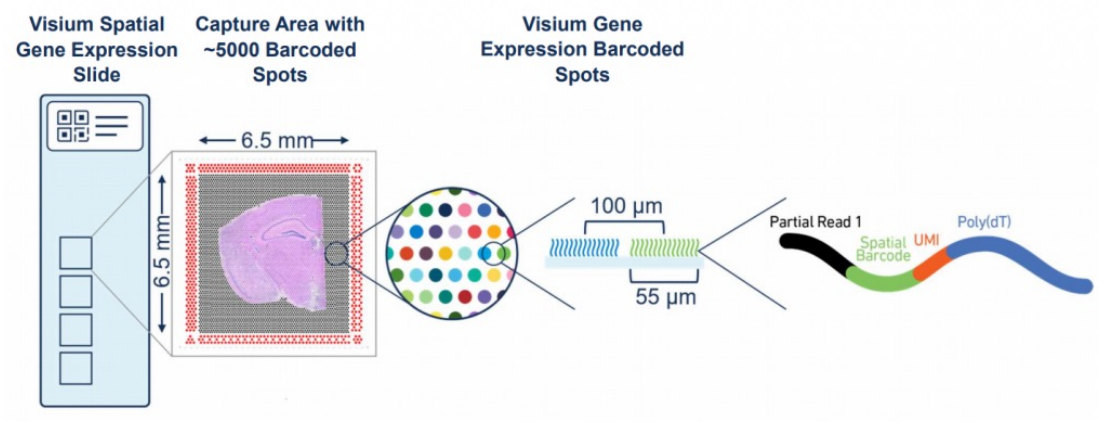
Visium HD

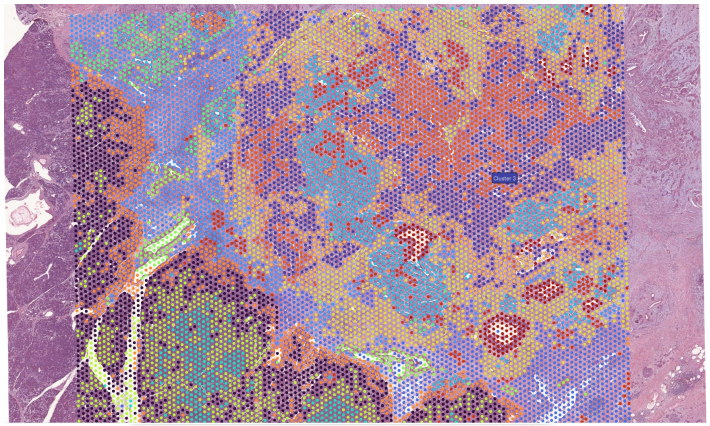

Visium Spatial Gene Expression Slide
Capture Area with ~5000 Barcoded Spots
Visium Gene Expression Barcoded Spots
6.5 mm
6.5 mm
100 µm
55 µm
Partial Read 1
Spatial Barcode
UMI
Poly(dT)
Fiducial frame 8 x 8 mm
8 x 8 µm bin
2 µm
2 µm
2 µm
2 µm
Visium HD Slide, 6.5 mm
Capture Area with continuous lawn of oligos, 6.5 x 6.5 mm
Grid of 2 x 2 µm barcoded squares, binned to 8 x 8 µm

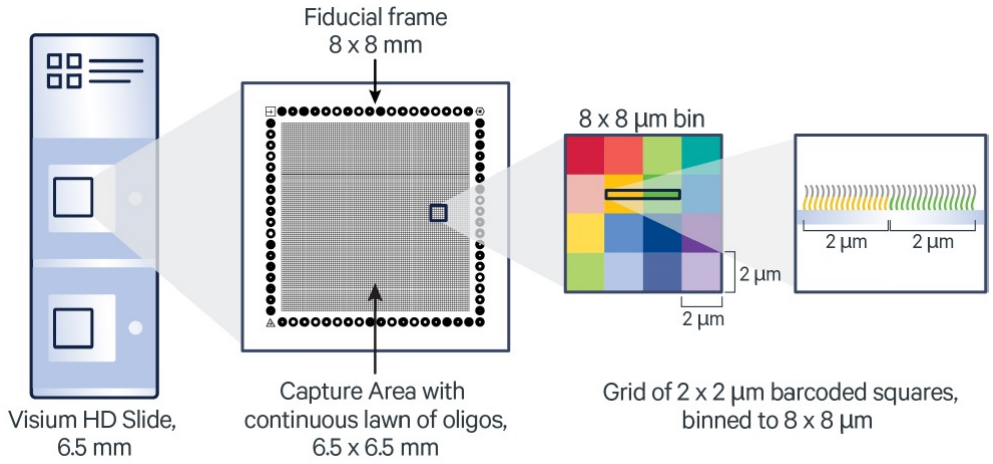

10 µm

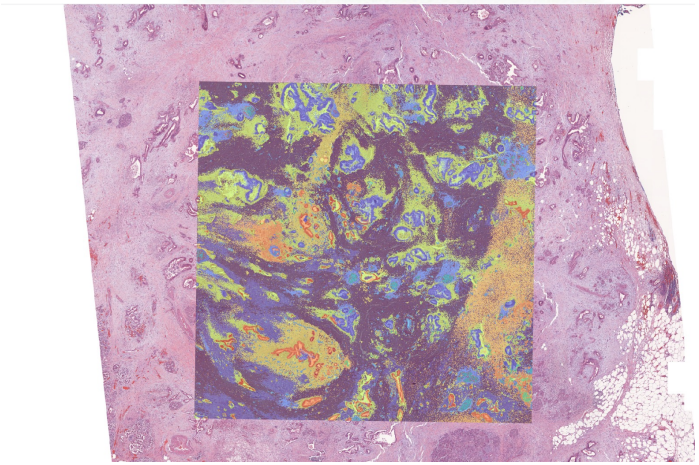

10 µm

Visium HD


Visium Spatial Gene Expression Slide
Capture Area with ~5000 Barcoded Spots
Visium Gene Expression Barcoded Spots
6.5 mm
6.5 mm
100 µm
55 µm
Partial Read 1
Spatial Barcode
UMI
Poly(dT)




Fiducial frame
8 x 8 mm
8 x 8 µm bin
2 µm
2 µm
2 µm
2 µm
Visium HD Slide, 6.5 mm
Capture Area with continuous lawn of oligos, 6.5 x 6.5 mm
Grid of 2 x 2 µm barcoded squares, binned to 8 x 8 µm



10x genomics Xenium/Atera
Imaging-based transcriptomics



Fig 1: Atera profiler dimensions

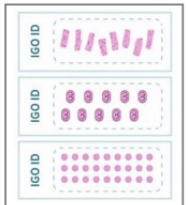


Fig. 2A (above): schematic of slides showing possible tissue placement including tissue micro-array on the bottom can maximize usage of the slide.

