生物物理学报(2006) 22(5), 360-366

PDF文档

基于滑动窗口的原核转录起始位点计算定位方法

杜耀华、王正志、倪青山
国防科技大学机电工程与自动化学院

转录起始位点的计算定位是基因转录调控研究的重要内容，但现有方法的识别性能较低。文章作者在已有原核启动子识别算法的基础上，提出了一种基于滑动窗口的原核转录起始位点计算定位方法，通过在合理限定的定位范围内对序列进行滑动扫描，来预测转录起始位点的位置。首先根据窗口序列的交迭组分特征和启动子其它特征分别建立二次判别分类器，用其计算对应位置的似然得分，再利用转录起始位点与翻译起始位点的间隔经验分布信息对似然得分进行修正，最后依照似然得分的分布情况由阈值定位算法确定预测位置。对大肠杆菌真实序列数据的测试结果表明，该定位算法可实现对真实转录起始位点位置的有效预测，与已有算法相比，当敏感性指标同为0.85左右时，特异性指标可从0.20提高至0.65，从而使得定位准确率提高了约20个百分点。

Computational Location of Transcrption Start Sites in Prokaryotic Genome Based on Sliding Window

Although a great deal of effort has been undertaken in the area of transcription start site (TSS) computational location due to its essential role in the research of transcription regulation, the problem has not yet been resolved. According to the previous work on prediction algorithm of prokaryotic promoters, a new computational location method for prokaryotic TSSs based on sliding window was proposed. At first, the authors limited the rational searching ranges in genomic sequences based on the prior information of TSSs occurrence. Then the TSS likelihood scores of each possible position in genomic sequences were calculated by two window classifiers which were trained by quadratic discriminant analysis on overlap content features and other promoter features, respectively. The empirical distribution of distances between TSSs and translation start sites (TLSs) was also utilized to amend the likelihood scores. Final location results were achieved through the procedure of threshold filtration on the likelihood score profiles. The testing results on E. coli datasets showed that the method could find the putative TSSs efficiently. Compared with other current algorithms, the specificity Sp could be improved from 0.20 to 0.65 when the sensitivity Sn was about 0.85, which made the location accuracy increasing by about 20 percents.

关键词

原核基因组(Prokaryotic genome)；转录起始位点(Transcription start site (TSS))；计算定位(Computational location)；滑动窗口(Sliding window)；交迭组分特征(Overlap content features)