sbv IMPROVER
SYSTEMS BIOLOGY VERIFICATION

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

# Harness The Power of the Crowd to Address Computational and Biological Challenges

*Dr Stéphanie Boué, Manager Scientific Transparency & Verification*
*Dr Vincenzo Belcastro, Scientist Systems Biology*

*Philip Morris Products S.A., Quai Jeanrenaud 5, CH-2000 Neuchatel, Switzerland (Part of Philip Morris International group of companies)*

CELEBRATING THE 25TH CONFERENCE ON
Intelligent Systems for Molecular Biology and the
16TH European Conference on Computational Biology
JULY 21-25 • PRAGUE CONGRESS CENTRE

# Outline

- **Introduction**: PMI R&D & Systems Toxicology
- **Transparency and verification in science**: INTERVALS & sbv IMPROVER
- **Past and active challenges**: diagnostic signature challenge, species translation challenge, network verification challenge, systems toxicology challenge
- **Datathons**: collaborate to innovate
- **Next computational challenge:** microbiomics challenge

# Outline

- **Introduction**: PMI R&D & Systems Toxicology
- **Transparency and verification in science**: INTERVALS & sbv IMPROVER
- **Past and active challenges**: diagnostic signature challenge, species translation challenge, network verification challenge, systems toxicology challenge
- **Datathons**: collaborate to innovate
- **Next computational challenge:** microbiomics challenge

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

# PMI R&D

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

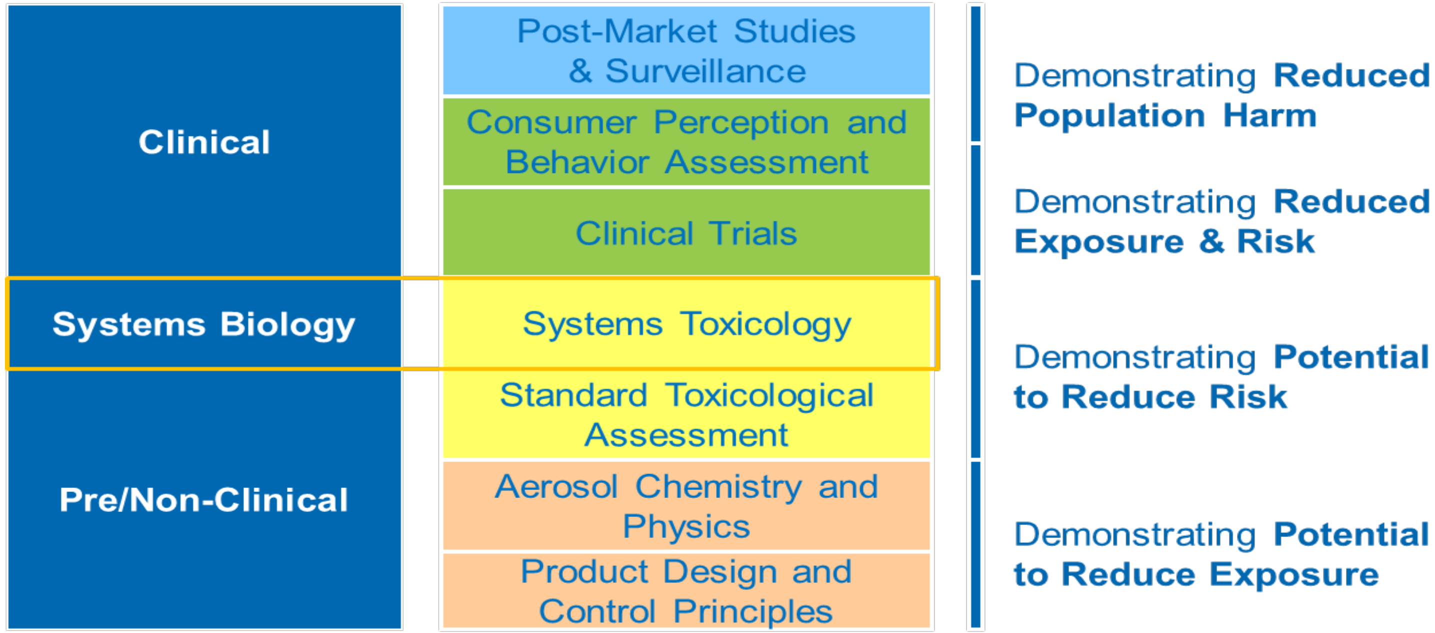
Philip Morris International is therefore developing novel products that may have the potential to reduce smoking-related disease risk compared to smoking cigarettes.

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.

Photo: Paul-André [illegible]

# Scientific Assessment Strategy & Systems Toxicology



**Clinical**

**Systems Biology**

**Pre/Non-Clinical**

Post-Market Studies & Surveillance

Consumer Perception and Behavior Assessment

Clinical Trials

Systems Toxicology

Standard Toxicological Assessment

Aerosol Chemistry and Physics

Product Design and Control Principles

Demonstrating **Reduced Population Harm**

Demonstrating **Reduced Exposure & Risk**

Demonstrating **Potential to Reduce Risk**

Demonstrating **Potential to Reduce Exposure**

## Program: Systems Toxicology

- Demonstrate and quantify the Risk Reduction Potential of RRPs* in *vitro* and *in vivo*.
- Develop methods for the quantitative mechanism-based comparison of the biological impact of RRPs aerosol as compared to 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 RRPs in various stages of development, scientific assessment and commercialization. Because our RRPs 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

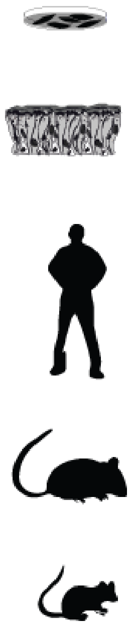
**Exposure**

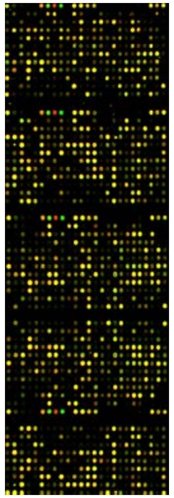
**Test items** | **Test system** | **Multi-Omics Data** | **Systems Response Profile** | **Biological Network models** | **Network Perturbation Amplitudes** | **Biological impact factor (BIF)**

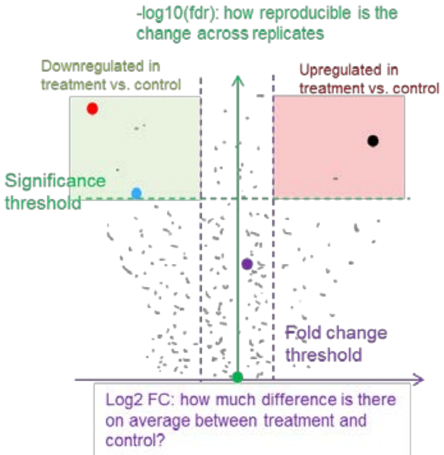
**3R4F**

**Sham**

**pRRP**







*1 per condition vs. control*

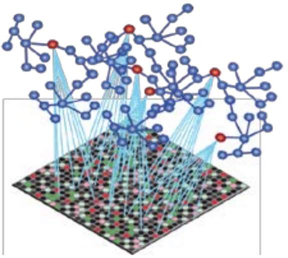
Inflammation (IPN)

Cell Proliferation (CPR)

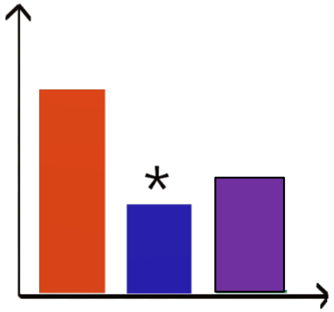
Cell Stress (CST)

Cell Fate (CFA)

Tissue repair and angiogenesis (TRA)



*1 per condition vs. control*



*comparing conditions*

$\sum$

Hoeng J *et al*. A network-based approach to quantifying the impact of biologically active substances. **Drug Discovery Today,** 2012, 17:413-418 (PMID: 22155224)

# Outline

- **Introduction**: PMI R&D & Systems Toxicology
- **Transparency and verification in science**: INTERVALS & sbv IMPROVER
- **Past and active challenges**: diagnostic signature challenge, species translation challenge, network verification challenge, systems toxicology challenge
- **Datathons**: collaborate to innovate
- **Next computational challenge:** microbiomics challenge

# Why do we need to share data?

- Data output is growing rapidly and most researchers would like to use the data of others (often shared privately)
- Despite significant investment, data is not being managed effectively
- Much of the data remains unverifiable
- Time and money is wasted, impacting on science and society
- Funders now require data management and sharing policies

- BUT Shared data is not always understandable and reusable
  - Context is important and good metadata is key
  - Data should be available from a central place, not (only) fragmented in specialist databases
  - Data curation is time consuming but necessary to harmonize data (=need of sufficient motivation, clear requirements from journals and funders)

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

http://blogs.lse.ac.uk/impactofsocialsciences/2014/01/21/five-reasons-to-protect-your-scientific-data/

# Scientific data transparency applied to Industry

Aim: establish a **community** and a public **repository** for 21st-century preclinical and clinical (systems) **inhalation toxicology assessment** data and results that supports open data principles.

Publications
Reports

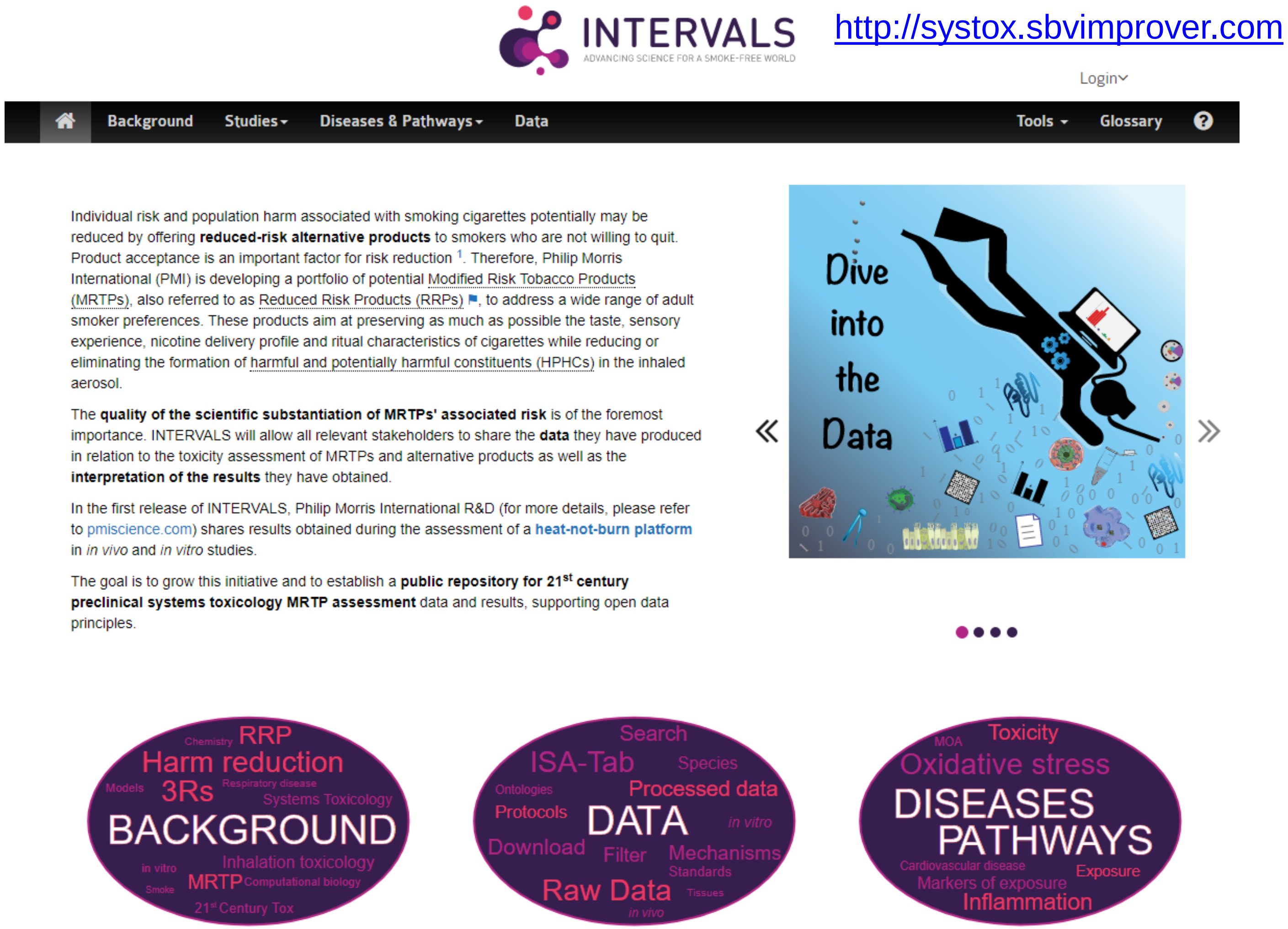
Protocols

Data

Peer review

Community

Education



http://systox.sbvimprover.com

Individual risk and population harm associated with smoking cigarettes potentially may be reduced by offering **reduced-risk alternative products** to smokers who are not willing to quit. Product acceptance is an important factor for risk reduction [1]. Therefore, Philip Morris International (PMI) is developing a portfolio of potential Modified Risk Tobacco Products (MRTPs), also referred to as Reduced Risk Products (RRPs), to address a wide range of adult smoker preferences. These products aim at preserving as much as possible the taste, sensory experience, nicotine delivery profile and ritual characteristics of cigarettes while reducing or eliminating the formation of harmful and potentially harmful constituents (HPHCs) in the inhaled aerosol.

The **quality of the scientific substantiation of MRTPs' associated risk** is of the foremost importance. INTERVALS will allow all relevant stakeholders to share the **data** they have produced in relation to the toxicity assessment of MRTPs and alternative products as well as the **interpretation of the results** they have obtained.

In the first release of INTERVALS, Philip Morris International R&D (for more details, please refer to pmiscience.com) shares results obtained during the assessment of a **heat-not-burn platform** in *in vivo* and *in vitro* studies.

The goal is to grow this initiative and to establish a **public repository for 21st century preclinical systems toxicology MRTP assessment** data and results, supporting open data principles.

