文章编号: 1007-7847(2000) 02-0106-06

DNA 芯片技术与基因表达研究*

祝 骥[1,2], 马文丽[1], 姚汝华[2]
(1.第一军医大学 分子生物学研究所,中国广东 广州 510515;
2.华南理工大学 生物工程研究中心,中国广东 广州 510460)

摘 要: 随着基因组计划的顺利实施,大量的生物信息被解析,基因表达及基因功能的研究将成为生命科学研究的热点. DNA 芯片技术是近年来出现的分子生物学与微电子技术相结合的最新 DNA 分析检测技术. 该技术将在生命科学与信息科学之间架起一道桥梁,因而成为后基因组时代基因功能分析的最重要的技术之一. 目前 DNA 芯片技术已在基因表达等研究中得到广泛的应用.

关键词: DNA 芯片; 基因表达; 生物信息学
中图分类号: Q786 文献标识码: A

DNA Chip Technology and the Studies of Gene Expression

ZHU Ji[1,2], MA Wen-li[1], YAO Ru-hua[2]
(1. Institute of Molecular Biology, First Military Medical University, Guangzhou 510515, Guangdong, China; 2. Bio-science & Engineering Research Center, South China University of Technology, Guangzhou 510640, Guangdong, China)

Abstract: With the implementing of Human Genomic Project and analyzing of a large amount of bioinformation, the studies of gene expression and gene function will become the hot spot of life sciences. DNA chip technology is the latest DNA analysis and examining technology which combined molecular biology with microelectronics. This technology will build up a bridge between bio-science and information science, and become one of the most important technologies for gene function analysis in the postgenomics for its treatments of parallel, large quantity, miniature and automation. DNA chip technology has been widely applied in the gene expression now.

Key words: DNA chips; gene expression; bioinformatics

研究基因表达的最好方法就是分析不同组织、不同发育阶段以及不同生理状况机体中

* 收稿日期: 2000-03-07
基金项目: 国家自然科学基金项目(39880032)
作者简介: 祝骥(1967-), 男, 湖北人, 第一军医大学助理研究员, 博士研究生, Tel: 020-8514836, E-mail: Zhuji@scib.ac.cn; 马文丽(1964-), 女, 第一军医大学教授, 博士生导师; 姚汝华(1933-), 男, 华南理工大学教授, 博士生导师.

© 1994-2011 China Academic Journal Electronic Publishing House. All rights reserved. http://

基因表达的差异, 但常规方法一次只能检测一个或少数几个基因. 然而由一个受精卵发育成生物体并保持个体的正常生存, 需要成千上万个基因, 其中每个基因的表达均有精确的时空性. 人类基因组含有 5 万个至 10 万个基因, 基因组 DNA 的全部 30 亿个核苷酸预期在 2005 年被确定[1]. 随着 DNA 测序技术的发展和几种同时监测大量基因表达的新技术出现, 人类基因组 DNA 序列分析可能很快完成, 并由此产生了生物信息学. 基因组的碱基序列仅仅代表了生物信息学复杂性的一个层次, 而其中所包含的时空表达信息则代表了更高的层次. 要揭示所有基因转录表达层次上的信息, 就必须有能高效快速且能同时测定基因组成千上万个基因活动的新方法. DNA 芯片技术应运而生, 该技术突出的特点就在于它能高度并行性、多样性、微型化和自动化进行 DNA 分析[2~5]. 这一技术允许研究者同时测定成千上万个基因的作用方式, 几周内获得的信息如用其他方法要几年才能得到. 生物芯片技术已被公认为将会给下个世纪的生命科学研究带来一场革命, 并已为各国学术界和工业界所瞩目.

1　DNA 芯片技术

DNA 芯片又称为 DNA 微集阵列[6], 因微集阵列的制备过程中采用了硅芯片, 故称为 DNA 芯片或基因芯片(gene chip), 其技术具体实现可分为 3 个步骤: 1) DNA 阵列的集成: 即以大量 DNA 特殊片段为探针, 采用物理及化学的方式, 将其固化于载体的表面, 形成致密、有序的 DNA 分子点阵; 2) 样品的制备与分子杂交: 利用常规方法, 纯化 DNA 或 RNA 待测样品, 并以荧光予以标记. 标记后的样品, 与基因芯片进行分子杂交. 由于在较小的面积上集成了大量的探针分子, 因而杂交效率大为提高; 3) 杂交信号的检测与分析: 经过杂交, 与探针互补的样品与之结合, 从而呈现阳性荧光信号进而被检测. 大量信号的并行采集, 需通过激光扫描仪, 并将采集的信号传送至计算机系统进行处理, 最后使样品中的基因成分被鉴定. 3 个步骤中, 基因微集阵列的集成是技术的关键[7], 它包括阵列的制作技术及探针的设计两部分.

