



www.sbvimprover.com



**The systems biology
verification endeavor –
harness the power of the
crowd to address
computational and
biological challenges**

*Stéphanie Boué, PhD
Manager Scientific Transparency and Verification
Philip Morris International R&D*

12th Career Fair in Horizons in Molecular Biology - September 2018



Crowdsourcing in biomedical research

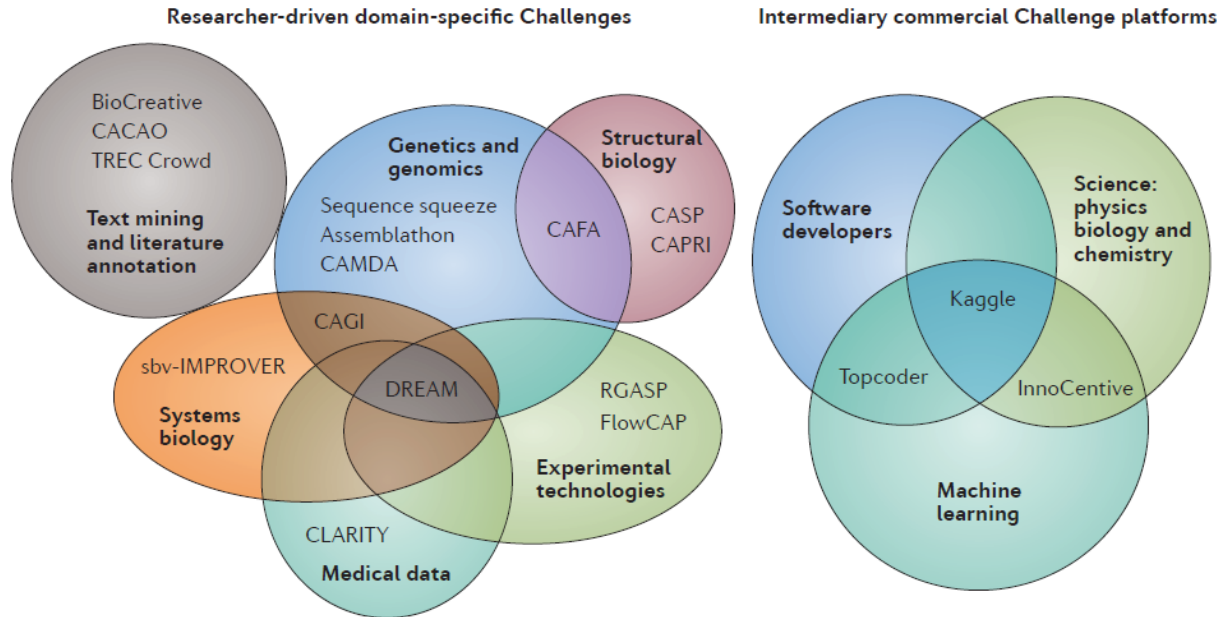


Figure 1 | **Challenge platforms and organizations.** The most popular researcher-driven Challenge initiatives in the life sciences (left) and the most popular commercial Challenge platforms (right) are shown. Initiatives, such as DREAM (Dialogue for Reverse Engineering Assessment and Methods), FlowCAP (Flow Cytometry Critical Assessment of Population Identification Methods), CAGI (Critical Assessment of Genome Interpretation) and sbv-IMPROVER (Systems Biology Verification combined with Industrial Methodology for Process Verification in Research), organize several Challenges per year; only the generic project and not the specific Challenges are shown. Among the most popular and successful commercial Challenge platforms are: InnoCentive, which crowdsources Challenges in science and technology (social sciences, physics, biology and chemistry); Topcoder, which serves the software developer community; and Kaggle, which administers Challenges to machine-learning and computer experts, addressing predictive analytics problems in a wide range of disciplines. The figure is not comprehensive, but highlights the most consistent and well-established Challenge initiatives. CAFA, Critical Assessment of Functional Annotation; CACAO, Cross-language Access to Catalogues And On-line libraries; CAMDA, Critical Assessment of Massive Data Analysis; CAPRI, Critical Assessment of PRediction of Interaction; CASP, Critical Assessment of protein Structure Prediction; CLARITY, Children's Leadership Award for the Reliable Interpretation and appropriate Transmission of Your genomic information; RGASP, RNA-seq Genome Annotation Assessment Project; TREC Crowd, Text REtrieval Conference Crowdsourcing Track.

