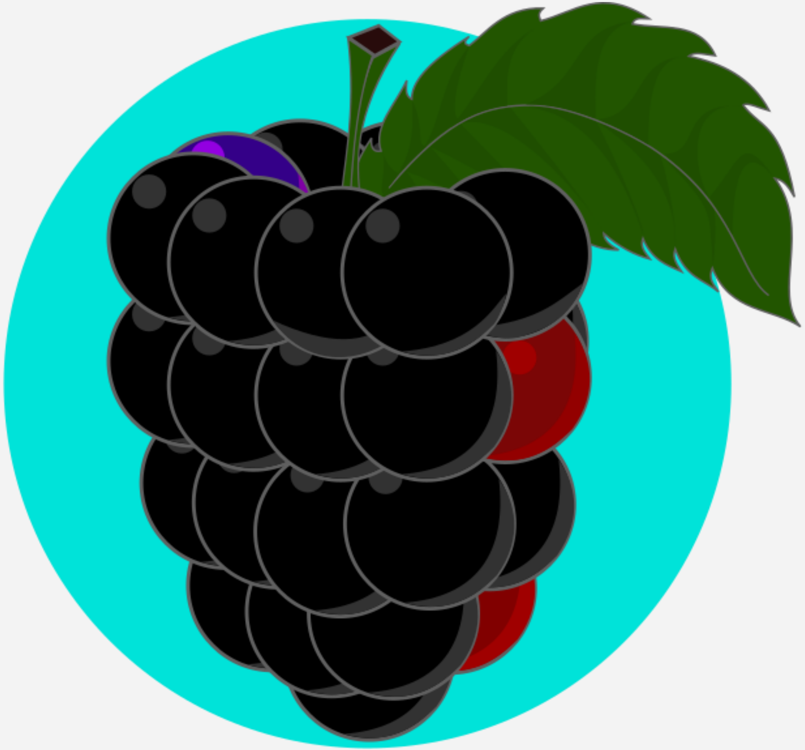


Druplet

Zeno Loi, Medical intern, Desbrest Institute of Public Health, June 2022

Transcriptome :

Quantitative expression of N genes
for I conditions

N	big	(~30k)
I	small	(~100)

→ Classical statistical analysis do not suit well… because of the curse of dimensionality

→ We need more adapted methods, more respectful of our transcription models

Group up co-regulated genes, 2 advantages :

- Tested hypotheses remain on a multiplicity of explanations
- Less hypotheses are tested

3 steps :

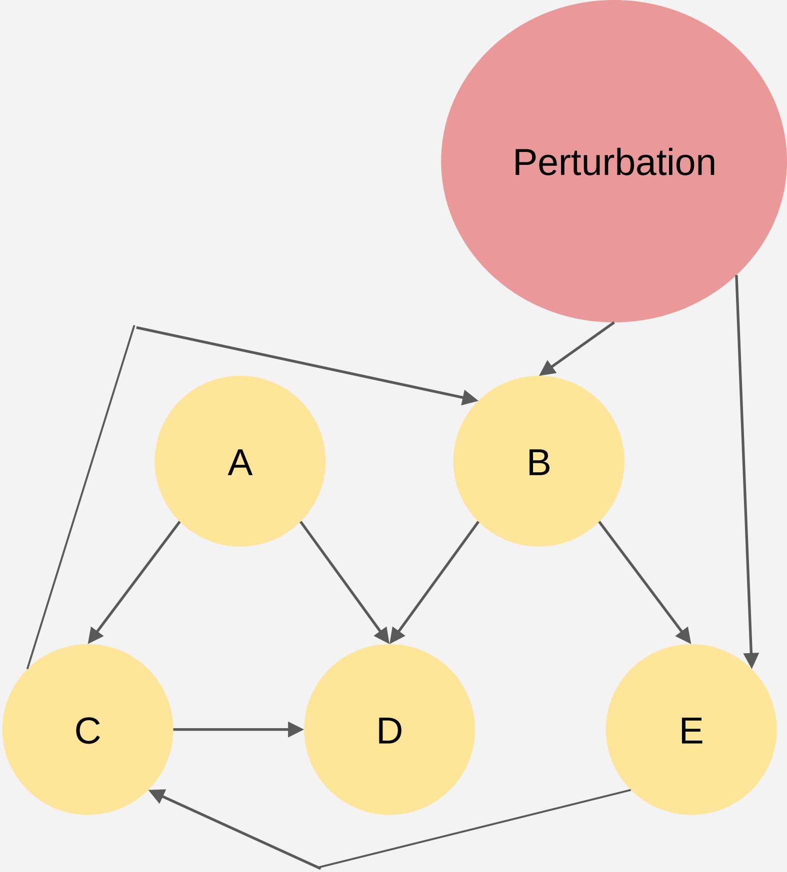
- Simulate artificial transcriptomes
- Identify co-regulated groups of genes
- Evaluate each group

