

Correlating overrepresented upstream motifs to gene expression: a computational approach to regulatory element discovery in eukaryotes

M.Caselle, P. Provero and F. Di Cunto

Universita' di Torino, Dip. di Fisica Teorica

Gene regulation in eukaryotes is mainly effected through transcription factors binding to rather short recognition motifs generally located upstream of the coding region. We present a novel computational method to identify regulatory elements in the upstream region of eukaryotic genes. The genes are grouped in sets sharing an overrepresented short motif in their upstream sequence. For each set, the average expression level from a microarray experiment is determined: if this level is significantly higher or lower than the average taken over the whole genome, then the overrepresented motif shared by the genes in the set is likely to play a role in their regulation.

The method was tested by applying it to the genome of *Saccharomyces cerevisiae*, using the publicly available results of a DNA microarray experiment, in which expression levels for virtually all the genes were measured during the diauxic shift from fermentation to respiration.

Several known motifs were correctly identified, and a new candidate regulatory sequence was determined.