WG2:
Benchmark datasets &
DREAM Challenge

Summary - June 8, 2023

Main themes for WG2:

1. Benchmark datasets
2. DREAM challenge
3. Supporting the other working groups

WG2: Benchmark datasets & DREAM Challenge

Milestones

M2.1: Agreement and description of benchmark data to acquire (Y1Q3)
M2.2: Uploaded benchmark data on the Web-portal (Y2Q2)
M2.3-4: DREAM Challenge announced (Y2Q2) and closed (Y2Q4)

Deliverables

D2.1: A list of suitable benchmark data to be acquired and their sources
D2.2: Public benchmark data at Web-portal with descriptive documentation
D2.3-4: Two publications/reports associated to the DREAM Challenge

1. Benchmark datasets

HUMAN GUT MICROBIOME ATLAS

GEOGRAPHY SPECIES DISEASES ABOUT DOWNLOADS

Introduction

The Human Gut Microbiome Atlas (HGMA) provides i) quantitative information of global shotgun metagenomics of human gut microbiome ii) function/phenotype information of the human gut-associated metagenomic species pan-genomes (MSP) and iii) provide global map of human gut microbiome. Global shotgun metagenomics of normal human gut microbiome from 20 countries are presented with i) species abundance, ii) gene richness and iii) enterotypes. In addition, comprehensive analysis enabled the identification of species enriched/depleted in 23 different diseases with functions/phenotype of respective species, detailing the dysbiosis in gut microbiome composition.

Geography

Global shotgun metagenomics provide a detailed understanding of regional variations of human gut microbiome similar to the genetic variations between different human populations. The Human Gut Microbiome Atlas presented the species abundance of 20 countries in five different continents. Navigating the list of region-enriched or disease-associated species, variations of normal and disease populations can be explored at the species summary page.

Processing using Metagenome ATLAS and various ML methods

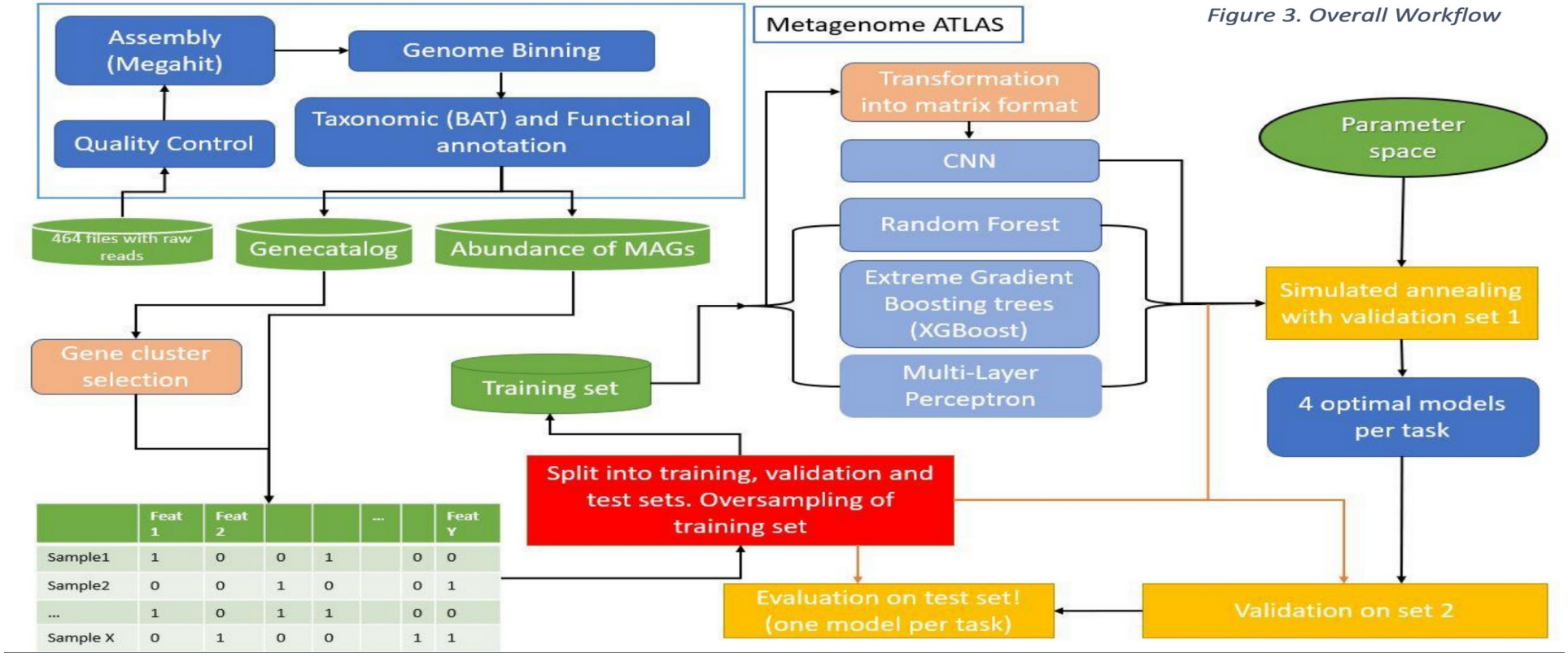


Figure 3. Overall Workflow

Available matrices: taxons *vs* samples and genes *vs* samples

https://www.microbiomeatlas.org/

Sensitivity of human biomedical data – sequencing and metadata - became a bottleneck for the collection and open sharing.

Implications:
→ Limitations on the web portal & training activities
→ Redesign of the DREAM challenge

Solutions for data availability limitations

Collecting complementary data for training activities:
- CRC data compilation (with the other WGs)
- R/Bioconductor open microbiome data resources

DREAM challenge
- synthetic data generation for competition
- executed using a partner data set where we could run the experiments in-house

2. DREAM challenges

TASK: time-to-event predictions of the heart failure using a combination of the information from fecal microbiome composition and clinical data

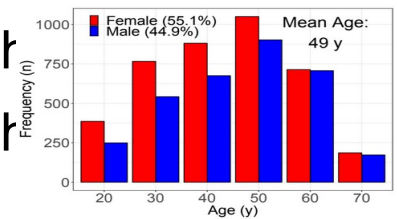
Data summary: FINRISK2002

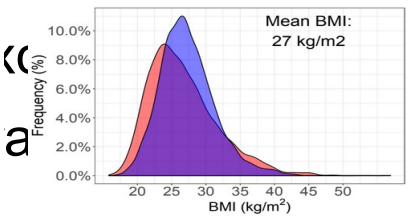
Cohort

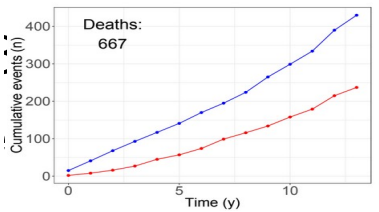
- 7231 adult stool samples (54% participation rate) *collected* in **2002**
- 17+ year follow up (in 2019)
- comprehensive health info from Finnish health registers (doctor visits, diagnoses, medication purchases..)

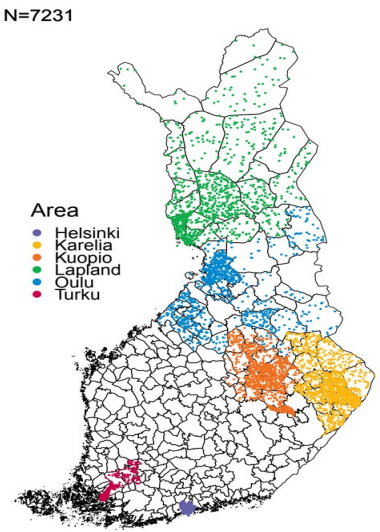
Omics and clinical measurements

- **shallow shotgun metagenome** (~1M reads/sample)
- 16S
- stool metabolome (NMR & MS)
- h
- h



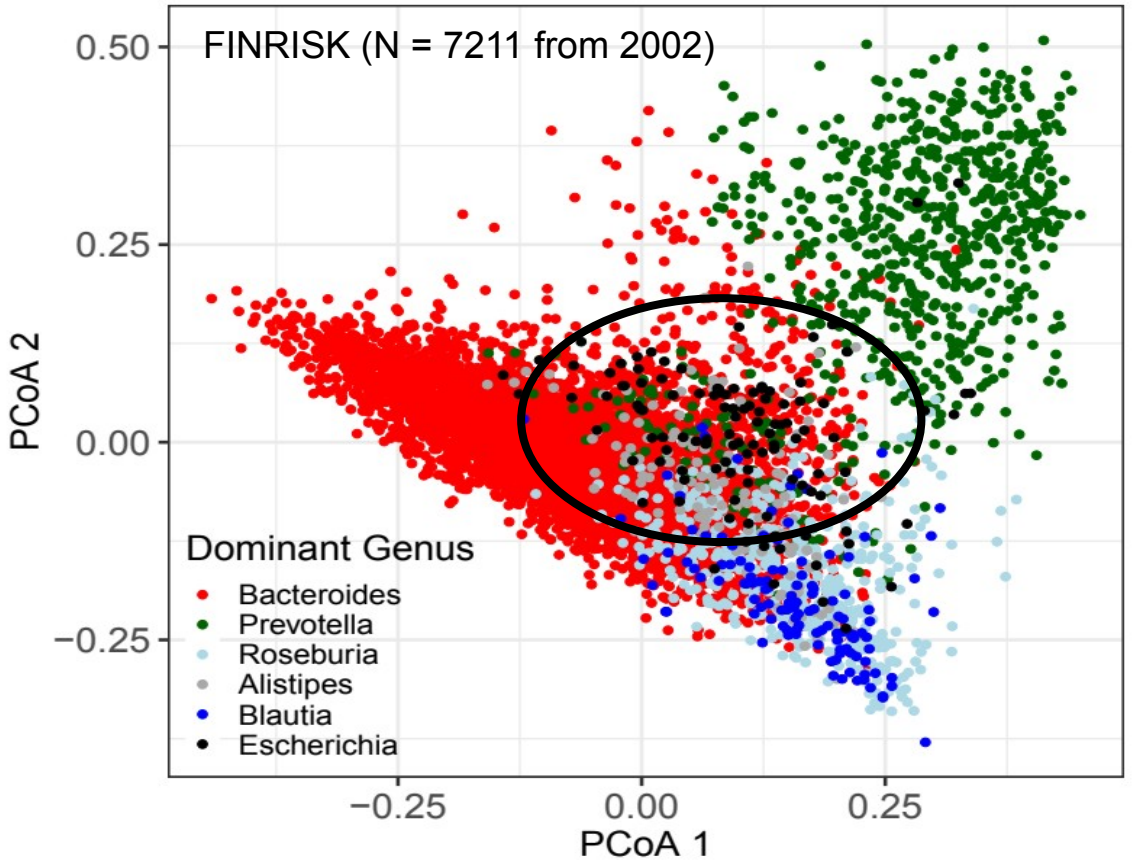


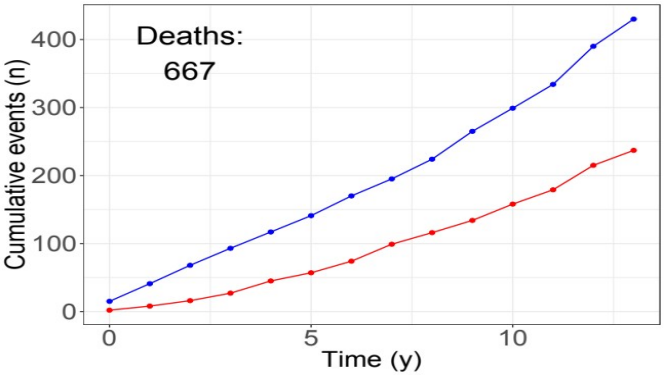




NATIONAL INSTITUTE FOR HEALTH AND WELFARE

Predicting overall mortality risk in Finnish adult population





Taxonomic signatures of cause-specific mortality risk in human gut microbiome

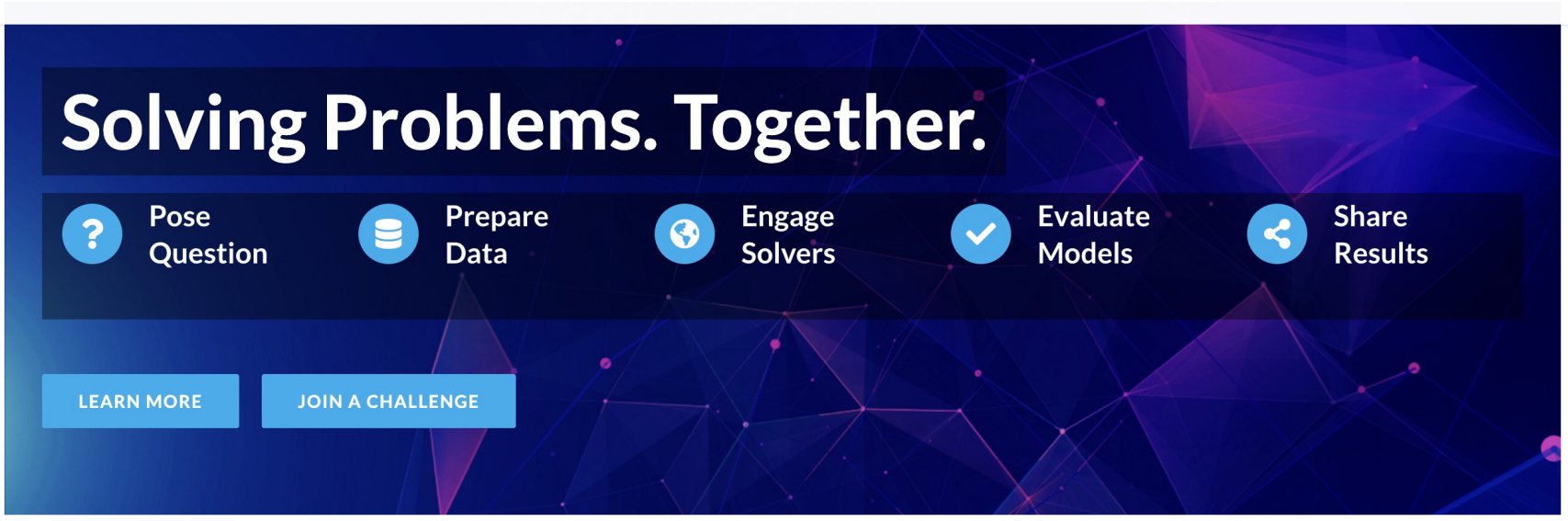
Aaro Salosensaari, Ville Laitinen, Aki S. Havulinna, Guillaume Meric, Susan Cheng, Markus Perola, Liisa Valsta, Georg Alfthan, Michael Inouye, Jeramie D. Watrous, Tao Long, Rodolfo A. Salido, Karenina Sanders, Caitriona Brennan, Gregory C. Humphrey, Jon G. Sanders, Mohit Jain, Pekka Jousilahti, Veikko Salomaa, Rob Knight, Leo Lahti ✉ & Teemu Niiranen ✉

Nature Communications **12**, Article number: 2671 (2021) | Cite this article

NATIONAL INSTITUTE FOR HEALTH AND WELFARE

DREAM CHALLENGES
powered by Sage Bionetworks

What is DREAM Challenge?



DREAM Challenges use crowd-sourcing to solve complex biomedical research questions

Open Challenges

COugh Diagnostic Algorithm for Tuberculosis (CODA TB) DREAM Challenge

By ambernelson | September 28, 2022 | Categories: Challenges, Infectious Disease, Open Challenge, Pre-Register

Register Tuberculosis (TB), a communicable disease

Heart Failure Prediction: Microbiome

By ambernelson | September 21, 2022 | Categories: Challenges, Open Challenge

Register Cardiovascular diseases are the leading cause of

Read More › 0

BraTS Continuous Evaluation

By ambernelson | June 6, 2022 | Categories: Cancer, Challenges, Open Challenge | Tags: Cancer

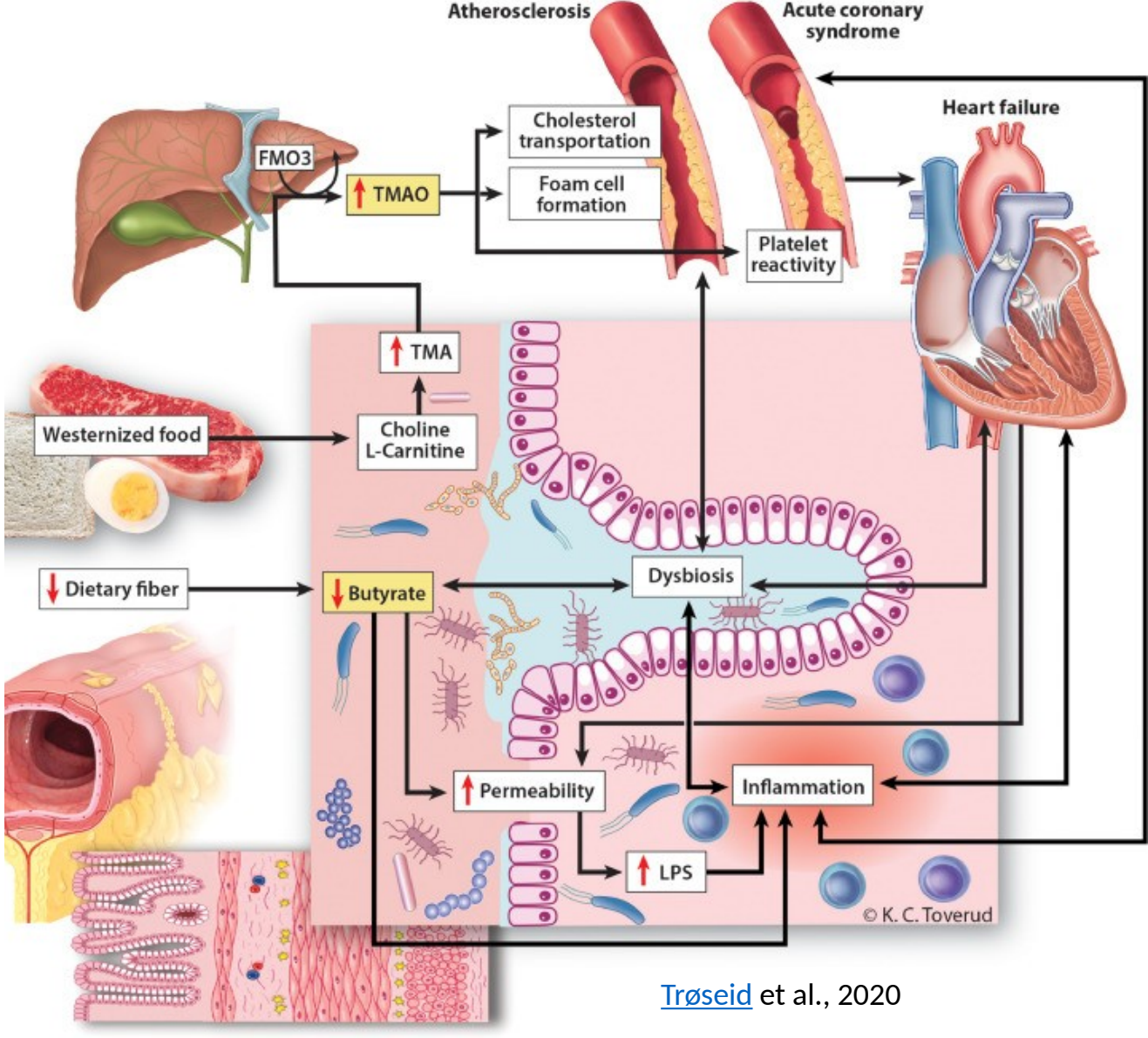
Register Brain tumors are among the deadliest types

Read More › 0

•More information: https://dreamchallenges.org/

Can microbiome improve the predictions of heart failure?

- Heart failure remains difficult to diagnose due to the heterogeneity of the disease and a lack of agreement of diagnostic criteria
- The link between heart failure and the microbiome has long been postulated.
- Lack of study with temporal follow up



Trøseid et al., 2020

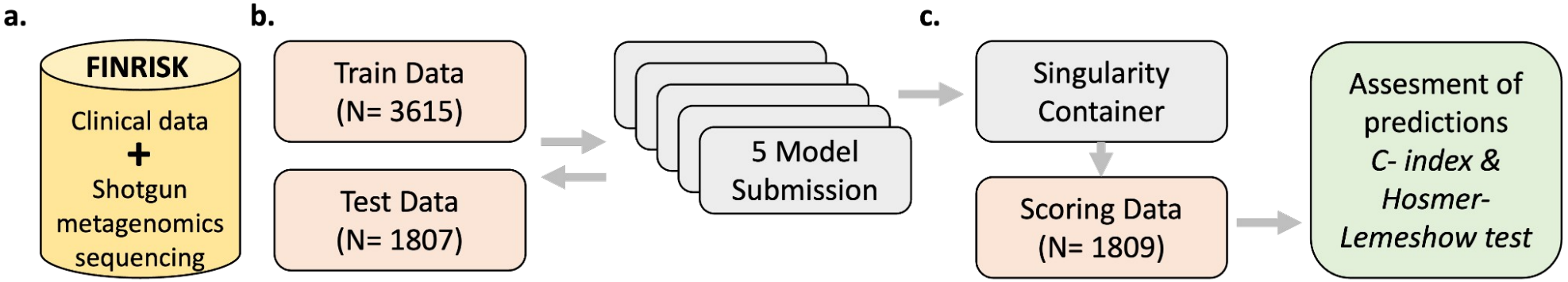
Heart failure: pilot results

Conventional risk factors / microbiome (diff): RSF

	Covariate set 1	Covariate set 2	Covariate set 3
HFAIL (N=478/6053)	76.3/79.1 (2.9)	80.3/81.4 (1.1)	82.8/83.0 (0.2)
HFAIL_STRICT (N=242/6094)	78.4/80.2 (1.9)	82.1/82.5 (0.4)	83.1/83.1 (0)
DEATH (N=715/6172)	77.0/79.4 (2.4)	80.9/81.8 (0.9)	81.8/81.9 (0.1)

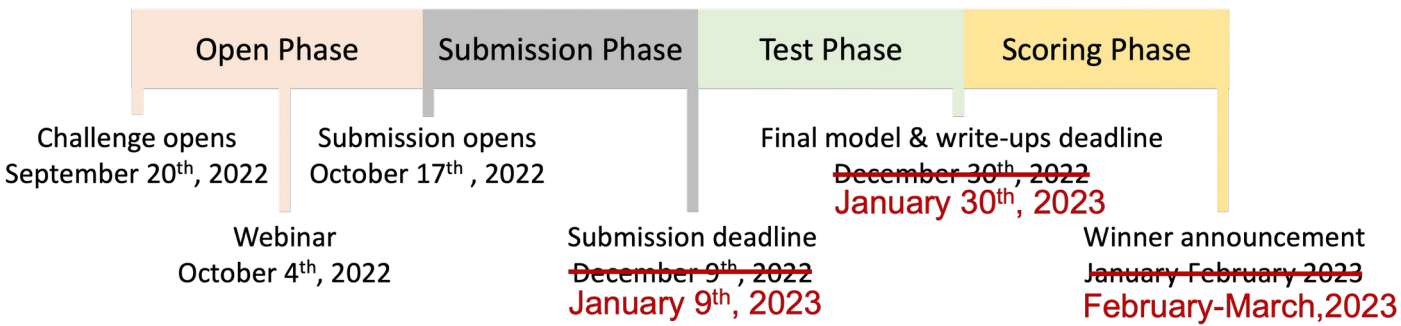
1) Age & Sex
2) + Sys BP, BP treatment, prev. CHD (N=181; excl?), smoker, nonHDL cholesterol
3) + BMI, prevalent Diabetes, prevalent heart problems, (alcohol consumption?)

FINRISK DREAM Challenge Workflow



Scan for more info

https://synapse.org/finrisk



Acknowledgements: DREAM

Initiated by:

MICROBIOME COST Action (CA18131)

Organized by:

Finnish institute for health and welfare

Aki Havulinna
Pekka Jousilahti
Veikko Salomaa

UNIVERSITY OF TURKU

Pande Putu Erawijantari
Rajesh Shigdel
Leo Lahti
Teemu Niiranen

HEIDELBERG FACULTY OF MEDICINE

Ece Kartal
Rebecca Levinson
Julio Saez-Rodriguez

UC San Diego

Rob Knight

Baker HEART & DIABETES INSTITUTE

Mike Inouye

With:

SageBionetworks

Funded by

Informatics for Life

cost EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

Funded by the European Union

3. Supporting the other working groups

Support for other WGs

WG 1 State-of-the-art evaluation and update

WG 2 Benchmark data & DREAM Challenge

WG 3 Optimisation and standardisation

WG 4 Dissemination and training

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*
License: *CC BY-NC-SA 3.0 US*
Copyright:
Source: *https://github.com/microbiome/OMA*

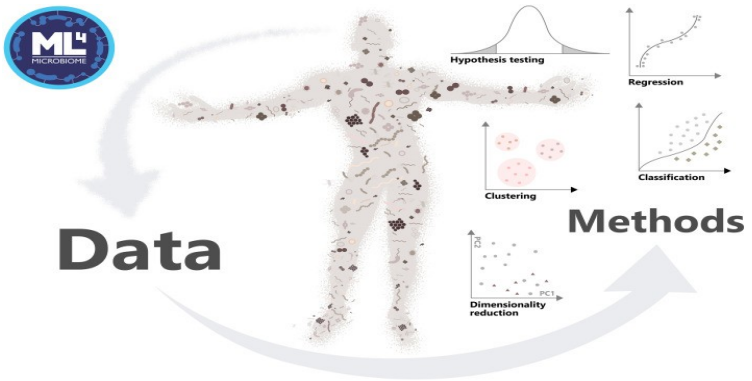


Figure source: Moreno-Indias et al. (2021) Statistical and Machine Learning Techniques in Human Microbiome Studies: Contemporary Challenges and Solutions. Frontiers in Microbiology 12:11.

WG2 contributions to training activities

<u>Courses & events:</u>

2022
- Tirana, Albania
- Oulu, Finland
- Nijmegen, NL
- Barcelona, Spain

2023
- Utrecht, NL
- Oulu, Finland
- Nijmegen, NL

ML4microbiome data case studies

Recent Outreach Activities

1) **International Human Microbiome Congress** (IHMC) 2023. FINRISK DREAM challenge. Ece Kartal.
2) (upcoming) **ISMB/EECB** Microbiome COSI track 23-27.7.2023, talk and poster(with Ece Kartal): Predicting Incident Heart Failure from the Microbiome: The FINRISK DREAM challenge
3) The 14th Finnish Gut Day 26.1.2023, poster presentations: Predicting Incident Heart Failure from the Microbiome: The FINRISK DREAM challenge
4) Finnish Bioinformatics Day 25.5.2023, talk: Predicting Incident Heart Failure from the Microbiome: The FINRISK DREAM challenge

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 (e..g Norway → Finland)
- Short research visits (e.g. Finland → NL)
- Training (Tirana, Oulu, Barcelona 10/22, Nijmegen 7/2021-2023, Utrecht)

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

5) Outreach: IHMC, ISMB, …

More detailed & extensive report in preparation.

Impact & Outputs

Data

- Human Gut Metagenome Atlas; CRC data compilation; R/Bioconductor public demo data sets; Synthetic data and baseline model generated for DREAM challenge

Preprints / in prep.

- Human Gut Microbiome Atlas; DREAM publications

Conferences

- IHMC, ISMB, local events

Training

- Contributions to 7 training events in 4 countries

Special thanks

D2.1-2.2: Benchmark data, documentation & web-portal
David Cabrero-Gomez, Saeed Shoaie, Laurent Falquet..

D2.3-D2.4 DREAM challenge
Pande Erawijantari, Ece Kartal, Rebecca Levinson, Julio Saez-Rodiguez, Aki Havulinna, Veikko Salomaa