Giới thiệu phương pháp phân tích dữ liệu phiên mã gene theo cấu trúc không gian và theo từng tế bào đơn

Phuc-Loi Luu, PhD
Bioinformatics Team Lead, Pacific Informatics
Loi.lp@pacificinformatics.com.vn
HCMC, 12/08/2023

Contents

Self-introduction and Acknowledgement

Acknowledgement

All patients who donate samples
All member of Prof. Susan Clark lab

- Prof. Susan Clark
- Dr Clare
- Dr Ruth
- Jenny
- Wenija
- Dilys
- Dr Qian
- Dr Amanda
- Dr Jo
- Dr Braydon
- Dr Shalima

Acknowledgement
Collaborators and Students in Viet Nam

- NGUYEN CANH HIEP, MD, PhD
- LUONG THI MY HANH, MD, PhD
- TRẦN THỊ THANH KHƯƠNG, PhD
- CAO THỊ TÀI NGUYÊN, PhD
- TRỊNH VẠN NGỮ, PhD
- LE MINH THONG, PhD
- Bsc. LÊ NHẤT THÔNG
- Bsc. PHẠM MAI TÂM
- Bsc. NGUYỄN TẤN THANH GIANG
- Bsc. NGUYỄN ANH XUÂN
- Bsc. LE VAN HIEU
- Bsc. NGUYỄN MINH HOÀNG
- Bsc. TRẦN BÁ THIÊN
- ONG PHUC THINH, MD
- NGUYEN HUY THINH, MD
- DAO NGOC BAC, MD, Msc
- NGUYEN PHAN XUAN TRUONG, MD
- TRAN NGUYEN TRONG PHU, MD

All member of VnPathoinformatics group

Pacific Informatics
Your comprehensive bioinformatics solution

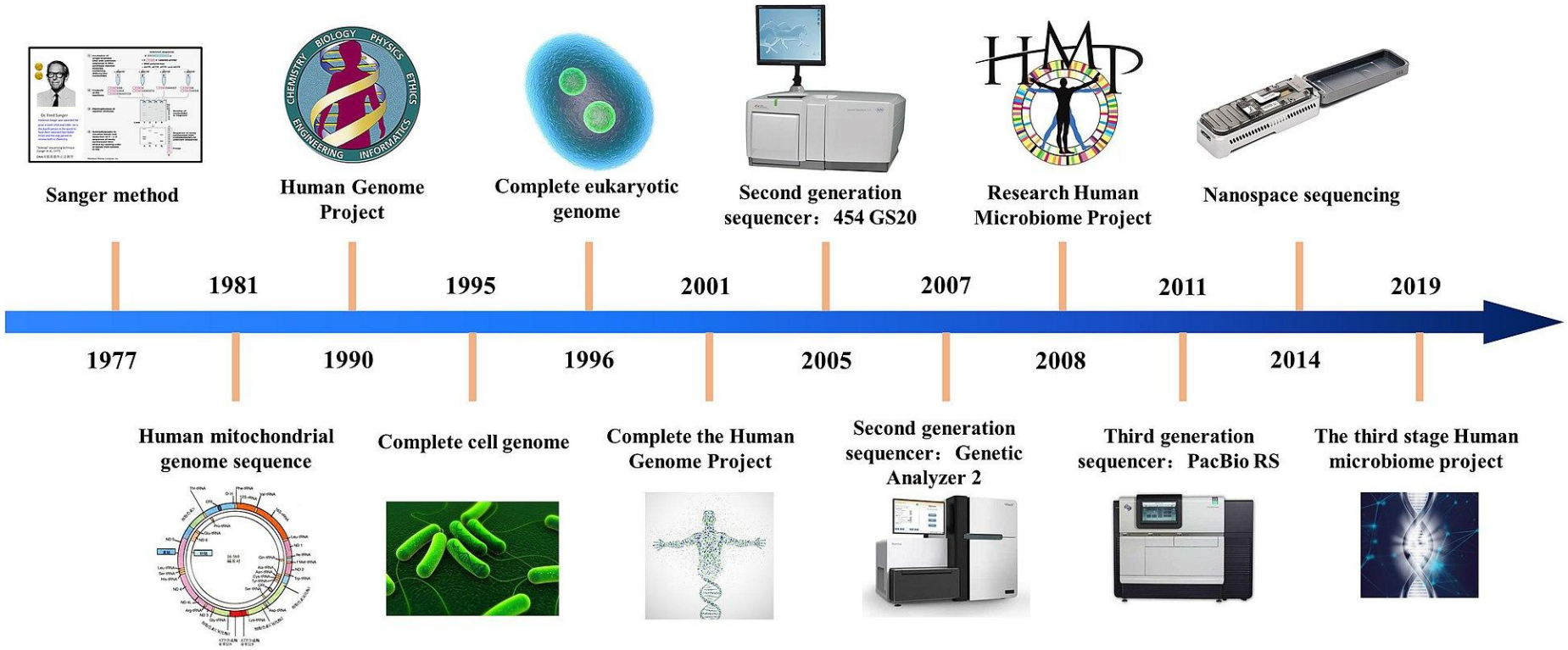
The Bioinformatics Team

ZYMO RESEARCH

Name and Role	Profile Picture	Profile Picture	Name and Role
Luu Phuc-Loi, PhD Team Lead			**Nguyen Viet Tuan**, PhD Consultant
Du Hoang Tien, Msc. Associate			**Phan Vo Thu Nga**, Bsc. Associate
Nguyen Le Duc Minh, Dr. Genetics Consultant			**Tran Thien Tan**, Bsc. Technician
Dao Khuong Duy, Bsc. Technician **Nguyen Minh Hoang**, Bsc. Technician	 		**Le Van Hieu**, Bsc. Technician

Human Genome and Next-Generation Sequencing

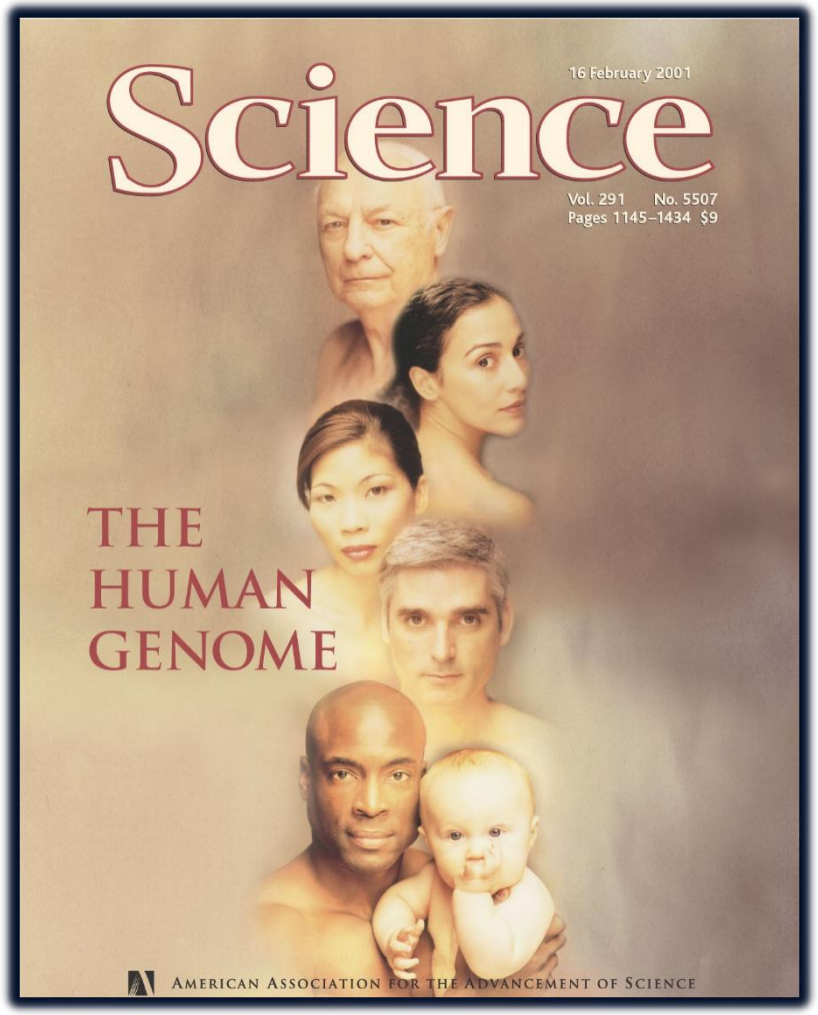
High-throughput sequencing (HTS) methods



https://en.wikipedia.org/wiki/DNA_sequencing

February 2001: Papers Reporting Draft Sequence of Human Genome

HGP Paper



Venter/Celera Paper

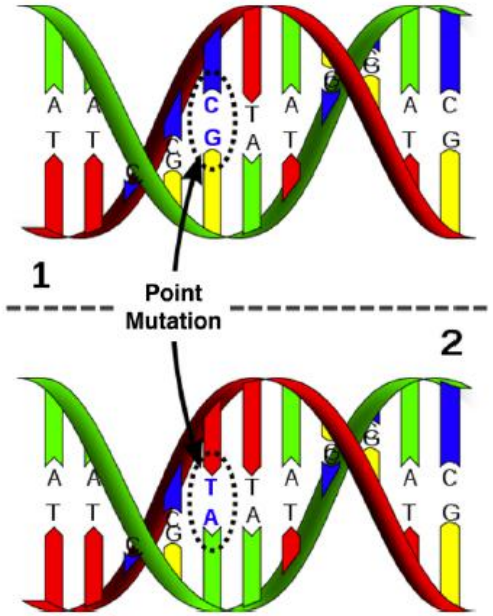
PACIFIC INFORMATICS
Your Trusted Partner in Bioinformatics

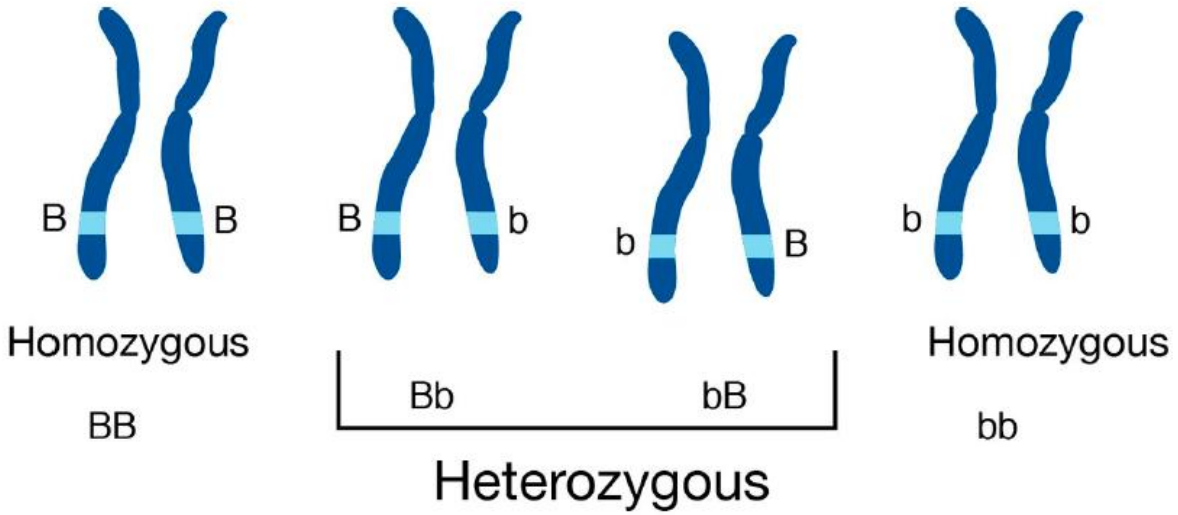

```
48541 agcccttcaa agaaatgttc tcagcaggca tggagcccag gacttgctcc ctttgggtag
48601 agagccgggt tgaaggtgac tgaagtgaaa tgggacagta gaggcggggg gggtggtgag
48661 ttcctggagg tgggggggtgt gggaacctgc tgtgtactga gatgcacccc tgccagttct
48721 gcctgaagat ttgaggcggg gggcaggggg gcggagtgaa gtcatttac tggtaagtaa
48781 ttttaaacct tttaatatta aagcaaacgt ggatatgtaa tgaatgaaat tcattctgga
48841 atgaaaaatt cacgtgatgt tgaaaaataa cacggggctt cagagaggac tttctggctg
48901 gcagcagact ccagattccc agggcccctg caccctcctc tgcccacagg gcaccttaat
48961 ggagaaggtg tgggaggaga gccaggccgg agtcagagca cactggtgac tccacatttg
49021 cagcgtgccc tgcctctctc ctgaggcttg gcaacgtgca atatgctaag caaactcccc
49081 ctgtccccgt ccagtttctg aggacaagag ccaccacctg tagcaaataa agacccagca
49141 acccttgac tcatctttgt gagtctctgg aatcagaggg tagccacatc gctgagaggt
49201 ggagtgaagc actcgggtga aaaggtacaa ggaagtcagg gacaggagtg tggggacatc
49261 acctagacaa tgacagagaa gaggggcaca gccgagtgag gggagagggg ccggcagtcc
49321 tacatcccct ggcctgaagc acgctccagg gcagaaggaa aaacactgtc tttggggtcc
49381 aagagacctg agttcaaatt ctggctccac cactgaccac ctgtgtaacc ttgaactgct
49441 gctgcctgaa cctcaggttc cccttctaaa aatagaggag aaaaggatgc atttctcctt
49501 gcccctgtga gaacgaaatg gtgcaagcac caaggagcct cagcaaaggt cgggcctgcc
49561 cccgcctggc caaacctttc ctcttcagga ggccacggca accgtagttt gacagaagag
49621 cagcaccttg atttaatgct tcccagcatg tgtccttgag caagtcacct aacctctctg
49681 ggctgcttcc tcattgggaa aatatggctg ccagtaaaac ctgccctgtc cacctcctgg
49741 ggcacttggc aaacagcaaa agagtccaaa tgtgcaggct gggccaggcg cagtggctca
49801 tgcctgtaat cccagcaatt taggaagcca aggtgggcgg atcacctgag gtcaggagtt
49861 tgagaccagc ctggccaaca tggtgaaacc ttgtctctac aaaaatacaa aaattagccg
49921 ggcatgatgg cgggtgcctg taatcccagt tactcgggag gctgaggcaa gagaatcgct
49981 tgaacccgga aggggaaggt tgcagtgagc caagattgtg ccactgcact ccagcctggg
50041 caacagagcg agactctgtc tcaaaaaaaa aaaaaaaaaa aaacaatgca gagctggctg
50101 tgtaaaaaac ctgttccact gcagggccca gtgtccacca ggctggggtg caggcctatg
50161 gggtgggggc ccagcatcag cctctcagca gccctgggag gcggggcgca tcccgtgccc
50221 ctcgtggtct ggatgtgttc tagcccaagt cctaggttac acctgccgtc gcctggcctc
50281 tcaggagagg cccagggtga ggaggagcat ggtaaaggtg aagctgattg ggaagtcagc
50341 tgttgggaaa gcaactcctt gcacattgga ggaaccgaga aagactgacc ccgaggacag
50401 cagccagcat ggcccttcct gggagcccat gttgggggat tcctgctgca gccaaggctc
50461 agcccttgtg gtcgcaggtg ctggtcctgg cctcttcccc tcccatgcag gagcacagga
50521 gagatggctt ctgaggacct gttgcagctg tggccctggg aatagatttg ccagggagct
50581 ttaaagcagc tgagtgtgtc atccagctaa gcctggggaa ggagcttggc tcaggtcctg
50641 acaggtgtga cagggatggg gactgggaag taagagatga aaccctggct ggaggctgtg
50701 agcttccaca gccagcgctg gacagggagg gtccagatat acccactagt gccctcacca
```

https://www.ncbi.nlm.nih.gov/nuccore/806904736

Human Genome Variation

- Humans genomes are >99% similar by sequence
- A typical human genome has ~5 million variants with 3-4 million single-nucleotide variants
- Humans are diploid

