sbv IMPROVER
SYSTEMS BIOLOGY VERIFICATION

www.sbvimprover.com

The Systems Toxicology Computational Challenge:

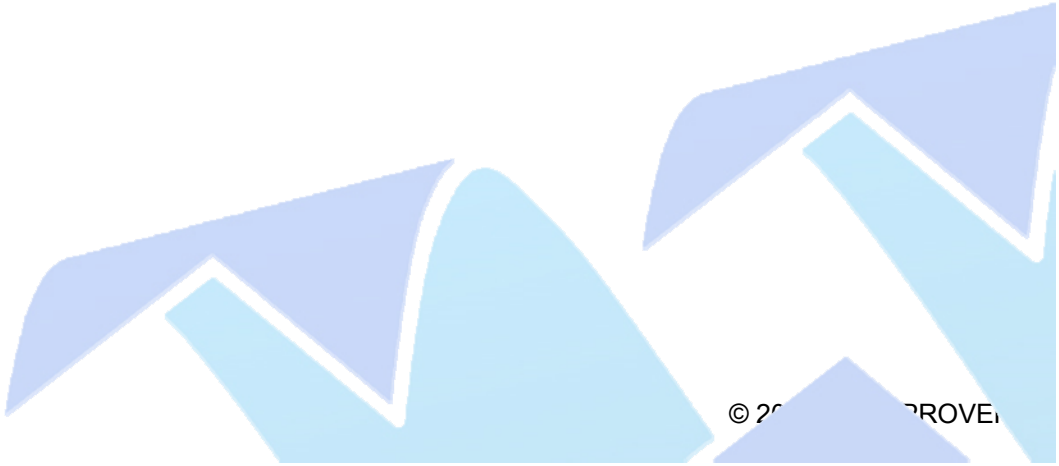
Marker of Exposure Response Identification

Overview

Carine Poussin, PhD

July 11th 2016, ISMB, Orlando

sbv IMPROVER in a Nutshell

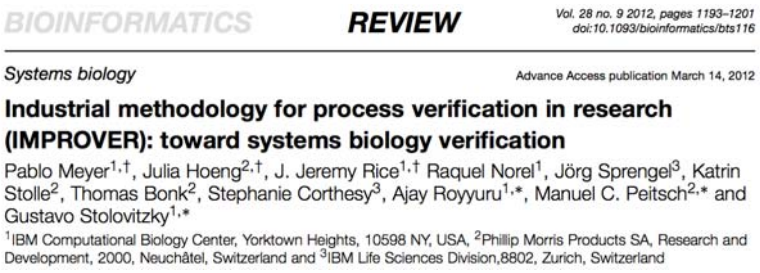


IMPROVER: Industrial Methodology for Process Verification in Research

Project initiated 6 years ago and funded by Philip Morris International R&D

Aims to verify methods & data in systems biology / toxicology

Verifies individual methods using double blind performance assessment



BIOINFORMATICS **REVIEW** Vol. 28 no. 9 2012, pages 1193–1201 doi:10.1093/bioinformatics/bts116

Systems biology Advance Access publication March 14, 2012

Industrial methodology for process verification in research (IMPROVER): toward systems biology verification

Pablo Meyer[1,†], Julia Hoeng[2,†], J. Jeremy Rice[1,†] Raquel Norel[1], Jörg Sprengel[3], Katrin Stolle[2], Thomas Bonk[2], Stephanie Corthesy[3], Ajay Royyuru[1,*], Manuel C. Peitsch[2,*] and Gustavo Stolovitzky[1,*]

[1]IBM Computational Biology Center, Yorktown Heights, 10598 NY, USA, [2]Phillip Morris Products SA, Research and Development, 2000, Neuchâtel, Switzerland and [3]IBM Life Sciences Division,8802, Zurich, Switzerland

Bioinformatics 2012 28(9):1193-1201

_computational BIOLOGY COMMENTARY

Verification of systems biology research in the age of collaborative competition

Pablo Meyer[1], Leonidas G Alexopoulos[2], Thomas Bonk[3], Andrea Califano[4], Carolyn R Cho[5], Alberto de la Fuente[6], David de Graaf[7], Alexander J Hartemink[8], Julia Hoeng[3], Nikolai V Ivanov[3], Heinz Koeppl[9], Rune Linding[10], Daniel Marbach[11], Raquel Norel[1], Manuel C Peitsch[3], J Jeremy Rice[1], Ajay Royyuru[1], Frank Schacherer[12], Joerg Sprengel[13], Katrin Stolle[3], Dennis Vitkup[4] & Gustavo Stolovitzky[1]

