



➤ Design d'une étude métabolomique

Binta Diémé, IR biologiste en analyse de données, PFEM

# > Plan

- I. Généralités: historique, définitions
- II. Les différents types d'études métabolomiques
- III. Les outils analytiques
- IV. Le prétraitement des données analytiques
- V. Les analyses statistiques
- VI. Interprétation des résultats

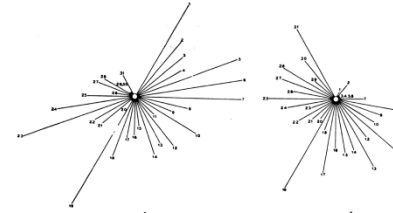


# ➤ Généralités: historique

Concept de  
« profil  
métabolique »  
1940

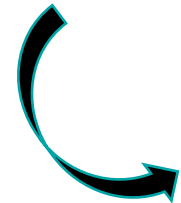


Roger Williams

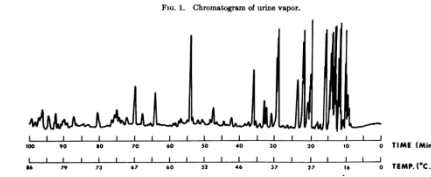


Profil métabolique urinaire sur  
papier chromatographique

Utilisation de « profil  
métabolique »  
(GC/MS)  
1971



Linus Pauling

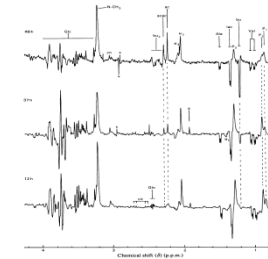


Profil métabolique urinaire  
par GC/MS

Utilisation de « profil  
métabolique » (RMN)  
1984



Jeremy K Nicholson



Profil métabolique du  
sérum par  $^1\text{H}$ -RMN

Utilisation du terme  
« métabolome »  
1988



Stephen G. Oliver

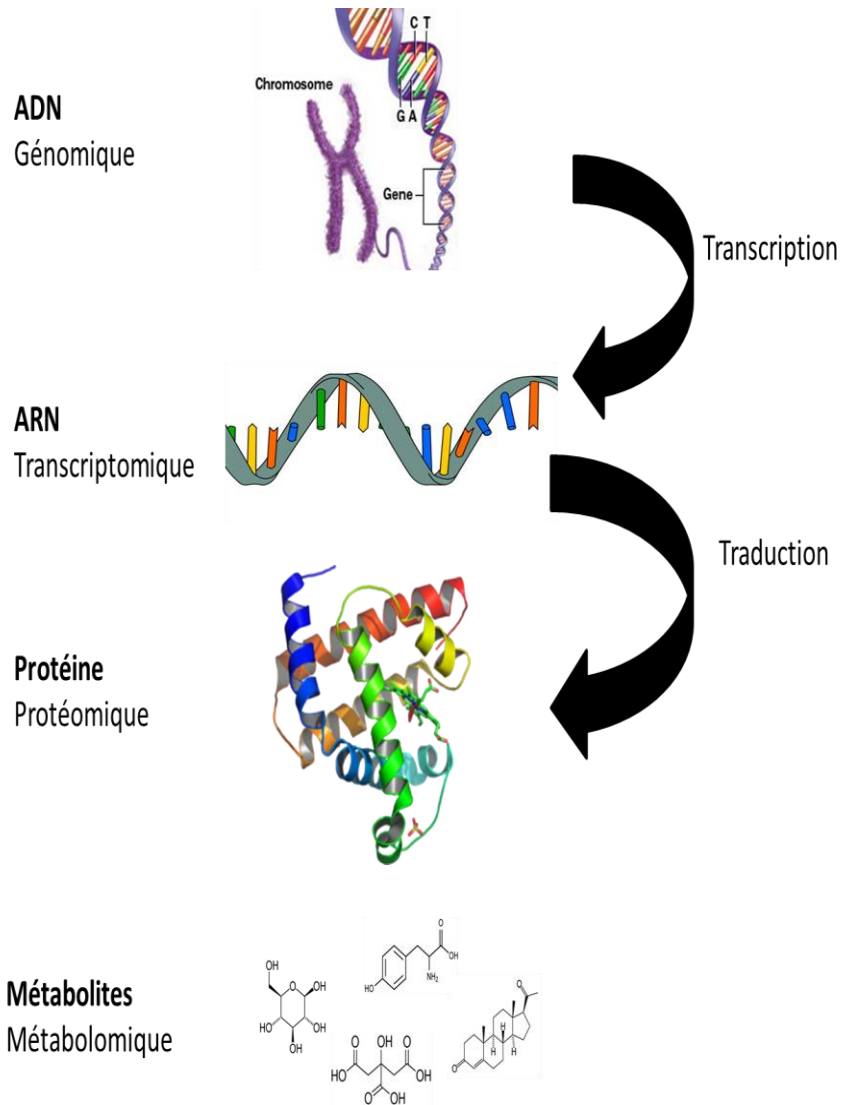


**INRAE**

Design d'une étude métabolomique

04-06-2020 / Production at analyse de données métabolomiques / Binta Diémé

# ➤ Généralités: définition



## Métabolome

Ensemble des petites molécules (< 1500 Da) présentes dans un échantillon biologique.

## Métabolomique

Définition (Fiehn *et al.* Nat. Biotechnol., 2000) :

→ Identification et quantification exhaustive et non sélective de tous les métabolites d'un système biologique.

INRAE

Design d'une étude métabolomique

04-06-2020 / Production et analyse de données métabolomiques / Binta Diémé

# ➤ Généralités: domaines d'application de la métabolomique

## Santé



- Diagnostique (précoce, prédictif)
- Physiopathologie (cancer, maladies cardio-vasculaires, diabète/obésité)
- Maladies métaboliques, maladies du métabolisme

## Pharmacologie



- Développement/validation de nouveaux médicaments, AMM
- Pharmacologie prédictive
- Suivi individualisé
- Pharmacocinétique
- Biomarqueurs de toxicité

## Agro-alimentaire



- Nutrition (prébiotiques, probiotiques)
- Suivi nutritionnel personnalisé
- Sécurité alimentaire, fraudes
- Traçabilité, OGM

## Microbiologie



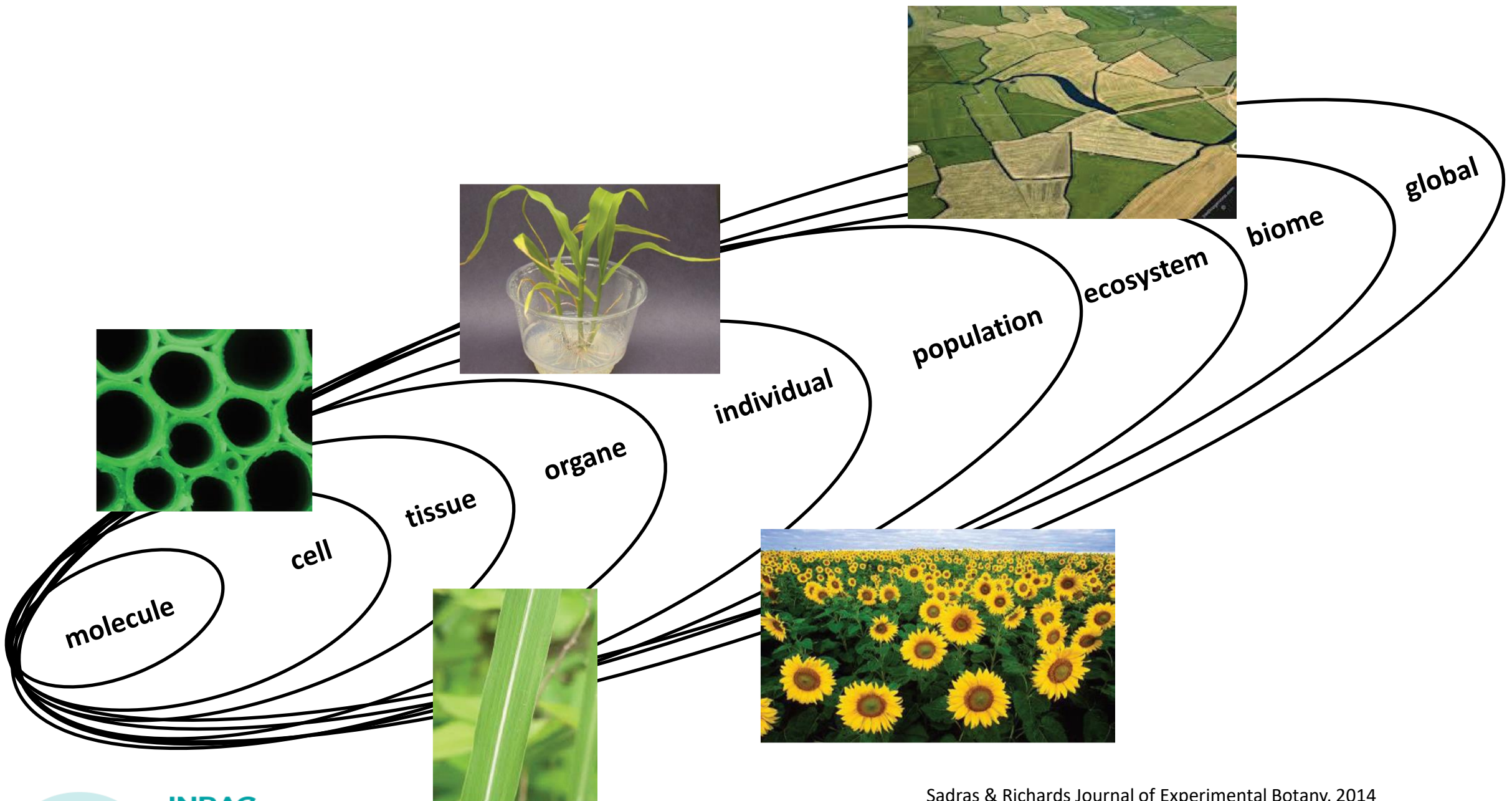
## Physiologie végétale



## Physiologie animale







**INRAE**

Design d'une étude métabolomique

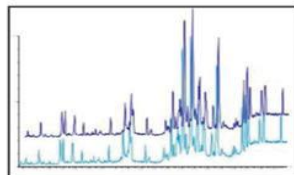
04-06-2020 / Production at analyse de données métabolomiques / Binta Diémé

Sadras & Richards Journal of Experimental Botany, 2014

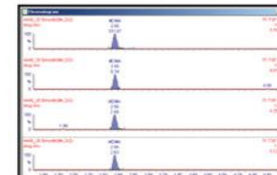
# Les différents types d'études métabolomiques

## Comprehensive Metabolomic Strategy

### Untargeted Analysis

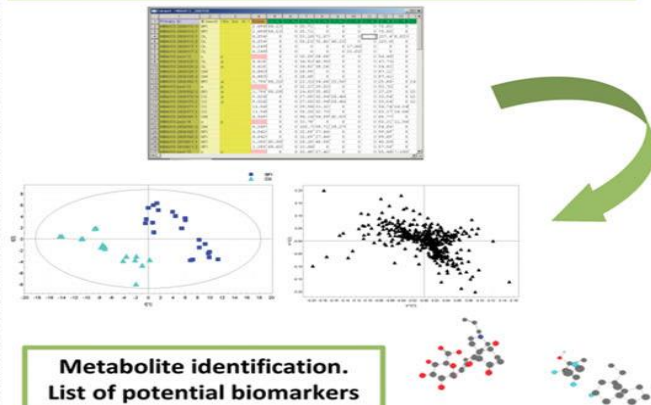


### Targeted Analysis



RP and HILIC chromatography...  
Data Pre-processing....

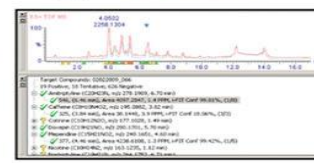
#### Multivariate Data Analysis (PCA, PLSDA..)



Metabolite identification.  
List of potential biomarkers

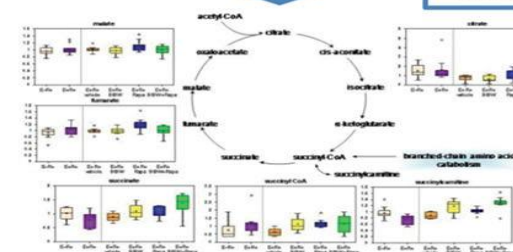
#### Target Metabolite list:

- Key metabolites of relevant pathways.
- Metabolite Class.

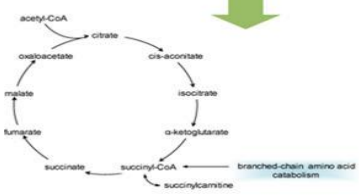


List of  
detected  
metabolites

#### Metabolite Quantification

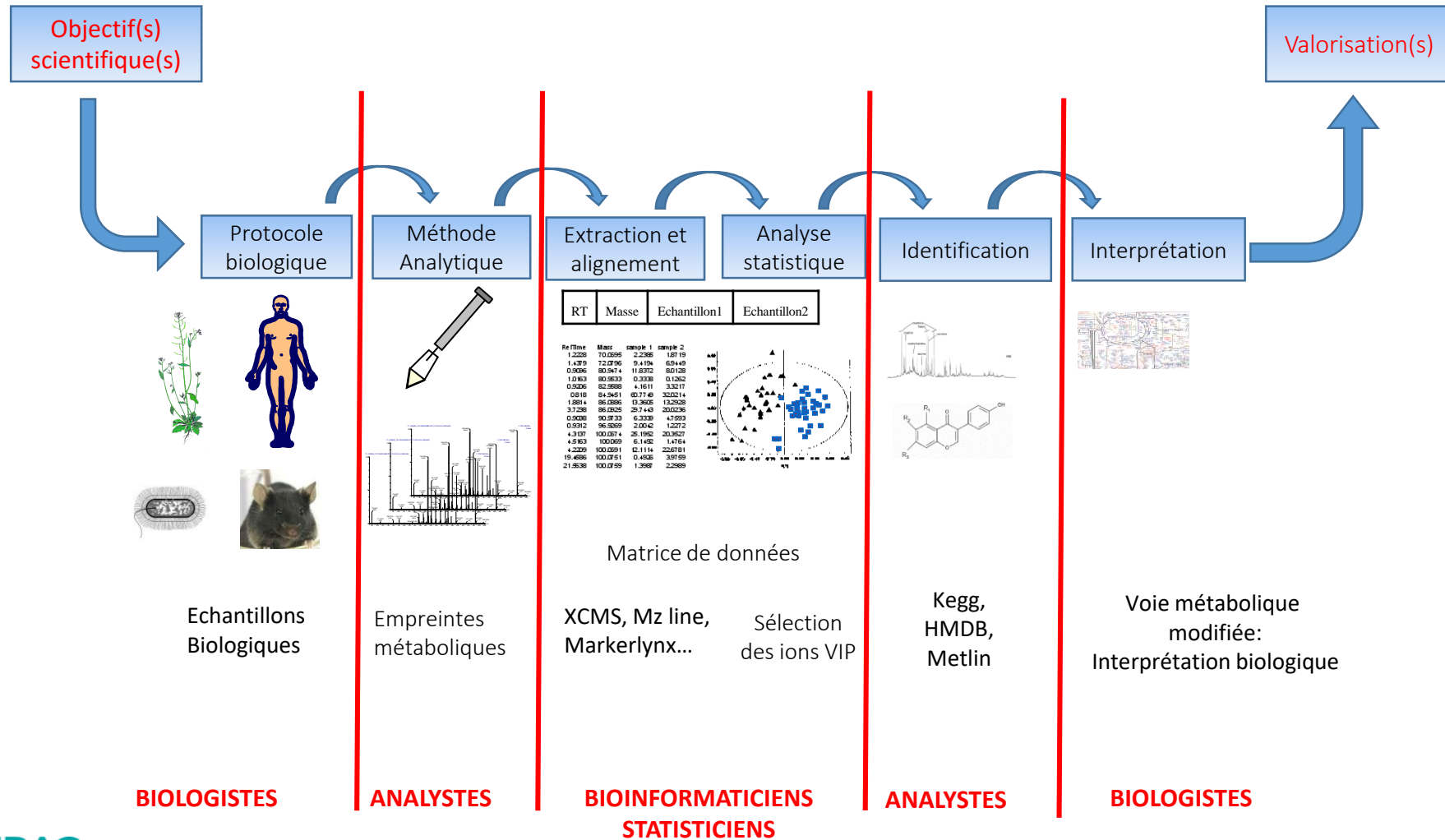


Biomarker Validation  
and  
Quantification



Biochemical interpretation

# Workflow d'une étude métabolomique



**INRAE**

Design d'une étude métabolomique

04-06-2020 / Production et analyse de données métabolomiques / Binta Diémé



# Etape 1: questions biologiques



## Etape 2: design de l'étude

### ➤ Mise au point du modèle expérimental:

- Modèle in vitro: cultures cellulaires, microorganismes
- Modèle animal

### ➤ Fluides biologiques: urine, plasma, sang, salive, LCR etc.

### ➤ biopsies

### ➤ Plantes



# Etape 1: questions biologiques



## Etape 2: design de l'étude

### ➤ Fluides et tissus biologiques humains: urine, plasma, LCR etc.

- Sexe
- Age
- Alimentation
- Prise médicamenteuse
- Exercice physique
- Heure du prélèvement

### ➤ Plantes

- Saison de prélèvement
- Lieu de prélèvement
- Conditions climatiques

# Etape 3: recueil et stockage des échantillons

## ➤ Prélèvement d'échantillons biologiques:

- Dépend de la nature de l'échantillon
- Doit être le plus court possible
- Puis l'aliquotage des échantillons (si fluides biologiques) permet de limiter les cycles congélations/décongélations

## ➤ Stockage

- Congélateurs -80°C



Etape 1: question biologique



Etape 2 : design de l'étude



Etape 3 : préparation des échantillons

- Randomisation de la préparation et de l'analyse des échantillons
- Garder au maximum la chaine de froid

## Etape 3 : préparation des échantillons

La préparation des échantillons dépend:

- du type de matrice à analyser
  - Solide: cellules, tissus, microorganismes
  - liquide
- du type de composés à analyser
  - Polaire, apolaire...
- de la plateforme analytique
  - RMN
  - MS (dérivation chimique pour la GC/MS par exemple)
  - Les deux?





## ➤ Points clés

- Renseigner au maximum les métadonnées:
  - caractéristiques de l'échantillon,
  - conditions de prélèvement ou de récolte
  - Conditions de stockage et d'analyses
  
- Extraire de manière répétable les métabolites
  
- Pas de protocole universel d'extraction des métabolites
  
- Outils analytiques complémentaires



# Les outils d'analyse: la spectrométrie de masse



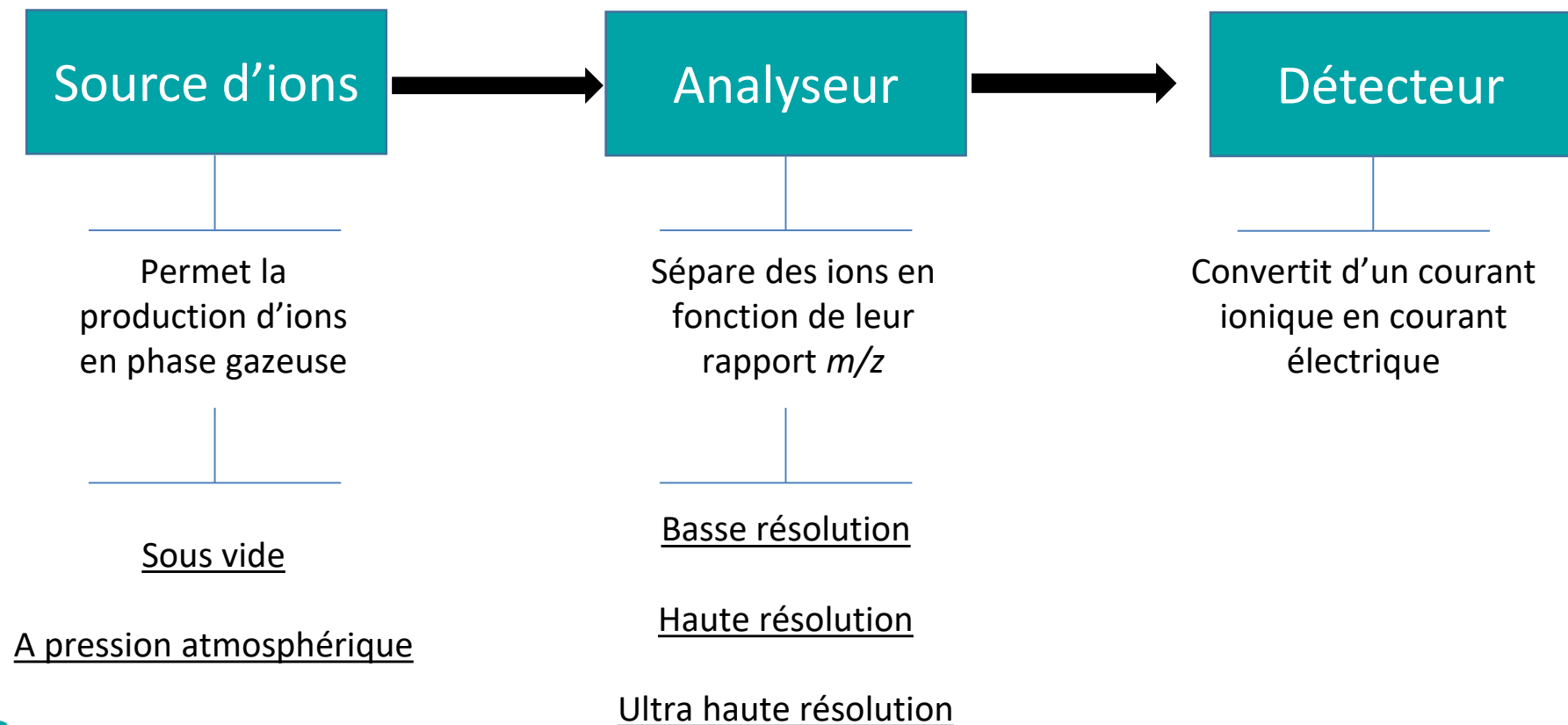
**INRAE**

Design d'une étude métabolomique

04-06-2020 / Production et analyse de données métabolomiques / Binta Diémé

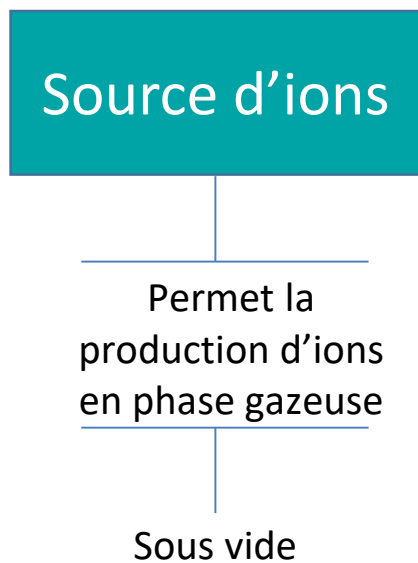
# ➤ La spectrométrie de masse

C'est une méthode de mesure des rapports masse sur charge de molécules ionisées



# ➤ La spectrométrie de masse : la source d'ions

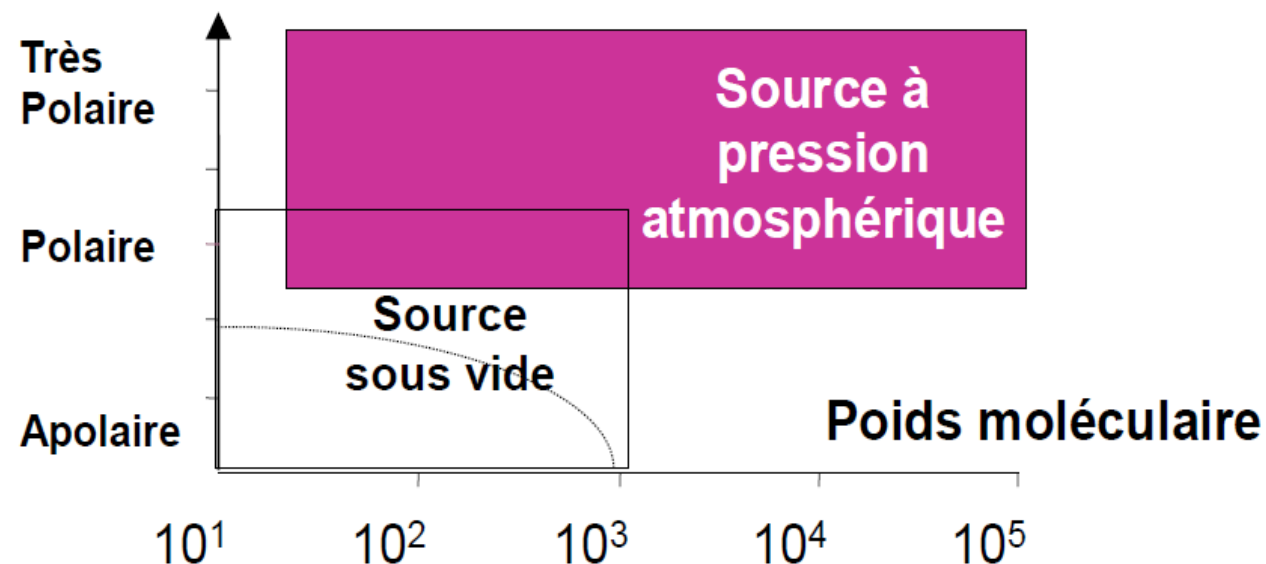
Objectif: ioniser et vaporiser les molécules



