

利用动态等位基因特异性杂交技术进行 单核苷酸多态高通量分型*

曾朝阳¹⁾ 熊 炜¹⁾ 沈守荣²⁾ 朱诗国¹⁾ 李小玲¹⁾ 李伟芳¹⁾ 李 江¹⁾
周 鸣¹⁾ 范松青¹⁾ 马 健¹⁾ 周 洁¹⁾ 肖炳焱¹⁾ 贺 林³⁾ 李桂源^{1)**}

(¹⁾中南大学湘雅医学院肿瘤研究所, 长沙 410078; (²⁾中南大学湘雅三医院消化内科, 长沙 410078;

(³⁾上海交通大学 Bio-X 生命科学研究中心, 上海 200031)

摘要 动态等位基因特异性杂交 (dynamic allele-specific hybridization, DASH) 是新发展起来的一种单核苷酸多态 (single nucleotide polymorphisms, SNP) 等位基因分型技术, 具有快速、经济、准确、高通量、重复性好等优点. 利用 DASH 技术, 对 96 份正常人外周血 DNA 样品成功地进行了两个 SNP 位点的基因分型, 并摸索实验条件, 对该技术进行了优化.

关键词 动态等位基因特异性杂交, 单核苷酸多态, 基因分型

学科分类号 Q53

单核苷酸多态 (single nucleotide polymorphisms, SNP) 是随着人类基因组计划的实施而发展起来的新一代遗传标记, 也是人类基因组多态的主要组成部分, 并已成为功能基因组研究领域的核心科学问题之一. 公共数据库中 SNP 的数目在最近几年中呈指数形式迅猛增长, 美国国家卫生研究院 (National Institutes of Health, NIH) 的 dbSNP (<http://www.ncbi.nlm.nih.gov/SNP>) 于 1998 年建立, 1999 年 9 月登录的 SNP 只有 18 250 个, 2000 年 8 月 15 日为 803 142 个, 到 2002 年 2 月 12 日, dbSNP 数据库中已登录 SNP 4 126 321 个, 大大超过原来预期的 3 000 000 个^[1], 在基因组中达到平均每 700 bp 左右就存在 1 个 SNP 的密度. 由于其分布广、密度高、便于高通量检测而被期望在包括连锁分析和基因定位、疾病的关联研究、多基因疾病的基因定位、发病机理的研究、个体识别和亲子鉴定以及研究生物进化、生物间相互关系等广泛领域中发挥重要作用.

目前各种高通量、自动化检测 SNP 的分型方法发展迅速. 动态等位基因特异性杂交 (dynamic allele-specific hybridization, DASH) 技术^[2,3], 就是最近发展起来的一种自动化、批量化 SNP 分型技术, 由于其快速、经济、准确、高通量、可重复性好而具有广阔的应用前景. 利用 DASH 技术, 我们对 96 份正常人外周血 DNA 样品成功地进行了两个 SNP 位点的基因分型, 并摸索实验条件,

对该技术进行了优化.

1 材料与方法

1.1 材料

1.1.1 样品. 本文所用 96 份正常人外周血样品均取自长沙市中心血站健康无偿献血员, 每份样品均取血 10 ml, ACD 抗凝, 依常规酚-氯仿法抽提高质量白细胞基因组 DNA, 并调配至终浓度 20 mg/L.

1.1.2 引物. DASH 分型引物为 rs1416840 (Forward primer: 5'-Bior caggacatagtttccacat-3', Reward primer: 5'-caaaggtgaagaagtgaagagg-3', Probe 1: 5'-aagaggacag-aattc-3', Probe 2: 5'-aagaggatagaattc-3') 和 rs2275477 (Forward primer: 5'-tctgtggttaagatagtcacac-3', Reward primer: 5'-Bior gagattaagaaccagtgacc-3', Probe 1: 5'-ccggaatctgggctcag-3', Probe 2: 5'-ccggaatccgggctcag-3'), 测序引物为 rs1416840 (Forward primer: 5'-ccttatgtttggcggttctctg-3', Reward primer: 5'-caaagcaattggcacataca-3') 和 rs2275477 (Forward primer: 5'-tttcactcctggtgggctatt-3', Reward primer: 5'-atgtgattgcctcatgtgg-3'), 各引物及探针均由上海博亚生物技术有限公司合成、标记.

