

Inferring Metabolic States from Omics Datasets with Graph Neural Networks

Marini Fabio 851977

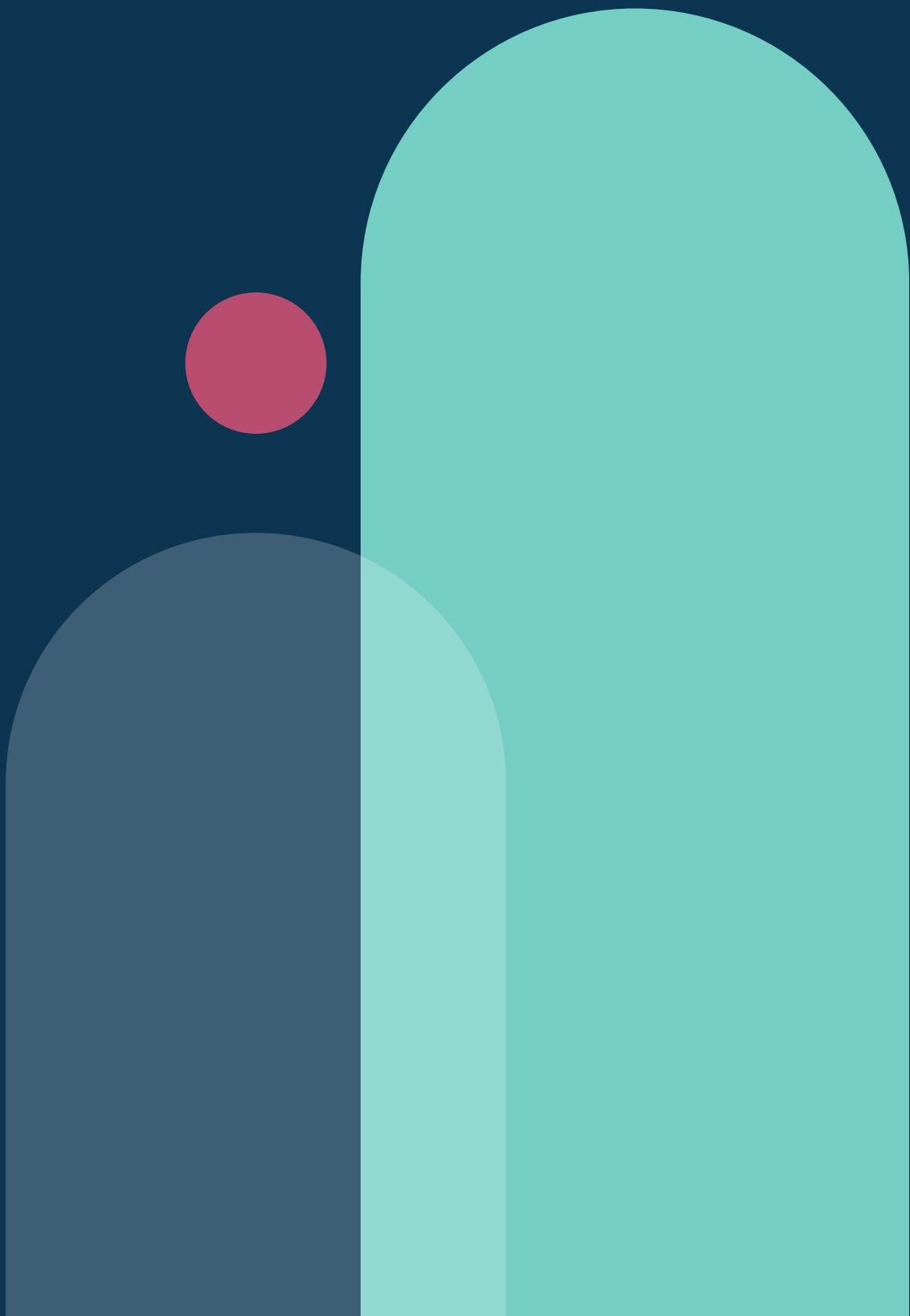
Supervisor: Prof. Damiani Chiara

Co-supervisors: Prof. Antoniotti Marco, Lapi Francesco

01

Introduction

Biological context and objective



Biological context

Fluxomics

- Study of metabolic fluxes
- Rate of conversion of **metabolites** in cellular reactions
- How cells regulate metabolism

What regulates fluxes

- Enzyme activity, determined by **gene expression**
- Metabolite concentration

Why study them

- **Dynamics** of cellular metabolism
- Identification of active pathways
- Biotechnology and medicine

Experimental limits

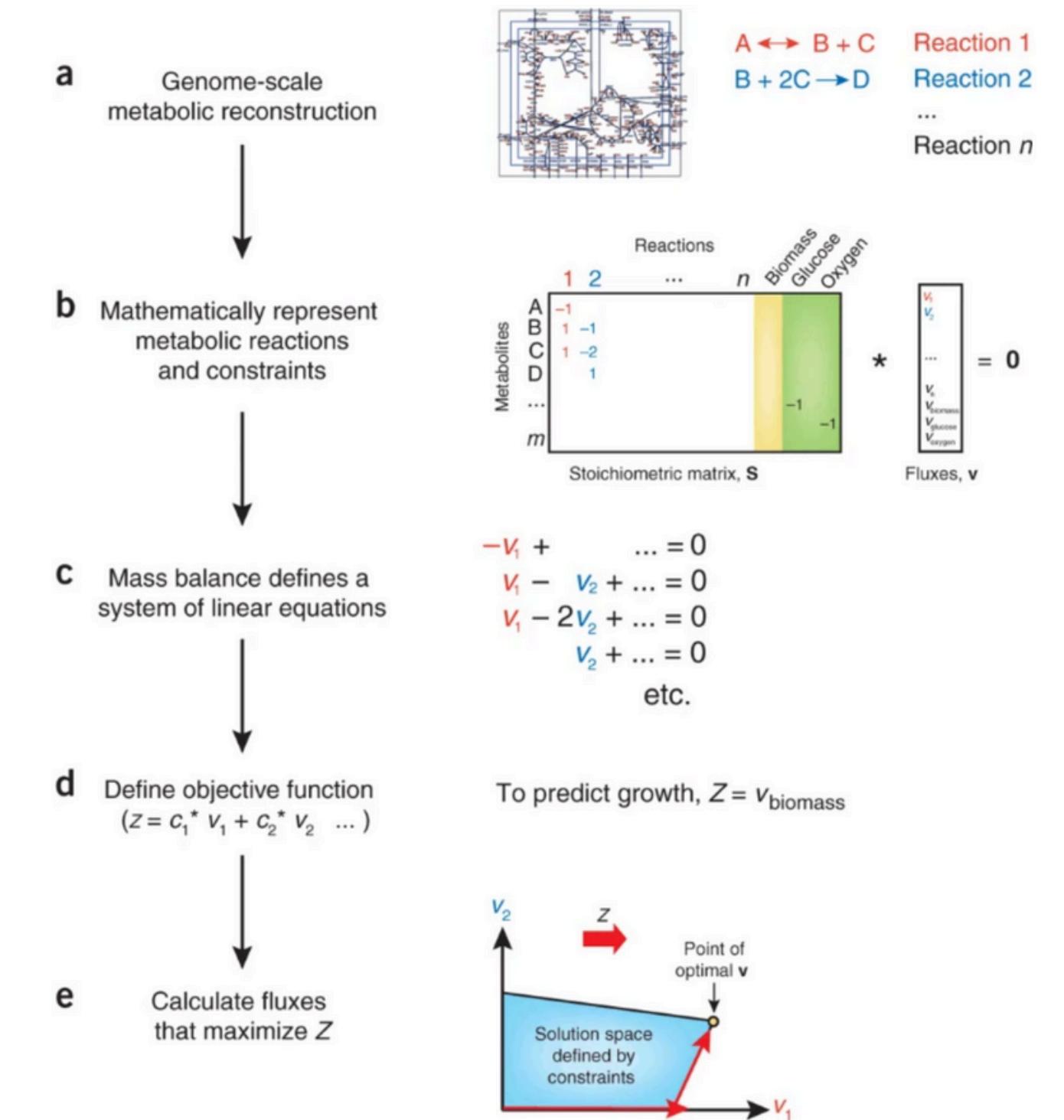
- **Not measurable** directly
- Computational approaches

Gold standard: Flux Balance Analysis

- We impose the **steady state constraint** $S \cdot v = 0$ to define the space of solutions
- **Biological constraints** define minimum and maximum limits $LB \leq v \leq UB$
- Among possible solutions, those that maximize an **objective function** are chosen. Typically, cell growth

Limiti:

- **Rigid steady-state constraint**
- Does not allow for integration of **metabolomics**



Objective

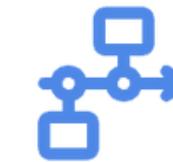
Evaluating methods based on **Deep Learning**



Computational time
contained



**Modelling
quasi-steady-state**



Non-linear relationships



Versatile objective
function



Natural modelling of
metabolism by **GNNs**

02

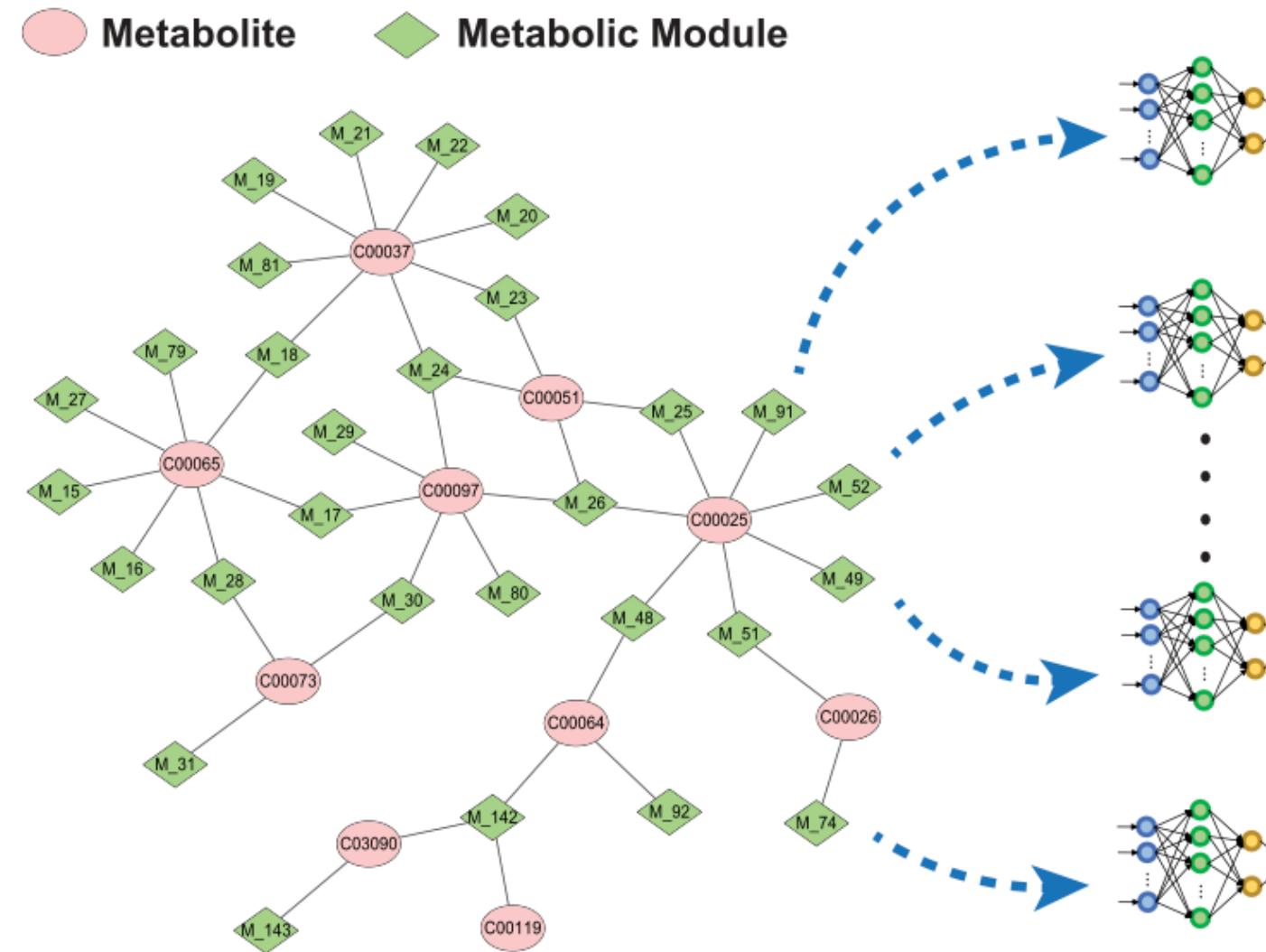
State of the art

single-cell Flux Estimation Analysis (scFEA)



scFEA

$$L = \sum_{\text{Cells Metabolize}} \sum_{\text{Metabolite}} \text{Flux Balance Loss} + \sum_{\text{Modules}} \text{Non-Negative Loss} + \sum_{\text{Modules}} \text{Inconsistency with Gene Expression} + \text{Flux Scale}$$



