



INSTITUT  
de MATHÉMATIQUES  
de TOULOUSE



# Intégration de données « omiques »

*Comment lier les ingrédients d'une recette*

Sébastien Déjean

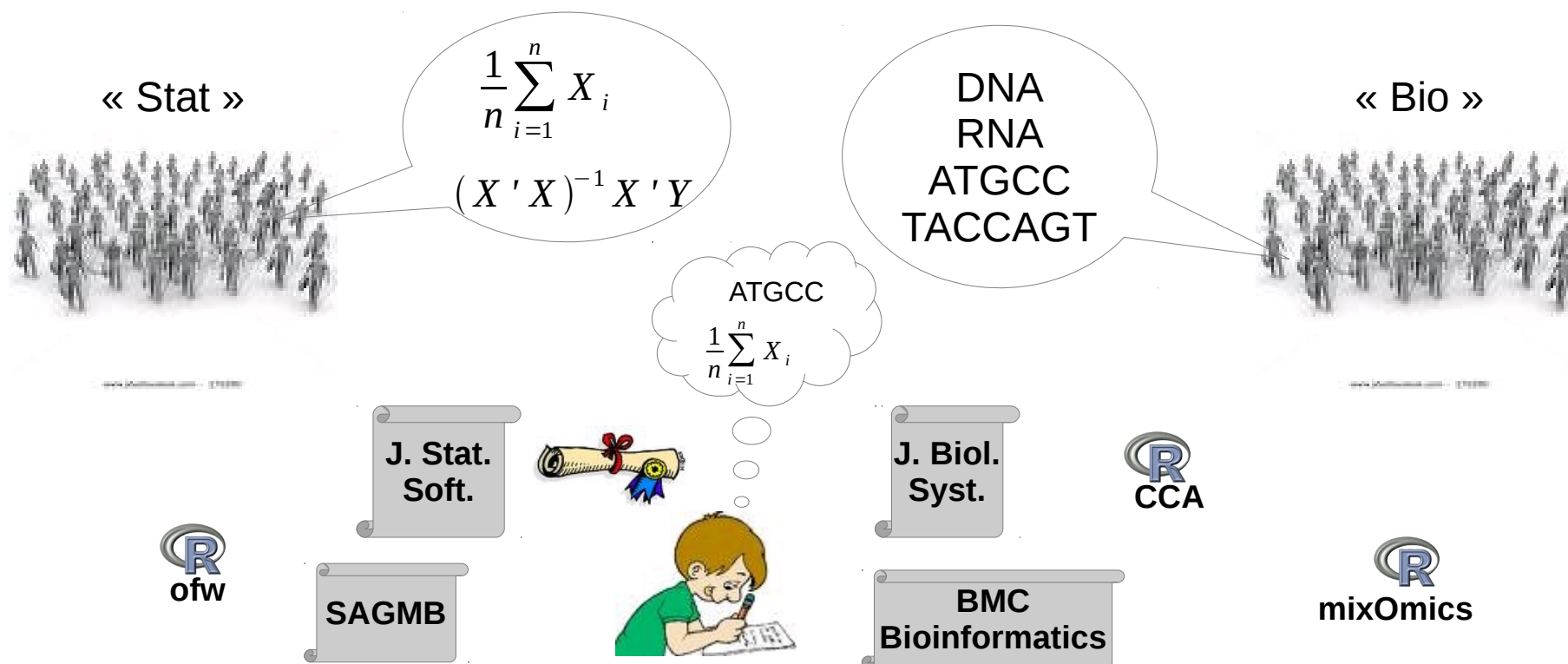
[www.math.univ-toulouse.fr/~sdejean/](http://www.math.univ-toulouse.fr/~sdejean/)

# Il était une fois...

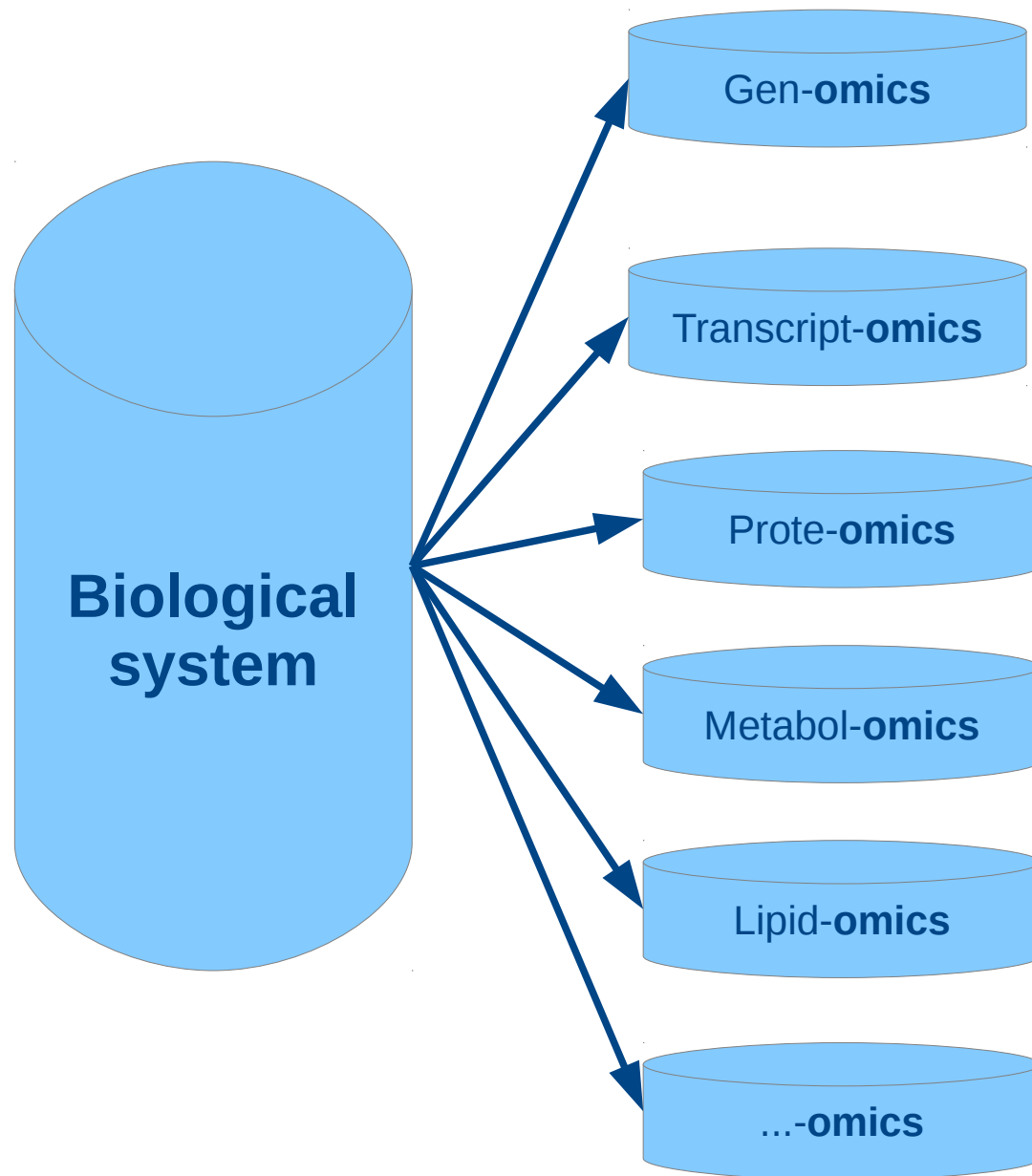
Once upon a time in Toulouse, a city in South West of France, two groups of scientists lived nearly together without talking to each other.



But one day, they decided to do so and to work together. They had Ph.D students, wrote articles and built R packages...



# Intégration de données ?



Chaque jeu de données « omic » peut être étudié séparément, mais...

... des informations pertinentes peuvent certainement être extraites d'une analyse conjointe de plusieurs jeux de données.

# <pub>



[www.math.univ-toulouse.fr/biostat/mixOmics](http://www.math.univ-toulouse.fr/biostat/mixOmics)

Package pour le logiciel  dédié à l'exploration et à l'intégration de données « omiques »

# </pub>

# Étude nutriginomique chez des souris soumises à différents régimes alimentaires



# Les données

Les données ont été acquises au laboratoire Toxalim de l'INRA de Toulouse. Elles proviennent d'une étude de nutrition chez la souris. Pour **40** souris, nous disposons :

- des données d'expression de **120** gènes recueillies sur membrane nylon avec marquage radioactif,
- des mesures de **21** acides gras hépatiques.

Par ailleurs, les **40** souris sont réparties selon deux facteurs :

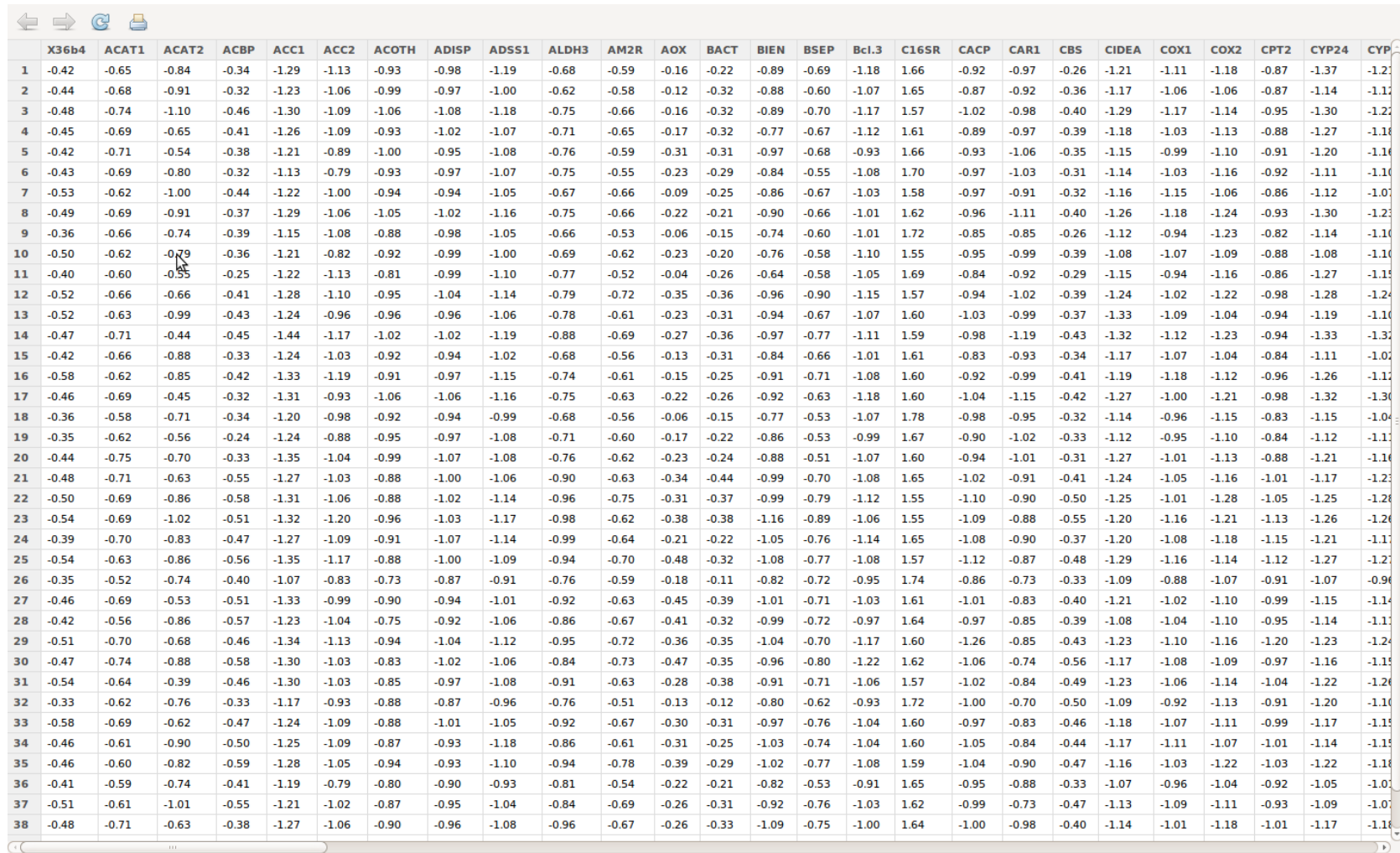
- Génotype (**2** modalités) : les souris sont soit de type sauvage (wt) soit génétiquement modifiées (PPAR) ; **20** souris dans chaque cas.
- Régime (**5** modalités) : les **5** régimes alimentaires sont notés *ref*, *coc*, *fish*, *lin*, *sun* ; **4** souris de chaque génotype sont soumises à chaque régime alimentaire.



✎ P. Martin, H. Guillou, F. Lasserre, S. Déjean, A. Lan, J-M. Pascussi, M. San Cristobal, P. Legrand, P. Besse, T. Pineau (2007). Novel aspects of PPARalpha-mediated regulation of lipid and xenobiotic metabolism revealed through a nutrigenomic study. *Hepatology*