* 国家重点基础研究发展规划 (973) 资助项目 (G1998051008) 和国家自然科学基金资助项目 (30100027).

** 通讯联系人.

Tel: 0731-4805446, E-mail: Ligy@public.cs.hn.cn

收稿日期: 2002-03-22, 接受日期: 2002-05-08

1.1.3 试剂、仪器及软件: Hotstar Taq DNA 聚合酶 (QIAGEN); Thermo Hybaid PCR Express PCR 仪、DASH 检测系统、链亲和素包被的微量板、DASH Buffer Gold 及 DASH 4.1C 软件 (Thermo Hybaid); SYBR 绿色荧光染料 I (Molecular Probes Inc.); NaOH (Sigma).

1.2 方法

1.2.1 DASH 技术的原理简介: 首先, 用生物素标记 1 条引物, PCR 扩增目的片段, PCR 产物经生物素结合于链亲和素包裹的 96 孔板壁; 用 0.1 mol/L NaOH 使 PCR 产物变性, 有生物素标记的单链仍结合在 96 孔板中而无标记的单链被洗掉; 加入探针和荧光染料, 探针与 SNP 的一种基因型完全匹配, 而正中一个碱基与另一基因型则不匹配, 荧光染料 SYBR Green I Dye 可特异性地插入 DNA 双链中, 在激发光的作用下发出荧光, 而在单链 DNA 中则不能; 探针和 PCR 产物单链杂交后, 在 DASH 仪中, 96 孔板持续缓慢升温, 同时连续检测各孔中荧光强度, 将温度及各孔中对应的荧光强度输入计算机, 并对荧光强度和温度求导. SNP 基因型与探针完全匹配的样品在较高温度才解链, 因此在较高温度才得到 $d(\text{Fluorescence})/dt$ 峰; 与探针完全不匹配的样品则在较低温度时就得到 $d(\text{Fluorescence})/dt$ 峰; 杂合子则得到两个峰 (参见 2.3).

1.2.2 PCR 反应: DASH 检测 PCR 反应总体积 25 μL , 包括 Tris-HCl (pH 8.3) 10 mmol/L、KCl 50 mmol/L、 MgCl_2 1.5 mmol/L、dNTPs 各 0.2 mmol/L、生物素标记引物 0.16 $\mu\text{mol/L}$ 、未标记引物 0.8 $\mu\text{mol/L}$ 、基因组 DNA 10 ng、Hotstar

Taq DNA 聚合酶 1 U. PCR 反应在 Thermo Hybaid PCR Express 上进行, 反应循环参数为: 首先 94 $^{\circ}\text{C}$ 12 min, 然后 94 $^{\circ}\text{C}$ 15 s、58 $^{\circ}\text{C}$ 30 s (30 个循环), 最后 72 $^{\circ}\text{C}$ 5 min. 测序验证 PCR 依常规条件.

1.2.3 DASH 检测: 将 20 μL PCR 产物与 30 μL 1 $\times$ DASH Buffer Gold 混合, 加入到链亲和素包被的 96 孔板中, 室温放置 1 h; 去掉 96 孔板各孔中的溶液, 加入 50 μL 0.1 mol/L NaOH 溶液, 室温保持 5 min, 移去 NaOH, 再加入 50 μL 0.1 mol/L NaOH 洗涤一次; 加 50 μL 1 $\times$ DASH Buffer Gold, 洗涤后移去, 并重复 1 次; 加 50 μL 含 1.2 $\mu\text{mol/L}$ 探针的荧光缓冲液, 马上封好 96 孔板迅速置于 PCR 仪上进行杂交, 温度 88~22 $^{\circ}\text{C}$, 每秒下降 0.06 $^{\circ}\text{C}$, 然后于 22 $^{\circ}\text{C}$ 保持 3 min, 去掉上步溶液, 加入无探针的荧光缓冲液, 放入 DASH 仪中运行, 温度 35~90 $^{\circ}\text{C}$, 每秒上升 0.03 $^{\circ}\text{C}$.

1.2.4 测序验证 DASH 结果: 每个 SNP 位点的三种等位基因型 (纯合子 1、杂合子、纯合子 2) 各随机抽取 3 份样品, 经常规 PCR、测序, 以验证 DASH 检测的准确性.