DNA 阵列的集成有两种不同的技术路线. 一种是采用光蚀刻技术(Photolithography)与固相化学相结合, 以化学合成的方法, 使 DNA 寡聚片段固化于硅芯片表面[8]. 采用这种方法制成的基因阵列, 其优点是集成度高(可达每平方厘米 10～40 万阵列), 要合成的序列信息可以直接从数据库得到, 不用自己去克隆和测序. 缺点是合成的分子只是寡聚 DNA 片段, 一般为 20～25 bp, 因而整体的遗传信息的总集成量会受到限制, 而且每个芯片要用大量的掩模(mask), 杂交及检测条件要求亦较高, 杂交后的信号需经过复杂的计算机分析处理过程, 虽然已接近产品阶段, 但整个系统的成本仍很高. 利用化学合成制作 DNA 阵列是美国 Affymetrix 公司的专利[9], 由于该技术将 DNA 分子合成于硅芯片的表面, 因而又被特称为 DNA 芯片. DNA 阵列集成的另一方式即微集阵列技术(DNA Microarray), 则采用自动化快速打印, 将纯化的 DNA 样品打印至常规纤维素膜或玻片的表面, 形成致密 DNA 微集阵列[10]. 采用这种方式制作的基因阵列, 集成度可达每平方厘米 2 万个 DNA 片段, 但由于对打印的基因探针无特殊要求, 可以用较长的基因片段甚至重组 DNA 的克隆, 因此, 实际集成度更高, 杂交与信号检测过程均较为简化.

2　DNA 芯片技术在基因表达研究中的应用

随着人类基因组计划的顺利进行, 对人的基因表达谱的研究, 疾病发生机理的分子水

© 1994-2011 China Academic Journal Electronic Publishing House. All rights reserved.　http://

平上的基因机制和预测突变发生的研究逐渐受到人们重视. 为了能够同时对多个可能的突变进行搜索、加快功能基因组学研究的进展,人们现已把越来越多的注意力放在能同时提供有关多个基因及其序列的所谓并行分子遗传学分析方法上. 功能基因组学所研究的是在特定组织中、发育的不同阶段或者是疾病的不同时期基因的表达情况. 因此它的要求是能在同一时刻获得多个分子遗传学分析的结果. 譬如,许多疾病引发基因可能会有数以百计的与表征有关的特定突变. 因而,需要能同时筛选这些突变的有效方法. 另外,任何一个细胞都会有数以千计的基因在表达,而细胞间基因表达的差异往往能反应出这些细胞是发育正常还是在朝病变甚至恶性肿瘤细胞方向发展. 采用芯片技术根据杂交情况对基因表达进行分析的好处是一次杂交便能提供有关多基因差异表达的信息,它将给基因克隆、基因表达、疾病诊断和药物筛选带来巨大的革命[1]. 利用芯片技术进行基因表达研究,根据集成在芯片上 DNA 片段的不同来源,主要有以下两种.

2.1 寡聚 **DNA** 片段

固定于芯片上用作探针的寡聚 DNA 片段一般是 8～20 bp(低于 50 bp)长度的 DNA. 这些 DNA 片段可以采用光蚀刻法原位合成,也可以预先合成寡聚核苷酸后再通过机械接触而固定在芯片上. 研究基因表达时,可根据序列分析,选择基因成分中特异的片段合成探针,通过样品 DNA 与芯片进行特异杂交,即可对相应基因表达进行有效分析. 这种方法还可以用于 DNA 序列的自动化分析与验证、基因组多态性分析或基因突变的诊断等方面. 但由于合成的 DNA 片段长度有限,因而杂交的条件要求很高,杂交结果分析较为复杂. 也可将不同条件下从某生物体中转录出来的所有 mRNA 经标记成为探针后,再与代表它所有基因而制成的寡核苷酸方阵杂交. 通过分析杂交位点及其信号强弱,就可以得出不同情况下某个基因是否已表达及表达多少. 由于转录情况分析直接涉及到功能基因,它已成为 DNA 芯片研究中的一个重点和热点. 众所周知,在人类基因组中只有大约 3% 的序列能有表达,直接通过测序等手段来了解功能基因的情况相当费时、费力. 改用功能基因转录出来的 mRNA 与方阵杂交以研究功能基因的表达,其效率可提高 30 倍以上[2～4].

