

## · 综述 ·

## 基因芯片在传染病分子流行病学中的应用

曾浔 张建中

基因芯片是指固定有许多有序列的寡核苷酸探针的载体,通过杂交检测,对各种基因进行分析<sup>[1]</sup>。由于它所具有的高通量性、快捷、便宜等优点,在各个领域起着越来越重要的作用,其中也包括分子流行病学领域。基因芯片主要用于菌株的地理流行病学分析、菌株的重要分子特征相关的流行病学分析——分子流行病学的病因学研究及传染病的诊断与鉴别诊断。本文主要从芯片的原理、应用及所存在问题等几方面,初步探讨基因芯片在传染病流行病学应用中的广阔前景。

## 一、基本原理

Fodor 等<sup>[2]</sup>于 1991 年首先提出了基因芯片的概念,进而随着分子生物学的突飞猛进,尤其是人类基因组计划的完成及后基因组研究工作的开展,基因芯片技术日趋成熟。它是一种通过微加工技术将成千上万个寡核苷酸或 cDNA 固定在一个很小的芯片上,用以对各种基因进行检测和分析。目前,基因芯片的种类主要包括寡核苷酸芯片、cDNA 芯片、肽芯片和芯片实验室(lab on a chip)。芯片技术由于其高通量分析特征,对于分子生物学、分子流行病学及其他相关学科的研究将产生难以估量的影响。

基因芯片的基本原理是通过原位杂交方法,用已知的 DNA 序列检测未知序列。目前,基因芯片主要用于 DNA 突变分析<sup>[3]</sup>、基因谱差异分析<sup>[4-8]</sup>、基因诊断分析<sup>[9-11]</sup>和大规模基因类型分析<sup>[12]</sup>。其研制技术主要包括以下几大步骤:寡核苷酸探针的合成、固定、杂交和检测。

制造基因芯片的载体材料必须是固体片状或薄膜,其表面硅烷化后可通过连接物将各种生物分子固定,应用最普遍的是玻璃片。

目前,连接物通常是在醇化、氯化、二硫化或丙烯酰胺化处理的载体表面与探针连接<sup>[13-18]</sup>。寡核苷酸探针主要是通过获取 mRNA、逆转录制备 cDNA 探针获得。逆转录过程中加入同位素或荧光素修饰的 dNTP 标记,然后将合成后 DNA 固定法固定在玻片上,再用标记的样品与芯片上固定的探针间进行的特异性选择反应,即杂交反应;最后,通过荧光显微镜检测各杂交点的荧光信号强弱或不同荧光素的颜色(双色荧光素标记),也可用同位素标记、化学发光检测、电化学发光检测及质谱法等进行检测。目前,基因芯片主要用于未知 DNA 片段检测和基因谱差异分析。

作者单位:102206 北京,中国疾病预防控制中心传染病预防控制所腹泻病研究室

国内外有许多网站对基因芯片的概念、制造、种类等方面有非常详细的报道<sup>[19-21]</sup>,本文不再赘述。

## 二、基因芯片在病原微生物及分子流行病学中的应用

目前,基因芯片在真核生物方面的应用较多,而在原核生物方面应用刚刚起步,但基因芯片在分子流行病学中的应用已显示出广阔的前景,主要应用于以下这些领域。

1. 菌株的地理流行病学分析:由于受方法学的限制,传统的病原地理流行病学分析一次只能对病原体的少数基因进行分析,在实际分析中不可避免地会导致一些重要的数据的遗漏,这有可能使实验结论与实际结果出现偏差,而芯片技术从根本上解决了这个问题。目前,应用的实例很多,最有代表性的例子是有关对卡介苗(BCG)的分析,卡介苗是计划免疫中非常重要的一种菌苗。由于卡介苗的原始出发菌株在一次世界大战中丢失,同时各国的卡介苗菌株通过多次传代,部分基因已发生改变或缺失,因此,各国的卡介苗菌株可能存在很大的不同,这就给全球范围内的计划免疫工作带来了很大的不便。Behr 等<sup>[23]</sup>用 DNA 芯片检测了多个国家卡介苗的子代菌株。他们通过人型结核杆菌 H37Rv 的基因框架来分析卡介苗。用 H37Rv 的全部 3 924 个开放阅读框(ORF)中的 3 902 个 ORF 制成芯片,对不同的卡介苗进行杂交。通过比较,发现了卡介苗菌株 16 个缺失区域,长度从 1 903 到 12 733 bp 不等,其中 4 个区域已有报道。他们将这些缺失区域进行统一命名,从 RD1 到 RD16。其中 9 个区域卡介苗及牛型结核杆菌(*M. bovis*)菌株全部缺失,2 个区域卡介苗及牛型结核杆菌菌株部分缺失,1 个区域所有卡介苗菌株全部缺失,4 个区域存在某种卡介苗菌株缺失。对于卡介苗菌株间的不同,Behr 认为那些在牛型结核杆菌存在而在卡介苗中缺失的 H37Rv 的区域是在传代过程丢失的。这些问题的解决对于疫苗评价以及改进有很大的意义,但用传统的分子生物学方法是不可能完成的。

