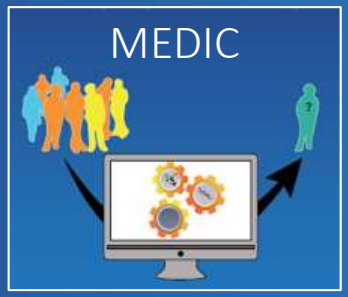
sbv IMPROVER

SYSTEMS BIOLOGY VERIFICATION

The sbv IMPROVER: Crowdsourcing platform for big data analysis in systems biology



www.sbvimprover.com

Lusine Khachatryan, PMI R&D, Philip Morris Products S.A., Neuchatel, Switzerland

October 30, 2020

PMI SCIENCE

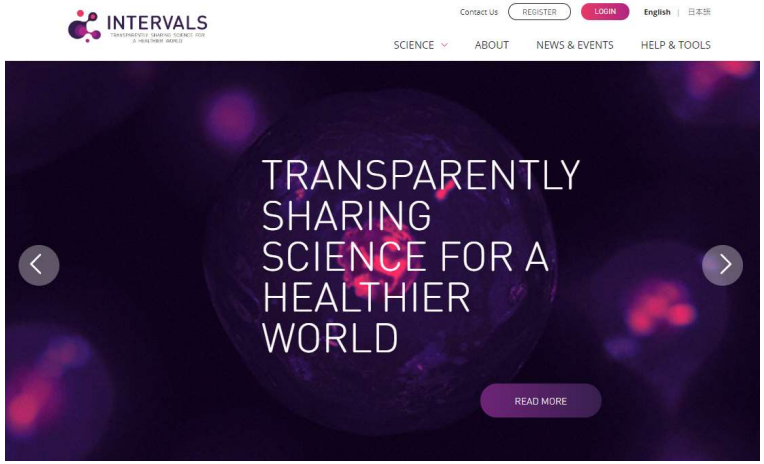
PHILIP MORRIS INTERNATIONAL

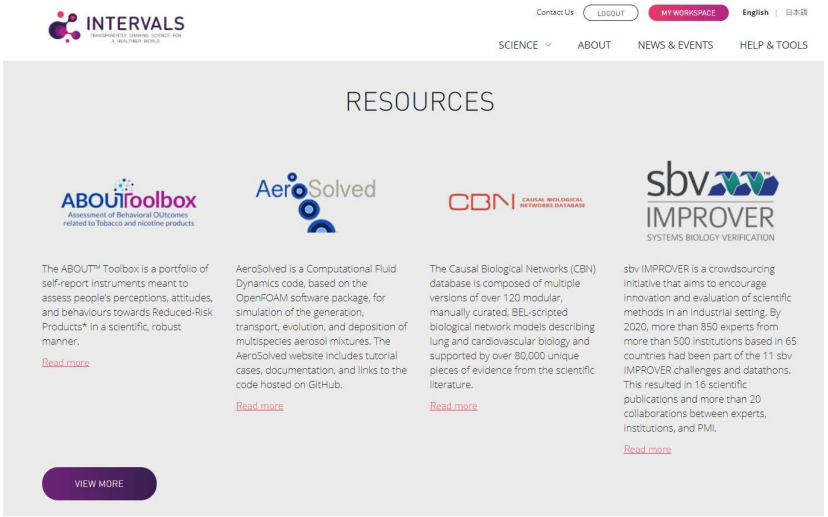
INTERVALS AND SBV IMPROVER

INTERVALS: Scientific data transparency applied to Industry

INTERVALS aims to establish a **community** and a public **repository** for 21st-century preclinical and clinical (systems) **inhalation toxicology assessment** data and results that support open-data principles.

