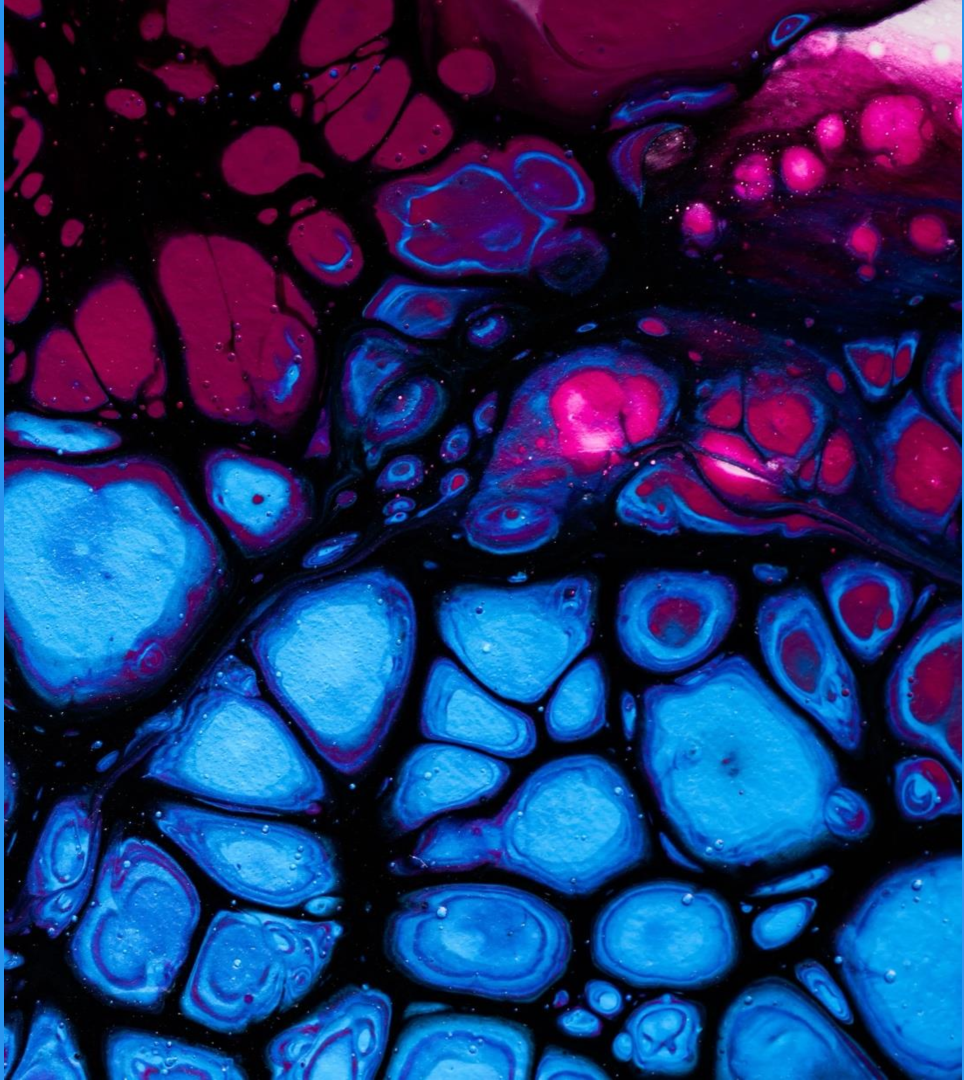
Modeling the Tumor Microenvironment with Graph ML

Marianna Rapsomaniki
Assistant Professor
Biomedical Data Science Center
UNIL | Université de Lausanne
CHUV
marianna.rapsomaniki@unil.ch

EMBO Spatial Biology of Cancer Workshop
December 3rd, 2024



Biomedical Data Science Center

Raphael Gottardo
Translational Data Science

Jacques Fellay
Precision Medicine

Jean-Louis Raisaro
Clinical Data Science

Marianna Rapsomaniki
AI/ML for Biomedicine

https://www.chuv.ch/en/bdsc/

Agenda

Introductions, Overview and Goals

Data | Single-cell and spatial omics

- Quick bio primer
- What is omics?
- Why *spatial* omics?

Break?

TME as a graph | Various applications

- Data exploration: ATHENA
- Biological Discovery: Concept learning

Exercise | Read and discuss paper: *Annotation of spatially resolved single-cell data with STELLAR*

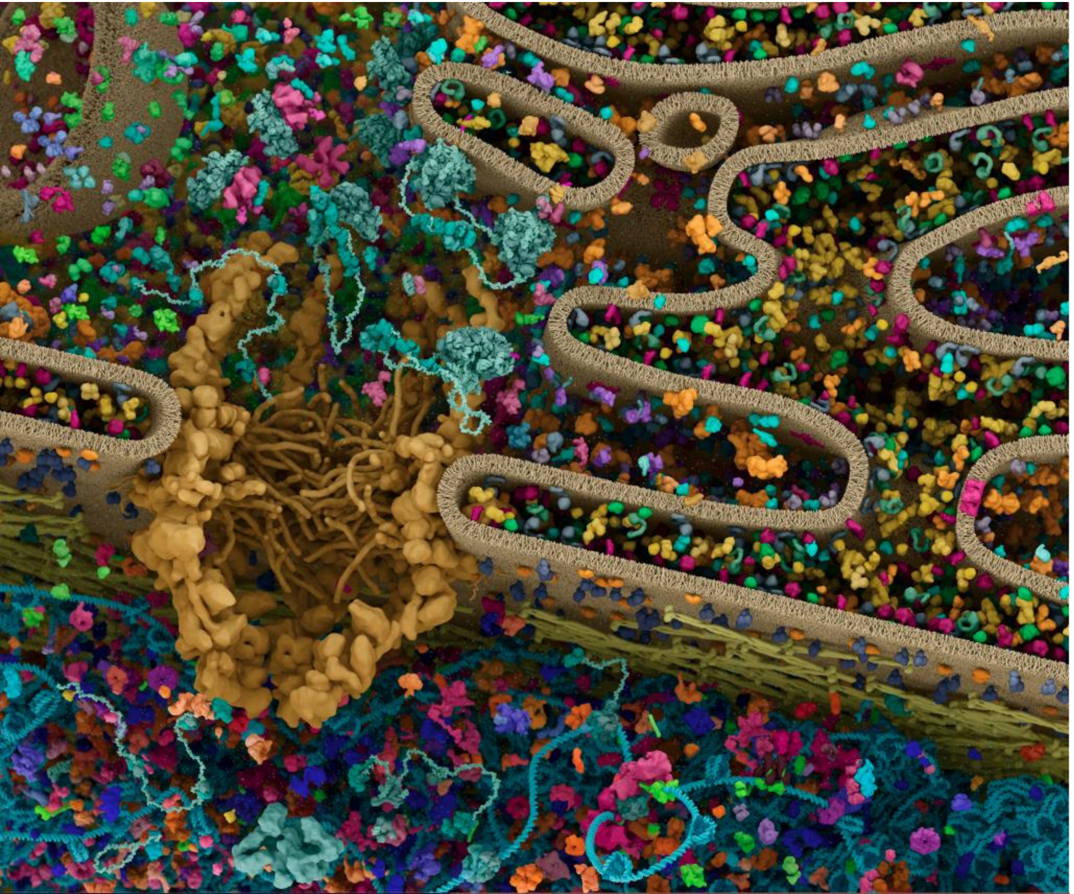
Recap

Conclusions, Q&A, and wrap-up

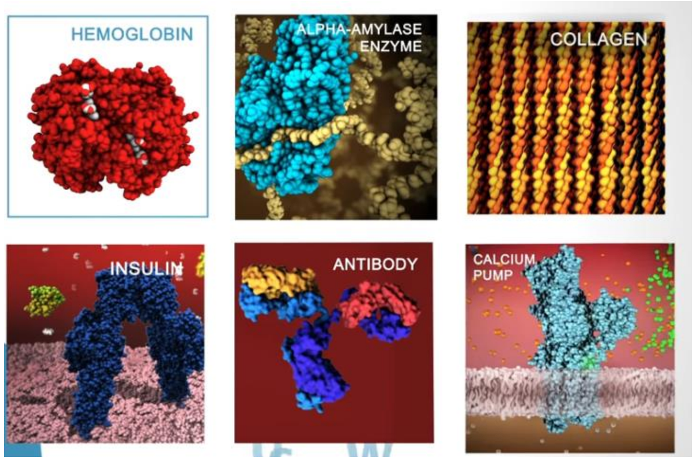
Goals? Questions?

https://web.speakup.info/room/join/73803

Proteins execute all essential functions of life

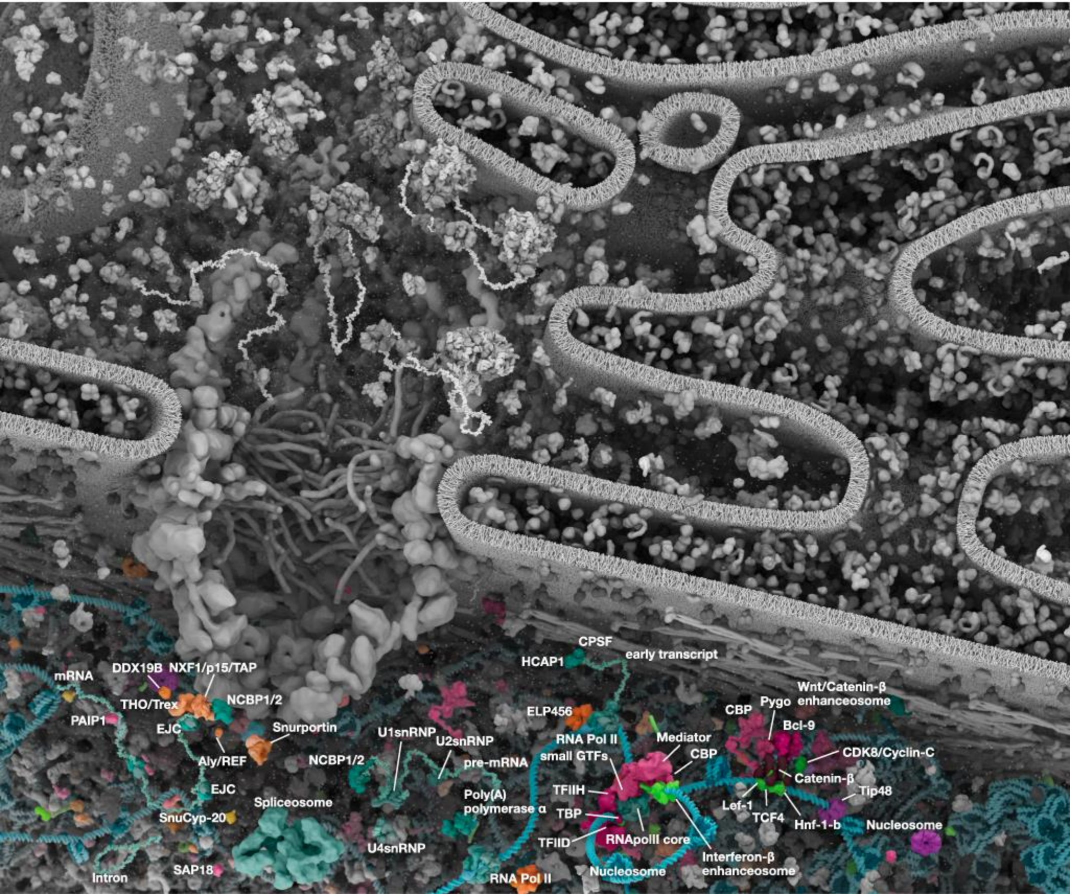


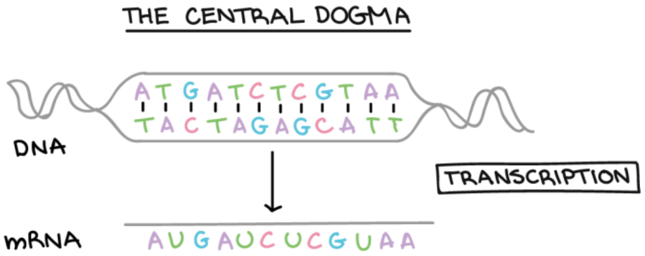
- Enzymes
- Structure
- Hormones
- Storage
- Response to stimuli
- Transport
- Immune response



Cell illustrations from https://www.cellsignal.com

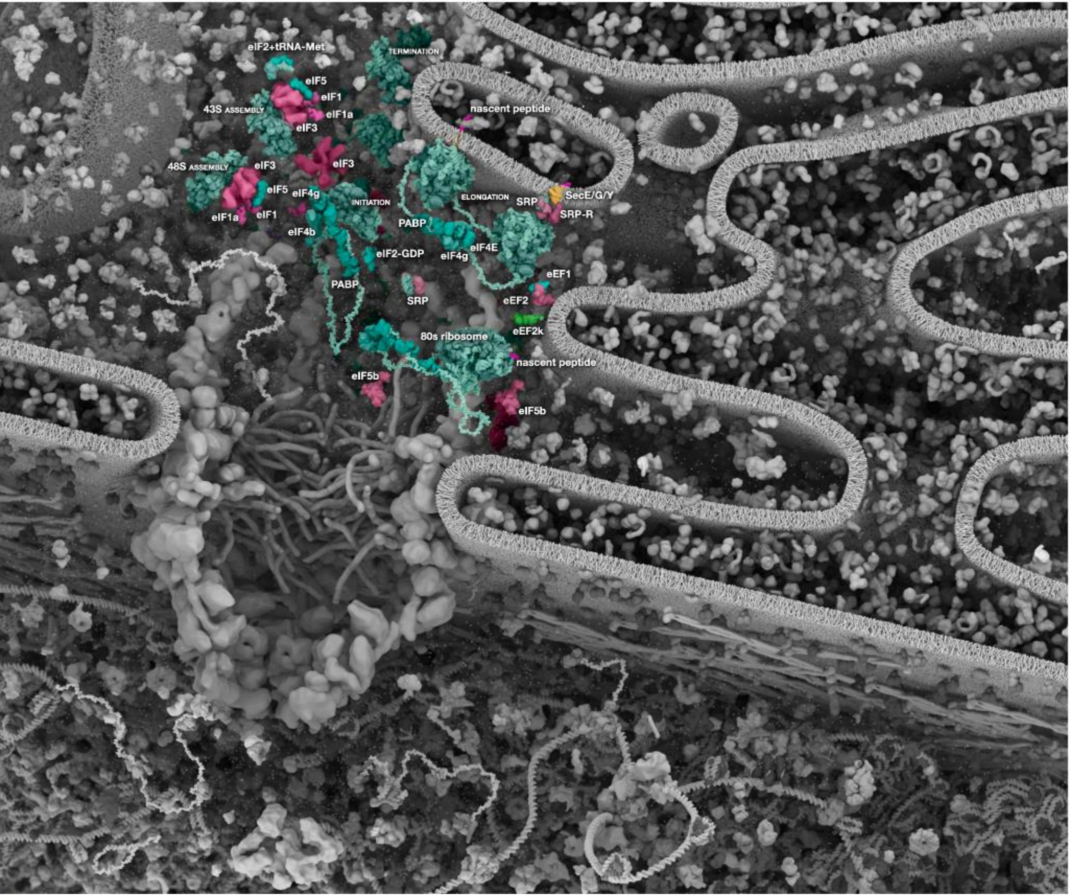
How is this achieved?

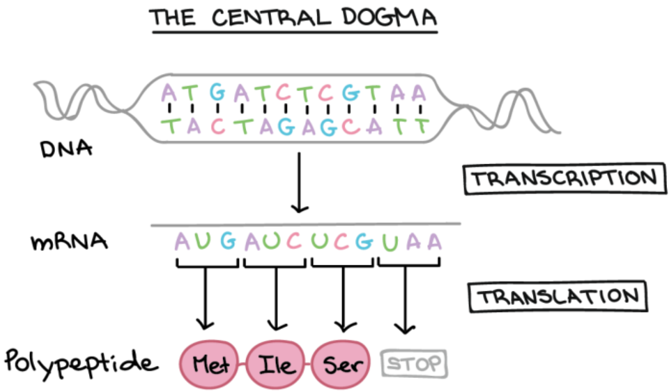




Cell illustrations from https://www.cellsignal.com

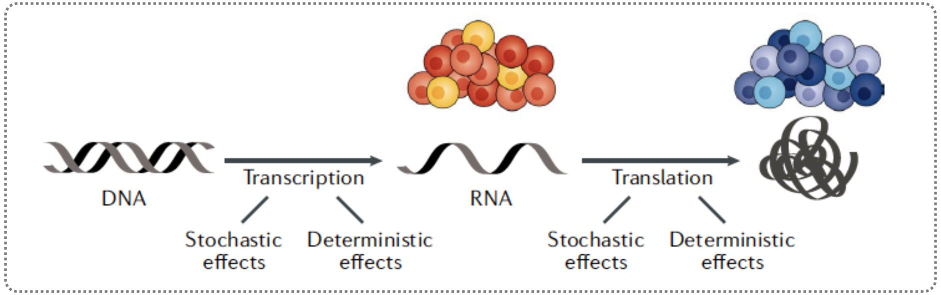
How is this achieved?





