ORIGINAL PAPERS

bit.ly/NPA-Protocol

causal biological network models • gene expression • network perturbation amplitude

Network-based approach for quantifying biological impact

- Causal biological network models organize disparate information into a structured presentation of knowledge.
- They are built within defined boundaries, describing biological molecules, modifications, and disease mechanisms with a high degree of granularity.
- Models enable interpretation of gene expression changes caused by perturbation of biological processes.
- Calculation of network perturbation amplitudes and biological impact factors facilitates a top-down approach to quantifying systems-wide responses to exposure to an external substance in a hierarchical and interpretable manner.

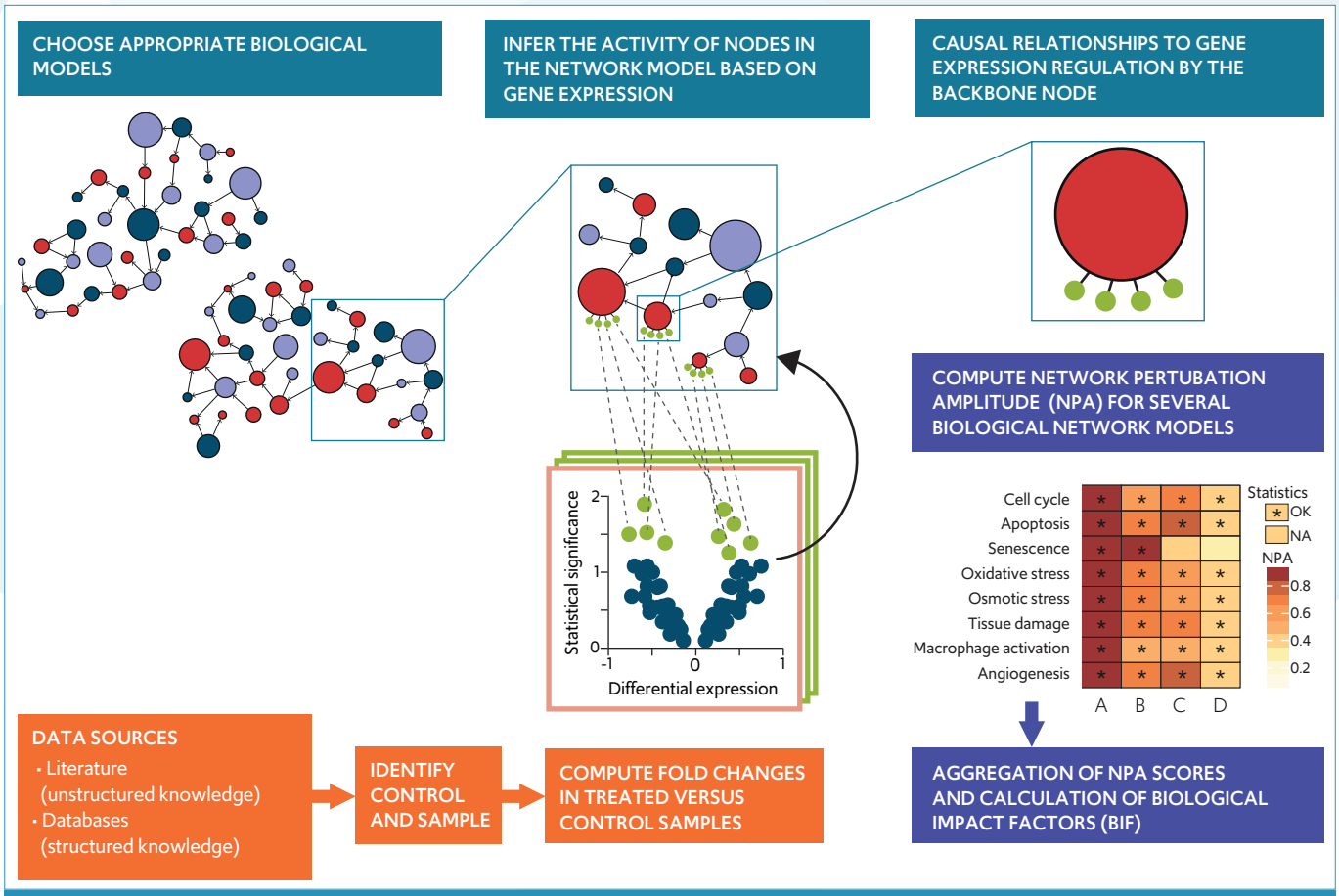

Figure 1. Workflow: Generating network perturbation amplitudes from gene expression data

Causal biological network models capture proteins, DNA variants, coding and non-coding RNA, chemicals, lipids, methylation states, or other modifications as well as disease mechanisms.

