TimeDelay-ARACNE: Reverse engineering of gene networks form time-course data by an Information theoretic approach

Zoppoli P[2], Morganella S[1-2], Ceccarelli M[1-2]

[1] Research Centre on Software Technologies - RCOST, University of Sannio, Benevento, Italy
[2] Biogem s.c.a.r.l., Ariano Irpino (Avellino), Italy

Motivation

Today one of Molecular biology's aims is the knowledge of the cell networks in order to understand gene function regulations and eventually modify part of it. Even if gene-gene interactions only partially describe the real networks because of post-transcriptional modifications and protein regulations, using the microarray technology it is possible to extract measurements of thousands of genes into a single analysis step having a picture of the cell gene expression. Inferring, or 'reverse-engineering', gene networks can be defined as the process of identifying gene interactions from experimental data through computational analysis. Actually, several methods have been developed to infer the gene networks from steady-state data, but very little literature is produced about time-course data, so the developing of algorithms to infer gene networks from time-series measurements is a current challenge into bioinformatics research area. Such an algorithm could be useful to understand gene network reaction to a new developed drug, to understand gene regulations in steam cells differentiation or to better investigate eucaryote cell cycle.

Methods

We propose an approach to infer the gene regulatory networks by time-series measurements modifying a previously published algorithm: ARACNE (an inference algorithm based on Mutual Information developed for steady-state experiments). TimeDelay ARACNE extends to time-course data ARACNE retrieving time statistical dependency between gene expression profiles. The TimeDelay ARACNE's idea is that the expression of a gene at a certain time could depend by the expression level of another gene at previous time point or at very few time points before. TD-ARACNE is also a 2 steps algorithm: first there is network construction than network pruning. Our algorithm identifies candidate interactions by Kernel estimating pairwise time delay gene expression profile mutual information. Mutual information is defined as the reduction in uncertainty about variable . after observing a second random variable .. Since MI is reparameterization invariant, TD-ARACNE, as ARACNE, uses copula-transform (i.e., rank-order) x and y for MI estimation; the range of these transformed variables is thus between 0 and 1, and their marginal probability distributions are manifestly uniform. This decreases the influence of arbitrary transformations involved in microarray data preprocessing and removes the need to consider position-dependent kernel widths which might be preferable for non-uniformly distributed data. After MIs estimation TD-ARACNE filters them using an auto-sets threshold in order to retrieve only statistical significant edges. TimeDelay ARACNE uses the data processing inequality (DPI) two times: first time uses the DPI to filters out indirect interactions from each time delay; then introduce the Influence concept as the max time-dependent MIs (gA, gB) in a user defined delay k.t to capture the more informative MI; at last filters out indirect Influences using the DPI a second time, retaining only best connections.

Results

TimeDelay ARACNE can infer small local networks of time regulated gene-gene interactions detecting their versus and also discovering some cyclic interactions. Our approach can infer regulatory interactions even when only a medium-small number of measurements are available. We test our approach both on synthetic networks and on an eleven-gene sub-network part of the cell cycle pathway in yeast S. cerevisiae. ARACNE's performance is also compared against TSNI and Banjo. The ability of our approach to infer the gene regulatory networks is tested for accuracy and recall. TimeDelay Aracne in-silico results show direct correlation with time points and seem to be not very strictly correlated with network gene numbers. On yeast data TD-ARACNE recovers many gene-gene edges with both a good accuracy and recall. In both cases TimeDelay ARACNE performs better than well known algorithms used for comparison.

Contact : zoppoli@biogem.it