Spatial Transcriptomic Data

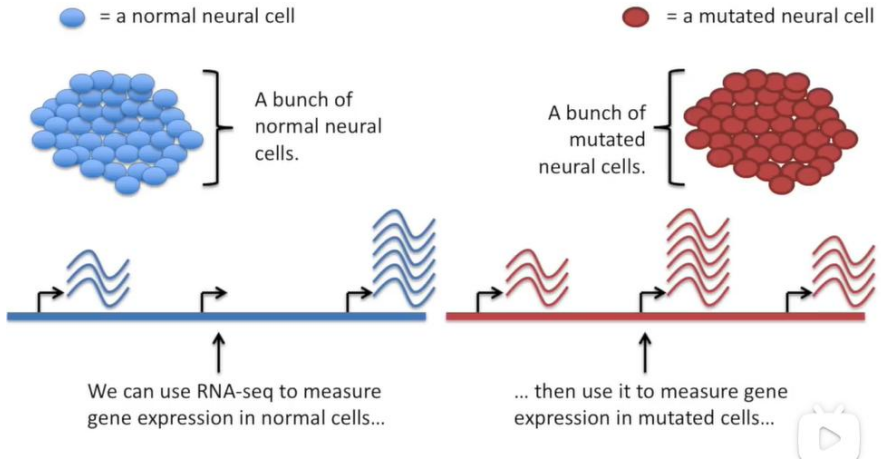
2022-01-09

CONTENTS

01

Introduction of STD

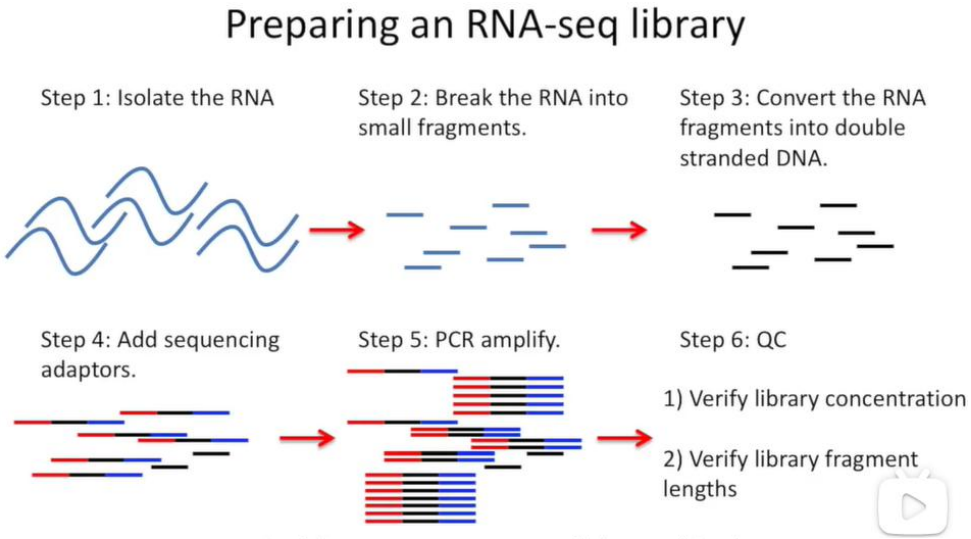
RNA-seq



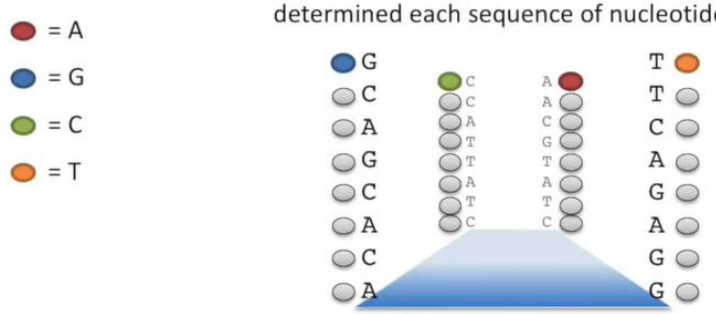
3 main steps for RNA-seq:
1. prepare a sequencing library
2. sequence
3. data analysis

https://www.bilibili.com/video/BV12441167CF?from=search&seid=12980852241798912825&spm_id_from=333.337.0.0