It also hosts or promotes a number of useful resources in various fields of biology and toxicology, including **sbv IMPROVER.**





Boué S, Exner T, Ghosh S et al. Supporting evidence-based analysis for modified risk tobacco products through a toxicology data-sharing infrastructure [version 2; referees: 2 approved] F1000Research 2017, 6:12 (doi: 10.12688/f1000research.10493.2)

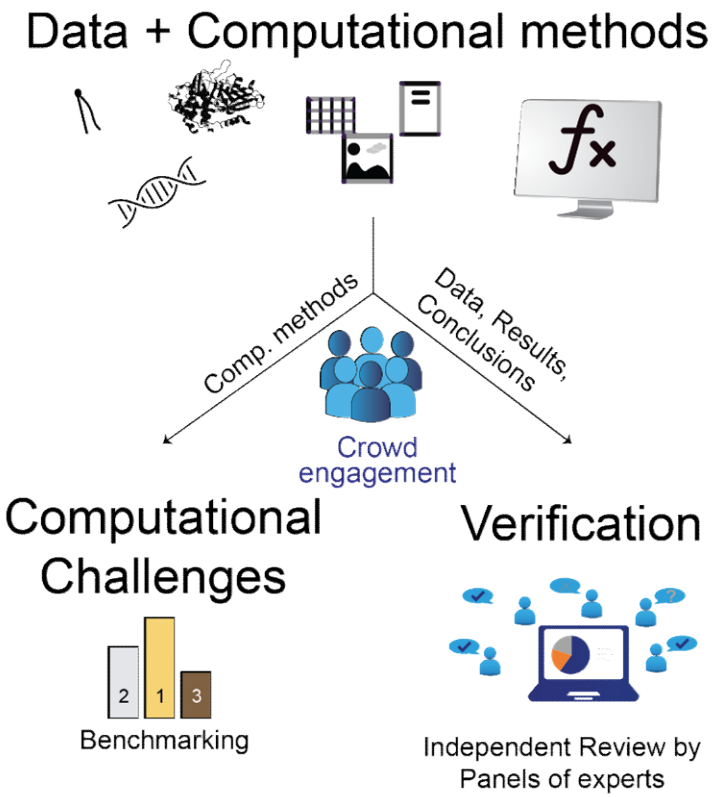
sbv IMPROVER

sbv IMPROVER stands for **S**ystems **B**iology **V**erification combined with **I**ndustrial **M**ethodology for **Pro**cess **Ve**rification in **R**esearch.

This approach aims to provide a measure of quality control in industrial research and development by verifying the methods used. It is complementary to the classical peer-review system.

Double-blind performance assessment to address the concern of the self-assessment trap (*Norel R, Molecular Systems Biology, 2011*)

The sbv IMPROVER project is a collaborative effort led and funded by PMI Research and Development.



Meyer, P., et al. 2011. Verification of systems biology research in the age of collaborative competition. Nature biotechnology 29, 811–5. https://doi.org/10.1038/nbt.1968

sbv IMPROVER Challenge: Build, run, score, analyze, and publish

Double-Blind Performance Assessment

- Predefined scoring strategy approved by a scoring review panel (SRP) of external experts
- Scoring metrics released after closure of the challenge
- Scoring of anonymized participants' submissions
- Final team ranking reviewed and approved by the SRP

Past challenges

COMPUTATIONAL CHALLENGES

2019-20

METAGENOMICS DIAGNOSIS FOR INFLAMMATORY BOWEL DISEASE CHALLENGE (MEDIC)

MEDIC aimed to investigate the diagnostic potential of metagenomics data to classify patients with Inflammatory Bowel Disease (IBD) and non-IBD subjects. The participants have attempted to classify Ulcerative Colitis (UC) and Crohn's Disease (CD) subjects.

Learn more

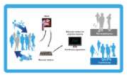
2018

MICROBIOTA COMPOSITION PREDICTION CHALLENGE

The first phase of the microbiomics challenge named "Microbiota composition prediction" aimed at identifying state-of-the-art computational microbiome analysis pipeline(s) that can be used as off-the-shelf solutions for scientists to best recover the composition and relative abundance of bacterial communities present in a sample.

Learn more

2014-15



SYSTEMS TOXICOLOGY (SYSTOX) CHALLENGE

The SysTox Challenge aimed at verifying that robust and sparse human-specific and species-independent gene signatures of exposure response can be extracted in whole blood gene expression data from human and rodent to predict exposed and non-exposed group labels.

