



Génomique

Guillaume LAMIRAU

guillaume.lamirault@univ-nantes.fr

UE M1S&S-Thorax, 10 octobre 2022, Nantes

Vue d'ensemble du génome humain (2004)

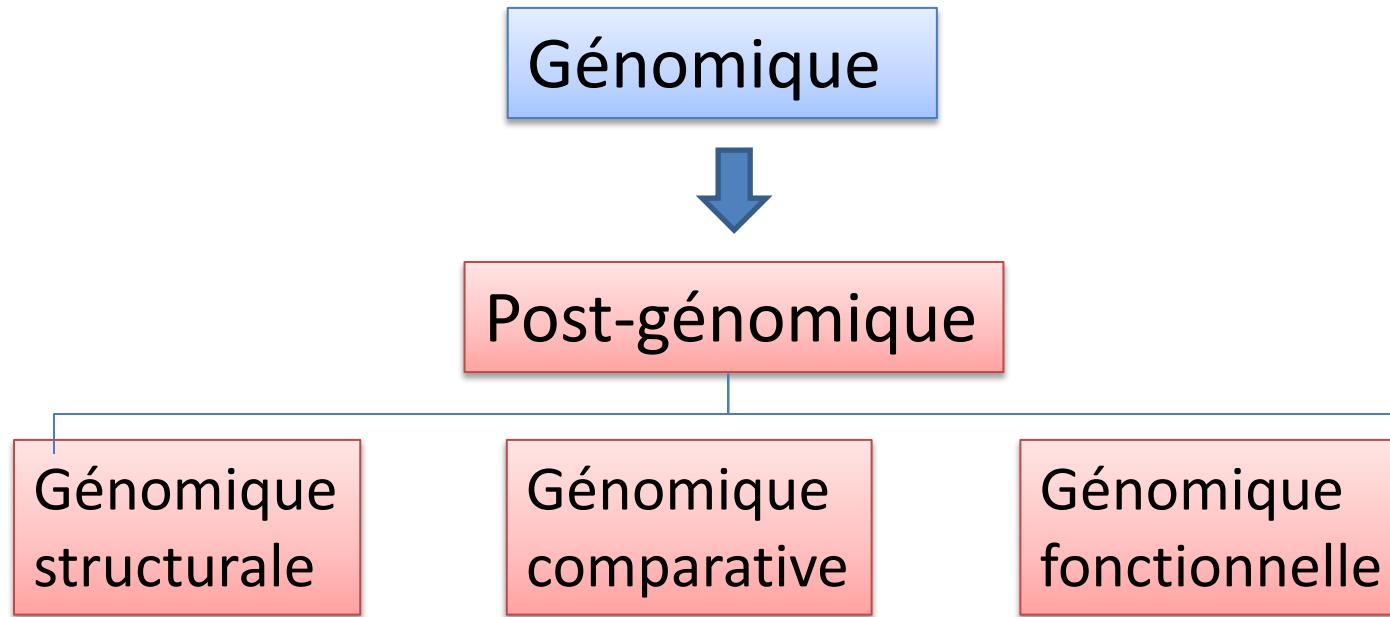
Taille du génome	3.08 Gbases
Taux d'erreur dans la séquence	1/100000 bases
Nombre de gènes identifiés	19438
Nombre de gènes (prédits et identifiés)	22287
Nombre moyen d'exon par gène	10,4
Régions traduites en protéines	34 Mbases (1,1%)
Régions transcrtes non traduites	21 Mbases (0.7%)

Génomique aujourd’hui

Définition initiale: Discipline scientifique dont l’objectif est de cartographier, séquencer et analyser les génomes

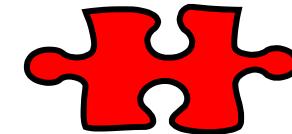
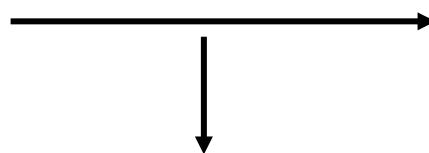
- Le séquençage à haut-débit des génomes est devenu une activité de routine.
- Développement de nouvelles disciplines scientifiques (transcriptomique, protéomique, bioinformatique, pharmacogénétique,...)
- Pour 40-60% des gènes identifiés, la fonction reste inconnue

La post-génomique



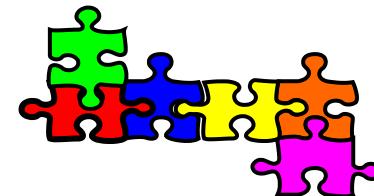
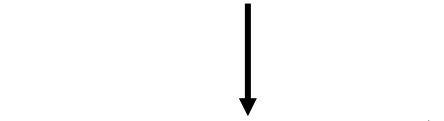
- Analyse globale du génome permettant d'assigner des fonctions aux gènes
- Définir l'organisation et le contrôle des processus génétiques qui s'associent pour faire le fonctionnement physiologique ou pathologique d'une cellule, d'un organe ou d'un organisme

approche
gène candidat

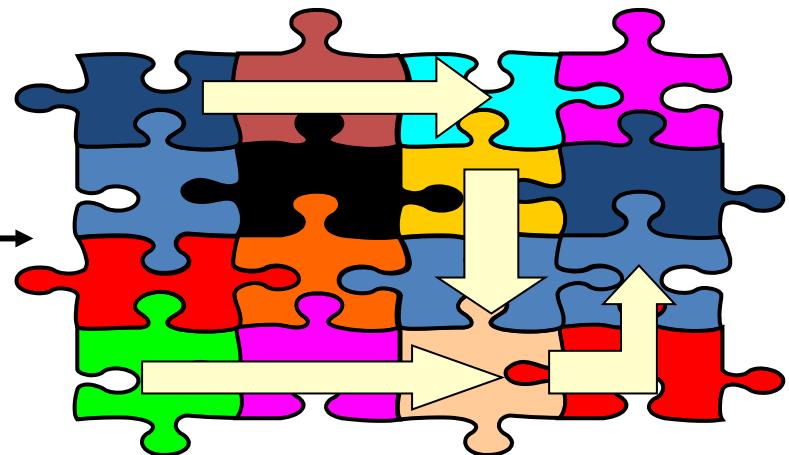


séquençage du génome humain

vision globale du
génome



Post génomique



*Tous les gènes sont des candidats
potentiels !*

Génomique structurale

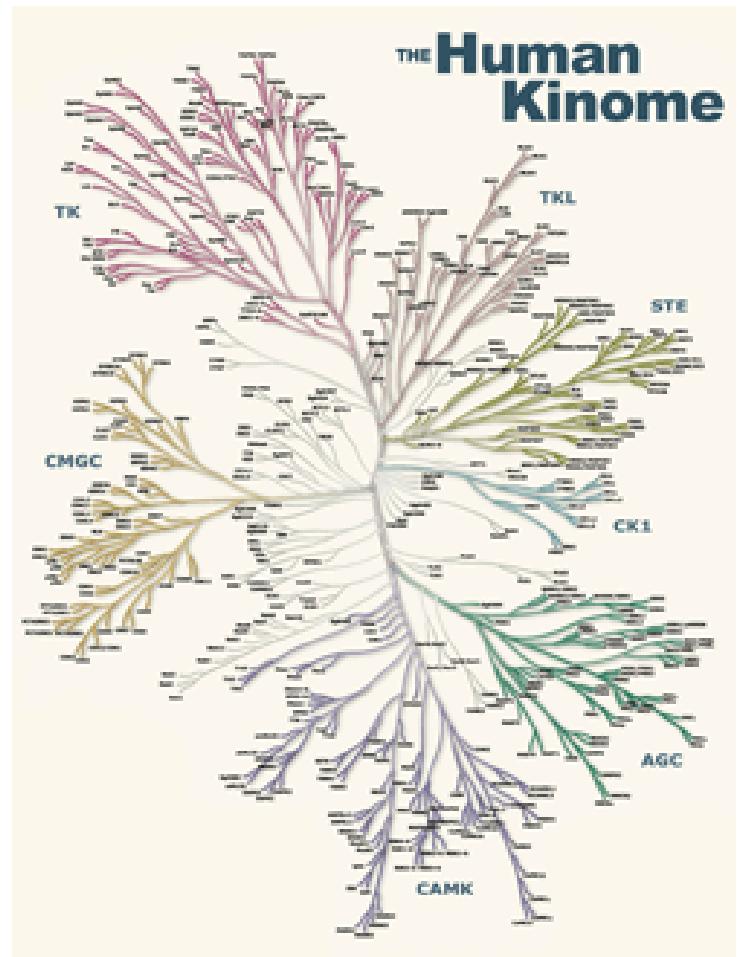
- Objectif: détermination de la structure tridimensionnelle des protéines à partir des données du génome
- Etudier la fonction des milliers de gènes « inconnus » à partir de la relation structure fonction des protéines

« Human kinome »

- = Ensemble des protéines kinase exprimées chez l'Homme

1- Définition d'une séquence consensus de site catalytique à partir des kinase connues

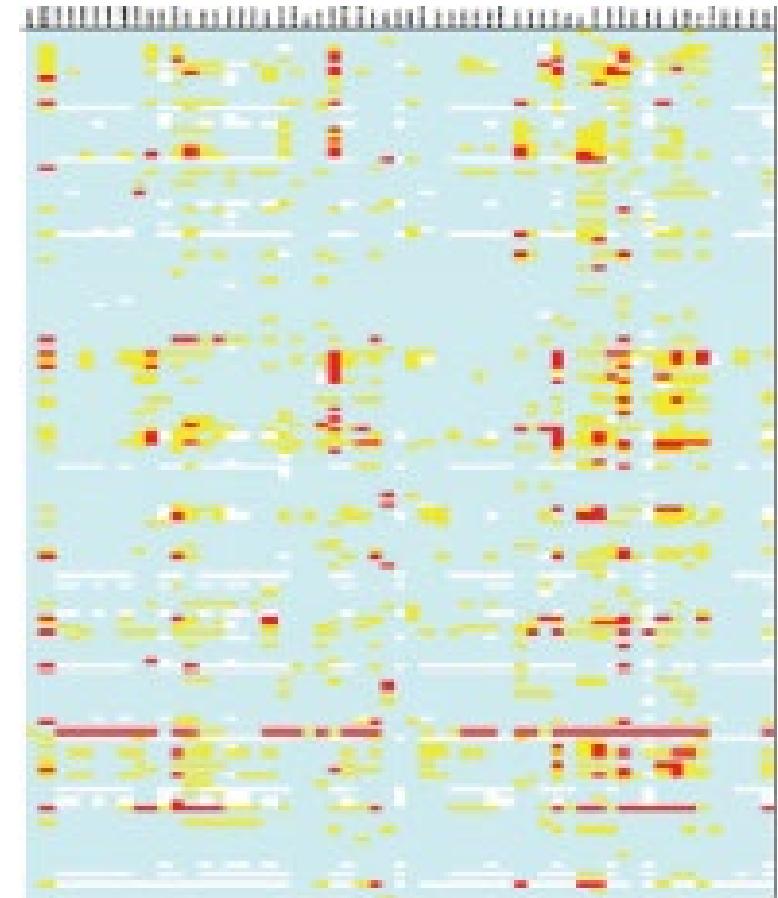
2- Identification de nouveaux gènes codant potentiellement des protéines kinases à partir de leur séquences



« human kinome »

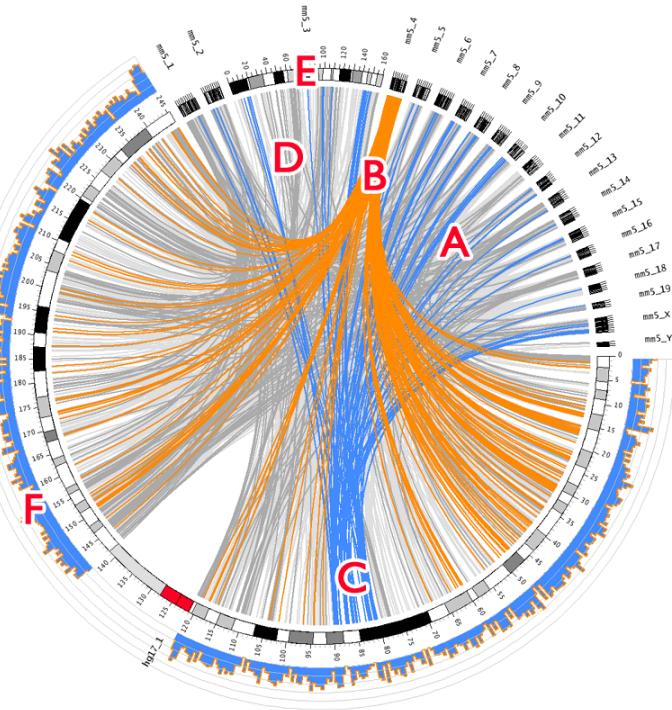
- Production des protéines à partir des séquences
- Test d'inhibiteurs connus des kinases
- Identification de cibles moléculaires

57 Ser/Thr Kinases



Génomique comparative

- Analyse et comparaison de l'ADN de différents organismes
- Conservation de l'ADN entre espèces en particulier au niveau des exons
Similarité Homme - singes 95%
Similarité Homme - souris 85%
- ~99% des gènes humains ont un équivalent (orthologue) chez la souris



Génomique comparative

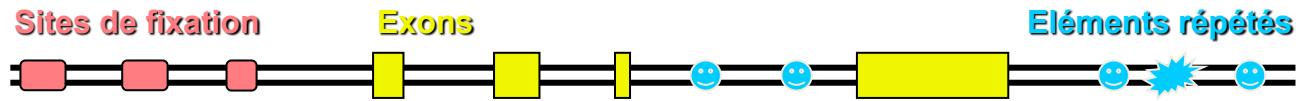
- Identification des séquences conservées entre différents organismes
 - Séquences codantes
- Identification de la fonction d'un gène à partir de la fonction des orthologues

	226	233	236	238	239	242	247
humain	L	I	L	V	V	I	L
souris	L	I	L	V	L	I	S
rat	L	I	L	V	L	I	S
mouton	L	I	L	V	L	I	L
bétail	L	I	L	V	L	I	L
cochon	L	I	L	V	L	I	L

