Single Nucleotide Polymorphism Profiling Assay to Confirm the Identity of Human Tissues

Ronald Huijsmans,* Jan Damen,[†]
Hans van der Linden,[†] and Mirjam Hermans*

From the Multidisciplinary Laboratory of Molecular Diagnostics*
and the Laboratory of Pathology,[†] Jeroen Bosch Hospital,
's-Hertogenbosch, The Netherlands

To identify issues of sample mix-ups, various molecular techniques are currently used. These techniques, however, are time consuming and require experience and/or DNA sequencing equipment or have a relatively high risk of errors because of contamination. Therefore, a quick and straightforward single nucleotide polymorphism (SNP) profiling assay was developed to link human tissues to a source. SNPs are common sequence variations in the human genome, and each individual has a unique combination of these nucleotide variations. Using potentially mislabeled paraffin-embedded tissues, DNA was extracted and SNP profiles were determined by real-time polymerase chain reaction analysis of the purified DNA using a selection of 10 commercially available SNP amplification assays. These profiles were compared with profiles of the supposed owners. All issues (34 in total) of potential sample mix-ups during the last 3 years were adequately solved, with six cases described here. The SNP profiling assay provides a quick (within 24 hours), easy, and reliable way to link human samples to a source, without polymerase chain reaction postprocessing. The chance for two randomly chosen individuals to have an identical profile is 1 in 18,000. Solving potential sample mix-ups will secure downstream evaluations and critical decisions concerning the patients involved. (*J Mol Diagn* 2007, 9:205–213; DOI: 10.2353/jmoldx.2007.060059)

The clinically significant diagnostic error rate in surgical pathology reported in the literature varies from 0.08 to 1.2%.^{1–6} One of the factors contributing to these errors is misidentification of paraffin-embedded tissue, the most commonly used material for diagnostic purposes. Although specimen identification is a carefully controlled factor in clinical laboratories, mislabeling still occurs. Unfortunately, sample mix-ups can have extremely serious consequences for the patients involved.

Various methods to detect specimen mix-ups have been described. Most methods rely on characterization of DNA repeats,^{7–9} human leukocyte antigen system

(HLA),^{10–15} or other polymorphic genetic loci.¹⁴ In addition, commercial identification kits are available.^{16–18} Because these methods are time consuming and/or require DNA sequencing equipment, which is not available in many laboratories, we developed an easy, quick, and reliable method to link human tissues to a source. The method relies on the analysis of carefully selected SNPs by means of real-time polymerase chain reaction (PCR). Single nucleotide polymorphisms (SNPs) are the most frequent sequence variations in the human genome, occurring approximately once every 100 to 300 bp.¹⁹ Apart from identical twins, each individual has a unique combination of nucleotides at these positions. Thus, a SNP profile provides a kind of fingerprint. SNP amplification assays have been commercially developed using DNA probes with conjugated minor groove binding (MGB) groups.²⁰ These probes form extremely stable duplexes with single-stranded DNA targets, thus allowing short probes to be used for hybridization-based assays.

In this article, we describe a SNP profiling test for human tissues using 10 SNP assays that were selected on the basis of allele frequency and chromosomal location. By comparison, of the SNP profiles obtained from relevant samples, the test can reliably confirm the identity of human materials. The test can be used to detect sample mix-ups and to evaluate quality surveillance of sample handling and storage.

Materials and Methods

Tissue Processing

Tissues were fixed in 0.01 mol/L buffered (0.005 mol/L disodium hydrogen phosphate anhydrous and 0.005 mol/L sodium dihydrogen phosphate dihydrate, pH 7.0) 10% formalin, and processed for paraffin embedding using a Tissue-Tek VIP 5 (Sakura, Torrance, CA). The program consisted of 14 steps of 1 hour under continuous agitation, pressure, vacuum, and heating. At 40°C, two 10% formalin steps were followed by one 70% (v/v) ethanol step, two 96% ethanol steps, three 100% ethanol steps, and two 100% xylene steps. Paraffin embedding was done at 60°C in four 100% paraffin steps.