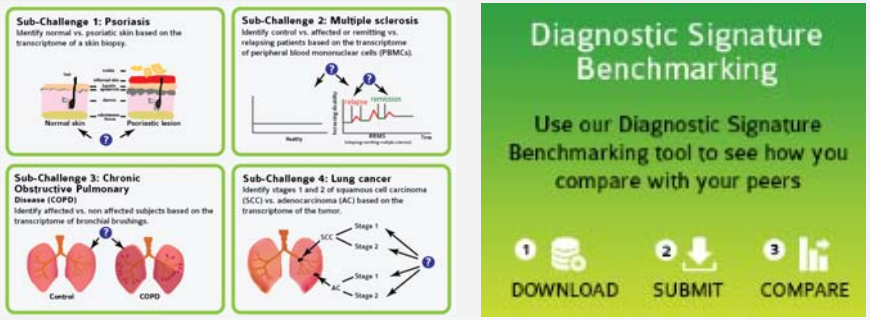
Collaborative competitions in which communities of researchers compete to solve challenges may facilitate more rigorous scrutiny of scientific results.

Nature Biotechnology 2011 Sep 8;29(9):811-5

Previous sbv IMPROVER crowd-sourcing challenges

1. Diagnostic Signature Challenge (2012)

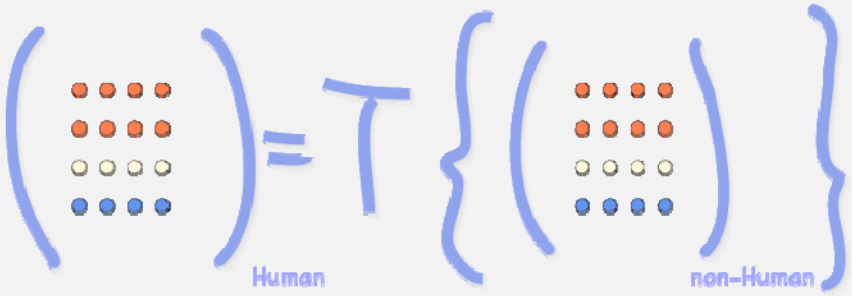
The identification of gene expression signatures and computational methods for diagnostic classification



https://sbvimprover.com/challenge-1

2. Species Translation Challenge (2013)

The translatability of biological system perturbations across species



3. Network Verification Challenge (2014-2015)

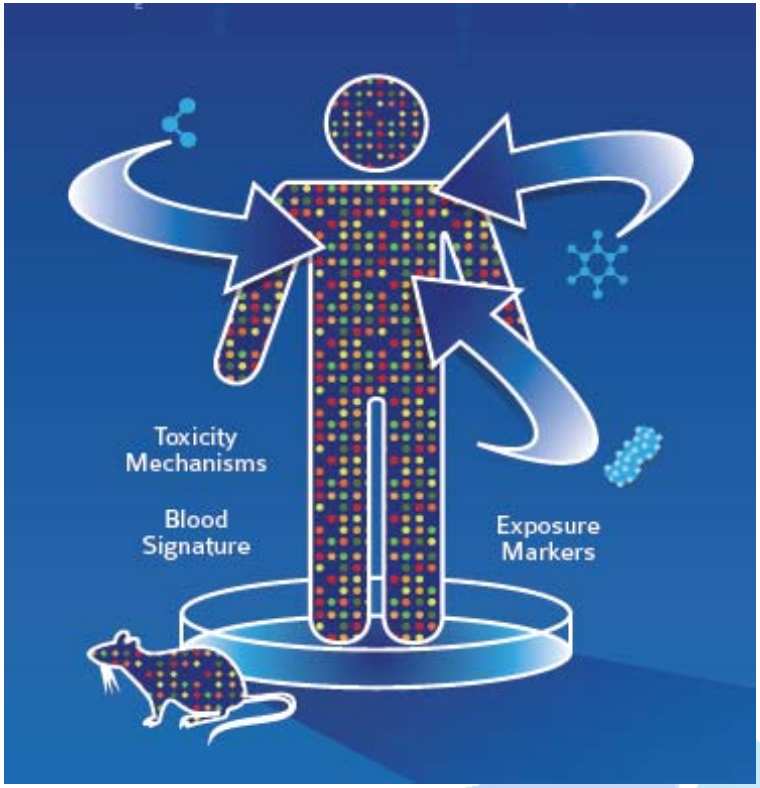
The verification and enhancement of biological causal networks representative of various biological processes

