

Genotyping of single nucleotide polymorphisms using the SNP-RFLP method

Saifullah, Toshifumi Tsukahara*

Area of Bioscience and Biotechnology, School of Materials Science, Japan Advanced Institute of Science and Technology (JAIST), Nomi, Ishikawa, Japan.

Summary

Genetic polymorphisms, including single nucleotide polymorphisms (SNPs), are responsible for inter-individual variability in susceptibility to cancer and other disorders. Both environmental factors (e.g., smoking or carcinogen exposure) and genetic variation underlie the development of cancer; however, studies of twins suggest that genetic variation is more important. Hence, the identification of SNPs makes an important contribution to cancer research. In this study, 13 SNPs in 12 genes were genotyped in HEK 293 and HeLa cells using the simple and inexpensive SNP-RFLP method. Sanger sequencing was performed for one SNP to validate the SNP-RFLP results. Of the 13 SNPs, 10 were homozygous and three were heterozygous (rs10937405, rs12296850, and rs3814113) in HEK 293 cells, while 12 were homozygous and one was heterozygous (rs995030) in HeLa cells. The cells carried eight disease-associated risk alleles (32% of typed alleles), including rs2853677, rs995030, rs2736100, and rs6010620 in HEK 293 cells, and rs10937405, rs3814113, rs4767364, and rs6010620 in HeLa cells. Four SNP loci were homozygous for different alleles in each cell line, with HEK 293 cells having a CC genotype at rs2853677, GG at rs2736100 and rs4767364, and TT at rs3819197, while HeLa cells had TT genotypes at rs2853677 and rs2736100, AA at rs4767364, and CC at rs3819197. In conclusion, these results are potentially applicable for testing of novel gene therapeutic approaches in future experiments where the non-risk alleles of the eight identified risk alleles are substituted into HEK 293 or HeLa cells.

Keywords: SNPs, carcinoma risk, RFLP, HEK 293 cells, HeLa cells

1. Introduction

Single nucleotide polymorphisms (SNPs), which are single base alterations at specific genome loci, are important, commonly bi-allelic, abundant (> 9 million), and the most frequent (1 in 1,000 bases) subtle genetic variation in the human genome (1). Depending on the location of a SNP within the genome, it can be associated with various phenomena at the level of gene expression. SNPs located in the 3'- or 5'-untranslated region (-UTR), or within introns, of genes can alter specific sequence motifs, including protein-binding sites, resulting in strong associations with various disorders, including

testicular neoplasms, glioma, and lung adenocarcinoma (2-4). SNPs in non-coding regions can also be useful as markers for population genetics and evolutionary studies (5) and may be associated with inherited susceptibility to cancer (6). Moreover, SNPs in the coding regions of genes have strong associations with other phenomena, such as alcohol consumption behavior (7), aero-digestive tract cancers (8), drug responses in pharmacogenomics (9), and inherited monogenic disorders (10), and are routinely analyzed for diagnostic purposes.

Single nucleotide polymorphism-restriction fragment length polymorphism (SNP-RFLP) is an enzymatic method of SNP genotyping in which a region containing a specific SNP allele is targeted and amplified by PCR and subsequently digested using an endonuclease enzyme. The results are interpreted based on length variation of the resulting digested PCR fragments, which are distinguished by gel electrophoresis. SNP-RFLP is a relatively simple, rapid, inexpensive, and convenient method for SNP genotyping, in contrast to more complex

*Address correspondence to:

Professor Toshifumi Tsukahara, Area of Bioscience and Biotechnology, School of Materials Science, Japan Advanced Institute of Science and Technology (JAIST), 1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan.
E-mail: tsukahara@jaist.ac.jp

SNP typing methods, such as sequencing, DNA chips, TaqMan assays, and hybridization, among others.

Cell lines are scientific tools that provide simplified models, allowing examination of the function of biological tissues under artificially regulated conditions, and without the ethical implications of experiments using whole organisms. Human embryonic kidney-derived epithelial cells (HEK 293) and the cervical cancer-derived epithelial cells (Henrietta Lacks; HeLa) are arguably the mammalian cell lines most extensively used in biological research, due to their rapid turnover time and ability to express heterologous proteins, including factors that undergo post-translational modifications (11,12). In this study, we characterized a number of disease-relevant SNPs in HEK 293 and HeLa genomic DNA using the simple and inexpensive SNP-RFLP method. The resulting information could assist other investigators performing SNP studies by providing control data for genome-wide association studies (GWAS) or laboratory diagnostic methods. In addition, the results will be of interest to other researchers for application in the development of new gene therapy, protein, or peptide-based drugs targeting these genetic variations.