Génomique Fonctionnelle

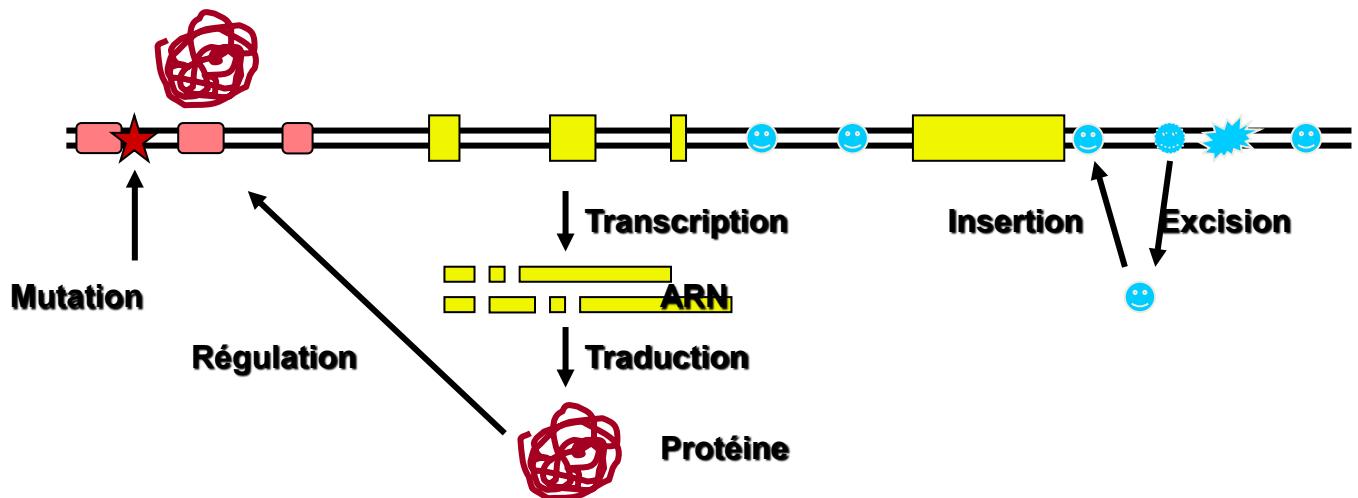
Génomique
=

Description
physique



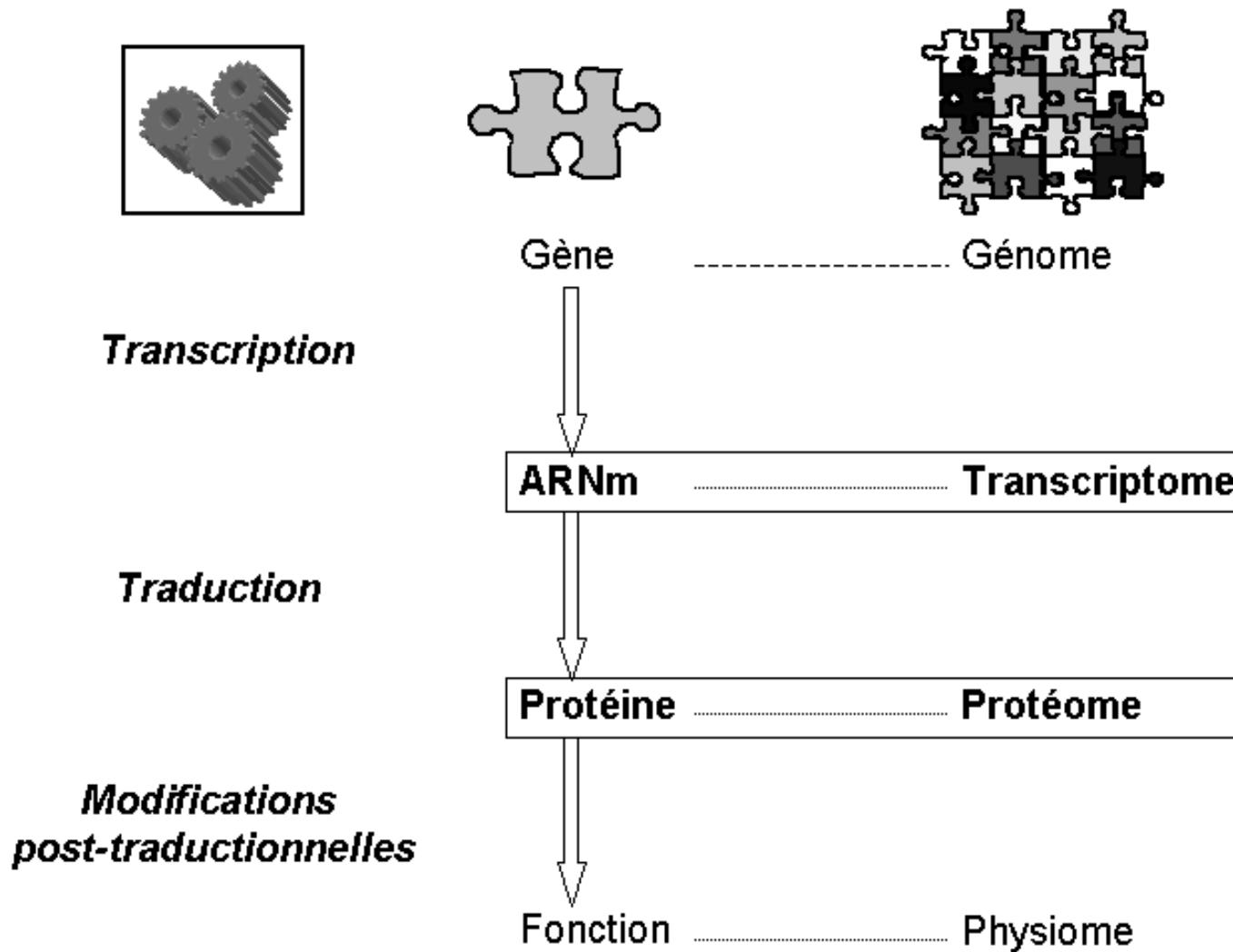
La Génomique
Fonctionnelle
=

fonctionnement
des génomes



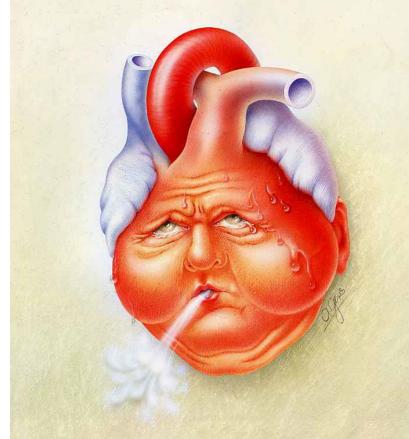
Ex: évaluer le fonctionnement global des gènes dans une maladie

Etude d'expression des gènes

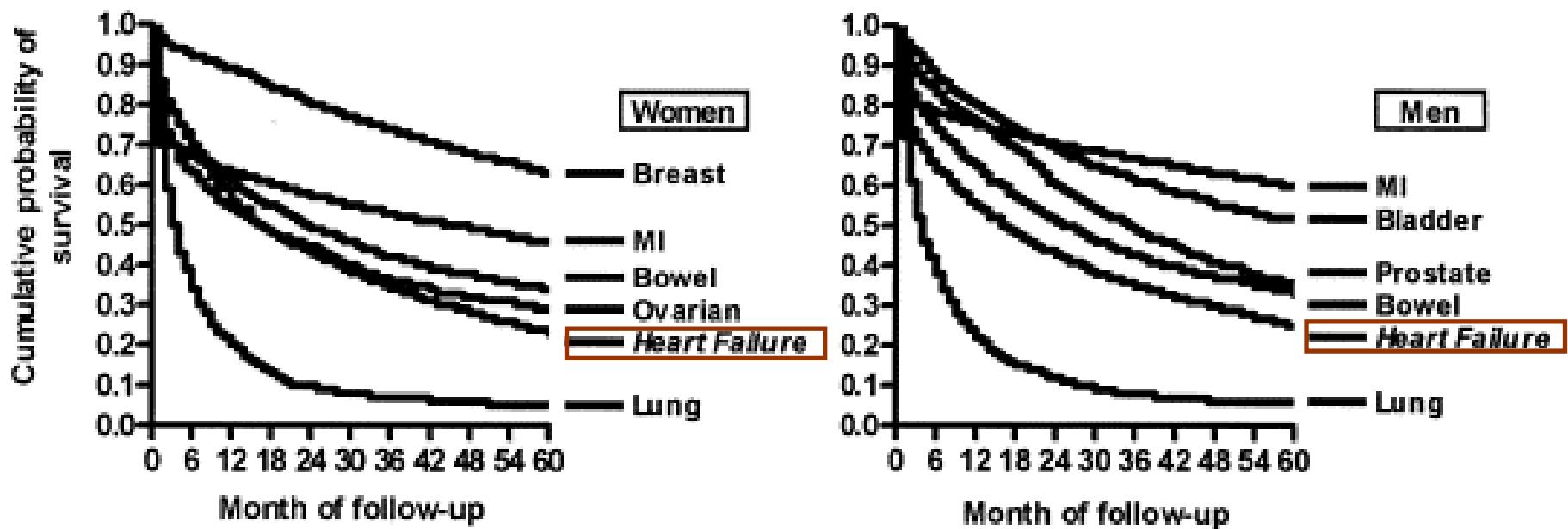


Pathologies cardiaques

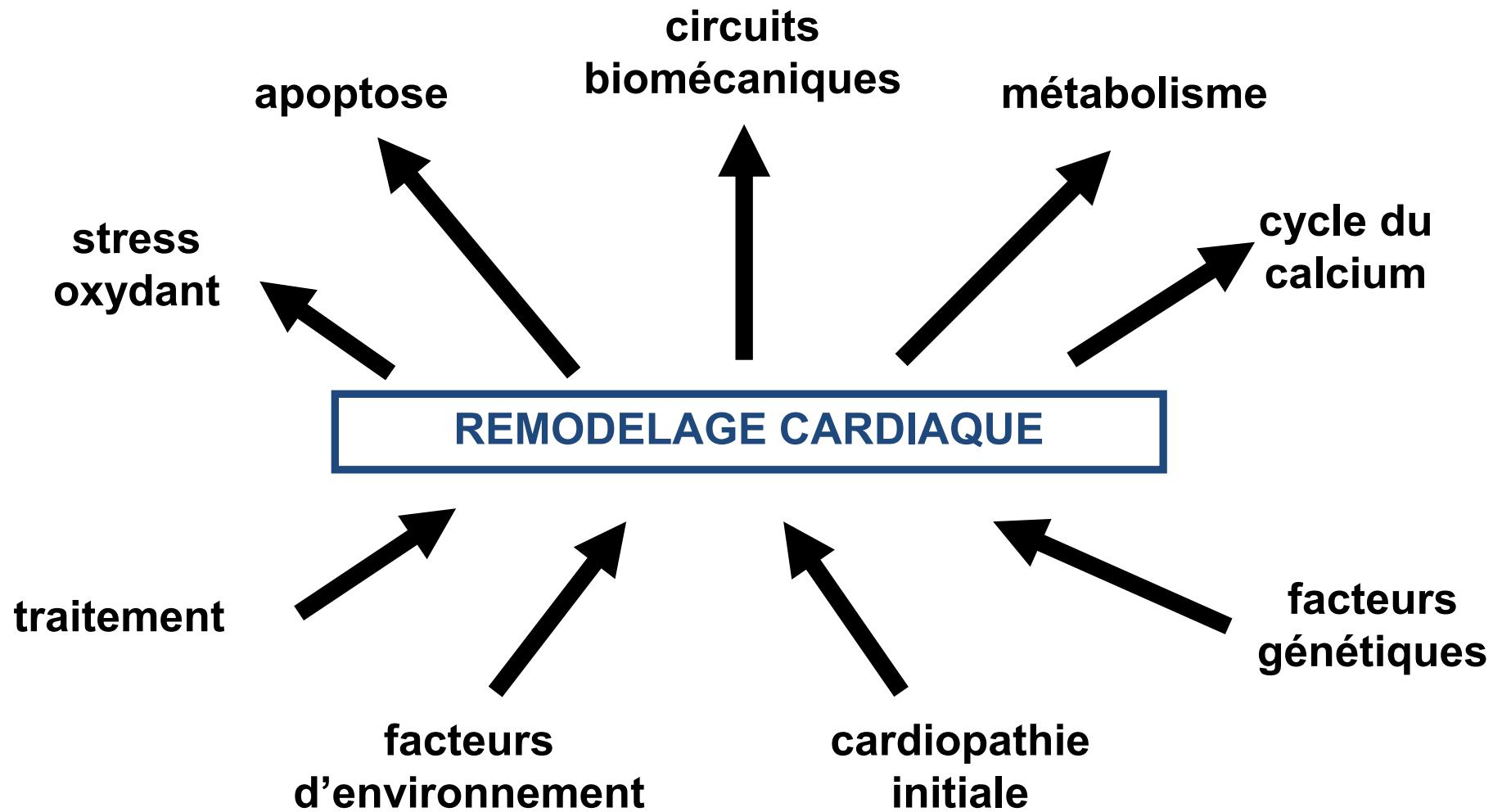
- **30% des causes de décès**
- **Forte prévalence**
- **Renforcement du poids de ces maladies avec le vieillissement de la population**
- **Facteurs de risque environnementaux et génétiques**



Insuffisance cardiaque: plus grave que le cancer?

