

# M2 Bio-Informatic Internship

03/10/2025 Presentation

Jordan Dutel

Université Claude Bernard Lyon 1  
Centre de Recherche en Cancérologie de Lyon (CRCL)

Team : Dr Pierre Saintigny

Tutor : Dr Pierre Martinez

March 10, 2025



## 1 Introduction

## 2 Workflow

## 3 Results

## 4 Outlook









# Research questions

- ① Determine specific markers of different tumor cell phenotypes (for diagnosis)
- ② Define the genes and pathways involved in trans-differentiation
- ③ Analyze the genomic and microenvironmental differences between compartments



# Pipeline

## Steps

- ① Load data
- ② Merging objects
- ③ Normalisation
- ④ PCA
- ⑤ Harmony (Batch effect correction)
- ⑥ UMAP
- ⑦ Find cell-clusters
- ⑧ Find cluster-specific markers

## Tools

Step	R Packages/Functions
Read Data	Seurat
Normalization	NormalizeData() + ScaleData() (No SCTransform !!!)
Batch effect correction	RunHarmony()

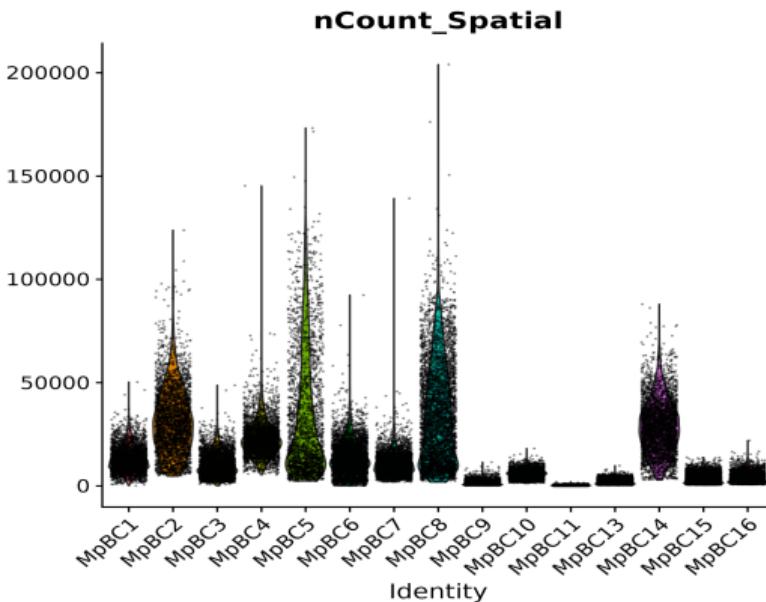
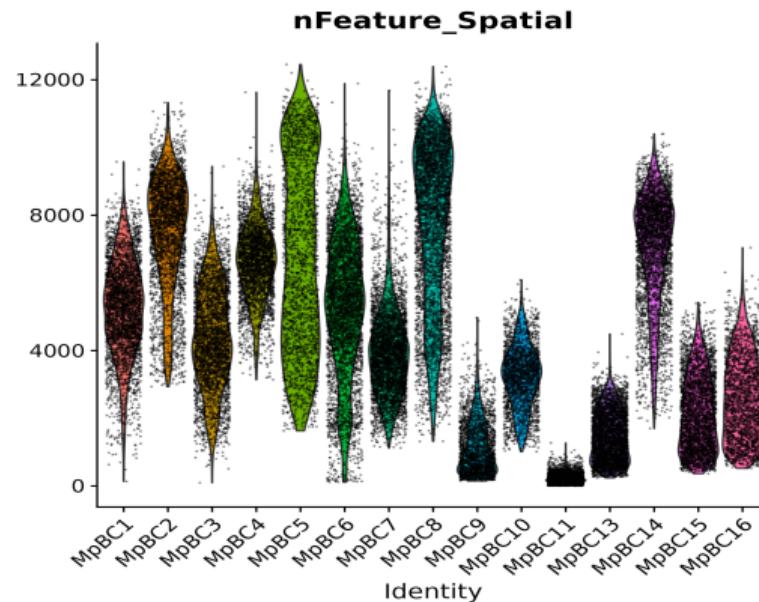
## 1 Introduction

## 2 Workflow

## 3 Results

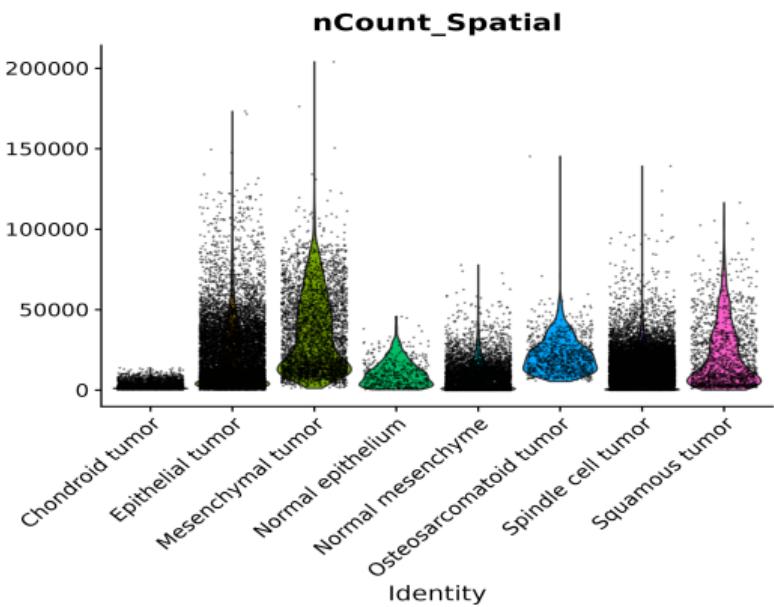
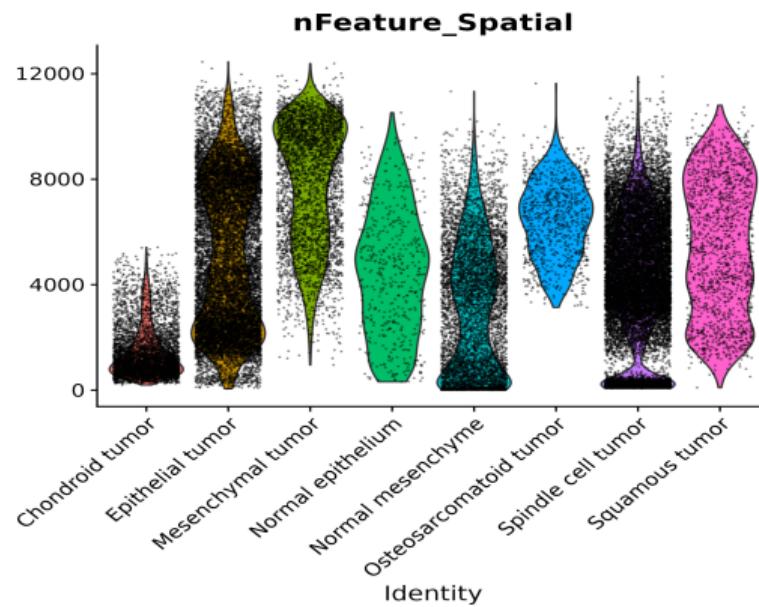
## 4 Outlook

