Assessment of dimensionality reduction methods for the detection of intermediate cancer phenotypes from ‘omic data

Poizat Jérôme

Supervised by:
Dr. Nicolas Alcala, Dr. Matthieu Foll and Dr. Lynnette Fernandez-Cuesta

Université Claude Bernard  Lyon 1

International Agency for Research on Cancer

World Health Organization

Context : Cancer classification

Cancer classification

Proper diagnostic

Adapted therapeutics

Dimension reduction

Expression matrix

	geneA	geneB	geneC	geneD	...
cancer1_a	42	58	12	47	...
cancer1_b	25	69	87	42	...
cancer2_a	35	91	25	81	...
cancer2_b	9	71	44	7	...

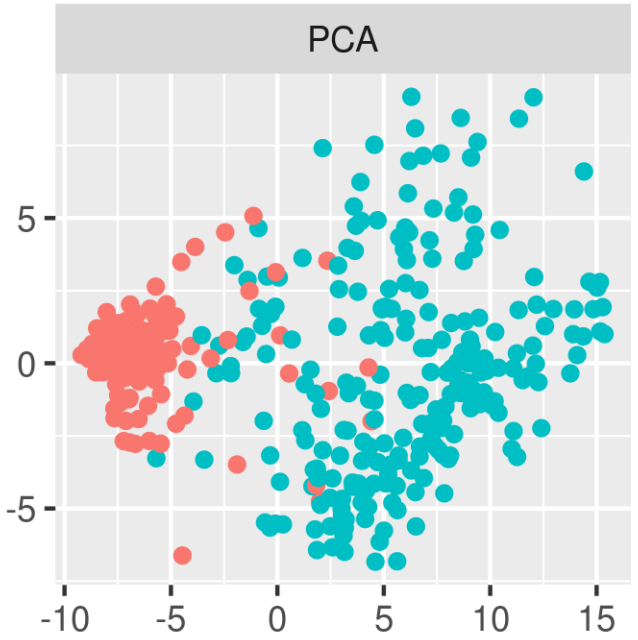
Dimension reduction

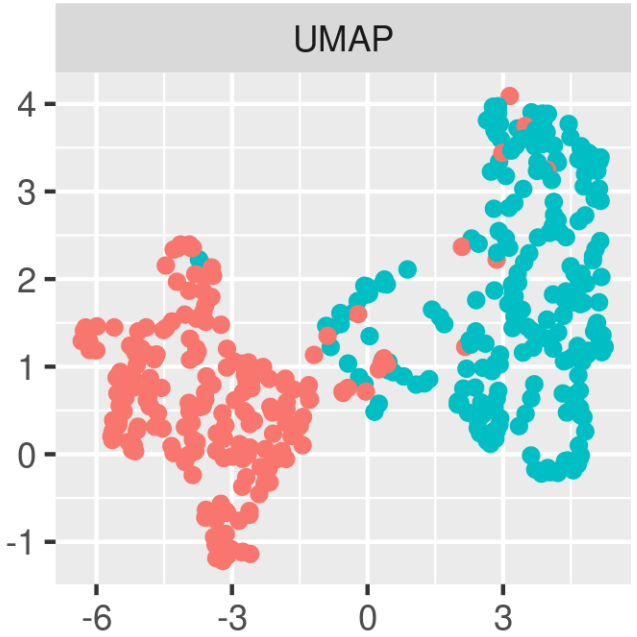
Expression matrix

	geneA	geneB	geneC	geneD	...
cancer1_a	42	58	12	47	...
cancer1_b	25	69	87	42	...
cancer2_a	35	91	25	81	...
cancer2_b	9	71	44	7	...

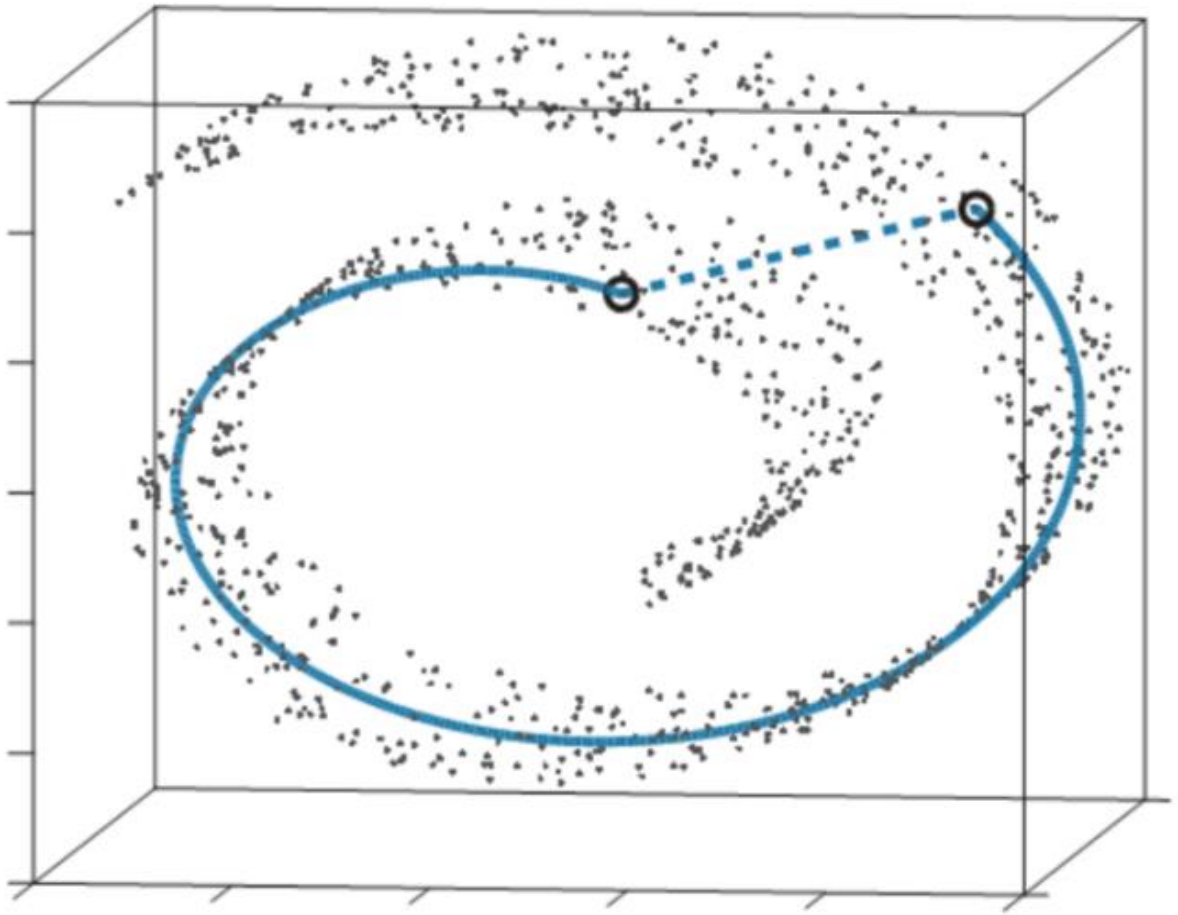
Reduce dimensions to :