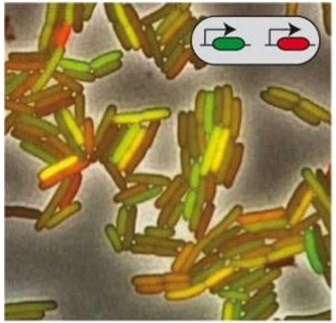
Cell illustrations from https://www.cellsignal.com

Biological stochasticity



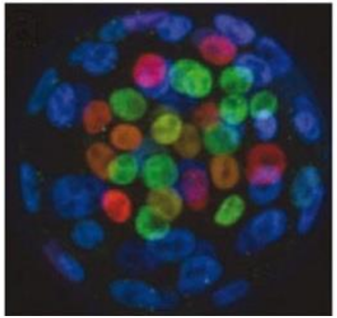
Eling *et al.*, 2019

Gene expression



Elowitz, *et al. Science* (2002)

Differentiation

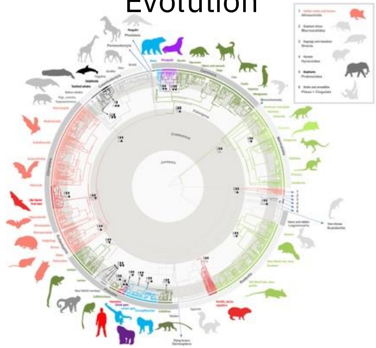


Eldar, *et al. Nature* (2010)

Pattern formation

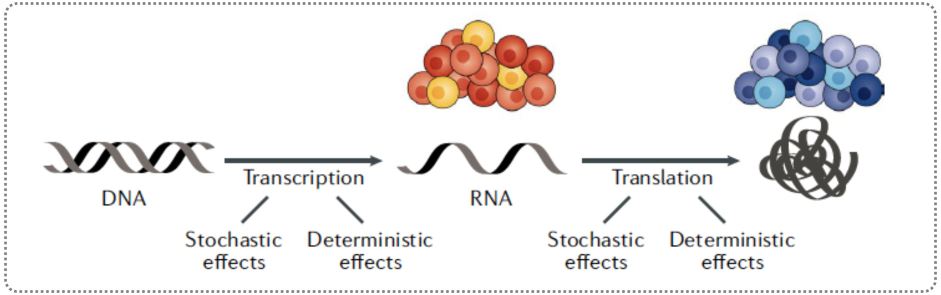
Image credit: https://ifisc.uib-csic.es/

Evolution

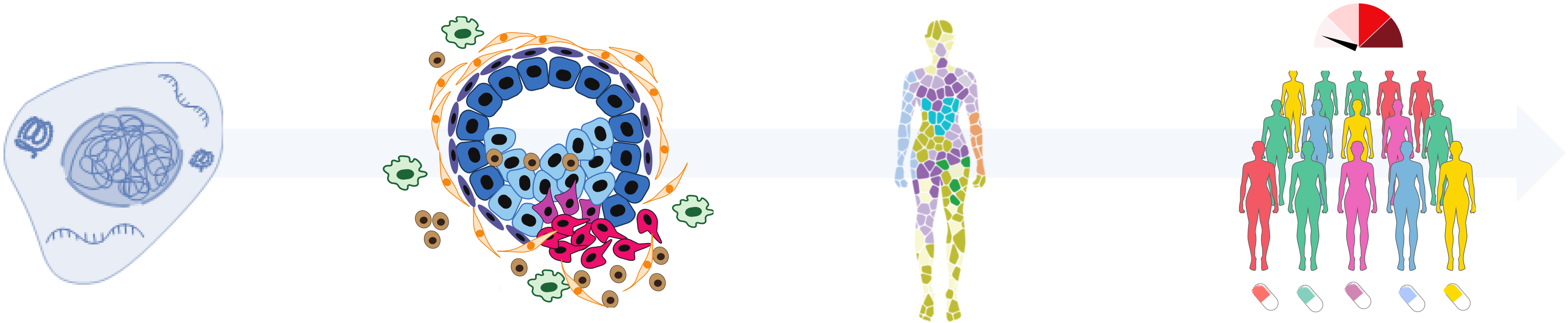


Smaers *et al., Science Advances*, 2021

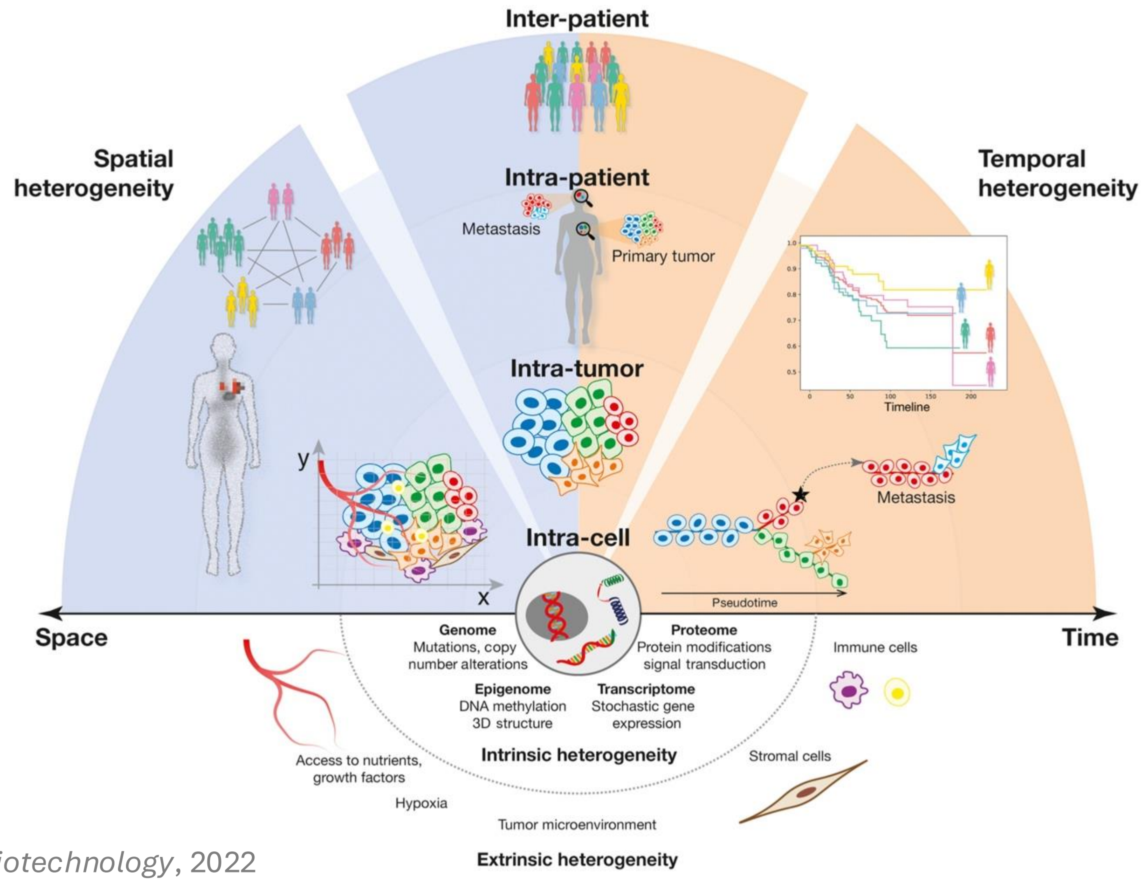
From stochasticity to tumor heterogeneity



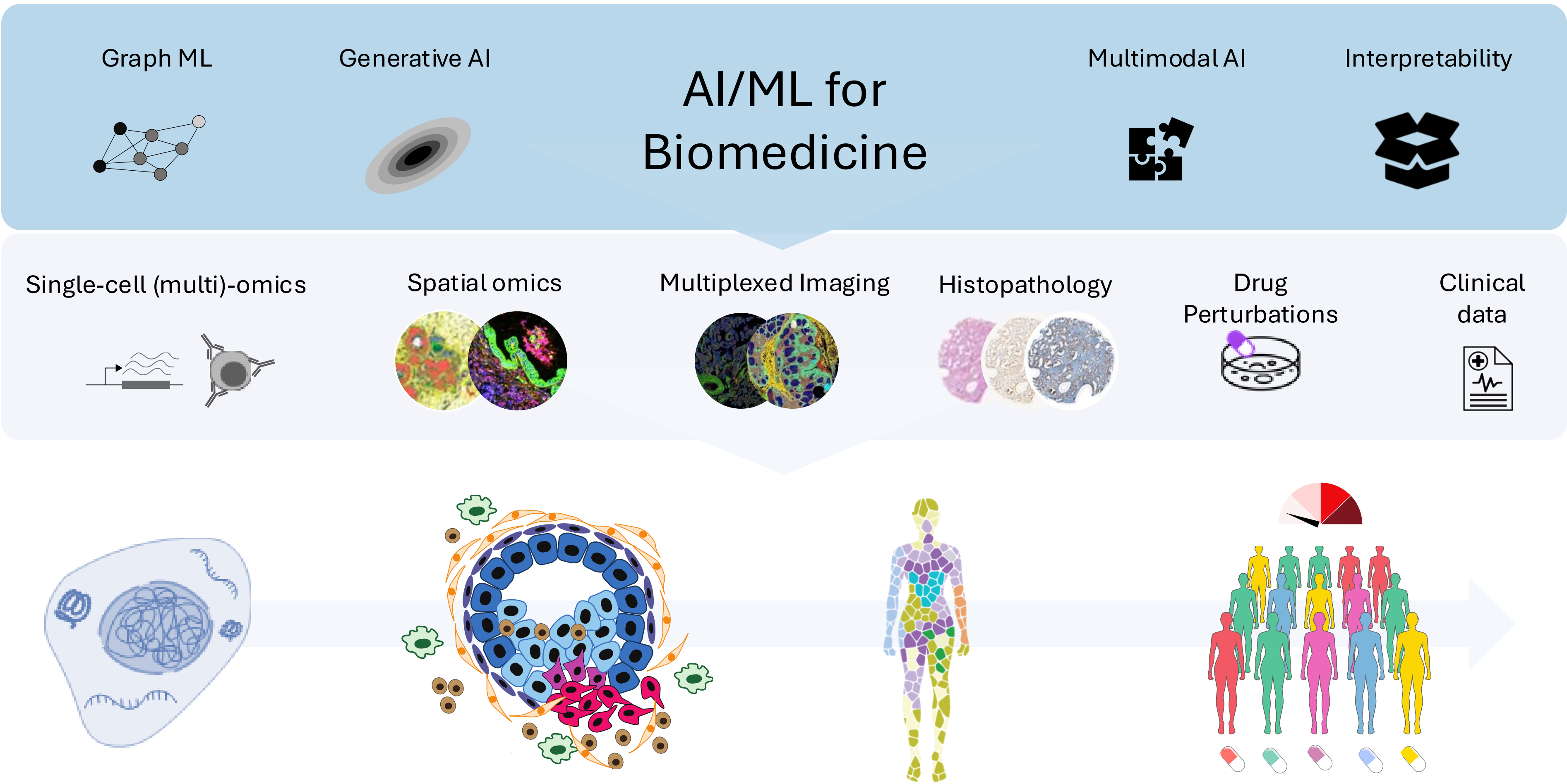
Eling *et al.*, 2019



From stochasticity to tumor heterogeneity

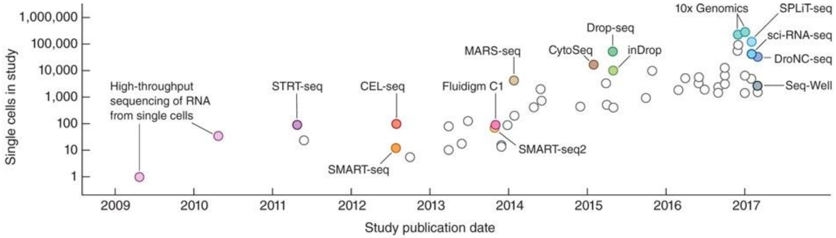


Kashyap et al., *Trends in Biotechnology*, 2022


Graph ML
Generative AI
AI/ML for Biomedicine
Multimodal AI
Interpretability
Single-cell (multi)-omics
Spatial omics
Multiplexed Imaging
Histopathology
Drug Perturbations
Clinical data

An emerging era for single-cell biology

Single-cell omics

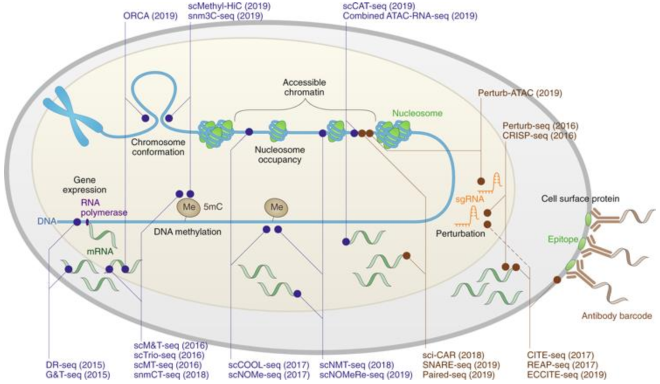


Svensson et al., Nat Protocols, (2018)

An emerging era for single-cell biology

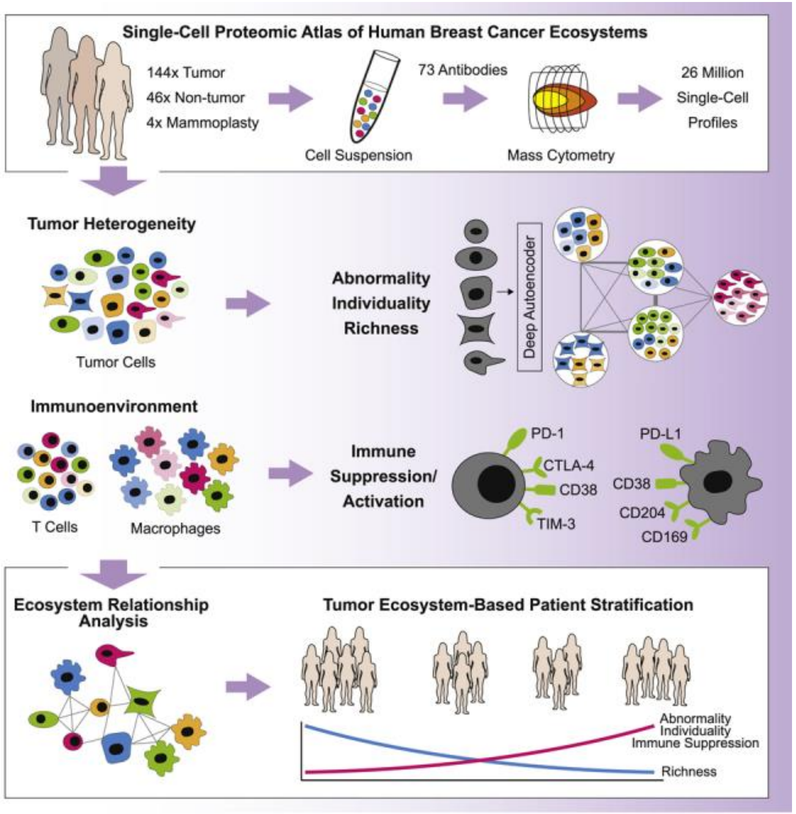
Single-cell omics

Single-cell multiomics



Zhu et al., *Nat Methods*, (2020)

A Single-Cell Atlas of Breast Cancer Ecosystems



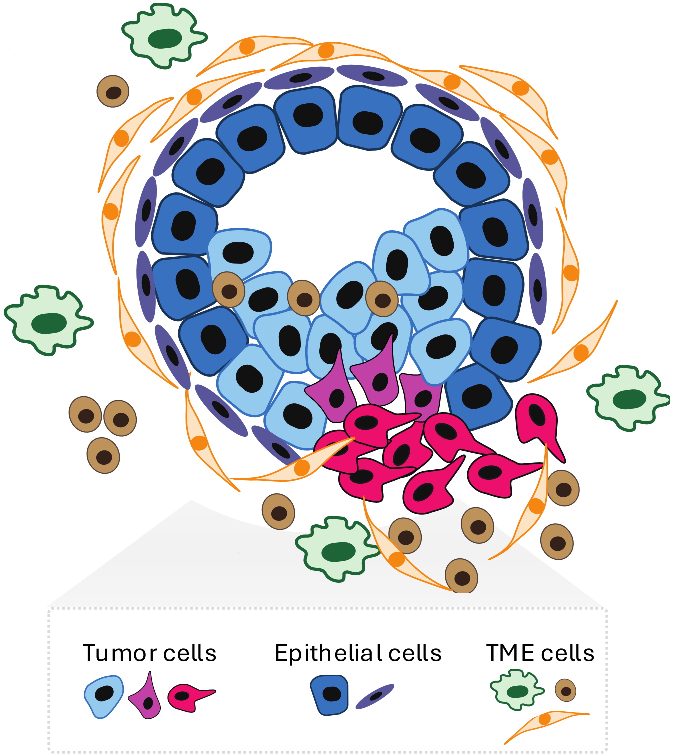
Wagner *et al.*, *Cell,* 2019

Machine Learning-based tumor heterogeneity scores

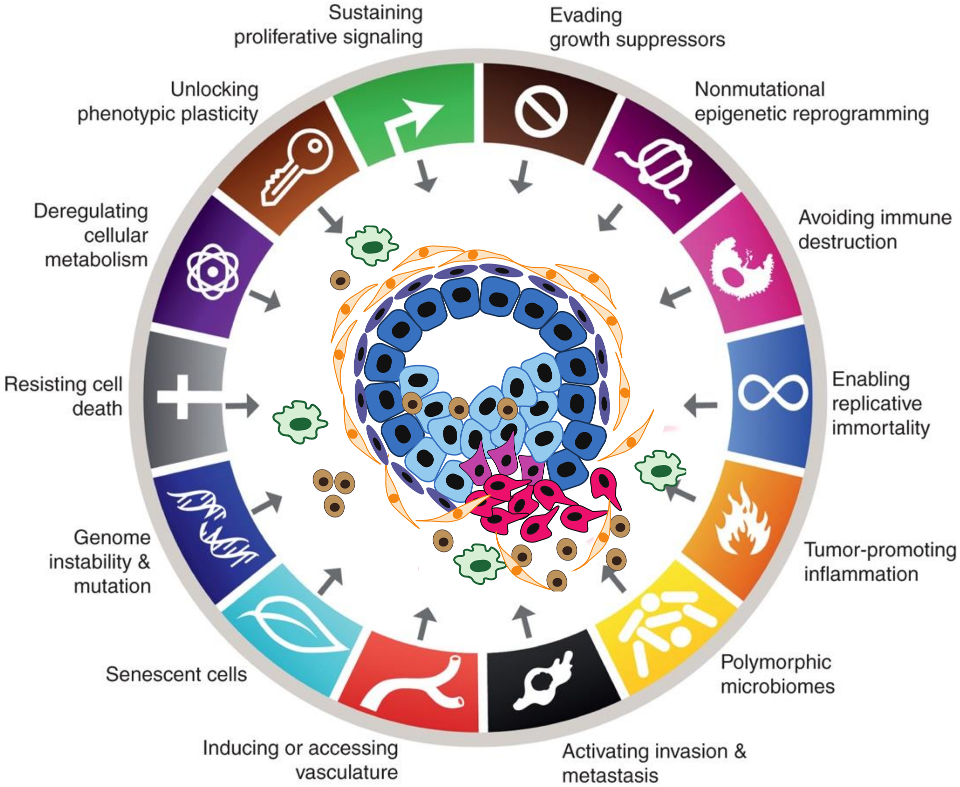
Abnormality	**Individuality**
How much have the tumor cells deviated from non-tumor?	How distinct is the tumor within the cohort?

