UMR PEGASE

GenPhySE
Génétique Physiologie et Systèmes d'Elevage

# Longitudinal analysis of adaption mechanisms to heat stress in pig, using omic data*

*** Ongoing work**

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

INRAE

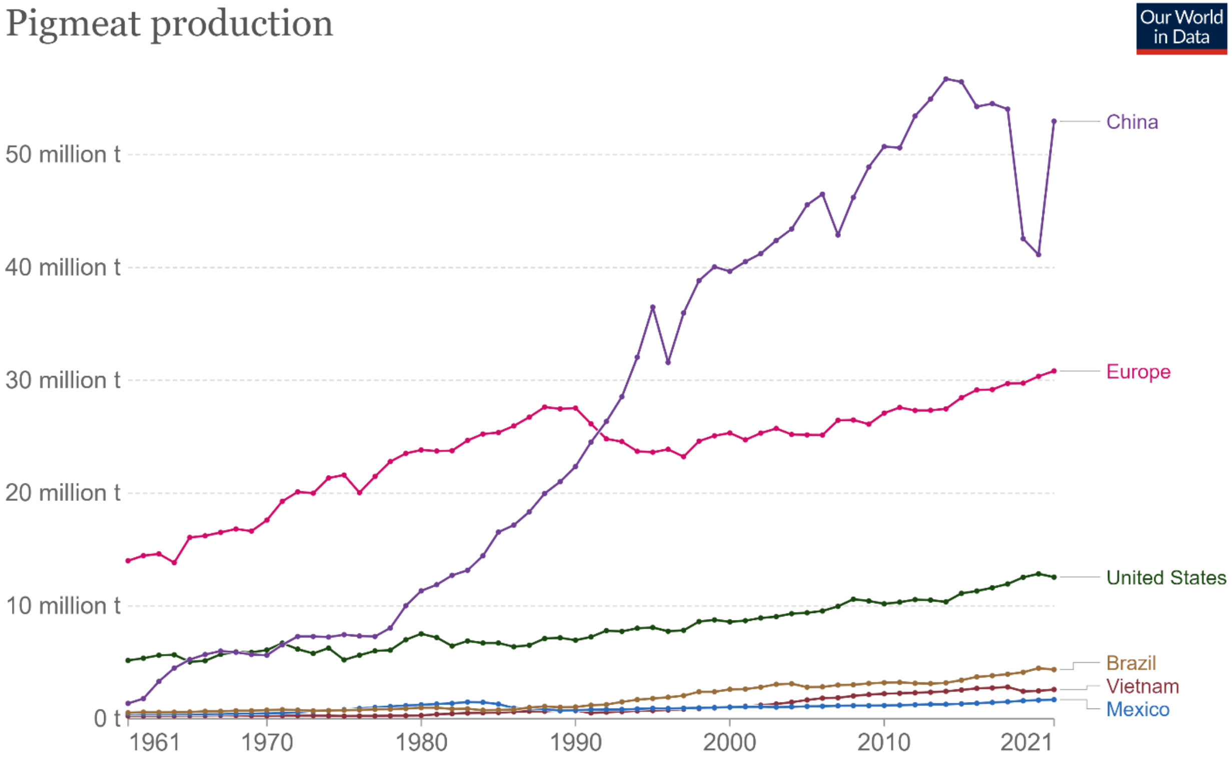
DOCTORAT BRETAGNE LOIRE
ECOLOGIE GEOSCIENCES AGRONOMIE ALIMENTATION

Phd student: **Guilhem Huau**
Co-director: Laurence Liaubet
Co-director: David Renaudeau

# Production ⇗
# in tropical areas ⇗

# Consumption ⇗

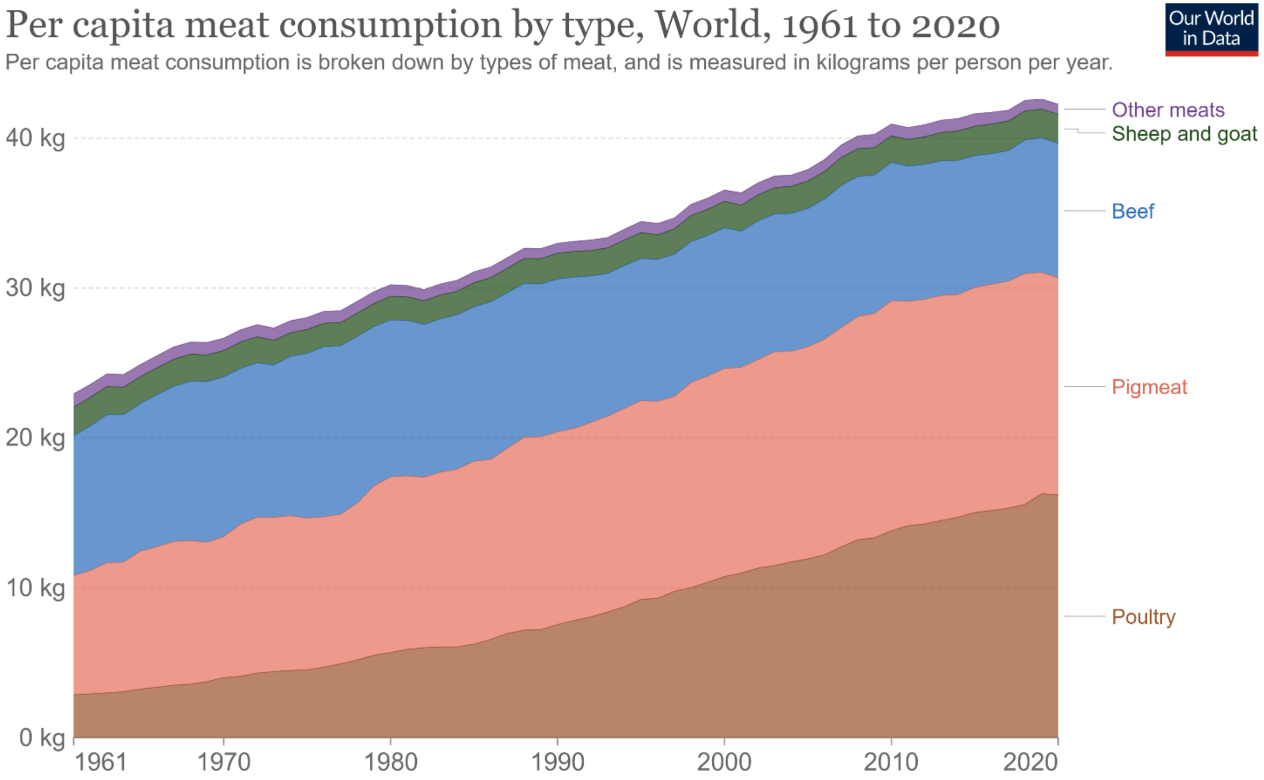
## Pigmeat production



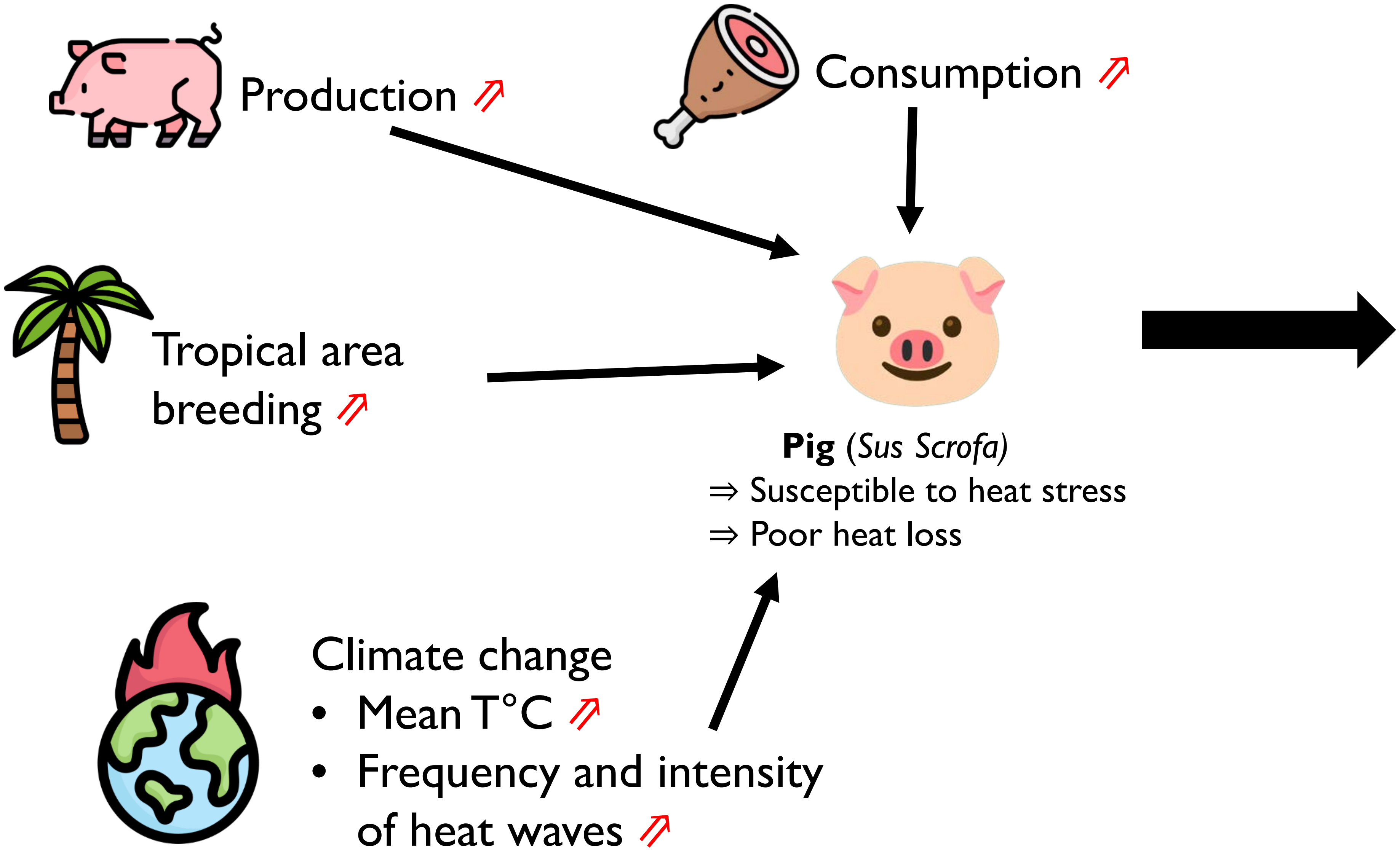
Source: Food and Agriculture Organization of the United Nations
OurWorldInData.org/meat-production • CC BY
Note: This refers to total meat production, from both commercial and farm slaughter. Data are given in terms of dressed carcass weight, excluding offal and slaughter fats.

## Per capita meat consumption by type, World, 1961 to 2020

Per capita meat consumption is broken down by types of meat, and is measured in kilograms per person per year.



Source: Food and Agriculture Organization of the United Nations
OurWorldInData.org/meat-production • CC BY
Note: Data does not include fish and seafood. Figures do not correct for waste at the consumption level so may not directly reflect the quantity of food finally consumed by a given individual.



Production ⇗

Consumption ⇗

Tropical area breeding ⇗

**Pig** (*Sus Scrofa*)
⇒ Susceptible to heat stress
⇒ Poor heat loss

Climate change

- Mean T°C ⇗
- Frequency and intensity of heat waves ⇗

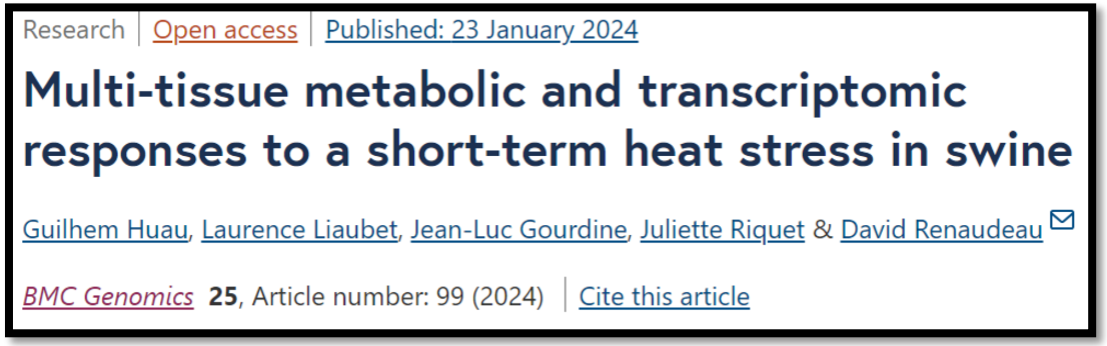
**Adaptation mechanisms**

# Variability of the adaptation mechanisms

- Animal related
  - **Breed**
  - Sex
  - Age
  - **Tissue-specific**

- Stress related
  - Intensity
  - **Duration**
  - Periodicity

- Mechanisms from one situation to the other
  - Similar
  - Contrasting
  - Absent
  - Etc.

