

Overview of computational analysis of spatially resolved omics data

Jovan Tanevski

Saez-Rodriguez Group



www.saezlab.org  @saezlab

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



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



Evaluation requirements

- Three questions in the two practicals marked with * and a number
- Choose one to answer
- The answer should be given in the form of a self-contained R notebook.
Feel free to use the provided notebooks for the practicals as a place to start.
- Upload the notebook containing your answer here -> <https://www.dropbox.com/request/CSkkwGYaazv5YIwNHb27> - you can find the link in the practicals too
- Name your notebook as follows
FirstLastName_programme_spatial_#.Rmd
where # is the number of the question you chose to answer



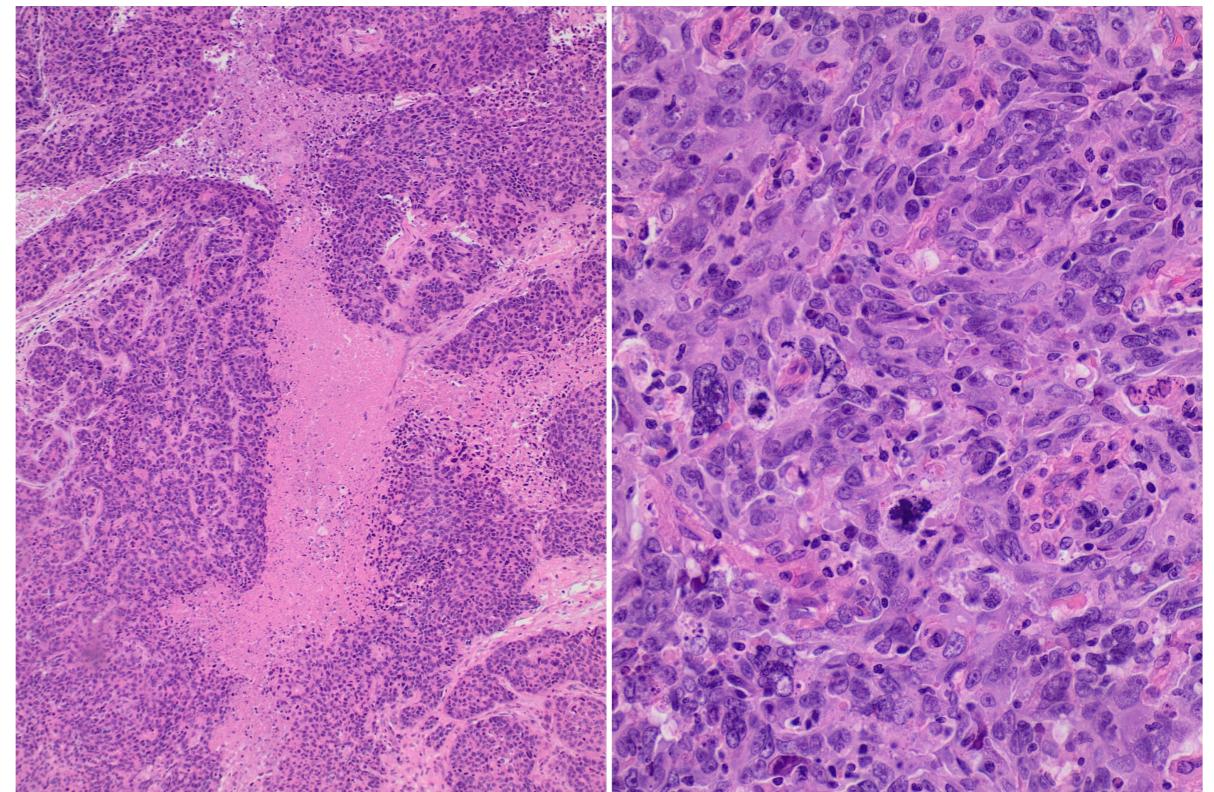
Frame of the lecture

- 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

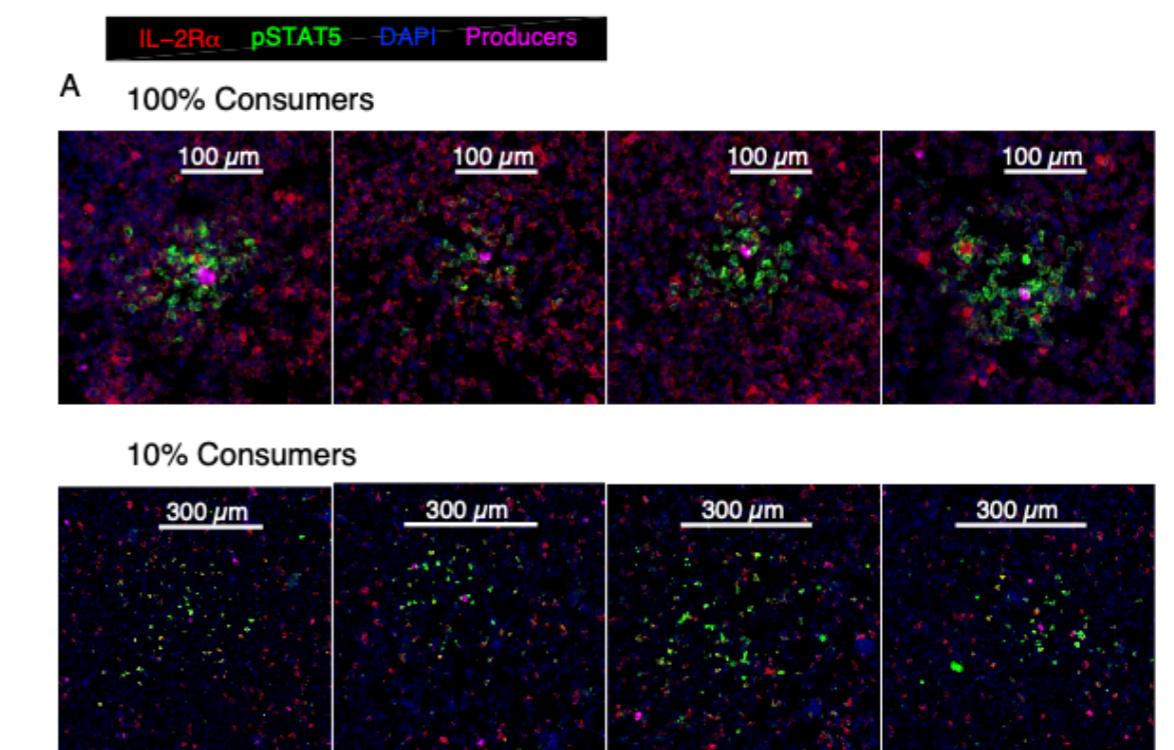


Why spatial data

- 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



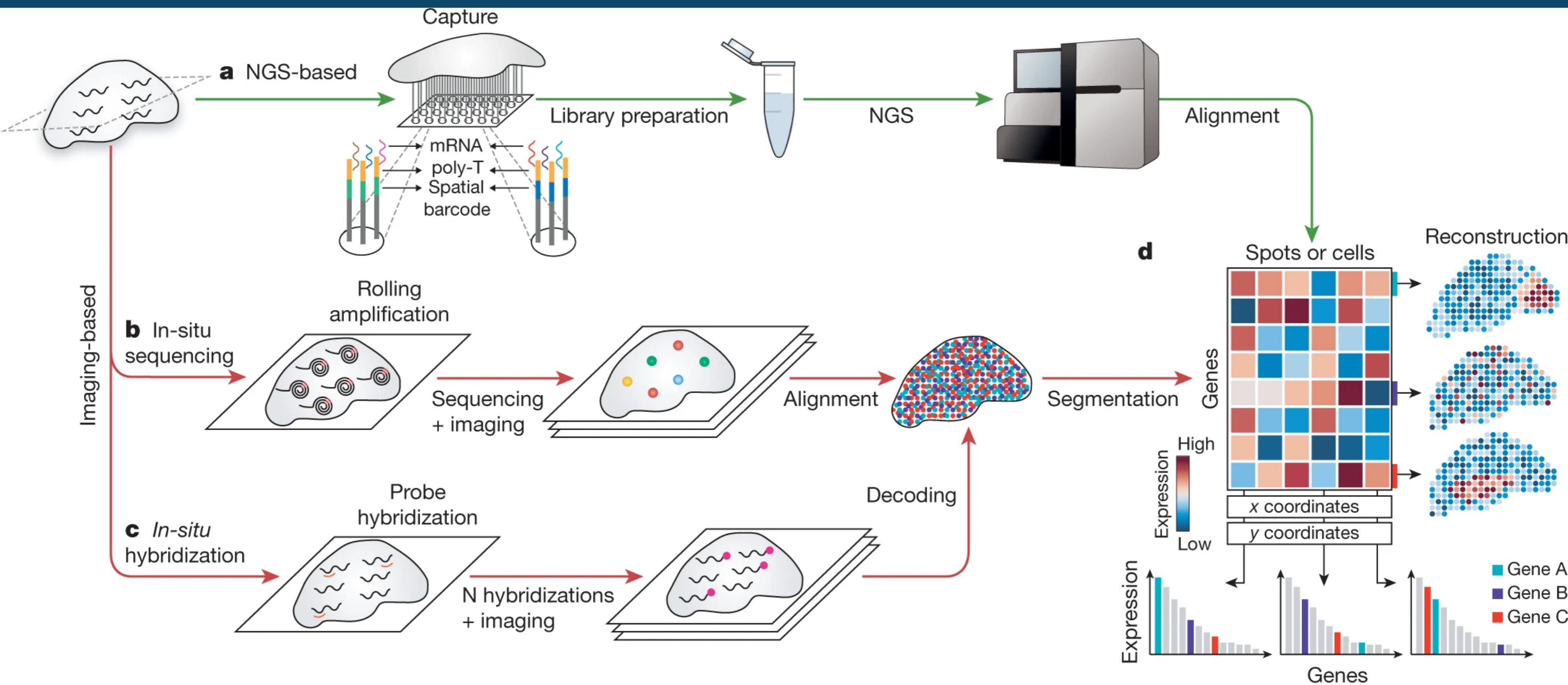
Wikimedia commons



Oyler-Yaniv et al. 2017, Immunity



How to get spatial data

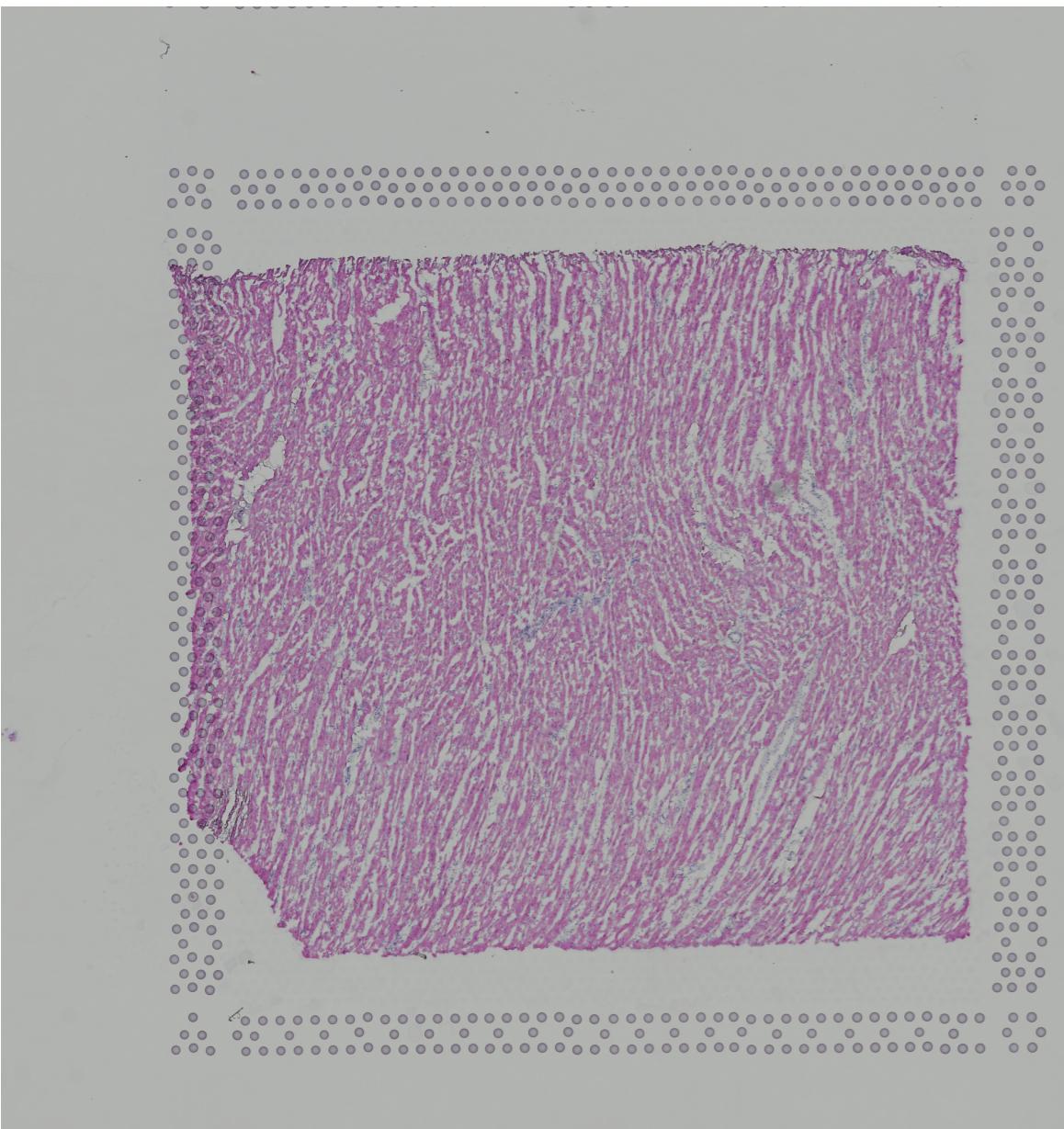




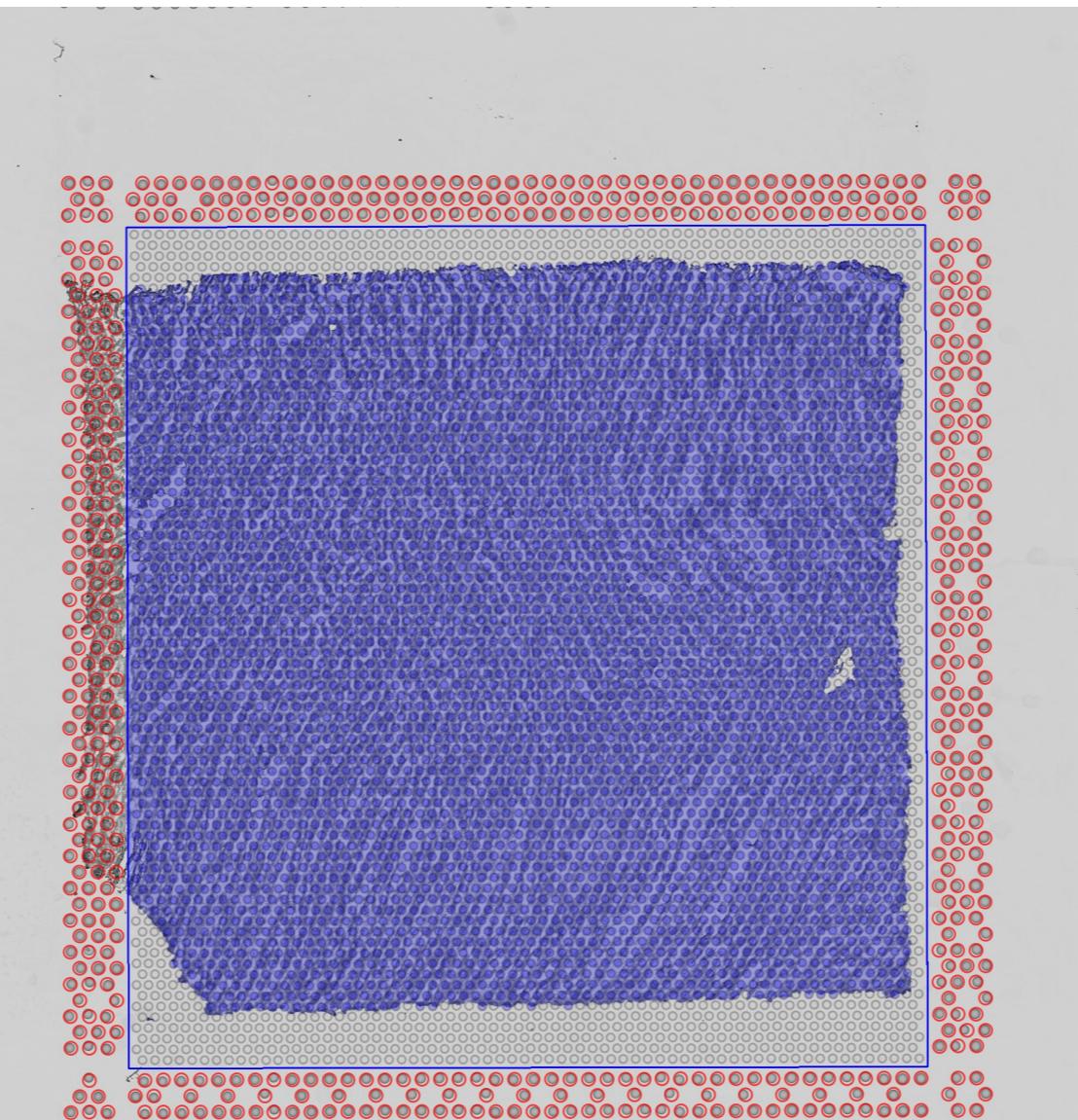
What does the data look like

Fresh frozen human heart tissue - 10x Genomics Visium

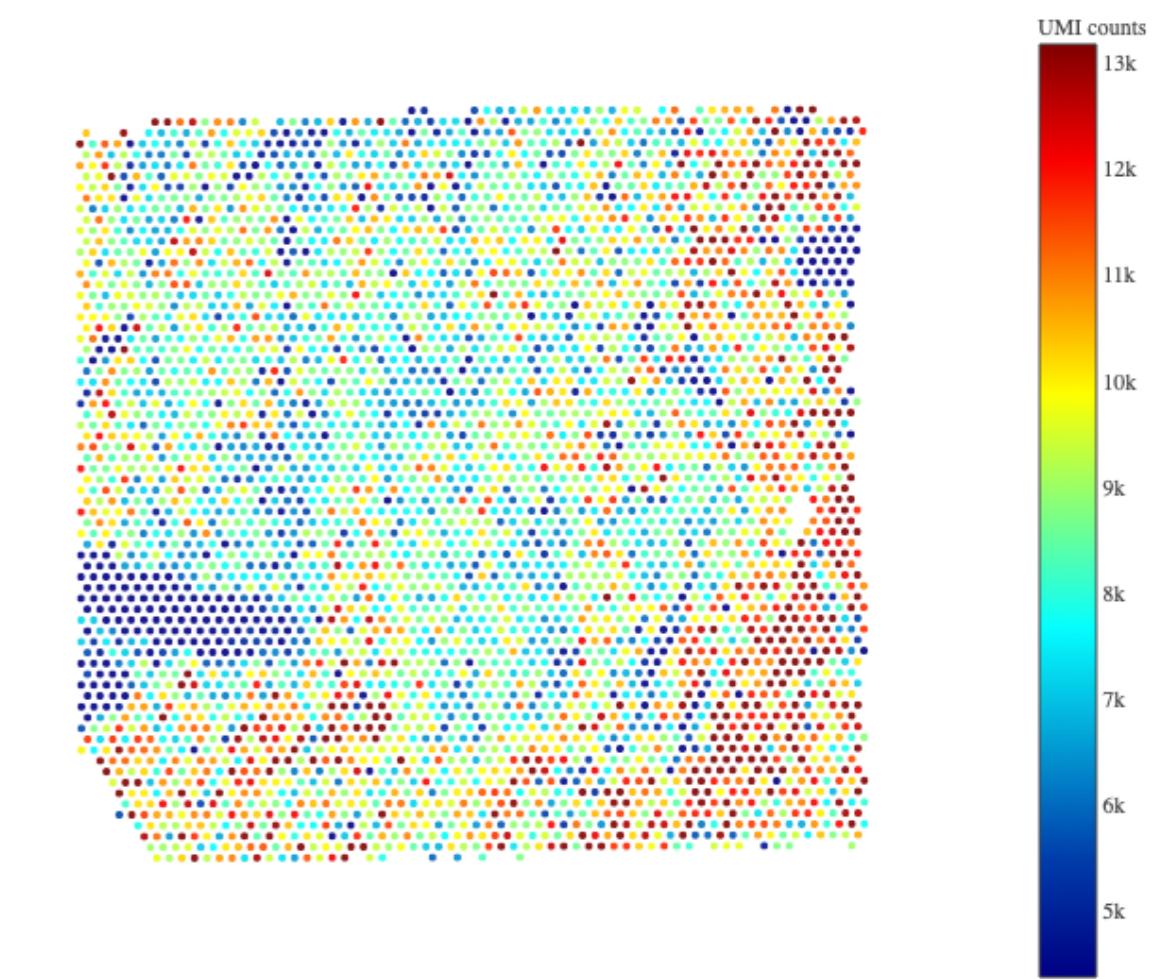
H&E



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

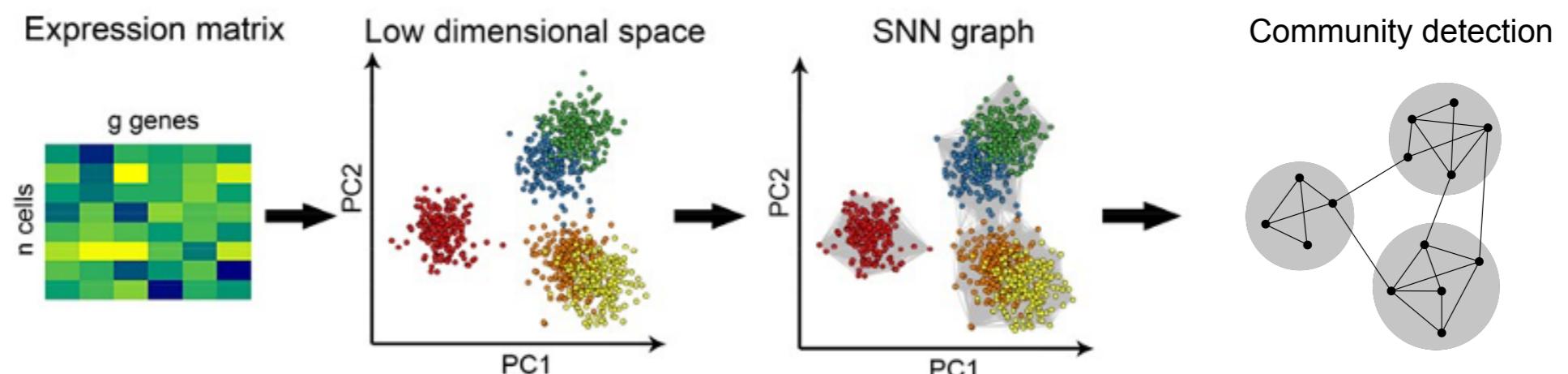
Part 1 - clustering , patterns and integration





