



[www.sbvimprover.com](http://www.sbvimprover.com)



# sbv IMPROVER: Systems Biology Verification combined with Industrial Methodology for Process Verification in Research

*Carine Poussin, PhD*  
*Senior Scientist, Philip Morris International R&D*

ISMB – July 2018

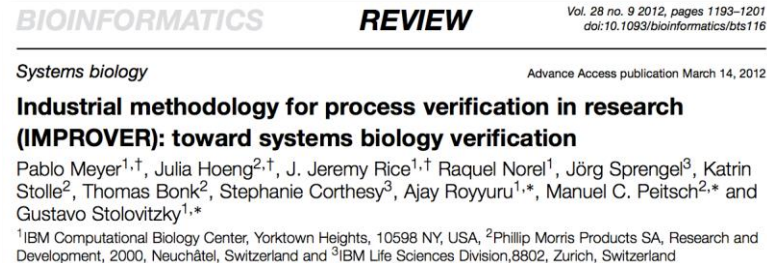
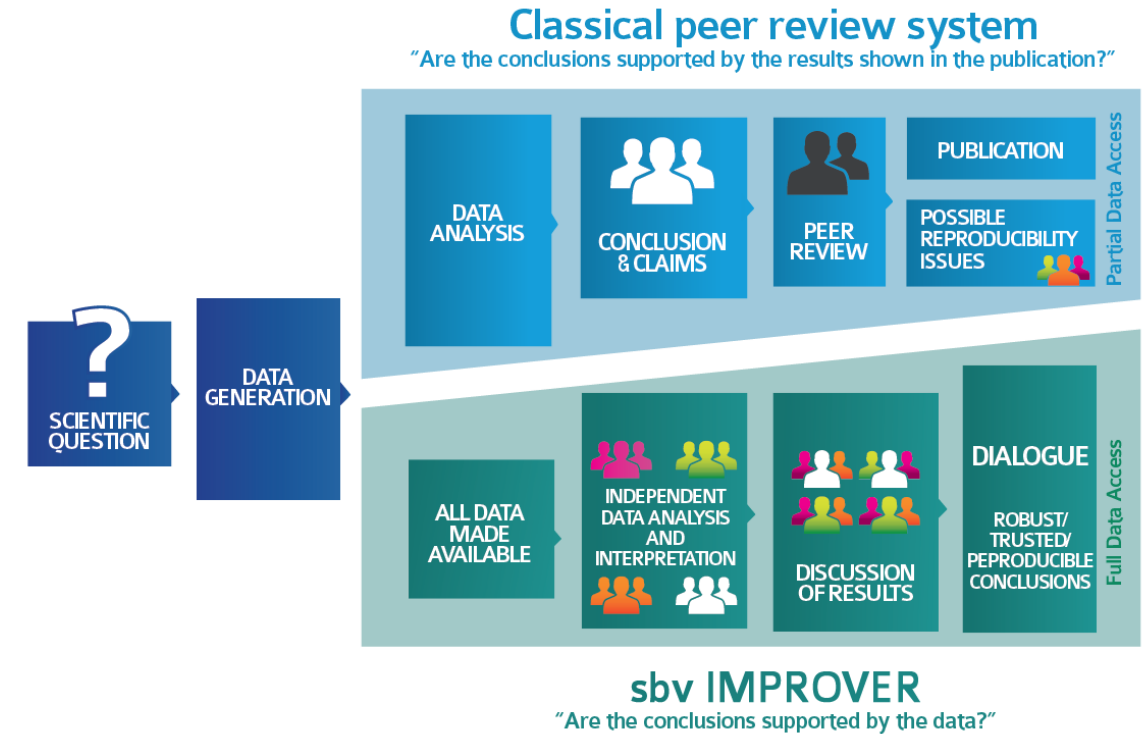
- What is sbv IMPROVER ?
- Example applications of sbv IMPROVER

# What is sbv IMPROVER?

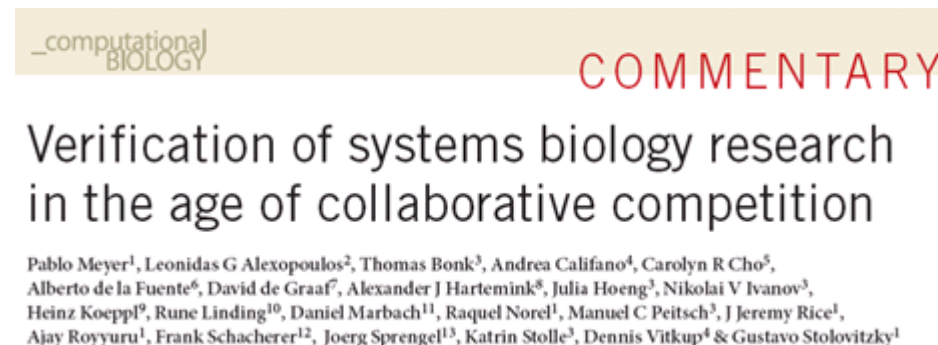
sbv IMPROVER stands for Systems Biology Verification combined with Industrial Methodology for Process Verification in Research.

This approach aims to provide a measure of quality control of industrial research and development by verifying the methods used.

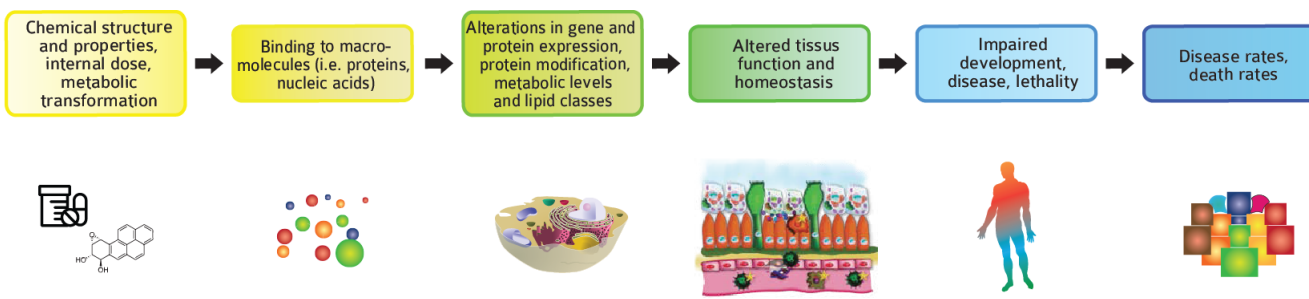
The sbv IMPROVER project is a collaborative effort led and funded by PMI Research and Development.