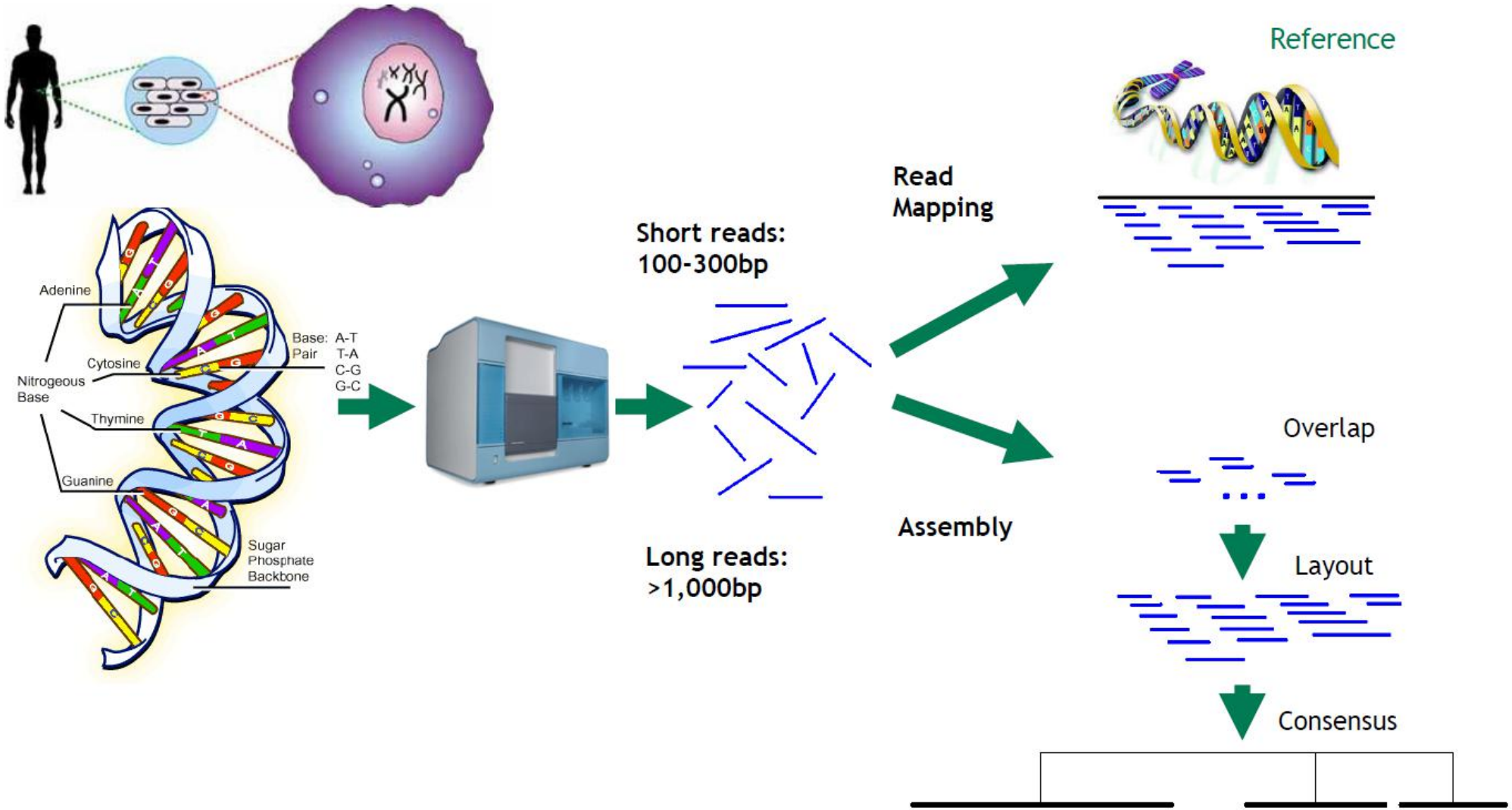




Source: https://rosalind.info/media/point_mutation.png, https://en.wikipedia.org/wiki/Zygosity

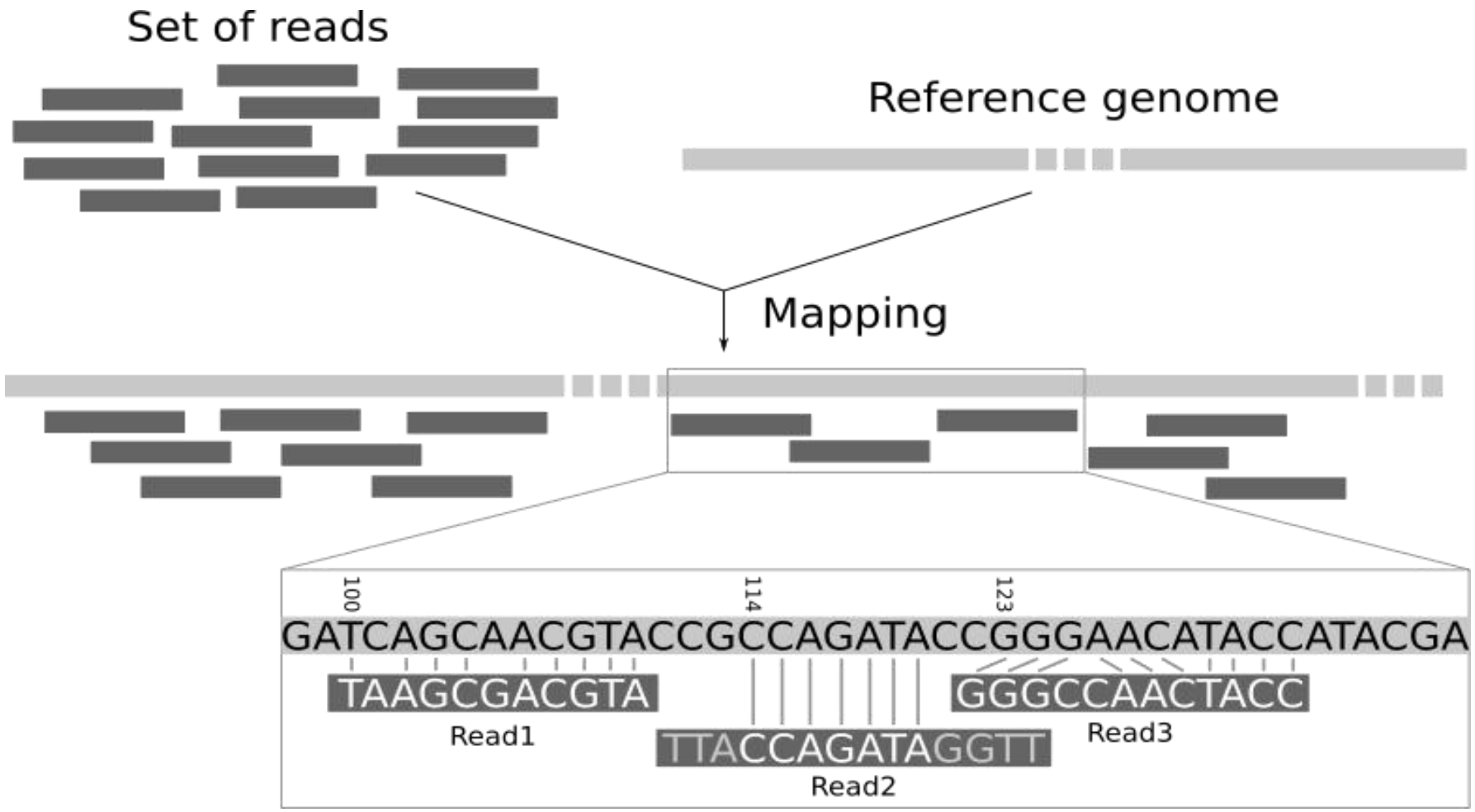
Tobias Rausch

Sequencing DNA



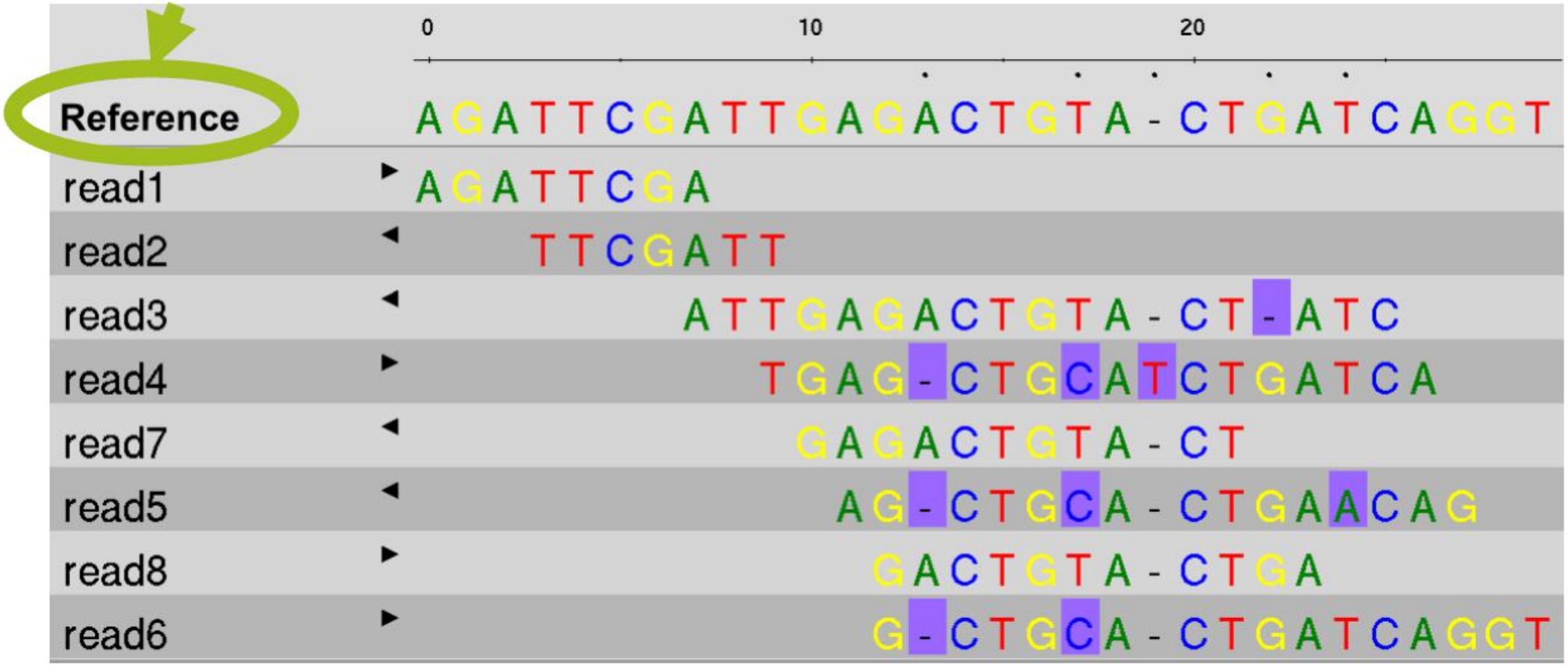
Tobias Rausch

Read Alignment



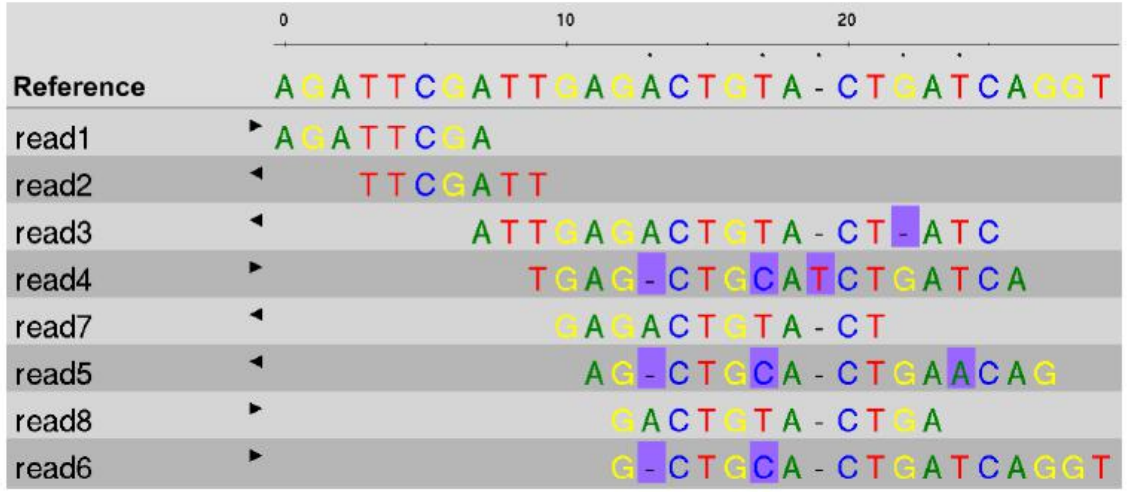
Read Alignment

Reference sequence



Variant Calling - Data Transformation

Alignment



SAM/BAM file

List of Variants

CHR	POS	ID	REF	ALT	GT
chr1	12	.	GA	G	0/1
chr1	17	rs123	T	C	0/1

[No Title]

