

# 发现新基因的高效方法——cDNA 文库的复性式均一化技术

王 强,刘秋云,李宝健

(中山大学生物工程研究中心和基因工程教育部重点实验室,广州 510275)

**摘要:**由复性式均一化技术制作的均一化 cDNA 文库(equalized cDNA library, normalized cDNA library)是近年来发展起来的一种获得 EST、发现新基因的高效平台。本文就该技术的原理、方法比较、存在问题和展望进行了阐述。

**关键词:**cDNA 文库;均一化;复性

**中图分类号:**Q75

**文献标识码:**A

**文章编号:**0253-9772(2002)03-0325-04

## An Efficient Approach to Gene Discovery—Normalization of cDNA Library by Reassociation

WANG Qiang, LIU Qiu-yun, LI Bao-jian

(Bioengineering & Research Center and the Key Laboratory of Gene Engineering of Ministry of Education, Zhongshan University, Guangzhou 510275, China)

**Abstract:** The cDNA library normalized by reassociation is a newly developed, effective platform for EST acquisition and gene discovery. This paper presents the principle, procedure, comparison, deficiencies, application and future of the technique of the normalization.

**Key words:** cDNA library; normalization; reassociation

### 1 概 述

在分子生物学中,研究方法决定我们能看到什么。由 mRNA 制成的 cDNA 文库可直接显示被测材料中基因表达的情况。基因表达,可以用“有无、强弱、关联”来描述,“有无”问题,即哪些种基因在起作用,是基因表达研究的基础和起点。然而,使用 cDNA 文库来显示基因表达的“有无”存在一定问题,这是因为在绝大多数细胞中,基因表达均存在差异。表现在 mRNA 上,就是在任意时刻,细胞内的各种 mRNA 的丰度都存在较大差别。一般,高丰度群占细胞总 mRNA 的 10%~20%,含 10~15 种 mRNA,平均每种占总量的 1%;低丰度群占总 mRNA 的 40%~45%,含 15000~20000 种 mRNA,平均每种占总量的 0.00002%<sup>[1,2]</sup>。这样的丰度分布,反映在使用通用方法<sup>[1,3]</sup>制成的 cDNA 文库,即直接 cDNA 文库(direct cDNA library)上,就是文库中大多数克隆都是重复的、冗余的,信噪比(signal-to-noise ratio)

过低。当然,冗余度可反映表达强度,但是,以这样的文库为平台来进行“有哪些种基因在起作用”这样问题的研究,效率比太低。由此,出现了 cDNA 文库的均一化技术。

cDNA 文库的均一化技术可分为两类:一是寡核苷酸序列指纹(Oligonucleotide sequence signatures, OSSs)均一化技术:使用数据处理软件和寡核苷酸探针直接 cDNA 文库进行处理以实现均一化<sup>[4~6]</sup>;先使用上百种标记的寡核苷酸探针与来自直接 cDNA 文库上万个克隆的 DNA 组成的芯片杂交,通过数据处理软件,就可以得到文库的 cDNA 目录,实现均一化。但是,该类方法工作量大,设备试剂要求高,自动化程度要求高,比较昂贵。二是复性式均一化技术<sup>[7~10]</sup>:基于复性动力学理论,通过变性—复性操作,使所有 cDNA 的丰度趋于一致。这类方法,操作简单,设备试剂要求比较低,均一化程度较好。以下就复性均一化技术的原理、制作方法和应用做一简要介绍。