# Aperçu des données "gènes"

120 colonnes



|    | X36b4 | ACAT1 | ACAT2 | ACBP  | ACC1  | ACC2  | ACOTH | ADISP | ADSS1 | ALDH3 | AM2R  | AOX   | BACT  | BIEN  | BSEP  | Bcl.3 | C16SR | CACP  | CAR1  | CBS   | CIDEA | COX1  | COX2  | CPT2  | CYP24 | CYP   |
|----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1  | -0.42 | -0.65 | -0.84 | -0.34 | -1.29 | -1.13 | -0.93 | -0.98 | -1.19 | -0.68 | -0.59 | -0.16 | -0.22 | -0.89 | -0.69 | -1.18 | 1.66  | -0.92 | -0.97 | -0.26 | -1.21 | -1.11 | -1.18 | -0.87 | -1.37 | -1.21 |
| 2  | -0.44 | -0.68 | -0.91 | -0.32 | -1.23 | -1.06 | -0.99 | -0.97 | -1.00 | -0.62 | -0.58 | -0.12 | -0.32 | -0.88 | -0.60 | -1.07 | 1.65  | -0.87 | -0.92 | -0.36 | -1.17 | -1.06 | -1.06 | -0.87 | -1.14 | -1.12 |
| 3  | -0.48 | -0.74 | -1.10 | -0.46 | -1.30 | -1.09 | -1.06 | -1.08 | -1.18 | -0.75 | -0.66 | -0.16 | -0.32 | -0.89 | -0.70 | -1.17 | 1.57  | -1.02 | -0.98 | -0.40 | -1.29 | -1.17 | -1.14 | -0.95 | -1.30 | -1.22 |
| 4  | -0.45 | -0.69 | -0.65 | -0.41 | -1.26 | -1.09 | -0.93 | -1.02 | -1.07 | -0.71 | -0.65 | -0.17 | -0.32 | -0.77 | -0.67 | -1.12 | 1.61  | -0.89 | -0.97 | -0.39 | -1.18 | -1.03 | -1.13 | -0.88 | -1.27 | -1.18 |
| 5  | -0.42 | -0.71 | -0.54 | -0.38 | -1.21 | -0.89 | -1.00 | -0.95 | -1.08 | -0.76 | -0.59 | -0.31 | -0.31 | -0.97 | -0.68 | -0.93 | 1.66  | -0.93 | -1.06 | -0.35 | -1.15 | -0.99 | -1.10 | -0.91 | -1.20 | -1.16 |
| 6  | -0.43 | -0.69 | -0.80 | -0.32 | -1.13 | -0.79 | -0.93 | -0.97 | -1.07 | -0.75 | -0.55 | -0.23 | -0.29 | -0.84 | -0.55 | -1.08 | 1.70  | -0.97 | -1.03 | -0.31 | -1.14 | -1.03 | -1.16 | -0.92 | -1.11 | -1.10 |
| 7  | -0.53 | -0.62 | -1.00 | -0.44 | -1.22 | -1.00 | -0.94 | -0.94 | -1.05 | -0.67 | -0.66 | -0.09 | -0.25 | -0.86 | -0.67 | -1.03 | 1.58  | -0.97 | -0.91 | -0.32 | -1.16 | -1.15 | -1.06 | -0.86 | -1.12 | -1.07 |
| 8  | -0.49 | -0.69 | -0.91 | -0.37 | -1.29 | -1.06 | -1.05 | -1.02 | -1.16 | -0.75 | -0.66 | -0.22 | -0.21 | -0.90 | -0.66 | -1.01 | 1.62  | -0.96 | -1.11 | -0.40 | -1.26 | -1.18 | -1.24 | -0.93 | -1.30 | -1.23 |
| 9  | -0.36 | -0.66 | -0.74 | -0.39 | -1.15 | -1.08 | -0.88 | -0.98 | -1.05 | -0.66 | -0.53 | -0.06 | -0.15 | -0.74 | -0.60 | -1.01 | 1.72  | -0.85 | -0.85 | -0.26 | -1.12 | -0.94 | -1.23 | -0.82 | -1.14 | -1.16 |
| 10 | -0.50 | -0.62 | -0.79 | -0.36 | -1.21 | -0.82 | -0.92 | -0.99 | -1.00 | -0.69 | -0.62 | -0.23 | -0.20 | -0.76 | -0.58 | -1.10 | 1.55  | -0.95 | -0.99 | -0.39 | -1.08 | -1.07 | -1.09 | -0.88 | -1.08 | -1.10 |
| 11 | -0.40 | -0.60 | -0.55 | -0.25 | -1.22 | -1.13 | -0.81 | -0.99 | -1.10 | -0.77 | -0.52 | -0.04 | -0.26 | -0.64 | -0.58 | -1.05 | 1.69  | -0.84 | -0.92 | -0.29 | -1.15 | -0.94 | -1.16 | -0.86 | -1.27 | -1.15 |
| 12 | -0.52 | -0.66 | -0.66 | -0.41 | -1.28 | -1.10 | -0.95 | -1.04 | -1.14 | -0.79 | -0.72 | -0.35 | -0.36 | -0.96 | -0.90 | -1.15 | 1.57  | -0.94 | -1.02 | -0.39 | -1.24 | -1.02 | -1.22 | -0.98 | -1.28 | -1.24 |
| 13 | -0.52 | -0.63 | -0.99 | -0.43 | -1.24 | -0.96 | -0.96 | -0.96 | -1.06 | -0.78 | -0.61 | -0.23 | -0.31 | -0.94 | -0.67 | -1.07 | 1.60  | -1.03 | -0.99 | -0.37 | -1.33 | -1.09 | -1.04 | -0.94 | -1.19 | -1.16 |
| 14 | -0.47 | -0.71 | -0.44 | -0.45 | -1.44 | -1.17 | -1.02 | -1.02 | -1.19 | -0.88 | -0.69 | -0.27 | -0.36 | -0.97 | -0.77 | -1.11 | 1.59  | -0.98 | -1.19 | -0.43 | -1.32 | -1.12 | -1.23 | -0.94 | -1.33 | -1.32 |
| 15 | -0.42 | -0.66 | -0.88 | -0.33 | -1.24 | -1.03 | -0.92 | -0.94 | -1.02 | -0.68 | -0.56 | -0.13 | -0.31 | -0.84 | -0.66 | -1.01 | 1.61  | -0.83 | -0.93 | -0.34 | -1.17 | -1.07 | -1.04 | -0.84 | -1.11 | -1.02 |
| 16 | -0.58 | -0.62 | -0.85 | -0.42 | -1.33 | -1.19 | -0.91 | -0.97 | -1.15 | -0.74 | -0.61 | -0.15 | -0.25 | -0.91 | -0.71 | -1.08 | 1.60  | -0.92 | -0.99 | -0.41 | -1.19 | -1.18 | -1.12 | -0.96 | -1.26 | -1.12 |
| 17 | -0.46 | -0.69 | -0.45 | -0.32 | -1.31 | -0.93 | -1.06 | -1.06 | -1.16 | -0.75 | -0.63 | -0.22 | -0.26 | -0.92 | -0.63 | -1.18 | 1.60  | -1.04 | -1.15 | -0.42 | -1.27 | -1.00 | -1.21 | -0.98 | -1.32 | -1.30 |
| 18 | -0.36 | -0.58 | -0.71 | -0.34 | -1.20 | -0.98 | -0.92 | -0.94 | -0.99 | -0.68 | -0.56 | -0.06 | -0.15 | -0.77 | -0.53 | -1.07 | 1.78  | -0.98 | -0.95 | -0.32 | -1.14 | -0.96 | -1.15 | -0.83 | -1.15 | -1.06 |
| 19 | -0.35 | -0.62 | -0.56 | -0.24 | -1.24 | -0.88 | -0.95 | -0.97 | -1.08 | -0.71 | -0.60 | -0.17 | -0.22 | -0.86 | -0.53 | -0.99 | 1.67  | -0.90 | -1.02 | -0.33 | -1.12 | -0.95 | -1.10 | -0.84 | -1.12 | -1.11 |
| 20 | -0.44 | -0.75 | -0.70 | -0.33 | -1.35 | -1.04 | -0.99 | -1.07 | -1.08 | -0.76 | -0.62 | -0.23 | -0.24 | -0.88 | -0.51 | -1.07 | 1.60  | -0.94 | -1.01 | -0.31 | -1.27 | -1.01 | -1.13 | -0.88 | -1.21 | -1.16 |
| 21 | -0.48 | -0.71 | -0.63 | -0.55 | -1.27 | -1.03 | -0.88 | -1.00 | -1.06 | -0.90 | -0.63 | -0.34 | -0.44 | -0.99 | -0.70 | -1.08 | 1.65  | -1.02 | -0.91 | -0.41 | -1.24 | -1.05 | -1.16 | -1.01 | -1.17 | -1.23 |
| 22 | -0.50 | -0.69 | -0.86 | -0.58 | -1.31 | -1.06 | -0.88 | -1.02 | -1.14 | -0.96 | -0.75 | -0.31 | -0.37 | -0.99 | -0.79 | -1.12 | 1.55  | -1.10 | -0.90 | -0.50 | -1.25 | -1.01 | -1.28 | -1.05 | -1.25 | -1.28 |
| 23 | -0.54 | -0.69 | -1.02 | -0.51 | -1.32 | -1.20 | -0.96 | -1.03 | -1.17 | -0.98 | -0.62 | -0.38 | -0.38 | -1.16 | -0.89 | -1.06 | 1.55  | -1.09 | -0.88 | -0.55 | -1.20 | -1.16 | -1.21 | -1.13 | -1.26 | -1.26 |
| 24 | -0.39 | -0.70 | -0.83 | -0.47 | -1.27 | -1.09 | -0.91 | -1.07 | -1.14 | -0.99 | -0.64 | -0.21 | -0.22 | -1.05 | -0.76 | -1.14 | 1.65  | -1.08 | -0.90 | -0.37 | -1.20 | -1.08 | -1.18 | -1.15 | -1.21 | -1.11 |
| 25 | -0.54 | -0.63 | -0.86 | -0.56 | -1.35 | -1.17 | -0.88 | -1.00 | -1.09 | -0.94 | -0.70 | -0.48 | -0.32 | -1.08 | -0.77 | -1.08 | 1.57  | -1.12 | -0.87 | -0.48 | -1.29 | -1.16 | -1.14 | -1.12 | -1.27 | -1.27 |
| 26 | -0.35 | -0.52 | -0.74 | -0.40 | -1.07 | -0.83 | -0.73 | -0.87 | -0.91 | -0.76 | -0.59 | -0.18 | -0.11 | -0.82 | -0.72 | -0.95 | 1.74  | -0.86 | -0.73 | -0.33 | -1.09 | -0.88 | -1.07 | -0.91 | -1.07 | -0.96 |
| 27 | -0.46 | -0.69 | -0.53 | -0.51 | -1.33 | -0.99 | -0.90 | -0.94 | -1.01 | -0.92 | -0.63 | -0.45 | -0.39 | -1.01 | -0.71 | -1.03 | 1.61  | -1.01 | -0.83 | -0.40 | -1.21 | -1.02 | -1.10 | -0.99 | -1.15 | -1.14 |
| 28 | -0.42 | -0.56 | -0.86 | -0.57 | -1.23 | -1.04 | -0.75 | -0.92 | -1.06 | -0.86 | -0.67 | -0.41 | -0.32 | -0.99 | -0.72 | -0.97 | 1.64  | -0.97 | -0.85 | -0.39 | -1.08 | -1.04 | -1.10 | -0.95 | -1.14 | -1.11 |
| 29 | -0.51 | -0.70 | -0.68 | -0.46 | -1.34 | -1.13 | -0.94 | -1.04 | -1.12 | -0.95 | -0.72 | -0.36 | -0.35 | -1.04 | -0.70 | -1.17 | 1.60  | -1.26 | -0.85 | -0.43 | -1.23 | -1.10 | -1.16 | -1.20 | -1.23 | -1.24 |
| 30 | -0.47 | -0.74 | -0.88 | -0.58 | -1.30 | -1.03 | -0.83 | -1.02 | -1.06 | -0.84 | -0.73 | -0.47 | -0.35 | -0.96 | -0.80 | -1.22 | 1.62  | -1.06 | -0.74 | -0.56 | -1.17 | -1.08 | -1.09 | -0.97 | -1.16 | -1.15 |
| 31 | -0.54 | -0.64 | -0.39 | -0.46 | -1.30 | -1.03 | -0.85 | -0.97 | -1.08 | -0.91 | -0.63 | -0.28 | -0.38 | -0.91 | -0.71 | -1.06 | 1.57  | -1.02 | -0.84 | -0.49 | -1.23 | -1.06 | -1.14 | -1.04 | -1.22 | -1.26 |
| 32 | -0.33 | -0.62 | -0.76 | -0.33 | -1.17 | -0.93 | -0.88 | -0.87 | -0.96 | -0.76 | -0.51 | -0.13 | -0.12 | -0.80 | -0.62 | -0.93 | 1.72  | -1.00 | -0.70 | -0.50 | -1.09 | -0.92 | -1.13 | -0.91 | -1.20 | -1.10 |
| 33 | -0.58 | -0.69 | -0.62 | -0.47 | -1.24 | -1.09 | -0.88 | -1.01 | -1.05 | -0.92 | -0.67 | -0.30 | -0.31 | -0.97 | -0.76 | -1.04 | 1.60  | -0.97 | -0.83 | -0.46 | -1.18 | -1.07 | -1.11 | -0.99 | -1.17 | -1.15 |
| 34 | -0.46 | -0.61 | -0.90 | -0.50 | -1.25 | -1.09 | -0.87 | -0.93 | -1.18 | -0.86 | -0.61 | -0.31 | -0.25 | -1.03 | -0.74 | -1.04 | 1.60  | -1.05 | -0.84 | -0.44 | -1.17 | -1.11 | -1.07 | -1.01 | -1.14 | -1.15 |
| 35 | -0.46 | -0.60 | -0.82 | -0.59 | -1.28 | -1.05 | -0.94 | -0.93 | -1.10 | -0.94 | -0.78 | -0.39 | -0.29 | -1.02 | -0.77 | -1.08 | 1.59  | -1.04 | -0.90 | -0.47 | -1.16 | -1.03 | -1.22 | -1.03 | -1.22 | -1.18 |
| 36 | -0.41 | -0.59 | -0.74 | -0.41 | -1.19 | -0.79 | -0.80 | -0.90 | -0.93 | -0.81 | -0.54 | -0.22 | -0.21 | -0.82 | -0.53 | -0.91 | 1.65  | -0.95 | -0.88 | -0.33 | -1.07 | -0.96 | -1.04 | -0.92 | -1.05 | -1.07 |
| 37 | -0.51 | -0.61 | -1.01 | -0.55 | -1.21 | -1.02 | -0.87 | -0.95 | -1.04 | -0.84 | -0.69 | -0.26 | -0.31 | -0.92 | -0.76 | -1.03 | 1.62  | -0.99 | -0.73 | -0.47 | -1.13 | -1.09 | -1.11 | -0.93 | -1.09 | -1.07 |
| 38 | -0.48 | -0.71 | -0.63 | -0.38 | -1.27 | -1.06 | -0.90 | -0.96 | -1.08 | -0.96 | -0.67 | -0.26 | -0.33 | -1.09 | -0.75 | -1.00 | 1.64  | -1.00 | -0.98 | -0.40 | -1.14 | -1.01 | -1.18 | -1.01 | -1.17 | -1.18 |

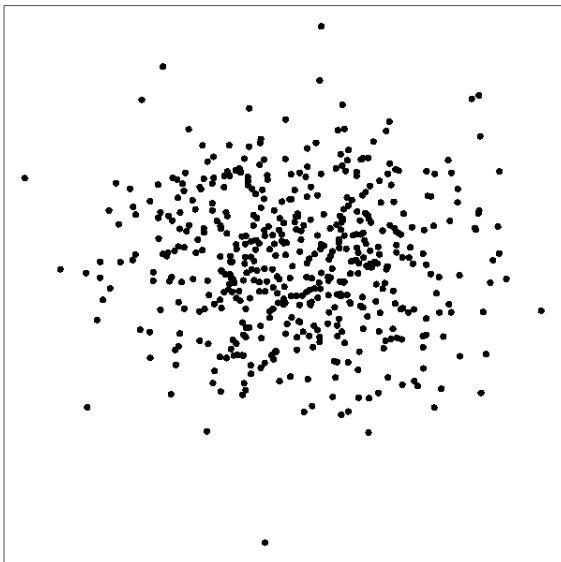
# Problème

Proposer une méthode permettant de mettre en évidence et de représenter les relations entre les gènes et les acides gras.

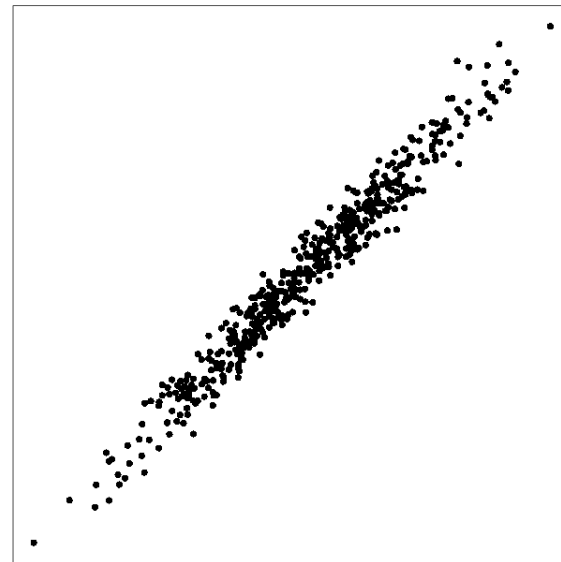


# Corrélation

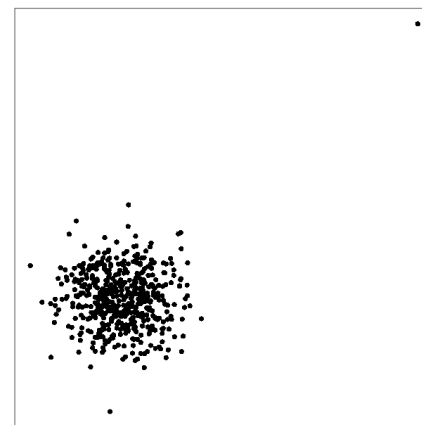
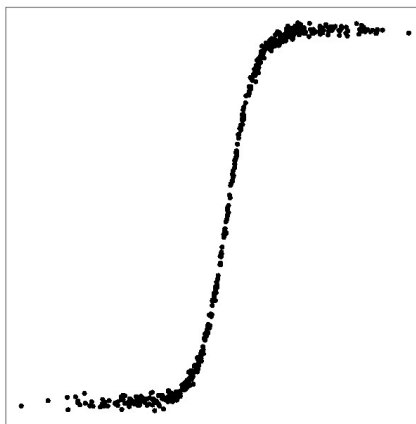
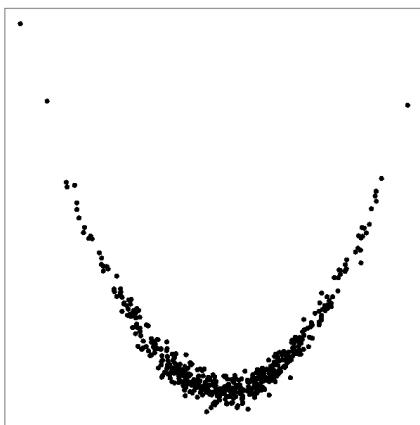
$\approx 0$



$\approx 1$

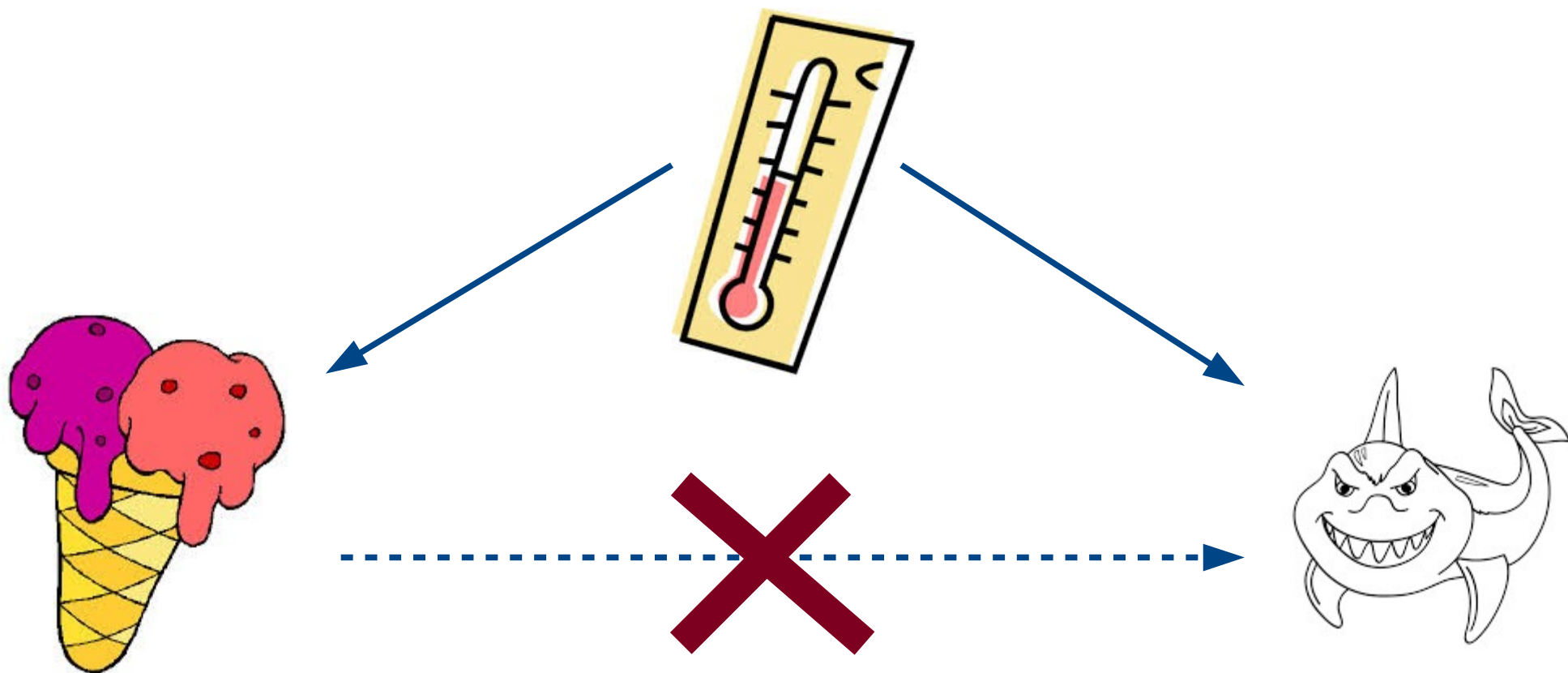


?

