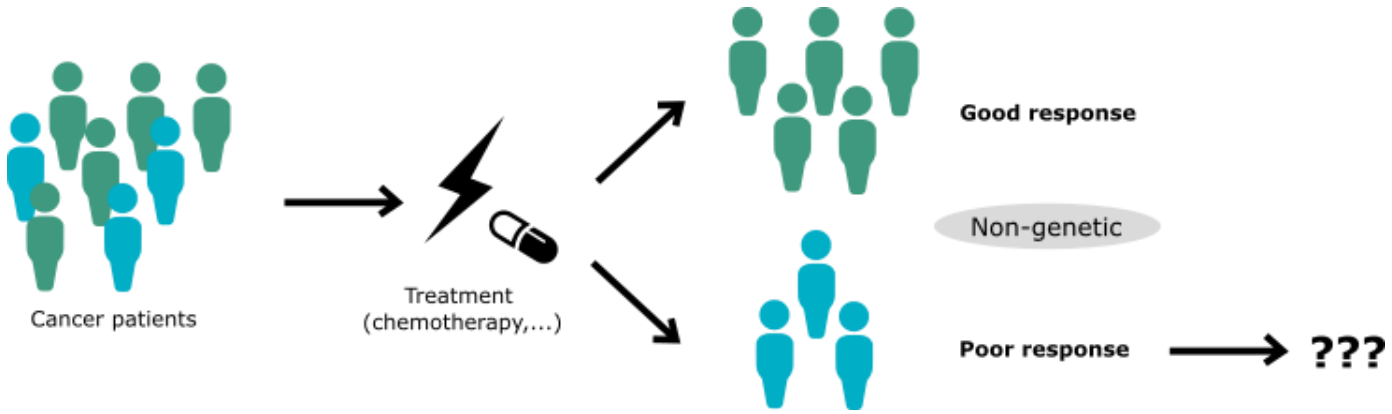
Multi-modal quantification of pathway activity with MAYA

Yuna LANDAIS - 2nd year PhD student
Dynamics of epigenetic plasticity in cancer (UMR3244)
One Biosciences

02/05/2023

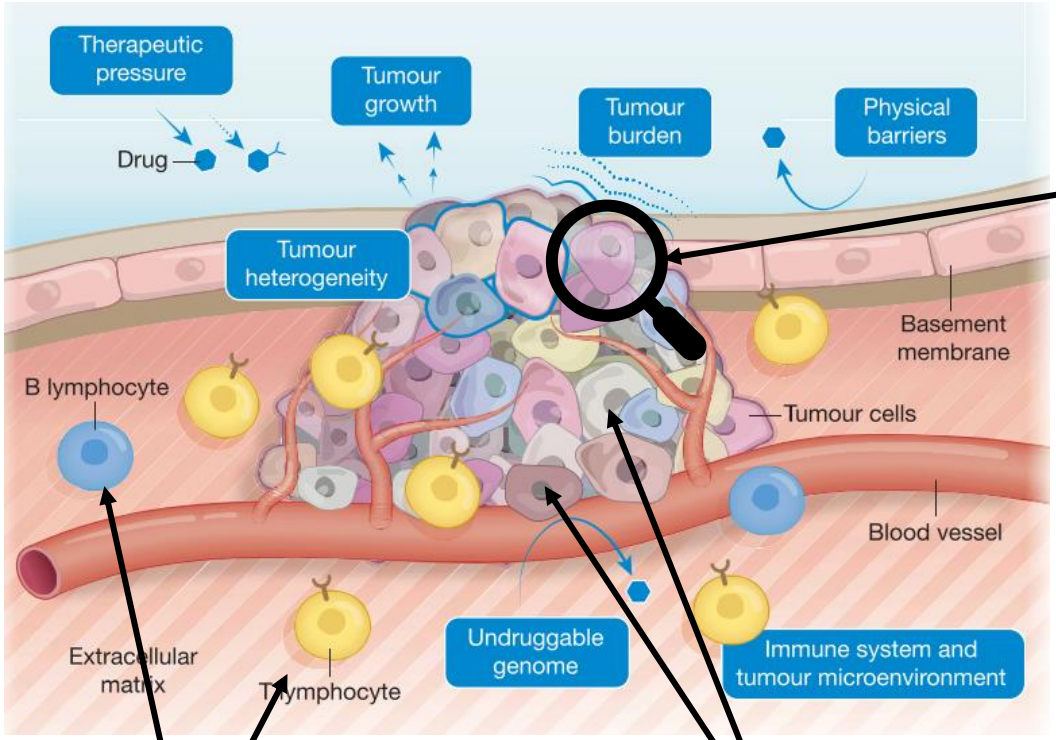
Resistance is a major challenge for cancer treatment



Can we predict patient's response to treatment?

Can we find new therapeutic approaches for non-responding patients?

Chemoresistance is a complex biological mechanism



Vasan et al., Nature, 2019

Regulated at different levels:
- Genetic
- Transcriptomic
- Epigenetic

+ How does the system change during the course of treatment?

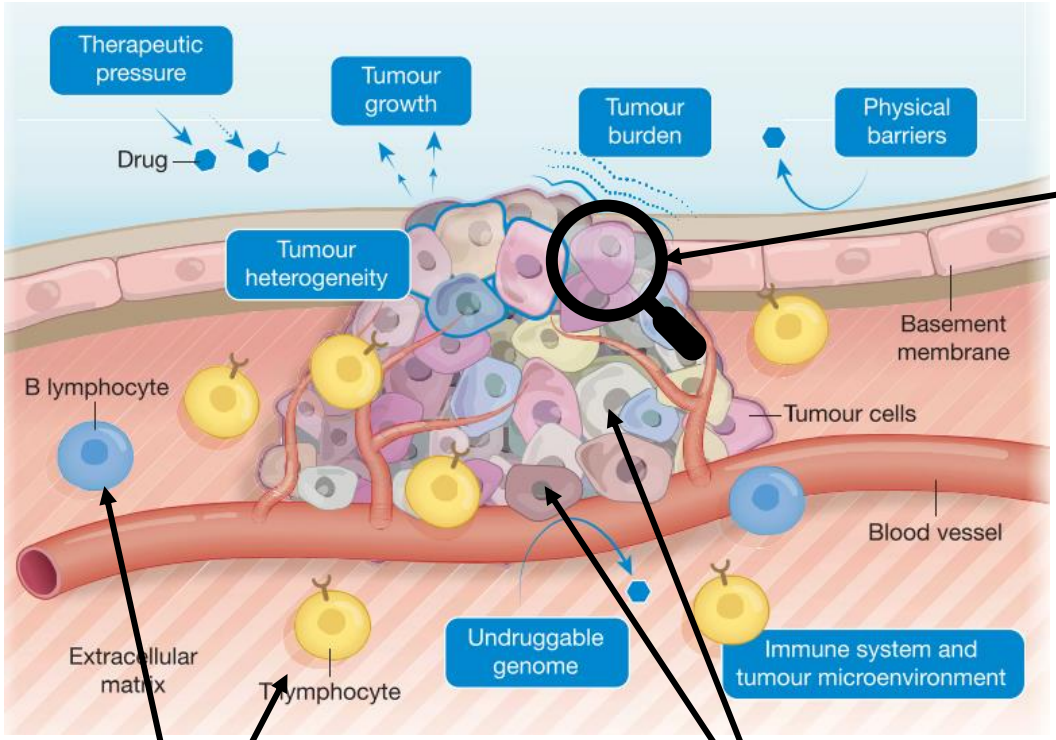
Different cell types

Inter-patient variability +
Intra-tumoral heterogeneity

Single-cell technologies are a great opportunity to precisely describe the tumor ecosystem.

BUT many variables here, **where to start**?

Chemoresistance is a complex biological mechanism



Vasan et al., Nature, 2019

Regulated at different levels:
- Genetic
- **Transcriptomic**
- Epigenetic

+ How does the system change during the course of treatment?