- Ionisation par impact électronique (EI): ionisation dure
- Ionisation chimique(CI): ionisation douce

A pression atmosphérique

- L'ionisation chimique à pression atmosphérique (APCI): composés apolaires
- Ionisation par électrospray (ESI): composés polaires



# ➤ La spectrométrie de masse : les analyseurs

Objectif: séparer les ions selon leur rapport masse/charge

Analyseur

Séparation des ions  
en fonction de leur  
rapport  $m/z$



Principales caractéristiques d'un analyseur:

- La résolution  $R$
- La gamme de masse ( $m/z$ ) qu'il peut analyser
- La rapidité de balayage en  $m/z$
- La sensibilité

Basse résolution

Haute résolution

Ultra haute résolution



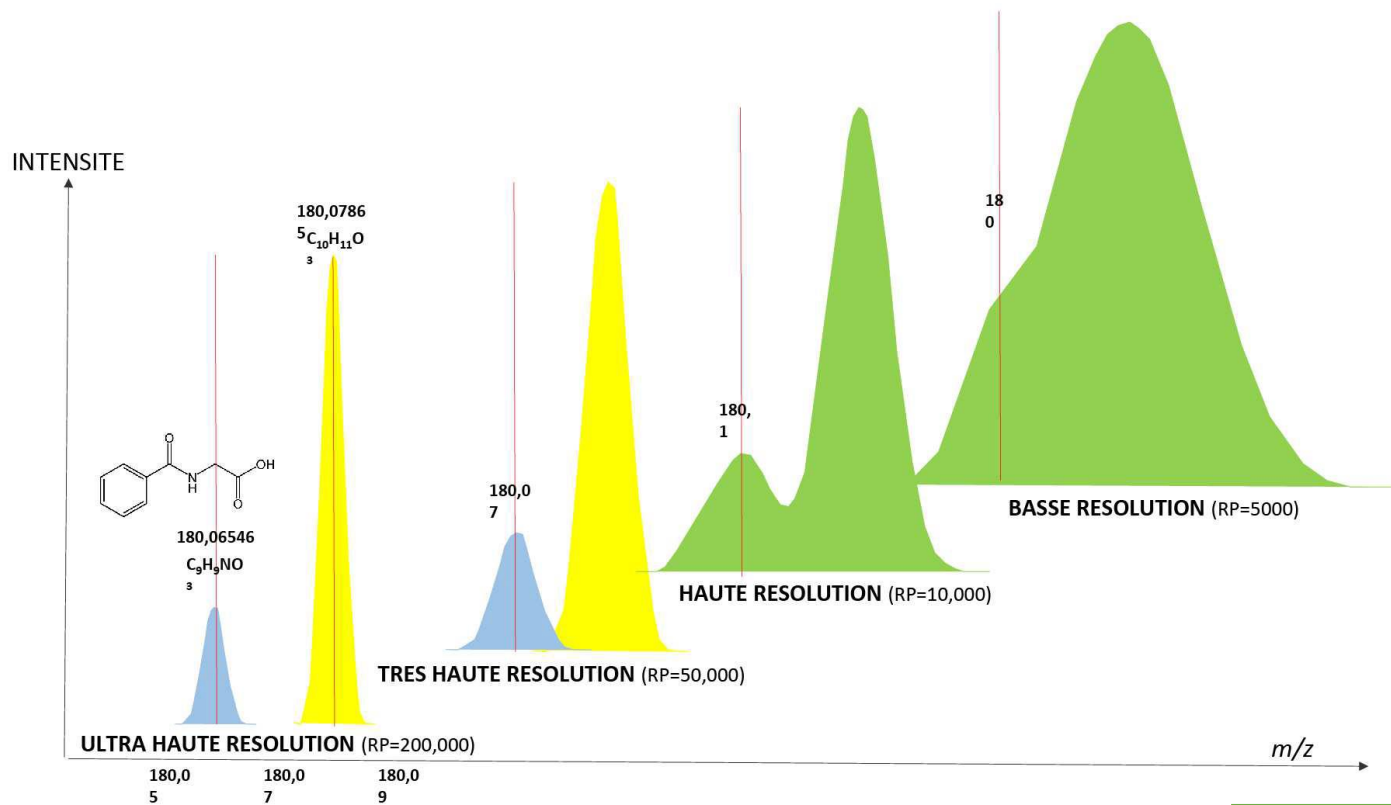
INRAE

Design d'une étude métabolomique

04-06-2020 / Production et analyse de données métabolomiques / Binta Diémé



# ➤ Les analyseurs: résolution



Tiré du MOOC de métabolomique

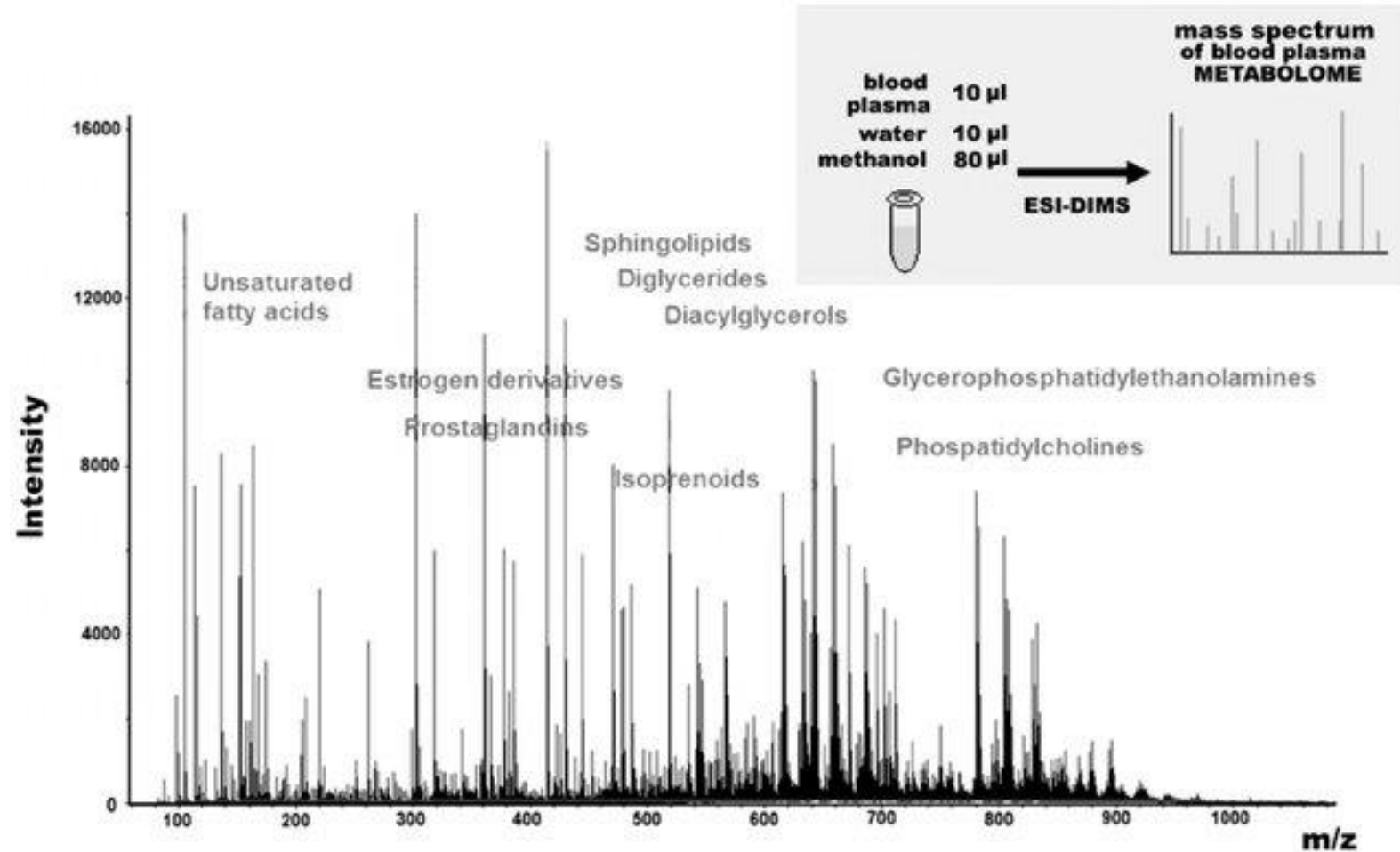
Analyseur	Résolution ( $m/\Delta m$ , FWHM)	Exactitude en masse (ppm)
Quadripôle	3 000	200
Piège à ions (2D/3D)	5 000	200
Temps de vol	8 000-80 000	1-10
Orbitrap®	Jusqu'à 280 000	<2
FT-ICR	Jusqu'à 10 000 000	<1

INRAE

Design d'une étude métabolomique

04-06-2020 / Production et analyse de données métabolomiques / Binta Diémé

# ➤ Complexité d'un spectre de masse: sans séparation chromatographique



Lokhov *et al.*, *PLoS ONE* (2014), 9 (9), e105343

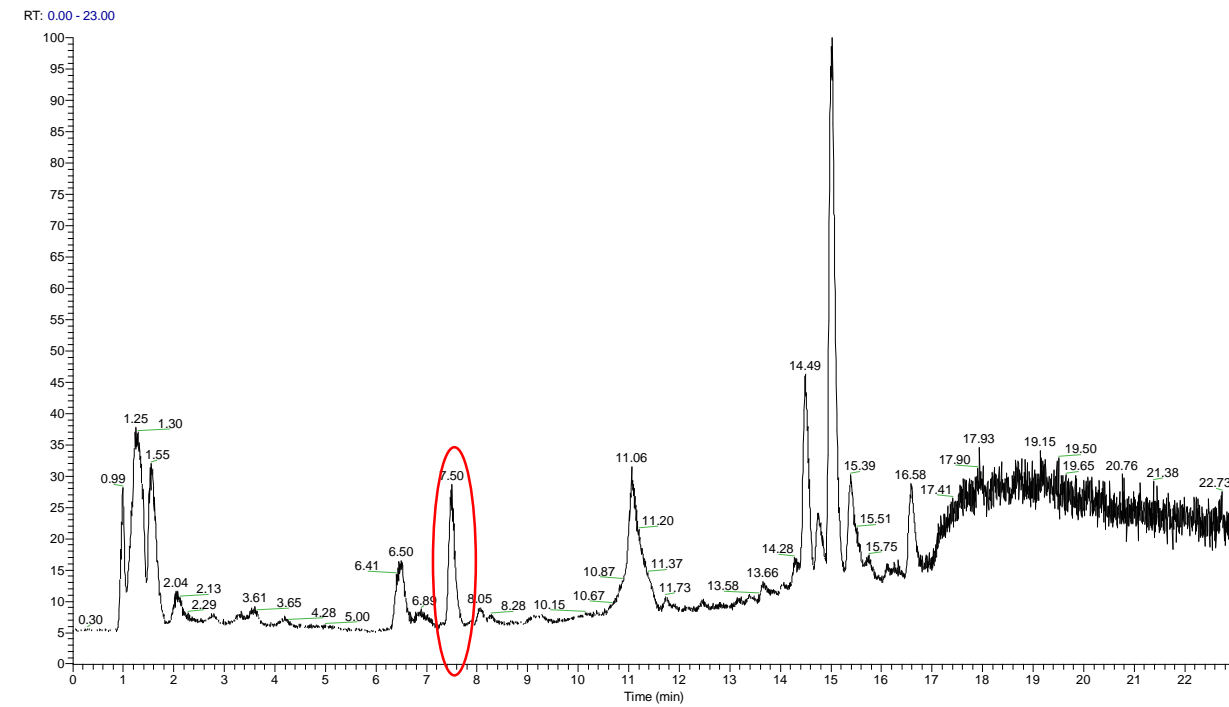


INRAE

Design d'une étude métabolique

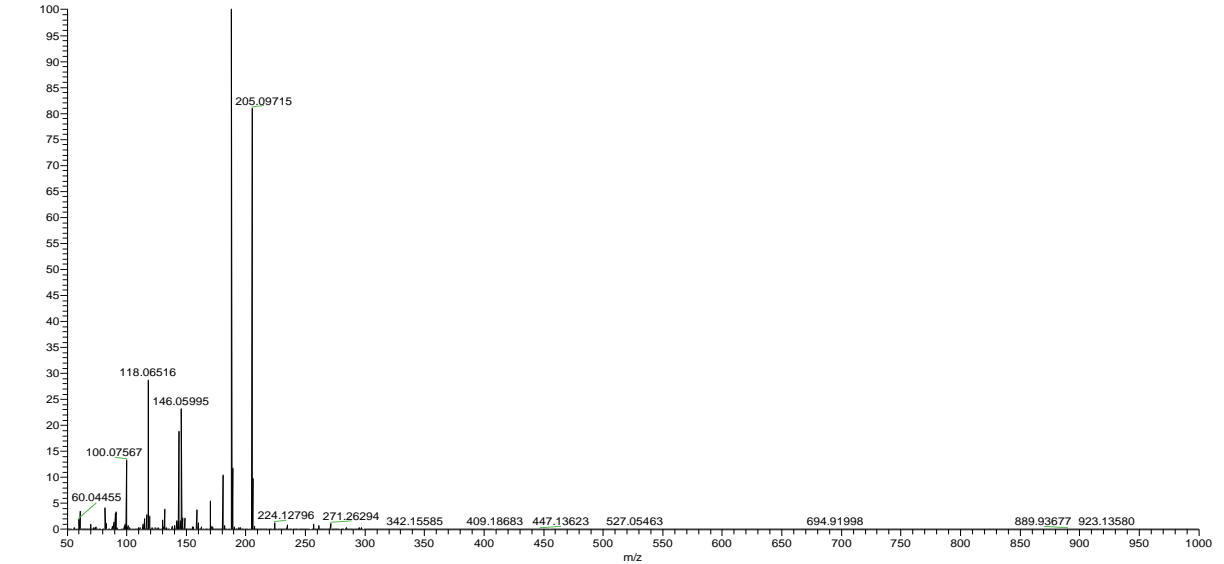
04-06-2020 / Production et analyse de données métaboliques / Binta Diémé

# Complexité d'un spectre de masse: avec séparation chromatographique



ScanPoolobi1\_pos #1353 RT: 7.47 AV: 1 NL: 4.57E6

F: FTMS + p ESI Full ms

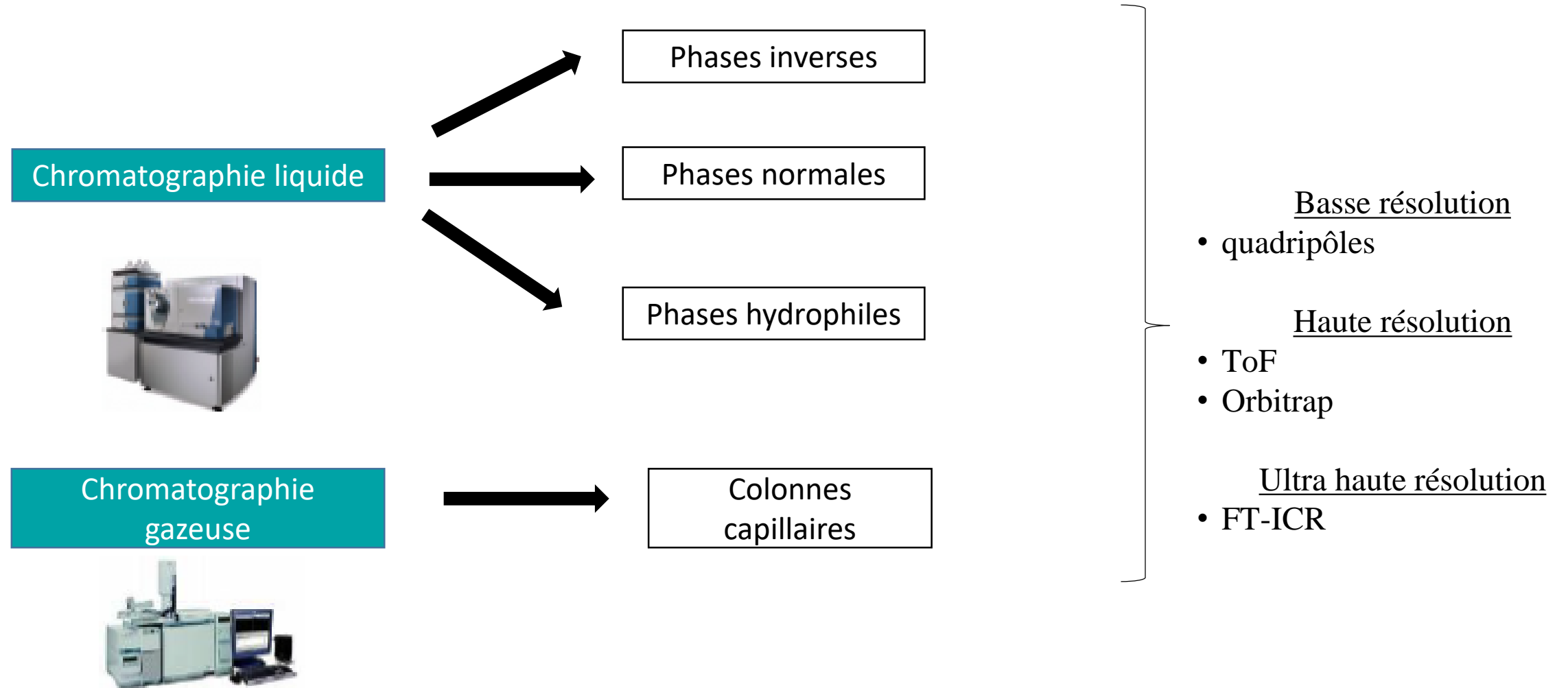


INRAE

Design d'une étude métabolique

04-06-2020 / Production et analyse de données métaboliques / Binta Diémé

# ➤ Les couplages chromatographiques



# ➤ Les couplages chromatographiques

## Chromatographie liquide

### ➤ Avantages de la LC-MS

- Analyse de composés polaires, thermolabiles et ioniques
- Accès à l'ion moléculaire
- Préparation de l'échantillon réduite (dans certains cas)

### ➤ Inconvénients de la LC-MS

- Ionisation douce
- Spectres de fragmentation peu reproductibles

## Chromatographie gazeuse

### ➤ Avantages de la GC-MS

- Analyse de composés apolaires et volatiles
- Spectres de fragmentation reproductibles

### ➤ Inconvénients de la GC-MS

- Préparation de l'échantillon plus longue
- Fragmentation intensive (EI), pas d'accès à l'ion moléculaire

### ➤ Inconvénients communs

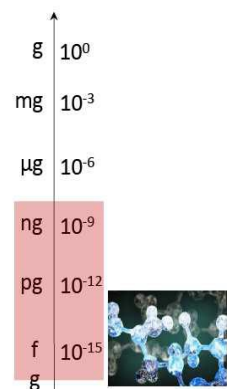
- Encrassement des sources
- Vieillessement des colonnes chromatographiques
- Phénomène de suppression d'ions





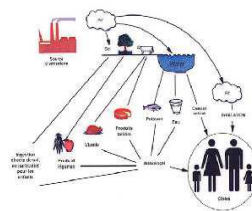
# ➤ Avantages de la spectrométrie de masse

Analyse possible d'ultra-traces



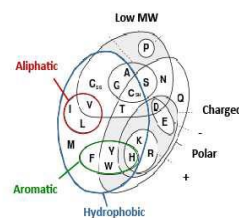
**SENSIBILITÉ**

Applicable à des matrices biologiques très diverses



**VERSATILITÉ**

Pris en charge d'analytes aux propriétés physico-chimiques très variées



**UNIVERSALITÉ**

Identification non ambiguë des analytes



**SÉLECTIVE**

Délivre des informations quantitatives très fiables



**PRÉCISION  
JUSTESSE**

Tiré du MOOC de métabolomique



**INRAE**

Design d'une étude métabolomique

04-06-2020 / Production et analyse de données métabolomiques / Binta Diémé

# Extraction des données en SM

Etape 4 : analyse des échantillons

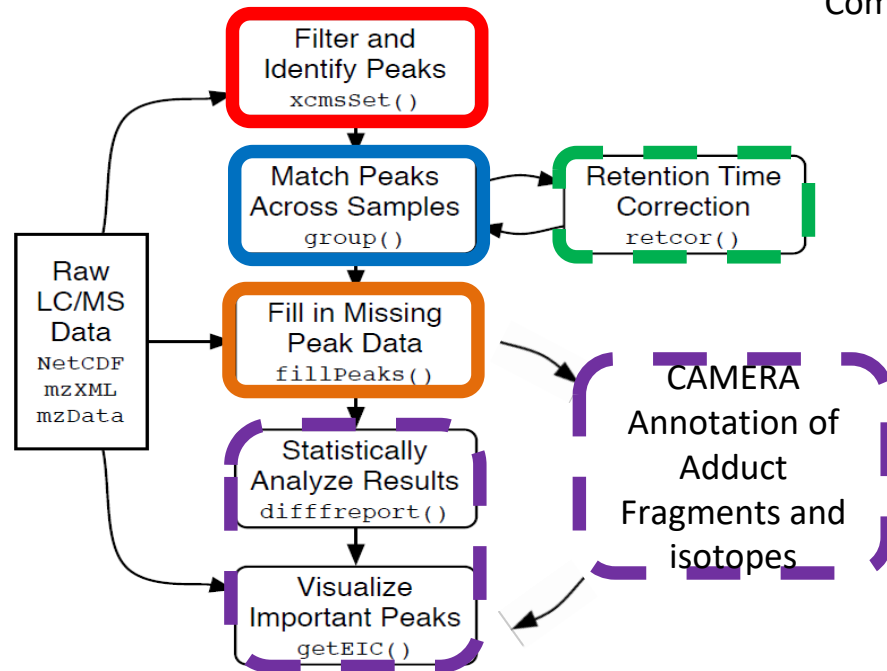
Etape 5 : Extraction des données

**Logiciels gratuits:**

- MZmine
- OpenMS
- MetAlign
- MSFACTS
- MSClust

**Logiciels commerciaux:**

- Progenesis®
- MS-Processor®
- Markerlynx®
- MassHunter®
- MassProfiler®
- Compound Discoverer®



Variable Echantillon	RT <sub>1</sub> -MS <sub>1</sub>	RT <sub>2</sub> -MS <sub>2</sub>	...	RT <sub>x</sub> -MS <sub>x</sub>
Patient 1	Int <sub>11</sub>	Int <sub>12</sub>		Int <sub>1x</sub>
Patient Y	Int <sub>Y1</sub>	Int <sub>Y2</sub>		Int <sub>Yx</sub>

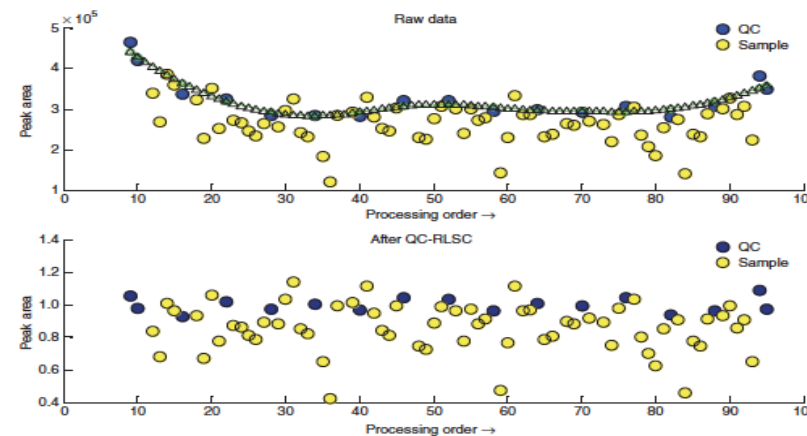


# ➤ Correction de la dérive du signal analytique en SM

Etape 5 : Extraction des données



Etape 6 : Correction dérive du signal



Dunn et al., (2011). *nature protocols*. 6, 1061-1083

**Figure 2 |** The QC-RLSC protocol for a metabolic feature detected in UPLC-MS (ES+) with signal attenuation across a given analytical batch. A cross-validated LOESS curve (upper plot) is fitted to the QC samples, the correction curve interpolated (triangles), to which the total data set for that peak is corrected (lower plot).