1



2



3

The raw data...

```
@NS500177:196:HFTTTAFXX:1:11101:10916:1458 2:N:0:CGCGGCTG
ACACGACGATGAGGTGACAGTCACGGAGGATAAGATCAATGCCCTCATTAAAGCAGCCGGTGTAA
+
AAAAAEEEEEEEEEEE//AEEEAEEEEEEEEEEE/EE/<<EE/AAEEAEE///EEEEAEEEAEA<
```

index; sequencing; ' +' ; quality score

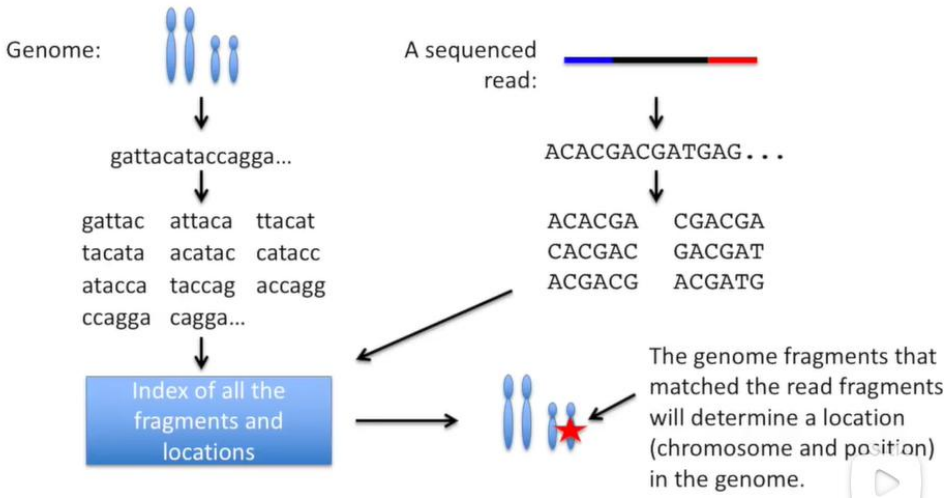
RNA-seq

Preprocession

Filter out garbage reads

Garbage reads are:
1. Reads with low quality base calls
2. Reads that are clearly artifacts of chemistry

Align the high quality reads to a genome



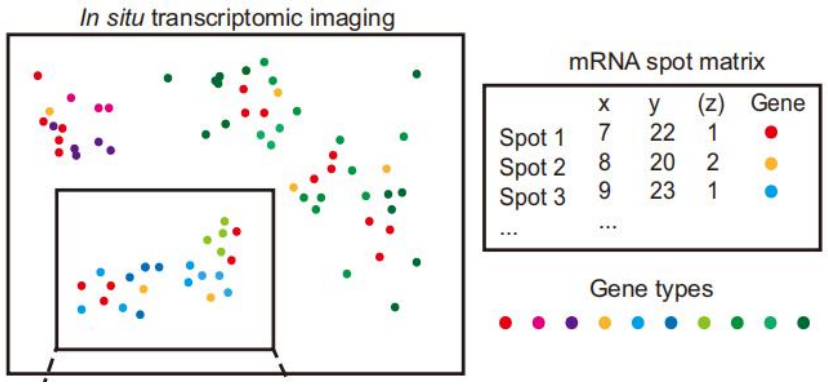
Count the number of reads per gene

Gene	Sample1	Sample2	Sample3...
A1BG	30	5	13...
A1BG-AS1	24	10	18...
A1CF	0	0	0...
A2M	5	9	7...
A2M-AS1	3563	5771	4123...
A2ML1	13	8	7...
...	...	...	...

