

**Clase para**

**“Introducción al estudio de los  
procesos evolutivos con énfasis en  
métodos de secuenciación masiva  
2024-1”**

**Dr. Andrés Parada  
Departamento Ecología y Evolución,  
Fcien, Udelar.**

**[andrespara@gmail.com](mailto:andrespara@gmail.com)**

- 
- timing de entrada a Sudamérica
  - sistemática
  - patrones macroevolutivos, e.g. factores influyendo en la tasa de diversificación

© Jorge Brito





SPRINGER BRIEFS IN ECOLOGY

Muhammad Saleem

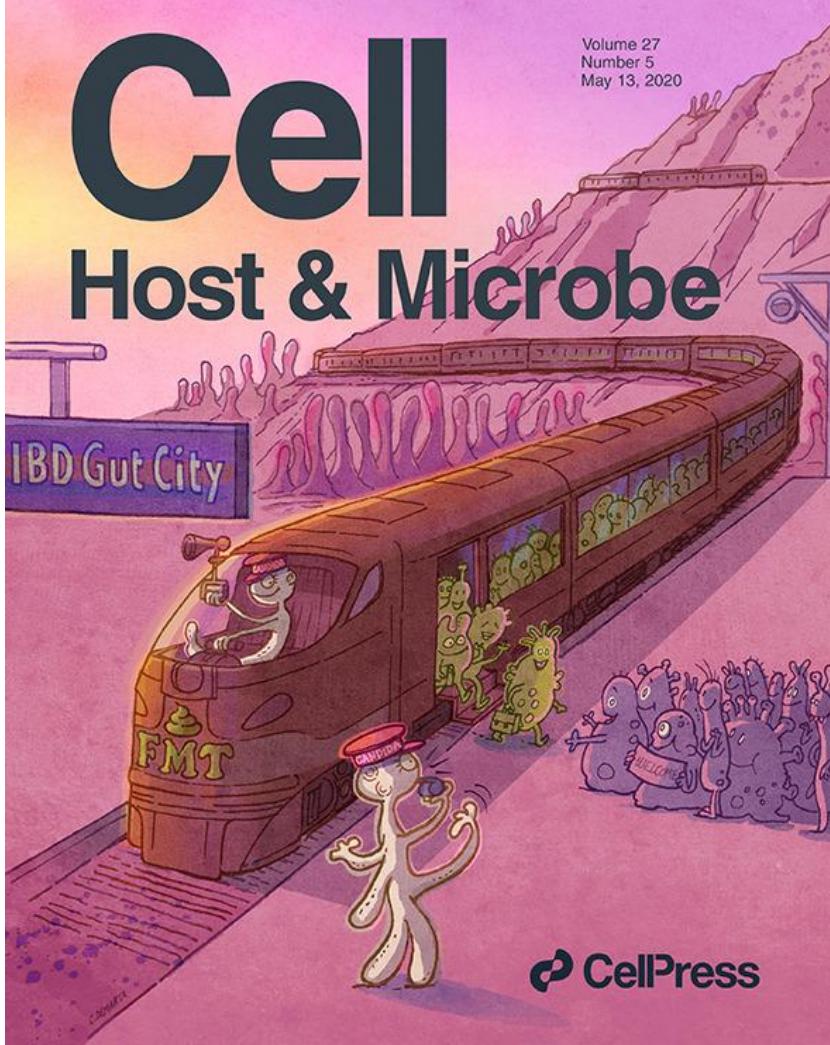
**Microbiome  
Community  
Ecology**  
Fundamentals and  
Applications

 Springer

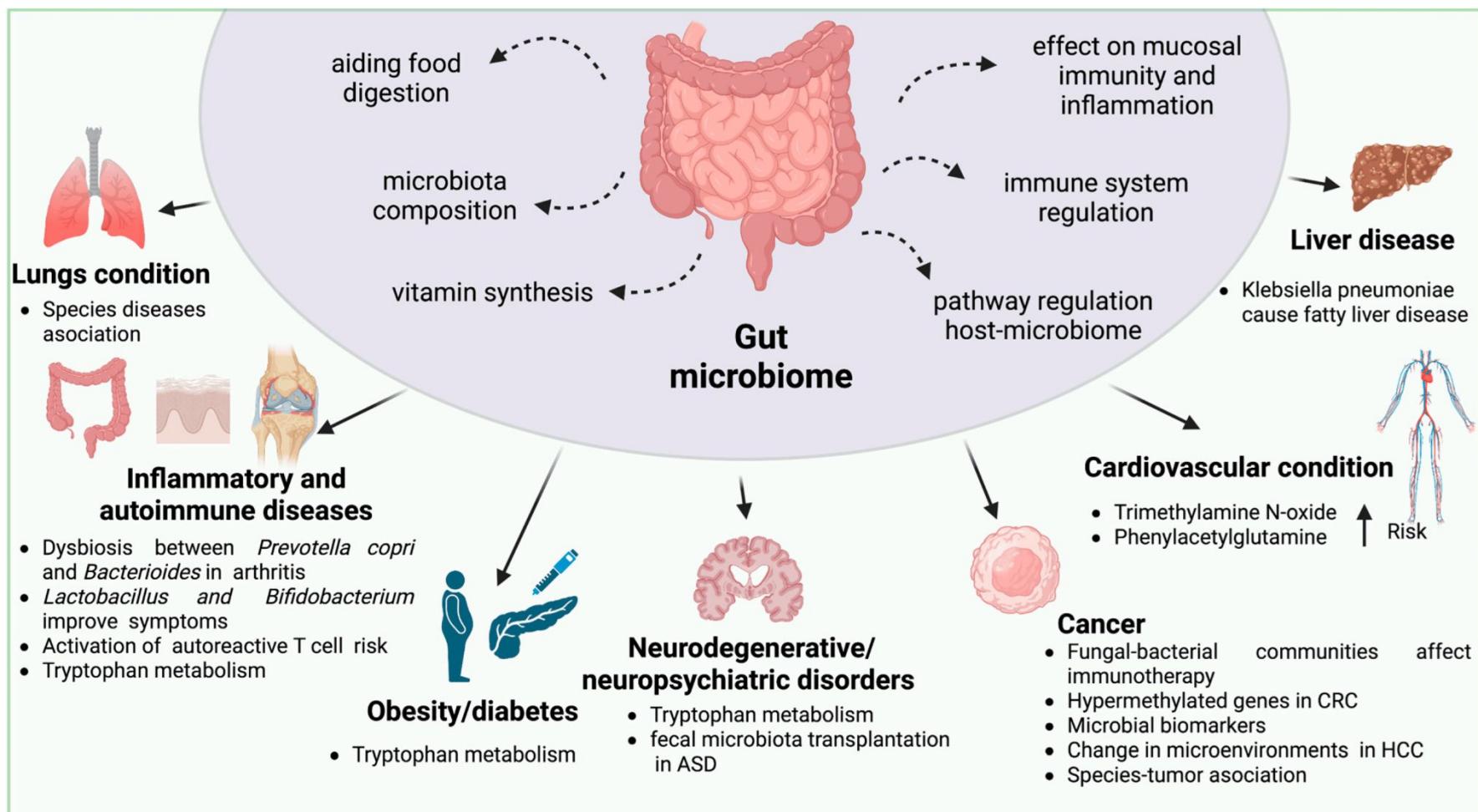
# Cell

## Host & Microbe

Volume 27  
Number 5  
May 13, 2020



CellPress



# La 'pipeline usual'



# La 'pipeline usual'



Article | [Open access](#) | Published: 23 February 2023

## Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4

Aitor Blanco-Miguel, Francesco Beghini, Fabio Cumbo, Lauren J. McIver, Kelsey N. Thompson, Moreno Zolfo, Paolo Manghi, Leonard Dubois, Kun D. Huang, Andrew Maltez Thomas, William A. Nickols, Gianmarco Piccinno, Elisa Piperni, Michal Punčochář, Mireia Valles-Colomer, Adrian Tett, Francesca Giordano, Richard Davies, Jonathan Wolf, Sarah E. Berry, Tim D. Spector, Eric A. Franzosa, Edoardo Pasolli, Francesco Asnicar, ... Nicola Segata + Show authors

[Nature Biotechnology](#) (2023) | [Cite this article](#)



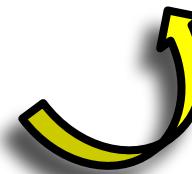
**bash**

**R**

# La 'pipeline usual'



Perfil taxonómico ~ análogo a conocer dos comunidades de vertebrados en dos sitios



Article | [Open access](#) | Published: 23 February 2023

**Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4**

Aitor Blanco-Miguez, Francesco Beghini, Fabio Cumbo, Lauren J. McIver, Kelsey N. Thompson, Moreno Zolfo, Paolo Manghi, Leonard Dubois, Kun D. Huang, Andrew Maltez Thomas, William A. Nickols, Gianmarco Piccinno, Elisa Piperni, Michal Punčochář, Mireia Valles-Colomer, Adrian Tett, Francesca Giordano, Richard Davies, Jonathan Wolf, Sarah E. Berry, Tim D. Spector, Eric A. Franzosa, Edoardo Pasolli, Francesco Asnicar, ... Nicola Segata + Show authors

[Nature Biotechnology \(2023\)](#) | [Cite this article](#)

**Segata Lab**  
Computational Metagenomics

bioBakery tools for meta'omic profiling

bioBakery is a suite of software, tutorials, and workflows of methods developed by the Huttenhower lab and the Segata lab for performing analyses on microbial communities from metagenomic data.

The following tools developed by the Segata Lab are included in bioBakery:

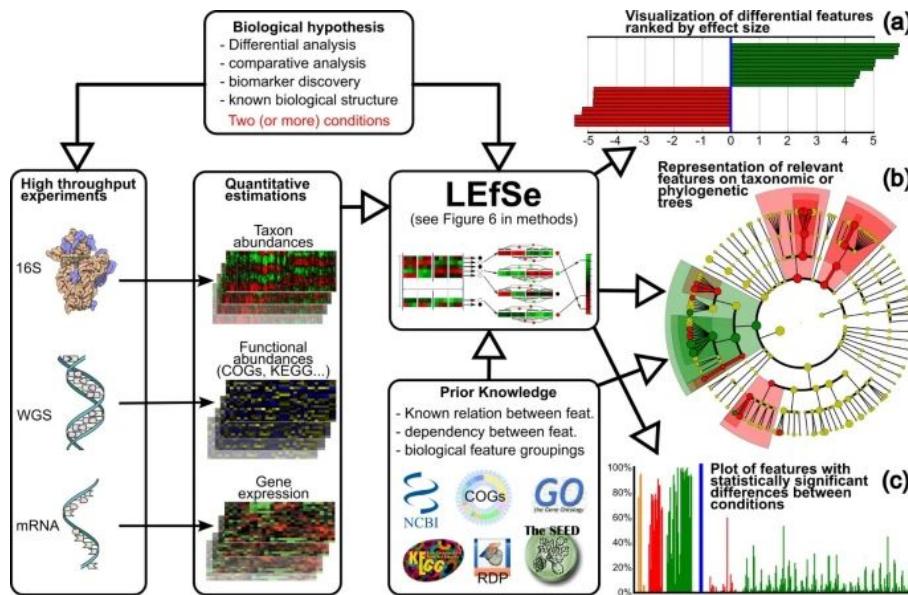
- MetaPhlAn
- StrainPhlAn
- PanPhlAn
- PhyloPhlAn
- GraPhlAn
- LEfSe

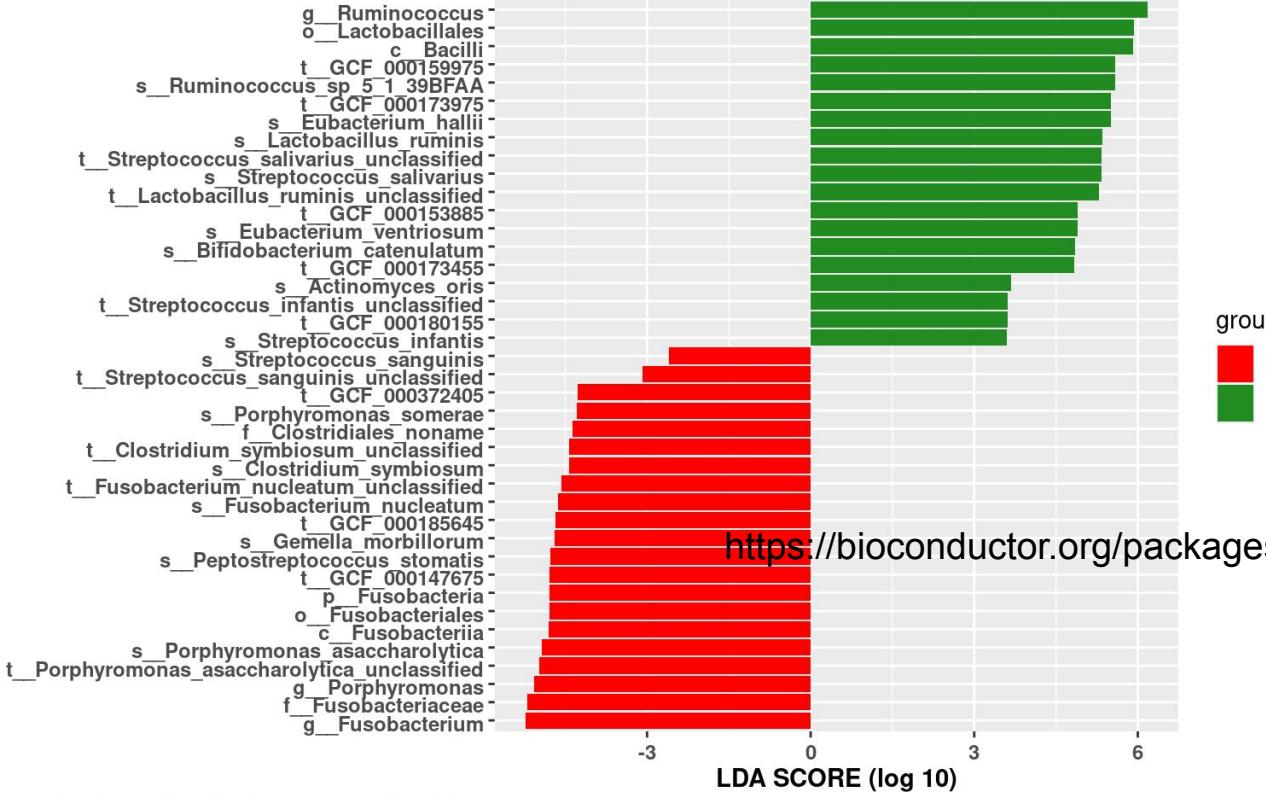
CIBIO  
UNIVERSITY OF TRENTO

## Metagenomic biomarker discovery and explanation

Nicola Segata, Jacques Izard, Levi Waldron, Dirk Gevers, Larisa Miropolsky, Wendy S Garrett & Curtis Huttenhower 

*Genome Biology* 12, Article number: R60 (2011) | [Cite this article](#)





Lefser finds features that have greatest differences between classes.

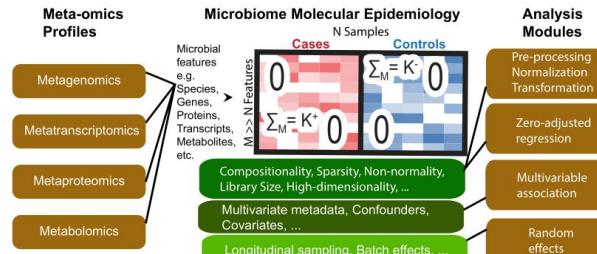
# MaAsLin2 User Manual

MaAsLin2 is the next generation of MaAsLin (Microbiome Multivariable Association with Linear Models).

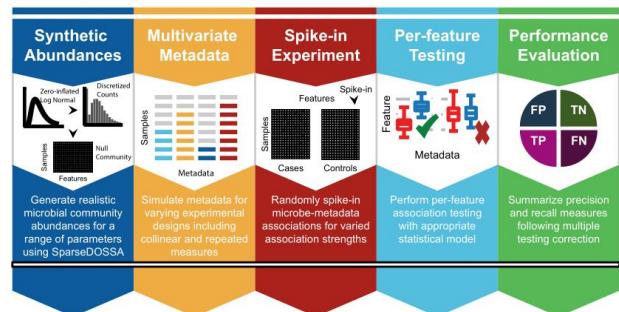
MaAsLin2 is comprehensive R package for efficiently determining multivariable association between clinical metadata and microbial meta-omics features. MaAsLin2 relies on general linear models to accommodate most modern epidemiological study designs, including cross-sectional and longitudinal, along with a variety of filtering, normalization, and transform methods.

Modelos para identificar features (bacterias en este caso) que permiten distinguir grupos (ejemplo condición sana de otra)

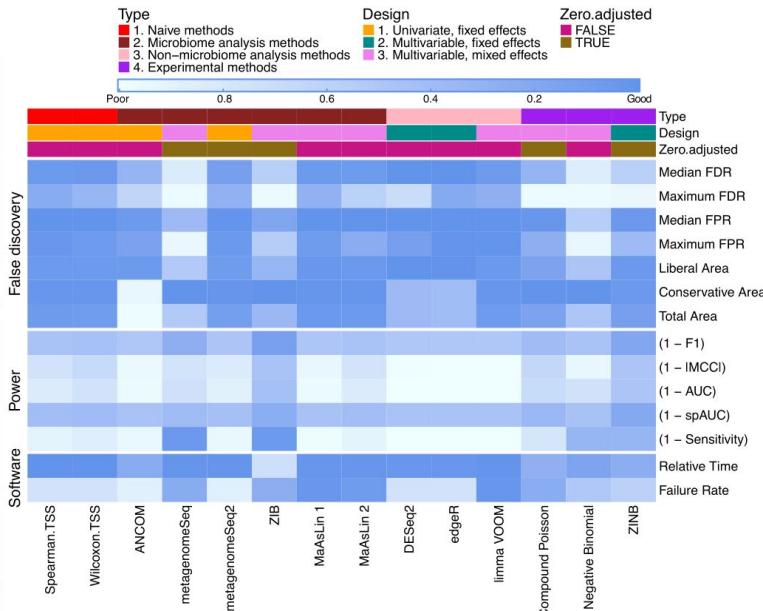
## A. MaAsLin 2 - A framework for microbiome epidemiology



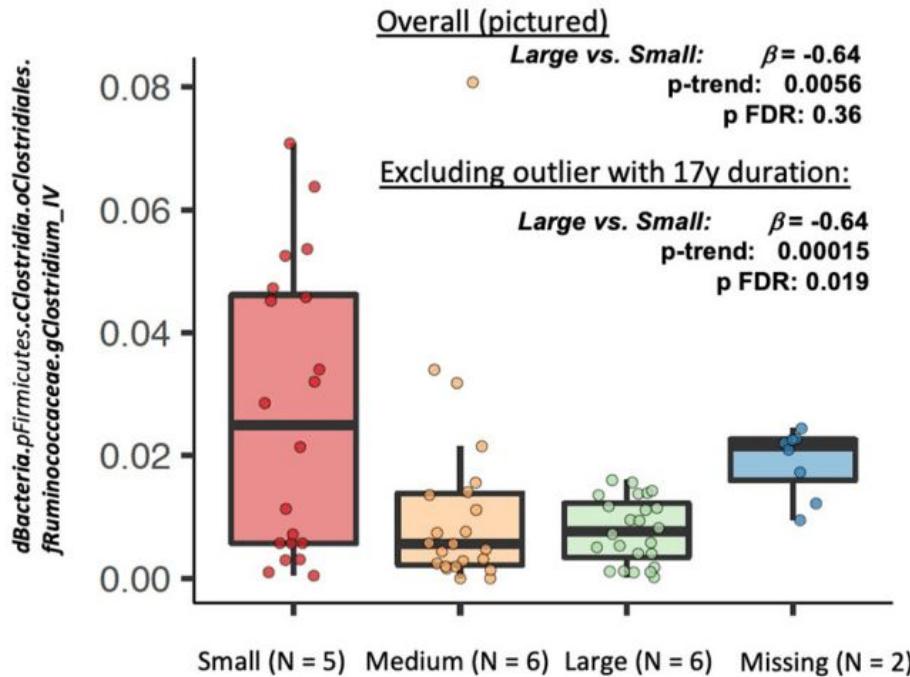
## B. Comprehensive benchmarking using synthetic abundances



## C. Large-scale evaluation of multivariable association methods



## Improvement in UPDRS score with Levodopa treatment.



abundancia relativa entre distintos grupos + significancia de diferencias

## Effect of Levodopa Initiation on the Gut Microbiota in Parkinson's Disease

Article

Full-text available

Mar 2021

Natalia Palacios · Anas Hannoun · Julie M. Flahive · [...] · Kara M. Smith

# data ya disponible accesible desde <https://waldronlab.io/curatedMetagenomicData/>

## curatedMetagenomicData

The `curatedMetagenomicData` package provides standardized, curated human microbiome data for novel analyses. It includes gene families, marker abundance, marker presence, pathway abundance, pathway coverage, and relative abundance for samples collected from different body sites. The bacterial, fungal, and archaeal taxonomic abundances for each sample were calculated with `MetaPhiAn3`, and metabolic functional potential was calculated with `HUMAN3`. The manually curated sample metadata and standardized metagenomic data are available as (Tree)SummarizedExperiment objects.

### Accessible, curated metagenomic data through ExperimentHub

[Edoardo Pasolli](#), [Lucas Schiffer](#), [Paolo Manghi](#), [Audrey Renson](#), [Valerie Obenchain](#), [Duy Tin Truong](#),  
[Francesco Beghini](#), [Faizan Malik](#), [Marcel Ramos](#), [Jennifer B Dowd](#), [Curtis Huttenhower](#), [Martin Morgan](#),  
[Nicola Segata](#) & [Levi Waldron](#)

[Nature Methods](#) **14**, 1023–1024 (2017) | [Cite this article](#)

8297 Accesses | 180 Citations | 40 Altmetric | [Metrics](#)

# con data ya disponible comparando distintos datasets

Performing meta-analyses of microbiome studies with MMUPHin

Siyuan Ma<sup>1,2\*</sup>

<sup>1</sup>Harvard T.H. Chan School of Public Health

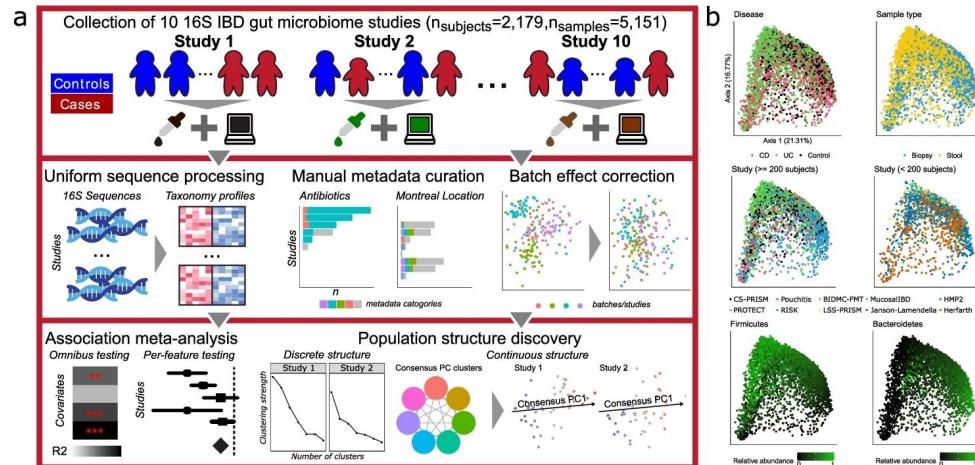
<sup>2</sup>Broad Institute

\*siyuanma@g.harvard.edu

10/09/2019

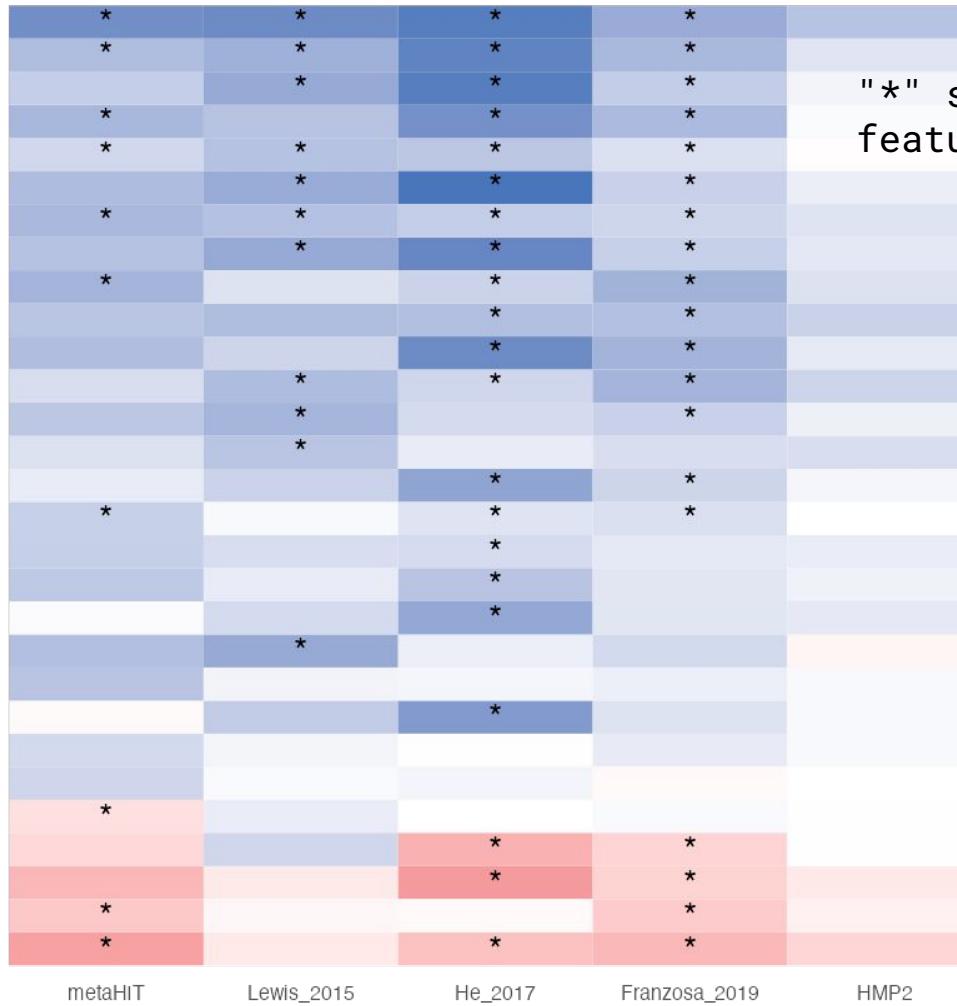
Package

MMUPHin 1.14.0



<https://www.bioconductor.org/packages/release/bioc/vignettes/MMUPHin/inst/doc/MMUPHin.html>

1263 Ruminococcus  
 NA Ruminococcaceae gen. incertae sedis  
 1730 Eubacterium  
 459786 Oscillibacter  
 1485 Clostridium  
 841 Roseburia  
 1501226 Romboutsia  
 NA Lachnospiraceae gen. incertae sedis  
 447020 Adlercreutzia  
 239759 Alistipes  
 216851 Faecalibacterium  
 397864 Barnesiella  
 283168 Odoribacter  
 1924093 Anaeromassilicibacillus  
 580596 Butyricicoccus  
 990721 Christensenella  
 244127 Anaerotruncus  
 61170 Holdemania  
 207244 Anaerostipes  
 838 Prevotella  
 1348911 Coprobacter  
 189330 Dorea  
 1017280 Pseudoflavonifractor  
 1935927 Massilioclostridium  
 816 Bacteroides  
 1505663 Erysipelatoclostridium  
 848 Fusobacterium  
 946234 Flavonifractor  
 1506553 Lachnoclostridium



"\*" significancia de este feature en cada dataset

Generalized fold change



# SIAMCAT

build canceled



## Overview

SIAMCAT is a pipeline for Statistical Inference of Associations between Microbial Communities And host phenoTypes. A primary goal of analyzing microbiome data is to determine changes in community composition that are associated with environmental factors. In particular, linking human microbiome composition to host phenotypes such as diseases has become an area of intense research. For this, robust statistical modeling and biomarker extraction toolkits are crucially needed. SIAMCAT provides a full pipeline supporting data preprocessing, statistical association testing, statistical modeling (LASSO logistic regression) including tools for evaluation and interpretation of these models (such as cross validation, parameter selection, ROC analysis and diagnostic model plots).

SIAMCAT is developed in the [Zeller group](#) and is part of the suite of computational microbiome analysis tools hosted at [EMBL](#).

## Starting with SIAMCAT

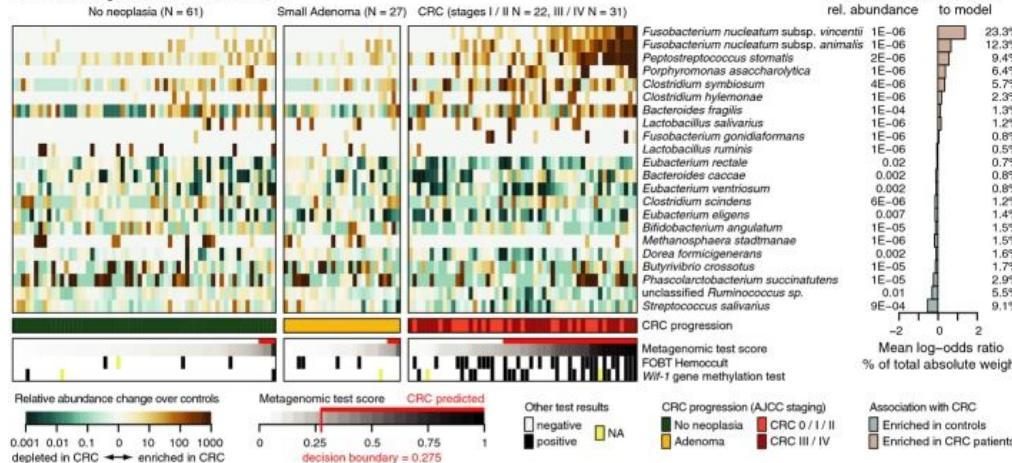


# Potential of fecal microbiota for early-stage detection of colorectal cancer.

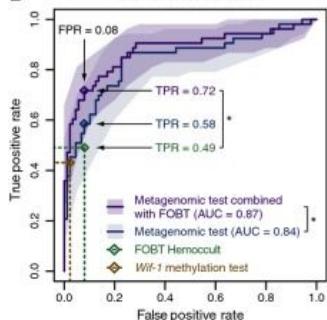
paper que ilustra utilidad de SIAMCAT

Zeller G<sup>1</sup> ●, Tap J<sup>2</sup> ●, Voigt AY<sup>3</sup> ●, Sunagawa S<sup>1</sup> ●, Kultima JR<sup>1</sup>, Costea PI<sup>1</sup> ●, Amiot A<sup>4</sup>, Böhm J<sup>5</sup>, Brunetti F<sup>6</sup>, Habermann

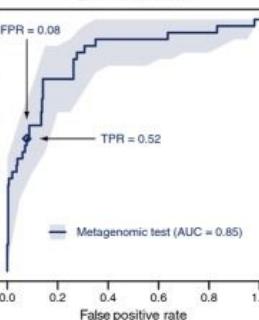
## A Microbial signature for CRC classification



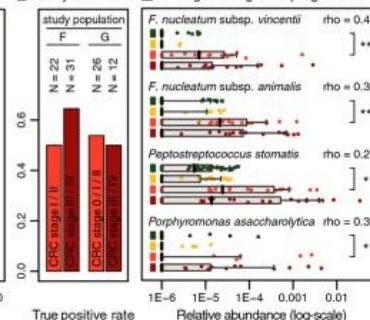
## B Internal cross validation



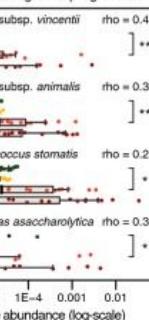
## C External validation

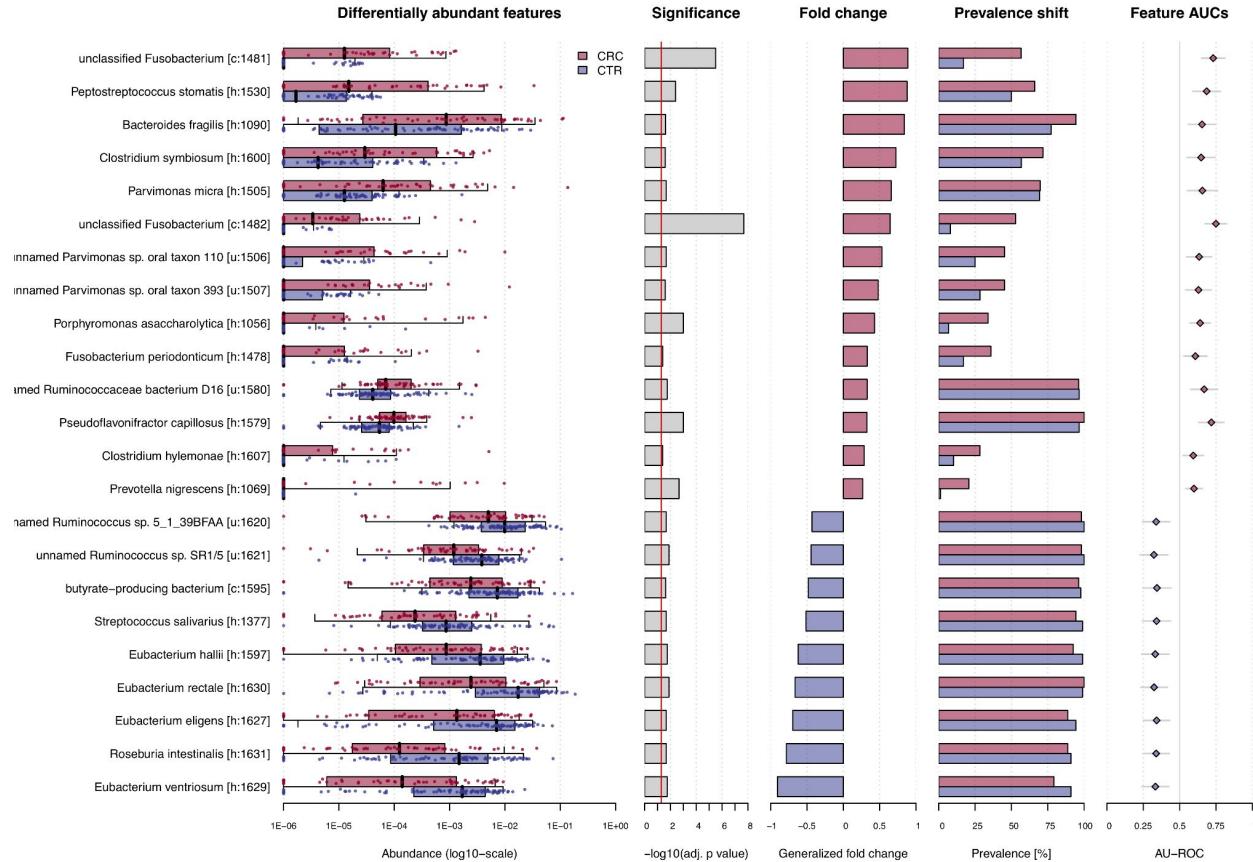


## D Early detection

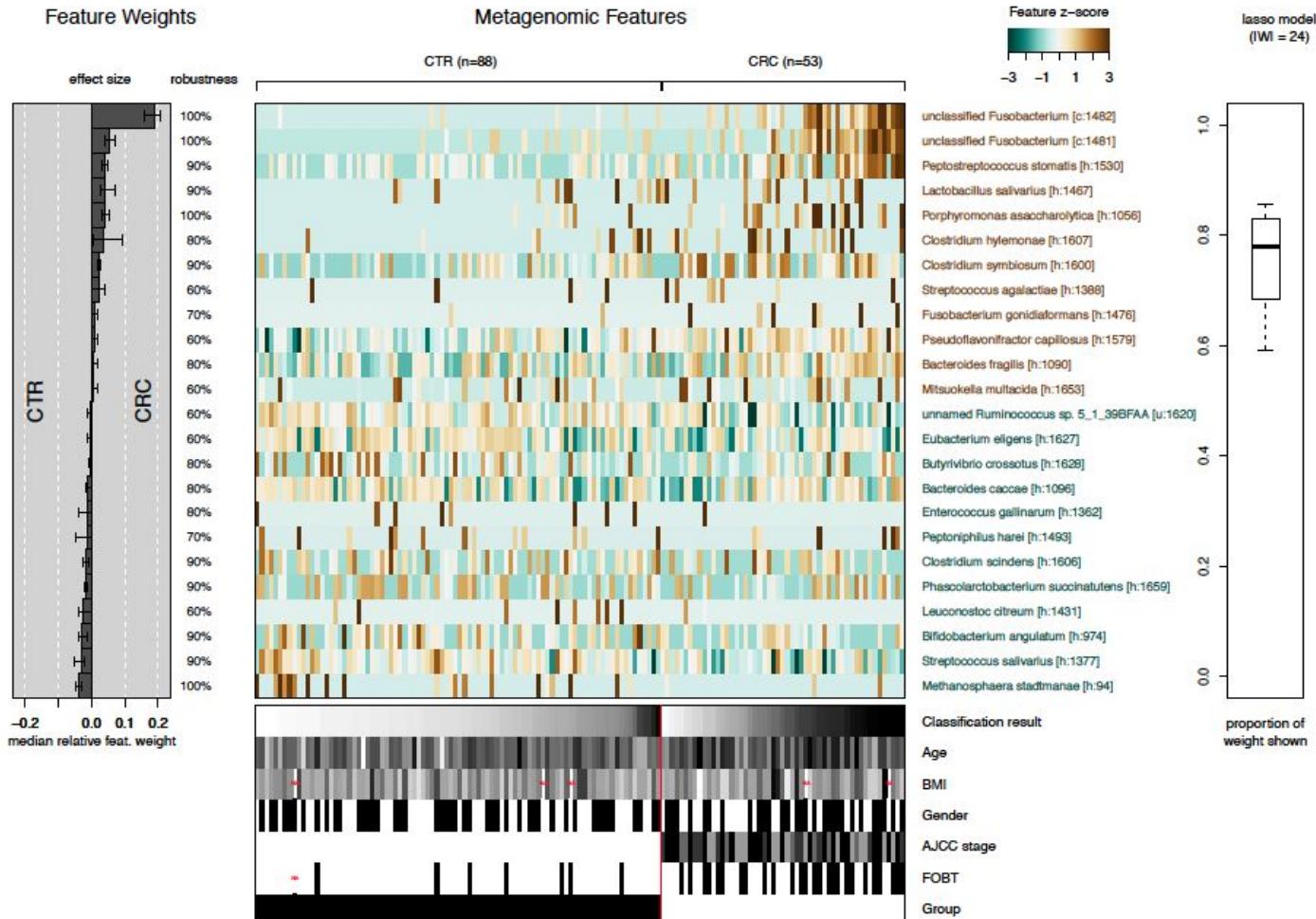


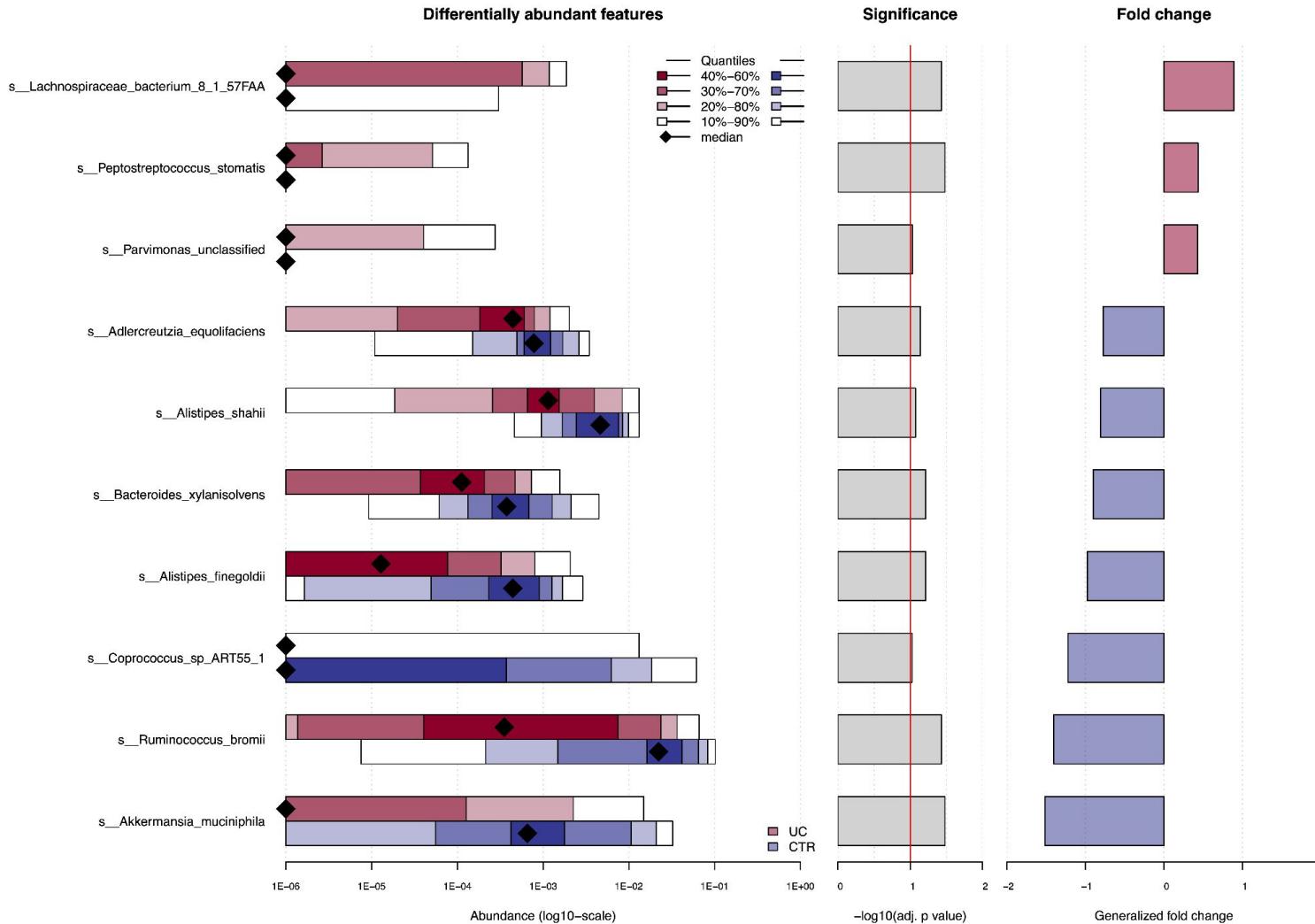
## E Changes along CRC progression

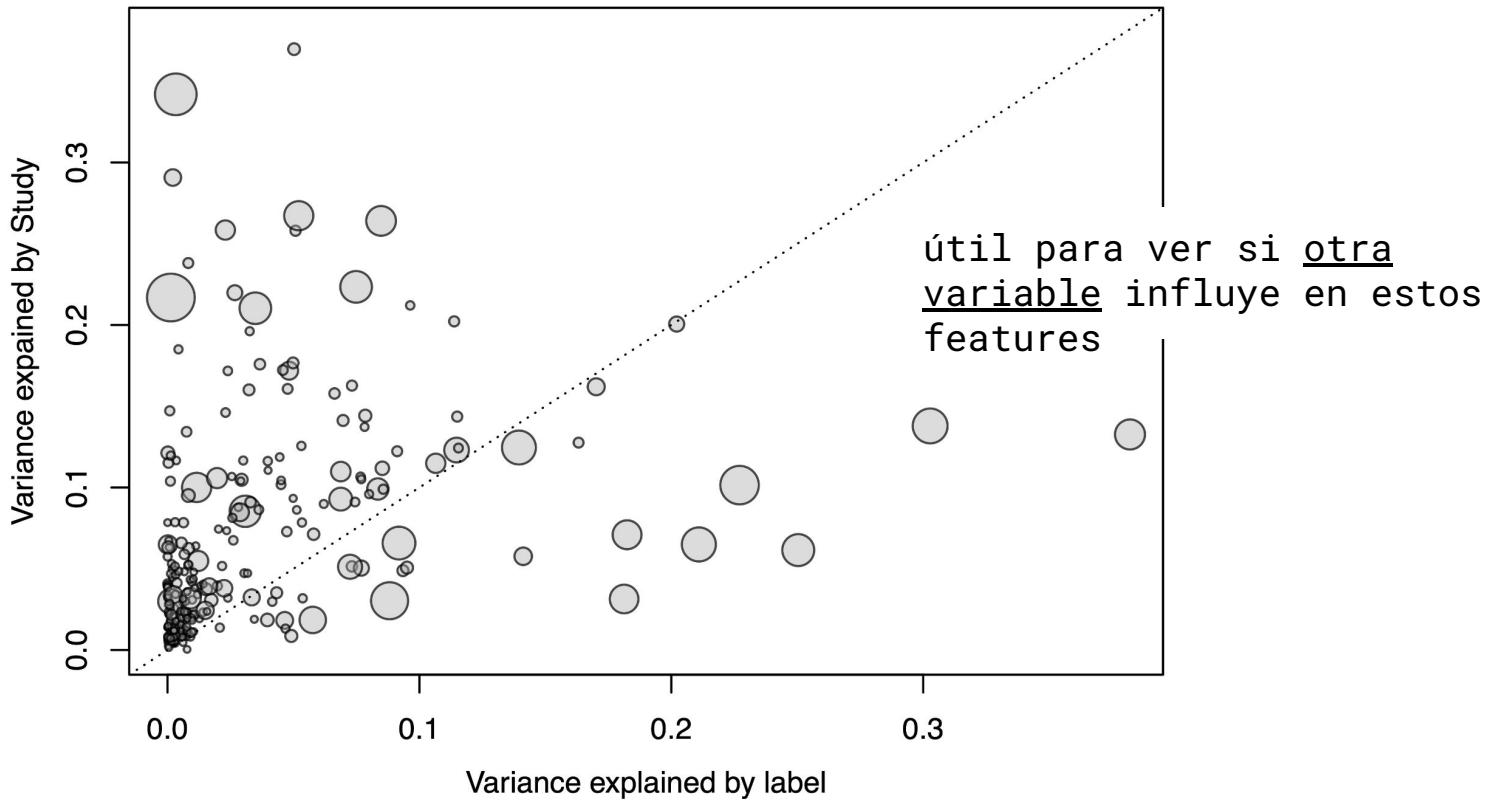




abundancia relativa entre distintos grupos y la evidencia a favor







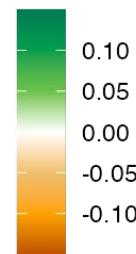
hay una gran utilidad de tratar de reconocer patrones en nuestro dataset entrenando modelos con otros datasets similares





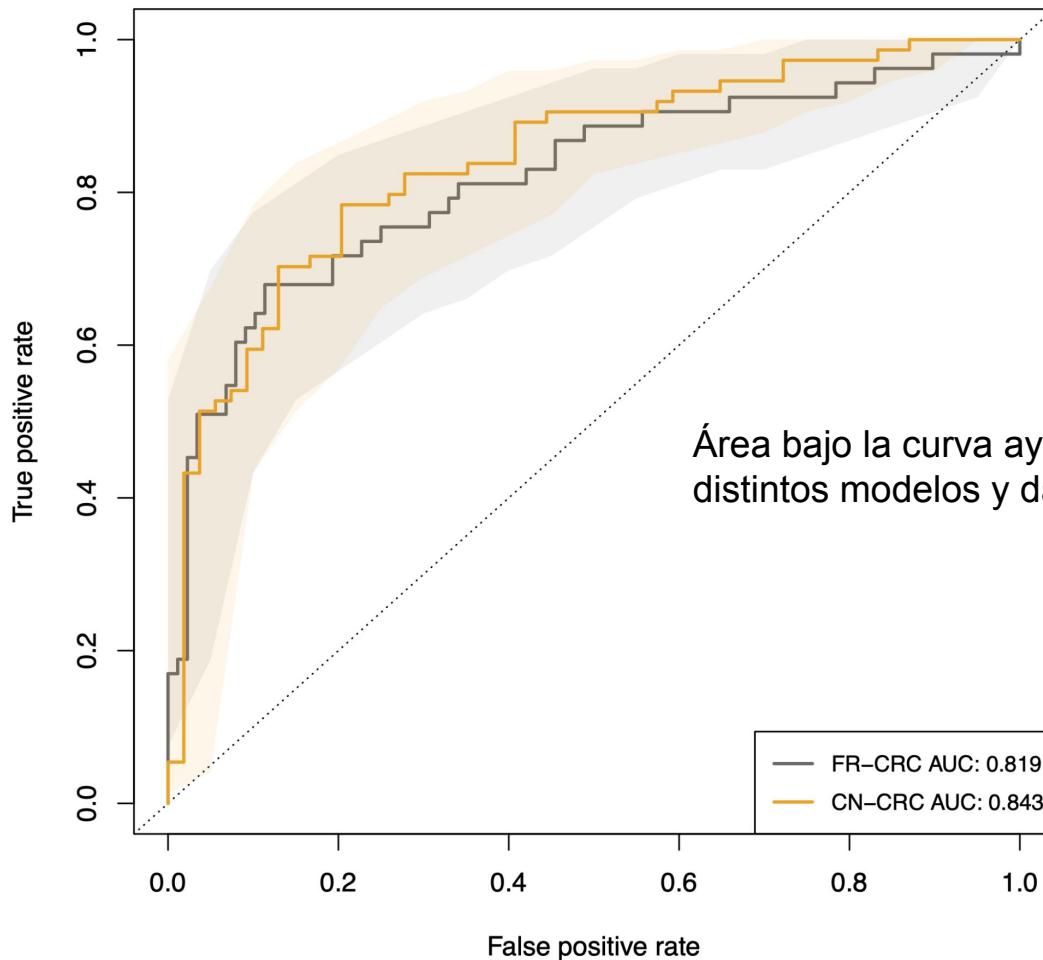
evidencia de cada feature y su utilidad para discriminar entre grupos en cada dataset

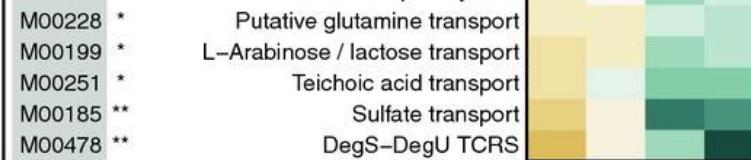
median.rel.weight



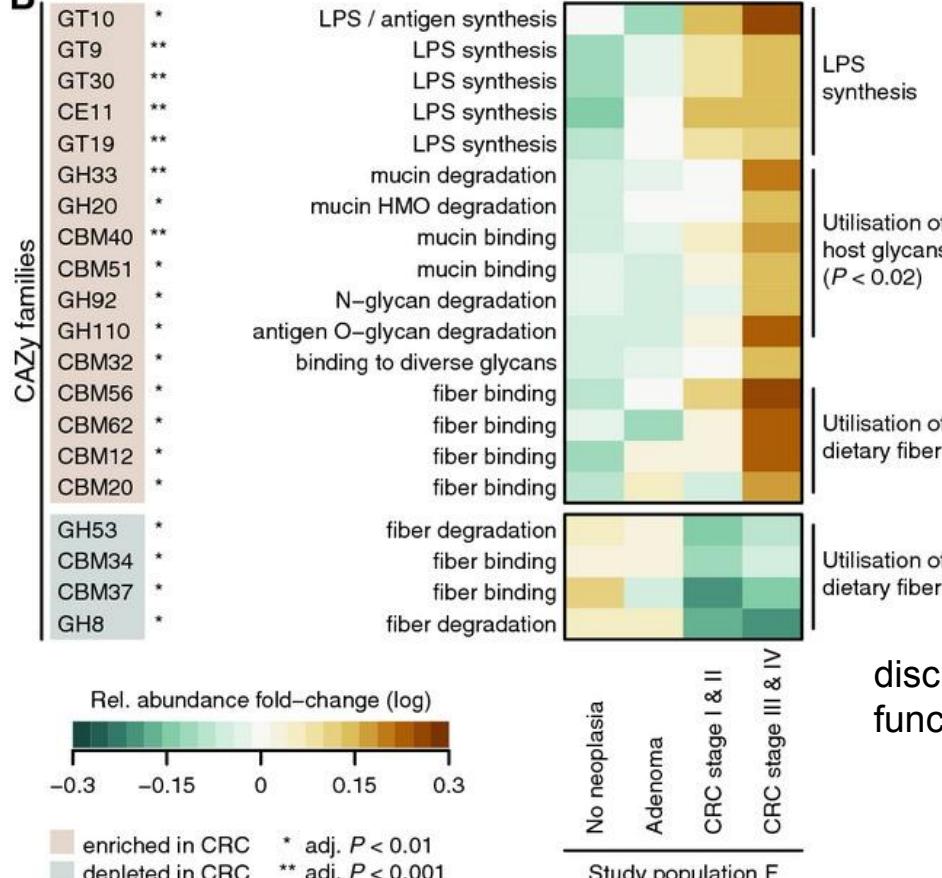
datasets!

### ROC curve for the model





B



discriminación entre datasets puede usar data funcional o de metabolismo también



# Gracias!