

二维凝胶电泳后,通过银染或放射自显影等显色,然后采用激光密度扫描(laser densitometer)、磷显影仪(phosphor-imager)等扫描装置,确立。

(2) 二维凝胶参照图谱(2D-gel referenced maps)

在各种蛋白的分子量和等电点,电子计算机分析构建二维凝胶参照图谱,最后通过肽测序、肽图(peptid mapping)、免疫和 cDNA 短暂超表达等研究,最后确立蛋白质结构,并与蛋白数据库及 DNA 数据库的信息进一步比较研究,从而可以为特定的疾病状态、遗传病、感染性疾病、以及不同的发育阶段,特别是宿主的肿瘤等的发生发展提供生物信息,并为相应疾病治疗提供依据。

(李孝主)

6.5 cDNA 捕捉法

6.5.1 概论

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 EST 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。在目前通过大规模基因组测序来发现新基因还未能实现前,即便开发出软件能 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(Monaco AP *et al.*, 1986; Romens JM *et al.*, 1989; Bird AP 1986; Hanson IM *et al.*, 1991; Elvin P *et al.*, 1990; Dyuk GM *et al.*, 1990; Buckler AF *et al.*, 1991; Corbo L *et al.*, 1990; Liu P *et al.*, 1989),但这些方法都有这样或那样缺点。

cDNA 捕捉法(cDNA capture)或叫 cDNA 直选法(direct cDNA selection)是一种直接用 cDNA 片段捕捉基因组 DNA 如 YACs 中所含人类基因的方法(Parimoo S *et al.*, 1991)。这种方法已成功地用于许多定位克隆(Vorechosky I *et al.*, 1993; The Huntington Disease collaborative Research Group, 1993; Hastbacka J *et al.*, 1994; Savitsky K *et al.*, 1995; Fann WF *et al.*, 1993; Wei H *et al.*, 1993)。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);②cDNA 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,再进行第二轮选择并将选出的 cDNA 片段克隆到载体上(图 6.11)。

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-III 区域分离了一些新基因(Fann WF *et al.*, 1993; Wei H *et al.*, 1993),并成功地在 >2Mb 的基因组区域构建详细的表达图谱(Simmons AD *et al.*, 1995; Simmons AD *et al.*, 1997; Di Mastro RG and Lovett, 1997)。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殊问题(Morgan JG *et al.*, 1992)。在对疾病基因进行定位克隆时,如果对疾病发病机理有

了解,可选择与疾病发生相关的组织 cDNA 文库来进行直选,这样可减少需分析基因数目。

6.5.2 cDNA 直选法的过程

图 6.11 是 cDNA 直选法操作原理图解,包括下列步骤:

- 用生物素(biotin)标记基因组目标 DNA。
- 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。
- 用 Cot1 DNA 或 gDNA 与扩增后的 cDNA 池杂交以封闭 cDNA 重复序列,重复序列被封闭在 $Cot1/2 = 20$ 条件下进行。

