Overview of computational analysis of spatially resolved omics data

Jovan Tanevski

Institute for Computational Biomedicine, Faculty of Medicine,
Heidelberg University and Heidelberg University Hospital

HEIDELBERG UNIVERSITY HOSPITAL
MEDIZINISCHE FAKULTÄT HEIDELBERG

Structure of the day

- One block lecture session in the morning divided in two parts (total 90 mins)
  - Overview of computational tasks for the analysis of spatially resolved data with focus on spatial transcriptomics
  - Identification of structural and functional relationships in spatially resolved data

- One block practical session in the afternoon divided in two parts (total 180 mins)
  - Analysis of spatial patterns of gene expression and functional annotation of sets of genes
  - Identification of functional relationships in spatially resolved data

Frame of the lecture

- Overview of the different tasks for the analysis of spatial data
- Get familiar with a subset of tools commonly used for them, some of their upsides and drawbacks
- What to use where and under what conditions

- Not an overview of technologies for spatial data acquisition
- Mostly limited to spatial transcriptomics

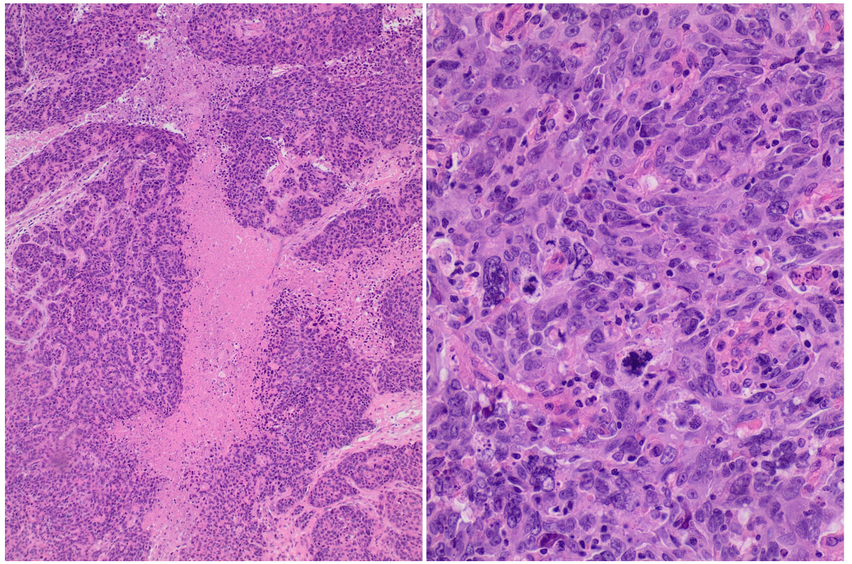
Evaluation

Written exam 15.05.2023 10:30 - 11:30

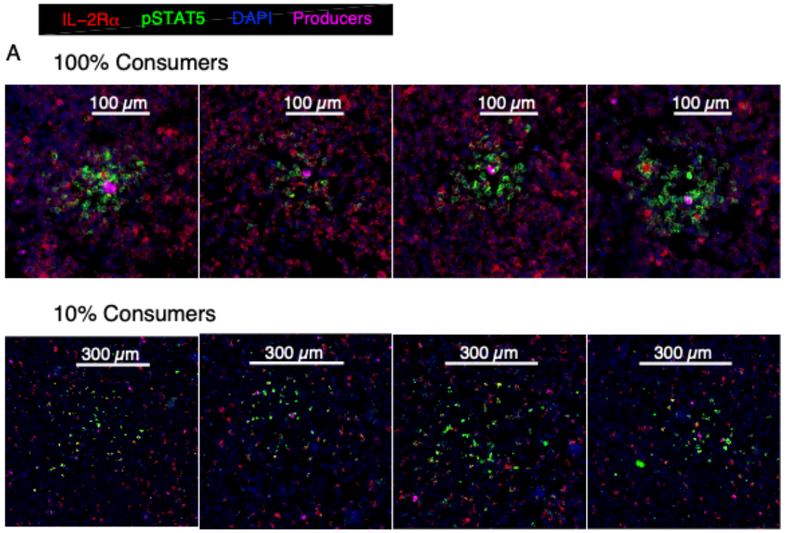
Location BioQuant CIP pool

Why spatial data

- Cells exist in space and not in a dissociated state
- Cells are arranged in specific way for a reason
- structure ↔ function
- Communication takes place in space
  - physical limitations

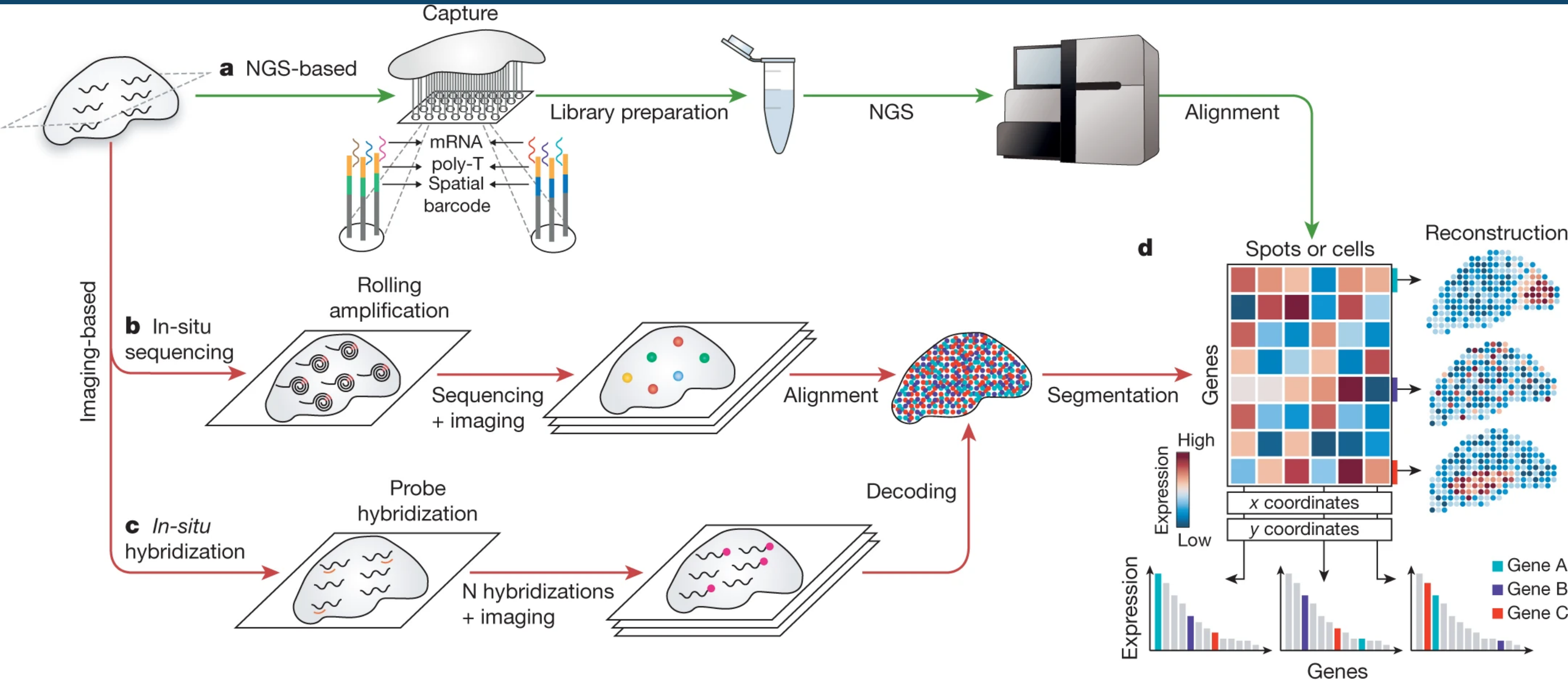


Wikimedia commons



Oyler-Yaniv et al. 2017, Immunity

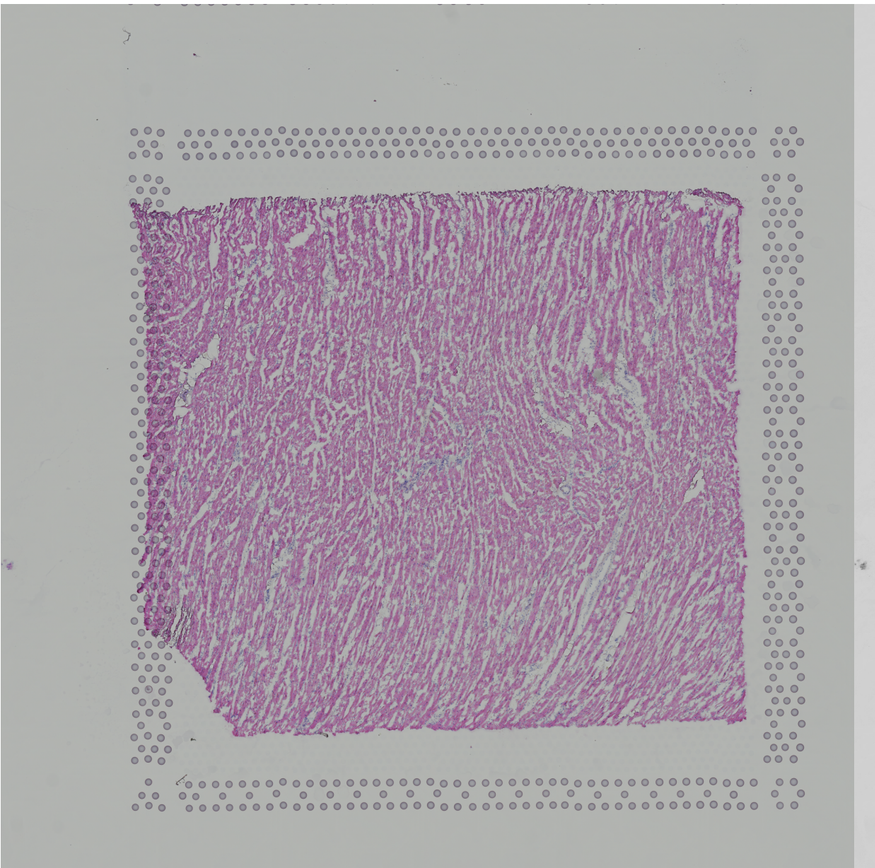
How to get spatial data



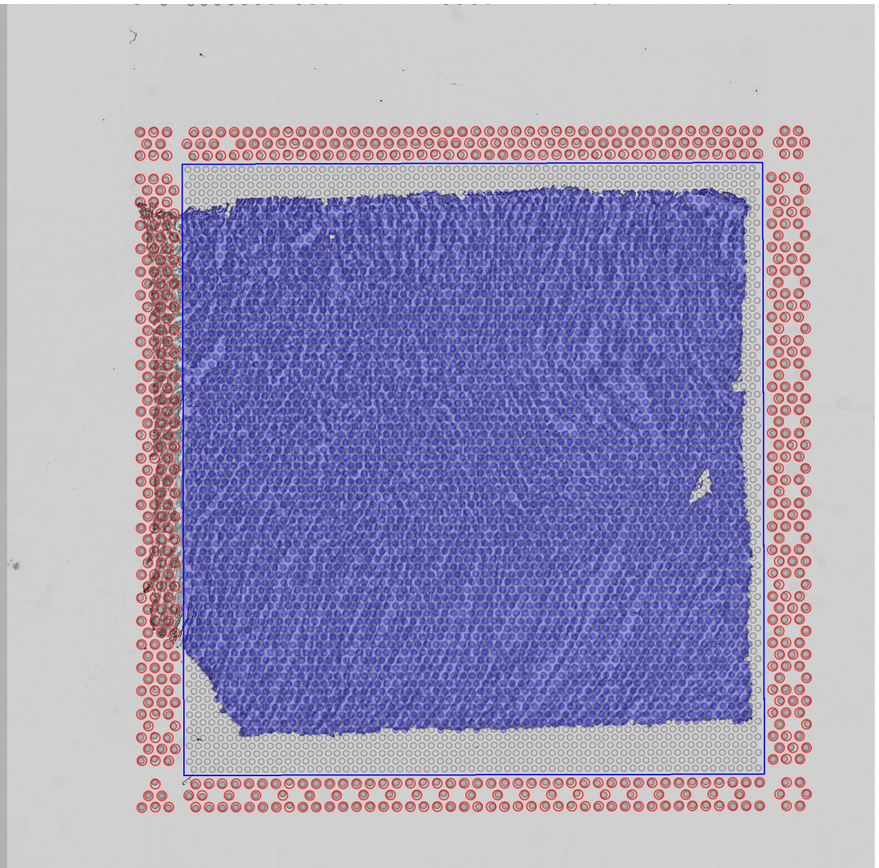
What does the data look like

Fresh frozen human heart tissue - 10x Genomics Visium

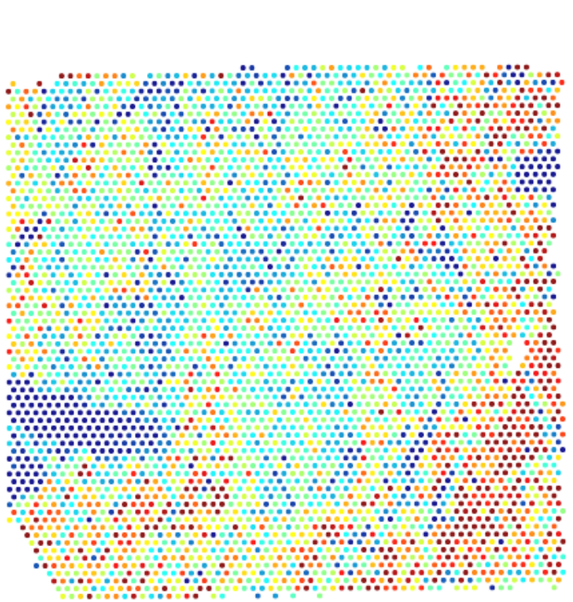
H&E



Spots - 70x70 55um diameter



UMI counts



Overview of some computational tasks

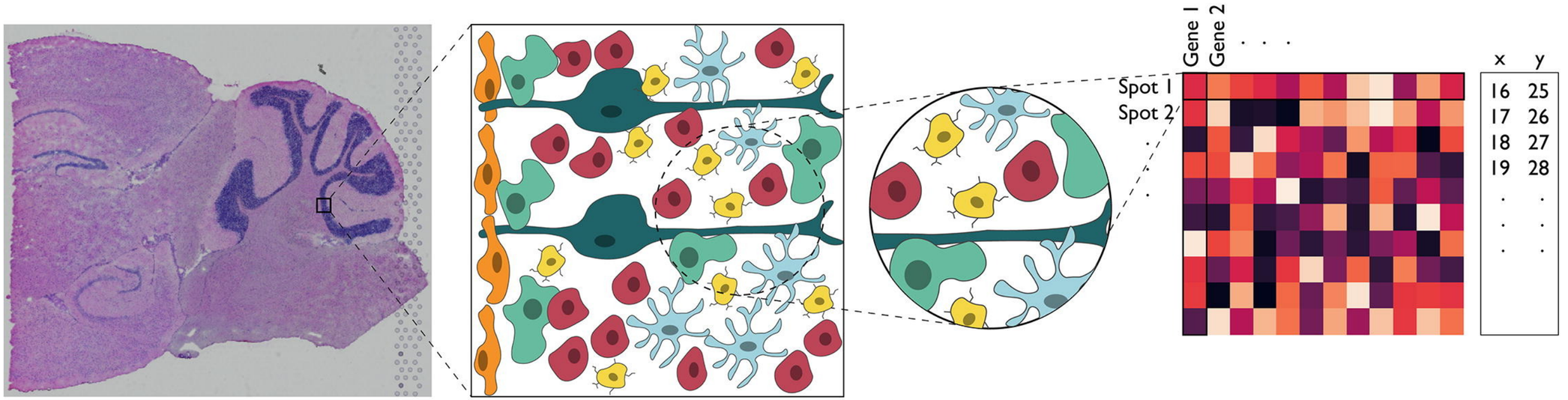
Data preprocessing usually by single cell pipelines

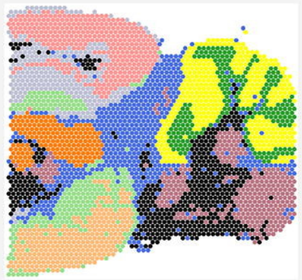
- Be careful! Does your data have the same properties as dissociated data?

What to do with the preprocessed data

- Clustering
- Spatial pattern detection
- Integration with richer dissociated data - deconvolution (low-resolution), localization (high resolution)
- Selection of regions of interest
- Characterization, annotation
- Analysis of spatial structure
- Analysis of communication/function

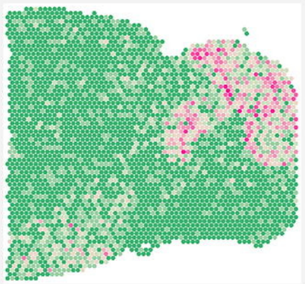
Overview of some computational tasks





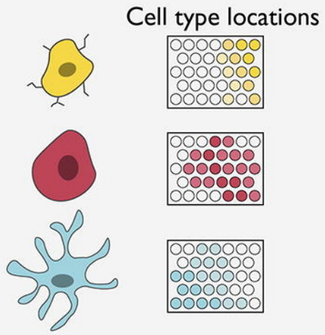
Spatial clustering

SpaCell
stLearn
BayesSpace
SpaGCN
.



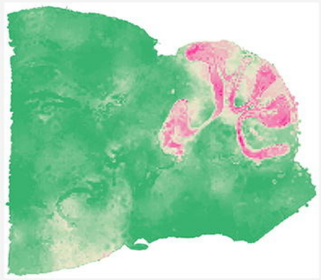
Spatially variable genes detection

Trendsceek
SpatialDE
SPARK
SpaGCN
.



Cell type deconvolution

Stereoscope
RCTD
SPOTlight
Cell2location
spatialDWLS



Enhancement of gene expression resolution

RCTD
BayesSpace
XFuse
.
.

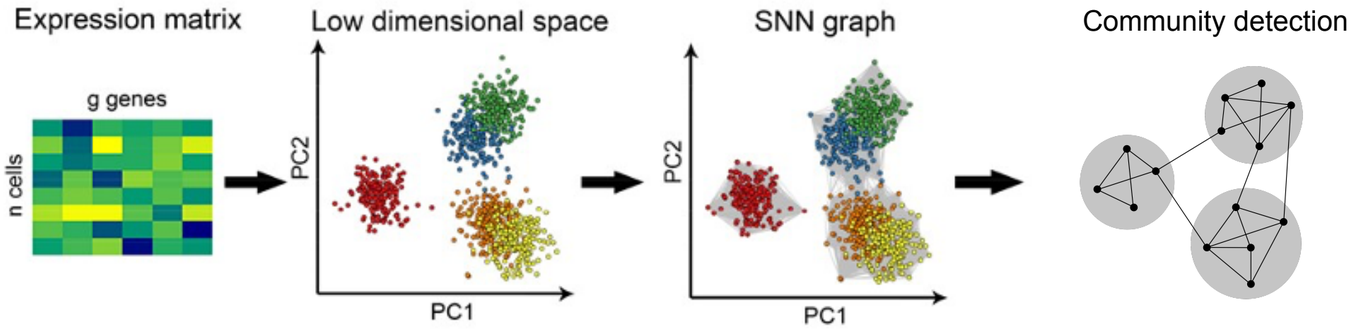
Identification of cellular interactions

Giotto
SpaOTsc
MISTY
.
.

