

# 双色荧光杂交芯片在检测 CYP1A1 MspI 基因多态性中的应用\*

梁戈玉 浦跃朴 尹立红

(东南大学公共卫生学院 江苏南京 210009)

**摘要 目的:**应用一种新的高通量 SNP 检测方法 - 双色荧光杂交芯片技术检测 CYP1A1 MspI 基因多态性。**方法:**收集江苏汉族人群原发性肺癌患者 75 例和相应对照 77 例,应用双色荧光杂交芯片技术检测了 152 例样本的 CYP1A1 基因 MspI 基因多态性,并应用 PCR-RFLP 技术验证双色荧光杂交芯片的特异性。**结果:**152 例样本的 CYP1A1 基因双色荧光杂交芯片技术分型结果与 PCR-RFLP 结果完全相符,两种方法的基因型分型结果具有很好的一致性。**结论:**双色荧光杂交芯片技术是一个高通量 SNP 检测的良好工具,特异性高,在大规模人群 SNP 筛选中具有良好的发展前景。

**关键词:**单核苷酸多态性;双色荧光杂交;CYP1A1 基因

**中图分类号:**Q503,R318,Q75 **文献标识码:**A

## Application of Dual-color Fluorescence Hybridization chip Technique in the detection of CYP1A1 MspI genetic polymorphism

LIANG Ge-yu, PU Yue-pu, YIN Li-hong

(Department of Occupational and Environmental Health, School of Public Health, Southeast University, Nanjing 210009, Jiangsu, China)

**ABSTRACT Objective:** To detect CYP1A1 MspI genetic polymorphism by a new high-throughput single nucleotide polymorphism (SNP) typing approach of dual-color fluorescence hybridization chip technique. **Methods:** 75 patients with lung cancer and 77 cancer-free controls were collected to detect CYP1A1 MspI polymorphism. Genotypes were analyzed by dual-color fluorescence hybridization chip technique and PCR-RFLP, respectively. The results of dual-color fluorescence hybridization chip were compared with PCR-RFLP to evaluate the genotyping validity. **Results:** The genotyping results of dual-color fluorescence hybridization chip on 152 samples matched the PCR-RFLP results. **Conclusion:** Dual-color fluorescence hybridization chip is a specific and high-throughput tool in SNP genotyping, which has a favorable applied foreground for SNP genotyping in a large population.

**Key words:** SNP; Dual-color fluorescence hybridization; CYP1A1 gene

基因多态性是指人类基因组中存在的基因序列的差异,单核苷酸多态性(Single Nucleotide Polymorphism, SNP)是其中的主要形式。SNP 分布广泛,估计整个人类基因组中共有 300 万个以上的 SNP<sup>[1]</sup>。SNP 作为分子遗传标记,具有高密度、相对稳定和易于检测的优势,因此在多基因疾病研究领域显示了巨大的优势。随着人类基因组计划的进行,越来越多的研究者发现一些 SNP 与肿瘤的易感性具有一定的相关性。从预防医学角度,确定基因与肿瘤的易感性间的关系有助于进行个体化预防。这需要对大量样本进行 SNP 分析,因此建立快速、准确和高通量的 SNP 检测方法十分必要。目前现有的检测人类 SNP 位点的方法有 PCR-限制性片段长度多态性分析(PCR-restriction Fragment Length Polymorphism, PCR-RFLP)、PCR-单链构象多态性分析(PCR single-strand Conformation Polymorphism, PCR-SSCP)等位基因特异性扩增(AAllele Specific Amplification, ASA)等,虽然在小规模人群中可以有效进行 SNP 分析检测,但是对于大量样本同时进行分型检测或者在大规模人群中进行 SNP 筛选,上述这些方法都存在局限性,显得费时费力。日前基

