Bioindicators ammonia

Olivier Chapleur

Table of Contents

# Objectives

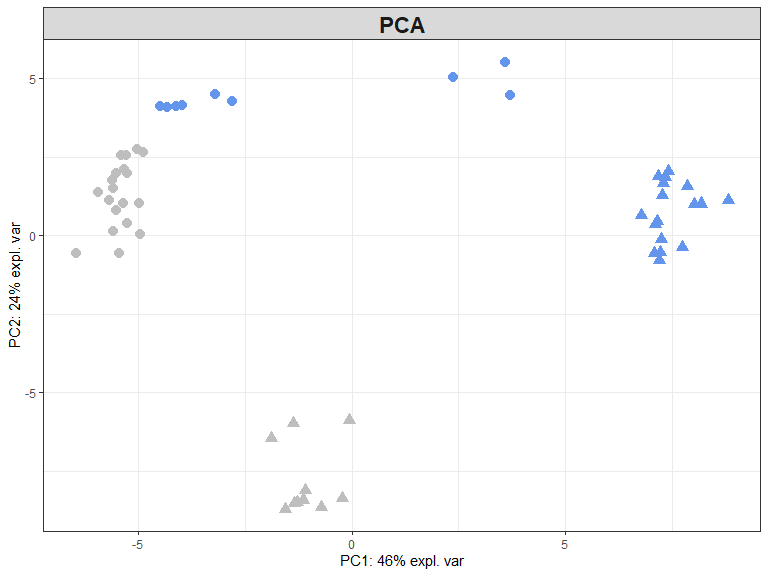
Different samples were produced in experiment testing the influence of ammonia (studies 1 to 4). The general purpose of this study is to assess if bioindicators of inhibition by ammonia can be identified within studies 1 and 2 and used to predict inhibition in samples of other studies (studies 3 and 4).

As 16S data was generated with different sequencing methodologies, it was agregated at the genus level.

# Data import

# Data transformation

# General view of the data with PCA



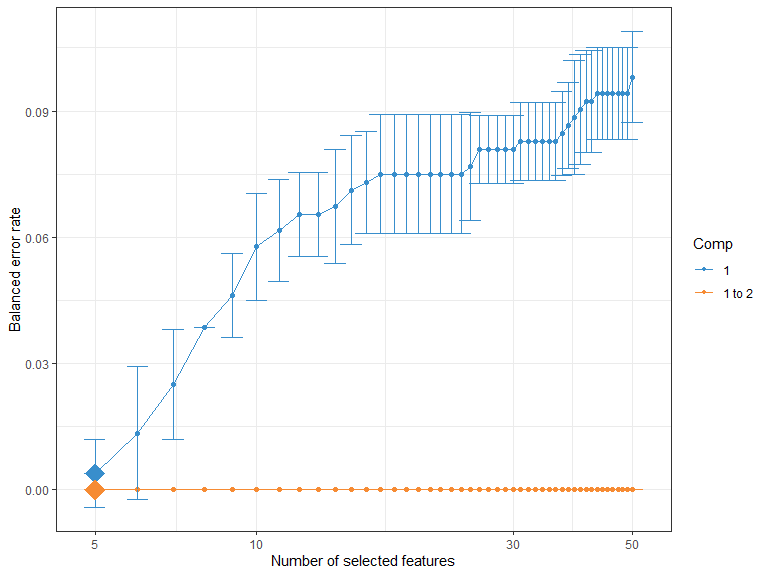
# Identification of biomarkers without taking into account the batch effect: sPLSDA and associated methods

Sparse PLSDA (sPLSDA) enables to select the variables discriminant of the groups of samples

## sPLSDA tuning (how many variables must be kept?)

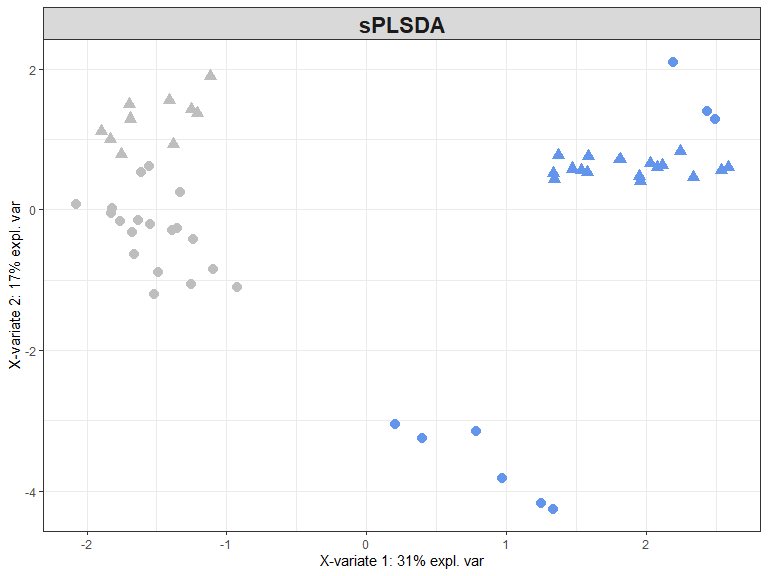
An error rate is calculated according to the number of variables kept on each component. Number of variables corresponding to the best error rate is selected.

