Heart Mass Spectrsocopy Data Analysis: Gender impact on HFD [Test Project]

Md Ismail Hossain, M. Sc.

2024-01-22

Key points to remember for metabolomics data analysis:

Raw data processing:

- The MS data converted and centroied by using MSconvert (https://proteowizard.sourceforge.io/) in mzML format.
- The mzML file then converted to feature matrix using Nextflow (metaboIGNITER). But metaboIGNITER cannot provide metabolites name for MS1 data, instead it gives ids only. Which is a major drawback of using metaboIGNITER.
- The alternative to metaboIGNITER is using MetaboanalystR for mzML to feateature matrix but that process also stacked at certain points (this function not executing when following their tutorial :PerformDBSearchingBatch).
- At this a result we could not find a smooth analysis pipeline to convert the raw mzML to a feature matrix to start the statistical analysis section.

Statistical Analysis:

- Using MetaboAnalystR we can run all the required statistical analysis in our local environment if someone provide us the feature matrix with the peaks read.
- For the pathway analysis, MetaboAnalystR webpage is very powerful and easy to use. Because it save time and give us a very well written repot of the analysis automatically. As a result for pathway analysis using webpage is more logical.
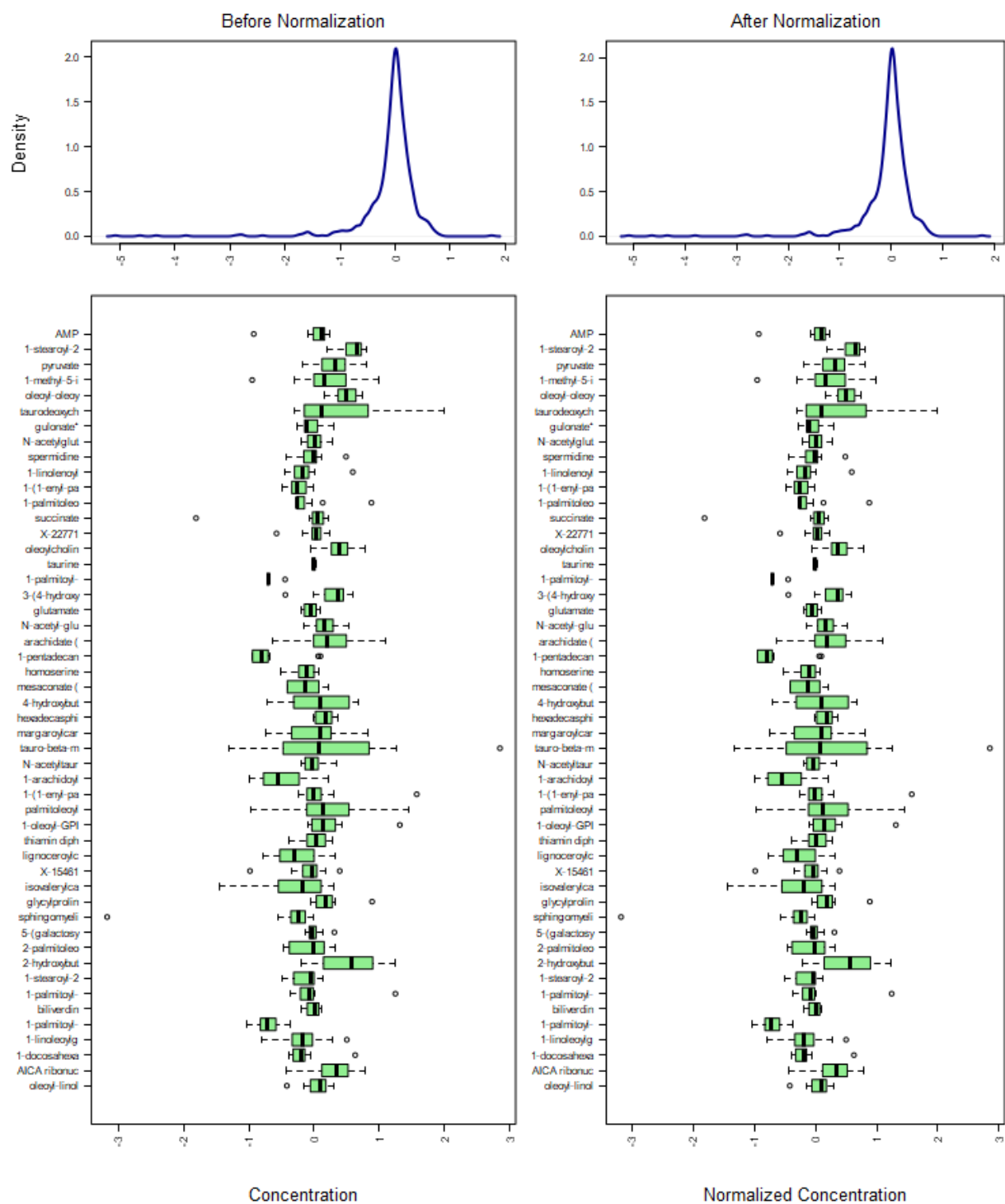
Future Direction

- We need to find alternative way to process the raw data which can give us metabolites name instead the ids.
- We can keep an eye on metaboIGNITER and also see some alternative solution in Python.

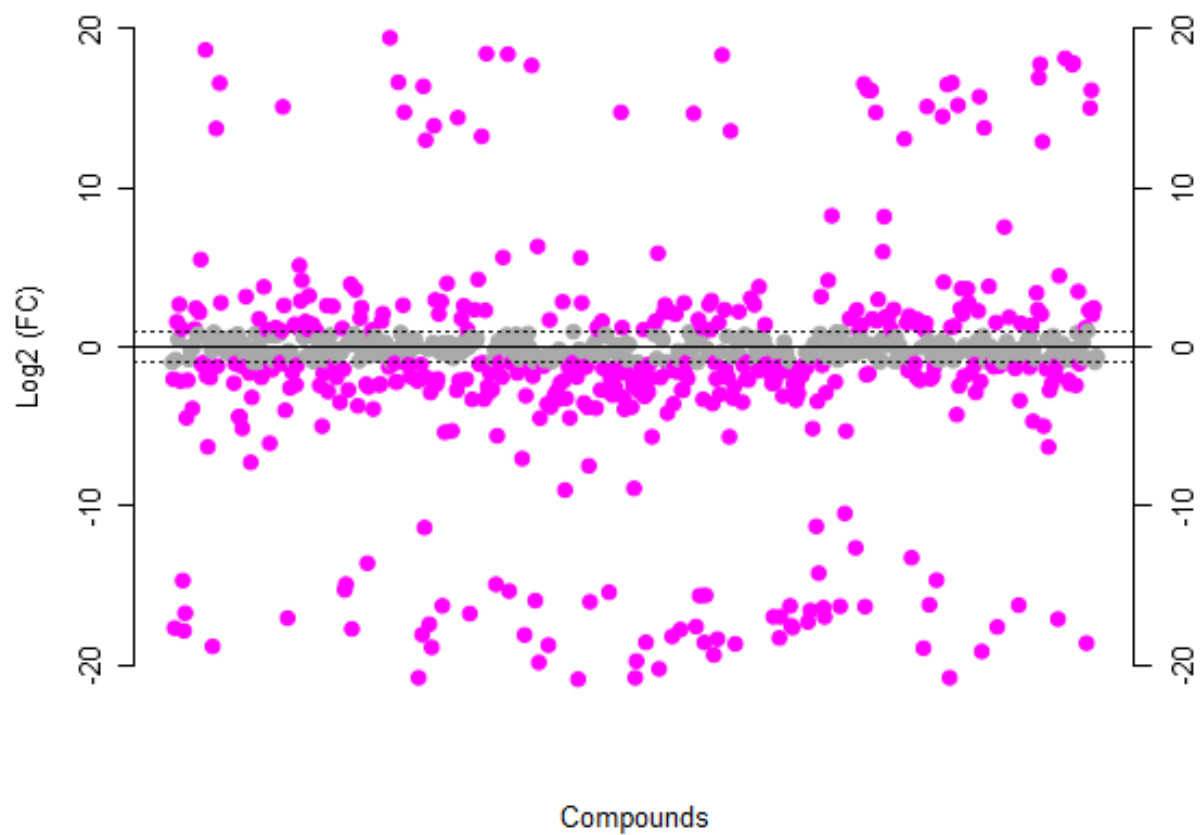# Loading the data and prepare for further steps:

We used a very specific data format to start the analysis procedure.

```
## Starting Rserve...
##  "C:\Users\mho004\AppData\Local\R\WIN-LI~1\4.3\Rserve\libs\x64\Rserve.exe" --no-save
## [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.
##  [6] "All data values are numeric."
##  [7] "A total of 0 (0%) missing values were detected."
##  [8] "<u>By default, missing values will be replaced by 1/5 of min positive values of their correspon
##  [9] "Click the <b>Proceed</b> button if you accept the default practice;"
## [10] "Or click the <b>Missing Values</b> button to use other methods."
```
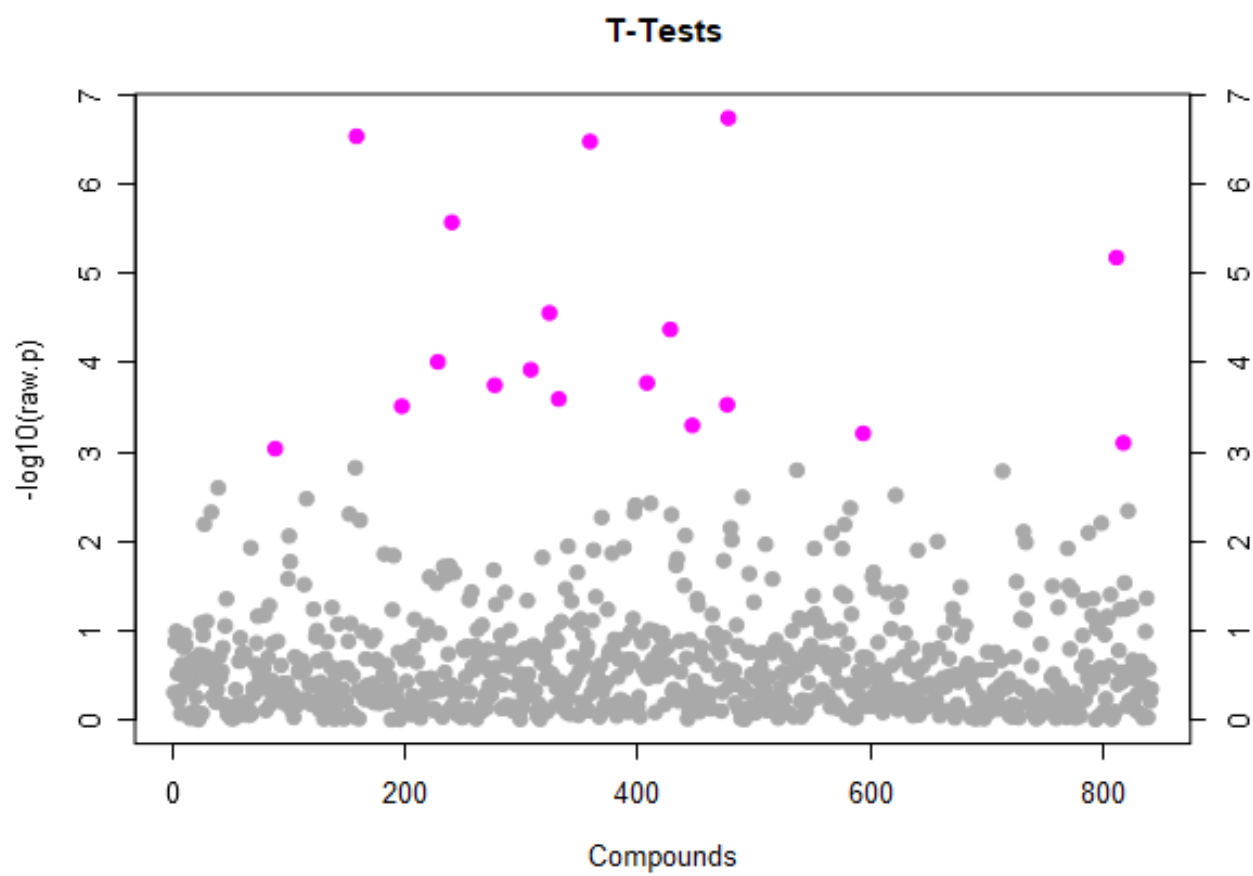
Before Normalization — Density vs. Concentration

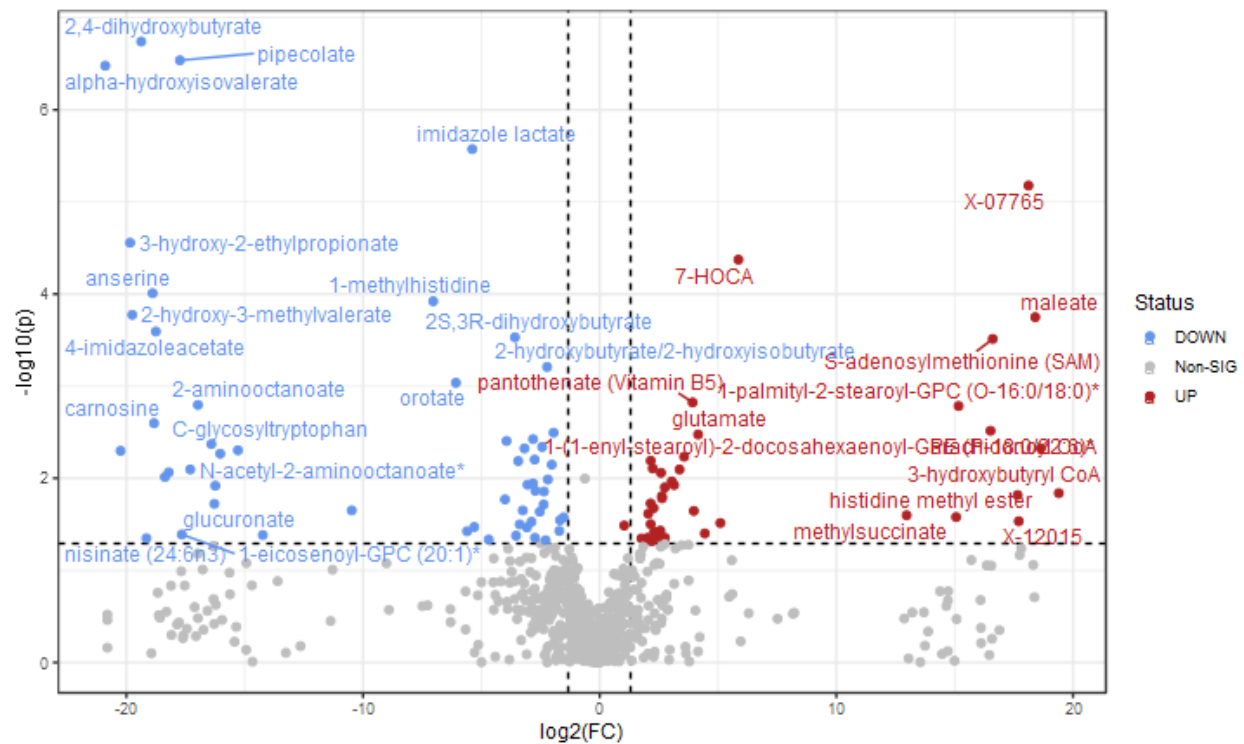After Normalization — Density vs. Normalized Concentration

3

Fold-change analysis

T-test (FDR-adjusted p-values)

```
## [1] "Performing regular t-tests ...."
## [1] "A total of 18 significant features were found."
```

T-Tests

Volcano Plot (FDR-adjusted p-values)

