PACFM (Pathway Analysis with Circos for Functional Metagenomics) User Manual

PACFM is a software tool developed for the analysis of biological pathways in metagenomics projects. It provides the users with an improved way of visualizing pathway abundance data in addition to presenting a wide array of methods for filtering and normalization.

Usage

Input files

The input file of PACFM is a tab separated text document with the first field assigned to the KEGG Orthology (KO) identifiers and the rest of the fields contain abundances or read counts in each sample.

```
Name    sample1 sample2 sample3 sample4 sample5
K00001  6       5       3       6       2
K00002  0       0       0       1       0
K00003  2       1       1       4       1
K00004  1       1       0       0       1
K00005  1       0       0       0       0
K00006  1       1       0       0       0
K00008  3       0       0       1       2
K00010  4       2       0       0       1
K00011  1       0       0       2       0
K00012  6       3       2       7       2
K00013  6       7       0       4       4
K00014  3       4       2       3       2
```

Graphical User Interface

i. Panels

a. Toolbar: The toolbar on top of the main panel of the GUI consists of 4 buttons. First button allows the user to select an input file from the file system. The second button resets the pathway analyses performed on the data and allows the user to restart analysis. Third and fourth buttons display the pathway analysis and the Circos option panels consecutively.

Figure 1. Pacfm toolbar.

b. Main display panel: The main display panel is divided into four sections where the raw data, representing the abundances of enzymes and the assigned pathways at the KEGG Brite database's hierarchy levels and the normalized data, representing the output of the pathway analysis module are shown. The final Pacfm plot output, drawn either by the raw or the normalized data is also found under the main display panel in order to leverage the comparative analysis between the default and manipulated pathway abundances.
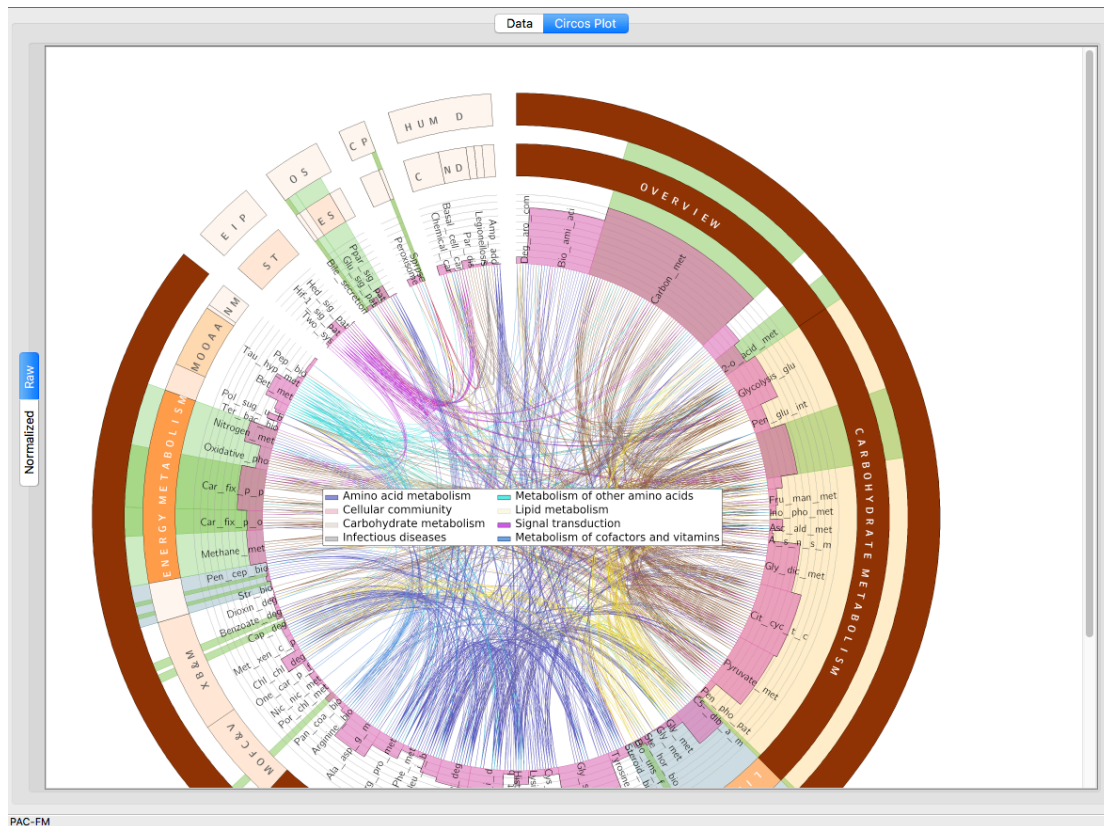
Figure 2: Main display panel

c. Pathway Analysis Panel

See Section II.b in the manuscript.