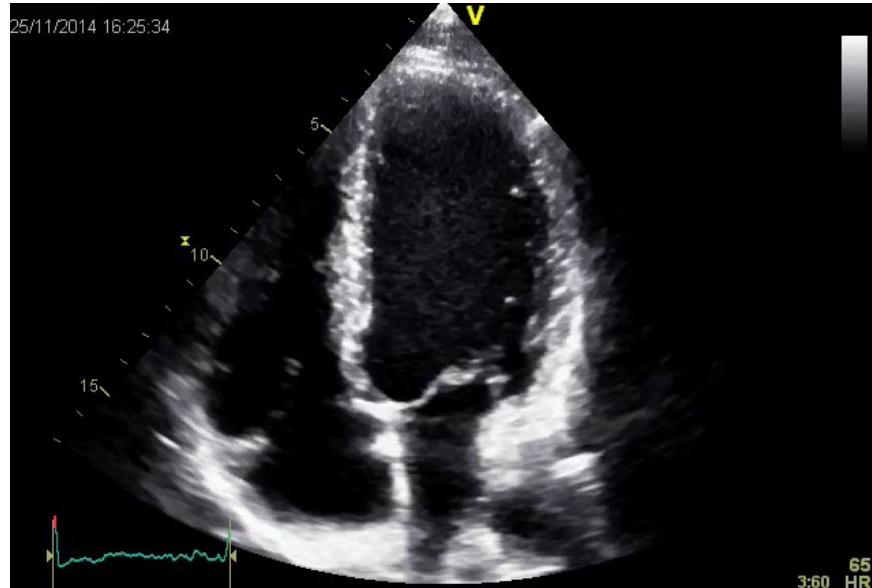


Insuffisance cardiaque: processus de remodelage complexe

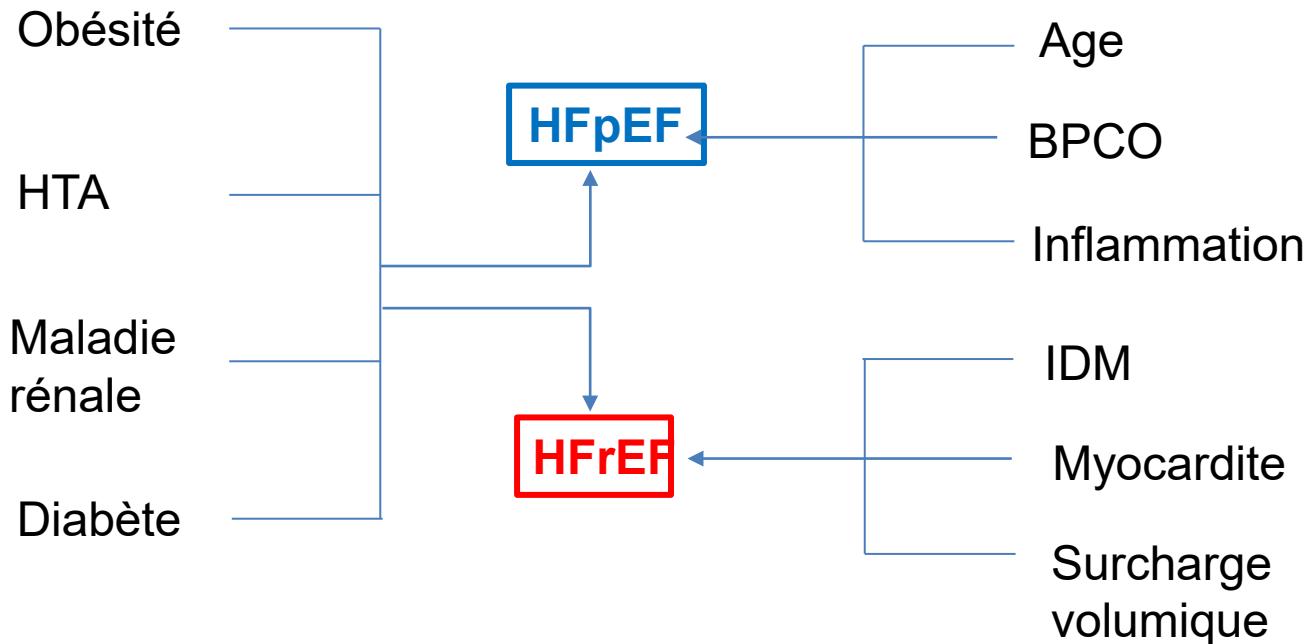


2 présentations d'IC

- Selon la fraction d'éjection ventriculaire gauche (FEVG) :
 - IC à FEVG altérée (ou insuffisance cardiaque systolique)
 - IC à FEVG préservée.



Facteurs de risque et comorbidités



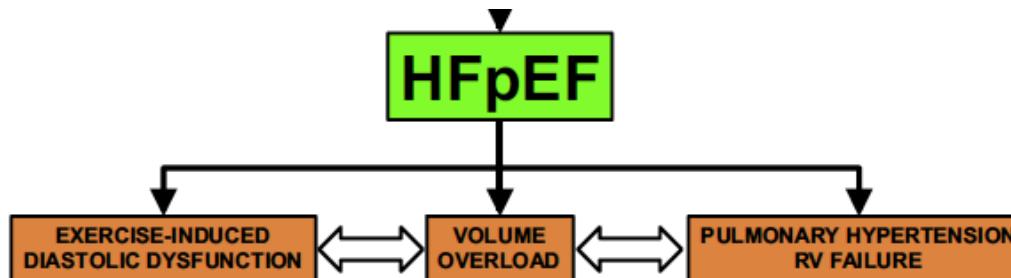
Les co-morbidités
diffèrent chez les pts IC
HFpEF et HFrEF

Pts + HFpEF : plus âgés,
femmes, prévalence↑
comorbidités non
cardiaques

Incidence hospitalisations
plus fréquentes en lien
avec les comorbidités si
HFpEF

La mortalité liée aux
comorbidités est
identique

HFpEF Phenotypes (« phenomapping »)



SAMPLE PATIENTS



- 72-year-old woman
- Long-standing HTN
- NYHA II
- Exercise intolerance
- Minimal fluid retention
- No HF hospitalizations
- LVEF 70%, 2+ LAE
- Grade I DD
- PASP 30 mmHg at rest
- Exercise E/e' = 14

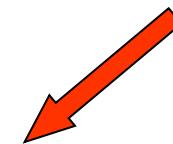
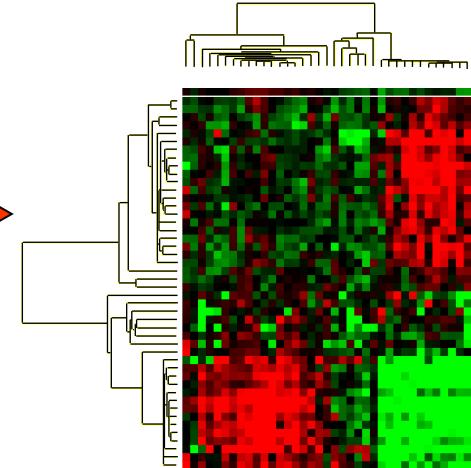
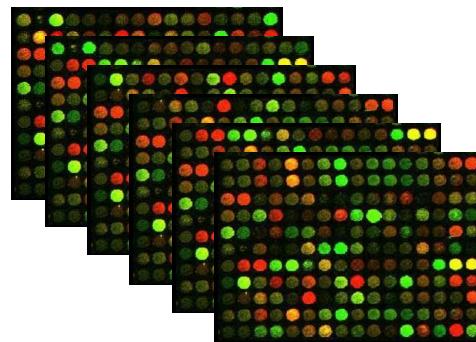


- 66-year-old woman
- HTN, CAD s/p CABG
- NYHA III
- Severe DOE
- 2+ LE edema
- Recent HF hospitalization
- LVEF 50%, 3+ LAE
- Grade III DD
- PASP 45 mmHg at rest
- 2+ MR, 2+ AR



- 59-year-old woman
- HTN, DM2, CKD, obese
- NYHA III
- Severe SOB, DOE
- 3+ edema, ascites
- Frequent HF hospitalizations
- LVEF 65%, 4+ LAE
- Grade II DD
- PASP 60 mmHg at rest
- RVH + RV dysfunction

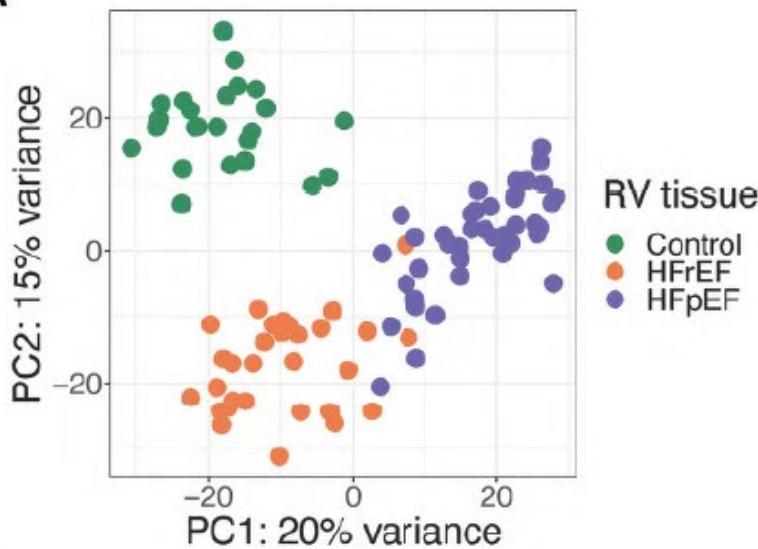
Profils d'expression génique



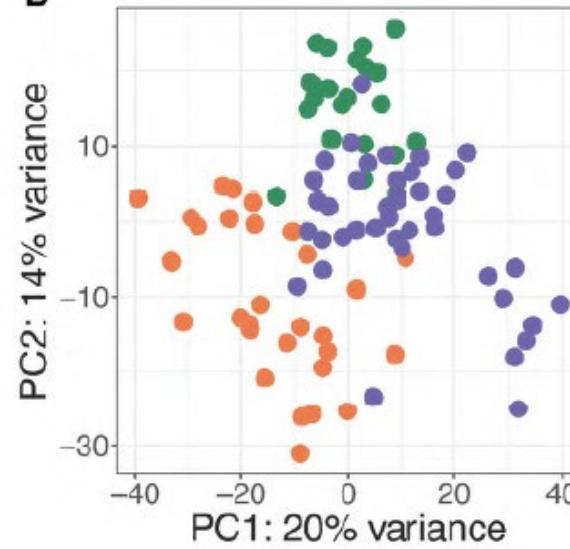
1. Découvertes physiopathologiques
2. Phénotypage moléculaire