- Represent data
- Extract information



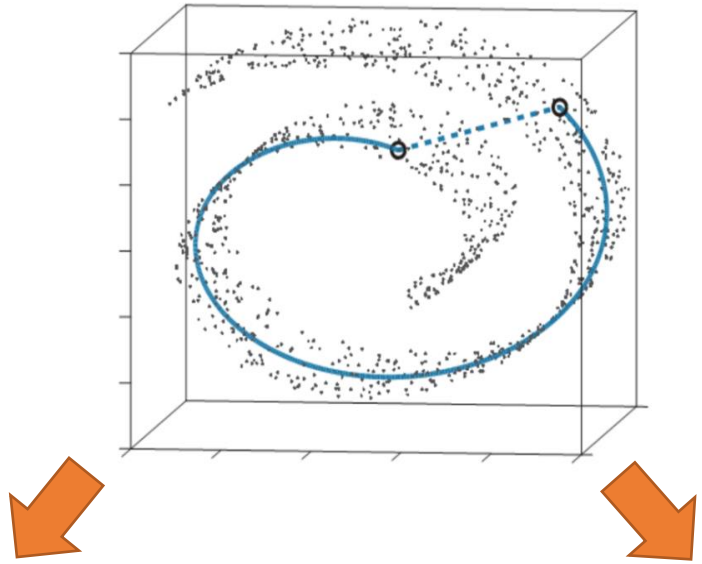


Dimension reduction

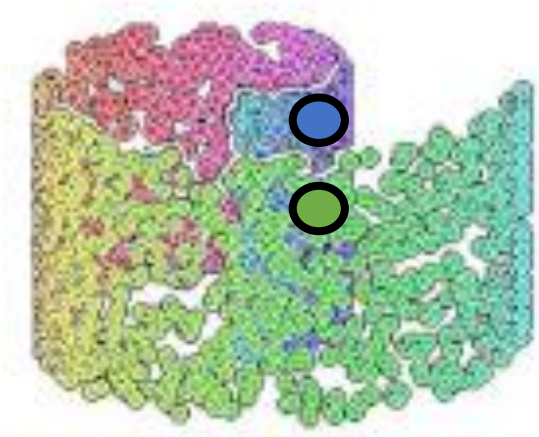


Global and local distances

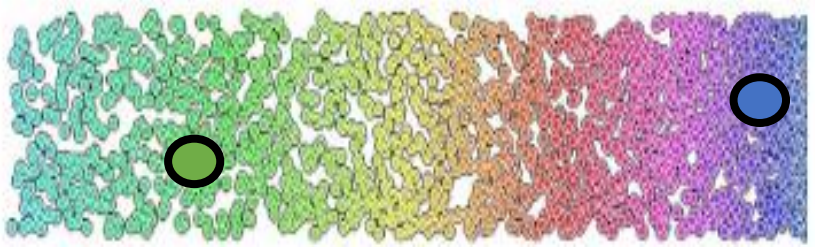
Dimension reduction



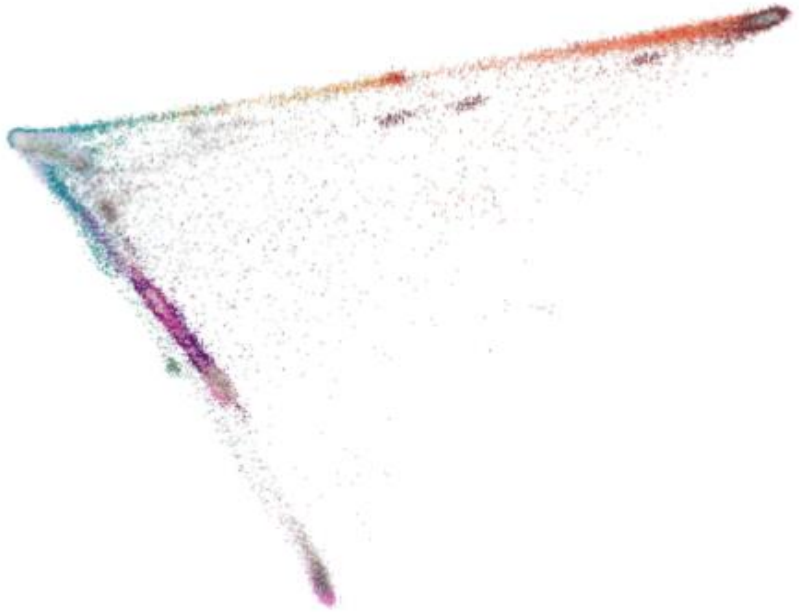
Linear reduction



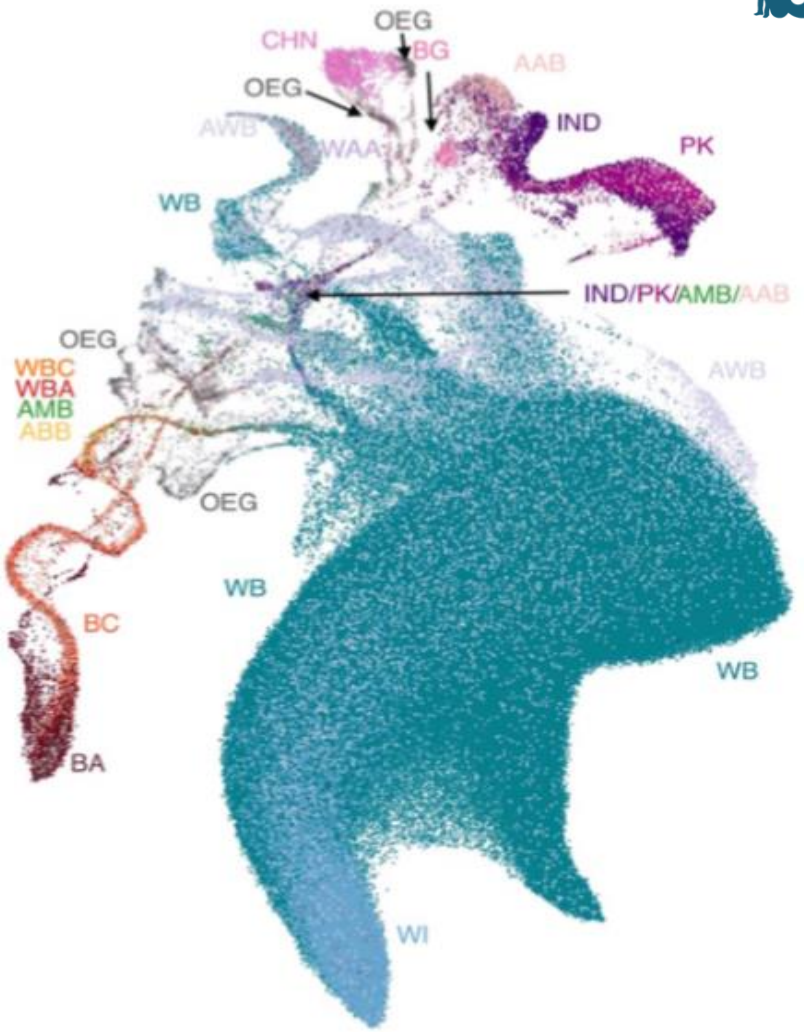
Non-Linear reduction



Dimension reduction



PCA **(a)** Principal components 1 and 2



UMAP **(b)** UMAP on first 10 principal components

Alex Diaz-Papkovich *et al., 2018*

Dimension reduction

Loss of information
Does not deform space

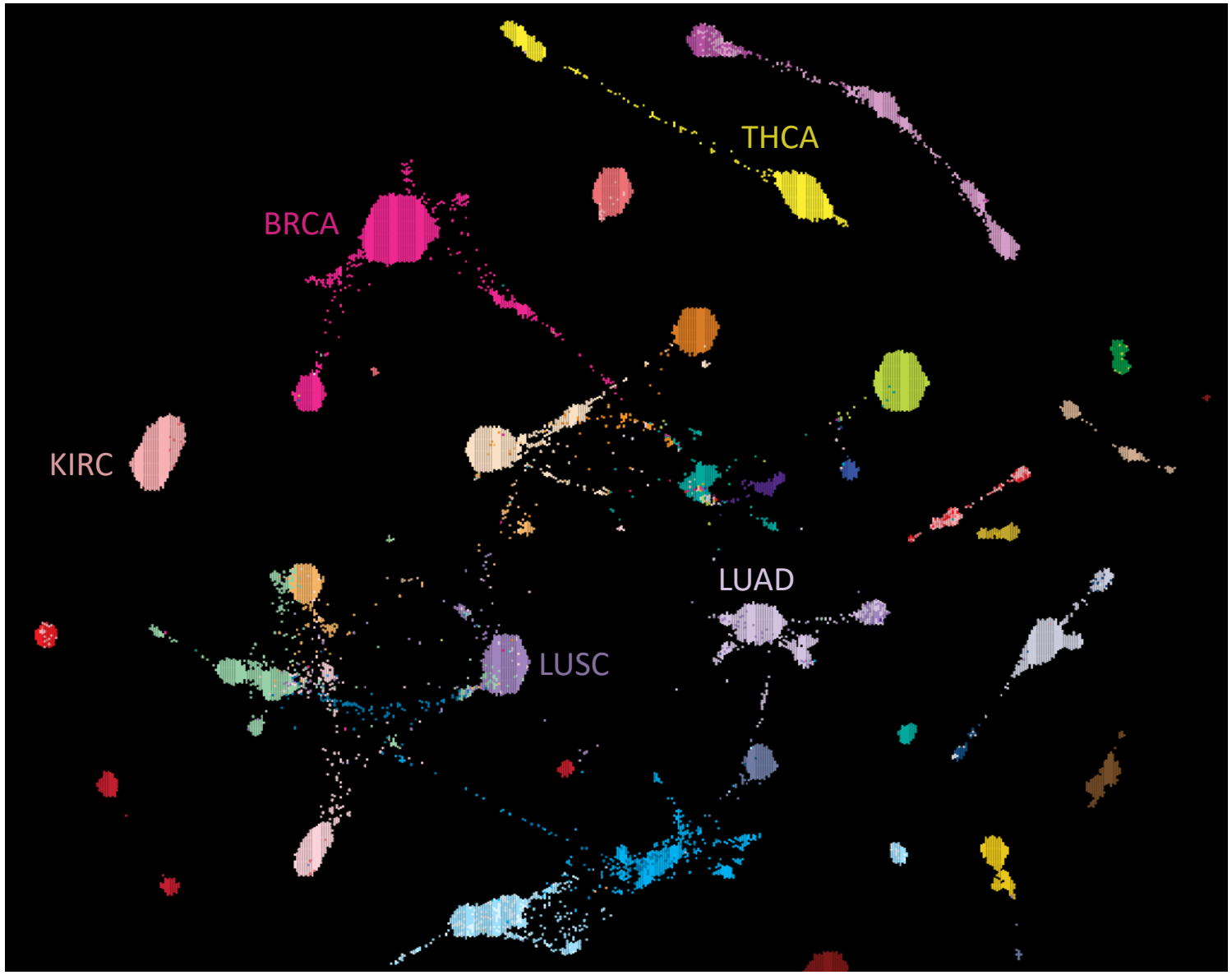
PCA (a) Principal components 1 and 2

Keeps information
Deforms space

UMAP (b) UMAP on first 10 principal components

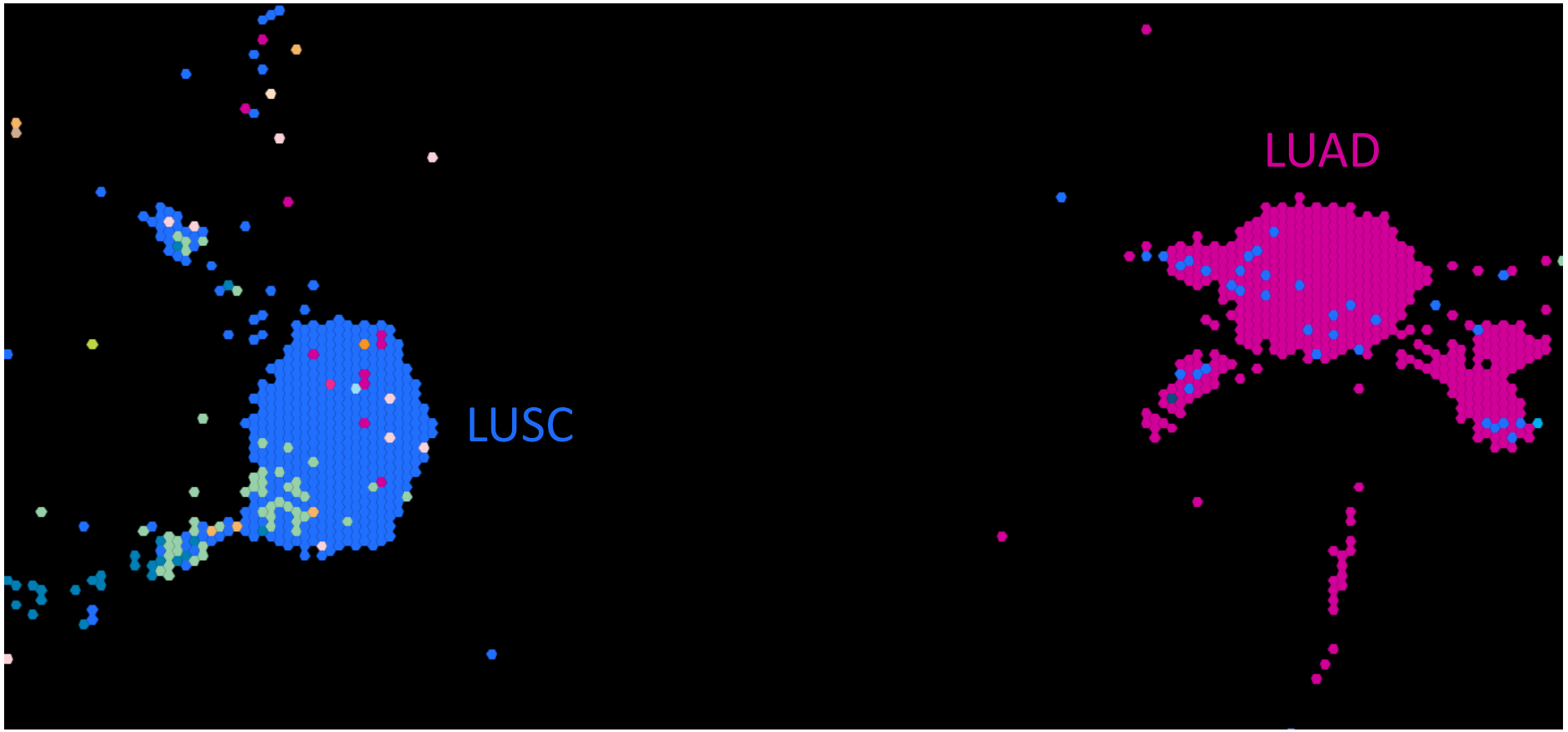
Alex Diaz-Papkovich *et al., 2018*

Dimension reduction

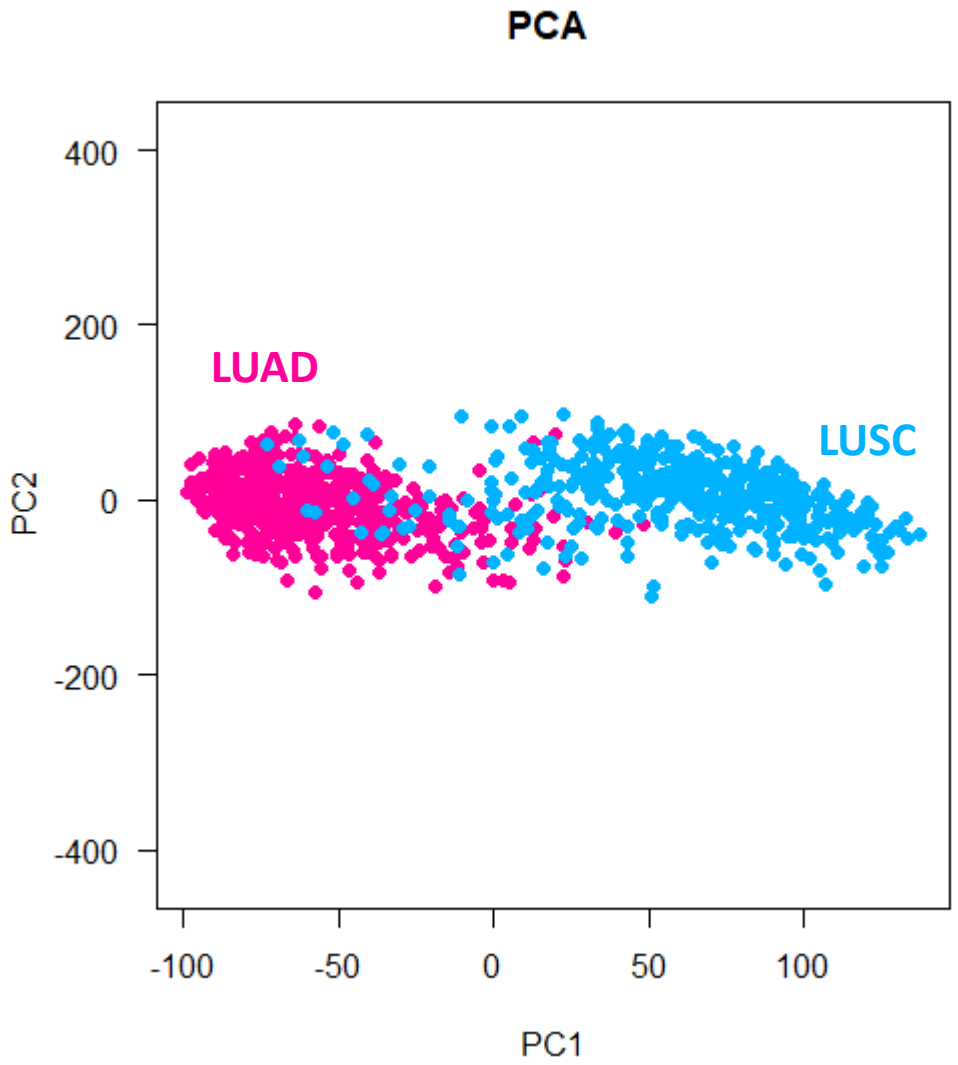


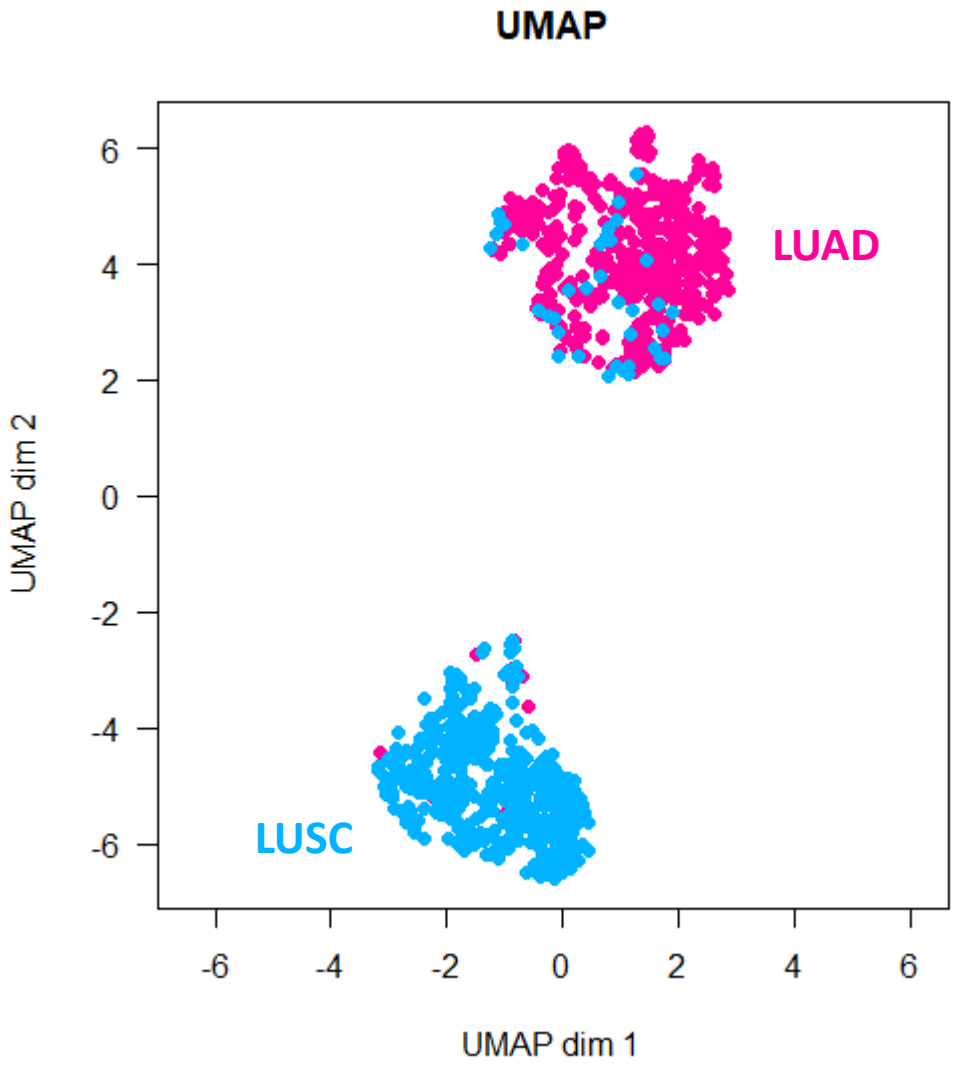
Katherine A. Hoedley *et al.*,2018

Dimension reduction



Dimension reduction

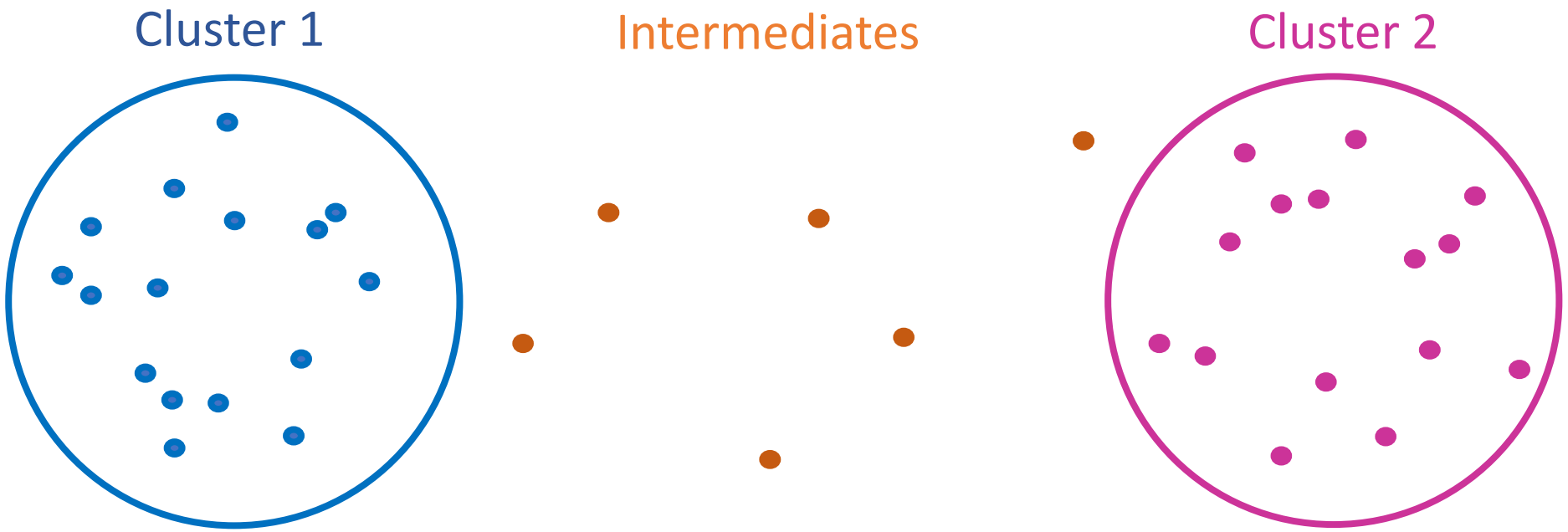




Objective :

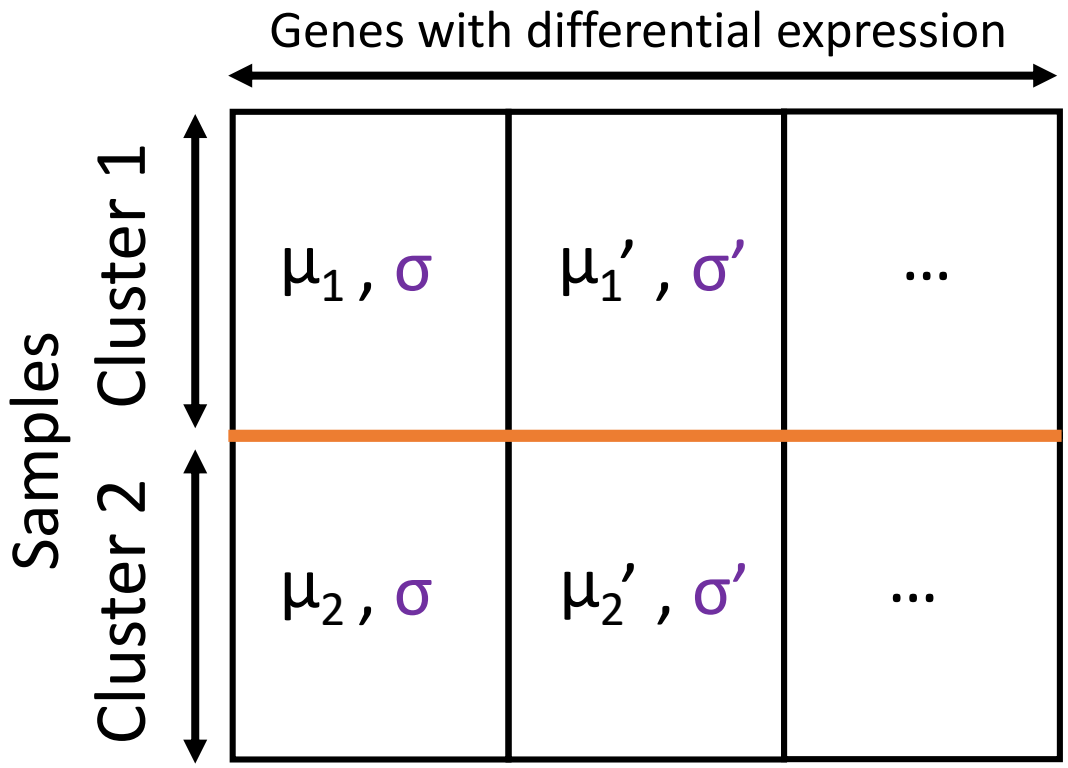
Assess dimensionality reduction (DR) methods for the detection of intermediate cancer phenotypes

