Multiomics data integration with DIABLO

Andrej Blejec

andrej.blejec@nib.si

National Institute of Biology

January 11, 2023

Situation

Four omics datasets

- metabolomics
- hormonomics
- proteomics
- qPCR

Several treatments: C, H, D, W and combination HD
Several time points: 1, 7, 8, 14, ...
Dimensionality issue: lots of variables, few samples

Goal

Integration across different omics datasets

1. Search for connections between datasets
2. Compare and find differences between treatments and control

Approach

regularized CCA

Finds maximally correlated canonical variates/components (combinations of original variables) for two datasets.

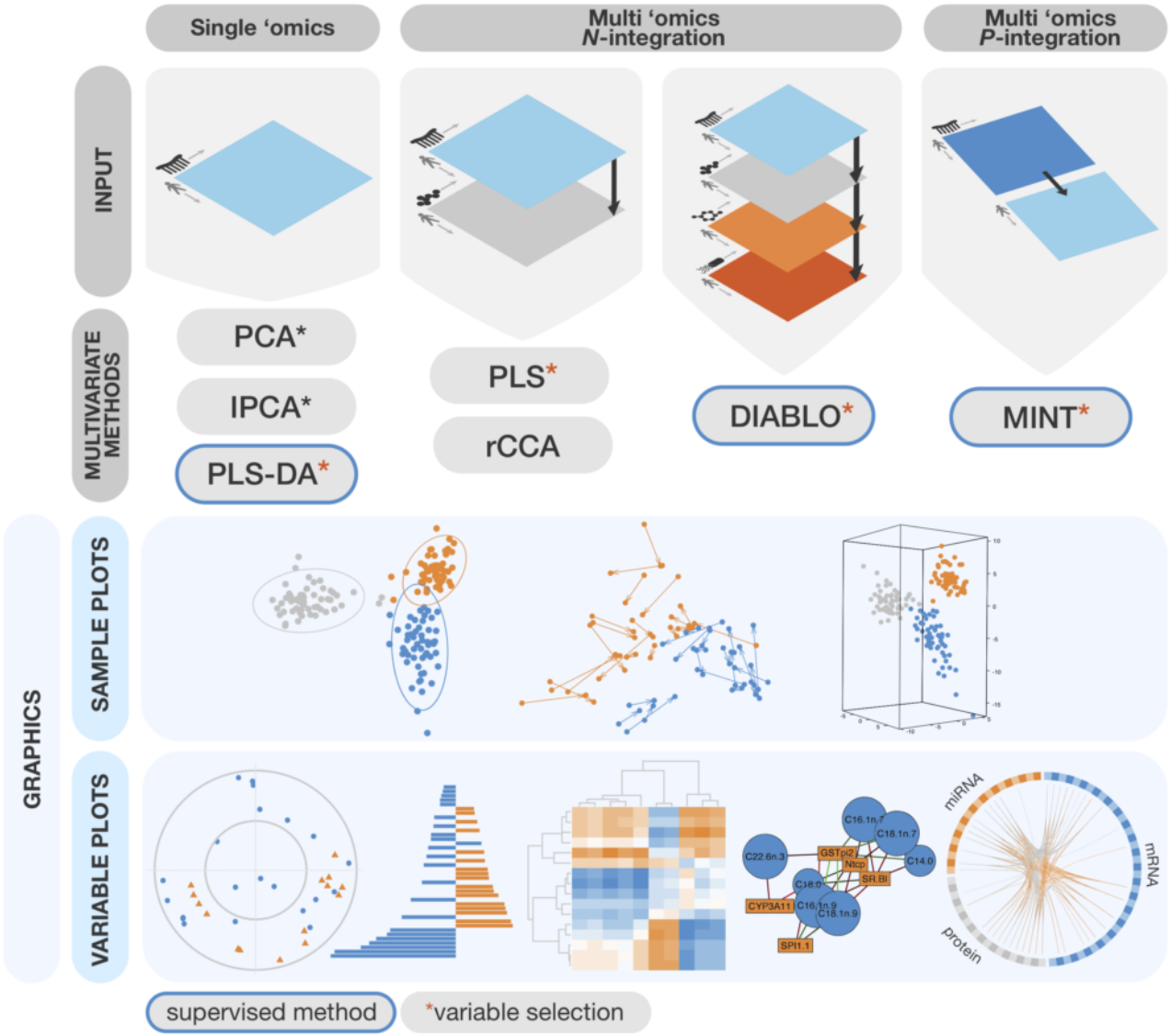
DIABLO
Data Integration Analysis for Biomarker discovery using Latent variable approaches for Omics studies

Can combine more than two datasets using PLS (Partial Least Squares, which is similar to CCA).

- seeks for linear combinations of the variables from each dataset in order to reduce the overall dimensionality
- can handle correlated variables
- it is efficient in 'low number of samples-high number of variates' situation

Both approaches are available in R package **mixOmics**

mixOmics

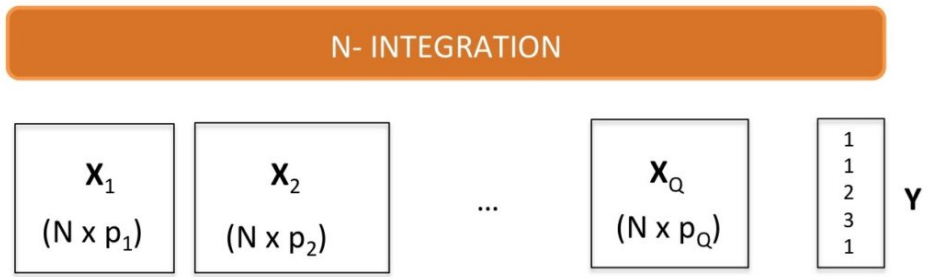

Single 'omics
Multi 'omics N-integration
Multi 'omics P-integration
INPUT
MULTIVARIATE METHODS
PCA*
IPCA*
PLS-DA*
PLS*
rCCA
DIABLO*
MINT*
GRAPHICS
SAMPLE PLOTS
VARIABLE PLOTS
miRNA
mRNA
protein
supervised method
*variable selection

DIABLO

Integrate multiple datasets while explaining their relationship with a categorical outcome variable.

It can also be referred as Multi block Sparse Partial Least Squares Discriminant Analysis , implemented as (`block.splsda()`).

- Four blocks (X_i, our datasets).
- Days of exposure to stress as status Y (`H_D_W_R: 0_0_0_0, 7_0_0_0, ...`).
- Analysis of pairs of stress combinations, e.g. C and H
- Needs the same samples in all datasets (within days and treatment pairs).