收稿日期:2001-06-11;修回日期:2001-10-16

基金项目:国家植物基因中心课题(J00-A-009)资助

作者简介:王 强(1972-),男,博士生;专业方向:遗传学。Tel:020-84110296;E-mail:lsbrc15@zsu.edu.cn

## 2 基本原理<sup>[11~14]</sup>

DNA 复性过程属于双分子二级反应。变性的 DNA 分子,随着复性时间  $t$  的增长,单链 DNA 浓度逐渐下降,遵循公式:  $C = C_0 / (1 + kC_0 t)$ 。其中  $C$  为时间  $t$  时单链 DNA 的浓度,  $C_0$  为复性起始时的单链 DNA 浓度,  $k$  为常数。上述公式又可表述为:  $-dC/dt = kC^2$ 。因此,在某一时刻,单链浓度越大,则其消失速度也就越快。如果体系中存在两种变性的 DNA 分子 A 和 B,若  $C_{0A} > C_{0B}$ ,则 A 单链消失的速度将大于 B 单链。随着时间延长,  $C_A$  将趋近于  $C_B$ 。在一个体系中存在多种 DNA 变性分子时情况也是一样:高丰度 DNA 的单链消失速度快于低丰度者,随着时间的延长,体系内各种 DNA 的单链浓度将趋于一致,这就是均一化 cDNA 文库的理论基础。

## 3 方 法

### 3.1 Ko 的方法

Ko 于 1990 年做出第一个均一化文库<sup>[7]</sup>,他使用寡聚 dT 反转录、超声波切断、PCR 直接扩增来制备杂交用 DNA 链。具体做法是:(1)提取材料中的 mRNA,使用带酶切位点的寡聚 dT 为引物,反转录,然后转换为双链 DNA。(2)用超声波处理双链 cDNA,电泳回收 200~400bp 的片段,补平;(3)加入含酶切位点的 Lone 接头。Lone 接头是这样一段双链 DNA:其一端为平端,一端 5' 凸出。这样,每一种 cDNA 片段分子只能以一个方向与接头相连。(4)PCR,回收纯化产物。(5)把这些产物加入杂交液中,变性,复性。(6)以羟基磷灰石色谱处理杂交液,收集其中的单链片段。(7)使用 PCR 扩增收集到的单链。(8)重复(5)~(7)的过程 2 次。(9)终产物克隆进载体,制成文库。为了检验文库的质量,Ko 使用步骤(2)的产物制作了一个直接 cDNA 文库,并自定义一个概念:丰度变异系数(abundance variation)——文库中基因丰度的最大值/最小值。克隆杂交和直接测序显示:丰度变异系数由均一化前的 7000 降至三轮均一化后的 180。

Takahashi 等<sup>[15]</sup>使用本方法建立了覆盖小鼠胚胎发育全程的 cDNA 文库。他们以 1.5、2.5、3.5、4.5 天的胚胎为材料,经三轮均一化制成了含 15000 个克隆的文库。克隆杂交显示丰度变异系数由  $10^6$  降至  $10^2$ 。定量 PCR 检测显示:六轮均一化的效果和三轮的区别不大。

Komiya 等<sup>[16]</sup>对本方法稍做变动,构建了一个大规模原位杂交系统。其改动是:原方法采用超声波处理来获得 200~400bp 的短片段,而他们采用控制反转录进程的方法来达到这一目的。

Inagaki 等<sup>[17]</sup>使用本方法,分别以致病性真菌 *Colletotrichum lagenarium* 的营养菌丝和处于感染发育阶段的分生孢子为材料,各自构建均一化文库。丰度变异系数由 276 降

至 10。通过两文库间的克隆杂交,找到 11 个仅在分生孢子中特异表达的基因。结合基因组文库,对其中 2 个基因的功能进行了预测。

### 3.2 Patanjali 的方法

Patanjali 等<sup>[8]</sup>使用随机引物引发的反转录、片段克隆进载体、PCR 扩增来制备杂交用 DNA 链。具体做法是:使用六碱基的随机引物来进行反转录。双链化的 cDNA 用电泳收集 400~1600bp 的部分,接上普通接头,制成文库。然后用 PCR 从文库中扩出 cDNA 片段,进行一轮变性-复性-羟基磷灰石色谱过程。其不同时间的复性程度以 Southern 印迹测定。克隆杂交和直接测序结果显示:丰度变异系数由均一化前的 10000 降至其后的 60。

Lamerdin 等<sup>[18]</sup>用本方法构建了成人胸腺组织的  $\lambda$ gt10 均一化文库。文库插入片段平均长度为 570bp,对其中 500 个克隆进行测序,333 个产生有价值信息,与公用数据库比对后,注册 298 个 EST,其中 136 个在人染色体上定了位。