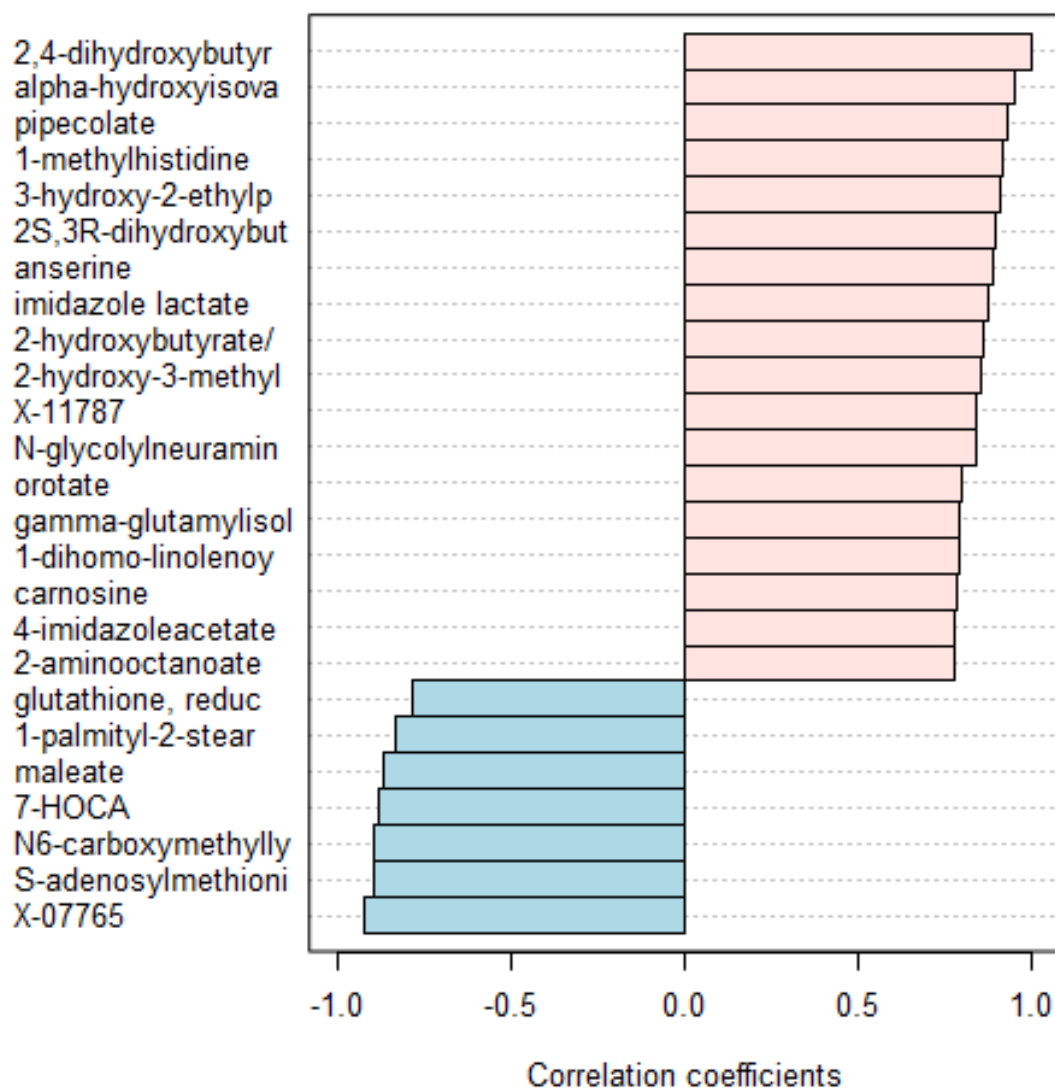```
[1] "A total of 106 significant features were found."
```

## Correlation Analysis

A small subset of metabolites is ideal to do correlation analysis otherwise it's hard to interpret the results.
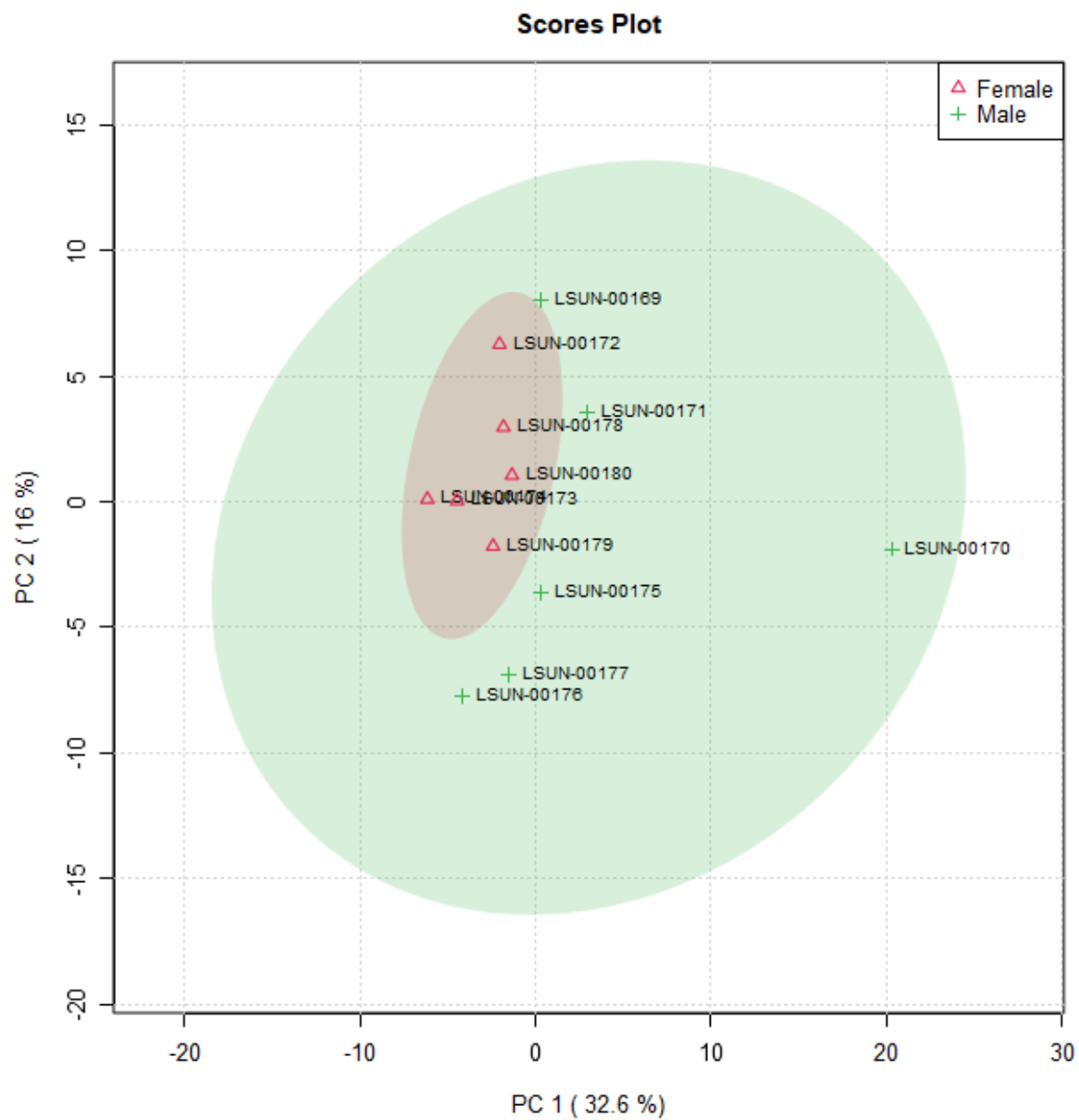
# Pattern Searching for 2,4-dihydroxybutyrate
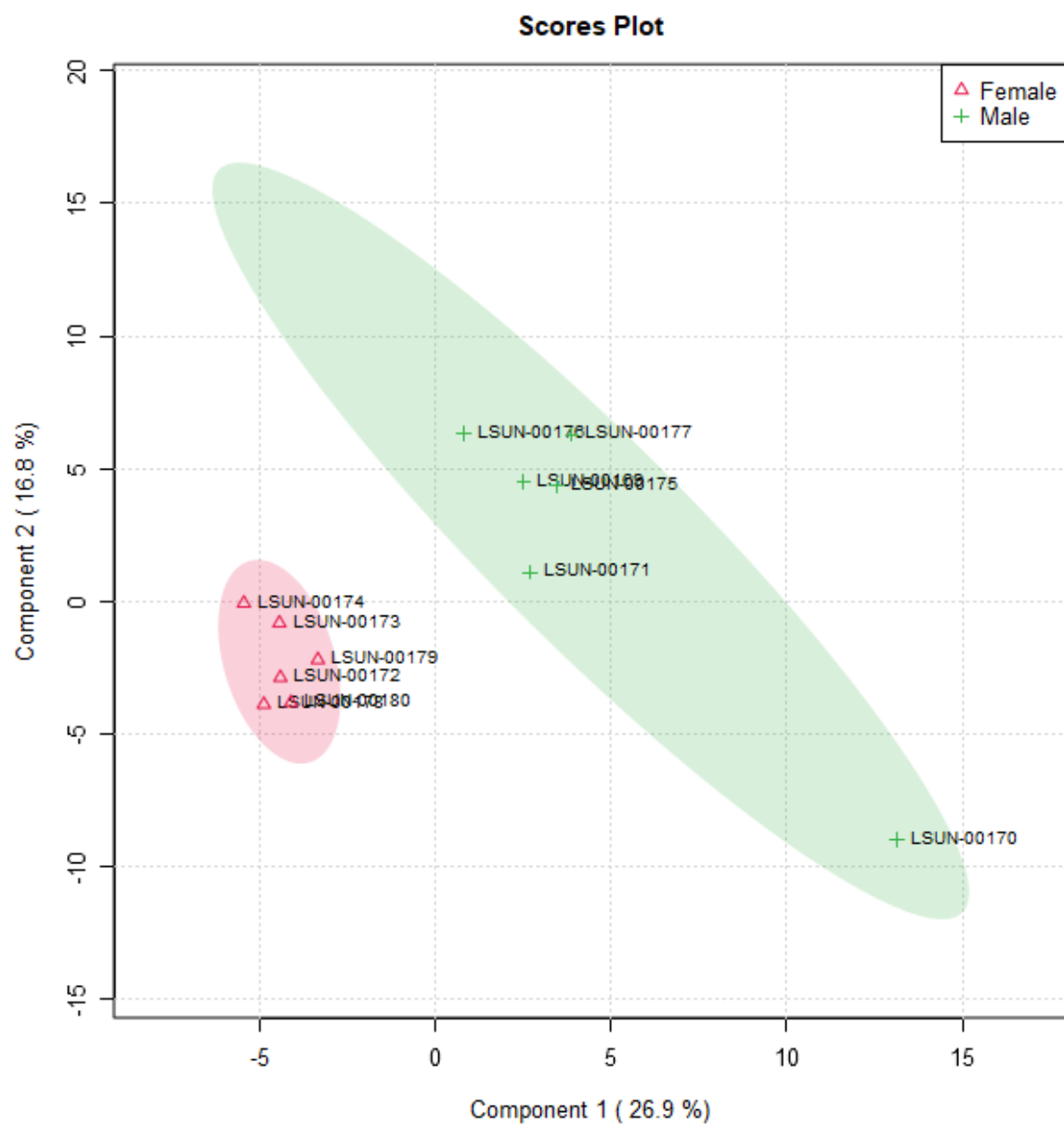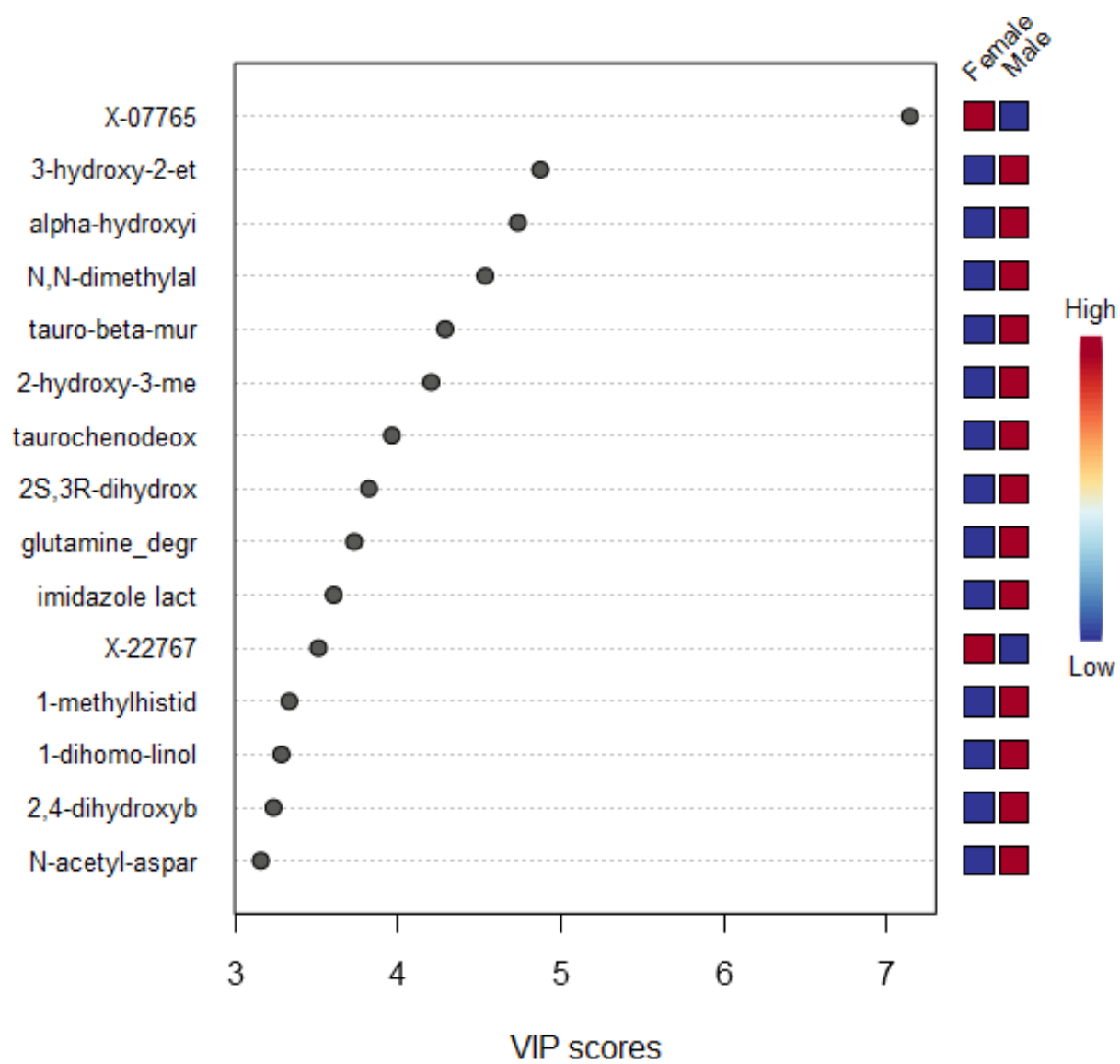
