

高通量飞行时间质谱基因分型方法的研究*

赵辉^{1,2)**} 王威¹⁾ 张清润¹⁾ 高扬^{1,2)} 赵洪斌^{1,2)} 周珺¹⁾ 林伟^{1,2)} 曾长青¹⁾

¹⁾中国科学院北京基因组研究所暨北京华大基因研究中心, 北京 101300;

²⁾中国科学院研究生院, 北京 100040)

摘要 为了考察飞行时间质谱基因分型方法 (MALDI-TOF) 的位点分型成功率和分型结果质量的关系, 分析了 96 个 SNPs 位点的近 10 000 个基因分型数据 (用 MALDI-TOF “4 重” 实验方法检测). 结果显示, 位点分型成功率和分型结果的质量显著正相关. 分型成功率低于 82% 的 SNP 位点, 其高质量结果占的比例开始逐渐降低. 提示 82% 的分型成功率可以作为衡量分型结果质量的数据点. 为了进一步提高通量并降低成本, 在 MALDI-TOF “4 重” 实验方法的基础上, 发展了两种 “准 8 重” 实验方法. 用新的实验方法检测了 95 个样本的 32 个 SNPs 位点. 结果显示 “混合准 8 重” 实验方法与 “4 重” 实验方法相比无显著差异, 而 “复点准 8 重” 的结果差于 “4 重” 分型方法.

关键词 单核苷酸多态性 (SNP), 飞行时间质谱, 基因分型

学科分类号 Q52

单核苷酸多态性 (single nucleotide polymorphism, SNP) 是基因组上广泛分布并能稳定遗传的 DNA 序列多态性. 大多数 SNP 位点只有两个等位基因, 易于自动化检测^[1]. 另外, 存在于编码区及调控区的 SNP 位点可以改变基因的表达水平和表达产物的功能, 因此成为遗传学、人类学、药物基因组学等领域常用的遗传多态位点^[2-6].

目前有很多 SNP 基因分型方法^[7]. 其中大规模高通量的基因分型方法以 Affimatrix 公司的 GeneChip^[8] 技术和 Illumina 公司的 BeadArray 技术为代表^[9]. 中等通量 SNP 分型的代表技术是基于 “引物延伸法” 的 SEQUENOM 公司的飞行时间质谱 (MALDI-TOF) 基因分型技术^[10-13] 和 PE 公司的荧光偏振光基因分型技术^[13,14]. 引物延伸法指延伸引物在待检测 SNP 位点上延伸一个或若干个碱基, 然后根据延伸产物所带荧光的不同或其分子质量的不同而确定其基因型. MALDI-TOF 是在引物延伸法的基础上, 结合 “基质辅助激光解吸电离飞行时间质谱技术” (matrix assisted laser desorption/ionization time of flight mass spectrometry, MALDI-TOF)^[15] 发展起来的 SNP 分型方法. 课题组用 MALDI-TOF “4 重” SNP 实验方法^[16], 对国际 “人类基因组单体型图计划” 协作组^[17] 随机抽取的 1 500 个 SNPs 位点进行分型检测, 有 1 183 个位点被成功分型, 分型准确率是 99.55%, 达到了国际同行的先进水平.

在 MALDI-TOF “4 重” SNP 分型实验中, 观察到不同位点有不同的分型成功率, 且高成功率位点的高质量分型结果较多. 因此, 分析了 96 个 SNPs 位点 (95 个样本) 的近 10 000 个分型数据 (4 重反应检测), 来研究位点分型成功率和位点质量之间的关系. 在 “4 重” 实验方法的基础上, 为了进一步提高通量和降低成本, 设计了新的高重反应分型方法. 新设计的方法被称为 “混合准 8 重” 和 “复点准 8 重” 实验方法. 用这两种 “准 8 重” 实验方法检测 95 个人 DNA 样本的 32 个 SNPs 位点, 并与 “4 重” 实验方法的相同样品和位点的结果做比较, 以验证两种 “准 8 重” 实验方法的实际可行性.