Minimize:

*Flux Balance
of Metabolite 1*

+

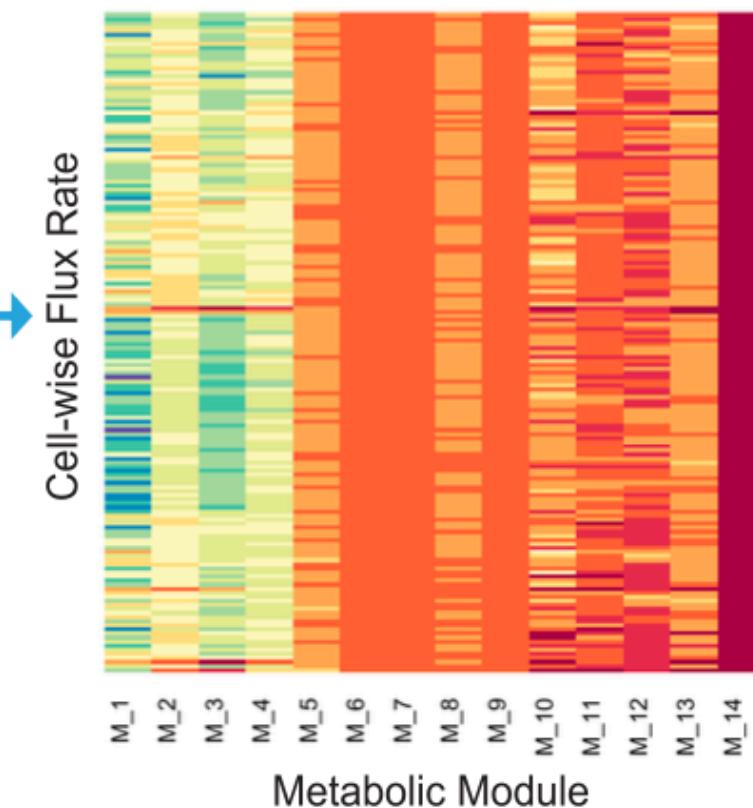
*Flux Balance
of Metabolite k*

+

⋮

⋮

**Predicted Cell-wise flux rate
of each metabolic module**



Limits of scFEA



Excludes reactions lacking coding genes



Based on a small and oversimplified network



Lack of metabolic constraints

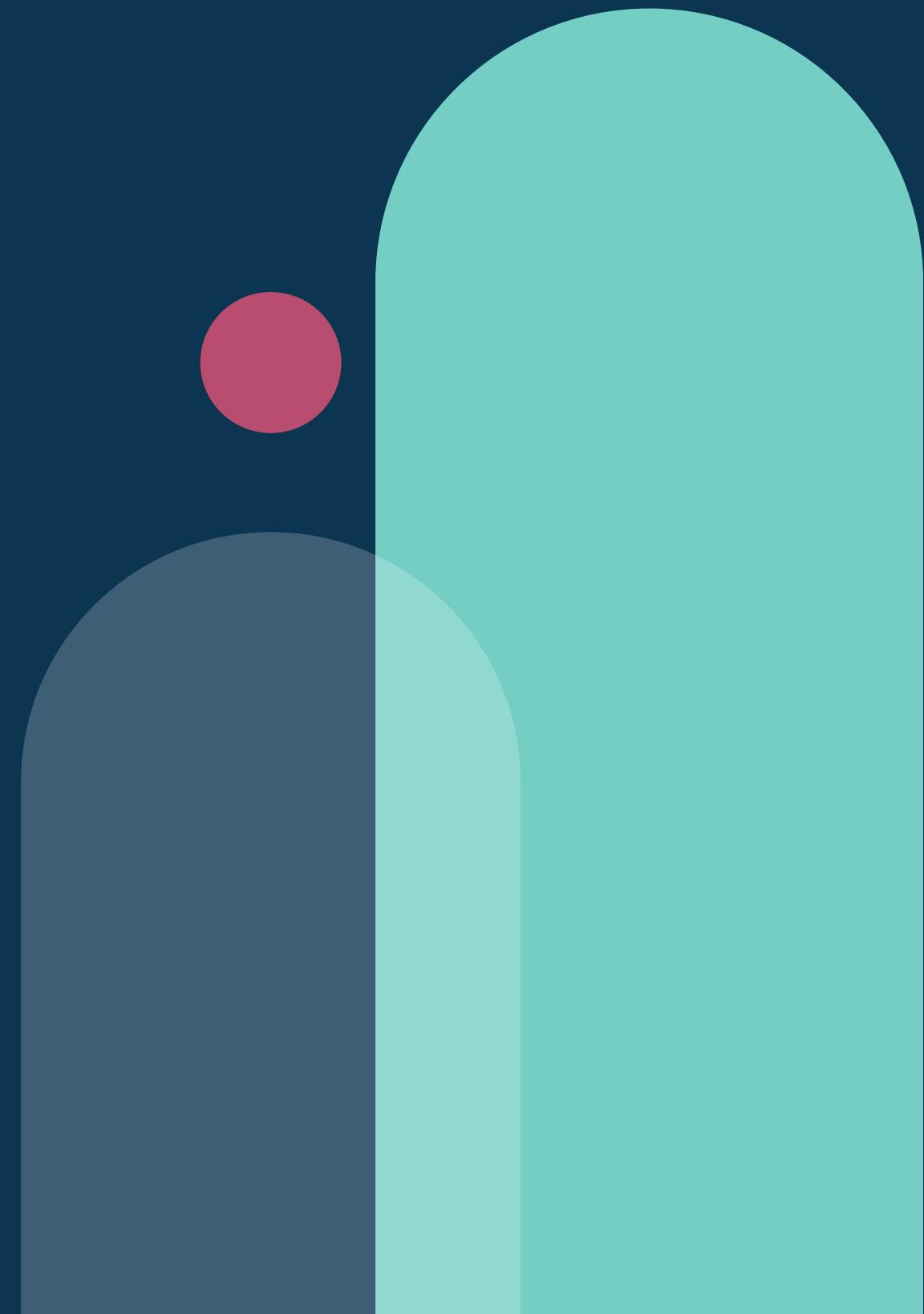


Uses only gene expression as input

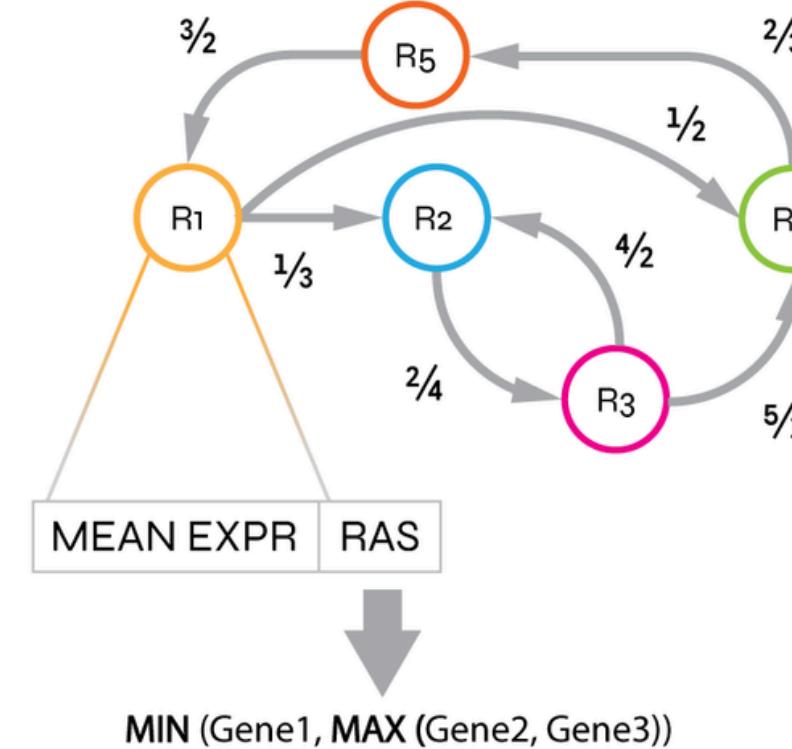
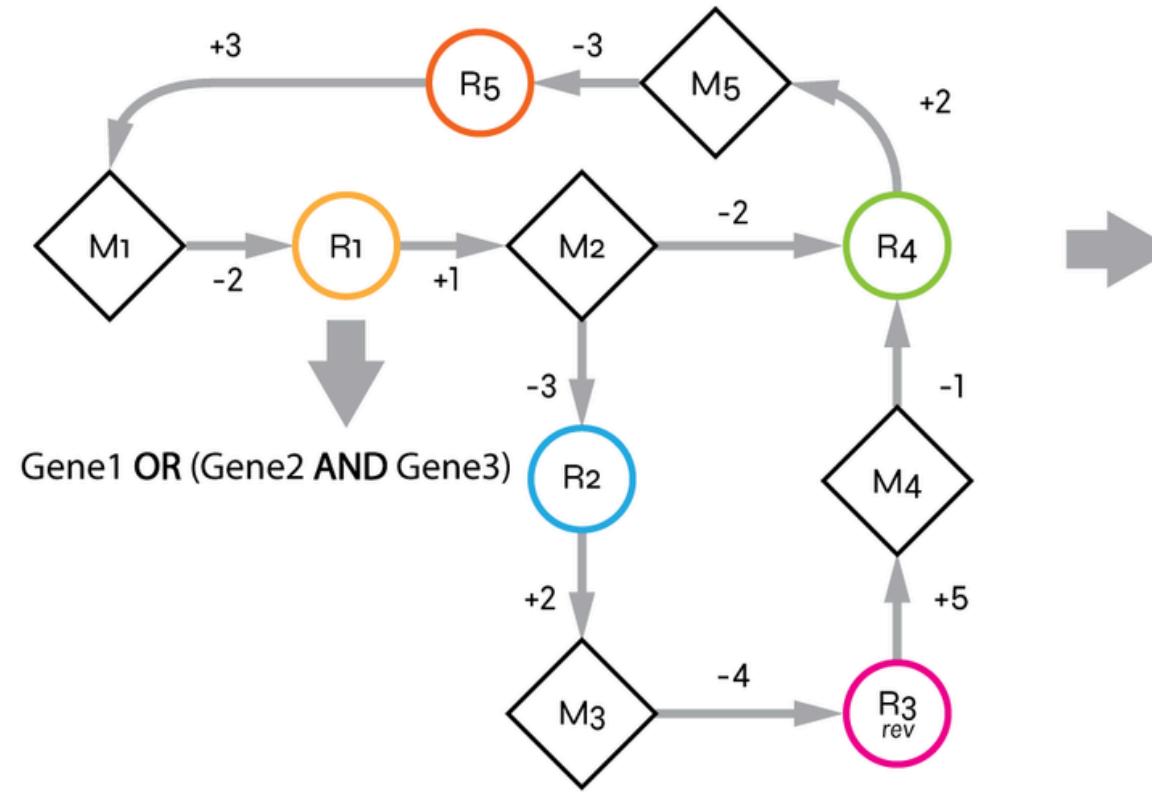
03

