The Systems Toxicology Computational Challenge: Markers of Exposure Response Identification – Insights gained

Vincenzo Belcastro, Carine Poussin, Stéphanie Boué, Yang Xiang, Florian Martin, Manuel C. Peitsch, Julia Hoeng and SBV IMPROVER challenge participants

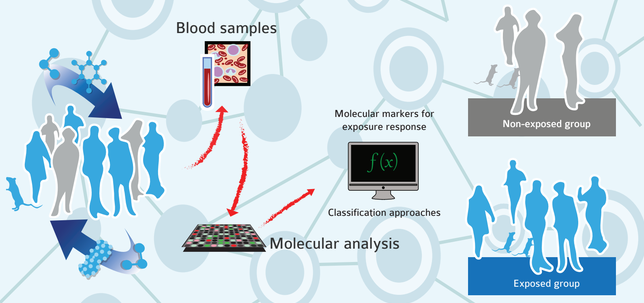
PMI R&D, Philip Morris Products S.A. Quai Jeanrenaud 5, 2000 Neuchatel, (part of Philip Morris International group of companies).

sbv IMPROVER
SYSTEMS BIOLOGY VERIFICATION

Markers of Exposure Response Identification

Background

Humans have evolved to respond to diverse environmental conditions including various chemical, mechanical, and pathogenic insults. Among these, exposure to chemicals (e.g., cigarette smoke, pollutants, pesticides) induce molecular changes in cells. A subset of exogenous chemicals, chemical-derived metabolites, and endogenous molecules produced by exposed organs (e.g., lung, gut) can pass into the blood stream and induce molecular changes in blood cells. While whole blood is an easily accessible matrix, its analysis is challenging owing to its compositional complexity. Moreover, most preclinical experiments are performed in animal models which raises a question of translatability and relevance of findings to humans.

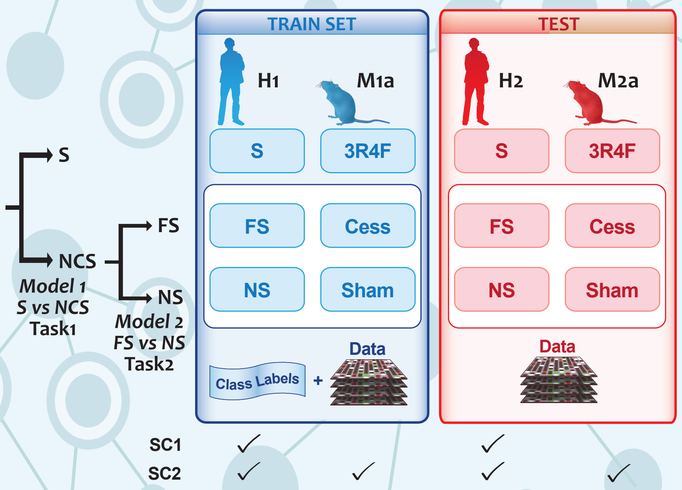


Goal

●To develop blood gene signature-based classification models to predict smoke exposure status or cessation status:

- **in human (Sub-challenge 1 - SC1)**
- **translatable across species (Sub-challenge 2 - SC2)**

Gene expression data generated from human and mouse blood samples



H1/H2: Human samples for training (H1) and for test (H2) datasets
M1a/M2a: Mouse samples for training (M1a) and for test (M2a) datasets

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

Timelines

- The challenge was open from November 2015 to April 2016 (5 months).

Rules

- To identify **robust** and **sparse** signatures (40 genes max).
- To develop **inductive** classification models that allow to predict class label of any new sample **without retraining** (unlike transductive models).

Scoring

- Submissions from participants were anonymized prior to scoring.
- Matthews correlation coefficient (**MCC**) and Area Under Precision Recall (**AUPR**) curve were metrics used to assign ranks to the team.
- Aggregated Score Ranks used to assess overall team performance.
- Results and final ranking presented to and approved by an external and independent Scoring Review Panel.

