Méthodologie de l'approche systémique de la biologie et de la médecine: Exemple de l'insuffisance cardiaque

Guillaume LAMIRAULT

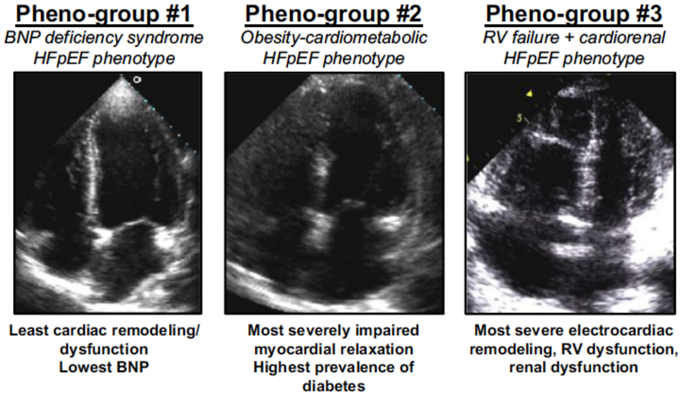
guillaume.lamirault@univ-nantes.fr

UE M1S&S-BioSyst, 13 décembre 2021, Nantes

L'unité de recherche de l'institut du thorax, Inserm UMR 1087 / CNRS UMR 6291
Nantes, France

Classification basée sur le “phenomapping »

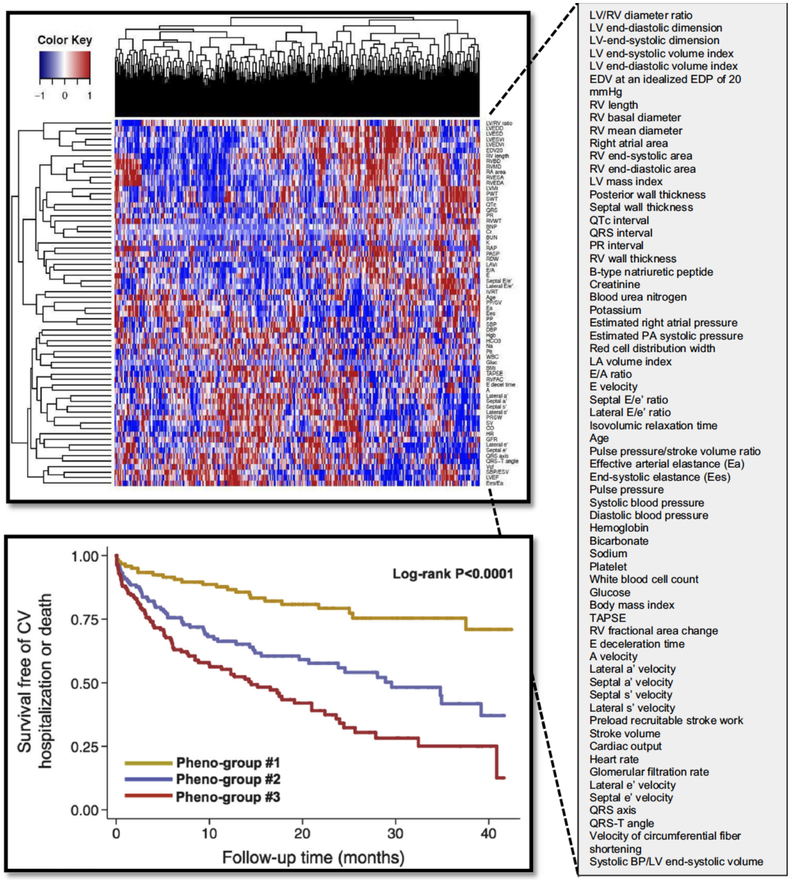
- analyses de données massives prenant en compte les caractéristiques physiques, les signes vitaux, les analyses ECG, biologiques et échocardiographiques
- Identification de 3 groupes ayant chacun un pronostic différent.
  - 1er groupe est caractérisé par un BNP abaissé
  - le 2ème par un phénotype associant obésité-syndrome cardiométabolique
  - le 3ème par une insuffisance cardiaque droite et syndrome cardio-rénal



Bi atrial enlargement / mild RV dysfunction

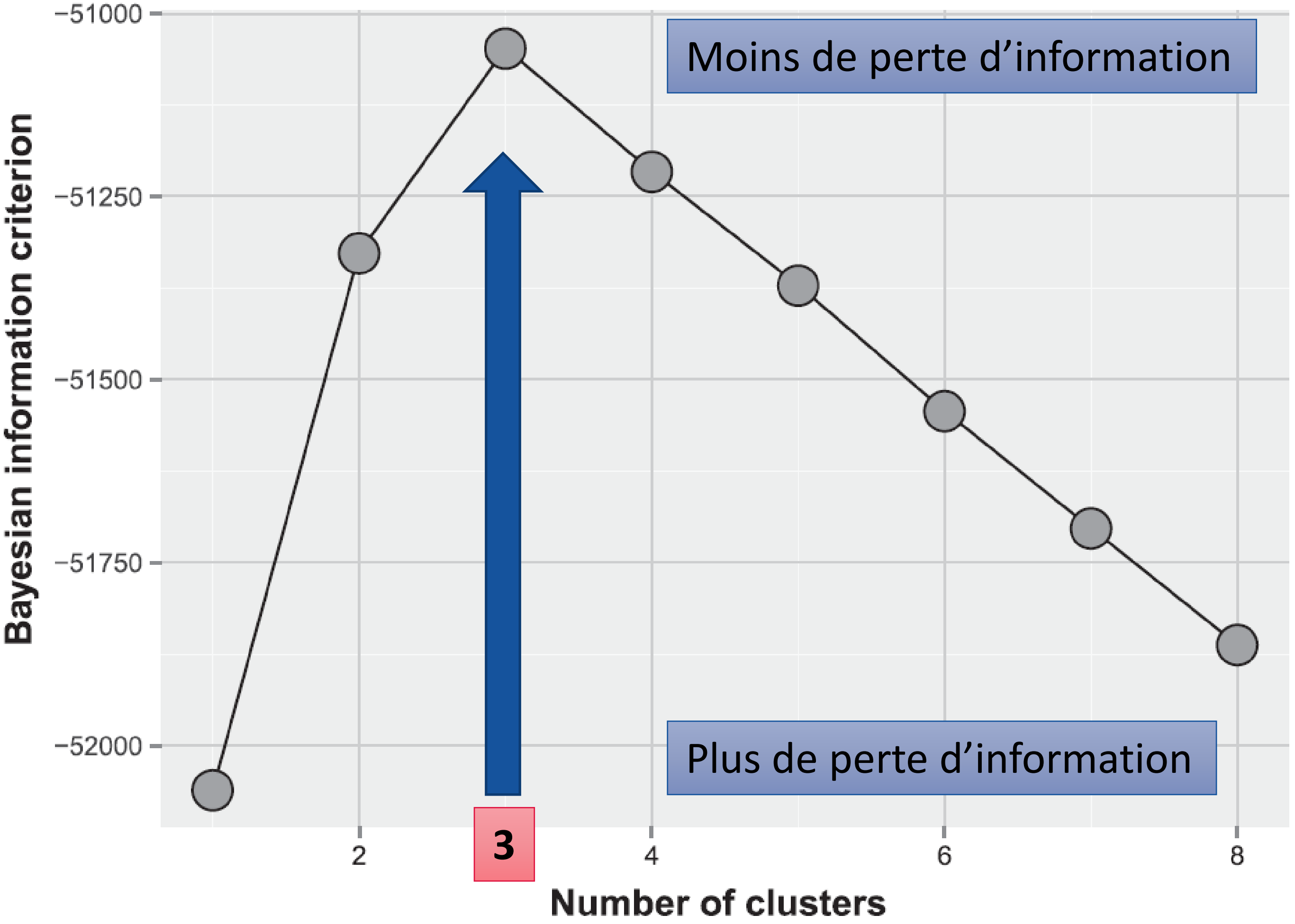
LA enlargement / normal size cardiac chamb.

Bi atrial enlargement / severe RV dysfunction

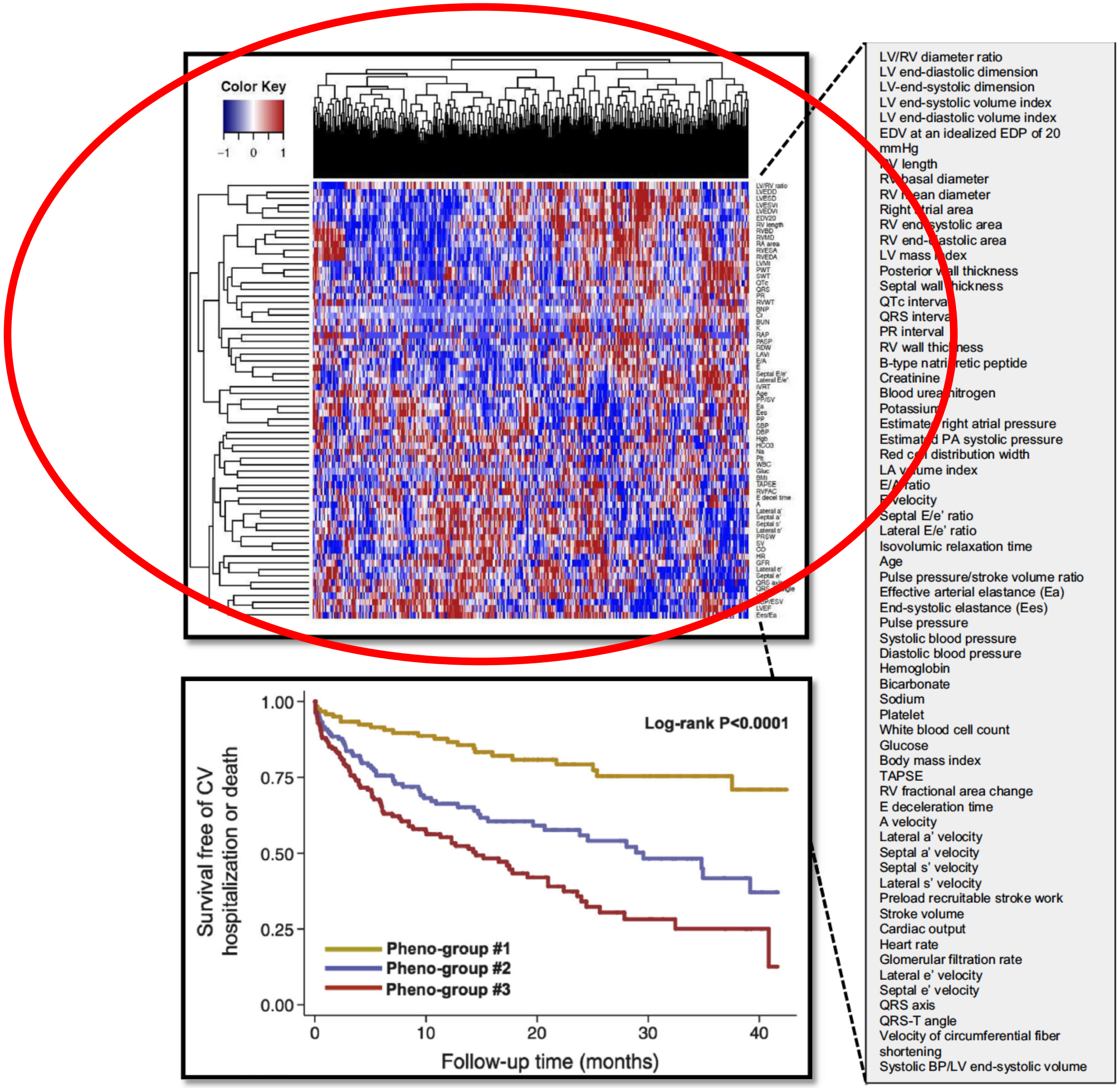


Pourquoi 3 groupes de patients

Estimation de la perte d'information lorsqu'on utilise un modèle donné pour représenter le processus qui génère les données



