Systems biology verification project: past, present, and future.

Elena Scotti, Stéphanie Boué, Julia Hoeng.
PMI R&D, Philip Morris Products S.A., Quai Jeanrenaud 5, CH-2000 Neuchâtel, Switzerland (Part of Philip Morris International group of companies)

Risk assessment in the context of 21st century toxicology relies on the development of relevant computational approaches for the extraction of mechanistic knowledge from big data.

Crowdsourcing is a powerful approach to solve scientific problem(s) and independently verify methods, results and conclusions. Relevant scientific contribution depends on the interest of the scientific questions formulated incentives put forward.

Based on these principles, the **sbv IMPROVER** organizes since 2011 crowd-sourced challenges covering a broad range of scientific questions. For instance, the **Diagnostic Signature Challenge** aimed to identify robust gene expression signatures and classification models in four disease areas. The **Species Translation Challenge** sought to refine our understanding of the limits of rodent models as predictors of human biology. The **Systems Toxicology Challenge** aimed to identify signatures predictive of smoking exposure or cessation status.

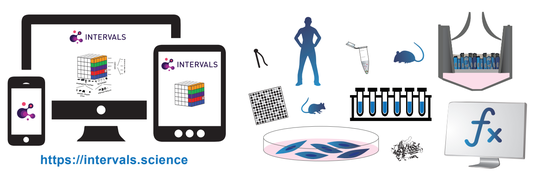
In addition to computational methods benchmarking, curation of scientific literature can also leverage crowdsourcing. Our **Network Verification Challenges (NVC)** have been designed to encourage broad participation in causal network models refinement with lower level of expertise required and participation encouraged through a live leaderboard.

A novel instance of NVC will focus on the verification of biological network models involved in the **xenobiotic transformation of toxicants in the liver**, which should be of great interest for toxicological and pharmacological assessment.

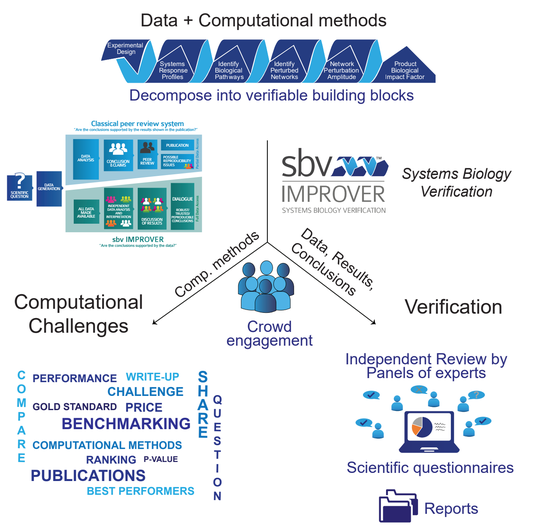
Finally, we are developing a step-wise **microbiomics challenge**, which should leverage the knowledge of diverse scientific communities to tackle in turn computational, biological, and medical topics related to the microbiome.

sbv IMPROVER

Based on the principles of **crowdsourcing** and **collaborative competition**, the sbv IMPROVER project is designed as a series of open **scientific challenges** where computational methods and conclusions related to scientific problems of interest in the systems biology and/or toxicology fields are rigorously scrutinized (Meyer 2011).



In strategically engaging the crowd, sbv IMPROVER **facilitates enhanced dialogue within the scientific community, transparency of research processes, and open innovation in scientific discovery**. The project advances the credibility of scientific techniques and complements the classical peer review process with a rigorous benchmarking of computational methods and assessment of conclusions. Computational challenges leveraging the wisdom of the crowd allow to benchmark methods for specific tasks, such as signature extraction or samples classification. Four challenges have already been successfully conducted and confirmed that the agglomeration of predictions often leads to better results than individual predictions and that methods perform best in specific contexts.



The challenge is posed by defining clear objectives and rules and providing data to the participants. Part of the data, called the "Gold Standard" (True values), to be predicted are kept hidden and are used in combination with pre-defined scoring metrics to assess the performance of anonymized participants' prediction submissions. Scoring results and team ranking are submitted to an external and independent Scoring Review Panel of experts for review and final approval. The results, conclusions and lessons learned from the challenge are shared with participants and with the scientific community through conference and symposium presentations and in peer-reviewed publications.

