

# EST 技术在寄生虫基因组研究中的应用

司 进 朱荫昌

无锡市血吸虫病防治研究所 无锡 214064

表达序列标记(expressed sequence tag, EST)技术是 1991 年由 Adams 等<sup>[1]</sup>提出的,用于寻找人类新基因,绘制人类基因组图谱,识别基因组序列编码区的一种技术。所谓 EST,实际上就是对应于 mRNA 的 cDNA 上的一条序列片段,这些短序列代表着染色体上独一无二的识别位点,所以它对应的序列标记位点(sequence-tagged site, STS)已成为绘制基因组物理图谱的标准标记(standard mark),长度超过 150 bp 的 EST 是一种最有用的遗传标记,并且适宜于相似基因的寻找。在美国分子生物学家经过两年的辩论决定于 1988 年开始对人类基因组进行核苷酸序列研究之后,cDNA 序列分析作为人类基因组计划的一部分,人们对其价值的看法却不尽相同。有人认为,基因组的编码区序列包含着基因组的绝大部分信息,却仅占 DNA 总量的 3% 左右,因此 cDNA 测序应优于基因组测序<sup>[2]</sup>。但有人提出<sup>[3]</sup>,要找出表达于所有组织,各种类型细胞及其不同发育阶段的所有的 mRNA 太困难,并且在 cDNA 测序时容易丢失来自内含子(intron)和基因间序列的信息,他们认为,基因编码区即 mRNA 序列很容易从基因组序列中推测出来,因而没有必要进行大规模的 cDNA 序列分析,事实上,基因组序列转录产物(transcripts)的推测<sup>[4]</sup>只有在相对较大的外显子(exon)中才有可能。无论如何,EST 技术从产生到现在短短数年,发现的新基因是已知基因的数倍。

寄生虫感染一直严重困扰着全世界,在全球范围流行的传染病中,寄生虫感染率可能超过其它任何病原所致的感染,估计占世界三分之一的人口有肠道蠕虫(包括鞭虫、蛔虫、旋毛虫等)的感染,有 2~3 亿人感染三种血吸虫之一,有 1.5 亿人以上感染丝虫,所以寄生虫感染的预防已是一个迫在眉睫的任务。然而,迄今为止还没有一种理想而有效的疫苗可用于免疫预防并阻断寄生虫病的传播,造成这种情况的原因是多方面的,最根本的原因还是我们对寄生虫的了解不够。据悉,目前我们所了解的寄生虫基因仅占寄生虫全部基因的 2%~3% 左右,在这

么少的基因基础上进行蛋白质或核酸疫苗的研制显然不够,EST 技术为我们从事寄生虫研究提出了一个全新的视角,为寄生虫新基因和未知基因的寻找提供了一个高效的工具。

## EST 技术

自从 70 年代中叶首例 cDNA 克隆问世以来,已发展了许多提高双链 cDNA 克隆合成效率的方法,并大大改进了载体系统,将 [poly(A)<sup>+</sup>mRNA] 在逆转录酶的作用下合成 cDNA,进而将此 DNA 导入原核载体这一流程已成为真核分子生物学的基本手段,而 EST 技术正是基于 cDNA 文库基础上的 cDNA 克隆的部分序列测定,因此 EST 技术有如下步骤:构建 cDNA 文库,随机挑选 cDNA 克隆,碱裂解法或 PCR 扩增制备模板,序列分析,与同种和异种生物已知的核酸和蛋白数据库进行比较分析,新基因及未知基因的基因库(GenBank)登录。真核生物基因组相当庞大,它的最大特点是含有大量的重复序列,而且功能 DNA 片段多被不编码蛋白质的非功能 DNA 片段所隔开,这就是“C 值反常规现象(C-value paradox)”。按重复频率高低,重复序列分为高度重复序列、中度重复序列和低度重复序列及单拷贝序列。为了增加获得新基因的可能性,必须提高 cDNA 文库的机率。可在起始 mRNA 或由它合成双链 cDNA 时进行富集,富集可使 cDNA 文库缩小并可降低获得新基因的人力物力。尽管如此,cDNA 文库还是存在大量的高丰度 cDNA 克隆,根据 Adams 等<sup>[1]</sup>的“试验(pilot project)”,来自人脑海马体 cDNA 文库的 30% 以上的克隆由 rRNA、线粒体 cDNA 和全部由多聚腺苷酸组成的插入序列。因此,一个理想的 cDNA 文库必须去除或尽量消除这些无信息克隆的影响,这就涉及到 cDNA 文库的前加工技术,例如:消减杂交<sup>[5]</sup>(subtractive hybridization),均等化<sup>[6]</sup>(normalization)。通过文库前加工,一方面可以减少高丰度 cDNA 克隆的数量,另一方面可以使文库中所有的基因序列所占的比例大致相等。理想的 cDNA 文库