J. Saez-Rodriguez, J. C. Costello, S. H. Friend, M. R. Kellen, L. Mangravite, P. Meyer, T. Norman, G. Stolovitzky, Crowdsourcing biomedical research: leveraging communities as innovation engines. *Nature reviews. Genetics* **17**, 470-486 (2016)

Benefits of crowdsourcing in science – for the organizers and participants

**Be part of a pioneering community,
working together to improve how
scientific research is verified**

Complement peer review

**Drive innovation in science
with crowdsourced solutions**

Monetary incentives

Peer recognition and self-esteem

**Contribute and work toward
consensus in the scientific
community**

Access network of experts

Co-author publications

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

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

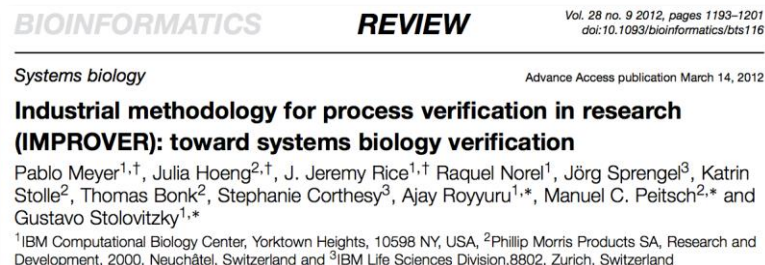
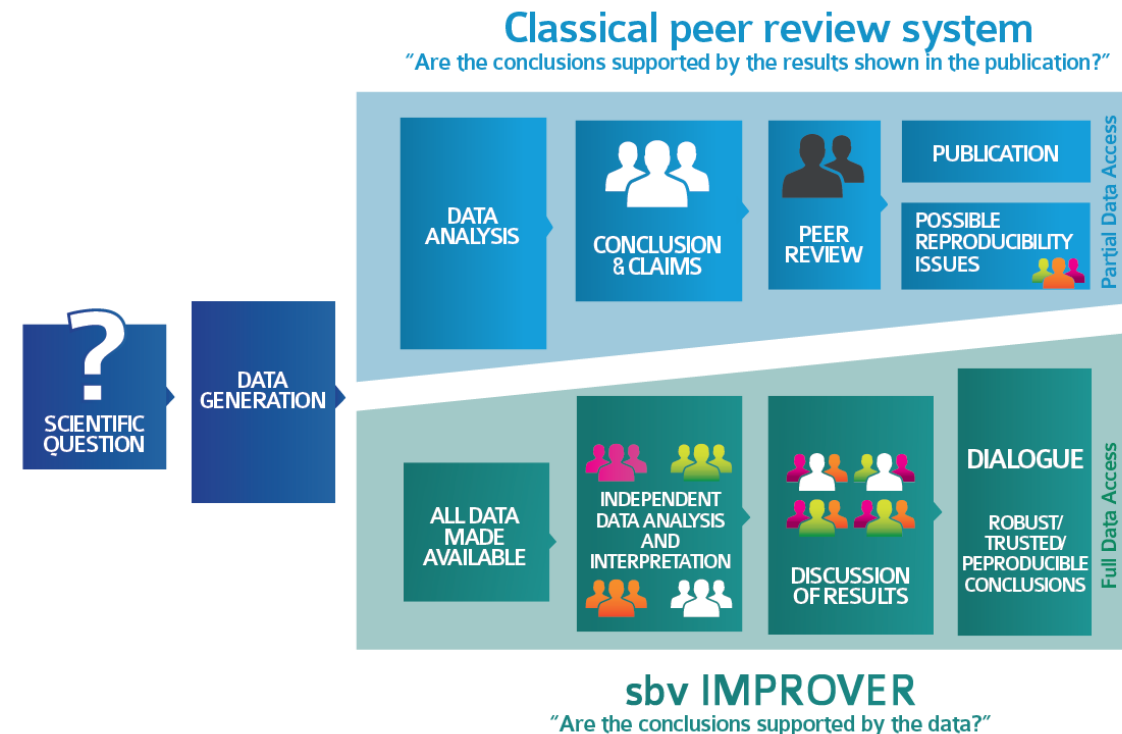
Scientific determination of the reduced risk potential of these products includes comparison of their biological impact with that of a 3R4F reference cigarette on a mechanism-by-mechanism basis.