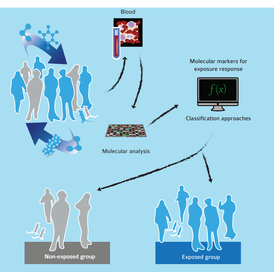
In particular, within the realm of the sbv IMPROVER project, **network verification challenges** are conducted to allow scientists to review a series of causal network models relevant for biological pathways and signaling known to be perturbed by cigarette smoke. In the past, two occurences focused on 50 network models relevant for lung biology. They culminated in 2 jamborees during which the new contributions were discussed and finalized and in peer reviewed publications (sbv IMPROVER, 2015 & 2016).

To learn more about the project and associated publications, please visit www.sbvimprover.com.

Challenges & Datathons

Systems Toxicology Challenge

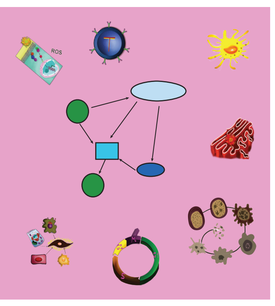
The SysTox Challenge aimed at verifying that robust and sparse human-specific and species-independent gene signatures of exposure response can be extracted in whole blood gene expression data from human and rodent to predict exposed and non-exposed group labels.



Special session at ISMB2016 and Symposium 2016, Orlando (USA)

Network Verification Challenge

The NVC Challenge aimed at verifying the biological network models to ensure their relevance to lung biology and COPD.



Symposium 2014, Montreux (CH)
Symposium 2015, Barcelona (ES)

Species Translation Challenge

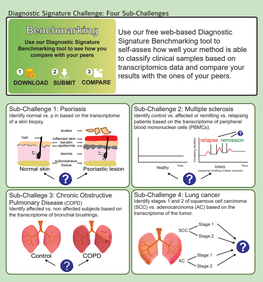
Changes in phosphorylation status and gene set activation induced by cellular response to 52 different perturbations in human cells can be predicted to a certain extent given responses generated in rat cells.

Symposium 2013, Athens (GR)

Diagnostic Signature Challenge

The goal of this Challenge was to assess and verify computational approaches that classify clinical samples based on transcriptomics data.

The high quality of predictions confirmed strongly the approach values.

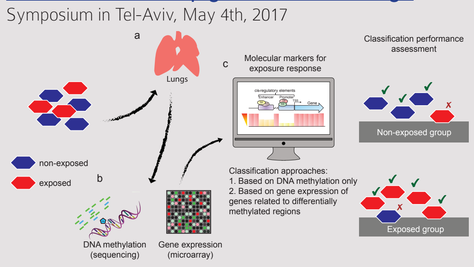


Symposium 2012, Boston (USA)

Datathons and Mini-computational Challenges

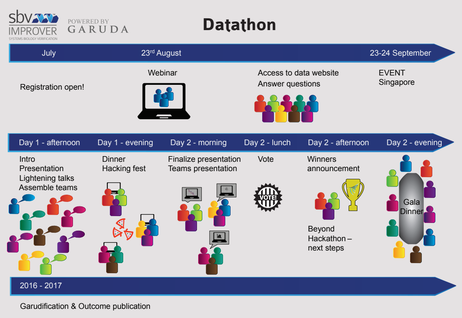
sbv IMPROVER Epigenomics Challenge

Symposium in Tel-Aviv, May 4th, 2017



The aim of this sbv IMPROVER Epigenomics Challenge was to apply computational approaches to assess the impact of different tobacco products in large methylome datasets obtained from a rodent inhalation study.

sbv IMPROVER Datathon - Singapore (Sept 2016)



During the Datathon in Singapore 2016, participants analyzed -omics data from an inhalation study in rodents.

References

Boue, S. et al. Causal biological network database: a comprehensive platform of causal biological network models focused on the pulmonary and vascular systems. Database2015, bav030 (2015)

Bradnam KR et al. Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. Gigascience. 2013; 2: 10.

