WORKING MANUAL OF MMIP

Notes:

1. → [#] *Red arrow and number box indicates additional functionality.*

2. ----▸ *Black dotted arrow indicates feature information.*

3. *All the comparison and the analyses are between two groups (not more).*

4. *By the term "significant", statistical significance is referred to.*

5. *All the analysis diagrams are made using Plotly and will feature the following tools (as shown below) on the top right corner of each diagram, allowing the user to access features like zoom, edit, download and other interesting tools.*

MMIP

HOME   OVERVIEW   SERVICES   MANUAL   RELEASE-NOTES   CONTACT US

Microbiome Metabolome Integration Platform

GET STARTED

Click to use the respective services.

Scroll

Overview

Welcome to MMIP Server.

This server is designed for prediction of the metabolites that could be produced by the microbial community during different health condition by using their metagenomics data (16S-rRNA). Furthermore, it also helps you to compare your real-time metabolomic (untargeted) data with predicted metabolite and helps to see what could be the most probable source of the metabolites.

⊘ Helps in predicting metabolites from the metagenomic data.

⊘ Helps in finding the most probable source of the metabolite (microbe responsible for the production).

⊘ Predict important feature at OTU, Compound, Enzyme level by machine learning approaches.

⊘ Helps in interlinking important feature.

You can access preprint here: Anupam Gautam, Debaleena Bhowmik, Sayantani Basu, Abhishake Lahiri, Wenhuan Zeng, Sandip Paul. Microbiome Metabolome Integration Platform (MMIP): a web-based platform for microbiome and metabolome data integration and feature identification. ***bioRxiv*** , 2023; doi: https://doi.org/10.1101/2023.04.04.535534

Services

Please Select any one of the two module to start your analysis.

Module-I

Perform general prediction of compounds from metagenomic data (16S-rRNA)

Module-II

Allow user to perform comparative analysis between real-time metabolic data and one being predicted

Click on either of the tabs to go to the respective "upload" pages.

(Fig. 1)

UPLOAD PAGES

Module-I:

Details of different releases

Returns to the homepage. ←- - - - - ┐

MMIP

HOME   MANUAL   **RELEASE-NOTES**   CONTACT US

Note: please read the manual before use or look at example dataset for appropriate input formats. Several new features have been implemented (please check Release-Notes). Please notify us if you encounter any bugs or issues.

Name

Enter name

Email

Enter Email

Institute Name

Enter Institute Name

Download example dataset for Greengenes 💾 Download example dataset for 'Other databases' 💾

Select OTU Table: [Choose file] No file chosen   Download example OTU file 💾

Select Metadata: [Choose file] No file chosen   Download example Metadata file 💾

Select Column for Grouping: [v] Select Group: [v]

Select Database: ○ 18may2012 ○ 13_5 ● 13_8 ○ Other Database - - - - - -

Do you want to provide below mentioned files: ○ Yes ● No - - - - - -

Select representative sequences: [Choose file] No file chosen   Or Paste Link: [____]

Select taxonomy file: [Choose file] No file chosen   Or Paste Link: [____]

Enter abundance value to filter biom table (1%=0.01,0.25%=0.0025): [0]

Select Model: ● SVM ○ RandomForest ○ DecisionTree

Submit Job	Example Result

Text

↳ *Select machine-learning model*
↳ *Add the taxonomy file* **(.tsv or .txt Format)**
↳ *Add rep-set file* **(.fasta Format)**   *Add abundance filtering value to filter OTU/ASV table*
↳ *Add the Metadata file* **(.txt Format)**
↳ *Add the OTU/ASV table* **(.biom Format)**   *Select in-case of databases other than Greengenes*
          (In order to select "Other database", kindly click "Yes" below and provide the corresponding files.)

Select Greengenes database version used

(Fig. 2)

Module-II:

(Fig. 3)

RESULT PAGES

USER AND JOB RELATED INFORMATION:

To search keywords or values within the table.

Each of the tabs takes the user to the respective result sections.

MMIP   JOB   METABOLITES   ALPHA-DIVERSITY   TAXONOMY   BETA-DIVERSITY   PATHWAY   FEATURE   CONTACT US

Job Summary

Show 10 ∨ entries   Search: [   ]

S.NO ▲	Content	Description
1	Job Id	215
2	User Name	Name of User
3	Biom File Name	otu_table_20000.biom
4	Meta Data File Name	mapping_file.txt
5	Grouping Column	Description
6	Total Number of Sample Group1 (CRC)	20

1

Showing 1 to 10 of 12 entries   Previous   1   2   Next

(*Fig. 4*)

1   *Scroll the table for additional information including number of samples in group 2, Institute name, email id, start date, time and status of the job.*

METABOLITES:

Add range of desired p values.

Click on the compound Ids to get to the KEGG page of the respective compound.

Metabolites

Minimum p-value: [   ]
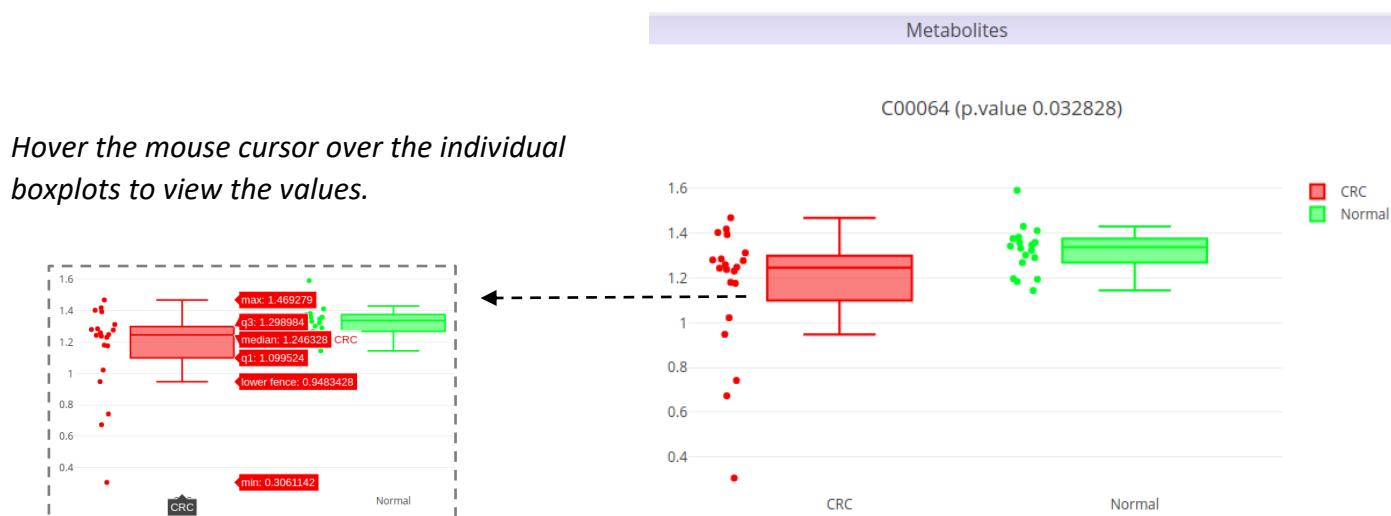
Maximum p-value: [   ]