Causal biological network models are being increasingly used in contemporary approaches to medicine and pharmacology and play a central role in PMI's systems toxicology assessment strategy for reduced-risk products (RRPs). The approach allows the use of data generated by in vitro, in vivo, and clinical studies to quantify the response of biological networks to exposure to an external substance, such as the aerosol generated by an RRP. This is achieved by calculating network perturbation amplitudes (NPAs) (Fig. 1), which, in turn, can be aggregated into an overall biological impact factor (BIF). This handout provides a brief introduction to our network-based approach for quantifying biological impact and the NPA/BIF methodology.[1,2]

CAUSAL BIOLOGICAL NETWORK MODELS

Causal biological network models are built by assembling cause and effect statements, which are encoded in Biological Expression Language (BEL, see Fig. 2 below). They are built within defined biological boundaries for inclusion of signaling pathways within specific contexts (species, tissues, cell types, and experiment types). These models provide a high degree of granularity in capturing biological molecules including proteins, DNA variants, coding and non-coding RNA, chemicals, lipids, methylation states, or other modifications (e.g., phosphorylation) as well as disease mechanisms. They are comprised of nodes and edges, with most causal edges supported by at least one literature reference (Fig. 3).

Browsable versions of the models have been published in the Causal Biological Network Database (http://causalbionet.com), reflecting a wide range of biological processes, including cell fate, cell stress, cell proliferation, inflammation, tissue repair, and angiogenesis, in the pulmonary and cardiovascular contexts.3 The networks are evolving, and all versions are made available in the Causal Biological Network Database to reflect their iterative refinement and keep a record of their historical development. While 98 first-generation networks describing lung and vascular biology have been published in six peer-reviewed articles, 50 network models have been consolidated from the 92 lung-relevant networks and enriched with COPD-relevant mechanisms, representing the second-generation network models. In addition to peer review, second-generation models on the Causal Biological Network Da-

BIOLOGICAL EXPRESSION LANGUAGE

The Biological Expression Language (BEL) was designed to capture biological cause and effect relationships from disparate sources. BEL converts existing knowledge derived from literature and content-rich biological datasets into a common language that is both human readable and computable. BEL statements specify protein abundance, activity or modification information and can be accompanied by rich contextual information including species, tissue, and experiment type. For more information see https://bel.bio/.