1 材料与方法

1.1 DNA 样品

实验样本由美国 Coriell 医学研究所 (Coriell Institute for Medical Research) 提供, 编号为 HAPMAPPT01. 样品包含 30 个欧裔三体家系和 5 个重复, 共计 95 份个体 DNA 样本和一个阴性对

* “十五” 国家重大科技专项 (2002BA711A09) 和中国科学院知识创新工程重大项目 (KSCX2-SW-207).

** 通讯联系人.

Tel: 010-80481102, Fax: 010-80498676

E-mail: zhaoh@genomics.org.cn

收稿日期: 2005-02-05, 接受日期: 2005-03-28

照. 详细情况参见网址 <http://ccr.coriell.org/nigms/nigms.cgi/panel.cgi?id=2&query=HAPMAP01>.

1.2 基因分型引物

本试验选取的 SNP 位点, 是国际“人类基因组单体型图计划”中国卷项目的工作位点. 分型引物用 MassARRAY Assay Designer 软件 (SEQUENOM), 按 8 重分型反应设计. 所设计的 8 重引物, 可按低重数任意组合, 如分成两套 4 重引物按 4 重实验方法做反应, 或直接做单重 SNP 分型反应. 我们选了 32 个 SNPs 位点的 8 套 4 重引物做新方法比较试验. 每个 SNP 位点对应两条 PCR 引物和一条延伸引物 (IDT 生产).

1.3 PCR 及引物延伸反应以及后续检测

使用热启动 Taq 酶 (Takara 公司, 大连), 在 384 孔板上, 常规扩增待测片段 (45 个循环). 然后在 PCR 产物中加入 2 μ l 虾碱性磷酸化酶 (SAP, SEQUENOM) 消化液 (双蒸水 1.53 μ l, hME 缓冲液 0.17 μ l, SAP 0.474 U), 以除去未用尽 dNTP. 消化反应在 PCR 仪 (Biomtra T-1) 上进行 (37°C 60 min; 85°C 10 min). 而后根据 SEQUENOM 程序进行引物延伸反应 (60 个循环) 和延伸产物纯化反应 (SEQUENOM 说明书 2002 版). 用纳升点样仪 (SpectroPOINT, SEQUENOM) 将纯化后的产物分点在质谱微阵列芯片上 (SpectroCHIP, SEQUENOM). 最后进行实验设计的“4 重”、“混合准 8 重”和“复点准 8 重”点样. 点出 2 块“4 重”反应芯片, 1 个“混合准 8 重”芯片和 1 个“复点准 8 重”芯片. 在 SpectroTYPER RT 操作程序 (SEQUENOM) 中, 编辑相对应的引物和样本, 并按顺序读取数据.

1.4 结果分析及处理

MassARRAY TYPER 软件 (SEQUENOM) 根据引物延伸反应产物质谱峰的位置确定其分子质量, 从而得出 SNP 分型结果; 同时根据质谱峰的信噪比、峰尖的位置、峰的宽度等参数判定分型结果的质量值, 即高质量 (conservative)、中等质量 (moderate) 和低质量 (aggressive) 三种结果. 然后使用 SPSS 12.0 软件进行统计学分析.

2 结 果

2.1 位点分型成功率和分型结果质量正相关

在 MALDI-TOF “4 重” SNP 分型实验中, 发现一些高分型成功率的位点, 其高质量结果也较

多. 为了考察飞行时间质谱基因分型方法的位点分型成功率和分型结果质量的关系, 分析了 96 个 SNPs 位点的近 10 000 个基因分型数据 (用“4 重” SNP 实验方法检测). 首先用每个位点的理论高质量结果和位点的分型成功率作散点图 (图 1). 然后选用 SPSS 12.0 拟合曲线功能的 Lowess 模块, 拟合散点图的曲线. 理论高质量结果用公式 $N_E = 95 \times (\frac{N}{N_T})$ 换算. 其中 N_E 指理论高质量结果, N_T 指该位点成功分型的样本数, N 是该位点实际的高质量结果数.

