Quantification of biological network perturbations: Impact assessment using causal biological network models

Florian Martin, Alain Sewer, Marja Talikka, Yang Xiang, Shen Yi*, Julia Hoeng, Manuel C. Peitsch.
Biological Systems Research, PMI Research & Development, Neuchatel, Switzerland.
*PMI Research Laboratories Pte Ltd #02-07, 50 Science Park Road The Kendall Singapore Science Park II Singapore

Objective

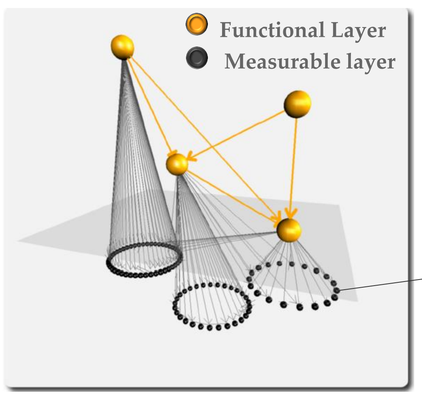
High-throughput profiling of gene expression has opened new avenues for the understanding of biological processes at the molecular level. However, the amount of information collected can be overwhelming, making interpretation of the data difficult and subsequent detailed biological understanding elusive. 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)[2-6]. 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 is advantageous, 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, which enable comparisons between different exposures and systems [7-9]. 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. As a conclusion, 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.

Methods

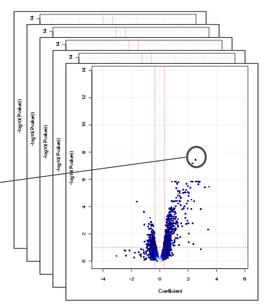
«CAUSE-AND-EFFECT» NETWORK MODEL



Biology is mechanistically captured by the molecular entities (**backbone**) of the model through "cause-and-effect" edges.
The nodes represent molecular entities such as catalytic activity of the protein IL1R1 (expressed in BEL language, catof(IL1R1)).

Gene expressions are experimental evidences for the activation of some nodes of the backbone.

GENE EXPRESSION DIFFERENTIAL DATA



Differential gene expression data e.g., estimated contrast of interest, or difference between individual patient profile and the average population.

Use differential effects (β) for each gene underlying the network model (>1000), β as boundary condition in the following optimization problem:

$$\operatorname{argmin}_{f \in l^2(V)} \sum_{x \to y} (f(x) - sign(x \to y) f(y))^2 w(x,y)$$

$$\text{such that } f|_{V_0} = \beta$$

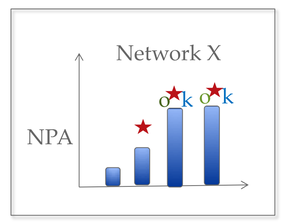
Where V is the set of nodes, and V_0 the nodes of the measurable layer

The solution is given as a linear transform of the boundary conditions, using the signed Laplacian of the graph.

NPA is the total graph "energy" of the solution:

$$NPA(\mathcal{G}, \beta) = \frac{1}{|\{x \to y\} \text{ s.t. } x,y \notin V_0|} \sum_{\substack{x \to y \\ \text{s.t } x,y \notin V_0}} (f(x) + sign(x \to y) f(y))^2 \cdot w(x,y)$$

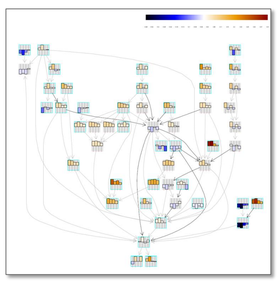
Network Perturbation Amplitude (NPA)



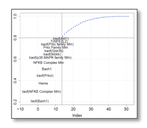
★Significance of the score with respect to biological variability;

O + **K**: Additional companion statistics (permutation tests) indicating if the biology described by the model is specifically perturbed.

