

基因克隆的常用方法介绍

孙春晓¹ 于常海^{1,2} (¹中国科学院上海生命科学研究中心, 上海 200031; ²香港科技大学生物系)

摘要 为能快速、准确地克隆出有意义的基因, 本文介绍了目前常用的一些基因克隆方法, 如差异显示 PCR、抑制性差减杂交、RAP2PCR、代表性差异显示、酵母双杂交系统、cDNA 直接捕捉法等; 并对这些方法作了简要的评价, 以利于大家选择适合自己的方法。

关键词 基因; 克隆; 差异显示

学科分类号 Q785

A Brief Introduction to the Methods for Novel Gene Cloning SUN Chun2Xiao¹, Albert C. H. Yu^{1,2} (¹ *Laboratory of Neuronal Injury and Regeneration, Shanghai Research Center of Life Sciences, Chinese Academy of Sciences, Shanghai 200031*; ² *Department of Biology, Hong Kong University of Sciences and Technology, Hong Kong*)

Abstract There are a lot of methods for novel gene cloning, but how to clone candidate gene(s) quickly and correctly? This is a brief introduction to methods of novel gene cloning, these methods includes: differential display reverse transcriptase polymerase chain reaction(DD RT2PCR), suppression subtractive hybridization (SSH), RNA arbitrarily primed PCR(RAP2PCR), representational difference analysis(RDA), yeast two2 hybrid system, cDNA capture, et al. We not only introduced these methods, but also discussed the advantages and disadvantages of them. However, no single method is omnipotent, one should pick up the method most suitable for a special purpose.

Key words Gene; Cloning; Differential display

人类 23 对染色体大约含有 3×10^9 bp DNA, 其中只有 3% ~ 5% 的基因能表达出有生理意义的蛋白质; 这些基因异常扩增、重排、缺失或突变等变化与人类众多疾病的发生发展有着密切关系。如何快速、准确地克隆出这些基因是人类基因组第二个五年计划的首要目标。目前基因克隆的方法多种多样, 本文就比较常用的几种方法分别作一简单介绍, 这些方法各有利弊, 各人应根据自己的需要选择不同的方法, 以取得理想的结果。

一、差异表达基因(片段)的获得

生物体之所以能表现出各种各样的特性, 展示出丰富多彩的表型, 主要是由于其内部基因表达的差异所致, 因此认识生物体内部基因有序地、时相性地表达是揭示遗传信息复杂性的一个极为重要的步骤。基因表达的差异表现为两个方面: 一是基因表达种类的不同; 二是基因表达水平的改变。如何检测这些差异表达的基因呢? 为此, 科学家们设计了以下几种主要方法:

(一) 差减杂交(subtractive hybridization, SH) 与抑制性差减杂交 (suppression subtractive hybridization, SSH) 差减杂交或消减杂交最早由 Lamar 和 Palmer 于 1984 年提出^[1], 他们先用超声波打断雌性小鼠的 DNA, 用 Mbo I 完全消化雄性小鼠 DNA; 将两者一起变性、复性 (其中被打断的雌性小鼠 DNA 过量 100 倍), 再将产物克隆入表达载体的 BamH I 位点中, 只

有那些两端均有 GATC 序列(即被 Mbo I 切割并自身复性的基因,也即雄鼠特有的基因)才能被克隆入载体,这样就达到了驱逐两者共有序列的目的,并最后得到了雄鼠 Y 染色体的 DNA。该方法的缺点是技术要求高、耗时、工作量大,并且有时往往不大可行或不可靠。然而经过多年的改进,它已发展成为众多基因克隆方法的基础,许多其它方法均是从该方法衍生发展而来。抑制性差减杂交(SSH)^[2]就是其中的一种。SSH 是先将 tester(样本) mRNA 和 driver(参照) mRNA 分别逆转录成 cDNA,用 4 碱基识别酶(Rsa I)酶切两种 cDNA 产生平端片段;tester cDNA 分别接上 adapter(接头)1 和 adapter 2,并与过量的经 Rsa I 消化的 driver 样本杂交。设计引物于 adapter 处,使具有差异表达的片段才能成为 PCR 扩增的模板,重复杂交以减少非特异性扩增片段,利用 adapter 上的酶切位点进行克隆、测序等。