2. 菌株分子流行病学及疾病病因学研究:传统流行病学对于传染病爆发中不同菌株的代谢分析、毒力分析、年代分析缺乏有效的手段。目前,大多数传染病及地方病的病因已比较明确,但是,对于它们的分子机制却了解很少。这主要是因为传统的方法学的限制,无法用有限的人力、物力对大样本量、多地区的信息在短时间内进行分析。由于方法上的欠缺,使防疫工作者很难从分子机制上探讨导致传染病爆发的菌株之间的异同,例如在历史上几次流行性感冒和霍乱的大流行,病原体有变异,但也有联系,这对于这些疾病的今后防治有重要的指导意义。基因芯片与传统的流行病学方

法相结合, 恰恰解决了这个矛盾。

Fitzgerald 等<sup>[24]</sup>用芯片技术对葡萄球菌的基因组的进化进行了分析, 发现大约 22% 的基因对葡萄球菌是可有可无的。同时这些基因的转移在葡萄球菌的进化中起很重要的作用。对甲氧苯青霉素耐药的葡萄球菌的 *mec* 基因(一种抗  $\beta$ -内酰胺的基因)被水平的转移了 5 次, 说明抗甲氧苯青霉素的菌株进化了多次, 而不是从单一的原始菌株进化而来。这就可以解释在 20 世纪 70 年代流行的中毒性休克综合征(TSS)是由宿主环境的改变造成, 而不是由一个基因谱的快速变化产生的一个新的亚型。此研究说明基因芯片能够对以往流行病学不能回答的问题(如造成 TSS 流行的具体原因)进行回答。芯片的出现是对传统流行病学缺陷的有力补充。

Lu 等<sup>[25]</sup>在中国贵州等地对砷中毒性肝病的分子机制进行了研究。他们用 cDNA 芯片对暴露在砷环境下 6~10 年、已经有部分砷中毒症状的人群进行研究。通过研究他们认为基因表达谱的变化在砷中毒性肝病及肝癌中起决定性作用。该研究对于大规模流行病的病因调查又开辟了新的途径。

3. 传染病的诊断与鉴别诊断: 传染病在流行病学中占有非常重要的地位, 尤其是近年来性传播疾病(STD)、结核病等的死灰复燃和大量新发现传染病的出现, 给各级防疫部门提出了严峻的考验。在传染病爆发中, 传统的流行病学研究从取样、培养病原体到鉴别诊断, 往往需要几天时间, 这对于传染病的治疗和控制极为不利。加之人员流动的频繁和传播机会的急剧增加, 需要找到一种快速、方便方法进行诊断和鉴别诊断, 而基因芯片恰恰能满足以上要求。目前, 在这方面已进行了初步的尝试。Chizhikov 等<sup>[9]</sup>将部分沙门菌、志贺菌及埃希菌的毒力基因作为芯片上的探针, 可以对以上三种菌进行检测。另外, 还出现了一些用于检测 HIV 及结核杆菌的芯片<sup>[10, 11]</sup>。但这还远远不够。这些基因芯片的共同特点是检测的探针少, 所得到的信息少, 只是对一种病原体或几种相关的病原体的检测, 对于多种病原体的鉴别诊断仍未进行, 应当说它们只是诊断芯片的雏形。但是, 由于基因芯片对于疫情爆发后快速的初步诊断、传染病的监控以及各地菌型变化的监测有得天独厚的优势, 它以诊断快速、易携带及储存方便而操作不需要太强的专业性等使基因芯片极易在地区及县一级防疫部门推广。目前, 在这一领域还处于探索阶段, 但这是传染病诊断中主要的发展方向之一。

### 三、存在的问题

尽管基因芯片技术已有许多方面的进步。但是在分子流行病学的具体应用中仍存在一些问题: 首先, 探针的杂交也有一定的错配率, 从而产生一定的背景, 如何区分这种背景所造成的假阳性和由于传染病样品中由于拷贝数太低所造成的弱阳性, 是传染病诊断 DNA 芯片面临的主要挑战, 由于在传染病诊断 DNA 芯片中采用的密度较低, 因此它的这些缺点就不如高密度芯片那样严重。但是, 由于其结果往往

