Exploring RNA-seq as a high throughput approach: Principles, Techniques, and Strategies for data analysis.

Ph.D. End Of the Year Seminar 2024/2025

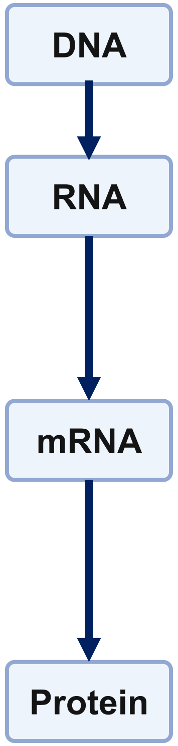
Ph.D. candidate Cecilia Riani XL cycle

Supervisors:

Professor Giorgio Baiocco

Professor Alessandro Lascialfari

Human Genome and Transcriptome

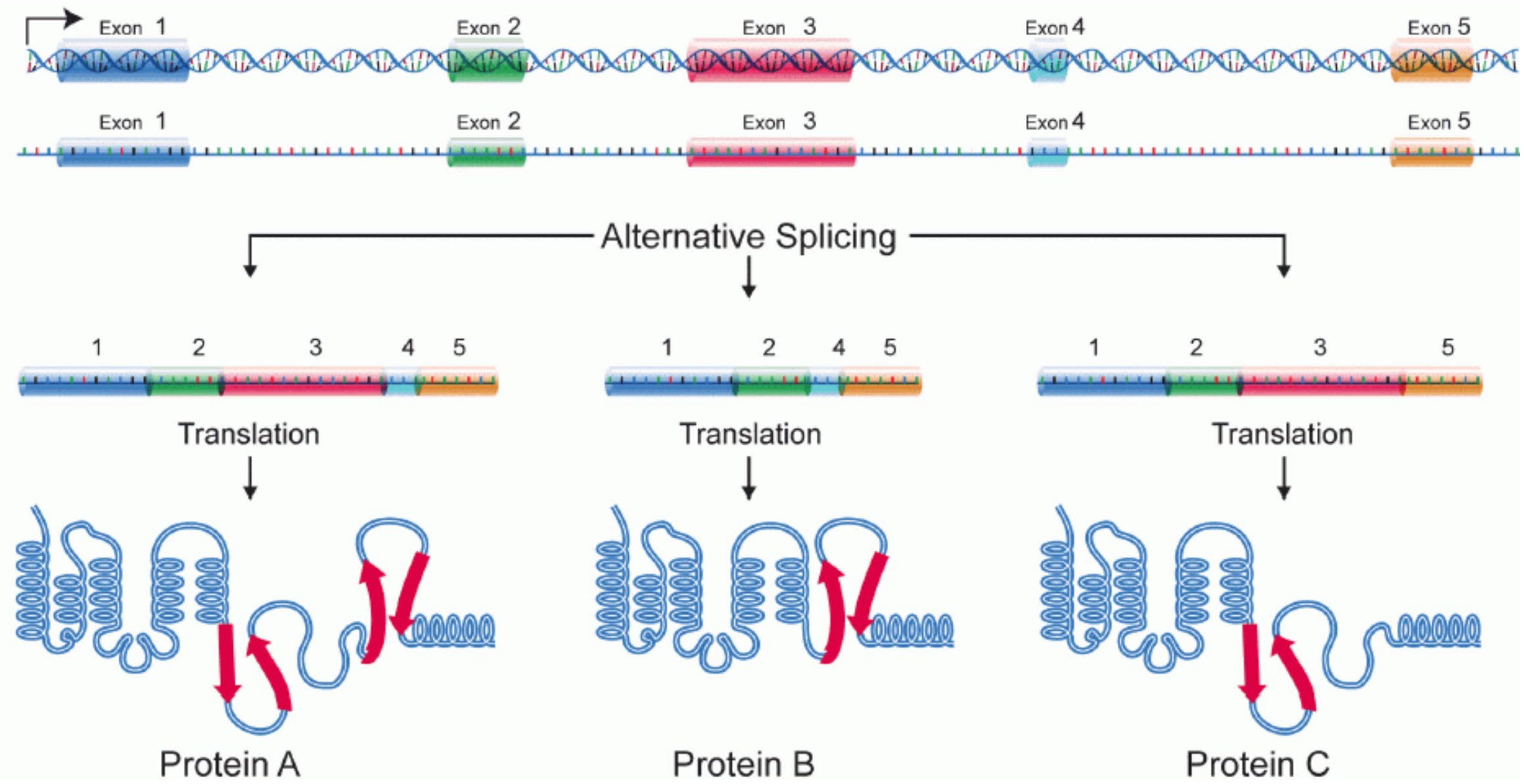


DNA

RNA

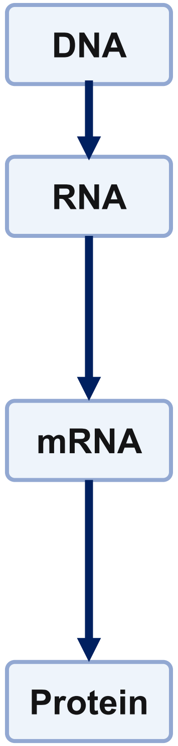
mRNA

Protein



DNA ≈ 3.1 billion base pairs
DNA ≈ 20000 coding genes (1-2% of the genome)
≈ million of mRNA molecules present in a cell

Human Genome and Transcriptome

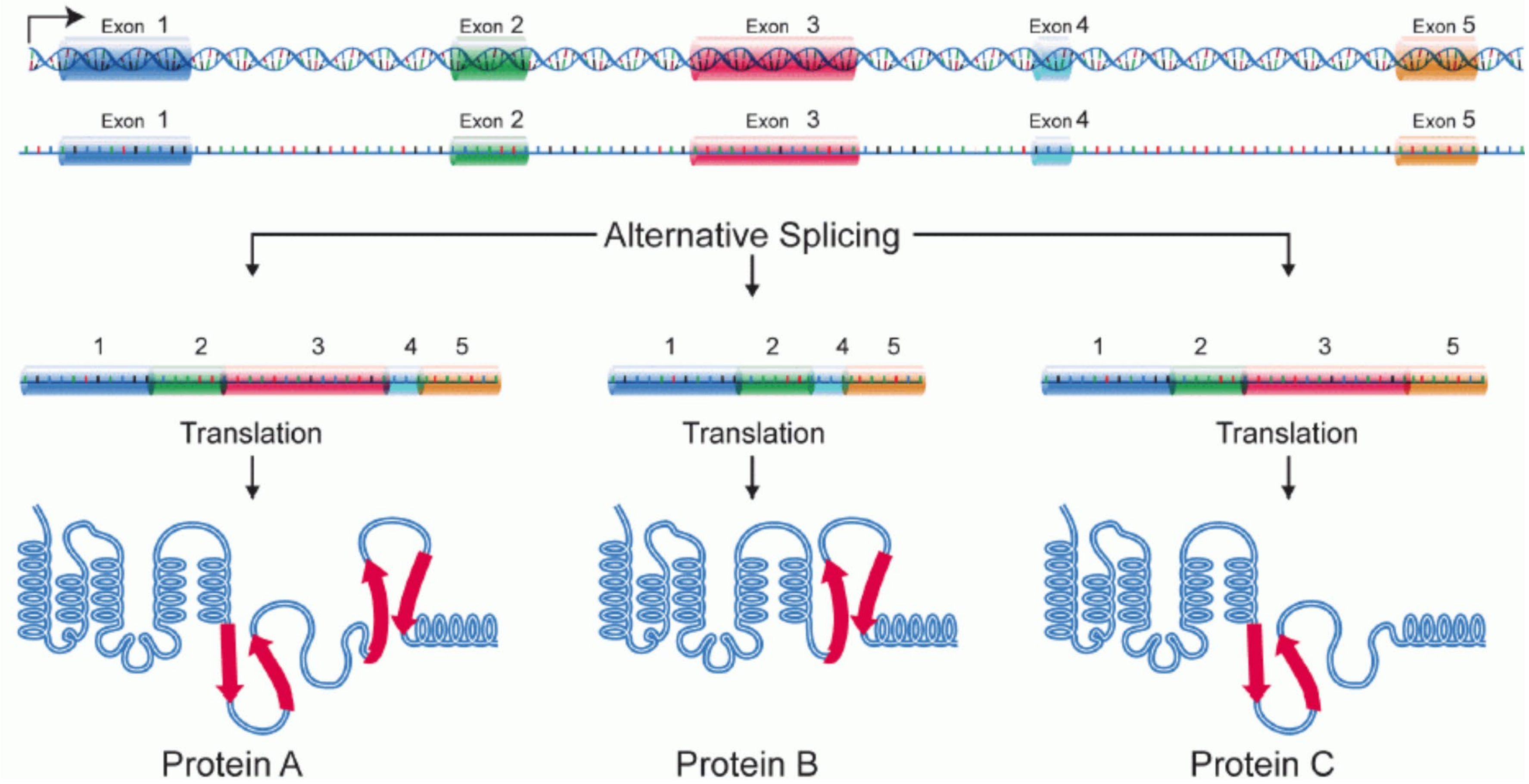


DNA

RNA

mRNA

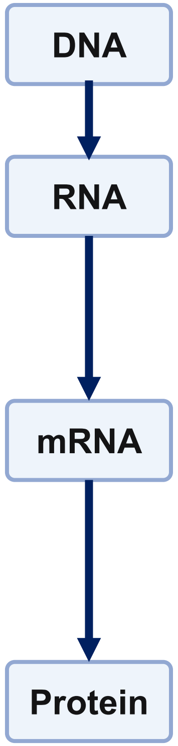
Protein

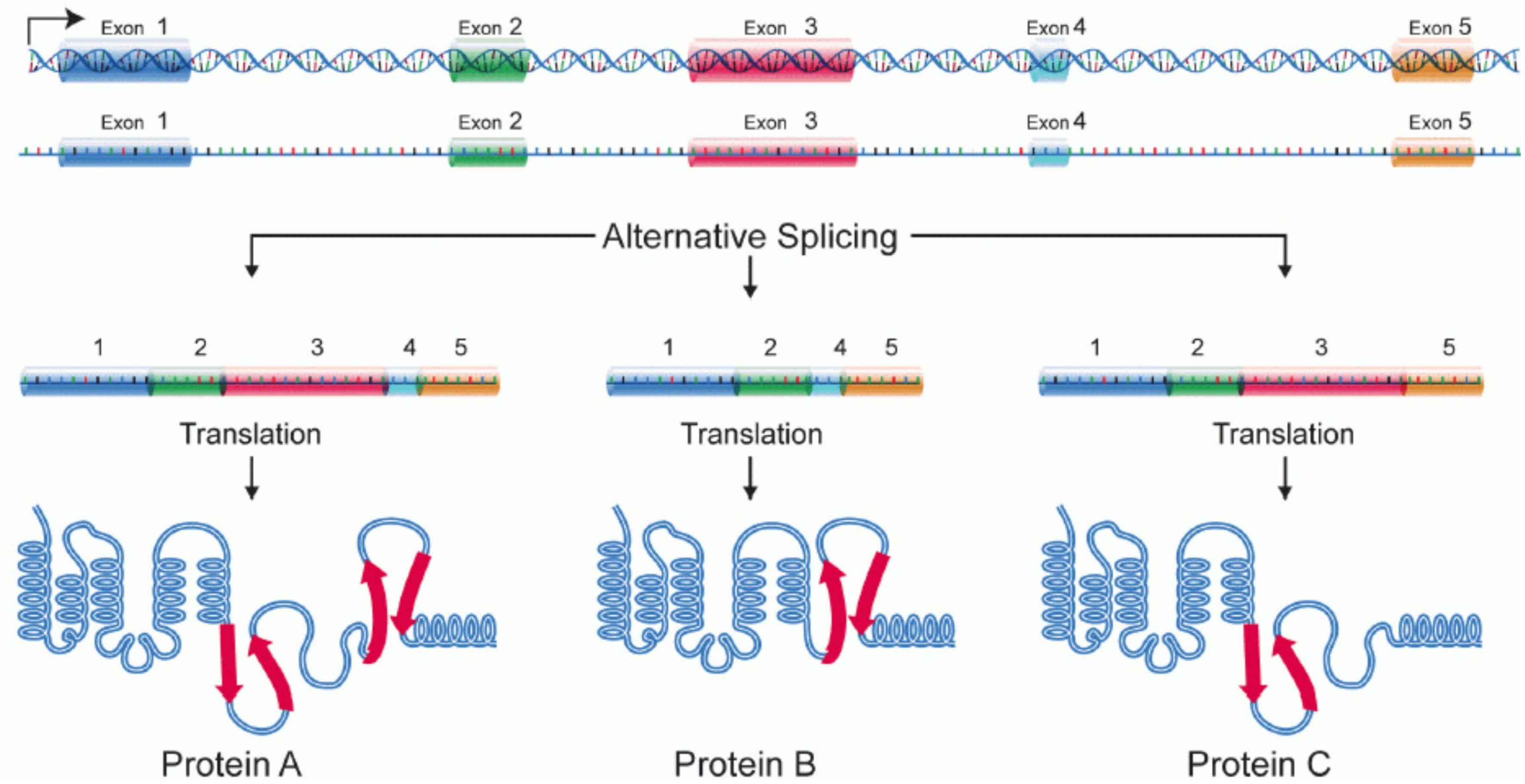


GENE

DNA ≈ 3.1 billion base pairs
DNA ≈ 20000 coding genes (1-2% of the genome)
≈ million of mRNA molecules present in a cell

Human Genome and Transcriptome





GENE

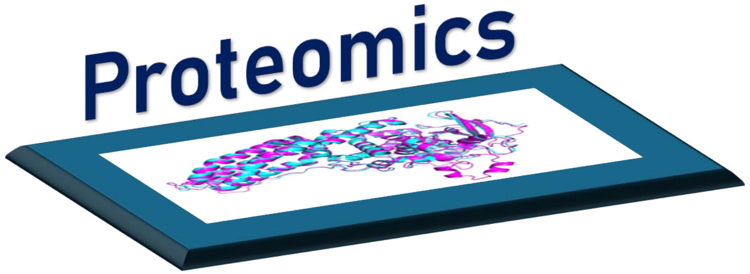
GENE EXPRESSION

DNA ≈ 3.1 billion base pairs
DNA ≈ 20000 coding genes (1-2% of the genome)
≈ million of mRNA molecules present in a cell

Multi-omics techniques

Genomics involves the study of the complete set of DNA in a cell

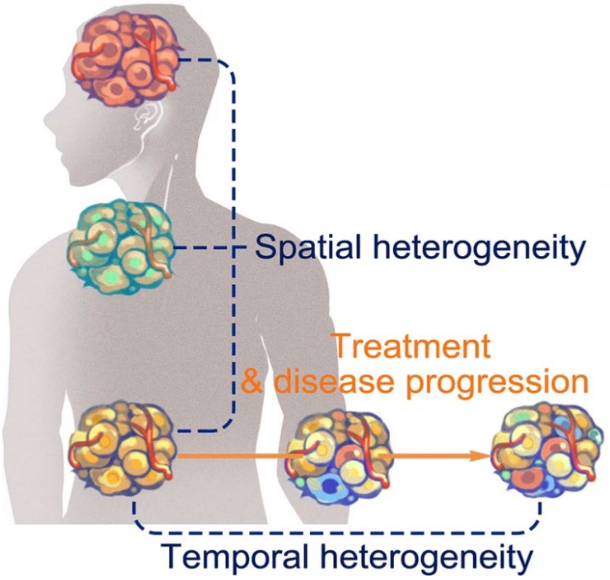
Transcriptomics involve the study of the complete set of RNAs (transcriptome) in a cell.

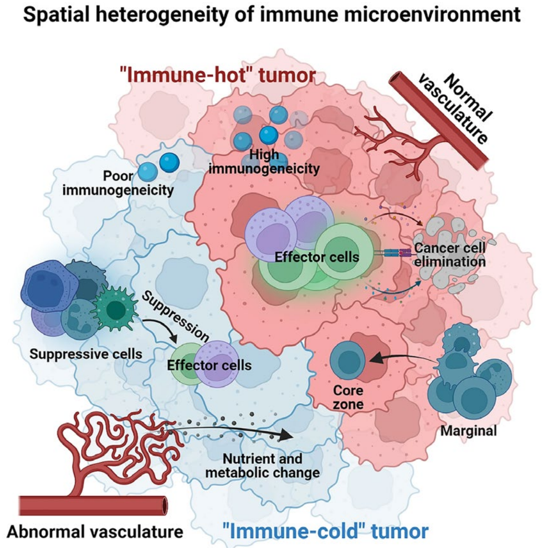


Proteomics involve the study of the entire set of proteins in a cell.

Cancer transcriptomics

Cancer Transcriptomics provides new insight into tumor heterogeneity and allows to decipher the biological pathway involved in tumor progression and tumor treatment response

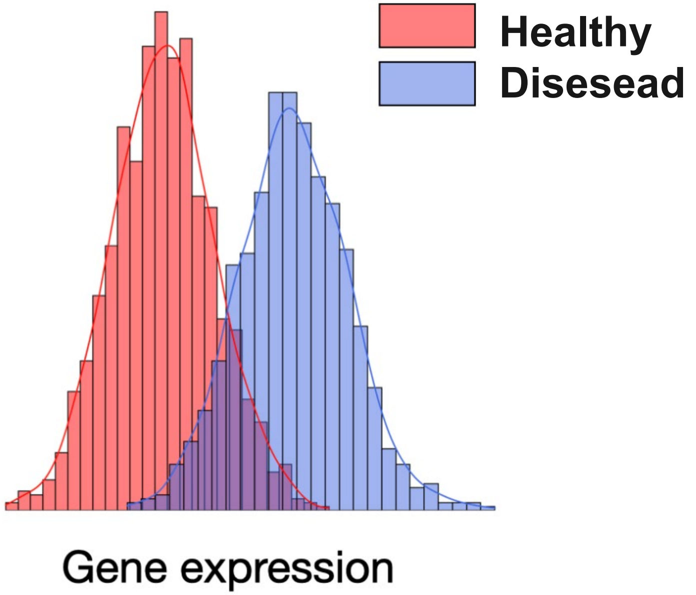



Spatial heterogeneity of immune microenvironment

Long et al. 2022

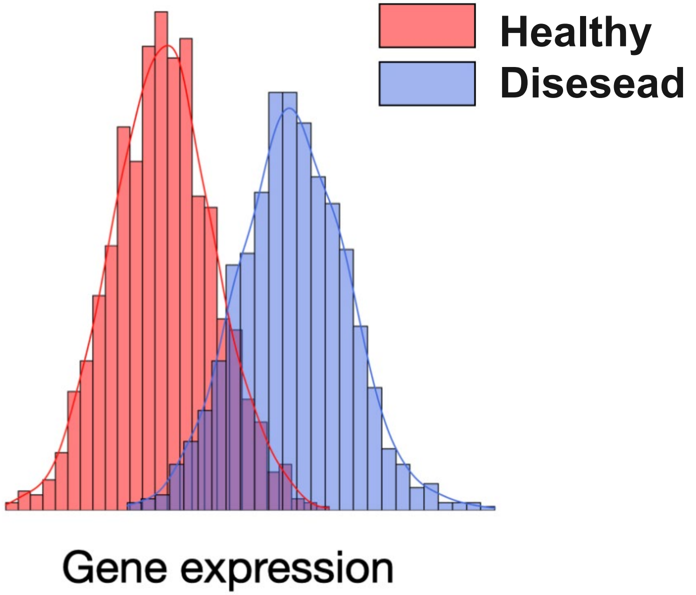
Differential gene expression (DGE) analysis

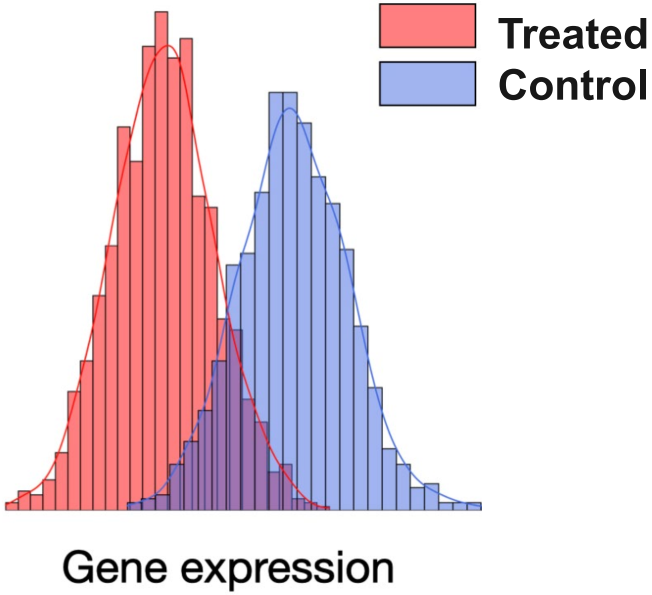
- ☐ *Is there a statistically significant change in gene expression between two groups?*



Differential gene expression (DGE) analysis

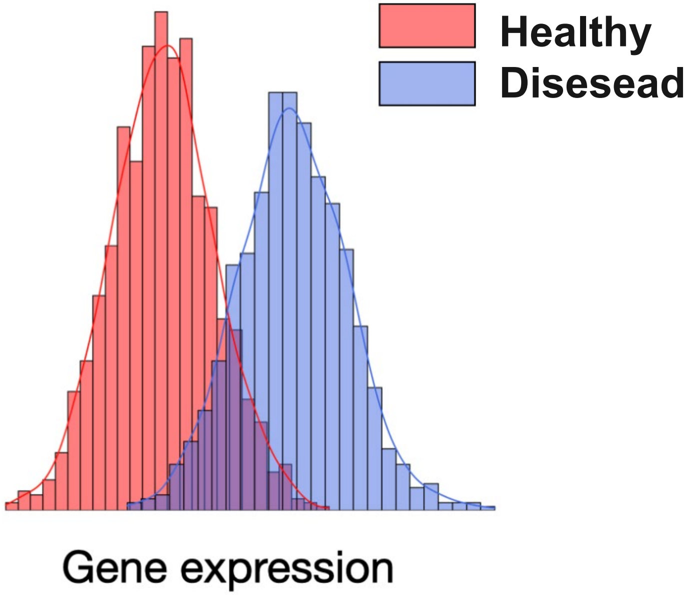
- *Is there a statistically significant change in gene expression between two groups?*

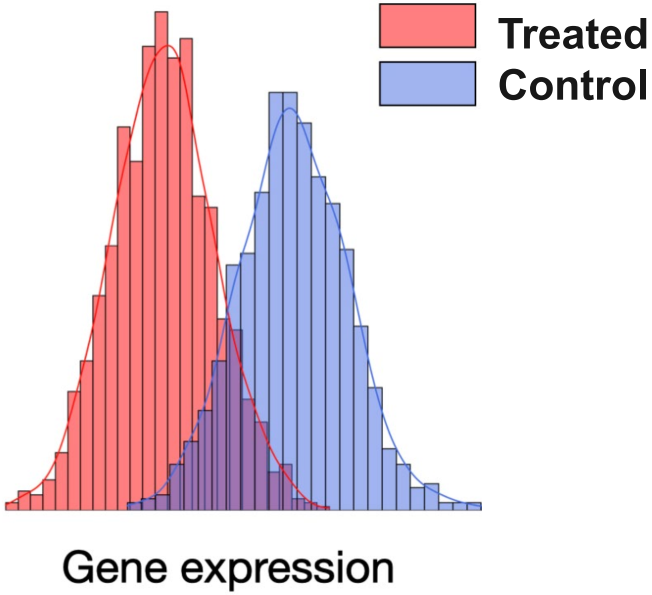




Differential gene expression (DGE) analysis

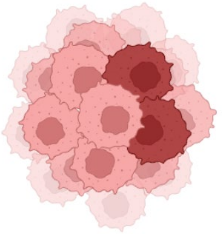
☐ *Is there a statistically significant change in gene expression between two groups?*





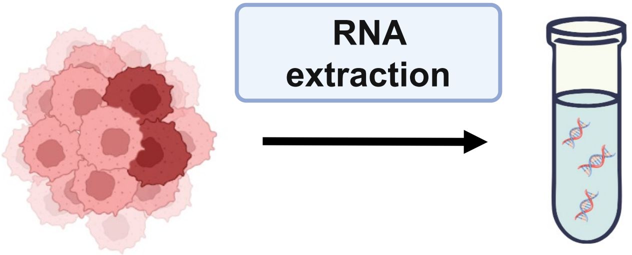
Quantify properly gene expression!!