Developments and Methodologies

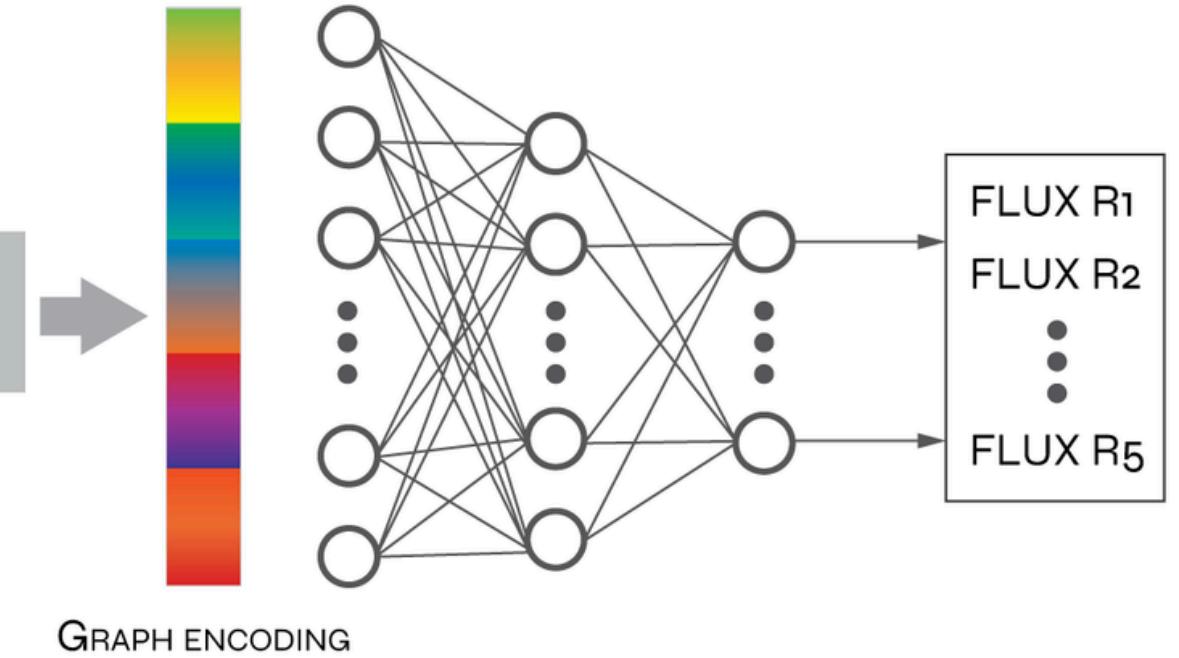
Developed model and experimental setup



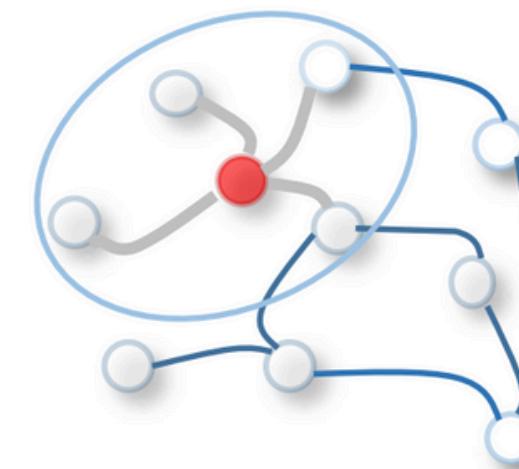
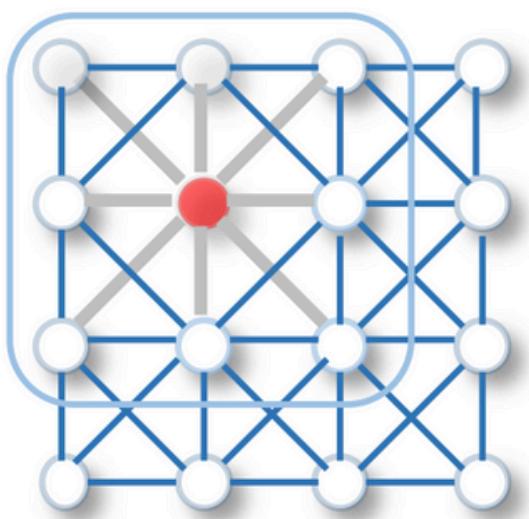
Metabolic Flux Predictor



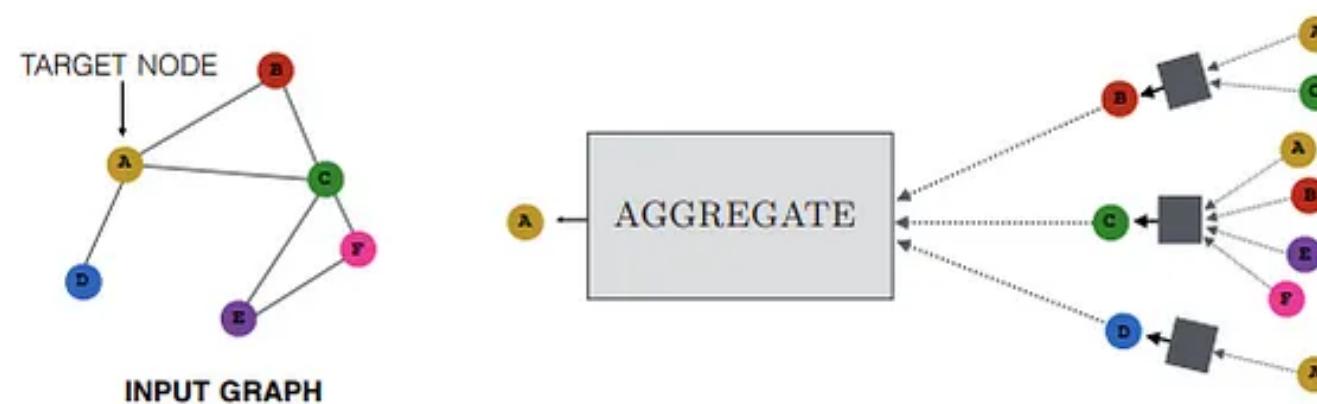
$$\mathcal{L} = \frac{1}{N} \sum_{c=1}^N \lambda_1 \mathcal{L}_{\text{flux balance}} + \lambda_2 \mathcal{L}_{\text{flux bounds}} + \lambda_3 \mathcal{L}_{\text{flux scaling}} + \lambda_4 \mathcal{L}_{\text{reaction variation}}$$



Graph Neural Networks

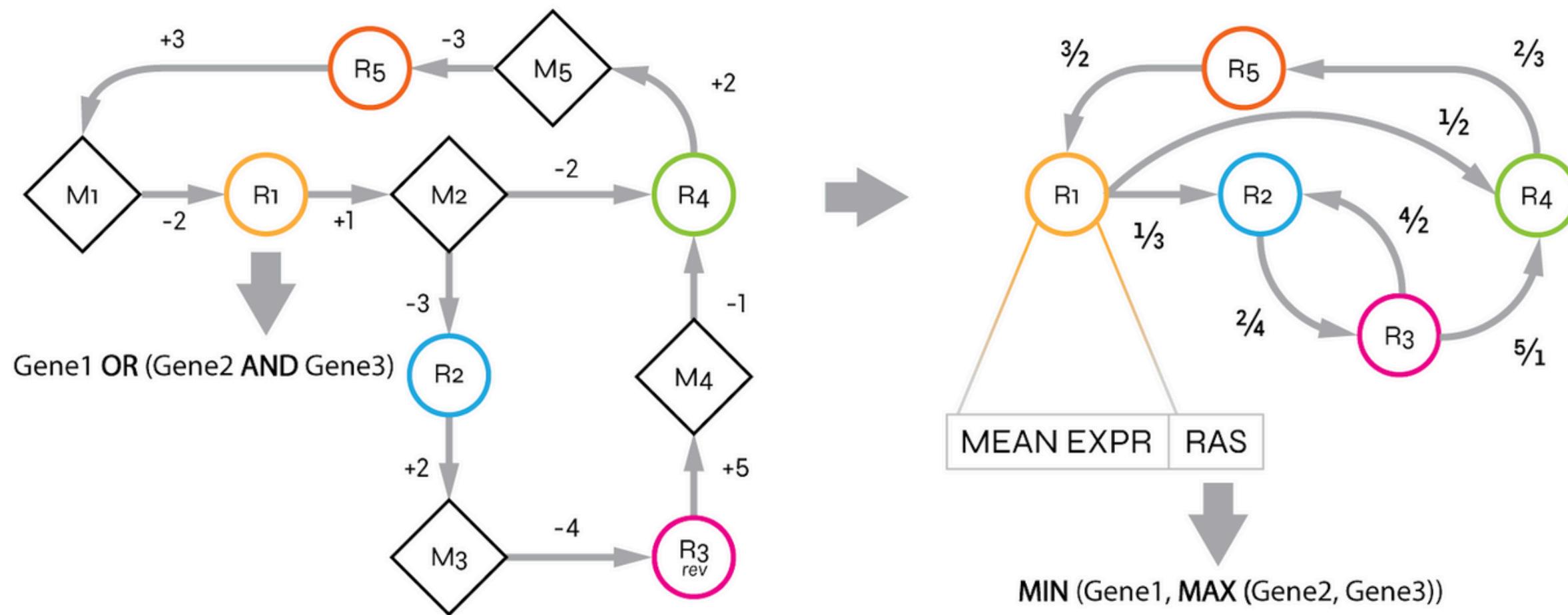


- Do not require fixed data structure
- They capture dependencies between nodes
 - **Message passing mechanism**
 - Representations that take into account the structure of the graph, but **do not depend on the order of the nodes**



Metabolic Flux Predictor – Input

Tumor transcriptomic dataset – ENGRO metabolic model



Metabolic Flux Predictor – Architecture



Graph Attention Network (GAT)

- Each node assigns a different **importance** (*attention*) to its neighbors
- In the **multi-head** case, multiple importations are processed for each neighbor

Metabolic Flux Predictor – Loss function

$$\mathcal{L} = \frac{1}{N} \sum_{c=1}^N \lambda_1 \mathcal{L}_{\text{flux balance}} + \lambda_2 \mathcal{L}_{\text{flux bounds}} + \lambda_3 \mathcal{L}_{\text{flux scaling}} + \lambda_4 \mathcal{L}_{\text{reaction variation}}$$

Flux balance  Steady-state deviation, $S \cdot v = 0$

Flux bounds  Violation of biological constraints LB and UB

Flux scaling  Distance to total gene expression in the sample

Reaction variation  Correlation with gene expression of the reaction

Experimental setup



scFEA with **ENGRO** metabolic network



scFEA with ENGRO metabolic network **including**
reactions without associated genes



Metabolic Flux Predictor



Metabolic Flux Predictor with encoding of **nodes**
passed **independently** to a single output MLP

04

Results

Analysis of approaches tested



Evaluation criteria

Tumor VS. Healthy



K-Means clustering of streams
compared with sample labels:
“tumor” and “healthy”

Comparison with transcriptomics



Consistency between transcripts
and streams predicted by K-Means
clustering, Louvain and Leiden

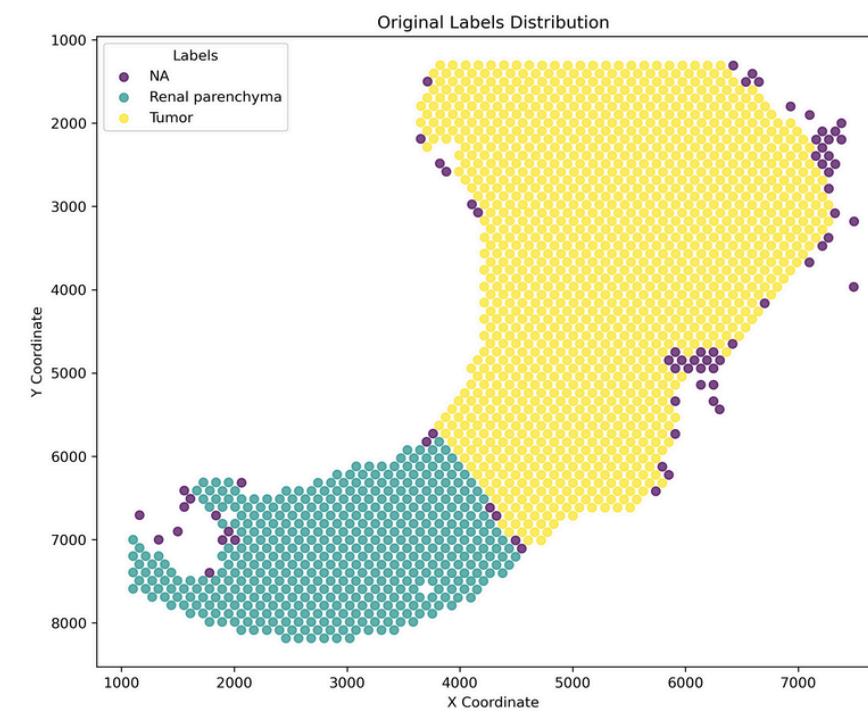
Warburg Effect



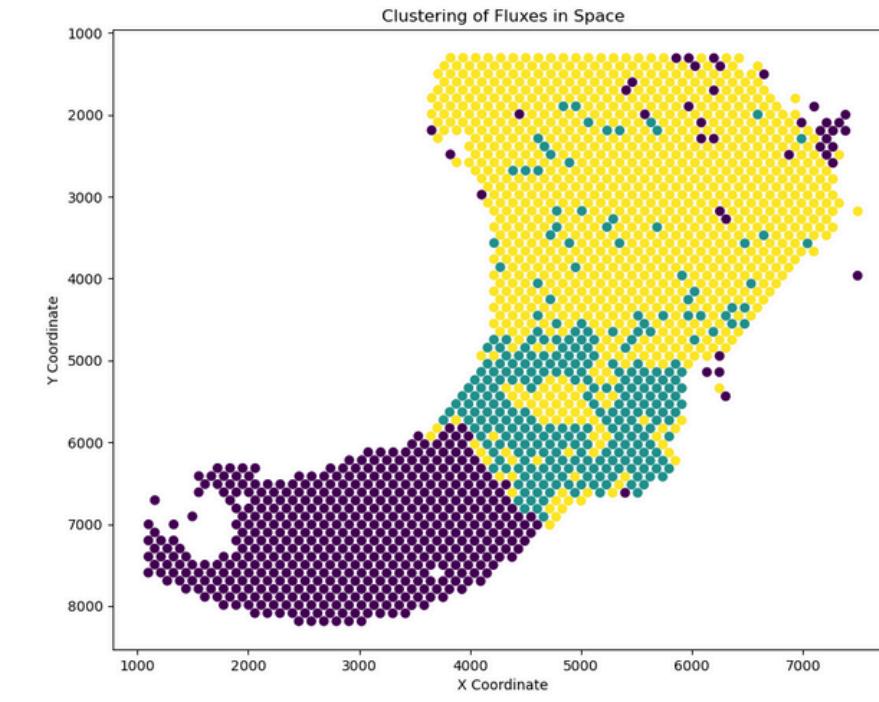
Higher predictions for *lactate* and *biomass*
production and *glucose* consumption are
expected in **tumor tissue**