Simulation approach

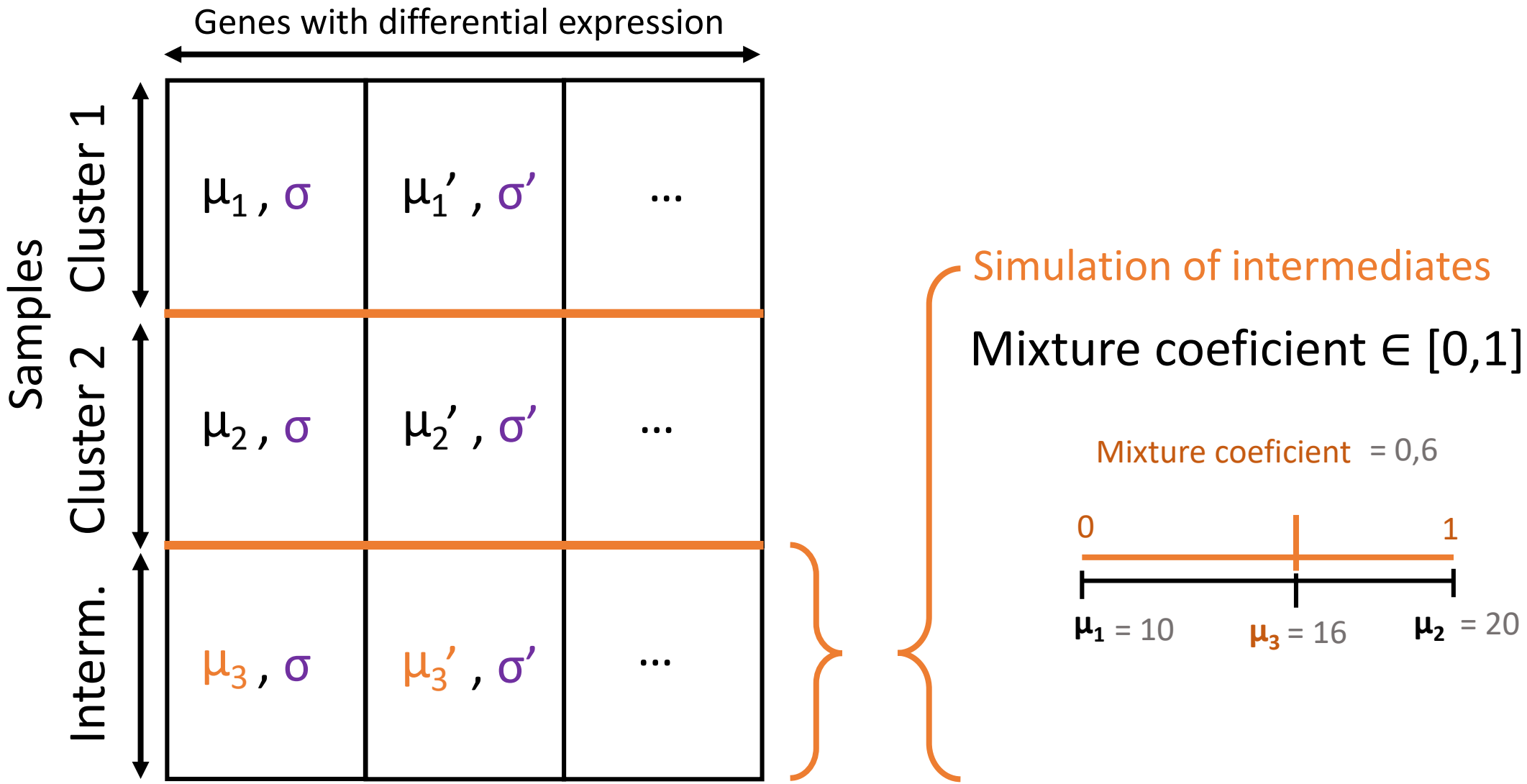


- Know the reality of the data
- Modify data parameters
- No limit in number of replicates