由于每一 mRNA 逆转录成的 cDNA 经 Rsa I 酶切可产生一个以上的 cDNA 片段,故 SSH 的检测率较高、二轮杂交和二轮 PCR 可大量扩增特异表达片段。但由于二次杂交中 driver2 cDNA 均为过量,tester2cDNA 中某些表达丰度有差异的 cDNA 可能被掩盖,本方法所需的 mRNA 量较大(达数微克),稀有样本检测较困难;并且酶切后的 cDNA 与 adapter 的连接效率是实验中的关键,连接效率不高就难以发现有表达差异的基因。

(二)差异显示 PCR (differential display reverse transcriptase PCR, DD RT2PCR) 本方法由 Liang 等于 1992 年报道^[3],目前该方法在实验室被广泛使用。其主要原理是:利用大多数真核细胞基因 mRNAs 结尾处有多聚腺苷酸[poly(A)]结构,在其 3' 端设计象 5' T₁₁CA 样引物,该引物可与 mRNAs 总数的十二分之一(即 poly(A) 前面二个碱基为 TG 的 mRNA)结合,从而使这部分基因得到逆转录;而一套(即 T₁₁MN, M、N 代表 4 种碱基中的一种,但 M 不为 T,共有 12 条)引物可使全部 mRNA 得到扩增。由于 PCR 能扩增的长度是 2~3kb,而 mRNA 的平均长度只有 1.2 kb;故在其 5' 端再设计一些随意碱基顺序的引物(20 条 102mer),可以使不同长度的基因得到扩增。如果在 PCR 扩增时,采用[²³⁵S]2dATP 代替 dATP,在凝胶电泳后的放射自显影照片上就可发现不同长度的 PCR 产物。针对使用同位素[²³⁵S]2dATP 容易在做 PCR 时衰减的缺点,Liang 等重新推荐使用[²³³P]2dATP;针对使用同位素在割胶时易发生差错的缺点,Chen 等对方法作修改的同时以 DIG 代替同位素,取得了不错的效果。

该方法具有简便、灵敏、高效及省时等优点,它能够快速地显示细胞 mRNA 的组成,可对各样本 mRNAs 的差异同时进行比较和展示,并且所用的 mRNA 量少;可以立即对扩增的 cDNAs 进行测序和亚克隆及探针标记、杂交、文库筛选等;也有报道认为不宜采用从 DD PCR 得到的不纯的片段作 Northern blot 的探针。它的缺点是假阳性条带多、对低丰度的基因表达不易检测、工作量大、无法定量研究及基因的克隆受 mRNA 表达时效的影响等。造成假阳性的原因有 PCR 反应管不好、RNA 降解和 DNA 污染等;而且在电泳胶上的单一条带有可能含有一种以上的 cDNA,其中一些可能是差异表达基因,另一些可能是组成表达基因。Vogeli2 Lange 等将对照 RNA 和样本 RNA 最初的 DD RT2PCR 产物分别与差异克隆逐一杂交,根据杂交类型,将两者分开^[4]。DD RT2PCR 还有一个比较大的问题是其扩增出来的条带往往是在基因 3' 端比较短的一段序列,这部分序列往往在基因的 3' UTR 区,所提供的信息比较少,为得到 cDNA 必须进行 cDNA 文库筛选等费时耗力的工作。

(三)DNA 代表性差异分析(DNA representational difference analysis, DNA RDA) 代表性差异分析方法有二种:一种是基因组 DNA (gDNA)的代表性差异分析(gDNA RDA);另一种是 cDNA 的代表性差异分析(cDNA RDA)。前一种方法是由 Lisitsyn 等于 1993 年在差减杂

交的基础上发明^[5]的。后一种是由 Hubank 等于 1994 对前一种方法的改良而来。它们在一定程度上具有差减杂交和 DD RT-PCR 两者的特点^[6]。

