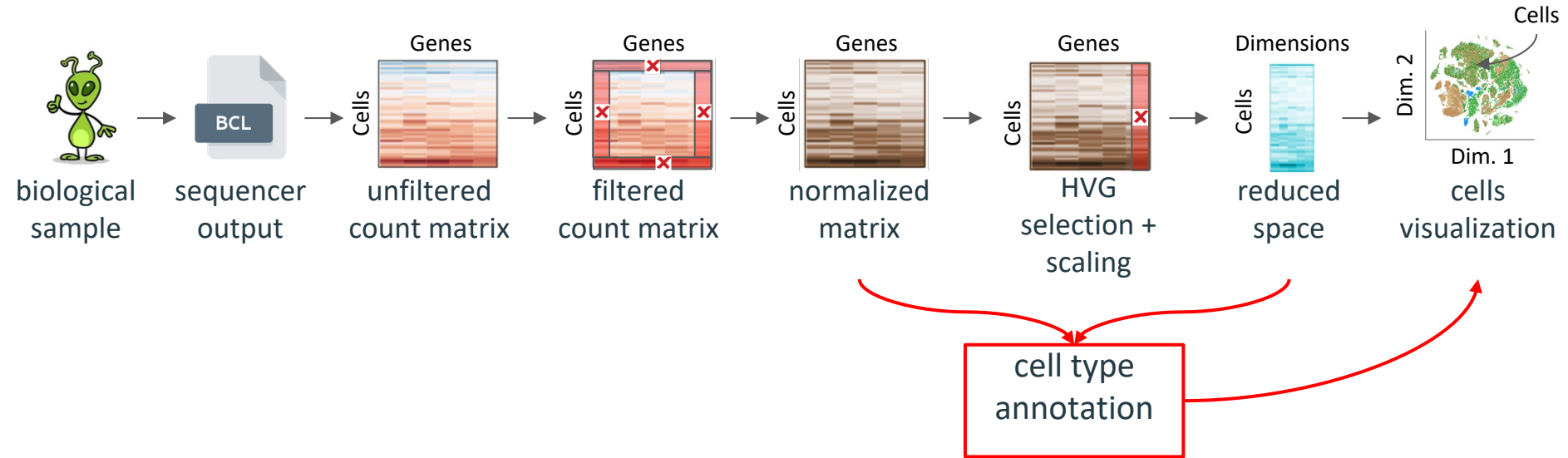




# scRNA-seq : cell type annotation

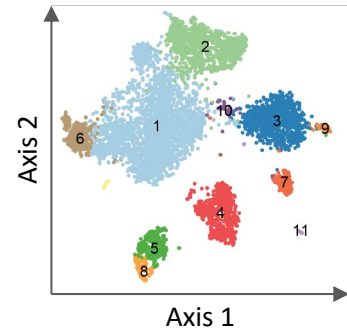
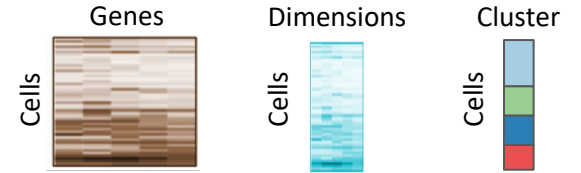
# scRNA-Seq pipeline overview



# What is available ?

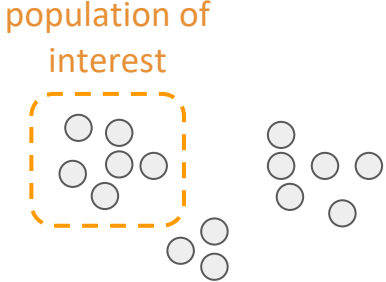
We have :

- **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**



# Specificity and representativeness

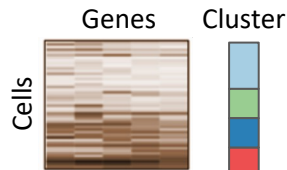
of a marker gene, for a cell population



		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 ?

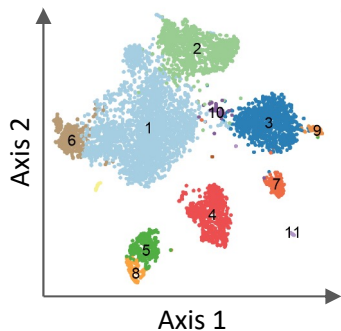
Method 1 : Manual cluster annotation using differential expression



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**



Cluster 1	A, B, C
...	...
Cluster n	X, Y, Z



Cluster 1	Cell type A
...	...
Cluster n	Cell type X

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

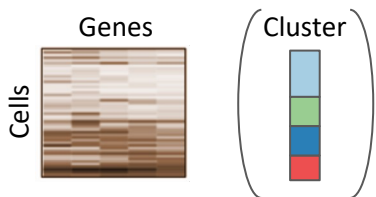


and / or



# How to annotate cells for cell type ?

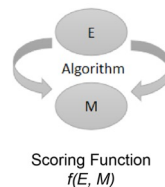
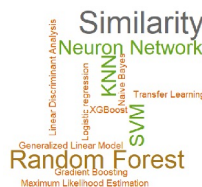
Method 2 : Automatic annotation using reference markers



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



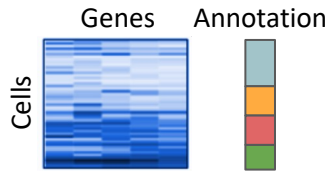
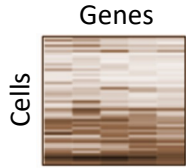
1) Annotate your dataset

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



# How to annotate cells for cell type ?

Method 3 : Automatic annotation using reference dataset

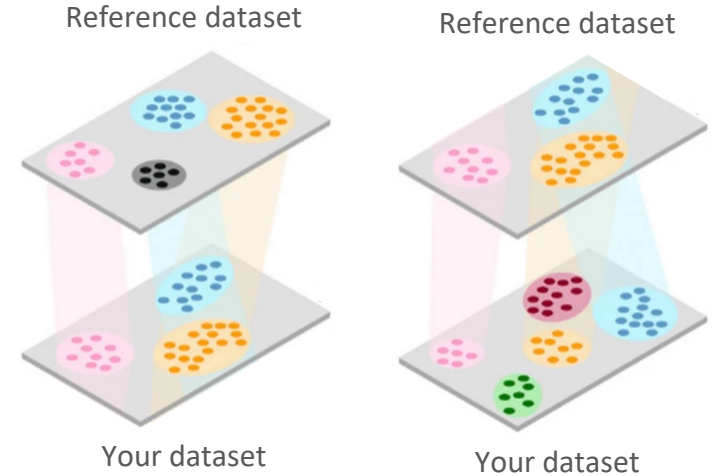


Reference dataset



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

# Take Home Messages

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

## 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