

Développements de modèles statistiques pour l'analyse de données métabolomiques longitudinales, application aux photogramules

Camille Guilmineau

11 septembre 2025

Rémi Servien, Marie Tremblay-Franco et Nathalie Vialaneix



► Plan

Introduction

Méthode proposée

Application à l'étude des photogramules

Conclusion et perspectives



INRAE

Séminaire biopuces

11/09/2025 / Camille Guilmineau

► Plan

Introduction

Contexte

État de l'art

Objectifs

Méthode proposée

Application à l'étude des photogramules

Conclusion et perspectives



INRAE

Séminaire biopuces

11/09/2025 / Camille Guilmineau

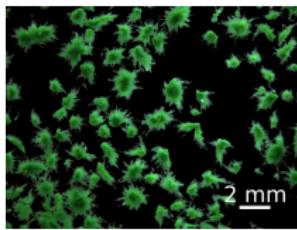
> Contexte

Objectif « zero pollution » du Pacte vert pour l'Europe : besoin d'améliorer le système de traitement des eaux usées

➤ Contexte

Objectif « zero pollution » du Pacte vert pour l'Europe : besoin d'améliorer le système de traitement des eaux usées

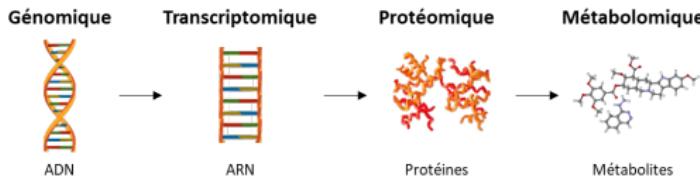
- ▶ Les **photogranules** : agrégats de micro-organismes



⚠ Manque de connaissances sur leur développement et leur cycle de vie

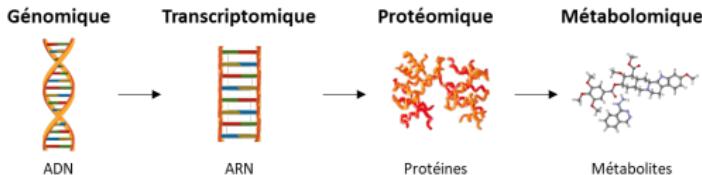
Contexte

Les données omiques :



- décrivent le fonctionnement global d'un organisme

Les données omiques :



- décrivent le fonctionnement global d'un organisme

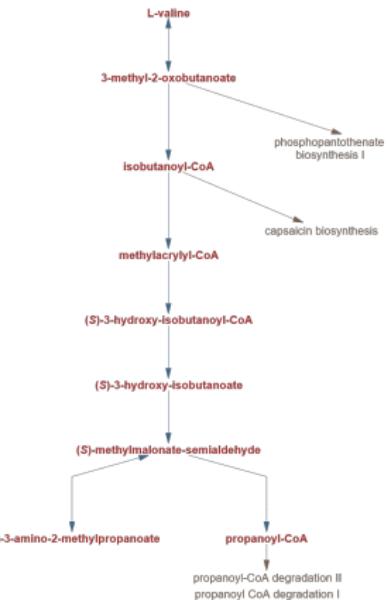
La **métabolomique** : étude des métabolites

- donne des informations fonctionnelles
- image du métabolome à un instant t
 - doit être répété pour suivre l'évolution du métabolome : étude **longitudinale**
- utile pour la recherche de biomarqueurs

Contexte

Les voies métaboliques :

- ▶ séquences de réactions entre métabolites
- ▶ associées à des fonctions biologiques



> Contexte

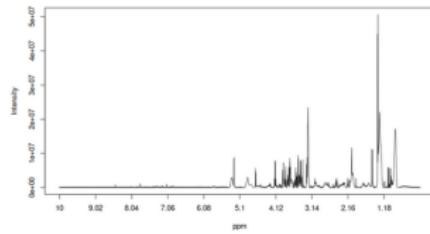
Génération des données métabolomiques :

- ▶ Spectrométrie de masse
- ▶ Résonnance magnétique nucléaire (RMN) :

➤ Contexte

Génération des données métabolomiques :

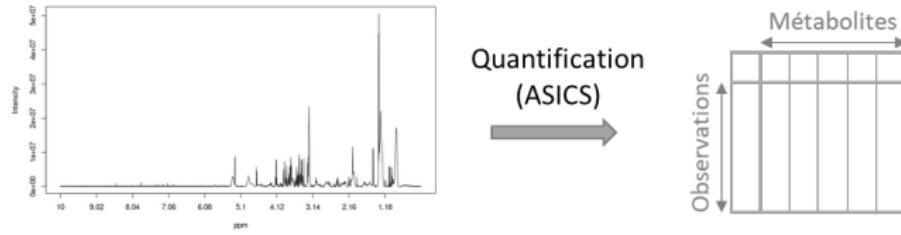
- ▶ Spectrométrie de masse
- ▶ Résonnance magnétique nucléaire (RMN) :
 - ▶ Sous forme de spectre



Contexte

Génération des données métabolomiques :

- ▶ Spectrométrie de masse
- ▶ Résonnance magnétique nucléaire (RMN) :
 - ▶ Sous forme de spectre
 - ▶ Pour obtenir un tableau de données, quantification des métabolites (avec ASICS)



Analyse statistique des données métabolomiques

X : matrice des quantifications de dimension ($n \times m$)

- ▶ n : nombre d'observations, structurées en K conditions
- ▶ m : nombre de métabolites

Analyse différentielle :

Avec μ_j^k la moyenne de la quantification du métabolite j dans la k -ème condition, on teste

$$H_0 : « \mu_j^1 = \mu_j^2 = \dots = \mu_j^K »$$

Correction des tests multiples

Analyse différentielle des données métabolomiques longitudinales

- ▶ Modèle linéaire mixte [Molenberghs and Verbeke, 2000]

$$x_{\cdot j} = \beta_0 1_{nT} + U \beta_j + D \alpha_j + \epsilon_j$$

- β_0 : niveau basal
- U : matrice des effets fixes
- β_j : coefficients des effets fixes
- D : matrice des effets aléatoires
- α_j : coefficients des effets aléatoires
- ϵ_j : résidus

Analyse différentielle des données métabolomiques longitudinales

- ▶ LiMM-PCA [Martin and Govaerts, 2020] : combine modèle mixte et réduction de dimension
 - ▶ ACP sur X , obtention de X^* qui contient m^* composantes principales.
 - ▶ Modèle linéaire mixte :

$$\forall j = 1, \dots, m^*, \quad x_j = U\beta_j + D\alpha_j + \epsilon_j$$

Test de la significativité des effets.