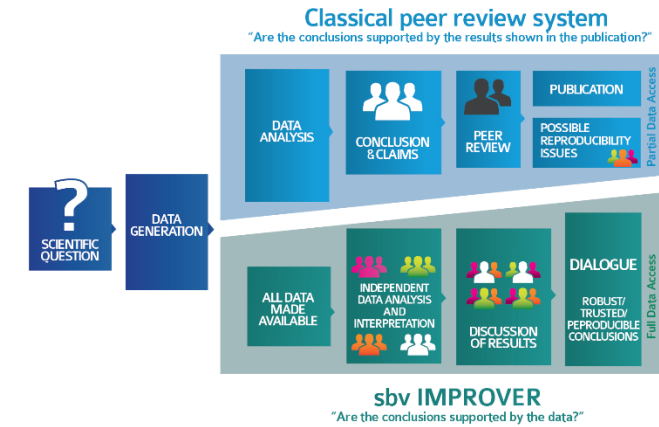
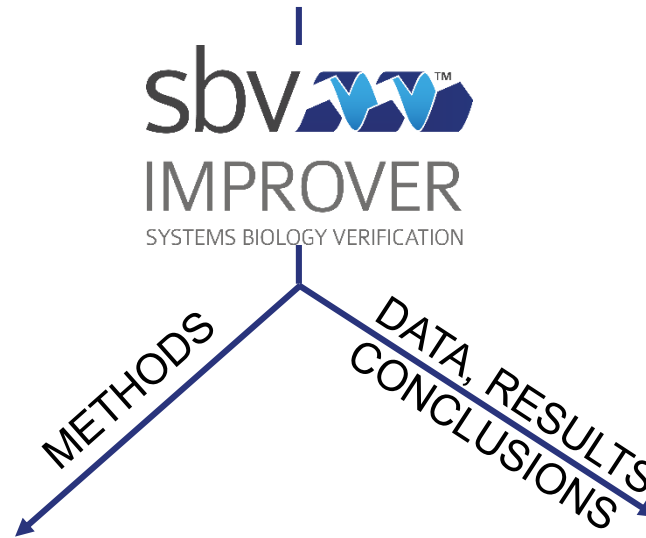
*Bioinformatics* 2012 28(9):1193-1201



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



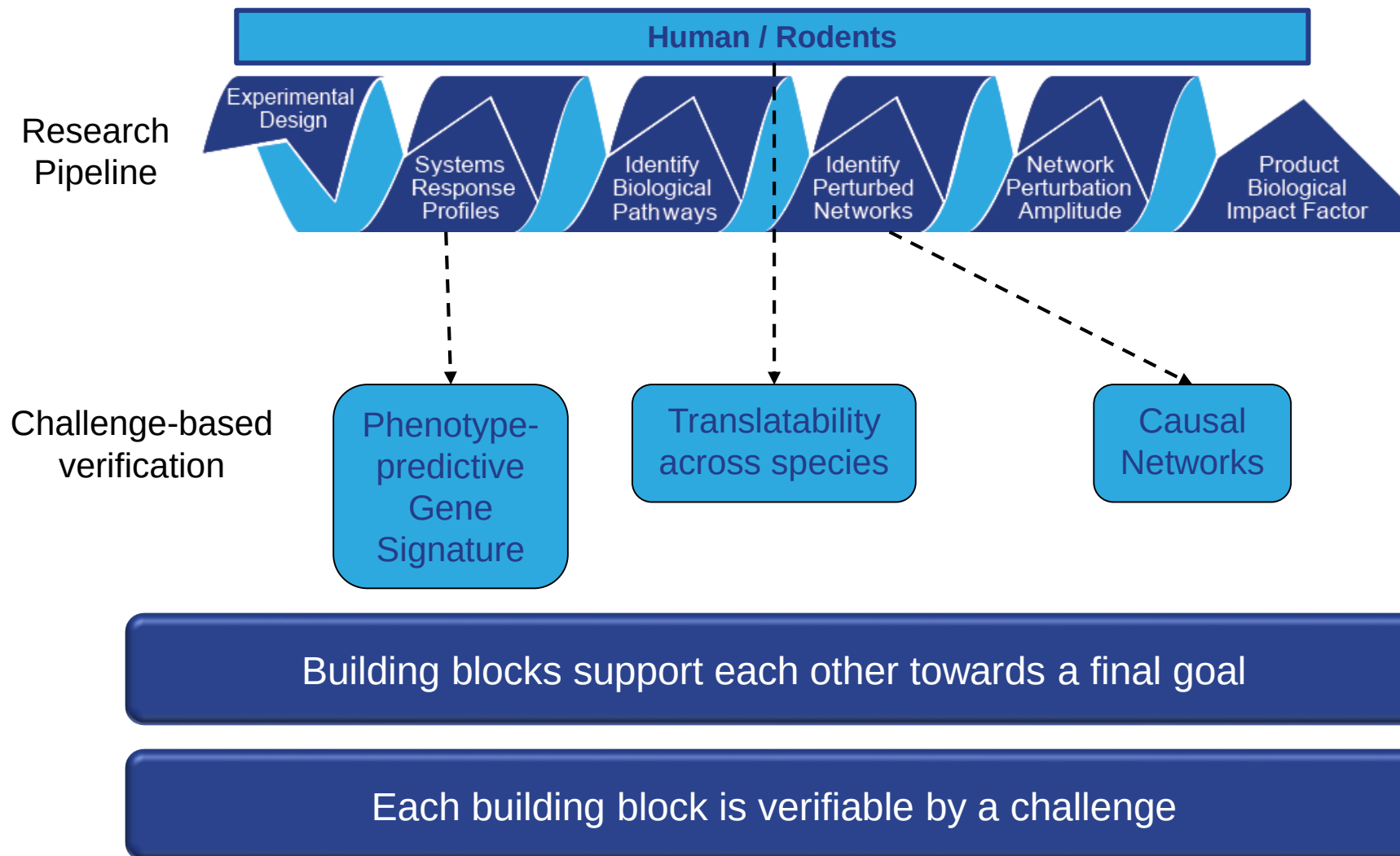
$$\pi \begin{smallmatrix} \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet \end{smallmatrix} \begin{smallmatrix} \neq \\ \pm \end{smallmatrix} \sqrt{f} x$$