### 3.3 Soares 方法

由于担心 PCR 对不同长度的 DNA 分子可能存在歧视,Soares 等<sup>[9]</sup>使用单链环状 DNA 和由它制成的小片段 DNA 作为杂交用 DNA 链。具体做法是:首先,按常规方法制作实验材料的以 phagemid 为载体的直接 cDNA 文库作为起始文库。Phagemid 在辅助噬菌体的帮助下,可产生含插入片段的单链环状拷贝。收集这些单链环,纯化,作为模板,以 DNA 聚合酶、制作直接 cDNA 文库时用于 mRNA 反转录的引物,以及混有 ddNTP(ddTTP 除外)的 dNTP 在载体上的插入片段区域进行延伸,延伸长度控制在  $200 \pm 20$  nts。这样就形成含部分双链的单链环状 DNA 分子。用羟基磷灰石色谱回收纯化这些分子,变性,用 (dT)<sub>25~30</sub> 以及延伸引物封闭杂交分子中的非特异性位点,复性。复性完成后用羟基磷灰石色谱回收纯化纯粹的单链环状分子,随机引发有限延伸后(该步有利于提高转化效率)转入受体菌,制成文库。重复上述过程可提高均一化程度。克隆杂交和测序显示:插入片段平均长度 1.7kb 左右,丰度变异系数由起始文库的  $> 7000$  降至一轮均一化后的 140 及两轮均一化后的 26。

Soares 方法的 Bonaldo 改进版(以下称为 Soares - Bonaldo 方法):

以上述 Soares 等 1994 年的方法<sup>[9]</sup>为基础,Bonaldo 等<sup>[10]</sup>提出两套均一化方案:方案一、首先,从上述的起始 cDNA 文库中提取双链质粒,分成两部分。一部分作为体外转录的模板,得到插入片段的 mRNA;另一部分用辅助噬菌体或者酶 Gene II、Exo III 制备单链环状 DNA。使用这两种制备物(mRNA 和单链环状 DNA)杂交,过羟基磷灰石色谱柱,用洗脱下来的单链环状 DNA 制作文库。或者,把吸附在羟基磷灰石色谱柱上的、含部分双链的单链环状 DNA 也洗脱下来,制成高丰度群 cDNA 文库,提取其质粒,体外转录,制备的 mRNA 与前述的以辅助噬菌体或者酶 Gene II、

Exo III 制备单链环状 DNA 杂交,羟基磷灰石色谱分离出纯粹的单链环状 DNA,制成高丰度 mRNA 丰度特别减弱的均一化文库。

方案二、首先,从直接 cDNA 文库中提取质粒,用酶 Gene II 和 Exo III 处理,得到单链环状 DNA。然后把它与插入片段的单链线状 DNA 杂交,羟基磷灰石色谱处理分离出完全单链的环状 DNA,制成文库。插入片段的单链线状 DNA 的制作方法有两种:一是从双链质粒 DNA 上切下插入片段,用 Exo III 处理成单链;一是用 PCR 以双链 DNA 为模板对插入片段部分扩增,扩增产物变性成为单链。

比较发现:方案二不但更好地保持了起始文库的代表性,而且均一化程度更高<sup>[10]</sup>。

Hiller 等<sup>[19]</sup>1996 年进行了一项大规模 EST 测定,使用 18 个 Soares-Bonaldo 方法构建的均一化文库和 8 个非均一化文库。共测 391154 个克隆,得到 280223 个 EST。这些 EST 中,243616 个来自均一化文库。其中,胎儿肝——脾均一化文库产生 82309 个 EST,而来自 8 个非均一化文库的 EST 总数是 36607。

Milosavljevic 等<sup>[20]</sup>使用寡核苷酸序列指纹法(Oligonucleotide sequence signatures, OSSs)检验了一个用 Soares-Bonaldo 方法构建的人胎儿脑均一化文库,结果显示:均一化处理对 cDNA 中、低丰度群的频率影响不大,而大大降低了高丰度群中 cDNA 分子的频率。

