Package 'MRS'

April 7, 2022

Type Package

Title Microbial Risk Score Framework

Version 0.1.0

Date 2022-04-07

Author Chan Wang (Chan.Wang@nyulangone.org) and Huilin Li (Huilin.Li@nyulangone.org)

Maintainer Chan Wang (Chan.Wang@nyulangone.org) and Huilin Li (Huilin.Li@nyulangone.org)

Description The microbial risk score (MRS) framework converts the high-dimensional microbiome data into a summarized risk score MRS that can be used to measure and predict disease susceptibility. We proposed to employ the existing sophisticated microbial association tests, such as ANCOMBC, ALDEx2, and Maaslin2 to identify microbial taxa associated with disease using the discovery samples. We proposed to a community-based MRS, which is defined the alpha diversity of the sub-community consisting of only the identified candidate taxa. Pruning and thresholding (P+T) method and AUC evaluation is used to determine the identified taxa based on the discovery samples.

License GPL (>= 3)

Encoding UTF-8

LazyData true

RoxygenNote 7.1.2

Depends R (>= 4.1.0)

Imports ALDEx2,
ANCOMBC,
magrittr,
Maaslin2,
phyloseq,
pROC

R topics documented:

GMHI *GMHI multi-study cohort*

Description

A datalist has an integrative discovery dataset and an independent validation dataset from the GMHI multi-study cohort.

Usage

```
GMHI
```

Format

A data list with two datasets:

discovery.data a phyloseq-class object, which consists of a feature table and a sample metadata. This dataset is extracted from the integrative discovery dataset in the GMHI multi-study cohort. Specifically, there are 4347 subjects and 313 species. There are two group statues in the sample data. Group variable indicates healthy or nonhealthy status for each subject. Group1 indicates healthy or disease condition in details.

validation.data a phyloseq-class object, which consists of a feature table and a sample metadata. This dataset is extracted from the independent validation dataset in the GMHI multi-study cohort. Specifically, there are 782 subjects and 576 species. There are two group statues in the sample data. Group variable indicates health or nonhealthy status for each subject. Group1 indicates healthy or disease condition in details.

MRS *Microbial risk score framework*

Description

`MRS` Construction and evaluation of MRS.

Usage

```
MRS(discovery.data, validation.data, GroupID, DA.method, measurement)
```

Arguments

`discovery.data`	A phyloseq-class object, which consists of a feature table (observed count table), a sample metadata, a taxonomy table (optional), and a phylogenetic tree (optional). See `phyloseq` for more details.
`validation.data`	A phyloseq-class object, which consists of a feature table (observed count table or relative abundance table), a sample metadata, a taxonomy table (optional), and a phylogenetic tree (optional). See `phyloseq` for more details.
`GroupID`	The name of the group variable of interest in metadata.

`DA.method` The microbial differential abundance method which is employed to identify the candidate taxa for MRS construction based on the discovery dataset (the first step of the MRS framework). Three top-performing methods ANCOMBC, ALDEx2, and Maaslin2 are available: c("ancombc", "ALDEx2","Maaslin2").

`measurement` The diverity index which is employed for MRS calculation. Three widely used indices are available: c("Shannon","Simpson","Observed").

Details

This function identifys the candidate taxa used for constructing MRS and calculates MRS in the discovery cohort, and evaluates MRS in both discovery and validation cohorts in terms of AUC. If there is no validation cohort, cross-validation can be performed at the pre-processing step to have both cohorts.

Reference: Wang C, Segal L, Hu J, Zhou B, Hayes R, Ahn J, and Li H (2022). Microbial risk score: capturing microbial characteristics, integrating multi-omics profiling, and predicting disease risk. *Microbiome*.

Value

A list with components:

- Cutoff: An optimal p-value cutoff (a numeric value) for identifying the taxa used for construction of MRS.
- Taxa used for MRS: The specific taxa used for construction of MRS.
- AUC: A numeric matrix which reports the AUC values and 95 the discovery and validation cohorts, respectively.

Examples

```
require(phyloseq)

## Evaluation of MRS in terms of comparison between Healthy and Nonhealthy ##
## using ANCOMBC method and Shannon index
 discovery=GMHI[[1]];

 validation=GMHI[[2]];

 res=MRS(discovery, validation, GroupID="Group", DA.method="ancombc", measurement="Shannon")

 AUC=res[[3]]

 ## using ALDEx2 method and Shannon index

 res=MRS(discovery, validation, GroupID="Group", DA.method="ALDEx2", measurement="Shannon")

 AUC=res[[3]]

## Evaluation of MRS in terms of comparison between Healthy and a specific disease ##
## Healthy vs. CA

 discovery.sub=prune_samples(sample_data(discovery)$Group1 %in% c("Healthy","CA"),discovery)
 validation.sub=prune_samples(sample_data(validation)$Group1 %in% c("Healthy","CA"),validation)
```

```
res=MRS(discovery.sub, validation.sub, GroupID="Group", DA.method="ALDEx2", measurement="Shannon")
AUC=res[[3]]
```

Index