Key references

sbv IMPROVER
- Meyer et al. Verification of systems biology research in the age of collaborative competition. Nat Biotechnol, 2011
- Meyer et al. Industrial methodology for process verification in research (IMPROVER): toward systems biology verification. Bioinformatics, 2012

Diagnostic signature challenge
- Tarca et al. Strengths and limitations of microarray-based phenotype prediction: lessons learned from the IMPROVER Diagnostic Signature Challenge. Bioinformatics, 2013

Blood signatures of smoke exposure
- Martin, et al. Identification of gene expression signature for cigarette smoke exposure response--from man to mouse. Human & experimental toxicology, 2015
- Beineke et al. A whole blood gene expression-based signature for smoking status. BMC medical genomics, 2012

Studies
- *H1*: The Queen Ann Street Medical Center (QASMC) clinical study - a non-interventional, observational, case–control study conducted in the UK (NCT01780298)
- *M1a*: Phillips, B. et al. A 7-month cigarette smoke inhalation study in C57BL/6 mice demonstrates reduced lung inflammation and emphysema following smoking cessation or aerosol exposure from a prototypic modified risk tobacco product. Food and Chemical Toxicology, 2015

Key Conclusions

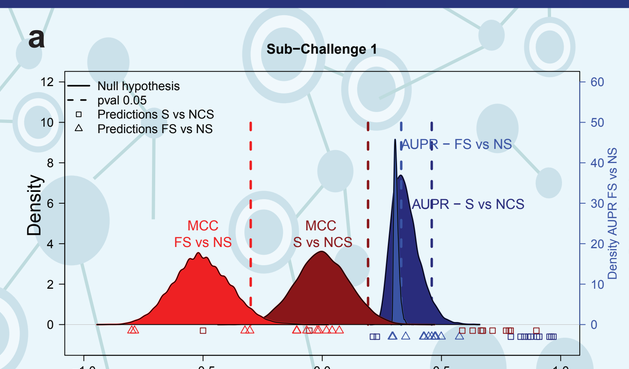
- Successful **worldwide participation** to the challenge.
- Gene expression changes measured in blood are **informative of exposure status**. Prediction of **smoking exposure status** (S vs NCS) is possible whereas prediction of **cessation status** (FS vs NCS) is more challenging.
- Best performers used **Random Forest** and **Linear Discriminant Analysis** as machine learning methods.
- Participants succeeded in development of **inductive** classification **models**.
- Confidence score aggregation improves performances

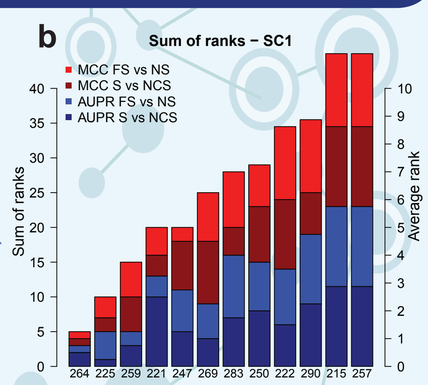
Participation and Results of the Challenge

Participation

Worldwide participation in the SBV IMPROVER SysTox computational challenge.
61 teams (grey dots) comprising of 135 participants registered to the challenge. Among those, 23 teams submitted predictions for at least one of the sub challenges. 12 teams for SC1 (out of 23) and 6 teams for SC2 (out of 15) submitted predictions that complied with all the rules and were qualified for scoring (blue dots). The top 3 performers were teams 264, 225, 259 and 219, 250 and 264 for SC1 and SC2 respectively (blue stars).

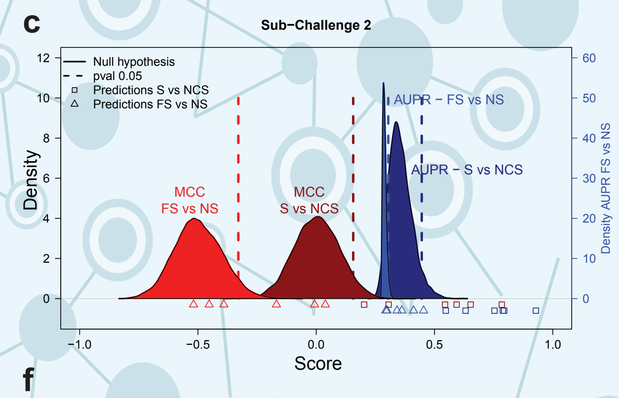
Participants' prediction MCC or AUPR scores and the sum of scores across the two metrics and tasks.
(a,c) Participants' scores (squares and triangles) relative to the null score distribution (in red and blue) calculated from 10'000 random predictions. triangles and squares: MCC and AUPR scores computed from participants' predictions; dotted lines: score corresponding to a P-value of 0.05 (threshold for significance of participants' prediction scores). (b,d) Barplots reporting the sum of ranks across all metrics and tasks for all the teams. Lower sum of ranks implies better performance. (e,f) Scores of top 3 best performing teams for each metrics and task; all scores were significant (p-value <= 0.05).