Verdun 等<sup>[21]</sup>使用 Soares-Bonaldo 方法构建了处于 CL Brener 短膜虫期的 *Trypanosoma cruzi* 的均一化 cDNA 文库,随机挑取了 1994 个克隆进行测序,得到 1305 个 EST。

Uremenyi 等<sup>[22]</sup>以 Soares-Bonaldo 方法,以处于短膜虫期的 *Trypanosoma cruzi* 为材料,构建均一化文库,随机挑取 49 个克隆进行测序,得到 41 个 EST。Porcel<sup>[23]</sup>等对该文库做了进一步分析,共测 5034 个克隆,得到 2016 个 EST。丰度变异系数为 107。

### 3.4 cDNA 微珠法

以上三种方法的复性体系中,互补的两条单链的浓度相等,可称为对称杂交。如果双方的浓度不等(可称为不对称杂交),则不遵守  $C = C_0 / (1 + kC_0 t)$  这个公式,而遵守公式:  $R = R_0 e^{-kD_0 t}$ <sup>[12,24]</sup>。其中,R 指互补链中浓度较低的那条链,D 指互补链中浓度较高的那条链。 $R_0$ 、 $D_0$  分别是这两条互补链的起始浓度。在这样的体系中,高丰度 DNA(RNA)种类的复性速度比  $R_0 = D_0$  系统中高丰度的种类复性速度更快<sup>[24]</sup>。基于此原理,Sasaki 等人<sup>[24]</sup>发展出一种半固相杂交制备均一化 cDNA 文库的办法。首先,把样品 mRNA 与表面挂有寡聚 dT 的乳胶微珠混合,反转录,加热变性除去 mRNA,制成外挂 cDNA 的微珠(以下简称 cDNA 微珠)。使用这种微珠,和样品的 mRNA(小量)混合杂交较短的一段时间。然后离心取上清,加入新的或再生的 cDNA 微珠,重复此过程。最后的上清液进行反转录。第二链合成后,转

入  $\lambda$ gt10 载体,制成文库,克隆杂交显示:丰度变异系数从均一化前的 3566 降至其后的 77。

### 3.5 方法间的比较

上述几种方法,都基于复性理论,又各有异同。Ko 的方法和 Soares-Bonaldo 方法都选用 mRNA 的 3' 端非编码序列来做杂交用序列,其依据是这个区域内的序列特异性和代表性能更高<sup>[7,9,25]</sup>。杂交片段的长度,Ko 的方法使用 200~400bp 之间,Soares-Bonaldo 方法为 200bp;而 Patanjali 等<sup>[9]</sup>发现:长度小于 200bp 的双链 DNA 在用羟基磷灰石处理时,会像单链 DNA 一样,被 0.1mol/L 的磷酸盐缓冲液洗脱下来,因此他们采用 400~1600bp 长的 DNA 片段来杂交。然而,这似乎不是绝对的:在 Ko 方法和 Soares 方法所采用的羟基磷灰石色谱方案<sup>[26]</sup>中,洗脱用的磷酸盐缓冲液的合适浓度要在 0.01mol/L~0.4mol/L 之间由实验来确定。

上述几种方法都采用克隆杂交和直接测序的手段来检查文库的均一化程度。在文库质量评估方面,都采用插入片段的平均长度、含插入片段克隆在文库中的频率以及丰度变异系数作为指标。总的来讲,Ko 的方法和 Patanjali 的方法由于采用 PCR,比较简便,而 Soares-Bonaldo 方法比较复杂,不过它可以提供更多的信息。cDNA 微珠法采用了半固相杂交技术,操作方便,易于自动化,也是一种很有前途的方法。

### 3.6 存在的问题<sup>[7~10]</sup>

DNA 复性并不完全遵守二级动力学规律。均一化文库构建过程中发现:某些高丰度 cDNA 在均一化后的降低程度比预期要大;某些高丰度 cDNA 丰度均一化处理后变化不大,如含人 Alu 重复的基因,或含鼠的 b1、b2 序列的基因。

