# The Systems Toxicology Challenge: Marker of Exposure Response Identification in Blood.

Carine Poussin, Vincenzo Belcastro, Stéphanie Boué, Akash Boda, Florian Martin, Alain Sewer, Bjoern Titz, Nikolai Ivanov, Manuel C. Peitsch and Julia Hoeng
Philip Morris International R&D, Neuchâtel, Switzerland

sbv IMPROVER — SYSTEMS BIOLOGY VERIFICATION

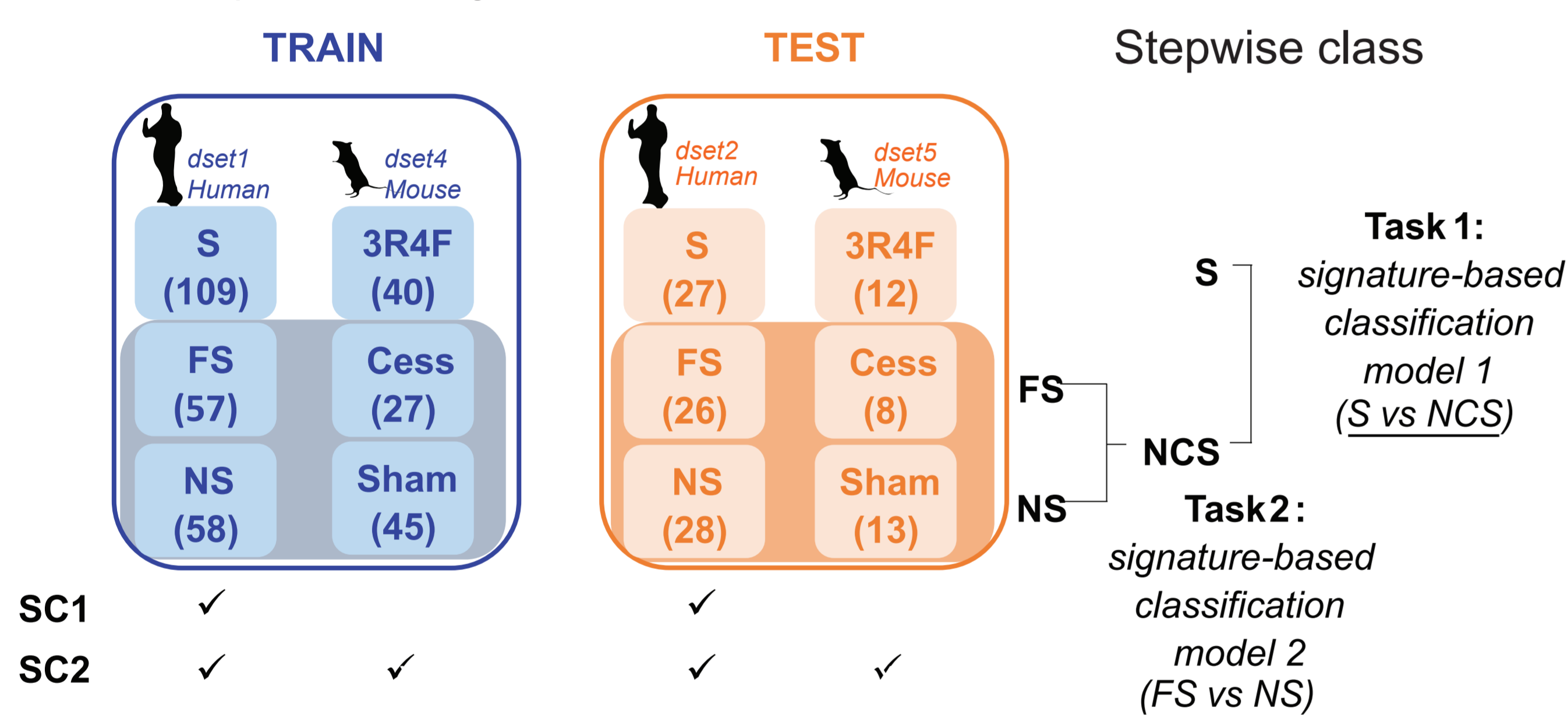## Markers of Exposure Response Identification

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 manifesting as a discrete exposure response fingerprint that can be examined. While whole blood is an easily accessible matrix, its analysis is challenging owing to its compositional complexity. Moreover, most preclinical in vivo experiments are performed in animals models which raises a question of translatability and relevance of findings to humans.