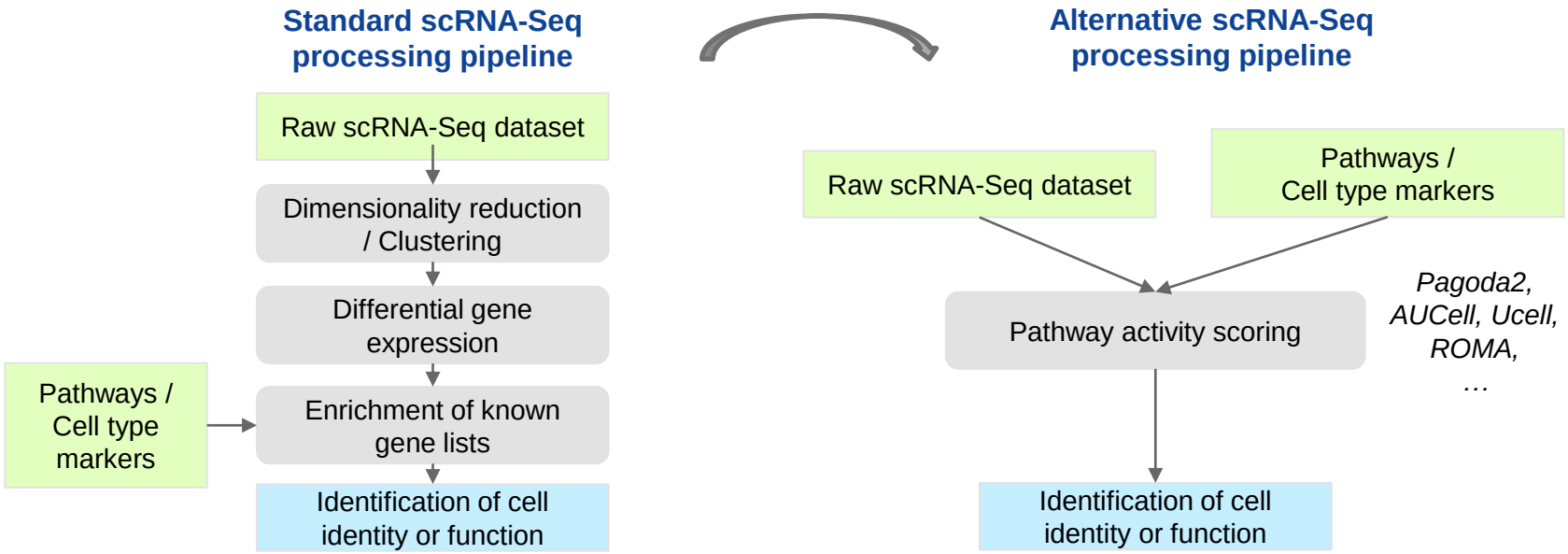
Different cell types

Inter-patient variability + Intra-tumoral heterogeneity

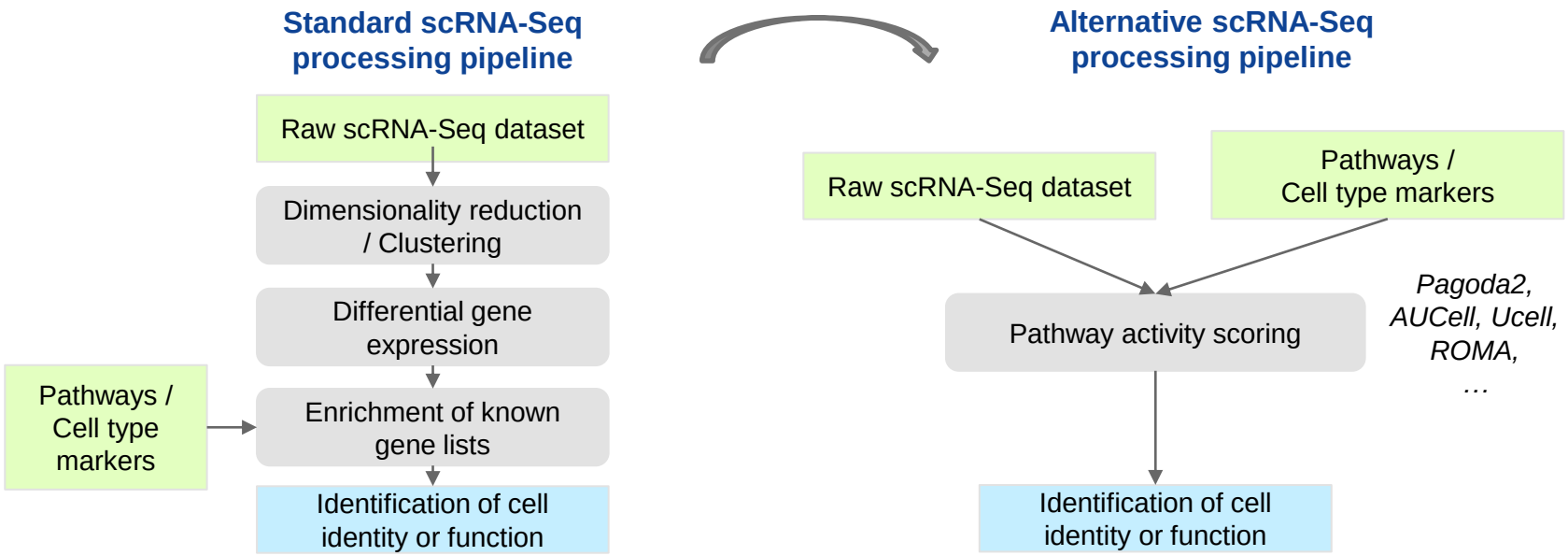
What does the tumor look like **before treatment** from **transcriptomic** perspective?

- **Who is who?**
- **Who does what?**
- **How?**

Current methods to study cell identity and function from scRNA-Seq data



Current methods to study cell identity and function from scRNA-Seq data

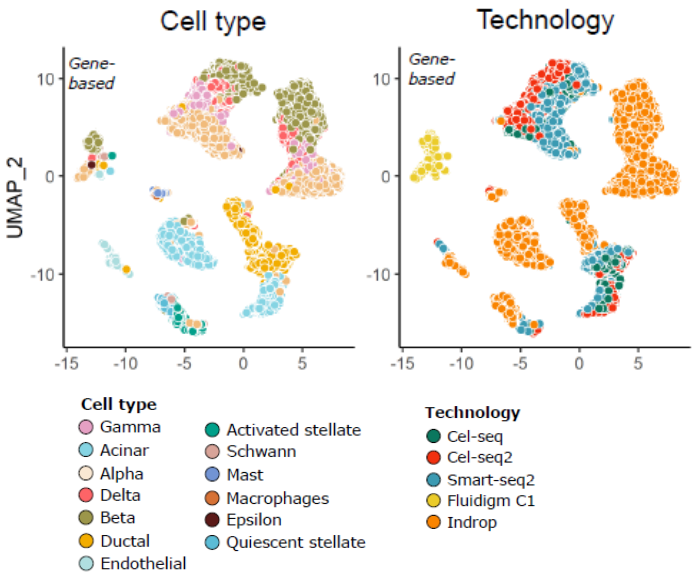


- ***Dependence to clustering parameters***
- ***Dependence to DEGs significance thresholds***

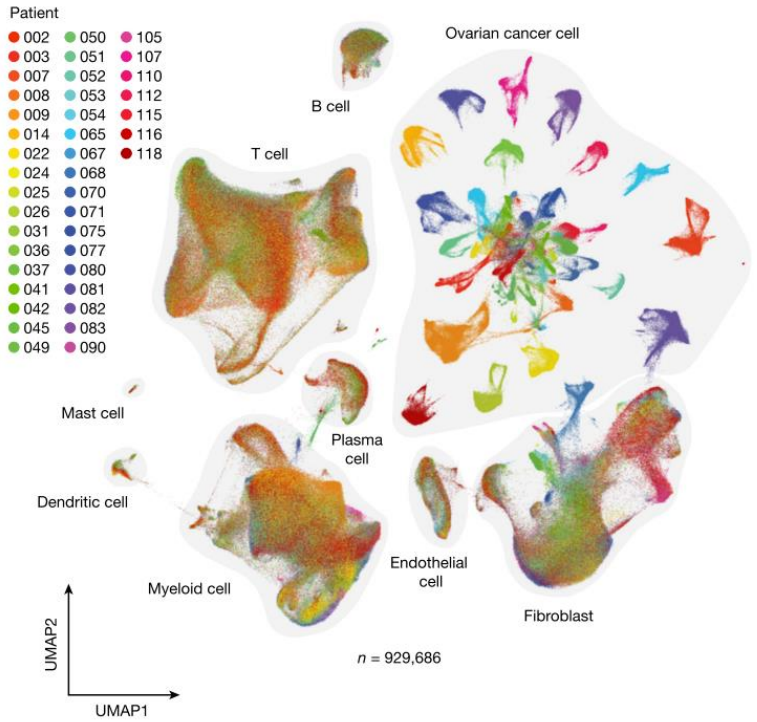
Strong hypothesis:
1 cluster = 1 cell identity = 1 cell function

Identity is not always the main driver of cell clustering

Batch effect



Patient specificity of tumor cells



Vazquez-Garcia et al., Nature, 2022

Current methods to study cell identity and function from scRNA-Seq data

Standard scRNA-Seq processing pipeline

Raw scRNA-Seq dataset

↓

Dimensionality reduction / Clustering

↓

Differential gene expression

↓

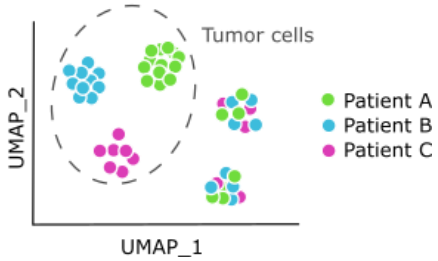
Enrichment of known gene lists ← Pathways / Cell type markers

↓

Identification of cell identity or function

- ***Dependence to clustering parameters***
- ***Dependence to DEGs significance thresholds***
- ***Inadequate for cancer datasets***

Tumor cells cluster by patients rather than by identity



Alternative scRNA-Seq processing pipeline

Raw scRNA-Seq dataset + Pathways / Cell type markers

↓

Pathway activity scoring (*Pagoda2, AUCell, Ucell, ROMA, …*)

↓

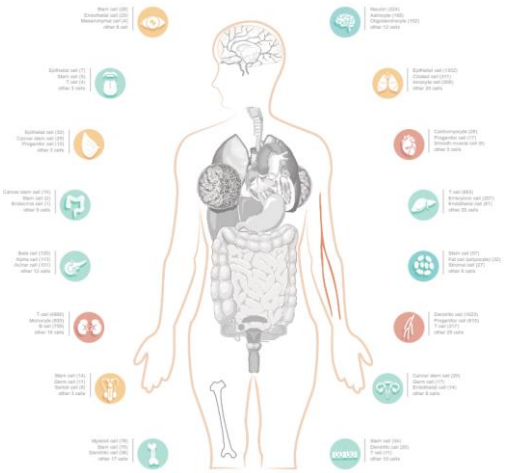
Identification of cell identity or function

Strong hypothesis:
1 pathway = 1 group of genes with coordinated expression across all cell populations

Current knowledge does not meet the granularity reached by single-cell data

Cell identity

PanglaoDB CellMarker



Capture of CellMarker Home page

Cell function

BROAD INSTITUTE MSigDB Molecular Signatures Database KEGG

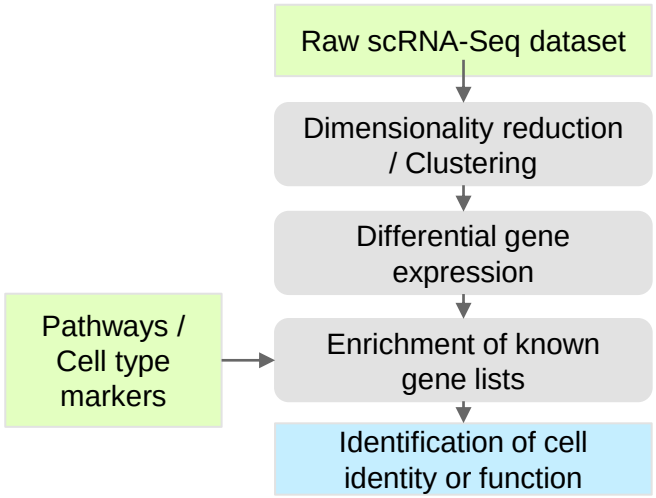
Category	Hallmarks
Cellular component	Apical_surface, apical_junction, peroxisome
Development	Adipogenesis, angiogenesis, epithelial_mesenchymal_transition, myogenesis,…
DNA damages	DNA_repair, UV_response_up, UV_response_down
Immune system	Coagulation, complement, allograft_rejection,…
Metabolism	Cholesterol_homeostasis, glycolysis…
Pathway	Apoptosis, hypoxia,…
Proliferation	G2M_checkpoint, MYC_targets,…
Signaling	Notch_signaling, KRAS_signaling_up,…

Example of "HALLMARK" pathways

Issue:
Pathway lists are often derived from bulk data
→ they do not describe cell-type-specific biological processes

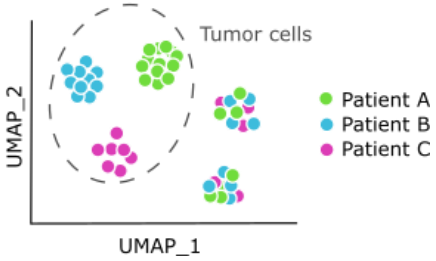
Current methods to study cell identity and function from scRNA-Seq data