2. Materials and Methods

2.1. Cell culture

HEK 293 cells and HeLa cells (RIKEN Cell Bank, Japan) were cultured in Dulbecco's modified Eagle's medium (DMEM, Nacalai Tesque, Inc., Japan) supplemented with 10% fetal bovine serum (Gibco, Life Technology, Canada). Cells were plated at a density of 2×10^6 per well in sterile 35×10 mm cell culture dishes (Falcon, USA) and cultured in an incubator at 37°C, 5% CO₂, and 95% humidity for 48-96 h, after which DNA samples were extracted.

2.2. Genomic DNA extraction

Extraction of genomic DNA from both HEK 293 and HeLa cell lines was performed using TRIzol reagent (Invitrogen, cat. no. 15596-018, USA), according to the manufacturer's instructions with minor modifications. Briefly, when cells were at 100% confluence, culture medium was removed and cells were rinsed with 1 mL of ice-cold PBS per 35×10 mm culture dish. Next, 1 mL of TRIzol reagent was added directly to the cells in the culture dish, followed by pipetting to lyse the cells, addition of 300 μ L of chloroform, and centrifugation at $12,000 \times g$ for 15 min at 4°C. The bottom layer after centrifugation contained protein, the interface included the DNA, and the supernatant contained RNA. Supernatants were removed, then 400 μ L of 100% ethanol was added to the interface and organic phase, and tubes were mixed by inversion several times,

followed by centrifugation at $2,000 \times g$ for 5 min at 4°C and the supernatant was removed (containing protein). Pellets were then washed twice with 1 mL of 0.1 M sodium citrate in 10% ethanol solution by incubation at room temperature for 30 min and centrifugation at $2,000 \times g$ for 5 min at 4°C. After discarding the wash solution, DNA pellets were precipitated with 1.8 mL of 75% ethanol, followed by incubation at room temperature for 10-20 min with periodic mixing and centrifugation at $2,000 \times g$ for 5 min at 2-8°C. Finally, the ethanol was removed, and DNA pellets were air dried for 5-10 min in an open tube and then dissolved in 100 μ L of 8 mM NaOH for long term storage at -20°C for use in future experiments. DNA concentrations were measured by UV spectroscopy (SmartSpec™ 3000, USA).

2.3. Primer design

One of the most important factors in optimizing SNP-RFLP assays is primer design, which is critical for successful genotyping. Primers should be designed so that PCR amplicons will be 250-500 bp in total, and each primer should be at different distances from the SNPs site, so that products of different length will be easily distinguishable after restriction digestion. In this study, we first input target 500-600 bp genomic DNA sequences, including the SNP alleles of interest, into the online NCBI primer BLAST software. Next, one pair of 18-35 bp nucleotide sequences were selected as primers to amplify a 250-500 bp genomic DNA target, with each primer a different distance from the SNP site; factors considered when designing primers included melting temperature (T_m), percentage GC content (GC%), and self-complementarity. PCR primer pairs were designed to amplify only the target genomic sequence with no cross-reaction with homologous non-target regions.

2.4. SNP-RFLP

Thirteen pairs of SNPs significantly associated with carcinoma development, alcoholism, or other conditions, according to previous publications, and that include restriction sites for commercially available enzymes, were selected from the NCBI SNP database (Table 1).

2.4.1. PCR amplification

PCR reactions were performed in a total volume of 20 μ L using a thermo cycler (GeneAmp PCR System 9700, Singapore). Reaction mixtures contained 8.3 μ L of autoclaved MilliQ water, 4 μ L of $5 \times$ Gold Taq buffer (Promega, USA), 2 μ L of 2 mM dNTPs, 2 μ L of 25 mM MgCl₂, 2 μ L of 91 ng/ μ L genomic DNA template, 0.8 μ L (10 pmol) of each primer, and 0.1 μ L of Taq polymerase (Promega, USA). PCR conditions were as