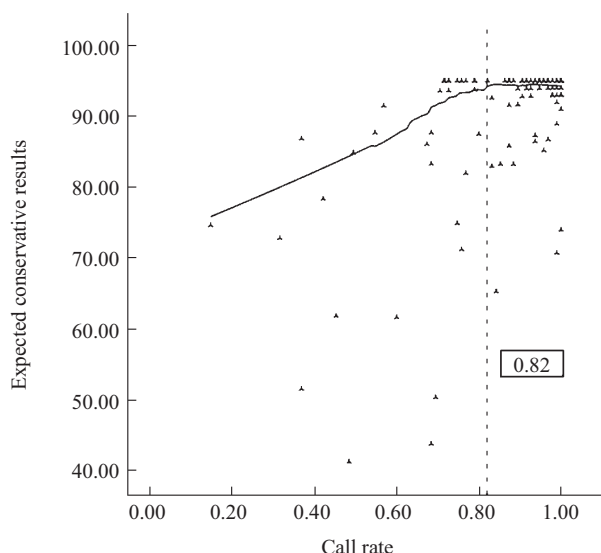


Fig.1 The count of expected conservative results is plotted against call rate

Y-axis is the count of expected conservative results, X-axis is call rate.

The fit line was predicted by SPSS 12.0 with Lowess model.

大多数观察点集中在图 1 的右上角, 说明高分型成功率的位点, 分型结果的质量较高. 进一步以每个位点的分型成功率为自变量, 取其理论高质量结果 (N_E) 为因变量, 做线性回归分析 (linear regression analysis). 得到回归系数为 0.535, P 值小于 0.001, 说明分型成功率与高质量结果显著正相关. 从拟合曲线上看, 82% 的成功率是一个分界点. 成功率高于 82% 的位点, 拟合曲线近似平行所对应的理论高质量结果数较高. 只是当位点的分型成功率低于 82% 后, 拟合曲线才开始逐渐降低, 意味着高质量分型结果数所占的比例开始降低.

2.2 质谱微阵列芯片点样的改进及结果比较

按 SEQUENOM 说明书的要求, 一块芯片只能点一块反应产物, 而且点样程序也只针对单次点样

设计. 新设计的高重反应改变了点样方法. “混合准 8 重”是将两板 4 重反应产物在一个 384 板内等比混合, 使其成为 “8 重” 反应产物, 然后按照正常的点样程序, 将混合后的反应产物点在质谱微阵列芯片上. “复点准 8 重”是将两板 4 重反应产物先后点在同一块质谱微阵列芯片上.

分型反应产物点在质谱微阵列芯片上后, 形成结晶的好坏直接和质谱峰图及分型结果质量判定相关. 两板 “4 重” 反应产物分别以 70 mm/s (点样体积检查程序确定) 的点样速度点在两块质谱微阵列芯片上, 形成的结晶体大小均匀, 外形较一致, 少有破碎点(图 2a), 读取的质谱峰图信噪比好, 分型结果明确, 如图 3a 和 3b 所示(8 个 SNPs 位点). “混合准 8 重” 反应产物以 75 mm/s (点样体积检查程序确定) 的点样速度点在一块质谱微阵列芯片上, 点样后的结晶体大小也很一致, 少有破碎点(图 2b), 质谱峰图信噪比良好, 峰高略低于 “4 重” (图 3c). 至于 “复点准 8 重”, 因为要在同一块芯片上重复点两次样品, 而 “点样体积检查程序” 只针对单次点样设计, 所以不能用点样程序来确定合适的点样速度. 经多次实验, 确定 “复点准 8 重” 的两次点样速度为 50 mm/s. “复点准 8 重” 第一