# Quality Control



**Some heterogeneity between samples**

# Quality Control



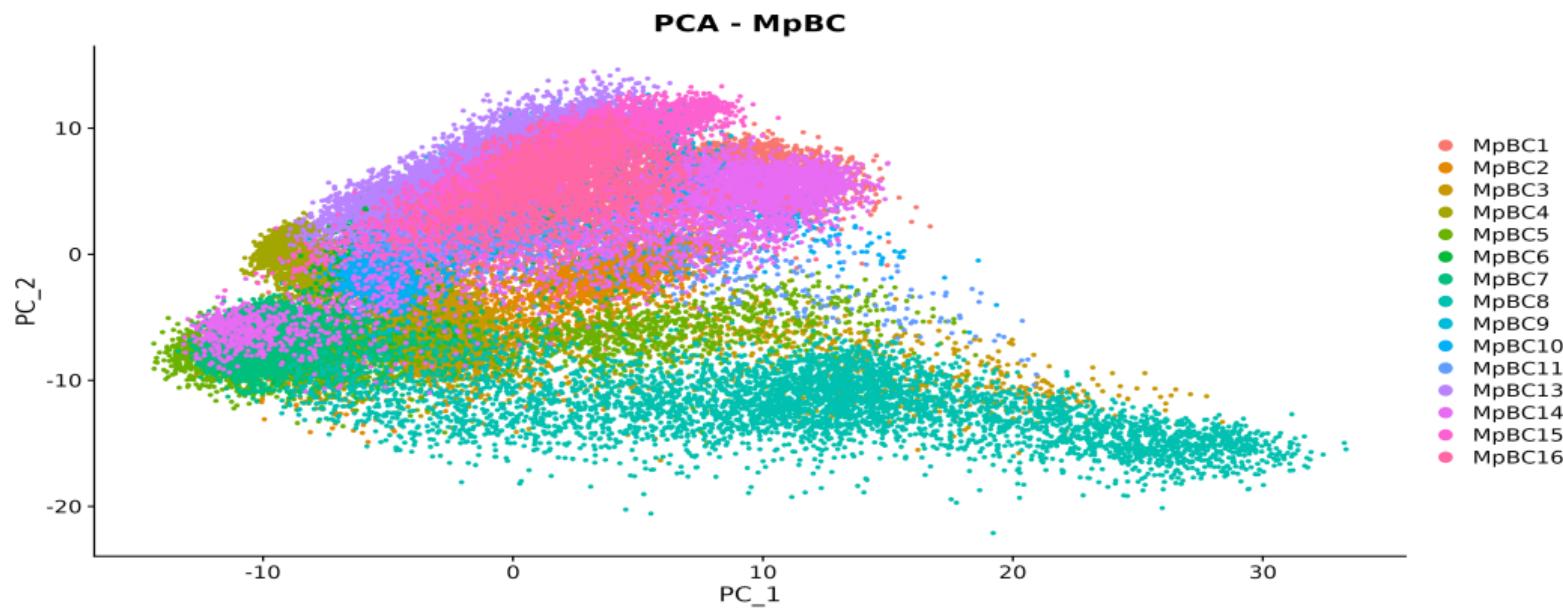
**Different transcriptional activities between tumor cell-types**

# Normalization

## Points to discuss

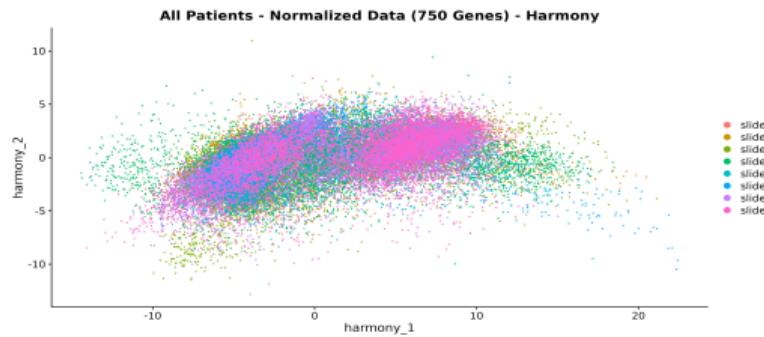
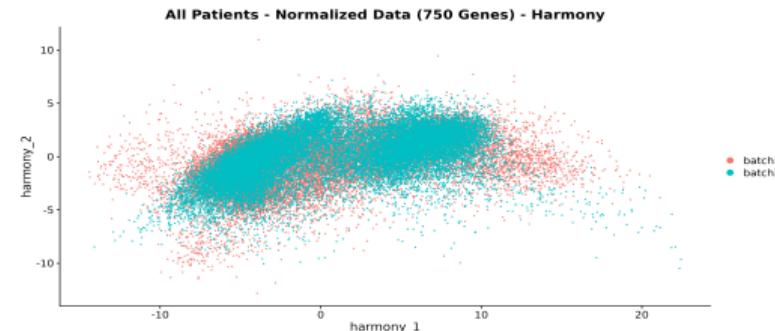
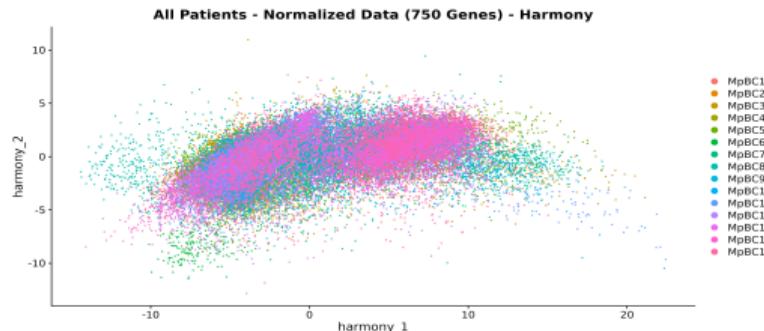
- Using Seurat::NormalizeData() + Seurat::ScaleData() (rather than Seurat::SCTransform())
- Log Normalization
- Several parameters (scale.factor, number of variable features...) empirically determined -> 750 genes
- Attempt to normalize individually each object (rather than normalize the whole object) -> No significant differences

## Dimension reduction : PCA



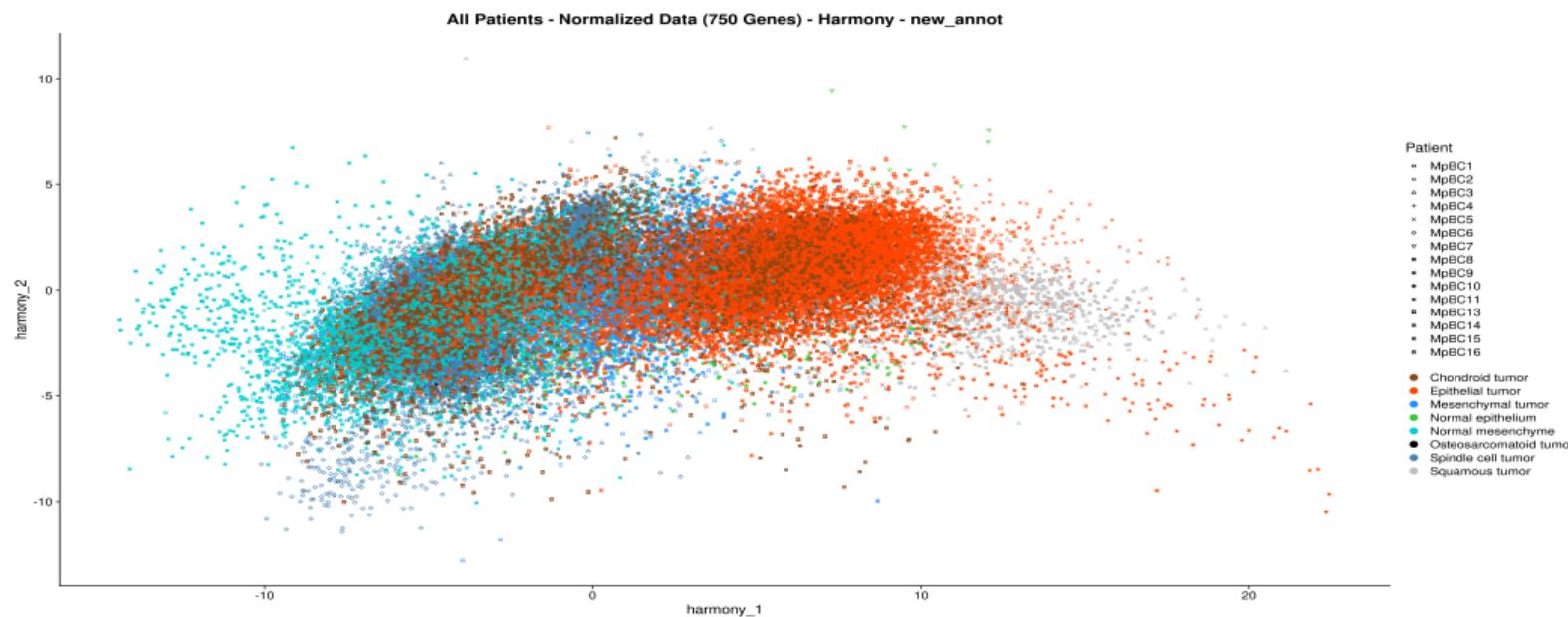
**Strong batch effect between samples to be corrected**

