



Intégration de données omiques sur une cohorte de patients obèses/diabétiques

Retour d'expériences d'utilisation de MixOmics::DIABLO et rgcca

21 Novembre 2023 – Journée StatOmique

Estelle Chatelain
Bilille - PLBS

En collaboration avec Guillemette Marot

Méthode RGCCA

- Méthode PLS, Partial Least Squares (deux blocs de données) (de Jong, 1993 ; M. Tenenhaus, 1998) :

$$\operatorname{argmax}_{\|a^h\|=1, \|b^h\|=1} \operatorname{cov}(Xa^h, Yb^h) \quad h = 1, \dots, H$$

- Méthode RGCCA, Regularised Generalised Canonical Correlation Analysis, qui permet de généraliser à J blocs de données (M. Tenenhaus, A. Tenenhaus, 2011) :

$$\operatorname{Maximize}_{\mathbf{a}_1, \dots, \mathbf{a}_J} \sum_{j,k=1, j \neq k}^J c_{jk} g(\operatorname{Cov}(\mathbf{X}_j \mathbf{a}_j, \mathbf{X}_k \mathbf{a}_k))$$

$$\text{subject to the constraints} \quad \tau_j \|\mathbf{a}_j\|^2 + (1 - \tau_j) \operatorname{Var}(\mathbf{X}_j \mathbf{a}_j) = 1, \quad j = 1, \dots, J.$$

DIABLO

- La méthode DIABLO, Data Integration Analysis for Biomarker discovery using Latent cOmponents, utilise la formule de la méthode SGCCA, Sparse Generalised Canonical Correlation Analysis (A. Tenenhaus, V. Guillemot, 2014) :

$$\left\{ \begin{array}{l} \operatorname{argmax}_{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_J} \sum_{j, k=1; j \neq k}^J c_{jk} g(\operatorname{cov}(\mathbf{X}_j \mathbf{a}_j, \mathbf{X}_k \mathbf{a}_k)) \\ \text{subject to } \|\mathbf{a}_j\|_2 = 1 \quad \text{and } \|\mathbf{a}_j\|_1 \leq s_j, \quad j = 1, \dots, J, \end{array} \right.$$

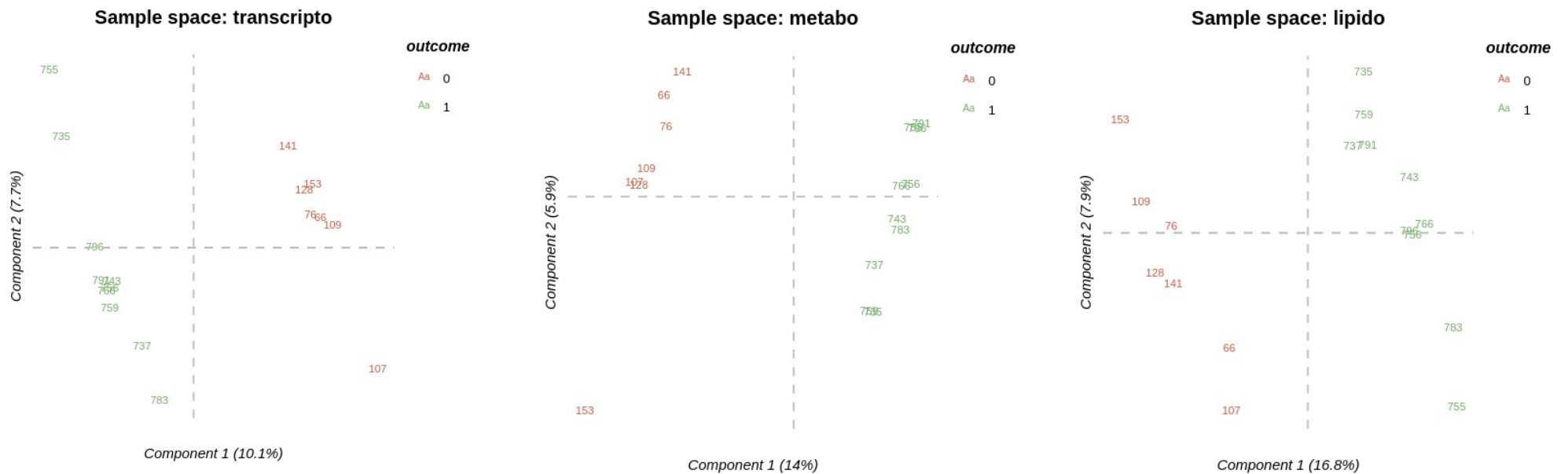
- Sélection de variables en utilisant la pénalisation L1.