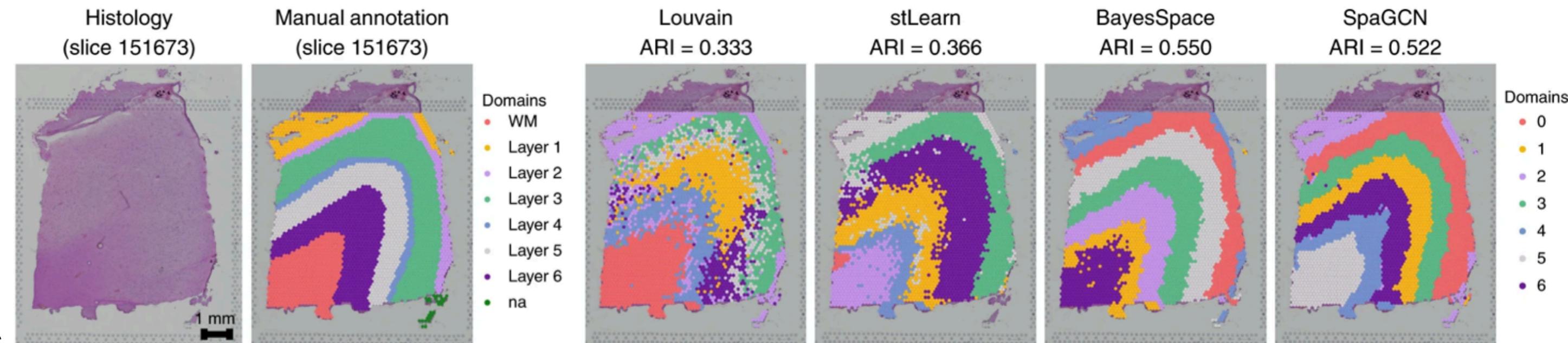
Clustering

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



Jarvis & Patrick 1973 *IEEE Trans. Comput.*

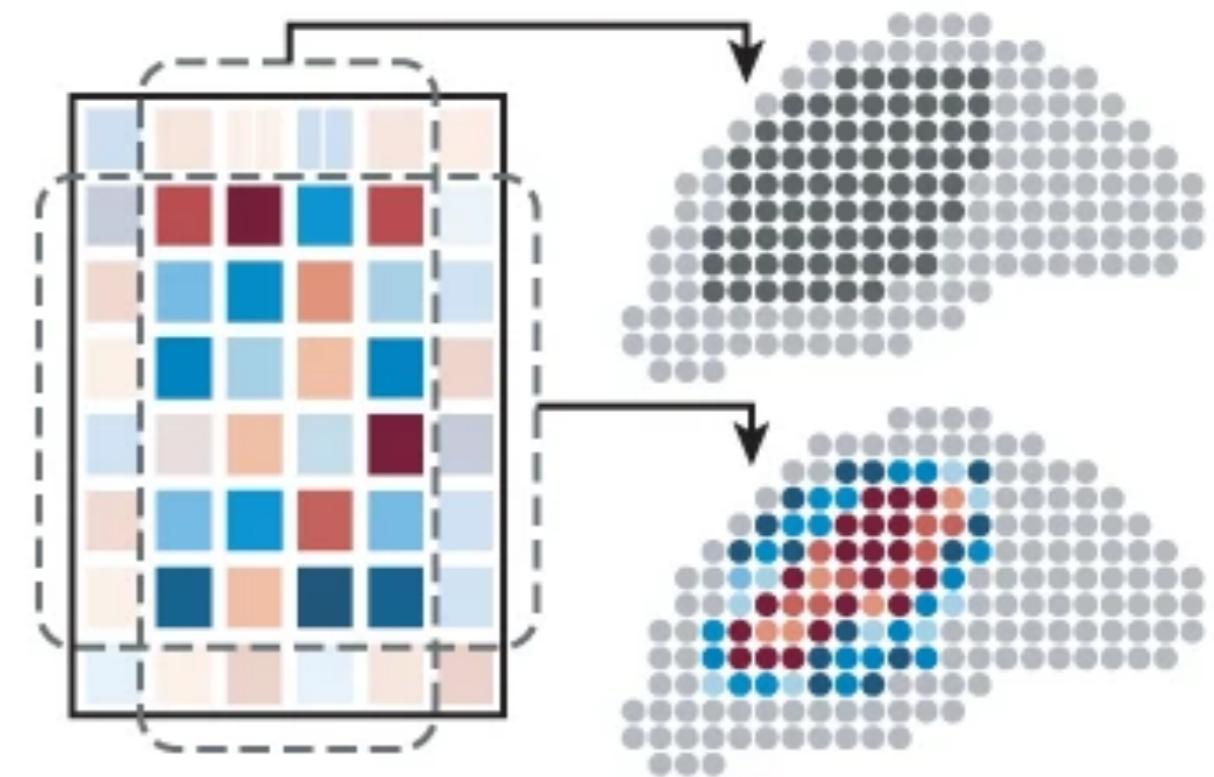
- There are also other methods that also take into account the spatial component of the data stLearn, SpaGCN, BayesSpace, although still not widely used





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





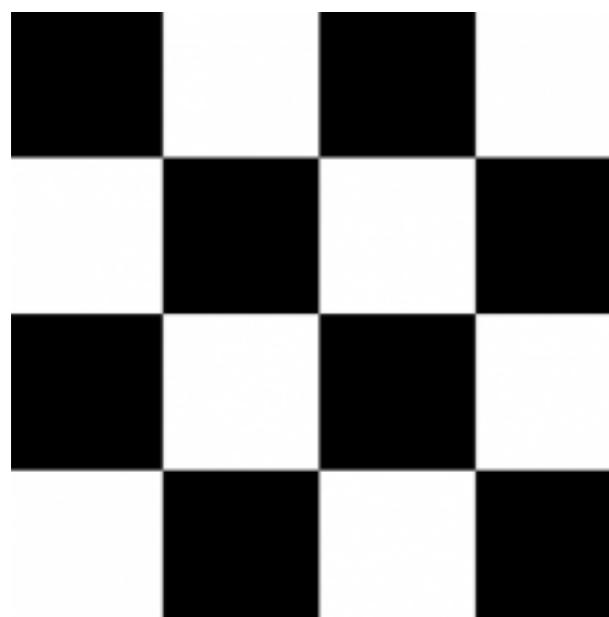
Spatial autocorrelation

Moran's I

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

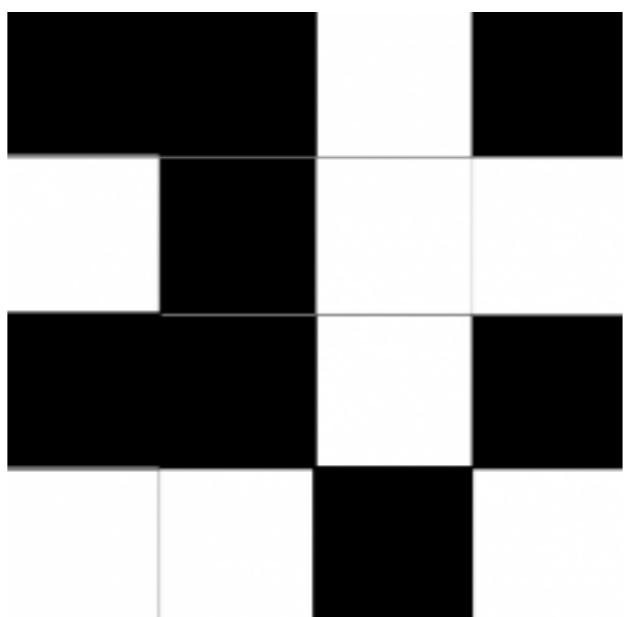
I = -1

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



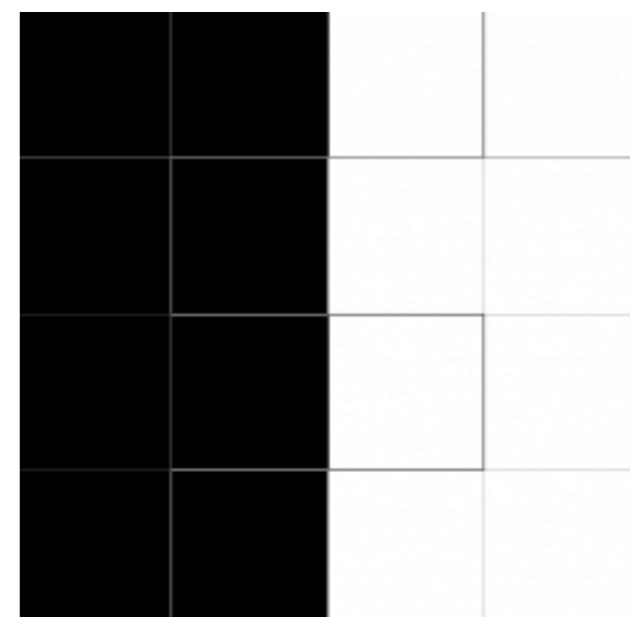
I = 0

Random



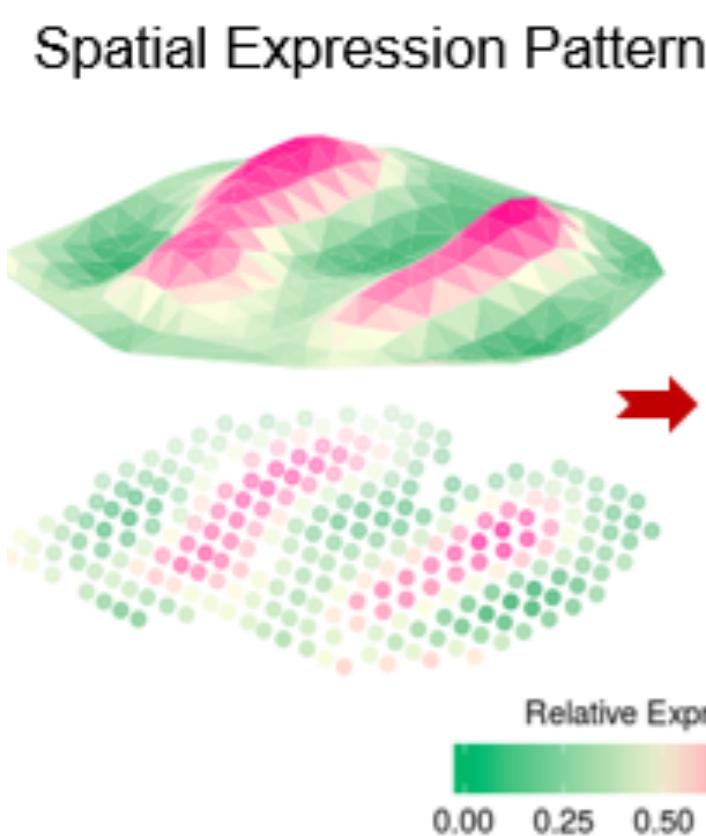
I = 1

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





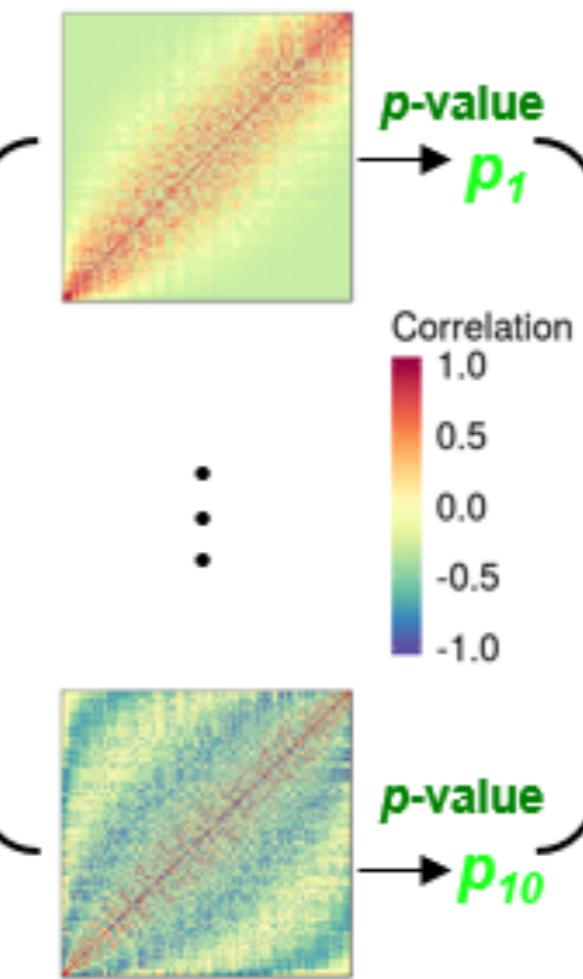
A more advanced approach. Why? Why not?



Generalized Linear Spatial Model

$$\begin{aligned}y_i &\sim Poi(N_i \lambda_i) \\ \log \lambda_i &= \mathbf{x}_i^T \boldsymbol{\beta} + b_i + \epsilon_i \\ \mathbf{b} &= (b_1, \dots, b_n)^T \sim MVN(0, \tau_1 \mathbf{K}) \\ \boldsymbol{\epsilon} &= (\epsilon_1, \dots, \epsilon_n)^T \sim MVN(0, \tau_2 \mathbf{I})\end{aligned}$$

Gaussian/Periodic Kernels (\mathbf{K})



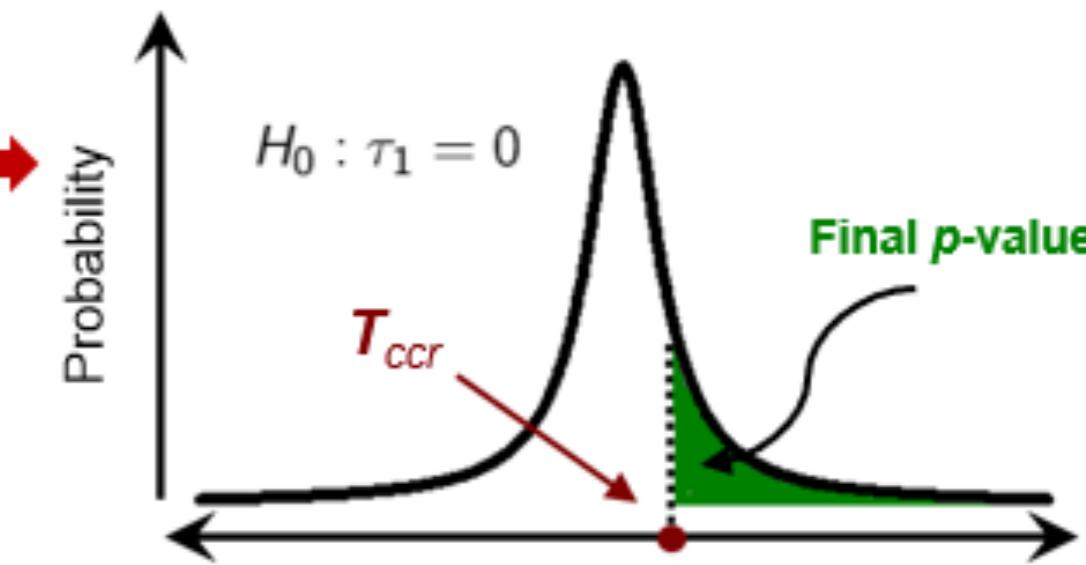
Cauchy Combination Rule

$$T_{ccr} = \sum_{i=1}^{10} w_i \tan \{(0.5 - p_i)\}$$

$$H_0 : \tau_1 = 0$$

$$T_{ccr}$$

Final $p\text{-value}$



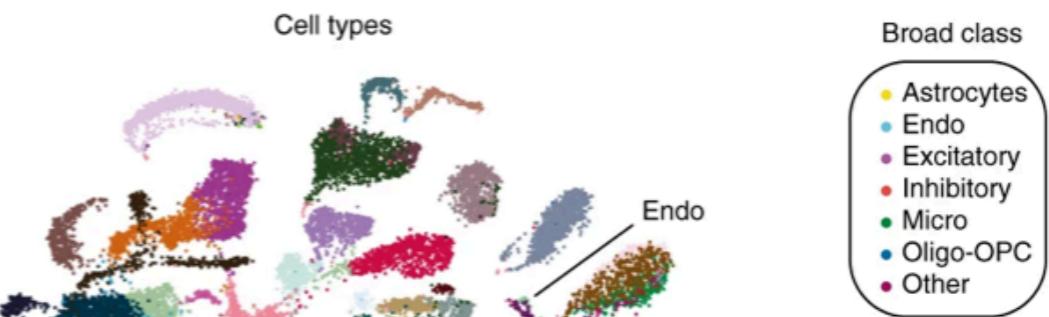


Integration with scRNAseq data

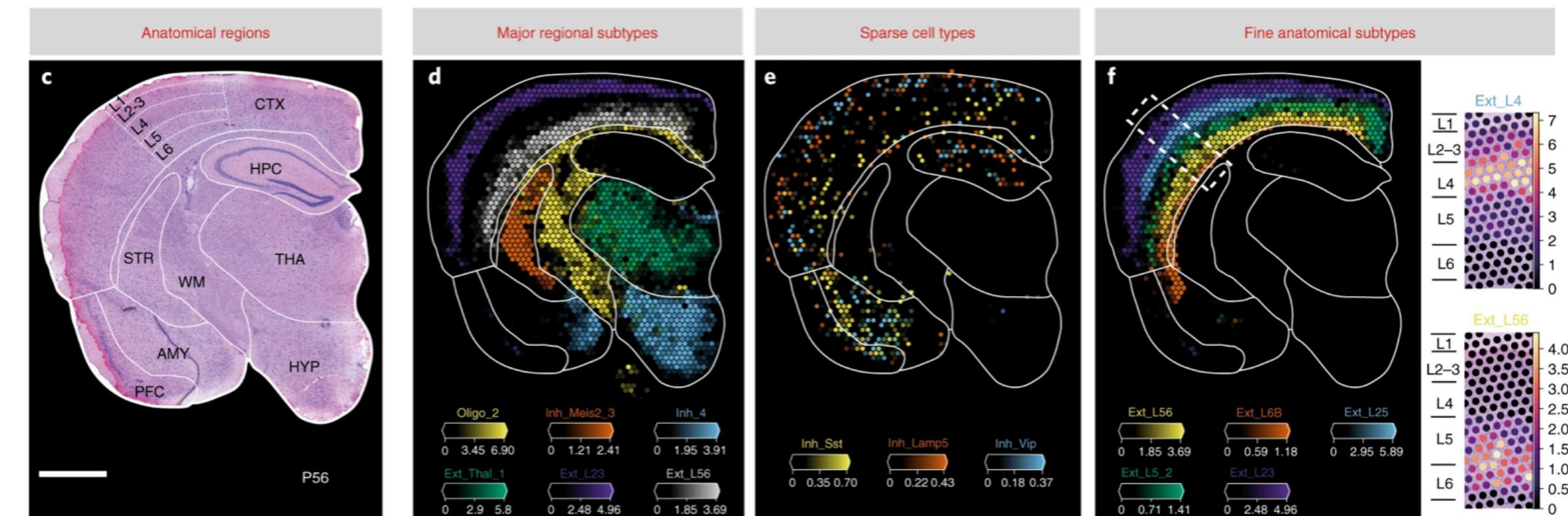
- Why integrate?