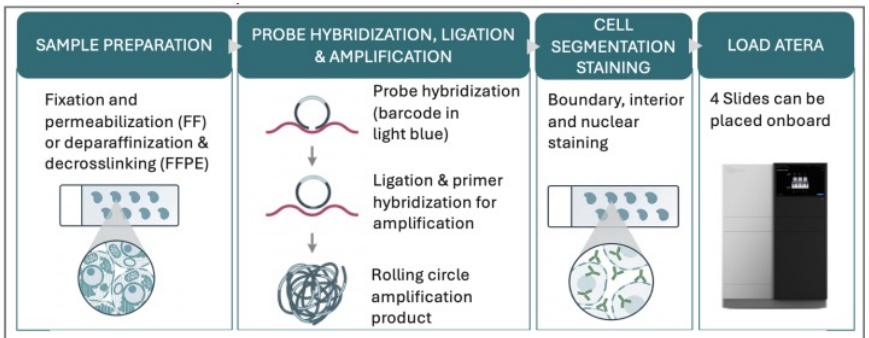


Fig. 2B (from left to right): in panel 1 slides are treated according to tissue preservation, then incubated with probes as seen in panel 2. Hybridization to the target (purple) is followed by rolling circle amplification. Panel 3 depicts staining with cell segmentation markers after which the slides are loaded on to the Atera (panel 4)

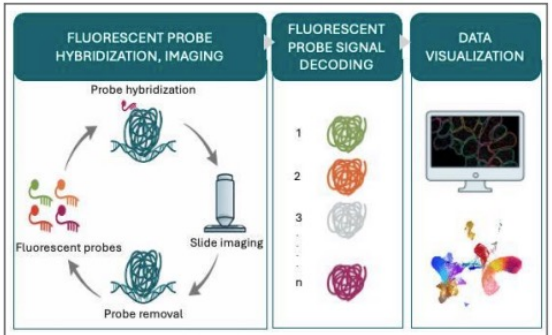
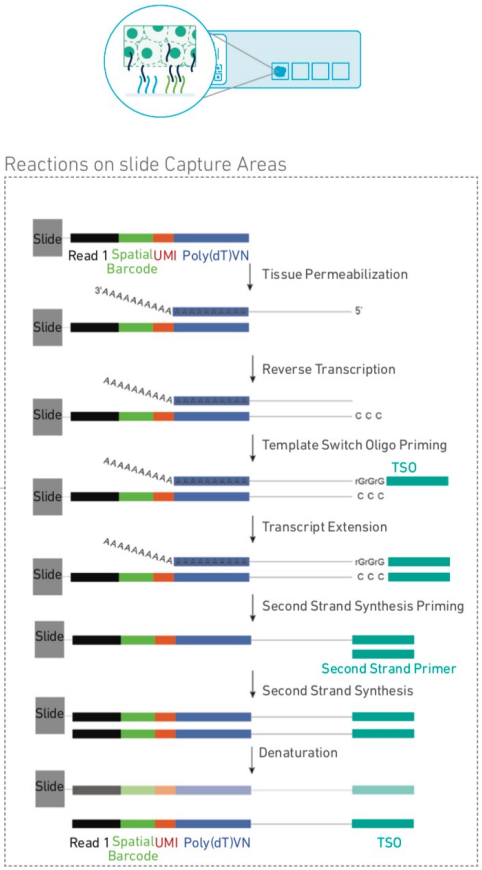


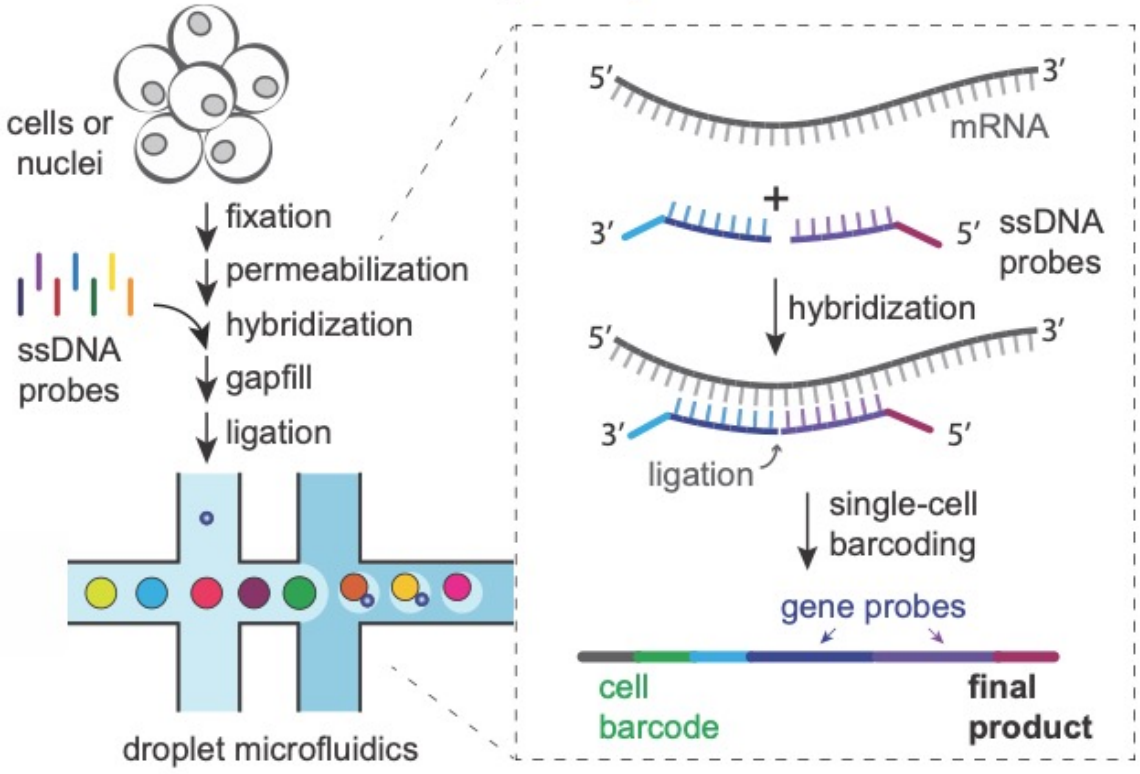
Fig. 3: probe detection workflow on board the Atera: the first panel shows cycles where fluorescent detection probes hybridize to the rolling amplification product. The signal from the probes is deconvoluted and assigned to an RNA target (middle panel). Data may then be visualized interactively using 10x Genomics software (also see Figure 4 below).

Can you use probes for fixed single cells?