- Aggressive tumors are characterized by **few but highly abnormal and unique** cell subtypes
- Ecosystem-based approach enables **patient stratification**

Tumors are **spatially** heterogeneous ecosystems

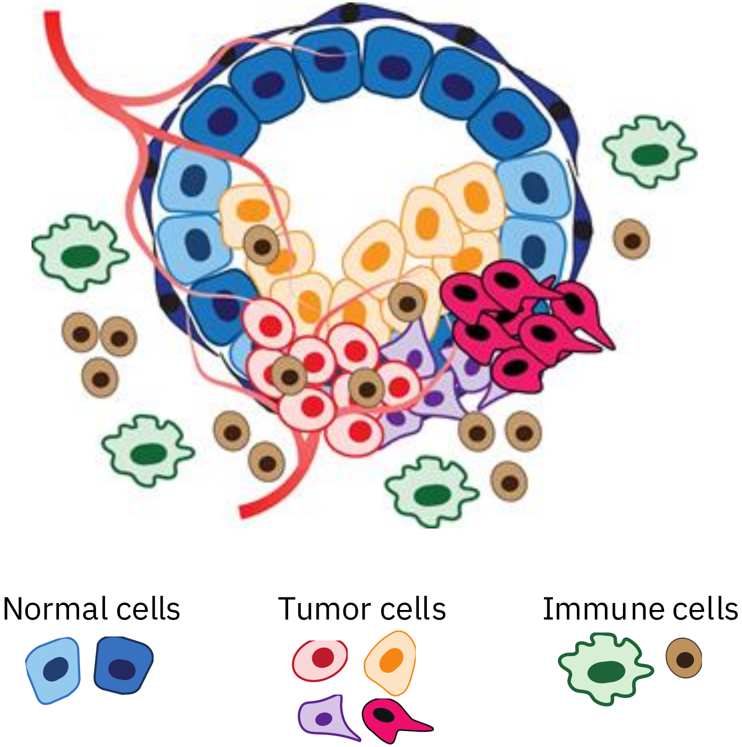


Tumors are **spatially** heterogeneous ecosystems

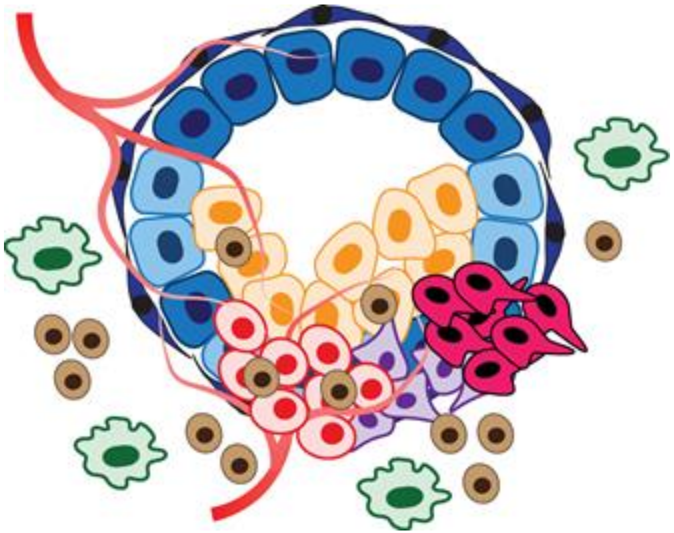


Hanahan *et al.*, 2022

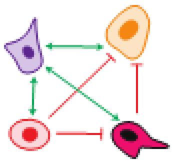
Tumors are **spatially** heterogeneous ecosystems



Tumors are **spatially** heterogeneous ecosystems



Tumor-tumor interactions:



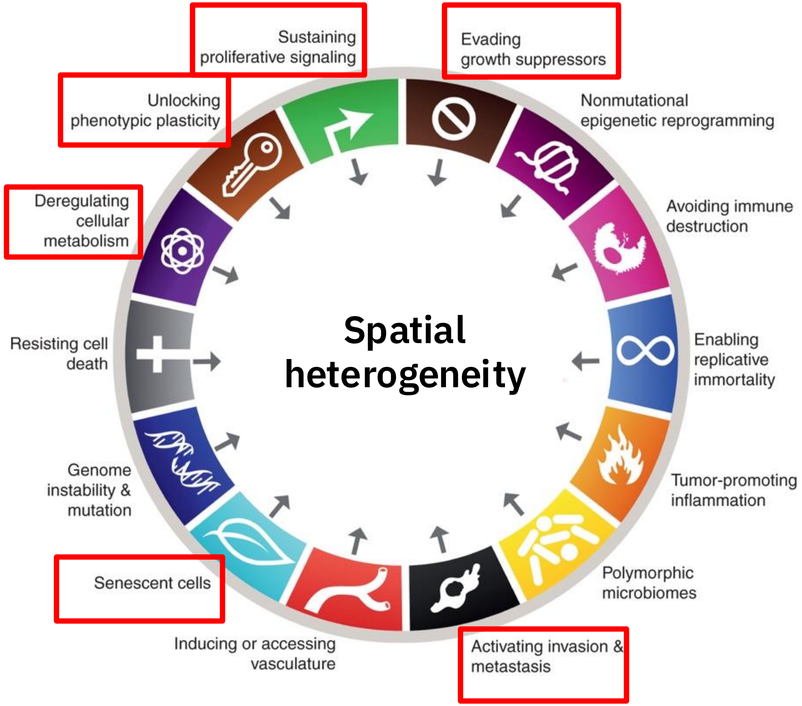
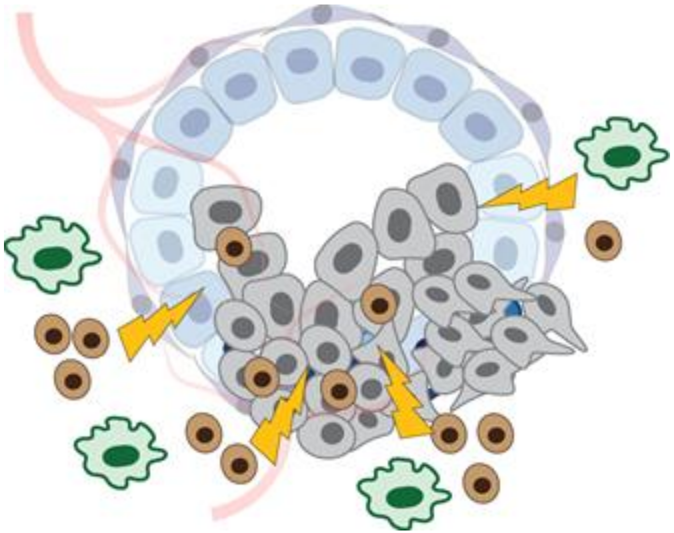
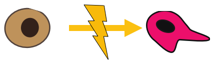


Image from: Hanahan et al., 2022

Tumors are **spatially** heterogeneous ecosystems



Immune attack: 

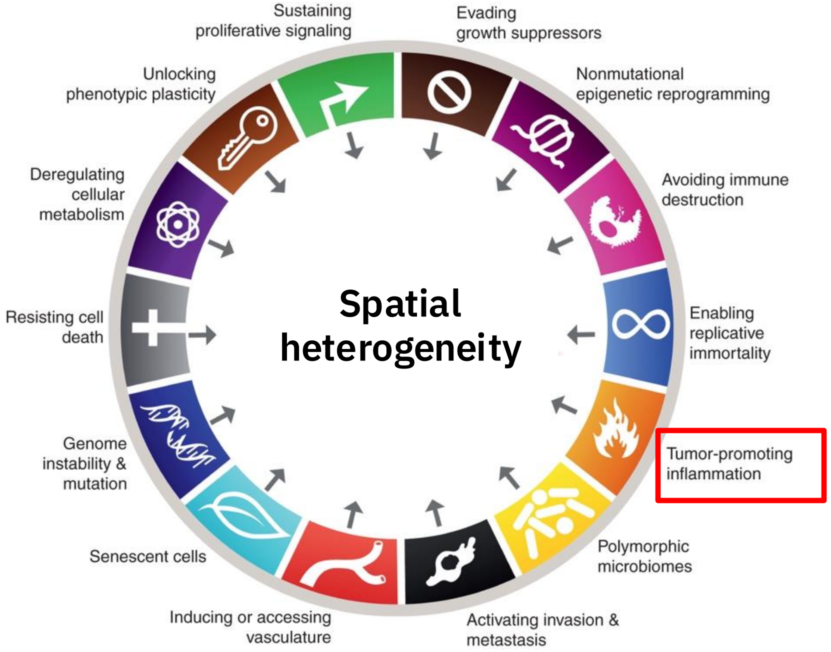
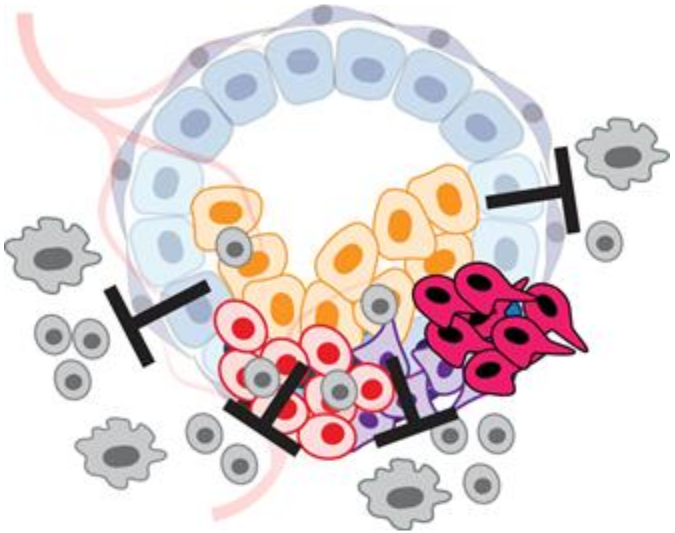


Image from: Hanahan et al., 2022

Tumors are **spatially** heterogeneous ecosystems



Immune suppresion: 

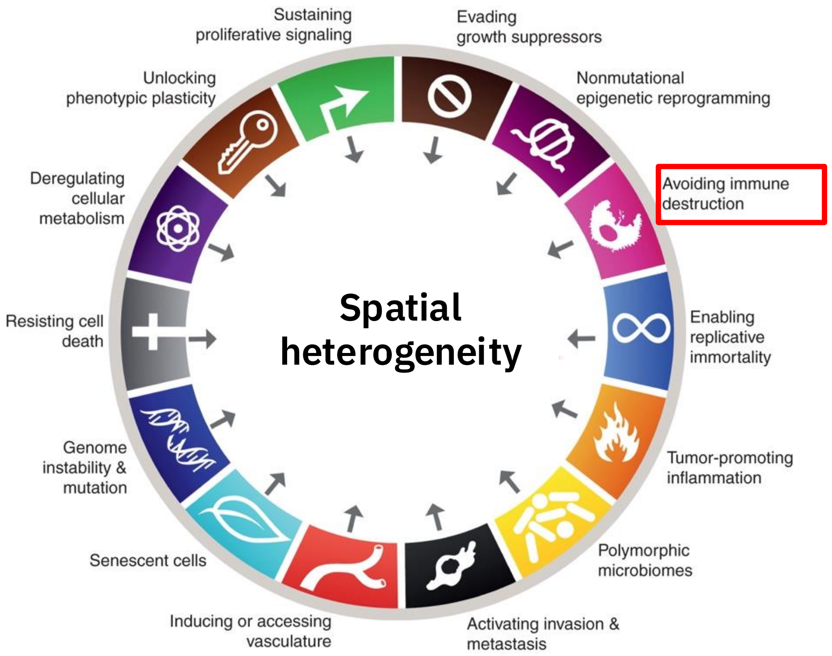
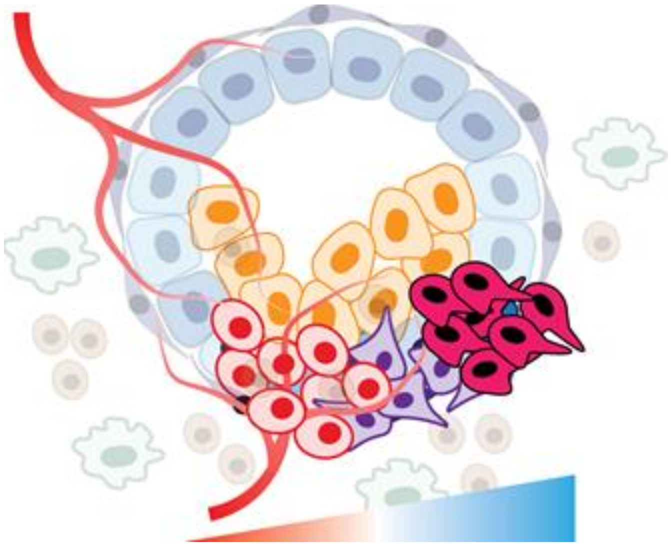


Image from: Hanahan et al., 2022

Tumors are **spatially** heterogeneous ecosystems



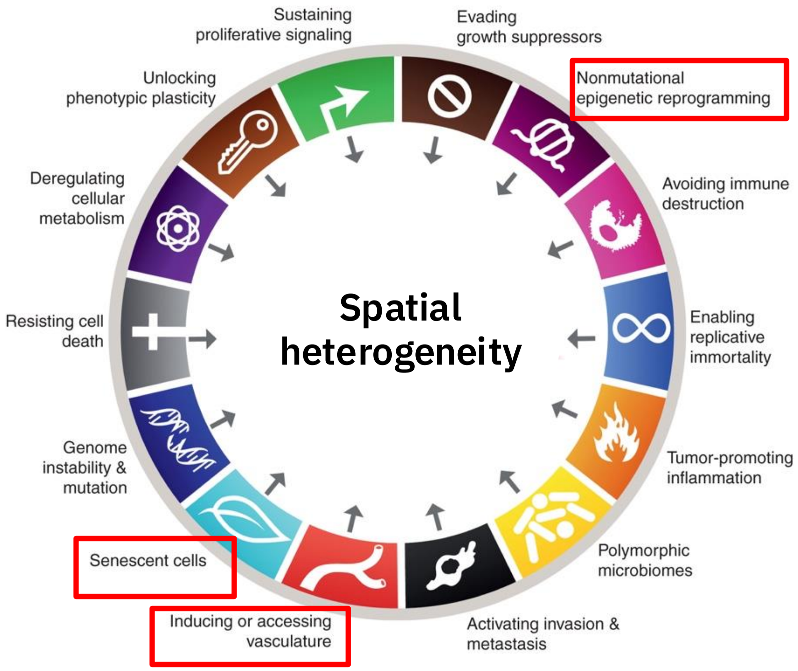
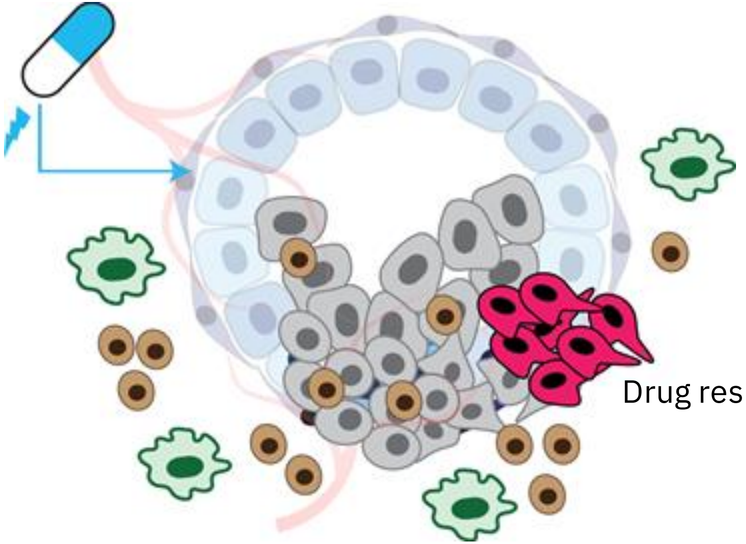


