ISSN 0582-9879 Acta Biochimica et Biophysica Sinica 2004, **36**(6): 379–384 **CN** 31-1300/Q

High Throughput SNP Genotyping with Two Mini-sequencing Assays

Chunqing LUO[1,2#], Libin DENG[1#], and Changqing ZENG[1,2*]

[1]*The Institute of Genetic and Developmental Biology, the Chinese Academy of Sciences, Beijing 100101, China;*
[2]*Beijing Genomics Institute, the Chinese Academy of Sciences, Beijing 101300, China*

Abstract Two mini-sequencing methods, FP-TDI (template-directed dye-terminator incorporation with fluorescence-polarization) and MassArray (matrix assisted laser desorption ionization time of flight detection mass spectrometry), were optimized. A numeric standard was introduced to evaluate the SNP scoring quality of FP-TDI assay, thus made the optimization work easier. At the same time, using multi-PCR technology, 8-plex genotyping of MassArray assay was successfully carried out, some softwares were developed and the data process of MassArray was highly automated. Then these two methods were applied to high throughput SNP genotyping, the accuracy, efficiency and robustness were compared. The result shows FP-TDI is more sensitive to the concentration of SNPprimer and PCR product, as well as extension cycles, the SNPprimer length of FP-TDI should be 24–30 bp long, whereas MassArray assay prefers to be as short as only 16 bp. Altogether 6440 SNP sites of human chromosome 3 were genotyped in a sample of 90 individuals, 4792 sites by FP-TDI assay and 1648 sites by MassArray assay, the success rates of FP-TDI and MassArray were 67.7% and 93.6% respectively. The throughput of MassArray was higher than FP-TDI, and the cost of MassArray was lower, MassArray was more suitable for high throughput SNP genotyping.

Key words SNP genotyping; FP-TDI; MassArray; high throughput

Single nucleotide polymorphisms (SNPs) are very important markers that can be used in many areas such as evolutionary genetics [1], disease-susceptibility genes [2,3], personalized medicine and forensics. Only about 20% of human polymorphisms are length polymorphisms, whereas about 80% of human polymorphisms are SNPs. Kruglyak *et al*. [4] reported that there were about 11,000,000 SNPs in the world population.

There are many kinds of SNP genotyping technology [5,6]: some are only suitable to low throughput genotyping [7], some are preferred when a small number of SNPs is tested in a large population [8], and some using modified primers are too expensive [9]. How to genotype large numbers of SNPs on thousands of individuals with a simple, robust, and cost-effective method is badly needed, but until now, no technology has become a widely accepted standard on this purpose.

Polymerase-mediated single base primer extension (mini-sequencing) is one of the most important SNP genotyping technologies [10]. In this method, first about 100 bp DNA fragment around the SNP site was amplified, then another primer immediately upstream the SNP site was designed, the primer was extended by one specific base as dictated by the target PCR product at the polymorphic site in the presence of DNA polymerase and appropriate ddNTP, and the alleles present in the target DNA could be inferred by determining which ddNTP was incorporated [11].

Many detection platforms have been developed for this method [12–14], including FP-TDI [15] and MassArray [13–15]. Both assays are easy and fast, do not need specialized probe and purification of PCR product, and have great promises [16,17]. FP-TDI uses two different fluorophore attached ddNTPs to carry out the homogenous primer extension reaction, fluorophore attached ddNTPs produce low polarization, but after incorporated into the primer, the fluorophore polarization (FP) strengthened, and the polarization change was different with different fluorophore. By detecting the polarization change, the

Received: February 6, 2004 Accepted: May 20, 2004
This work was supported by the grants from the Knowledge Innovation Program (KIP), Important Project of Chinese Academy of Science (KSCX-SW-207) and the Special Funds for Major State Basic Research of China (2002BA711A09)
[#]These two authors contributed equally to this work
*Corresponding author: Tel, 86-10-80484976-1; Fax, 86-10-80494199; E-mail, czeng@genomics.org.cn

incorporated ddNTP can be determined. MassArray uses three ddNTPs and one dNTP to carry out the primer extension reaction according to the two alleles of the polymorphic site, one allele will extend only one base, and the other allele will extend two or more bases at the presence of appropriate ddNTP, the mass of the primer extension products are different, which can be determined by mass spectrometry.

