

筛选差异表达基因和蛋白质的方法进展*

王吉村 药立波** 赵忠良

(第四军医大学生物化学与分子生物学教研室, 西安 710032)

摘要 分离和鉴定差异表达基因和蛋白质不仅有助于发现基因和蛋白质的功能, 更有助于揭示某些疾病的发生机理. 目前筛选差异表达基因的方法主要有差异显示 PCR 方法 (differential display RT-PCR, DDRT-PCR)、消减杂交法 (subtractive hybridization, SH)、基因芯片技术 (DNA chip technique) 和基因表达的系统分析 (serial analysis of gene expression, SAGE) 等, 其中消减杂交法中又先后建立了代表性差异分析技术 (representational difference analysis, RDA)、抑制消减杂交法 (suppression subtractive hybridization, SSH) 和获得全长基因的消减杂交法 (full length gene obtainable subtractive hybridization). 筛选差异表达蛋白质的方法主要有双向电泳技术 (two dimensional gel electrophoresis) 和噬菌体全套抗体库技术 (phage display antibody repertoire library technique). 这些方法各有特点, 各有利弊, 研究者可根据自己的需要选择适合于自己的方法.

关键词 筛选, 差异表达基因, 差异表达蛋白质

学科分类号 Q78

对于一个生物个体来讲, 虽然所有细胞的基因组组成是相同的, 但在不同的细胞、同一细胞的不同分化阶段和同一细胞受到不同的外界刺激时, 有关基因的表达及蛋白质的功能都会有所不同. 研究两种处于不同状态的细胞在基因表达上的差异, 有助于了解基因和蛋白质的功能. 分离与疾病相关的基因或蛋白质更具有实用价值. 目前筛选差异分子的研究可从两个水平进行: 一是从基因水平筛选差异表达基因, 二是从蛋白质水平筛选差异表达蛋白质, 下面分别综述.

1 筛选差异表达基因

在基因水平, 也就是在 mRNA 水平或 cDNA 水平筛选差异基因, 可以通过引入 PCR 技术, 将目的片段扩增, 使其容易克隆、保存和进行序列分析, 目前从基因水平筛选差异基因主要有以下几种方法:

1.1 差异显示 PCR 方法

1992 年 Liang 等^[1] 创建了差异显示 PCR (differential display RT-PCR, DDRT-PCR) 方法, 就此揭开了大规模筛选差异基因的序幕. 它是利用可以扩增所有哺乳类 mRNA 的几条 5' 武断引物和几条 3' 锚定引物随机组合, 将 cDNA 进行 PCR 扩增, 这样对于相同的 cDNA, PCR 产物的种类多少和分布应该是完全一样的, 而对于不同的 cDNA 则不同. 将所有产物电泳分离后比较, 就可以找到差

异条带, 它们所代表的 cDNA 就有可能与细胞状态的改变相关^[2].

但实践证明, 此方法虽有很多优点如快速、所需 RNA 量少、可以同时显示多种生物性状的差异及可以同时获得高表达和低表达的基因等, 但也有很大的局限性: a. 假阳性率高, 即差异片段常不能在 RNA 印迹中被证实, Sun 等^[3] 报道, 这种假阳性率可达 70%; b. 对高拷贝基因有偏向性; c. 获得的片段太短, 多为 300 bp 左右, 多为 3' 非编码区, 很难直接判断其功能和意义.

1.2 消减杂交法

经典的消减杂交法 (subtractive hybridization, SH) 是将实验组 mRNA 与对照组 (也叫驱动组或扣除组) cDNA 杂交, 或将实验组 cDNA 与对照组 mRNA 杂交, 去除杂交体和未杂交上的对照成分 (cDNA 或 mRNA), 得到的为实验组独有的 mRNA 或 cDNA, 这些基因即为差异表达基因. 传统的消减杂交法对 mRNA 的质和量要求都很高, 而且还很难获得低丰度的差异基因. 为此, 人们引入了 PCR 技术, 先后建立了几种以 PCR 为基础的消减杂交方法.

* 国家自然科学基金杰出青年基金 (39825113) 和重点课题 (39730430-II) 资助项目.