Earl D et al. Assemblathon 1: a competitive assessment of de novo short read assembly methods. Genome Res. 2011

Martin et al. Quantification of biological network perturbations for mechanistic insight and diagnostics using two-layer causal models. BMC Bioinformatics. 2014

Meyer et al. Verification of systems biology research in the age of collaborative competition. Nat Biotechnol. 2011

Omiecinski CJ et al. Xenobiotic metabolism, disposition, and regulation by receptors: from biochemical phenomenon to predictors of major toxicities. Toxicological sciences : an official journal of the Society of Toxicology. 2011

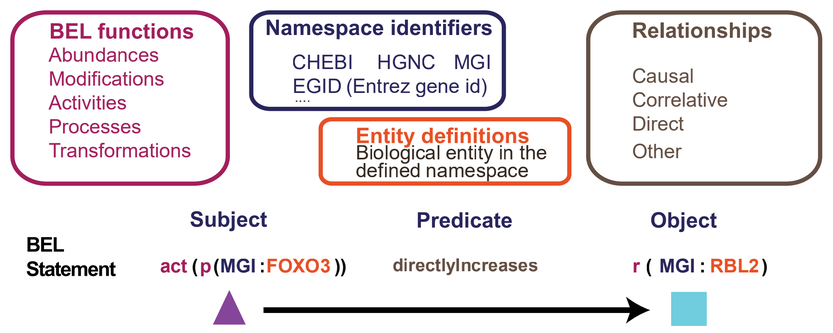
sbv IMPROVER team et al. Enhancement of COPD biological networks using a web-based collaboration interface. F1000Res. 2015

sbv IMPROVER team and NVC best performers. Community-reviewed biological network models for toxicology and drug discovery applications. Gene Regulation and Systems Biology. 2016

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

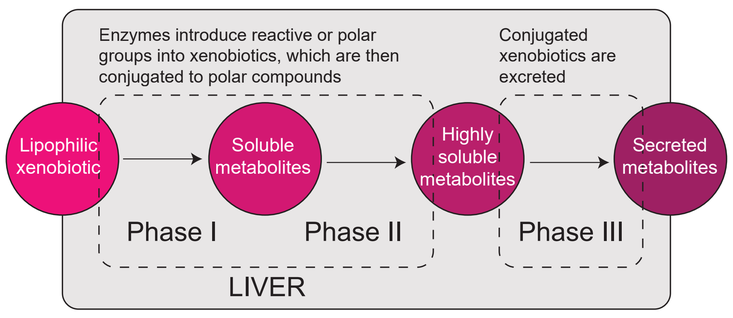
Network Verification Challenge

Philip Morris International (PMI) has invested in the development of a computational methodology to predict early toxicity using causal biological network models and high throughput molecular data (Martin 2014, Boue 2015).



Biological network models gather scattered knowledge from a myriad of publications into a structured presentation of biology, in the **biological expression langage (BEL)**, describing causal interactions between key molecular players in the context of a given biological process.

Manual construction of a network model is a labor intensive process. However, since scientific publications are written in a natural (human) language, computers cannot readily utilize the information as such. Hence, today's network building efforts take advantage of biological languages that are both human understandable and computable.



The 3rd edition of the network verification challenge (**NVC3**) focuses on three network models that represent biotransformation and chemical elimination involved in xenobiotic metabolism in the liver.

The **metabolism of xenobiotics** includes oxidative reactions by phase I enzymes that convert lipophilic chemical compounds into their hydrophilic forms, followed by phase II conjugation enzymes, and finally the phase III membrane transporters (Omiecinski 2011). The second and the last play a role in the elimination of xenobiotic metabolites.

The networks will be released for editing sequentially and the challenge will be open until end of December 2017.

NVC3 at a glance

Liver metabolism network models are available on bionet.sbvimprover.com for verification.

Users can extend networks with new edges and provide additional evidence for edges.

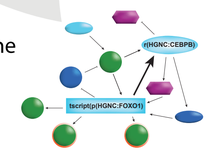
Evidence that has been posted in support of network edges can then be approved...

... or rejected.

The responses from the scientific community to your actions will define how you rank in the leaderboard and, ultimately, your success.

When the challenge closes, participants will be ranked and the winners will be rewarded.

After consolidation of the final changes, an updated version of the network models will be released on CBN website (causalbionet.com).



Collaborate. Contribute. Compete.