# Batch effect correction : Harmony



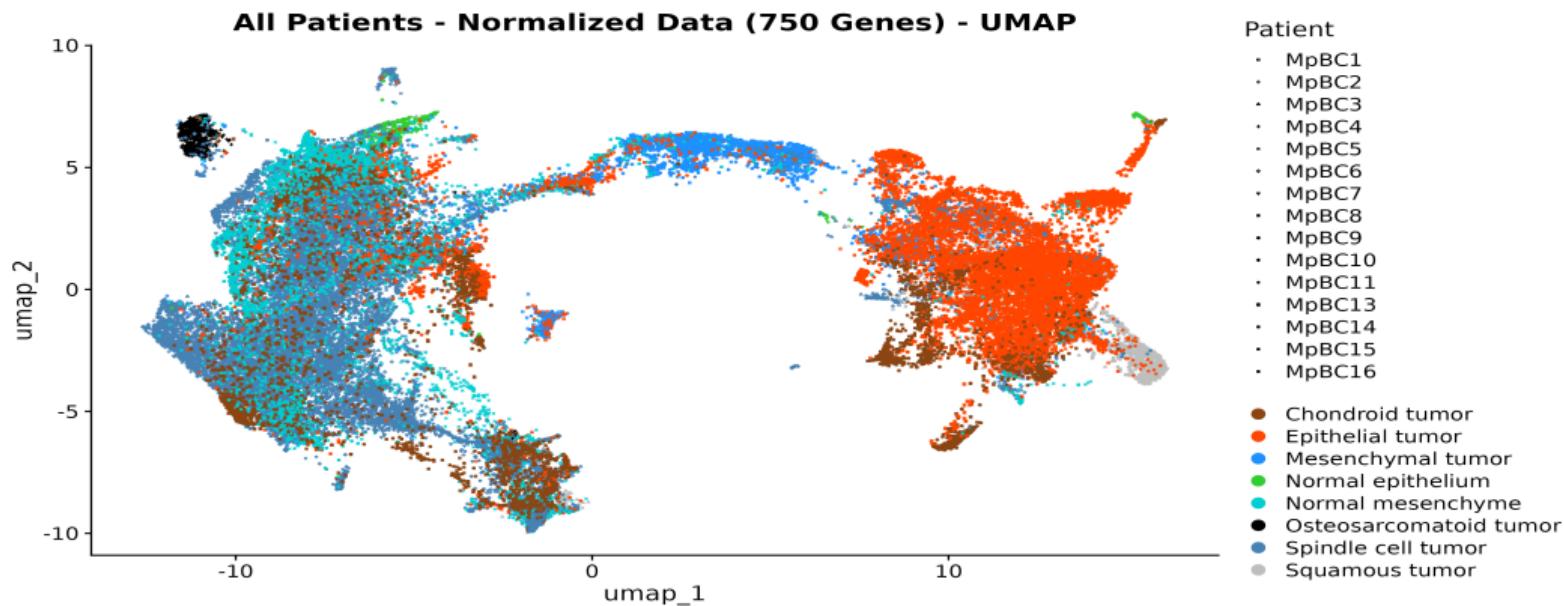
**Batch effect corrected between patients,  
batch sequencing and slide sequencing**

## Batch effect correction : Harmony



**Good separation between epithelial and mesenchymal phenotypes**

# UMAP Projection



**Clear epithelial-mesenchymal axis (umap1), with few ambiguous cells**

# UMAP Projection

## Points to discuss

- Some points/cells are still patient-specific (Chondroid and Epithelial cells).
  - Enhance Harmony correction (but jeopardize biological signal) ?
- Chondroid cells difficult to regroup and Spindle cells really heterogenous...

## Solutions

- Single nuclei and spot deconvolution to improve cell groups according to cell types/annotations

# UMAP Projection

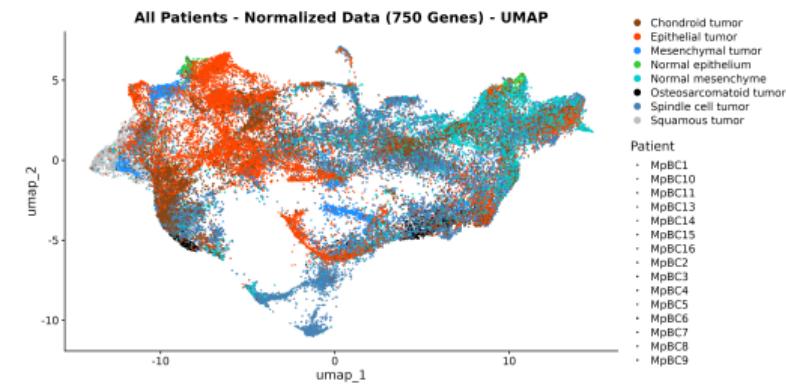
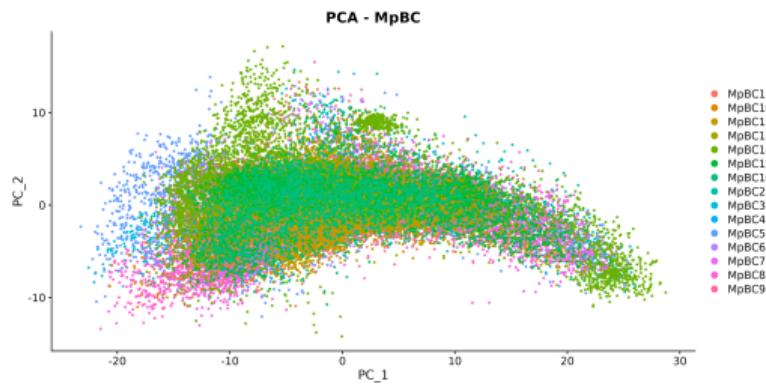
## Question

- As samples have different tumor composition, how do we manage the risk of overcorrection with Harmony and loss of biological signal ?

