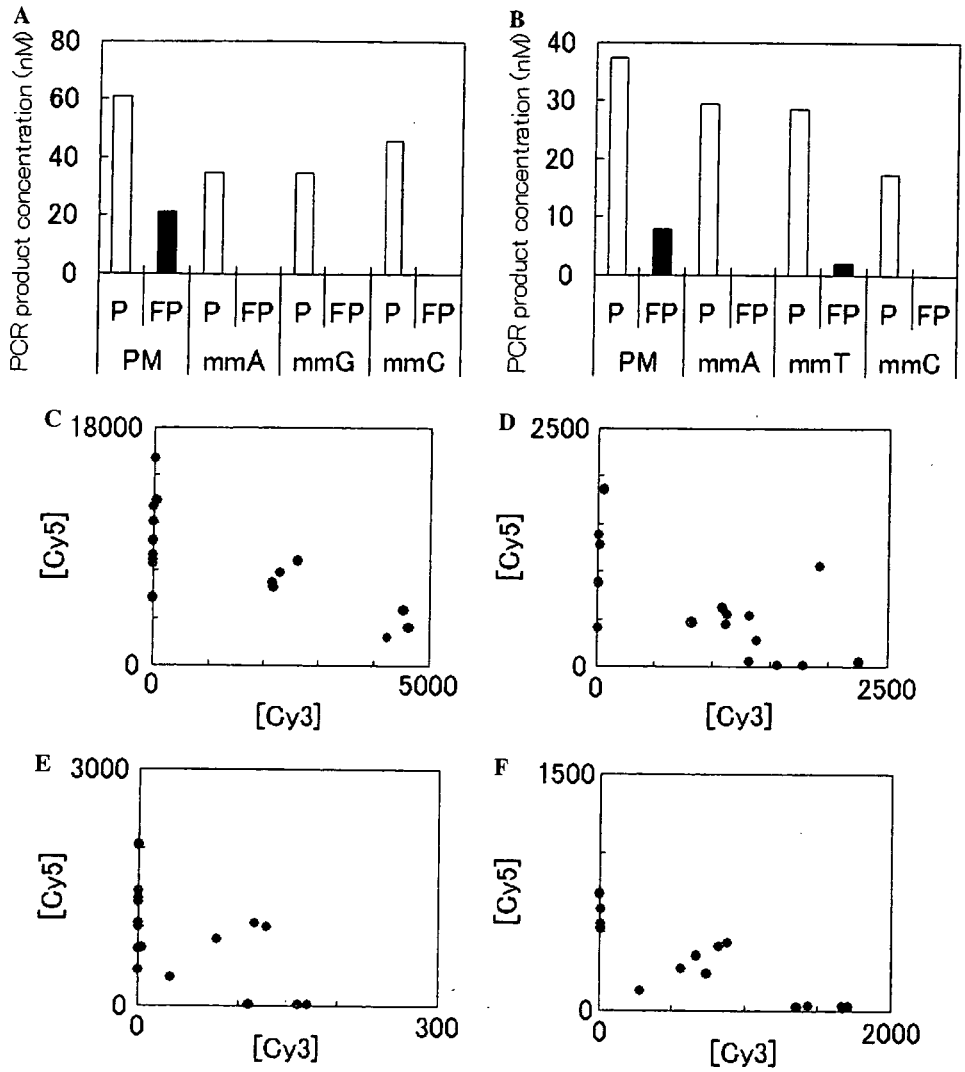

Fig. 4. Suppression of false-positive signals by mismatch-induced 5′-query probes. (A and B) Both graphs show the singleplex typing result of SNP 6 and SNP 9, respectively. PM indicates the perfect-match pair of 5′-query probe and target oligonucleotide. Three types of mismatch-induced 5′-query probes which had an artificial mismatch base at the fourth position from the SNP base were prepared (mmA, mmG, and mmC for SNP 6; mmA, mmT, and mmC for SNP 9). P and FP mean positive signals and false-positive signals, respectively. Scatter diagrams show the multiplex typing result of SNP 6 and SNP 9, respectively: (C and D) using perfect-match 5′-query probes and (E and F) using mismatch-induced 5′-query probes (mmC for SNP 6 mmA for SNP 9).

reference data from direct sequencing was 100%. The reproducibility of this assay was examined by duplicate experiments and was observed higher than 0.99 in r^2 except for 2 SNPs (SNP 2, 0.96; SNP 20, 0.96).

Discussion

We developed DigiTag assay and performed multiplex typing for 28 SNPs using 40 genomic DNA samples. One of the 28 SNPs (SNP 13) was revealed to be monomorphic in 40 samples and was therefore excluded from further analysis. This assay was found to be successful in SNP genotyping, giving a high success rate (24 of 27 SNPs) for randomly chosen SNPs. Three SNPs of 27 SNPs showed indistinct clusters, presumably resulting from missligation in the encoding step (SNP 6 and SNP 9) and insufficient amplification in the multiplex PCR (SNP 19).

The missligation was reported to be prone to occur when mismatched pairs are G-T, G-A, G-G, and A-G [21]. In our results, one of two missligated SNPs had G-G mismatch (SNP 6) and another one had G-T mismatch (SNP 9) between the 5′-query probe and the target fragment. The missligation results in increasing the false-positive signals and then leads to ambiguous genotyping. To suppress the false-positive signals, we designed mismatch-induced 5′-query probes, which had an artificial mismatched base at the fourth position from the SNP base. The mismatch-induced 5′-query probes could effectively suppress the false-positive signals without diminishing the positive signals.

Multiplex PCR has the potential to reduce the complexity fraction of the genome by selectively collecting the target SNP sites from the genome, which would lead to successful genotyping [22]. However, it was also reported that the presence of multiple primer pairs increases the chance of obtain-

ing spurious amplification products such as primer dimers [23]. To avoid spurious amplification products, we designed primer pairs to have relatively long-length (average length 40-mer) and performed multiplex PCR by a two-step protocol with an elongated extension step for 6 min. This optimization of multiplex PCR leads to all of the target fragments being obtained. In contrast, insufficient amplification was observed for SNP 19 in this assay, such target fragment would be reamplified using a different primer pair.

The newly developed DigiTag assay described in this paper has the potential to accurately analyze almost all kinds of SNPs by applying mismatch-induced 5′-query probes and redesigned primer pairs. In this assay, the genotype information is encoded to one of the well-designed oligonucleotides called DCNs, which enable performance of SNP genotyping with high accuracy and reproducibility. Moreover, this assay has several other advantages: (i) any set of SNPs can be analyzed by the same DNA capillary array having the same set of probes, because the same set of DCNs can be unrestrainedly assigned to any set of target SNPs, thus reducing the cost of genotyping; (ii) oligonucleotide ligation assay, used in the encoding step, enables one to execute multiplex SNP typing; (iii) the same set of DCNs can also be used for other applications, such as gene expression profiling, thus ensuring low-cost analysis of genomic information [19,24]; and (iv) this assay can be readily adapted to high-throughput typing using automated equipment or a robot.

We have already prepared more than 100 available DCNs to perform SNP genotyping. We are currently attempting to scale up this technique to perform multiplex SNP typing for 100–500 SNPs using an improved DNA capillary array with more capillaries and probes. Our preliminary results show that this assay could perform multiplex typing for more than 100 SNPs if the target fragments are sufficiently amplified in the multiplex PCR. It would be essential to find an upper limit on the multiplex PCR, which forms a bottleneck for multiplexing. The simplification and automation of the assay protocols would lead to execution of multiplex genotyping in a high-throughput form.

Acknowledgments

This study was supported by a Grant-in-Aid for Scientific Research on Priority Areas (C) and by 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.2005.08.007.

References

[1] X. Zhu, Y.-P.C. Chang, D. Yan, A. Weder, R. Cooper, A. Luke, D. Kan, A. Chakravarti, Associations between hypertension and genes in the renin–angiotensin system, Hypertension 41 (2003) 1027–1034.

[2] T. Nakajima, L.B. Jorde, T. Ishigami, S. Umemura, M. Emi, J.-M. Lalouel, I. Inoue, Nucleotide diversity and haplotype structure of the human angiotensinogen gene in two populations, Am. J. Hum. Genet. 70 (2002) 108–123.

[3] Y. Horikawa, N. Oda, N.J. Cox, X. Li, M. Orho-Melander, M. Hara, Y. Hinokio, T.H. Lindner, H. Mashima, P.E.H. Schwarz, L. del Bosque-Plata, Y. Horikawa, Y. Oda, I. Yoshiuchi, S. Colilla, K.S. Polonsky, S. Wei, P. Concannon, N. Iwasaki, J. Schulze, L.J. Baier, C. Bogardus, L. Groop, E. Boerwinkle, C.L. Hanis, G.I. Bell, Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus, Nat. Genet. 26 (2000) 163–175.

[4] D. Altshuler, J.N. Hirschhorn, M. Klannemark, C.M. Lindgren, M.-C. Vohl, J. Nemesh, C.R. Lane, S.F. Schaffner, S. Bolk, C. Brewer, T. Tuomi, D. Gaudet, T.J. Hudson, M. Daly, L. Groop, E.S. Lander, The common PPARγ Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes, Nat. Genet. 26 (2000) 76–80.

[5] A. Suzuki, R. Yamada, X. Chang, S. Tokuhiro, T. Sawada, M. Suzuki, M. Nagasaki, M. Nakayama-Hamada, R. Kawaida, M. Ono, M. Ohtsuki, H. Furukawa, S. Yoshino, M. Yukioka, S. Tohma, T. Matsubara, S. Wakitani, R. Teshima, Y. Nishioka, A. Sekine, A. Iida, A. Takahashi, T. Tsunoda, Y. Nakamura, K. Yamamoto, Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis, Nat. Genet. 34 (2003) 395–402.

[6] N. Tsuchiya, J. Ohashi, K. Tokunaga, Variations in immune response genes and their associations with multifactorial immune disorders, Immunol. Rev. 190 (2002) 169–181.

[7] 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.

[8] 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.

[9] 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.

[10] R. Yamada, S. Tokuhiro, X. Chang, K. Yamamoto, SLC22A4 and RUNX1: identification of RA susceptible genes, J. Mol. Med. 82 (2004) 558–564.

[11] D.G. Wang, J.-B. Fan, C.-J. Siao, A. Berno, P. Young, R. Sapolsky, G. Ghandour, N. Perkins, E. Winchester, J. Spencer, L. Kruglyak, L. Stein, L. Hsie, T. Topaloglou, E. Hubbell, E. Robinson, M. Mittmann, M.S. Morris, N. Shen, D. Kilburn, J. Rioux, C. Nusbaum, S. Rozen, T.J. Hudson, R. Lipshutz, M. Chee, E.S. Lander, Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome, Science 280 (1998) 1077–1082.