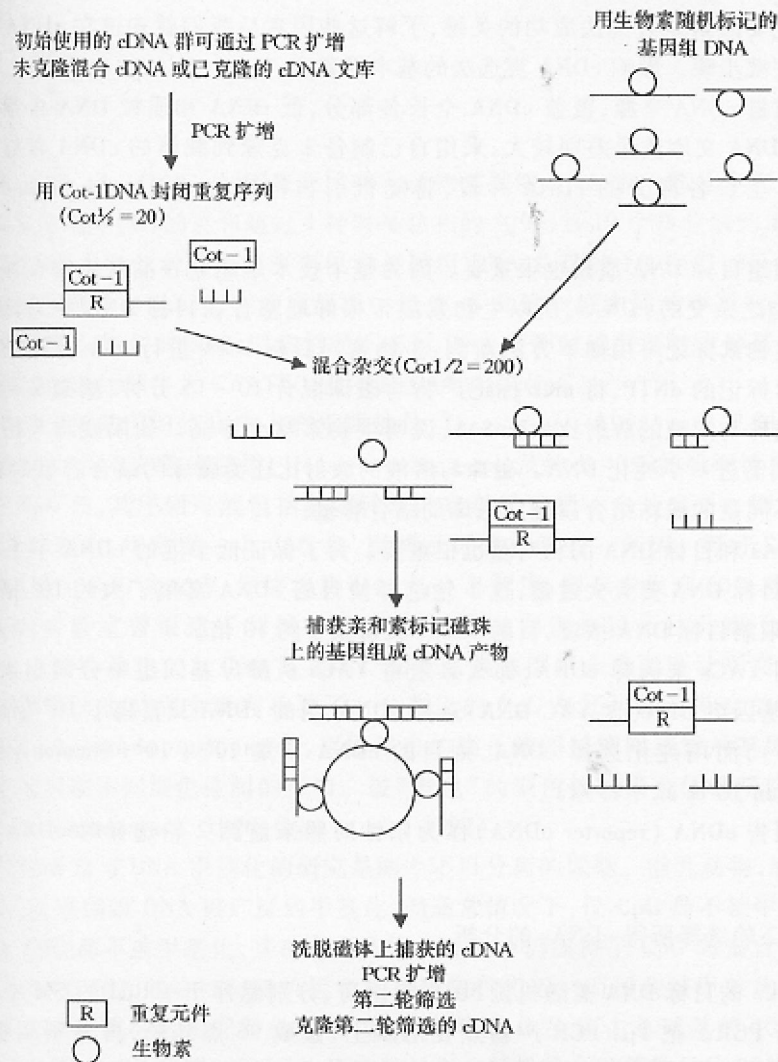


图 6.11 cDNA 直选法的流程图

- 封闭了重复序列的 cDNA 与生物素标记的目标 DNA 混合,在 $Cot1/2 = 100 \sim 200$ 条件下进行杂交。

- 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。

- 洗去非特异结合的 cDNA 后,再将目标 DNA 捕捉的 cDNA 洗脱下来。
- 用 PCR 扩增第一轮选择所得的 cDNA,并用同样方法进行第 2 轮选择。
- 将第 2 轮选择得到的 cDNA 克隆到载体上并进行分析。

6.5.3 影响 cDNA 直选法的基本因素

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:

- 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 全长各部分的 cDNA 片段,将随机引物构建的 cDNAs 与 Oligo dT 构建的 cDNAs 混合。

- 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。

- 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。

- 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(Parimoo S *et al.*, 1993),而用纯化选择 cDNA,则目的 cDNA 富集 $10^4 \sim 10^5$ (Simmons AD *et al.*, 1995; Morgan JG *et al.*, 1992)。

- 用报告 cDNA (reporter cDNA)作为阳性对照来监测 2 轮选择时 cDNA 的富集情况。

6.5.4 第 2 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 中后,取 1 μ l 噬菌体进行 PCR。把 1 μ l PCR 产物点在尼膜上,做成 96 点阵膜,再分别用报告 cDNA (reporter DNA)、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

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 2 轮选择所得的 cDNA 进行分析,发现获得至少另外 31 个 HLA 基因。用这些基因在 MHC-I 区域作图可得到详细的表达图(Wei H *et al.*, 1993)。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。

(杨新平 于常海)

6.6 CpG 岛

CpG 岛(CpG island)这一词用来描述哺乳动物基因组 DNA 中一部分序列,其特点是胞嘧啶(C)与鸟嘌呤(G)的总和超过 4 种碱基总和的 50%,每 10 个核苷酸约出现一次双核苷酸序列 CG。具有这种特点的序列仅占基因组 DNA 总量的 1% 左右,每一个单位的长度约为一千个碱基对(kb)。在此顺便提一句,CpG 岛这一词中间的小写“p”表示连接两个脱氧核苷酸 C 和 G 的磷酸,通常情况下书写 DNA 序列时磷酸基被省掉了,例如我们写 EcoRI 酶切位点的顺序为 GAATTC,而不是 GpApApTpTpC。

CpG 岛在分析基因组结构中占有重要地位(Cross *et al.*, 1995)。从已知的 DNA 序列统计发现,几乎所有的管家基因(House-keeping gene)及约占 40% 的组织特异性基因的 5'末端含有 CpG 岛,其序列可能包括基因转录的启动子及第一个外显子。因此,在大规模 DNA 测序计划中,每发现一个 CpG 岛,则预示可能在此有一基因。基于以上统计数字,如果可以建立 CpG 岛文库,最终将所有的 CpG 岛测序,则至少发现 60% 的基因的 5'端结构。我们实验室曾报道了利用特异 DNA 结合蛋白亲核层析柱分离 CpG 岛序列(Cross *et al.*, 1994)及建立人 CpG 岛文库。如果人类基因组全部测序计划完成,建立人 CpG 岛文库的意义也就完全没有必要了。分离出的 CpG 岛还可作为杂交的探针,进行“染色体油画(chromosome painting)”,这样得出的染色体图谱能间接指示基因在染色体中的分布及基因在不同染色体间的分布。被“着色”的染色体或染色体区带应富含基因(图 6.12)(Bickmore *et al.*, 1996; Cross *et al.*, 1997)。

CpG 岛的研究与 DNA 甲基化的研究是两个不可分离的课题。哺乳动物,或广延到有脊椎的动物,其基因组 DNA 被广泛的甲基化,但通常情况下,仅 CpG 岛不被甲基化修饰。可能也正是 CpG 岛不被甲基化,其在进化及传代过程中仍保持了 CpG 岛富含 C 和 G,实际的双核苷酸 CG 序列的数目与随机组合的预测数目相近。而那些被甲基化的区域,由于甲基 C 易变成 T 或与之相配对的 G 突变成 A,导致 C 及 G 占总碱基的 40% 左右,并且受影响最大的是 CG 序列,其 CG 出现的数目仅为随机组合计算数目的 20% 左右。

CpG 岛在某些特定情况下可发生甲基化。最明显的例子是正常雌性个体细胞的两条 X 染色体的一条处于非活性状态,几乎所有的 CpG 岛(除几个有活性的基因)被甲基化。