的另一特征是定向克隆(directional cloning), 这样可选择性的得到编码序列或是 3' 端非编码序列。Hoog<sup>[7]</sup>采用差异 cDNA 文库筛选策略, 使获得新基因的概率从随机挑选的 63% 上升至 84%。

为了能快速分析大量的 cDNA 克隆, 减少准备测序模板的时间, 可采用 PCR 扩增与线性 PCR 序列分析相结合的方法, 同时用琼脂糖凝胶电泳进行 PCR 扩增产物分析, 例如无扩增产物出现, 可以排除那些插入序列很短、无插入序列甚至无扩增产物的克隆, 这样就大大优化了测序效果。获得 EST 序列数据后, 首先与 GenBank 核酸数据库进行相似性检测, 假如没有精确匹配基因, 将 EST 序列数据按正反 6 种阅读框翻译成蛋白质, 按着与蛋白质序列数据库进行比较分析, 基因分析的结果大致有 3 种: 第一是已知基因, 是研究对象为人类已鉴定和了解的基因; 第二是以前未经鉴定的新基因; 第三是未知基因, 这部分基因无同种或异种基因的匹配。新基因和未知基因将进一步用于生物学研究。

## 在寄生虫方面的应用

### 1 在血吸虫研究中的进展

在血吸虫研究方面, 目前 EST 技术主要对曼氏血吸虫 (*Schistosoma mansoni*, Sm) 进行了研究。Franco 等<sup>[8,9]</sup>用了不到两年的时间分离到大量的新基因, 超过已知的 Sm 基因的两倍, 这些新的 Sm 基因广泛分布于细胞浆结构蛋白和调节蛋白、酶、膜核和分泌蛋白。同时 Franco 等<sup>[10]</sup>还用 EST 技术对曼氏血吸虫不同发育阶段基因表达的全貌进行了比较研究, 共获得 1 401 个 EST。7 个不同的 cDNA 文库来自生活史的 4 个不同阶段, 它们中无效克隆少于 20%, 新基因却多于 50%, 对每个库的富有基因研究表明, 成虫 cDNA 文库含有少量的高频基因。比较一下不同文库的 EST 可以发现: 大多数的基因仅在某一文库出现, 但也有一些基因在一个以上的不同发育阶段表达, 这些少数基因或许代表 Sm 的管家基因(housekeeping gene)。将那些在一种以上的文库中仅出现一次的基因规定为独特基因(unique gene), 在这些基因中, 大约 20% 与来自其它生物体的基因有同源性, 8% 与 Sm 已知基因相同, 有 70% 属未知基因, 在这些基因中或许有 Sm 种属特异性基因。另外, 通过 EST 测序, Franco 等<sup>[11]</sup>还从成虫 cDNA 文库中获得编码胸基本保守蛋白(breast basic conserved protein)/核糖体 L13 的基因, 并通过与 Sm 有序粘粒文库杂交将该基因定位

于 3 号和 W 染色体上。

### 2 在锥虫研究中的进展

以前对非洲锥虫的分子遗传学研究仅集中在少数的基因及基因产物上, 而且大多数基因只涉及其表面抗原。为了获得新基因并分析它们的表达情况, 产生表达序列标记绘制锥虫基因组物理图谱, Djikeng 等<sup>[12]</sup>对从罗德西亚布氏锥虫文库中随机挑选的 cDNA 克隆进行了部分序列分析, 2 128 个 EST 的分析表明: 它与来自人、啮齿动物、动基体目原生动物、酵母和植物基因表达的蛋白质有显著的相似性。这些 EST 编码的同源蛋白具有各种各样的功能, 例如: 信号接收、转导、促细胞分离、基因调节、DNA 修复和复制、维护结构完整性等等。36% 的 EST 与非洲锥虫以外的生物基因有显著的同源性, 仅 15% 为锥虫已知基因, 52% 的 EST 与已知生物基因无显著同源性。Mathieu-Daude 等<sup>[13]</sup>采用基于随机引物 PCR (random arbitrary primed polymerase chain reaction, RAP-PCR) 的 RNA 指印技术对表达于布氏锥虫生活史过程中的基因进行鉴定, 标准的 RAP-PCR 分两步: 用根据 5' 端最小外元设计的引物进行 PCR 扩增, 可确保大多数扩增产物含有 mRNA 的 5' 末端; 不同的产物被再次扩增, 并通过单链结构多态性(single strand conformation polymorphism) 凝胶电泳分离, 然后进行克隆、测序, 共获得 32 个 EST, 通过逆转录 PCR (RT-PCR) 证实了 24 个 EST 表达于锥虫生活史的不同阶段, 这些转录产物包括编码细胞表面蛋白、代谢酶类、热休克蛋白的基因。通过随机挑选 cDNA 克隆和部分序列分析, El-Saved 等<sup>[14]</sup>从罗德西亚布氏锥虫 cDNA 文库中获得 518 个 EST, 其中 205 个 EST 在已知数据库中找到匹配基因, 主要编码代谢酶、信号转导蛋白、转录因子、核糖体蛋白、组蛋白、增殖相关蛋白等, 313 个 EST 无任何匹配, 这些来自 cDNA 的 EST 为寻找新的锥虫特异基因提供了一条新途径。