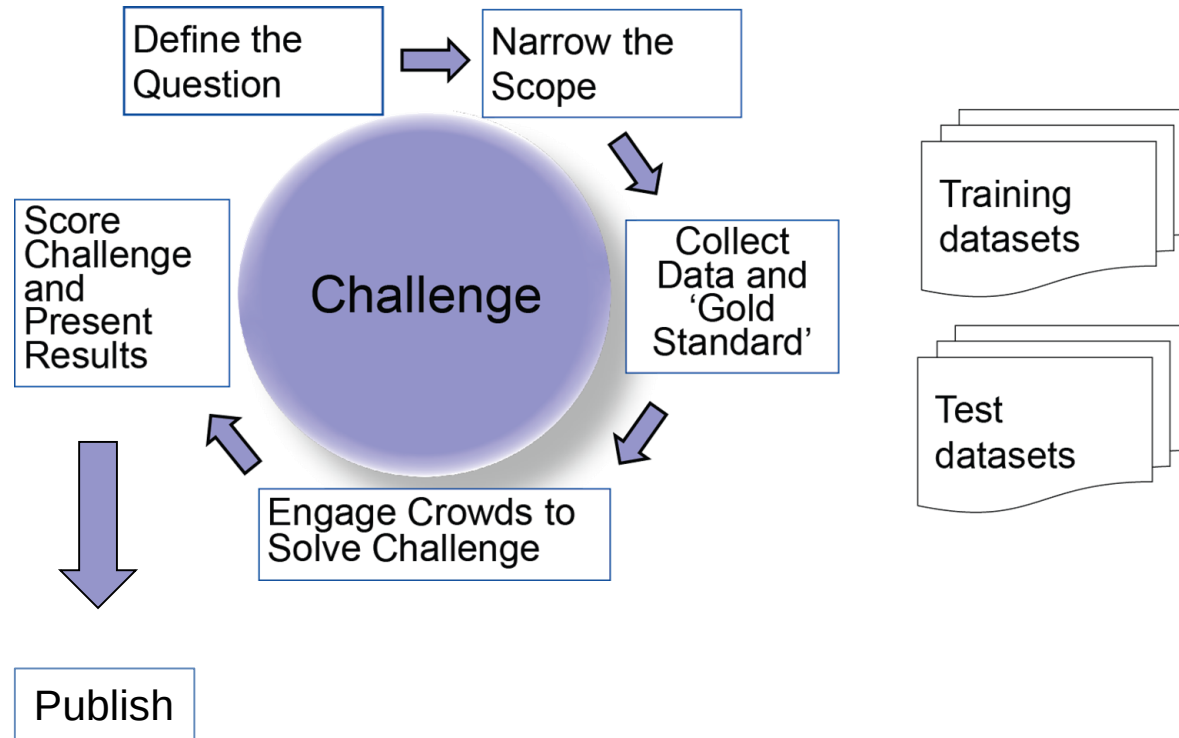


COMPUTATIONAL CHALLENGE

VERIFICATION

# Complex industrial research pipeline/workflow divided into verifiable building blocks





## Double Blind Performance Assessment

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



# sbv IMPROVER: engagement with the scientific community

## Engagement with potential participants during scientific conferences:

- Oral presentations
- Poster presentations
- Booth

## Webinars

## Posters and flyers

## Website

## Social media campaigns

## Ambassadors & Lectures



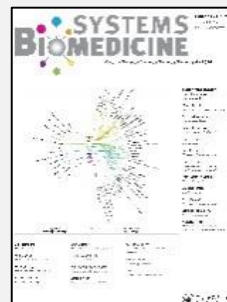
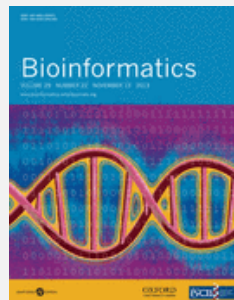
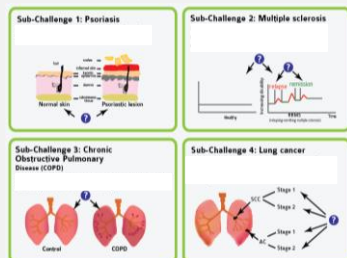


# Example applications of sbv IMPROVER

# Past sbv IMPROVER computational challenges

## Diagnostic signature challenge - DSC (2012)

To identify gene signatures for diagnostic classification in four disease area

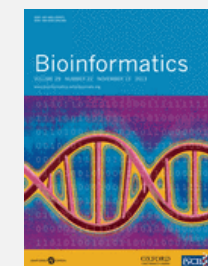


## Species translation challenge - STC (2013)

To identify and quantify a function of translatability of biological perturbations across human and rodent species

$$\begin{pmatrix} \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet \end{pmatrix} = T \left\{ \begin{pmatrix} \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet \end{pmatrix} \right\}$$

Human                      non-Human



## Network verification challenge – NVC (2014-2015)

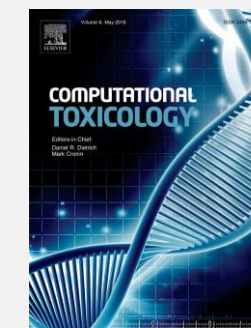
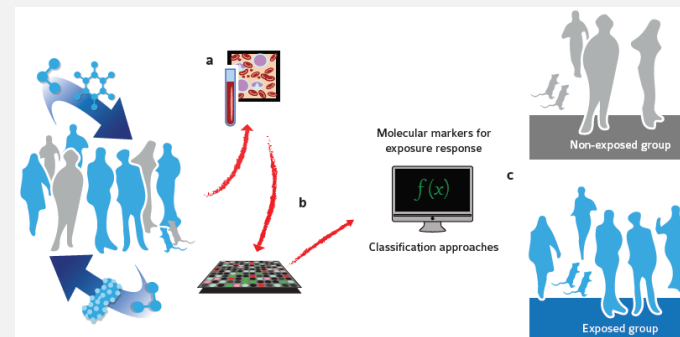
To review biological network models that are suitable for drug discovery, toxicological and mechanistic research in respiratory and cardiovascular diseases



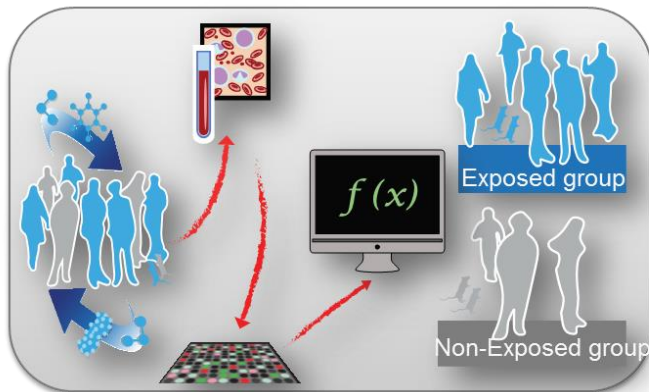
Pacific Symposium on Biocomputing 2015

## Systems Toxicology challenge (2015-2016)

To identify robust blood-based gene signatures as predictors for smoking and cessation status



# Systems Toxicology Challenge – Lessons learned



- 135 registered participants
- 61 international teams
- 23 valid submissions for SC1 and SC2

## Publications

- (1) Poussin et al. *Chem. Res. Toxicol.* 2017
- (2) Belcastro et al. *Comput. Toxicol.* 2017
- (3) Tarca AL et al, *Comput. Toxicol.* 2017 (BP SC1)
- (4) Sarac OS et al, *Comput. Toxicol.* 2017 (BP SC2)