RNA sequencing through Next-Generation Sequencing



Sequencing depth and coverage: key considerations in genomic analyses. Ponting et al. 2014
Coming of age: ten years of next-generation sequencing technologies Richard McCombine 2016

RNA sequencing through Next-Generation Sequencing



Sequencing depth and coverage: key considerations in genomic analyses. Ponting et al. 2014
Coming of age: ten years of next-generation sequencing technologies Richard McCombine 2016

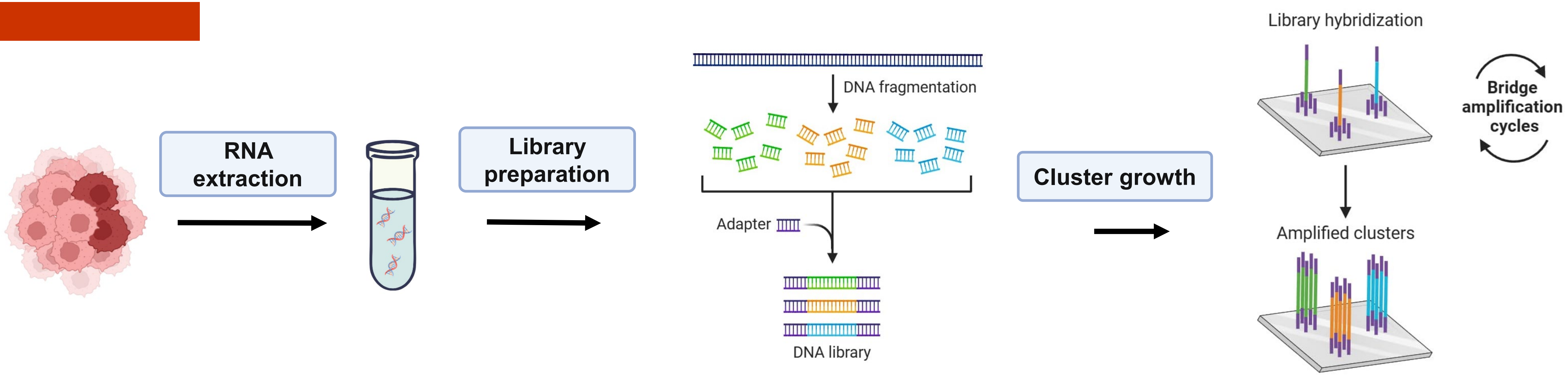
RNA sequencing through Next-Generation Sequencing

RNA extraction

Library preparation

Sequencing depth and coverage: key considerations in genomic analyses. Ponting et al. 2014
Coming of age: ten years of next-generation sequencing technologies Richard McCombine 2016

RNA sequencing through Next-Generation Sequencing



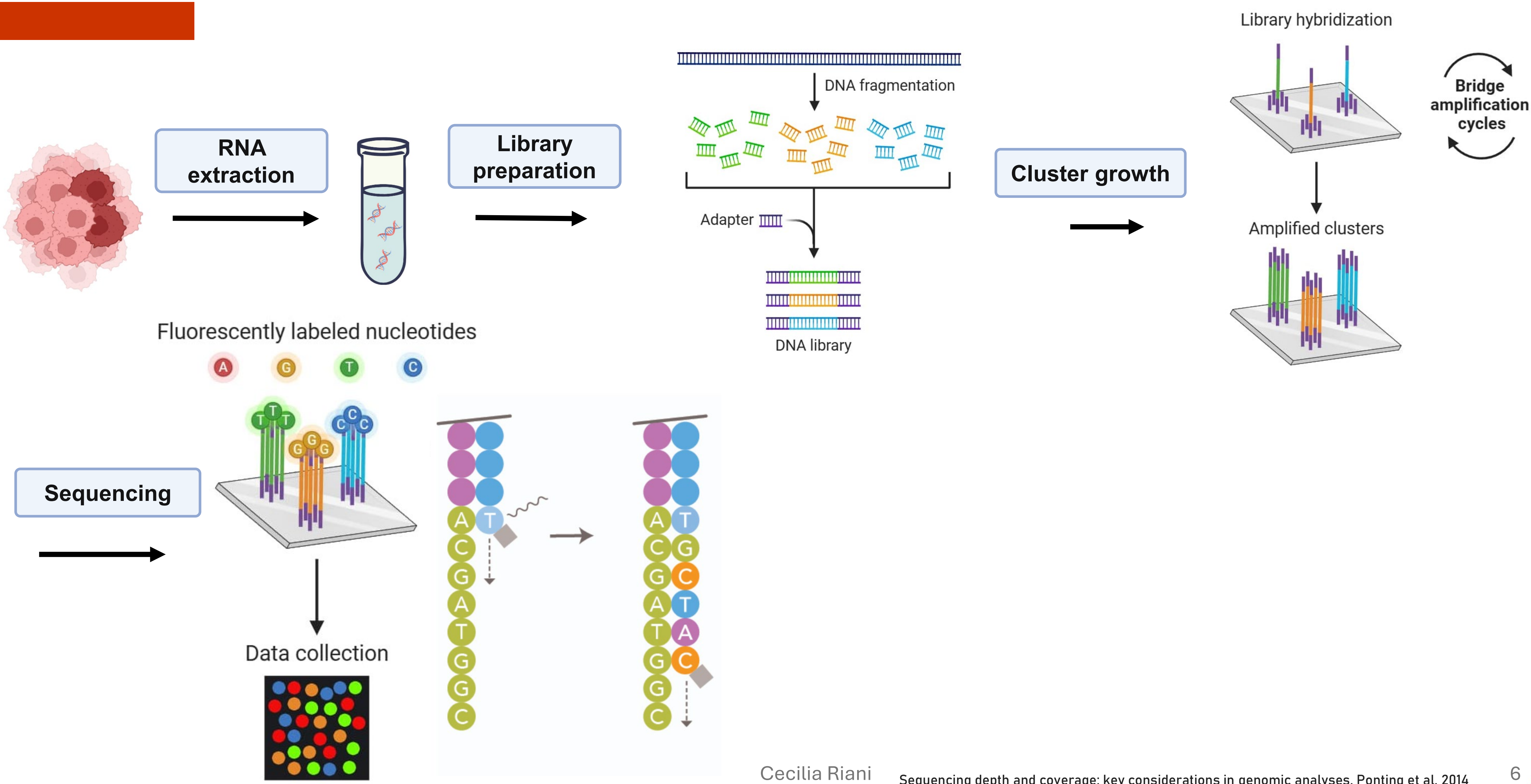
RNA extraction

Library preparation

Cluster growth

Sequencing depth and coverage: key considerations in genomic analyses. Ponting et al. 2014
Coming of age: ten years of next-generation sequencing technologies Richard McCombine 2016

RNA sequencing through Next-Generation Sequencing



RNA extraction

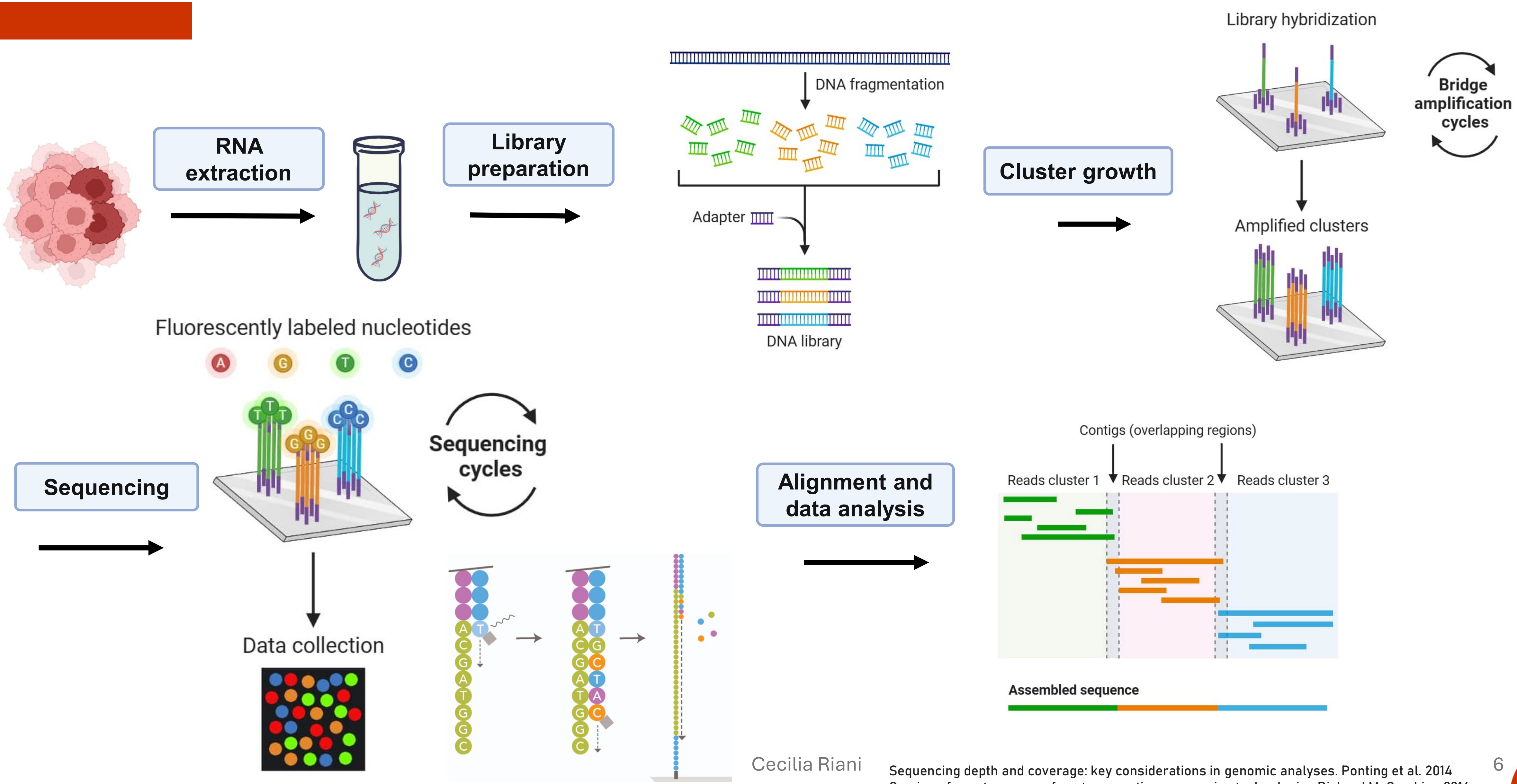
Library preparation

Cluster growth

Sequencing

Sequencing depth and coverage: key considerations in genomic analyses. Ponting et al. 2014
Coming of age: ten years of next-generation sequencing technologies Richard McCombine 2016

RNA sequencing through Next-Generation Sequencing



RNA extraction

Library preparation

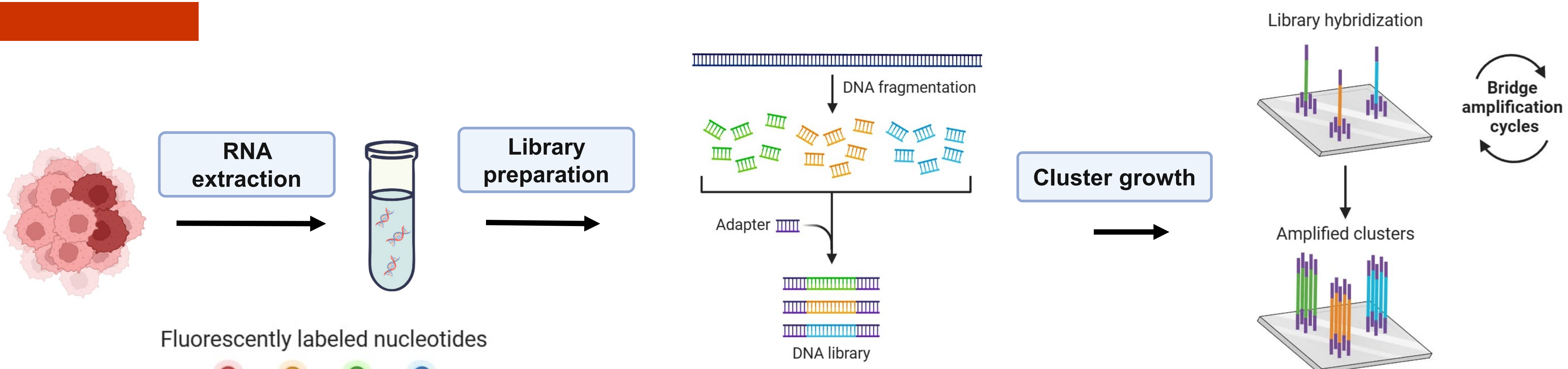
Cluster growth

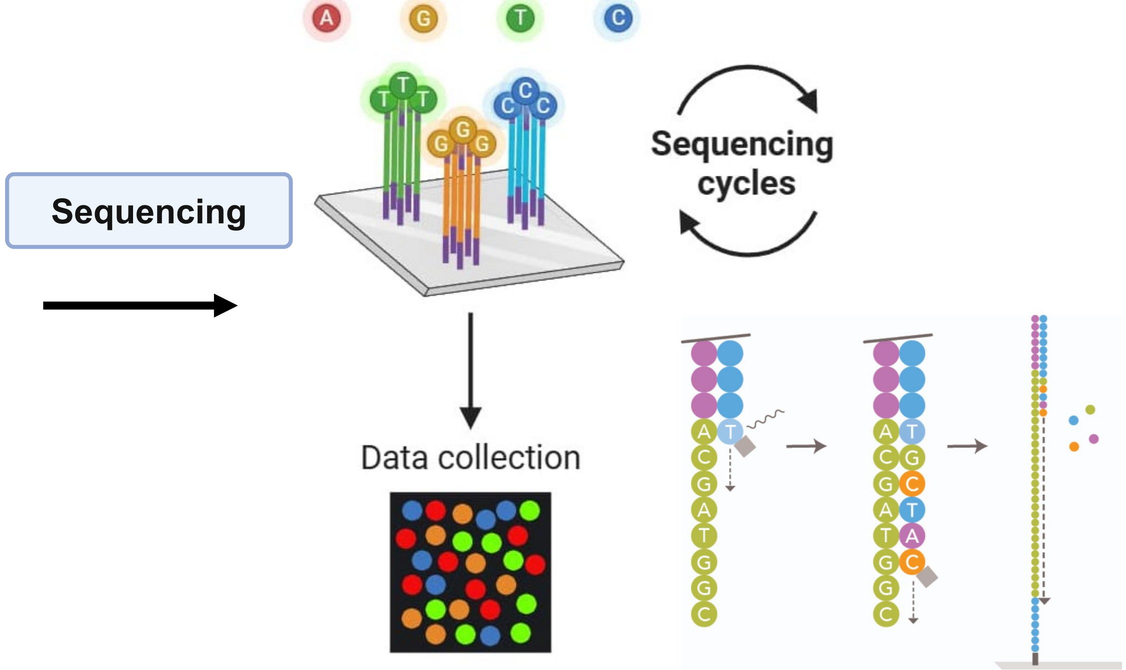
Sequencing

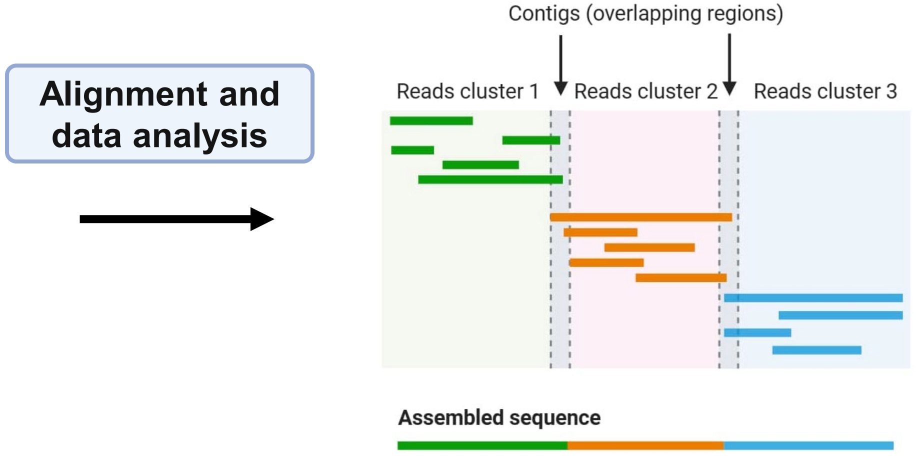
Alignment and data analysis

Sequencing depth and coverage: key considerations in genomic analyses. Ponting et al. 2014
Coming of age: ten years of next-generation sequencing technologies Richard McCombine 2016

RNA sequencing through Next-Generation Sequencing



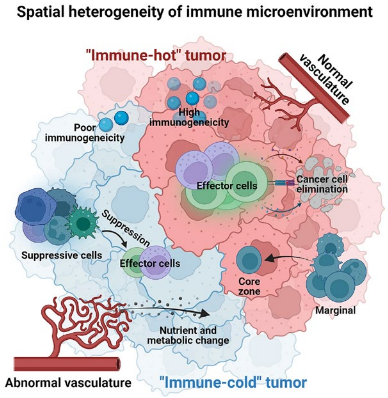




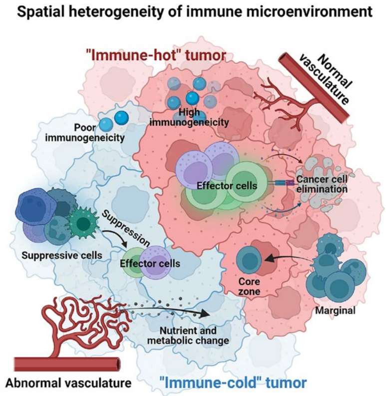
GENES	EXPRESSION
Gene A	4
Gene B	10
Gene C	2
...	...

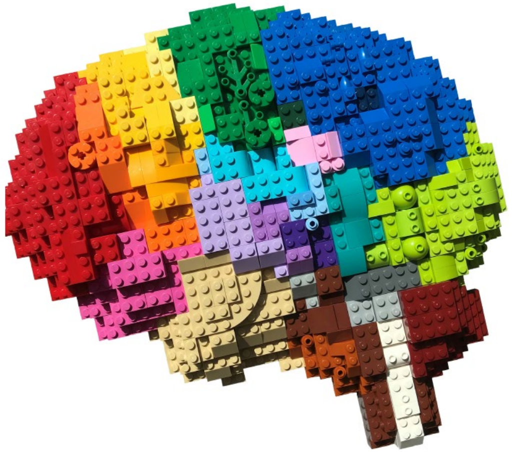
Sequencing depth and coverage: key considerations in genomic analyses. Ponting et al. 2014
Coming of age: ten years of next-generation sequencing technologies Richard McCombine 2016

Cancer Transcriptomics

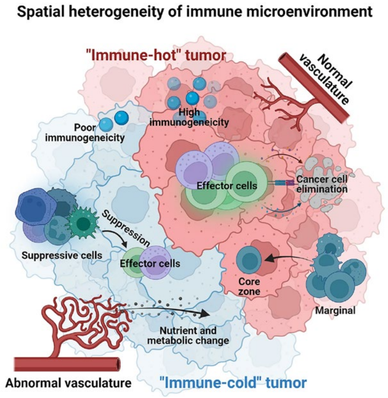

Spatial heterogeneity of immune microenvironment

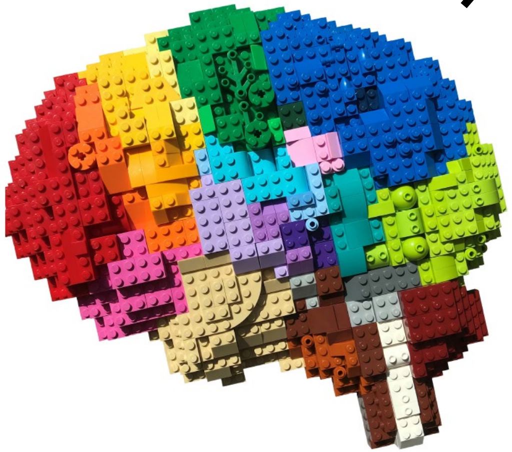
Cancer Transcriptomics



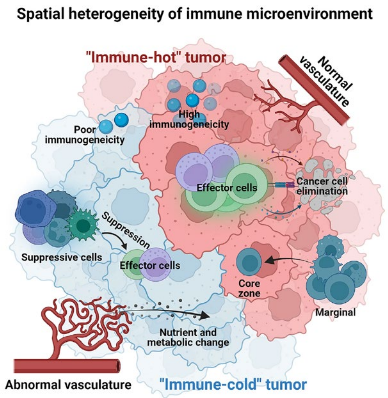


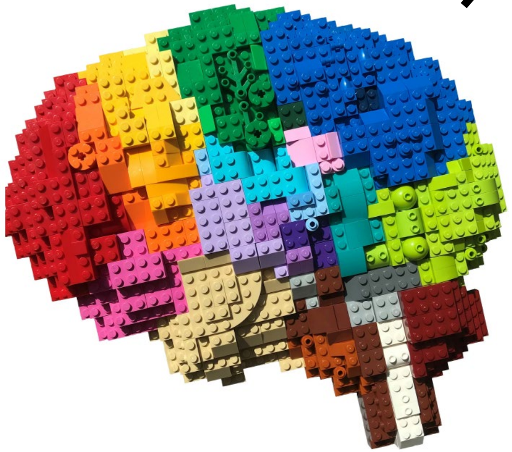
Cancer Transcriptomics





Cancer Transcriptomics

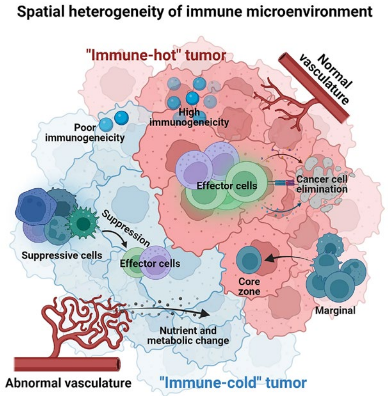


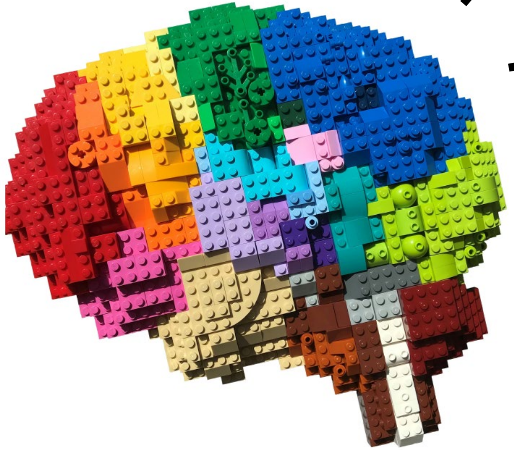


Bulk RNAseq

GENES	EXPRESSION
Gene A	8
Gene B	10
Gene C	2
…	…

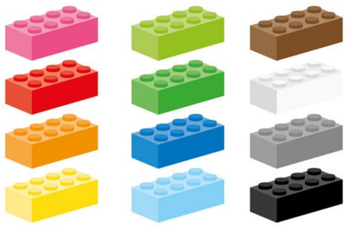
Cancer Transcriptomics



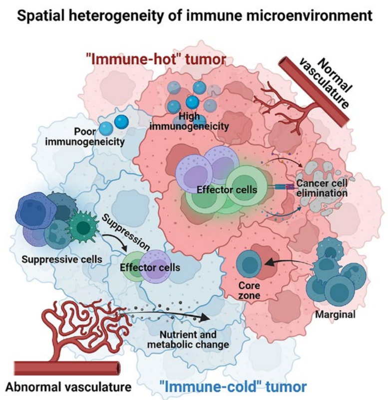


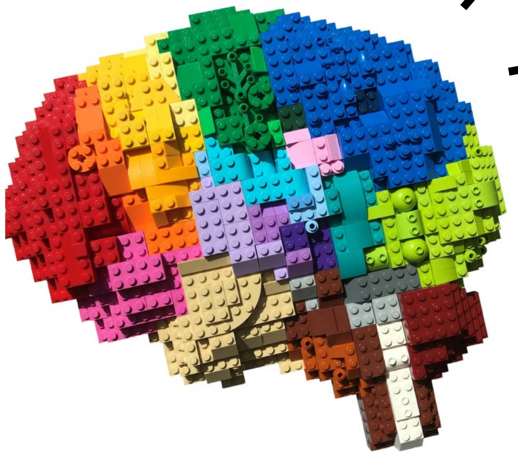
Bulk RNAseq

GENES	EXPRESSION
Gene A	0
Gene B	10
Gene C	2
...	...



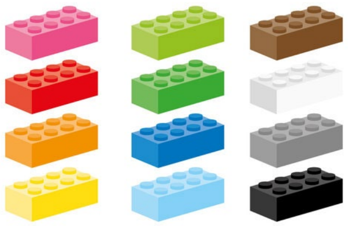
Cancer Transcriptomics





Bulk RNAseq

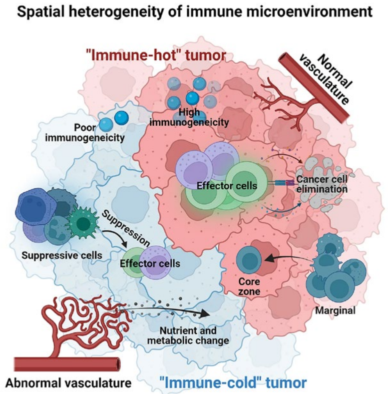
GENES	EXPRESSION
Gene A	0
Gene B	10
Gene C	2
...	...

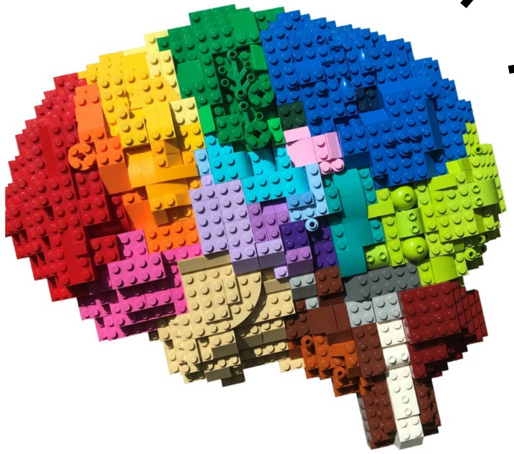


scRNAseq

GENES	Cell 1	Cell 2	Cell 3	...	Cell N
Gene A	0	1	4		5
Gene B	10	5	3		4
Gene C	2	10	0		0
......	...				

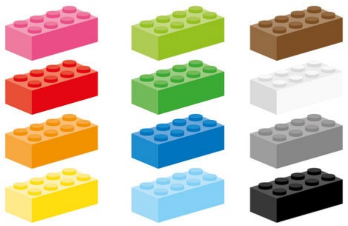
Cancer Transcriptomics





Bulk RNAseq

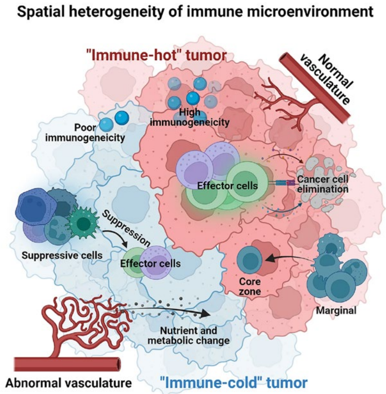
GENES	EXPRESSION
Gene A	0
Gene B	10
Gene C	2
...	...

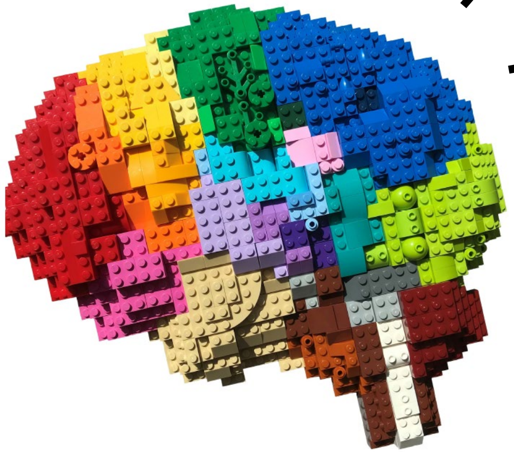


scRNAseq

GENES	Cell 1	Cell 2	Cell 3	...	Cell N
Gene A	0	1	4		5
Gene B	10	5	3		0
Gene C	2	10	0		0
……	...				

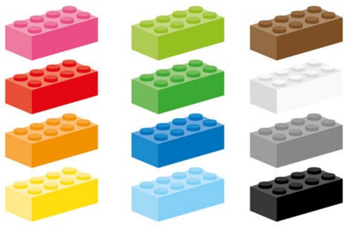
Cancer Transcriptomics


Spatial heterogeneity of immune microenvironment



Bulk RNAseq

GENES	EXPRESSION
Gene A	0
Gene B	10
Gene C	2
...	...



scRNAseq

GENES	Cell 1	Cell 2	Cell 3	...	Cell N
Gene A	0	1	4		5
Gene B	10	5	3		0
Gene C	2	10	0		0
......	...				