## Alternatives

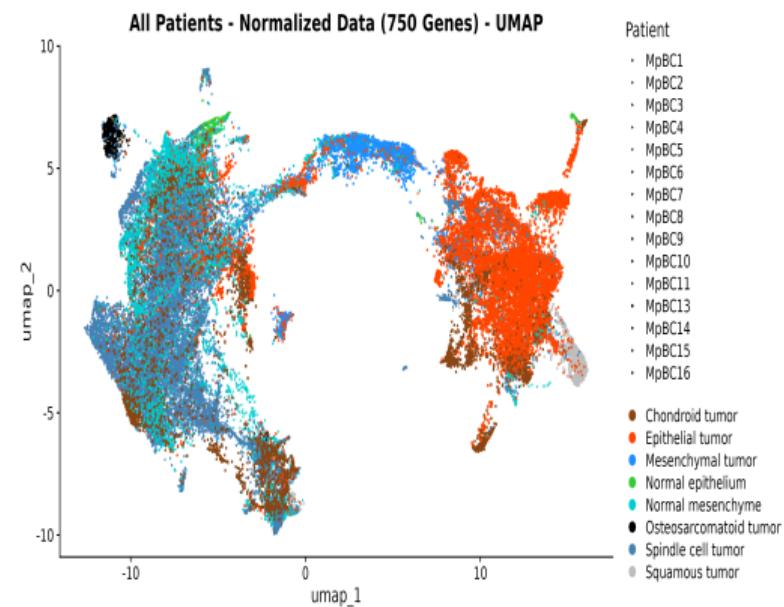
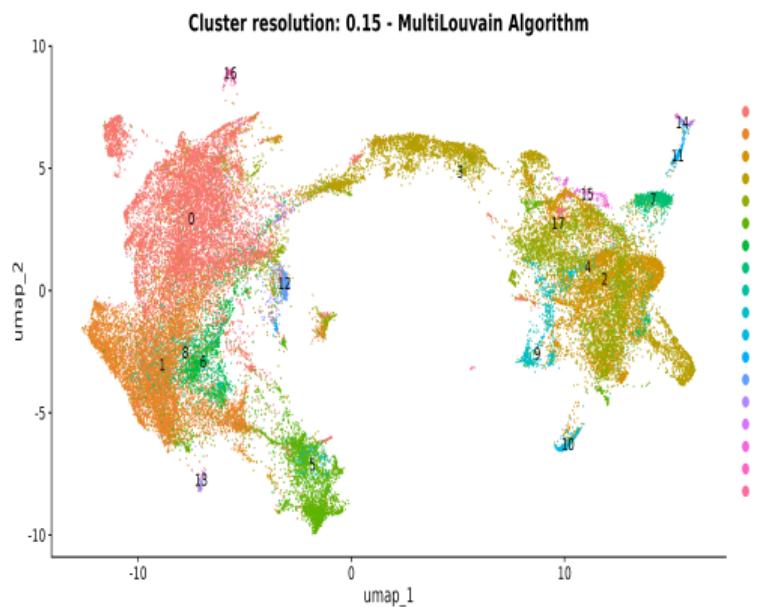
- Seurat Integration (Seurat::FindAnchors())
  - Not conclusive (for our dataset)
  - Samples with different tumor cell types -> Too few common anchors (features) shared between all samples

# UMAP Projection



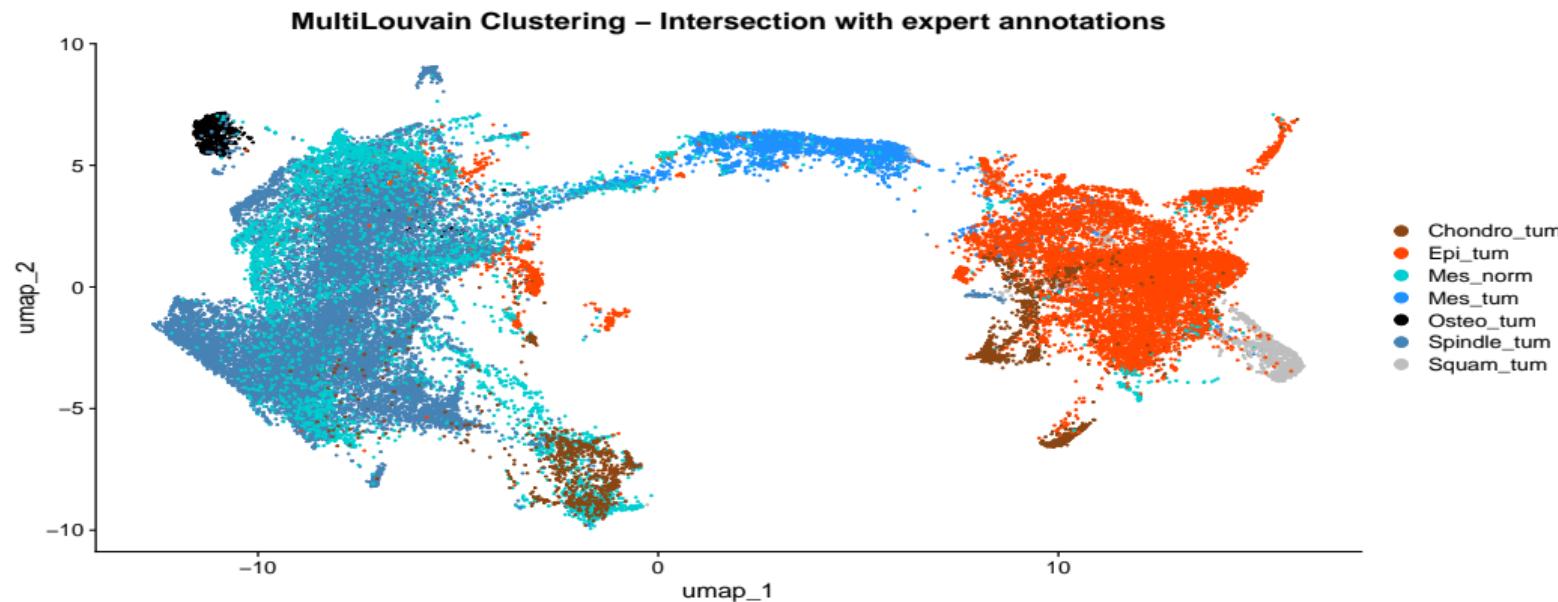
**UMAP projection with Seurat integration not good as UMAP with Harmony correction**

# Clustering



**Multi-Louvain algorithm : 18 clusters for 7 tumor cell-types**

# Clustering



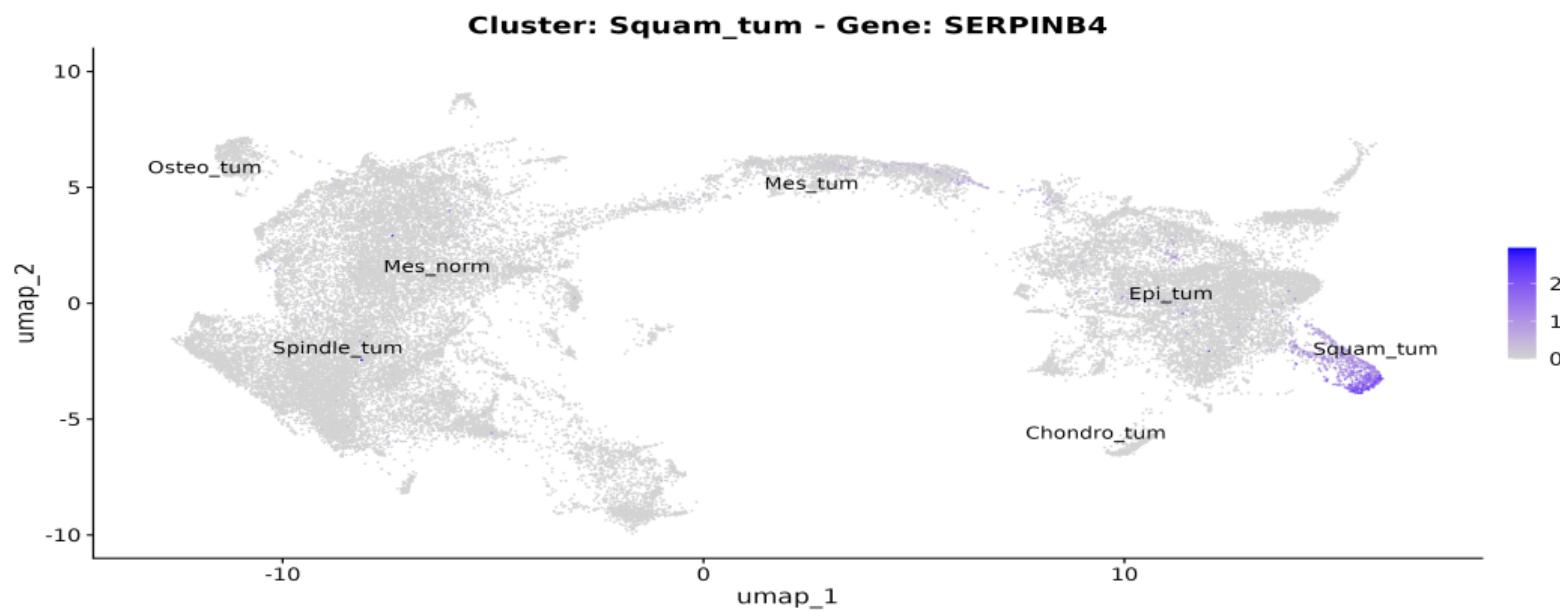
**Combine the two informations (Louvain clustering and expert annotations) to work with "purified" clusters**

