

Nonidet P-40, 50% glycerol, and 2mM each dNTP (TaKaRa, Shiga, Japan) with 6.0 pmol of Cy3-labeled ED-1 (Cy3-ED-1), 6.0 pmol of Cy5-labeled ED-2 (Cy5-ED-2), 30 fmol each of the 96 D1s, and 1.5 U *Ex Taq* polymerase. The reaction initially was incubated at 95 °C for 1 min, followed by 30 cycles of 95 °C for 30 s, 55 °C for 6 min, and 72 °C for 30 s, using a Bio-Rad PTC-200 Peltier thermal cycler. The reaction was stopped by holding the temperature at 10 °C.

#### Hybridization and detection on DNA microarray

We purchased a DNA microarray (NovusGene, Tokyo, Japan) that had 24 separated areas on the same slide glass. Each of the separated areas contained 100 types of oligonucleotide probe (96 probes for 96 SNPs and 4 probes for validation controls of the assay) identical to D1 sequences. Of the 4 validation control probes, 3 were not used in the DigiTag2 assay because these probes were prepared to validate the washing step with magnetic beads in the previous version of the DigiTag assay. The ready-to-use DNA microarrays were stored in a desiccator at room temperature until use.

A hybridization mixture was prepared by mixing 5  $\mu$ l of labeled products in 12  $\mu$ l of hybridization buffer containing 0.5 $\times$  SSC, 0.1% sodium dodecyl sulfate (SDS), 15% formamide, and 1 mM EDTA with 1  $\mu$ l of hybridization control. The hybridization control was prepared with 2.5 fmol of Cy3-labeled D1\_100 and Cy5-labeled D1\_100. Then 8  $\mu$ l of the hybridization mixture was applied to each area on the DNA microarray. Hybridization was carried out for 30 min at 37 °C in a hybridization oven (ThermoStat plus, Eppendorf, Hamburg, Germany). After hybridization, DNA microarrays were washed in washing buffer (0.1 $\times$  SSC and 0.1% SDS) with shaking at 60 rpm for 5 min. DNA microarrays were consecutively washed in distilled water with shaking at 60 rpm for 1 min and were then dried by centrifugation at 2000 rpm for 1 min. Hybridization images were scanned at photomultiplier voltages of 400 V for Cy3 and 480 V for Cy5 using a commercially available DNA chip scanner, and fluorescence image analysis was performed using commercially available software (GenePix 4000B unit and GenePix Pro 4.1 software package, Axon Instruments, Foster City, CA, USA). The genotype calls were determined using the SNPStar software (version 0.0.0.8, Olympus, Tokyo, Japan).

## Results and discussion

#### DigiTag2 assay scheme

We previously reported a multiplex SNP typing method, the DigiTag assay, in which all of the SNP genotypes are encoded to the well-designed oligonucleotides, named DNA coded numbers (DCNs) [20]. The assignment of the DCNs to the SNPs is unconstrained; therefore, the DNA chips prepared to read out the types of DCN are univer-

sally available for any types of SNP. We revealed that the DigiTag assay has the potential to analyze nearly all kinds of SNP with high accuracy and reproducibility. However, the DigiTag assay needs the washing step with magnetic beads, which is a laborious step in manipulation. Also, the biotinylated probes, which are necessary for the washing step, are expensive. For the next version of the assay, we improved the protocol to exclude the washing step and named it the DigiTag2 assay.

The DigiTag2 assay involves four steps to accomplish the genotyping: target preparation, encoding, labeling, and detection (Fig. 1). During target preparation, target fragments (including target SNP sites) are prepared by multiplex PCR from genomic DNA. For multiplex PCR, we designed 40-mer primers (average length) and performed multiplex PCR using a two-step protocol (denature and extension steps) with a 6-min extension step. For encoding, we prepared two 5' query probes and one 3' query probe for each SNP site. The 5' query probes have a sequence

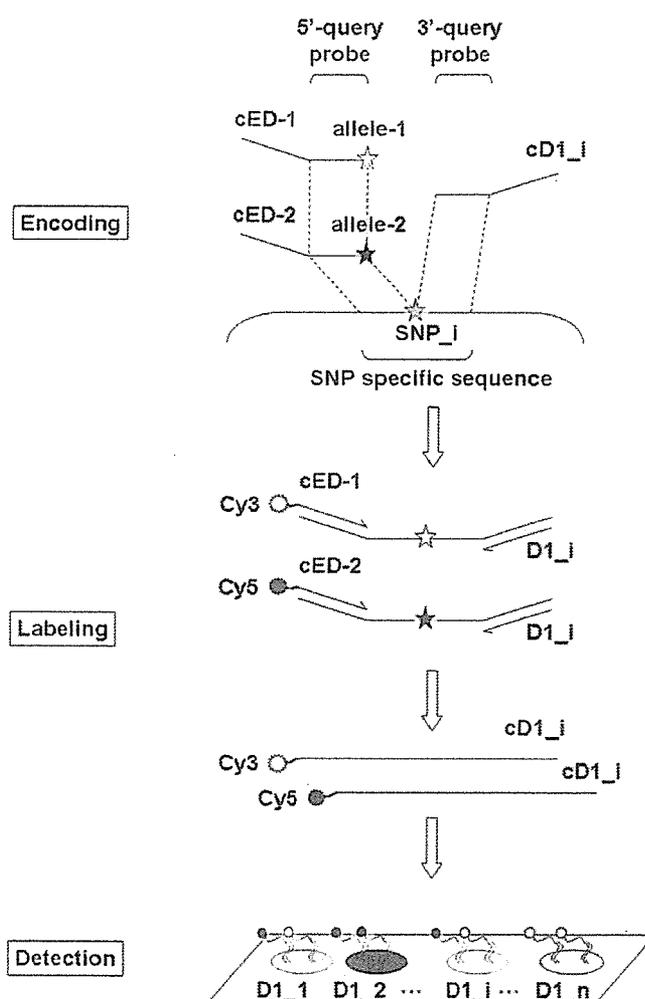


Fig. 1. Schematic representation of DigiTag2 assay. This assay involves four steps to accomplish SNP typing: target preparation, encoding, labeling, and detection. The 5' query probes have EDs (cED-1 and cED-2) corresponding to each allele, and the 3' query probes have a variable sequence (cD1\_i) for each SNP. Each reverse complement sequence is depicted by a lowercase "c" before the sequence name.

complementary to the 5'-flanking region of the target SNP, and each of the probes has an allele-specific sequence. Two types of ED (ED-1 and ED-2) were attached to each of the 5' query probes (see Materials and Methods), and we incorporated a mismatch base into the 5' query probes at the fourth position from the SNP site to improve the precision of allele discrimination [20]. The 3' query probe has a sequence complementary to the 3'-flanking region of the target SNP, and each of the probes has a D1 on its 3' end. In the encoding step, the 5' query and 3' query probes are successfully concatenated by *Taq* DNA ligase, and the probes are fully complementary to adjacent regions on the target fragment [22]. The genotype is then converted into a type of ED and a type of D1. The types of ED and D1 designate the type of allele and SNP, respectively. After the encoding step, fluorescence is incorporated into the ligated products by asymmetric PCR using fluorescent-labeled primers (Cy3-ED-1 and Cy5-ED-2) and D1 primers. The D1 primers are a mixture of all D1s used in the assay. The Cy3- and Cy5-labeled PCR products are directly hybridized with the D1 probes on the DNA microarray to reveal SNP genotypes by reading signals from the various D1s. If the genomic DNA sample is homozygous for a certain SNP, a single color signal from Cy3 or Cy5 is detected from the corresponding spot on the DNA microarray. In contrast, both signals are present when the genomic DNA sample is heterozygous.

#### SNP selection and probe design

In a previous report, we investigated the ligation conditions in the encoding step using an SNP located in the *PLOD* gene on human chromosome 1p36 as a model SNP (JSNP ID IMS-JST068774) and determined the parameters for 5' query and 3' query probes [20]. We then randomly selected 96 SNPs from a 610-kb region, including the *IL-4* and *IL-13* genes on human chromosome 5q31-33, which contains various candidate genes related to immune and allergic disorders. We subsequently designed probes for the 96 SNP sites to have a uniform melting temperature as that of *PLOD* SNP so as to give similar ligation efficiency among the 96 SNP sites to be analyzed in a single tube. We also incorporated a mismatch base into the 5' query probe at the fourth position from the SNP site for all target SNPs. The 20-mer mismatch-induced 5' query probes and 3' query probes (average length) had melting temperatures of  $50.7 \pm 2.1$  °C and  $52.4 \pm 1.5$  °C, respectively. Here we found that the length of the 3' query probe influences the ligation efficiency in the encoding step; when a longer 3' query probe was used in the encoding step, stronger signal intensities were acquired on microarray detection (data not shown). Therefore, we used the lengthened 3' query probes to 30-mer, and the average melting temperature of the lengthened 3' query probes was  $66.1 \pm 3.5$  °C. The sequence information for 5' query probes and lengthened 3' query probes is listed in Supplementary Table 3.

#### Optimization of reaction conditions

When we used the mismatch-induced 5' query probes, indistinct clusters were observed from 5 SNPs (SNP 7, SNP 9, SNP 18, SNP 49, and SNP 93) (Fig. 2A). However, these 5 SNPs can be discriminatively genotyped with perfect

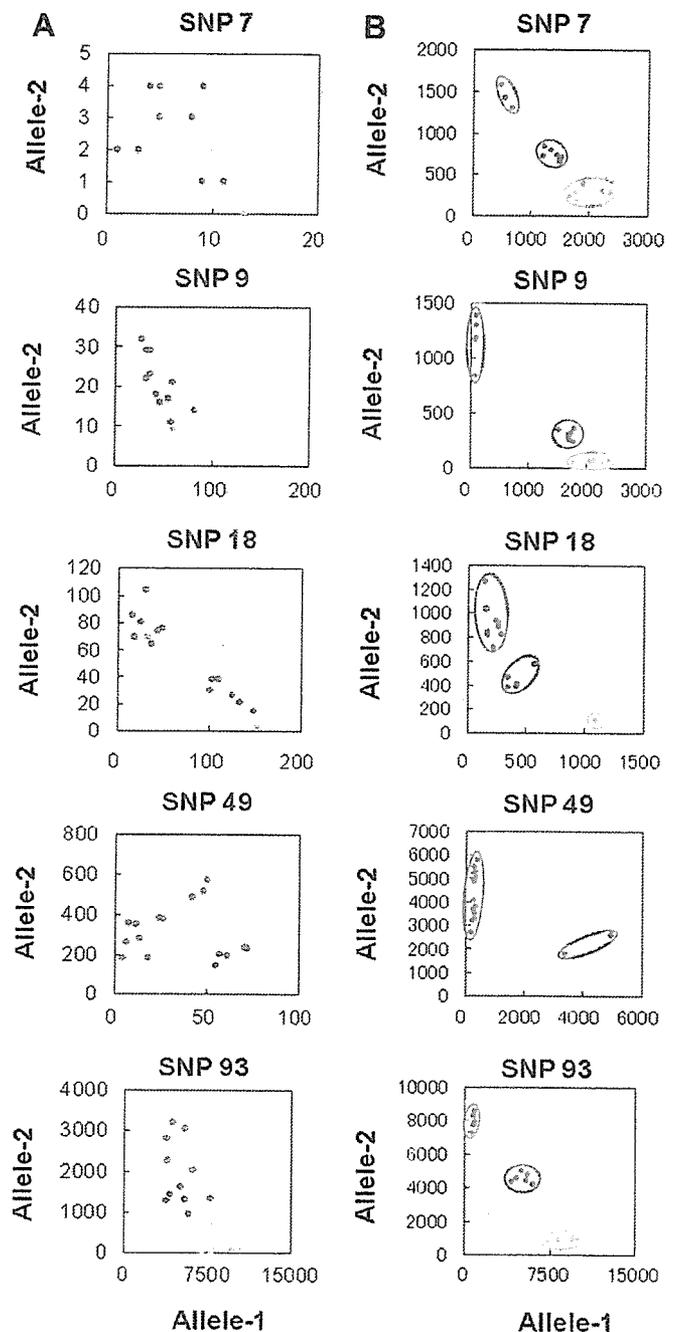


Fig. 2. Treatment of 5 failed SNPs with mismatch-induced 5' query probes. Green dots and circle show allele-1 homozygous samples, red dots and circle show allele-2 homozygous samples, and blue dots and circle show heterozygous samples. (A) Mismatch-induced 5' query probes, which have a mismatched base incorporated into the fourth position from the SNP base, were used. (B) Here 5' query probes, which have a perfect match sequence for the target SNP, were used instead of mismatch-induced 5' query probes. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

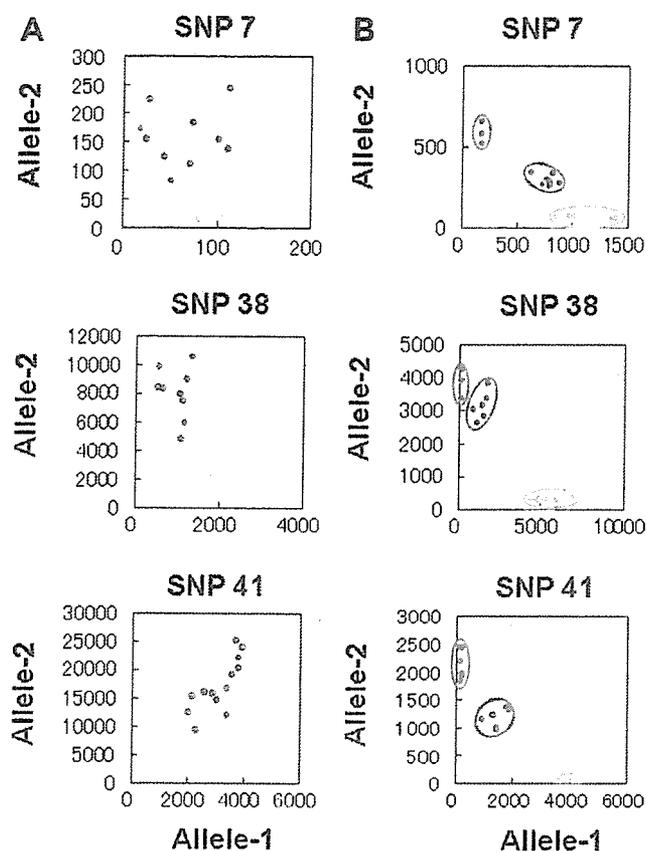


Fig. 3. Effects of D1 primer concentration in the labeling step. Green dots and circle show allele-1 homozygous samples, red dots and circle show allele-2 homozygous samples, and blue dots and circle show heterozygous samples. (A) Here 5 nM D1 primers was used with 500 nM fluorescent-labeled primers. (B) Here 2.5 nM D1 primers was used with 500 nM fluorescent-labeled primers. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

match 5' query probes (Fig. 2B). For these 5 SNPs, the mismatch base incorporated into the fourth position from the SNP site has a drastic effect on the hybridization stability between 5' query probes and the target multiplex PCR products in the encoding step and leads to signal loss on microarray detection. Therefore, we performed a 96-plex oligonucleotide ligation assay with the mismatch-induced 5' query probes for 91 target SNPs in combination with perfect match 5' query probes for these 5 target SNPs.

To incorporate the fluorescent label into the ligated products, we performed asymmetric PCR with fluorescent-labeled primers and D1 primers (mixture of D1s). Using the mixture of all D1s, instead of the single primer pair mentioned in the DigiTag assay [20], would make it possible to uniformly acquire all target fragments. However, the concentration of D1 primers used in the labeling step was found to exert an influence on the cluster distribution in scatter diagrams. When we used 5 nM D1 primers with 500 nM fluorescent-labeled primers, dispersed and/or indiscrete clusters were observed for several SNPs (Fig. 3A). However, the dispersed and/or indiscrete clusters became convergent and/or discrete clusters when we used 2.5 nM D1 primers with 500 nM fluorescent-labeled primers

(Fig. 3B). The D1 primers share the fluorescent-labeled primers in the labeling step, and the ratio of each D1 primer to fluorescent-labeled primer was approximately 1:2 at 2.5 nM and 1:1 at 5 nM. When the amount of each D1 primer was greater than the amount of fluorescent-labeled primer, strong false-positive signals were observed on microarray detection, leading to indiscrete clusters on scatter diagrams (data not shown). On the other hand, when the amount of each D1 primer was less than the amount of fluorescent-labeled primer, insufficient amplification occurred in a number of target SNPs, leading to weak signal intensities on microarray detection (data not shown). We found that the optimal ratio of D1 primer to fluorescent-labeled primer is approximately 1:2, irrespective of the multiplicity of the assay (number of SNPs to be analyzed).

### Genotyping results

Multiplex PCR products, including the 96 SNP sites, showed similar band patterns as 48 individual DNA samples, although it was difficult to clearly discern all 96 PCR products due to limitations in electrophoretic resolution (Fig. 4A). We then performed a multiplexed oligonucleotide ligation assay using the multiplex PCR products as targets. To incorporate the fluorescent label into the ligated products, asymmetric PCR was performed using the fluorescent-labeled and D1 primers. DNA microarray revealed hybridization images of 24 individual samples from each of the 24 separated areas having 100 spots (96 probes for 96 SNPs and 4 probes for validation controls) (Fig. 4B). The hybridization image was analyzed using a DNA chip scanner, and the Cy3 and Cy5 signal intensities of each spot were plotted to produce a scatter diagram. The SNP genotypes of 16 genomic DNA samples, randomly selected from the 48 samples, were alternatively determined by direct sequencing and were used as reference data.

As a result of 96-plex genotyping under optimal labeling conditions using the mismatch-induced 5' query probes in combination with the perfect match 5' query probes for 5 SNPs, three distinct clusters corresponding to two homozygous genotypes and one heterozygous genotype were observed from 84 SNPs (exceptions were SNP 31, SNP 37, SNP 60, SNP 61, and SNP 87) (Fig. 4C). The remaining 7 SNPs (SNP 12, SNP 22, SNP 27, SNP 33, SNP 67, SNP 88, and SNP 91) were found to be monomorphic in 48 genomic DNA samples and were excluded from further analysis. For SNP 37, SNP 60, and SNP 87, drastically attenuated signal intensities were observed on microarray detection (Fig. 4D). Signal loss was caused by insufficient amplification of the target fragments on multiplex PCR because no amplified products were observed on singleplex PCR, even when the second candidate primer pairs were used. There were some structural obstacles in the target region, although we could not identify any characteristic structures. SNP 31 and SNP 61 were found to have strong false-positive signals, leading to indistinct clusters on scatter diagrams (Fig. 4E). The false-positive signals would be caused

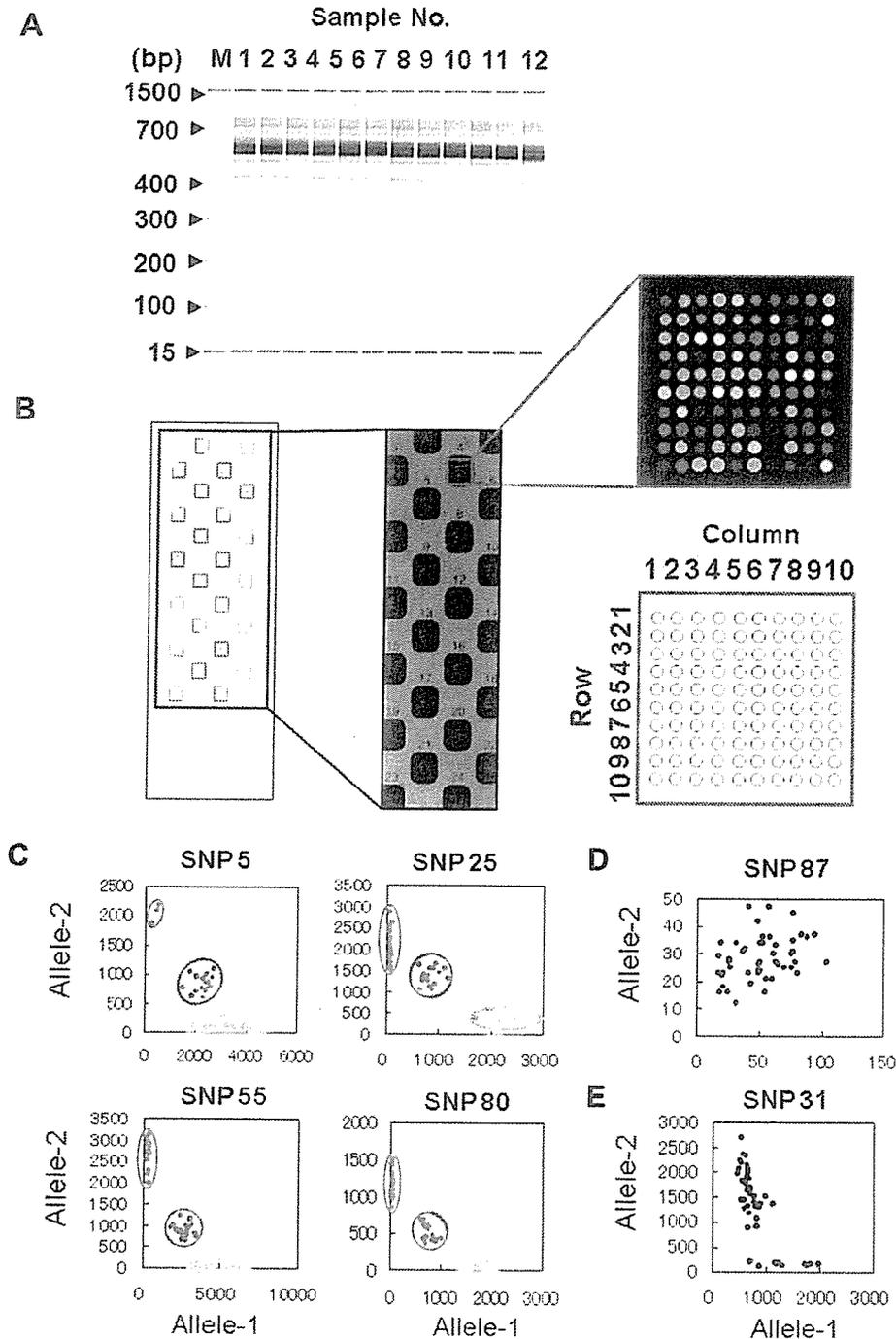


Fig. 4. Multiplex SNP typing for 96 SNPs using 48 individual genomic DNA samples. (A) Gel images of multiplex PCR products with different samples. In all sample lanes, sample bands were observed between two inner markers: 15 and 1500 bp. (B) Hybridization images of DNA microarray. (C) Scatter plot diagrams for 4 randomly selected SNPs from 84 working SNPs. Green dots and circle show allele-1 homozygous samples, red dots and circle show allele-2 homozygous samples, and blue dots and circle show heterozygous samples. (D) Example of the typing-failed SNP caused by insufficient amplification of target fragment in multiplex PCR. (E) Example of the typing-failed SNP that was found to have strong false-positive signals. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

by the misligation in the encoding step that was reported to be prone to occur when the mismatched pairs are G-T, G-A, G-G, A-G, and T-G [23,24]. Of the 2 misligated SNPs, 1 had an A-G mismatch (SNP 31) and the other had a T-G mismatch (SNP 61) between the 5' query probe and the target fragment. These 2 SNPs were undetectable, even when

5' query probes with the mismatched base incorporated into different positions were used (data not shown). Although there were other G-T, G-A, G-G, A-G, and T-G mismatches within the set of 84 working SNPs, we consider that these mismatches would increase the likelihood of misligation in some cases. In the future, we will be able to

search for the cause of misligation by accumulating data from failed analyses of numerous target SNPs.

#### *Conversion rate, call rate, accuracy, and reproducibility*

We investigated the feasibility of the DigiTag2 assay by performing 96-plex SNP typing using 48 human genomic DNA samples, and we found that the DigiTag2 assay has the potential to analyze all types of SNP with high accuracy and reproducibility. Here we excluded 7 SNPs from further analysis because they were revealed to be monomorphic in 48 samples. The DigiTag2 assay was found to have a 94.4% (84/89) conversion rate, which is defined as the proportion of successfully genotyped SNPs among the total number of SNPs examined. The call rate, which is defined as the number of genotype calls among the total number of samples examined, was 99.95% (4030/4032). The typing results were 100% identical to the results of direct sequencing. The reproducibility of this assay was examined by duplicate experiments, and it was found that genotype calls were 100% identical between duplicate experiments.

#### *Advantages of DigiTag2 assay*

The DigiTag2 assay performs multiplex PCR to excise target regions, including SNP sites from genomic DNA, prior to oligonucleotide ligation assay. Reducing the complexity of the genome by selectively collecting target SNP sites from the genome would lead to successful genotyping [25]. In designing the genotyping probes for oligonucleotide ligation assay, there are no alternatives because the SNP sequence is included in the probe sequences. Therefore, multiplex PCR prior to oligonucleotide ligation assay has an important role in analyzing SNPs that have highly homogeneous regions in the genome. Based on 96-plex SNP typing using 48 individual genomic DNA samples, the DigiTag2 assay has the potential to analyze all types of SNP with high accuracy and reproducibility. Moreover, the DigiTag2 assay uses unmodified primers and probes for target SNPs, thereby reducing assay cost, and requires only simple assay protocols without specialized equipment. We estimated that the running cost for the DigiTag2 assay (for oligonucleotides, reagents, DNA microarrays, etc.) is less than \$0.06/genotype. The DigiTag2 assay can use the same set of D1s and EDs for any set of target SNPs, thereby enabling 96-plex genotyping with the same assay protocols and the same microarray having the same set of probes. However, hybridization products, which are prepared in the labeling step, may cross-hybridize to the wrong D1 probes on DNA microarray due to SNP-specific sequences being introduced into the hybridization products. With regard to the 96 target SNPs selected in this study, there is no evidence of cross-hybridization between D1 probes and SNP-specific sequences. Cross-hybridization may be avoided by predicting the interaction between D1 probes and SNP-specific sequences. In the future, we will attempt to predict cross-

hybridization by accumulating data from failed analyses of numerous target SNPs.

#### **Acknowledgments**

This study was supported by a Grant-in-Aid for Scientific Research on Priority Areas and the New Energy and Industrial Technology Development Organization.

#### **Appendix A. Supplementary data**

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ab.2007.02.005.

#### **References**

- [1] L. Kruglyak, D.A. Nickerson, Variation is the spice of life, *Nat. Genet.* 27 (2001) 234–236.
- [2] M.W. McBride, D. Graham, C. Delles, A.F. Dominiczak, Functional genomics in hypertension, *Curr. Opin. Nephrol. Hypertens.* 15 (2006) 145–151.
- [3] M. Ochi, H. Osawa, H. Onuma, A. Murakami, T. Nishimiya, F. Shimada, K. Kato, I. Shimizu, K. Shishino, M. Murase, Y. Fujii, J. Ohashi, H. Makino, The absence of evidence for major effects of the frequent SNP +299 G > A in the resistin gene on susceptibility to insulin resistance syndrome associated with Japanese type 2 diabetes, *Diabetes Res. Clin. Pract.* 61 (2003) 191–198.
- [4] B.I. Freedman, D.W. Bowden, M.M. Sale, C.D. Langefeld, S.S. Rich, Genetic susceptibility contributes to renal and cardiovascular complications of type 2 diabetes mellitus, *Hypertension* 48 (2006) 8–13.
- [5] R. Yamada, S. Tokuhiko, X. Chang, K. Yamamoto, SLC22A4 and RUNX1: identification of RA susceptible genes, *J. Mol. Med.* 82 (2004) 558–564.
- [6] K. Yamamoto, R. Yamada, Genome-wide single nucleotide polymorphism analyses of rheumatoid arthritis, *J. Autoimmun.* 25 (2005) 12–15.
- [7] N. Tsuchiya, J. Ohashi, K. Tokunaga, Variations in immune response genes and their associations with multifactorial immune disorders, *Immunol. Rev.* 190 (2002) 169–181.
- [8] J. Ohashi, K. Tokunaga, The power of genome-wide association studies of complex disease genes: Statistical limitations of indirect approaches using SNP markers, *J. Hum. Genet.* 46 (2001) 478–482.
- [9] A. Wille, J. Hoh, J. Ott, Sum statistics for the joint detection of multiple disease loci in case-control association studies with SNP markers, *Genet. Epidemiol.* 25 (2003) 350–359.
- [10] C.S. Carlson, M.A. Eberle, L. Kruglyak, D.A. Nickerson, Mapping complex disease loci in whole-genome association studies, *Nature* 429 (2004) 446–452.
- [11] N. Hu, C. Wang, Y. Hu, H.H. Yang, C. Giffen, Z-Z. Tang, X-Y. Han, A.M. Goldstein, M.R. Emmert-Buck, K.H. Buetow, P.R. Taylor, M.P. Lee, Genome-wide association study in esophageal cancer using GeneChip Mapping 10 K Array, *Cancer Res.* 65 (2005) 2542–2546.
- [12] D.J. Schaid, J.C. Guenther, G.B. Christensen, S. Hebring, C. Rosenow, C.A. Hilker, S.K. McDonnell, J.M. Cunningham, S.L. Slager, M.L. Blute, S.N. Thibodeau, Comparison of microsatellites versus single-nucleotide polymorphisms in a genome linkage screen for prostate cancer-susceptibility loci, *Am. J. Hum. Genet.* 75 (2004) 948–965.
- [13] T. Arinami, T. Ohtsuki, H. Ishiguro, H. Ujike, Y. Tanaka, Y. Morita, M. Mineta, M. Takeichi, S. Yamada, A. Imamura, K. Ohara, H. Shibuya, K. Ohara, Y. Suzuki, T. Muratake, N. Kaneko, T. Someya, T. Inada, T. Yoshikawa, T. Toyota, K. Yamada, T. Kojima, S. Takahashi, O. Osamu, T. Shinkai, M. Nakamura, H. Fukuzako, T. Hashiguchi, S. Niwa, T. Ueno, H. Tachikawa, T. Hori, T. Asada, S. Nanko, H. Kunugi, R. Hashimoto, N. Ozaki, N. Iwata, M. Harano, H. Arai, T. Ohnuma, I. Kusumi, T. Koyama, H. Yoneda, Y. Fukumaki, H. Shi-

- bata, S. Kaneko, H. Higuchi, N. Yasui-Furukori, Y. Numachi, M. Itokawa, Y. Okazaki, Japanese Schizophrenia Sib-Pair Linkage Group, Genome-wide high-density SNP linkage analysis of 236 Japanese families supports the existence of schizophrenia susceptibility loci on chromosomes 1p, 14q, and 20p, *Am. J. Hum. Genet.* 77 (2005) 937-944.
- [14] P.M. Holland, R.D. Abramson, R. Watson, D.H. Gelfand, Detection of specific polymerase chain reaction product by utilizing the 5' → 3' exonuclease activity of *Thermus aquaticus* DNA polymerase, *Proc. Natl. Acad. Sci. USA* 88 (1991) 7276-7280.
- [15] N. Pourmand, E. Elahi, R.W. Davis, M. Ronaghi, Multiplex pyrosequencing, *Nucleic Acids Res.* 30 (2002) e31.
- [16] K. Lindroos, U. Liljedahl, M. Raitio, A.-C. Syvänen, Minisequencing on oligonucleotide microarrays: Comparison of immobilisation chemistries, *Nucleic Acids Res.* 29 (2001) e69.
- [17] J. Tost, I.G. Gut, Genotyping single nucleotide polymorphisms by mass spectrometry, *Mass Spectrom. Rev.* 21 (2002) 388-418.
- [18] M.S. Bray, E. Boerwinkle, P.A. Doris, High-throughput multiplex SNP genotyping with MALDI-TOF mass spectrometry: Practice, problems, and promise, *Hum. Mutat.* 17 (2001) 296-304.
- [19] A.R. Tobler, S. Short, M.R. Andersen, T.M. Paner, J.C. Briggs, S.M. Lambert, P.P. Wu, Y. Wang, A.Y. Spoonde, R.T. Koehler, N. Peyret, C. Chen, A.J. Broome, D.A. Ridzon, H. Zhou, B.S. Hoo, K.C. Hayashibara, L.N. Leong, C.N. Ma, B.B. Rosenblum, J.P. Day, J.S. Ziegler, F.M. De La Vega, M.D. Rhodes, K.M. Hennessy, H.M. Wenz, The SNPlex genotyping system: A flexible and scalable platform for SNP genotyping, *J. Biomol. Tech.* 16 (2005) 396-404.
- [20] N. Nishida, T. Tanabe, K. Hashido, K. Hirayasu, M. Takasu, A. Suyama, K. Tokunaga, DigiTag assay for multiplex single nucleotide polymorphism typing with high success rate, *Anal. Biochem.* 346 (2005) 281-288.
- [21] H. Yoshida, A. Suyama, Solution to 3-SAT by breadth first search, *DIMACS Ser. Discrete Math. Theor. Comput. Sci.* 54 (2000) 9-22.
- [22] F. Barany, Genetic disease detection and DNA amplification using cloned thermostable ligase, *Proc. Natl. Acad. Sci. USA* 88 (1991) 189-193.
- [23] J.N. Housby, E.M. Southern, Fidelity of DNA ligation: A novel experimental approach based on the polymerisation of libraries of oligonucleotides, *Nucleic Acids Res.* 26 (1998) 4259-4266.
- [24] J. Luo, D.E. Bergstrom, F. Barany, Improving the fidelity of *Thermus thermophilus* DNA ligase, *Nucleic Acids Res.* 24 (1996) 3071-3078.
- [25] H. Matsuzaki, H. Loi, S. Dong, Y.-Y. Tsai, J. Fang, J. Law, X. Di, W.-M. Liu, G. Yang, G. Liu, J. Huang, G.C. Kennedy, T.B. Ryder, G.A. Marcus, P.S. Walsh, M.D. Shriver, J.M. Puck, K.W. Jones, R. Mei, Parallel genotyping of over 10,000 SNPs using a one-primer assay on a high-density oligonucleotide array, *Genome Res.* 14 (2004) 414-425.