基于信息量的调控元件预测方法

谢雪英、孙啸*、谢建明、陆祖宏
东南大学吴健雄实验室

　　设计基于信息含量的调控元件识别算法，对酵母的基因表达数据聚类结果进行分析，旨在预测共表达基因上游非编码区可能存在的转录因子结合位点。分析已知受相同调控因子作用的基因上游序列的结果表明，算法能正确识别具有单一保守核心序列的调控元件和具有间隔子(spacer)的保守序列。通过分析共表达基因，算法提取出的候选调控元件，部分可能具有生物学意义，这还有待于生物学实验的进一步验证。

RECOGNITIONS OF PUTATIVE BINDING SITES BASED ON INFORMATION CONTENT

　　Understanding the mechanism of genes expression and regulation is a necessary and challenging problem for biology genetics. Gene expression is affected by many factors, especially the interactions between regulatory factors and corresponding DNA binding sites. This paper has developed an algorithm for automatic discovery of putative binding sites from yeast genome, which is called information content index method (ICIM)on basis of information theory. ICIM can accurately extract the binding sites from the gene's upstream sequences regulated by known transcriptional factors. It has also recognized the putative binding sites from the upstream regions of two representative gene's clusters based on gene co-expression. Some of those sites have explicit biological functions.

关键词