生物物理学报(2004) 20(3), 225-232

PDF文档

与实验条件相关的基因功能模块聚类分析方法

喻辉[1]、郭政*[1,2]、李霞[1]、屠康[1]
1 哈尔滨医科大学医学数学与生物医学工程学教研室
2 哈尔滨基太生物芯片开发有限责任公司

针对细胞内基因功能模块化的现象，定义了“基因功能模块”和“特征功能模块”两个概念，并基于这两个概念提出一种“与实验条件相关的基因功能模块聚类算法”。该算法综合利用基因功能知识与基因表达谱信息，将基因聚类为与实验条件相关的基因功能模块。向基因表达谱中加入水平逐渐升高的数据噪音，根据基因功能模块对数据噪音的抵抗力，确定最稳定的基因功能模块，即特征功能模块。加噪音实验显示，在基因芯片技术可能发生的噪音范围内，该算法对噪音的稳健性优于层次聚类和模糊C均值聚类。将模块聚类算法应用在NCI60数据集上，发现了8个与实验条件高度相关的特征功能模块。

CONDITION-RELATED GENE FUNCTIONAL MODULE CLUSTER ANALYSIS

Given the phenomenon of gene functional modulization in a cell, the concepts of gene functional module and characteristic functional module were formally defined. In view of these two novel concepts, the condition-related gene functional module clustering algorithm was developed, which clusters genes into conditional-related gene functional modules based on both gene function knowledge and gene expression data. Increasing level of artificial noise was added into the original gene expression dataset and the stability of various functional modules was compared, and those functional modules which were most resistant to the data noise were extracted as the characteristic functional modules. It was demonstrated with the adding-noise experiment that the condition-related gene functional module clustering algorithm was superior to both hierarchical clustering and fuzzy C-means clustering with respect to their stability against data noise that was commonly found in microarray technology. Eight characteristic functional modules were extracted when the algorithm was applied to NCI60 gene expression data.

关键词