# Clustering

## Points to discuss

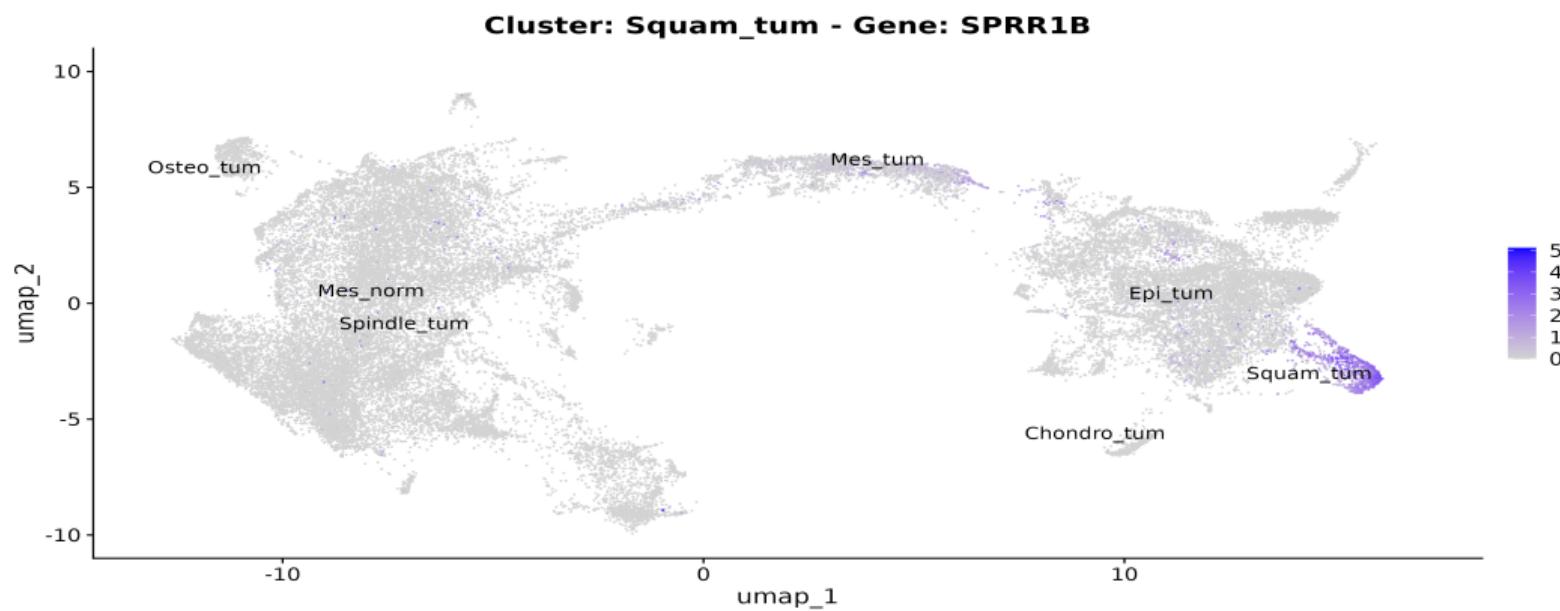
- Mesenchymal tumor cluster close to Squamous tumor...
  - Strange, because phenotypically very different
  - Come from the same patient (MpBC8)
  - Harmony correction not enough ?
- Some clusters are still patient-specific more than cell-types-specific (e.g. Epi-tum)

# Feature Plots



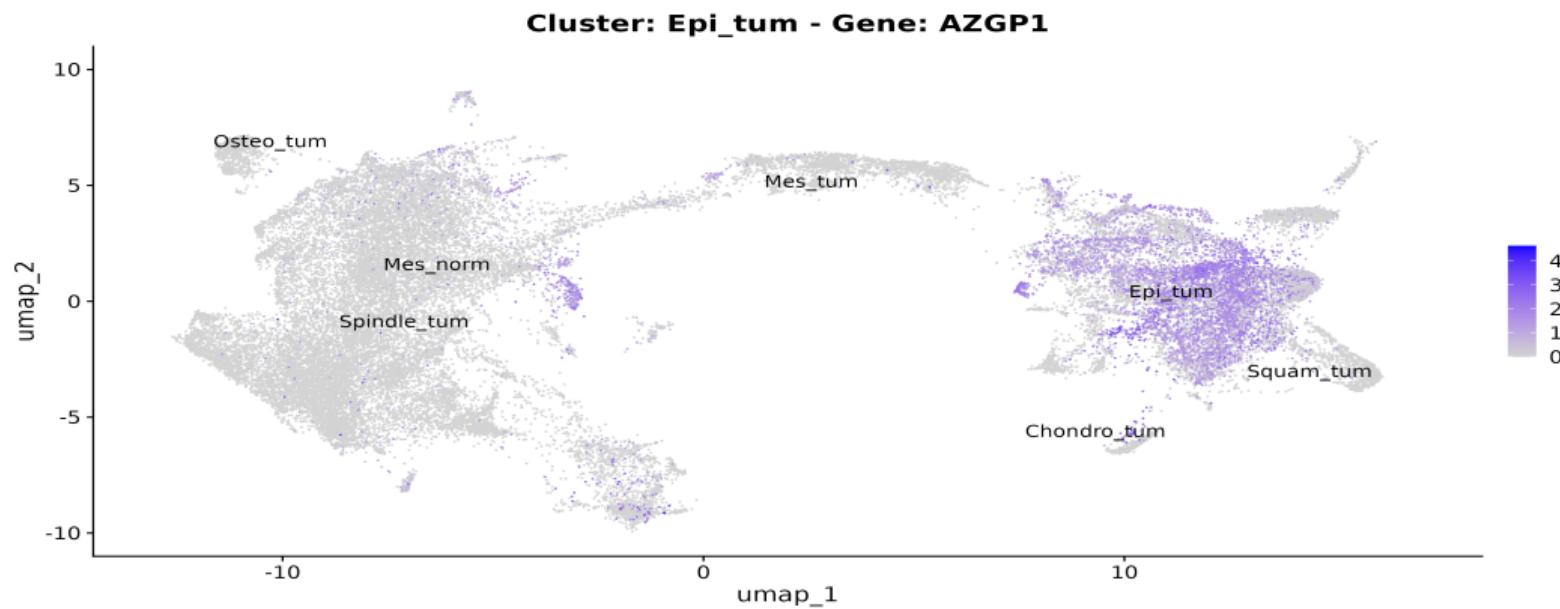
**Feature Plot of SERPINB4 gene (Squamous cluster)**