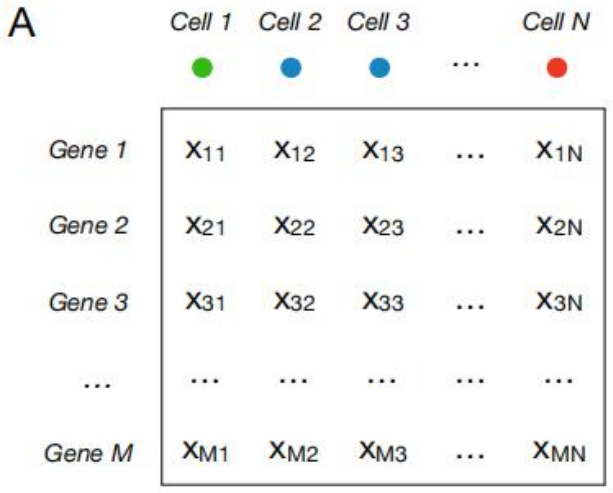
https://www.bilibili.com/video/BV12441167CF?from=search&seid=12980852241798912825&spm_id_from=333.337.0.0

STD: Data structure

https://doi.org/10.1038/s41467-021-26044-x



Gene type & Coordinate



How to get 2D(3D) coodinate:

1. spatial barcoding
2. in situ hybridization
3. in situ sequenceing

Method	Type	Resolution	Genes	Reference
Visium	Spatial barcoding	55μm	Whole transcriptome	(16)
Slide-seq	Spatial barcoding	10μm	Whole transcriptome	(17, 18)
HDST	Spatial barcoding	2μm	Whole transcriptome	(19)
DBiT-Seq	Spatial barcoding	10μm	Whole transcriptome	(20)
Seq-scope	Spatial barcoding	0.6μm	Whole transcriptome	(21)
Stereo-seq	Spatial barcoding	500nm	Whole transcriptome	(22)
seqFISH	*in situ* hybridization	single-molecule	>10,000	(23, 24)
merFISH	*in situ* hybridization	single-molecule	100 – 1,000	(25, 26)
STARmap	*in situ* sequencing	single-cell	160 – 1020	(27)
FISSEQ	*in situ* sequencing	subcellular	~ 8000	(28)

http://arxiv.org/abs/2110.07787

STD: Data analysis

data filter

1. genes and cells may be filtered based on threshold specific to the dataset.

2. gene expression per cell may be normalized to have the same total library size so that expression levels are comparable across cells.

http://arxiv.org/abs/2110.07787

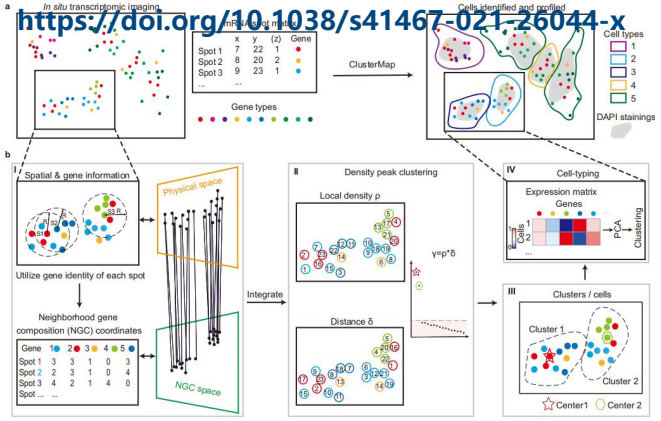
analysis and visualization in expression domain

A first step in the spatial transcriptomic analysis is to identify the cell type (for datasets of single-cell resolution) or cell mixture (for datasets of multicellular resolution) of each spatial unit or spot. Cell type identification usually starts with dimensionality reduction technique to reduce time and space complexity for downstream analysis. The reduced representations are used to cluster cells based on the assumption that cell of the same type falls into the same cluster.