Table 1. SNPs used in this study

SNP no.	SNP site	Accession number	Gene	Chromosome: position	Molecular consequence [risk allele]	Phenotype; possible risks
1	CATACACT[A/G]AAGTGAAA	rs671	ALDH2	12:111803962	c.1510G>A (p.Glu504Lys); missense [A]	alcohol consumption behavior; cardiovascular diseases (7)
2	TATAATCT[C/T]GTGTTCAA	rs10937405	TP63	3:189665394	g.189665394C>T; intron variant [C]	lung adenocarcinoma (4)
3	GAGACCCG[C/T]CTGGTGGA	rs2853677	TERT	5:1287079	g.12969C>T; intron variant [C]	lung adenocarcinoma (4)
4	GGGCTTAC[A/G]GCTATTGA	rs12296850	SLC17A8	12:100426307	g.100820085A>G; downstream gene variant [G]	lung squamous cell carcinoma (15)
5	CAAATGGT[A/G]CATGCAGT	rs995030	KITLG	12:88496894	g.88496894A>G; 3'-UTR variant [G]	testicular neoplasms (2)
6	TCCATATC[C/T]TCTGGACC	rs3814113	BNC2	9:16915023	g.16915021T>C; intergenic variant [T]	serous ovarian cancer (16)
7	GTGTTCT[G/T]TAGCTTTG	rs2736100	TERT	5:1286401	g.129673211T>G; intron variant [G]	idiopathic pulmonary fibrosis (21)
8	AAAAGCCA[G/T]GGGAACCTA	rs4295627	CCDC26	8:129673211	g.129673211T>G; intron variant [G]	glioma (3)
9	GCCTTGG[C/T]CCAGGAGA	rs498872	PHLDB1	11:118606652	g.118606652C>T; 5'-UTR variant [T]	glioma (3)
10	TGACTCAA[C/T]TGTTTTCT	rs1494961	HELQ	4:83453327	g.83453327C>T; missense [C]	aero-digestive tract cancers (8)
11	TGGGTAC[A/G]TATACTCC	rs4767364	NA425	12:112083644	g.112083644G>A; intron variant [A]	aero-digestive tract cancers; type 1 diabetes (8)
12	TCCTTCCA[C/T]CTTTTCAT	rs3819197	ADH1A	4:99279352	g.99279352C>T; intron variant [none]	alcohol consumption behavior (17)
13	GGAGGAG[C/A/G]CCTGCTTT	rs6010620	RTEL1	20:63678486	g.63678486A>G; intron variant [G]	glioma; atopic dermatitis (3)

follows: pre-denaturation at 95°C for 3 min; followed by 23-28 cycles of 95°C for 30 s, 58-63°C for 30 s, and 72°C for 30 s; and a final extension at 72°C for 7 min. Polyacrylamide gel electrophoresis (PAGE) was used to analyze PCR amplicons.

2.4.2. RFLP

For rapid detection of SNP sites, PCR products were cleaved using specific restriction enzymes (NEB, Takara, and Thermo Scientific™). The restriction endonucleases for SNP-RFLP were chosen using two online tools, Watcut (13) and SnapGene Viewer software (14). RFLP reaction mixtures contained 6 µL of autoclaved MilliQ water, 10 µL of PCR product, 1 µL of bovine serum albumin, 2 µL of 10 × reaction buffer, and 1 µL of endonuclease enzyme. Reactions were carried out at 37°C for 2 h.

2.4.3. PAGE

PCR amplicons and products of restriction enzyme cleavage were analyzed by PAGE; gels consisted of autoclaved MilliQ water (6 mL), 40% acrylamide/bis mixed solution (1.2 mL), 10 × tris borate EDTA stock solution (0.8 mL), 10% ammonium persulfate (120 µL), and TEMED (8 µL). PCR amplicons and cleaved fragments were run alongside a 100 kb ladder marker at 200 V, 2.0 A, for 30 min. Gel staining was performed with SYBR Green dye (Lonza, Cat. 50513, USA) for 25 min. Gel images were obtained using a LAS-3000 Image Analyzer (Fujifilm, Japan).

2.4.4. Mechanism of SNP-RFLP

The SNP-RFLP method for genotyping used in this study is exemplified by SNP No. 10 (Accession No. rs1494961) where the target SNP (T/C), was designated position 0; comparatively, the forward primer was designed at -241 bp upstream and the reverse primer +100 bp downstream, to amplify a 341 bp genomic DNA region flanking the target SNP. Following PCR, amplicons were treated using the MnlI restriction endonuclease. Where the sequence 5'-CAATTG-3' was included among the amplicons, MnlI cleaved it to produce fragments of 100 and 241 bp. No other MnlI recognition sites were included in the 341 bp fragment, other than the target SNP within it. Product length variation was subsequently analyzed by PAGE.