2 结果

2.1 SNP 位点选择及引物设计

DASH 技术的关键是探针与 PCR 产物一条链杂交、解链, 因而要求 PCR 产物单链自身不能形成明显的二级结构. 一般来说, 如果 PCR 产物单链分子内形成发夹结构连续超过 4 个碱基, 或者在解链温度下, 分子内二级结构自由能 $\Delta G < -3.0$ 就很难对 SNP 准确分型. 因此, 设计引物时一般尽可能缩短 PCR 产物长度, 使其控制在 45~69 bp

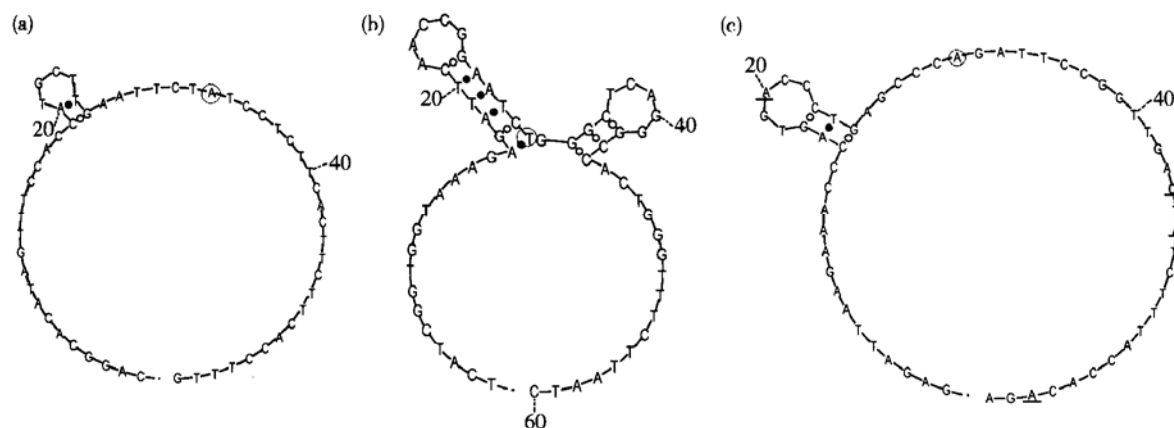


Fig. 1 The most stable target sequence secondary structures (predicted by MFOLD at 50 $^{\circ}\text{C}$)

(a) rs1416840; (b) rs2275477 before modified; (c) rs2275477 modified. The cycle indicated the SNP site and the under lined indicated the modified base.

之间,以减少二级结构形成的机会,引物设计后,将 PCR 产物序列输入 MFOLD (<http://bioinfo.math.rpi.edu/~mfold/dna/form1.cgi>) 中预测二级结构及 ΔG 值,退火温度设为 50°C (探针的解链温度一般在 50°C 左右),其余参数均用默认值。

从 dbSNP 中选取了两个 SNP 位点 rs1416840 和 rs2275477,设计引物后 PCR 产物序列与人类基因组进行 BLAST 匹配均只有一个结果,说明该片段在人类基因组中为单拷贝;将 PCR 产物序列输入 MFOLD, rs1416840 未见明显的二级结构 (图 1a),而 rs2275477 则有明显的发夹结构 $\Delta G = -4.2$ (图 1b),对引物序列进行调整,在尽可能远离 3' 端改动少数碱基后 PCR 产物已符合 DASH 检测要求 (图 1c)。

我们还发现 SNP 的不同等位基因型及同一 PCR 产物的正反链均有可能形成不同的二级结构,有不同的 ΔG 值 (结果未显示),因此在引物设计时均应予以考虑并通过 MFOLD 预测。

2.2 利用 DASH 快速检测 PCR 产物

为了避免 PCR 产物单链形成二级结构,有的引物序列进行了调整,优化 PCR 反应条件时需要检测 PCR 产物质量。Thermo Hyaid 公司推荐用 2% 低熔点琼脂糖胶点样电泳检测,然而实验过程中我们发现,由于 PCR 产物太小 ($45 \sim 69 \text{ bp}$),很难做出准确判断。若不经检测,直接做 DASH,不仅可能由于 PCR 反应不成功而浪费试剂和 96 孔板,而且整个 DASH 程序做完也需要近 2 h。

