WG2:
Benchmark datasets & DREAM Challenge

Progress update up to now

MICROBIOME

cost
EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

CA18131 - Statistical and machine learning techniques in human microbiome studies

https://www.ml4microbiome.eu/

WG2: Benchmark datasets & DREAM Challenge

1) Deliverables
D2.1-2.2: Benchmark data, documentation & web-portal
(some) key people: David Cabrero-Gomez, Saeed Shoaie, Laurent Falquet..

2) DREAM challenge
In preparation; more on this on the next slides

3) Support for the other WGs
- Training & R/Bioconductor Multi-omic data science framework
- STSM / virtual mobility

HUMAN GUT MICROBIOME ATLAS

microbiomeatlas.org

Research Square

Preprints are preliminary reports that have not undergone peer review. They should not be considered conclusive, used to inform clinical practice, or referenced by the media as validated information.

Global and temporal state of the human gut microbiome in health and disease

Saeed Shoaie (saeed.shoaie@kcl.ac.uk)
Centre for Host-Microbiome Interactions, Faculty of Dentistry, Oral & Craniofacial Sciences, King's College London https://orcid.org/0000-0001-5834-4533

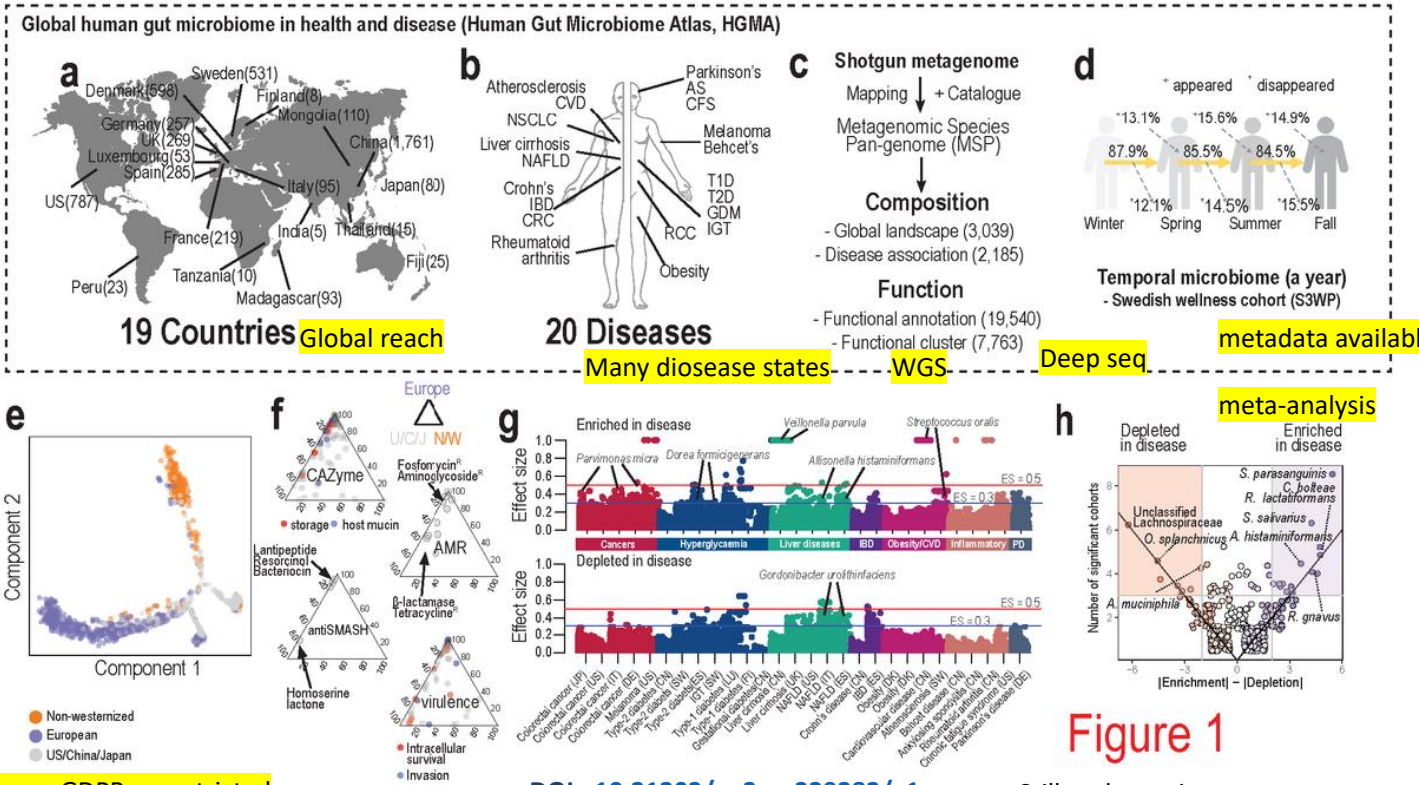
WG2 data progress

Benchmarking data:

- Preprint on HGMA is out
- Continuing to process 5000 samples from the human gut health index metastudy: taxonomic information, functional genes, enzymatic reactions, metabolic pathways and predicted metabolites.
- Additional work on a coherent high-quality CRC data subset.

Processing:

- Metagenome processing: Biobakery Metaphlan, Humann & Melonnpan
- 16S processing: taxonomic profiles functional genes enzymatic reactions metabolic pathways



Global reach | Many diosease states | WGS | Deep seq | metadata available | meta-analysis

Figure 1

open -> GDPR unrestricted

DOI: 10.21203/rs.3.rs-339282/v1

Still under review

DREAM CHALLENGES
powered by Sage Bionetworks

Designed and run by a community of researchers from a variety of organizations, DREAM Challenges invite participants to propose solutions to fundamental biomedical questions — fostering collaboration and building communities in the process. Sage Bionetworks provides the expertise and infrastructure to host challenges via their Synapse platform. Together, we share a vision allowing individuals and groups to collaborate openly so that the "wisdom of the crowd" provides the greatest impact on science and human health.

If you have an idea for a DREAM challenge, please contact us here.

SageBionetworks

EXAMPLES

Tumor Deconvolution DREAM Challenge

Launch: June 3, 2019 (Challenge Now Open!)
Webinar: June 5, 2019

The goal of this Challenge is to evaluate the ability of computational methods to deconvolve bulk expression data, reflecting a mixture of cell types, into individual immune components.

Preterm Birth Prediction: Transciptomics DREAM Challenge

Launch: May 4, 2019 (Register now)
Submission Opens: June 3, 2019
Submission Closes: August 15, 2019
Winners Announced: September 13, 2019

Using whole blood gene expression data collected from pregnant women, participants will develop models to predict gestational age at blood draw and risk of preterm birth.

Ongoing talks and discussions with SAGE

Search for datasets that would be:
- Public (GDPR lifted!)
-high density in terms of samples and diseases
-random samples from population
-ideally single study
-WGS
-sequenced deep-enough @ 5 mio+ reads / sample
-cost-free
-could be supported by COST A postdocs + PhDs
…

Support for other WGs

WG1 State-of-the-art evaluation and update

WG2 Benchmark data & DREAM Challenge
- Including STSM & virtual mobility period (6 months)

WG3 Optimisation and standardisation

WG4 Dissemination and training
- ML4microbiome training schools 2021&2022

Orchestrating Microbiome Analysis

Authors: *Leo Lahti [aut], Sudarshan Shetty [aut], Felix GM Ernst [aut, cre]*

Version: *0.98.9*
Modified: *2021-04-10*
Compiled: *2021-07-29*
Environment: *R version 4.1.0 (2021-05-18), Bioconductor 3.14*

Source: *https://github.com/microbiome/OMA*

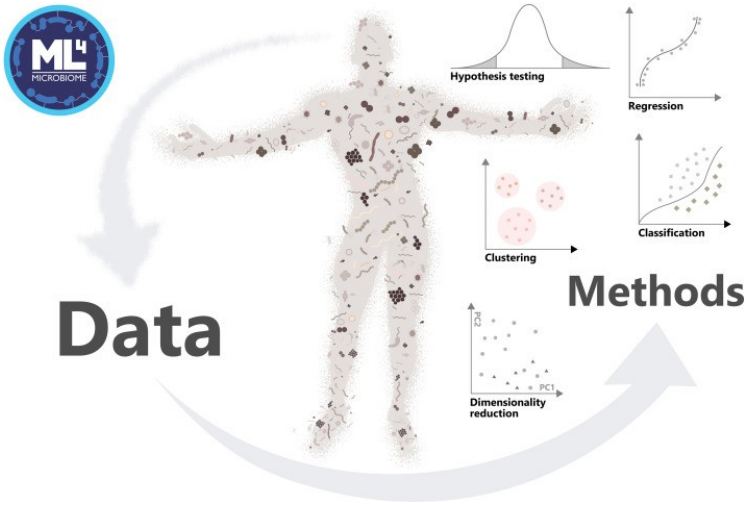


Figure source: Moreno-Indias et al. (2021) Statistical and Machine Learning Techniques in Human Microbiome Studies: Contemporary Challenges and Solutions. Frontiers in Microbiology 12:11.

Online tutorial (beta)

microbiome.github.io

Courses & events:

May, Tirana, Albania (ML4microiome lecture)

June 20-23, Oulu, Finland (ML4microbiome)

Oct 5-7, Barcelona, Spain (ML4microbiome)

Gut Microbiome Atlas as a demonstration data set

WG2 overview of the completed activities

1) Deliverables
D2.1-2.2: Benchmark data, documentation & web-portal

2) DREAM challenge preparations

3) Support for the other WGs
- Hosted STSM / virtual mobility
- Training & R/Bioconductor Multi-omic data science framework (Tirana, Oulu, Barcelona)

4) August 2022 MC & WG meeting coordination
Aug 29-31 Turku, Finland