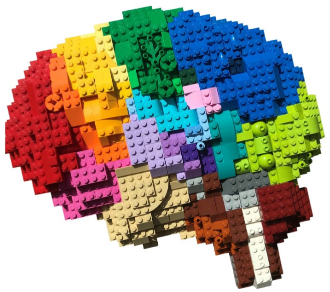
Spatial RNAseq

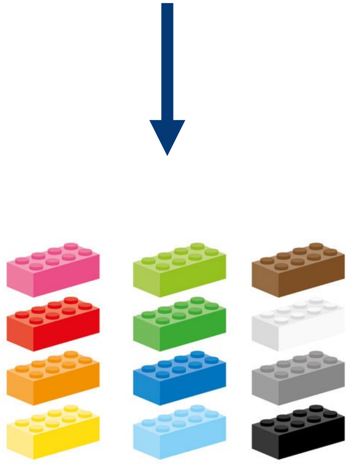
X	Y

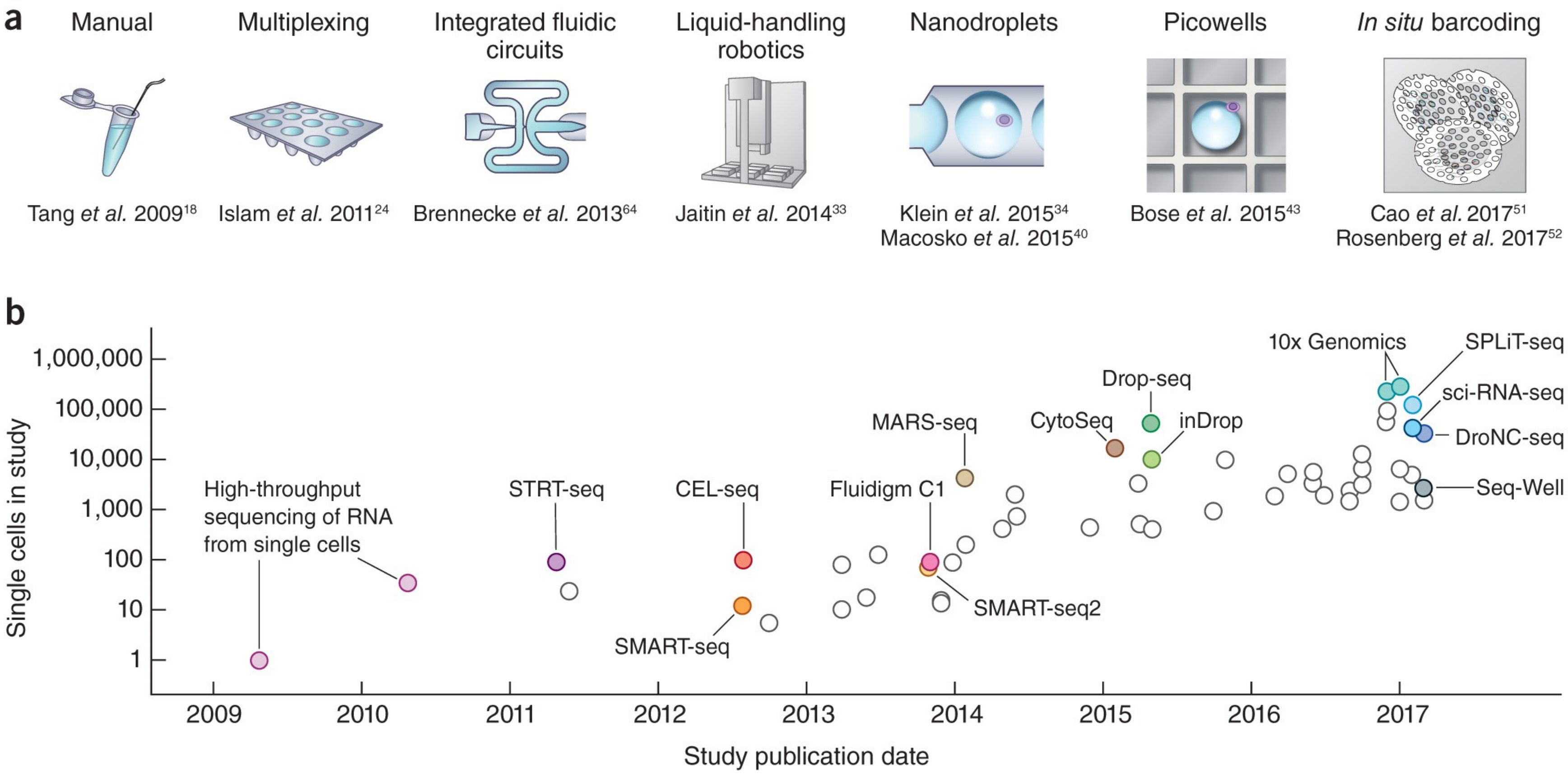
GENES	EXPRESSION
Gene A	0
Gene B	10
Gene C	2
...	...

RNA seq techniques

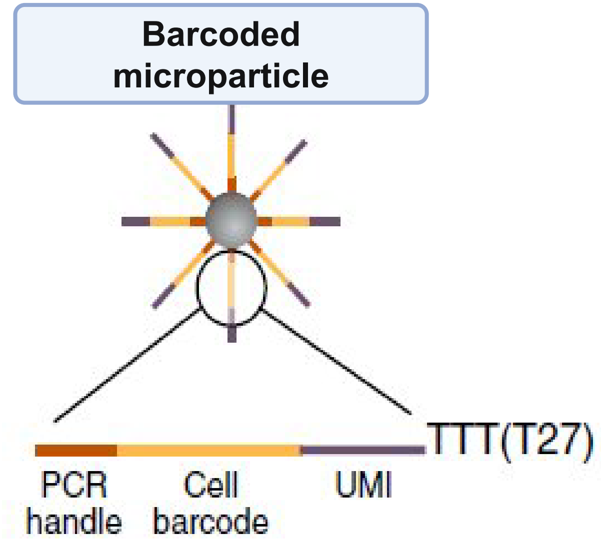
Single cell RNA sequencing







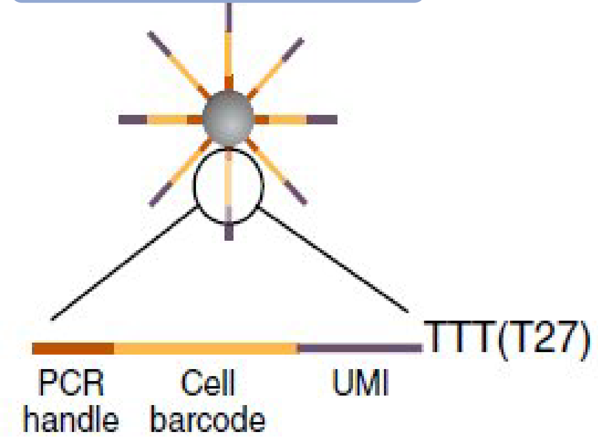
Single cell RNA sequencing: Drop-seq



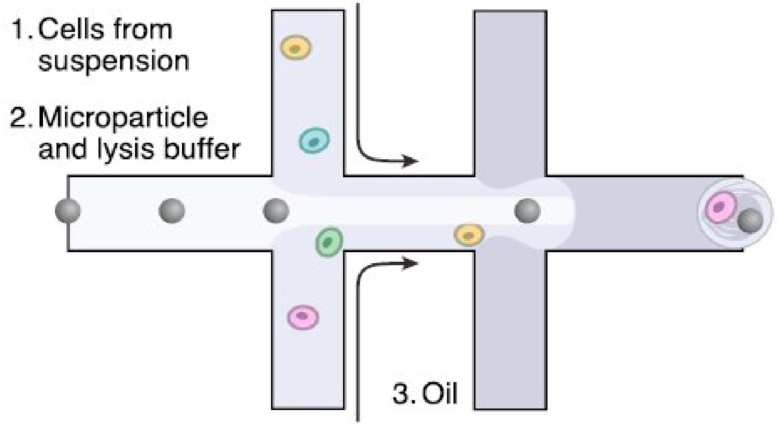
Drop-seq:Macosko *et al.* (2015) Cell

Single cell RNA sequencing: Drop-seq

Barcoded microparticle



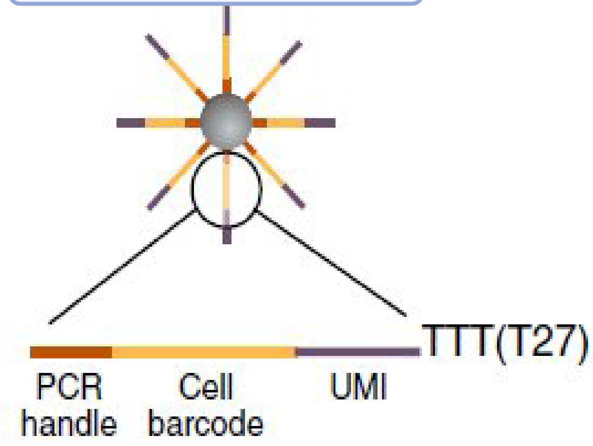
Microfluidic device



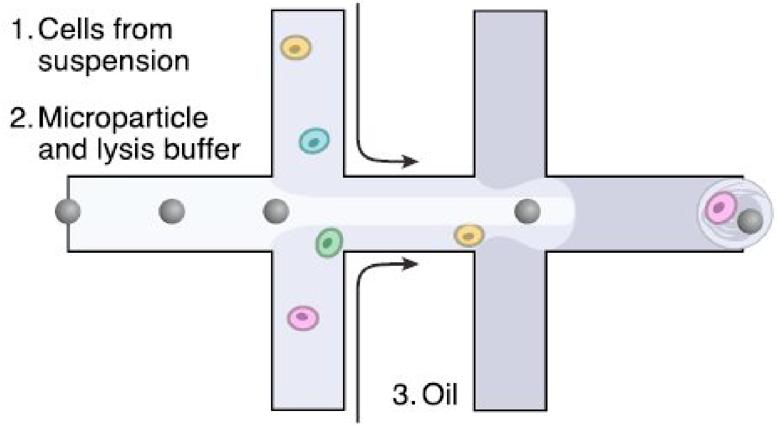
Drop-seq:Macosko *et al.* (2015) Cell

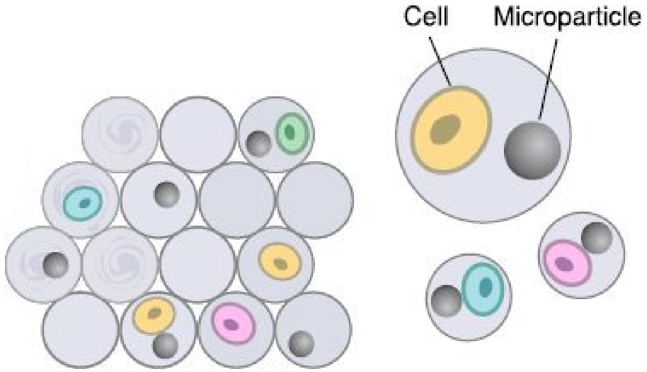
Single cell RNA sequencing: Drop-seq

Barcoded microparticle



Microfluidic device

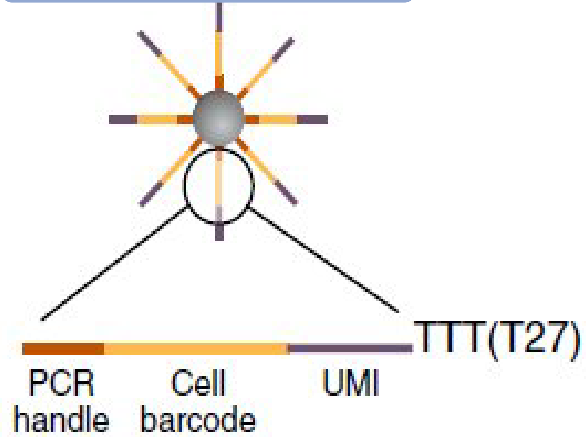




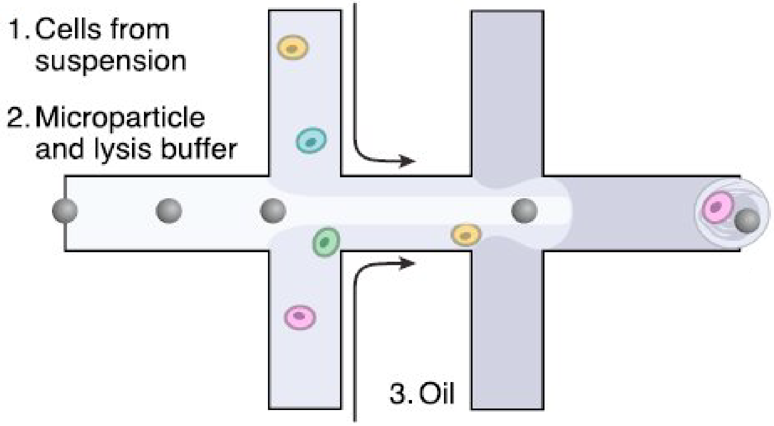
Drop-seq:Macosko *et al.* (2015) Cell

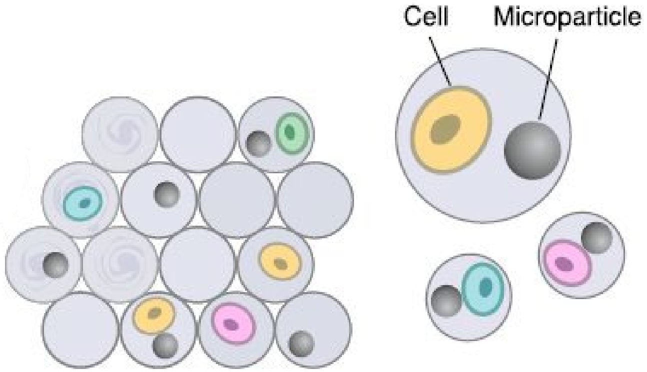
Single cell RNA sequencing: Drop-seq

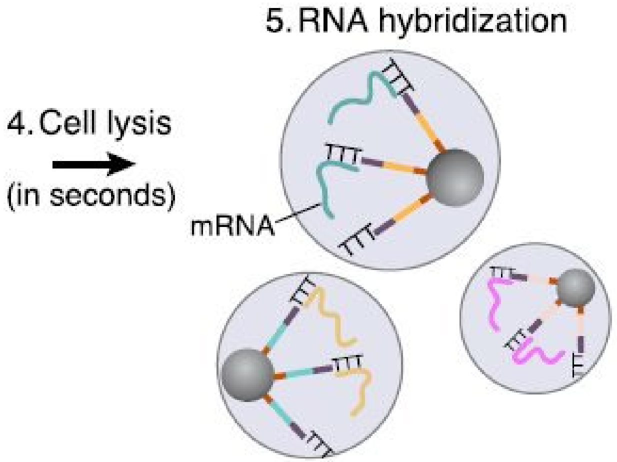
Barcoded microparticle



Microfluidic device



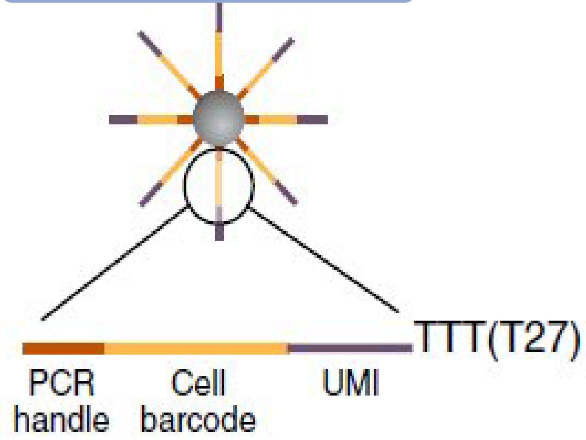




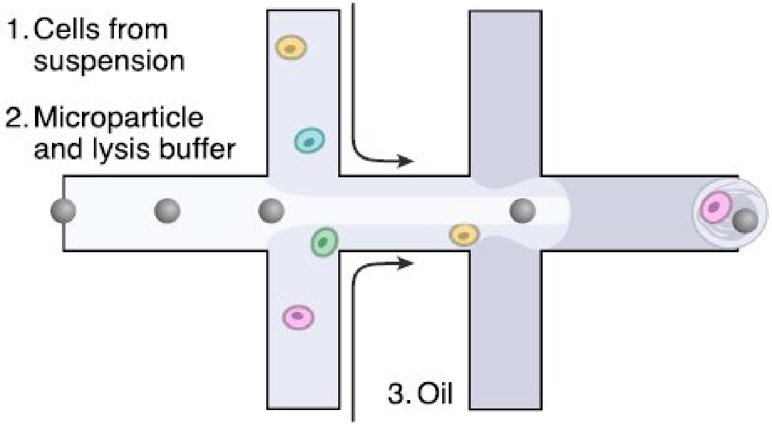
Drop-seq:Macosko *et al.* (2015) Cell

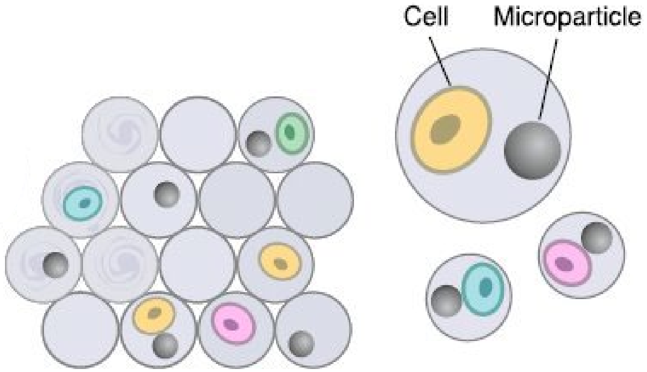
Single cell RNA sequencing: Drop-seq

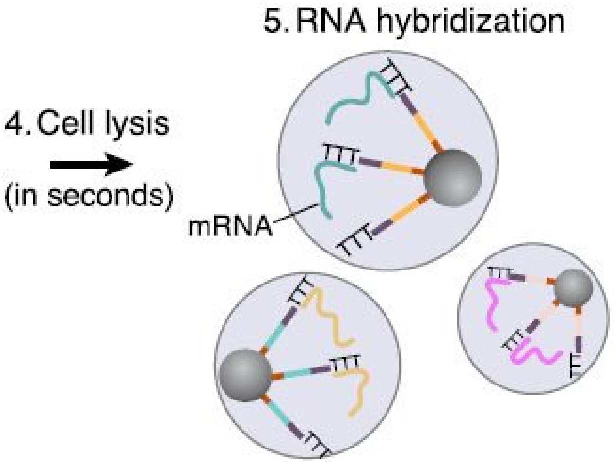
Barcoded microparticle

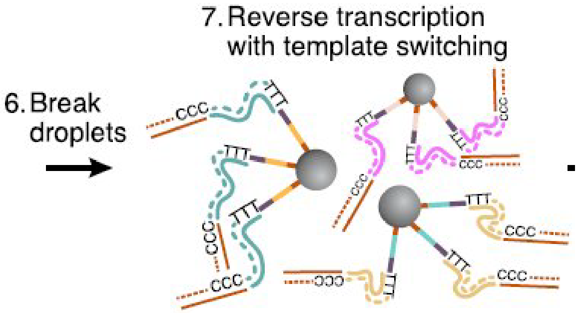


Microfluidic device





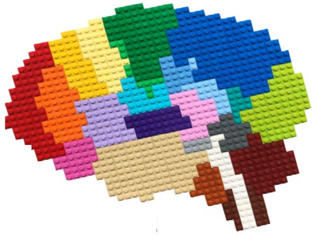




→ 9. Sequencing and analysis

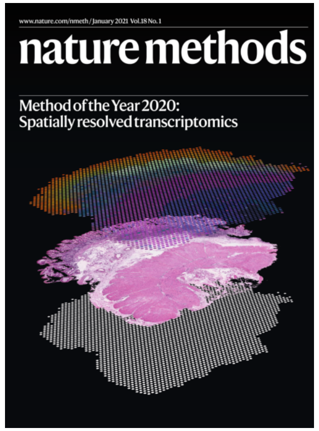
- Each mRNA is mapped to its cell-of-origin and gene-of-origin
- Each cell's pool of mRNA can be analyzed

Drop-seq:Macosko *et al.* (2015) Cell



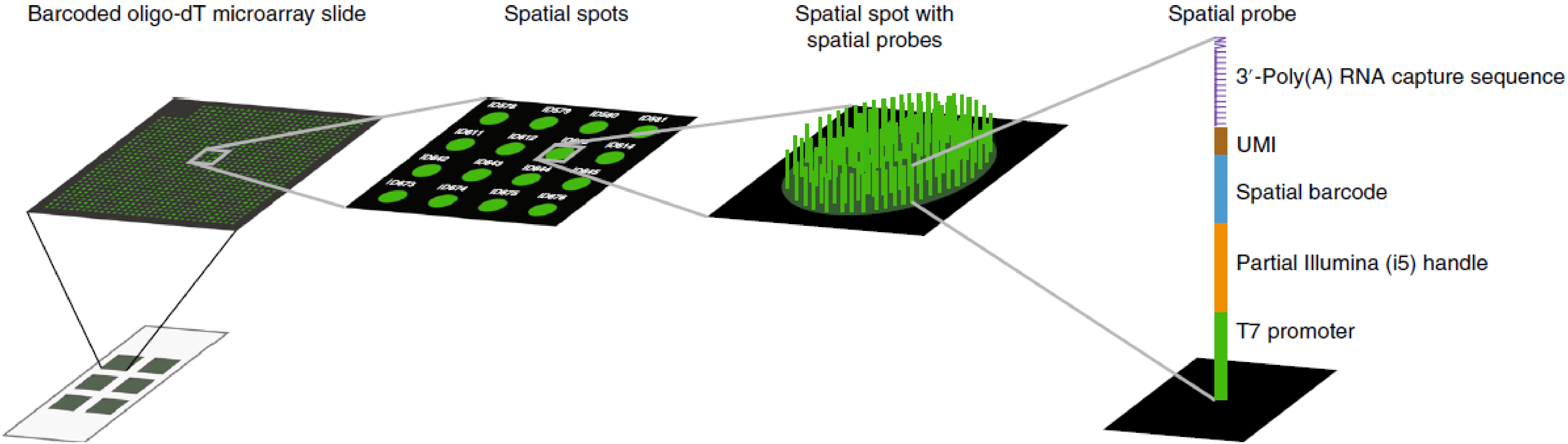
Spatial RNA sequencing

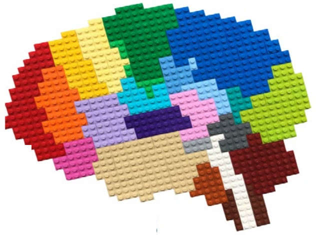
Stahl et al. 2016 Science



nature methods

Method of the Year 2020:
Spatially resolved transcriptomics





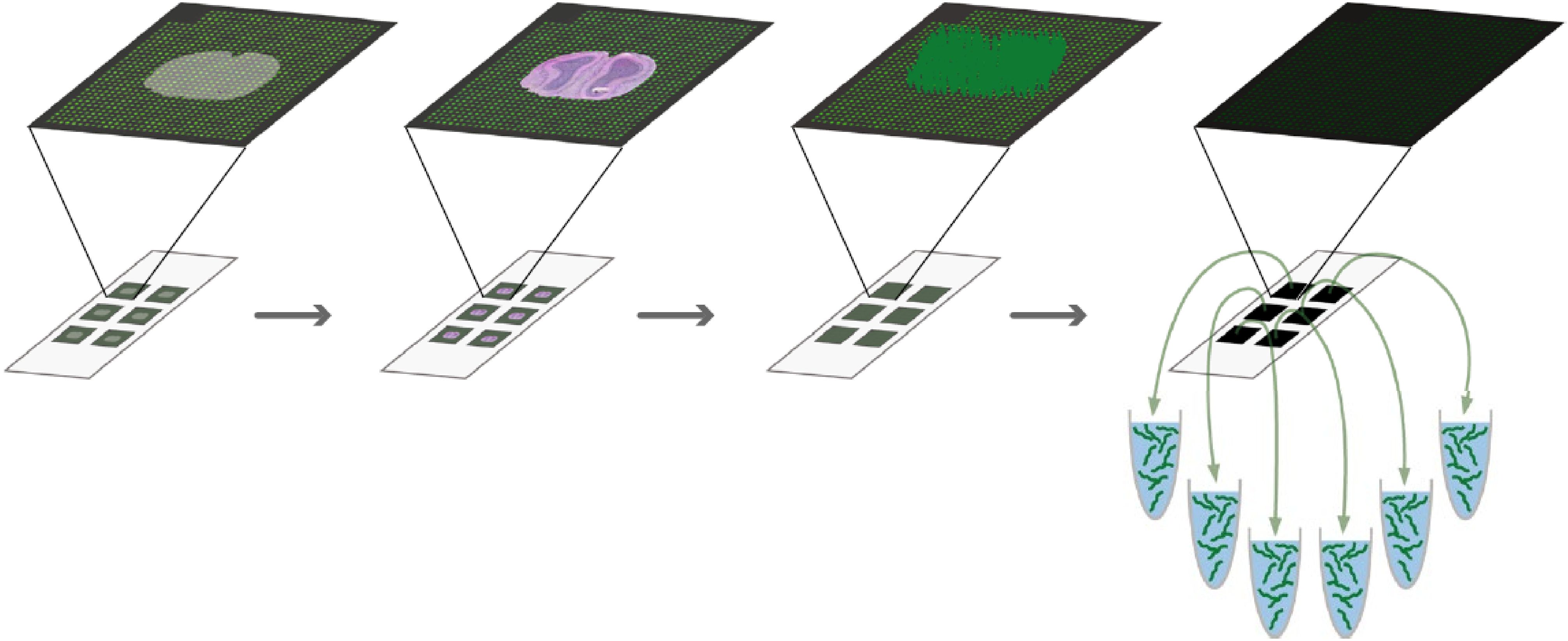
Spatial RNA sequencing

Stahl et al. 2016 Science

www.nature.com/nmeth / January 2021 Vol.18 No.1

nature methods

Method of the Year 2020:
Spatially resolved transcriptomics

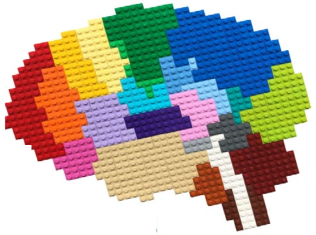


(1) Tissue attachment and fixation

(2) Tissue staining and imaging

(3) Reverse transcription and tissue removal

(4) Release of cDNA and collection in tubes



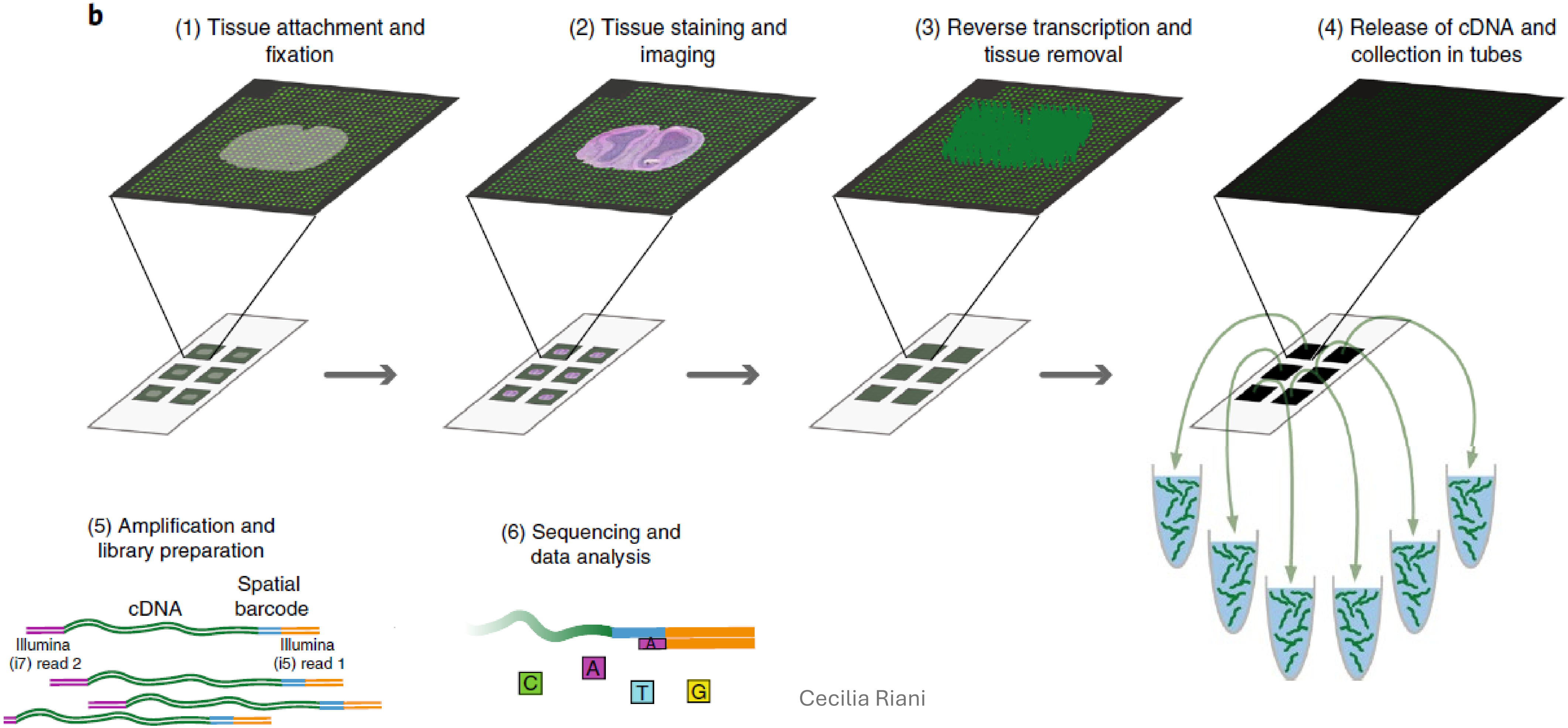
Spatial RNA sequencing

www.nature.com/nmeth / January 2021 Vol.18 No.1

nature methods

Method of the Year 2020:
Spatially resolved transcriptomics

Stahl et al. 2016 Science



(1) Tissue attachment and fixation

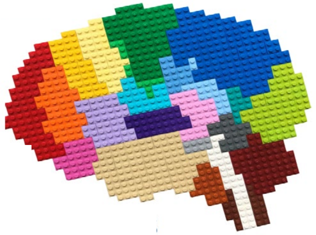
(2) Tissue staining and imaging

(3) Reverse transcription and tissue removal

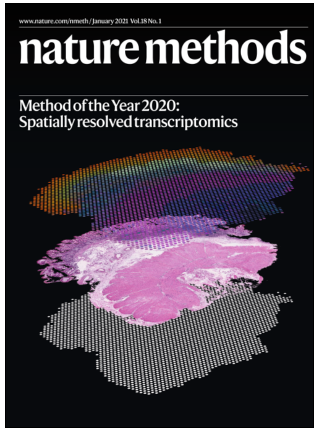
(4) Release of cDNA and collection in tubes

