

Atelier scRNA-seq

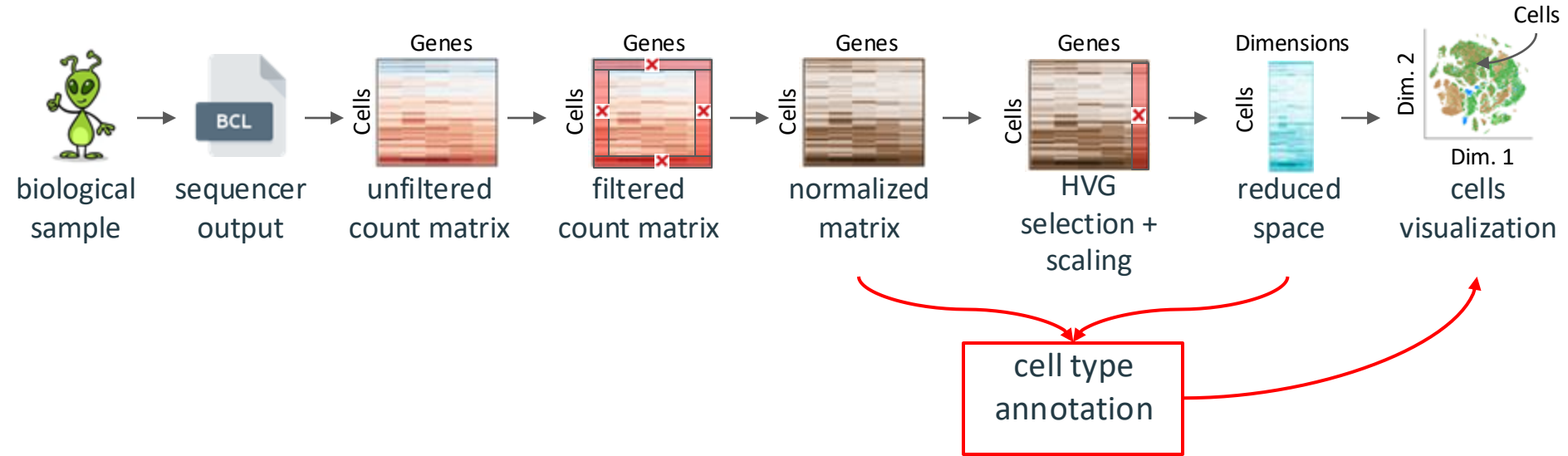
Cell type annotation

Sébastien Mella - Institut Pasteur, Paris

Margot Tragin - Institut Imagine, Paris

D'après Audrey Onfroy
et Eulalie Liorzou

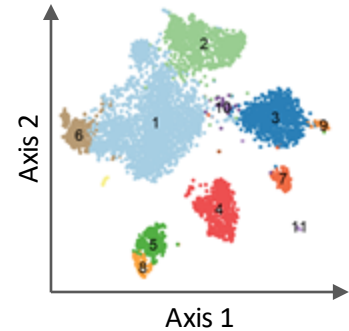
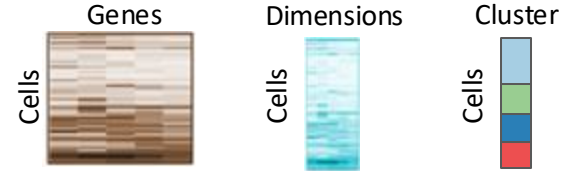
scRNA-Seq pipeline overview



What is available ?

We have :

- **gene expression matrix** : for each cell, gene expression is available
- **reduced space** : gene expression matrix is summarized in N dimensions
- **clustering** : each cell belongs to a specific cluster
- **2D space** : cells can be visualized on a 2D representation
- **biologist knowledge**
- **internet connection**



Objectives

Method 1 : **Manual cluster annotation** using differential expression

- **Input data** : which parts of the Seurat analysis are necessary to annotate clusters ?
- **Analysis tools** : know and understand the functions used to define marker genes
- **Visualisation** of marker genes

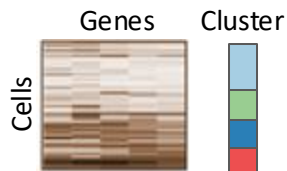
Method 2 : **Automatic** annotation

Optional

- Understand why it can be useful
- Know the limits

How to annotate cells for cell type ?

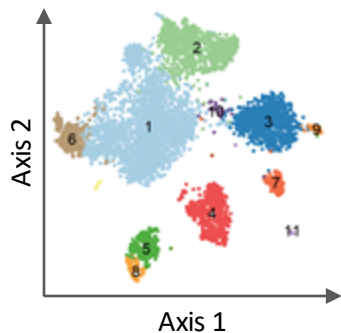
Method 1 : Manual cluster annotation using differential expression



For each cluster :

Exemple using **cluster 4**

- 1) Differential expression between **cluster 4** and all others
- 2) Look at gene expression on the 2D projection, to validate **specificity** and **representativeness**
- 3) Find the cell population corresponding to your gene set
- 4) Annotate **cluster 4**



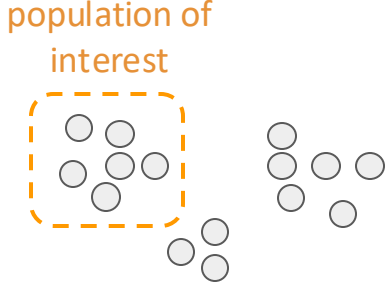
and / or



Advantages	Limits
<ul style="list-style-type: none">● Easy to implement● May be the only solution● Everything is possible	<ul style="list-style-type: none">● Clustering : resolution, merged clusters, “bio-informatic” cluster● Change clustering ? Change annotation...● Knowledge : time-consuming

Specificity and representativeness

of a marker gene, for a cell population



		Specific ?	
		Yes	No
Representative ?	Yes	<p>Gene A</p>	<p>Gene B</p>
	No	<p>Gene C</p>	<p>Gene D</p>

How to annotate cells for cell type ?