Challenge outcome including best performing methods and lessons learned were:

- Presented in symposium
- Published in peer-reviewed journals

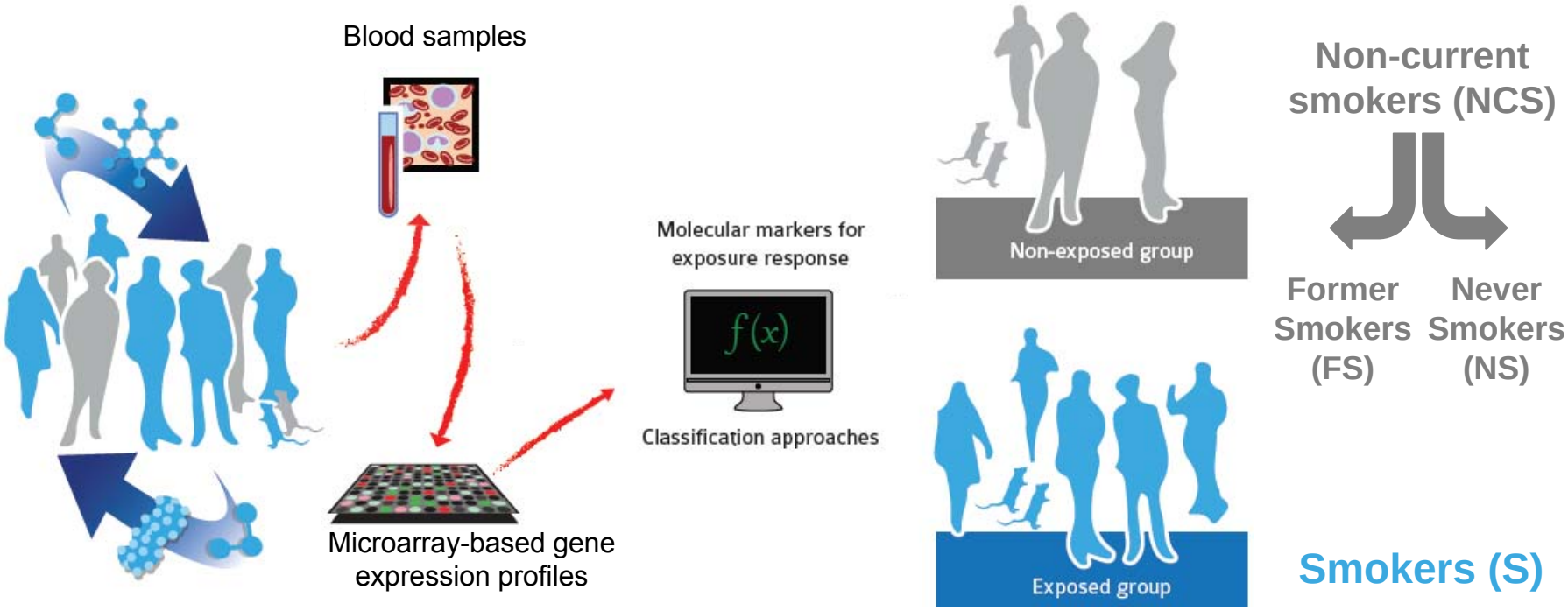
https://sbvimprover.com/sbv-improver/publications/

2016 - The Systems Toxicology Computational Challenge



https://sbvimprover.com/challenge-4

Overview



To develop a classification approach that identifies a blood gene signature capable of associating subjects to the correct exposure group

The Systems Toxicology Computational Challenge
Marker of Exposure Response Identification

Goal: to develop **inductive** blood gene signature-based classification models to predict **smoking exposure** or **cessation status**

Sub-challenge 1 (SC1)
Human blood signature as exposure response marker

⇩

Human signatures

Sub-challenge 2 (SC2)
Species translatable blood gene signature as exposure response marker