# Feature Plots



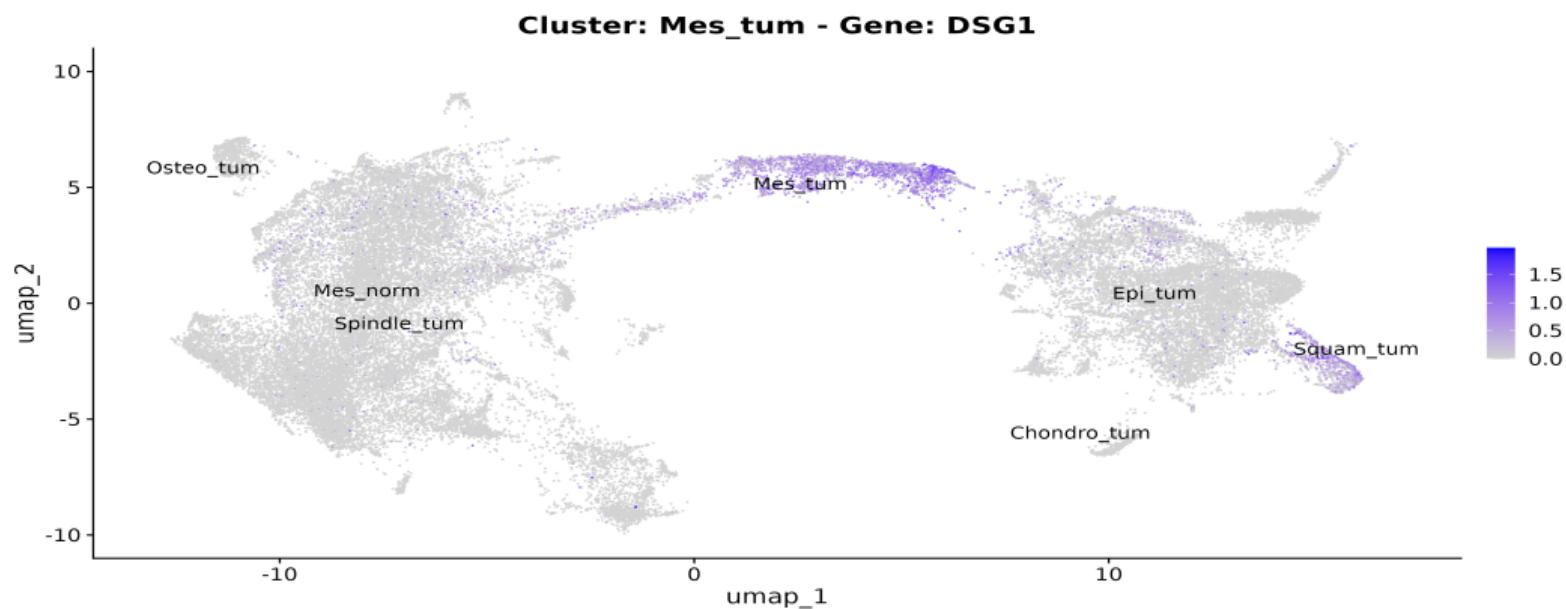
**Feature Plot of SPRR1B gene (Squamous cluster)**

# Feature Plots



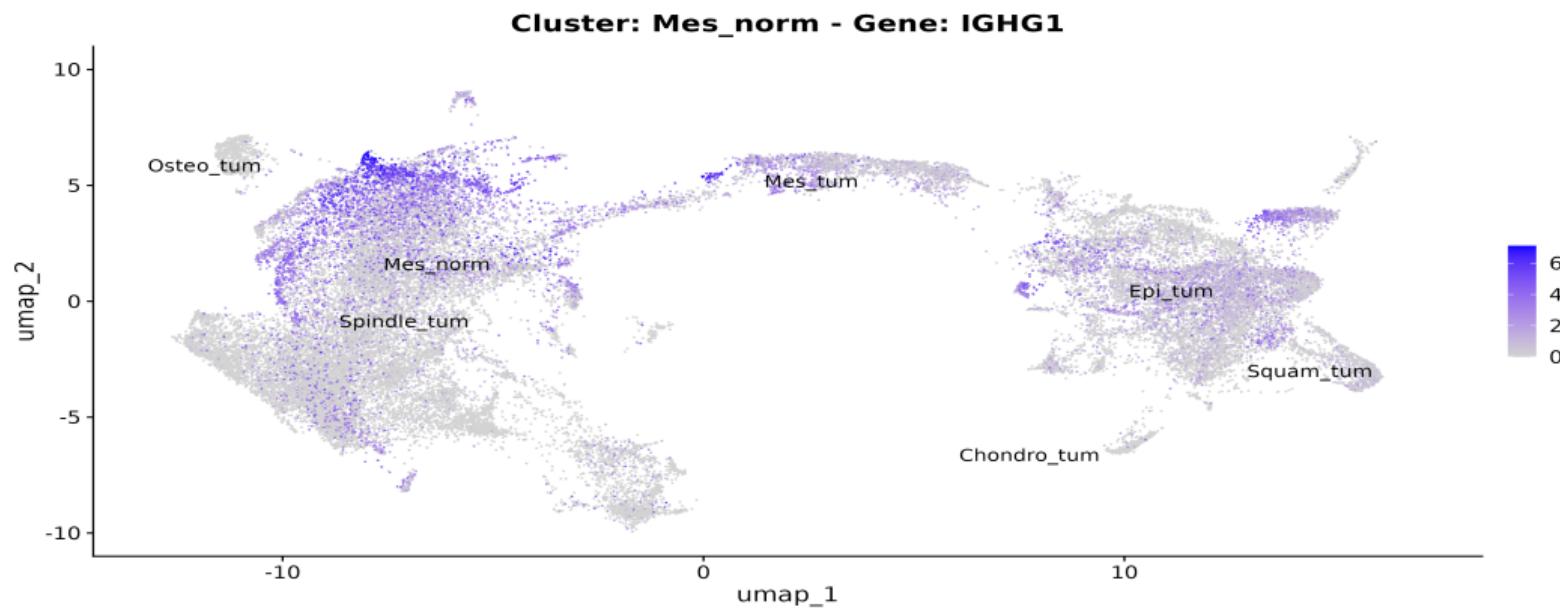
**Feature Plot of AZGP1 gene (Epithelial cluster)**

