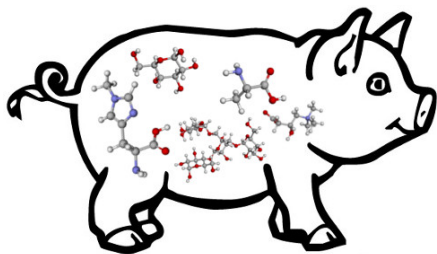


Projet SUSoSTRESS : analyses de données métabolomiques



Gaëlle Lefort

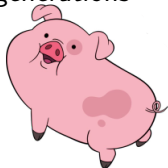
Biopuces

24 novembre 2016



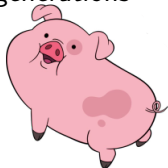
Projet SUSoSTRESS : protocole expérimental

2 générations



Projet SUSoSTRESS : protocole expérimental

2 générations

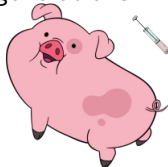


Génome



Projet SUSoSTRESS : protocole expérimental

2 générations



Stress (ACTH, contrainte, Lipopolysaccharide (LPS))



Génome



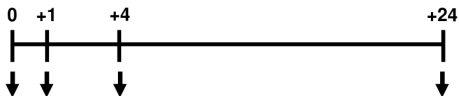
Projet SUSoSTRESS : protocole expérimental

2 générations



Stress (ACTH, contrainte, Lipopolysaccharide (LPS))

Tissu prélevé : sang total



Génome



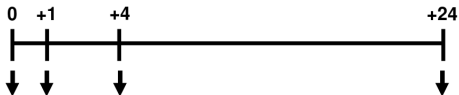
Projet SUSoSTRESS : protocole expérimental

2 générations



Stress (ACTH, contrainte, Lipopolysaccharide (LPS))

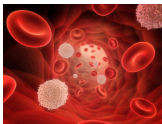
Tissu prélevé : sang total



Génome



Biologie clinique



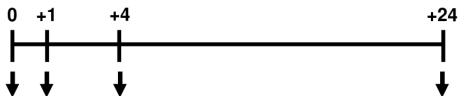
Projet SUSoSTRESS : protocole expérimental

2 générations



Stress (ACTH, contrainte, Lipopolysaccharide (LPS))

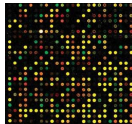
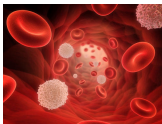
Tissu prélevé : sang total



Génome

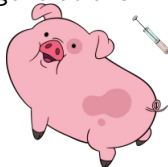


Biologie clinique Transcriptome



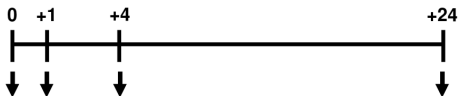
Projet SUSoSTRESS : protocole expérimental

2 générations



Stress (ACTH, contrainte, Lipopolysaccharide (LPS))