Accepted for publication November 15, 2006.

Address reprint requests to M.H.A. Hermans, Multidisciplinary Laboratory of Molecular Diagnostics, Jeroen Bosch Hospital, P.O. Box 90153, 5200 ME 's-Hertogenbosch, The Netherlands. E-mail: m.hermans@jzbz.nl.

Table 1. Assays Selected from Applied Biosystems (ABI)

SNP	SNP ID	Predeveloped assay reagent ID	Chromosome position	LOH hits	SNP type (gene)	Genotype Vic/Fam	Observed MAF	HapMap MAF	ABI MAF
<u>1</u>	<u>rs2283839</u>	C__2474321_1_	22q12	21	Intron (TPST2)	A/C	0.48(<i>n</i> =122)	0.43	0.5
2	rs1860300	C__11445512_10	17p13	218	Intergenic/unknown	A/C	0.40(<i>n</i> = 78)	Unknown	0.5
<u>3</u>	<u>rs2400077</u>	C__15792496_10	5p15.2	14	Intron (CTNND2)	A/C	0.44(<i>n</i> =124)	0.45	0.5
4	rs663528	C__1098888_10	13q12	59	Intergenic/unknown	G/T	0.38(<i>n</i> = 72)	0.46	0.5
<u>5</u>	<u>rs2239508</u>	C__1315434_1_	18p11	10	Intron (RAB31)	A/C	0.45(<i>n</i> =112)	0.45	0.5
6	rs2658509	C__1523697_10	4p15.3	6	Intron (LDB2)	A/C	0.49(<i>n</i> = 76)	0.42	0.49
7	rs1610180	C__1967983_10	3p25	81	Intron (SLC6A11)	C/A	0.45(<i>n</i> =104)	Unknown	0.5
<u>8</u>	<u>rs1202295</u>	C__2511899_10	6p22	7	Intergenic/unknown	A/C	0.43(<i>n</i> =114)	Unknown	0.5
9	rs1150964	C__3142524_10	12p11.2	2	Intergenic/unknown	G/T	0.44(<i>n</i> = 90)	Unknown	0.49
10	<u>rs2237958</u>	C__2574166_1_	11p15	116	Intron (USH1C)	A/C	0.40(<i>n</i> =100)	0.45	0.49
<u>11</u>	<u>rs2807845</u>	C__1235820_10	1q41	1	FLJ22390/intergenic	G/T	0.46(<i>n</i> = 46)	0.4	0.48
12	rs2305749	C__16191595_10	19q13.1	24	Intron (HPN)	G/T	0.27(<i>n</i> = 30)	Unknown	0.49
<u>13</u>	<u>rs1255912</u>	C__11462447_10	14q23	5	Intron (SYNE2)	A/C	0.45(<i>n</i> = 20)	0.49	0.47
<u>14</u>	<u>rs690054</u>	C__1616502_10	15q21	5	Intron (WDR72)	C/T	0.35(<i>n</i> = 20)	0.5	0.5
15	rs816852	C__3240268_10	10q22	14	Intron (KCNMA1)	A/G	0.40(<i>n</i> = 20)	0.5	0.5
<u>16</u>	<u>rs12996989</u>	C__1174290_10	2q21	6	Intron (LRP1B)	C/G	0.40(<i>n</i> = 20)	0.5	0.5
<u>17</u>	<u>rs656255</u>	C__2952210_10	12q24	2	Intron (KSR2)	A/C	0.40(<i>n</i> = 20)	0.5	0.5

Underlined SNP assays are recommended. Other SNP assays are excluded because of LOH (SNP2, SNP4, SNP7, and SNP10), allele frequencies significantly different from 0.5 (SNP4, SNP10, and SNP12), or technical limitations (SNP6 and SNP9). MAF, minor allele frequency.

DNA Isolation

Blood

DNA was extracted from 200 μ l of blood with the QIAamp DNA blood mini kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions and eluted in 150 μ l of buffer AE (Qiagen).