Part 1 - clustering , patterns and integration

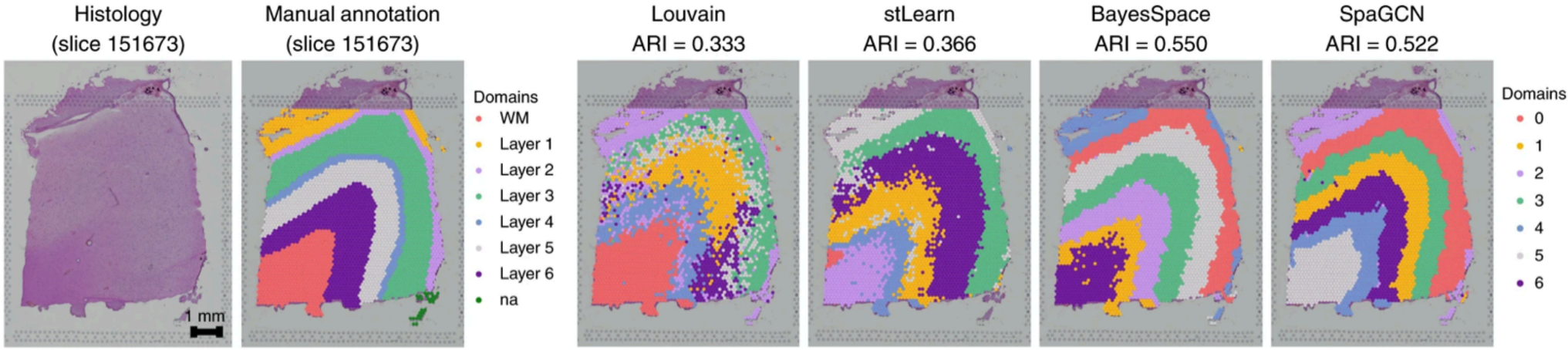
Clustering

- Why cluster?
- Usually follows approaches applied also to sc data
- Spatial component here unfortunately serves only for visualization



Jarvis & Patrick 1973 IEEE Trans. Comput.

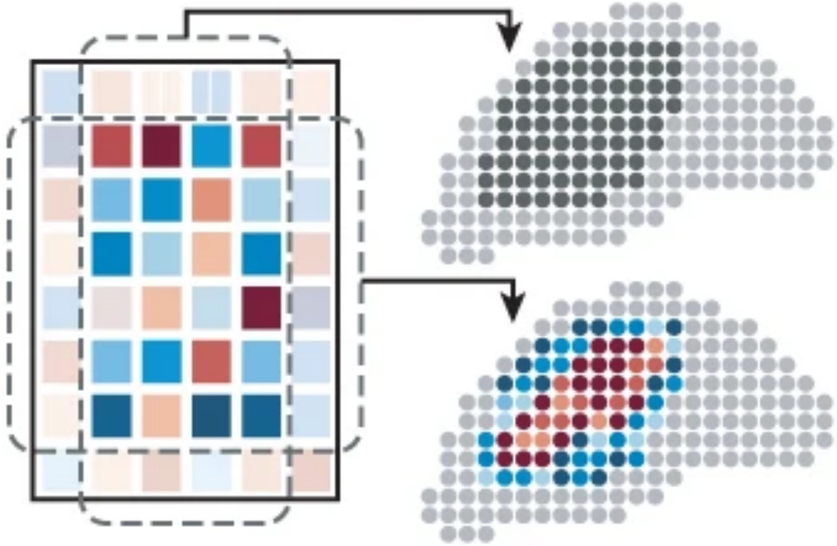
- There are also other methods that also take into account the spatial component of the data stLearn, SpaGCN, BayesSpace, still not very widely used



Hu et al. 2021 Nature Methods

Patterns

- Why identify patterns?
- Highly variable and differentially expressed genes → genes with significant spatial patterns, spatially variable features
- Initial motivation came from spatial analysis in geography, so we'll also look at the basics



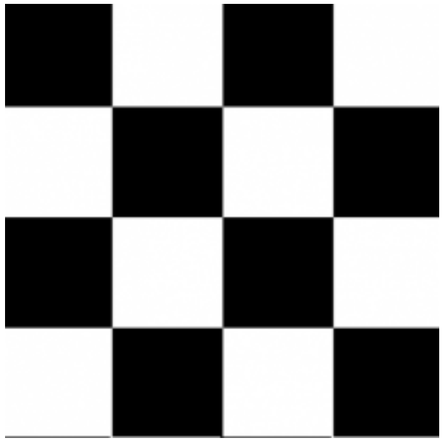
Spatial autocorrelation

Moran's I

$$I = \frac{N}{W} \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{N} (x_i - \bar{x})^2}$$

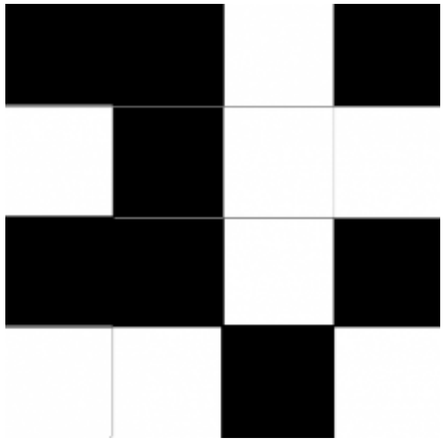
I = -1

Dispersion:
High values are never close to high and low values are never close to low



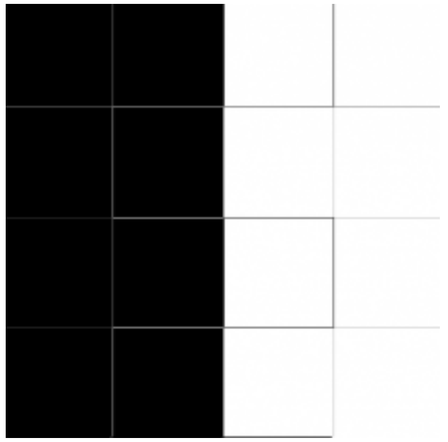
I = 0

Random



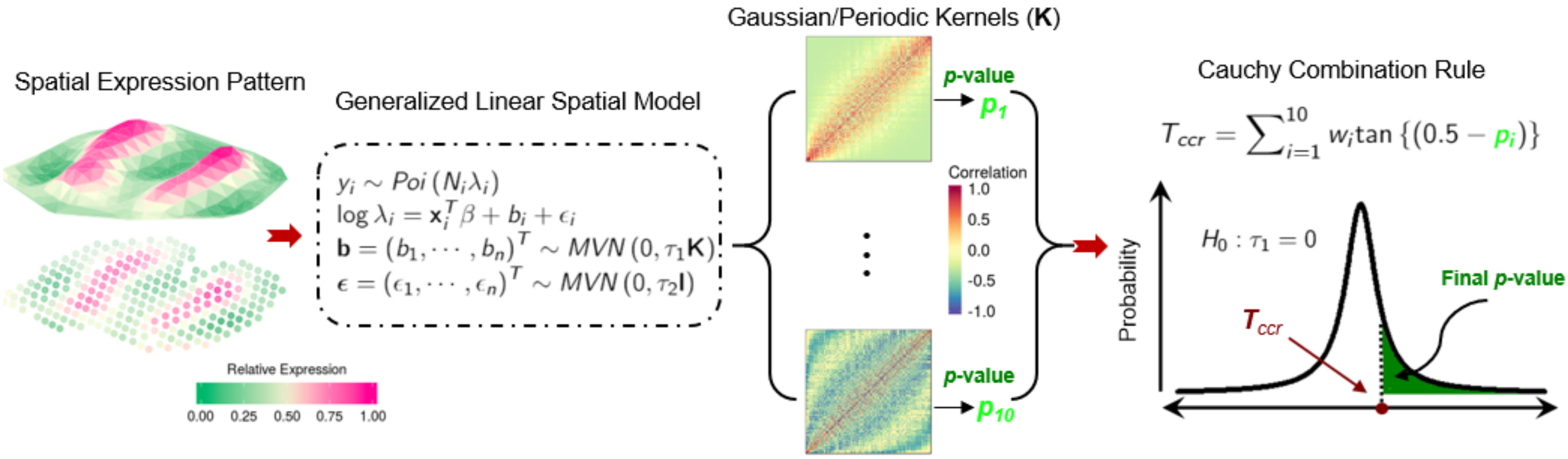
I = 1

Clustering:
High values are close to high and low values are close to low



Moran, P. A. P. 1950 Biometrika

A more advanced approach. Why? Why not?



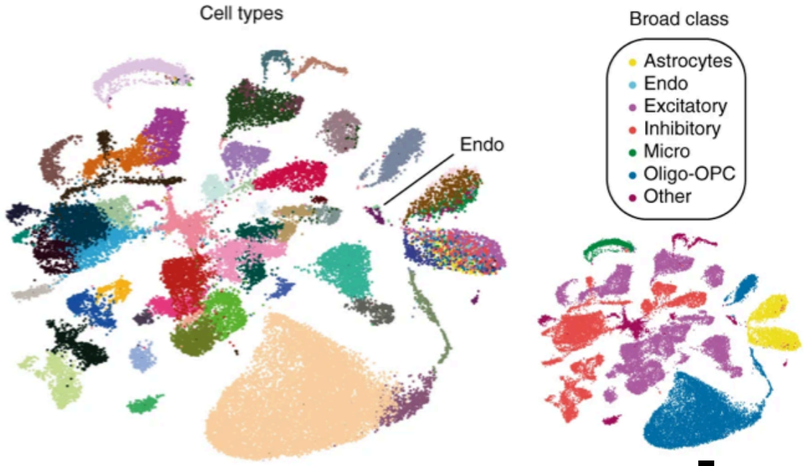
Sun et al. 2022 Nature Methods

Integration with scRNAseq data

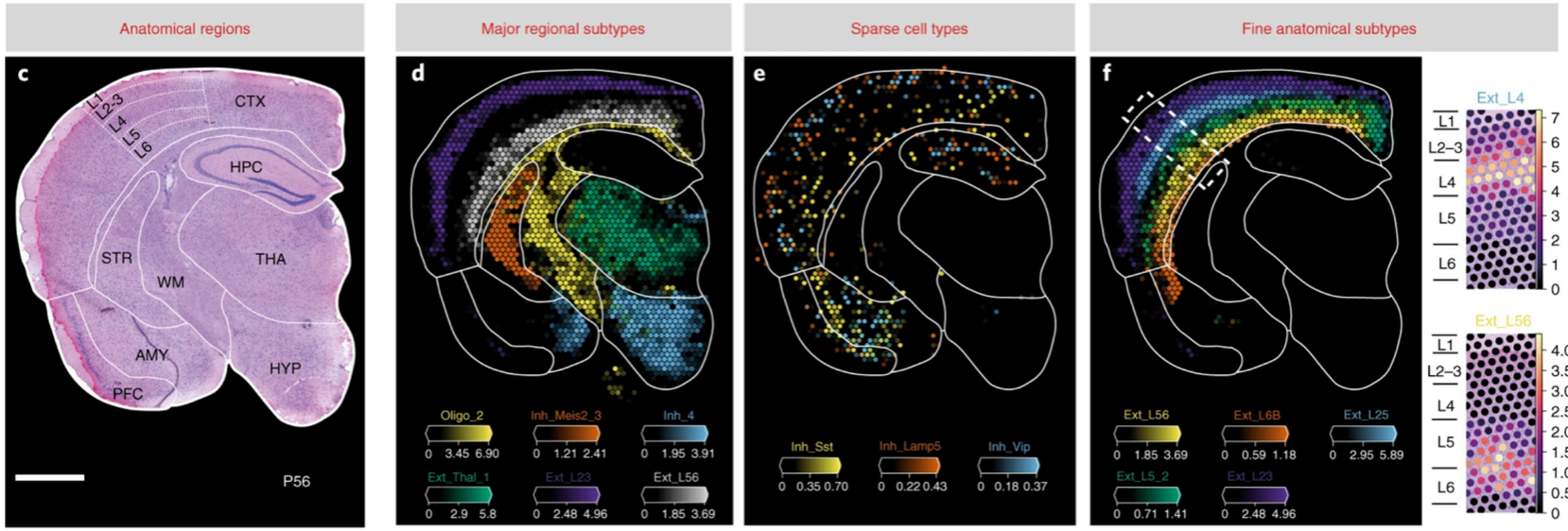
- Why integrate?

- Different tasks:
  - Label transfer
  - Deconvolution
  - Localization

sc(sn)RNAseq “atlas”



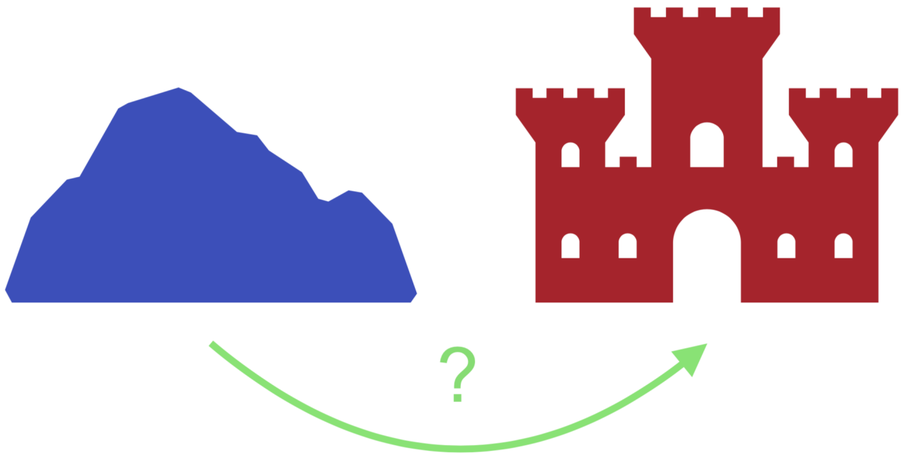
visium slide with deconvoluted cell-type abundance



Optimal Transport

- How to transport the soil during the construction of the building of forts and roads with minimal transport expenses?

- Nowadays we also call this cost the **Earth Mover's Distance** ➔ Cost of matching two probability distributions



Optimal Transport: Then and Now

Earth Move

Optimal Tr
Nob

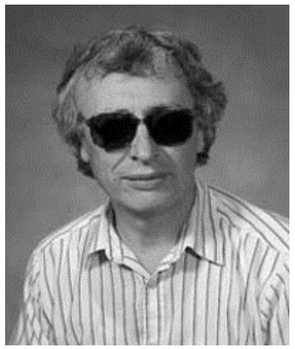
An algorithmi
Linear Progr

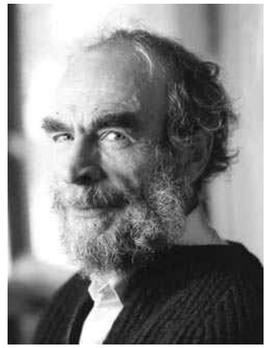
Wasserstein
measuring dist
probability di

His th
for gec

R
w
le

Approximate Optimal Transport
and its applications in ML



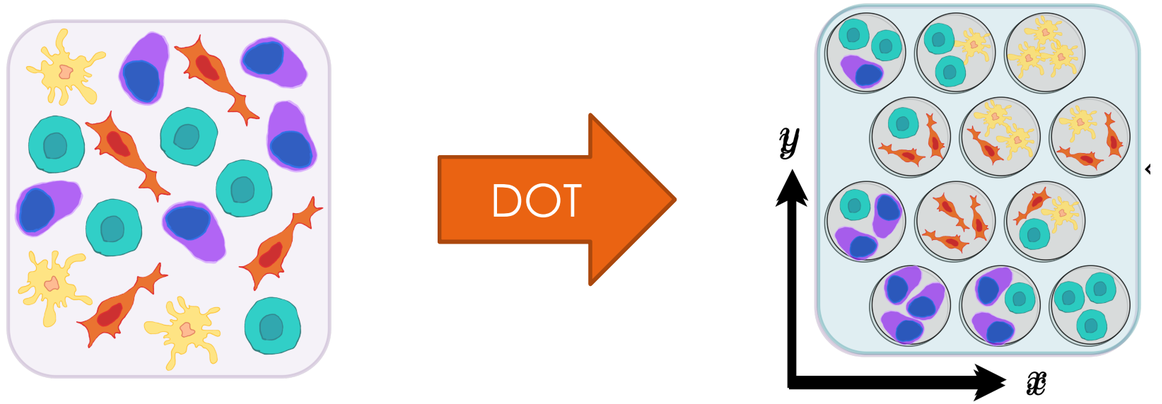


Monge	Kantorovich	Dantzig	Wasserstein	Gromov	Villani	Cuturi
1781	1939	1947	1970	1981	2003	2011

DOT

- A pile of annotated cells
- Locations in spatial omics
- **Move the annotations to locations in space**

- **Cost of moving?**
  - Mitigating platform effects by introducing scale-invariant distance functions
  - Cell type heterogeneity
  - Sparsity of transport map to accommodate different spatial resolutions
  - Abundance of cell types in the tissue.