VISUALIZE ALL COMPOUND

2

Show [10 ▾] entries

Search: [   ]

S.NO ▲	KEGG Compound ID ⇕	p-value ⇕	Box Plot ⇕
1	C00064	0.032828	🖼
2	C00117	0.035297	🖼
3	C00120	0.014059	🖼
4	C00167	0.046764	🖼
5	C00185	0.003461	🖼
6	C00194	0.024378	🖼

| S.NO ▲	File Name ⇕	Download ⇕
1	Statistically Significant Compound Detail (kruskal-walis test)	🔍💾
2	Statistically Significant Compound Detail (Wilcox test)	🔍💾
3	Statistically Significant Compound Detail (Anova-1-way test)	🔍💾
4	Statistically Significant Compound Detail (t-test)	🔍💾
5	All Compound CMP Score	🔍💾

3

4

Showing 1 to 5 of 5 entries

Click to get the boxplot of the compound in the same page (above the table).

Metabolites

C00064 (p.value 0.032828)

Hover the mouse cursor over the individual boxplots to view the values.

max: 1.469279
q3: 1.298984
median: 1.246328 CRC
q1: 1.099524
lower fence: 0.9483428

min: 0.3061142

CRC   Normal
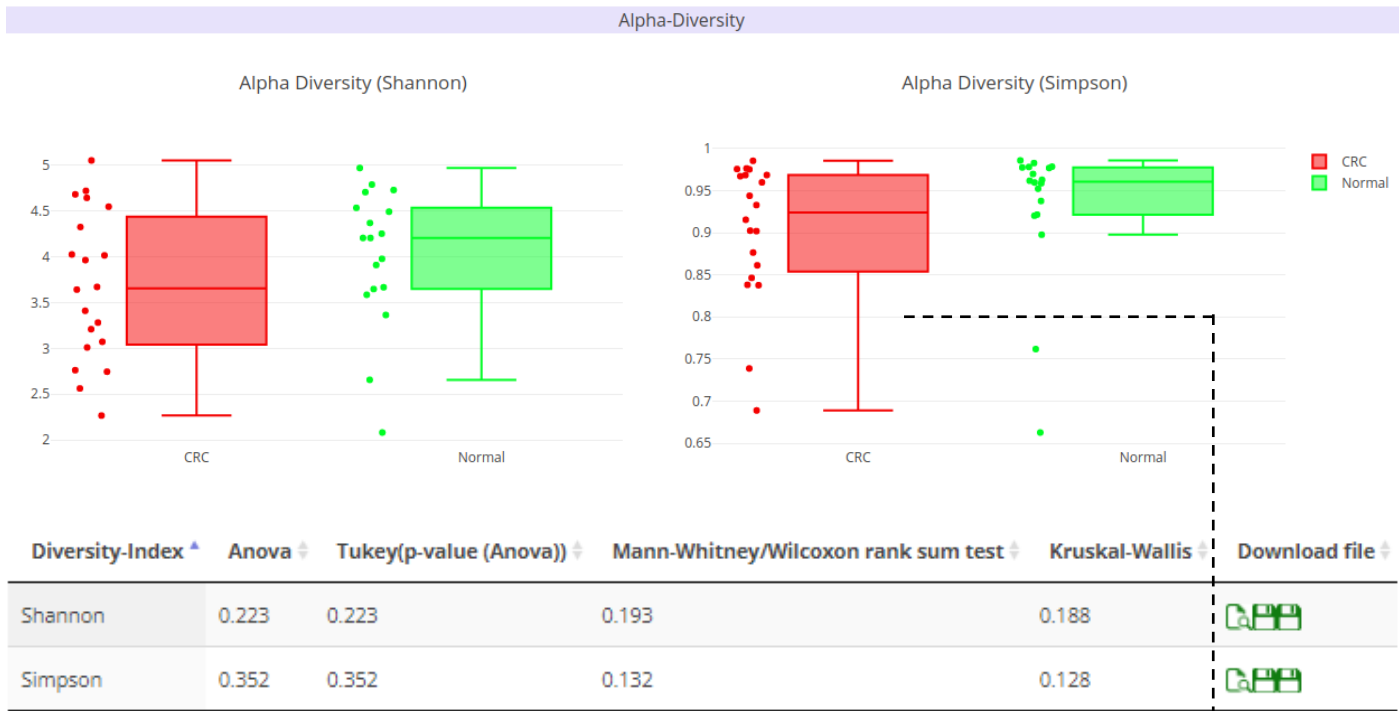
CRC
Normal

CRC   Normal

(Fig. 5)

2 *Click the tab to view all the compounds and more. The following page will open with the significant compounds. The boxplot shows the median values and differences between the two groups.*