Boué S, Exner T, Ghosh S et al. Supporting evidence-based analysis for modified risk tobacco products through a toxicology data-sharing infrastructure [version 1; referees: awaiting peer review]. F1000Research 2017, 6:12 (doi: 10.12688/f1000research.10493.1)

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

---

* *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 RRPs in various stages of development, scientific assessment and commercialization. Because our RRPs do not burn tobacco, they produce far lower quantities of harmful and potentially harmful compounds than found in cigarette smoke.*

# Building an infrastructure supporting scientific excellence

- Strong emphasis on transparency and publication of methods and data to enable reproducibility (inspired by F1000Research, data journals, and protocols.io)
- Guiding the user to facilitate upload of relevant and reusable data and methods
- Highlighting key results
- Facet search to ease navigation
- Educational material to better contextualize studies and results
- Community portal features: comments, peer review (phase 2)
- Data analytics integrated on the site (phase 2)

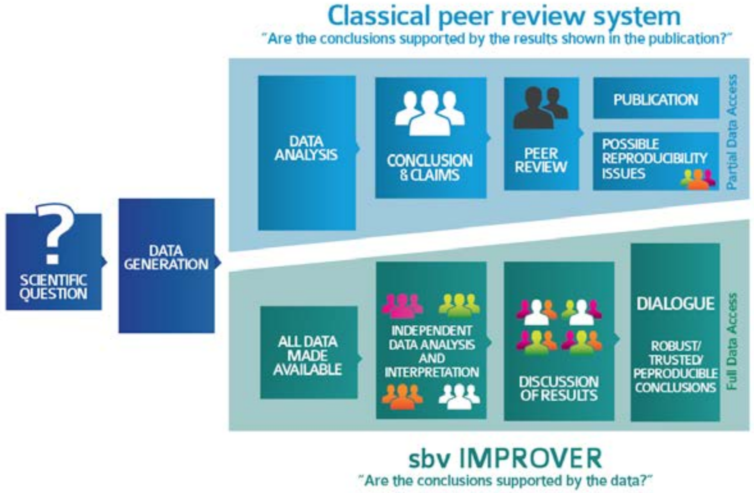
# IMPROVER: Industrial Methodology for Process Verification in Research

Project initiated 6 years ago and funded by Philip Morris International

Aims to provide a measure of quality control in R&D by identifying the building blocks that need verification in a complex industrial research pipeline

Aims to verify methods & data in systems biology / toxicology using double blind performance assessment

Complements the classical peer review system



_computational BIOLOGY

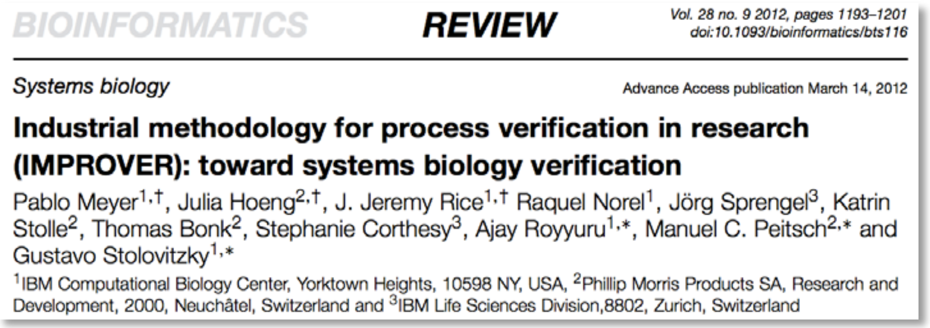
COMMENTARY

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



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

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

# sbv IMPROVER solutions to verify methods and results

**Interactive poster session**

On site: 2-3 hours
Lead time: 1-2 months

**Datathon/ Symposium**

Geo-restricted scientists

Offline 2-4 months
On site: 1-2 days
Lead time: 6-8 months

**Challenge**

Worldwide scientific community

Live phase: 6-8 months
Lead time: 1-2 years

**Independent verification**

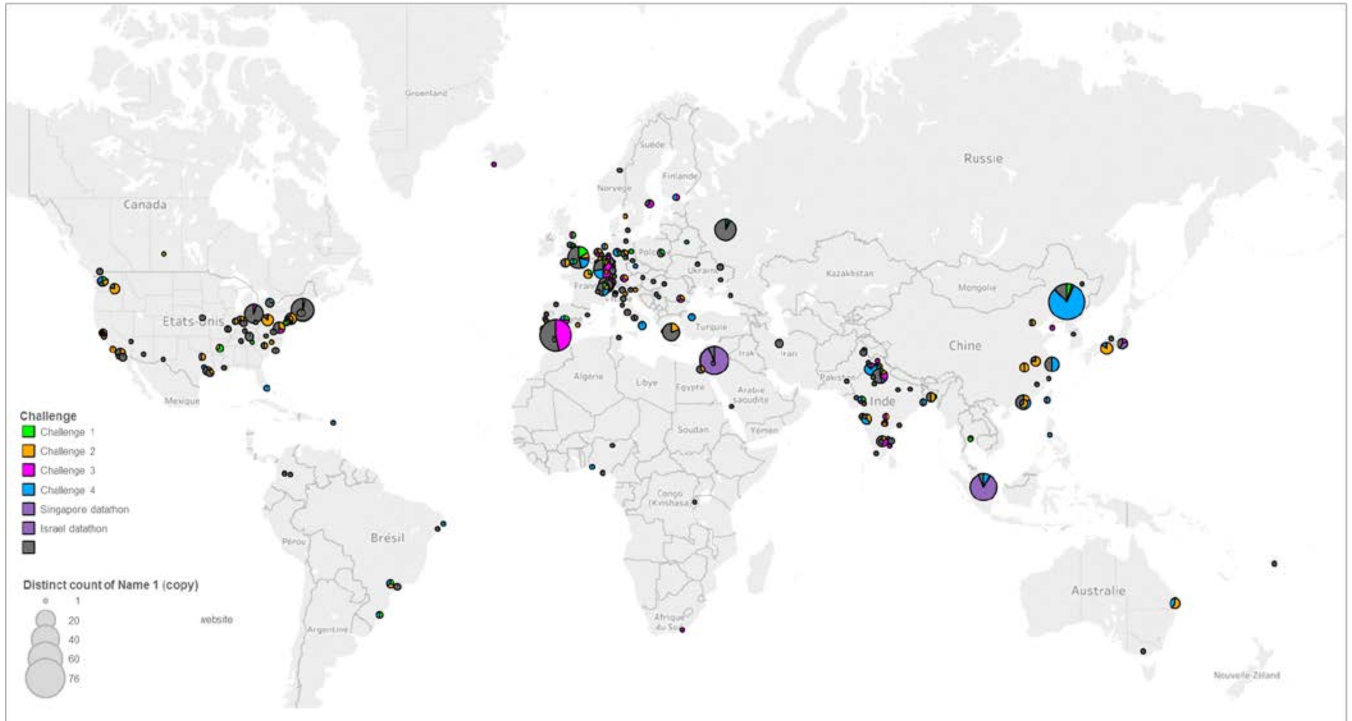
5-10 experts

Live phase: 6-8 weeks
Lead time: 4-6 months

**SHARE**
**LEARN**
**CONTRIBUTE**
**COLLABORATE**
**COMPETE**
**PUBLISH**

# sbv IMPROVER – Achievements to date

- 4 challenges
- 2 datathons
- > 600 scientists
- ~15 scientific publications in peer reviewed journals



Map based on Longitude (generated) and Latitude (generated). Color shows details about Challenge. Size shows distinct count of Name 1 (copy). Details are shown for Country and City. The view is filtered on Country and Challenge. The Country filter keeps 62 of 62 members. The Challenge filter keeps 7 of 7 members.

**Transparency**
- BioSpectrum Asia
- Technology Networks
- Tobacco Reporter
- Technology Networks

**SysTox Computational Challenge**
- American Laboratory
- Bio IT World
- BioSpectrum Asia
- News Medical
- Clinical Omics
- Front Line Genomics

**Network Verification Challenge 2**
- American Laboratory
- BioSpectrum Asia
- Asia Pacific Biotech News
- American Laboratory
- Biospectrum Asia

**Network Verification Challenge 1**
- Asia Pacific Biotech News
- Science 2.0
- UPO
- Genome Web
- BioSpectrum Asia
- HT Syndication
- EN CPhI
- Bio IT World
- Asia Pacific Biotech News
- Drug Discovery News
- American Laboratory

**Species Translation Challenge**
- Bio IT World
- GEN
- American Laboratory
- About.com
- Technonomy
- American Laboratory
- Noozhawk
- Genome Web
- American Laboratory
- Bio IT World

• 2017

• 2016

**Network Verification Challenge 2**
- News Medical
- Drug Discovery Today
- American Laboratory
- Libertas Academica Blog
- Drug Discovery News

**Datathon Singapore**
- Technology Networks
- American Laboratory

• 2015

**SysTox Computational Challenge**
- Bio IT World
- Genome Web

• 2014

• 2013

**Network Verification Challenge 1**
- American Laboratory
- BioSpectrum Asia
- Genome Web
- Bio IT World
- Pharmaceutical Marketing
- Drug Discovery News
- Science Reporter

• 2012

**Diagnostic Signature Challenge**
- Genome Web
- Genome Web
- Bio IT World

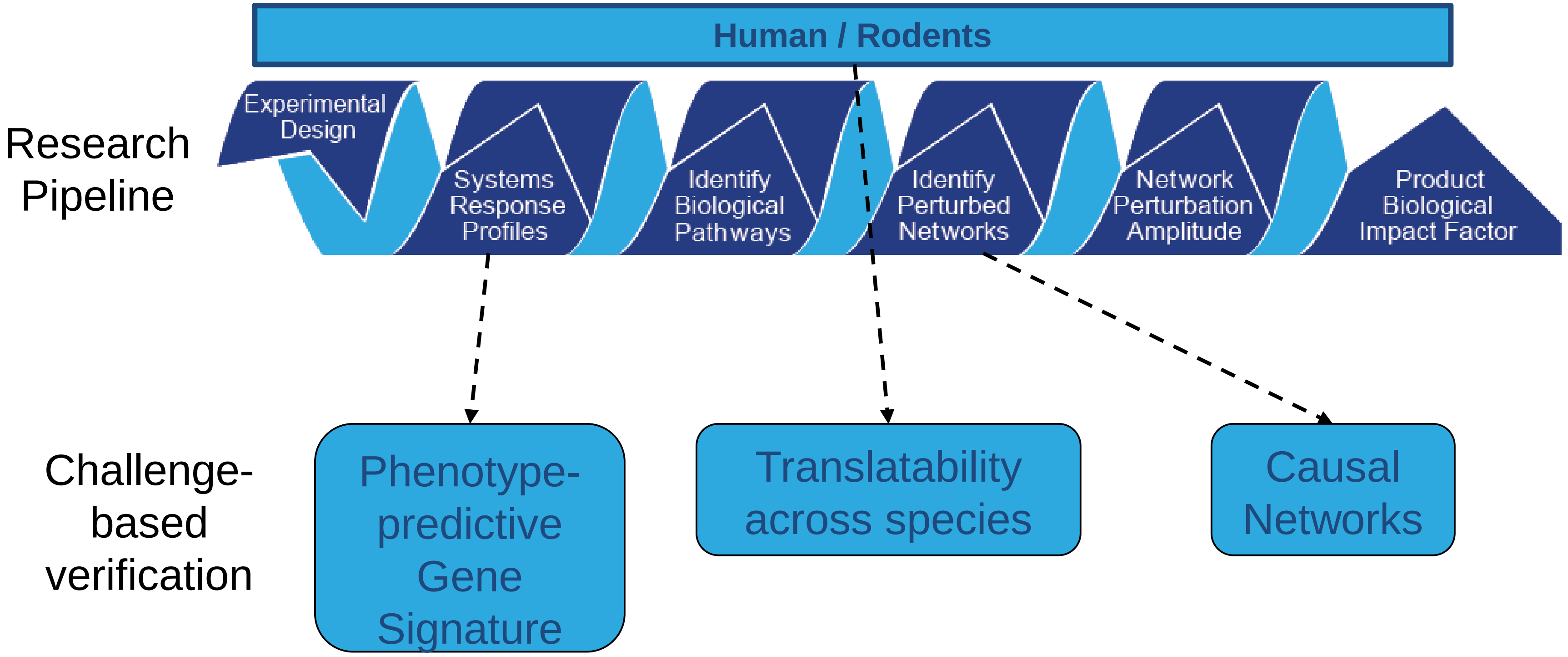
Challenge annoucement

✓ Challenge outcome

# Outline

- Introduction: PMI R&D & Systems Toxicology
- Transparency and verification in science: INTERVALS & sbv IMPROVER
- **Past and active challenges**: diagnostic signature challenge, species translation challenge, network verification challenge, systems toxicology challenge
- Datathons: collaborate to innovate
- Next computational challenge: microbiomics challenge

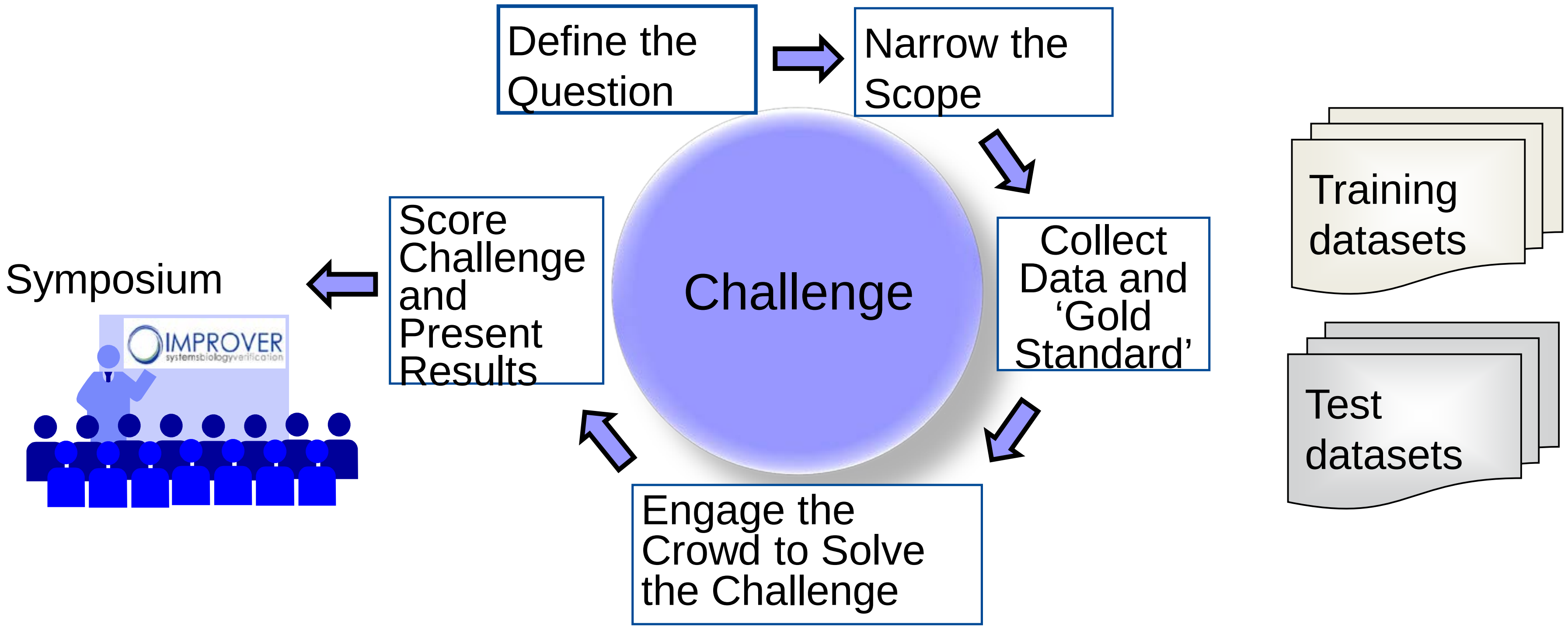
# Complex industrial research pipeline/Workflow divided into verifiable building blocks