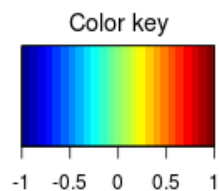


# Corrélation & Causalité ?

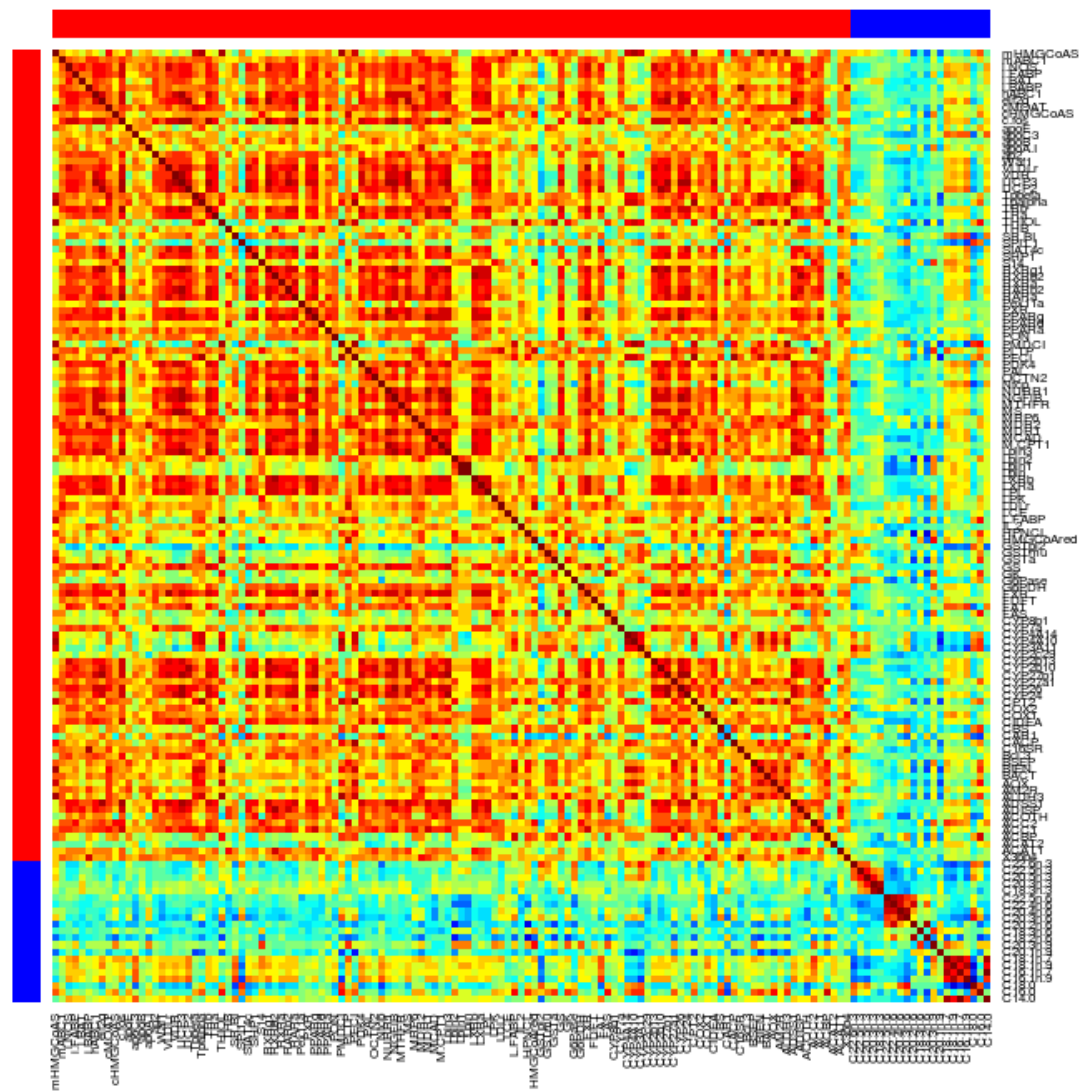
La fréquence des attaques de requins est fortement corrélée à la vente de glaces sur les pages ! Manger des glaces nous rend-il plus appétissant pour les requins ?



# Matrice de corrélation complète

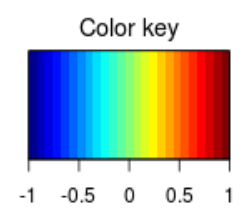


[X,Y] correlation matrix

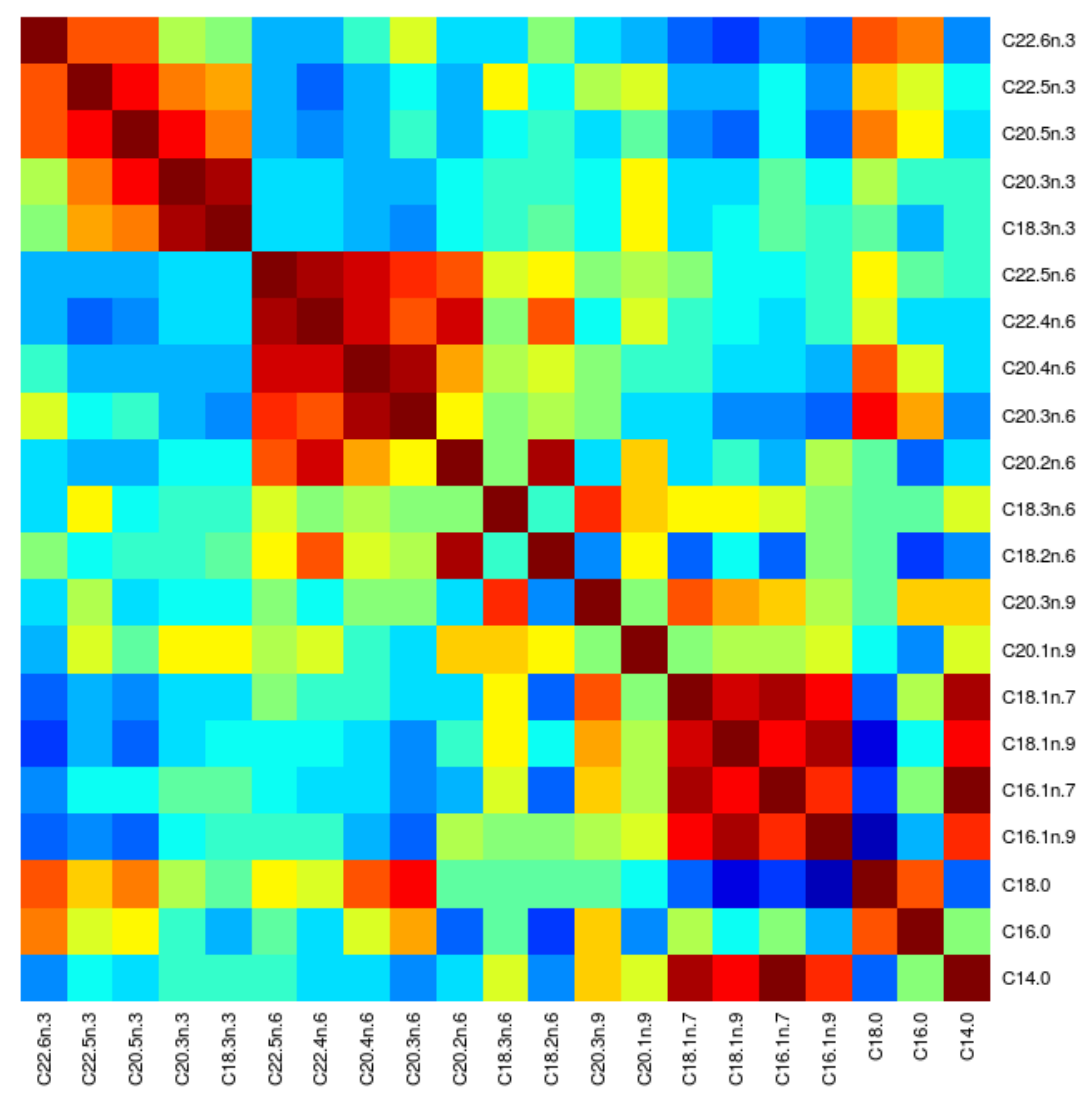


A horizontal color bar legend labeled 'Color key' at the top. It shows a gradient from dark blue on the left to dark red on the right, passing through green and yellow in the middle. Below the bar are numerical tick marks at -1, -0.5, 0, 0.5, and 1.

# Matrice de corrélation (lipides)



X correlation matrix



Color key



-0.5 0 0.5

Heatmap showing the expression of 150 genes across 15 cell lines. The color scale ranges from blue (low expression) to red (high expression). The genes are listed on the left, and the cell lines are listed on the right.

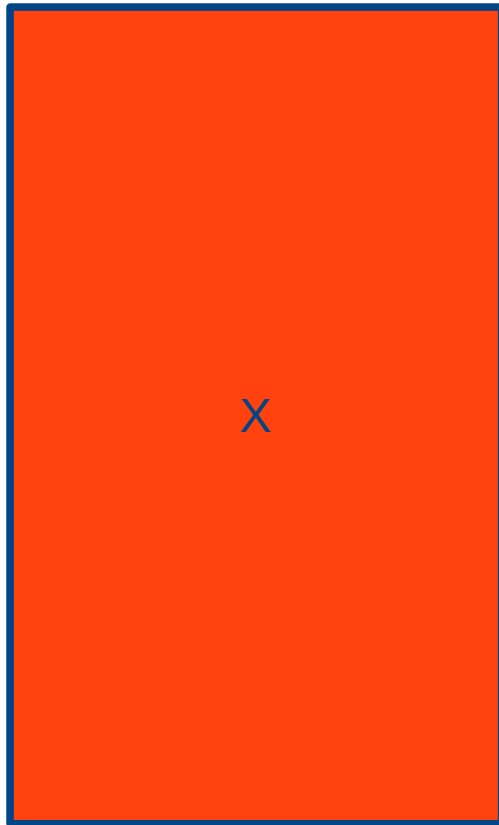
Genes (left):

- mHMC6.1AS
- mHMC6.1
- IFABP2
- IFABP1
- IFABP3
- IFABP4
- IFABP5
- IFABP6
- IFABP7
- IFABP8
- IFABP9
- IFABP10
- IFABP11
- IFABP12
- IFABP13
- IFABP14
- IFABP15
- IFABP16
- IFABP17
- IFABP18
- IFABP19
- IFABP20
- IFABP21
- IFABP22
- IFABP23
- IFABP24
- IFABP25
- IFABP26
- IFABP27
- IFABP28
- IFABP29
- IFABP30
- IFABP31
- IFABP32
- IFABP33
- IFABP34
- IFABP35
- IFABP36
- IFABP37
- IFABP38
- IFABP39
- IFABP40
- IFABP41
- IFABP42
- IFABP43
- IFABP44
- IFABP45
- IFABP46
- IFABP47
- IFABP48
- IFABP49
- IFABP50
- IFABP51
- IFABP52
- IFABP53
- IFABP54
- IFABP55
- IFABP56
- IFABP57
- IFABP58
- IFABP59
- IFABP60
- IFABP61
- IFABP62
- IFABP63
- IFABP64
- IFABP65
- IFABP66
- IFABP67
- IFABP68
- IFABP69
- IFABP70
- IFABP71
- IFABP72
- IFABP73
- IFABP74
- IFABP75
- IFABP76
- IFABP77
- IFABP78
- IFABP79
- IFABP80
- IFABP81
- IFABP82
- IFABP83
- IFABP84
- IFABP85
- IFABP86
- IFABP87
- IFABP88
- IFABP89
- IFABP90
- IFABP91
- IFABP92
- IFABP93
- IFABP94
- IFABP95
- IFABP96
- IFABP97
- IFABP98
- IFABP99
- IFABP100
- IFABP101
- IFABP102
- IFABP103
- IFABP104
- IFABP105
- IFABP106
- IFABP107
- IFABP108
- IFABP109
- IFABP110
- IFABP111
- IFABP112
- IFABP113
- IFABP114
- IFABP115
- IFABP116
- IFABP117
- IFABP118
- IFABP119
- IFABP120
- IFABP121
- IFABP122
- IFABP123
- IFABP124
- IFABP125
- IFABP126
- IFABP127
- IFABP128
- IFABP129
- IFABP130
- IFABP131
- IFABP132
- IFABP133
- IFABP134
- IFABP135
- IFABP136
- IFABP137
- IFABP138
- IFABP139
- IFABP140
- IFABP141
- IFABP142
- IFABP143
- IFABP144
- IFABP145
- IFABP146
- IFABP147
- IFABP148
- IFABP149
- IFABP150

Cell Lines (right):

- C22.6n.3
- C22.5n.3
- C20.5n.3
- C20.3n.3
- C18.3n.3
- C22.5n.6
- C22.4n.6
- C20.4n.6
- C20.3n.6
- C20.2n.6
- C18.3n.6
- C18.2n.6
- C20.3n.9
- C20.1n.9
- C18.1n.7
- C18.1n.9
- C16.1n.7
- C16.1n.9
- C18.0
- C16.0
- C14.0

# Analyse des Corrélations Canoniques



**Objectif de la méthode** : explorer les relations linéaires entre deux ensembles de variables quantitatives observées sur les mêmes individus.

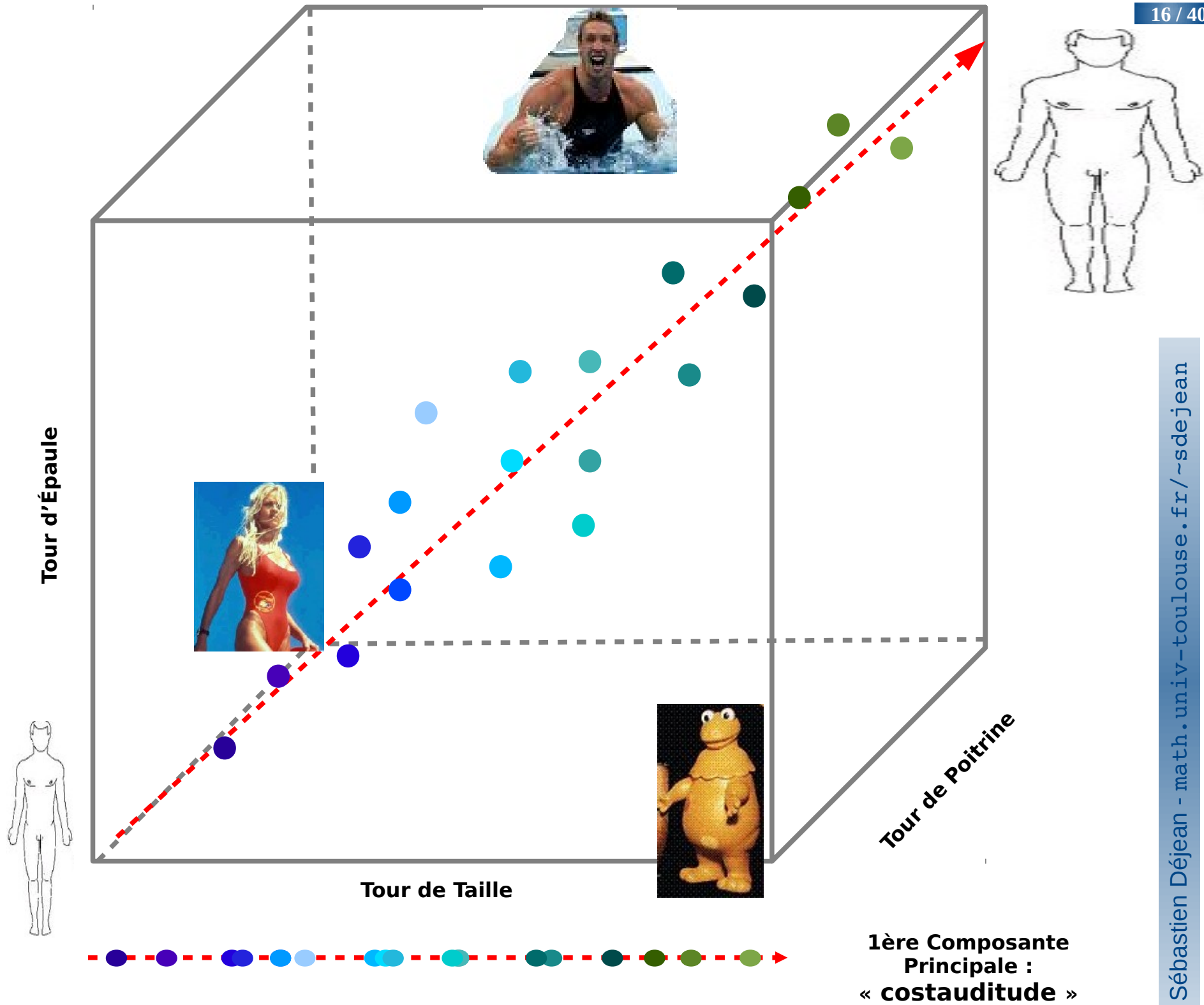
**Principe** : trouver la corrélation maximale entre une combinaison linéaire des variables X et une combinaison linéaire des variables Y pour avoir la première paire de variables canoniques. Itérer pour les suivantes.

