

基于连接酶检测反应的基因芯片分型技术的建立

肇旭 秦胜营 冯国鄞 贺林

【摘要】 目的 建立基于连接酶检测反应的基因芯片分型技术。方法 将连接酶检测反应、通用芯片技术及片基异硫氰基化从头活化方法相结合。结果 应用测序技术验证,该方法分型准确性达到 100%。另外也不会产生非模板依赖性信号,对未纯化模板的分型效果也较好。结论 连接酶检测反应能够在高温下进行,显著减少了探针非特异性连接,其通用性使得研究者可基于一张芯片自由选择所研究的位点,无需针对每一组探针进行操作条件的单独优化,并且保证了杂交的特异性;片基从头活化技术更使其摆脱了对商品化片基的依赖。这些特点都极大地提高了该技术的分型准确性,并且降低了费用,使其应用于高通量 SNP 分型成为可能。

【关键词】 基因芯片; 连接酶检测反应; 分型

Establishment of DNA Chip-based Genotyping Technology Mediated by Ligase Detection Reaction ZHAO Xu^{1,2}, QIN Sheng-ying^{1,2}, FENG Guo-yin³, HE Lin². (¹Shanghai Jiao Tong University, Shanghai 200030; ²Institute for Nutritional Sciences, Shanghai Institute of Biological Sciences, Chinese Academy of Sciences, Shanghai 200031; ³Shanghai Institute of Mental Health, Shanghai 200030, P. R. China)

Corresponding author: HE Lin. E-mail: helin@nhgg.org

【Abstract】 Objective To establish a high-discrimination and cost-effective technology with significant promise for future high-throughput SNP genotyping. **Methods** Five probes including two discriminating probes, one common probe and two corresponding Zips, which were coupled to the slides at known locations, were designed for each SNP. In addition, the positive and negative marker probes were designed for result evaluation. The standard microscope glass slides were activated by PDITC. The PCR and LDR were all performed using multiplex amplifying method. After hybridization and scanning, the average signals after subtraction of local background were used for calculating the signal ratios of spot pairs corresponding to different alleles. **Results** Genotyping accuracy of this technology was validated as 100% by comparing with sequencing method. The homozygote and heterozygote could be distinguished easily by this method. The signal ratio of positive and negative dots was higher than 10. No signal was detected in the experiment without DNA template, which demonstrated that the detection result relied on the template directly. When analyzing the same subject one more times, the results could be repeated very well, indicating that the possibility of variation with uncertain conditions was very low. In addition, genotyping of non-purified templates revealed good results. **Conclusion** LDR can be performed at a relatively high temperature, which helps to eliminate non-specific ligations. The characteristic of universal gives researchers freedom to detect specified polymorphic loci based on the same array, with no need to optimize the operation condition of each probe groups and ensures the specificity of hybridization. Self-activating method makes us avoid the dependence on commercial slides. These improve the genotyping accuracy to a high level and reduce the cost greatly, which make it possible for high-throughput genotyping.

【Key words】 DNA chip; Ligase detection reaction; Genotyping

基金项目:国家高技术研究发展计划(863 计划)(No. 2002AA223021)

作者单位 200030,上海交通大学(肇旭、秦胜营);200031,中科院上海生命科学研究所以营养学院(肇旭、秦胜营、贺林);200030,上海市精神卫生中心(冯国鄞)

通讯作者 贺林(E-mail: helin@nhgg.org)

基因芯片是进行高度并行性基因组分析的工具,由固定在载体上的若干按照一定规则排列的 DNA 探针组成^[1,2]。基因芯片是最有效的 SNP 分型方法之一,迄今为止,人们已经开发了多种 SNP 分

型芯片技术。

早期的分型芯片应用等位基因特异性寡核苷酸探针,基于杂交反应,通过严格控制杂交及洗脱的条件进行区分不同的基因型^[3~5]。但由于此法是直接在不同的基因组序列上进行探针的设计,所以序列之间的差距较大,这样在进行高通量的检测时便难以寻找到适合众多探针序列的杂交条件,从而导致其结果的特异性较差,据报道在高通量实验中该方法准确率仅为 70% 左右。Affymetrix 公司采用对每一位点设计多条探针来进行纠错,这也势必带来了分型成本的提高。后来,酶促反应被引入芯片分型技术中,取得了较好的特异性,如单碱基延伸法、等位基因特异性引物延伸法、连接酶法等^[6~16]。这些为研究者自制芯片并自由选择研究位点提供了极大的便利,但是这些方法大多仍依赖于商品化片基,因此费用依然昂贵,难以在普通实验室广泛使用。

