

Report

1 User settings

1.1 Main method

PCMA1

1.2 Input files

	File direction	Variable type
X(Independent Variable)	/home/huai/Project/PCMA/Data/test/Bacteria.csv	Discrete variable
M(mediator)	/home/huai/Project/PCMA/Data/test/Metabolite.csv	Discrete variable
Y(Dependent Variable)	/home/huai/Project/PCMA/Data/test/Diagnosis.csv	Binary Variable

1.3 Feature selection

Usage of correlation analysis:

	Method	Parameter
X~Y	nan	nan
M~Y	SCC; p	SCC >= 0.2; p <= 0.05
X~M	SCC; p	SCC >= 0.2; p <= 0.05

1.4 Normalization

There are variables of normalized

Normalization variable	Method
X	NA
M	MinMaxScaler

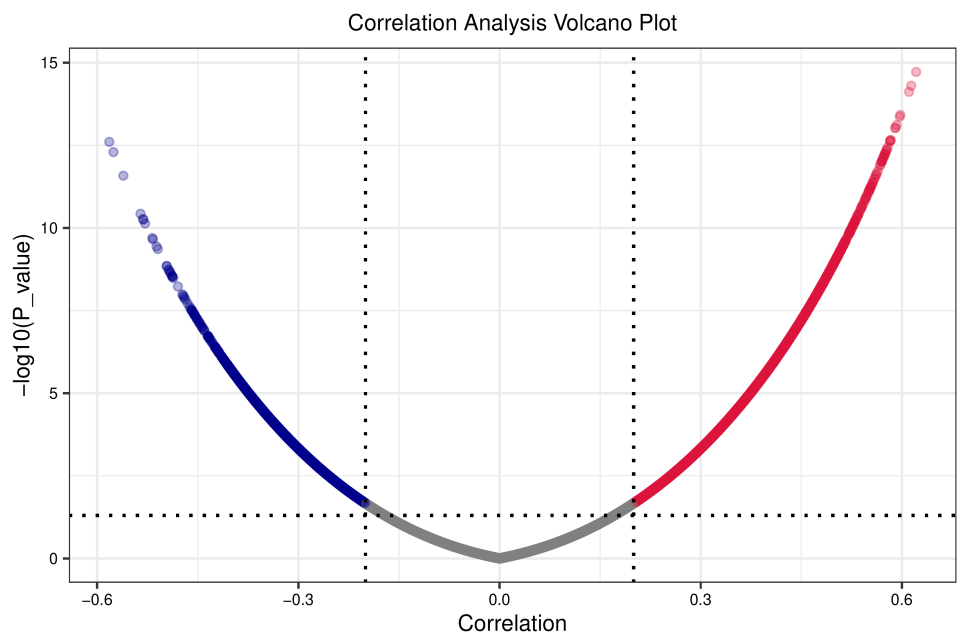
2 Feature selection

2.1 Results of correlation analysis

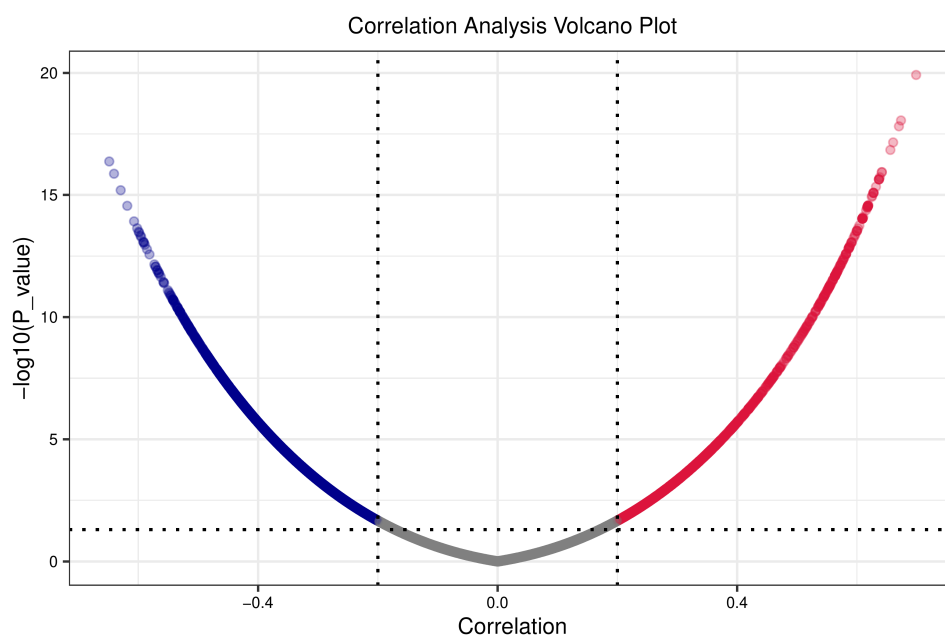
To select the mediator variable M, a total of 8848 features were considered, of which 3464 were retained based on the aforementioned selection criteria.

Volcano plot for paired variable:

Paired variable	Volcano	File direction
X~Y	NA	NA
M~Y	Volcano	/home/huai/Project/PCMA/Data/pcma1/result/data/Metabolite_Diagnosis.csv
X~M	Volcano	/home/huai/Project/PCMA/Data/pcma1/result/data/Bacteria_Metabolite.csv



Bacteria-Metabolite Correlation Analysis Volcano Plot

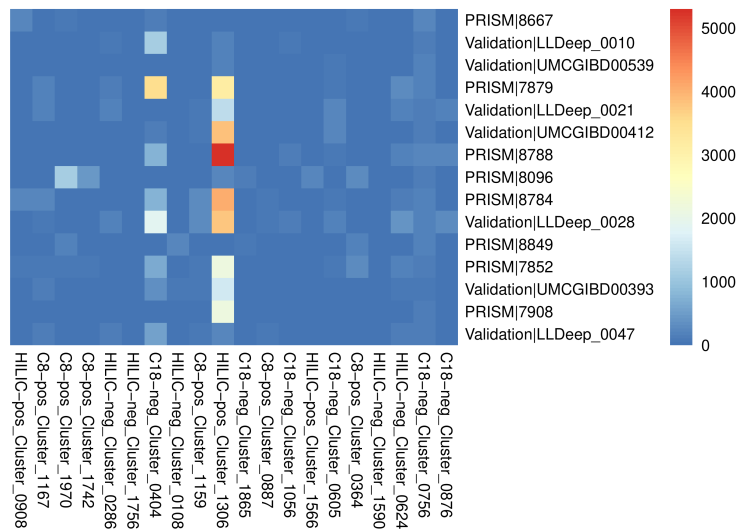


Metabolite-Diagnosis Correlation Analysis Volcano Plot

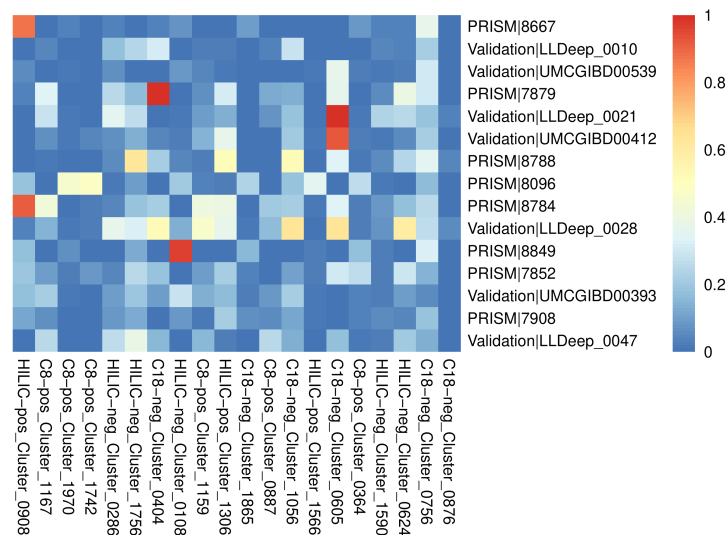
2.2 Results of normalization

The following heatmap visualizes the changes in mediator variable M before and after normalization (selected features and samples randomly).

Variable	Visualization	File direction
X	NA	NA
M	Heatmap	/home/huai/Project/PCMA/Data/pcma1/result/ data/Metabolite_scale_data.csv



Heatmap (Before Normalization)



Heatmap (After Normalization)

3 PCA

Main method: PCMA1

3.1 X

Description: The heatmap reflects the principal components (PCs) derived from the independent variable X after PCA processing. Each PC is linearly composed of the original variables with certain coefficients.

heatmap:

Not available in this method

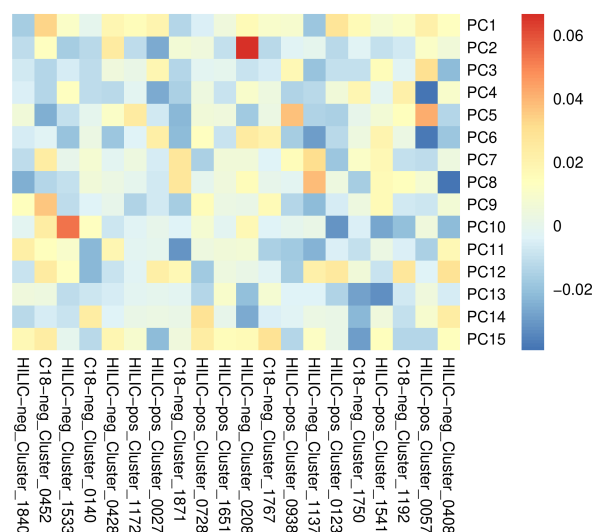
PCA Bacteria Heatmap (PC1 to PC15)

Data	Description	File
Result (Values of each sample's PC)	It shows the values of each principal component (PC) in each sample after the independent variable X has been processed by PCA.	nan
Components	It shows the principal components (PCs) derived from the independent variable X after PCA processing, with each PC linearly composed of the original variables with certain coefficients.	nan
Explained Variance Ratio	-	nan

3.2 M

Description: The heatmap reflects the principal components (PCs) derived from the mediator variable M after PCA processing. Each PC is linearly composed of the original variables with certain coefficients.

heatmap:



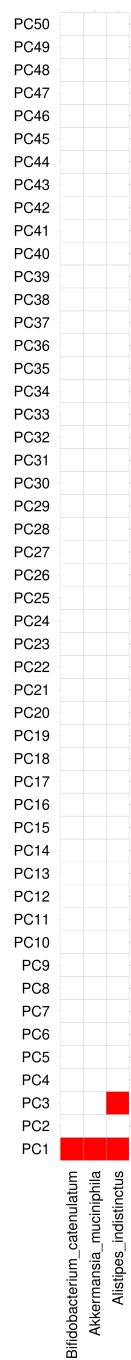
PCA Metabolite Heatmap (PC1 to PC15)

Data	Description	File
Result (Values of each sample's PC)	It shows the values of each principal component (PC) in each sample after the mediator variable M has been processed by PCA	/home/huai/Project/PCMA/Data/pcma1/result/data/Metabolite_pca.csv
Components	It shows the principal components (PCs) derived from the mediator variable M after PCA processing, with each PC linearly composed of the original variables with certain coefficients.	/home/huai/Project/PCMA/Data/pcma1/result/data/Metabolite_pca_components.csv
Explained Variance Ratio	-	/home/huai/Project/PCMA/Data/pcma1/result/data/Metabolite_pca_contribution.csv

4 Mediation analysis

Main method: PCMA1

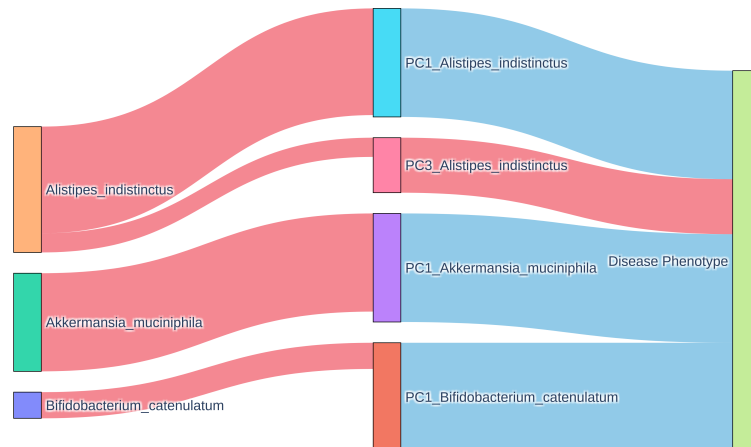
Significant results:



Significant PC Plot. X-axis: bacteria (PC), Y-axis: metabolites (PC)

Sankey Plot:

Sankey Plot



Sankey Plot. From left to right, the three layers are Bacteria (PC), Metabolites (PC), and Disease Phenotype.

File:

Significant PC: /home/huai/Project/PCMA/Data/pcma1/result/data/
Significant_PC.csv

Coefficient mediation: /home/huai/Project/PCMA/Data/pcma1/result/data/
coefficient.csv

5 Further analysis

5.1 Fisher analysis

Optional. The file of metabolite function should be provided.

After performing principal component analysis (PCA) on metabolite data, each principal component (PC) is a linear combination of the original metabolites.

This module analyzes the functional annotations of each metabolite (functional annotations need to be provided) to observe which functions are enriched in each principal component.

N50

For each principal component, select the top 50% of metabolites based on the sum of positive and negative coefficients, and perform functional annotation.

barplot_fisher.pdf

The figure will be attached at the end of the document.

The chart compares the functional annotations of the top 50% of metabolites (by coefficient sum) with those of all metabolites before selection. The comparison is visually represented using a stacked bar chart.

Functional Analysis

In significant pathways involving bacteria (or bacterial PCs), metabolite PCs, and diseases, use Fisher's exact test to identify metabolites significantly enriched in the corresponding PCs.

Attachment 1: significant_pathway.csv

Across all PCs, the results of Fisher's exact test are provided:

Attachment 2: Fisher_results.csv

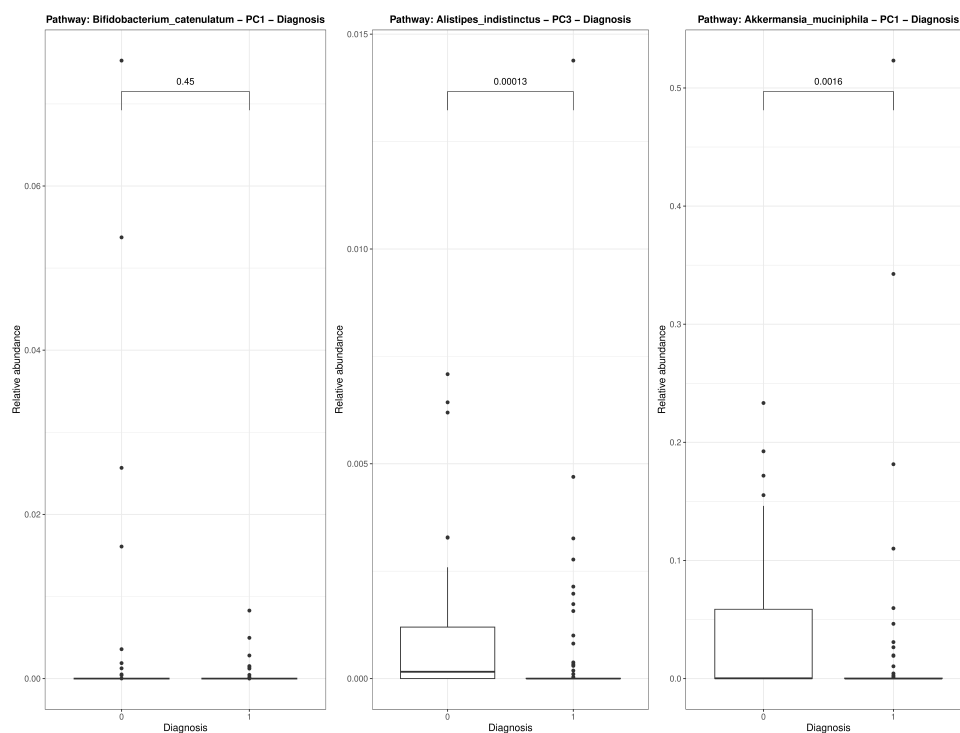
5.2 Boxplot analysis

Boxplot analysis for Bacteria (PC) and Metabolite (PC)

Bacteria PC:

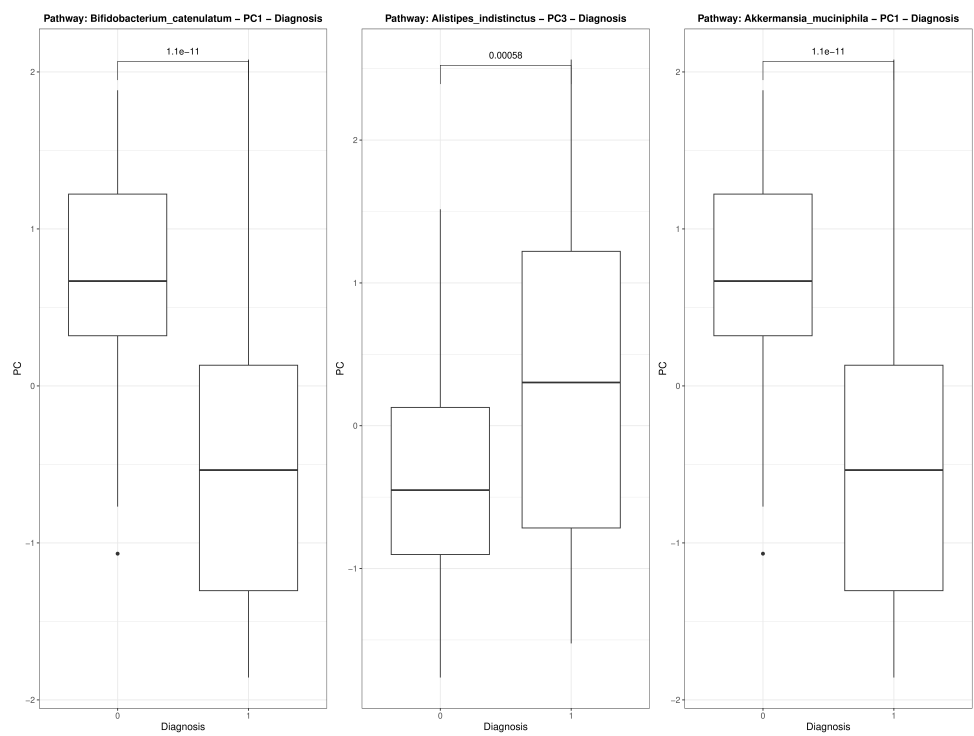
Not available in this method.

Single Bacteria:



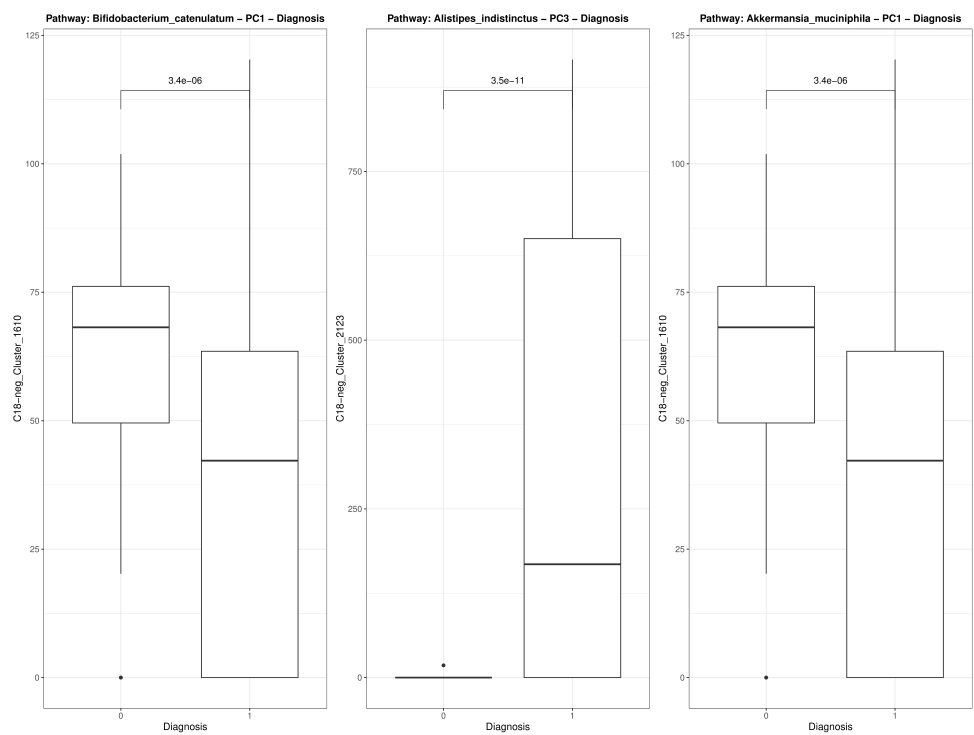
Boxplot for Bacteria Relative abundance.

Metabolite PC:

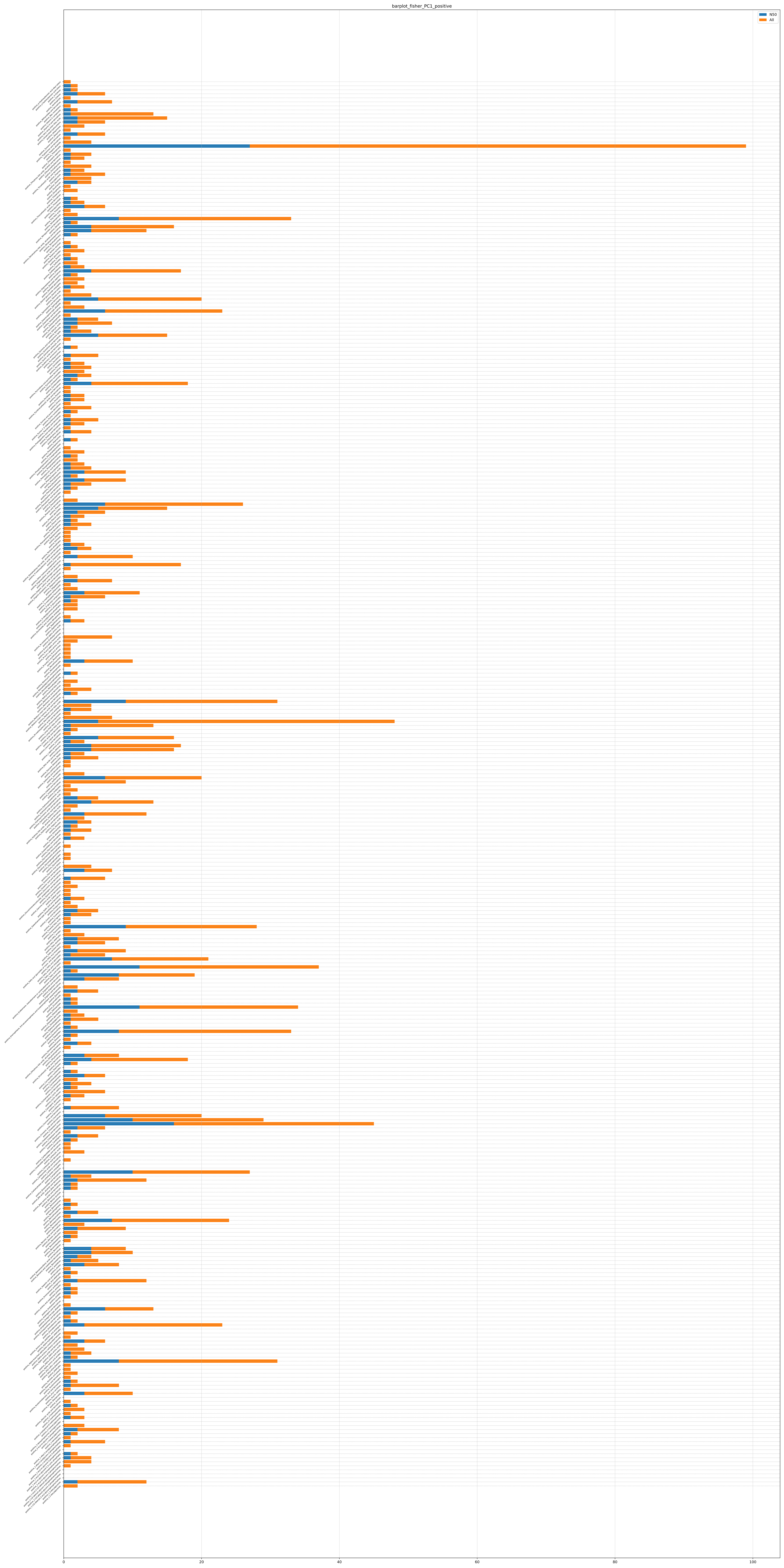


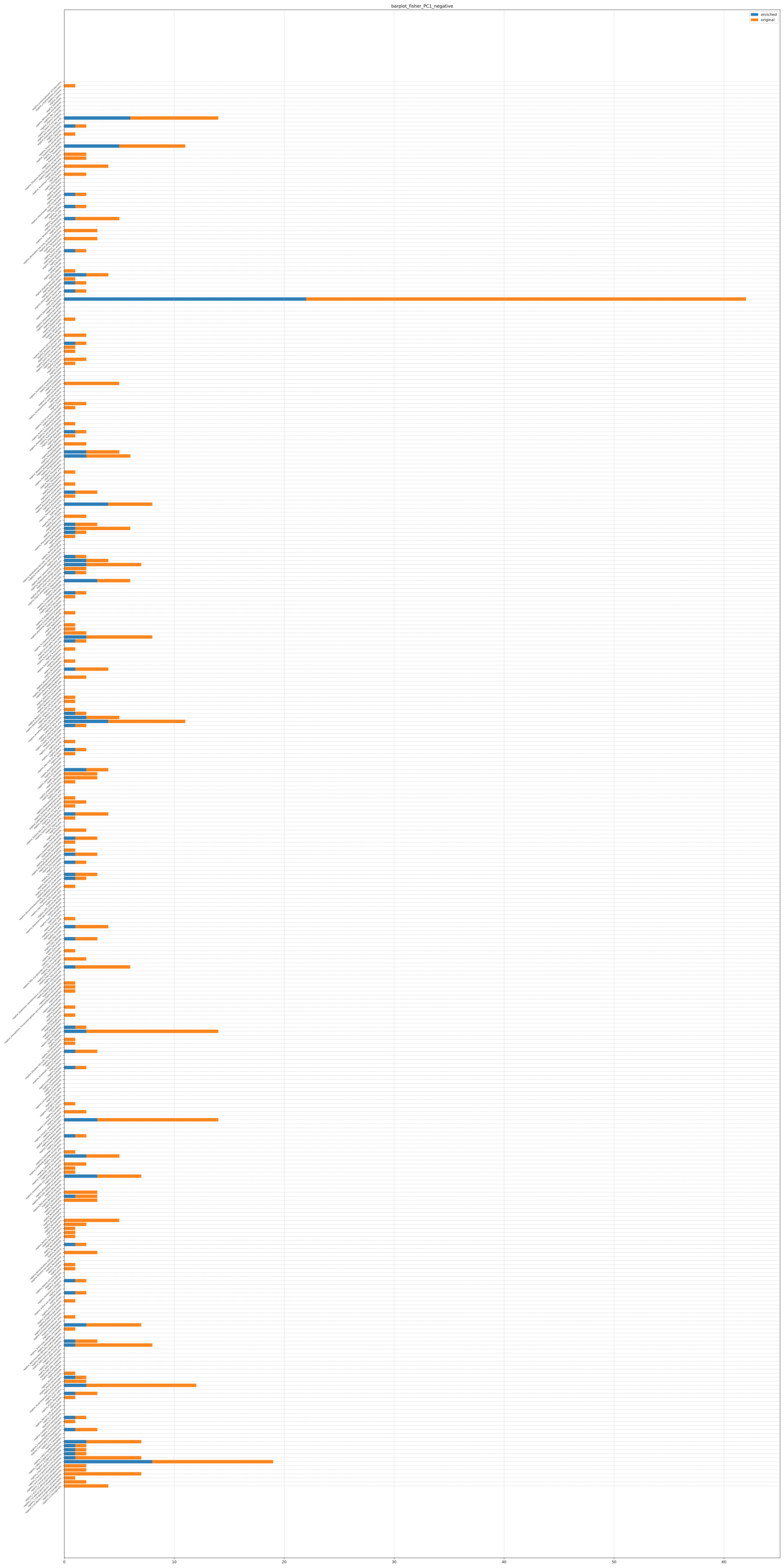
Boxplot for Metabolite Significant PC

Single Metabolite:



Boxplot for metabolite in Significant Metabolite PC.





barplot_fisher_PC1_positive