因芯片技术日益发展,它具有自动化、集成化和高速等特点,因此正被逐步广泛应用到基因序列差异性研究中。双色荧光杂交芯片是新近发展起来的一种基于 DNA 微阵列技术进行 SNP 分析的新方法,可以对大量样本进行同时分型。CYP1A1 是非常重要的细胞色素氧化酶,它主要编码芳烃羟化酶(AHH),可激活多环芳烃类(PAH)致癌物,形成酚类和环氧化合物。CYP1A1 基因是目前被研究的最多的一个可能与肺癌易感性相关的基因。本研究应用双色荧光杂交芯片技术原理对 CYP1A1 基因 3'端非编码区 MspI 酶识别的 3801T→C 突变位点进行了分析,并用传统 PCR-RFLP 方法加以验证,探讨双色荧光杂交芯片技术在人群中筛选肺癌易感基因的可行性。

### 1 材料与方法

#### 1.1 研究对象

在南京市胸科医院、鼓楼医院、中大医院收集病理诊断明确的原发性肺癌病例 75 例,男性占 74.7%(56 例)、女性占 25.3%(19 例),平均年龄 62.2 岁。选择除呼吸系统疾病以外的

\* 基金项目:国家自然科学基金(30170791);江苏省社会发展基金(BS2000034);

江苏省高等学校研究生创新计划资助项目(XM04-70)

作者简介:梁戈玉,(1976-),女,博士,讲师,主要研究方向:肺癌相关环境基因组学;

E-mail:lianggcyu@yahoo.com.cn, Tel: 13851720165

通讯作者:浦跃朴教授, E-mail: yppu@seu.edu.cn

(收稿日期 2006-09-27 接受日期:2006-10-23)

非肿瘤同期住院患者 77 例作为对照, 其中男性占 70.1%(54 例), 女性占 29.9%(23 例), 平均年龄 61.1 岁。两组性别年龄分布无显著性差异。

1.2 基因多态性分析

1.2.1 DNA 提取

取静脉血 3ml, 采用改良醋酸钠法<sup>[9]</sup>提取基因组 DNA。

1.2.2 双色荧光杂交芯片原理

双色荧光杂交芯片检测 SNP 的原理<sup>[10]</sup>是将含有所检测 SNP 位点的目的片段进行 PCR 扩增, 扩增后的产物纯化并将其打印在氨基修饰的玻璃片上, 固定处理后与所含 SNP 位点相匹配的荧光标记的探针进行杂交, 最后进行扫描。针对目的 SNP 本研究设计了两条 13 个碱基长度的寡核苷酸探针, 分别用 cy3 和 cy5 进行荧光标记。经过杂交扫描后, 可以看到野生型和突变型样本分别获得一个较强的 cy3 (绿色荧光) 或 cy5 (红色荧光), 杂合型样本既可获得 cy3 荧光又可获得 cy5 荧光, 经过叠加后能显示一个较强的“黄色”荧光。

1.2.3 PCR 扩增引物和杂交探针的设计

根据 CYP1A1 基因序列运用 Primer5.0 软件设计 PCR 扩增用引物和杂交用探针 (表 1)。3801TT 和 3801CC 分别与 CYP1A1 基因 Msp I 酶切位点的野生型和突变型匹配。含有 CYP1A1 Msp I 多态性位点的部分基因序列(表 2)。

表 1 CYP1A1 的引物和探针序列

Table 1 Primers and probes sequences of CYP1A1

CYP1A1		序列 * Sequence
引物 Primers	upstream	5'-AGG GAG GAA GAA GAG GAG-3'
	downstream	5'-TGA GGT GGG AGA ATC GTG-3'
探针 Probes	3801 TT	5'-Cy3-ACCTCCT GGGCTC
	3801 CC	5'-Cy5-ACCTCCC GGGCTC

\* 带下划线字母表示待测的 SNP 位点

\*The letters with underline indicate local mutation of SNP

1.2.4 PCR 扩增

CYP1A1 PCR 反应体积为 100 μl, 内含上、下游引物各 100pmol, 200 μM dNTP, MgCl<sub>2</sub> 1.5 mM, Taq 酶 5 U, DNA 模板 250ng。扩增条件为 95℃ 预变性 5min, 然后经 30 次循环(95℃ 30s, 55℃ 30s, 72℃ 30s), 然后 72℃ 延伸 7min。扩增产物为 182bp。PCR 产物纯化后沉淀干燥, 溶于点样缓冲液中。

