Statistics report for HMP2

Author: LM

Date: 05/25/2021

Contents

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 other data types : HAllA tests all possible associations of each feature in a data set against all features in a second data set. It is run to compare each of the data sets provided for the study against all others, eg taxonomy vs pathways.

All against all: Mantel test

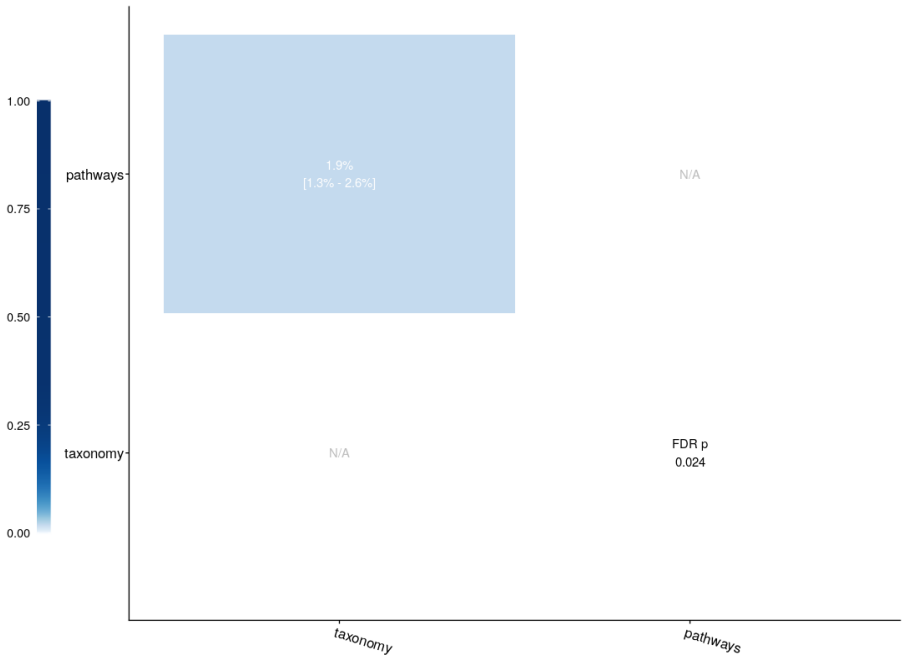

Figure 1: Mantel test using Bray-Curtis dissimilarity

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

Univariate

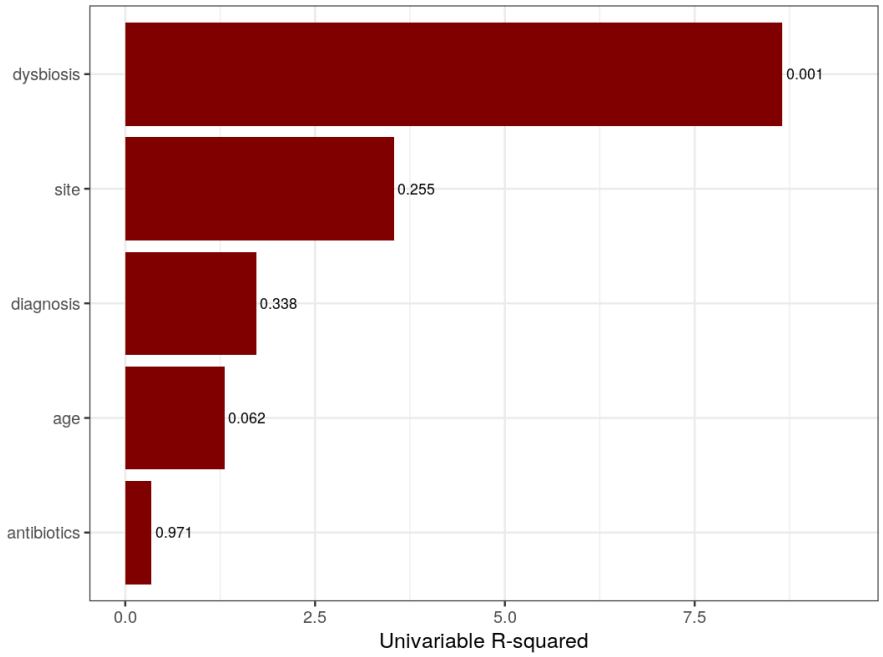

Figure 2: Taxonomy - Bar plot of R-squared value, annotated with the FDR adjusted p-value.

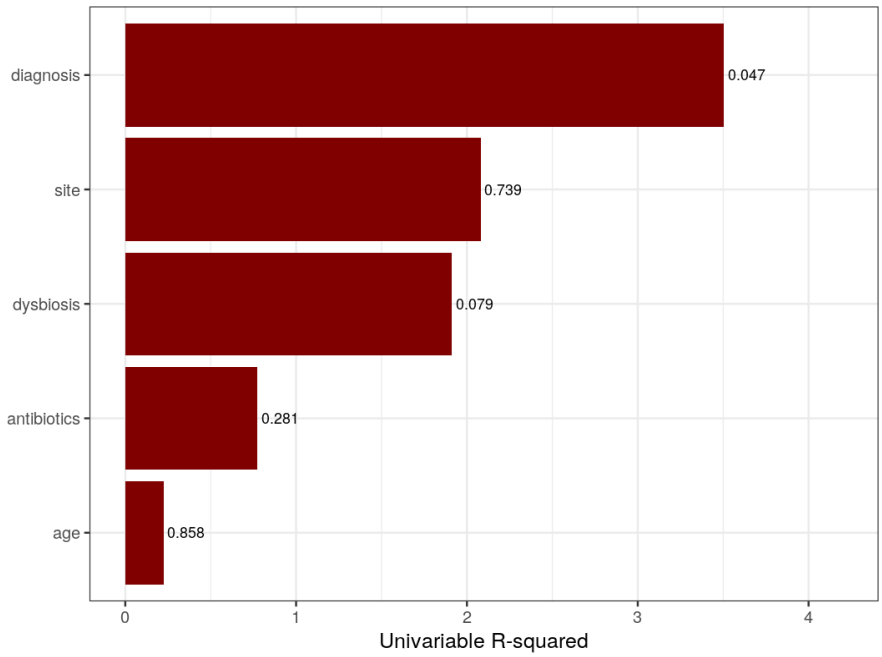

Figure 3: Pathways - Bar plot of R-squared value, annotated with the FDR adjusted p-value.

Pairwise

Each categorical variable is run against all other categorical variables in a set of pairs.

	Df	SumOfSqs	R2	F	Pr(>F)
age	1	0.445059249965635	0.0127700773106879	1.78641669151055	0.057
dysbiosis	1	3.00590134371938	0.0862482749219595	12.0653426119531	0.001
Residual	126	31.3910331011587	0.900702366246995	NA	NA
Total	128	34.851727138186	1	NA	NA

Figure 4: Taxonomy - Table of R-squared value and the FDR adjusted p-value.

	Df	SumOfSqs	R2	F	Pr(>F)
age	1	0.0105887434674865	0.00314361666215819	0.401900045951796	0.722
dysbiosis	1	0.0674692665228376	0.0200304701946973	2.56082333084409	0.066
Residual	125	3.29333859691712	0.977735849318378	NA	NA
Total	127	3.36833164009793	1	NA	NA

Figure 5: Pathways - Table of R-squared value and the FDR adjusted p-value.

Multivariate

For the multivariate model, applying settings so the order of the variables does not affect the results, the following covariate equation was provided: 'bray ~ + diagnosis + antibiotics + age' .

	Df	SumOfSqs	R2	F	Pr(>F)
diagnosis	2	0.607445667598441	0.017429427964644	1.11819104939256	0.289
antibiotics	1	0.108696068993893	0.00311881441522014	0.400177260078756	0.983
age	1	0.449795500015682	0.0129059744509148	1.65597461304826	0.068
Residual	124	33.6808557102675	0.966404206503857	NA	NA
Total	128	34.851727138186	1	NA	NA

Figure 6: Taxonomy - Table of R-squared value and the FDR adjusted p-value.

	Df	SumOfSqs	R2	F	Pr(>F)
diagnosis	2	0.113542500557436	0.0337088246316905	2.16767388909073	0.051
antibiotics	1	0.0199440176260848	0.00592103740280898	0.761514429211645	0.423
age	1	0.0086853748862552	0.00257853911499126	0.331630187206879	0.785
Residual	123	3.22136268717587	0.956367433903335	NA	NA
Total	127	3.36833164009793	1	NA	NA

Figure 7: Pathways - Table of R-squared value and 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.

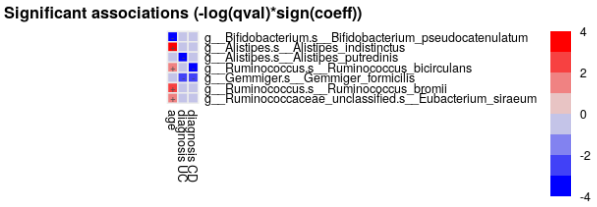

Figure 8: taxonomy heatmap

MaAsLin2 Plots

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

diagnosis

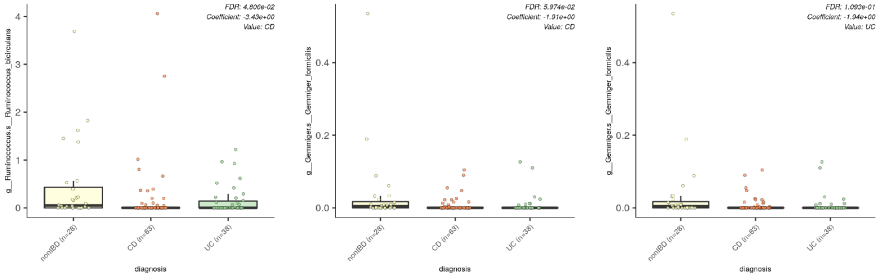

Figure 9: Top diagnosis associations for taxonomy

age

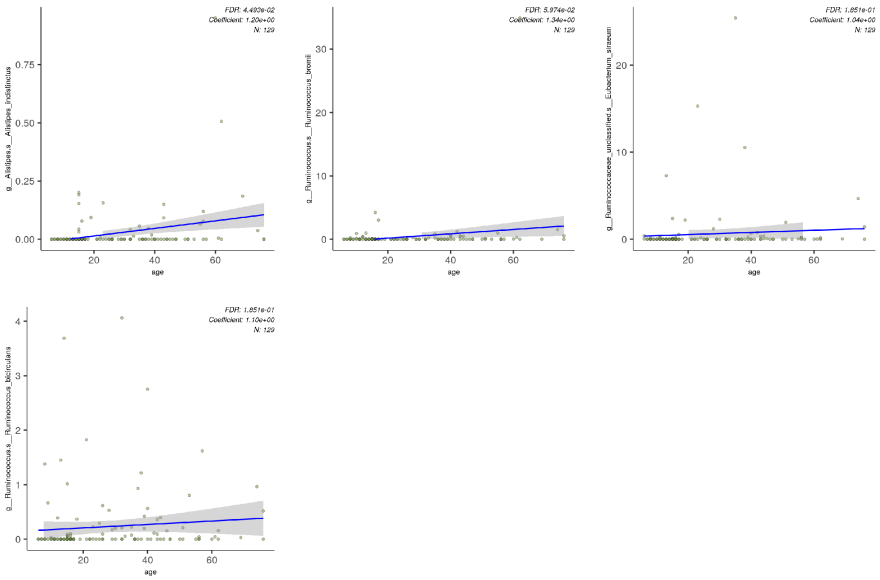

Figure 10: Top age associations for taxonomy

Pathways

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

Not enough significant associations for a heatmap.

MaAsLin2 Plots

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

No significant associations.

MaAsLiN2 stratified pathways plots

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

No significant associations for pathways with categorical metadata found.

Each data type against all other data types: 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 possible set of pairs from 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 vs. pathways

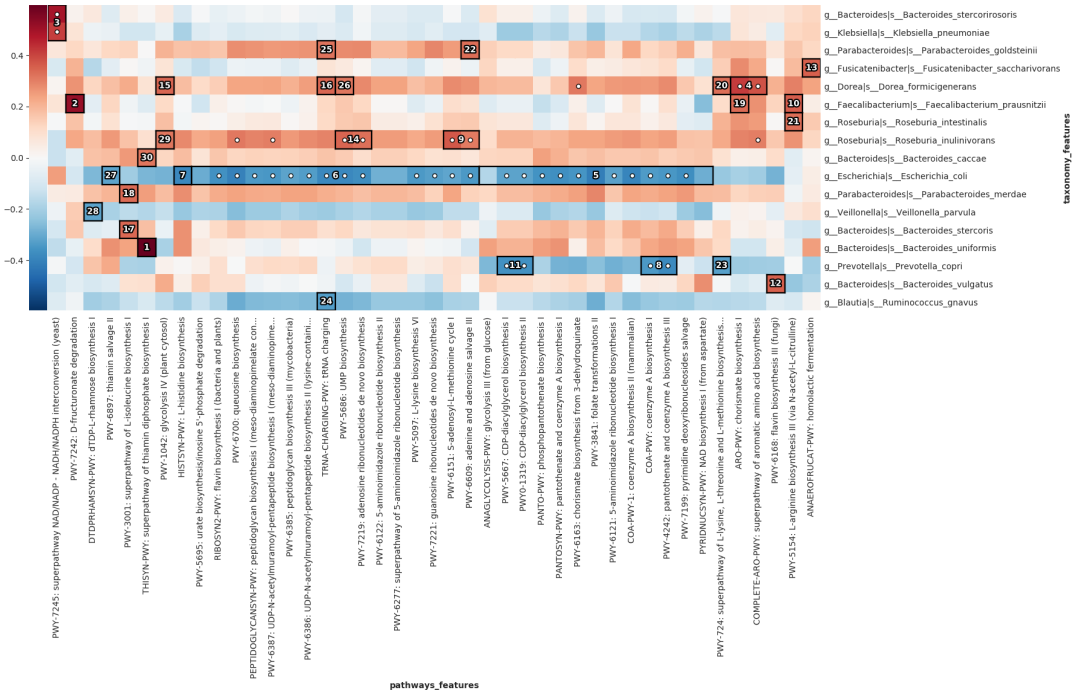

Figure 11: taxonomy pathways heatmap