Filtres

- CV échantillons > CV CQ
- CV CQ < 30 %
- Ratio signal Echantillon/Blanc > 3

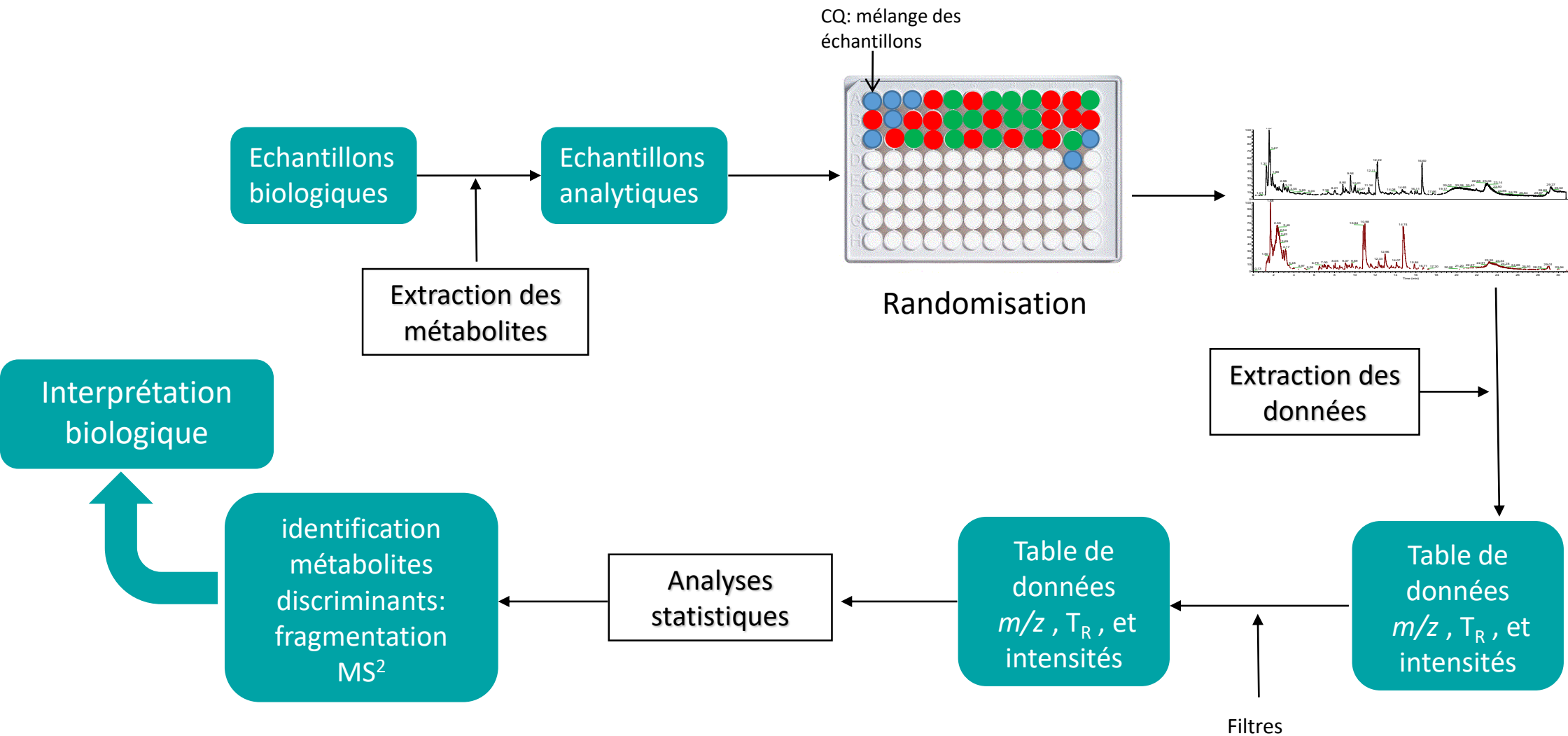
Table de données

INRAE

Design d'une étude métabolomique

04-06-2020 / Production at analyse de données métabolomiques / Binta Diémé

# Workflow en spectrométrie de masse (SM)



# TP conversion de fichiers



**INRAE**

Design d'une étude métabolomique

04-06-2020 / Production et analyse de données métabolomiques / Binta Diémé



# Les outils d'analyse: la résonance magnétique nucléaire

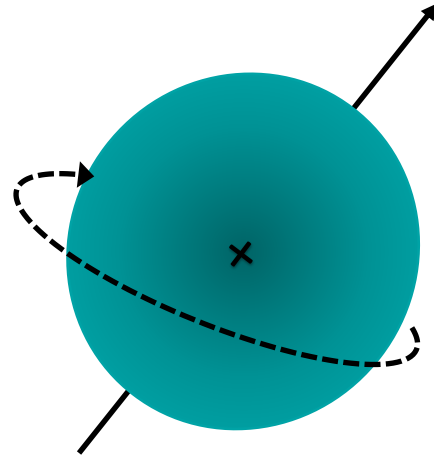


**INRAE**

Design d'une étude métabolomique

04-06-2020 / Production et analyse de données métabolomiques / Binta Diémé

## ➤ Concepts Physiques: spin



- Le spin est une des propriétés des particules, au même titre que la charge et la masse
- Le spin caractérise le magnétisme du noyau
- Le spin total dépend du nombre de particules dans le noyau (protons et neutrons) de l'isotope

## > Concepts Physiques: spin

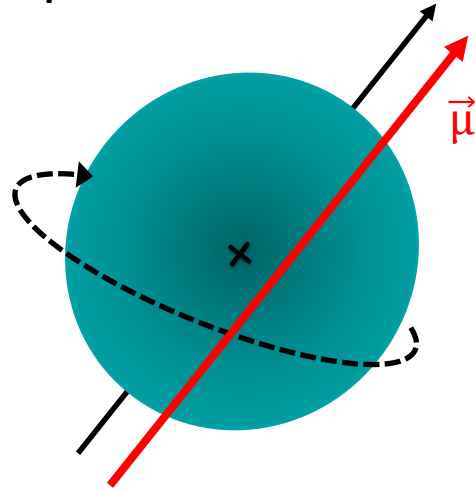
s	Masse atomique	Numéro atomique	
Demi-entier	impaire	Pair/impair	$^1\text{H}$ , $^{13}\text{C}$
Entier	paire	impair	$^2\text{D}$ , $^{14}\text{N}$
Nul	paire	pair	$^{12}\text{C}$ , $^{16}\text{O}$

Noyau	Spin
$^1\text{H}$	1/2
$^2\text{H}$	1
$^{14}\text{N}$	1
$^{15}\text{N}$	-1/2
$^{13}\text{C}$	1/2
$^{17}\text{O}$	-5/2



## ➤ Concepts Physiques: moment magnétique

Au moment cinétique de spin  $\vec{S}$  est associé un moment magnétique dipolaire  $\vec{\mu}$

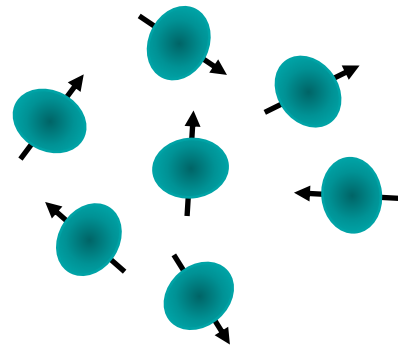


$$\vec{\mu} = \gamma \vec{S}$$

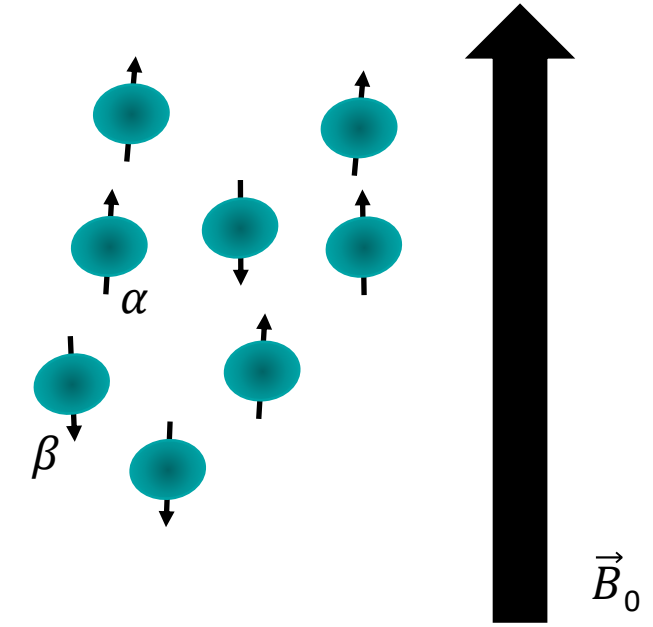
Le rapport gyromagnétique ( $\gamma$ ) est une caractéristique de chaque noyau atomique

# ➤ Concepts Physiques: moment magnétique vs champ magnétique

Absence champ magnétique



présence champ magnétique

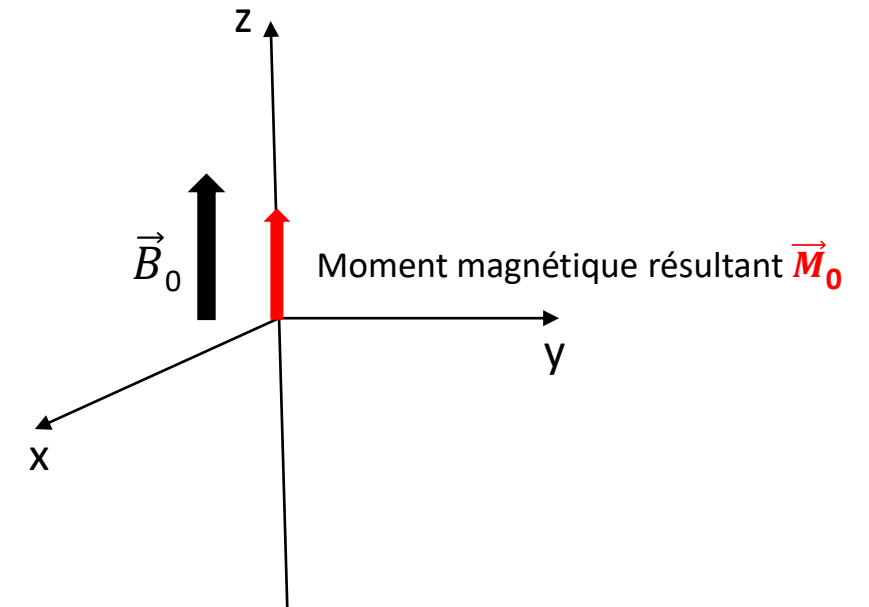
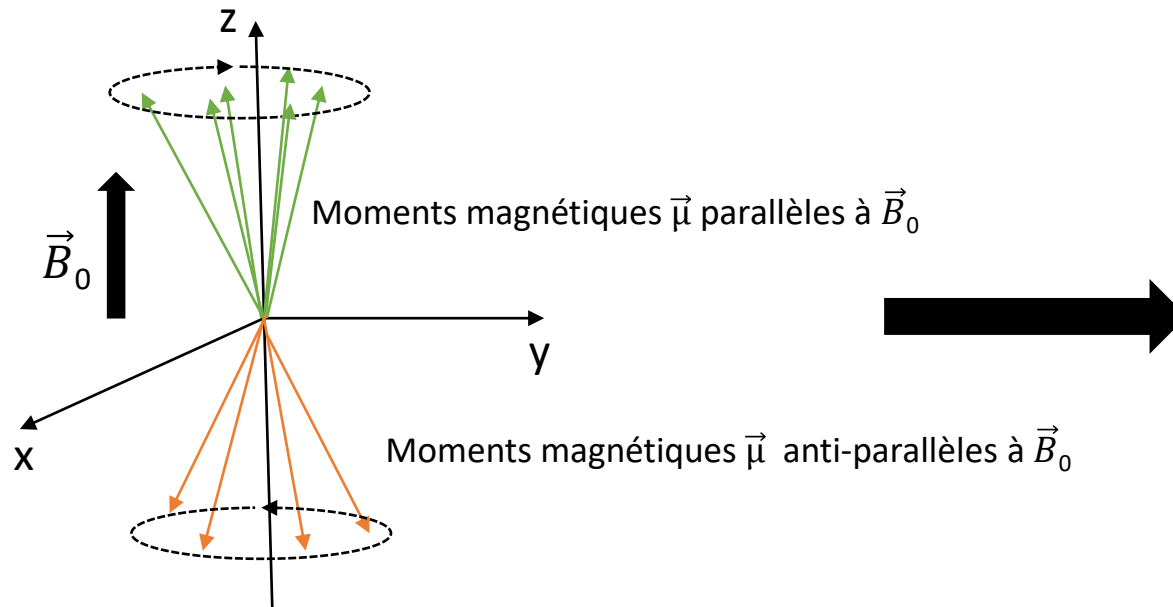


INRAE

Design d'une étude métabolomique

04-06-2020 / Production et analyse de données métabolomiques / Binta Diémé

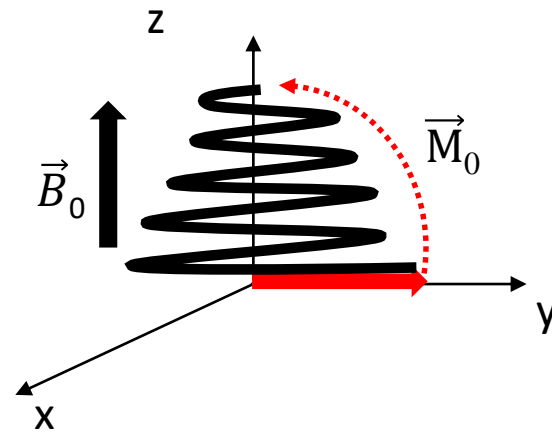
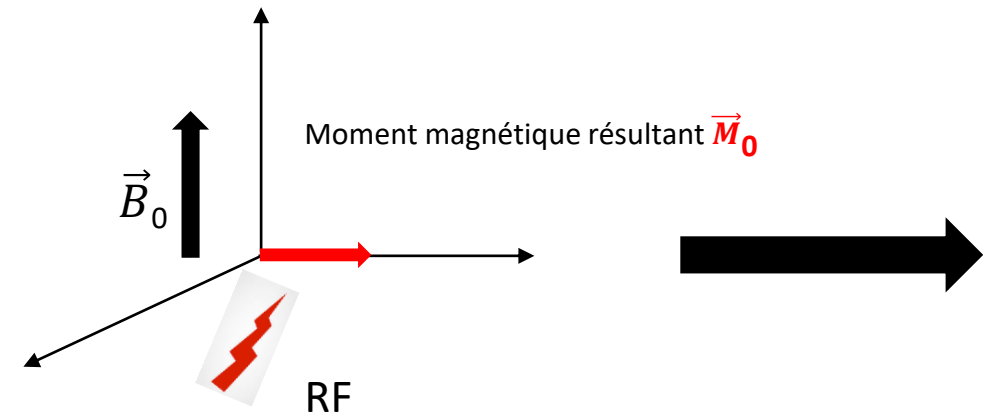
# ➤ Concepts Physiques: magnétisation



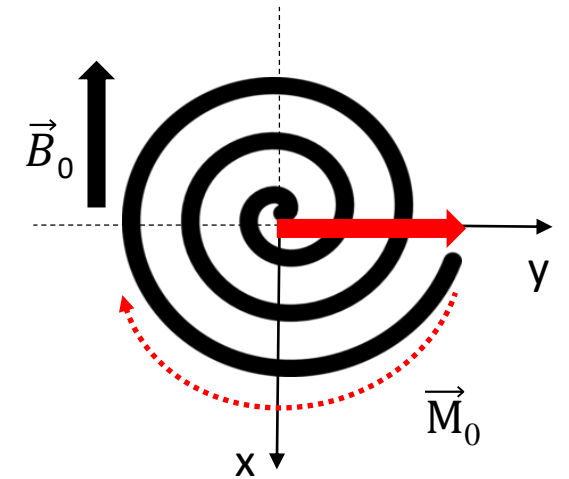
$$\vec{M} = \sum_N \vec{\mu} = \vec{M}_0$$



# ➤ Principe d'une expérience de RMN standard

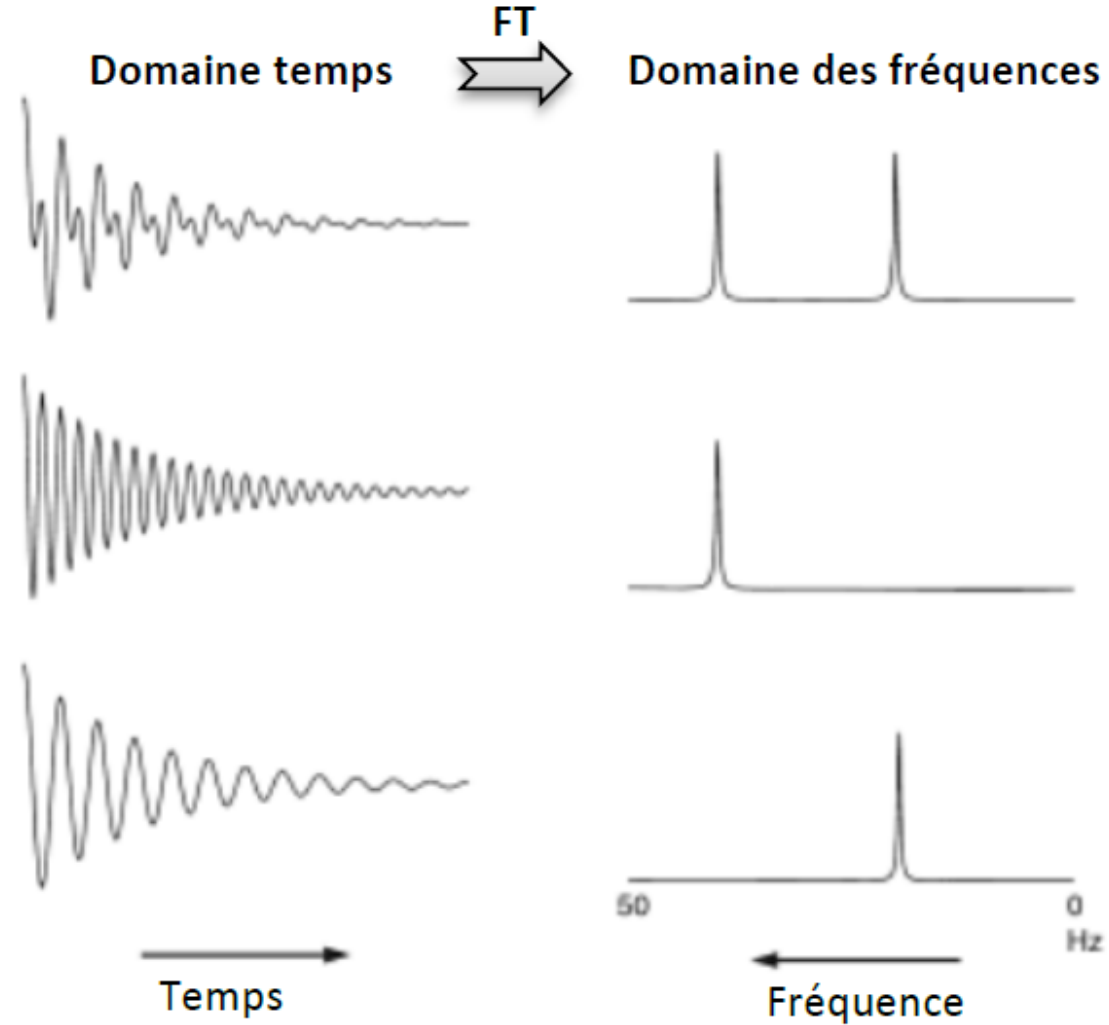


Relaxation longitudinale  $T_1$



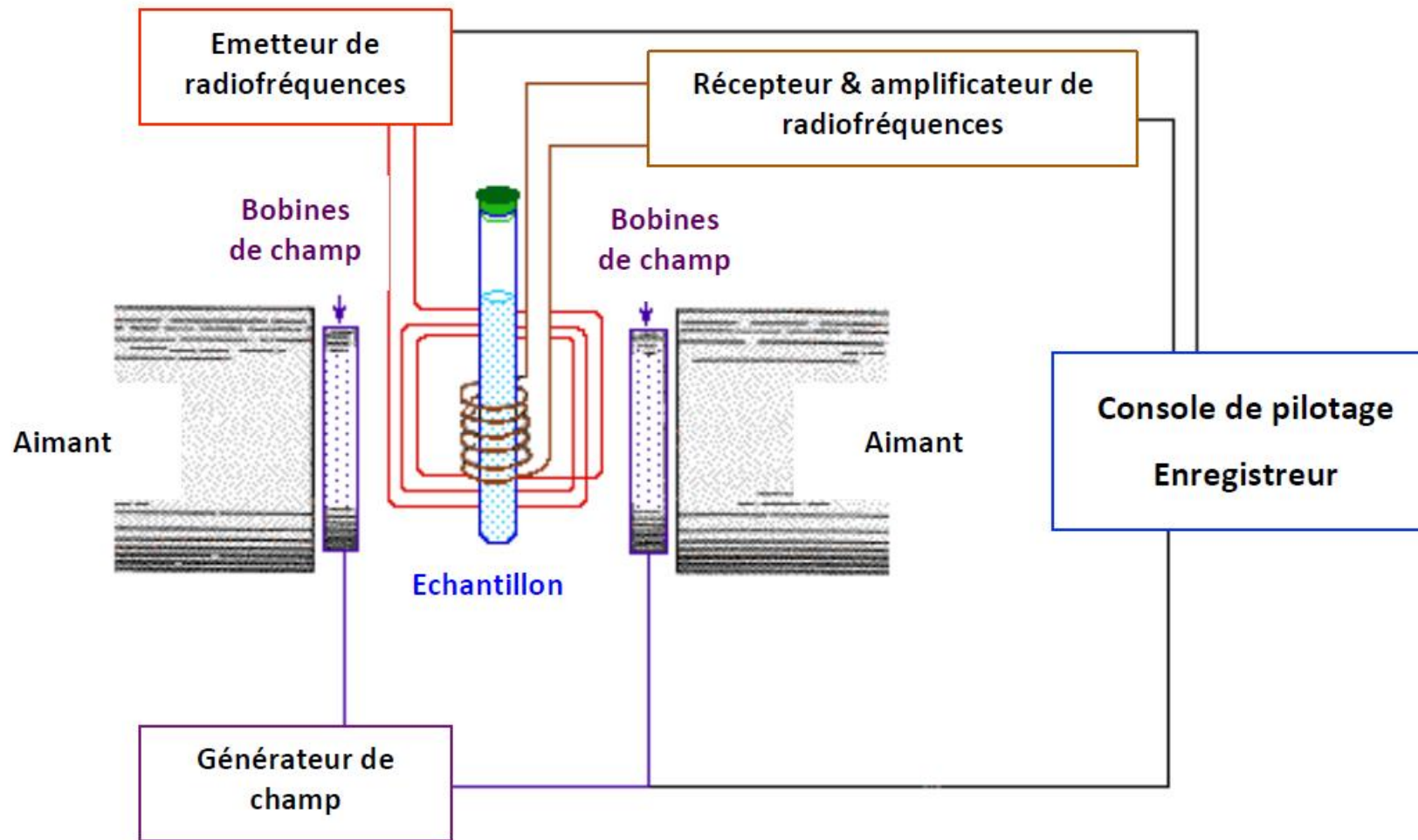
Relaxation transversale  $T_2$

# ➤ Le signal RMN

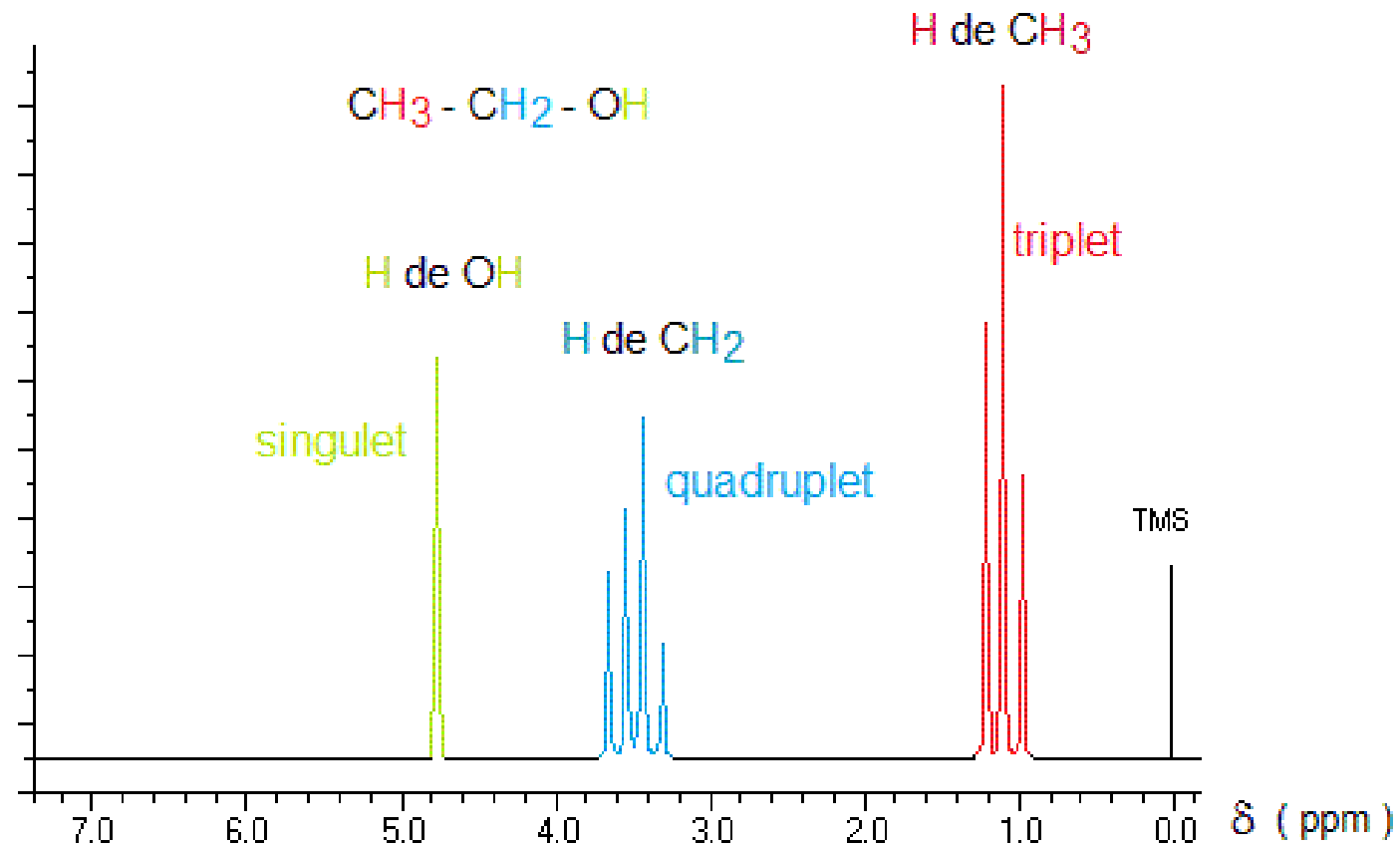
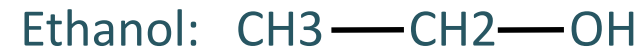