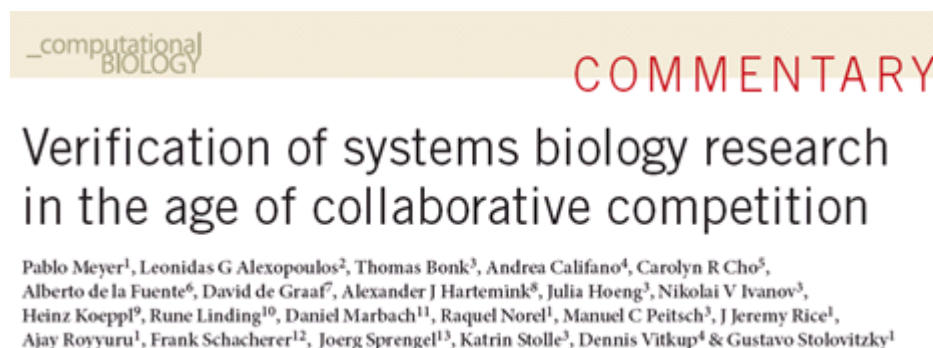
sbv IMPROVER stands for **S**ystems **B**iology **V**erification combined with **I**ndustrial **M**ethodology for **P**rocess **V**erification in **R**esearch.

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

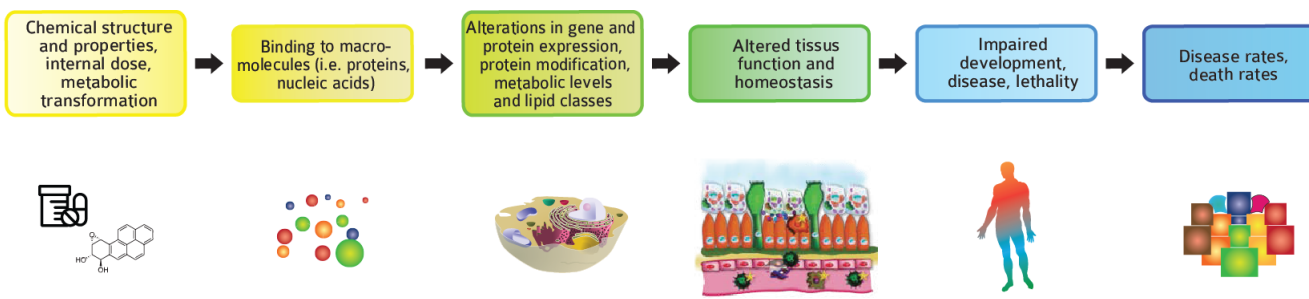
The sbv IMPROVER project is a collaborative effort led and funded by PMI Research and Development.



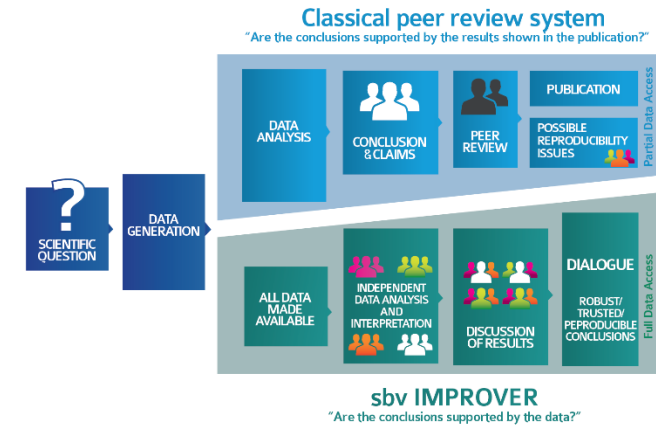
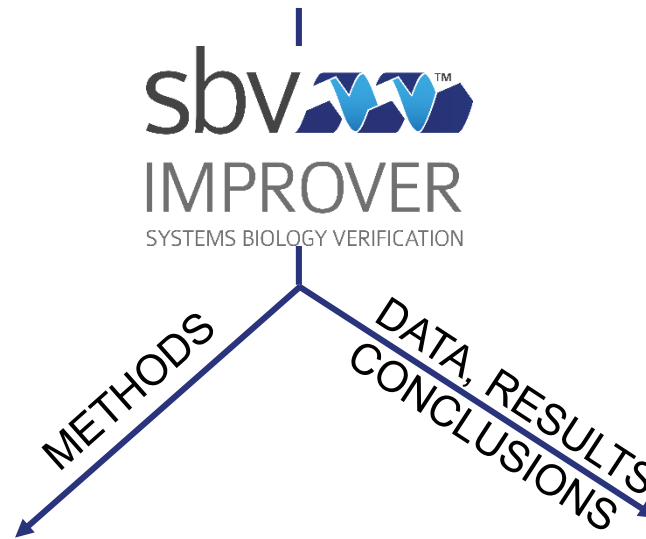
Bioinformatics 2012 28(9):1193-1201



