Supplementary file 6: Multivariate Data analysis methods in iVikodak

Description of data analysis methods pertaining to inferred function data for deciphering Core Functions, Top functions, Differentiating functions and Function driven correlation networks are provided in this file

Index

Core Functions

Core functions refer to those set of functions which appear consistently in most samples of a given metagenomic environment (e.g samples pertaining to Healthy individuals) at a minimum threshold of abundance.

In the context of iVikodak, Core Functions are computed using a bootstrapped approach as described below:

a. A random set of samples, equal to 75% of the total population size, is picked from the total population of samples pertaining to a given class of metagenomic environment.

b. Median abundance of all the inferred functions for the given environment is computed using the random set.

c. A minimum prevalence of 0.2 x Highest median abundance is set for core assignment.

d. For each function in the given randomly picked environment, assessment is performed for maximum prevalence proportion. Maximum prevalence proportion refers to the percentage of samples (in the random set) in which the given function had abundance greater than minimum prevalence.

e. All those functions are tagged as 'Tentative-Core' for the random set, which pass the prevalence proportion of 75%.

f. This process (between step a – e) is repeated 100 times and a boostrap score is assigned to all the Tentative Cores. Bootstrap score refers to the number of iterations in which a function appeared as

Tentative Core.

g. Those functions which have a minimum bootstrap score of 95% are tagged as Core Set of Functions for the given environment.

Visualization of the core functions is accomplished using a heatmap, which is generated using the rank normalized abundance profile of the union of core functions for all the classes of samples.

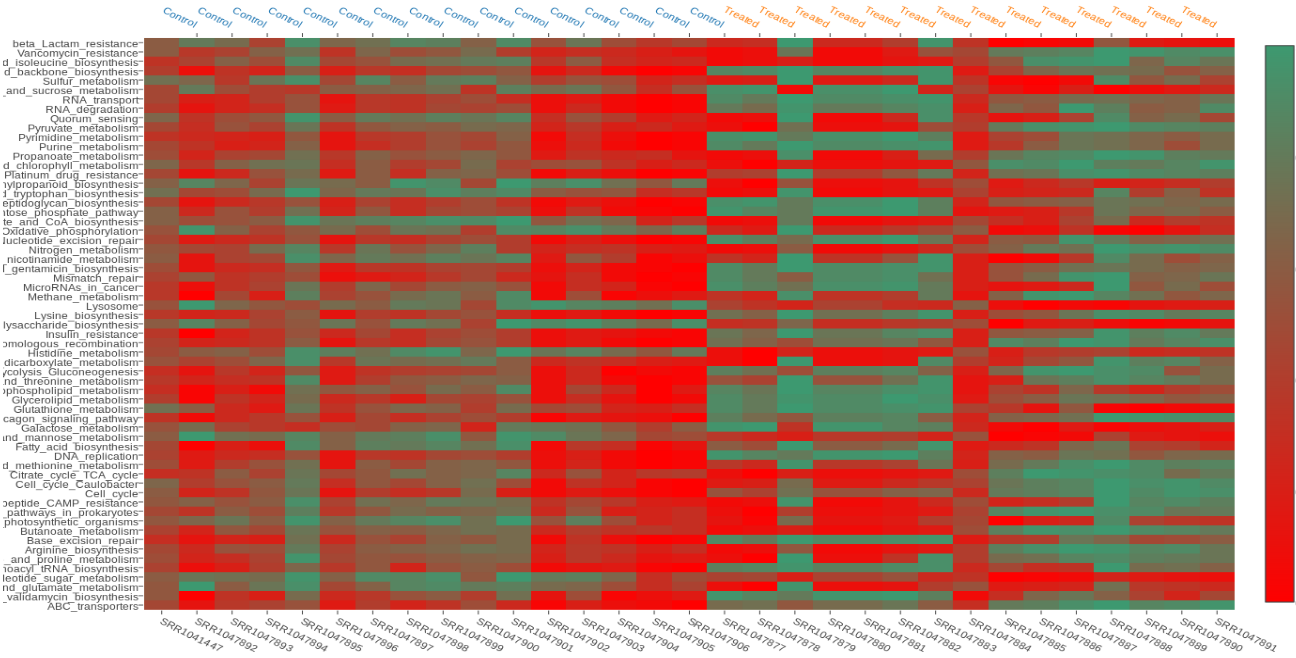


Figure SP1: Example of Core Function Visualization

Top Functions

Top Functions are computed using the median abundances of each inferred function in the environment. For the purpose of clarity, only Top 5 functions are selected from each class of samples in a population (e.g Top functions for Healthy class or Control class of samples).

Computation of Top Function set is done for Level 3 functions (most specific functional level, e.g Pyruvate metabolism) as well as for Level 2 functions (broader level in functional hierarchy, e.g Carbohydrate Metabolism).

Box plots are computed using the union set of functions deciphered for all classes of samples in the submitted taxonomic data.

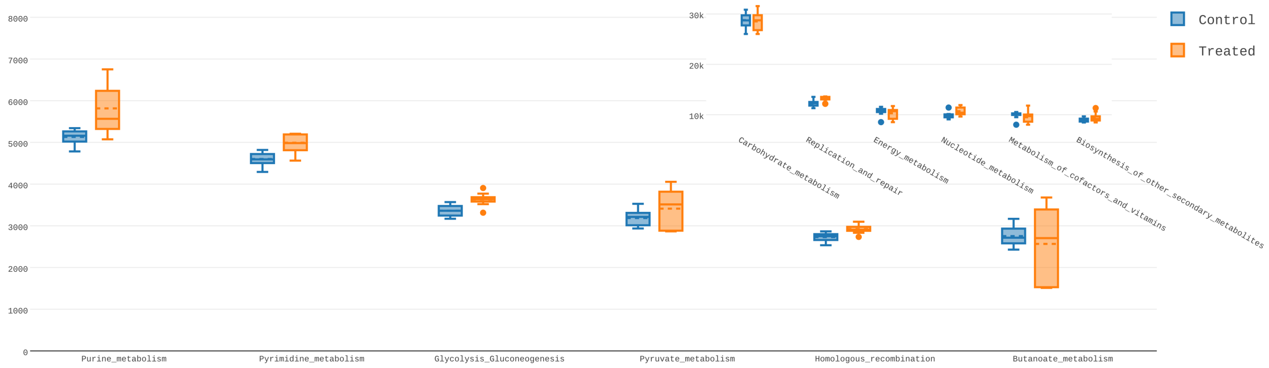


Figure SP2: Example of Top Functions Visualization (Main plot represents Level 3 functions, while inset represents Level 2 functions). Individual graphs for each level are also available in the dashboards

Home: Go to index

Differentiating Functions

Differentiating functions are those which can potentially act as signatures of distinction between multiple classes of metagenomic datasets (as supplied in the meta-data).

Differentiating features analysis is performed using classical tests:

1. Kruskal walis test (for multiple groups)

2. Wilcoxon rank sum test (for pair-wise differentiation)

As performed for the Top Functions, differentiating feature analysis is performed for both Level 3 and Level 2 functions are the hierarchy of differentiating functions is thereafter mapped to the Sankey based cladogram in the dashboard.

Uncorrected and BH corrected p-values are reported in the downloadable results, while the visualization module uses only BH corrected p-values.

As described in the main manuscript, Global Mapper module can infer functions using Pathway Exclusion Cut-off (PEC) thresholds. The ISFA module of iVikodak employs the functional inference at various PEC thresholds to arrive at differentiating functions of high-confidence. The Batch module of ISFA performs the statistical tests for the functional abundance matrices of all PEC values and then reports the PEC profile for each differentiating function. PEC profile is generated using binary information of 0/1 for each function, where 0 refers to absence of signature trait for a function at a given PEC value, while 1 refers to the presence of signature trait. A function which has signature trait at all or most PEC values is a potentially most significant differentiator.

Visualization of the PEC profile is accomplished using a heatmap, where green cells indicate presence of signature trait, while red cells indicate absence of signature trait.

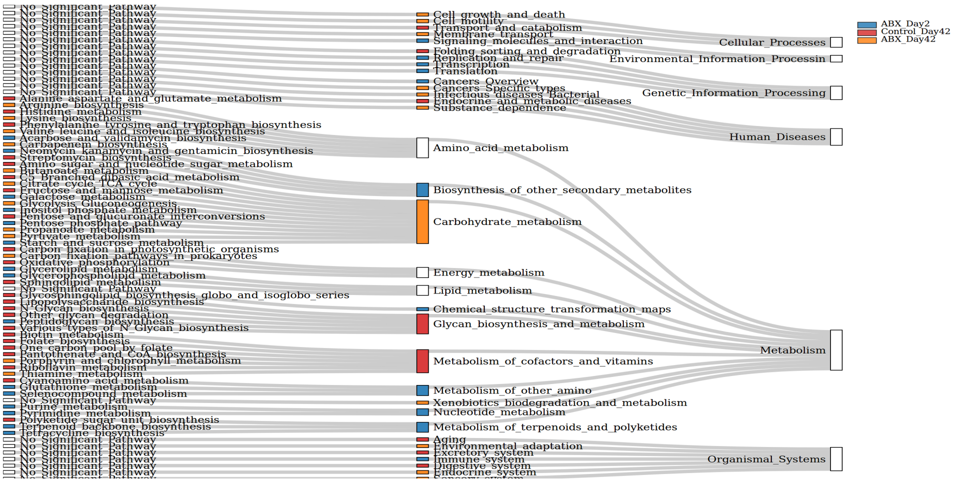

No_Significant_pathway
No_Significant_Pathway
No_Significant_Pathway
No_Significant_Pathway
No_Significant_Pathway
No_Significant_Pathway
No_Significant_Pathway
No_Significant_Pathway
No_Significant_Pathway
No_Significant_Pathway
No_Significant_Pathway
No_Significant_Pathway
No_Significant_Pathway
No_Significant_Pathway
Alanine_aspartate_and_glutamate_metabolism
Arginine_biosynthesis
Histidine_metabolism
Lysine_biosynthesis
Phenylalanine_tyrosine_and_tryptophan_biosynthesis
Valine_leucine_and_isoleucine_biosynthesis
Acarbose_and_validamycin_biosynthesis
Carbapenem_biosynthesis
Neomycin_kanamycin_and_gentamicin_biosynthesis
Streptomycin_biosynthesis
Amino_sugar_and_nucleotide_sugar_metabolism
Butanoate_metabolism
C5_Branched_dibasic_acid_metabolism
Citrate_cycle_TCA_cycle
Fructose_and_mannose_metabolism
Galactose_metabolism
Glycolysis_Gluconeogenesis
Inositol_phosphate_metabolism
Pentose_and_glucuronate_interconversions
Pentose_phosphate_pathway
Propanoate_metabolism
Pyruvate_metabolism
Starch_and_sucrose_metabolism
Carbon_fixation_in_photosynthetic_organisms
Carbon_fixation_pathways_in_prokaryotes
Oxidative_phosphorylation
Glycerolipid_metabolism
Glycerophospholipid_metabolism
Sphingolipid_metabolism
No_Significant_Pathway
Glycosphingolipid_biosynthesis_globo_and_isoglobo_series
Lipopolysaccharide_biosynthesis
N_Glycan_biosynthesis
Other_glycan_degradation
Peptidoglycan_biosynthesis
Various_types_of_N_Glycan_biosynthesis
Biotin_metabolism
Folate_biosynthesis
One_carbon_pool_by_folate
Pantothenate_and_CoA_biosynthesis
Porphyrin_and_chlorophyll_metabolism
Riboflavin_metabolism
Thiamine_metabolism
Cyanoamino_acid_metabolism
Glutathione_metabolism
Selenocompound_metabolism
No_Significant_Pathway
Purine_metabolism
Pyrimidine_metabolism
Polyketide_sugar_unit_biosynthesis
Terpenoid_backbone_biosynthesis
Tetracycline_biosynthesis
No_Significant_Pathway
No_Significant_Pathway
No_Significant_Pathway
No_Significant_Pathway
No_Significant_Pathway
No_Significant_Pathway
No_Significant_Pathway
Cell_growth_and_death
Cell_motility
Transport_and_catabolism
Membrane_transport
Signaling_molecules_and_interaction
Folding_sorting_and_degradation
Replication_and_repair
Transcription
Translation
Cancers_Overview
Cancers_Specific_types
Infectious_diseases_Bacterial
Endocrine_and_metabolic_diseases
Substance_dependence
Amino_acid_metabolism
Biosynthesis_of_other_secondary_metabolites
Carbohydrate_metabolism
Energy_metabolism
Lipid_metabolism
Chemical_structure_transformation_maps
Glycan_biosynthesis_and_metabolism
Metabolism_of_cofactors_and_vitamins
Metabolism_of_other_amino
Xenobiotics_biodegradation_and_metabolism
Nucleotide_metabolism
Metabolism_of_terpenoids_and_polyketides
Aging
Environmental_adaptation
Excretory_system
Immune_system
Digestive_system
Endocrine_system
Sensory_system
Cellular_Processes
Environmental_Information_Processin
Genetic_Information_Processing
Human_Diseases
Metabolism
Organismal_Systems
ABX_Day2
Control_Day42
ABX_Day42

Figure SP3: Example of Cladogram view for differentiating functions. Color of nodes represents the class in which the differentiating function is most abundant

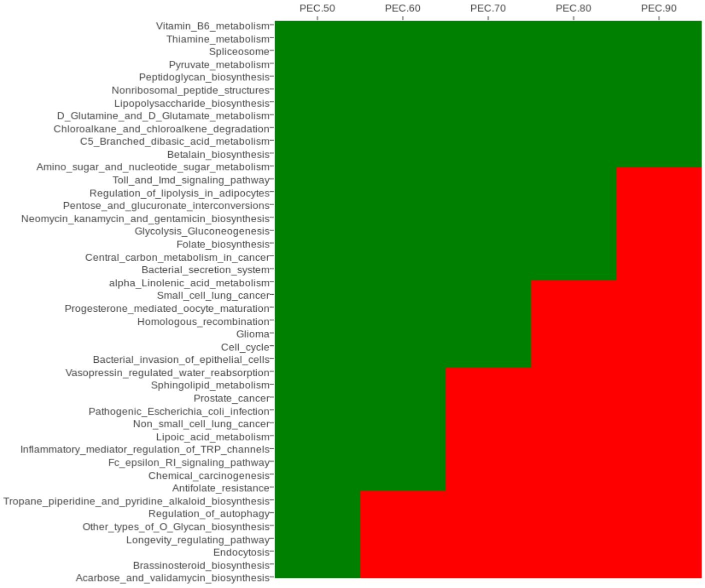
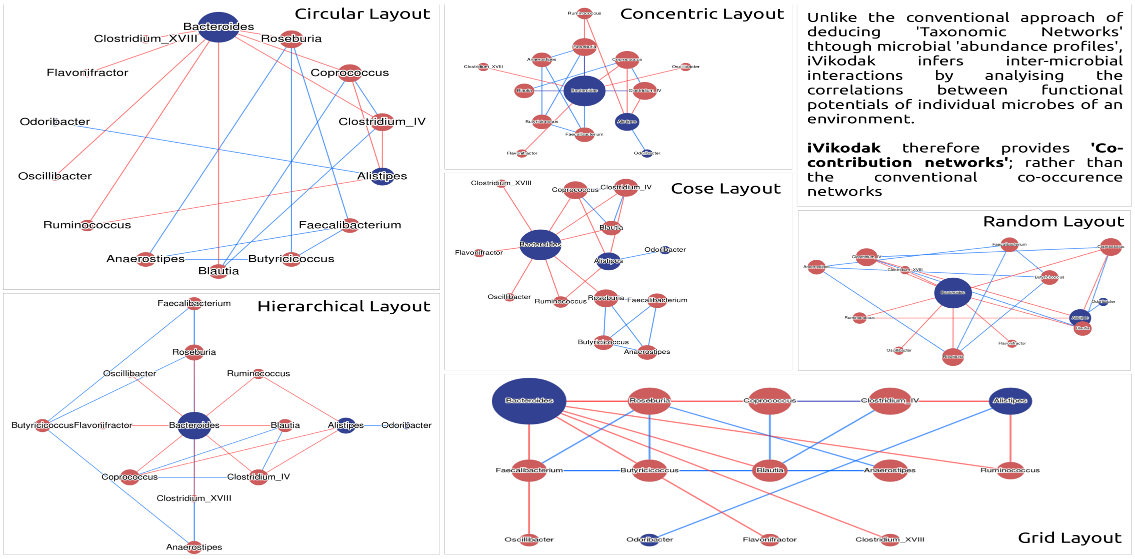


Figure SP4: Example of PEC profile for differentiating functions

Function Driven Correlation Networks

Vikodak introduced a novel concept of inferring inter-microbial interaction patterns using the correlations between their functional potentials (*Nagpal et al*., 2016). iVikodak has realized that concept through automatic computation of correlations between contributing bacteria in a given environment. These networks are termed as 'Co-contribution' networks, because unlike co-occurrence networks derived from the abundance profiles of microbes in an environment, co-contribution networks, as detailed in Vikodak (*Nagpal et al*., 2016), are derived from functional contribution profiles of resident microbes of an environment towards various inferred functions. iVikodak generates contribution profiles at all PEC values as well, thereby enabling the end users to probe function driven microbial interaction patterns using algorithms of choice. Visualization of such networks is enabled using Cytoscape.js and through the use of AJAX and in-house javascripting, various kinds of layouts are enabled for the end users, for all levels of meta-data.



Unlike the conventional approach of deducing 'Taxonomic Networks' thtough microbial 'abundance profiles', iVikodak infers inter-microbial interactions by analyising the correlations between functional potentials of individual microbes of an environment.

iVikodak therefore provides **'Co-contribution networks'**; rather than the conventional co-occurence networks

Figure SP5: Example of network visualizations possible in iVikodak