Crowdsourced enhancement of causal network models – new Network Verification Challenge enabling liver xenobiotic metabolism models verification and refinement

Stéphanie Boué, Justyna Szostak, Marja Talikka, Florian Martin, 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)

Biological networks capturing mechanisms of toxicity are useful for understanding and quantifying the effects of exposure to compounds or mixtures thereof on organisms or *in vitro* models. As research progresses, recently published biological interactions need to be taken into account, to either expand or modify parts of the network.

Crowdsourcing can be used as a powerful tool to gather feedback from a wide community on a broad range of biological expertise. The **sbv IMPROVER** (www.sbvimprover.com) project, led and funded by Philip Morris International, leverages the power of the crowd to **verify methods and data relevant to systems biology**. Crowdsourcing is also increasingly used to perform tasks in both natural language processing and biocuration. Therefore, a few years ago, sbv IMPROVER launched the Network Verification Challenges (NVC) to **leverage crowdsourcing for the verification and enhancement of biological network models**. A **web platform** (www.bionet.sbvimprover.com) was custom-built to share networks, facilitate collaboration, and allow commenting and verification of scientific evidence supporting edges of biological networks. A **reputation-based system** and a leaderboard encouraged participation and allowed to highlight the most relevant contributions.

To extend beyond past use cases focused on inhalation toxicology, we have constructed new biological network models that represent the **mechanisms involved in the xenobiotic transformation of toxicants in the liver**. Publications with causal relationships related to xenobiotic metabolism in the liver were identified and curated using the **BELIEF** knowledge extraction workflow in the biological expression language (**BEL**) to form the network model backbone.

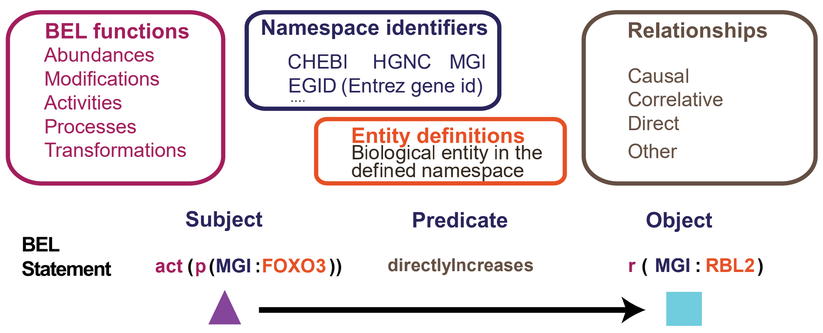
A new network verification phase has been launched with an initial focus on **liver xenobiotic metabolism Phase I**. This model contains nuclear receptors, such as the nuclear receptor subfamily 1 group I member 2 (Nr1i2), receptor subfamily 1, group I, member 3 (Nr1i3), nuclear factor, erythroid derived 2, like 2 (Nfe2l2), and aryl hydrocarbon receptor (Ahr), as well as signaling molecules that mediate the response to xenobiotic compounds. It also includes the main CYP450 enzymes that detoxify or activate xenobiotic compounds by catalyzing chemical reactions, such as oxidation, reduction or hydrolysis with the aim to make them more water soluble. Liver xenobiotic metabolism Phases II and III will also become available for verification and refinement.

This poster describes learnings from the past NVC, **algorithms that enable to quantify perturbation of such network models**, and our latest challenge focused on the enhancement of liver xenobiotic phase I metabolism models, which will facilitate mechanistic understanding of liver toxicity in response to xenobiotics.

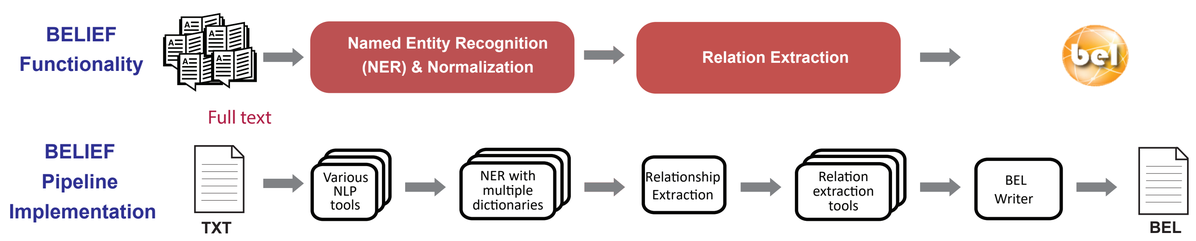
Building Causal Network Models - BEL & BELIEF

Biological network models gather scattered knowledge from a myriad of publications into a structured presentation of biology, 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 network models are encoded in the Biological Expression Language (BEL), which represents, as illustrated in the figure, scientific knowledge in human readable and computable format as semantic triples (subject, a predicate, and an object).



