

cDNA 捕捉法

——从大的基因组区域直接分离编码序列

杨新平^{1,2} 于常海² 贺林^{1,2}

(1 中国科学院上海生理研究所 上海 200031; 2 中国科学院上海生命科学研究中心 上海 200031)

摘要 cDNA 捕捉法或 cDNA 直选法是一种以表达为基础的基因分离技术, 直接利用目的区域的基因组 DNA 捕捉 cDNA, 快速从大的基因组区域分离表达序列。该法已成功地应用于定位克隆和详尽的转录图谱的构建。

关键词 cDNA 捕捉法, cDNA 直选法, 定位克隆, 转录图谱

分类号 Q75 Q343.1

cDNA capture

—— Isolation of coding sequences from large genomic region

YANG Xin-Ping^{1,2} Albert C H Yu² HE Lin^{1,2}

(1 Shanghai Institute of Physiology, Chinese Academy of Sciences, Shanghai 200031;

2 Shanghai Research Center of Life Sciences, Chinese Academy of Sciences, Shanghai 200031)

Abstract cDNAs capture (or direct cDNAs selection) is an expression-based gene isolation technique which uses the genomic DNA to capture the cDNAs directly and thus rapidly identifies expressed sequences from large genomic region. This method thus has been successfully applied to perform positional cloning and to construct detailed transcription maps.

Key words cDNA capture, direct cDNA selection, positional cloning, transcription map

在定位克隆 (positional cloning) 和制作表达图谱 (expression map) 中, 从大的基因组区域分离编码序列 (coding sequence) 或表达序列标签 (expressed sequence tags) 是一项十分艰巨的工作, 因为在人类基因组 DNA 中只有很小部分编码 mRNA, 而大部分则为内含子、基因间序列和重复序列。目前通过大规模基因组测序来发现新基因还不可行, 即使开发出的软件能 100% 地找出基因组序列中的新基因, 也还要通过分离 cDNA 来证实。因此, 分离表达序列和制作表达图仍是人类基因组计划中的必要工作。虽然有一些分离表达序列的方法^[1-9], 但 these 方法都有这样或那样的缺点。

cDNA 捕捉法 (cDNA capture) 或叫 cDNA 直选法 (direct cDNA selection) 是一种直接用基因组 DNA 如 YACs 捕捉 cDNA 片段的方法^[10]。这种方法已成功地用于许多定位克隆^[11-16]。该项

技术包括: (1) 用 PCR 扩增来自不同组织的混合 cDNA 池 (complex cDNA pools); (2) cDNA 片段与基因组目标 DNA (genomic target) 杂交, 选择出与基因组目标 DNA 杂交的 cDNA 片段并进行 PCR 扩增; (3) 进行第二轮选择并将选出的 cDNA 片段克隆到载体上 (图 1)。

cDNA 直选法特别适合于大的基因组区域内表达序列的分离, 如在 1Mb 的 MHC-I 区域分离了一些新基因^[15,16], 并成功地在 >2Mb 的基因组区域构建详细的表达图谱^[17-19]。为了在基因组目标区域内找出尽量多的转录单位, 可将多种组织来源的 cDNA 混合, 进行一次直选。这同时也解决了基因在不同组织中表达的特异性问题^[20]。在对疾病基因进行定位克隆时, 如果对疾病发病机理有所了解, 可选择与疾病发生相关的组织 cDNA 文库来进