### 3 在弓形虫研究中的进展

为了加速刚地弓形虫的基因发现和遗传图谱绘制, Ajoka 等<sup>[15]</sup>从弓形虫 cDNA 文库中随机挑选的 cDNA 克隆的 5' 末端获得 7 000 多个 EST, 通过与数据库比较, 其中约有 500 多个新的刚地弓形虫基因, 其功能主要有转录、翻译、蛋白质分泌、信号转导、代谢等, 50% 以上的 EST 其对应的基因功能不明。刚地弓形虫是一个有着复杂生活史的顶复合门寄生虫, 其生活史包括速殖子这一快速分裂期, 据报

道速殖子表面有 5 个主要抗原,含量最大的是表面抗原 1 (surface antigen 1, SA G1),其它 4 个中至少有 SA G3 和新近描述的小抗原 (SA G1-related sequence 1, SRS1),它们与 SA G1 有结构上的相关性,为了证实是否有更多的 SA G1 同系物存在, Manger 等<sup>[16]</sup>对 EST 数据库进行了搜索,发现在大量的 EST 相应基因中至少有 3 个新基因与 SA G1 相关,象 SA G1 一样,这些新的 SRS 基因编码含有若干基序和一系列半胱氨酸残基的糖基化 1-磷脂酰肌醇磷酸三酯酶锚定蛋白,免疫荧光试验证实这些 SRS 表达于弓形虫速殖子的表面,目前已搞清楚, SA G1 家族有 10 个成员,其中包括 5 个主要 SA G 中的 3 个,以及仅表达于缓殖子的一个抗原,该家族的功能是提供大量与宿主细胞相互作用的受体系统和/或通过免疫反应限制刚地弓形虫感染。除了速殖子之外,刚地弓形虫还有缓殖子这一缓慢分裂的无性形式,为了进一步研究这两种无性形式的差异, Manger 等<sup>[17,18]</sup>从速殖子和缓殖子文库中获得了大量的 EST,与以前从速殖子中得到的大约 7 400 个 EST 结合起来分析,发现了一个缓殖子特异的烯醇化酶,考虑到以前发现的缓殖子乳酸脱氢酶,提示缓殖子期糖酵解流量的显著差异,丰富了对弓形虫生长和代谢的认识。

#### 4 在利什曼原虫研究中的进展

目前,硕大利什曼原虫已有约 500 个 EST 序列进入数据库<sup>[19]</sup>,均是从含有引导序列的全长 cDNA 的 5' 端测出的序列,对利什曼原虫的目标是测出至少 1 500 个新序列,包括寄生于脊椎动物中的鞭毛体构建的含引导序列的 cDNA 库的 EST 序列。Patrick Bastien 等在法国已利用 EST 技术绘制利什曼虫种核型图谱。Wincker 等<sup>[19]</sup>用 66 个 EST、41 个已知基因和 137 个未知 DNA 序列做成特异性探针与婴儿利什曼原虫梯度分离的染色体杂交,进行染色体物理图谱的研究。结果表明:利什曼原虫含有 36 条染色体,大小 0.35~3 mb,将婴儿利什曼原虫与硕大利什曼原虫、热带利什曼原虫及埃塞俄比亚利什曼原虫进行基因组结构比较,发现在所有利什曼原虫中都有连锁群。提示:利什曼原虫结构上的保守性也许是致病的关键。目前研究中获得的高密度标记将加速利什曼原虫详细物理图谱的构建,从而促进利什曼原虫的遗传学研究。