Mirzabekov 实验室利用寡核苷酸链长为 10～12 的微芯片与具荧光标记的 DNA 杂交对 β-地中海贫血症的各种突变进行了探查. 由于双链的稳定性和 AT 含量不同,通过选择解链曲线中分辨能力强的温度,可以提高杂交的专一性. 对同一样品进行两种不同荧光标记,并与同一微芯片杂交,在多波长荧光显微镜下平行检测两者的激发光强度即可以提高突变检测的精度[12]. Guschin 等[13]用与某类微生物的小亚基 rRNA 序列的互补序列为寡核苷酸探针,所选择的群体包括了硝化细菌的主要种类,可以有选择地保留目标序列. 通过改变探针的量可以使信号强度处于检测范围,而对标记的靶序列的选择性也极高. 利用多色荧光检测法同时定量测定多个目标探针群体是一条行之有效的途径.

Lockhart 等人采用固定有 6.5 万个不同序列的长度为 20 个核苷酸的探针芯片定量地分析了一个小鼠 T 细胞系中整个 RNA 群体中 21 个不相同的信使 RNA 表达. 这些专门设计的探针能与 114 个已知的小鼠基因杂交. 分析发现在诱发生成细胞分裂后,另外有 20 个信使 RNA 的表达也发生了改变. 检测结果表明该系统对 RNA 的检出率为 1∶30 万,对信使 RNA 的定量基准为 1∶300[2].

2.2 克隆的 **cDNA**

© 1994-2011 China Academic Journal Electronic Publishing House. All rights reserved. http://

将克隆的 cDNA 固定在芯片上就成为 cDNA 芯片. 该法是将 mRNA 反转录 cDNA 然

后对 cDNA 进行 PCR 扩增, 并分别等量转入微量滴定板的小孔内, 利用微量液体转移器, 将 cDNA "转印"至玻璃板或其它载体上, 经化学和热处理使 cDNA 附着于载体表面并使之变性, 制作成 cDNA 阵列杂交板. 从待检测物中提取 mRNA, 单周期反向转录制备荧光标记的 cDNA 混合物与上述制成的 cDNA 阵列杂交板进行严格杂交, 然后用激光扫描仪记录荧光强度, 由于荧光标记 cDNA 探针与阵列板中的 cDNA 相比数量稀少, 不能产生饱和现象, 所以每一阵列点的荧光强度决定了相应基因的转录强度, 其结果应与 RNA 印迹结果相符. 若用于对同一基因的表达进行分析时, 可用两个不同来源的 mRNA 分别以不同荧光染料作标记, 制备 cDNA 探针, 两者等量混合并与同一阵列杂交后, 分别扫描两种不同激发态的发光强度即可确知不同来源细胞的某一基因的转录水平, 能直观地反映出基因组中该基因的表达图谱. 若阵列板中包含的 cDNA 的数目接近整个基因组的表达基因数目, 那么杂交的结果就能反应出基因组全部基因的表达概貌. 目前微集阵列板可容纳至 2 万个 cDNA 探针. 据估计, 一个由 10 万个不同 cDNA 构成的微方阵可通过一次杂交对整个人类基因组的表达情况进行检测[14].