本研究拟采用高温连接酶检测反应(LDR)与通用芯片技术相结合的方法,并将之与玻片异硫氰基化方法结合,从而保证分型的准确率,并且摆脱对商品化片基依赖,建立一种特异性高、稳定性好、费用低廉的基因芯片分型技术。

1 材料与方法

1.1 玻片活化:

1.1.1 预处理 用碱液清洗玻片,然后用 ddH₂O 清洗 2min,随后放置在装有 1mol/L NaOH 的玻片缸内浸泡过夜,取出后用 ddH₂O 清洗 2min,再放置于装有 1% 的 HCl 的玻片缸中浸泡 10min,最后用 ddH₂O 彻底清洗后备用。

1.1.2 氨基化 将清洗后的玻片放置在含 2% 的 4-氨基丁基-3-乙氧基硅烷及 95% 的丙酮溶液中浸泡 15min。然后立即用新鲜丙酮清洗 5 次,每次 5min。置于 110℃ 烘箱中烘烤 45min。冷却后储存待用。

1.1.3 异硫氰基化 将苯异硫氰酸酯溶解于 10% 无水吡啶、90% 无水 DMF 的溶液中,调整浓度至 1mmol/L 并充分混匀。然后将氨基化玻片放置在苯异硫氰酸酯溶液中浸泡 1.5h。反应结束后分别用无水 DMF 和二氯乙烷溶液清洗 3 次,迅速用氮气将玻片吹干,最后将活化好的片基 4℃ 避光保存于充氮容器中。

1.2 探针设计:现以 NMDA 受体基因 *GRIN1* 及 *GRIN2B* 为研究对象,选取了 5 个 SNP 位点。其中 G1001C 位于 *GRIN1* 基因上;G366C、C2664T、T4196C 及 T5988C 位于 *GRIN2B* 基因上。

对于任一待研究 SNP 位点,在位点的 5' 方向根据 SNP 的碱基类型设计两条特异性探针,每条探针的 3' 末端碱基分别与此位点的一种等位基因型相对应,同时在该探针的 5' 末端设计一段与芯片上 Zip 序列互补的一段序列,称之为 cZip 序列。此外在该位点的 3' 方向序列中设计一条末端带有 Cy3 修饰的通用探针。即每一个 SNP 位点分别具有 3 条探针,同时芯片上还有两条与之对应的 Zip 探针。Zip 系列探针设计根据 Gerry 等^[15]提供的探针设计方法,将其提供的 4 个碱基对随机组合出长度为 24bp 的寡核苷酸序列,这些序列具有相同或相似的 T_m 值。然后利用 NCBI 网站的 BLAST 程序将这些序列与基因组序列库进行同源性比较,挑选出与基因组序列库同源性程度最低的一组寡核苷酸片段序列作为芯片上的 Zip 探针序列。*GRIN1* 及 *GRIN2B* 基因分型探针序列见表 1。

为了使分型结果更加准确可行,需设计阳性对照探针和阴性对照探针,其中阳性对照探针是末端带有 Cy3 荧光标记的寡核苷酸片段,可直接固定于片基上面,此探针可评价芯片制作质量并同时作为芯片图像分析时的区域识别标记,序列为:5'-NH₂-TTTTTTTTTTTTTTTTTTTCCCAGGTGATACTTGAAAC-Cy3-3'。阴性对照探针对芯片杂交特异性进行评价,序列为:5'-ATCGACCTGGTAATCGTGCAGCCTTTTTTTT-TT-NH₂-3'。

1.3 芯片制备:

1.3.1 点样 将 Zip 序列探针溶解于点样液中配成 20μmol/L 溶液。将其转移到与点样仪配套的 96 孔板内,使用 GMS417 点样仪(GMS, USA),根据实验需要设计点样程序,然后运行程序将探针点到活化好的片基上;点样完毕后,将芯片置于 37℃,90% 湿度的恒温保湿箱中过夜,使探针与玻片充分交联,随后分别用水和甲醇清洗芯片 2min,离心干燥。

1.3.2 封闭 将固定好探针的芯片在 10% 的氨水中浸泡两次,每次 1h,使未参与反应的活性基团充分封闭,最后用蒸馏水清洗 5min,离心干燥,4℃ 避光保存。