Network Perturbation Shape



Leading Nodes



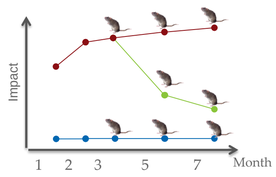
Each mechanism is scored in the context of the network and the most contributing nodes are identified as *leading nodes*. The leading nodes are central nodes in the network, accounting for up to 80% of the NPA score.

Use case : Smoking cessation study using the C57/BL6 emphysema model

Background Chronic Obstructive Pulmonary Disease (COPD) is one of the leading causes of chronic morbidity and mortality in the world. C57/Bl6 mice exposed to cigarette smoke (CS) provide valuable insight into emphysema initiation and progression, although this mimics only some aspects of early human COPD, characterized by reduced lung function, abnormal inflammatory response in the airways, small airway remodeling, and the destruction of lung alveolar tissue. Previously, we have applied our mechanism-based systems toxicology strategy to gain mechanistic insights and quantify the activation of various biological processes in the lungs from smoke-exposed mice [10].

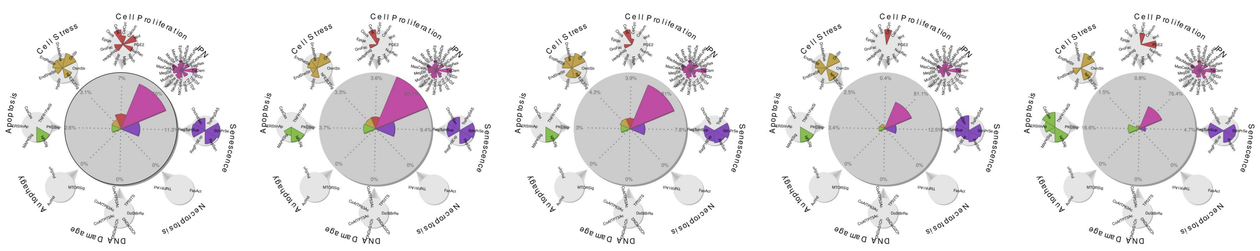
Whole lung samples were collected from 8 mice for each Month x Arm

Study Arms:
Sham
CS
CS + Sham (Cessation)



Objective We have used transcriptomic data from mice exposed to mainstream CS for 3, 5 and 7 months as well as data after 3 months of smoke exposure followed by a cessation period of 2 month and 4 months, respectively. Similar to the findings from a study by Beckett et al. [11], the mice developed emphysema in the lungs after just 3 months of CS exposure (M. Peck et al., unpublished), characterized by lung morphometry and histopathological analysis.

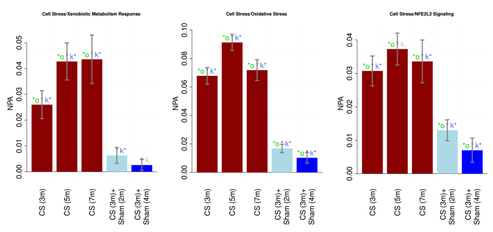
Figure 1 We computed NPA's for the Pulmonary Inflammatory Processes Network (IPN), Cell Proliferation Network , Cell Stress Network and the networks that constitute the DNA damage, Autophagy, Cell death (apoptosis and necroptosis) and Senescence Network (DACS). The starplots show a summary of the NPAs of lungs from CS-exposed versus sham-exposed mice at each time point. Only *O*K* networks are displayed and aggregated by biological processes. The outer starplot is relative to the corresponding inner segment.



