**Introduction**

Inference of gene regulatory networks (GRNs) is one of the key problems in systems biology and it holds the promise of a better understanding of crucial functions of living organisms. Despite the research efforts devoted to this task, the reconstruction accuracy of the proposed solutions is far from being optimal in many situations due to drawbacks related to the methods and the available data.

RegANN is a novel method based on an ensemble of multilayer perceptrons for reverse engineering gene networks. This inference algorithm builds a regressor for each gene in the network. It estimates the neighborhood of each gene (the correlations among one gene and all the others) independently and then it joins these neighborhoods to form the overall network. The main feature of the novel method here presented is that it makes no assumptions about the nature of the relationships between the variables, potentially capturing high-order and non-linear dependencies between expression patterns.

**RegANN: Network inference using ANN**

To infer gene regulatory networks we adopt an ensemble of feed-forward multilayer perceptrons trained using the back-propagation algorithm. Each member of the ensemble is essentially a multi-variable regressor (one to many) trained using an input expression matrix to learn the relationships (correlations) among a target gene and all the other genes in the network.

From each row of the gene expression matrix a set of input and output patterns is built to train a selected multilayer perceptron. Each input pattern corresponds to the expression value for the selected gene of interest. The output pattern is the row-vector of expression values for all the other genes for the given row in the gene expression matrix.

The topology of gene regulatory networks is obtained by applying a second procedure. The correlation of each gene with all the others is extracted by passing a purposely made test pattern to the regressor: considering separately each multilayer perceptron in the ensemble, a value of 1 is passed to its input neuron, consequently recording its output values. In this way, the correlation between the corresponding gene with all the others is obtained as a vector of values in [−1, 1].

By cycling through all the members of the regression system, we obtain the adjacency matrix of the sought gene network.

As comparison methods we select three alternative algorithms widely used in the literature:

- ARAANE [3], CLR [4], and KELLER [5].

**MCC: Matthews Correlation Coefficient**

The MCC is in essence a correlation coefficient between the observed and predicted binary classifications: it returns a value between −1 and +1. A coefficient value equal to +1 represents a perfect prediction. 0 indicates an average random prediction while −1 an inverse prediction. In the context of network topology inference the observed class is the true network adjacency matrix, while the predicted class is the inferred one. The Matthews Correlation Coefficient is obtained by the following equation:

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}.$$

**Results: Toy Experiments**

- Multiple Interaction
  - Original network
  - RegANN – inferred network
- Indirect Interaction
  - Original network
  - RegANN – inferred network

RegANN captures the correlation among genes: C → A, C → B and A → C. No correlation (less than 0.1) is recorded among D and the other genes.

**Conclusions**

The proposed method makes no assumptions about the nature of the relationships between the variables, capturing high-order dependencies between expression patterns and the direction of the interaction, as shown on selected synthetic toy examples. The evaluation indicates that the newly introduced RegANN shows accuracy scores that compare very favorably with all the other inference methods tested. Often outperforming the reference algorithm in the case of fixed binarization threshold. On the other hand, considering all the possible thresholds for the binarization of the inferred adjacency matrix (the AUC score) the differences among the methods tend to become irrelevant.

**Funding**

Funding by the European Union FP7 Project HipercDART, by the Italian Ministry of Health Project ISISAT (RF2007 conv. 42), and by the PAT funded Project ENVIRECHANGE.