Method 2 : **Automatic** annotation using **reference markers**



Steps :

- 1) Find a good marker gene reference (PanglaoDB, CellMarker, CancerSEA...)
- 2) Select a tool / model : **classifier**, **scoring function** ...

Cell type A	A, B, C
...	...
Cell type X	X, Y, Z



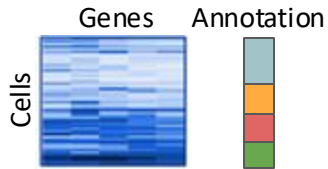
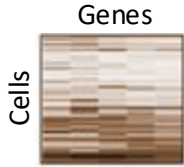
1) Annotate your dataset

Advantages	Limits
<ul style="list-style-type: none"> ● Single cell level is possible ● Design your own reference ● Made with human, mouse dataset 	<ul style="list-style-type: none"> ● Find the good reference markers ● Cell types arborescence ● Limited number of cell types : all cells are annotated, or “unknown” is possible ? ● Made with human, mouse dataset



How to annotate cells for cell type ?

Method 3 : Automatic annotation using reference dataset

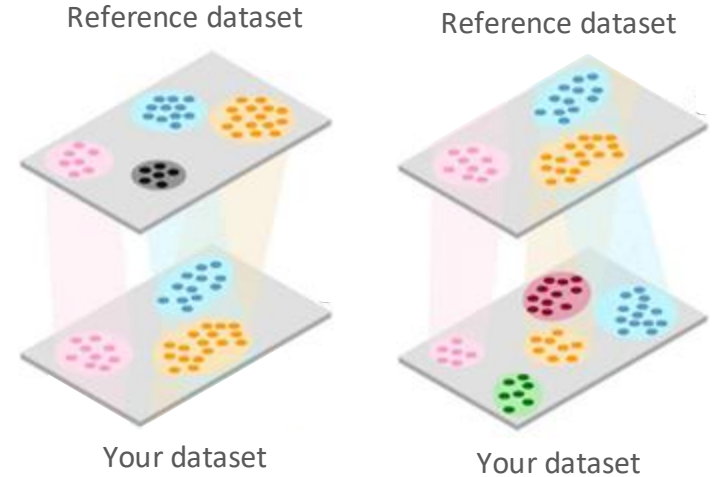


Reference dataset



Steps :

- 1) Find a good reference dataset : several bulk RNA-seq, one scRNA-seq...
- 2) Select a tool to transfer annotation
- 3) Annotate your dataset



Advantages	Limits
<ul style="list-style-type: none">● Single cell level● Design your own reference	<ul style="list-style-type: none">● Find the good reference dataset● Limited number of cell types

Take Home Messages

Method	Advantages	Limits
Manual cluster annotation using differential expression	<ul style="list-style-type: none"> ● Easy to implement ● May be the only solution ● Everything is possible 	<ul style="list-style-type: none"> ● Clustering : resolution, merged clusters, “bio-informatic” cluster ● Change clustering ? Change annotation... ● Knowledge : time-consuming
Automatic annotation using reference markers	<ul style="list-style-type: none"> ● Single cell level is possible ● Design your own reference 	<ul style="list-style-type: none"> ● Find the good reference markers ● Cell types arborescence ● Limited number of cell types : all cells are annotated, or “unknown” ?
Automatic annotation using reference dataset	<ul style="list-style-type: none"> ● Single cell level ● Design your own reference 	<ul style="list-style-type: none"> ● Find the good reference dataset ● Limited number of cell types : all cells are annotated, or “unknown” ?

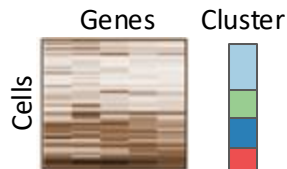
Advice :

1. Use manual cluster annotation to identify quickly your cell populations
2. Identify good markers for each cell populations → your reference markers
3. Use automatic cell annotation using your set of marker → your reference dataset
4. Use your references to annotate new dataset

How to annotate cells for cell type ?

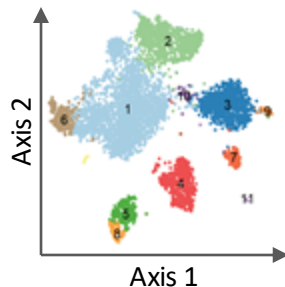
Method 1 : **Manual cluster annotation** using differential expression

Input data :



Analysis :

Exemple using **cluster 4**



Visualisation :

Dataset with :

- **Normalised** count matrix
- Pick **one clustering resolution**

For each cluster :

- Differential expression between **cluster 4** and all others

In Seurat the function used is : *FindAllMarkers()*

For each cluster :

- Validate the specificity and representativeness of your marker
- Annotate your cluster