Here we applied FP-TDI and MassArray assay to high throughput SNP genotyping, after optimization, the accuracy, efficiency and robustness were compared.

Materials and Methods

Materials

Primers were from IDT. *Taq* DNA polymerase was from TaKaRa. *Exo*I-SAP, R110 and TAMRA attached Terminator Mix, and acyclopol were from Perkin-Elmer. H2OBIT™ thermal cycler was from Abgene. Thermosequenase, Bruker AutoFlex mass spectrometry and SpectroChip were from Sequenom. MassArray nanodispenser and Clean resin were from Sequenom.

DNA samples

96 samples including 90 genomic DNA, 5 repeat control and one negative control were purchased from Coriell Institute for Medical Research, USA.

SNP markers and PCR

SNP markers of human chromosome 3 were selected randomly from the dbSNP database (http://www.ncbi.nlm.nih.gov/SNP). Some SNP markers selected and the detection primers are as following Table 1.

PCR was carried out on H2OBIT™ thermal cycler using *Taq* DNA polymerase 0.01 u/μl, dNTP (Promega) 0.1 mM each, $MgCl_2$ 2.5 mM, forward and reverse primer 1 μM each, and genomic DNA 0.3 ng/μl. Cycling program was: 95 °C 2 min; 95 °C 20 s, 55 °C 30 s, 72 °C 40 s (45 cycles); 72 °C 10 min. Products were checked on a 2% agarose gel.

Primer extension on FP-TDI

After reaction, 8 μl PCR product was incubated with 4 μl *Exo*I-SAP for 1 h at 37 °C to remove excess primers and dNTPs, followed by 20 min at 85 °C for enzyme inactivation; then added SNPprimer 2 μM, Terminator

Table 1 Some markers and related primers used in FP-TDI or MassArray assay

Marker	Forward primer	Reverse Primer	SNPprimer
rs1447153	TGTACTTAAAAGGCATCATCTTT	GCCAAAGTAAGTCACATAGTCAA	AACACTTTATGCATTTTC-AGTAAGTTAAG
rs6800929	ACGTTGGATGCCTGCTATTAGAC-TGCCAAC	ACGTTGGATGGGTTCCACAAAAA-TCAGAGC	CTGCCAACCAAACACTTC
rs1429768	ACGTTGGATGATCAGGGACTTTA-CAGGACG	ACGTTGGATGCCTCAACCTGCTTT-TCCATC	GGAAGAATTTAATTGTTT-GGAG
rs1947555	ACGTTGGATGCCTCCTTTTGTCA-TCTTGCC	ACGTTGGATGGATCTACCTTTGTC-CCTTGG	CATCTTGCCATTAATGCC-AAA
rs201674	ACGTTGGATGCATTGCTGTGATG-AAGTAGG	ACGTTGGATGAGCTGTGTAGTAGA-GAGATG	AGGATGATGTTCTGTGC
rs201739	ACGTTGGATGGCACTTAATCCAC-AACCTGC	ACGTTGGATGCATCCTTGCCAAGT-CTTGAG	CAACCTGCACCGCTGTC
rs2303994	ACGTTGGATGCTCTAATGGGATG-CAACCAC	ACGTTGGATGGTGGATTAATTACC-AGTCGG	TTGACAGATGCTAATTTC-TTGAC
rs3850168	ACGTTGGATGGTTATGAGACTTC-ATCTCAG	ACGTTGGATGACATGGCTCTAAGG-TAGCAC	GACTTCATCTCAGAGGGG-TTT
rs4266132	ACGTTGGATGCTCATTGCGCAAG-GCAAATC	ACGTTGGATGATGCTCCTTCTGGG-ATTATG	TGAGTTATGGAGAGAAGC
rs4336059	ACGTTGGATGCAGGTCCAAATCA-TAGTTTTC	ACGTTGGATGGGCTCCAGAAGCT-GAAAAAC	ATCAAACCAGTAGCCATG