在 gDNA RDA 中,先用相对不足的 6 碱基识别酶对 gDNA 进行消化,再与寡核苷酸接头连接,用 PCR 扩增 20 个循环,只有那些 1 kb 以下的片段被扩增,这些被扩增的片段称为扩增子(amplicon)(不同的内切酶可产生不同的扩增子)。在检测扩增子(tester amplicon)和驱逐扩增子(driver amplicon)产生之后,再将 tester amplicon 的 5' 端与新的 adaptor 连接并与过量的 driver amplicon 混合一起变性、复性以去除两者的共有部分(称为消减富集);加入 Taq DNA 多聚酶后,只有那些自身复性的 5' 端带有接头的双链 DNA 3' 才能被补平,这样只有这些自身复性的 tester 片段才能被随后的 PCR 扩增(称为动力学富集),而 driver amplicon 只为与 tester amplicon 的同源片段竞争复性而存在。重复杂交 2 扩增步骤可使共有序列最大程度地被去除,剩余的差异表达片段可以被用于克隆和测序。

该方法可以用于新基因的发现、肿瘤中基因的正常和遗传分析等;但由于步骤本身的随机性,消减扩增之后的 PCR 产物并不一定代表有意义的目的基因,且在样本制备中的污染也可能产生假阳性;此外,我们不清楚什么情况下,某一器官或组织会发生基因重组;也不知道组织中基因嵌合的程度。

cDNA RDA 的基本原理与 gDNA RDA 基本类似,均是通过消减富集和动力学富集而实现。Hunbank 等认为 gDNA 代表性差异分析对样本的处理太复杂,因此他们提出改进:如果先用识别 6 个碱基的酶对 gDNA 进行消化,用 PCR 进行扩增,只有 150~1000 bp 的片段可以被扩增;而大部分片段不能被扩增,其代表性片段占总数的 2%~10%;而用酶对 cDNA 消化后 PCR 扩增的结果其代表性片段只占原先的 1%~2%。少量而又充足的序列就不需要再降低其复杂性。而用识别 4 个碱基的酶(Dpn II)消化产生的代表性片段的平均长度是 256 bp,它可以保证每种主要的 DNA 片段至少有一条代表性片段被扩增。它的主要步骤是:将 tester 的 mRNA 逆转录为 cDNA,用 4 碱基识别酶消化,与 R2Bgl212/24 adapter(linker)连接,PCR 扩增,去除 R2adapter,接上 J2Bgl212/24 adapter,与 driver amplicons 杂交,补平后 PCR 扩增,将 tester 的 J2adapter 换成 N2adapter,再与 driver 杂交,再扩增,代表性片段克隆、测序。为了提高 cDNA RDA 的敏感性,O'Neill 等对此作了改进:合成 cDNA 用的 poly(A) mRNA 的量为 100~150 ng;用一种特殊的滤膜(microcon 30)除去 cDNA 中被消化的和过量的接头;在不改变 driver 和 tester 比例的前提下,将 driver 的量改为 25 ng。

gDNA RDA 由于需用 6 碱基识别酶消化样本,其样本的结构比较复杂且代表性相对较低,这样部分 tester 中可扩增的片段不能被 driver 中不可扩增的同源长片段所去除,从而易产生假阳性。与 gDNA RDA 相比,cDNA RDA 无需降低样本的复杂程度,由于用识别 4 个碱基的酶消化样本,其代表性也更完全且较大的片段也可以被扩增,因此 cDNA RDA 中杂交消减的基础是信息的有或无,而不是酶切片段的长短差异。cDNA RDA 可对不表达的基因进行检测,而 gDNA RDA 只能对大范围的缺失检测;cDNA RDA 还可以对不同阶段的基因的不同表达进行监测,可用于比较发育阶段、细胞周期的时相、药物处理前后基因表达的变化等;此外,cDNA RDA 还能检测到影响某一表型的多个基因及其下游的调节基因。

与 DD2PCR 一样,cDNA RDA 可以对一定时间内的基因表达进行检测,且所检测的样品含量可以很低。然而 cDNA RDA 还有比 DD2PCR 更好之处,cDNA RDA 不会产生模棱两可的结果,从而降低了随后分析的难度;且它几乎没有假阳性;cDNA RDA 采用全长(24 mer)的引