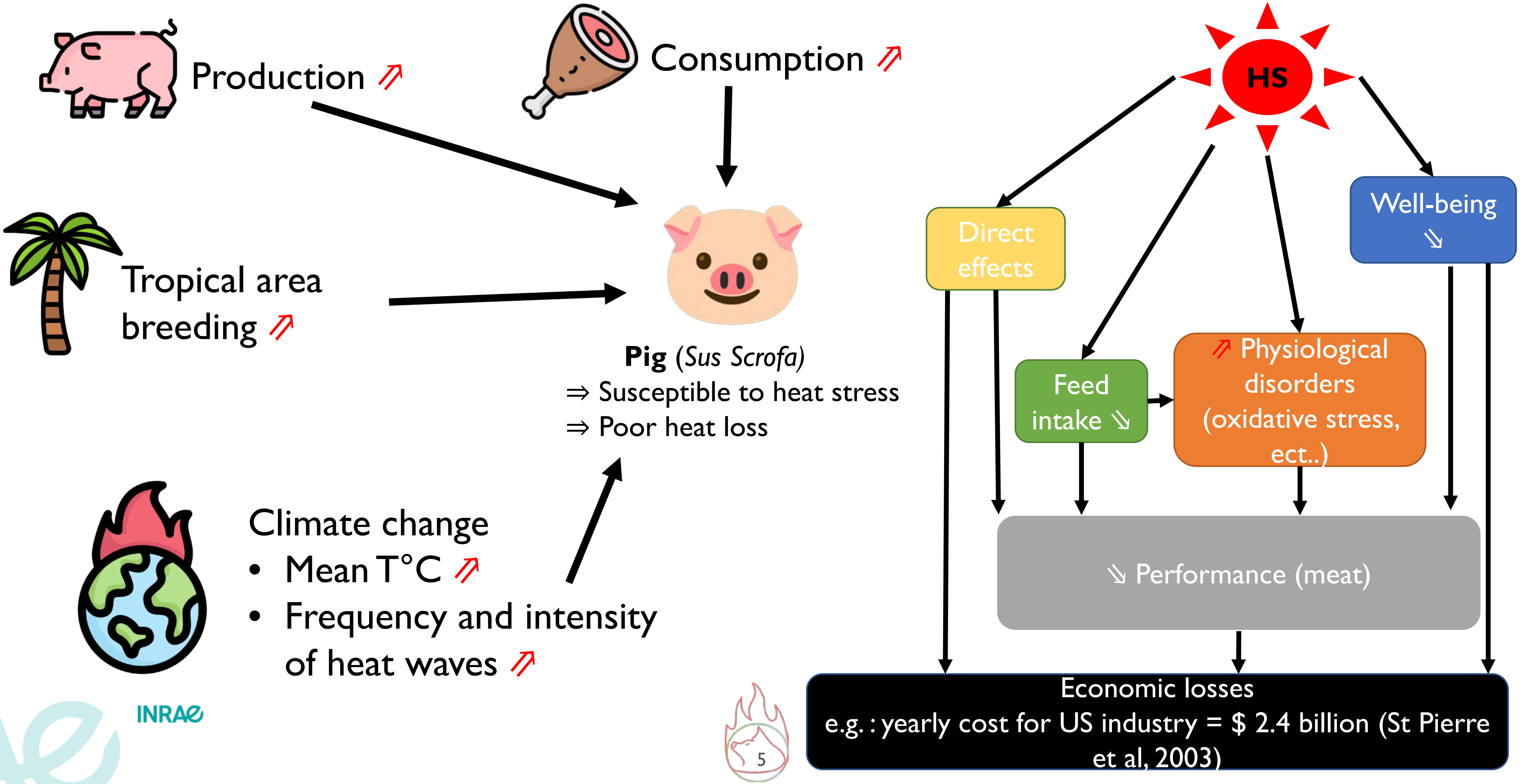


Research | Open access | Published: 23 January 2024

**Multi-tissue metabolic and transcriptomic responses to a short-term heat stress in swine**

Guilhem Huau, Laurence Liaubet, Jean-Luc Gourdine, Juliette Riquet & David Renaudeau

*BMC Genomics* **25**, Article number: 99 (2024) | Cite this article



Production ⇗

Consumption ⇗

Tropical area breeding ⇗

**Pig** (*Sus Scrofa*)
⇒ Susceptible to heat stress
⇒ Poor heat loss

Climate change
- Mean T°C ⇗
- Frequency and intensity of heat waves ⇗

HS

Direct effects

Feed intake ⇘

⇗ Physiological disorders (oxidative stress, ect..)

Well-being ⇘

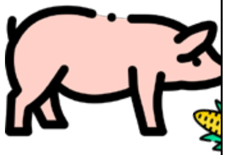
⇘ Performance (meat)

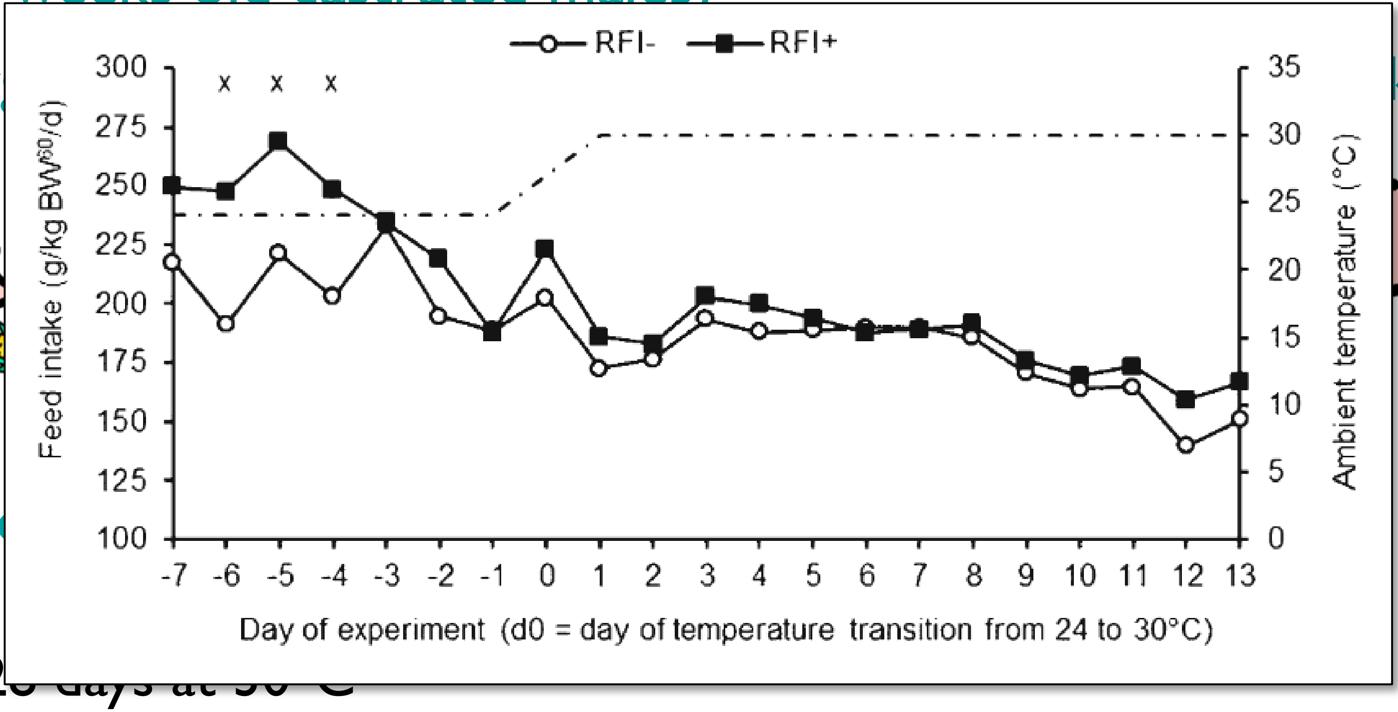
Economic losses
e.g. : yearly cost for US industry = $ 2.4 billion (St Pierre et al, 2003)

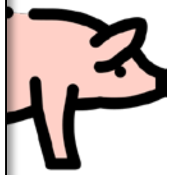
# How duration of HS affect adaptation mechanisms in pigs ?

# Experimental design

- 34 pigs (6 weeks old castrated males)
- Two diver ke ~ Feed efficiency







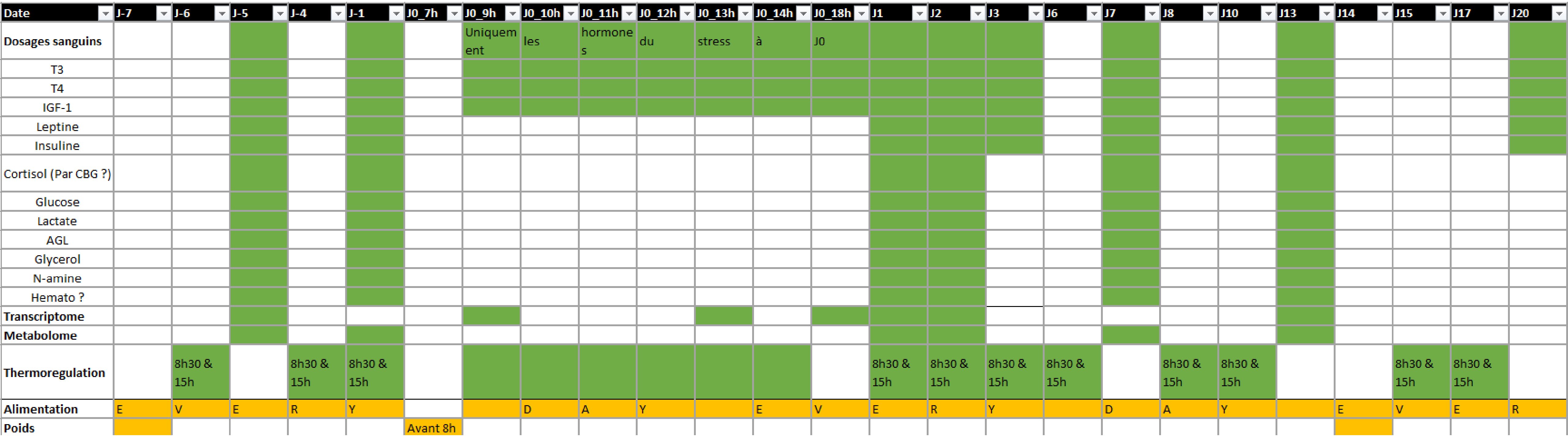
- In a contr
  - 7 days
  - Then 2 days at 30 °C
- Two replicates

Campos *et al.* Thermoregulatory responses during thermal acclimation in pigs divergently selected for residual feed intake. *Int J Biometeorol* **58**, 1545–1557 (2014). https://doi.org/10.1007/s00484-013-0759-3

Effect of heat stress duration

Effect of the line (CMJR+/CMJR-)

# Data available

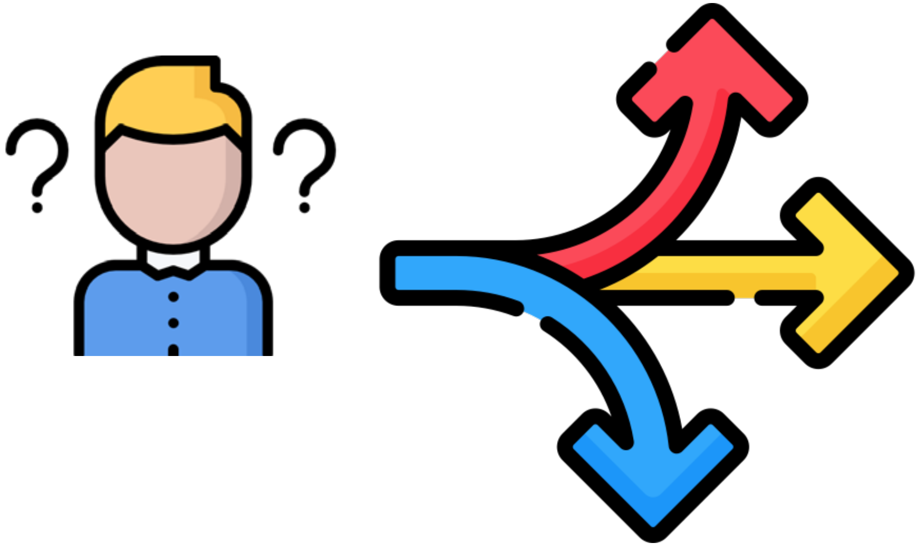


| Date | J-7 | J-6 | J-5 | J-4 | J-1 | J0_7h | J0_9h | J0_10h | J0_11h | J0_12h | J0_13h | J0_14h | J0_18h | J1 | J2 | J3 | J6 | J7 | J8 | J10 | J13 | J14 | J15 | J17 | J20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Dosages sanguins** | | | 🟩 | | 🟩 | | Uniquem ent | les | hormone s | du | stress | à | J0 | 🟩 | 🟩 | 🟩 | | 🟩 | | | 🟩 | | | | 🟩 |
| T3 | | | 🟩 | | 🟩 | | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | | 🟩 | | | 🟩 | | | | 🟩 |
| T4 | | | 🟩 | | 🟩 | | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | | 🟩 | | | 🟩 | | | | 🟩 |
| IGF-1 | | | 🟩 | | 🟩 | | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | | 🟩 | | | 🟩 | | | | 🟩 |
| Leptine | | | 🟩 | | 🟩 | | | | | | | | | 🟩 | 🟩 | 🟩 | | 🟩 | | | 🟩 | | | | 🟩 |
| Insuline | | | 🟩 | | 🟩 | | | | | | | | | 🟩 | 🟩 | 🟩 | | 🟩 | | | 🟩 | | | | 🟩 |
| Cortisol (Par CBG ?) | | | 🟩 | | 🟩 | | | | | | | | | 🟩 | 🟩 | | | 🟩 | | | 🟩 | | | | |
| Glucose | | | 🟩 | | 🟩 | | | | | | | | | 🟩 | 🟩 | | | 🟩 | | | 🟩 | | | | |
| Lactate | | | 🟩 | | 🟩 | | | | | | | | | 🟩 | 🟩 | | | 🟩 | | | 🟩 | | | | |
| AGL | | | 🟩 | | 🟩 | | | | | | | | | 🟩 | 🟩 | | | 🟩 | | | 🟩 | | | | |
| Glycerol | | | 🟩 | | 🟩 | | | | | | | | | 🟩 | 🟩 | | | 🟩 | | | 🟩 | | | | |
| N-amine | | | 🟩 | | 🟩 | | | | | | | | | 🟩 | 🟩 | | | 🟩 | | | 🟩 | | | | |
| Hemato ? | | | 🟩 | | 🟩 | | | | | | | | | 🟩 | 🟩 | | | 🟩 | | | 🟩 | | | | |
| **Transcriptome** | | | 🟩 | | | | 🟩 | | | | 🟩 | | 🟩 | 🟩 | 🟩 | | | | | | 🟩 | | | | |
| **Metabolome** | | | 🟩 | | 🟩 | | | | | | | | | 🟩 | 🟩 | | | 🟩 | | | 🟩 | | | | |
| **Thermoregulation** | | 8h30 & 15h | | 8h30 & 15h | 8h30 & 15h | | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | 🟩 | | 8h30 & 15h | 8h30 & 15h | 8h30 & 15h | 8h30 & 15h | | 8h30 & 15h | 8h30 & 15h | | | 8h30 & 15h | 8h30 & 15h | |
| **Alimentation** | E | V | E | R | Y | | 🟧 | D | A | Y | 🟧 | E | V | E | R | Y | 🟧 | D | A | Y | 🟧 | E | V | E | R |
| **Poids** | 🟧 | | | | | Avant 8h | | | | | | | | | | | | | | | | 🟧 | | | |

- Transcritptome : 7 time step x 13 pigs
- Metabolome : 6 time step x 16 pigs
- Blood dosages (hormones and metabolites) : ~5 to 15 time steps x 19 pigs
- Thermoregulation : Nearly everyday x 34 pigs

