

A combined computational/experimental strategy to map phosphatases on growth pathways

P.F. Gherardini¹, F. Sacco¹, S. Paoluzi¹, J. Saez-Rodriguez², M. Helmer-Citterich¹, A. Ragnini-Wilson^{1,3}, L. Castagnoli¹, G. Cesareni^{1,4}✉

¹Department of Biology, University of Rome "Tor Vergata", Italy

²EMBL-EBI, Hinxton, UK and EMBL-Genome Biology Unit, Heidelberg, Germany

³High-throughput Microscopy facility; Department of Translational and Cellular Pharmacology, Consorzio Mario Negri Sud, SM. Imbaro, Italy

⁴IRCCS Fondazione Santa Lucia, Rome, Italy

Motivations

Large-scale screenings allow linking the function of poorly characterized genes to phenotypic readouts. According to this strategy, genes are associated to a function of interest if the alteration of their expression perturbs the phenotypic readouts. However, given the intricacy of the cell regulatory network, such screenings often provide mostly qualitative results, as it is difficult to identify the molecular mechanisms underlying the observed phenotype. In recent years computational modeling has emerged as a powerful tool to investigate biological signaling. In general these models are necessarily limited in size due to their complexity. Conversely large-scale perturbation screenings provide a higher-level view of a larger number of proteins. In this work we aim to bridge the gap between these two worlds in order to obtain a higher-detail mapping of gene products onto complex pathways on a large scale. This strategy was applied to map the poorly characterized family of human phosphatases on pathways related to cell growth.

Methods

The strategy we have developed is based on two complementary datasets, which are conceptually linked by "sentinel proteins". These are defined as a number of molecular readouts that define the state of the cell, given specific experimental conditions. The datasets we used are: 1) A detailed mechanistic model describing a pathway of interest. The model includes as entities the sentinel proteins. This is constructed starting from low/medium-throughput experiments. 2) A high-throughput perturbation screening where the molecular readout is defined by the sentinel proteins. The core of the strategy is to use the signaling model to simulate the result of up/down regulating each protein in the signaling pathway. Each perturbation results in a

predicted cell state defined as the calculated activity of the sentinel proteins. By matching this signature with the experimental profile obtained in the high-throughput screening it is possible to infer the target and effect (activating/inhibitory) of each protein screened, thus defining its "entry-point" in the network. For instance, if the silencing of a gene results in the same cell state obtained when protein A is up-regulated in the simulation, we can infer that the gene modulates the activity of protein A. The biological modeling was performed using CellNetOptimizer. This software allows the construction of boolean logic models which are represented as signed direct graphs, representing activating/inhibitory relationship between proteins. CNO then optimizes the topology of the model against a dataset of experimental data in order to remove connections that are not relevant in the specific cell system and to integrate using AND/OR logic gates multiple stimuli acting on the same protein. The optimization procedure is repeated 1000 times both because of its stochastic nature and also because the training data are generally not sufficient to fully constrain the model. By performing subsequent simulations on this family of 1000 models it is possible to average out the inconsistencies present in any single model. Moreover by this approach one obtains quantitative predictions even though a given node can only be on (1) or off (0) in any single model. We also extended this strict boolean approach to allow the simulation of three different states for a protein, i.e. control, up and down regulation.

Results

We assembled from the literature a network describing cell growth pathways. This model includes 34 species and 59 stimulatory or inhibitory interactions and was optimized using CNO against a dataset of experimental data ob-

tained in HeLa cells. The model also includes five sentinel proteins, whose activation status defines the "cell state". We used the results of a siRNA screening of the human phosphatase family that yielded 58 proteins whose silencing affect the activity of one of the sentinel proteins. The interference of 35 of the 58 phosphatase hits (60%) results in a profile that matches one of those inferred by inactivating or activating in silico one of the nodes of the model. The correctness of this mapping, and thus the predictive capabilities of the model, was demonstrated in a number of experiments. In particular one experiment confirmed that the over-expression of four

phosphatases, which were mapped to different positions of the signaling network (i.e. upstream and downstream of AKT and ERK), differentially affect the activity of two readouts (RAF1 and AKT activation) that were not considered in the mapping procedure. In conclusion in this study we developed a novel strategy to map perturbations screening onto complex pathways. The proposed mapping strategy is general and could be used in combination with the results of such large screenings to achieve a more detailed mechanistic description of the molecular mechanisms by which genes or small molecules determine phenotype modulation.