收稿日期: 1998-11-25

行直选，这样可减少需分析的基因数目。

cDNA 直选法操作原理图解如图 1 所示。

1. cDNA 直选法的过程

该法操作过程有如下步骤：

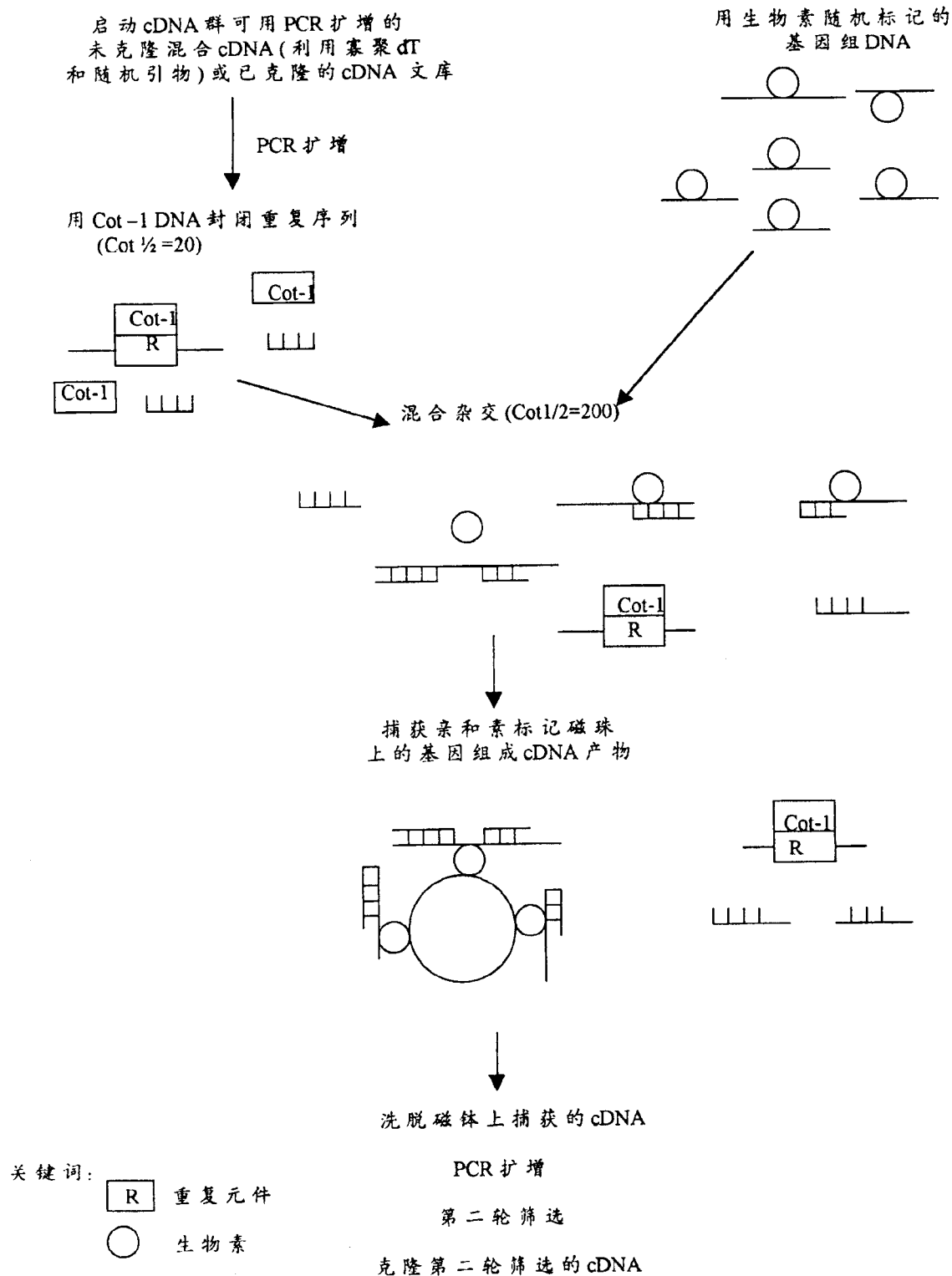


图 1 直选法的流程图

(1) 用生物素 (biotin) 标记基因组目标 DNA。

(3) 用 Cot1 DNA 或 gDNA 与扩增后的 cDNA

(2) 将不同组织来源的 cDNA 混合，用随机引物和 oligo dT 引物 PCR 扩增 cDNA 池。

池杂交以封闭 cDNA 重复序列，重复序列的封闭在 Cot 1/2=20 条件下进行。

(4) 封闭了重复序列的 cDNA 与生物素标记的目标 DNA 混合, 在 $Cot\ 1/2=100\sim 200$ 条件下进行杂交。

(5) 用表面有 streptavidin 的磁珠吸附生物素标记的目标 DNA, 从而将与目标 DNA 杂交的 cDNA 保留下来。

(6) 洗去非特异结合的 cDNA 后, 再将被目标 DNA 捕捉的 cDNA 洗脱下来。

(7) 用 PCR 扩增第一轮选择所得的 cDNA, 并用同样方法进行第二轮选择。

(8) 将第二轮选择得到的 cDNA 克隆到载体上并进行分析。

2 影响 cDNA 直选的基本因素

有些重要因素是直选法成功的关键, 了解这些因素后我们就知道在 cDNA 直选过程中哪些是关键步骤。影响 cDNA 直选法的基本因素有:

(1) 高质量 cDNA 来源: 覆盖 cDNA 全长各部分, 低 rRNA 和质粒 DNA 含量等。由于商业化的 cDNA 文库质量差别较大, 采用自己制备未克隆到载体的 cDNA 较好。为了保证含 cDNA 全长各部的 cDNA 片段, 将随机引物构建的 cDNAs 与 Oligo dT 构建的 cDNAs 混合。

(2) 基因组目标 DNA 质量也很重要。因为整个技术是建立在捕获生物素标记的目标 DNA 以及与之杂交的 cDNAs, 所以生物素能否很好地整合在目标 DNA 中是该技术的关键之一。生物素标记可用如下方法监测: 生物素对目标 DNA 进行 nick 标记的同时加入一种同位素标记的 dNTP, 将 nick 标记产物与磁珠混合 10~15 分钟, 测磁珠与溶液的放射比。若磁珠与溶液的放射比大于 8:1, 说明生物素整合率高; 如果磁珠与溶液的放射比小于 8:1, 则需进一步纯化 DNA。磁珠与溶液的放射比还受磁珠的结合容量的影响, 将标记产物与不同量的磁珠结合以确定磁珠的结合容量。

(3) cDNAs 和目标 DNA 的相对量也很重要。为了保证低丰度的 cDNA 不丢失, 在第一轮选择时, 目标 DNA 要大大过量, 这一轮选择使目的 cDNA 富集了大约 10^3 倍。第二轮选择时, 则限制目标 DNA 浓度, 目的 cDNA 又富集了约 10 倍。

(4) 如用 YACs 来选择 cDNA, 那么必须将 YACs 从酵母基因组中分离出来。实验表明, 用酵母基因组 DNA (含 YAC DNA) 选择 cDNA, 目的

cDNA 只富集了 10^2 倍^[21], 而用纯化选择 cDNA, 则目的 cDNA 富集 $10^4\sim 10^5$ 倍^[17,20]。

(5) 用报告 cDNA (reporter cDNA) 作为阳性对照来监测二轮选择时 cDNA 的富集情况。

3 第二轮选择所得 cDNAs 的分析

每 100kb 的目标 DNA 需随机挑 50 个噬菌斑, 分别悬浮于 100 μ l 的 SM 中, 取 1 μ l 噬菌体进行 PCR。取 PCR 产物 1 μ l 点在尼膜上, 做成 96 点阵的膜, 分别用报告 cDNA、高拷贝重复序列、质粒 DNA、rDNA 和已知基因的 DNA 探针杂交, 除去背景克隆和已知基因克隆。

除掉冗余克隆后, 剩下的克隆可用于杂交和测序。这些克隆与目标 DNA 的重叠群 (通常为某区域的 YAC 或 BAC 重叠群) 杂交以确定它们确实来源于目标基因组 DNA。一般来说, >70% 的 cDNAs 可杂交回去。

用覆盖人 MHC-I 区域约 1Mb 的 YACs (A231G12, B30H3 和 190C8) 对几个 cDNA 文库进行直选, 通过对第二轮选择所得 cDNA 进行分析, 发现获得至少 31 个外 HLA 基因。这些基因在 MHC-I 区域作图得到详细的表达图^[16]。如果把 HLA 基因和假基因都计算在内的话, 那么基因密度为约 20kb 一个基因, 这与 MHC-II、MHC-III 区域的基因密度相似。这说明应用 cDNA 直选法基本上能找出基因组目标区域的所有基因。

参 考 文 献

- 1 Monaco A P, Neve R L, Colletti-Feener C, et al. *Nature*, 1986; **323**:646-650
- 2 Romens J M, Ianuzzi M C, Kerem B S, et al. *Science*, 1989; **245**: 1059-1065
- 3 Bird A P. *Nature*, 1986; **321**:209-213
- 4 Hanson I M, Poustka A, Trowsdale J. *Genomics*, 1991; **10**: 417-424
- 5 Elvin P, Slynn G, Black D, et al. *Nucleic Acids Res*, 1990; **18**: 3913-3917
- 6 Dyuk G M, Kim S, Myers R M, et al. *Proc Natl Acad Sci USA*, 1990; **87**: 8995-8999
- 7 Buckler A F, Chang D D, Graw S L, et al. *Proc Natl Acad Sci USA*, 1991; **88**: 4005-4009
- 8 Corbo L, Maley J A, Nelson D L, et al. *Science*, 1990; **249**: 652-655
- 9 Liu P, Legerski R, Siciliano M J. *Science*, 1989; **246**: 813-815
- 10 Parimoo S, Patanjini S R, Shukla H, et al. *Proc Natl Acad Sci USA*, 1991; **88**: 9623-9627
- 11-21 略