

# 基因芯片技术在微生物学研究中的应用<sup>\*</sup>

饶志明<sup>\*\*</sup> 张荣珍 王正祥 方慧英 诸葛健

(江南大学工业微生物研究中心 无锡 514036)

**摘要** 近年来,基因芯片技术的诞生使得在一个实验中就可以同时对成千上万个基因进行转录水平的表达和 DNA 同源性分析成为可能。该项技术已被应用于揭示许多微生物体的转录表达和基因组的差异,随着越来越多的微生物基因组全序列测定的完成,基因芯片正逐渐成为许多微生物学研究领域中的一项常规技术。归纳了该技术在微生物生理、致病性、流行病学、生态、进化、代谢工程及发酵优化等研究中的应用。

**关键词** 基因芯片 基因表达谱 基因组 微生物

基因芯片是 20 世纪 90 年代发展起来的一项前沿生物技术。随着生物技术的迅猛发展,20 世纪人类最宏伟的计划之一——人类基因组计划的启动和工作框架图已基本完成<sup>[1]</sup>,以及拟南芥<sup>[2]</sup>等模式生物的全基因组序列的测定,最近我国科学家又独立完成并公布了籼稻的全基因组框架序列<sup>[3]</sup>。模式生物的结构基因组计划将由此进入后基因组——功能基因组时代,研究的重点将由发现基因转向探索基因的功能。人们将逐步阐明从基因到蛋白再到生命现象这一过程的奥秘。如何大规模研究基因的功能,特别是基因相互作用和调控关系,迫切需要一种新的方法能以大规模、高通量的方式进行成千上万个基因在各种生理状态下表达状况的研究<sup>[4]</sup>。传统的 Northern 印迹杂交或点杂交方法,以及以电泳为基础的基因表达、序列测定、突变和多态性分析等研究方法显然不能适应上述的要求。基因芯片技术由此应运而生。

## 1 基因芯片的原理及其产生背景

分子杂交实验一直是遗传学研究的一个重要和有效的手段,它对于揭示基因和一些重要的生命现象之间的内在联系具有重要的意义。基因芯片的概念可追溯到 Southern 印迹杂交技术,即 DNA-DNA 之间通过碱基配对机制形成互补链。随后又

发展出 Northern 印迹和点杂交技术。这几种技术都是将核酸样品固定在滤膜上。为了提高点样密度和检测灵敏度,降低探针用量,以玻璃、硅等材料为载体的基因芯片技术逐步得到了发展。

基因芯片是指采用原位合成或直接点样的方法将 DNA 片段或寡核苷酸片段排列在滤膜、硅片、玻璃等介质上形成微矩阵,待检样品用同位素或荧光分子标记后,与微矩阵杂交,通过扫描及计算机分析即可获得样品中大量的基因序列及表达信息,以达到快速、高效、高通量地分析生物信息的目的。

美国 Affymetrix 公司 20 世纪 80 年代末至 90 年代初,率先开展了这方面的研究。1991 年,该公司生产了世界上第一块寡核苷酸基因芯片。同时,探针的荧光标记、激光共聚焦扫描和计算机分析等技术也随之发展。1995 年,第一块以玻璃为载体的基因微矩阵芯片在美国 Stanford 大学诞生,并首先被应用于人类基因组的研究<sup>[5]</sup>,最近,基因芯片也被用来检测模式植物如水稻的基因表达谱的改变<sup>[6]</sup>,标志着基因芯片技术步入了广泛研究和应用的时期。

尽管基因芯片的基本技术一样,但当该项技术应用于原核生物和真核生物时却有很大的区别。通常,在细菌芯片的试验中,被标记的探针是总 RNA,而在真核生物中,mRNA 却被用作探针来进行标记。这是由于细菌 mRNA 的半衰期要比真核生物短得多,使得在细菌中分离完整的 mRNA 变得十分艰难。更为重要的是,基因芯片在微生物中具有许多特殊的应用。下面即对基因芯片在微生物