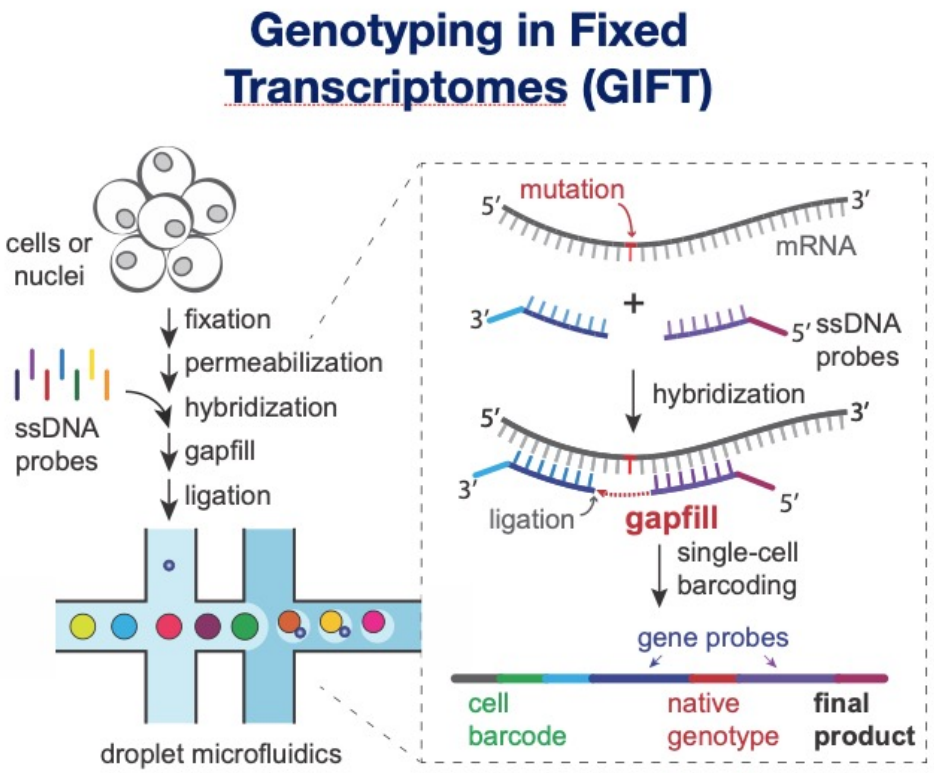


Probe-based gene expression – 10x Flex



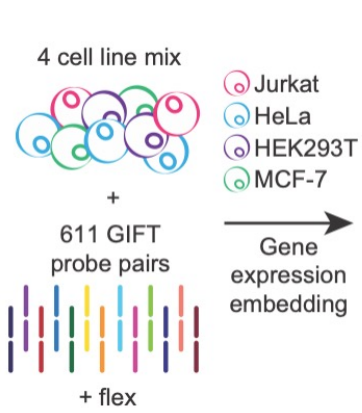
Slides by Sydney Blattman

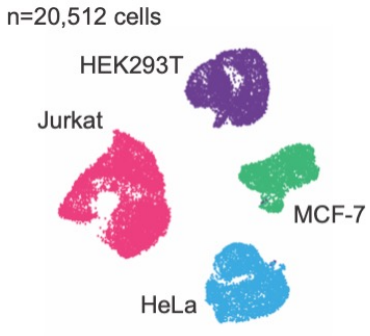
Blattman, Maslah, Varela et al Biorxiv 2026

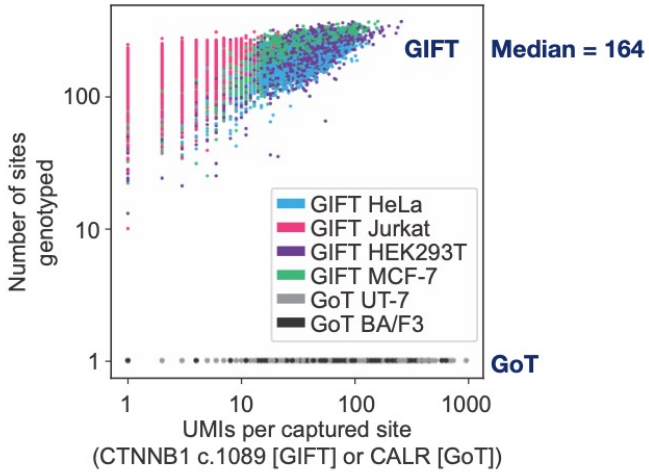


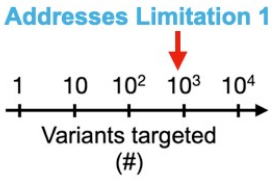
Slides by Sydney Blattman

GIFT readily scales to hundreds of targets



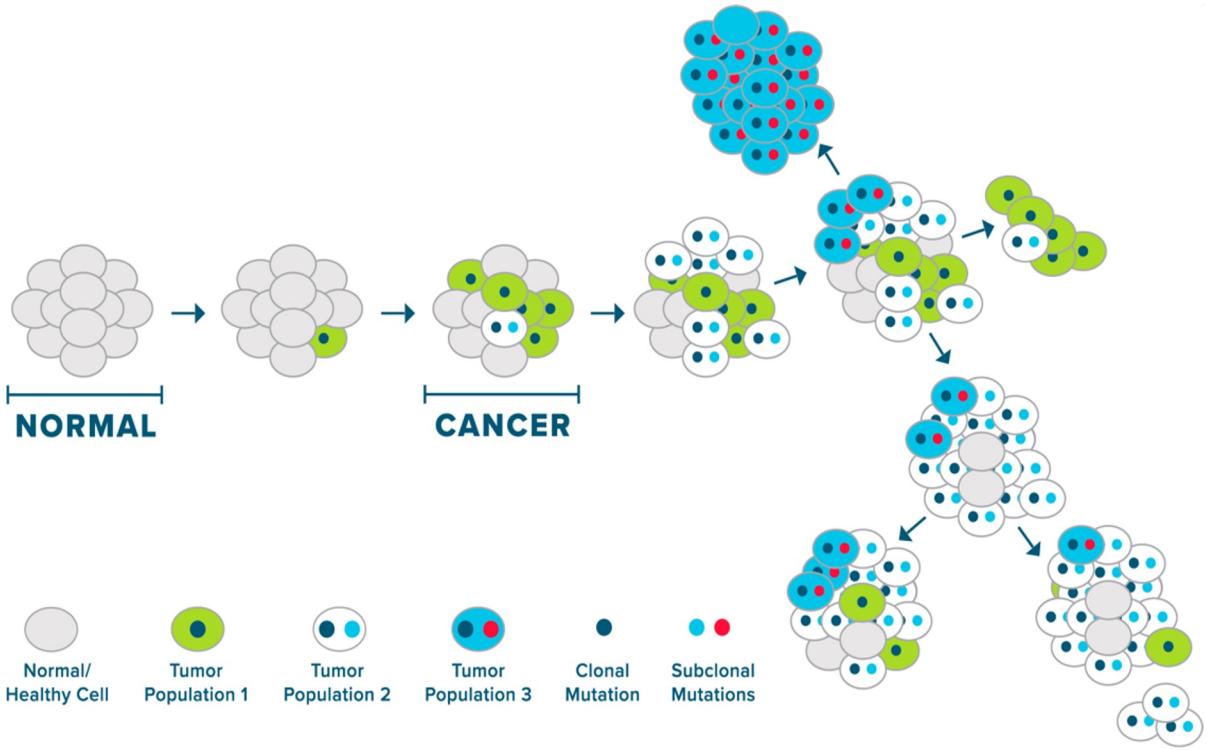






Slides by Sydney Blattman

Single cell DNA studies



Direct Library Prep

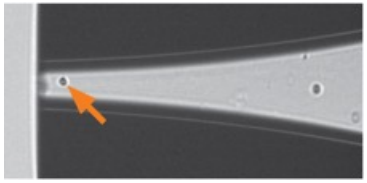


Contactless piezo electric dispenser

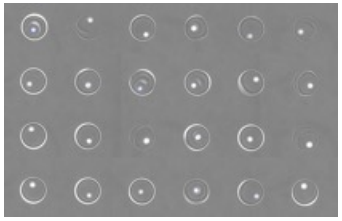
high throughput dispensing of individual cells from cell suspensions
50 pl to 1 ul