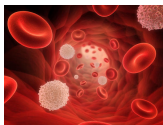
Tissu prélevé : sang total



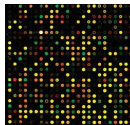
Génome



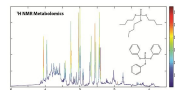
Biologie clinique



Transcriptome

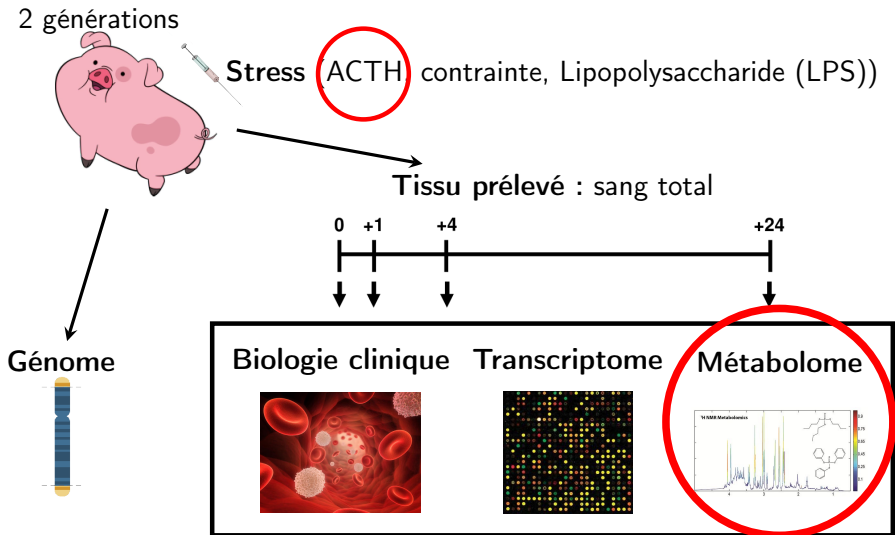


Métabolome



Source : thèse de V. Sautron

Projet SUSoSTRESS : protocole expérimental



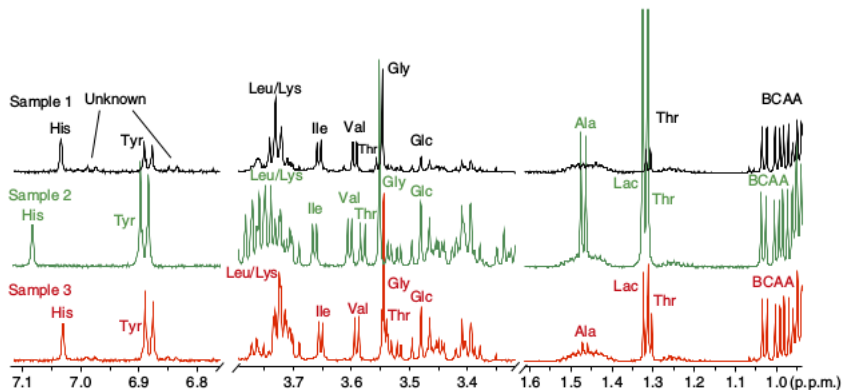
Source : thèse de V. Sautron

Objectifs

- Etude de la réponse au stress au niveau du métabolome :
 - Quels sont les métabolites dont l'abondance change après un stress ?
 - Quelles sont les différences de réponse au stress entre les générations et les lignées ?
 - Quelles sont les différences dans le métabolome entre les générations et les lignées à *baseline* ?
- Identification et quantification des métabolites dans un spectre
- Intégration du métabolome avec le génome puis les autres données omics

Données métabolomiques

Spectre obtenu par résonance magnétique nucléaire (RMN) :



Hao, J., Liebecke, M., Astle, W., De Iorio, M., Bundy, J. G., & Ebbels, T. M. (2014). Bayesian deconvolution and quantification of metabolites in complex 1D NMR spectra using BATMAN. *Nat. Protoc.* 9(6), 1416-1427.

Sommaire

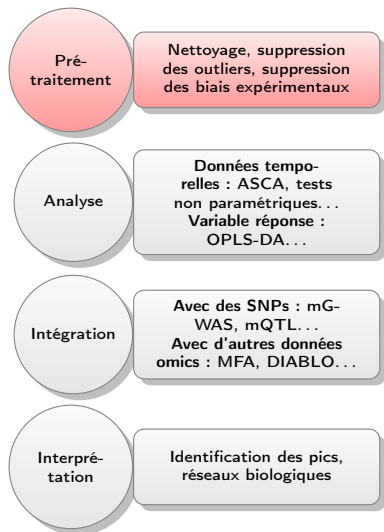
Étapes de l'analyse de données métabolomiques

Recherche des métabolites dont l'abondance évolue après un stress

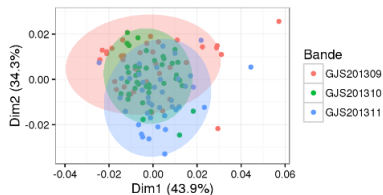
Identification (et quantification) des métabolites