**Top 25 compounds correlated with the 2,4-dihydroxybutyra**



| | |
|---|---|
| 2,4-dihydroxybutyr | |
| alpha-hydroxyisova | |
| pipecolate | |
| 1-methylhistidine | |
| 3-hydroxy-2-ethylp | |
| 2S,3R-dihydroxybut | |
| anserine | |
| imidazole lactate | |
| 2-hydroxybutyrate/ | |
| 2-hydroxy-3-methyl | |
| X-11787 | |
| N-glycolylneuramin | |
| orotate | |
| gamma-glutamylisol | |
| 1-dihomo-linolenoy | |
| carnosine | |
| 4-imidazoleacetate | |
| 2-aminooctanoate | |
| glutathione, reduc | |
| 1-palmityl-2-stear | |
| maleate | |
| 7-HOCA | |
| N6-carboxymethylly | |
| S-adenosylmethioni | |
| X-07765 | |

Correlation coefficients
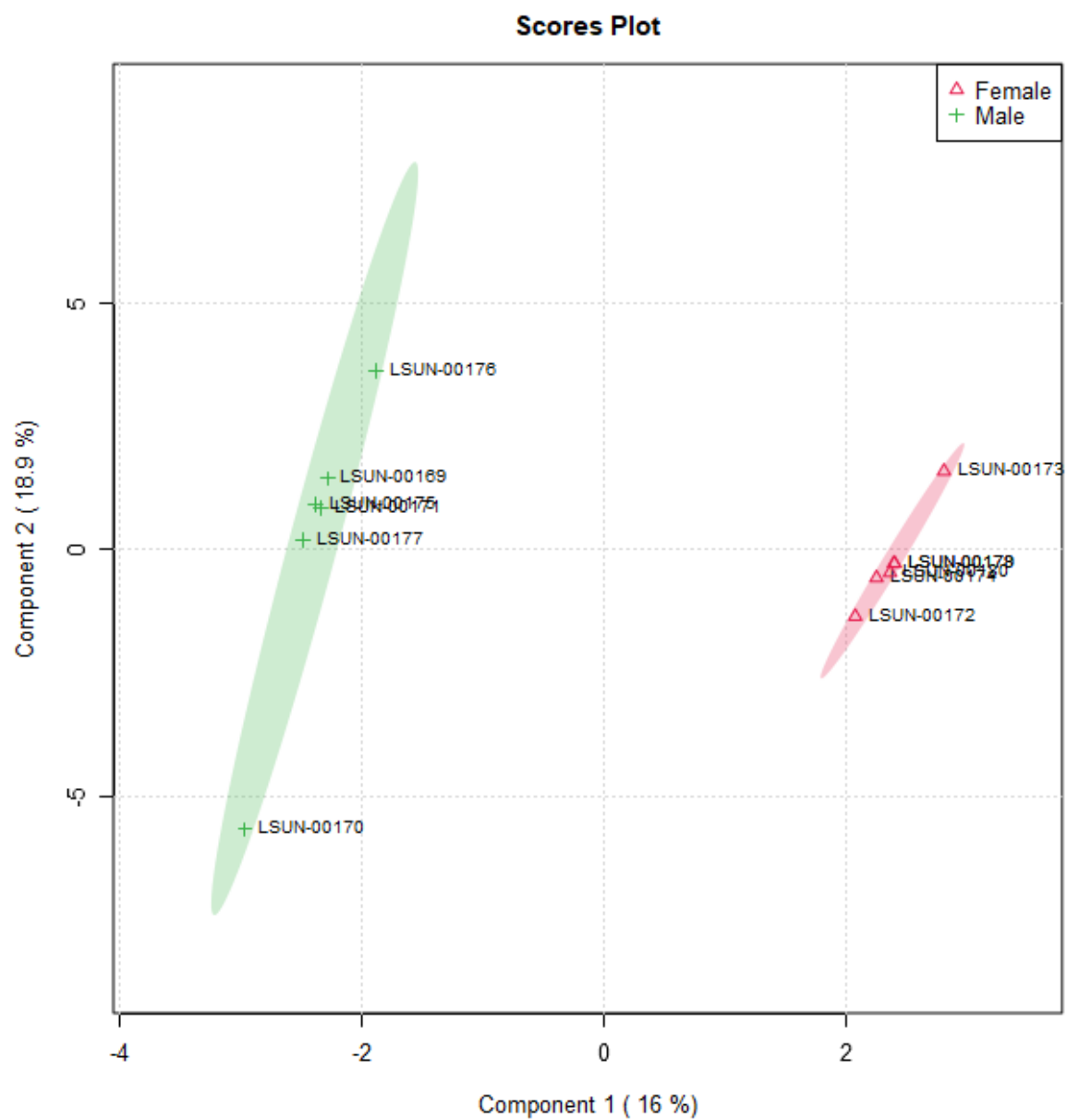
# Principal component analysis



Scores Plot

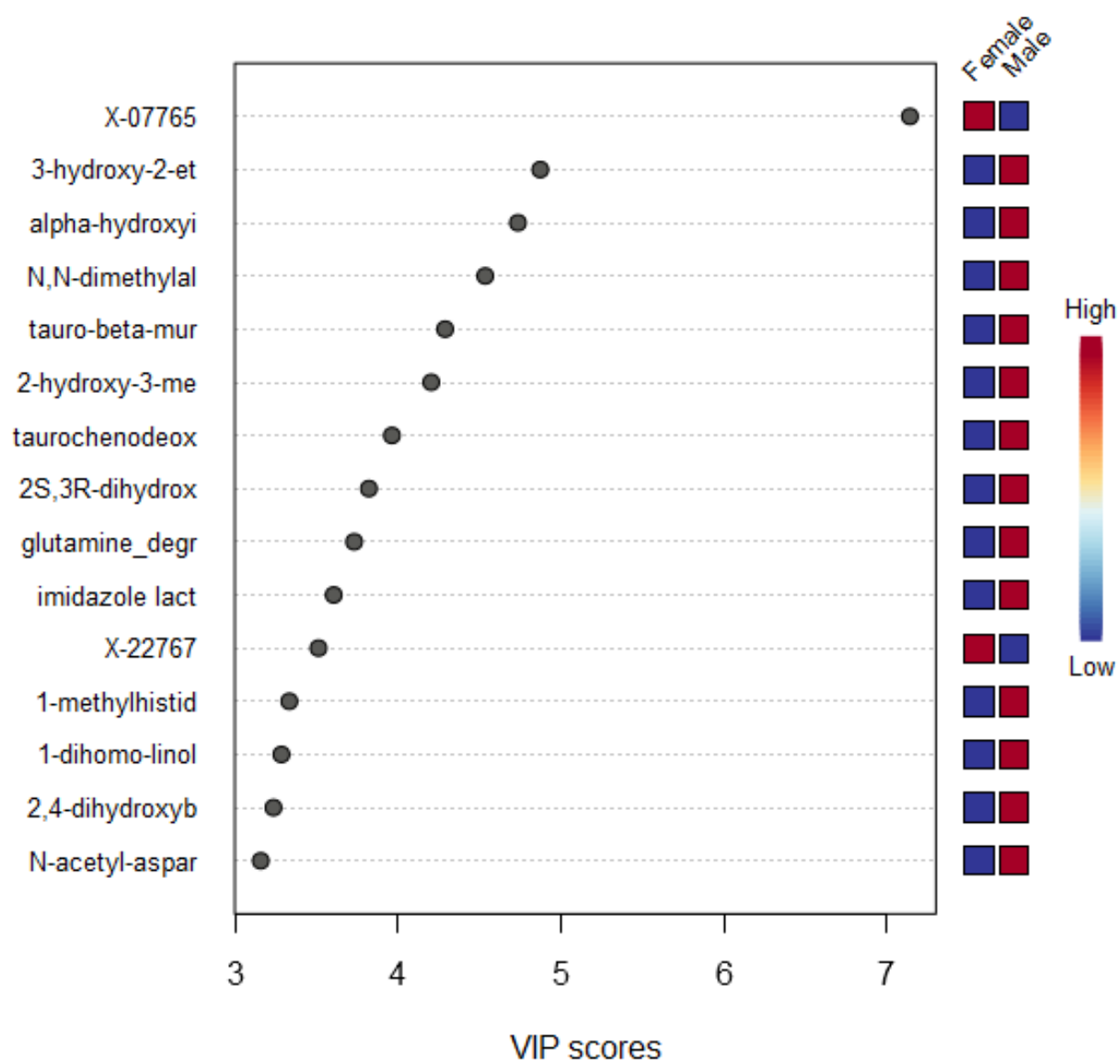# Partial Least Squares - Discriminant Analysis (PLS-DA)

```
[1] "performing 100 permutations ..."
[1] "Empirical p value: p = 1 (100/100)"
```
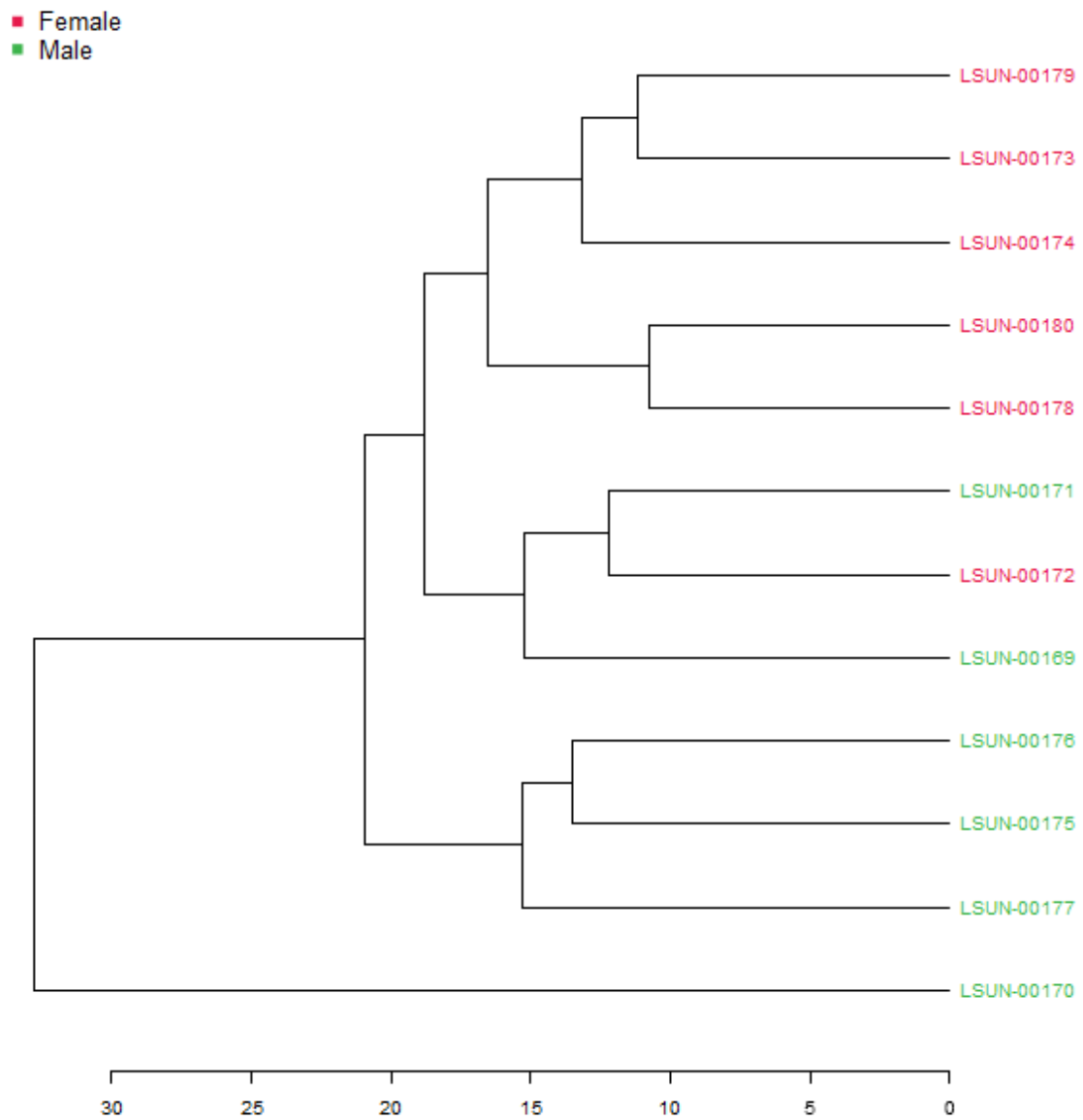
**Scores Plot**

△ Female
+ Male

LSUN-00176 LSUN-00177

LSUN-00171

LSUN-00174
LSUN-00173
LSUN-00179
LSUN-00172
LSUN-00180

LSUN-00170

Component 1 ( 26.9 % )

Component 2 ( 16.8 % )

# Sparse Partial Least Squares - Discriminant Analysis (sPLS-DA)

**Scores Plot**

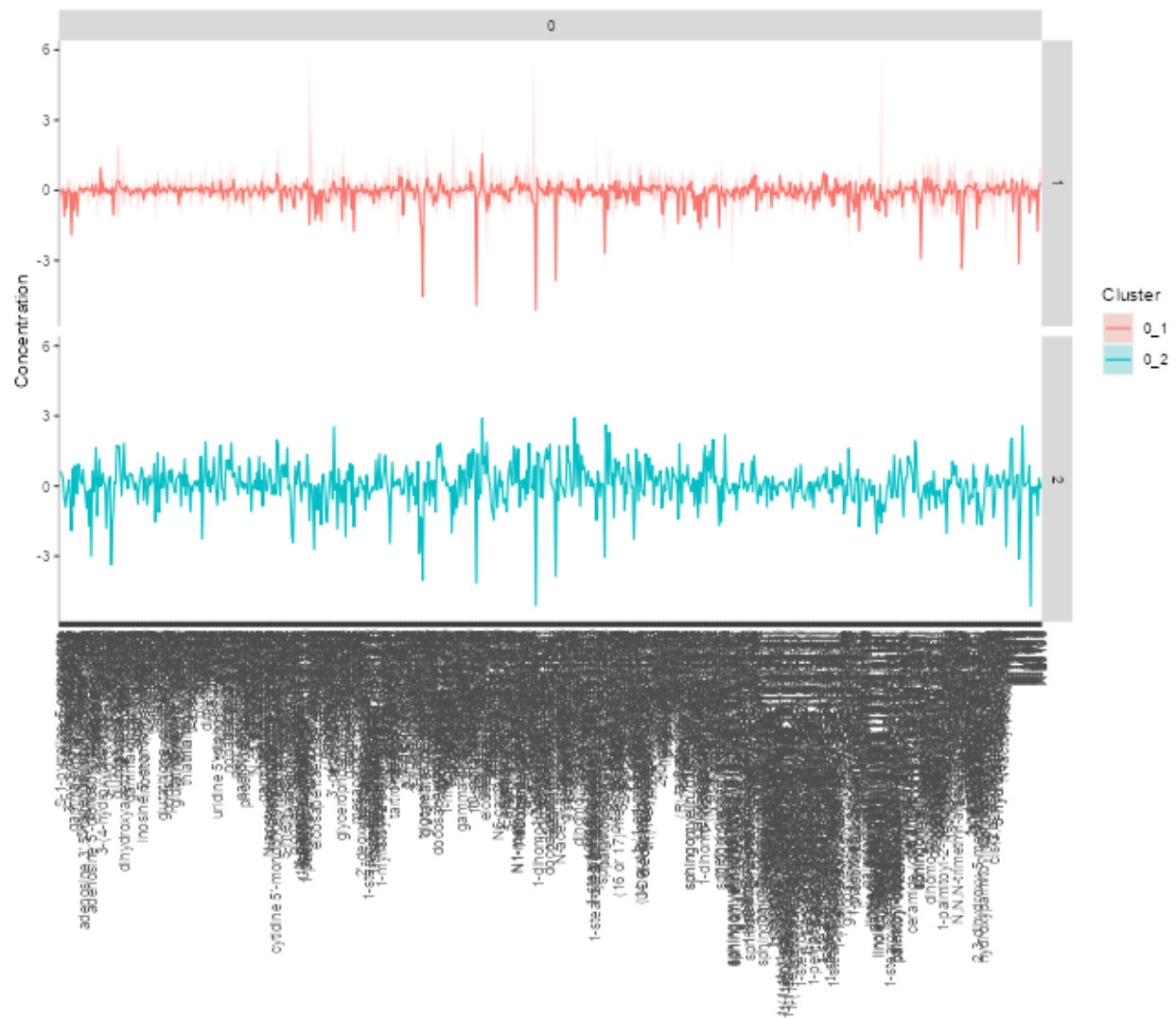# Hierarchical Clustering: Dendogram

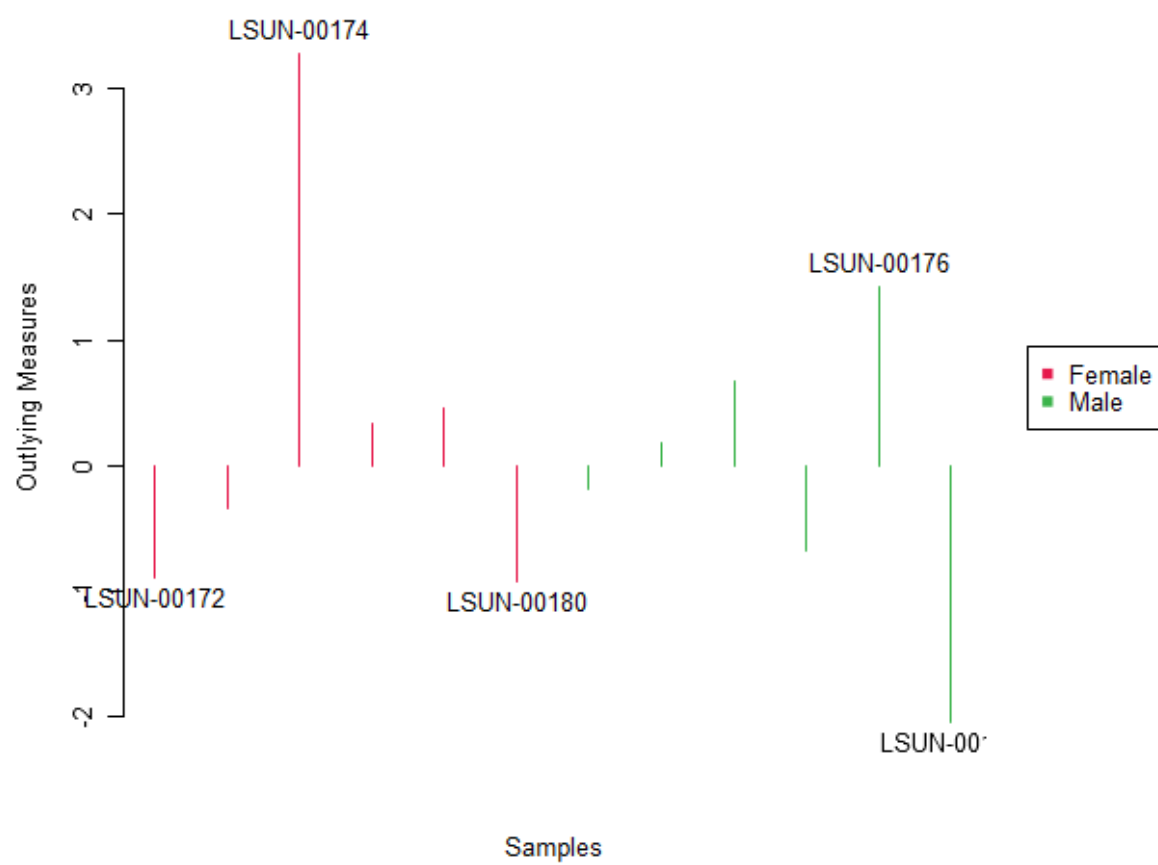# Hierarchical Clustering: Heatmaps
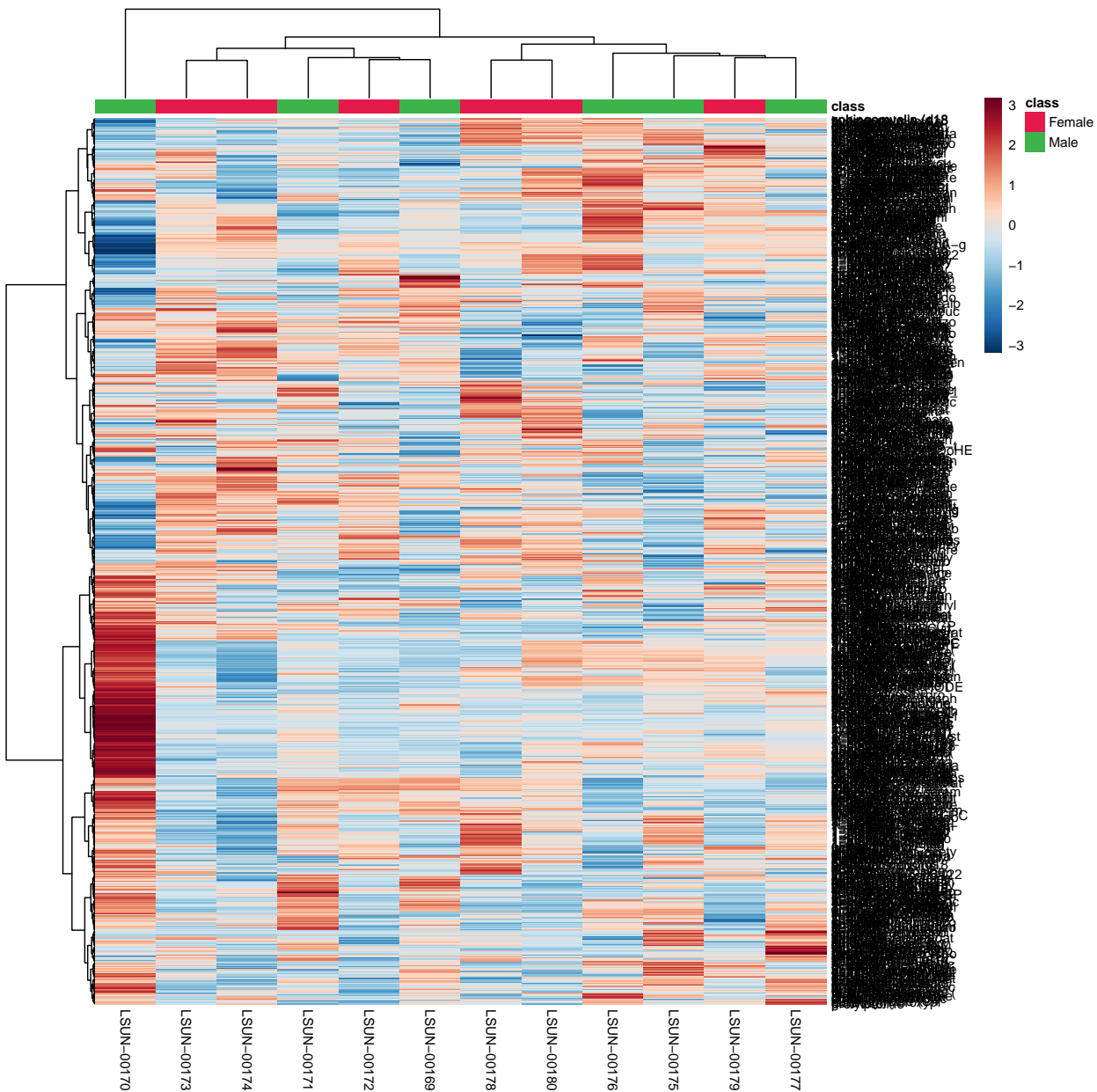
# Partitional Clustering: K-Means
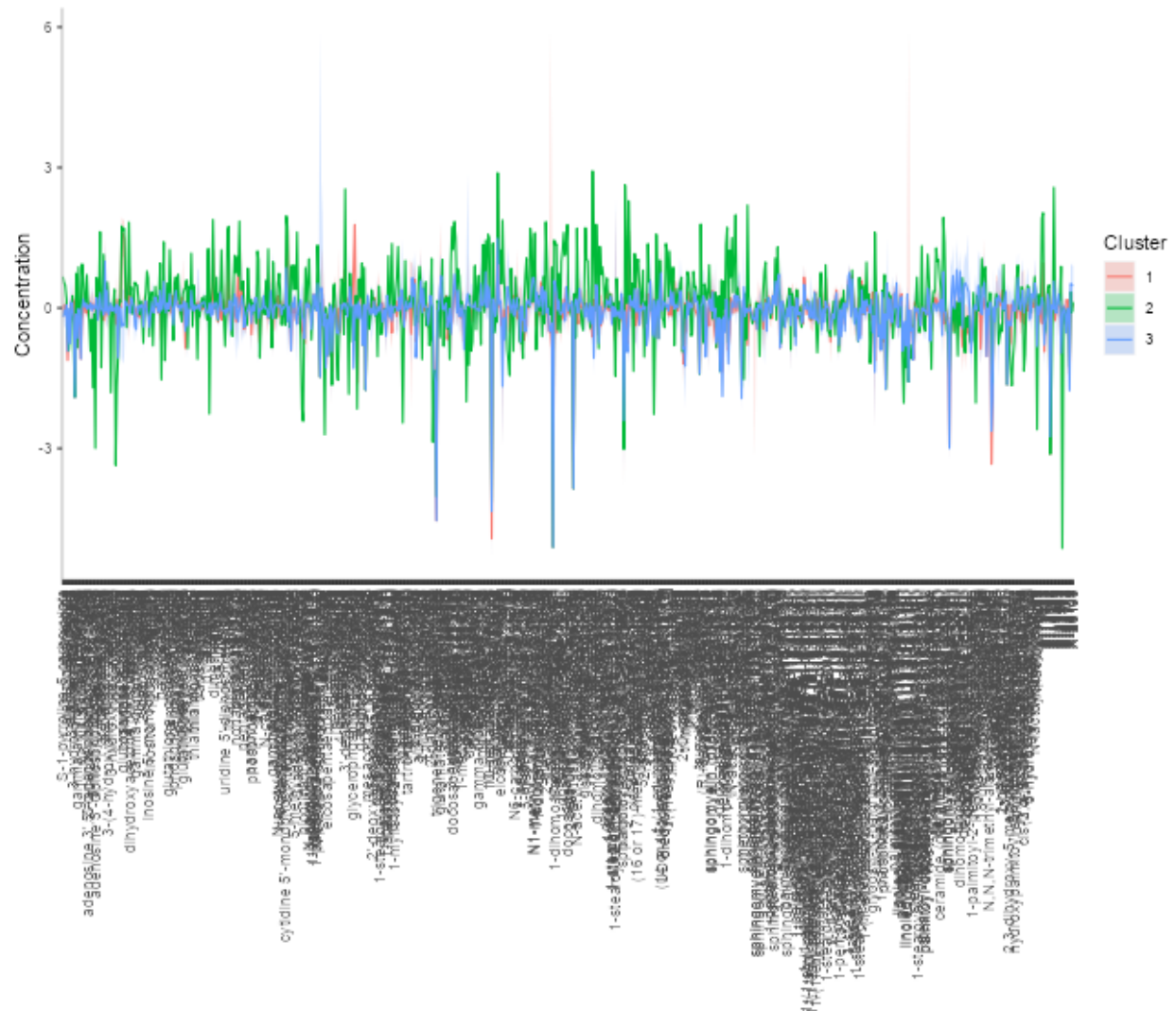
# Partitional Clustering: Self Organizing Maps (SOM)

# Random Forest

**Random Forest classification**
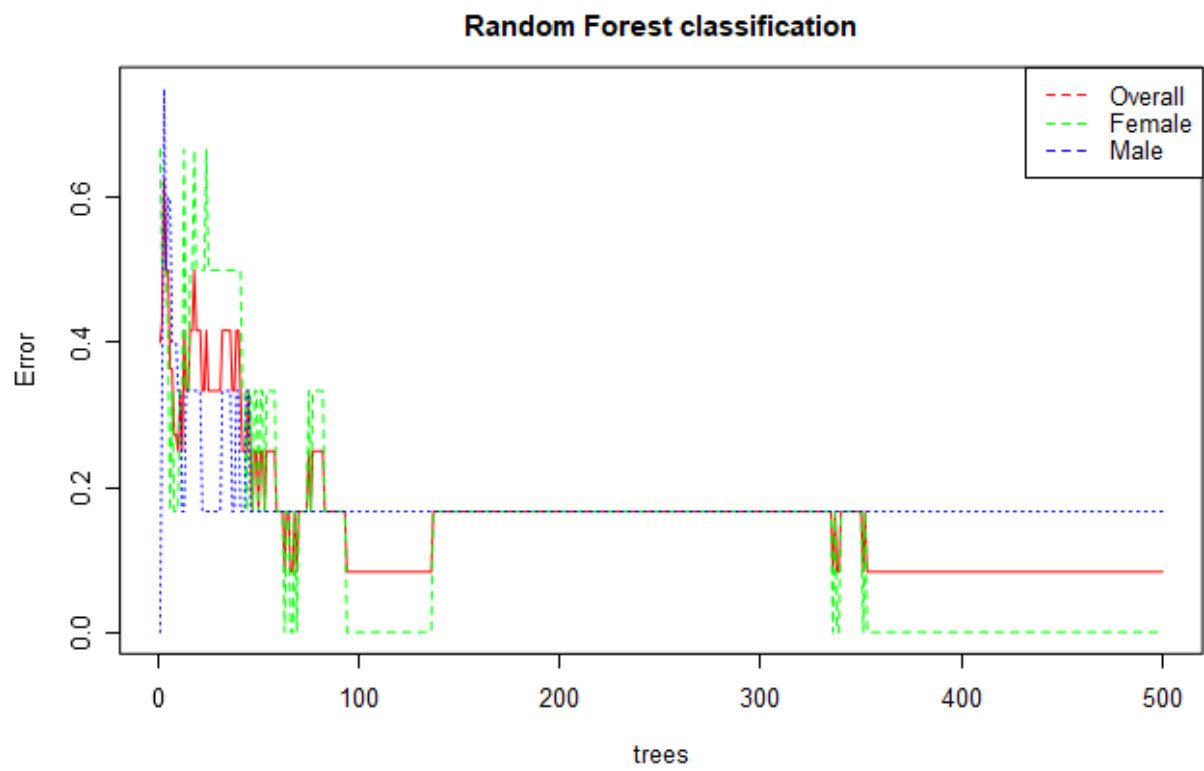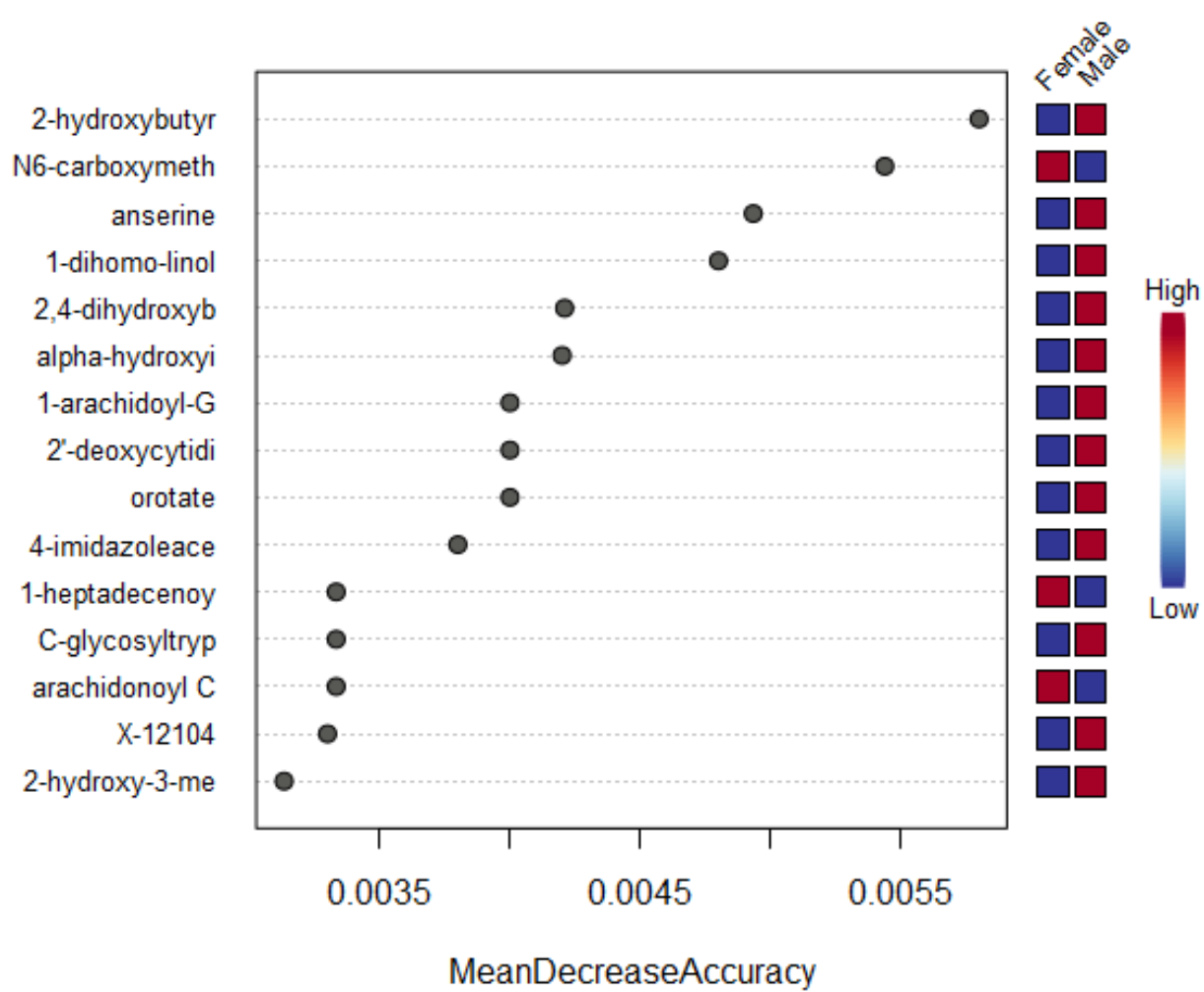
# Support Vector Machine (SVM)



Recursive SVM classification