Direct Library Prep

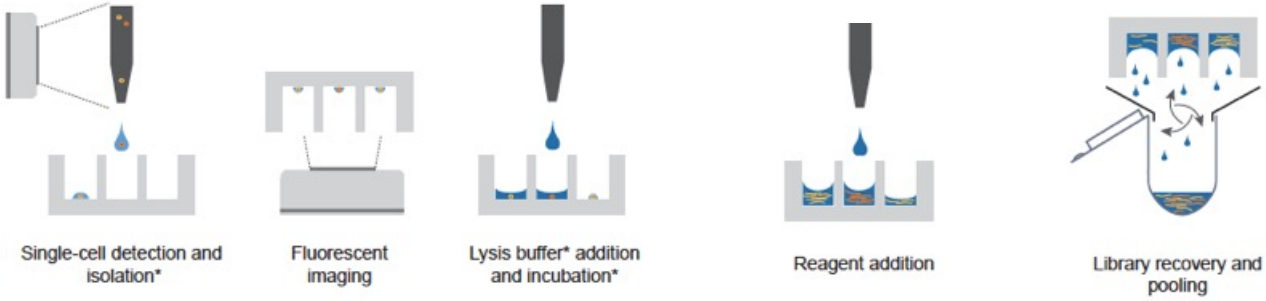


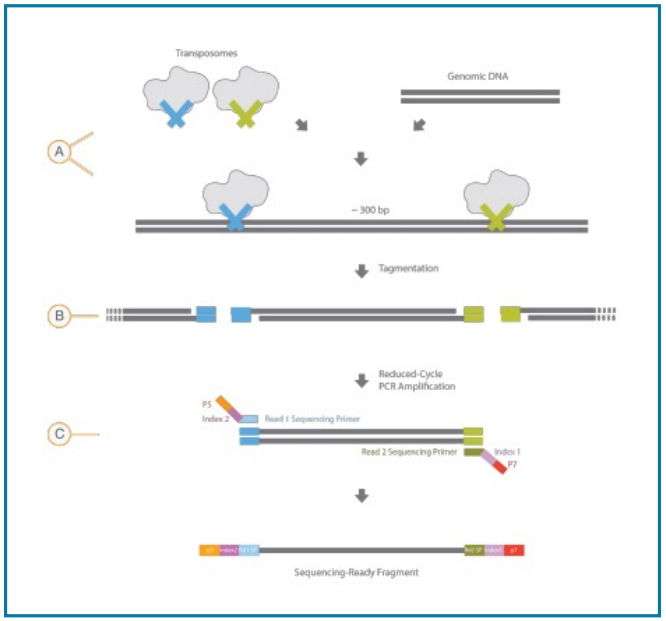
Real-time object recognition during dispensation
On-board fluorescent module



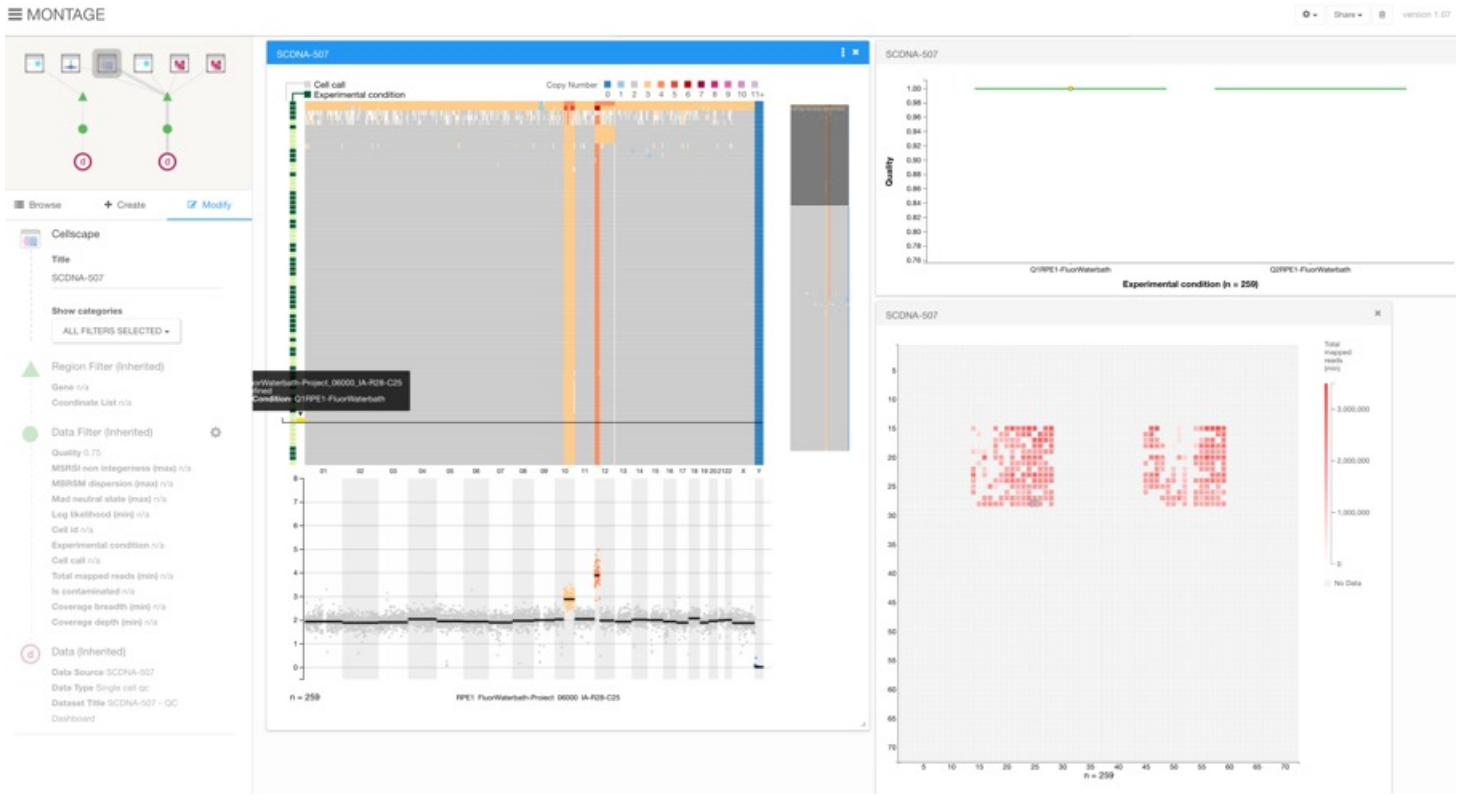
Imaging of cells on chip
Link cell state to sequencing data

