

# Joint estimation of causal effects from observational and intervention gene expression data

Gregory Nuel

**Background:** In recent years, there has been great interest in using transcriptomic data to infer gene regulatory networks. For the time being, methodological development in this area has primarily made use of graphical Gaussian models for observational wild-type data, resulting in undirected graphs that are not able to accurately highlight causal relationships among genes. In the present work, we seek to improve the estimation of causal effects among genes by jointly modeling observational transcriptomic data with arbitrarily complex intervention data obtained by performing partial, single, or multiple gene knock-outs or knock-downs.

**Results:** Using the framework of causal Gaussian Bayesian networks, we propose a Markov chain Monte Carlo algorithm with a Mallows proposal model and analytical likelihood maximization to sample from the posterior distribution of causal node orderings, and in turn, to estimate causal effects. The main advantage of the proposed algorithm over previously proposed methods is its flexibility to accommodate any kind of intervention design, including partial or multiple knock-out experiments. Using simulated data as well as data from the Dialogue for Reverse Engineering Assessments and Methods (DREAM) 2007 challenge, the proposed method was compared to two alternative approaches: one requiring a complete, single knock-out design, and one able to model only observational data.

**Conclusion:** The proposed algorithm was found to perform as well as, and in most cases better, than the alternative methods in terms of accuracy for the estimation of causal effects. In addition, multiple knock-outs proved to contribute valuable additional information compared to single knock-outs. Finally, the simulation study confirmed that it is not possible to estimate the causal ordering of genes from observational data alone. In all cases, we found that the inclusion of intervention experiments enabled more accurate estimation of causal regulatory relationships than the use of wild-type data alone.

**Keywords:** Causal inference, Gaussian Bayesian network, intervention calculus, Metropolis-Hastings, maximum likelihood.