2.5. Sanger sequencing

To validate the PCR-RFLP results, SNP No. 13 (Accession No. rs6010620) was also analyzed by Sanger sequencing using a 3130XL Genetic Analyzer (Applied Biosystems, Japan). The thermocycling reactions for direct sequencing comprised 10 µL, containing 5 µL of

Table 2. Primer sequences and optimum PCR conditions

SNP no.	Accession no.	Primers (5' to 3')	^a PCR conditions
1	rs671	¹ GATGTGTTTGGAGCCCAGTC ² AGCAGACCCCTAAATCCCTGG	63°C ^b , 23°C
2	rs10937405	¹ ACTTATGCTACGGTTCAGGGA ² CCATCCTCTGCATGTTAGGTTT	63°C ^b , 23°C
3	rs2853677	¹ TCTGGCCTAATTTCAACACTTTTA ² GATCAACACACACTCGGCAG	63°C ^b , 23°C
4	rs12296850	¹ TGCTGCCAGTCAGAATCATATC ² TAGAGGTGCTCGTTACAGGTG	58°C ^b , 28°C
5	rs995030	¹ GCTGACATGTTGCCAAATCC ² AACTTGCATGGAGCAGGACT	63°C ^b , 23°C
6	rs3814113	¹ ACCGTGTAGGATGGTCTCG ² CCAGTTGGAGGTGAGTTCGT	63°C ^b , 23°C
7	rs2736100	¹ TTGCTACCCTTGCCTGAGC ² ACGTTGCTGCTCACTACTGG	63°C ^b , 23°C
8	rs4295627	¹ TGGCTATTTTGAAGGAAAAGTG ² TCCTCACTTGGCAATCTGGT	63°C ^b , 23°C
9	rs498872	¹ AGATTGGGAGATGGAGGCAG ² AGGCTTTCTGCTCCTTCTT	63°C ^b , 23°C
10	rs1494961	¹ GGAATGGAAGGGGCTGAG ² GACTGGAATGCGAAGGC	63°C ^b , 23°C
11	rs4767364	¹ AGCTGGTGTGAATTGCAC ² CGGTTTCTAACTCAAAGACAAAAT	63°C ^b , 23°C
12	rs3819197	¹ TGGCTTCAGCACAAATAGGAA ² CCCAAAACCTGTGGCTGATT	63°C ^b , 23°C
13	rs6010620	¹ TCCTCGACCCACAGTGATCC ² AGTGTGTGCTGCTCCTCTCA	63°C ^b , 23°C

¹Forward primer; ²Reverse primer; ^aOptimal conditions, denaturation at 95°C for 30 s, extension at 72°C for 30 s, 91 ng/μL concentration genomic DNA template, 0.8 μL of 10 pmol/μL each primer; ^bPrimer annealing temperature; ^cCycles.

water, 1.5 μL of 5 × Big Dye buffer, 1 μL of 2.5 × Big Dye version 1.1, 3.1 (Applied Biosystems, USA), 0.5 μL of 3.2 pmol/μL reverse primer, and 2 μL of purified PCR fragment from agarose gel. The thermocycling conditions were as follows: pre-denaturation at 96°C for 1 min; followed by 25 cycles of 96°C for 10 s, 50°C for 5 s, and 72°C for 4 min; and then a hold step at 4°C for an indefinite period. Reaction products were ethanol precipitated and resuspended in HiDi formamide (Applied Biosystems, USA). Products were denatured at 95°C for 3 min followed by chilling on ice for 3 min. Then, samples were mounted on the Genetic Analyzer system to extract nucleotide sequence information by capillary electrophoresis. Finally, results were analyzed using Seq_Scanner software (Applied Biosystems, Japan).

3. Results

3.1. DNA extraction

Genomic DNA was isolated from HEK 293 and HeLa cell lines, and sample concentrations were 1,520 and 704 ng/μL, with purity ratios (absorbance at 260 nm divided by that at 280 nm) of 1.83 and 1.77, respectively.

3.2. Primer sequences

The primer sequences used in this genotyping study

and their optimum PCR conditions are presented in Table 2. These data could be useful to other researchers to facilitate convenient genotyping of these SNPs by enzymatic methods.

3.3. SNP-RFLP

Images of PAGE separated PCR and products of restriction digestion are shown in Figure 1. The results of genotyping of the 13 SNPs in 12 genes in genomic DNA from both cell lines are presented in Table 3. In HEK 293 cells, 10 SNPs were homozygous, while three (rs10937405, rs12296850, and rs3814113) were heterozygous; among the homozygous SNPs, eight comprised alleles cut by the specific endonucleases, whereas two were the uncut allele. In HeLa cells, 12 SNPs were homozygous and one (rs995030) was heterozygous; among the homozygous SNPs, five comprised alleles containing specific endonuclease sites and seven were the uncut alleles. Although both cell types had similar allele distributions, some distinct alleles were observed at a number of SNP positions; for example, HEK 293 cells had a CC genotype at rs2853677, GG at rs2736100 and rs4767364, and TT at rs3819197, whereas HeLa cells had TT genotypes at rs2853677 and rs2736100, AA at rs4767364, and CC at rs3819197. These variations may reflect the different origins of the cells; HEK 293 cells are an embryonic cell line originating from the Netherlands, while HeLa cells are mature cancerous cells derived from