Application

- Les données
 - confidentielles
 - 7 patients malades et 10 patients sains
 - 3 blocs de données omiques
 - Transcriptomiques du foie (22 K variables)
 - Métabolomiques (1,4 K variables)
 - Lipidomiques (800 variables)
- Les méthodes
 - MixOmics::DIABLO (version 6.20.0)
 - RGCCA via l'application Shiny (<https://iconics.shinyapps.io/rsgcca/>)

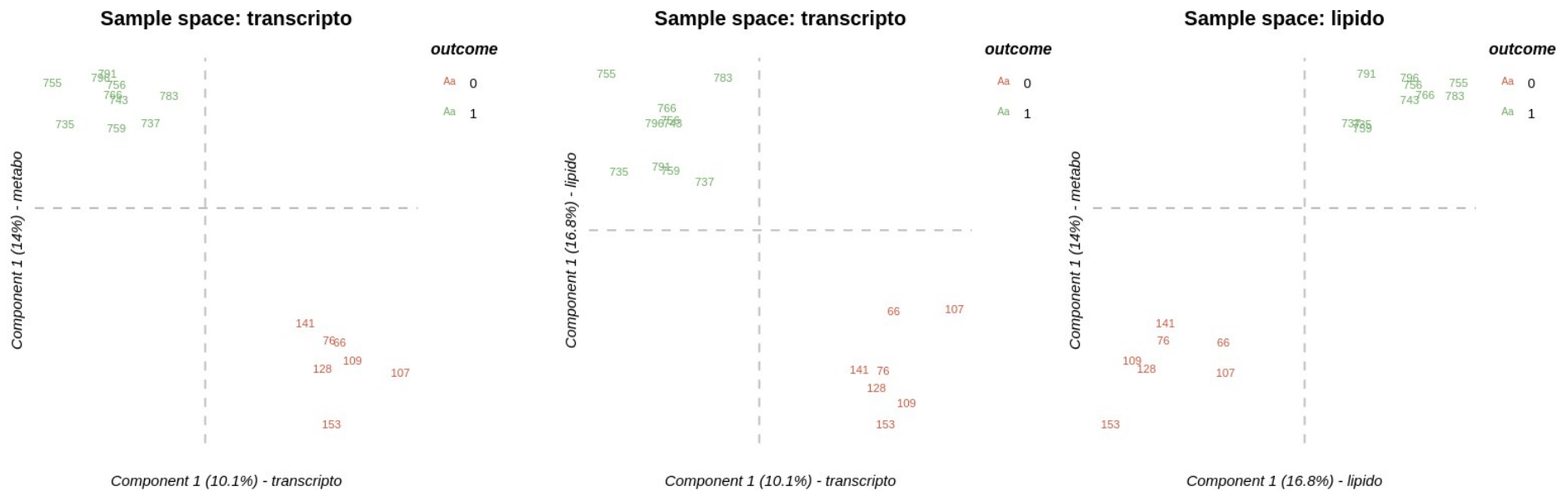
Représentation des individus

RGCCA - Shiny



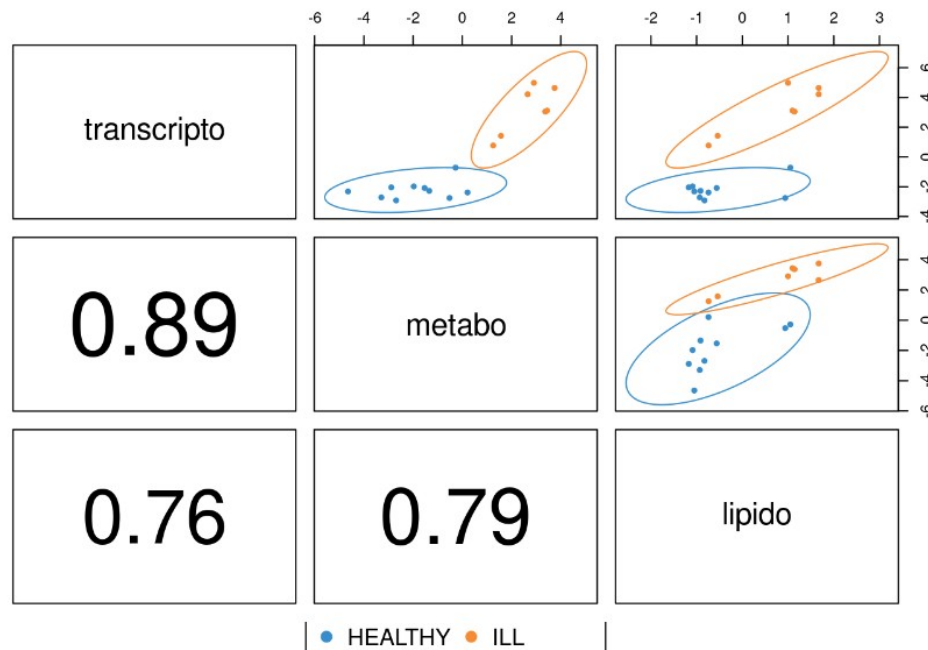
Représentation des individus

RGCCA - Shiny

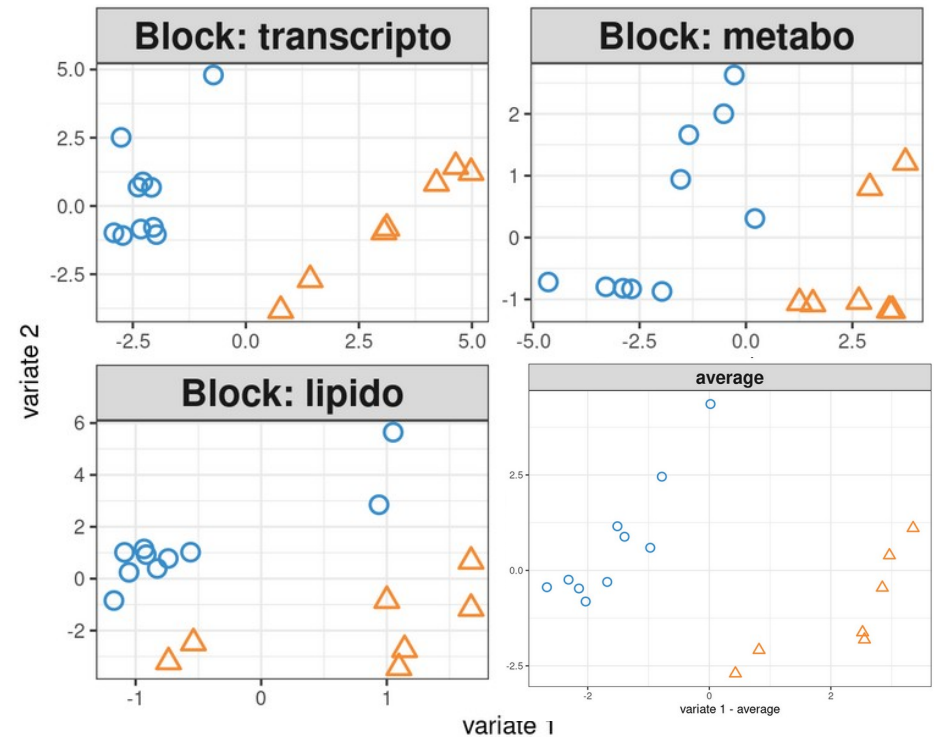


Représentation des individus

DIABLO



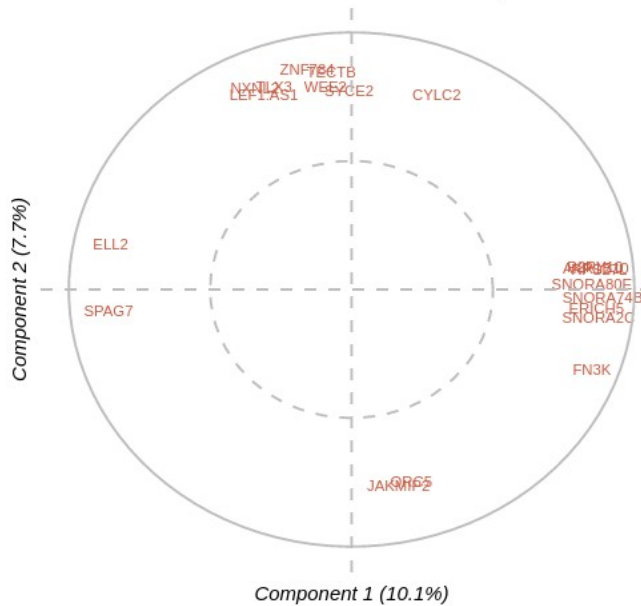
* Graphique corrigé après présentation informelle



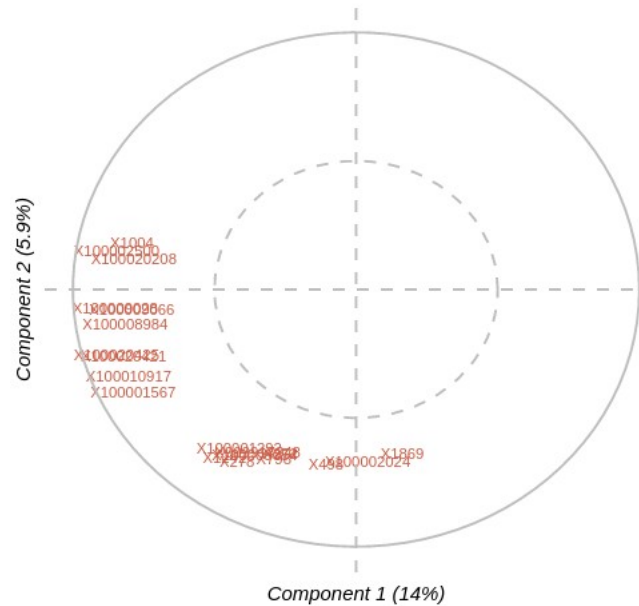
Représentation des variables

RGCCA - Shiny

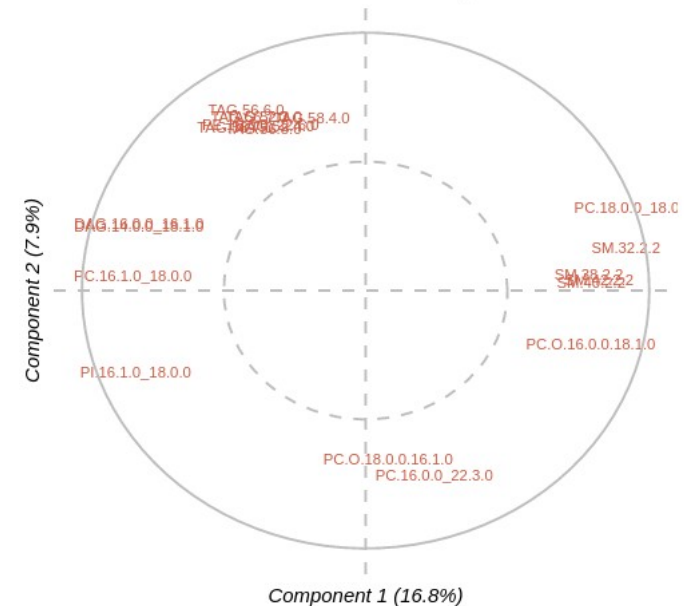
Correlation circle: transcripto



Correlation circle: metabo



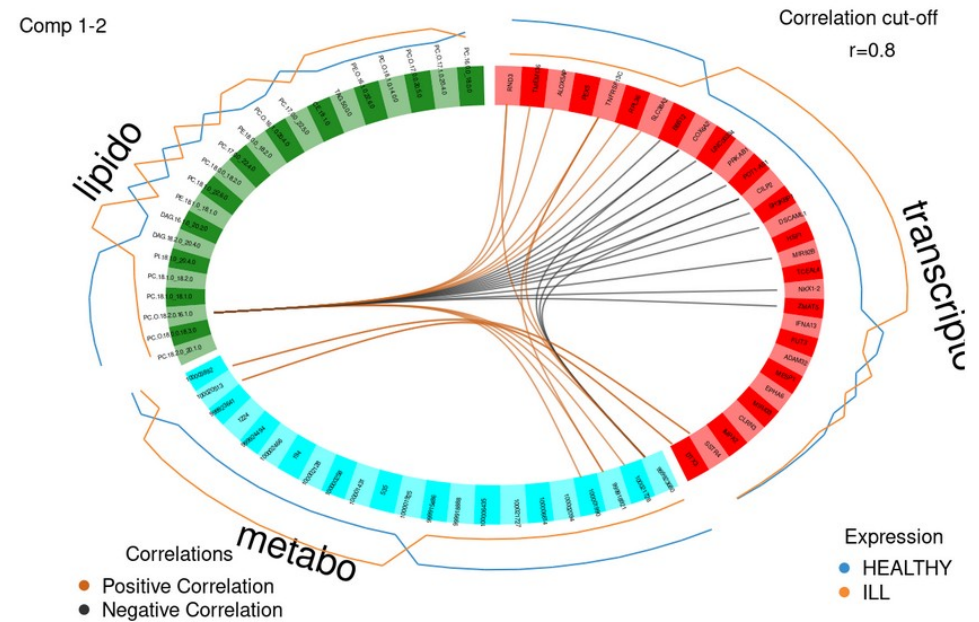
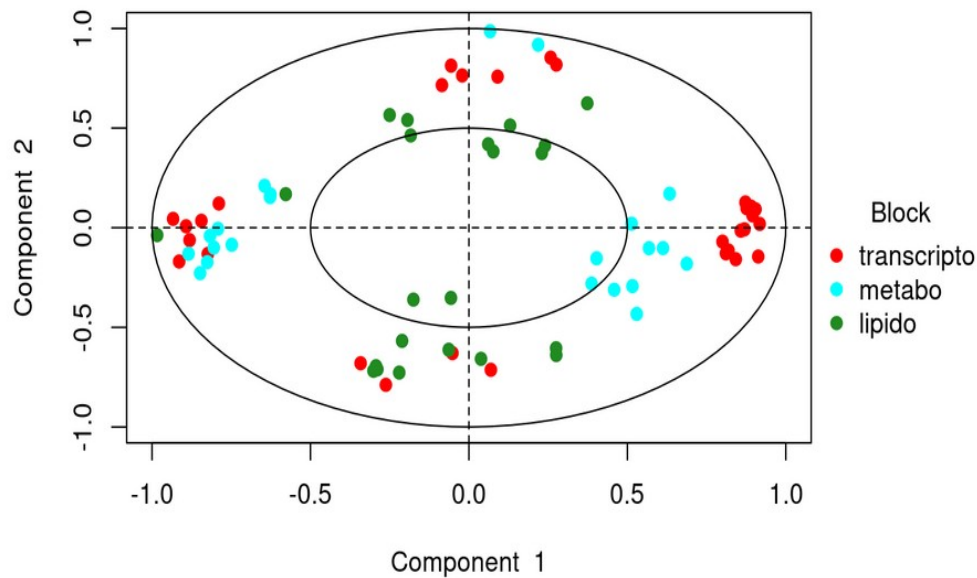
Correlation circle: lipido



Représentation des variables

DIABLO

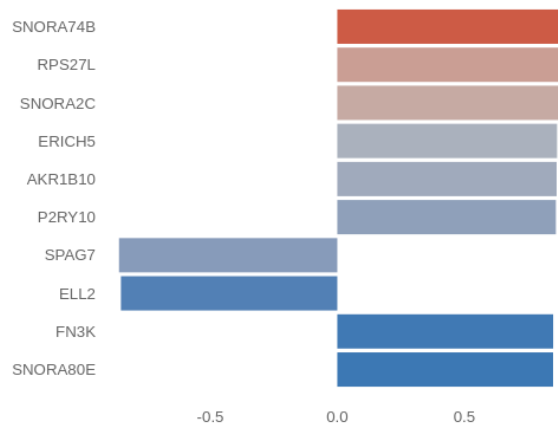
ABOS, DIABLO comp 1 - 2



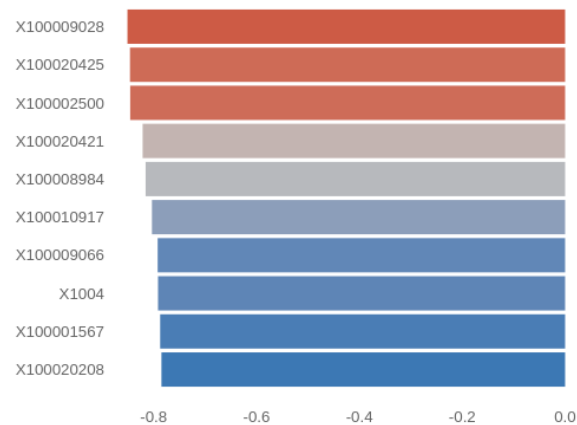
Contribution des variables

RGCCA - Shiny

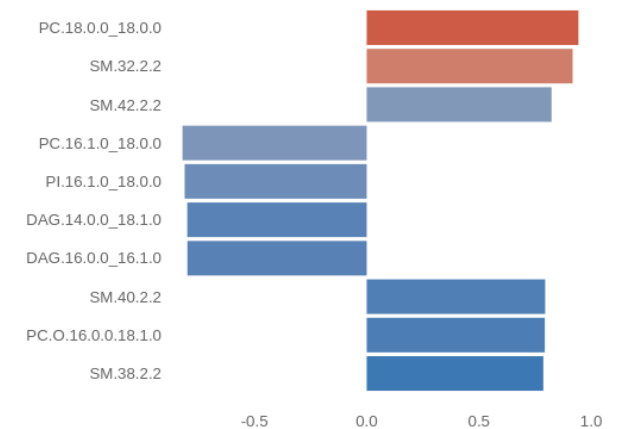
Variable correlations: transcripto with component 1 (10.1%)



Variable correlations: metabo with component 1 (14%)



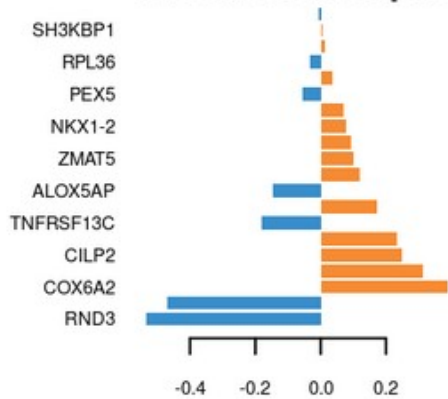
Variable correlations: lipido with component 1 (16.8%)



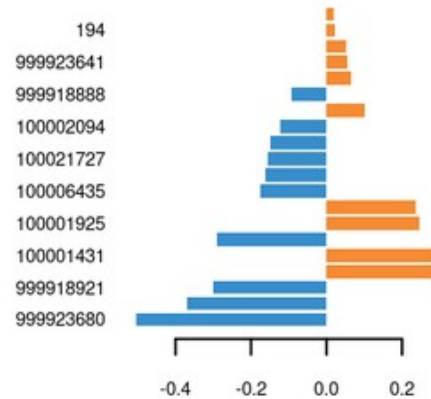
Contribution des variables

DIABLO

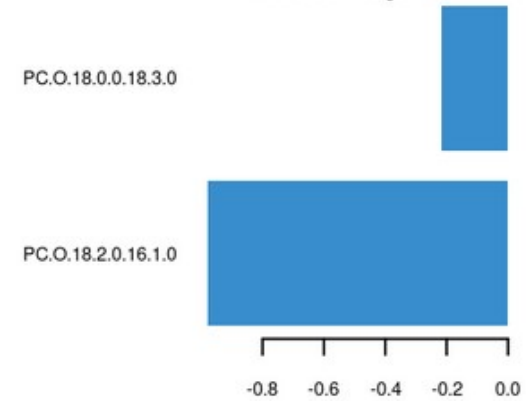
Contribution on comp 1
Block 'transcripto'



Contribution on comp 1
Block 'metabo'



Contribution on comp 1
Block 'lipido'



Conclusion

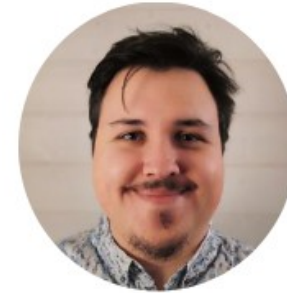
	mixOmics::DIABLO	shiny::RGCCA
scatter plot	yes	yes
correlation circle	yes	yes
circos plot	yes	no
network	yes	no
heatmpap	yes	no
loadings coordinates plot	yes	yes
shiny interface	no	yes

- Avantages de l'application Shiny :
 - Permet des zooms, graphes dynamiques, exports
 - Utilisable par les moins familiers de R (eg biologistes)
- Perspectives : comparer avec les graphes obtenus avec le package RGCCA, sans se restreindre à l'application Shiny.

Remerciements



U1190



StatOmique



Inria