Analyse différentielle des données métabolomiques longitudinales

- ▶ LiMM-PCA [Martin and Govaerts, 2020] : combine modèle mixte et réduction de dimension
 - ▶ ACP sur X , obtention de X^* qui contient m^* composantes principales.
 - ▶ Modèle linéaire mixte :

$$\forall j = 1, \dots, m^*, \quad x_j = U\beta_j + D\alpha_j + \epsilon_j$$

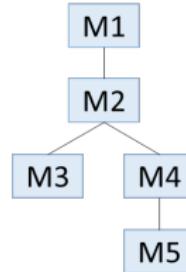
Test de la significativité des effets.

Limite : ne tient pas compte des voies métaboliques.

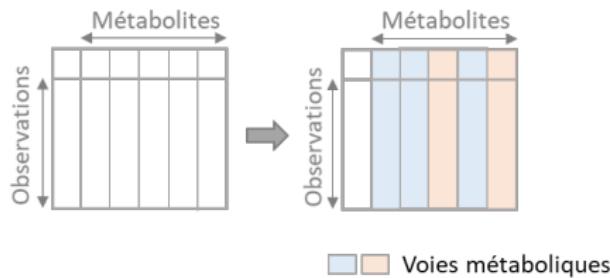
Voies métaboliques

Analyse des voies métaboliques :

- ▶ Faciliter l'interprétation biologique
- ▶ Déetecter plus d'effets



Identification des voies métaboliques avec une base de données.



► État de l'art

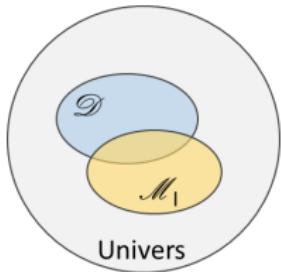
Analyse différentielle des voies métaboliques

- ▶ Analyse d'enrichissement (*over-representation analysis, ORA*)

► État de l'art

Analyse différentielle des voies métaboliques

- ▶ Analyse d'enrichissement (*over-representation analysis*, ORA)



\mathcal{D} : ensemble des métabolites différentiels

\mathcal{M}_l : ensemble des métabolites de la l -ème voie

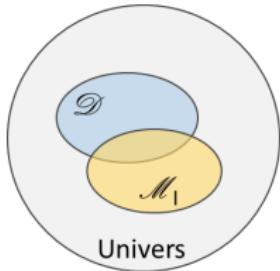
	Métabolites différentiels	Métabolites non différentiels
Métabolites dans la voie \mathcal{M}_l	$ \mathcal{M}_l \cap \mathcal{D} $	$ \mathcal{M}_l \cap \bar{\mathcal{D}} $
Métabolites hors de la voie \mathcal{M}_l	$ \bar{\mathcal{M}}_l \cap \mathcal{D} $	$ \bar{\mathcal{M}}_l \cap \bar{\mathcal{D}} $

L'enrichissement de \mathcal{M}_l est testé avec un test exact de Fisher.

► État de l'art

Analyse différentielle des voies métaboliques

- ▶ Analyse d'enrichissement (*over-representation analysis*, ORA)



\mathcal{D} : ensemble des métabolites différentiels

\mathcal{M}_l : ensemble des métabolites de la l -ème voie

	Métabolites différentiels	Métabolites non différentiels
Métabolites dans la voie \mathcal{M}_l	$ \mathcal{M}_l \cap \mathcal{D} $	$ \mathcal{M}_l \cap \bar{\mathcal{D}} $
Métabolites hors de la voie \mathcal{M}_l	$ \bar{\mathcal{M}}_l \cap \mathcal{D} $	$ \bar{\mathcal{M}}_l \cap \bar{\mathcal{D}} $

L'enrichissement de \mathcal{M}_l est testé avec un test exact de Fisher.

Limite : n'utilise pas la corrélation entre métabolites par voie.

Analyse différentielle des voies métaboliques

- ▶ *global test* [Goeman et al., 2004] : basé sur le modèle linéaire généralisé

$$E(u_i|\xi) = h^{-1}(\kappa + \sum_{j=1}^{m_I} x_{ij}\xi_j)$$

- κ : niveau basal
- ξ_j : coefficients de régression
- h : fonction de lien
- u_i : condition de l'observation i

On teste $H_0 : \xi_1 = \xi_2 = \dots = \xi_{m_I}$.

► État de l'art

Analyse différentielle des voies métaboliques

- ▶ ktest [Ozier-Lafontaine et al., 2024] : test non linéaire basé sur les méthodes à noyaux

On teste $H_0 : \mathbb{P}_1 = \mathbb{P}_2$.

- ▶ Projection non linéaire des données dans \mathcal{H} via une fonction implicite ϕ
- ▶ Noyau q :

$$q(x_{I,i..}, x_{I,i'..}) = \langle \phi(x_{I,i..}), \phi(x_{I,i'..}) \rangle_{\mathcal{H}}$$

- ▶ Revient à un test d'égalité des moyennes dans \mathcal{H} avec

$$\forall k \in \{1, 2\}, \quad \mu_k = \mathbb{E}_{X_I \sim \mathbb{P}_k} [\phi(X_I)]$$

Analyse différentielle des voies métaboliques

- ▶ ktest [Ozier-Lafontaine et al., 2024] : test non linéaire basé sur les méthodes à noyaux

On teste $H_0 : \mathbb{P}_1 = \mathbb{P}_2$.

- ▶ Projection non linéaire des données dans \mathcal{H} via une fonction implicite ϕ
- ▶ Noyau q :

$$q(x_{I,i..}, x_{I,i'..}) = \langle \phi(x_{I,i..}), \phi(x_{I,i'..}) \rangle_{\mathcal{H}}$$

- ▶ Revient à un test d'égalité des moyennes dans \mathcal{H} avec

$$\forall k \in \{1, 2\}, \quad \mu_k = \mathbb{E}_{X_I \sim \mathbb{P}_k} [\phi(X_I)]$$

Limite : actuellement limité au test entre 2 conditions uniquement.

► Objectifs de la thèse

Objectif : analyse différentielle longitudinale des voies métaboliques

- ▶ Développer une méthode qui combine l'analyse longitudinale et l'analyse des voies métaboliques.
- ▶ Utiliser la méthode pour étudier le métabolome des photogranules.