- Different tasks:
 - Label transfer
 - Deconvolution
 - Localization

sc(sn)RNAseq “atlas”



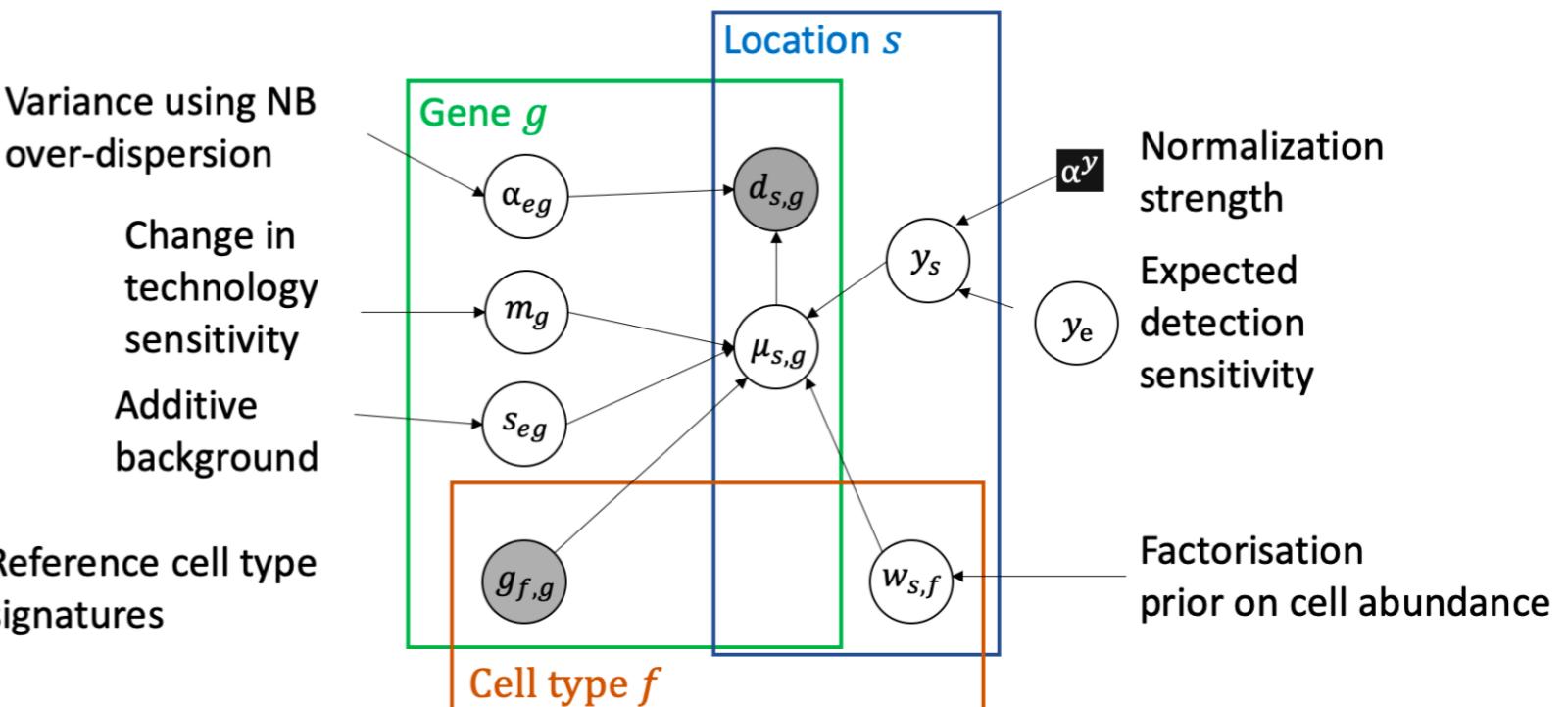
visium slide with deconvoluted cell-type abundance



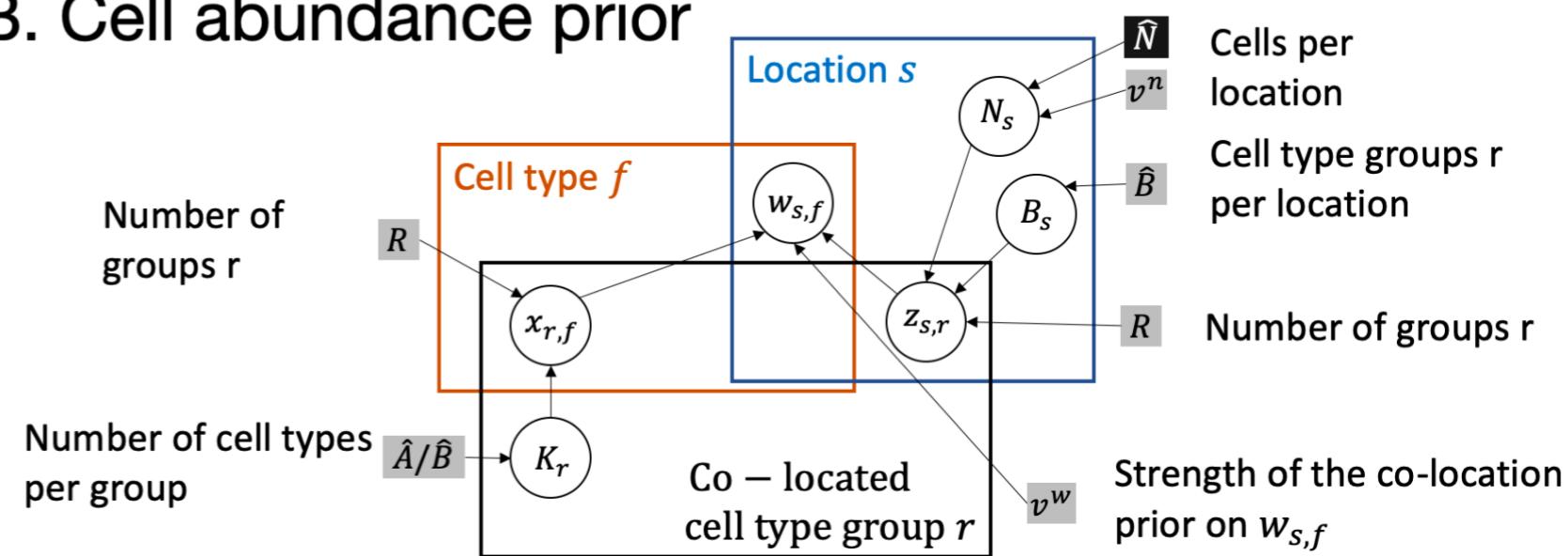
Mapping fine-grained cell types in spatial transcriptomics

$$\mu_{s,g} = \left(\underbrace{m_g}_{\text{technology sensitivity}} \cdot \underbrace{\sum_f w_{s,f} g_{f,g}}_{\text{cell type contributions}} + \underbrace{s_{e,g}}_{\text{additive shift}} \right) \cdot \underbrace{y_s}_{\text{per-location sensitivity}}$$

A. mRNA counts model



B. Cell abundance prior



Observed

 Unobserved variable

<i>Name</i>	Hyper-parameter
<i>Name</i>	Hyper-parameter

Part 2 - analysis of structure and function

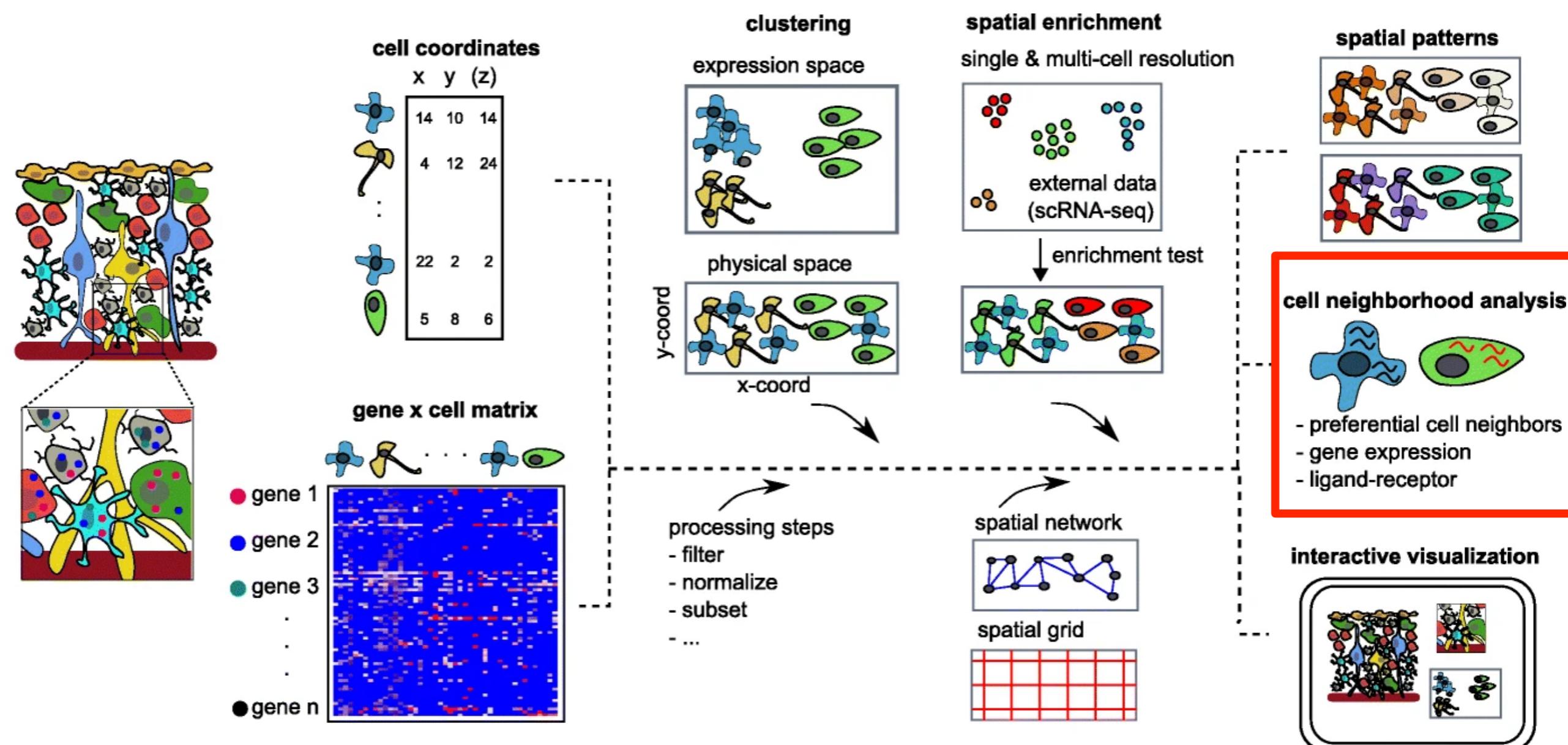




Neighborhood analysis

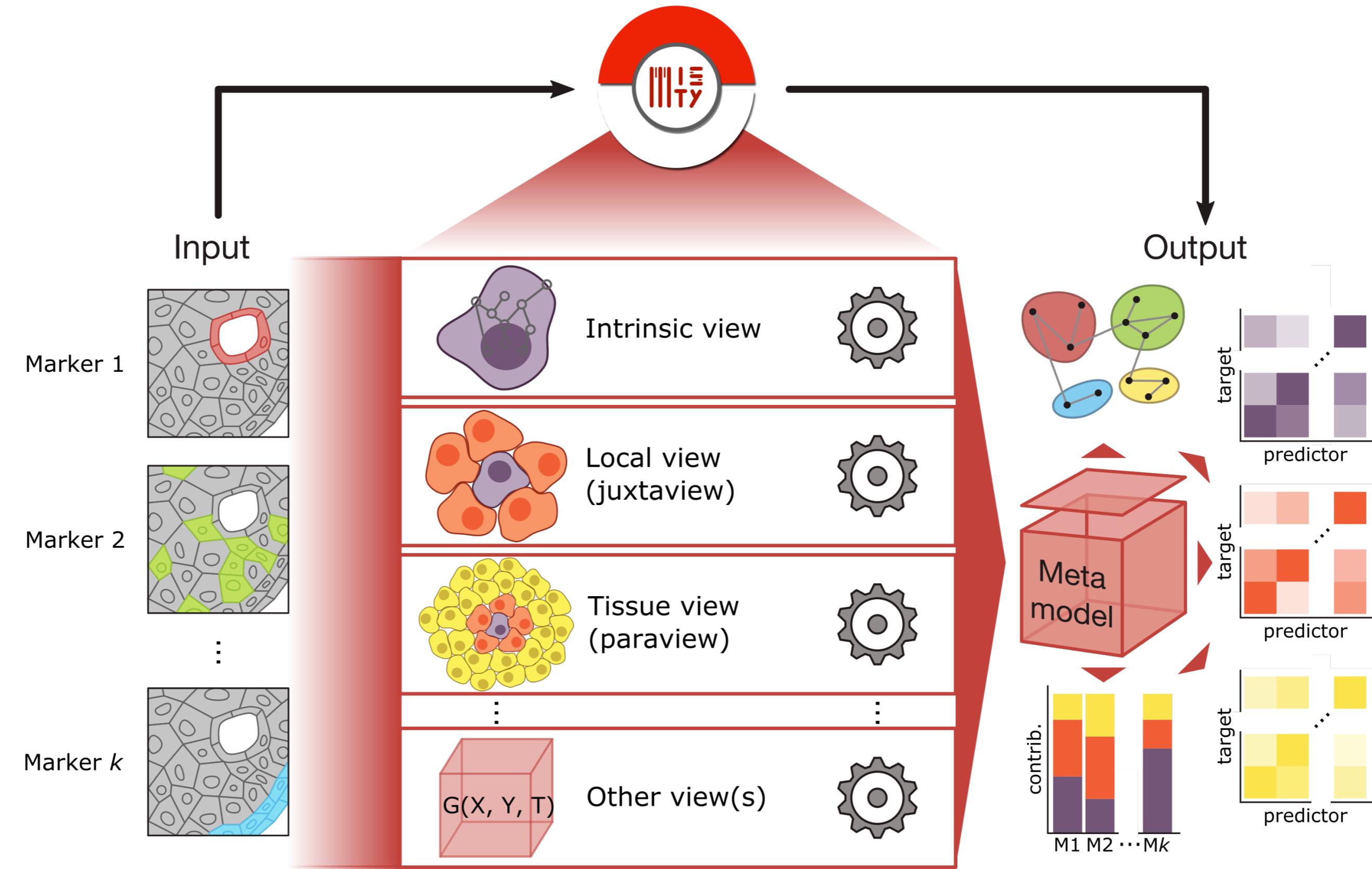
Toolboxes for analysis of spatial data

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





MISTy





Central resource



<https://saezlab.github.io/mistyR/>

mistyR **1.2.1** Get started Reference Articles ▾ Changelog

MISTy - Multiview Intercellular SpaTial modeling framework

Overview

The advancement of technologies for measurement of highly multiplexed spatial data require the development of scalable methods that can leverage the availability of the spatial context. Multiview Intercellular SpaTial modeling framework (MISTy) is an explainable machine learning framework for knowledge extraction and analysis of single-cell, highly multiplexed, spatially resolved data.



Links

[View on Bioconductor](#)

[Browse source code](#)

[Report a bug](#)

License

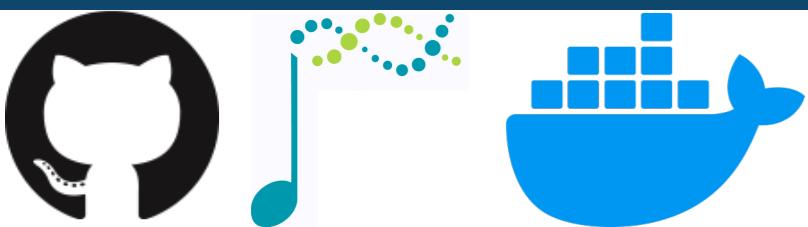
[GPL-3](#)

Community

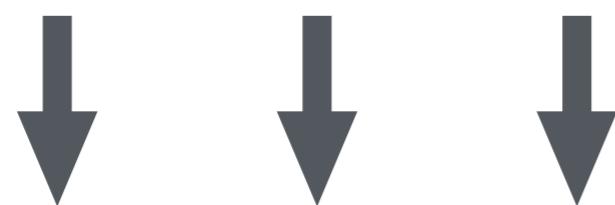
[Code of conduct](#)



Central resource



<https://saezlab.github.io/mistyR/>



mistyR 1.2.1 [Get started](#) Reference [Articles ▾](#) Changelog

MISTy - Multiview Intercellular SpaTial modeling framework

Overview

The advancement of technologies for measurement of highly multiplexed spatial data require the development of scalable methods that can leverage the availability of the spatial context. Multiview Intercellular SpaTial modeling framework (MISTy) is an explainable machine learning framework for knowledge extraction and analysis of single-cell, highly multiplexed, spatially resolved data.



Links

[View on Bioconductor](#)

[Browse source code](#)

[Report a bug](#)

License

[GPL-3](#)

Community

[Code of conduct](#)

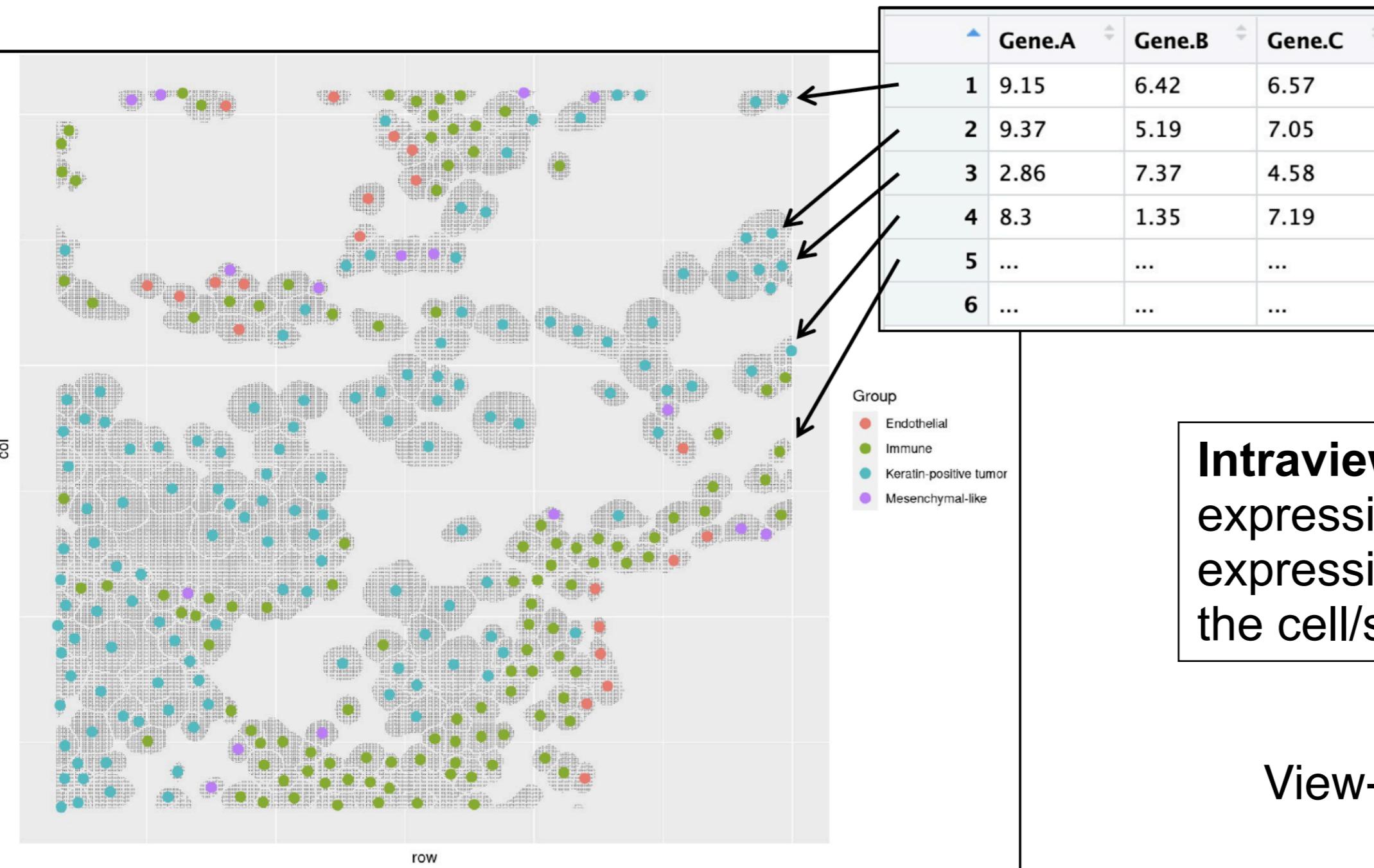


Caveats

The results from MISTy are not causal and shouldn't be interpreted as such!

MISTy is data-driven but it is also a knowledge-driven data exploration and hypothesis generation method.

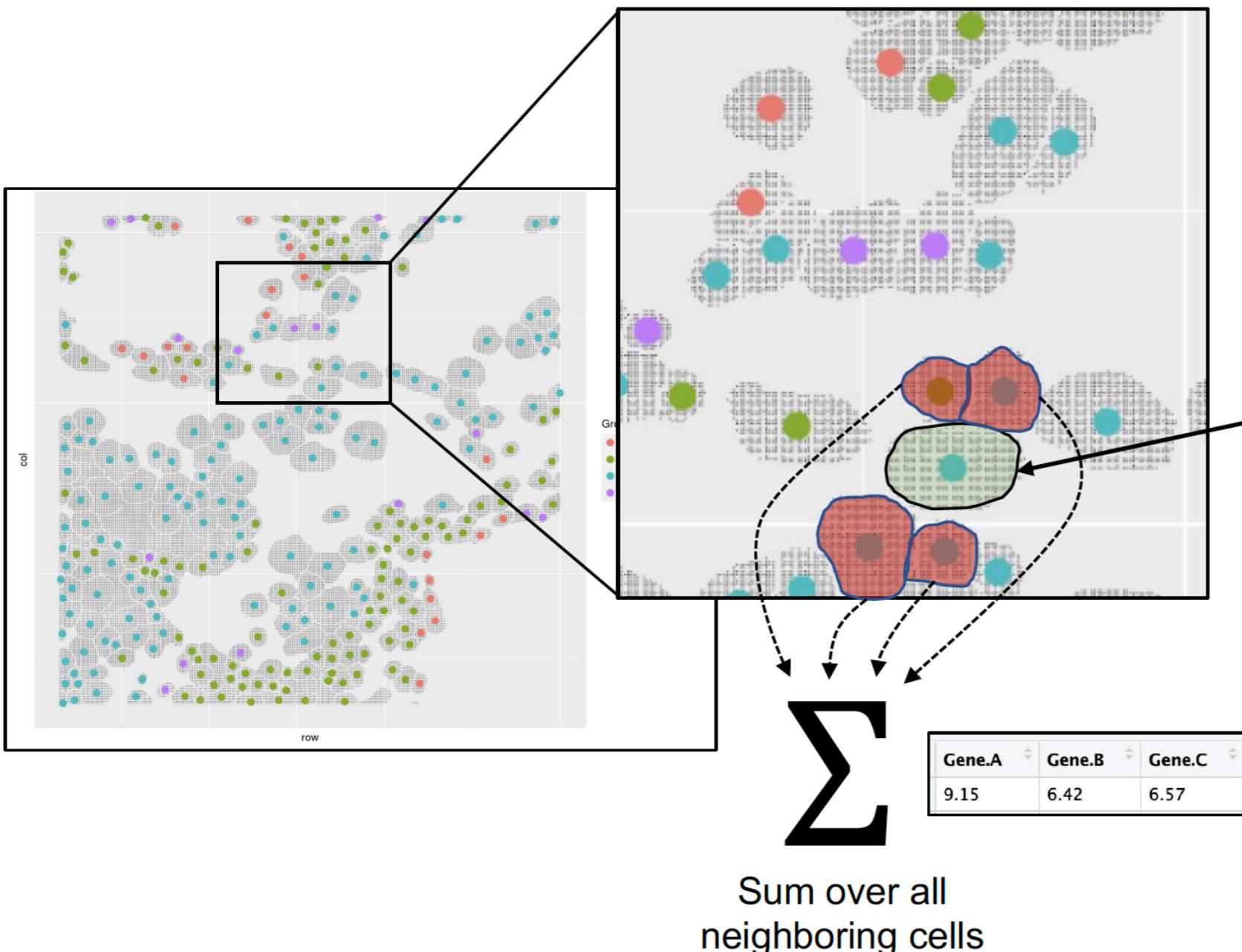
MISTy supports encoding domain knowledge in the form of views capturing different aspects of the data, steering(biasing) the analysis and the resulting hypotheses.