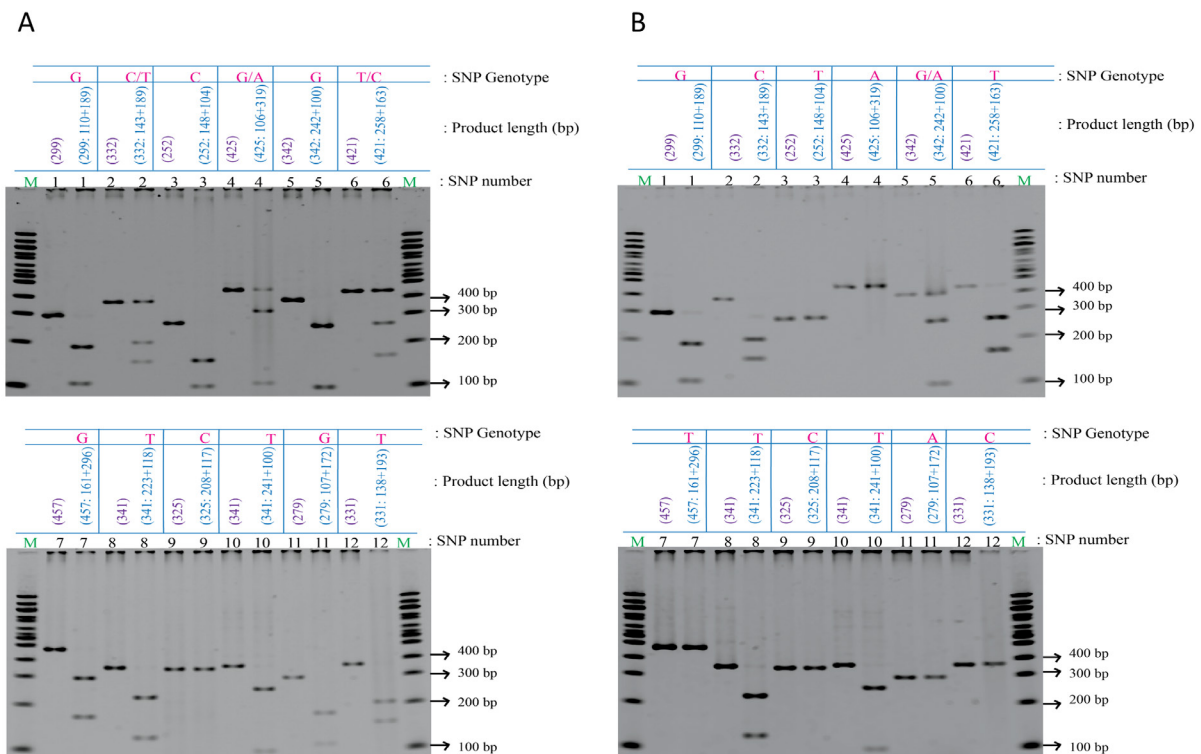


Figure 1. PAGE images for SNP genotyping of (A) HEK 293 cells and (B) HeLa cells by SNP-RFLP. Samples were run in pairs with PCR products to the left of restriction digest reactions; for example, 'SNP No. 1 (299)', PCR products for SNP No. 1; and 'SNP No. 1 (299:110 + 189)', restriction digestion reaction (RFLP) for the same SNP.