Fluxes clustering vs. pathological annotations

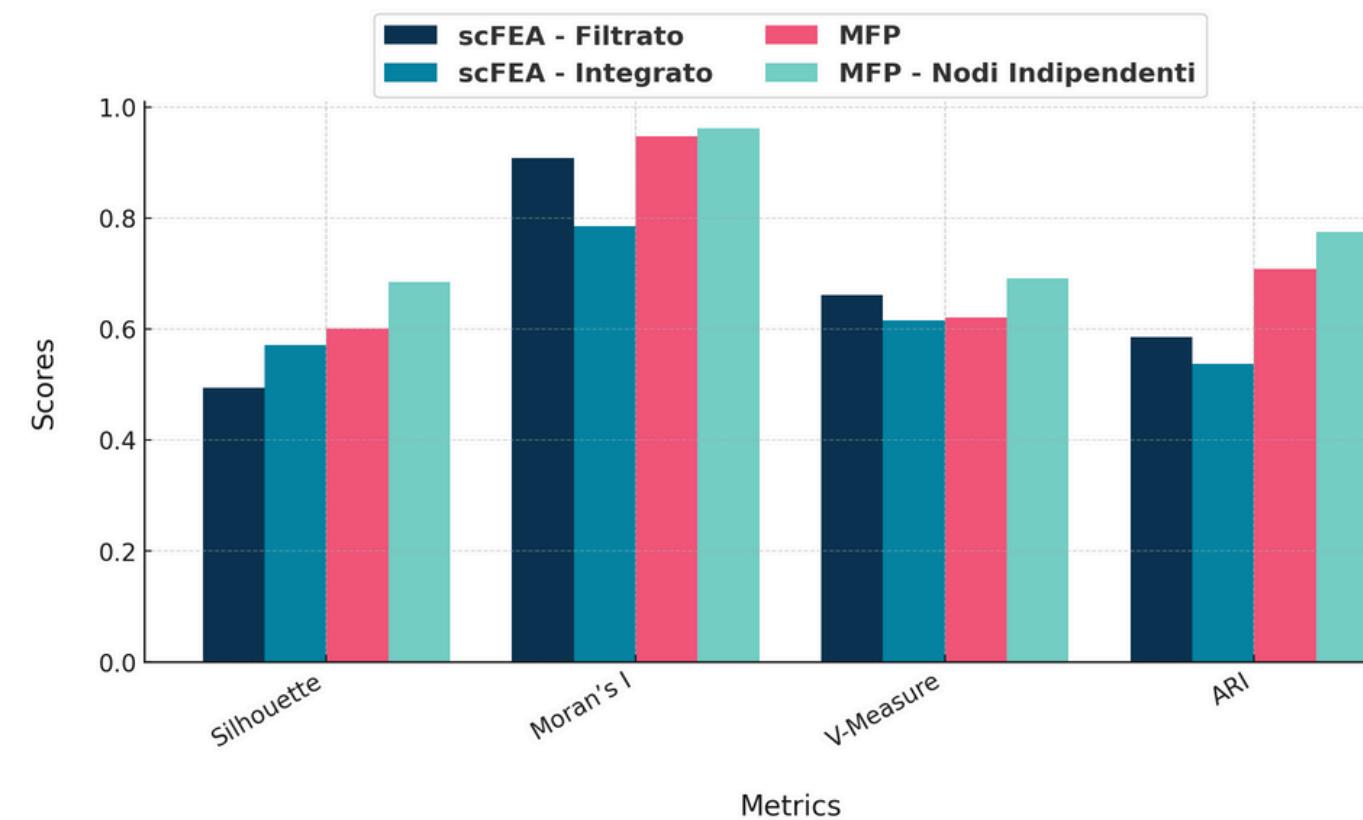
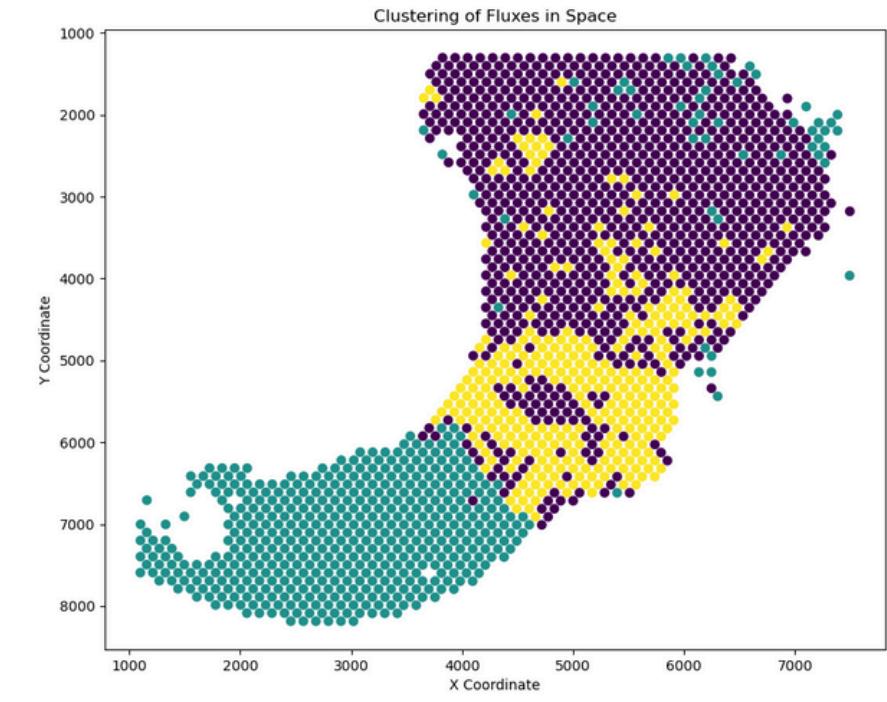
Tissue annotations