Image from: Hanahan et al., 2022

Tumors are **spatially** heterogeneous ecosystems



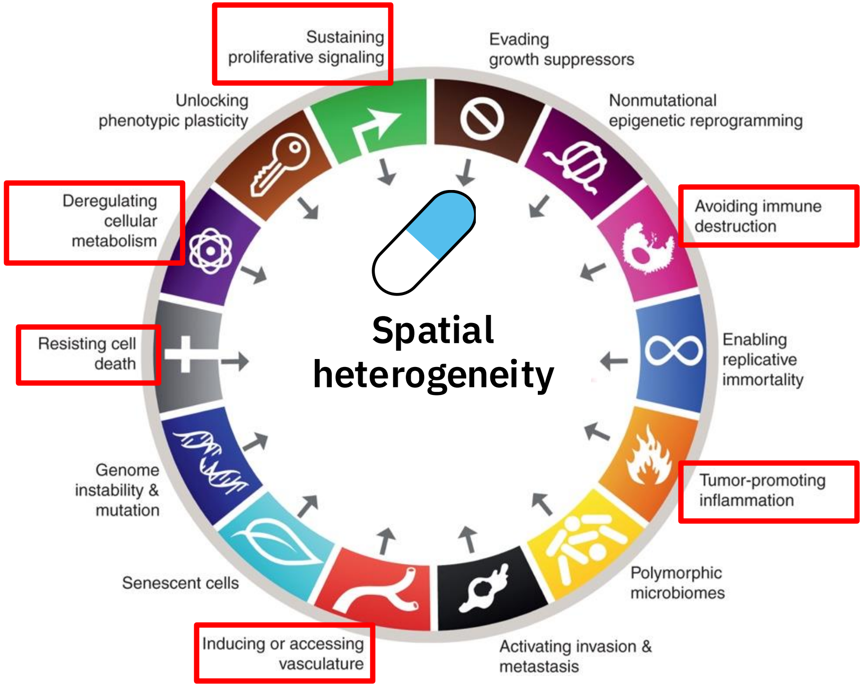


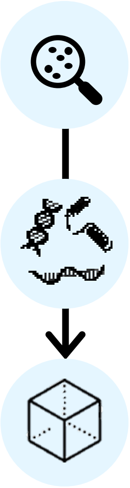
Image from: Hanahan et al., 2022

An emerging era for spatial biology

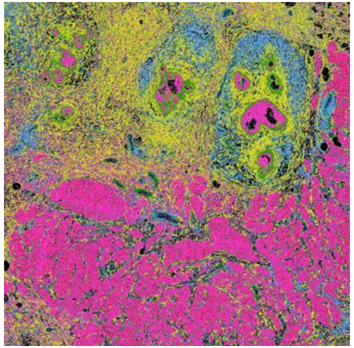
Single-cell omics

Single-cell multiomics

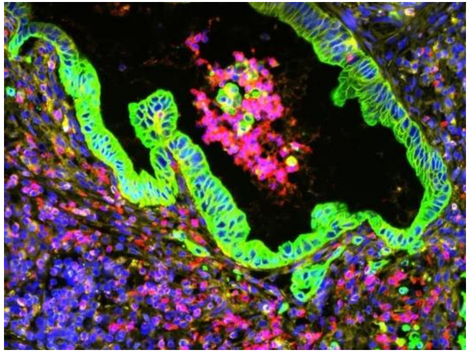
Spatial (single-cell) (multi) omics



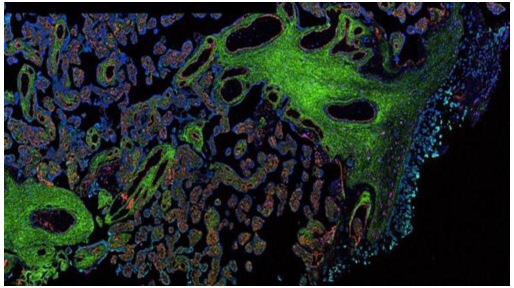
10X Xenium - Breast Cancer



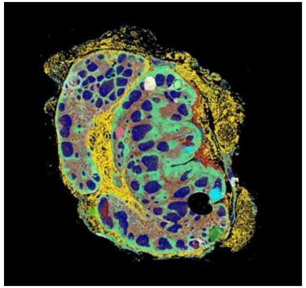
Nanostring CosMx – Human Lung



Imaging Mass Cytometry - Human Placenta



CODEX – Lymph node



Images from https://hubmapconsortium.org, 10Xgenomics.com and Nanostring.com

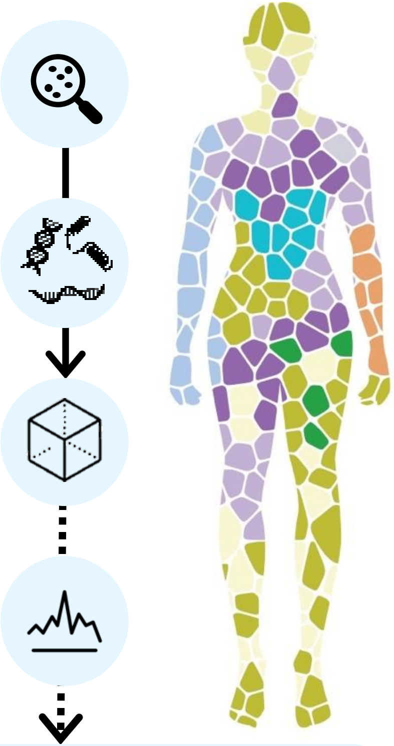
An emerging era for spatial biology

Single-cell omics

Single-cell multiomics

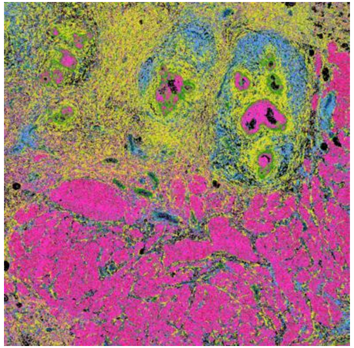
Spatial (single-cell) (multi) omics

Spatio-temporal (single-cell) (multi) omics
Drug or Genetic Perturbations

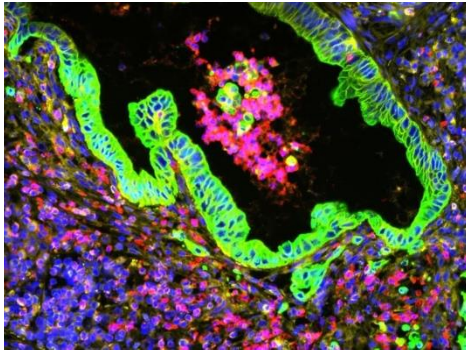


Whole tumors in health and disease, over time or under treatment down to the last single-cell, for millions of cells

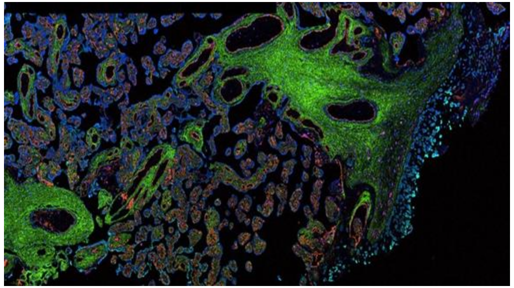
10X Xenium - Breast Cancer



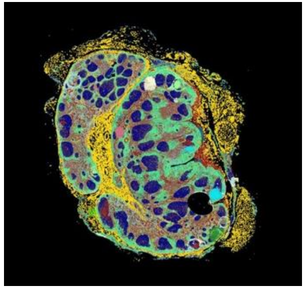
Nanostring CosMx – Human Lung



Imaging Mass Cytometry - Human Placenta

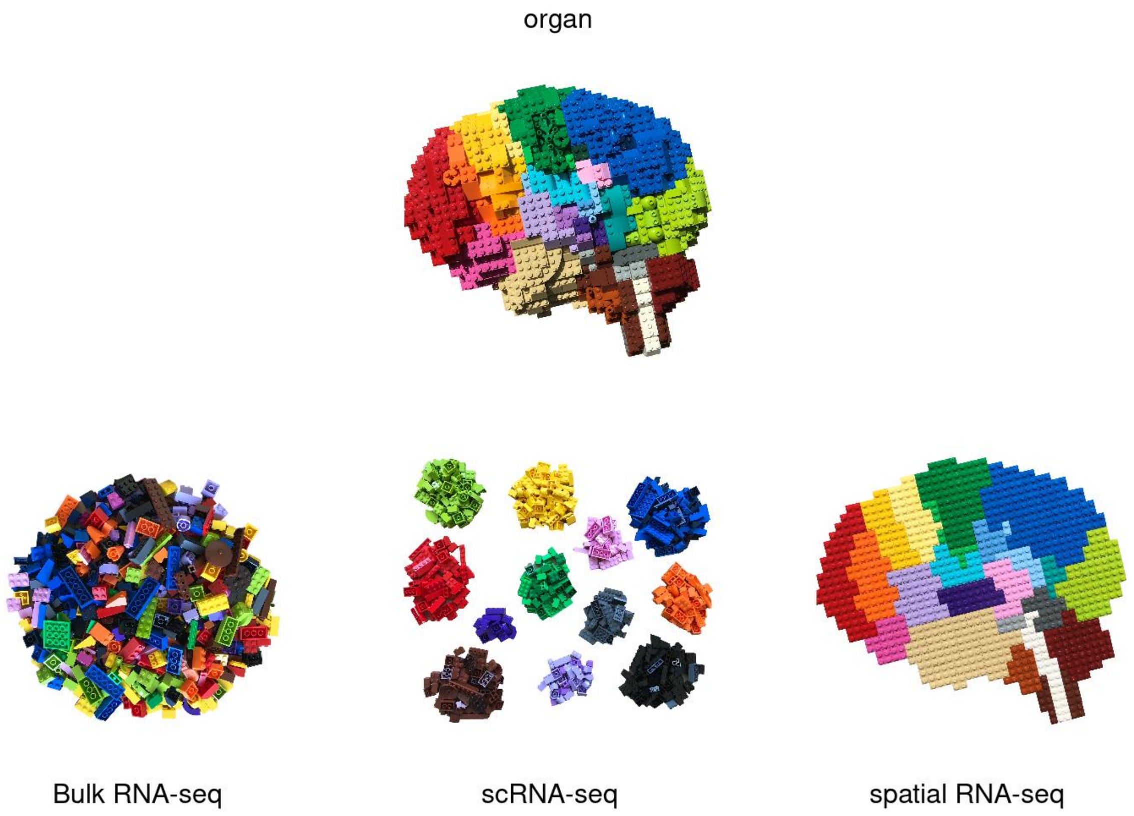


CODEX – Lymph node



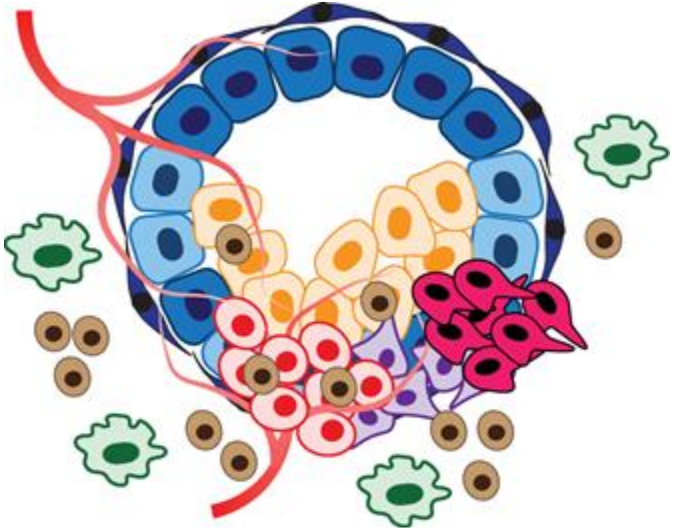
Images from https://hubmapconsortium.org, 10Xgenomics.com and Nanostring.com

Bulk vs. single-cell vs. spatial



https://statomics.github.io/SGA/sequencing_singleCell.html

Spatial omics: the new frontier



Spatial transcriptomics:
e.g., smFISH, MERFISH, seqFISH, LCM, Visium, FISSEQ

Spatial proteomics:
e.g., mIHC, CycIF, CODEX, ImmunoSABER, IMC, MIBI

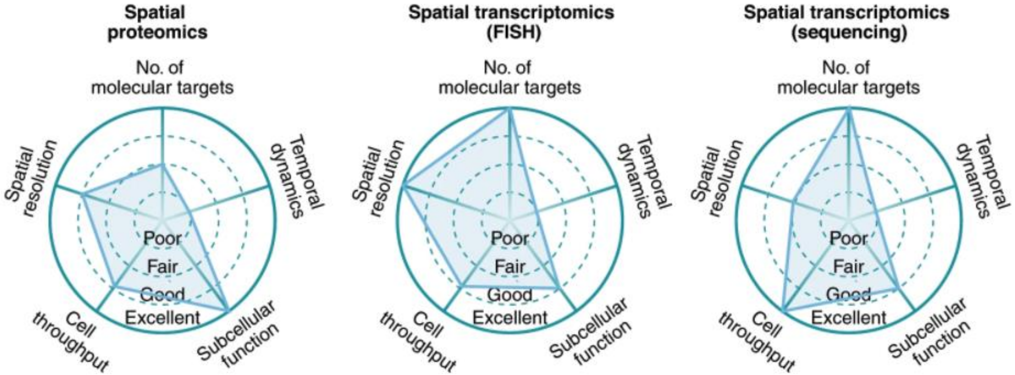
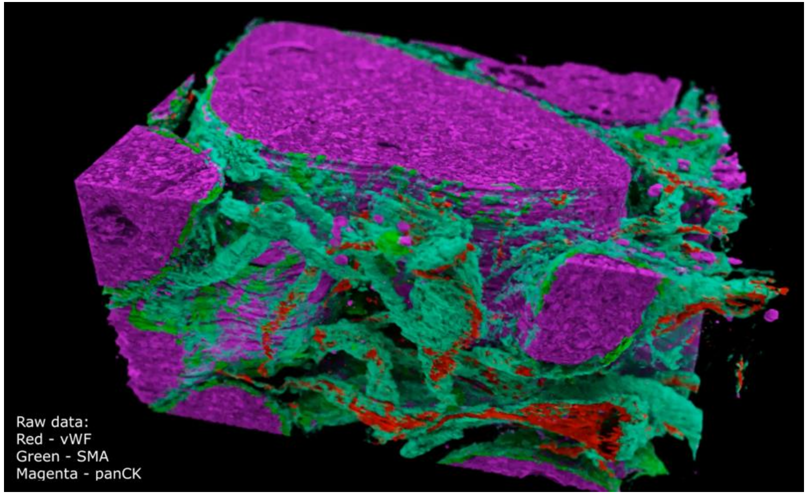


Figure from: Lewis *et al.*, *Nature Methods*, 2021



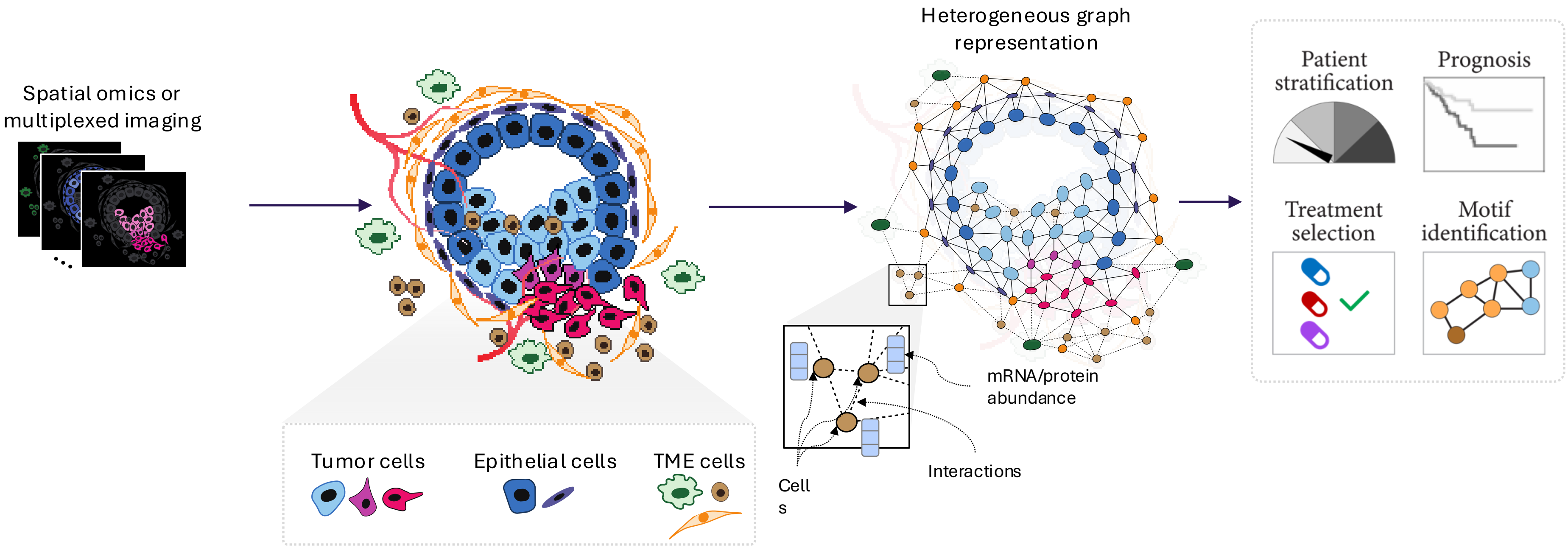
How do cancer subpopulations interact with each other and the TME?

How do they evade immune destruction and treatments?

How does the TME influence tumor growth?

Spatial heterogeneity as an **opportunity** to enable novel biomarker and therapeutic discovery

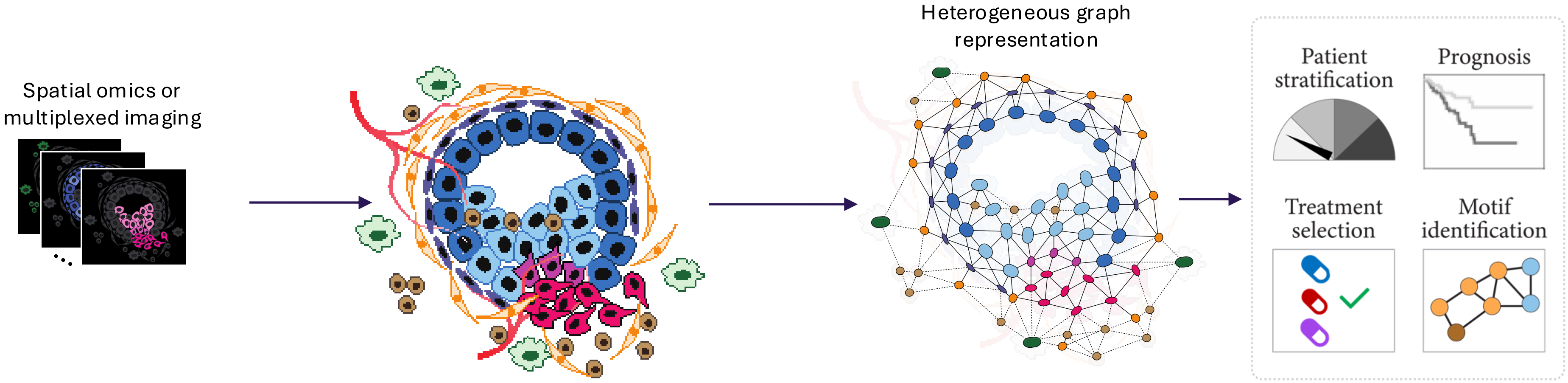
Modeling the TME with **Graph Representation Learning**



ATHENA

Analysis of Tumor Heterogeneity from Spatial Omics

Adriano
Martinelli



Martinelli and Rapsomaniki, *Bioinformatics* (2022)

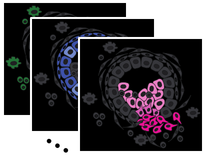
https://github.com/AI4SCR/ATHENA

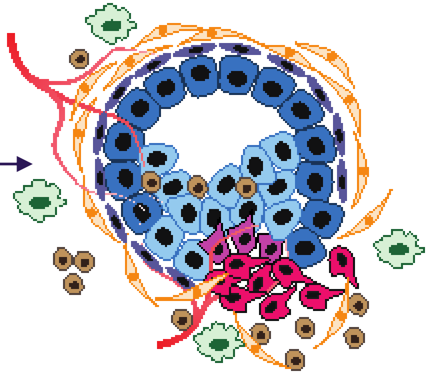
ATHENA

Analysis of Tumor Heterogeneity from Spatial Omics

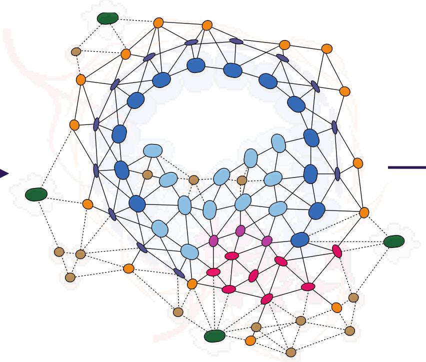
Adriano Martinelli

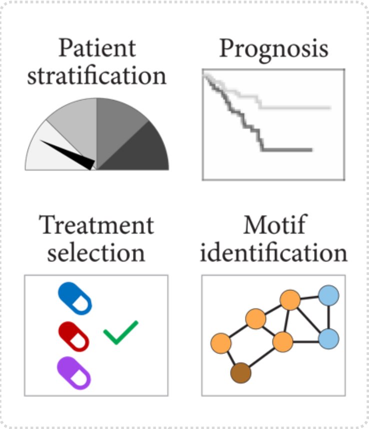
Spatial omics or multiplexed imaging

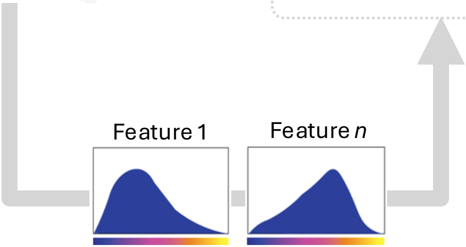




Heterogeneous graph representation







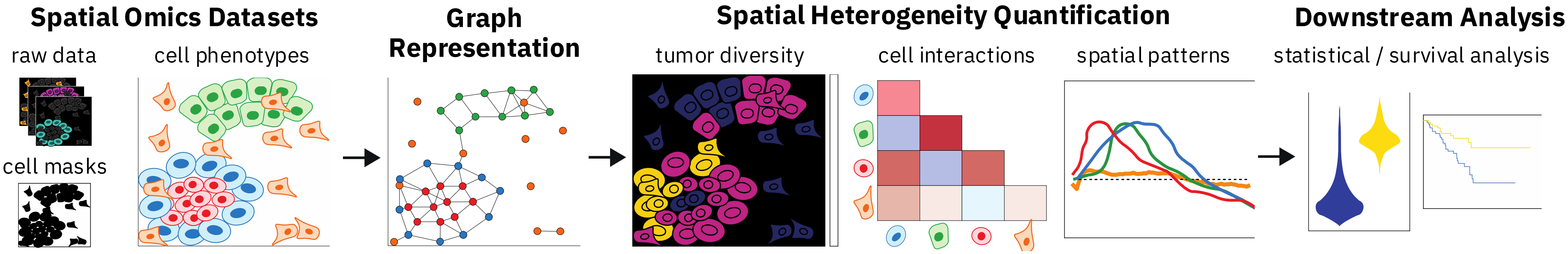
Interpretable spatial heterogeneity scores