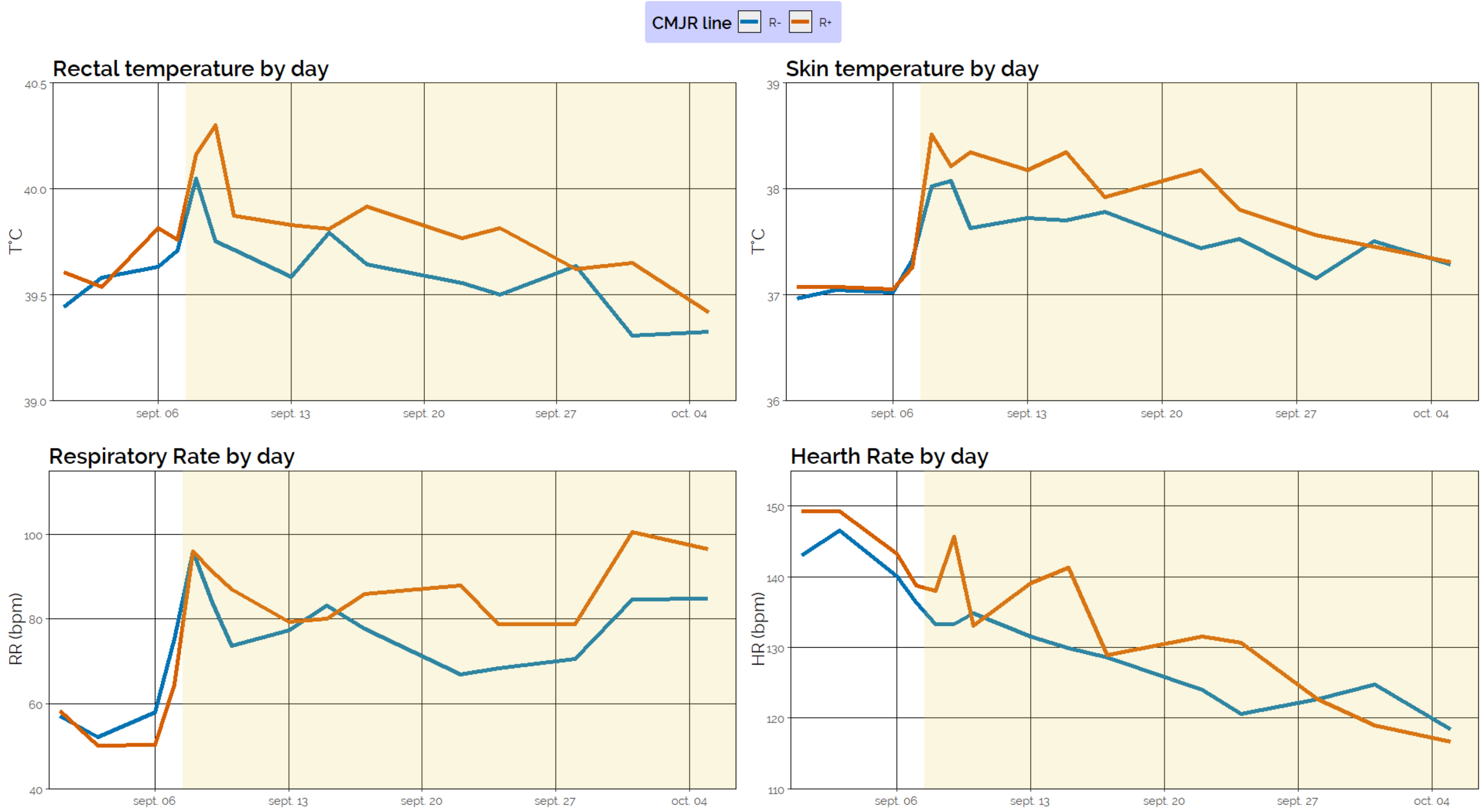
# A few questions with these data…

- Longitudinal data
- Small number of animals for some omics
- High dimension of the data
- Multiple omics
- Integration of heterogenous data
- Blood = agregator of responses from the whole body



# Thermoregulation

- Data on 34 pigs
- Bi-modal reaction to heat stress
    - Increase in ST, RR
    - Then smaller decrease
    - Short-term vs long-term HS
- Mean by day


Rectal temperature by day

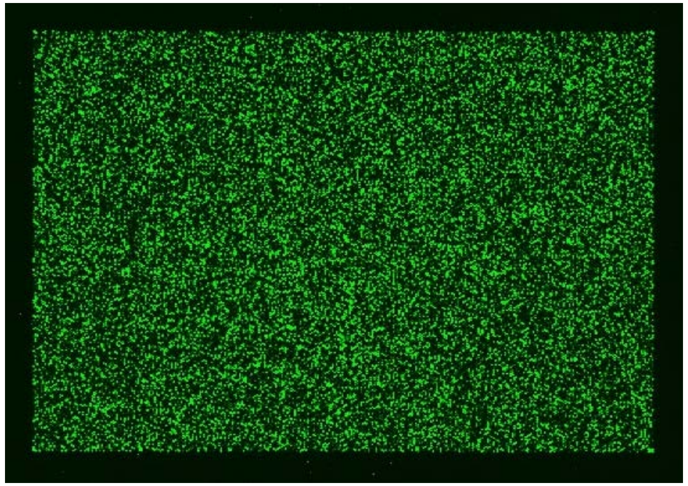
Skin temperature by day

Respiratory Rate by day

Hearth Rate by day

# Transcriptome

# Transcriptome

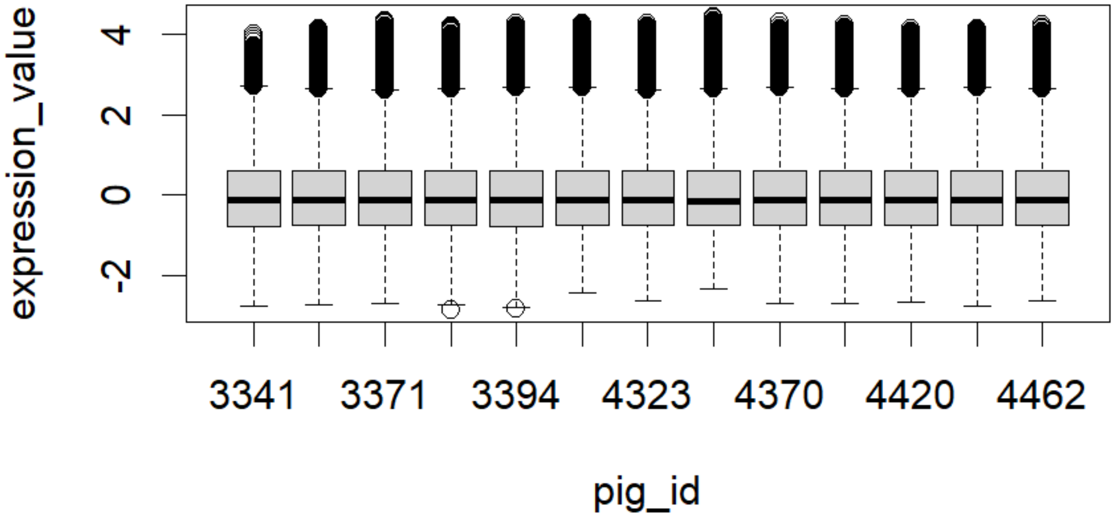


- Microarray data for 60k probes

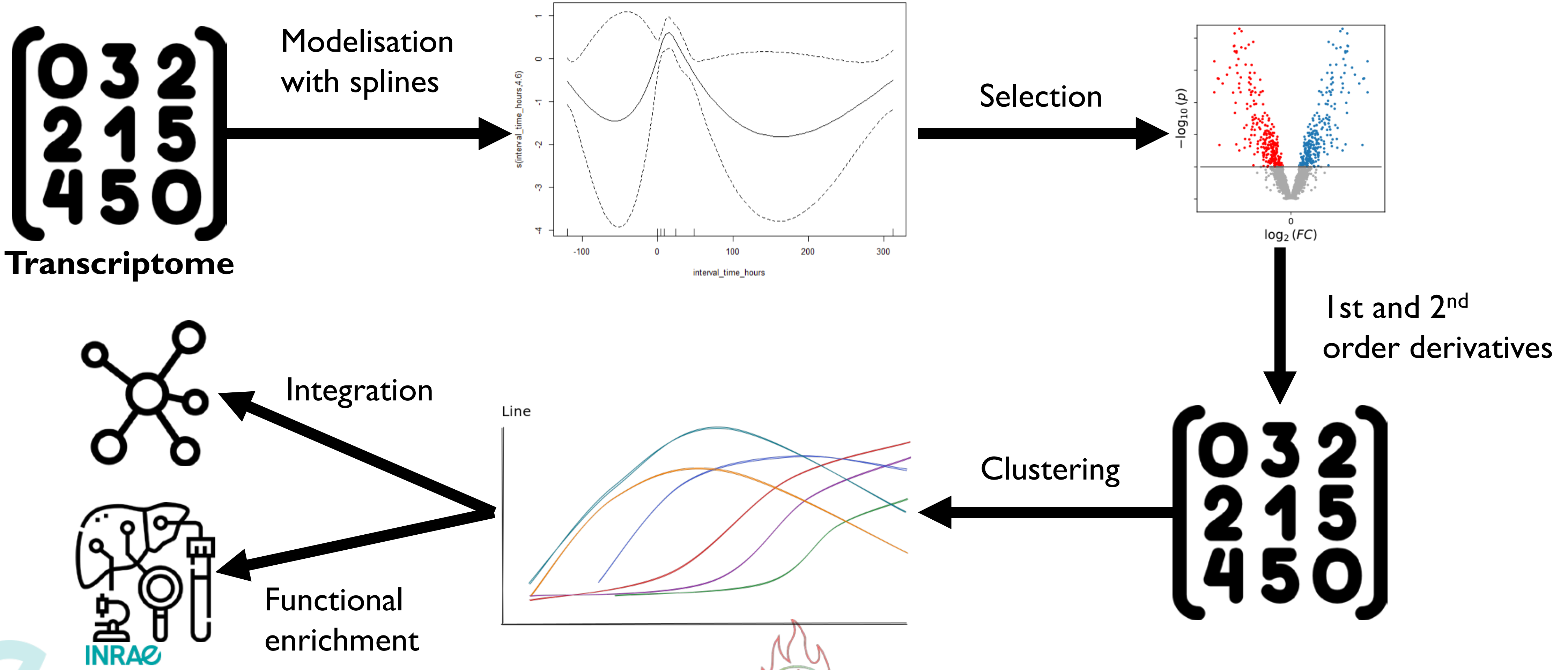
| 13 pigs | Replicate 1 | Replicate 2 |
|---|---|---|
| **CMJR-** | 3 | 4 |
| **CMJR+** | 2 | 4 |

- In total : 2,502,224 records of transcript expression

- Data already preprocessed with 32924 probes



# A « promising » first idea of pipeline



**Transcriptome**

Modelisation with splines

Selection

1st and 2nd order derivatives

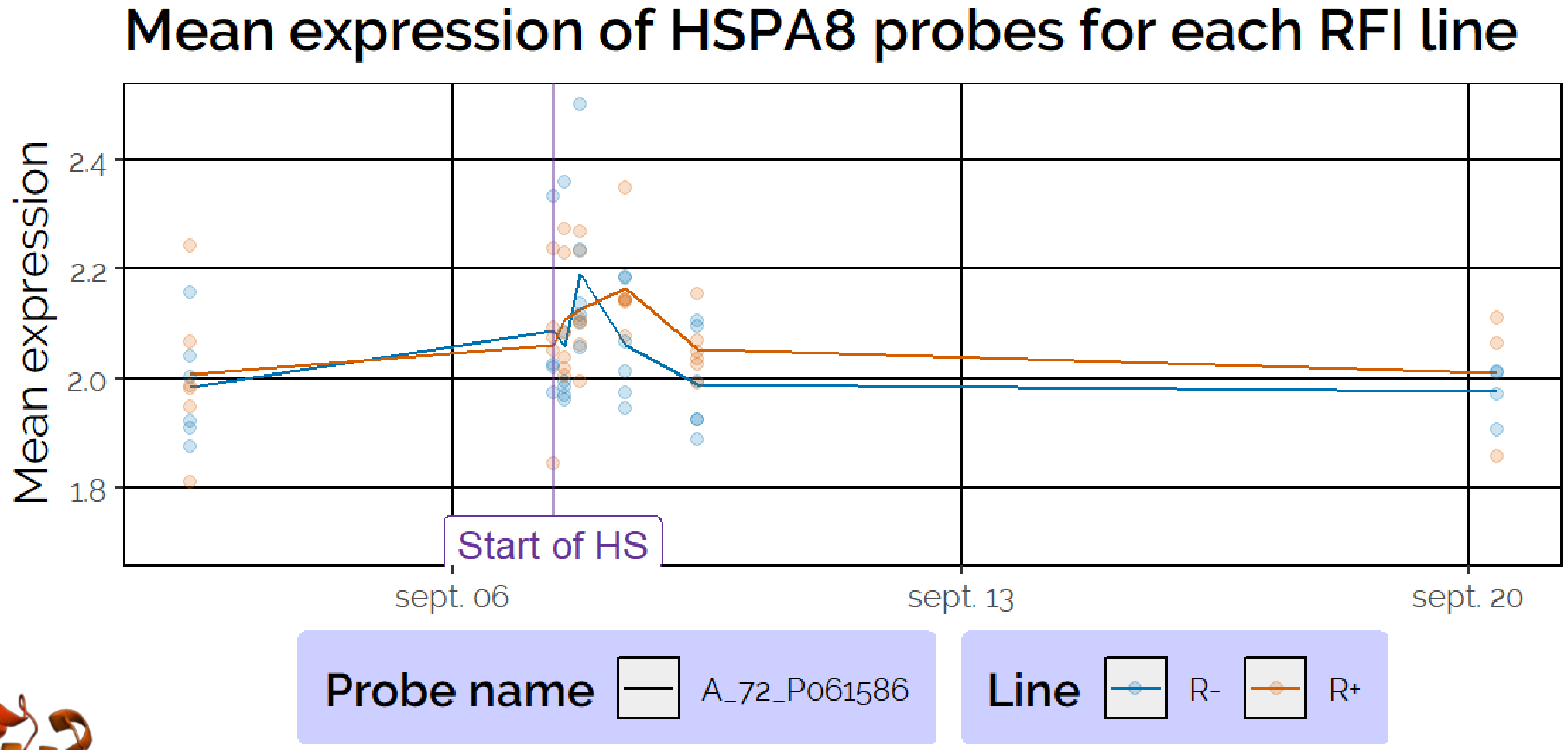
Clustering

Integration

Functional enrichment

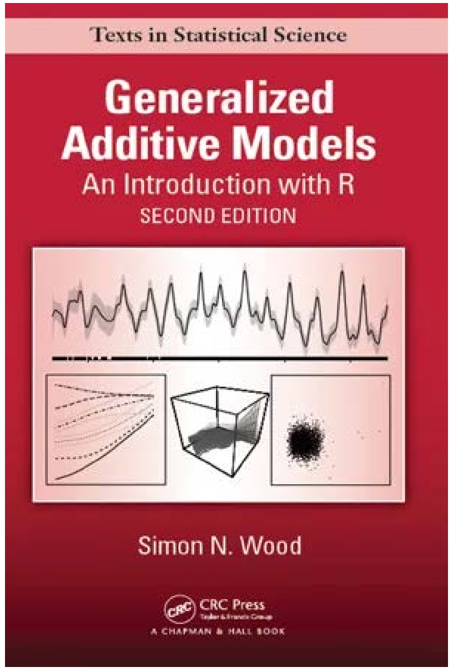
# Test on one probe

- Selection of the HSPA8 (Hsp70) gene
- Heat Shock Protein
  - Chaperone protein (act on the folding of protein)
  - Presence during cellular stress
  - Known response during HS


Mean expression of HSPA8 probes for each RFI line

# Test on one probe

- For spline modelisation (or GAM for Generalized Additive Model)
  - Package R *mgcv*
  - Splines, fixed or random effects
- Model taking into account
  - time
  - line (CMJR+/CMJR-)
  - replicate
  - individual effect
- Time term expressed as hours from start of HS