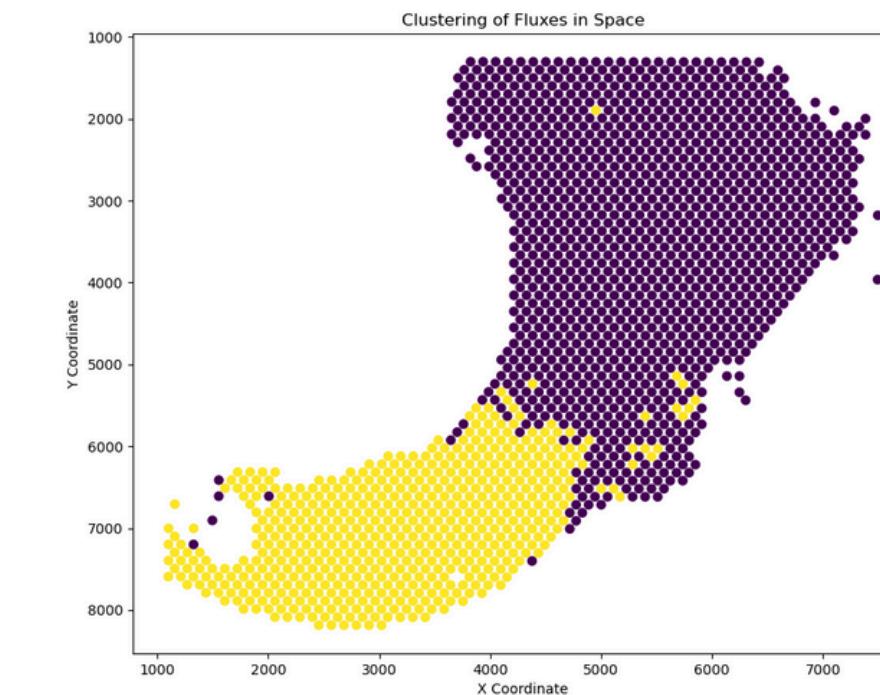
scFEA - Filtered, VM=0.66



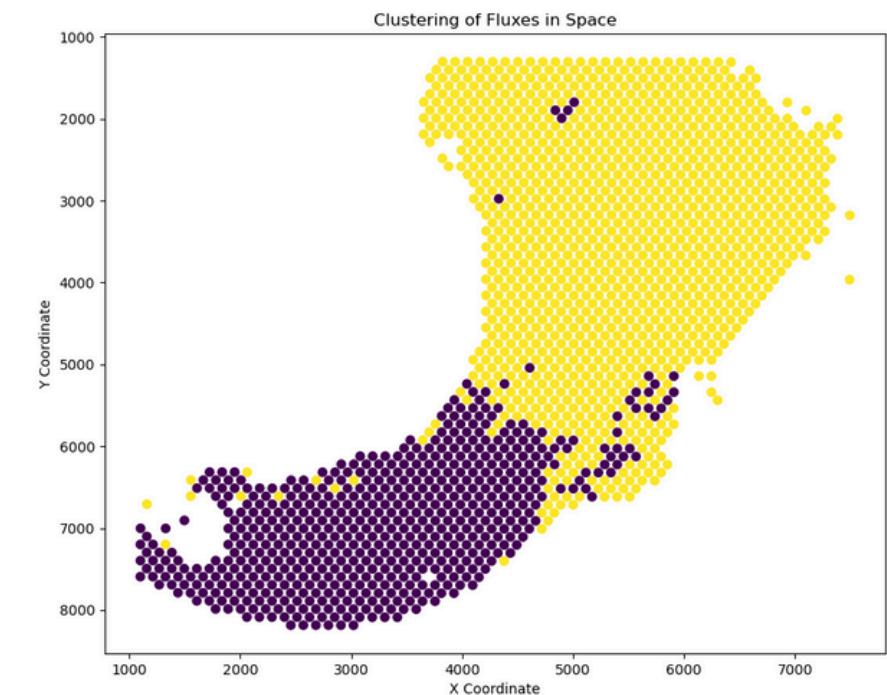
scFEA - Integrated, VM=0.61



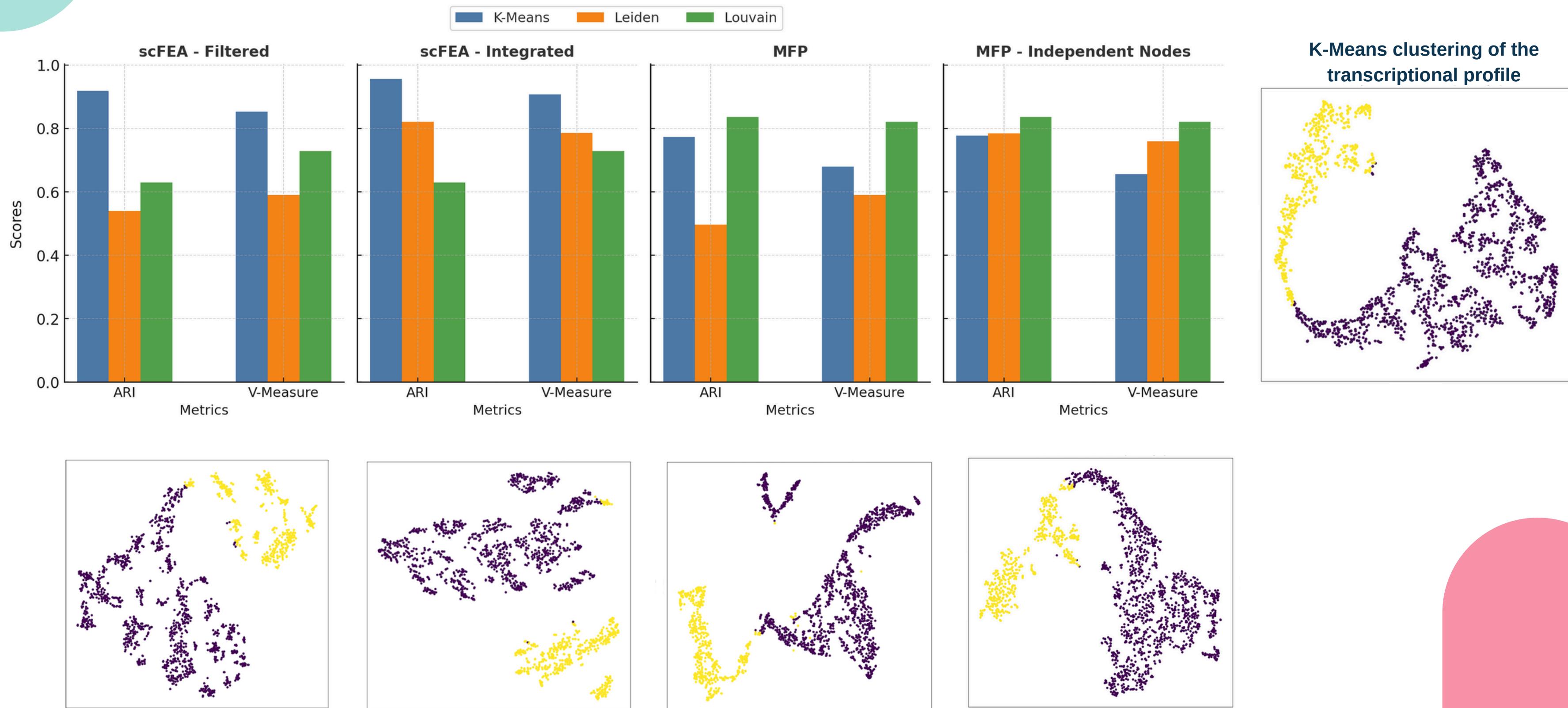
MFP - Independent Nodes, VM=0.69



MFP, VM=0.62

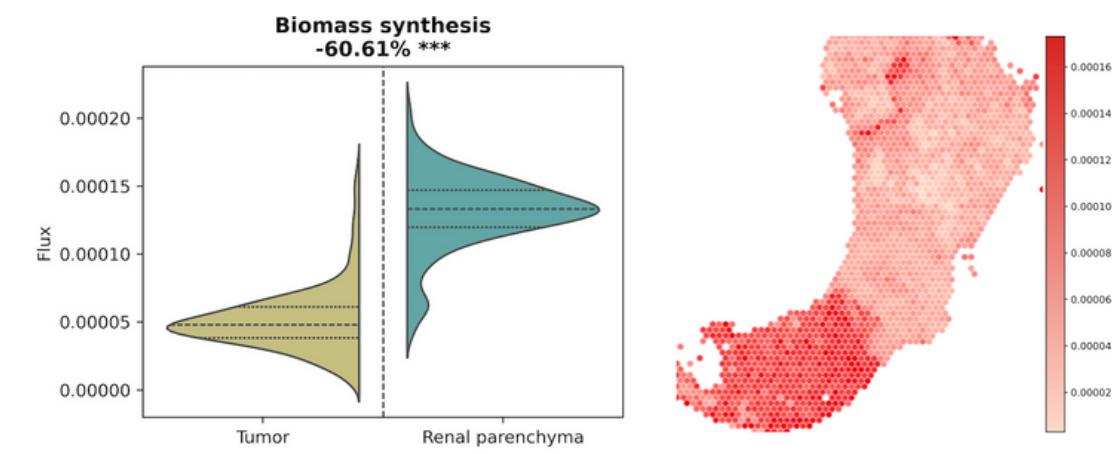
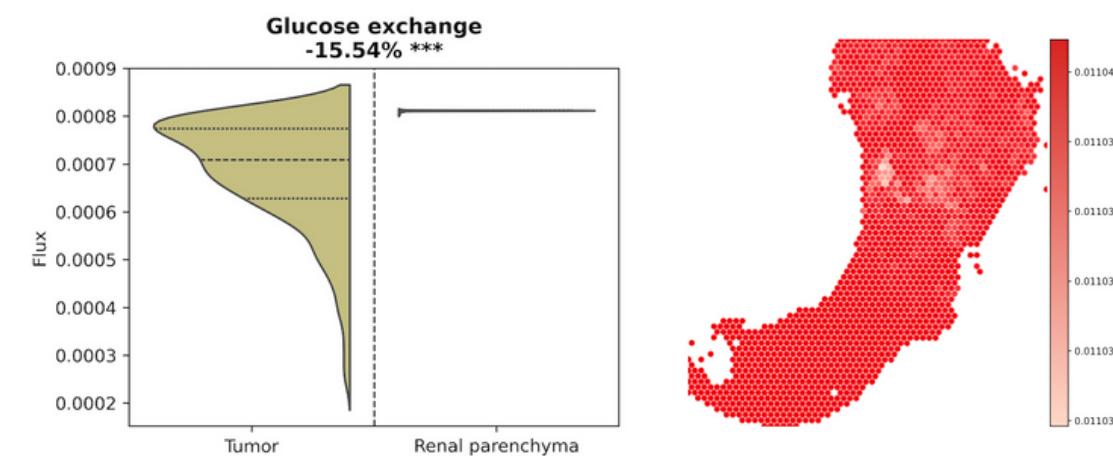
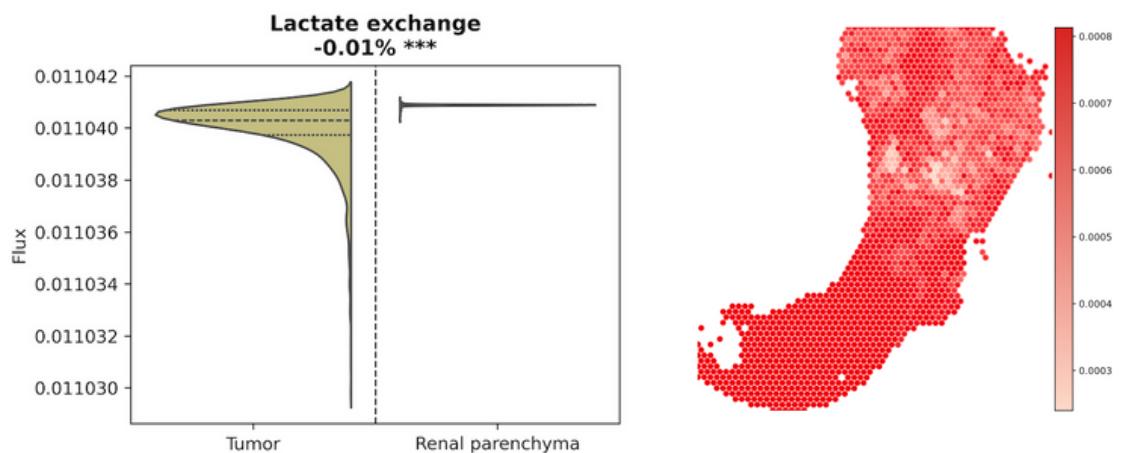


Fluxes clustering vs. transcriptional profiling

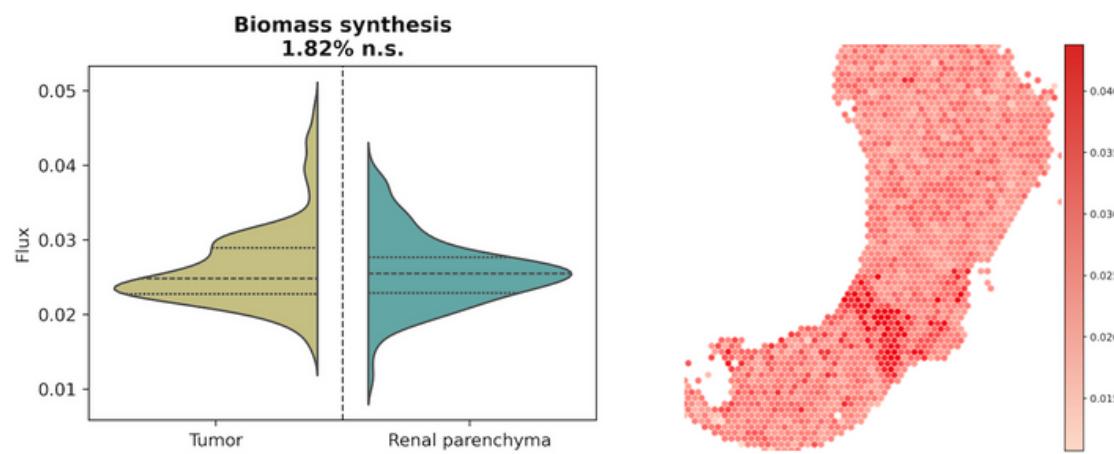
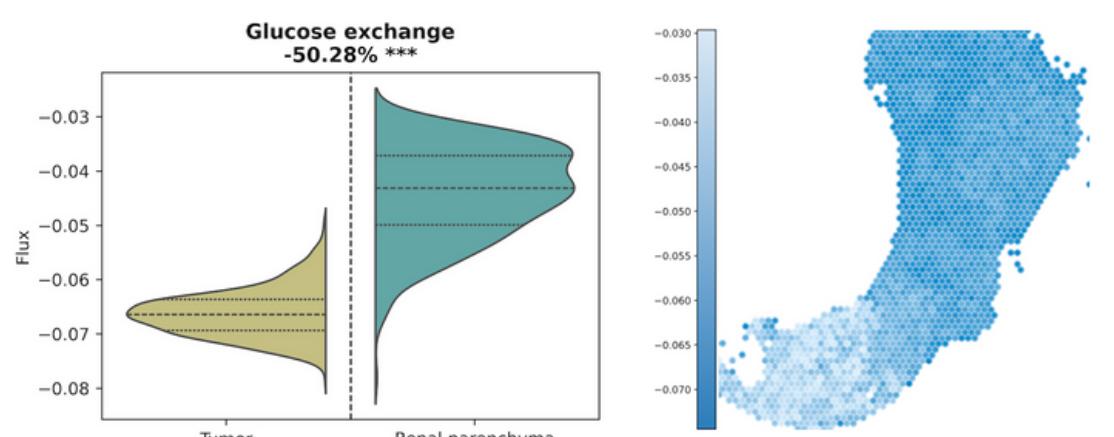
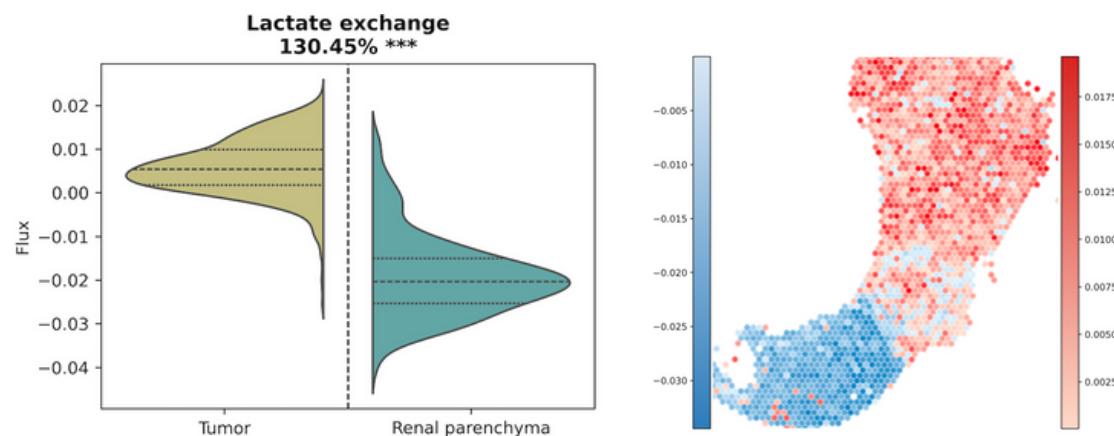