Presentation of results

Variable plots

- Plot variables (loadings)
- Correlation Circle Plot

Sample plots

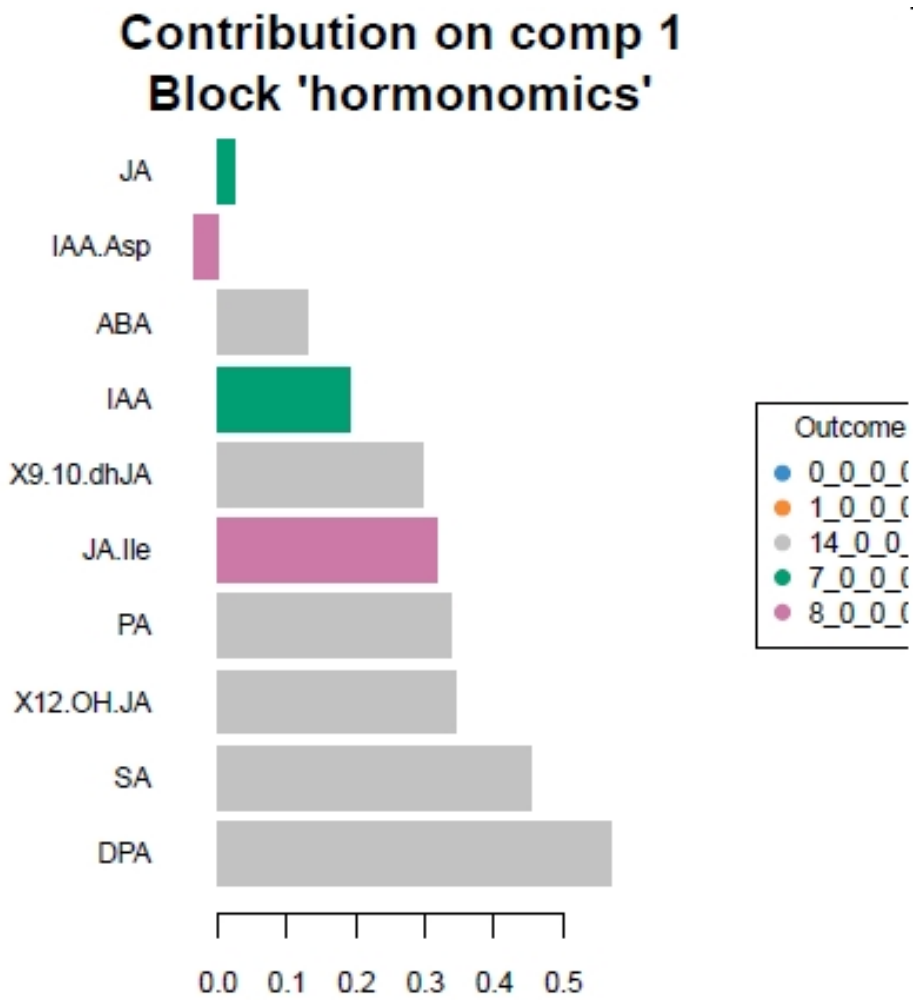
- Correlation of components in blocks
- Scatterplots, overall and by block

Networks

- Circos plot
- Relevance networks
- Bi- and Multipartite networks
- Differential networks

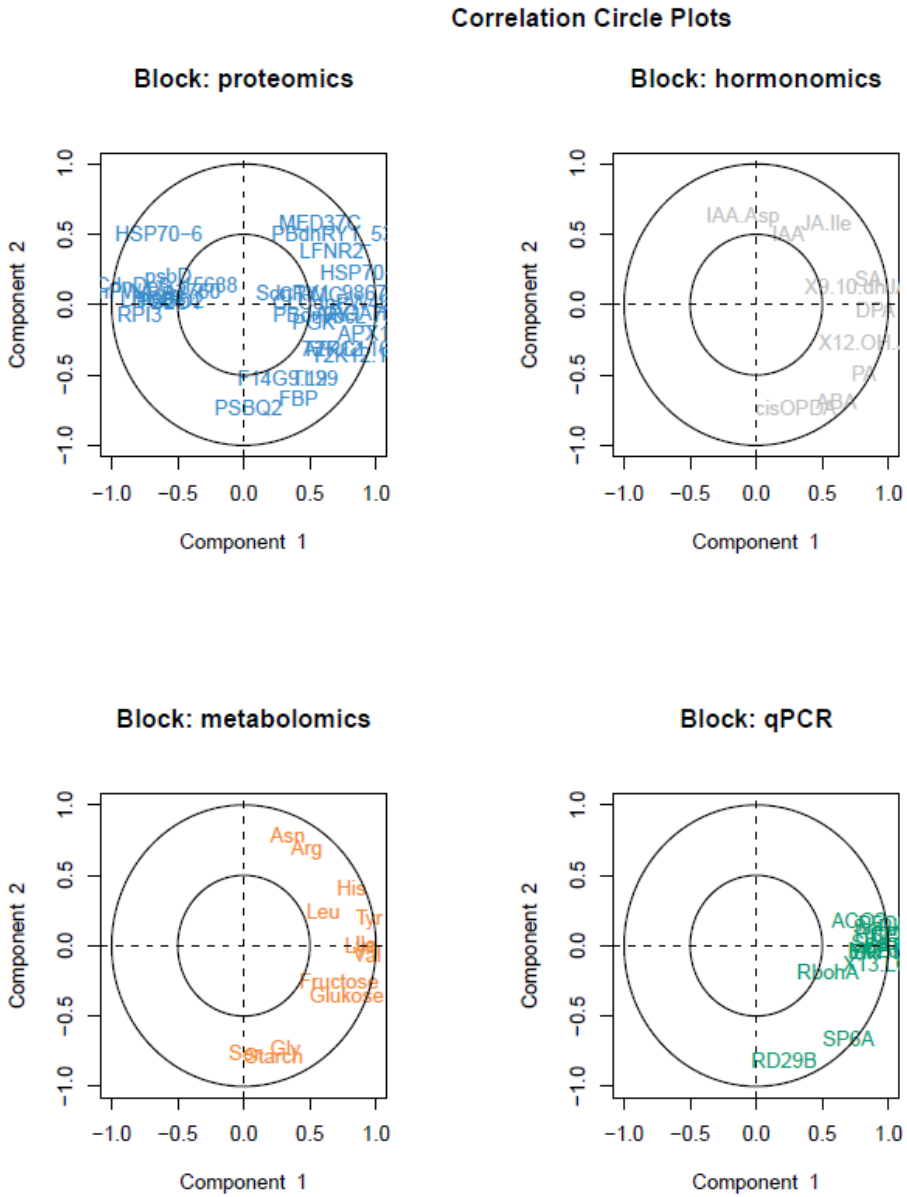
Loadings - contribution to components

```
$hormonomics
                comp1       comp2
IAA        0.19008176  0.28372756
oxIAA      0.00000000 -0.12122695
IAA.Asp   -0.03460993  0.36048002
ABA        0.13076992 -0.49456865
PA         0.33766916 -0.35189668
DPA        0.56727896  0.00000000
SA         0.45197412  0.03313652
JA         0.02374963 -0.02423041
JA.Ile     0.31501155  0.49330159
X9.10.dhJA 0.29529073  0.00000000
X12.OH.JA  0.34419600 -0.11864106
cisOPDA    0.00000000 -0.38382305
```