我们利用 DASH 技术的原理建立了一种简单、

经济、快捷的检测方法,即在用过洗净后的 96 孔板中加入 $2 \mu\text{l}$ PCR 产物和 $30 \mu\text{l}$ 荧光缓冲液,混匀后直接上 DASH 仪运行,若 PCR 反应成功,则较短而均一的 PCR 产物会在 86°C 左右解链,得到一个较高的峰 (图 2); 否则只能得到一系列较低的杂峰。实验证明此法快速、准确,耗费的只有少量荧光缓冲液,20 min 左右就可以得到结果。

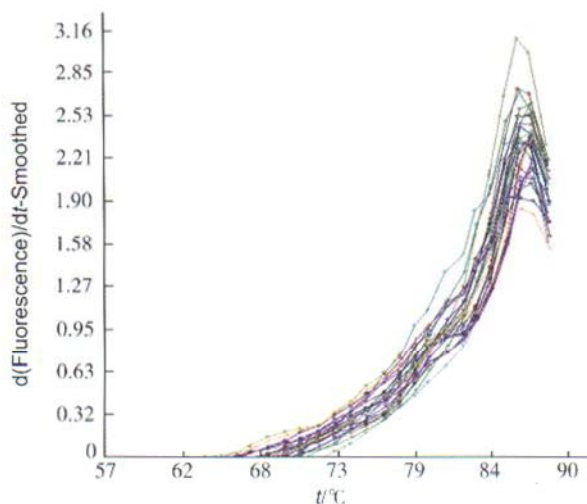


Fig. 2 Testing PCR production by DASH

High quality PCR production denatured at about 86°C and shown a peak of first derivative of the fluorescence versus temperature.

2.3 SNP 分型

利用 DASH 技术对两个 SNP 位点采用对应于不同等位基因型的两条探针,在 96 份样品中均成功地进行了基因分型 (图 3),且同一位点两种探针

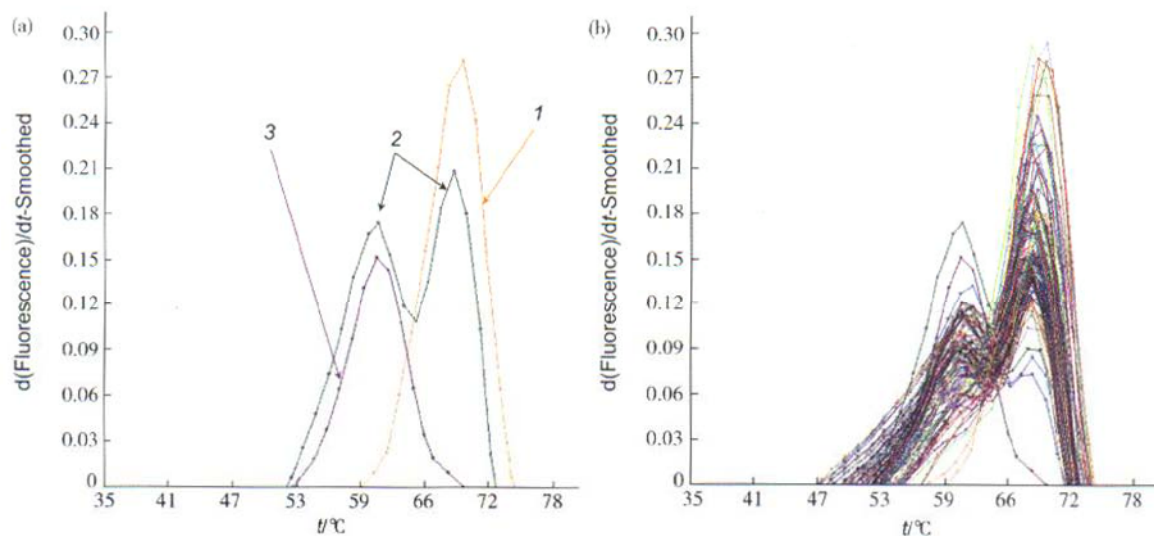


Fig. 3 A demonstration of DASH genotyping curves

(a) Genotyping result of homozygous sample perfectly matched to the utilized probe (1), homozygous sample with a single-base mismatched to the probe (3) and heterozygote (2); (b) 96 samples genotyped by DASH.