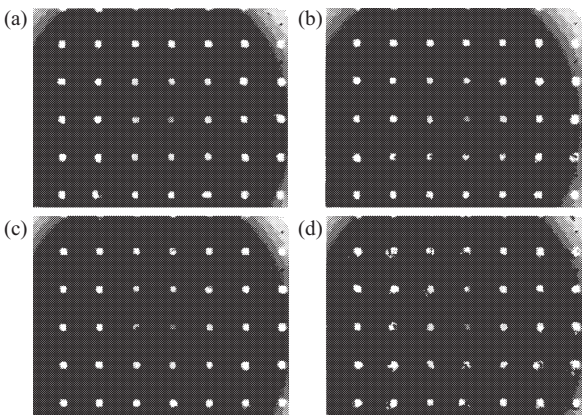


Fig.2 The images of crystal tabs on part of SpectroCHIP for three protocols

(a) The crystal tabs of “4-plex” ; (b) The crystal tabs of “Mix 8-plex” ; (c) The crystal tabs of “Double-spotting 8-plex” when reaction products were dispensed firstly; (d) The crystal tabs of “Double-spotting 8-plex” when reaction products was dispensed secondly.

次点样后的结晶形态较好(图 2c), 但第二次点样后结晶体的形态各异, 大小明显不一致, 部分晶体破碎(图 2d). 从质谱峰图也可以看出 “复点准 8 重” 的产物峰高低于 “4 重” 和 “混合准 8 重” (图 3d).

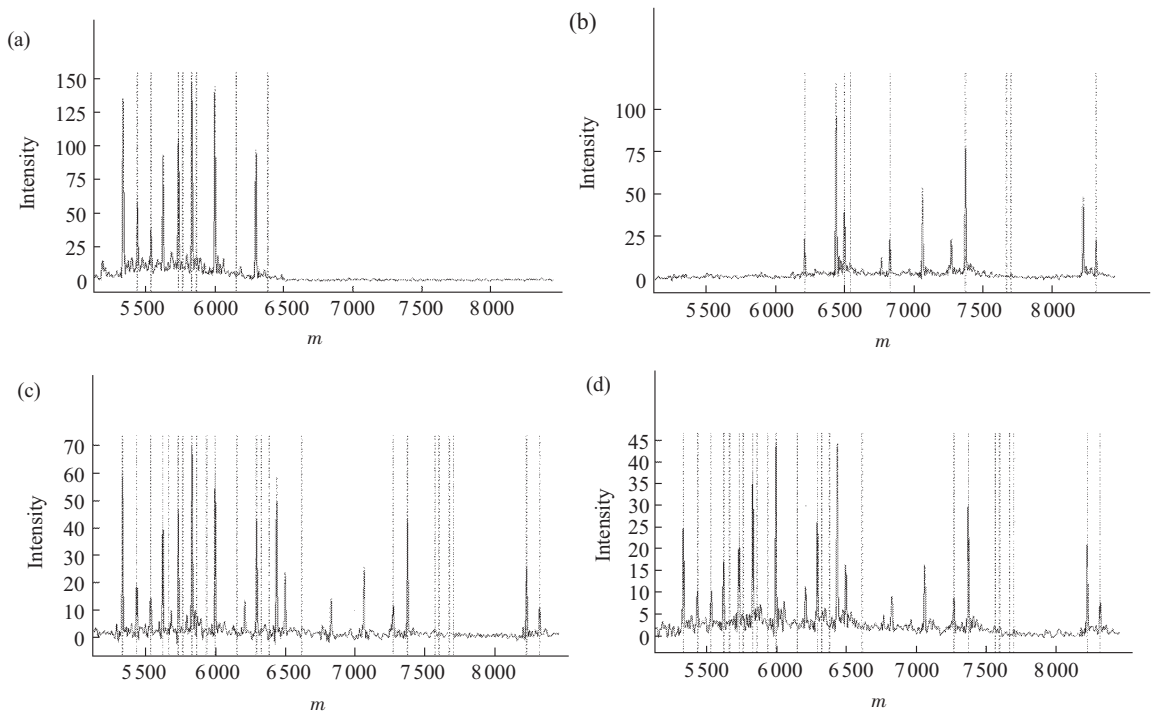


Fig.3 Mass spectrums of three protocols

(a),(b) Mass spectrum of “4-plex” ; (c) Mass spectrum of “Mix 8-plex” ; (d) Mass spectrum of “Double-spotting 8-plex” .