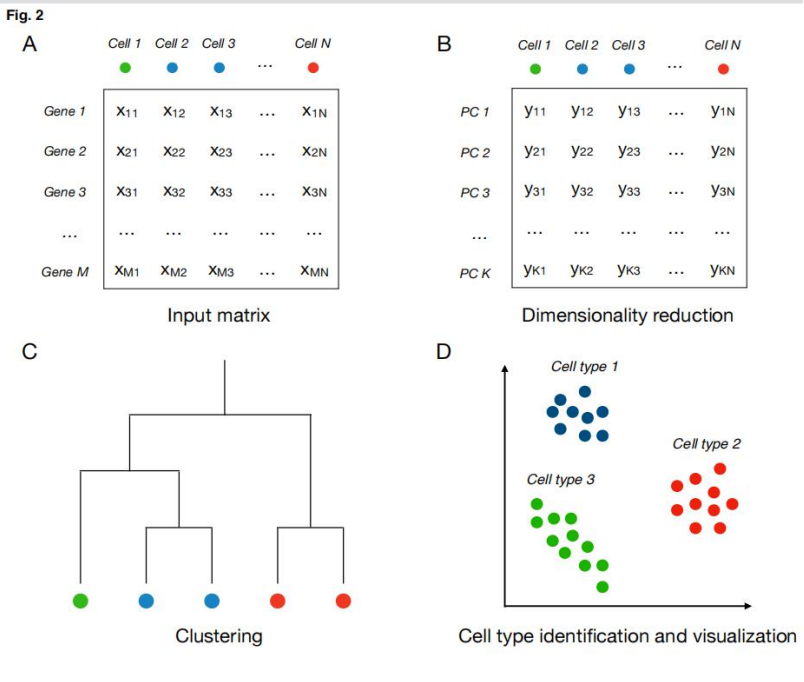
Clustering

Agglomerative clustering : a class of methods that iteratively aggregate data points into clusters.

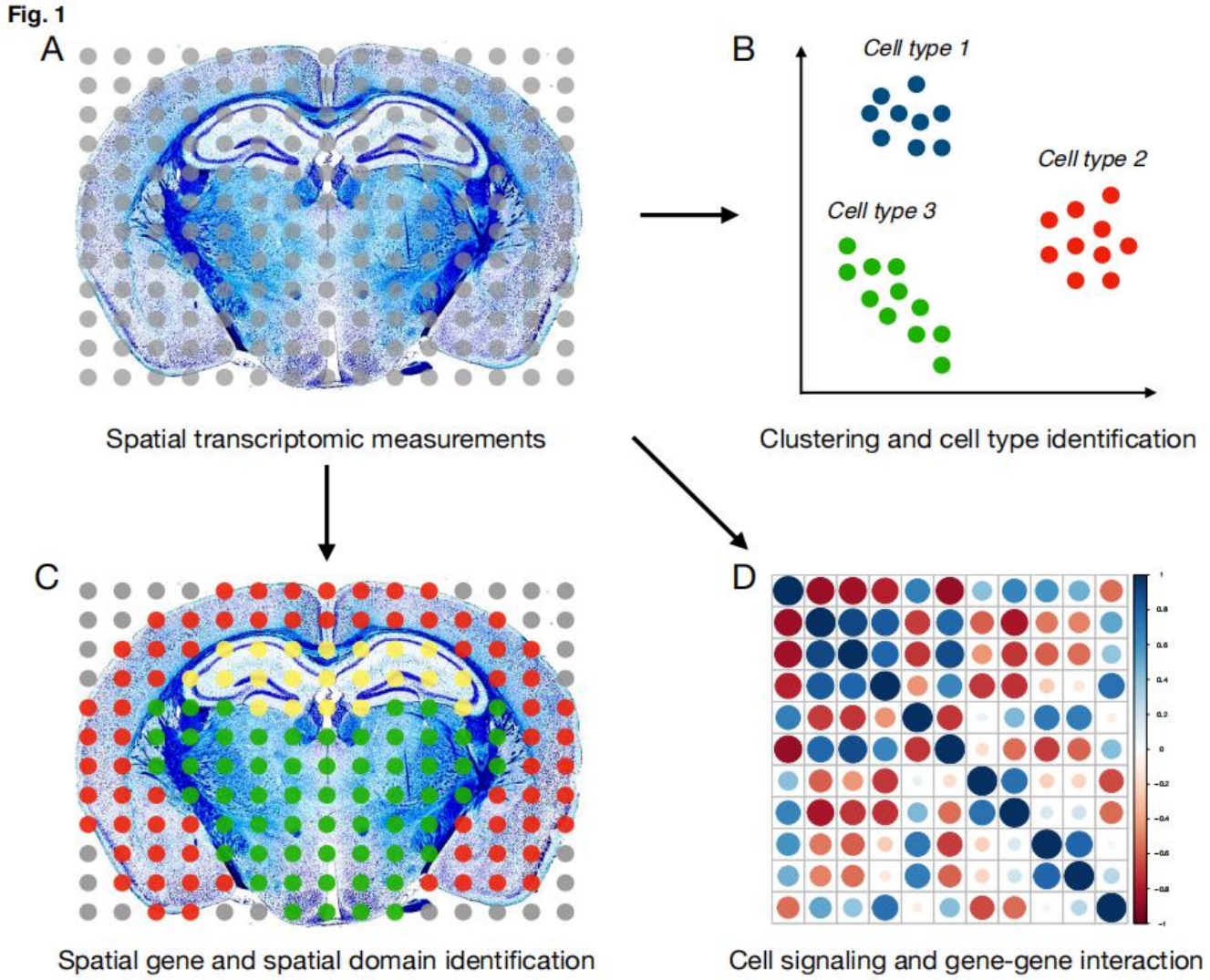
https://doi.org/10.1038/s41467-021-26044-x