得到的结果完全互补。对两个位点共 4 种探针分别随机抽取 24 份 (25%) 样品重复检测, 结果也与第一次完全吻合, 说明利用 DASH 技术对 SNP 进行分型, 结果可重复性好, 较为可靠。

2.4 DASH 的重复检测

对于同一个位点使用分别对应于 SNP 等位基因型的两种探针重复检测, 可以提高 SNP 检测的可靠性。Thermo Hybaid 公司推荐在用第一条探针做完 DASH 检测后, 再用 NaOH 变性将第一条探针洗涤下来, 加入第二条探针, 退火、DASH 检测, 也就是做一次 PCR, 用一块链亲和素包裹的 96 孔板先后对两条探针进行检测。然而我们发现, 第二次检测时荧光强度已经很弱, 不能准确分型, 说明经过 4 次 NaOH 处理, 4 次升温、退火过程, 与 96 孔板壁链亲和素结合的 PCR 产物单链大部分已被洗脱; 不过, DASH 第一次检测后少数孔可能结果不能令人满意, 将 96 孔板缓慢降温退火 (不做任何处理), 再在 DASH 仪上重复检测一次以作为第一次检测的补充, 往往可以得到令人满意的结果。

2.5 DASH 结果测序验证

为了验证 DASH 检测结果的准确性, 我们重

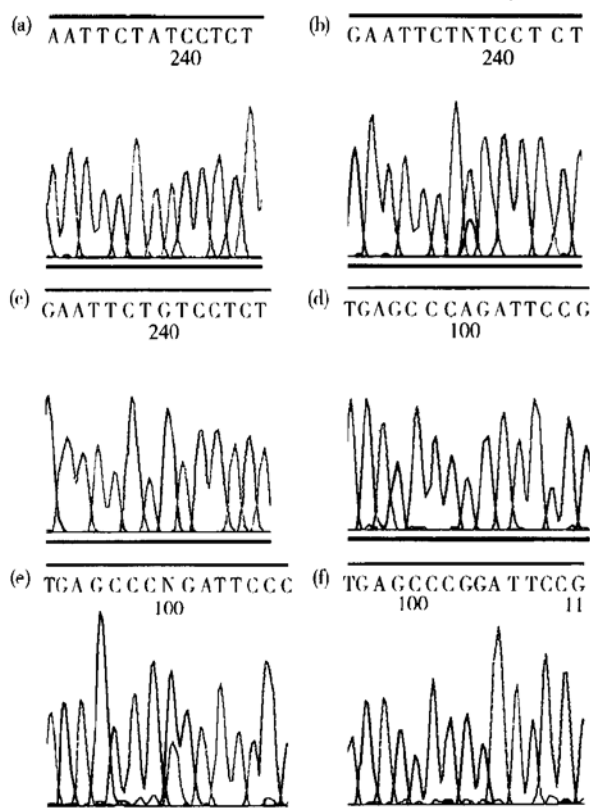


Fig. 4 Genotyping results of SNPs by sequence

(a) ~ (c) Genotyping result of rs1416840. (a) genotype AA; (b) genotype AG; (c) genotype GG. (d) ~ (f) Genotyping result of rs2275477. (d) genotype AA; (e) genotype AG; (f) genotype GG.

新合成了引物, 两个 SNP 位点的 3 种不同等位基因型各随机抽取 3 份样品进行了测序验证, 测序结果见图 4, 测序结果与 DASH 检测完全一致, 进一步证实了利用 DASH 技术进行 SNP 基因分型的准确性。

3 讨 论

SNP 的检测技术已有很多, 如国内比较常用的 PCR-RFLP、单链构象多态性 (single strand conformation polymorphism, SSCP) 分析、异源双链分析 (heteroduplex analysis)、变性梯度凝胶电泳 (denaturing gradient gel electrophoresis, DGGE)、等位基因专一性扩增 (allele specific amplification, ASA) 等, 这些方法相对来说比较简单, 也无需特殊的仪器设备, 一般分子生物学实验室均可以开展。然而, 它们操作步骤较多, 均无法摆脱凝胶电泳, 因而基因分型效率低下, 不利于高通量分析, 且存在对 SNP 位点有特殊要求 (PCR-RFLP、ASA) 及准确性较差 (SSCP、heteroduplex analysis、DGGE、ASA) 等缺点, 因而未能得到大规模应用。