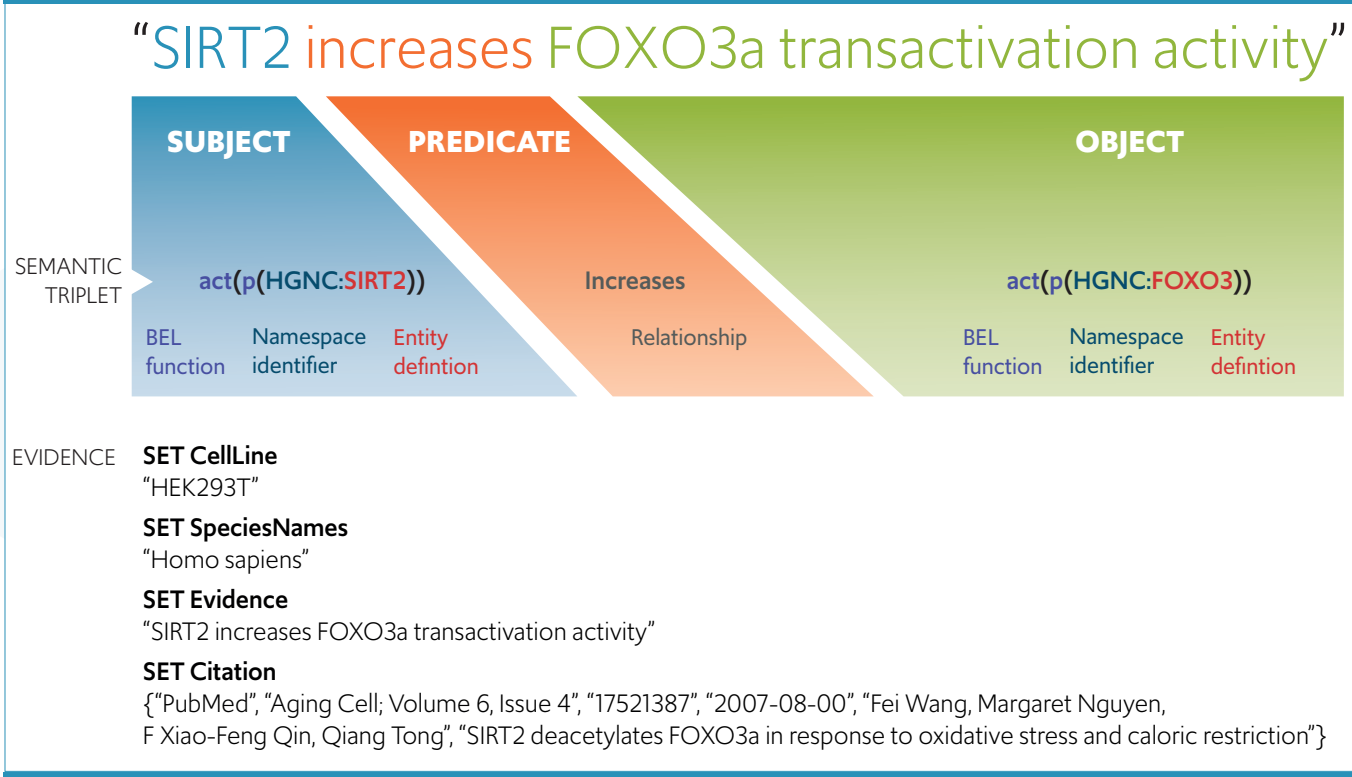

Figure 2. Example BEL statement with annotations

act: protein activity; **p:** protein abundance; **HGNC:** HUGO Gene Nomenclature Committee