Conclusion

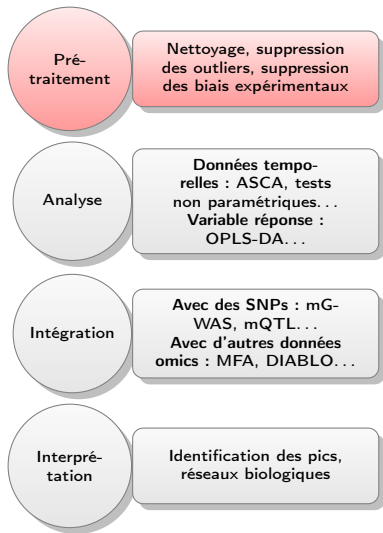
Pré-traitement



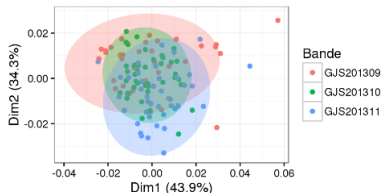
Identification des problèmes avec des ACPs :



Pré-traitement



Identification des problèmes avec des ACPs :



Normalisation avec la méthode ComBat (package sva)

Johnson, W. E., Li, C., & Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*, 8(1), 118-127.

Analyses multivariées

Pré-
traitement

Nettoyage, suppression
des outliers, suppression
des biais expérimentaux

Analyse

Données tempo-
relles : ASCA, tests
non paramétriques. . .
Variable réponse :
OPLS-DA. . .

Intégration

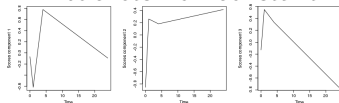
Avec des SNPs : mG-
WAS, mQTL. . .
Avec d'autres données
omics : MFA, DIABLO. . .

Interpré-
tation

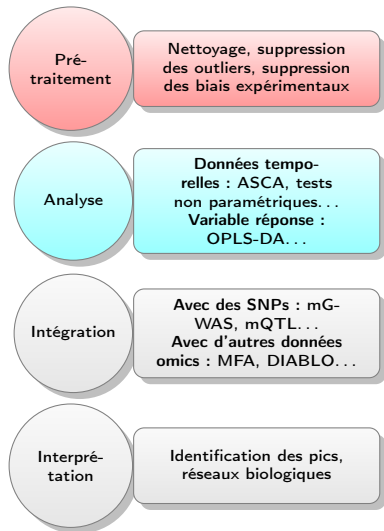
Identification des pics,
réseaux biologiques

Données temporelles

ASCA faire ressortir les
buckets non constant

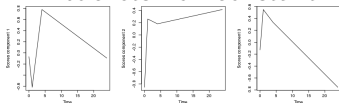


Analyses multivariées



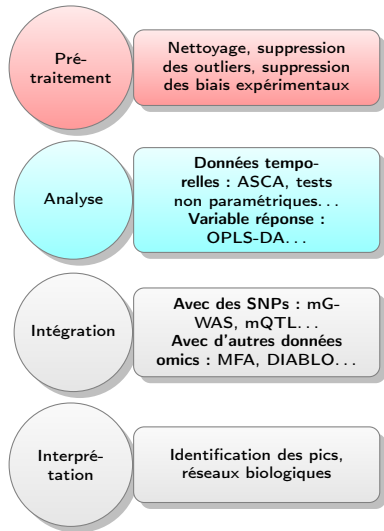
Données temporelles

ASCA faire ressortir les buckets non constant



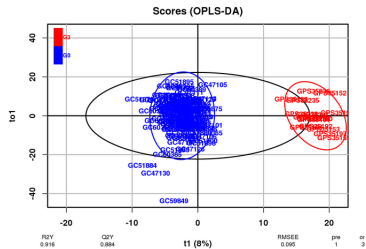
Tests tests de Wilcoxon deux à deux par rapport à baseline

