Introduction

microRNAs (miRNAs) are important gene regulators. They control a wide range of biological processes and are involved in several types of cancers. Thus, exploring miRNA functions is important for diagnostics and therapeutics. However, there are still no feasible experimental techniques to discover miRNA regulatory mechanisms.

Existing computational methods discover the statistical relationships, such as correlations and associations between miRNAs and mRNAs at data level, such statistical relationships are not necessarily the real causal regulatory relationships that would ultimately provide useful insights into the causes of gene regulations (see the below figure).

The standard method for determining causal relationships is randomized controlled perturbation experiments. In practice, however, such experiments are expensive and time consuming. Our motivation for this study is to discover the miRNA–mRNA causal regulatory relationships from observational data.

Algorithm of inferring miRNA–mRNA causal effects [1]

Step 1: Identify differentially expressed genes, i.e. genes whose expression values vary significantly across the conditions (categories) of the dataset. The dataset will be the expression profiles of the differentially expressed miRNAs and mRNAs.

Step 2: Use the PC algorithm to estimate the causal structure, CPDAG (Completed Partially Directed Acyclic Graph).

Step 3: Estimate the causal effects of each miRNA on each mRNA using IDA [2].

Step 4: For each miRNA, rank the targets (mRNAs) based on their causal effect values for validation. We can also have a ranking for all the regulators (miRNAs) of a particular mRNA.

Results and application

Data:

We use the NCI-60 dataset for Epithelial to Mesenchymal Transition (EMT).

The input dataset is a 47 x 1678 matrix for 47 samples and 1678 variables (43 miRNAs and 1635 genes)

Ground Truths:

We conduct the follow-up controlled experiments for the miR-200 family to validate the computational results.

We measure the gene expression level in the MDA-MB-231 samples transfected with miR-200 family and the MDA-MB-231 samples without miR-200 family (controls). Genes that are differentially expressed between the control and transfected samples are considered as miR-200 targets.

Results:

<table>
<thead>
<tr>
<th></th>
<th>miRNA</th>
<th>mRNA</th>
<th># confirmed genes for miR-200a</th>
<th>p-value</th>
<th># confirmed genes for miR-200b</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 20</td>
<td></td>
<td></td>
<td>10</td>
<td>0.0544</td>
<td>14</td>
<td>0.0339</td>
</tr>
<tr>
<td>Top 50</td>
<td></td>
<td></td>
<td>22</td>
<td>0.0131</td>
<td>32</td>
<td>0.0125</td>
</tr>
<tr>
<td>Top 100</td>
<td></td>
<td></td>
<td>42</td>
<td>0.0103</td>
<td>53</td>
<td>0.1442</td>
</tr>
</tbody>
</table>

In [3], we integrate expression data and miRNA target information to find direct miRNA–mRNA causal regulatory relationships.

We also investigate the causal feedforward patterns of TF-miRNA-mRNA. The following networks shows an example of strong causal regulatory relationships between miR-200 family, TFs, and target genes.

Further Information

[1] Thuc Duy Le*, Lin Liu, Anna Tsykin, Gregory J. Goodall, Bing Liu, Bing-Yu Sun, and Jiuyong Li. Inferring miRNA-mRNA causal regulatory relationships from expression data, Bioinformatics 2013, 29(6), 765-771, 2013