Building blocks support each other towards a final goal

Each building block is verifiable by a challenge

# The elements of a challenge



Define the Question

Narrow the Scope

Collect Data and 'Gold Standard'

Engage the Crowd to Solve the Challenge

Score Challenge and Present Results

Challenge

Symposium

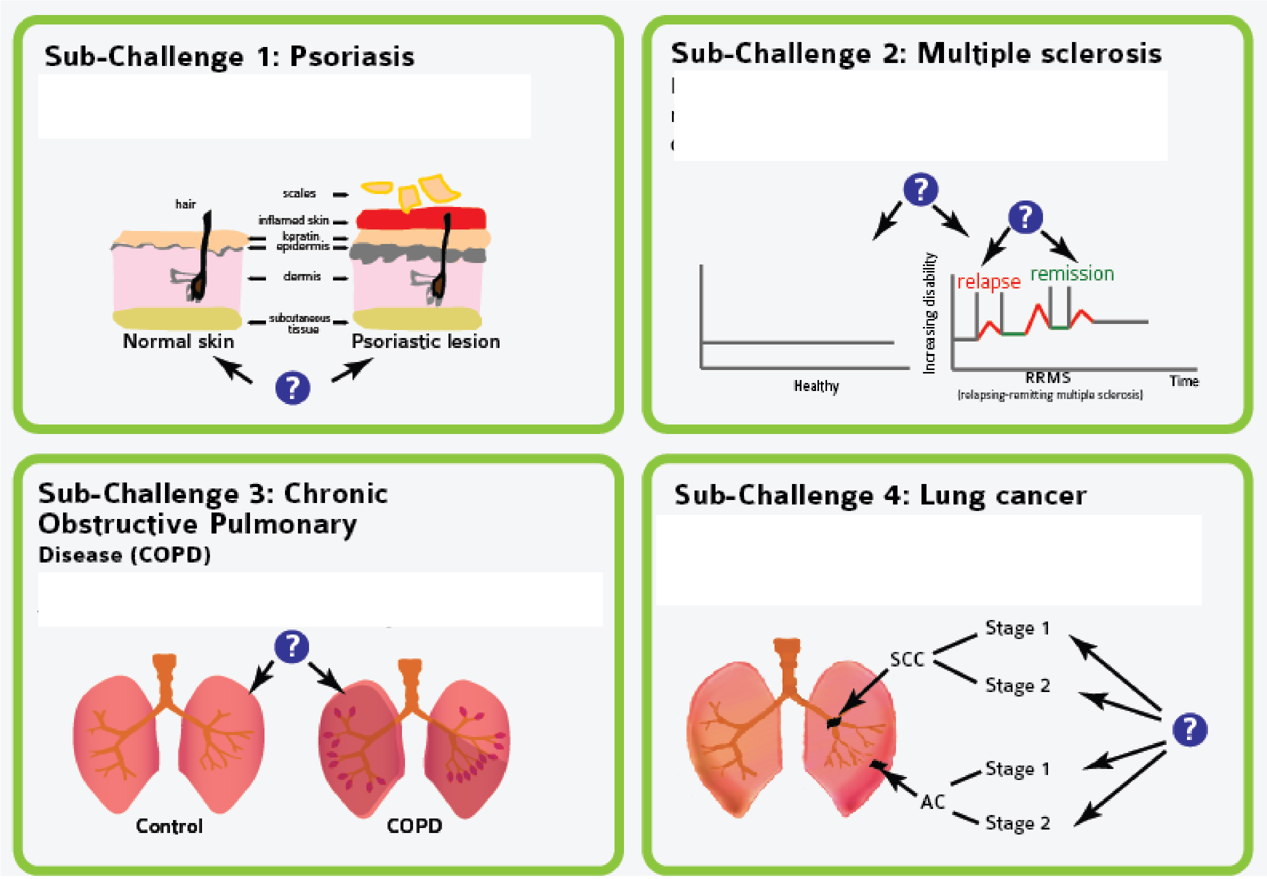
Training datasets

Test datasets

Crowd sourcing brings new ideas and leverages the wisdom of crowds

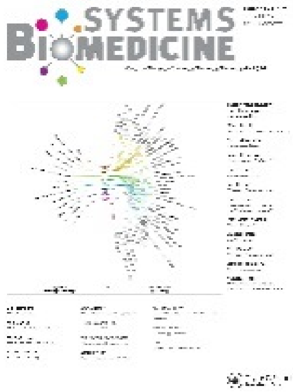
Clear challenge description and forum discussion in a user friendly website

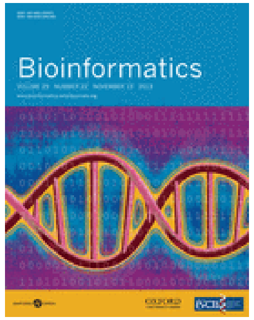
# Diagnostic Signature Challenge (2012)



Aim: Verify computational approaches that classify clinical samples based on transcriptomics data in 4 disease areas.

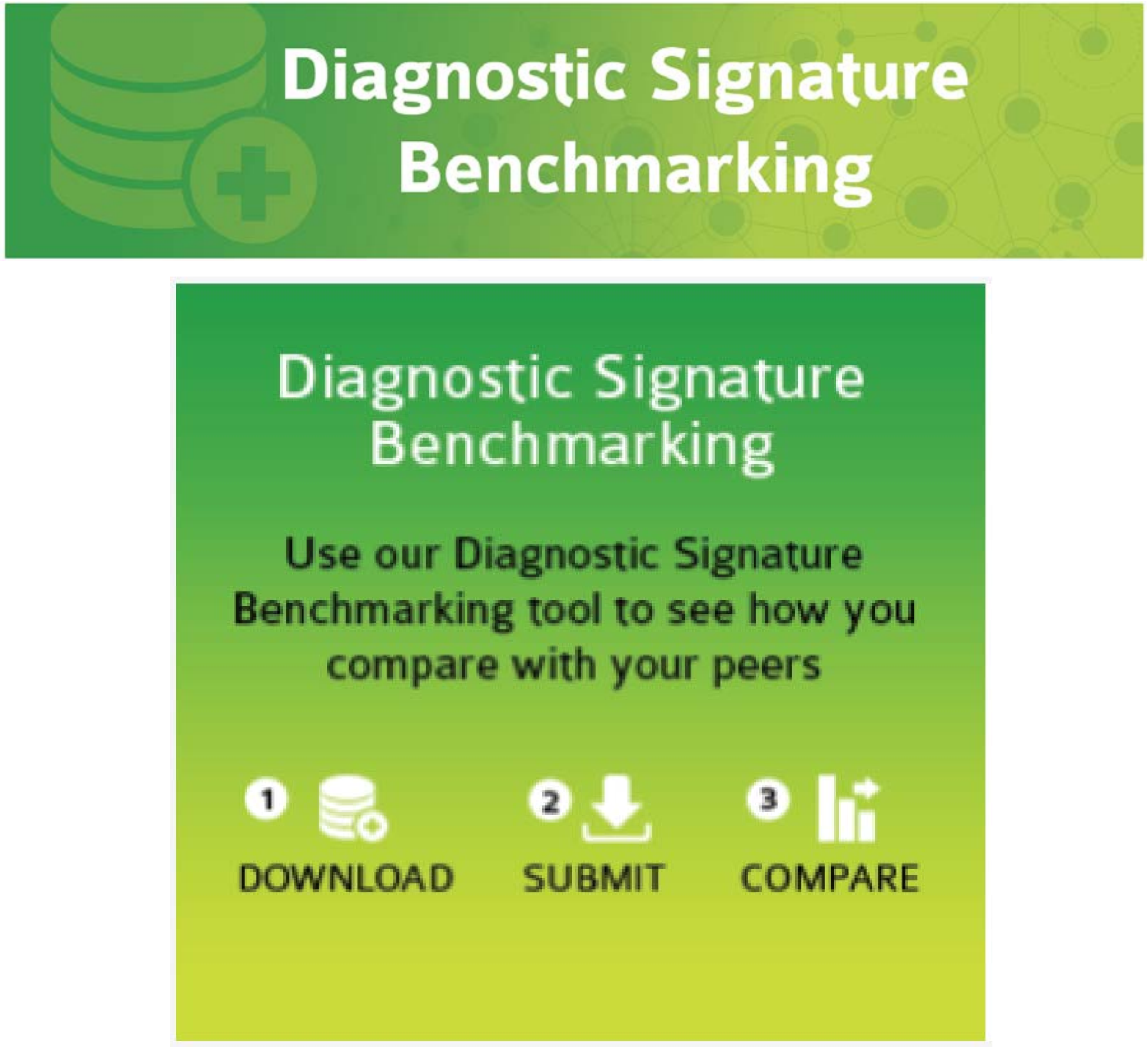
The Diagnostic Signature Challenge was successfully closed at the end of 2012 after 54 scientific teams from across the world had participated.
The overall quality of predictions was good and confirmed the predictive value of the approach.


SYSTEMS BIOMEDICINE


Bioinformatics

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

# Diagnostic Signature Challenge – Continue to benchmark your methods!



Diagnostic Signature Benchmarking

Diagnostic Signature Benchmarking

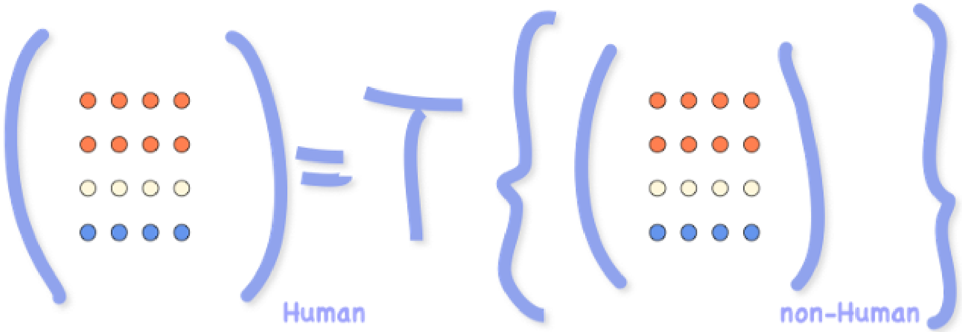
Use our Diagnostic Signature Benchmarking tool to see how you compare with your peers

Use our free web-based Diagnostic Signature Benchmarking tool to self-assess how well your method is able to classify clinical samples based on transcriptomics data and compare your results with the ones of your peers.
[https://sbvimprover.com/challenge-1](https://sbvimprover.com/challenge-1)

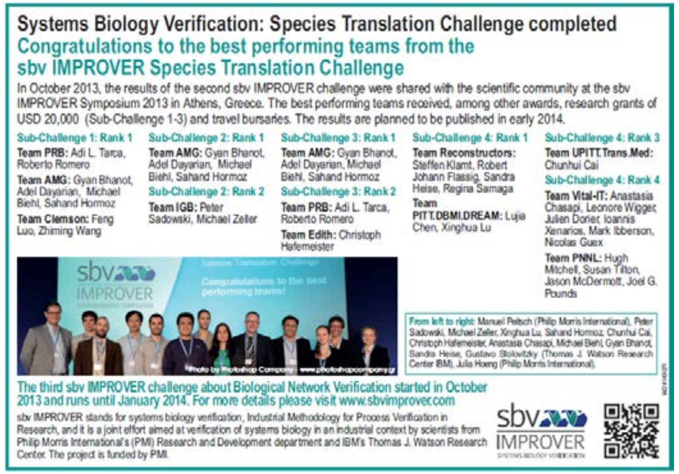
PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

# Species Translation Challenge (2013)

- Identify a function which maps measurements derived from systematic perturbations in one species to another
- Understand the system boundaries of the translatability concept
- Quantify the translatability between species



Best-performers announcement



**Systems Biology Verification: Species Translation Challenge completed**

**Congratulations to the best performing teams from the sbv IMPROVER Species Translation Challenge**

In October 2013, the results of the second sbv IMPROVER challenge were shared with the scientific community at the sbv IMPROVER Symposium 2013 in Athens, Greece. The best performing teams received, among other awards, research grants of USD 20,000 (Sub-Challenge 1-3) and travel bursaries. The results are planned to be published in early 2014.