伴随着许多重大疫情的处理, 诊断必须十分慎重。

另一种对应用的限制来自芯片研究的成本。在芯片研究过程中, 对样品的标记及探针的制备都非常耗时、耗力。对病原微生物来说, 由于绝大部分的病原微生物都是原核生物, 它们的共同特点是没有 polyA 尾巴, 因此无法象真核生物那样通过 polyA 尾巴纯化 mRNA, 并且用逆转录方法建立 cDNA 文库。这就给制造芯片带来很大的问题。Tao 等<sup>[26]</sup>报道直接在混合的 RNA 中(mRNA、tRNA、rRNA)加入 3' 端的特异引物扩增 cDNA 探针。但是, 另一方面, 与真核生物相比, 病原微生物(原核生物或病毒)的研究又有得天独厚的优势。因为这些微生物的基因组都较小, 只有几千个 ORF, 同时又没有内含子, 这样就可以通过对每个 ORF 进行扩增, 用扩增产物代替 cDNA, 制备探针。这种探针的另外一个好处就是它包含了病原微生物所有的基因, 这就避免了 cDNA 文库可能使某些基因遗漏的潜在问题。这样虽然解决了建立 cDNA 文库的问题, 但是, 以大肠埃希菌为例, 4 290 个 ORF 就需要 4 290 个特殊的引物, 这大大的增加了科研难度和成本。因此, 如何方便、快捷的建立病原微生物的检测探针是基因芯片研究中亟待突破的瓶颈。随着新技术的不断出现, 基因芯片的成本将会越来越低。

第三, 样品的获得与处理。对于传染病诊断芯片来说, 样品的来源广泛并且混杂有各种各样的干扰因素。在实际推广过程中, 要使广大防疫部门在很短的时间内, 在广大农村地区很快的获得象实验室中获得的样品是不可能的。因此, 这就需要传染病诊断芯片有很高的特异性和敏感性。如何正确的对样品进行简单的处理, 使混杂因素降到最低是广大防疫工作者所面临的主要挑战。

第四, 对于数据的处理和应用仍存在问题。基因芯片的出现对生物信息学提出了新的要求, 对于大量的数据需要新的方法来处理。同时, 对于基因芯片所提供的信息也要求更精确的判断。Tao 等<sup>[26]</sup>对大肠埃希菌的全基因序列制成的芯片检测细菌生长在微量葡萄糖环境和含葡萄糖的 Luria 肉汤环境中基因组表达方式。他们发现绝大多数基因的翻译表达明显上升, 而细菌在葡萄糖缺乏的环境中一些基因的表达也上升。这么多的基因表达发生了变化, 这使数据的分析和应用变的很困难。更重要的是, 许多基因同时变化可能使真正的原因被掩盖, 因为一个基因的变化往往导致一连串下游基因或调控基因全都发生变化, 因此, 如何找到真正引起这些连锁反应的基因是科研人员需要注意的问题。另一个问题是, 由于绝大多数病原微生物都是原核生物, 它们的 mRNA 多为多顺反子, 即一条 mRNA 可翻译成多种相关蛋白, 所以它的 cDNA 就代表了多个基因。而且编码这些蛋白的基因相互交错, 即使以每个基因作为探针, 在杂交时也很容易发生假阳性。这就对结果的判断带来很大的不便。这就需要其他相关领域的共同研究和帮助, 如基因组草图的绘制、二维蛋白电泳等。

### 四、展望

从 1991 年 Fodor 提出芯片这个概念到现在,已有 2 000 多篇各类与芯片有关的文章发表,其中绝大多数是在 1999 年以后发表的。这说明基因芯片在最近二三年发展迅速,应用广泛。目前,对于基因芯片的研究朝着快速、高质量、准确的方向高速发展。虽然基因芯片仍有许多问题有待解决,但其前景一片光明。在未来流行病学领域中,基因芯片将在以下几个方面具有广泛的应用前景。

1. 传染病的快速诊断及监测:未来的传染病爆发后(尤其是大规模自然灾害爆发后)的检测主要依靠基因芯片的快速初筛及诊断,做到 24 h 内明确病因,为防疫工作者提供可靠的数据,是指导控制疫情的必要手段。

2. 分子流行病学资料分析:对于一些病因机制不清、病原体相关资料缺乏的传染病可以用基因芯片进行研究,为历史上爆发的传染病之间的联系提供新的线索。

3. 病因及流行规律的探究:随着人类基因组计划的完成,使一些遗传性疾病病因的分子机制得到了明确。基因芯片的使用将对这些遗传性疾病的流行机制的探讨会有很大帮助。同时,基因芯片还可用于一些慢病(如肿瘤、慢性传染病、原发性高血压等)的早期筛查及普查。这比用传统的方法将更加准确、便宜、方便。