- Simulate artificial transcriptomes

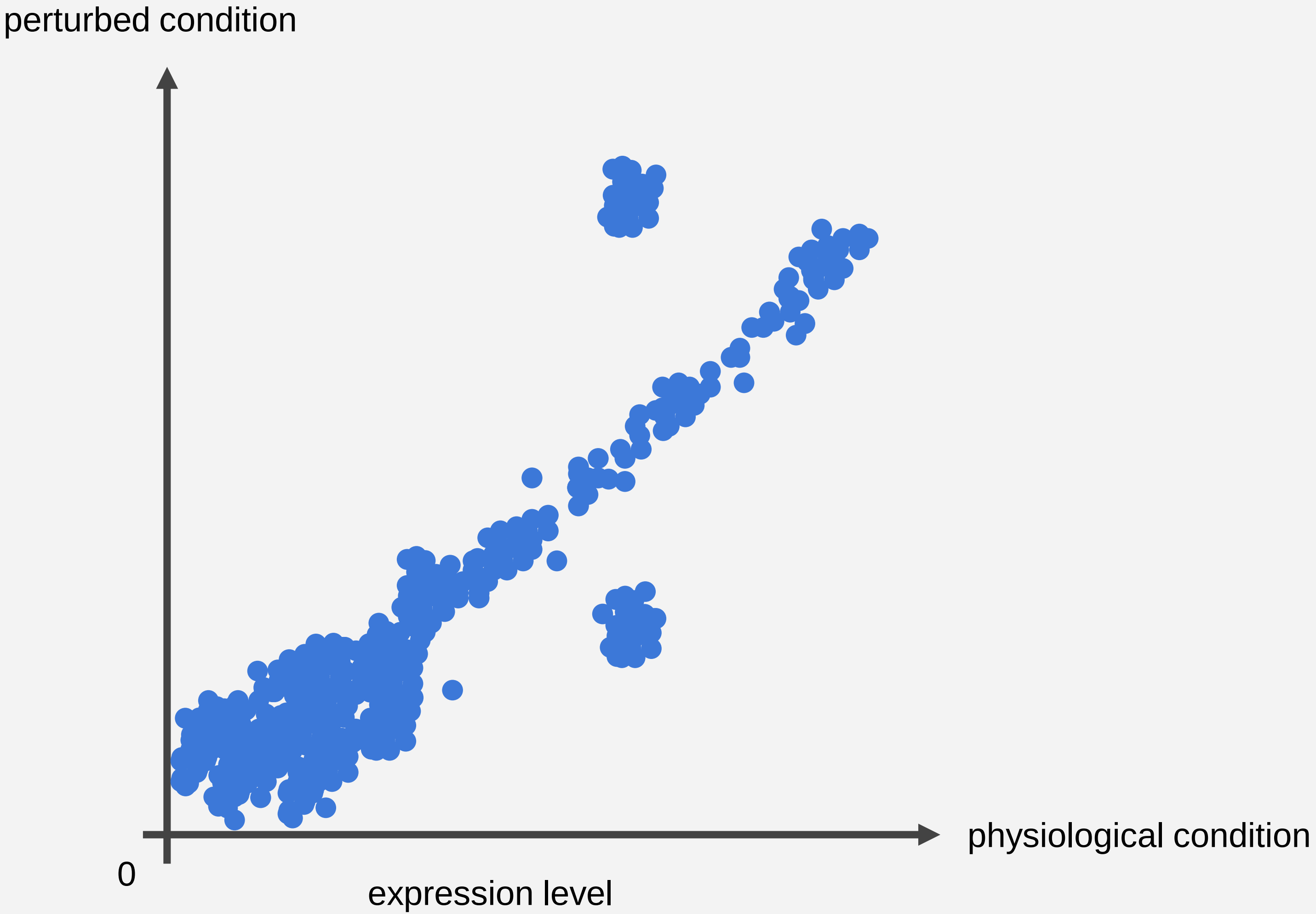
Regulation network (Dwight Kuo 2006) by simulating transcription in vivo

