Gene Co-regulation vs. Co-expression

Ya Zhang[1], Hongyuan Zha[2], James Z. Wang[3], Chao-Hsien Chu[4]
The Pennsylvania State University, University Park, PA 16802

Keywords: Microarray, gene expression, co-regulation, regulon, regulator

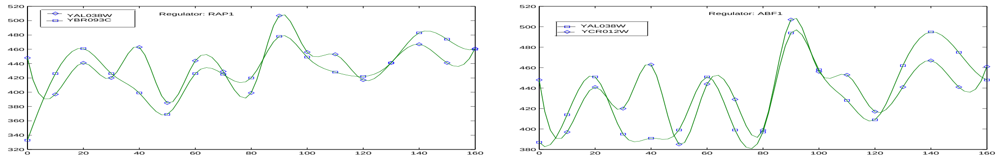
One important goal of analyzing gene expression data is to discover co-regulated genes. For a relatively long time, it has been assumed that similar patterns in gene expression profiles usually suggest relationships between the genes. According to [3], genes targeted by the same transcription factors tend to show similar expression patterns along time. Analyzing expression profiles of genes targeted by the same transcription factors revealed complex relationships between co-regulated gene pairs, including co-expression, time shifted, inverted, and inverted and time-shifted relationships. To investigate how co-regulation corresponds to co-expression, we retrieved regulator-regulon pairs from the Yeast Promoter Database [1] and examined the expression profiles of the regulons with the same regulator. We plotted expression profiles of target genes regulated by several regulators, respectively (Fig. 2). The gene expression data were generated by Cho et al[2]. They sampled 17 time points at 10 minutes time interval, covering nearly two full cell cycles of yeast *Saccharomyces cerevisiae*. As shown from the plots, the relationships among co-regulated genes are very complex and beyond the description of the four relationships identified by Yu et al [3]. We further identified the partial co-expression relationship between genes: gene profiles may simultaneously rise and fall in a sub-range of the time course rather than the overall time course. For example, among the set of genes regulated by $ABF1$, genes $CDC19$ ($YAL038W$) and $PGK1$ ($YCR012W$) show similar expression profile in the second half of the time course (Figure 1(B)). Partial time-shift relationship is exemplified by genes $PGK1$ ($YCR012W$) and $CDC9$ ($YDL164C$) (Figure 1(C)). Moreover, we also observed partial inverted and partial co-expression relationship between genes $PDR3$ ($YBL005W$) and $SNQ2$ ($YDR011W$) (Figure 1(D)). The above observation suggests that a regulator may only function in some particular stage of cell development, and therefore genes may be co-regulated in part of the time course, which corresponds to certain phase of cell cycle or cell development.

A gene may be regulated by multiple regulators. When genes sharing multiple regulators, their expression profiles are more likely to be similar. For example, genes HO ($YDL227C$), $CLN1$ ($YMR199W$), and $CLN2$ ($YPL256C$), who share regulator $CCBF$, SCB, and $SWI6$, exhibit close expression patterns (Figure 2). However, for some genes regulated by several regulators, its expression profile may only reflect the effect of a dominate regulator. For example, regulators of gene HO ($YDL227C$) include $BAS1$, $CCBF$, $MATalpha2$, and SCB. However, according to the plots of gene expression profiles, HO ($YDL227C$) is highly correlated genes regulated by $CCBF$ and SCB (Figure 2). Similar phenomena is shown by gene $PGK1$. The regulator of $PGK1$ ($YCR012W$) includes $GCR1$, $CPF1$, and $ABF1$. In this case, $PGK1$ ($YCR012W$) showed similar expression profile to some genes regulated by $GCR1$ and $ABF1$ in the second half of the time course (Figure 2).

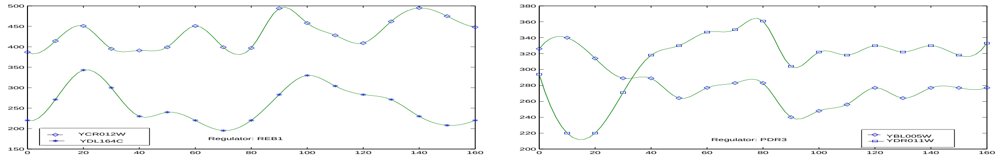
Current analysis of microarray gene expression focuses on relationships based on overall correlation in expression profile, identifying clusters of genes whose expression levels simultaneously rise and fall throughout a time course. However, genes may be regulated by different regulators over a time course. Co-regulating in part of the time course does not guarantee

---

[1] School of Information Sciences and Technology. E-mail: `yzhang@ist.psu.edu`
[2] Department of Computer Science and Engineering. E-mail: `zha@cse.psu.edu`
[3] School of Information Sciences and Technology. E-mail: `jwang@ist.psu.edu`
[4] School of Information Sciences and Technology. E-mail: `chu@ist.psu.edu`



(A) Partial co-expressed (B) Partial co-expressed



(C) Partial time shifted (D) Partial inverted and co-expressed

Figure 1: Expression profiles for co-regulated genes. The time-series gene expression data are from [2].

a global similarity in gene profiles. Therefore, new clustering algorithms are needed to address this issue. Several biclustering algorithms have been proposed to discover sets of genes that co-regulated in only part of the experiments conditions under study. However, these algorithms are not applicable to clustering gene expression time-series data because they ignored the internal relationship between time points. When analyzing the time-course gene expression data, it is necessary to consider the internal connection between time points and preserve the time locality in time-course gene expression data.

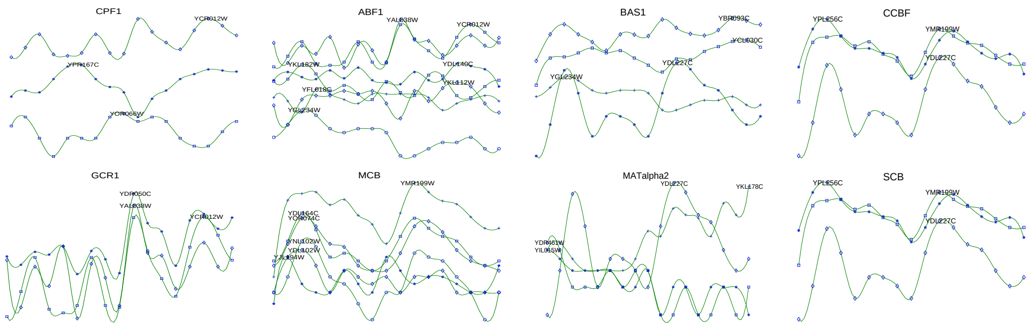


Figure 2: Expression profiles of co-regulated gene groups. Each curve represents the expression profile of a gene. Each sub-plot represents gene expression profiles of a co-regulated gene group. The gene expression time-series data are obtained from [2]. The time range is from 0 min to 160 min.

References

[1] http://cgsigma.cshl.org/jian/

[2] Cho,R.J., Campbell,M.J., Winzeler,E.A., Steinmetz,L., Conway,A., Wodicka,L. et al. (1998) A genome-wide transcriptional analysis of the cell cycle. *Mol. Cell*, 2, pp.65-73.

[3] Yu,H., Luscombe,N., Qian,J., and Gerstein,M. (2003) *Trends Genet* 19(8):422-7.