Sensitizing GO Tools Using Multiple Ratio-Cutoffs for Gene Selection

Kenichi Hirotani¹
hirotani@wis02.ee.t.kanazawa-u.ac.jp

Yoichi Yamada¹
youichi@is.t.kanazawa-u.ac.jp

Miyuki Onda³
onda@cb.k.u-tokyo.ac.jp

Ken-ichiro Muramoto¹
muramoto@t.kanazawa-u.ac.jp

Kenji Satou²
ken@jaist.ac.jp

Takashi Ito³
ito@k.u-tokyo.ac.jp

¹ Department of Information and Systems Engineering, Faculty of Engineering, Kanazawa University, Kakuma-machi, Kanazawa, Ishikawa 920-1192, Japan
² Graduate School of Knowledge Science, Japan Advanced Institute of Science and Technology (JAIST), 1-1 Asahidai, Tatsunokuchi, Ishikawa 923-1292, Japan
³ Department of Computational Biology, Graduate School of Frontier Sciences, University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8561, Japan

Keywords: Gene Ontology, ratio-cutoff, p-value

1 Introduction

The computational analysis tools using Gene Ontology (GO) annotations [4] give us biologically meaningful contexts of a large amount of gene expression from microarray data. The most common requirement of such tools [1, 3] is to pre-extract a group of induced or suppressed genes from expression data for statistical tests. For this reason, the user has to apply a ratio-cutoff to the expression levels to select genes as a representative of the transcriptome. We here emphasize the importance of using multiple ratio-cutoffs for such analyses.

The yeast Pdr1 and Pdr3 are paralogous transcription factors (TFs) regulating drug-resistance genes. Based on the microarray data of the cells expressing artificially-activated Pdr1 and Pdr3, we previously revealed that the former and latter serve as a negative and positive regulators of salt tolerance, respectively [2]. We reached to this finding via a rather complex analysis using both the expression data and literature mining, because the conventional GO tools using several different ratio-cutoffs failed to find any GO terms related to salt tolerance. To cover such GO terms, we improved a tool as applying a series of continuous ratio-cutoffs to select genes for GO analysis automatically.

2 Methods and Results

2.1 Microarray Data of the Yeast Cells

We used the microarray data of the yeast cells expressing artificially-activated TFs [2]. For each gene, we calculated the ratio between the fold-inductions by the artificially-activated Pdr3 and Pdr1 to generate a dataset termed as Pdr3/Pdr1 data. Then, we applied a cutoff to the data to select genes that are induced more efficiently by Pdr3 than Pdr1. These genes are analyzed by a GO tool, GO term finder [1].

2.2 An Automated Application of Multiple Ratio-cutoffs

To examine multiple cutoffs, we developed a tool that automatically increases the ratio-cutoff by 0.5 from 1.0 to possible maximum value to select genes for GO analysis. We applied the tool to the Pdr3/Pdr1 data using a p-value cutoff of 0.05 to define significant GO terms. Consequently, we identified the GO term "response to salt stress" as a significant one at a ratio-cutoff of 1.5 but not at the generally-used cutoffs such as 3.0–5.0. Furthermore, we used the tool to analyze all the possible binary combinations of the 8 microarray data reported in [2] (Fig 1, 2). The results indicate that Pdr3 induces the genes related to "response to salt stress" more efficiently than any other transcription factors but Leu3. Although Leu3 regulates genes involved in the biosynthesis of branched chain amino acids, the results indicate its possible contribution to salt resistance.
Figure 1: Pair-wise comparison of TFs in terms of relative induction of genes assigned to the GO term "response to salt stress". A series of cutoffs were applied to the ratio of fold-inductions by each TF to select genes to be examined for significant association with the GO term. The gray cells indicate the comparison that showed a significant association with the GO term (P < 0.05) and each number is the maximum-ratio-cutoff to detect significance.

<table>
<thead>
<tr>
<th></th>
<th>Pdr3</th>
<th>Pdr3</th>
<th>Yrr1</th>
<th>Yor172</th>
<th>Ylr266</th>
<th>Ykl222</th>
<th>Put3</th>
<th>Leu3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pdr1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pdr3</td>
<td>1.5</td>
<td>1.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.5</td>
<td></td>
</tr>
<tr>
<td>Yrr1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yor172</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ylr266</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ykl222</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>1.5</td>
</tr>
<tr>
<td>Put3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu3</td>
<td></td>
<td>1.5</td>
<td>1.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2: p-values of GO term (response to salt stress) in ratio-cutoff shown in Figure 1.

3 Discussion

The GO analysis using an automated application of multiple ratio-cutoffs for gene selection allowed us to easily identify "response to salt stress" as a GO term significantly enriched in the Pdr3-regulated but not Pdr1-regulated genes. Since Pdr3 induced most of these genes ~1.5-fold better than Pdr1, one may overlook this trend without using multiple ratio-cutoffs. Thus, the use of multiple ratio-cutoffs for gene selection may sensitize GO tools to help us uncover biology hidden behind the huge expression data.

References