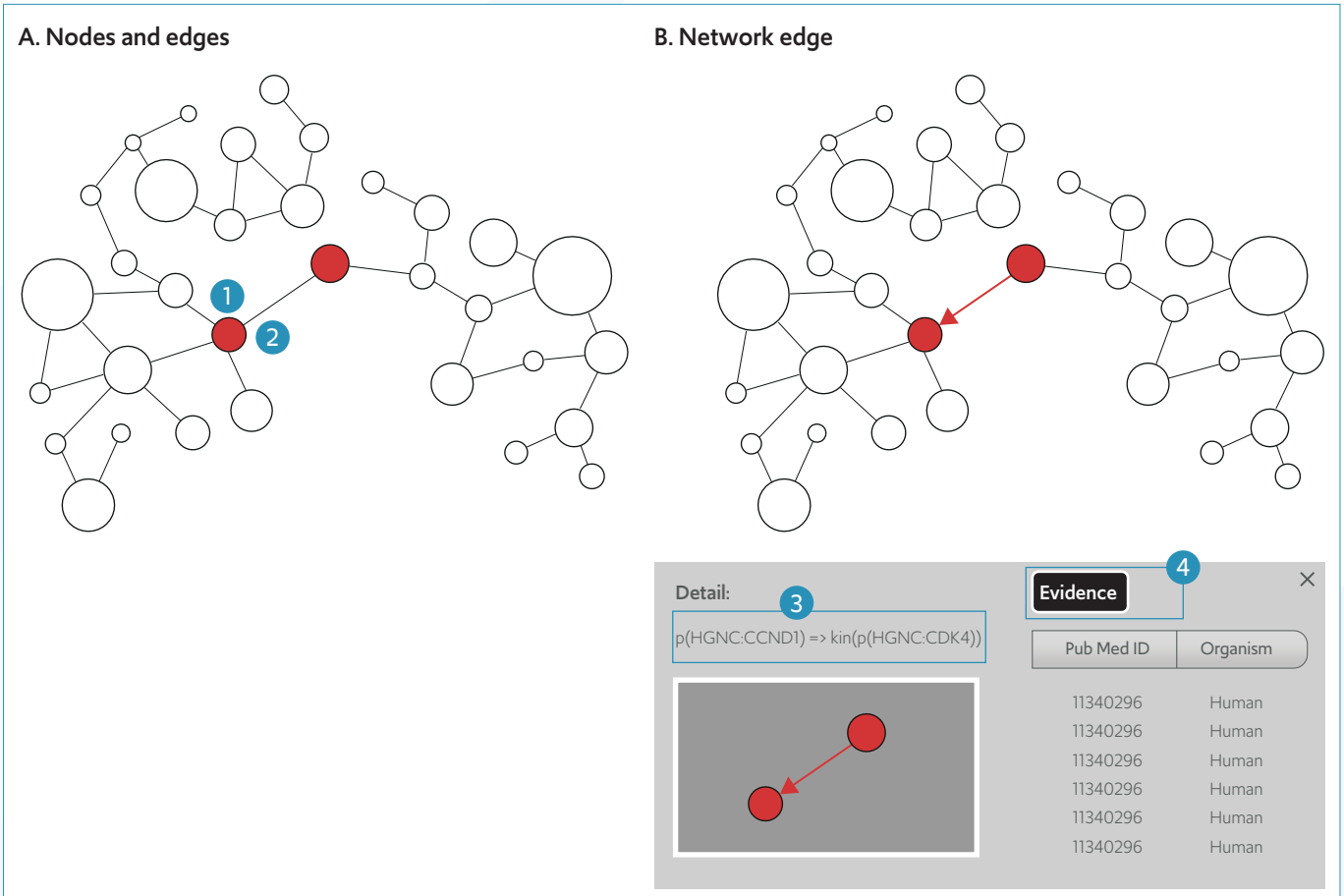

Figure 3. Causal biological network models

1. Node, e.g. p(HGNC:EGFR), bp(GO:"Oxidative Process"), a(CHEBI:Water)
2. Network edge
3. Network edge, consisting of a node connected to another node by a signed, directed relationship
4. Evidence statement: edges are generally supported by one or more evidence (BEL) statements that provide the evidence for the edge, including the citation (usually PubMed ID) and experimental context

tabase have undergone refinement by experts from the scientific community via the sbv IMPROVER Network Verification Challenge (see www.sbvimprover.com/challenge-3).

DIFFERENTIAL GENE EXPRESSION

Gene expression, the production of RNA, is central to a cell's functioning, acting as a crucial step in the production of active proteins that perform the tasks needed by the cell. Perturbation of biological processes in cells usually results in changes to the gene expression levels of specific genes. As such, causal biological network models are extended in a downstream layer to include cause and effects statements from the molecular entities described in the model (network backbone) to gene expression changes. The extended cause and effect models therefore have a two-layer structure: 1) the biological network itself and 2) the gene expression changes related to it.

Systems toxicology exposure experiments use at least two groups of samples (tissue from clinical studies, tissue culture, or cells). One group is exposed to an external substance, such as the aerosol generated by an RRP, while the other serves as a control and is exposed only to a negative control, such as fresh air. For each group, both stimulated and unstimulated, RNA is extracted at defined times post-exposure, and the level of gene expression is measured. Measuring the levels of gene expression in a group of stimulated cells and comparing those to the levels in a group of unstimulated cells helps determine the differential gene expression linked to the stimulus.

High-throughput technologies allow us to measure the gene expression levels of more than 20,000 genes in a given sample, enabling estimation of differential gene expression and the statistical sgnificance for every gene.

NETWORK PERTURBATION AMPLITUDES

A meaningful systems toxicology approach for studying the response of a cell system or organism to exposure to bioactive substances requires a quantitative measure of those responses at a network level. This goes beyond an understanding of the differential expression of genes. Therefore, we developed a method that quantifies network response in an interpretable manner. It enables integration of differential gene expression and causal network models to quantify network perturbation and unravel activation patterns. This is achieved by computing the values of the nodes in the model (layer 1) that best fit the

Cigarette smoke (CS) is the leading modifiable risk factor for many human diseases. Complete smoking cessation is the best approach to reduce the risks of tobacco products. However, while the prevalence of cigarette smoking has been steadily declining over the years, millions of individuals across the globe continue to smoke. Smoking cessation has proven difficult for many smokers, who might benefit from using alternative products that reduce the harm caused by CS. While health professionals should vigorously promote smoking cessation in their daily practices, it is also important to investigate and understand the potential effects of emerging tobacco products on health. For smokers who would otherwise continue smoking cigarettes, PMI's goal is to offer smoke-free alternatives, reduced-risk products (RRPs),* that have the potential to reduce the risk of developing smoking-related diseases as compared to continued smoking.

cause and effect relationships while being constrained by the genes underlying the extended network (layer 2). The computed values are summarized into a network perturbation amplitude (NPA) score, quantifying the extent to which a network has been affected by the exposure.

