Method for Identifying Transcription Factor Binding Sites in Yeast

Huai-Kuang Tsai
Genomics Research Center, Academia Sinica, Taipei, 115 Taiwan

Grace Tzu-Wei Huang
Genomics Research Center, Academia Sinica, Taipei, 115 Taiwan

Meng-Yuan Chou
Institute of Information Science, Academia Sinica, Taipei, 115 Taiwan

Henry Horng-Shing Lu
Institute of Statistics, National Chiao Tung University, 1001 Ta Hsueh Road, Hsinchu 30050, Taiwan

Wen-Hsiung Li
Genomics Research Center, Academia Sinica, Taipei, 115 Taiwan and

Department of Ecology and Evolution, University of Chicago, 1101 East 57th Street, Chicago, IL, 60637, U.S.A.

Identifying transcription factor binding sites (TFBSs) is helpful for understanding the mechanism of transcriptional regulation. The abundance and diversity of genomic data provide an excellent opportunity for identifying TFBSs. Developing methods to integrate various types of data has become a major trend in this pursuit. We develop a TFBS identification method, TFBSfinder, which utilizes several data sources, including DNA sequences, phylogenetic information, microarray data and ChIP-chip data. For a TF, TFBSfinder rigorously selects a set of reliable target genes and a set of non-target genes (a background set) to find over-represented and conserved motifs in target genes. A new metric for measuring the degree of conservation at a binding site across species and methods for clustering motifs and for inferring position weight matrices are proposed. For synthetic data and yeast cell cycle TFs, TFBSfinder identifies motifs that are highly similar to known consensuses. Moreover, TFBSfinder outperforms well-known methods.

[Wen-Hsiung Li, Department of Ecology and Evolution, University of Chicago, 1101 East 57th Street, Chicago, IL, 60637, U.S.A.; whli@uchicago.edu]