⇩

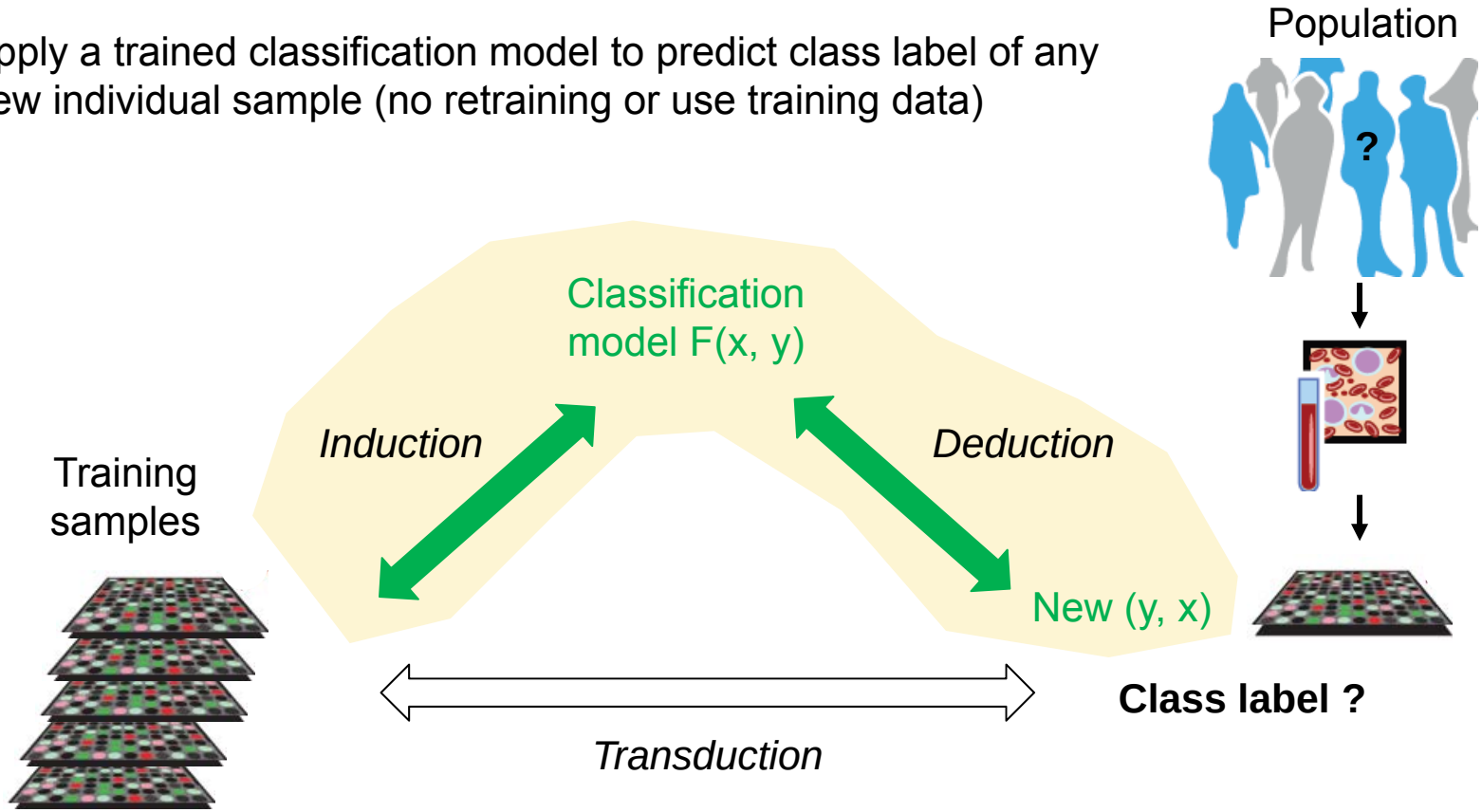
Species-independent signatures

Robust and **sparse** gene signatures → do not exceed 40 genes

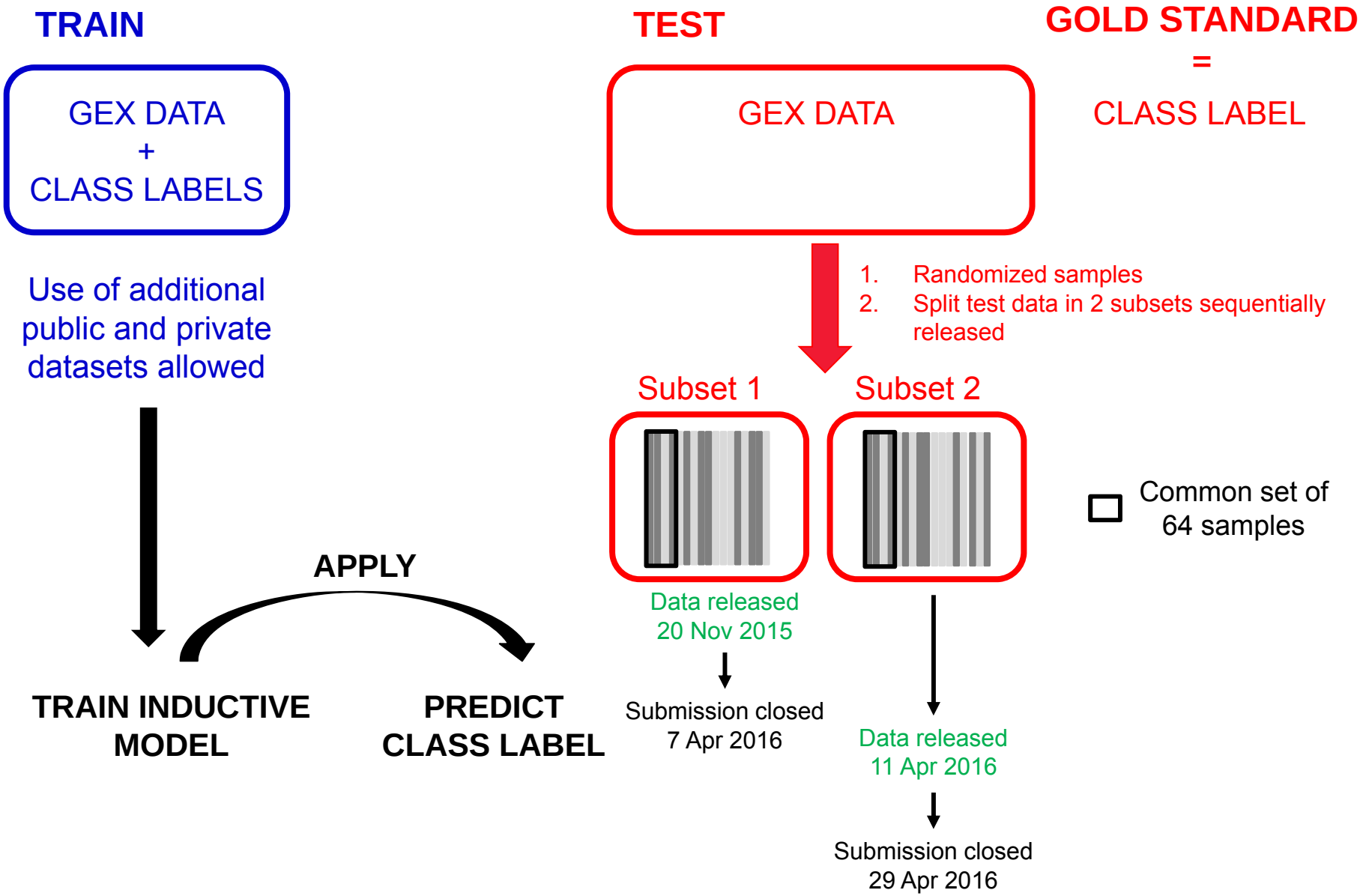
Classification models

Inductive classification model as opposed to transductive

Apply a trained classification model to predict class label of any new individual sample (no retraining or use training data)

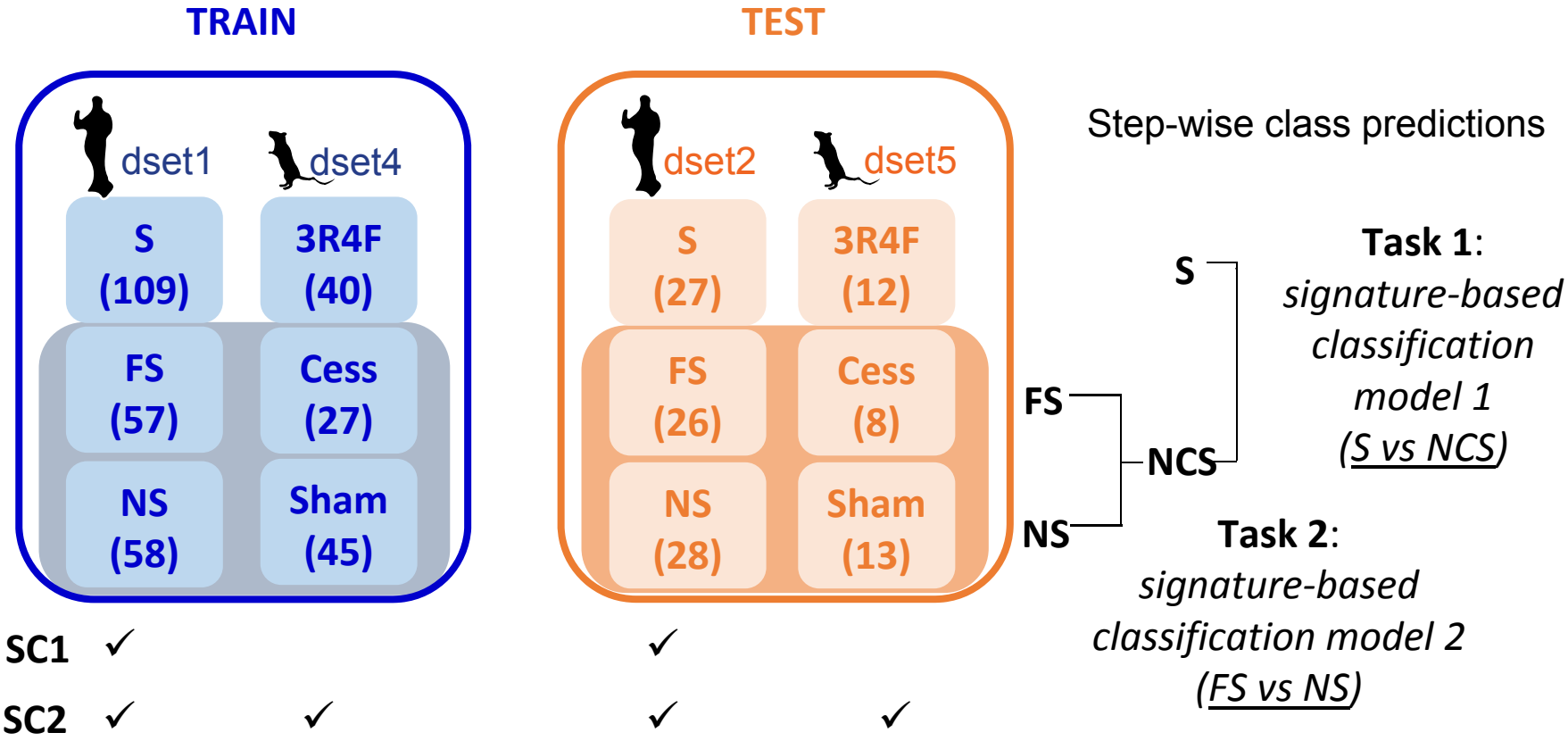


Training and Test data



Study datasets

Gene expression data generated from human and mouse blood samples



Step-wise class predictions

S
FS
NS
NCS

Task 1: *signature-based classification model 1 (S vs NCS)*

Task 2: *signature-based classification model 2 (FS vs NS)*

S/3R4F: Smokers / 3R4F (exposure to smoke from a reference cigarette)
FS/Cess: Former smokers / Cessation
NS/Sham: Never smokers / Sham
NCS: Non-current smoker

Freedom to use two separate models for 2-class prediction for each step, or directly a 3-class prediction model

https://sbvimprover.com/challenge-4/the-computational-challenge/hbs

Submissions and Timelines

Submissions for each sub-challenge

- 4 class label prediction files
- 2 gene lists
- 1 write up

Sample ID	Smoker	Non-current smoker
Sample 1	P1	P2
Sample 2	0.95	0.05
Sample 3	0.94	0.04
….		
Sample M	0.85	0.15

Sample ID	Former Smoker	Non-current smoker
Sample 1	P1	P2
Sample 2	0.95	0.05
Sample 3	0.94	0.04
….		
Sample M	0.85	0.15

- P_x= confidence value that a sample belongs to a class
- P1+P2=1 ; P1≠P2