Table 3. Results of SNP-RFLP analysis for SNP genotyping in both cell lines

SNP no.	Accession no.	Gene	Endonuclease	Recognition site	Product (restriction fragment) length (bp)	HEK 293 genotype		HeLa genotype	
						Cut/uncut	Genotype	Cut/uncut	Genotype
1	rs671	<i>ALDH2</i>	AclI	CTGAAGN16 [^]	299 (110 + 189)	+	GG	+	GG
2	rs10937405	<i>TP63</i>	BssSalI	C [^] ACGAG	332 (143 + 189)	+/-	C/T	+	CC
3	rs2853677	<i>TERT</i>	MvaI	CC [^] WGG	252 (148 + 104)	+	CC	-	TT
4	rs12296850	<i>SLC17A8-NR1H4</i>	BceAI	ACGCGN12 [^]	425 (106 + 319)	+/-	G/A	-	AA
5	rs995030	<i>KITLG</i>	SphI	GCAATG [^] C	342 (242 + 100)	+	GG	+/-	G/A
6	rs3814113	<i>BNC2</i>	MboII	GAAGAN8 [^]	421 (258 + 163)	+/-	T/C	+	TT
7	rs2736100	<i>TERT</i>	BfmI	C [^] TRYAG	457 (161 + 296)	+	GG	-	TT
8	rs4295627	<i>CCDC26</i>	NcoI	C [^] CATGG	341 (223 + 118)	+	TT	+	TT
9	rs498872	<i>PHLDB1</i>	PfoI	T [^] CCNGGA	325 (208 + 117)	-	CC	-	CC
10	rs1494961	<i>HELQ</i>	MunI	C [^] AATTG	341 (241 + 100)	+	TT	+	TT
11	rs4767364	<i>NAA25</i>	Hpy166II	GTN [^] NAC	279 (107 + 172)	+	GG	-	AA
12	rs3819197	<i>ADH1A</i>	BceI	CCATCN4 [^]	331 (138 + 193)	+	TT	-	CC
13	rs6010620	<i>RTEL1-TNFRSF6B</i>	BspMI	ACCTGCN4 [^]	284 (242 + 42)	-	GG	-	GG

'+', cut; '-', uncut; 'W', A or T; 'R', A or G; 'Y', C or T.

an American woman. Identified SNP risk alleles are presented in Table 1 ("phenotype; possible risks").

3.4. Validation of SNP-RFLP by Sanger sequencing

Sanger sequencing of one marker (SNP No. 13; Accession No. rs6010620) was performed to validate the SNP-RFLP results. The sequencing results for both cell lines indicated that they carried the G rather than the A allele at this locus (Figure 2B-C), consistent with the results of SNP-RFLP analysis (rs6010620 in Table 3; Figure 2A). These data indicate that the PCR-RFLP

results were correct, supporting the conclusion that the generated SNP-RFLP results confirm the genotypes of 13 SNPs in 12 different genes. These results have the potential to assist other researchers and inform clinical diagnostics to identify SNPs associated with cancer risk using a convenient enzymatic method.

4. Discussion

In this study, we profiled 13 SNPs associated with disease risk in 12 genes in the HEK 293 and HeLa cell lines using the cost-effective SNP-RFLP method. Across

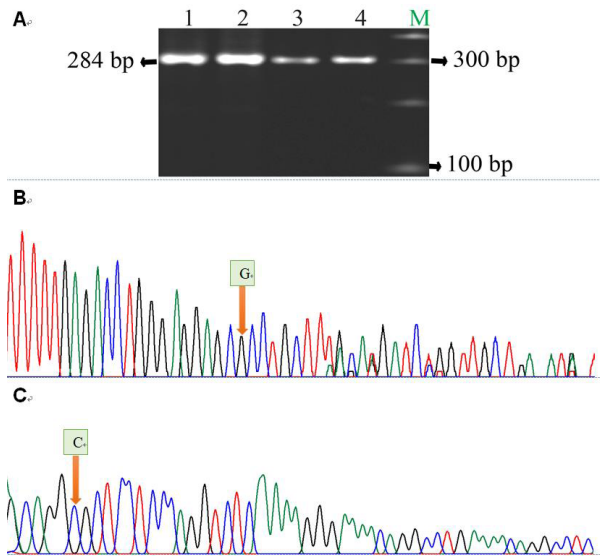


Figure 2. Sequencing validation results for SNP-RFLP of SNP No. 13 (Accession No. rs6010620). **A:** Image of PAGE of SNP No. 13 genotyping; 1: HEK 293 cells PCR product, 2: HEK 293 cells RFLP, uncut, G allele; 3: HeLa cells PCR product; 4: HeLa cells RFLP, uncut, G allele; M: marker. **B:** Sequencing electropherogram results for SNP rs6010620 in HEK 293 cells; G allele is indicated by an arrow. **C:** Sequencing electropherogram results for SNP rs6010620 in HeLa cells; G allele (reverse primer 5'-AGTGTGTGCTGCTCCTCTCA-3' was used) was also present.

the two cell lines, 22 alleles contributed to homozygous genotypes, and four were in the heterozygous state (Figure 3). Eight risk alleles associated with disease, comprising 32% of typed alleles, were identified at the following SNPs: rs2853677, rs995030, rs2736100, and rs6010620 in HEK 293 cells and rs10937405, rs3814113, rs4767364, and rs6010620 in HeLa cells. We identified heterozygosity at three SNPs, rs10937405, C/T; rs12296850, G/A; and rs3814113, T/C, in HEK 293 cells; and one in HeLa cells (rs995030, G/A).