Stochastic production of 50 observations of a physiological condition

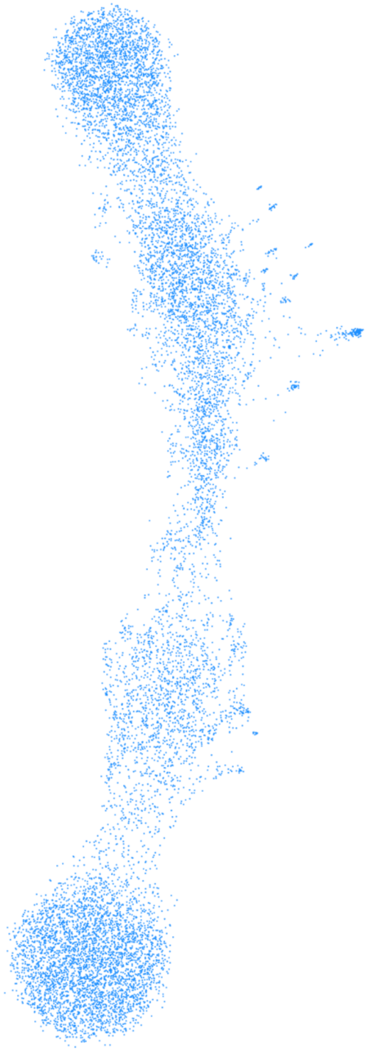
Stochastic production of 50 observations of a condition containing an external perturbation