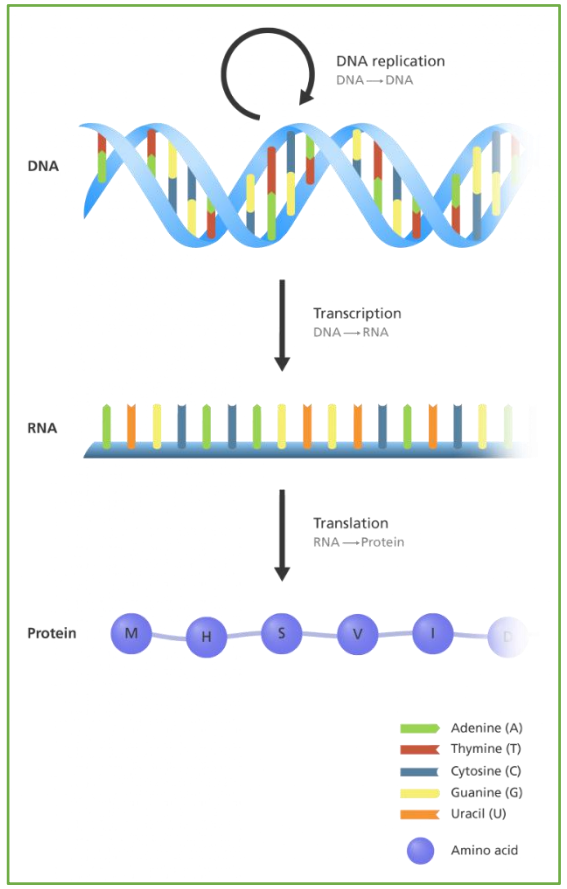
Genotype (GT):

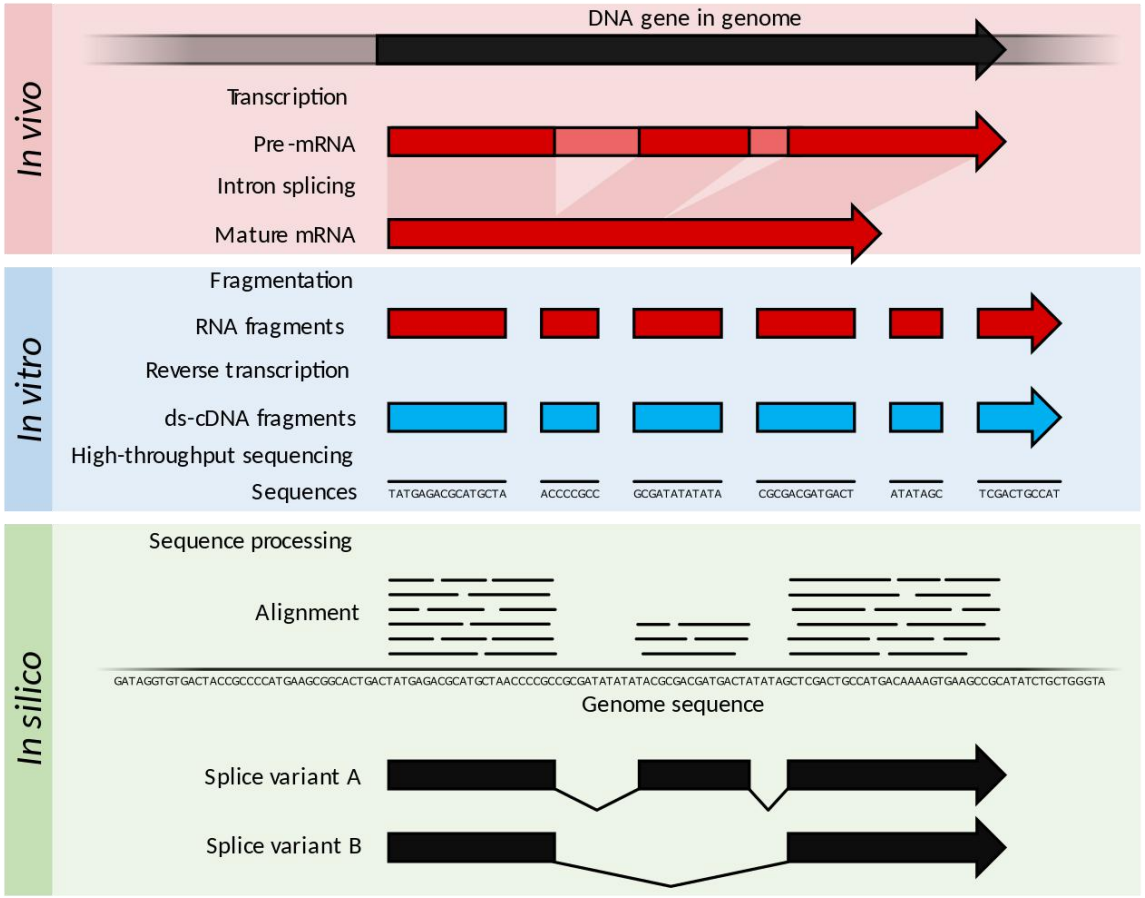
0/0: hom. reference
0/1: heterozygous
1/1: hom. alternative

VCF/BCF file

Bulk RNA Sequencing (RNA-seq)

What is RNA sequencing (RNA-seq)?





https://www.yourgenome.org/facts/what-is-the-central-dogma/

https://en.wikipedia.org/wiki/RNA-Seq

What is RNA sequencing (RNA-seq)?

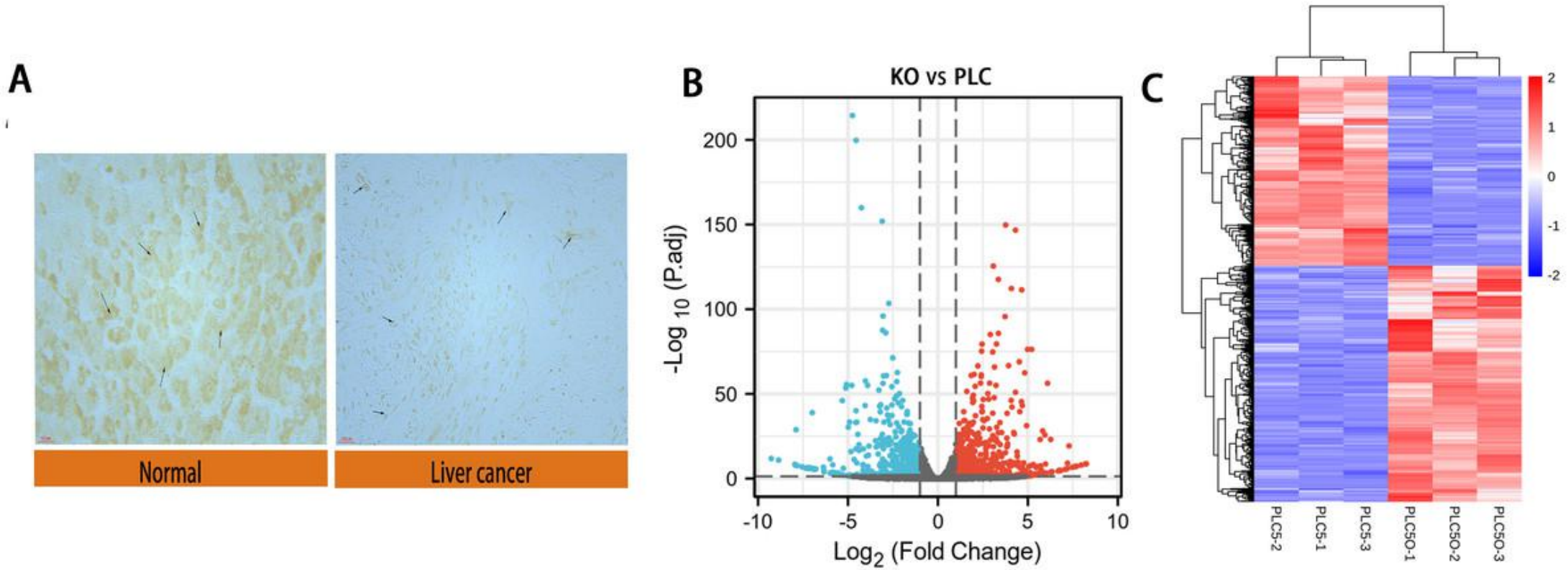


Figure: Differential expression of DAPK1 and volcano plot and heat maps of differential genes.

(A) Relative expression of DAPK1 compared to normal liver tissue samples. (B, C) Volcano plot and heat map showing the 732 differentially expressed genes. The color red indicates upregulated genes, and blue indicates downregulated genes. PLC, PLC/PRF/5 cells; KO, DAPK1–knockout PLC/PRF/5 cells. padj <0.5, logFC 1.

https://www.yourgenome.org/facts/what-is-the-central-dogma/

What is RNA sequencing (RNA-seq)?

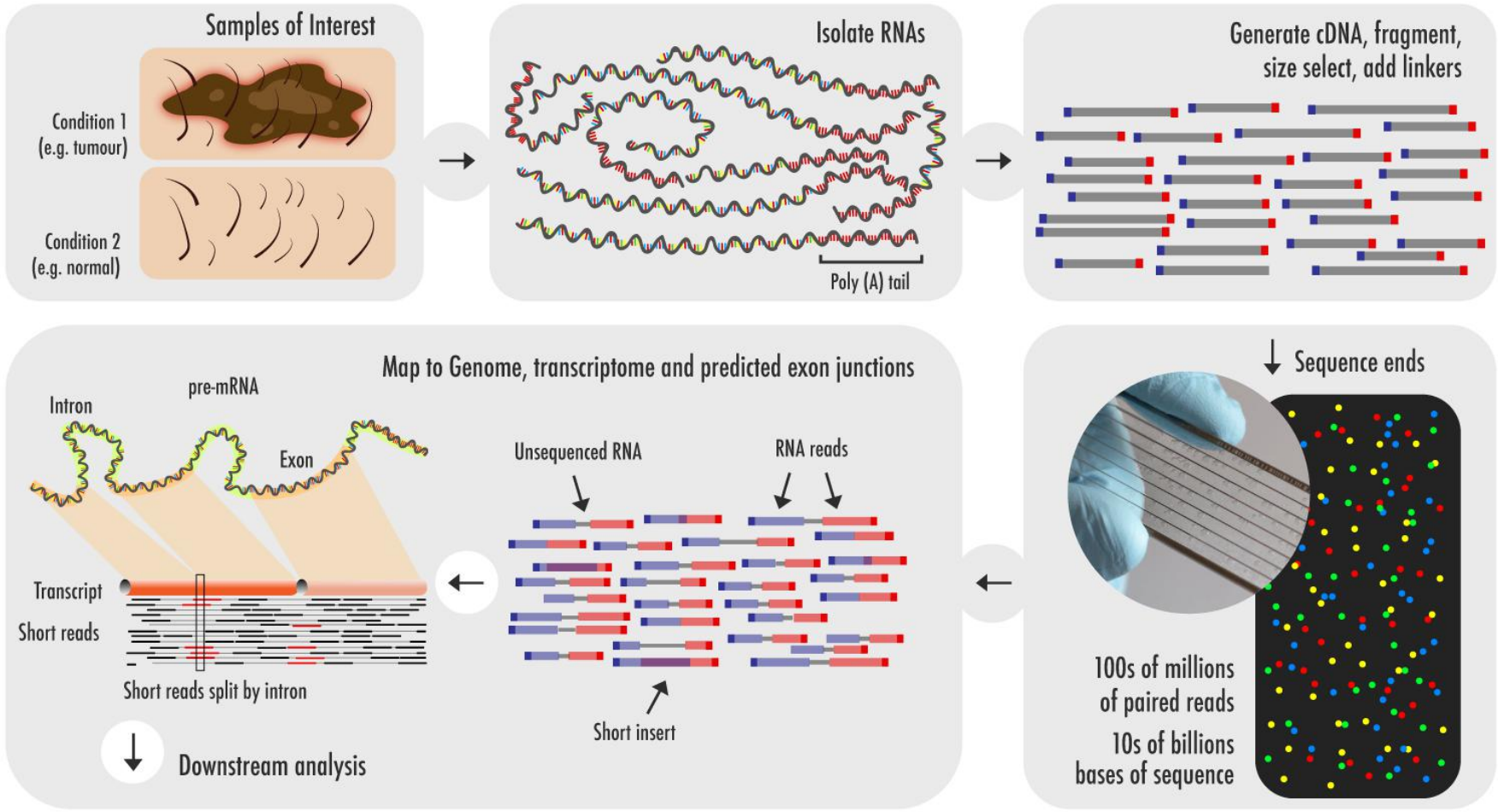
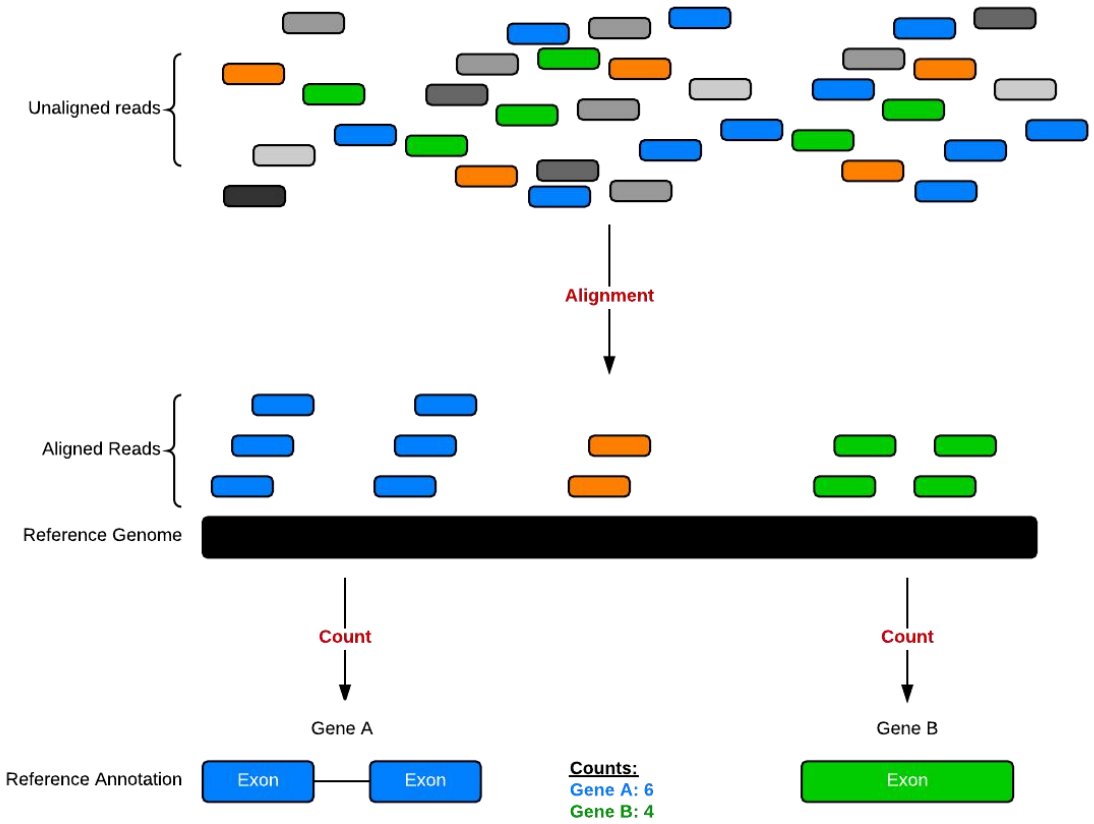