Distribution of fluxes - Effetto Warburg

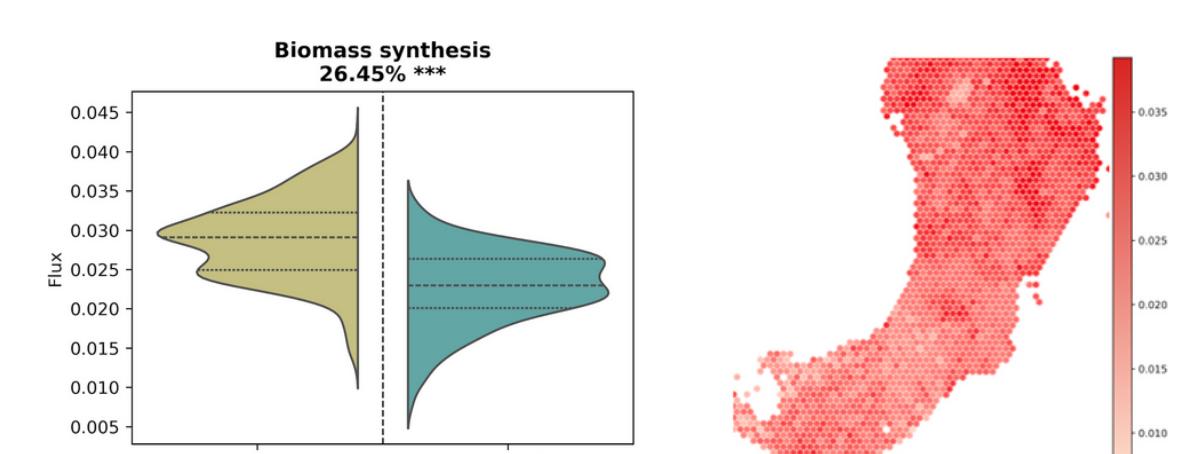
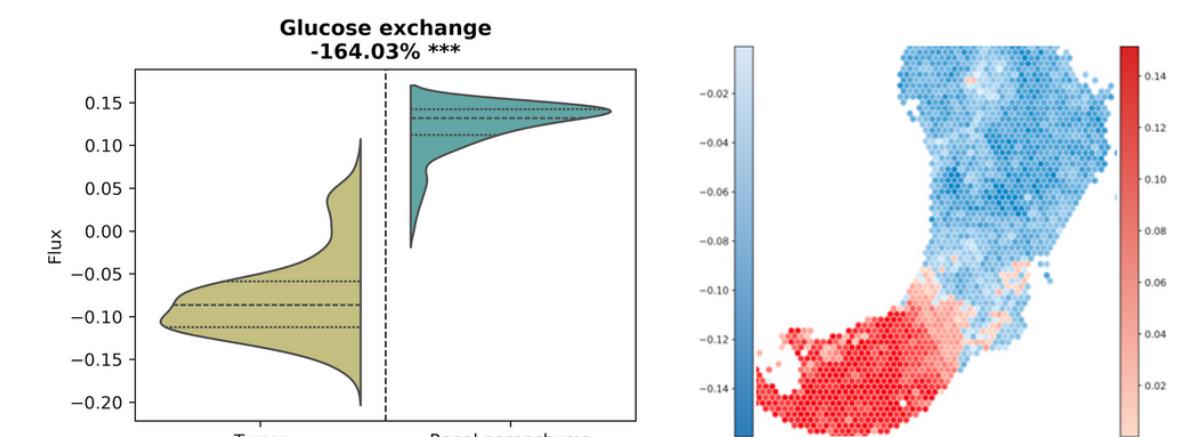
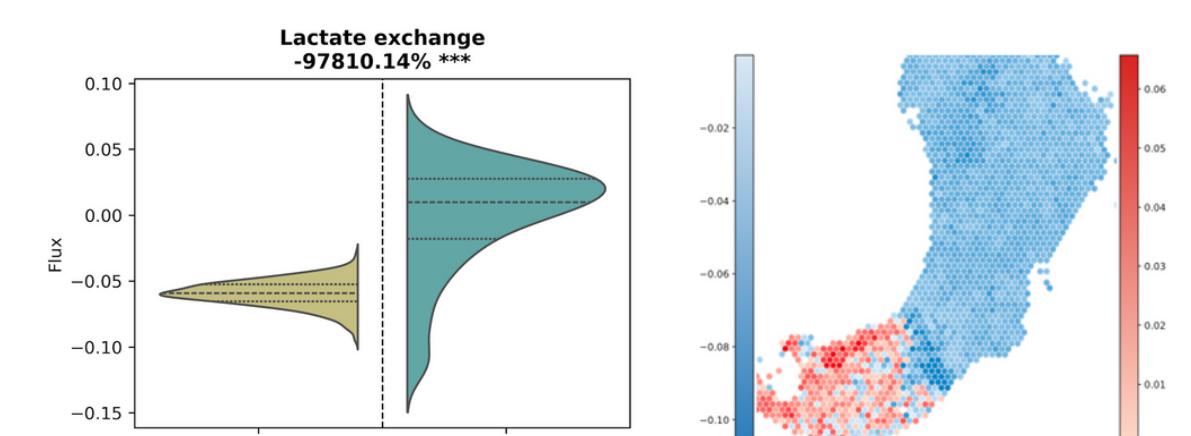
scFEA - Integrato



MFP



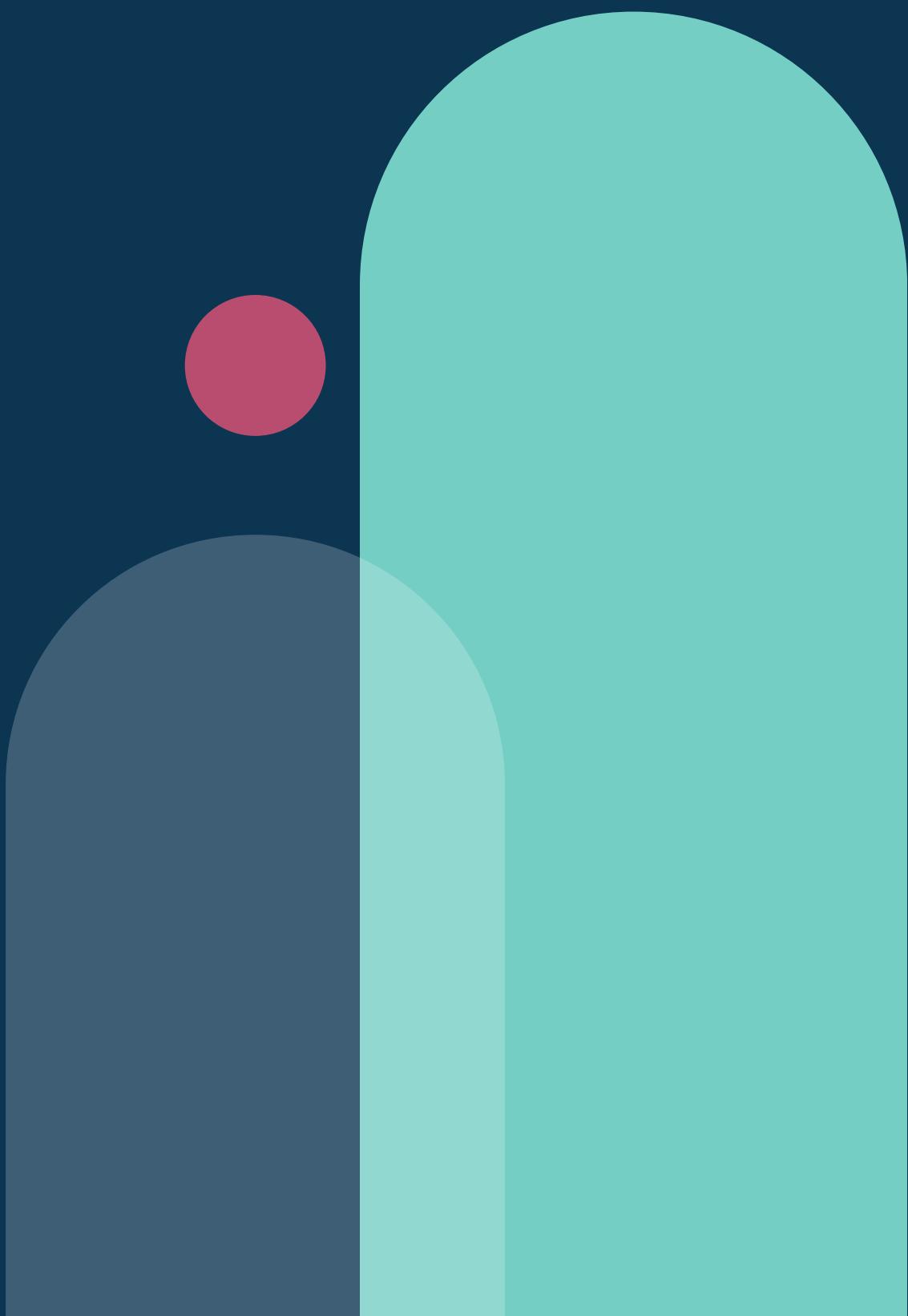
MFP - Nodi Indipendenti



05

Conclusions

Final considerations and future developments



Final considerations

- Constraint-based models such as **FBA** are effective, but limited by **rigid assumptions** that do not reflect the complexity of real biological systems
- The possibility of using approaches based on **deep learning** was explored
- **scFEA** is the only existing method that stands out for not being FBA-based, but it suffers from important limitations
- To overcome these limitations, a new model was developed based on **GNNs**
- The model demonstrated **promising performance** and biologically consistent modeling