物而不是 DD2PCR 中 5' 很短的引物,从而避免了因引物的错配而产生的问题。此外,由于 cDNA RDA 中接头是在全长 cDNA 的消化产物上,因此它只要一套引物就可产生代表性片段,而不象 DD2PCR 一套引物只能产生十二分之一的代表性片段(DD2PCR 需要 20~25 套 5' 端引物和 4 套 3' 端引物结合才能覆盖全部 cDNA)。况且,cDNA RDA 并不依赖于 mRNA 的分离,因此,它可以用于非 poly(A) 结尾的 mRNA 的检测。但 cDNA RDA 也不是全能的,它要求样本新鲜,且它不能用于点突变、小的缺失或插入、转录末端片段及缺乏合适酶切位点的基因的检测;酶消化的结果也可以使 cDNA 两端的信息量丢失。

(四) RNA 指纹技术(RNA arbitrarily primed PCR, RAP2PCR) 该方法于 1992 年几乎与 DD RT2PCR 同时报道,且两者的原理基本一致^[7]。先从样本中提取总 RNA,用随机引物合成 cDNA 的第一条链,再用同样的引物 PCR 合成 cDNA 的第二条链,凝胶电泳可以显示全部 PCR 产物,再从中回收差异表达的基因片段,进行克隆、测序。

该方法采用随机引物而不是 oligo(dT) 逆转录 mRNA,可将非 poly(A) 结尾的 mRNA 也转录出来,避免了信息的丢失;有实验表明,RAP PCR 在一定的反应条件下,重复性相当好,RNA 的浓度、DNA 的污染等均对结果无明显影响;组间条带的亮度差异在 20% 以上即可定为基因有差异表达^[8]。但该方法需要较高的变性温度和较低的离子强度;并且随机引物有可能正好在某一类基因的保守区或分散的基因重复序列区,这样有可能只扩增这一类基因。然而,RAP PCR 在 RNA 的多态性和不同发育阶段 RNA 表达类别的差异显示上却有其用武之地。

象其它方法一样,回收 RAP2PCR 的电泳产物克隆、测序是一项累人的工作,而且有些大小类似的条带不大容易区分开来,这给以后的工作带来了极大的麻烦,为了方便电泳条带的回收,有人采用 SSCP(single strand conformation polymorphism) 凝胶分离 PCR 产物,在 SSCP 胶上可将条带分为三种:(1) 可以直接克隆相对较纯的条带;(2) 克隆前需再扩增的条带;(3) 结构太复杂的条带。这大大节省了随后的工作量。最近又有人根据 DD RT2PCR 和 RAP2PCR 两者重复性上的缺点,对其作了改进并同时采用非同位素和非变性胶,成了另一种基因差异显示方法(DEmRNA PCR)。

(五) 酵母双杂交系统(yeast two-hybrid system) 该方法由 Chien 等于 1991 年创立,首先构建二个表达杂交体的质粒:一是酵母转录活化蛋白 GAL4 的 DNA 结合区和已知蛋白 SIR4 的融合体;二是 GAL4 的活化区与酵母基因组 DNA 库编码的蛋白的融合体。将这二种质粒同时转入酵母细胞中,如果已知蛋白和文库的一种编码蛋白相互作用,则可使与 GAL4 有结合位点的报告基因转录活化,通过筛选 GAL4 依赖的报告基因的转录就可以发现与已知蛋白相互作用的蛋白的编码基因^[9]。该法用途有三:(1) 验证已知基因间的相互作用;(2) 可以快速发现编码蛋白与蛋白间相互作用的特定区域;(3) 通过扫描文库与活化区域的作用可以发现编码相互作用的蛋白,只需一个质粒就可以直接得到编码蛋白的基因,又无需制备抗体和纯化蛋白。由于双杂交系统应用已十分广泛,在此不多作介绍。有关它的操作步骤可参阅文献^[10];有关其在蛋白连锁中的作用可参阅文献^[11];对该方法中假阳性的防止可参阅文献^[12]。

(六) cDNA 捕捉法(cDNA capture) 它是一种用 YAC 探针或基因组探针直接捕捉 cDNA 的方法。最先由 Elvin 等报道,他们用 YACs 探针直接筛选 cDNA 文库,获得了目的基因。后来,Parimoo 等又提出了另一种 cDNA 直选法,其基本原理与前者类似,他们将 gDNA 固定于尼龙盘上,用两侧带有载体序列的短的随机片段与其杂交,再用载体引物做 PCR 和巢式 PCR,再酶切克隆。随后,Rouquier 等又提出了第三种 cDNA 直选法^[13],他们先从组织中提