Soares 等认为 PCR 会在文库构建中引入因模板长短不一而造成的歧视性效应<sup>[9]</sup>,似乎根据不足。因为(1)Ko 的方法<sup>[7]</sup>中 PCR 模板长度为 200~400bp,存在一定程度的模板长度多态性,但是,应用 Ko 方法的文献中均无有关此效应的报道<sup>[15~17]</sup>;(2)Soares-Bonaldo 方法<sup>[9]</sup>(方案二),以起始文库质粒中插入片段两侧序列为引物,使用 PCR 来扩增制备杂交所需 DNA 中的一种。由于该方法的插入片段是使用酶切处理双链 cDNA 后,未经凝胶分部而是直接克隆进载体制成的,本身也存在长度多态性,因此如果 Soares 所谓的“PCR 长度依赖的歧视效应(length-dependent differential PCR amplification)”在使用 PCR 来构建均一化 cDNA 文库过程中比较显著的话,在该研究(Soares-Bonaldo 方法<sup>[9]</sup>(方案二))也应该显示出来。然而,该研究并未提及此效应。因此,该效应在均一化文库制作过程中似可忽略。

## 4 展 望

由人类基因组计划所带动的先进的测序技术,已经获得了越来越多的基因组序列数据,使得从全基因组角度来综合看待生命现象成为可能。“全基因组学”的首要任务,就是要

获得全部基因的清单(gene index<sup>[27]</sup>, gene catalogue<sup>[15]</sup>, gene inventories<sup>[28]</sup>或 gene periodic<sup>[29]</sup>)。然而,由于信噪比(即外显子和内含子序列长度之比)等问题,使用这些数据来直接寻找新的基因尚存在较大困难,对于高等哺乳动物来讲尤其如此<sup>[27]</sup>。而使用比较基因组学方法来寻找新基因的手段又有很大的局限性。目前多采用序列数据分析和其他方法获得的基因信息(如 EST、同源性比较等)相结合的方法来进行。cDNA 文库均一化技术的不断完善,无疑会对这项工作起巨大的推动作用。