Ins. cardiaque: étude du transcriptomique cardiaque

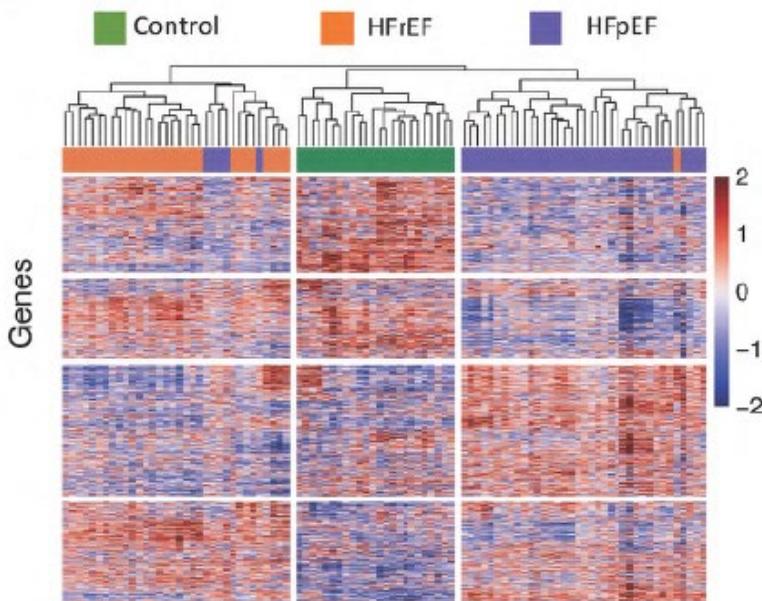
A



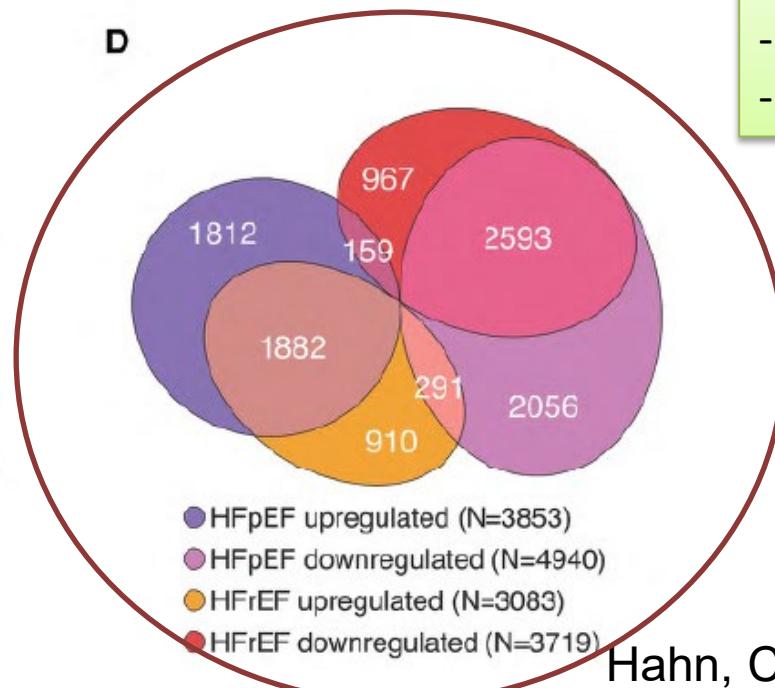
B



C



D

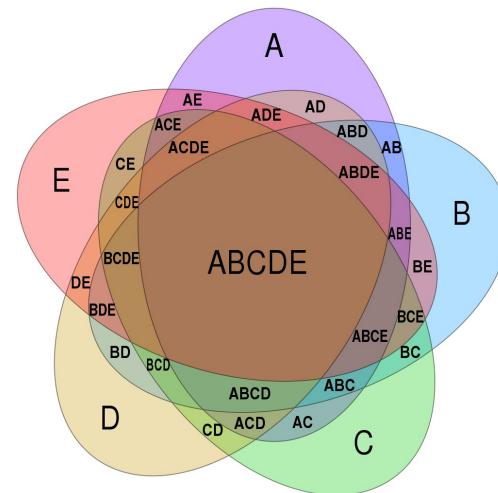
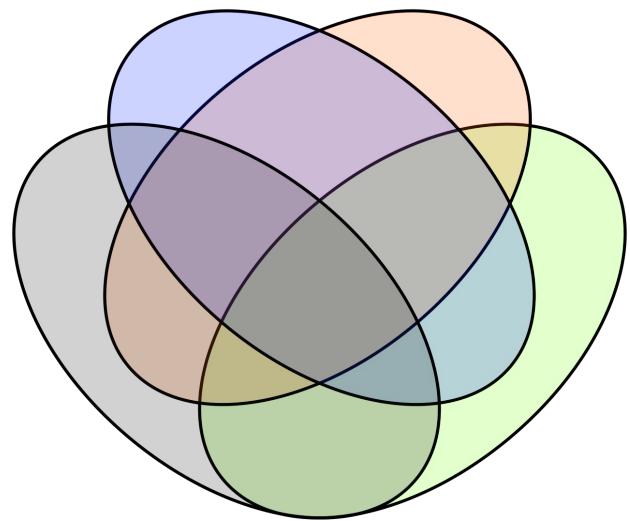
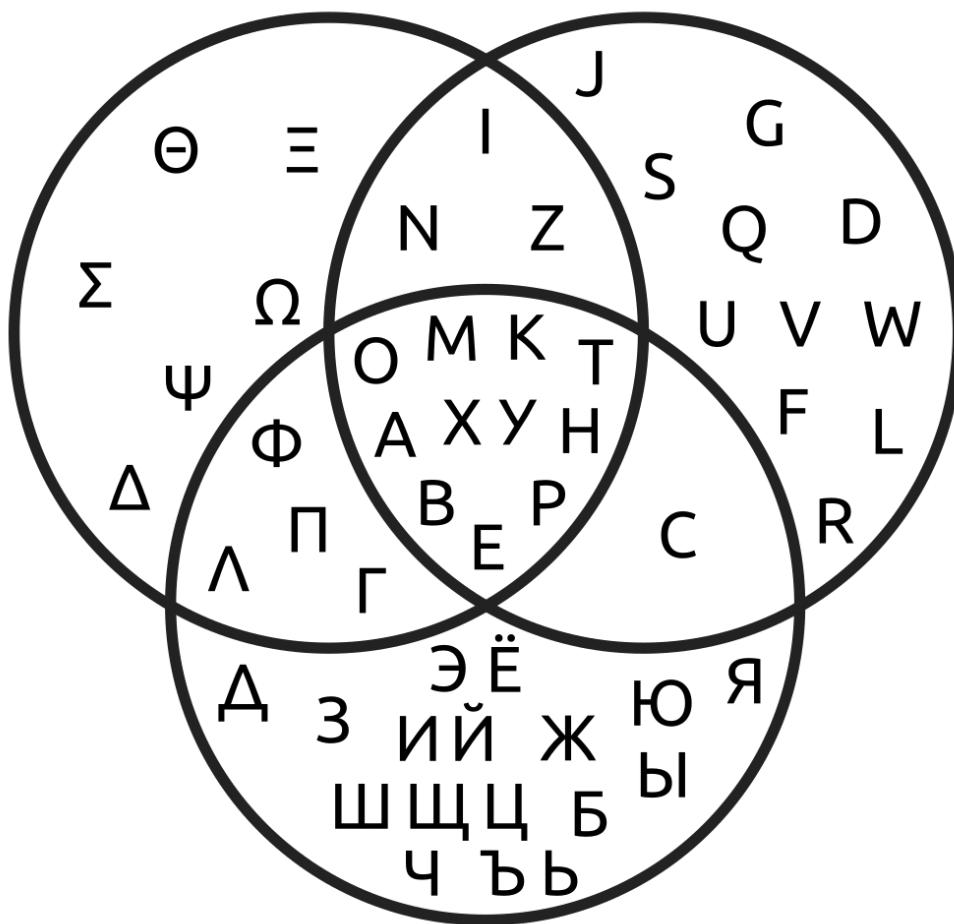


Biopsies de VD:
- Contrôle
- HFpEF
- HFrEF

Interprétation des données

- **Detection de gènes différentiels**
Méthodes statistiques
- **Detection de groupes de gènes co-regulés**
- **Annotation fonctionnelle**

Diagrammes de Venn



Type I and II Errors

		Actual Situation “Truth”	
		H_0 True	H_0 False
Decision	Do Not Reject H_0	Correct Decision $1 - \alpha$	Incorrect Decision Type II Error β
	Reject H_0	Incorrect Decision Type I Error α	Correct Decision $1 - \beta$

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

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?

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

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

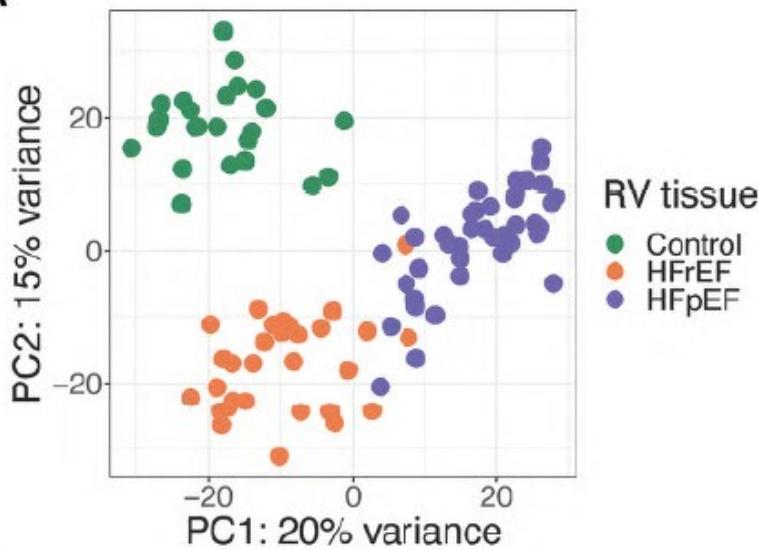
$$P(\text{Not making an error in } m \text{ tests}) = (1 - \alpha)^m$$

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

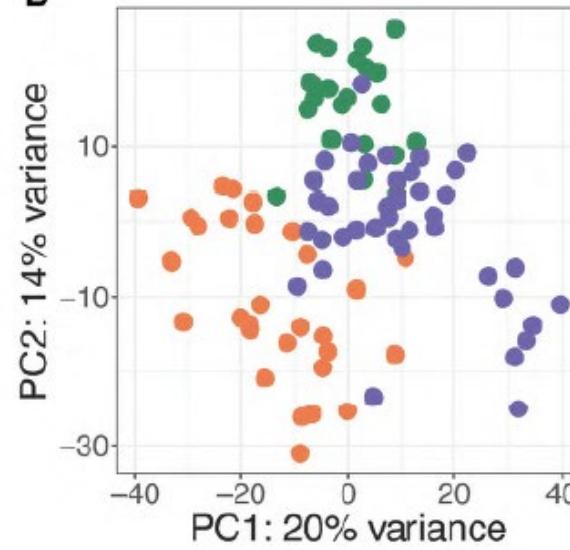
Il faut des méthodes pour contrôler l'erreur de type I

Ins. cardiaque: étude du transcriptomique cardiaque

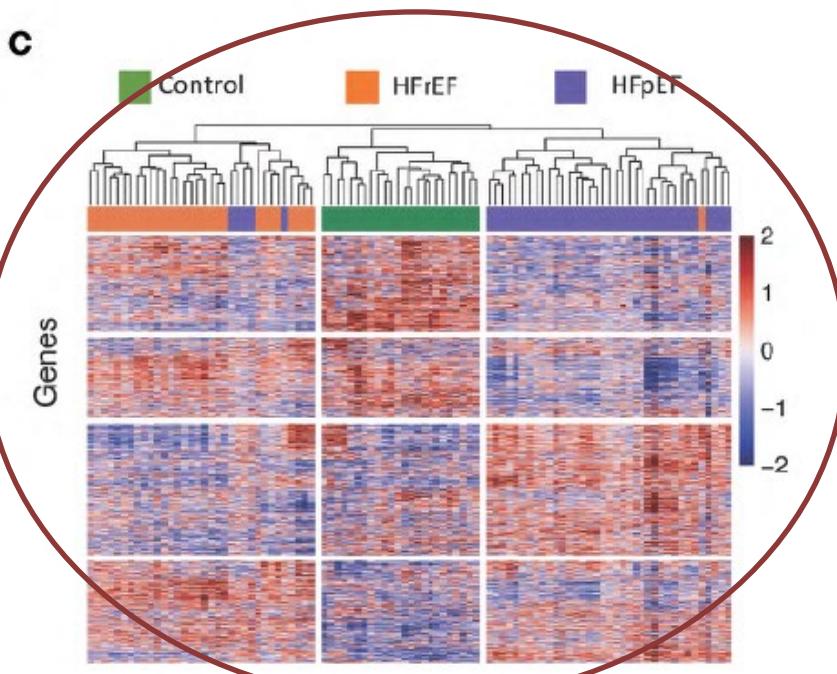
A



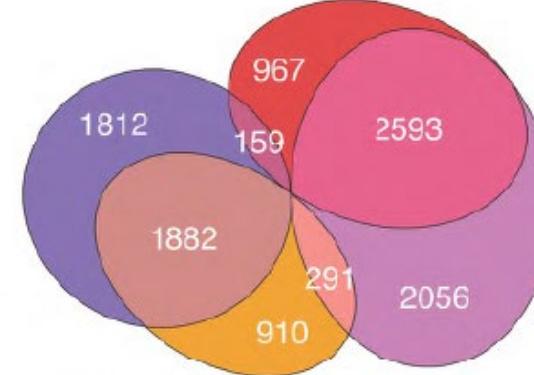
B



C



D



- HFpEF upregulated (N=3853)
- HFpEF downregulated (N=4940)
- HFrEF upregulated (N=3083)
- HFrEF downregulated (N=3719)

Biopsies de VD:

- Contrôle
- HFpEF
- HFrEF

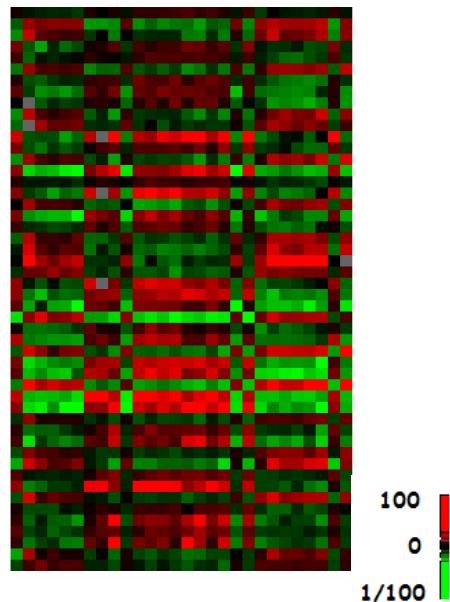
Interprétation des données

- **Detection de gènes différentiels**
- **Detection de groupes de gènes co-regulés**
Méthodes de « clustering »
- **Annotation fonctionnelle**

Clustering - Eisen et al. (1998)