[12] P. Hardenbol, J. Banér, M. Jain, M. Nilsson, E.A. Namsaraev, G.A. Karlin-Neumann, H. Fakhrai-Rad, M. Ronaghi, T.D. Willis, U. Landegren, R.W. Davis, Multiplexed genotyping with sequence-tagged molecular inversion probes, Nat. Biotechnol. 21 (2003) 673–678.

[13] 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.

[14] N. Pourmand, E. Elahi, R.W. Davis, M. Ronaghi, Multiplex pyrosequencing, Nucleic Acids Res. 30 (2002) e31.

[15] K. Lindroos, U. Liljedahl, M. Raitio, A.-C. Syvänen, Minisequencing on oligonucleotide microarrays: comparison of immobilisation chemistries, Nucleic Acids Res. 29 (2001) e69.

[16] J. Tost, I.G. Gut, Genotyping single nucleotide polymorphisms by mass spectrometry, Mass Spectrom. Rev. 21 (2002) 388–418.

[17] 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.

[18] H. Yoshida, A. Suyama, Solution to 3-SAT by breadth first search, DIMACS Series in Discrete Mathematics and Theoretical Computer Science 54 (2000) 9–22.

[19] N. Nishida, M. Wakui, K. Tokunaga, A. Suyama, Highly specific and quantitative gene expression profiling based on DNA computing, Genome Inform. 12 (2001) 259–260.
[20] F. Barany, Genetic disease detection and DNA amplification using cloned thermostable ligase, Proc. Natl. Acad. Sci. USA 88 (1991) 189–193.
[21] 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.
[22] 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.
[23] P. Markoulatos, N. Siafakas, M. Moncany, Multiplex polymerase chain reaction: a practical approach, J. Clin. Lab. Anal. 16 (2002) 47–51.
[24] Y. Sakakibara, A. Suyama, Intelligent DNA chips: Logical operation of gene expression profiles on DNA computers, Genome Inform. 11 (2000) 33–42.

3.1.6. Multiple-displacement amplification (MDA) provides greater accuracy in downstream genotyping assays

Recently, commercial kits (GenomiPhi, Amersham Biosciences; REPLI-g, Qiagen) employing MDA (see *Chapters 8–11*) have been used for WGA. When the amount of genomic DNA is sufficient (10 ng) for all samples, SNP typing using DOP-PCR products as template provides accurate results (see *Fig. 2*), comparable to those obtained using MDA products (see *Fig. 3*). However, we occasionally observed cases where MDA-generated DNA gave greater genotyping accuracy in mass SNP typing than DOP-PCR-generated DNA (see *Fig. 4*). In such cases, it is possible that suboptimal amounts of DNA are present in the DOP-PCR-amplified sample. When we compared DOP-PCR and MDA-amplified DNA for SNP typing, we succeeded in typing 82.1% (348 of 424) and 95.8% (68 out of 71) of the SNPs, respectively, when using the same genomic DNA template for WGA. For example, in *Fig. 4*, 34 out of 36 samples situated between clusters AA and CA (indicated by crosses) by DOP-PCR (see *Fig. 4a*) were classified in the CA cluster by MDA (see *Fig. 4b*).

3.2. Conclusion

SNP typing can be successfully performed using DOP-PCR-amplified DNA. However, it is important to ensure that sufficient starting template is used in the DOP-PCR and that an appropriate amount of DOP-PCR product is used for any subsequent PCRs. The genotypes determined by SSP-PCR and FCS using DOP-PCR samples were 100% in agreement with those determined by direct sequencing of genomic samples. Under these conditions, for most if not all cases, there should be no or very little biased amplification by DOP-PCR.

4. REFERENCES

★★★ 1. **Telenius H, Carter NP, Bebb CE, Nordenskjold M, Ponder BA & Tunnacliffe A** (1992) *Genomics*, **13**, 718–725. – *First report of DOP-PCR.*

2. **Dietmaier W, Hartmann A, Wallinger S, *et al.*** (1999) *Am. J. Pathol.* **154**, 83–95.

3. **Hirose Y, Aldape K, Takahashi M, Berger MS & Feuerstein BG** (2001) *J. Mol. Diagn.* **3**, 62–67.

4. **Huang Q, Schantz SP, Rao PH, Mo J, McCormick SA & Chaganti RS** (2000) *Genes Chromosomes Cancer*, **28**, 395–403.

5. **Kittler R, Stoneking M & Kayser M** (2002) *Anal. Biochem.* **300**, 237–244.

6. **Telenius H, Pelmear AH, Tunnacliffe A, *et al.*** (1992) *Genes Chromosomes Cancer*, **4**, 257–263.

7. **Umayahara K, Numa F, Suehiro Y, *et al.*** (2002) *Genes Chromosomes Cancer*, **33**, 98–102.

8. **Kallioniemi A, Kallioniemi OP, Sudar D, *et al.*** (1992) *Science*, **258**, 818–821.

★ 9. **Cheung VG & Nelson SF** (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93**, 14676–14679. – *The use of DOP-PCR-amplified DNA in SNP genotyping.*

10. **Grant SF, Steinlicht S, Nentwich U, Kern R, Burwinkel B & Tolle R** (2002) *Nucleic Acids Res.* **30**, e125.

11. **Bannai M, Higuchi K, Akesaka T, *et al.*** (2004) *Anal. Biochem.* **327**, 215–221.

12. **Jordan B, Charest A, Dowd JF, *et al.*** (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99**, 2942–2947.

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Analytical Biochemistry 346 (2005) 281–288
ANALYTICAL BIOCHEMISTRY
www.elsevier.com/locate/yabio

DigiTag assay for multiplex single nucleotide polymorphism typing with high success rate

Nao Nishida [a,b,*], Tetsuya Tanabe [c], Kento Hashido [b], Kouyuki Hirayasu [a], Miwa Takasu [a], Akira Suyama [d], Katsushi Tokunaga [a]

[a] *Department of Human Genetics, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan*
[b] *Biomedical Business Incubation Division, Olympus Corp., 2-3 Kuboyama-cho, Hachioji, Tokyo 192-8512, Japan*
[c] *R&D Division, NovusGene Inc., 2-3 Kuboyama-cho, Hachioji, Tokyo 192-8512, Japan*
[d] *Department of Life Sciences, Graduate School of Arts and Sciences, The University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8902, Japan*

Received 29 June 2005
Available online 31 August 2005

Abstract

As a consequence of Human Genome Project and single nucleotide polymorphism (SNP) discovery projects, several millions of SNPs, which include possible susceptibility SNPs for multifactorial diseases, have been revealed. Accordingly, there has been a strong drive to perform the investigation with all candidate SNPs for a certain disease without decreasing the number of analyzed SNPs. We developed DigiTag assay, which uses well-designed oligonucleotides called DNA coded numbers (DCNs) in multiplex SNP genotype analysis. During the analysis, the information of a genotype is converted to one of the DCNs in a one to one manner using oligonucleotide ligation assay (encoding). After the encoding reaction, only the DCNs regions and not the SNP specific regions are amplified using the universal primers and then SNP genotype is read out using DNA capillary arrays. DigiTag assay was found to be successful in SNP genotyping, giving a high success rate (24 of 27 SNPs) for randomly chosen SNPs. Moreover, this assay has the potential to analyze almost all kinds of the target SNPs by applying mismatch-induced probes and redesigned primer pairs at a low-cost.

Keywords: Genotyping method; Multiplex genotyping; SNPs; Mutation; Oligonucleotide ligation assay

Numerous single nucleotide polymorphisms (SNPs)[1] are considered to be candidate susceptibility or resistance genetic factors for multifactorial diseases such as hypertension, diabetes, and rheumatic diseases [1–5]. Large-scale case–control analyses of SNPs in candidate genes have revealed associations between various diseases and SNPs with the highest detection power [6–8]. Moreover, genome-wide association studies using SNPs have become important in the search for susceptibility and/or resistance genes [9,10]. Accordingly, large-scale whole-genome genotyping projects need high-throughput, cost-effective, and highly reliable technology to identify primary genes or SNPs. At present, there are a variety of SNP genotyping applications including microarray technology [11], molecular inversion probe genotyping [12], BeadArray genotyping technology (Illumina), 5′ exonuclease fluorescence-based assay (TaqMan) [13], pyrosequencing [14], single-base extension [15], matrix-assisted laser desorption/ionization time-of-flight mass spectrometry [16,17], and SNPlex (Applied Biosystems). However, many applications need to select relevant SNPs for their assay by in silico assay design, and then a portion of candidate SNPs will be excluded from investigation. To accomplish successful typing for all candidate SNPs at a low-cost, new technologies must be developed. In this study, we developed a new multiplex SNP typing method, named DigiTag assay, and performed typing for 28 SNPs using 40 genomic DNA samples. This approach

* Corresponding author. Fax: +81 3 5802 8619.
E-mail address: nishida-75@umin.ac.jp (N. Nishida).

[1] *Abbreviations used:* SNPs, single nucleotide polymorphisms; DCNs, DNA coded numbers; DTT, dithiothrecitol; SSC, standard saline citrate.

0003-2697/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.ab.2005.08.007

uses well-designed oligonucleotides called DNA coded numbers (DCNs), which enable performance of multiplex SNP genotyping with high accuracy and reproducibility.

Materials and methods

DNA samples

Genomic DNA samples from 40 unrelated healthy donors were obtained from the Japan Health Science Foundation (Osaka, Japan). All donors provided written informed consent and samples were anonymized. One hundred nanograms of purified genomic DNA was dissolved in 20 μl TE buffer, pH 8.0 (Wako, Osaka, Japan), for use and stored at 4 °C.

Preparation of DNA coded numbers

DCNs were designed to be 69-mer oligonucleotides (Fig. 1). DCNs consist of three parts, designated SD (start-digit), D1 (first-digit), and ED (end-digit). SD and ED are the common DNA sequences prepared at both edges in all DCNs and are used for priming sites in the labeling step. D1s are different DNA sequences among DCNs and are used to identify SNPs of interest. We prepared two EDs (ED-1 and ED-2) for two alleles at each SNP. The sequences of the three DCN components have the same length of 23-mer and the uniform melting temperature of 60.5 ± 0.9 °C. The assignment of DCNs to the SNPs analyzed in this study is listed in the Supplementary Table.

We designed DCNs (i) to have a uniform melting temperature and length, (ii) to ensure specific hybridization only to complementary DCNs, (iii) to minimize interaction with other DCNs, and (iv) to prevent the formation of secondary structures [18,19]. Therefore, all DCNs can be uniformly amplified in a multiplex manner using the common priming sites (SD, ED-1, and ED-2). Furthermore, we can perform precise hybridization on the DNA capillary array using a set of DCNs with high reproducibility.

Multiplex PCR from sample DNA

We designed multiplex PCR primers for each of the 28 SNP sites to have relatively long-length (average length 40-mer) and to give PCR product lengths of between 200 and 800 bp (average PCR product length 464 bp). To avoid spurious amplification products, we performed multiplex PCR by a two-step protocol (denature and extension steps) with elongated extension step for 6 min using specifically designed primer pairs.