(5) Amplification and library preparation

(6) Sequencing and data analysis



Spatial RNA sequencing

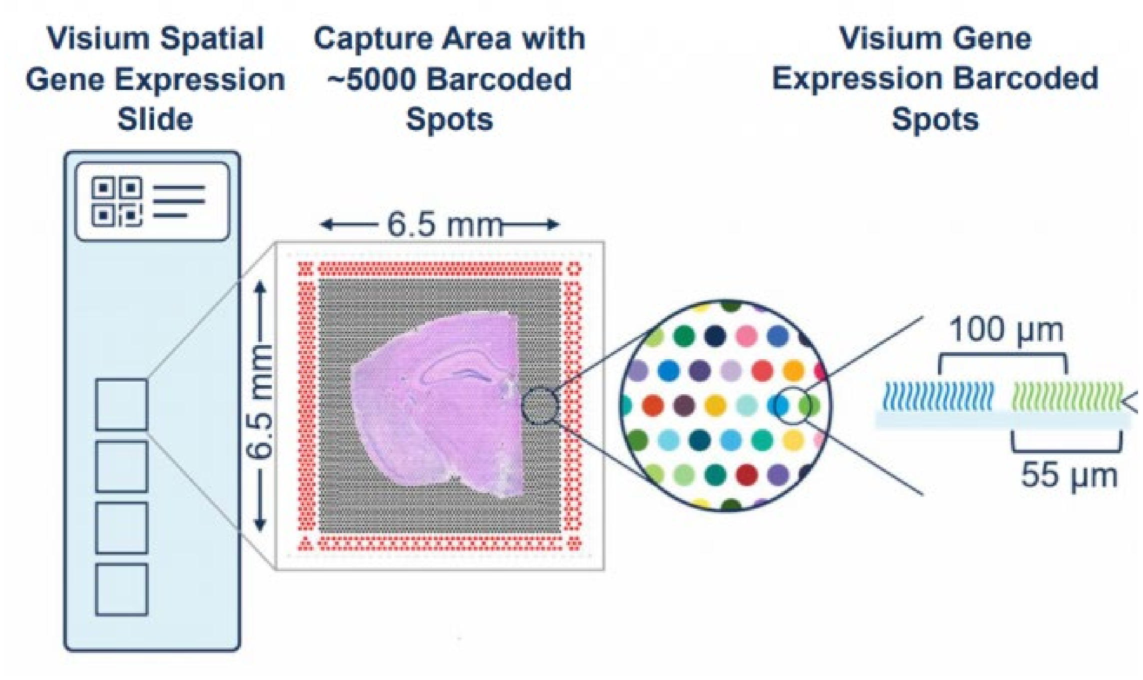


www.nature.com/nmeth / January 2021 Vol.18 No.1

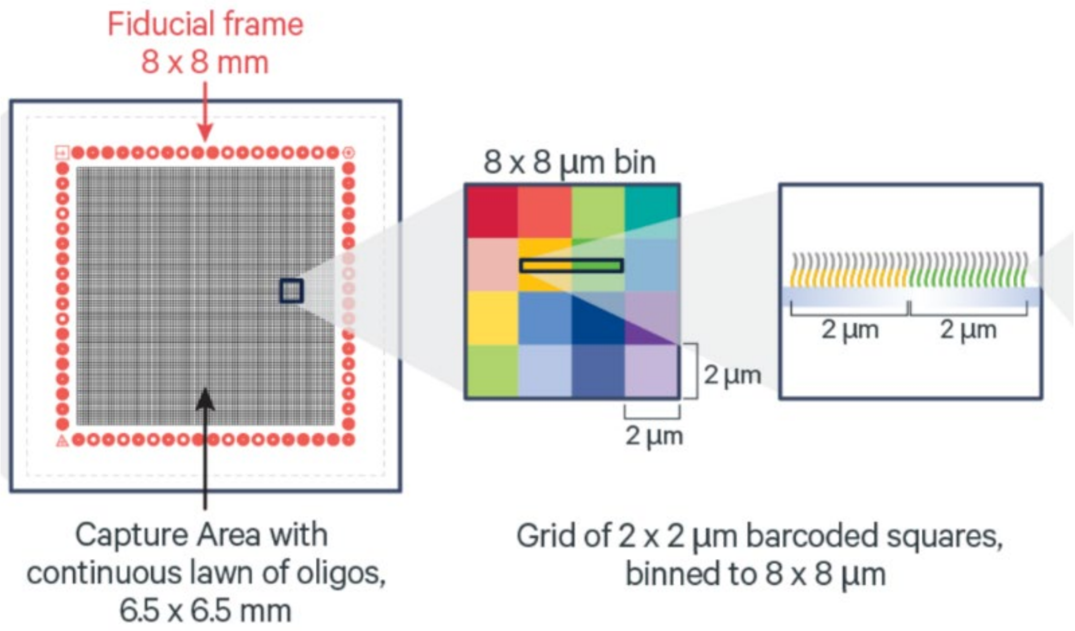
nature methods

Method of the Year 2020:
Spatially resolved transcriptomics

VISIUM



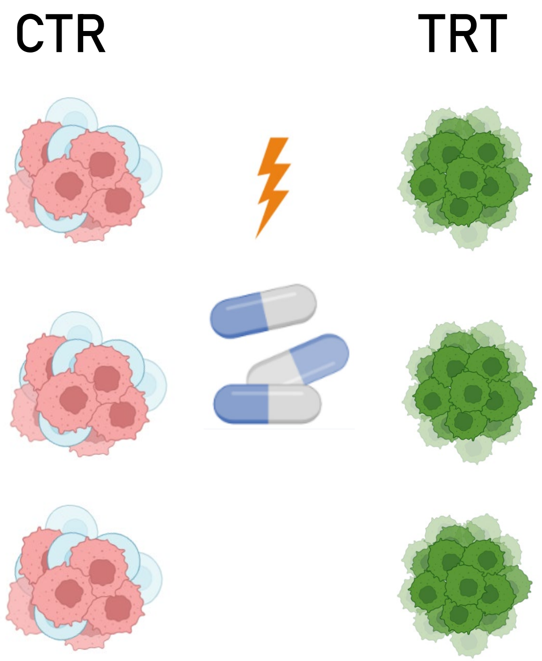
VISIUM HD



RNA seq data analysis:
Differential gene expression analysis

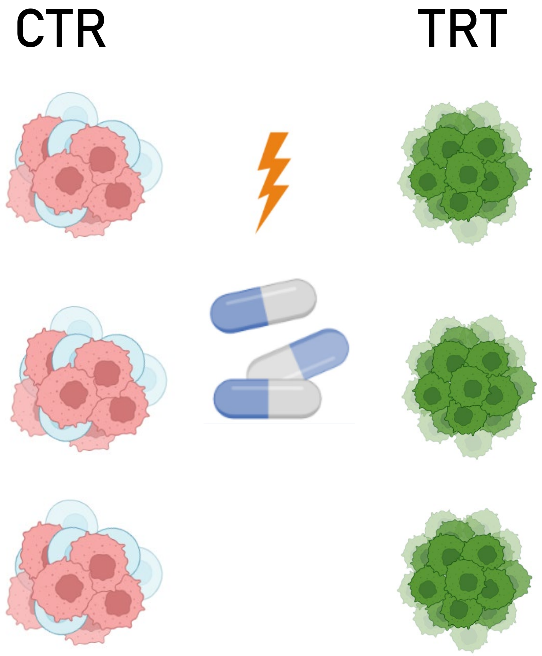
Differential gene expression (DGE) analysis

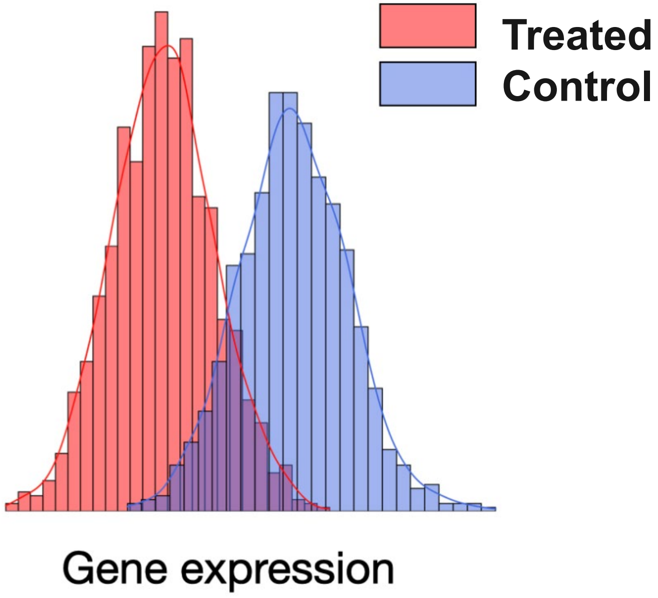
☐ *Is there a statistically significant change in gene expression between two groups?*



Differential gene expression (DGE) analysis

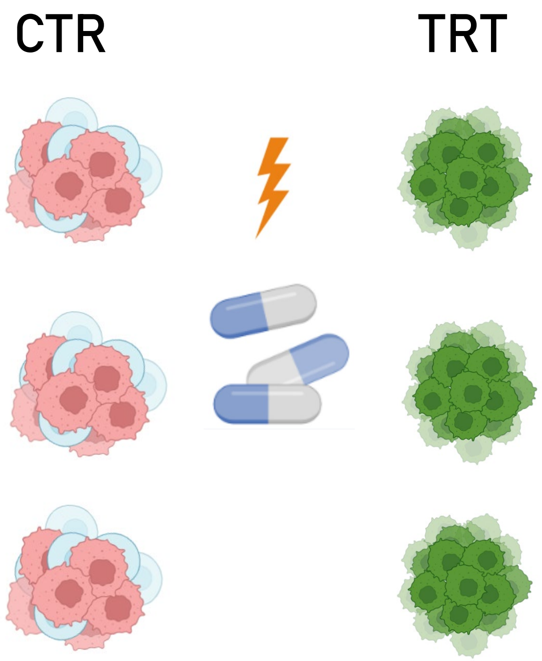
☐ *Is there a statistically significant change in gene expression between two groups?*

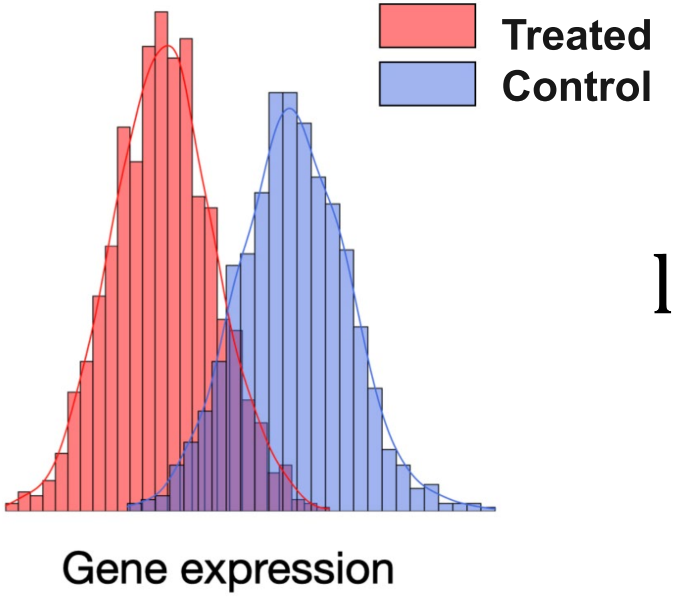




Differential gene expression (DGE) analysis

❑ *Is there a statistically significant change in gene expression between two groups?*





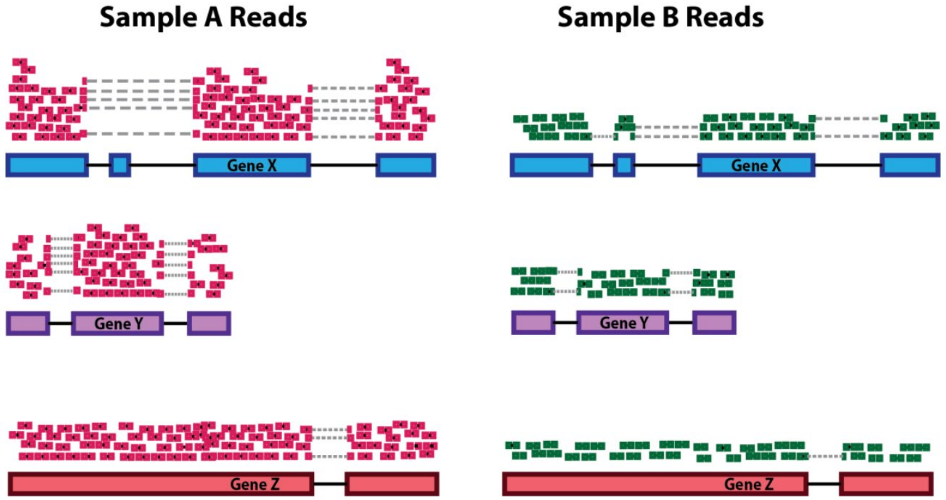
$$\log_2 FC = \log_2 \frac{\overline{\mu_T}}{\overline{\mu_C}}$$

Differential gene expression (DGE) analysis

Modelling counts

Observed counts are non negative.
They depend on

- Natural Abundance of mRNA molecules
- Gene length
- Sequencing depth

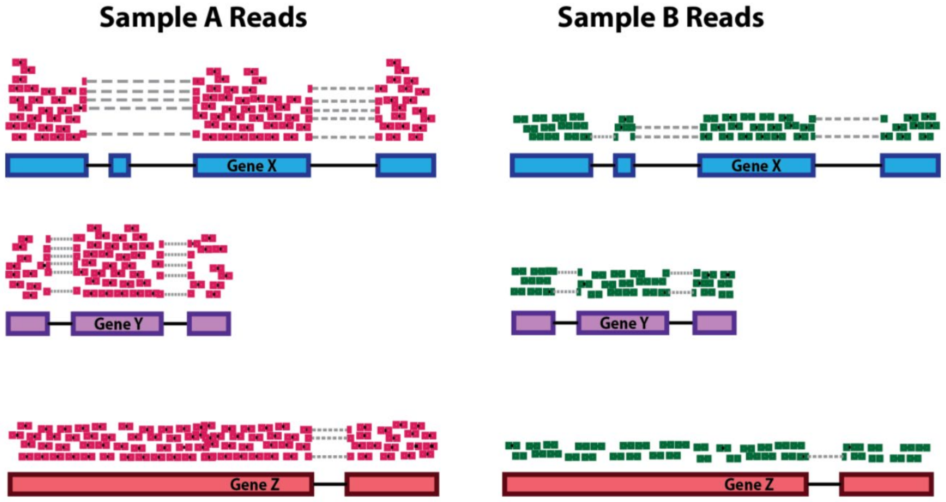


Differential gene expression (DGE) analysis

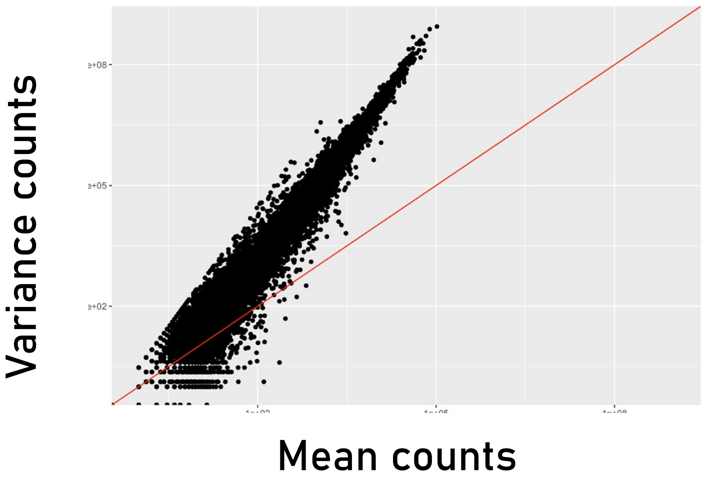
Modelling counts

Observed counts are non negative.
They depend on

- Natural Abundance of mRNA molecules
- Gene length
- Sequencing depth



VARIANCE depends on the mean count
(biological or technical)

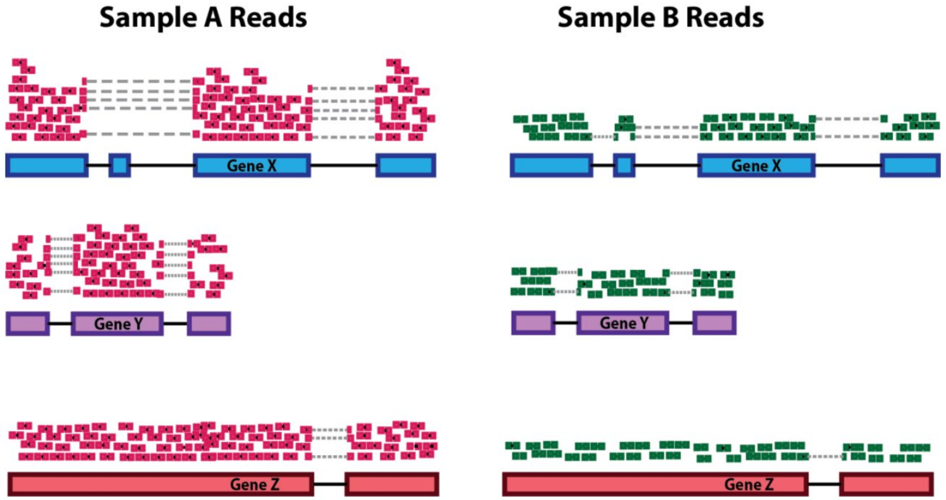


Differential gene expression (DGE) analysis

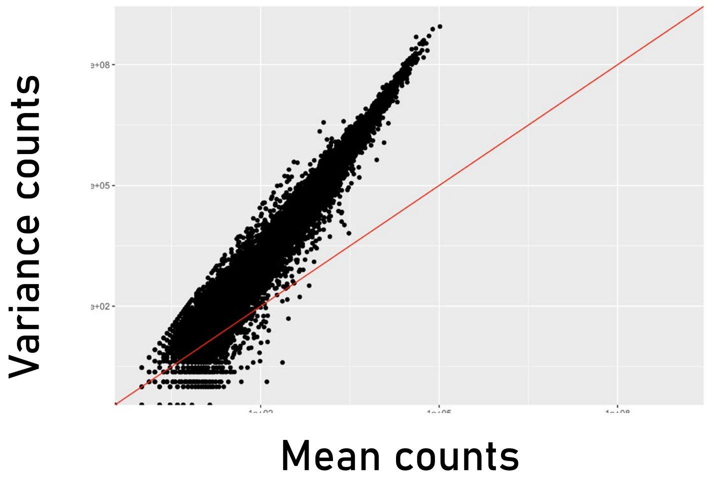
Modelling counts

Observed counts are non negative.
They depend on

- Natural Abundance of mRNA molecules
- Gene length
- Sequencing depth



VARIANCE depends on the mean count
(biological or technical)



Challenges for RNA-seq data analysis

- ☐ Choice of statistical distribution
- ☐ Normalization between sample
- ☐ Few sample -> difficult to estimate parameters

Deseq2 algorithm for DGE analysis

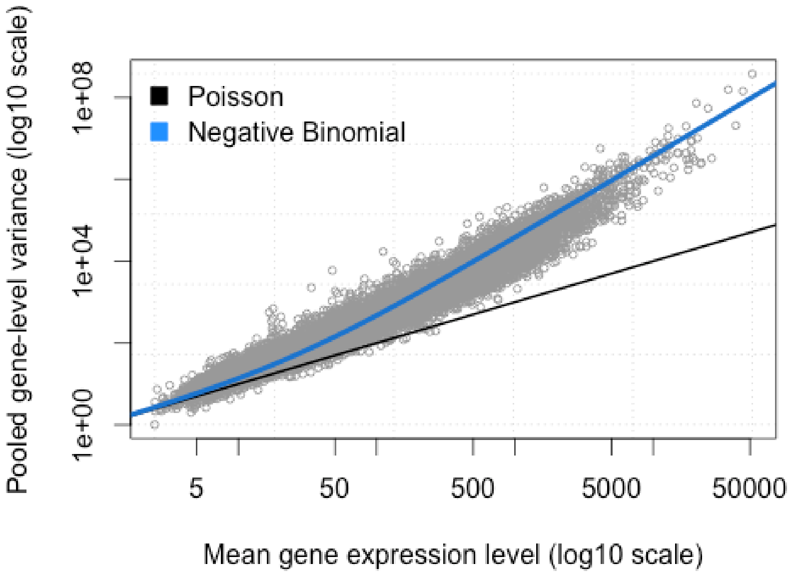
DESeq *Anders et al. Genome Biology 2010*
Deseq2 *Love et al. Genome Biology 2014*
edgeR. MD Robinson Bioinformatics 2009
Limma+voom *Law CW et al. Genome Biology 2014*

Deseq2 algorithm for DGE analysis

Modelling counts- the Negative Binomial Distribution (overdispersed Poisson distribution)

$$\Pr(X = k) = \binom{k+r-1}{k}(1-p)^k p^r$$

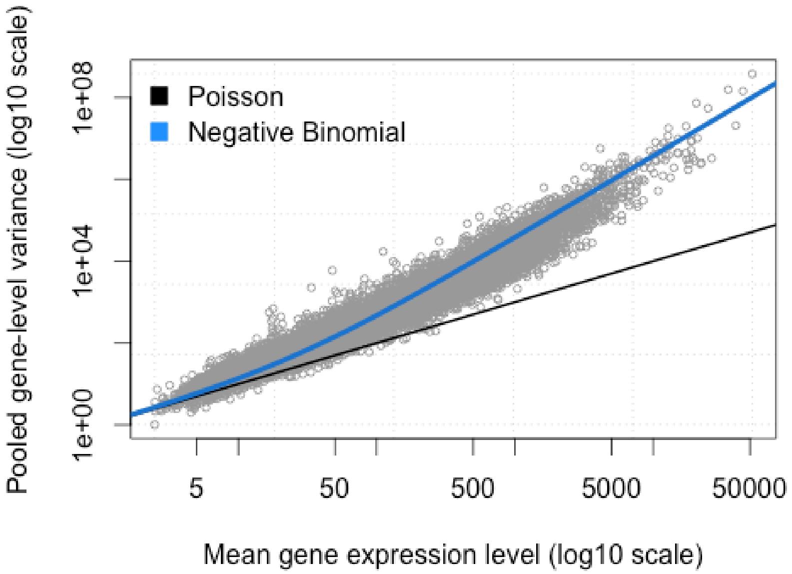
K number of failures
r is the number of successes
p is the probability of success



DESeq *Anders et al. Genome Biology 2010*
Deseq2 *Love et al. Genome Biology 2014*
edgeR. MD Robinson Bioinformatics 2009
Limma+voom *Law CW et al. Genome Biology 2014*

Deseq2 algorithm for DGE analysis

Modelling counts- the Negative Binomial Distribution (overdispersed Poisson distribution)



- *Set up a null hypothesis for each gene. No differential expression across the two sample groups (LFC == 0)*
- *Modelling raw count for each gene* $\sim NB\left(\mu_{ij} = s_{ij}q_{ij}, \alpha = \mu_{ij} + s_j^2 v_{i,\rho(j)}\right)$

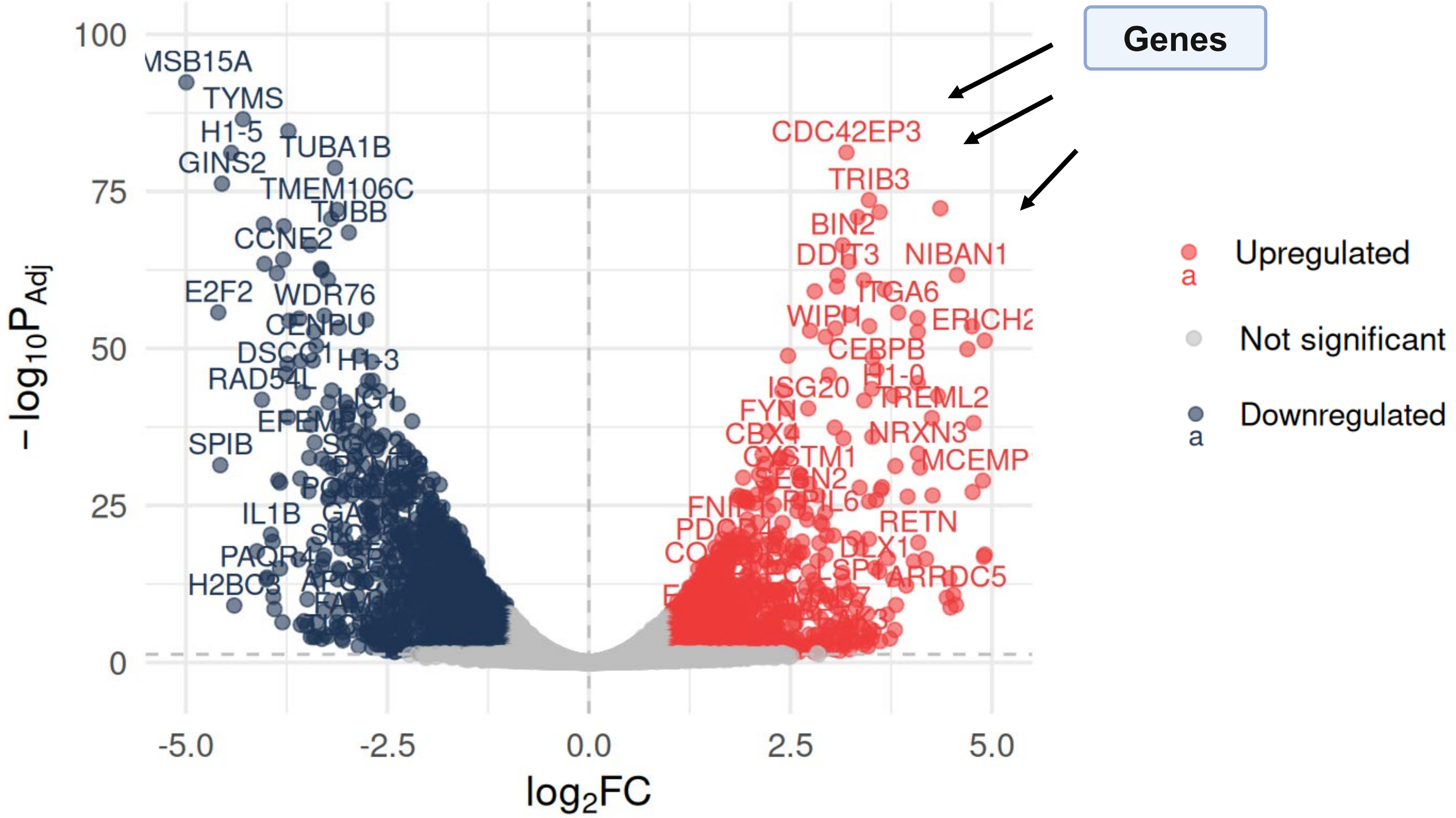
1. Estimate size factor $\hat{s_j} = median \frac{c_{ij}}{\left(\prod_{v=1}^{m} c_{iv}\right)^{1/m}}$ → **Pseudo-reference sample**
2. Normalization through size factor
3. Estimate gene-wise dispersion (Maximum likelihood)

- *Testing for differential expression (Wald test)*
- *Multiple correction*

DESeq *Anders et al. Genome Biology 2010*
Deseq2 *Love et al. Genome Biology 2014*
edgeR. MD Robinson Bioinformatics 2009
Limma+voom *Law CW et al. Genome Biology 2014*

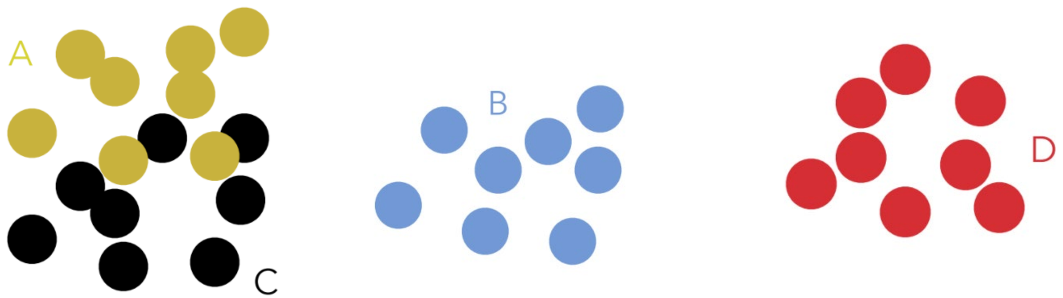
Deseq2 algorithm for DGE analysis

Volcano plot

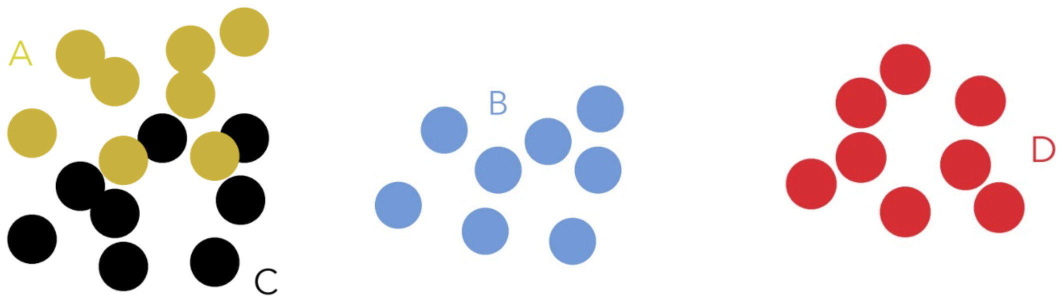


scRNAseq data analysis

1. Marker gene detection
- Which genes are differentially expressed between cell types A and B?
- Which genes are specifically expressed in cell type A?



scRNAseq data analysis



1. Marker gene detection

- Which genes are differentially expressed between cell types A and B?
- Which genes are specifically expressed in cell type A?