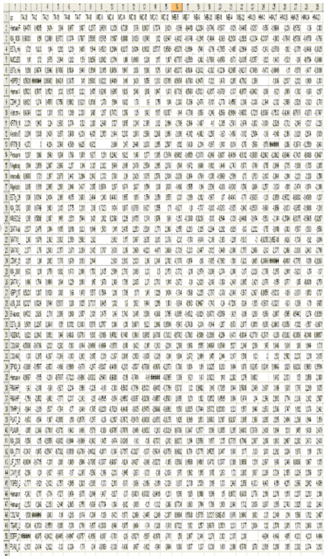
Représentation par technique de clustering

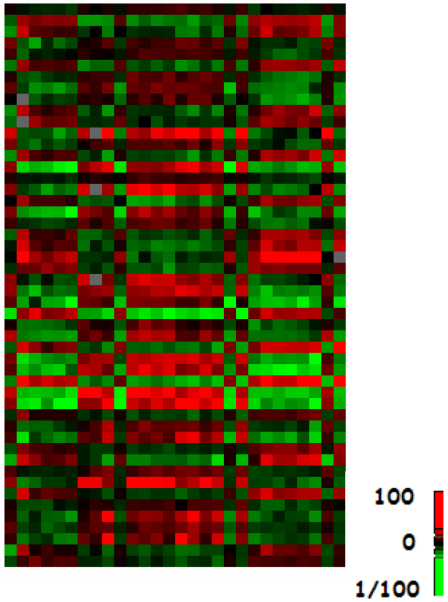


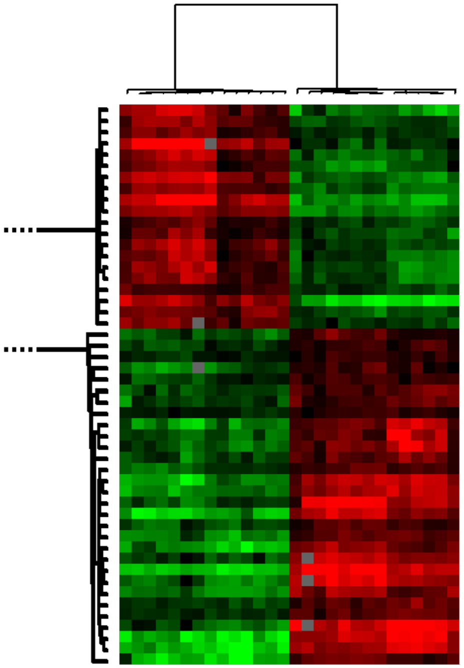
Clustering

échantillons

biomarqueurs







Données → **code couleur** → **clustering hiérarchique**

Approche multiparamétrique en biologie

Etude d'expression des gènes

Niveau d'expression ⟷ Activité biologique

<u>Objectif</u>: Vue d'ensemble de l'expression du génome
(profil d'expression génique)

Profil d'expression génique dynamique

↕

Activité biologique modifiée
→ fonctions des gènes

Génomique fonctionnelle et maladies

Phénotype malade

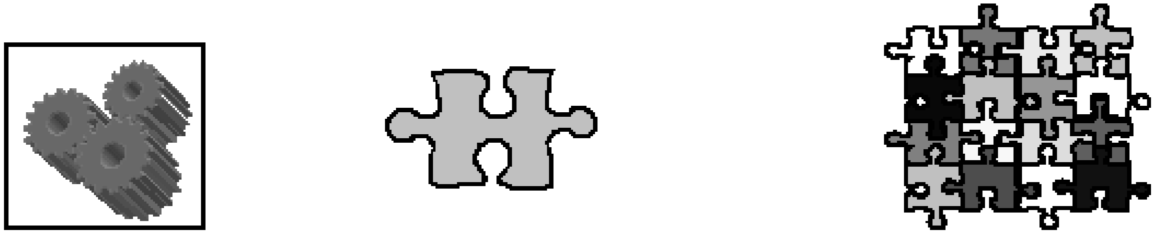
Variations génétiques + **facteurs d'environnement et facteurs de risque**

- Effet individuel très modeste
- Accumulation et combinaison d'un nombre important de variations

- Modulation de l'effet des variations génétiques

Faible contribution de chaque variations génétique au phénotype → Intérêt d'une analyse plus globale