2.3 “混合准 8 重”基因分型结果与“4 重”实验方法的结果相比无显著差异

本试验所用样本为置于同一个 96 孔板的 90 个不同个体 DNA (30 个三体家系), 以及 5 个重复和一个阴性对照. 其中 5 个重复和 30 个三体家系可分别作为两个阳性对照, 以检查分型结果是否有重复样本不一致, 以及三体家系样本分型结果是否符合孟德尔遗传规律. 一个阴性对照可以判断实验过程中是否有交叉污染. 我们用自己编写的质控软件对“4 重”、“混合准 8 重”和“复点准 8 重”的分型结果进行分析, 未发现三种实验方法的分型结果有阴性污染、重复错误和家系检验错误, 且分型结果均符合哈代 - 温伯格平衡. 对于同一样本的同一个位点, 我们亦未发现三种方法分型结果的不符.

进而分析三种实验方法的 SNP 分型结果, 以判断两种“准 8 重”实验方法的结果与“4 重”相比是否显著不同. 具体是考察每一个位点的分型成功率(call rate)和分型结果的质量. 分型成功率 (call

rate) 是成功分型样本数和总样本数的比值, 而分型结果被 SpectroTYPER 软件自动分为高质量、中等质量和低质量三种. 结果显示, 三种实验方法的平均分型成功率是 88.5% (4 重), 87.5% (混合准 8 重) 和 79.5% (复点准 8 重), 分型结果一致性为 100%. 这里提到的一致性指不同实验方法检测同样的样本和位点得到相同结果的概率. “混合准 8 重”和“4 重”一样, 分型结果达到了国际“人类单体型图计划”协作组所要求的大于 80% 的分型成功率和大于 99% 的一致性 (concordance). 而“复点准 8 重”实验方法的平均分型成功率较低 (表 1). 用“配对 T 检验”方法两两比较三种实验方法的分型数据, 结果显示: “混合准 8 重”的分型成功率和高质量结果与“4 重”相比无显著差异 (P 值均大于 0.05). 而“复点准 8 重”的分型成功率和高质量结果均显著低于“4 重”实验方法 (P 值均小于 0.01) (表 2).

Table 1 Genotyping results of three protocols

Descriptive statistics	The results of “4-plex” protocol			The results of “mix 8-plex”			The results of “Double-spotting 8-plex”		
	Conservative	Moderate and aggressive	Total	Conservative	Moderate and aggressive	Total	Conservative	Moderate and aggressive	Total
Mean	82.44	1.66	84.09	81.25	1.91	83.16	70.59	4.94	75.53
Std. deviation	12.16	2.71	11.08	11.56	3.42	10.20	24.90	9.79	19.88

Table 2 The difference between the genotyped results of three protocols

Pairs of variables	Paired-samples T test		Paired-samples correlations	
	T value	2-tailed P value	Correlation	2-tailed P value
All results: “4-plex” vs. “Mixed 8-plex”	1.602	0.119	0.9555	<0.001
All results: “4-plex” vs. “Double-spotting 8-plex”	2.911	0.007	0.548	<0.001
Conservative results: “4-plex” vs. “mixed 8-plex”	1.836	0.075	0.954	<0.001
Conservative results: “4-plex” vs. “Double-spotting” 8-plex	2.931	0.006	0.405	0.021

3 讨 论

在 SNP 基因分型实践中, 位点的分型成功率差异很大. 那么哪些位点被使用, 哪些位点重做, 哪些位点放弃, 需要综合考虑分型成功率, 数据缺失和结果的质量. 从对近 10 000 个 SNPs 分型数据的分析中可以看出, 分型成功率与高质量结果显著正相关. 提示在研究实践中, 尽量采用高分型成功率的位点, 能减少错误分型的数据被使用的概率. 从图 1 拟合的曲线中, 观察到分型成功率高于