Takeuchi *et al.*, 2011, genotyped the SNP, rs671, in the Japanese population and found a strong association with alcoholism and subsequently also with cardiovascular diseases, particularly for individuals with the AA genotype at this locus (7). Here, we found that both cell lines were GG homozygotes (wild-type) for rs671. Shiraishi *et al.*, 2012, and Li Xiaoting *et al.*, 2016, reported that lung adenocarcinoma is associated with rs10937405 in the *TP63* gene, and rs2853677 and rs2736100 in *TERT*, in Japanese and Chinese populations (4,6). One study identified rs12296850, in the *SLC17A8-NR1H4* gene, as significantly associated with the risk of lung squamous cell carcinoma at the genome-wide significance level (15). Rapley *et al.*, 2009, studied men with infertility and found an association with the genetic variant, rs995030, in *KITLG*, with testicular germ cell tumors in a UK population (2). Song *et al.*, 2009, and Shete *et al.*, 2009, genotyped rs3814113 for ovarian cancer, rs4295627, rs498872, and rs6010620 for glioma in populations from the UK, the USA, and France (3,16). MacKay *et*

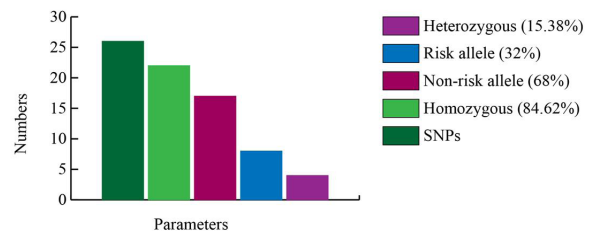


Figure 3. Allele distribution chart for the 13 SNPs typed in this study.

al., 2011, studied upper aero-digestive tract cancer and identified associations with rs1494961 and rs4767364, while Han *et al.*, 2007, reported an association of rs3819197 with alcoholism (8,17). These GWAS indicate that SNPs can be associated with susceptibility to certain diseases, even when the variant is in a non-coding, intronic, or 3'/5'-UTR region.

All of the SNPs mentioned above were characterized using BeadChip, TaqMan, or pyrosequencing methods. By contrast, we used the simple and cost-effective SNP-RFLP method to conduct SNP genotyping. SNP-RFLP is an accurate method for SNP genotyping, useful for small basic research studies of complex genetic variation, such as HLA allele typing or ABO blood group determinants (18). One disadvantage of this method is that target SNP nucleotide sequences may not always span commercially available endonuclease enzyme recognition sites. Another limitation is that sequences may contain too many palindromic recognition sites to allow a single endonuclease cleavage.

The findings presented here could be useful for the development of new gene therapy techniques, which could be tested directly in HEK 293 or HeLa cell lines as model mammalian expression systems. For example, Liu *et al.*, 2017, developed A to I editing (the conversion of A.T to G.C base pairs in DNA) at the genomic level by combining the dCas9 and TadA enzyme systems (19). We may apply this system in future to replace the pathogenic A allele with a G nucleotide (I) at the rs4767364 SNP, using HeLa cells as a test system. Furthermore, these data could be useful for the preparation of allele-specific probes for the development of probes-on-carrier DNA chips (20) for SNP genotyping.

In summary, in the present study 13 SNPs in 12 genes were genotyped in HEK 293 and HeLa cells using the SNP-RFLP technique. We detected eight risk alleles and four distinguish loci in both cell lines. HEK 293 cells were heterozygous at three loci, while there was a single heterozygous locus in HeLa cells. Further profiling of other carcinoma-related SNPs in HEK 293 and HeLa cells will be necessary to facilitate the study of the relationship between SNPs and cancer, to enable the development of gene therapy for treatment of those disorders. These findings will also assist mechanistic

studies of the functions of these SNPs in HEK 293 or HeLa cell models in the near future.

Acknowledgements

The authors thankfully acknowledge the scholarship from the Ministry of Education, Culture, Sports, Science, and Technology (MEXT), Japan.