Learn more

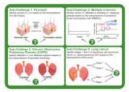
2013

SPECIES TRANSLATION CHALLENGE (STC)

The Species Translation Challenge aimed at verifying that changes in phosphorylation status and gene set activation induced by cellular response to 52 different perturbations in human cells can be predicted to a certain extent given responses generated in rat cells.

Learn more

2012



DIAGNOSTIC SIGNATURE CHALLENGE (DSC)

The goal of this Challenge was to assess and verify computational approaches that classify clinical samples based on transcriptomics data.

Learn more

NETWORK VERIFICATION CHALLENGES

2018-19

NETWORK VERIFICATION CHALLENGE 3: LIVER XENOBIOTIC METABOLISM NETWORKS

NVC3 aimed at verifying three biological network models to ensure their relevance to liver xenobiotic metabolism.

Learn more

2014-15

NETWORK VERIFICATION CHALLENGE 2: LUNG BIOLOGY

NVC2 aimed at verifying fifty biological network models describing cell fate, cell stress, cell proliferation, inflammation, and tissue repair and angiogenesis to ensure their relevance to lung biology and COPD.

Learn more

2013-14

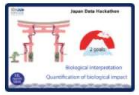
NETWORK VERIFICATION CHALLENGE 1: LUNG BIOLOGY

NVC1 aimed at verifying fourteen biological network models describing cell fate, cell stress, and inflammation to ensure their relevance to lung biology.

Learn more

MINI-CHALLENGES AND DATATHONS

2017

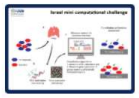


SBV IMPROVER DATATHON - JAPAN

The datathon organized in Japan addressed how to evaluate qualitatively (on which pathways and biological processes) and quantitatively the extent of biological impact of a system to specific exposure.

Learn more

2017



SBV IMPROVER EPIGENOMICS CHALLENGE - ISRAEL

The challenge organized in Israel aimed at answering the questions whether a smoke exposure signature can be extracted (i) from DNA methylation levels of DNA cis-regulatory elements (CRE) or (ii) from expression data of genes controlled by differentially methylated DNA CRE.