> Plan

Introduction

Méthode proposée

Méthode

Évaluation de la méthode

Résultats

Conclusion

Application à l'étude des photogrammes

Conclusion et perspectives



INRAE

Séminaire biopuces

11/09/2025 / Camille Guilmineau

> Notations

X : matrice des quantifications de dimension ($nT \times m$)

- ▶ n : nombre d'individus
- ▶ T : nombre de dates
- ▶ m : nombre de métabolites

> Notations

X : matrice des quantifications de dimension ($nT \times m$)

- ▶ n : nombre d'individus
- ▶ T : nombre de dates
- ▶ m : nombre de métabolites

K : nombre de conditions d'intérêt

> Notations

X : matrice des quantifications de dimension ($nT \times m$)

- ▶ n : nombre d'individus
- ▶ T : nombre de dates
- ▶ m : nombre de métabolites

K : nombre de conditions d'intérêt

\mathcal{M}_l : voie métabolique (avec $l = 1, \dots, p$ où p est le nombre total de voies métaboliques)

> Notations

X : matrice des quantifications de dimension ($nT \times m$)

- ▶ n : nombre d'individus
- ▶ T : nombre de dates
- ▶ m : nombre de métabolites

K : nombre de conditions d'intérêt

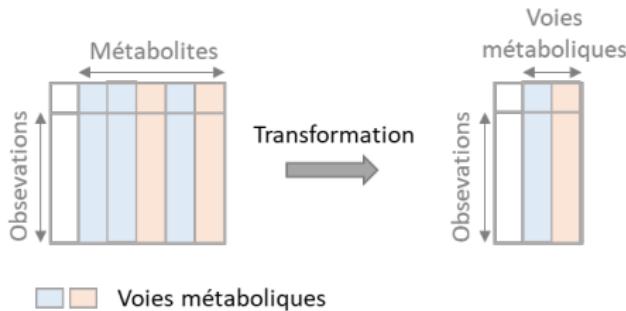
\mathcal{M}_l : voie métabolique (avec $l = 1, \dots, p$ où p est le nombre total de voies métaboliques)

$X_l = (x_{ij})_{i=1, \dots, n, j \in \mathcal{M}_l}$: matrice des quantifications de la voie métabolique \mathcal{M}_l

> Méthode phoenics : phoenicsPCA

Étape 1 : calcul des « scores » des voies métaboliques

- ▶ ACP sur X_I
- ▶ m_I^* composantes principales extraites dans les matrices A_I de dimension $(nT \times m_I^*)$



► Méthode phoenics : phoenicsPCA

Étape 2 : test des voies métaboliques

- ▶ Estimation d'un modèle mixte pour chaque composante principale a :

$$a = \beta_t + \theta_k + \alpha_i + \epsilon_{tki}$$

- ▶ β_t l'effet fixe du temps t ;
- ▶ θ_k l'effet fixe de la condition k ;
- ▶ $\alpha_i \sim \mathcal{N}(0, G)$ l'effet aléatoire de l'individu i , avec G la matrice de covariance ;
- ▶ $\epsilon_{tki} \sim \mathcal{N}(0, \sigma_\epsilon^2)$ les résidus du modèle, supposés i.i.d.

► Méthode phoenics : phoenicsPCA

Étape 2 : test des voies métaboliques

- ▶ Estimation d'un modèle mixte pour chaque composante principale a :

$$a = \beta_t + \theta_k + \alpha_i + \epsilon_{tki}$$

- ▶ β_t l'effet fixe du temps t ;
- ▶ θ_k l'effet fixe de la condition k ;
- ▶ $\alpha_i \sim \mathcal{N}(0, G)$ l'effet aléatoire de l'individu i , avec G la matrice de covariance ;
- ▶ $\epsilon_{tki} \sim \mathcal{N}(0, \sigma_\epsilon^2)$ les résidus du modèle, supposés i.i.d.
- ▶ Agrégation des p -valeurs par voie métabolique avec la procédure de Simes : contrôle l'erreur de Type I de
$$H_0 = \cap_{j=1}^{m_i^*} H_{0j}.$$

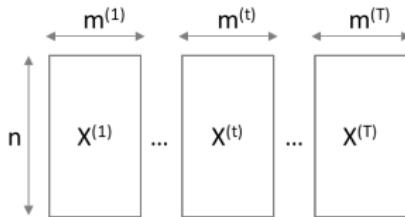
► Méthode phoenics : une variante phoenicsMFA

Limite de phoenicsPCA : ne prend pas en compte la structure des mesures répétées.

► Méthode phoenics : une variante phoenicsMFA

Limite de phoenicsPCA : ne prend pas en compte la structure des mesures répétées.

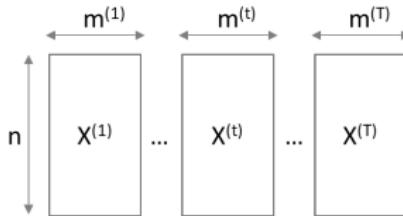
Analyse factorielle multiple (MFA)



► Méthode phoenics : une variante phoenicsMFA

Limite de phoenicsPCA : ne prend pas en compte la structure des mesures répétées.

Analyse factorielle multiple (MFA)



phoenicsMFA :

- ▶ X_I est vue comme un ensemble de matrices
 $X_I^{(t)} = (x_{ij})_{i=n(t-1)+1, \dots, nt, j \in \mathcal{M}_I}$
- ▶ MFA sur l'ensemble des matrices $X_I^{(t)}$ pour extraire les m_I^* composantes principales
- ▶ Modèle mixte

Simulation de données

Données semi-synthétiques simulées à partir de données réelles.

Simulation d'un effet :

- ▶ Condition
- ▶ Temps
- ▶ Condition et temps simultanément

Différents scénarios : variation de la taille de l'effet.