References

- Kim S, Misra A. SNP genotyping: Technologies and biomedical applications. *Annu Rev Biomed Eng.* 2007; 9:289-320.
- Rapley EA, Turnbull C, Al Olama AA, *et al.* A genome-wide association study of testicular germ cell tumor. *Nat Genet.* 2009; 41:807-810.
- Shete S, Hosking FJ, Robertson LB, *et al.* Genome-wide association study identifies five susceptibility loci for glioma. *Nat Genet.* 2009; 41:899-904.
- Shiraishi K, Kunitoh H, Daigo Y, *et al.* A genome-wide association study identifies two new susceptibility loci for lung adenocarcinoma in the Japanese population. *Nat Genet.* 2012; 44:900-904.
- Jorde LB, Watkins WS, Bamshad MJ, Dixon ME, Ricker CE, Seielstad MT, Batzer MA. The distribution of human genetic diversity: A comparison of mitochondrial, autosomal, and Y-chromosome data. *Am J Hum Genet.* 2000; 66:979-988.
- Li X, Xu X, Fang J, Wang L, Mu Y, Zhang P, Yao Z, Ma Z, Liu Z. Rs2853677 modulates Snail1 binding to the TERT enhancer and affects lung adenocarcinoma susceptibility. *Oncotarget.* 2016; 7:37825-37838.
- Takeuchi F, Isono M, Nabika T, Katsuya T, Sugiyama T, Yamaguchi S, Kobayashi S, Ogihara T, Yamori Y, Fujioka A, Kato N. Confirmation of ALDH2 as a Major locus of drinking behavior and of its variants regulating multiple metabolic phenotypes in a Japanese population. *Circ J.* 2011; 75:911-918.
- McKay JD, Truong T, Gaborieau V, *et al.* A genome-wide association study of upper aerodigestive tract cancers conducted within the INHANCE consortium. *PLoS Genet.* 2011; 7:e1001333.
- Jittikoon J, Mahasirimongkol S, Charoenyingwattana A, Chaikledkaew U, Tragulpiankit P, Mangmool S, Inunchot W, Somboonyosdes C, Wichukchinda N, Sawanpanyalert P, He Y, McLeod HL, Chantratita W. Comparison of genetic variation in drug ADME-related genes in Thais with Caucasian, African and Asian HapMap populations. *J Hum Genet.* 2015; 61:119-127.
- Syvänen A-C. Accessing genetic variation: Genotyping single nucleotide polymorphisms. *Nat Rev Genet.* 2001; 2:930-942.
- Thomas P, Smart TG. HEK293 cell line: A vehicle for the expression of recombinant proteins. *J Pharmacol Toxicol Methods.* 2005; 51:187-200.
- Moorhead J. Henrietta Lacks: The mother of modern medicine. *The Guardian* 2010. <https://www.theguardian.com/science/2010/jun/23/henrietta-lacks-cells-medical-advances> (accessed May 2, 2018).
- Palmer M. WutCut: A web application for restriction analysis and related tasks. University of Waterloo. http://wutcut.uwaterloo.ca/template.php?act=snp_new (accessed May 2, 2018).
- SnapGene® software (from GSL Biotech; available at [snapgene.com](http://www.snapgene.com/products/snapgene_viewer/)). http://www.snapgene.com/products/snapgene_viewer/ (accessed May 2, 2018).
- Dong J, Jin G, Wu C, *et al.* Genome-wide association study identifies a novel susceptibility locus at 12q23.1 for lung squamous cell carcinoma in Han Chinese. *PLoS Genet.* 2013; 9:e1003190.
- Song H, Ramus SJ, Tyrer J, *et al.* A genome-wide association study identifies a new ovarian cancer susceptibility locus on 9p22.2. *Nat Genet.* 2009; 41:996-1000.
- Han Y, Gu S, Oota H, Osier MV, Pakstis AJ, Speed WC, Kidd JR, Kidd KK. Evidence of positive selection on a class I ADH locus. *Am J Hum Genet.* 2007; 80:441-456.
- Ota M, Fukushima H, Kulski JK, Inoko H. Single nucleotide polymorphism detection by polymerase chain reaction-restriction fragment length polymorphism. *Nat Protoc.* 2007; 2:2857-2864.
- Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, Bryson DI, Liu DR. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature.* 2017; 551:464-471.
- Tsukahara T, Nagasawa H. Probe-on-carriers for oligonucleotide microarrays (DNA chips). *Sci Technol Adv Mater.* 2004; 5:359-362.
- Mushiroda T, Wattanapokayakit S, Takahashi A, Nukiwa T, Kudoh S, Ogura T, Taniguchi H, Kubo M, Kamatani N, Nakamura Y. A genome-wide association study identifies an association of a common variant in TERT with susceptibility to idiopathic pulmonary fibrosis. *J Med Genet.* 2008; 45:654-656.

(Received May 15, 2018; Revised June 26, 2018; Accepted June 27, 2018)