Single cell DNA libraries

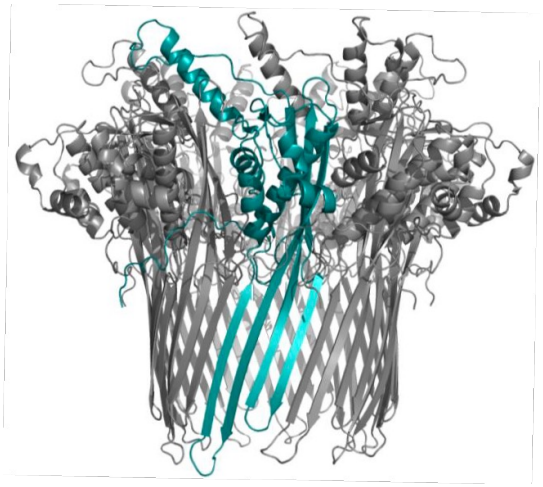


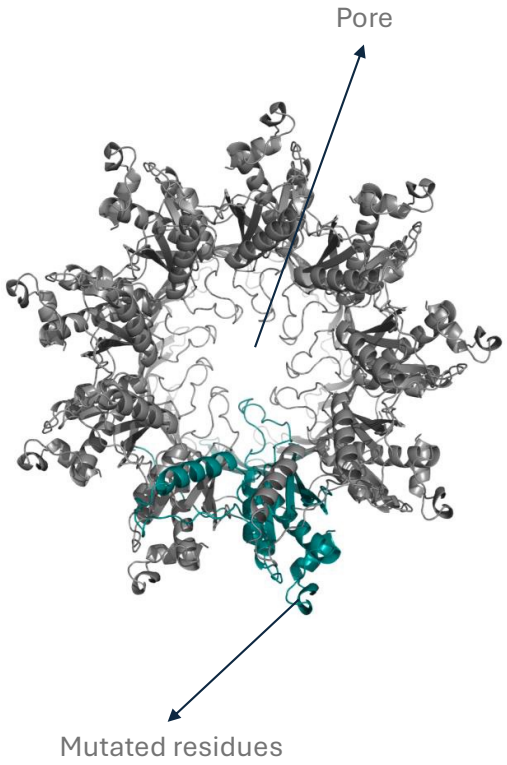


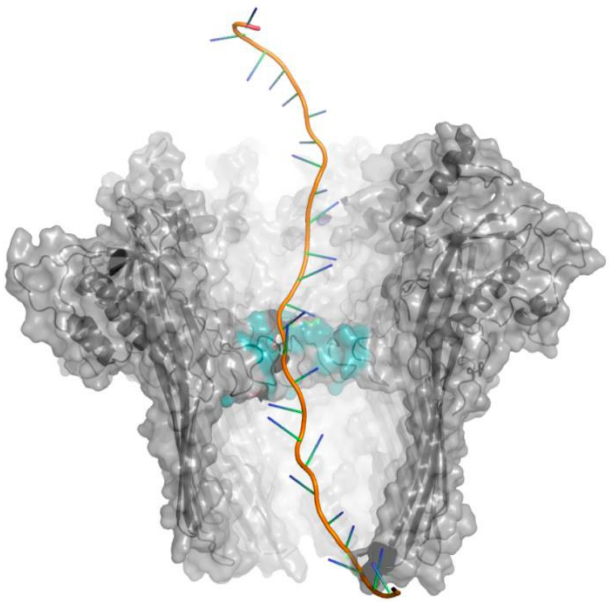
Data analysis to look at copy number aberrations



The nanopore

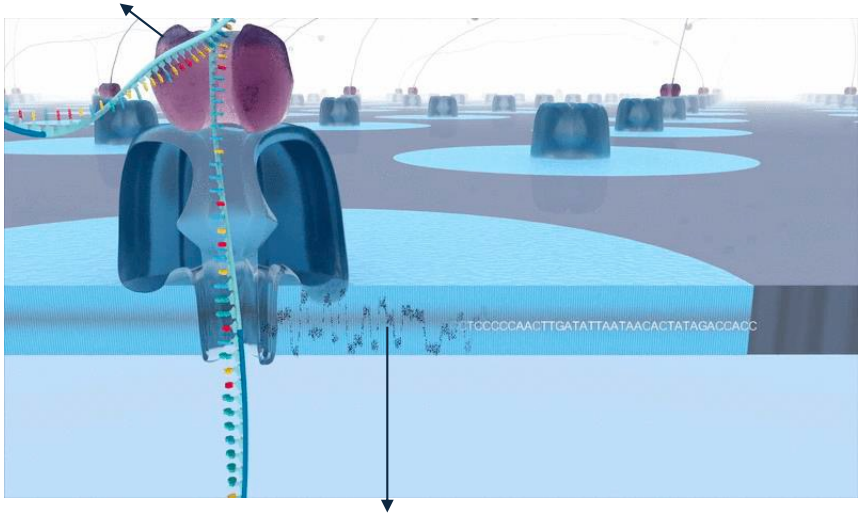






Nanopore DNA/RNA sequencing

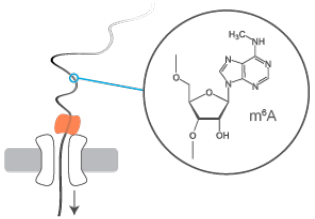
Motor protein



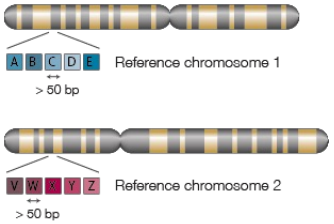
Nanopores, allowing ion flow, are created in a membrane

Bases caused characteristic disruption in the current which is read and analyzed in real time

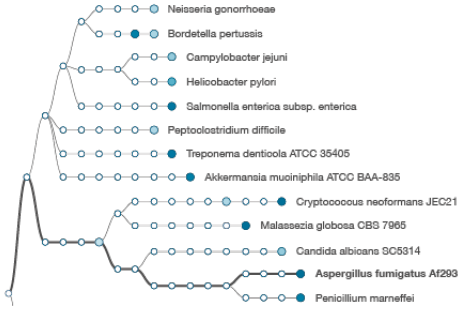
Why nanopore sequencing?