Identification of cell types



STD: Data analysis



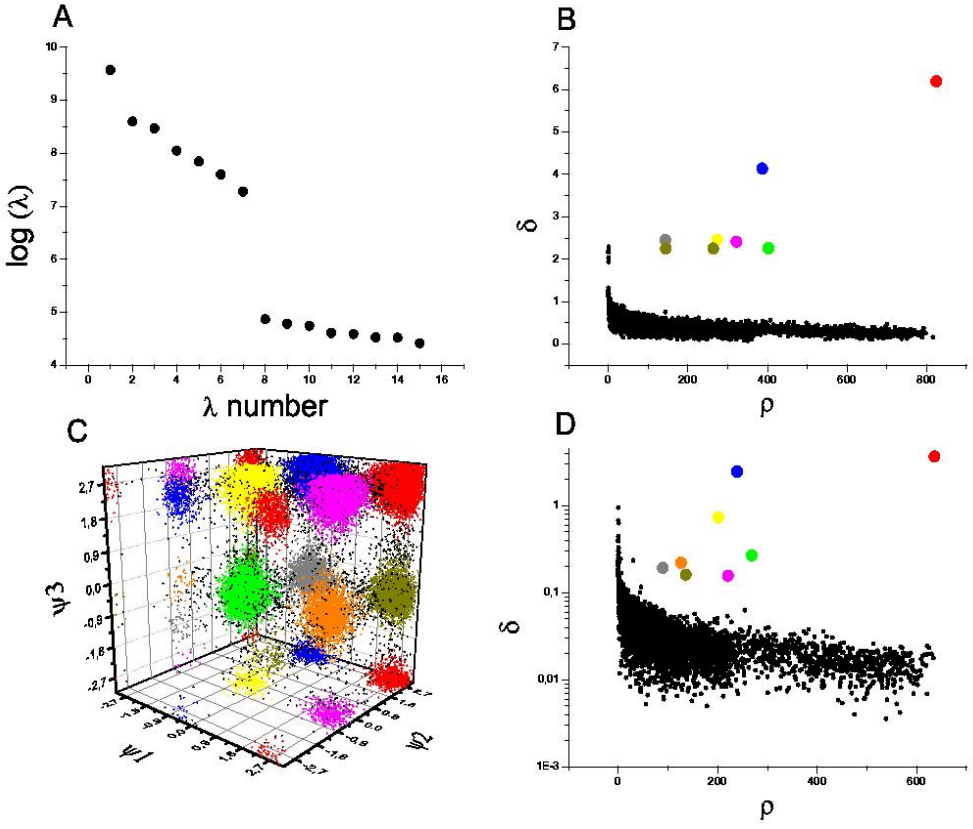
Data analysis:
1. Clustering and cell type identification
2. Spatial gene and spatial domain identificaiton
3. Cell signaling and gene-gene interation

http://arxiv.org/abs/2110.07787

02

ClusterMap

Density peak clustering (DPC)



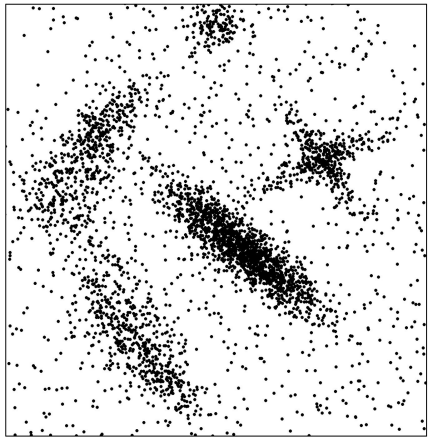
When applied to Molecular Dynamics trajectories, the results are coherent with a much more complex kinetic model, even when employing different similarity measures like RMSD or Dihedral distance.