BIOLOGICAL IMPACT FACTOR

The final step in estimating the biological impact of a perturbing substance is to aggregate NPA scores into a biological impact factor (BIF), which requires normalization of scores between networks and weighting of the contribution of each network involved (Fig. 1). The BIF expresses the overall impact of a substance on a biological system as a single holistic value, while allowing for a top-down interpretation of biological impact from groups of networks to single networks to individual nodes. Ultimately, the approach offers the ability to quantitatively describe the long-term impact of network perturbations.

APPLICATIONS

The NPA/BIF systems-toxicology approach described here has been extensively used in assessment of Tobacco Heating System (THS) 2.2, one of the portfolio of reduced-risk products being developed and assessed by PMI. In combination with traditional toxicological endpoints in various experimental systems, the approach has quantitatively demonstrated the reduced biological impact of exposure to THS 2.2 aerosol relative to cigarette smoke.[1]

Causal biological network models have also been used in a meta-analysis of in vitro assessment studies that focused on human organotypic cultures of the aerodigestive tract (buccal, bronchial, and nasal epithelial cultures) exposed to THS 2.2 aerosol and cigarette smoke. Thousands of simultaneously measured data points were put into the context of known biological processes, and their perturbation effects were deciphered by using computational evaluation strategies and causal network modelling.[1]

While there are few true systems toxicology meta-analyses available, causal biological network models have also recently been used in a meta-analysis of independent studies of engineered nanomaterials to gain mechanistic insights into how they impact pulmonary biological processes. An NPA heatmap was generated, which provided a quantitative view of the network impact across the experiments, while leading edge node analysis showed the directionality of the inferred impact on each node.[1]

The NPA/BIF systems toxicology approach has further been demonstrated to reliably group drugs into those that are safe and those that are likely to cause adverse effects, with potential for prediction of personalized adverse drug reactions upon adding clinical data that considers the heterogeneity of the human population.[4]

USING THE APPROACH

Network models are freely available via the Causal Biological Network Database: **bit.ly/CBN-Database**. These networks reflect a range of biological processes including cell fate, cell stress, cell proliferation, inflammation, tissue repair, and angiogenesis in the pulmonary and cardiovascular context.

A packaged computational NPA/BIF tool is available for free via the GitHub platform: **bit.ly/NPA-BIF**. The tool allows users to understand the mechanisms behind and predict the biological effects of exposure to an external substance based on transcriptomics datasets. Researchers are able to translate gene expression fold-changes into differential values for biological network nodes and to summarize this at the network level to provide a quantitative assessment of the degree of perturbation of a network model. Combining multiple relevant network models, the overall biological impact of a perturbing agent can then be calculated.

REFERENCES

1. Talikka M, et al. Systems toxicology meta-analysis: from aerosol exposure to nanotoxicology. *Curr Opin Toxicol* 2019; **16:** 39–48.
2. Hoeng J, et al. A network-based approach to quantifying the impact of biologically active substances. *Drug Discov Today* 2012; **17:** 413–8. https://doi.org/10.1016/j.drudis.2011.11.008.
3. Boué S, et al. Causal biological network database: a comprehensive platform of causal biological network models focused on the pulmonary and vascular systems. *Database* 2015. https://doi.org/10.1093/database/bav030.
4. Talikka M, et al. Applying systems toxicology methods to drug safety. 2019. In press.
5. Martin F, et al. NPA: an R package for computing network perturbation amplitudes using gene expression data and two-layer networks. *BMC Bioinformatics* 2019; **20:** 1–9.
6. Martin F, et al. Quantification of biological network perturbations for mechanistic insight and diagnostics using two-layer causal models. *BMC Bioinformatics* 2014; **15:** 238.

NPA PROTOCOL

bit.ly/NPA-Protocol