1.2.5 样品微阵列的制备

用 PixSys5500 点样仪将 PCR 产物点于制备好的氨基玻片上, 100℃ 热台迅速干燥, 然后 400mJ 条件下紫外交联, 85℃ 烘 2h。

表 2 CYP1A1 DNA 序列

Table 2. DNA sequence of CYP1A1

基因 Gene	序列 5'→3' 1) Sequence 5'→3'
CYP1A1 (T3801C)	1 acaactcctc tattctagcc tgcattgagc ttgcatgctt gcataagagc ttaagaaacc
	61 attgatttaa tgtaataggg aaaattctaa cccaggtatc caaaaatgtg taagaacaac
	121 tacctgagct aaataaagat attgttcaga aaatcctata ggtagagatt tttgaaatca
	181 taaatgatte atcactcgtc taaatactca cctgaacc cattctgtgt tgggttttac
	241 tgtagggagg aagaagagga ggtagcagtg aagaggtgta gccgctgcac ttaagcagtc
	301 tgtttgaggg acaagactct attttttgag acaggggcc caggtcatcc aggctggagt
	361 gcactggtac cattttgtt cactgtaacc tccacctcct gggtcaccac gattctccca
	421 cctcagcctc tgagtagttg gggcccagc agcccaccac agctttttt tttttttt
	481 ttttttttg tagagatggg gtttcacat gttgccagc ctggctcaca actcctgagc
	541 tcaagtgate cacctgcctc agcctcccaa agtgcctgga ttacaggcac gagacaagac
	601 tctaatacac tgtgctgtct tagcgcctc tctaactat cacaaatga

\* 带下划线字母表示碱基变化位点

\*The letters with underline indicate the location of base change

1.2.6 杂交

点样玻片置 94℃ 水浴 3 min, 95% 冰乙醇 10s 后吹干。将 cy3 和 cy5 荧光标记的探针 (1.5pmol/μL) 与杂交液按 3:1 混合, 加到点阵上。置入杂交箱避光杂交 37℃ 2h, 33℃ 2h。杂交后依次用 2XSSC-0.1%SDS, 0.2XSSC-0.1%SDS 清洗玻片, 最后用灭菌去离子水充分清洗玻片, 吹干。

1.2.7 图像扫描与数据处理

将玻片置入芯片分析系统, 分别在 cy3 和 cy5 波长下扫描获得图像, 图像经叠加后分析并判断基因型。CYP1A1 基因型判断: 绿色荧光 (cy3) 为野生型 (TT), 红色荧光 (cy5) 为突变型 (CC), 黄色 (cy3 和 cy5 叠加) 为杂合型 (TC)。

1.2.8 PCR-RFLP 方法验证实验

将 152 例样本全部用传统 PCR-RFLP 方法进行基因型分型, 并将双色荧光杂交芯片技术结果与 PCR-RFLP 结果进行比较。CYP1A1 (Msp I) PCR-RFLP 反应条件: 引物序列为 5'-CAG TGA AGA GGT GTA GCC GCT-3', 5'-TAG GAG ICT TGT CTC ATG CCT-3'; PCR 扩增反应总体积 25 μl, 含有 DNA 75ng, dNTP 200 μmol/l, 引物各 25pmol/l, Mgcl<sub>2</sub> 2.0mmol/l, Taq 酶 1.5 个单位, 1× Buffer, 扩增长度为 340bp; 热循环条件为 95℃ 预变性 5min, 然后经 30 次循环 (95℃ 1min, 60℃ 1min, 72℃ 1min), 最后 72℃ 延伸 7min; PCR 产物用 10 个单位的 Msp I 限制性内切酶 37℃ 消化 4 小时; 用 3.0% 琼脂糖 (含 0.5 μg/ml EB) 将酶切产物进行检测, 电泳后在图像分析仪上观察并判断基因型; 当基因型为纯合野生型时 (TT), 则