Multiplex PCR was performed with 2.5 μl of genomic DNA and 1.25 pmol of primer pairs for 28 SNP sites in 50 μl of Multiplex PCR buffer, 0.2 μM dNTPs, and HotStarTaq DNA polymerase (Qiagen Multiplex PCR Kit, Qiagen, CA, USA). Cycling was performed as follows: 95 °C for 15 min, followed by 40 cycles of 95 °C for 30 s and 68 °C for 6 min. When necessary, the fragment length of 28 PCR products was confirmed by capillary electrophoresis (Agilent 2100 Bioanalyzer, Agilent, CA, USA) to evaluate the PCR efficiency.

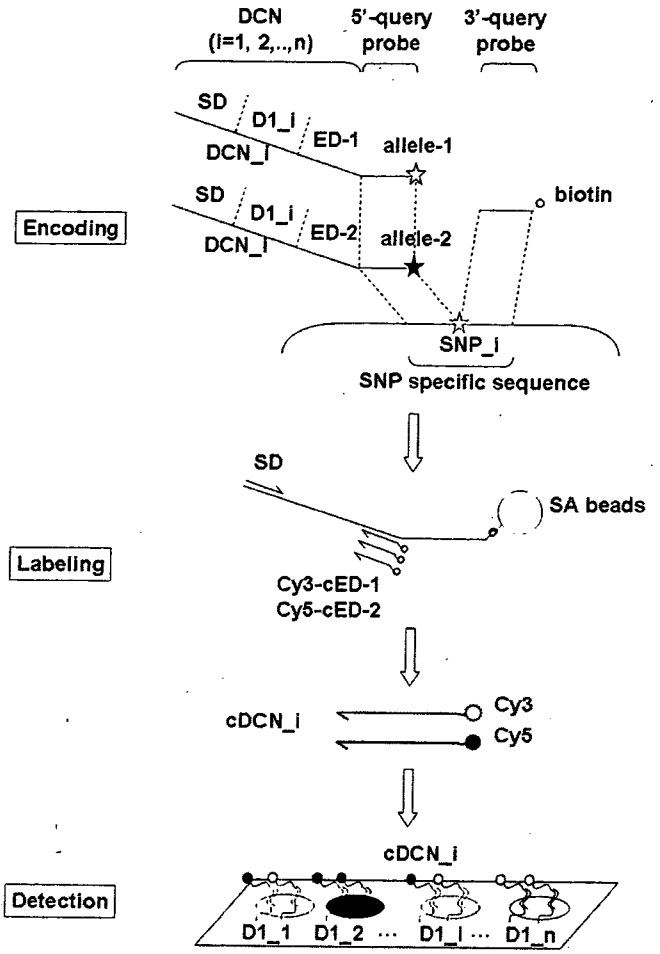

Fig. 1. Schematic representation of DigiTag assay. This assay has four steps to accomplish SNP typing: target preparation, encoding, labeling, and detection. DCNs are composed of three parts: SD, D1, and ED. SD and ED are the common DNA sequences prepared at both edges in all DCNs and are used for priming sites in the labeling step. Two EDs (ED-1 and ED-2) are prepared for two alleles at each SNP. D1s are different DNA sequences among DCNs and are used to identify SNPs of interest. DCN_i includes the common priming sites (SD, ED-1, and ED-2) and variable sequence (D1_i). Reverse.complement sequences are written by attaching the character c before the sequence name.

Encoding reaction

We prepared two 5′-query probes and one 3′-query probe for a single SNP site (Fig. 1). The 5′-query probes have the sequence complementary to that of the 5′ flanking of the target SNP (average length 20-mer) and each of the probes has an allele-specific sequence. Two types of DCNs, which have ED-1 and ED-2, were attached to each of the 5′-query probes. The 3′-query probe has the sequence complementary to that of the 3′ flanking of the target SNP (average length 20-mer) and has a phosphate group on its 5′ end and a biotin molecule on its 3′ end.

Two microliters of multiplex PCR product was mixed in 30 μl of 20 mM Tris–HCl, pH 7.6, 25 mM potassium acetate, 10 mM magnesium acetate, 10 mM DTT, 1 mM NAD, 0.1% Triton X-100 (*Taq* DNA ligase buffer, New England BioLabs, MA, USA) with 10 fmol of each probe (56 5′-query and 28 3′-query probes), 0.1 μl of control mix, and 20 U *Taq* DNA ligase. The control mix was prepared to assess each step of this assay including (i) 10 fmol of control target oligonucleotides and 10 fmol of two 5′-query probes and one 3′-query probe (assigned to DCN_29) for positive control of entire step, (ii) 0.1 fmol of 3′ end biotinylated DCN_30 for positive control of washing step with streptavidin-coupled magnetic beads, and (iii) 10 fmol of nonlabeled DCN_31 for negative control of washing step with streptavidin-coupled magnetic beads. All components of the encoding reaction were mixed on ice. The encoding reaction was first held at 95 °C for 5 min, followed by 58 °C for 15 min. The reaction was stopped by holding temperature at 10 °C.

Labeling of DCNs

The ligated products were washed with 1 × binding and washing buffer (1 M NaCl, TE, pH 8.0) twice at room temperature after binding to streptavidin-coupled magnetic beads (Dynabeads M-280 streptavidin, Dynal, Oslo, Norway), following the manufacturer's protocol. Alkali denaturation was performed to remove the multiplex PCR product and then asymmetric PCR was performed with single-strands of the ligated products binding to streptavidin-coupled magnetic beads, 1.0 pmol of SD, 10.0 pmol of Cy3-labeled reverse complement of ED-1 (Cy3-cED-1), 10.0 pmol of Cy5-labeled reverse complement of ED-2 (Cy5-cED-2), 2.5 U of *Ex Taq* polymerase in a 20 μl of 20 mM Tris–HCl, pH 8.0, 100 mM KCl, 0.1 mM EDTA, 1 mM DTT, 0.5% Tween 20, 0.5% Nonidet P-40, 50% glycerol, 2 mM each dNTP (*Ex Taq* Buffer, TaKaRa, Shiga, Japan). Asymmetric PCR was first held at 95 °C for 1 min, followed by 20 cycles of 95 °C for 30 s, 55 °C for 30 s, and 72 °C for 30 s.

Hybridization and detection on DNA capillary array

The DNA capillary array is a DNA detection device integrating oligonucleotide probes attached to specific locations in eight-parallel capillaries on a slide glass (see Fig. 3B, Olympus, Tokyo, Japan). Thirty-two types of oligonucleotide probes (28 probes for 28 SNPs and 4 probes for validation controls of the assay) identical to D1 sequences of DCNs were immobilized in each capillary. The ready-to-use DNA capillary arrays were stored in a desiccator at room temperature until use.

A hybridization mixture was prepared by mixing the supernatant of asymmetric PCR mixture in 24 μl of hybridization buffer containing 0.5 × SSC, 0.1% SDS, 15% formamide, 1 mM EDTA, with 2 μl of hybridization control. The hybridization control for ensuring the hybridization step was prepared with 5 fmol of Cy3-labeled D1_32 and Cy5-labeled D1_32. Twenty microliters of the hybridization mixture was applied to each capillary on the DNA capillary array. Hybridization was carried out for 30 min at 37 °C in a hybridization oven. After hybridization, the glass slides were washed in a washing buffer (0.1 × SSC, 0.1% SDS) by shaking at 60 rpm for 5 min. The glass slides were consecutively washed in distilled water by shaking at 60 rpm for 1 min and then dried by centrifugation at 2000 rpm for 1 min. Hybridization images were scanned at photomultiplier voltages of 400 V for Cy3 and 520 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, CA, USA).

Results

Schematic representation of DigiTag assay

This assay is performed in four steps: target preparation, encoding, labeling, and detection (Fig. 1). In this assay, multiplex PCR is performed with genomic DNA to prepare target fragments before the encoding step. For multiplex PCR, we designed the primer pairs to have 40-mer in average length and performed multiplex PCR by a two-step protocol (denature and extension steps) with elongated extension step for 6 min. The long-length primers and elongated extension step are essential for multiplex PCR to uniformly amplify all of the target fragments. In the encoding step, the 5′-query probe and 3′-query probe are successfully concatenated by *Taq* DNA ligase when two probes are fully complementary to adjacent regions on the target fragment [20]. The information of genotype is converted to one of the DCNs by a one to one manner in the encoding step. After the encoding step, single-strand forms of alkali denatured ligation products serve as templates in asymmetric PCR using Cy3- and Cy5-labeled primer pairs (SD, Cy3-cED-1, and Cy5-cED-2). The Cy3- and Cy5-labeled PCR products are gathered as single-strand forms of complementary DCNs and are then hybridized with the D1 probes on the DNA capillary array to reveal SNP genotypes by reading the 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 capillary array. In contrast, both signals are present when the sample is heterozygous.

Optimization of reaction condition in encoding step and DCN amplification rate

We first investigated the ligation conditions in the encoding step using a SNP located in the *PLOD* gene on human chromosome 1p36 as a model SNP (JSNP ID IMSJST068774). We prepared four types of 5′-query probes with four types of DCNs, each of which had one of the original SNP bases G and C and the two artificial SNP bases A

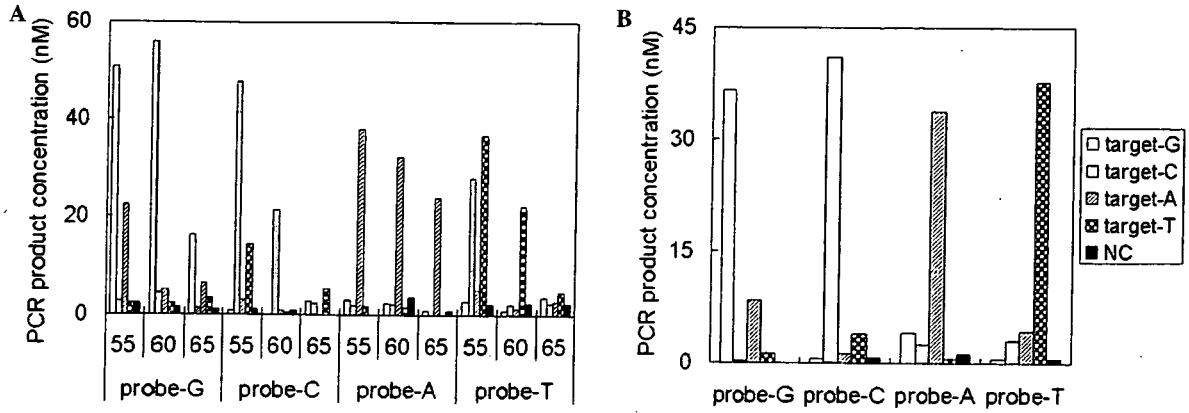

Fig. 2. Encoding rates of four SNP bases at different ligation temperatures. Encoding rate was compared as the amount of PCR product in the combination with four types of 5′-query probes and four types of target oligonucleotides. NC means that no target was included in the encoding reaction. Average amount of PCR product was calculated from three independent experiments. (A) The ligation reaction was performed at three temperatures: 55, 60, and 65 °C for 15 min. (B) The ligation reaction was performed at 58 °C for 15 min.