Non-current smokers (NCS)
Former Smokers (FS) — Never Smokers (NS)
Smokers (S)
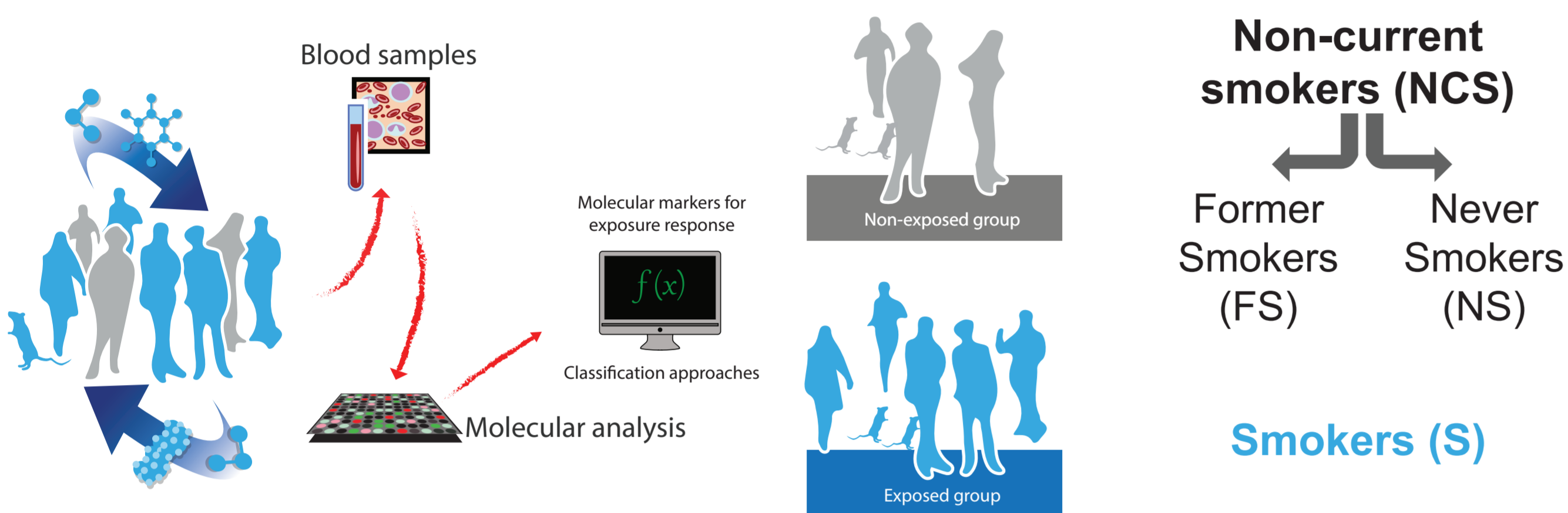
Goal

- 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

	TRAIN		TEST	
	dset1 Human	dset4 Mouse	dset2 Human	dset5 Mouse
S	S (109)	3R4F (40)	S (27)	3R4F (12)
FS	FS (57)	Cess (27)	FS (26)	Cess (8)
NS	NS (58)	Sham (45)	NS (28)	Sham (13)

SC1 ✓ ✓
SC2 ✓ ✓ ✓ ✓

Stepwise class

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 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 (in contrast to transductive models).

Scoring

- Submissions from participants were **anonymized** prior to scoring.
- Results and final ranking presented to and approved by an **external and independent Scoring Review Panel.**
- **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.

Participation and Results of the Challenge

Participation

Number of Team Members: 1 2 3 4 5

Submission:
- Registration but no submission
- Registration and submission
- ★ Best Performers

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).

