Replication Using MetaboAnalystR

Replicating PLSDA Analysis Using NMR Spectral Data

# 1. MetaboAnalystR

The purpose of this repository is to serve as the home of the replication project. All relevant files, data, and code will be made available here in order to facilitate transparency and further reproducibility of this work. In this regard, the aim will be to minimize barriers and time required to perform future analyses.

There are a number of data processing challanges germane to metabolomics. Some notable ones tend to include specialized knowledged and software, which is why there is much value to be had in lowering the accessibility barrier for analysis. To help simplify the process for some of the more common metabolomic data analyses, MetaboAnalyst offers an accessible tool capable of handling most kinds of metabolomic data. At least in regards to NMR/MS spectral data, our project intends replicate the work used to support this claim.

# 2. Partial Least Square Discriminant Analysis (PLSDA)

# 3. Replicating PLSDA Analysis of NMR Spectral Bin Data

## here() starts at /Users/medhasagar/Documents/ReproProj

## Loading required package: lattice

## Loading required package: pls

##   
## Attaching package: 'pls'

## The following object is masked from 'package:stats':  
##   
## loadings

## Loading required package: data.table

## Warning in fun(libname, pkgname): mzR has been built against a different Rcpp version (1.0.2)  
## than is installed on your system (1.0.3). This might lead to errors  
## when loading mzR. If you encounter such issues, please send a report,  
## including the output of sessionInfo() to the Bioc support forum at   
## https://support.bioconductor.org/. For details see also  
## https://github.com/sneumann/mzR/wiki/mzR-Rcpp-compiler-linker-issue.

## Warning: replacing previous import 'xcms::plot' by 'graphics::plot' when loading  
## 'CAMERA'

## Warning: replacing previous import 'CAMERA::annotate' by 'ggplot2::annotate'  
## when loading 'MetaboAnalystR'

## Warning: replacing previous import 'MSnbase::normalize' by 'igraph::normalize'  
## when loading 'MetaboAnalystR'

## Warning: replacing previous import 'reshape::recast' by 'reshape2::recast' when  
## loading 'MetaboAnalystR'

## Warning: replacing previous import 'reshape::melt' by 'reshape2::melt' when  
## loading 'MetaboAnalystR'

## Warning: replacing previous import 'reshape::colsplit' by 'reshape2::colsplit'  
## when loading 'MetaboAnalystR'

## Warning: replacing previous import 'igraph::normalize' by 'xcms::normalize' when  
## loading 'MetaboAnalystR'

## Warning: replacing previous import 'igraph::groups' by 'xcms::groups' when  
## loading 'MetaboAnalystR'

## [1] "MetaboAnalyst R objects initialized ..."  
## [1] "Successfully passed sanity check!"   
## [2] "Samples are not paired."   
## [3] "2 groups were detected in samples."   
## [4] "Only English letters, numbers, underscore, hyphen and forward slash (/) are allowed."   
## [5] "<font color=\"orange\">Other special characters or punctuations (if any) will be stripped off.</font>"  
## [6] "All data values are numeric."   
## [7] "A total of 0 (0%) missing values were detected."   
## [8] "<u>By default, these values will be replaced by a small value.</u>"   
## [9] "Click <b>Skip</b> button if you accept the default practice"   
## [10] "Or click <b>Missing value imputation</b> to use other methods"   
## [1] " Further feature filtering based on Interquantile Range"  
## [1] " Further feature filtering based on Interquantile Range"

## [1] "The Interactive 3D PLS-DA plot has been created, please find it in mSet$imgSet$plsda.3d."

## Loading required package: ggplot2

##   
## Attaching package: 'caret'

## The following object is masked from 'package:MetaboAnalystR':  
##   
## splsda

## The following object is masked from 'package:pls':  
##   
## R2

## Warning in RColorBrewer::brewer.pal(25, colorpalette): n too large, allowed maximum for palette RdYlGn is 11  
## Returning the palette you asked for with that many colors

## Warning in RColorBrewer::brewer.pal(25, colorpalette): n too large, allowed maximum for palette RdYlGn is 11  
## Returning the palette you asked for with that many colors