扩增产物不能被消化,为 340bp;如基因型为杂合子(TC)时则为 140bp、200bp 和 340bp 三个片段;基因型为纯合突变型(CC)时,则为 140bp 和 200bp 两个片段。

2 结果

2.1 双色荧光杂交芯片技术与传统 PCR-RFLP 技术对 CYP1A1 进行基因分型的结果比较

CYP1A1 基因 152 例样本应用传统 PCR-RFLP 技术进行

分型并与双色荧光杂交芯片法进行对比,结果(图 1)。图 1 中 A 为 CYP1A1 双色荧光杂交芯片结果,每个样重复 8 次,B 为 1-16 号样本对应的 PCR-RFLP 电泳结果,其中泳道 3,6,7,8,12,13 为 CC 基因型图谱,泳道 1,2,9,10,11,14,15,16 为 CT 基因型图谱,泳道 4,5 为 TT 基因型图谱,M 为 100bpMarker。152 例样本 CYP1A1 基因型双色荧光杂交芯片判断结果与 PCR-RFLP 判断结果完全吻合(表 3)。

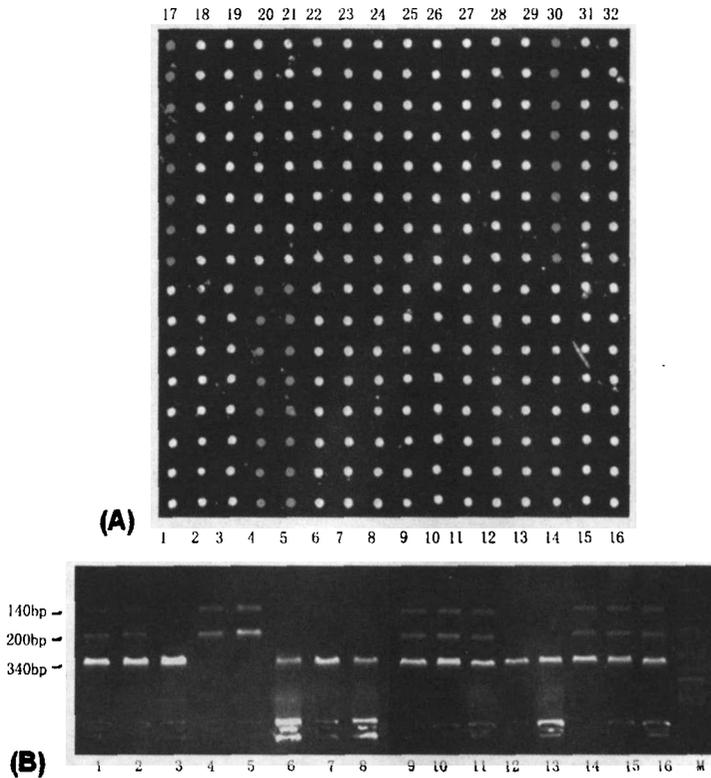


图 1 CYP1A1 芯片杂交结果及 PCR-RFLP 电泳结果对比

Figure 1 The results of dual-color fluorescence hybridization chip and corresponding gel electrophoretograms of PCR-RFLP of CYP1A1

表 3 双色荧光杂交芯片技术与 PCR-RFLP 技术对 CYP1A1 进行基因分型的结果比较

Table 3. Compare the results of dual-color fluorescence hybridization chip to the results of PCR-RFLP in CYP1A1 genotyping

PCR-RFLP 技术 PCR-RFLP method	双色荧光杂交芯片技术 dual-color fluorescence hybridization chip method			合计 Total
	TT	TC	CC	
TT	56	0	0	56
TC	0	78	0	78
CC	0	0	18	18
合计 Total	56	78	18	152

3 讨论

SNP 是指基因组内 DNA 中某一特定核苷酸位置上存在转换、颠换、插入、缺失等变化,人类基因组平均每 500-1000bp 就有 1 个 SNP 发生。人类大部分 SNP 是共有的,但在不同种群中发生的频率不同,位于编码区和调控区的 SNP 能够导致蛋白功能性改变,对表型产生很大的影响,因此 SNP 可以被用来阐明疾病的发病机制。目前普遍认为,SNP 是一个具有高度稳定性的遗传标记,它决定了不同人群疾病易感性和药物治疗敏感性的差异,已成为进行疾病预防和药物筛选的基础<sup>[4,5]</sup>。对 SNP 的检测方法多种多样,传统的方法有限制性酶切片段长度