**Sub-Challenge 1: Rank 1**

**Team PRB:** Adi L. Tarca, Roberto Romero

**Team AMG:** Gyan Bhanot, Adel Dayarian, Michael Biehl, Sahand Hormoz

**Team Clemson:** Feng Luo, Zhiming Wang

**Sub-Challenge 2: Rank 1**

**Team AMG:** Gyan Bhanot, Adel Dayarian, Michael Biehl, Sahand Hormoz

**Sub-Challenge 2: Rank 2**

**Team IGB:** Peter Sadowski, Michael Zeller

**Sub-Challenge 3: Rank 1**

**Team AMG:** Gyan Bhanot, Adel Dayarian, Michael Biehl, Sahand Hormoz

**Sub-Challenge 3: Rank 2**

**Team PRB:** Adi L. Tarca, Roberto Romero

**Team Edith:** Christoph Hafemeister

**Sub-Challenge 4: Rank 1**

**Team Reconstructors:** Steffen Klamt, Robert Johann Flassig, Sandra Heise, Regina Samaga

**Team PITT.DBMI.DREAM:** Lujia Chen, Xinghua Lu

**Sub-Challenge 4: Rank 3**

**Team UPITT.Trans.Med:** Chunhui Cai

**Sub-Challenge 4: Rank 4**

**Team Vital-IT:** Anastasia Chasapi, Leonore Wigger, Julien Dorier, Ioannis Xenarios, Mark Ibberson, Nicolas Guex

**Team PNNL:** Hugh Mitchell, Susan Tilton, Jason McDermott, Joel G. Pounds

Photo by Photoshop Company - www.photoshopcompany.gr

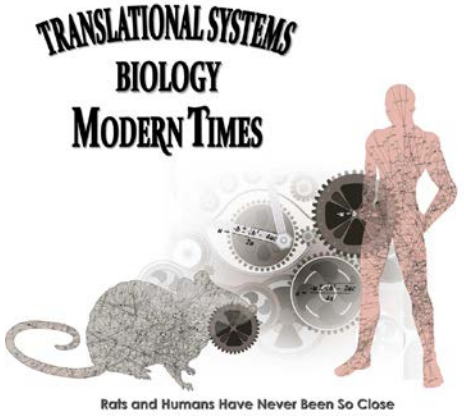
From left to right: Manuel Peitsch (Philip Morris International), Peter Sadowski, Michael Zeller, Xinghua Lu, Sahand Hormoz, Chunhui Cai, Christoph Hafemeister, Anastasia Chasapi, Michael Biehl, Gyan Bhanot, Sandra Heise, Gustavo Stolovitzky (Thomas J. Watson Research Center IBM), Julia Hoeng (Philip Morris International).

**The third sbv IMPROVER challenge about Biological Network Verification started in October 2013 and runs until January 2014. For more details please visit www.sbvimprover.com**

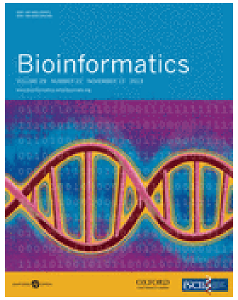
sbv IMPROVER stands for systems biology verification, Industrial Methodology for Process Verification in Research, and it is a joint effort aimed at verification of systems biology in an industrial context by scientists from Philip Morris International's (PMI) Research and Development department and IBM's Thomas J. Watson Research Center. The project is funded by PMI.

sbv IMPROVER

*Nature, 21 Nov. 2013*



- Symposium in Athens,Greece 2013
- Special Issue:

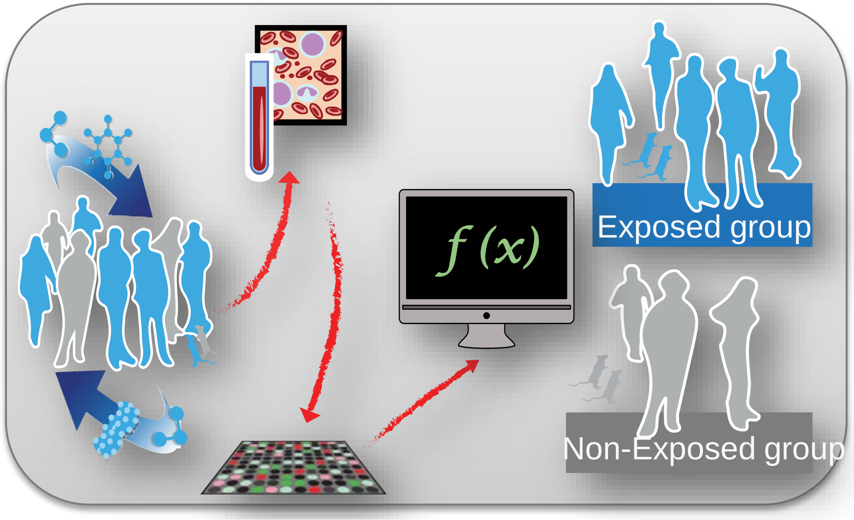


SCIENTIFIC DATA

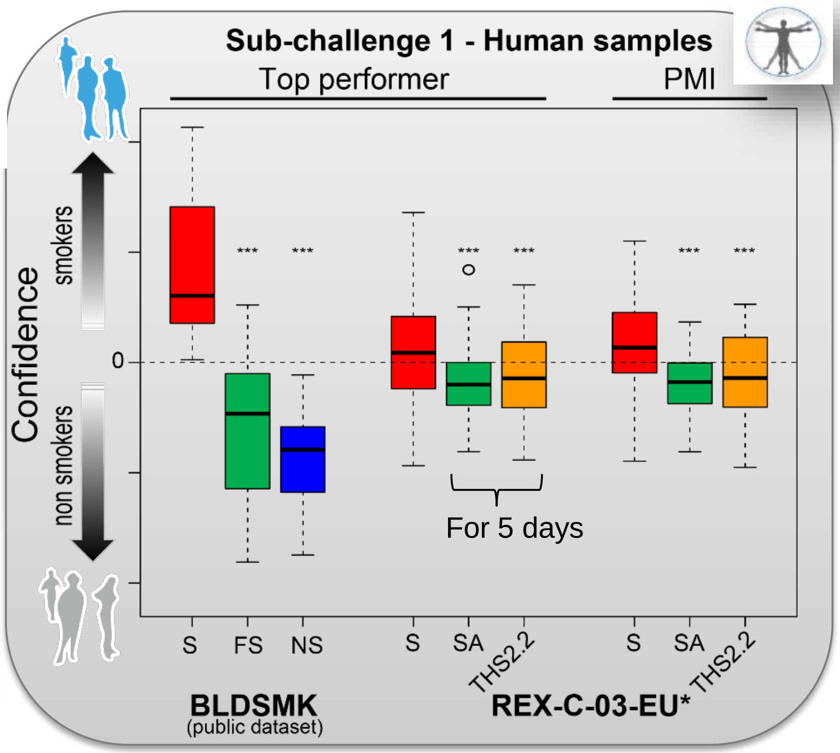
PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

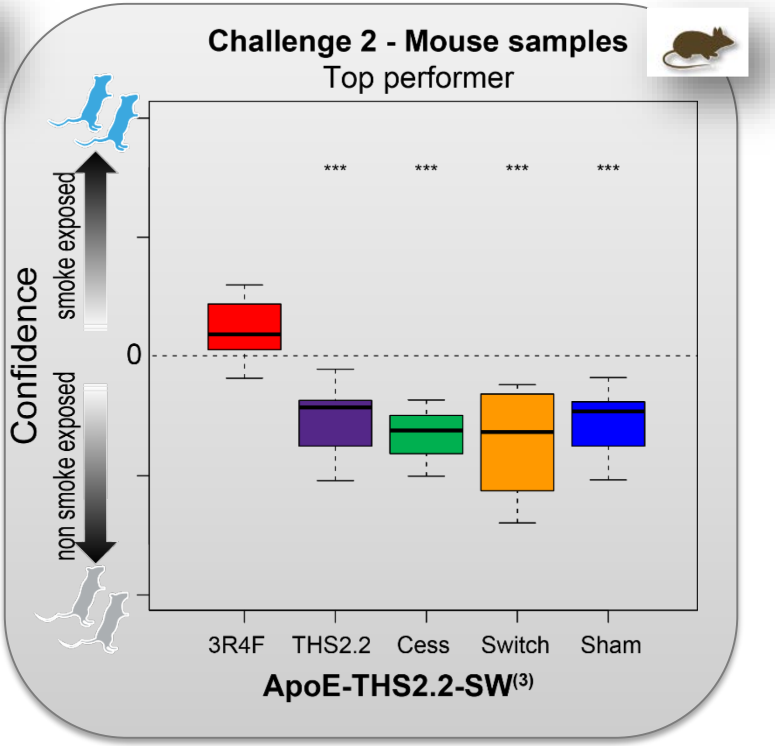
# Whole blood exposure marker: Crowdsourcing validation of PMI's gene expression signature

## sbv IMPROVER Systems Toxicology Computational Challenge(1)



- 135 registered participants
- 61 international teams
- 23 valid submissions





Gene Signature(2)

***LRRN3*, *SASH1*, *PALLD*, *RGL1*, *TNFRSF17*, *CDKN1C*, *IGJ*, *RRM2*, *ID3*, *SERPING1*, *FUCA1***

*In black genes common to the crowdsourcing signature***

S / 3R4F: Smoker / 3R4F exposed
FS / Cess: Former Smoker / Cessation
NS / Sham: Never Smoker / exposed to fresh air

SA: Smoking Abstinence
THS2.2: THS2.2

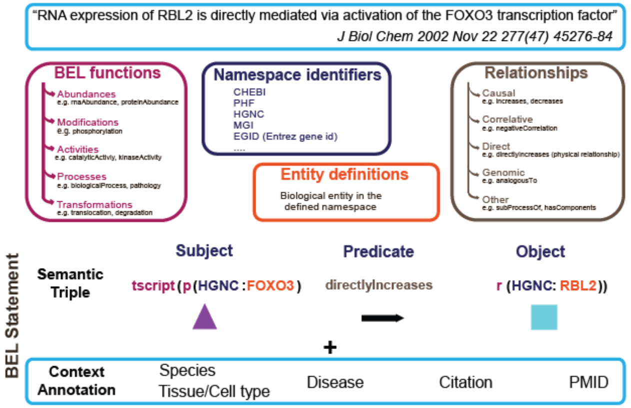
Switch: Switch to THS2.2

(1) *Poussin et al. Chemical research in toxicology. 2017*
(2) *Martin et al. Human & Experimental Toxicology. 2015*
(3) *Phillips et al. Toxicol Sci. 2016*

** similar results were obtained for REX-C-04-JP (not shown)*
*** selected by at least 2 teams*
**** significantly different from CC smoke exposed group*

# Leveraging biological knowledge from scientists for network models verification and refinement

## A. The Biological Expression Language



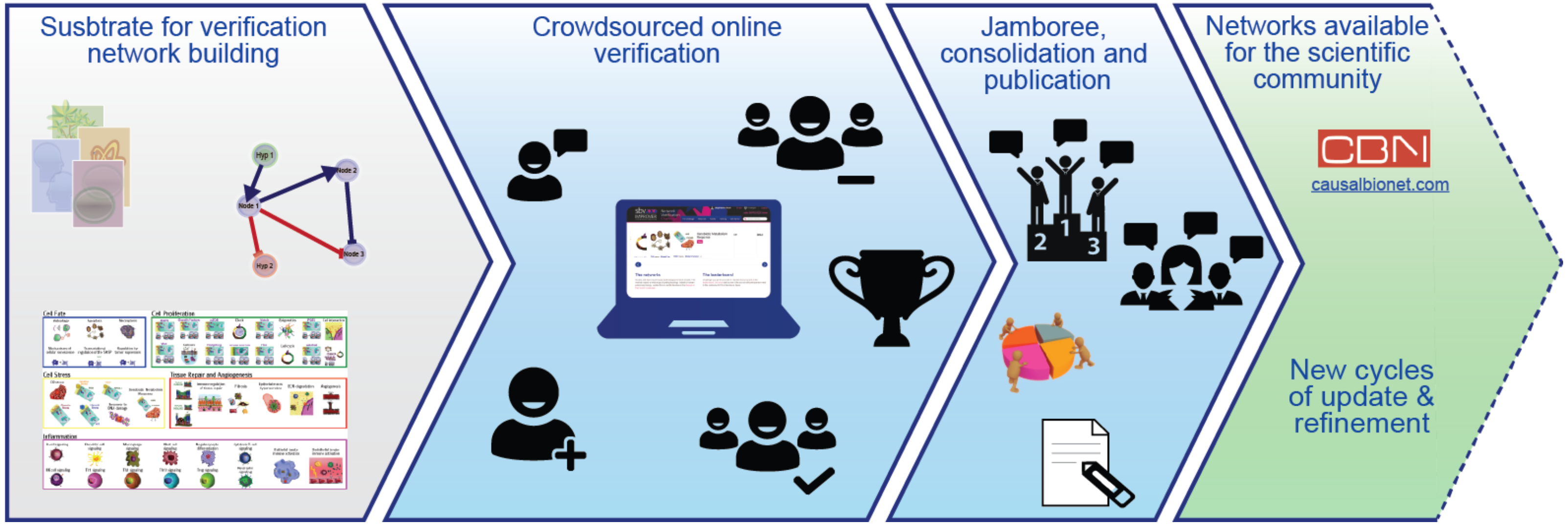
## B. The Bionet platform for crowd verification

1 online platform

- ▶ Challenge explanation
- ▶ Network visualization
- ▶ Evidence visualization
- ▶ Comment and vote on evidence
- ▶ Create new edge / new supporting evidence
- ▶ Leaderboard / badges
- ▶ Activity feed

Participate at:
bionet.sbvimprover.com

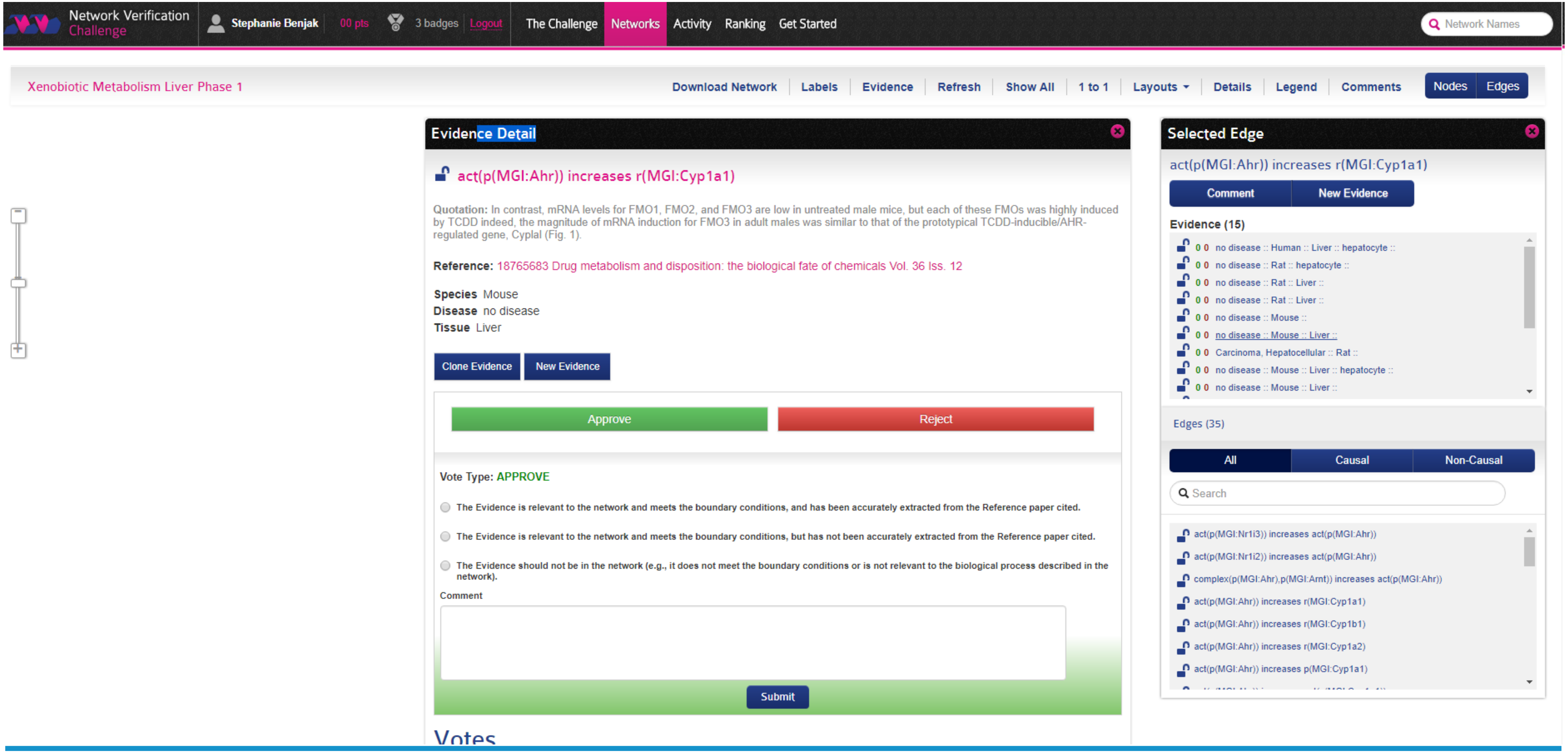
## C. The network verification challenge (NVC)



PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

From: Crowdsourcing and curation: perspectives from biology and natural language processing
Database (Oxford). 2016;2016. doi:10.1093/database/baw115

# Models in Bionet for refinement/verification



act(p(MGI:Ahr)) increases r(MGI:Cyp1a1)

**Quotation:** In contrast, mRNA levels for FMO1, FMO2, and FMO3 are low in untreated male mice, but each of these FMOs was highly induced by TCDD indeed, the magnitude of mRNA induction for FMO3 in adult males was similar to that of the prototypical TCDD-inducible/AHR-regulated gene, Cyplal (Fig. 1).

**Reference:** 18765683 Drug metabolism and disposition: the biological fate of chemicals Vol. 36 Iss. 12

**Species** Mouse
**Disease** no disease
**Tissue** Liver

**Vote Type:** APPROVE

○ The Evidence is relevant to the network and meets the boundary conditions, and has been accurately extracted from the Reference paper cited.

○ The Evidence is relevant to the network and meets the boundary conditions, but has not been accurately extracted from the Reference paper cited.

○ The Evidence should not be in the network (e.g., it does not meet the boundary conditions or is not relevant to the biological process described in the network).

Comment

Votes

act(p(MGI:Ahr)) increases r(MGI:Cyp1a1)

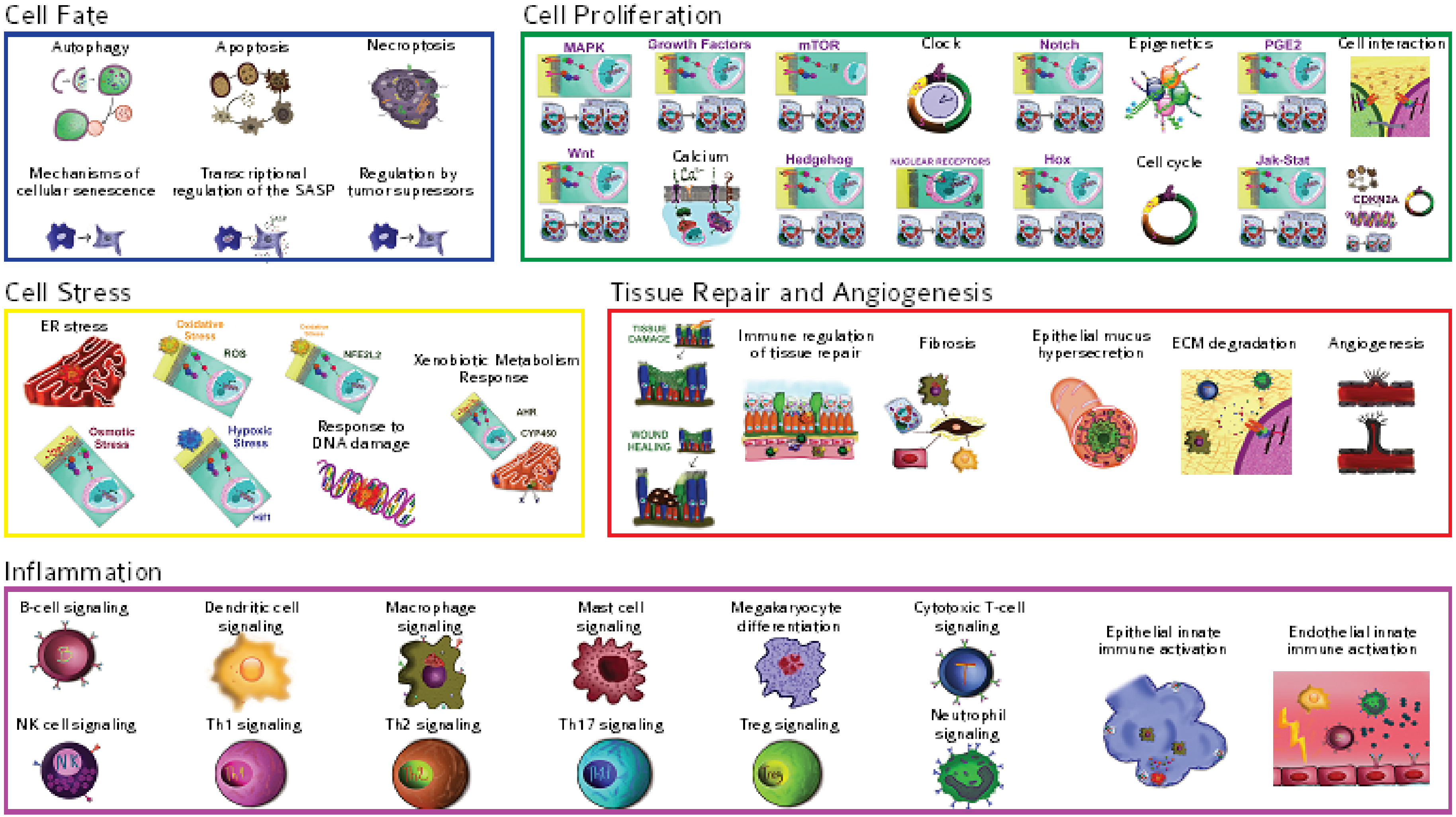
**Evidence (15)**

0 0 no disease :: Human :: Liver :: hepatocyte ::
0 0 no disease :: Rat :: hepatocyte ::
0 0 no disease :: Rat :: Liver ::
0 0 no disease :: Rat :: Liver ::
0 0 no disease :: Mouse ::
0 0 no disease :: Mouse :: Liver ::
0 0 Carcinoma, Hepatocellular :: Rat ::
0 0 no disease :: Mouse :: Liver :: hepatocyte ::
0 0 no disease :: Mouse :: Liver ::

Edges (35)

act(p(MGI:Nr1i3)) increases act(p(MGI:Ahr))
act(p(MGI:Nr1i2)) increases act(p(MGI:Ahr))
complex(p(MGI:Ahr),p(MGI:Arnt)) increases act(p(MGI:Ahr))
act(p(MGI:Ahr)) increases r(MGI:Cyp1a1)
act(p(MGI:Ahr)) increases r(MGI:Cyp1b1)
act(p(MGI:Ahr)) increases r(MGI:Cyp1a2)
act(p(MGI:Ahr)) increases p(MGI:Cyp1a1)

# Network models included in Network Verification Challenges 1 and 2

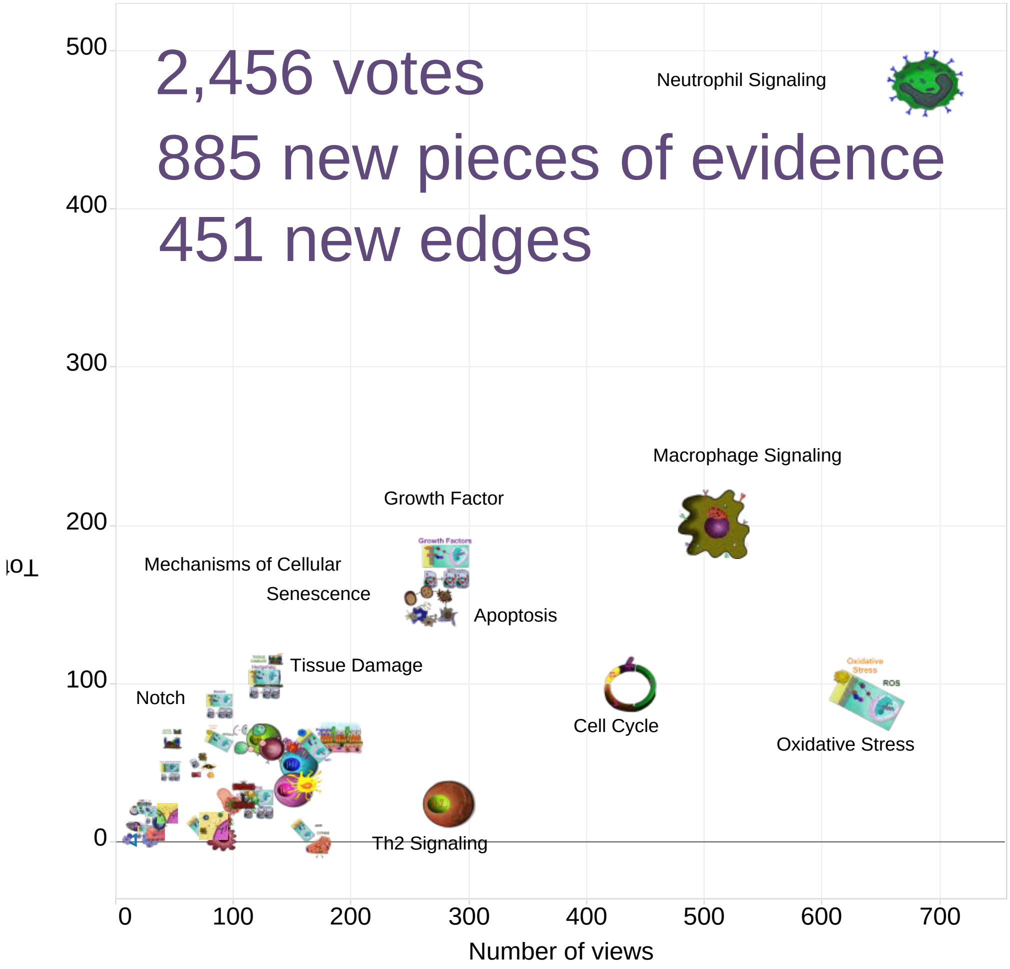


Models can be downloaded from [www.causalbionet.com](http://www.causalbionet.com)

Boué S, Talikka M, Westra JW, Hayes W, Di Fabio A, Park J, et al. Causal biological network database: a comprehensive platform of causal biological network models focused on the pulmonary and vascular systems. Database. 2015;2015:bav030.

# The first network verification challenge in numbers

## Activity during the open phase (10/2013 – 02/2014)



2,456 votes

885 new pieces of evidence

451 new edges

## 150 participants from 18 countries

**26 Best Performers invited to the Jamboree in Montreux, Switzerland with a travel bursary.**

- The sbv IMPROVER project team and the Challenge Best Performers. *Enhancement of COPD biological networks using a web-based collaboration interface. F1000Research.* ***4: 32 (2015)***
- Binder J, Boue S, DI Fabio A, et al. *Reputation-based collaborative network biology. Pacific Symposium on Biocomputing Pacific Symposium on Biocomputing*. ***270-81 (2014)***

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

# The second network verification challenge in numbers

Activity during the open phase
(04/2014 – 05/2015)

9,286 votes

2,225 new pieces of evidence

1,289 new edges

1,000 new nodes

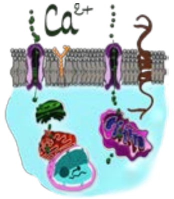
173 participants
from 26 countries

**20 Best Performers invited to the Jamboree in Barcelona, Spain with a travel bursary.**

- sbv, IMPROVER team and challenge best performers. *Community-Reviewed Biological Network Models for Toxicology and Drug Discovery Applications. Gene regulation and systems biology* ***10, 51-66, (2016).***
- Hirschman, L. *et al. Crowdsourcing and curation: perspectives from biology and natural language processing. Database* ***2016, baw115, (2016).***