# ➤ Représentation schématique de l'appareillage RMN



## ➤ Structure d'un spectre 1D $^1\text{H}$



Spectre de RMN de l'éthanol

## ➤ Les informations contenues dans un spectre RMN

### ➤ Nature du noyau

- Fréquence de résonance
- Déplacement chimique

### ➤ Nombre de noyaux résonant à la même fréquence

- Concentration de la molécule

### ➤ Nombre de noyaux des groupements voisins

- Multiplicité des signaux



## ➤ La RMN: les spectromètres

Les plus utilisés sont les spectromètres à haut champ

Caractéristiques d'un spectromètre:

- La sensibilité et la résolution augmentent avec le champ magnétique
- Grand nombre de noyaux observables
- Expériences en 2D

Ils permettent d'analyser:

- Des biofluides
- Des extraits solubilisés
- Tissus ou des biopsies



# ➤ Les techniques analytiques

## RMN



RMN

SM

## SM



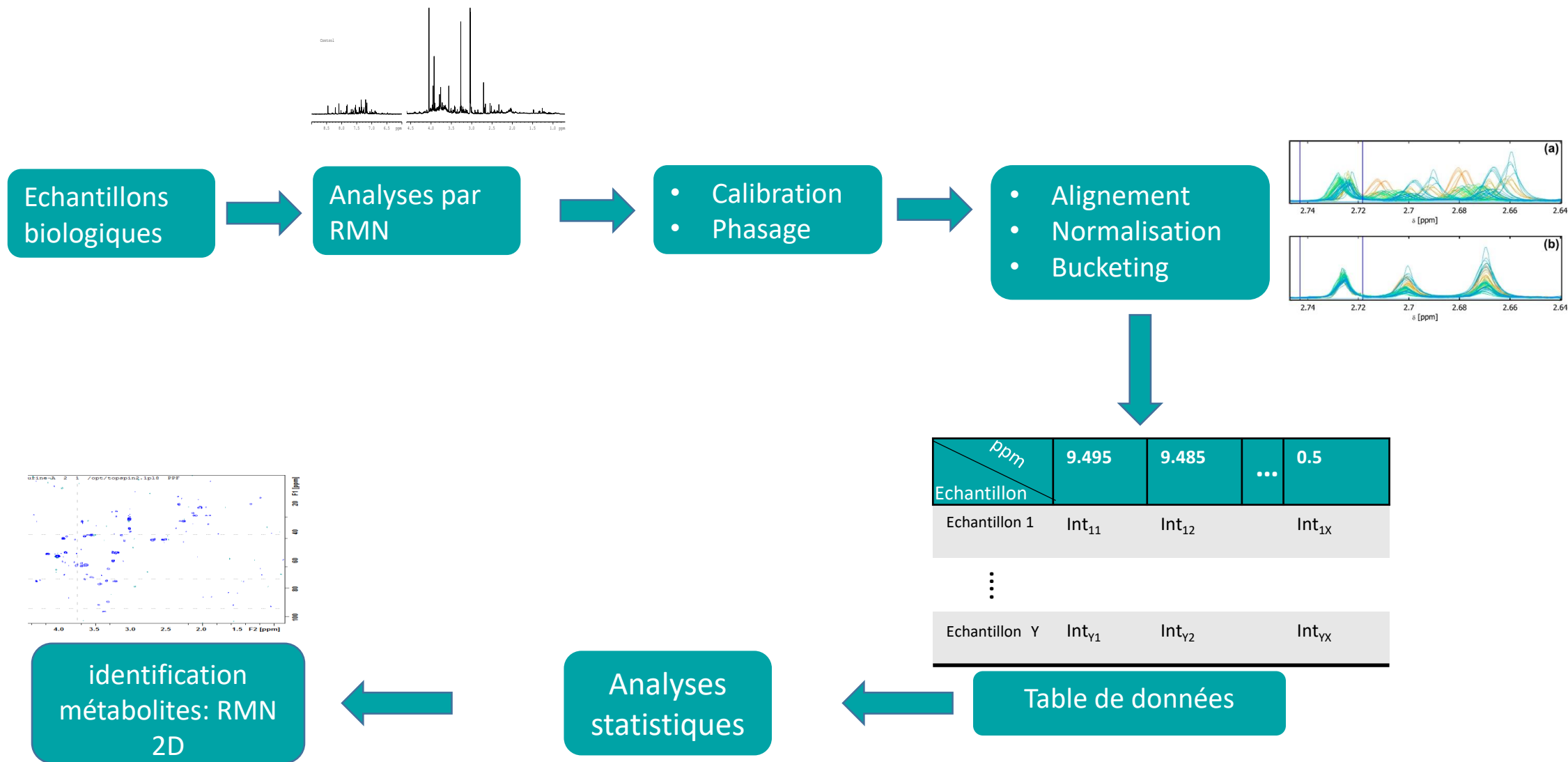
- Robuste
- Pas de préparation de l'échantillon
- Quantitative

- Moins robuste
- préparation de l'échantillon ++
- Semi-quantitative

Sensibilité faible

Sensibilité élevée

# Workflow en RMN



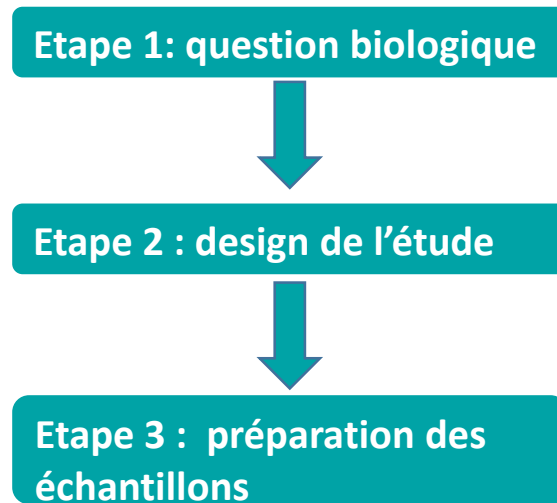
**INRAE**

Design d'une étude métabolomique

04-06-2020 / Production at analyse de données métabolomiques / Binta Diémé

Euceda and al., *Scandinavian Journal of Clinical and Laboratory Investigation* 2015, 75, 193-203.

## ➤ Préparation d'échantillons biologiques en RMN



- Ajout de solvant deutéré à l'échantillon solide (tissus, cellules) ou liquide (urines)
- Utilisation de solutions tampons
- Ajout d'un standard externe (TSP, DSS...)



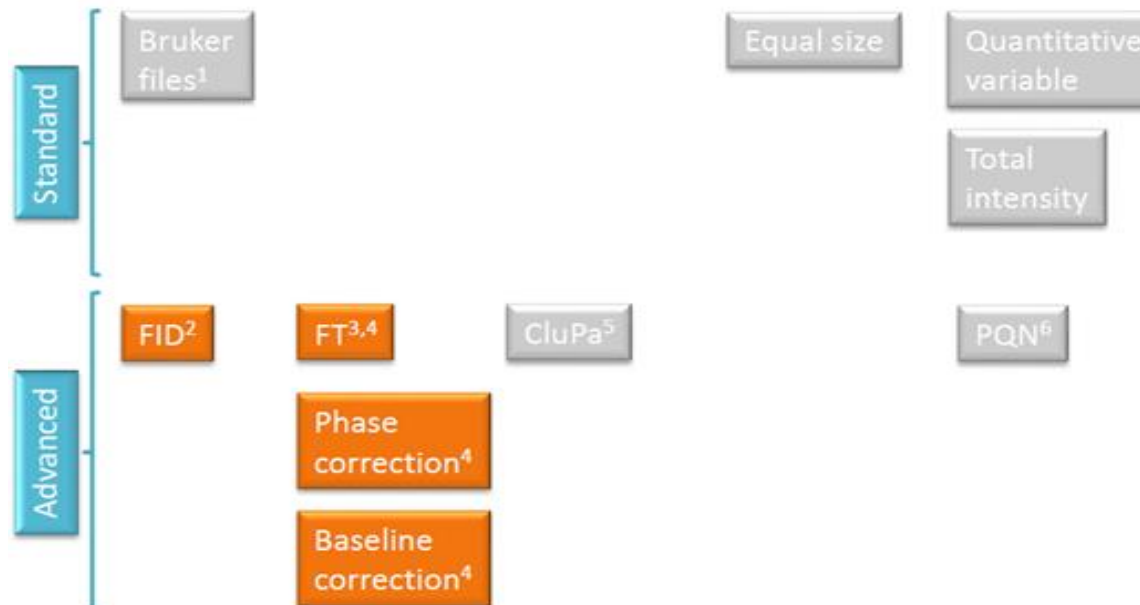
# ➤ Extraction des données en RMN

**Solutions gratuites:**  
NMRProcFlow  
W4M

Etape 4 : analyse des échantillons

Etape 5 : Extraction des données

**Solutions commerciales:**  
AMIX®  
Icoshift® sous Matlab  
MestRenova®



**INRAE**



## ➤ Etude de LCR de patients d'Alzheimer par $^1\text{H}$ -RMN

Analyse par  $^1\text{H}$ -RMN de 77 liquides céphalo-rachidiens (LCR) de différents patients (collaboration du Dr Pariente Jérémie et Christian Vincent; et hôpitaux de Toulouse)

Contrôle (Ctr) (23 patients)	Démence (DE) non MA (27 patients)	MA (27 patients)
RAS	Démence sémantique	Maladie d'Alzheimer (MA)
Maladie psychiatrique	Démence fronto temporale	MA + angiopathie amyloïde
Neuropathie périphérique	CADASIL	MA + démence OH
Alcoolisme	Dégénérescence cortico-basale	MA + HPN
Comitialité (épilepsie)	Vascular cognitive impairment	
Maladie psychiatrique	Paralysie supranucléaire progressive	
AVC thalamique antérieur	Maladie de Creutzfeldt-Jakob	
Psychose chronique	Dégénérescence lobaire fronto-temporale	
Hydrocéphalie à pression normale	Sclérose latérale amyotrophique (=maladie de Charcot)	
	Maladie neuro-dégénérative non étiquetée	
	Démence alcoolique + épilepsie	
	Démence vasculaire	

Cruz, T. et al.(2014). *Metabolites*. 4, 115-128.



INRAE

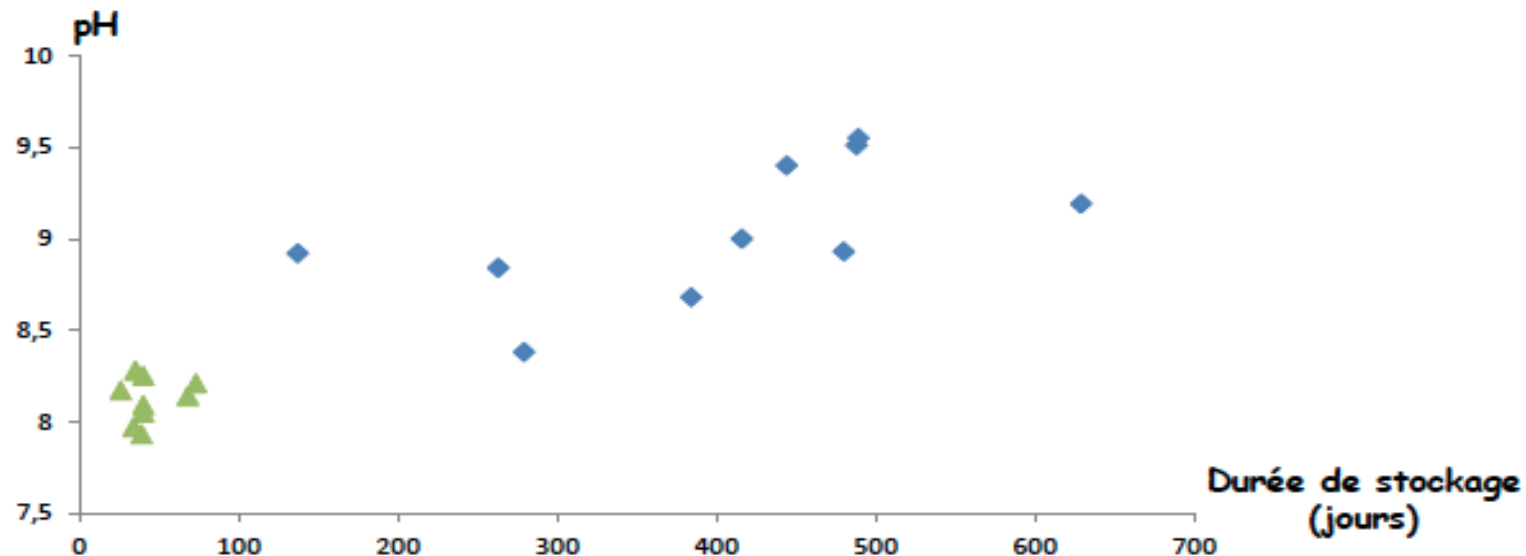
Design d'une étude métabolomique

04-06-2020 / Production et analyse de données métabolomiques / Binta Diémé

## ➤ Etude de LCR de patients d'Alzheimer par $^1\text{H}$ -RMN

Le pH du LCR évolue naturellement dès son prélèvement

- Au moment du prélèvement il est entre 7.4 et 7.6
- Après 2h à 4°C il peut atteindre 9.6



Des variations de pH entraînent des déplacements chimiques de certaines molécules



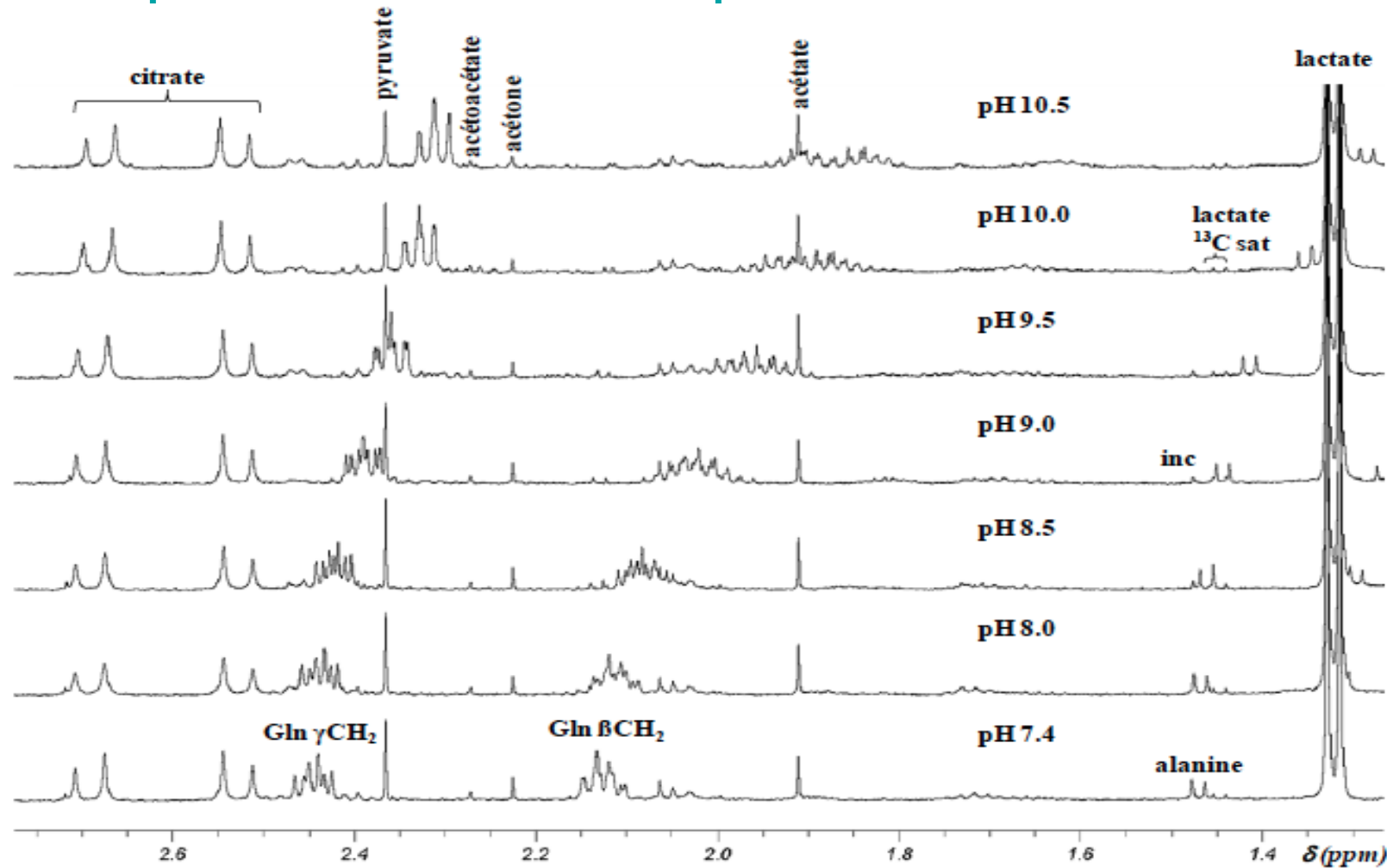
INRAE

Design d'une étude métabolomique

04-06-2020 / Production et analyse de données métabolomiques / Binta Diémé

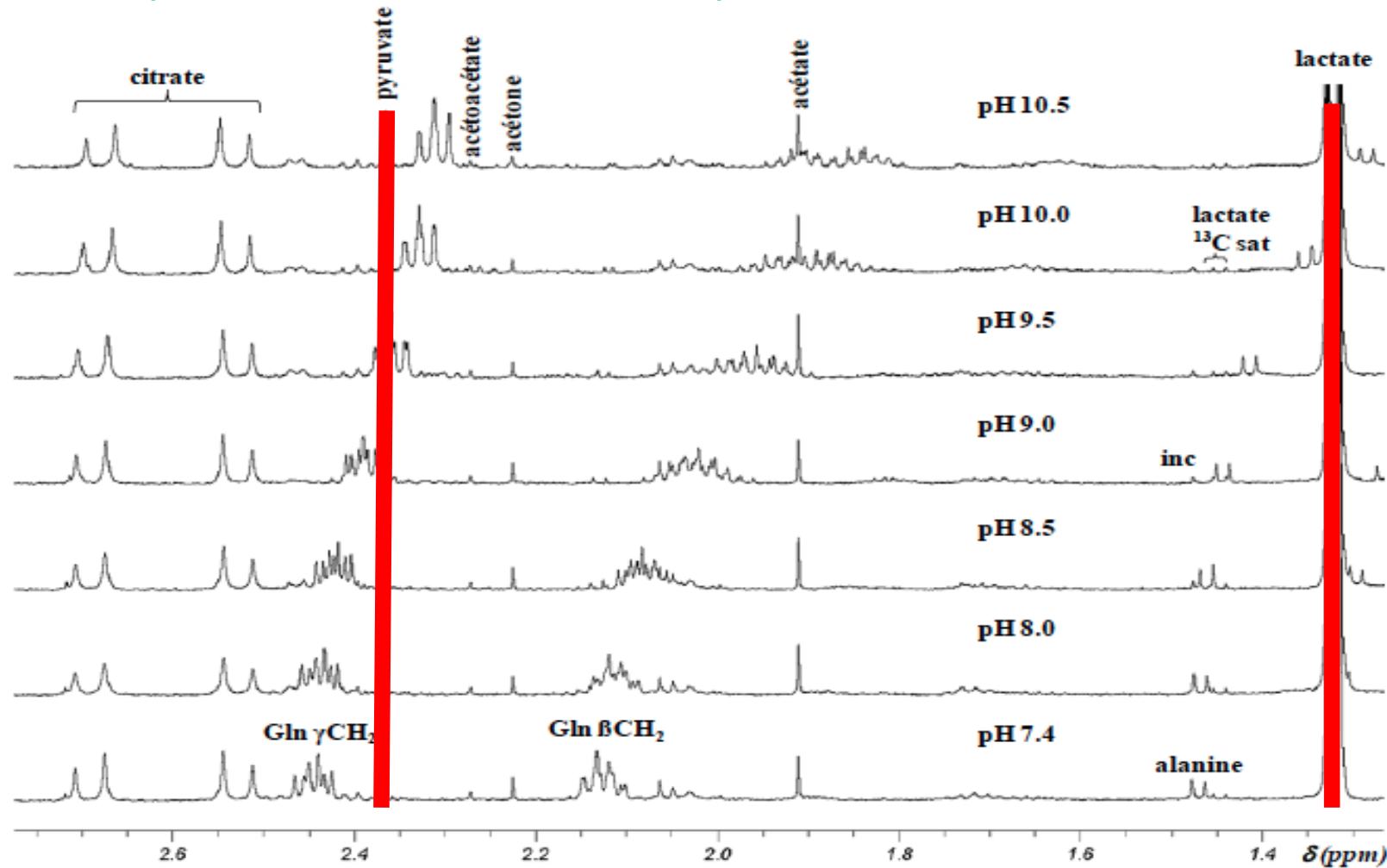
Cruz, T. et al. (2014). *Metabolites*. **4**, 115-128.

## ➤ Etude de LCR de patients d'Alzheimer par $^1\text{H}$ -RMN



**Figure 4.** Representative  $^1\text{H}$  NMR spectra (1.25–2.8 ppm region) of a CSF sample containing 33%  $\text{D}_2\text{O}$  whose pH was adjusted at different pH between 7.4 and 10.5. Gln: glutamine; sat: satellite; unk: unknown.