*Principe similaire à l'Analyse en Composantes Principales (ACP) pour 1 jeu de données.*

**Limites** : ne peut fonctionner qu'avec un nombre « suffisant » d'individus.

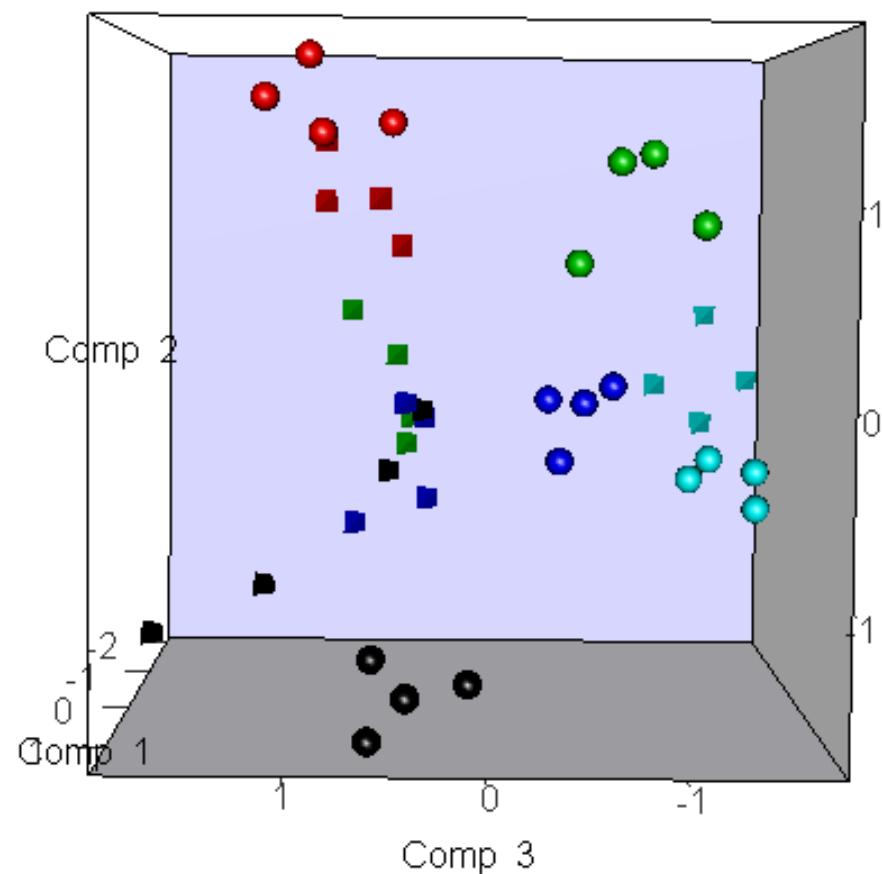
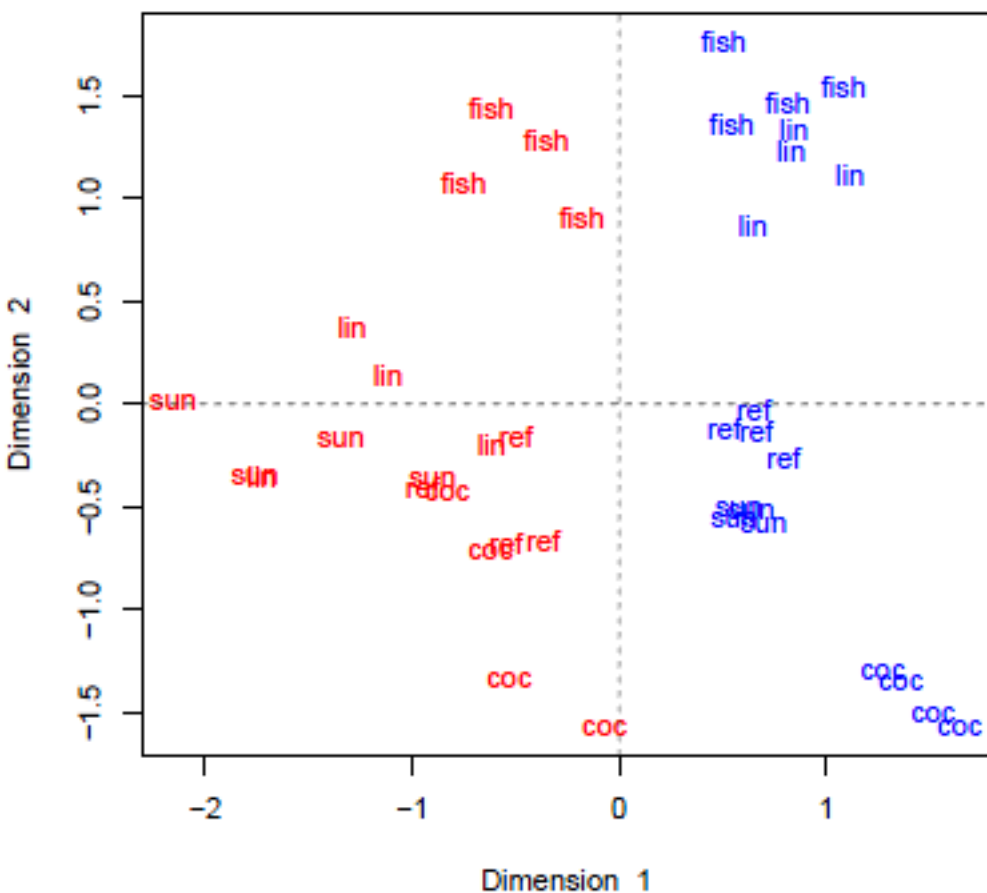
**Alternatives** : version régularisée (*ridge*) ou parcimonieuse (*sparse*).

✎ I. González, S. Déjean, P.G.P. Martin, O. Gonçalves, P. Besse, A. Baccini (2009). *Highlighting Relationships Between Heterogeneous Biological Data Through Graphical Displays Based On Regularized Canonical Correlation Analysis*. *Journal of Biological Systems*, 17(2), 173-199.

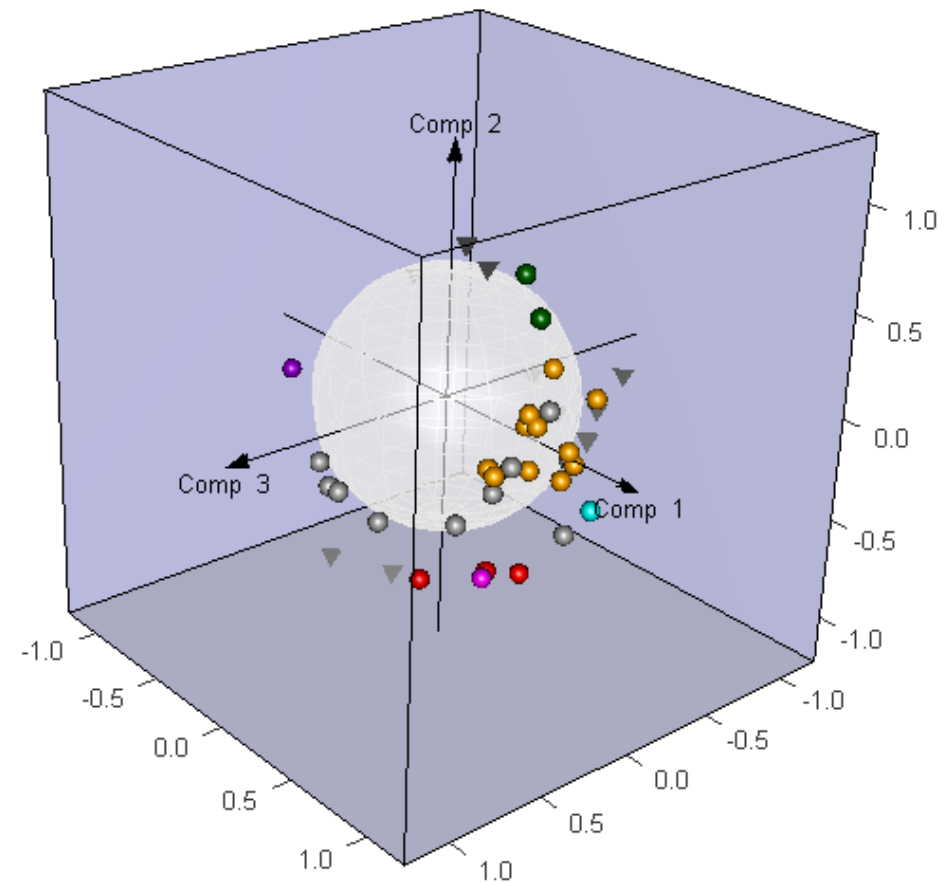




# ACC régularisée : représentation des individus (souris)



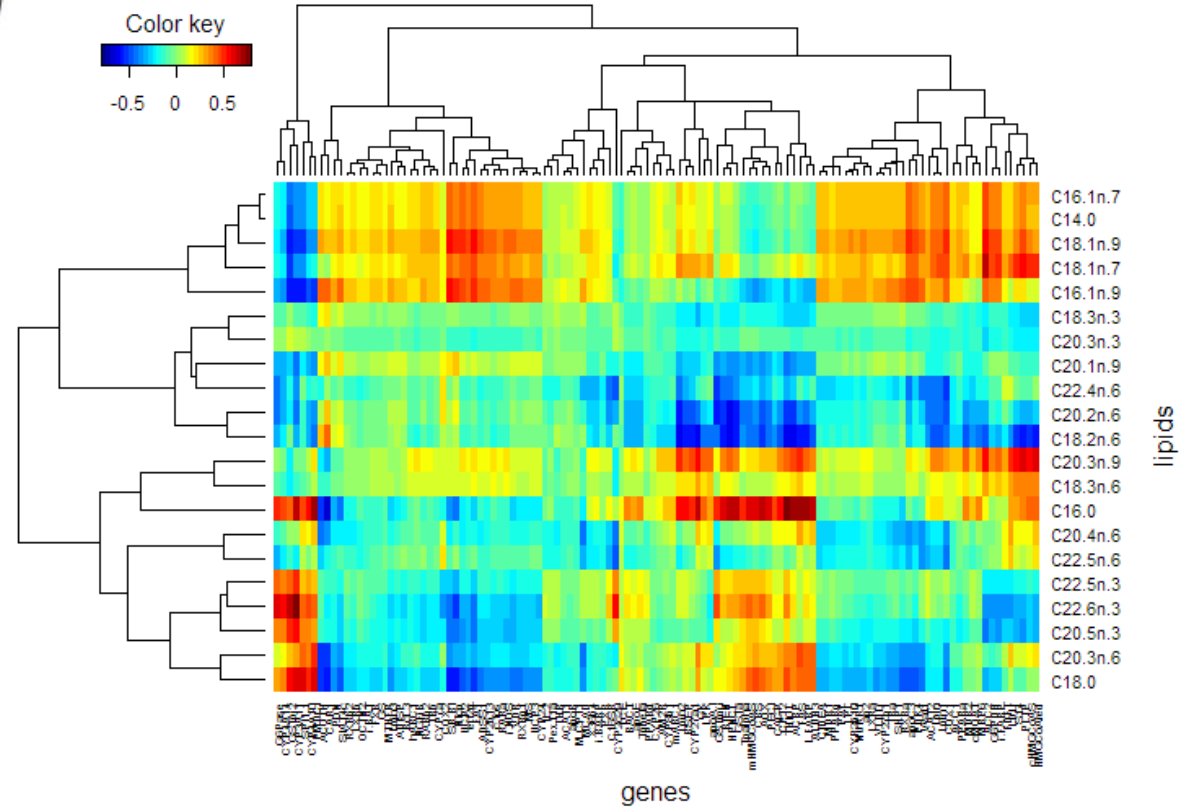
📖 Lê Cao K.-A., González I. and Déjean S. (2009) *mixOmics: an R package to unravel relationships between two omics data sets*. *Bioinformatics*, 25(21):2855-2856



Color key



-0.78 0.78



# Analyse sensorielle



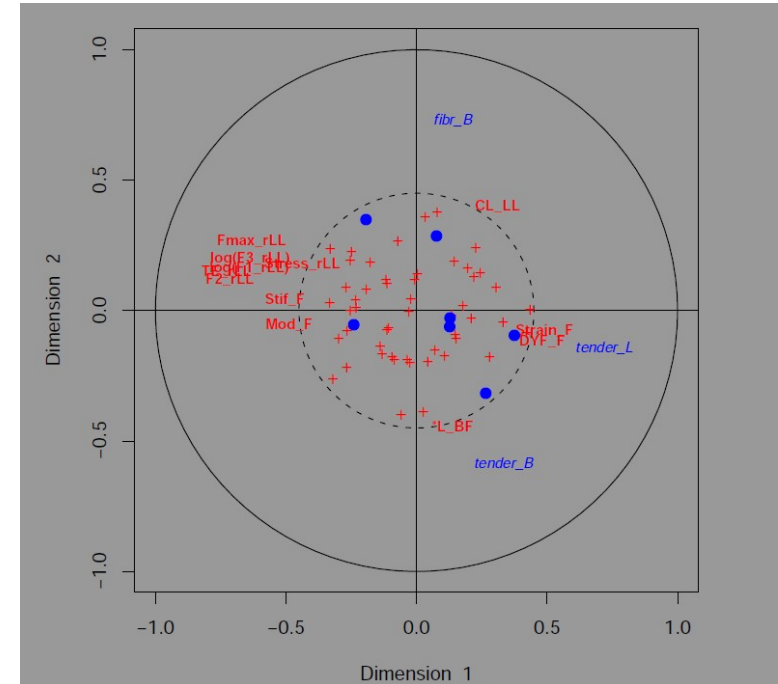
**96** experimental units (rabbits) on which were acquired 2 sets of variables

**Sensory analysis: 10** variables evaluated in back and leg by trained tasters

- tenderness
- sticky
- fibrous
- Juiciness
- flavour

**Physicochemical measurements: 63** variables concerning

- pH
- weight
- color
- water content
- cooking loss
- shear test
- ...



- ➡ Negative correlations between tenderness and measurements acquired from shear test.
- ➡ Positive correlations between tenderness in the leg and variables representing bone elasticity.