Direct sequencing



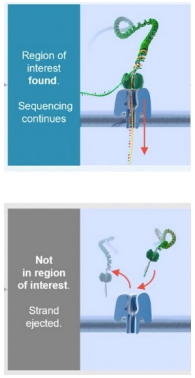
Ultra-long reads



Real-time analysis

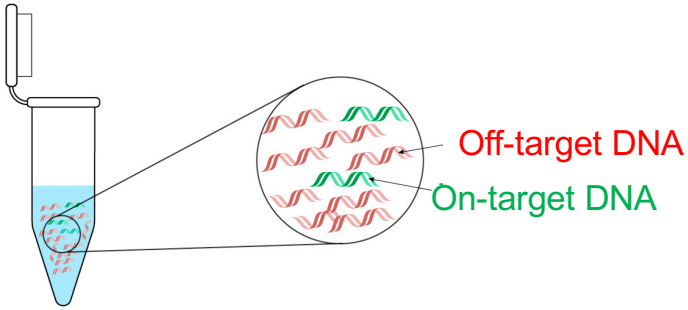


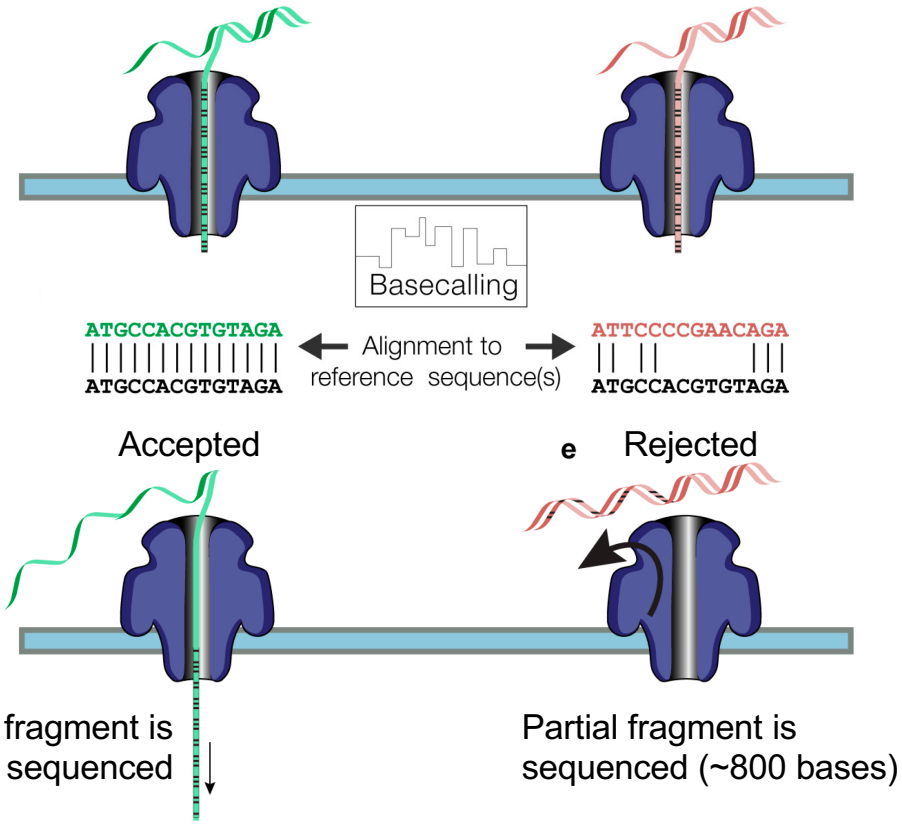
Scalable technology



Enrichment via “adaptive sampling”

ADAPTIVE SAMPLING





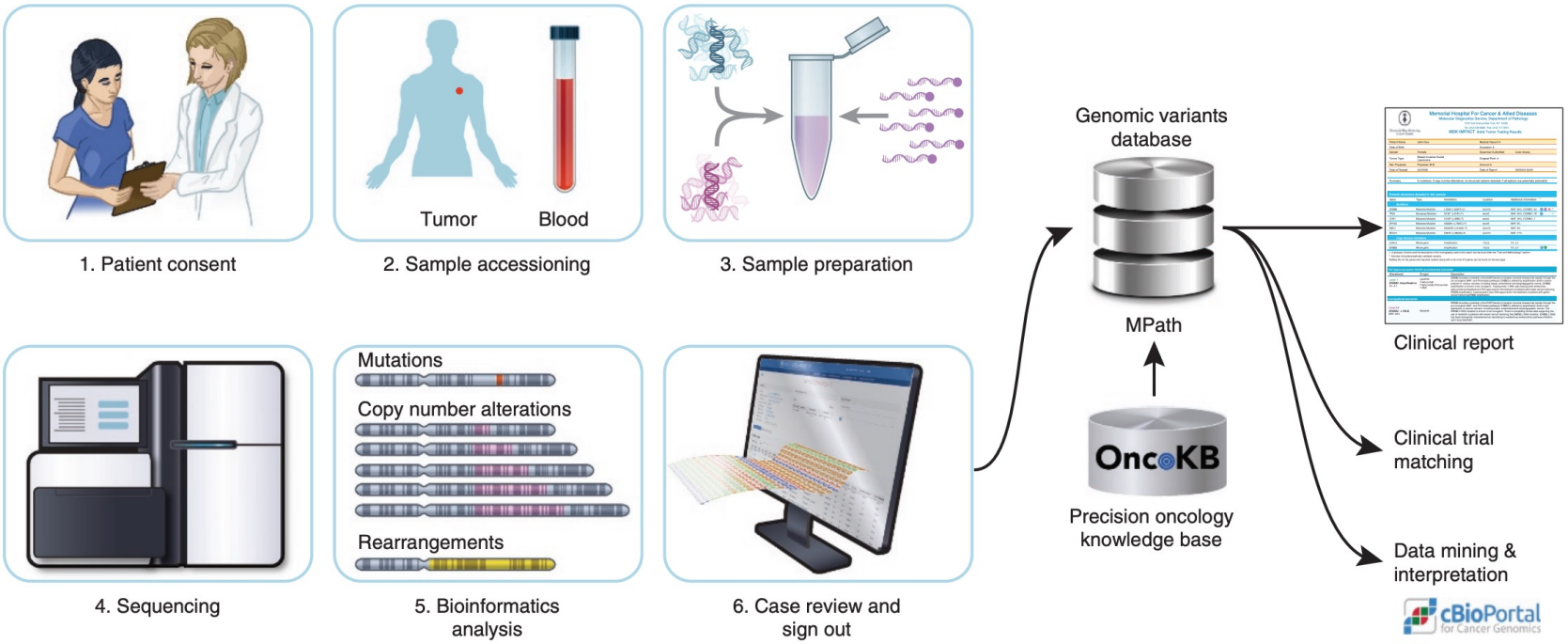
Requirements

✓ Target .bed file
(genomic coordinates)