## The plot shows the average error rate with respect to the keepX values tested:



After tuning the parameters, sPLSDA can be computed.

## sPLSDA visualisation



### sPLS-DA performance (prediction error rate)

splsda\_perf = perf(splsda, validation = 'Mfold', folds = 5,   
 progressBar = FALSE, nrepeat = 20)  
cat("Perf with Mfold validation, nrepat=20, folds=5")

## Perf with Mfold validation, nrepat=20, folds=5

splsda\_perf$error.rate

## $overall  
## max.dist centroids.dist mahalanobis.dist  
## comp1 0.006363636 0.004545455 0.004545455  
## comp2 0.000000000 0.000000000 0.000000000  
##   
## $BER  
## max.dist centroids.dist mahalanobis.dist  
## comp1 0.006730769 0.004807692 0.004807692  
## comp2 0.000000000 0.000000000 0.000000000

#plot(splsda\_perf)

# Identification of biomarkers while taking into account the batch effect: MINT sPLSDA

## study  
## Y Study 1 Study 2  
## Ammonia 9 17  
## No inhibition 19 10

### Tuning: how many components and variable to keep in the model?

The tuning uses leave one out cross validation.

## Calling 'tune.mint.splsda' with Leave-One-Group-Out Cross Validation (nrepeat = 1)

## $ncomp  
## NULL  
##   
## $values  
## comp1 comp2 comp3 comp4 comp5 comp6  
## Study 1 0.02631579 0.02631579 0.02631579 0.02631579 0.02631579 0.02631579  
## Study 2 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000

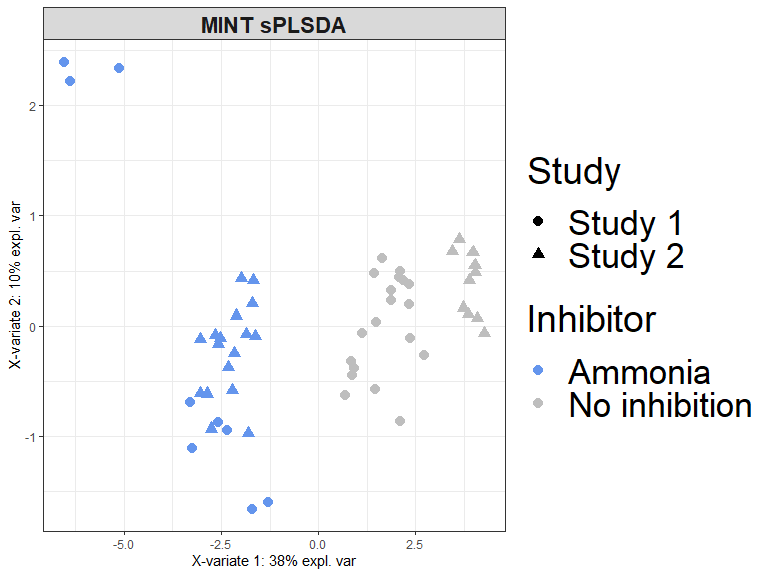
## comp1 comp2 comp3 comp4 comp5 comp6   
## 17 1 1 4 1 1

### Computation of MINT sPLSDA and projection of the samples on components 1 and 2

mint.splsda.res = mint.splsda(X = X, Y = Y, study = study, ncomp = 3,   
 keepX = tune.mint$choice.keepX)

## keepX is of length 6 while ncomp is 3   
## trimming keepX to [33mc(17,1,1)[39m

# cat("Variables kept on comp 1")  
# selectVar(mint.splsda.res, comp = 1)  
#   
# cat("Variables kept on comp 2")  
# selectVar(mint.splsda.res, comp = 2)



### Perfomance of the MINT sPLSDA model

The performance of the MINT sPLS-DA model can be assessed with the function. (Since the previous tuning of the MINT sPLS-DA model was conducted with the distance , the same distance is used to assess the performance of the final model.)