Figure. *A workflow for RNA-seq. Credit: Technology Networks.*

How can RNA sequencing quantify gene expression?



countData

	ctrl_1	ctrl_2	exp_1	exp_1
geneA	10	11	56	45
geneB	0	0	128	54
geneC	42	41	59	41
geneD	103	122	1	23
geneE	10	23	14	56
geneF	0	1	2	0
…	…	…	…	…
…	…	…	…	…
…	…	…	…	…

colData

	treatment	sex
ctrl_1	control	male
ctrl_2	control	female
exp_1	treatment	male
exp_2	treatment	female

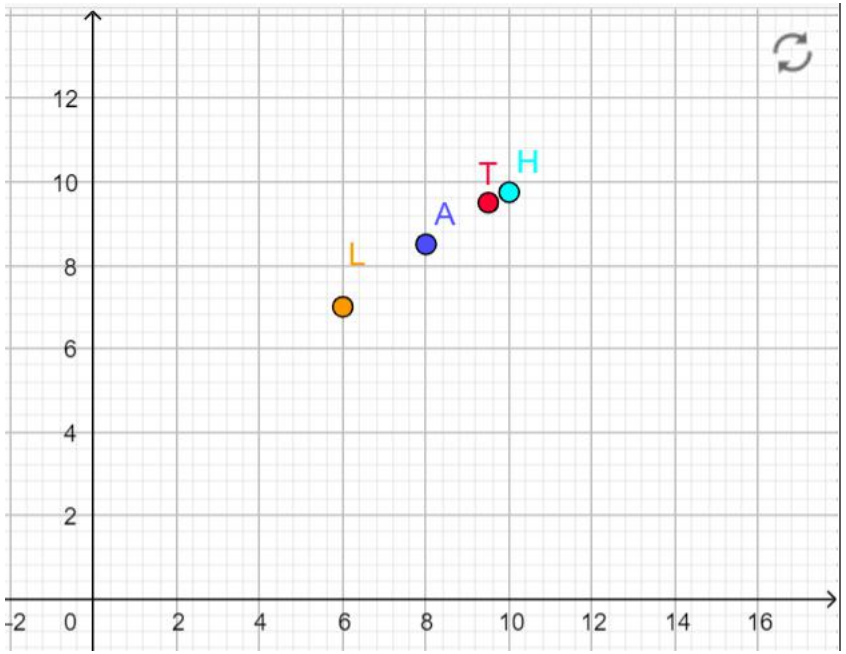
Sample names:
ctrl_1, ctrl_2, exp_1, exp_2

http://bioconnector.github.io/bims8382/r-rnaseq-airway.html

Data of life

	Name	Math	English
1	Kim Anh	8	8.5
2	Hiep	10	9.75
3	Hanh	9.75	10 (French)
4	Loi	6	7
5	Truong	9.5	9.5

English

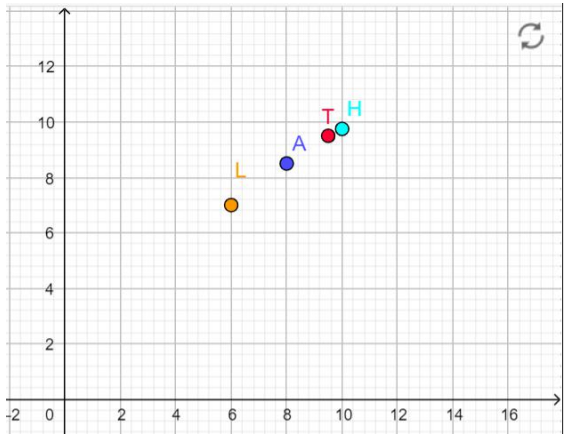


Math

Data of life

	Name	Gender	Age	Math	English	History	isDoc
1	Kim Anh	f	16	8	8.5	10	No
2	Hiep	m	39	10	9.75	6	yes
3	Hanh	f	18	9.75	10 (French)	9.5	yes
4	Loi	m	41	6	7	8	No
5	Truong	m	34	9.5	9.5	9	yes

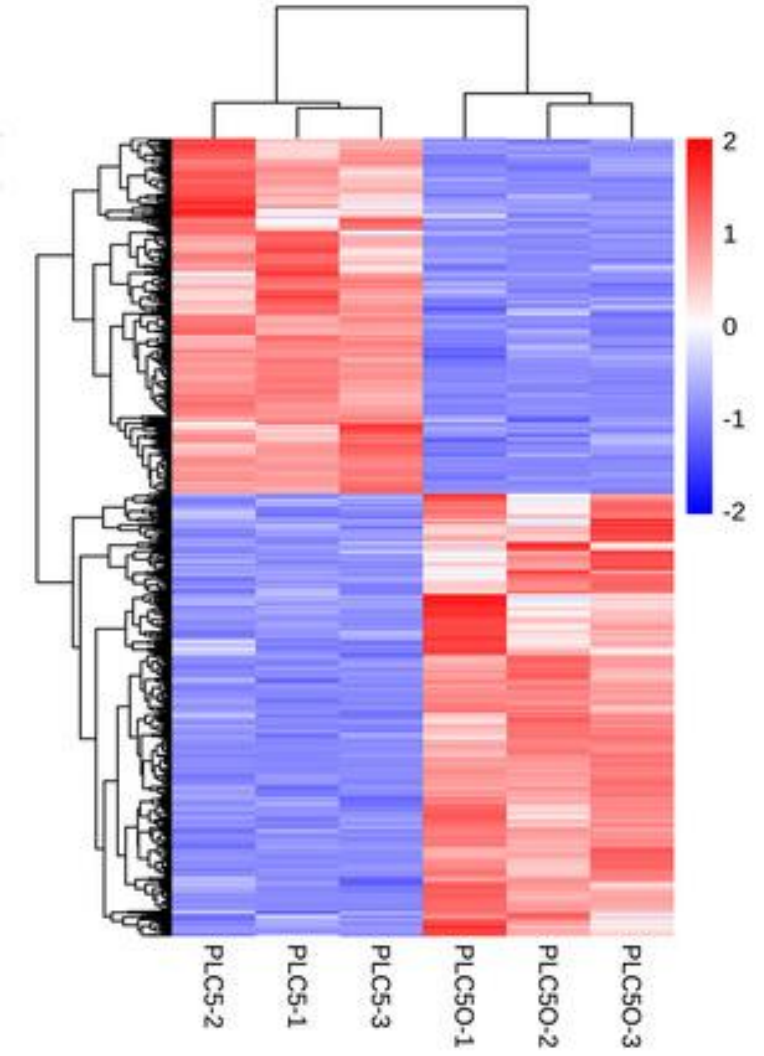
PC2 (14%)



PC1 (62%)

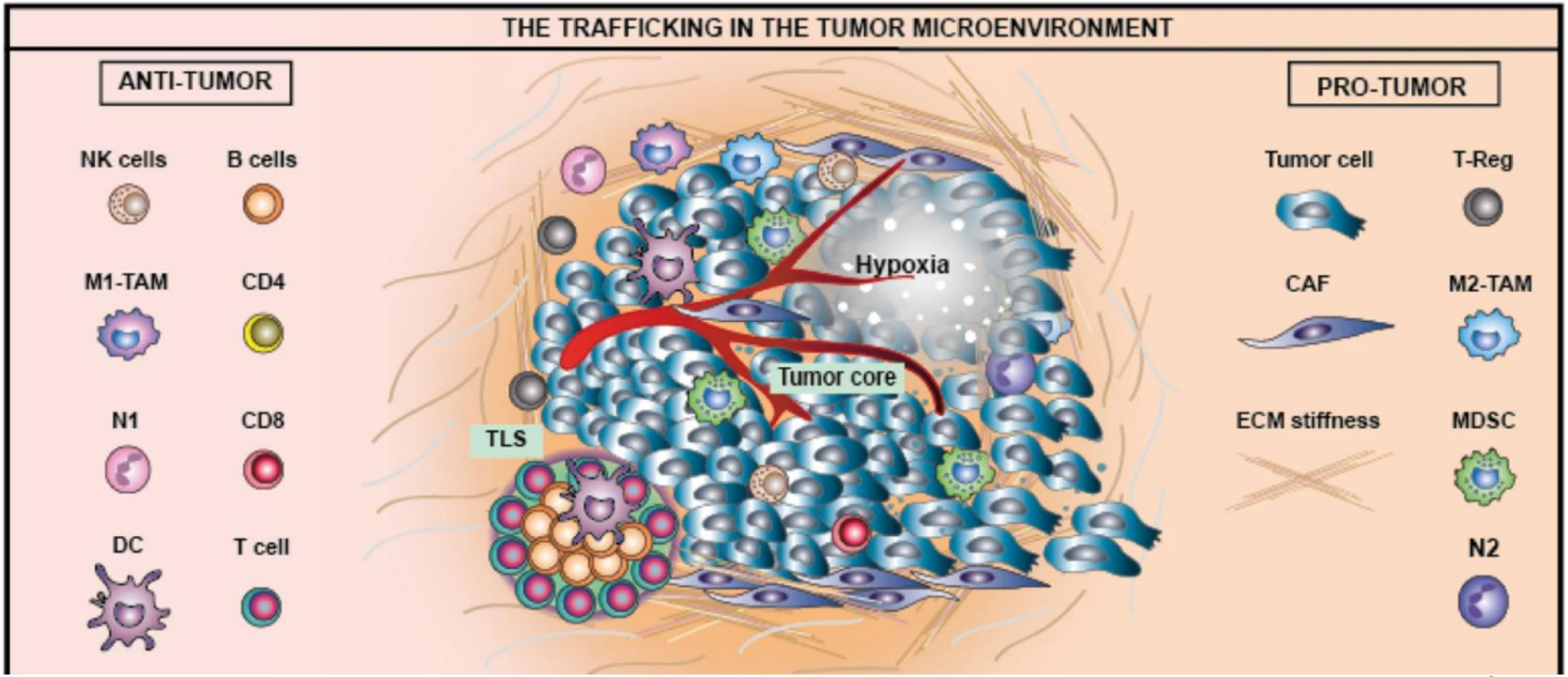
Heatmap

	A	B	C	D
1		2014	2015	2016
2	January	600	708	594
3	February	607	984	749
4	March	901	886	908
5	April	608	615	835
6	May	715	833	734
7	June	520	663	618
8	July	731	521	950
9	August	709	663	987
10	September	633	863	979
11	October	533	651	841
12	November	996	958	749
13	December	792	717	875



Single cell RNA Sequencing (scRNA-seq)

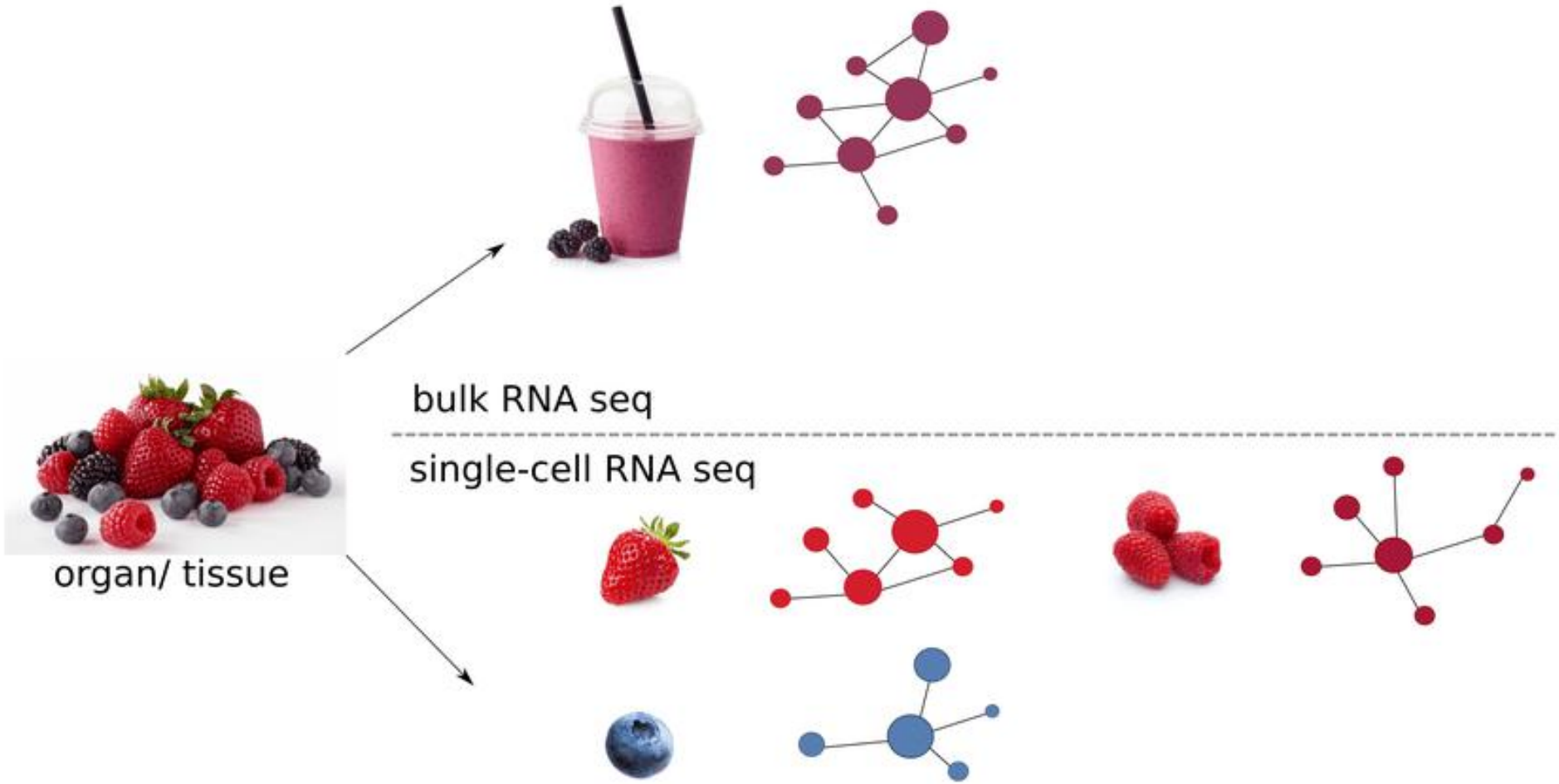
Single-cell applications: Tumor microenvironment