Martinelli and Rapsomaniki, *Bioinformatics* (2022)

https://github.com/AI4SCR/ATHENA

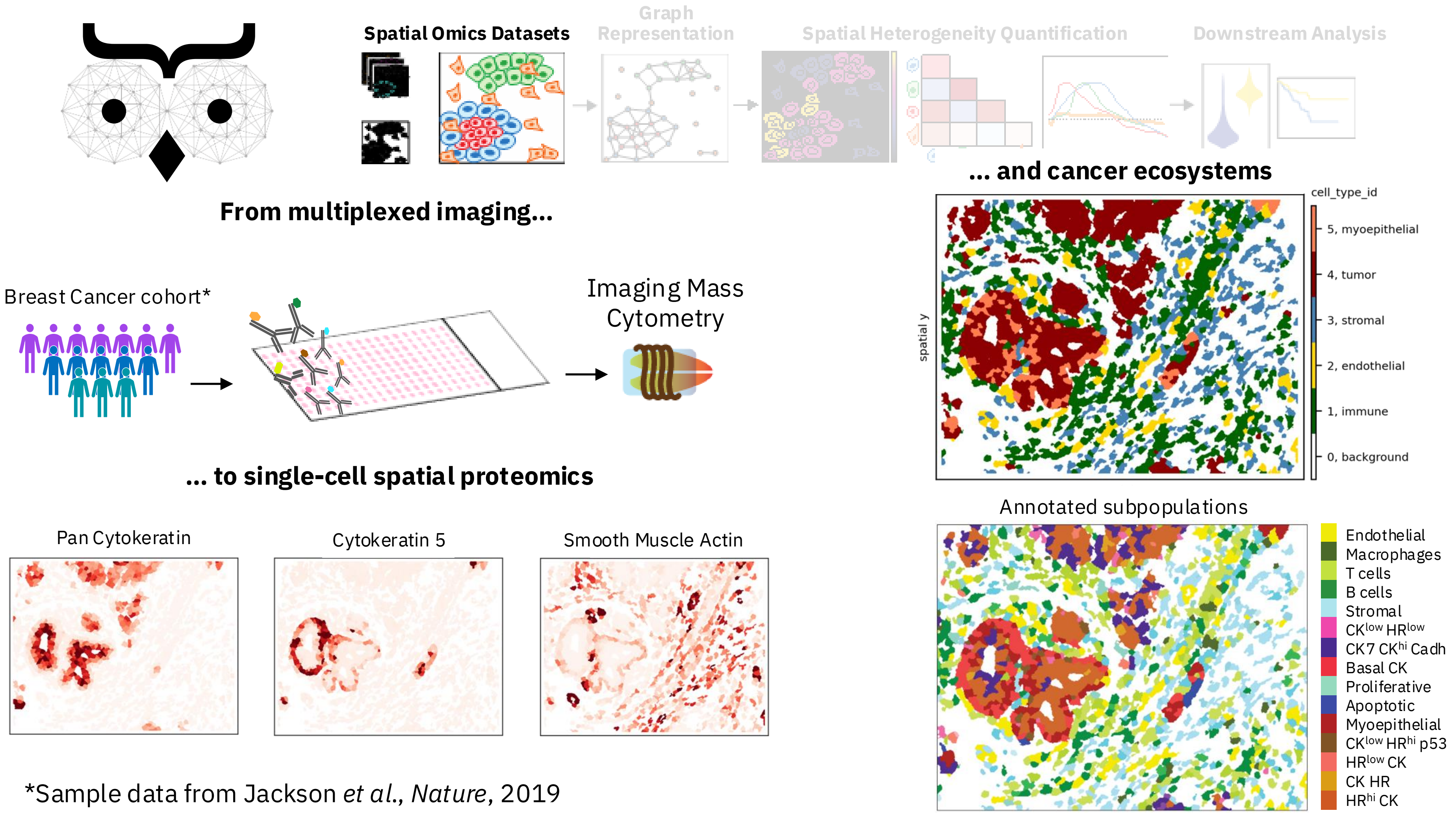
ATHENA

Analysis of Tumor Heterogeneity from Spatial Omics

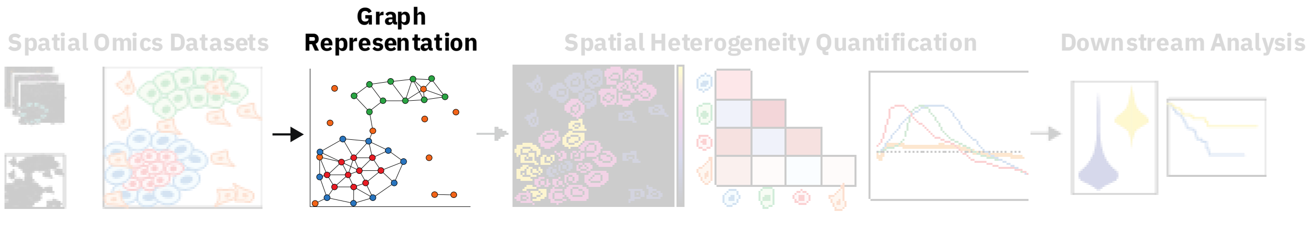


Martinelli and Rapsomaniki (2022) ATHENA: analysis of tumor heterogeneity from spatial omics measurements, *Bioinformatics*, doi: [10.1093/bioinformatics/btac303](https://doi.org/10.1093/bioinformatics/btac303)

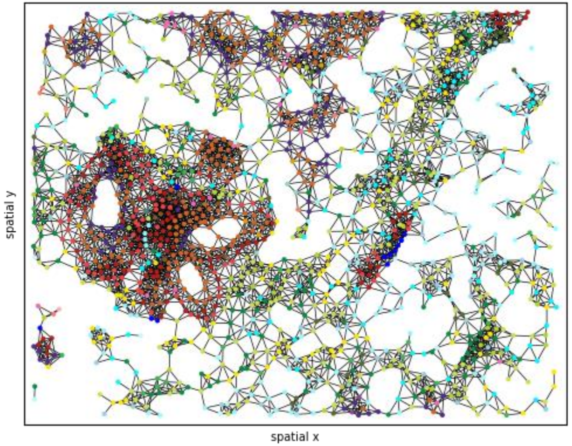
https://github.com/AI4SCR/ATHENA



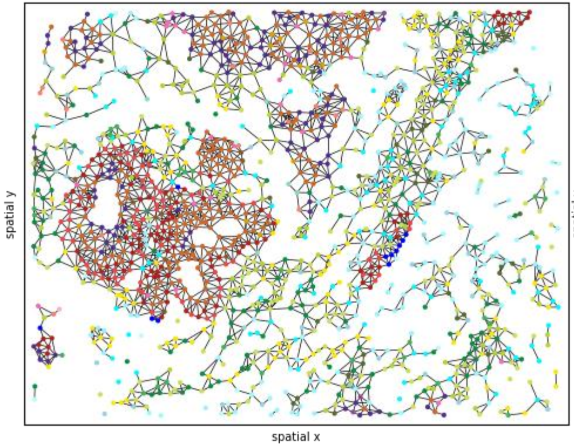
Spatial Omics Datasets
Graph Representation
Spatial Heterogeneity Quantification
Downstream Analysis
From multiplexed imaging...
Breast Cancer cohort*
Imaging Mass Cytometry
... to single-cell spatial proteomics
Pan Cytokeratin
Cytokeratin 5
Smooth Muscle Actin
*Sample data from Jackson et al., Nature, 2019
... and cancer ecosystems
spatial y
cell_type_id
5, myoepithelial
4, tumor
3, stromal
2, endothelial
1, immune
0, background
Annotated subpopulations
Endothelial
Macrophages
T cells
B cells
Stromal
CK^{low} HR^{low}
CK7 CK^{hi} Cadh
Basal CK
Proliferative
Apoptotic
Myoepithelial
CK^{low} HR^{hi} p53
HR^{low} CK
CK HR
HR^{hi} CK



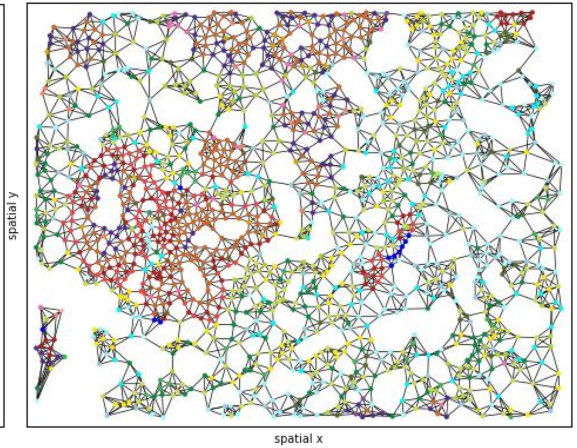
Radius Graph

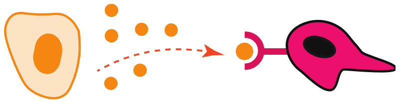


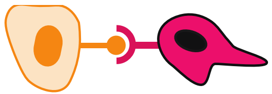
Contact Graph

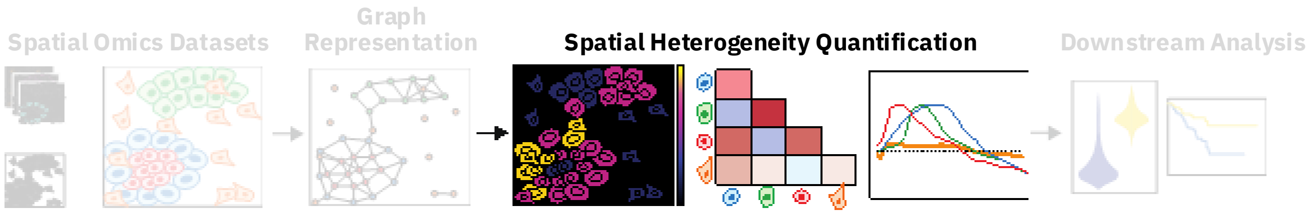


k-NN Graph





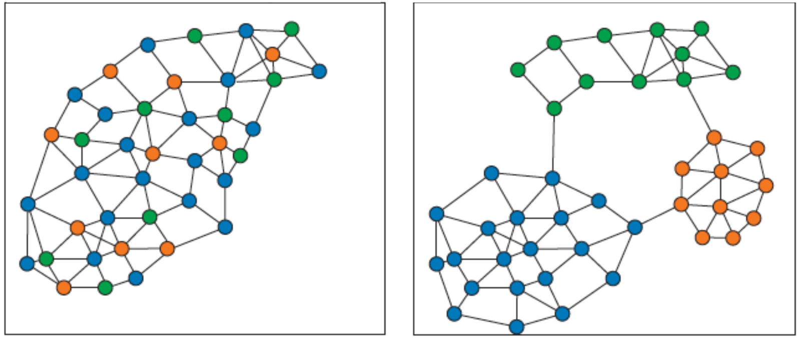


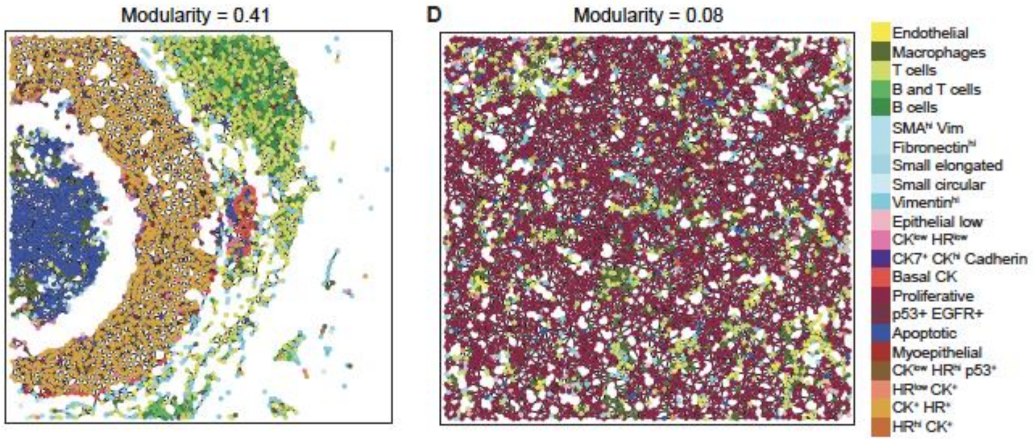


Graph-theoretic scores

low modularity

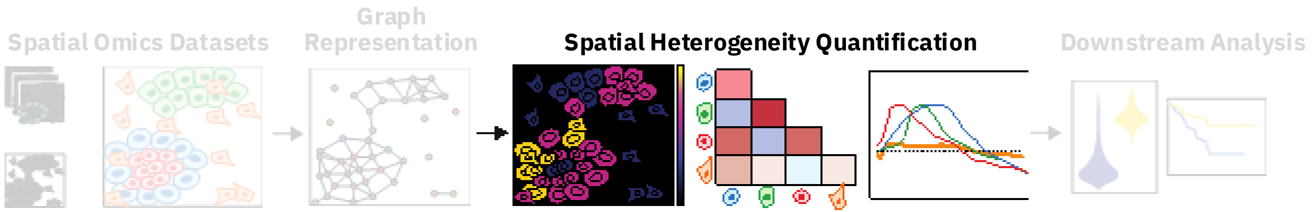
high modularity



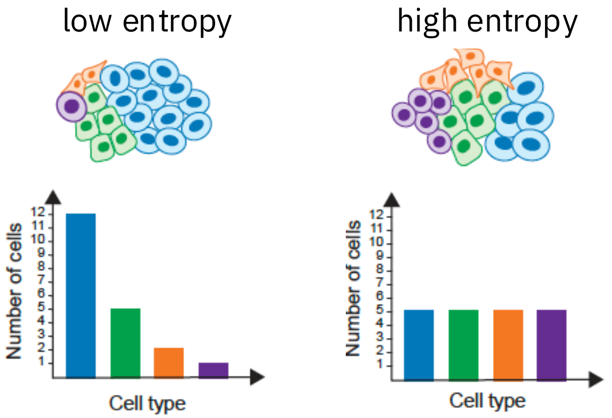


Modularity: $$Q = \sum_{c=1}^{n}\left[\frac{L_c}{m} - \gamma\left(\frac{k_c}{2m}\right)^2\right]$$

where m: #connections, γ: resolution, L_c: inter-community edges, k_c: sum of degrees of all nodes in community c

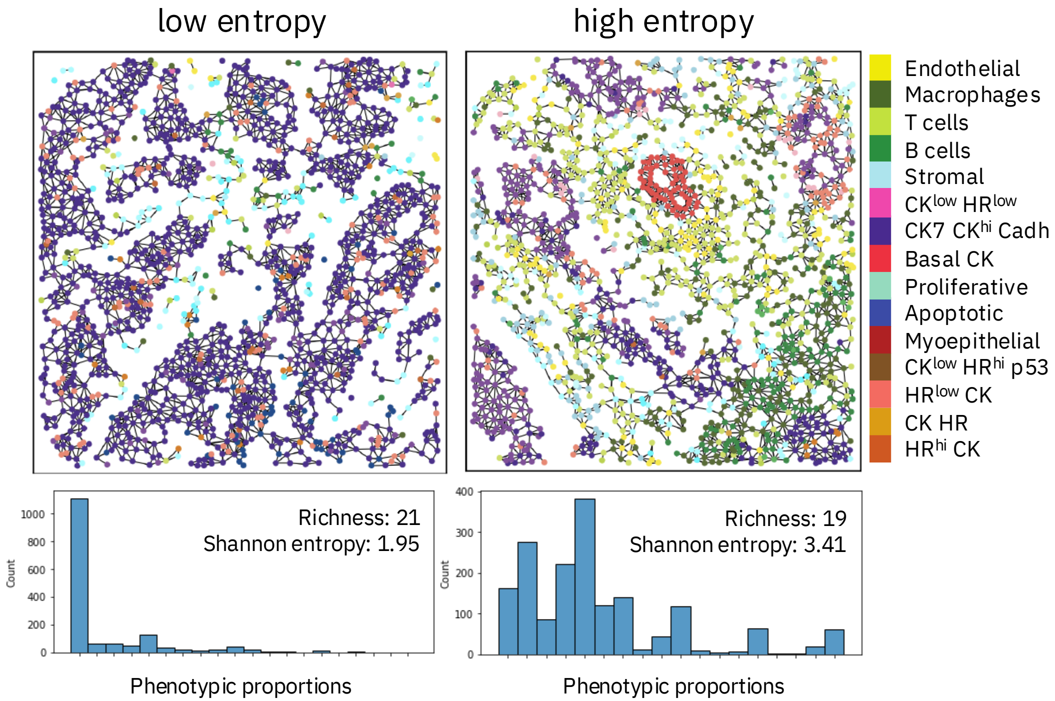


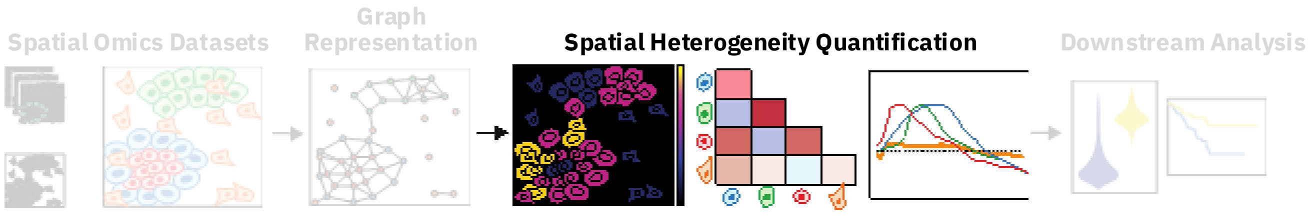
Information-theoretic scores



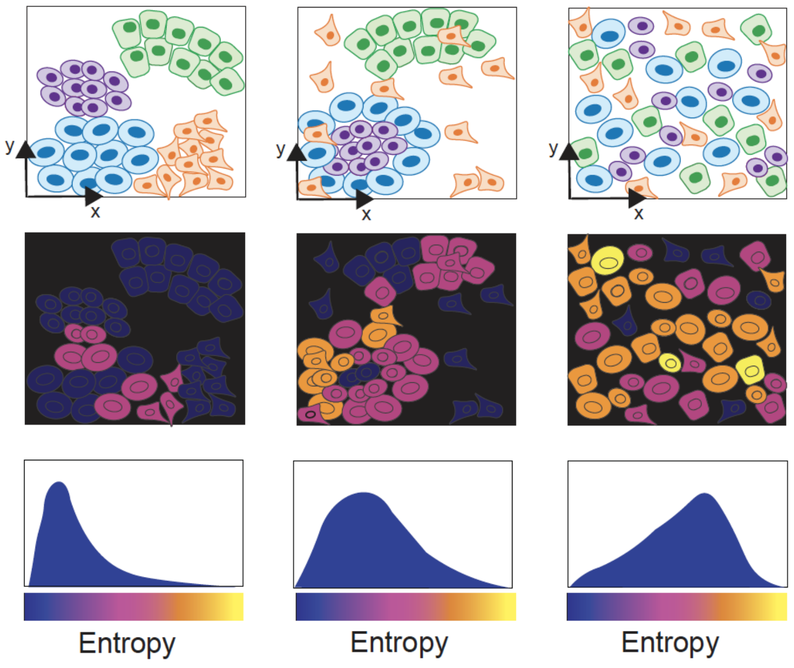
Shannon entropy: $$-\sum_{i=1}^{S} p(x_i)\log p(x_i)$$