DOT: Fast Cell Type **D**ecomposition of Spatial Omics by **O**ptimal **T**ransport

Rahimi et al. 2023 arXiv

DOT

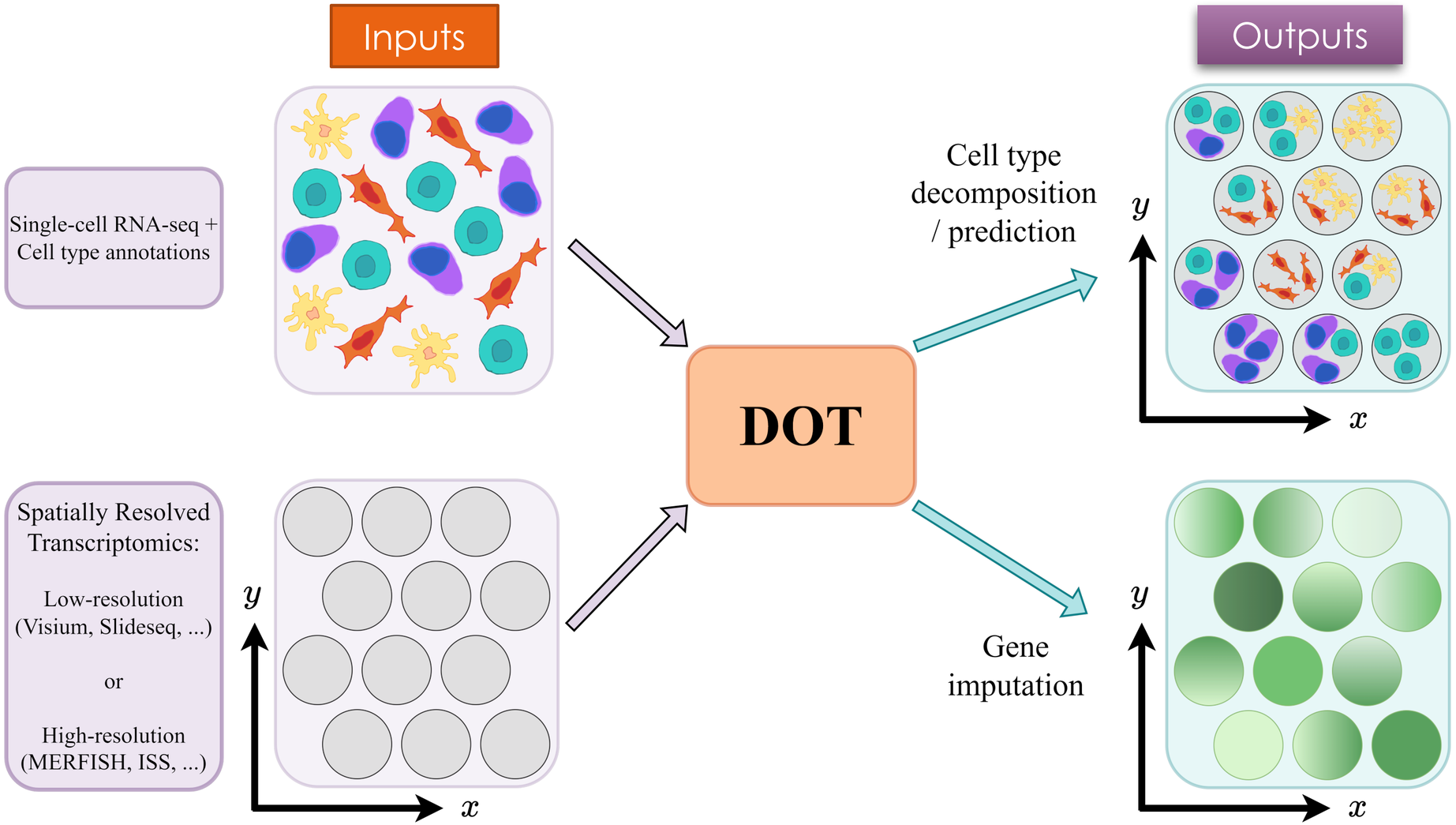
- **Mathematical model:**

$$\min d_{spots}(Y) + d_{genes}(Y) + d_{spatial}(Y) + \dots$$

$$\text{subject to} \sum_{c \in C} Y_{ic} = 1 \quad \forall i \in I, \qquad Y \geq 0$$

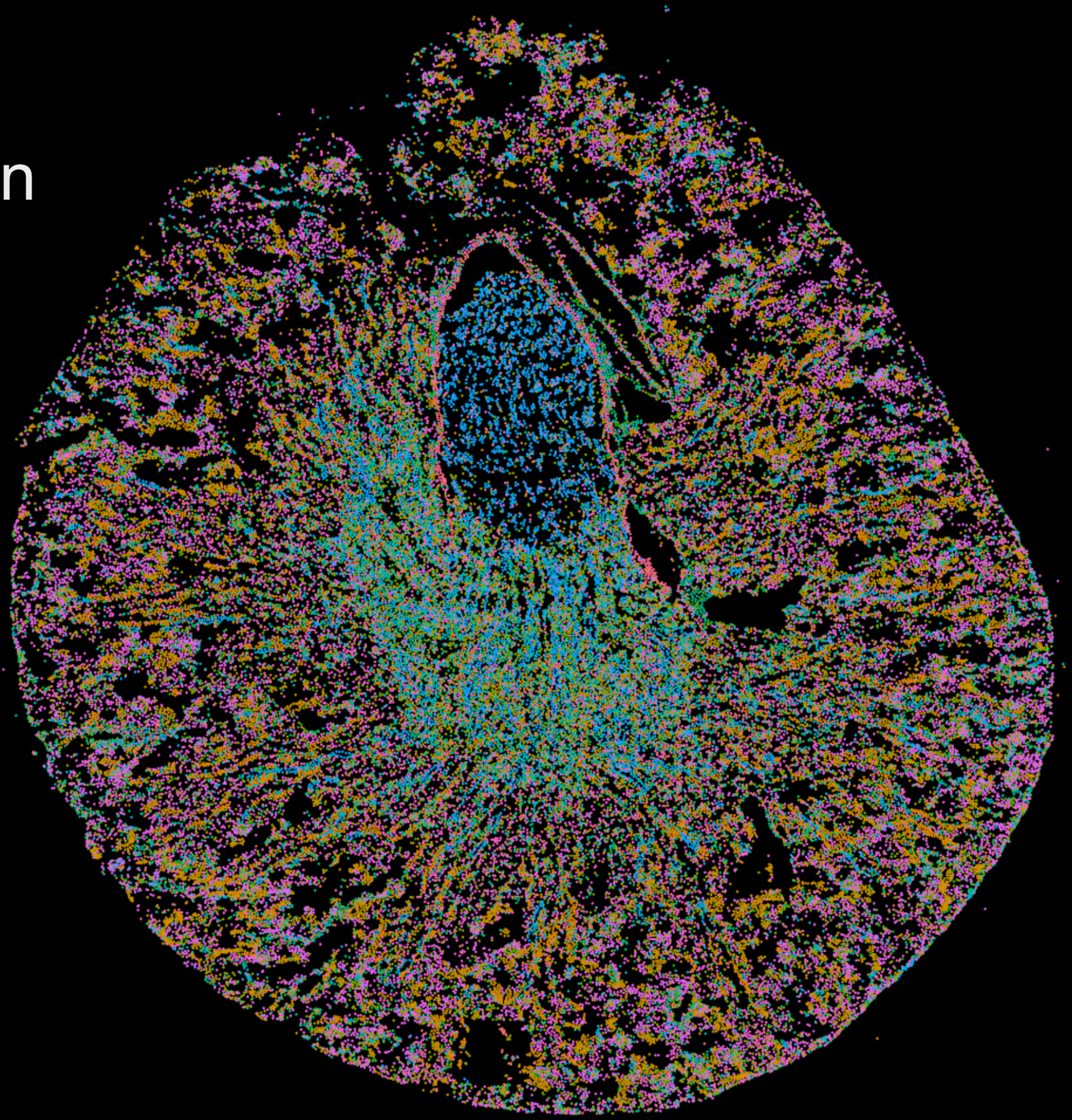
- The objective function is a composite of several individual distance (loss) functions. Each evaluates the quality of mapping from a different perspective:
  - d_{spots}: Measures dissimilarity of gene expressions mapped to spots from scRNA-seq and their own gene expressions.
  - d_{genes}: Measures dissimilarity of expression map of genes and the ones constructed using the reference scRNA-seq data.
  - $d_{spatial}$: Uses spatial information to encourage adjacent spots with similar expression profiles to have similar cell types.
- We want to find a mapping Y that is good for all of them.

How DOT Works



DOT:
Cell Type Prediction

- Female Mouse Kidney
- ISS:
  - 86,255 cells
  - 190 genes
- snRNA-seq:
  - 26 cell types
- DOT:
  - 178 clusters
  - ~ 10 minutes

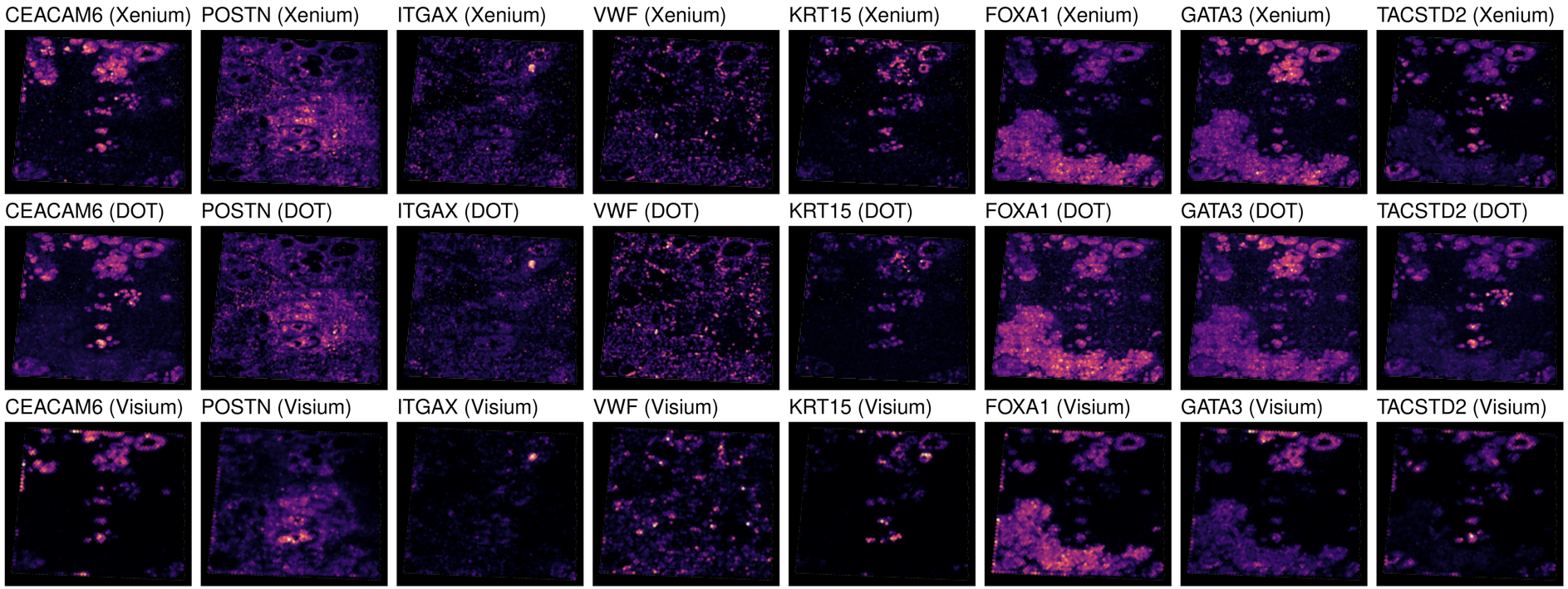


DOT: Gene Expression Prediction

High-resolution (Ground truth)

Mapped to high resolution via DOT

Low-resolution (Secondary ground truth)

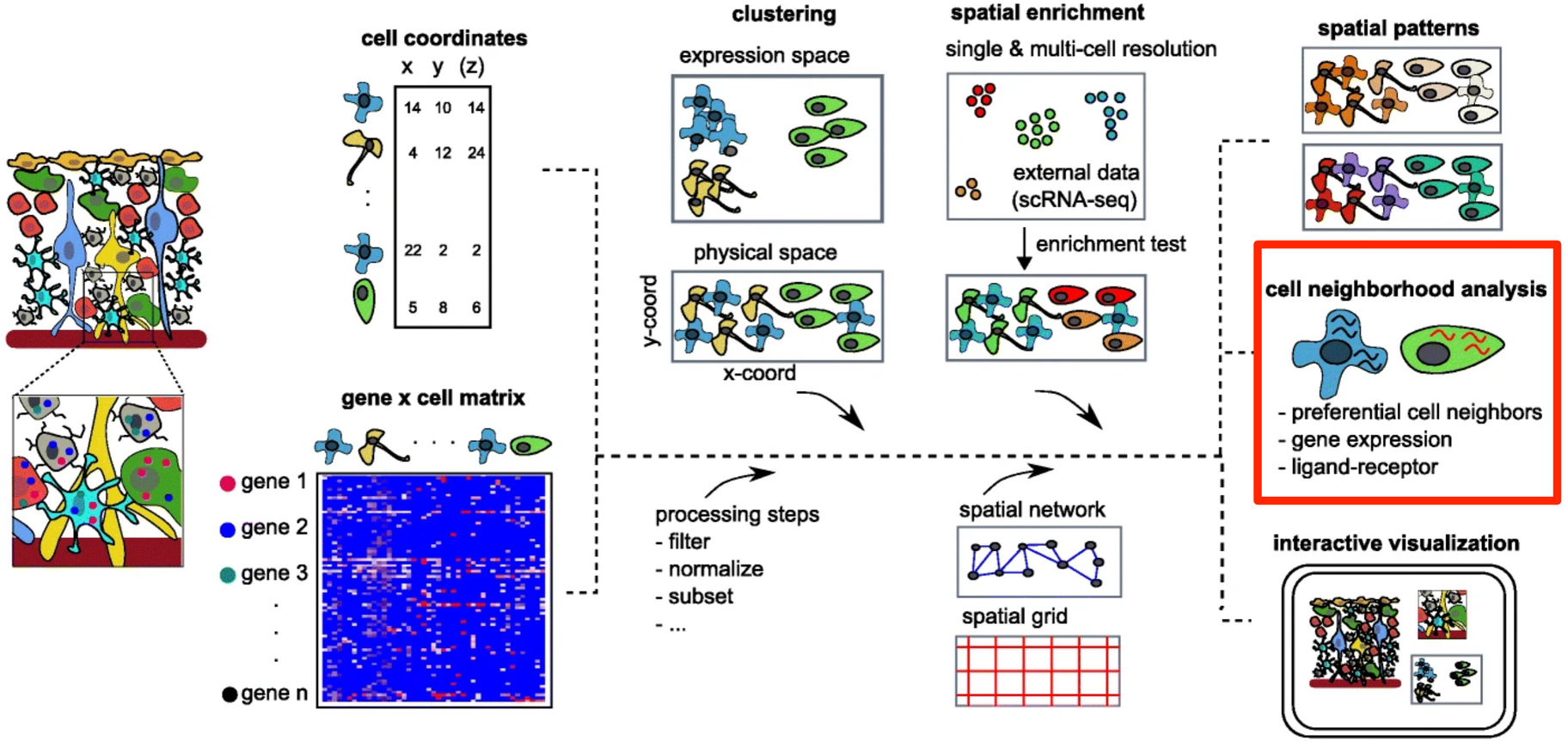


Part 2 - analysis of structure and function

Neighborhood analysis

Toolboxes for analysis of spatial data

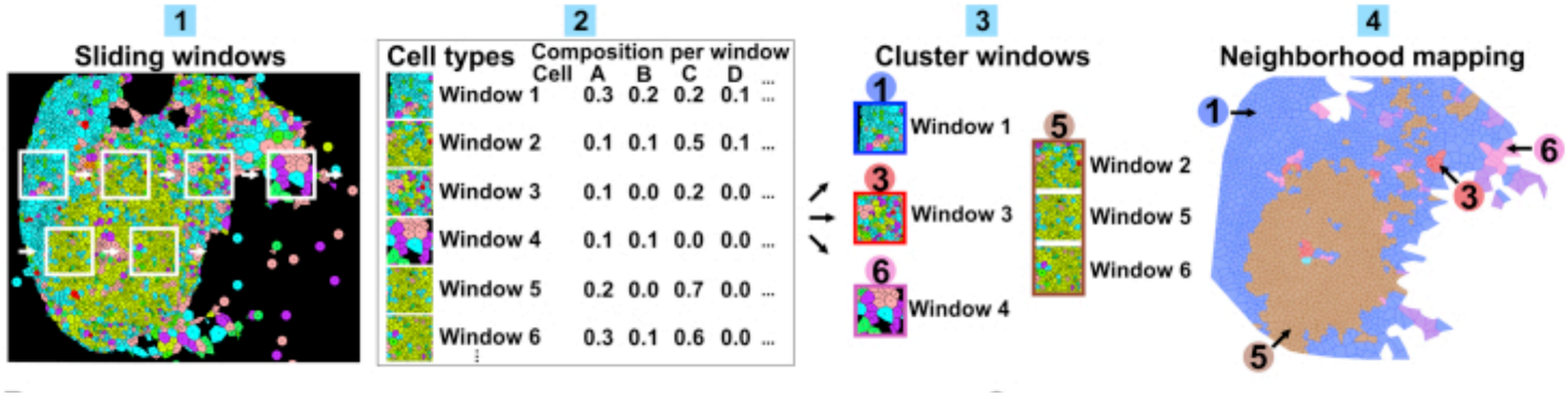
- Some focused on a specific technology - histoCAT
- Others try to be more general and incorporate several tasks - Giotto



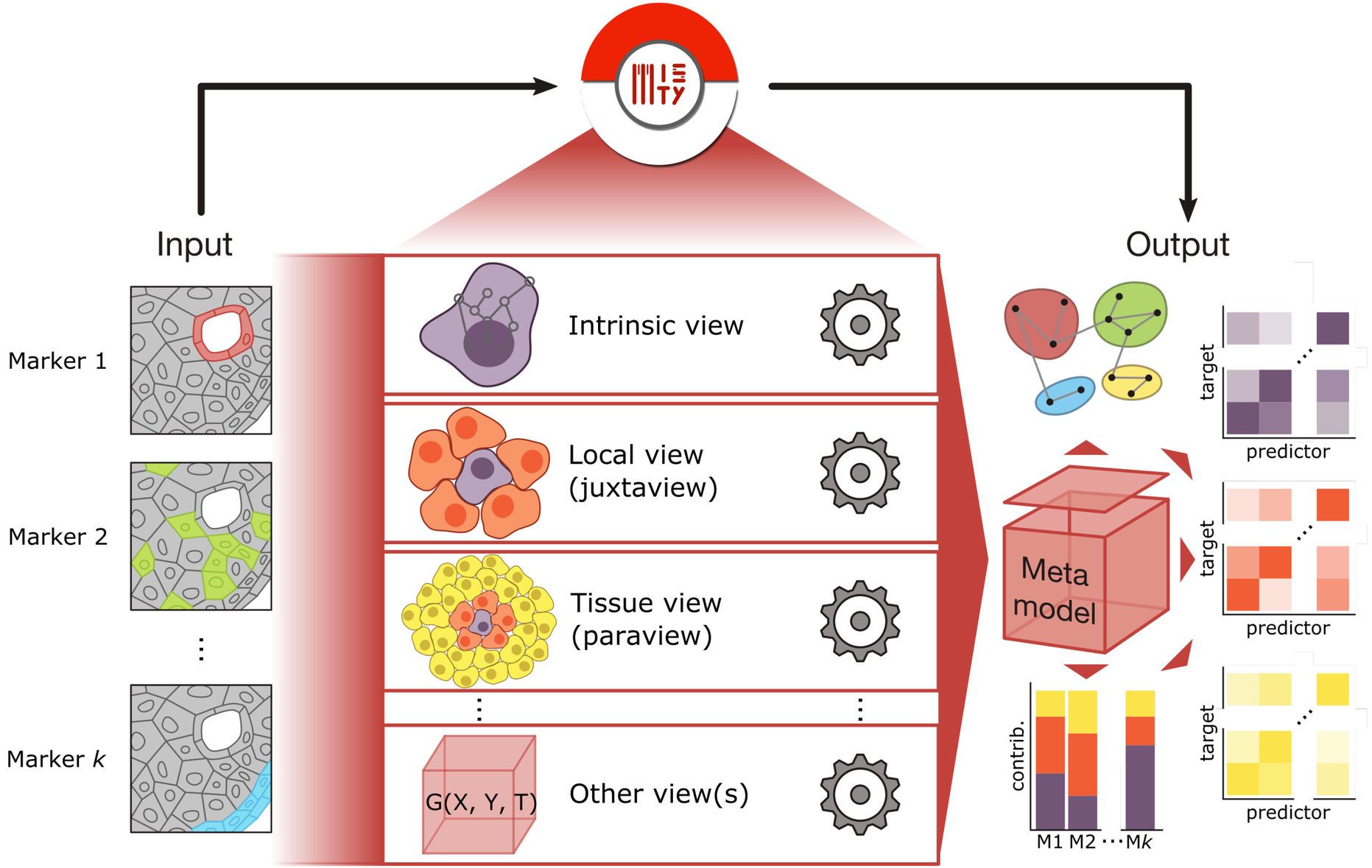
Dries et al. 2021 Genome Biology

Niches and cellular neighborhoods

- Identification of local structural patterns
- Focusing on cell types
- iNiche, SpatialLDA, Cellular Neighbourhoods
- By extension also GNN based methods (require large amounts of data)