Learn more

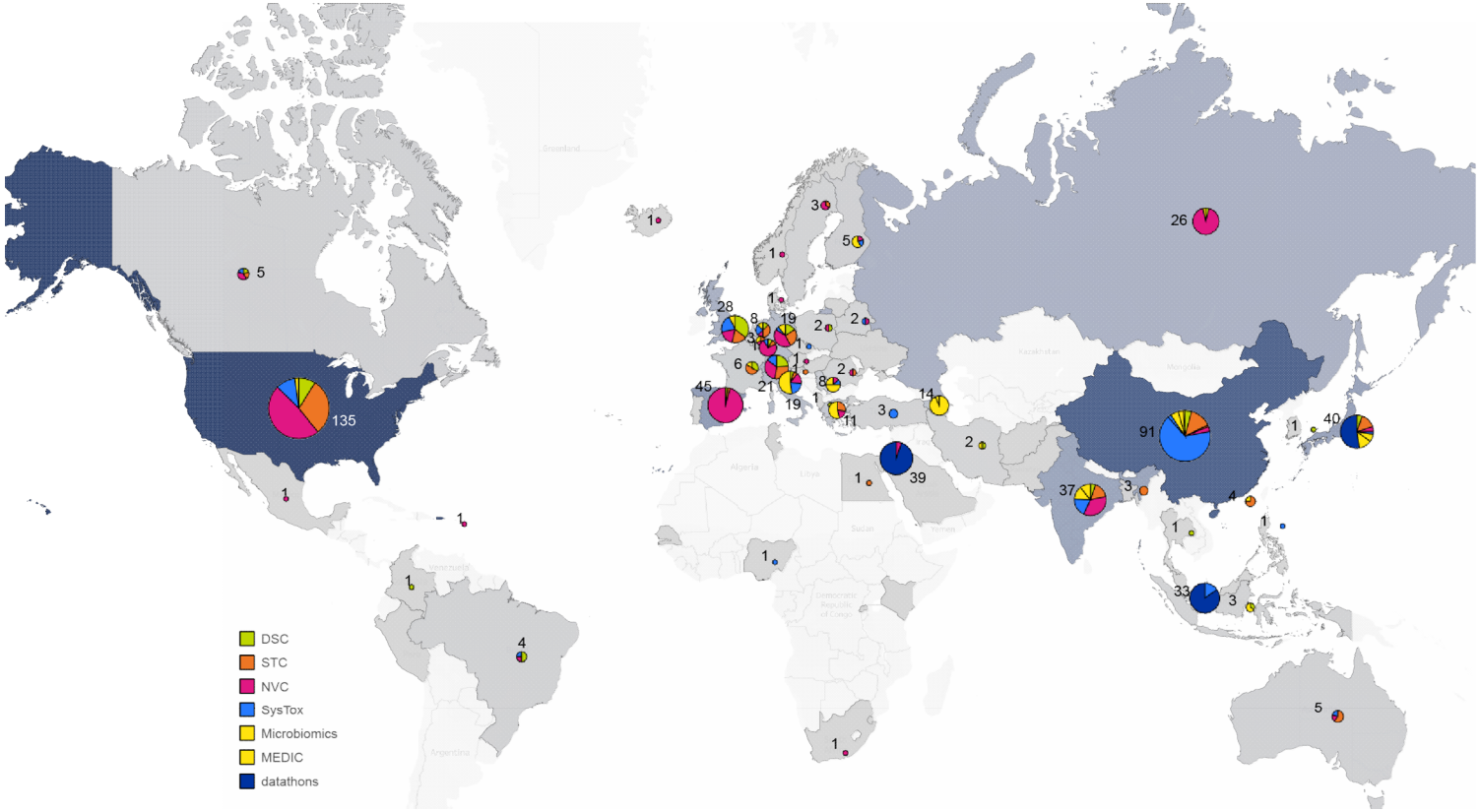
2016

SBV IMPROVER DATATHON - SINGAPORE

Omics datasets and functional measurements from a 7-month inhalation toxicology study were provided to encourage scientists creating applications to analyze the datasets that could be later on included into Garuda.

Learn more

Participation map – All challenges and datathons



THE METAGENOMICS DIAGNOSIS FOR IBD CHALLENGE (MEDIC)

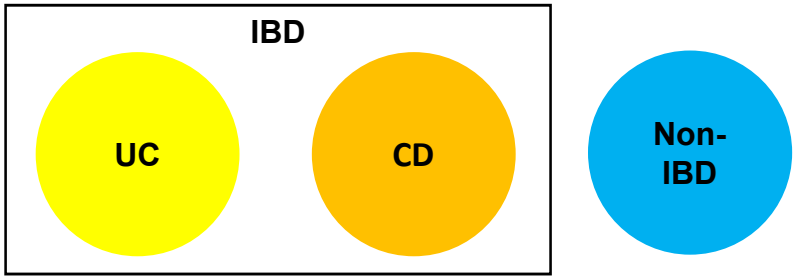
Aim – MEDIC

The challenge aims to **investigate the diagnostic potential of metagenomics data**

1) to classify IBD patients and non-IBD subjects
2) within the IBD category, to attempt to classify subjects with ulcerative colitis (UC) and Crohn's disease (CD)