## $BER  
## max.dist centroids.dist mahalanobis.dist  
## comp1 0.01724138 0.2473475 0.2473475  
## comp2 0.01724138 0.2281167 0.1896552  
## comp3 0.01724138 0.2281167 0.1896552  
##   
## $overall  
## max.dist centroids.dist mahalanobis.dist  
## comp1 0.01818182 0.2545455 0.2545455  
## comp2 0.01818182 0.2363636 0.2000000  
## comp3 0.01818182 0.2363636 0.2000000  
##   
## $error.rate.class  
## $error.rate.class$max.dist  
## comp1 comp2 comp3  
## Ammonia 0.00000000 0.00000000 0.00000000  
## No inhibition 0.03448276 0.03448276 0.03448276  
##   
## $error.rate.class$centroids.dist  
## comp1 comp2 comp3  
## Ammonia 0.1153846 0.07692308 0.07692308  
## No inhibition 0.3793103 0.37931034 0.37931034  
##   
## $error.rate.class$mahalanobis.dist  
## comp1 comp2 comp3  
## Ammonia 0.1153846 0.0000000 0.0000000  
## No inhibition 0.3793103 0.3793103 0.3793103

## max.dist centroids.dist mahalanobis.dist  
## comp1 0.01724138 0.2473475 0.2473475  
## comp2 0.01724138 0.2281167 0.1896552  
## comp3 0.01724138 0.2281167 0.1896552

# Prediction of Lufan and Peng data, model on one component

## Calling 'tune.mint.splsda' with Leave-One-Group-Out Cross Validation (nrepeat = 1)

## Summary of the results of the prediction.

## predicted.as.Ammonia predicted.as.No inhibition  
## Ammonia 20 6  
## No inhibition 3 17

## Detailed results of the prediction.

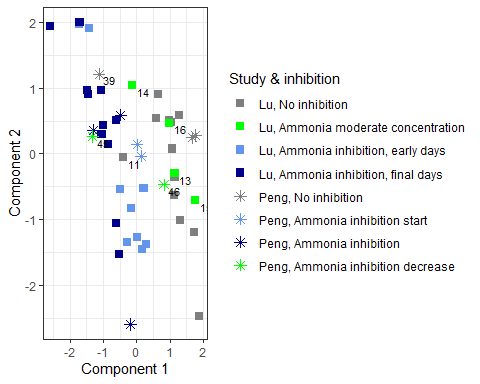
# here I try to extract the coordinate of initial and predicted samples to draw a figure with everything

## Calling 'tune.mint.splsda' with Leave-One-Group-Out Cross Validation (nrepeat = 1)

background2=background.predict(object=mint.splsda.res, comp.predicted = 1, dist='max.dist', xlim=c(-3,3),ylim = c(-3,3))

ordered\_factor=factor( factor(metadata\_studies\_3\_4$prediction), levels(factor(metadata\_studies\_3\_4$prediction))[c(4,3,1,2,8,7,5,6)])  
  
# for(i in 1:length(background2))  
# {  
# if(!is.null(background2[[i]]))  
# background2[[i]]=data.frame(id=i,col=names(background2)[i],  
# background2[[i]])  
# }  
#   
# background2 = do.call(rbind,background2)  
  
#tiff(filename = "figure\_5.tiff", width = 3000, height = 1800, res = 400,pointsize=12)  
  
  
 ggplot(data=predit, aes(x=dim1, y=dim2))+  
 # geom\_polygon(data = background2,aes(x=Var1, y=Var2,fill = col), inherit.aes = FALSE, show.legend =TRUE)+  
 scale\_fill\_manual(values = c("lavender", "grey95"), name='Prediction')+  
 geom\_point(aes(color=ordered\_factor, shape=ordered\_factor), size=2.5)+  
 theme(legend.title=element\_blank()) +   
 scale\_color\_manual(name = "Study & inhibition", values=c("grey50", "green","cornflowerblue", "darkblue","grey50" , "cornflowerblue", "darkblue", "green" ), labels=levels(droplevels(ordered\_factor))) +   
 scale\_shape\_manual(name = "Study & inhibition", values=c(15,15,15, 15, 8,8,8,8), labels=levels(droplevels(ordered\_factor))) +  
 labs( y="Component 2", x = "Component 1",color="Inhibition")+  
 theme\_bw()+  
theme(strip.text= element\_text(size=16, face='bold'))+  
 geom\_text(label=metadata\_studies\_3\_4$prediction\_code\_plot,hjust=-0.4,vjust=1.2, cex=3)

## Warning: Removed 38 rows containing missing values (geom\_text).



#dev.off()