and T (see Supplementary Table). To investigate the encoding conditions, we prepared four types of 30-base target oligonucleotides identical to the SNP specific sequences. The encoding conditions were investigated by varying the ligation temperature from 55 to 65 °C (Fig. 2A). The encoding rate was compared as the amount of PCR products after performing 25 cycles of PCR with SD and reverse complement of ED (cED) primers using the ligated products as templates. When the ligation reaction was performed at 58 °C for 15 min, the signal intensities from perfect-match pair of 5′-query probe and target oligonucleotide were substantially higher than those from non-perfect-match pairs (Fig. 2B). False-positive signals were also suppressed at this ligation condition among the four SNP bases. False-positive signals increased at lower ligation temperatures, particularly when G-T mismatch occurred between 5′-query probe and target oligonucleotide. The intensities of positive signals decreased with ligation temperatures above 58 °C. Moreover, it became clear that the SNP base should be located at the 3′ end of the 5′-query probes to ensure precise discrimination between alleles (data not shown). When the SNP base was located at the 5′ end of the 3′-query probe, false-positive signals were significantly high and therefore resulted in incorrect genotyping.

PCR amplification rate was investigated by real-time PCR among DCNs used in multiplex SNP typing. A single DCN was added to a separate tube in 10-fold dilutions between 100 and 1 pM. PCR amplification rates of each DCN were calculated based on the results of real-time PCR. As expected, PCR amplification rate was found to be uniform at about 1.8 at concentrations between 100 and 1 pM (data not shown).

Multiplex typing for 28 SNPs using 40 genomic DNA samples

We then randomly selected 28 SNPs from a 500-kb region including the *IL-4* and *IL-13* genes on human chromosome 5q31-33, which contains several candidate genes related to immune disorders. We subsequently designed probes for the 28 SNP sites to give properties similar to those for *PLOD* SNP to obtain similar ligation efficiency among the 28 SNP sites to be analyzed in a single tube. The 5′-query probes and 3′-query probes were designed to have the uniform melting temperatures, 52.9 ± 1.8 and 55.0 ± 1.4 °C, respectively. The SNP genotypes of 40 genomic DNA samples were alternatively determined by direct sequencing and were used as reference data.

Multiplex PCR products including the 28 SNP sites showed similar band patterns with different individual DNA samples, although it was difficult to clearly discern all 28 PCR products due to the limitation of the electrophoretic resolution (Fig. 3A). Multiplex SNP typing for 28 SNPs was then performed using the multiplex PCR products as targets. The DNA capillary array demonstrated hybridization images of each sample in each capillary having 32 spots (28 probes for 28 SNPs and 4 probes for validation controls, see Fig. 3B). 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 (Fig. 3C). The SNP 13 was monomorphic in 40 genomic DNA samples and was excluded from further analysis. For 24 SNPs (except for SNP 6, SNP 9, and SNP 19), three distinct clusters corresponding to two homozygous and one heterozygous genotypes were observed, although the average signal intensity for the 24 SNP sites fluctuated between 100 and 16,000. The fluctuated intensities presumably result from different efficiencies of PCR in the target preparation step by the multiplex PCR.

Indistinct clusters were observed for SNP 6, SNP 9, and SNP 19. For SNP 6 and SNP 9, the false-positive signals were detected when we performed singleplex SNP typing using target oligonucleotides identical to the SNP specific sequences (Figs. 4A and B). To suppress the false-positive signals observed from SNP 6 and SNP 9, we prepared

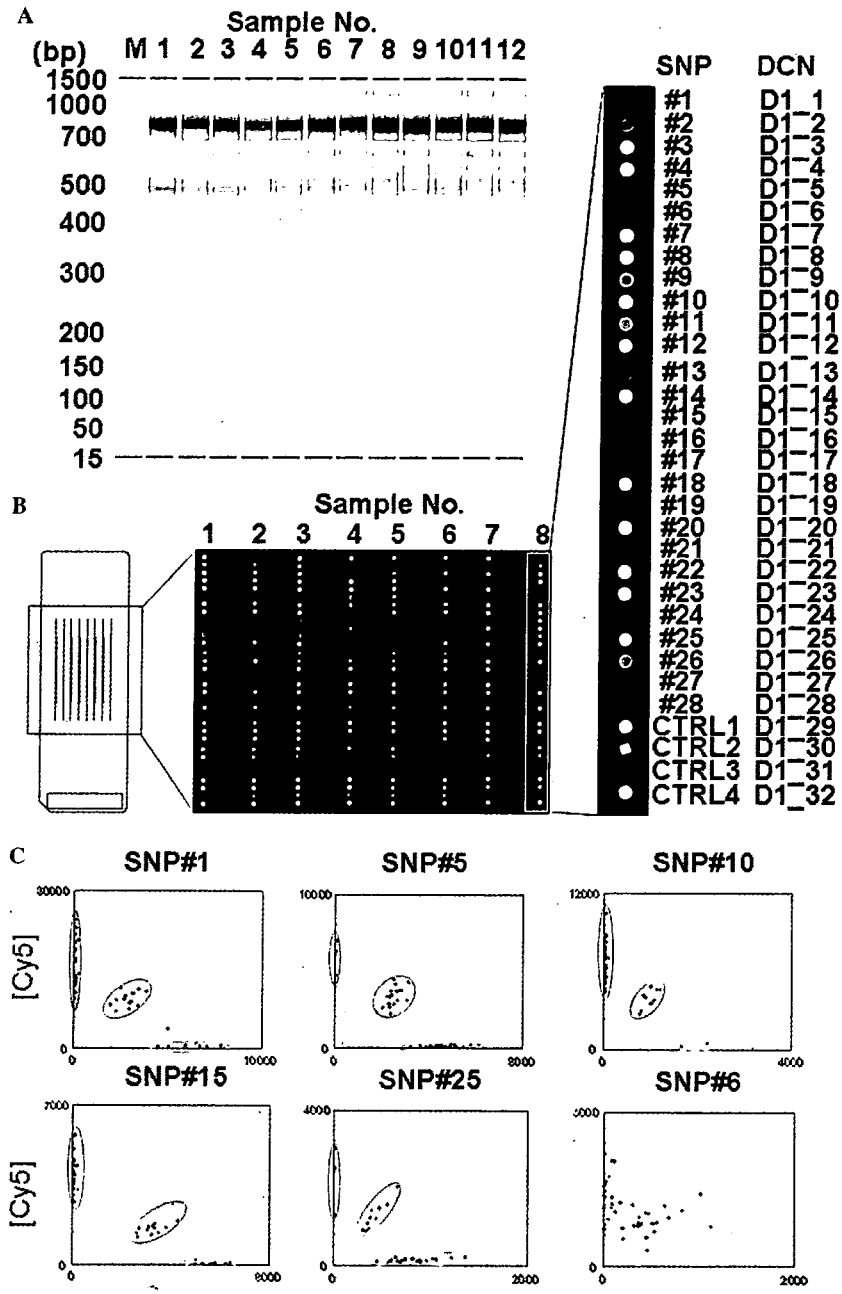

Fig. 3. Multiplex SNP typing for 28 SNPs using 40 genomic DNA samples. (A) Gel images of multiplex PCR products with different individual samples. In all sample lanes, the sample bands were observed between two inner markers; 15 and 1500 bp. (B) Hybridization images of the DNA capillary array. (C) Scatter diagrams for randomly chosen 6 SNPs from 28 SNPs.

three types of mismatch-induced 5′-query probes, which had an artificial mismatch at the fourth position from the SNP base. The encoding rate was investigated using mismatch-induced 5′-query probes and target oligonucleotides identical to the SNP specific sequences by comparing the amount of PCR products after performing 25 cycles of PCR using the ligated products as templates. All of the three mismatch-induced 5′-query probes can effectively suppress the false-positive signals without diminishing the positive signals from the perfect-match pair of the mismatch-induced 5′-query probe and target oligonucleotide (Figs. 4A and B). We then performed multiplex typing using one of the mismatch-induced 5′-query probes for each of two SNPs. Scatter diagrams showed that three clusters from perfect-match 5′-query probes were indistinctively observed (Figs. 4C and D) as compared to the mismatch-induced 5′-query probes (Figs. 4E and F). For SNP 19, many samples did not belong to any clusters, presumably due to insufficient amplification of the target fragment in multiplex PCR.

In 24 successfully genotyped 24 SNPs, 8 miscallings were observed among 960 genotypes. The percentage of calling rate (number of identified SNPs divided by the total number examined) was 99.2% and the concordance rate with