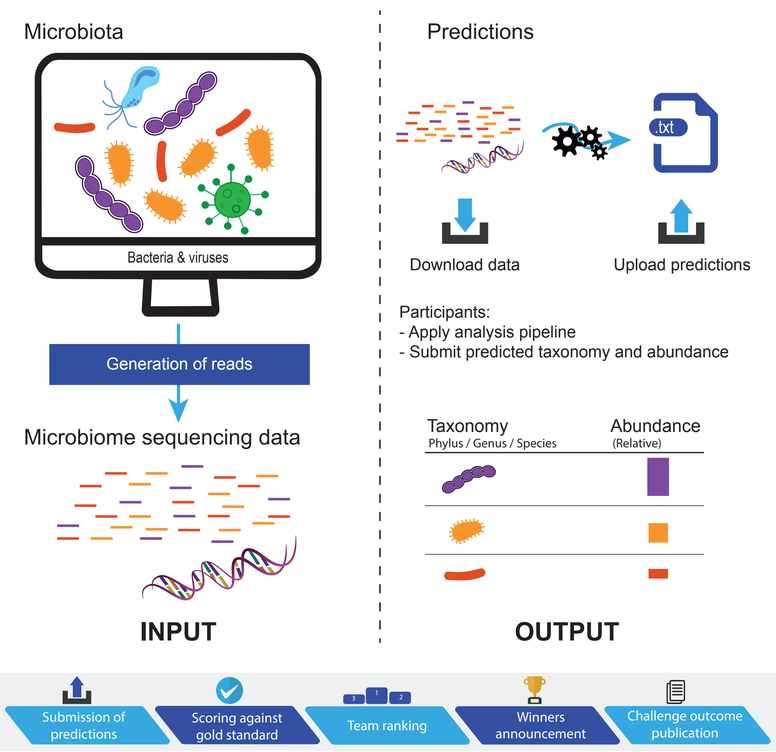
- Join your peers as they unite to verify and enhance existing biological network models that will then be released to the community for use in research applications such as drug discovery, personalized medicine, and toxicological assessment.
- Collaborate: have fun competing and collaborating with others.
- Test and expand your knowledge.
- 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.
- A gift card of 150 USD will reward participants reaching 3000 points in the leaderboard (see Challenge rules on bionet.sbvimprover.com).
- At the end of the challenge the best performing participants will be rewarded with a travel grant of up to 2,000 USD (see Challenge rules on bionet.sbvimprover.com).

Microbiomics Challenge

The **human microbiome** is known to have a beneficial role for homeostasis, assisting for example in the bioconversion of nutrients and detoxification, supporting immunity, protecting against pathogenic microbes, and maintaining host development, metabolism and physiology. A good and sensitive balanced interaction of microbes with the host is essential to health.

The function of the indigenous microbiota can be influenced by many factors, including genetics, diet, age, and toxins. The disruption of this balance, called **dysbiosis**, is associated with a plethora of diseases, including cancers, immune-related diseases, metabolic diseases, inflammatory bowel disease, pulmonary pathologies, oral diseases, skin problems, and neurological disorders.

Advances in genome sequencing technologies have enabled progress in the characterization of the microbial diversity, leading to a rapid expansion of the field known as **metagenomics**: the study of DNA of a microbial community.



An accurate analysis of microbiome sequencing data (e.g. accurate taxonomic assignment and relative abundance estimates) relies on computational methods, which have partly been scrutinized by initiatives such as the Assemblathon (Bradnam 2013) and CAMI (http://www.cami-challenge.org/).

To build on and go beyond what has been done by CAMI and the Assemblathon, namely assessing individual steps of the workflow, **the SBV-5 "Microbiota composition prediction" challenge, aims to assess objectively the performance of metagenomics computational analysis pipeline(s) as a whole, i.e. from quality control to taxonomy profiling, the end result being the recovery of relative abundance and taxonomy assignment of bacterial communities**.

Participants will be provided with shotgun DNA sequencing data for several microbiome samples and asked to predict, at the phylum, genus, and species level, the relative abundance of bacterial communities present in each sample. The participants have the freedom to use any private/public datasets to set up and test their approach.

Collaborate. Contribute. Compete.

- Receive an independent assessment of your method(s)
- Contribute to writing peer-reviewed scientific article(s) describing the outcome of the challenge
- Grow your professional network by engaging with researchers from around the world
- Win a travel bursary to a symposium taking place at the end of the next phase of the challenge (venue and time to be confirmed)

BC2, Basel
September 2017

Competing Financial Interest
The research described in this poster was sponsored by Philip Morris Products SA