sbv IMPROVER
SYSTEMS BIOLOGY VERIFICATION

ECCB 2014

Overview:

The sbv IMPROVER project, the website 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 current challenges, website and biological network models were developed and are maintained as part of a collaboration among Selventa, OrangeBus and ADS. 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.pmi.com.

A reputation-based web application (sbv IMPROVER Network Verification Challenge) that facilitates collaboration on biological network models

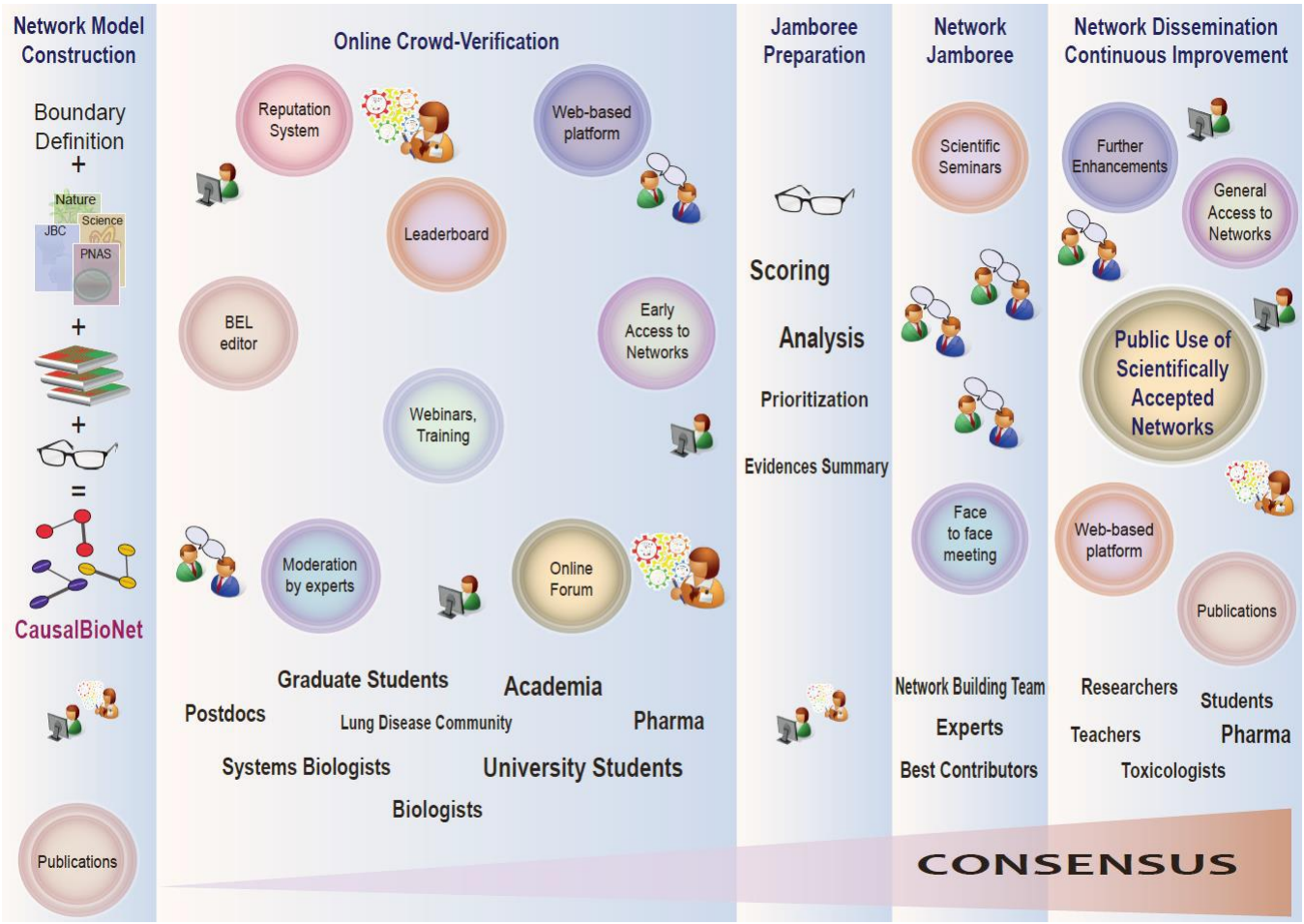
The sbv IMPROVER network verification challenge (NVC) aims to verify and enhance existing biological network models. Such models are captured using a structured syntax (Biological Expression Language, BEL) and serve as a powerful way of representing biological information generated from systems biology data.

In the first phase (NVC1) Crowdsourcing principles enabled participants from 60 institutions across 12 countries to communally annotate these relationships based on literature evidences. Collaborative competition principles were incorporated to further engage domain experts from various fields of biology and medicine to gather robust peer-reviewed information from which relationships were identified and evaluated. The 34 resulting network models present the current status of biological knowledge within the defined boundaries related to human lung disease.

By creating smarter solutions to complement peer review with collaborative crowd sourcing, the NVC uses innovative concepts of reputation-based engines and state of art web application for online network verification.

This presentation will discuss the crowd-verification approach that allowed the visualization and expansion of biological networks in NVC1, and is currently being used to further review biological network models that are suitable for drug discovery, toxicological and mechanistic research in respiratory diseases (NVC2).

