

of activity assays showed that the chimeric protein could significantly promote the survival of spinal cord neurons and had a higher neurotrophic activity than ACTH (4-10) and GDNF respectively.

Key words adrenocorticotropin (4-10), glial cell line derived neurotrophic factor, chimeric protein, gene expression

* This work is supported by a grant from National Natural Sciences Foundation of China (30000048).

** Corresponding author. Tel: 86-21-65492132, E-mail: chengzyk@online.sh.cn

Received: December 04, 1999 Accepted: April 04, 2000

小知识

表达序列标签及其应用

王忠华¹⁾ 周美园²⁾ 夏英武^{1, 2)}

(^{1)**} 浙江大学农业与生物技术学院核农所, 杭州 310029; ²⁾ 浙江华川专修学院, 杭州 310020)

宏伟的人类基因组计划自 20 世纪 90 年代初正式启动以来, DNA 序列数据库一直处于指数级水平增长之中, GenBank 的容量每隔 14 个月就翻一番. 如何充分利用这些已有的数据资源, 加速基因克隆研究, 已成为一个急迫而富有挑战性的课题摆在我们面前. 越来越多的基因克隆建立在生物信息学分析的基础之上, 进行同源性比较与计算机辅助结构和功能分析, 尤其是 1995 年以来, 表达序列标签 (简称 EST) 的广泛应用, 为大规模进行基因克隆和表达分析提供了强大的动力, 也为生物信息学功能的充分发挥提供了广阔的空间.

通过从 cDNA 文库中随机挑取的克隆进行测序所获得的部分 cDNA 的 3' 或 5' 端序列称为表达序列标签 (EST), 一般长度为 300~500 bp 左右. 利用 EST 作为标记所构建的分子遗传图谱称为转录图谱. EST 技术最早由 Sikela 和 Matsubara 提出, 并被 Venter 大规模应用. EST 测序的飞速发展, 使得 dbEST 中的记录已经超出一百多万条. 这种方法不仅可以判断一段 DNA 是否包含开放阅读框架 (ORF), 而且能精确地给出该基因的内含子和外显子的剪切模式. 植物 EST 计划主要集中在拟南芥和水稻上, 其他植物的 EST 相对较少. 1998 年 12 月美国国家生物技术信息中心 (NCBI) 公布的各种植物的 EST 数目总和已达几万条. 这些 EST 不仅为植物基因组遗传图谱的制作提供了大量的分子标记, 也为基因的功能研究提供了有价值的信息. 此外, EST 计划还为基因的鉴定提供了候选基因.

在各类数据中, 目前应用最多的仍为 EST, 其应用范围主要包括以下几个方面: a. 不同剪接型的识别; b. 构建重叠群 Contigs, 加速全长基因的克隆; c. 基因结构分析 (内含子/外显子识别); d. 基因定位分析; e. 基因表达分析. 限于篇幅, 本文从生物信息学的角度来简要阐述 EST 的应用.

生物信息学 (bioinformatics) 是用数理和信息科学的观点、理论和方法来研究生命现象, 组织和分析呈指数增长的生物学数据的一门学科. 生物信息学由数据库、计算机网络和应用软件三大部分组成.

EST 信息分析的基本思路: 基因编码区序列 (简称 CDS) 与 EST 数据比较 → 寻找感兴趣 ESTs (标准: 长度 \geq 100 bp, 同源性介于 50% ~ 85% 之间) → 所选 ESTs 与 GenEmble 数据库比较 → 找出未克隆 ESTs → 再与 dbEST、dsSTS、dbHTGs、MGD 及 UniGene 数据库比较搜寻重叠群 Contigs → 设计引物进行 PCR 扩增或筛选 cDNA 文库或索取 cDNA 克隆号进行电子拼接获取全长 cDNA → 基因定位、结构、功能检测分析等.

随着“后基因组”时代的到来, 生物信息学将在基因功能研究中发挥愈来愈重要的作用. 而 EST 分析是生物信息学分析的核心, 它为新基因的克隆和功能分析提供了出发点. 通过 EST 的综合分析, 克隆、鉴定和定位新基因, 对于我们获取和克隆候选基因有重要参考价值.