## Key networks

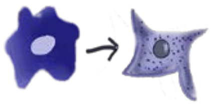
Calcium signaling



Macrophage signaling

Epigenetics

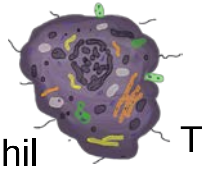
Senescence



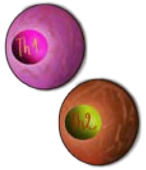
Oxidative Stress

Neutrophil Signaling

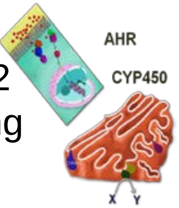
Necroptosis



Th1-Th2 Signaling



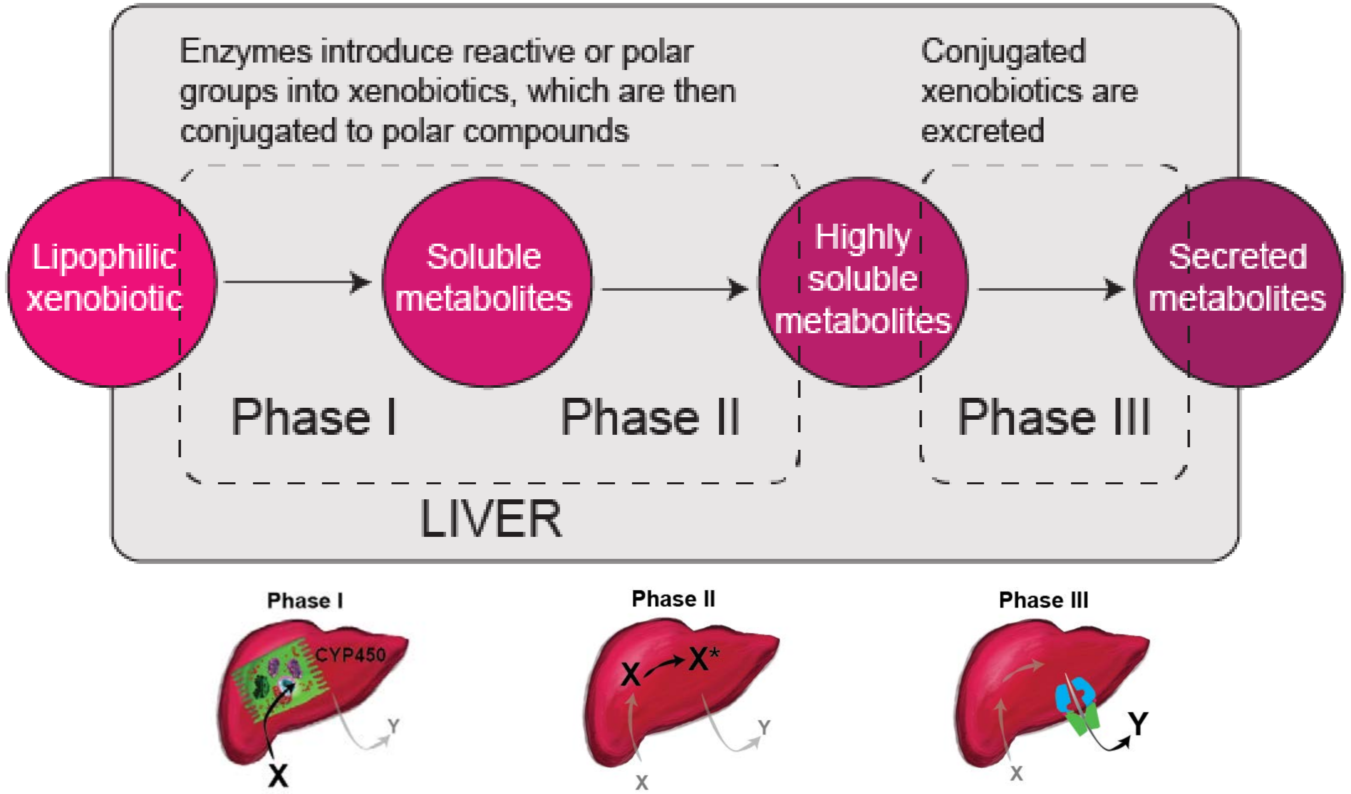
Xenobiotic Metabolism



PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

# New network models under review in NVC3

Collaborate. Contribute. Compete.



Enzymes introduce reactive or polar groups into xenobiotics, which are then conjugated to polar compounds

Conjugated xenobiotics are excreted

Lipophilic xenobiotic

Soluble metabolites

Highly soluble metabolites

Secreted metabolites

Phase I

Phase II

Phase III

LIVER

- **Verify and enhance** existing biological network models that will then be released to the community for use in research applications such as drug discovery, personalized medicine, and toxicological assessment.
- **Collaborate**: have fun competing and collaborating with others.
- **Learn** the Biological Expression Language, and use BELIEF, a curation tool to create BEL statements from text extracted from scientific publications.
- **Challenge your peers** and see in real time how you rank in the **leaderboard**.
- **Earn a gift card of 150 USD** when reaching 3000 points in the leaderboard (see Challenge rules on bionet.sbvimprover.com).
- As best performer, **win a travel grant** of up to 2,000 USD (see Challenge rules on bionet.sbvimprover.com).

# Outline

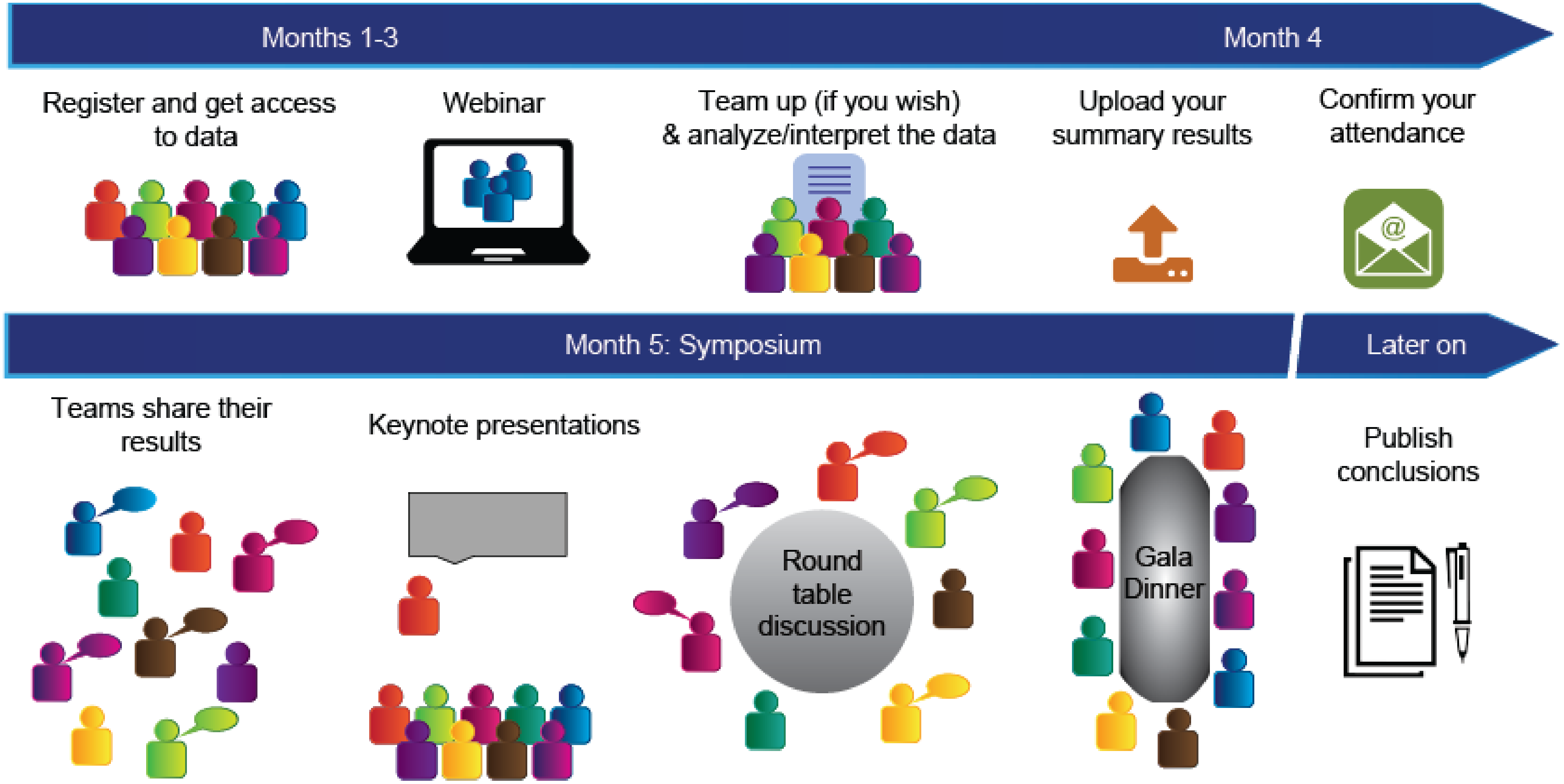
- **Introduction**: PMI R&D & Systems Toxicology
- **Transparency and verification in science**: INTERVALS & sbv IMPROVER
- **Past and active challenges**: diagnostic signature challenge, species translation challenge, network verification challenge, systems toxicology challenge
- **Datathons**: collaborate to innovate
- **Next computational challenge:** microbiomics challenge

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

# Tackling specific challenges as a Community

sbv IMPROVER
SYSTEMS BIOLOGY VERIFICATION

## Datathon scientific topic



Months 1-3

Month 4

Register and get access to data

Webinar

Team up (if you wish) & analyze/interpret the data

Upload your summary results

Confirm your attendance

Month 5: Symposium

Later on

Teams share their results

Keynote presentations

Round table discussion

Gala Dinner

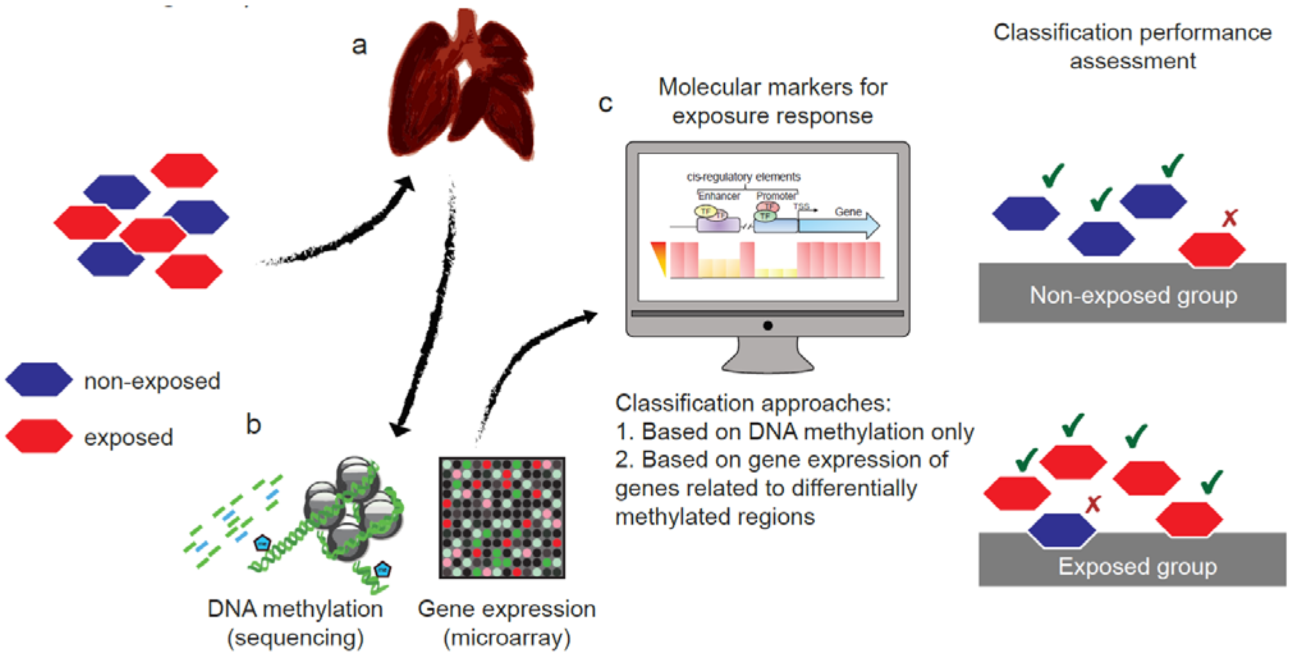
Publish conclusions

- Singapore, Oct 2016 - Development of methods for analysis of DNA methylation data
- Tel Aviv, May 2017 - Epigenetics challenge and networking symposium
- Tokyo, Oct 2017 - Biological interpretation of omics data & quantification of biological impact

# Epigenomics Challenge Israel

- The aim of this sbv IMPROVER Challenge was to **apply computational approaches** to assess the impact of tobacco smoke or aerosol in large **methylome** datasets obtained from rodent inhalation studies
- The challenge has culminated in a **multi-disciplinary symposium** in Tel Aviv, Israel which has attracted computational scientists, bioinformaticians and clinicians working across a variety of medical fields

*First prize winner Hagit Philip, Systems BioMedicine Lab, Faculty of Life Science, Bar-Ilan University*

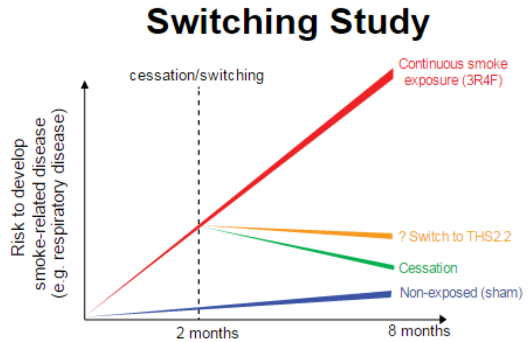


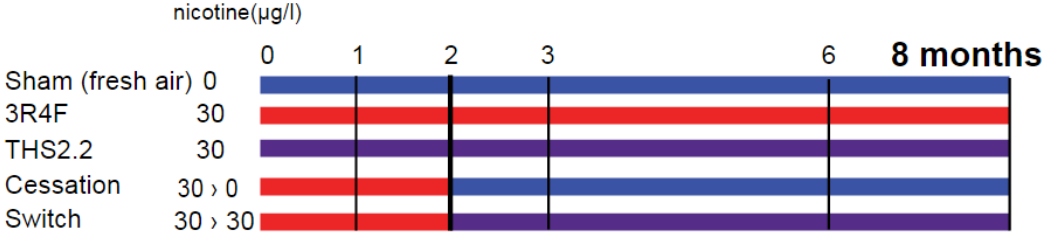
a. Samples of lung tissue were collected from mice belonging to the exposed or non-exposed groups.
b. Gene expression profiles (GEX) and DNA methylation (DNA-Meth) were measured using microarray based technology and Illumina sequencing by synthesis technology, respectively.
c. Participants are provided with GEX and DNA-Meth and asked to develop a classification approach capable of associating subjects to the correct exposure group

# Datathon Japan – Fall 2017 - Studies & Data

## Mouse cardiovascular & respiratory model – 8 month inhalation study

Apoe-/-





**Cardiovascular, respiratory, and liver -related measurements covering apical and systems toxicology endpoints**

SOT Society of Toxicology
www.toxsci.oxfordjournals.org

TOXICOLOGICAL SCIENCES

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

Blaine Phillips,* Emilija Veljkovic,† Stéphanie Boué,† Walter K. Schlage,‡ Gregory Vuillaume,† Florian Martin,† Bjoern Titz,† Patrice Leroy,† Ansgar Buettner,§ Ashraf Elamin,† Alberto Oviedo,* Maciej Cabanski,†,§ Héctor De León,† Emmanuel Guedj,† Thomas Schneider,† Marja Talikka,† Nikolai V. Ivanov,† Patrick Vanscheeuwijck,† Manuel C. Peitsch,† and Julia Hoeng,†,2

SOT Society of Toxicology
www.toxsci.oxfordjournals.org

TOXICOLOGICAL SCIENCES, 149(2), 2016, 441–457

Advance Access Publication Date: November 17, 2015
Research Article

ffects of Cigarette Smoke, Cessation, and Switching to Two Heat-Not-Burn Tobacco Products on Lung Lipid Metabolism in C57BL/6 and $Apoe^{-/-}$ Mice—An Integrative Systems Toxicology Analysis