82% 的位点, 其高质量结果所占比例较高, 而成功率低于 82% 的位点其高质量结果开始逐渐降低. 因此, 作者认为 82% 的分型成功率是衡量数据质量的一个数据点. 应当指出, 本文中提到的高质量结果、中等质量结果和低质量结果是 MassARRAY TYPER 基因分型软件的自动分类, 并不意味着中等质量和低质量结果的正确率很低. 实际上, 本次实验中, 没有发现三种方法分型结果的不符, 而且三种实验方法的结果均未出现阴性污染、重复错误

和家系检验错误. 说明了 MALDI-TOF 基因分型技术的高可靠性.

MALDI-TOF 多重反应的原理是通过调整不同 SNP 位点的引物, 设计出一组分子质量不同的延伸产物, 通过质谱测定确认每个位点的基因型. 在“4 重”反应基础上直接增加反应重数较为直接, 但有以下困难. 首先, 过多的引物在一个反应体系中容易形成引物二聚体, 既影响 PCR 扩增的效率, 也会增加引物设计的难度. 其次, 高重反应的 PCR 循环次数也要增加, 从而延长扩增时间. 还有, 反应重数的提高要求不同延伸引物的浓度高度一致, 而大多数商品引物难以达到这个要求. 因此, 直接提高反应重数的瓶颈主要是前期的生化反应, 即目标序列扩增和引物延伸反应. 新的实验方法依然按“4 重”方法做 PCR 扩增和引物延伸反应, 但通过混合或重点的技术将点样和检测的重数提高到 8 重. 这样的设计既保留了成熟的 4 重生化反应的高成功率, 又降低了价格昂贵的质谱微阵列芯片的用量. 从实验结果可以看出, “混合准 8 重”点样后芯片的结晶图、质谱峰图和实际的基因分型结果与“4 重”实验方法比较无显著差异, 说明了即使反应产物被稀释 1 倍(两个 4 重反应等比混合), 被检测位点依然能被准确分型, 体现了质谱检测的高灵敏度. 至于“复点准 8 重”, 分析其结果较差的原因有以下两点: 第一, “复点准 8 重”无法确定合适的点样量. 第二, 重复点样时点样针对样品点的机械撞击, 会损伤第一次点样后形成的结晶体, 从而影响质谱峰图的质量.

参 考 文 献

- Brookes A J. The essence of SNPs. *Gene*, 1999, **234** (2): 177~186
- Collins A, Lonjou C, Morton N E. Genetic epidemiology of single-nucleotide polymorphisms. *Proc Natl Acad Sci USA*, 1999, **96** (26): 15173~15177
- Kruglyak L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat Genet*, 1999, **22** (2): 139~144
- Lander E S, Linton L M, Birren B, *et al.* Initial sequencing and analysis of the human genome. *Nature*, 2001, **409** (6822): 860~921
- Riley J H, Allan C J, Lai E, *et al.* The use of single nucleotide polymorphisms in the isolation of common disease genes. *Pharmacogenomics*, 2000, **1** (1): 39~47
- John S, Shephard N, Liu G, *et al.* Whole-genome scan, in a complex disease, using 11 245 single-nucleotide polymorphisms: comparison with microsatellites. *Am J Hum Genet*, 2004, **75** (1): 54~64
- Syvanen A C. Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nat Rev Genet*, 2001, **2** (12): 930~942
- Matsuzaki H, Loi H, Dong S, *et al.* Parallel genotyping of over 10 000 SNPs using a one-primer assay on a high-density oligonucleotide array. *Genome Res*, 2004, **14** (3): 414~425
- Gunderson K L, Kruglyak S, Graige M S, *et al.* Decoding randomly ordered DNA arrays. *Genome Res*, 2004, **14** (5): 870~877
- Jurinke C, Oeth P, van den Boom D. MALDI-TOF mass spectrometry: a versatile tool for high-performance DNA analysis. *Mol Biotechnol*, 2004, **26** (2): 147~164
- Storm N, Darnhofer-Patel B, van den Boom D, *et al.* MALDI-TOF mass spectrometry-based SNP genotyping. *Methods Mol Biol*, 2003, **212**: 241~262
- Garcia B A, Heaney P J, Tang K. Improvement of the MALDI-TOF analysis of DNA with thin-layer matrix preparation. *Anal Chem*, 2002, **74** (9): 2083~2091
- Luo C, Deng L, Zeng C. High throughput SNP genotyping with two mini-sequencing assays. *Acta Biochim Biophys Sin (Shanghai)*, 2004, **36** (6): 379~384
- Luo C, Deng L, Zhou J, *et al.* High throughput SNP genotyping with FP-TDI. *Progress in Biochemistry and Biophysics*, 2004, **31** (8): 731~735
- Karas M, Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10 000 daltons. *Anal Chem*, 1988, **60** (20): 2299~2301
- Nakai K, Habano W, Fujita T, *et al.* Highly multiplexed genotyping of coronary artery disease-associated SNPs using MALDI-TOF mass spectrometry. *Hum Mutat*, 2002, **20** (2): 133~138
- The International HapMap Consortium. *Nature*, 2003, **426** (6968): 789~796

The Study of High Throughput MALDI-TOF Genotyping Assay*

ZHAO Hui^{1,2)**}, WANG Wei¹⁾, ZHANG Qing-Run¹⁾, GAO Yang^{1,2)},
ZHAO Hong-Bin^{1,2)}, ZHOU Jun¹⁾, LIN Wei^{1,2)}, ZENG Chang-Qing¹⁾

¹⁾Beijing Genomics Institute, Beijing 101300, China;

²⁾Graduate School of The Chinese Academy of Sciences, Beijing 100040 China)

Abstract MALDI-TOF is a simple and robust method for the analysis of single nucleotide polymorphism (SNP), which combines proven high-fidelity enzymatic procedures and matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry. This technology is among the most promising SNP typing methods for delivering accurate, effective, flexible and high throughput analysis of SNPs. “4-plex” genotyping protocol, which could detect 4 SNPs in one reaction, was performed in Beijing Genomics Institute with accuracy of 99.55%. To learn the correlation between genotyping call rates and quality, nearly ten thousands of typing results (“4-plex”) were analyzed. Significant positive correlation between call rate and the counts of conservative results were observed. SNPs with call rate lower than 82% have less conservative results, suggesting 82% was a cutoff to evaluate the genotyping results of MALDI-TOF assay. Aiming to further improve genotyping throughput and reduce the cost, “Mix 8-plex” and “Double-spotting 8-plex” genotyping protocols were developed. For new protocols, PCR and primer extension reactions were still performed under reliable “4-plex” protocol. For “Mix 8-plex” protocol, two sets of “4-plex” primer extension products were mixed and dispensed onto one SpectroCHIP at the same time. As to “Double-spotting 8-plex” protocol, two sets of “4-plex” products were consecutively dispensed onto one SpectroCHIP. 32 SNPs were genotyped in 95 human DNA samples to test the feasibility of these new protocols. The results showed that the performance of “Mix 8-plex” was as good as “4-plex” protocol, while the performance of “Double-spotting 8-plex” was poor.

Key words single nucleotide polymorphism (SNP), MALDI-TOF, genotyping

*This work was supported by grants from The Mega-projects of Science Research for the 10th Five-year Plan (2002BA711A09) and The Knowledge Innovation Program (KIP), Important Project of The Chinese Academy of Sciences.

**Corresponding author . Tel: 86-10-80481102, Fax: 86-10-80488676, E-mail: zhaoh@genomics.org.cn

Received: February 5, 2005 Accepted: March 28, 2005