Source: https://doi.org/10.1186/s13046-020-01586-y

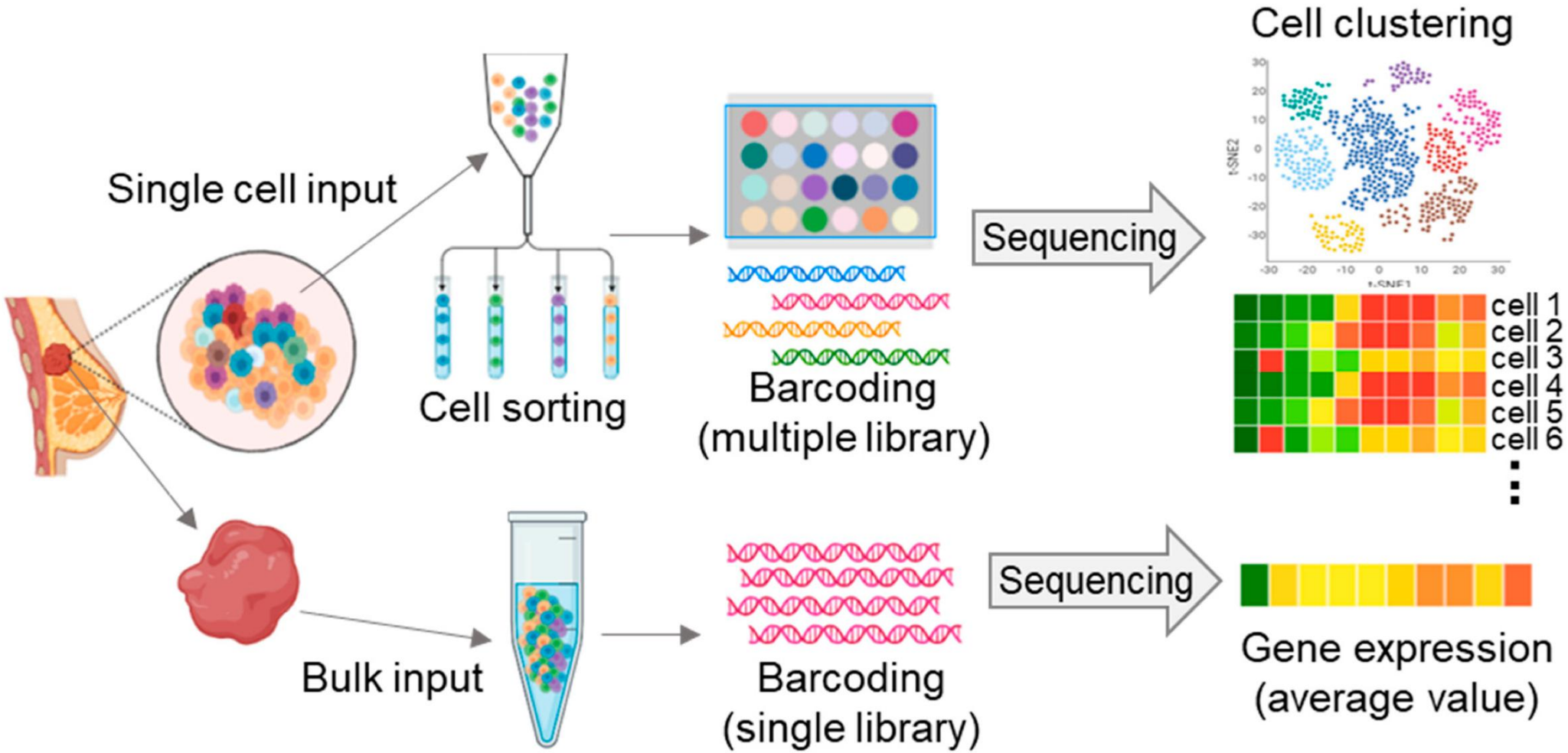
Tobias Rausch

Bulk RNA-seq vs Single-cell RNA-seq



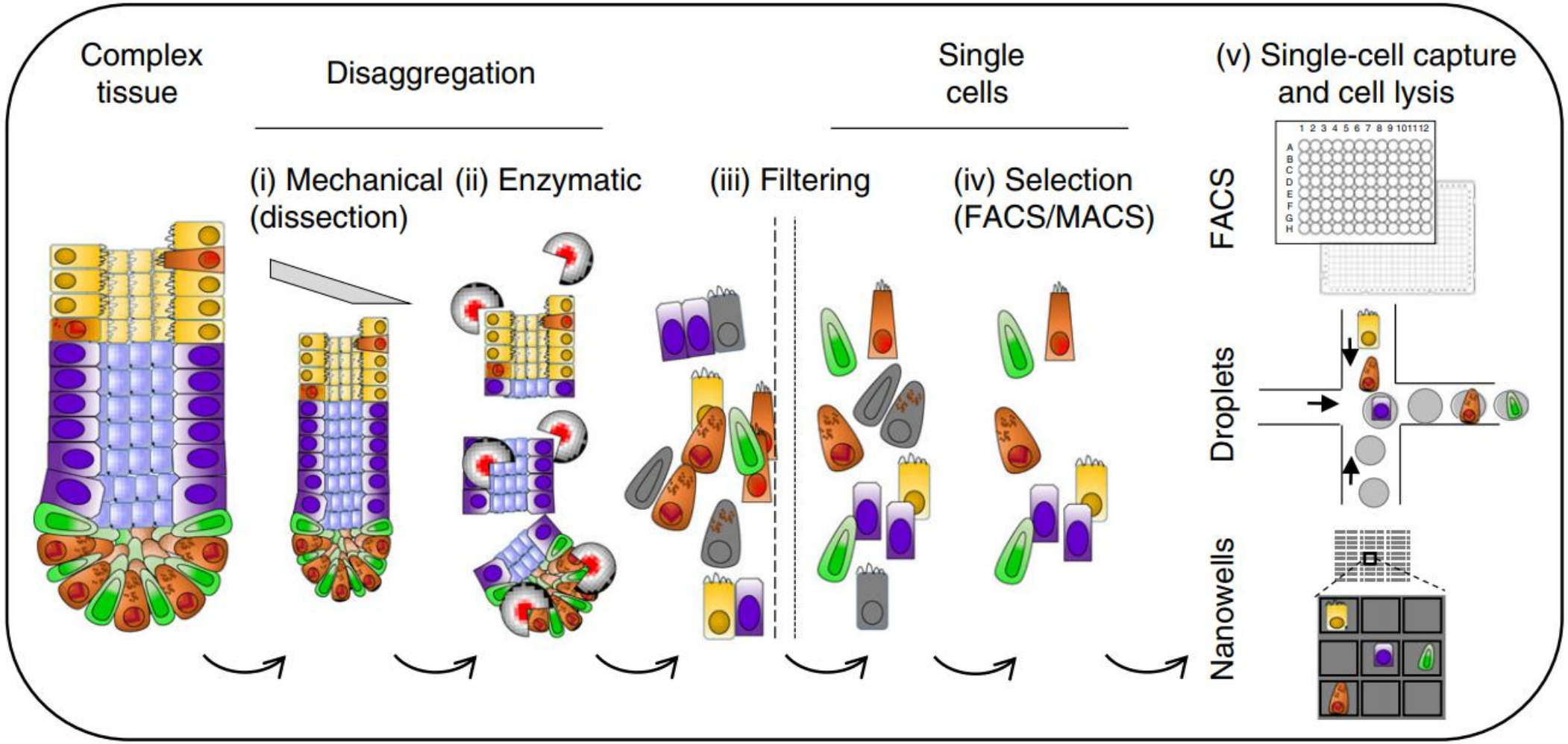
DOI:10.1101/2021.04.02.438193

Bulk RNA-seq vs Single-cell RNA-seq



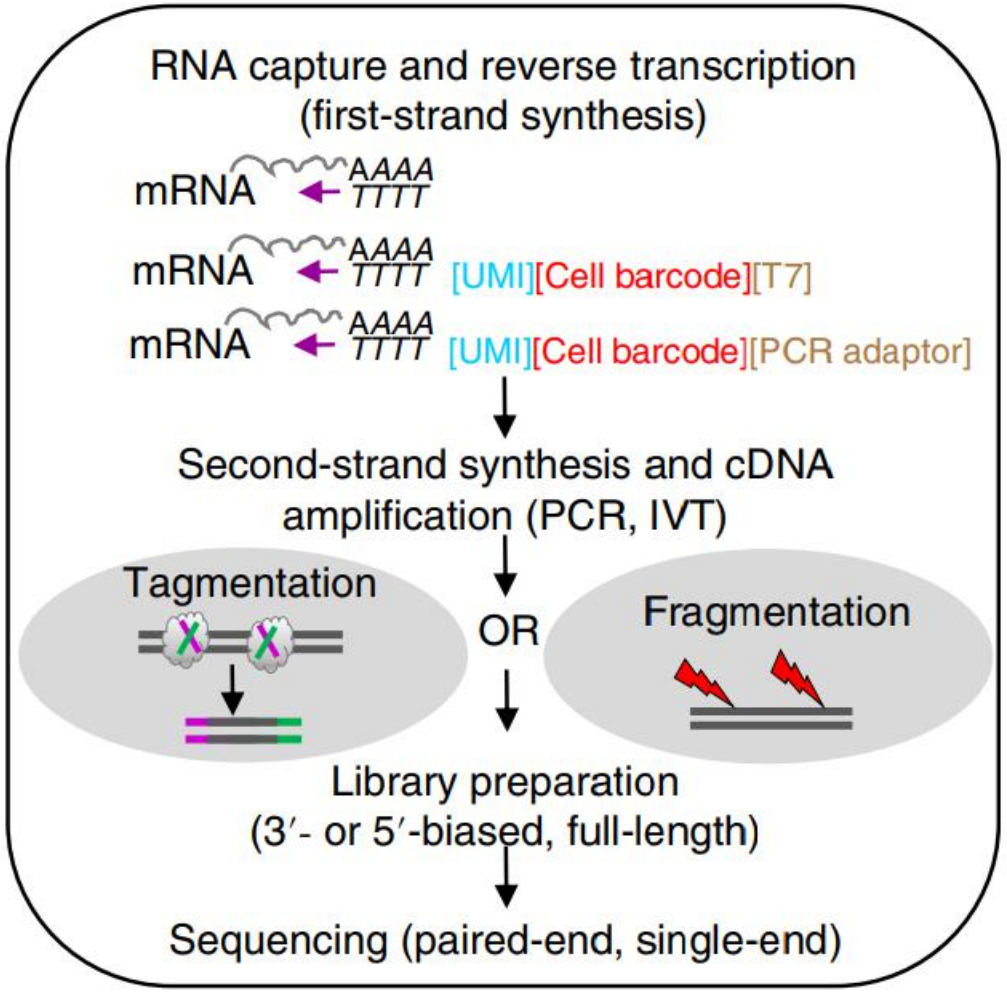
DOI:10.3390/ijms23010216

Single-cell RNA-seq: sample prepraration



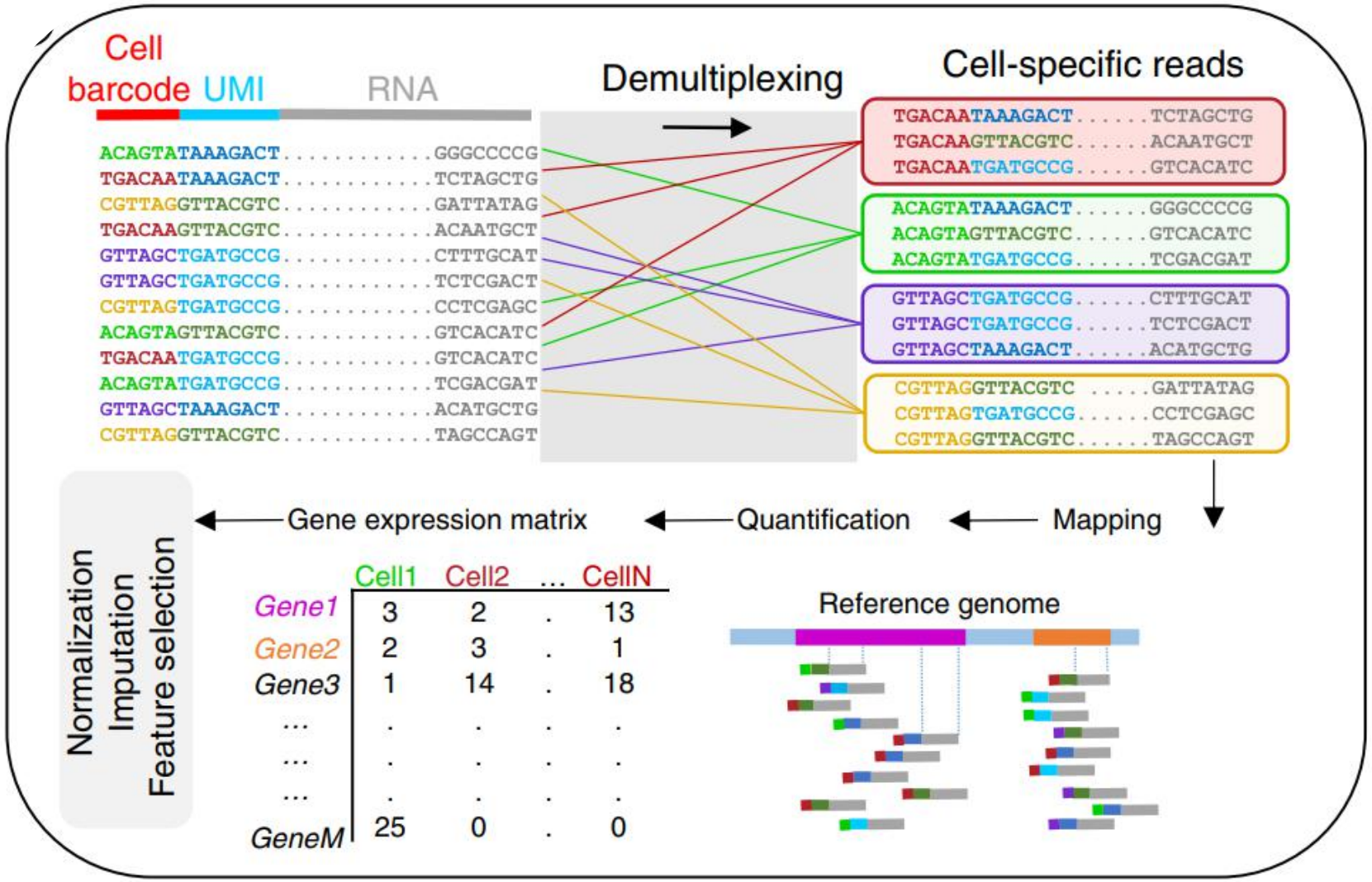
DOI:10.1038/s41596-018-0073-y

Single-cell RNA-seq: sequencing



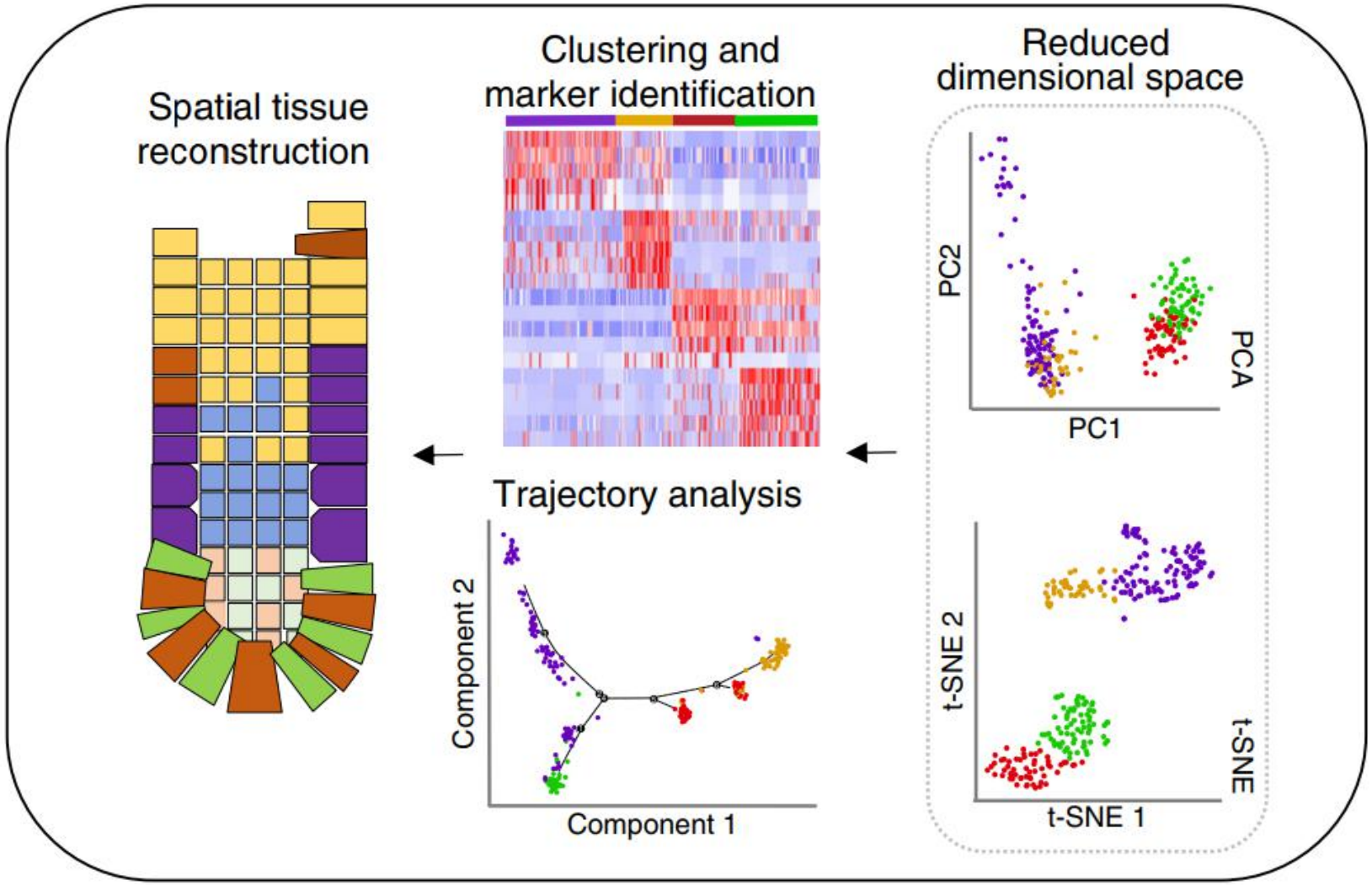
DOI:10.1038/s41596-018-0073-y

Single-cell RNA-seq: data processing



10.1038/s41596-018-0073-y

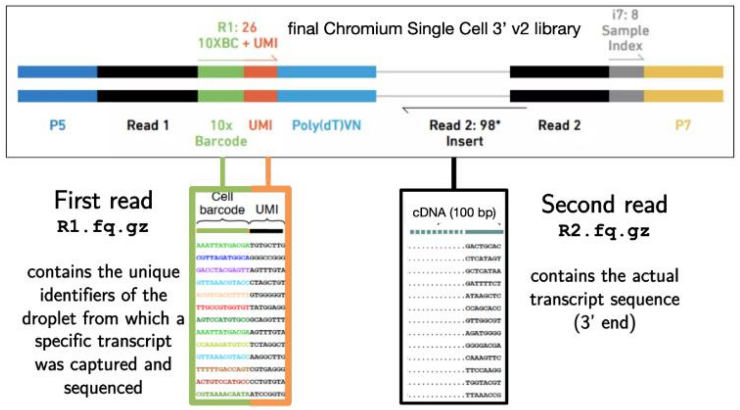
Single-cell RNA-seq: data processing



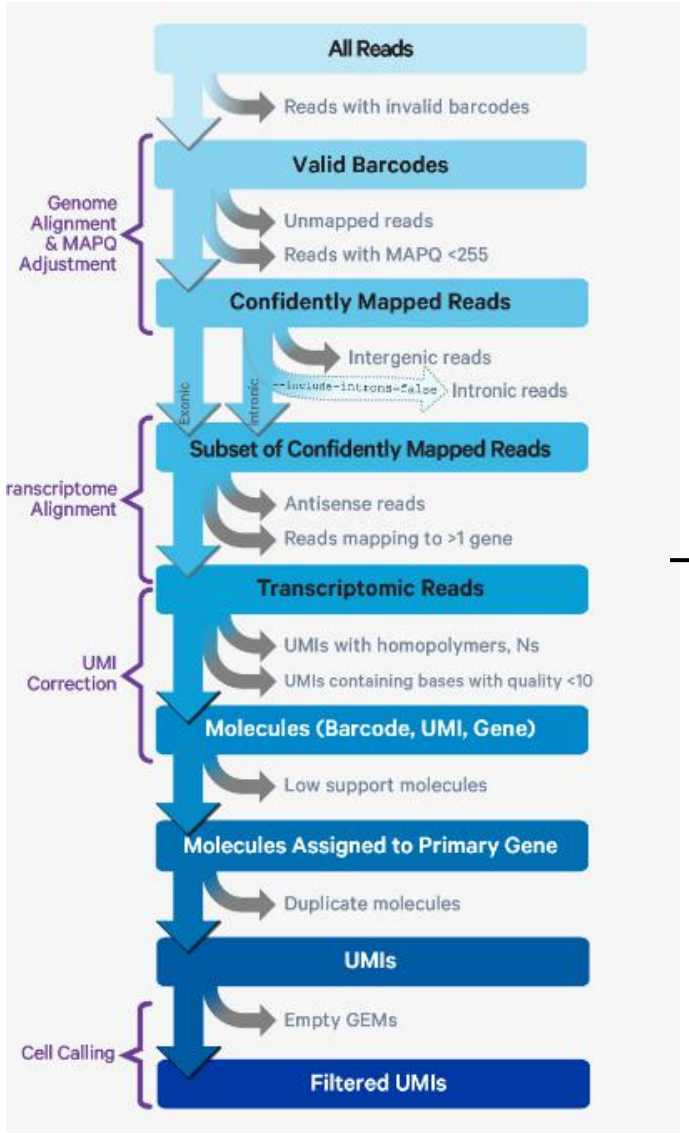
10.1038/s41596-018-0073-y

Upstream Analysis Process

Input: Fastq



Sequencing multi-QC



Cellranger Pipeline

Individual sample CellRanger

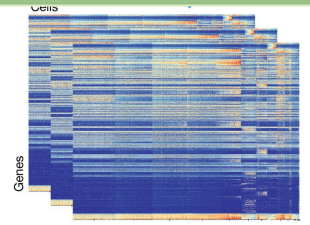
