

# Package ‘Maaslin2’

November 21, 2024

**Title** “Multivariable Association Discovery in Population-scale  
Meta-omics Studies”

**Year** 2021

**Version** 1.20.0

**Depends** R (>= 3.6)

**Description** 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. MaAsLin2 is the next generation of MaAsLin.

**License** MIT + file LICENSE

**LazyData** false

**Imports** robustbase, biglm, pcaPP, edgeR, metagenomeSeq, pbapply, car,  
dplyr, vegan, chemometrics, ggplot2, pheatmap, logging,  
data.table, lmerTest, hash, optparse, grDevices, stats, utils,  
glmmTMB, MASS, cplm, pscl, lme4, tibble

**Suggests** knitr, testthat (>= 2.1.0), rmarkdown, markdown

**VignetteBuilder** knitr

**Collate** fit.R utility\_scripts.R viz.R Maaslin2.R

**URL** <http://huttenhower.sph.harvard.edu/maaslin2>

**biocViews** Metagenomics, Software, Microbiome, Normalization

**BugReports** <https://github.com/biobakery/maaslin2/issues>

**git\_url** <https://git.bioconductor.org/packages/Maaslin2>

**git\_branch** RELEASE\_3\_20

**git\_last\_commit** 1cad4c

**git\_last\_commit\_date** 2024-10-29

**Repository** Bioconductor 3.20

**Date/Publication** 2024-11-20

**Author** Himel Mallick [aut],  
Ali Rahnavard [aut],  
Lauren McIver [aut, cre]

**Maintainer** Lauren McIver <[lauren.j.mciver@gmail.com](mailto:lauren.j.mciver@gmail.com)>

## Contents

Maaslin2 .....	2
<b>Index</b>	<b>4</b>

---

Maaslin2	<i>MaAsLin2 is the next generation of MaAsLin, a multivariable statistical framework for finding associations between clinical metadata and potentially high-dimensional microbial multi-omics data.</i>
----------	--

---

### Description

MaAsLin2 finds associations between microbiome meta-omics features and complex metadata in population-scale epidemiological studies. The software includes multiple analysis methods (including support for multiple covariates and repeated measures), filtering, normalization, and transform options to customize analysis for your specific study.

### Usage

```
Maaslin2(
  input_data,
  input_metadata,
  output,
  min_abundance = 0.0,
  min_prevalence = 0.1,
  min_variance = 0.0,
  normalization = "TSS",
  transform = "LOG",
  analysis_method = "LM",
  max_significance = 0.25,
  random_effects = NULL,
  fixed_effects = NULL,
  correction = "BH",
  standardize = TRUE,
  cores = 1,
  plot_heatmap = TRUE,
  heatmap_first_n = 50,
  plot_scatter = TRUE,
  max_pngs = 10,
  save_scatter = FALSE,
  save_models = FALSE,
  reference = NULL
)
```

### Arguments

<code>input_data</code>	The tab-delimited input file of features.
<code>input_metadata</code>	The tab-delimited input file of metadata.
<code>output</code>	The output folder to write results.
<code>min_abundance</code>	The minimum abundance for each feature.

min_prevalence	The minimum percent of samples for which a feature is detected at minimum abundance.
min_variance	Keep features with variance greater than.
max_significance	The q-value threshold for significance.
normalization	The normalization method to apply.
transform	The transform to apply.
analysis_method	The analysis method to apply.
random_effects	The random effects for the model, comma-delimited for multiple effects.
fixed_effects	The fixed effects for the model, comma-delimited for multiple effects.
correction	The correction method for computing the q-value.
standardize	Apply z-score so continuous metadata are on the same scale.
plot_heatmap	Generate a heatmap for the significant associations.
heatmap_first_n	In heatmap, plot top N features with significant associations.
plot_scatter	Generate scatter plots for the significant associations.
max_pngs	Set the maximum number of scatter plots for significant associations to save as png files.
save_scatter	Save all scatter plot ggplot objects to an RData file.
cores	The number of R processes to run in parallel.
save_models	Return the full model outputs and save to an RData file.
reference	The factor to use as a reference for a variable with more than two levels provided as a string of 'variable,reference' semi-colon delimited for multiple variables.

### Value

List containing the results from applying the model.

### Author(s)

Himel Mallick<himel.stat.iitk@gmail.com>  
 Ali Rahnavard<gholamali.rahnavaard@gmail.com>  
 Maintainers: Lauren McIver<lauren.j.mciver@gmail.com>

### Examples

```
input_data <- system.file(
  'extdata', 'HMP2_taxonomy.tsv', package="Maaslin2")
input_metadata <- system.file(
  'extdata', 'HMP2_metadata.tsv', package="Maaslin2")
fit_data <- Maaslin2(
  input_data, input_metadata, 'demo_output', transform = "AST",
  fixed_effects = c('diagnosis', 'dysbiosisnonIBD', 'dysbiosisUC', 'dysbiosisCD', 'antibiotics', 'age'),
  random_effects = c('site', 'subject'),
  normalization = 'NONE',
  reference = 'diagnosis,nonIBD',
  standardize = FALSE)
```

# Index

Maaslin2, [2](#)