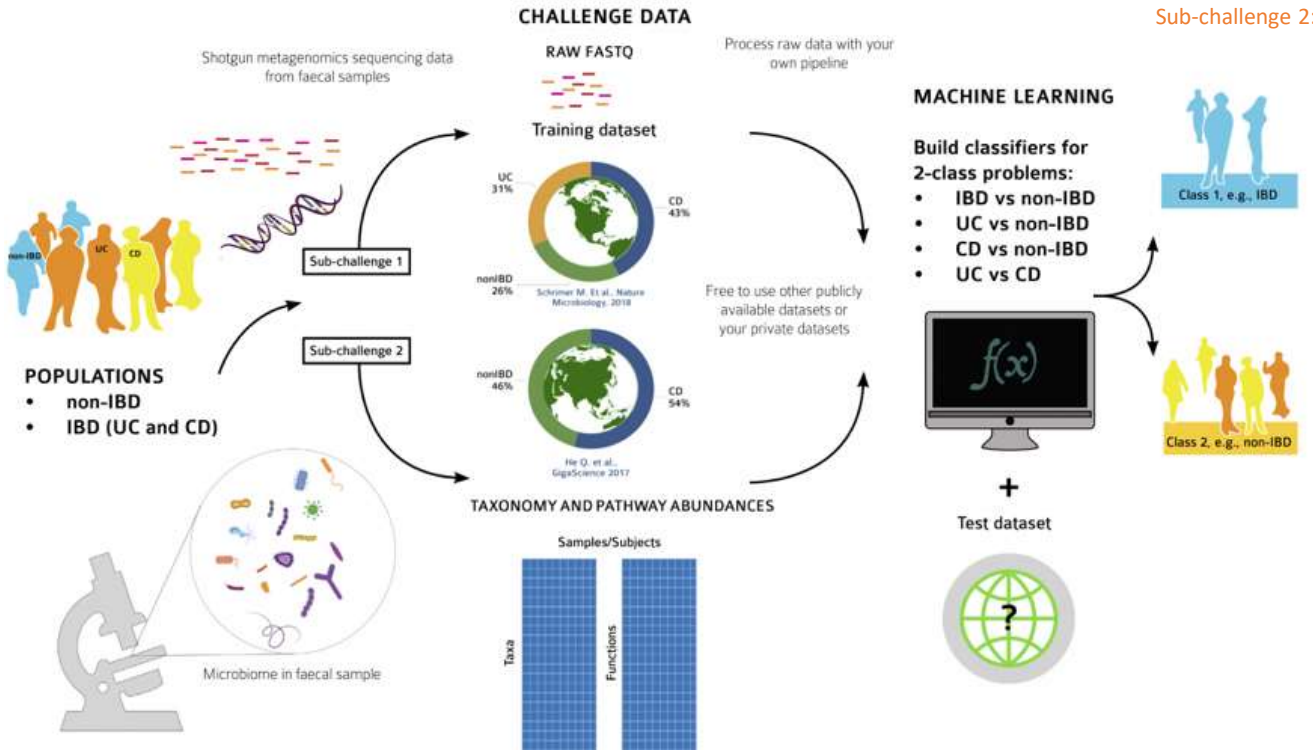
More specifically, the challenge poses four 2-class problems