CellRanger Multi-QC report

https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/algorithms/overview#cell_calling

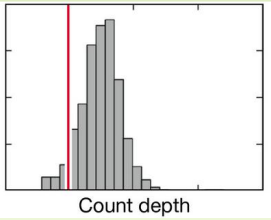
DATA PRE-PROCESSING

Raw data process

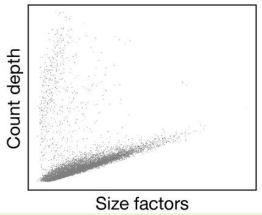
Alignment to transcriptome



Quantification into Raw Counts Matrix

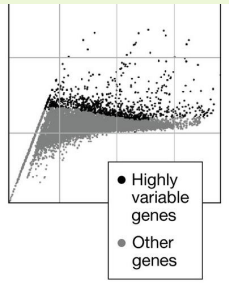


Quality control



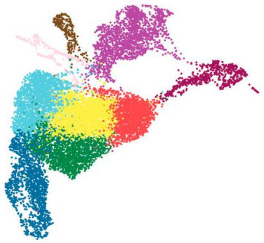
Normalization

Feature selection

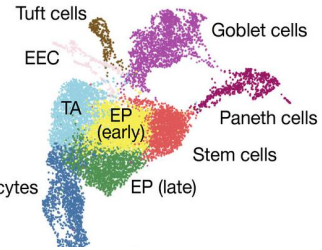


Visualization

Clustering



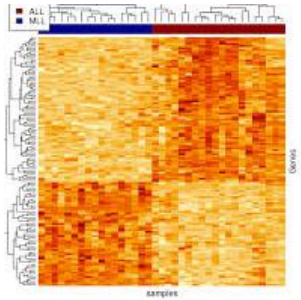
Cluster annotation



SUBSETING

Differentially gene expressed features (cluster biomarkers)

Pseudo-bulk differential gene expression analysis



Gene Ontology analysis

DOWNSTREAM ANALYSIS

Bulk ATAC-seq

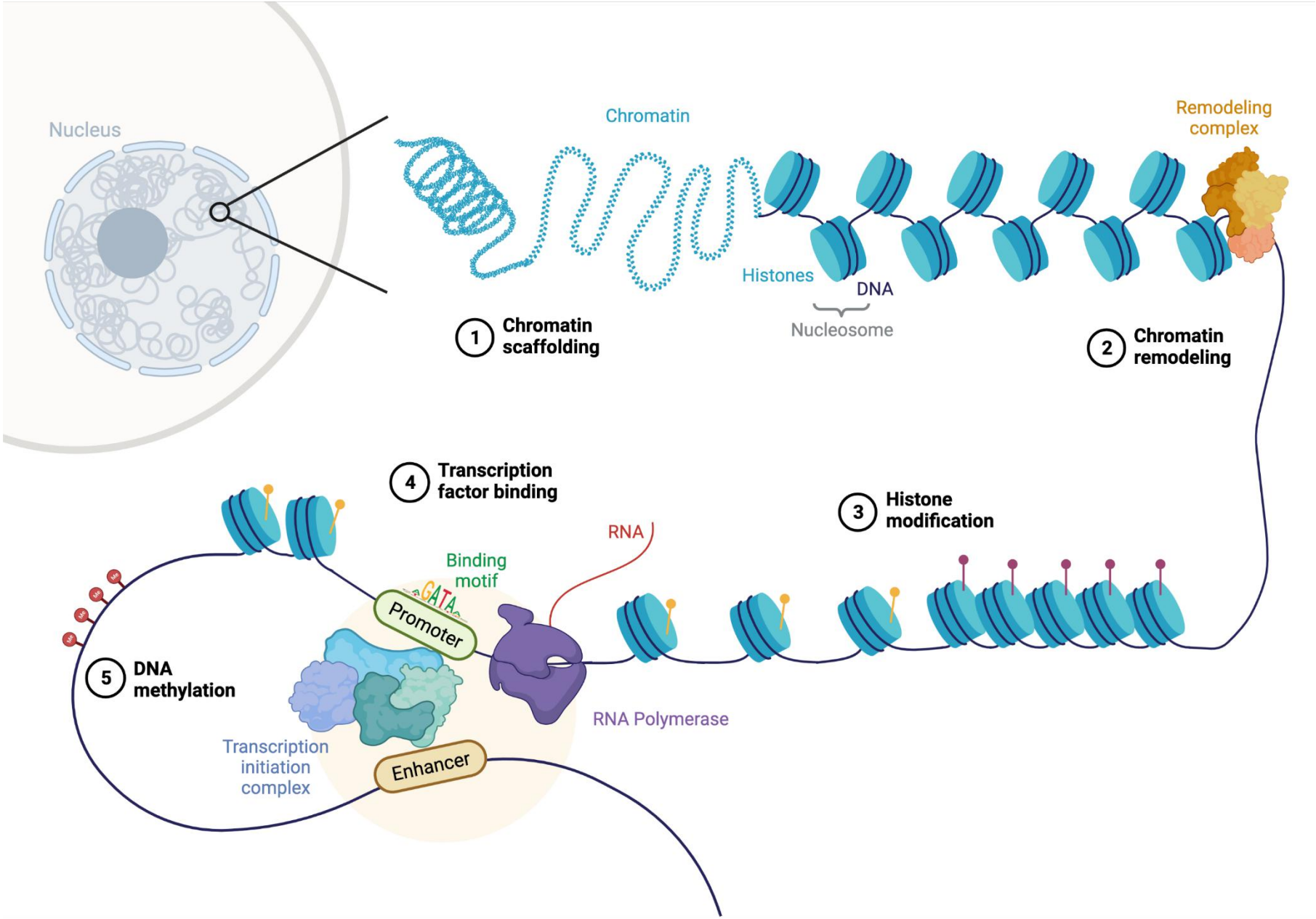


Figure. Overview of mechanisms influencing chromatin accessibility. Created with BioRender.com.