# Feature Plots



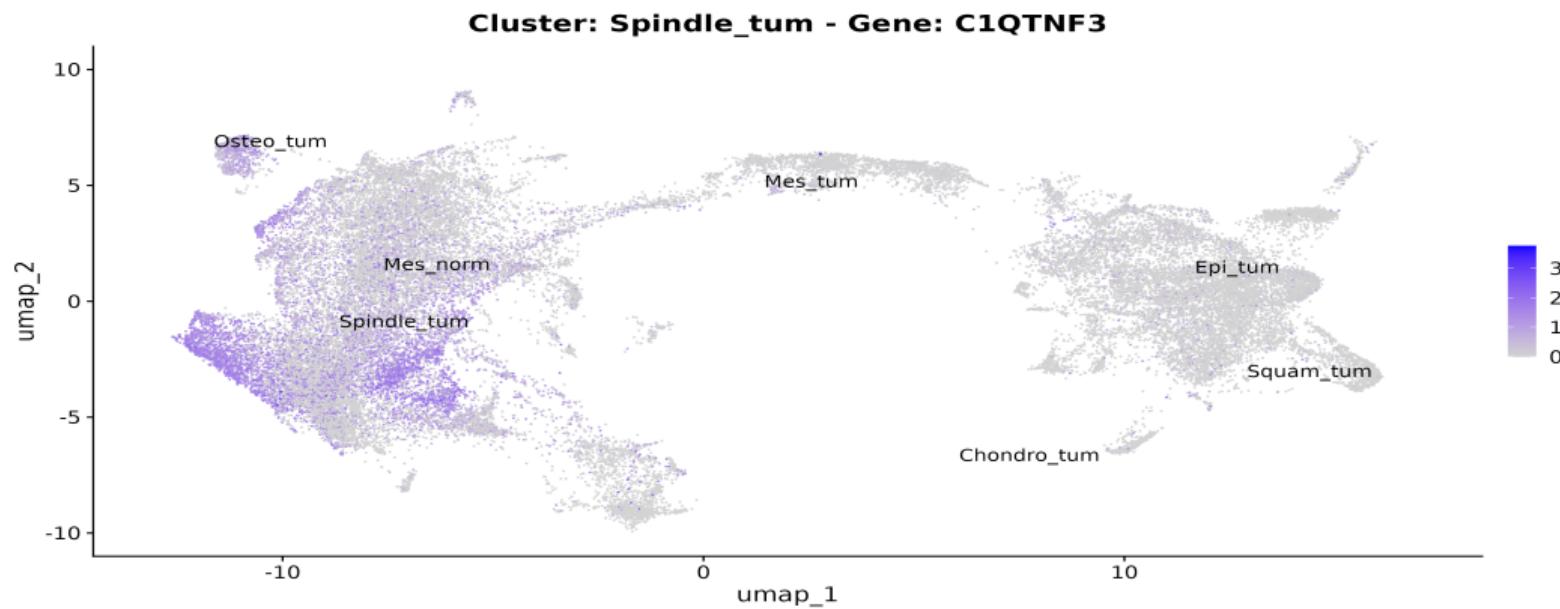
**Feature Plot of DSG1 gene (Mesenchymal tumor cluster)**

# Feature Plots



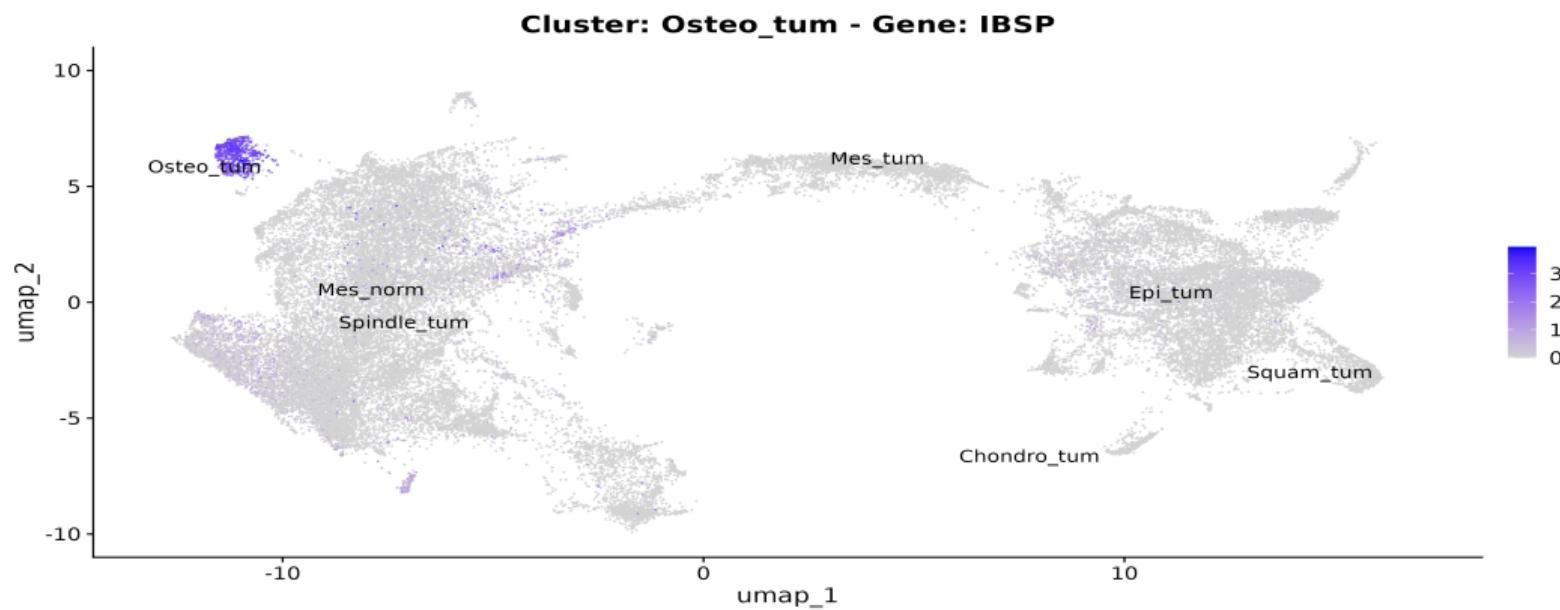
**Feature Plot of IGHG1 gene (Mesenchymal normal cluster)**

# Feature Plots



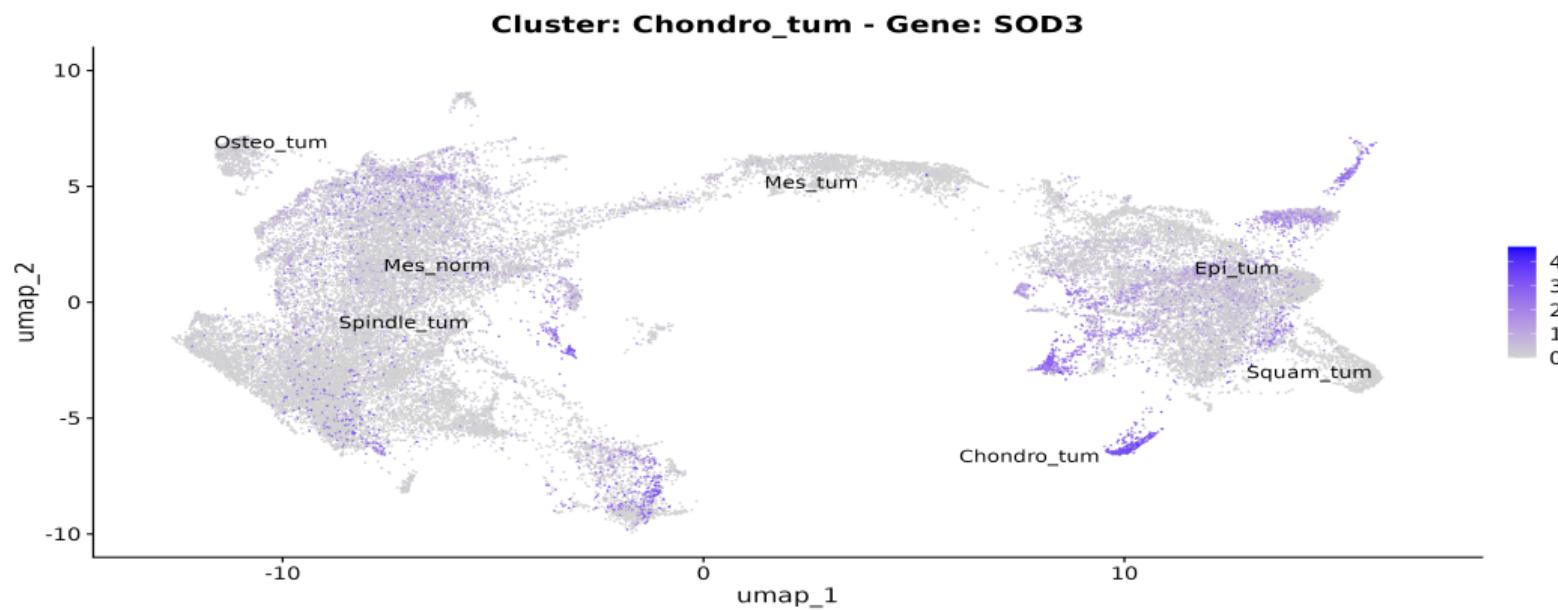
**Feature Plot of C1QTNF3 gene (Spindle cluster)**