Texts in Statistical Science

Generalized Additive Models

An Introduction with R

SECOND EDITION

Simon N. Wood

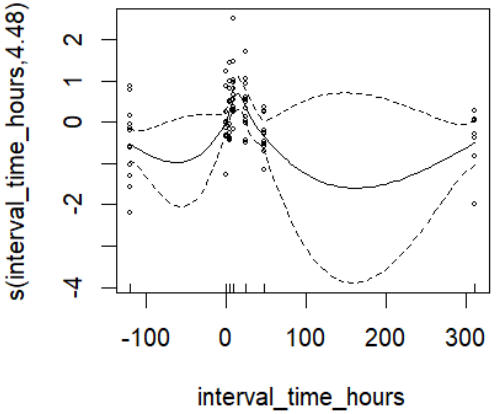
CRC Press

A CHAPMAN & HALL BOOK

Cross Validated

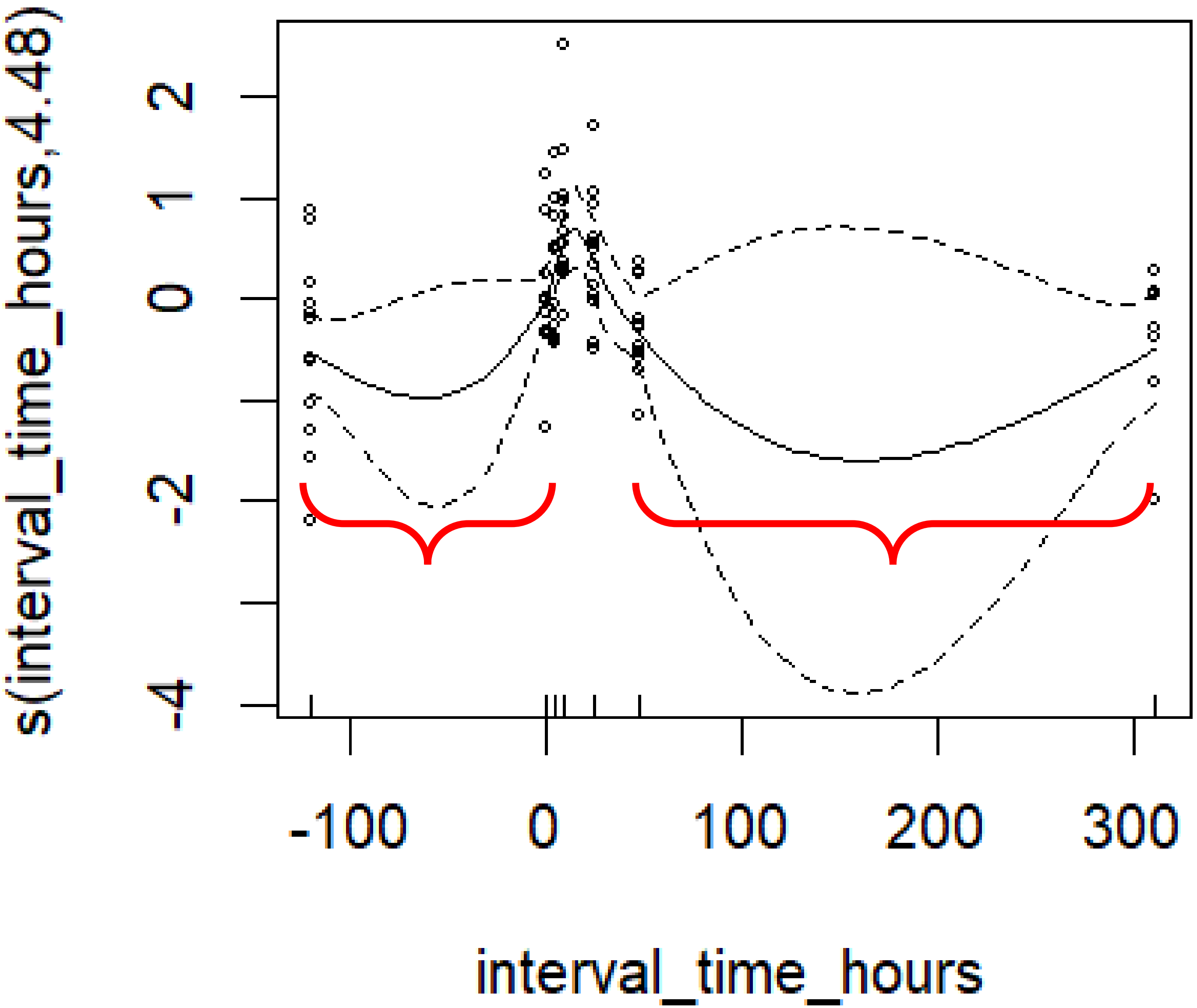
stackoverflow

```
gam(formula = expression_value_scale ~ 1 + s(interval_time_hours, bs = "tp",
k = 4, pc = 0) + s(pig_id, bs = "re") + s(Rep, bs = "re") + lignee,
data = dat, gamma = 0.7)
```



# But…

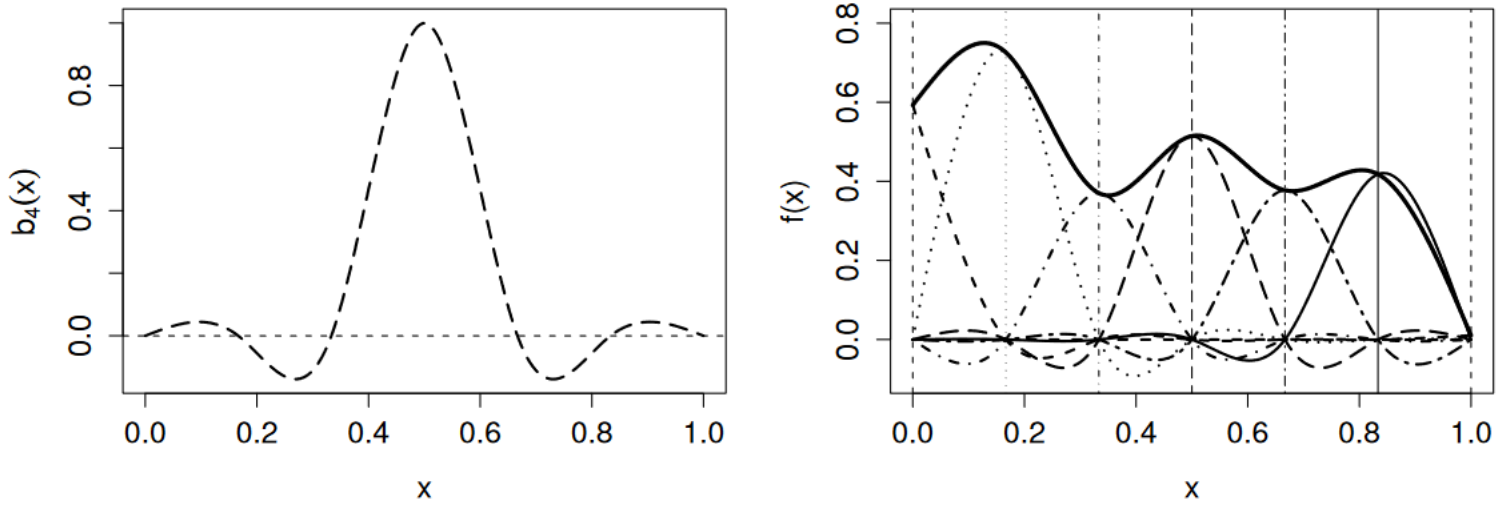
- For spline modelisation (or GAM for Generalized Additive Model)
  - Package R *mgcv*
  - Splines, fixed or random effects
- Model taking into account
  - time
  - line (CMJR+/CMJR-)
  - replicate
  - individual effect
- Time term expressed as hours from start of HS



INRAE

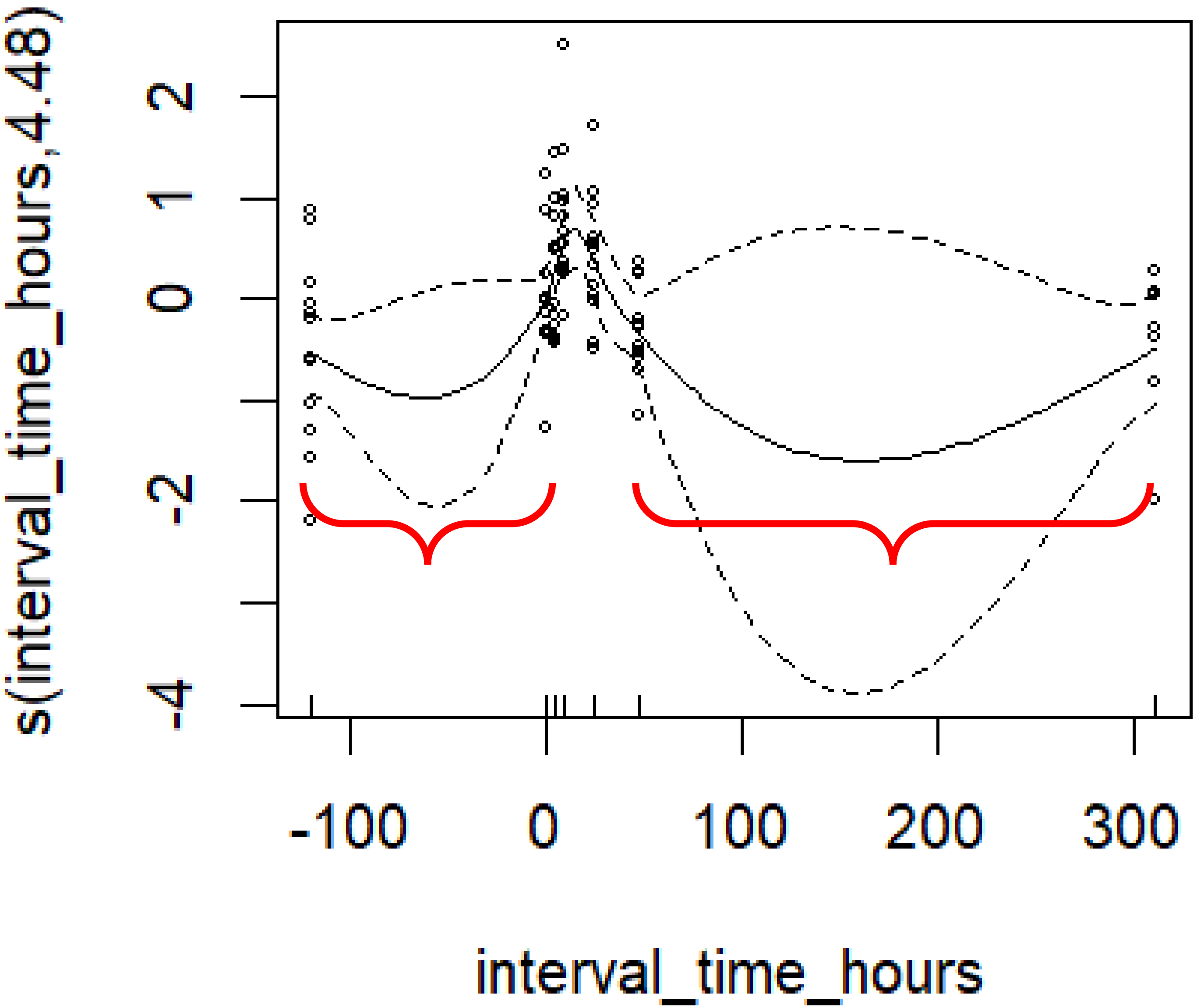
# Back to the spline model

- How does it work ?
  - Select a basis
  - Select the basis dimension
  - $f(x) = \sum_{i=1}^{k} b_i(x)\beta_i$

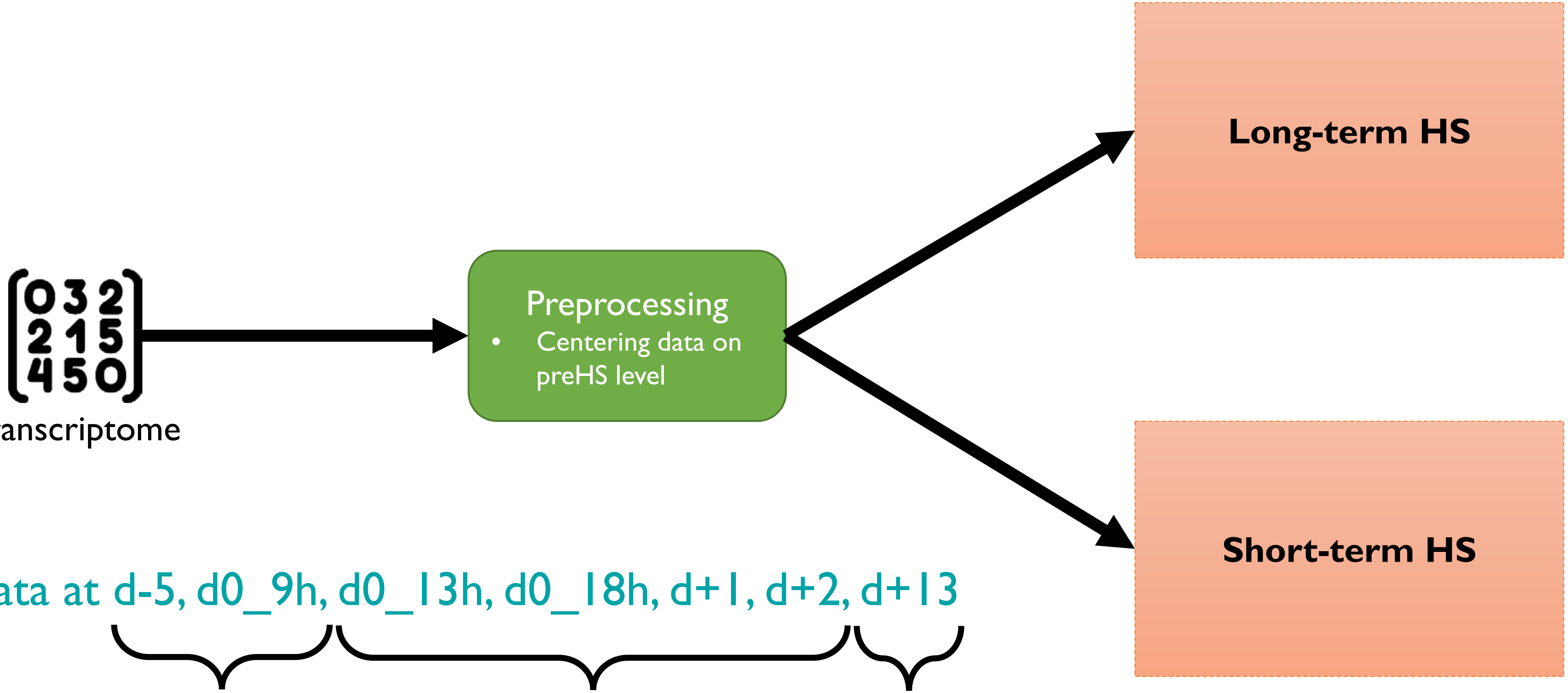


# But…

- For spline modelisation (or GAM for Generalized Additive Model)
    - Package R *mgcv*
    - Splines, fixed or random effects
- Model taking into account
    - time
    - line (CMJR+/CMJR-)
    - replicate
    - individual effect
- Time term expressed as hours from start of HS