Standard scRNA-Seq processing pipeline

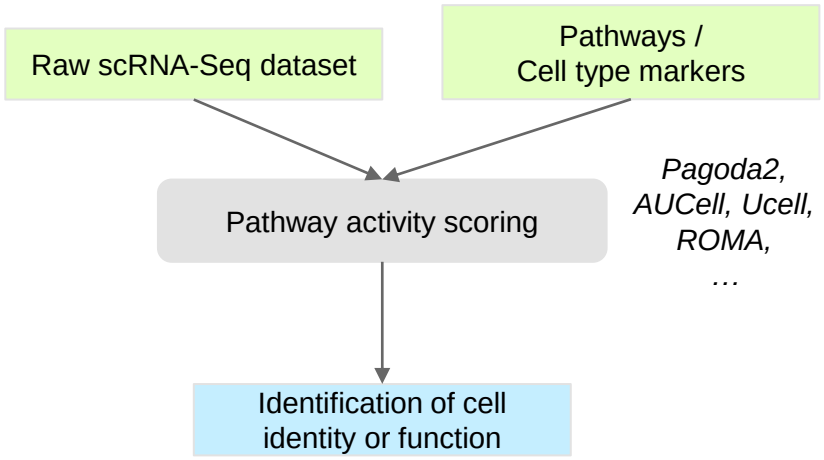


- ***Dependence to clustering parameters***
- ***Dependence to DEGs significance thresholds***
- ***Inadequate for cancer datasets***

Tumor cells cluster by patients rather than by identity

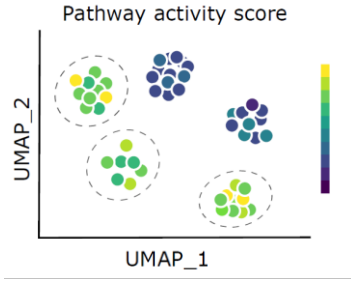


Alternative scRNA-Seq processing pipeline

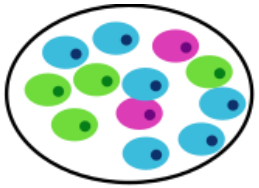


- ***Current knowledge does not meet the granularity reached by single-cell data***
- ***Information can be incomplete for drug discovery purposes***

Different populations could activate the same pathway through different genes and we would not know



Precise description of pathway activation in specific populations is essential for drug discovery



Tumor sample

Routine

Cell type identification

= Macrophage

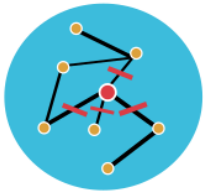
= Tumor cell

= Fibroblast

Are there differences in cell type composition and/or proportions between samples?

Functional analysis

Gene regulatory networks



gene

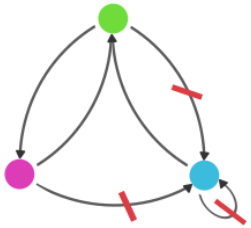
potential target

Cancer cell

What biological functions are activated in the cells?

How are they regulated (both internal and external signals)?

Cell-cell communication



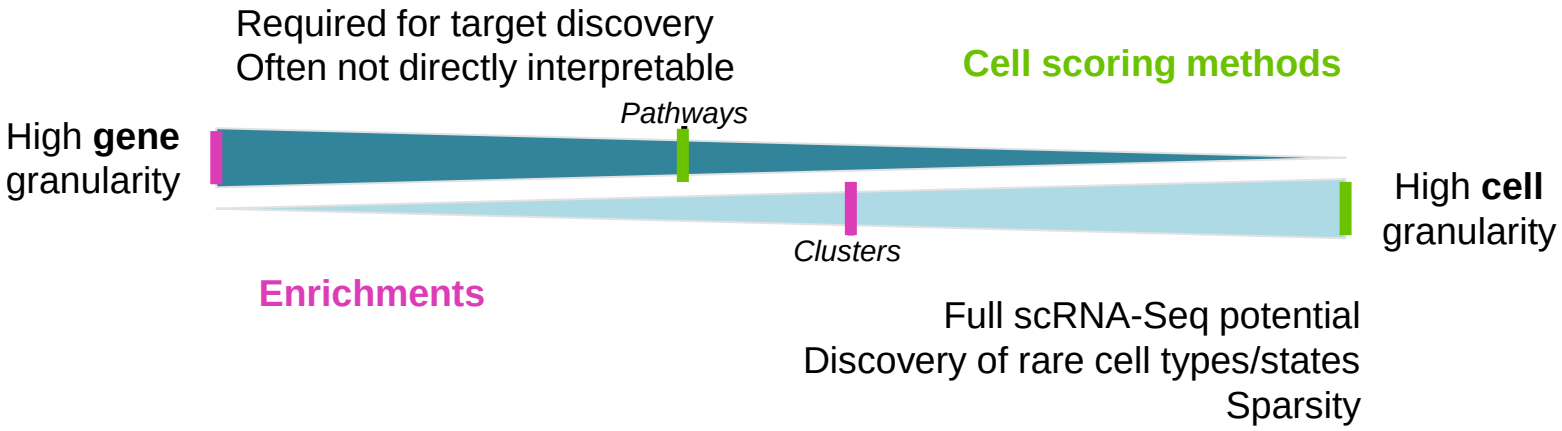
Cancer cell

What **genes** are key in the regulation mechanisms?

Target inherent properties

Cut off supplies

Can we find a compromise?

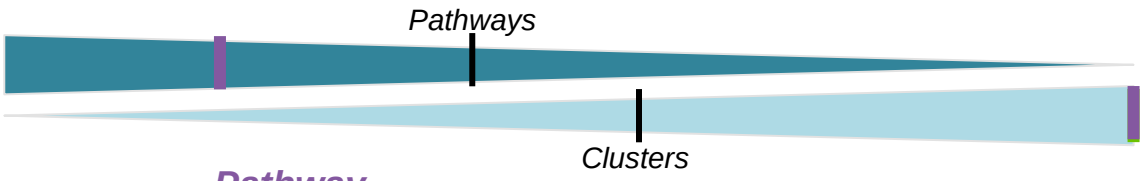


Difficult-to-reach balance between using the full potential of the **cell level granularity** of the data and having **precise information about pathway activation**

Can we find a compromise?

Required for target discovery
Often not directly interpretable

High **gene** granularity

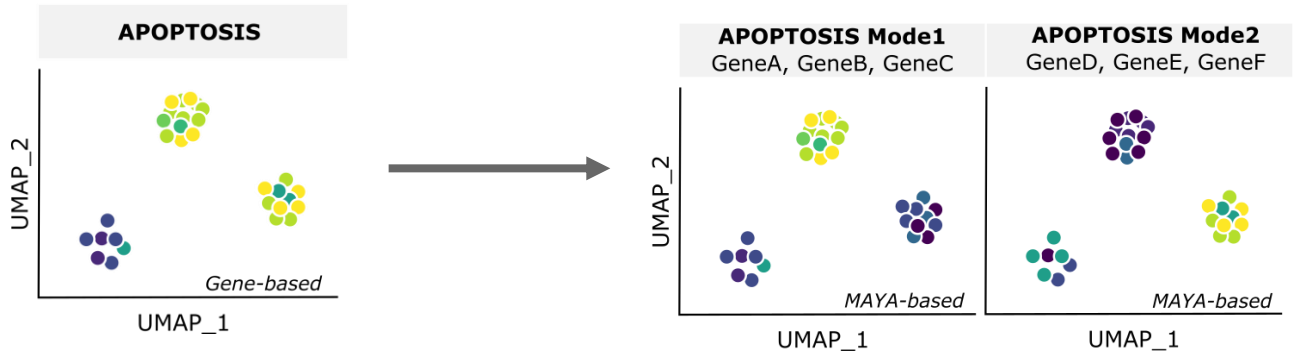


High **cell** granularity

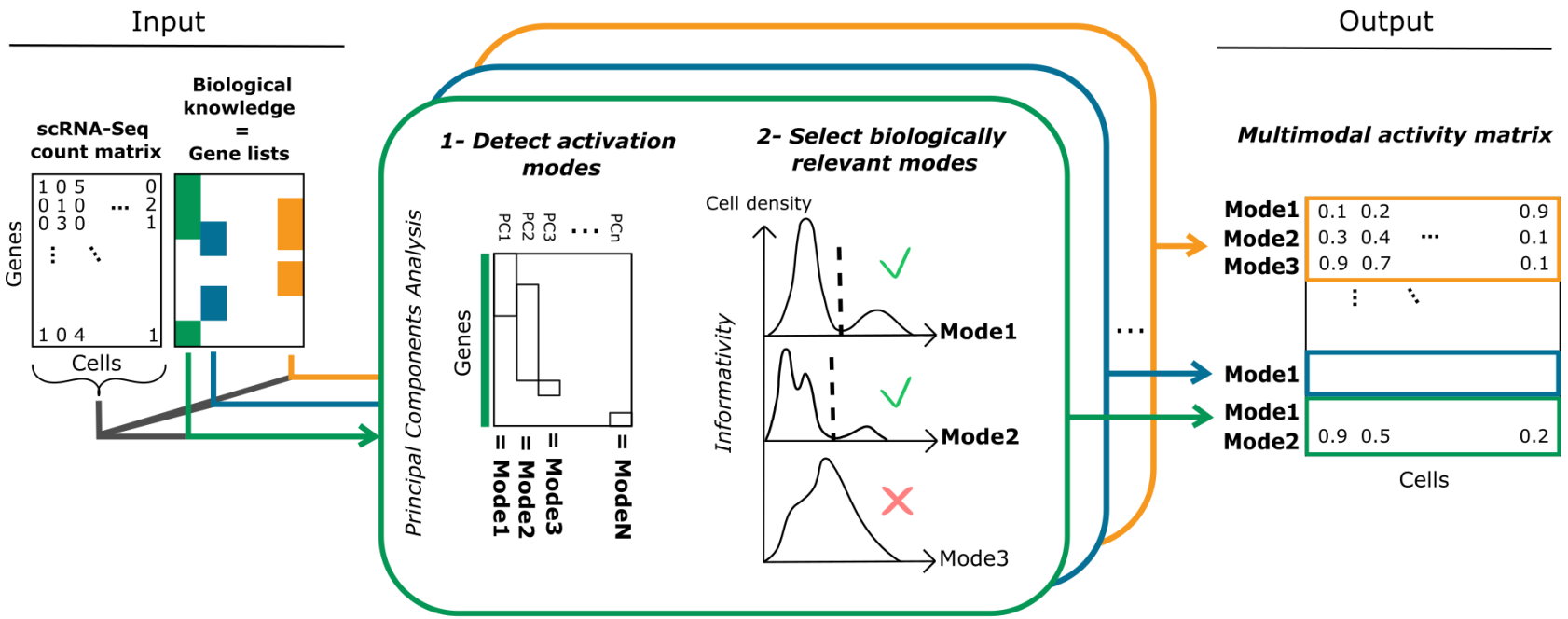
Pathway modes of activation

Full scRNA-Seq potential
Discovery of rare cell types/states
Sparsity

Difficult-to-reach balance between using the full potential of the **cell level granularity** of the data and having **precise information about pathway activation**



MAYA: Multimodes of pathwAY Activation



MAYA detects **subgroups of genes within reference pathways**, each characteristic of a cell population and how it activates a pathway.