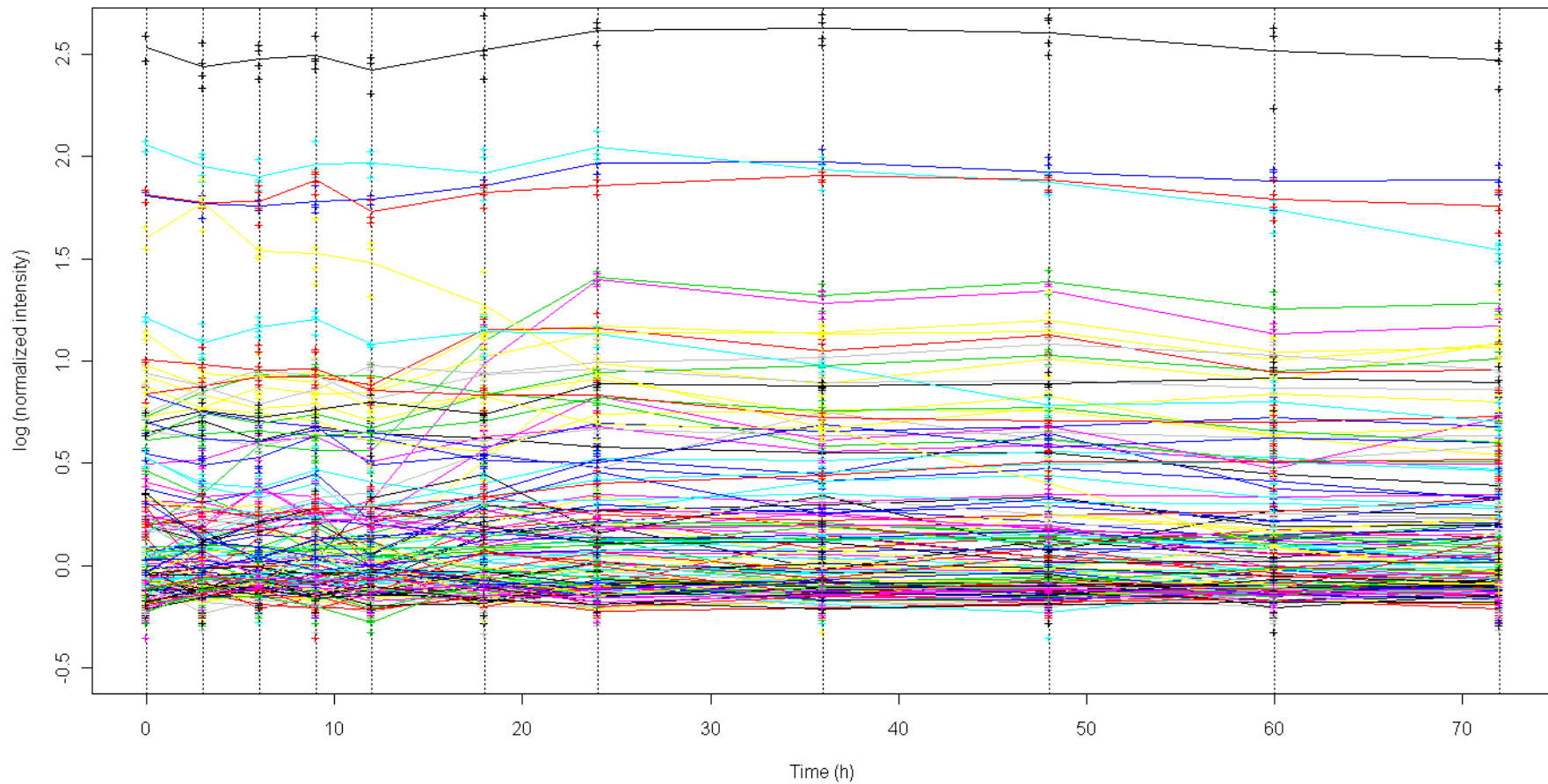
📖 S. Combes, I. González, S. Déjean, A. Baccini, N. Jehl, H. Juin, L. Cauquil, B. Gabinaud, F. Lebas, C. Larzul (2008). Relationships between sensory and physicochemical measurements in meat of rabbit from three different breeding systems using canonical correlation analysis. *Meat science*, 80(3).

# Étude cinétique du transcriptome du foie de souris soumises à un jeûne



# Les données

**44** souris (11 cages  $\times$  4 souris/cage) sont soumises à des périodes de jeûne différentes allant de 0 à 72 heures. A chaque temps sélectionné (0, 3, 6, 9, 12, 18, 24, 36, 48, 60 et 72 heures), 4 souris sont euthanasiées. Pour chaque souris, l'expression de **120** gènes est mesurée dans des cellules du foie.

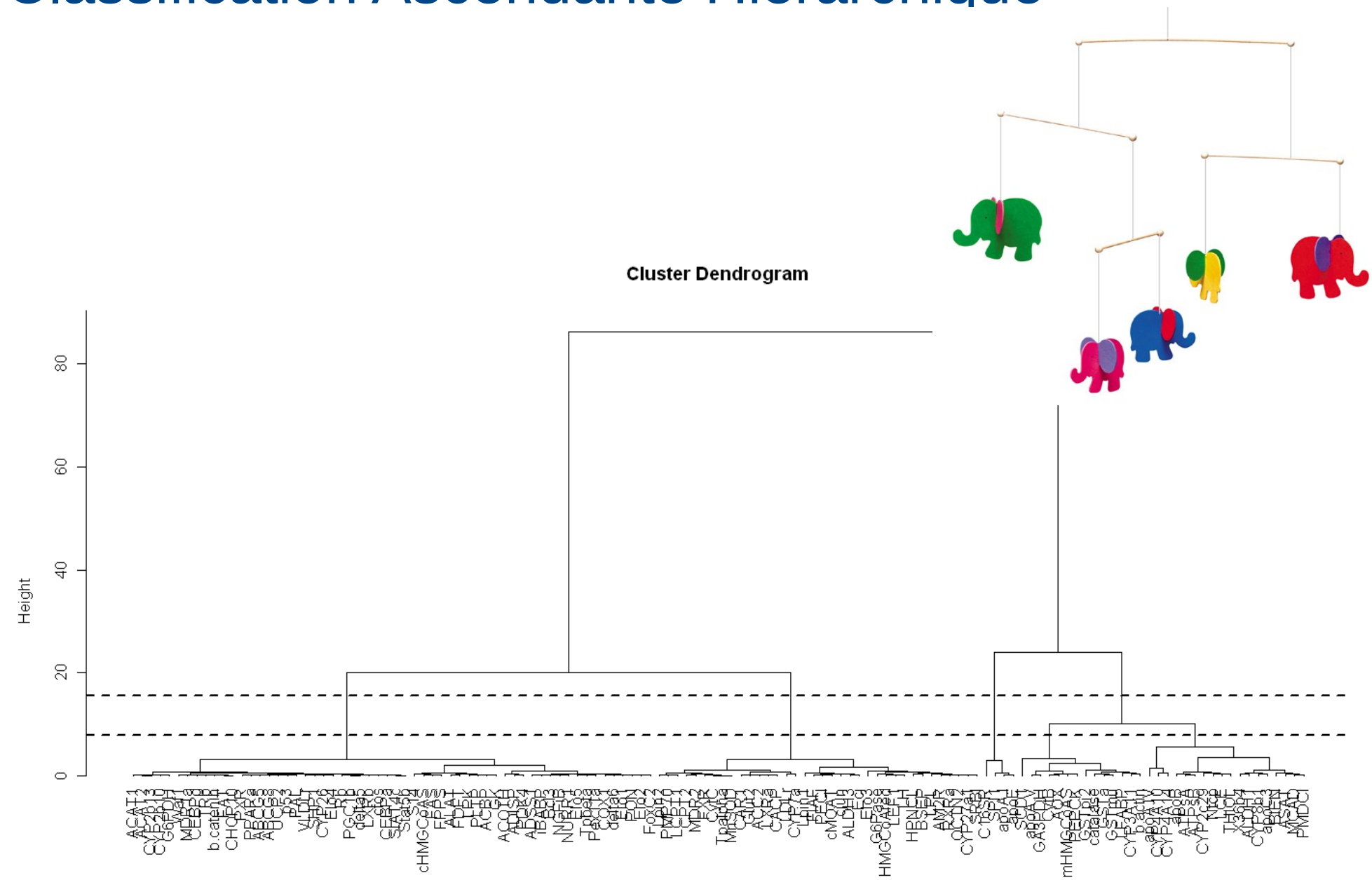


 S. Déjean, P. Martin, A. Baccini, P. Besse (2007). Clustering time series gene expression data using smoothing spline derivatives. *EURASIP Journal on Bioinformatics and Systems Biology*

# Problème

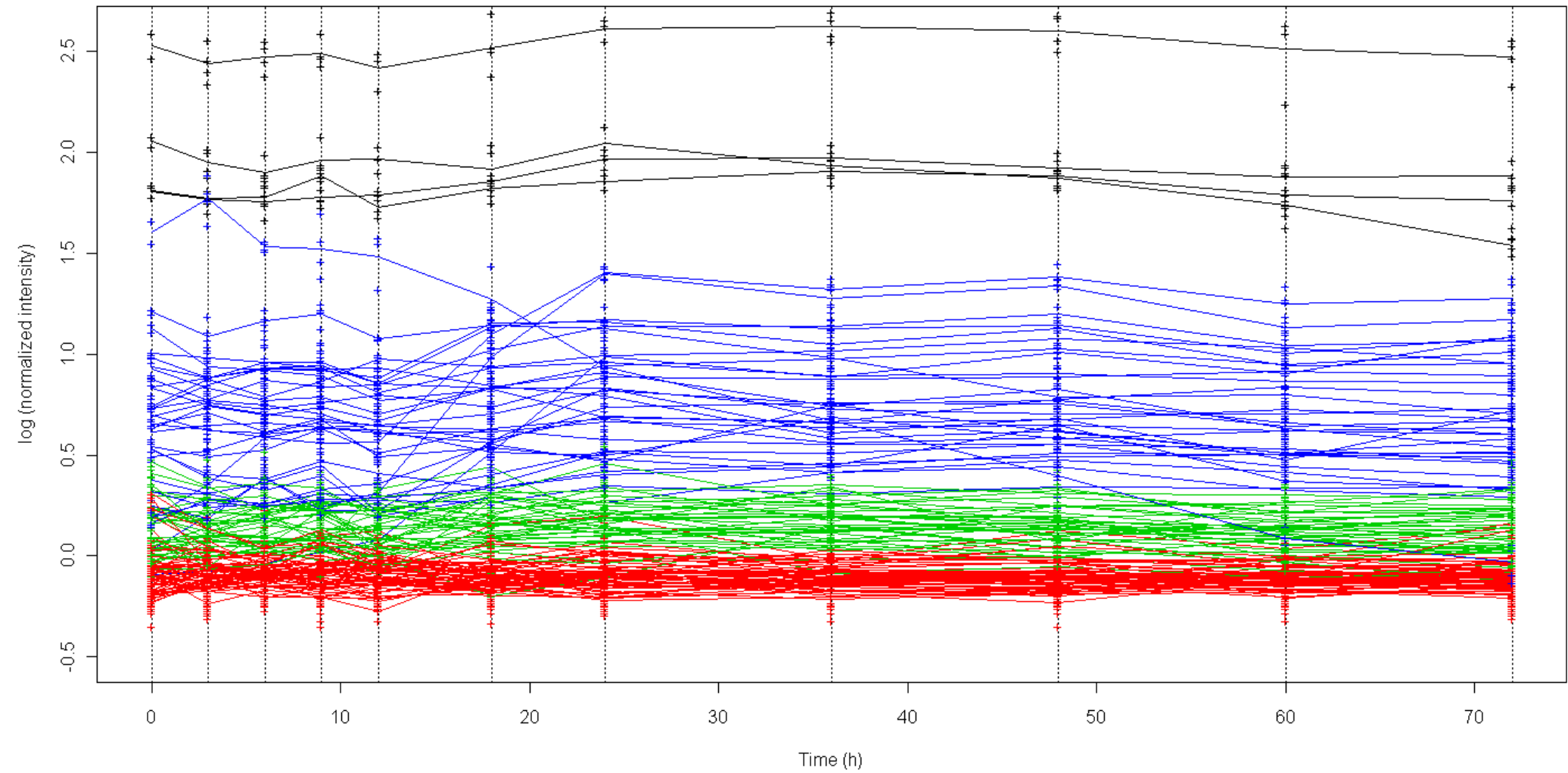
Proposer une méthode permettant de regrouper les gènes selon leur profil d'expression au cours du temps.