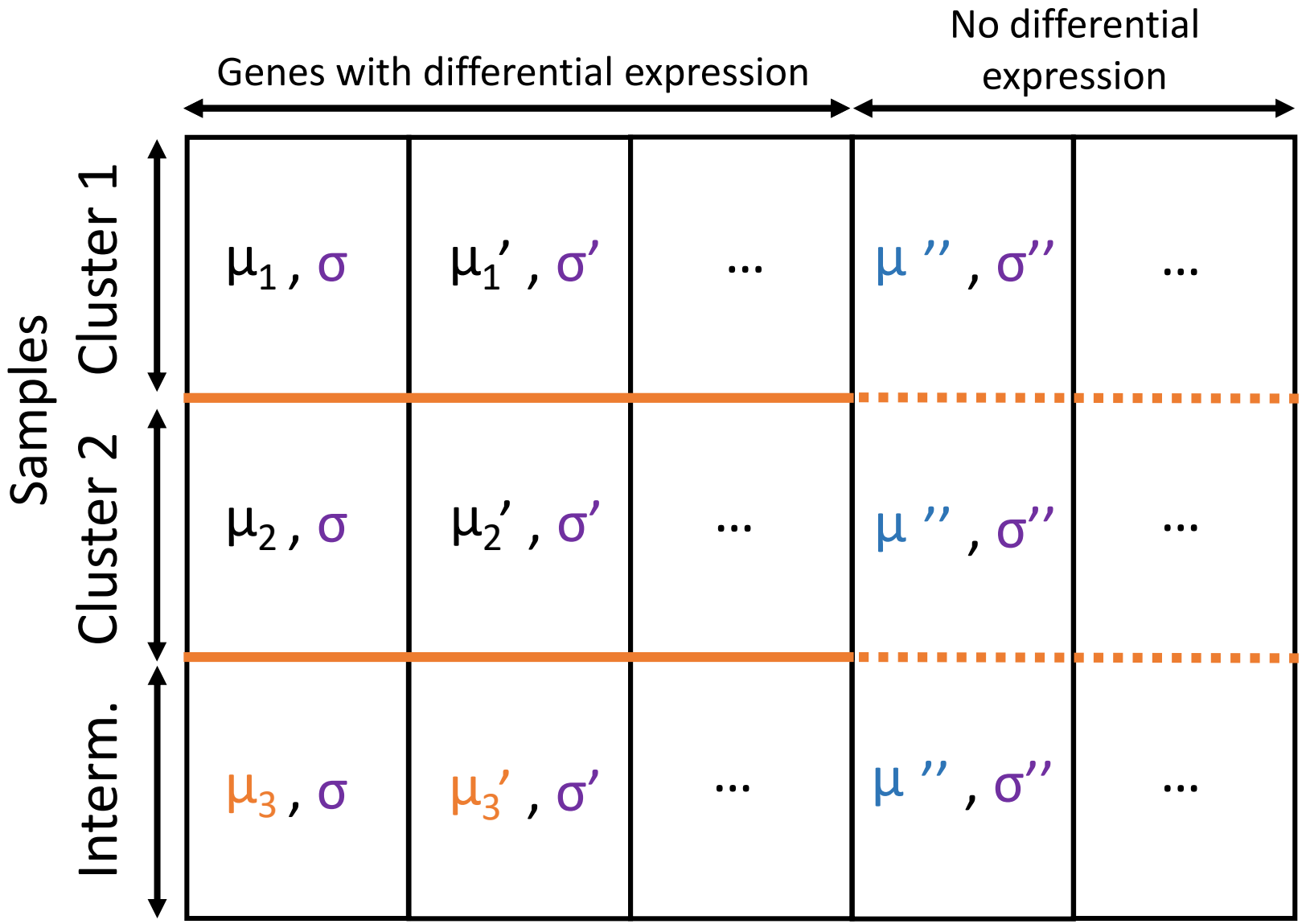
Model : Building an expression matrix



Model : Building an expression matrix

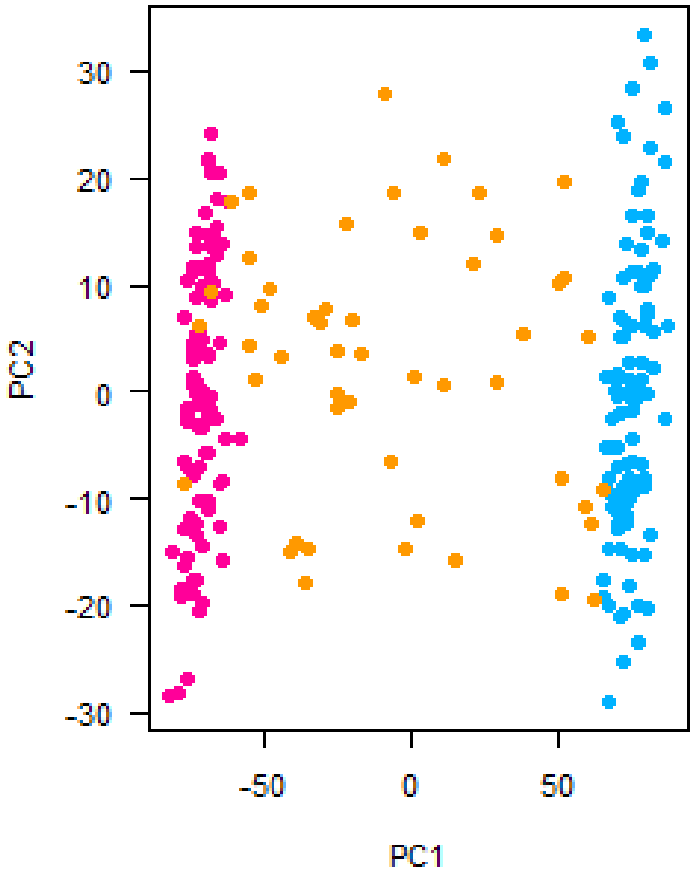


Model : Building an expression matrix

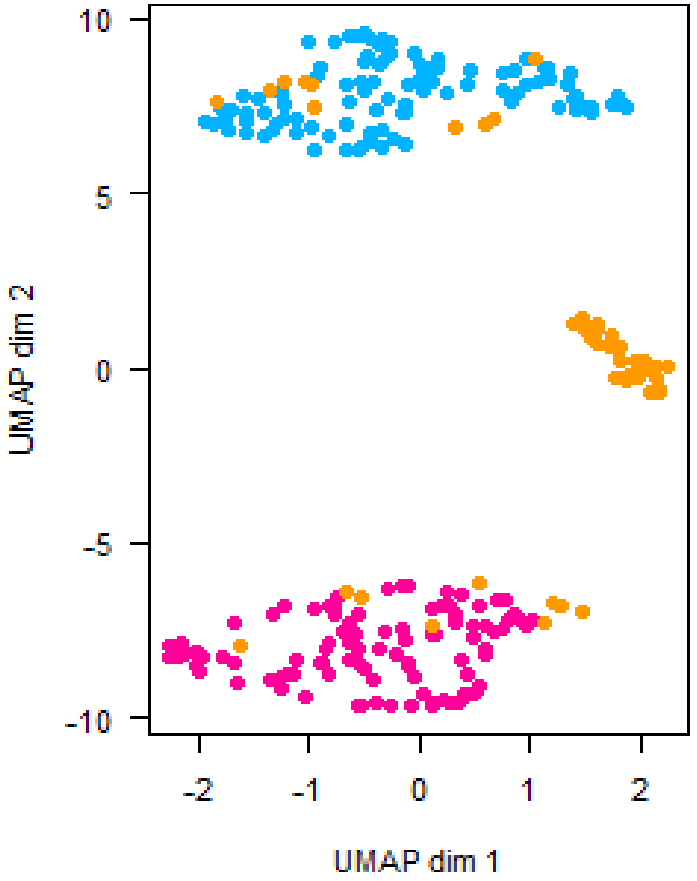


Simulation example

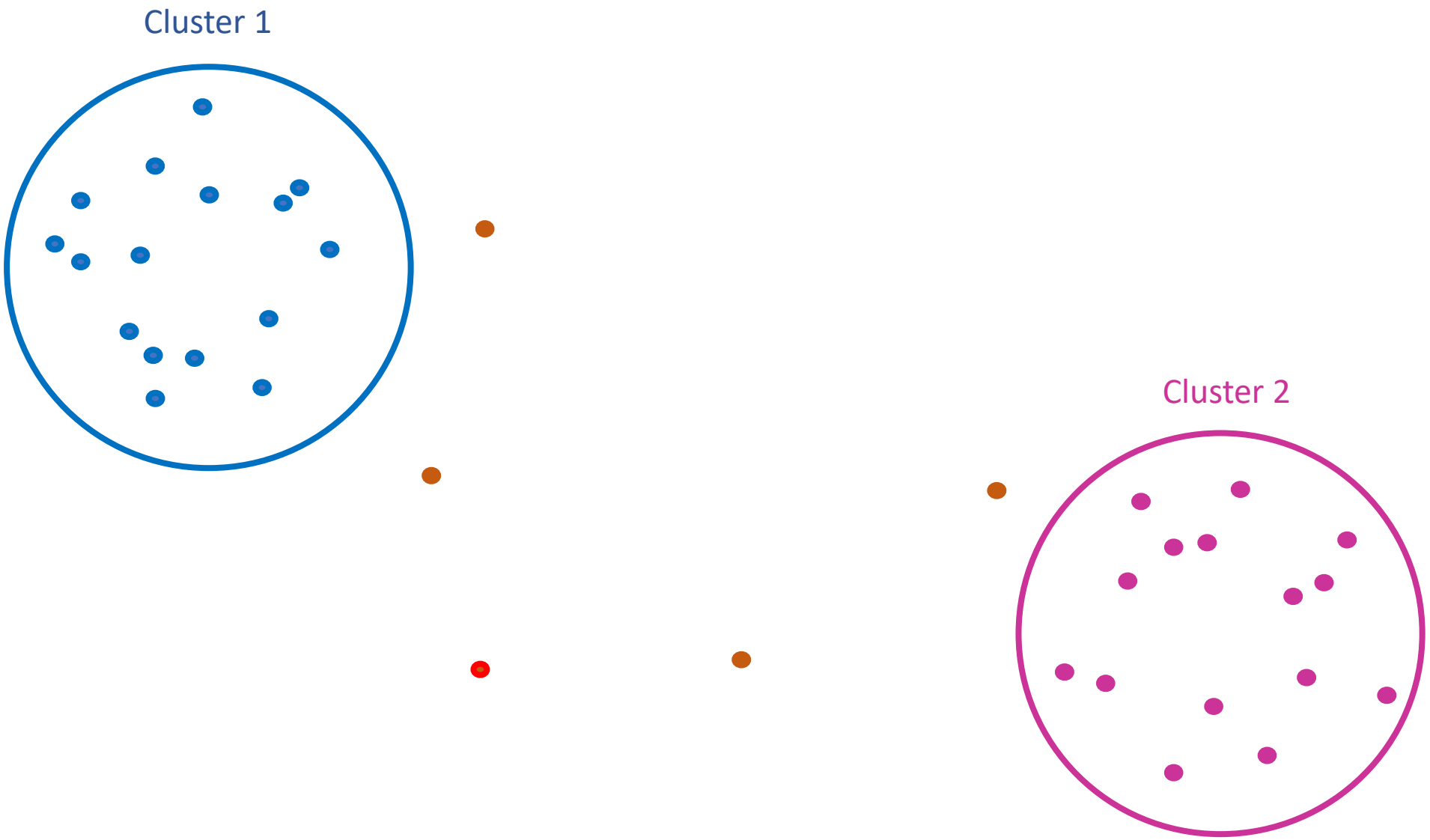
PCA



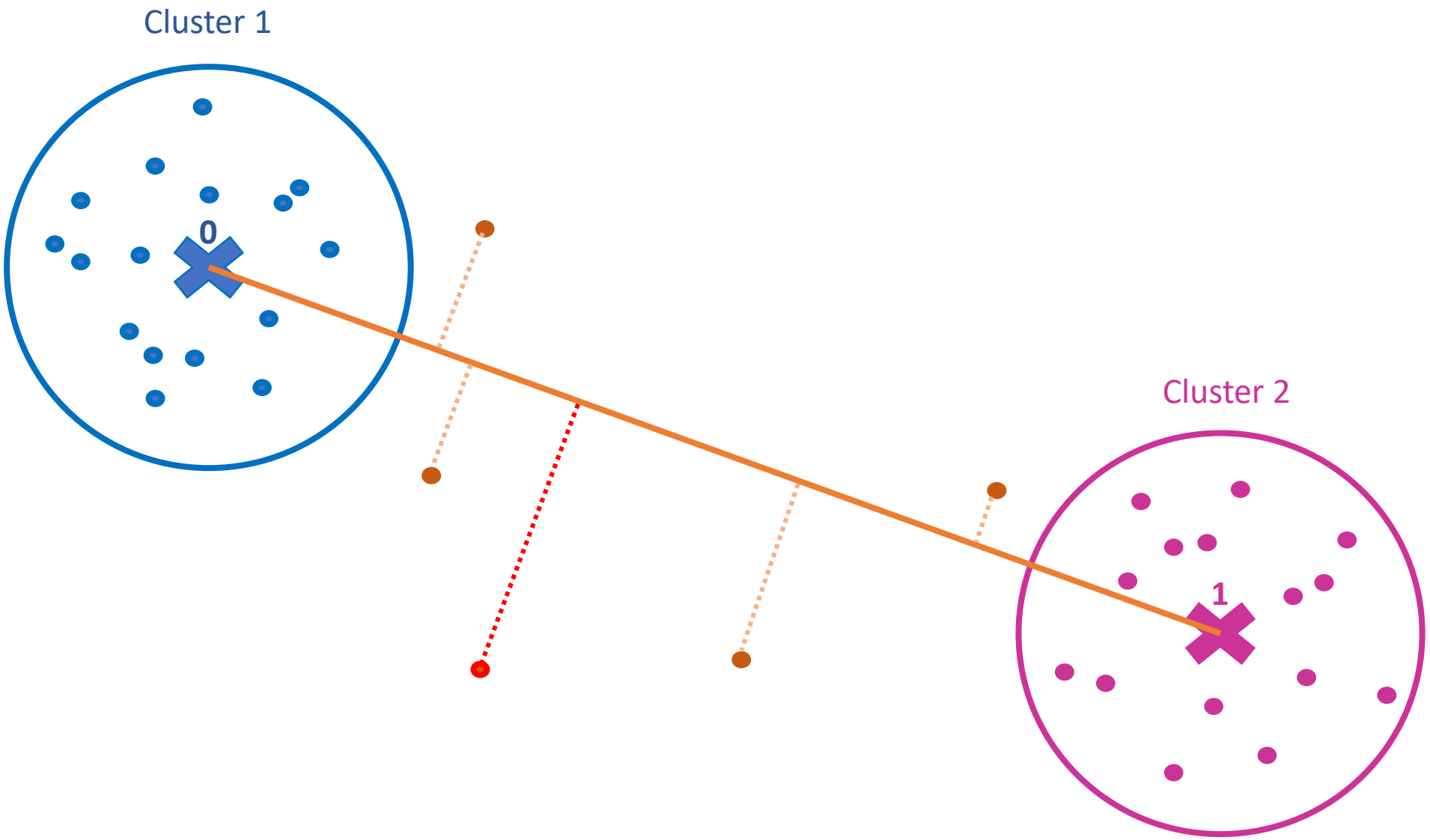
UMAP



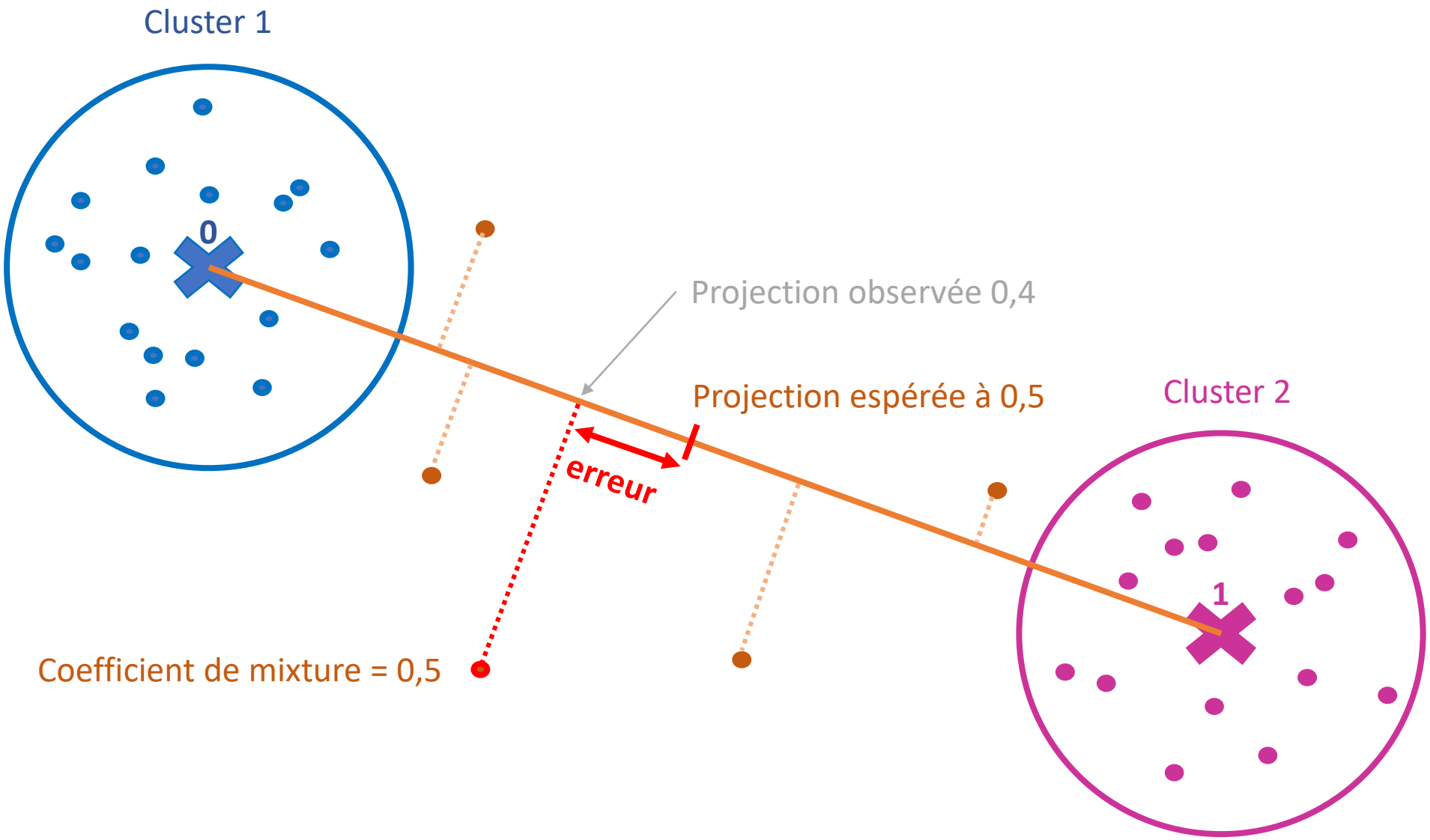
Evaluation : Projection of intermediates



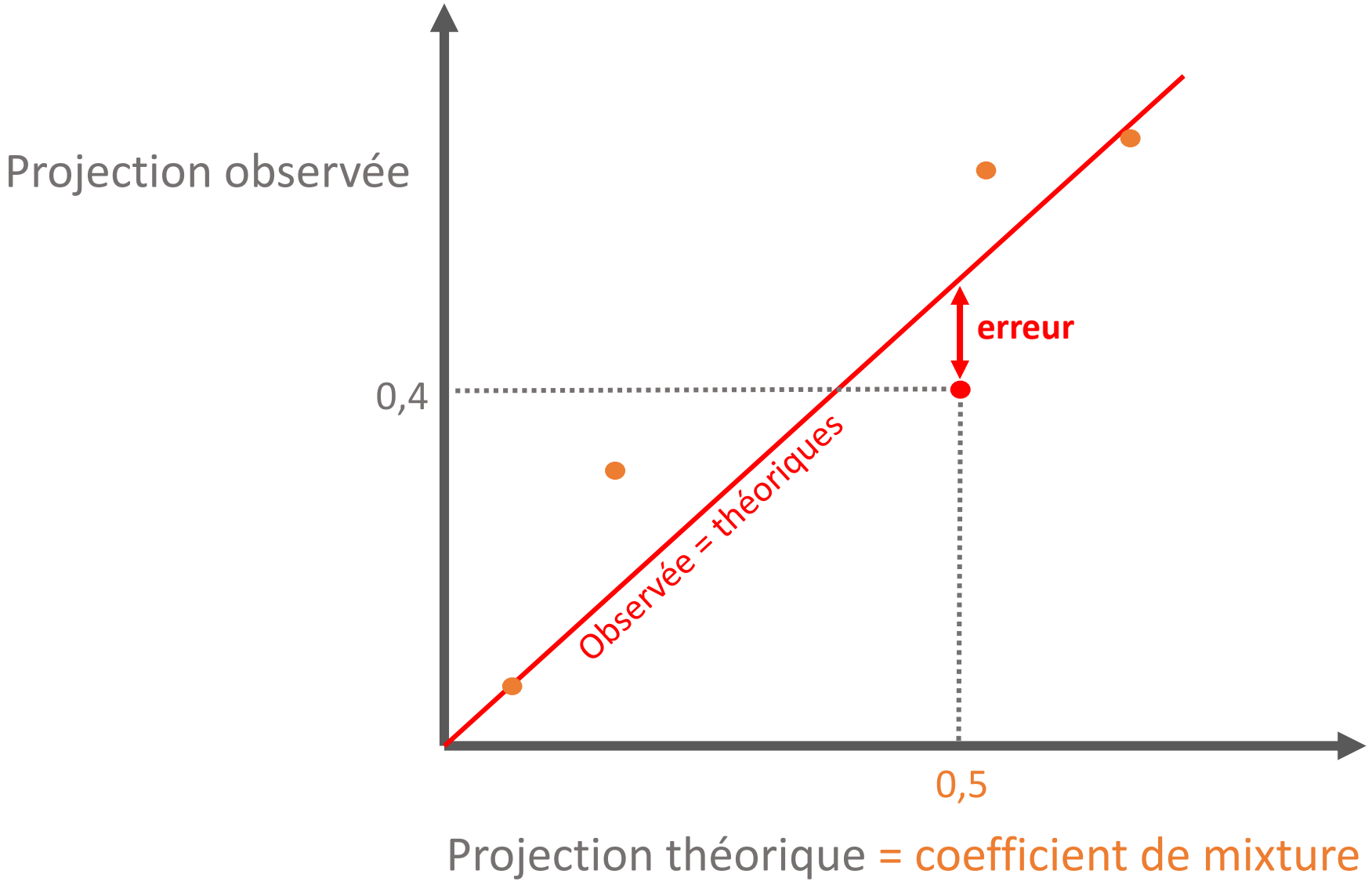
Evaluation : Projection of intermediates



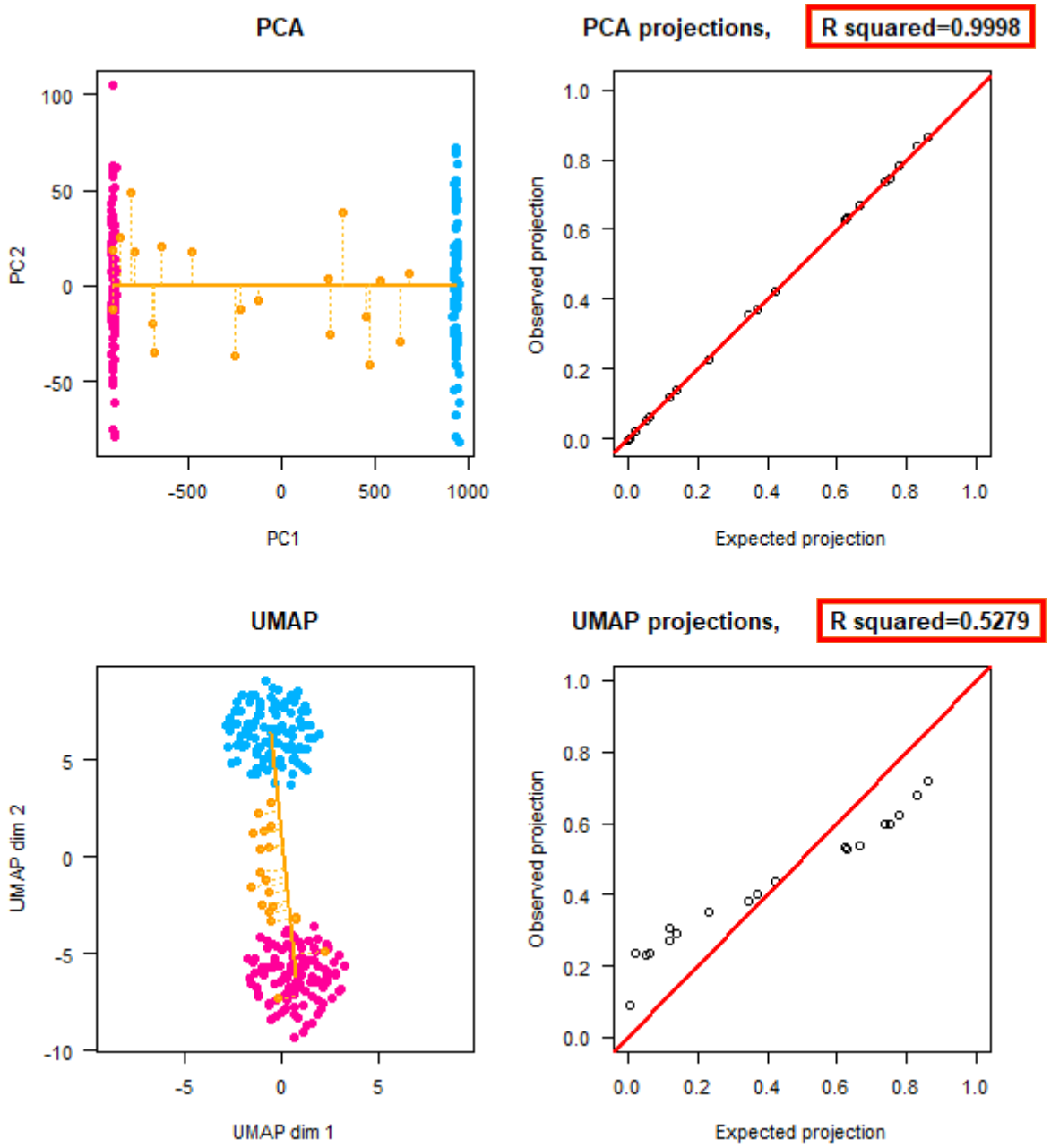
Evaluation : Projection of intermediates



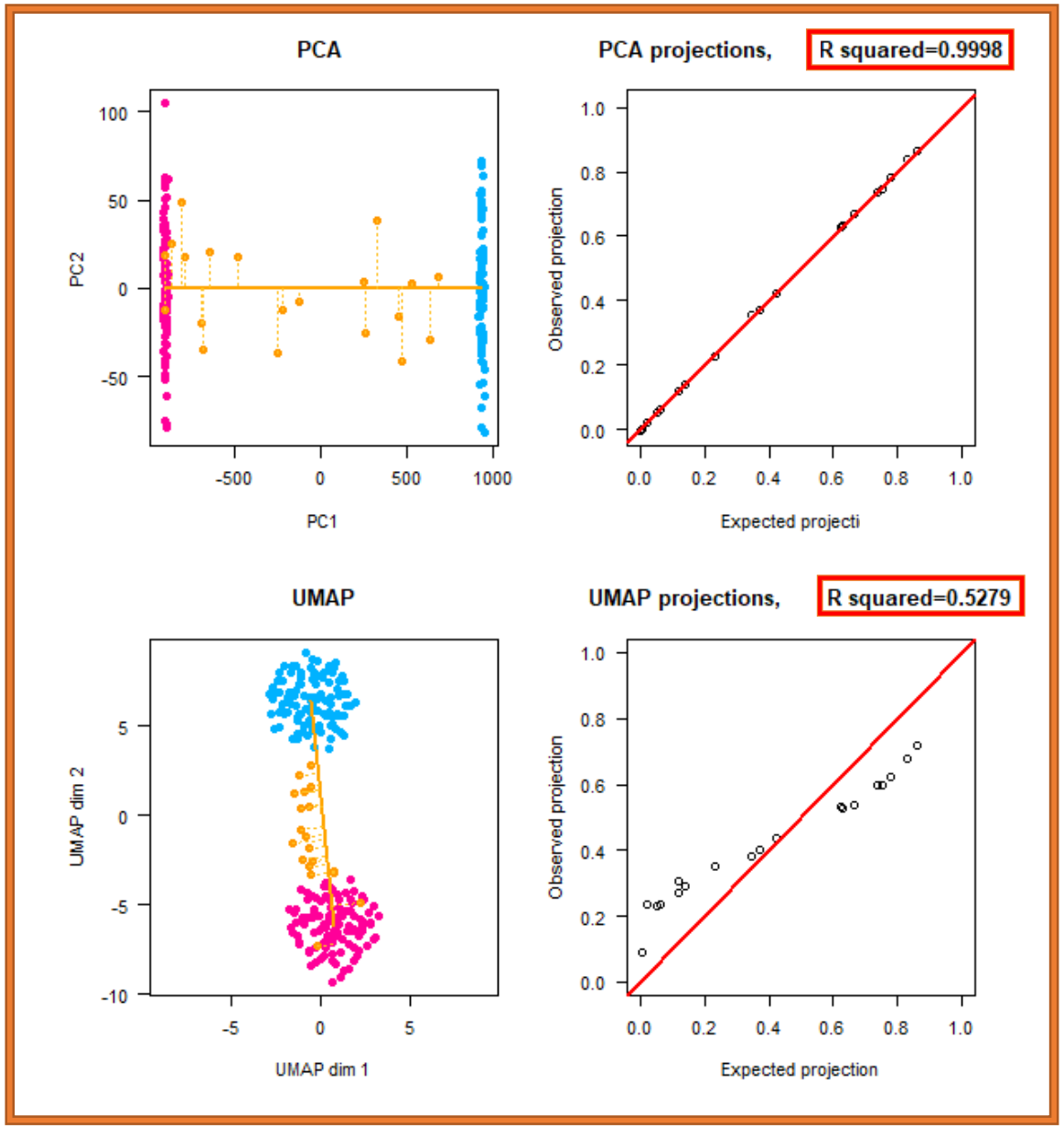
Evaluation : Projection of intermediates