Analyses multivariées

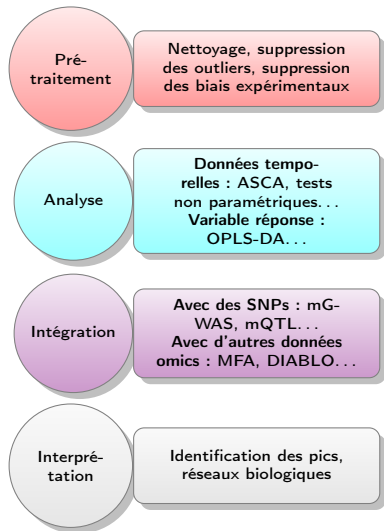


Variable réponse

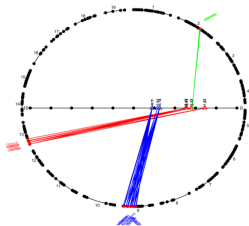
OPLS-DA faire ressortir les buckets créant la différence entre les groupes



Intégration avec d'autres type de données

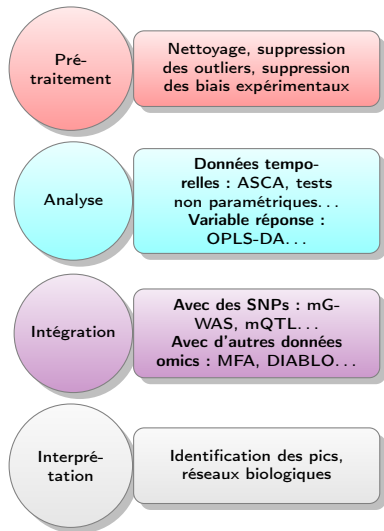


Intégration de SNPs :

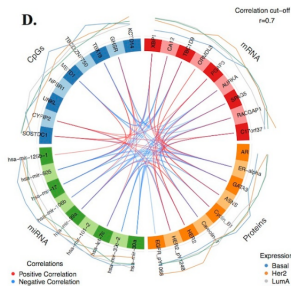


Hedjazi, L., Gauguier, D., Zalloua, P. A., Nicholson, J. K., Dumas, M. E., & Cazier, J. B. (2015). mQTL. NMR : an integrated suite for genetic mapping of quantitative variations of ^1H NMR-based metabolic profiles. *Analytical chemistry*, 87(8), 4377-4384.

Intégration avec d'autres type de données

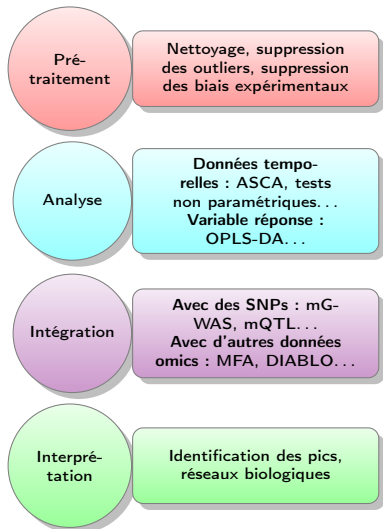


Intégration de données multi-omics :

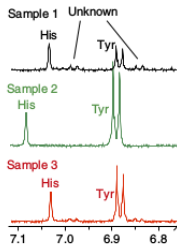


A Singh, B Gautier, C Shannon, M Vacher, F Rohart, S Tebbutt, K-A. Lê Cao. DIABLO – an integrative, multi-omics, multivariate method for multi-group classification. Submitted.

Interprétation biologique



Identification des pics :



Hao, J., Liebeke, M., Astle, W., De Iorio, M., Bundy, J. G., & Ebbels, T. M. (2014). Bayesian deconvolution and quantification of metabolites in complex 1D NMR spectra using BATMAN. *Nat. Protoc*, 9(6), 1416-1427.

Interprétation biologique

Pré-
traitement

Nettoyage, suppression
des outliers, suppression
des biais expérimentaux

Analyse

Données tempo-
relles : ASCA, tests
non paramétriques. . .
Variable réponse :
OPLS-DA. . .