表 1 GRIN1 及 GRIN2B 基因分型探针

Table 1 The probes for SNP genotyping in GRIN1 and GRIN2B gene

SNP	Zip(5'-3')	通用探针(Common probe)(5'-3')	特异性探针(Discriminating probe)(5'-3')
G1001C	TGCGACCTCAGCATCGACCTCAGC- TTTTTTTTTT-NH2	PGGGGGTCCCAAGTTACATGGC-Cy3	GCTGAGGTCGATGCTGAGGTCGCA- CTATTCTCTTAGCCCGAGGAG
	CAGCGGTAGACCACCTATCGTGCG- TTTTTTTTTT-NH2		CGCACGATAGGTGGTCTACCGCTG- CTATTCTCTTAGCCCGAGGAG
G366C	GACCACCTTGGCATCGGGTACAGC- TTTTTTTTTT-NH2	PGGGGTGAGAGTCTGTGCTGAAA-Cy3	GCTGTACCCGATCGCAAGGTGTC- CCCCGTGGATGCCCAGGATG
	TGCGATCGCAGCGGTAACCTGACC- TTTTTTTTTT-NH2		GGTCAGGTTACCGCTCGCATCGCA- CCCCGTGGATGCCCAGGATG
C2664T	GACCGGTATGCGACCTGATGCG- TTTTTTTTTT-NH2	PGCAACCATGAACAACACACTC-Cy3	CGCATACCAGTTCGCATACCGGTC- AGTCTGTAATGAACTCCCCCACC
	CAGCATCGACCGGTAATCGGACC- TTTTTTTTTT-NH2		GGTCCGATTACCGGTCCGATGCTG- AGTCTGTAATGAACTCCCCACT
T4197C	GACCATCGTCCGGTAGGTAGACC- TTTTTTTTTT-NH2	PGGCAGCAAATCCTACTTCTCAG-Cy3	GGTCTACCTACCCGCACGATGGTC- GGACGACCACTGCTTGCTCCAT
	GGTAACCTACCTATCGCAGCCAGC- TTTTTTTTTT-NH2		GCTGGCTGGCATAGGTAGGTTACC- GGACGACCACTGCTTGCTCCAC
T5988C	CAGCACCTGACCATCGATCGCAGC- TTTTTTTTTT-NH2	PGCAACACCTTGTTCCTCTCAG-Cy3	GCTGCGATCGATGGTCAGGTGCTG- GGTAGAAGCCCTTCTACCTTCCT
	TGCGGGTACAGCACCTACCTTGCG- TTTTTTTTTT-NH2		CGCAAGGTAGGTGCTGTACCCGCA- GGTAGAAGCCCTTCTACCTTCCT

1.4 样品制备:为达到芯片检测的灵敏度,基因组 DNA 首先需要经 PCR 扩增,扩增采用的是多重 PCR 反应。应用 MIT Whitehead Insitute 的 Primer3.0 软件针对每个 SNP 位点设计两条引物,要尽量保证所有 SNP 位的引物的 T_m 值相同^[15]。PCR 反应体系为(25 μ L):10ng 基因组 DNA,10mmol/L Tris-HCl (PH8.3),50mmol/L KCl,3.0mmol/L MgCl₂,200 μ mol/L dNTP,引物各 0.6 μ mol/L,3.0U Taq DNA 聚合酶。反应条件为:95 $^{\circ}$ C 预变性 3min,95 $^{\circ}$ C 变性 30s,55~60 $^{\circ}$ C 退火 1min(根据具体扩增引物调整退火温度),72 $^{\circ}$ C 延伸 1min,循环 35 次,72 $^{\circ}$ C 延伸 10min。反应在 PTC-100 型 PCR 仪上进行,完成后在产物中加入 1 μ L 蛋白酶 K(20mg/mL),混匀后于 70 $^{\circ}$ C 反应 10min,随后于 94 $^{\circ}$ C 加热 15min 灭活蛋白酶 K。

待扩增产物经蛋白酶消化后,即可进行 LDR 反应。反应体系为(20 μ L):10 \times Taq DNA 连接酶 buffer 2 μ L,总探针混合液(各探针浓度均为 1 μ mol/L)1 μ L, Taq DNA 连接酶(NEB,USA 40U/ μ L)0.5 μ L,PCR 产物 3 μ L,最后用 ddH₂O 补至 20 μ L。反应条件为:94 $^{\circ}$ C 变性 30s,63 $^{\circ}$ C 退火 4min,循环 40 次。反应在 PTC-100 型 PCR 仪上进行。