# A possible solution in two steps



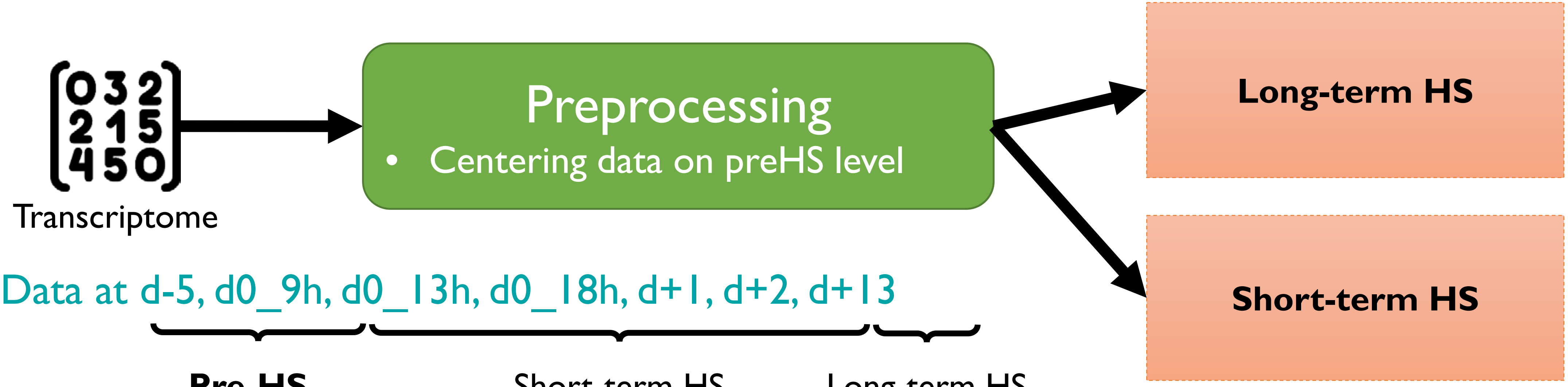
- Data at d-5, d0_9h, d0_13h, d0_18h, d+1, d+2, d+13

**Pre-HS** (d-5, d0_9h)

**Short-term HS** (d0_13h, d0_18h, d+1, d+2)

**Long-term HS** (d+13)

# A possible solution



$$\begin{bmatrix} 0 & 3 & 2 \\ 2 & 1 & 5 \\ 4 & 5 & 0 \end{bmatrix}$$

Transcriptome

Preprocessing

- Centering data on preHS level

**Long-term HS**

**Short-term HS**

- Data at d-5, d0_9h, d0_13h, d0_18h, d+1, d+2, d+13

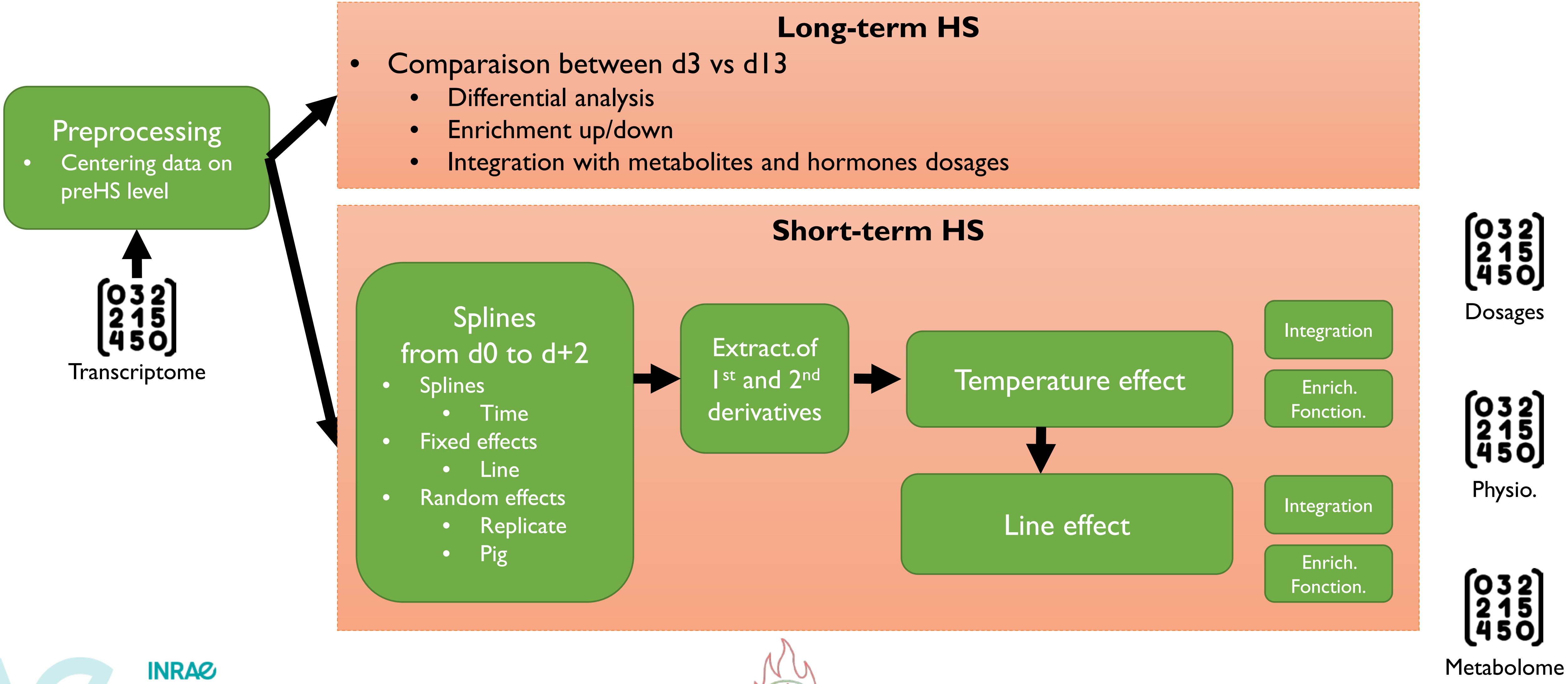
**Pre-HS** Short-term HS Long-term HS

Is d-5 expression level similar to d0_9h ?

- Paired t-test with multiple test correction (fdr)
  - 1656 probes have significantly different values, out of 32924 probes (~5%)

- Possible sources of differences
  - Age
  - Possible experimental stress at d0

# New pipeline idea



Preprocessing
- Centering data on preHS level

Transcriptome

**Long-term HS**
- Comparaison between d3 vs d13
  - Differential analysis
  - Enrichment up/down
  - Integration with metabolites and hormones dosages

**Short-term HS**

Splines from d0 to d+2
- Splines
  - Time
- Fixed effects
  - Line
- Random effects
  - Replicate
  - Pig

Extract.of 1st and 2nd derivatives

Temperature effect

Line effect

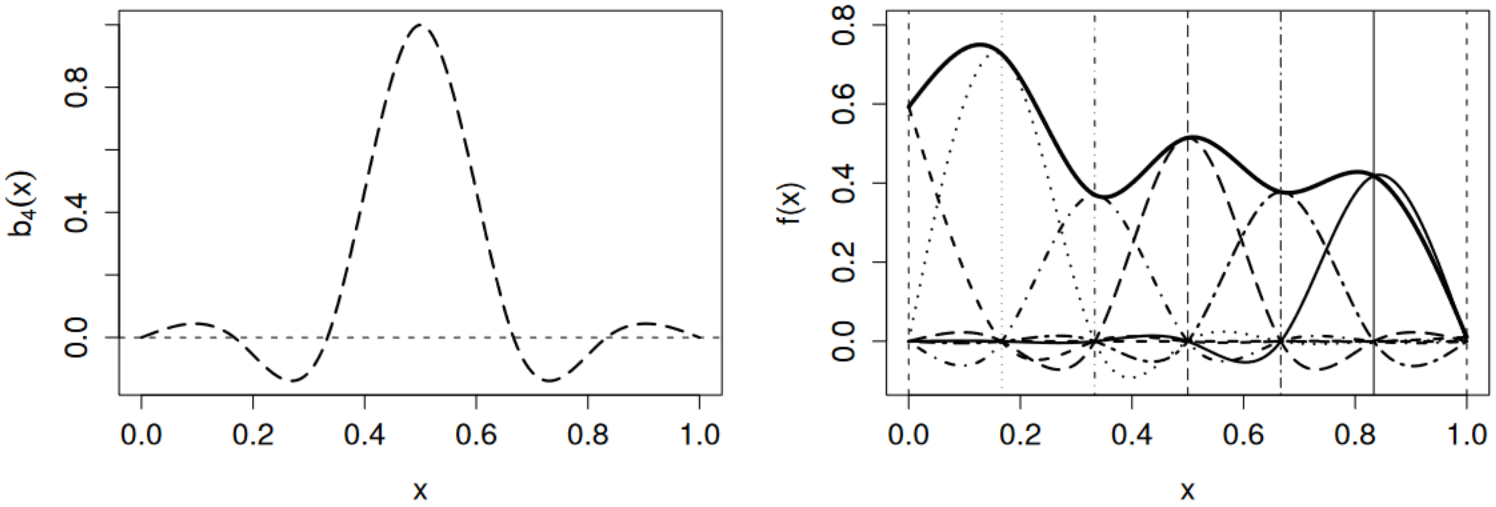
Dosages

Physio.

Metabolome

# Back to the spline model

- How does it work ?
  - Select a basis
  - Select the basis dimension
  - $f(x) = \sum_{i=1}^{k} b_i(x)\beta_i$
  - Possibility to had some penalty term to play on the smoothness of the curve

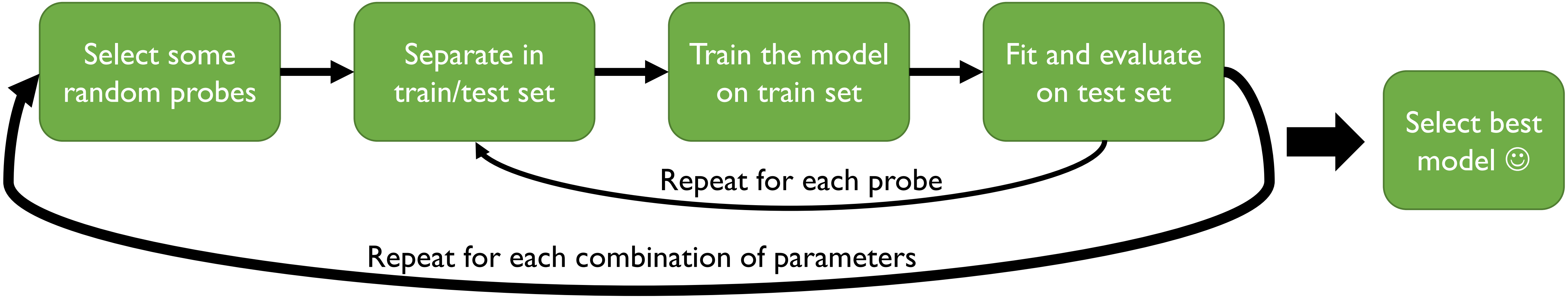


# Selection of the model' parameters

- We choose to perform a grid.search (i.e. exploring all the possible sets of parameters)
  - Basis : Thin-plate spline, cubic regression spline (each with an alternative)
  - Basis dimension : 3 to 5
  - Gamma (penalty) : 0.5 to 2.5 by 0.2 increments
  - In total 4 x 3 x 10 = 120 possibilities
- Approach by k-fold cross-validation difficult based on the small number of animals

# Selection of the model' parameters

- We choose to perform a grid.search (i.e. exploring all the possible sets of parameters)
- Approach by k-fold cross-validation difficult because of the small number of animals
- We tried a kind of similar home-made method



Select some random probes

Separate in train/test set

Train the model on train set

Fit and evaluate on test set

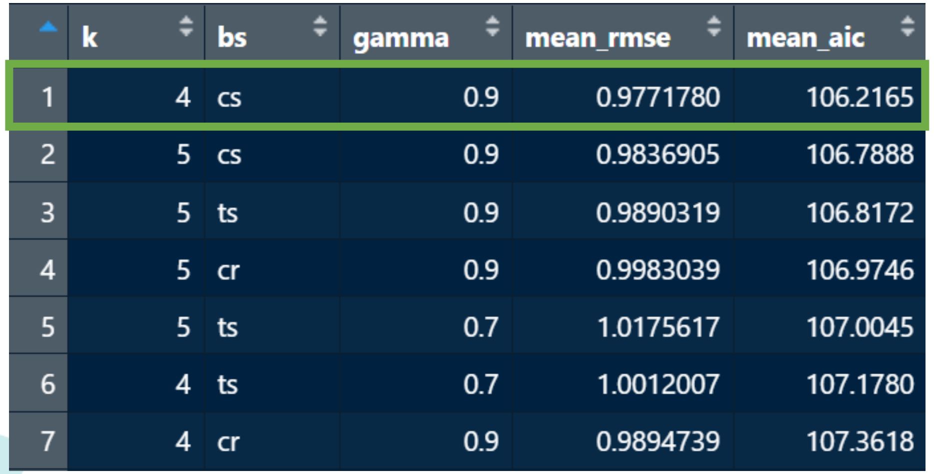
Select best model ☺

Repeat for each probe

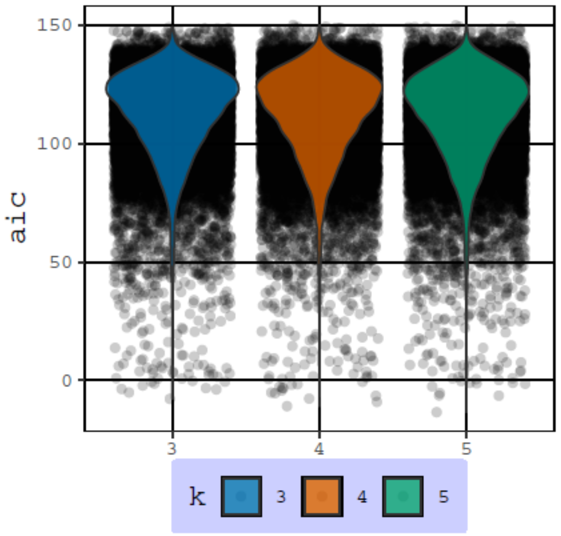
Repeat for each combination of parameters