取 mRNA,合成双链 cDNA,用 Cot 1 DNA 预杂交封闭重复序列;流式细胞仪分选染色体, DOP2PCR (degenerate oligonucleotide primed PCR) 并生物素化,与已封闭的 cDNA 杂交,再用链亲和素包被的磁珠捕捉 cDNA,严格洗脱后,PCR 扩增后进行第二轮筛选,最后对 cDNA 进行染色体作图。这些方法比较适用于基因组区域内基因表达序列的分离,可在大至 2Mbp 的基因组区域内构建详细的表达图谱。在采用这些方法时,cDNA 和基因组目标 DNA 的质量、cDNAs 和目标 DNA 的相对量均十分重要;如采用 YACs 探针,但最好能纯化后使用。

(七)其它

1. 扩增限制性片段长度多态性 (amplified restriction fragment length polymorphism, AFLP):其最大的好处是只用一 4 碱基识别酶即可筛选大量不同的 cDNA,它由 Vos 等率先报道。其基本步骤是:将 DNA 用一种常用酶和另一种稀有酶分别消化后,接上接头(该接头由三部分组成,分别为核心序列、酶特异序列和选择延伸部分),先用 PCR 预扩增,再作 PCR 扩增(后一轮 PCR 的一种引物标有同位素),变性测序胶电泳。后来有人将其基本步骤略作了修改:poly(A) mRNA 逆转录成 cDNA 后锚定 oligo(dT) 引物,Taq I 酶消化,加上接头,预杂交,选择性扩增,再在测序胶上展示。

2. cDNA 微量排列法 (microarray)^[14]:该方法采用高速机器人技术将 1046 条未知序列的人 cDNA 和 10 个阴性对照点在经过多聚赖氨酸处理的玻片上,制成 1.0 cm² 的 DNA 芯片 (DNA chips)。从组织 mRNA 中提取 Jurkat mRNA 逆转录成 cDNA 探针,该探针标有 Cy52 dCTP;再将对照标本标上 fluorescein,实验样本标上 Cy52dCTP。而在第二套标记反应中,对照样本和实验样本却分别被标上 fluorescein 和 Cy52dCTP,每一种荧光探针均与 1056 点阵玻片杂交,根据荧光强度的变化,挑取变化明显的克隆,测序分析。该方法杂交体积小、分析密度大、采用荧光标记、可同时检测 1000 条以上的基因,商业化 cDNA chips 的出现大大促进该方法的使用,而大规模 ESTs 的涌现使 cDNA chips 有了丰富的资源,但其灵敏度有待于进一步提高。

3. 外显子捕获法 (exon capture):由于外显子的 5' 和 3' 端均有功能性剪接位点,将含有未知基因的基因组 DNA 插入真核表达载体的 HIV 1 tat 基因的内含子中,该内含子的二侧均为 RNA 剪接位点,将载体导入 COS27 细胞中,利用细胞内的转录剪接机制,产生成熟的 RNA,该 RNA 含有未知序列的外显子,可利用 RT2PCR 将其扩增、克隆。该方法高效、特异、简便,即使内含子大于 10 kb 也不影响效果。但它也有缺点:有些目的外显子不能有效地接入载体的外显子之间;目的外显子两侧的内含子序列可能由于剪接位点的活化而延长,从而使 RNA 的产生只有正常的 1/10 ~ 1/5,实验中的假阳性也不容忽视。

4. 序列基因表达测定 (serial analysis of gene expression, SA GE)^[15]:SA GE 方法基于二条原则:一是来自转录体特定位置的 9 - 10 bp 的核苷酸序列已含有足够的信息来确定转录体;二是通过对一个克隆内一系列标签的测序并将这些标签联系起来,就可以对转录体进行一系列研究。它的基本思路是:将 mRNA 逆转录成 cDNA,用 4 碱基识别酶(称为锚定酶 anchoring enzyme, AE)消化,将切割后的 3' 与链亲和素微珠 (streptavidin beads) 结合后分成二组,分别与接头 A 和接头 B 连接,再用标签酶 (tagging enzyme, TE) 切割产生平末端,将 A 与 B 连接后用引物 A、B 扩增,再用 AE 消化后分离片段进行克隆。该方法不仅可提供已知基因丰度的定量信息,还能发现新基因。为了减少接头对该方法效率的影响,有人提出用生物素修饰 PCR 引物,尔后再用链亲和素与之结合以去除接头,这样可大大提高该方法的效率。

5. 此外尚有接头连接与“vectorette PCR”相结合的抑制性 PCR (suppression PCR) 和从细

胞杂交体的转录本中直接克隆等方法。总之,基因克隆方法多种多样,不一而足。

二、基因全长的获取

差异表达基因片段的获得是认识遗传信息复杂性的第一步和基础,其目的是为了获取基因的全长、研究其结构,并最终认识其功能。目前从基因片段获得其全长的方法主要有:

(一) 基因文库的筛选 该方法已成经典,读者可参阅其它书籍。

(二) 3' RACE or 5' RACE (rapid amplification of cDNA ends) cDNA 末端快速扩增法也是用得比较多的一种方法,它分成二种类型,即 3' RACE 和 5' RACE,二种类型原理一样,操作步骤略有不同^[16]。3' RACE:用 35 bp 的引物(17 个 dT,含三个 gDNA 中稀有内切酶识别位点的接头序列)将 mRNA 逆转录成 cDNA 负链;再用基因特异的引物(3' amp)合成 cDNA 的正链;最后用接头引物和 3' amp 同时扩增。5' RACE:用基因特异的引物(5' amp)将 mRNA 逆转录成 cDNA 负链,将多余的引物去除后用 dATP 和末端脱氧核糖转移酶接上 poly(A) 尾巴,用(dT)₁₇2接头引物合成正链,再用 5' amp 和接头引物同时扩增。目前,有许多基因的全长就是采用 3' RACE 和 5' RACE 的方法获得的,GDNF 就是其中的代表,而且现在已经有 3' RACE 和 5' RACE 的试剂盒可供购买。

(三) 计算机杂交(computer hybridization)^[17] 计算机杂交是近几年才发展起来的新兴方法,它是基因克隆的蓬勃发展和计算机技术日益成熟的必然结果。它指的是利用计算机对现有的基因序列、ESTs、蛋白序列等进行扫描,同时与目的片段进行比较,清除垃圾片段,从而确定目的片段与已知基因、基因片段的同源性和成为新基因的可能性,并利用现有的基因序列,对目的基因片段进行拼接加工,以完成基因全长的克隆。计算机杂交的理论基础是绝大多数功能基因的编码区比较保守,即同源性较高。目前有众多软件收有大量的基因序列可供查询,国际上最著名的三家核酸序列数据库是:GenBank、EMBL 和 DDBJ (the DNA Database of Japan)。这三家公司每天交换信息,它们的网点分别是:www.ncbi.nlm.nih.gov, www.embl2ebi.ac.uk, sakura.ddbj.nig.ac.jp;而最主要的蛋白质数据库在:GenBank、EMBL、PIR 和 Swiss2PORT。后二者的网址为:www2nbrf.georgetown.edu/pir/ 和 expasy.hcuge.ch/sprot/sprot2top.html。而 ESTs 的数据库也主要在 GenBank、EMBL 和 DDBJ。用户可以通过 E-mail、匿名 FTP、及 Internet 网等进行查询;更详尽的资料可以从 Nucleic Acids Research 杂志中获取,现在每年的第一期均为数据库专辑(database issue)。

用户查询数据库的软件也多种多样,主要是 BLAST (Basic Local Alignment Search Tool), 它有基本的(basic)和高级的(advanced)二种查询方式,核酸序列与核酸数据库的比较在 BLASTN;核酸序列与蛋白质数据库的比较在 BLASP;核酸的六种翻译序列与蛋白质数据库的比较在 BLASTX;蛋白质序列与核酸数据库所有序列六种翻译的比较在 tBLASTN;核酸序列的六种翻译与核酸数据库所有序列六种翻译的比较在 tBLASTX。将核酸序列以 FASTA 形式输入,选定查询方式即可进行查询,并可利用已有 ESTs 资料对目的序列进行加工,根据基因的同源性可以获得目的基因的全长或者基因的间断序列,再利用 PCR 等即可得到全长。

三、直接测序(direct sequencing)或大规模测序

这是目前大多数基因组测序中心在进行大规模基因测序时最常采用的方法,其基本步骤是:先将样本 DNA 随机切成 1.5 kb 大小左右的片段并克隆入合适的测序载体;再对每 kb 的 DNA 进行 10 - 30 个亚克隆的高覆盖率的测序;然后根据测出的相互重叠序列组装成连续的多序列重叠线;最后从质量最高的测得序列中获得一致序列(consensus sequence)。近来有人

提出了全基因组随机测序法,即全基因组鸟枪战略 (whole2genome short2gun strategy)^[18]。该方法的基本步骤是:用机械的方法将基因组 DNA 随机打断后,装入适当的测序载体,然后对插入的 DNA 片段的两端各 500 bp 进行测序。这个方法的提出将传统的基因组测序思路倒了过来,它的战略思想是先测序,后作图。是对不完整 DNA 序列的巧妙应用,它可以大大加快基因组测序的步伐,而且毛细管电泳测序仪的问世使该方法如虎添翼。由于这两种方法主要用于人类基因组的研究,一般的实验室不大会做到,这里不多作讨论。

四、定位克隆 (positional cloning)

基因定位克隆作为基因克隆方法的一种,主要用于与遗传病相关基因等致病基因的克隆。它先根据已知的遗传标记进行连锁分析和系谱分析,先确定候选基因所在的位置,再通过其它方法获得基因的全部序列。基因的遗传连锁分析的原理可参阅文献^[19]。大规模遗传连锁分析所需计算机软件可从下列网址中获得: <http://www.genome.wi.mit.edu/genome2software>; 突变表型资源库在: <http://www.mgu.har.mrc.ac.uk> 和 <http://www.jax.org>。小鼠基因组微卫星标记可从: <http://www.resgen.com> 和 <http://www.genome.wi.mit.edu> 中获得。

基因定位克隆中获得基因序列的方法大致有^[20]: (1) 对关键部位进行直接测序,目前已经可以对 500 kb 左右的区域进行直接测序; (2) 比较基因组作图和测序,有关的信息可从 <http://www.ncbi.nlm.nih.gov/XREFdb/> 中得到; (3) 位置候选分析 (positional candidate analysis), 从某种程度上讲,它将成为基因定位克隆的标准步骤,它是在被克隆的基因 (往往是 ESTs 形式) 和它们相应的染色体位置日益清楚的基础上迅速发展起来的,有关这方面的信息收录在: <http://www.2shgc.stanford.edu/cgi2bin/smag#GOTO>; (4) 基因结构特征分析,主要有三种方法: HTF 岛 (HTF island) 作图 (非甲基化 CpG 二核苷酸)、进化保守区分析和外显子捕获。HTF 作图法是根据稀有切割酶对基因组 DNA 的特征性切割,产生可识别的标记,如基因的 5' 端和 CpG 二核苷酸等,有人又称之为限制性标记基因组扫描 (restriction landmark genome scanning, RLGS); (5) cDNA 捕获,它是根据如果基因组 DNA 和已知 cDNA 序列同源,它们就能形成异二聚体。这样将二者接上接头杂交后,再经过 PCR 扩增等即可得到未知基因。

参 考 文 献

- 1 Lamar EE, Palmer E. Y2encoded, species2specific DNA in mice: evidence that the Y chromosome exists in two polymorphic forms in inbred strains. *Cell*, 1984, 37: 171 ~ 177.
- 2 Diatchenko L, Lau YFC, Campell AP, et al. Sup2pression subtractive hybridization: a method for generating differentially regulated or tissue2specific cDNA probes and libraries. *Proc Natl Acad Sci*, 1996, 93: 6025 ~ 6030.
- 3 Liang P, Pardee AB. Differential display of eukary2otic messenger RNA by means of the polymerase chain reaction. *Science*, 1992, 257: 967 ~ 970.
- 4 Vogeli2Lange R, Burckert N, Boller T, et al. Rapid selection and classification of positive clones generated by mRNA differential display. *Nucl Acids Res*, 1996, 24: 1385 ~ 1386.
- 5 Lisitsyn N, Lisitsyn N, Wigler M. Cloning the dif2ferences between two complex genomes. *Science*, 1993, 259: 946 ~ 951.
- 6 Hubank M, Schatz DG. Identifying difference in mRNA expression by representational difference analysis of cDNA. *Nucl Acids Res*, 1994, 22: 5640 ~ 5648.
- 7 Welsh J, Chada K, Dalal SS, et al. Arbitrarily primed PCR fingerprinting of RNA. *Nucl Acids Res*, 1992, 20: 4965 ~ 4970.

8 Tótola S, Capell ÆG, Marcuello E, et al. Analysis of differential gene expression in human colorectal tumor tissues by RNA arbitrarily primed PCR: a technical assessment. *Lab Invest*, 1998, 78: 309 ~ 317.

9 Chien CT, Bartel PL, Sternalanz R, et al. The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. *Proc Natl Acad Sci*, 1991, 88: 9578 ~ 9582.

10 Getz RD, Triggs-Raine B, Robbins A, et al. Identification of proteins that interact with a protein of interest: application of the yeast two-hybrid system. *Mol Cell Biochem*, 1997, 172: 67 ~ 79.

11 Hua SB, Luo Y, Qiu M, et al. Construction of a modular yeast two-hybrid cDNA library from human EST clones for the human genome protein linkage map. *Gene*, 1998, 215: 143 ~ 152.

12 Wong C, Naumovski L. Method to screen for relevant yeast two-hybrid-derived clones by coimmunoprecipitation and colocalization of epitope-tagged fragments: application to Bcl-2. *Anal Biochem*, 1997, 252: 33 ~ 39.

13 Rouquier S, Trask BJ, Taviaux S, et al. Direct selection of cDNAs using whole chromosomes. *Nucl Acids Res*, 1995, 23: 4415 ~ 4420.

14 Schena M, Shalon D, Heller R, et al. Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc Natl Acad Sci*, 1996, 93: 10614 ~ 10619.

15 Velculescu VE, Zhang L, Vogelstein B, et al. Serial analysis of gene expression. *Science*, 1995, 270: 484 ~ 487.

16 Frohman MA, Dush MK, Martin GR. Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc Natl Acad Sci*, 1988, 85: 8998 ~ 9002.

17 Mao M, Fu G, Wu JS, et al. Identification of genes expressed in human CD34⁺ hematopoietic stem/progenitor cells by expressed sequence tags and efficient full-length cDNA cloning. *Proc Natl Acad Sci*, 1998, 95: 8175 ~ 8180.

18 Venter JC, Adams MD, Sutton GG, et al. Shotgun sequencing of the human genome. *Science*, 1998, 280: 1540 ~ 1542.

19 Boyd Y. Genetic mapping of mouse genome. *Methods*, 1998, 14: 120 ~ 134.

20 Boehm T. Positional cloning and gene identification. *Methods*, 1998, 14: 152 ~ 158.

丙肝研究取得新进展

丙肝病毒(HCV)对公众的健康影响很大,全世界估计已有 1.7 亿人被它感染。科学家们正在努力研究这种病毒,希望治愈由它引起的慢性肝炎。目前研究的最大障碍是缺少研究 HCV 的细胞培养系。没有这些实验体系,科学家们就无法在分子水平上对 HCV 的生活周期进行详细研究,也不能检测新的抗病毒治疗法。Lohmann 等通过对 HCV 基因组的改造已能“强迫”HCV 在培养的人体细胞中复制。研究人员可以利用这一体系向许多 HCV 研究的新领域进发。Taylor 等揭示了 HCV 躲避干扰素抗病毒效应的可能机制。目前干扰素是治疗 HCV 唯一可行的方法。它通常和另一种药物,三(氮)唑核苷(ribavirin)结合使用。干扰素对大多数人虽然没有效果,但当它起作用时,就会激活一种名为 PKR 的酶。PKR 可以攻击某些细胞蛋白,导致蛋白合成终止。没有蛋白合成,病毒不能成功地感染。Taylor 等发现一种叫 E2 的 HCV 蛋白在体外可与 PKR 结合并抑制它的活性。现在研究人员已经找到一种 HCV 抵抗干扰素的可能机制,科学家们可以借助该成果找到更好的治疗方法。

(*Science*, 1999, 285: 110 ~ 113) (车 笛)