

Regulatory Network Dependencies From Quantitative Trait Loci

Manjunatha Jagalur,¹ David Kulp,²

Keywords: Gene Regulatory Network, QTL Analysis, Bayesian Network

1 Introduction.

Jansen and Nap (2001) introduced the concept of "genetical genomics" in which putative disease-associated loci are identified through linkage analysis of gene expression data treated as quantitative phenotypes. Since then works by Brem, et al (2002), Schadt, et al (2003), Bystrykh, et al (2005), and others have shown that QTL analysis of large scale gene expressions data can implicate trans-acting loci as putative upstream modulators of gene expression in regulatory pathways. Inference of regulatory networks from large-scale gene expression data sets has received much attention, with some success in lower organisms such as yeast, since Friedman, et al (2000) demonstrated the use of Bayesian Networks (BNs) to deduce causal relations among genes from correlations in expression. Recently, Zhu, et al (2004) described a method to integrate genetic marker data into such gene expression BNs as priors. The key insight being that genetic marker data provides explicit directionality information that is generally indeterminate using the conventional BN method. We expand on this integrative approach by proposing a simple gaussian linear model of gene expression that explicitly incorporates additive genetic effect from one or more latent modulators. We show through simulation that such a model can correctly predict small regulatory modules. We introduce an alternative linkage map to visualize QTGs (quantitative trait genes) and propose an EM strategy to optimally fit model parameters when the genotypes of the putative modulator genes are unknown.

2 Genetic effect and gene expression modeling

While constructing the bayesian network using expression data, target gene is assumed to be a function of its regulators. If the underlying regulation is assumed to be linear gaussian, then the function will be linear. In multiple interval mapping, usually relation between target gene and the genotype of the regulators is assumed to be linear. Linear gaussian model does not take into account genotype of the regulators and interval mapping approach does not take into account the expression levels of the regulators. Also interval mapping does not provide with accurate location of the regulator. Our goal is to combine these approaches to come up with a model which uses both genetic and expression data of the genes to infer regulatory modules. We tried modeling the target gene as sum of weighted sum expression of the regulators and weighted sum of product of expression and genotype of the regulator. Depending on the values of these weights, different scenarios of regulation can be simulated. If the weight related to only expression is high, then it represents a scenario where genotype of the regulator is of less importance and if the weight related to expression and genotype is high, it represents a scenario where genotype plays a major role in the regulation.

¹Department of Computer Science, University of Massachusetts Amherst, 140 Governors Drive, Amherst, MA. 01003-9264. E-mail: manju@cs.umass.edu

²Department of Computer Science, University of Massachusetts Amherst, 140 Governors Drive, Amherst, MA. 01003-9264. E-mail: dkulp@cs.umass.edu

We simulated various scenarios and tried recovering the regulator set. Initial results were optimistic.

3 Quantitative trait genes

Interval mapping provides us with a subset of genome in which the regulator may reside. We need to use further analysis to find the actual genes residing in that subset. Also in many cases as expression data of the regulator has not been used, relations found using interval mapping can be weak. To address the problem we propose *Quantitative Trait Gene Mapping*(QTG) technique which uses both genetic and expression data of the regulator. In this approach we assume expression of the target gene to be a function of expression and genotype of the regulator. Then we calculate the LOD score, which is the log of ratio of conditional probability of target gene and probability target given null hypothesis. We simulated expression of a gene using the approach mentioned in previous section.

References

- [1] N Friedman, M Linial, I Nachman, D Peer. *Using Bayesian networks to analyze expression data*, Journal of Computational Biology, vol. 7, pp 601-620. (2000).
- [2] Perez-Enciso M, Toro MA, Tenenhaus M, Gianola D. *Combining gene expression and molecular marker information for mapping complex trait genes: a simulation study*, Genetics, 164(4), pp 1597-606. (2003).
- [3] RB Brem, G Yvert, R Clinton, L Kruglyak. *Genetic dissection of transcriptional regulation in budding yeast*, Science, 26;296(5568), pp 752-755. (2002).
- [4] Schadt EE, Monks SA, Drake TA, Lusk AJ, Che N, Colinayo V, Ruff TG, Milligan SB, Lamb JR, Cavet G, Linsley PS, Mao M, Stoughton RB, Friend SH. *Genetics of gene expression surveyed in maize, mouse and man*, Nature, Mar 20;422(6929), pp 297-302.(2003).
- [5] Jansen RC, Nap JP. *Regulating gene expression: surprises still in store*, Trends in Genetics, 20(5), pp 223-225.(2004).