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



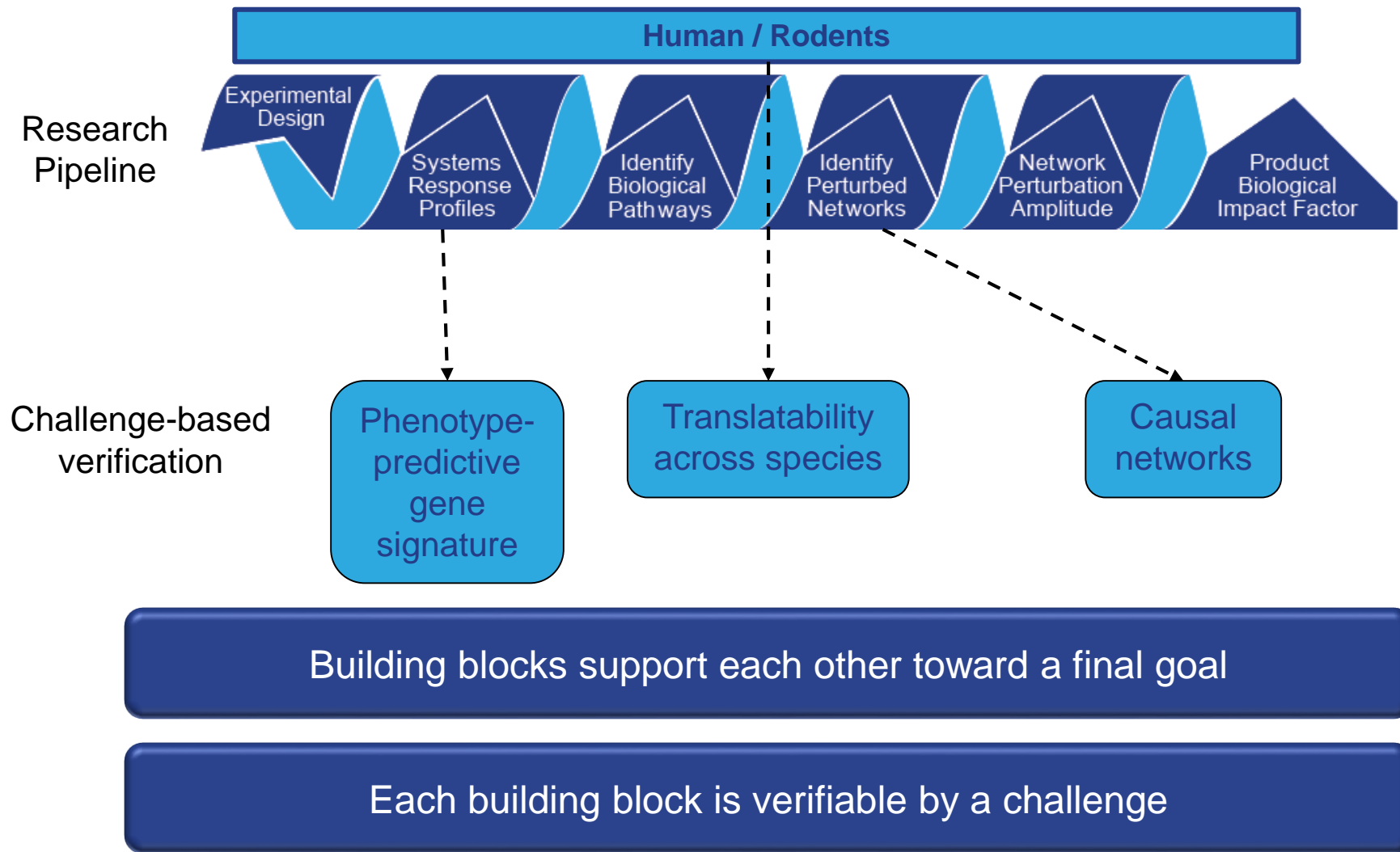
$$\pi \begin{smallmatrix} \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet \end{smallmatrix} \begin{smallmatrix} \neq \\ \pm \end{smallmatrix} \sqrt{f} x$$

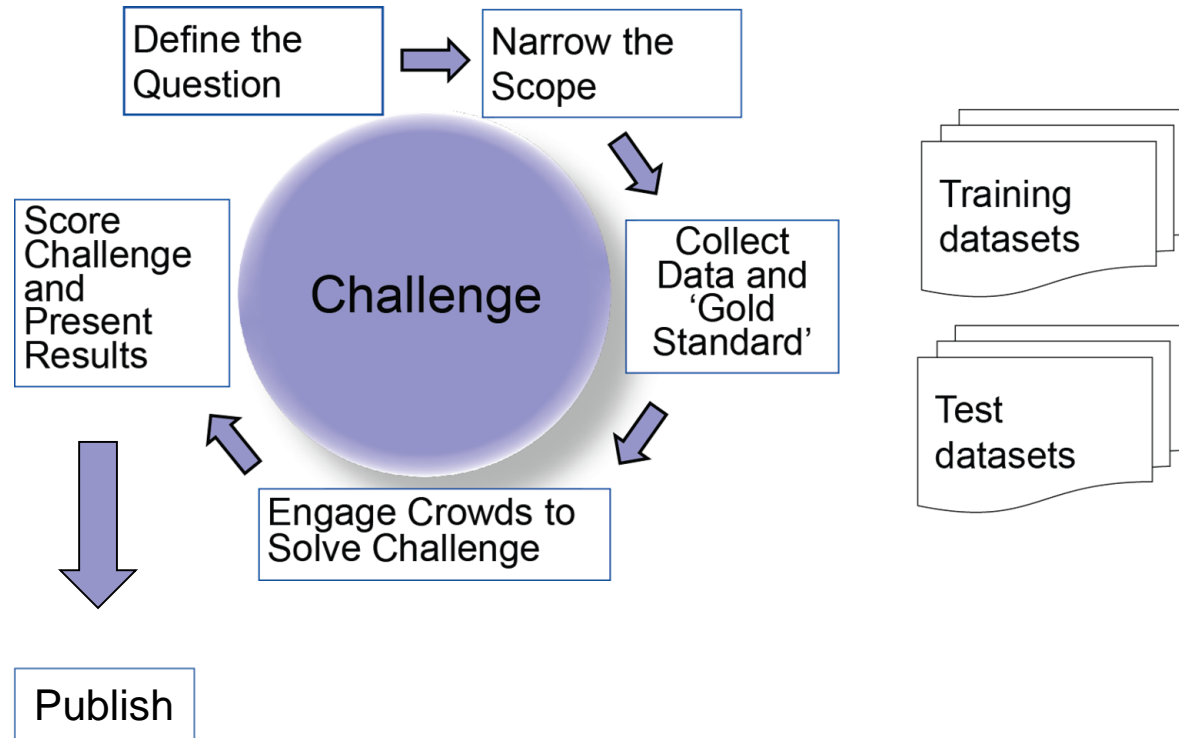


COMPUTATIONAL CHALLENGE

VERIFICATION

Complex industrial research pipeline/workflow divided into verifiable building blocks





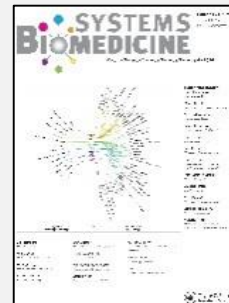
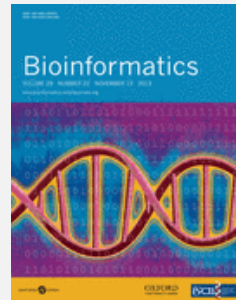
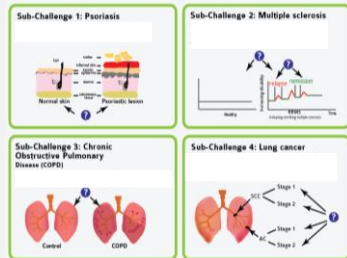
Double-blind performance assessment

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

Past sbv IMPROVER Computational Challenges

Diagnostic Signature Challenge - DSC (2012)

To identify gene signatures for diagnostic classification in four disease areas

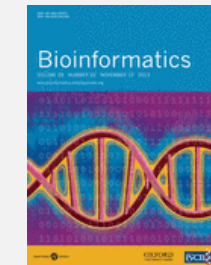


Species Translation Challenge - STC (2013)

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)

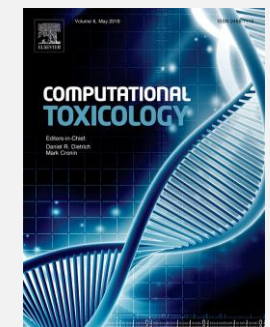
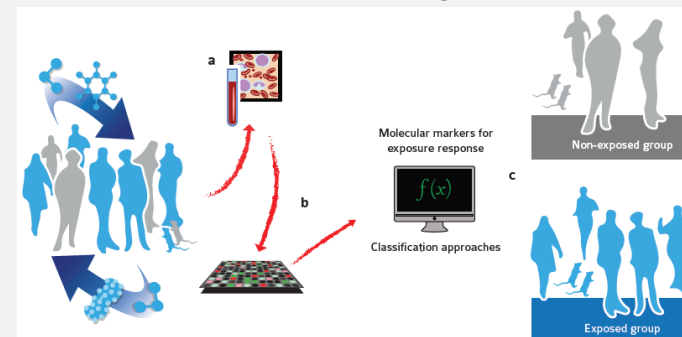
To review biological network models that are suitable for drug discovery and toxicological and mechanistic research in respiratory and cardiovascular diseases



Pacific Symposium on Biocomputing 2015 

Systems Toxicology Challenge (2015-2016)

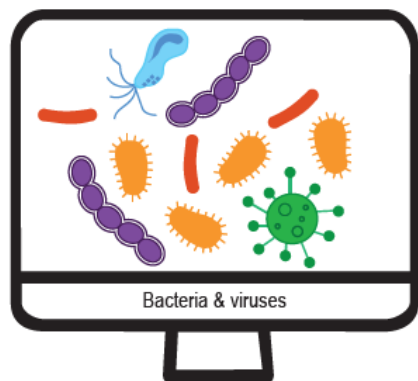
To identify robust blood-based gene signatures as predictors for smoking and cessation status



Microbiomics Computational Challenge (2017-2018)

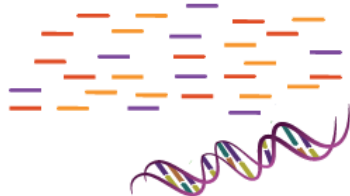
<https://www.sbvimprover.com/challenge-5>

Microbiota



Generation of reads

Microbiome sequencing data



INPUT

Predictions









Download data

Upload predictions

Participants:

- Apply analysis pipeline
- Submit predicted taxonomy and abundance

Taxonomy	Abundance
Phylum / Genus / Species	(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 SRP

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

CBN CAUSAL BIOLOGICAL NETWORKS DATABASE

HOMEABOUTHELP

ABOUT CBN


The Causal Biological Networks (CBN) database is composed of multiple versions of over 120 modular, manually curated, BEL-scripted biological network models supported by over 80,000 unique pieces of evidence from the scientific literature.

They represent causal signaling pathways across a wide range of biological processes including cell fate, cell stress, cell proliferation, inflammation, tissue repair and angiogenesis in the pulmonary and vascular systems.

Search for a Network

Human ▼ proliferation


SEARCH RESULTS (34)



Cell Proliferation (CPR)
Calcium

The Calcium network depicts the causal mechanisms that regulate calcium signaling in the context of cell proliferation. The network describe...