收稿日期:2003-04-01 修回日期:2002-05-28

<sup>\*</sup>江南大学人才引进专项基金资助项目(101000-21053500)

<sup>\*\*</sup>电子信箱:raozm@sina.com

学研究中的应用进行综述。

## 2 基因芯片在微生物学研究中的应用

自从基因芯片技术诞生之后,便得到了广泛的应用,尤其是在揭示许多微生物体的转录表达和基因组的差异方面。

### 2.1 全基因组转录水平的检测

细胞的功能是从基因的表达开始的,基因芯片可以用来检测微生物不同的发育阶段或不同的生理状态下基因转录水平的差异。例如: Selinger 等<sup>[7]</sup>对大肠杆菌处于静止期和对数生长期的基因表达谱进行了比较,发现共有 1529 个基因的表达水平发生了变化。其中许多是以前所未发现的生长期调节相关的基因。Stephens<sup>[8]</sup>对 *Caulobacter crescentus* 参与细胞周期调节的全基因组转录水平和蛋白质稳定性进行了研究,结果令人吃惊的表明一大部分基因和蛋白受到细胞的生长和分裂的影响。CtrA 是一个信号传导家族中成员之一,却被发现其直接或间接的参与了细胞周期调节基因中的 26%。另一项基因芯片研究还被应用于检测枯草芽孢杆菌在厌氧条件下的转录水平的变化,发现超过 100 个基因的转录表达受到供氧限制条件的影响,这些基因的功能包括参与能量代谢、离子吸收、抗生素合成、碳代谢和逆境胁迫等生理代谢过程<sup>[9]</sup>。

除了应用于比较不同样品中间的 mRNA 表达水平之外,基因芯片还可以应用于同一样品中某一 mRNA 分子相对于其他 mRNA 分子的相对表达量的研究。这种研究有助于揭示某一代谢途径中某个基因在特定条件下是否表达及其转录水平的相对高低,它为探明细胞的代谢活力提供了有力的线索。Wei 等<sup>[10]</sup>在大肠杆菌中曾对一些 mRNA 转录水平进行了半定量的研究,有趣的是许多结果与大肠杆菌中已知的生理代谢非常一致。但研究中也发现用基因组 DNA 作为参照的一个主要问题是,基因组 DNA 与 mRNA 标记效率存在差异。因为不同的 mRNA 具有稳定的二级结构,所以要对这些 mRNA 进行一致的标记就比较困难,这导致对芯片中 mRNA 进行精确定量也较为困难。所以应用基因芯片方法获得的许多结果也只能简单的反映转录水平的变化,但对一些假设可以提供有益的信息。

全基因组转录水平的检测还有助于揭示微生物

物在不同生长条件、发育阶段或药物处理情况下的基因表达模式。这些特定表达模式在新药品的发现和发酵过程优化等领域中具有很重要的应用价值。而一个可以分析、贮存和比较特异转录本的软件的开发则将使该项研究变得更加便利。

### 2.2 定义调节子

调节子通常指的是被一个调节因子控制的一套基因。要定义一个新的调节子,通常是从一些调控区域发生突变并得到表达,再从突变的细胞中提取 RNA,通过基因芯片进行确认。Zimmer 等<sup>[11]</sup>用一个基因芯片将来源于一个超水平表达的 NtrC 激活的突变菌株和一个相应的野生型菌株的 RNA 水平进行比较,发现了大肠杆菌中的调节子 NtrC,研究还表明,大肠杆菌基因组中大约有 2% 的基因的表达受到该调节子的控制。

一个潜在的问题是基因突变或超表达之后的多重效应,这可能会导致一些与特定调节子没有直接关联的基因的转录水平下降的现象。因而难以发现这些低表达水平的基因调节或受调节子微弱影响的基因。在这种情况下就有必要结合基因组学和传统方法来定义新的调节子。