2. Differential abundance analysis

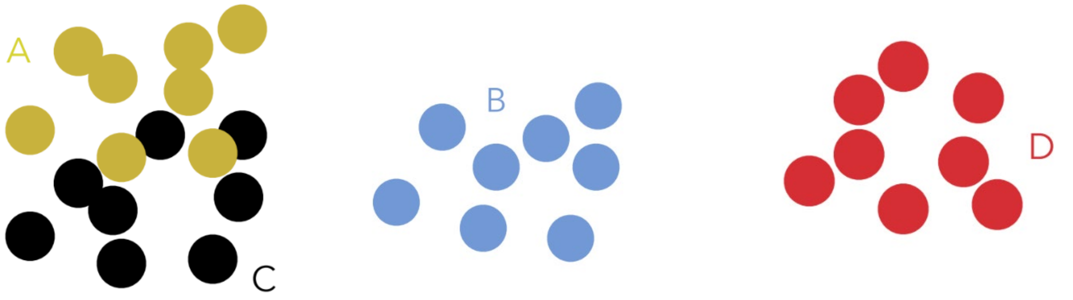
- Are some cell types more/less abundant in one condition compared to another

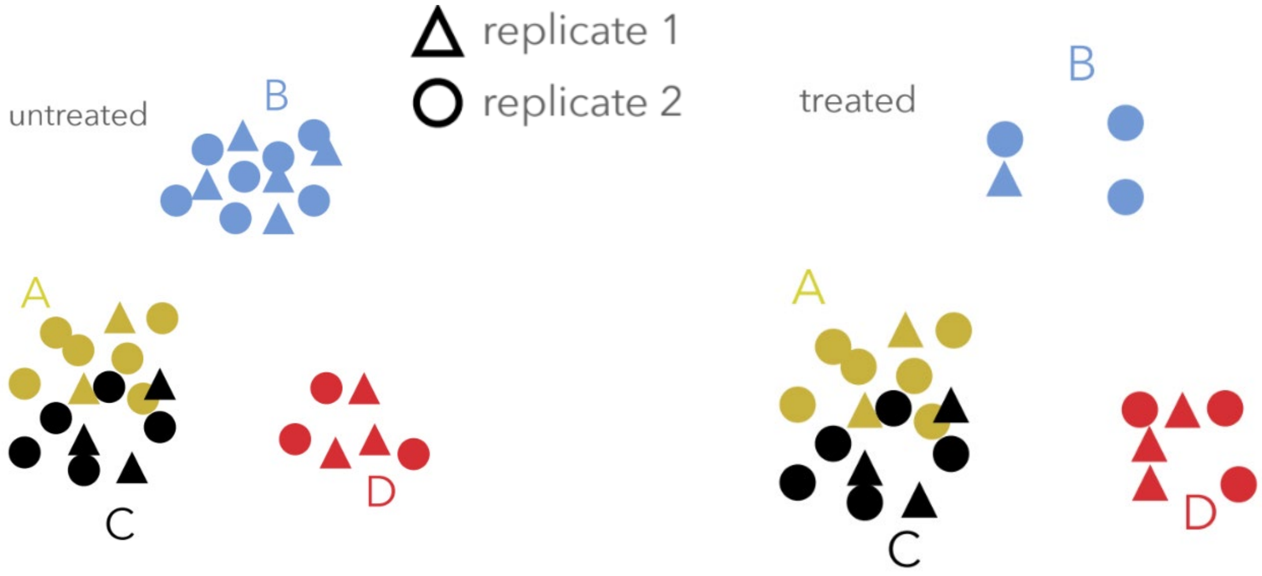
scRNAseq data analysis

1. **Marker gene detection**
   - Which genes are differentially expressed between cell types A and B?
   - Which genes are specifically expressed in cell type A?

2. **Differential abundance analysis**
   - Are some cell types more/less abundant in one condition compared to another

3. **Differential state analysis**
   - Are any genes differentially expressed between conditions within a given cell type?





scRNAseq data analysis workflow

$\sim$15000 genes

GENES	Cell 1	Cell 2	...	Cell N
Gene A	0	1		5
Gene B	10	5		0
Gene C	2	0		0
……	...			

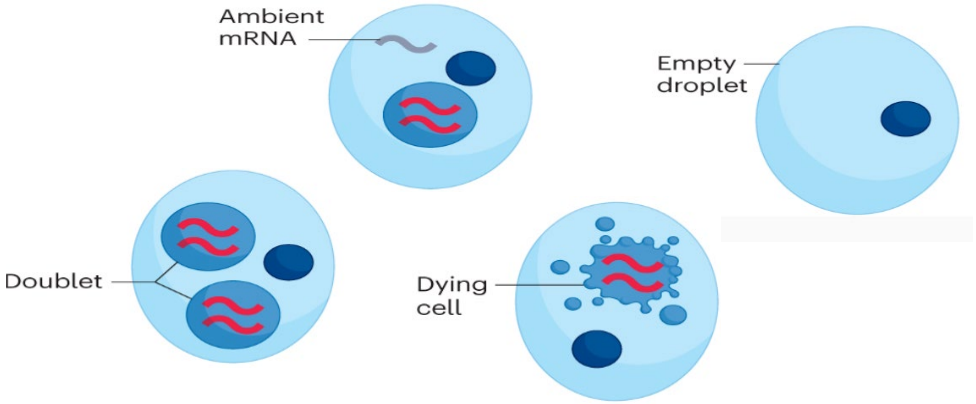
$\sim$20000 cells

scRNAseq data analysis workflow

~15000 genes

GENES	Cell 1	Cell 2	...	Cell N
Gene A	0	1		5
Gene B	10	5		0
Gene C	2	0		0
……	...			

~20000 cells



Quality Control

Take care when filtering.

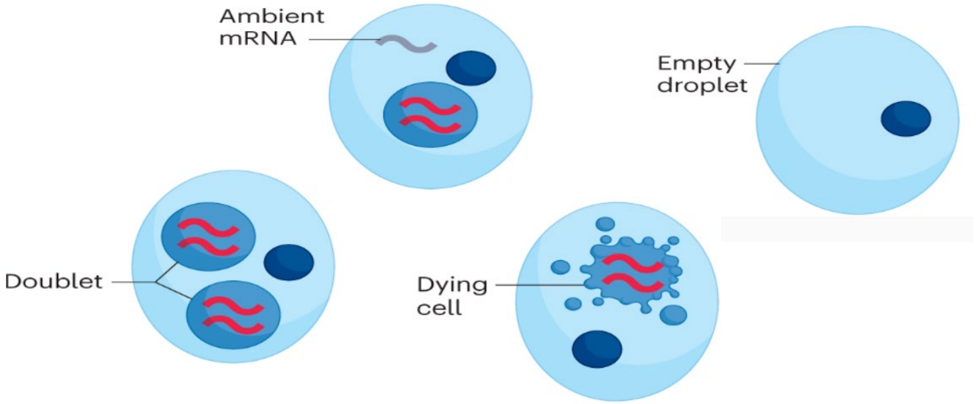
- ❑ Heavy filtering may result in the loss of rare cell populations,
- ❑ minimal filtering, may impact our ability to annotate cell populations downstream.

scRNAseq data analysis workflow

~15000 genes

GENES	Cell 1	Cell 2	...	Cell N
Gene A	0	1		5
Gene B	10	5		0
Gene C	2	0		0
……	...			

~20000 cells

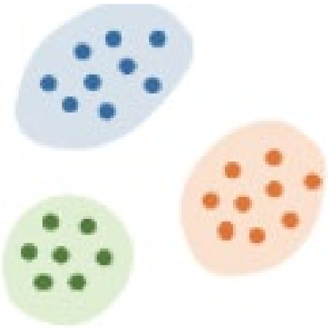


Quality Control

Dimensionality reduction & Clustering

Take care when filtering.

- ❑ Heavy filtering may result in the loss of rare cell populations,
- ❑ minimal filtering, may impact our ability to annotate cell populations downstream.

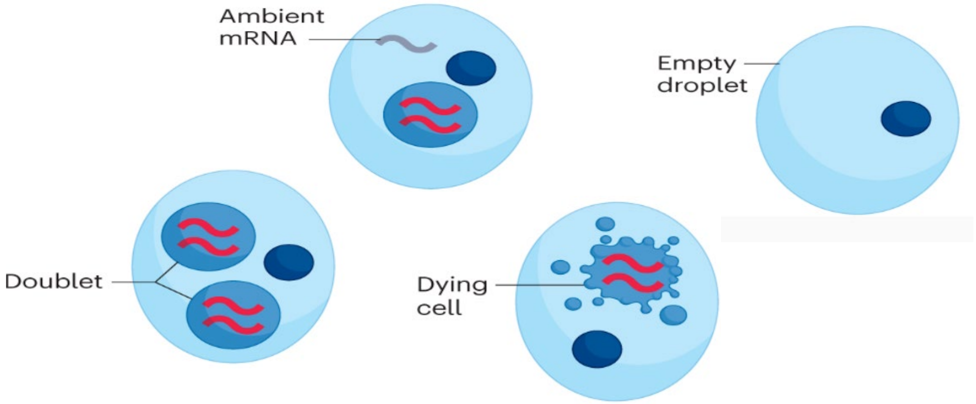


scRNAseq data analysis workflow

~15000 genes

GENES	Cell 1	Cell 2	...	Cell N
Gene A	0	1		5
Gene B	10	5		0
Gene C	2	0		0
......	...			

~20000 cells

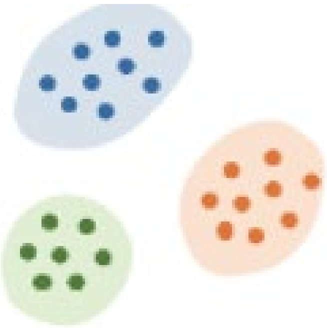


Quality Control

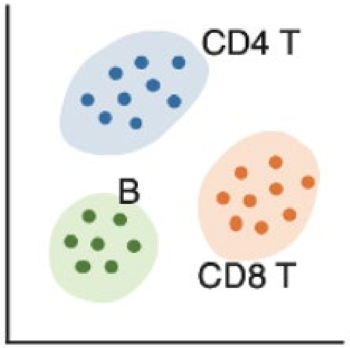
Take care when filtering.

- ❑ Heavy filtering may result in the loss of rare cell populations,
- ❑ minimal filtering, may impact our ability to annotate cell populations downstream.

Dimensionality reduction & Clustering



Cell Type Annotation

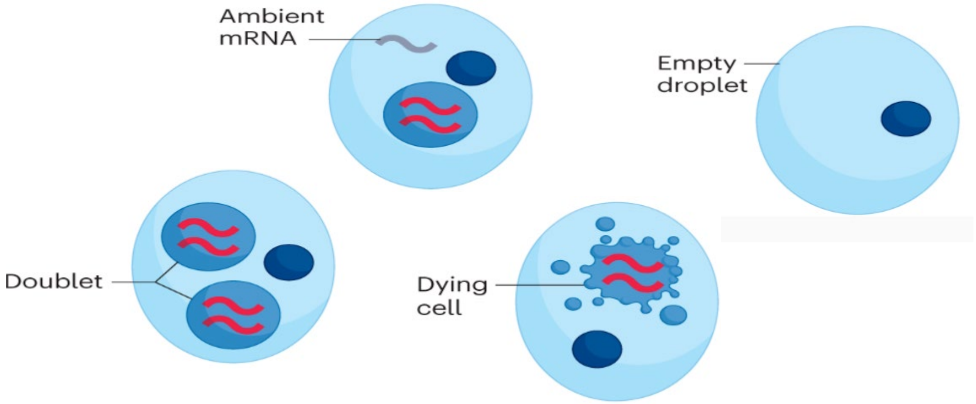


scRNAseq data analysis workflow

~15000 genes

GENES	Cell 1	Cell 2	...	Cell N
Gene A	0	1		5
Gene B	10	5		0
Gene C	2	0		0
......	...			

~20000 cells

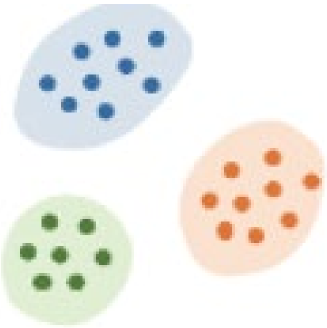


Quality Control

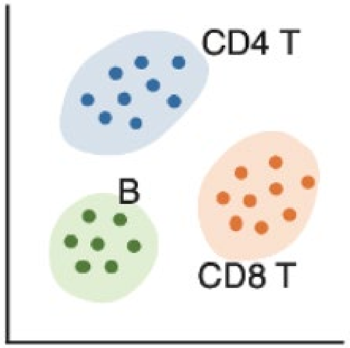
Take care when filtering.

- ❑ Heavy filtering may result in the loss of rare cell populations,
- ❑ minimal filtering, may impact our ability to annotate cell populations downstream.

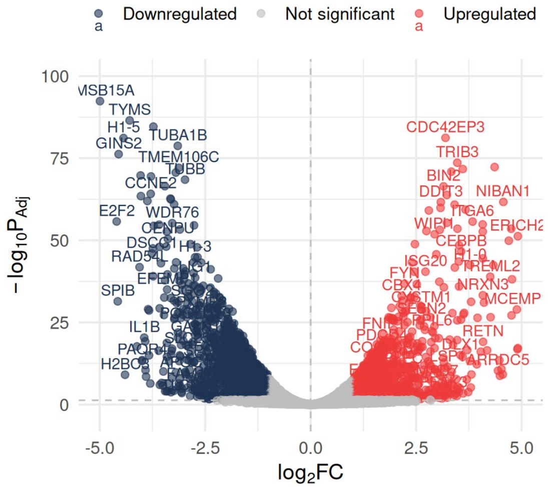
Dimensionality reduction & Clustering



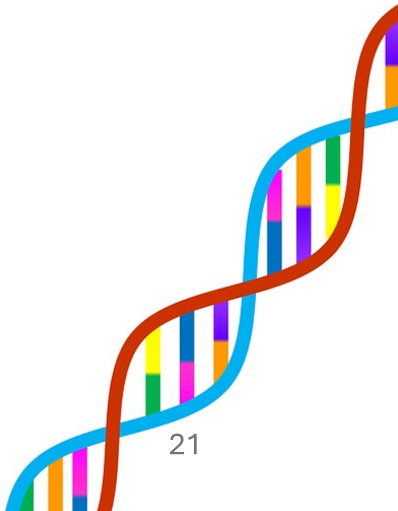
Cell Type Annotation



DGE



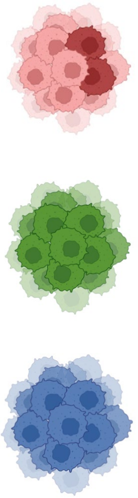
Results



scRNAseq analysis of L-asparaginase Treatment Response of leukemia cell line RS4;11

RS4;11

CTR



Rep 1

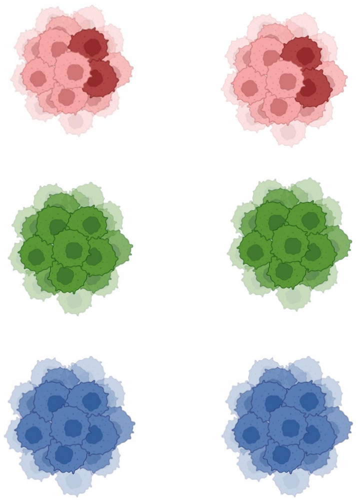
Rep 2

Rep 3

scRNAseq analysis of L-asparaginase Treatment Response of leukemia cell line RS4;11

RS4;11

L-asparaginase
0.001 U/ml
→ Blocks RS4;11 cells in G1 cell cycle phase



CTR TRT

Rep 1

Rep 2

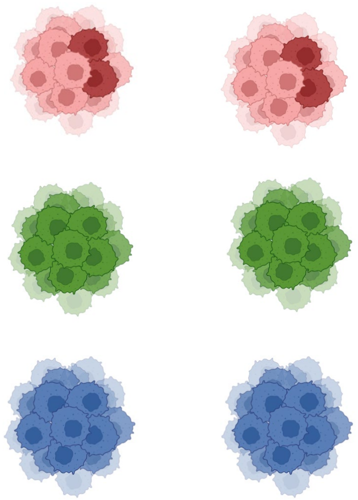
Rep 3

scRNAseq analysis of L-asparaginase Treatment Response of leukemia cell line RS4;11

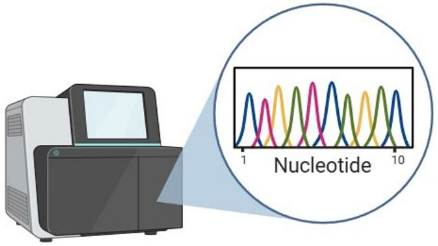
RS4;11

L-asparaginase
0.001 U/ml

Blocks RS4;11 cells in G1 cell cycle phase



Single cell mRNA Sequencing
72 h after treatment

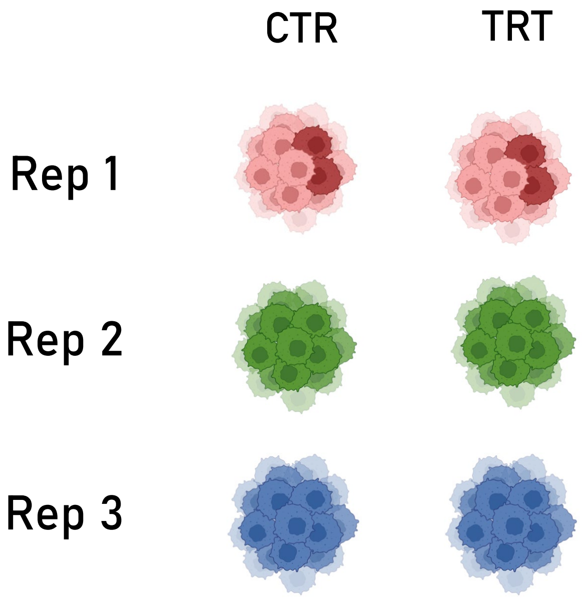


scRNAseq analysis of L-asparaginase Treatment Response of leukemia cell line RS4;11

RS4;11

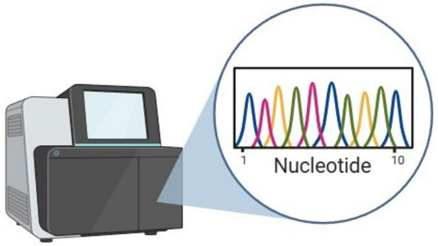
L-asparaginase 0.001 U/ml →

Blocks RS4;11 cells in G1 cell cycle phase

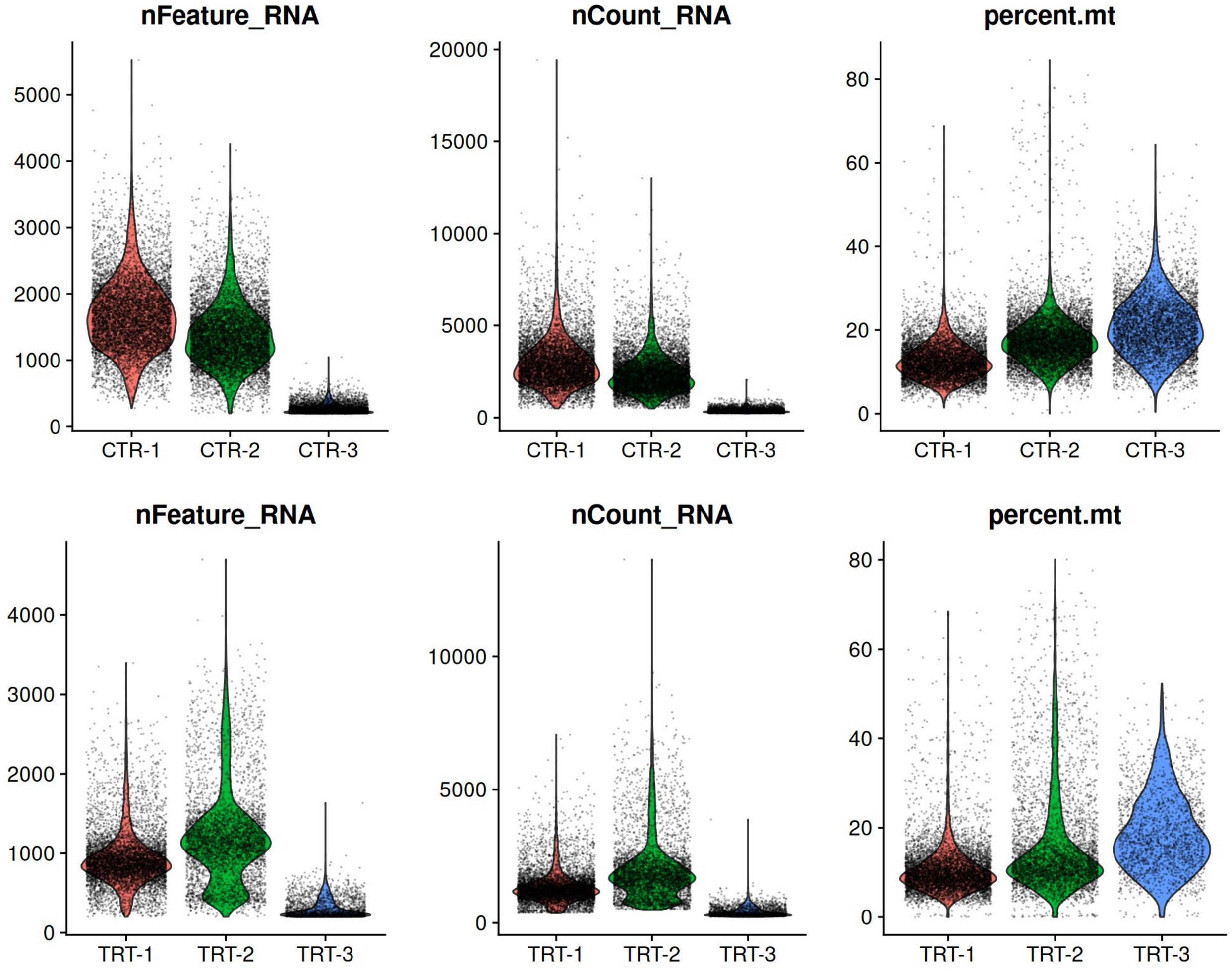


Single cell mRNA Sequencing

72 h after treatment



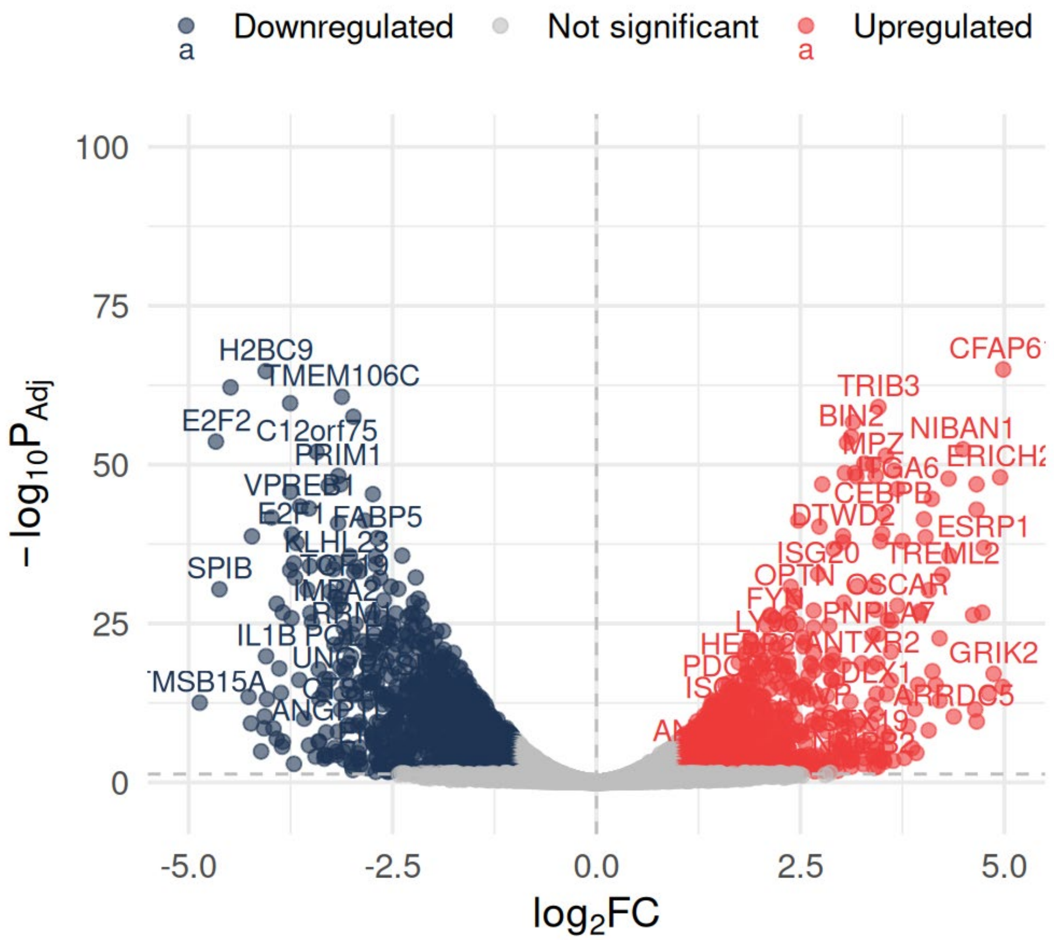
Quality control



scRNAseq analysis of L-asparaginase Treatment Response of leukemia cell line RS4;11

