Drug Target Identification from Inferred Gene Networks: a computational and experimental approach

Diego Di Bernardo(1), Timothy S Gardner(2), James J Collins(2)

(1) Telethon Institute of Genetics and Medicine, Via Pietro Castellino 111, Naples, Italy
(2) Department of Biomedical Engineering, Boston University, MA, USA
dibernardo@tigem.it

Keywords: gene network, drug target, computational biology

Introduction

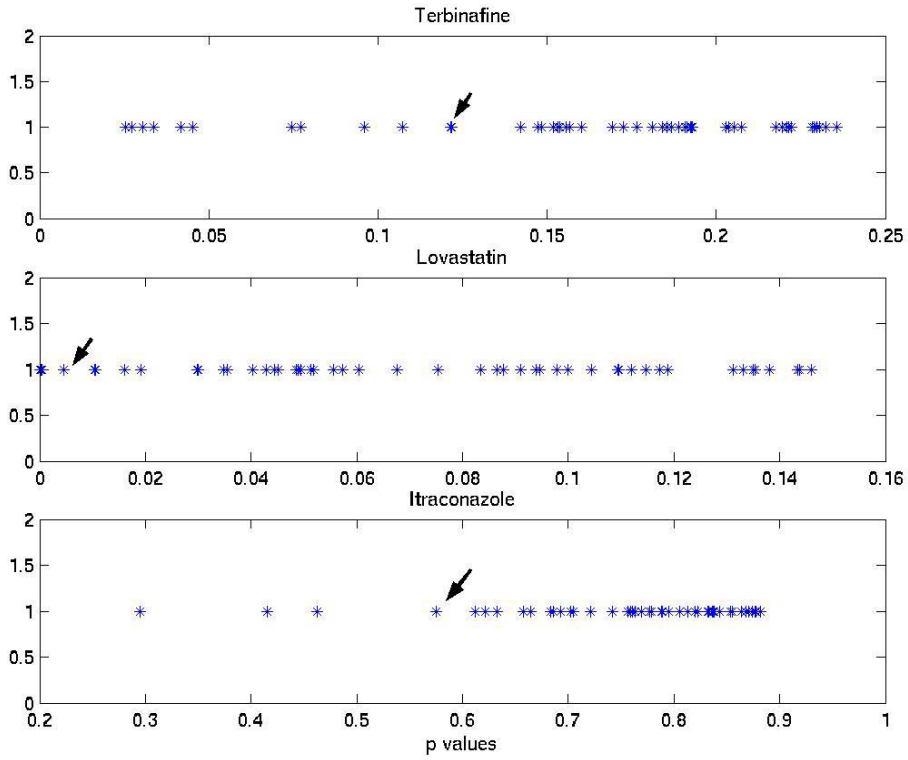
Genome-wide gene expression profiles provide a means to discover the direct mediators of biologically active compounds. We have already shown that it is possible to infer a predictive model of a genetic network by overexpressing each gene of the network and measuring the resulting expression at steady state of all the genes in the network[1]. This approach however requires the perturbation of each gene and the measurement of the perturbation magnitude. In this work we explored the possibility of inferring predictive models of large genetic networks without requiring the knowledge of which genes have been perturbed and by what amount. The network identification algorithm here described allows to infer a model of a genetic network from perturbation experiments for which the perturbed genes are not known. This model can be used to identify the target gene, or genes, of a given drug.

Algorithm and Methods

We assume to have a data matrix **X** (NxM) containing in each row ***j*** and column ***k*** the change (ratio) of mRNA concentration of gene ***j*** following the ***k***-th perturbation experiment, compared to untreated cells. The algorithm is divided in two parts. In part (1) it estimates the most likely genes that have been perturbed in each perturbation experiment. In part (2) it used the data matrix **X** and the knowledge of which genes have been perturbed in each experiment to infer a connectivity matrix **A** describing the regulatory network among the N genes. The algorithm is recursive: once an estimate of matrix **A** has been found, this is used to estimate the perturbations to each gene. Using the estimate of the perturbations, a more accurate estimate of **A** can be obtained, and so on. Once matrix **A** has been inferred, this can be used to identify the gene target of a given drug, starting from the steady-state mRNA change (ratio) of the N genes following the treatment with the drug. We tested our approach on a publicly available ”compendium” dataset [2] consisting of 300 different steady-state genome-wide expression profiles in S. cerevisiae measured using 2 color cDNA microarrays. Transcript levels of mutated or compound treated cultures where compared to wild type or mock treated cultures. 276 KO mutants, 11 tetracycline-regulatable alleles of essential genes and 13 drugs were profiled. To test the algorithm, we selected from the 14 compound expression profiles in the data set, the three compounds corresponding to the ones with the best characterised target in literature: Itraconazole an antifungal drug that binds to erg11p, Lovastatin, used to treat hypercholesterolemia, that acts by inhibiting HMG-CoA reductase (hmg1,hmg2), and Terbinafine that acts specifically on erg1 in the ergosterol biosynthesis pathway. Our data matrix **X** consisted on 6316 rows (number of genes) by 287 columns (number of experiments) containing the steady state gene expressions following the 276 KO experiments and the 11 tet-regulatable genes experiments.

Results and Conclusion

Using the algorithm we inferred the network connectivity matrix **A** from the data matrix **X**. Using **A** we were able to predict correctly the direct targets for all three drugs. The algorithm assigns a p value to each of the 6316 gene. Those genes with a low p-value are the identified targets of the drug. In Figure 1 we show the 50 genes with the smallest p value for each of the three drugs tested. The arrows show the real target of the drug as reported in literature. The real target is always among the top 10 most significant targets. Our algorithm, after further validation, could represent an original and innovative method to drug target discovery.



References

[1] Gardner TS, di Bernardo D, Lorenz D and Collins JJ. (2003) *Inferring gene networks and identifying compound mode of action via expression profiling.* Science 301, 102-5.

[2] Hughes TR, et al (2000) *Functional discovery via a compendium of expression profiles.*
Cell 102,109-26.