### 2.3 描绘操纵子结构

基因芯片有助于阐明那些位于同一区域并具有同一转录方向的一套基因,如果它们具有相同的转录模式,这表明了操纵子的存在。如果它们在不同的条件下具有不同的表达模式,这提示该操纵子中可能有多个启动子存在,当然也有可能是在同一区域存在多个简单的操纵子。尽管基因芯片并不一定是发现操纵子的最佳方法,但该技术的高通量特性对操纵子的阐明无疑具有重要的验证作用。最近一项利用大肠杆菌的寡核苷酸芯片的研究就成功地获得了大量的关于操纵子转录结构的起始和终止的信息<sup>[7]</sup>。

### 2.4 分析基因组的未知区域

目前,虽然许多微生物基因组测序工作已经完成,但在基因组中仍然有一大部分基因序列的功能没有阐明,主要是因为这些序列在数据库中找不到与之同源的信息或者这些基因序列被突变后却没有明显的表现型。而基因芯片技术则可以从几个不同的方面来加速对未知基因的功能分析。例如,(1)分析功能不祥的调控区域的调节子就有可能导致其生理功能的发现;(2)对一些未知基因进行突变后,就有可能引起与之相关的一些功能已知基因

表达水平的变化,而基因芯片技术则能够很简单的检测这些功能已知基因的转录变化;(3)如果未知基因和已知基因在相同的生长条件或处于同一发育阶段具有相似的表达模式,则提示这些未知基因可能具有与已知基因相似的功能,至少有助于揭示该未知基因在这种特定条件的功能。当然基因芯片技术与其他基因组学方法的结合将大大加速我们对微生物基因组未知功能区域的功能的阐明。

## 2.5 阐明 DNA—蛋白的互作

许多蛋白质可以结合到基因组染色体的特定区域,对调节细胞的功能和维护基因组结构的稳定具有重要作用。Ren 等<sup>[12]</sup>应用基因芯片对酵母中 DNA 结合蛋白的功能及其基因组的结合区域进行了研究,该方法结合了应用包含酵母全部基因组信息的基因芯片和经过修正的染色质免疫沉淀技术,结果表明该技术能够发现酵母细胞在体内被转录激活子直接控制的基因。

## 2.6 比较基因组学和基因分型研究

全基因组芯片的杂交可以用来检测其它微生物是否含有某一特定 DNA 序列的存在,并在全基因组水平上比较不同种的遗传差异,目前已经有几篇报道将基因芯片应用于微生物基因组差异比较的研究,这些研究对于加速了解近缘种间遗传差异,发现致病因子,阐明分子进化,开展分子诊断以及发展新疫苗等都有极大的帮助<sup>[13]</sup>。

基因芯片技术还可应用于同一菌种经过长期的菌种改良后其遗传成分的改变。Riehle 等<sup>[14]</sup>利用基因芯片技术对大肠杆菌在经过连续 200 代 41.1 高温逆境胁迫条件下的 DNA 片段的重复和缺失的全基因组水平进行了检测,共发现了 5 个 DNA 片段重复和缺失事件,有力的证明了基因重复在微生物适应逆境胁迫中具有重要作用。

## 2.7 微生物病原菌中致病因子的检测

许多致病相关基因受到特定条件的调节,检测候选致病因子的策略之一就是应用基因芯片技术对致病菌在与寄主作用的过程进行全基因组水平的检测,当然,比较基因组研究亦可应用于病原菌中致病因子的检测。Israel 等<sup>[15]</sup>在一项应用基因芯片技术对 *Helicobacter pylori* 菌种的比较基因组研究中,发现了一些致病候选基因,并对两个具有显著不同致病性的 *H. pylori* 菌株间的基因差异进行研究,结果表明该两个菌株调节上皮细胞对炎症反应的能力依赖于一个完整的致病岛的存在。