✓ Reference .fasta file

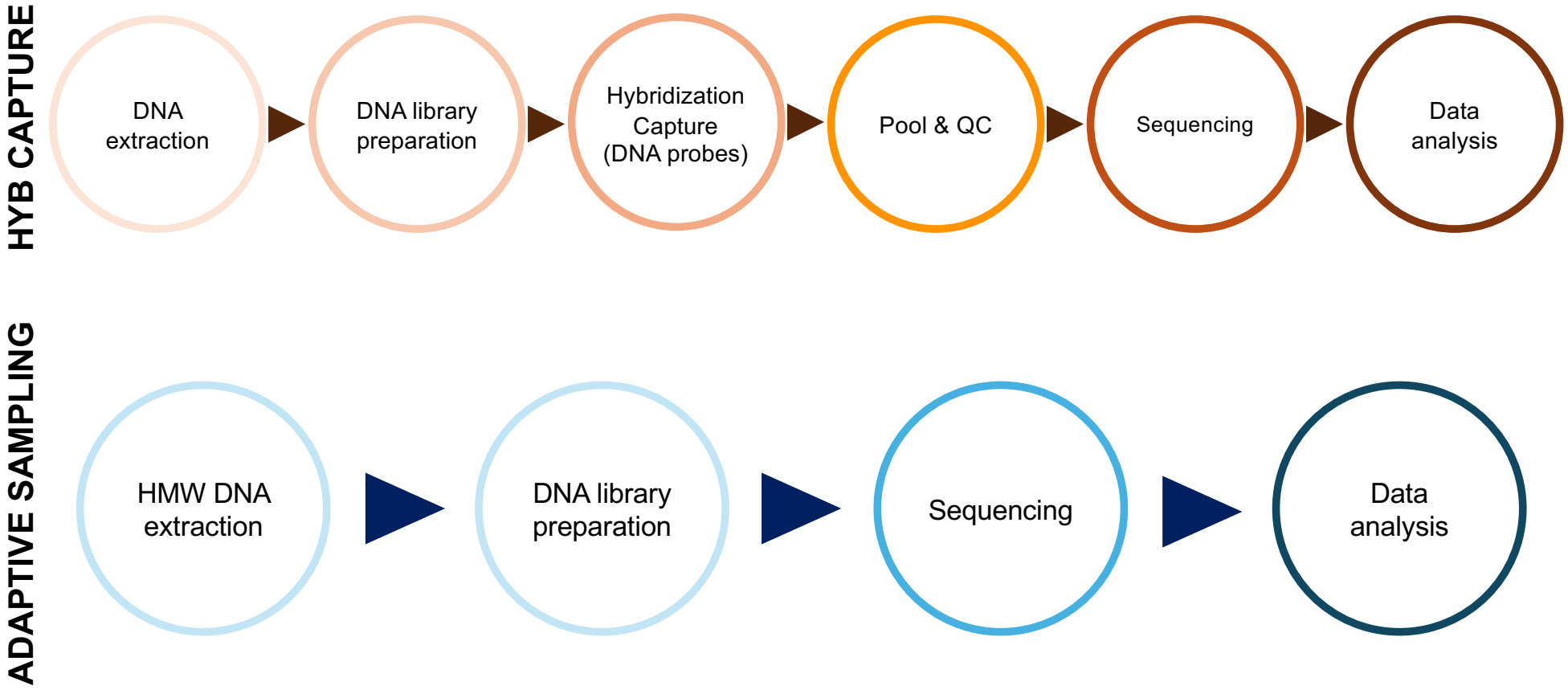
BACKGROUND

Overview of MSK-IMPACT clinical workflow.