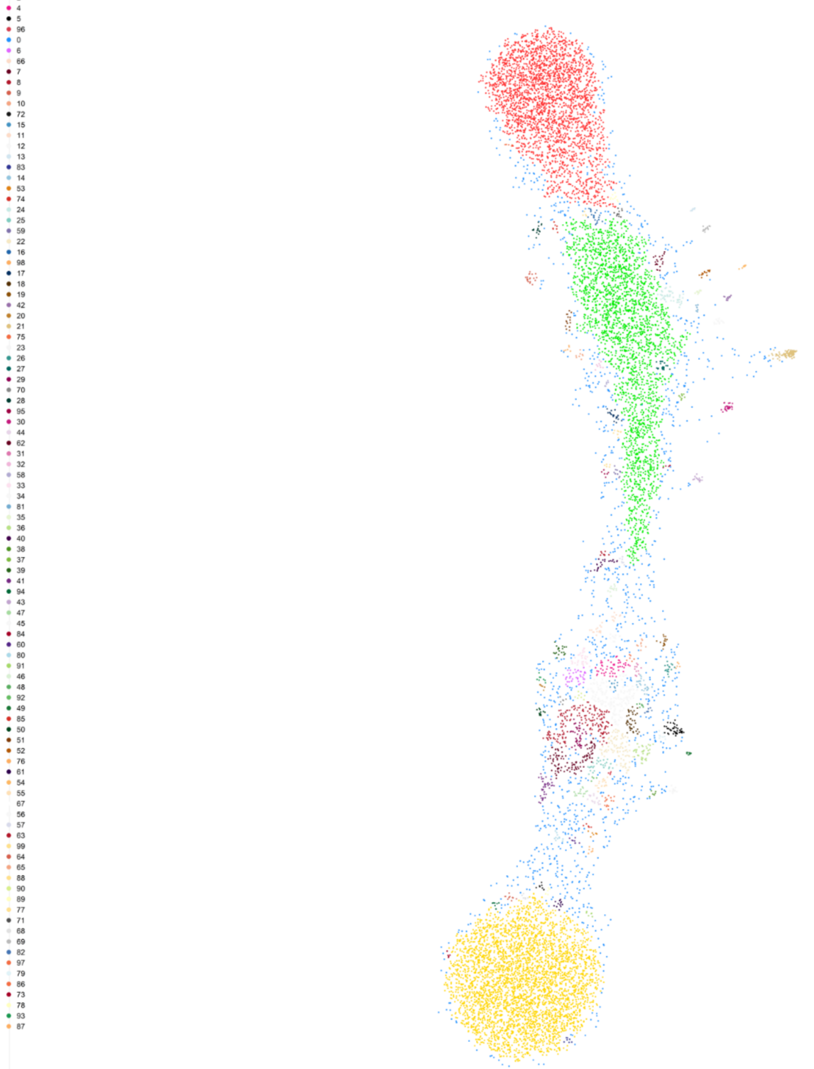
- Identify co-regulated genes


perturbed condition
physiological condition
0
expression level

UMAP



DB-scan

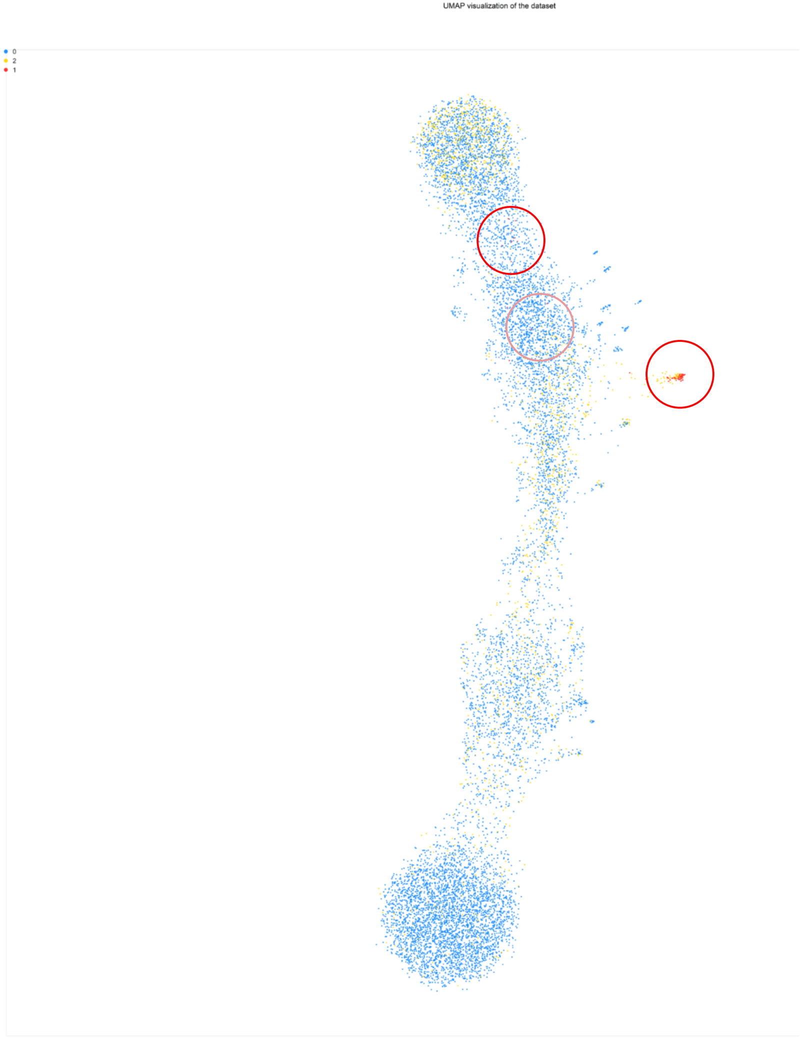

4
5
96
0
6
66
7
8
9
10
72
15
11
12
13
83
14
53
74
24
25
59
22
16
98
17
18
19
42
20
21
75
23
26
27
29
70
28
95
30
44
62
31
32
58
33
34
81
35
36
40
38
37
39
41
94
43
47
45
84
60
80
91
46
48
92
49
85
50
51
52
76
61
54
55
67
56
57
63
99
64
65
88
90
89
77
71
68
69
82
97
79
86
73
78
93
87