Pseudo-bulking and DGE

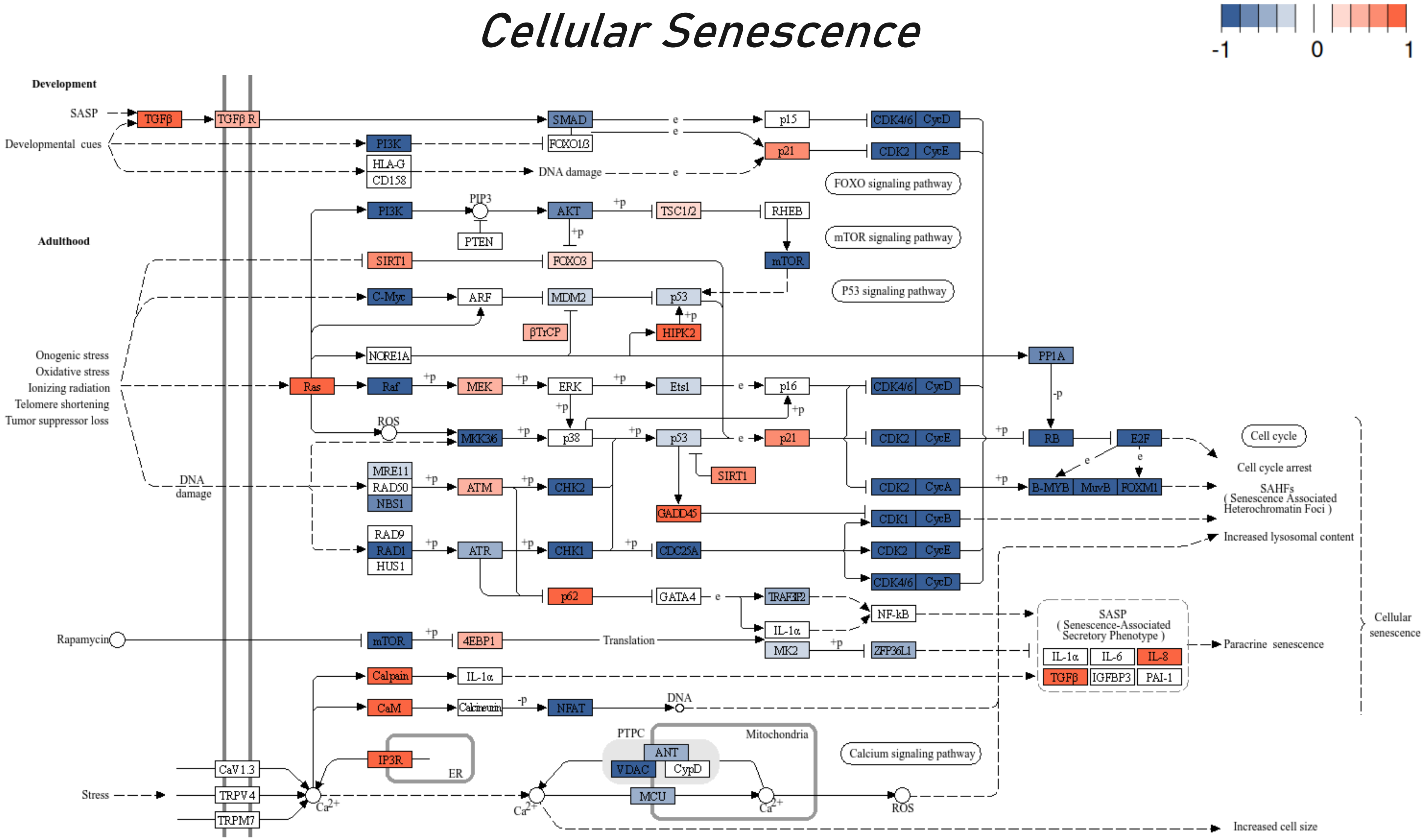
(deseq2)



scRNAseq analysis of L-asparaginase Treatment Response of leukemia cell line RS4;11

Cellular Senescence

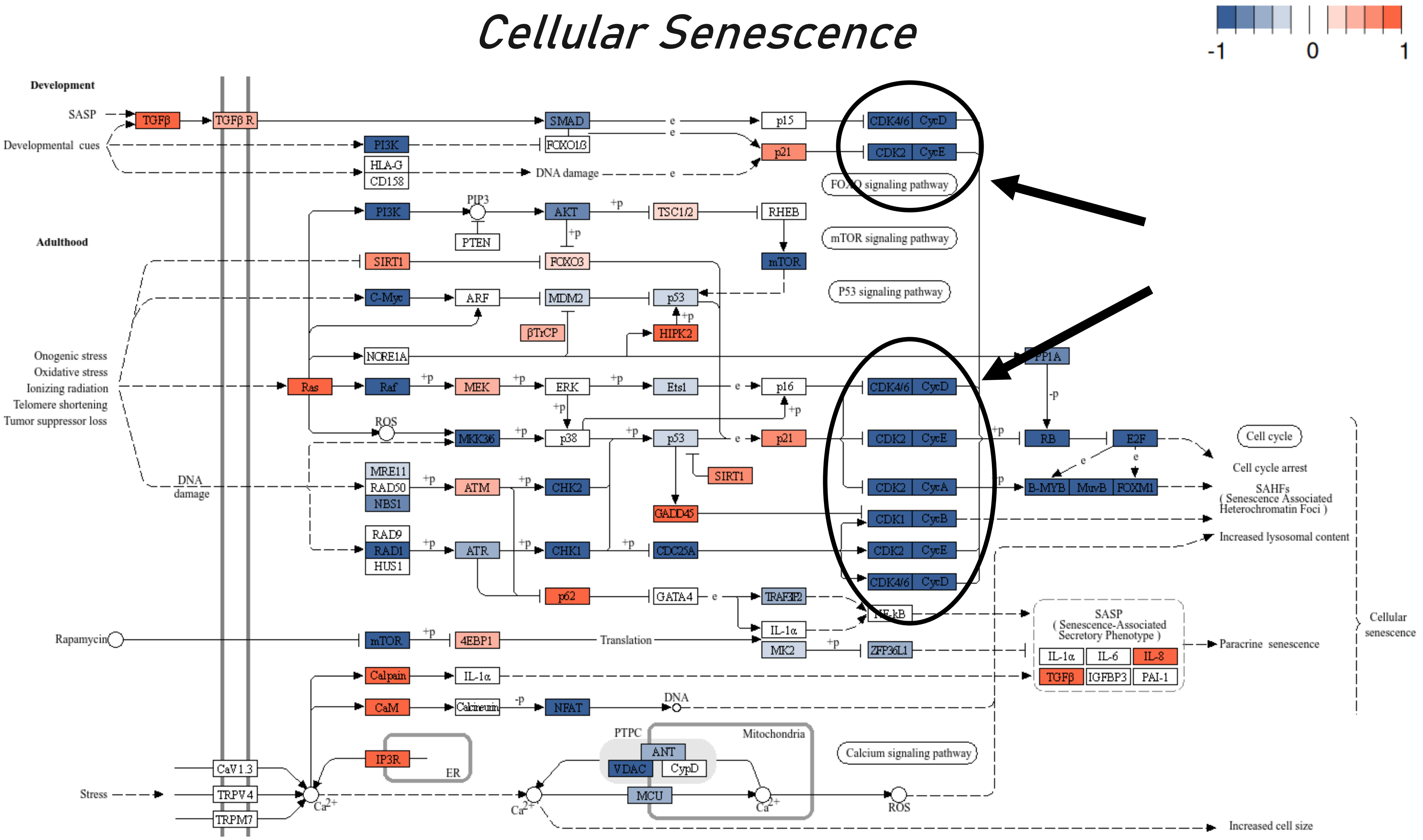
KEGG
Kyoto Encyclopedia of Genes and Genome



scRNAseq analysis of L-asparaginase Treatment Response of leukemia cell line RS4;11

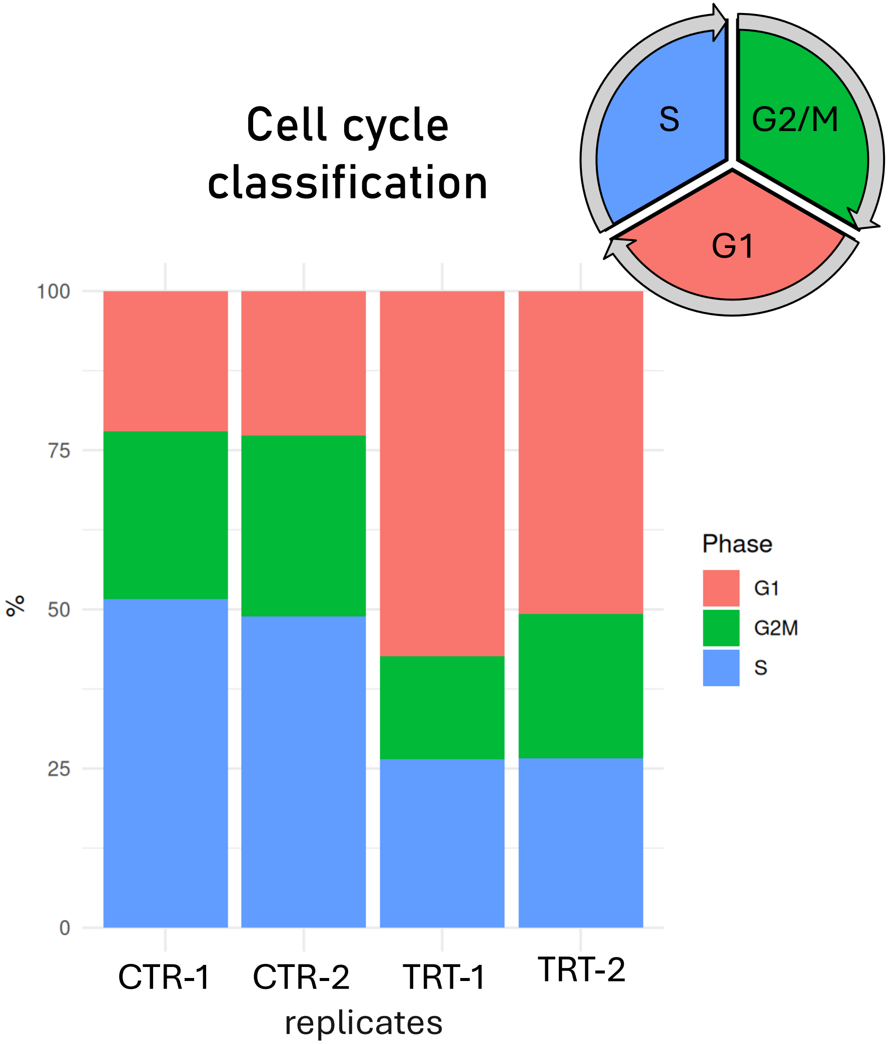
Cellular Senescence

KEGG
Kyoto Encyclopedia of Genes and Genome



scRNAseq analysis of L-asparaginase Treatment Response of leukemia cell line RS4;11

Cell cycle classification



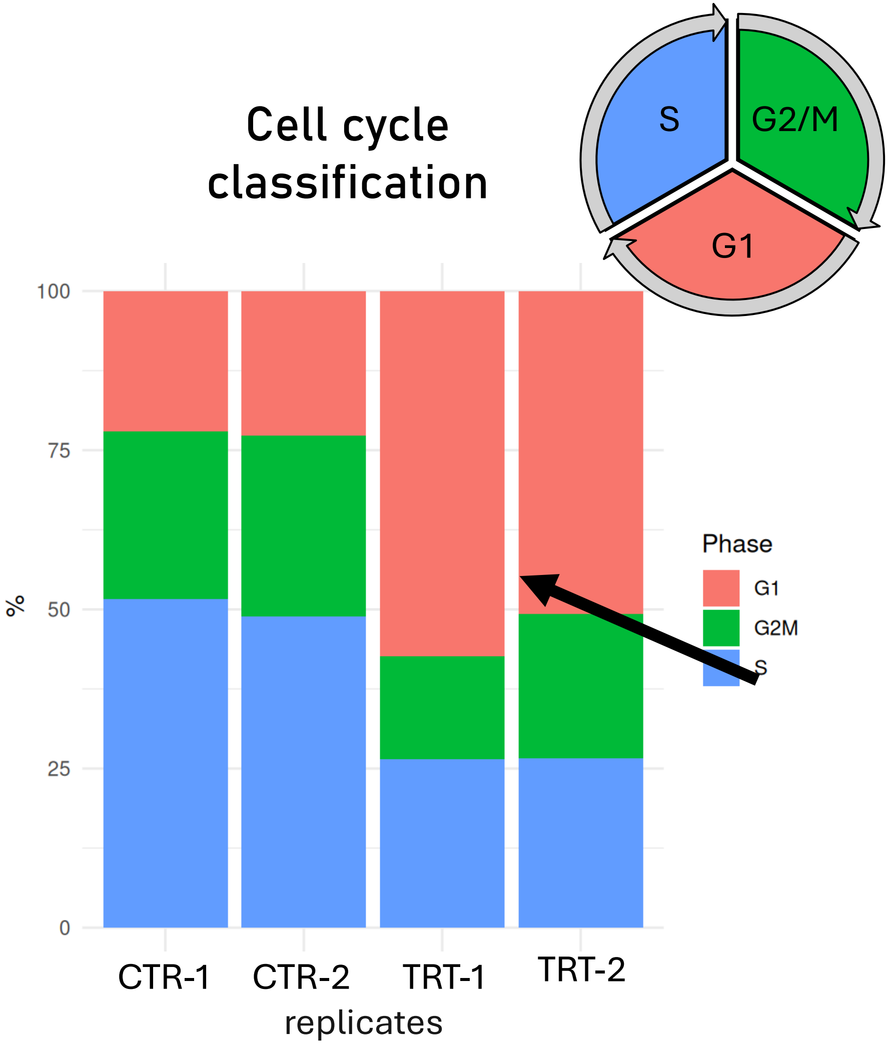
G1 - Growth

S - DNA synthesis

G2 - Growth and preparation for mitosis

M - Mitosis (cell division)

scRNAseq analysis of L-asparaginase Treatment Response of leukemia cell line RS4;11



Cell cycle classification

G1 - Growth

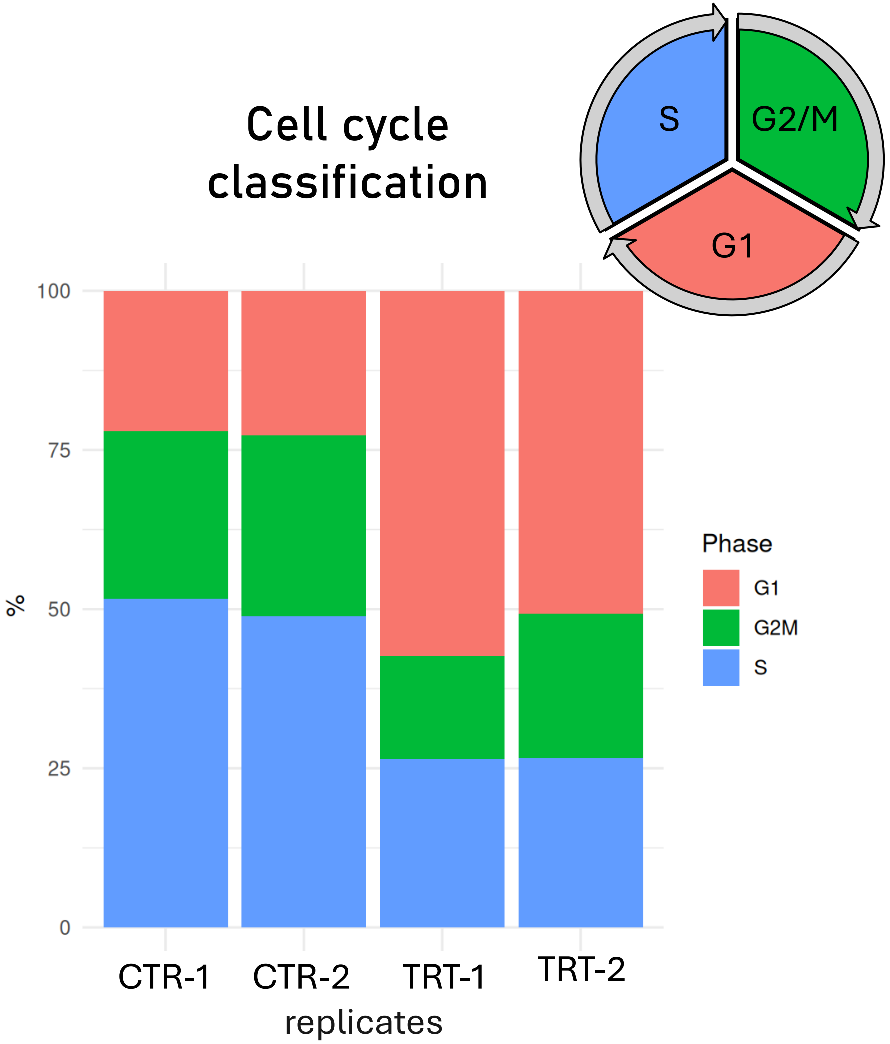
S - DNA synthesis

G2 - Growth and preparation for mitosis

M - Mitosis (cell division)

scRNAseq analysis of L-asparaginase Treatment Response of leukemia cell line RS4;11

Cell cycle classification



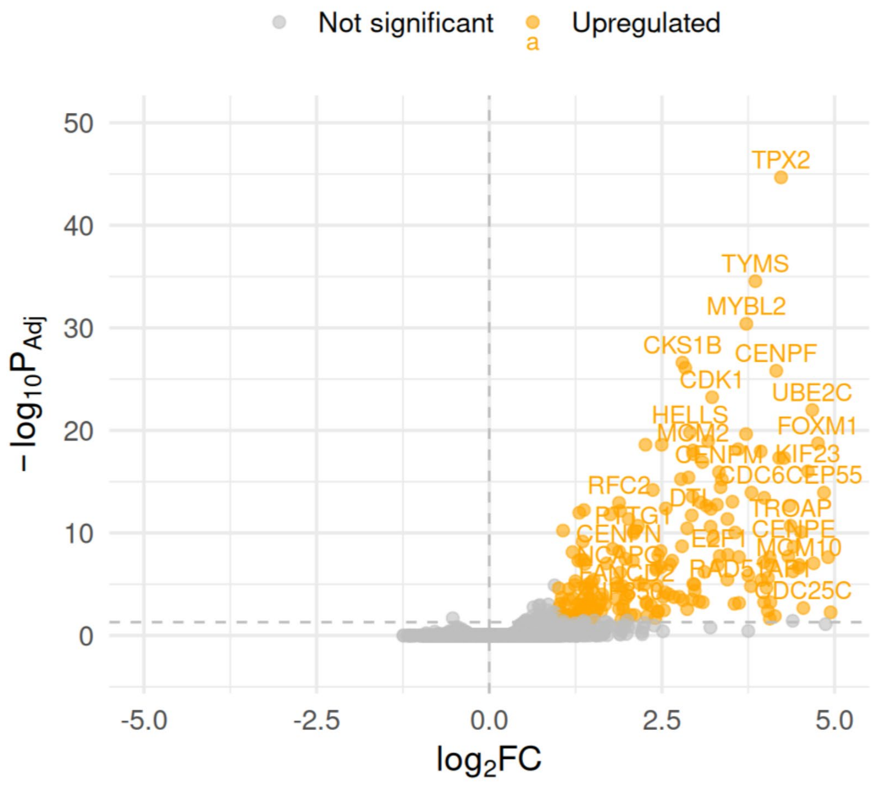
G1 - Growth

S - DNA synthesis

G2 - Growth and preparation for mitosis

M - Mitosis (cell division)

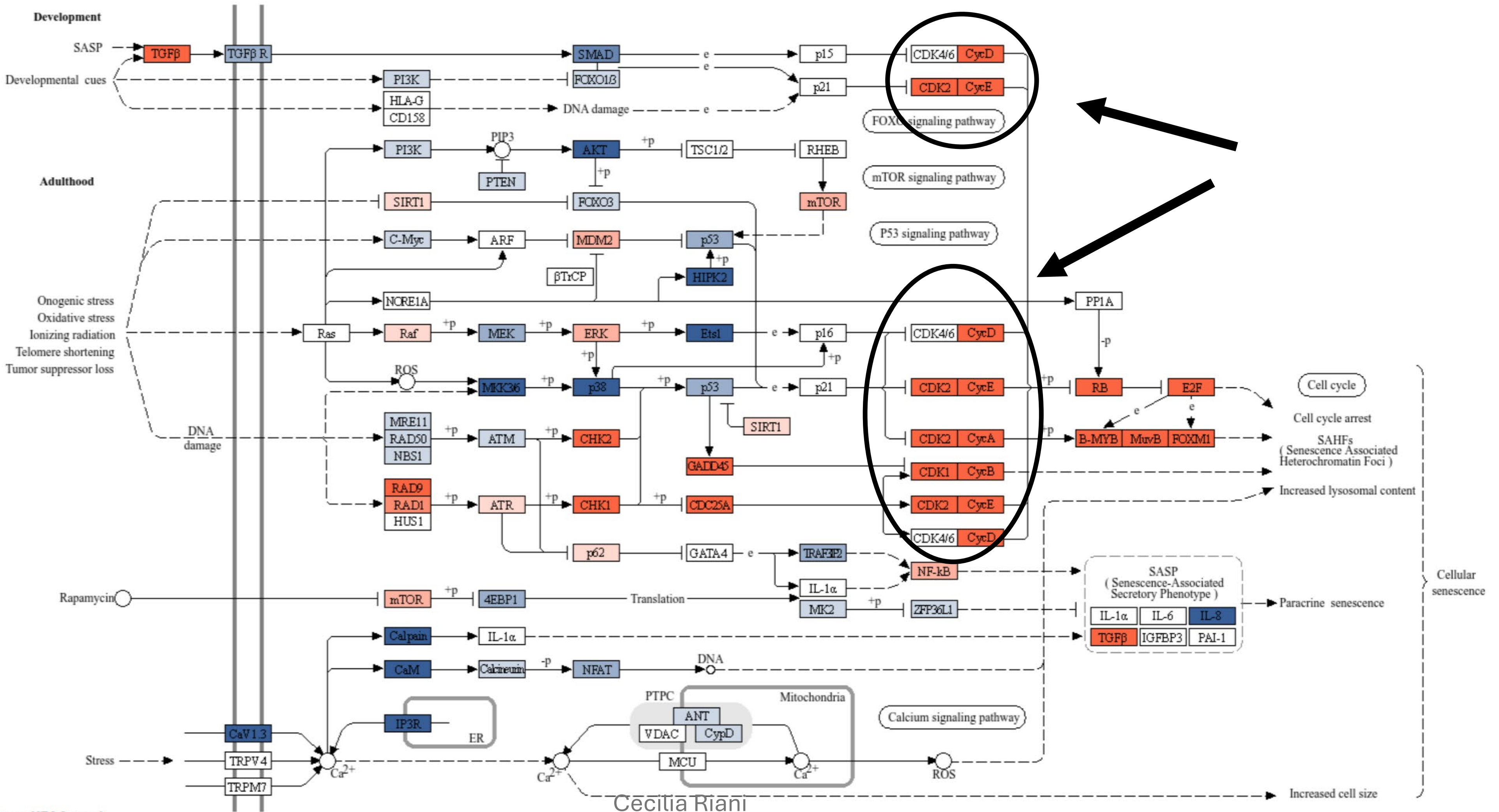
Cells in S/G2M phase (non-responders) against cells in G1 phase (responders)



scRNAseq analysis of L-asparaginase Treatment Response of leukemia cell line RS4;11

Cellular Senescence

KEGG
Kyoto Encyclopedia of Genes and Genome



Data on KEGG graph
Rendered by Pathview

IMAGEOMICS

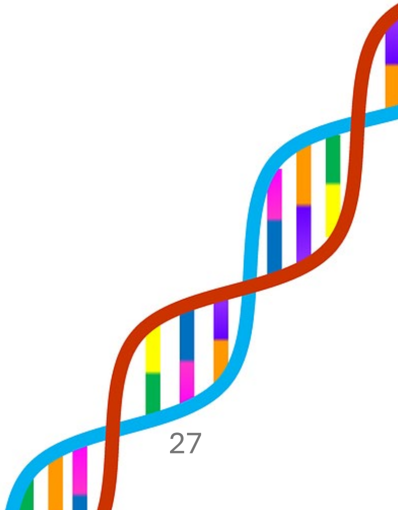
PIANOFORTE
EUROPEAN PARTNERSHIP FOR
RADIATION PROTECTION RESEARCH

UNIVERSITÀ
DI PAVIA

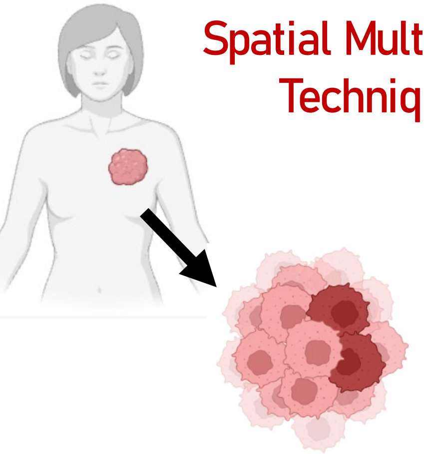
ISGlobal Barcelona
Institute for
Global Health

OTTO VON GUERICKE
UNIVERSITÄT
MAGDEBURG

NNGYK



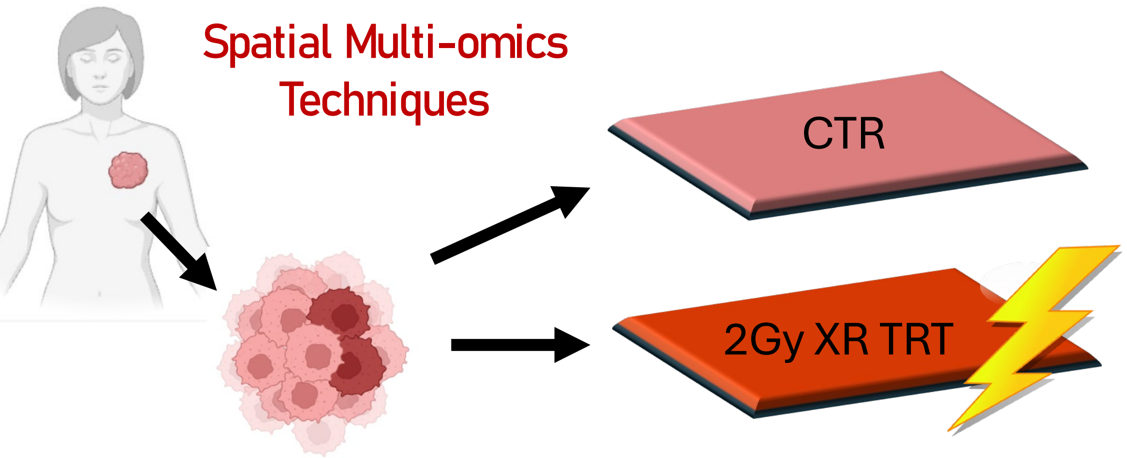
Spatial Multi-omics Techniques



IMAGEOMICS

*Optimizing Benefit/Risk Ratio in **Breast Cancer** Diagnosis and Radiotherapy:*

Identifying Molecular, Cellular and Imaging Signatures of Breast Cancer Heterogeneity to Improve Personalized Therapeutic Strategies for Synergistic Treatment Combinations