Evaluation of dimension reduction (DR) methods

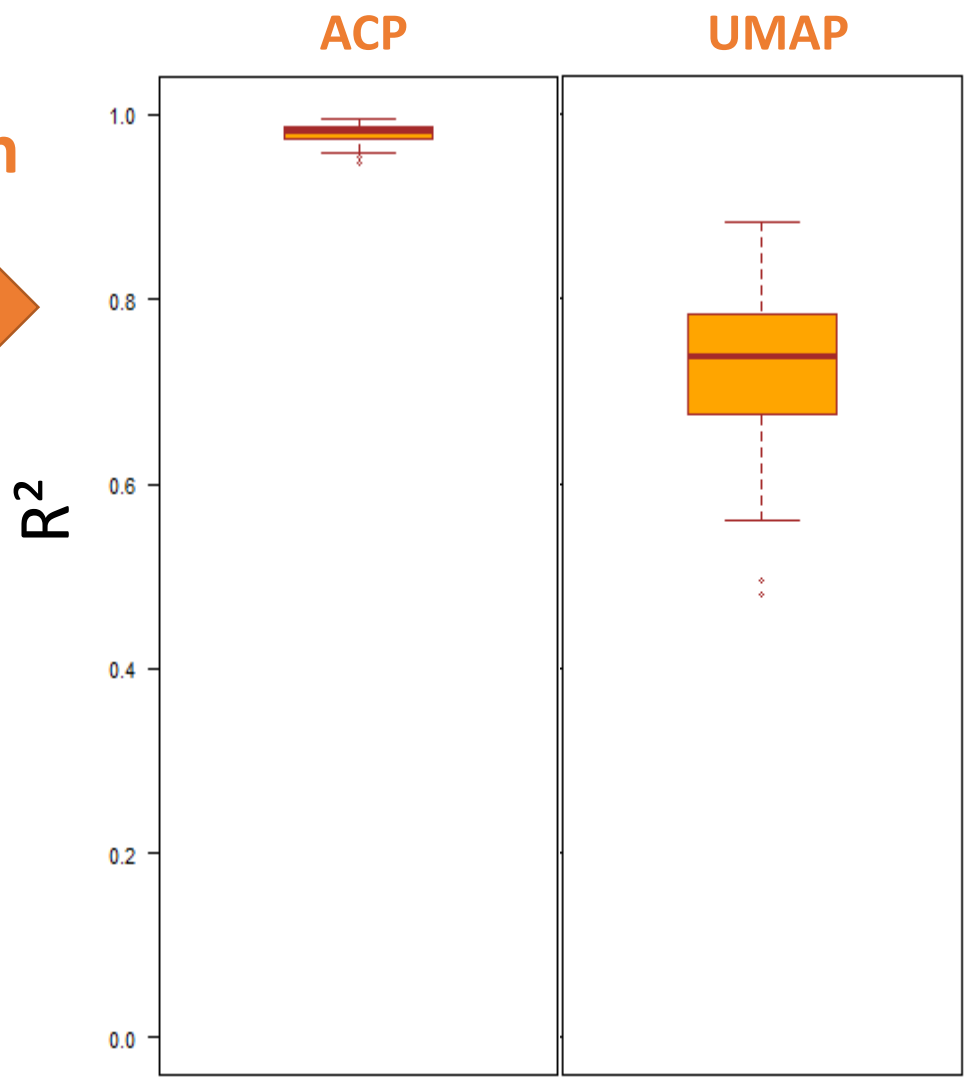


Evaluation of dimension reduction (DR) methods



Simulation

X 100



Exploration of parameters

Data dependent parameters :

- Number of samples (Data size)
- Proportion of intermediates
- Distance between clusters (δ)

DR parameters :

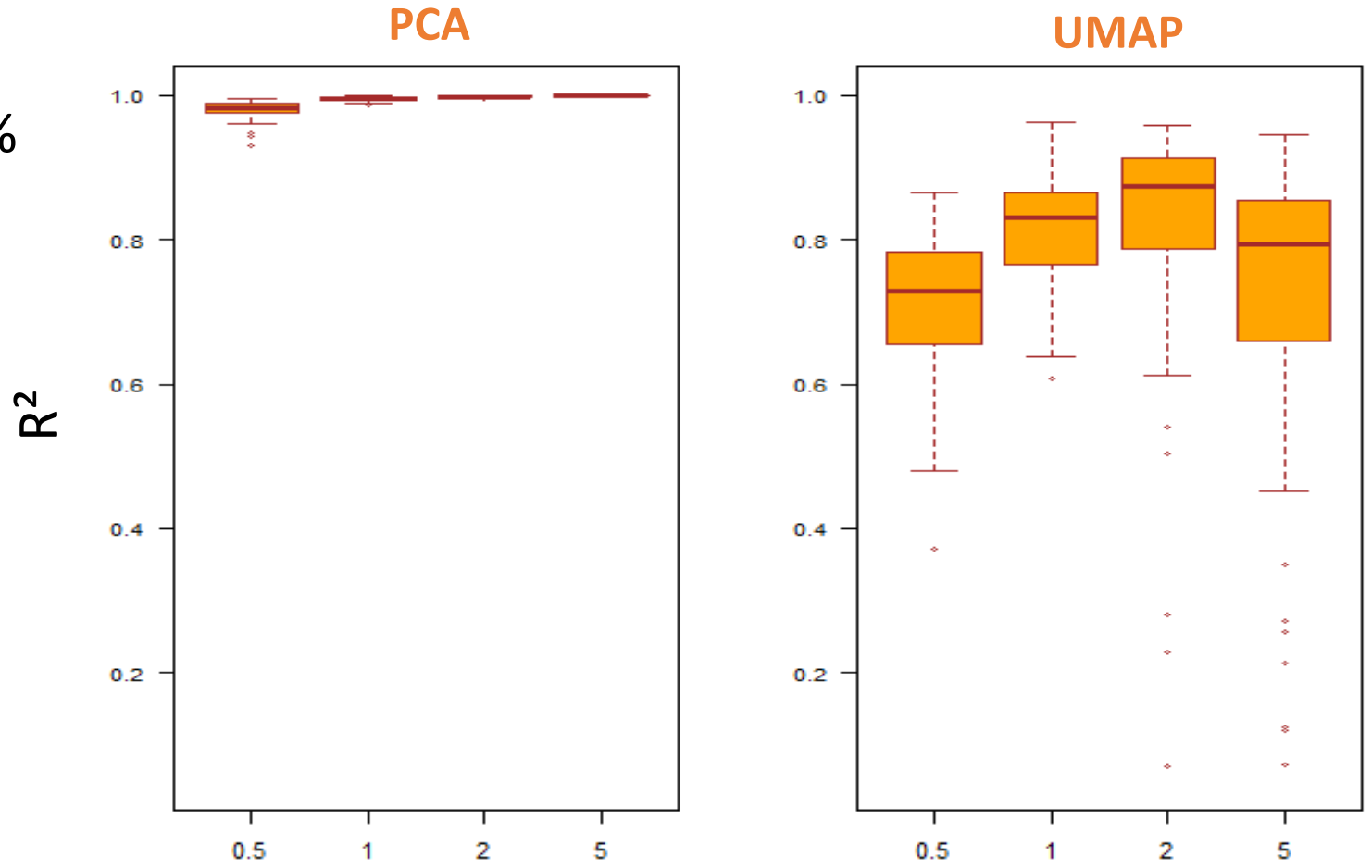
- UMAP Min_dist
- UMAP N_neighbours

Exploration of parameters

Proportion of intermediates : 10%

UMAP min_dist : 0.75

Data size : 200

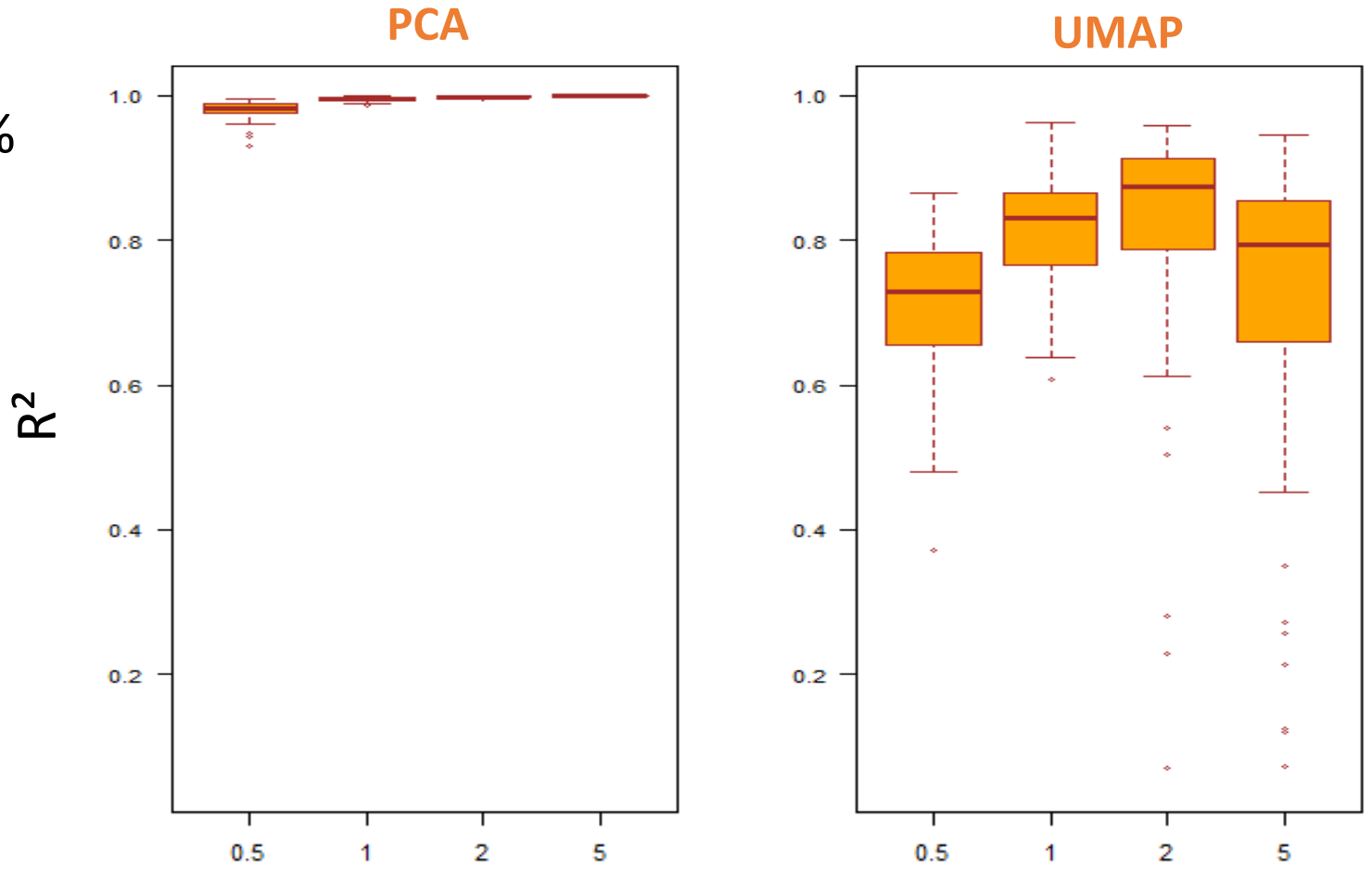


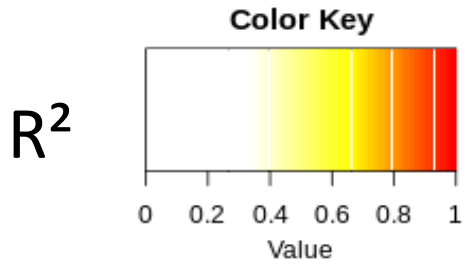
Exploration of parameters

Proportion of intermediates : 10%

UMAP min_dist : 0.75

Data size : 200



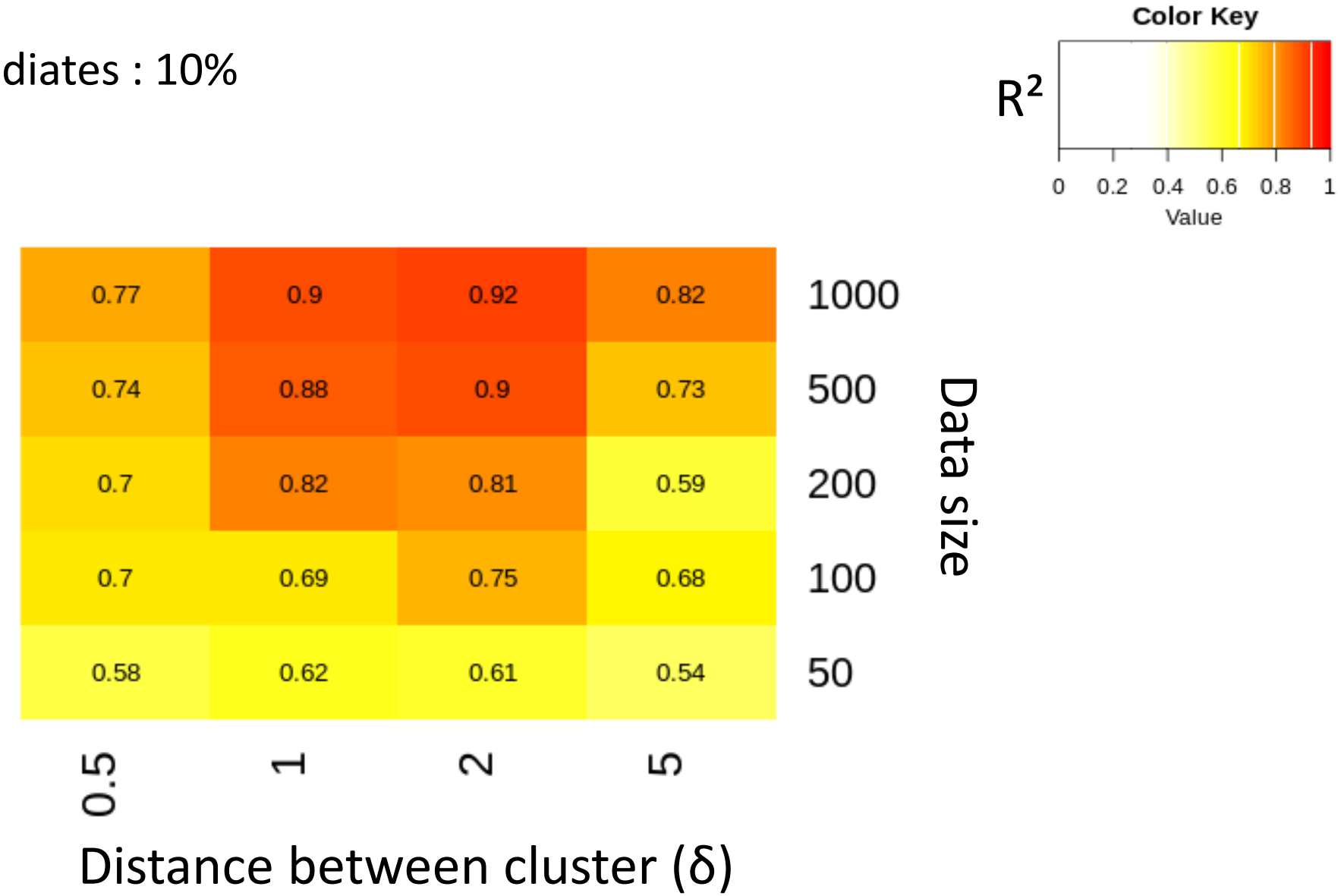


Distance between clusters (δ)