Future developments

Model extension

- Temporal GNNs
- Self-attention

Additional biological context

- Integration of further omics data
- Impose flux values for specific reactions

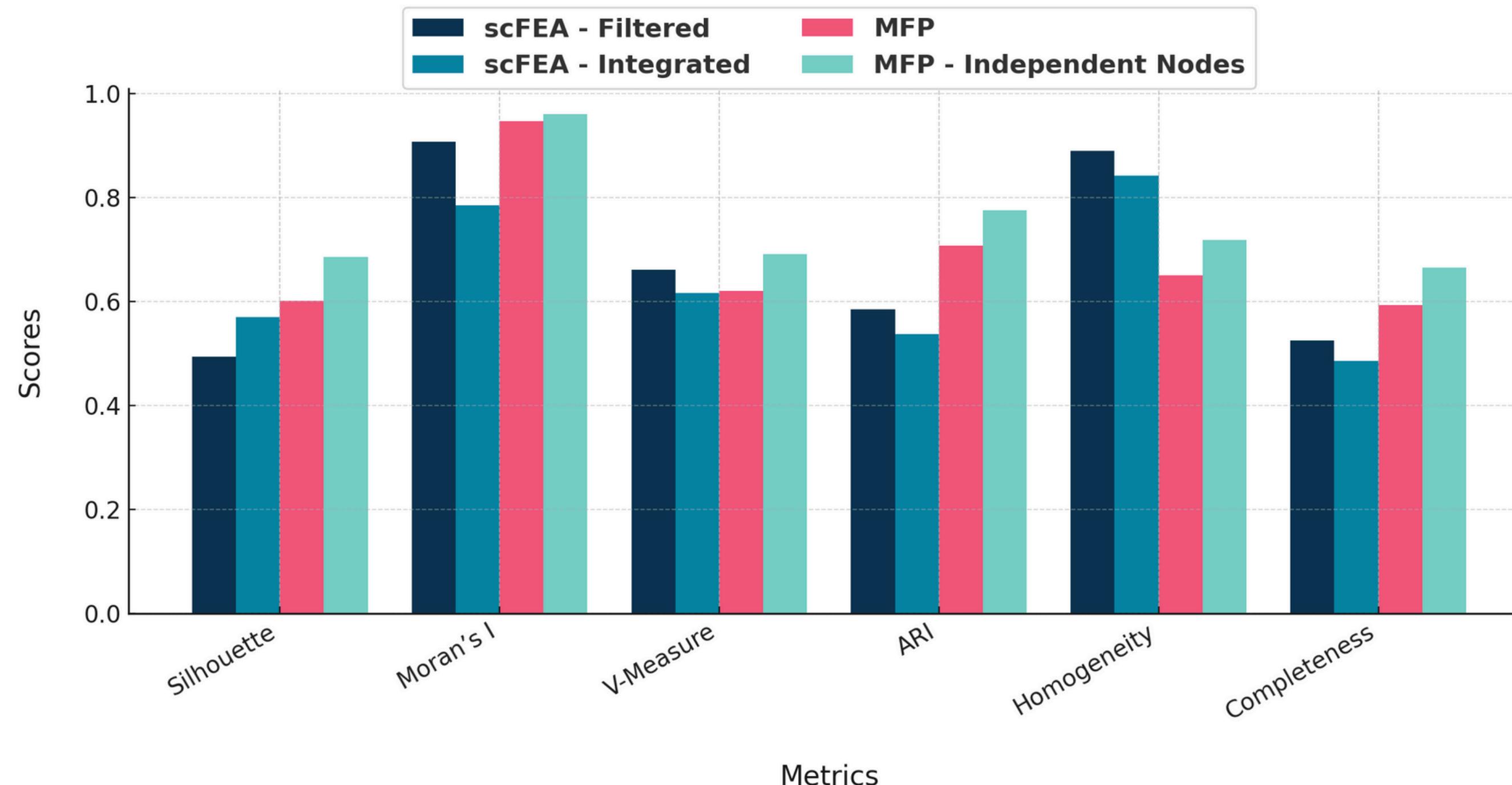
Alternative strategies

- GNN for fluxes
- Variational Graph Auto Encoder
- Graph Pooling

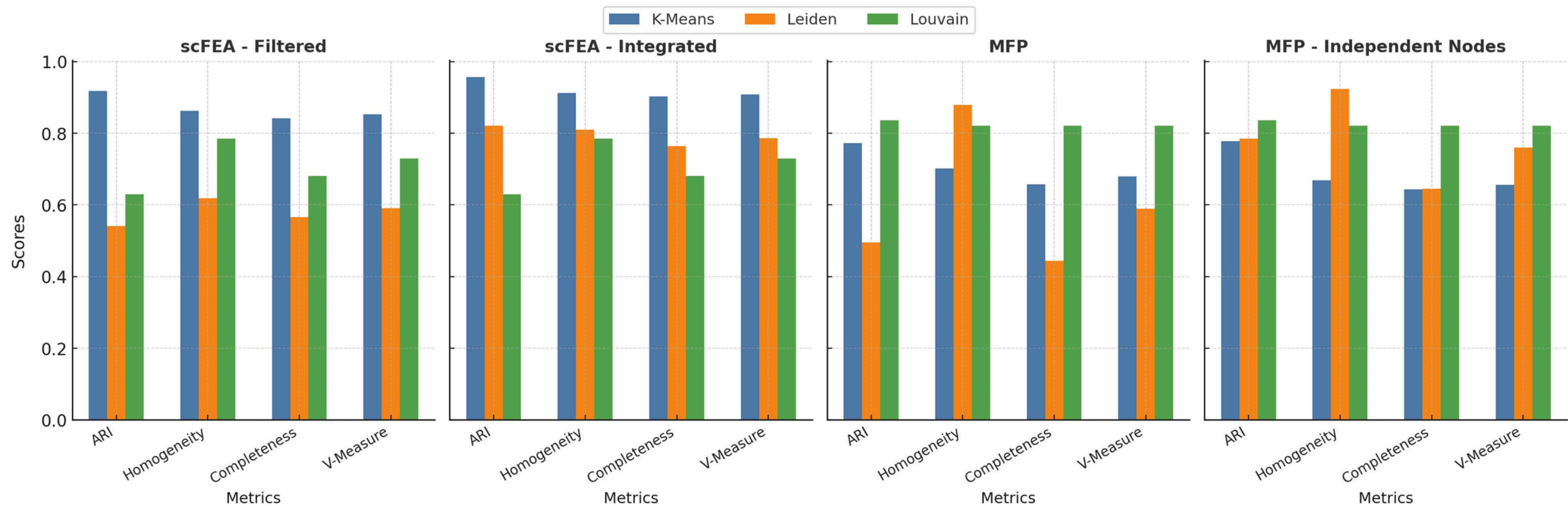
Thanks for the
attention

Marini Fabio 851977

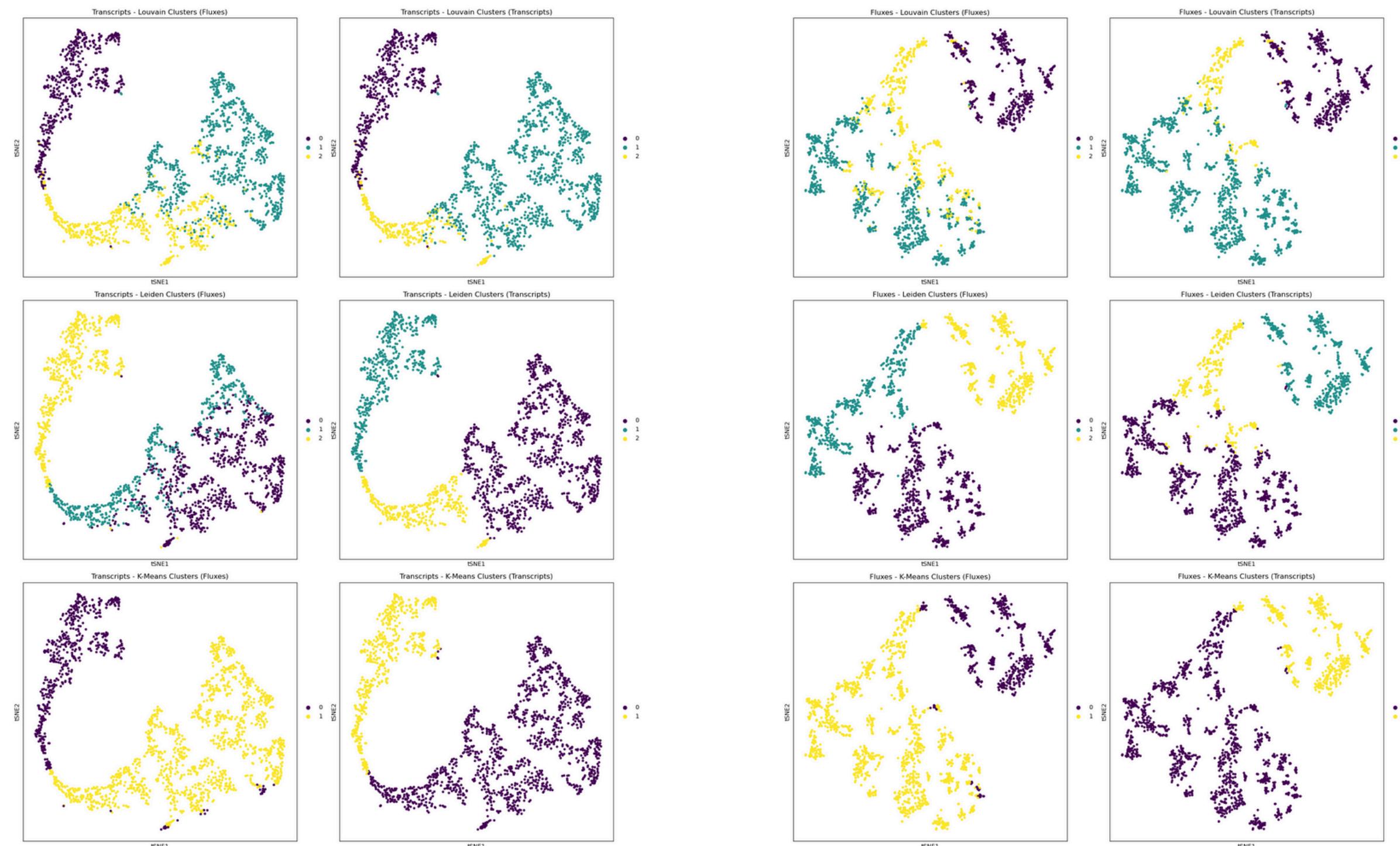
Fluxes clustering vs. pathological annotations – Comprehensive metrics



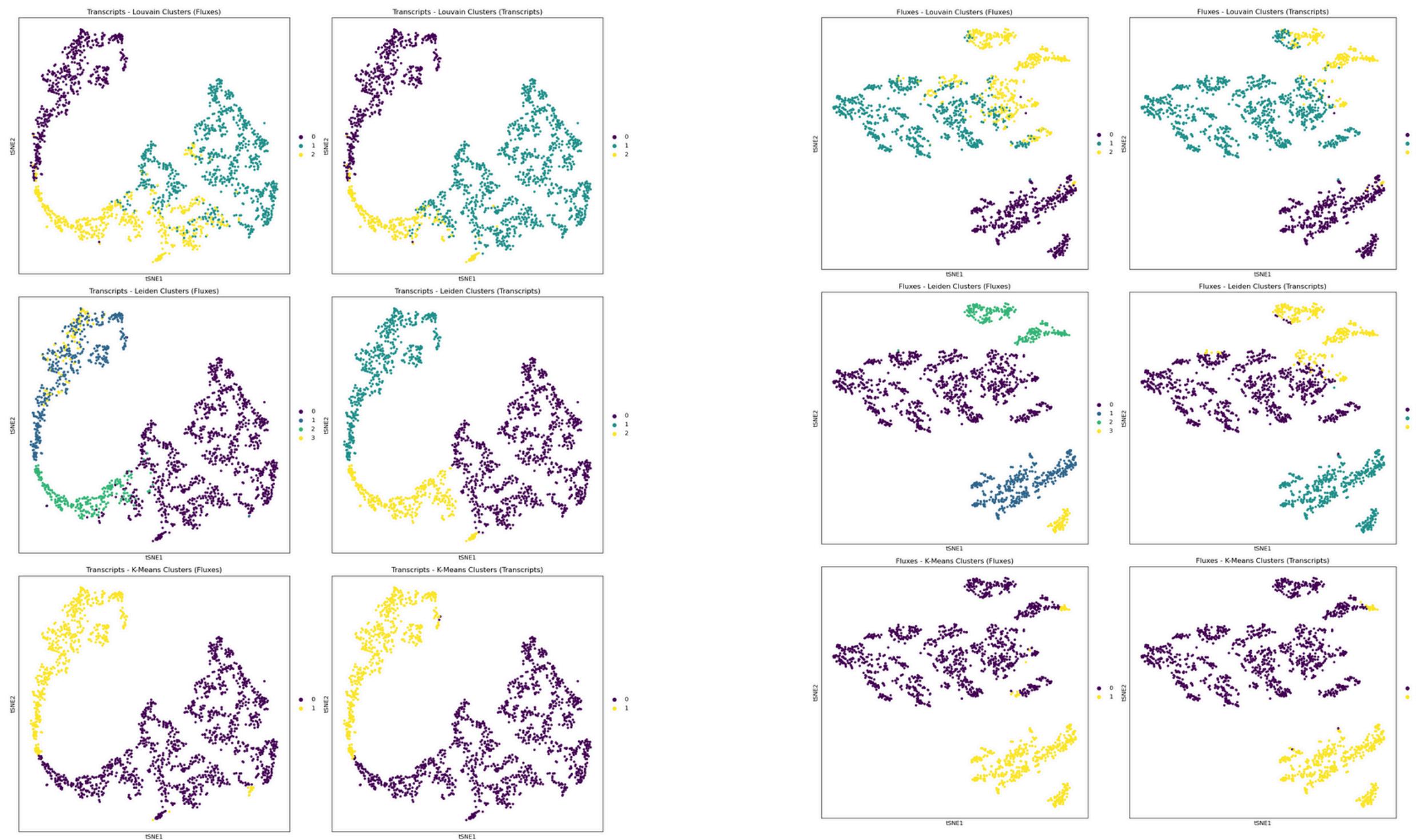
Fluxes clustering vs transcriptional profiling - Comprehensive metrics



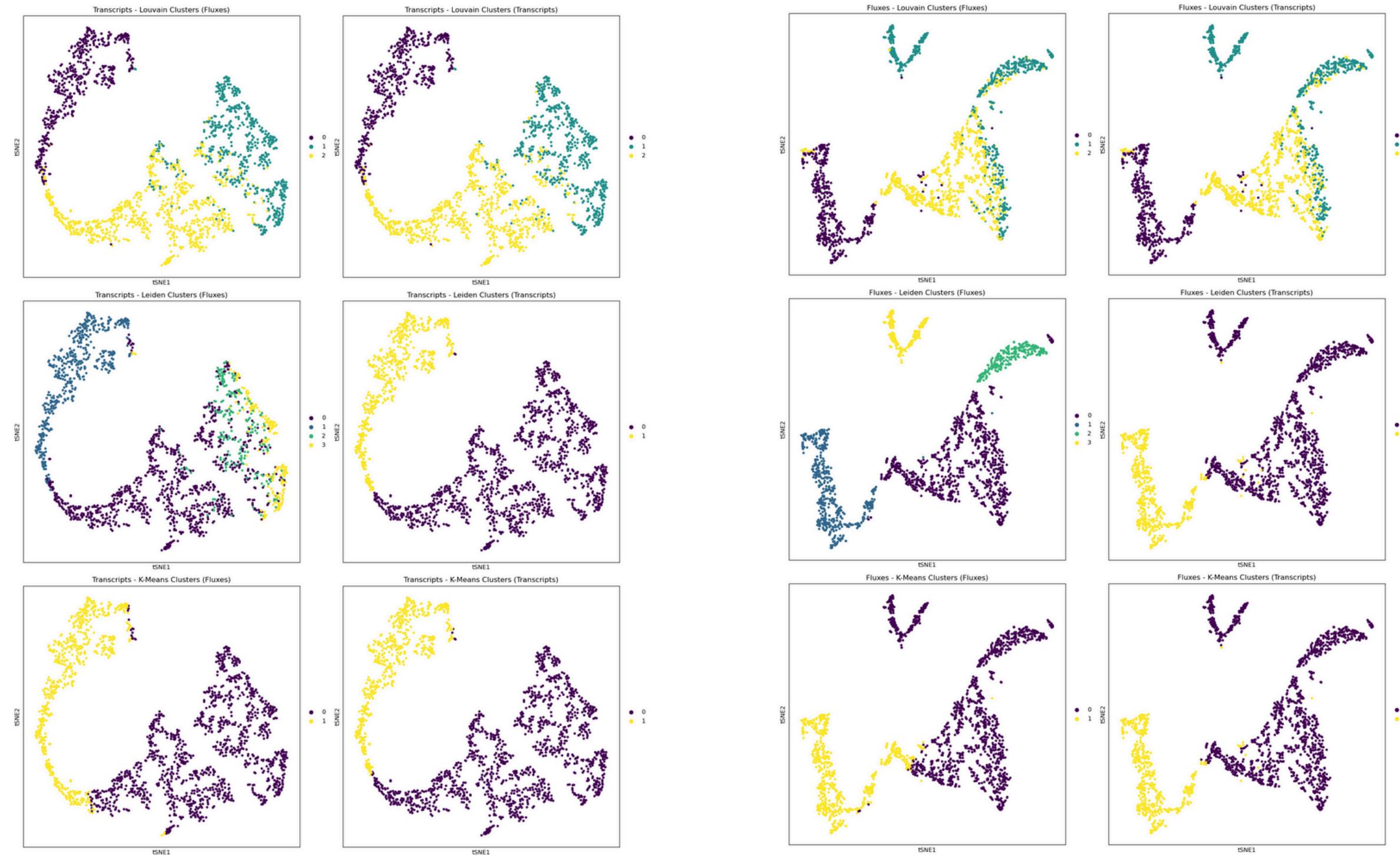
Fluxes clustering vs transcriptional profiling - scFEA Filtered