Paraffin Sections

Paraffin-embedded tissues were trimmed of paraffin excess and cut into 3- μ m-thick sections. As a control, a paraffin block without tissue, which had undergone the same treatment as the tissue containing blocks, was sectioned. Digestion solution was made by adding 100 μ l of proteinase K (20 mg/ml; Roche Diagnostics GmbH, Mannheim, Germany) and 10 μ l of Tween 20 (Merck BV, Amsterdam, The Netherlands) to 2 ml of TE buffer (1 mmol/L ethylenediaminetetraacetic acid, and 10 mmol/L Tris-HCl buffer, pH 8.0). Approximately 1 to 1.5 cm² of sectioned tissue (a single section or short ribbons depending on the surface per section) was put in 250 μ l of digestion solution and incubated overnight at 45°C. Proteinase K was inactivated by incubation at 100°C for 15 minutes. Samples were centrifuged for 2 minutes at 14,000 rpm. DNA was extracted from the supernatants with QIAamp DNA blood mini kit (Qiagen), according to the manufacturer's instructions (Qiagen protease digestion was not performed), and eluted in 150 μ l of buffer AE.

TaqMan Assays-on-Demand SNP Genotyping Products

TaqMan Assays-on-Demand SNP genotyping products were purchased from Applied Biosystems (Foster City, CA). These assays are predesigned and validated by analyzing DNA samples from 180 individuals of four ethnic origins. The assays are supplied in single tube format

and contain two unlabeled primers, a VIC dye-labeled TaqMan MGB probe with the sequence of allele 1 and a FAM dye-labeled TaqMan MGB probe with the sequence of allele 2.

Real-Time PCR

SNP amplification assays were used according to the manufacturer's instructions. Twenty-five μ l of PCR contained 20 mmol/L Tris-HCl, pH 8.4, 50 mmol/L KCl, 3 mmol/L MgCl₂ (prepared from 10 $\times$ PCR buffer and 50 mmol/L MgCl₂ solution delivered with Platinum Taq polymerase), 0.75 U of Platinum Taq polymerase (Invitrogen BV, Breda, The Netherlands), 4% glycerol (molecular biology grade; Calbiochem, VWR International BV, Amsterdam, The Netherlands), 200 μ mol/L of each dNTP (Invitrogen BV), 0.5 μ l of Rox reference dye (Invitrogen BV), 1.25 μ l of predeveloped assay reagent from the Assays-on-Demand SNP genotyping products (Applied Biosystems) containing two primers and two MGB TaqMan probes (5' VIC for allele 1, 5' FAM for allele 2, a 3' black hole quencher), and 11.25 μ l of target DNA. Real-time PCR was performed in an ABI Prism 7000 SDS (Applied Biosystems) for 2 minutes at 50°C, 10 minutes at 95°C, followed by 40 cycles of 15 seconds at 95°C and 1 minute at 60°C.

Results

Selection of SNP Assays

From an internet database containing more than 150,000 human TaqMan Assays-on-Demand SNP genotyping products (Applied Biosystems), we selected 17 SNP assays (Table 1). The principle of the assays is explained in Figure 1. An important criterion of selection was that minor (and major) allele frequency of the SNP should be ~0.5. TaqMan Assays-on-Demand SNP genotyping

A

Figure 1. SNP detection with MGB probes. **A:** Principle of SNP detection using hydrolysis probes. When the probe is intact, the quencher quenches the fluorescent signal from the reporter. A/C SNP: in the presence of A, the VIC probe, complementary to A, is hydrolyzed during the PCR by the Taq polymerase. The VIC reporter is no longer quenched by the close proximity of the quencher and can be measured. In the presence of C, the VIC probe is removed from the DNA strand without being hydrolyzed. The FAM probe is complementary to C and, according to the same principle, generates a fluorescent signal only in the presence of C. **B:** Fluorescent signals accompanying the different genotypes (SNP10; genotype is indicated in top left corner).

