

Statistics report for HMP2

Author: LM

Date: 06/22/2021

Contents

Introduction	2
All against all: Mantel test	3
One against all: PERMANOVA	4
Each metadata variable against all features: MaAsLin 2	7
Taxonomy	7
MaAsLin2 Plots	8
Pathways	11
MaAsLin2 Plots	12
MaAsLiN2 stratified pathways plots	15
Each data type against all metadata variables: HAllA	19
HAllA taxonomy	19
HAllA pathways	20

Introduction

The data for this project was run through the standard stats workflow. The workflow is composed of four sections.

1. All against all : A mantel test, which computes the correlation between two matrices of the same dimension, is run to compare all points of each data set provided with all points of all other data sets.
2. One against all : A PERMANOVA, used to compare groups of objects, is run with two different approaches to the statistical analysis for each of the data types, eg taxonomy, provided for the study. First the permanova is run to compare a single metadata variable with the data set. Next, in a multivariable analysis, all of the metadata variables are used for the analysis against the data set. For a longitudinal study, where there are multiple time points for each individual, the metadata co-variates that do not vary for each individual are factored into the analysis.
3. Each metadata variable against all features individually: MaAsLin 2 filters, transforms, and then performs a linear model to fit metadata variables to feature data (e.g. taxonomy, pathways), one at a time. If both taxonomy and pathways are provided, plots for the significant pathways, stratified by species are generated.
4. Each data type against all metadata variables : HAllA tests all possible associations of each feature in a data set against all metadata variables. It is run to compare each of the data sets provided for the study against all variables.

All against all: Mantel test

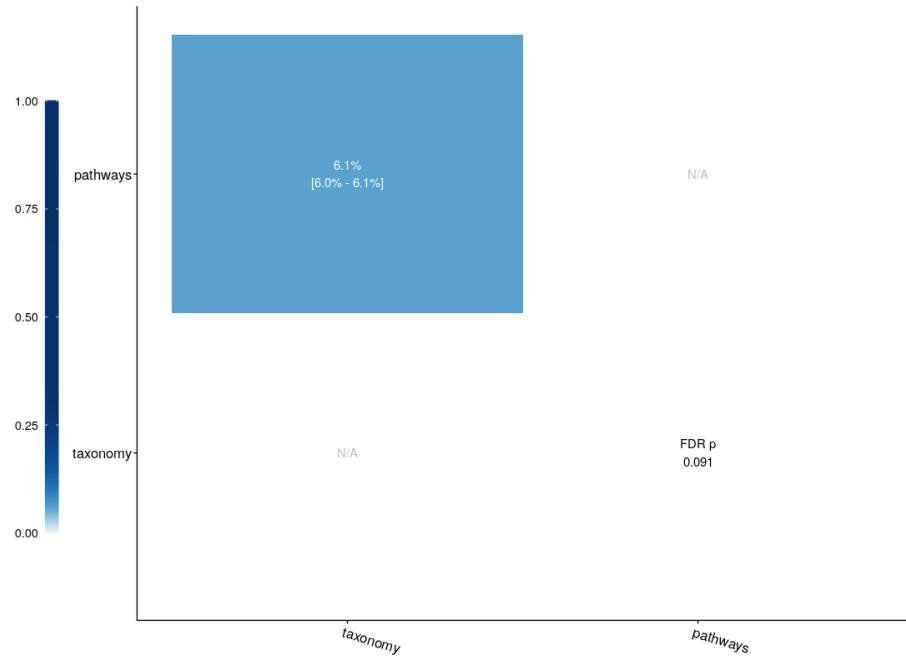


Figure 1: Mantel test using Bray-Curtis dissimilarity

Shown are the observed correlation and bootstrap sample quantiles at 2.5% and 97.5% probabilities in the upper triangle and the Benjamini–Hochberg FDR p-value in the lower triangle using Bray-Curtis dissimilarity.

One against all: PERMANOVA

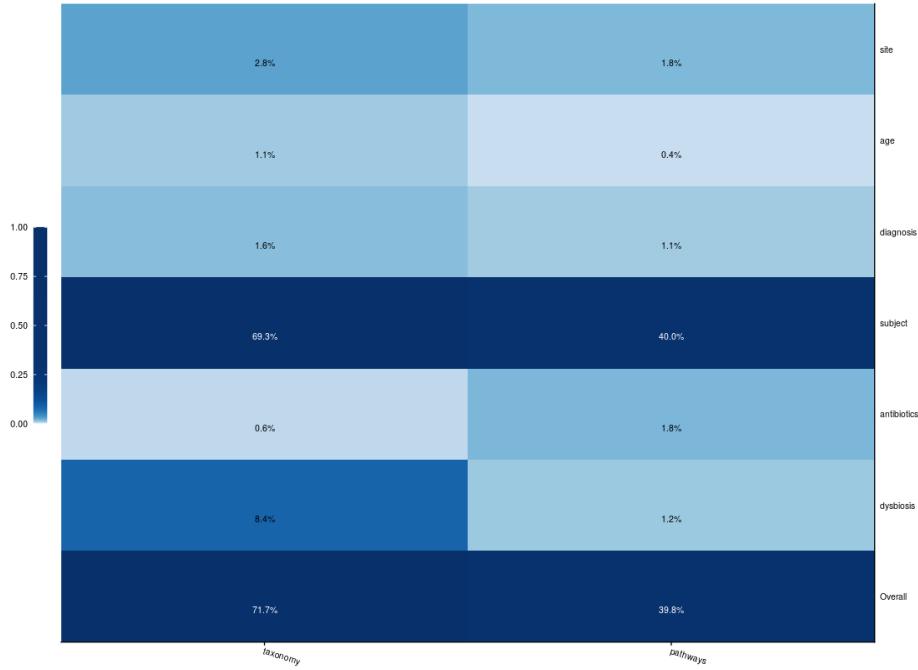


Figure 2: Heatmap of univariable R-squared value with repeated measures

The results are visualized as a heatmap of the percentage of variation explained for each metadata variable (rows) in each meta'omic matrix tested against (columns) to highlight similarities between tests using Bray-Curtis dissimilarity. The results of each PERMANOVA are displayed in a separate barplot as well with the R-squared and FDR adjusted p-values for each metadata variable. Text files of these results are included in the output folder.

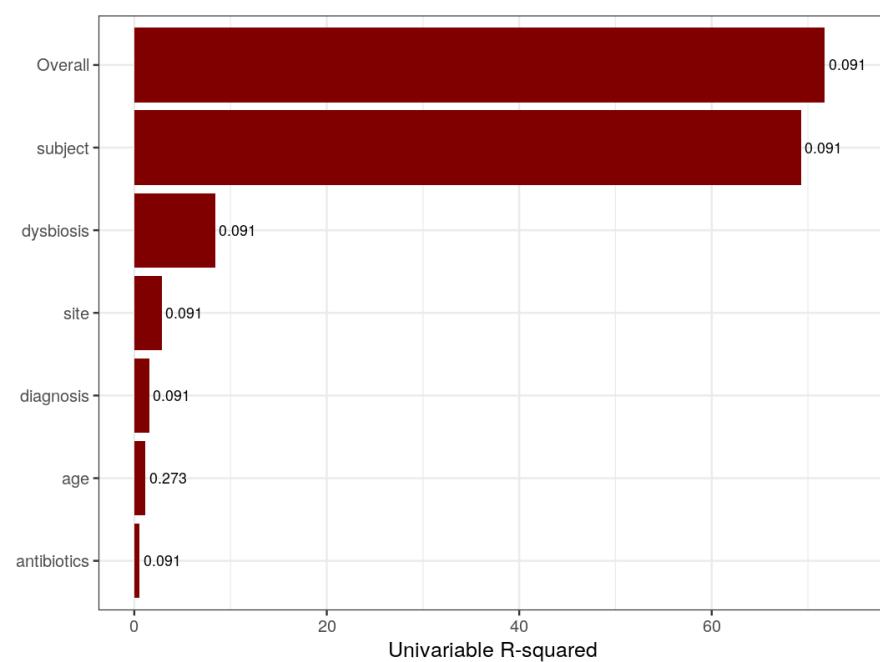


Figure 3: Taxonomy - Bar plot of R-squared value, annotated with the FDR adjusted p-value.

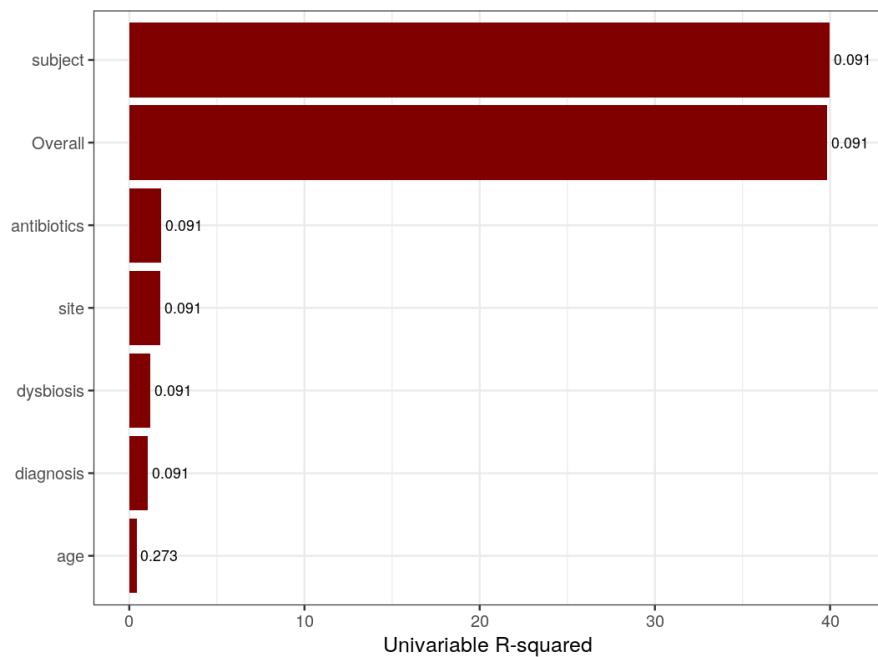


Figure 4: Pathways - Bar plot of R-squared value, annotated with the FDR adjusted p-value.

Each metadata variable against all features: MaAsLin 2

MaAsLin2 is comprehensive R package for efficiently determining multivariable association between clinical metadata and microbial meta'omic features. MaAsLin2 relies on general linear models to accommodate most modern epidemiological study designs, including cross-sectional and longitudinal, and offers a variety of data exploration, normalization, and transformation methods. More detailed information may be found in the [MaAsLin2 User Manual](#).

See the stats workflow log for the MaAsLiN2 commands run. Also note these are a subset of the outputs. Check out the MaAsLin2 results folder for the complete set.

Taxonomy

This report section contains the results from running the taxonomy data through MaAsLin2.

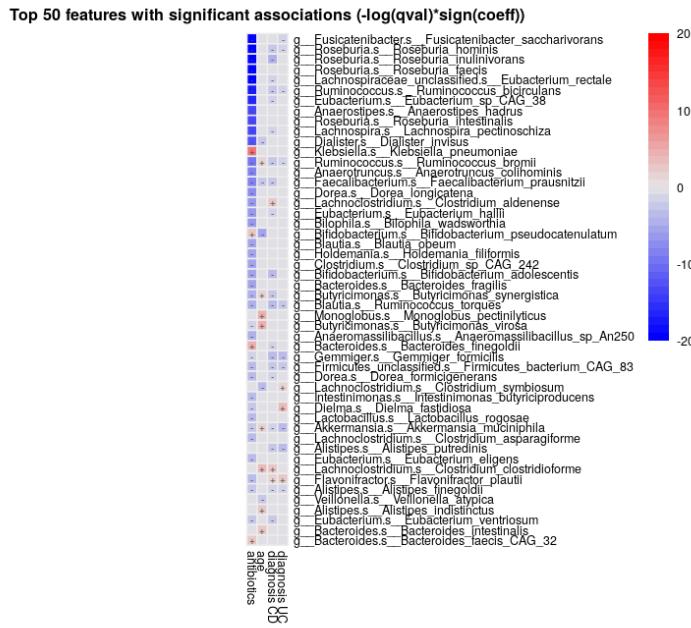


Figure 5: taxonomy heatmap

MaAsLin2 Plots

The most significant association for each metadata are shown. For a complete set of plots, check out the MaAsLin2 results folders.

antibiotics

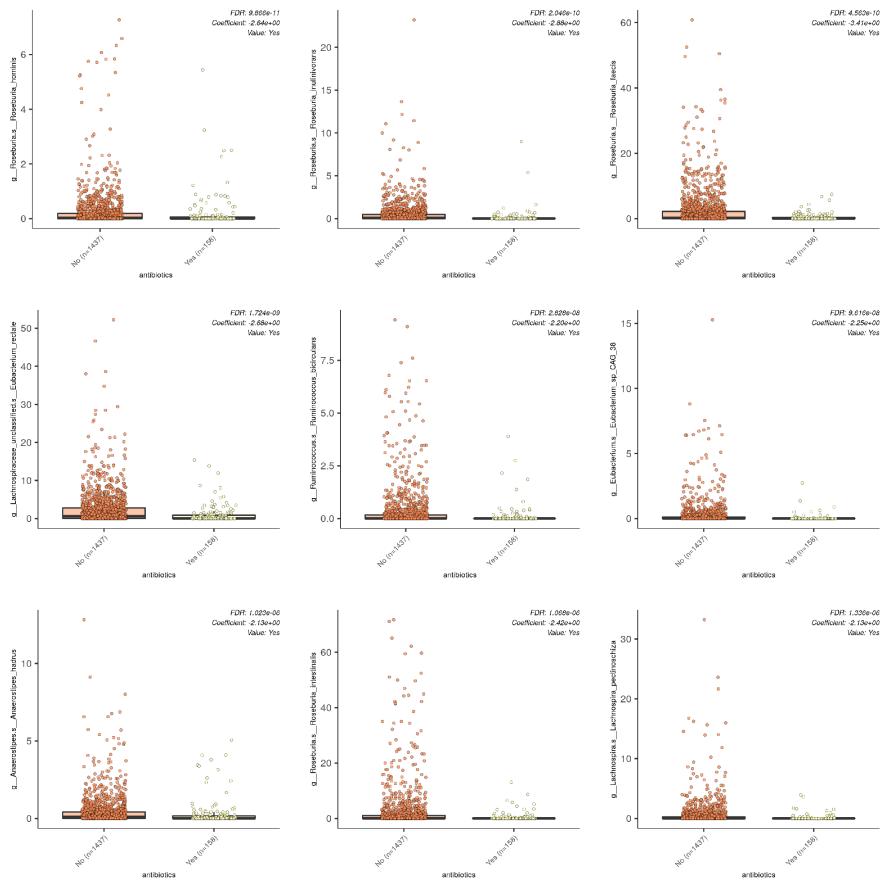


Figure 6: Top antibiotics associations for taxonomy

diagnosis

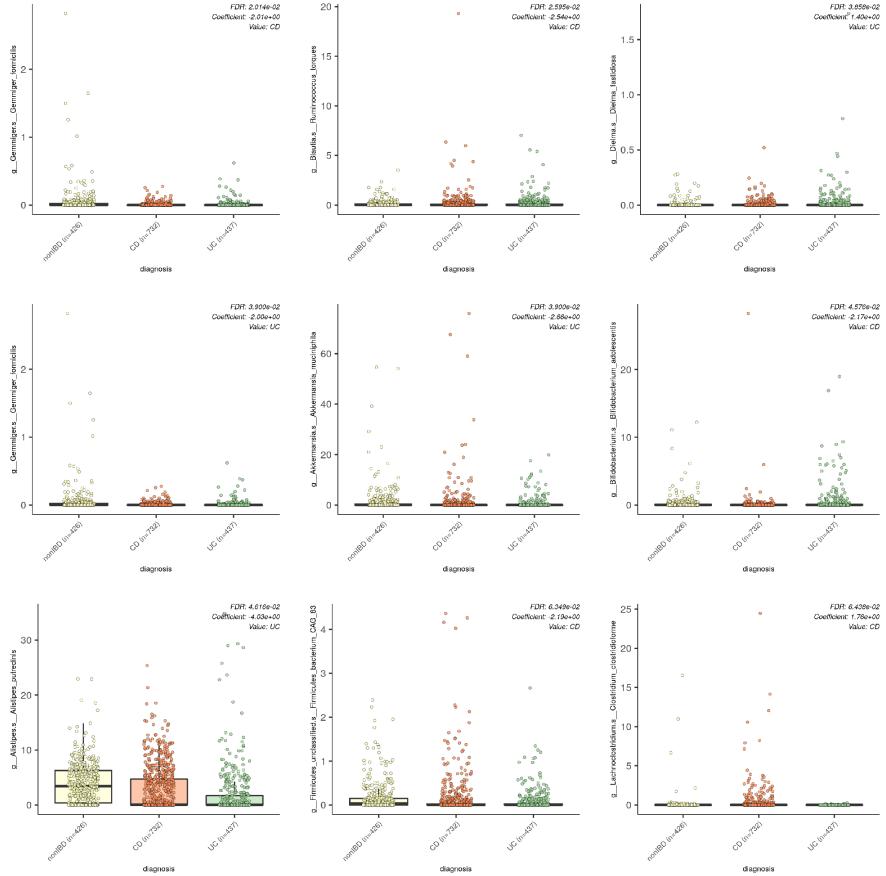


Figure 7: Top diagnosis associations for taxonomy

age

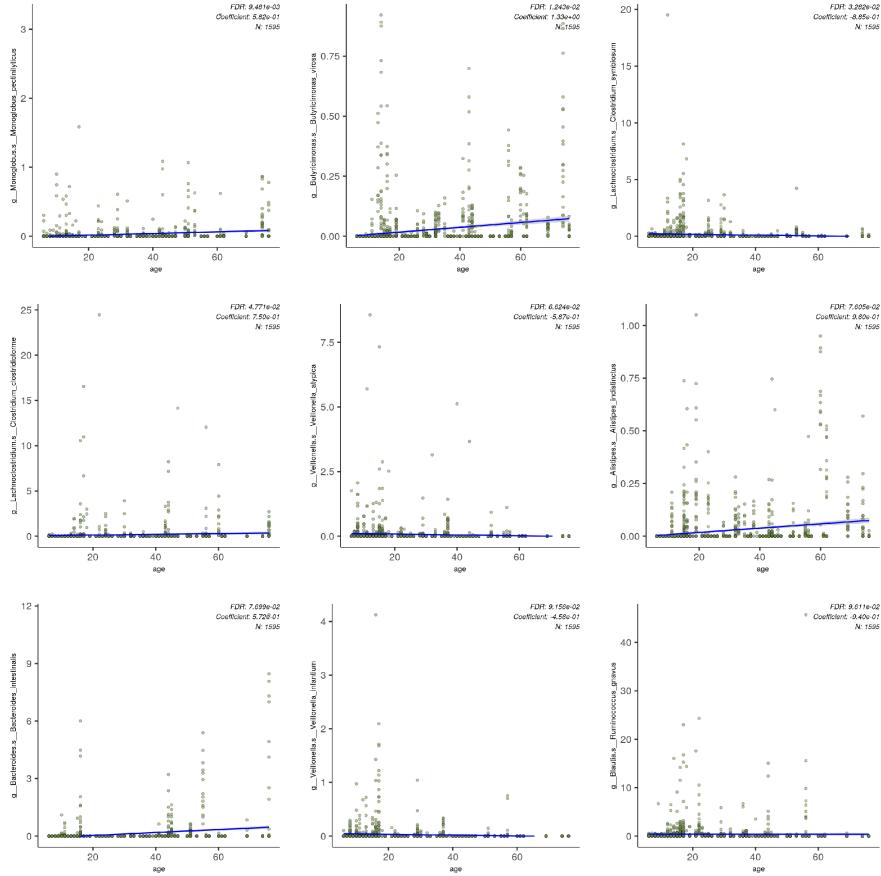


Figure 8: Top age associations for taxonomy

Pathways

This report section contains the results from running the pathways data through MaAsLin2.



Figure 9: pathways heatmap

MaAsLin2 Plots

The most significant association for each metadata are shown. For a complete set of plots, check out the MaAsLin2 results folders.

antibiotics

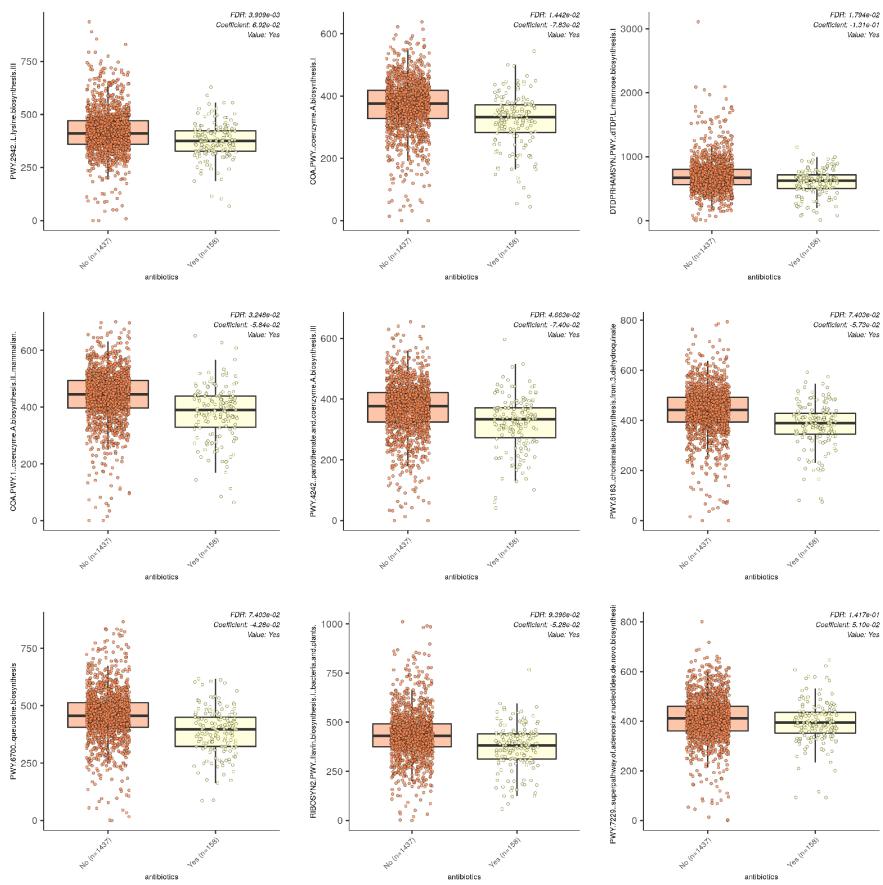


Figure 10: Top antibiotics associations for pathways

diagnosis

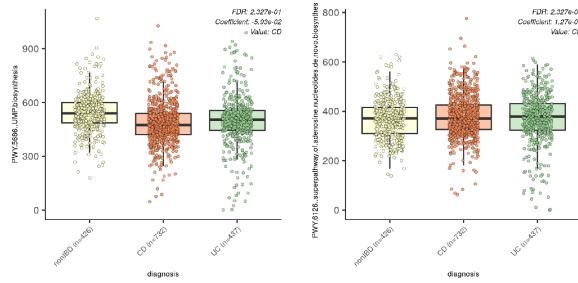


Figure 11: Top diagnosis associations for pathways

age

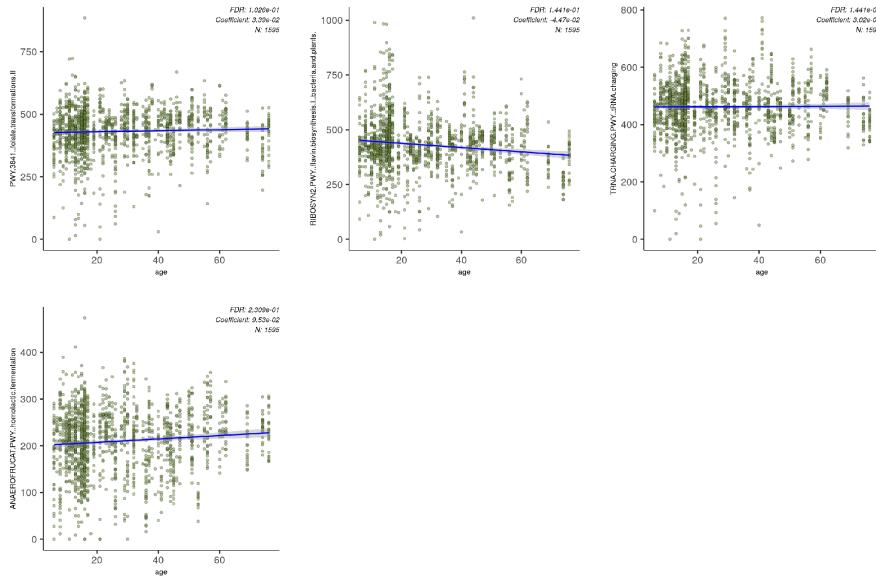


Figure 12: Top age associations for pathways

MaAsLiN2 stratified pathways plots

The abundance for the 3 most significant associations for each of the categorical features are plotted stratified by species. These plots were generated with the utility script included with HUMAnN named humann_barplot.

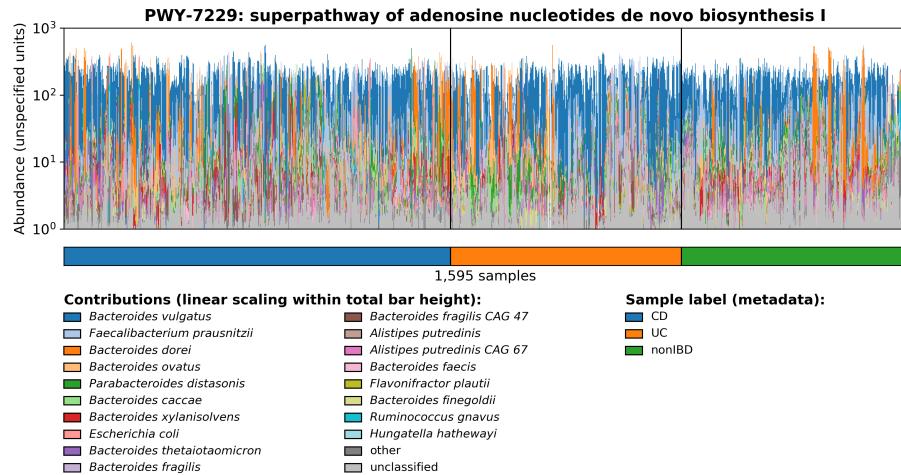


Figure 13: Significant association pathway #1 for metadata focus diagnosis

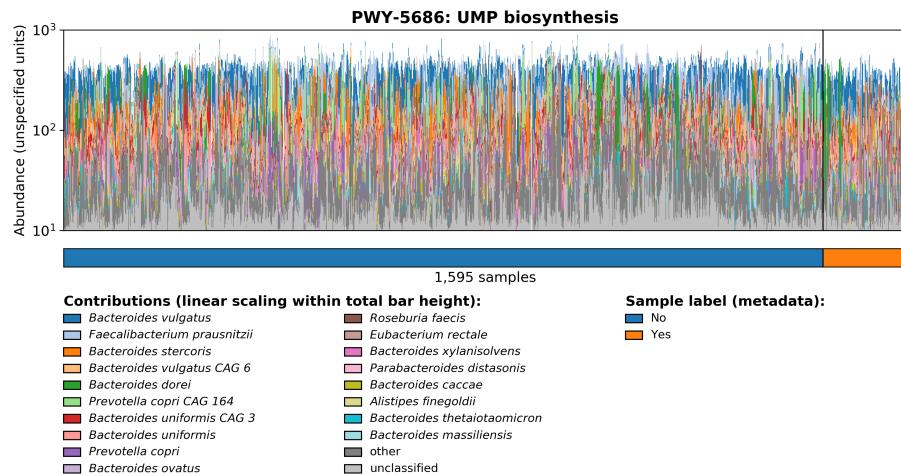


Figure 14: Significant association pathway #1 for metadata focus antibiotics

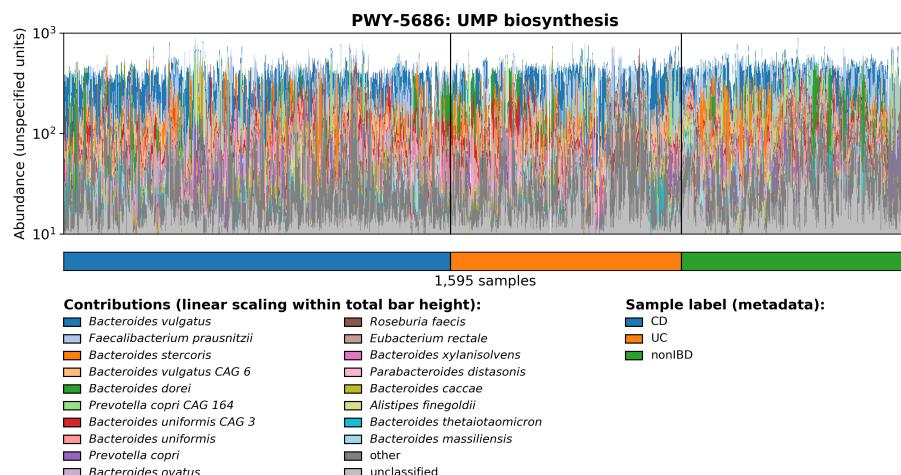


Figure 15: Significant association pathway #2 for metadata focus diagnosis

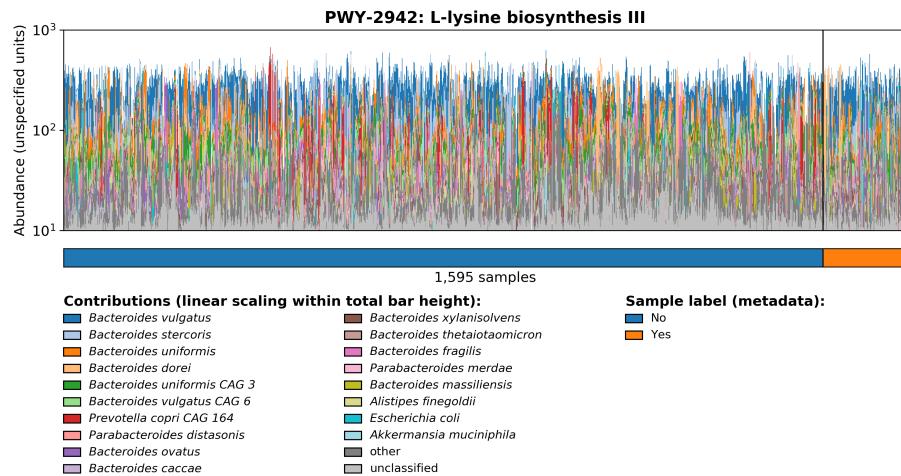


Figure 16: Significant association pathway #2 for metadata focus antibiotics

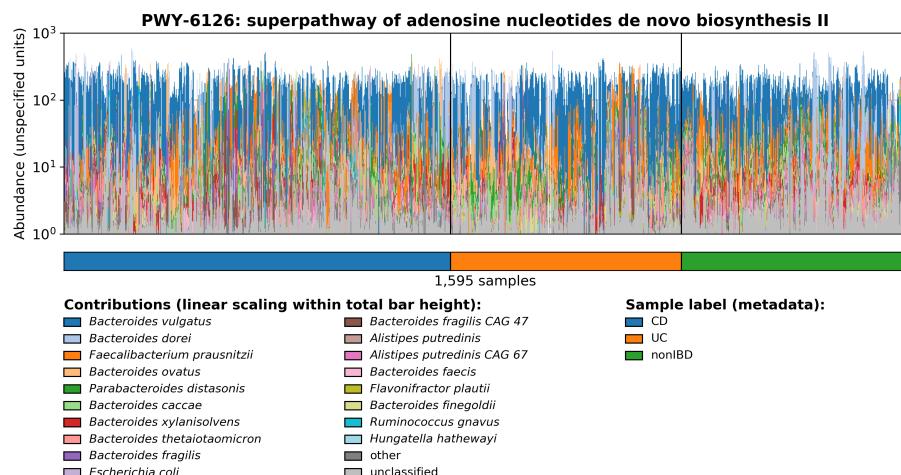


Figure 17: Significant association pathway #3 for metadata focus diagnosis

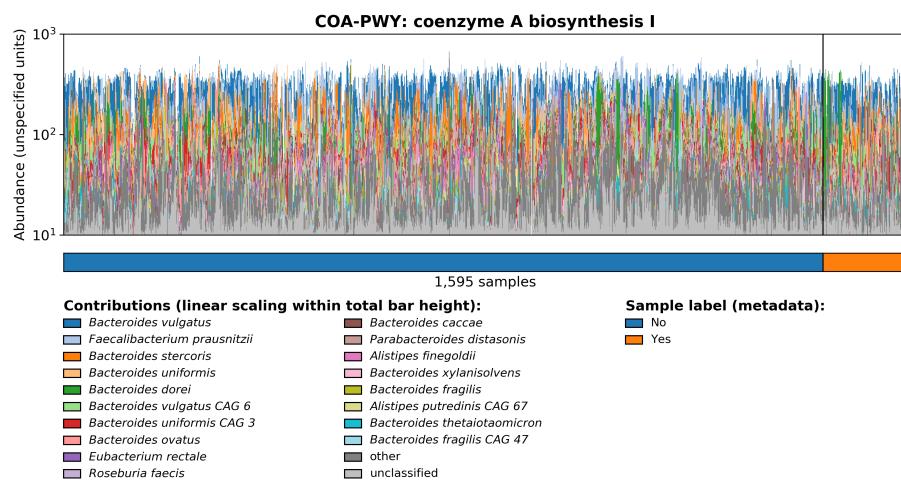


Figure 18: Significant association pathway #3 for metadata focus antibiotics

Each data type against all metadata variables: HAllA

HAllA (Hierarchical All-against-All Association Testing) discovers densely-associated blocks of features between two high-dimensional 'omics datasets. HAllA was run on each of the data sets provided. The heatmaps for each run type are shown. For more information from each HAllA run, check out the HAllA results folders for a complete set of output files.

HAllA taxonomy

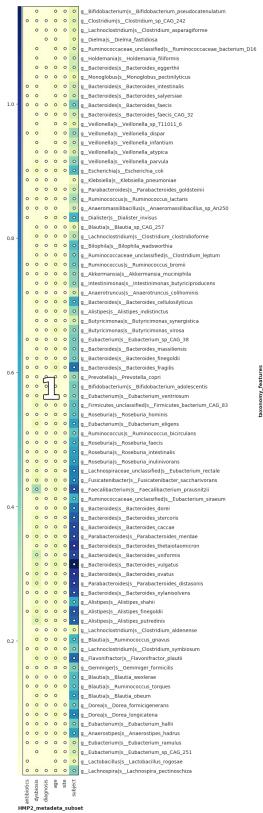


Figure 19: taxonomy heatmap

HAllA pathways

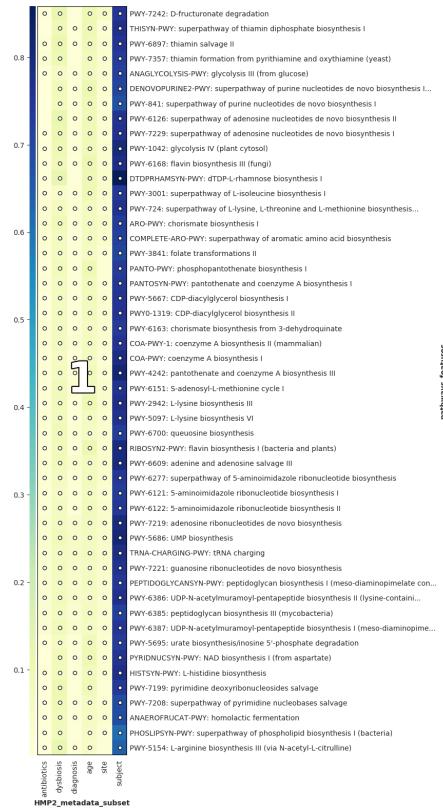


Figure 20: pathways heatmap