génés

échantillons



Données d'expression



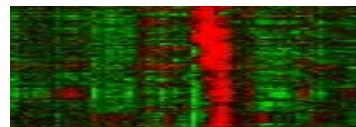
code couleur



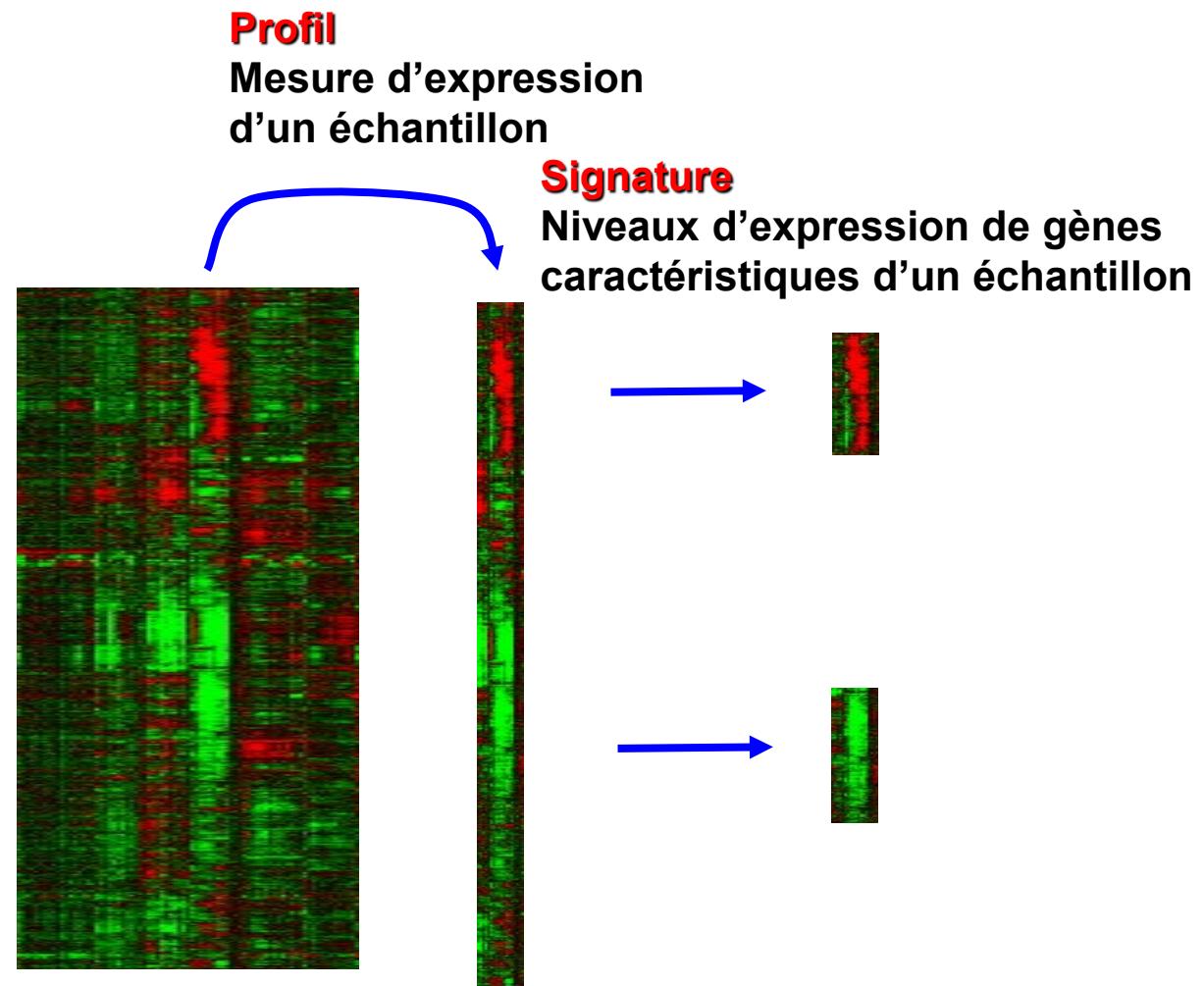
clustering hiérarchique

Définitions

Cluster
Groupe de gènes co-régulés

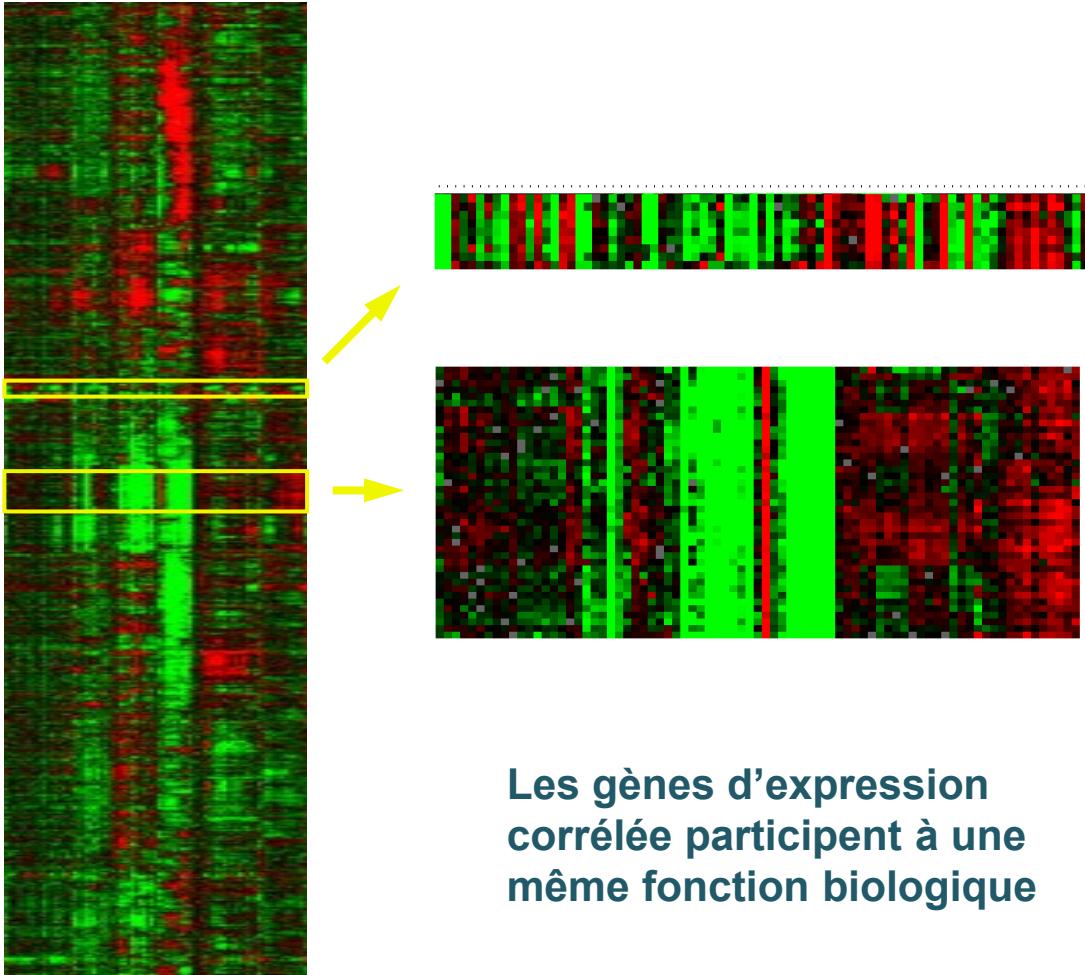


↓
Fonction
↓
Régulation



Correlation = Fonction

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

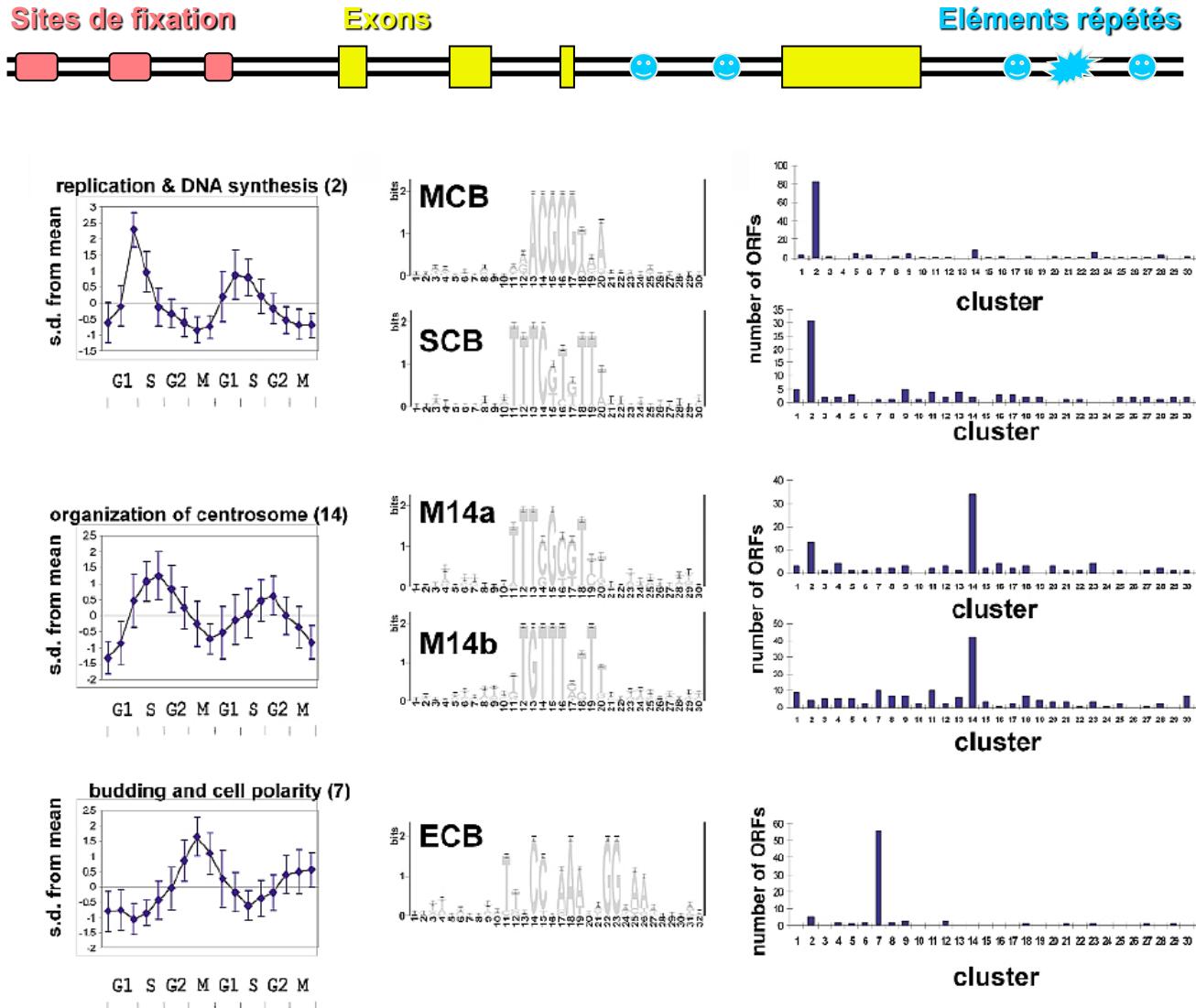


Les gènes d'expression
corrélée participent à une
même fonction biologique

HTB2	CHROMATIN STRUCTURE	HISTONE H2B
HHT1	CHROMATIN STRUCTURE	HISTONE H3
HMF1	CHROMATIN STRUCTURE	HISTONE H4
HTA1	CHROMATIN STRUCTURE	HISTONE H2R
HTB1	CHROMATIN STRUCTURE	HISTONE H2B
HMF2	CHROMATIN STRUCTURE	HISTONE H4
HTA2	CHROMATIN STRUCTURE	HISTONE H2R
HHT2	CHROMATIN STRUCTURE	HISTONE H3
HH01	CHROMATIN STRUCTURE	HISTONE H1
RPS5	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN S5
RPS4A	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN S4B
RPL26B	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN L26B
RPS7A	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN S7A
RPS24B	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN S24B
RPS21A	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN S21B
RPL14A	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN L14B
RPL38	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN L38
RPL24A	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN L24B
RPS1A	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN S1B
RPS1B	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN S1B
RPL24B	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN L24B
RPS23A	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN S23B
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 S19B
RPL8B	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN L8B
RPS19B	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN S19B
RPS26B	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN S26B
RPL11A	PROTEIN SYNTHESIS	RIBOSOMAL PROTEIN L11B
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 L13B

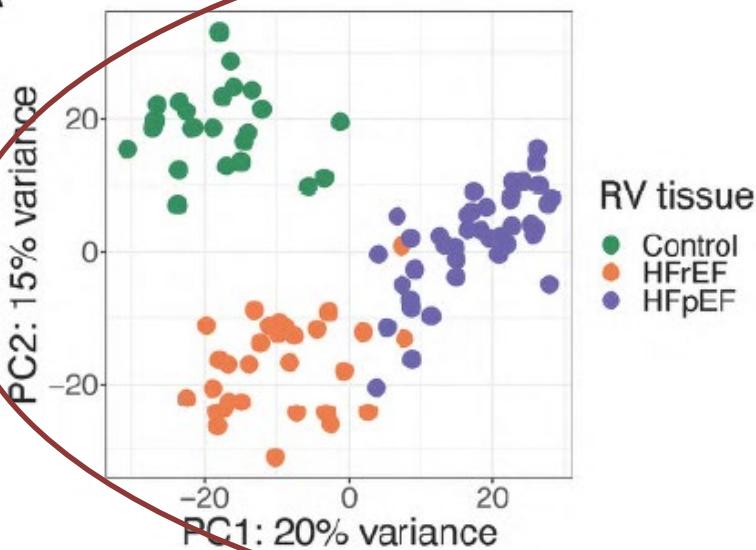
Cluster = Régulation

Tavazoie et al. (1998) Nature Genetics

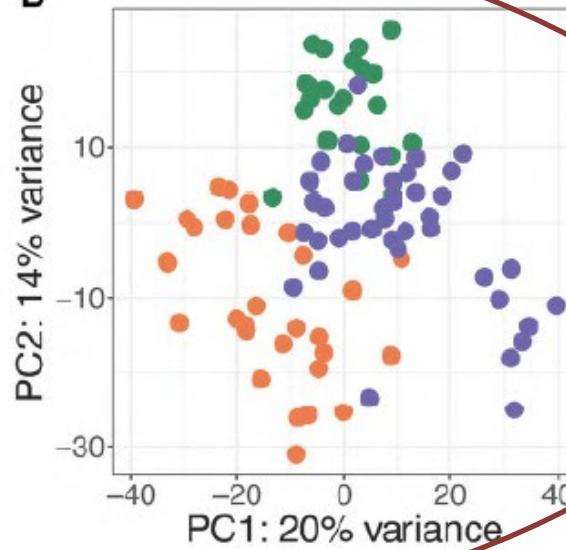


Ins. cardiaque: étude du transcriptomique cardiaque

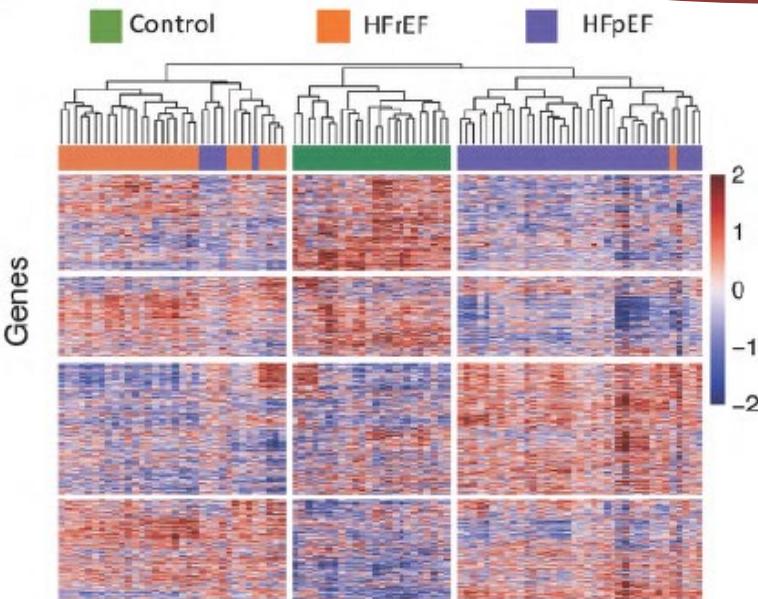
A



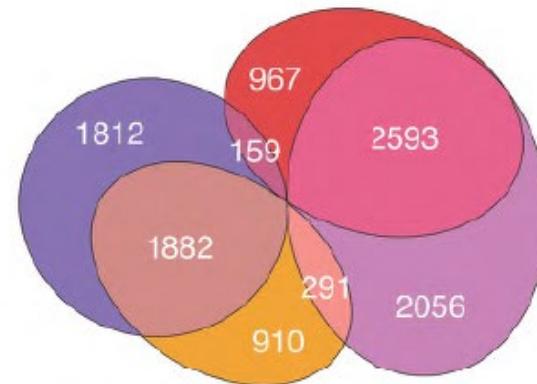
B



C



D



- HFpEF upregulated (N=3853)
- HFpEF downregulated (N=4940)
- HFrEF upregulated (N=3083)
- HFrEF downregulated (N=3719)