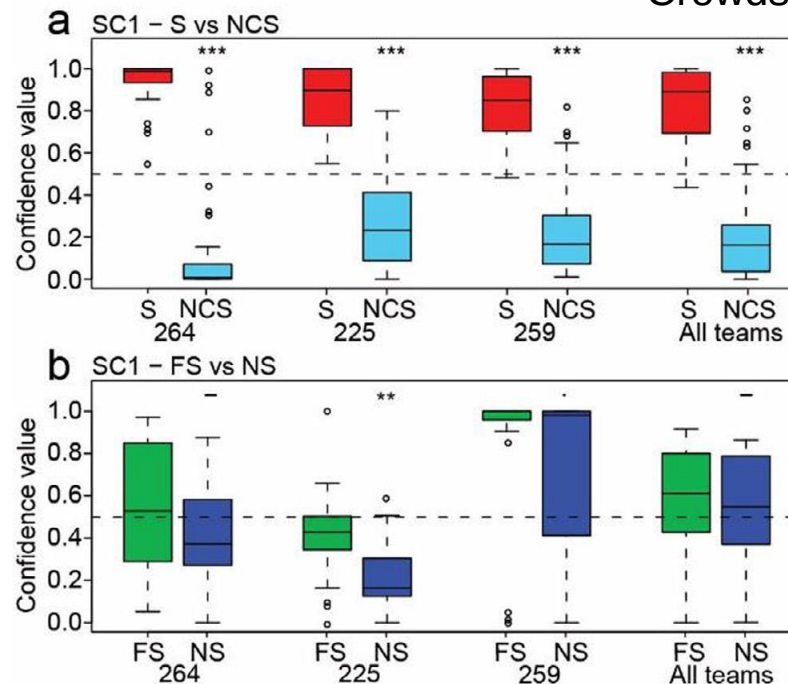
BP: best performers  
SC 1/2: sub-challenge 1 or 2

S: smokers  
NCS: non-current smokers (former and never smokers)

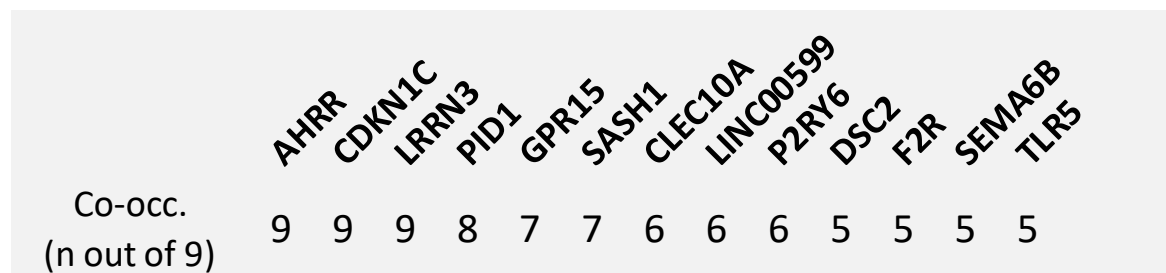
DEG: differentially expression genes

Inductive predictive models  
(as opposed to transductive)

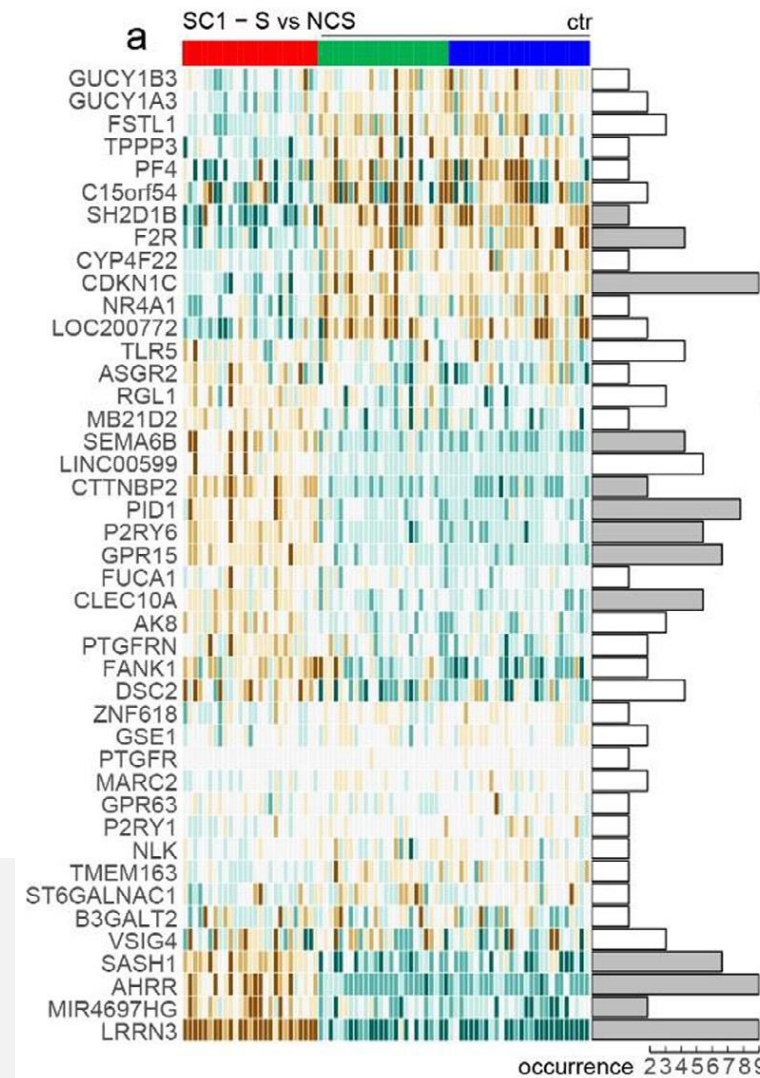
Wisdom of  
Crowds



Consistent core gene signature (2)  
Wisdom of Crowds



Not necessarily DEG



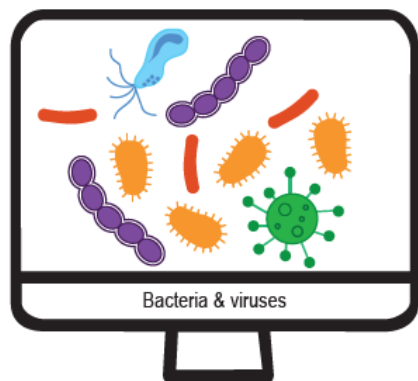
Fold change FDR < 0.05

© 2018 Philip Morris Products S.A.

# Microbiomics computational challenge (2017-2018)

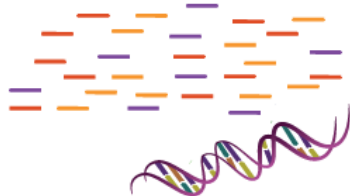
<https://www.sbvimprover.com/challenge-5>

Microbiota



Generation of reads

Microbiome sequencing data



INPUT

Predictions



Download data

Upload predictions

Participants:

- Apply analysis pipeline
- Submit predicted taxonomy and abundance

| Taxonomy<br>Phylum / Genus / Species                                                | Abundance<br>(Relative)                                                              |
|-------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------|
|    |    |
|   |   |
|  |  |

OUTPUT

## Scientific questions

- Which pipelines best recover bacterial community composition and relative sequence read abundance at phylum, genus and species taxa rank?
- Do technical biases and specific microbial composition affect the performance?

## Datasets

- Simulated and real shotgun sequencing metagenomics data

## Scoring

- Scoring of anonymized teams' predictions against the gold standard = known relative abundances of bacteria
- Binary classification and abundance metrics computed using the OPAL software (<https://github.com/CAMI-challenge/OPAL>)
- Score aggregation: weighted sum of ranks
- Final team ranking approved by an external Scoring Review Panel

November 2017



Challenge  
opens



Submission of  
predictions

17 June 2018



Challenge  
closes



Scoring against  
gold standard

June 2018



Team ranking



Winners  
announcement

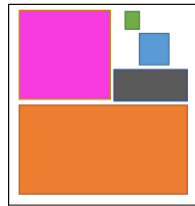


Challenge outcome  
publication



# Microbiomics computational challenge (2017-2018) - Summary

## Datasets



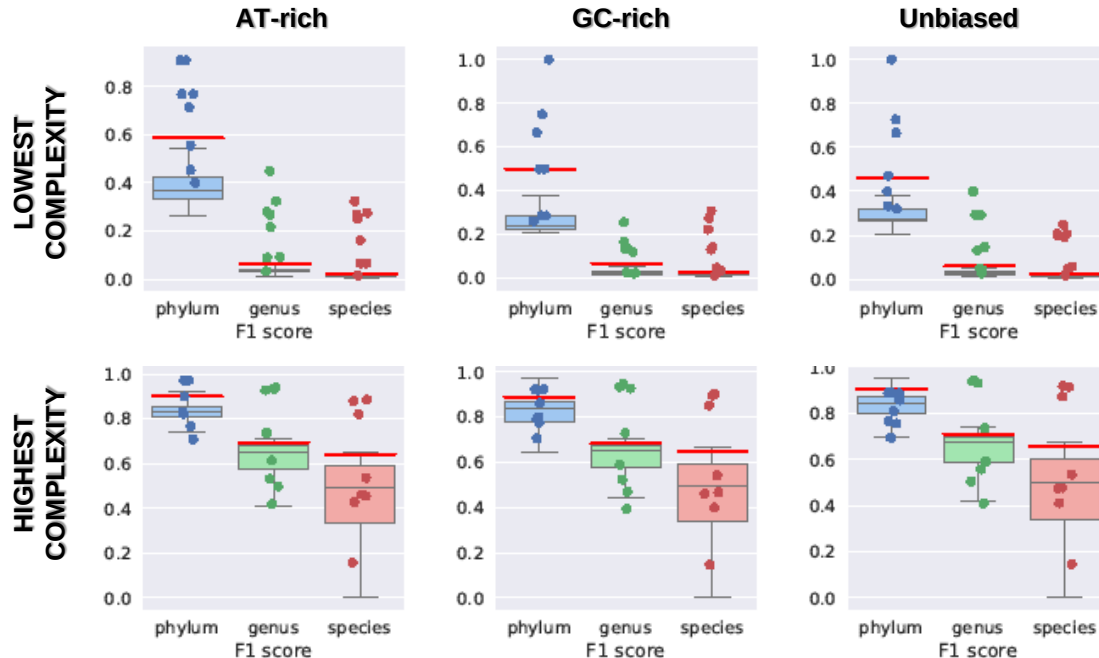
~2'000 genomes

| Sequencing Reads | # Species | # Samples  | Characteristic                                        |
|------------------|-----------|------------|-------------------------------------------------------|
| Simulated        | 10        | 3          | unbiased                                              |
|                  | 40        | 3          | GC-rich bacteria                                      |
|                  | 120       | 3          | AT-rich bacteria                                      |
|                  | 500       | 3          | biased                                                |
|                  | 1000      | 3          |                                                       |
| Real             | 10*       | 2 (lib. 1) | DNA control (ZymoBIOMICS)<br>- 2 library preparations |
|                  | 10*       | 2 (lib. 2) |                                                       |

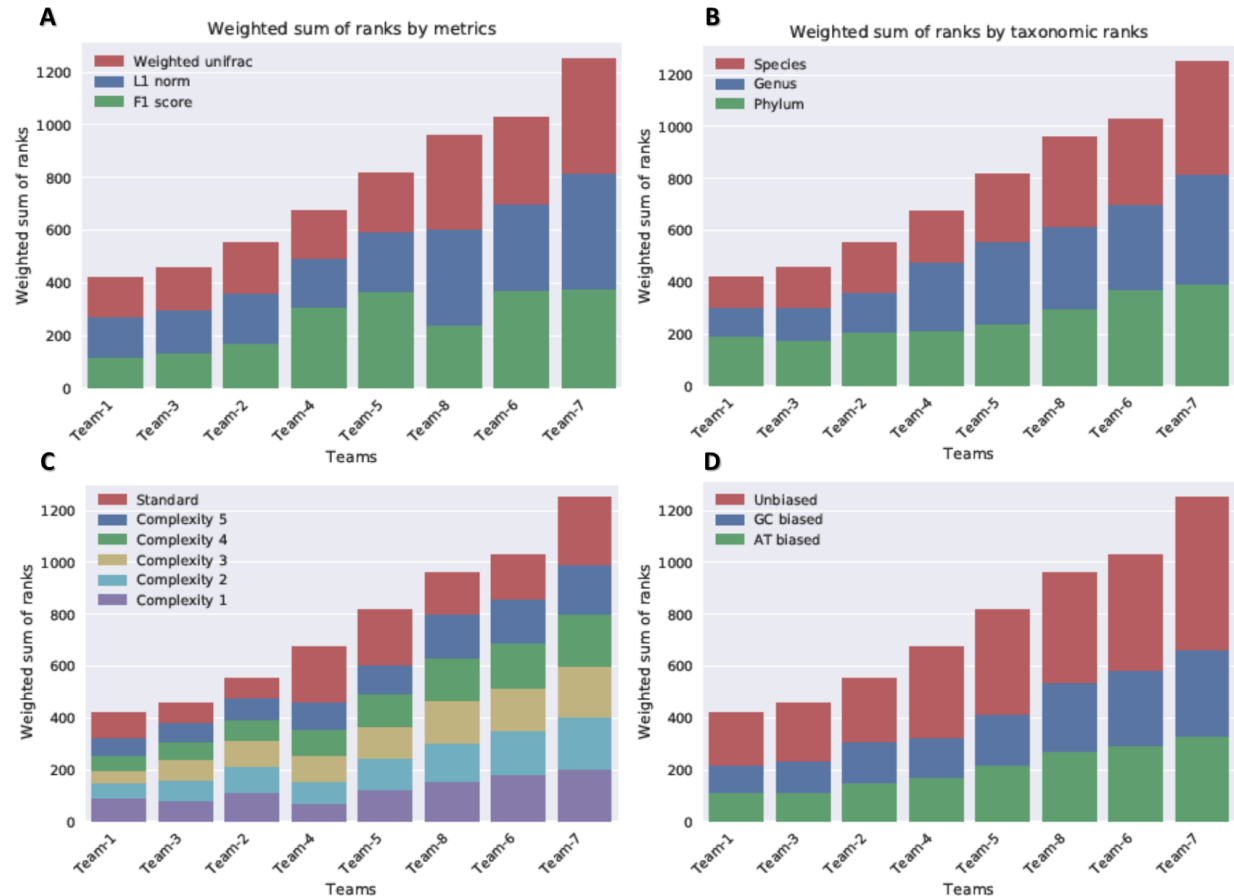
(\*8 bacteria species, 2 yeast species)