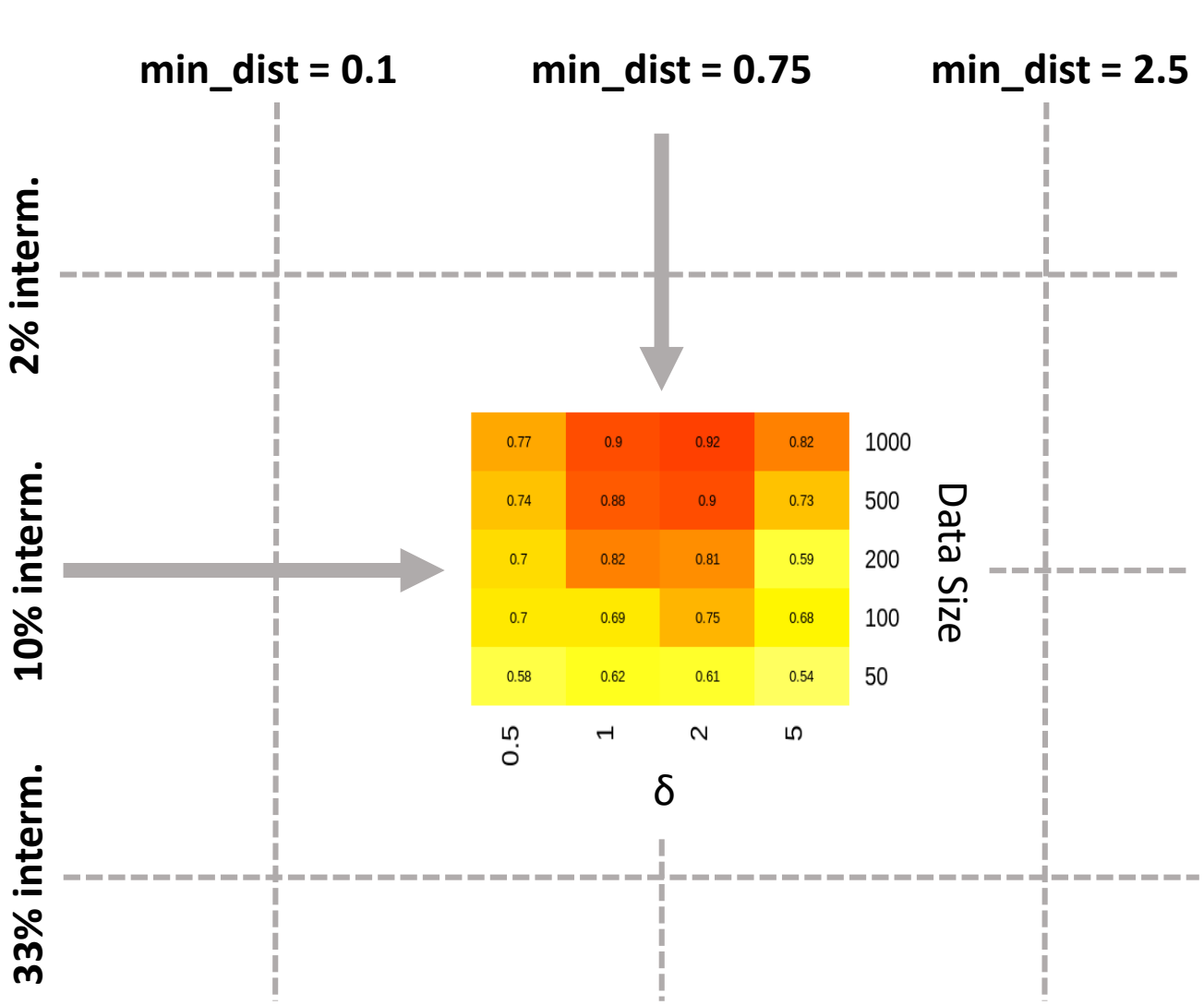
Exploration of parameters

Proportion of intermediates : 10%

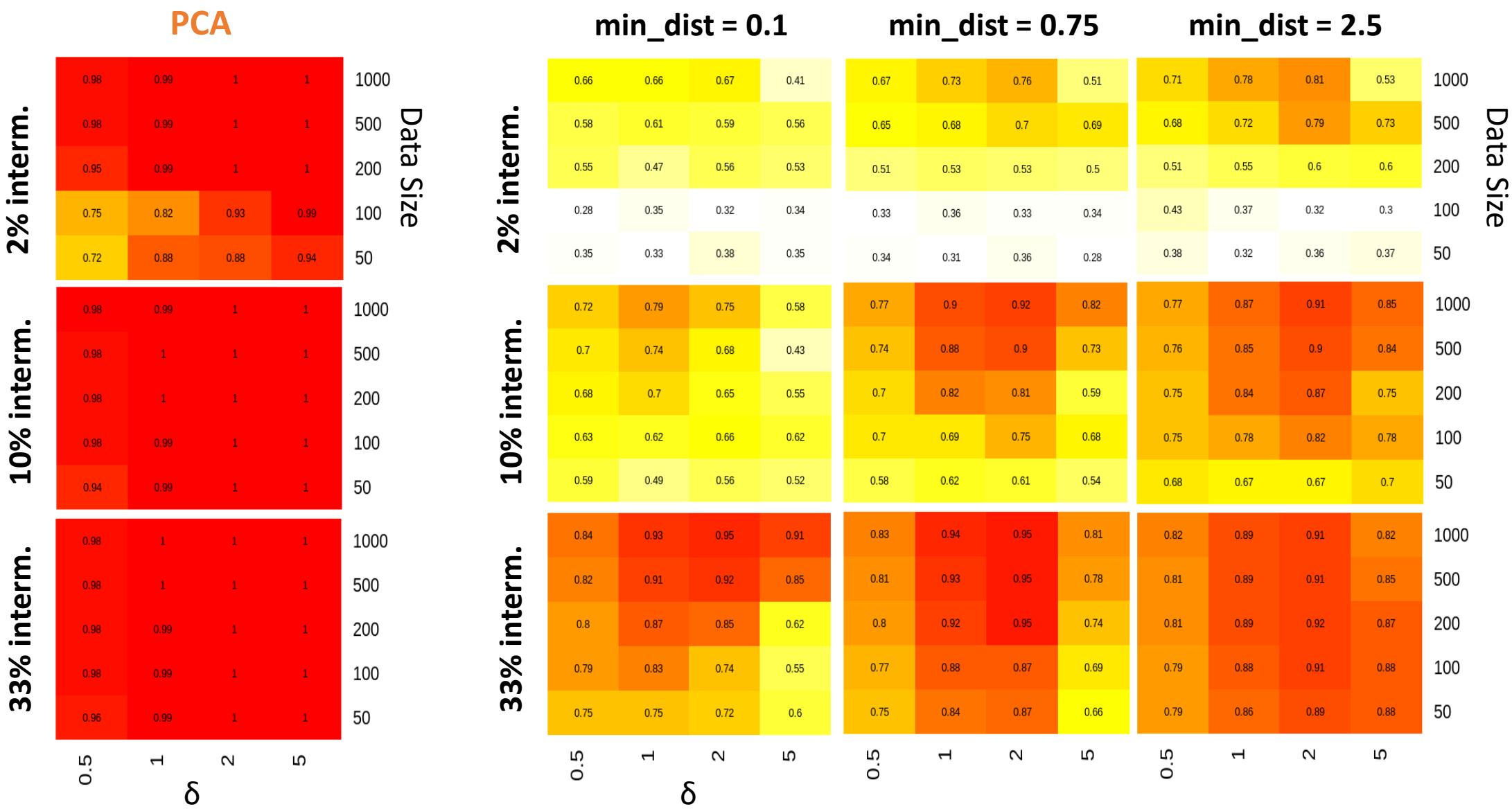
UMAP min_dist : 0,75



Exploration of parameters

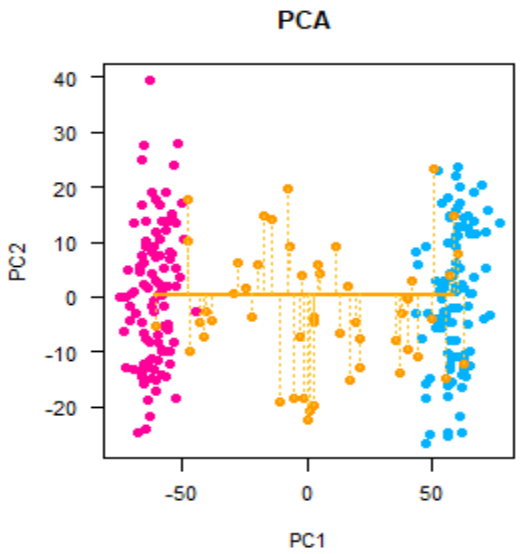


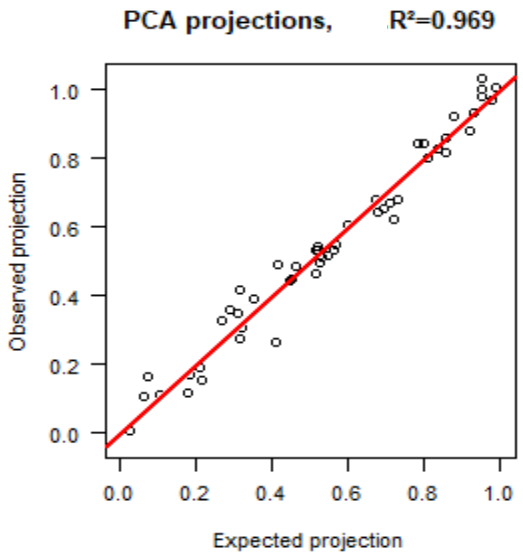
Results:

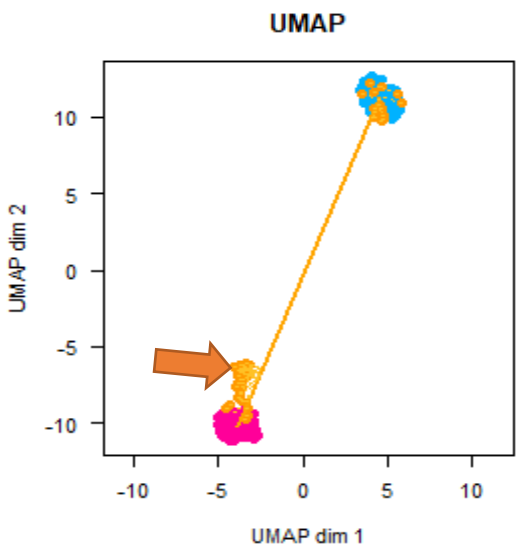


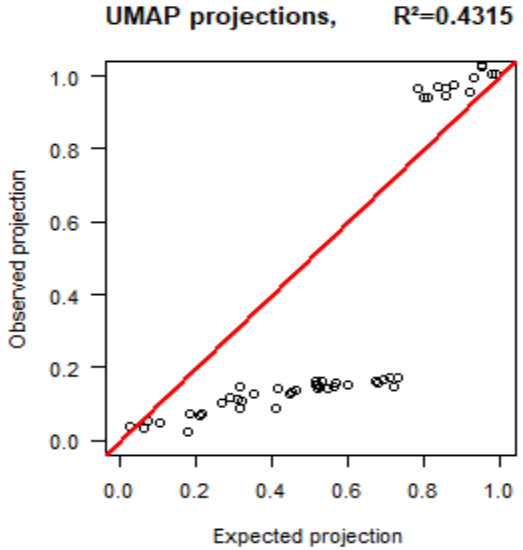
Exploration of parameters and evaluation of DR methods