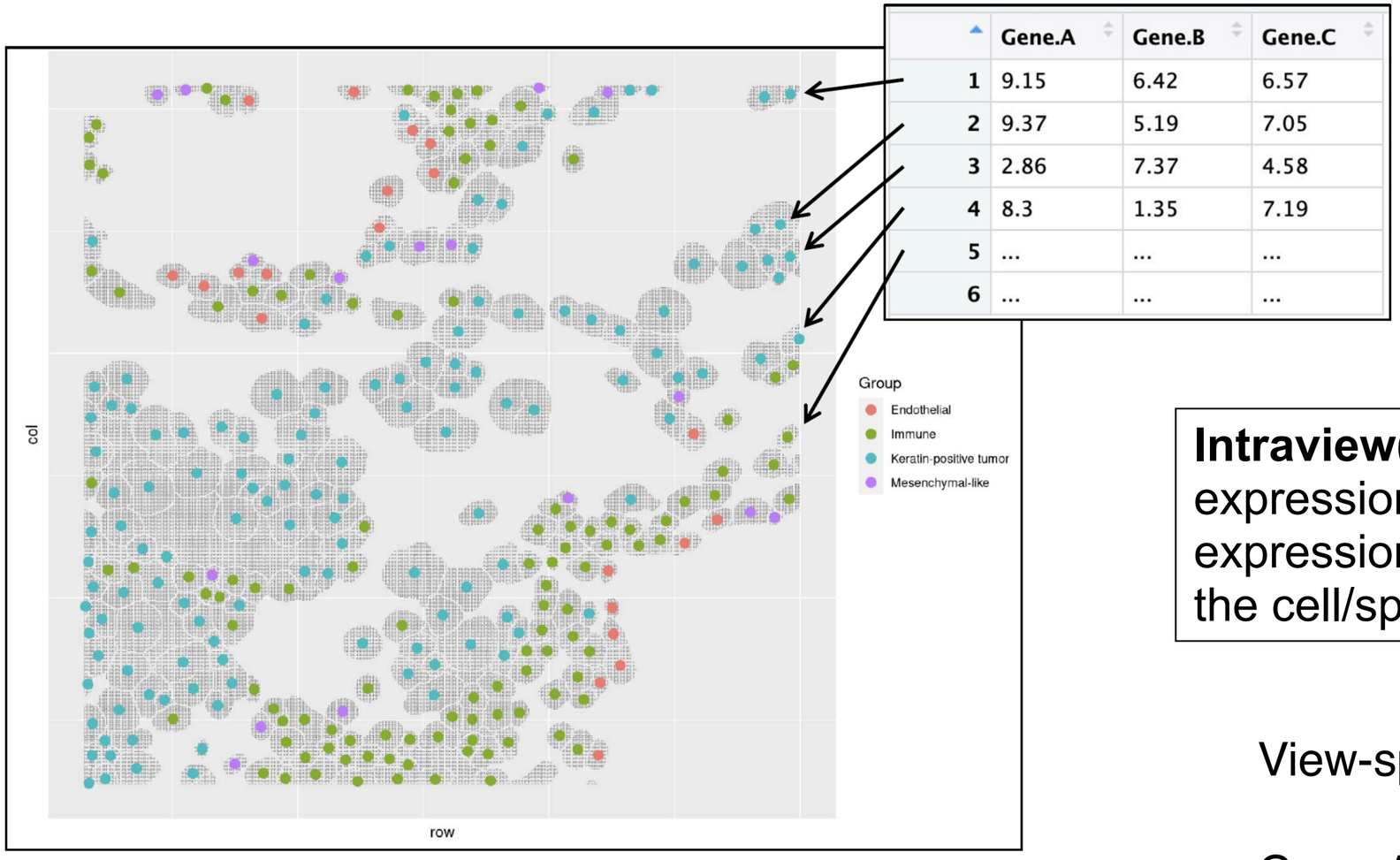


MISTy



Tanevski et al. 2022 Genome Biology

Intraview



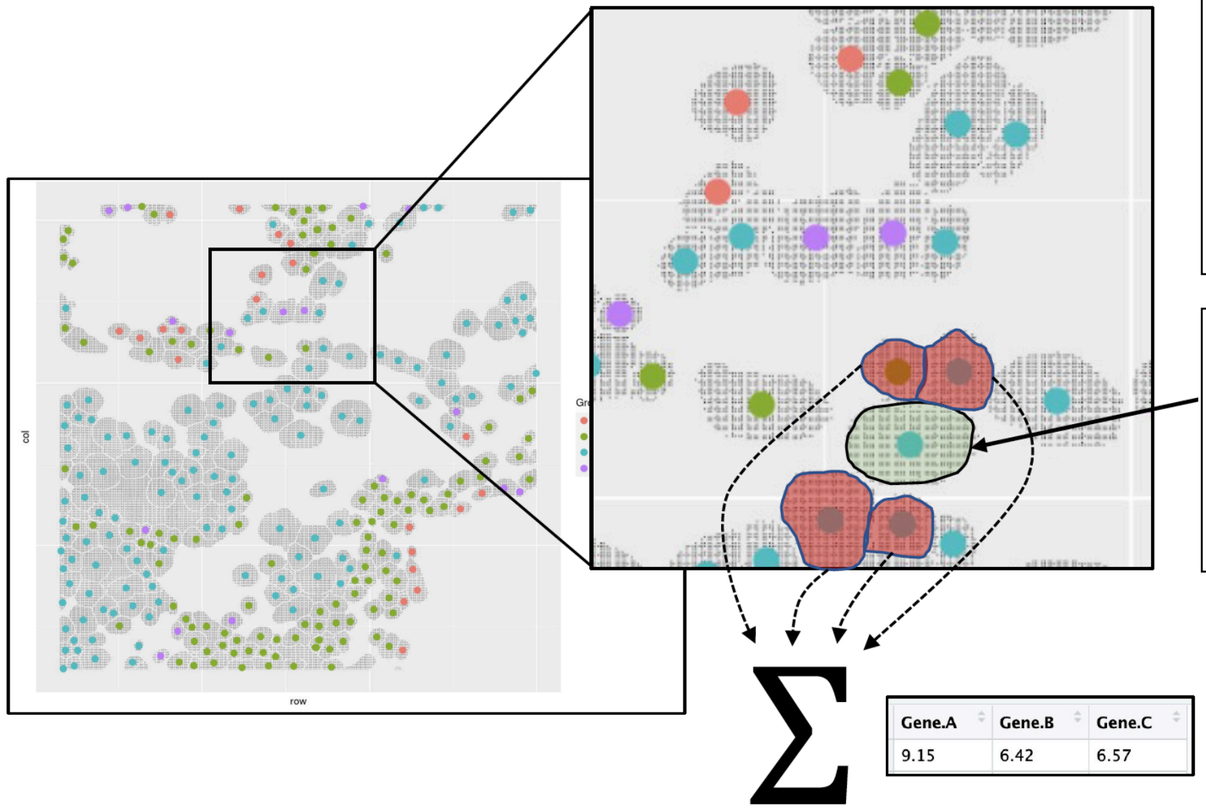
	Gene.A	Gene.B	Gene.C
1	9.15	6.42	6.57
2	9.37	5.19	7.05
3	2.86	7.37	4.58
4	8.3	1.35	7.19
5	...	...	...
6	...	...	...

Intraview(baseline): Explain the expression of a gene by the expression of all other genes within the cell/spatial unit.

View-specific model

Gene.A = **f**(Gene.B, Gene.C)

Juxtaview



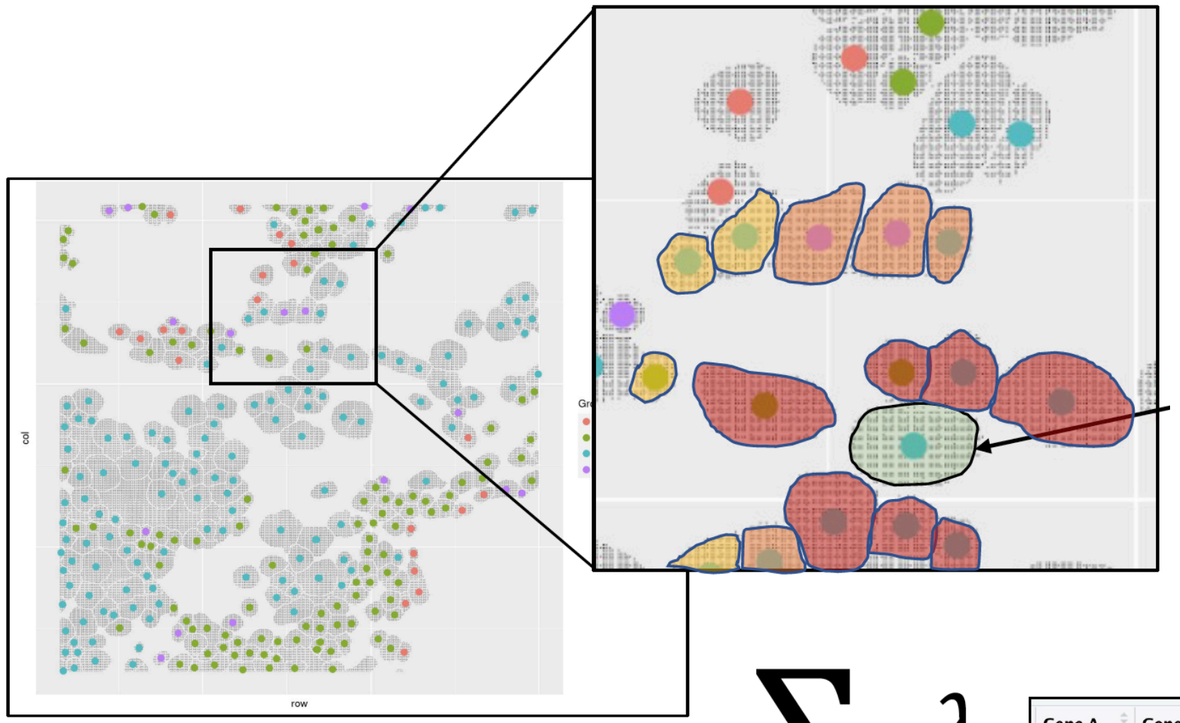
Gene.A	Gene.B	Gene.C
9.15	6.42	6.57

Sum over all neighboring cells

Question: Can we explain the expression of a gene better if we take into account information coming from different spatial contexts?

Juxtaview: Explain the expression of a gene by the expression of all other genes in its immediate neighborhood.

Paraview



Question: Can we explain the expression of a gene better if we take into account information coming from different spatial contexts?

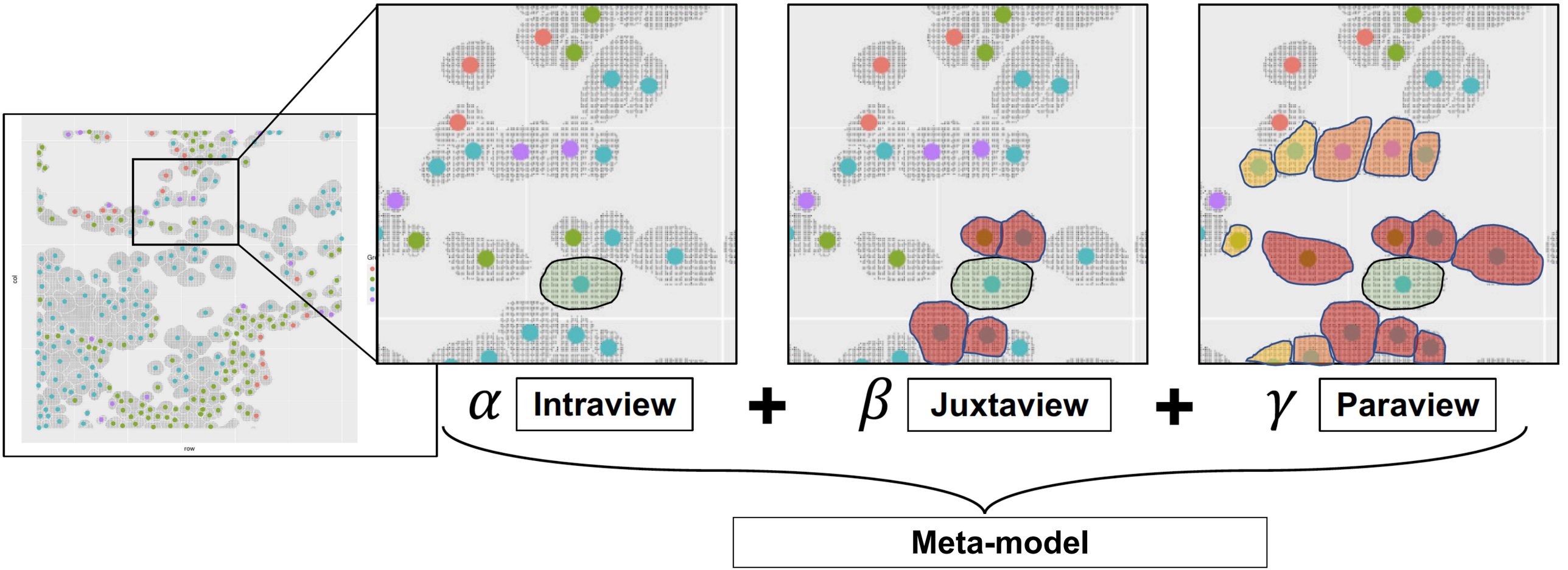
Paraview: Explain the expression of a gene by the expression of all other genes in the (weighted) broader spatial context.

$$\sum \lambda$$

Weighted sum

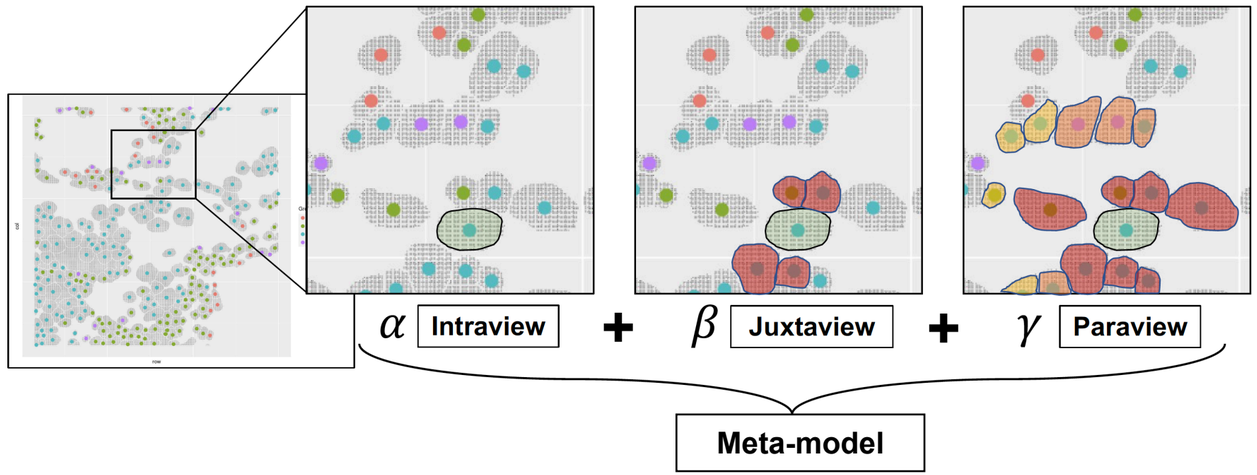
Gene.A	Gene.B	Gene.C
9.15	6.42	6.57

Meta-model



Output is composed of the view-specific models plus the meta model

MISTy pipeline

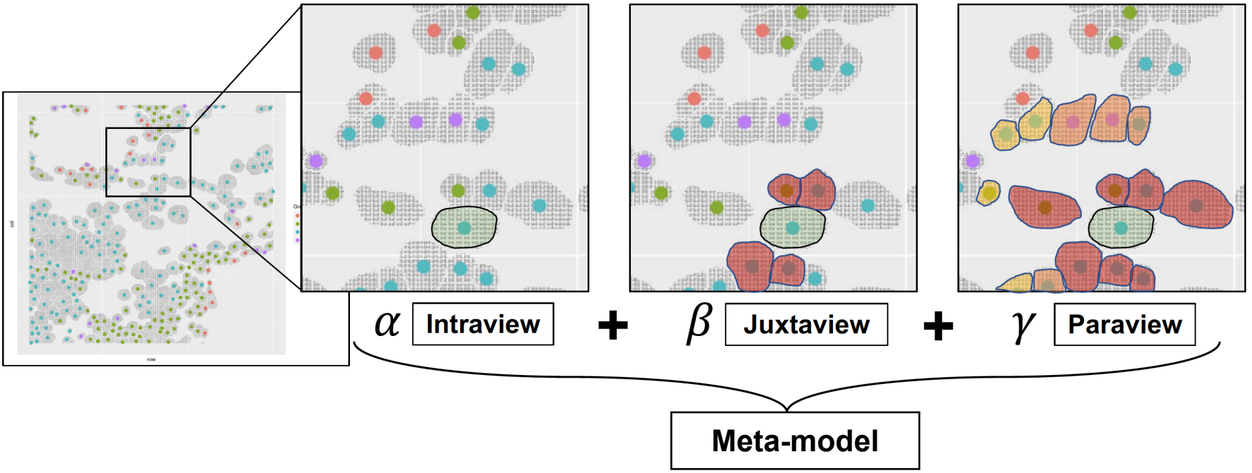


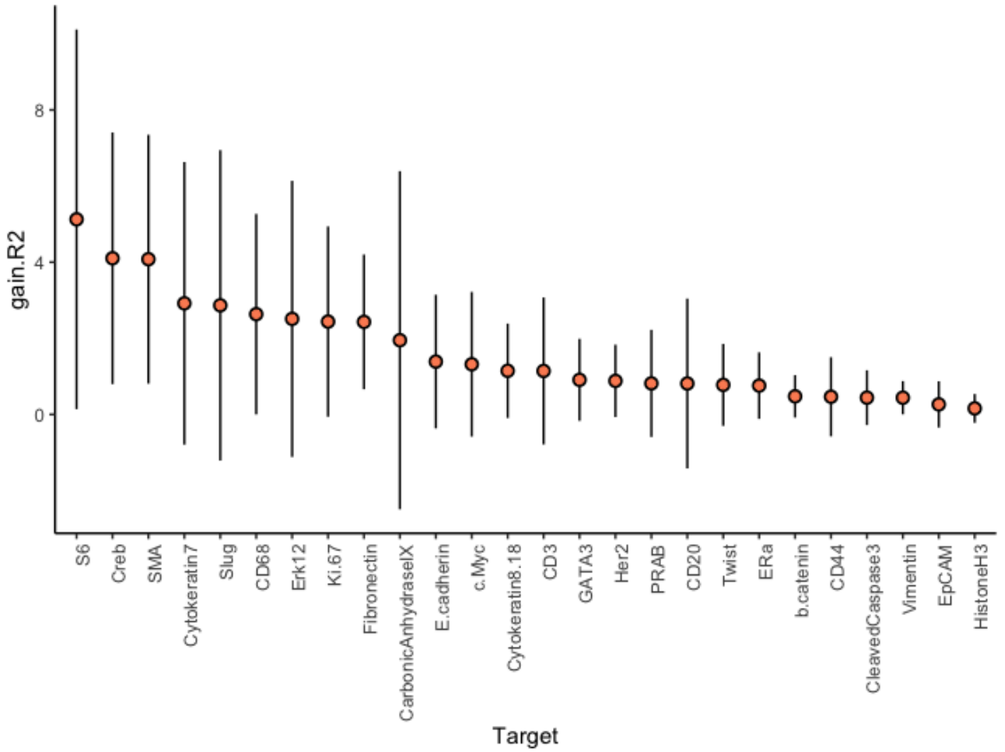

```
views <- create_initial_view(expr) %>% add_juxtaview(pos, threshold) %>% add_paraview(pos, l)

output <- run_misty(views, results.folder = “sample_result”)

misty.results <- collect_results(output)
```

https://github.com/saezlab/misty_pipelines

Results

```
misty.results %>% plot_improvement_stats() %>% plot_view_contributions()
```