Steps in the sbv IMPROVER Network Verification Challenge



The sbv IMPROVER project, the website 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 current challenges, website and biological network models were developed and are maintained as part of a collaboration among Selventa, OrangeBus and ADS. 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.pmi.com.

Quantification of biological network perturbations: Impact assessment and diagnostic using causal biological networks

Florian Martin, Alain Sewer, Marja Talikka, Yang Xiang, Julia Hoeng, Manuel C. Peitsch. Biological Systems Research, R&D, Philip Morris International, Neuchatel, NE, Switzerland.

High-throughput profiling of gene expression has opened new avenues for the understanding of biological processes at the molecular level. However, the amount of information collected can be overwhelming, making interpretation of the data difficult and subsequent detailed biological understanding elusive.
Reducing the complexity of such data by evaluating them in a relevant biological context is required to gain meaningful insight. We propose that "cause-and-effect" network approaches to pharmacology and toxicology are valuable to quantify network perturbations caused by bio-active substances, and to identify mechanisms and biomarkers modulated in response to exposure[1]. The underlying concept is that transcriptional changes are the consequences of the biological processes described in the network.
We have recently built an ensemble of network models that consist of "cause and effect" relationships (typically activation or inhibition) between molecular entities and activities (e.g. kinase activation or increased protein abundance)[2-5]. The description of the biological context has been manually built into the network models using prior knowledge extracted from both relevant literature and published datasets after a large-scale knowledge mining effort. Some network nodes are also related to mRNA abundance entities that they positively or negatively regulate.
Thus, our biological network models have a two-layer structure, where the functional level is explicitly distinguished from the transcriptional level. Using transcriptional downstream effects to infer the activity of upstream entities has its advantages, because the activity of a node is inferred based on the differential expression of many genes known to be regulated by a given entity, even the ones encoding proteins with unknown functions. This is unlike the networks derived from other pathway databases, which rely upon the "forward assumption" stating that changes in gene expression induce changes in the activity and abundance of the gene product.
We present a novel framework for the quantification of the amplitude of network perturbations to enable comparisons between different exposures and systems. Also, our approach enables quantification of each biological entity (nodes) in the network, among which key contributors, referred to as leading nodes, can be identified to unravel biological mechanisms. It efficiently integrates transcriptomics data and network models to enable a mathematically coherent framework from quantitative impact assessment to data interpretation and mechanistic hypothesis generation. The gene expression fold-changes are translated into differential values for each node of the network (denoted by f) by fitting the functional layer relationships with respect to the boundary constraint given by the observed fold-changes. The node differential values are in turn summarized into a quantitative measure of network perturbation amplitude (NPA). The NPA is computed as a Sobolev (semi-)norm on the signed directed graph underlying the network, which can be expressed as a quadratic form $f_T Q f$. In addition to the confidence intervals of the NPA scores, which account for experimental error, companion statistics were derived to inform on the specificity of the NPA score with respect to the biology described in the network. The network is considered to be specifically perturbed if all P-values are low (typically < 0.05). A unique property of our methodology allows the mapping of transcriptomics data observed in individual samples onto the network nodes which enables the generation of diagnostic network signatures signature that can be applied on the context of sbv IMPROVER Diagnostic Signature Challenge[6, 7]. These signatures are coherent with the overall quantification of the perturbation amplitude.
The set of networks built-to-date provide a coherent framework for investigating the impact of exposures at the molecular, pathway and process levels. Multiple network scores can be combined to calculate an overall biological impact factor (BIF) that can then be compared to other treatments, time points or doses in the experiment.
Example scored data sets will be given including in vitro systems with simple exposures, and an in vivo system with a complex exposure. These examples will illustrate that various fields of human disease research, from drug development to consumer product testing and environmental impact analysis, could benefit from using this quantitative network scoring methodology. The presented framework efficiently integrates transcriptomics data and "cause and effect" network models to enable a mathematically coherent framework from quantitative impact assessment, data interpretation and mechanistic hypothesis generation to patient stratification for diagnosis purposes.

[1] Hoeng, J., et al. A network-based approach to quantifying the impact of biologically active substances. Drug Discov. Today 17(9-10), 413–418 (2012)
[2] Schlage, W.K, et al. A computable cellular stress network model for non-diseased pulmonary and cardiovascular tissue. BMC Syst Biol 5, 168 (2011)
[3] Westra, J.W, et al. Construction of a computable cell proliferation network focused on non-diseased lung cells. BMC Syst Biol 5, 105 (2011)
[4] Gebel, S., et al. Construction of a computable network model for DNA damage, autophagy, cell death, and senescence. Bioinformatics and Biology Insights 7, 97–117 (2013)
[5] Westra, J., et al. A modular cell-type focused inflammatory process network model for non-diseased pulmonary tissue. Bioinform Biol Insights 7, 167–192 (2013)
[6] Meyer, P., et al. Verification of systems biology research in the age of collaborative competition. Nature biotechnology, 29(9), 811 (2011)
[7] Tarca, A.L., et al. Strengths and limitations of microarray-based phenotype prediction: lessons learned from the IMPROVER Diagnostic Signature Challenge. Bioinformatics 29(22), 2892–2899 (2013)