Simulation des données : effet condition

Données :

Temps	Condition	Métabolites des voies différentielles	Autres métabolites
1	Y		
2	Y	G	
3	Y		
4	Y	G	
5	Y		
6	Y	G	
7	Y		
8	Y	G	
9	Y		
10	R	G	
11	R		
12	R	G	
13	R		
14	R	G	
15	R		



Simulation des données : effet condition

Simulation sous H_0 :



Permutation des
conditions



Simulation des données : effet condition

Simulation sous H_0 :

	Temps	Condition	Métabolites des voies différentielles	Autres métabolites
1	Yellow	Green	Red	Blue
2	Orange	Green	Red	Blue
3	Yellow	Green	Red	Blue
4	Red	Green	Red	Blue

Permutation des conditions

Simulation sous H_1 :

	Temps	Condition	Métabolites des voies différentielles	Autres métabolites
1	Yellow	Green	Red	Red
2	Orange	Green	Red	Red
3	Yellow	Green	Red	Red
4	Red	Green	Red	Red

- * Introduction d'un effet condition
- $$\tilde{x}_{ij} = x_{ij} \times \gamma_k$$

Simulation des données : effet temps

Simulation sous H_0 :

Temps	Condition	Métabolites des voies différentielles	Autres métabolites
1	Yellow	Blue	
2	Yellow	Green	
3	Yellow	Blue	Red
4	Yellow	Green	
5	Red	Blue	
6	Green		



Remplacement par la quantification à la date 1,
ajout d'un bruit

Simulation des données : effet temps

Simulation sous H_0 :

Temps	Condition	Métabolites des voies différentielles	Autres métabolites
Yellow	Blue		
Yellow	Green		
Yellow	Blue		
Yellow	Green		
Red	Blue		
Red	Green		

Remplacement par la quantification à la date 1,
ajout d'un bruit



Simulation sous H_1 :

Temps	Condition	Métabolites des voies différentielles	Autres métabolites
Yellow	Blue		
Yellow	Green		
Yellow	Blue		
Yellow	Green		
Red	Blue		
Red	Green		

- Introduction d'un effet temps
 $\tilde{x}_{ij} = x_{ij} \times \gamma_t$

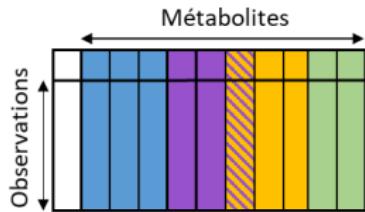
Évaluation de la méthode sur des données simulées

Evaluation de 3 critères :

- ▶ Contrôle de l'erreur de Type I
- ▶ Puissance statistique
- ▶ Puissance statistique par rapport au pourcentage de métabolites différentiels

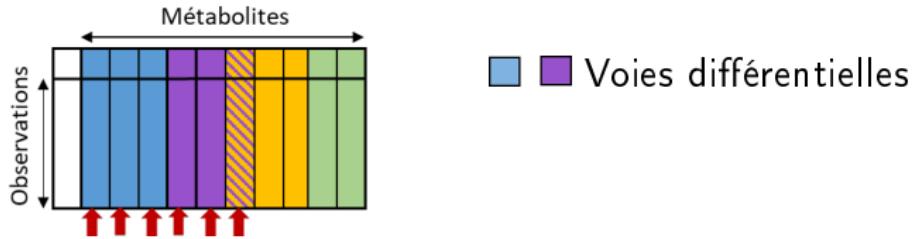
➤ Évaluation de la méthode sur des données simulées

Sous H_1 , différents types de voies simulées.



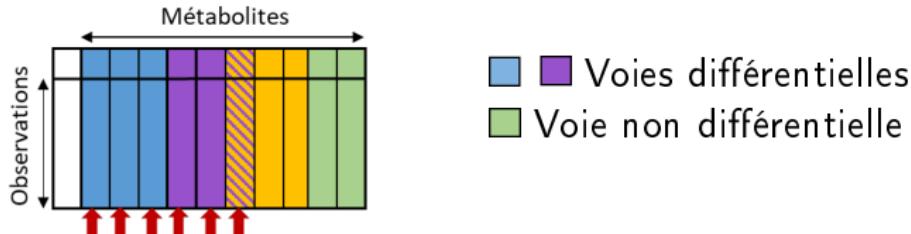
➤ Évaluation de la méthode sur des données simulées

Sous H_1 , différents types de voies simulées.



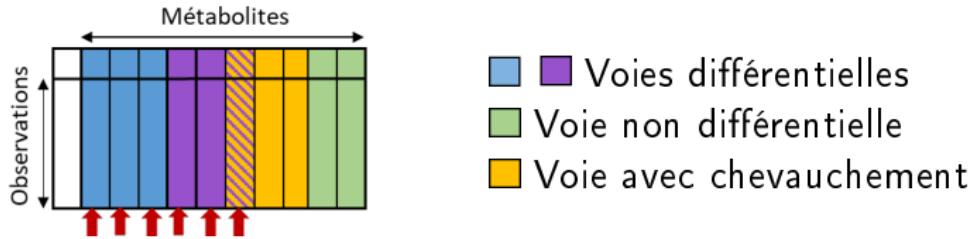
► Évaluation de la méthode sur des données simulées

Sous H_1 , différents types de voies simulées.



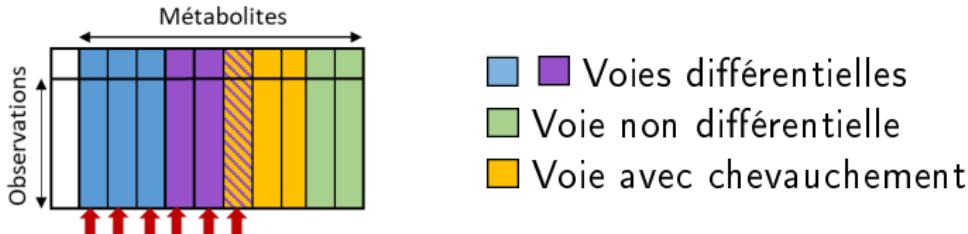
➤ Évaluation de la méthode sur des données simulées

Sous H_1 , différents types de voies simulées.



► Évaluation de la méthode sur des données simulées

Sous H_1 , différents types de voies simulées.



	Significative	Non significative
Différentielle	Vrai positif	Faux négatif
Non différentielle	Faux positif	Vrai négatif
Chevauchement	Positif	Négatif

► Évaluation de la méthode sur des données simulées

	Significatif	Non significatif
Différentiel	Vrai positif	Faux négatif
Non différentiel	Faux positif	Vrai négatif
Chevauchement	Positif	Négatif

- Comptage du nombre de voies dans chaque catégorie
- Valeur prédictive (PPV) = $\frac{\text{Nombre de vrais positifs}}{\text{Nombre de voies déclarées positives}}$
- Sensibilité = $\frac{\text{Nombre de vrais positifs}}{\text{Nombre de voies différentielles}}$

► Résultats : test de l'effet condition

Contrôle de l'erreur de Type I

Pourcentage de tests positifs sous H0 :

	Condition
phoenicsMFA	1.10%
phoenicsPCA	4.08%
Enrichissement	2.04%
ktest	0.00%
globaltest	0.00%

► Résultats : test de l'effet condition

Contrôle de l'erreur de Type I

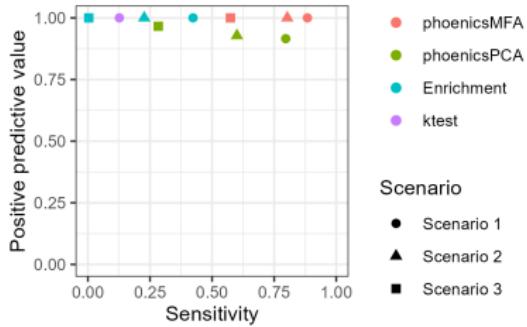
Pourcentage de tests positifs sous H0 :

	Condition
phoenicsMFA	1.10%
phoenicsPCA	4.08%
Enrichissement	2.04%
ktest	0.00%
globaltest	0.00%

- Toutes les méthodes contrôlent l'erreur de Type I.

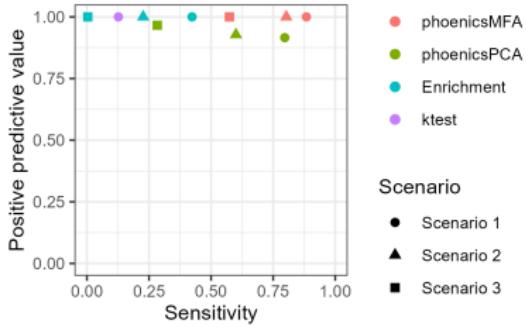
► Résultats : test de l'effet condition

Puissance statistique



► Résultats : test de l'effet condition

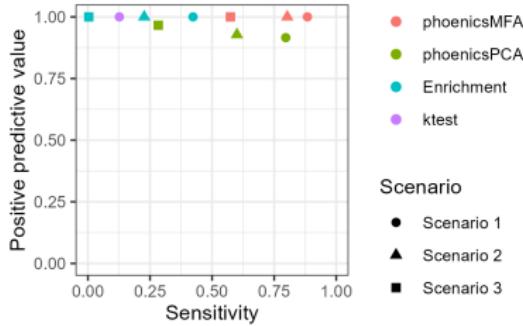
Puissance statistique



- ▶ La sensibilité diminue lorsque la taille de l'effet diminue.

Résultats : test de l'effet condition

Puissance statistique

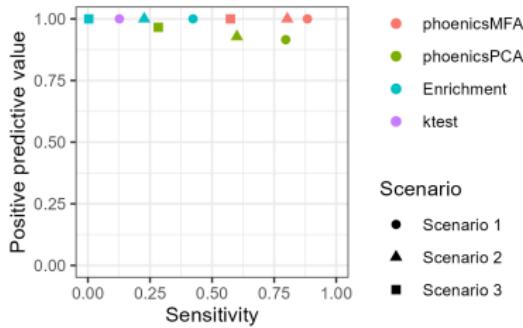


- ▶ La sensibilité diminue lorsque la taille de l'effet diminue.
- ▶ phoenicsMFA et phoenicsPCA ont une meilleure sensibilité.



Résultats : test de l'effet condition

Puissance statistique



- ▶ La sensibilité diminue lorsque la taille de l'effet diminue.
- ▶ phoenicsMFA et phoenicsPCA ont une meilleure sensibilité.
- ▶ phoenicsMFA a une meilleure sensibilité et valeur prédictive que phoenicsPCA.



► Résultats : test de l'effet condition

Puissance statistique

Pourcentage de voies avec chevauchement déclarées positives :

	Scénario 1	Scénario 2	Scénario 3
phoenicsMFA	51%	39%	27%
phoenicsPCA	44%	30%	16%
Enrichissement	9%	6%	0%
ktest	3%	0%	0%
globaltest	0%	0%	0%

- phoenicsPCA et phoenicsMFA déclarent positives un plus grand pourcentage de voies que les autres méthodes.

► Résultats : test de l'effet condition

Puissance statistique

Pourcentage de voies avec chevauchement déclarées positives :

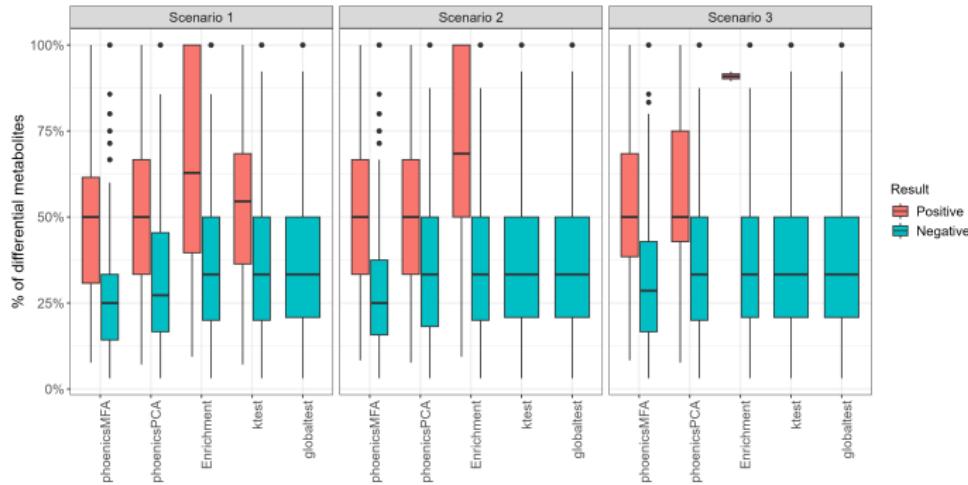
	Scénario 1	Scénario 2	Scénario 3
phoenicsMFA	51%	39%	27%
phoenicsPCA	44%	30%	16%
Enrichissement	9%	6%	0%
ktest	3%	0%	0%
globaltest	0%	0%	0%

- ▶ phoenicsPCA et phoenicsMFA déclarent positives un plus grand pourcentage de voies que les autres méthodes.
- ▶ Est-ce que ces voies sont pertinentes ?



Résultats : test de l'effet condition

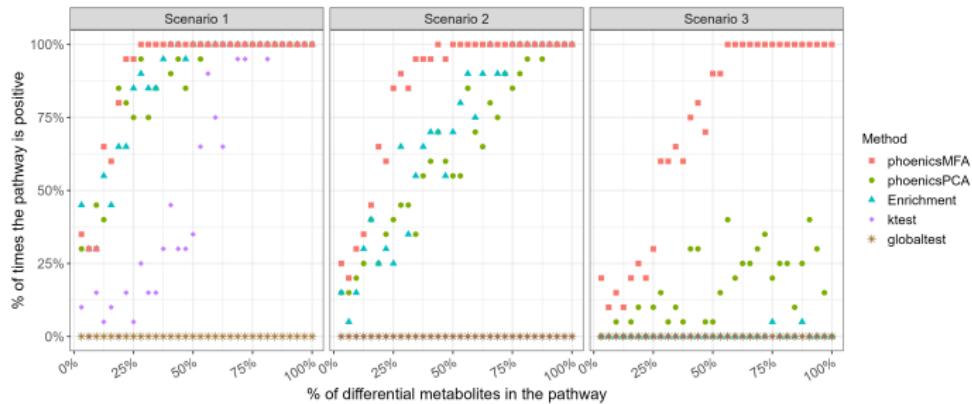
Puissance statistique



- ▶ Plus de métabolites différentiels dans les voies déclarées positives que négatives.
- ▶ Les voies positives avec phoenicsMFA ont un pourcentage de métabolites différentiels distinct des voies négatives.

Résultats : test de l'effet condition

Puissance de détection en fonction du pourcentage de métabolites différentiels dans une voie



- ▶ Pourcentage de voies positives plus élevé avec phoenicsMFA.
- ▶ Les différences entre les méthodes augmentent quand la taille de l'effet diminue.

► Résultats : test de l'effet temps

Contrôle de l'erreur de Type I :

- ▶ Toutes les méthodes sauf globaltest contrôlent l'erreur de Type I.

► Résultats : test de l'effet temps

Contrôle de l'erreur de Type I :

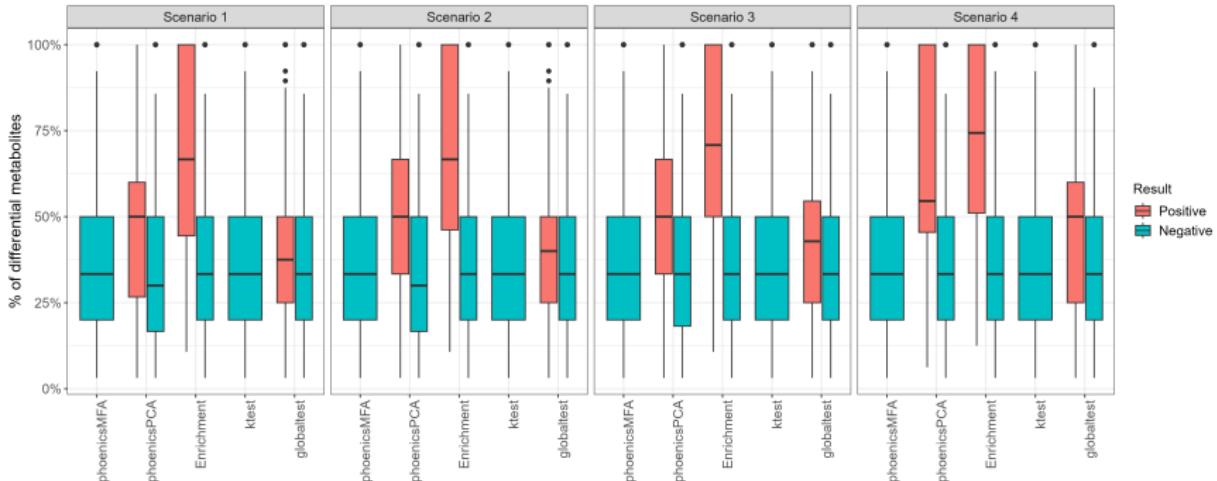
- ▶ Toutes les méthodes sauf globaltest contrôlent l'erreur de Type I.

Puissance statistique :

- ▶ phoenicsMFA n'identifie pas de voie positive.
- ▶ phoenicsPCA et globaltest ont une meilleure sensibilité que les autres méthodes.



Résultats : test de l'effet temps



- ▶ Les voies déclarées positives et négatives par globaltest ont un pourcentage de métabolites différentiels similaire.

⚠ Doute sur la pertinence des voies positives par globaltest.

► Résultats : test des effets condition et temps simultanément

Test de l'effet temps :

- ▶ phoenicsPCA a les meilleures performances.

Résultats : test des effets condition et temps simultanément

Test de l'effet temps :

- ▶ phoenicsPCA a les meilleures performances.

Test de l'effet condition :

- ▶ phoenicsPCA a une meilleure sensibilité et valeur prédictive que phoenicsMFA.

➤ Conclusion sur les données semi-synthétiques

Méthode phoenics la plus performante :

- ▶ Un seul effet : phoenicsMFA
- ▶ Plusieurs effets ou effet temps : phoenicsPCA

► Résultats : données réelles

Étude de l'effet d'antibiotiques sur des souris :

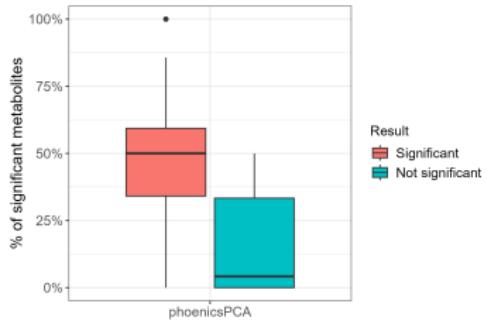
- ▶ 8 souris
- ▶ 2 conditions
- ▶ 3 dates

Nombre de voies métaboliques significatives, sur 98 voies testées :

	Temps	Condition
phoenicsPCA	72	1
phoenicsMFA	0	5
Enrichissement	0	0

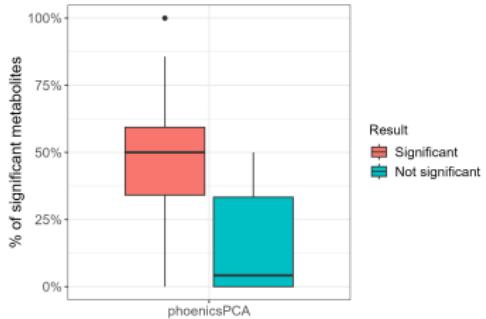
> Résultats : données réelles

Pourcentage de métabolites différenciels pour l'effet temps :



> Résultats : données réelles

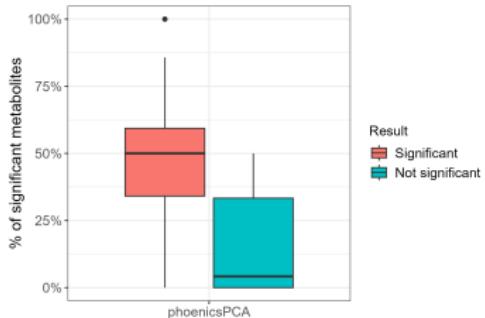
Pourcentage de métabolites différenciels pour l'effet temps :



- ▶ Pourcentage de métabolites différenciels distinct.