# Feature Plots



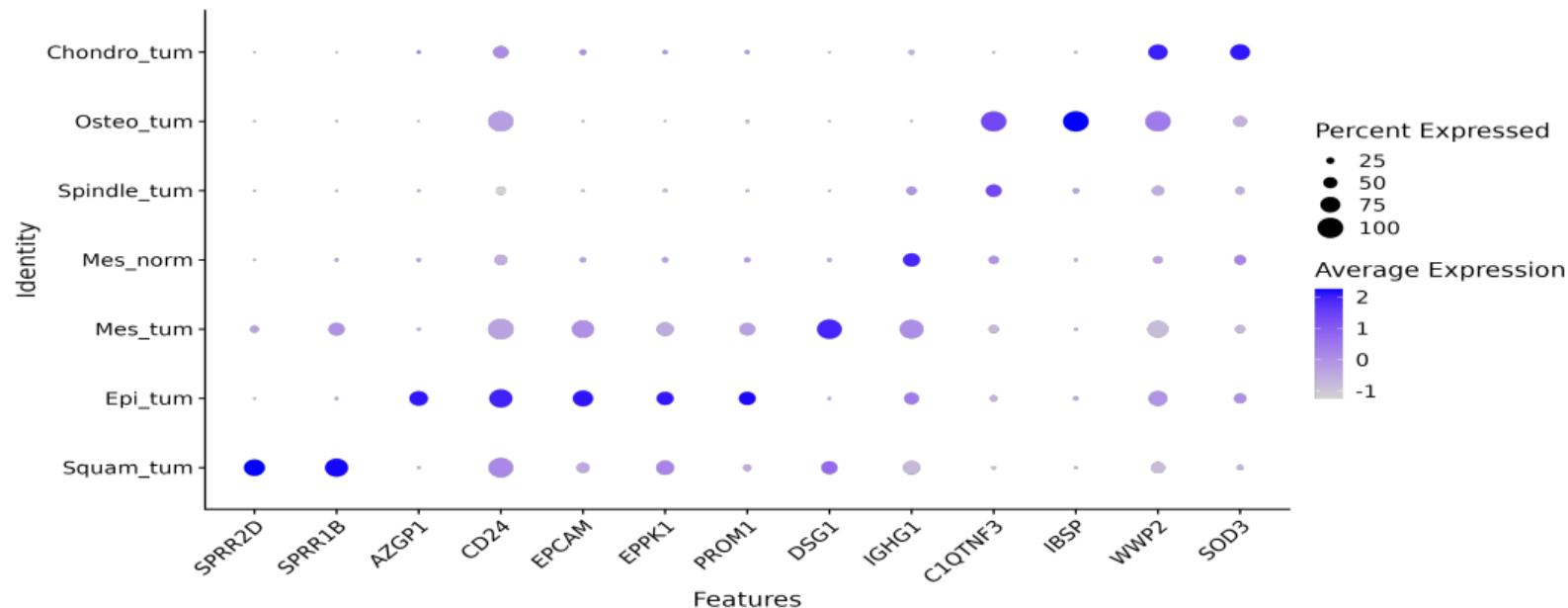
**Feature Plot of IBSP gene (Osteosarcomatoid cluster)**

# Feature Plots



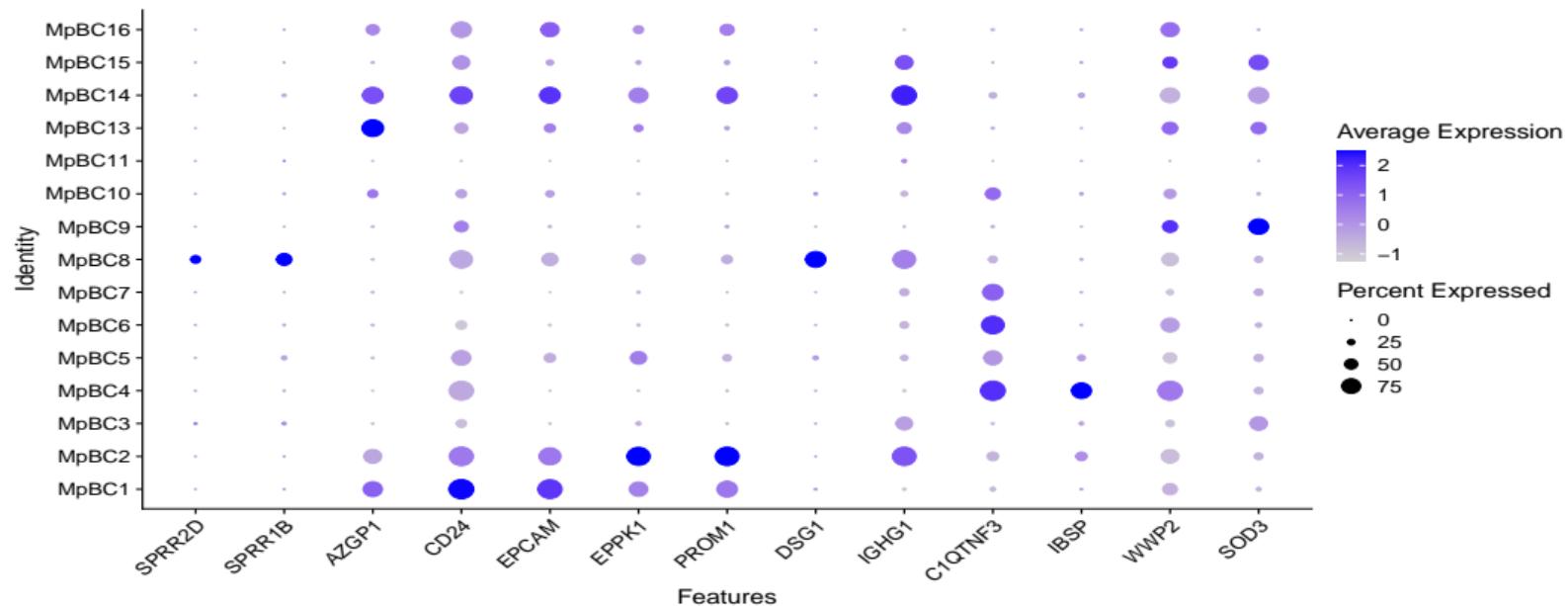
**Feature Plot of SOD3 gene (Chondroid cluster)**

## Dot Plot



Some markers seems to be overexpressed only in specific clusters...

## Dot Plot



But some cluster-specific markers are also patient-specific markers

## 1 Introduction

## 2 Workflow

## 3 Results

## 4 Outlook

In the next months,

- Search for **Copy Number Alteration (CNA)** in specific tumor cell types -> *InferCNVplus*
- Perform cell-deconvolution with **sNuc-seq** (improve cluster resolutions and markers) + Realize cell-trajectory analysis (*Slingshot*)
- **Epigenomic** and tumoral **microenvironment analysis**
- ...And more fun stuff !

**Thank you for listening !**

Tutor : Dr Pierre Martinez

jordan.dutel@lyon.unicancer.fr