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

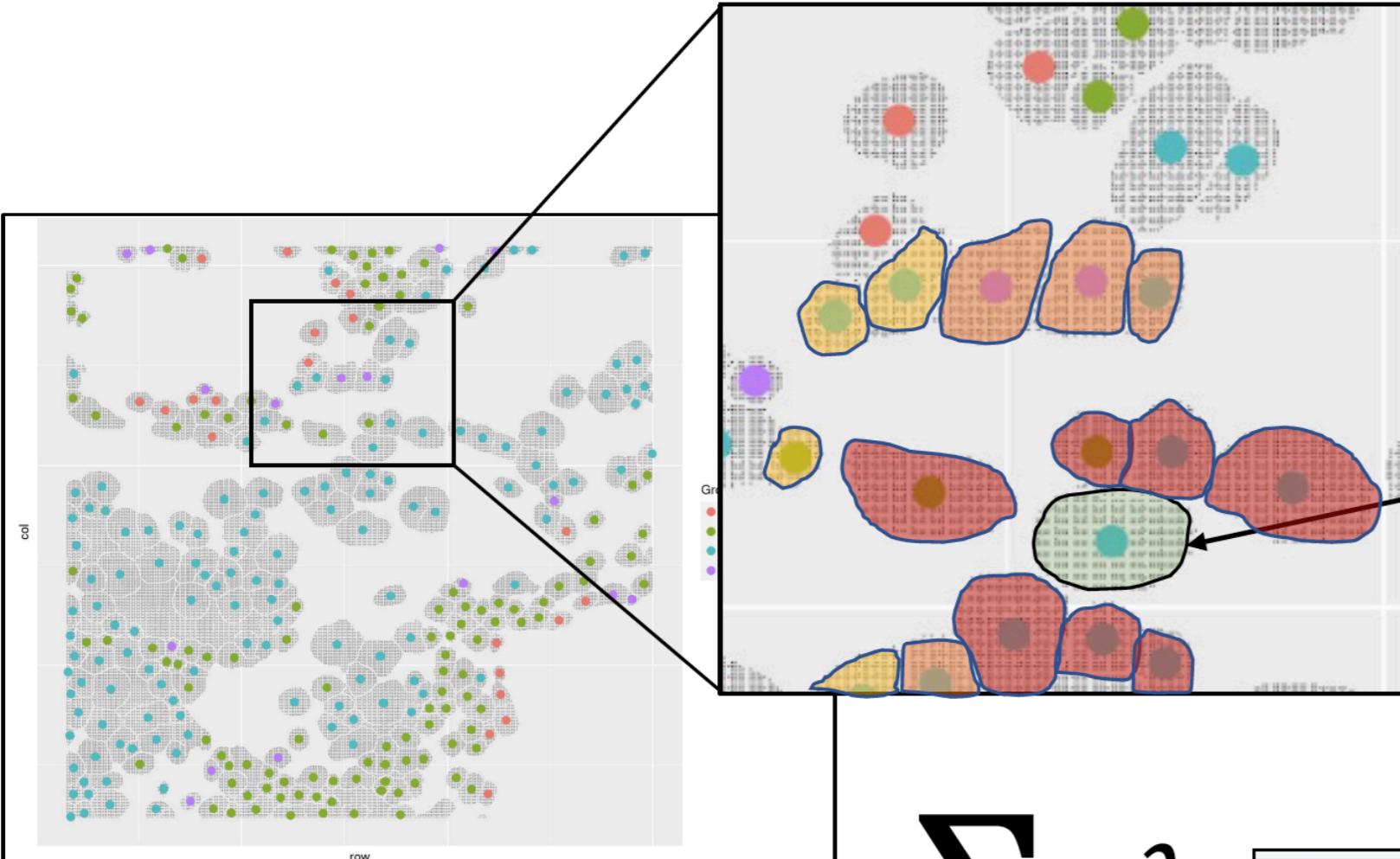
View-specific model

$$\text{Gene.A} = f(\text{Gene.B}, \text{Gene.C})$$



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.



$$\sum \lambda$$

Weighted sum

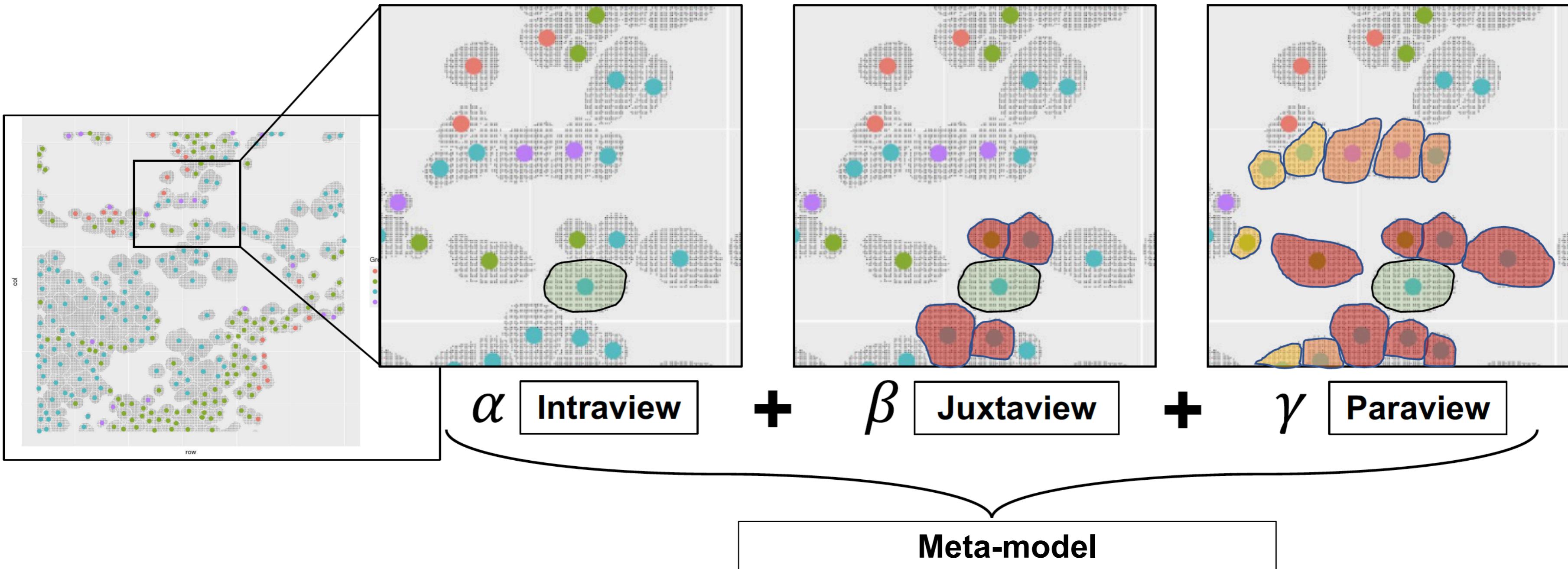
Gene.A	Gene.B	Gene.C
9.15	6.42	6.57

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.



Meta-model



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



Imaging Mass Cytometry

Proof of concept and added value of spatial component

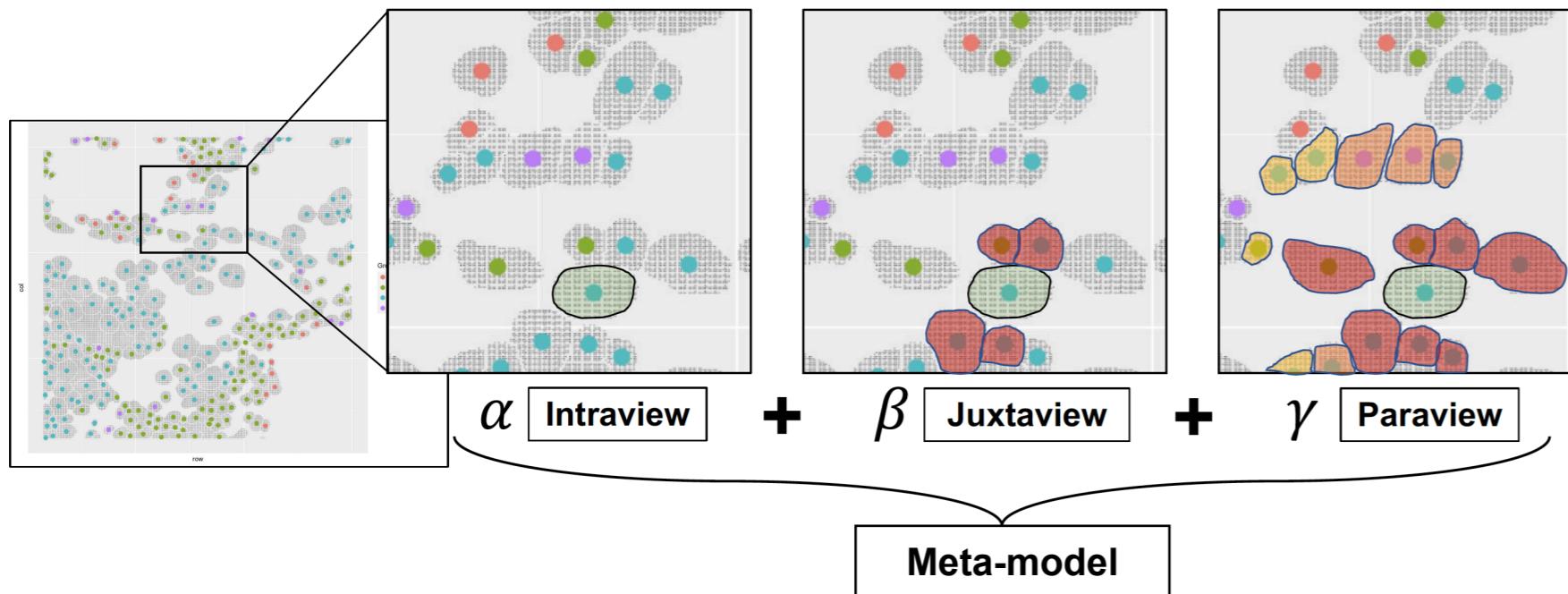
Tissue from breast cancer biopsies - 46 samples from 27 patients
Bodenmiller lab (Giesen et al. 2014, Schapiro et al. 2017)

267 - 1455 cells per image (average 900)
26 marker proteins

Single cell expression - median count per cell



MISTy pipeline

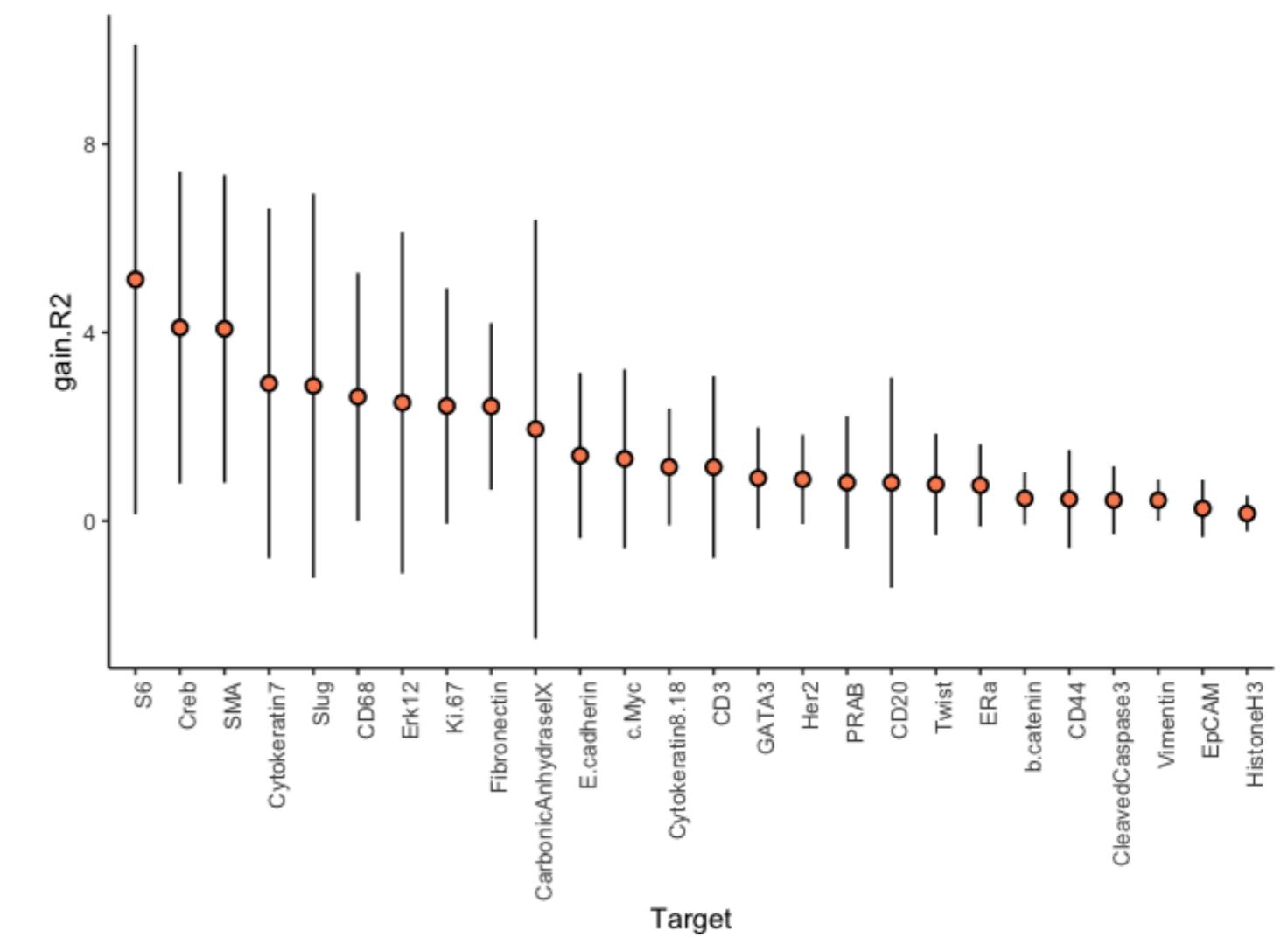
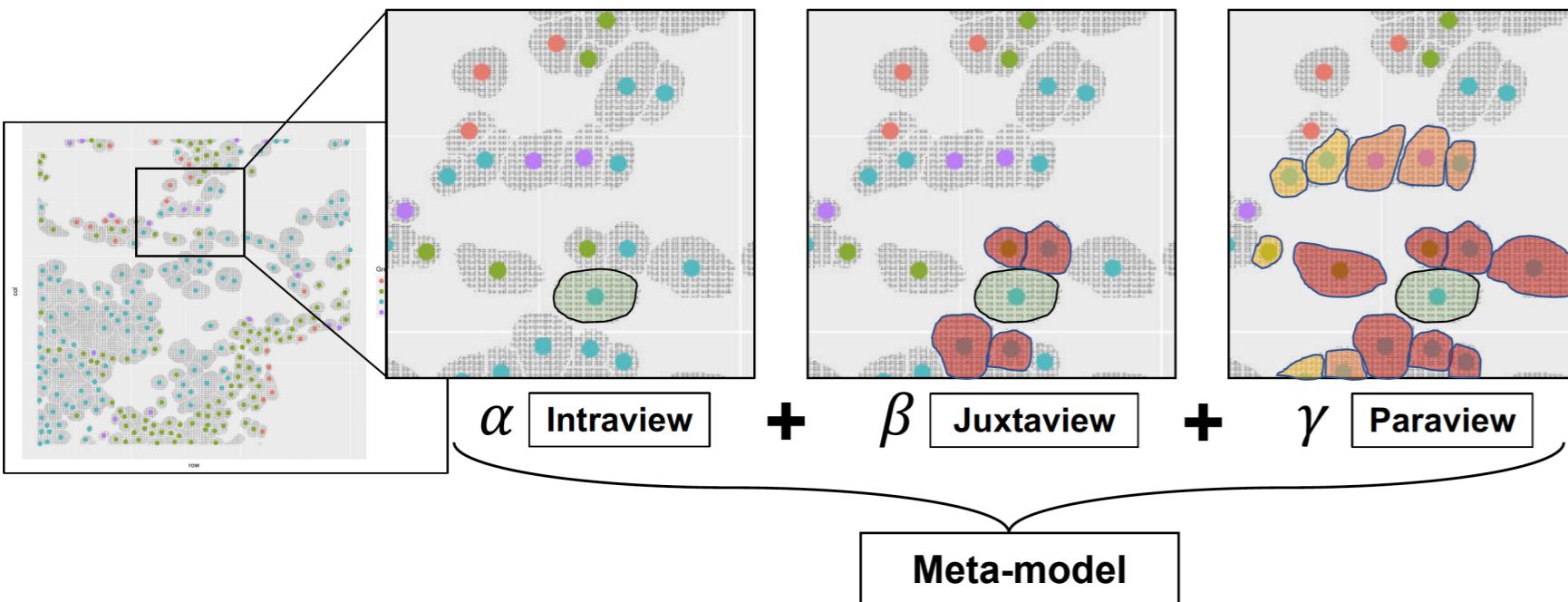


```
views <- create_initial_view(expr) %>% add_juxtaview(pos, threshold) %>% add_paraview(pos, l)
output <- run_misty(views, results.folder = "sample_result")
misty.results <- collect_results(output)
```