最近也涌现出不少新的高通量 SNP 分型技术, 如实时焦磷酸 DNA 测序 (real-time pyrophosphate DNA sequencing)^[4]、通用能量转移标记引物等位基因特异性 PCR (allele-specific PCR with universal energy-transfer-labeled primers)^[5] 等。然而, 它们或者要求实时检测反应过程, 需要特殊的昂贵仪器, 或者对引物及 PCR 反应要求较高, 而且所用的试剂相对较贵。

DASH 技术则是一种相对准确、经济、高通量的 SNP 分型新方法。DASH 仪操作简单, 检测过程已经自动化, 只要将 96 孔板升温、监测荧光强度而无需依赖酶促反应, 因而干扰因素较少; 对 PCR 无苛刻要求, 只要能扩增出目的片段, 一般都能得到准确的结果; 实验材料和试剂的耗费也相对低廉, 大规模 SNP 检测每份样品费用仅合人民币 4 元左右, 因此不失为一种很有前途的大规模 SNP 分型筛查方法。

致谢 本文曾得到上海市精神卫生中心冯国鄞老师、中国科学院上海生命科学研究杨建东博士、樊金波博士、上海交通大学 Bio-X 生命科学研究谈峥博士、陈刚博士等的热心帮助与指导, 特此致谢!

参 考 文 献

- 1 Anthony J B. The essence of SNPs. *Gene*, 1999, **234** (2): 177 ~ 186
- 2 Mathias H, Magnus J, Ulf G, *et al.* Dynamic allele-specific hybridization a new method for scoring single nucleotide polymorphisms. *Nat Biotech*, 1999, **17** (1): 87~ 88
- 3 Prince J A, Feuk L, Howell W M, *et al.* Robust and accurate single nucleotide polymorphism genotyping by dynamic allele-specific hybridization (DASH): design criteria and assay validation. *Genome Res*, 2001, **11** (1): 152~ 162
- 4 Anders A, Anna K, Ulf H. Determination of single nucleotide polymorphisms by real time pyrophosphate DNA sequencing. *Genome Res*, 2000, **10** (8): 1249~ 1258
- 5 Maxim V M, Yuri K, Stella H, *et al.* High-throughput SNP genotyping by allele-specific PCR with universal energy-transfer labeled primers. *Genome Res*, 2000, **11** (1): 163~ 169

High-throughput Single Nucleotide Polymorphisms Genotyping by Dynamic Allele-specific Hybridization*

ZENG Zhao-Yang¹⁾, XIONG Wei¹⁾, SHEN Shou-Rong²⁾, ZHU Shi-Guo¹⁾, LI Xiao-Ling¹⁾,
LI Wei-Fang¹⁾, LI Jiang¹⁾, ZHOU Ming¹⁾, FAN Song-Qing¹⁾, MA Jian¹⁾,
ZHOU Jie¹⁾, XIAO Bing-Yi¹⁾, HE Lin³⁾, LI Gui-Yuan^{1)**}

¹⁾ Cancer Research Institute, Xiang-Ya Medical School of Central South University, Changsha 410078, China;

²⁾ 3rd Xiang-Ya Hospital Central South University, Changsha 410078, China;

³⁾ Bio-X Life Science Research Center, Shanghai Jiaotong University, Shanghai 200031, China)

Abstract Dynamic allele-specific hybridization (DASH), a new method for rapid, economical, accurate, repeatable and high-throughput genotyping of single nucleotide polymorphisms (SNP), has been used successfully for genotyping 2 SNPs in 96 samples. Experimentation technology has been optimized. Primer design criteria, assay validation and its advantage have also been discussed.

Key words dynamic allele-specific hybridization, single nucleotide polymorphisms, genotyping

* This work was supported by grants from The Special Funds for Major State Basic Research of China (G1998051008) and The National Natural Sciences Foundation of China (30100027).

** Corresponding author. Tel: 86-731-4805446, E-mail: Ligy@public.cs.hn.cn

Received: March 22, 2002 Accepted: May 8, 2002