多态性(RFLP)、单链构象多态性(SSCP)、等位基因特异性扩增(ASA)等。这些技术虽在某种程度上能完成对 SNP 的检测,但应用上也有些不足。RFLP 是目前最常用 SNP 检测方法,该法通过限制性内切酶识别 DNA 序列上的特异位点,然后通过电泳后图谱中不同长度的片段来判断基因型。虽然该法较为廉价,也无需特殊的仪器,但是它的应用受到 SNP 突变位点前后序列的限制,因为并不是所有的 SNP 突变都造成酶切位点的改变,在天然的 SNP 突变中大约只有一半能别限制性内切酶所识别。此外,有些限制性内切酶较为稀有,不易获得,会造成检测成本大为提高。SSCP 的原理是当单链上某一碱基发生改变时会影响该链的构象,可通过非变性凝胶电泳进行分离。

SSCP 较适合 150-450bp 片段 SNP 位点的检测, 检出率可达 90%。但是, 该法结果所受影响因素众多, 如电泳温度、凝胶浓度等均可影响其灵敏度, 因而此法需要一个额外的电泳装置以维持其电泳时的温度。ASA 最早在 1989 年由 Newton 等<sup>[6]</sup>建立, 主要是根据突变位点设计两个 3' 末端不同的引物, 分别与正常 DNA 和突变 DNA 碱基互补, 应用时加入这两种引物之一与互补端引物同时进行两个平行 PCR, 通过 PCR 反应产物判断基因型。此方法是基于耐热 TaqDNA 聚合酶缺乏 3'→5' 外切校正活性的特点进行设计的。基本原理是, 引物 3' 端的特异碱基分别互补于野生型和突变型等位基因的相对碱基, 若此碱基对形成错配, 链延伸反应就会因 3', 5'-磷酸二酯键形成障碍而受阻。该法虽省了酶切步骤, 但是每个样本需要两步 PCR 反应, 而且该法的检出率依赖于反应条件的优化。综上所述, 目前所应用的 SNP 检测方法中, 有些检测过程较为繁琐, 有些影响因素较多, 因此不适合大规模人群的高通量的检测。

近十年来, 基因芯片技术在各研究领域发展迅速并初步得以应用, 该技术工作原理是应用已知核酸序列作为探针与互补的靶核苷酸序列杂交, 通过随后的信号检测进行分析, 从而检测特定基因<sup>[7,8]</sup>。基因芯片技术具有高通量, 高灵敏度, 快速自动等特点, 可以高效地大规模获取相关生物信息。目前基因芯片技术在 SNP 的检测应用中主要分为两种, 一个是对一个研究对象的若干各位点同时进行分析, 一个是对大量样研究对象的同一个位点进行分析。在肿瘤易感基因的研究中需要的是后一种技术, 双色荧光杂交芯片技术即是在针对此新开发的一种用于对大样本同时进行分析的高通量检测工具。CYP1A1 是 CYP450 家族的主要成员, 在外源性化学物质的氧化代谢方面起着重要的作用, 许多化学物质经过 CYP1A1 的氧化作用后毒性增强, 如多环芳烃类(PAH)可通过 CYP1A1 代谢激活具有致癌活性。PAH 广泛存在于香烟烟气和污染的空气中, 是肺癌的主要致病因素之一。CYP1A1 基因位于 15q22-q24, 含 7 个外显子和 6 个内含子。CYP1A1 基因 3' 端非编码区存在 MspI 酶识别的 3801T→C 突变, 突变型个体酶诱导活性显著高于野生型<sup>[9]</sup>。日前许多研究表明 MspI 多态性可能与肺癌易感性有关, 其作为遗传易感性的重要指标日益受到重视<sup>[10-12]</sup>。本次研究中, 我们运用双色荧光杂交芯片技术在对 152 例肺癌和对照样本检测了 CYP1A1 MspI 多态性位点, 并应用 PCR-RFLP 技术对其结果进行验证, 结果表明 152 例样本的 CYP1A1 基因双色荧光杂交芯片技术分型结果与 PCR-RFLP 结果完全相符, 两种方法的基因型分型结果具有很好的一致性。此结果证实双色荧光杂交芯片技术法具有很高的特异性和准确度, 可以替代传统 PCR-RFLP 方法进行基因型检测。该方法最主要的优点是高通量, 可以同时分析上千个样品的基因型, 增强了效率。该法在样