cDNA 芯片技术是以定量的方式同时监测大量基因相对表达的强有力的新方法. 它已成功地用于同时检测 1 000 多个基因的表达[15,16], 如测定植物[14]、酵母[17,18]、人类癌基因[3]和病毒样品的基因表达水平研究. 这些 DNA 芯片是由大量 PCR 扩增的基因片段以许多不连续的位点固定于玻璃片上构成的[10]. 来自两种不同类型的细胞 RNA 的 cDNA 探针, 各用不同颜色的荧光标记, 与芯片上的扩增片段杂交. 芯片上某一基因片段的 2 种不同颜色的信号的相对强度提供了该基因在 2 种细胞中的相对表达量的信息. 用 DNA 芯片检测的许多基因表达的变化中, 有一种热休克蛋白基因, hsp70, 在 HIV 急性感染 H9 细胞后发生改变[19], 其芯片数据被 Northern blot 分析进一步证实, 两者的结果非常吻合. Schena 等人用含 48 种成对 cDNA 材料的 DNA 芯片测定了拟南芥菜(*Arabidopsis thaliana*) 中的差异表达. Schena 等人还用热休克或佛波酯调节基因的双色差异表达监测分析了 1 046 种人的未知序列的 cDNA. 这一技术被用于炎症疾病, 如风湿性关节炎和肠炎的特异性基因表达研究[5]. 最近 Heller 等应用该技术对风湿性关节炎病变细胞与正常细胞间基因表达的差异进行了研究, 发现了一些与炎症相关的新基因, 并经对风湿性关节炎与炎性肠道疾病的组织比较确证了白细胞介素-3、趋化因子 Gro-α、金属蛋白酶基质的胰肽酶与这两种疾病的新关系. 从这两种病症的末梢血库得出的信息也揭示了金属蛋白酶-1 的组合子抑制物、铁蛋白的轻链和锰超氧化物歧化酶的基因表达具有显著性差异. Derisi 等人不仅制成了包含酵母所有基因的 cDNA 芯片[16], 而且还应用芯片技术研究了肿瘤产生的遗传基础[4], 他将一个恶性肿瘤细胞系中得到的 870 个不同的 cDNA 探针通过机手"印刷"到载玻片上观察癌基因的表达情况. 在比较两个标有不同荧光标记的细胞 mRNA 群的杂交结果之后, 他们对引入染色体后肿瘤基因受到抑制的细胞中基因表达结果进行了分析. Shoemaker 等报道了一种利用生物芯片技术来确定许多新近发现酵母基因的生物功能的所谓分子条形码技术. 该技术的好处是它能让我们以并行的方式定量地分析很复杂的核酸混合物[20].

3　展望

DNA 芯片的出现与快速发展并不是偶然的, 随着大量基因序列的确定, 分子生物学的研究越来越变成为一种生物信息学的分析过程. DNA 芯片由于采用了高通量、自动化与并

© 1994-2011 China Academic Journal Electronic Publishing House. All rights reserved. http://

行化过程,将为分子生物学探索提供全新的技术,使相应的研究速度、效率与规模大为提高.由于其巨大的分析能力,样品的用量极少,使用简便,还因具有可在同一杂交板上同一杂交反应中进行不同样品的比较等优点,此法在临床检测与基因表达研究中具有广阔的应用前景. 尽管目前尚有许多技术上的问题,相信随着研究的深入和技术的进步,这些问题都将得到解决. 正如《财富》[21]杂志所言:在写 20 世纪科技史时有两件事值得大书特书. 一是微电子芯片,它是社会与自然信息的总集成,是计算机和许多家电的心脏,它改变了我们的经济和文化生活,并已进入到每一个家庭;另一个就是 DNA 芯片,它是生命信息的集成,将给生命科学的研究方式带来重大改变,开辟一个生命信息研究和应用的新纪元.

参考文献:

[1] SCHULER G D, BOGUSKI M S, STEWART E A, *et al*. Life with 6000 genes[J]. Science, 1996, 274: 546-567.

[2] LOCKHART D J, DONG H, BYRNE M C, *et al*. Expression monitoring by hybridization to high-density oligonucleotide arrays[J]. Nature Biotechnology, 1996, 14: 1675-1680.

[3] SAIZIEU A de, CERTA U, WARRINGTON J, *et al*. Bacterial transcript imaging by hybridization of total RNA to oligonucleotide arrays[J]. Nature Biotechnology, 1998, 16: 45-48.

[4] DERISI J, PENLAND L, BROWN P O, *et al*. Use of a cDNA microarray to analyze gene expression patterns in human cancer[J]. Nature Genetics, 1996, 14: 457-460.

[5] HELLER R A, SCHENA M, CHAI A, *et al*, Discovery and analysis of inflammatory disease related genes using cDNA microarray[J]. Proc Natl Acad Sci USA, 1997, 94: 2150-2155.

[6] FODOR S P. Massively Parallel Genomics[J]. Science, 1997, 277: 393-395.

[7] CHEE M, YANG R, FODOR S P, *et al*. Accessing genetic information with high-density DNA arrays[J]. Science, 1996, 274: 610-614.

[8] MCGALL G, LABADIE J, NGUYEN T, *et al*. Light-directed synthesis of high-density oligonucleotide arrays using semiconductor photoresists[J]. Proc Natl Acad Sci USA, 1996, 3: 13555-13560.

[9] FODOR S P, RICHARD P R, XIAOHUA C H, *et al*. Multiplexed biochemical assays with biological chips[J]. Nature, 1993, 364: 555-556.