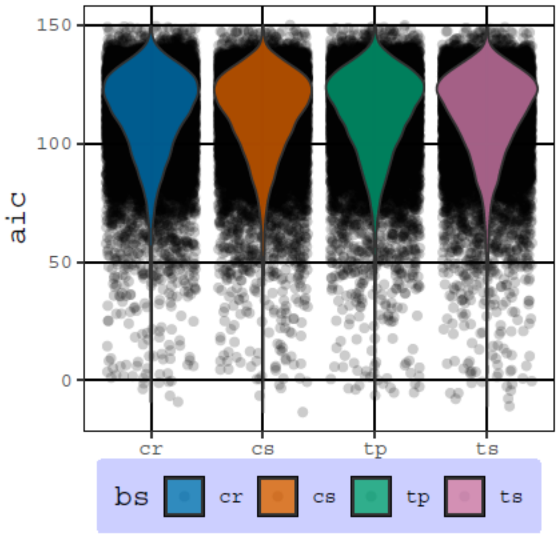
# We have a winner

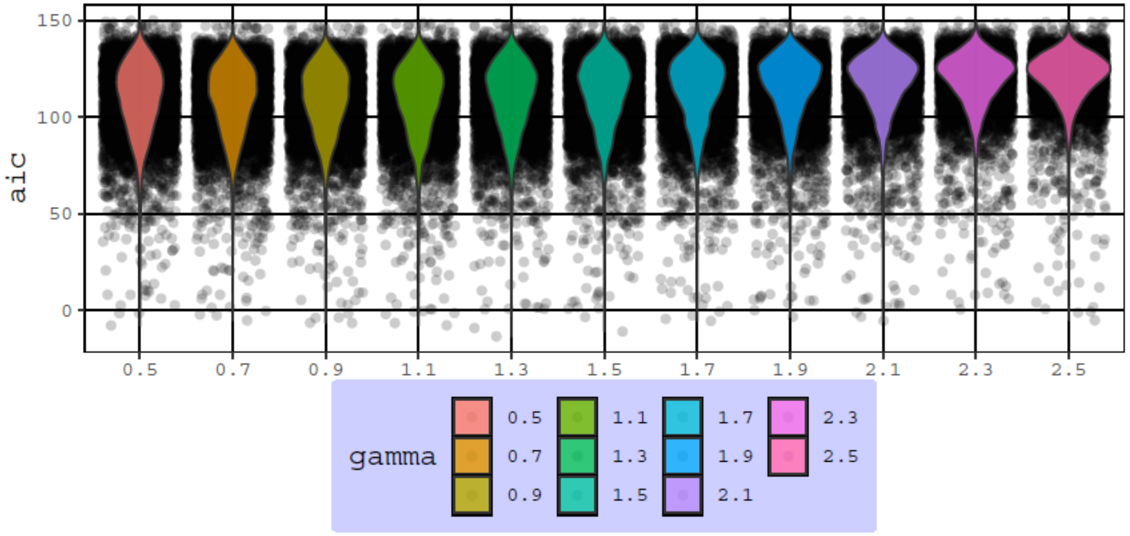
- We choose to perform a grid.search (i.e. exploring all the possible sets of parameters)
  - Basis : Thin-plate spline, cubic regression spline (each with an alternative)
  - Basis dimension : 3 to 5
  - Gamma (penalty) : 0.5 to 2.5 by 0.2 increments
  - In total 4 x 3 x 10 = 120 possibilities



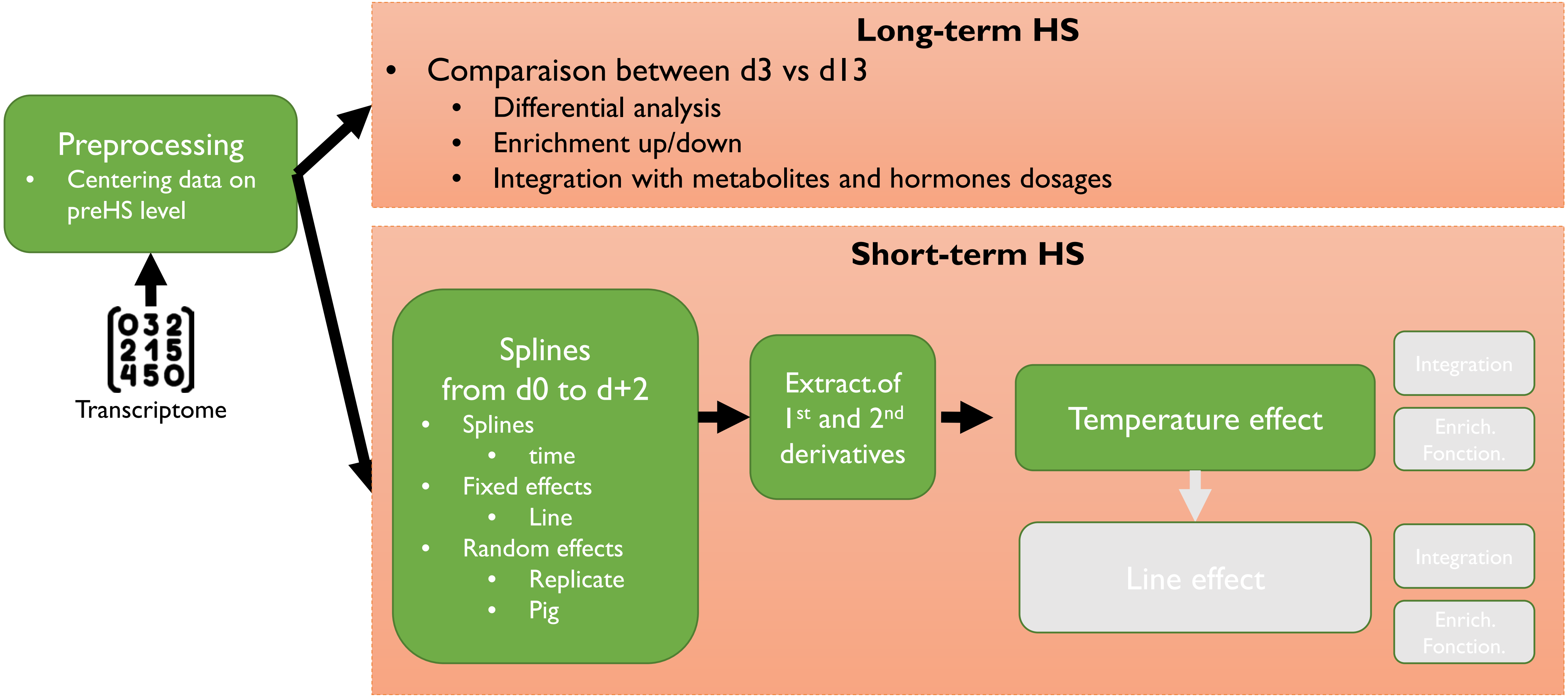
| | k | bs | gamma | mean_rmse | mean_aic |
|---|---|---|---|---|---|
| 1 | 4 | cs | 0.9 | 0.9771780 | 106.2165 |
| 2 | 5 | cs | 0.9 | 0.9836905 | 106.7888 |
| 3 | 5 | ts | 0.9 | 0.9890319 | 106.8172 |
| 4 | 5 | cr | 0.9 | 0.9983039 | 106.9746 |
| 5 | 5 | ts | 0.7 | 1.0175617 | 107.0045 |
| 6 | 4 | ts | 0.7 | 1.0012007 | 107.1780 |
| 7 | 4 | cr | 0.9 | 0.9894739 | 107.3618 |







# New pipeline



Transcriptome

**Preprocessing**
- Centering data on preHS level

## Long-term HS

- Comparaison between d3 vs d13
  - Differential analysis
  - Enrichment up/down
  - Integration with metabolites and hormones dosages

## Short-term HS

**Splines from d0 to d+2**
- Splines
  - time
- Fixed effects
  - Line
- Random effects
  - Replicate
  - Pig

Extract.of 1st and 2nd derivatives

Temperature effect

Line effect

Integration

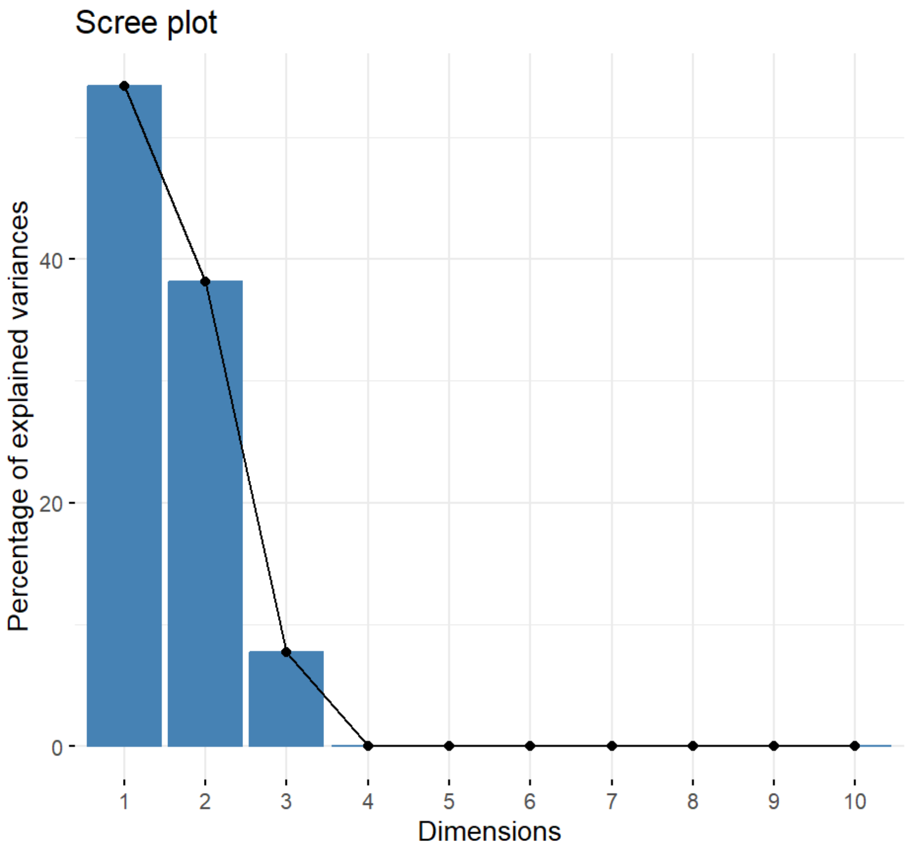
Enrich. Fonction.

Integration

Enrich. Fonction.

# PCA on the 1st and 2nd derivatives

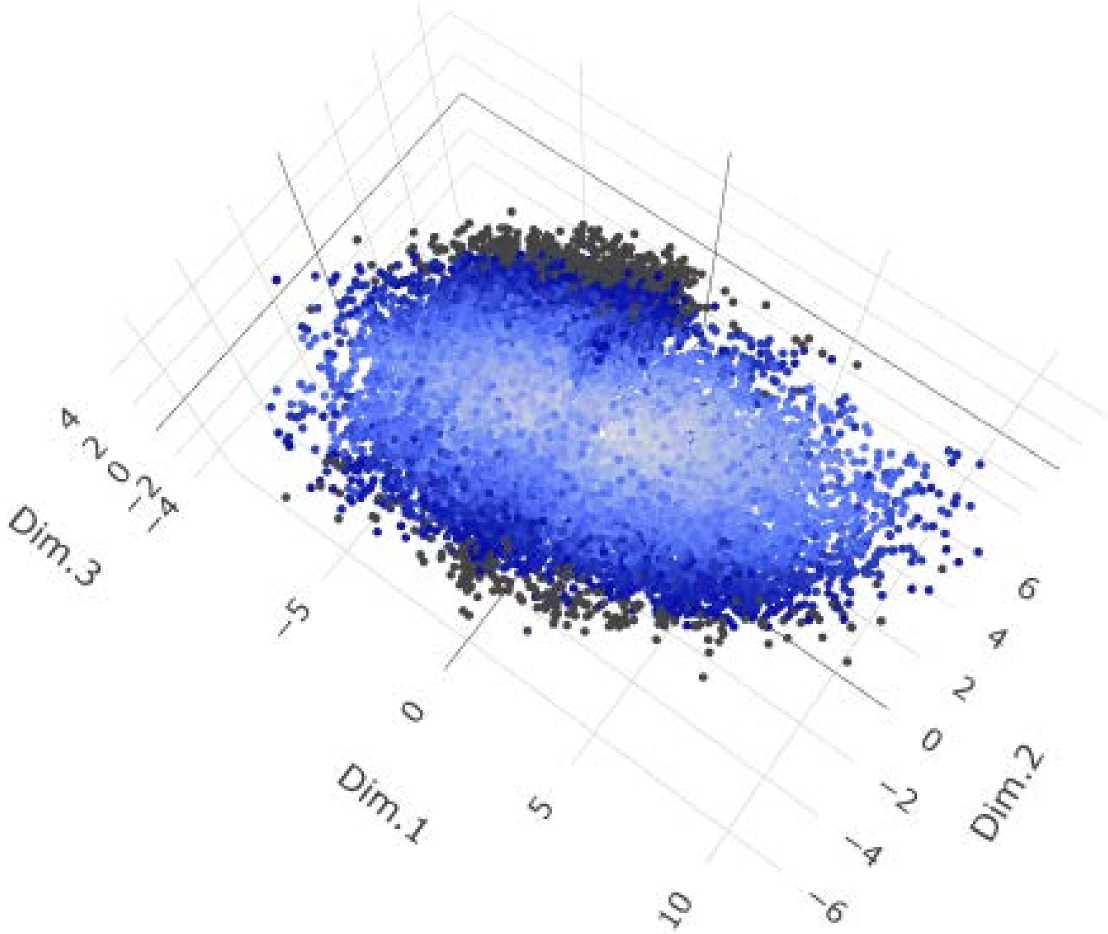
- We extract derivatives at times of prelevment
    - 5 point for each derivatives
    - From 2.5M observations to 32k models to a matrix of size 32k x 10
- We perform a PCA to visualize the data
- Using FactoMineR package