Spatial Multi-omics Techniques

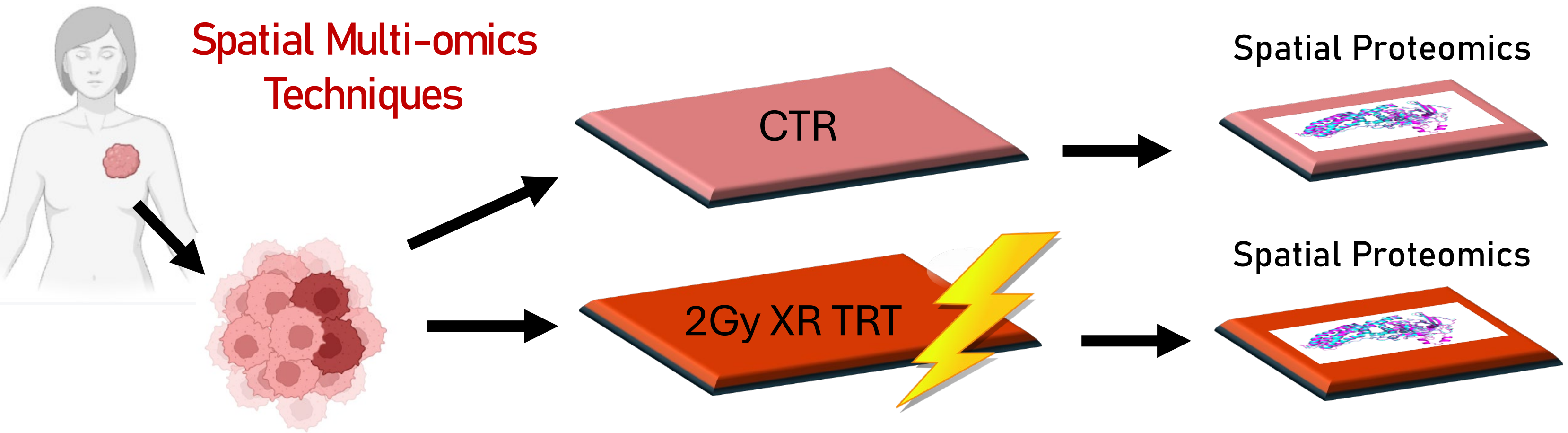
CTR

2Gy XR TRT

Surgery Department, San Matteo Hospital
Anatomy Pathology Unit, San Matteo Hospital
Medical Physics Department, San Matteo Hospital

IMAGEOMICS

Optimizing Benefit/Risk Ratio in ***Breast Cancer*** *Diagnosis and Radiotherapy:*
Identifying Molecular, Cellular and Imaging Signatures of Breast Cancer Heterogeneity to Improve Personalized Therapeutic Strategies for Synergistic Treatment Combinations

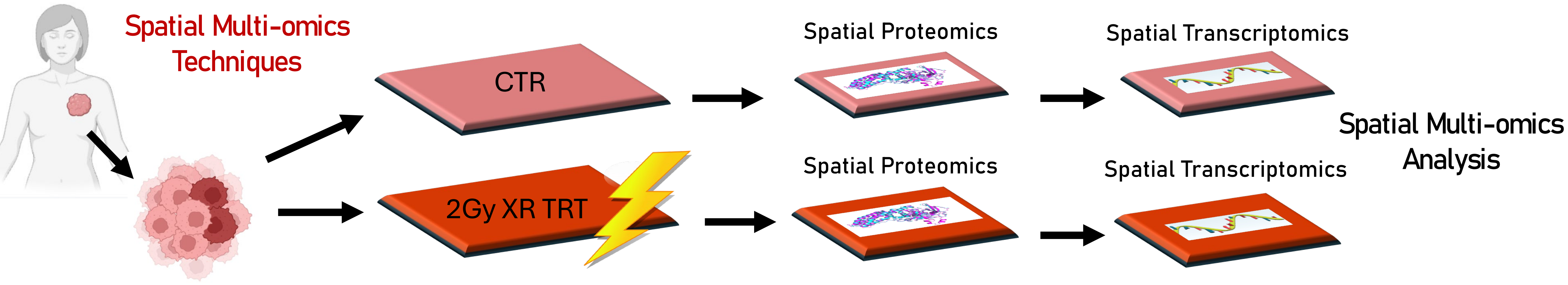


Surgery Department, San Matteo Hospital
Anatomy Pathology Unit, San Matteo Hospital
Medical Physics Department, San Matteo Hospital

Cecilia Riani

IMAGEOMICS

*Optimizing Benefit/Risk Ratio in **Breast Cancer** Diagnosis and Radiotherapy: Identifying Molecular, Cellular and Imaging Signatures of Breast Cancer Heterogeneity to Improve Personalized Therapeutic Strategies for Synergistic Treatment Combinations*



Spatial Multi-omics Techniques

CTR

2Gy XR TRT

Spatial Proteomics

Spatial Transcriptomics

Spatial Proteomics

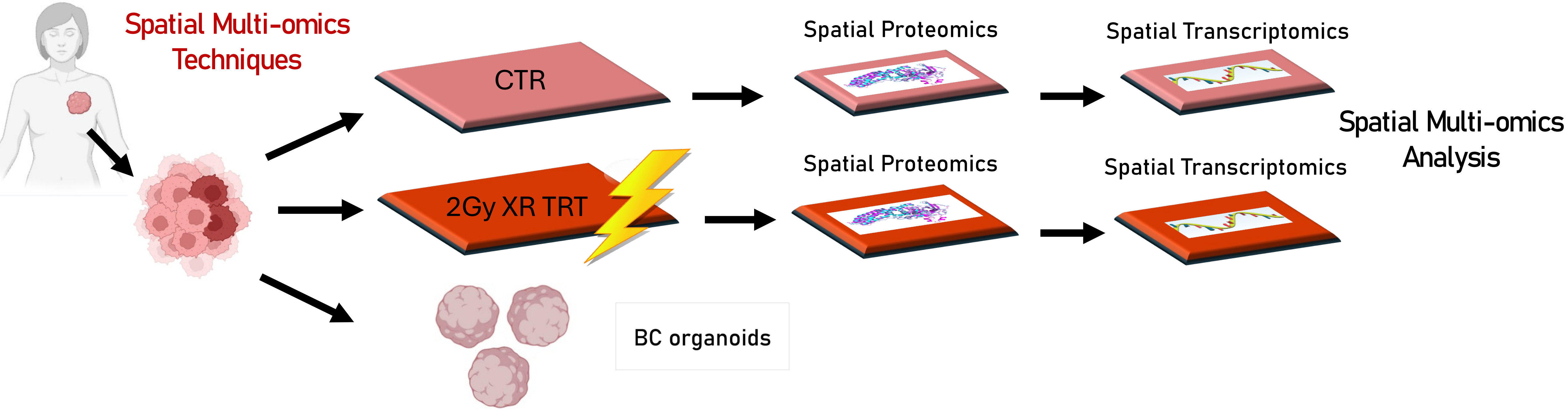
Spatial Transcriptomics

Spatial Multi-omics Analysis

IMAGEOMICS

*Optimizing Benefit/Risk Ratio in **Breast Cancer** Diagnosis and Radiotherapy:*

Identifying Molecular, Cellular and Imaging Signatures of Breast Cancer Heterogeneity to Improve Personalized Therapeutic Strategies for Synergistic Treatment Combinations



Spatial Multi-omics Techniques

CTR

2Gy XR TRT

BC organoids

Spatial Proteomics

Spatial Transcriptomics

Spatial Proteomics

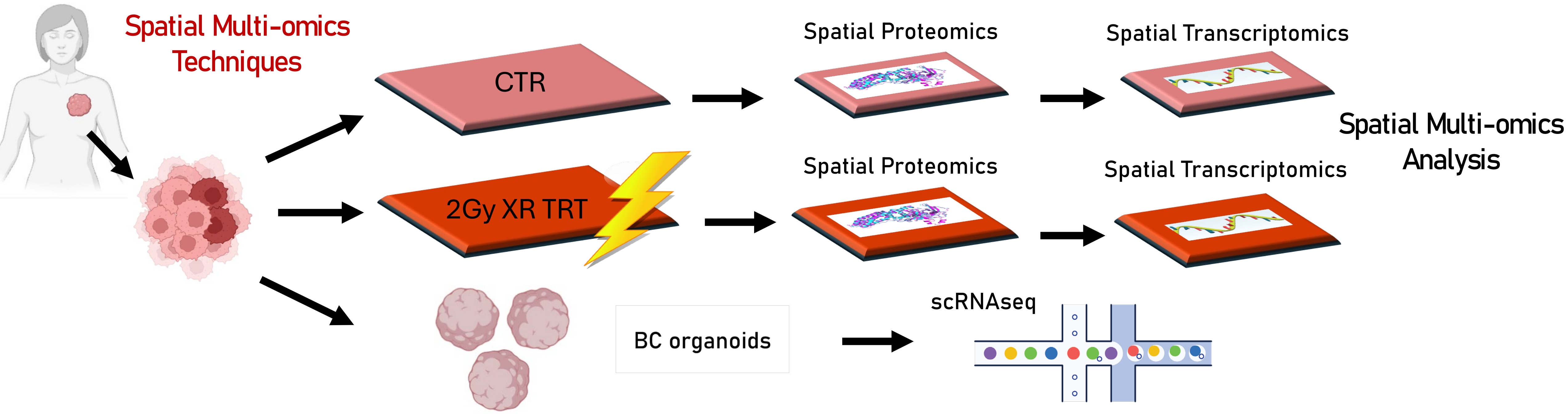
Spatial Transcriptomics

Spatial Multi-omics Analysis

PIANOFORTE
EUROPEAN PARTNERSHIP FOR RADIATION PROTECTION RESEARCH

IMAGEOMICS

*Optimizing Benefit/Risk Ratio in **Breast Cancer** Diagnosis and Radiotherapy: Identifying Molecular, Cellular and Imaging Signatures of Breast Cancer Heterogeneity to Improve Personalized Therapeutic Strategies for Synergistic Treatment Combinations*



Spatial Multi-omics Techniques

CTR

2Gy XR TRT

BC organoids

Spatial Proteomics

Spatial Transcriptomics

Spatial Proteomics

Spatial Transcriptomics

scRNAseq

Spatial Multi-omics Analysis

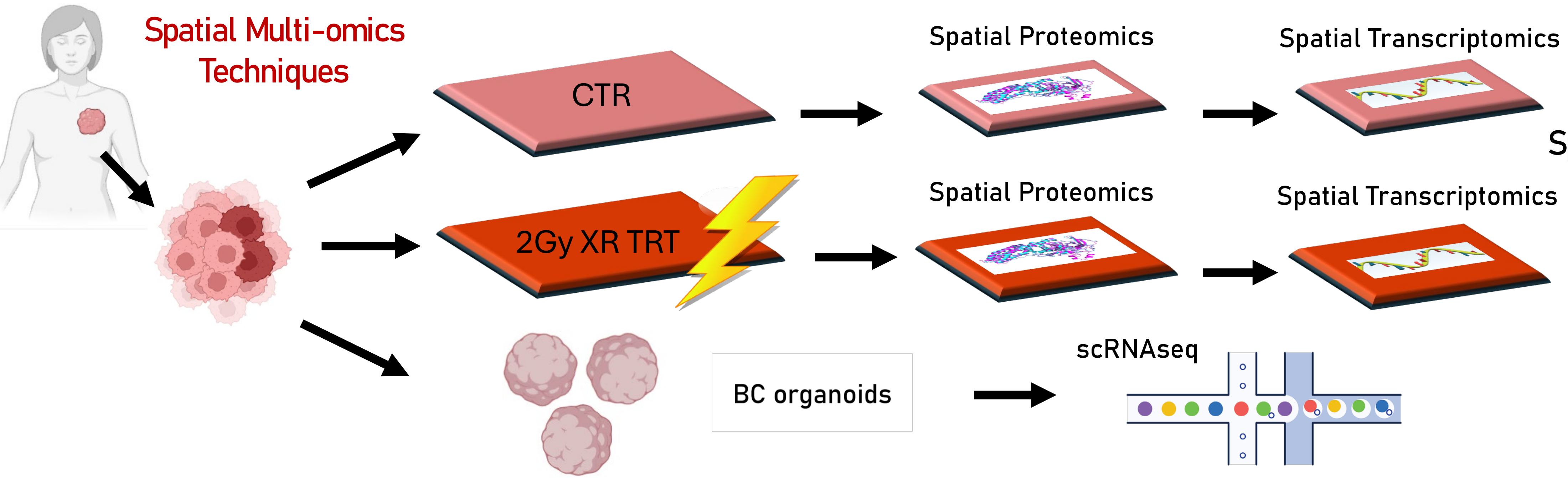
Surgery Department, San Matteo Hospital
Anatomy Pathology Unit, San Matteo Hospital
Medical Physics Department, San Matteo Hospital

PIANOFORTE
EUROPEAN PARTNERSHIP FOR RADIATION PROTECTION RESEARCH

IMAGEOMICS

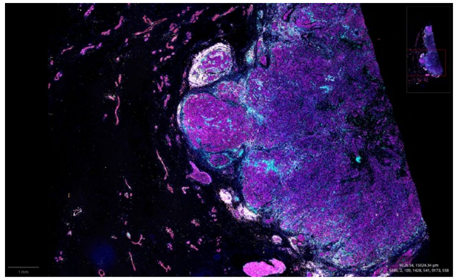
*Optimizing Benefit/Risk Ratio in **Breast Cancer** Diagnosis and Radiotherapy:*

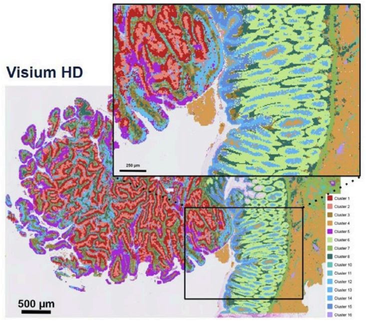
Identifying Molecular, Cellular and Imaging Signatures of Breast Cancer Heterogeneity to Improve Personalized Therapeutic Strategies for Synergistic Treatment Combinations



Spatial Multi-omics Analysis

Akoya Pheno Cycler

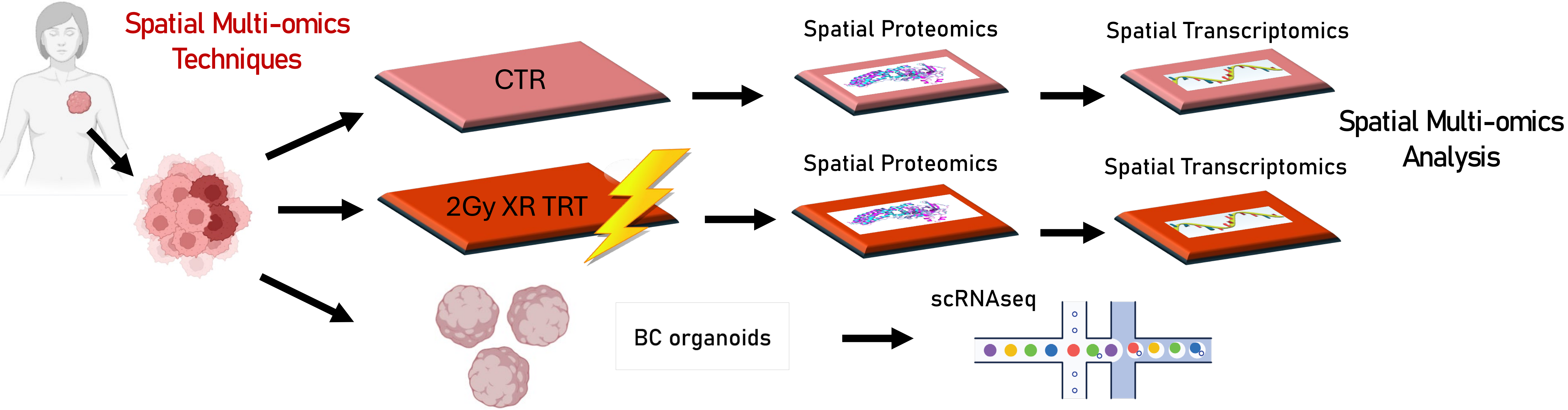




Surgery Department, San Matteo Hospital
Anatomy Pathology Unit, San Matteo Hospital
Medical Physics Department, San Matteo Hospital

IMAGEOMICS

*Optimizing Benefit/Risk Ratio in **Breast Cancer** Diagnosis and Radiotherapy: Identifying Molecular, Cellular and Imaging Signatures of Breast Cancer Heterogeneity to Improve Personalized Therapeutic Strategies for Synergistic Treatment Combinations*



Spatial Multi-omics Techniques

CTR

2Gy XR TRT

Spatial Proteomics

Spatial Transcriptomics

Spatial Proteomics

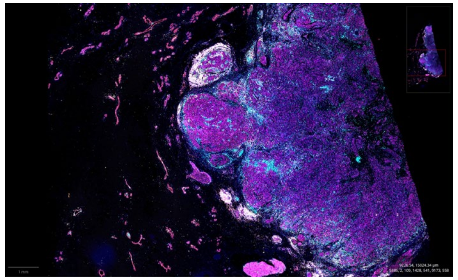
Spatial Transcriptomics

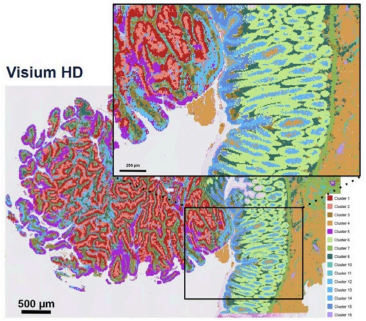
Spatial Multi-omics Analysis

BC organoids

scRNAseq

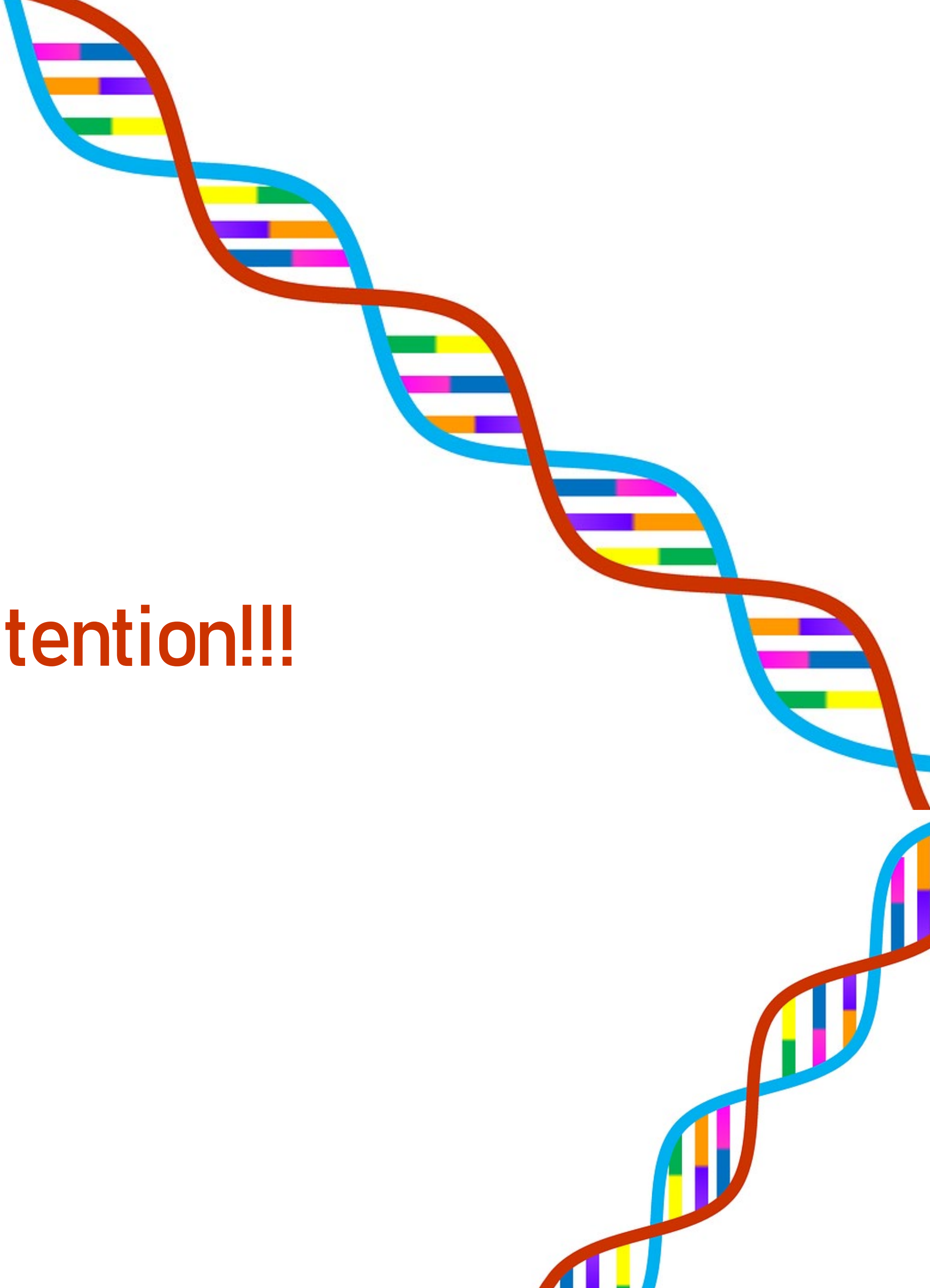
Akoya Pheno Cycler



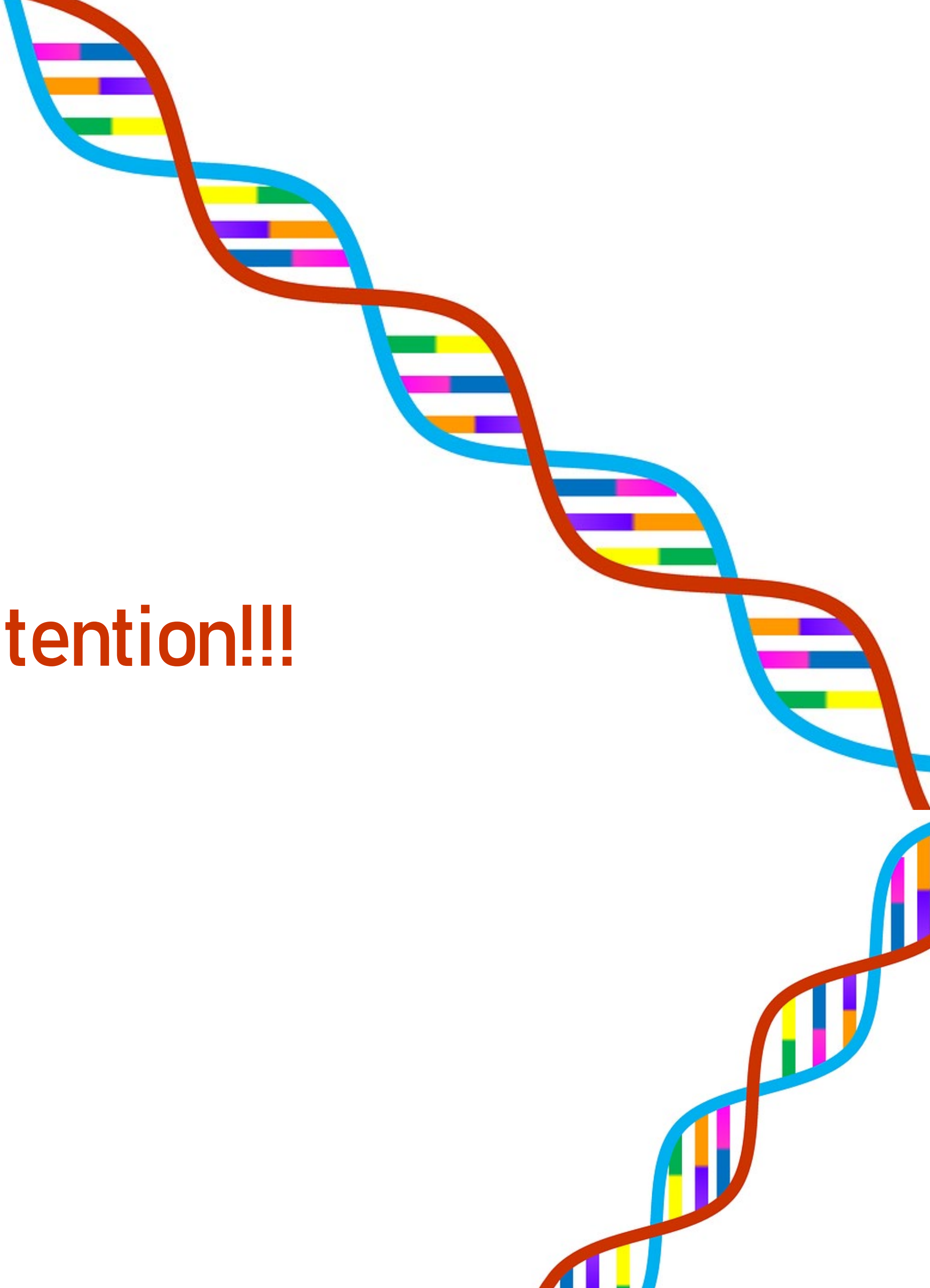


- *Exploit protein marker to classify BC and identify different cell types*
- *Identify spatially variable genes through proper DGE methods*
- *Identify local marker of response to radiation treatment*
- *Decipher tumor microenvironment interactions*
- *Identify biological mechanism up-regulated or depleted*

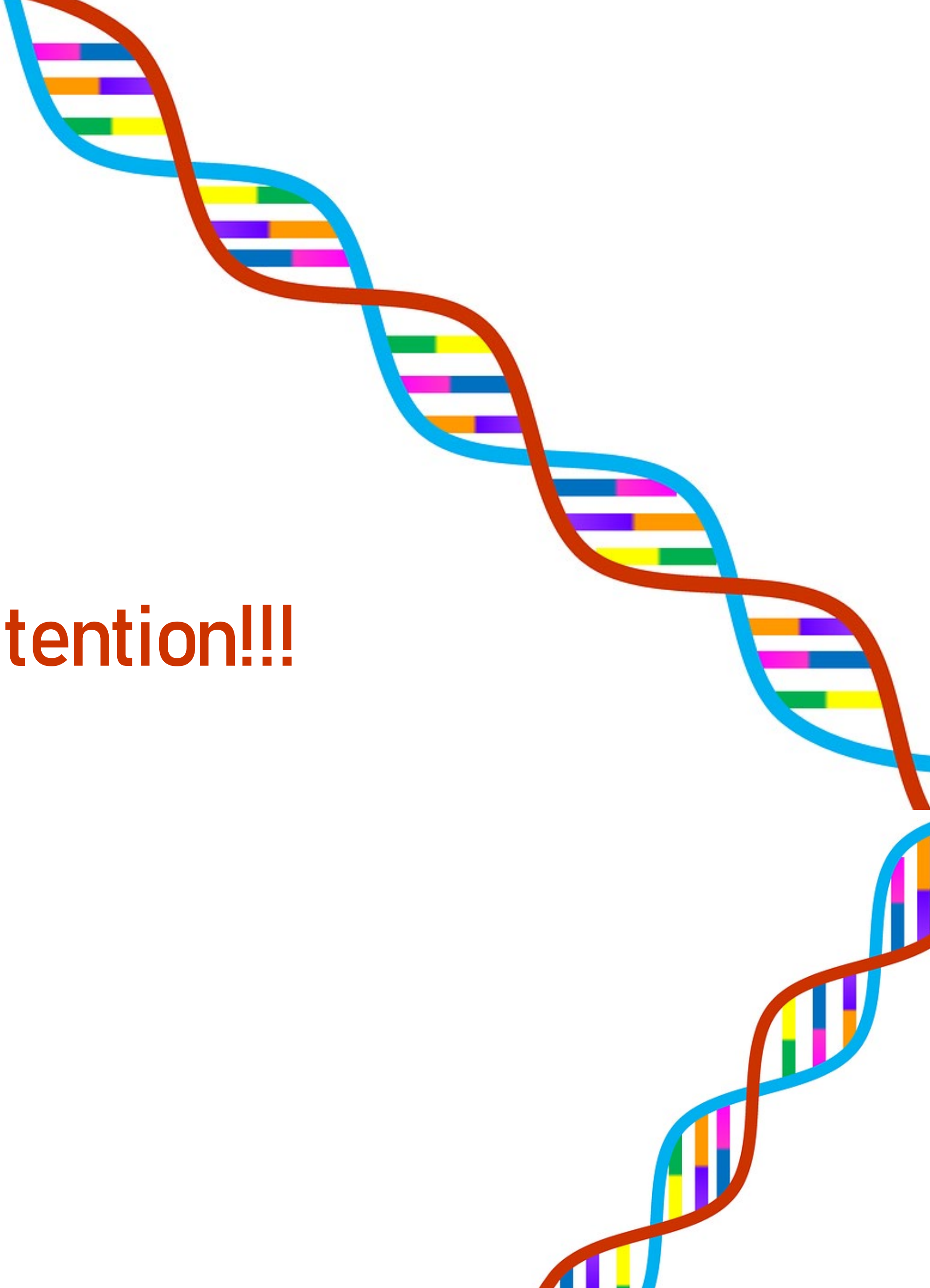
Surgery Department, San Matteo Hospital
Anatomy Pathology Unit, San Matteo Hospital
Medical Physics Department, San Matteo Hospital



THANK you for the attention!!!