=19 samples in total

## F1-score

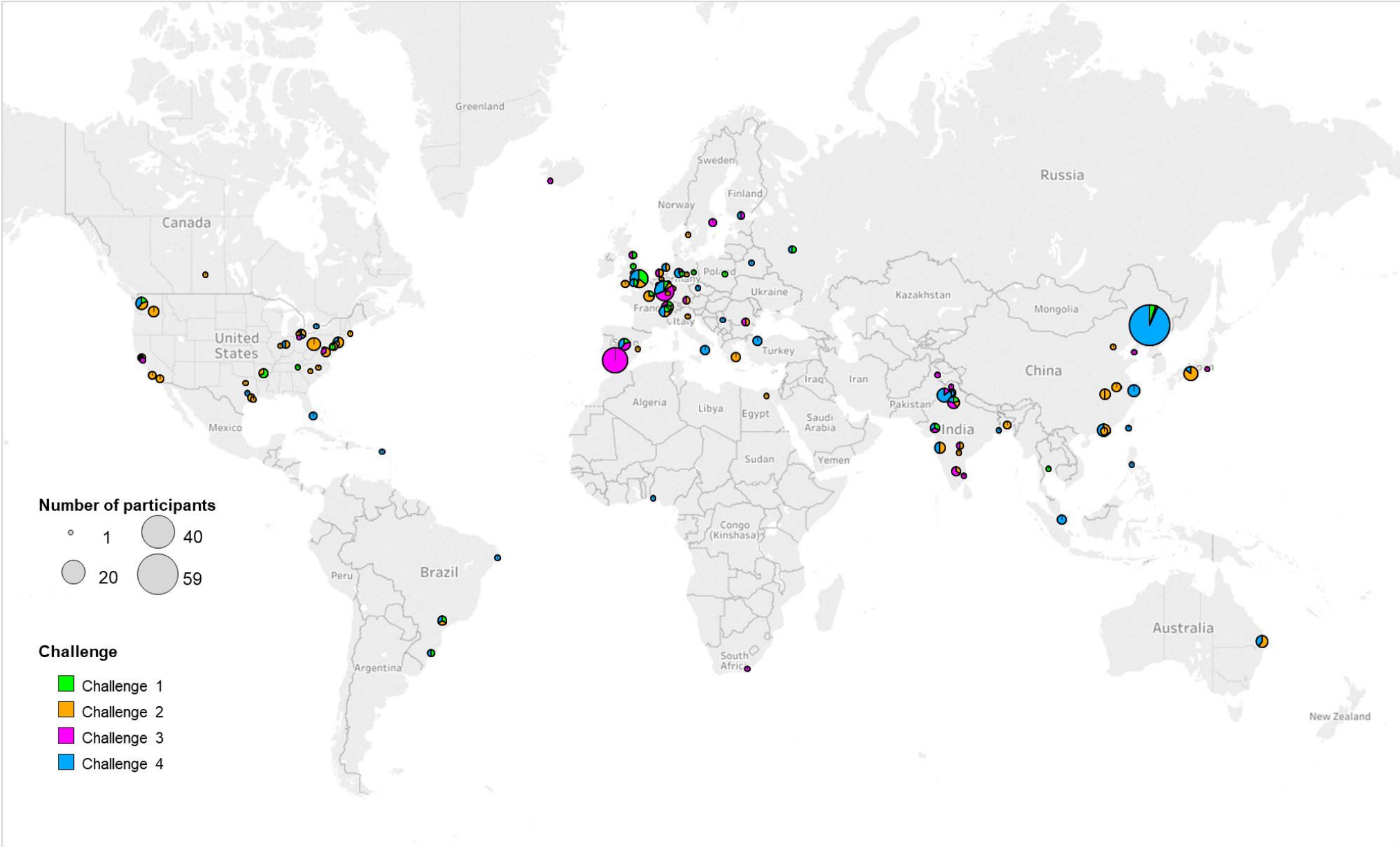


## Weighted sum of ranks



- 8 complete submissions from Japan, India and Armenia
- Scoring: F1-score (binary metric), L1-norm and weighted UniFrac (abundance metrics)
- Final ranking approved by our external Scoring Review Panel: Prof. Alice McHardy (Germany) and Dr Luisa Cutillo (Italy)

# Participation map (challenges 1-4)







Boston, 2012



Montreux, 2014



Athens, 2013



Barcelona, 2015



Orlando, 2016



Singapore, 2016



Tel Aviv, 2017



- Accessing unpublished datasets (through collaborations) used as testing set and Gold Standard
- Low number of participants (submissions) depending on the scientific topic
- Finding the right incentives
- Raising awareness about the challenge
- Define the right duration for the challenge