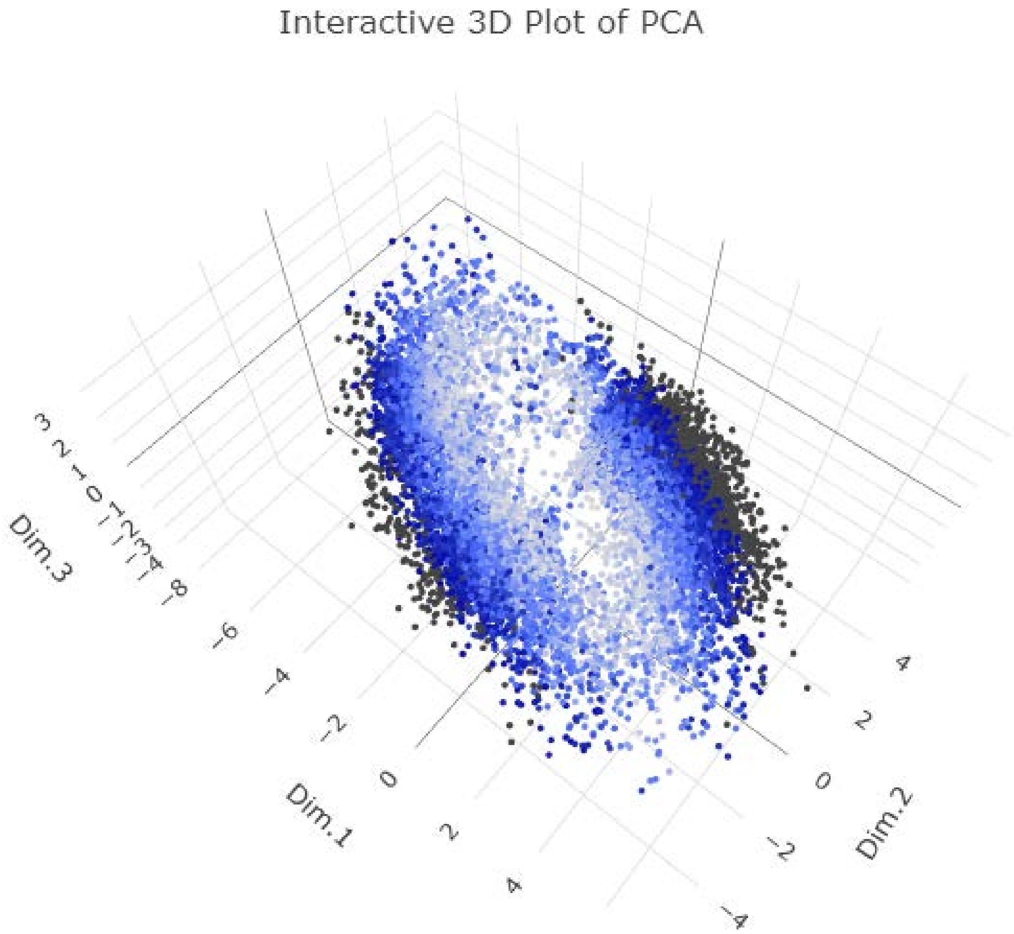
# Visualization of the first 3 axis

- Colored with the p-value of the temporal term
  - Darker = Smaller p-value

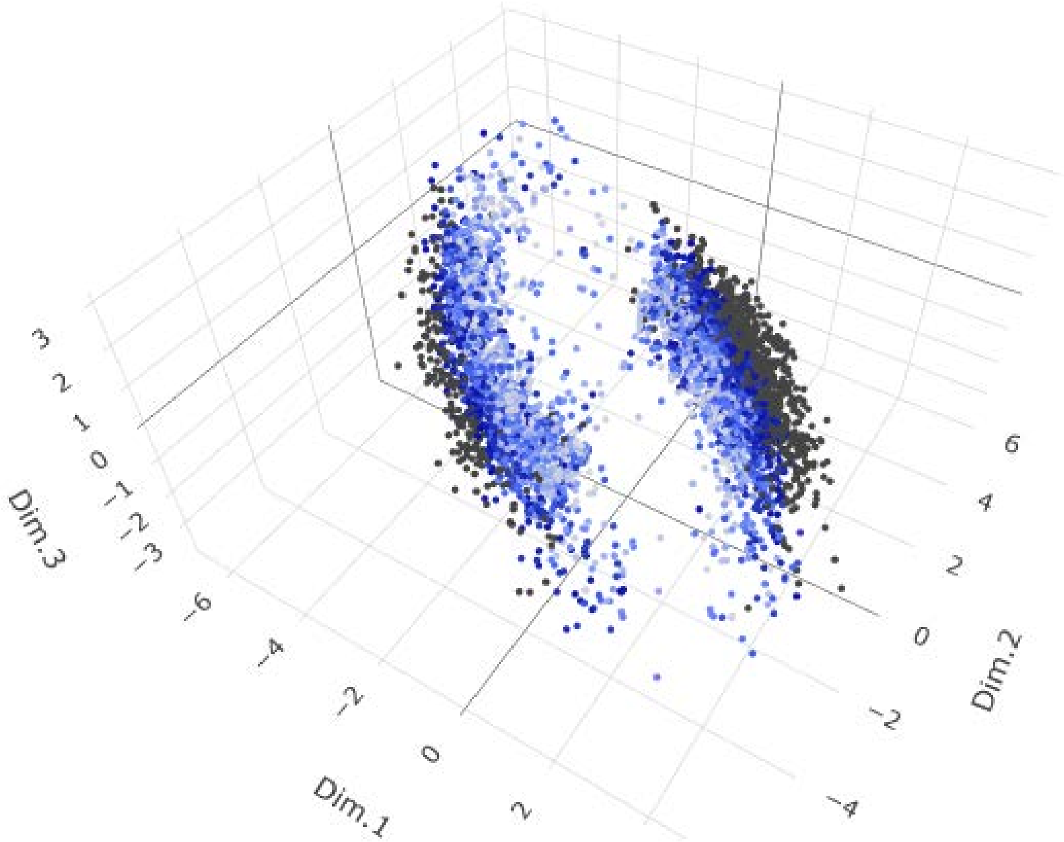
- Seems to be some structure on the data
- We can try to filter the probes affected by time



# Visualization of the first 3 axis



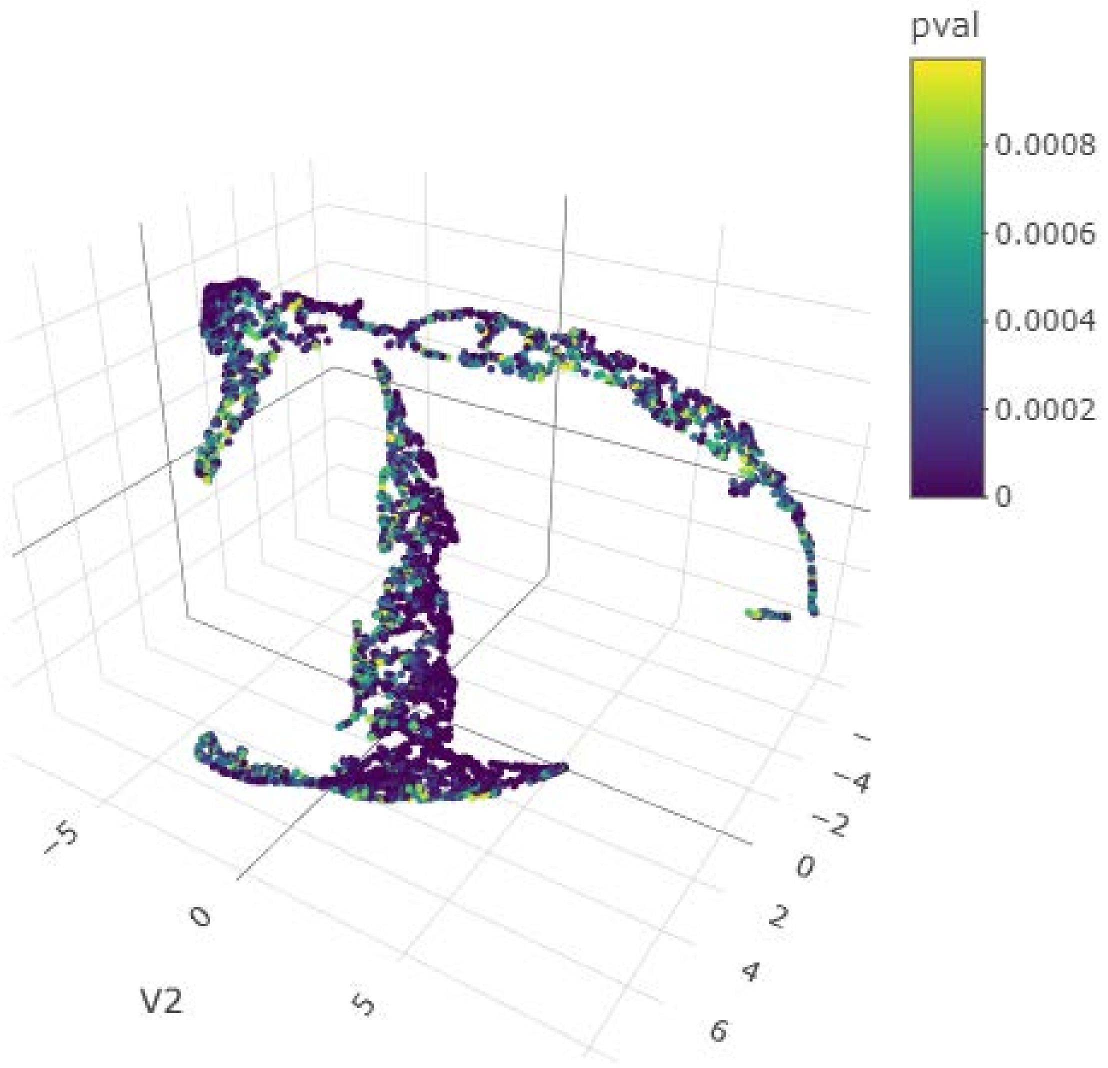
p.value < 0.05



p.value < 0.001

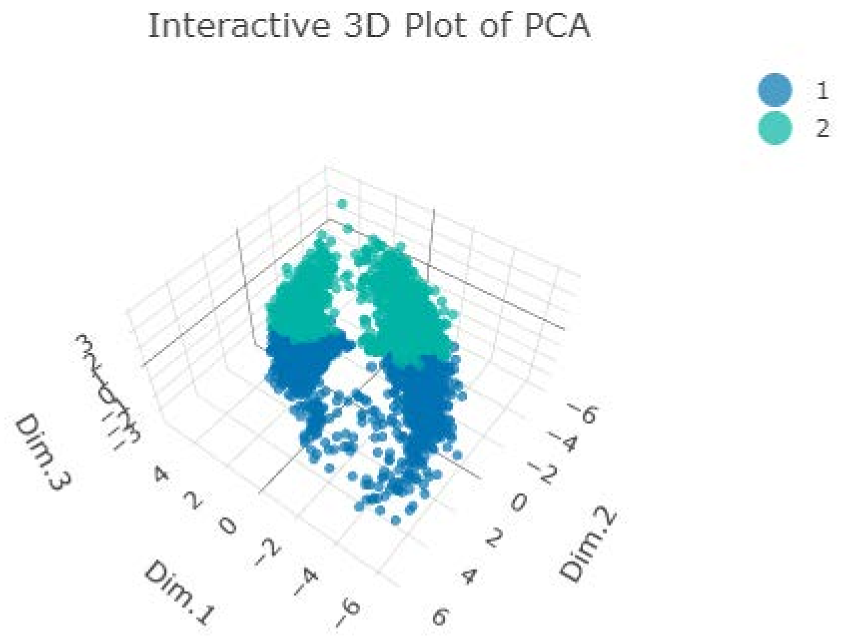
# Or maybe a umap can give even better results

- There definitly seems to be some structure found in the data
  - Some clustering might be interesting



# Trying some clustering

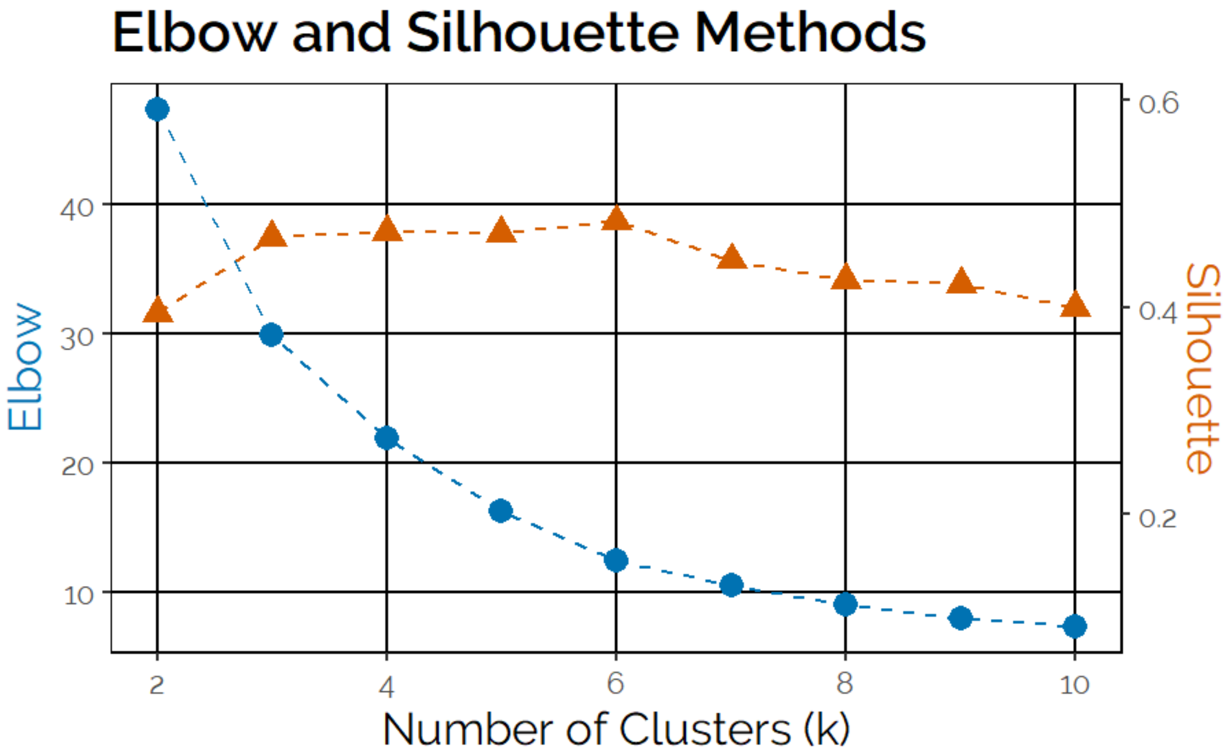
- According to what we have seen, we want 2 clusters
- But when we try a simple kmeans or hclust clustering =>



Not exactly what we want /:

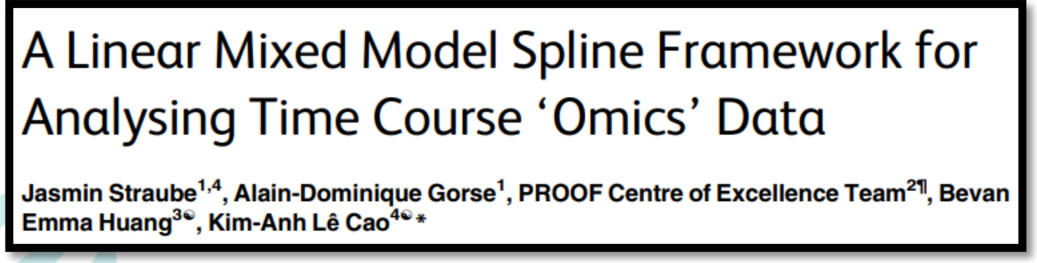
# Finding the right number of clusters

- We want to :
  - maximize inter-clusters distance
  - minimize intra-cluster distance
- We can use the elbow and silhouette plot to decide
- Between 5 or 6 clusters seems a good compromise



# Visualisation of PCA results with clustering

- Look interesting but maybe better results could be obtained
- Few ideas to improve
  - Clustering directly on the curves
  - Other clustering techniques
  - More derivative points
  - Consensus clustering like in *Straube et al, 2015*