Results

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

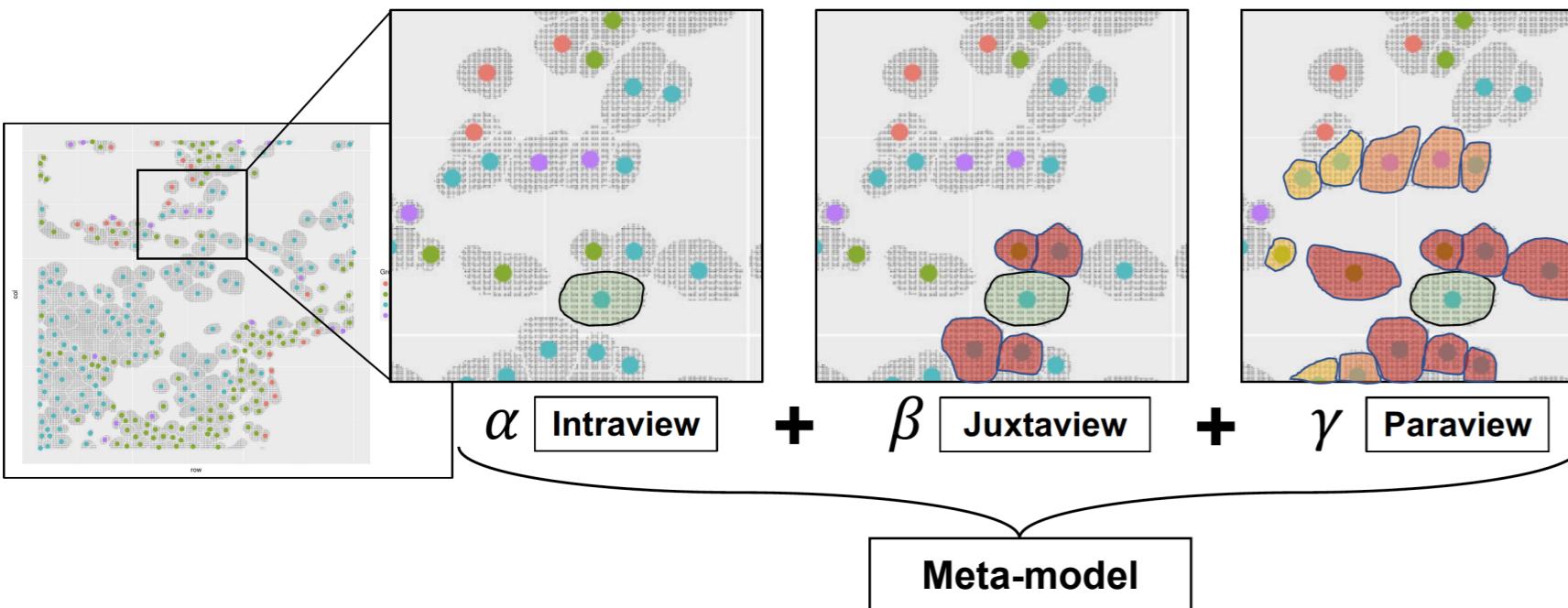


1. How much can the intercellular (spatial) context explain expression (in contrast to intracellular)?

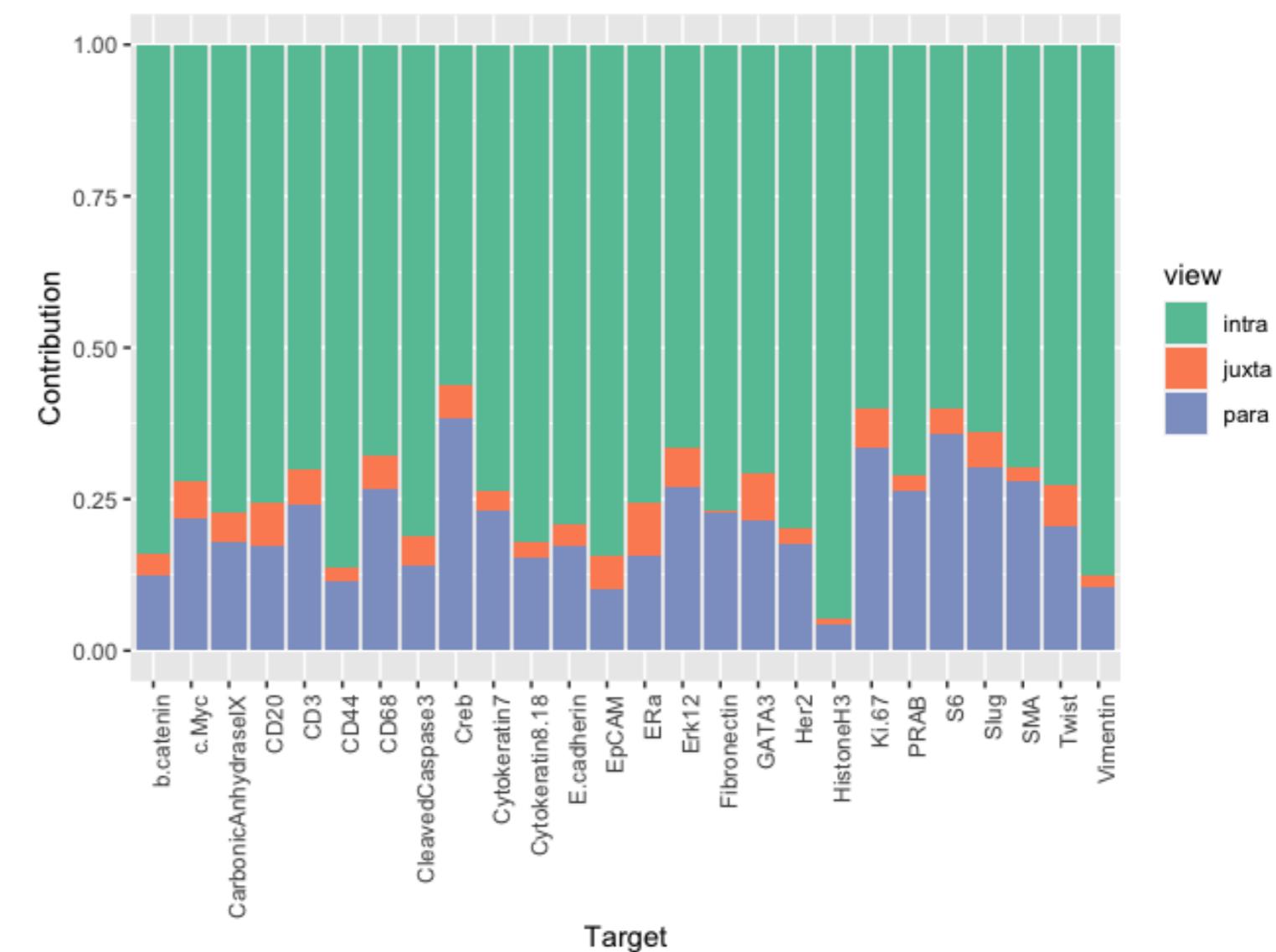


Results

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



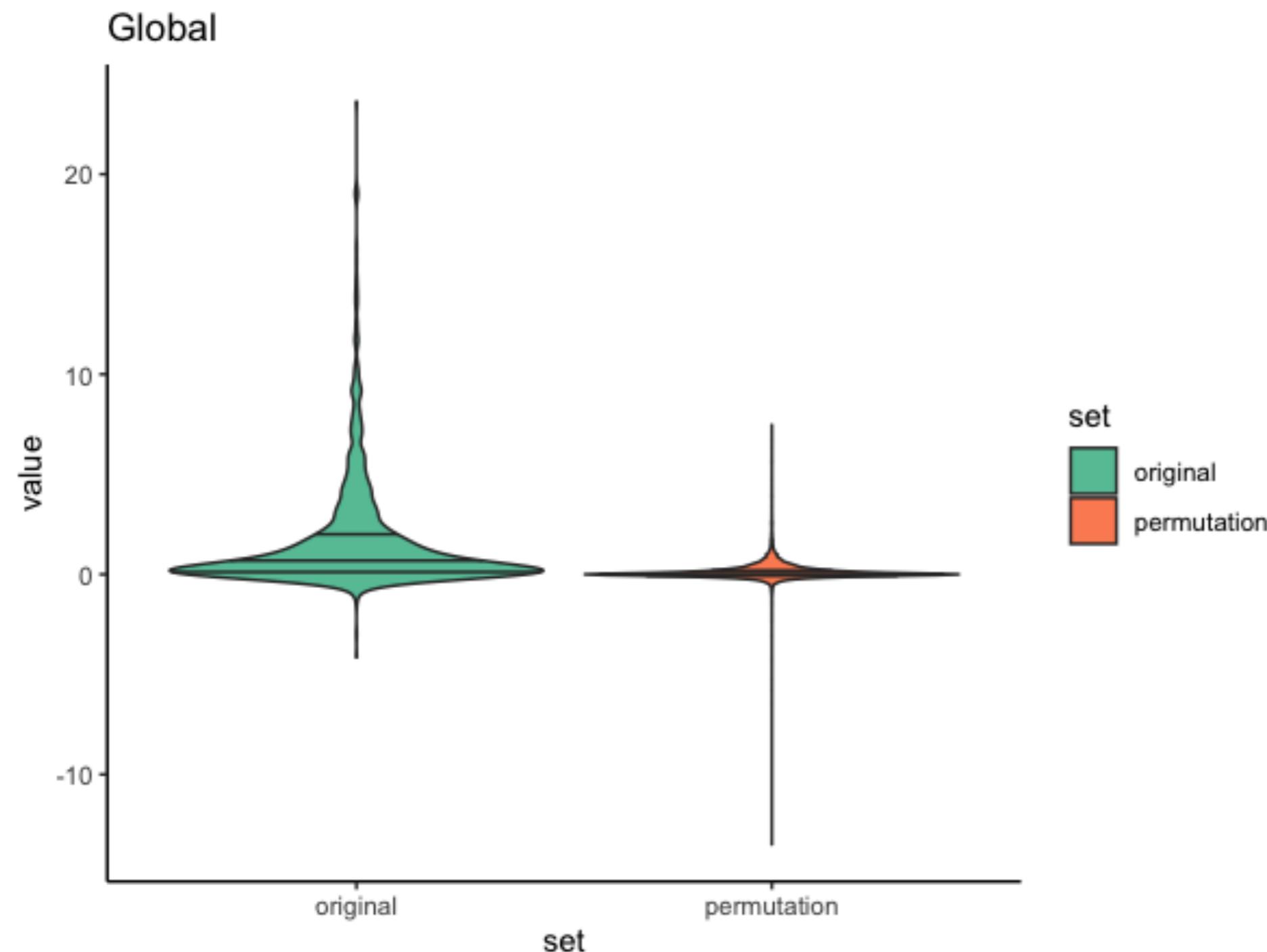
2. How much do different view components contribute to explaining the expression?





Permutation analysis

Permute the cell locations and rerun MISTy 10 times



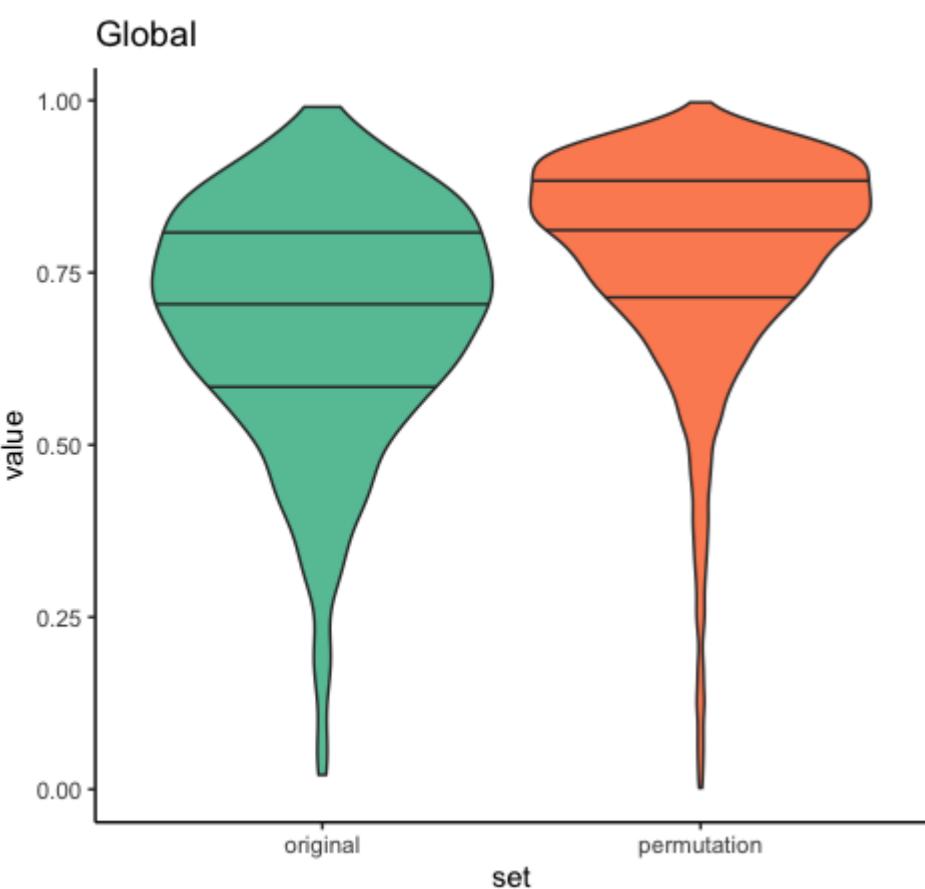
Distribution of gain in variance explained
across all targets and samples