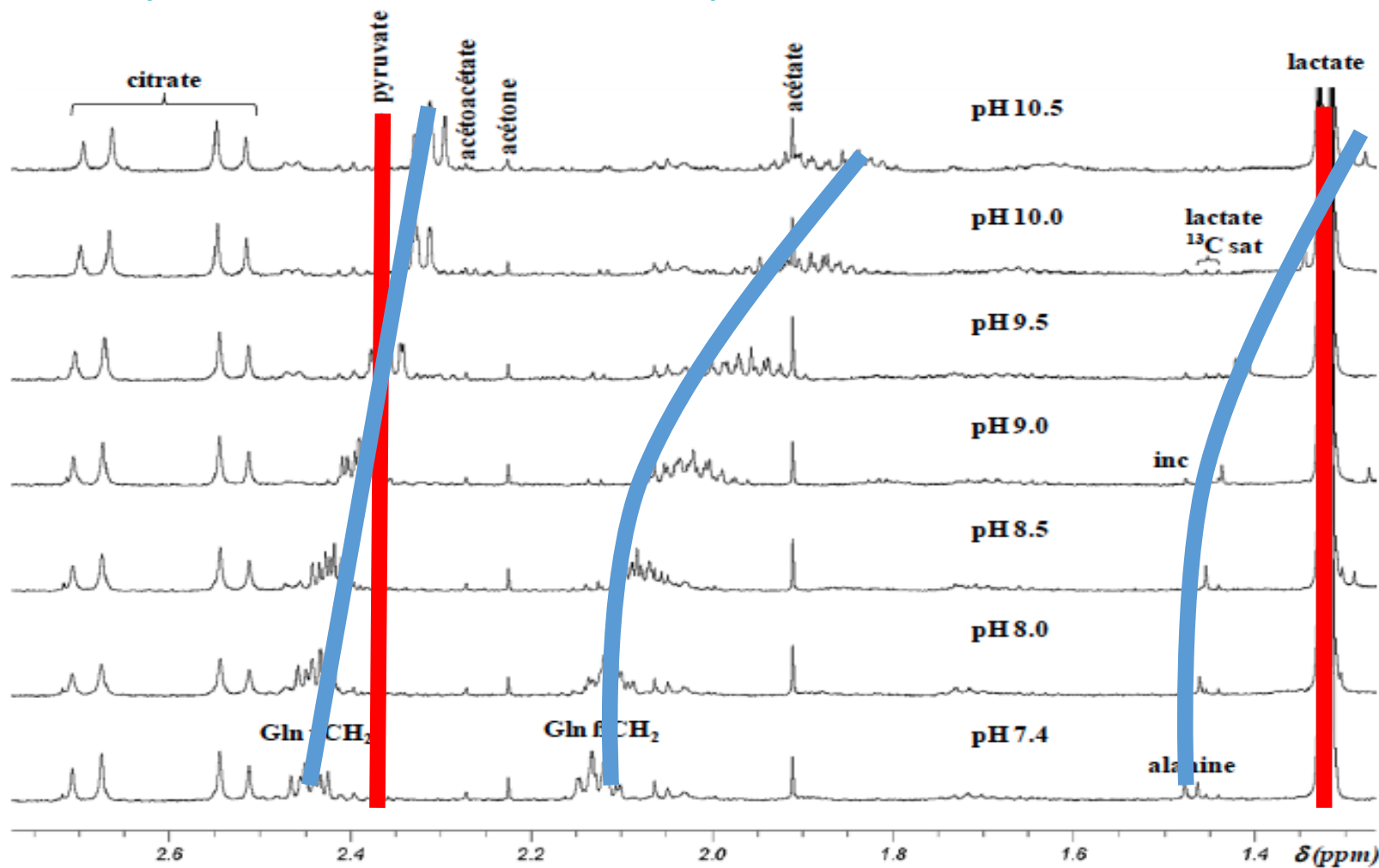
# ➤ Etude de LCR de patients d'Alzheimer par $^1\text{H}$ -RMN



**Figure 4.** Representative  $^1\text{H}$  NMR spectra (1.25–2.8 ppm region) of a CSF sample containing 33%  $\text{D}_2\text{O}$  whose pH was adjusted at different pH between 7.4 and 10.5. Gln: glutamine; sat: satellite; unk: unknown.

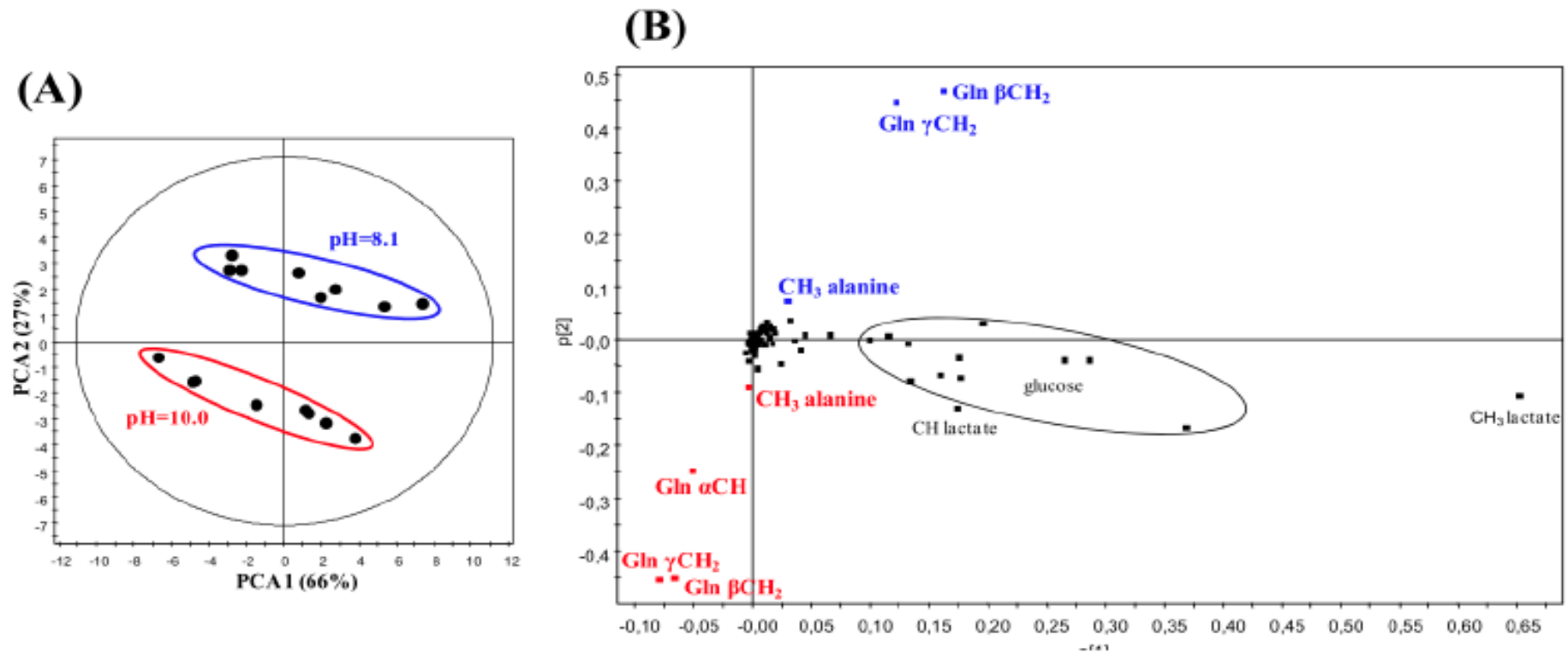


# ➤ Etude de LCR de patients d'Alzheimer par $^1\text{H}$ -RMN



**Figure 4.** Representative  $^1\text{H}$  NMR spectra (1.25–2.8 ppm region) of a CSF sample containing 33%  $\text{D}_2\text{O}$  whose pH was adjusted at different pH between 7.4 and 10.5. Gln: glutamine; sat: satellite; unk: unknown.

# ➤ Etude de LCR de patients d'Alzheimer par $^1\text{H}$ -RMN



Cruz, T. et al. (2014). *Metabolites*. **4**, 115-128.



INRAE

Design d'une étude métabolomique

04-06-2020 / Production et analyse de données métabolomiques / Binta Diémé

## ➤ Etude de LCR de patients d'Alzheimer par $^1\text{H}$ -RMN

- Le contrôle du pH est primordial dans le LCR
- Mauvaises interprétations des résultats biologiques



Importance d'éviter tout biais analytique

# Les analyses statistiques



**INRAE**

Design d'une étude métabolomique

04-06-2020 / Production et analyse de données métabolomiques / Binta Diémé



# Tableaux d'informations

## Matrice de Design

Echantillon, Individu

Label	Age	BMI	Sexe
HU_011	29	19.75	M
HU_014	59	22.64	F
HU_015	42	22.72	M
HU_017	41	23.03	M
HU_018	34	20.96	M
HU_019	35	23.41	M
HU_020	59	17.1	M
HU_021	34	23.36	M
HU_022	51	28.23	F
HU_023	51	29.55	M
HU_024	57	29.86	M
HU_025	53	21.6	M
HU_026	34	23.46	F
HU_027	37	24.82	M

Identifiant

## Tableau des Données métabolomiques

Label	MTB 1	MTB 2	MTB 3	MTB 4	MTB 5	MTB 6	MTB 7	MTB 8	MTB 9	MTB 10	MTB 11	MTB 12	MTB 13	MTB 14
HU_011	3.02	3.888	3.869	3.717	3.535	3.325	4.205	4.08	4.56	3.467	4.298	3.8	3.503	4.673
HU_014	3.814	4.277	3.838	3.777	3.933	4.025	5.182	4.359	3.158	3.943	3.834	4.192	4.37	4.49
HU_015	3.52	4.196	4.102	4.292	3.955	3.973	3.886	4.249	4.291	3.54	4.199	4.147	3.991	4.416
HU_017	2.562	4.324	4.539	4.432	4.228	4.11	4.238	4.231	3.645	3.152	3.865	4.233	4.151	3.961
HU_018	3.782	4.629	4.179	4.117	4.006	4.024	1.853	4.324	4.096	4.49	4.271	4.755	4.373	4.248
HU_019	4.161	4.412	4.517	4.562	4.321	4.327	4.081	4.244	4.428	4.157	3.832	4.334	1.446	4.164
HU_020	3.199	4.175	4.18	4.253	4.014	4.198	4.379	4.124	3.885	3.92	3.818	3.891	4.162	4.552
HU_021	3.378	4.635	4.036	3.988	3.667	3.997	3.954	4.262	2.298	3.824	4.117	4.535	4.176	3.971
HU_022	3.265	3.847	4.295	4.071	4.57	4.459	4.258	4.163	4.705	4.177	4.01	4.367	4.333	4.509
HU_023	3.787	3.872	4.398	4.469	4.686	4.636	3.482	4.143	3.659	3.384	3.714	4.381	3.604	3.902
HU_024	5.364	4.418	4.53	4.498	4.442	4.421	4.359	4.287	3.656	4.059	4.115	4.108	4.065	4.31
HU_025	3.202	4.329	4.11	4.125	4.009	3.945	4.191	4.18	4.169	4.361	3.962	4.23	4.007	3.891
HU_026	3.98	4.03	3.74	4.159	3.199	4.255	4.207	4.159	3.712	3.207	3.957	4.064	1.462	4.474
HU_027	3.02	4.152	3.784	3.748	3.703	4.132	4.007	4.092	3.905	3.89	3.708	4.271	4.138	4.723

n observations

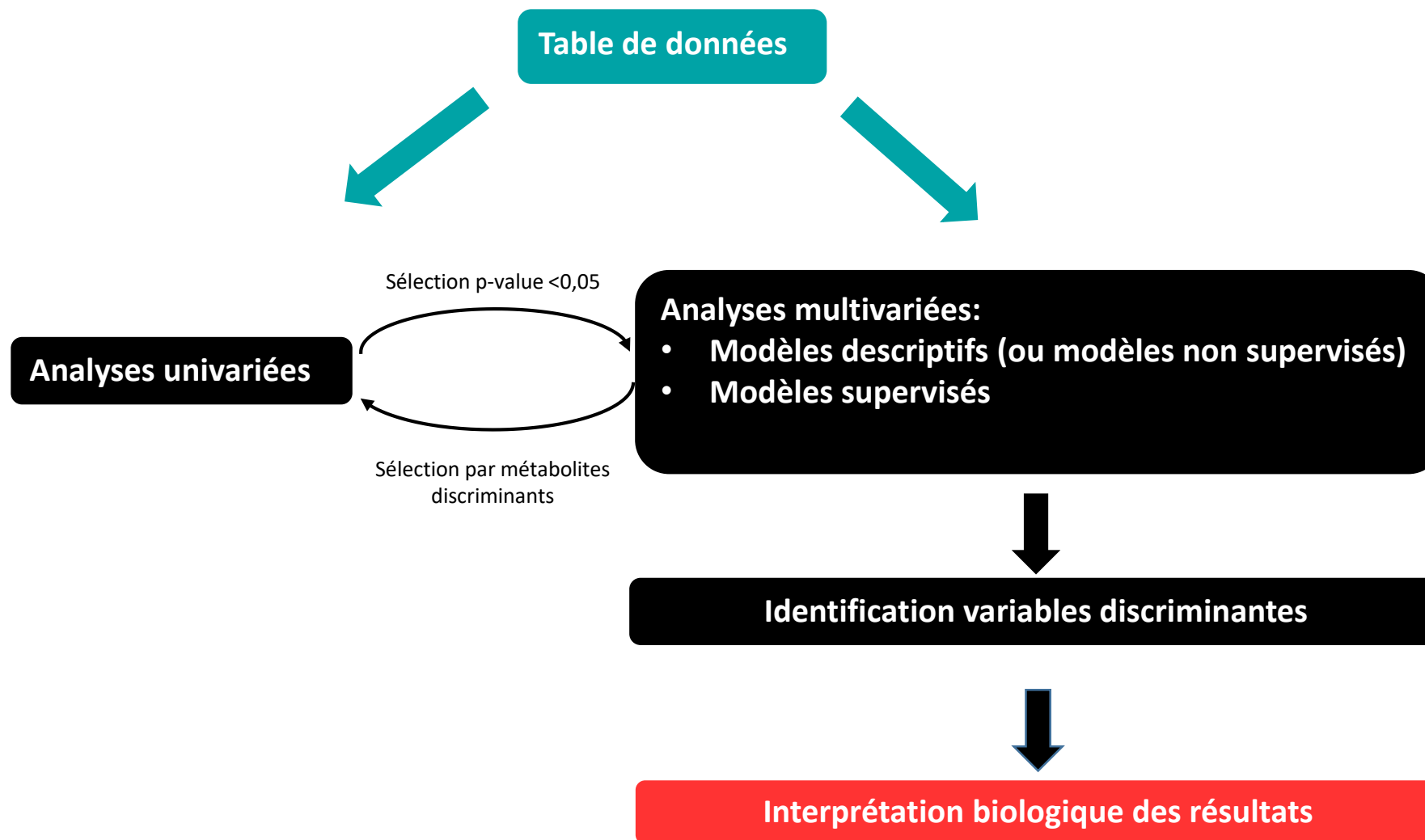
Variable, Descripteur métabolomique

p variables

Données Sacurine



# ➤ Analyses statistiques



## ➤ Analyses statistiques

➤ **Statistiques descriptives:** décrire, résumer, synthétiser l'information contenue dans les données

- Univariées
- Multivariées (modèles non supervisés)

➤ **Statistiques inférentielles ou décisionnelles:** extrapoler à la population entière les propriétés mises en évidence sur l'échantillon

- Univariées
- Multivariées (modèles supervisés)



# Statistiques univariées

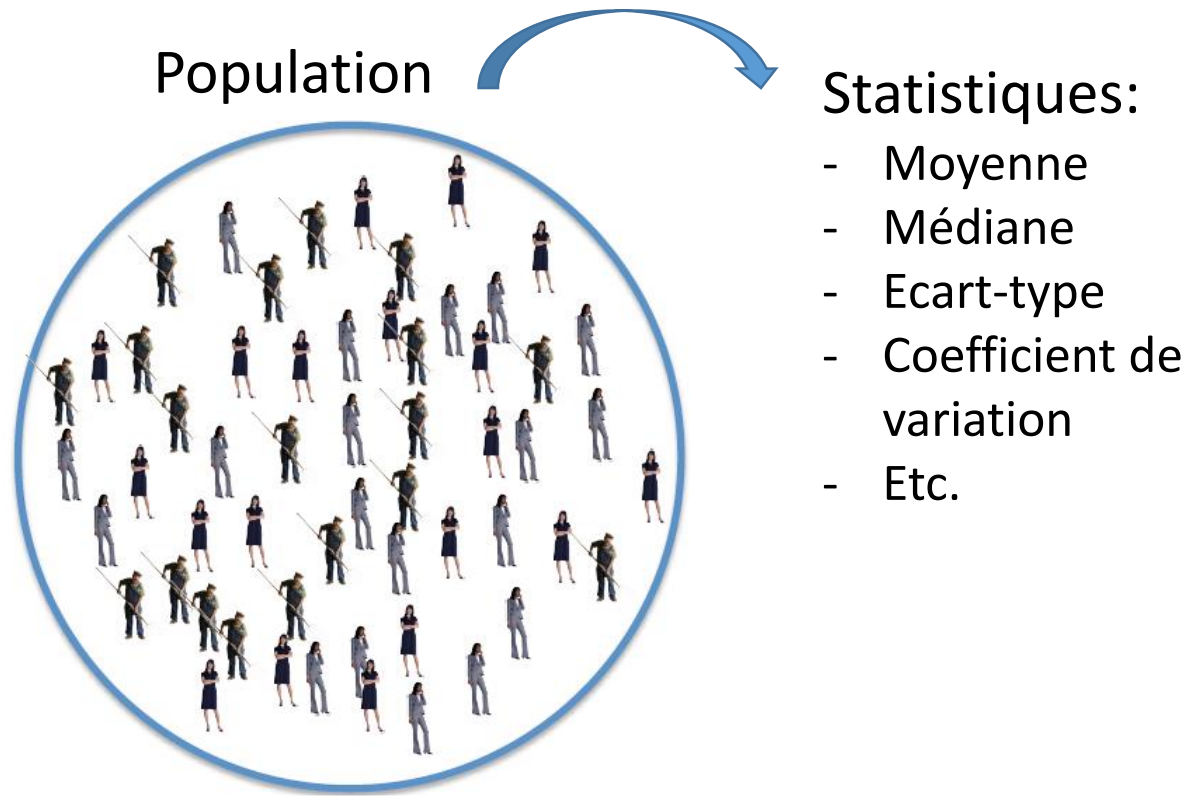


**INRAE**

Design d'une étude métabolomique

04-06-2020 / Production et analyse de données métabolomiques / Binta Diémé

# ➤ Statistiques univariées : statistiques descriptives



## ➤ Statistiques univariées : statistiques descriptives

		Type de variable	
		Variable Catégorielle Ex: Traitement, Sexe...	Variable Quantitative Ex: Teneur en métabolite ou intensité du signal
Classe d'outil	Indice numérique	Tableau de contingence Tableau de fréquence	Moyenne, Médiane Variance, Ecart-type, Erreur-standard, Coefficient de corrélation
	Graphique	Graphe en barres	Graphe en points, Graphe X-Y Graphe à barres d'erreurs, Box-plot, Histogramme, Graphe quantile-normal.

Tiré du MOOC de métabolomique



**INRAE**

Design d'une étude métabolomique

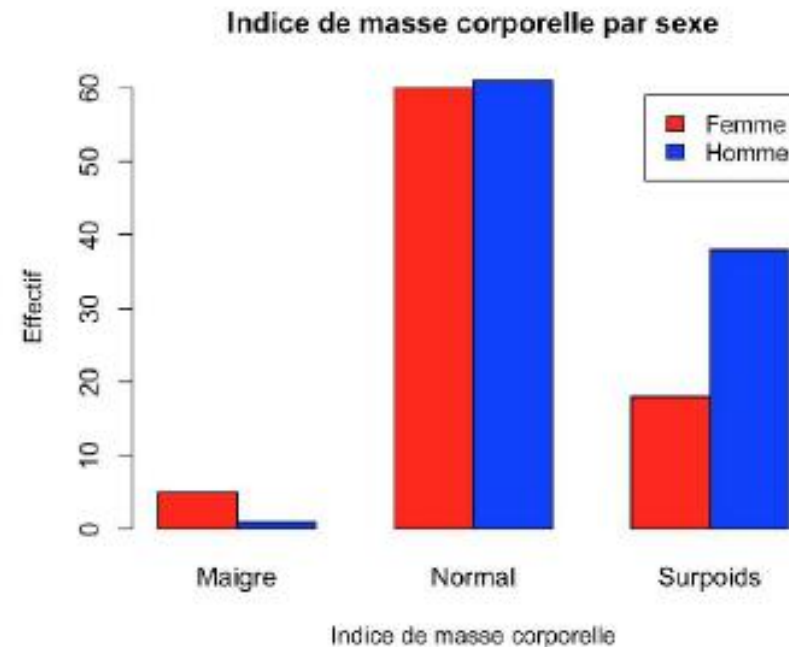
04-06-2020 / Production et analyse de données métabolomiques / Binta Diémé

# ➤ Statistiques univariées : statistiques descriptives

Tableau de fréquence (en %)

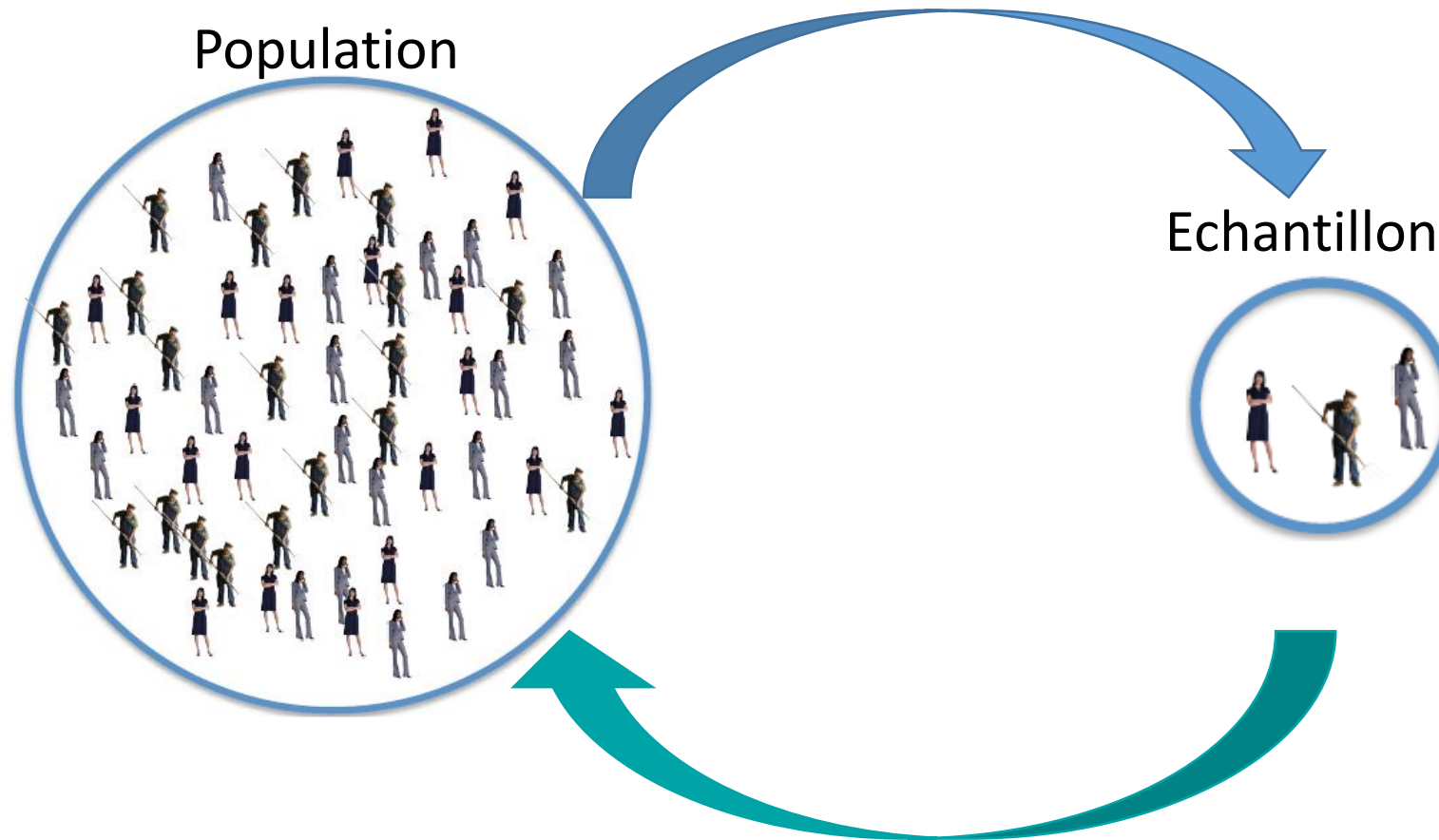
	Maigre	Normal	Surpoids	Total
Femme	2.7%	32.8%	9.8%	45.4%
Homme	0.5%	33.3%	20.8%	54.6%
Total	3.3%	66.1%	30.6%	100%

Graphe en barres



Données Sacurine

# ➤ Statistiques univariées



**Statistiques inférentielles**



**INRAE**

Design d'une étude métabolomique

04-06-2020 / Production et analyse de données métabolomiques / Binta Diémé



## ➤ Statistiques inférentielles ou décisionnelles

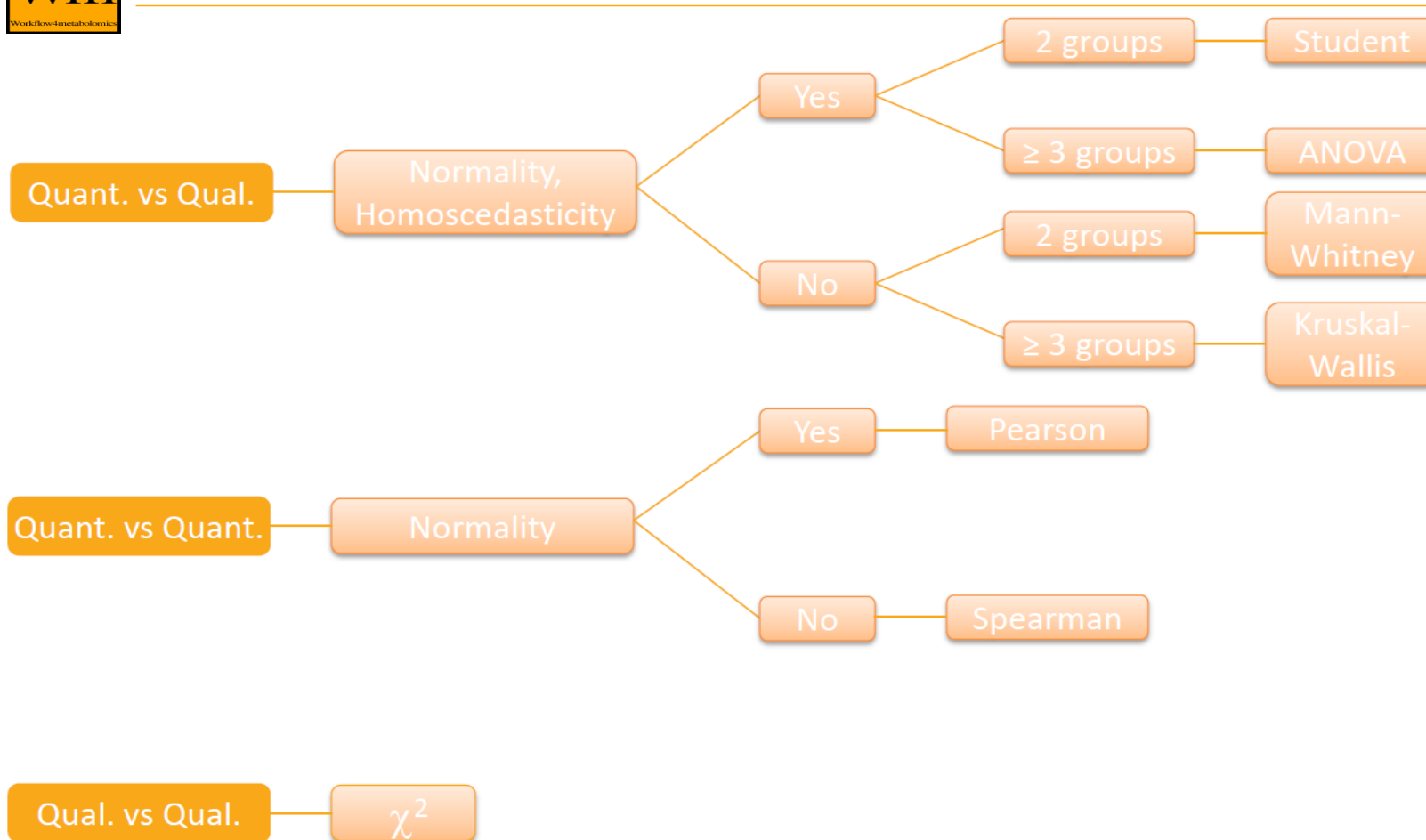
➤ Objectifs: extrapoler à la population entière les propriétés mises en évidence sur l'échantillon

- Tests d'hypothèses
  - La loi de distribution



- Méthodes d'échantillonnage: cruciales pour éviter tout biais

# Statistiques inférentielles ou décisionnelles



# Statistiques multivariées



**INRAE**

Design d'une étude métabolomique

04-06-2020 / Production et analyse de données métabolomiques / Binta Diémé

# ➤ Tableau de données métabolomiques

Label	MTB 1	MTB 2	MTB 3	MTB 4	MTB 5	MTB 6	MTB 7	MTB 8	MTB 9	MTB 10	MTB 11	MTB 12	MTB 13	MTB 14
HU_011	3.02	3.888	3.869	3.717	3.535	3.325	4.205	4.08	4.56	3.467	4.298	3.8	3.503	4.673
HU_014	3.814	4.277	3.838	3.777	3.933	4.025	5.182	4.359	3.158	3.943	3.834	4.192	4.37	4.49
HU_015	3.52	4.196	4.102	4.292	3.955	3.973	3.886	4.249	4.291	3.54	4.199	4.147	3.991	4.416
HU_017	2.562	4.324	4.539	4.432	4.228	4.11	4.238	4.231	3.645	3.152	3.865	4.233	4.151	3.961
HU_018	3.782	4.629	4.179	4.117	4.006	4.024	1.853	4.324	4.096	4.49	4.271	4.755	4.373	4.248
HU_019	4.161	4.412	4.517	4.562	4.321	4.327	4.081	4.244	4.428	4.157	3.832	4.334	1.446	4.164
HU_020	3.199	4.175	4.18	4.253	4.014	4.198	4.379	4.124	3.885	3.92	3.818	3.891	4.162	4.552
HU_021	3.378	4.635	4.036	3.938	3.667	3.997	3.954	4.262	2.298	3.824	4.117	4.535	4.176	3.971
HU_022	3.265	3.847	4.295	4.071	4.57	4.459	4.258	4.163	4.705	4.177	4.01	4.367	4.333	4.509
HU_023	3.787	3.872	4.398	4.469	4.686	4.636	3.482	4.143	3.659	3.384	3.714	4.381	3.604	3.902
HU_024	5.364	4.418	4.53	4.498	4.442	4.421	4.359	4.287	3.656	4.059	4.115	4.108	4.065	4.31
HU_025	3.202	4.329	4.11	4.125	4.009	3.945	4.191	4.18	4.169	4.361	3.962	4.23	4.007	3.891
HU_026	3.98	4.03	3.74	4.159	3.199	4.255	4.207	4.159	3.712	3.207	3.957	4.064	1.462	4.474
HU_027	3.02	4.152	3.784	3.748	3.703	4.132	4.007	4.092	3.905	3.89	3.708	4.271	4.138	4.723

Echantillon, Individu

Identifiant

Variable, Descripteur métabolomique

p variables

n observations

Données Sacurine



## ➤ Statistiques multivariées descriptives ou non supervisées

### Objectifs:

- résumer et visualiser l'information pertinente contenue dans un grand tableau de données
- Mettre en évidence des liens potentiels entre les variables
- Identifier de regroupements d'individus similaires
- Mettre en évidence des individus atypiques



## ➤ Statistiques multivariées descriptives ou non supervisées

Méthodes descriptives	Variables quantitatives	Variables qualitatives/mixtes
Les méthodes factorielles: Visualiser les corrélations multiples entre les variables et/ou les individus	Analyses en composantes principales (ACP)	Analyses factorielles des correspondances (AFC, AFCM)
Les méthodes de classification: réduire le nombre d'individus en groupes homogènes	Classification hiérarchique ascendante	AFC ou AFCM et classification

## ➤ Statistiques multivariées descriptives

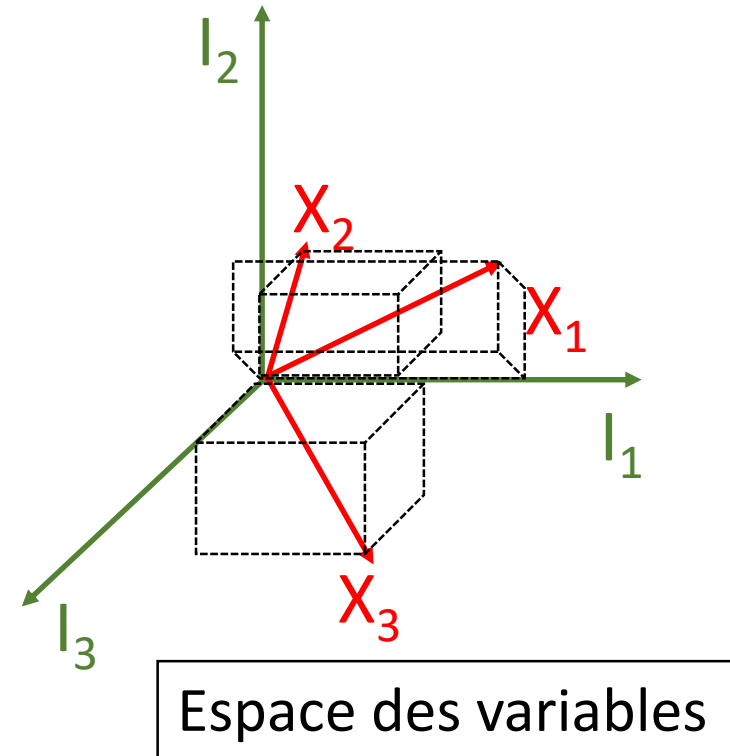
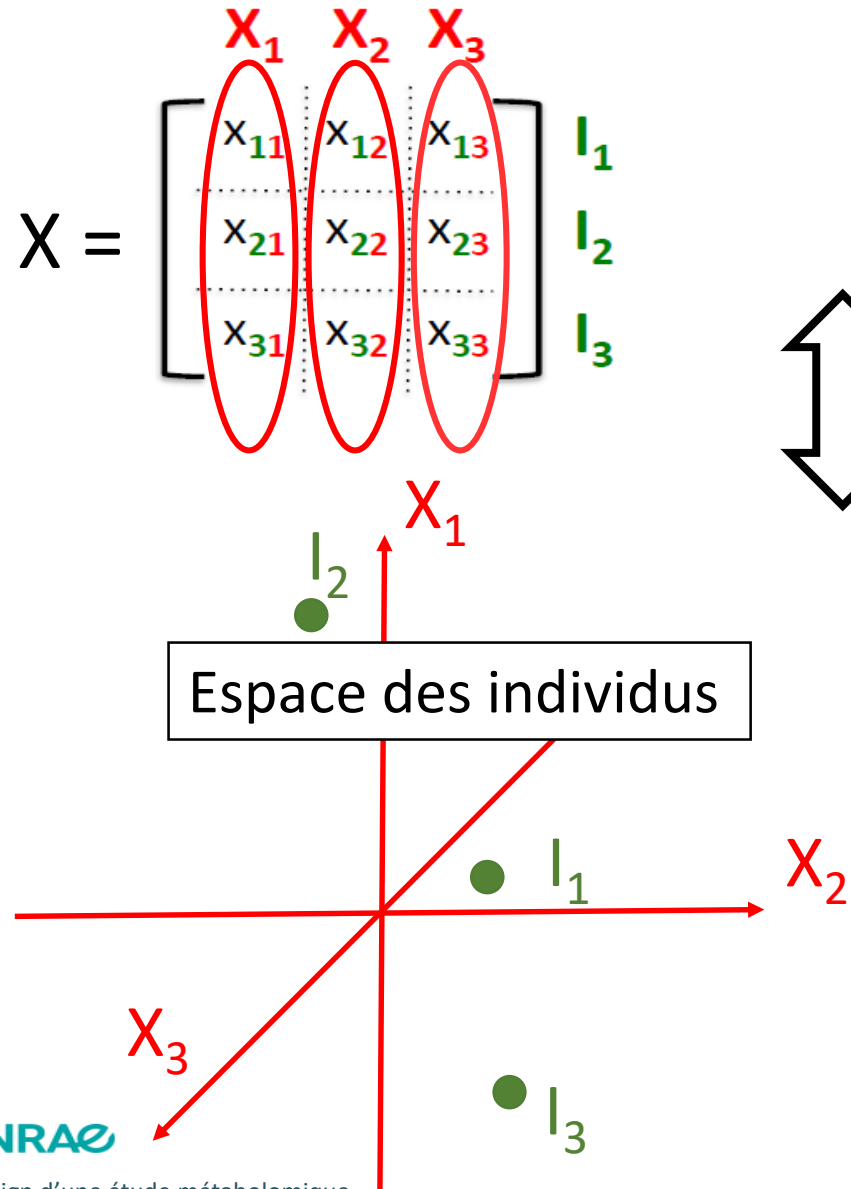
L'analyse en composante principale (ACP):

➤ réduire la dimensionnalité en perdant le moins d'information possible

- les composantes principales sont des combinaisons linéaires des variables d'origines centrées
- Projection du nuage de points sur une surface plane



# ➤ Statistiques multivariées descriptives : ACP

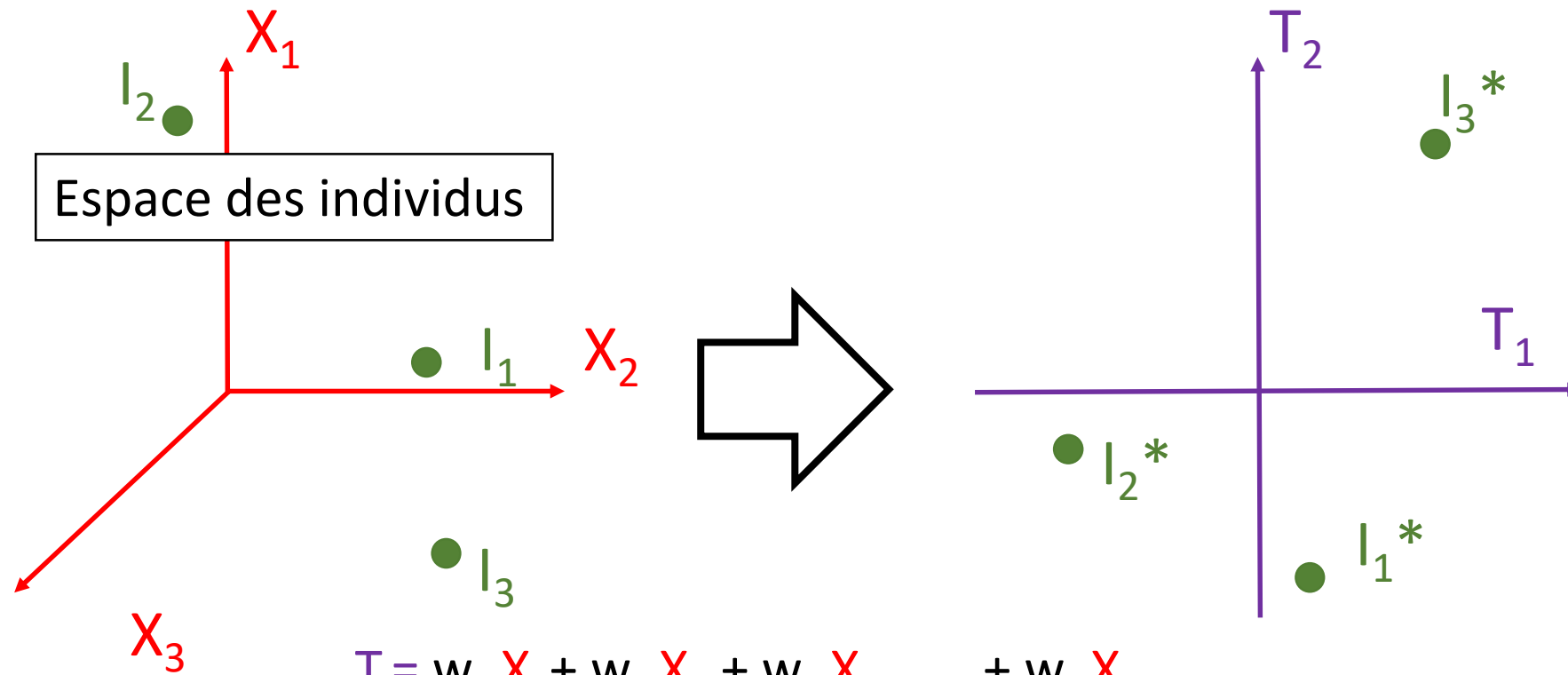




# ➤ Analyse en Composantes Principales

➤ Composante principale = somme pondérée des variables d'origine centrée

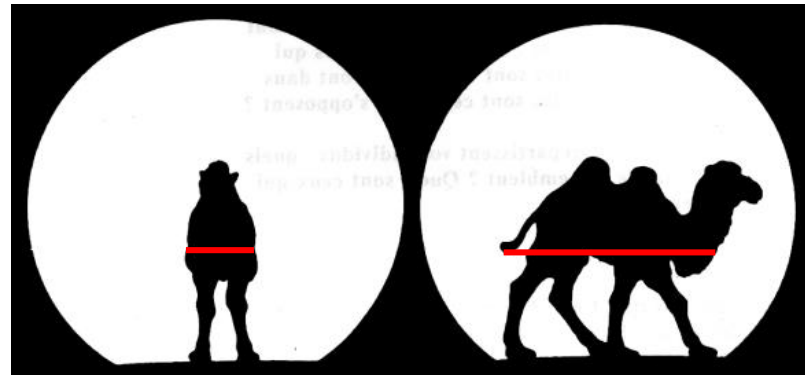
➤ Un poids est associé à chaque variable, ce poids peut être positif ou négatif



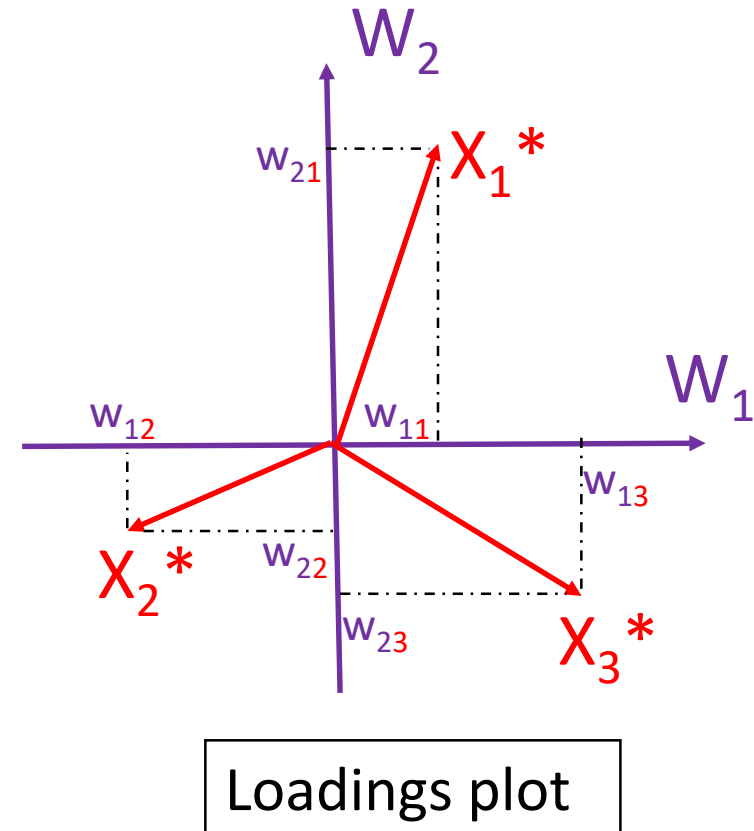
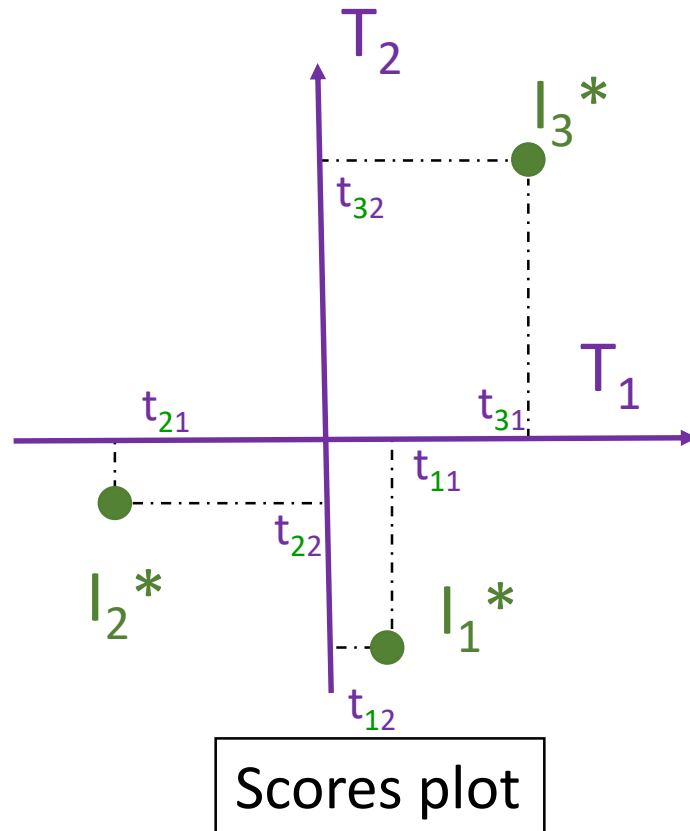
$$T_i = w_{i1}X_1 + w_{i2}X_2 + w_{i3}X_3 \dots + w_{ik}X_k$$

## ➤ Statistiques multivariées descriptives: ACP

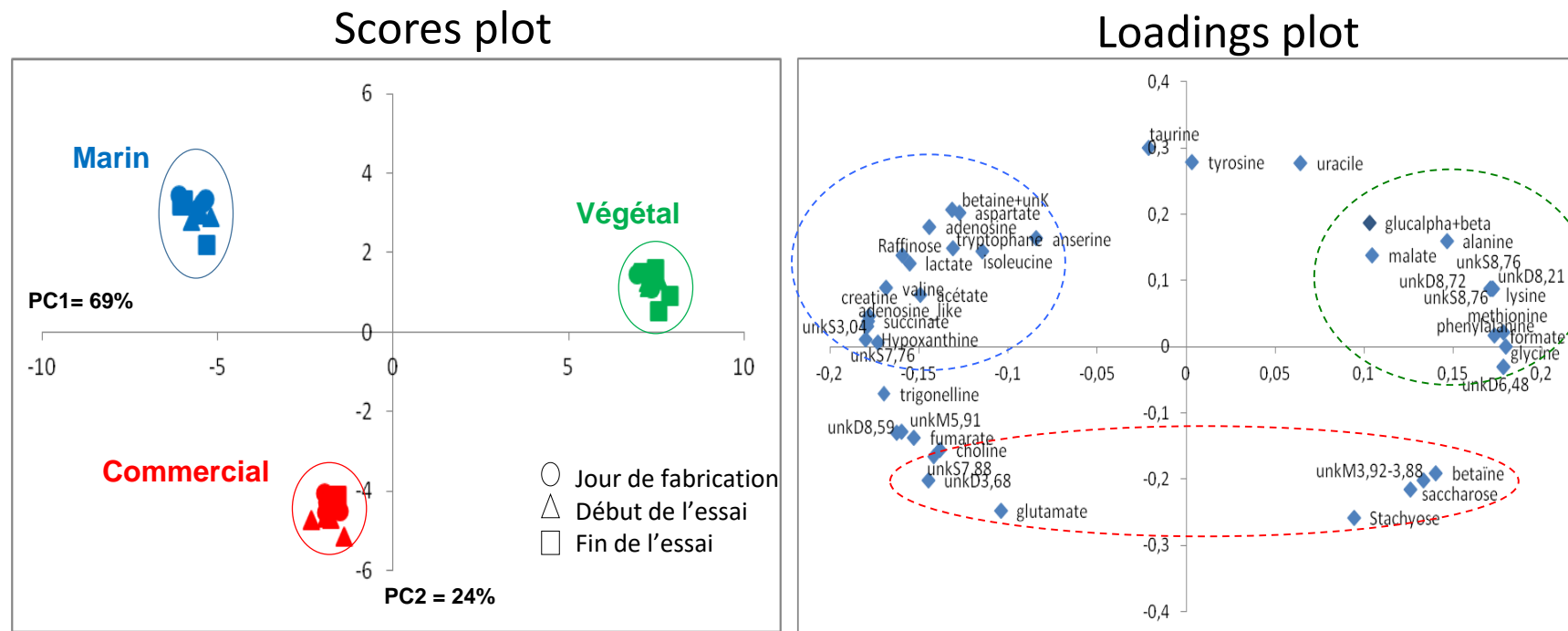
- **D'un point de vue géométrique** : construction d'une série d'axes orthogonaux permettant de projeter le maximum de variabilité des données sur un plan à 2 dimensions. (techniquement : maximise la variance projetée sur les axes)



# ➤ Analyse en Composantes Principales



# ➤ ACP sur les concentrations des métabolites d'extraits polaires d'aliments



45 variables ( Concentrations des métabolites), 36 échantillons, Normalisation centrée réduite

Stage Master 2 Blandine Madji Hounoum

- Discrimination des trois aliments (97 % de la variabilité totale)
- Mise en évidence des métabolites majeurs discriminants les aliments
- Effet spécifique de chaque aliment sur le métabolisme des truites?

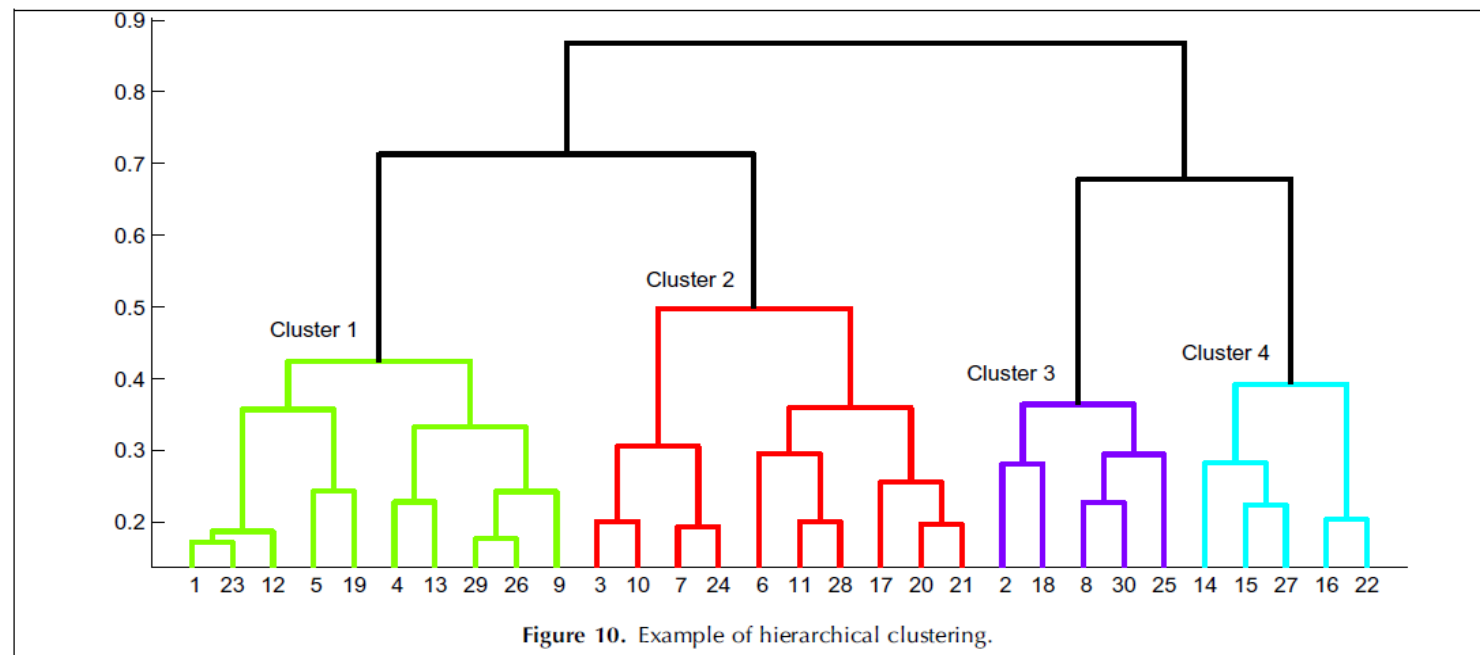


# ➤ Statistiques multivariées descriptives: classification ascendante hiérarchique

## ➤ Classification hiérarchique ascendante

- Mesure de distance ou dissimilarité entre les individus

## ➤ Objectifs: regrouper les individus en groupes homogènes



Kristian Hovde Liland, *Trends in Analytical Chemistry*, Vol. 30, No. 6, 2011

## ➤ Statistiques multivariées supervisées

➤ Objectif: expliquer la variable Y à partir des variables explicatives X

$$X = \begin{array}{c} \begin{array}{ccc} X_1 & X_2 & X_3 \end{array} \\ \left[ \begin{array}{ccc} X_{11} & X_{12} & X_{13} \\ X_{21} & X_{22} & X_{23} \\ X_{31} & X_{32} & X_{33} \end{array} \right] \begin{array}{l} l_1 \\ l_2 \\ l_3 \end{array} \end{array}$$

Variables explicatives

$$Y = \begin{bmatrix} Y_{11} \\ Y_{21} \\ Y_{31} \end{bmatrix}$$

$$\boxed{f(X) = Y}$$

# ➤ Statistiques multivariées supervisées

## Partial Least Squares (PLS):

- réduire la multidimensionnalité en perdant le moins d'information possible (comme pour l'ACP)
  - Variables latentes: maximisent la covariance entre X et Y
  - Projection des nuages de points sur une surface plane

# ➤ Différents types d'études métabolomiques

## ➤ Études méthodologiques:

- Optimisation du protocole de préparation des échantillons
- Optimisation de(s) méthode(s) statistique(s)

## ➤ Études non ciblées

- Détection du maximum de composés afin de :
  - Rechercher un ou plusieurs biomarqueurs (modèles prédictifs)
  - Rechercher des voies métaboliques perturbées (modèles explicatifs)

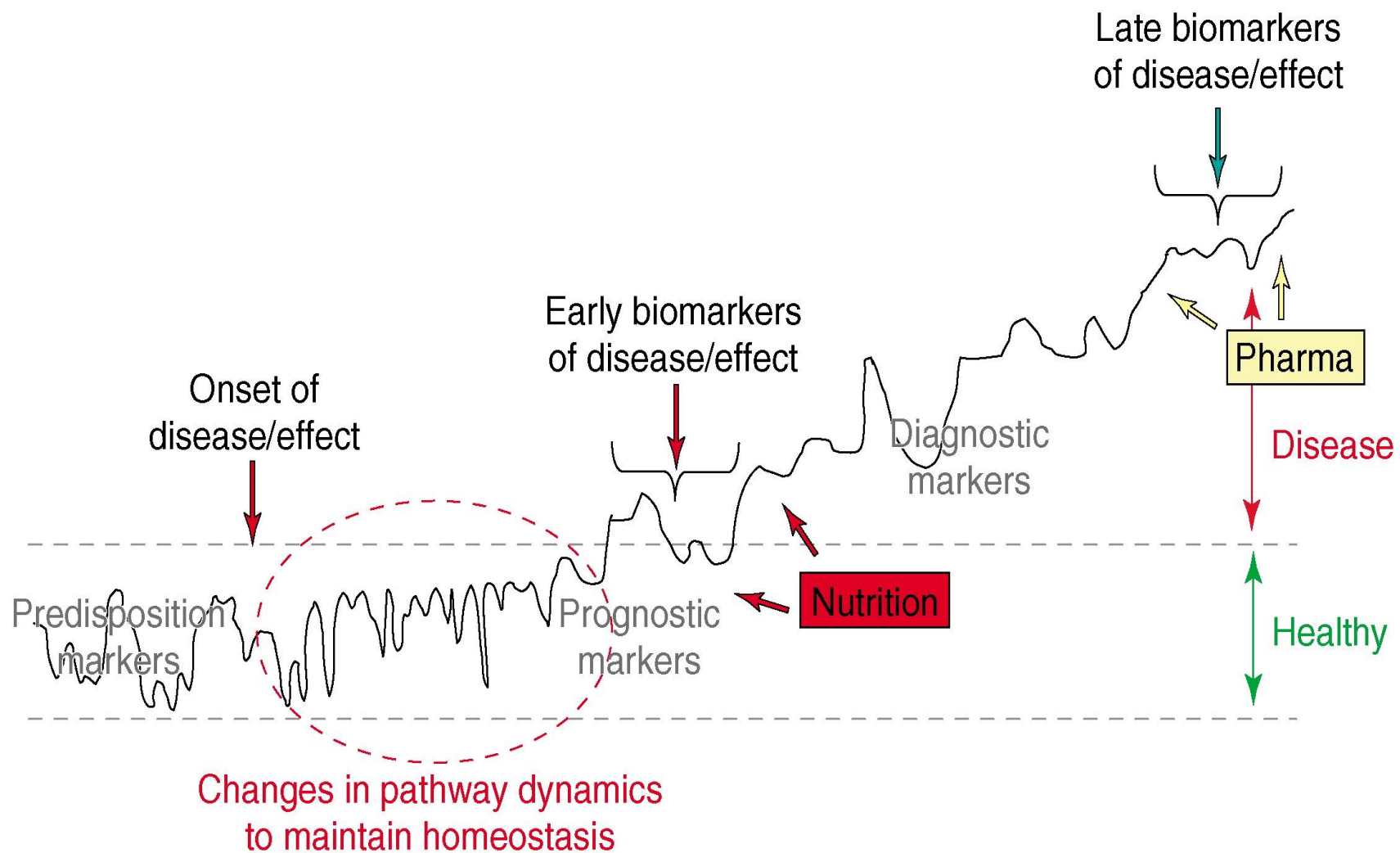
## ➤ Études ciblées:

- Recherche spécifique de composés appartenant à une ou plusieurs voies métaboliques
- Quantification de composés biomarqueurs

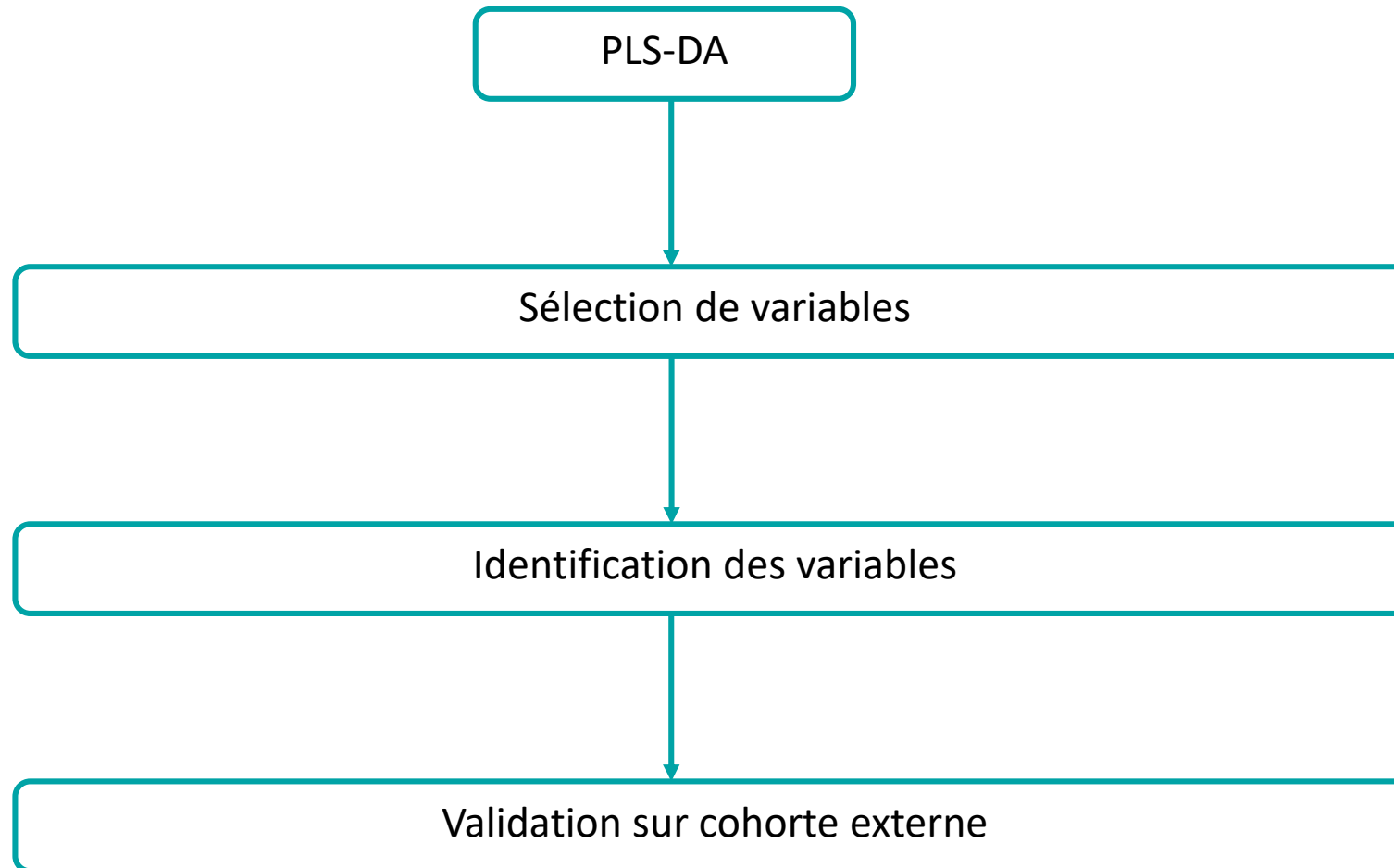




# ➤ Différents types de biomarqueurs



## ➤ Modèles prédictifs: recherche de biomarqueurs



## ➤ Exemple : recherche de biomarqueurs précoces de la boiterie chez les vaches

Boiterie chez les vaches:

- Forte prévalence chez les vaches laitières
- Souffrance animale
- Pas de symptômes avant l'apparition de la maladie

**Objectifs:** mettre en évidence des biomarqueurs précoces avant l'apparition de la boiterie chez les vaches

Protocole d'expérience:

- Prélèvement d'échantillons de sang de la vache à différents temps pendant la gestation et après la mise bas.

Zhang et al., *Animals* (2015), 5, 717-747.

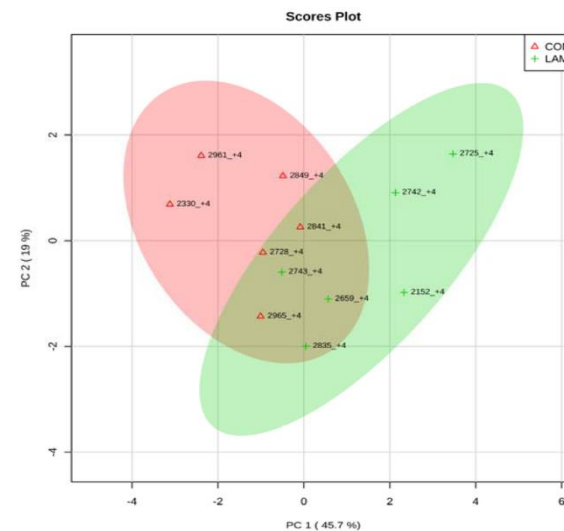
# ➤ Exemple : recherche de biomarqueurs précoces de la boiterie chez les vaches

Construction d'une série d'axes (composantes principales) permettant de maximiser la variance entre les individus **ACP**

➔ Visualisation du lien entre les variables quantitatives X et les variables qualitatives (groupes).

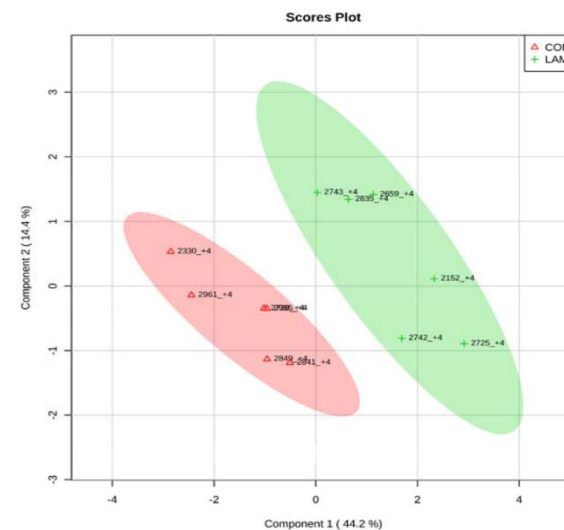
➔ Classification des variables (score VIP) par ordre décroissant

A



B

**PLS-DA**



Zhang et al., *Animals*, (2015), 5, 717-747.



**INRAE**

Design d'une étude métabolomique

04-06-2020 / Production et analyse de données métabolomiques / Binta Diémé

# ➤ Différents types d'études métabolomiques

## ➤ Études méthodologiques:

- Optimisation du protocole de préparation des échantillons
- Optimisation de(s) méthodes statistique(s)

## ➤ Études non ciblées

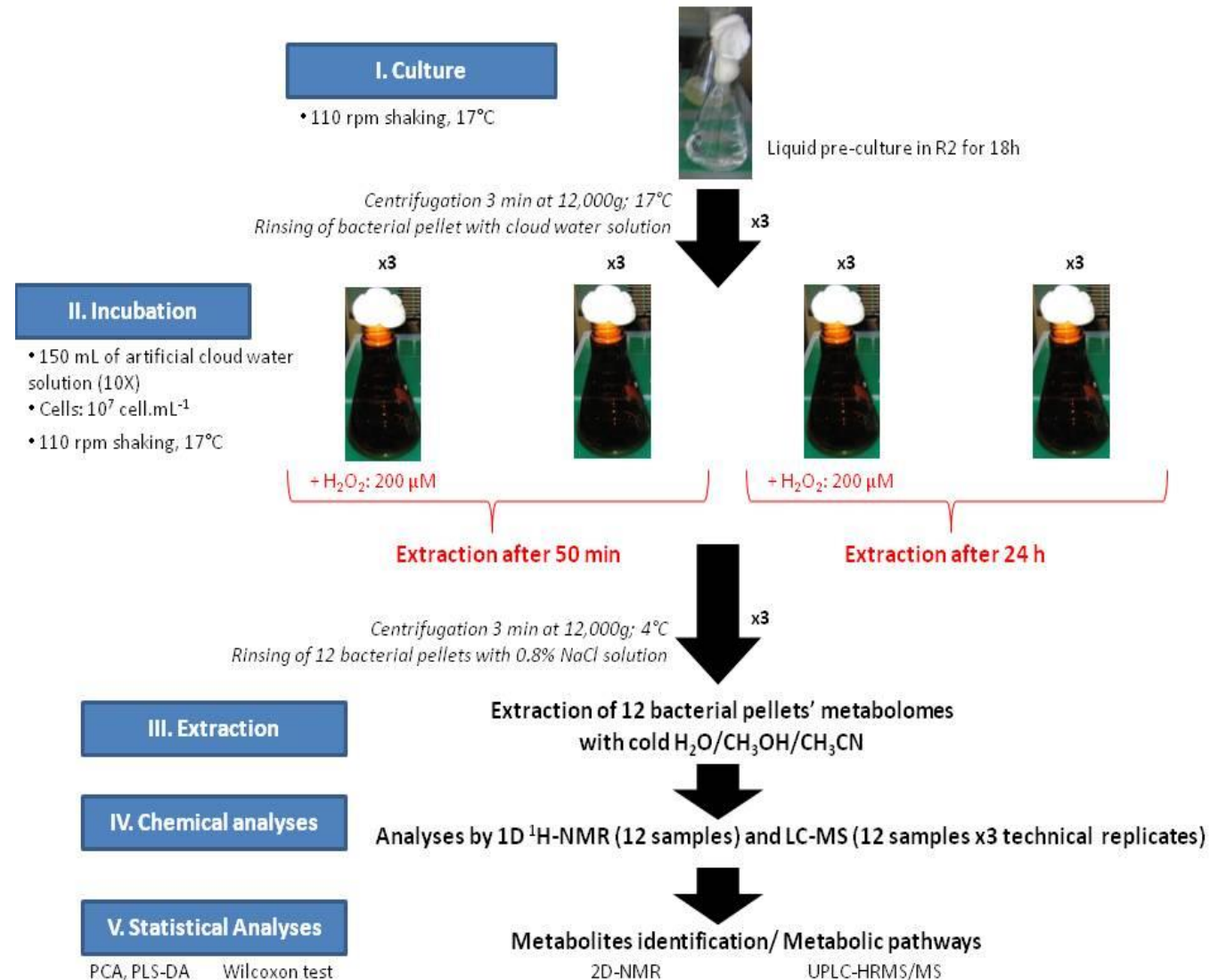
- Détection du maximum de composés afin de :
  - Rechercher un ou plusieurs biomarqueurs (modèles prédictifs)
  - Rechercher des voies métaboliques perturbées (modèles explicatifs)

## ➤ Études ciblées:

- Recherche spécifique de composés appartenant à une ou plusieurs voies métaboliques
- Quantification de composés biomarqueurs

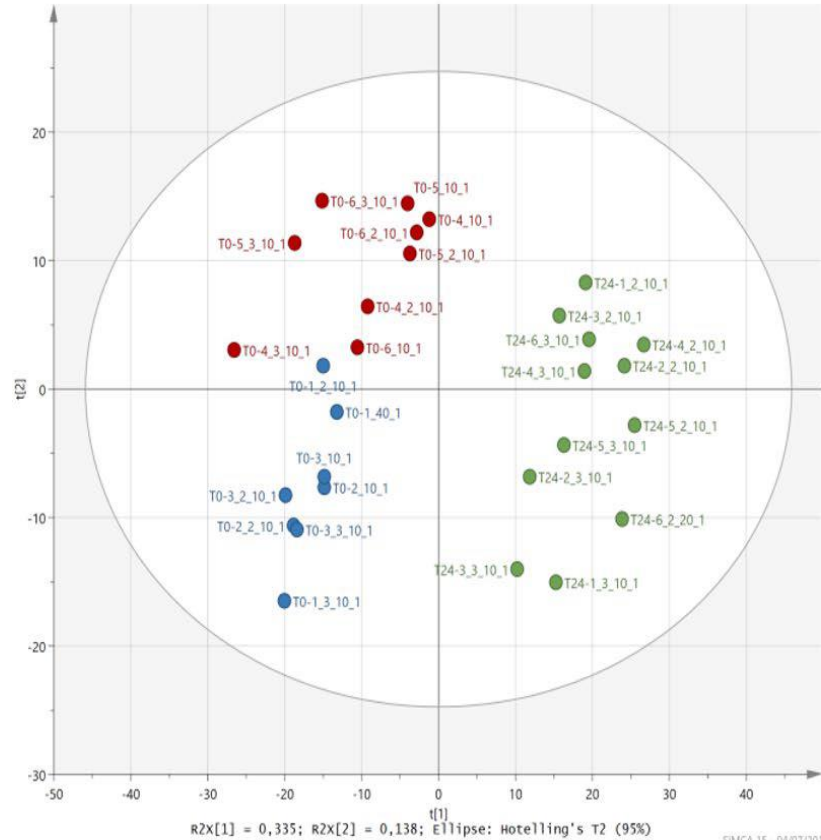


# Adaptations métaboliques de *Pseudomonas graminis* au stress H<sub>2</sub>O<sub>2</sub> dans l'eau de nuage



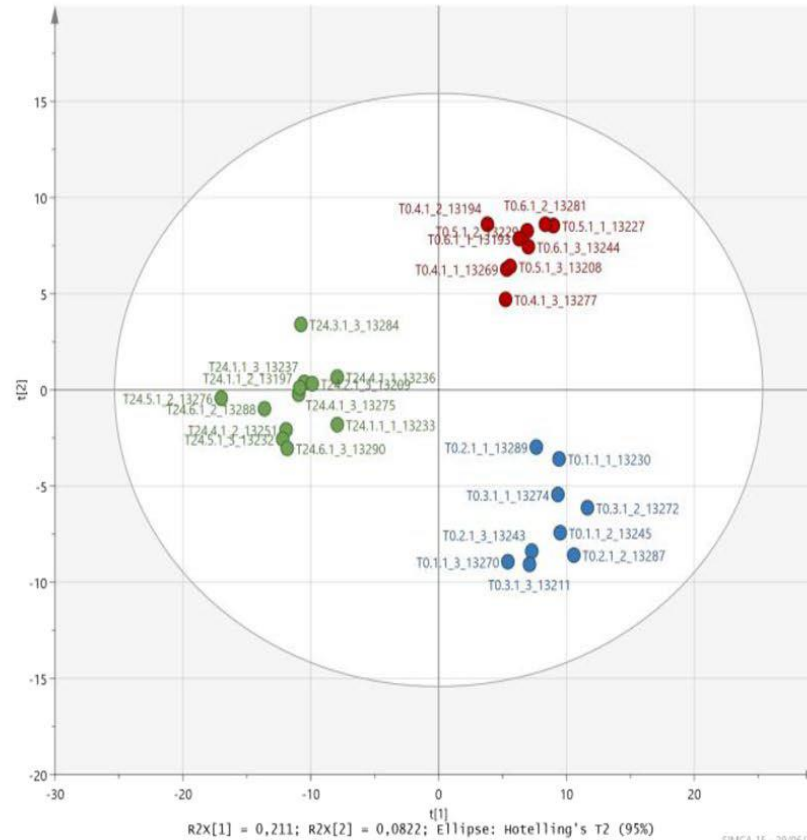
# ➤ Adaptations métaboliques de *Pseudomonas graminis* au stress H<sub>2</sub>O<sub>2</sub> dans l'eau de nuage