CS (3m) *CS (5m)* *CS (7m)* *CS (3m)+Sham (2m)* *CS (3m)+Sham (4m)*

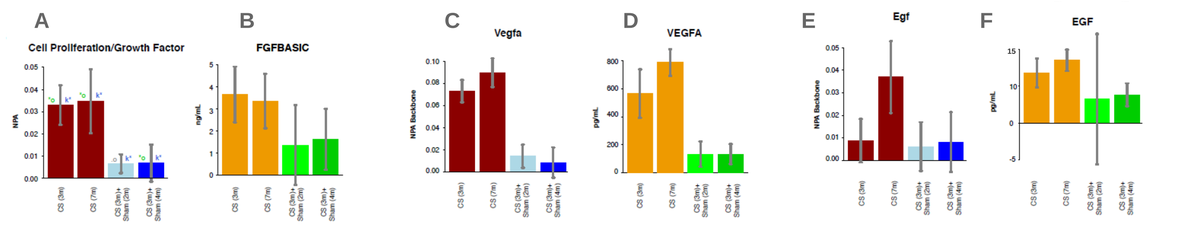
The impact of CS on biological processes in the lung is increased over month of CS exposure and is significantly decreased after cessation. Similar to endpoints related to lung function and pathology, the extent of impact depends on both the duration of smoke exposure and length of cessation.

Figure 2 The comparison of the perturbation of two subnetworks within the cellular stress network, xenobiotic metabolism response oxidative stress and NFE2L2 signaling, across the treatment regimens.



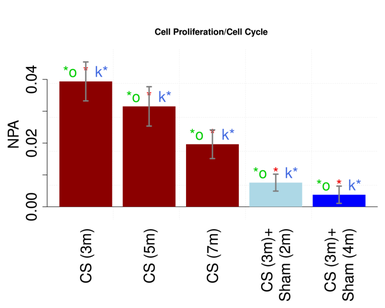
Xenobiotic metabolism , Oxidative stress and NFE2L2 signaling networks show CS induced perturbation that is released upon smoking cessation.

Figure 3. Computation of NPA for processes related to cell proliferation and comparison to growth factors measurement in BALF (available at 3 and 7 month).

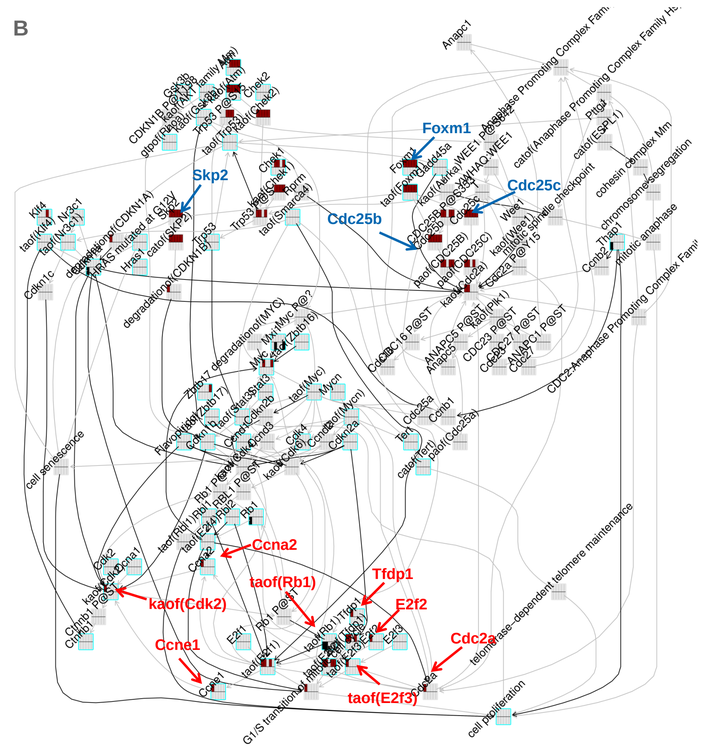


The growth factor subnetwork is clearly perturbed following CS exposure and released from perturbation upon smoking cessation (A). Broncho Alveolar Lavage Fluid (BALF) concentrations for bFGF is similarly altered in CS exposed mice (B). The Vegfa node in the network follows the same trend as the analyte level in the BALF, being most affected at 7 months of exposure (C). On the other hand, Egf signalling shows no consistent behaviour at the network node level (E) and similarly, the protein measured in BALF fails to serve as a marker for disease progression/reversal (F).