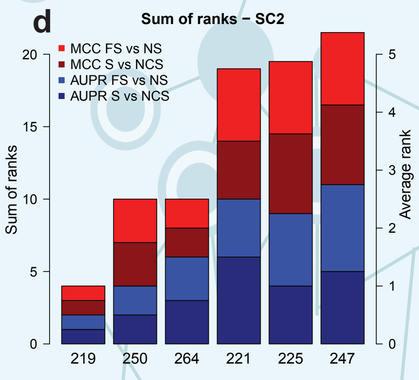




e

Team	Team Rank	AUPR S vs NCS	AUPR FS vs NS	MCC S vs NCS	MCC FS vs NS
264	1	0.96	0.58	0.90	0.07
225	2	0.97	0.50	0.77	0.02
259	3	0.95	0.47	0.79	-0.02



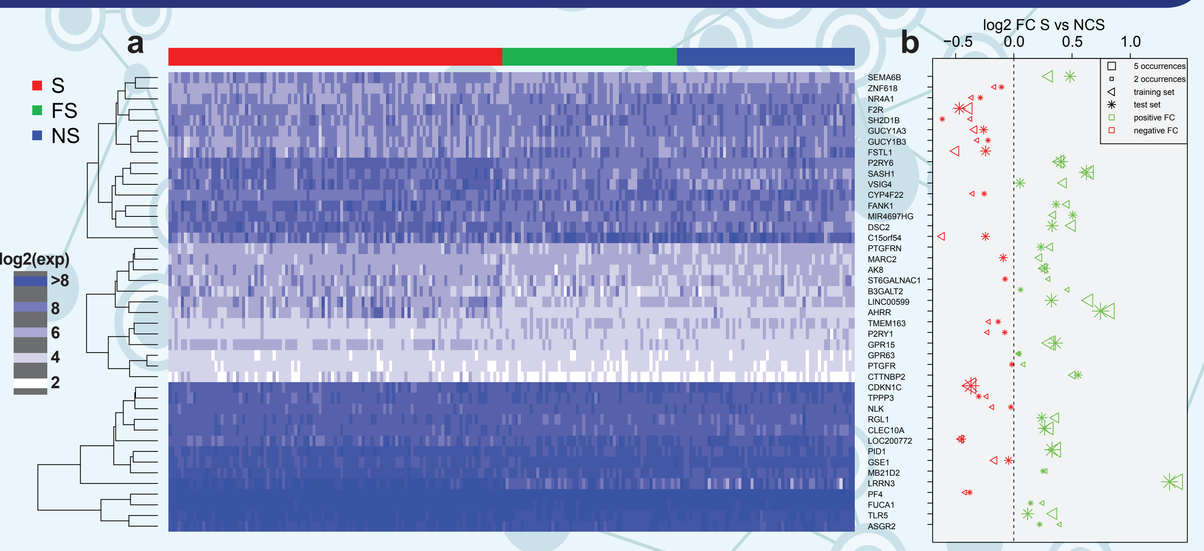


f

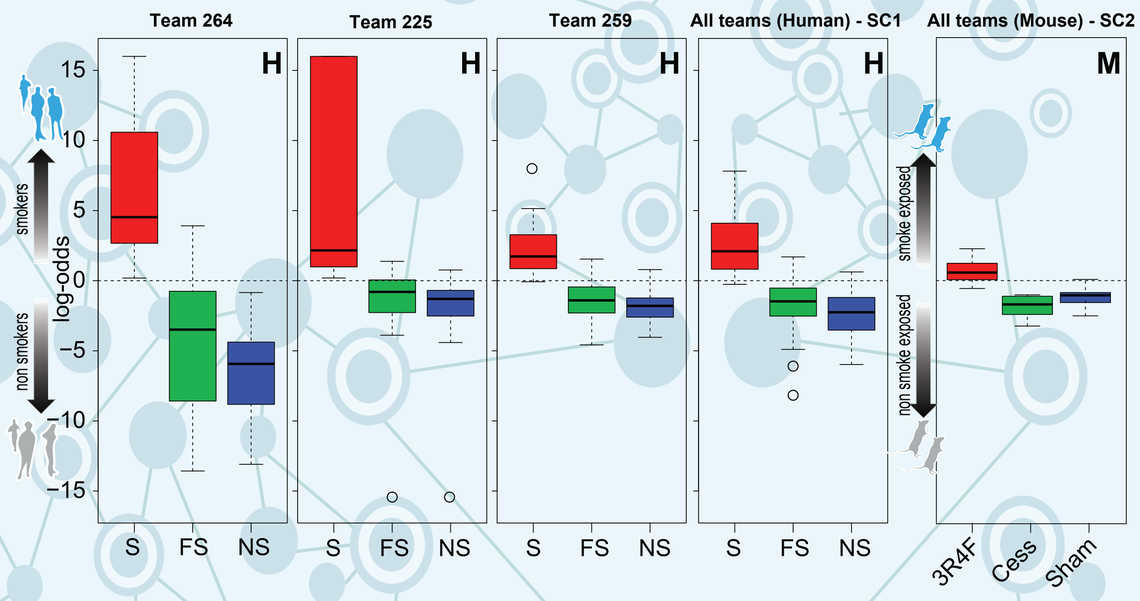
Team	Team Rank	AUPR S vs NCS	AUPR FS vs NS	MCC S vs NCS	MCC FS vs NS
219	1	0.93	0.45	0.78	0.04
250	2	0.79	0.36	0.65	-0.17
264	2	0.79	0.41	0.59	-0.01