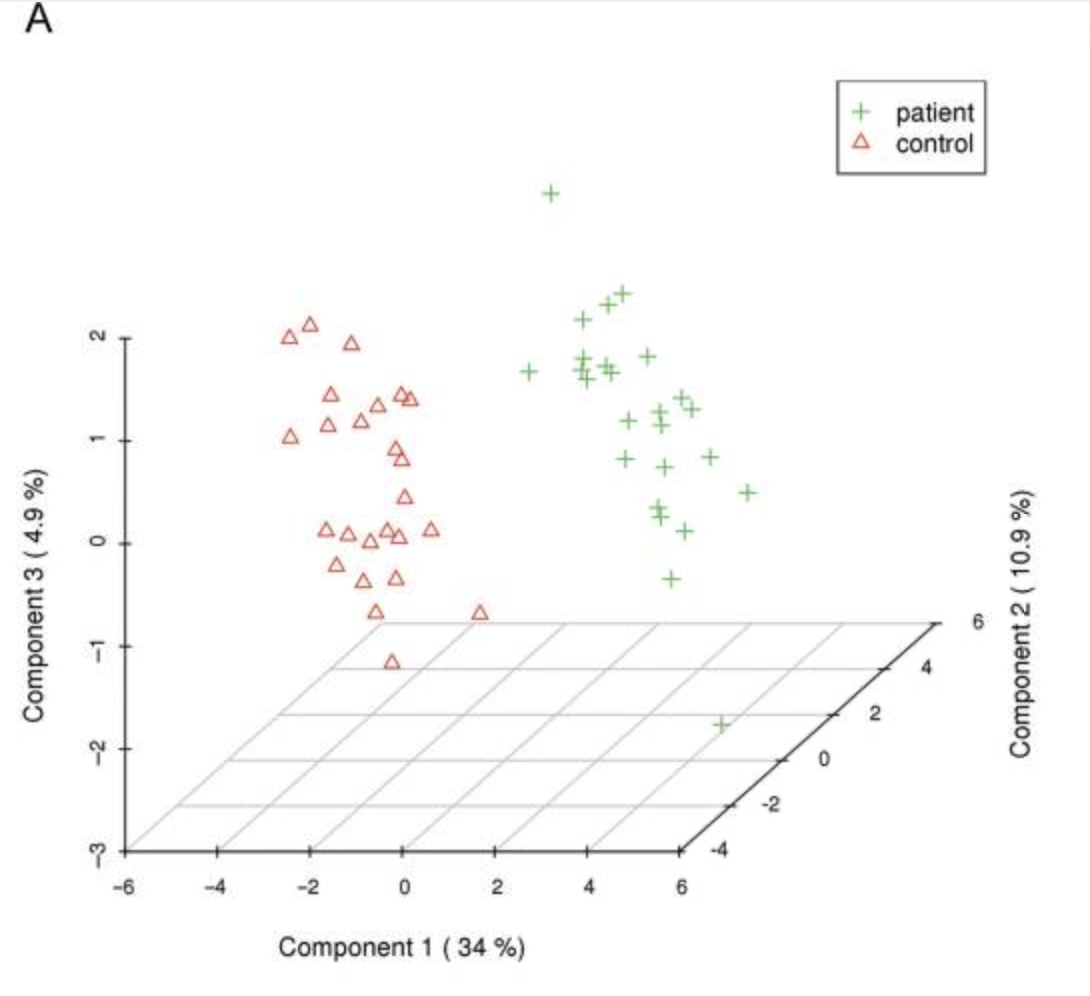


Image to be replicated.