[10] SHALON D, SMITH S J, BROWN P O, *et al*. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization[J]. Genome Research, 1996, 6: 639-645.

[11] SCANGOS G. Drug discovery in the postgenomic era[J]. Nature Biotechnology, 1997, 15: 1220-1221.

[12] DROBYSHEV A, MOLOGINA N, SHIK V, *et al*. Sequence-analysis by hybridization with olignucleotide microchip-identification of beta-thalassemia mutations[J]. Gene, 1997, 188(1): 45-52.

[13] GUSCHIN D Y, MOBARRY B K, PROUDNICKOV D, *et al*. Oligonucleotide microchips as genosensors for determinative and environmental studies in microbiology[J]. Applied and Environmental Microbiology, 1997, 63(6): 2397-2402.

[14] SCHENA M, SCHALON D, DAVIS R W, *et al*. Quantitative monitoring of gene expression patterns with a complementary DNA microarray[J]. Science, 1995, 270(20): 467-470.

[15] SCHENA M, SCHALON D, HELLER R, *et al*. Parallel human genome analysis: microarray-

© 1994-2011 China Academic Journal Electronic Publishing House. All rights reserved. http://

based expression monitoring of 1000 genes[J]. Proc Natl Acad Sci USA, 1996, 93: 10614–10619.

[16] DERIS J, VISHWANATH R L, BROWN P O. Exploring the metabolic and genetic control of gene expression on a genomic scale[J]. Science, 1997, 278: 680–686.

[17] LASHKARI D A, MCCUSKER J H, DAVIS R W, *et al*. Whole genome analysis: experimental access to all genome sequenced sgments through larger-scale efficient oligonucleotide synthesis and PCR[J]. Proc Natl Acad Sci USA, 1997, 94: 8945–8947.

[18] LASHKARI D A, DERISI J L, MCCUSKER J H, *et al*. Yeast microarray for genome wide parallel gene expression analysis. Proc Natl Sci USA, 1997, 94: 13057–13062.

[19] WAINBERG Z, OLIVEIRA M, LERNER S, *et al*. Modulation of stress protein (hsp27 and hsp70) expression in CD4 + lymphocytic cells following acute infection with human immunodeficiency virus type[J]. Virology, 1997, 233: 364–373.

[20] SHOEMAKER D D, LASHKARI D A, MORRIS D, *et al*. Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar coding strategy[J]. Nature Genetics, 1996, 14: 450–456.

[21] STIPP D. Gene chip breakthrough[M]. Fortune, 1997, 135(6): 56.

·简讯·

《Nature》中关于生命科学研究进展的最新报道

文献标识码: E

编者按: 为了适应我国现代化建设的需要,促进生命科学的学术交流,本刊从本期起陆续介绍世界著名杂志《Nature》中刊登的关于生命科学研究进展的最新报道,以飨读者.

人类的祖先是怎样行走的

Evidence that humans evolved from a knuckle-walking ancestor (B G Richmond, D S Strait)

过去,曾有人提出过人类有一个像猩猩一样用指节着地行走的祖先,但这种观点缺少化石证据支持. 现在, Richmond 和 Strait 在早期南方古猿(australopithecines)的腕部发现了以前未曾发现的指节行走特征,这说明,双足行走这一将类人猿与其他灵长目动物区分开来的进化特征是从一个已经适应了陆地行走的祖先演化来的,而这种演化也许是新的进食方式或手的使用所产生的结果. 本期刊登的是一种名叫 Pan troglodytes(黑猩猩属的穴居类人猿)的指节行走者的腕部X-光照片(由哈佛大学的 Farish Jenkins 和纽约州立大学石溪分校的 John Fleagle 提供)(见 Nature, 2000, 404(6776): 382).

帕金森氏症的新模型

A Drosophila model of Parkinson's disease (M B Feany, W W Bender)

一种在果蝇身上开发出的帕金森氏症新模型,被认为对于这种疾病的研究具有重大价值,因为科学家可利用专为果蝇开发的一系列强大的基因工具来研究这种疾病. 虽然果蝇与人类相差悬殊,但该模型却能很好地重现人类帕金森氏症的主要特征. 为 a-synuclein(一种丰富的、功能未知的神经蛋白)编码的基因发生突变,可在家族性帕金森氏症患者身上导致神经发生,而在果蝇身上产生这一动物模型的正是同样的基因突变(见 Nature, 2000, 404(6776): 394).

© 1994-2011 China Academic Journal Electronic Publishing House. All rights reserved. http://

(下转第 129 页)