1.5 杂交:在 20 μ L 的 LDR 反应产物中加入 15 μ L ddH₂O 和 35 μ L 2 \times 杂交液,充分混匀后于 94 $^{\circ}$ C 变性 2min,然后立即置于冰上,随后 4 000r/min 离心 5min。将 Gene Frame 杂交框(ABgene,UK)粘贴于芯

片的点样区域,吸取 65 μ L 杂交混合液加至杂交框内,然后将之封闭,封闭时应使杂交混合液均匀分布于杂交框内并防止有气泡进入,将芯片放置于 65 $^{\circ}$ C 培养箱中杂交 3h。杂交结束后将杂交框除去,然后将芯片在 1 \times SSC 0.1%SDS 洗脱液中清洗 2min,再于 0.05 \times SSC 中清洗 1min,最后 1 000r/min 离心甩干。

1.6 扫描与数据分析:应用 GMS418 型扫描仪(GMS,USA)对芯片结果进行扫描,扫描得到的 16-bit TIFF 杂交图像用 GenePix 2.0(Axon Instruments, Inc.)软件进行分析得到各个芯片点的荧光信号强度值和背景值,将荧光信号强度值扣除相应的背景值即得此样品点的绝对信号强度。最后计算同一位点不同等位基因型探针点的绝对信号强度比值,根据比值所处范围确定基因型。

2 结果

我们建立了对 GRIN1 及 GRIN2B 基因上 5 个 SNP 位点进行分型的芯片,扫描结果见图 1。

为减少由于点样遗漏以及片基活化不均一性带来的结果偏差,对同一探针重复点样 5 次。该扫描结果表明 5 个 SNP 位点的基因型情况为:1001G/G, 366C/C, 2664C/T, 4197T/T, 5988T/T。从图中可以看出杂合子及纯合子均具有较高的区分度,阳性点与阴性点信号强度比值平均达到 10 以上。我们也任意选择 20 个样本分别应用测序技术及芯片分型

技术进行了 5 个 SNP 的分型,两种方法分型结果完全一致,芯片分型技术的准确率达 100%。

此外我们还做了无模板的对照反应,其中一个反应中未加任何检测模板,而另一个反应中只加 3 个位点的扩增模板,结果见图 2。

从图中可以看出在未加任何检测模板时,没有检测到假阳性信号,同时在只有 3 个位点模板的反应中缺少模板的位点也未检测到任何信号,这些结果说明此法不会产生非模板依赖性信号。这一点使得结果具有较高的模板忠实性。Erdogan 等^[17]在使用芯片等位基因特异性延伸法进行 SNP 分型时发

现在芯片上进行聚合酶延伸时极易产生非模板依赖性的信号。在这一点上本文所建立的技术比利用聚合酶进行分型的芯片技术具有更大的优势。

为简化整个分型技术的操作步骤,我们也分析了 PCR 产物纯化前后分型结果的变化。考虑到在某些位点 PCR 扩增反应过程中有时需要加入一些优化剂,我们选择了一个各位点均为纯合子的模板,对经过纯化 PCR 产物、未经纯化无优化剂 PCR 产物、未经纯化含 DMSO 优化剂 PCR 产物以及未经纯化含 BSA 优化剂 PCR 产物的分型结果分别进行了比较和分析,结果如图 3 所示。

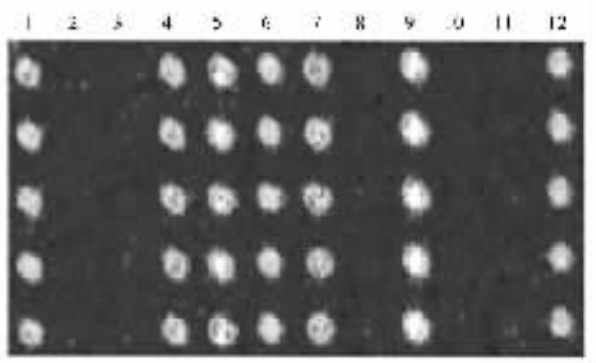


图 1 基因芯片分型结果图

Fig. 1 Results of DNA chip-based genotyping

分型图谱中每列含 5 个重复点且每列与一种基因型相对应。具体对应方式为:

1-1001G ;2-1001C ;3-366G ;4-366C ;5-2664C ;6-2664T ;7-4197T ;8-4197C ;9-5988T ;10-5988C ;11-阴性对照点 ;12-阳性对照点。

Each rank of the array corresponds to a polymorphism allele and contains five-spot replicates. Deposition scheme :

1-1001G ;2-1001C ;3-366G ;4-366C ;5-2664C ;6-2664T ;7-4197T ;8-4197C ;9-5988T ;10-5988C ;11-negative marker ;12-positive marker.

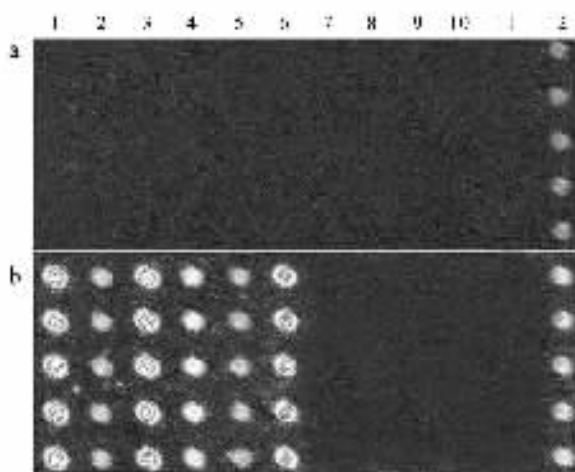


图 2 非模板依赖性假阳性验证

Fig. 2 False-positive test of non-template dependence

a :反应液中未加入任何模板 ;b :反应液中仅加入 3 个位点的模板。

a : Reaction solutions with no template ; b : Reaction solutions with templates of 3 SNPs

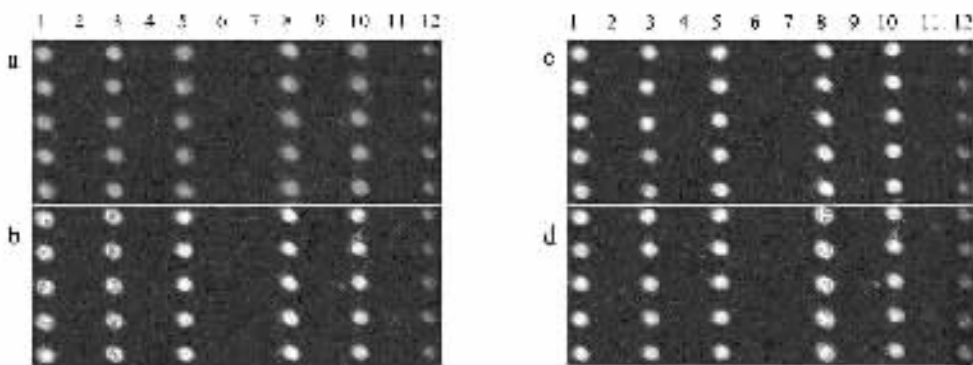


图 3 不同质量模板的分型结果图

Fig. 3 Genotyping results of templates with different qualities

a 纯化的模板 ;b 未纯化且无优化剂的模板 ;c 未纯化且含 DMSO 的模板 ;d 未纯化且含 BSA 的模板。

a : Purified template ; b : Unpurified template without optimization reagent ; c : Unpurified template with DMSO ; d : Unpurified template with BSA.

可以明显的看出 PCR 扩增产物经过蛋白酶 K 消化后即使不进行纯化对分型结果的特异性也没有影响。同时还可以看出未经纯化的信号强度均强于经过纯化的结果,这主要可能是纯化过程会带来部分的产物损失而造成的。据此我们可以省略掉 PCR 扩增产物单独纯化的步骤,从而可以减少操作繁琐程度以及时间。并且也可避免在纯化步骤中产生的人为污染。

3 讨论

对于任何一种分型技术而言,特异性是其中最关键的因素之一,将直接影响到它的推广应用。在相关的几种芯片分型技术中,直接杂交法与芯片上等位基因特异性延伸法的特异性影响因素较复杂,有时极难寻找到其最佳反应条件,而且阳性点与阴性点的荧光信号强度比值并不是很高,从而会给最终的结果判断带来较大的困难。此外杂交法与延伸法在检测重复序列时极易产生假阳性结果,以至限制了它们的适用范围。与此相比,本研究建立的分型技术表现出极高的特异性,且其中阳性点与阴性点的信号强度比值可达 10 以上,以至直接用眼睛即可准确的进行结果判断。同时由于此法利用两条探针的杂交与连接酶的忠实性使分型准确性得以双重确定,故而可进行重复序列区域的多态性检测,其中包括对插入和缺失多态性的检测,从而使得此法具有较广的适合范围。

另外由于芯片上面的 Zip 探针序列并不是和特异性的位点分析直接相关联的,它的互补序列可以添加到任何一组 LDR 特异性探针的后面,因此这种芯片被称为通用(universal)芯片。这种通用性给予研究者很大的自由度去选择所研究的位点,从而无需针对每一探针组进行杂交条件的单独优化。片基从头活化技术更使其摆脱了对商品化片基的依赖,且经优化后探针用量极少。此外由于在一个反应体系内许多探针可以彼此不受干扰的进行各自的连接反应,使得数十个乃至数百个反应可以在一个反应管内完成,保证了其高通量检测的顺利进行。

综上所述,本研究所建立的基于连接检测反应的基因芯片分型技术具有特异性高、通用性好、适用范围广等特点,适用于高通量基因分型。

参 考 文 献

- 1 Marshall A, Hodgson J. DNA chips: an array of possibilities. *Nat Biotechnol*, 1998, 16: 27-31.
- 2 Sevenet N, Cussenot O. DNA microarrays in clinical practice: past, present, and future. *Clin Exp Med*, 2003, 3: 1-3.
- 3 Hacia JG, Brody LC, Chee MS, et al. Detection of heterozygous mutations in BRCA1 using high density oligonucleotide arrays and two-colour fluorescence analysis. *Nat Genet*, 1996, 14: 441-447.
- 4 Fotin AV, Drobyshev AL, Proudnikov DY, et al. Parallel thermodynamic analysis of duplexes on oligodeoxyribonucleotide microchips. *Nucleic Acids Res*, 1998, 26: 1515-1521.
- 5 Hacia JG, Sun B, Hunt N, et al. Strategies for mutational analysis of the large multiexon ATM gene using high-density oligonucleotide arrays. *Genome Res*, 1998, 8: 1245-1258.
- 6 Pastinen T, Kurg A, Metspalu A, et al. Minisequencing: A specific tool for DNA analysis and diagnostics on oligonucleotide arrays. *Genome Research*, 1997, 7: 606-614.
- 7 Head SR, Rogers YH, Parikh K, et al. Nested genetic bit analysis (NGBA) for mutation detection in the p53 tumor suppressor gene. *Nucleic Acids Res*, 1997, 25: 5065-5071.
- 8 Fan JB, Chen X, Halushka MK, et al. Parallel genotyping of human SNPs using generic high-density oligonucleotide tag arrays. *Genome Res*, 2000, 10: 853-860.
- 9 Fan JB, Chen X, Halushka MK, et al. Simple two-color array-based approach for mutation detection. *Eur J Hum Genet*, 2000, 8: 884-894.
- 10 Kurg A, Tonisson N, Georgiou I, et al. Arrayed primer extension: solid-phase four-color DNA resequencing and mutation detection technology. *Genet Test*, 2000, 4: 1-7.
- 11 Pastinen T, Raitio M, Lindroos K, et al. A system for specific, high-throughput genotyping by allele-specific primer extension on microarrays. *Genome Res*, 2000, 10: 1031-1042.
- 12 Pastinen T, Perola M, Ignatius J, et al. Dissecting a population genome for targeted screening of disease mutations. *Hum Mol Genet*, 2001, 10: 2961-2972.
- 13 Busti E, Bordoni R, Castiglioni B, et al. G. Bacterial discrimination by means of a universal array approach mediated by LDR (ligase detection reaction). *BMC Microbiol*, 2002, 2-27.
- 14 Delrio-Lafreniere SA, Browning MK, McGlennen RC. Low-density addressable array for the detection and typing of the human papillomavirus. *Diagn Microbiol Infect Dis*, 2004, 48: 23-31.
- 15 Gerry NP, Witowski NE, Day J, et al. Universal DNA microarray method for multiplex detection of low abundance point mutations. *J Mol Biol*, 1999, 292: 251-262.
- 16 Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol*, 2000, 132: 365-386.
- 17 Erdogan F, Kirchner R, Mann W, et al. Detection of mitochondrial single nucleotide polymorphisms using a primer elongation reaction on oligonucleotide microarrays. *Nucleic Acids Res*, 2001, 29: E36-E42.

(收稿日期 2006-03-23)

(本文编辑:高巍)