Etude d'expression des gènes



Gène ------------------- Génome

Transcription

ARNm .... **Transcriptome**

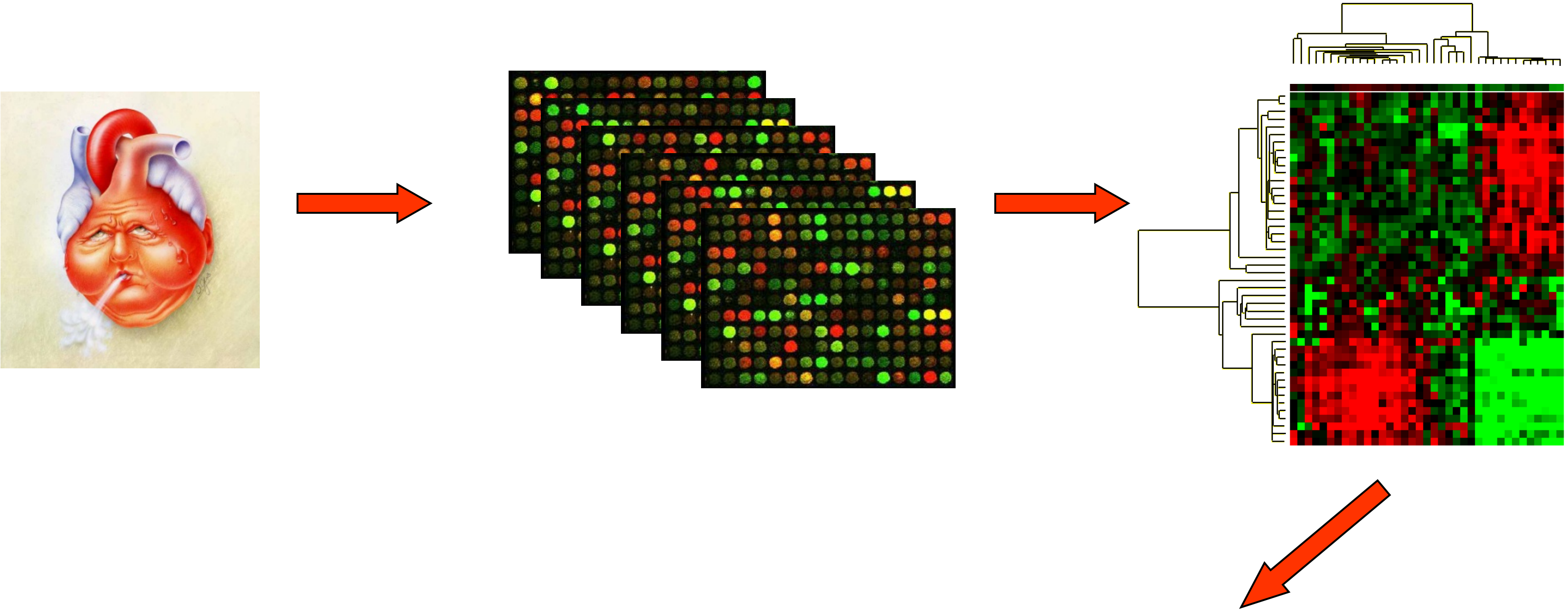
Traduction

Protéine .... **Protéome**

Modifications post-traductionnelles

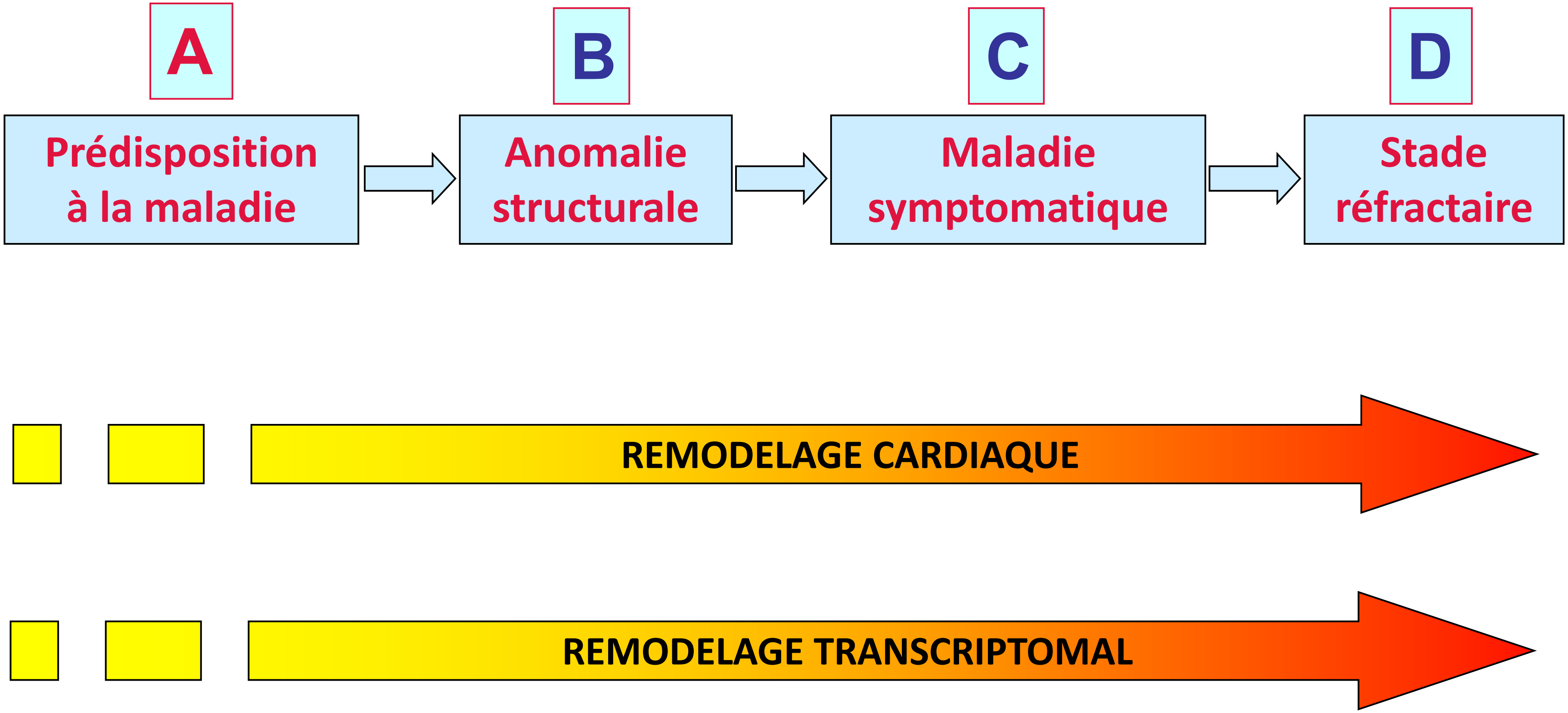
Fonction .... Physiome

Profils d’expression génique



1. **Découvertes physiopathologiques**
2. **Phénotypage moléculaire**

Insuffisance cardiaque: remodelage myocardiaque



A
Prédisposition à la maladie

B
Anomalie structurale

C
Maladie symptomatique

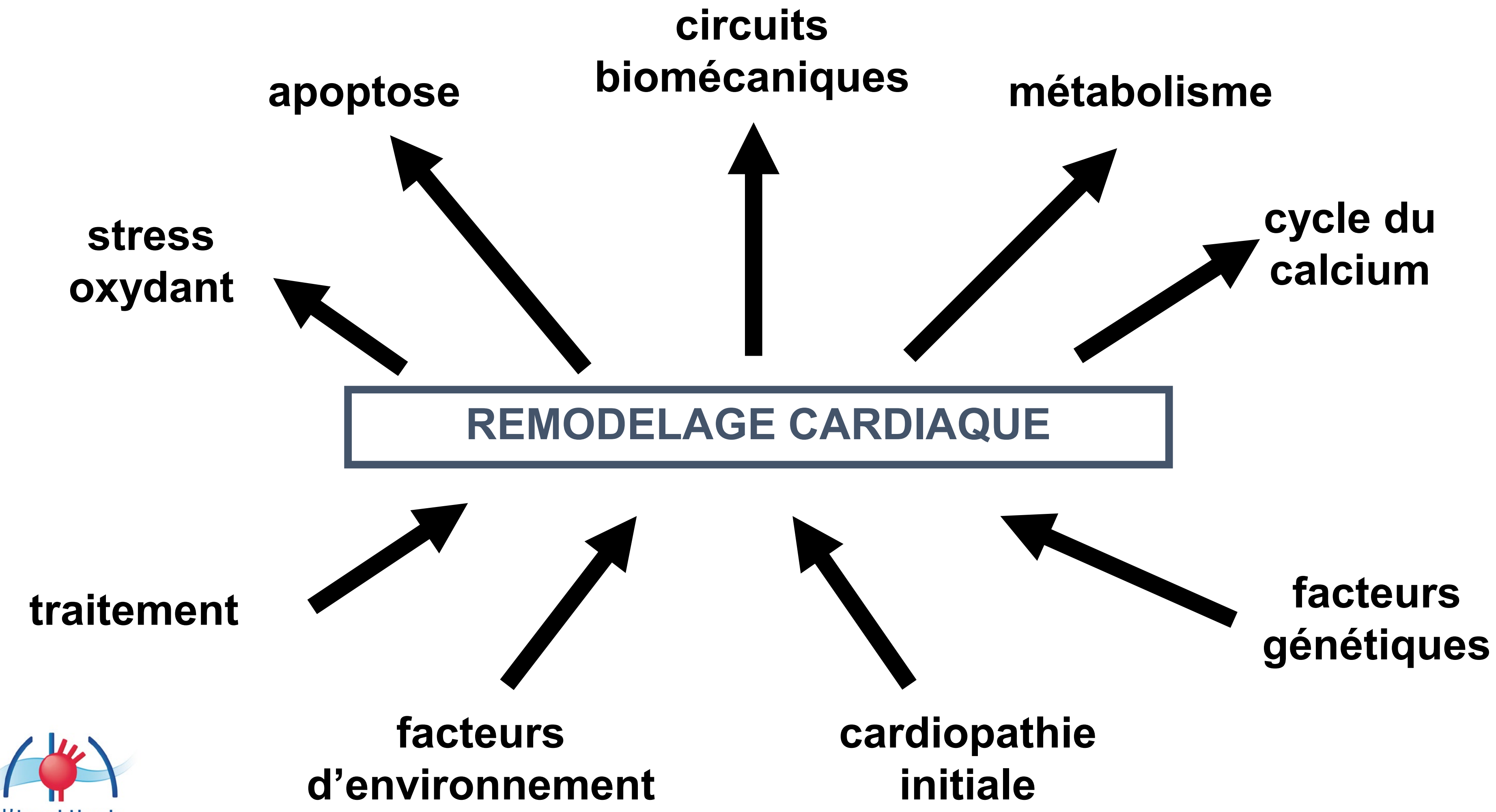
D
Stade réfractaire

REMODELAGE CARDIAQUE

REMODELAGE TRANSCRIPTOMAL

l'institut du thorax

Insuffisance cardiaque: processus de remodelage complexe



circuits biomécaniques

apoptose

métabolisme

stress oxydant

cycle du calcium

REMODELAGE CARDIAQUE

traitement

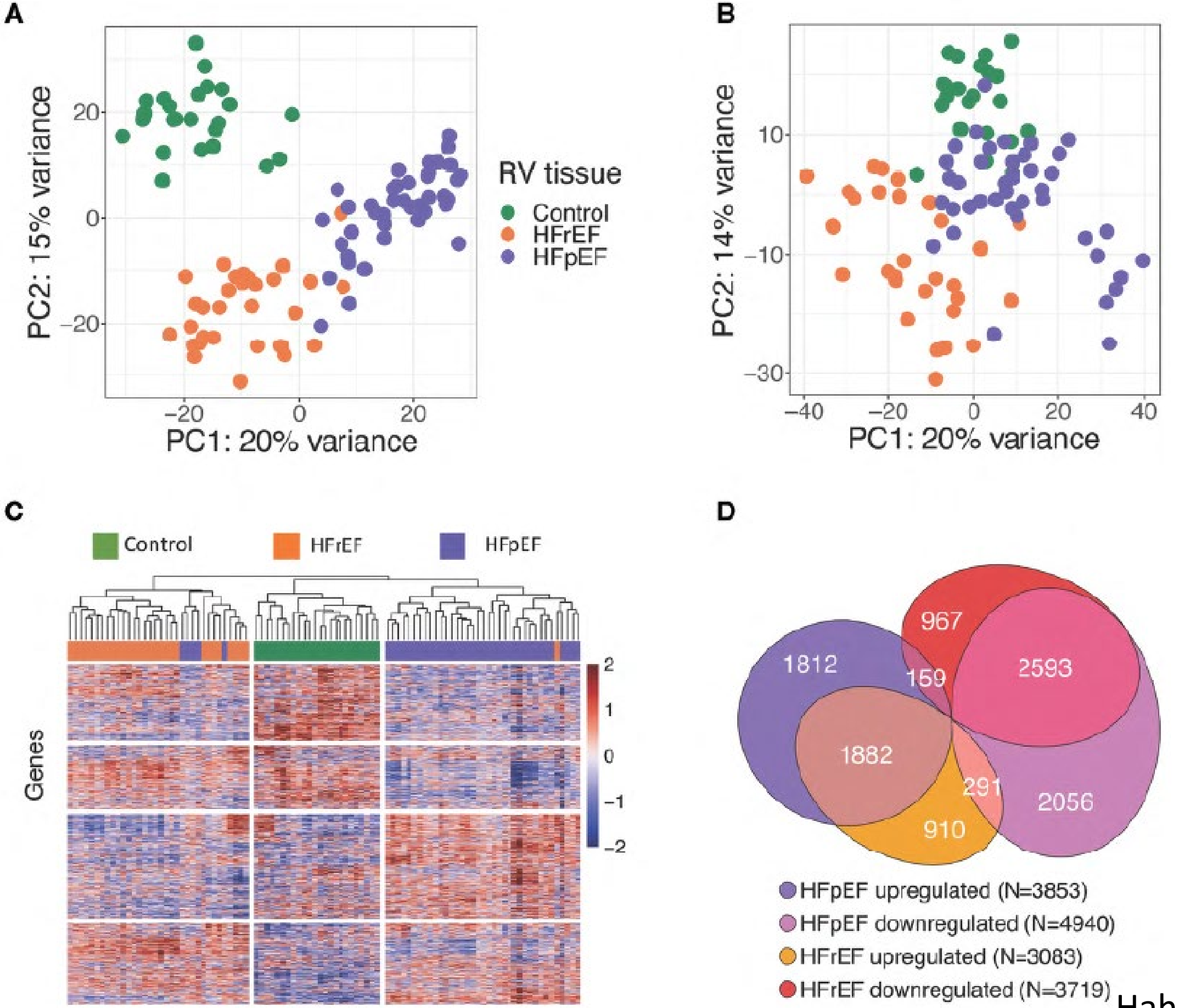
facteurs génétiques

facteurs d'environnement

cardiopathie initiale

l'institut du thorax

Insuffisance cardiaque: Etude du transcriptomique cardiaque



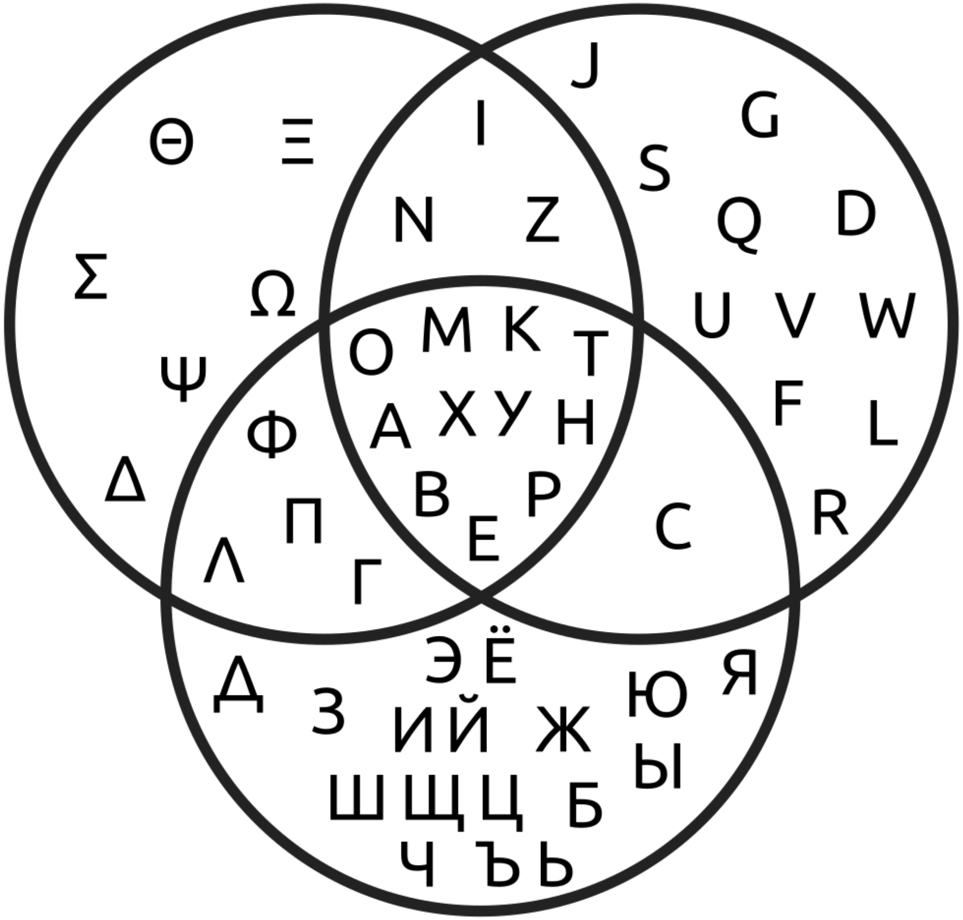
Biopsies de VD:
- Contrôle
- HFpEF
- HFrEF

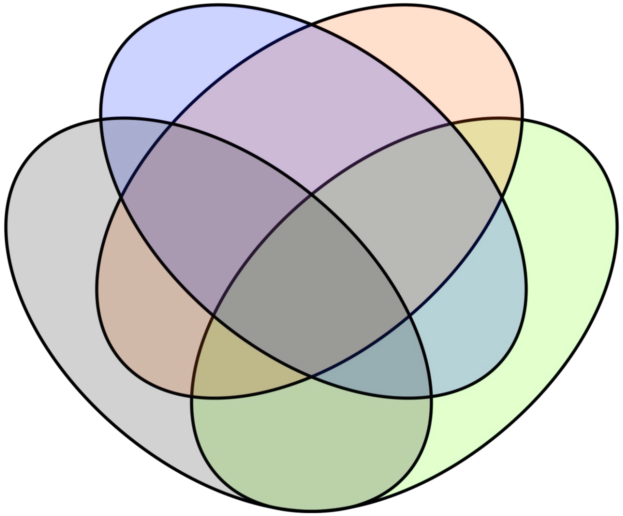
Hahn, Circulation (2021)

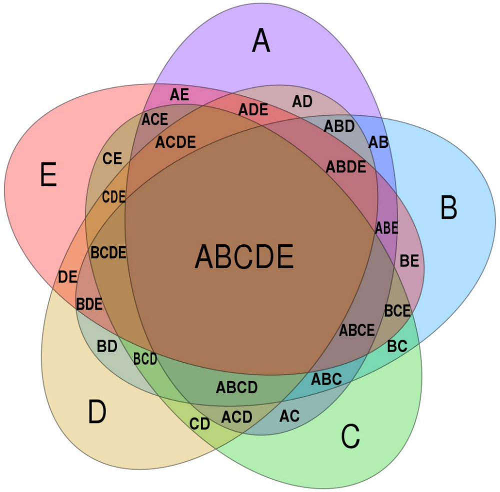
Interprétation des données

- Detection de gènes différentiels
  - Méthodes statistiques
- Detection de groupes de gènes co-regulés
  - Méthodes de « clustering » et visualisation
- Annotation fonctionnelle
  - Banques d'information (www)
    - Gene Ontology, Reactome,…

Diagrammes de Venn







Type I and II Errors

Decision	**Actual Situation “Truth”**	
	H_0 True	**H_0 False**
Do Not Reject H_0	**Correct Decision** **1 - α**	**Incorrect Decision** **Type II Error** **β**
Rejct H_0	**Incorrect Decision** **Type I Error** **α**	**Correct Decision** **1 - β**

$$\alpha = P(\text{Type I Error}) \quad \beta = P(\text{Type II Error})$$

Why Multiple Testing Matters

Genomics = Lots of Data = Lots of Hypothesis Tests

- Environ 10000-15000 gènes s'expriment en même temps dans une cellule ou un tissu
- Analyser des différences d'expression entre 2 situations (sains *vs.* contrôle) pour chaque gène fait réaliser 10000 à 15000 tests statistiques
- Avec un seuil de p-value à 0,05, on s'attend donc à avoir 500 à 750 gènes déclarés comme significatifs par erreur.

Why Multiple Testing Matters

- In general, if we perform m hypothesis tests, what is the probability of at least 1 false positive?

$$P(\text{Making an error}) = \alpha$$

$$P(\text{Not making an error}) = 1 - \alpha$$

$$P(\text{Not making an error in m tests}) = (1 - \alpha)^m$$

$$P(\text{Making at least 1 error in m tests}) = 1 - (1 - \alpha)^m$$

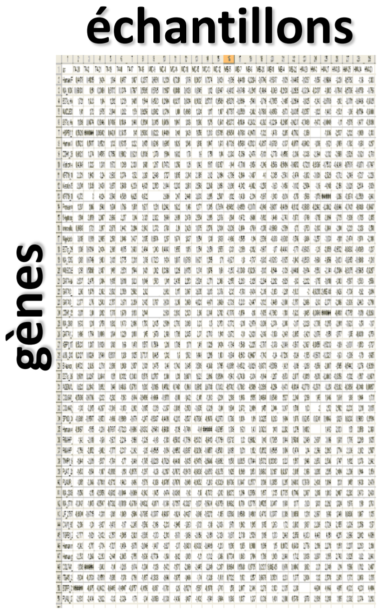
Il faut des méthodes pour contrôler l’erreur de type I

Interprétation des données

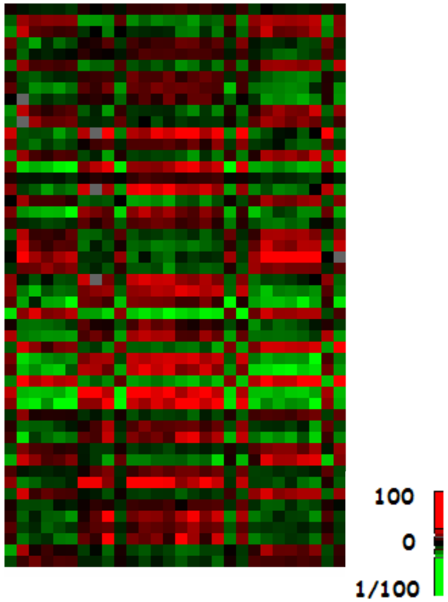
- **Detection de gènes différentiels**
  - **Méthodes statistiques**
- **Detection de groupes de gènes co-regulés**
  - **Méthodes de « clustering »**
- **Annotation fonctionnelle**
  - **Banques d'information (www)**
    - **Gene Ontology, Reactome,…**

l'institut
du thorax

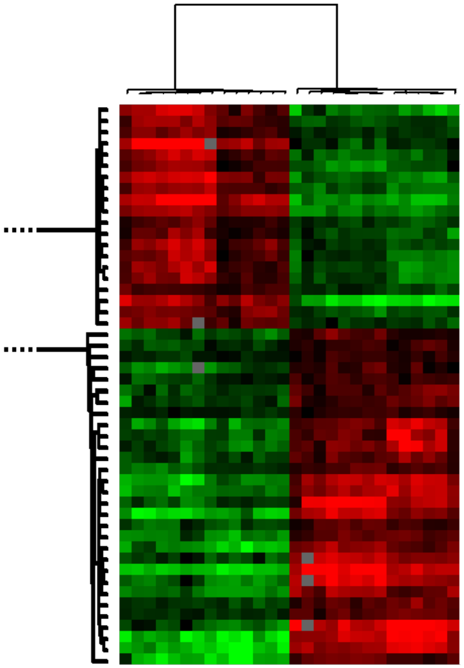
Clustering - Eisen et al. (1998)



Données d'expression

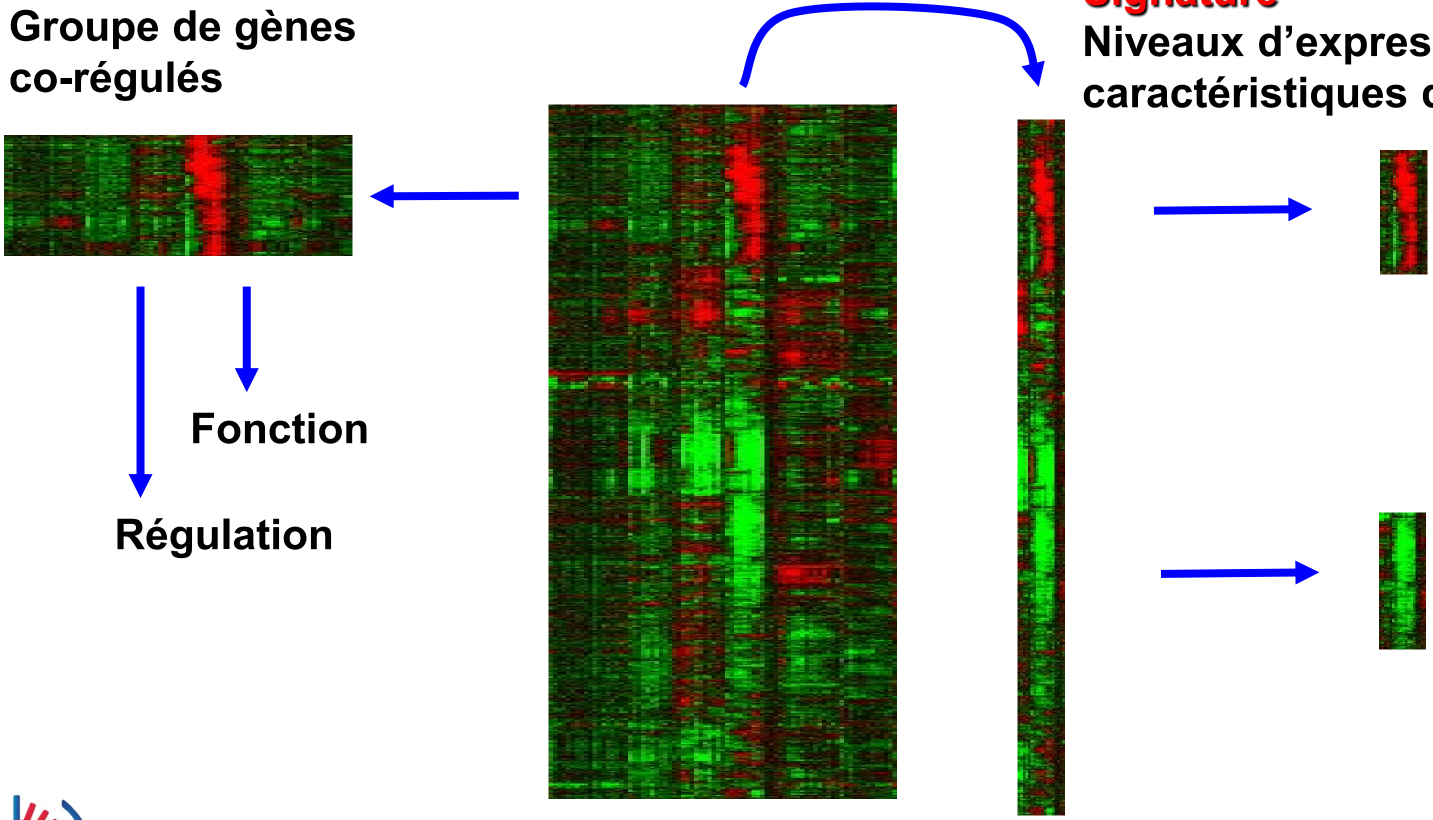


code couleur



clustering hiérarchique

Définitions



Profil
Mesure d’expression d’un échantillon

Cluster
Groupe de gènes co-régulés

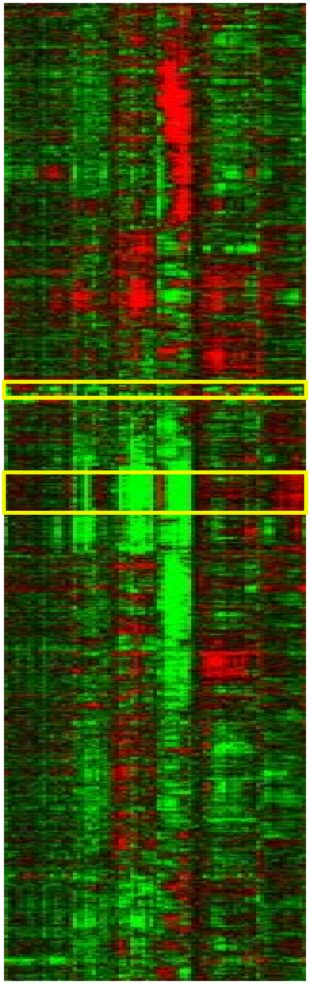
Signature
Niveaux d’expression de gènes caractéristiques d’un échantillon

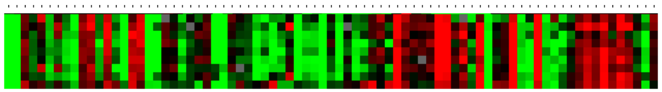
Fonction

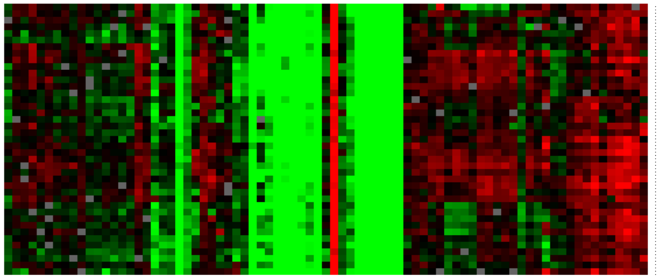
Régulation

Correlation = Fonction

Eisen et al. (1998) PNAS 95: 14863-14868







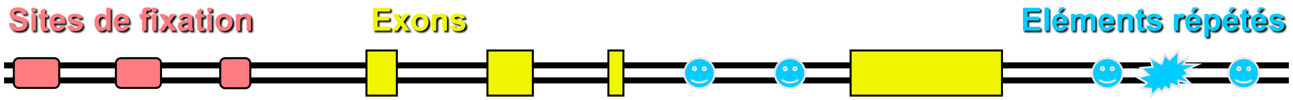
HTB2	CHROMATIN STRUCTURE	HISTONE H2B
HHT1	CHROMATIN STRUCTURE	HISTONE H3
HHF1	CHROMATIN STRUCTURE	HISTONE H4
HTA1	CHROMATIN STRUCTURE	HISTONE H2A
HTB1	CHROMATIN STRUCTURE	HISTONE H2B
HHF2	CHROMATIN STRUCTURE	HISTONE H4
HTA2	CHROMATIN STRUCTURE	HISTONE H2A
HHT2	CHROMATIN STRUCTURE	HISTONE H3
HHO1	CHROMATIN STRUCTURE	HISTONE H1

RPS5	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN S5
RPS4A	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN S4A
RPL26B	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN L26B
RPS7A	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN S7A
RPS24A	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN S24A
RPS21A	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN S21A
RPL14A	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN L14A
RPL38	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN L38
RPL24A	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN L24A
RPS1A	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN S1A
RPS1B	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN S1B
RPL24B	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN L24B
RPS23A	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN S23A
RPS23B	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN S23B
RPS3	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN S3
RPS4B	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN S4B
RPS6B	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN S6B
RPL5	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN L5
RPL17B	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN L17B
RPS19A	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN S19A
RPL8B	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN L8B
RPS19B	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN S19B
RPS26B	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN S26B
RPL11A	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN L11A
RPL11B	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN L11B
RPL34B	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN L34B
RPL10	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN L10
RPL12B	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN L12B
RPS15	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN S15
RPS20	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN S20
RPS0A	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN S0A
RPL18B	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN L18B
RPL2B	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN L2B
RPL16B	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN L16B
RPS0B	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN S0B
RPL7B	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN L7B
RPL7A	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN L7A
RPL33B	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN L33B
RPL6B	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN L6B
RPL33A	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN L33A
RPL13A	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN L13A

Les gènes d’expression corrélée participent à une même fonction biologique

Cluster = Régulation

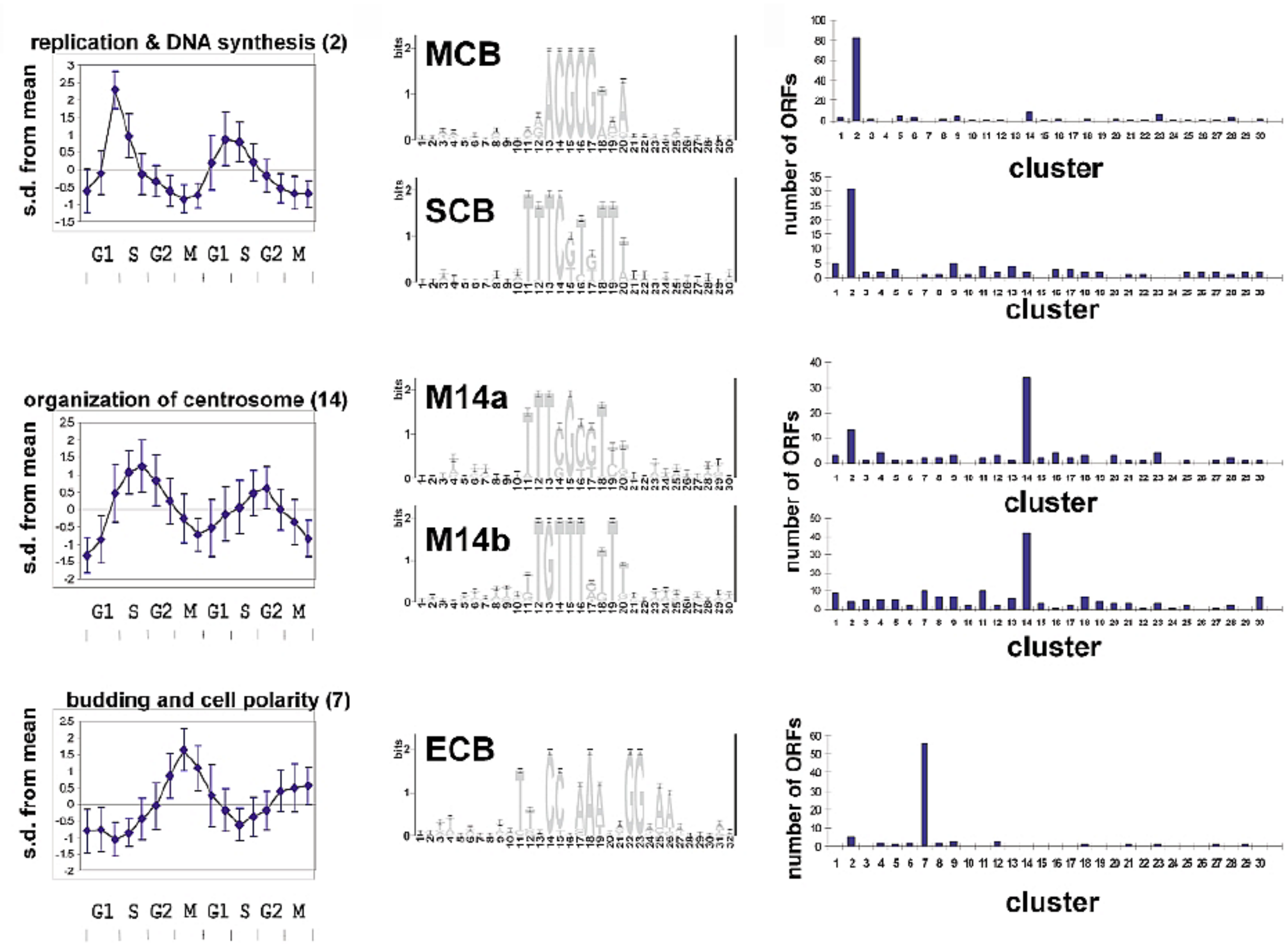
Tavazoie et al. (1998) Nature Genetics



Sites de fixation

Exons

Eléments répétés



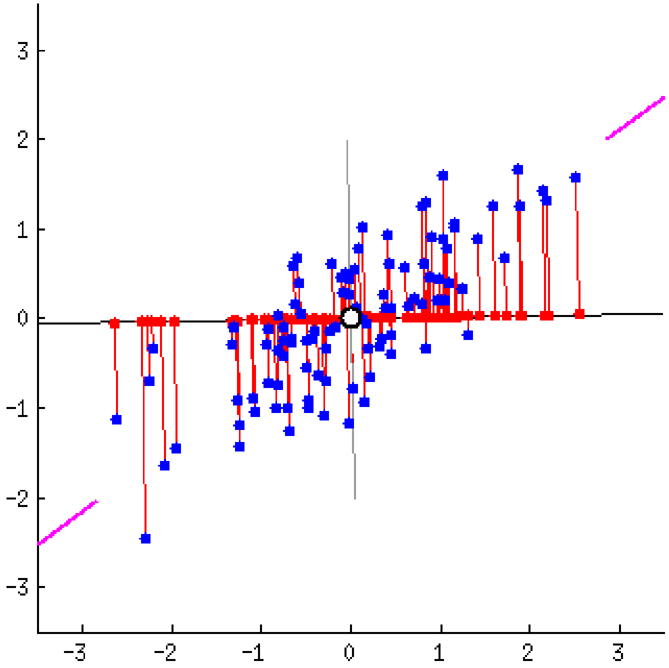
Interprétation des données

- **Detection de gènes différentiels**
  - **Méthodes statistiques**
- **Detection de groupes de gènes co-regulés**
  - **Visualisation = réduction de dimension**
- **Annotation fonctionnelle**
  - **Banques d'information (www)**
    - Gene Ontology, Reactome,…

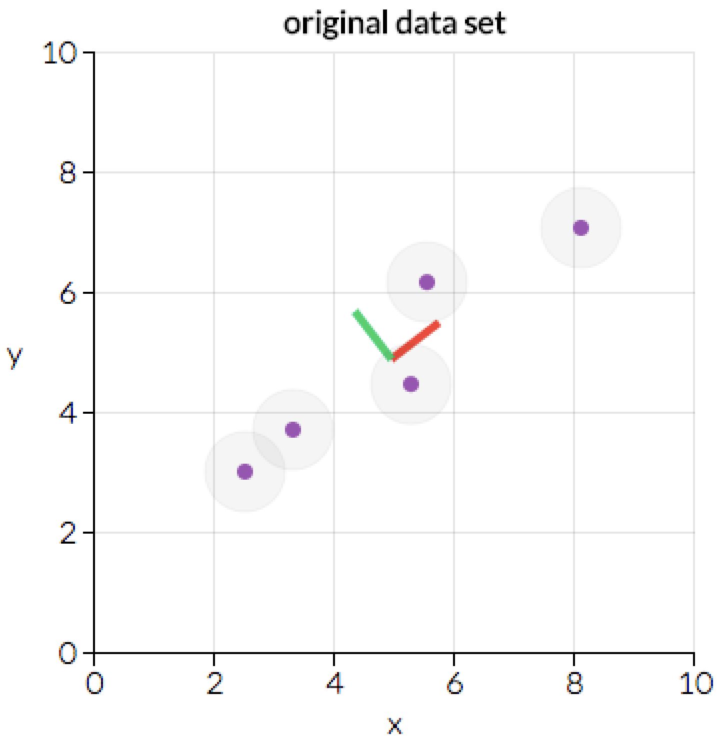
l'institut du thorax

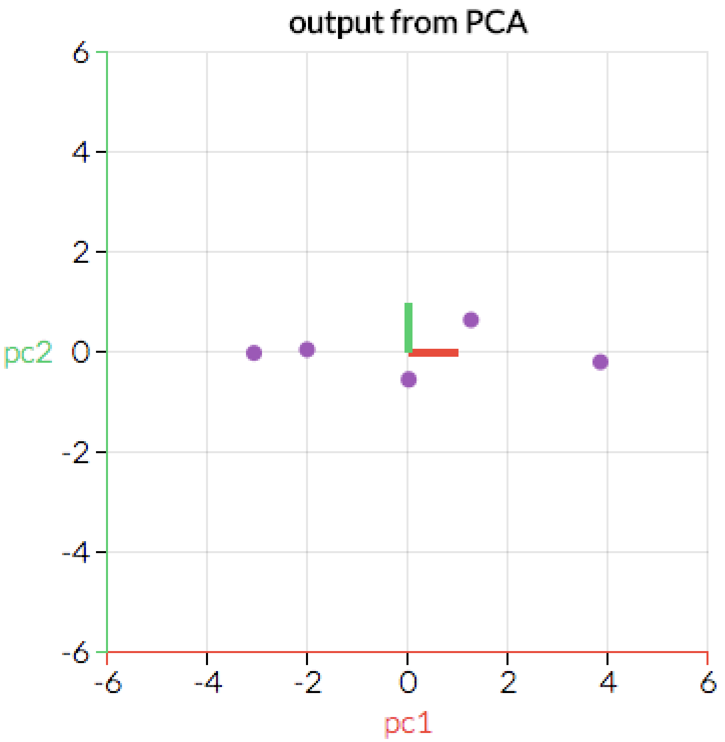
Analyse en composante principale

- Une méthode de réduction du nombre de variables utiles en préservant au maximum la quantité d'information
- Création séquentielle de nouvelles variables (« pseudogènes ») qui capturent à chaque fois le plus possible d'information discriminante
- Plus de simplicité/lisibilité en limitant la perte de précision/information

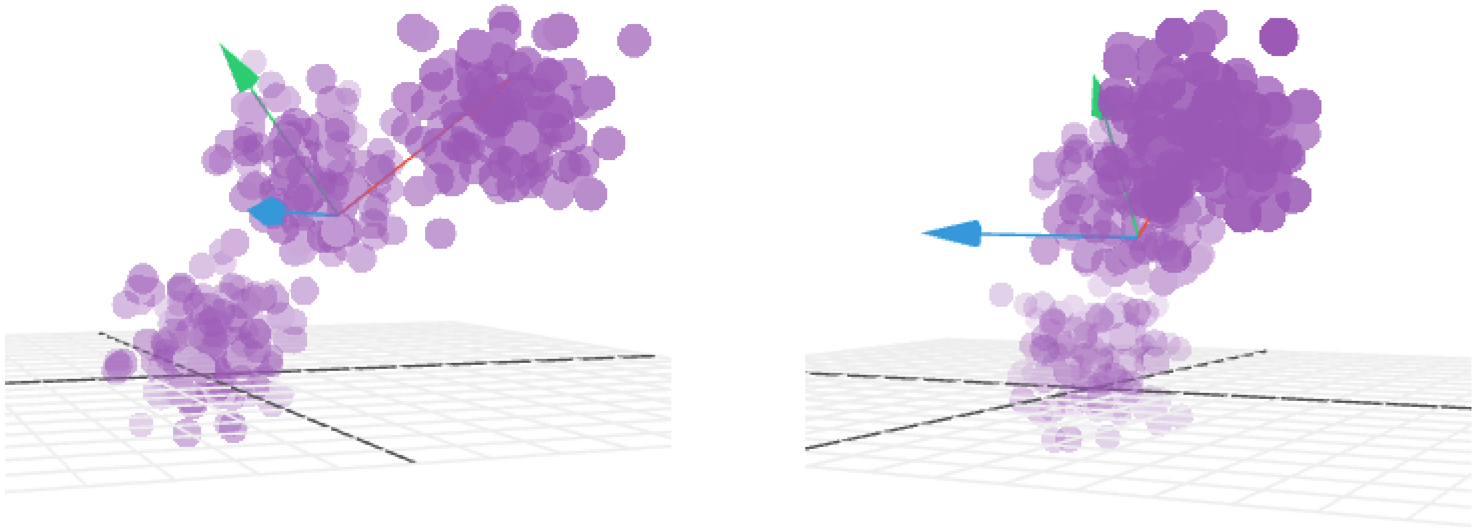


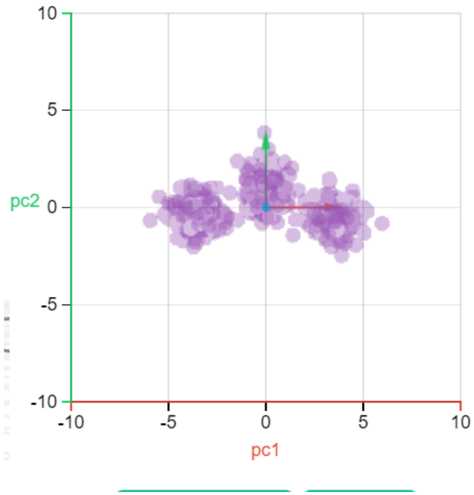
5 échantillons et 2 variables

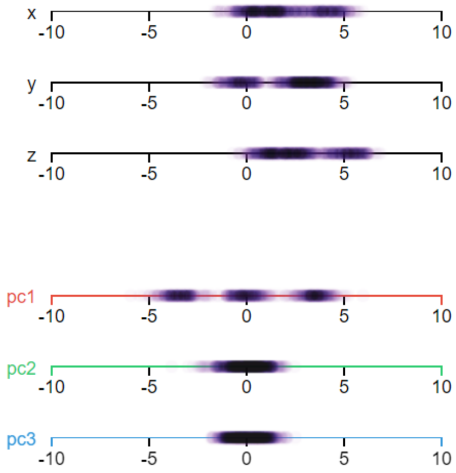




Nombreux échantillons et 3 variables

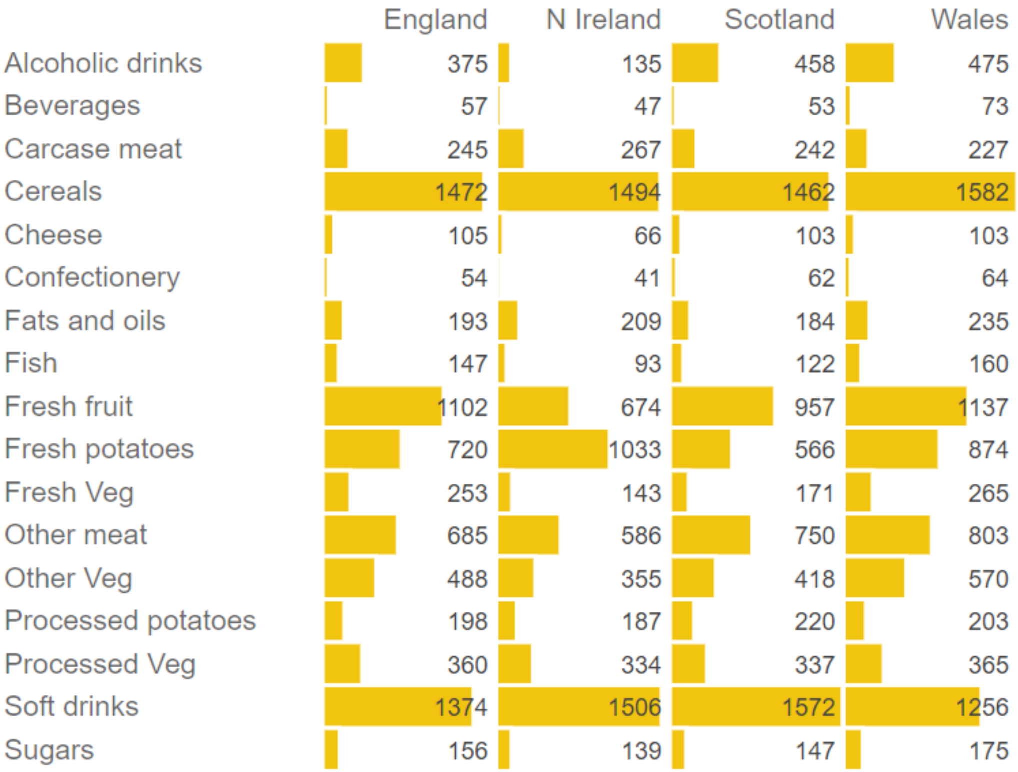




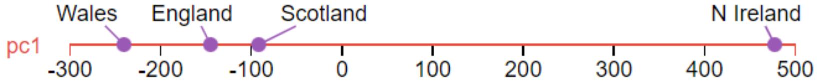


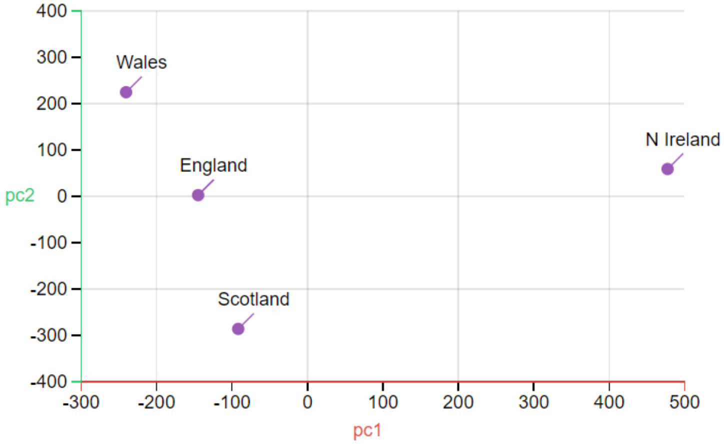
https://setosa.io/ev/principal-component-analysis/

4 groupes et 17 variables



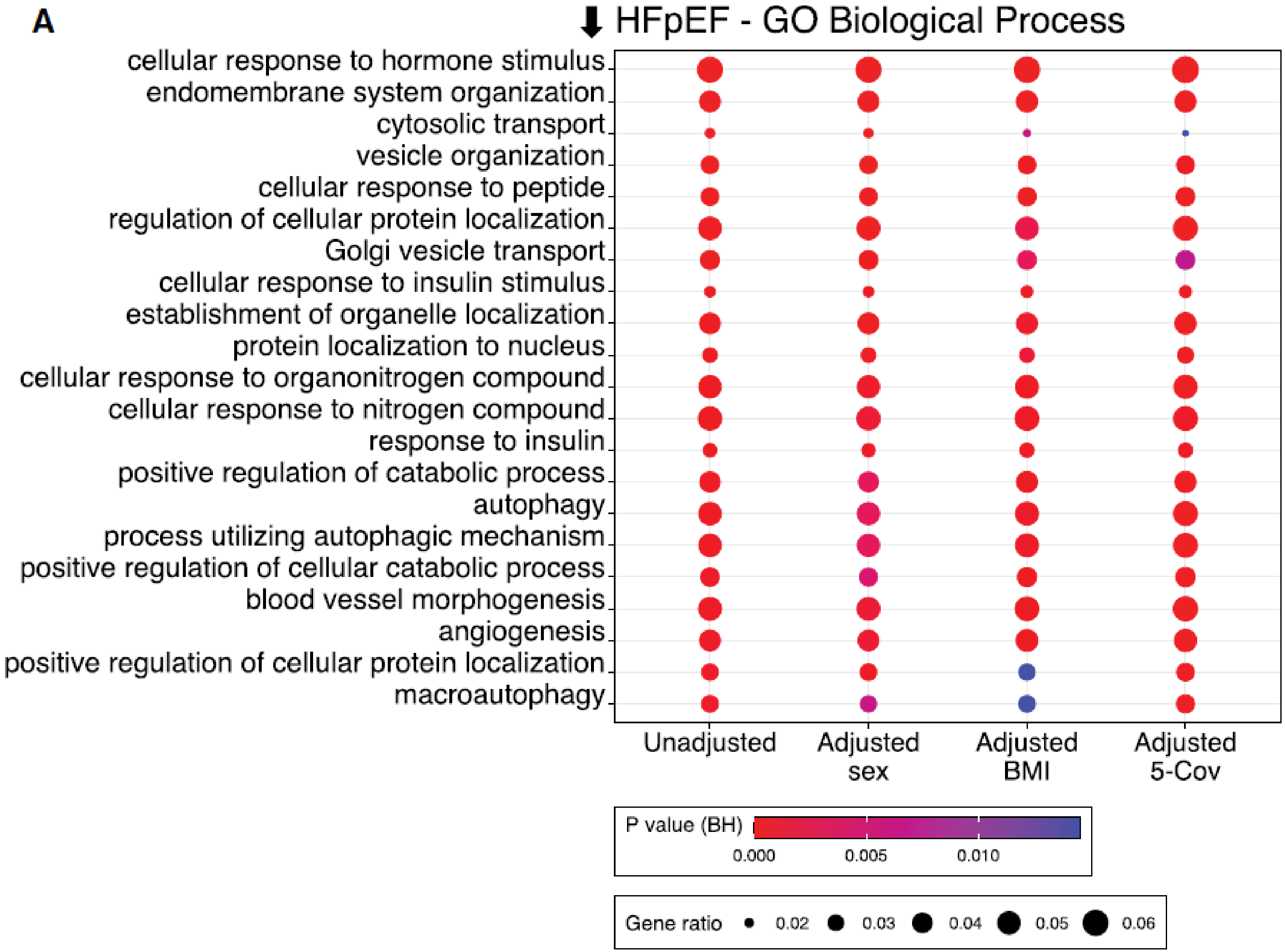
	England	N Ireland	Scotland	Wales
Alcoholic drinks	375	135	458	475
Beverages	57	47	53	73
Carcase meat	245	267	242	227
Cereals	1472	1494	1462	1582
Cheese	105	66	103	103
Confectionery	54	41	62	64
Fats and oils	193	209	184	235
Fish	147	93	122	160
Fresh fruit	1102	674	957	1137
Fresh potatoes	720	1033	566	874
Fresh Veg	253	143	171	265
Other meat	685	586	750	803
Other Veg	488	355	418	570
Processed potatoes	198	187	220	203
Processed Veg	360	334	337	365
Soft drinks	1374	1506	1572	1256
Sugars	156	139	147	175





https://setosa.io/ev/principal-component-analysis/

A quelles fonctions biologiques participent les gènes sous-exprimés



Hahn, Circulation (2021)

Interprétation des données

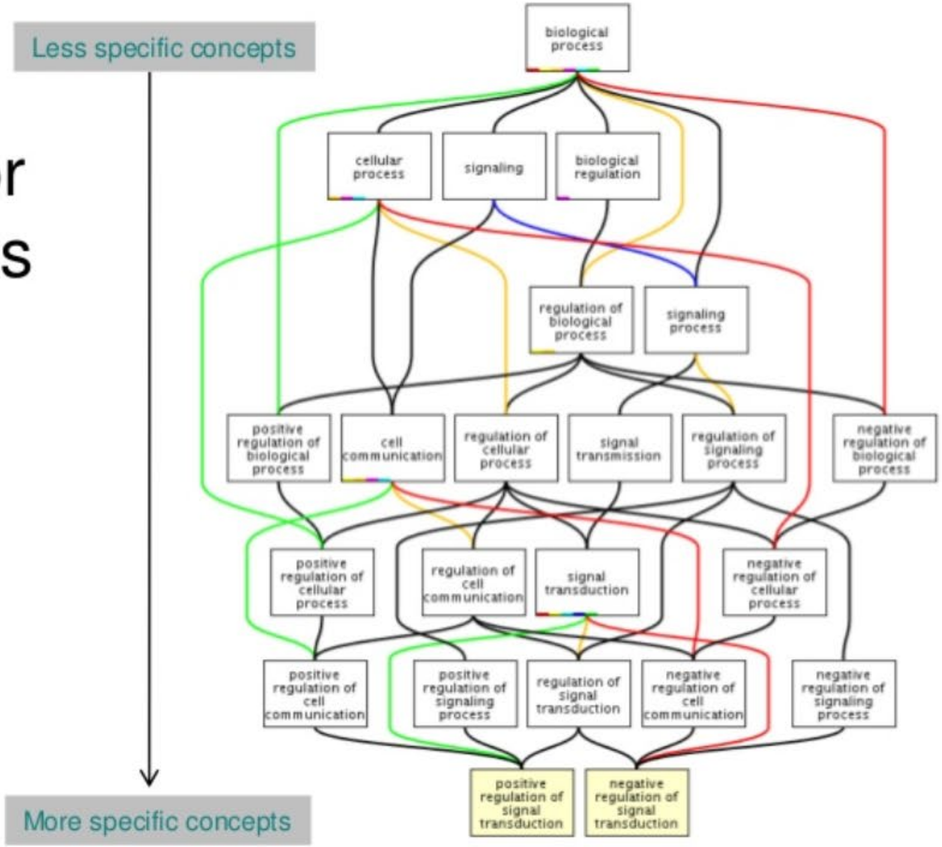
- **Détection de gènes différentiels**
  - **Méthodes statistiques**
- **Détection de groupes de gènes co-regulés**
  - **Méthodes de « clustering » et visualisation**
- **Annotation fonctionnelle**
  - **Banques d'information (www)**
    - **Gene Ontology, Reactome,…**

l'institut du thorax

The Gene Ontology

- A way to capture biological knowledge for individual gene products in a written and computable form

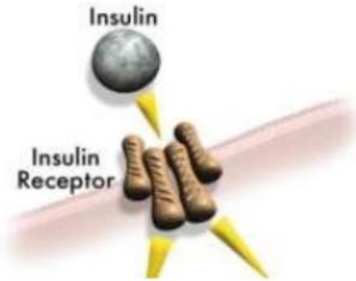
- A set of concepts and their relationships to each other arranged as a hierarchy



http://www.ebi.ac.uk/QuickGO

1. Molecular Function

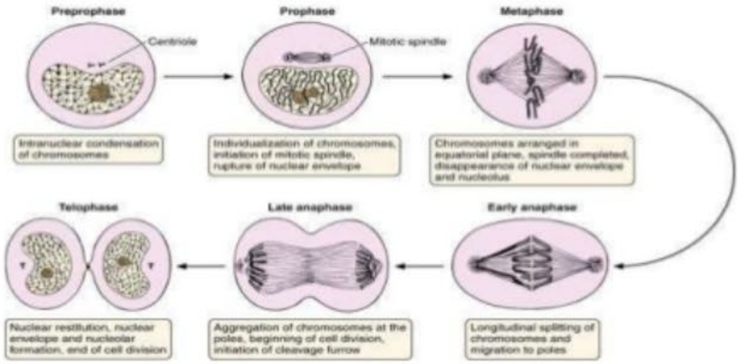
An elemental activity or task or job



- protein kinase activity
- insulin receptor activity

2. Biological Process

A commonly recognized series of events



Intranuclear condensation of chromosomes

Individualization of chromosomes, initiation of mitotic spindle, rupture of nuclear envelope

Chromosomes arranged in equatorial plane, spindle completed, disappearance of nuclear envelope and nucleolus

Longitudinal splitting of chromosomes and migration to poles

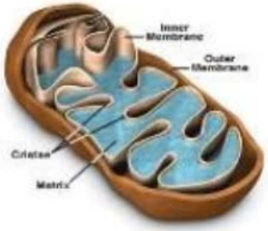
Aggregation of chromosomes at the poles, beginning of cell division, initiation of cleavage furrow

Nuclear restitution, nuclear envelope and nucleolar formation, end of cell division

- cell division

3. Cellular Component

Where a gene product is located



- mitochondrion
- mitochondrial matrix
- mitochondrial inner membrane

A GO annotation is …

…a statement that a gene product;

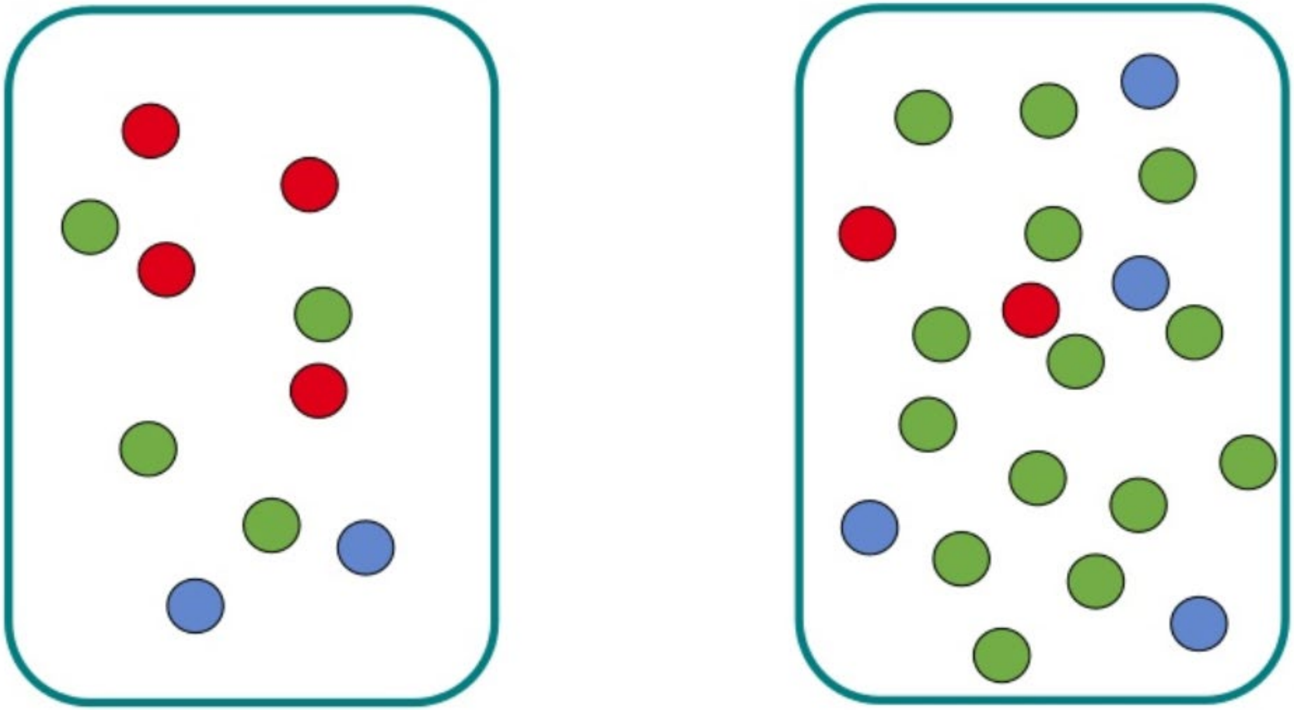
1. has a particular molecular function
   or is involved in a particular biological process
   or is located within a certain cellular component

2. as described in a particular reference

3. as determined by a particular method

Accession	*Name*	*GO ID*	*GO term name*	*Reference*	*Evidence code*
P00505	GOT2	GO:0004069	aspartate transaminase activity	PMID:2731362	IDA

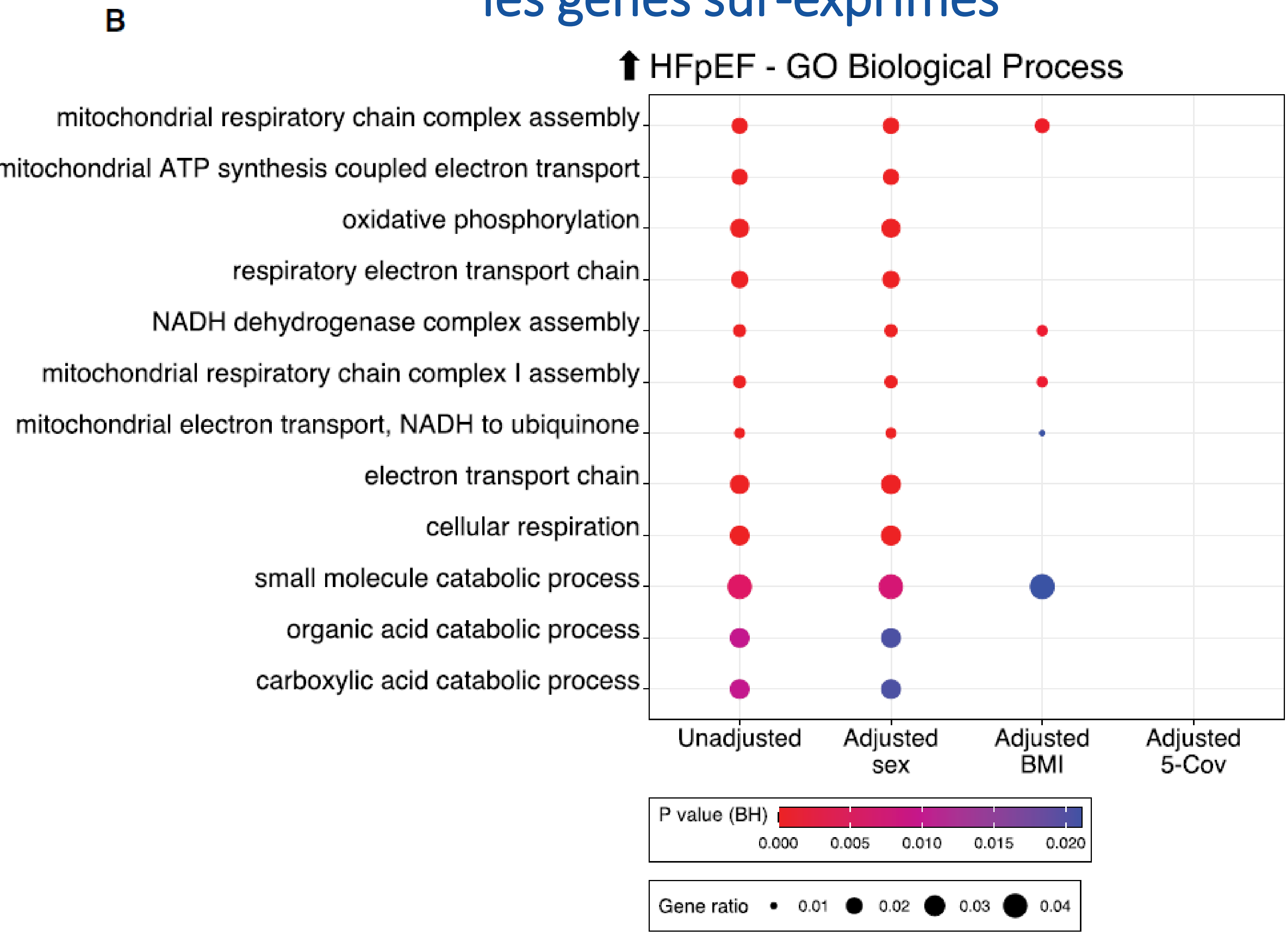
Enrichment analysis



	Sample	Reference
🔴	40%	20%
🔵	20%	20%

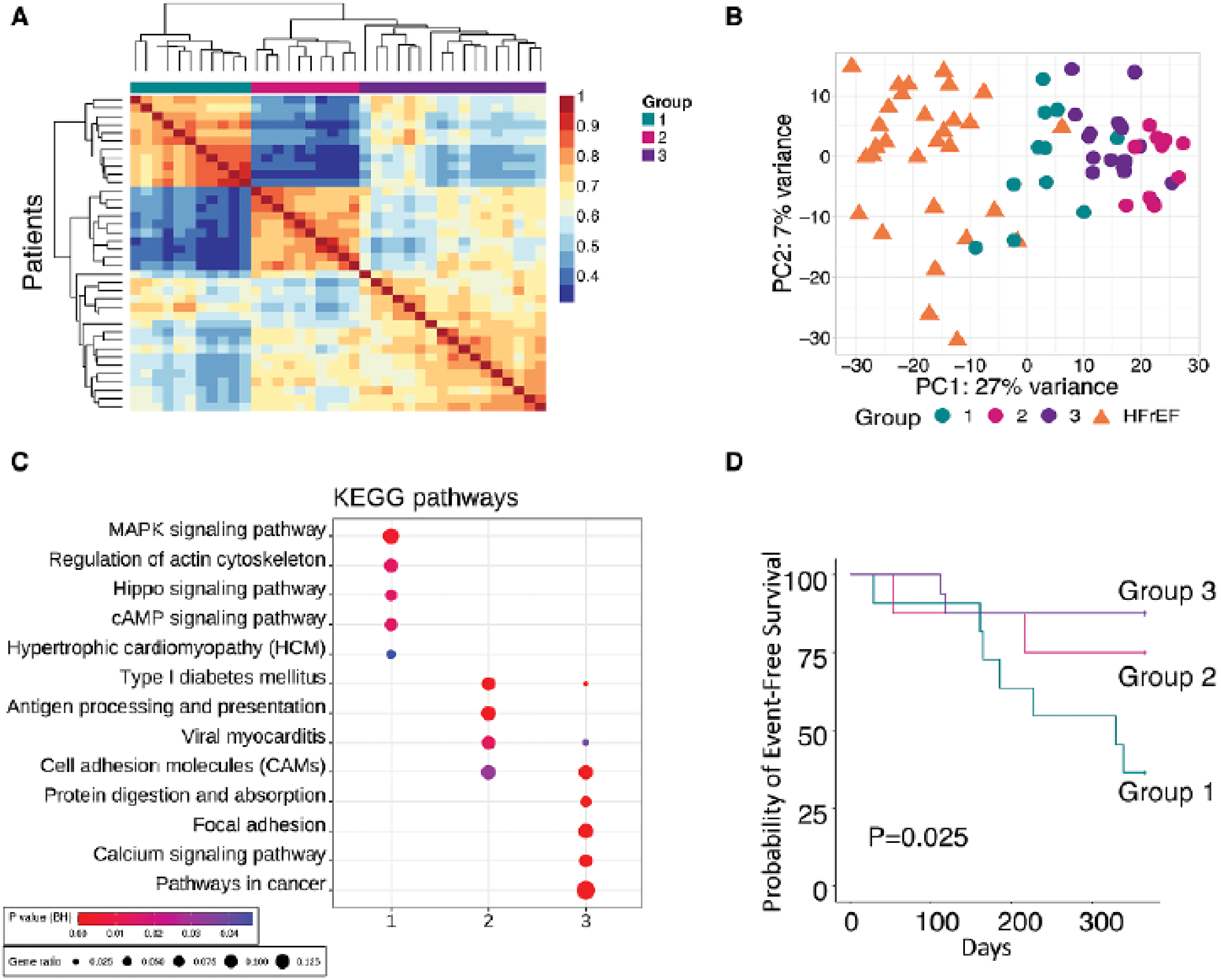
=> The sample is over-enriched for 🔴

A quelles fonctions biologiques participent les gènes sur-exprimés



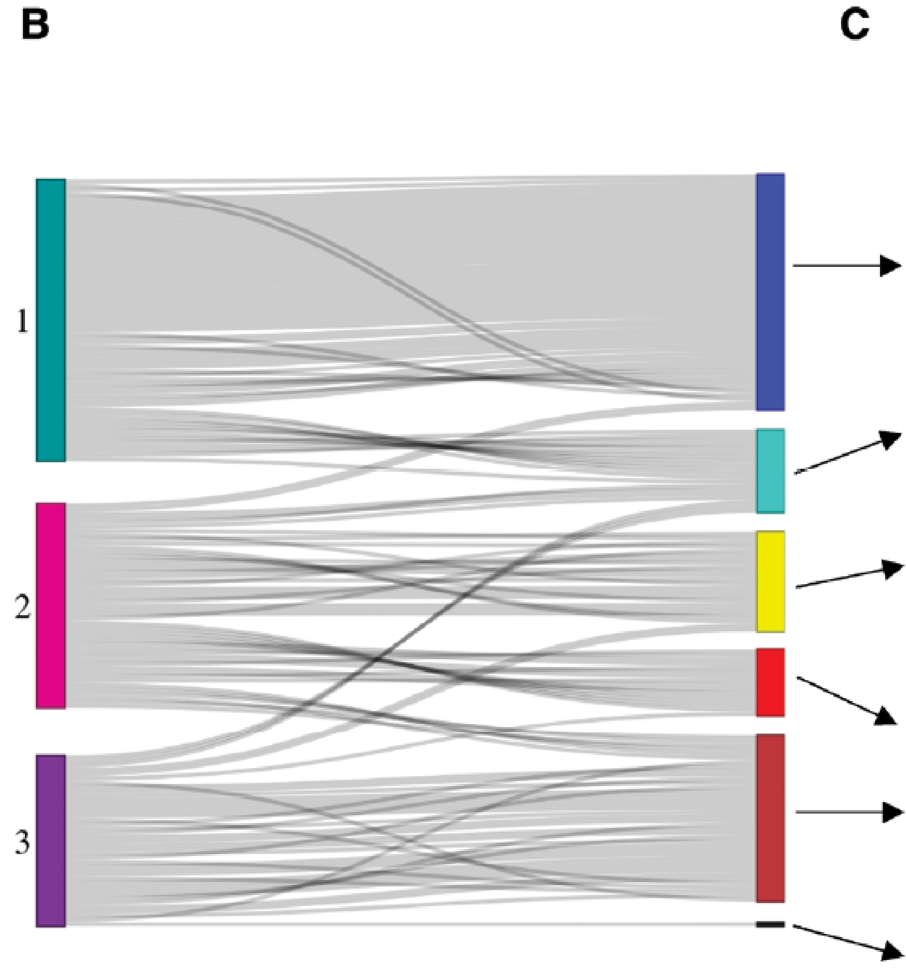
Hahn, Circulation (2021)

Sous-groupes de patients HFpEF



Hahn, Circulation (2021)

Sous-groupes de gènes et sous-groupes de patients



C

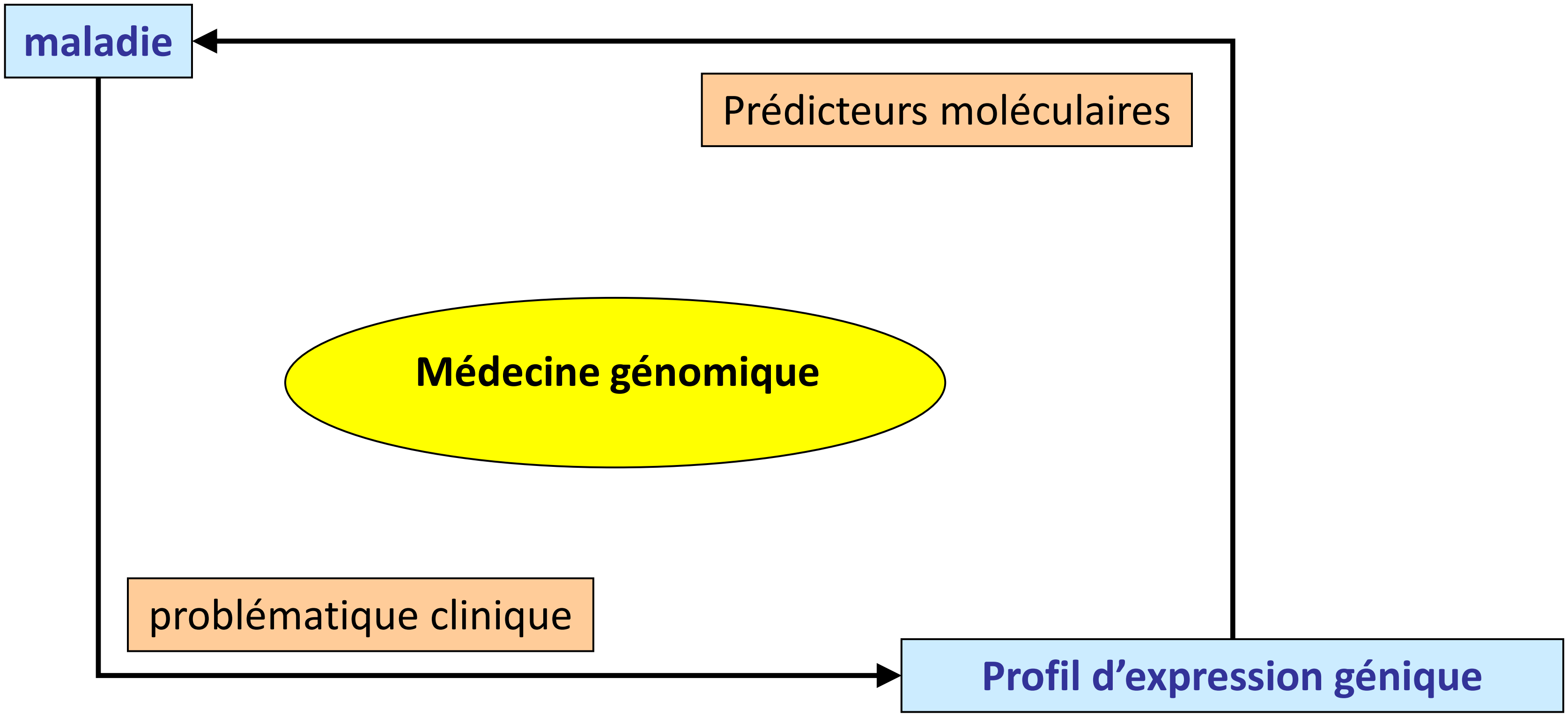
	Clinical Characteristics	Gene Ontology Biological Processes
Blue	Higher LV wall thickness and myocyte hypertrophy, Higher RV afterload and pulmonary vascular resistance, higher NTproBNP	Regulation of cytoskeleton organization; Striated muscle cell development; Sarcomere organization; Actin filament organization; Activation of protein kinase activity; Stress-activated protein kinase signaling cascade
Turquoise	Lower BMI, higher LV wall thickness and myocyte hypertrophy	Histone modification; Golgi organization; Regulation of gene expression, epigenetic; Negative regulation of microtubule depolymerization
Yellow	More female, less RV load and hypertrophy	Oxidative phosphorylation; ATP biosynthesis; electron transport chain; mitochondrial translation and membrane organization; cell response to hypoxia; T-cell and antigen receptor signaling
Red	Female and higher +CD68	Protein targeting to endoplasmic reticulum; Negative regulation of protein ubiquitination
Brown	Worse NYHA functional class, smaller heart and lower cardiac output, less diabetes, better renal function, lower NTproBNP, higher +CD68	Extracellular matrix organization; Regulation of actin cytoskeleton organization; Blood vessel morphogenesis; Inflammatory response; Chemotaxis
Black	Lower blood pressure, less LV hypertrophy	mRNA processing; RNA splicing

L'expression des gènes comme pont
entre données cliniques et processus biologiques

l'institut du thorax

Hahn, Circulation (2021)

Perspectives



maladie

Prédicteurs moléculaires

Médecine génomique

problématique clinique

Profil d’expression génique

l’institut du thorax

Merci de votre attention!

Questions?

Guillaume LAMIRAULT

guillaume.lamirault@univ-nantes.fr
L'unité de recherche de l'institut du thorax
Inserm UMR 1087 / CNRS UMR 6291
Nantes, France

www.umr1087@univ-nantes.fr