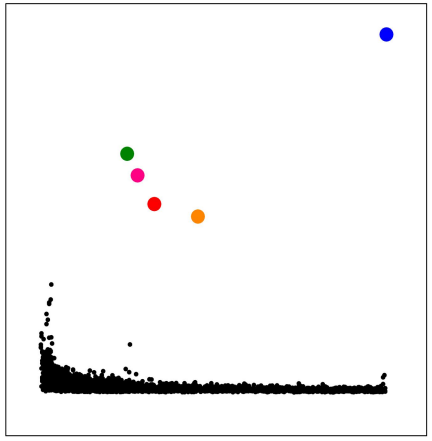
k-NN by ρ

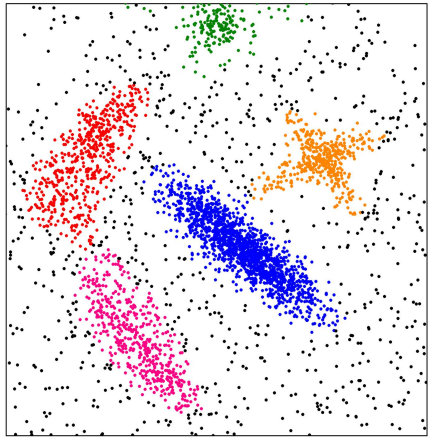
$$\rho_i = \sum \chi\left(d_{ij} - d_c\right)$$

$$\chi(x) = \begin{cases} 1 & \text{if } x < 0 \\ 0 & \text{if } otherwise \end{cases}$$