Mann-Whitney U
 $p \approx 0$



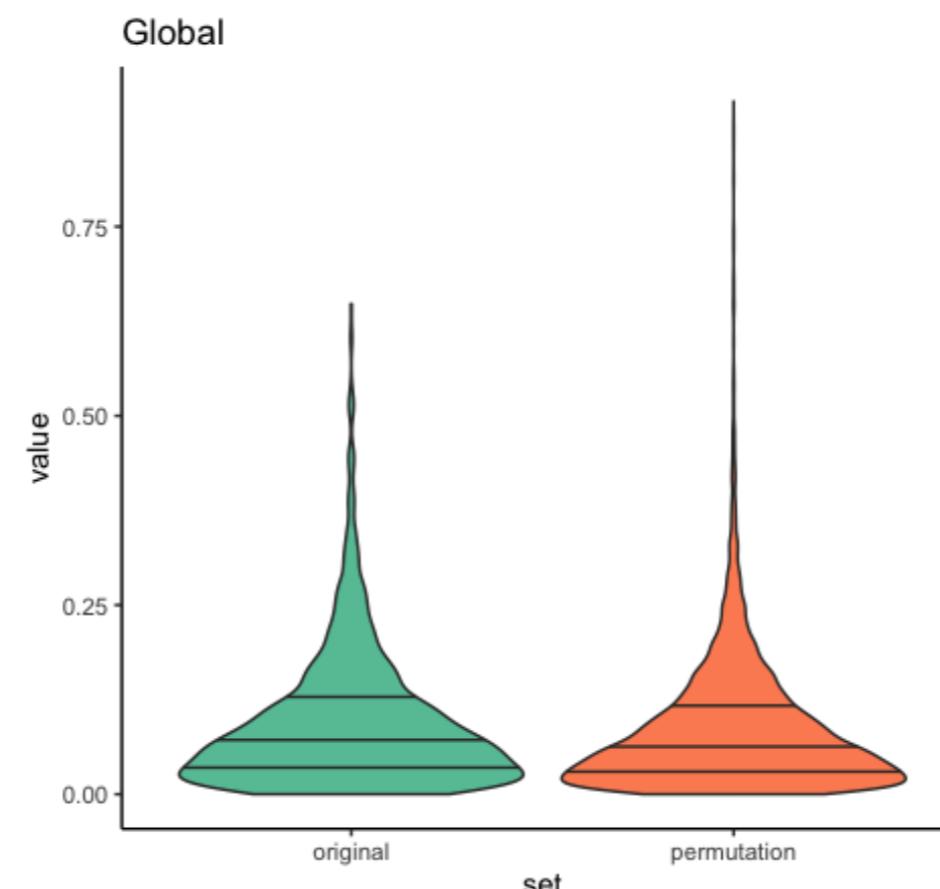
Distribution of view contributions

intra



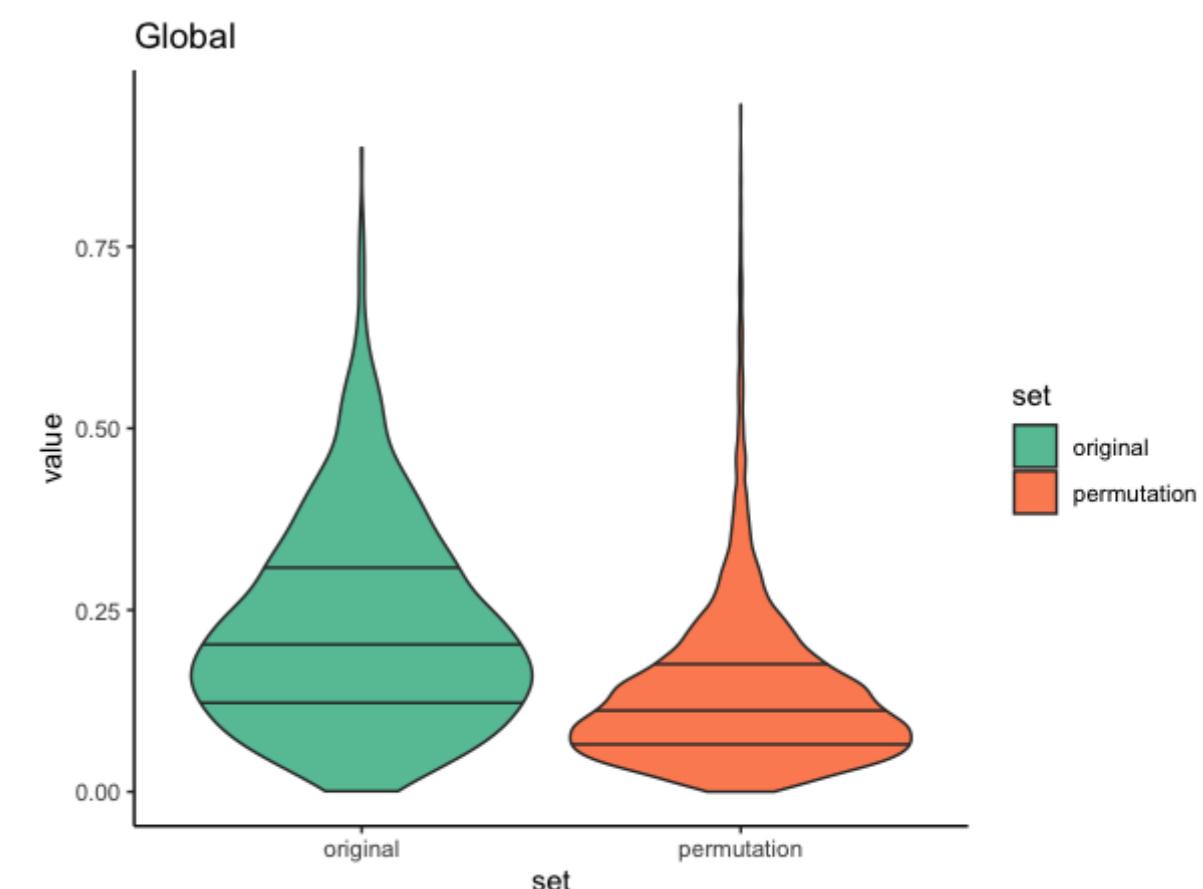
Mann-Whitney U
 $p \approx 0$

juxta



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

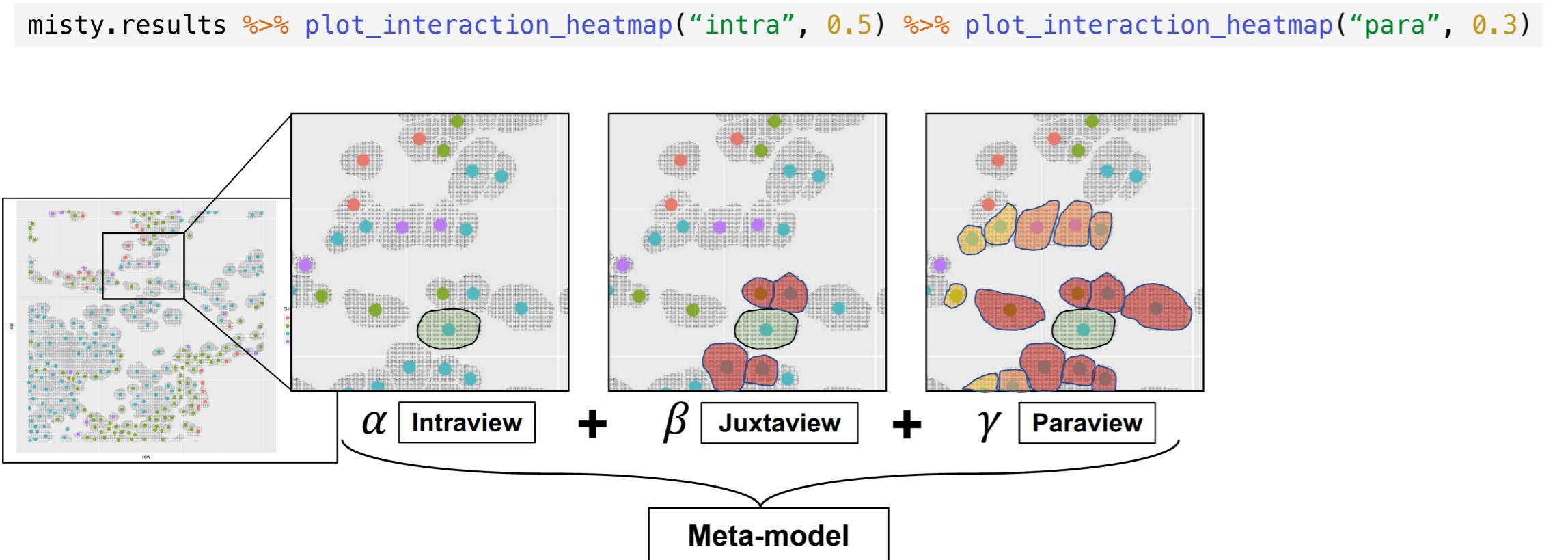
para



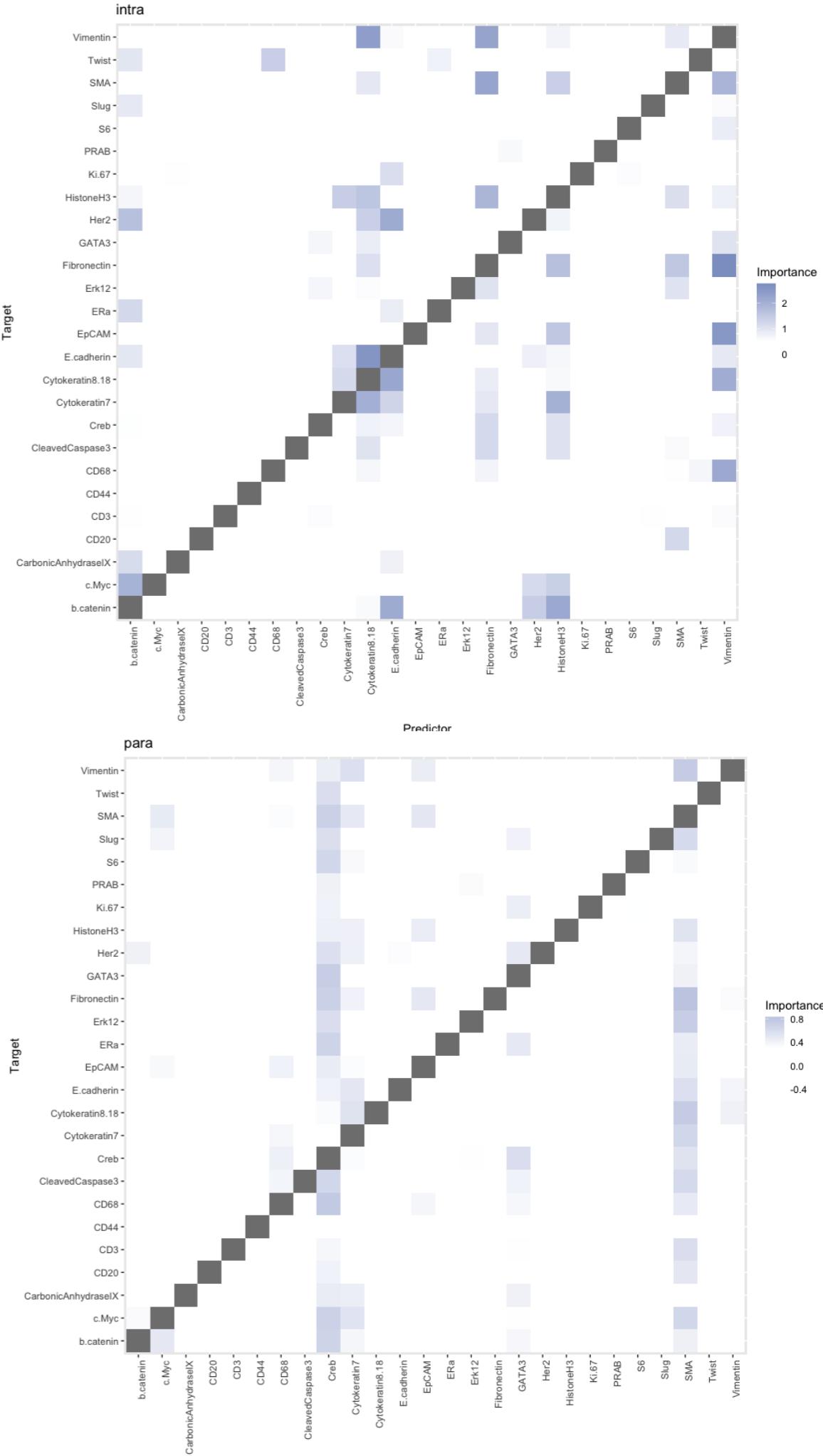
Mann-Whitney U
 $p \approx 0$



Results

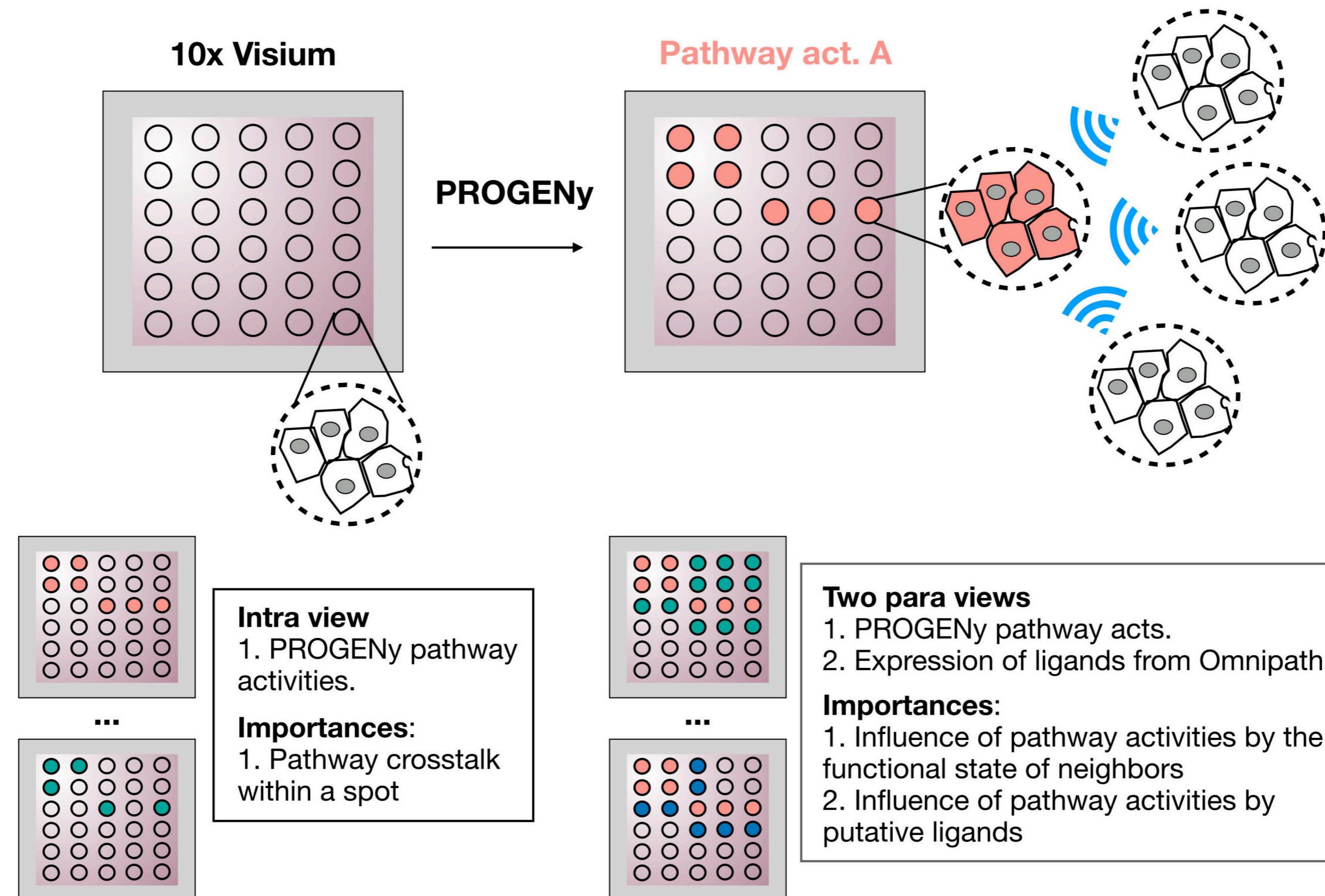


3. What are the specific relations that can explain the contributions?





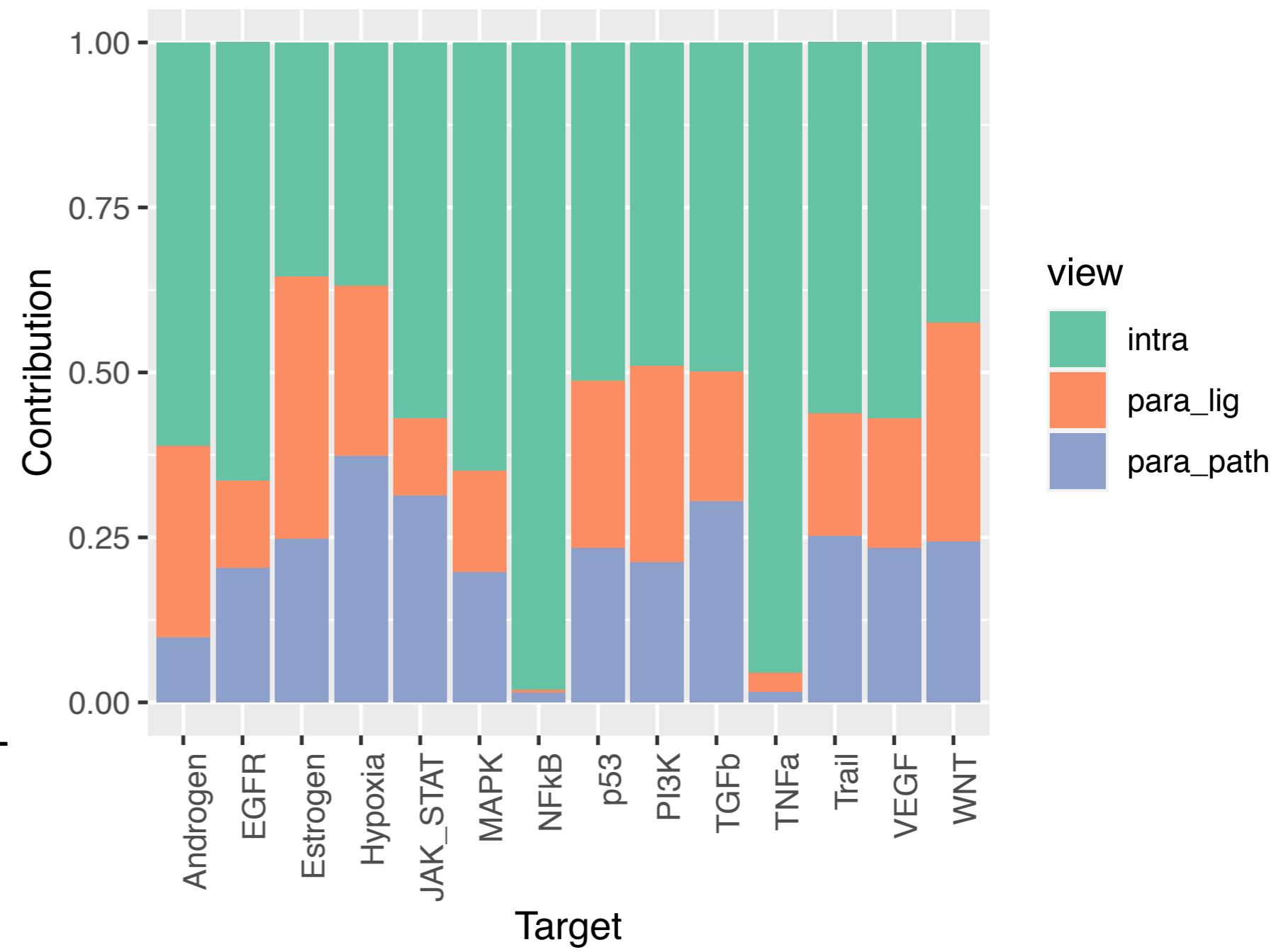
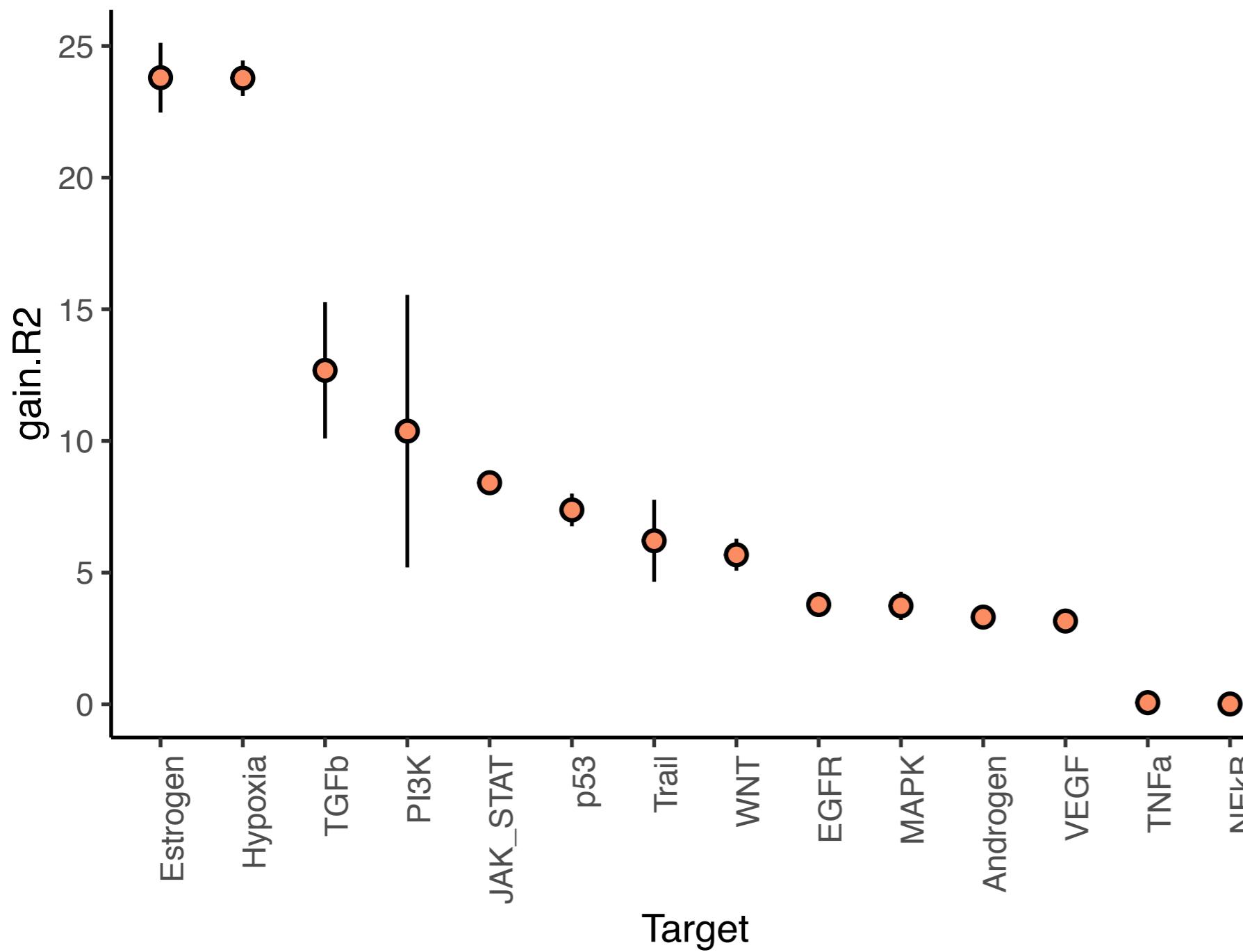
Functional pipeline, Visium data





MISTy results

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

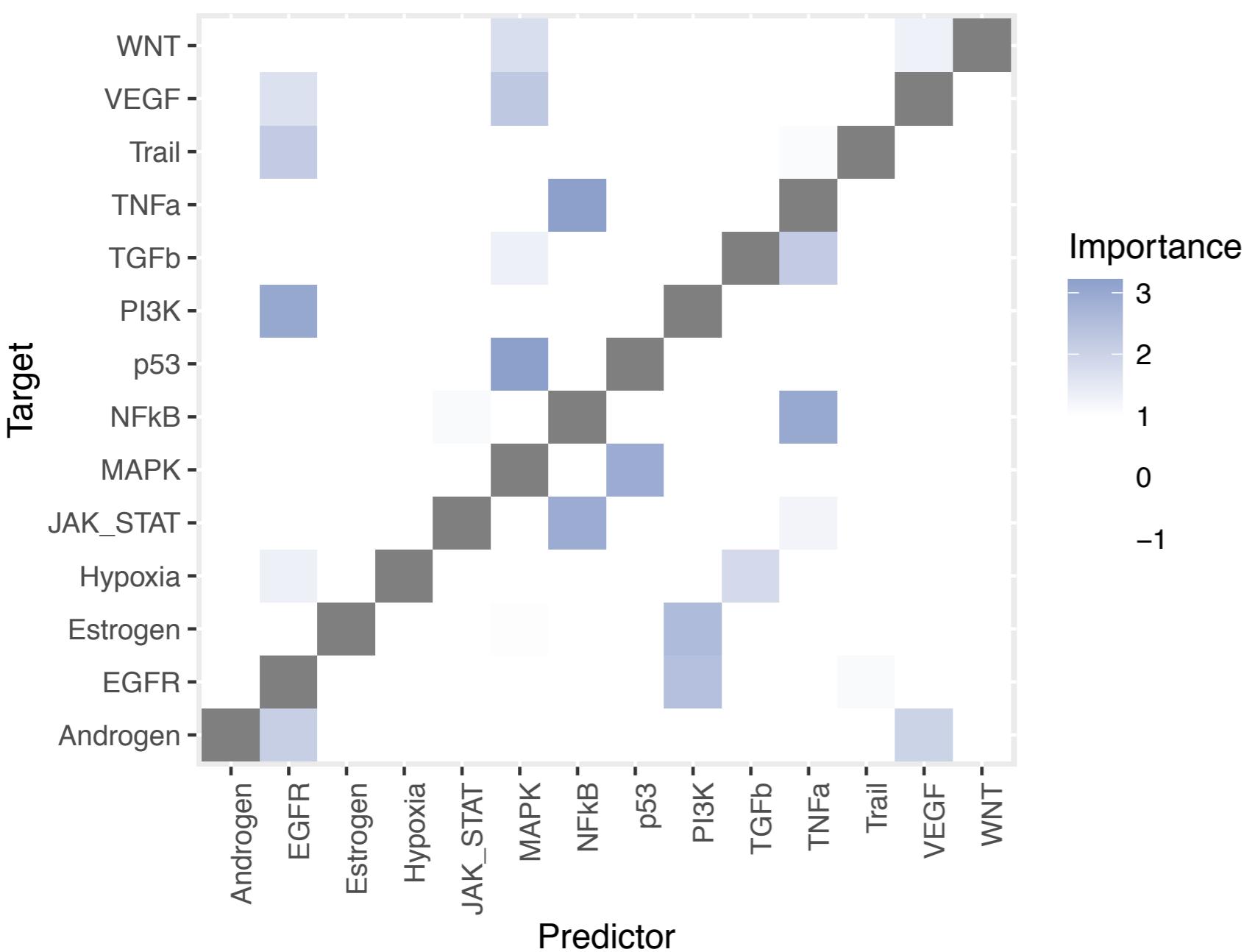




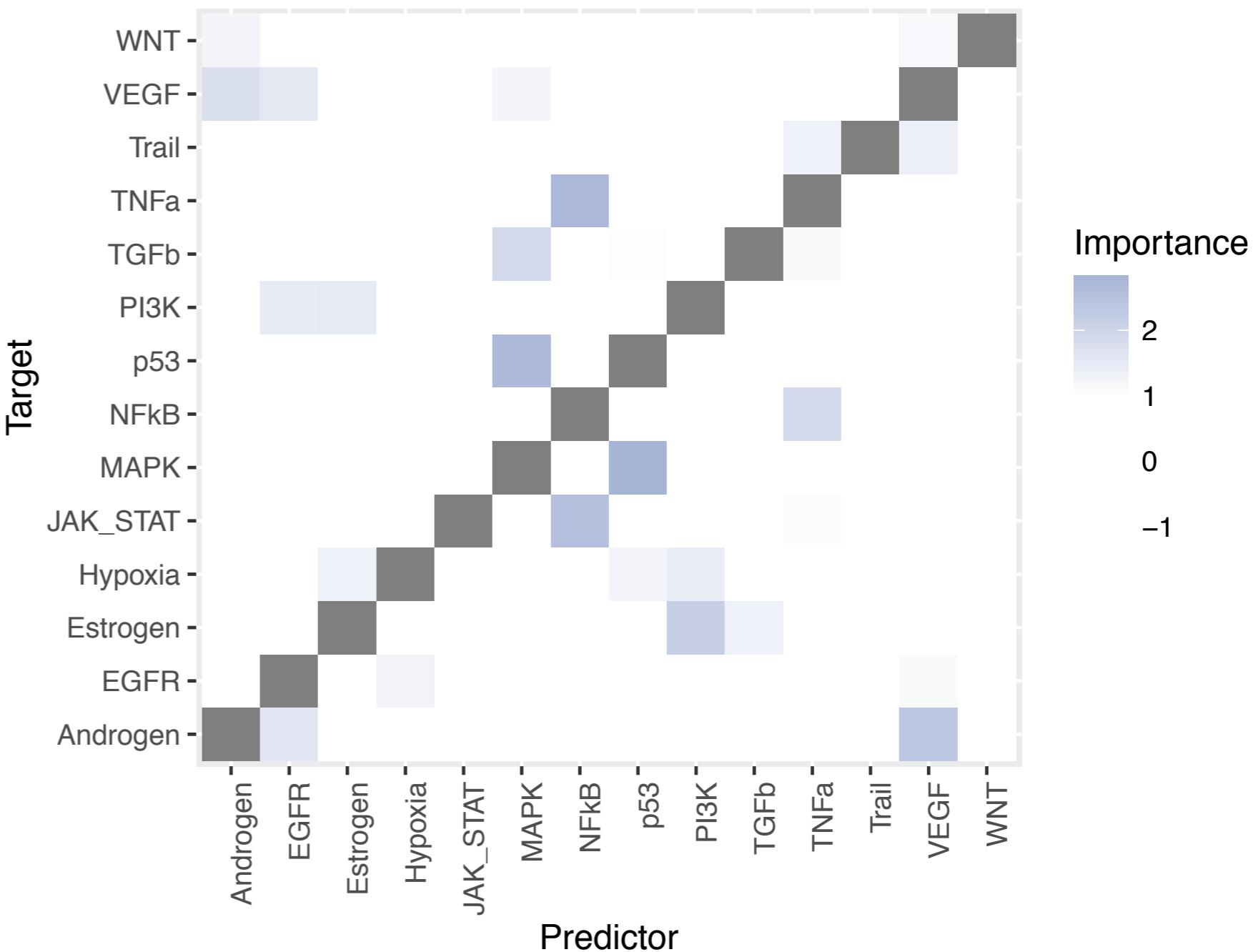
Pathway crosstalk

```
misty.results %>% plot_interaction_heatmap("intra", 1) %>% plot_interaction_heatmap("para_path", 1)
```