- Ansari, S. et al. On crowd-verification of biological networks. *Bioinformatics and biology insights* **7** (2013).
- Belcastro, V. et al. The sbv IMPROVER Systems Toxicology computational challenge: Identification of human and species-independent blood response markers as predictors of smoking exposure and cessation status. *Computational Toxicology*, doi:<https://doi.org/10.1016/j.comtox.2017.07.004> (2017).
- Bilal, E. et al. A crowd-sourcing approach for the construction of species-specific cell signaling networks. *Bioinformatics* **31**, 484-491, doi:[10.1093/bioinformatics/btu659](https://doi.org/10.1093/bioinformatics/btu659) (2015).
- Binder, J. et al. in *Pacific Symposium on Biocomputing*. Pacific Symposium on Biocomputing. 270-281.
- Boue, S. et al. Enhancement of COPD biological networks using a web-based collaboration interface. *F1000Research* **4** (2015).
- Hoeng, J., Peitsch, M. C., Meyer, P. & Jurisica, I. Where are we at regarding species translation? A review of the sbv IMPROVER challenge. *Bioinformatics* **31**, 451-452, doi:[10.1093/bioinformatics/btv065](https://doi.org/10.1093/bioinformatics/btv065) (2015).
- Meyer, P. et al. Verification of systems biology research in the age of collaborative competition. *Nature biotechnology* **29**, 811-815, doi:[10.1038/nbt.1968](https://doi.org/10.1038/nbt.1968) (2011).
- Meyer, P. et al. Industrial methodology for process verification in research (IMPROVER): toward systems biology verification. *Bioinformatics* **28**, 1193-1201, doi:[10.1093/bioinformatics/bts116](https://doi.org/10.1093/bioinformatics/bts116) (2012).
- Poussin, C. et al. Crowd-Sourced Verification of Computational Methods and Data in Systems Toxicology: A Case Study with a Heat-Not-Burn Candidate Modified Risk Tobacco Product. *Chemical research in toxicology* **30**, 934-945, doi:[10.1021/acs.chemrestox.6b00345](https://doi.org/10.1021/acs.chemrestox.6b00345) (2017).
- Poussin, C. et al. The species translation challenge-a systems biology perspective on human and rat bronchial epithelial cells. *Scientific data* **1**, 140009, doi:[10.1038/sdata.2014.9](https://doi.org/10.1038/sdata.2014.9) (2014).
- Rhissorakrai, K. et al. Understanding the limits of animal models as predictors of human biology: lessons learned from the sbv IMPROVER Species Translation Challenge. *Bioinformatics* **31**, 471-483, doi:[10.1093/bioinformatics/btu611](https://doi.org/10.1093/bioinformatics/btu611) (2015).
- sbv IMPROVER project team et al. On Crowd-verification of Biological Networks. *Bioinformatics and biology insights* **7**, 307-325, doi:[10.4137/BBI.S12932](https://doi.org/10.4137/BBI.S12932) (2013).
- sbv IMPROVER project team et al. Reputation-based collaborative network biology. *Pacific Symposium on Biocomputing*. Pacific Symposium on Biocomputing, 270-281 (2015).
- sbv IMPROVER project team et al. Enhancement of COPD biological networks using a web-based collaboration interface. *F1000Research* **4**, 32, doi:[10.12688/f1000research.5984.2](https://doi.org/10.12688/f1000research.5984.2) (2015).
- sbv IMPROVER project team et al. Community-Reviewed Biological Network Models for Toxicology and Drug Discovery Applications. *Gene regulation and systems biology* **10**, 51-66, doi:[10.4137/GRSB.S39076](https://doi.org/10.4137/GRSB.S39076) (2016).
- Tarca, A. L. et al. Strengths and limitations of microarray-based phenotype prediction: lessons learned from the IMPROVER Diagnostic Signature Challenge. *Bioinformatics* **29**, 2892-2899, doi:[10.1093/bioinformatics/btt492](https://doi.org/10.1093/bioinformatics/btt492) (2013).

# Thank you!

Questions? Contact Us  
[sbvimprover.RD@pmi.com](mailto:sbvimprover.RD@pmi.com)

**Website: [sbvimprover.com](http://sbvimprover.com)**

**Scoring Review Panel of Experts**

Douglas Connect  
*Working communities*

**SBX**

**orangebus.**



**Sci Pinion**

**ADVANTAGE**  
Integral



**PMI SCIENCE**  
PHILIP MORRIS INTERNATIONAL

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 crowd sourcing method for verification of scientific data and results. The project is led and funded by Philip Morris International.

# BACK-UP SLIDES



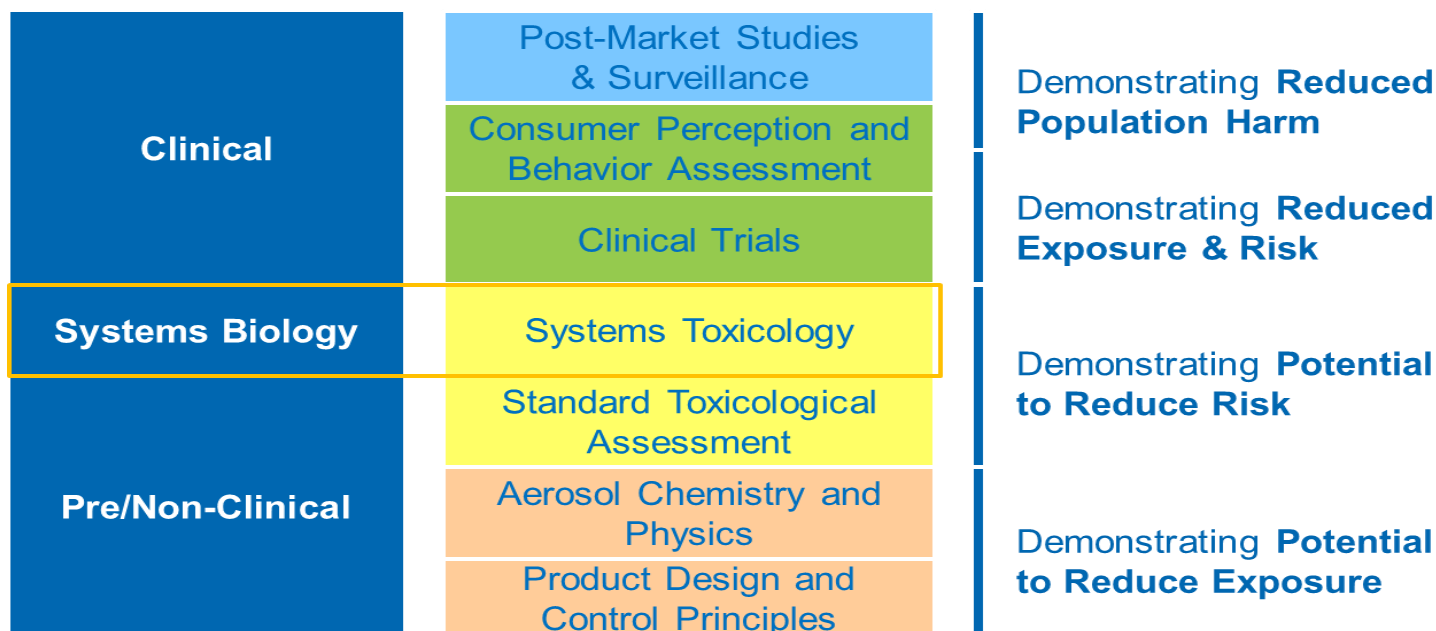
Smoking is one of the causes of serious diseases, such as cardiovascular diseases, lung cancer, and chronic obstructive pulmonary disease.

Philip Morris International is therefore developing novel products that may have the potential to reduce smoking-related disease risk compared with that of smoking cigarettes.

Scientific determination of the reduced risk potential of these products includes comparison of the biological impact with that of a reference cigarette (3R4F) on a mechanism-by-mechanism basis.

We want to share this data and encourage other stakeholders in inhalation toxicology to also share their data on the same platform.