Figure 3: Pathway analysis panel

d. Circos Panel

Circos panel comprises options to configure colors and plot types of each hierarchy level to be displayed. The default plotting scheme includes heat-map representations for the two broadest levels of the KEGG Brite database and a bar chart representation at the pathway level. The interdependence settings between different pathways are also specified in the Circos panel. Users have the ability to include or exclude dependencies to be shown between pathways, represented by the links in the center of the final plot. Circos panel also allows the users to set up their own specialized pathway annotation profiles within the KEGG Brite database by highlighting independent pathways with a color coded label. This option allows the users to focus their attention on the pathways that are directly relevant for their study of interest, for example in targeted metagenomics.
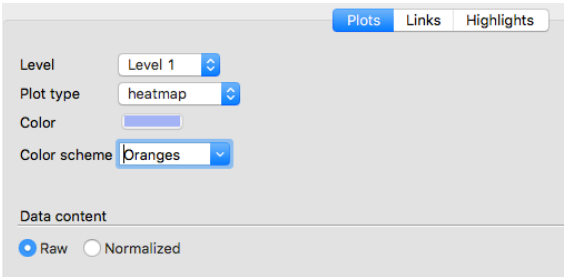

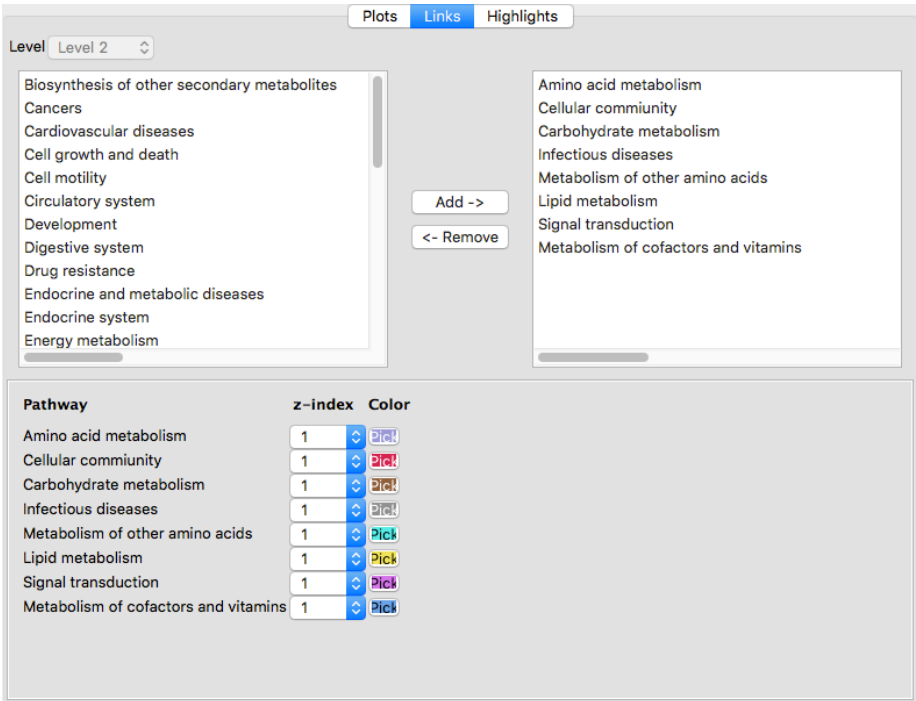
Figure 4: Circos panel plot settings.
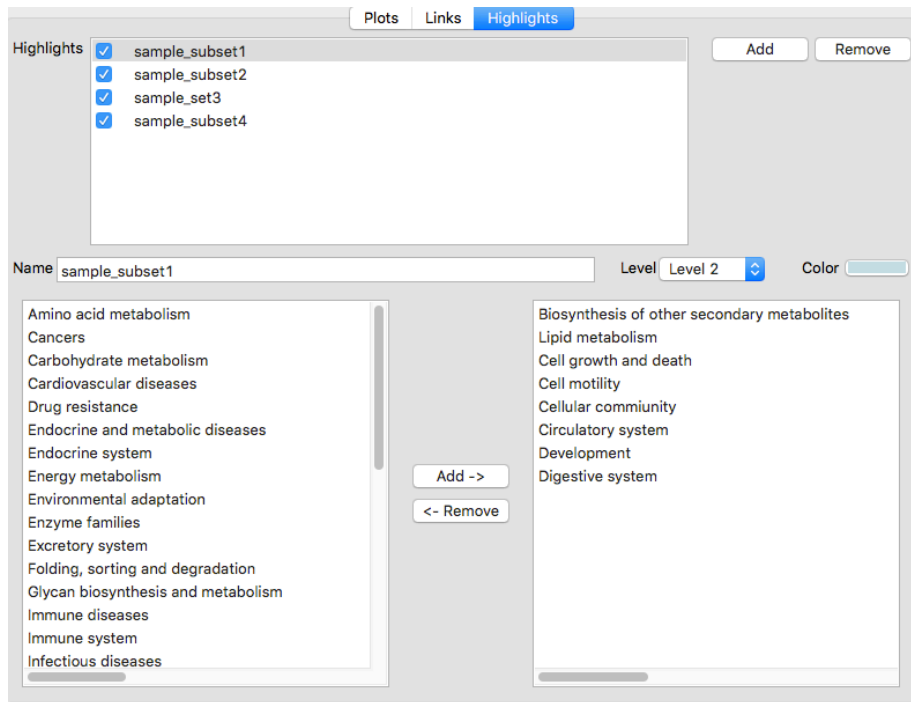

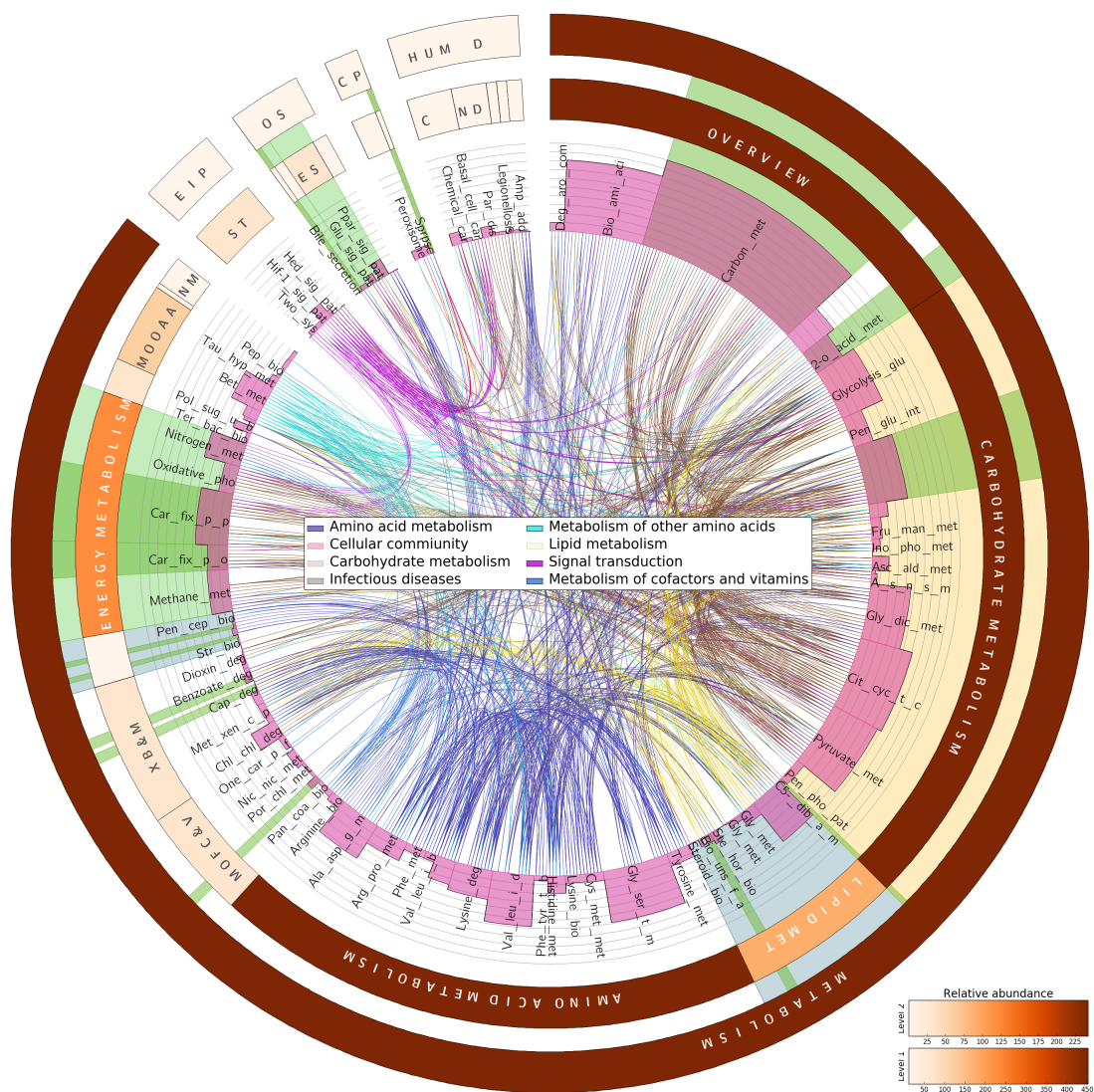Figure 5: Circos panel link settings.

Figure 6: Circos panel highlight (manual database subset curation) settings

Figure 7: Final Pacfm output.