## 2.8 寄主对病原菌的反应

Belcher 等<sup>[16]</sup>应用高密度基因芯片对 *Bordetella pertussis* 与寄主人支气管上皮细胞系 (BEAS-2B) 互作的转录本进行了研究,发现对该病菌的早期应答受编码原浆移动, DNA 结合蛋白和 NF $\kappa$ B 调节因子的控制。另一项应用基因芯片对寄主肺气肿细胞系 A549 受 *Pseudomonas aeruginosa* 感染后差异表达基因的研究中,发现许多具有不同细胞功能的基因的表达受到改变,其中一个基因编码转录因子干扰素调节因子 1<sup>[17]</sup>。上述两个研究显示了基因芯片技术在寄主受病原菌侵染后,寄主基因组转录水平的变化研究中应用的巨大潜力。当然将该技术与寄主受病原菌侵染后的常规分析手段相结合,将有助于进一步阐明寄主与病原菌互作的分子机理。

## 2.9 揭示药物、抑制剂及有害成分的基因作用模式

特定细胞代谢过程的抑制将产生反馈调控,并导致基因表达模式的改变。应用基因芯片技术将有助于揭示药物、抑制剂及有害成分的基因作用模式。Wilson 等<sup>[18]</sup>应用基因芯片技术对肺结核菌受抗肺结核药物异烟肼处理的应答基因进行了研究,发现异烟肼可以诱导一些编码具有与该药品作用方式相关生理功能蛋白的表达,例如,编码二型脂肪酸合成酶的 5 个基因构成的操纵子簇以及编码海藻糖 dimycolyl 转移酶基因。表明该项技术对发现新药靶并开发相应新药也大有帮助。

药物除了会引起与之作用的相关基因的表达水平发生改变之外,还会诱导该药物毒副作用相关的基因的转录改变。药物的另一个效果是产生抗药性,而开展基因芯片的研究有助于设计开发一些毒副作用较小,而病原菌又难以产生抗性的药物。

药物的每一类成分有特定的表达模式,建立这些表达模式的数据库将有助于分析那些尚未了解其毒副作用的成分。

## 2.10 微生物进化与流行病学的分析

基因芯片可以用来对微生物特定种及其相关种的分子进化和基因转录本的多样性进行分析,它还有助于阐明微生物多样性和流行病学的分子基础。应用高密度寡核苷酸芯片对临床上已经详细阐明过的 19 个 DNA 片段缺失突变的种进行了全基因组水平的比较,结果表明,一些存在于祖先基因中缺失的片段,似乎对保存下来的种不再那么重要,而那些从未缺失过的基因就构成了微生物的最小基因组。似乎随着基因组缺失程度的增加,病原

菌导致肺气肿的概率减少,提示病原菌突变的积累也可能会降低其致病性<sup>[19]</sup>。

### 2.11 基因芯片可以作为诊断手段

基因芯片可以解决医学和环境微生物种的鉴定,它通过独立的 16S rRNA 序列、抗生素相关的标记以及致病性相关的 DNA 区域来确定他们的种类。Troesch 等<sup>[20]</sup>利用基因芯片对 70 株来自 27 个不同种的利福平抗性菌株进行了鉴定,发现其中的 26 个种的 68 个菌株得到了确认。

### 2.12 途径工程和过程优化

传统的生物催化优化方法包括随机筛选、突变及工程改良。尽管这些方法仍然十分有效,但基因组学技术的渗入将使我们从分子水平了解这些生理过程发生变化的根本原因。基因芯片对途径工程和过程优化的帮助主要体现在以下几个方面:(1)基因芯片可以对微生物在不同的生长条件下、不同代谢途径中的基因表达调控进行检测;(2)细胞在发酵过程中的生理状态可以通过全基因组转录本而进行检测;(3)基因芯片有助于发现一些代谢过程中协同表达的基因;(4)基因芯片还可以用来比较野生型和改良型菌株之间的遗传信息和转录本的差异;(5)芯片的数据还可以整合到描述细胞代谢的数学模型中<sup>[21~22]</sup>。

目前将基因芯片用于研究途径工程和生物加工还处于初期阶段,Gill 等<sup>[21]</sup>利用一个包含编码中心代谢、生物合成关键酶、调节因子以及胁迫反应相关的基因芯片,对大肠杆菌超高量表达蛋白的代谢反应进行了分析。Tao 等<sup>[22]</sup>应用基因芯片技术对大肠杆菌的亲本菌株以及工程菌株 KO11 中与木糖代谢相关的 30 个基因进行了表达分析,发现参与木糖代谢的基因表达水平的提高有可能是 KO11 的生长速率和糖酵解水平提高的原因。

## 3 展 望

随着许多微生物全基因序列测定的完成,基因芯片技术已经变成研究全基因组转录本和遗传分型的强大工具。由于基因芯片只能对转录本的丰度和基因组上 DNA 片段存在与否进行检测,因而在没有其它实验证据下,对芯片数据要进行详尽的阐述比较困难。尤其是当一些代谢过程还没有被研究清楚的时候更是如此。此外对于一些基因表达水平的改变到底是由于突变,或者是由于处在不同生长条件下,还是由于其他相关基因表达变化尚

难以明了。因而只有将这项技术与其他高通量基因组学方法以及生理生化方法和遗传方法相结合才能发挥更大的作用。这些方法的结合已经被成功的应用于构建检测和优化酵母细胞代谢途径模型<sup>[23]</sup>。尽管应用基因芯片对两个样本的转录本进行分析的结果相对可靠,但是芯片技术本身并不能直接定量 RNA 实际丰度。因而有必要进一步发展在同一个样本中特定 mRNA 丰度的检测技术。

目前,寡核苷酸芯片和基因芯片都可以应用于微生物的研究,但将来哪一种芯片应用更加广泛将取决于自动性能、价格、可行性和实验目的。例如短链寡核苷酸芯片就不太适合于亲缘关系较远的微生物的比较基因组学分析。

值得一提的是基因芯片在微生物系统进化和菌种鉴定分析中的作用,因为 16S rRNA 分型的方法并不能对许多微生物进行很好的分类,而全基因组芯片却能弥补这一缺陷,可以预见将来会有这样一种芯片出现,这种芯片上具有许多代表性微生物种的 16S rRNA 序列、23S rRNA 序列和许多具有重要功能的独立基因。显然,这种芯片将在食品、医学、环境和农业中得到广泛应用。

随着基因芯片的广泛应用和大量基因序列数据的产生,构建数据库并将这些基因数据进行详尽的功能注释将是该技术发展的新领域。而这些包括 DNA 序列和芯片结果的基因组数据,必将进一步大大推动基因组学研究的迅猛发展。

## 参考文献

- [1] International Human Genome Sequencing Consortium. Analysis and sequencing of the human genome [J]. Nature, 2001, 409(6822): 860~921
- [2] The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana* [J]. Nature, 2000, 408: 796~815
- [3] Jun Yu, Songnian Hu, Jun Wang, et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *Indica*) [J]. Science, 2002, 296: 79~91
- [4] Shoemaker DD, Schadt EE, Armour CD, et al. Experimental annotation of the human genome using micarray technology [J]. Nature, 2001, 409 (6822): 922~927
- [5] Schena M, Shalon D, Davis R, et al. Quantitative monitoring of gene expression patterns with a complementary DNA microarray [J]. Science, 1995, 270 (5235): 467~470
- [6] 饶志明,董海涛,庄杰云,等. 水稻抗稻瘟病近等基因系的 cDNA 微阵列分析[J]. 遗传学报, 2002, 29(10): 887~893