Since neither purely manual nor automated literature curation yields satisfactory results, in terms of curation time and quality, we have developed, in collaboration with the Franhofer Institute, the BEL Information Extraction workFlow (BELIEF, accessible at http://belief.scai.fraunhofer.de/BeliefDashboard/), a semi-automated workflow for BEL network creation [1].
It embeds an information extraction workflow with state-of-the-art named entity recognition and relation extraction methods.

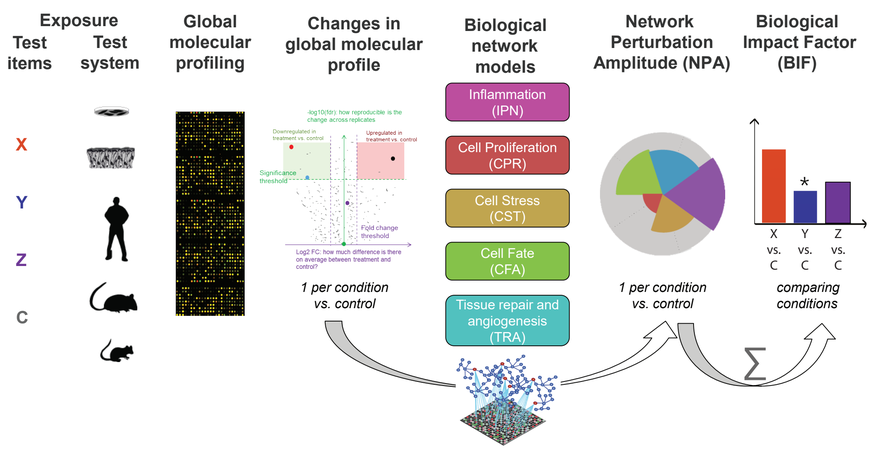


BELIEF is now routinely used to build biological network models in 3 steps:

- **Boundary criteria and literature selection**. Set of specific boundary criteria for the biological process of interest and selection of scientific articles
- **Literature curation**. Submission of the scientific articles to the text mining pipeline in the BELIEF platform. It automatically identifies causal relationships & proposes BEL statements. BELIEF platform also supports the user to check the validity of BEL statements and annotations.
- **Network model creation**. The extracted causal relationships are then compiled into a causal network model with the OpenBEL framework 3.0: the nodes in the network model are connected by causally related edges.

Quantifying Perturbation of a Biological System with Network Models

Based on causal network models, we have developed computational methods that allow to understand the mechanisms behind and predict the effect of exposure based on transcriptomics datasets [2].
This methodology enables to translate the gene expression fold-changes into differential values for each network node, and to summarize this at the network level to provide a quantitative assessment of the degree of perturbation of the network model, the Network Perturbation Amplitude (NPA) [2,3].
Combining multiple relevant network models, the overall biological impact of a perturbing agent, the Biological Impact Factor (BIF) [3], can be calculated by aggregating individual NPA scores.



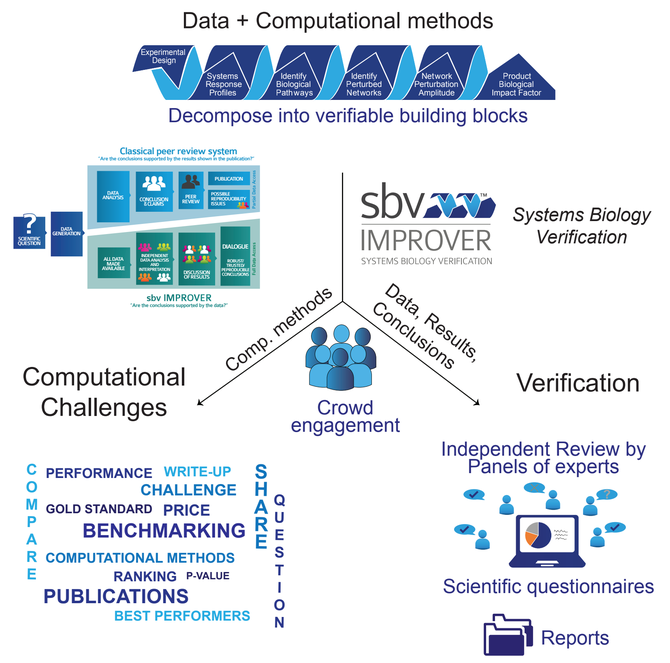
sbv IMPROVER

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 [4].