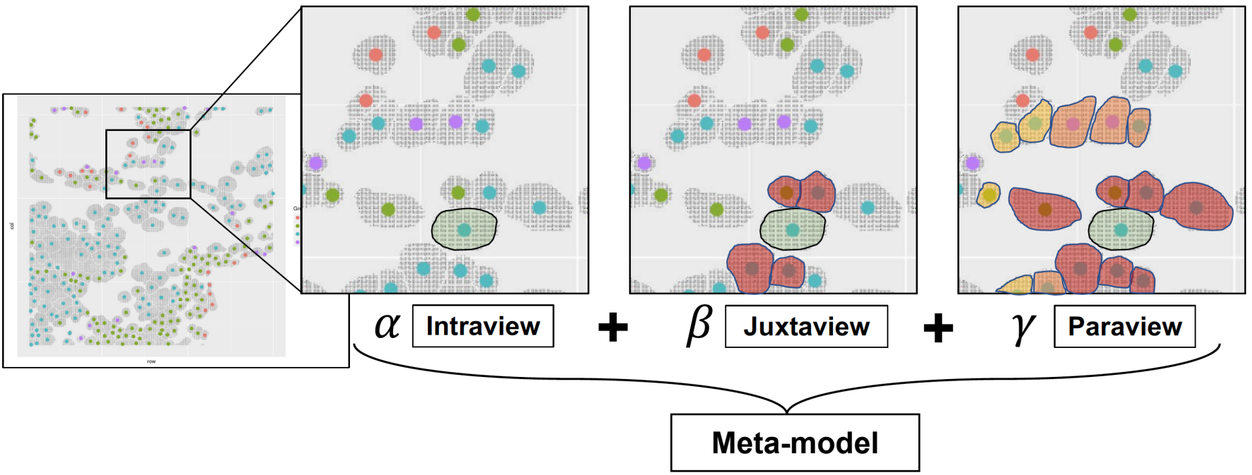


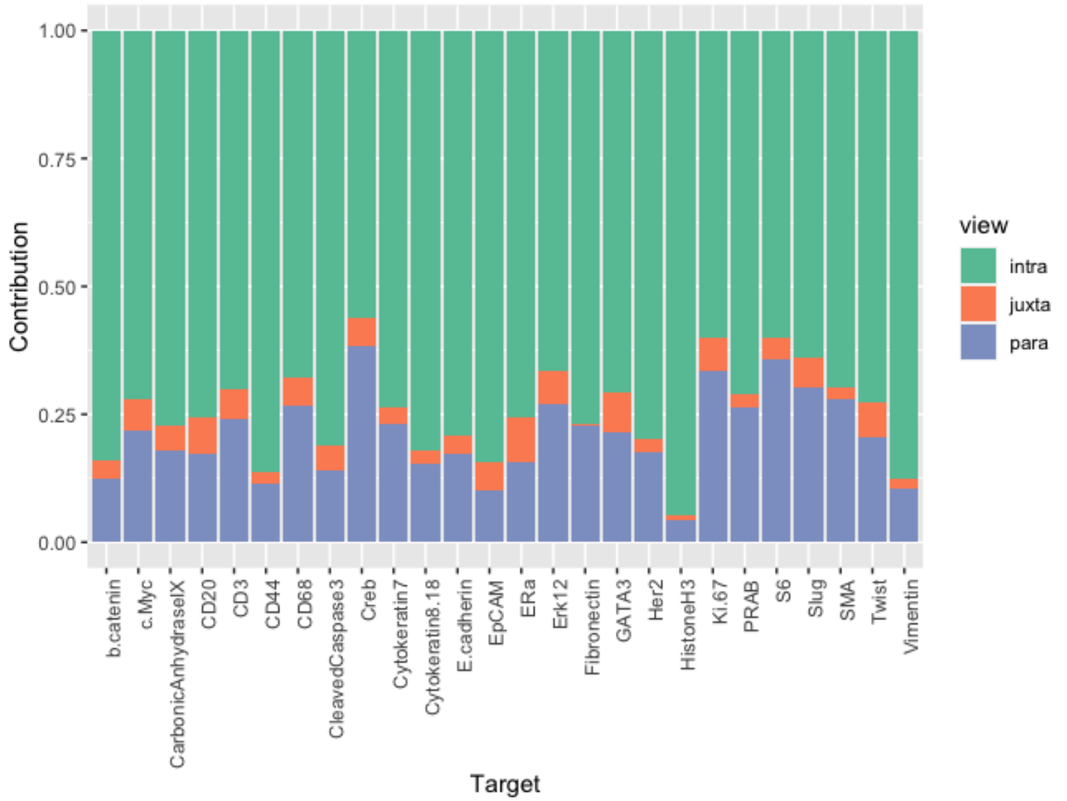


1. How much can the intercellular (spatial) context explain expression (in contrast to intracellular)?

Results

```
misty.results %>% plot_improvement_stats() %>% plot_view_contributions()
```