Mix 0.75 μl, acyclopol 0.05 μl to carry out the primer extension reaction with 20 μl total volume. Cycling program was: 95 °C 2 min; 95 °C 15 s, 55 °C 30 s (15 cycles). The extension product was centrifuged RPMI 1600 for 3 min at 20 °C, detected by Envision™. SNPscoring was used to assign the SNP types.

Primer extension on MassArray

After reaction, 5 μl PCR product was incubated with 2 μl SAP (PerkinElmer) for 20 min at 37 °C to remove excess dNTPs, followed by 7 min at 90 °C for enzyme inactivation; then added SNPprimer 600 nM, TSQ 0.036 μl, hME Mix (1×, Sequenom) 0.2 μl to carry out the primer extension reaction with 9 μl total volume. Cycling program was: 95 °C 2 min; 95 °C 5 s, 55 °C 5 s, 72 °C 5 s (45 cycles).

The extension product was first deionized by adding 3 mg dry Clean Resin and 15 μl MQ water per well, rotated on a rotator for 20 min, centrifuged for 5 min in RPMI 1600. MassArray Nanodispenser was used to transfer the samples to SpectroChip. Mass spectra were recorded on a Bruker Autoflex operated in the linear mode, and SNP genotypes were analyzed by software MassArray Typer.

Results

SNP assignment evaluation of FP-TDI assay

After selecting SNP marker rs1447153 and A/G type Terminator Mix (containing TAMRA attached ddATP and R110 attached ddGTP), FP-TDI genotyping was carried out, the result was as following Fig. 1.

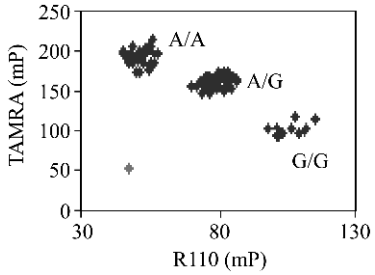

Fig. 1 FP-TDI assay genotyping result of rs1447153

FP-TDI assigns the genotype by clustering samples according to their different FP. A/A homozygotes have higher FP of TAMRA, clustered in up-left, G/G homozygotes have higher FP of R110, clustered in down-right, A/G heterozygotes clustered in the middle (Fig. 1). But the picture quality of sample clustering can not be judged, so Formula (1–3) as following were introduced here:

$$d = \frac{\sum_i \sqrt{(p_{ti} - \bar{p}_t)^2 + (p_{ri} - \bar{p}_t)^2}}{n} \quad (1)$$

$$D = \sum_{k=1}^{3} d_k \quad (2)$$

$$E = \frac{\frac{\sum_i \sqrt{(P_{Ti} - \bar{P}_T)^2 + (P_{Ri} - \bar{P}_R)^2}}{n} - D}{\frac{\sum_i \sqrt{(P_{Ti} - \bar{P}_T)^2 + (P_{Ri} - \bar{P}_{Ri})^2}}{n}} \quad (3)$$

P_{Ti}, TAMRA FP value of sample i; P_{Ri}, R110 FP value of sample i; P_R, average R110 FP value of all samples; P_T, average TAMRA FP value of all samples; P_r, average R110 FP value within a cluster; P_t, average TAMRA FP value within a cluster; P_{ti}, TAMRA FP value of sample i within a cluster; P_{ri}, R110 FP value of sample i within a cluster.

Samples with same genotypes clustered tighter, the D value was smaller, and the E value bigger, the picture quality better, the SNP assignment by FP-TDI more believable, if E value was less than 0.6, the SNP assignment would be unacceptable.

Concentration of SNPprimers and PCR products

Different concentration of SNPprimer and PCR products were tested (SNP marker rs1447153), and SNP assignment quality of FP-TDI was evaluated. The result showed that FP-TDI assay was very sensitive to the concentration of SNPprimer and PCR products. Using the extension cycle as a window under which successful SNP assignment could be achieved, when the concentration of PCR products were 30 ng/μl, 10 ng/μl and 3 ng/μl, the SNP assignment windows were 9–18 cycles, 9–30 cycles, 27–33 cycles respectively [Fig. 2(A)]; when the concentration of SNPprimers were 1 μM, 0.5 μM and 0.125 μM, the SNP assignment windows were 9–40 cycles, 9–22 cycles, 18–22 cycles respectively [Fig. 2(B)].

MW (molecular weight) of SNPprimers and extension cycles

SNPprimers with different MW were tested in each assay. When the MW of SNPprimers were 4000 Da, 5000 Da, 6000 Da, 7000 Da, 8000 Da, 9000 Da, 10,000 Da, the corresponding call rates were as following Fig. 3(A). The MW of SNPprimer was positively related to the call rate in FP-TDI assay: when the MW was less than 6000 Da (about 20 bp), the call rate was less than 50%; while the

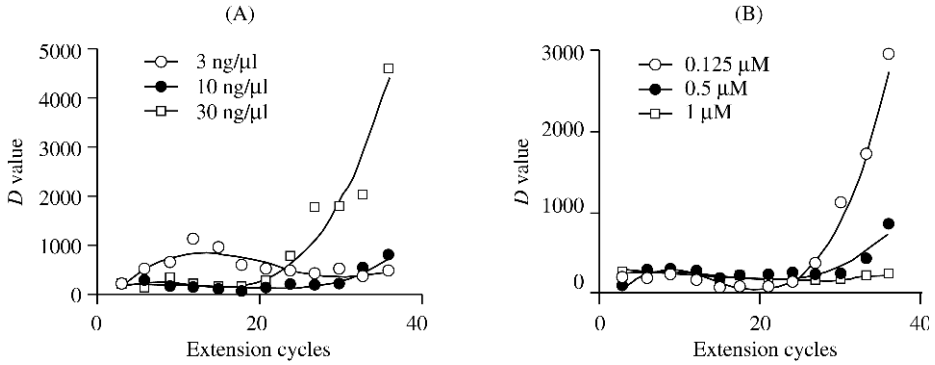

Fig. 2 SNP assignment quality and it's relationship with PCR product concentration (A) and SNPprimer concentration (B) in FP-TDI assay

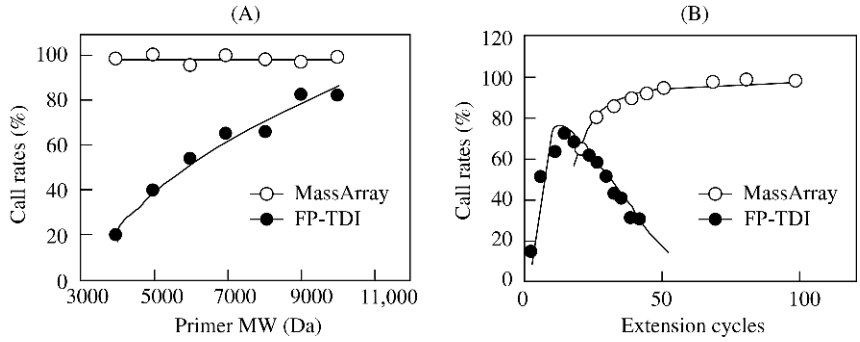

Fig. 3 The relationships between call rates and MW of SNPprimers (A) and between call rates and extension cycles (B) in FP-TDI and MassArray assays

MW was greater than 8000 Da (about 25 bp), the call rate could reach over 70%, the most suitable MW was 8000–10,000 Da (24–30 bp). In MassArray assay, the length of the SNPprimer was not so sensitively related to the call rate, the length of SNPprimer could be as short as about 16 bp.

Different extension cycles were tested using randomly selected SNP sites of human chromosome 3 (data not shown), the relationship between the call rate and extension cycles was as Fig. 3(B). In FP-TDI, the window of higher call rate was very narrow, with the extension cycles between 6–15, the call rate could reach above 60%; while in MassArray, when extension cycles was more than 30, the call rate could reach >80%, with some more cycles to lead to a higher call rate.

Single and multiplex genotyping

Both multiplex PCR and multiplex extension reaction were generated. In FP-TDI, 4-plex PCR was carried out first, then the PCR products were divided into 4 parts, and successfully followed by 4 uniplex extension reactions (data not shown).

In MassArray assay, with carefully selected SNPprimers (length of which not vary so much), 8-plex PCR followed by 8-plex extension reaction was successfully carried out and detected (Fig. 4), the PCR volume was reduced to 5 μl, and all steps of the sample preparation were done in a single vessel.

High throughput, cost and accuracy

Using one H2OBIT™ thermal cycler and 384 microplates, 30 plates could be prepared in 6 h. The main factor that affected the output of FP-TDI and MassArray was the instrument detection ability. Using Envision™, FP-TDI would spend 2 h in reading 30 plates, altogether 34,560 samples could be processed within 24 h, about 360 sites (96 samples per site); using Bruker AutoFlex, MassArray would spend 20 h in reading 30 plates, altogether 85,070 samples could be processed in 24 h (8-plex), about 880 sites (96 samples per site), if extending robotic spotting into the off hours, the throughput could be further improved. With several softwares developed, the data

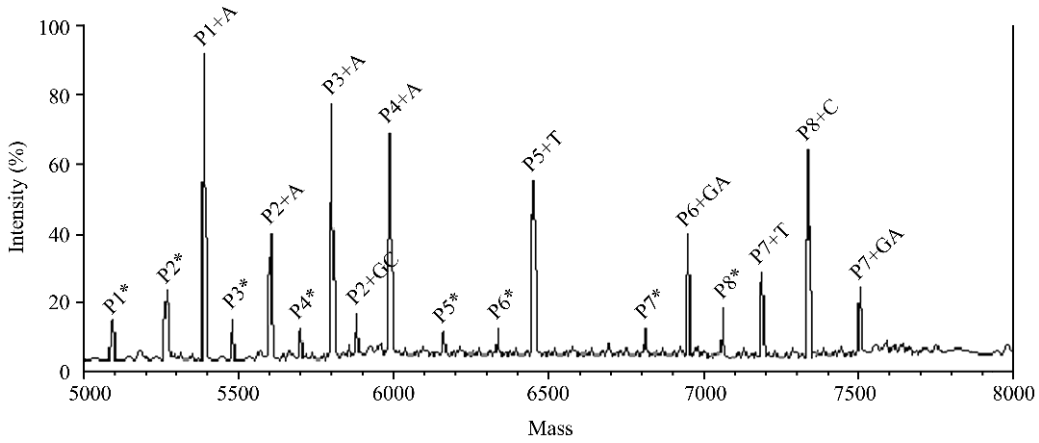

Fig. 4 8-plex MassArray genotyping

"*" indicated: P1, rs201739 (A/GC); P2, rs201674 (A/GC); P3, rs4336059 (A/GT); P4, rs4266132 (A/GGT); P5, rs3850168 (T/GC); P6, rs1947555 (A/GA); P7, rs1429768 (T/GA); P8, rs2303994 (C/GGGC).

processing of MassArray was automated, while the data processing of FP-TDI kept being time consuming. In addition, 8-plex MassArray considerably reduced the cost compared with the FP-TDI method.

Following the standard protocol, 6440 SNP sites selected from dbSNP of human chromosome 3 were genotyped, 4792 sites by FP-TDI, 1648 sites by MassArray (data not shown). The success rates of each assay were 67.7% and 96.5% respectively, including 26% homozygotes sites detected by FP-TDI and 18.9% by MassArray.

Discussion

Mini-sequencing is a relatively simple SNP genotyping method, including only three steps: PCR, digestion and single base extension. The main difference between FP-TDI and MassArray is the detection strategy.

FP-TDI detects the FP change of the fluorescent molecule, which has been successfully used in clinical diagnosis recently [18,19]. The degree of FP is determined by temperature, solvent viscosity, and fluorescent molecule volume [16,18], this is why FP-TDI assay is very sensitive to the concentration of extension primer and PCR product. The volume of the fluorescent molecule is directly related to its MW. When the MW is about 10,000 Da, the FP is most effective [20]. In addition, the detection instrument is very sensitive for detecting small amounts of fluorophore, and any nonspecific products will increase the noise in the signal, so the longer SNPprimer will lead to better specificity of the SBE product to decrease the noise. According to the MW of TAMRA and R110, SNPprimer would be better with length longer than 24 bp (about 7200 Da) and ranging between 24–30 bp.

Based on FP change, FP-TDI assigns the genotype by clustering the samples. Until now there has no numeric standard to evaluate the clustering result, the *E* value introduced here can efficiently judge the tightness of the cluster, thus make the optimization work easier.

FP-TDI assay is very sensitive to the extension cycles, this is because the incorporation speed of TAMRA and R110 are unequal. R110 is much faster, and can be used up quickly, then the continue incorporation of TAMRA will lead to misincorporation, this is why the homozygotes call rate of the FP-TDI assay is more higher than MassArray, while in high throughput genotyping, this also makes the extension reaction very difficult to control.

MassArray assay directly measures the physical property—MW of the product, because the mass of the products are fixed, and the software will recognize the background noise, slight contamination and non-specificity of the SNPprimer can be neglected, if possible, the MW of SNPprimer can be only 5000 Da (about 16 bp), higher MW will lead to less efficient desorption and decreasing sensitivity of the detector [16], our results showed that even the length of the SNPprimer was 30 bp, the success rate was not affected, meaning that the detector was very efficient, which even could detect the product at very low level, this also made the MassArray not so sensitive to the concentration of the SNPprimer and PCR product, the concentration of SNPprimer and DNA product could vary in a wider range.

The most important advantages of using MassArray assay for analysis of mini-sequencing products are the

accuracy of the masses from genotyping reactions and being able to analyze multiplexed samples, by multiplexing, the throughput was improved, and the cost was decreased considerably. The mass spectrophotometric process does not label the primer, the protocol is rather stable and needs not to change time to time. In MassArray assay the misincorporation was not so serious, it was less than 0.03%, this might be due to the TSQ enzyme used by MassArray which was more specific than the acyclopol enzyme used by FP-TDI. One disadvantage of MassArray is that the mass spectrometry and the SpectroChip are very expensive, whereas the detection instrument of FP-TDI assay is cheap and easy to use.

Although these two assays are practicable in high through-put SNP genotyping, they are still under development. Daniel *et al*. [20] decreased the rate of misincorporation by adding proofreading polymerases, and Thorarinn *et al*. [21] modified the MassArray by only add 3 dNTPs to the extension mix, thus the requirement for modified ddNTP incorporating DNA polymerases was alleviated, and reduced the complexity compared to the original mini-sequencing methodology by utilizing single DNA polymerase.

In general, compared to FP-TDI assay, the MassArray assay is more faster, accurate, promising, and suitable for high throughput SNP genotyping.

References

1 Miki Y, Swensen J, Shattuch ED, Futreal PA, Harshman K, Tavtigian S, Liu Q *et al*. A strong candidate for the breast cancer and ovarian cancer susceptibility gene BRCA1. Science, 1994, 266: 66–71

2 Kirk BW, Feinsod M, Fanis R, Klimen RM, Barany F. Single nucleotide polymorphism seeking long term association with complex disease. Nucleic Acids Res, 2002, 30(15): 3295–3311

3 Erlich HA, Gelfand D, Sninsky JJ. Recent advances in the polymerase chain reaction. Science, 1991, 252: 1643–1651

4 Kruglyak L, Nickerson DA. Variation is the spice of life. Nature Genetics, 2001, 27(3): 234–236

5 Kwok PY. Methods for genotyping single nucleotide polymorphisms. Annu Rev Genomics Hum Genet, 2001, 2: 235–258

6 Suzanne Jenkins, Neil Gibson. High-throughput SNP genotyping. Comp Funct Genom, 2002, 3: 57–66

7 Zeng ZY, Xiong W, Shen SR, Zhu SG, Li XL, Li WF, Li J *et al*. High-throughput single nucleotide polymorphisms genotyping by dynamic allele-specific hybridization. Prog Biochem Biophys, 2002, 29(5): 806–810

8 Ranade K, Chang MS, Ting CT, Pei D, Hsiao CF, Olivier M, Pesich R *et al*. Highthroughput genotyping with single nucleotide polymorphisms. Genome Res, 2001, 11(7): 1262–1268

9 le Hellard S, Ballereau SJ, Visscher PM, Torrance HS, Pinson J, Morris SW, Thomson ML *et al*. SNP genotyping on pooled DNAs: Comparison of genotyping technologies and a semi automated method for data storage and analysis. Nucleic Acids Res, 2002, 30(15): e74

10 Syvanen AC. From gels to chips: 'Mini-sequencing' primer extension for analysis of point mutations and single nucleotide polymorphisms. Hum Mutat, 1999, 13: 1–10

11 Chen X, Levine L, Kwok PY. Fluorescence polarization in homogeneous nucleic acid analysis. Genome Res, 1999, 9(5): 492–498

12 Chen X, Kwok PY. Template-directed dye-terminator incorporation (TDI) assay: A homogeneous DNA diagnostic method based on fluorescence resonance energy transfer. Nucleic Acids Res, 1997, 25: 347–353

13 Chen X, Levine L, Kwok PY. Fluorescence polarization in homogeneous nucleic acid analysis. Genome Res, 1999, 9: 492–498

14 Haff LA, Smirnov IP. Single-nucleotide polymorphism identification assays using a thermostable DNA polymerase and delayed extraction MALDI-TOF mass spectrometry. Genome Res, 1997, 7: 378–388

15 Sun X, Ding H, Hung K, Guo B. A new MALDI-TOF based mini-sequencing assay for genotyping of SNPs. Nucleic Acids Res, 2000, 28(12): E68

16 Bray MS, Boerwinkle E, Doris PA. High-throughput multiplex SNP genotyping with MALDI-TOF mass spectrometry: Practice, problems and promise. Hum Mutat, 2001, 17: 296–304

17 Kwok PY. SNP genotyping with fluorescence polarization detection. Hum Mutat, 2002, 19: 315–323

18 Gao YE, Zhang J, Wu J, Chen ZC, Yan XJ. Detection and genotyping of human papillomavirus DNA in cervical cancer tissues with fluorescence polarization. Acta Biochim Biophys Sin, 2003, 35(11): 1029–1034

19 Freeman BD, Buchman TG, McGrath S, Tabrizi AR, Zehnlauer BA. Template directed dye-terminator incorporation with fluorescence polarization detection for analysis of single nucleotide polymorphisms implicated in sepsis. J Mol Diagn, 2002, 4(4): 209–215

20 di Giusto D, King GC. Single base extension (SBE) with proofreading polymerases and phosphorothioate primers: Improved fidelity in single-substrate assays. Nucleic Acids Res, 2003, 31(3): e7

21 Blondal T, Waage BG, Smarason SV, Jonsson F, Fjalldal SB, Stefansson K, Gulcher J *et al*. A novel MALDI-TOF based methodology for genotyping single nucleotide polymorphisms. Nucleic Acids Res, 2003, 31(24): e155

Edited by
You-Xin JIN