Intra



Para

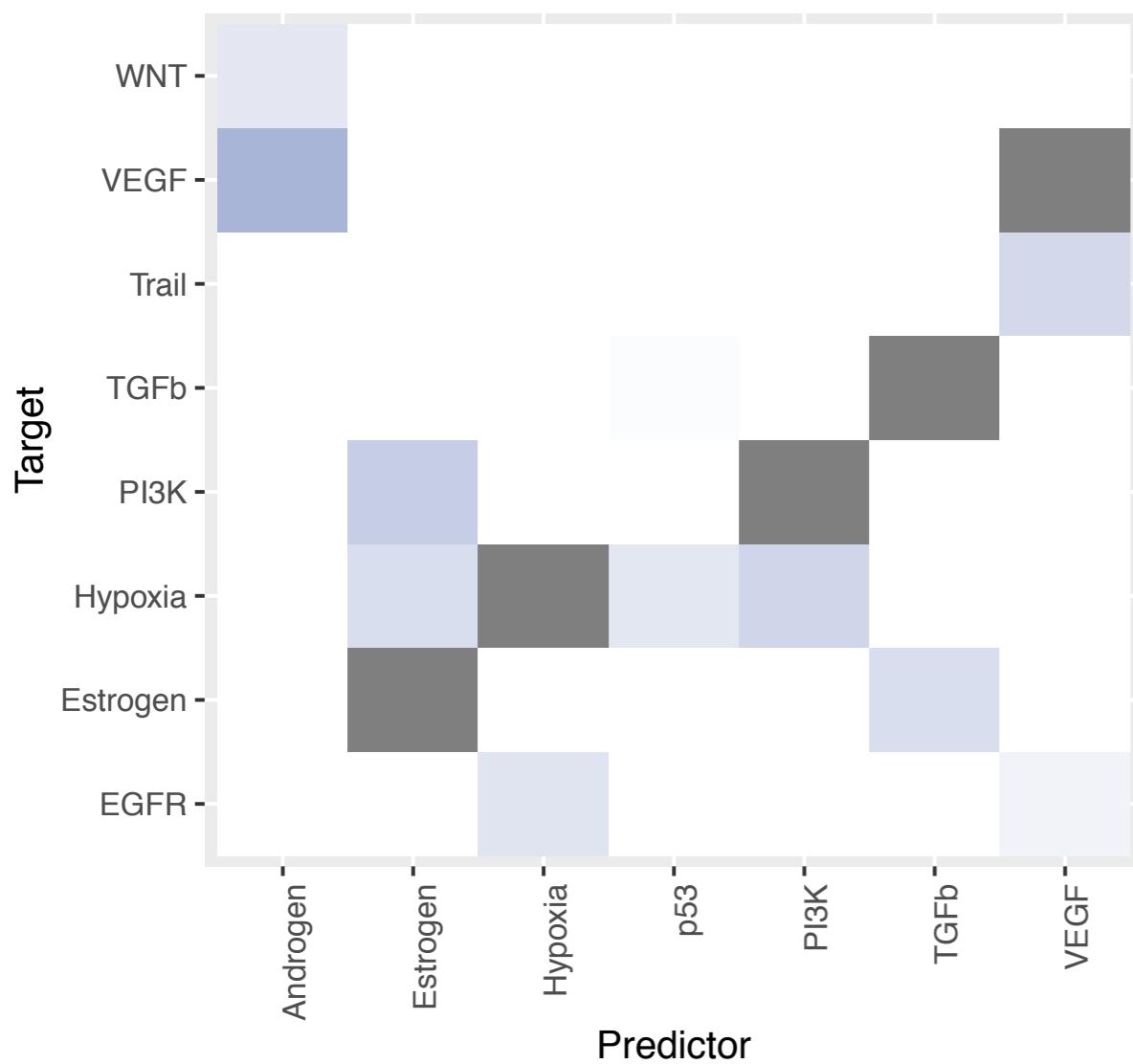




Pathway crosstalk coming from the tissue structure view

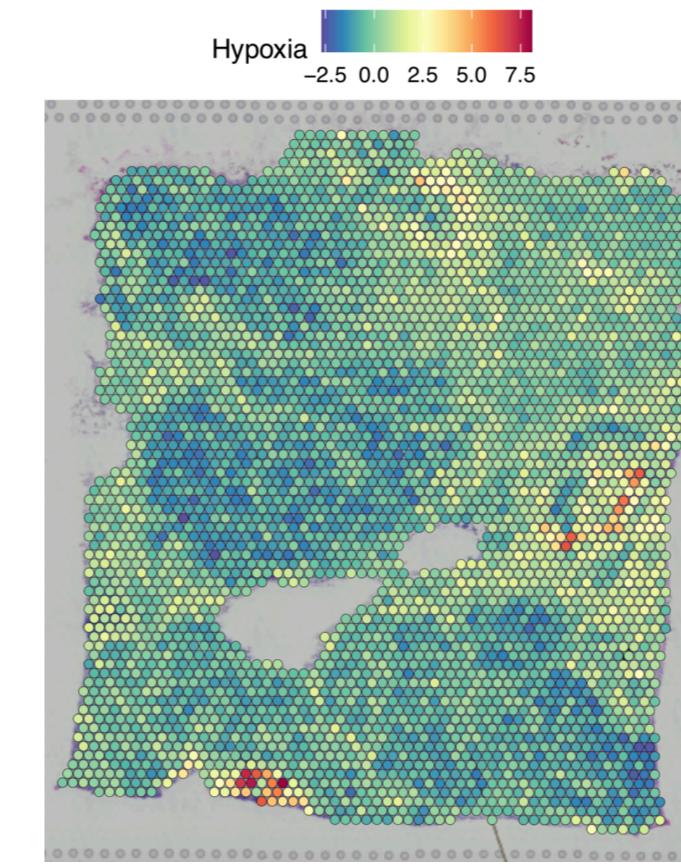
```
misty.results %>% plot_contrast_heatmap("intra", "para_path", 1)
```

para_path - intra

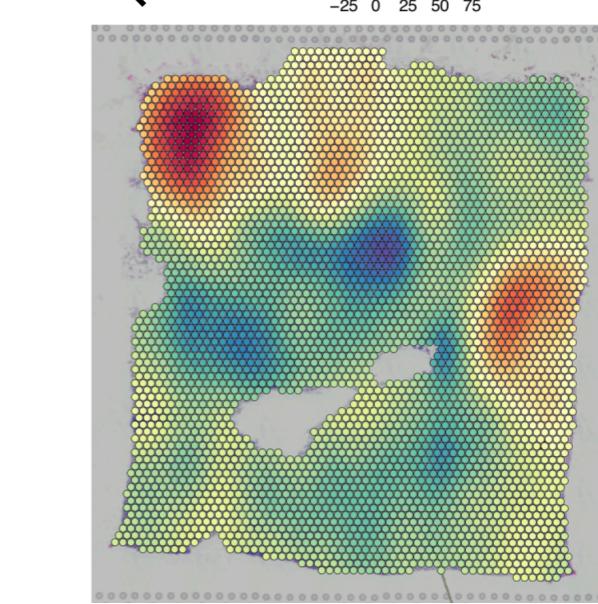


Importance

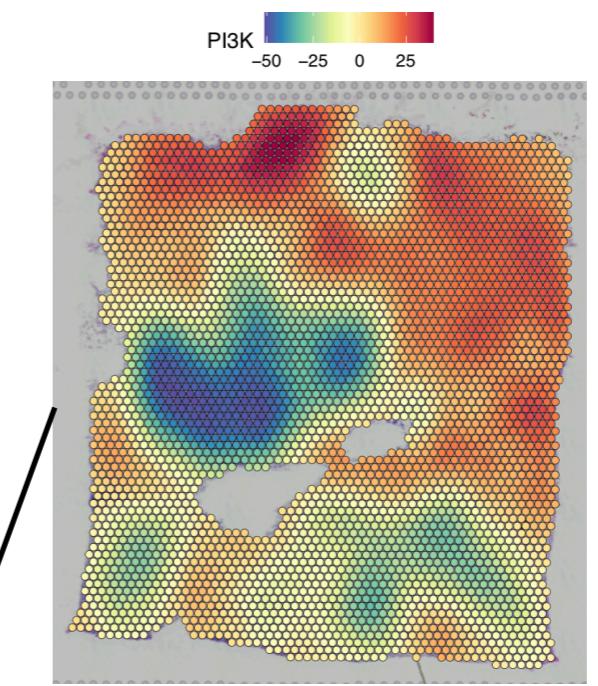
1.5
1.0
0.5
0.0



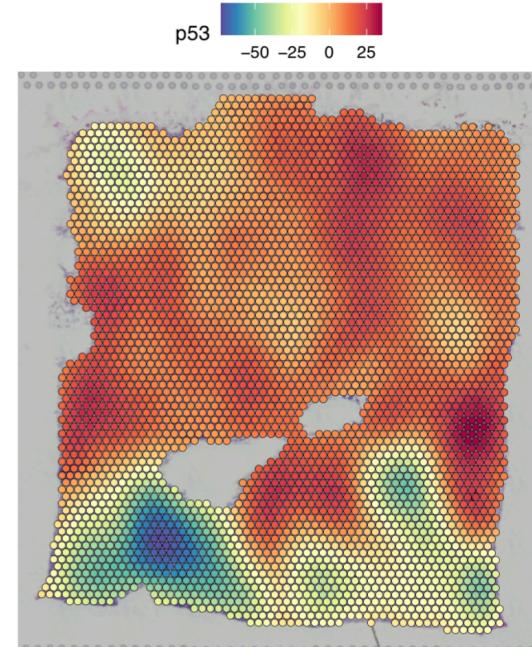
Hypoxia
-2.5 0.0 2.5 5.0 7.5



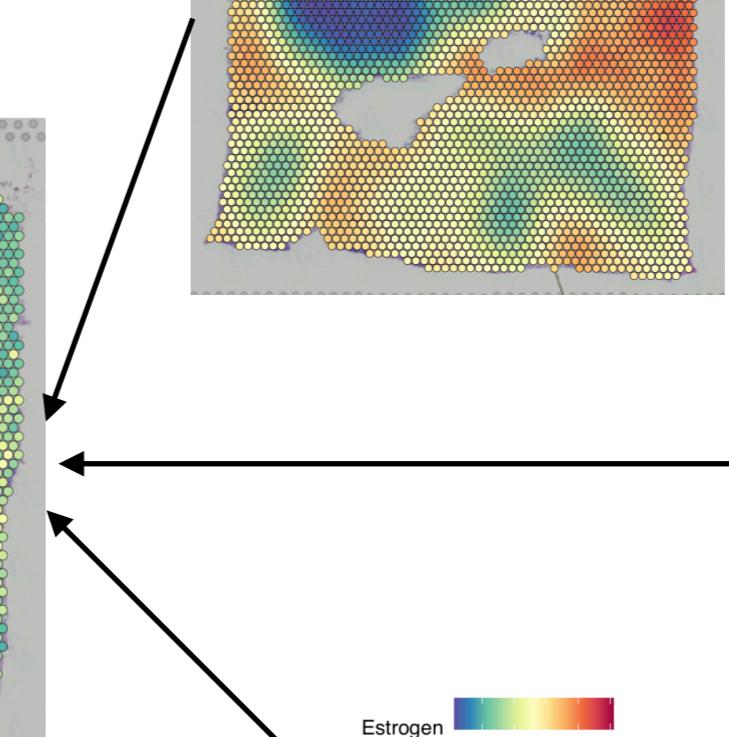
Estrogen
-25 0 25 50 75



PI3K
-50 -25 0 25



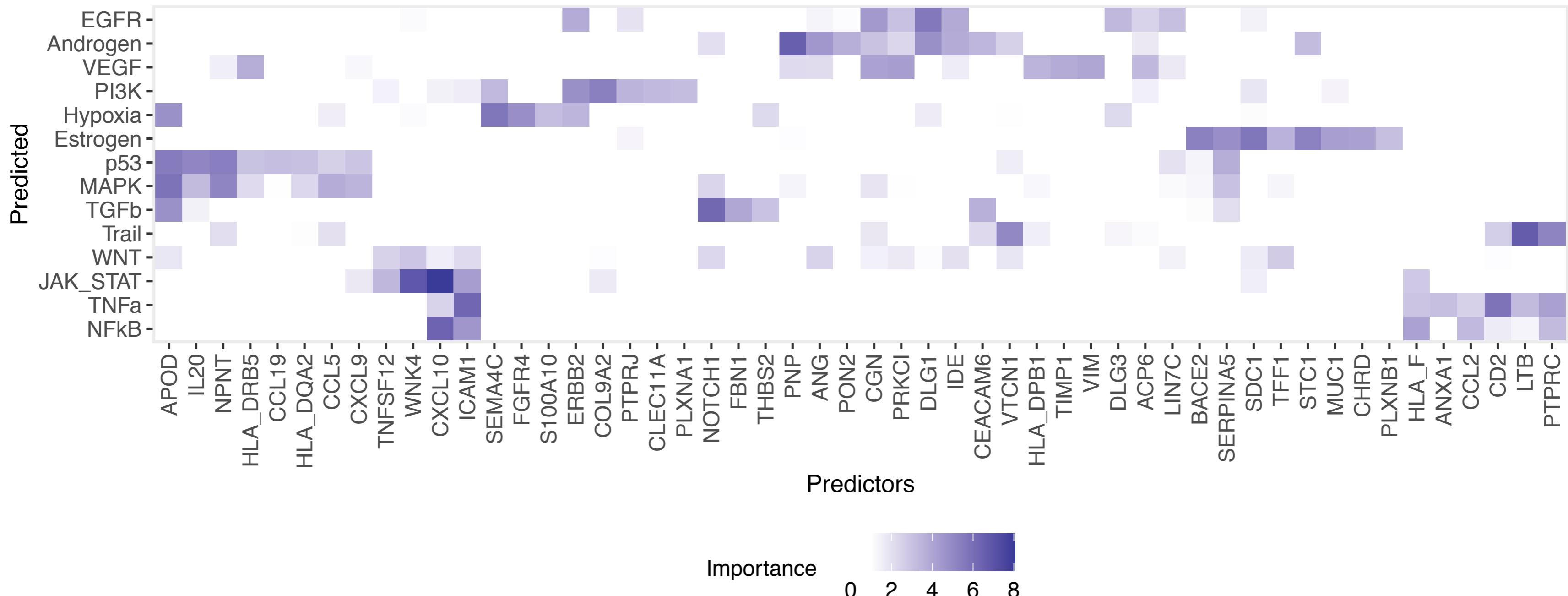
p53
-50 -25 0 25





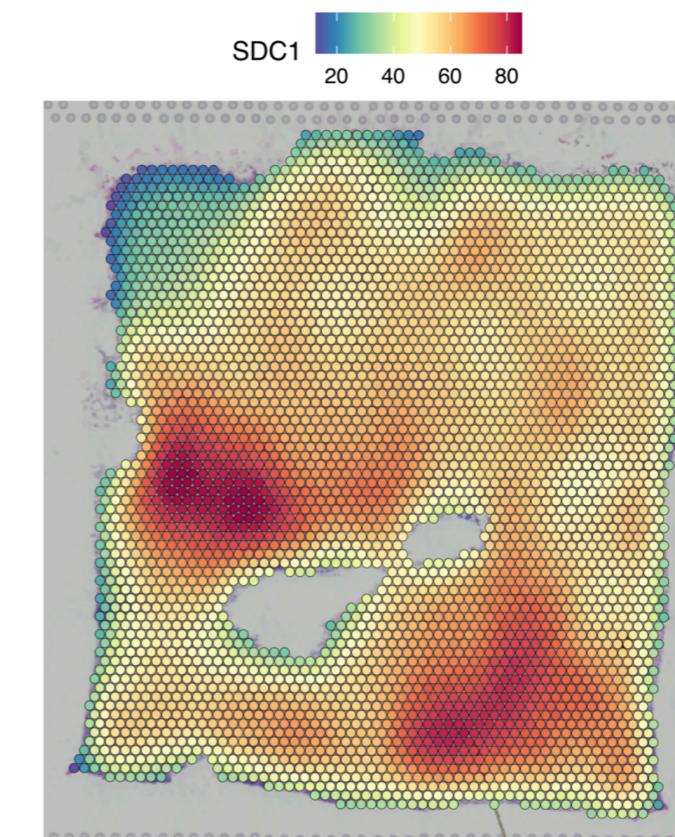
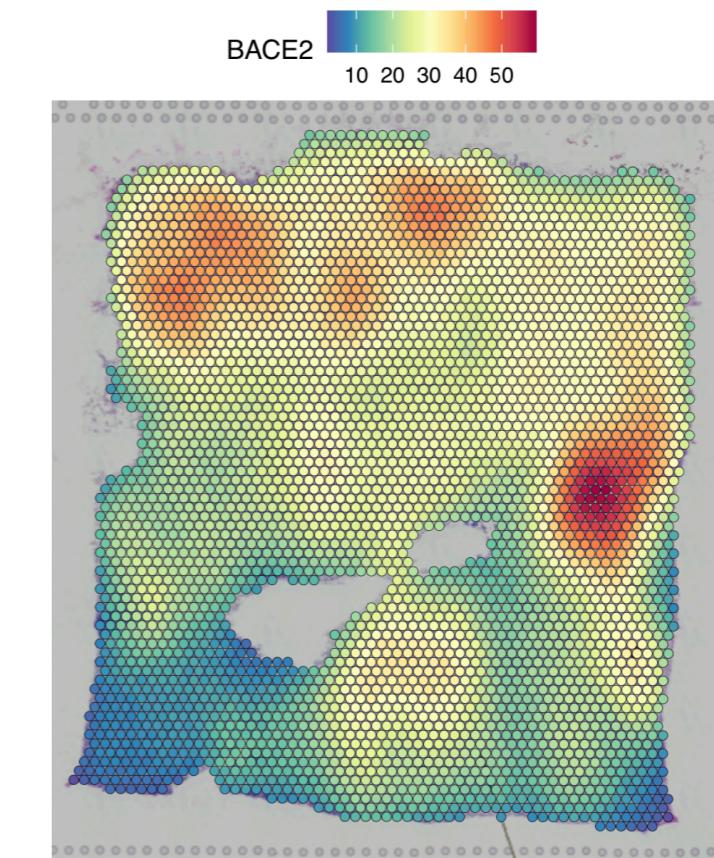
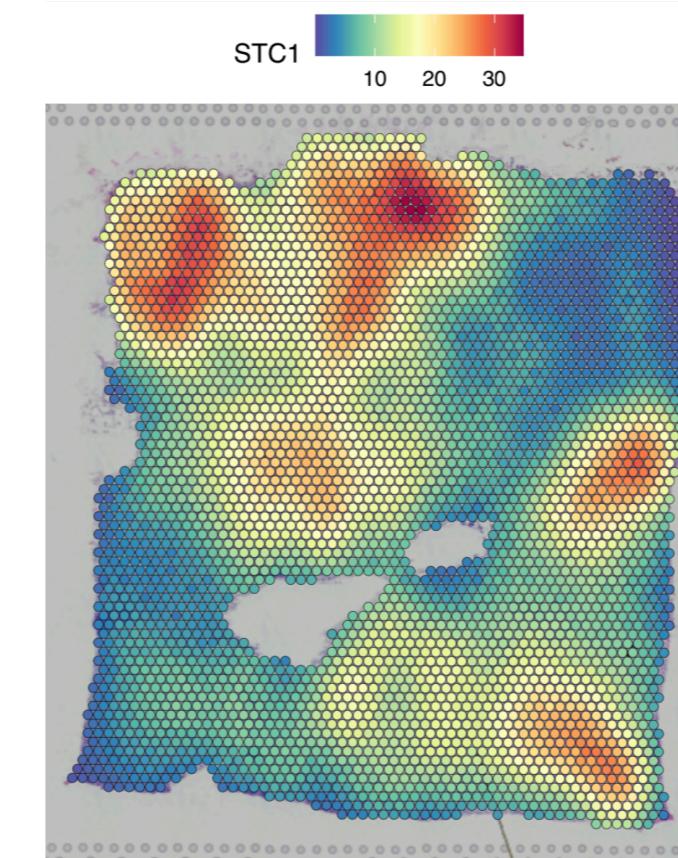
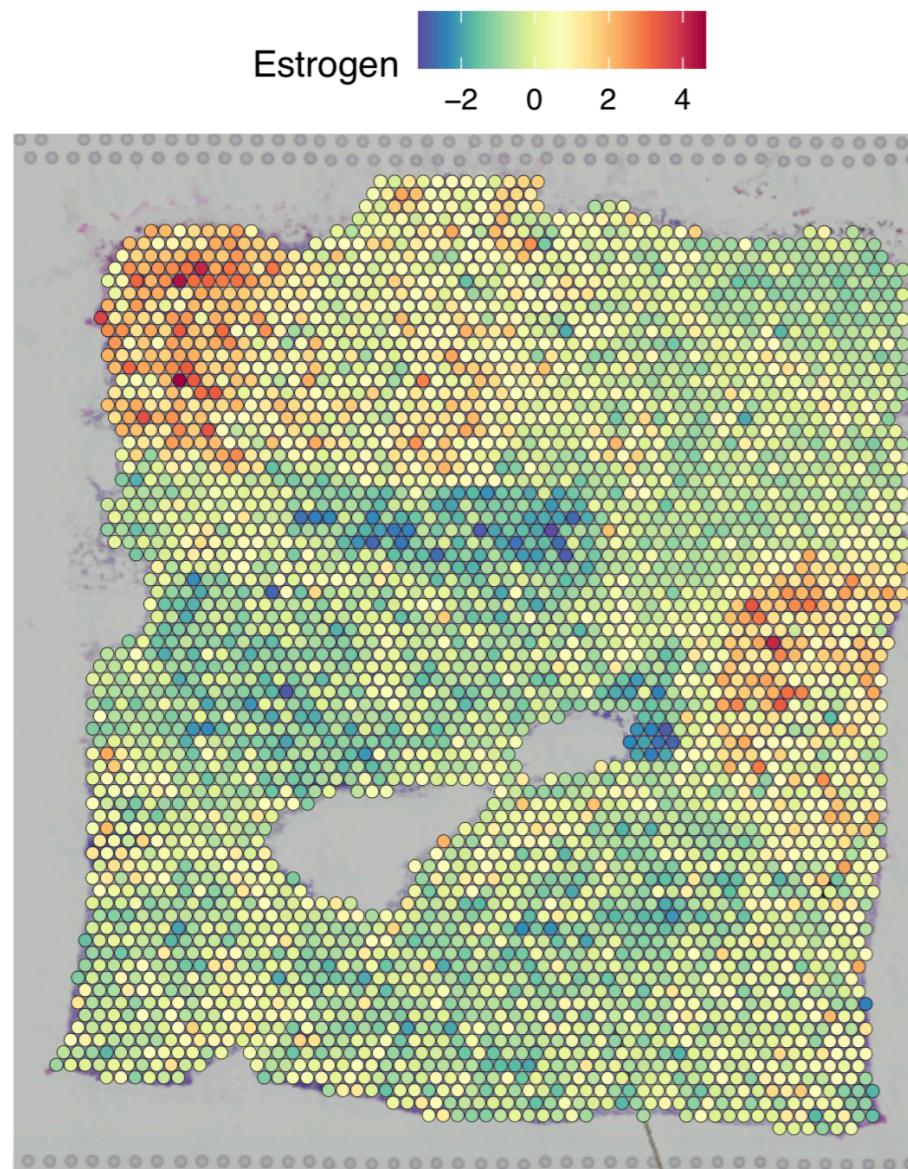
Ligand expression in the tissue structure explaining pathway activity

```
misty.results %>% plot_interaction_heatmap("para_lig", 1)
```



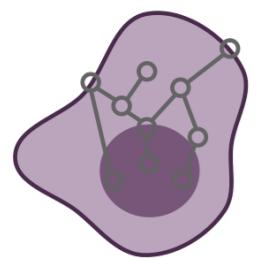


The interactions are not simply linear or independent



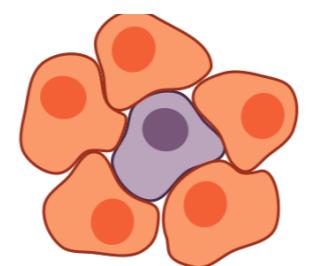


Structural pipeline, IMC data



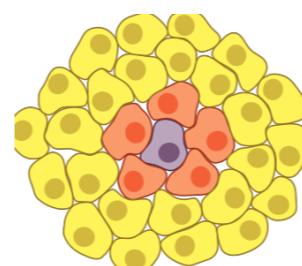
Intrinsic view

+



Local view
(juxtapview)

+



Tissue view
(paraview)