A Linear Mixed Model Spline Framework for Analysing Time Course 'Omics' Data

**Jasmin Straube[1,4], Alain-Dominique Gorse[1], PROOF Centre of Excellence Team[2¶], Bevan Emma Huang[3©], Kim-Anh Lê Cao[4©*]**

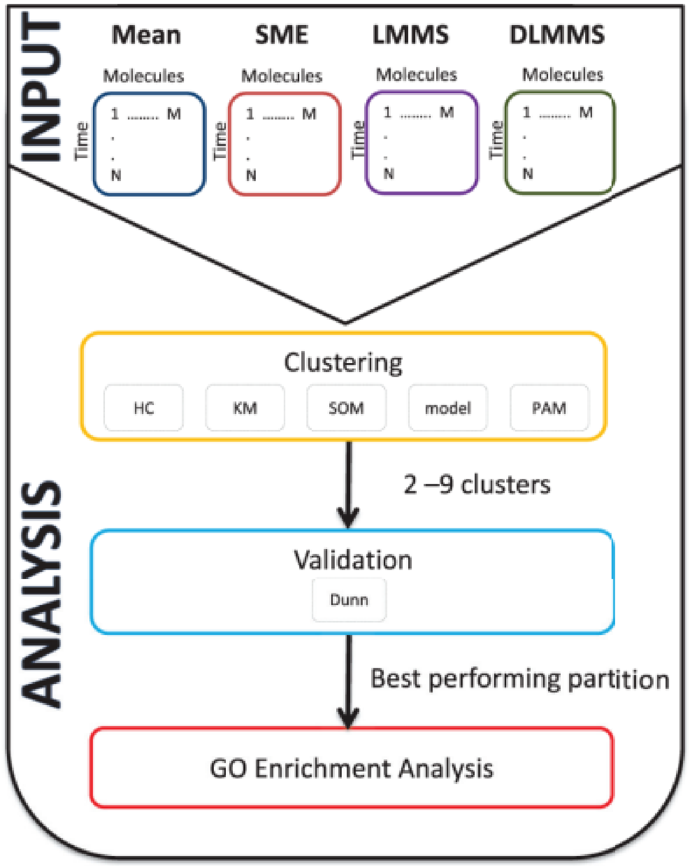


**Fig 3. Workflow for the profile cluster analysis.** Trajectories derived from Linear Mixed Model Spline (LMMS) and Derivative Linear Mixed Model Spline (DLMMS) were compared to trajectories derived either from the mean or Smoothing Splines Mixed Effects (SME) models. Five clustering algorithms—hierarchical clustering (HC), kmeans (KM), Self-Organizing Maps (SOM), model-based (model) and Partitioning Around Medoids (PAM) were then applied on modelled trajectories using a range of two to nine clusters. The performance of each algorithm was assessed using the Dunn index. Gene Ontology (GO) term enrichment analysis was performed on each of the obtained clusters.

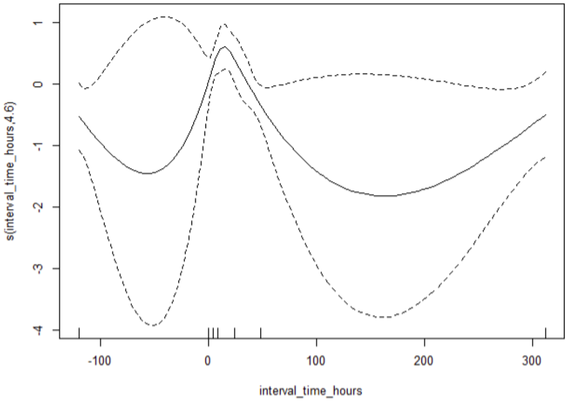
doi:10.1371/journal.pone.0134540.g003

# A « promising » first idea of pipeline

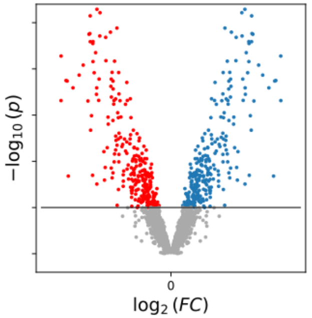
$$\begin{bmatrix}0&3&2\\2&1&5\\4&5&0\end{bmatrix}$$

**Transcriptome**

Modelisation with splines



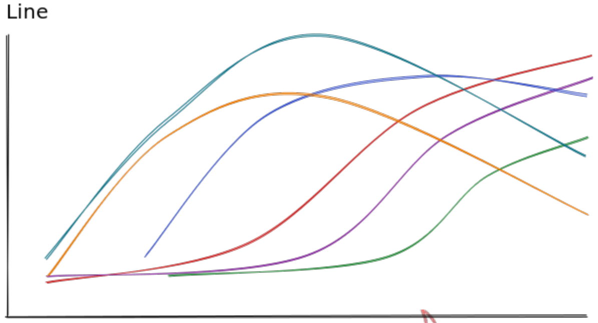
Selection



1st and 2nd order derivatives

$$\begin{bmatrix}0&3&2\\2&1&5\\4&5&0\end{bmatrix}$$

Clustering by consensus

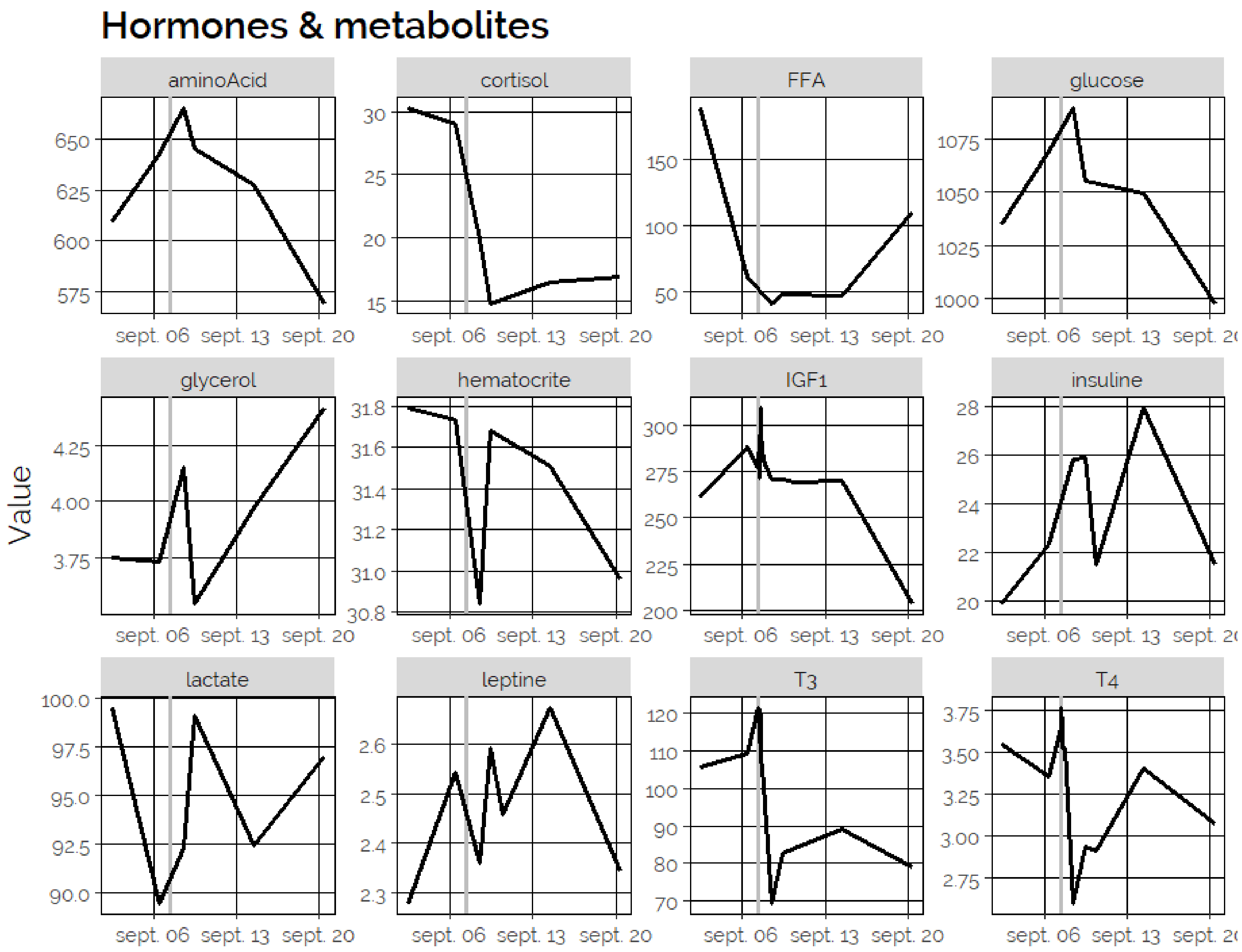


Integration

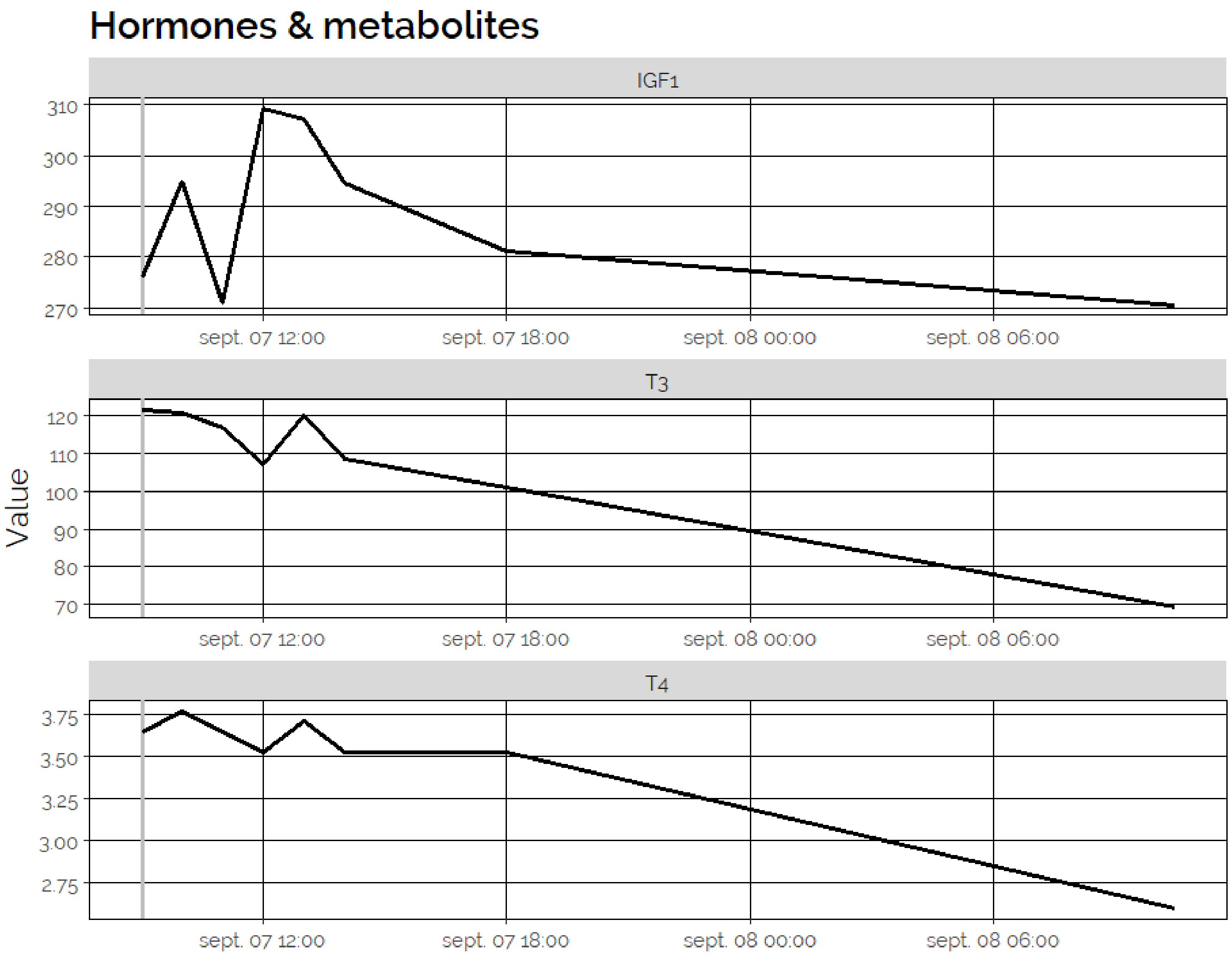
Functional enrichment

# Other omics

# Hormones & Metabolites



# Hormones – Start of stress



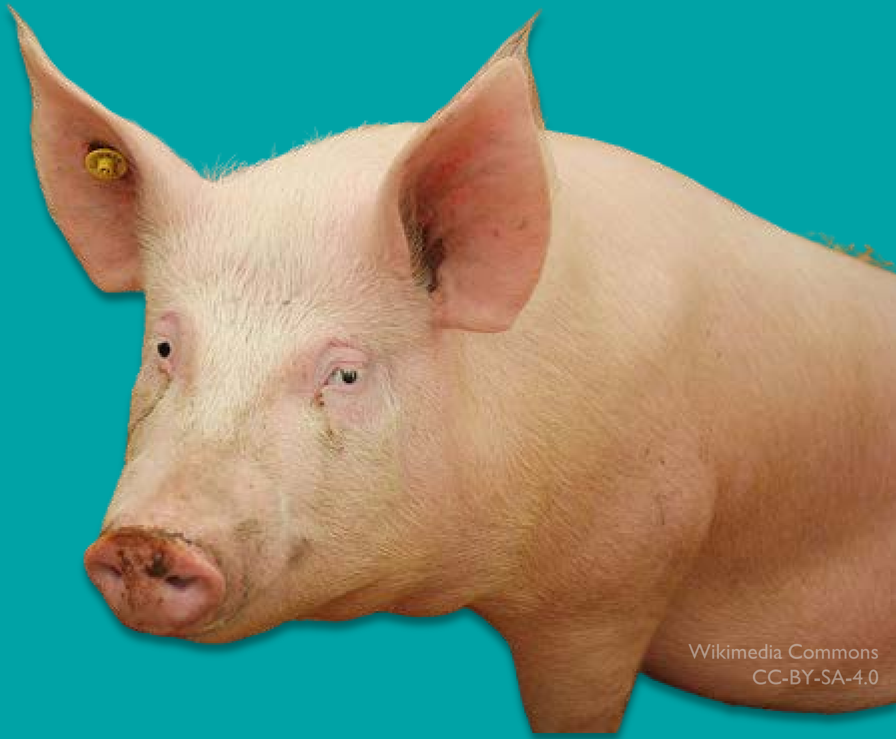
INRAE

# Metabolome

- Normally, 1H-NMR spectra
- But for now, we only have buckets
- Impossible to use ASICS for automatic identification and quantification of metabolites

# Integration ?

# Thanks for your attention ! 🐷


Wikimedia Commons
CC-BY-SA-4.0

Any questions ? 🐷