## 结 语

感染人类寄生虫的基因组研究计划主要针对 7

种主要的人体寄生虫病:疟疾、利什曼病、非洲锥虫病、美洲锥虫病、弓形虫病、血吸虫病和丝虫病。由 WHO 资助, Martin A Slett (欧洲生物信息研究所,英国) 建立了一个寄生虫主页— <http://www.ebi.ac.uk/parasites-genome.html><sup>[20]</sup>, 该网络与 7 个基因组相连, 主要涉及丝虫基因组数据库, 目前收录的马来丝虫表达序列标签已超过 4 000 个, 此外有 300 多个旋盘尾丝虫的 EST 亦被收录到该数据库中。美国佛罗里达大学生物技术研究中心的 John Dame 则专门为恶性疟原虫建立了一个基因序列标签数据库站址— <http://Parasite.arl.nfl.edu/malaria.html> WHO、UNDP (联合国开发计划署) 和世界银行共同创建了一个血吸虫基因组网络 (*Schistosoma* Genome Network, SGN)<sup>[21]</sup>, 该网络收录的血吸虫 EST 约 3 000 个。此外, 利什曼原虫和格氏锥虫也各自建立了基因组计划数据库。

EST 测序所获得的基因中约 70% 在数据库中找不到有匹配关系的序列, 但随着更多的其它生物序列的获得, 这种情况将有所改变, 或许可能这些基因是更好的化疗目标和候选疫苗分子, 目前有人开始探索一种新方法, 先将这些 DNA 分子作为疫苗而不必先测定其蛋白结构或功能, 如果它可以作核酸疫苗使用, 再进一步测序和功能分析。

可以预见, 随着 EST 技术的进一步完善, EST 测序的不断进展, 更多的寄生虫新基因的鉴定, 将丰富人类对寄生虫的认识, 促进寄生虫疫苗的研制, 为阻断寄生虫感染的途径并最终控制和消灭寄生虫病打下坚实的基础。

## 参 考 文 献

- 1 Adams MD, Kelly JM, Gocayne JD, et al Complementary DNA sequencing: expressed sequence tags and human genome project Science 1991; 252: 1631~ 56
- 2 Brenner S. The human genome: the nature of enterprise Ciba Found Symp 1990; 149: 6
- 3 Baltimore D. Report of the Committee on Mapping and Sequencing the Human Genome. National Academy Press, Washington DC, 1988
- 4 Fickett J. Recognition of protein coding regions in DNA sequences Nucleic Acids Res 1982; 10: 5303
- 5 Patanjali SR, Parimoo S, Weissman SM, et al Construction of a uniform-abundance (normalized) cDNA library. Proc Natl Acad Sci 1991; 88: 1943~ 1947
- 6 Fargnoli J, Holbrook NJ, Fomace Jr, et al Low-ratio hybridization subtraction Anal Biochem 1990; 187: 364~ 373
- 7 Hoog C. Isolation of a large number of novel mammalian genes by a differential cDNA library screening strategy. Nucleic Acids Res 1991; 19: 6123~ 6127
- 8 Franco GR, Simpson AJ, Pena SD, et al Sequencing and identification of expressed *Schistosoma mansoni* genes by random selection of cDNA clones from a directional library. Mem Inst Oswaldo Gruz 1995; 90: 215~ 216

- 9 Franco GR, Adams MD, Soares MB, et al Identification of new *Schistosoma mansoni* genes by EST strategy using a directional cDNA library. *Gene* 1995; 152: 141~147
- 10 Franco GR, Rabelo EM, Azevedo V, et al Evaluation of cDNA libraries from different developmental stages of *Schistosoma mansoni* for production of expressed sequence tags (ESTs). *DNA Res* 1997; 4: 231~240
- 11 Franco GR, Tanaka M, Simpson AJ, et al Characterization of a *Schistosoma mansoni* homologue of the gene encoding of the gene encoding the breast basic conserved protein 1/L13 ribosomal protein. *Comp Biochem Physiol* 1998; 120: 701~708
- 12 Djikeng A, Agnif C, Ponelson JE, et al Generation of expressed sequence tags as physical landmarks in the genome of *Trypanosoma brucei*. *Gene* 1998; 221: 93~106
- 13 Mathieu-Daude F, Weiss J, Davis C, et al Differentially expressed genes in the *Trypanosoma brucei* life cycle identified by RNA fingerprinting. *Mol Biochem Parasitol* 1998; 92: 15~28
- 14 El-Sayed NM, Alaroon CM, Beck JC, et al cDNA expressed sequence tags of *Trypanosoma brucei rhodesiense* provide new insights into the biology of the parasites. *Mol Biochem Parasitol* 1995; 73: 75~90
- 15 Ajioka JW, Boothroyd JC, Brunk BD, et al Gene discovery by EST sequencing in *Toxoplasma gondii* reveals sequences restricted to the Apicomplexa. *Genome Res* 1998; 8: 18~28
- 16 Manger D, Hehl AB, Boothroyd JC, et al The surface of *Toxoplasma* tachyzoites is dominated by a family of glycosylphosphatidylinositol-anchored antigens related to SAG1. *Infect Immun* 1998; 66: 2237~2244
- 17 Manger D, Hehl AB, Pamley S, et al Expressed sequence tag analysis of the bradyzoite stage of *Toxoplasma gondii*: identification of developmentally regulated genes. *Infect Immun* 1998; 66: 1632~1637
- 18 Blackwell JM. Parasite genome analysis progress in the *Leishmania* genome project. *Trans R Soc Trop Hyg* 1997; 91: 107~110
- 19 Wincker P, Ravel C, Blaineau C, et al The *Leishmania* genome comprises 6 chromosome conserved across widely divergent human pathogenic species. *Nucleic Acids Res* 1996; 24: 1688~1694
- 20 Tavame J. The human genome: the nature of enterprise. *Parasitol Today* 1996; 12: 463~464
- 21 Johnston DA. Recognition of protein coding regions in DNA sequences. *Parasitol Today* 1997; 13: 45~46