Figure 4. (A) Activation of cell cycle submodel within the cell proliferation network [4]. This subnetwork has been carefully verified to reflect the correct biology with a tailor-made dataset that reflects the activation of the cell cycle after arrest to G1 phase [12]. (B) To identify the network model entities that are the most important to define the difference between CS exposed and sham animals, we have compared the leading nodes across the exposure regimens (1 through 7 months). Leading nodes are highlighted (red if the node is up regulated, and black if it is down regulated).



The cell cycle network is more perturbed at 3 months as compared to 7 months of CS exposure and the perturbation is decreased upon smoking cessation.



Several nodes were important after both short and long CS exposures as well as after cessation. Many leading nodes that disappear with prolonged exposure are linked to S-phase entry (red arrows) and the leading nodes that persist in all exposure scenarios are more related to G2 phase of the cell cycle (blue arrows). Interestingly, while the overall perturbation of cell cycle after cessation is low as compared to 7 month exposure (A), the leading nodes that persist are essentially the same (B). As a conclusion, leading node investigation enables a detailed understanding of the perturbation.

References

1. Hoeng, J. et al. A network-based approach to quantifying the impact of biologically active substances. Drug Discov Today 17, 2012.
2. Gebel, S. et al. Construction of a Computable Network Model for DNA Damage, Autophagy, Cell Death, and Senescence Bioinformatics and Biology Insights 7, 2013.
3. Schlage, W. K. et al. A computable cellular stress network model for non-diseased pulmonary and cardiovascular tissue. BMC Syst Biol 5, 2011.
4. Westra, J. W. et al. Construction of a computable cell proliferation network focused on non-diseased lung cells. BMC Syst Biol 5, 2011.
5. Westra, J. W. et al. A Modular Cell-Type Focused Inflammatory Process Network Model for Non-Diseased Pulmonary Tissue. Bioinformatics and Biology Insights 7, 2013.
6. Park, et al. Construction of a Computable Network Model of Tissue Repair and Angiogenesis in the Lung. J Clinic Toxicol 2013, S12-002, 2013.
7. Martin, F. et al. Quantification of biological network perturbations: Impact assessment and diagnosis using causal biological networks. (2013).
8. Martin, F. et al. Assessment of network perturbation amplitude by applying high-throughput data to causal biological networks. BMC Syst Biol 6, 2012.
9. Thomson, T. M. et al. Quantitative Assessment of Biological Impact Using Transcriptomic Data and Mechanistic Network Models. Toxicology and Applied Pharmacology (2013).
10. Hoeng, J. et al. Case study: the role of mechanistic network models in systems toxicology. Drug discovery today (2013).
11. Beckett, E. L. et al. A new short-term mouse model of chronic obstructive pulmonary disease identifies a role for mast cell tryptase in pathogenesis. J Allergy Clin Immunol 131, 2013.
12. Belcastro, V. et al. Systematic Verification of Upstream Regulators of a Computable Cellular Proliferation Network Model on Non-Diseased Lung Cells Using a Dedicated Dataset. Bioinformatics and Biology Insights 7, 2013.

Conclusion

The hierarchical structure of the network models offers mechanistic understanding on the biological impact of the CS exposure, revealing multiple network models, subnetworks and nodes, whose NPA scores are consistent with measured experimental endpoints. Moreover, even when there is no phenotypic information available, network scoring provides valuable mechanistic insight and a testable hypothesis.

Cigarette smoke is known to affect the cell cycle, but its role is not clear in emphysema development. Some of the mechanisms identified here might shed light into the cell cycle processes that are involved in emphysema development and recovery in the mouse model of COPD.

PMI RESEARCH & DEVELOPMENT

CMSB 2013
Kleusterneuburg, Austria
23-25 September 2013

Philip Morris International Research & Development, Quai Jeanrenaud 5, 2000 Neuchâtel, Switzerland
T: +41 58 242 21 11, **F**: +41 58 242 28 11, **W**: www.pmi.com