Biopsies de VD:

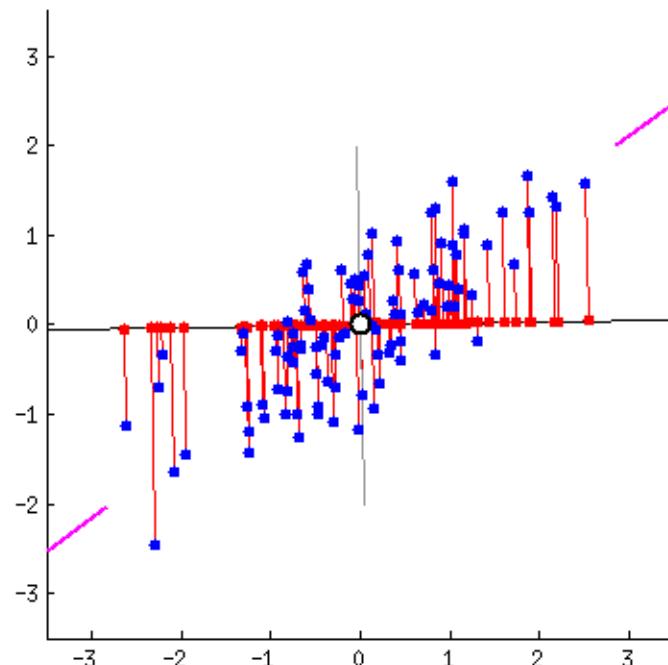
- Contrôle
- HFpEF
- HFrEF

Interprétation des données

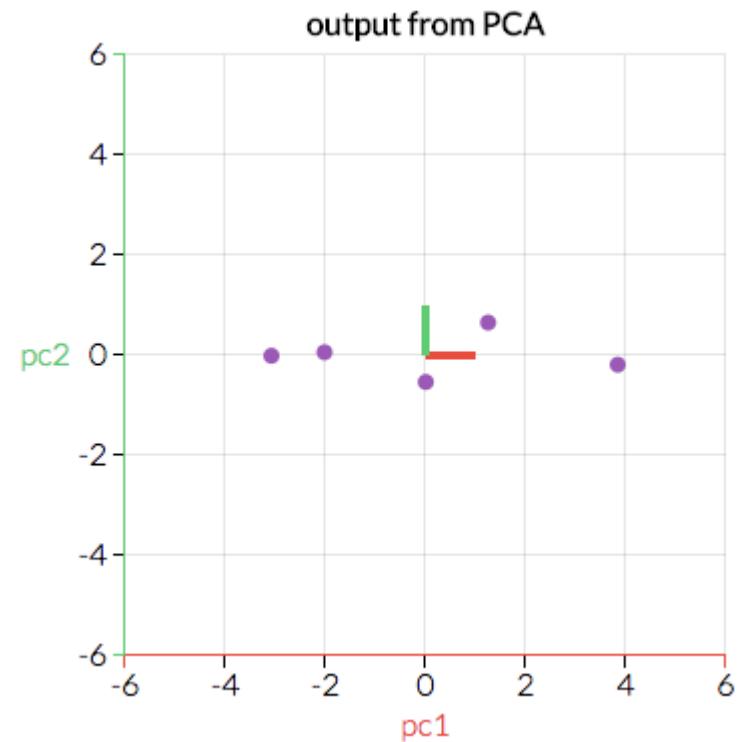
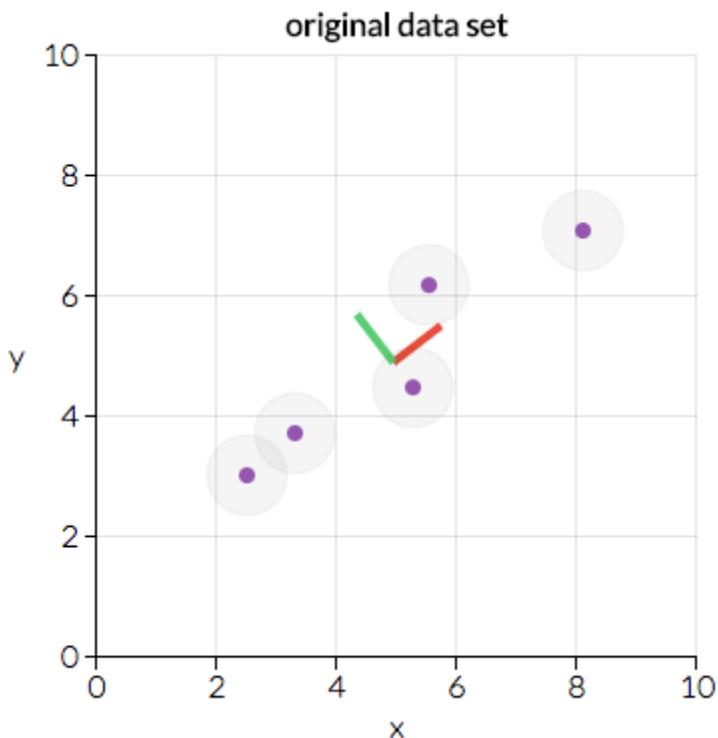
- **Detection de gènes différentiels**
- **Detection de groupes de gènes co-regulés**
Réduction de dimension
- **Annotation fonctionnelle**

Analyse en composante principale

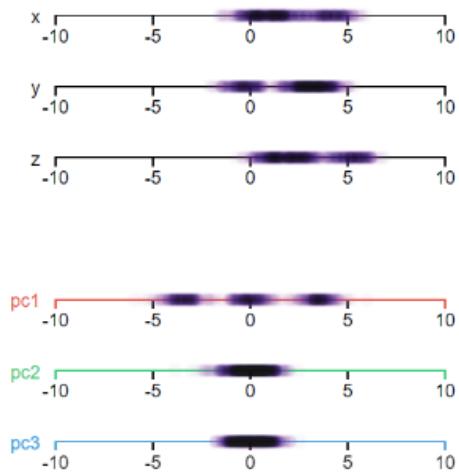
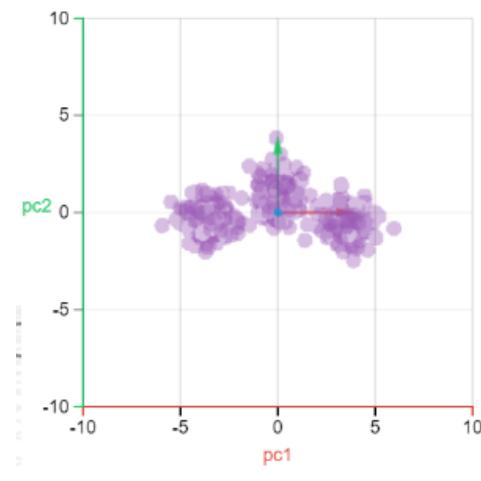
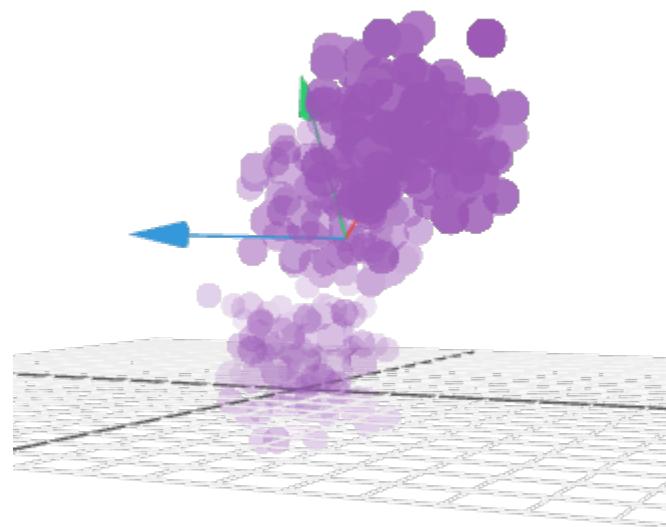
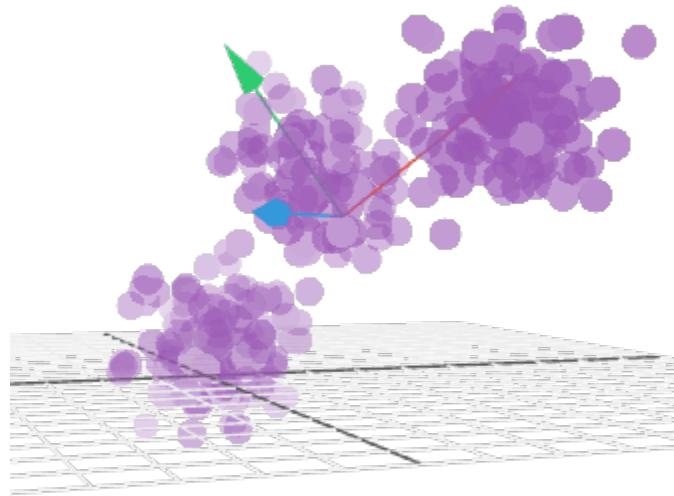
- Une méthode de réduction du nombre de variables utiles en préservant au maximum la quantité d'information
- Création séquentielle de nouvelles variables (« pseudogènes ») qui capturent à chaque fois le plus possible d'information discriminante
- Plus de simplicité/lisibilité en limitant la perte de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