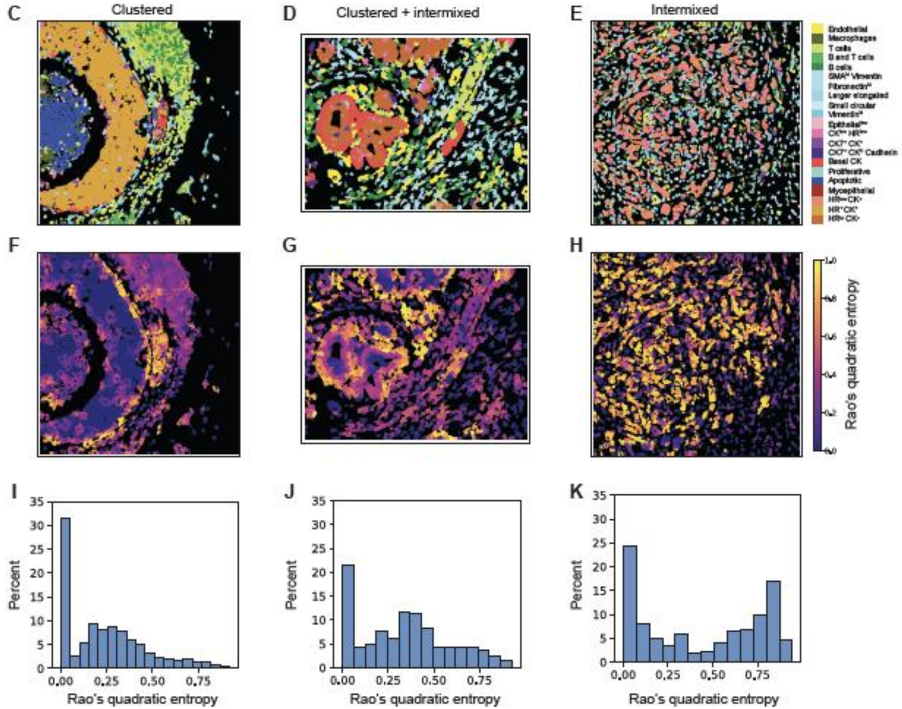
Quadratic entropy: $$\sum_{i=1}^{S-1}\sum_{j=i+1}^{S} d(x_i,x_j)p(x_i)p(x_j)$$

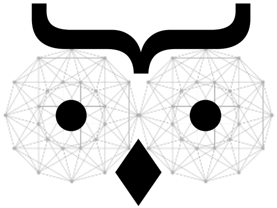


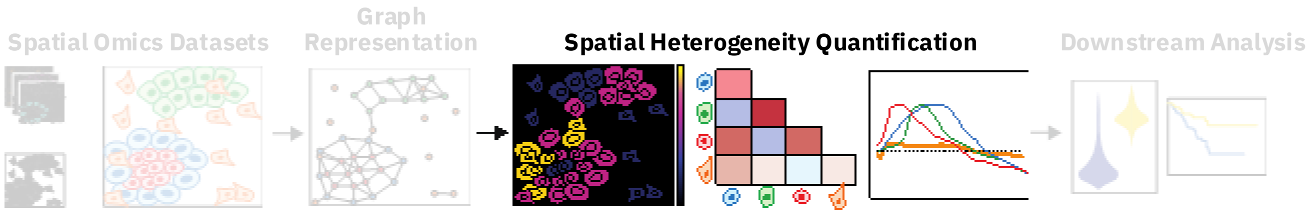


Local information-theoretic scores

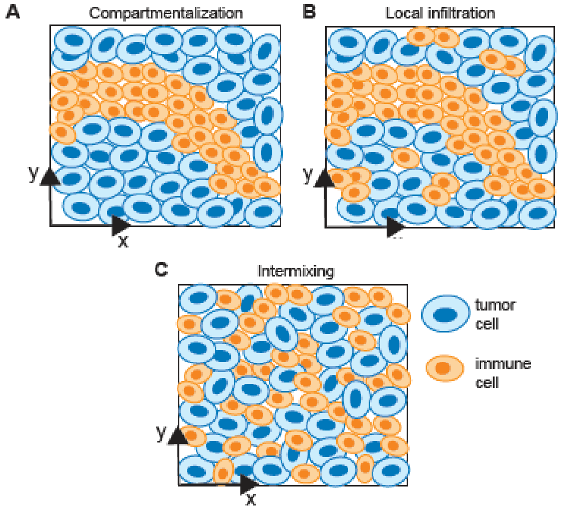


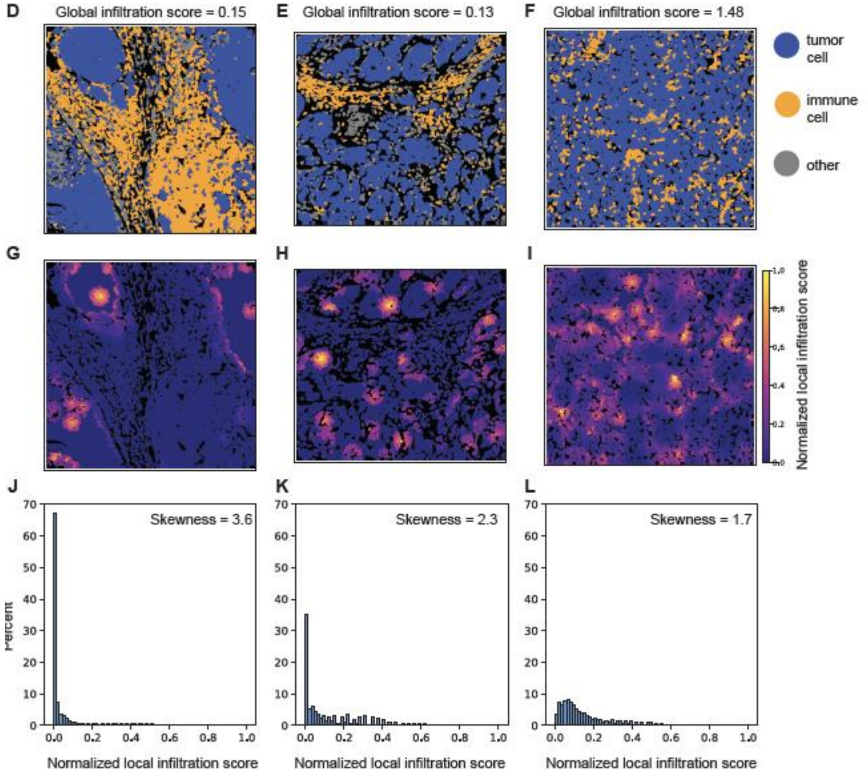






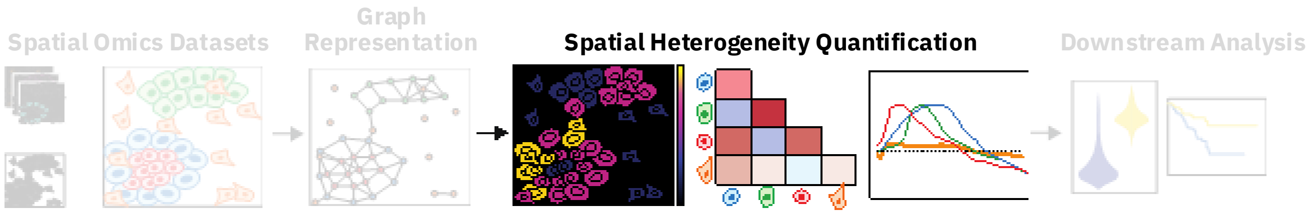
Immune infiltration scores



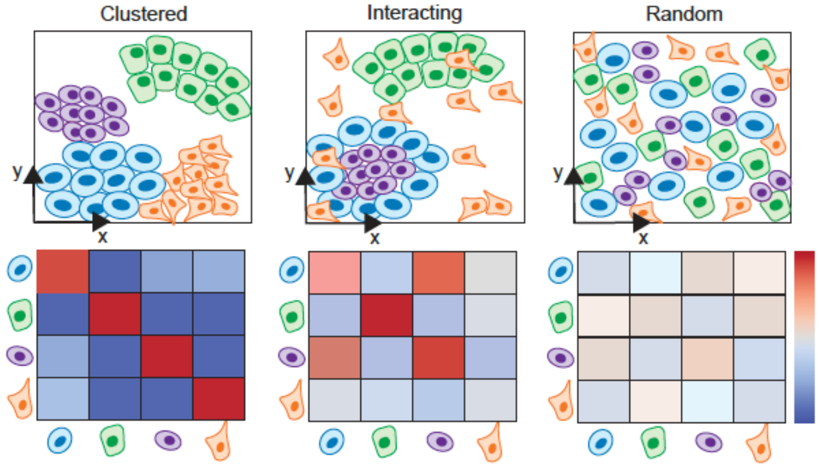


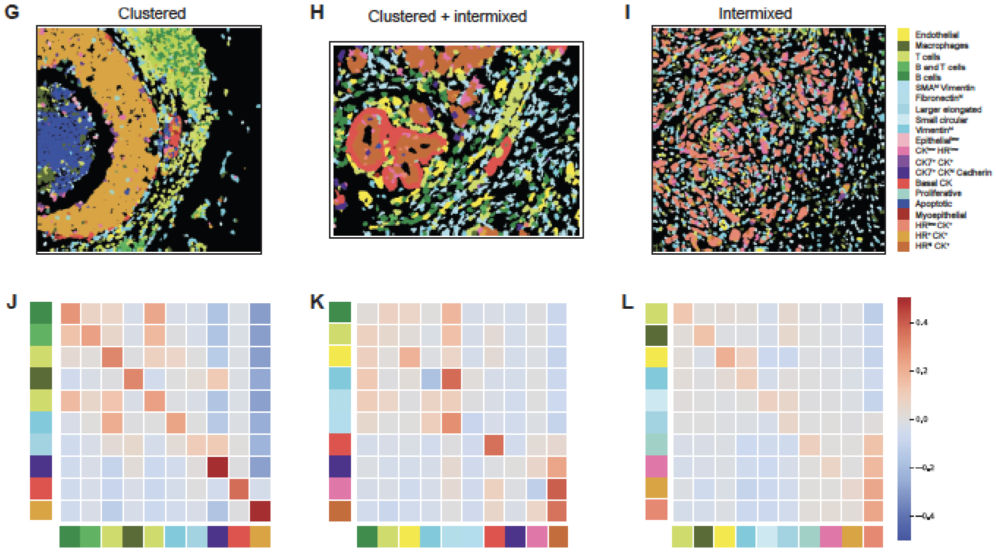
Infiltration score: $$\frac{\text{tumor}-\text{immune interactions}}{\text{immune}-\text{immune interactions}}$$

Keren et al., *Cell*, 2018



Cell interaction scores





Neighborhood analysis - Spatial enrichment methods

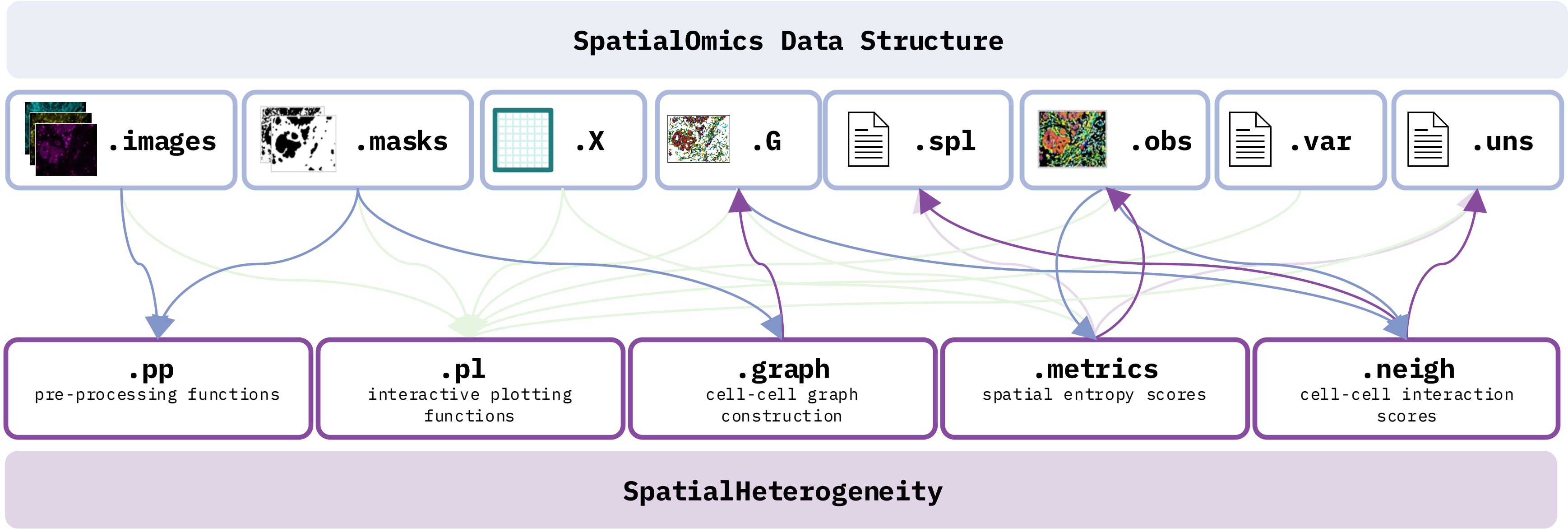
Schapiro et al., *Nature Methods*, 2017
Keren et al., *Cell*, 2018

ATHENA

Implementation

Adriano Martinelli

Theo Maffei



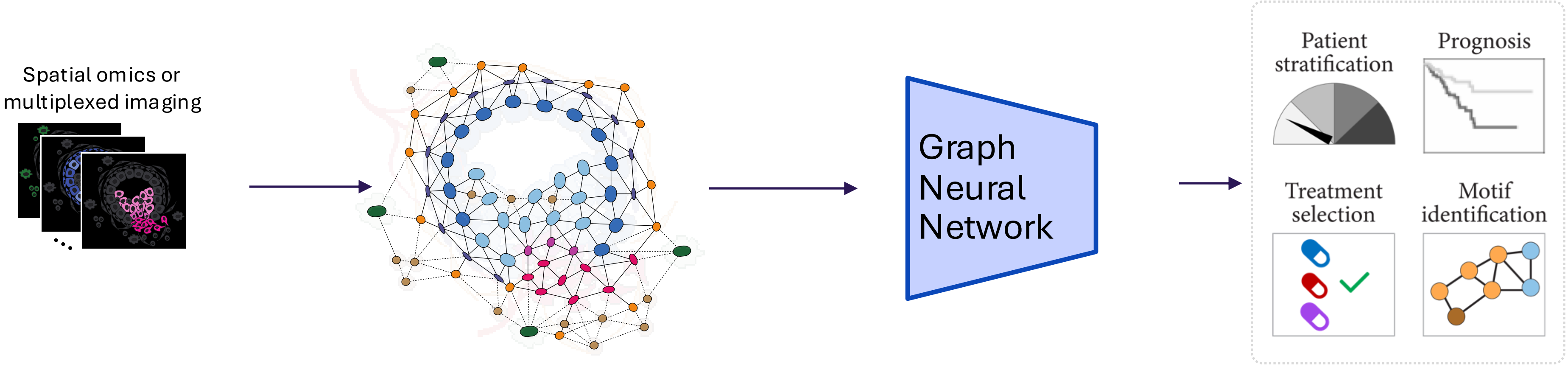
Code, tutorials, datasets:

https://github.com/AI4SCR/ATHENA

Compatible with popular analysis frameworks

squidpy NetworkX Network Analysis in Python seaborn

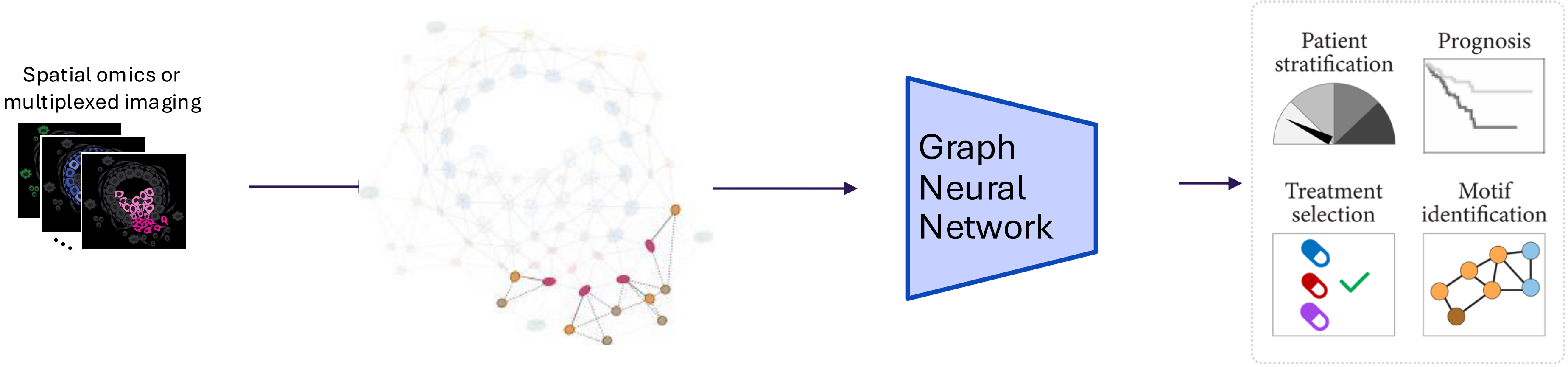
Modeling the TME with Graph Concept Learning



- Limited interpetability
- Is the model able to pick up the signal in the noisy data?

