Comparison with previous studies

A comprehensive comparison of several published studies showing the fraction of significant modules (Bonferroni corrected P-value < 0.01) according to Gene Ontology annotations and the transcription factor binding data published by Lee et al (calculations done as described in the current paper).

The table below compares numbers from the other studies to the numbers in Table 2 (Lee et al.) and Table 3 (Gene Ontology) in the paper. The best scores for each test are given in bold italics (i.e. all scores in our current/previous studies that are better than the best score for the previous /current study).

	Study	Molecular function	Biological Process	Cellular compartment	Binding data by Lee et al.
Previously published studies	Segal et al. (2003b), Cell Cycle, 17 modules	0.235	0.353	0.353	*0.647*
	Segal et al. (2003b), Stress, 20 modules	0.400	0.600	0.200	0.450
	Segal et al. (2003a), 48 modules	0.250	0.359	0.229	0.229
	Beer et al. (2004)	0.413	0.587	0.426	0.306
Current study	Cell cycle	0.308	0.462	0.410	0.538
	Sporulation	0.262	0.535	*0.442*	0.133
	Diauxic shift	0.302	0.429	*0.444*	0.291
	Heat and cold shock	*0.538*	*0.635*	*0.596*	0.520
	Pheromone	*0.512*	*0.667*	*0.600*	0.388
	DNA-damage agents	*0.386*	*0.638*	*0.614*	0.351

Conclusion

Our approach performs on par or better than the previously published analyses when it comes to extracting significant binding site modules. Furthermore, since our modules are smaller (contain fewer genes), a larger fraction of the genes need to be co-annotated or

bound by the same transcription factors than in the previous studies to yield the same P-values. This clearly points to a stronger coherence in modules discovered in the current study.

(1) Segal, E., R. Yelensky, and D. Koller. 2003b. Genome-wide discovery of transcriptional modules from DNA sequence and gene expression. *Bioinformatics* **19 Suppl 1:** I273-I282.
(2) Segal, E., M. Shapira, A. Regev, D. Pe'er, D. Botstein, D. Koller, and N. Friedman. 2003a. Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat Genet* **34:** 166-176.
(3) Beer, M.A. and S. Tavazoie. 2004. Predicting gene expression from sequence. *Cell* **117:** 185-198.