Results

Sub-challenge 1 (SC1): panels a, b
Sub-challenge 2 (SC2): panels c, d
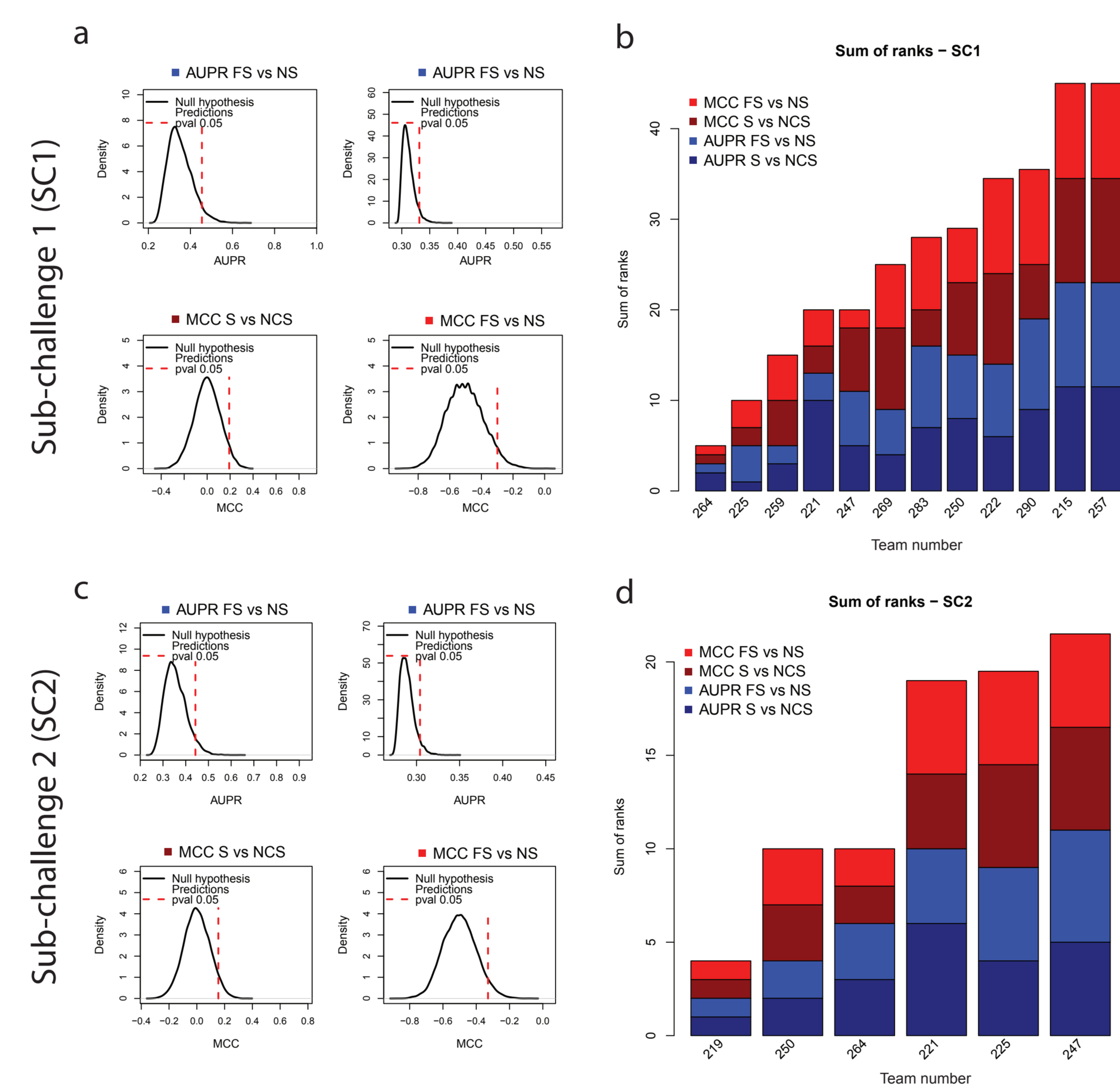
Team	Team Rank	AUPR S vs NCS	AUPR FS vs NS	MCC S vs NCS	MCC FS vs NS
264	1	0.96	0.58	0.90	0.07
225	2	0.97	0.50	0.77	0.02
259	3	0.95	0.47	0.79	-0.02

(panel e)

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

(panel f)