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

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

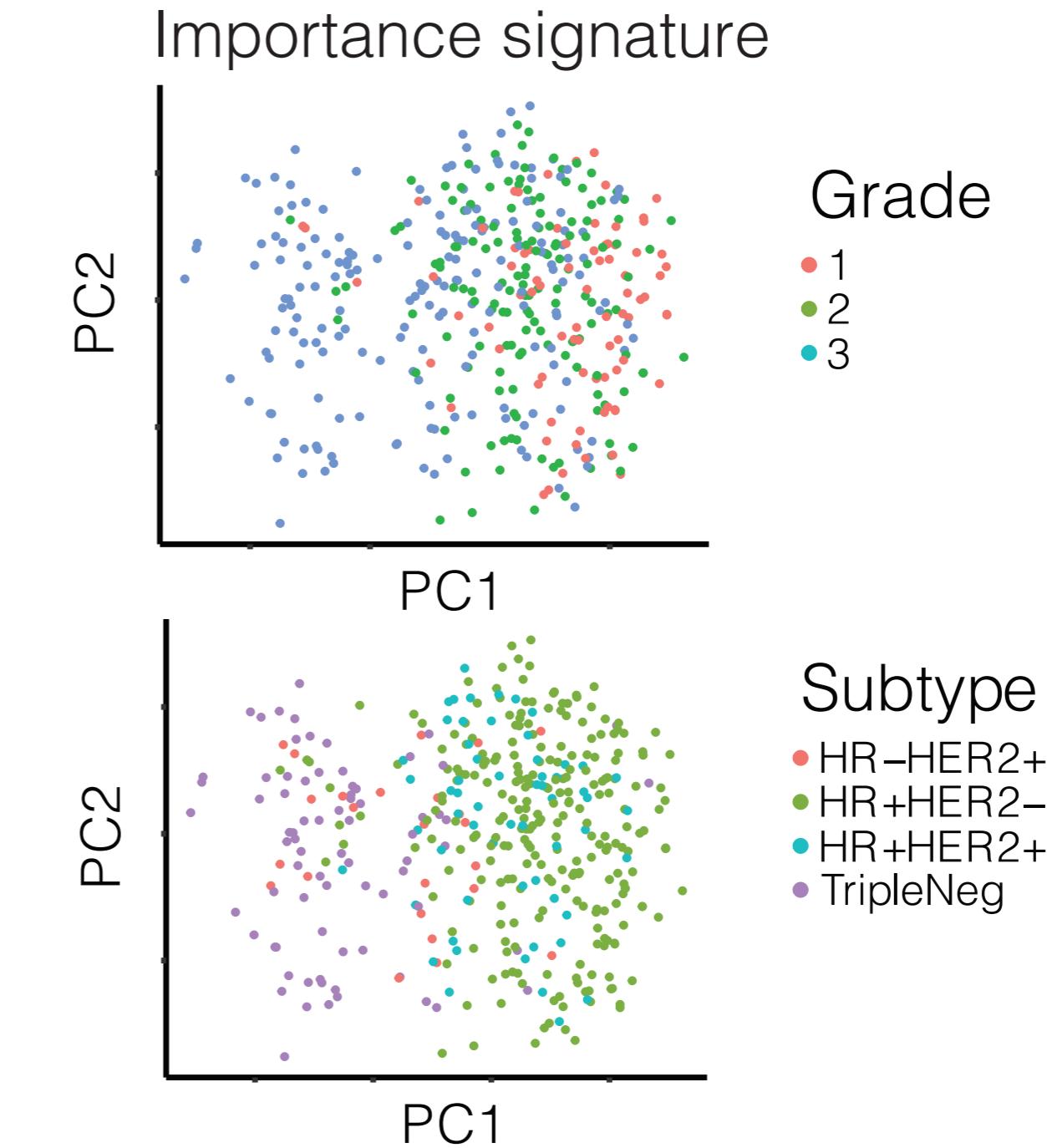
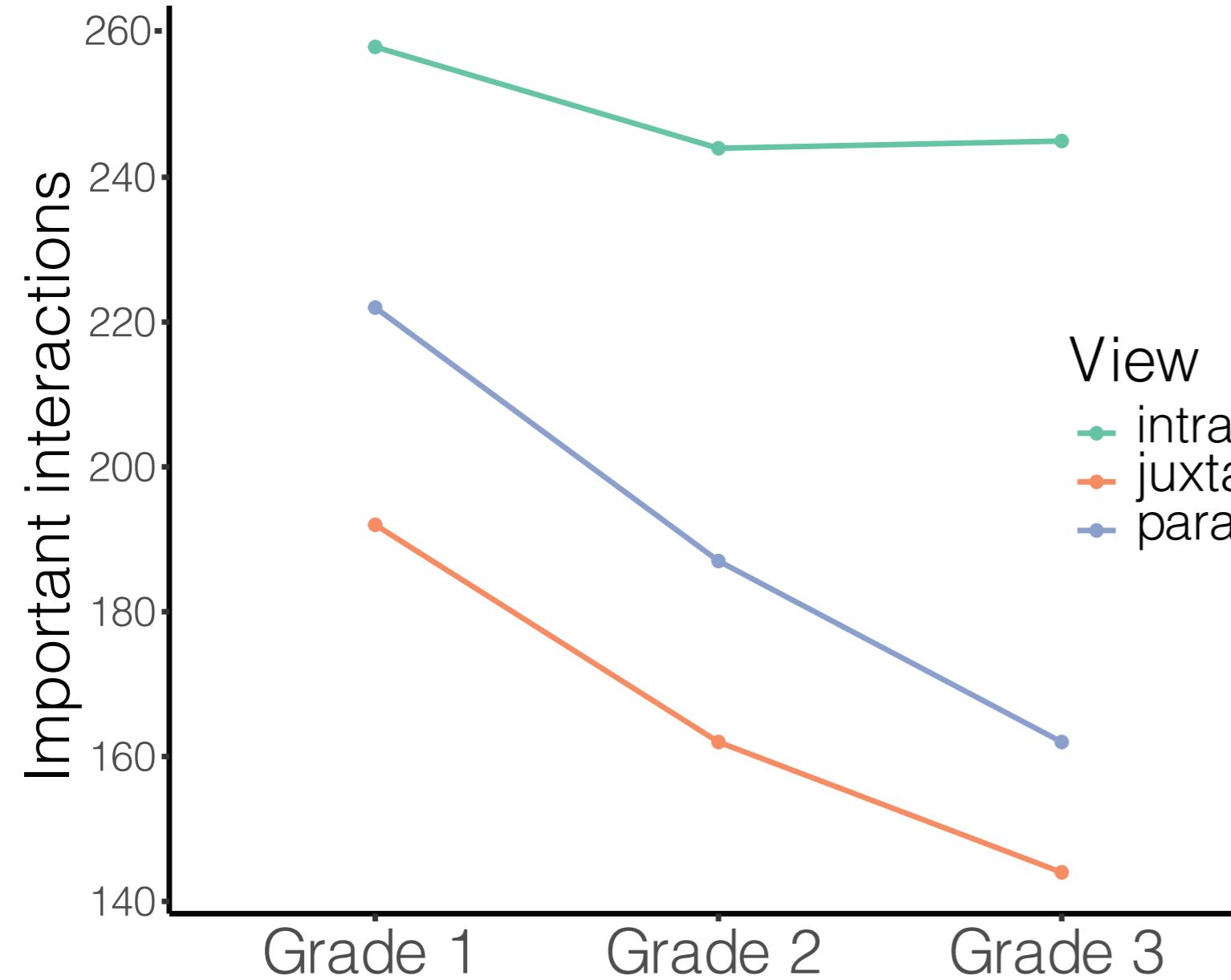
misty.results <- collect_results(output)
```



Further analyses and hypothesis generation

Large breast cancer IMC cohort. Analysis of results based on estimated importances of interaction.

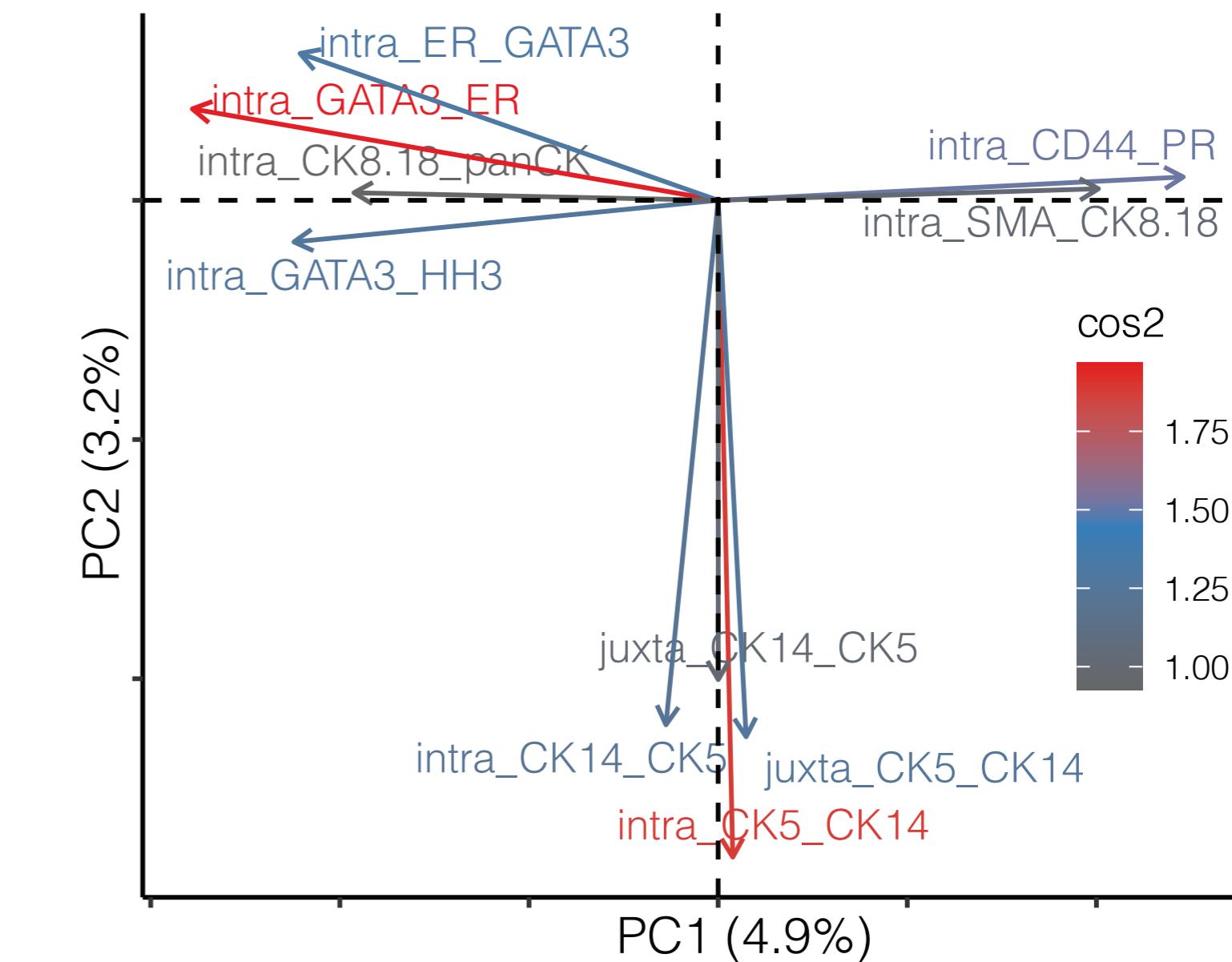
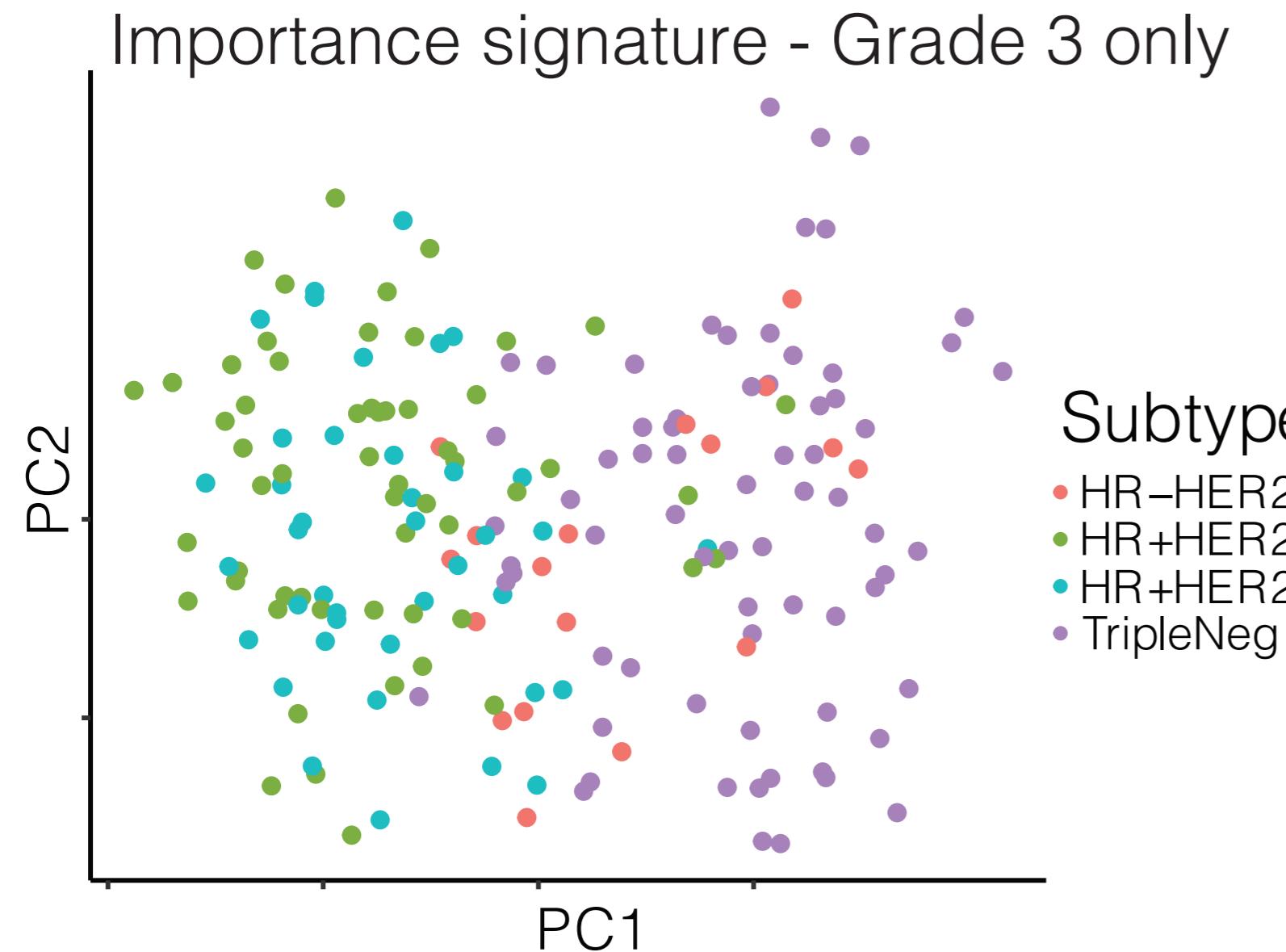
```
misty.results %>% extract_signature("importance")
```





Further analyses and hypothesis generation

Large breast cancer IMC cohort. Analysis of results based on estimated importances of interaction.

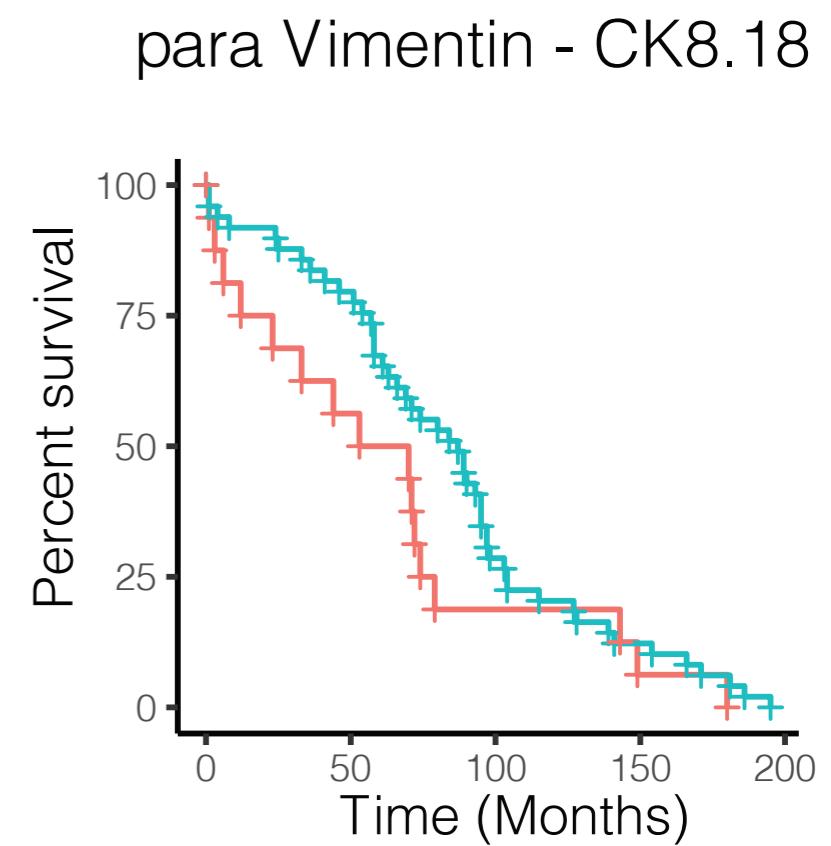
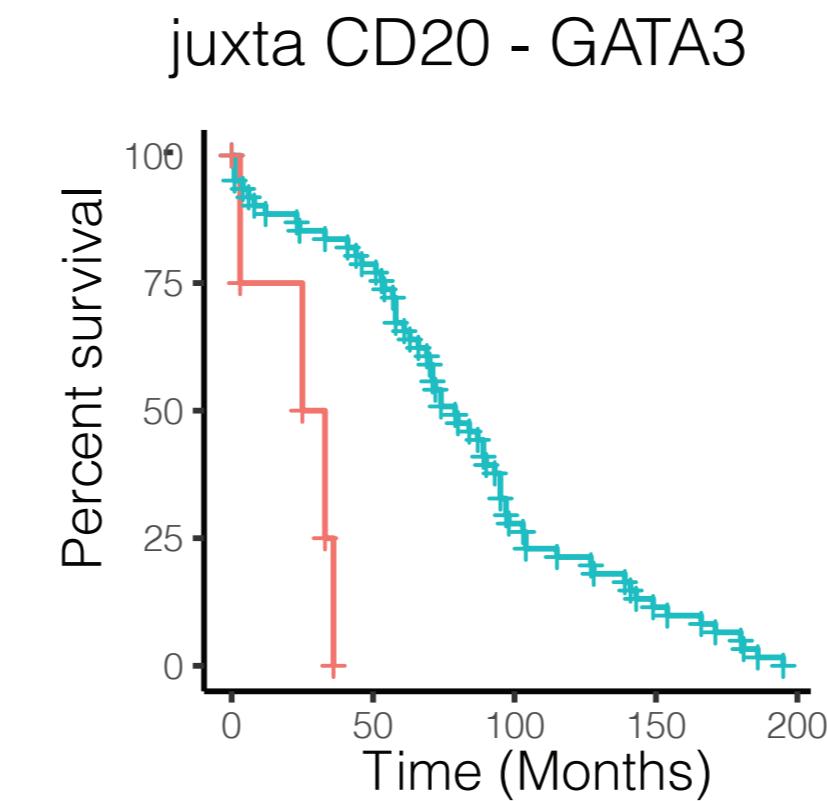
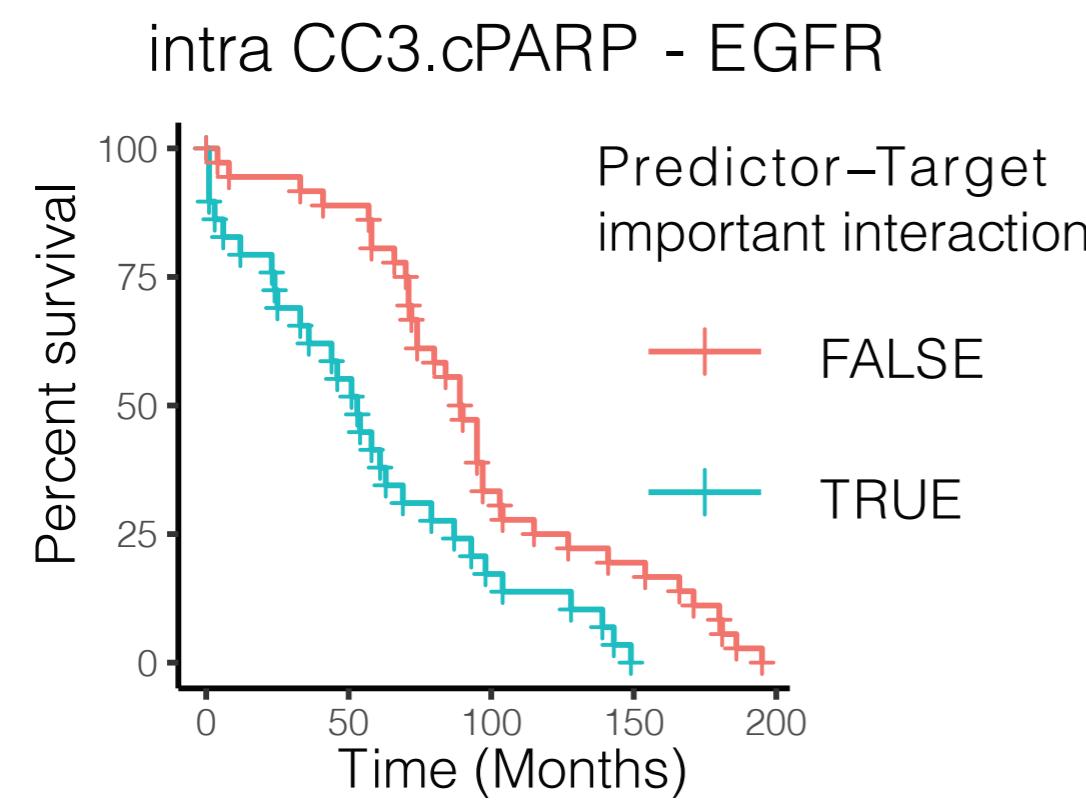




Further analyses and hypothesis generation

Large breast cancer IMC cohort. Association of results to clinical features.

Overall survivability in grade 3 triple negative patients.





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>



Available positions

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

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

Check <https://saezlab.org> for more details