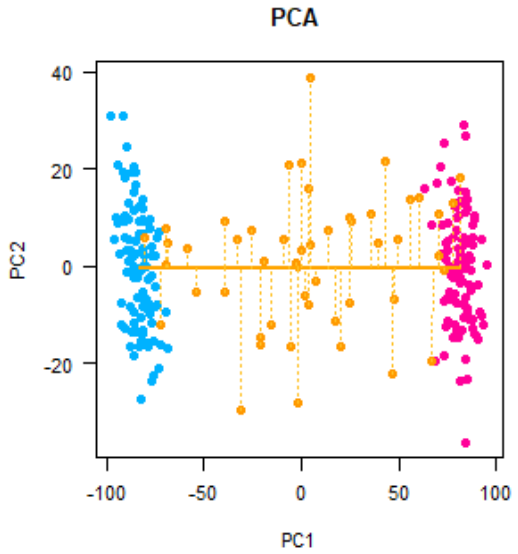
UMAP behavior

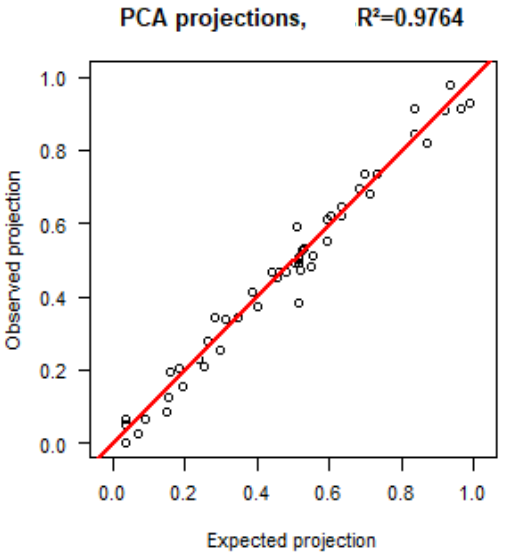


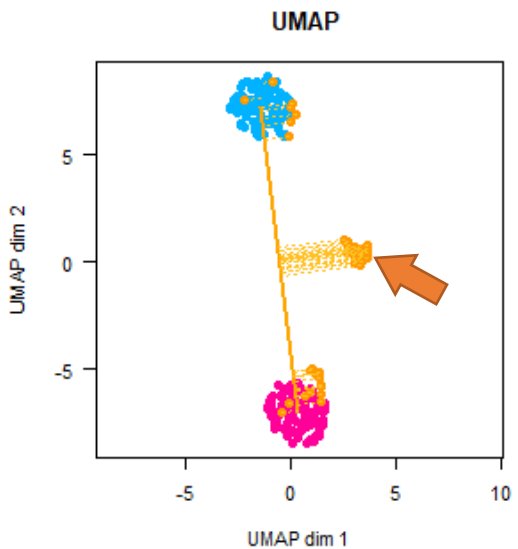


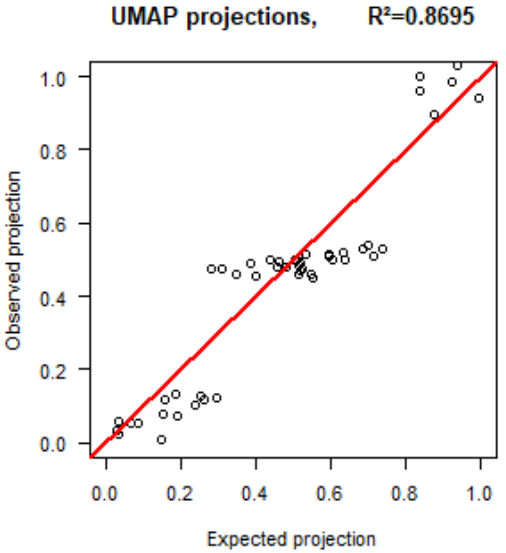






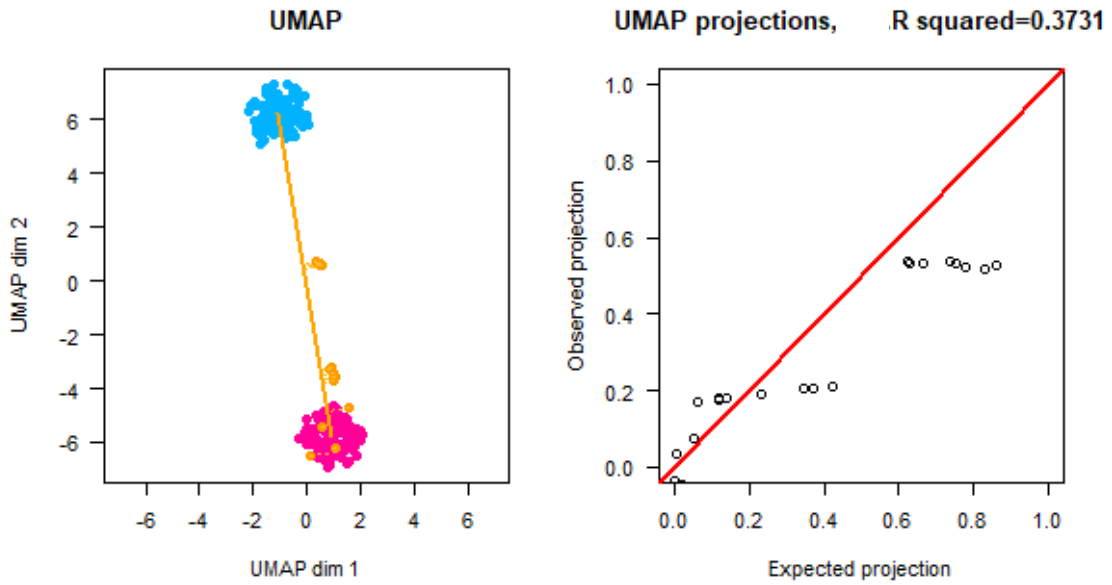




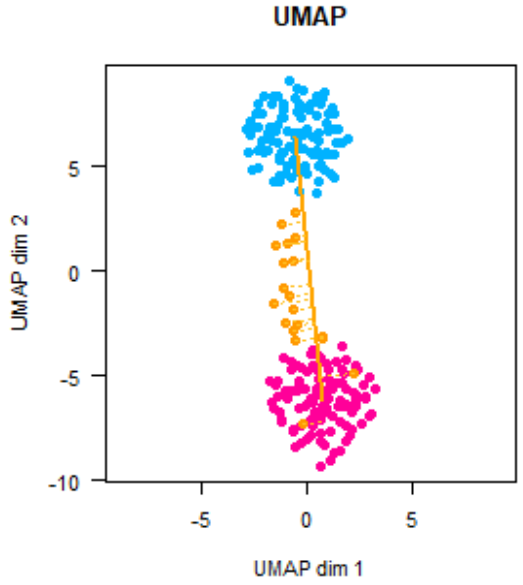


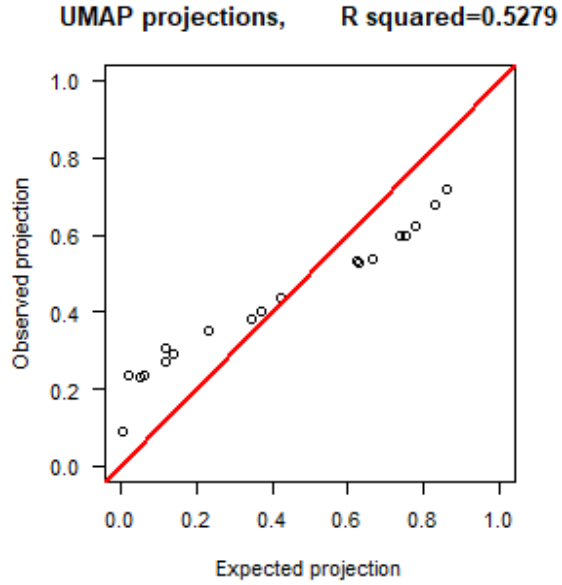
UMAP min_dist

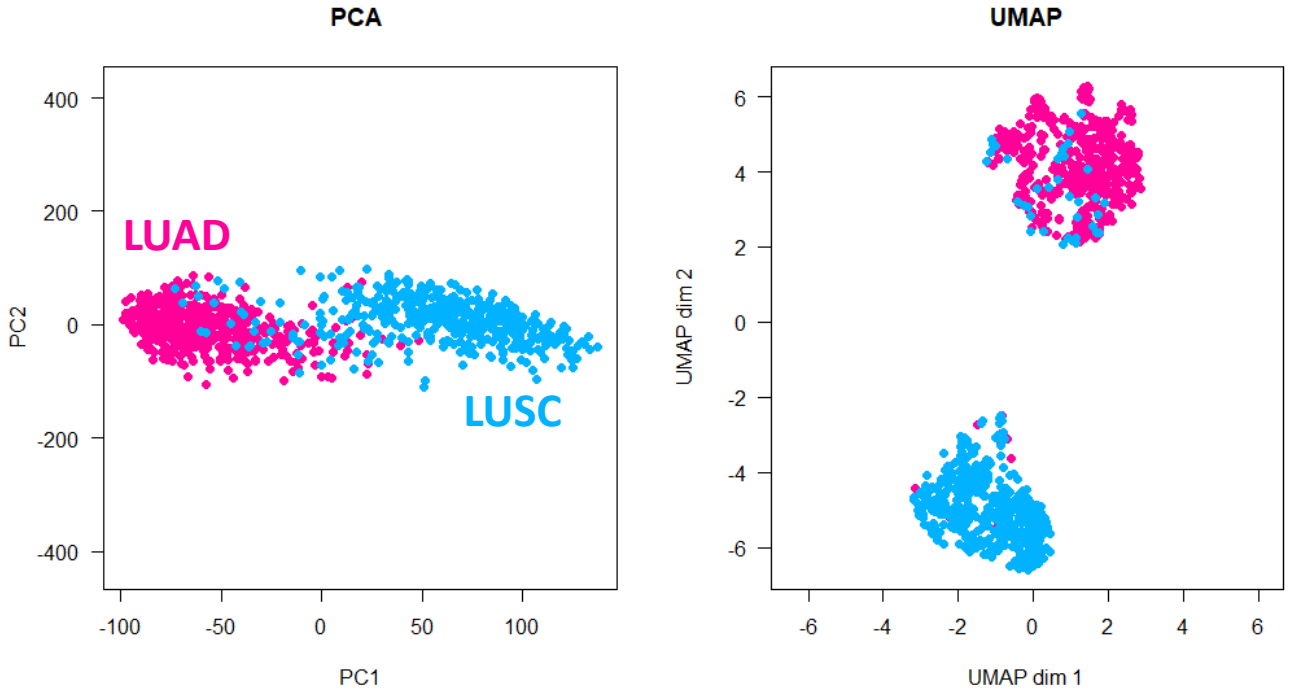
UMAP min_dist = 0



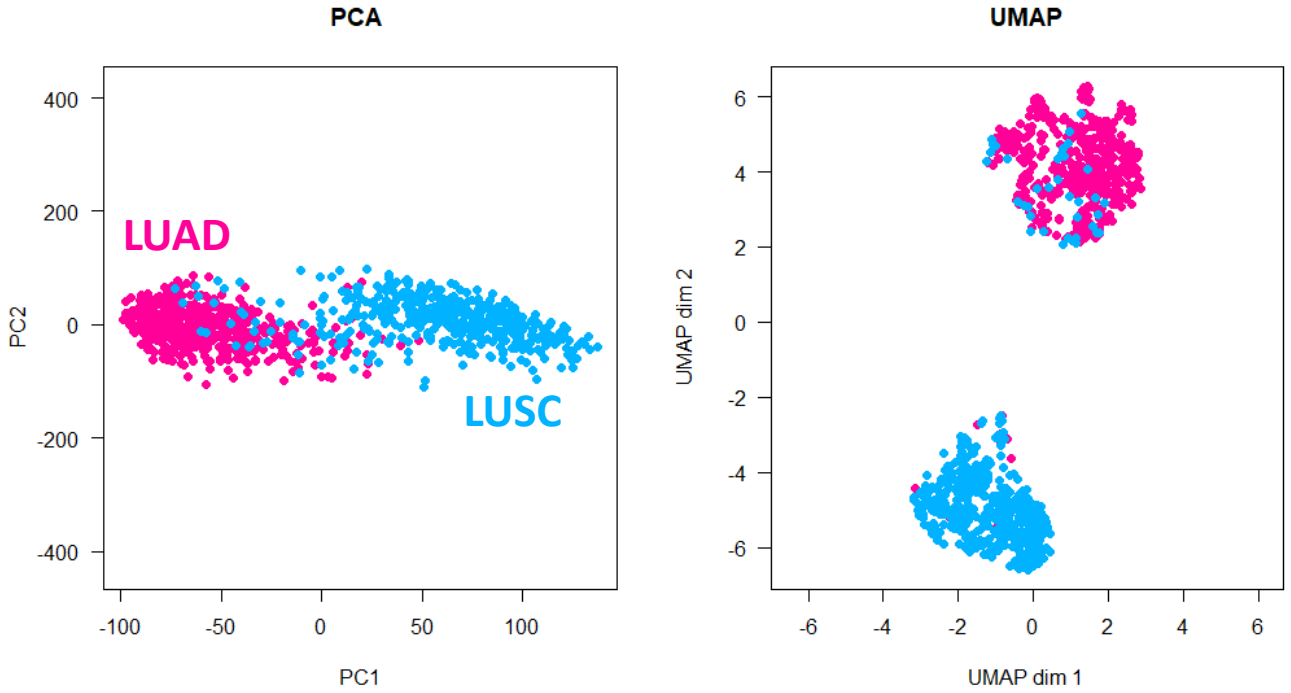
UMAP min_dist = 1



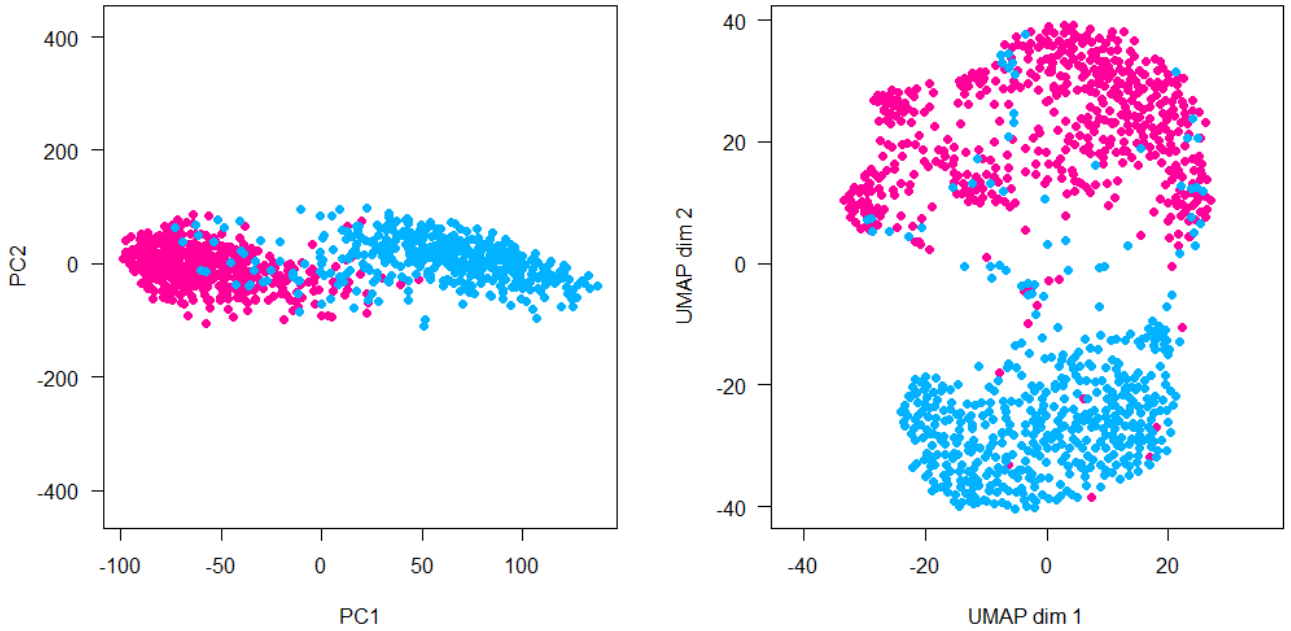




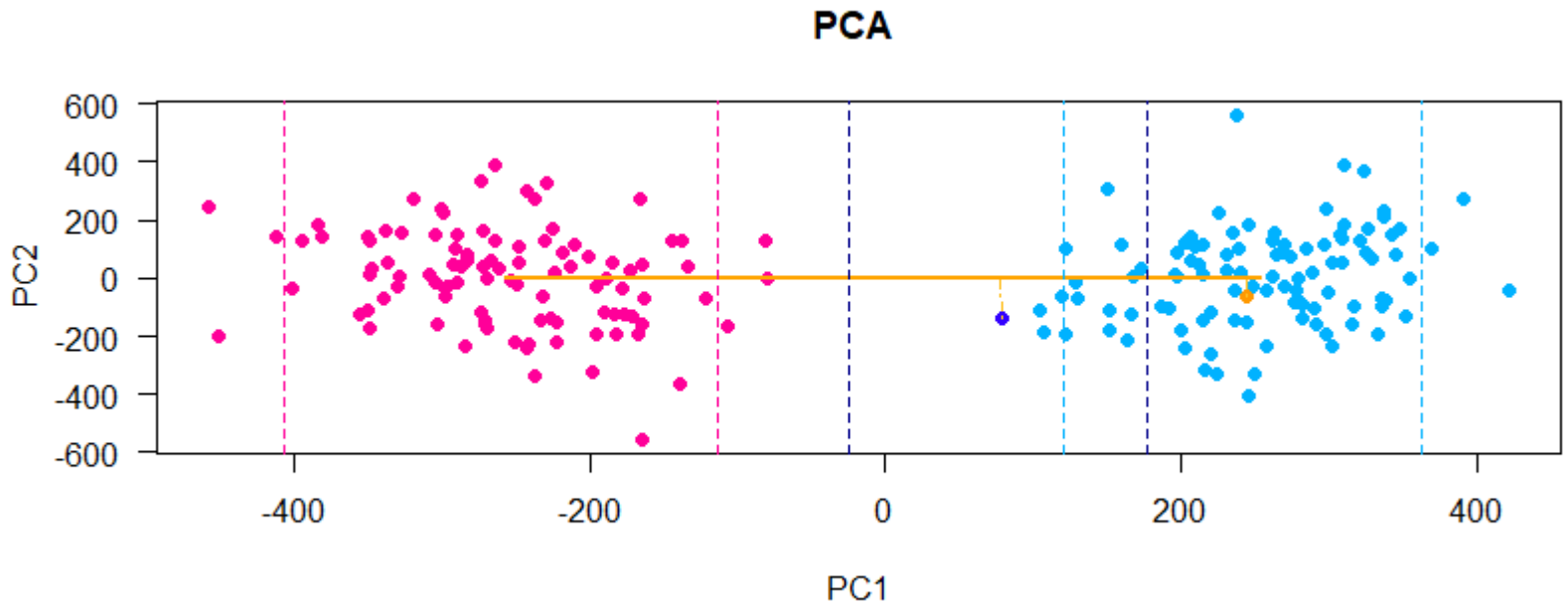
UMAP min_dist = 0

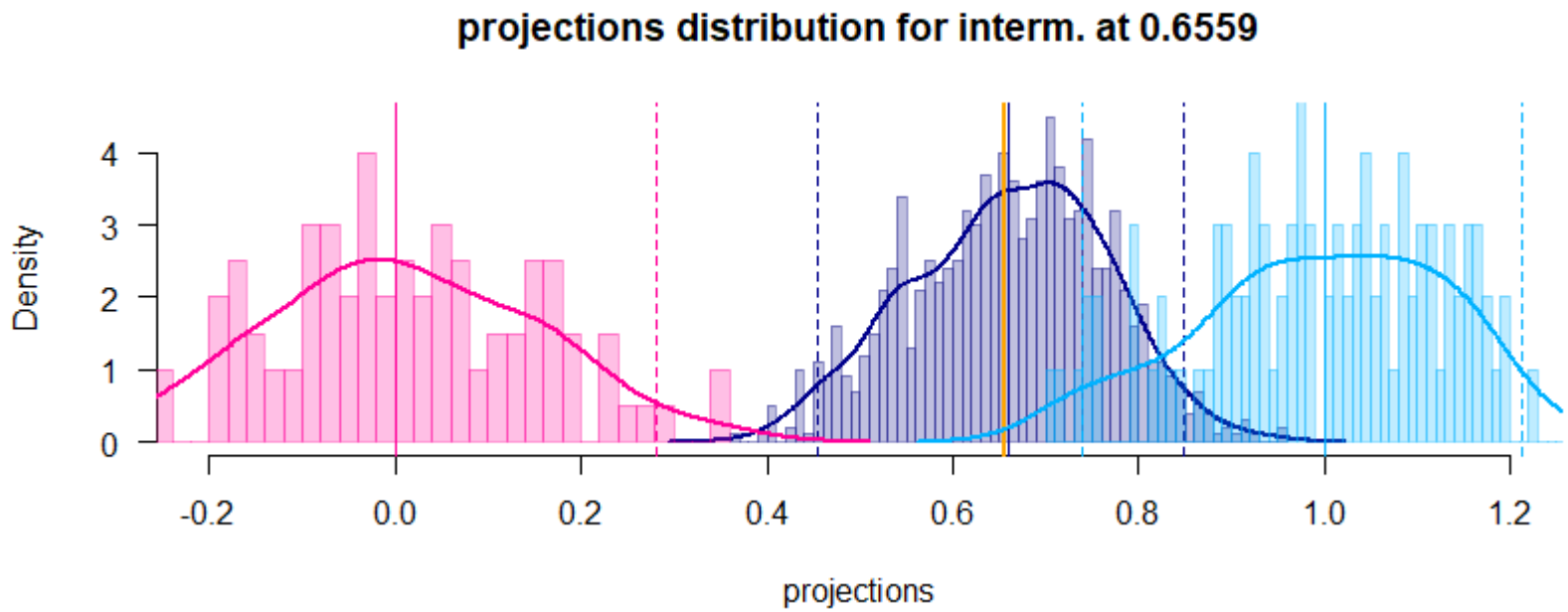


UMAP min_dist = 0

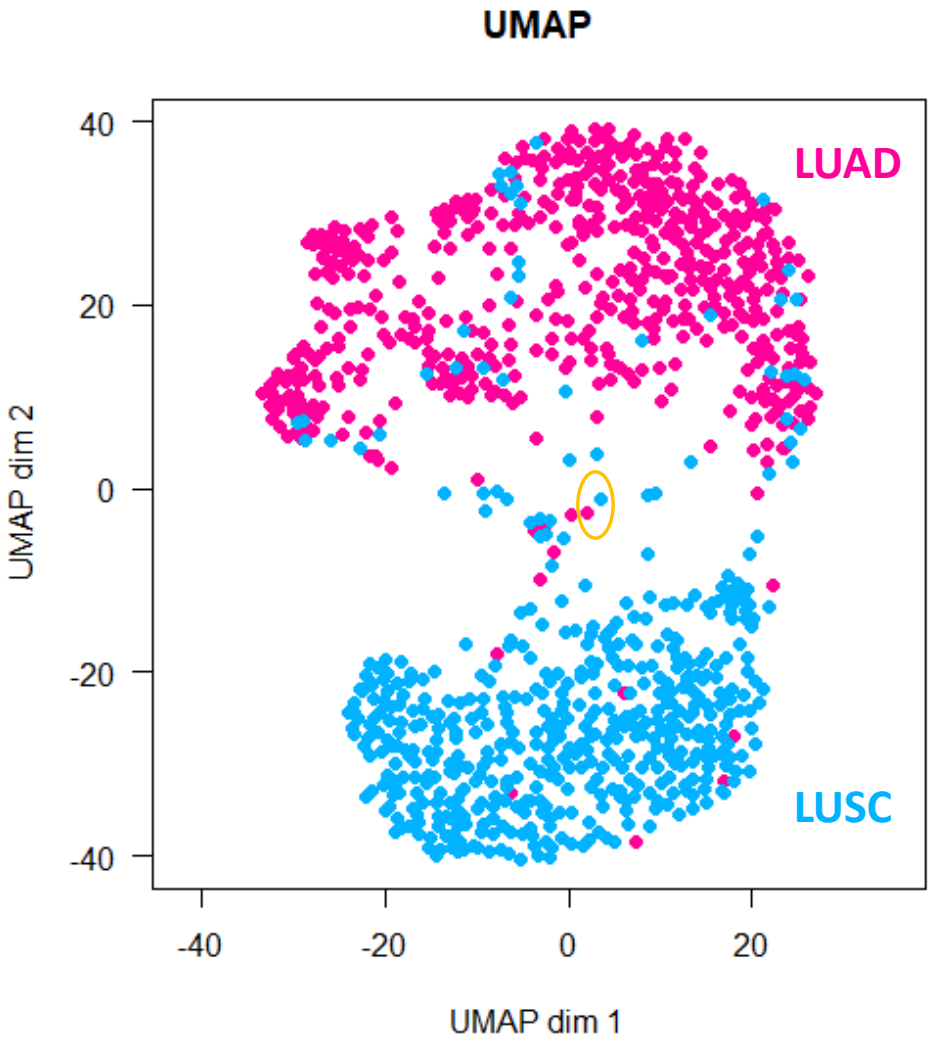


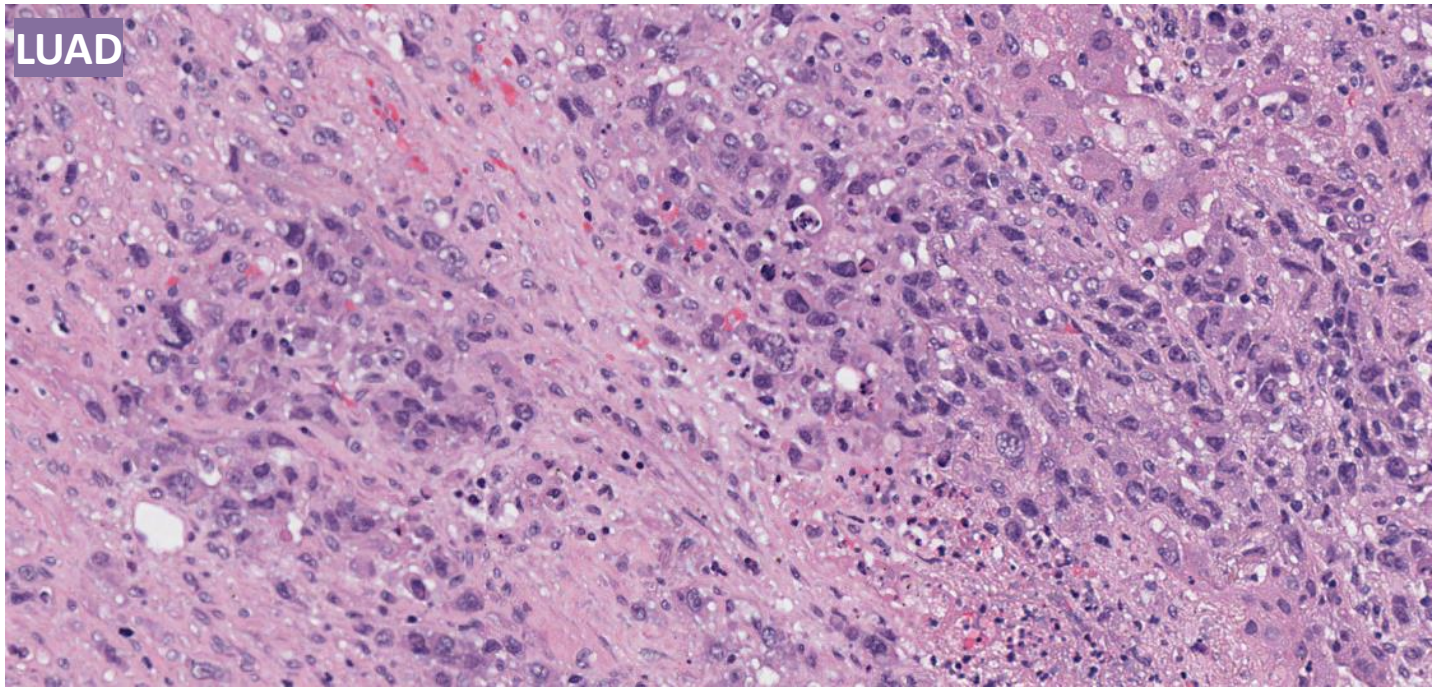
UMAP min_dist = 2.5

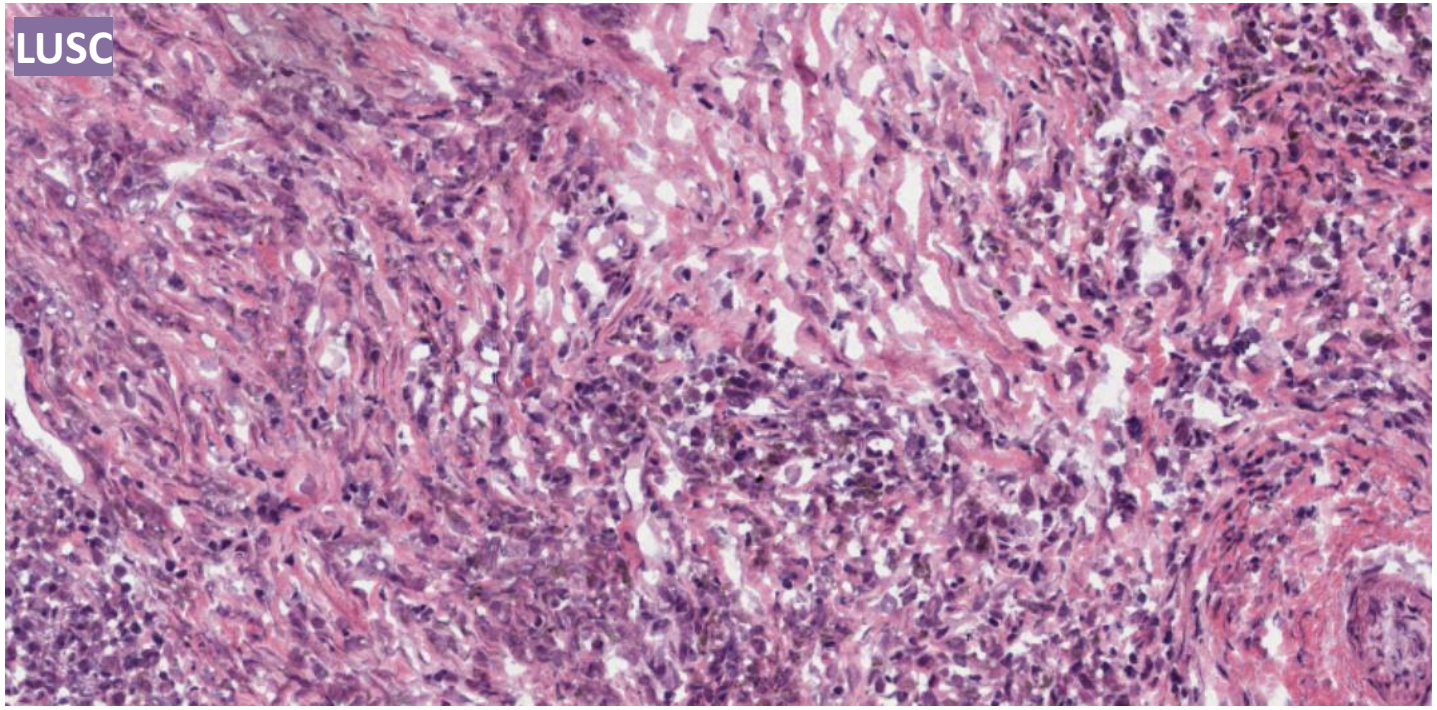

PCA
600
400
200
0
-200
-400
-600
PC2
-400
-200
0
200
400
PC1


projections distribution for interm. at 0.6559
4
3
2
1
0
Density