$$\delta_i = \min_{j:\rho_j > \rho_i} \left(d_{ij}\right)$$

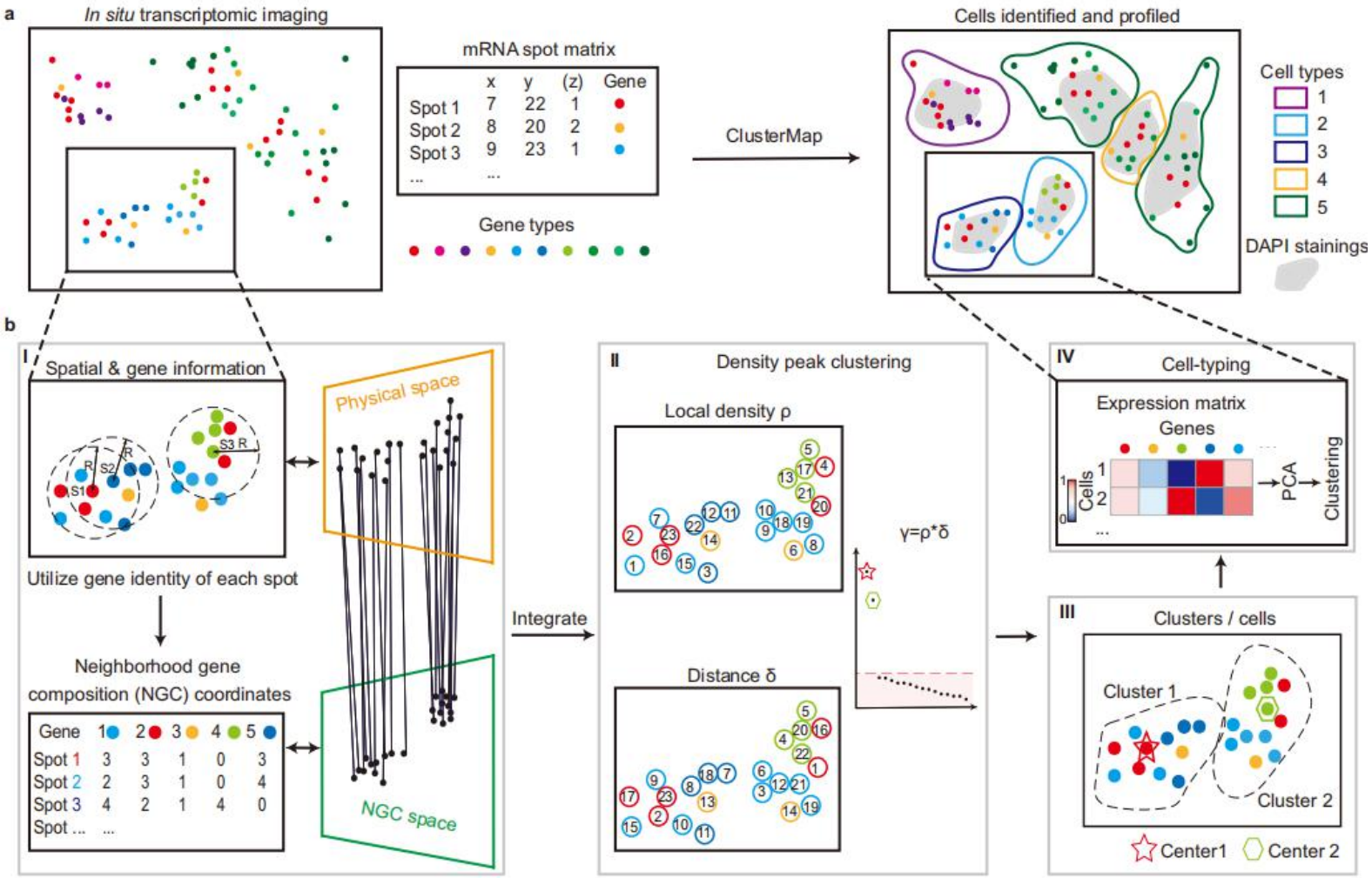






https://people.sissa.it/~laio/Research/Res_clustering.php

ClusterMap



https://doi.org/10.1038/s41467-021-26044-x

https://github.com/wanglab-broad/ClusterMap

Integration of the physical and NGC coordinate

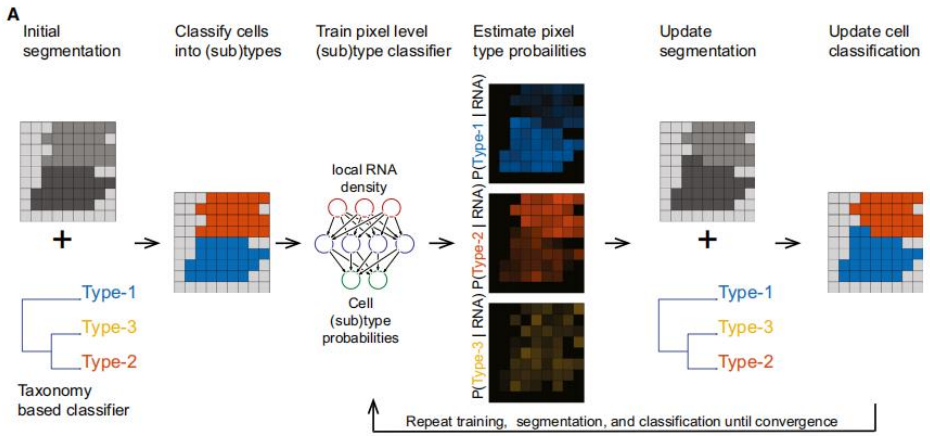
$$P(i) = <x_i, y_i, (z_i)> \quad (6)$$

$$NGC(i) = <Num_{Gene\,1}, Num_{Gene\,2}, \ldots, Num_{Gene\,t}, \ldots, Num_{Gene\,T}> \quad (7)$$

$$d_{ij} = \frac{Distance\{P(i), P(j)\}}{SpearmanCorr\{\mathrm{NGC}(i), \mathrm{NGC}(j)\}}$$

$$\rho_i = \sum_j I(d_{ij} - d_{\max})^* e^{-(d_{ij}/R)^2} \quad (4)$$

$$\delta_i = min(d_{ij}),\ j: \rho_j > \rho_i \quad (5)$$



https://doi.org/10.15252/msb.202010108