1999年8月10日收稿 1999年9月30日修回

(编辑: 庄兆农)

## 以牛带绦虫六钩蚴静脉注射小牛后检出牛囊尾蚴

新疆医科大学寄生虫学教研室 乌鲁木齐 830054 康金凤 吾拉木·马木提  
旭川医科大学寄生虫学教研室 日本 伊藤亮

为适应寄生虫学教学及科研的需要,我们曾给小牛经口灌喂大量的牛带绦虫虫卵,甚至整节妊娠节片,结果均未感染成功。今以虫卵孵化出的六钩蚴,经静脉途径注射感染小牛获得成功。

### 材料与方法

- 1 感染动物 断奶1个月的雌性小牛1只。
- 2 牛带绦虫虫卵 采自用槟榔、南瓜籽、硫酸镁驱出的活虫体,用自来水洗净,取末端妊娠节片3~5节,剪碎,使虫卵散出,用无菌生理盐水漂洗,直至除净虫体碎片,留取纯净虫卵备用。
- 3 孵化液 蒸馏水配制,内含1.5%胰酶,1%NaHCO<sub>3</sub>,5%羊胆汁。
- 4 虫卵孵化 将上述纯净的虫卵置3ml孵化液中,37℃水浴振荡,15min后吸取混匀的该孵化液镜下观察,待全部虫卵胚膜破裂,六钩蚴孵出后,即加入10倍~20倍的无菌生理盐水,混匀,室温自然沉淀。然后弃去上清液,留取白色的沉淀物(即聚集的六钩蚴),再加无菌生理盐水至原量,并混匀后镜下计数(1000个六钩蚴/ml,共30ml),待用。
- 5 牛体感染 用9号输血头皮针头,经颈静脉将上述孵化的六钩蚴缓慢注入牛体。注射过程中注意活动针管,使六钩蚴保持悬浮状态,避免因六钩蚴聚集而形成栓塞。

### 结果

感染4个月后将小牛宰杀,进行活检。发现牛体全身肌

肉均有牛囊尾蚴寄生,约有上万个,以心肌、膈肌、咀嚼肌及舌肌等处较多。虫体为典型的透明/半透明囊状物,卵圆形,长径2mm~4mm,个别可达6mm。囊外有宿主纤维组织形成的外囊,囊内充满清晰囊液,并可见卷缩在内的白色头节。压片后镜下观察到头节的4个吸盘发育完好,头节部位布满石灰小体。

### 讨论

此法简单、易行、成功率高,且可控性强,能满足有关的科研及教学需要。补充说明的是:选取小牛,是因它对寄生虫易感,而小牛的雌雄性别应不受限制。活的虫体遇刺激后即收缩,子宫主干不易找到,为获取较多虫卵,剪碎虫体,使子宫中虫卵散出;死亡虫体节片松弛,可顺着子宫主干剪开,在生理盐水中荡洗,待虫卵散出后挑出节片,即可获取纯虫卵。为减少孵化液中的胆汁及酶类等进入血液引起不良反应,所用孵化液的量,以能将虫卵彻底孵化的最小量为适宜,并且用10倍~20倍的生理盐水对孵化出的六钩蚴清洗1次,清洗时采用自然沉淀法,以免对六钩蚴造成伤害。所有与注入牛体有关的操作环节,应做到无菌,以避免接种注射引起细菌感染。未孵化的虫卵不应注入血循环,因其直径31μm~43μm,有可能引起微细血管堵塞,影响虫体播散。

1999年4月8日收稿 1999年9月27日修回

(编辑: 任燕芬)