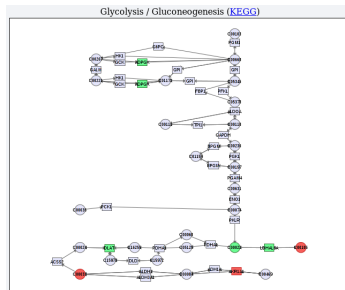
Intégration

Avec des SNPs : mG-
WAS, mQTL. . .
Avec d'autres données
omics : MFA, DIABLO. . .

Interpré-
tation

Identification des pics,
réseaux biologiques

Réseaux biologiques :



Xia, J., & Wishart, D. S. (2011). Web-based inference of biological patterns, functions and pathways from metabolomic data using MetaboAnalyst. *Nature protocols*, 6(6), 743-760.

Sommaire

Étapes de l'analyse de données métabolomiques

Recherche des métabolites dont l'abondance évolue après un stress

Identification (et quantification) des métabolites

Conclusion

Données analysées

Données disponibles :

- 120 porcs Large White
- 4 temps de mesure
- 613 buckets

Pré-traitement :

- 1 spectre ré-aligné
- effet de bande corrigé

Tests statistiques

- Tests de Wilcoxon appariés (non paramétriques) et correction de Bonferroni

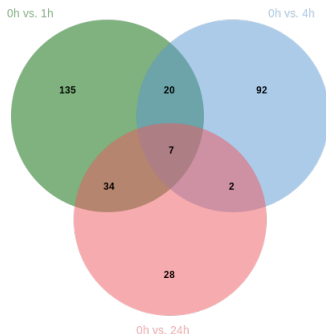


Figure – Nombre de buckets significatifs ($\alpha = 5\%$)

Tests statistiques

- Tests de Wilcoxon appariés (non paramétriques) et correction de Bonferroni
- Classification de courbes pour les buckets mis en évidence

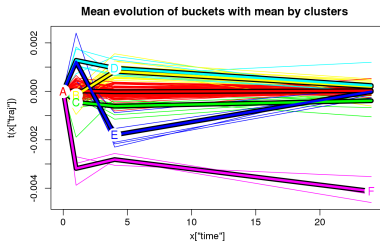


Figure – Clusters obtenus grâce à la méthode kml

Source : Genolini, C., & Falissard, B. (2010). KmL : k-means for longitudinal data. Computational Statistics, 25(2), 317-328

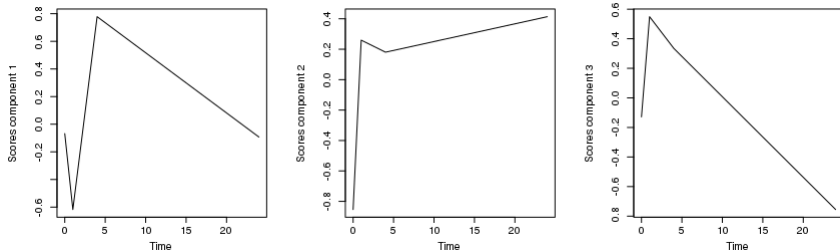
ANOVA-simultaneous component analysis (ASCA)

- ACP sur les coefficients d'une ANOVA

ANOVA-simultaneous component analysis (ASCA)

- ACP sur les coefficients d'une ANOVA

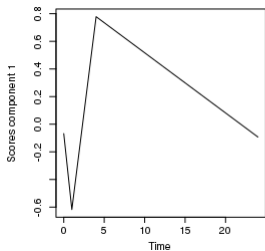
Figure – Profils identifiés par l'ASCA



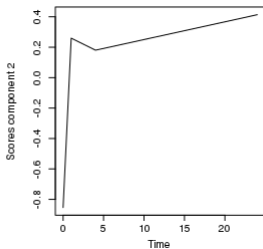
ANOVA-simultaneous component analysis (ASCA)