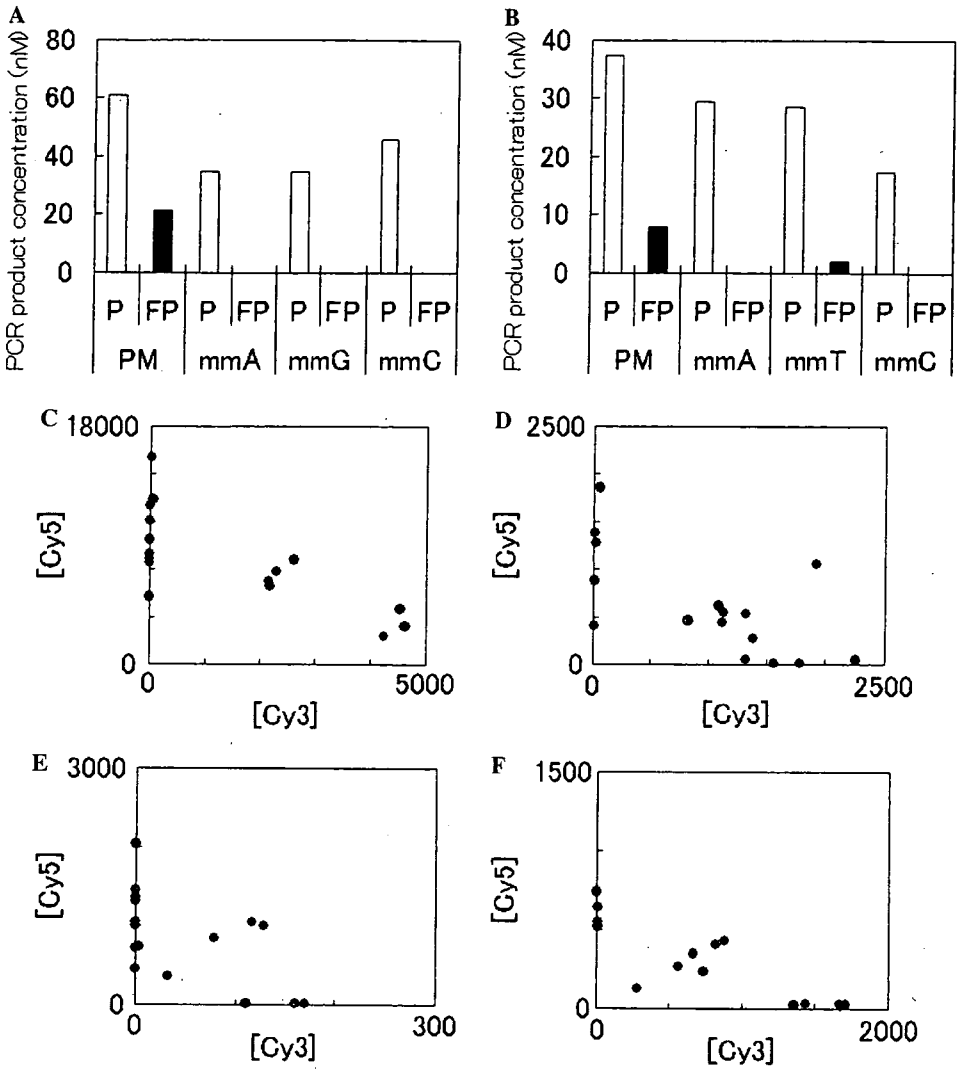

Fig. 4. Suppression of false-positive signals by mismatch-induced 5′-query probes. (A and B) Both graphs show the singleplex typing result of SNP 6 and SNP 9, respectively. PM indicates the perfect-match pair of 5′-query probe and target oligonucleotide. Three types of mismatch-induced 5′-query probes which had an artificial mismatch base at the fourth position from the SNP base were prepared (mmA, mmG, and mmC for SNP 6; mmA, mmT, and mmC for SNP 9). P and FP mean positive signals and false-positive signals, respectively. Scatter diagrams show the multiplex typing result of SNP 6 and SNP 9, respectively: (C and D) using perfect-match 5′-query probes and (E and F) using mismatch-induced 5′-query probes (mmC for SNP 6 mmA for SNP 9).

reference data from direct sequencing was 100%. The reproducibility of this assay was examined by duplicate experiments and was observed higher than 0.99 in r^2 except for 2 SNPs (SNP 2, 0.96; SNP 20, 0.96).

Discussion

We developed DigiTag assay and performed multiplex typing for 28 SNPs using 40 genomic DNA samples. One of the 28 SNPs (SNP 13) was revealed to be monomorphic in 40 samples and was therefore excluded from further analysis. This assay was found to be successful in SNP genotyping, giving a high success rate (24 of 27 SNPs) for randomly chosen SNPs. Three SNPs of 27 SNPs showed indistinct clusters, presumably resulting from missligation in the encoding step (SNP 6 and SNP 9) and insufficient amplification in the multiplex PCR (SNP 19).

The missligation was reported to be prone to occur when mismatched pairs are G-T, G-A, G-G, and A-G [21]. In our results, one of two missligated SNPs had G-G mismatch (SNP 6) and another one had G-T mismatch (SNP 9) between the 5′-query probe and the target fragment. The missligation results in increasing the false-positive signals and then leads to ambiguous genotyping. To suppress the false-positive signals, we designed mismatch-induced 5′-query probes, which had an artificial mismatched base at the fourth position from the SNP base. The mismatch-induced 5′-query probes could effectively suppress the false-positive signals without diminishing the positive signals.

Multiplex PCR has the potential to reduce the complexity fraction of the genome by selectively collecting the target SNP sites from the genome, which would lead to successful genotyping [22]. However, it was also reported that the presence of multiple primer pairs increases the chance of obtain-

ing spurious amplification products such as primer dimers [23]. To avoid spurious amplification products, we designed primer pairs to have relatively long-length (average length 40-mer) and performed multiplex PCR by a two-step protocol with an elongated extension step for 6 min. This optimization of multiplex PCR leads to all of the target fragments being obtained. In contrast, insufficient amplification was observed for SNP 19 in this assay, such target fragment would be reamplified using a different primer pair.

The newly developed DigiTag assay described in this paper has the potential to accurately analyze almost all kinds of SNPs by applying mismatch-induced 5′-query probes and redesigned primer pairs. In this assay, the genotype information is encoded to one of the well-designed oligonucleotides called DCNs, which enable performance of SNP genotyping with high accuracy and reproducibility. Moreover, this assay has several other advantages: (i) any set of SNPs can be analyzed by the same DNA capillary array having the same set of probes, because the same set of DCNs can be unrestrainedly assigned to any set of target SNPs, thus reducing the cost of genotyping; (ii) oligonucleotide ligation assay, used in the encoding step, enables one to execute multiplex SNP typing; (iii) the same set of DCNs can also be used for other applications, such as gene expression profiling, thus ensuring low-cost analysis of genomic information [19,24]; and (iv) this assay can be readily adapted to high-throughput typing using automated equipment or a robot.

We have already prepared more than 100 available DCNs to perform SNP genotyping. We are currently attempting to scale up this technique to perform multiplex SNP typing for 100–500 SNPs using an improved DNA capillary array with more capillaries and probes. Our preliminary results show that this assay could perform multiplex typing for more than 100 SNPs if the target fragments are sufficiently amplified in the multiplex PCR. It would be essential to find an upper limit on the multiplex PCR, which forms a bottleneck for multiplexing. The simplification and automation of the assay protocols would lead to execution of multiplex genotyping in a high-throughput form.

Acknowledgments

This study was supported by a Grant-in-Aid for Scientific Research on Priority Areas (C) and by 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.2005.08.007.

References

[1] X. Zhu, Y.-P.C. Chang, D. Yan, A. Weder, R. Cooper, A. Luke, D. Kan, A. Chakravarti, Associations between hypertension and genes in the renin–angiotensin system, Hypertension 41 (2003) 1027–1034.

[2] T. Nakajima, L.B. Jorde, T. Ishigami, S. Umemura, M. Emi, J.-M. Lalouel, I. Inoue, Nucleotide diversity and haplotype structure of the human angiotensinogen gene in two populations, Am. J. Hum. Genet. 70 (2002) 108–123.

[3] Y. Horikawa, N. Oda, N.J. Cox, X. Li, M. Orho-Melander, M. Hara, Y. Hinokio, T.H. Lindner, H. Mashima, P.E.H. Schwarz, L. del Bosque-Plata, Y. Horikawa, Y. Oda, I. Yoshiuchi, S. Colilla, K.S. Polonsky, S. Wei, P. Concannon, N. Iwasaki, J. Schulze, L.J. Baier, C. Bogardus, L. Groop, E. Boerwinkle, C.L. Hanis, G.I. Bell, Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus, Nat. Genet. 26 (2000) 163–175.

[4] D. Altshuler, J.N. Hirschhorn, M. Klannemark, C.M. Lindgren, M.-C. Vohl, J. Nemesh, C.R. Lane, S.F. Schaffner, S. Bolk, C. Brewer, T. Tuomi, D. Gaudet, T.J. Hudson, M. Daly, L. Groop, E.S. Lander, The common PPARγ Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes, Nat. Genet. 26 (2000) 76–80.

[5] A. Suzuki, R. Yamada, X. Chang, S. Tokuhiro, T. Sawada, M. Suzuki, M. Nagasaki, M. Nakayama-Hamada, R. Kawaida, M. Ono, M. Ohtsuki, H. Furukawa, S. Yoshino, M. Yukioka, S. Tohma, T. Matsubara, S. Wakitani, R. Teshima, Y. Nishioka, A. Sekine, A. Iida, A. Takahashi, T. Tsunoda, Y. Nakamura, K. Yamamoto, Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis, Nat. Genet. 34 (2003) 395–402.

[6] N. Tsuchiya, J. Ohashi, K. Tokunaga, Variations in immune response genes and their associations with multifactorial immune disorders, Immunol. Rev. 190 (2002) 169–181.

[7] 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.

[8] 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.

[9] 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.

[10] R. Yamada, S. Tokuhiro, X. Chang, K. Yamamoto, SLC22A4 and RUNX1: identification of RA susceptible genes, J. Mol. Med. 82 (2004) 558–564.

[11] D.G. Wang, J.-B. Fan, C.-J. Siao, A. Berno, P. Young, R. Sapolsky, G. Ghandour, N. Perkins, E. Winchester, J. Spencer, L. Kruglyak, L. Stein, L. Hsie, T. Topaloglou, E. Hubbell, E. Robinson, M. Mittmann, M.S. Morris, N. Shen, D. Kilburn, J. Rioux, C. Nusbaum, S. Rozen, T.J. Hudson, R. Lipshutz, M. Chee, E.S. Lander, Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome, Science 280 (1998) 1077–1082.

[12] P. Hardenbol, J. Banér, M. Jain, M. Nilsson, E.A. Namsaraev, G.A. Karlin-Neumann, H. Fakhrai-Rad, M. Ronaghi, T.D. Willis, U. Landegren, R.W. Davis, Multiplexed genotyping with sequence-tagged molecular inversion probes, Nat. Biotechnol. 21 (2003) 673–678.

[13] 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.

[14] N. Pourmand, E. Elahi, R.W. Davis, M. Ronaghi, Multiplex pyrosequencing, Nucleic Acids Res. 30 (2002) e31.

[15] K. Lindroos, U. Liljedahl, M. Raitio, A.-C. Syvänen, Minisequencing on oligonucleotide microarrays: comparison of immobilisation chemistries, Nucleic Acids Res. 29 (2001) e69.

[16] J. Tost, I.G. Gut, Genotyping single nucleotide polymorphisms by mass spectrometry, Mass Spectrom. Rev. 21 (2002) 388–418.

[17] 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.

[18] H. Yoshida, A. Suyama, Solution to 3-SAT by breadth first search, DIMACS Series in Discrete Mathematics and Theoretical Computer Science 54 (2000) 9–22.

[19] N. Nishida, M. Wakui, K. Tokunaga, A. Suyama, Highly specific and quantitative gene expression profiling based on DNA computing, Genome Inform. 12 (2001) 259–260.
[20] F. Barany, Genetic disease detection and DNA amplification using cloned thermostable ligase, Proc. Natl. Acad. Sci. USA 88 (1991) 189–193.
[21] 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.
[22] 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.
[23] P. Markoulatos, N. Siafakas, M. Moncany, Multiplex polymerase chain reaction: a practical approach, J. Clin. Lab. Anal. 16 (2002) 47–51.
[24] Y. Sakakibara, A. Suyama, Intelligent DNA chips: Logical operation of gene expression profiles on DNA computers, Genome Inform. 11 (2000) 33–42.

