Computational Causal Reasoning Models of Mechanisms of Androgen Stimulation in Prostate Cancer

D. Pratt, W. Hahn, Matthews, P. Febbo, R. Berger, B. Duckworth,
J. Levy, T. Segaran, J. Sun, B. Ladd, and K. Elliston

Abstract- High-throughput transcriptional analyses of tissue samples can yield datasets describing significant differences in the expression of hundreds - or even thousands - of genes. In principle, this rich source of data can provide a systems-level view of the biological processes in an experiment, leading to testable hypotheses describing the mechanisms that led to the observed changes. But typically, the integration of hundreds of observations to infer the active biological networks is an unmanageable task, limiting the analysis to categorization of the changed genes by annotations and by patterns of modulation. To identify disease mechanisms, compound mechanisms, and biomarkers from high-throughput systems biology experiments requires the development of a model of biology. We describe the development of a very large-scale causal, computable model of biology and its specific application in the identification of molecular cause and effect hypotheses of mechanisms underlying the effects of androgen stimulation in the LNCaP prostate carcinoma cell line. In contrast to previous LNCaP studies in which genes have been hierarchically clustered by their pattern of response to androgen, our causal reasoning methodology identifies possible explanations in terms of discrete and testable molecular mechanisms. We have inferred changes in cell proliferation and fatty-acid synthesis transcriptional control mechanisms based on gene expression changes in transcriptional targets of proteins such as RB1, E2F1,2,3, and SREBF1,2. Further analysis has identified multiple causal pathways linking the activity of the androgen receptor (AR) to these processes, providing succinct, testable hypotheses for subsequent experimentation. Our causal reasoning methodology operates on significant differences in molecular states measured in the biological systems being compared, in this case the RNA expression measurements of LNCaP cells with and without androgen stimulation. For purposes of the analysis, differential measurements are abstracted to "State Changes", qualitative changes in state: increased, decreased, or measured but unchanged. The ontology of the model distinguishes details of biological entities and their activities such that State Changes derived from different modalities of measurement map to distinct objects in the model. For example, the RNA expression differences are mapped to objects representing gene expression processes, such as "exp(AR)" representing the expression of the AR gene. In other studies, differential proteomic measurements have been mapped to objects representing the abundance of specific proteins, such as NRDG1. The analysis of the LNCaP State Changes was performed using a model of human biology encoding specific cause – effect findings from public and proprietary sources, including peer-reviewed scientific literature and databases, further augmented by findings from rat and mouse biology translated to their human homologs. The model currently contains information derived from over 25,000 literature citations on over 35,000 biological concepts such as transcripts, proteins and metabolites and over 188,000 cause – effect relationships. The analysis generated discrete hypotheses by determining, for each observed State Change, the possible events that could lead to that change. This Causal Reasoning process starts by determining the most immediate possible causes and then works backward through the model to identify entire upstream networks, resulting in thousands of hypotheses, ranked by the concordance of their predictions with observed State Changes. In a critical step, multiple additional criteria prioritize and filter the hypotheses to produce a small set of likely explanations, reducing the analysis to a human scale. Overall, the methodology is an example of computer-aided reasoning enhancing and extending the abilities of researchers. This study identified new targets for RNAi-mediated gene knockdowns to study androgen control of cell proliferation in prostate cancer. Causal Reasoning and modeling provides a valuable tool for the analysis of molecular profiling data, can identify distinct mechanisms of action and can be used to design experiments to probe those mechanisms. Causal Reasoning is effective in the study of cellular models of prostate cancer, and this systems approach is broadly applicable to identifying the molecular mechanisms of cancer.

D. Pratt, A. Matthews, B. Duckworth, J. Levy, T. Segaran, J. Sun, B. Ladd, and K. Elliston are with Genstruct Inc. Cambridge, MA, 02140 (617-547-5421, dpratt@genstruct.com)
W. Hahn, P. Febbo and R. Berger were with the Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02115 USA

1-4244-0033-3/06/$20.00 ©2006 IEEE.

This page was left blank intentionally