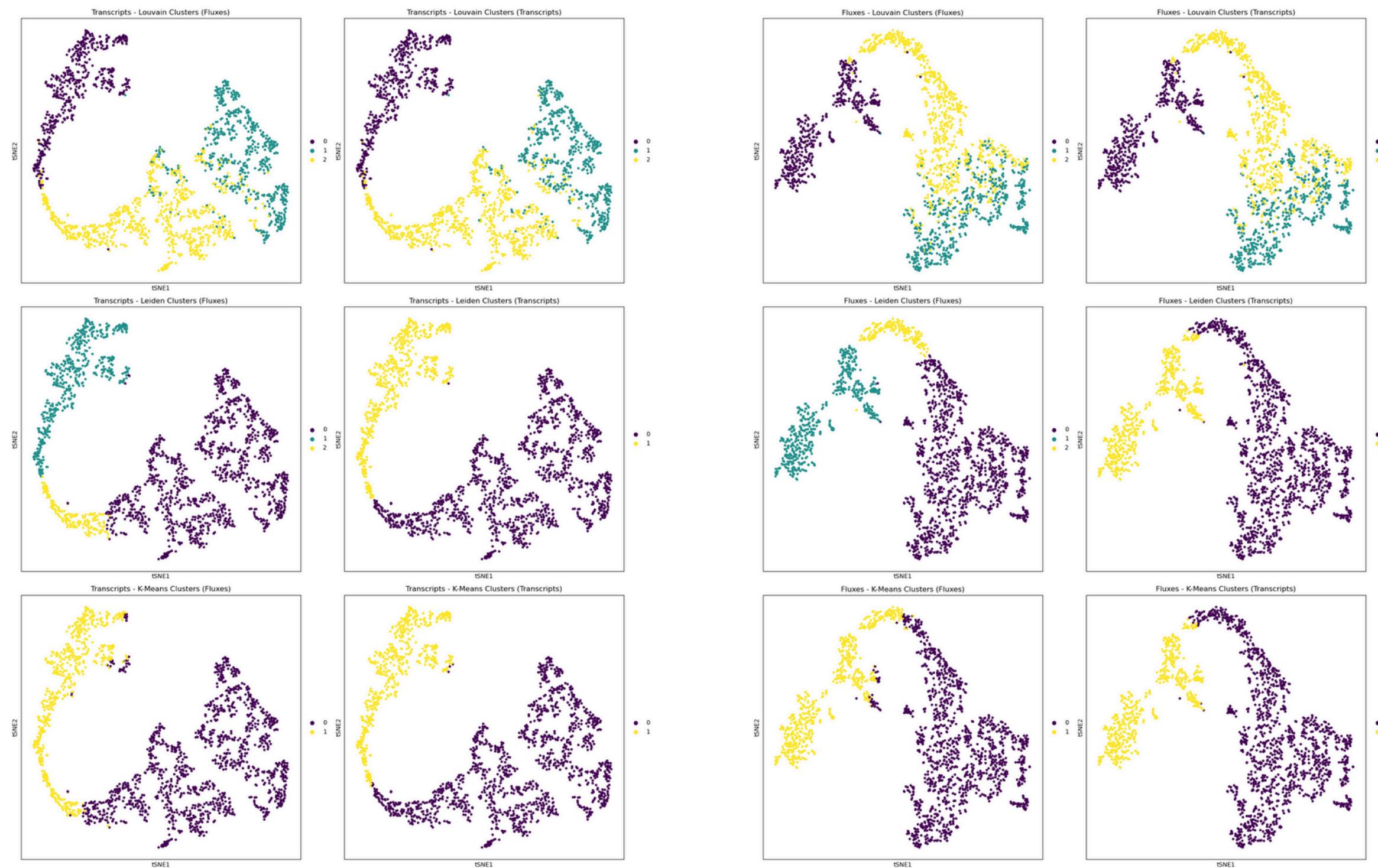
Fluxes clustering vs. transcriptional profiling - scFEA Integrated



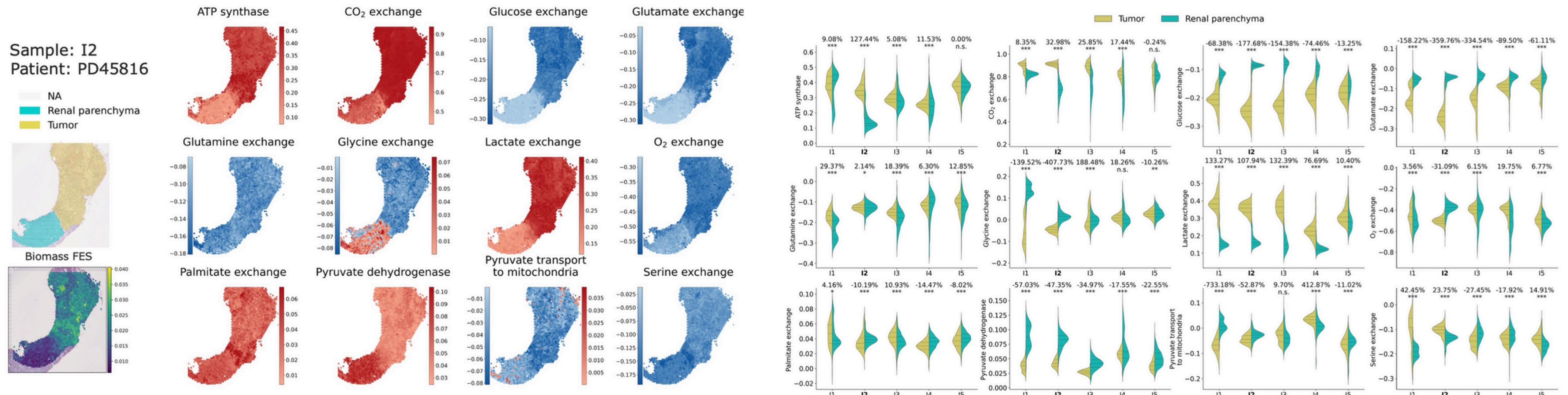
Fluxes clustering vs transcriptional profiling - MFP



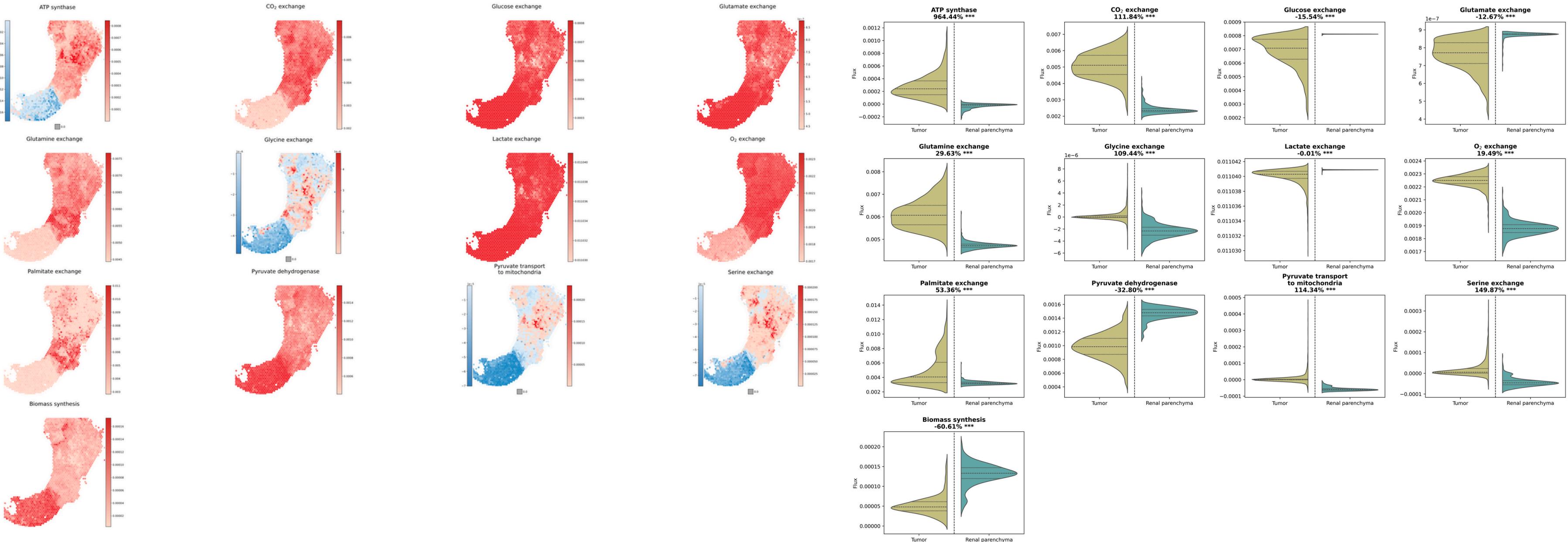
Fluxes clustering vs transcriptional profiling - MFP Independent Nodes



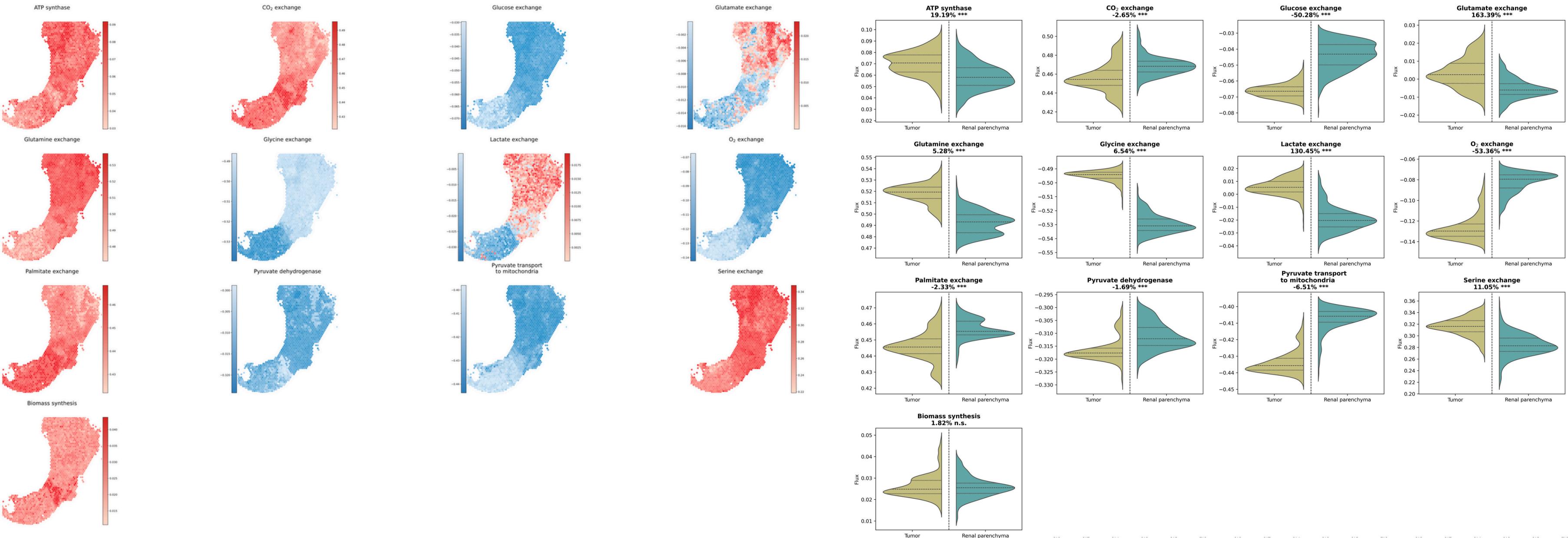
Distributions of fluxes - SpatialFBA



Distributions of fluxes – scFEA Integrated



Distributions of fluxes - MFP



Distributions of fluxes - MFP Independent Nodes