可以预见,基因芯片随着成本的不断降低、准确性的不断提高,将成为流行病学研究中一个有力的工具。

### 参 考 文 献

- 1 Schena M, Editer. DNA microarrays-a practical approach. Oxford University Press, 1999. 211.
- 2 Fodor SP, Read JL, Pirrung MC, et al. Light-directed, spatially addressable parallel chemical synthesis. Science, 1991, 251:767-773.
- 3 Hacia JG, Collins FS. Mutational analysis using oligonucleotide microarrays. J Med Genet, 1999, 36:730-736.
- 4 Oh MK, Liao JC. Gene expression profiling by DNA microarrays and metabolic fluxes in *Escherichia coli*. Biotechnol Prog, 2000, 16:278-286.
- 5 Rosenberge CM, Scott MG, Gold MR, et al. *Salmonella typhimurium* infection and lipopolysaccharide stimulation induce similar changes in macrophage gene expression. J Immunol, 2000, 164:5894-5904.
- 6 Han J, Yoo HY, Choi BH, et al. Selective transcriptional regulations in the human liver cell by *Hepatitis B*, viral X protein. Biochem Biophys Res Commun, 2000, 272, 525-530.
- 7 Richmond CS, Glasner JD, Mau R, et al. Genome-wide expression profiling in *Escherichia coli* k-12. Nucleic Acids Res, 1999, 27:3821-3835.
- 8 Stingley SW, Ramirez JJ, Aguilar SA, et al. Global analysis of herpes simplex virus type 1 transcription using an oligonucleotide-based DNA microarray. J Virol, 2000, 74:9916-9927.
- 9 Chizhikov V, Rasooly A, Chumakov K, et al. Microarray analysis of microbial virulence factors. Appl Environ Microbiol, 2001, 67: 3258-

3263.

- 10 Kozal MJ, Shah N, Shen N, et al. Extensive polymorphisms observed in the HIV-1 cladeB protease gene using high-density oligonucleotide arrays. Nature Med, 1996, 2:753-759.
- 11 Gingeras TR, Ghandour G, Wang E, et al. Simultaneous genotyping and species identification using hybridization pattern recognition of generic mycobacterium DNA arrays. Genome Res, 1998, 8:435-448.
- 12 Huang F, Adelman J, Jiang H, et al. Identification and temporal expression pattern of genes modulated during irreversible growth arrest and terminal differentiation in human melanoma cells. Oncogene, 1999, 18: 3546-3552.
- 13 Markus B, Jorg DH. Versatile derivatisation of solid support media for covalent bonding on DNA-microchips. Nucleic Acids Res, 1999, 27: 1970-1977.
- 14 Proudnikov D, Timofeev E, Mirzabekov A. Immobilization of DNA in polyacrylamide gel for the manufacture of DNA and DNA-oligonucleotide microchips. Anal Biochem, 1998, 259:34-41.
- 15 Rehman FN, Audeh M, Abrams ES, et al. Immobilization of acrylamide-modified oligonucleotides by co-polymerization. Nucleic Acids Res, 1999, 27:649-655.
- 16 Rogers YH, Jiang-Baucom P, Huang ZJ, et al. Immobilization of oligonucleotides onto a glass support via disulfide bonds: a method for preparation of DNA microarrays. Anal Biochem, 1999, 266:23-30.
- 17 Weiler J, Hoheisel JD. Combining the preparation of oligonucleotide arrays and synthesis of high-quality primers. Anal Biochem, 1996, 243: 218-227.
- 18 Kumar A, Larsson O, Parodi D, et al. Silanized nucleic acids: a general platform for DNA immobilization. Nucleic Acids Res, 2000, 14:e71.
- 19 www.DNA.net.cn
- 20 www.37c.com.cn
- 21 www.affymetrix.com
- 22 www.genecore.com
- 23 Behr MA, Wilson MA, Gill WP, et al. Comparative genomics of BCG vaccines by whole-genome DNA microarray. Science, 1999, 284: 1520-1522.
- 24 Fitzgerald JR, Sturdevant DE, Mackie SM, et al. Evolutionary genomics of *Staphylococcus aureus*: insights into the origin of methicillin-resistant strains and the toxic shock syndrome epidemic. Proc Natl Acad Sci, 2001, 98:8821-8826.
- 25 Lu T, Liu J, LeCluyse EL, et al. Application of cDNA microarray to the study of arsenic-induced liver diseases in the population of Guizhou, China. Toxicol Sci, 2001, 59:185-192.
- 26 Tao H, Bausch C, Richmond C, et al. Functional genomics: expression analysis of *Escherichia coli* growing on minimal and rich media. J Bacteriol, 1999, 181:6425-6440.

(收稿日期 2001-11-30)

(本文编辑:尹廉)