** 通讯联系人.

Tel: 029-3374513, E-mail: wangjicun@fmmu.edu.cn

收稿日期: 1999-12-31, 接受日期: 2000-02-29

1.2.1 代表性差异分析技术 (representational difference analysis, RDA): 1993年 Lisitsyn 等^[4]建立了第一种以 PCR 技术为基础的消减杂交技术, 随后 Hubank 等又将其改进为 cDNA-RDA. 原理是, 首先用 *Dpn* II 或 *Sau* 3A 限制性内切酶分别消化两组双链 cDNA, 两端接上单链接头 R, 补平后, 用含接头 R 序列的引物扩增, 再用 *Dpn* II 或 *Sau* 3A 去除双链 cDNA 两端的接头 R; 然后单给实验组双链 cDNA 两端再接上单链接头 J, 实际上接头是加在 cDNA 每条链的 3' 端. 将实验组与驱动组 cDNA 杂交, 杂交反应体系中只有实验组自身杂交形成的双链 cDNA 才两端都带接头 J, 补平后可被含有接头 J 序列的引物几何级数扩增, 而只有一端带接头 J 的杂交体只能被线性扩增, 并可被大豆核酸酶去除. 将此杂交产物再行第二轮酶切、加接头、杂交和 PCR 扩增, 重复两次后, 可确保从实验组中彻底去除与驱动组共有的序列.

这种方法 mRNA 用量少, 特异性高. 但有三个缺点: a. 实验组低丰度的分子自身杂交的几率很低, 所以被扩增和克隆的几率仍很低; b. 酶切连接步骤太多, cDNA 会损失很多, 所以可能漏掉关键基因; c. 得到的不是 cDNA 全长而是其酶切片段.

1.2.2 抑制消减杂交法 (suppression subtractive hybridization, SSH): 1996年, Diatchenko 等^[5]建立了另一种应用 PCR 方法的消减杂交技术. 这种方法与 RDA 方法有些类似, 也是先将双链 cDNA 酶切为短片段 (如 *Rsa* I 或 *Hea* III), 但后续工作是将实验组分为两组加不同的单链接头 (这些接头实际上是连于 DNA 每条链的 3' 端), 分别为接头 1 和接头 2, 它们分别具有一段反向末端重复序列. 将实验组的两组样品分别与驱动组过量的酶切产物杂交, 第一轮杂交后, 两组杂交产物混合, 将这些产物再与驱动组进行第二轮杂交, 杂交产物补平后作为模板, 用含反向末端重复序列的引物进行 PCR 扩增. 杂交产物中驱动组自身杂交产物和实验组与驱动组的杂交体因末端不带接头或只有一端有接头而不能被扩增. 杂交产物中还有三种两端带接头的产物, 它们都是实验组 cDNA: A 产物为两端带接头 1 的双链 cDNA, 每条单链 cDNA 来源于第一次杂交的同一体系; B 产物为两端带接头 2 的双链 cDNA, 每条单链 cDNA 来源于第一次杂交的同一体系; C 产物为两端分别带接头 1 和 2 的双链 cDNA, 每条单链 cDNA 来源于第一次杂交的不同

体系, 更具有特异性. 而 A 产物和 B 产物带含反向末端重复序列的相同接头, 所以在 PCR 变性和复性过程中, 每条单链 cDNA 可自身头尾互补结合形成发卡结构而不能被扩增, 只有 C 产物才能被扩增出. 所以这种方法称为抑制消减杂交法.

这种方法的优点是, 因为被扩增和克隆的双链 cDNA 的每条链来源于第一次杂交的不同体系而且又进行了第二次杂交, 所以具有很高的特异性. 但其缺点与 RDA 法相似: a. 实验组中带不同接头的低丰度单链 cDNA 与互补链杂交的几率很低, 不易被扩增克隆; b. 加接头过程中会有 cDNA 损失; c. 不能直接获得全长 cDNA.

1.2.3 获得全长基因的消减杂交法 (full-length-gene-obtainable subtractive hybridization): 这种方法是本室 Zhao 等^[6] 1998 年首先建立的, 其原理是, 分别将实验组和扣除组的 mRNA 反转录, 其中对于实验组, 利用 mRNA 的 3' poly (A) 尾和 5'-GGG 帽结构分别在 cDNA 两端加上序列相同的锚定引物, 对于扣除组, 将其 cDNA 进行随机引物 PCR 扩增制备扣除探针, 同时掺入生物素 (biotin-dUTP); 然后将实验组的 cDNA 与掺有生物素的扣除探针进行液相杂交, 杂交反应物用连有链亲和素的磁珠吸附掉杂交体及未杂交上的扣除探针, 得到实验组独有的 cDNA, 然后用锚定引物将其扩增并克隆.