B

products (Applied Biosystems) are validated by analyzing DNA samples from 180 individuals of four ethnic origins. Because the Dutch population is mainly of Caucasian origin, we used the Applied Biosystems indicated allele frequencies for Caucasians for the selection of SNP assays. Assuming the genotype frequencies to be in Hardy-Weinberg equilibrium, the chance to have identical outcome of two events (allele 1 versus allele 2) is $p^2 + (1 - p)^2$.^{21,22} To minimize this chance, the derivative of this equation equals 0. Thus, $2p - 2 + 2p = 0$ and $p = 0.5$. Another criterion of selection was that all SNPs should be situated on different chromosomes (c.q. autosomes). In addition, the SNPs should not reside within a coding region or the regulatory sequence of a gene. From the 17 SNP assays, the most suitable were selected for a panel of 10 SNP assays that we recommend to be used for the SNP profiling assay (Table 1). Beside the SNP assays, for identity confirmation we included the detection of one sequence specific to the human Y chromosome²³ to resolve male/female sample mix-ups.

SNP Profiling

Because blood is an easy source of high-quality DNA, we tested the assays by analyzing four blood samples from four different patients. Figure 2 shows the fluorescent signals of three representative SNP assays; cycle thresholds (Ct values) were ~ 23 . The presence or absence of

alleles 1 and 2 could easily be deduced from the signals (Table 2). All four individuals could easily be distinguished based on their SNP profiles.

Validation of the SNP Assays on Paraffin Tissues

To validate the SNP assays on paraffin materials, 35 paraffin-embedded tissues were analyzed in a total of

Figure 2. Fluorescent signals (light gray line, VIC; black line, FAM) of three SNP assays performed on DNA isolated from blood of four patients.

Table 2. SNP Profiles: Presence (+) or Absence (–) of Allele 1 (V) and Allele 2 (F)

Case	Pathology number	Tissue	SNP 1		SNP 3		SNP 5		SNP 8		SNP 11		SNP 13		SNP 14		SNP 15		SNP 16		SNP 17	
			V	F	V	F	V	F	V	F	V	F	V	F	V	F	V	F	V	F	V	F
1	1-A-I	Blood 1	–	+	+	+	+	–	+	+	+	+	+	–	+	+	+	+	–	+	+	–
		Blood 2	–	+	+	–	–	+	+	–	+	+	–	+	+	–	+	–	+	+	+	–
		Blood 3	–	+	+	+	+	–	+	+	+	+	–	+	+	–	+	+	+	+	+	+
		Blood 4	+	–	–	+	–	+	+	+	+	–	+	+	–	+	+	–	+	+	+	–
	1-A-II	Mammary tissue	–	+	+	–	–	+	+	+	–	+	+	–	+	–	+	+	–	+	–	–
		Mammary tissue	–	+	+	+	+	+	+	+	+	+	–	+	+	+	+	–	+	+	+	–
	1-B	Cervical polyp	–	+	+	–	–	+	+	+	–	+	+	–	+	–	+	+	–	+	–	–
	1-C	Infiltrating carcinoma	–	+	+	–	–	+	+	+	–	+	+	–	+	–	+	+	–	+	–	–
	1-D	Mammary tissue	–	+	+	+	+	+	+	+	+	+	–	+	+	+	+	–	+	+	+	–
	1-E	Mammary tissue	–	+	+	+	+	+	+	+	+	+	–	+	+	+	+	–	+	+	+	–
2	2-A	floaters	+	+	–	+	+	–	+	–	+	–	+	+	+	+	+	+	+	–	+	+
3	2-B	Mediastinoscopy	+	+	–	+	+	–	+	–	–	+	+	+	+	+	+	+	+	–	+	+
	3-A	Skin	–	+	+	+	+	+	+	–	–	+	–	–	+	+	–	+	–	+	+	+
	3-B	Skin	–	+	+	+	+	–	+	+	+	+	+	–	+	+	+	+	+	+	+	+
	3-C	Skin	+	–	+	+	–	+	+	–	+	+	+	+	+	–	+	–	–	+	+	+
	3-D	Skin	+	+	–	+	+	+	+	–	+	+	+	–	–	+	+	+	+	+	+	–
	3-E	Vocal cord	–	+	+	+	+	–	+	+	+	+	+	–	+	+	+	+	+	+	+	+
	3-F	Skin	+	+	–	+	+	+	+	–	+	+	+	–	–	+	+	+	+	+	+	–
4	3-G	Curettage	+	–	+	+	–	+	+	–	+	+	+	+	+	–	+	–	–	+	+	+
	4-A	Skin	+	–	–	+	+	+	–	+	+	+	+	–	+	–	+	–	+	–	+	+
5	4-B	Skin	+	–	–	+	+	+	–	+	+	+	+	–	+	–	+	–	+	–	+	+
	5-1-1	Prostate	–	+	+	+	+	–	+	+	+	+	–	+	+	+	–	+	+	+	+	–
	5-2-1	Prostate	+	–	+	+	+	+	+	–	+	+	–	+	+	–	+	+	+	+	–	+
	5-A	Prostate	–	+	+	+	+	–	+	+	+	+	–	+	+	+	–	+	+	+	+	–
	15 specimens	Prostate	–	+			+	–														
	14 specimens	Prostate	+	–			+	+														
	5-1-12	Prostate	?	?			?	?														
6	5-1-13	Prostate	+	–			?	?														
	5-1-16	Prostate	–	+			?	?														
	6-A	Mammary biopsy	+	–	+	–	–	–	+	+	+	+	–	+	+	+	+	+	+	–	+	+
	6-B	Mammary tissue	+	–	+	–	–	–	+	+	+	+	–	+	+	+	+	+	+	–	+	+
	6-C	Mammary tissue	+	–	+	–	–	–	+	+	+	+	–	+	+	+	+	+	+	–	+	+

?, result not conclusive.

266 SNP assays. For the 35 tissues, the mean Ct value for fluorescent signals that were considered positive was 27.8 ± 4.2 (mean $\pm$ SD). A Ct value of 27.8 is approximately comparable with 10^4 cells in one PCR. During the processing of paraffin embedding, sectioning, and further analysis by real-time PCR, small traces of foreign DNA may contaminate the tissue under investigation. We therefore included a paraffin block in which no tissue was embedded in all analysis as a negative control. This paraffin block was processed identically to tissue containing blocks. The negative controls generated Ct values of 38.6 FAM and 41.9 (VIC) in 1 of 122 SNP assays. Based on the above findings, we defined the cutoff level for reliable Ct values at 33. Thus, Ct values higher than 33 were rejected, whereas Ct values lower than 33 were accepted. If Ct values exceeded 33, a larger surface of new sections was cut and analyzed.

Allele Frequencies of the SNP Assays

To test statistically whether the observed allele frequencies in the SNP assays differed from the frequencies indicated by Applied Biosystems for the Caucasian population, we used the χ^2 test. For each SNP, materials from a minimum of 10 individuals (maximum 62 individuals) were analyzed. With a probability of exceeding the critical value of 0.05, the observed frequencies for SNP4, 10,

and 12 differed significantly from the expected frequencies (data not shown). We therefore excluded these SNP assays from the recommended panel of SNP assays (Table 1).

SNP Assays and Loss of Heterozygosity (LOH)

LOH may lead to allelic loss and may therefore complicate the analyses of tumor tissues. During the development of the SNP profiling assay, we encountered fluorescent signals indicative of LOH in one lung tumor tissue (case 2, described below). We therefore investigated *in silico* whether the SNP assays were located in regions implicated in LOH by using the chromosomal region together with LOH as search terms in PubMed (Table 1). 3p25, 11p15, 13q12, and 17p13 yielded between 59 and 218 hits, rendering them less suitable for analysis of tumor tissues. The number of hits for the 10 recommended SNPs varied from 1 to 21.

SNP Profiling to Resolve Potential Sample Mix-Ups

Within the last 3 years, we processed 182,695 specimens. We had 34 potential sample mix-ups in which 111

patient materials were involved. Thus, the potential error percentage was 0.06%. To solve these issues, a total of 29 archived tissue blocks were analyzed. