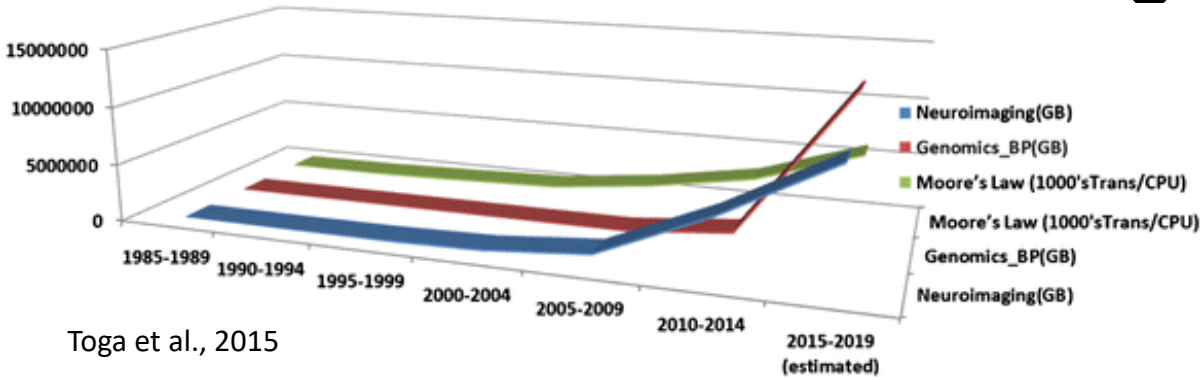
Heart Failure Prediction: Microbiome
FINRISK DREAM Challenge

Pande Putu Erawijantari
Turku Collegium for Science, Medicine, and Technology – Postdoctoral researcher
Department of Computing, Faculty of Technology, University of Turku,
Finland
pande.erawijantari@utu.fi

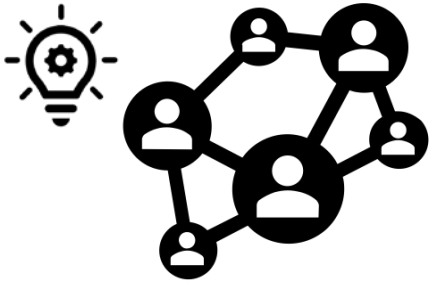
What motivates crowd-sourcing science?



Toga et al., 2015

Exponential growth of data in biomedical research

Growth goes beyond the capacity and expertise of individuals or classic research groups



joint effort of a large number of scientists with a diverse set of skills → accelerate discovery?

DREAM CHALLENGES
powered by Sage Bionetworks

Critical Assessment of Genome Interpretation

MICROBIOME
COMMUNITY OF SPECIAL INTEREST

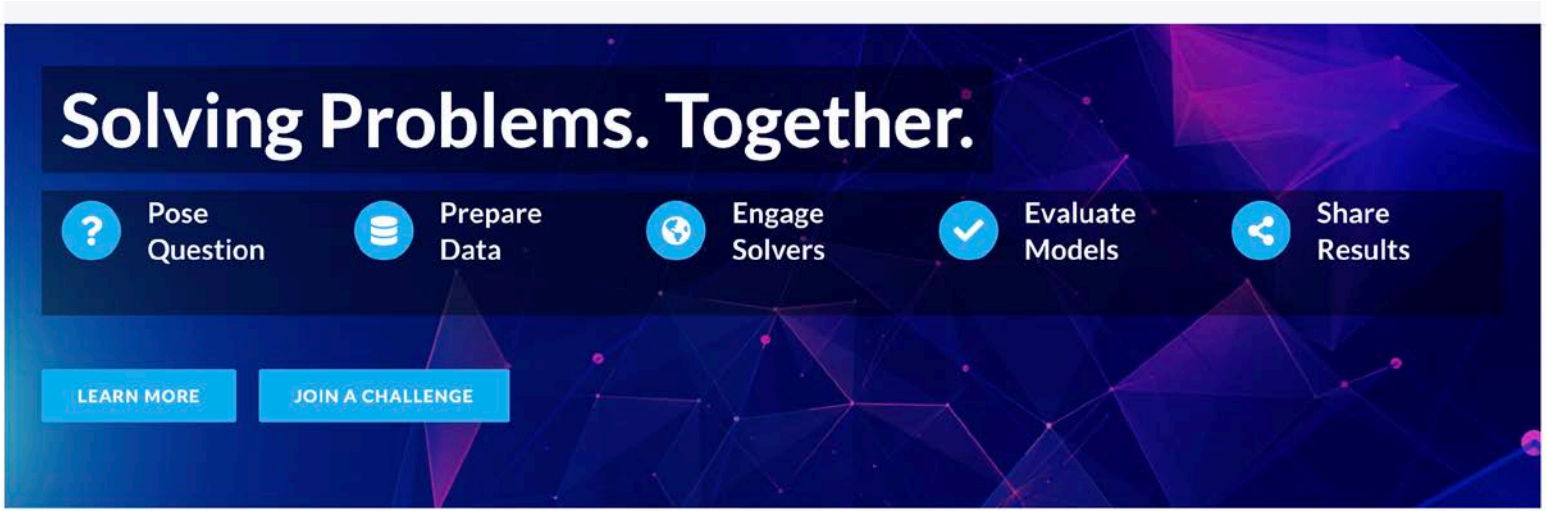
kaggle

CASP15

Community challenge in biomedical research

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

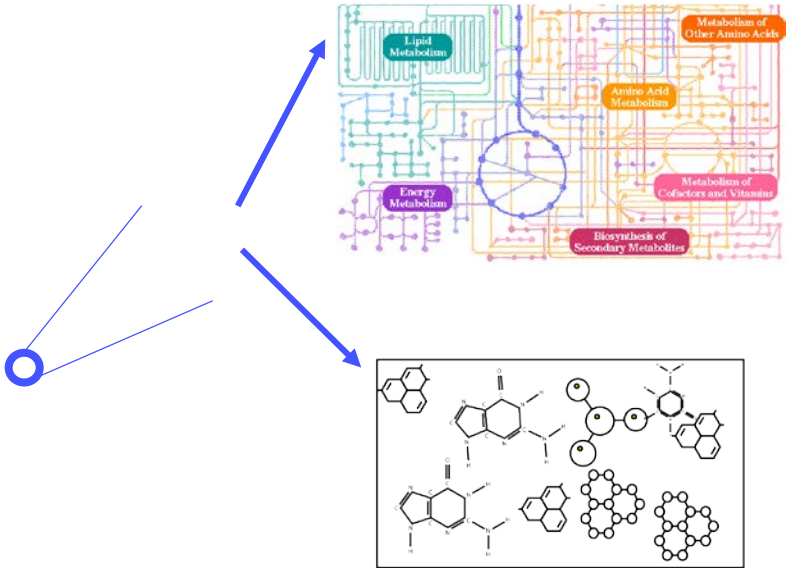
By ambernelson | June 6, 2022 | Categories: Cancer, Challenges, Open Challenge | Tags: Cancer

Register Brain tumors are among the deadliest types

Read More › 0

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

Overview of human gut microbiome



Factors

- Diet
- Lifestyle
- Geography
- Age
- Medication
- Diseases
- Genetics

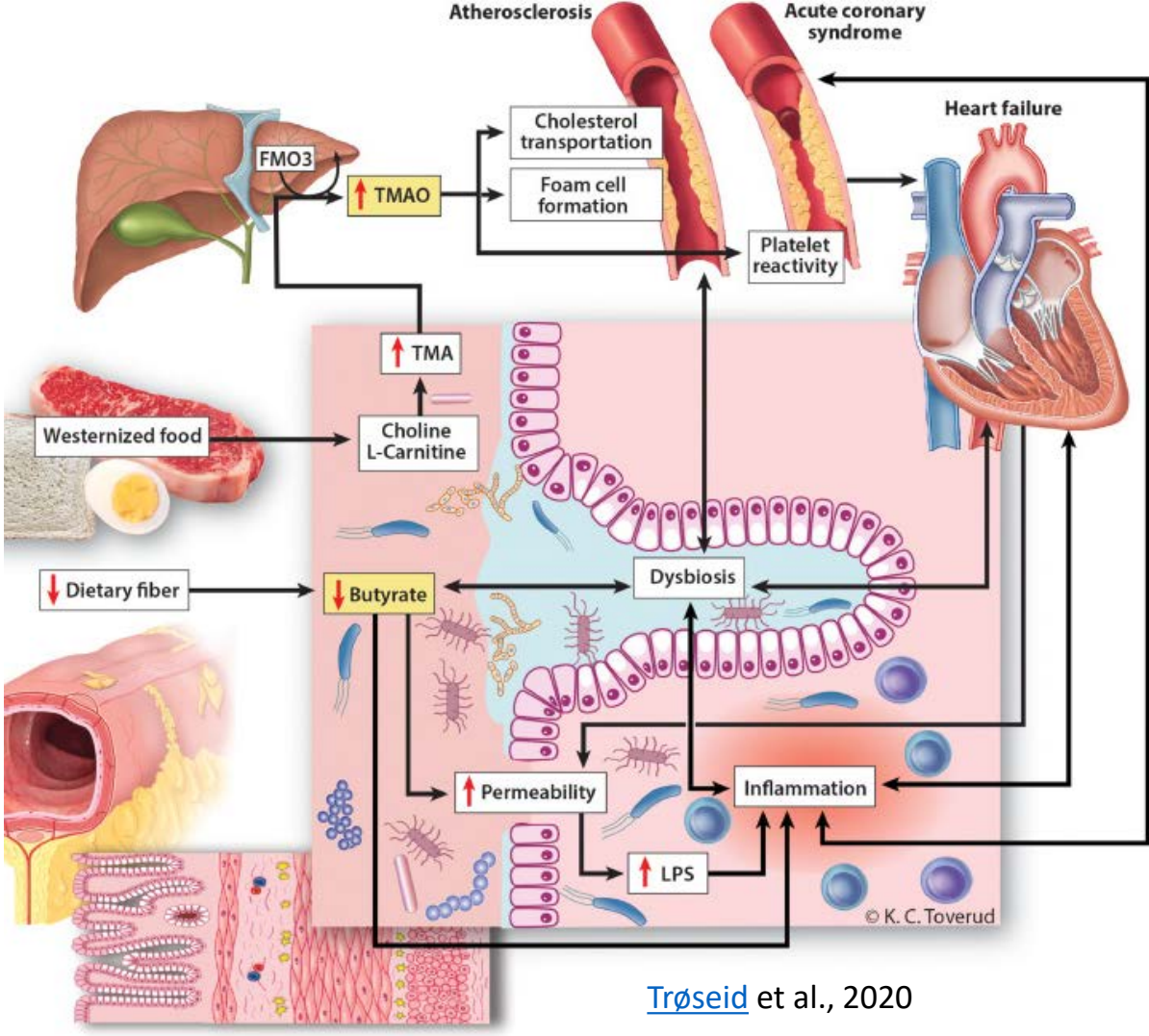
Roles

- Protecting host against pathogenic microbes
- Modulating immunity
- Regulating metabolic processes

Li et al., Nature Biotechnology, 2014, Human body picture designed by freepik.com

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



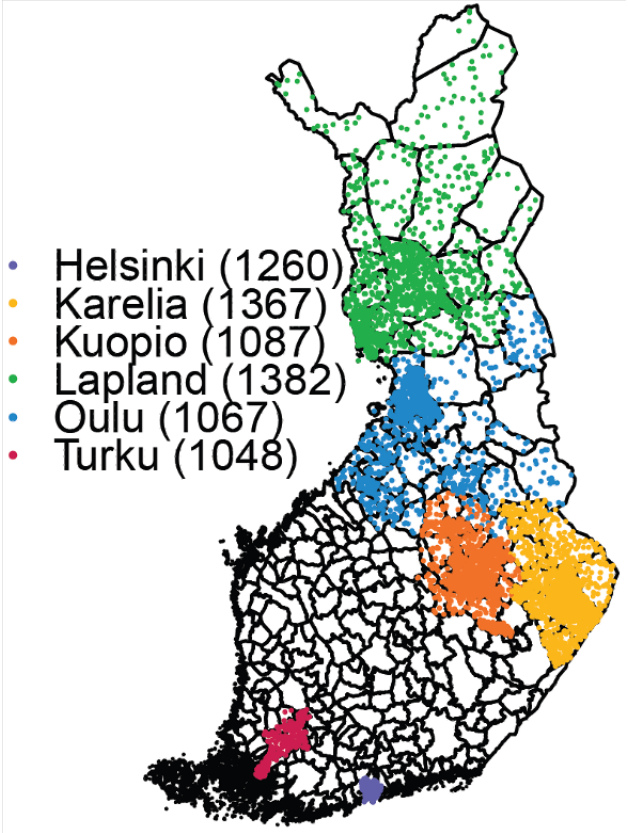
Trøseid et al., 2020

FINRISK 2002

13,498 invitees

↓

8,738 participated → 7,231 donated Fecal sample (54% participation rate) → DNA extractions And shallow metagenomic sequencing (~1M reads/sample)

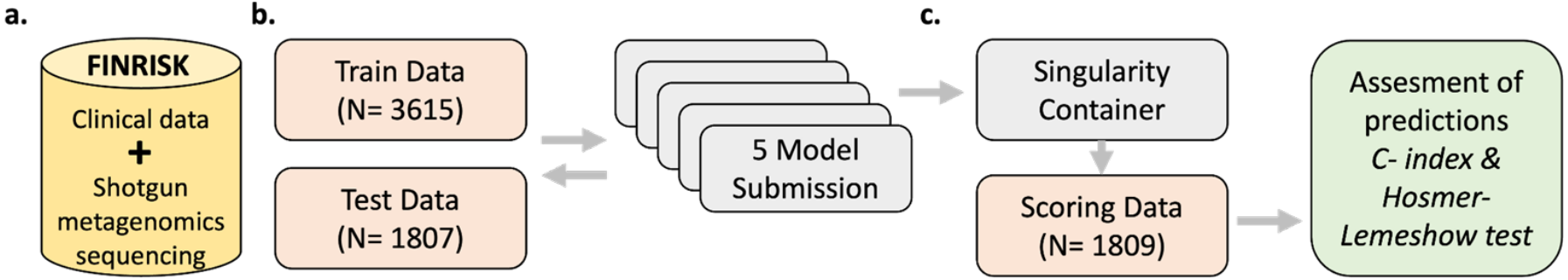


Participants were followed through for 15+ years.
There are **493 incident heart failure** among 6902 participants with complete phenotype

Borodulin et al., 2018

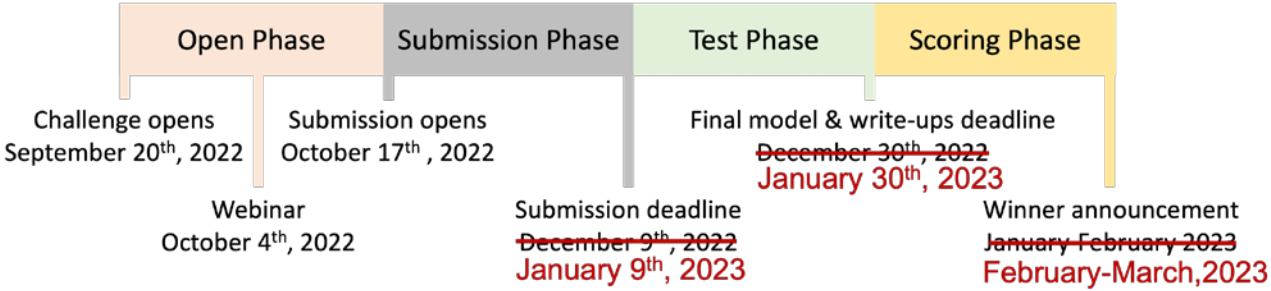
TASK: time-to-event predictions of the heart failure using a combination of the information from fecal microbiome composition and clinical data

FINRISK DREAM Challenge Workflow



Scan for more info

https://synapse.org/finrisk



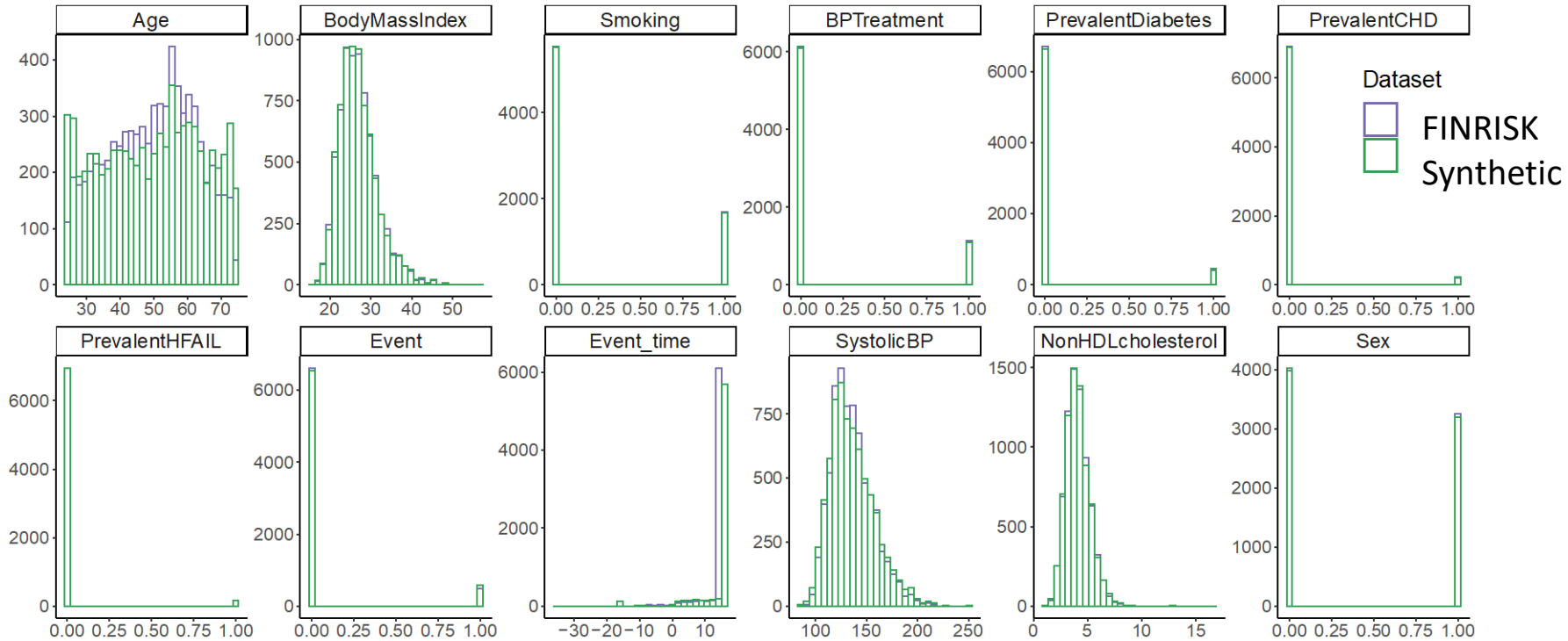
Synthetic data generation: Why is it necessary?

IMPORTANT! The FINRISK data contains sensitive personal information from healthcare registers which cannot be shared **without formal agreement with THL biobank**

Criteria: preserve the covariance structure while privatizing the data

Rank-based Inverse Normal Transformation (by Aki Havulinna, 2022)

Synthetic vs Real data on clinical covariates

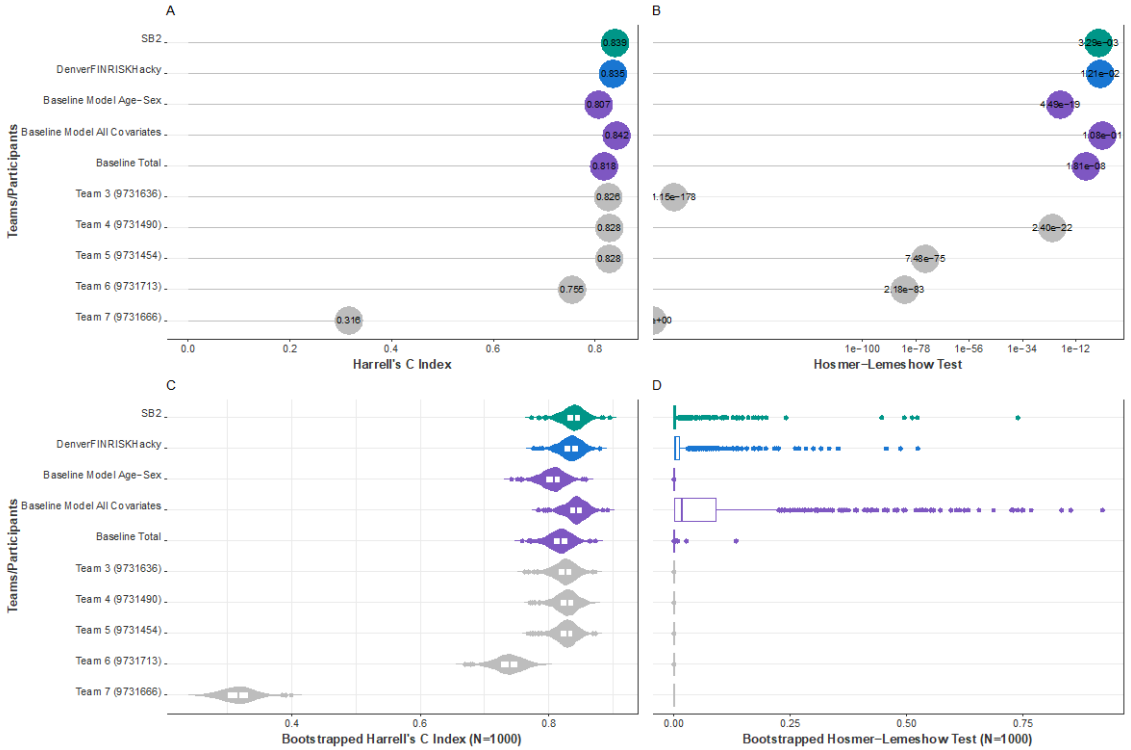


Received submissions

- 35 VALID submissions from 9 participants were accepted in submission phase
- 7 teams submitted final model

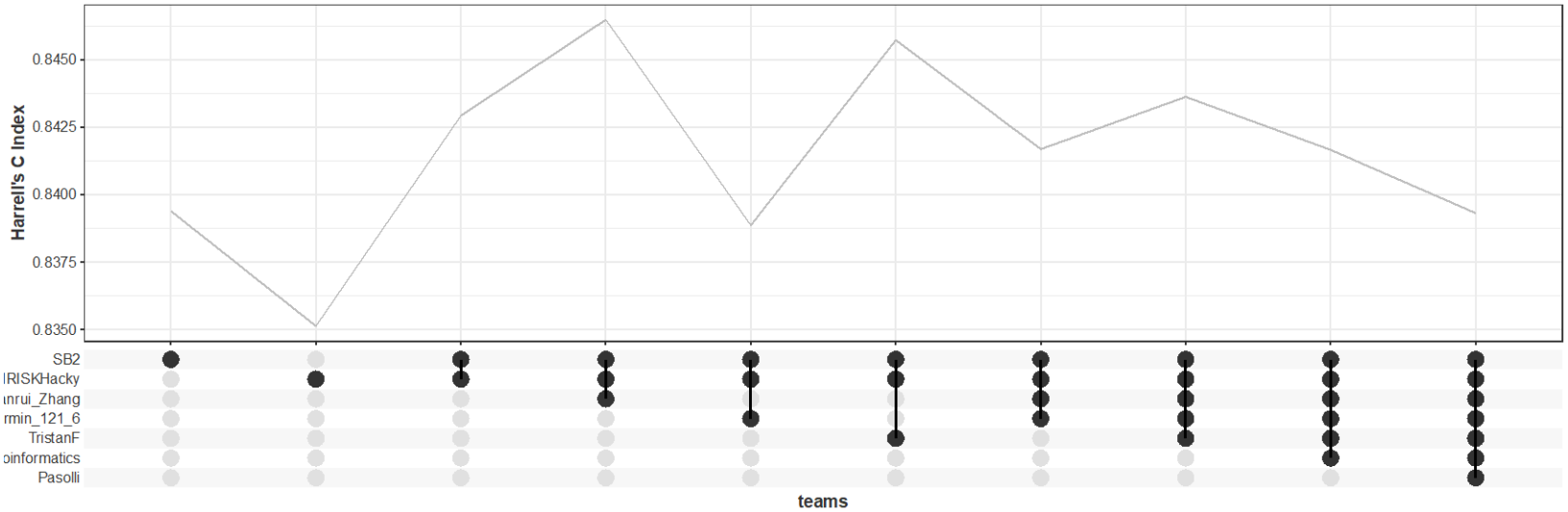
id	createdOn	submitterid	status	harrell_c	hoslem_test
9731637	1/29/2023 11:29 PM	SB2	SCORED	0.839395379197954	0.0032924949185753
9731625	1/29/2023 7:12 PM	DenverFINRISKHacky	SCORED	0.835136923343354	0.0120891360271361
9731490	1/27/2023 1:26 AM		SCORED	0.827974369335565	2.40190686358651e-22
9731454	1/26/2023 11:55 AM		SCORED	0.827752343482875	7.48389464769304e-75
9731636	1/29/2023 11:02 PM		SCORED	0.825842921149739	1.14999761860238e-178
9731713	1/31/2023 12:33 AM		SCORED	0.754830172425277	2.18334641086665e-83
9731666	1/30/2023 9:35 PM		SCORED	0.315609749599243	0

Received model performance vs Baseline model

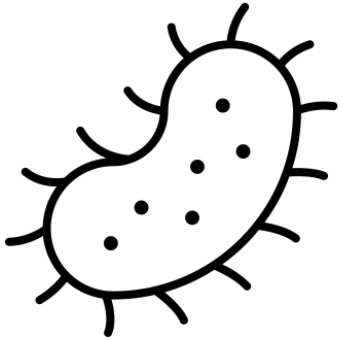


Post Challenge Phase

- Model refinement from each winning team
- Extract knowledge through “wisdom of the crowd” approach by assessing the ensemble model performance
- Investigating the most predictive features to discuss the clinical relevancy of the model



Perspective of the challenge



- Open avenue to explore the potential of microbiome-based biomarkers to complement clinical risk factors in identifying individuals with an elevated risk of Heart Failure.
- The crowdsourced *Challenge* provides a unique platform for solving scientific problem.
- The synthetic dataset has proven beneficial for challenge participants and could become an option for data sharing while ensuring privacy protection.

Acknowledgements

Initiated by:

MICROBIOME COST Action (CA18131) lead by Leo Lahti

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
Mohit Jain

Baker HEART & DIABETES INSTITUTE

Mike Inouye

With:

SageBionetworks

Thanks to all FINRISK DREAM Challenge participants especially SB2 and DenverFINRISKHacky who has been active in the post-challenge phase

Funded by

Informatics for Life

cost EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

Funded by the European Union