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
1.2
projections

Conclusion


UMAP
LUAD
LUSC
UMAP dim 2
UMAP dim 1
40
20
0
-20
-40
-40
-20
0
20


LUAD


LUSC

Acknowledgement

Genetic Cancer Susceptibility Group

Dr. Nicolas Alcala

Dr. Matthieu Foll

Dr. Lynnette Fernandez-Cuesta

Dr. Arnaud Poret

Emilie Mathian

Aurelie Gabriel

Lise Mangiante

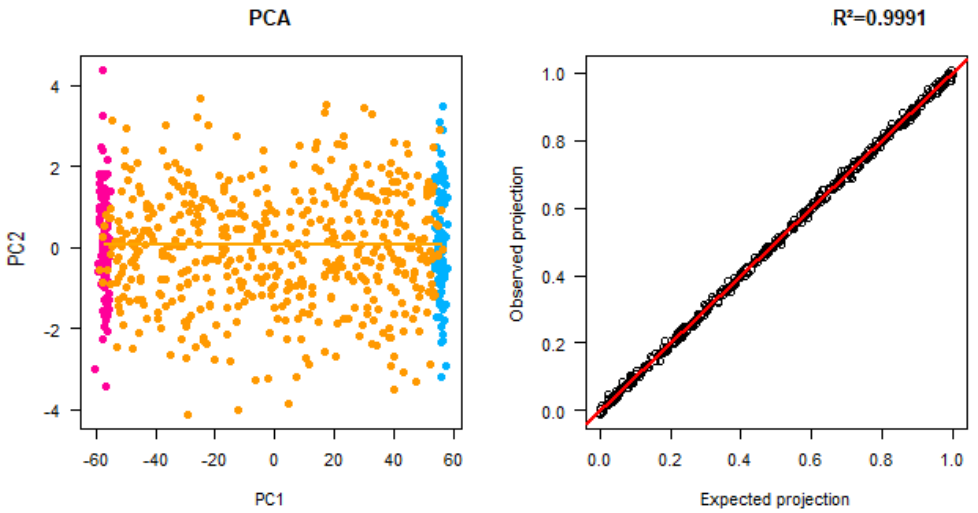
International Agency for Research on Cancer

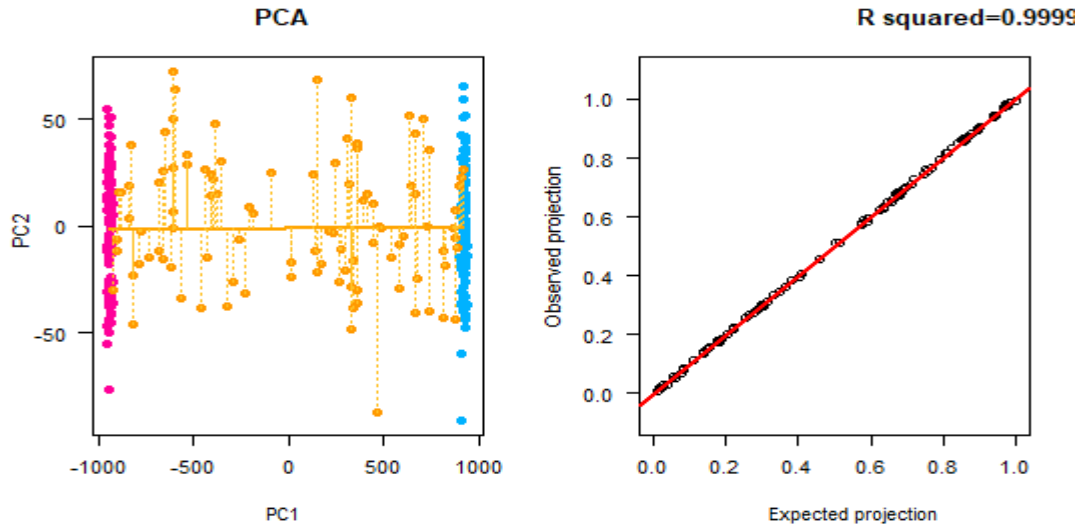
World Health Organization

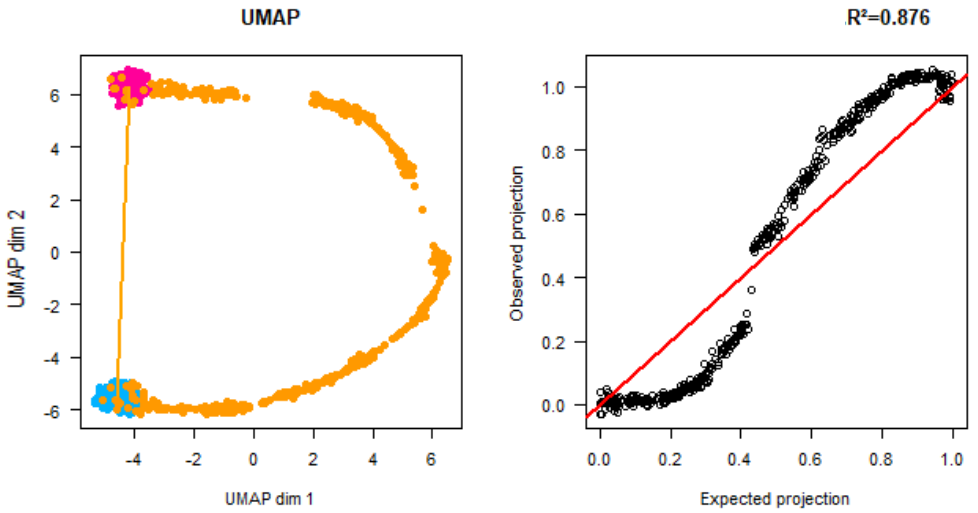
INSTITUT NATIONAL DU CANCER

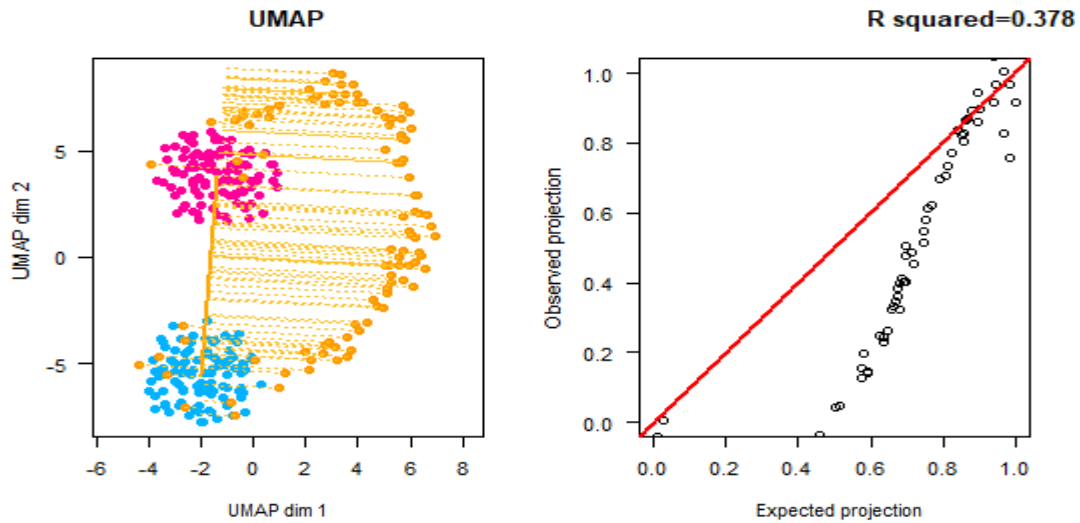
LA LIGUE CONTRE LE CANCER

UMAP


PCA
PC2
PC1
R²=0.9991
Observed projection
Expected projection


PCA
PC2
PC1
R squared=0.9999
Observed projection
Expected projection


UMAP
UMAP dim 2
UMAP dim 1
R²=0.876
Observed projection
Expected projection


UMAP
UMAP dim 2
UMAP dim 1
R squared=0.378
Observed projection
Expected projection