Discovering drug mode of action using reverse-engineered gene network models

Bansal M.(1,2), Della Gatta G.(1,3), Wierzbowski J.(5), Ambesi-Impiombato A.(1,4), Gardner T.(5) and Di Bernardo D.(1,2)

(1) Telethon Institute of Genetics and Medicine, Via Pietro Castellino 111, Naples, Italy (2) European school of Molecular Medicine (SEMM), Naples, Italy (3) Second University of Naples, Italy. (4) Department of Neurosciences, University of Naples "Federico II", Italy (5) Department of Biomedical Engineering, Boston University

Motivation

A critical step in drug development is the optimization of the efficacy and specificity of candidate therapeutic compounds. Ideally, optimization is carried out using knowledge of the drug's mode of action, i.e., the gene products with which a drug functionally interacts (drug targets). These drug targets may include genes that mediate the therapeutic effects of the drug, as well as genes that mediate undesirable side-effects. However, for many drug candidates the targets are unknown and difficult to identify among the thousands of genes in a typical genome. Previously, we developed an algorithm to identify drug targets in yeast using multiple perturbations to a cell and by measuring the response at steady-state (di Bernardo et al, Nature Biotechnology, in press). Here, we report a novel computational approach for rapidly identifying drug targets using time-course gene expression profiles. The approach filters expression profiles using a reverse-engineered gene-network model to distinguish the targets of compounds from the genes that exhibit only secondary responses. We tested this approach experimentally in E coli and show that it can overcome some of the experimental and computational limitations of existing chemogenetic approach for identifying a drug's mode of action.

Methods

Algorithm: We modeled the network as a system of ordinary differential equations describing the time evolution of the mRNA concentration in the network of N genes: $X'(t) = A X(t) + B U(t)$ (1) where X(t) is a Nx1 vector representing the measured concentrations of mRNAs at time t following the treatment of the cells with a compound with unknown mode of action; X' represents the measured rate of change of X at time t. U(t) is a 1x1 vector representing the external perturbation (i.e. drug treatment) to the rate of accumulation of X at time t, B is Nx1 vector which represents the effect of the perturbation (U) on the genes, and A, the connectivity matrix, is an NxN matrix of coefficients describing the regulatory interactions between the elements in X. The unknowns are the connectivity matrix A and B. Specifically, the element i of B will be different from 0 if gene i is a direct target of the compound, and 0 otherwise. The magnitude of element i of B is related to the strength of the interaction between the compound and the gene. To solve equation (1) for A and B in our algorithm, we first apply a smoothing filter thus reducing the fluctuations in the data introduced by the noise. Second, we increase the number of time points by interpolating the smoothed data. Third, we apply Principle Component Analysis (PCA) to reduce the number of unknown parameters to be estimated from the expression data. Experimental setup: The bacterial strain MG1655 was grown over night at 37 C. The time course experiment consisted in the induction with 10 microg/ml of Norfloxacin and extraction of the total RNA at the following time points: 0min, 12min, 24min, 36min, 48min and 60min from the drug treatment. Norfloxacin acts by damaging the DNA and therefore it directly

activates the recA protein and subsequently the SOS pathway. Each experiment was done in triplicate; positive controls were done at 24min and 60min from the induction with Norfloxacin. We extracted total RNA at each time point and measured with Affymetrix E. coli GeneChip. We then processed the microarray data using the dChip algorithm that performs normalization and gene expression estimation using the algorithm described by (Gautier et al, 2004).

Results

Simulation Results: In order to test the algorithm, we generated 100 random stable networks of N = 1000 genes with an average of 100 connections per gene. For each network, we randomly applied a constant perturbation to 100 (out of 1000) of the genes, thus simulating the treatment with a compound. The results of the simulations showed that the algorithm could correctly identify most of the 100 genes directly perturbed by the compound (Sensitivity=50% Positive Predictive Value=40%). Experiment Results: By applying our algorithm on the preprocessed microarray data 300 genes were found to respond significantly to the perturbation experiment. The algorithm correctly predicted recA to be the direct target of the drug (ranked as the 5th strongest element of vector B). Also the top 50 genes with the highest magnitude according to their value in the vector B genes were found to belong to SOS pathway, in agreement with the known mode of action of the drug. Our algorithm may represent a powerful tool to speed up drug development.

Contact email: dibernardo@tigem.it
URL:
http://www.tigem.it/Research/Personal%20Web%20Page_files/dibernardo/Website/Group_Frame_blue.htm