这种方法有两大优点: a. 杂交后未杂交上的实验组单链 cDNA 可直接作为模板被扩增, 所以低丰度的 cDNA 被扩增克隆的几率大大提高; b. 可以直接获得全长 cDNA, 使获得功能基因的速度大大加快. 这两个优点都是上述方法不可比拟的, 而且方法操作简便, 快速, 值得推广.

1.3 基因芯片技术

基因芯片技术 (DNA chip technique) 也叫 cDNA 微阵列杂交技术 (cDNA microarray hybridization), 就是将大量探针分子固定于支持物上, 后与标记的样品 cDNA 杂交, 杂交信号阳性的探针分子就是在组织或细胞中表达的分子片段^[7]. 若同时将两组生物性状相似但又有不同的组织或细胞 cDNAs 分别与芯片杂交, 芯片上信号分子的分布就会有所不同. 基因芯片由于同时将大量的探针固定于支持物上, 所以可以一次性对细胞内多种分子进行检测和分析, 从而找到差异分子. 这种方法灵敏度高, 重复性好, 通过购买商品化的芯片和采用自动化分析技术, 使寻找差异基因和功能基因的

工作变得简便快速。

1.4 基因表达的系统分析

上述几种方法虽然各有优势,但都只能提供部分信息,不能全面反映组织或细胞的基因表达变化。Velculescu 等^[8]建立的基因表达系统分析(serial analysis of gene expression, SAGE),可以快速和详细地分析成千上万个基因。方法是将 mRNA 用生物素标记的 oligo dT 合成双链 cDNA,用识别 4 个碱基的内切酶切割得到多个长 250 bp 左右的片段。然后用连有链亲和素的磁珠将带 poly (A) 尾,也就是 cDNA 3' 短片段分离出来,每个 3' 片段代表一个 cDNA。把 cDNA 分成两份,分别连上包含标签内切酶(这种酶属 II 类限制酶,能在距识别位点 20 碱基位置切割双链 DNA,产生平末端)位点的接头(A 和 B),再用标签酶消化,产生单侧平末端 cDNA。然后将两份 cDNA 混合连接。随机两种分子靠平末端连接,产生两端分别带 A 和 B 接头的 cDNA 连接体,即双标签。这些双标签再被标签酶酶切后,靠其平末端再首尾连接串联起来。然后将多标签的串联体一同扩增连入载体。这种方法可以全面提供生物体基因表达谱信息。它还可用来定量比较不同状态下的组织或细胞的所有差异表达基因。

2 筛选差异表达蛋白质

以上是从 RNA 水平研究差异表达的基因,而实际上有些 DNA 或 RNA 的改变,并不影响到其编码的蛋白质的改变,而且某些蛋白质被合成后需经“翻译后修饰”才呈现功能,所以从蛋白质水平找差异分子更准确直接。这里主要介绍双向电泳技术和全套抗体库技术。

2.1 双向电泳技术

等电聚焦技术是利用蛋白质的等电点将细胞全蛋白质分离,SDS-聚丙烯酰胺凝胶电泳技术是按分子质量将细胞全蛋白质分离,用这两种技术结合起来的 双向电泳技术(two-dimensional gel electrophoresis)就可以基本上将细胞内所有蛋白质一一分开。将两组细胞或组织样品的蛋白质成分在完全相同的条件下,进行双向电泳分离,比较胶上的蛋白质分布,就可以找到差异表达的蛋白质点,经氨基酸序列分析,就可知差异蛋白质的特点和功能。目前已有整套的双向电泳设备(Pharmacia 公司),保证了分离条件的完全一致,也就保证了筛选差异蛋白质的可靠性和重复性。

这种方法不仅可以将细胞全成分进行分离和比较,也可以将某一类蛋白质进行分离和比较,也就是功能蛋白质组学。如结合超速离心技术可以进行细胞膜蛋白质的分析;还可以在分离前,将细胞全成分进行磷酸化反应,然后进行双向电泳,放射自显影后,得到的图谱为所有被磷酸化的分子,这类分子的表达差异可能在细胞病变中具有更重要的意义^[9]。