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. The outcome of each challenge is published in peer reviewed journals, so it may benefit the scientific community as a whole.

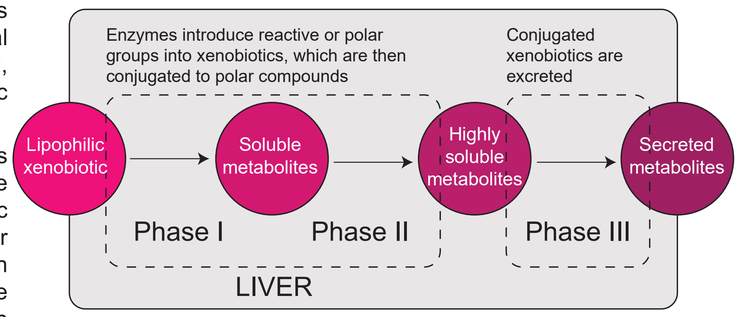
In particular, within the realm of the sbv IMPROVER project, **network verification challenges** are conducted to allow scientists to review a series of causal network models relevant for biological pathways and signaling known to be perturbed by cigarette smoke. In the past, two occurences focused on 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 [5-8].

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



Liver Xenobiotic Metabolism Network Models

The NVC3 focuses on three network models that represent biotransformation and chemical elimination involved in Phase I, Phase II, and Phase III, respectively, of the xenobiotic metabolism in the liver:



- **Phase I** Xenobiotic Metabolism encompasses the biochemical reactions that introduce reactive and polar groups into xenobiotic compounds by oxidation, reduction, or hydrolytic reactions. One of the most common modifications is hydroxylation catalyzed by the cytochrome P-450-dependent mixed-function oxidase system. The phase I network model describes signaling that leads to the **activation of P450 family enzymes**. Important nuclear receptors, such as nuclear receptor subfamily 1 group I member 2 (Nr1i2), nuclear receptor subfamily 1 group I member 3 (Nr1i3), the aryl hydrocarbon receptor (Ahr), and the peroxisomal proliferator-activated receptors alpha (Ppara) mediate this signal.
- In subsequent **phase II** reactions, these activated xenobiotic metabolites are **conjugated** with charged species. The phase II Xenobiotic Metabolism Network Model describes the activation of UDP-glucuronosyltransferases, sulfotransferases, N-acetyltransferases, glutathione S-transferases and highlights the importance of nuclear factor, erythroid 2 like 2 (Nfe2l2), and nuclear receptor subfamily 3 group C member 1 (Nr3c1) in the transformation process.
- In the final step, the **phase III** Xenobiotic Metabolism Network Model describes the xenobiotic **excretion** from cells. It is mediated by the activation of molecular transporters, such as ATP-binding cassette subfamily B (Abcb), ATP-binding cassette subfamily G (Abcg), ATP-binding cassette subfamily C (Abcc), and ATP-binding cassette subfamily A (Abca).

Network Verification Challenge

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.

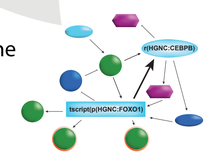
Evidence that has been posted in support of network edges can then be approved...

... 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).



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

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).

1. Szostak et al. Construction of biological networks from unstructured information based on a semi-automated curation workflow". Database. 2015
2. Hoeng et al. A network-based approach to quantifying the impact of biologically active substances. Drug Discovery Today. 2012
3. Martin et al. Quantification of biological network perturbations for mechanistic insight and diagnostics using two-layer causal models. BMC Bioinformatics. 2014
4. Meyer et al. Verification of systems biology research in the age of collaborative competition. Nat Biotechnol. 2011
5. sbv IMPROVER team et al. On Crowd-verification of Biological Networks Bioinformatics and Biology Insights. 2013
6. sbv IMPROVER team et al. Reputation-based collaborative network biology PSB Proceedings. 2015
7. sbv IMPROVER team et al. Enhancement of COPD biological networks using a web-based collaboration interface. F1000Res. 2015
8. 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

ISMB, Prague
July 2017

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