Bjoern Titz,*,1 Stéphanie Boué,*,1 Blaine Phillips,† Marja Talikka,* Terhi Vihervaara,‡ Thomas Schneider,* Catherine Nury,* Ashraf Elamin,* Emmanuel Guedj,* Michael J. Peck,* Walter K. Schlage,* Maciej Cabanski,*,2 Patrice Leroy,* Gregory Vuillaume,* Florian Martin,* Nikolai V. Ivanov,* Emilija Veljkovic,* Kim Ekroos,‡ Reijo Laaksonen,‡ Patrick Vanscheeuwijck,* Manuel C. Peitsch,* and Julia Hoeng*,3

Inhalation Toxicology
International Forum for Respiratory Research

Effects of cigarette smoke, cessation and switching to a candidate modified risk tobacco product on the liver in $Apoe^{-/-}$ mice – a systems toxicology analysis

Giuseppe Lo Sasso, Bjoern Titz, Catherine Nury, Stéphanie Boué, Blaine Phillips, Vincenzo Belcastro, Thomas Schneider, Sophie Dijon, Karine Baumer, Daruisz Peric, Remi Dulize, Ashraf Elamin, Emmanuel Guedj, Ansgar Buettner, Patrice Leroy, Samuel Kleinhans, Gregory Vuillaume, Emilija Veljkovic, Nikolai V. Ivanov, Florian Martin, Patrick Vanscheeuwijck, Manuel C. Peitsch & Julia Hoeng

To cite this article: Giuseppe Lo Sasso, Bjoern Titz, Catherine Nury, Stéphanie Boué, Blaine Phillips, Vincenzo Belcastro, Thomas Schneider, Sophie Dijon, Karine Baumer, Dariusz Peric, Remi Dulize, Ashraf Elamin, Emmanuel Guedj, Ansgar Buettner, Patrice Leroy, Samuel Kleinhans, Gregory Vuillaume, Emilija Veljkovic, Nikolai V. Ivanov, Florian Martin, Patrick Vanscheeuwijck, Manuel C. Peitsch & Julia Hoeng (2016) Effects of cigarette smoke, cessation and switching to a candidate modified risk tobacco product on the liver in $Apoe^{-/-}$ mice – a systems toxicology analysis, Inhalation Toxicology, 28:5, 226-240, DOI: 10.3109/08958378.2016.1150368

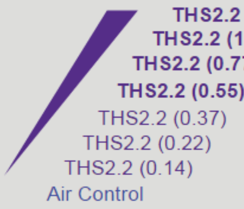
To link to this article: http://dx.doi.org/10.3109/08958378.2016.1150368

## Assessment of acute THS2.2 aerosol exposure in human nasal epithelial cultures

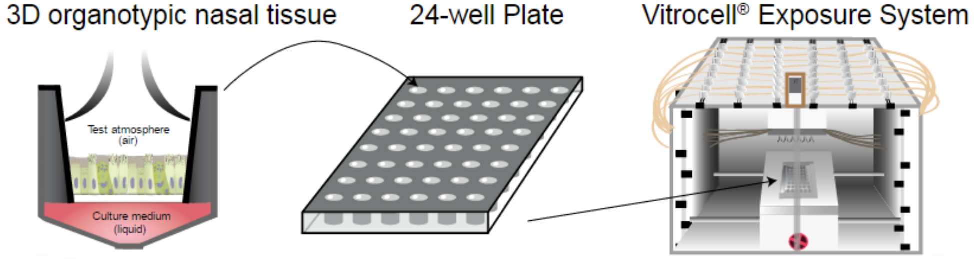
### 5 Repetitions - Parallel exposure

CYP1A1/1B1 activity
Secretion of mediators
Ciliary beating frequency
mRNA and miRNA profiles
AK-based cytotoxicity
Histology / Immunostaining

### 1 Repetition - THS2.2 Dose Range



AK-based cytotoxicity
Histology



Toxicology in Vitro 39 (2017) 29–51

Contents lists available at ScienceDirect

Toxicology in Vitro

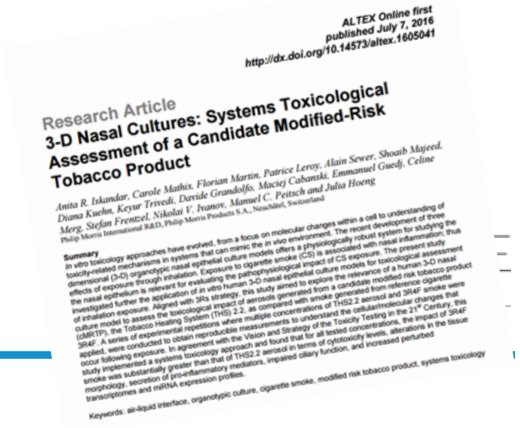
journal homepage: www.elsevier.com/locate/toxinvit

A systems toxicology approach for comparative assessment: Biological impact of an aerosol from a candidate modified-risk tobacco product and cigarette smoke on human organotypic bronchial epithelial cultures

Anita R. Iskandar, Carole Mathis, Walter K. Schlage, Stefan Frentzel, Patrice Leroy, Yang Xiang, Alain Sewer, Shoaib Majeed, Laura Ortega-Torres, Stephanie Johne, Emmanuel Guedj, Keyur Trivedi, Gilles Kratzer, Celine Merg, Ashraf Elamin, Florian Martin, Nikolai V. Ivanov, Manuel C. Peitsch, Julia Hoeng

ARTICLE INFO

ABSTRACT



ALTEX Online first
published July 7, 2016
http://dx.doi.org/10.14573/altex.1605041

Research Article

3-D Nasal Cultures: Systems Toxicological Assessment of a Candidate Modified-Risk Tobacco Product

Anita R. Iskandar, Carole Mathis, Florian Martin, Patrice Leroy, Alain Sewer, Shoaib Majeed, Diana Kuehn, Keyur Trivedi, Davide Grandolfo, Maciej Cabanski, Emmanuel Guedj, Celine Merg, Stefan Frentzel, Nikolai V. Ivanov, Manuel C. Peitsch and Julia Hoeng

Philip Morris International R&D, Philip Morris Products S.A., Neuchâtel, Switzerland

Summary

# Datathon – Summer/Fall 2017, Japan

Datasets (accessed through the systox.sbvimprover.com platform):

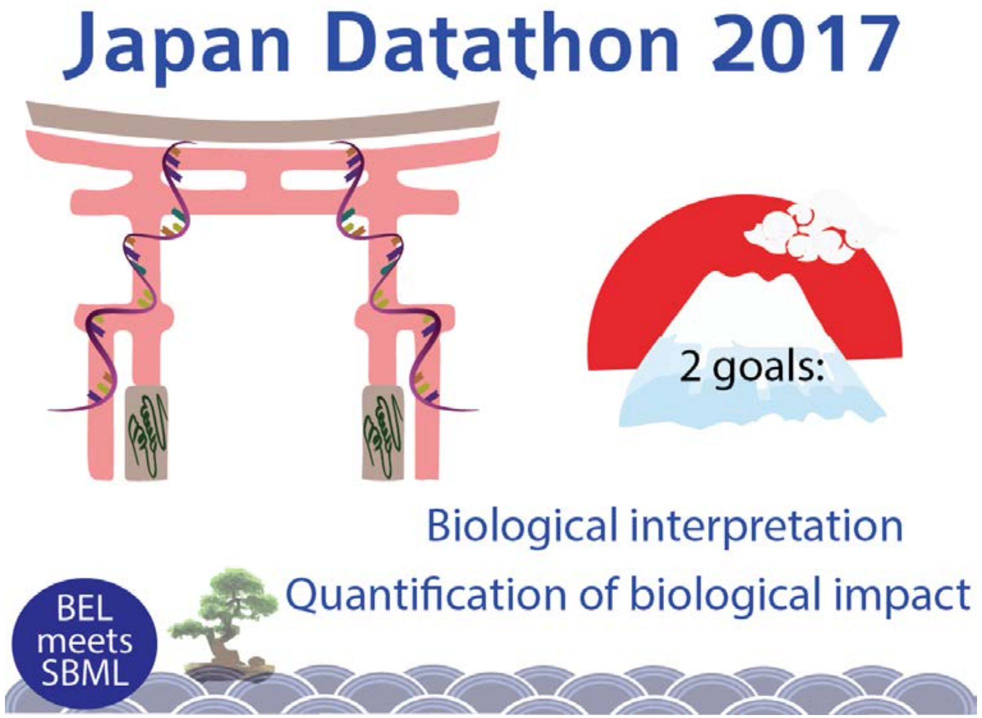
- Transcriptomics & proteomics datasets from the *in vivo* study (Apoe$^{-/-}$ 8 month inhalation)
- Transcriptomics dataset from the *in vitro* study (acute THS2.2 aerosol exposure of organotypic nasal epithelial cultures)

Questions:

- **What are the pathways/biological processes perturbed by exposure to cigarette smoke or to THS2.2 aerosol, a heat-not-burn product? After smoking cessation? After switching to THS2.2 aerosol?**

  ➔ Participants are expected to submit a write up with their biological interpretation
- **How would you quantify relative perturbations in the respective groups?**

  ➔ Innovative proposals will be conversed during the round table discussion
- Bonus track: Biological network models are at the center of systems biology. The Biological Expression Language (BEL) is a language for representing scientific findings in the life sciences in a computable form. SBML is a data format for encoding models of biological processes. What are the proposals/interfaces one might consider to convert between BEL statements and SBML models?

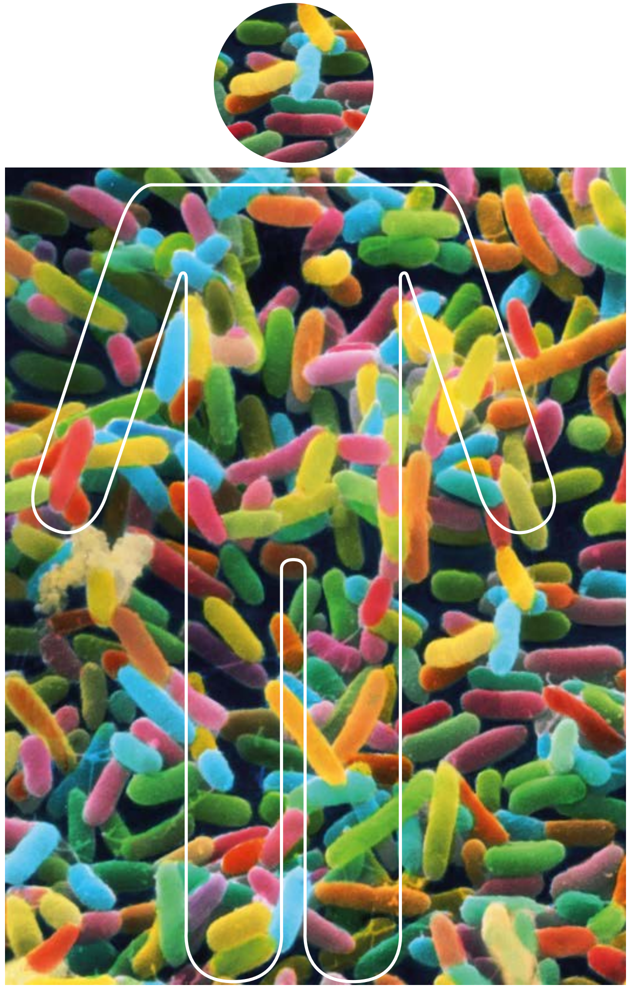


# Outline

- **Introduction**: PMI R&D & Systems Toxicology
- **Transparency and verification in science**: INTERVALS & sbv IMPROVER
- **Past and active challenges**: diagnostic signature challenge, species translation challenge, network verification challenge, systems toxicology challenge
- **Datathons**: collaborate to innovate
- **Next computational challenge:** microbiomics challenge

# Microbiome and host

- **Microbiome** is the **set of microorganisms** in an environment (e.g., the human body).
- The microbiome is **involved in multiple functions** and is known to **affect many physiological processes**.
- **Dysbiosis**, defined as the disturbance of microbial equilibrium, may lead to **aberrant physiological responses** in the host species.
- The identification of the **microbial composition** - or **taxonomic profile** – may be helpful in the diagnosis or prognosis of specific diseases.



PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

# Microbiome and technology

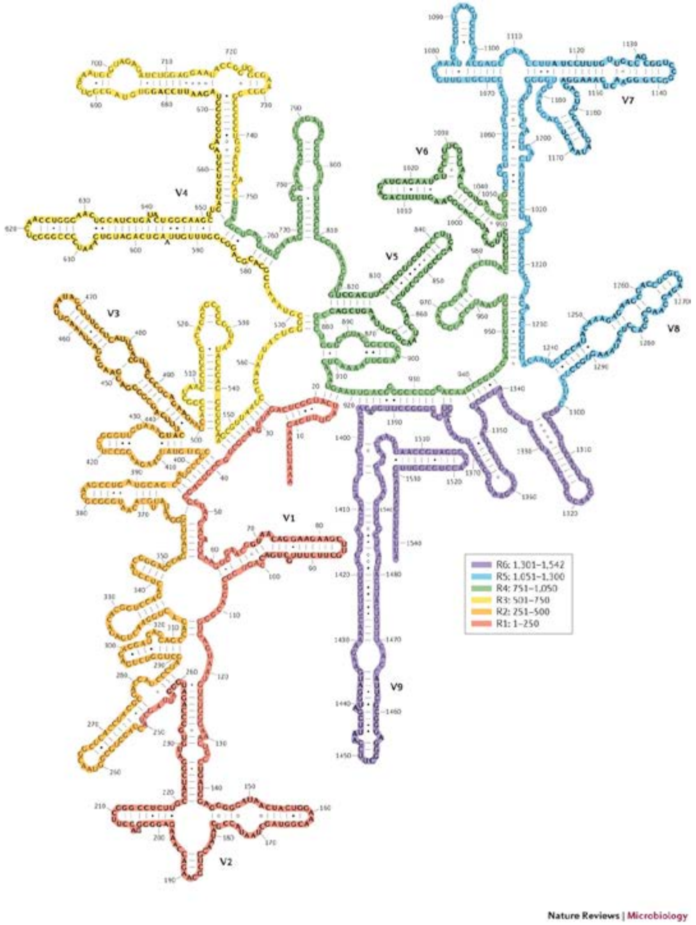
- “Metagenomics“ is defined as the study of DNA of a microbial community.
- Advances in genome sequencing technologies enabled a rapid expansion of microbial studies.
- Computational methods for the analysis of DNA samples are required to facilitate downstream biological data interpretation.

http://igoscience.com

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

# 16S vs shotgun sequencing (metagenetics vs metagenomics) – 1/2

- 16S and shotgun are two popular methods to identify microbial communities.
- 16S ribosomal RNA is a structural component of the prokaryotic ribosome.
- All prokaryotes have 16S that is required for protein synthesis.
- Why 16S rRNA is used for phylogeny reconstruction?
  - (i) slow rate of evolution, (ii) functional constancy, (iii) size
- The 16S rRNA gene is about ~1500bp long, it consists of 9 variable regions flanked by conserved regions.
- Variable regions evolve at faster rate allowing differentiation between taxa.