2.2 噬菌体全套抗体库技术

噬菌体表面全套抗体库技术(phage display antibody repertoire library technique)是将全套抗体基因克隆入噬菌粒表达载体,使全套抗体分子片段被呈现表达于噬菌体蛋白质 III 上^[10]。用全套抗体库筛选差异蛋白质时,先用目的细胞对噬菌体表面全套抗体库进行筛选,称“阳性选择”,再用阳性选择得到的噬菌体抗体对对照细胞进行“阴性选择”,去除识别目的细胞与对照细胞共同抗原的噬菌体抗体,从而得到仅识别目的细胞特有的差异蛋白质的抗体,依此抗体可以确定或捕获差异蛋白质^[11]。

参 考 文 献

- 1 Liang P, Pardee A B. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science*, 1992, **257** (5072): 967~ 971
- 2 Liang P, Pardee A B. Differential display. A general protocol. *Mol Biotechnol*, 1998, **10** (3): 261~ 267
- 3 Sun Y, Hegamyer G, Colburn N H. Molecular cloning of five messenger RNAs differentially expressed in preneoplastic or neoplastic JB6 mouse epidermal cells: one is homologous to human tissue inhibitor of metalloproteinases 3. *Cancer Res*, 1994, **54** (5): 1139~ 1144
- 4 Lisitsyn N, Lisitsyn N, Wigler M. Cloning the differences between two complex genomes. *Science*, 1993, **259** (5097): 946~ 951
- 5 Diatchenko L, Lau Y F, Chenchik A, *et al.* Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc Natl Acad Sci USA*, 1996, **93** (12): 6025~ 6030
- 6 Zhao Z L, Huang X, Li N, *et al.* Direct cloning of cell differential expression genes with full-length by a new strategy based on the multiple rounds of “long distance” polymerase chain reaction and magnetic beads mediated subtraction. *J Biotechnol*, 1999, **73** (1): 35~ 41
- 7 Kurian K M, Watson C J, Wyllie A H. DNA chip technology. *J Pathol*, 1999, **187** (3): 267~ 271
- 8 Velculescu V E, Zhang L, Vogelstein B, *et al.* Serial analysis of gene expression. *Science*, 1995, **270** (5235): 484~ 487

- 9 Guy G R, Philip R, Tan Y H. Analysis of cellular phosphoproteins by two-dimensional gel electrophoresis: Applications for cell signaling in normal and cancer cells. *Electrophoresis*, 1994, **15** (3~4): 417~440
- 10 Sheets M D, Amersdorfer P, Finnern R, *et al*. Efficient construction of a large nonimmune phage antibody library: The production of high-affinity human single-chain antibodies to protein antigens. *Proc Natl Acad Sci USA*, 1998, **95** (11): 6157~6162
- 11 Burioni R, Plaisant P, Bugli F, *et al*. A new subtraction technique for molecular cloning of rare antibody specificities from phage display libraries. *Res Virol*, 1998, **149** (5): 327~330

The Progress of the Methods for Screening Differentially Expressed Genes and Proteins*

WANG Ji-Cun, YAO Li-Bo**, ZHAO Zhong-Liang

(Department of Biochemistry and Molecular Biology, Fourth Military Medical University, Xi'an 710032, China)

Abstract Cloning and identification of differentially expressed genes or proteins is helpful not only for finding the functions of genes and proteins, but also for discovery of the mechanism of some diseases. Some methods have been developed for screening differentially expressed genes, such as differential display RT-PCR (DDRT-PCR), subtractive hybridization (SH), DNA chip technique, and serial analysis of gene expression (SAGE). In subtractive hybridization, there have advanced three improved methods which include representational difference analysis (RDA), suppression subtractive hybridization (SSH), and full-length gene-obtainable subtractive hybridization. For obtaining differentially expressed proteins, scientists have only two choices so far. One is two-dimensional gel electrophoresis. The other is phage display antibody repertoire library technique. Since all of the methods above have their own advantages and disadvantages, they should be used according to different needs.

Key words screening, differentially expressed genes, differentially expressed proteins

* This work was supported by a grant from National Science Foundation of China (39825113 and 39730430-II).

** Corresponding author. Tel: 86-29-3374513, E-mail: wangjicun@fmmu.edu.cn

Received: December 31, 1999 Accepted: February 29, 2000