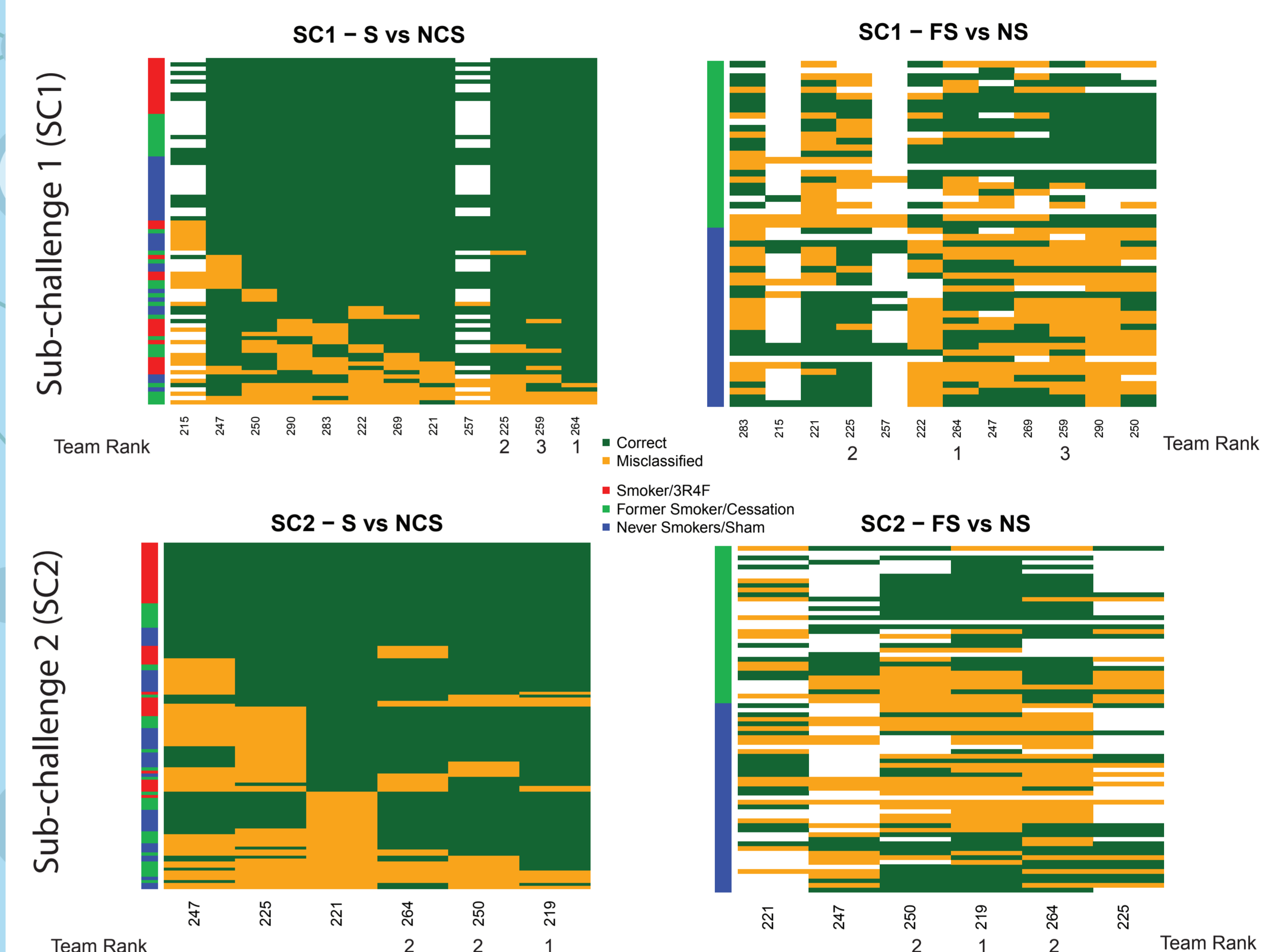**Participants' prediction MCC or AUPR scores and the sum of scores across the two metrics and tasks.**
(a,c) Participants' scores relative to the null score distribution (black curve) calculated from 10'000 random predictions. blue dot: MCC or AUPR scores computed from participants' predictions; red dotted line: 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).