Post-challenge analysis

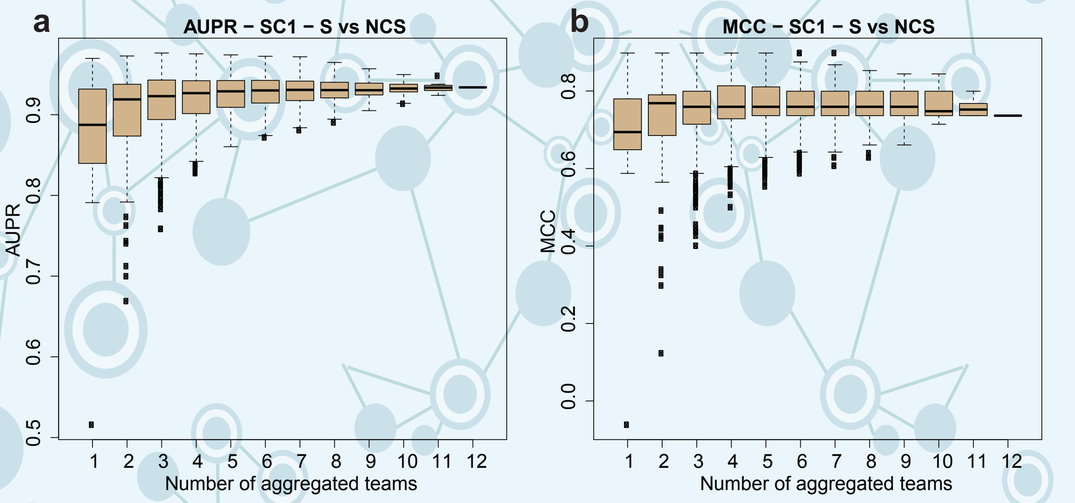
Human smoking exposure gene signature
(a) Heatmap of expression values (training set) for genes selected as exposure response markers by at least two teams.
(b) Plot of fold changes (FCs). FCs were computed as differential expression values between the Smoker group and the Former and Never Smokers groups combined. The size of the symbols is proportional to the number of times the gene is selected.



Human (H) and Mouse (M) blood sample class predictions by the crowd.
Boxplot showing the distribution of the log-odds (and median log-odds for "All teams") values for all samples of each class: Smoker (S)/3R4F, Former Smoker (FS)/Cess, Never Smoker (NS)/Sham.
Higher positive log-odds imply higher confidence that the sample belongs to the smoker/3R4F group. Lower negative log-odds imply higher confidence that the sample belongs to the non-current smoker (FS+NS/Sham) group.



Wisdom of crowd
Boxplot showing the distribution of the AUPR curve and MCC values for S vs NCS classification upon team aggregation.
AUPR and MCC values are reported on the y-axis of panels (a) and (b), respectively. Number of aggregated teams is reported on the x-axis. The barplot shows the AUPR (a) and MCC (b) distributions when confidence scores for all possible combinations of k teams are aggregated. Median performance tends to increase as the number of aggreated teams increases until a plateau is reached as observed in panel (a).



Acknowledgement: SBV IMPROVER Challenge participants

The sbv IMPROVER project, the website and the symposia are part of a collaborative project designed to enable scientists to learn about and contribute to the development of a new crowd sourcing method for verification of scientific data and results. The current challenges, website and biological network models were developed and are maintained as part of a collaboration among Philip Morris International, Selventa, OrangeBus, and ADS. 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.pmi.com.

www.sbvimprover.com