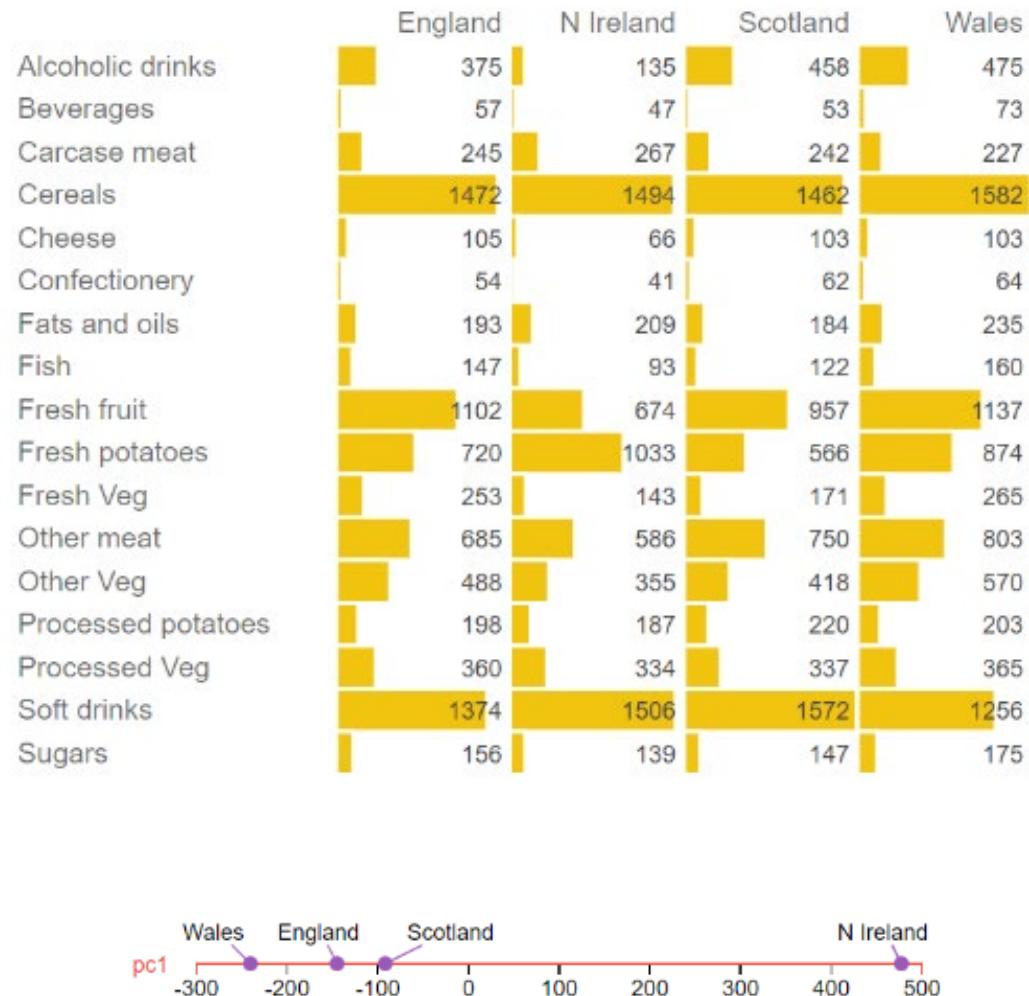
5 échantillons et 2 variables



Nombreux échantillons et 3 variables

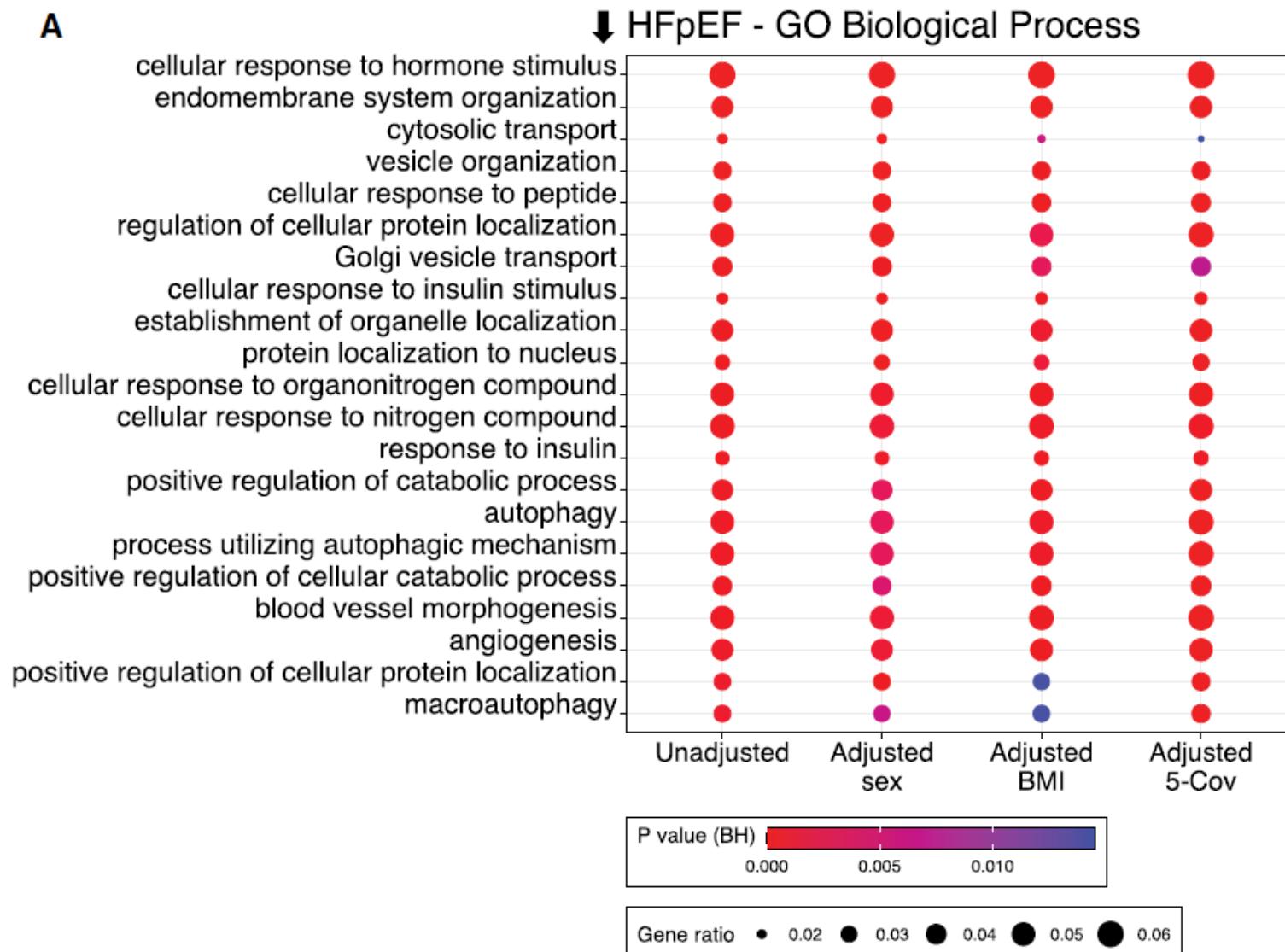


4 groupes et 17 variables



A quelles fonctions biologiques participent les gènes sous-exprimés

A

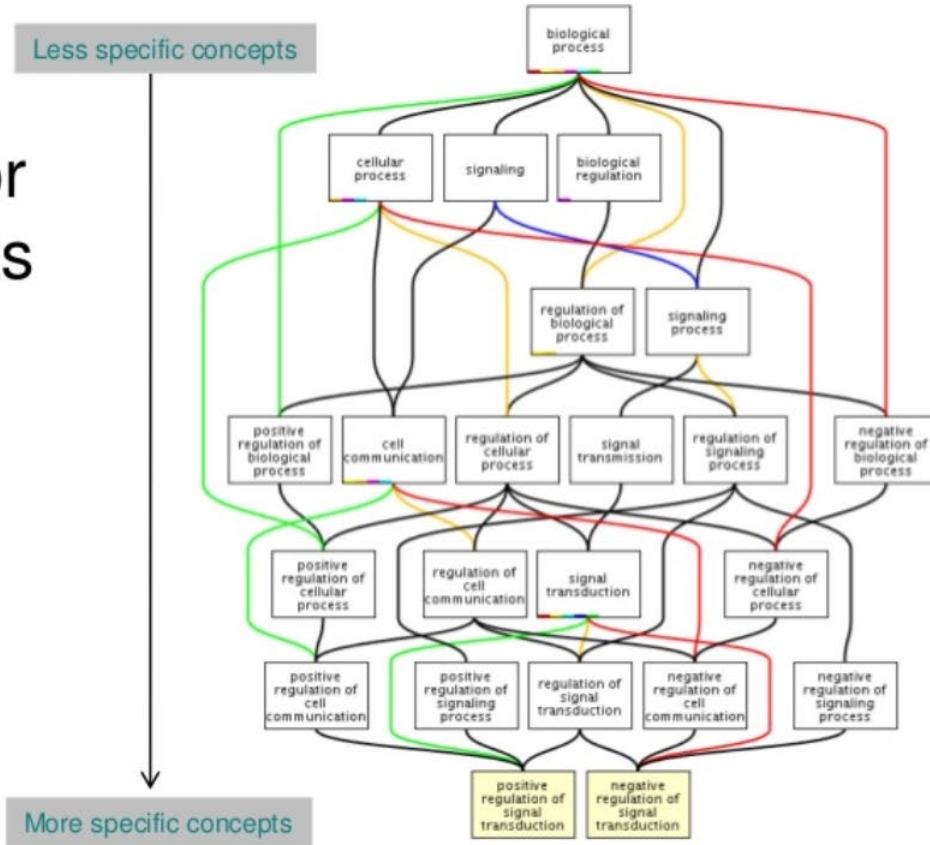


Interprétation des données

- Détection de gènes différentiels
- Détection de groupes de gènes co-regulés
- Annotation fonctionnelle
 - Banques d'information (www)
 - Gene Ontology, Reactome,...

The Gene Ontology

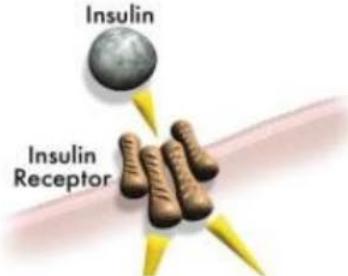
- A way to capture biological knowledge for individual gene products in a written and computable form
- A set of concepts and their relationships to each other arranged as a hierarchy



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

1. Molecular Function

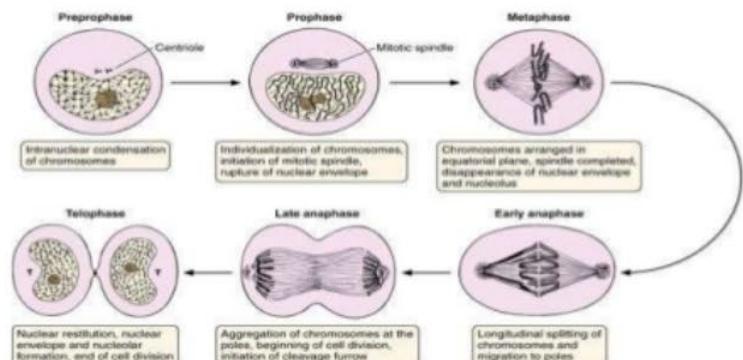
An elemental activity or task or job



- protein kinase activity
- insulin receptor activity

2. Biological Process

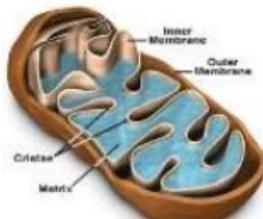
A commonly recognized series of events



- cell division

3. Cellular Component

Where a gene product is located

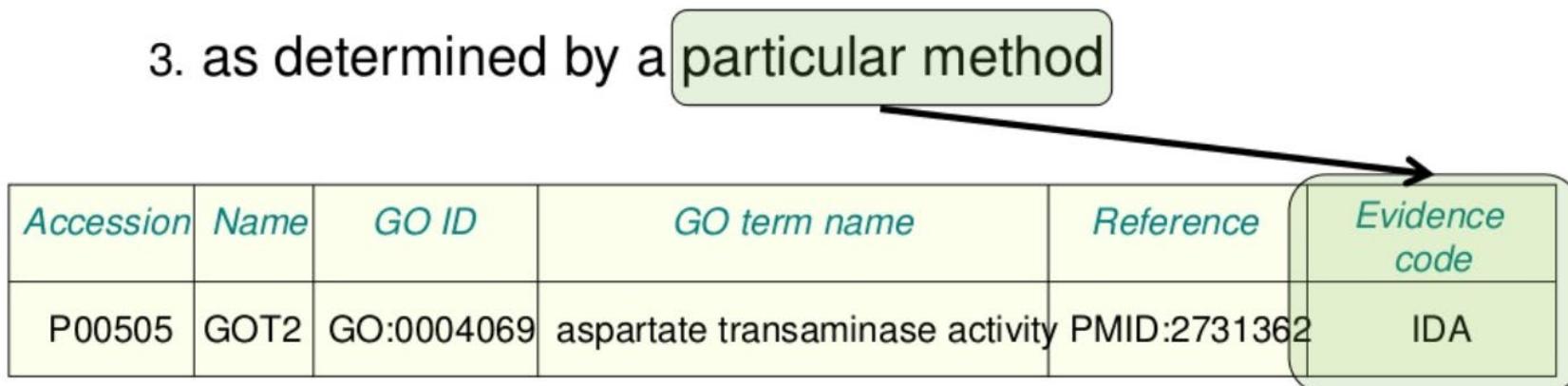


- mitochondrion
- mitochondrial matrix
- mitochondrial inner membrane

A GO annotation is ...

...a statement that a gene product;

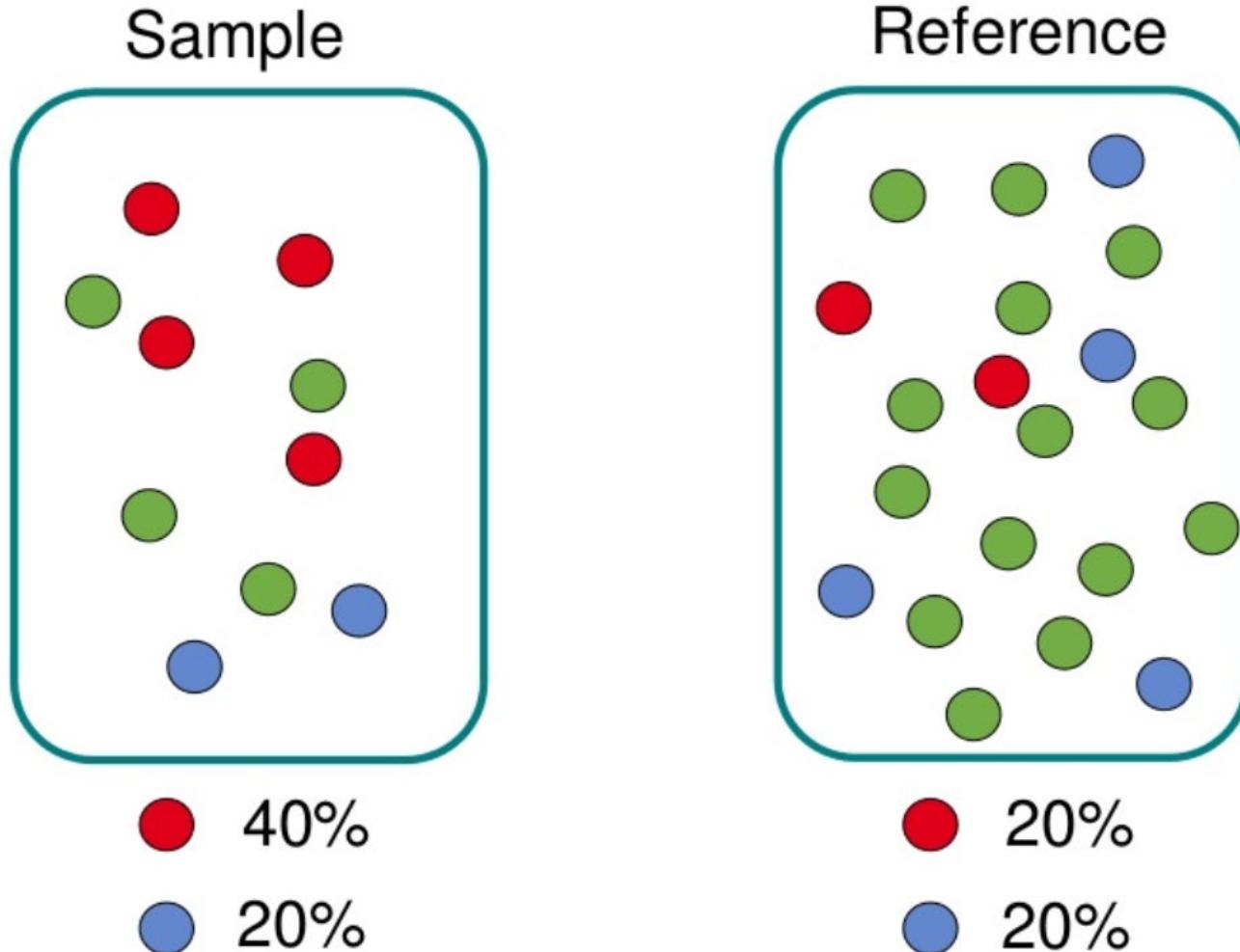
1. has a particular molecular function
or is involved in a particular biological process
or is located within a certain cellular component
2. as described in a particular reference
3. as determined by a **particular method**



The diagram illustrates a GO annotation table. The table has six columns: Accession, Name, GO ID, GO term name, Reference, and Evidence code. The 'Evidence code' column is highlighted with a green rounded rectangle and contains the code 'IDA'. An arrow points from the 'particular method' in the list above to this 'Evidence code' column.