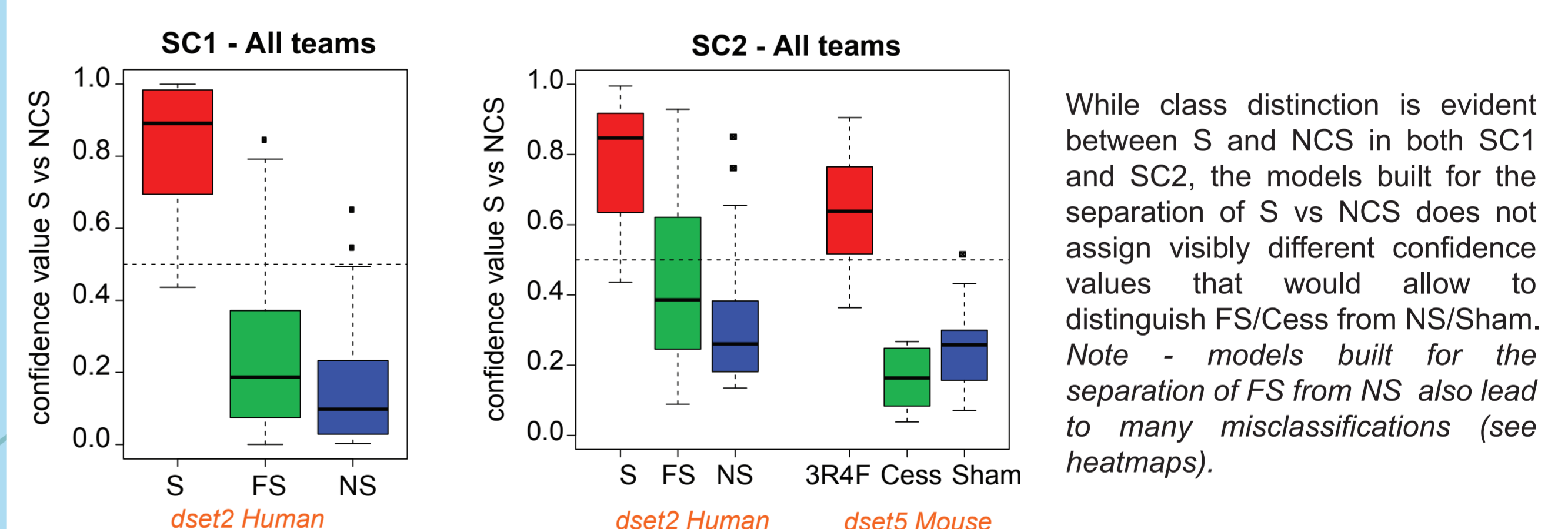## Post-Challenge Analysis

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**.
- Samples from **former smokers** tend to be **more frequently misclassified**.

Misclassification of samples across predictions

SC1 – S vs NCS / SC1 – FS vs NS
SC2 – S vs NCS / SC2 – FS vs NS

Legend:
- Correct
- Misclassified
- Smoker/3R4F
- Former Smoker/Cessation
- Never Smokers/Sham

HeatMap depicting subject misclassification by all teams in SC1 and SC2.
Each cell of the HeatMap shows if the corresponding team in the column, correctly (dark green) or incorrectly (orange) predicted the subject in the row. White cells represent subject-team pairs for which no predictions were available. The subjects are labelled on the left side according to their classes i.e. smokers in red, former smokers in light green, and never smokers in blue. Top 3 best performing teams per sub-challenge are indicated.

Median confidence value for all samples in S vs NCS (Task1) across all teams

SC1 - All teams (dset2 Human) / SC2 - All teams (dset2 Human, dset5 Mouse)

While class distinction is evident between S and NCS in both SC1 and SC2, the models built for the separation of S vs NCS does not assign visibly different confidence values that would allow to distinguish FS/Cess from NS/Sham. Note - models built for the separation of FS from NS also lead to many misclassifications (see heatmaps).

Boxplot showing the distribution of the median confidence value across all teams for all samples of each class: Smoker (S/3R4F), Former smoker (FS/Cess), Never smoker (NS/Sham))
A confidence value of 1 means full confidence that the sample is from a smoker. Red boxplots show the distribution of confidence values of samples from smokers while green and blue boxplots show confidence distribution values associated to samples from former and never smokers, respectively. Samples with confidence value > 0.5 are considered as smokers.

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
- dset1: The Queen Ann Street Medical Center (QASMC) clinical study - a non-interventional, observational, case–control study conducted in the UK (NCT01780298)
- dset5: Phillips, B. et al. An 8-Month Systems Toxicology Inhalation/Cessation Study in Apoe-/- Mice to Investigate Cardiovascular and Respiratory Exposure Effects of a Candidate Modified Risk Tobacco Product, THS 2.2, Compared With Conventional Cigarettes. Toxicological sciences, 2016
- dset4: 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

www.sbvimprover.com