- [ 7 ] Selinger DW, Cheung KJ, Mei R, et al. RNA expression analysis using a 30-base pair resolution *Escherichia coli* genome array [J]. Nat Biotechnol, 2000, 18:1262 ~ 1268
- [ 8 ] Stephens C. Bacterial cell cycle ; seeing the big picture with microarrays [J]. Curr. Boil, 2001, 11:R222 ~ R225
- [ 9 ] Yu RW, Tao W, Bedzyk I, et al. Global gene expression profile of *Bacillus subtilis* grown under anaerobic conditions [J]. J Bacteriol, 2000, 182:4458 ~ 4465
- [10] Wei Y, Lee JM, Richmond C, et al. High-density microarray-mediated gene expression profiling of *Escherichia coli* [J]. J Bacteriol, 2001, 183:545 ~ 556
- [11] Zimmer DP, Soupene E, Lee HL, et al. Nitrogen regulatory protein C controlled genes of *Escherichia coli* scavenging as a defense against nitrogen limitation [J]. Proc Natl Acad Sci USA, 2000, 97:14674 ~ 14679
- [12] Ren B, Robert F, Wyrick JJ, et al. Genome-wide location and function of DNA-binding proteins [J]. Science, 2000, 290:2306 ~ 2309
- [13] Behr MA, Wilson MA, Gill WP, et al. Comparative genomics of BCG vaccines responses of respiratory epithelial cells to *Bordetella pertussis* reveal host defensive and pathogen counter-defensive strategies [J]. Proc Natl Acad Sci USA, 1999, 97:13847 ~ 13852
- [14] Riehle MM, Bennett AF, Long AD. Genetic architecture of thermal adaptation in *Escherichia coli* [J]. Proc Natl Acad Sci USA, 2001, 98:525 ~ 530
- [15] Israel DA, Salama N, Arnold CN, et al. *Helicobacter pylori* strain-specific differences in genetic content, identified by microarray, influence host inflammatory responses [J]. J Clin Invest, 2001, 107:611 ~ 620
- [16] Belcher CE, Drenkow J, Kehoe B, et al. The transcriptional responses of respiratory epithelial cells to *Bordetella pertussis* reveal host defensive and pathogen counter-defensive strategies [J]. Proc Natl Acad Sci USA, 2000, 97:13847 ~ 13852
- [17] Ichilawa JK, Norris A, Banger MG, et al. Interaction of *Pseudomonas aeruginosa* with epithelial cells: identification of differentially regulated genes by expression microarray analysis of human cDNAs [J]. Proc Natl Acad Sci USA, 2000, 97:9659 ~ 9664
- [18] Wilson M, Derisi J, Kirsrensens HH, et al. Exploring drug-induced alterations in gene expression in *Mycobacterium tuberculosis* by microarray hybridization [J]. Proc Natl Acad Sci USA, 1999, 96:12833 ~ 12838
- [19] Kato-Maeda M, Rhee JT, Gingeras TR, et al. Comparing genomes within the species *Mycobacterium tuberculosis* [J]. Genome Res, 2001, 11:547 ~ 554
- [20] Tioesch A, Nguyen H, Miyada CG, et al. Mycobacterium species identification and rifampin resistance testing with high-density DNA probe arrays [J]. J Clin Microbiol, 1999, 37:49 ~ 55
- [21] Gill RT, Delisa MP, Valdes JJ, et al. Genomic analysis of high-cell density recombinant *Escherichia coli* fermentation and cell condition for improved recombinant protein yield [J]. Biotechnol Bioeng, 2001, 72:85 ~ 95
- [22] Tao H, Gonzalez R, Martinez A, et al. Engineering a homoethanol pathway in *Escherichia coli* increased glycolytic flux and levels of expression of glycolytic genes during xylose fermentation [J]. J Bacteriol, 2001, 183:2979 ~ 2988
- [23] Ideker T, Thorsson V, Raniish JA, et al. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network [J]. Science, 2001, 292:929 ~ 934

## Applications of Gene Chip in Microbial Systems

Rao Zhiming Zhang Rongzhen Wang Zhengxiang Fang Huiying Zhuge Jian

(Research Centre of Industrial Microorganisms Southern Yangtze University Wuxi 214036)

**Abstract** Gene chip technology allows a parallel analysis of RNA abundance and DNA homology for thousands of genes in a single experiment. Over the past years, this powerful technology has been used to explore transcriptional profiles and genome differences for a variety of microorganisms, greatly facilitating our understanding of microbial metabolism. With the increasing availability of complete microbial genomes, gene chips are becoming a common tool in many areas of microbial research, including microbial physiology, pathogenesis, epidemiology, ecology, phylogeny, pathway engineering and fermentation optimization.

**Key words** Gene chip Gene expression profiling Genome Microbial systems