Hover the mouse cursor over the individual boxplots to view the values.
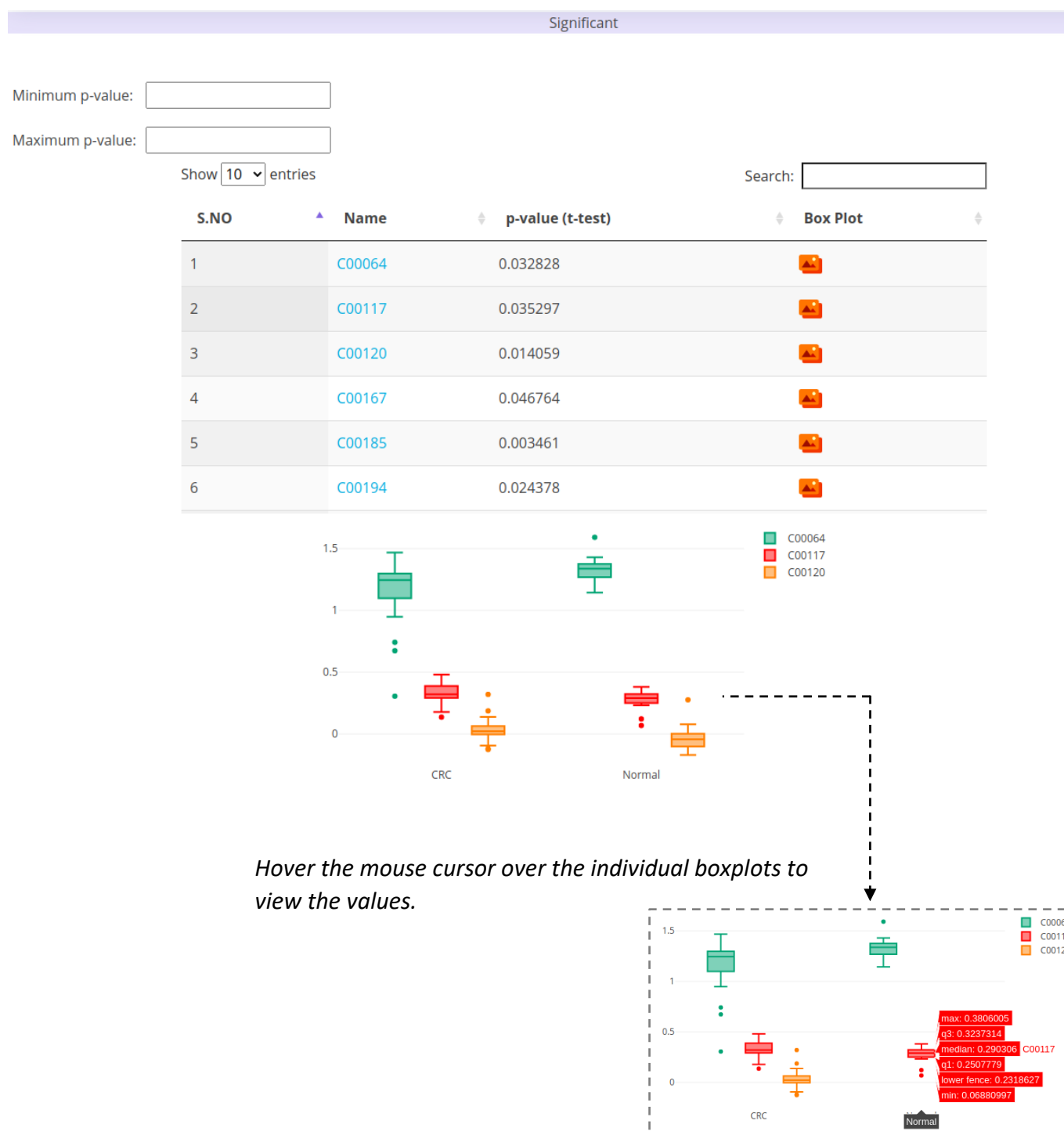
(*Fig. 6*)

3 *Click to download the data table (tsv format).*

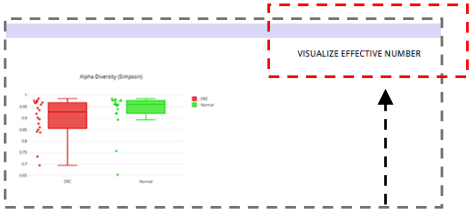**4** *Click to open the data table in a new webpage (tsv format).*

ALPHA DIVERSITY:

Diversity-Index ▲	Anova ⇕	Tukey(p-value (Anova)) ⇕	Mann-Whitney/Wilcoxon rank sum test ⇕	Kruskal-Wallis ⇕	Download file ⇕
Shannon	0.223	0.223	0.193	0.188	
Simpson	0.352	0.352	0.132	0.128	

Showing 1 to 2 of 2 entries

Hover the mouse cursor over the individual boxplots to view the values.
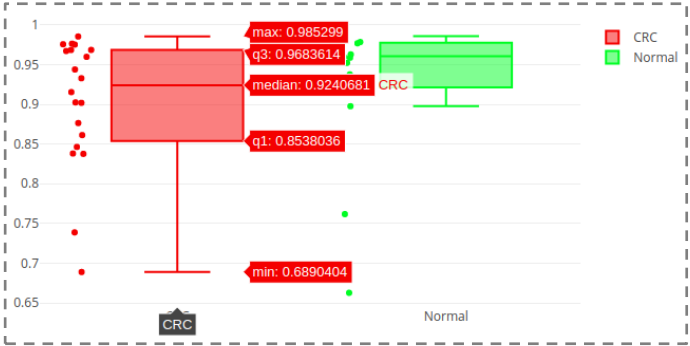
Note: **EFFECTIVE NUMBER** added.
*(Click to go to a new page and view.)*

(*Fig. 7*)

TAXONOMIC DIVERSITY:

5

Sample-wise taxonomic profile of constituent microbes.

6

Taxonomy

TAXONOMICALLY SIGNIFICANT ∨

VISUALIZE KRONA PLOT ∨

Phylum Level Taxonomy Plot for CRC

Phylum Level Taxonomy Plot for Normal

Genus Level Taxonomy Plot for CRC

Genus Level Taxonomy Plot for Normal
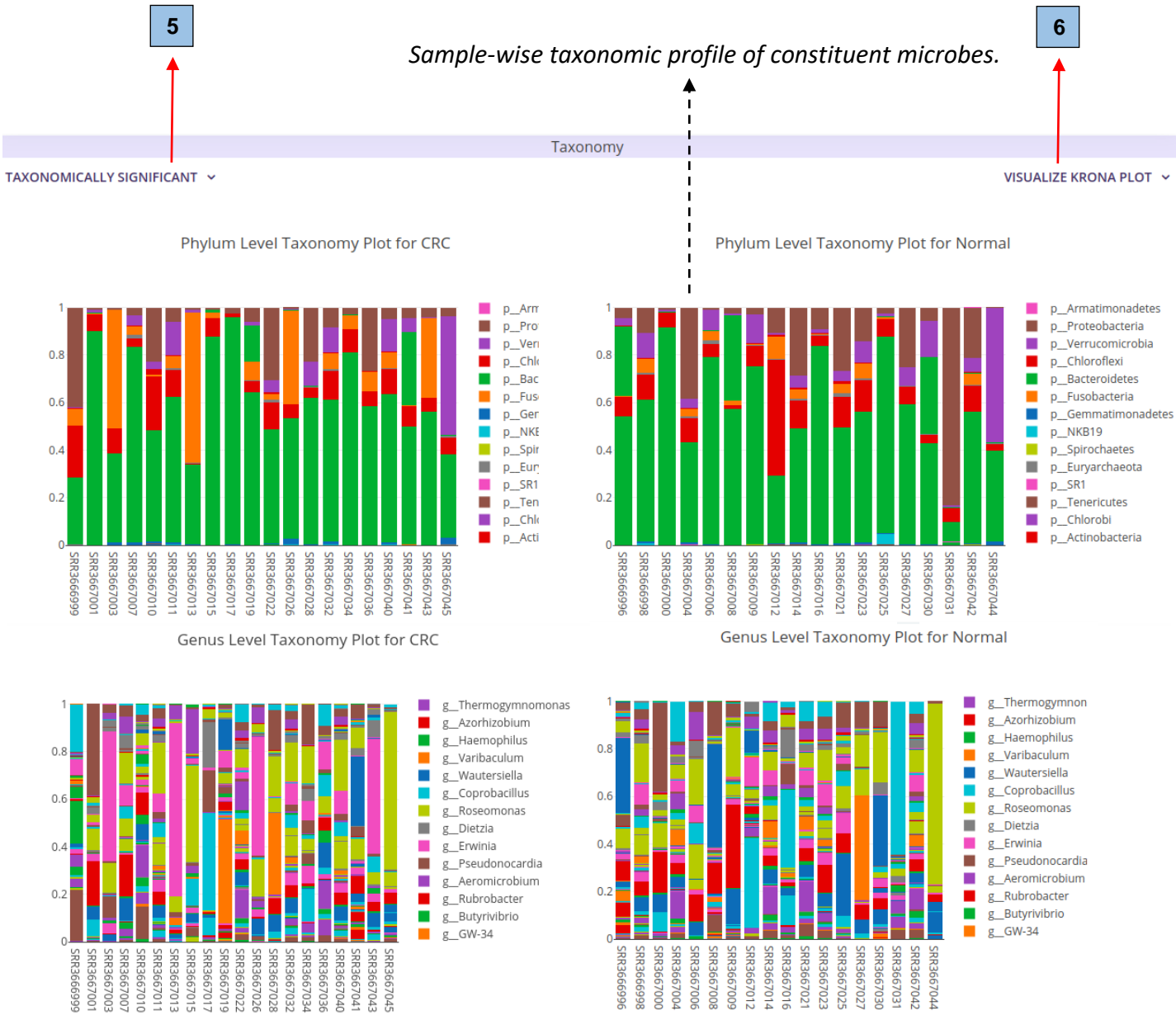
S.No ▲	File Description	Download File
1	Phylum Level Taxonomy Data Group1	
2	Phylum Level Taxonomy Data Group2	
3	Genus Level Taxonomy Data Group1	
4	Genus Level Taxonomy Data Group2	
5	Statistically Significant OTUs Group1	
6	Statistically Significant OTUs Group2	

Showing 1 to 6 of 6 entries

(Fig. 8)

5 *Click on this tab to find a drop-down list (shown below), from which user can select the desired taxonomic level to check significant ones and is visualized in a new webpage, also depicted below.*
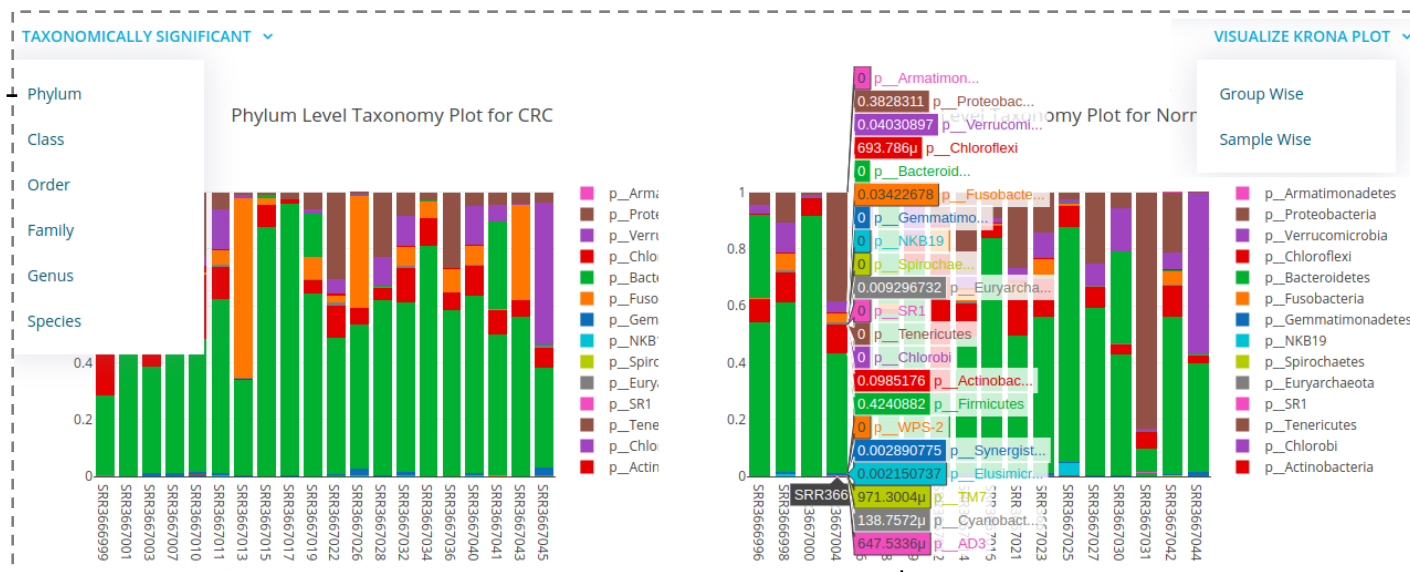
(Fig. 9)

Click on one of the taxonomic level to go to the following page showing the taxa that are significant.

Hover the mouse cursor over the individual samples to view the different taxa and their relative abundance.

(Fig. 10)

6 *Click on this tab to open the Krona plots representing the taxonomic abundance profile, both group-wise and sample-wise (as shown in fig. 9). This is an interactive plot, and is another way of depicting the taxonomic profiles.*

Click on the different segments to see its constituent taxa and at different levels. The ring can be collapsed back into its higher levels of taxa.

(*Fig. 11*)

BETA DIVERSITY:

..	F.Model	p-value
Bray-Curtis (Adonis test)	0.760695687284777	0.797

Showing 1 to 1 of 1 entries

File Name	Download
Beta-Diversity p-value detail (adonis)	
Beta-Diversity Plot Data & Distance Matix	
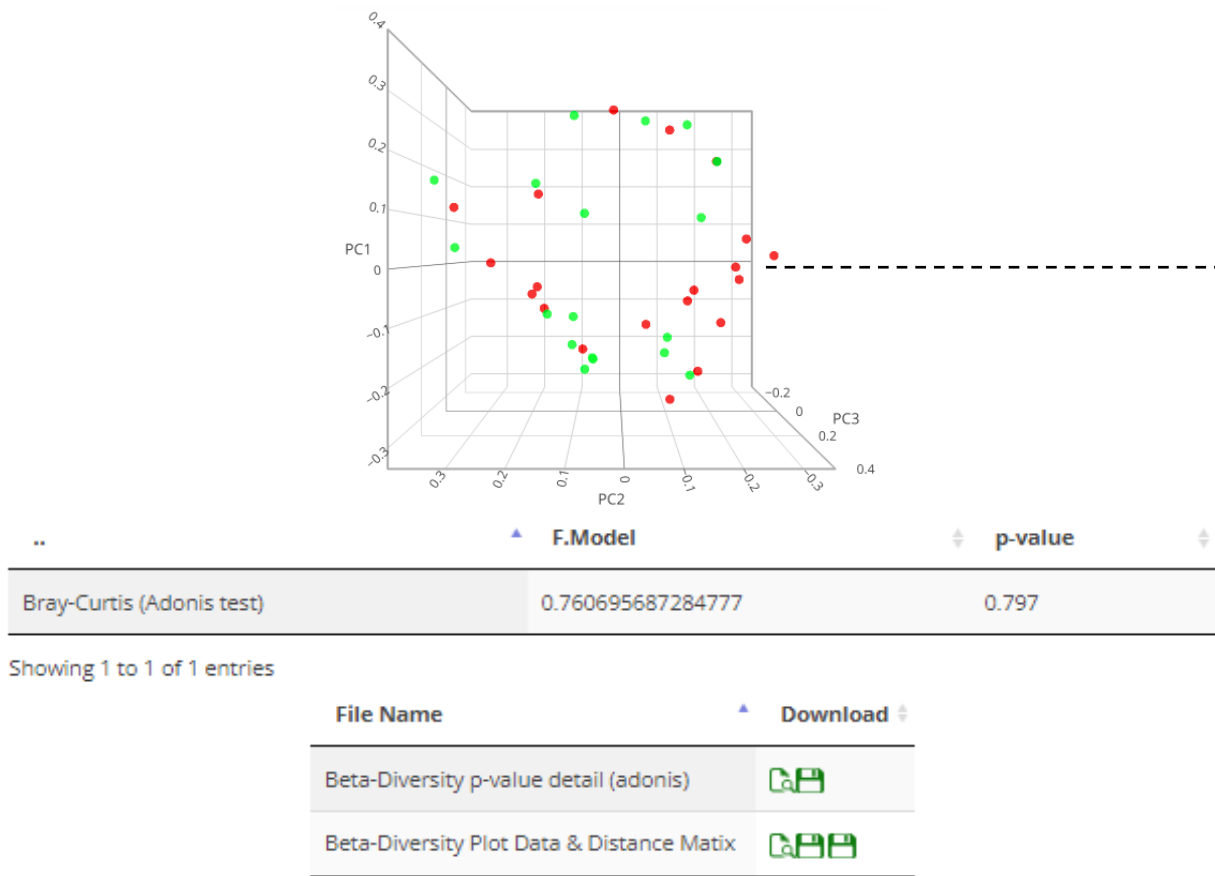
Hover the mouse cursor over the individual dots to view the values and sample number.
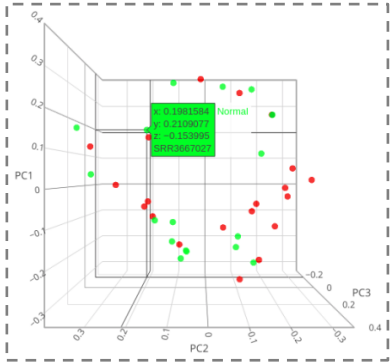
Note: **UNIFRAC (WEIGHTED/UNWEIGHTED)** added.
(Click to go to a new page and view.)

(Fig. 12)

PATHWAY ANALYSIS:

Pathway

SIGNIFICANT PATHWAY ˅   VISUALIZE KRONA PLOT
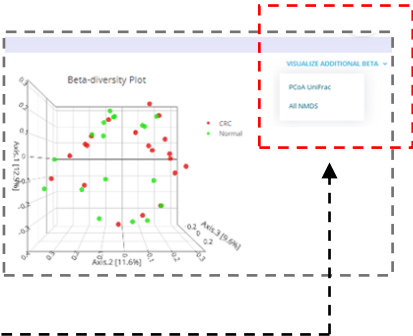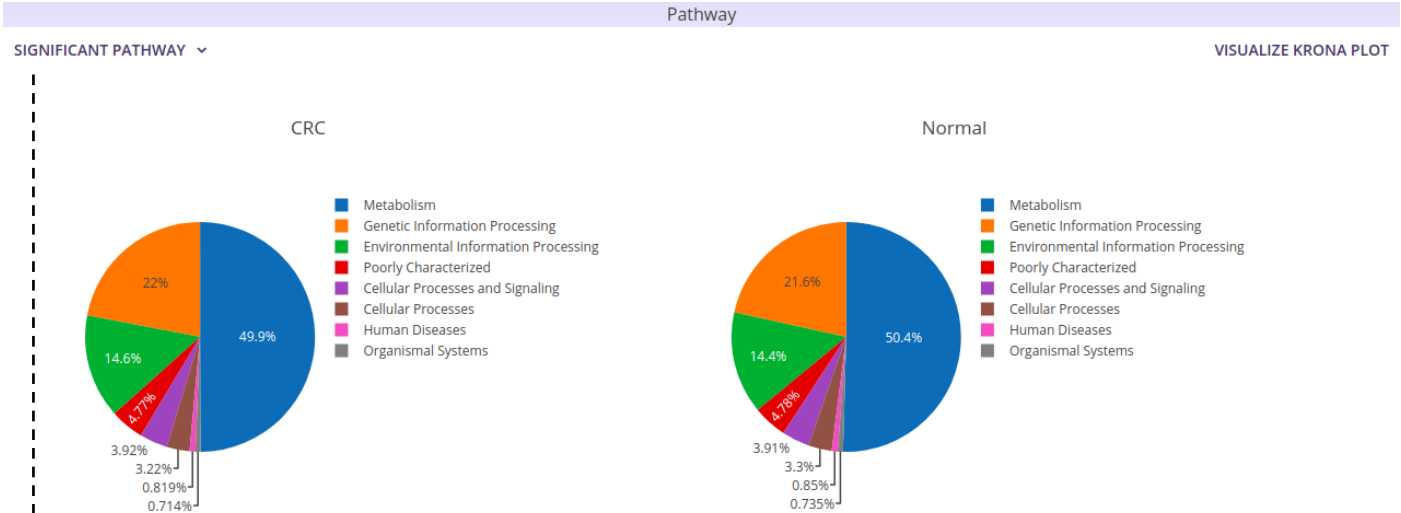
Click on the tab for a drop-down list showing the different pathway levels as found in KEGG database.

7

8
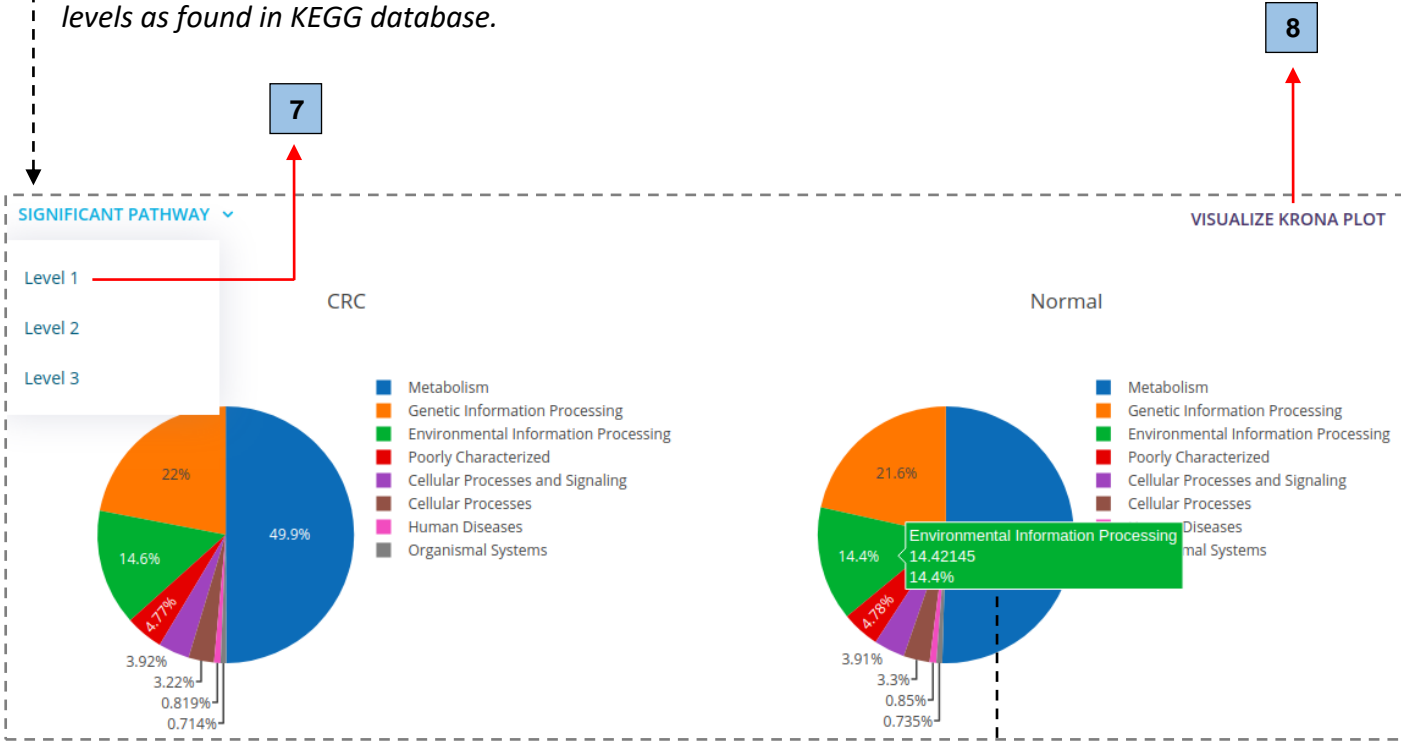
Hover the mouse cursor over the different section to view the pathway name and the percentage of its contribution to overall metabolism in that group.

(Fig. 13)

7 *Click on one of the levels to check the significant pathways in that level and it opens in a new page.*

Significant level1

Minimum p-value: [   ]

Maximum p-value: [   ]

Show [10 ▼] entries   Search: [   ]

S.NO ▲	Name	p-value (t-test)	Box Plot
1	Cellular Processes	0.771470	📊
2	Cellular Processes and Signaling	0.913329	📊
3	Environmental Information Processing	0.976748	📊
4	Genetic Information Processing	0.983255	📊
5	Human Diseases	0.710409	📊
6	Metabolism	0.822901	📊
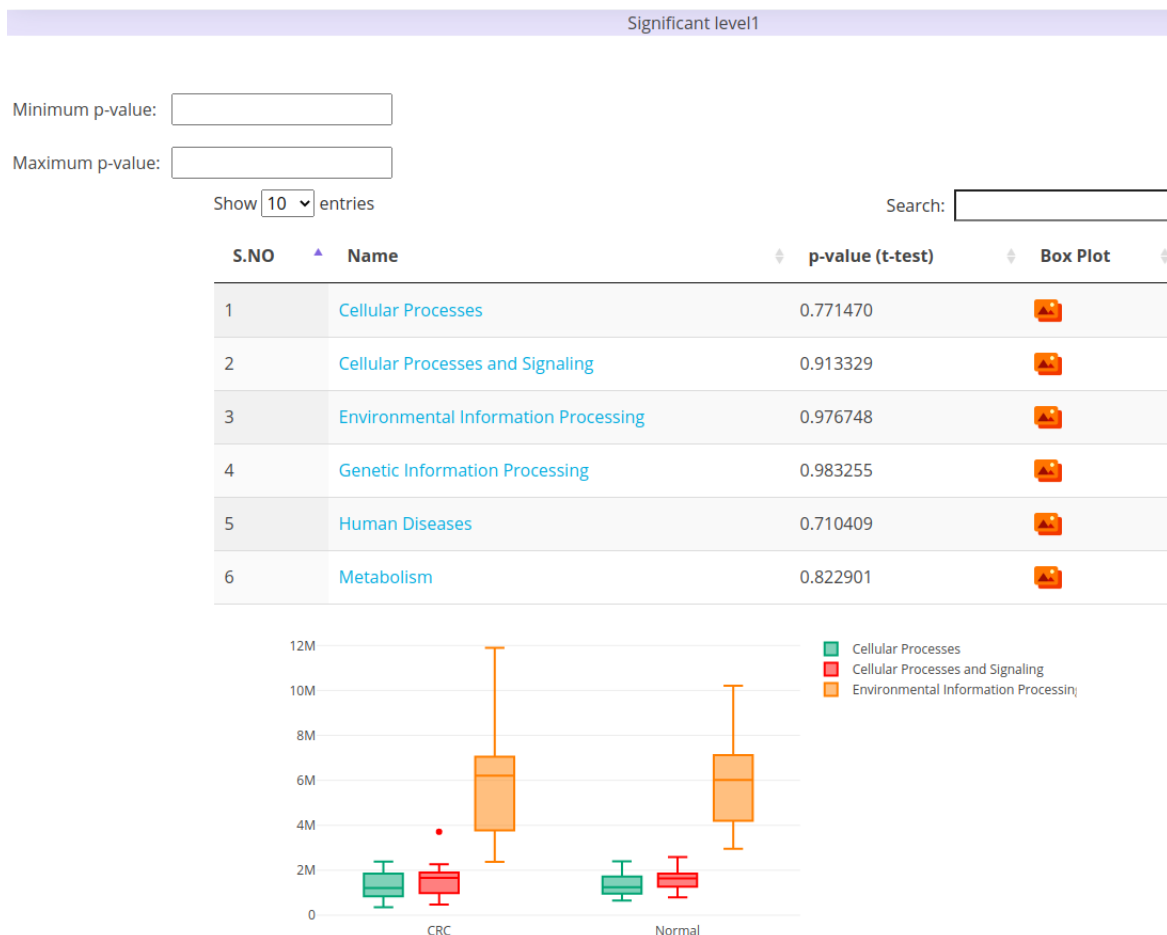
(Fig. 14)

8 *Click on this tab to visualize the Krona plot for metabolic pathway abundances between two groups.*
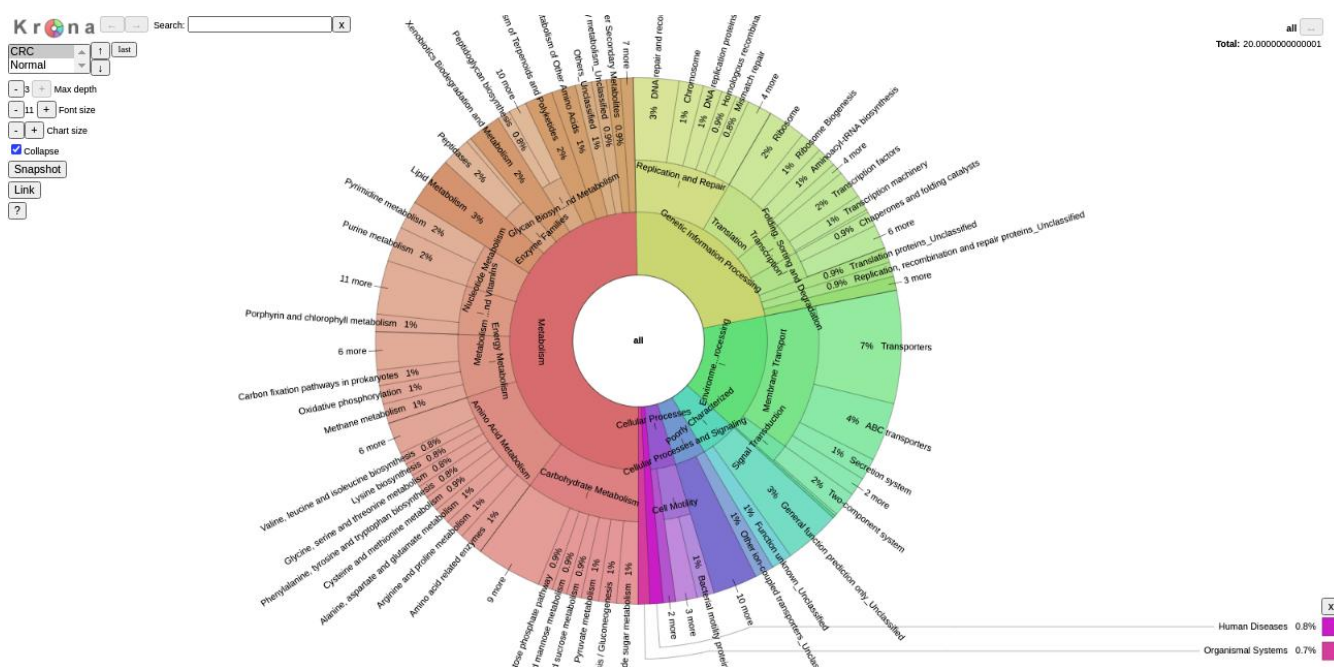
(Fig. 15)

REAL TIME V/S PREDICTED METABOLITES:

In case of MMIP- MASS this additional tab is observed.

(Fig. 16)

(Fig. 17)

FEATURE PREDICTION:

Feature ROC Curve	

ROC Curve For Normal Feature   ROC Curve For CRC Feature

Mean ROC curve (area =0.510)   Mean ROC curve (area =0.510)

Download ROC Curve Data for Normal Feature   Download ROC Curve Data for CRC Feature

Show 10 entries   Search:

Model Evaluation metric ▲	F1 Score	Accuracy Score	Recall Score	Precision	ModelParamter & ConfusionMatrix
Model Name: SVM	0.313	0.421	0.278	0.357	

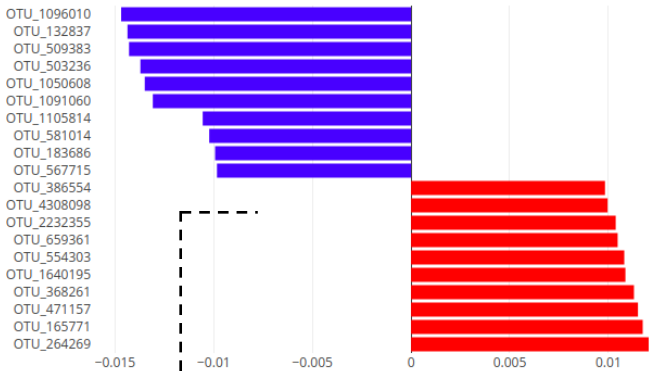Showing 1 to 1 of 1 entries   Previous   1   Next

ROC curve and evaluation metrics of the selected machine learning model shown.

(Fig. 18)

Feature Identified

Community Structure

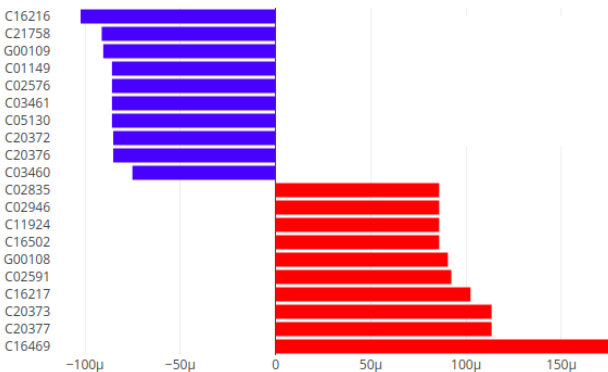**KEGG**

Download All Feature For Community Structure 💾
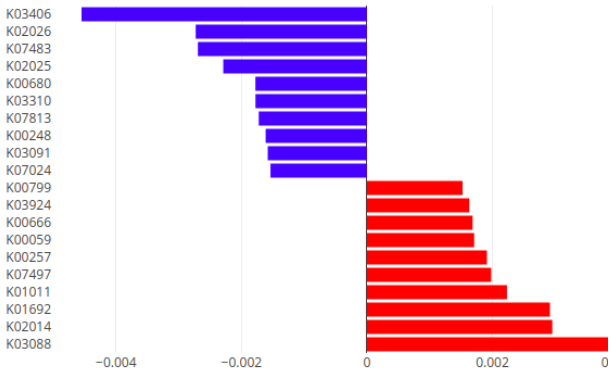
Download All Feature For KEGG 💾

The feature table can be downloaded (.tsv format).

CMP

The top 10 features each of positive (in red) and of negative (in blue) are shown here.
Also hover mouse over diagram to view values.

Download All Feature For CMP 💾

Feature Category

Normal Feature		
Show 10 entries		Search
KEGG Compound ID ▲	**p-value (Anova)** ⇕	**Visualize cor**
C02591	0.56764712027688	LINK
C02835	0.5254858214149772	LINK
C02946	0.5254858214149772	LINK
C11924	0.5254858214149772	LINK
C16217	0.5107272046458358	LINK
C16469	0.6895104179008892	LINK

CRC Feature		
Show 10 entries		Search
KEGG Compound ID ▲	**p-value (Anova)** ⇕	**Visualize cor**
C01149	0.5254858214149752	LINK
C02576	0.5254858214149752	LINK
C03460	0.6294221232575476	LINK
C03461	0.5254858214149752	LINK
C05130	0.5254858214149752	LINK
C16216	0.5107272046458358	LINK

9

(Fig. 19)

9 *Click on the link to visualize the following page which shows the predicted microbial source of metabolites (predicted as feature by our algorithm), in a tree as well as tabular form. [all p values provided in adjacent brackets]*

Expanded view of the tree.

Save the tree (found on the top of the diagram).

The adjacent feature links represented in a tabular format and can be downloaded.

Predicted Metabolite

The pathway that the predicted metabolite is associated with.

All the OTU ids (microbes) that could probably be the source of the predicted metabolite.

(Fig. 20)