- ACP sur les coefficients d'une ANOVA

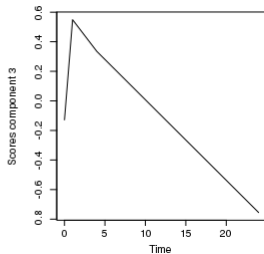
Figure – Profils identifiés par l'ASCA



37 buckets



2 buckets



1 buckets

ANOVA-simultaneous component analysis (ASCA)

■ ACP sur les coefficients d'une ANOVA

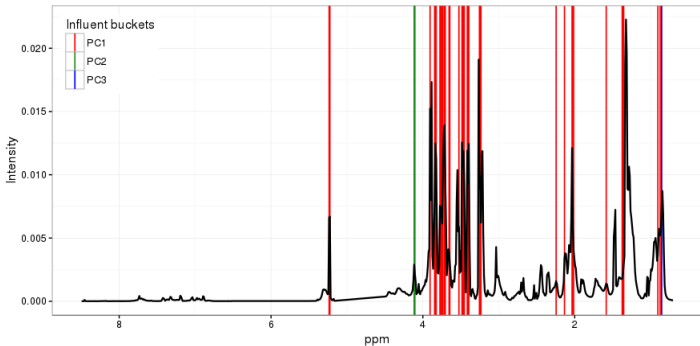


Figure – Buckets influents sur la courbe moyenne

Comparaison entre les deux méthodes

Buckets mis en évidence par chaque méthode :



Sommaire

Étapes de l'analyse de données métabolomiques

Recherche des métabolites dont l'abondance évolue après un stress

Identification (et quantification) des métabolites

Conclusion

Plusieurs méthodes

■ Identification à la main

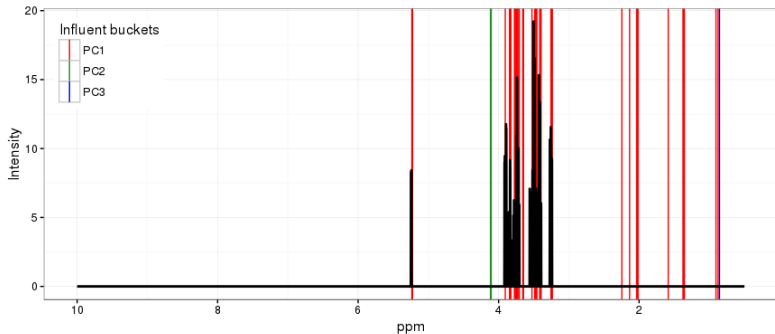
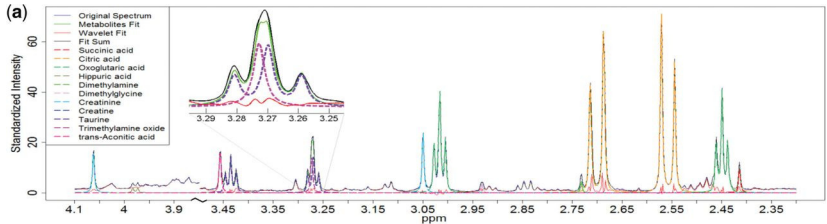


Figure – Spectre du glucose

Plusieurs méthodes

- Identification à la main
- Modèle de mélange complexe



Source : Hao, J., Liebeke, M., Astle, W., De Iorio, M., Bundy, J. G., & Ebbels, T. M. (2014). Bayesian deconvolution and quantification of metabolites in complex 1D NMR spectra using BATMAN. *Nat. Protoc.* 9(6), 1416-1427.

1^{re} méthode : Bayesian AuTomated Metabolite Analyser for NMR spectra (BATMAN)

- Avantages**
- grande bibliothèque de métabolites (environ 700)
 - fournit des intervalles de crédibilité aux estimations

- Inconvénients**
- difficile à paramétrer
 - très couteux en temps
 - ne semble pas prendre en compte les décalages de pics

Source : Hao, J., Liebeke, M., Astle, W., De Iorio, M., Bundy, J. G., & Ebbels, T. M. (2014). Bayesian deconvolution and quantification of metabolites in complex 1D NMR spectra using BATMAN. *Nat. Protoc.*, 9(6), 1416-1427.