# Classification Ascendante Hiérarchique



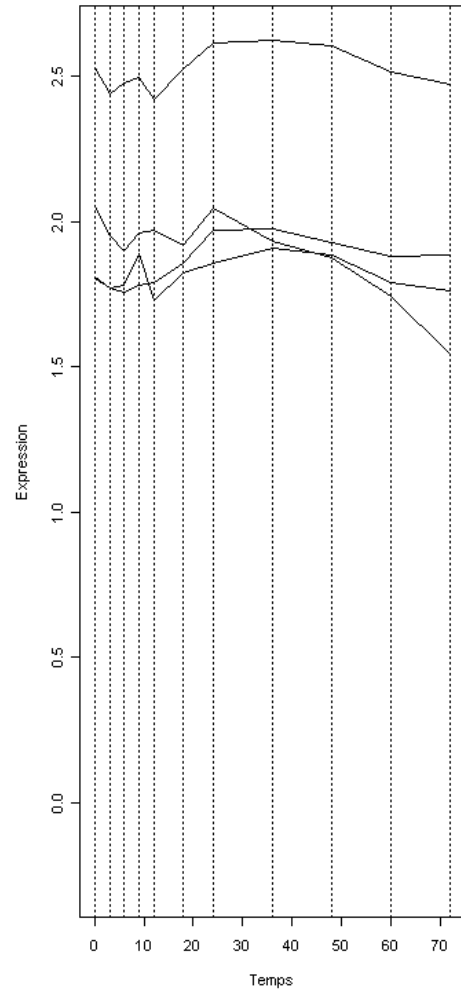


# CAH - 4 groupes

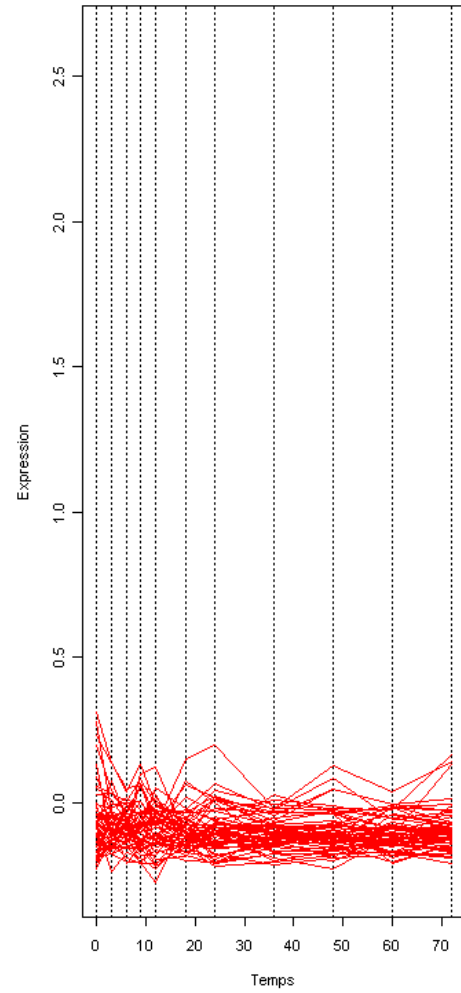


# CAH - 4 groupes

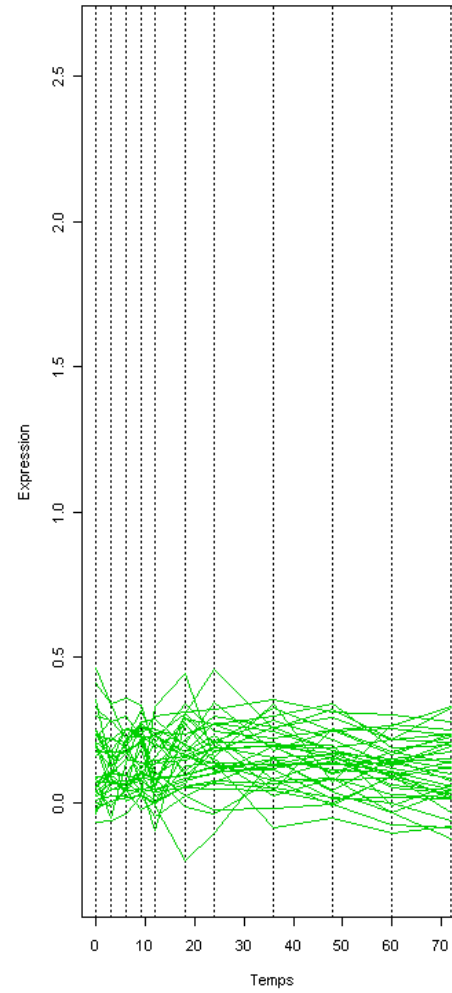
Groupe 1



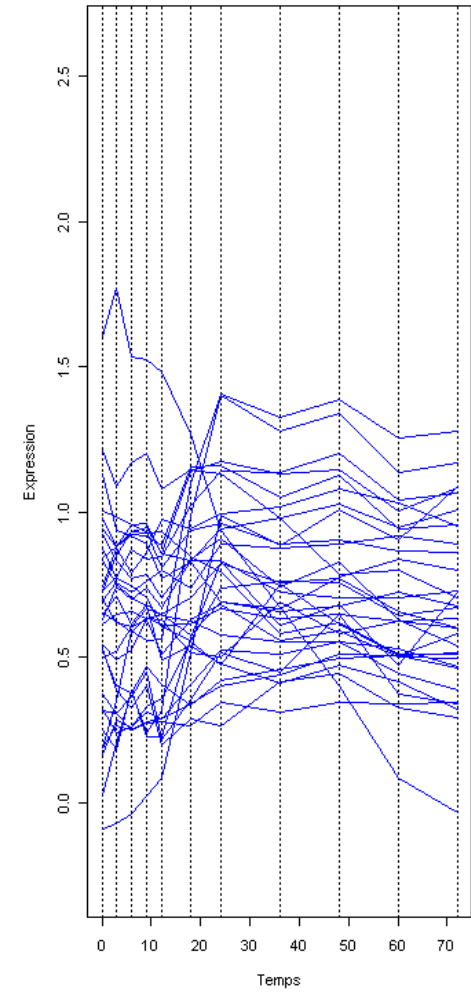
Groupe 2



Groupe 3



Groupe 4



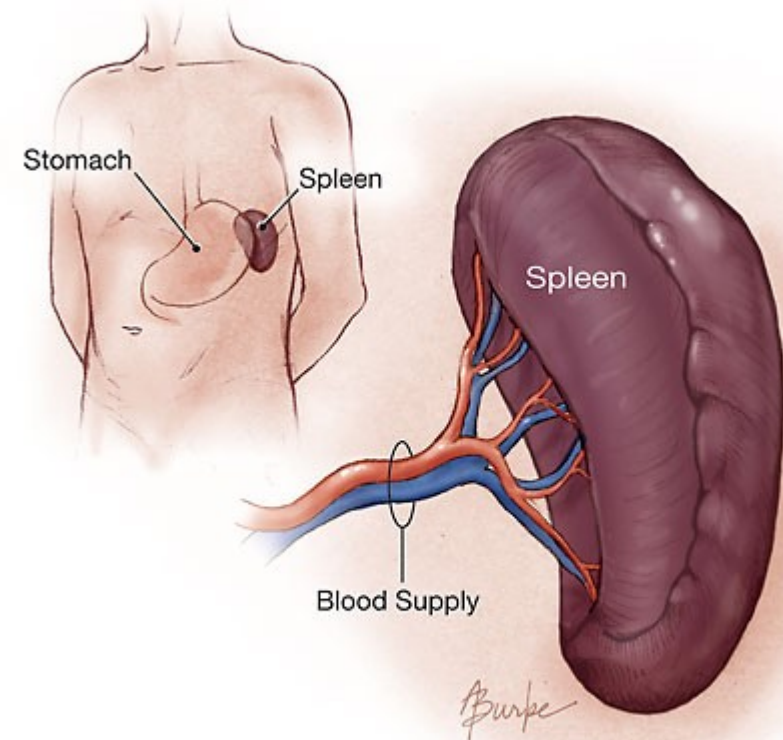
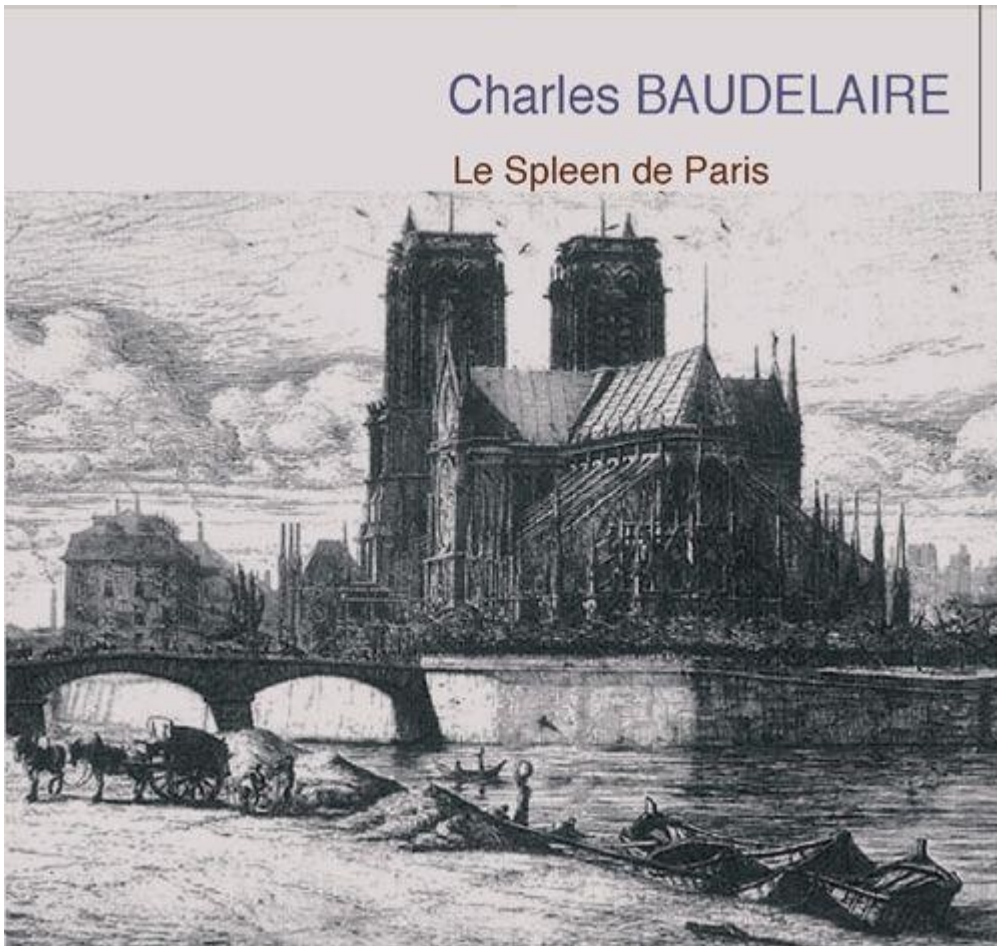
Avons-nous répondu à la question ?

# Lissage *spline*

## spline $\neq$ spleen

Charles BAUDELAIRE

Le Spleen de Paris



# Lissage

✍ S. Déjean, P. Martin, A. Baccini, P. Besse (2007). Clustering time series gene expression data using smoothing spline derivatives. *EURASIP Journal on Bioinformatics and Systems Biology*

## 3.1. Signal extraction

Rather than directly computing means of the observed values as in Figure 1, we tried a somewhat more realistic approach based on two essential assumptions:

(i) the values at each time point are noisy observations of the “true” value (obviously unknown),

(ii) this type of biological phenomenon should be a regular, and so differentiable, function of time. This means for us without singularities or any chaotic behavior. This is a sensible assumption when data are acquired at a macroscopic level; it may be false at a molecular or a single-cell scale. Furthermore, in this study, fasting is typically a progressive stimulus where hormonal changes take place progressively and should not imply biological thresholds.

This led us to consider the following nonparametric model for each gene expression:

$$y_i^j = f(t_j) + \varepsilon_{ij}, \quad i = 1, \dots, 4, j = 1, \dots, 11, \quad (1)$$

This problem is classically solved by a nonparametric estimation of  $f$ . Kernel smoothing or spline smoothing both achieve this objective, but we naturally preferred spline smoothing since we needed to estimate both the function and its derivative. This is quite easy using cubic spline smoothing. The estimation of any gene expression curve according to this model is then the solution to the following optimization problem [17]:

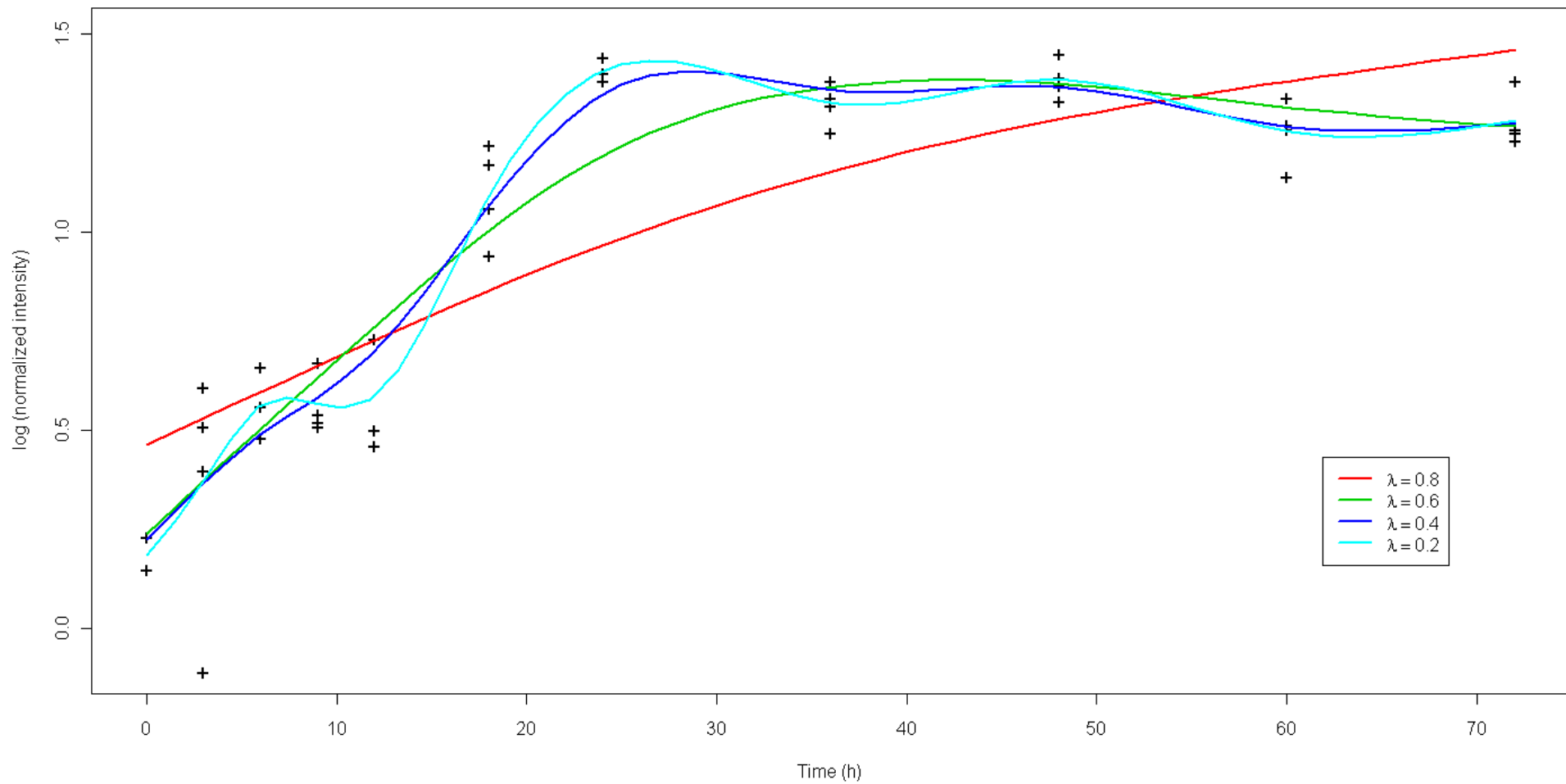
$$\min_{f \in H_1} \frac{1}{4 \times 11} \sum_{i=1,4; j=1,11} [y_i^j - f(t_j)]^2 + \lambda \int_{t_1}^{t_{11}} [f''(u)]^2 du, \quad (3)$$

where  $f$  belongs to  $H_1$ , the Sobolev space of continuous functions with integrable squared second derivative, and  $\lambda$  is the smoothing parameter. This parameter balances the influence between the left-hand term of (3), which forces solutions to be close to mean values, and the right-hand one, which controls the regularity of the function.

The solution  $\hat{f}$  of (3) is a piecewise function which is defined on the basis of cubic polynomials. The solution shape and its smoothness depend directly on  $\lambda$ . On the one hand, as  $\lambda$  grows, the solution converges to a trivial linear regression since the integral in the right-hand term of (3) tends to zero (with the second derivative). On the other hand, if  $\lambda$  decreases towards zero, the solution becomes a piecewise polynomial interpolating function of the means of the four values at each time point since the left-hand term reaches its minimum value.

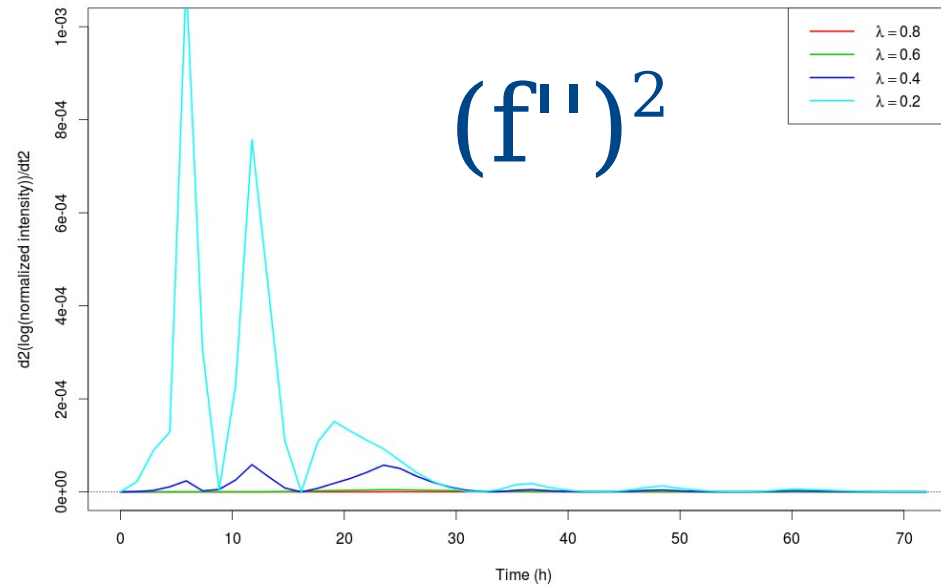
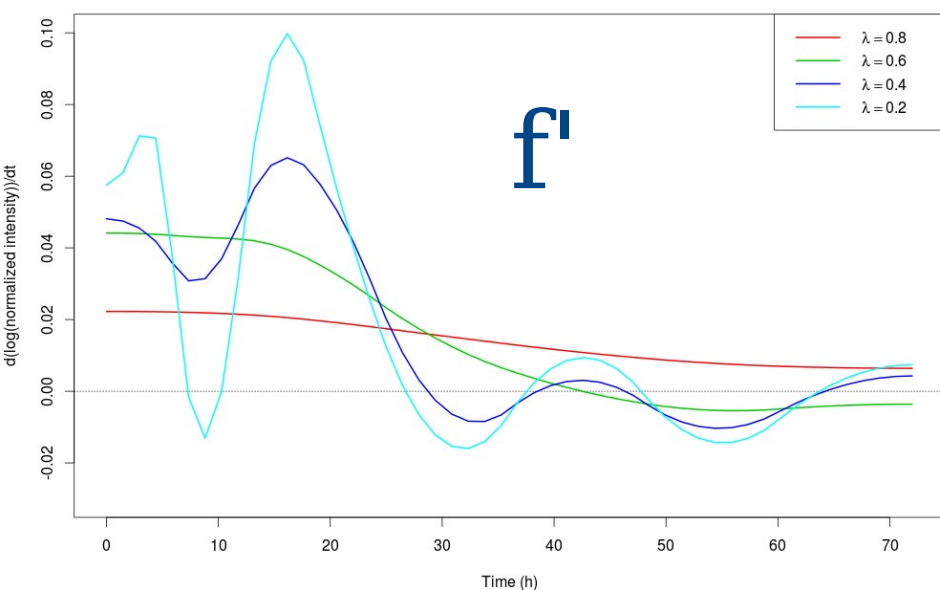
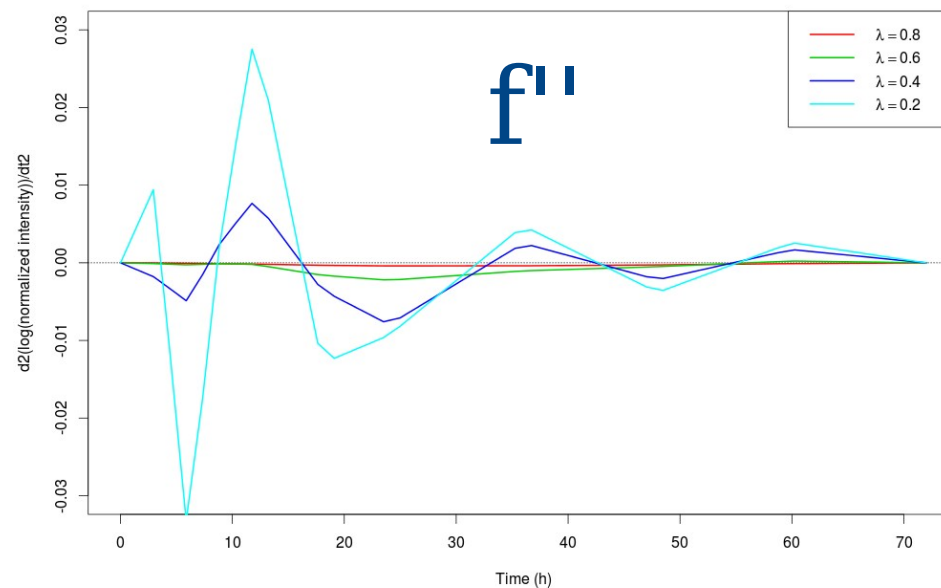
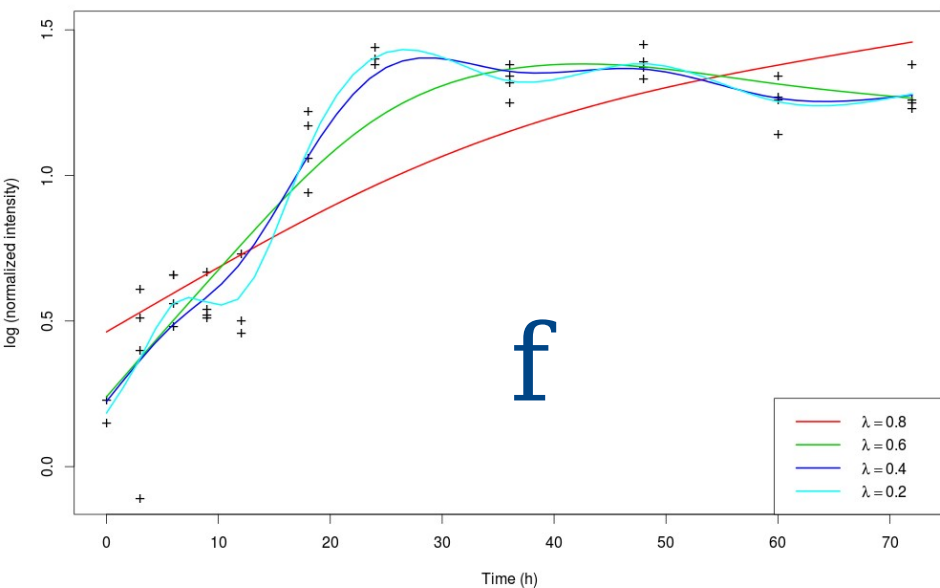