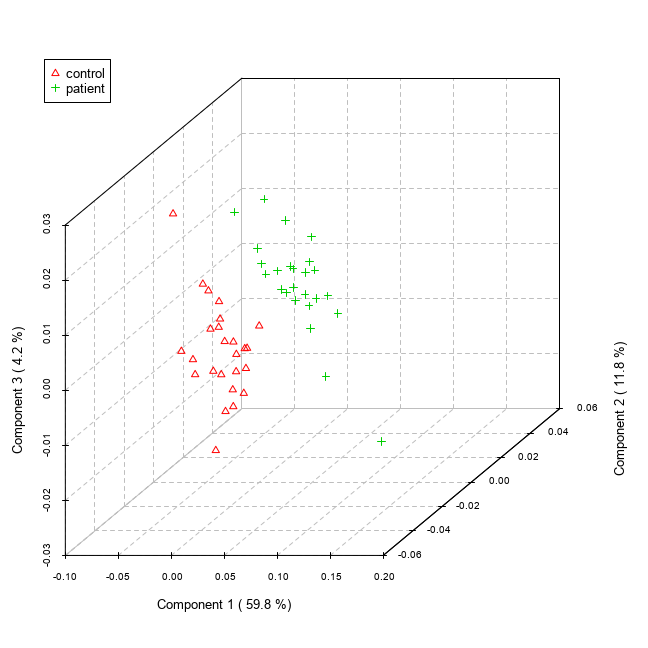


Image replicated.

# 4. Change

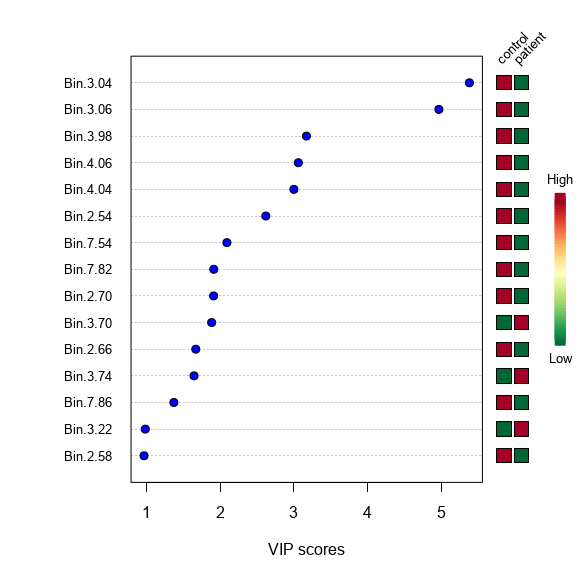
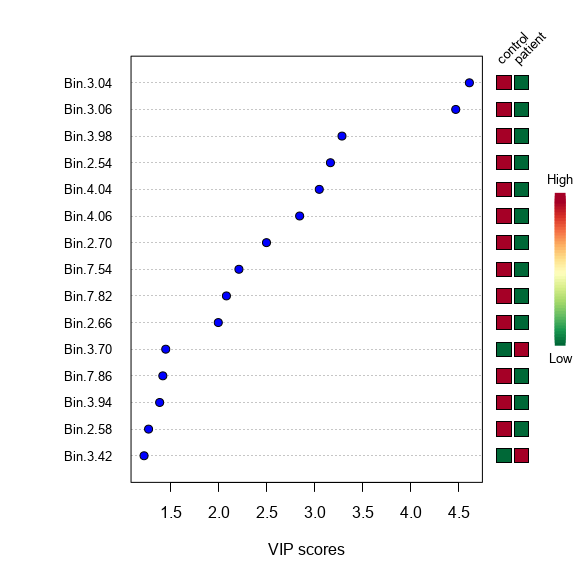
## 4.1 Change in imputation method

We will be changing two technical details in our analysis. First, the default imputation method for columns with less than 50% of missing values is set to “min”. This method assigns half of the column’s minimum positive value to the missing observations. The assumption of this approach is that most missing values occurred because the levels of abundance metabolites are below the detection limit. This assumption is generalized for all the possible datasets used in the MetaboAnalyst website. We decided to change the imputation method to “mean”, a more widely used method that imputes the mean value of the column to the missing values. This change shouldn’t impact the output in a significant manner. Due to the nature of the dataset used, data provided for the testing of the tool, we believe that it should not make a difference.

## 4.2 Change in filtering method

The second change we implemented was in the filtering of the data. The goal of this step is to identify the variables that are less likely to be used in the final model. The default method to filter the variables is by Interquantile Range (IQR). What we propose is to change this method to Standard Deviation (SD) as, like IQR, it performs well when identifying non-informative variables that mantain near-constant values throughout the entire experiment.

## 4.3 Change in Component for Variable Importance

# 5. Conclusion

# 6. References