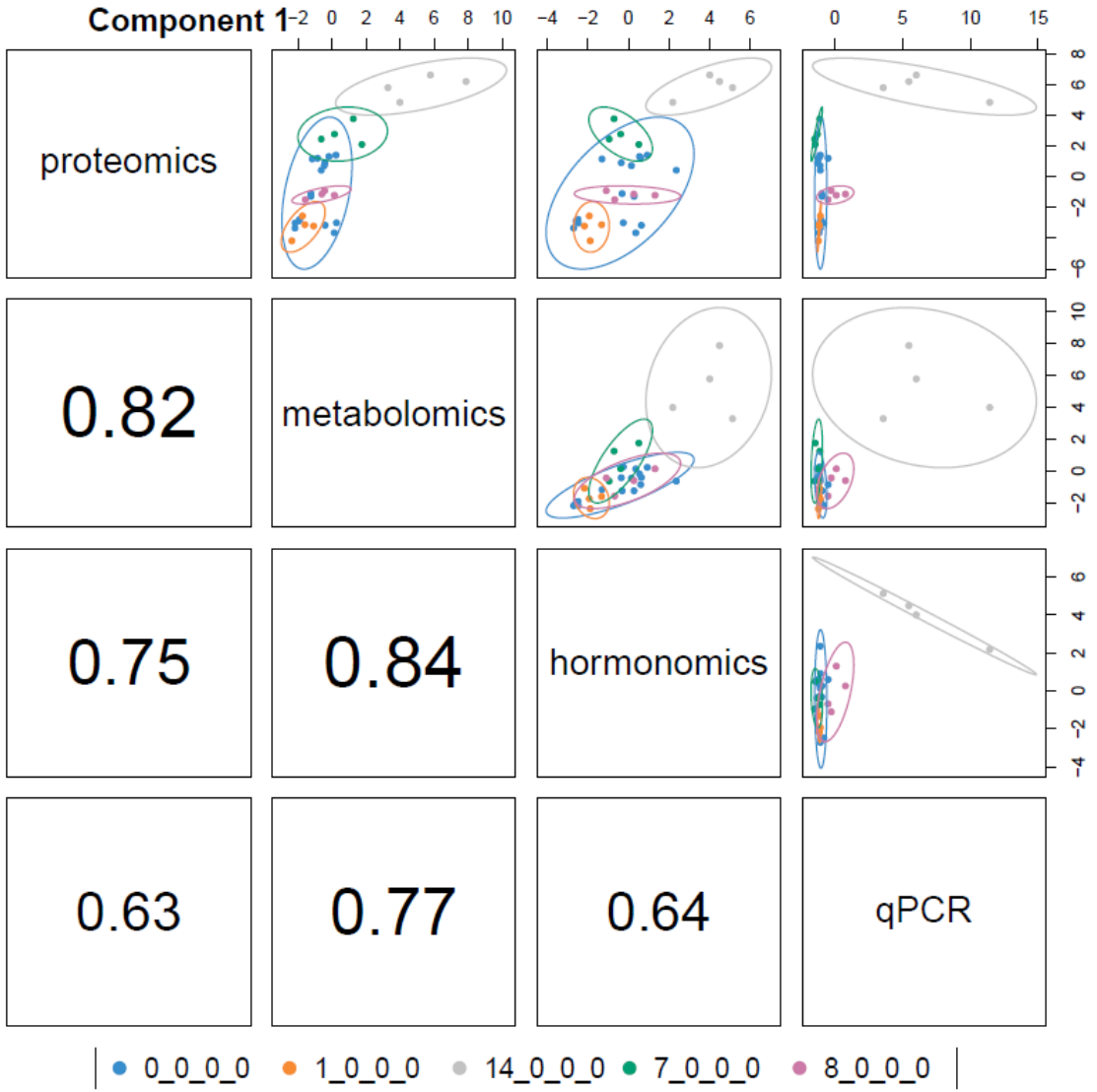



```
plotLoadings()
```

Correlation Circle Plot

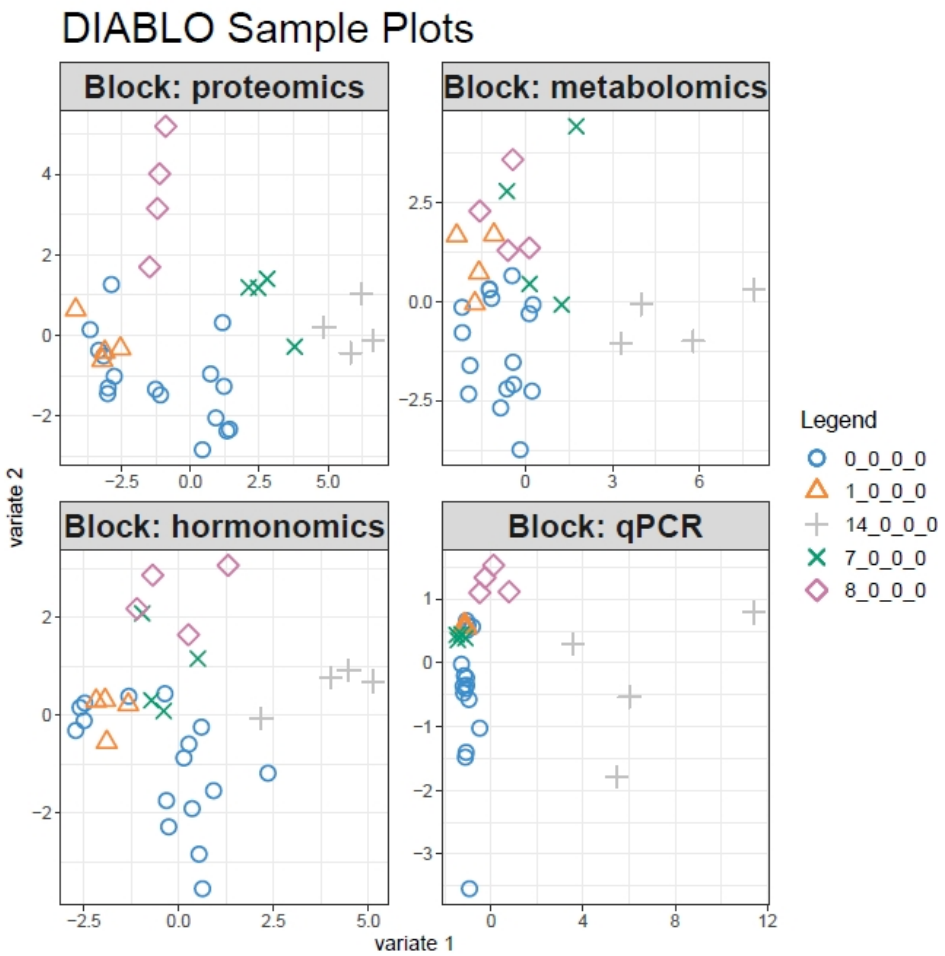


Correlation of components in blocks



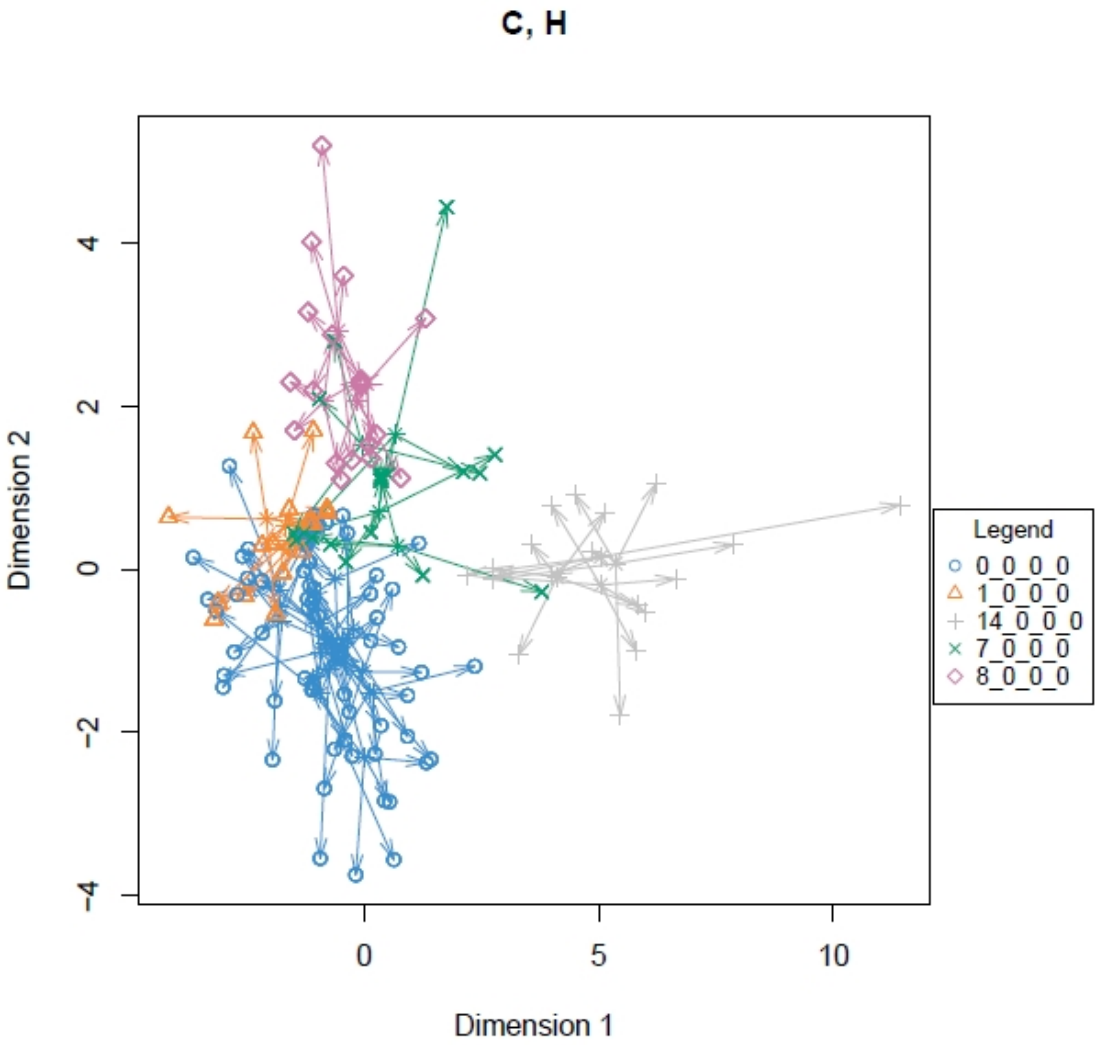
`plotDiablo()`

Scatterplots by block



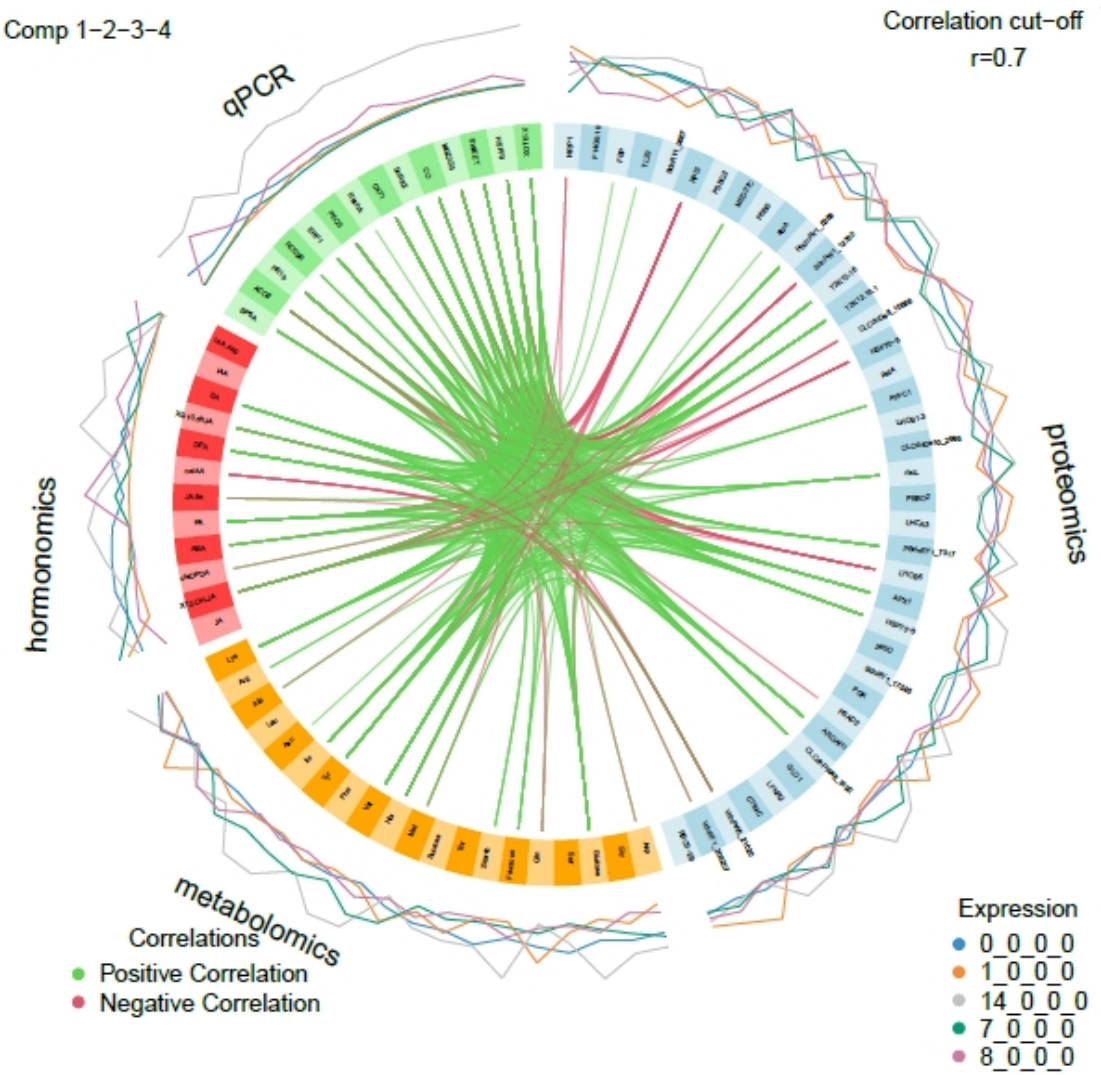
`plotIndiv()`

Overall scatterplot



`plotArrow()`

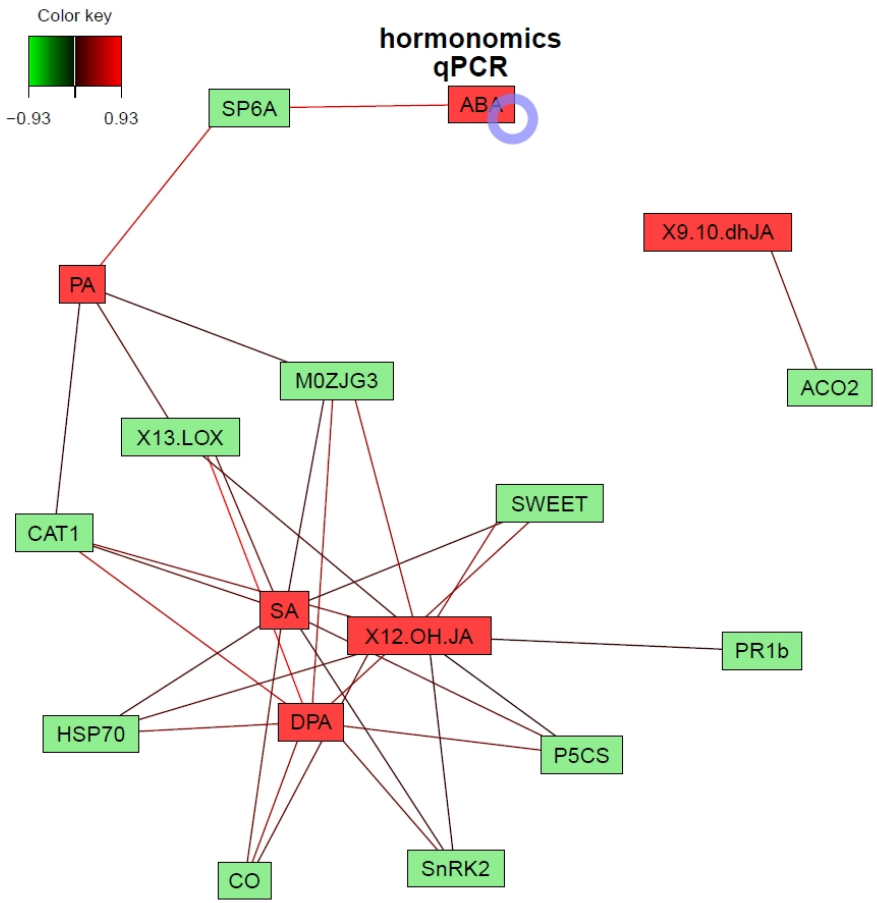
Circos plot



`circosPlot()`

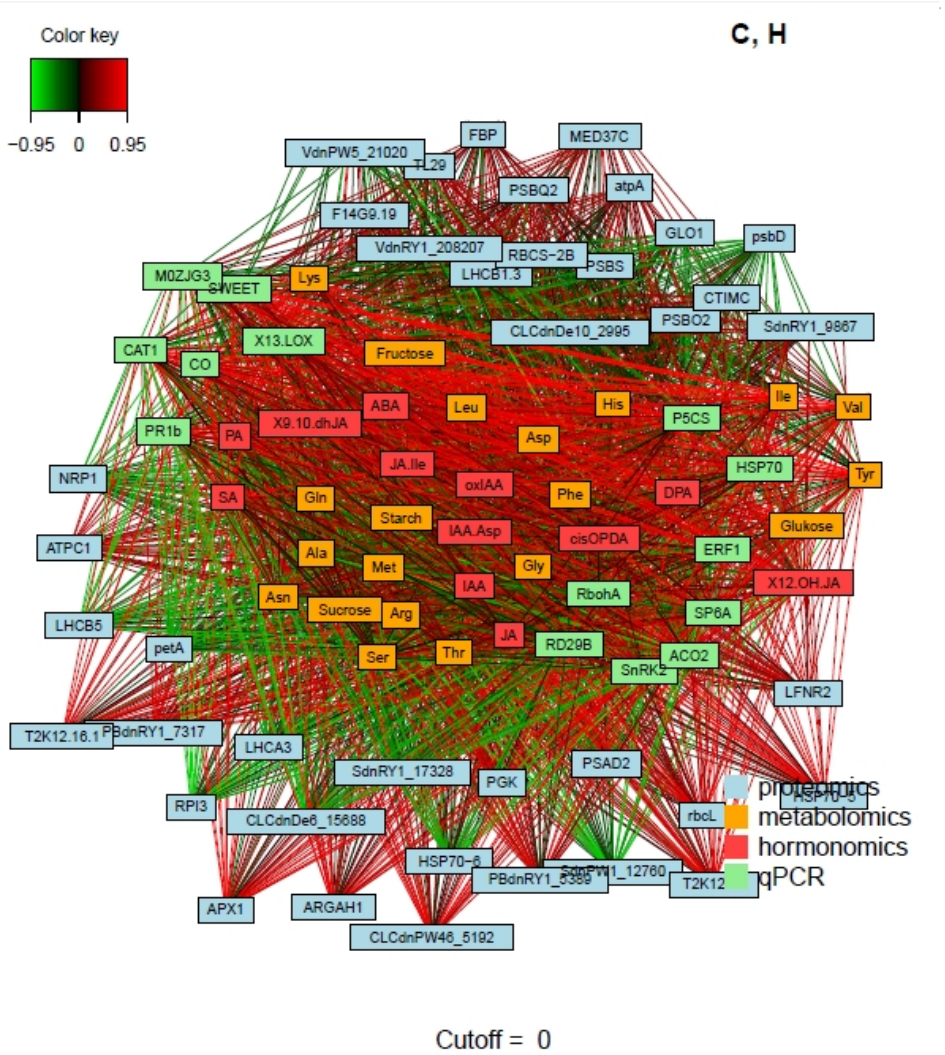
Bipartite network

cutoff = 0.8



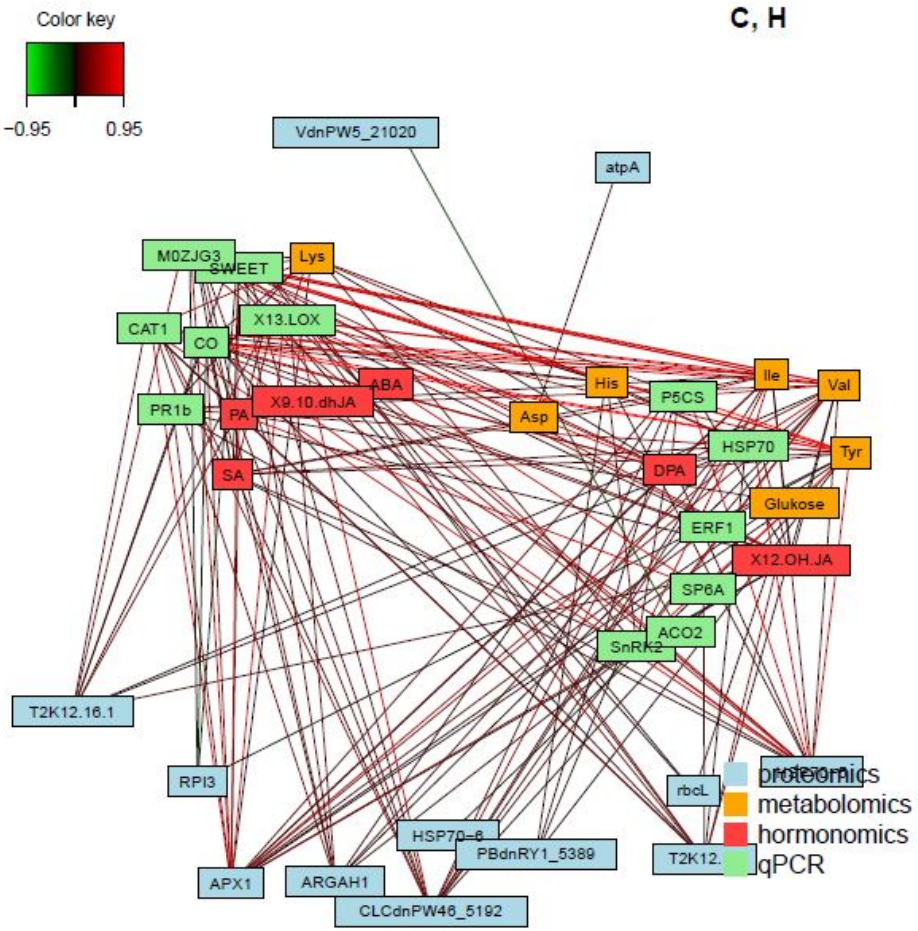
`network()`

Multipartite network



Cutoff = 0

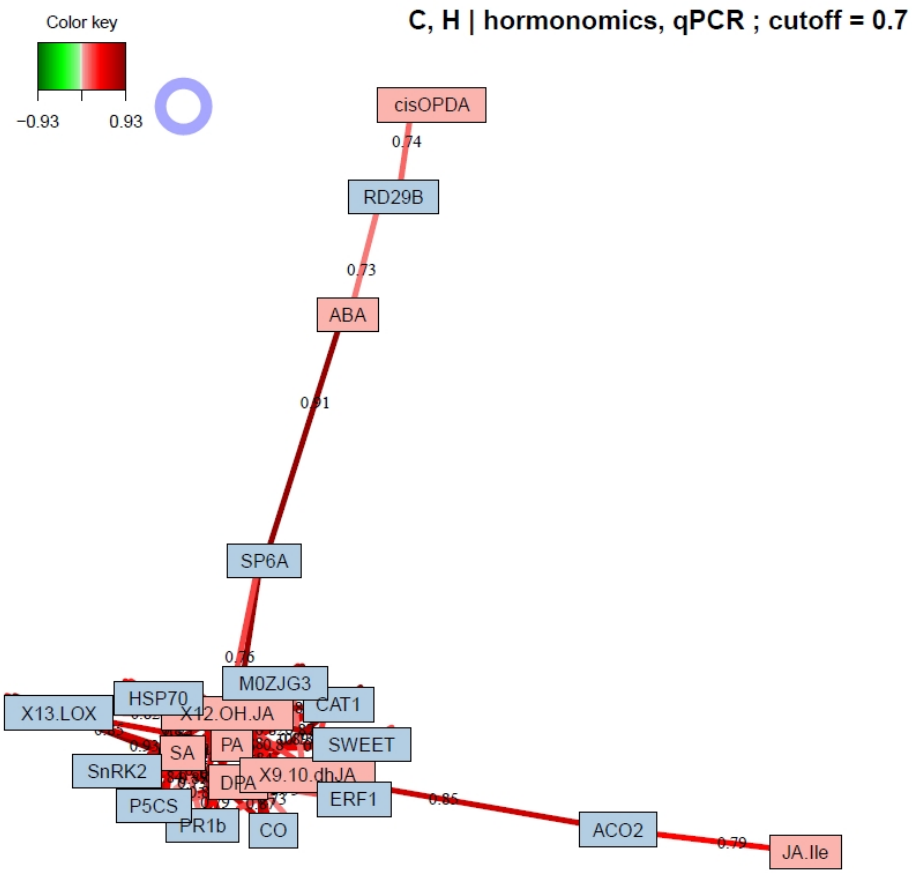
Multipartite network



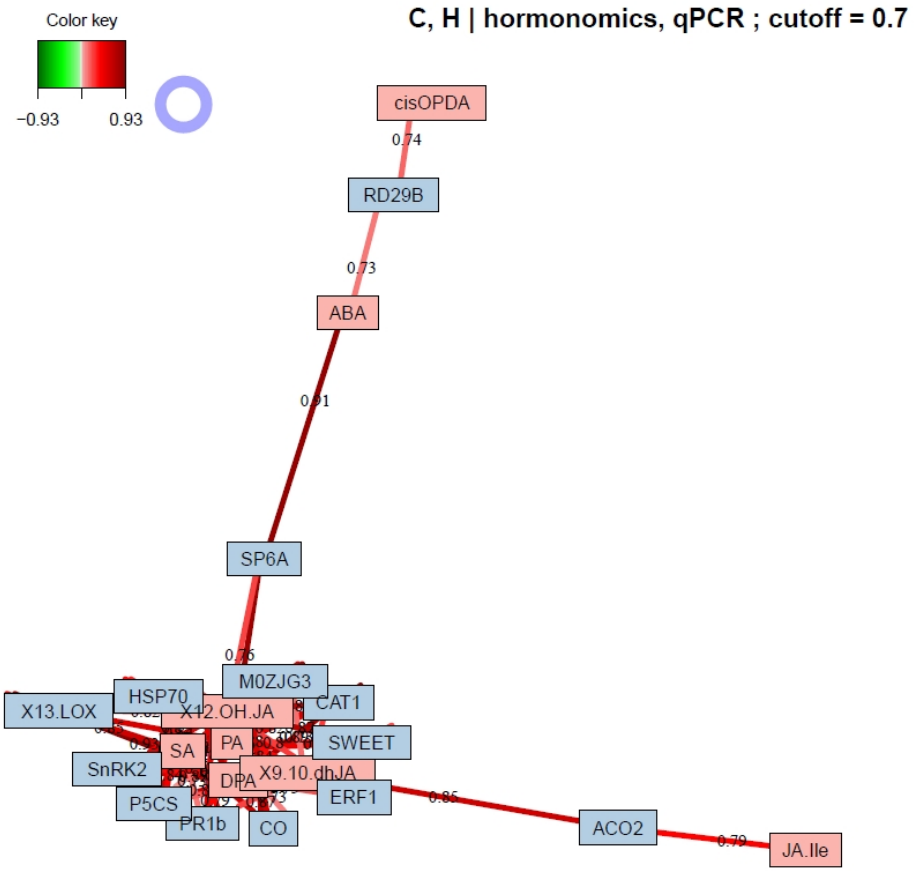
Cutoff = 0.8

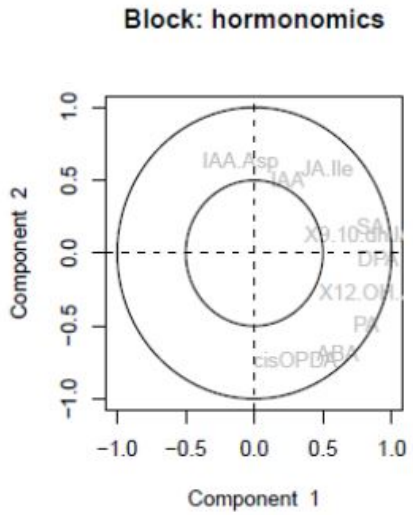
Differential networks (Bipartite)

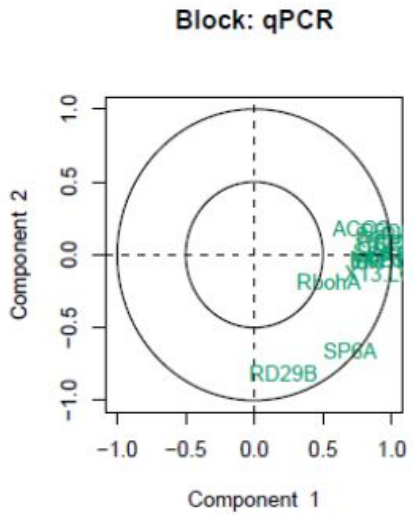
To compare treatment and control, we try to discover edges that are present only in treatment (or control).



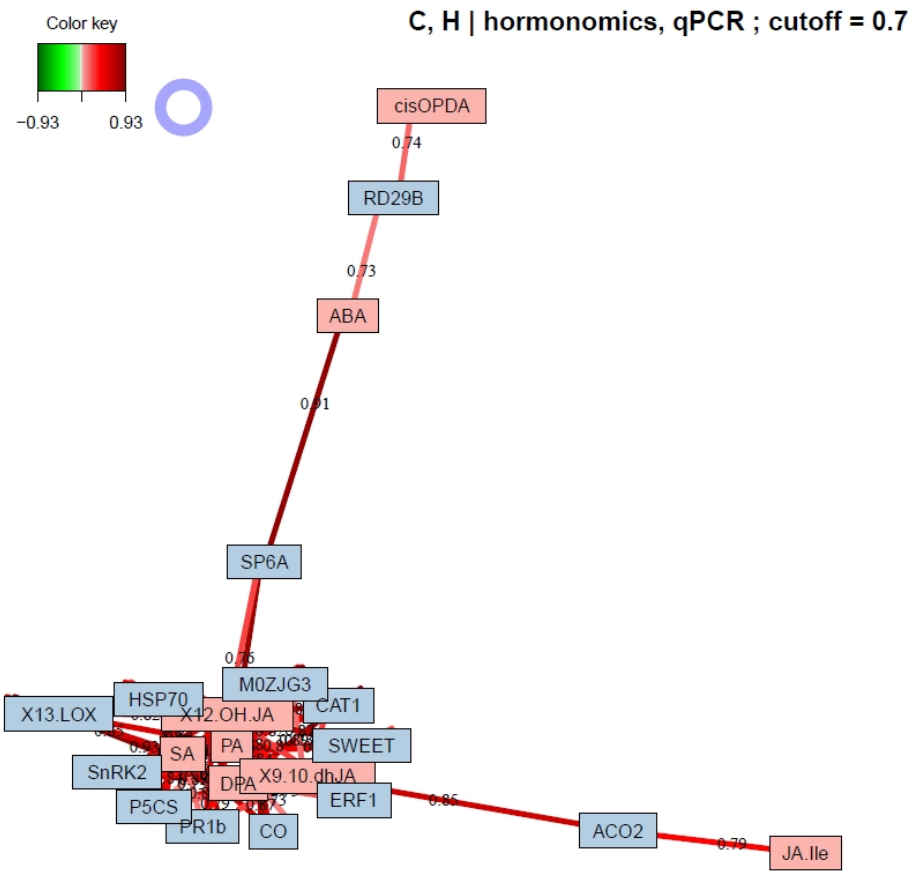
Heatmap of used correlation matrix

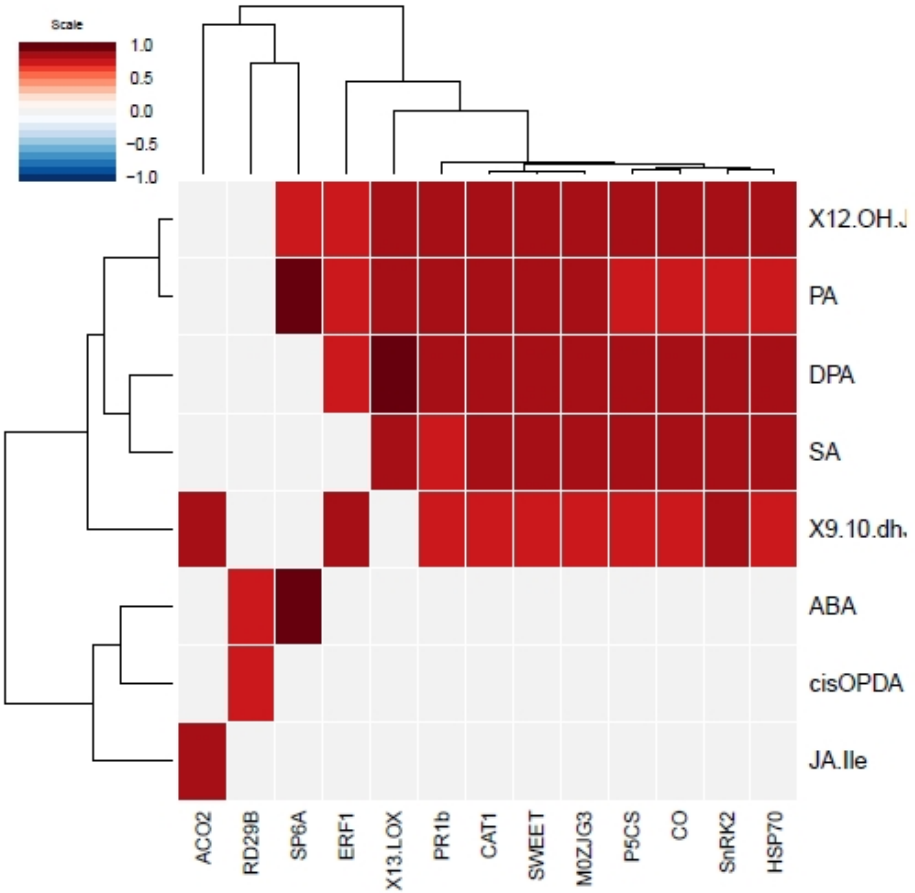






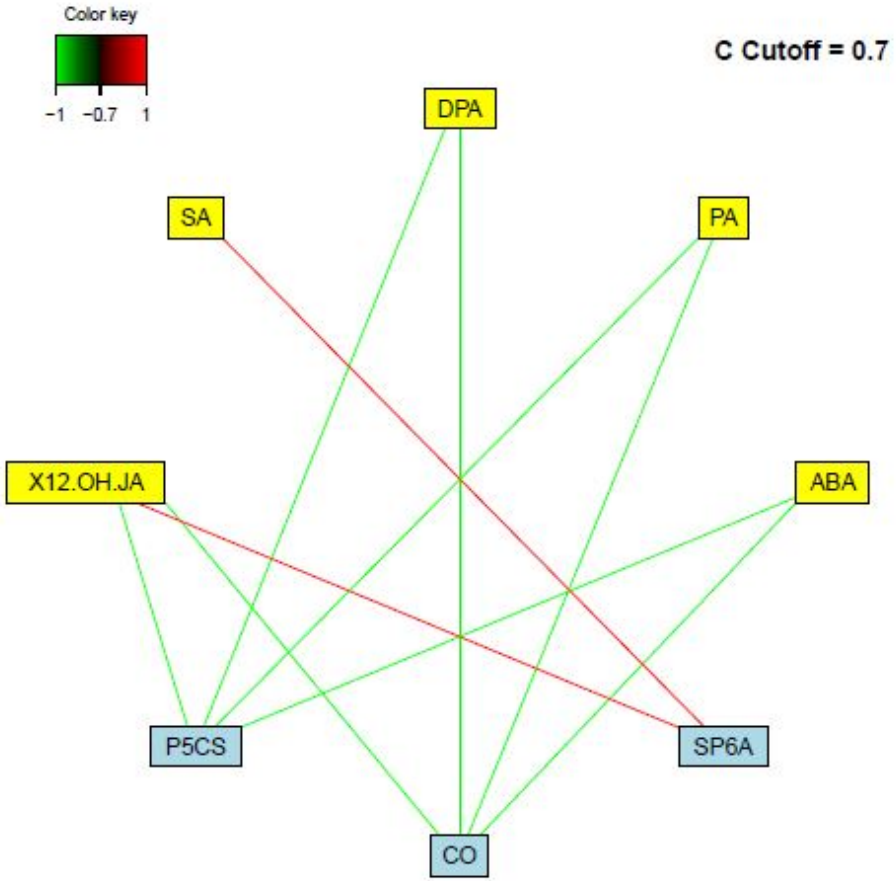
Heatmap of used correlation matrix



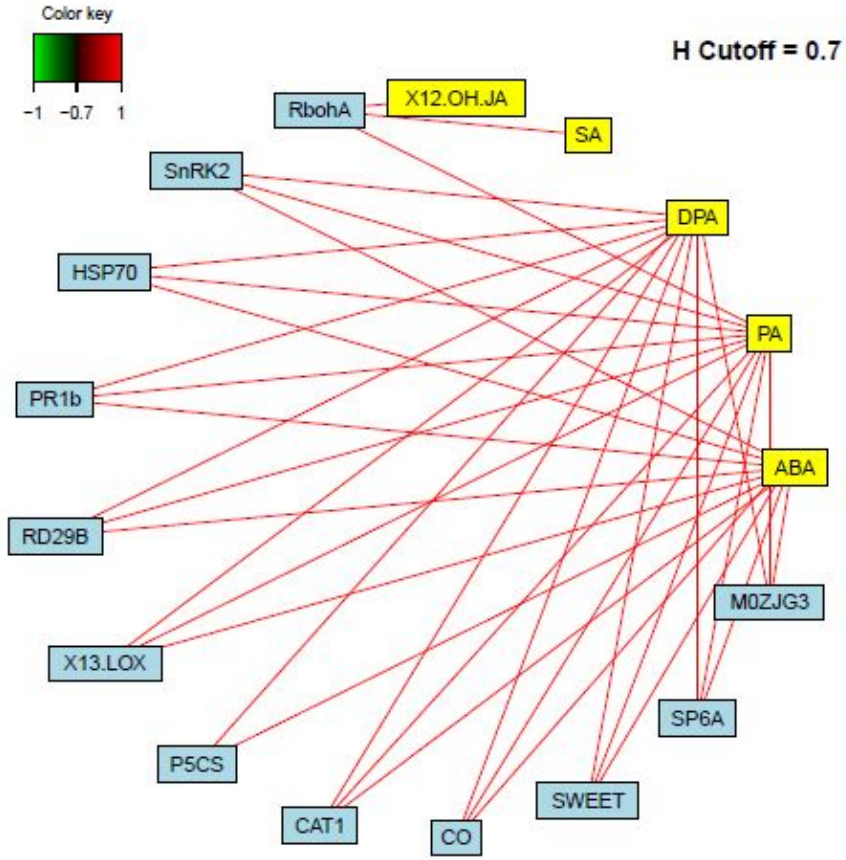


Networks for single condition (treatment)

Network for Control

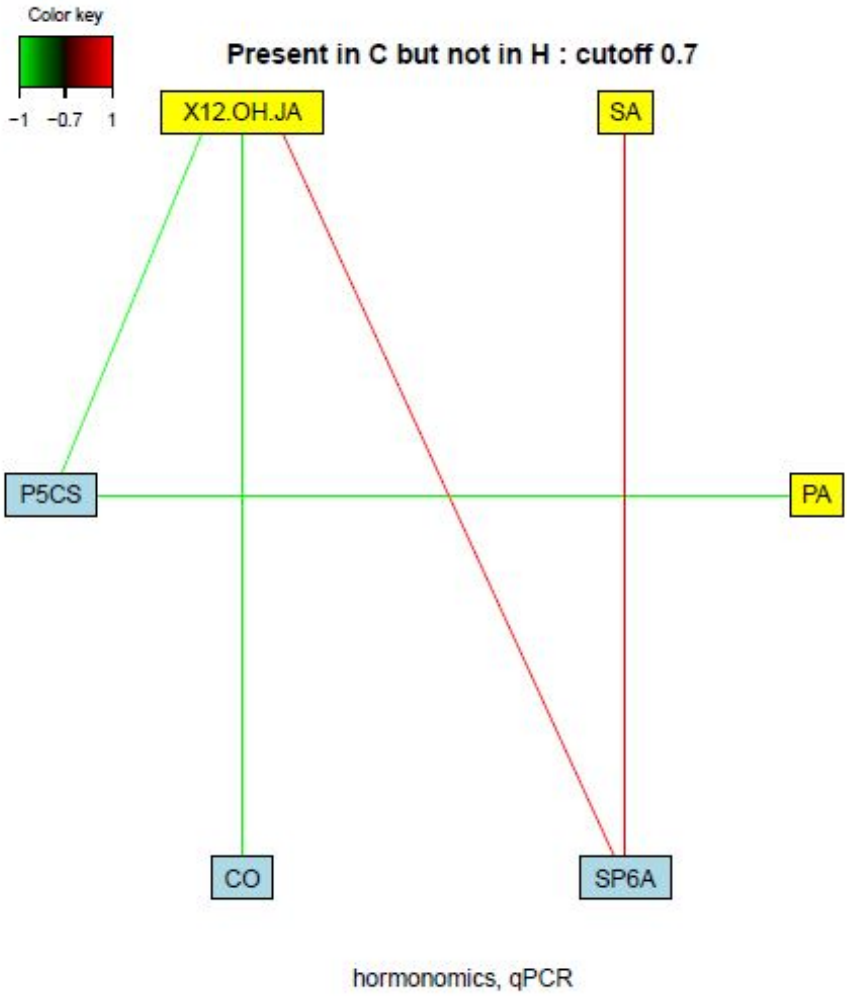


Network for Heath

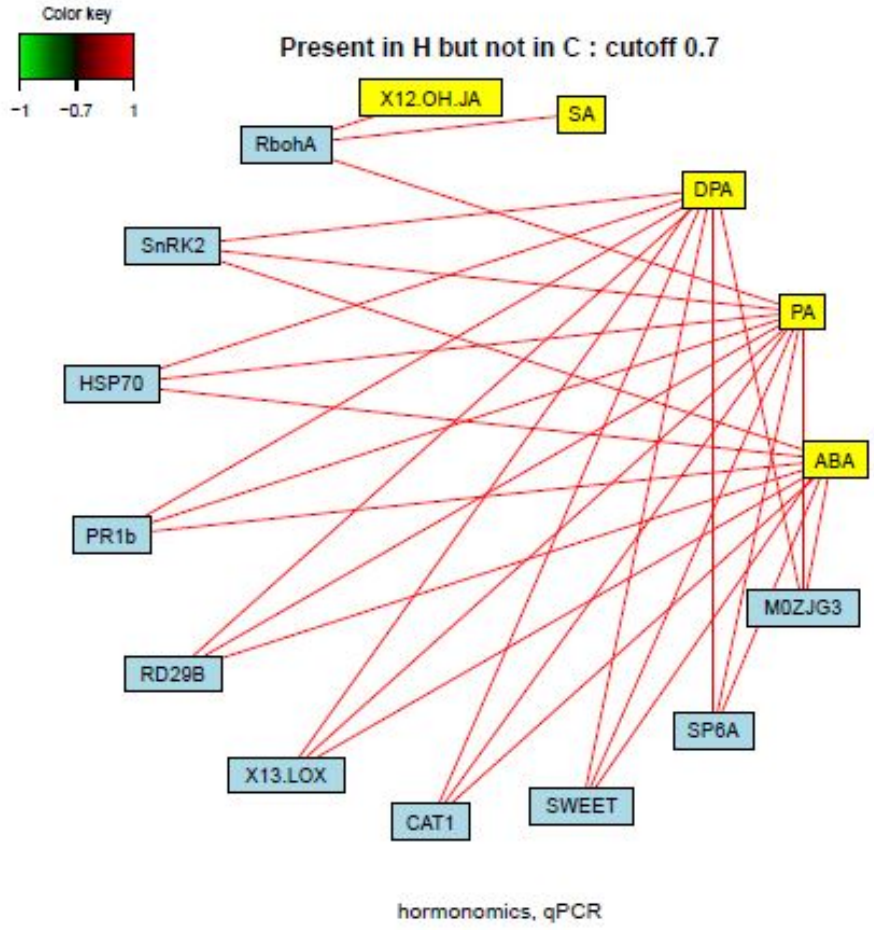


Unique edges

Control, absent in Treatment



Treatment, absent in Control



Parallel network plots