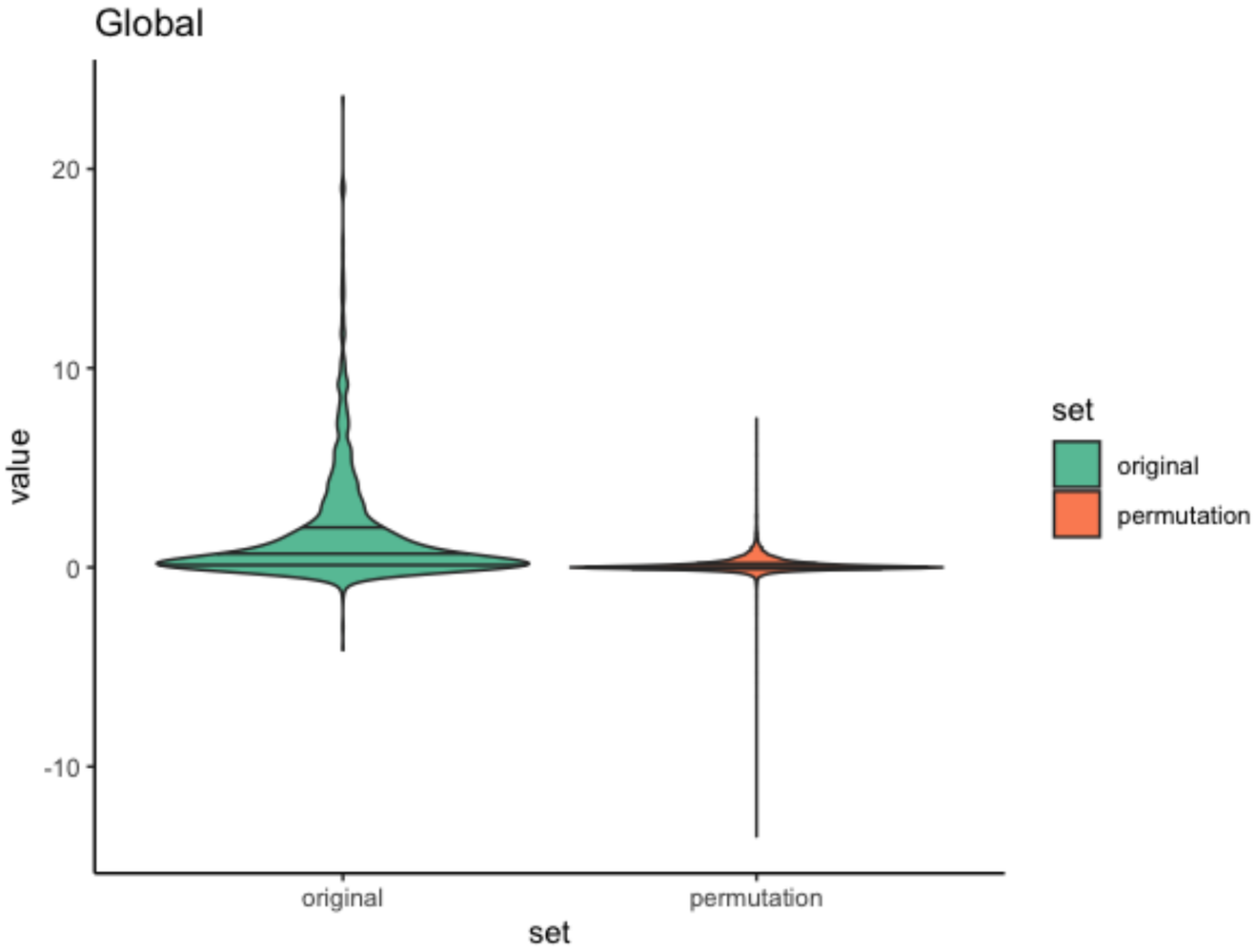




2. How much do different view components contribute to explaining the expression?

Permutation analysis

Permute the cell locations and rerun MISTy 10 times

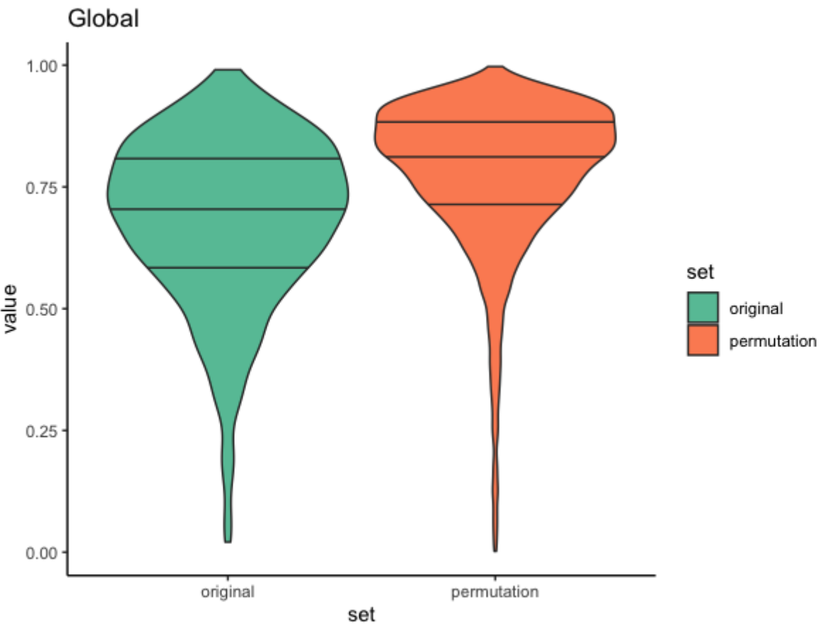


Distribution of gain in variance explained across all targets and samples

Mann-Whitney U
$p \approx 0$

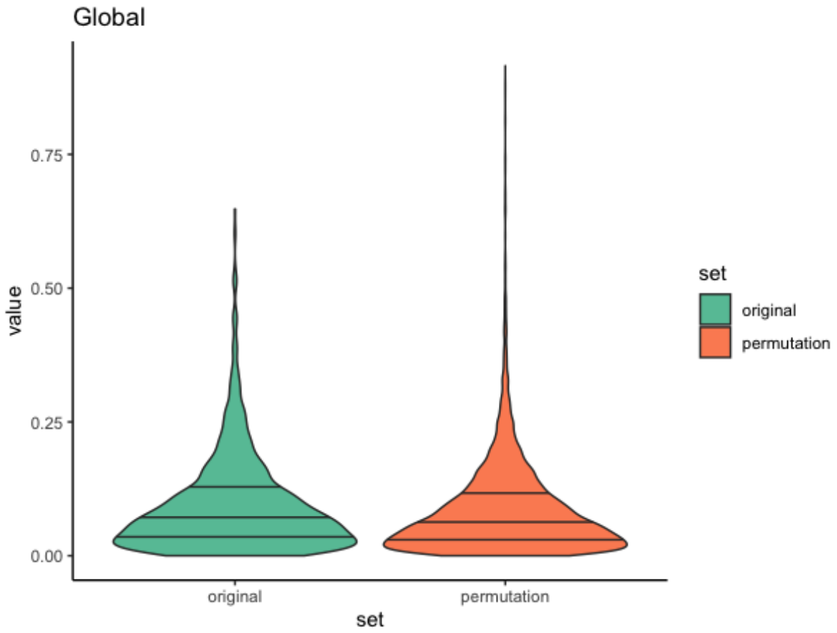
Distribution of view contributions

intra



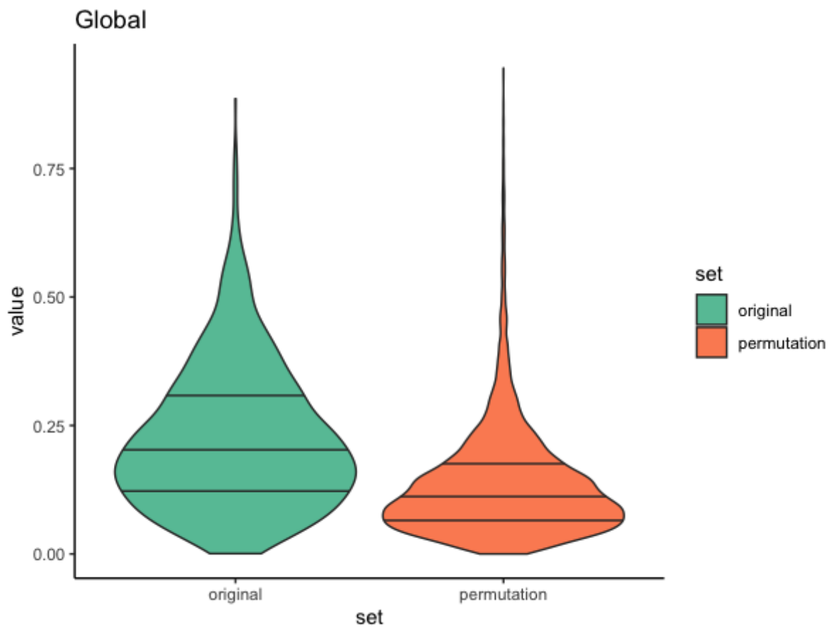
Mann-Whitney U
$p \approx 0$

juxta



Mann-Whitney U
$p < 10^{-2.5}$

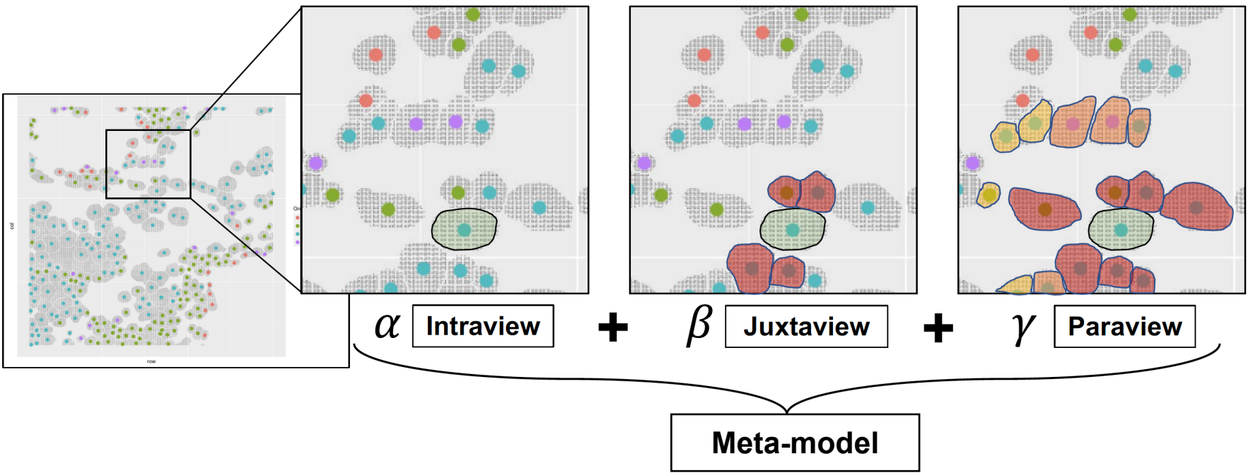
para

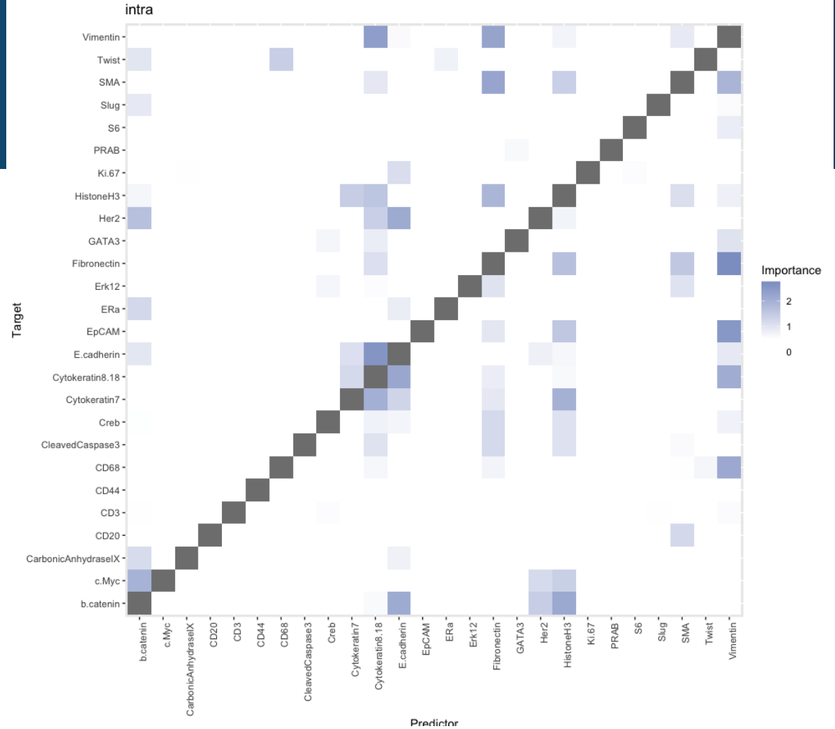


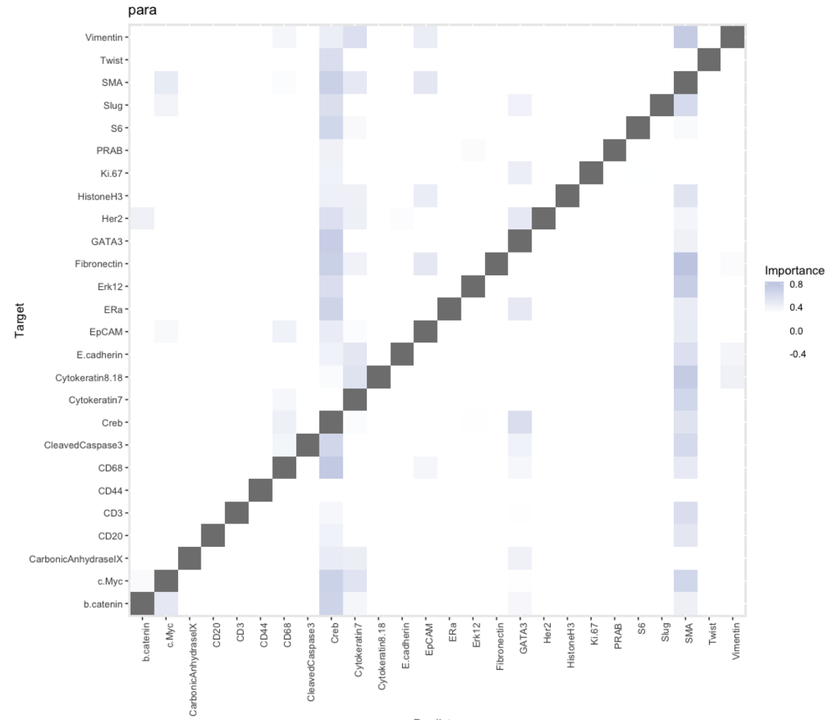
Mann-Whitney U
$p \approx 0$

Results

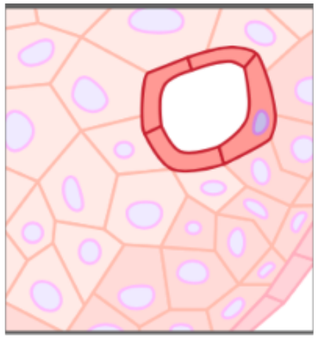
```
misty.results %>% plot_interaction_heatmap("intra", 0.5) %>% plot_interaction_heatmap("para", 0.3)
```



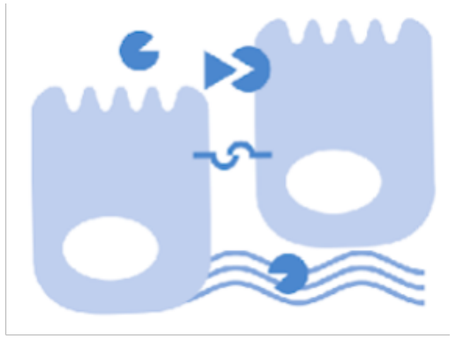




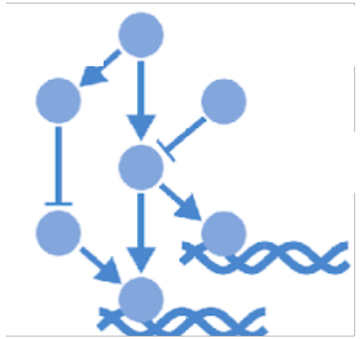
3. What are the specific relations that can explain the contributions?



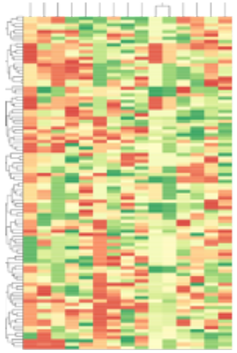
Structure



Communication



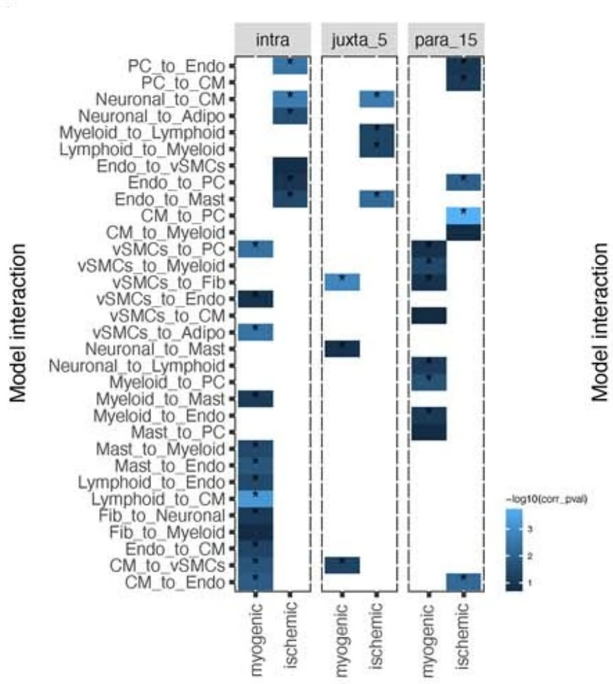
Function

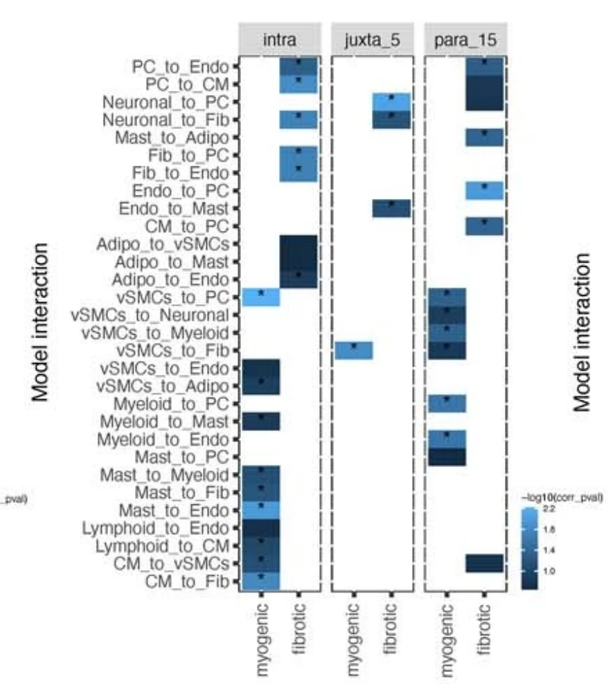


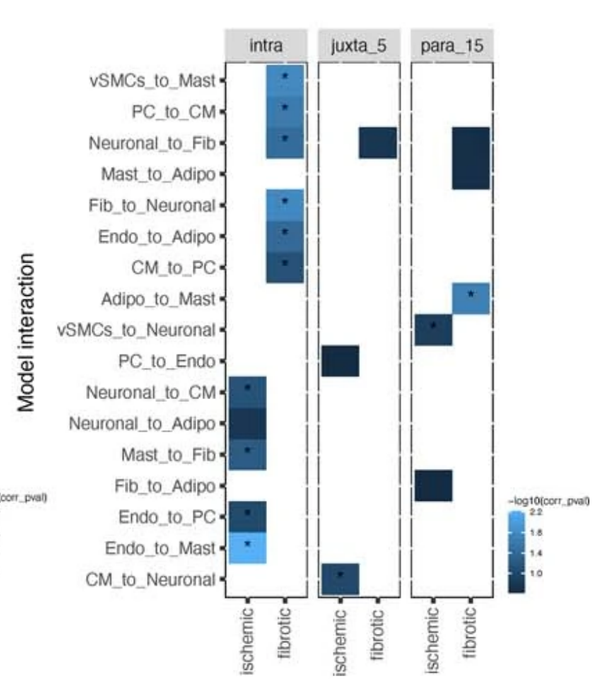
Integration

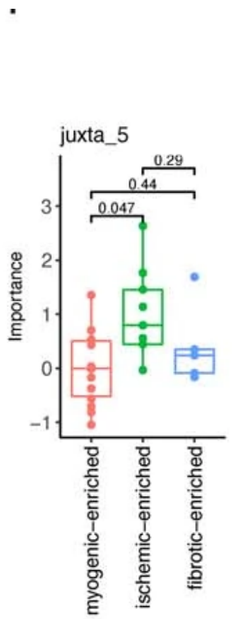
Structure: cell type composition changes during hear failure and recovery

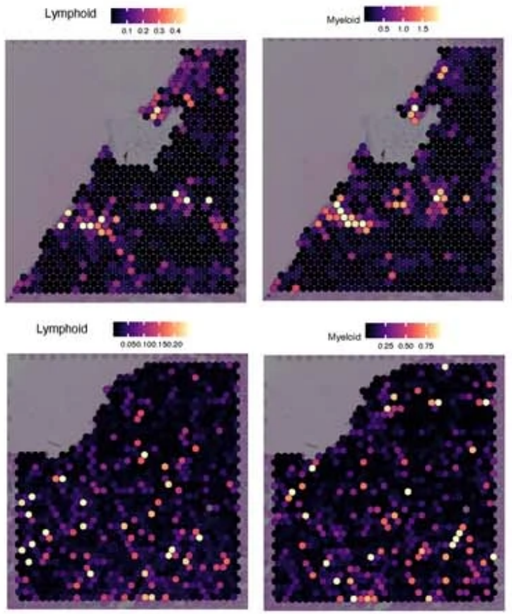
Intraview captures cell type/composition
Relationships capture persistent patterns across samples in different spatial contexts





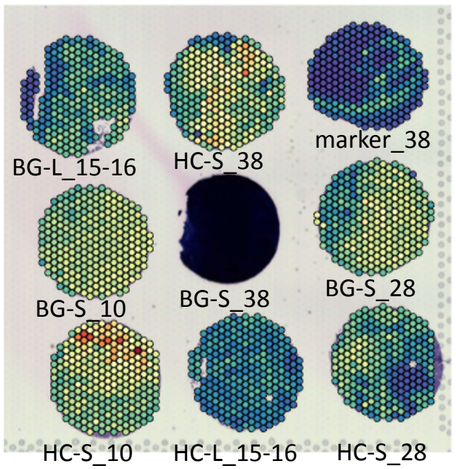


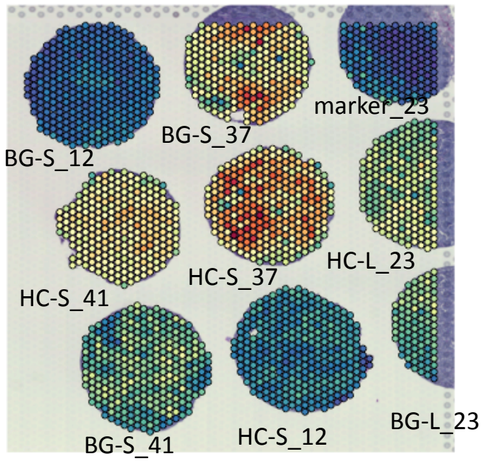


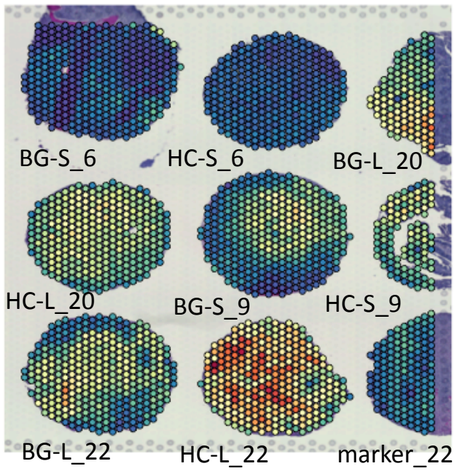


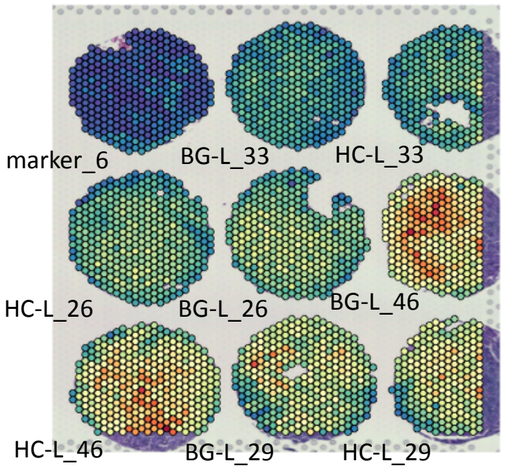
Kuppe, Ramirez Flores, Li et al. 2022 Nature

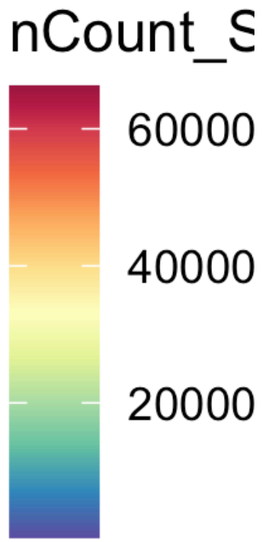
Structure: responders vs nonresponders









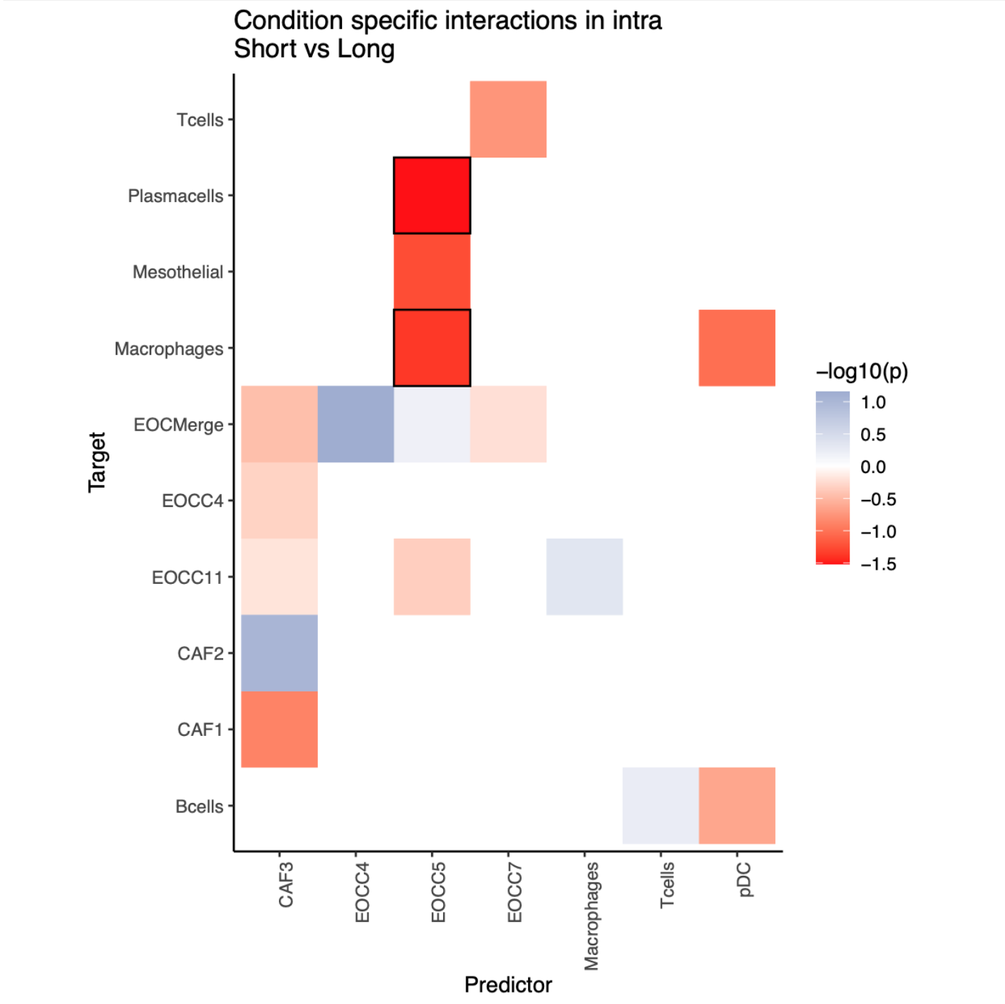


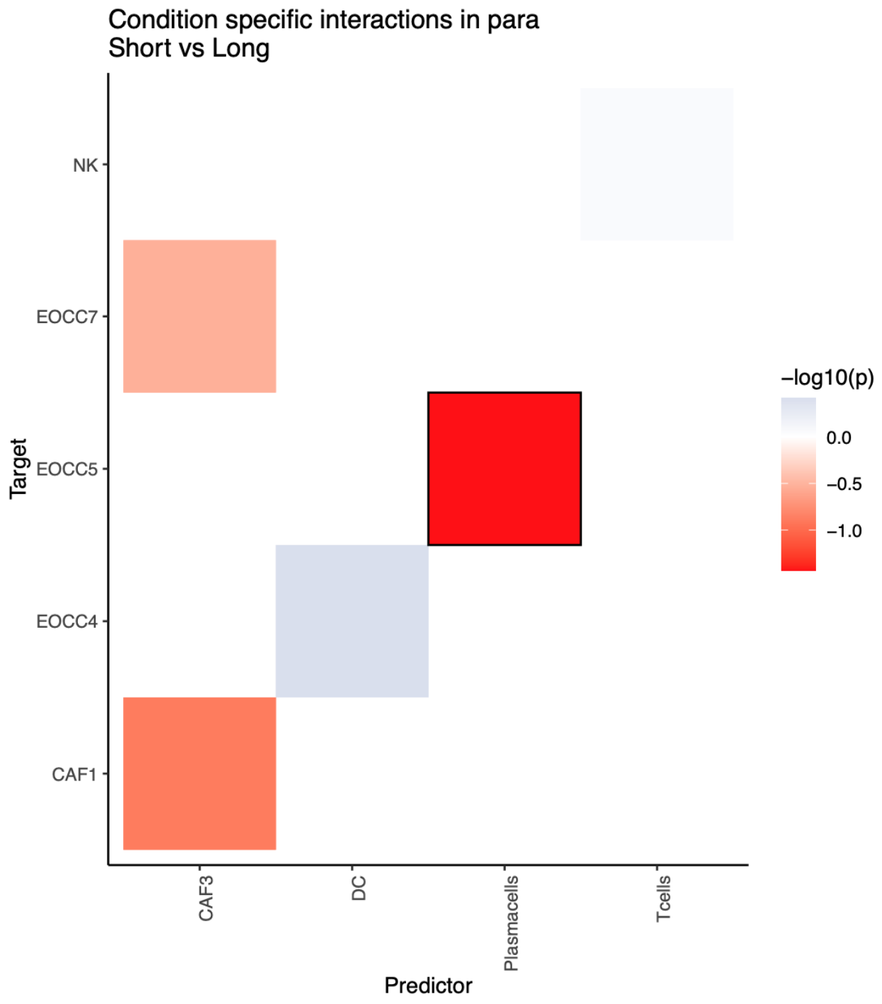
Ovarian cancer Visium TMA:

- Long and short PFI (~survival)
  - Histologically similar samples
- Benign tissue

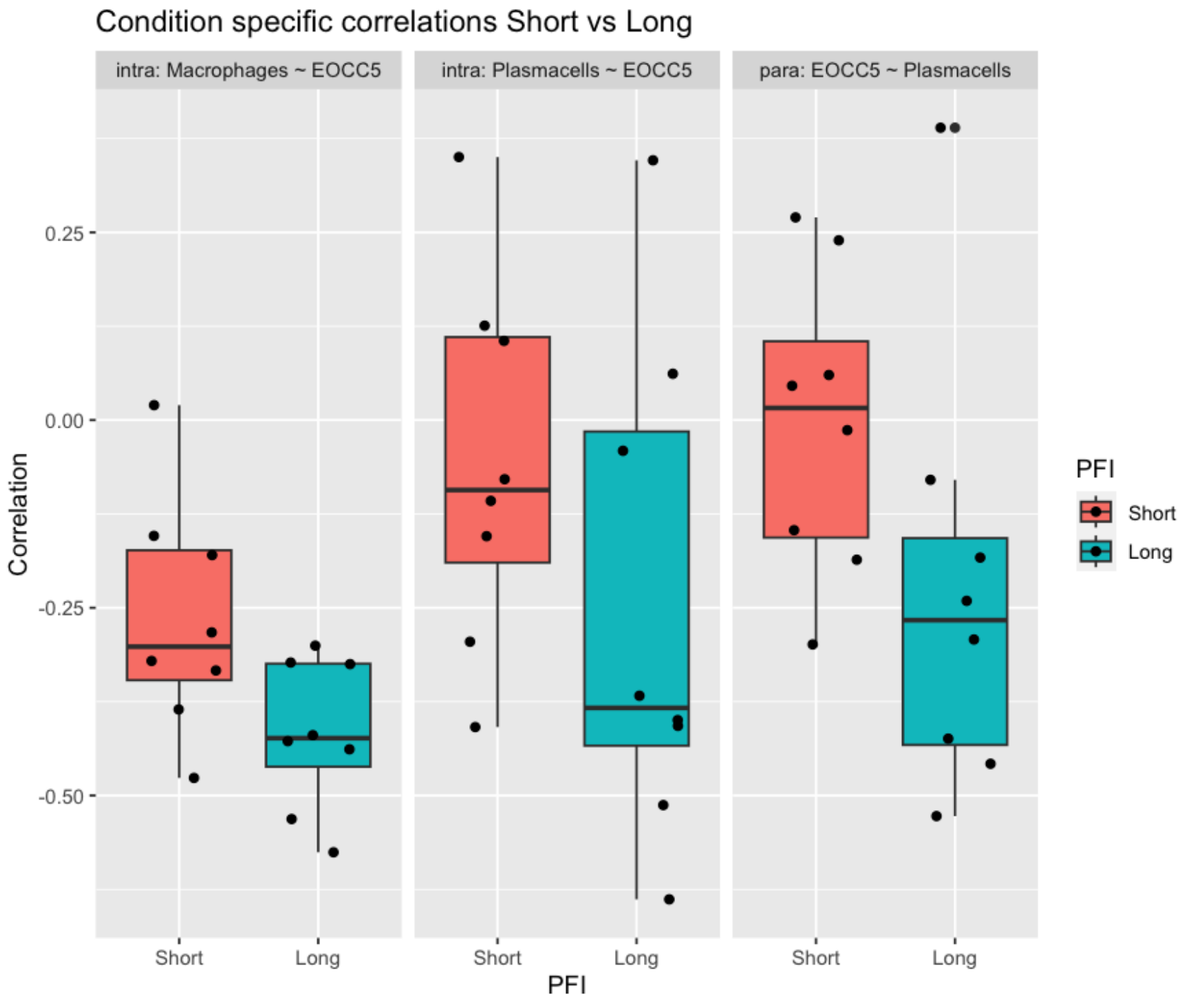
Robin Fallegger et al.

Ovisium results



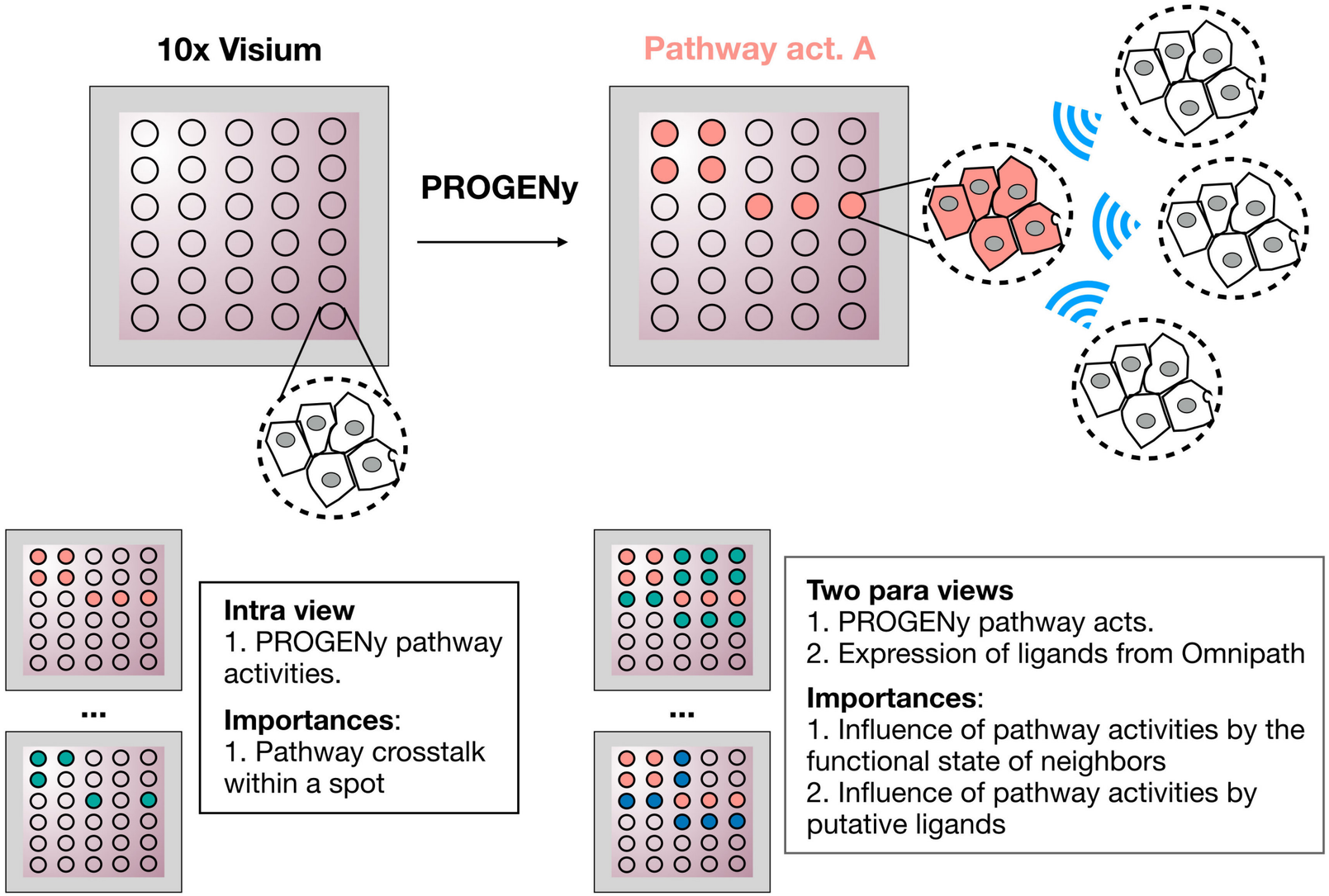


Ovisium analysis



- EOCC5 has a gene signature associated with proliferative DNA repair (Zhang et al, 2022)
- EOCC5 is more negatively correlated with plasmacells and macrophages in responders than in non-responders in the same spots (intra) and the local neighbourhood (para)

Function



Intra view
1. PROGENy pathway activities.

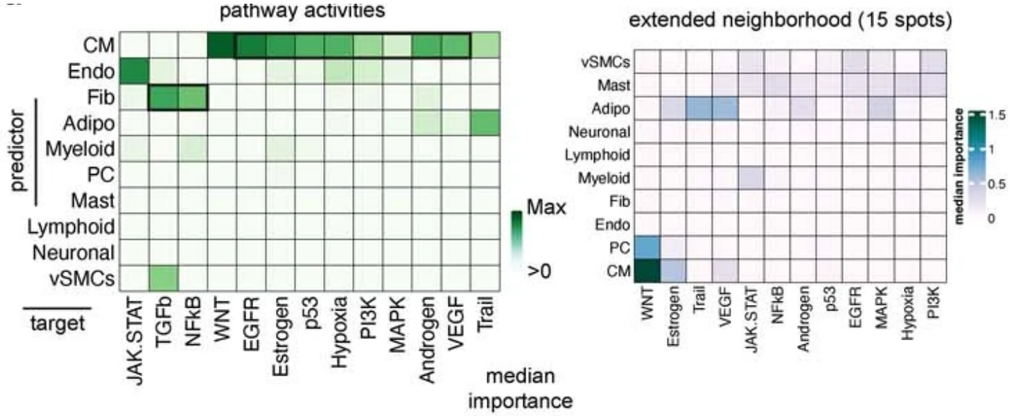
Importances:
1. Pathway crosstalk within a spot

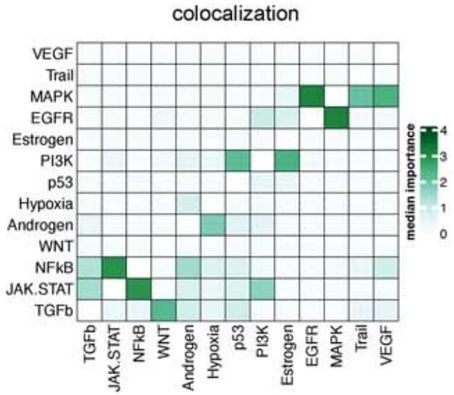
Two para views
1. PROGENy pathway acts.
2. Expression of ligands from Omnipath

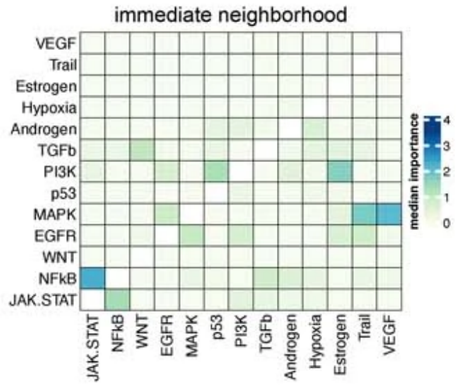
Importances:
1. Influence of pathway activities by the functional state of neighbors
2. Influence of pathway activities by putative ligands

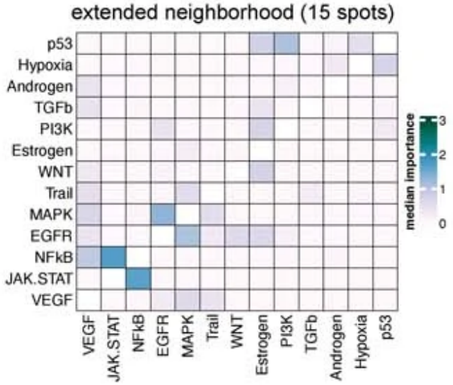
Function

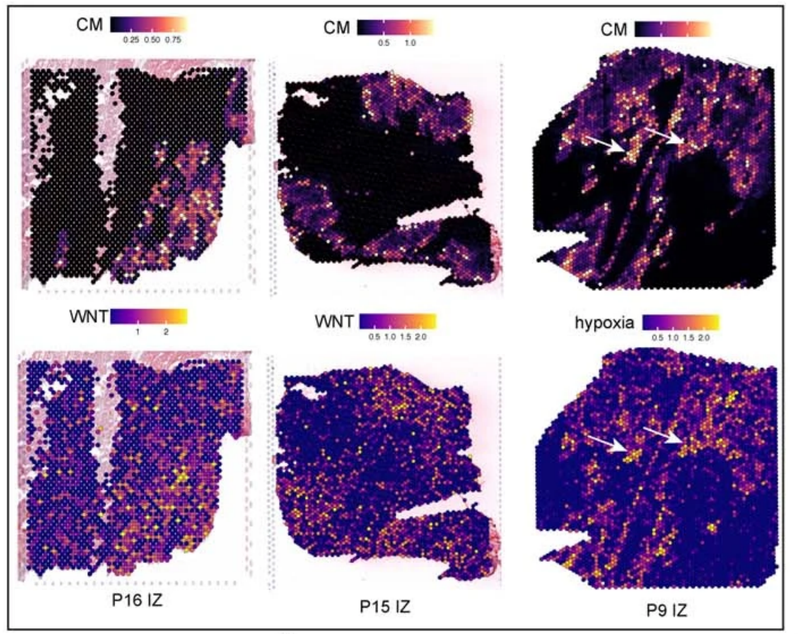
Myocardial infarction - Visium

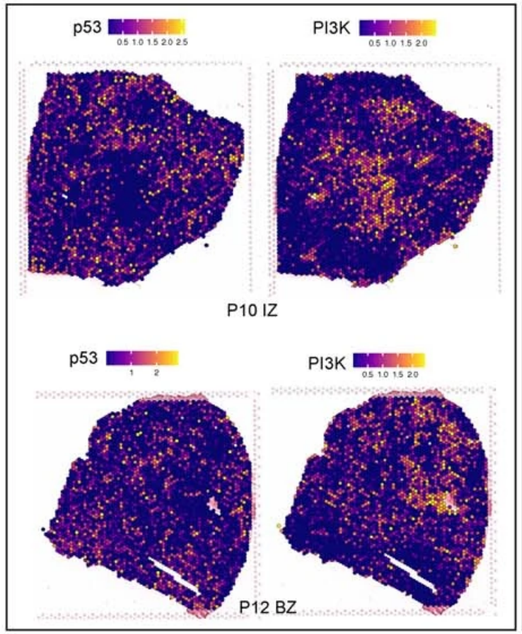


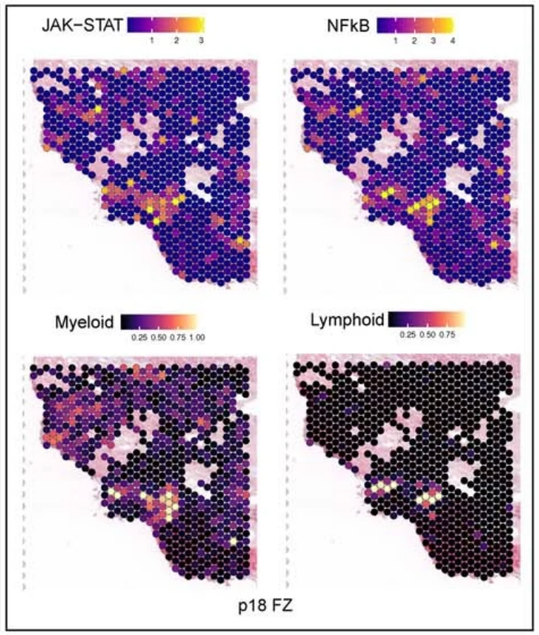








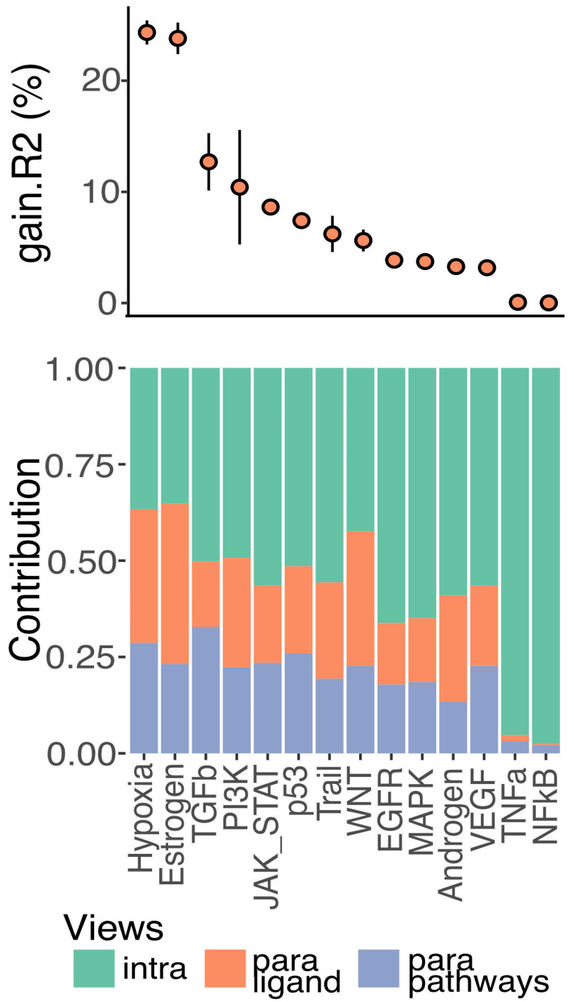


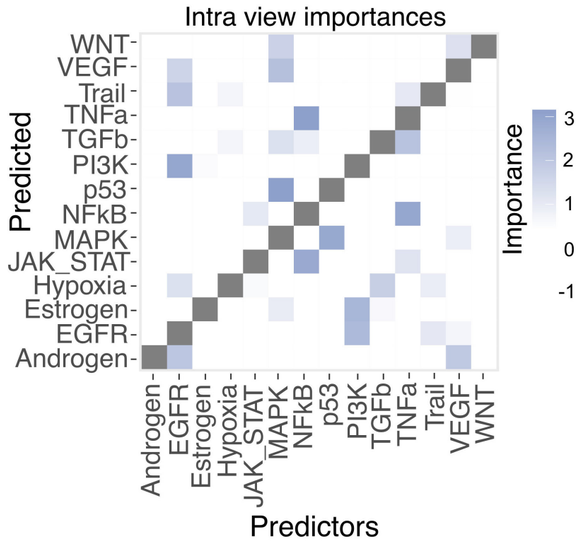


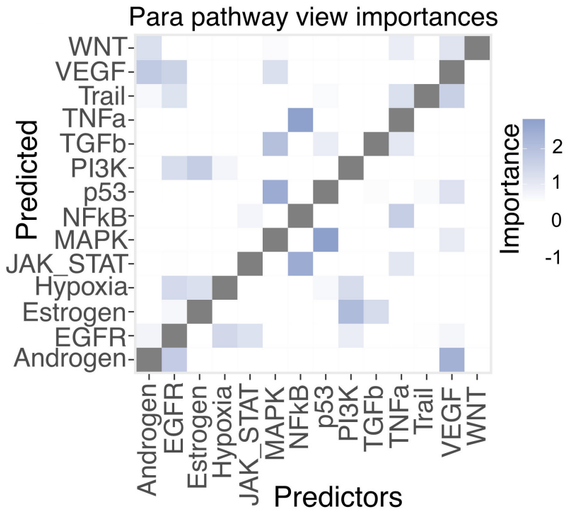
Kuppe, Ramirez Flores, Li et al. 2022 Nature