大規模SNPタイピングによる多因子疾患遺伝子の探索

西田奈央，徳永勝士

ヒトゲノム計画をはじめとする遺伝情報解析の成果としてデータベースに蓄積された600万種を超える一塩基多型（SNP）のうち，約50万種のSNPを同時にタイピングすることのできる手法が近年になって実用化された．われわれは，500K Human Mapping Array Set（Affymetrix社）によるSNPタイピングを効率的に行うためのシステムを構築し，大規模なケース・コントロール関連分析を行っている．一例として，睡眠障害の1つであるナルコレプシーの疾患感受性遺伝子の探索研究を紹介する．

はじめに

現在，ヒトのさまざまな多因子疾患に関わる遺伝子を探索する戦略として最も注目を浴びているのが，本稿で紹介するゲノムワイド関連分析法である．例を挙げると，米国NIH（国立衛生研究所）はGWAS（genomewide association study）計画を提案し，いくつかのcommon diseasesについて大規模な研究チームを公募した．また，大規模な疫学研究として知られるFramingham Heart Study（フラミンガム研究）で収集された試料のうち9,000検体について，本稿で紹介する技術を用いて解析することにより，心，肺，血液，睡眠疾患に関与する遺伝子変異を探索する計画も発表された．

[キーワード&略語]
一塩基多型（SNP），SNPタイピング，ケース・コントロール関連分析，疾患感受性遺伝子，ナルコレプシー
GWAS：genomewide association study
LD：linkage disequilibrium（連鎖不平衡）
SNP：single nucleotide polymorphism（一塩基多型）

このような動向の背景には，多因子疾患感受性遺伝子を探索する統計遺伝学的手法の選択がある．大別して連鎖分析（linkage analysis）と関連分析（association analysis）があり，従来のゲノムワイド探索のほとんどは連鎖分析法が用いられてきた．多数の成功例があるものの，連鎖分析のみで疾患感受性遺伝子のほとんどを同定するのは困難であると考えられるようになり，また後述する大規模なSNP解析技術の進展も相まって，ゲノムワイド関連分析が大きく取り上げられることとなった．そもそも連鎖分析は患者家族を対象として文字通り疾患遺伝子と多型マーカーの連鎖を検出する手法であり，ゲノム全域に分布する1万から数万種のSNPを用いればよい．一方，関連分析の代表であるケース・コントロール関連分析法※1は，非血縁の患者集団と健常者集団を対象として疾患遺伝子と多型マーカーの連鎖不平衡（LD）を検出する手法であり，ゲノム全域に分布する数十万種以上のSNPを解析することが必要となる[1]（統計遺伝学手法の原理ならびに集団遺伝学的基礎については別書に解説した[2]．

Search for susceptibility genes to multifactorial diseases by large-scale SNP typing
Nao Nishida/Katsushi Tokunaga：Department of Human Genetics, Graduate School of Medicine, University of Tokyo（東京大学大学院医学系研究科人類遺伝学分野）

ヒトゲノム上には，マイナーアレル頻度が10％以上のSNPsが500万種，マイナーアレル頻度が1％以上のSNPsとなると1,100万種も存在すると試算されており[3]，これらのSNPsを大規模にかつ正確にタイピングのできる手法の確立が求められてきた．

SNPタイピング法として最もよく知られている方法は，個々の多型部位を含むゲノム断片を特異的にPCR増幅した後にアレルを識別する方法である[4)~7)]．これらの方法では，1,000種程度のSNPのタイピングであれば，PCRプライマーはじめ各種試薬にかかるコストを考えて実用可能であるといえるが，数千，数万種を超える数のSNPをタイピングすることは困難である．一方，近年になって多型部位特異的なPCRを行わずに大規模なSNPタイピングを行う方法が実用化された[8) 9)]．その一つであるAffymetrix社によって確立された方法では，まず制限酵素反応でゲノムDNAの断片化を行い，続いてそれら断片の両端にアダプター配列を付加し，まとめて増幅した後にマイクロアレイを用いたアレル特異的なハイブリダイゼーションを行う[8)]．現在では，この手法を用いて50万種を超えるSNPを同時にタイピングするキットが市販されている．われわれは500K Human Mapping Array Setを用いる大規模なSNP解析システムを構築し，いくつかの多因子疾患についてゲノムワイド関連分析を実施している．その一例としてナルコレプシーの疾患感受性遺伝子の探索研究を紹介する．

1 技術原理

500K Human Mapping Array Set（以下，500K Array Set）は，制限酵素によるゲノムDNAの断片化とマイクロアレイによるタイピングの手法に改良を加えることにより，大規模なタイピングを行える手法として確立された[8)]．解析対象となるSNPは，公共のSNPデータベースおよびPerlegen社に登録されている約220万種のSNPから遺伝学的情報量が最大化されるように，また連鎖不平衡（LD）やHapMapプロジェクトからの情報も考慮して選択された約50万種のSNPである．

500K Array SetによるSNPタイピングは，ゲノムの複雑さを低減しマイクロアレイへのハイブリダイゼーション効率を上げるための酵素反応ステップと，洗浄・染色装置およびマイクロアレイ用スキャナーを用いた検出ステップに分けることができる（図1）．50万種のSNPタイピングは，2種類の制限酵素（*Sty*Ⅰ，*Nsp*Ⅰ）を用いてそれぞれ約25万種のSNPを独立にタイピングすることで実現される．制限酵素によるゲノムDNAの断片化を行った後，断片化されたゲノムDNAの両末端にアダプター配列をライゲーション反応により付加する．アダプター配列は，続くPCRで使用されるプライマーと相同な配列をもち，また制限酵素認識配列を突出端としてもつ二本鎖DNAである．2種類の制限酵素（*Sty*Ⅰ，*Nsp*Ⅰ）のそれぞれに対して用意されるアダプター配列は，制限酵素認識配列を除いて共通の配列をもっているので共通のプライマーを使用してPCRを行うことができる．PCRでは，目的の長さをもったゲノムDNA断片（200～1,100 bp）だけが選択的に増幅される．ここまでの酵素反応により，もともと30億塩基対のゲノムDNAが5億塩基対程度のPCR混合産物となる．マイクロアレイへの効率的なハイブリダイゼーションには，ゲノムの複雑さを低減することが大きな役割を果たすと考えられている[10) 11)]．続いて，PCR産物の精製を行った後，DNaseⅠ制限酵素によりPCR産物の断片化を行う．断片化されたPCR産物は平均長で180 bp以下となる．マイクロアレイへの効率的なハイブリダイゼーションには，ゲノムの複雑さを低減することに加えてPCR産物の断片化が重要になる．最後にterminal deoxynucleotidyl transferase酵素反応により，断片化されたPCR産物の末端にビオチンを導入する．

続いて，専用のマイクロアレイを用いてハイブリダイゼーションを行う．マイクロアレイには解析対象となる各SNPに対して合計24本のプローブが用意されている．プローブは25塩基長のオリゴDNAで，SNP部位を含む塩基配列をもっている．2種類のアレルに対して完全に相補的な塩基配列をもつプローブ（PMプローブ）と1塩基のミスマッチを含むプローブ（MMプローブ）を用意し，4種類のプローブを1組のプローブセットとしている．SNP部位を25塩基長のプローブの中心に置いたプローブセットを基本として，SNP

※1 ケース・コントロール関連分析

ある疾患に罹っている患者群（ケース）と健常者群（コントロール）とで遺伝子・ゲノム多型の頻度に差があるかどうかを検定することにより，疾患関連遺伝子を探索するための統計遺伝学的方法．

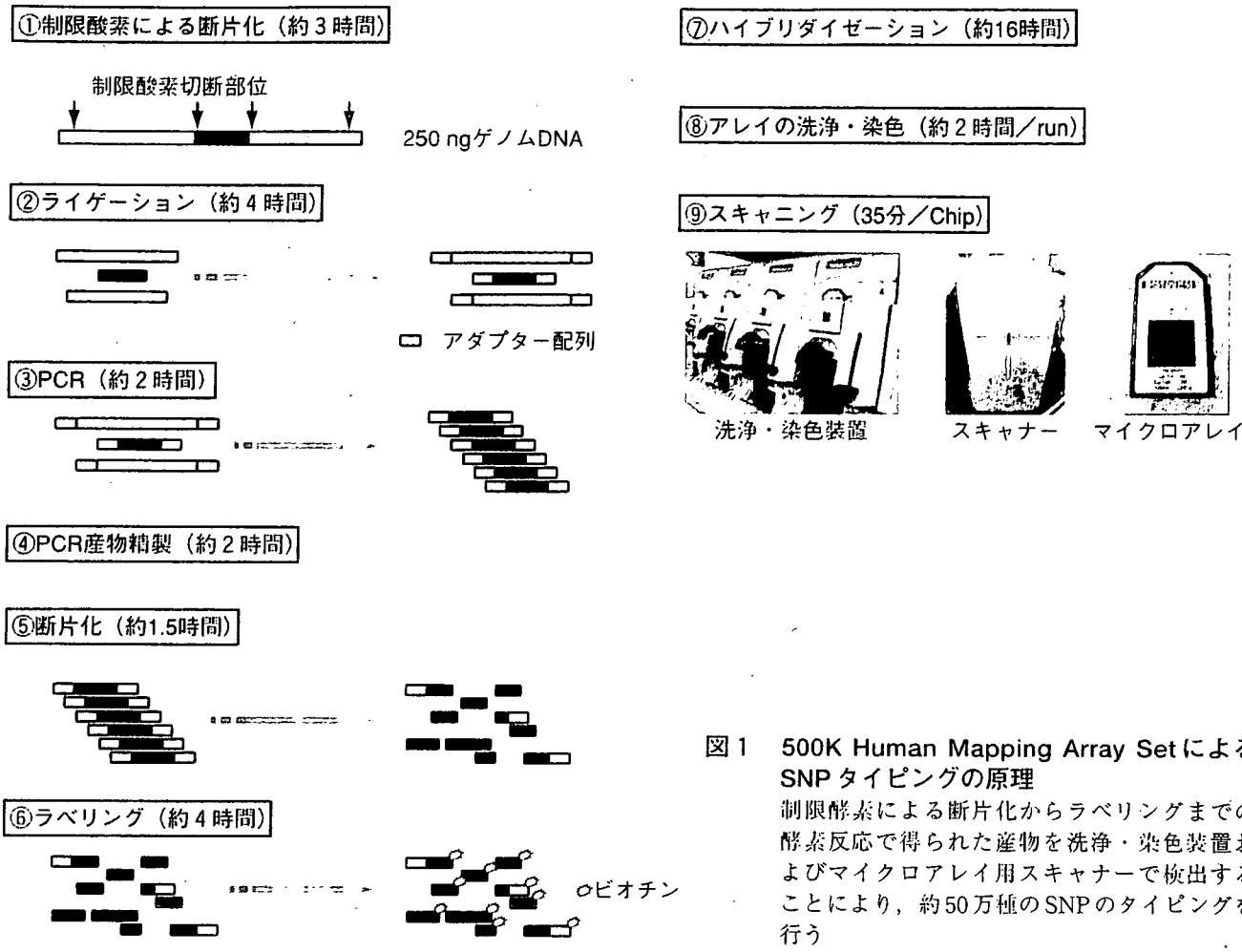

図1　500K Human Mapping Array SetによるSNPタイピングの原理
制限酵素による断片化からラベリングまでの酵素反応で得られた産物を洗浄・染色装置およびマイクロアレイ用スキャナーで検出することにより，約50万種のSNPのタイピングを行う