- IBD vs non-IBD
- UC vs non-IBD
- CD vs non-IBD
- UC vs CD



The challenge structure

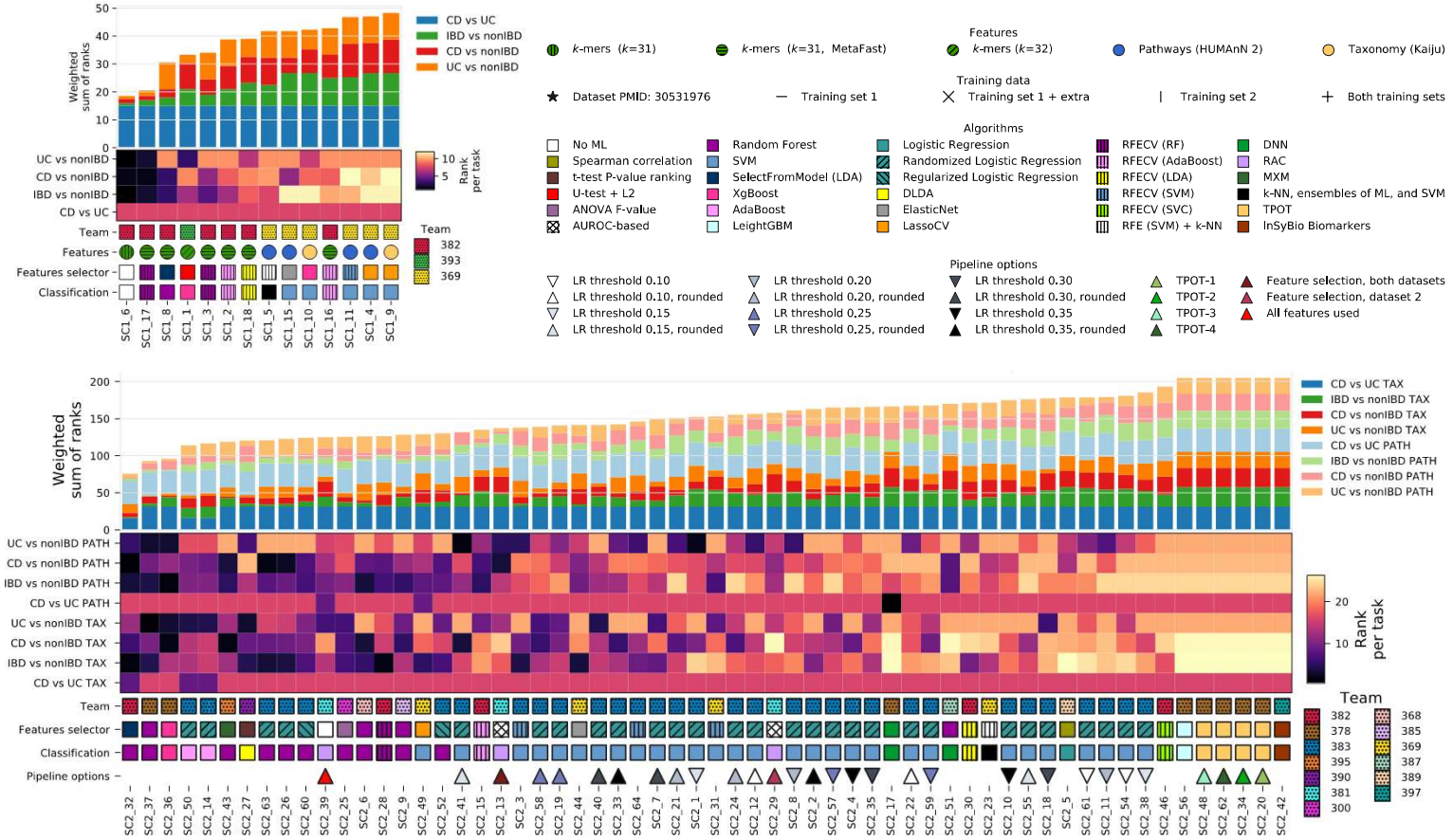
Sub-challenge 1: Requires access to HPC
Sub-challenge 2: Can be completed on a PC



Participants could choose to solve **either one or both sub-challenges.**

Submissions summary

- SC1: **14 submissions** from **3 teams**
- SC2**: 60 submissions** from **13 teams**



CD – Crohn's Disease
IBD – Inflammatory bowel disease
UC – Ulcerative colitis

ML – Machine learning
LDA – Linear discriminant analysis
RF – Random forest
SVM – Support vector machine
k-NN – k-nearest neighbours
SVC – Support vector classifier
DNN – Deep neural networks
LR – Logistic regression

Teams ranking

SUBCHALLENGE 1: MEDIC-RAW

Overall subchallenge 1	IBD vs. non-IBD	CD vs. non-IBD	UC vs. non-IBD
1. CTLAB@ITMO	1. CTLAB@ITMO	1. CTLAB@ITMO	1. CTLAB@ITMO
2. GG-GUMC	2. GG-GUMC	2. CDS-Lab	2. GG-GUMC
3. CDS-Lab	3. CDS-Lab	3. GG-GUMC	3. CDS-Lab

SUBCHALLENGE 2: MEDIC-PROCESSED

Overall subchallenge 2	IBD vs. non-IBD - Taxonomy	CD vs. non-IBD - Taxonomy	UC vs. non-IBD - Taxonomy
1. CTLAB@ITMO	1. CTLAB@ITMO	1. mignon	1. mignon
2. mignon	2. GiGi	2. Gnosis DA	2. JC
3. GiGi	3. CDS-Lab	3. GiGi	3. GiGi

Thank you!

Website: sbvimprover.com

Questions? Contact Us

sbvimprover.RD@pmi.com

The sbv IMPROVER project, the websites, and the symposia are part of a collaborative project designed to enable scientists to learn about and contribute to the development of a new crowdsourcing method for verification of scientific data and results. The project is led and funded by Philip Morris International.

For more information on the focus of Philip Morris International's research, please visit www.pmiscience.com.