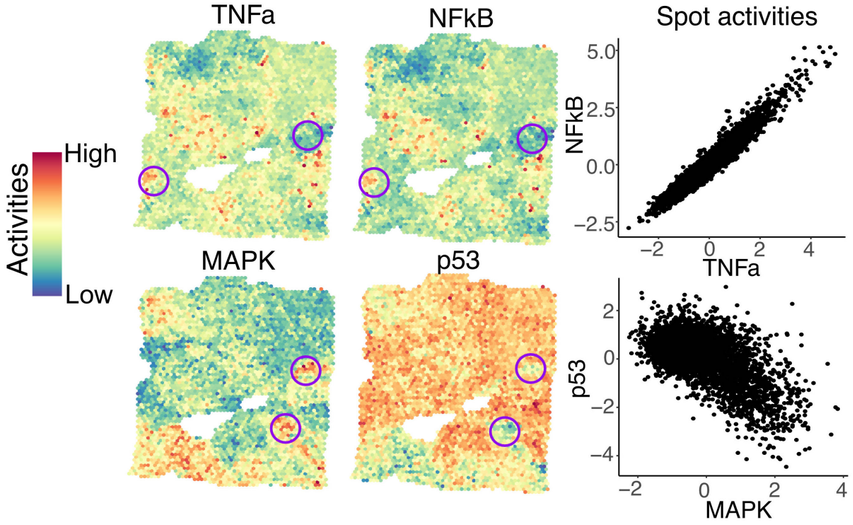
Function: Pathway relationships in different spatial contexts

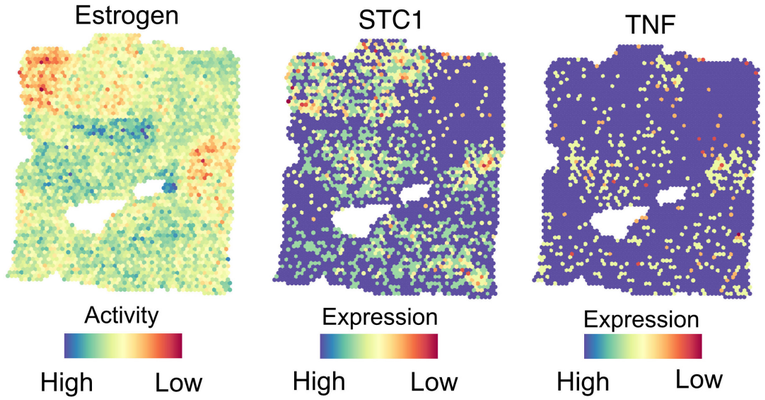
Breast cancer - Visium





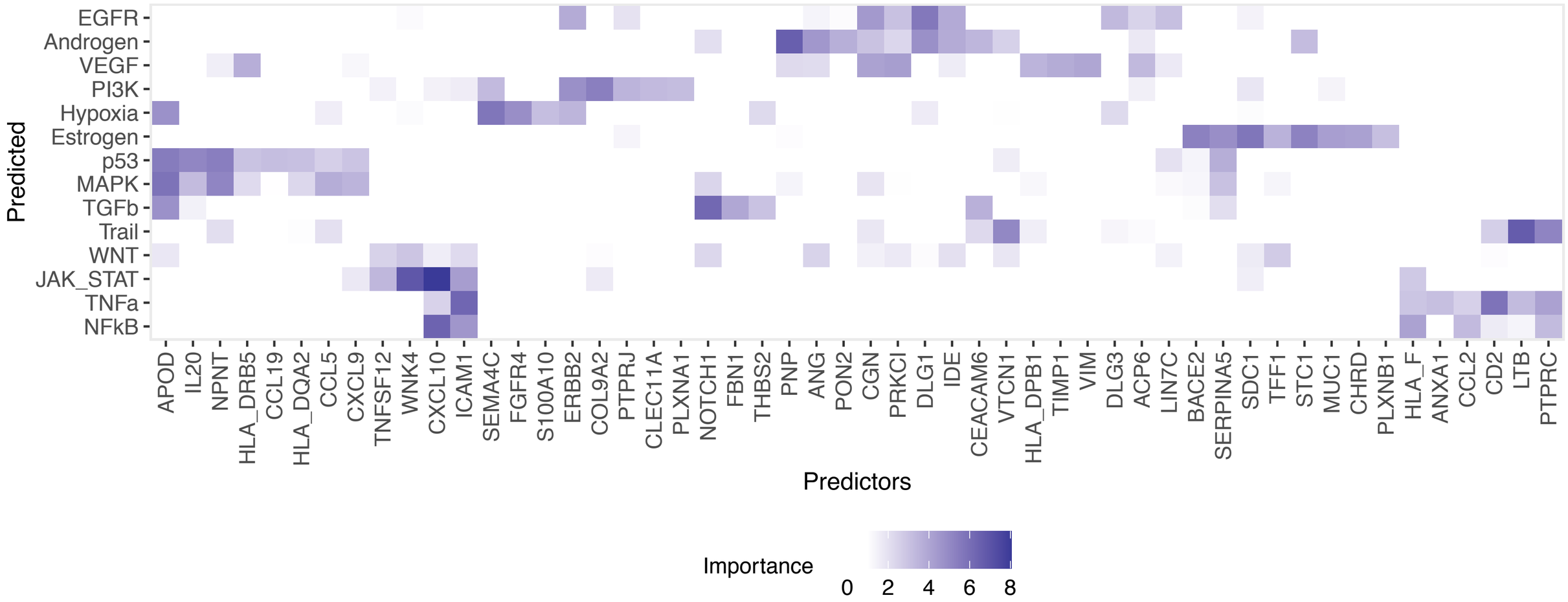






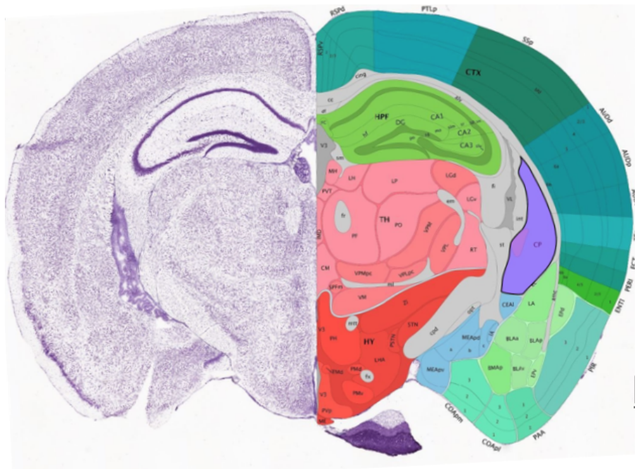
Tanevski et al. 2022 Genome Biology

Communication: Relationship between ligand expression and pathway activities



Tanevski et al. 2022 Genome Biology

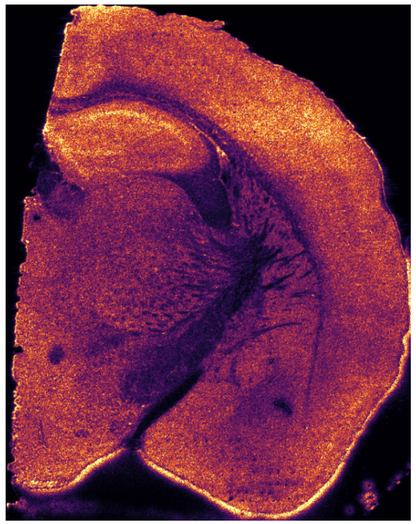
Integration



44 ions

MALDI

[C6H9NO5H]-



Two “consecutive” slides

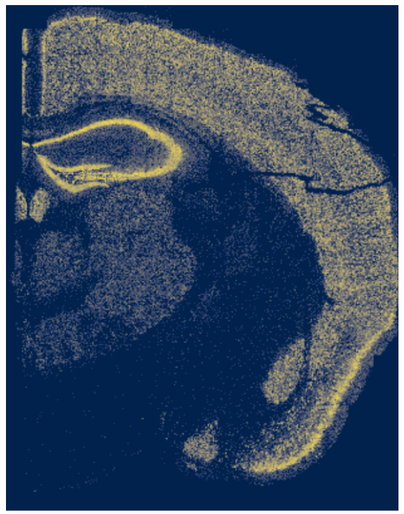
Image registration

Spatial multiomics dataset

resolve biosciences

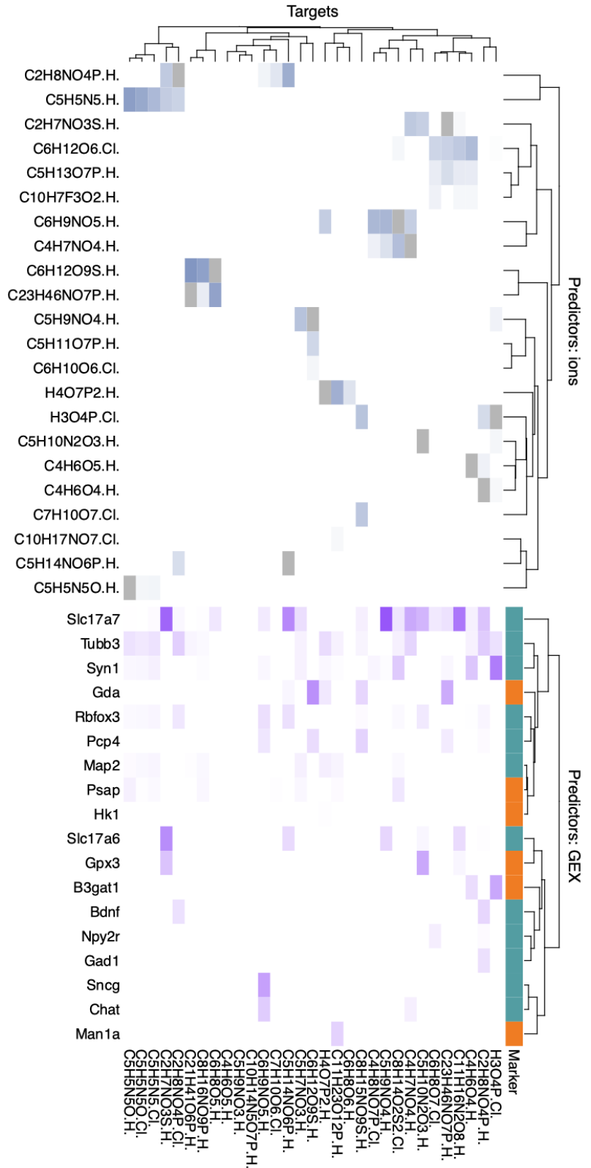
84 genes: celltype markers and metabolic genes

Slc17a7



Bishoy Wadie, Robin Fallegger, Aurelien Dugourd et al.

Integration MISTy to COSMOS

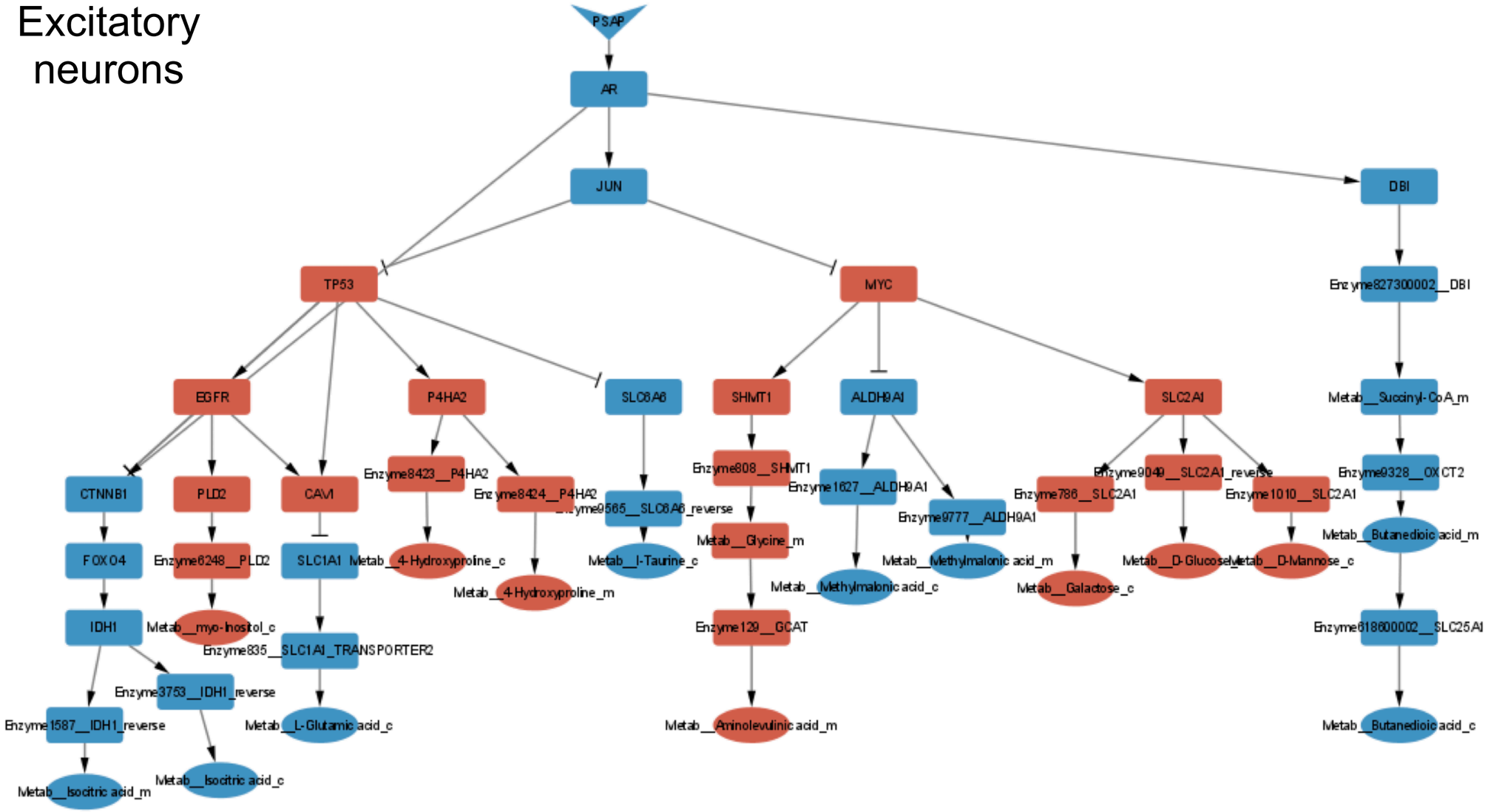


Predictor	Target	Importance	corr
Adcyap1	C2H7NO3S.H.	1.51630113961409	-0.4748476472322239
Aldoc	C5H5N5.Cl.	1.05427732016435	-0.021509317307028073
Aldoc	C5H5N5O.H.	1.05687010901289	-0.011006793731087785
Aldoc	C5H9NO4.H.	1.11910578224508	-0.25702779837839307
Aldoc	C6H8O7.Cl.	1.01617893664255	-0.2065348944352434
Aldoc	C8H16NO9P.H.	1.05932851965315	0.007087207730339929
B3gat1	C23H46NO7P.H.	1.30075511763017	-0.15249549751865749
B3gat1	C4H6O4.H.	2.82040998032327	-0.29353900043522824
B3gat1	C5H10N2O3.H.	1.83852354063271	0.2801797220819485
B3gat1	C6H12O9S.H.	1.65202953456576	-0.19236949330252925
B3gat1	C6H8O7.Cl.	1.17956050311113	-0.23298549398270263
B3gat1	C6H9NO5.H.	1.0257102495121	-0.23899803850770648
B3gat1	C8H14O2S2.Cl.	1.66996696275153	0.23145021364002877
B3gat1	H3O4P.Cl.	4.19092617966786	-0.404564254434519
Bdnf	C10H14N5O7P.H.	1.01884815258267	-0.09147276523215933
Bdnf	C23H46NO7P.H.	1.48479471905082	-0.17340405291050784
...	...	...	...

node	sign
Adcyap1	1
Aldoc	1
B3gat1	1
Bdnf	-1
C5H10N2O3-H-	1
C5H11O7P-H-	1
C5H13O7P-H-	1
C5H14NO6P-H-	1
C5H5N5+Cl-	-1
C5H5N5-H-	-1
C5H5N5O+Cl-	-1
C5H5N5O-H-	-1
C5H9NO3-H-	1
C5H9NO4-H-	-1
C6H10O6+Cl-	1
C6H12O6+Cl-	1
C6H12O9S-H-	1
C6H8O7+Cl-	-1
...	...

Integration: cell-type specific multi-omic network

Excitatory neurons



Learn more

Rao, A., Barkley, D., França, G.S. *et al.* Exploring tissue architecture using spatial transcriptomics. *Nature* **596,** 211–220 (2021). https://doi.org/10.1038/s41586-021-03634-9

Zeng, Z., Li, Y., Li, Y. et al. Statistical and machine learning methods for spatially resolved transcriptomics data analysis. *Genome Biol* **23**, 83 (2022). https://doi.org/10.1186/s13059-022-02653-7

Palla, G., Fischer, D.S., Regev, A. et al. Spatial components of molecular tissue biology. *Nat Biotechnol* **40**, 308–318 (2022). https://doi.org/10.1038/s41587-021-01182-1

Moses, L., Pachter, L. Museum of spatial transcriptomics. *Nat Methods* (2022). https://doi.org/10.1038/s41592-022-01409-2

Tanevski, J., Flores, R.O.R., Gabor, A. et al. Explainable multiview framework for dissecting spatial relationships from highly multiplexed data. *Genome Biol* **23**, 97 (2022). https://doi.org/10.1186/s13059-022-02663-5

Rahimi, A., Vale-Silva, L.A., Faelth Savitski, M., Tanevski, J., Saez-Rodriguez, J. DOT: Fast Cell Type Decomposition of Spatial Omics by Optimal Transport. arXiv (2023). https://doi.org/10.48550/arXiv.2301.01682

Available positions

Spatial data seems interesting?
Interested in getting some experience analyzing it?

Inquire about internship/rotation at jobs.saez@uni-heidelberg.de

Check https://saezlab.org for instructions and more details