部位を中心から4塩基上流（＋4）にずらしたプローブセットから4塩基下流（－4）にずらしたプローブセットまで7組のプローブセット（－4，－2，－1，0，＋1，＋3，＋4）の中から3組のプローブセットを選択する．3組のプローブセットはゲノムDNAの両鎖に対して用意されているので，合計6組のプローブセットが各SNPに対して用意されている．また，遺伝学的に重要だとされる約52,000 SNPsに関してはプローブセットを合計10組とし，計40本のプローブを用意している．すべてのプローブセットからのPMプローブとMMプローブのシグナル強度差を検出することで遺伝子型の判定を行うことができる．

マイクロアレイへのハイブリダイゼーションが終了した後，洗浄・染色装置を用いてマイクロアレイの洗浄および蛍光染色を行う．蛍光染色は，蛍光分子で標識されたストレプトアビジンが，前述のビオチン導入されたPCR断片に結合することにより行われる．また，洗浄・染色装置内ではビオチン修飾された抗ストレプトアビジン抗体を用いてシグナルの増強が行われる．最後に蛍光染色されたマイクロアレイを専用のスキャナーで画像データとして読み取り，続いて専用の画像解析ソフトウェアを用いて各SNPの遺伝子型を決定する．

2 システム構築

1）ハードウェアの整備

500K Array SetによるSNPタイピングを効率的に行うために，環境，装置を整備し，作業マニュアルを作成した．まず，ゲノムDNAへのPCR産物のコンタミネーションを防ぐために，試料調製室とSNP解析室を設けた．試料調製室にはゲノムDNAを保管し，PCR以前の酵素反応を行うのに必要な装置（サーマルサイクラーなど）を用意した．制限酵素による断片化からPCRの反応溶液の調製までは試料調製室で行い，PCR以降の酵素反応はSNP解析室で行った．また，3台の洗浄・染色装置を用意し，1回のランで計

12枚のマイクロアレイを洗浄・染色することができるようにした．すべてのマイクロアレイはバーコードで管理され，洗浄・染色が終了したマイクロアレイはオートローダー付きのマイクロアレイ用スキャナーに装填され画像データが読み込まれる．オートローダー付きのマイクロアレイ用スキャナーは計64枚のマイクロアレイを装填することができ，バーコードを参照しながらすべてのマイクロアレイの画像データを自動的に読み込むことができる．

SNPタイピング作業のルーチン化にあたって，制限酵素による断片化からラベリングまでの5つの酵素反応ステップで使用するすべてのマイクロタイタープレートをバーコードで管理する（図2）．また，1枚のマイクロタイタープレートで32検体分の酵素反応を行うこととし，制限酵素による断片化で使用するマイクロタイタープレートのウェル位置をサンプルと対応させることでサンプルのID化を行った．制限酵素による断片化以降の酵素反応で使用するマイクロタイタープレートでもレイアウトを変えずに酵素反応を行うことで，ウェル位置をサンプルIDとして解析結果を得ることができる．また，酵素反応の各工程を管理するためにチェックシートを作成し，反応工程の進行を随時チェックシートで確認しながら進める．PCRおよび断片化の酵素反応の後にはアガロースゲル電気泳動を行い，PCR産物および断片化産物の平均長がそれぞれ200～1,100 bp，180 bp以下となっていることを確認する．加えて，同一のマイクロプレート上で酵素反応を行った32検体のうちから4検体だけを先行してハイブリダイゼーションを行い，SNPコール率に問題がないことを確認した後で残り28検体のハイブリダイゼーションを行った．28検体のハイブリダイゼーションを行う際に，次のマイクロタイタープレートから4検体を加えて合計32検体のハイブリダイゼーションを順次行っていくこととした．

2）ソフトウェアの開発

500K Array SetによるSNPタイピングではGeneChip® Operating Software（GCOS）とGeneChip® Genotyping Analysis Software（GTYPE）という2種類のソフトウェア（ともにAffymetrix社）を使用する．GCOSソフトウェアは洗浄・染色装置およびマイクロアレイ用スキャナーを操作する際に使用し，またGTYPEソフトウェアはマイクロアレイの画像データから遺伝子型を判定する際に使用する．GTYPEソフトウェアで決定された約25万種のSNPの遺伝子型は，*Sty*I，*Nsp*Iごとにテキストファイルとして転送することができる．

われわれは500K Array SetによるSNPタイピングから得られる約50万SNPsの遺伝子型情報を用いてケース・コントロール関連分析を行うためのソフトウェアを開発した．ケース・コントロール関連分析をするにあたって，*Sty*Iおよび*Nsp*Iごとにまとめられた約25万SNPsの遺伝子型データを検体ごとに統合し，さらに検体をケース群およびコントロール群に分けて新たなテキストファイルとして作成する機能をソフトウェアに加えた．続いて，作成したケース群およびコントロール群の解析結果のテキストファイルを使ってケース・コントロール関連解分析を行った．この際，各コントロール群における各遺伝子型の観察数からも，ハーディー・ワインベルク平衡※2の検定を行うこととした．ケース・コントロール関連分析の結果はレポートファイルとしてまとめられ，専用のビューアーを用いて表示することができる．

3 ナルコレプシー感受性領域のゲノムワイド探索

われわれは上に述べた大規模SNPタイピングシステムを用いて，文部科学省科学研究費特定領域研究「基盤ゲノム」におけるSNPタイピングセンターとして，数種の疾患のゲノムワイド関連分析を実施しており，すでに2種の疾患については約200名ずつの患者試料と約200名の健常者試料の解析を終了している．

またわれわれは睡眠障害の1つナルコレプシーの感受性・抵抗性遺伝子をゲノムワイド関連分析法によって探索している．すでに2万3千種のマイクロサテライト多型を用いたゲノムワイド関連解析を行い，11カ所の候補領域を検出するとともに，その1つから新たな疾患抵抗性遺伝子を同定した[12]．現在，他の候補領域についても詳細な解析を行っているが，これと平行して，新たに50万種のSNPsを用いたゲノムワイド関

※2 ハーディー・ワインベルク平衡
自然淘汰が働かず，突然変異によって新たな対立遺伝子が生じず，また移住や混血などが起こらない十分に大きな規模の集団においては，対立遺伝子の頻度は世代を経ても変化しないという集団遺伝学の基本的法則．

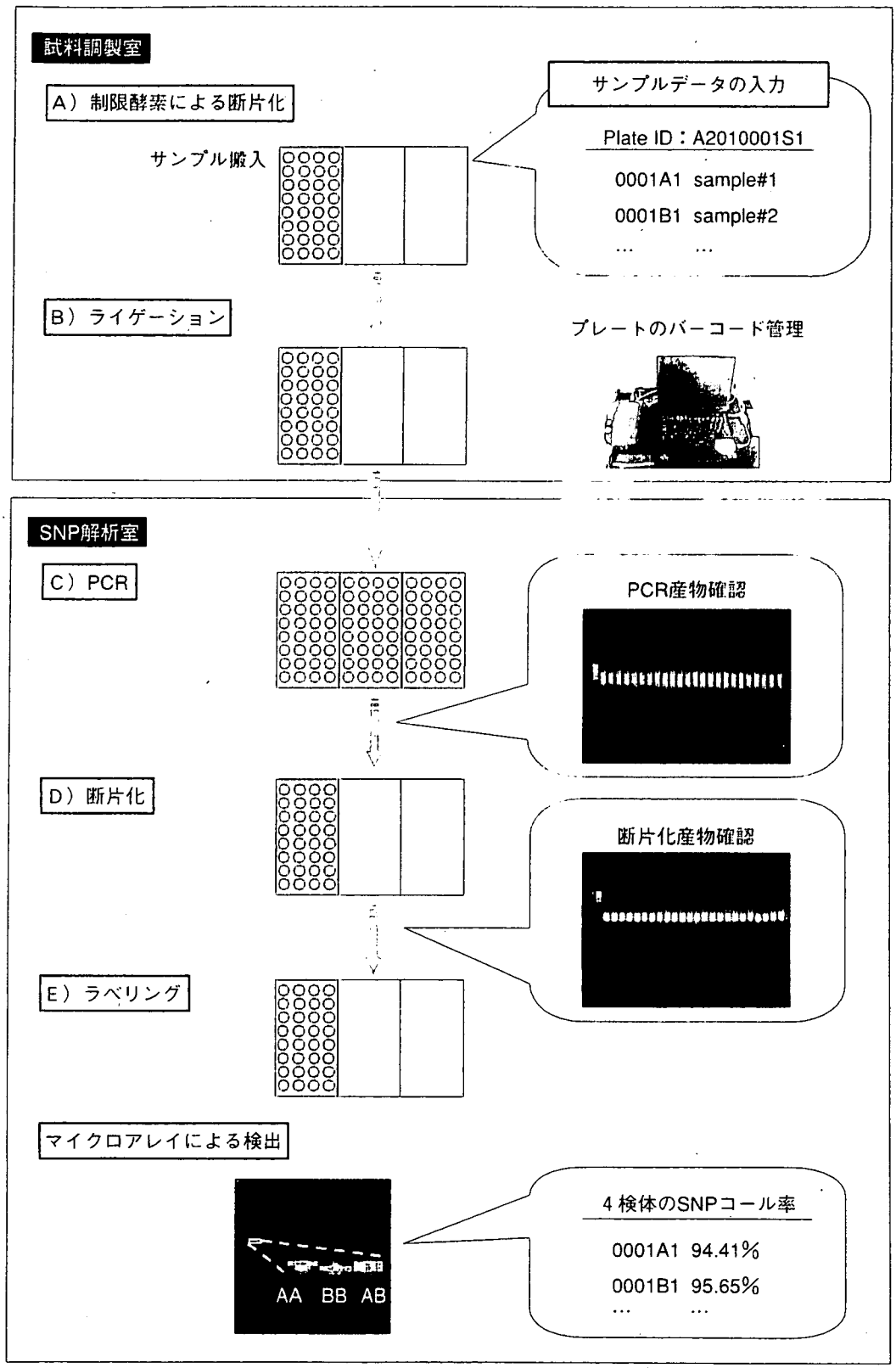

図2　SNPタイピングシステムの構築

500K Human Mapping Array Setを用いたSNPタイピングを効率的に行うためのシステム環境を構築し，作業マニュアルを作成した．96ウェルマイクロタイタープレートを使用して32検体ずつ解析を行う