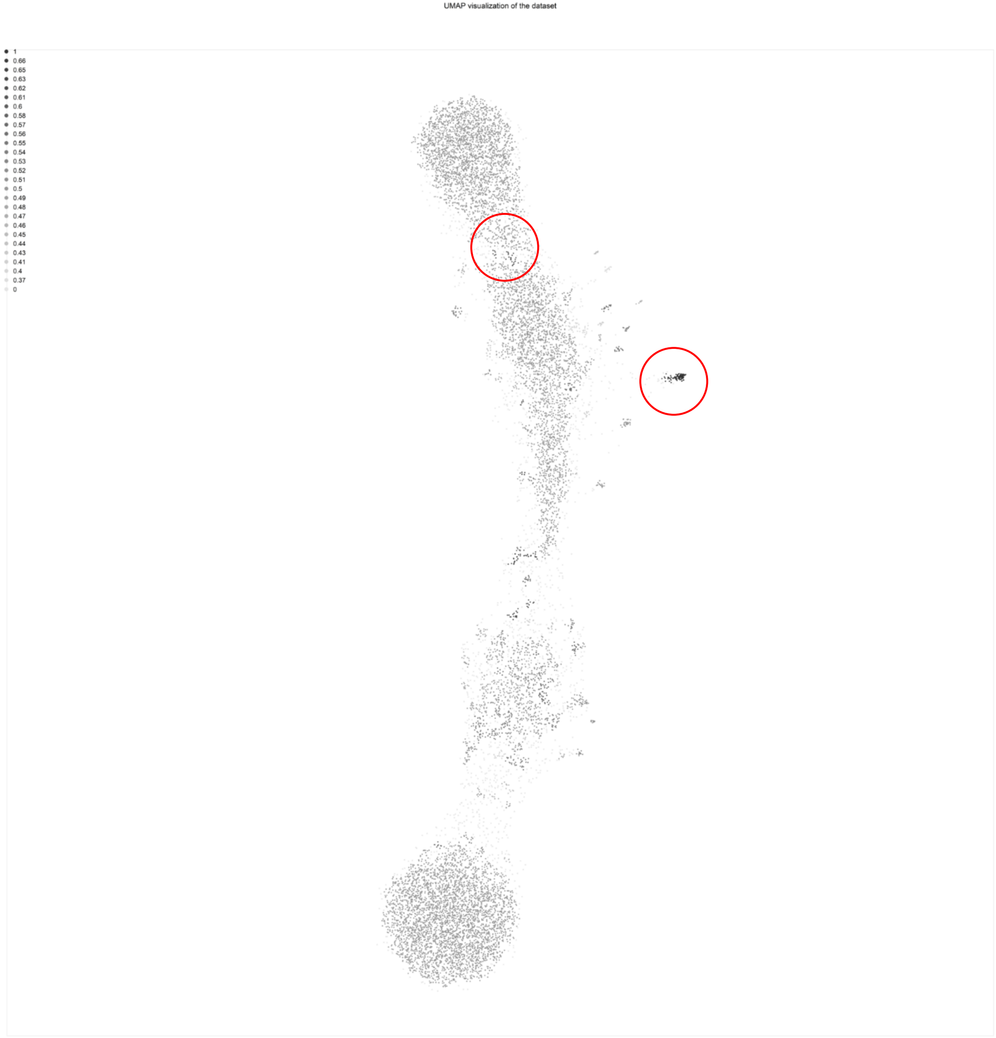
- Evaluate each group of genes

Divide the database in sub sets, one for each cluster

Each subset attempts to predict the condition state of each observation


0.63
0.62
0.61
0.6
0.58
0.57
0.56
0.55
0.54
0.53
0.52
0.51
0.5
0.49
0.48
0.47
0.46
0.45
0.44
0.43
0.41
0.4
0.37
0


UMAP visualization of the dataset


UMAP visualization of the dataset

Data base	Method	Genes directly impacted in the database	Selected impacted genes	Selected genes directly regulated by impacted genes	Wrongly selected genes
D1 (strong perturbation)	Druplet	100	**84**	**34**	**0**
D1	Logistic	100	5	8	85
D1	Lasso	100	3	2	1
D2 (weak perturbation)	Druplet	97	**68**	**4**	**0**
D2	Logistic	97	32	24	42
D2	Lasso	97	25	51	21