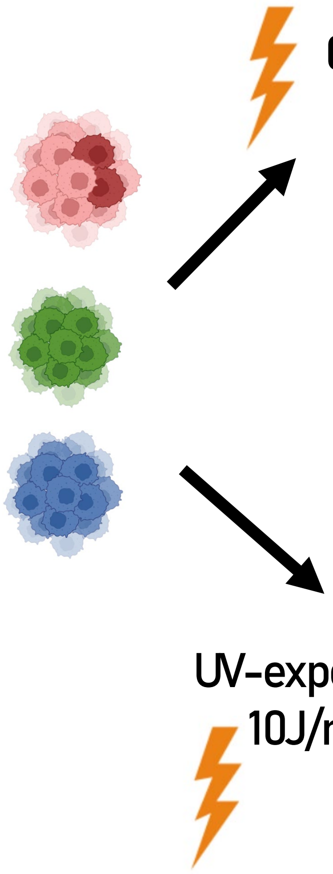


THANK you for the attention!!!



THANK you for the attention!!!

Investigation of CPD and $\gamma H2AX$ markers after Xrays and UV exposure

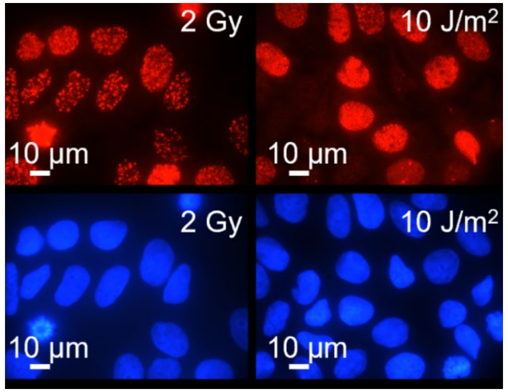

HeLa

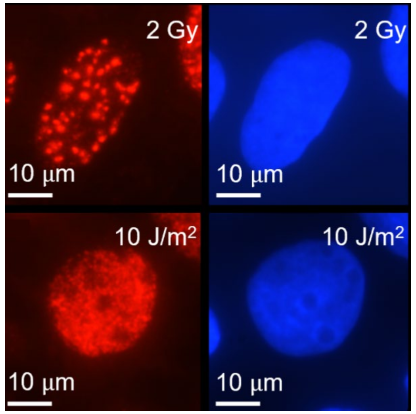
HaCat

Caco-2

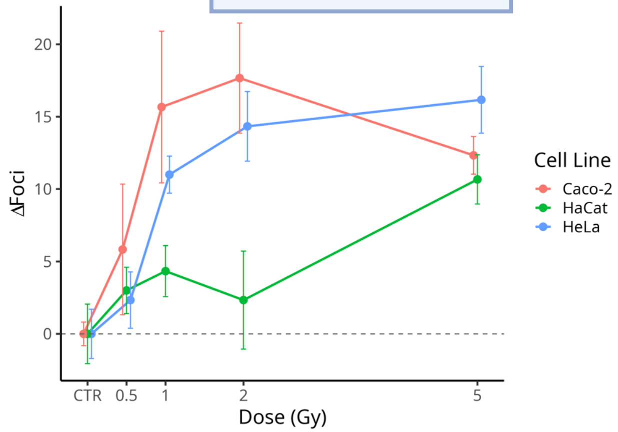
Xrays
0.5,1,2,5 Gy

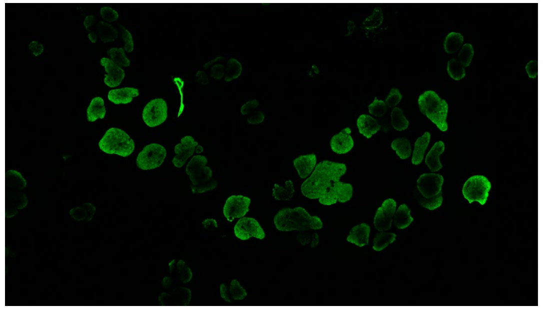
UV-exposure
10J/m2



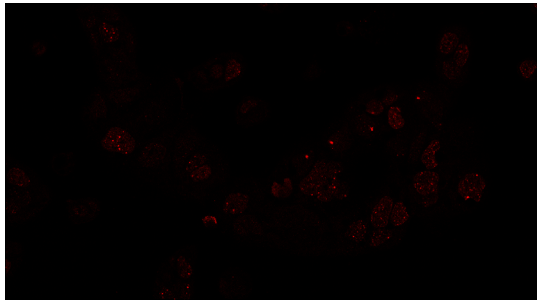


Nuclei & foci segmentation

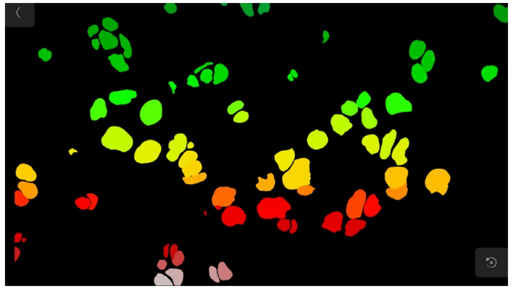


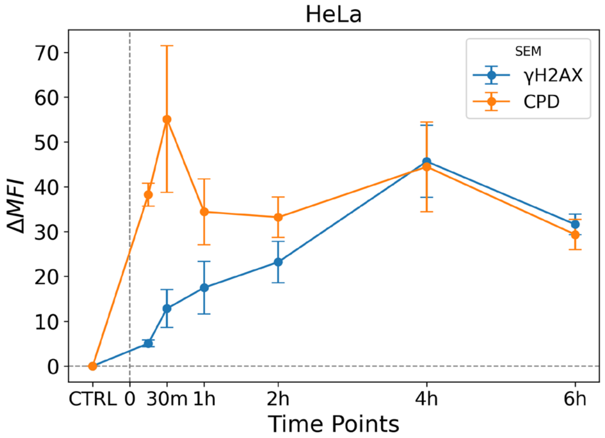


Confocal microscopy acquisition



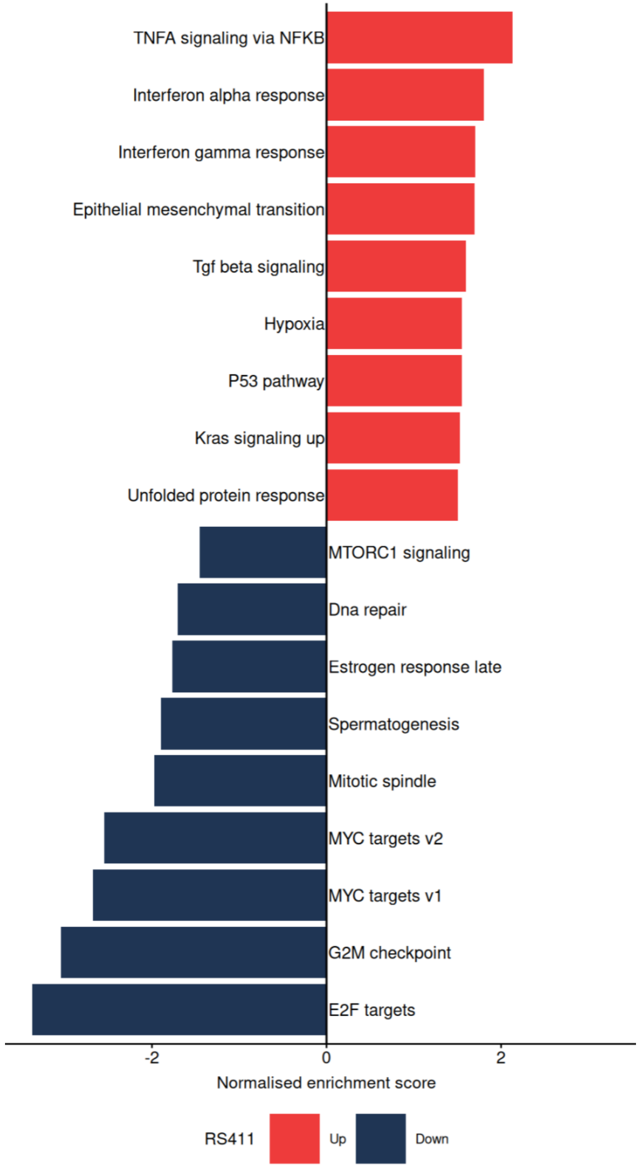
DL based nuclei segmentation

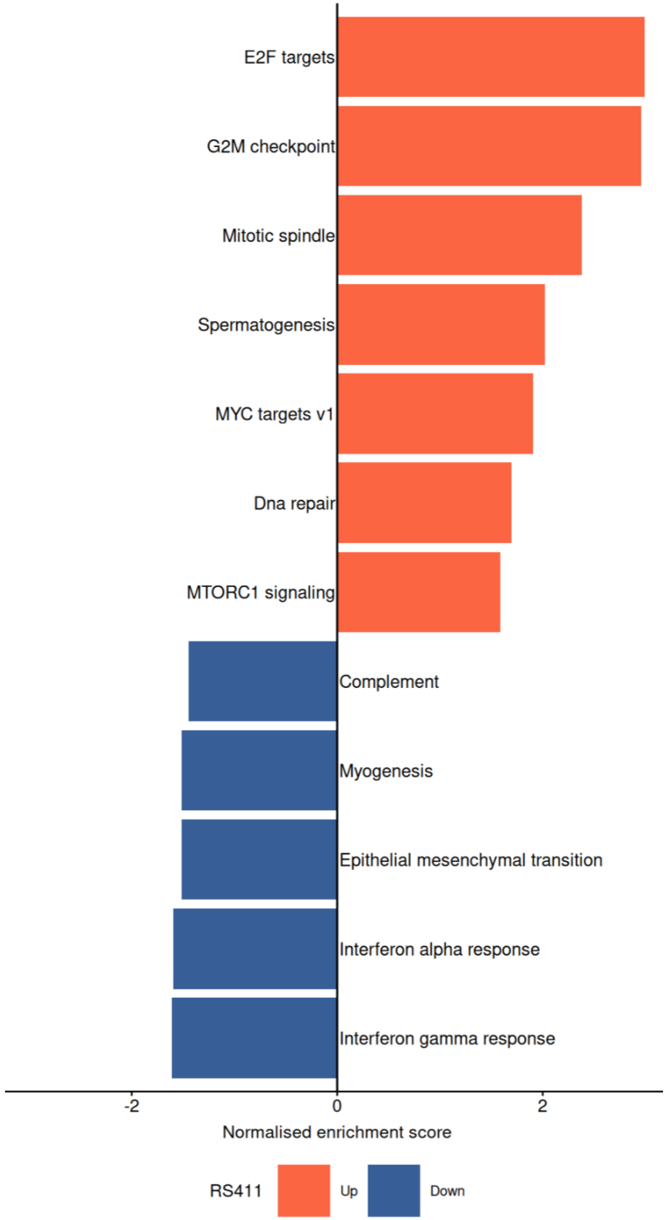




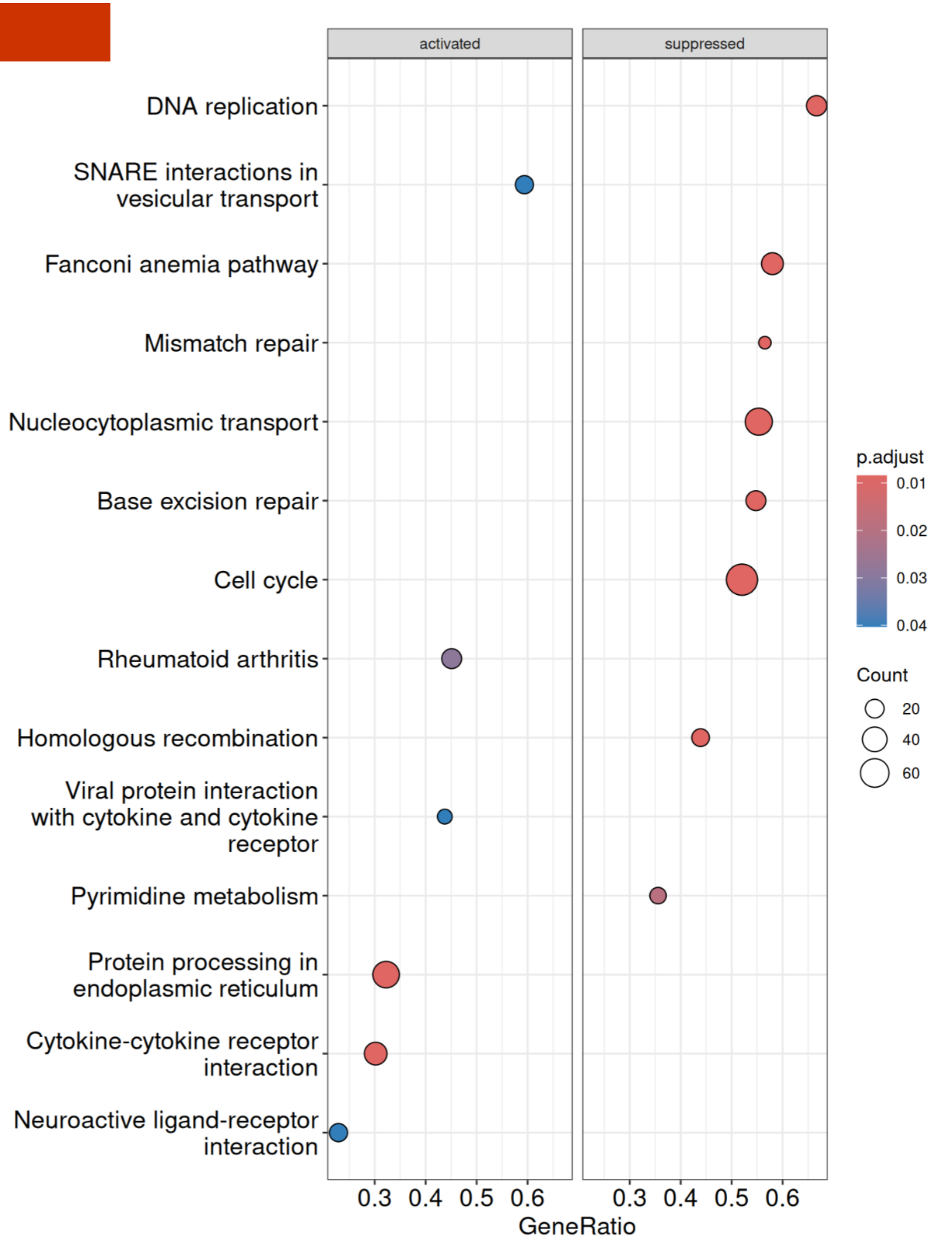
Cecilia Riani

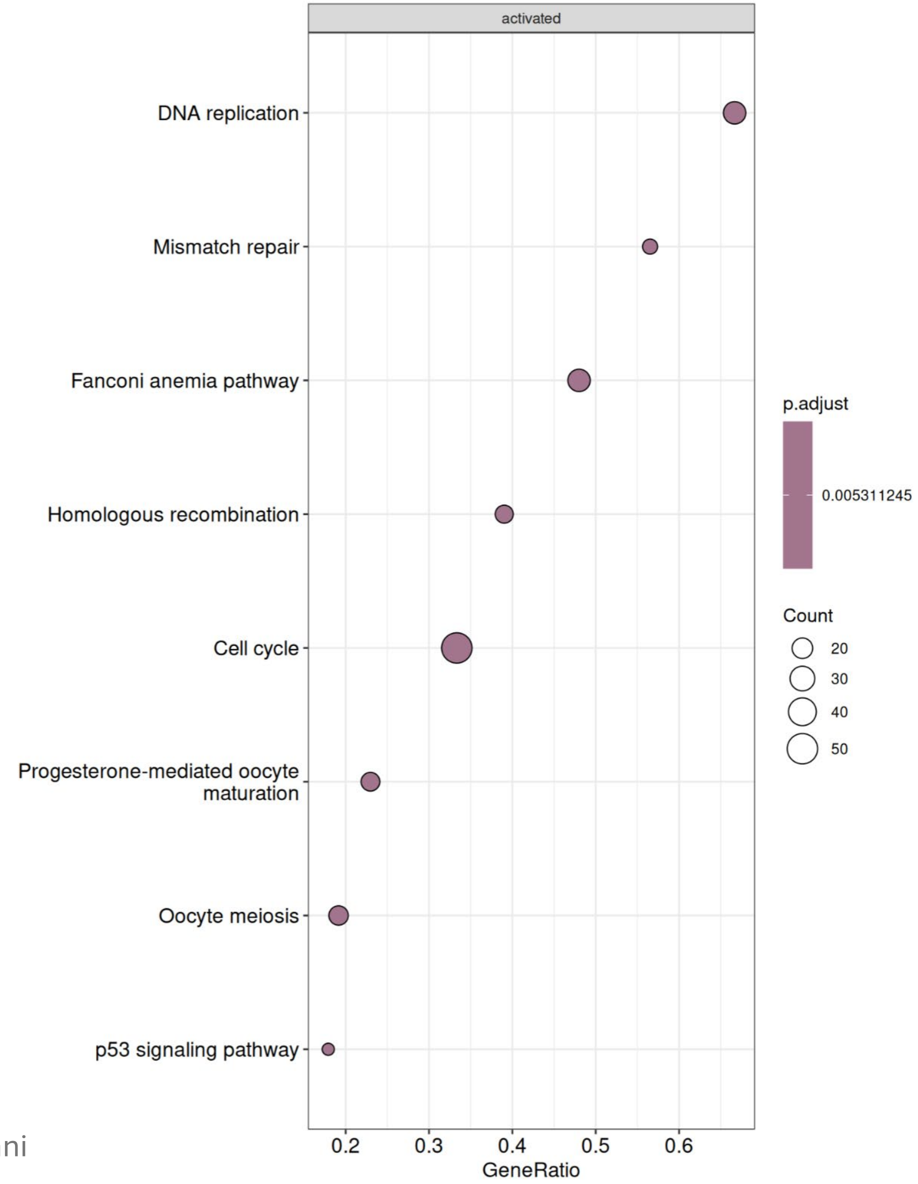
Pathway Enrichment Analysis





Gene set Enrichment Analysis





Dimensionality reduction techniques

- PCA (Principal Component Analysis) linear methods
- ICA (indipendent component analysis) linear methods can be used for visualization and subsequent clustering
- T-sne (t-distributed stochastic neighbou embedding): non linear methods can be used only for visualization
- UMAP (Uniform Manifold approximation and projection): non linear methods can be used only for visualization

Clustering techniques

- K-means clustering
- Hierarchical clustering
- Graph based clustering (Louvain and Leiden methods)

Annotation process

1. Automated cell-type annotation, classifier-based methods and reference mapping (well-curated existing atlases). The quality of the transferred annotations depends on the quality of the reference data, the model and the suitability to the data set.
2. Manual annotation, leverages gene signatures of each cluster to annotate cell clusters. These gene signatures are commonly known as marker genes
3. Verification by experts

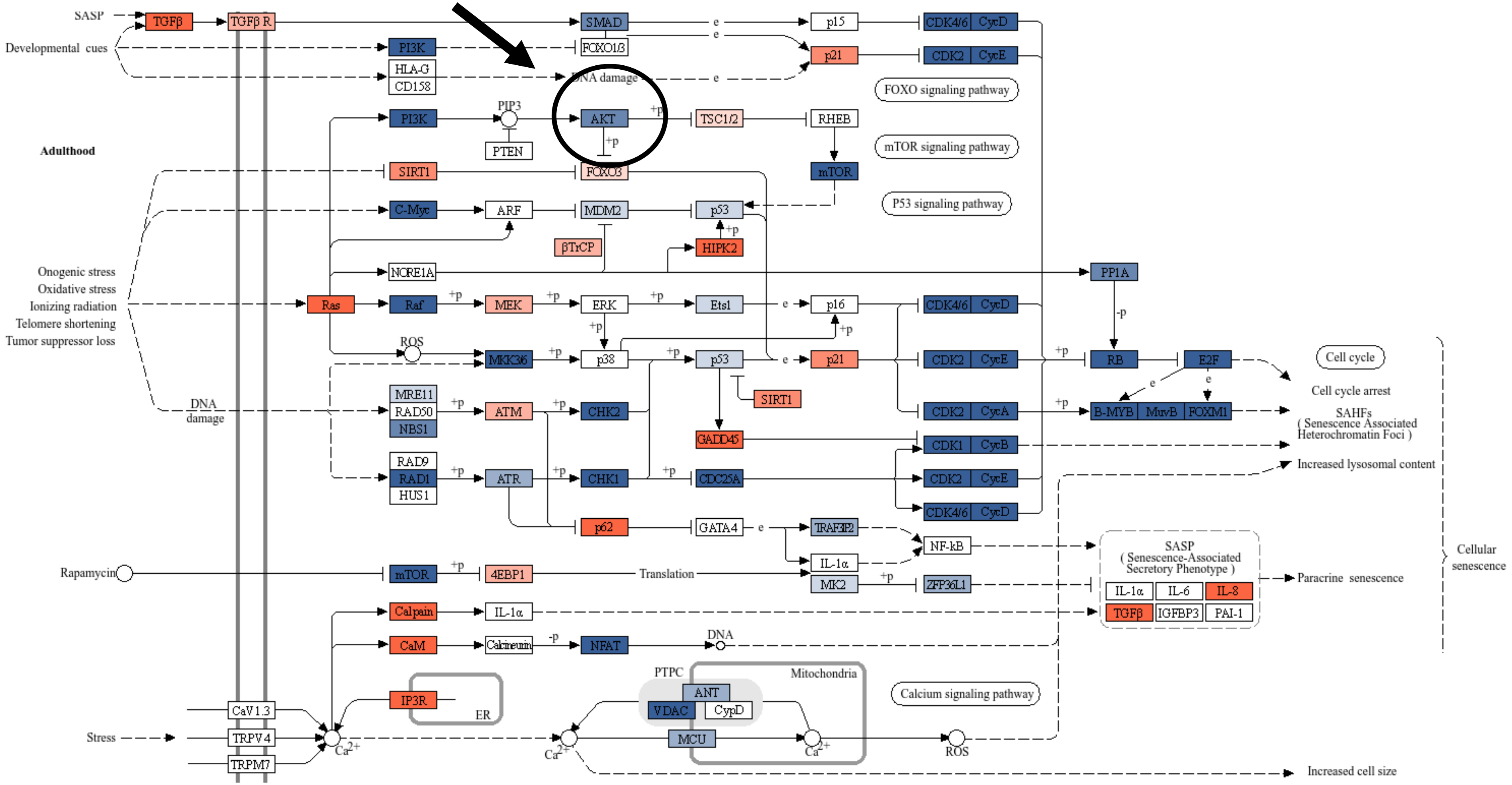
scRNAseq analysis of L-asparaginase Treatment Response of leukemia cell line RS4;11

AKT

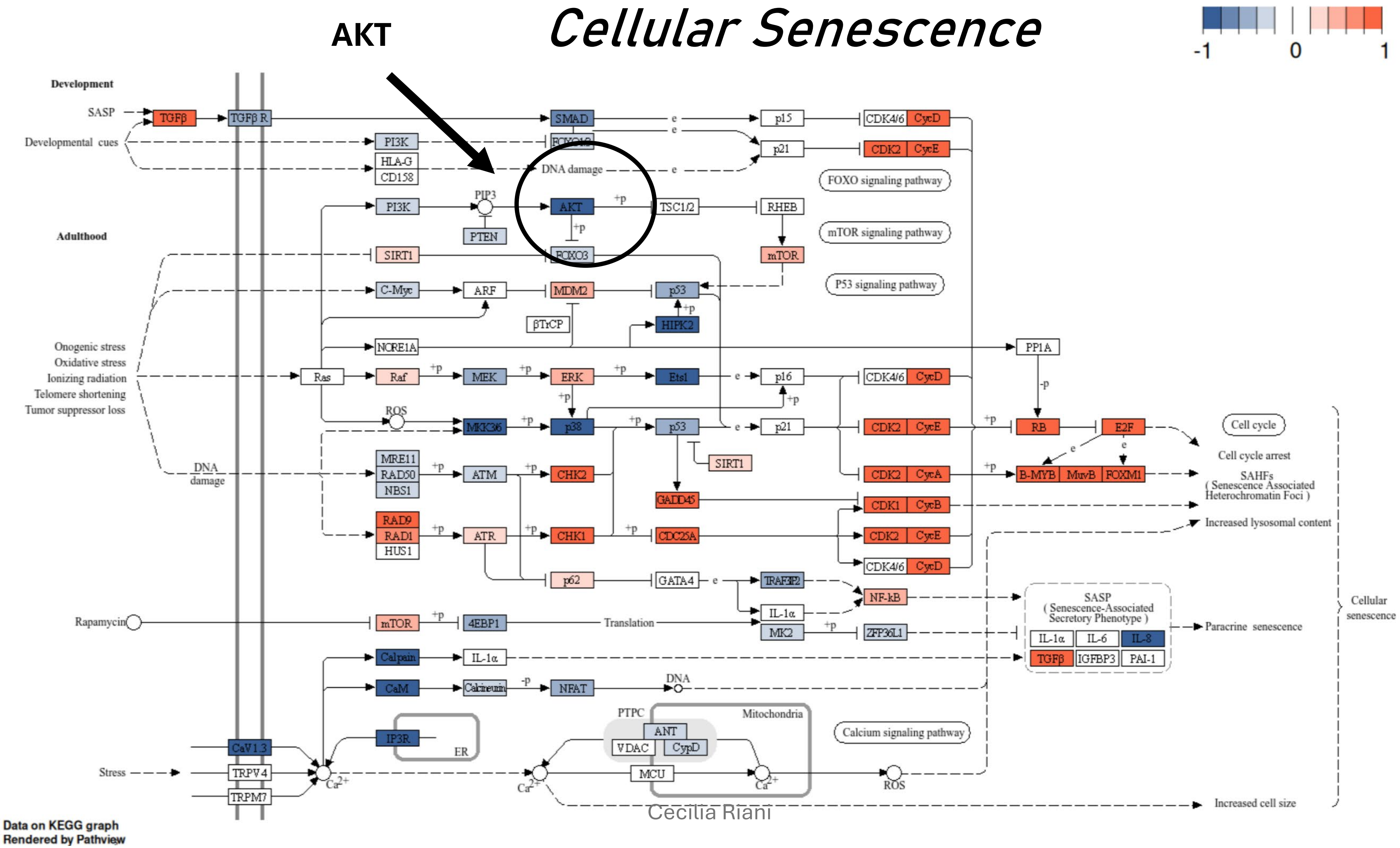
Cellular Senescence

-1 0 1

KEGG
Kyoto Encyclopedia of Genes and Genome



scRNAseq analysis of L-asparaginase Treatment Response of leukemia cell line RS4;11



AKT

Cellular Senescence

Cecilia Riani