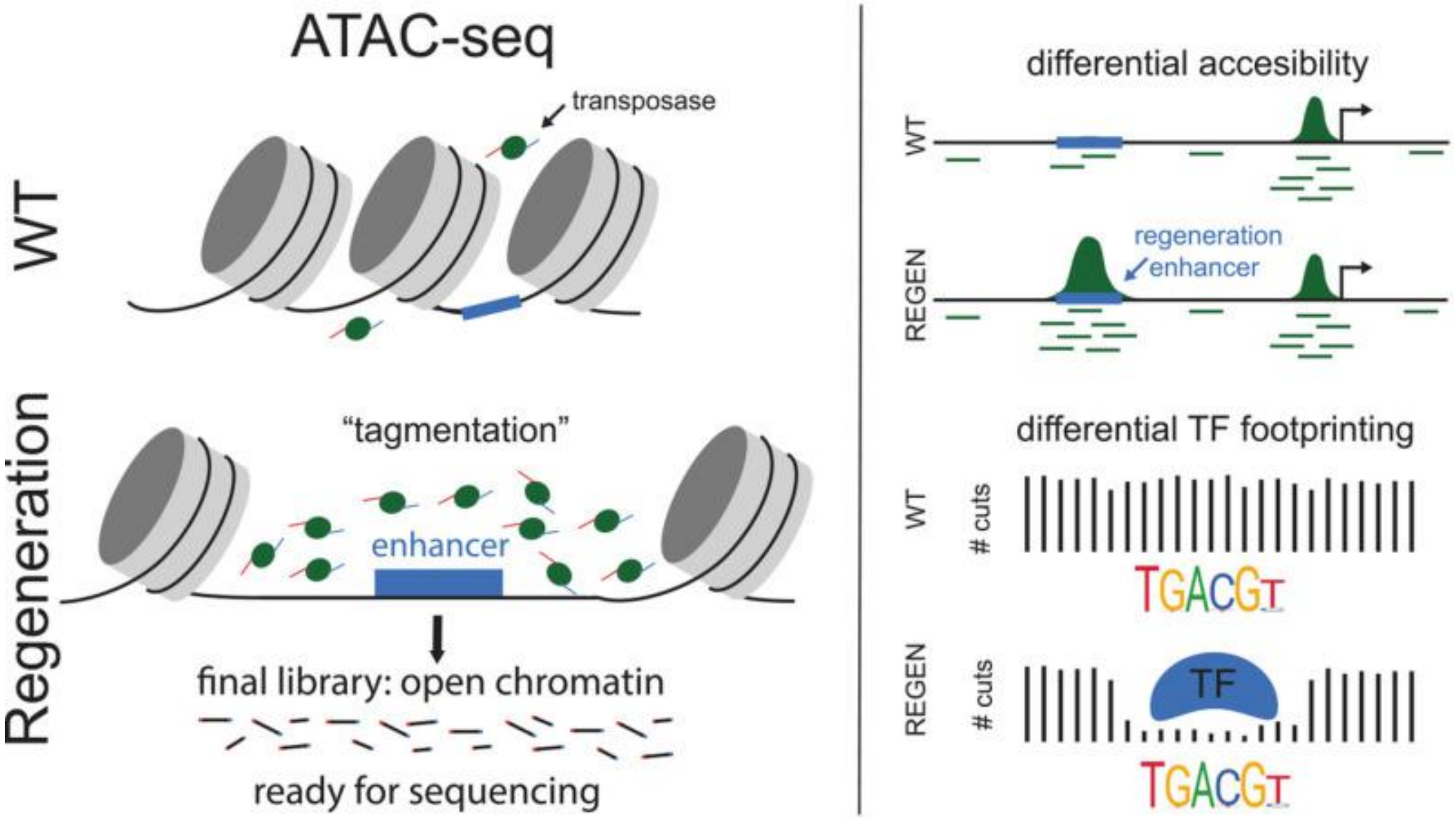


Figure. Overview of an ATAC-seq seq experiment to assay regeneration-responsive chromatin.

https://www.ncbi.nlm.nih.gov/books/NBK586944/figure/ch29.Fig1/?report=objectonly

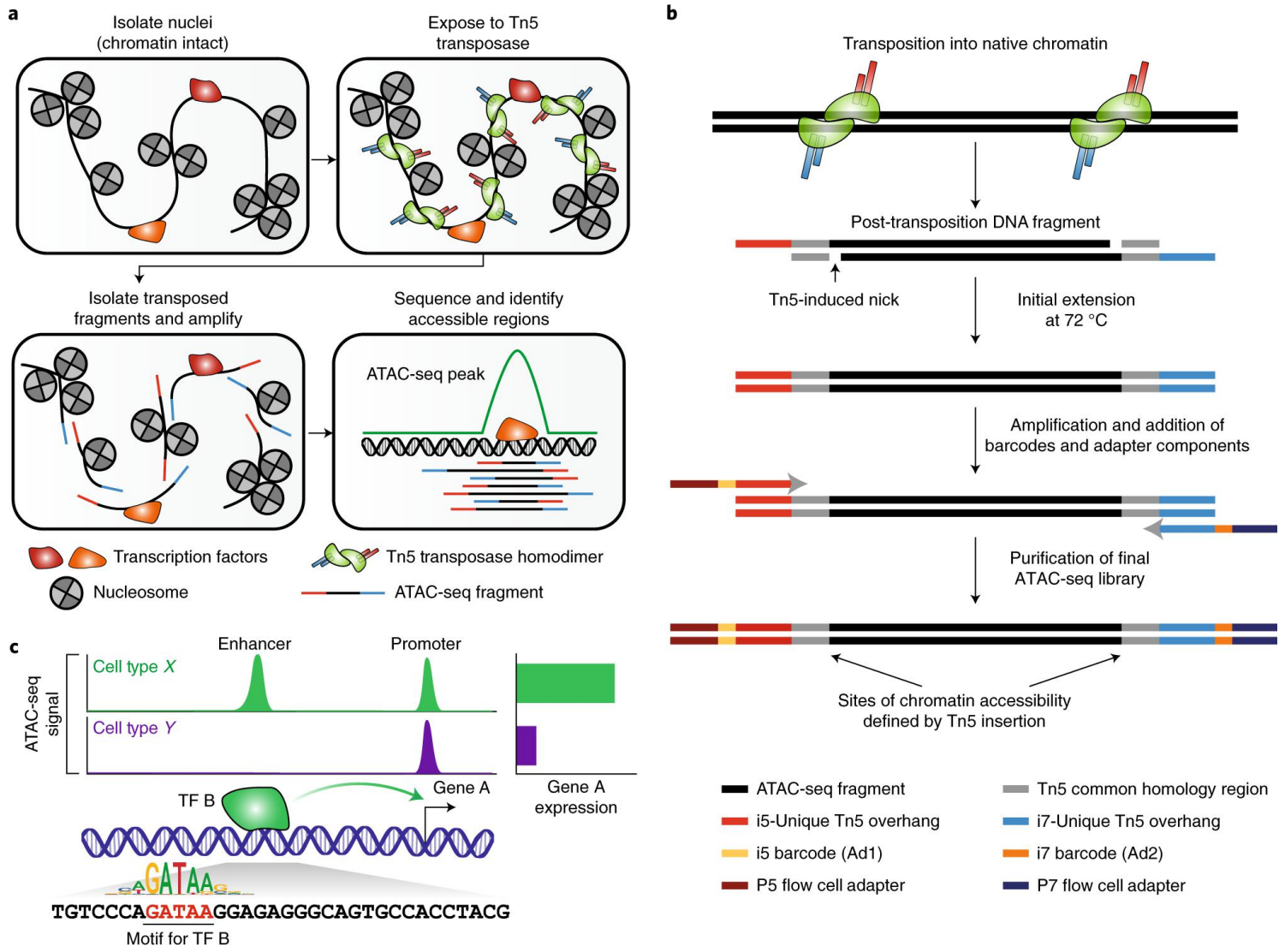
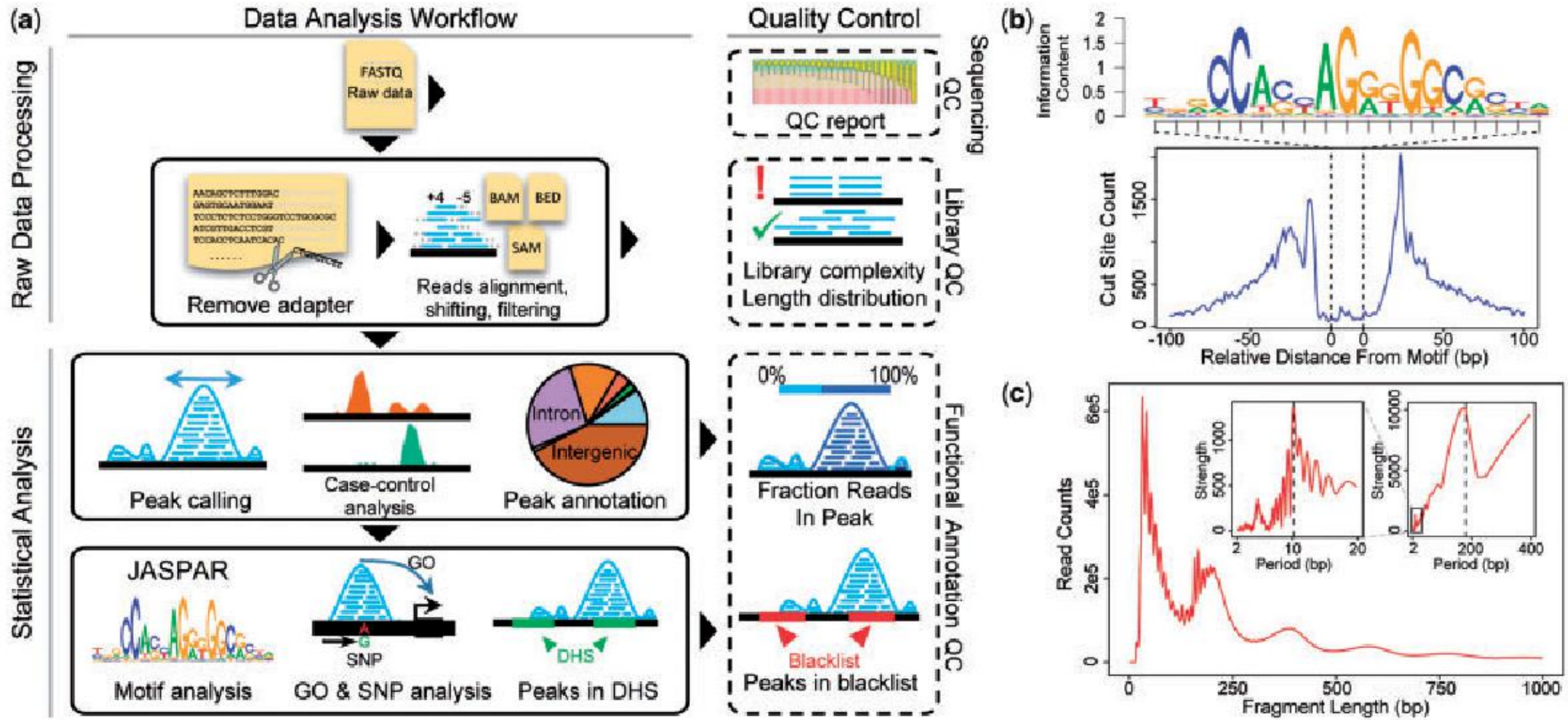


Figure. Schematic of the ATAC-seq transposition reaction and library preparation.