表1　50万SNPsによるヒトナルコレプシーのゲノムワイド関連解析：関連SNP数の分布

有意水準	SNP数
$p < 0.0001$	214
$p < 0.001$	631
$p < 0.01$	3,665
$p < 0.05$	15,443
解析SNP総数	335,811

連分析も開始している．すでにタイピングおよび統計解析を終了した110名の患者試料と200名の健常者試料のデータから得られた関連SNPsの数の分布を表1に示す．call rate（＞90％）およびハーディ・ワインベルク平衡からのずれ（$p > 0.001$）に基づいて選別された約33万6千種のSNPsのうち，$p < 0.0001$の関連を示したSNPsが約200種検出された．なお，この解析には従来の遺伝子型判定ソフト（GTYPE4.0）を用いたデータを使用している．現在，新しい判定ソフト（GTYPE4.1）によってcall rateが向上していることから，より多くのSNPsについて統計解析することが可能になっている．ヒトのナルコレプシーについては，すでに確立した感受性遺伝子として6番染色体上の*HLA-DQB1*遺伝子が知られている．図3は今回の解析から*HLA*遺伝子領域について得られた結果であるが，予想通り*HLA-DQB1*遺伝子近傍をピークとする強い関連が認められた．現在われわれは，解析規模を2倍に拡大して新たなナルコレプシー感受性・抵抗性候補領域を検出している．

おわりに

数十万種以上のSNPを一挙にタイピングできる技術の実用化によって，従来は存在しなかった広範かつ詳細なヒトゲノム多型情報が得られる時代となった．このような情報は，疾患遺伝子探索研究のみならず，人類の進化や人類集団の歴史を解明する糸口を提供し，ヒトゲノム多様性に関連するさまざまな研究分野に画期的なインパクトを及ぼすことは疑いない．

しかしながら同時に，われわれはまだ得られる膨大な多型情報を十分に活用できるノウハウをもっていないことも指摘しておきたい．500K Array Setによる SNPタイピングで得られる1検体当たりのファイルデータのサイズは約2 Gbであるため，何百，何千検体のデータを保管し，必要な時に取り出して解析するためのコンピュータ環境を整備することは容易ではない．また，われわれの統計解析ソフトウェアはおのおののSNPについて関連分析できるものの，まだSNPハプロタイプについて関連分析することはできない．市販

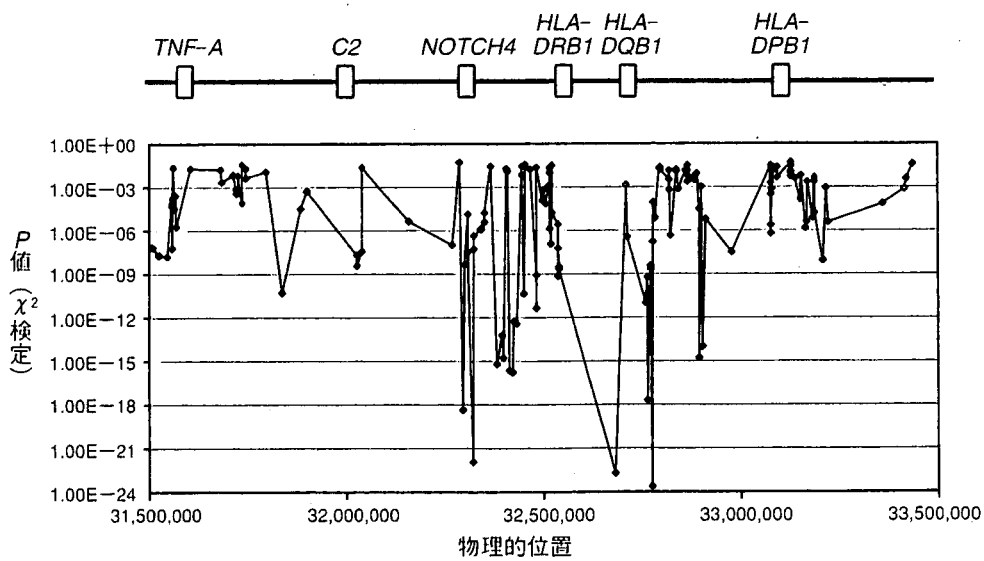

図3　ヒトナルコレプシーのゲノムワイド関連解析から
HLA-DQB1，*HLA-DRB1*遺伝子近傍領域におけるSNP関連マッピングによって強い疾患関連が検出された

のソフトウェアにも，限定した領域でハプロタイプ関連分析できるものはあるものの，ゲノム全域にわたって一挙に分析できるものはない．さらに，多数の検体について得られた50万SNPsデータから，これまで全く知られていなかった新しい遺伝子－遺伝子相互作用が見出される可能性がある．しかし残念ながら，従来の統計学的手法や計算アルゴリズムでは，このように膨大なデータを実用的に処理できない．このように，ゲノムワイド多型解析情報はバイオインフォマティクスに関わるさまざまな研究者にとって挑戦に値する多くの課題を提供してくれるとともに，その達成によって従来にない実り豊かな成果をもたらしてくれるに違いない．

文献

1）Ohashi, J. & Tokunag, K. : J. Hum. Genet., 46 : 478-482, 2001
2）徳永勝士：「人類遺伝学ノート」，南山堂（印刷中）
3）Kruglyak, L. & Nickerson, D. A. : Nat. Genet., 27 : 234-236, 2001
4）Syvänen, A. -C. : Nat. Rev. Genet., 2 : 930-942, 2001
5）Kwok, P. -Y. & Chen, X. : Curr. Issues Mol. Biol., 5 : 43-60, 2003
6）Nishida, N. et al. : Anal. Biochem., 346 : 281-288, 2005
7）Rachagani, S. et al. : BMC Genetics, 7 : 31, 2006
8）Matsuzaki, H. et al. : Genome Research, 14 : 414-425, 2004
9）Oliphant, A. et al. : BioTechniques, 32 : S56-S61, 2002
10）Grant, S. F. et al. : Nucleic Acids Res., 30 : e125, 2002
11）Jordan, B. et al. : Proc. Natl. Acad. Sci. USA, 99 : 2942-2947, 2002
12）Kawashima, M. et al. : Am. J. Hum. Genet., 79 : 252-263, 2006

＜著者プロフィール＞

西田奈央：東京大学大学院総合文化研究科で博士号を取得後，東京大学大学院医学系研究科人類遺伝学分野（徳永勝士教授）にて研究員として従事，研究課題は遺伝子多型解析手法の開発．

徳永勝士：東京大学理学部，同医学部附属病院，日本赤十字中央血液センターを経て，1995年より東京大学大学院医学系研究科教授．研究課題はヒトゲノム多様性および複合疾患の遺伝要因とその機能．

Available online at www.sciencedirect.com
ScienceDirect
Analytical Biochemistry 364 (2007) 78–85
ANALYTICAL BIOCHEMISTRY
www.elsevier.com/locate/yabio

Further development of multiplex single nucleotide polymorphism typing method, the DigiTag2 assay

Nao Nishida [a,*], Tetsuya Tanabe [b], Miwa Takasu [a], Akira Suyama [c], Katsushi Tokunaga [a]

[a] Department of Human Genetics, Graduate School of Medicine, University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan
[b] Bio Business Division, Olympus Corporation, Hachioji, Tokyo 192-8512, Japan
[c] Department of Life Sciences, Graduate School of Arts and Sciences, University of Tokyo, Meguro-ku, Tokyo 153-8902, Japan

Received 21 December 2006
Available online 13 February 2007

Abstract

A number of single nucleotide polymorphisms (SNPs) are considered to be candidate susceptibility or resistance genetic factors for multifactorial disease. Genome-wide searches for disease susceptibility regions followed by high-resolution mapping of primary genes require cost-effective and highly reliable technology. To accomplish successful and low-cost typing for candidate SNPs, new technologies must be developed. We previously reported a multiplex SNP typing method, designated the DigiTag assay, that has the potential to analyze nearly any SNP with high accuracy and reproducibility. However, the DigiTag assay requires multiple washing steps in manipulation and uses genotyping probes modified with biotin for each target SNP. Here we describe the next version of the assay, DigiTag2, which works with simple protocols and uses unmodified genotyping probes. We investigated the feasibility of the DigiTag2 assay by genotyping 96 target SNPs spanning a 610-kb region of human chromosome 5. The DigiTag2 assay is suitable for genotyping an intermediate number of SNPs (tens to hundreds of sites) with a high conversion rate (> 90%), high accuracy, and low cost.

Keywords: Multiplex genotyping; SNPs; Mutation; Oligonucleotide ligation assay

As a consequence of the Human Genome Project and single nucleotide polymorphism (SNP)[1] discovery projects, several million SNPs have been uploaded onto public SNP databases. It is estimated that there are 5 million SNPs with a greater than 10% minor allele frequency and 11 million SNPs with a greater than 1% minor allele frequency in the human genome [1]. Among these SNPs, many are candidate susceptibility or resistance genetic factors for multifactorial diseases and have been identified based on linkage analysis in families or association analysis with unrelated patients (cases) and healthy controls [2–6]. Large-scale case–control analyses using a dense set of SNP markers across the human genome have revealed associations between various diseases and SNPs with the highest detection power [7–9].

During recent years, genome-wide association studies using SNP markers have attempted to search for susceptibility and/or resistance genes by using emerging genome-wide SNP typing technologies such as Affymetrix GeneChip arrays and Illumina BeadArray genotyping technology [10–13]. These genome-wide SNP typing technologies would detect candidate regions, including susceptibility or resistance genes. However, to identify primary SNPs or genes, it is necessary to perform association analysis using an intermediate number of SNPs (tens to hundreds of sites) located within the candidate regions. Currently, there are a variety of SNP genotyping methods that are suitable for genotyping large numbers of samples for a modest number of SNPs such as 5′ exonuclease

* Corresponding author. Fax: +81 3 5802 8619.
E-mail address: nishida-75@umin.ac.jp (N. Nishida).

[1] *Abbreviations used:* SNP, single nucleotide polymorphism; MALDI–TOF MS, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry; ED, end digit; D1, first digit; PCR, polymerase chain reaction; dNTP, deoxynucleoside triphosphate; ATP, adenosine triphosphate; DTT, dithiothreitol; NAD, nicotinomide adenosine dinucleotide; EDTA, ethylenediaminetetraacetic acid; Cy3–ED-1, Cy3-labeled ED-1; Cy5–ED-2, Cy5-labeled ED-2; SDS, sodium dodecyl sulfate; DCN, DNA coded number.

0003-2697/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.ab.2007.02.005