> Résultats : données réelles

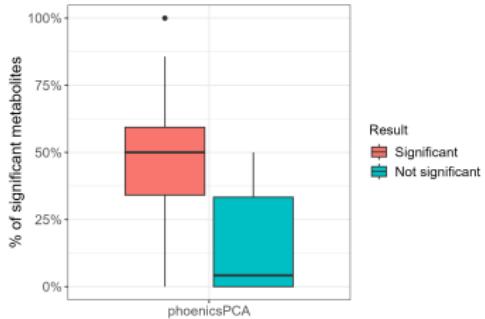
Pourcentage de métabolites différentiels pour l'effet temps :



- ▶ Pourcentage de métabolites différentiels distinct.
- ▶ Voies significatives avec peu de métabolites différentiels grâce à l'analyse multivariée.

> Résultats : données réelles

Pourcentage de métabolites différentiels pour l'effet temps :



- ▶ Pourcentage de métabolites différentiels distinct.
- ▶ Voies significatives avec peu de métabolites différentiels grâce à l'analyse multivariée.
- ▶ Voies positives pertinentes biologiquement.

> Conclusion

Approche proposée pour l'analyse différentielle longitudinale des voies métaboliques :

- ▶ Méthode en deux étapes basée sur la réduction de dimension et le modèle mixte.

> Conclusion

Approche proposée pour l'analyse différentielle longitudinale des voies métaboliques :

- ▶ Méthode en deux étapes basée sur la réduction de dimension et le modèle mixte.

Résultats :

- ▶ phoenicsMFA est adapté au test de n'importe quel effet sauf celui associé aux mesures répétées.
- ▶ phoenicsPCA permet de tester cet effet.
- ▶ Les méthodes phoenics ont une meilleure puissance de détection que les autres méthodes, tout en contrôlant l'erreur de Type I.
- ▶ Sur des données réelles, les méthodes phoenics permettent d'identifier des voies pertinentes.

> Conclusion

Productions :

- ▶ Package R **phoenics** publié sur le CRAN.
- ▶ Article publié dans *BMC Bioinformatics*.

► Plan

Introduction

Méthode proposée

Application à l'étude des photogranules

 Introduction

 Données

 Méthode

 Résultats

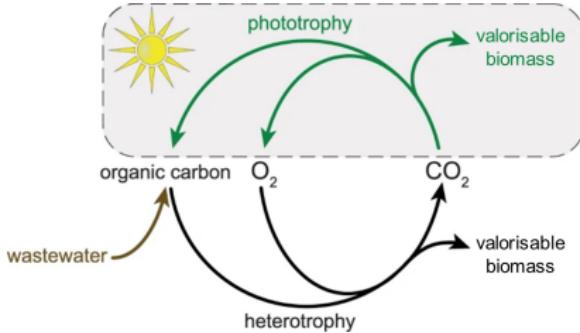
 Conclusion

Conclusion et perspectives

▶ Introduction

Composition des photogranules :

- ▶ Organismes phototrophes
- ▶ Organismes hétérotrophes



Production d'oxygène qui pourrait remplacer l'aération mécanique pour le traitement des eaux usées, capacités épuratoires.

Les variations des conditions environnementales peuvent :

- ▶ Influencer le métabolisme des photogranules
- ▶ Modifier leurs performances

> Introduction : le projet PANORAMICS

Objectif : identifier des biomarqueurs liés à des stress environnementaux.

Étude de l'impact sur le métabolome des variations de :

- ▶ Disponibilité en carbone
- ▶ Intensité lumineuse



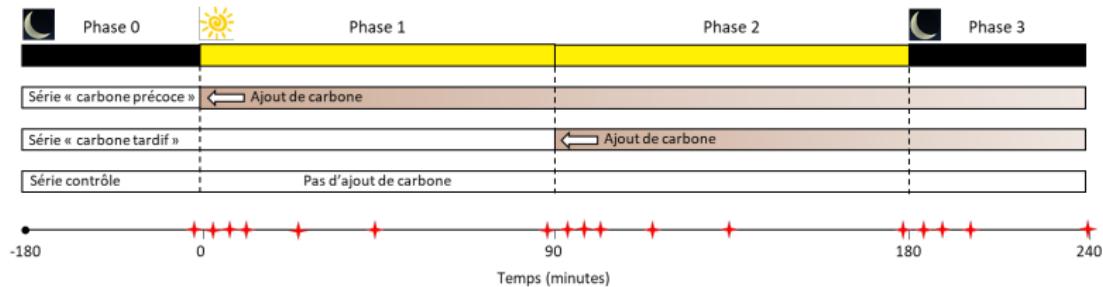
▶ Introduction : le projet PANORAMICS

Objectif : identifier des biomarqueurs liés à des stress environnementaux.

Étude de l'impact sur le métabolome des variations de :

- ▶ Disponibilité en carbone
- ▶ Intensité lumineuse

Plan expérimental :



► Données et notations

Matrice des quantifications : 532 observations × 143 métabolites
67 voies métaboliques

► Données et notations

Matrice des quantifications : 532 observations \times 143 métabolites
67 voies métaboliques

Conditions expérimentales :

- ▶ Ajout de carbone :
 $k \in \{\text{carbone précoce, carbone tardif, contrôle}\}$
- ▶ État lumineux : $d \in \{\text{lumière, obscurité}\}$
- ▶ Date de prélèvement :
 $t \in \{-2, 5, 10, 15, 45, 88, 95, 100, 105, 120, 135, 178, 185, 190, 200, 240\}$

> Méthode

Test de l'effet de l'ajout de carbone avec phoenicsPCA.

Données : séries « carbone précoce » et « carbone tardif » dans les phases 1 et 3 séparément.

> Méthode

Test de l'effet de l'ajout de carbone avec phoenicsPCA.