(a)



a) RMN

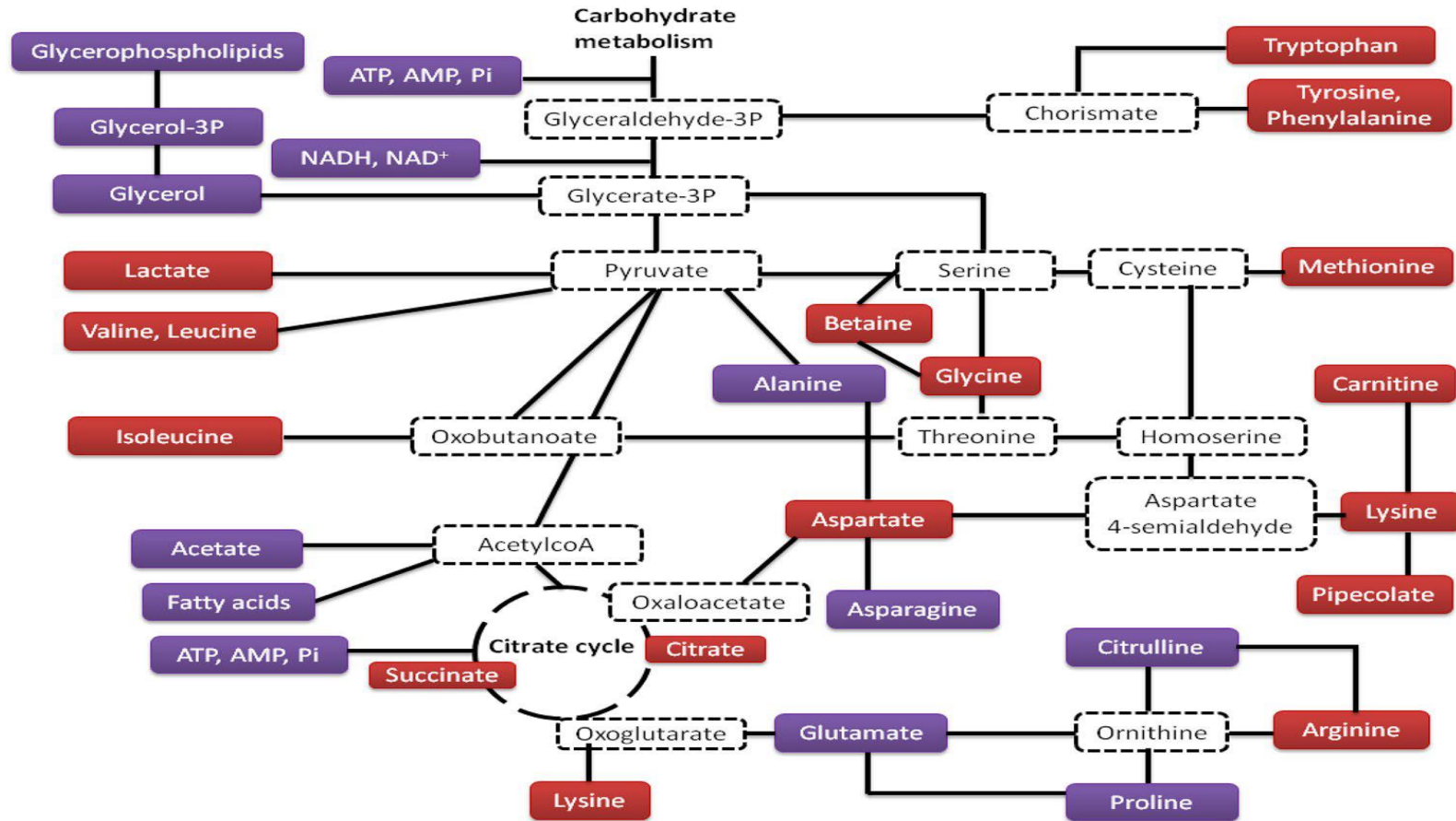
(b)



b) LC-MS mode +



# ➤ Adaptations métaboliques de *Pseudomonas graminis* au stress H<sub>2</sub>O<sub>2</sub> dans l'eau de nuage

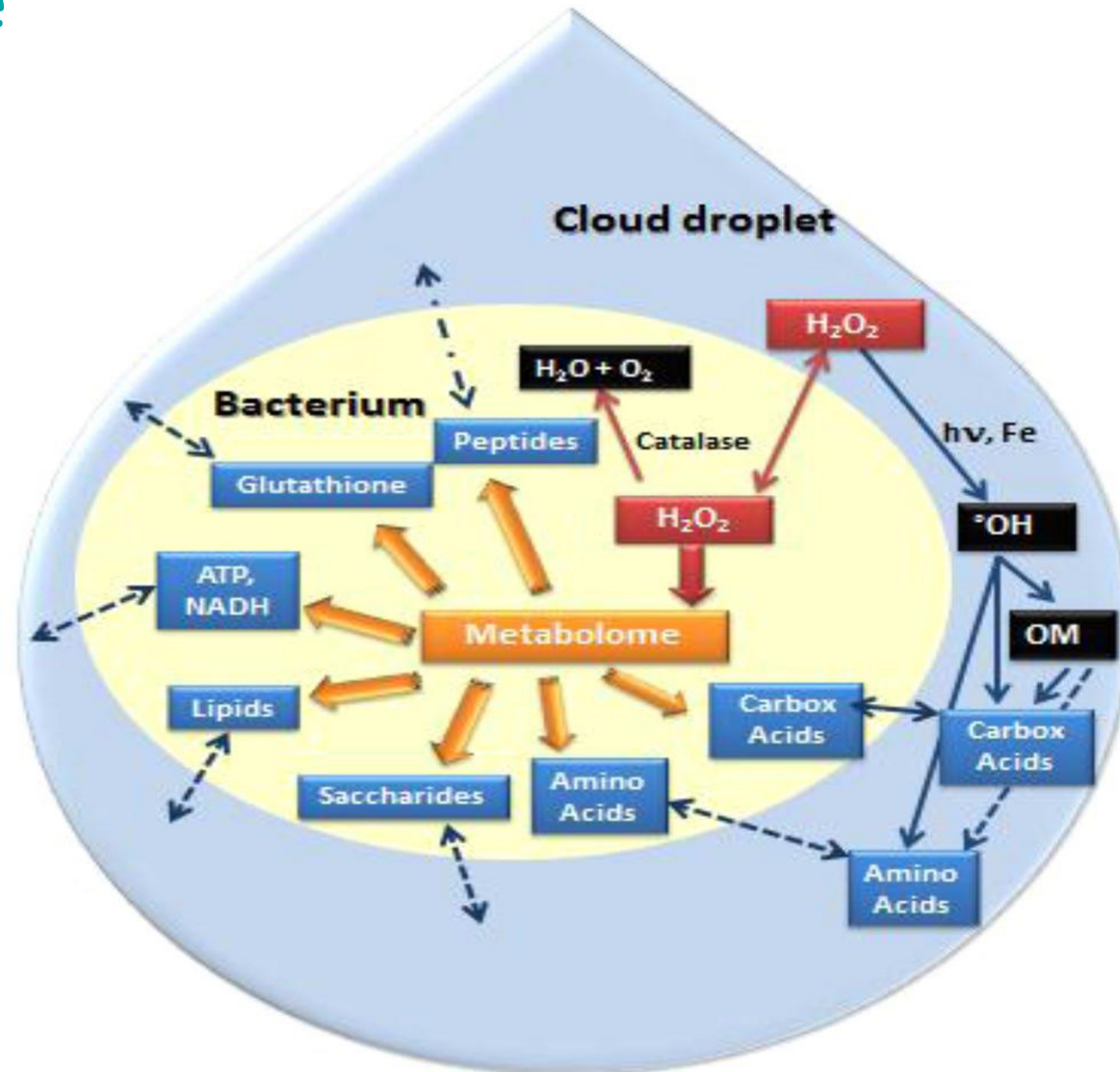


Ile/Leu-His; Ile/Leu-Asp; Val-Ile/Leu; Val-Asp; [Gln-Asp-Thr-Pro]  
Ala-Ala; Val-Ser; [Arg-Cys-Ser-Trp];

Wirgot et al., *Scientific Reports* (2019), 9, 12799



# ➤ Adaptations métaboliques de *Pseudomonas graminis* au stress $H_2O_2$ dans l'eau de nuage



## ➤ Modèles explicatifs: réseaux métaboliques

### ➤ Plusieurs logiciels de mise en réseaux des métabolites

- MetExplore: plusieurs organismes dans BDD
- MetaboAnalyst: comparaison de différence de profil d'expression mais peu d'organismes disponibles pour mapper
- KEGG
- iPath 3 .....



# ➤ Différents types d'études métabolomiques

## ➤ Études méthodologiques:

- Optimisation du protocole de préparation des échantillons
- Optimisation de(s) méthodes statistique(s)

## ➤ Études non ciblées

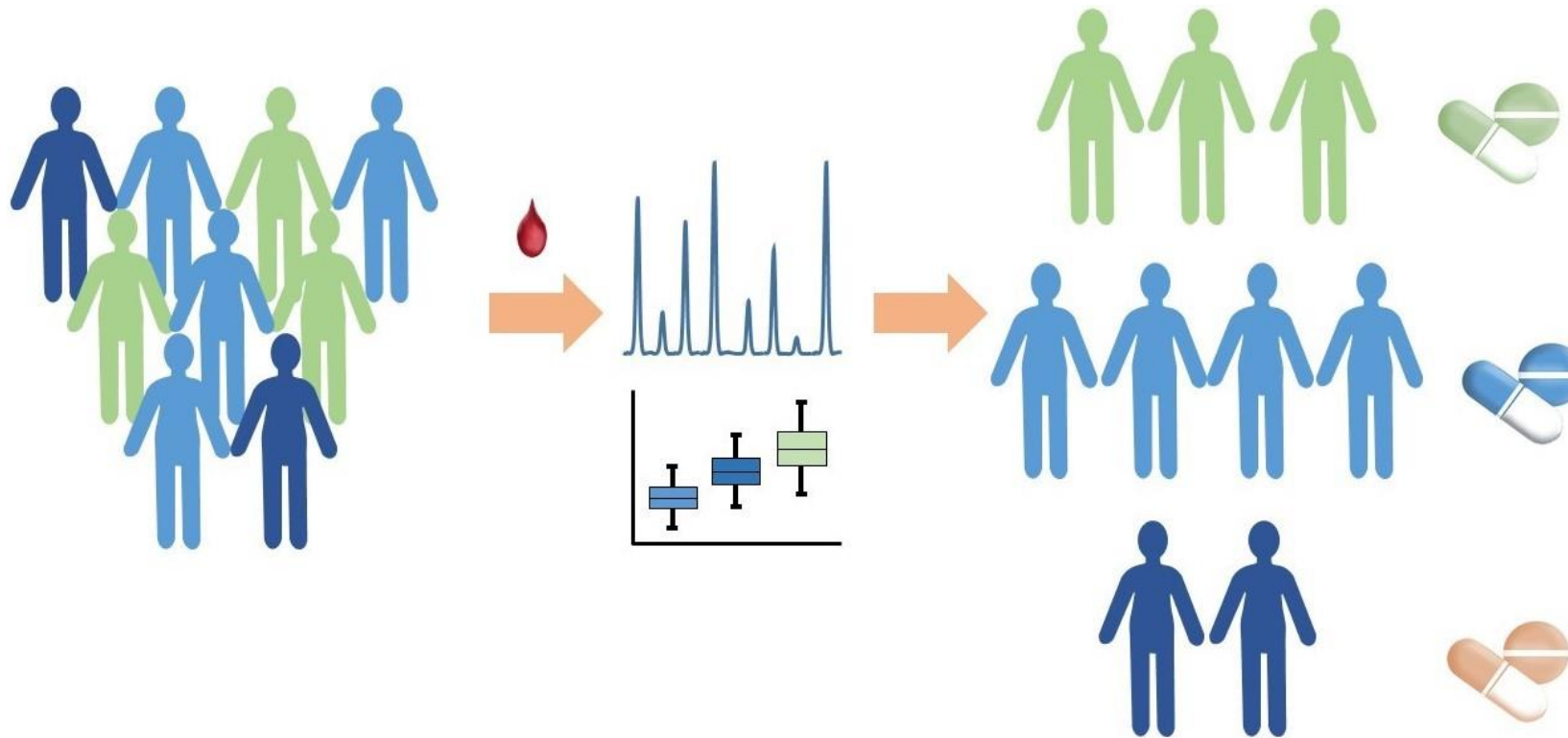
- Détection du maximum de composés afin de :
  - Rechercher un ou plusieurs biomarqueurs (modèles prédictifs)
  - Rechercher des voies métaboliques perturbées (modèles explicatifs)

## ➤ Études ciblées:

- Recherche spécifique de composés appartenant à une ou plusieurs voies métaboliques
- Quantification de composés biomarqueurs



# ➤ Etude clinique ciblée de métabolites pour la stratification de patients diabétiques



Ahonen et al., *Metabolites*, (2019), 9, 184.

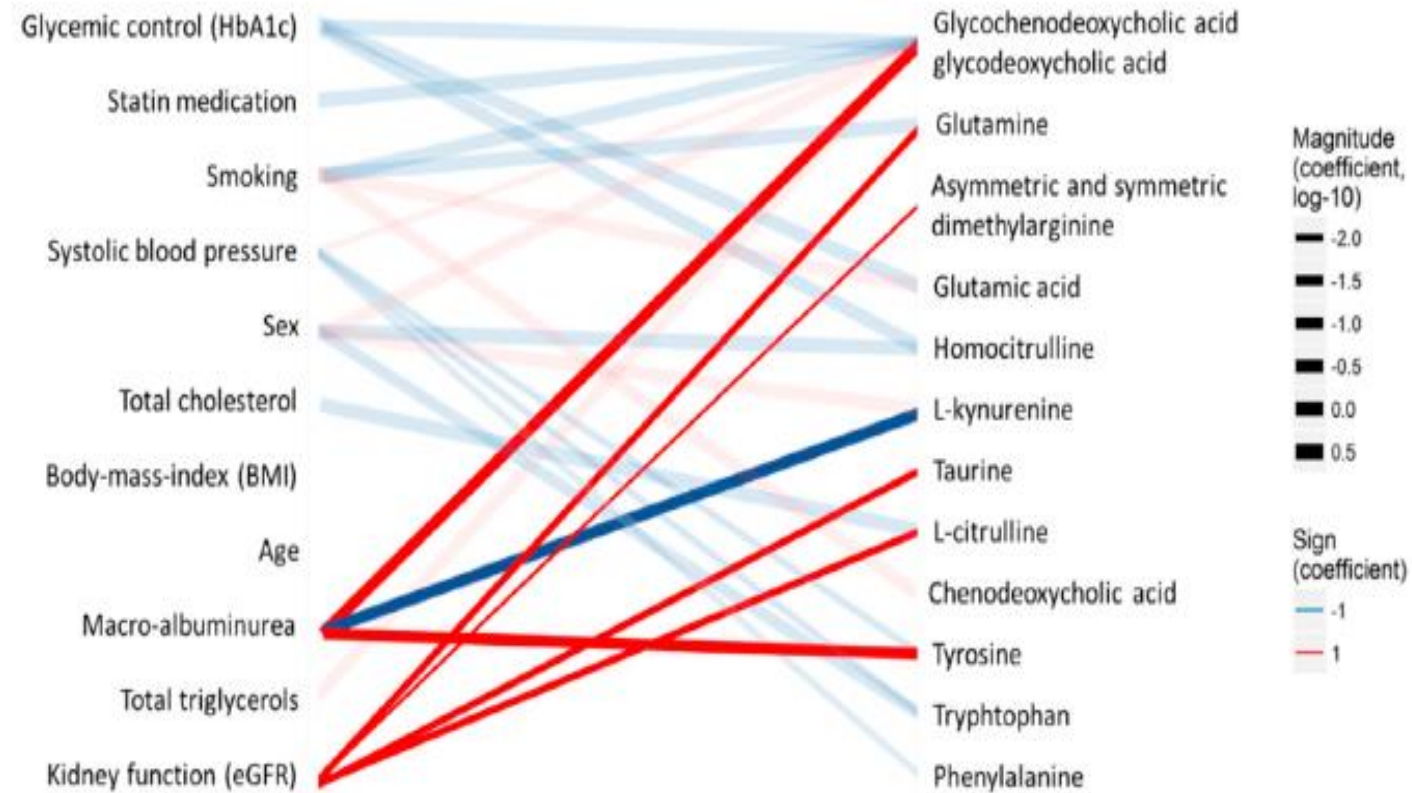


**INRAE**

Design d'une étude métabolomique

04-06-2020 / Production et analyse de données métabolomiques / Binta Diémé

# ➤ Etude clinique ciblée de métabolites pour la stratification de patients diabétiques



Ahonen et al., *Metabolites*, (2019), 9, 184.



INRAE

Design d'une étude métabolomique

04-06-2020 / Production at analyse de données métabolomiques / Binta Diémé

# Omiques intégratives

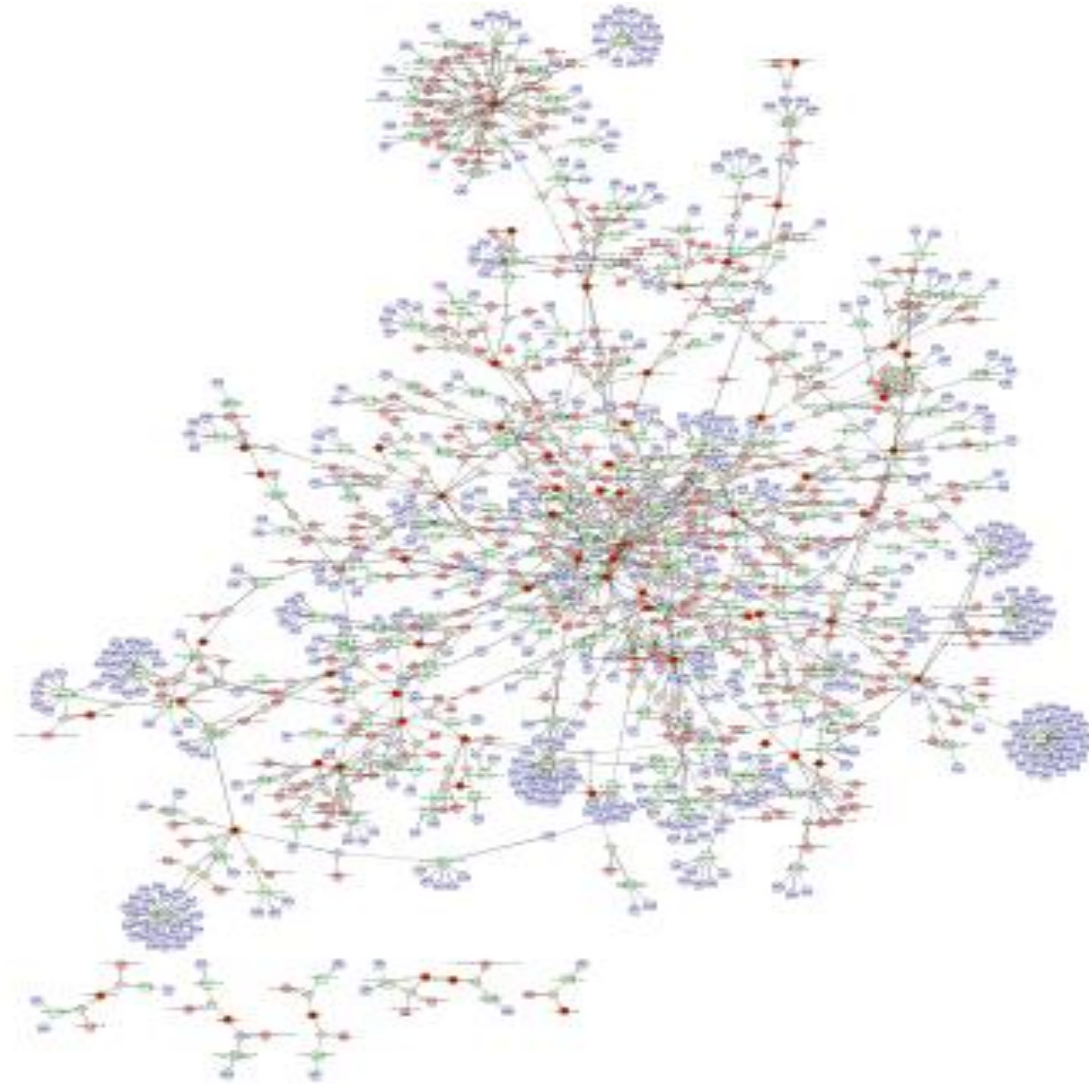


**INRAE**

Design d'une étude métabolomique

04-06-2020 / Production et analyse de données métabolomiques / Binta Diémé

## ➤ Réseaux métabolites-enzymes-gènes



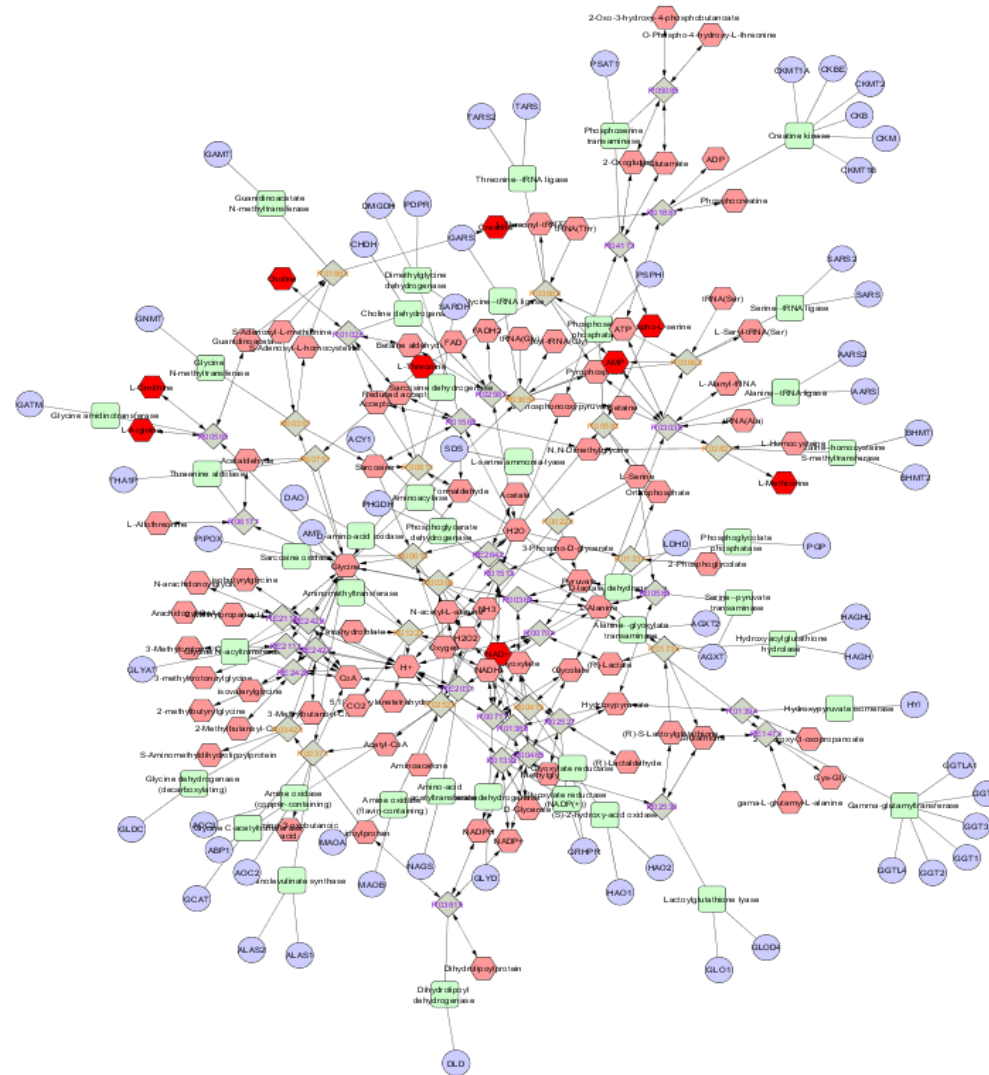
**INRAE**

Design d'une étude métabolomique

04-06-2020 / Production et analyse de données métabolomiques / Binta Diémé



# ➤ Métabolisme de la glycine, sérine, thréonine, alanine



**INRAE**

Design d'une étude métabolomique

04-06-2020 / Production et analyse de données métabolomiques / Binta Diémé



## ➤ Conclusions générales

### Métabolomique:

- Technologie multidisciplinaires
- Applicable a des nombreux domaines de recherche

### Cependant

- Précautions à prendre à toutes les étapes du workflow pour éviter les biais