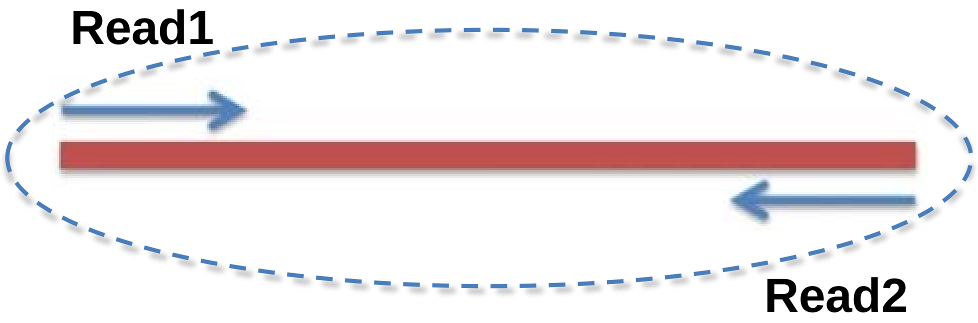


Yarza et al. Nature Reviews Microbiology, 2014

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

# 16S vs shotgun sequencing (metagenetics vs metagenomics) – 2/2

- Shotgun metagenomics utilizes sequencing data generated from random fragments from the whole genome.
- Shotgun metagenomics allows for species abundance determination and functional analysis.



# Shotgun data analysis (overview)

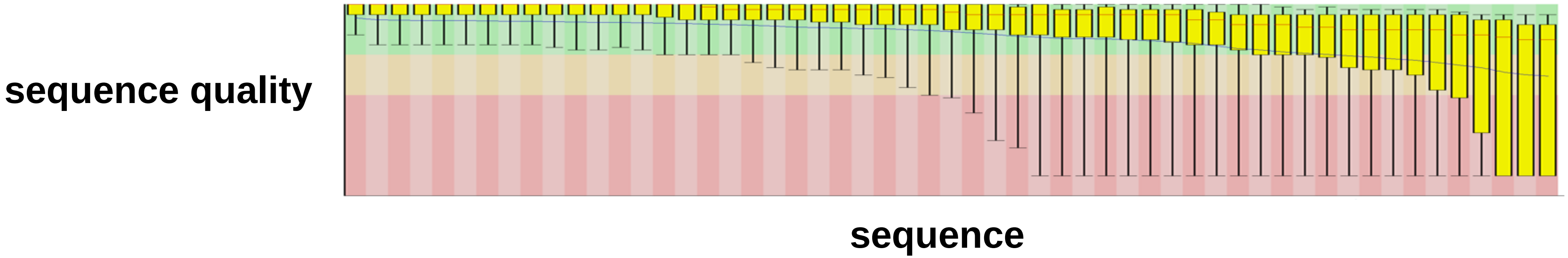
- Shotgun data analysis involves different steps. One of the possible analysis pipeline includes the following:
  - Sample collection and DNA extraction (wet lab)
  - Sequencing and raw data collection (dry lab)
  - ✓ Quality control data (include host contamination removal)
  - ✓ Metagenome assembly (only to annotate unknown genomes)
  - ✓ Taxonomic profiling: determine species abundances
  - ✓ Functional profiling: pathway abundance and completeness per species
  - ✓ Comparative analysis

# Shotgun data analysis (Quality control)

- Quality control is the process of remove bad quality reads (and host genome contamination)
- Different tools are available to perform QC (FastQC, Trimmomatic, Trim_galore)

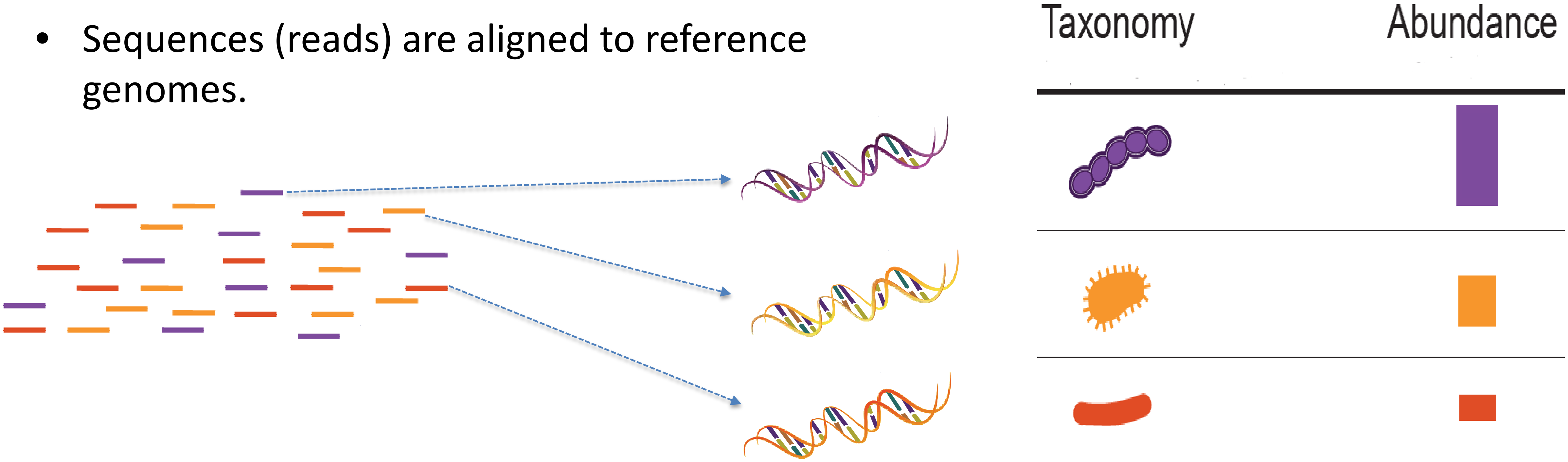
***fastq format entry***

```
@sequence001-R1
ATGGTTCTCTTAAGAATACGCTTTTGGAATTCTGGTAATTCATATGTGGGACGAGCATTCAAAATAATA…
+
=CCGGGGGGGGGGCCJGCCJJGJJJCtoto1GJ=GJJJJJJJGJJG=JGCJG=JJGGJGGJGC(GJGCGGC…
```

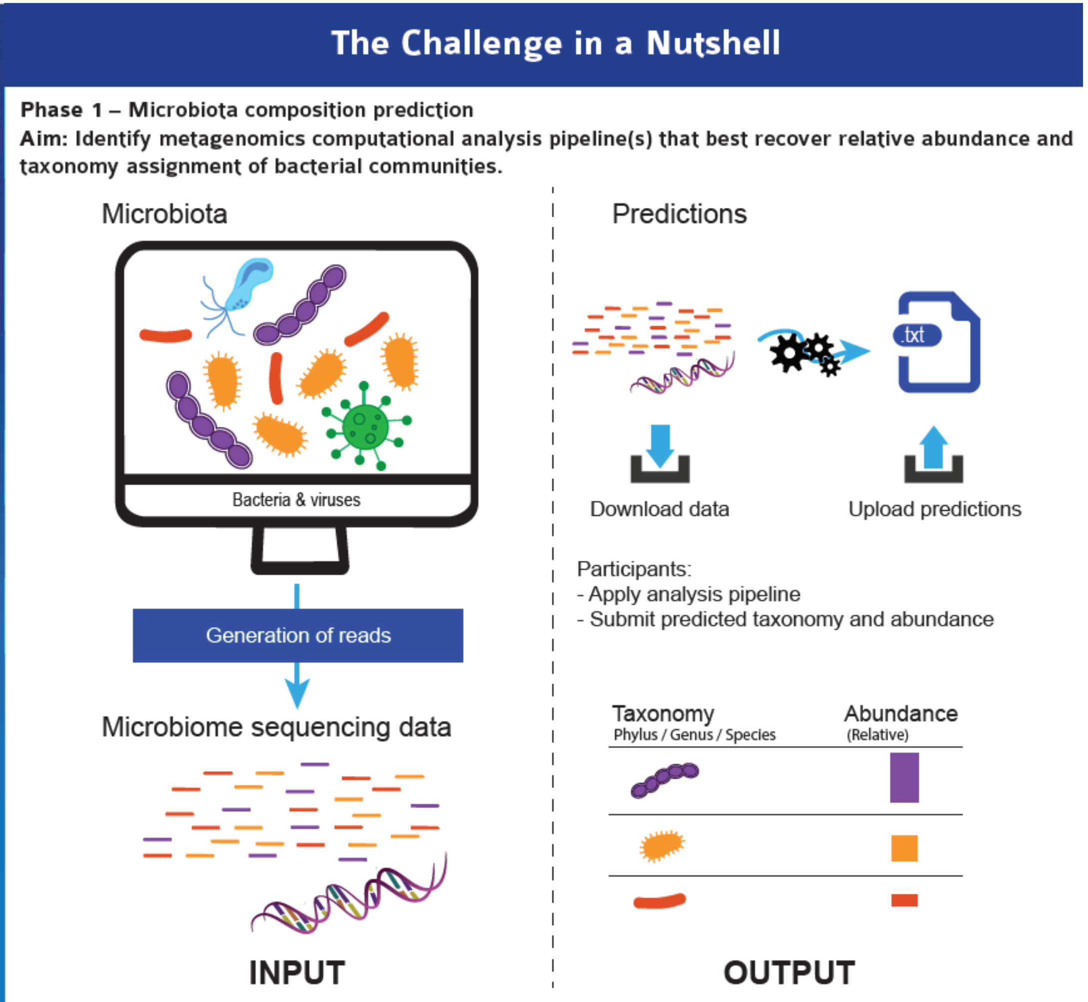


# Shotgun data analysis (Taxonomic profiling)

- Taxonomic profiling is the process of identifying species abundance.
- Sequences (reads) are aligned to reference genomes.



# Metagenomics computational challenge



## The Challenge in a Nutshell

**Phase 1 – Microbiota composition prediction**

**Aim: Identify metagenomics computational analysis pipeline(s) that best recover relative abundance and taxonomy assignment of bacterial communities.**

Microbiota

Bacteria & viruses

Generation of reads

Microbiome sequencing data

**INPUT**

Predictions

Download data

Upload predictions

Participants:
- Apply analysis pipeline
- Submit predicted taxonomy and abundance

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

**OUTPUT**

Submission of predictions — Jan 2018

Scoring against gold standard

Team ranking

Winners announcement — Feb 2018

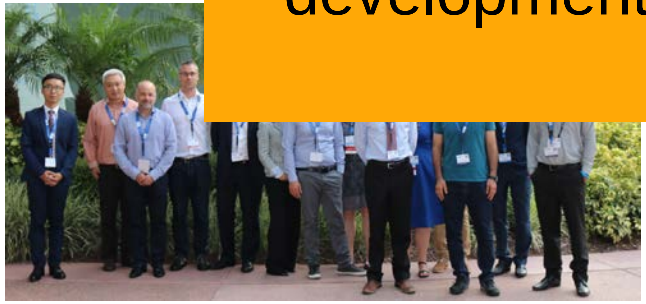
Challenge outcome publication

# Join the sbv IMPROVER community!

- **Challenges**: Microbiomics challenge (multiple phases), NVC3 (liver xenobiotic metabolism)
- **Datathons**: Biological interpretation of omics data (Japan)
- **Transparency in science**: release of new INTERVALS platform allowing the community to share data on RRPs, development/integration of analytics tools

Boston, 201

Athens, 20



Orlando, 2016

Singapore, 2016

Tel Aviv, 2017

# Curious? Want to learn more

## Posters

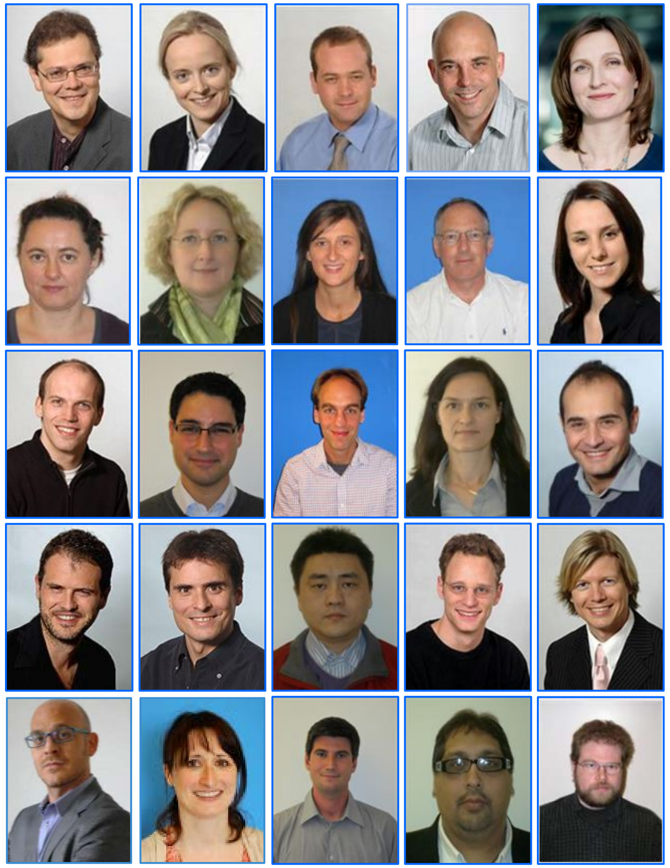
**Crowdsourced enhancement of causal network models – results from past network verification challenges and new application enabling liver phase I xenobiotic metabolism model refinement.**

Sunday 23rd July, 6:00pm - 7:00pm
#A-340

**High Content Screening data storage and analysis platform - An open source solution**

Saturday 22nd July 22, 6:00pm - 7:00pm
#A-117

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL



# Thank you!

SBX

SciPinion

orangebus.

Douglas Connect
*Working communities*

ADVANTAGE
Integral

ADS

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.

For more information on the focus of Philip Morris International’s research, please visit www.pmiscience.com.

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL