

The TAO-Gen Algorithm for Identifying Gene
Interaction Networks with Application to SOS
Repair in E. coli

Takeharu Yamanaka, Hiroyoshi Toyoshiba, Hideko
Sone, Frederick M. Parham and Christopher J. Portier
doi:10.1289/txg.7105 (available at <http://dx.doi.org/>)
Online 21 July 2004



The National Institute of Environmental Health Sciences
National Institutes of Health
U.S. Department of Health and Human Services

The TAO-Gen Algorithm for Identifying Gene Interaction Networks with Application to SOS Repair in *E. coli*

Takeharu Yamanaka¹, Hiroyoshi Toyoshiba¹, Hideko Sone¹, Frederick M. Parham¹
and Christopher J. Portier¹

¹Laboratory of Computational Biology and Risk Analysis, National Institute of Environmental
Health Sciences, Research Triangle Park, NC 27709 USA

Send correspondence to:

Christopher J. Portier, Ph.D.
Laboratory of Computational Biology and Risk Analysis
National Institute of Environmental Health Sciences
P.O. Box 12233, MD: A3-06
111 Alexander Drive
Research Triangle Park, NC 27709

919-541-3802 (Phone)
919-541-3647 (Fax)
portier@niehs.nih.gov

Short Running Title: TAO-Gen Algorithm and SOS Repair Pathways

Key words: gene networks, microarray, Bayesian model selection, SOS repair, toxicogenomics

Acknowledgement: We would like to thank Drs. Thomas Darden, Schaaper Roel and Nigel Walker for helpful comments and advices.

Title, Authors, Author's affiliations, Name and Address of corresponding author

Short running title, Keywords, Acknowledgement

Contents

Abbreviations

Abstract

Introduction

Material and Method

Results

Discussion

Conclusion

References

Tables

Figure legends

Figures

Abbreviations Used

E. coli – Escherichia coli

DNA - deoxyribonucleic acid

ssDNA – single stranded deoxyribonucleic acid

mRNA – messenger ribonucleic acid

MMC - mitomycin C

MCMC – Markov Chain Monte-Carlo

NIR – Network Identification by multiple linear Regression

TAO-Gen - Theoretical Algorithm for identifying Optimal GENE interaction networks

Abstract

One of the major unresolved issues in the analysis of gene expression data is the identification and quantification of gene regulatory networks. Several methods have been proposed for identifying gene regulatory networks, but these methods predominantly focus on the use of multiple pairwise comparisons to identify the network structure. In this paper, a method is developed for analyzing gene expression data to determine a regulatory structure consistent with an observed set of expression profiles. Unlike other methods, this method goes beyond pairwise evaluations by using likelihood-based statistical methods to obtain the network that is most consistent with the complete data set. The proposed algorithm performs accurately for moderate-sized networks with most errors being minor additions of linkages. However, the analysis also indicates that sample sizes may need to be increased to uniquely identify even moderate-sized networks. The method is used to evaluate interactions between genes in the SOS signaling pathway in *E. coli* using gene expression data where each gene in the network is over-expressed using plasmids inserts.