人类全基因组范围的CpG岛的预测与分析

庄海滨[1]、朱景德[2]、刘湘军[*1]
1　清华大学
2　上海交通大学肿瘤研究所

　　CpG岛的甲基化是表观遗传中基因表达调控的重要机制。虽然目前已存在几个从DNA序列判别CpG岛的标准，但如何在标准中选择合适的参数仍是研究的焦点。文章通过分析比较两种经典CpG岛判定标准与三种预测方法，提出了改进的CpG岛预测方法———CpGI Seeker。应用该预测方法，结合判定标准中的三个基本参数组合出的13组组合参数，在人类全基因组范围内进行了CpG岛预测，并统计分析了CpG岛的重复序列组成以及相对于基因转录起始位点的位置分布情况。分析结果表明CpGI Seeker具有更精确判定CpG岛的特性；同时还提示，随着判定标准严格性的增加，CpG岛的重复序列含量降低，与基因转录起始位点的相关性提高。将CpG岛最小尺寸为500 bp、GC含量为60％、CpG出现率达到0.65的组合参数作为标准，是目前预测CpG岛的最佳方式。

Prediction and Analysis of CpG Islands in the Human Genome

　　Methylation of CpG Islands is one of the most important mechanisms in the epigenetic regulation of gene expression. Although there exist some criteria for CpG Islands prediction from DNA sequences, the appropriate selection of parameters is still a challenging problem. After comparing two classic criteria of CpG Islands and three prediction algorithms, the authors proposed an improved algorithm, named CpGI Seeker, which showed the better performance than other algorithms. Through using the CpGI Seeker, they did the prediction of CpG Islands in the human genome using 13 combinatorial parameters from 3 basic parameters. Moreover, they analyzed the repeat sequences appeared on the CpG Islands and the location of CpG Islands relative with the transcription initial sites (TISs) of genes. The results demonstrated that more strict criteria of CpG Islands lead to less repeat sequences and more associated with TISs of genes. It is found that the combinatorial parameters of size=500nt, GC%=60% and CpG observed/expected=0.65 is the best criterion for predicting CpG Islands at present.

关键词

CpG岛(CpG Island)；甲基化(Methylation)；表观遗传学(Epigenetics)