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

Enrichment analysis

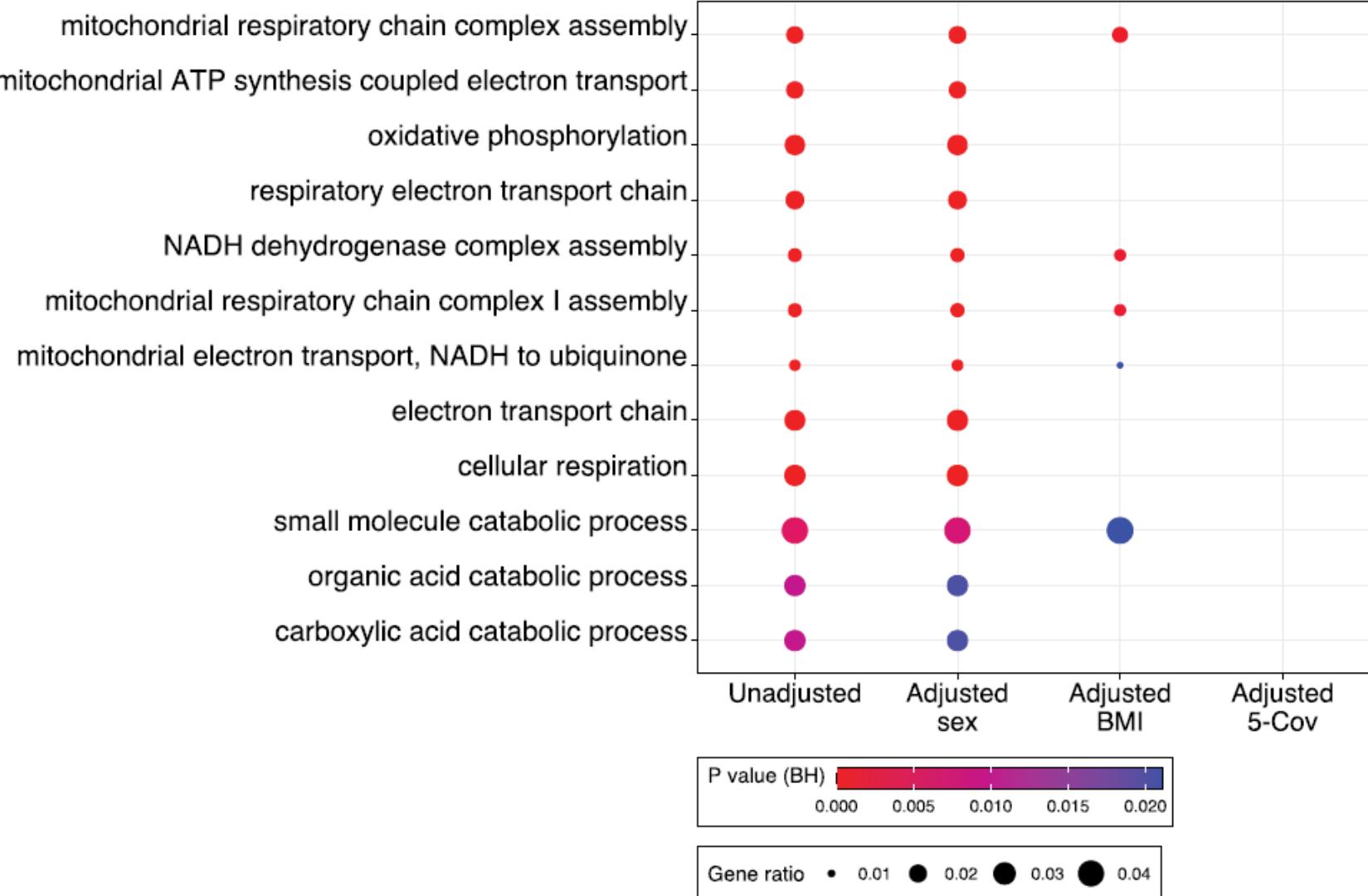


=> The sample is over-enriched for ●

A quelles fonctions biologiques participent les gènes sur-exprimés

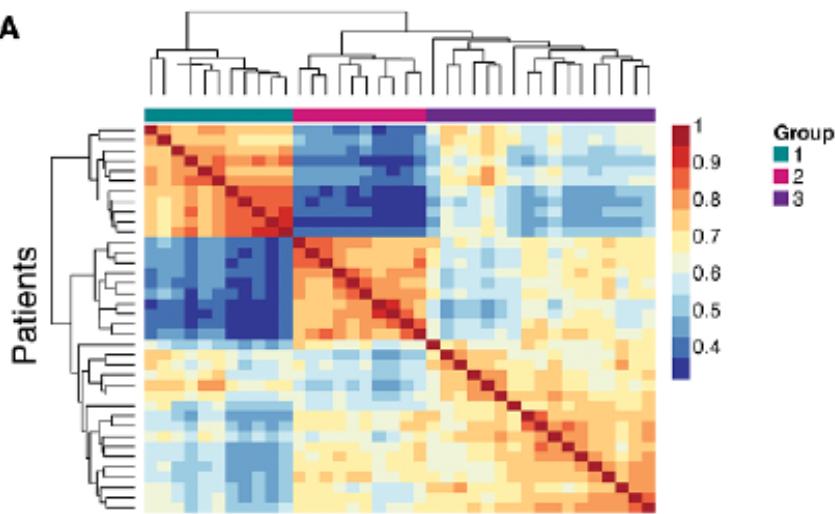
B

↑ HFpEF - GO Biological Process

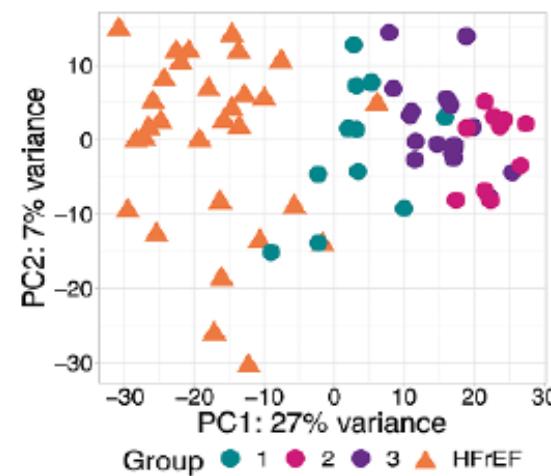


Sous-groupes de patients HFrEF

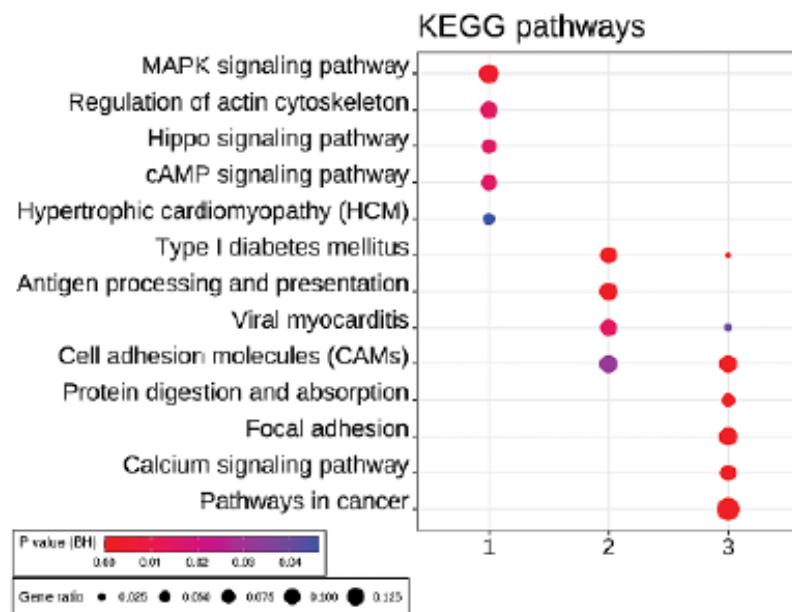
A



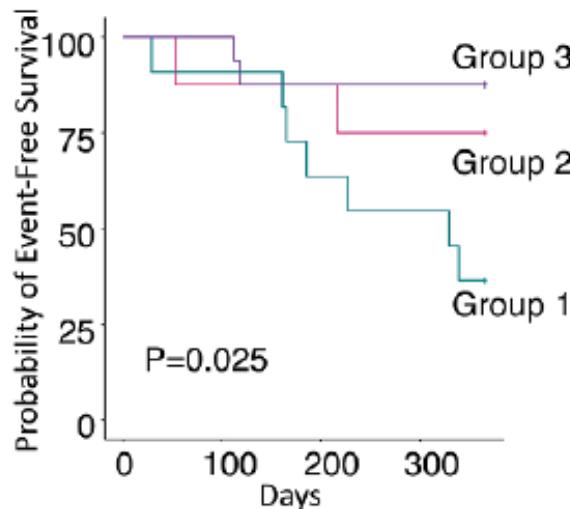
B



C

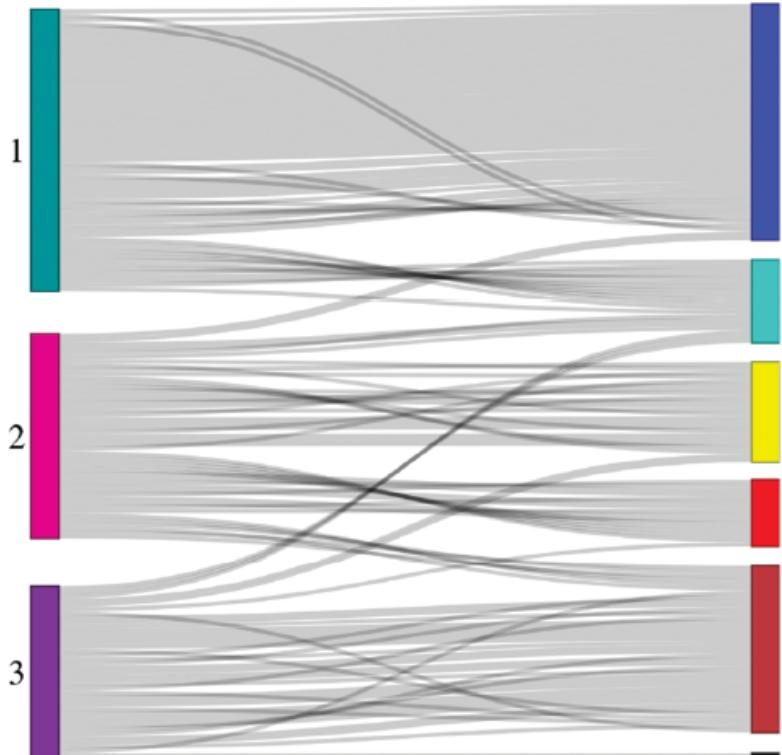


D



Sous-groupes de gènes et sous-groupes de patients

B

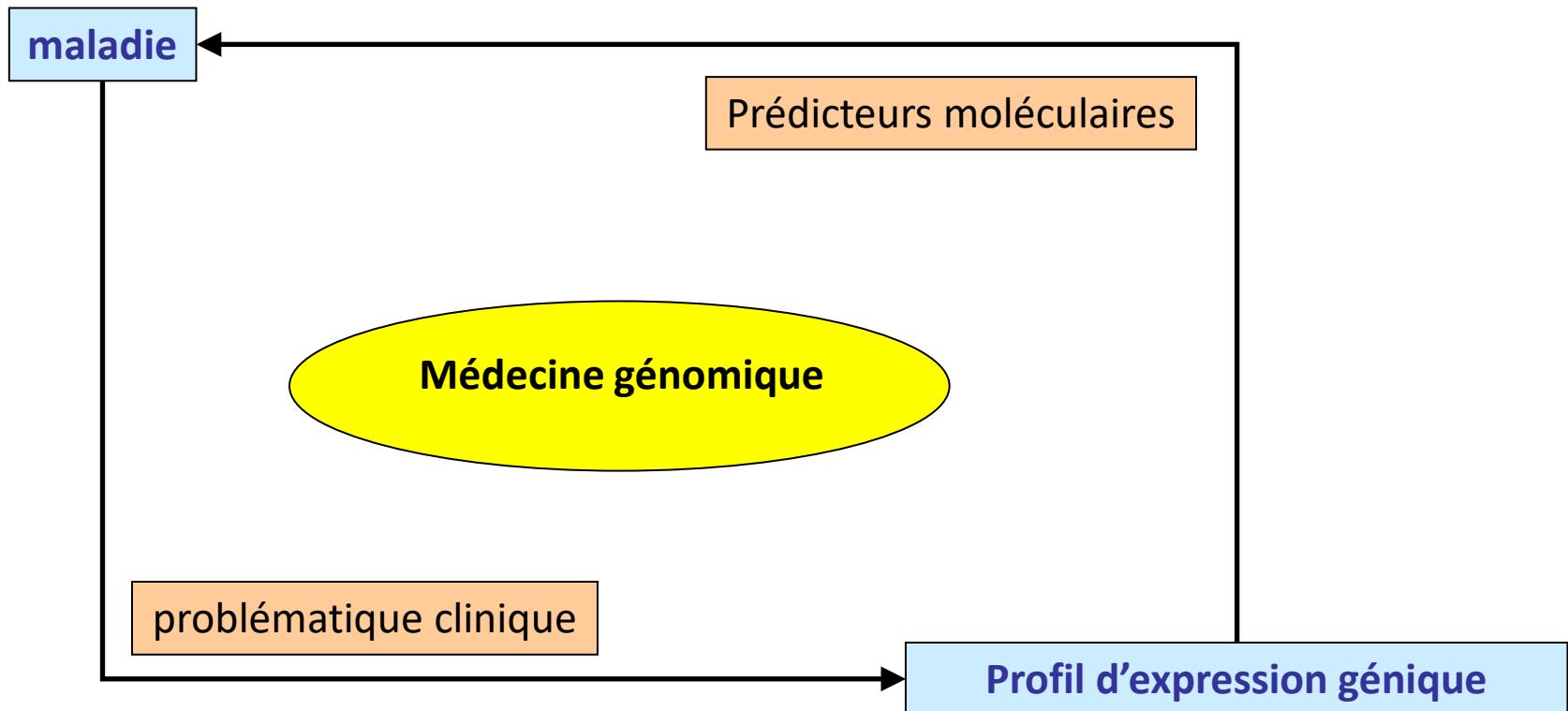


C

	Clinical Characteristics	Gene Ontology Biological Processes
1	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 signalling cascade
2	Turquoise: Lower BMI, higher LV wall thickness and myocyte hypertrophy	Histone modification; Golgi organization; Regulation of gene expression, epigenetic; Negative regulation of microtubule depolymerization
3	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 signalling
	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

Perspectives



Merci de votre attention!

Questions?

Guillaume LAMIRAUT

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



UNIVERSITÉ DE NANTES
CENTRE HOSPITALIER
UNIVERSITAIRE DE NANTES