品制备和数据分析上都相对简单, 只需要一个 PCR 和一个杂交过程, 最后扫描后数据处理和分析也并不复杂。该法的另一优点是该法的设计思路可广泛用于其它 SNP 点突变研究, 与 PCR-RFLP 相比没有酶切特异性序列的限制, 因此本法对于普及 SNP 位点的检测具有重要意义。此外, 该法成本相对较低, 虽然荧光标记的探针价格较为昂贵, 但是仅仅需要少量的探针就能对成千上万个样品进行分型, 因此, 样品越多其成本越低。该法的缺点是需要特定的杂交仪器和芯片扫描仪器才能进行试验。此外, 在日前芯片技术并不是十分完善和自动化程度较低的情况下, 该法并不十分节省时间。但随着芯片技术的日趋成熟及其操作的日益自动化和简单化, 该技术必将在大规模人群 SNP 筛选中具有有良好的应用前景。

#### 参考文献

- [1] Sherry ST, Ward MH, Kholodov M, et al. dbSNP: the NCBI database of genetic variation[J]. *Nucleic Acids Res*, 2001, 29:308-311
- [2] 宋雅辉, 尹立红, 蒲跃朴, 等. 南京市人群 DNA 修复基因 XRCC1 多态性与肺癌易感性的关系[J]. *环境与职业医学*, 2004, 21(1):18-21
- [3] Ji MJ, Hou P, Li S, et al. Microarray-based method for genotyping of functional single nucleotide polymorphisms using dual-color fluorescence hybridization[J]. *Mutation Research*, 2004, 548:97-105
- [4] Davidson S. Research suggests importance of haplotypes over SNPs[J]. *Nature Biotechnol*, 2000, 18: 1134-1135
- [5] Brookes AJ. The essence of SNPs[J]. *Gene*, 1999, 234(2): 177-186
- [6] Newton CR, Graham A, Heptinstall L E, et al. Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS)[J]. *Nucleic Acids Res*, 1989, 17: 2503-2516
- [7] Marshall A, Hodgson J. DNA chips: an array of possibilities[J]. *Nature Biotechnol*, 1998, 16:27-31
- [8] 马立人, 蒋中华主编. 生物芯片[M]. 北京: 化学工业出版社, 2000
- [9] Kiyohara C, Nakanishi Y, Inutsuka S, et al. The relationship between CYP1A1 aryl hydrocarbon hydroxylase activity and lung cancer in a Japanese population [J]. *Pharmacogenetics*, 1998, 8(4): 315-323
- [10] Taioli E, Gaspari L, Benhamou S, et al. Polymorphisms in CYP1A1, GSTM1, GSTT1 and lung cancer below the age of 45 years [J]. *Int J Epidemiol*, 2003, 32: 60-63
- [11] Sobti RC, Sharma S, Joshi A, et al. CYP1A1 and CYP2D6 polymorphism and risk of lung cancer in a North Indian population [J]. *Biomarkers*, 2003, 8(5): 415-428
- [12] 梁戈玉, 蒲跃朴, 尹立红. 南京汉族群体肺癌易感性相关基因的研究[J]. *遗传*, 2004, 26(5): 584-588
- [13] Vineis P, Veglia F, Anttila S, et al. CYP1A1, GSTM1 and GSTT1 polymorphisms and lung cancer: a pooled analysis of gene-gene interactions[J]. *Biomarkers*, 2004, 9(3): 298-305