Zehir A, et al., Nat Med. 2017

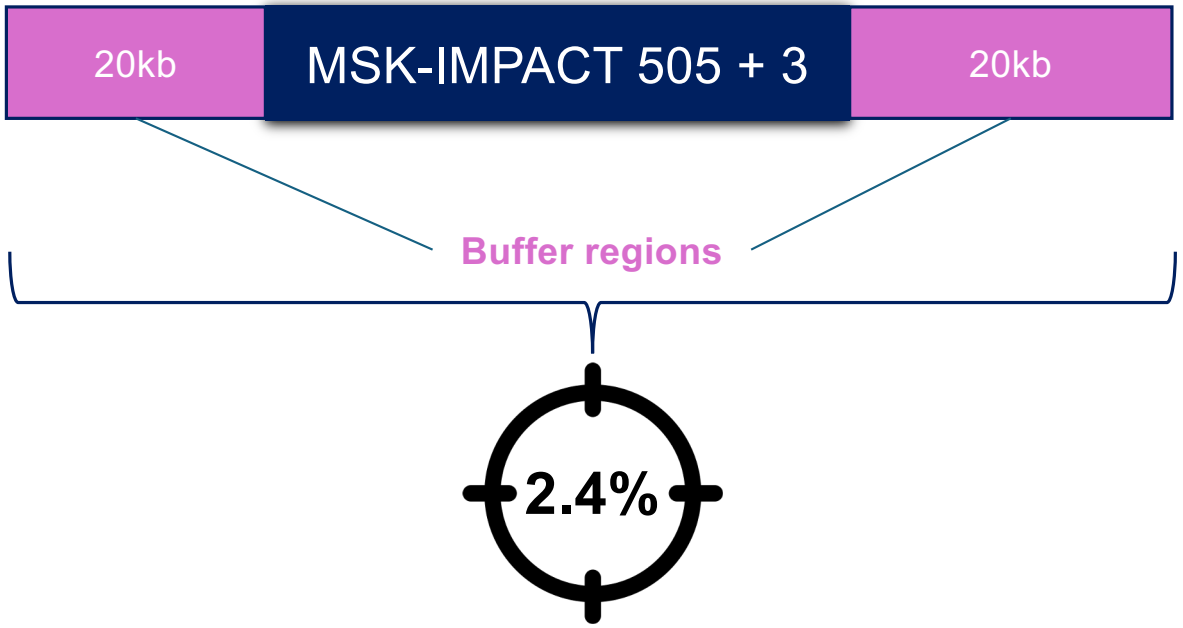
WORKFLOW COMPARISON



ADAPTIVE SAMPLING WORKFLOW

Pre-Sequencing

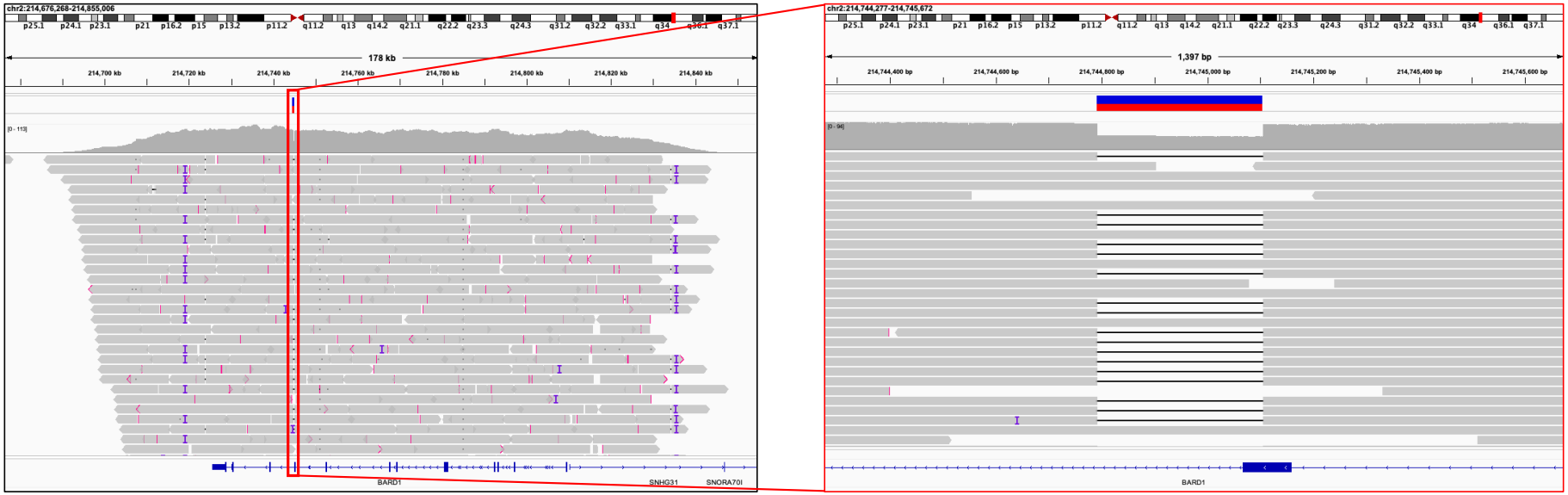
- Target .bed file



- Reference .fasta file

  GRCh38

Patient (R-1925) with a germline alteration in BARD1 gene



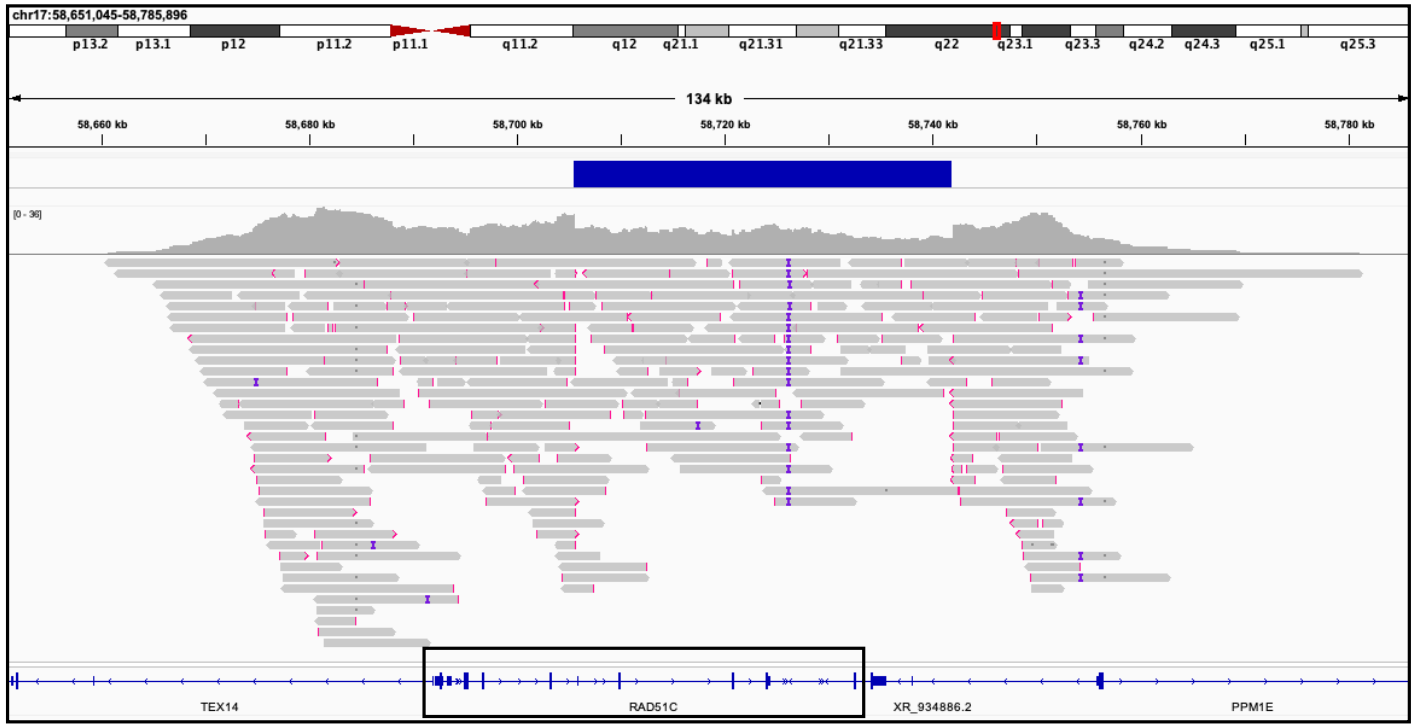
313 bp deletion

Pathogenic **BARD1 SV**
Chr2: Exon 9 deletion (SVLEN=313)

Patient (R-1933) with multiple germline alterations

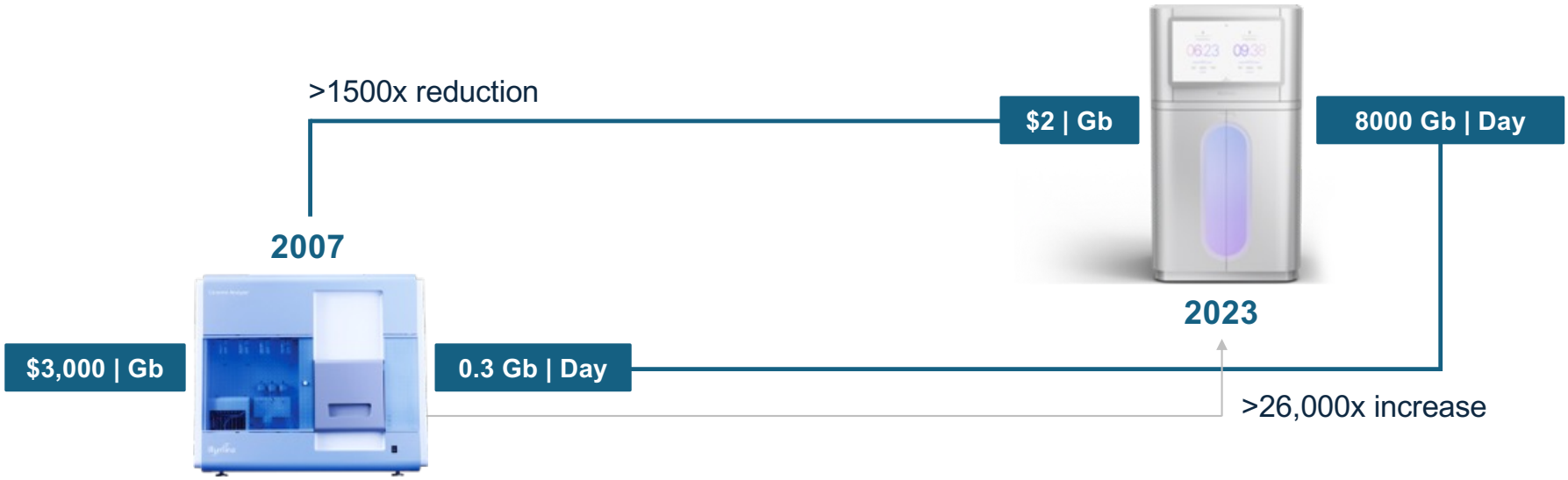
III. Pathogenic **RAD51C SV**
Chr17: 5-exon deletion (SVLEN=36372)

KolmogorovLab/Severus



RAD51C

Sequencing by synthesis - cost



New York City
$2,000 | 6 hours

Paris, France
$1.33 | 1 sec