## 参 考 文 献 (References):

- [1] Bishop J O, Morton J G, Rosbash M, Richardson M. Three abundance classes in HELA cell messenger RNA [J]. *Nature*, 1974, 250:199~204.
- [2] Davison E H, Britten R J. Regulation of gene expression: possible role of repetitive sequences [J]. *Science*, 1979, 204:1052~1059.
- [3] J. 萨姆布鲁克, E. F. 弗里奇, T. 曼阿蒂斯(金冬雁等译). 分子克隆实验指南[M]. 北京: 科学出版社, 第二版, 1999, 396~456.
- [4] Poustka A J, Herwig R, Krause A, *et al.* Toward the gene catalogue of sea urchin development: the construction and analysis of an unfertilized egg cDNA library highly normalized by oligonucleotide fingerprinting [J]. *Genomics*, 1999, 59:122~133.
- [5] Eickhoff H, Schuchhardt J, Ivanov I, *et al.* Tissue gene expression analysis using arrayed normalized cDNA library [J]. *Genome research*, 2000, 10:1230~1240.
- [6] Meier-ewert S, Lange J, Gerst H, *et al.* comparative gene expression by oligonucleotide fingerprinting [J]. *Nucleic acids research*, 1998, 26:2216~2223.
- [7] Ko M S H. An 'equalized cDNA library' by the reassociation of short double-stranded cDNAs [J]. *Nucleic acids research*, 1990, 18(19):5705~5711.
- [8] Patanjali S R, Parimoo S, Weissman S M. Construction of a uniform-abundance (normalized) cDNA library [J]. *Proc Natl Acad Sci USA*, 1991, 88:1943~1947.
- [9] Soares M B, Bonaldo M D F, Jelene P, *et al.* construction and characterization of a normalized cDNA library [J]. *Proc Natl Acad Sci USA*, 1994, 91:9228~9232.
- [10] Bonaldo M D F, Lennon G, Soares M B. Normalization and subtraction: two approaches to facilitate gene discovery [J]. *Genome research*, 1996, 6:791~806.
- [11] 孙乃恩, 孙东旭, 朱德熙. 分子遗传学 [M]. 南京: 南京大学出版社, 1990:11~12.
- [12] Britten R J, Graham D E, Neufeld B R. Analysis of repeating DNA sequences by reassociation [J]. *Methods Enzymol*. 1974, 29(0):363~418.
- [13] Britten R J, Davidson E H. Hybridization strategy. In: Harres B D, Higgins S J. *Nucleic Acid Hybridization— A Practical Approach* [M]. IRL Press, Oxford, 1985, 3~15.
- [14] Young B D, Anderson M L M. Quantitative analysis of solution hybridization. In Harres B D, Higgins S J. *Nucleic Acid Hybridization— A Practical Approach* [B]. IRL Press, Oxford, 1985, 47~71.
- [15] Takahashi N, Ko M S H. Toward a whole cDNA catalog: construction of an equalized cDNA library from mouse embryos [J]. *Genomics*, 1994, 202~210.
- [16] Komiya T, Tanigawa Y, Hirohashi S. A large-scale in situ hybridization system using an equalized cDNA library [J]. *Analytic Biochemistry*, 1997, 254:23~30.
- [17] Inagaki A, Takano Y, Kubo Y, *et al.* Construction of an equalized cDNA library from *Colletotrichum lagenarium* and its application to the isolation of differentially expressed genes [J]. *Can. J. Microbiol*. 2000, 46:150~158.
- [18] Lamerdin J E, Athwal P S, Kansara M S, *et al.* Chromosomal localization and expressed sequence tag generation of clones from a normalized human adult thymus cDNA library [J]. *Genome Research*, 1995, 5:359~367.
- [19] Hillier L, Lennon G, Becker M, *et al.* Generation and analysis of 280,000 human expressed sequence tags [J]. *Genome Research*, 1996, 6:807~828.
- [20] Milosavljevic A, Zeremski M, Strezoska Z, *et al.* Discovering distinct genes represented in 29,570 clones from infant brain cDNA libraries by applying sequencing by hybridization methodology [J]. *Genome Research*, 1996, 6:132~141.
- [21] Verdun R E, Paolo N D, Urmenyi T P, *et al.* Gene Discovery through expressed sequence tag sequencing in *trypanosoma cruzi* [J]. *Infection and Immunity*, 1998, 5393~5398.
- [22] Urmenyi T P, Bonaldo M D F, Soares M B, *et al.* Construction of a normalized cDNA library for the *trypanosoma cruzi* genome project [J]. *J Eukaryot Microbiol*, 1999, 46(5):542~544.
- [23] Porcel B M, Tran A, Tammi M, *et al.* Gene survey of the pathogenic protozoan *Trypanosoma cruzi* [J]. *Genome Research*, 2000, 10:1103~1107.
- [24] Sasaki Y F, Ayusawa D, Oishi M. Construction of a normalized cDNA library by introduction of a semi-solid mRNA—cDNA hybridization system [J]. *Nucleic Acids Research*, 1994, 22(6):987~992.
- [25] Hawkins J D. A survey on intron and exon lengths [J]. *Nucleic Acids Res*, 1988, 16:9893~9908.
- [26] Sambrook J, Fritsch E F, Maniatis T. *Molecular Cloning: A Laboratory Manual* (second edition) [B]. Cold Spring Harbor University Press, Cold Spring Harbor, 1989.
- [27] Lander E S, Linton L M, Birren B, *et al.* Initial sequencing and analysis of the human genome [J]. *Nature*, 2001, 409:860~921.
- [28] Venter J C, Adams M D, Myers E W, *et al.* The sequence of the human genome [J]. *Science*, 2001, 291:1304~1350.
- [29] Lander E S. *The New Genomics: Global Views of Biology* [J]. *Science*, 1996, 274:536~539.