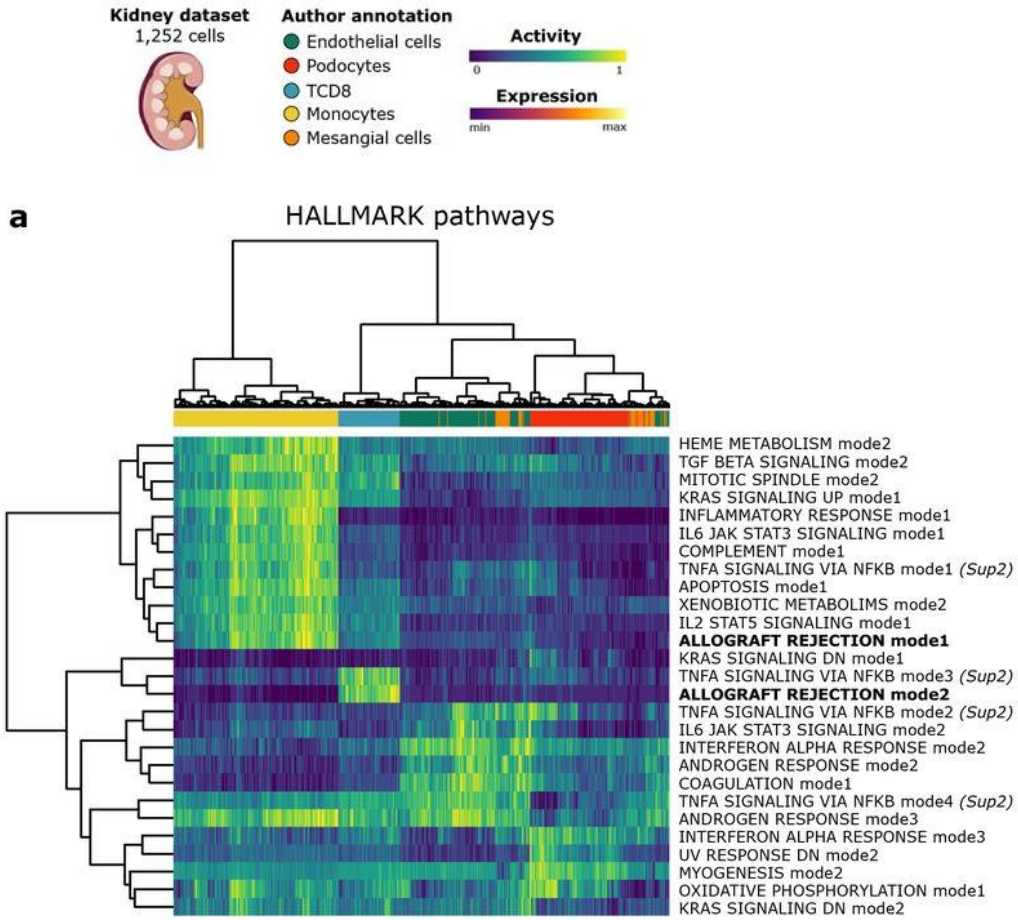
MAYA's output can be used to:

- Identify cell populations sharing similar modes of pathway activation → they can be different from the ones found with standard gene-based pipeline
- Compute a **new visualization** of cells in 2D space

Landais, Y., Vallot, C. Multi-modal quantification of pathway activity with MAYA. *Nat Commun* **14**, 1668 (2023).

ONE BIOSCIENCES

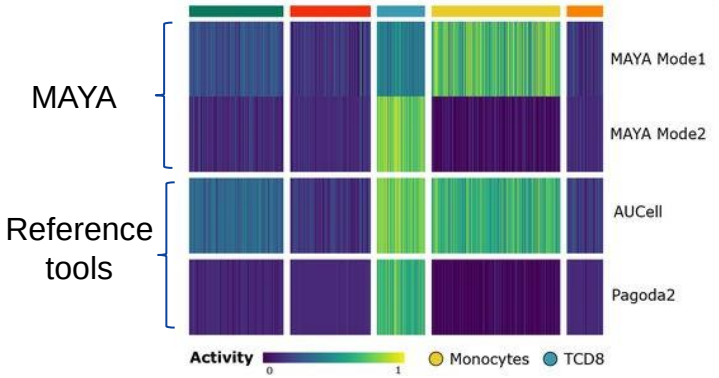
MAYA detects biologically relevant **multimodal** pathway activity in kidney



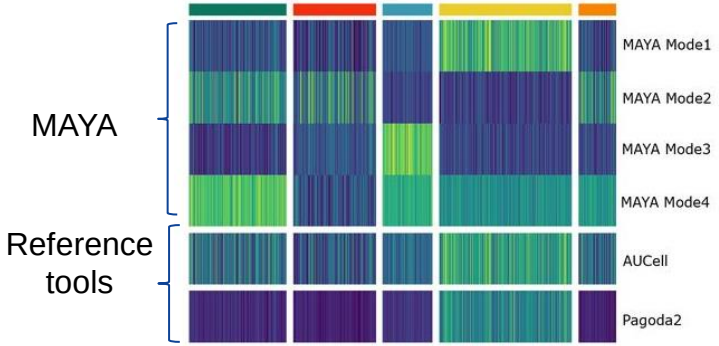
Heatmap representing MAYA activity matrix

Example of two multimodal pathways

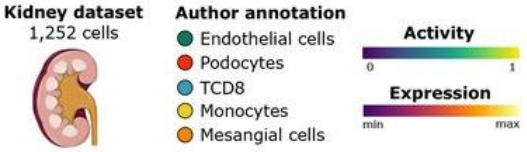
ALLOGRAFT REJECTION



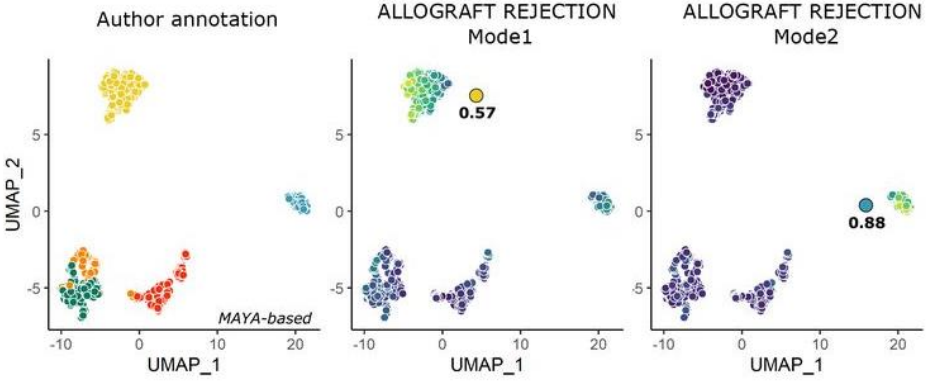
TNFA SIGNALING VIA NFKB

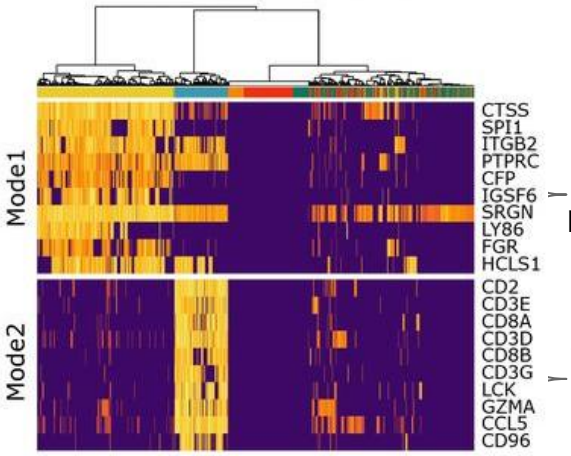


MAYA detects cell-type specific modes



ALLOGRAFT REJECTION

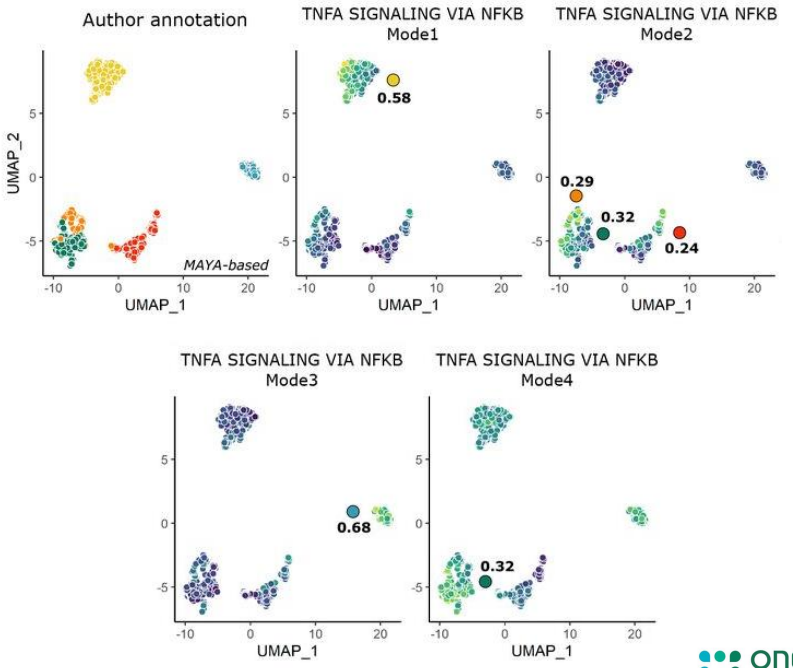


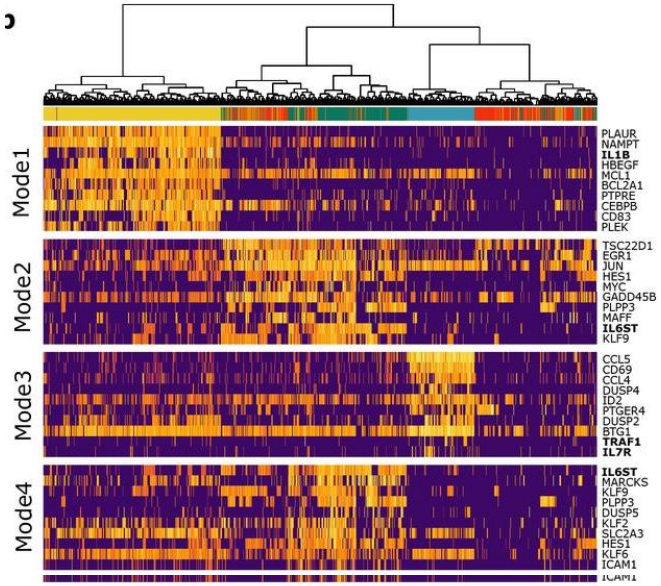


Antigen presentation related

Coding for T cell surface proteins

TNFA SIGNALING VIA NFKB





IL1B = lymphocyte activated factor

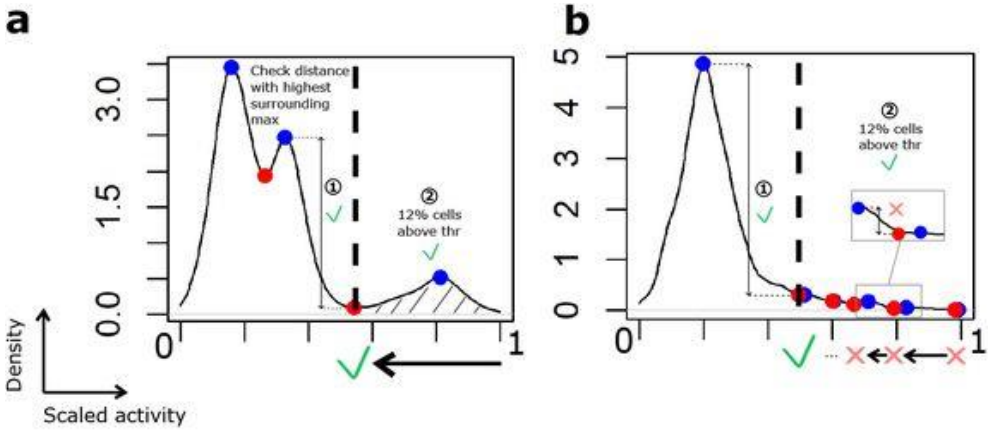
IL6ST = signal transducer activated in endothelial cells

IL7R = T cell differentiation

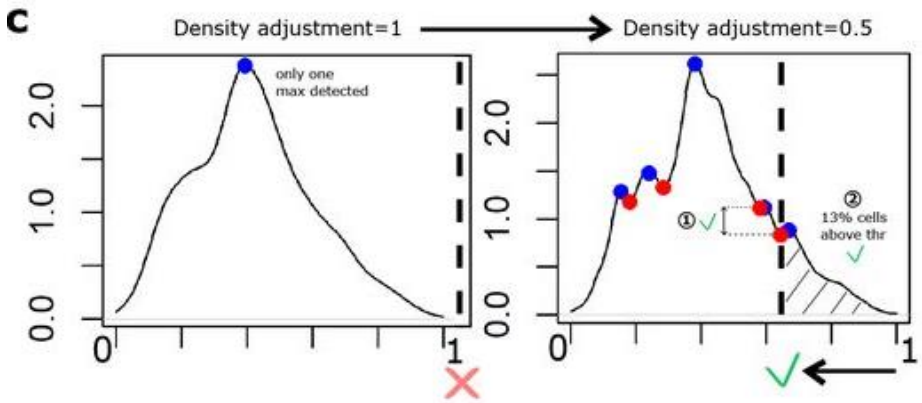
ONE BIOSCIENCES

Informativity to select modes: bimodal detection

Identifying a density minima separating two cell populations with different activity levels

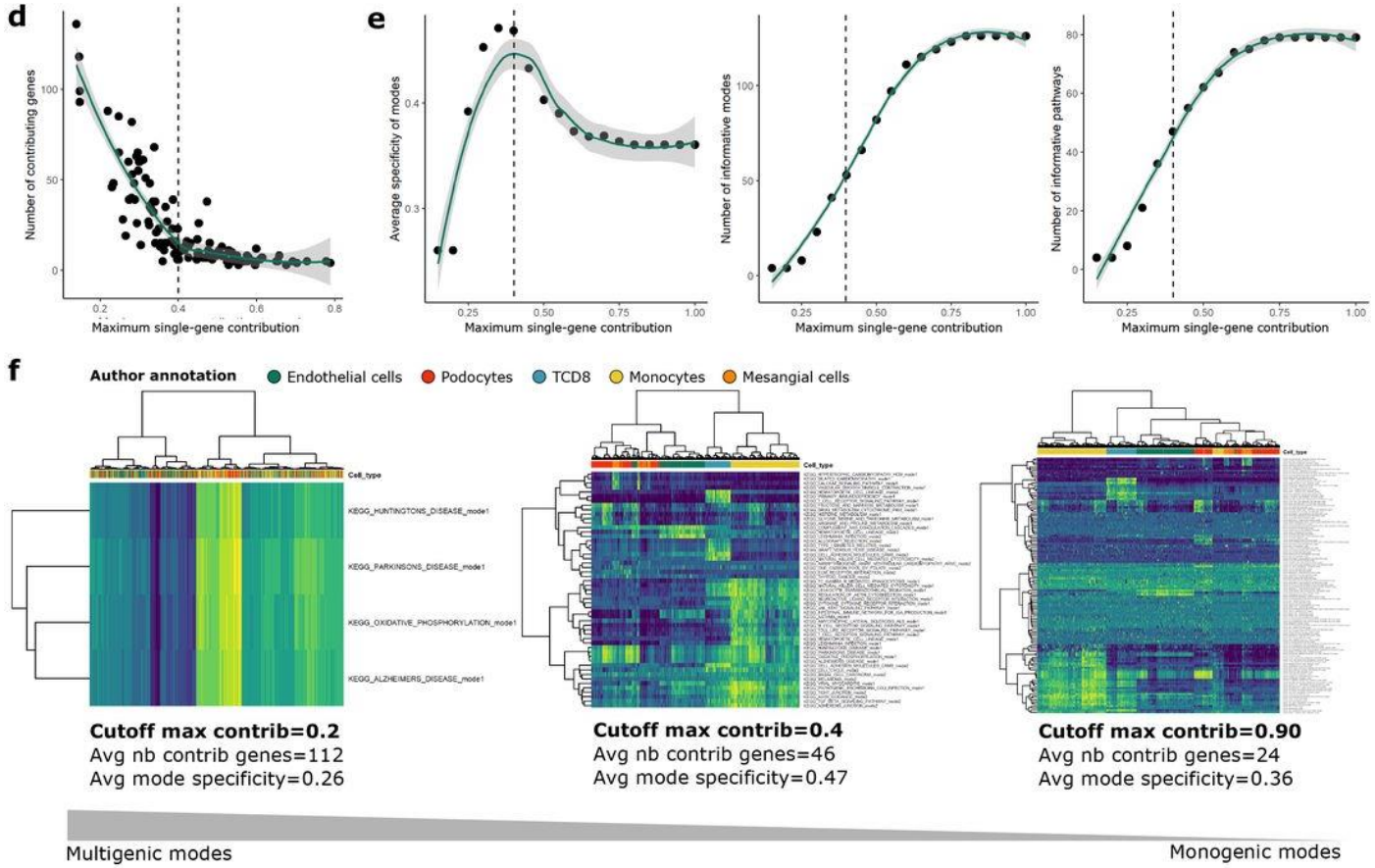


Testing a more fitted density curve to ensure we don't miss more subtle variations in activity



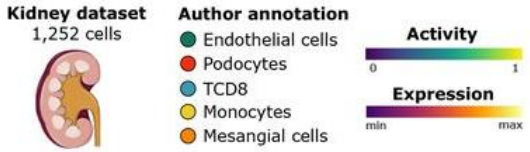
The minimal percentage of cells expected to activate a pathway is a parameter of the function and can be tuned by the user based on prior knowledge or arbitrarily.

Informativity to select modes: maximum gene contribution as an indicator for monogenic modes



Users can also tune the cutoff on maximum single-gene contribution, depending on their tolerance to monogenic modes

Testing MAYA robustness to noise



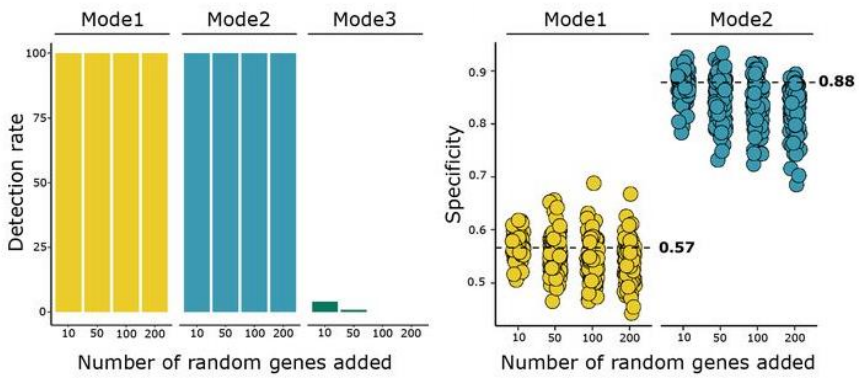
Allograft rejection pathway (200 genes) → 2 cell-type specific modes

- + 10 genes → ?
- + 50 genes → ?
- + 100 genes → ?
- + 200 genes → ?

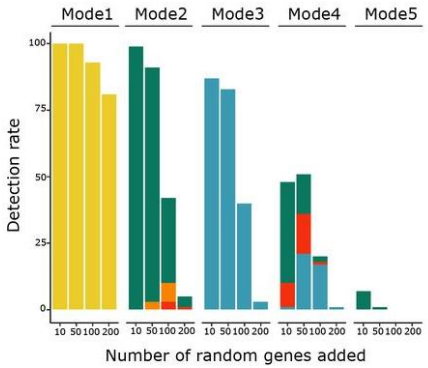
TNFA signaling pathway (200 genes) → 4 cell-type specific modes

- + 10 genes → ?
- + 50 genes → ?
- + 100 genes → ?
- + 200 genes → ?

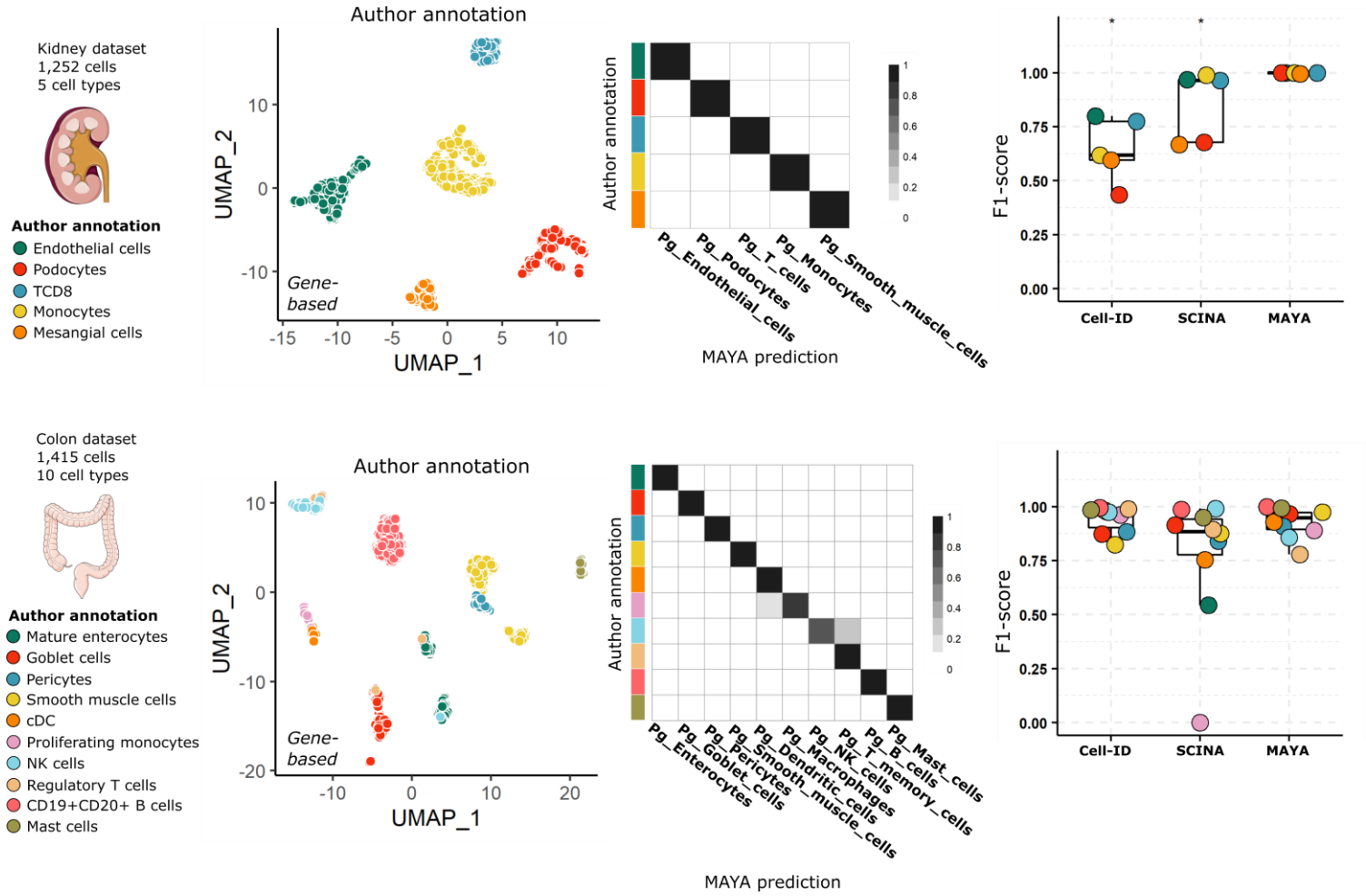
ALLOGRAFT REJECTION



TNFA SIGNALING VIA NFKB

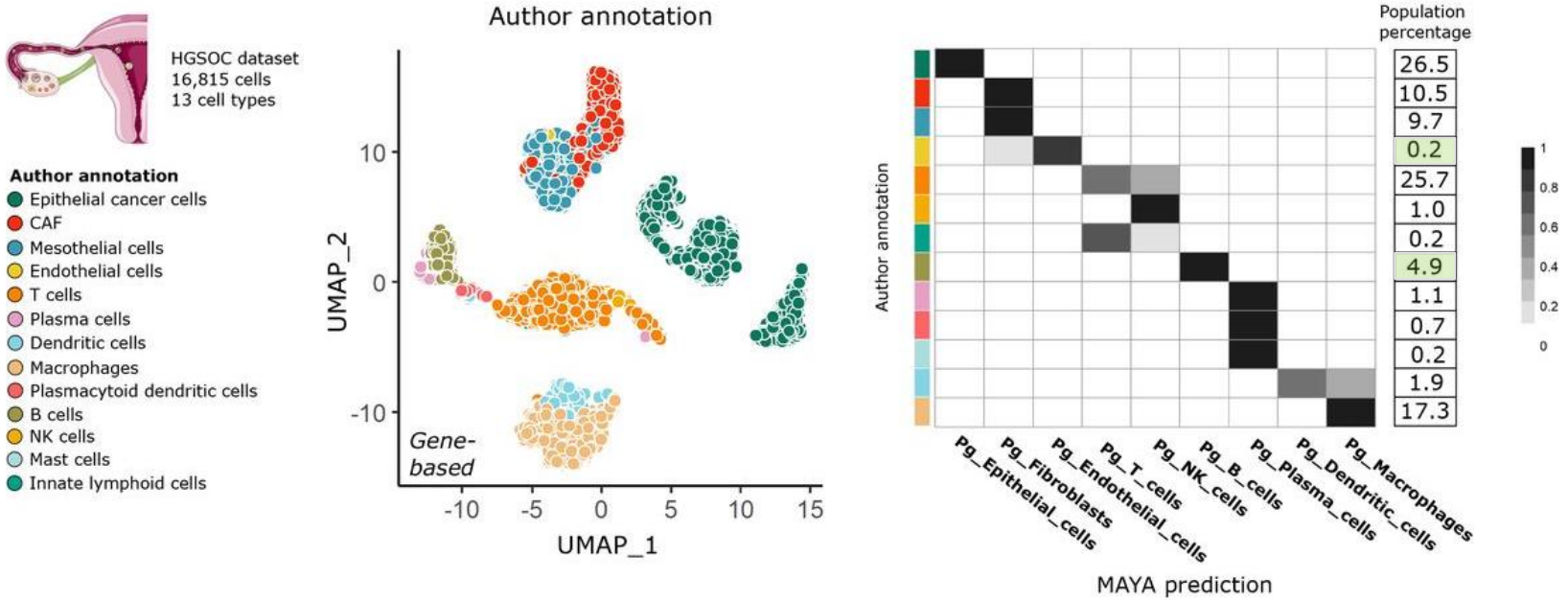


MAYA annotates cell types



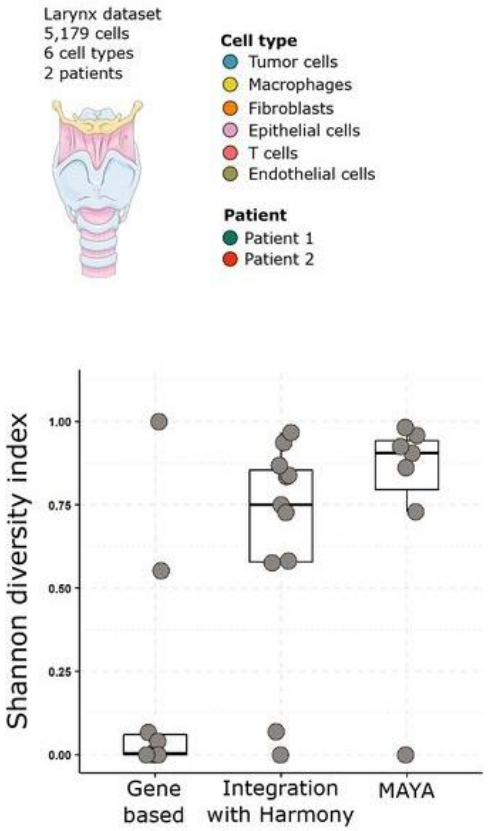
MAYA presents rates of precision and recall comparable to other specialized cell-type annotation methods.

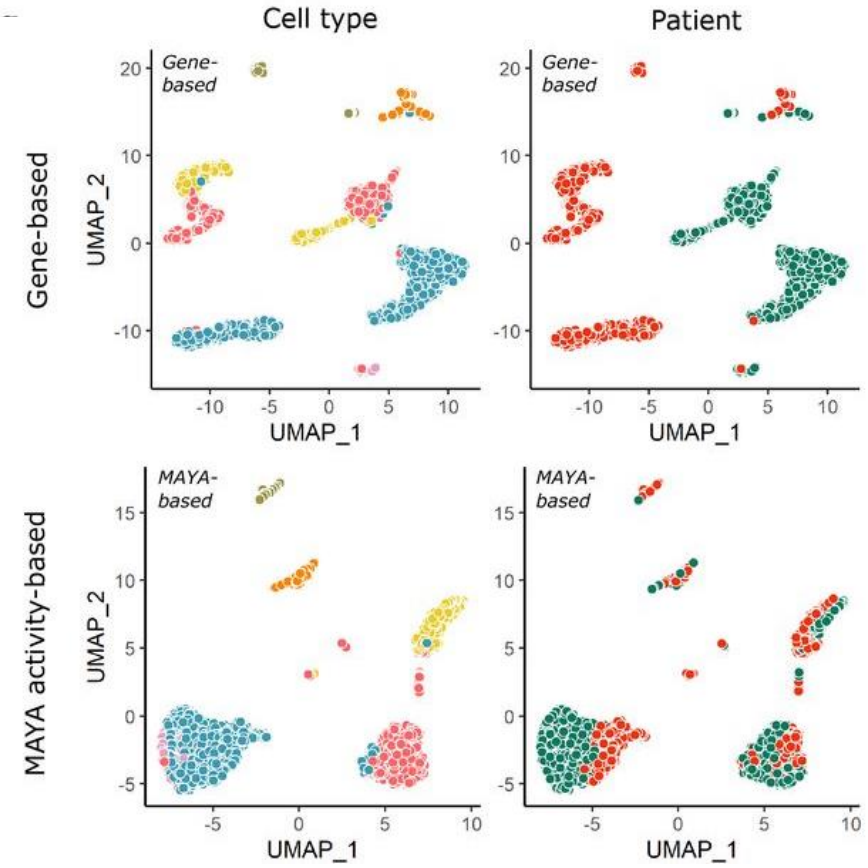
MAYA detects and annotates rare cell populations



- MAYA can also accurately annotate very rare populations in a dataset.
- However, a limitation is the overlap between cell-type markers list that sometimes make it difficult to distinguish between two cell types (NK/T cells, DC/macrophages, endothelial/fibroblasts)

MAYA allows batch effect free cell-type annotation

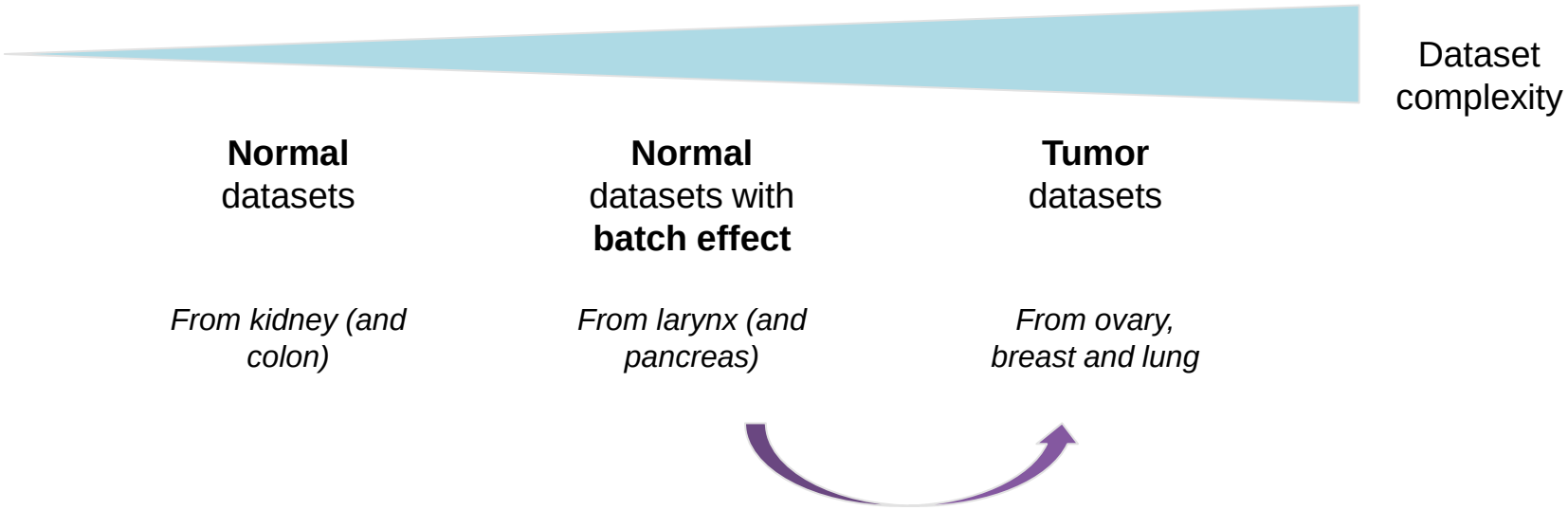




MAYA UMAP embeddings can be used for batch-effect-free visualization of datasets

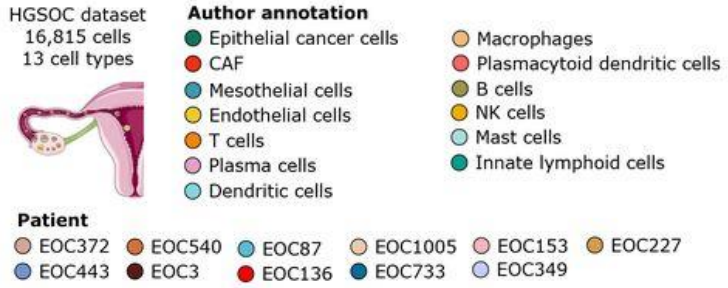
Summary

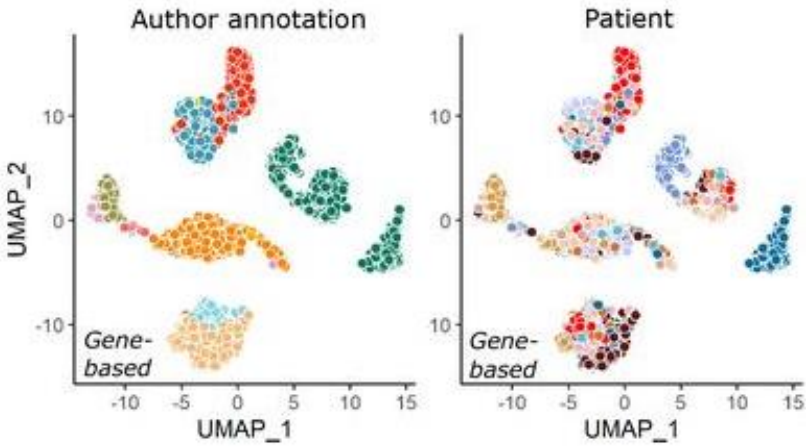
- All-in-one tool for cell identity and pathway study and easy to use (few parameters, with default values)
- Robust to noise: very useful as manual curation of gene lists is time-consuming
- Scalable: annotation of 125,000 cells in ~15 minutes
- Interpretable way to correct batch effect for visualization

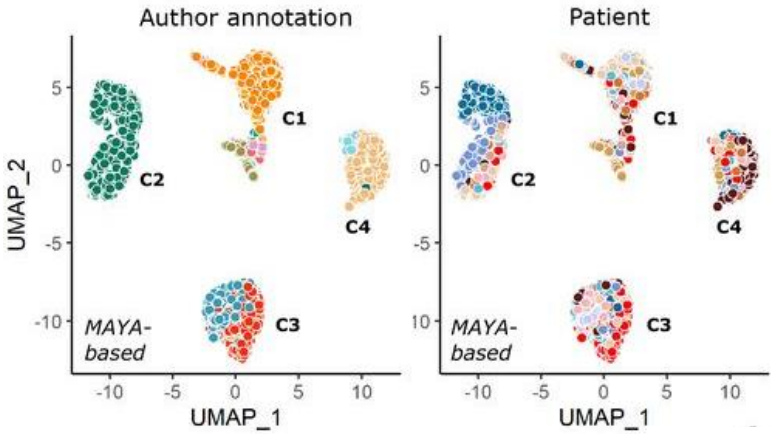


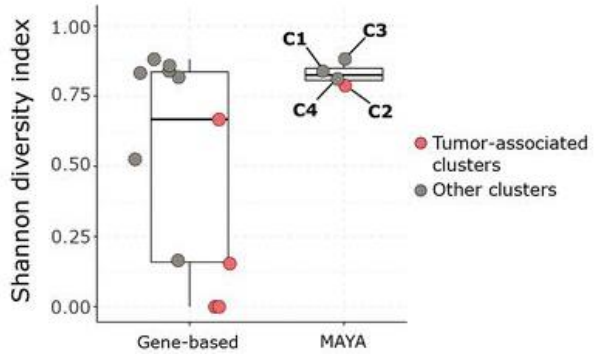
MAYA detects common modes of pathway activation across patients

With HALLMARK pathways



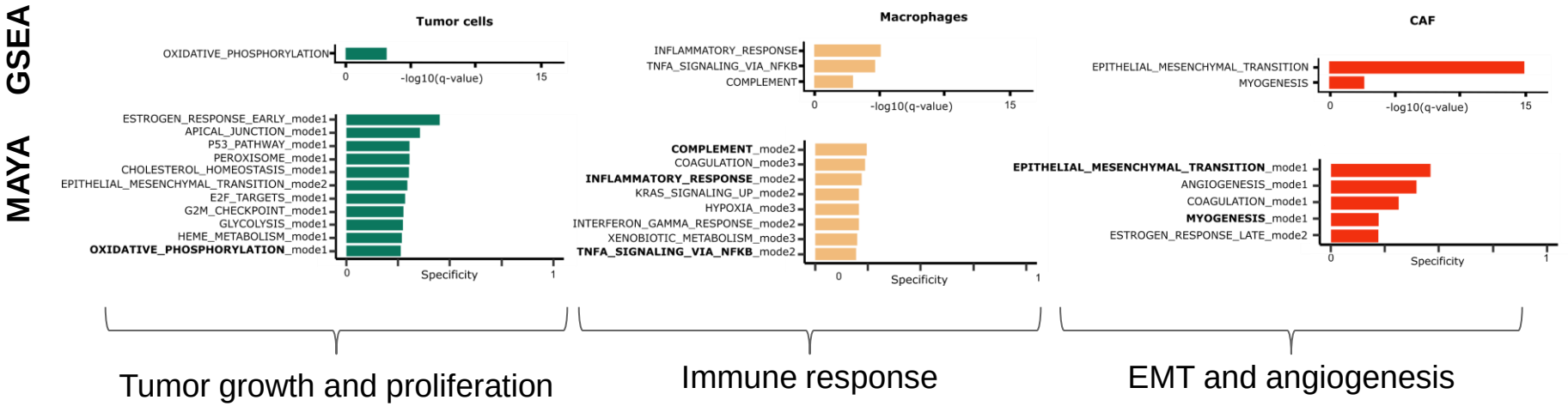






Tumor cells tend to cluster by patient rather than by identity with standard analyses
→ MAYA identifies shared modes of pathway activation across patient

MAYA detects common modes of pathway activation across patients that are not detected by GSEA



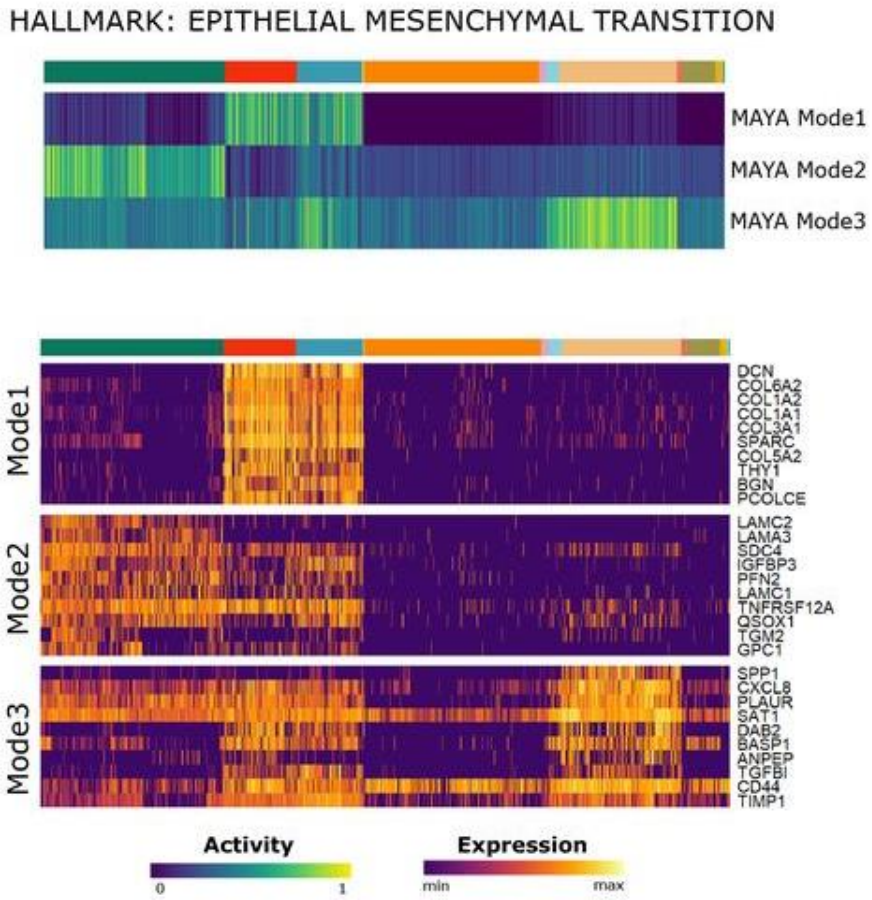
Tumor growth and proliferation

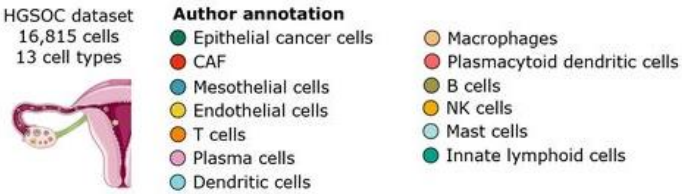
Immune response

EMT and angiogenesis

Identification of relevant top-specific modes of pathway activation for both tumor cells and the microenvironment
Identification of more functions than with classical GSEA analysis

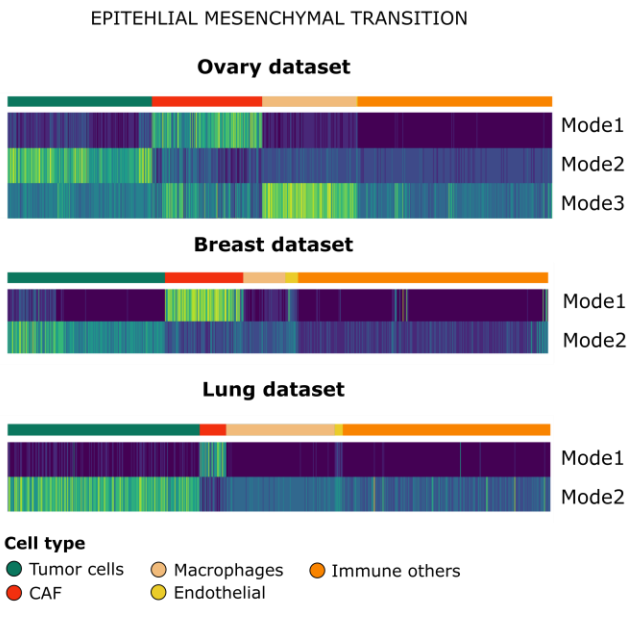
MAYA detects three cell-type specific modes of EMT activation

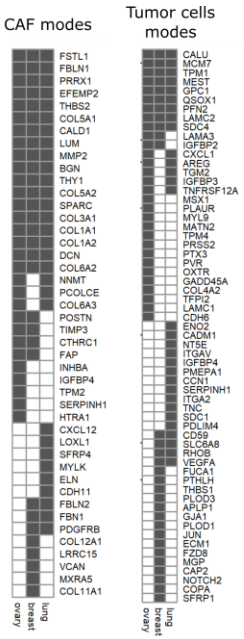


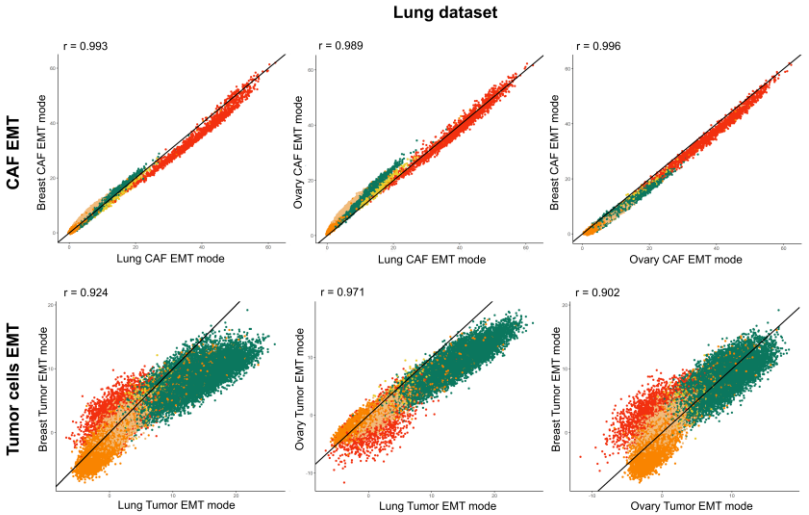


- Identification of a set of genes that characterizes EMT occurring in epithelial cells driven by genes coding for subunit of laminin 332 (component of epithelial basement membrane, that promotes tumor motility).
- EMT in fibroblasts driven mainly by ECM genes
- Macrophages shown to be involved in EMT in several cancer types.

MAYA EMT modes are reproducible across datasets

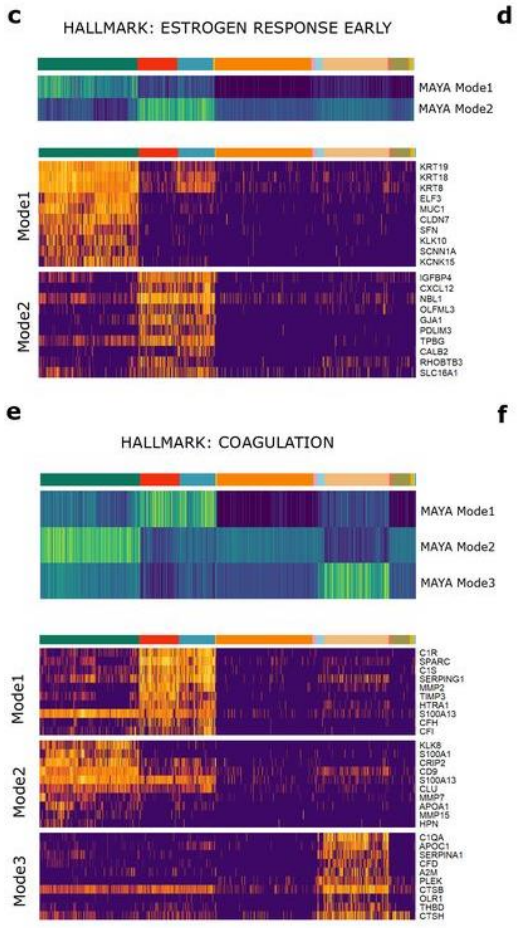


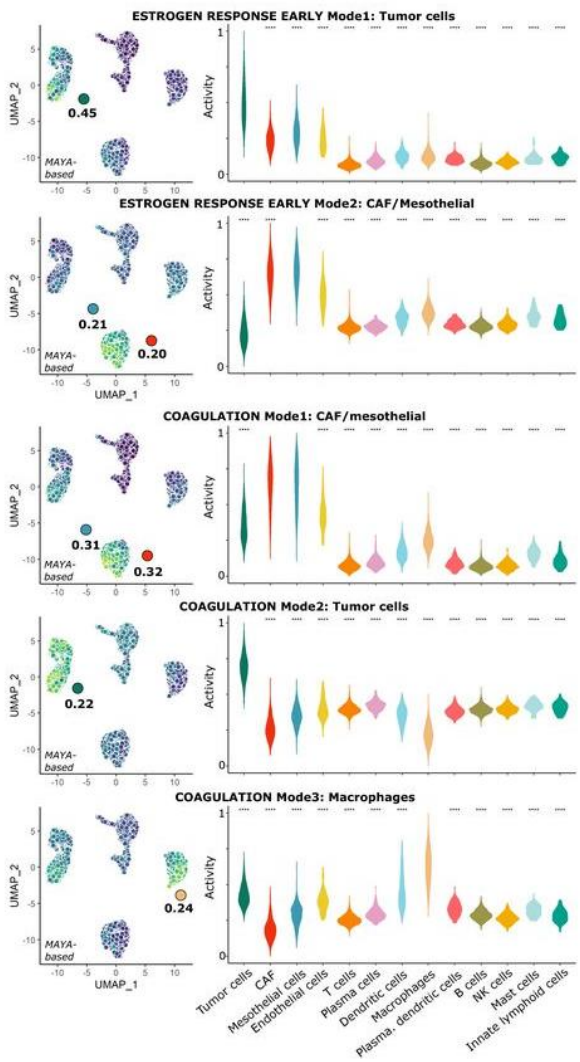


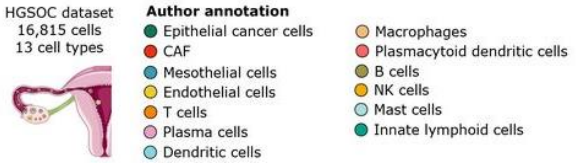


CAF and tumor cell specific modes of EMT were detected in two additional cancer datasets and display high correlation when evaluated in the same cells.

Other examples of multimodal pathway activation

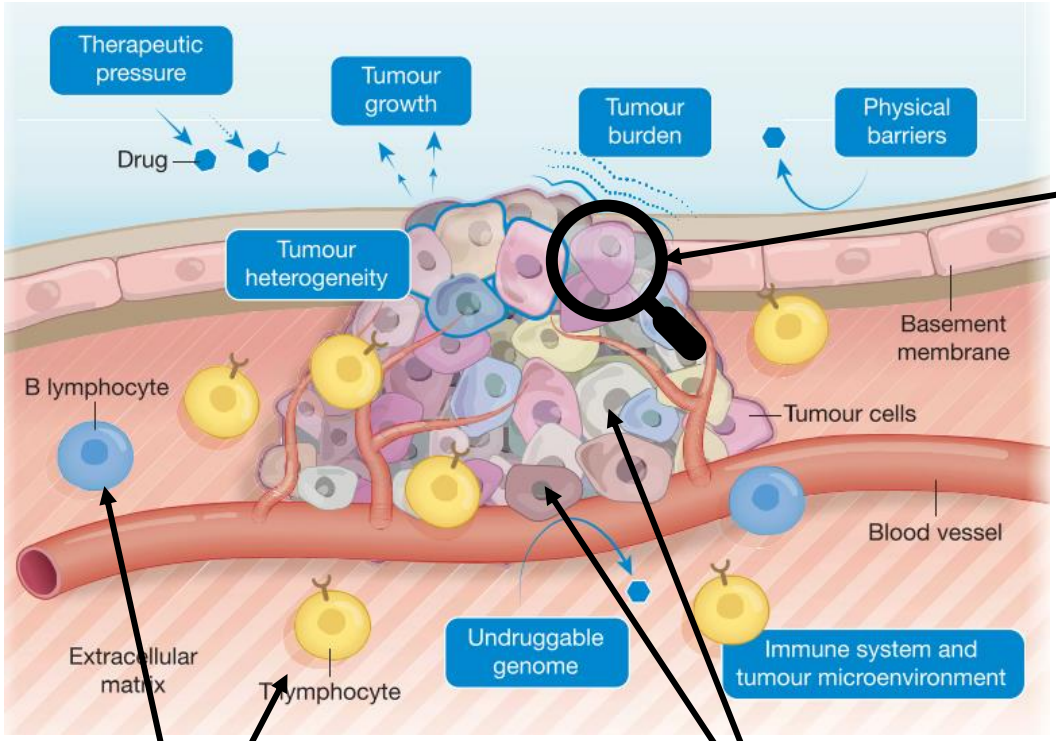






- CAFs can use ER-mediated pathways to promote tumor cell proliferation
- MAYA helps to untangle the respective contributions of cancer cells cells and its microenvironment to the hemostatic imbalance observed in cancer

Can MAYA help us address other challenges related to the study of chemoresistance?



Vasan et al., Nature, 2019

Regulated at different levels:
- Genetic
- Transcriptomic
- Epigenetic

+ How does the system change during the course of treatment?

How far does MAYA precision go?
Can we detect subtle changes in pathway activation:
- within a tumor?
- Between paired samples pre/post-treatment?

Different cell types

Inter-patient variability + **Intra-tumoral heterogeneity**

Acknowledgments

Vallot Lab

Léa Baudre
Juliette Bertorello
Adeline Durand
Grégoire Jouault
Marthe Laisne
Justine Marsolier
Mélissa Saichi
Mathias Schwartz
Céline Vallot

Pacôme Prompsy
Camille Landragin

One Biosciences

Virginie Bouche
Hugo Cornu
Marianne Delville
Marine Gaillard
Cécile Hui-bon-hoa
Marceau Quatredeniers
Magali Richard
Baptiste Simon
Céline Vallot

Paul Cadorin

Thank you for your attention !