Données : séries « carbone précoce » et « carbone tardif » dans les phases 1 et 3 séparément.

Modèle mixte estimé sur les scores des voies métaboliques, notés génériquement a :

$$a = \beta_t + \theta_k + \alpha_i + \epsilon_{tki}$$

- ▶ β_t : effet fixe de la date de prélèvement t
- ▶ θ_k : effet fixe de la série k
- ▶ α_i : effet aléatoire correspondant à la date d'extraction du métabolome de l'individu i
- ▶ ϵ_{tki} : résidus du modèle

> Méthode

Test de l'effet de la lumière avec phoenicsPCA.

Données : séries « carbone précoce » et « carbone tardif »
séparément dans les phases 2 et 3.

> Méthode

Test de l'effet de la lumière avec phoenicsPCA.

Données : séries « carbone précoce » et « carbone tardif »
séparément dans les phases 2 et 3.

Modèle mixte estimé sur les scores des voies métaboliques, notés
génériquement a :

$$a = \beta_t + \delta_d + \alpha_i + \epsilon_{tdi}$$

- δ_d l'effet fixe correspondant à l'état lumineux d

> Méthode

Test de l'effet de la lumière avec phoenicsPCA.

Données : séries « carbone précoce » et « carbone tardif »
séparément dans les phases 2 et 3.

Modèle mixte estimé sur les scores des voies métaboliques, notés
génériquement a :

$$a = \beta_t + \delta_d + \alpha_i + \epsilon_{tdi}$$

- ▶ δ_d l'effet fixe correspondant à l'état lumineux d

Pour le test des deux effets : correction des tests multiples des voies
métaboliques avec la procédure BH.

Test de l'effet de l'ajout de carbone

Phase 1 : 8 voies significatives, dont 7 contenant le métabolite acétate.

- ▶ Voie « *Amino sugar and nucleotide sugar metabolism* » : le métabolite N-acetylneuraminate peut être utilisé comme source de carbone.

Test de l'effet de l'ajout de carbone

Phase 1 : 8 voies significatives, dont 7 contenant le métabolite acétate.

- ▶ Voie « *Amino sugar and nucleotide sugar metabolism* » : le métabolite N-acetylneuraminate peut être utilisé comme source de carbone.

Phase 3 : 10 voies significatives.

- ▶ Voie « *Carbon fixation by Calvin cycle* »

Test de l'effet de la lumière

Pas de voie significative.

3 métabolites significativement différents entre les phases 2 et 3, dans une des séries.

- ▶ La réponse des photogranules aux changements de lumière est influencée par l'historique de l'apport en carbone.

> Conclusion

- ▶ phoenics a permis d'identifier des voies métaboliques impactées par les modifications de l'environnement.
- ▶ Les métabolites de certaines voies identifiées ont des variations faibles mais cohérentes.
- ▶ En cas d'apport de carbone, les activités phototrophiques et hétérotrophiques semblent maintenues.
- ▶ En cas de limitation en carbone et lumière, d'autres sources de carbone disponibles sont sollicitées.

► Plan

Introduction

Méthode proposée

Application à l'étude des photogranules

Conclusion et perspectives



INRAE

Séminaire biopuces

11/09/2025 / Camille Guilmineau

> Conclusion

Méthode phoenics :

- ▶ Méthode d'analyse différentielle longitudinale des voies métaboliques.
- ▶ Meilleures performances que les autres méthodes sur données semi-synthétiques et réelles.

> Conclusion

Méthode phoenics :

- ▶ Méthode d'analyse différentielle longitudinale des voies métaboliques.
- ▶ Meilleures performances que les autres méthodes sur données semi-synthétiques et réelles.

Étude des photogranules :

- ▶ Utilisation de la méthode phoenics pour étudier les variations du métabolome des photogranules.
- ▶ Identification de voies et métabolites impactées par les modifications de l'environnement des photogranules.
- ▶ Les métabolites des voies identifiées ont des variations faibles mais cohérentes.

Perspectives

Méthode phoenics :

- ▶ Utiliser une méthode qui utilise la structure en graphe des voies métaboliques.
- ▶ Se baser sur les distances entre les métabolites dans le graphe pour éviter le découpage arbitraire des voies métaboliques.

Perspectives

Méthode phoenics :

- ▶ Utiliser une méthode qui utilise la structure en graphe des voies métaboliques.
- ▶ Se baser sur les distances entre les métabolites dans le graphe pour éviter le découpage arbitraire des voies métaboliques.

Étude des photogranules :

- ▶ Étudier le rôle des indicateurs identifiés.
- ▶ Utiliser phoenicsMFA pour prendre en compte la structure longitudinale des données.
- ▶ Intégrer les données métabolomiques avec les données métatranscriptomiques mesurées sur les mêmes échantillons.

Perspectives

Méthode phoenics :

- ▶ Utiliser une méthode qui utilise la structure en graphe des voies métaboliques.
- ▶ Se baser sur les distances entre les métabolites dans le graphe pour éviter le découpage arbitraire des voies métaboliques.

Étude des photogranules :

- ▶ Étudier le rôle des indicateurs identifiés.
- ▶ Utiliser phoenicsMFA pour prendre en compte la structure longitudinale des données.
- ▶ Intégrer les données métabolomiques avec les données métatranscriptomiques mesurées sur les mêmes échantillons.

Méthode d'intégration multi-omiques :

- ▶ Développer une méthode d'analyse longitudinale de voies multi-omiques.

► Remerciements

Merci aux membres du jury, à mes encadrant.es, à l'ensemble du LBE et à toutes les personnes qui ont participé à cette thèse.



 Goeman, J. J., van de Geer, S. A., de Kort, F., and van Houwelingen, H. C. (2004).

A global test for groups of genes : testing association with a clinical outcome.

Bioinformatics, 20(1) :93–99.

 Martin, M. and Govaerts, B. (2020).

LiMM-PCA : Combining ASCA+ and linear mixed models to analyse high-dimensional designed data.

Journal of Chemometrics, 34(6).

 Molenberghs, G. and Verbeke, G. (2000).

Linear Mixed Models for Longitudinal Data.

Springer New York.

 Ozier-Lafontaine, A., Fourneaux, C., Durif, G., Arsenteva, P., Vallot, C., Gandrillon, O., Gonin-Giraud, S., Michel, B., and Picard, F. (2024).

Kernel-based testing for single-cell differential analysis.
Genome Biology, 25 :114.



INRAE

Séminaire biopuces

11/09/2025 / Camille Guilmineau

p. 47