# Lissage

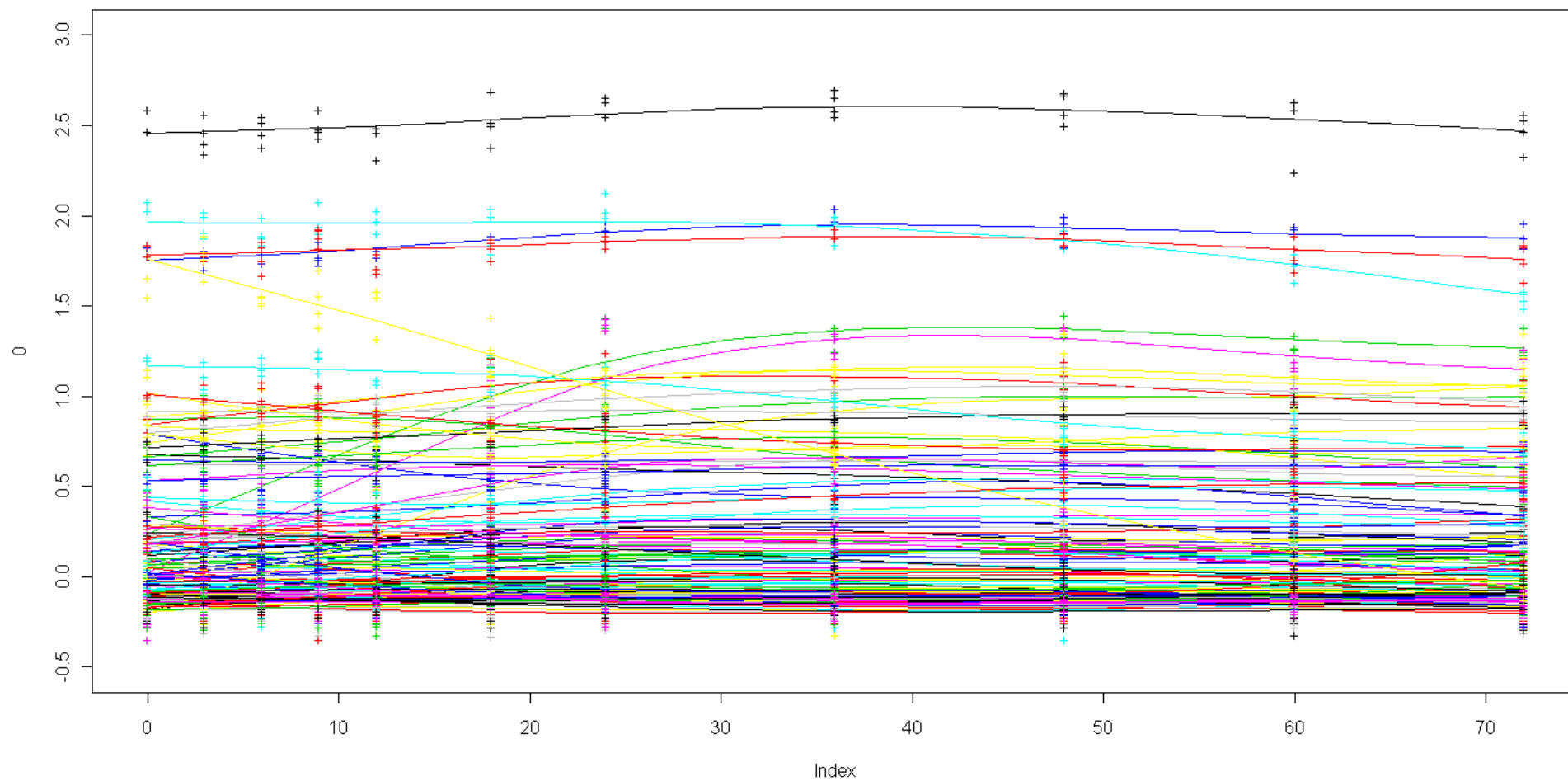


$$\min_{f \in H_1} \frac{1}{4 \times 11} \underbrace{\sum_{i=1,4; j=1,11} [y_i^j - f(t_j)]^2}_{\text{Proche des données}} + \underbrace{\lambda \int_{t_1}^{t_{11}} [f''(u)]^2 du}_{\text{Régulière}},$$

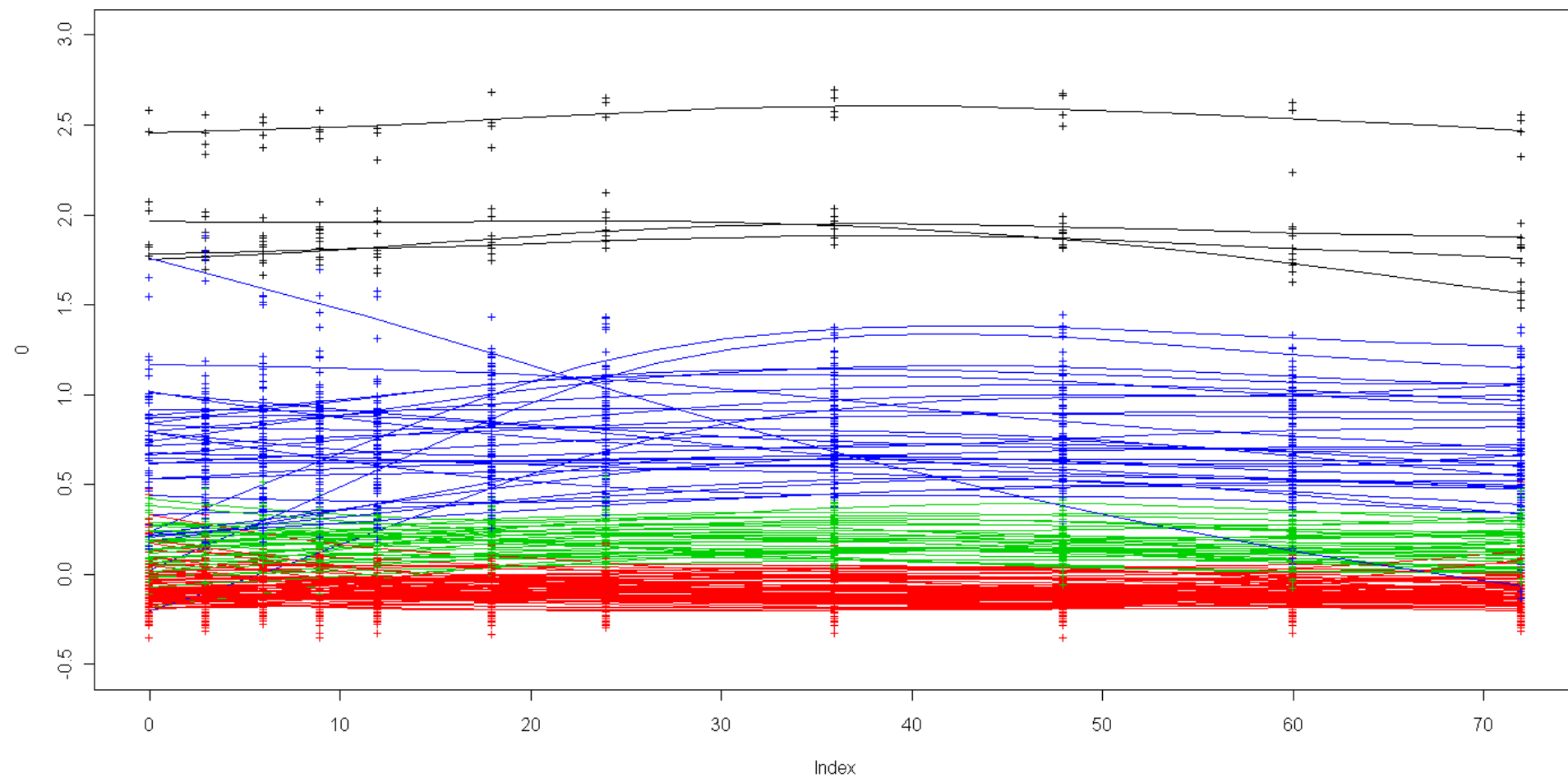
# Dérivée seconde ?!?



# Données lissées

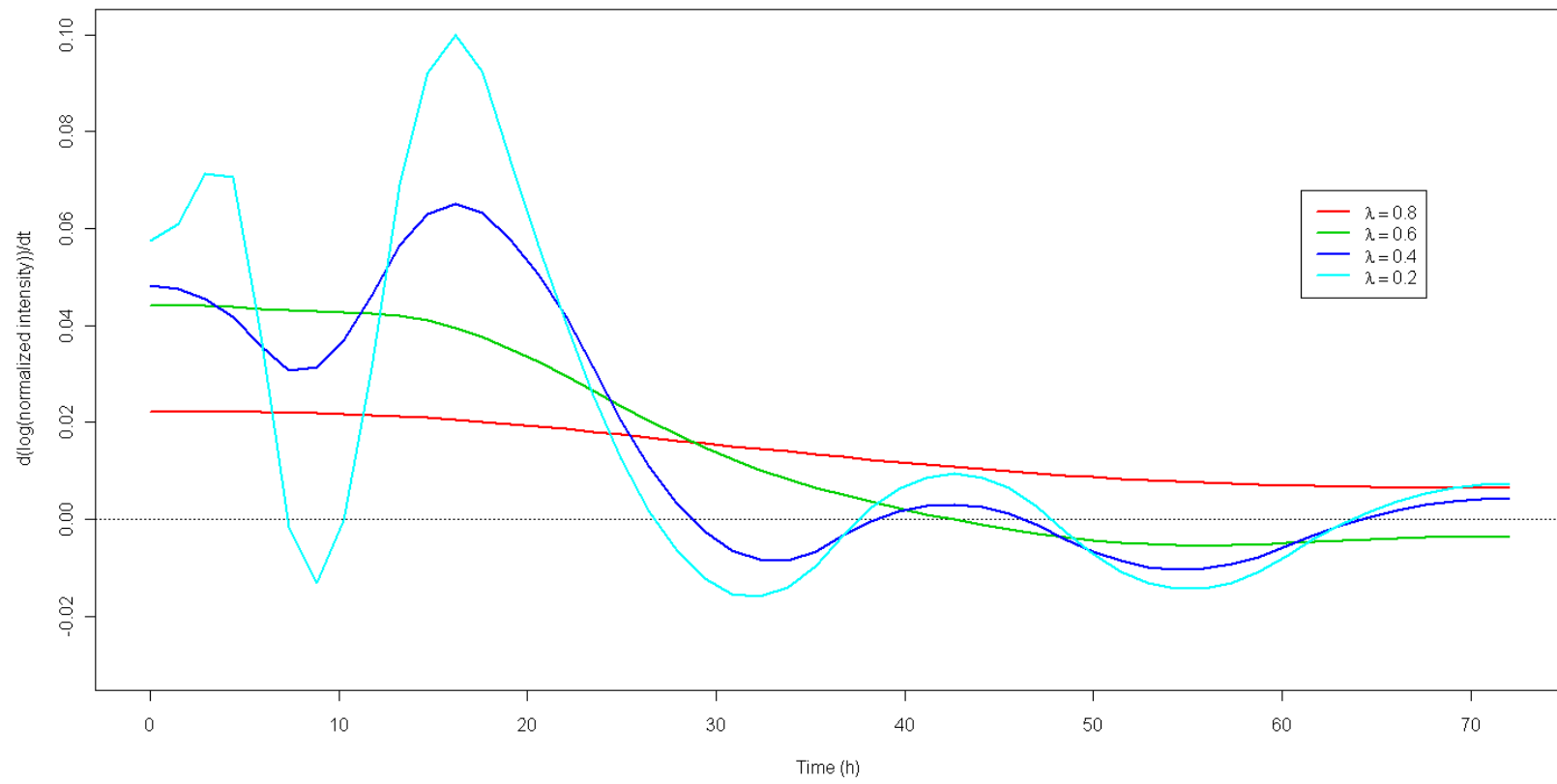
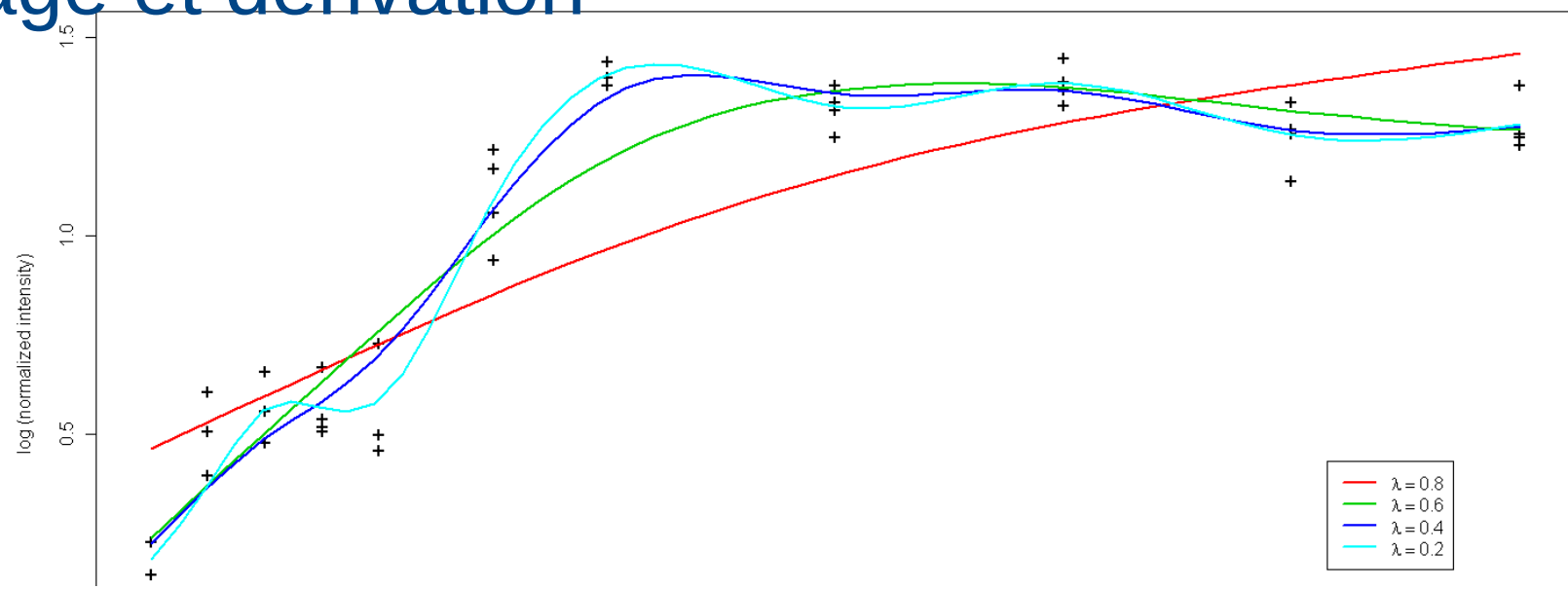




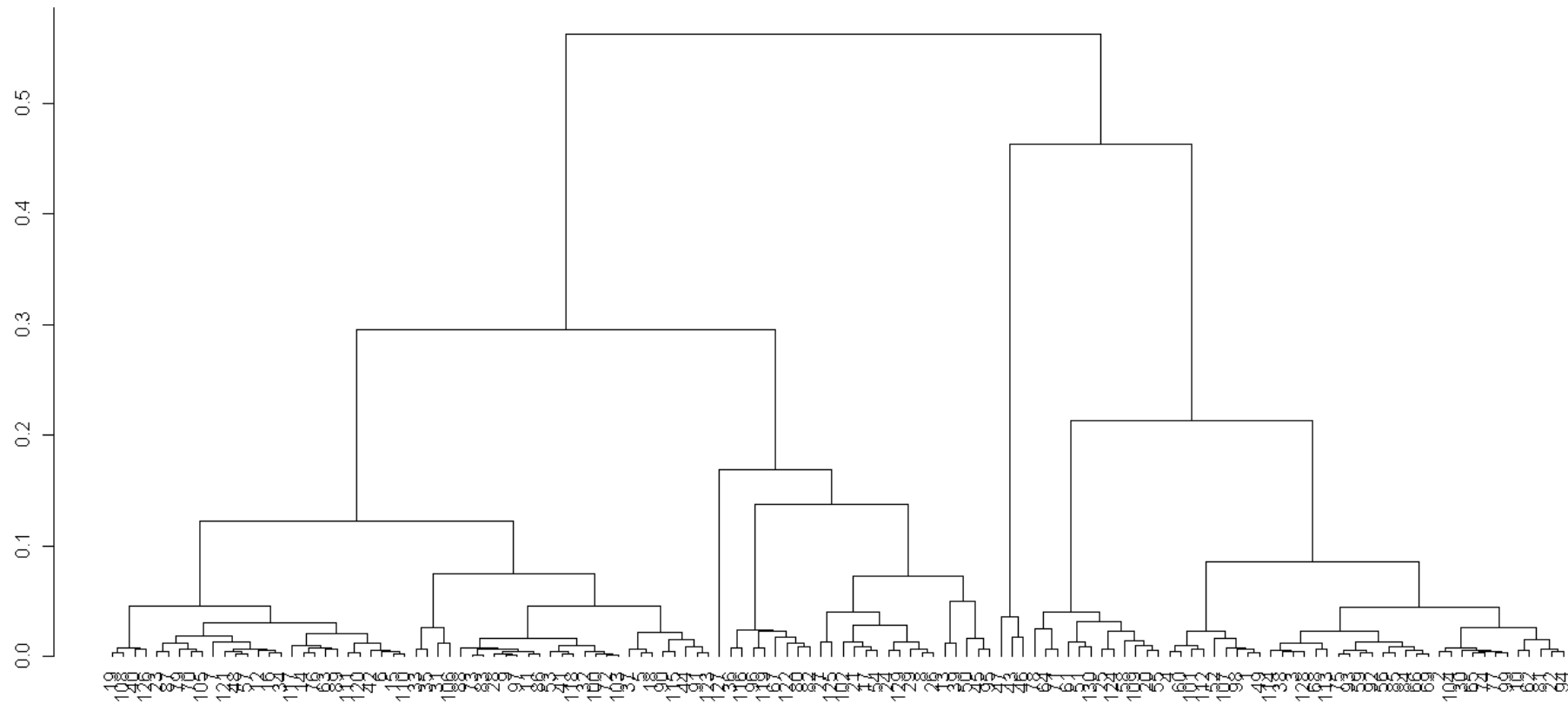


Et là, c'est mieux ?