https://www.nature.com/articles/s41596-022-00692-9

Bulk ATAC-seq analysis



DOI:10.1093/bioinformatics/bty141

Single-cell ATAC-seq

scATAC-seq analysis

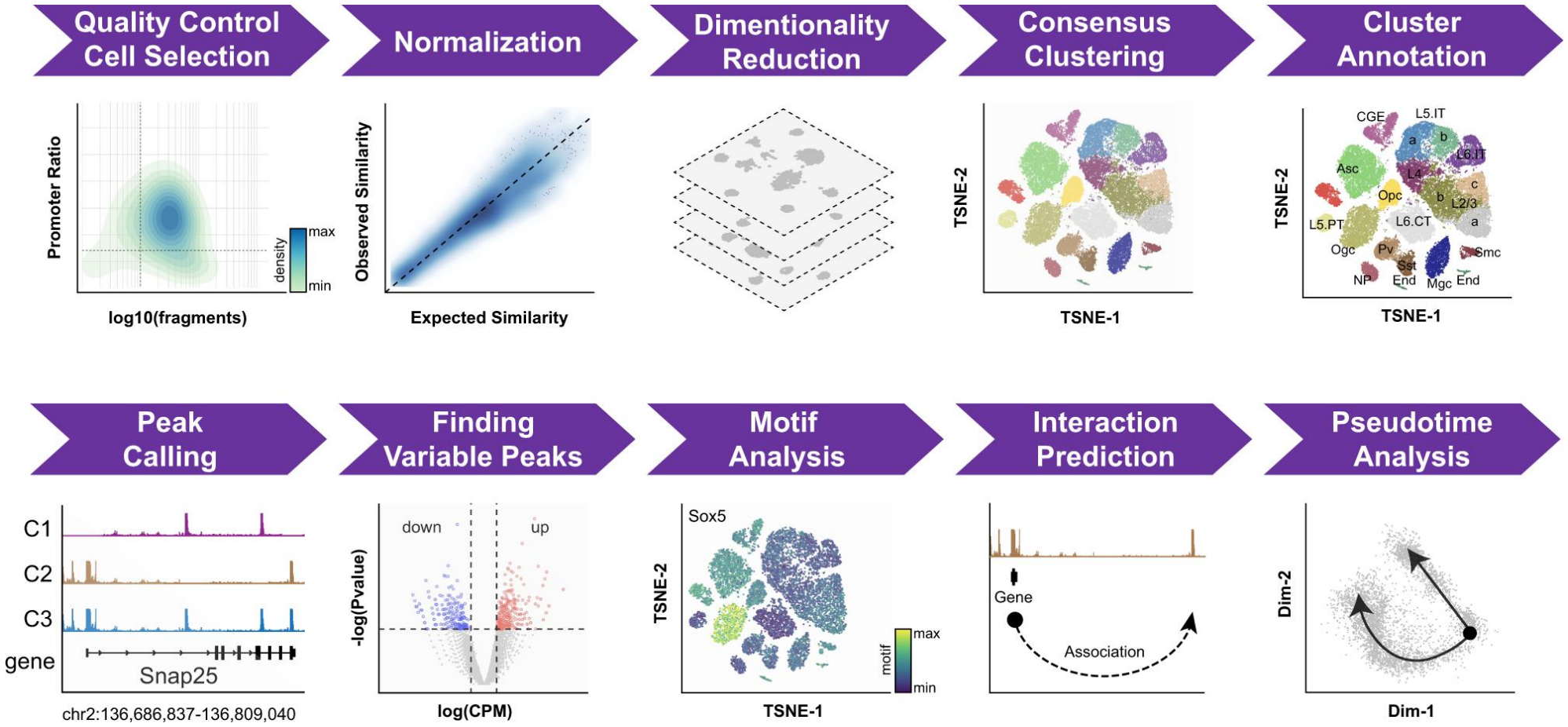
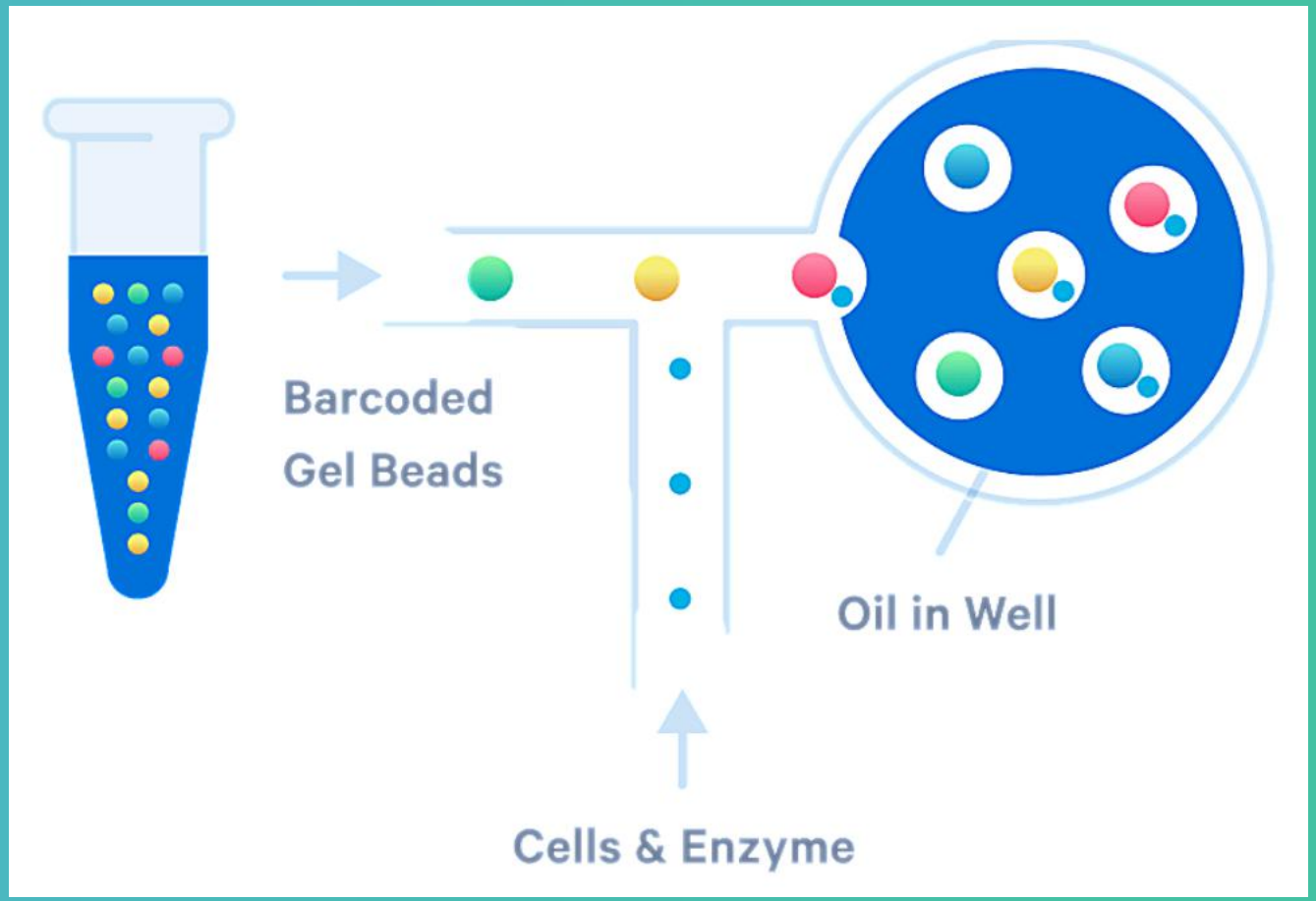
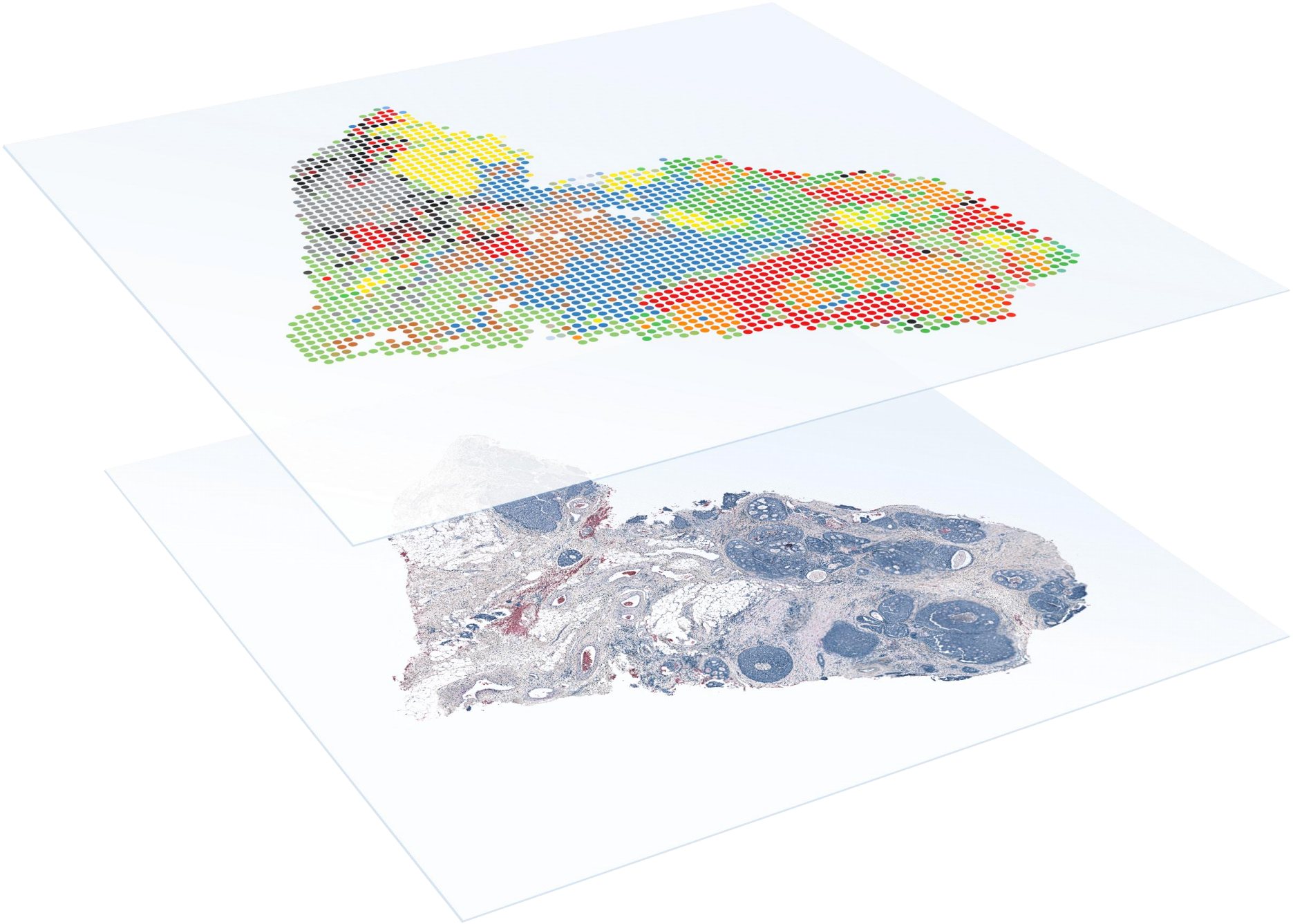


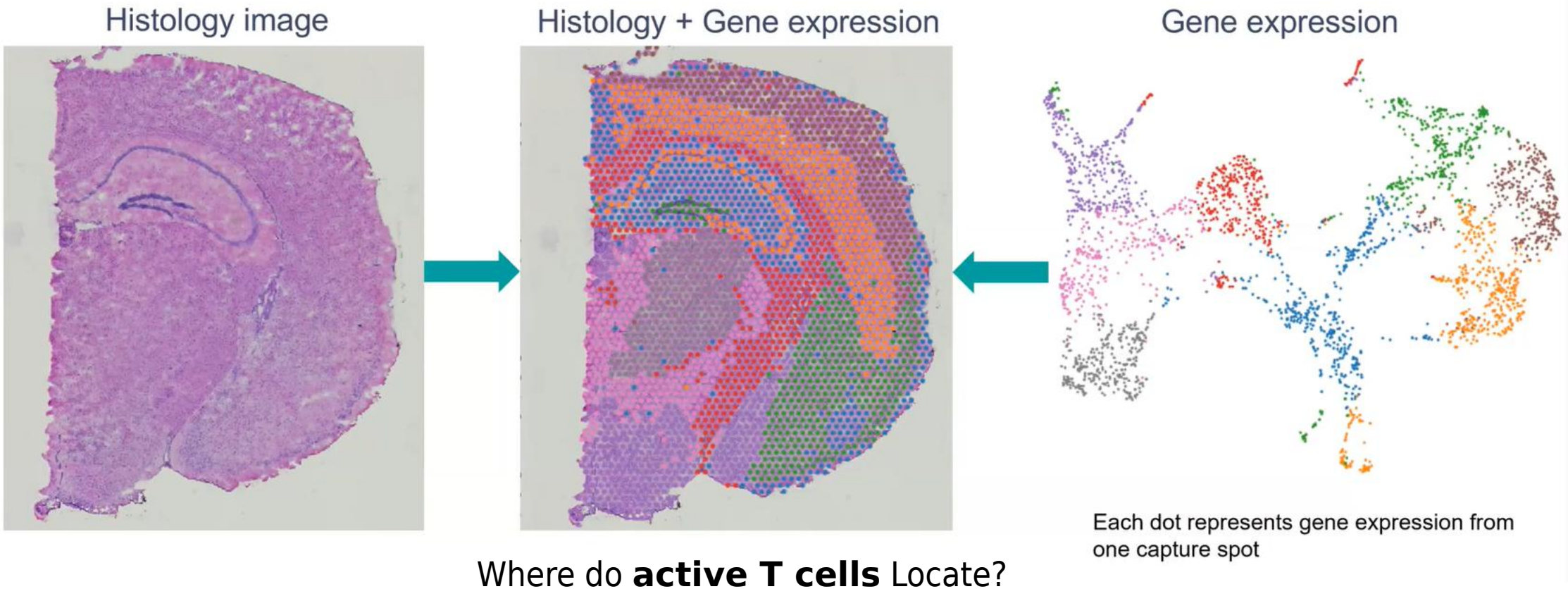
Figure: Schematic overview of SnapATAC analysis workflow.

Chromium Single Cell



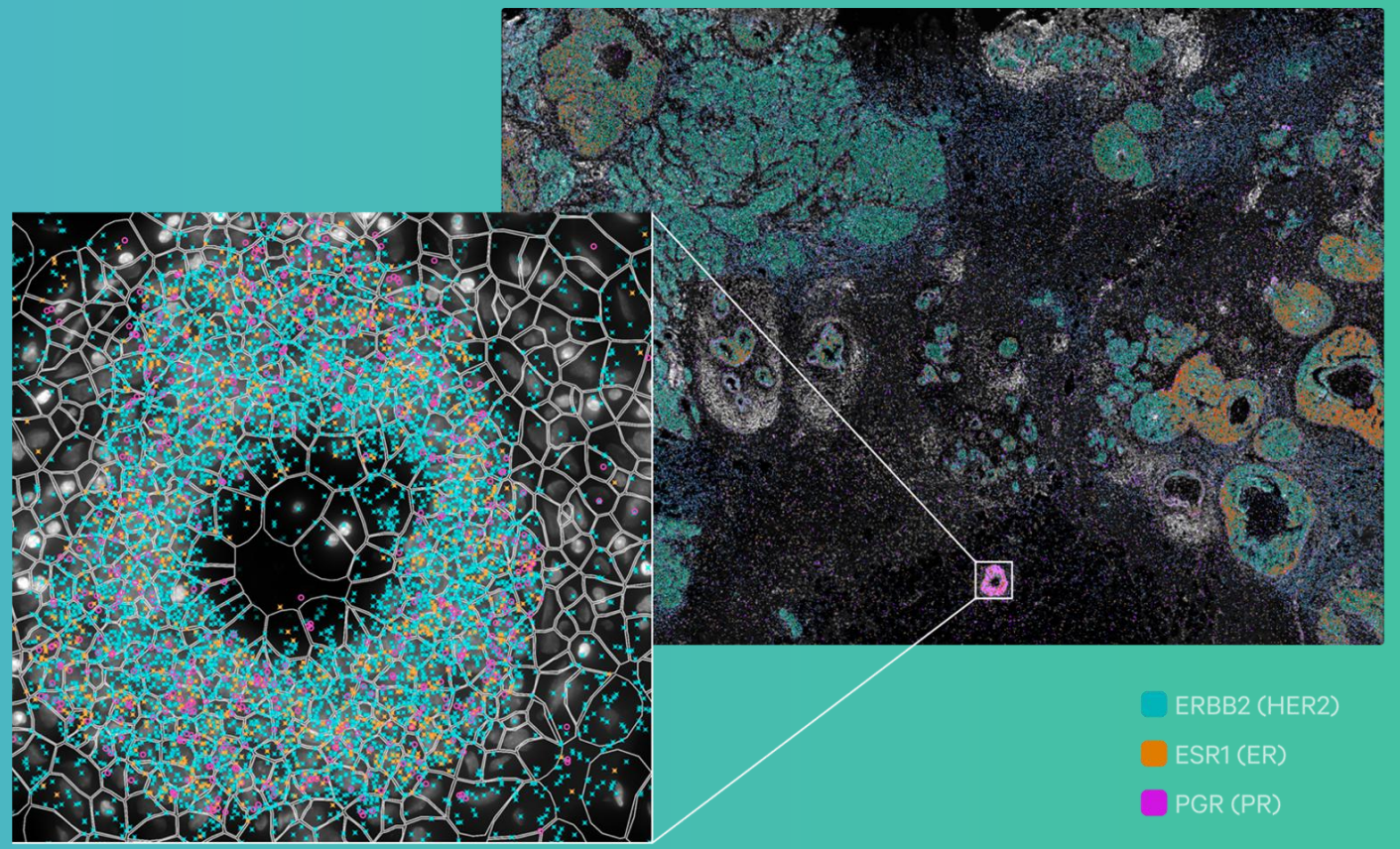


Spatial transcriptomics with Visium



Where do **active T cells** Locate?

Xenium In Situ



In this breast cancer sample
(Stage II-B, ER+/PR−/HER2+), Xenium identifies a
previously unknown triple-positive region.

Pacific Informatics
Your comprehensive bioinformatics solution

ZYMO RESEARCH

THANK YOU FOR YOUR ATTENTION and Q&A!

Please contact:loi.lp@pacificinformatics.com.vn
For further information!