Santiago Castro Dau

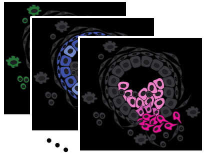
Modeling the TME with Graph Concept Learning

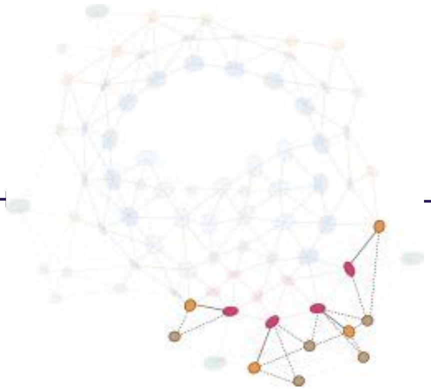


- What if we focused on the relevant part of the graph?

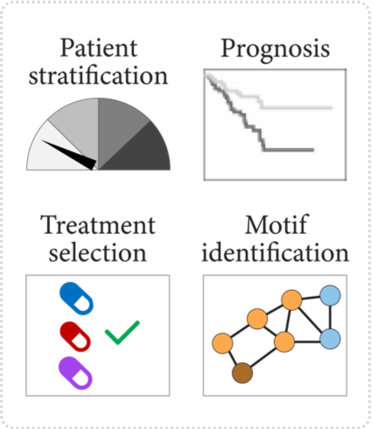
Modeling the TME with Graph Concept Learning

Spatial omics or multiplexed imaging





Graph Neural Network



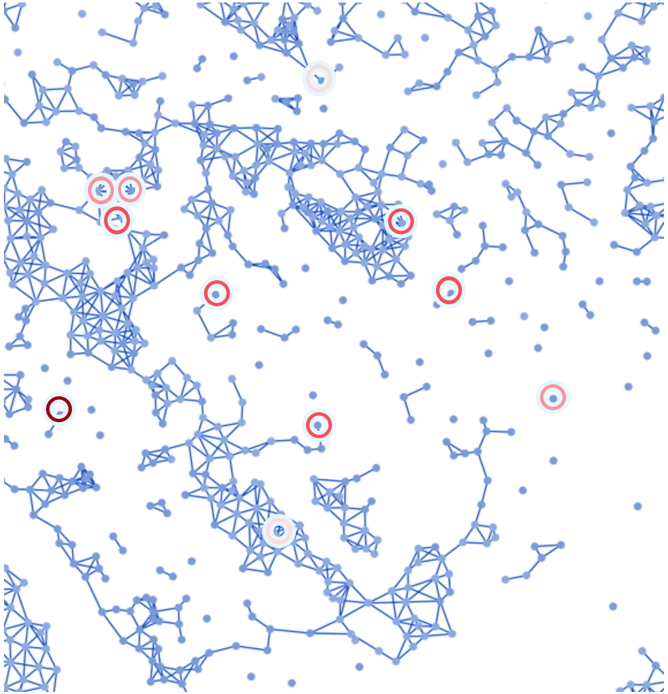
+

Attention scores

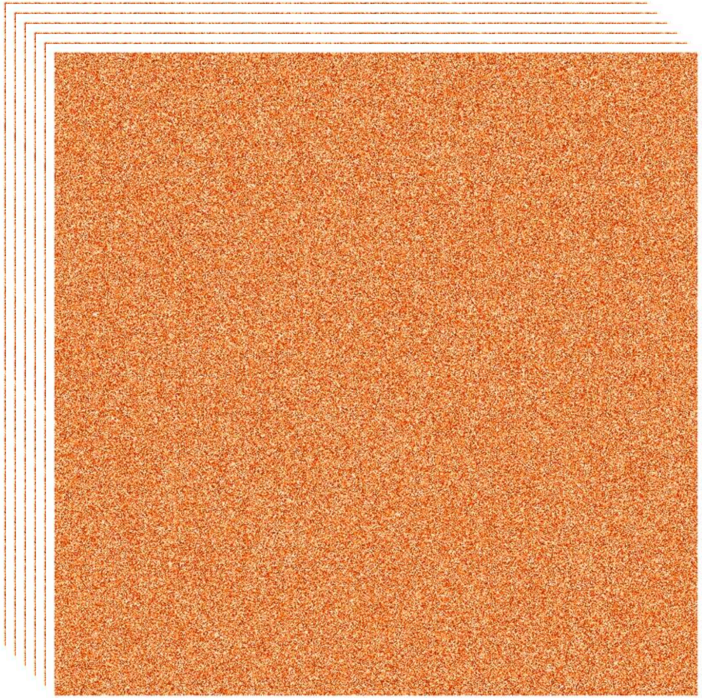
- Attention as interpretation

Limited Interpretability

Simple attention:

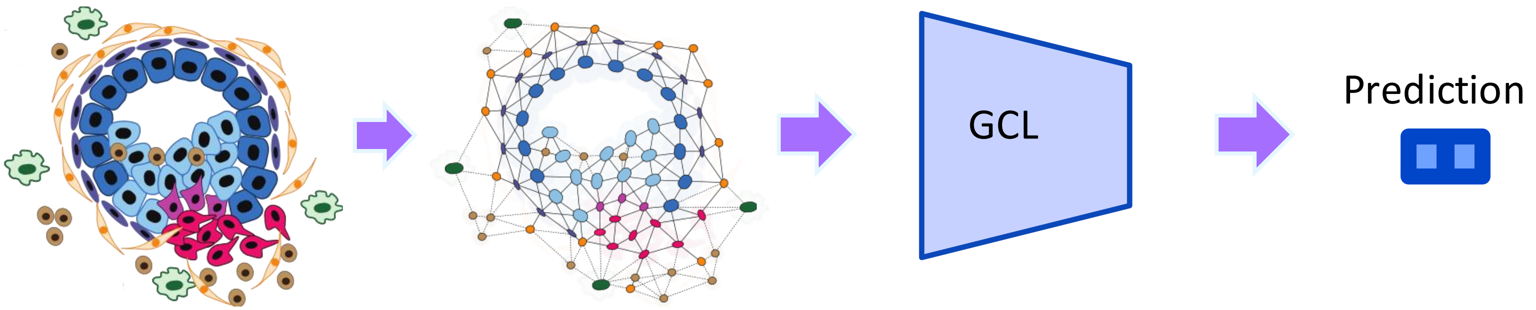


Multi-head self attention:



The Graph Concept Learner

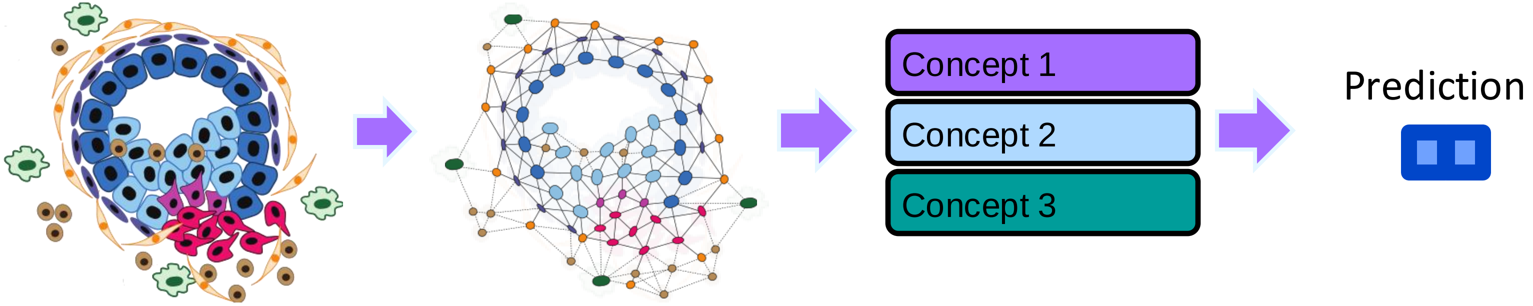
A Novel Machine Learning Framework



- Mode for inclusion of prior knowledge
- Improved interpretability

The Graph Concept Learner

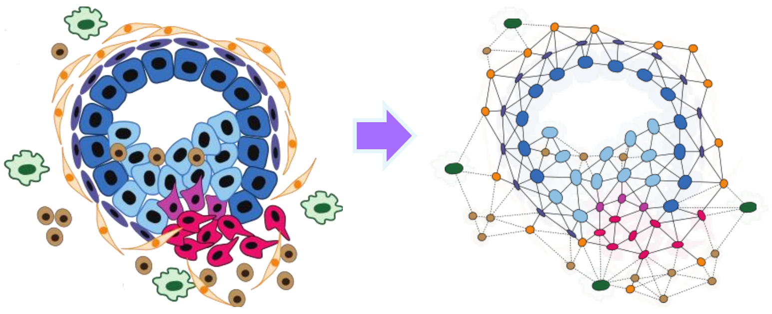
A Novel Machine Learning Framework

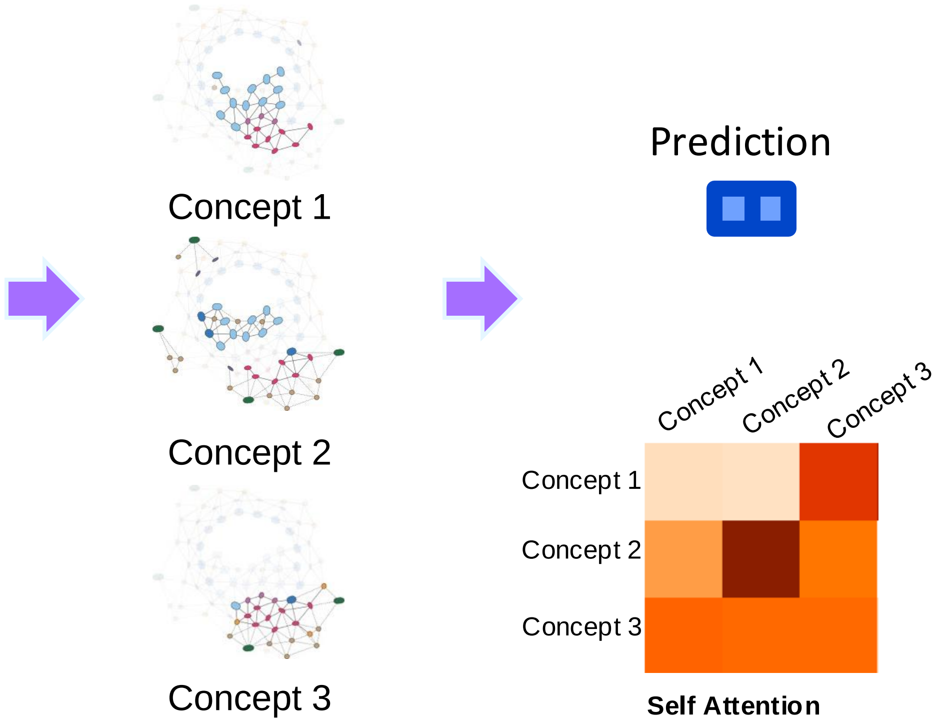


- Mode for inclusion of prior knowledge
- Improved interpretability

The Graph Concept Learner

A Novel Machine Learning Framework

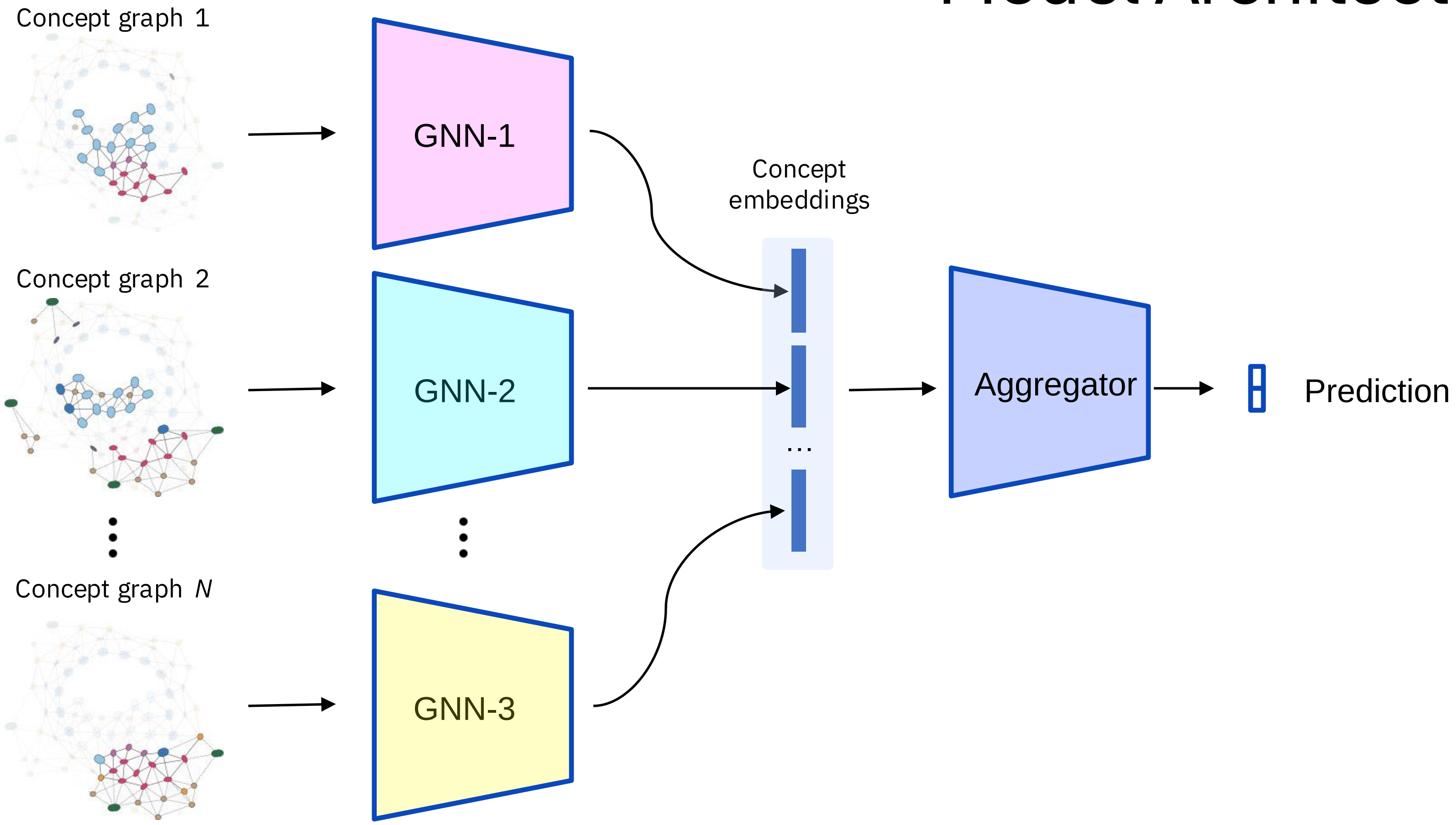




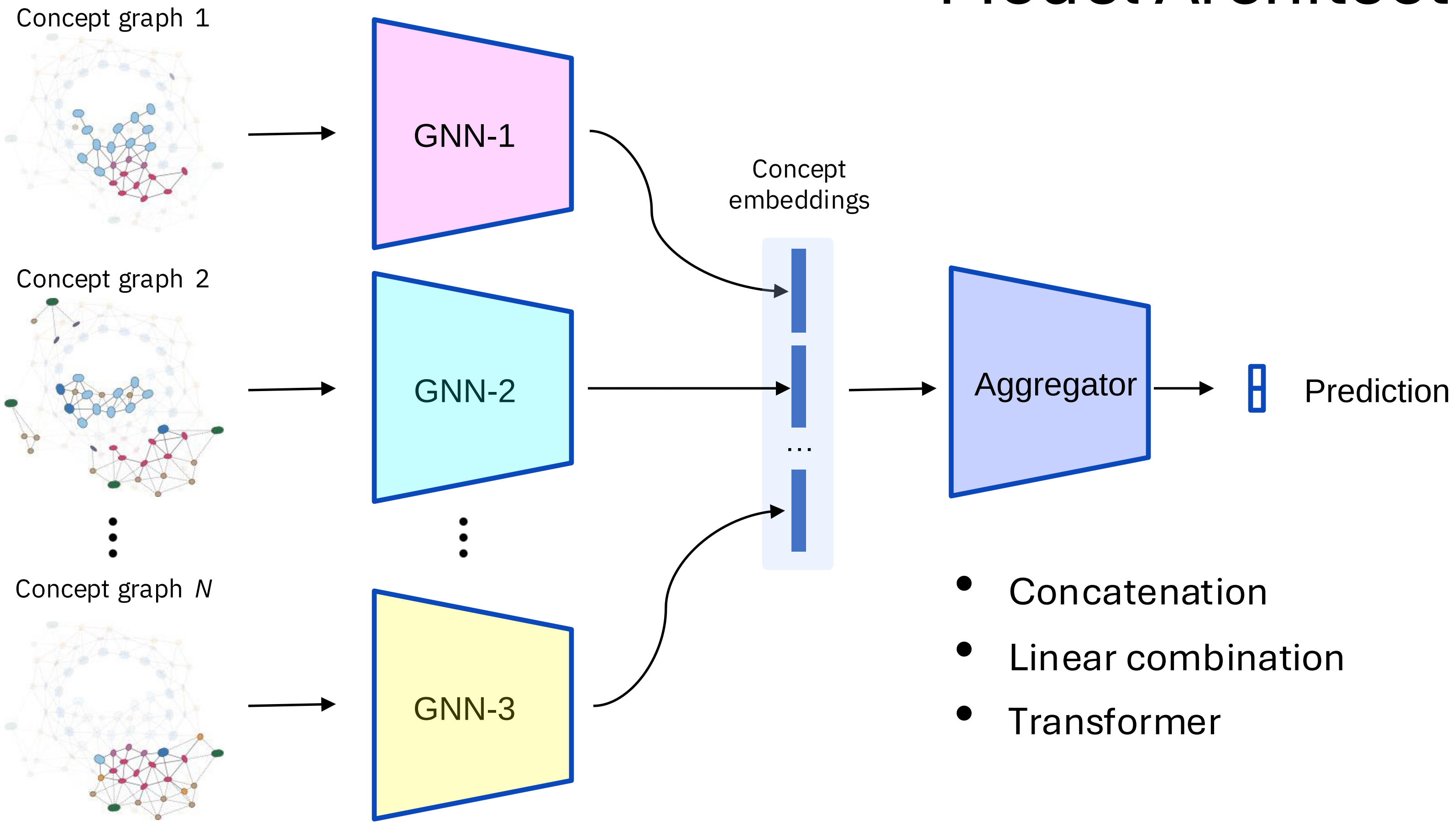
Deconvolving the input:

- Improve interpretability
- Place inductive bias in the model

Model Architecture

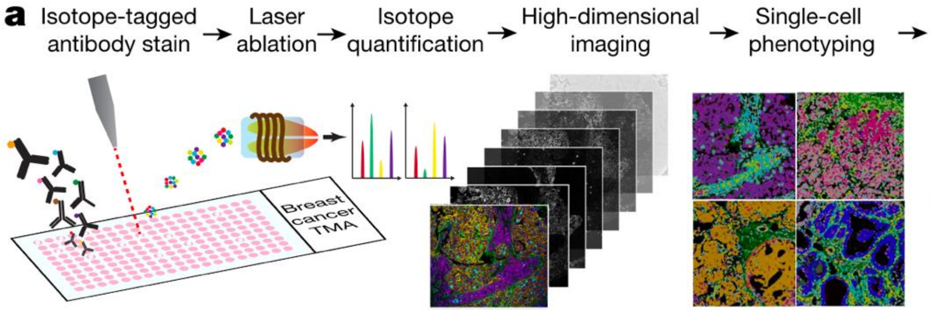


Model Architecture



- Concatenation
- Linear combination
- Transformer

Proof of Concept:
Estrogen Receptor Status Prediction



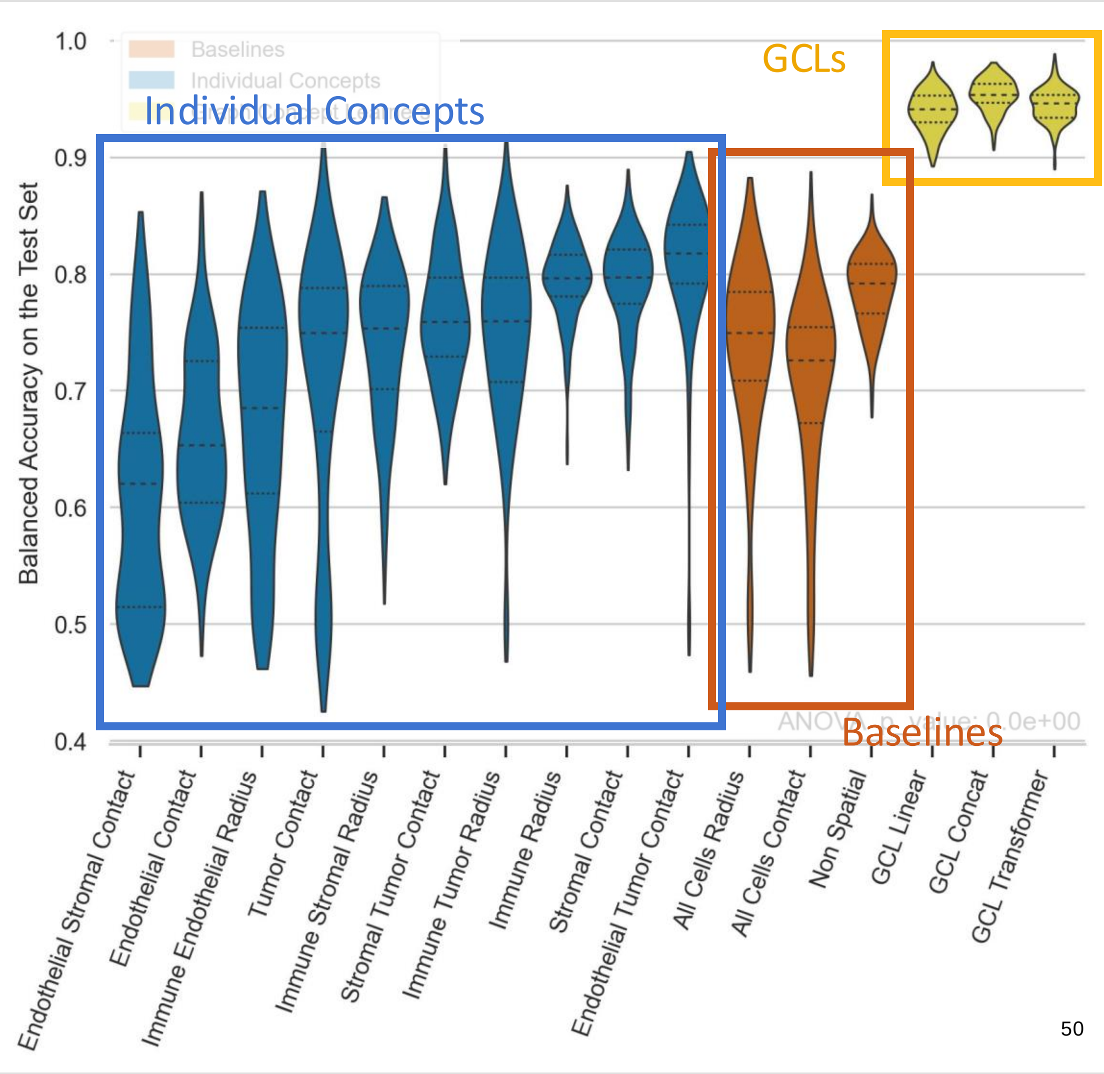
- The Jackson dataset: over 700 35-plex Imaging Mass Cytometry breast cancer images.

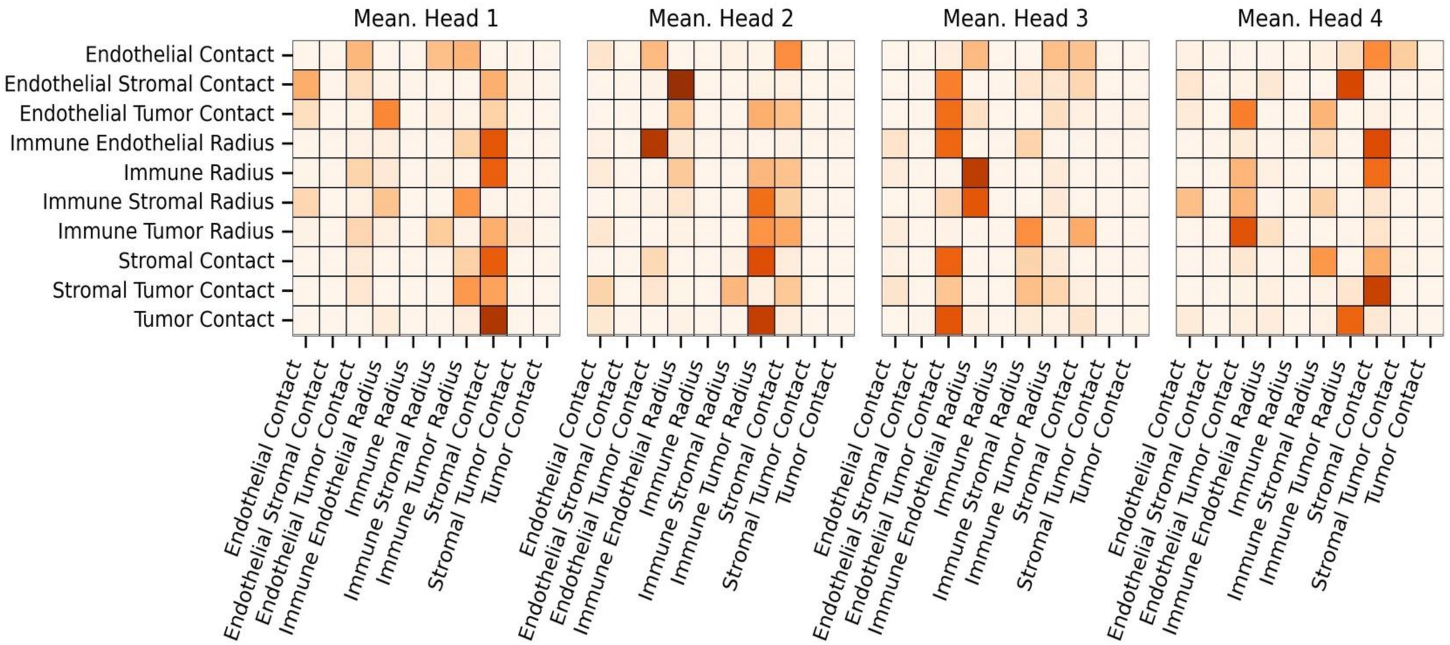
- Why ER status? Clear understanding of how labels are generated.
- Expectation: Concepts that include tumor cells would be the most attended to.

"The single-cell pathology landscape of breast cancer", Jackson *et.al.*

Performance

- GCL models excel
  - Increased performance
  - Reduced variance
- The effect size of aggregation strategy on performance seems small.
- Tumor concepts not clearly outperforming others.
- Concept decomposition can improve performance.
- Non-geometric baseline outperforms most concepts.

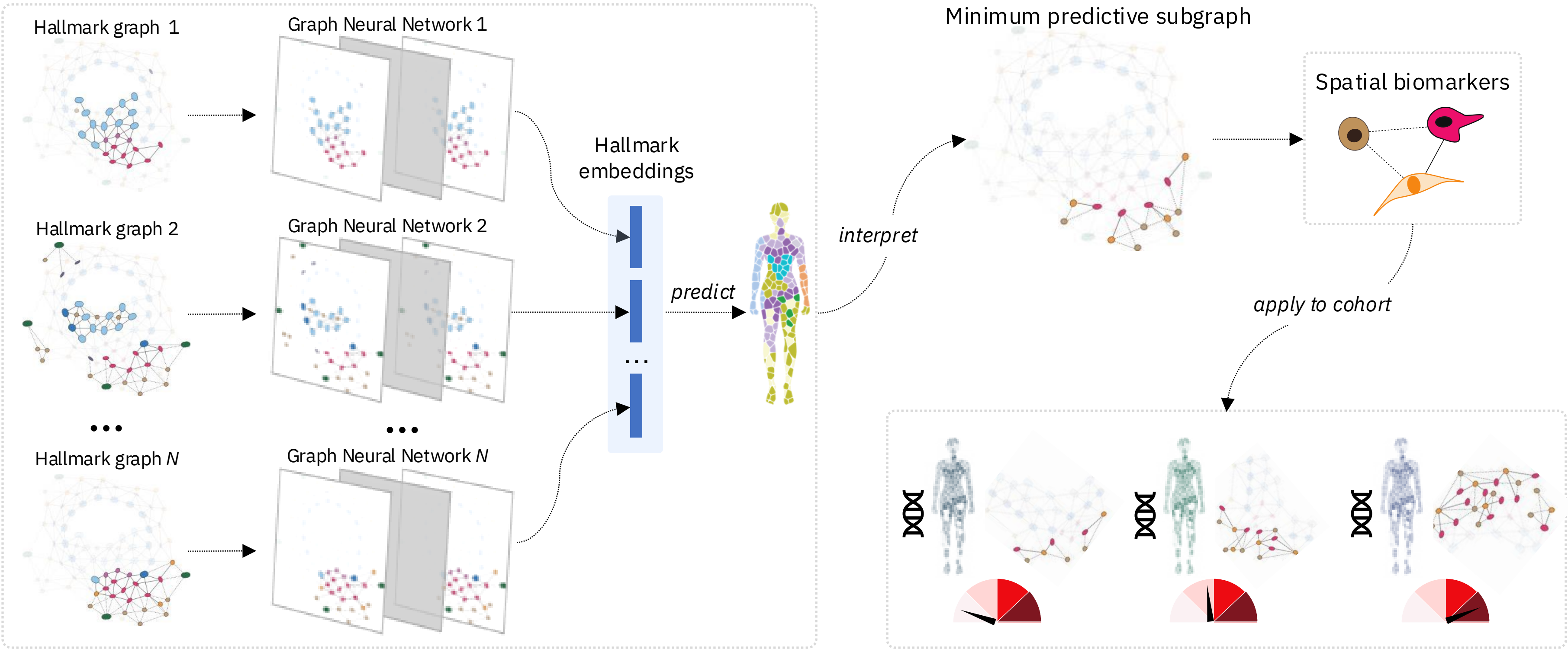



Mean. Head 1
Mean. Head 2
Mean. Head 3
Mean. Head 4
Endothelial Contact
Endothelial Stromal Contact
Endothelial Tumor Contact
Immune Endothelial Radius
Immune Radius
Immune Stromal Radius
Immune Tumor Radius
Stromal Contact
Stromal Tumor Contact
Tumor Contact

Future work


Theo Maffei


Louis McConnell



EXERCISE

Article | Published: 24 October 2022

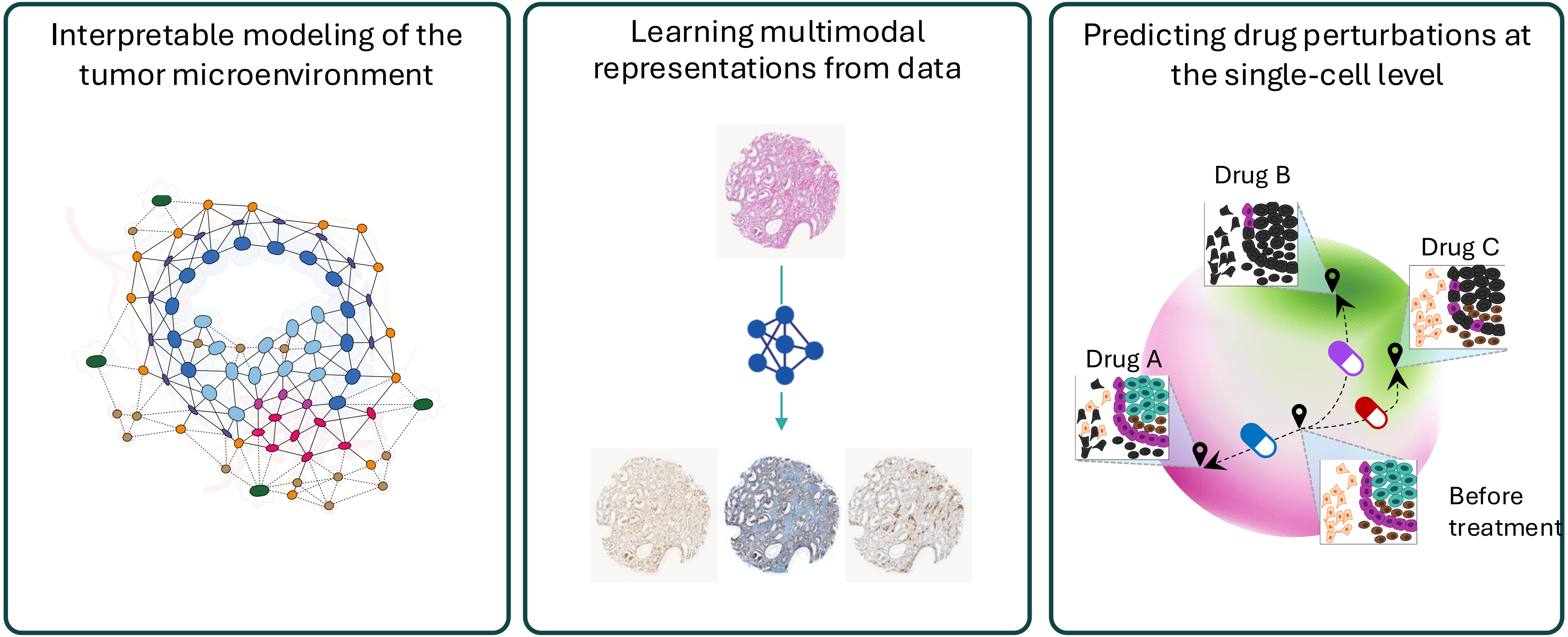
Annotation of spatially resolved single-cell data with STELLAR

Maria Brbić, Kaidi Cao, John W. Hickey, Yuqi Tan, Michael P. Snyder, Garry P. Nolan & Jure Leskovec

Nature Methods **19**, 1411–1418 (2022) | Cite this article

17k Accesses | **31** Citations | **91** Altmetric | Metrics

30 mins to read the paper, 30 mins discussion


Interpretable modeling of the tumor microenvironment
Learning multimodal representations from data
Predicting drug perturbations at the single-cell level
Drug B
Drug C
Drug A
Before treatment

KEY PAPERS AND CODE

- **ATHENA**: Martinelli & Rapsomaniki, *Bioinformatics,* 2022
- **Ovchinnikova et al.**, *npj Precision Oncology,* 2023
- **MatchCLOT**: Gossi *et al*., *Briefings in Bioinformatics,* 2023
- **VirtualMultiplexer:** Pati *et al*., *Nature Machine Intelligence*, 2024
- **CMonge**: Harsanyi *et al*., ICML MLXGen Workshop, 2024
- **CAROT**: Driessen et al., bioRxiv 2024.11.11.622906

https://github.com/AI4SCR/

AI/ML for Biomedicine


Adriano Martinelli


Theo Maffei


Louis McConnell


Melissa Ensmenger

Biomedical Data Science Center

Gottardo Group: Translational Data Science
Raisaro Group: Clinical Data Science
Fellay Group: Precision Medicine

UNIL | Université de Lausanne

IBM **Research** Zurich


Pushpak Pati
Santiago Castro
Federico Gossi


Alice Driessen
Benedek Harsanyi
Jannis Born

UNIVERSITÄT BERN

Marianna Kruithof-de Julio
Sofia Karkampouna
Francesco Bonollo
Katja Ovchinnikova

Funding

Swiss National Science Foundation

CHAN ZUCKERBERG INITIATIVE

Contact:
marianna.rapsomaniki@unil.ch