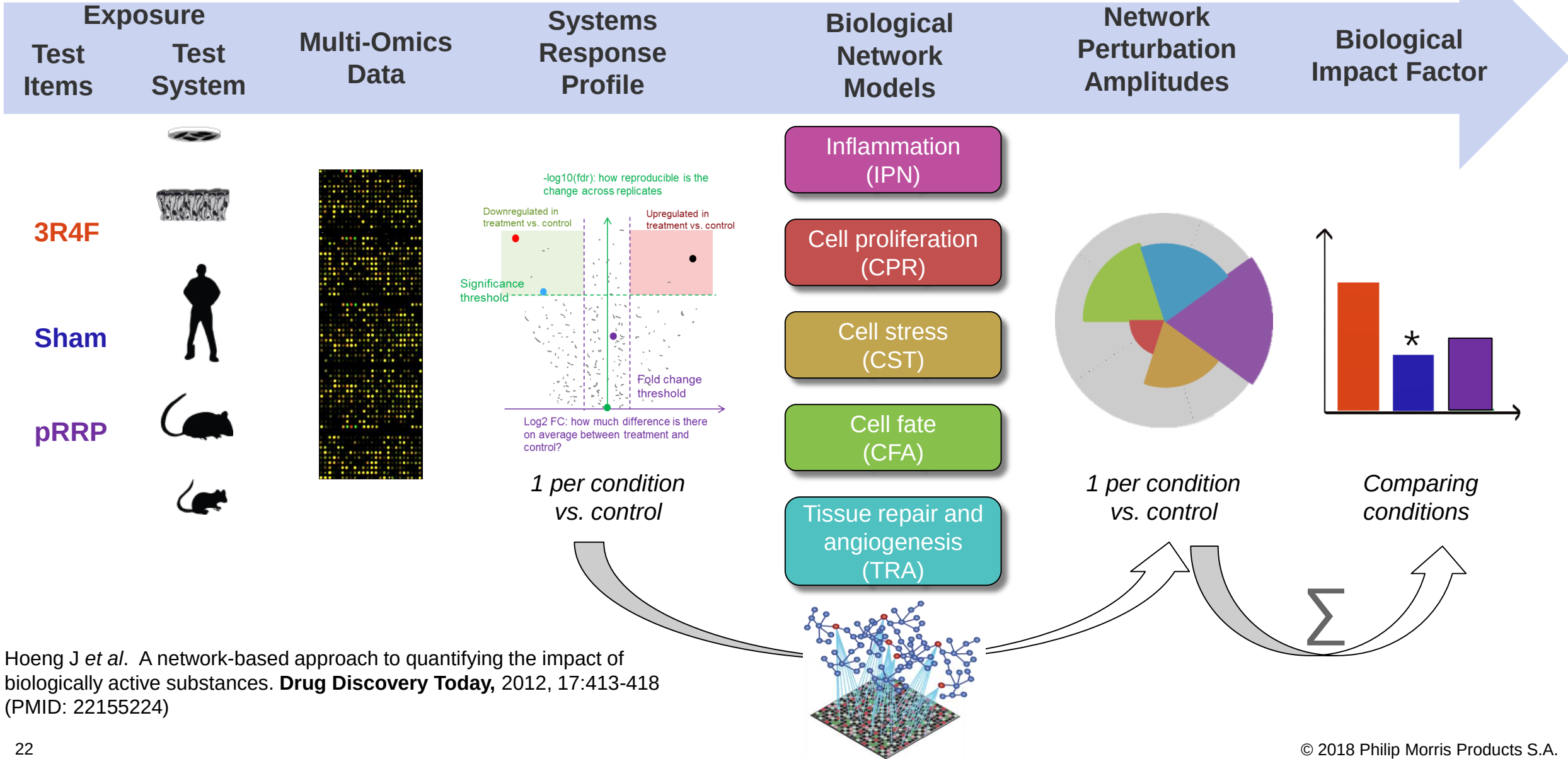


## Program: Systems Toxicology

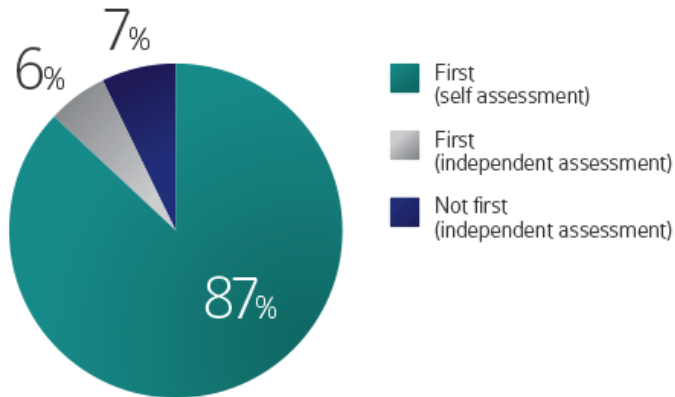
- Demonstrate and quantify the risk-reduction potential of RRP<sup>s</sup>\* *in vitro* and *in vivo*.
- Develop methods for the quantitative, mechanism-based comparison of the biological impact of RRP aerosol as compared with that of cigarette smoke.
- Further grow our mechanistic understanding of cigarette smoke-induced diseases.
- Independently verify our findings using community-based approaches.

\* *Reduced-Risk Products ("RRPs") is the term we use to refer to products that present, are likely to present, or have the potential to present less risk of harm to smokers who switch to these products versus continued smoking. We have a range of RRP<sup>s</sup> in various stages of development, scientific assessment, and commercialization. Because our RRP<sup>s</sup> do not burn tobacco, they produce far lower quantities of harmful and potentially harmful compounds than found in cigarette smoke.*

# Systems Toxicology Assessment: Use Disease Mechanism Understanding for Product Assessment



# Meeting the Needs of Industry – The Role of sbv IMPROVER



*The self-assessment trap: can we all be better than average?*

- ✓ Researchers wishing to publish their methods are usually required to compare their methods against others
- ✓ Authors' method tends to be the best in an unreasonable majority of cases
  - ✓ Selective reporting of performance: inadvertent or disingenuous
  - ✓ Choice of only one, best metric

Mol Syst Biol. 2011 Oct 11;7:537. doi: 10.1038/msb.2011.70.

Develop a robust methodology that verifies systems biology-based approaches



[Login](#)  [Register](#) 

sbv IMPROVER

## Microbiomics

## SysTox

## Network Verification

## Diagnostic Signature

## Species Translation

[Home](#)

## About sbv IMPROVER

## Conferences

## News & Publications

## Datathons

## Symposia

# Microbiomics Challenge

## Benchmark your methods!

Contribute. Collaborate. Compete.

[Read More](#)

## sbv IMPROVER at a Glance

sbv IMPROVER stands for Systems Biology Verification combined with Industrial Methodology for Process Verification in Research. This approach aims to provide a measure of quality control of industrial research and development by verifying the methods used. The sbv IMPROVER project is a collaborative effort led and funded by PMI Research and Development. For more information please see [Nature Biotechnology \(2011\)](#) or [Bioinformatics \(2012\)](#).

It is different from other scientific crowdsourcing approaches as it focuses on the verification of processes in an industrial context and not just on basic questions regarding science. The shv IMPROVER approach allows an

## Perennial Diagnostic Signature Benchmarking

Use our diagnostic signature benchmarking tool to find out how you compare to your peers.

**1**  **DOWNLOAD**

2  SUBMIT

**3 COMPARE**