2^e méthode : Automatic Statistical Identification of metabolites in Complex Spectra (ASICS)

- Avantages**
- élimination des métabolites qui ne peuvent pas être dans le mélange
 - filtrage des métabolites trop peu concentrés
 - facile à utiliser et rapide

- Inconvénients**
- assez peu de métabolites dans la librairie
 - l'initialisation peut faire varier les résultats

Source : Tardivel, P. J., Servien, R., & Concordet, D. (2016). Non-asymptotic active set properties of lasso-type estimators in small-dimension.

Comparaison

Données Mélange de 21 métabolites en concentrations connues

Pseudo R^2 $R^2_{ASICS} = 0,94 < R^2_{BATMAN} = 0,54$

Comparaison

Données Mélange de 21 métabolites en concentrations connues

Pseudo R^2 $R^2_{ASICS} = 0,94 < R^2_{BATMAN} = 0,54$

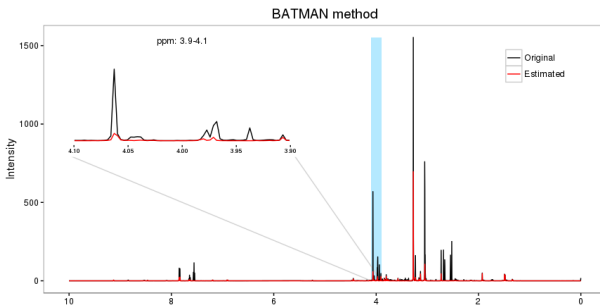


Figure – Estimation du spectre par la méthode BATMAN

Comparaison

Données Mélange de 21 métabolites en concentrations connues

Pseudo R^2 $R^2_{ASICS} = 0,94 < R^2_{BATMAN} = 0,54$

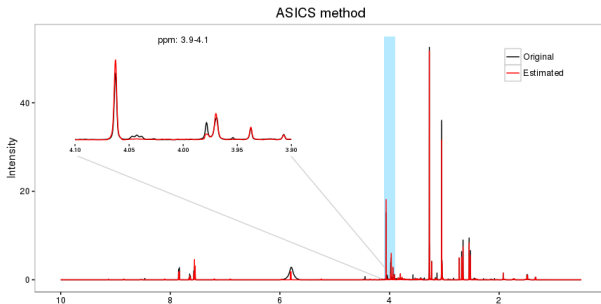


Figure – Estimation du spectre par la méthode ASICS

Comparaison

Données Mélange de 21 métabolites en concentrations connues

Pseudo R^2 $R^2_{ASICS} = 0,94 < R^2_{BATMAN} = 0,54$

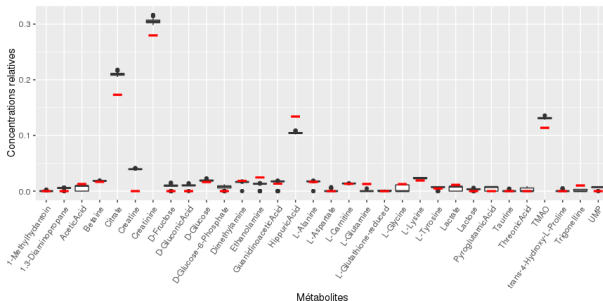


Figure – Concentration relatives estimées par la méthode ASICS

Sommaire

Étapes de l'analyse de données métabolomiques

Recherche des métabolites dont l'abondance évolue après un stress

Identification (et quantification) des métabolites

Conclusion

Analyses réalisées

- Caractérisation de la réponse au stress pour les 3 expériences
- Différence de réponse au stress entre les lignées et les générations
- Différences entre les générations et les lignées à *baseline*

Perspectives

Métabolome Refaire les analyses en utilisant les concentrations relatives obtenues grâce à ASICS

Intégration du génome Recherche de mQTL (métabolome) et de eQTL (transcriptome)

Interprétation biologique Interprétation à l'aide de réseaux biologiques (quelle méthode?)