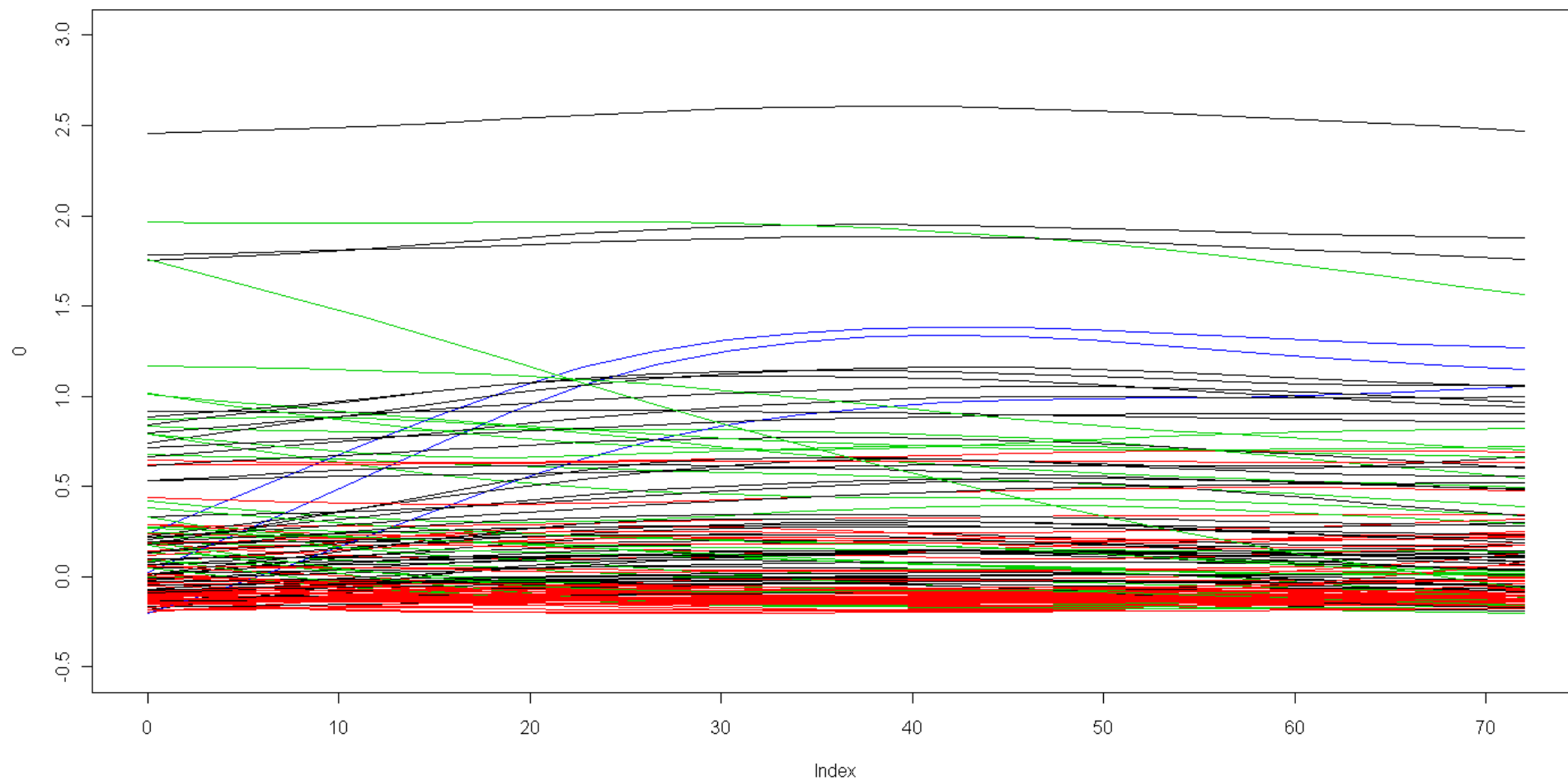
# Lissage et dérivation



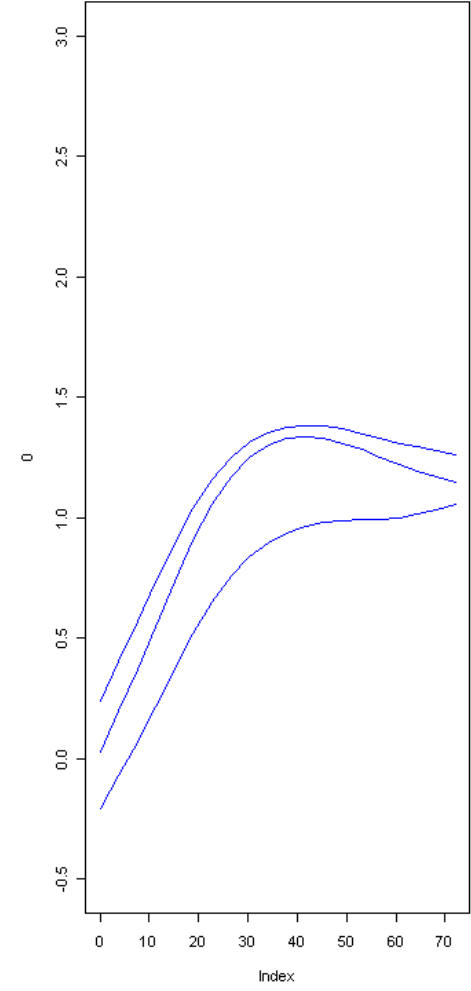
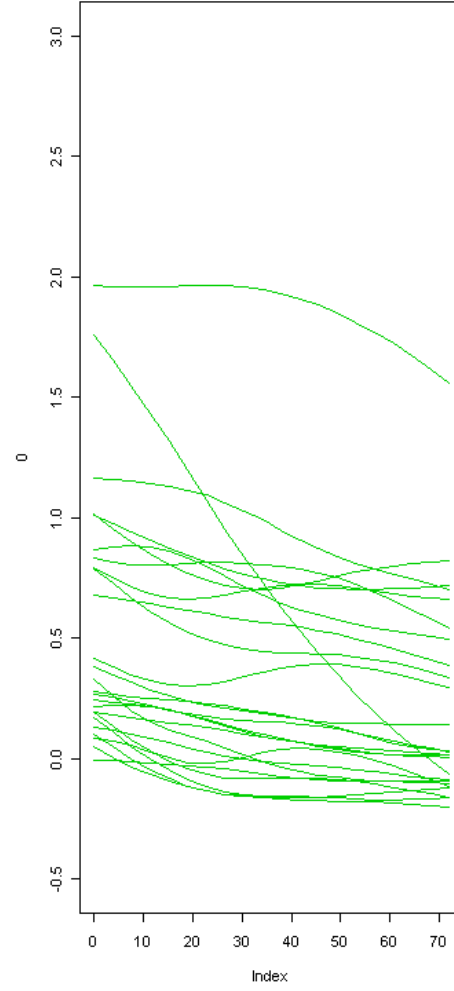
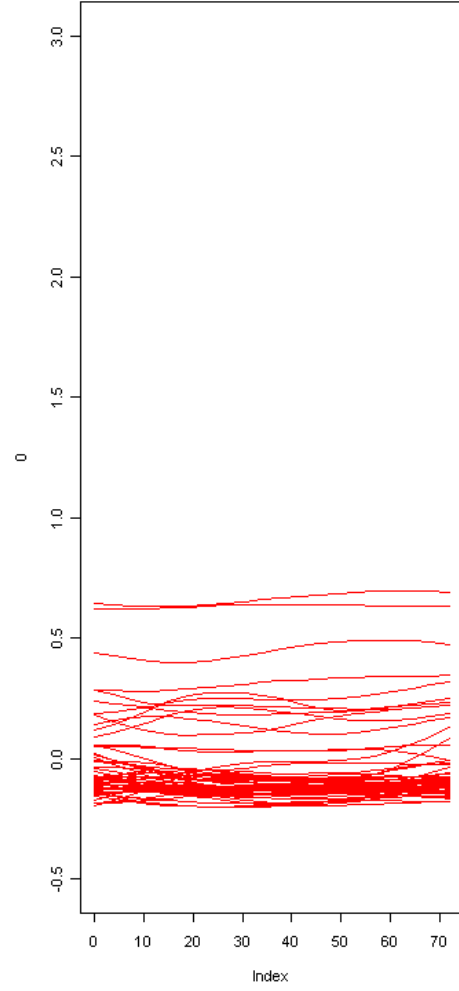
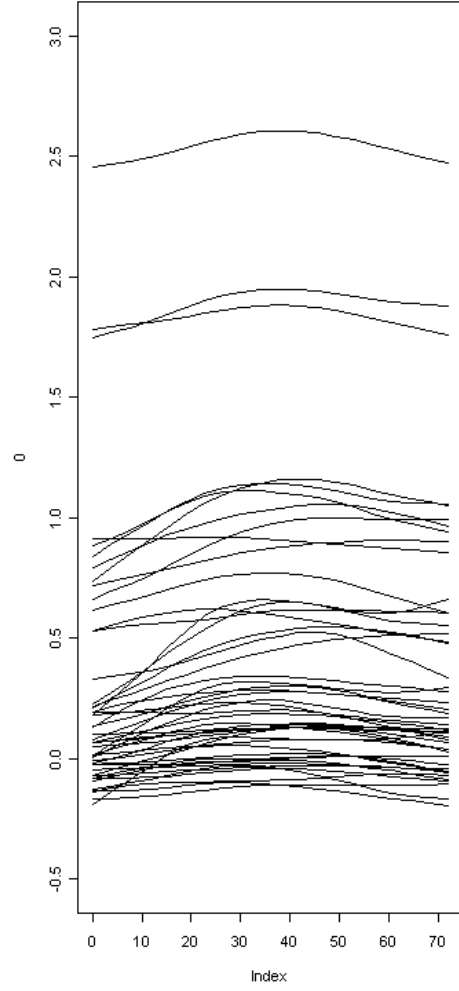
# CAH sur données lissées et dérivées



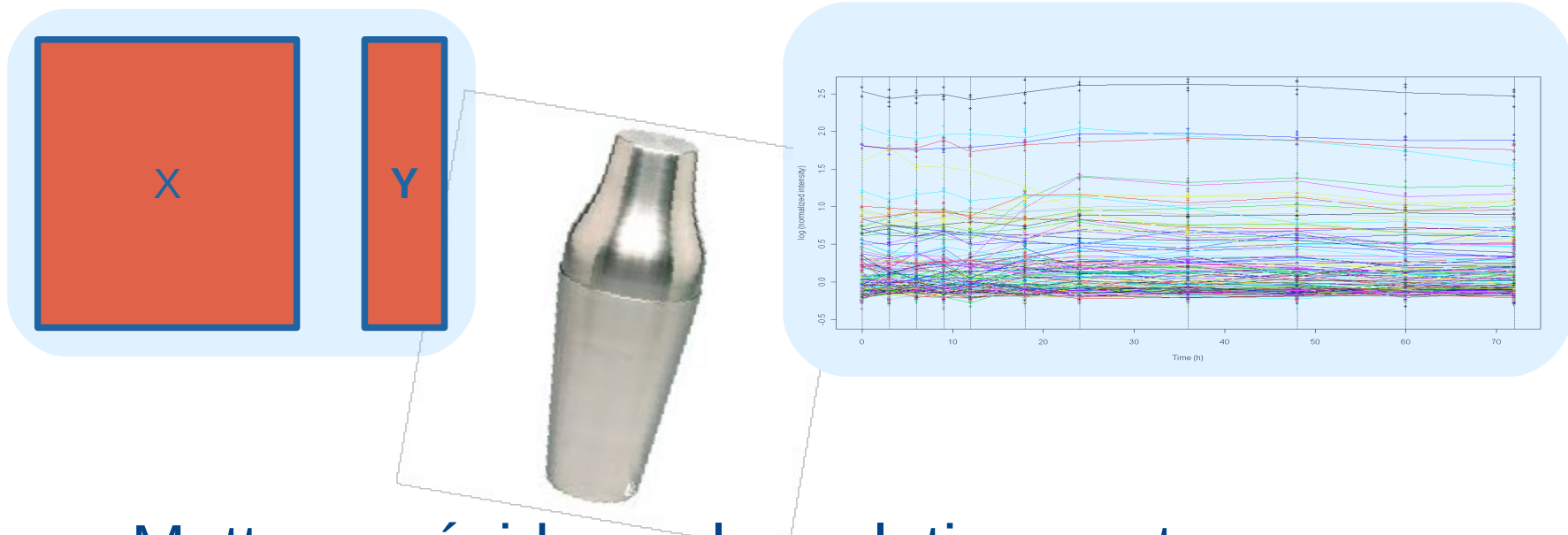
# CAH – 4 groupes



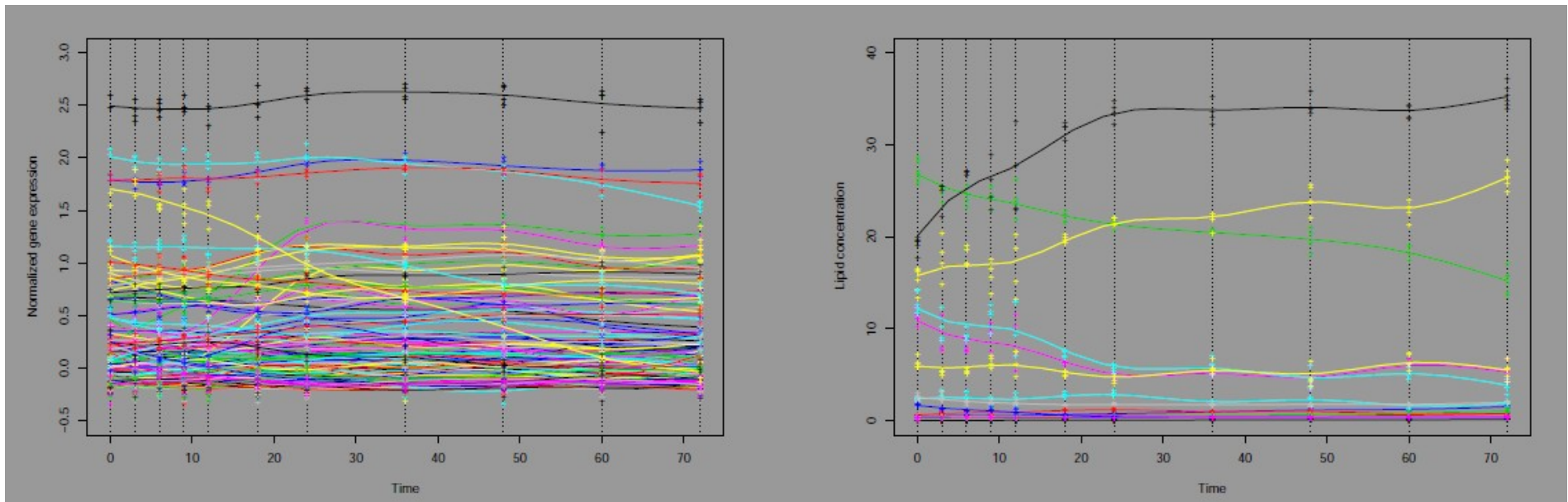
# CAH – 4 groupes



# Pour lier le tout...



Mettre en évidence les relations entre  
2 ensembles de profils temporels





INSTITUT  
de MATHÉMATIQUES  
de TOULOUSE



# Intégration de données « omiques »

*Comment lier les ingrédients d'une recette*

Sébastien Déjean

[www.math.univ-toulouse.fr/~sdejean/](http://www.math.univ-toulouse.fr/~sdejean/)