[VIEW](#) | [DETAILS](#) | [EXPORT](#) ▼
2.0 (Other versions: [Hs_1.1](#), [Hs_1.0](#))

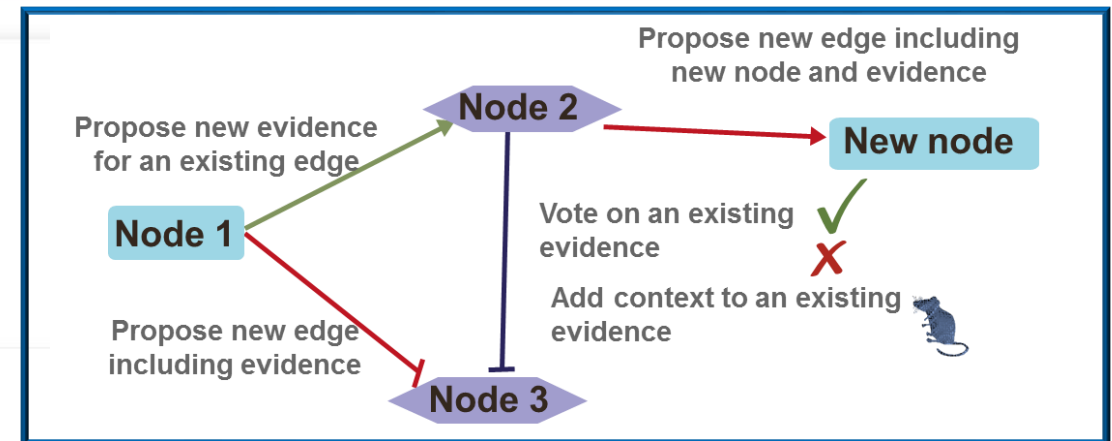
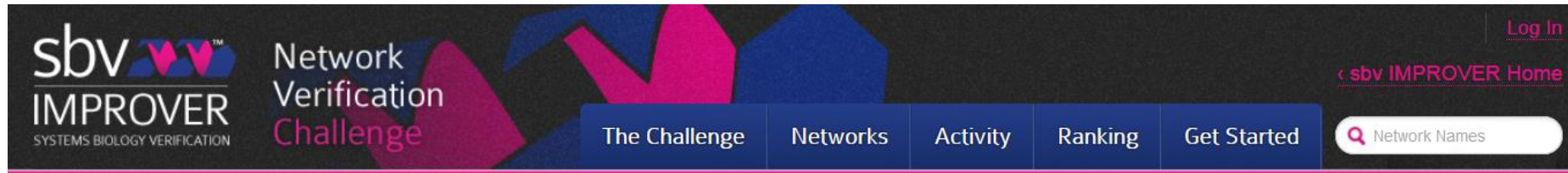


Cell Proliferation (CPR)
Cell Cycle

The Cell Cycle network depicts the causal mechanisms that regulate cell cycle including canonical elements of the core machinery regulating ...

[VIEW](#) | [DETAILS](#) | [EXPORT](#) ▼
1.2 (Other versions: [Hs_1.0](#), [Hs_1.1](#))

Network Verification Challenges



	NVC1	NVC2	NVC3
Number of networks	40 (lung biology)	8 (lung biology)	3 (liver xenobiotic metabolism)
Open phase	10/2013-02/2014	04/2014-05/2015	07/2017-06/2018
Number of participants / countries	150 / 18	173 / 26	39 / 6
Number of votes	2,456	9,286	2,599
New pieces of evidence	885	2,225	423
Number of new edges	451	1,289	250

Testimonials from participants on NVC3

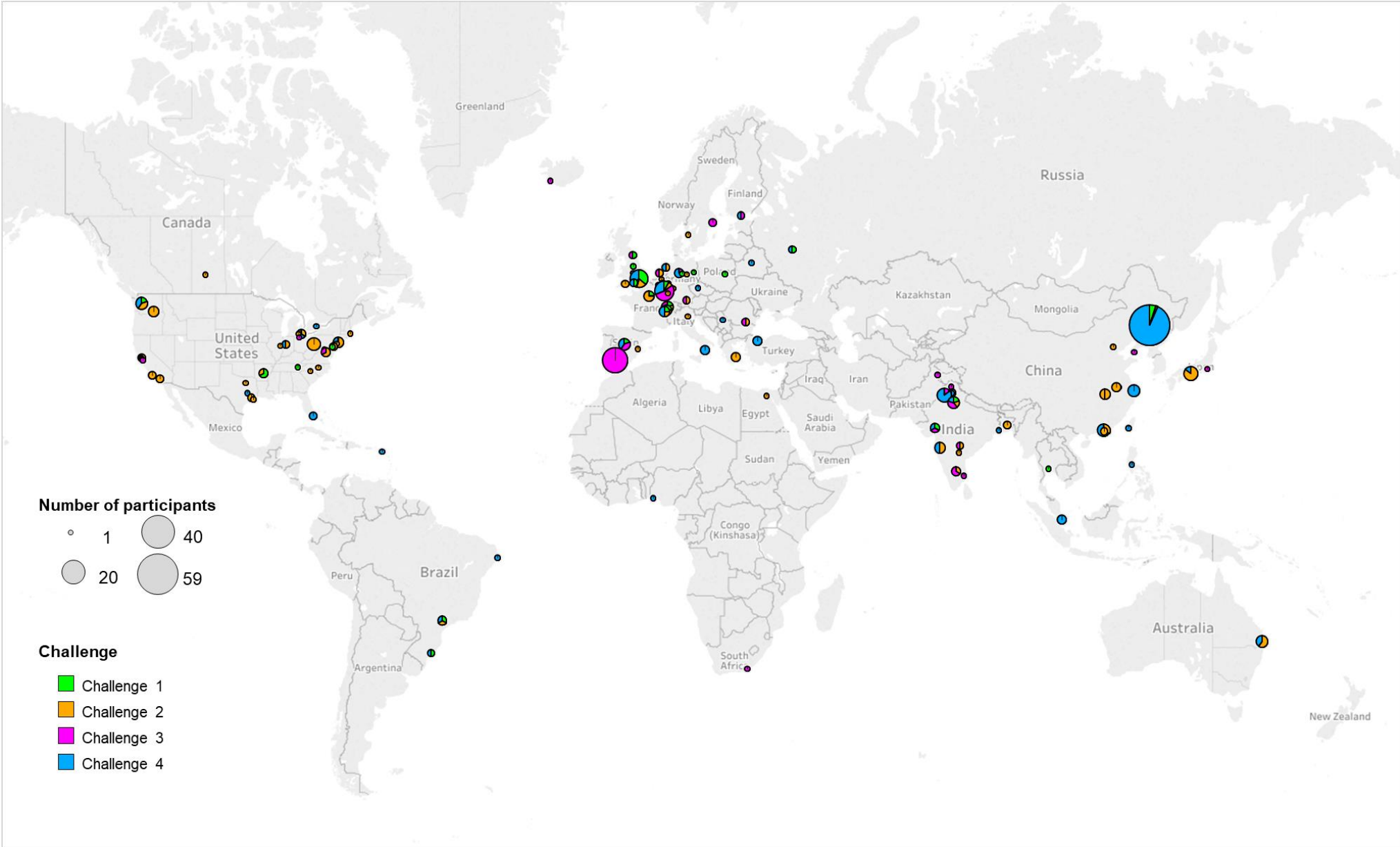
As a recent biology graduate, an interesting topic caught my attention. During my work I was exploring many research papers and the platform itself. I enriched my knowledge in field of Xenobiotic metabolism in the liver, but most importantly I learned the basics of BEL. In the future, I will definitely use network models in my scientific activity, learn more and contribute in collaborative projects. I would recommend participation to young scientists because beside rewards, I gained great knowledge and new acquaintances.

Tijana Samardzic from the Faculty of Sciences, University of Novi Sad, Serbia

One of the great things about this challenge is that we learned a lot during it. The BEL syntax for example, which enables the easiest way of representing scientific findings in life sciences, lot of us was not familiar with it before. I met a lot of incredible people during the challenge, and after the challenge on NVC3 workshop as well, and I'm so glad that I have been a part of project like this. I would recommend my colleagues to participate in future challenges, and would also like to participate again.

Ana Predojevic from the Faculty of Sciences, University of Novi Sad, Serbia

Participation map (Challenges 1-4)



sbv IMPROVER Symposia



Boston, 2012



Athens, 2013



Montreux, 2014



Orlando, 2016



Singapore, 2016



Barcelona, 2015



Tel Aviv, 2017



Neuchâtel, 2018

Publications resulting from the sbv IMPROVER project

- 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

Website: <https://www.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 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.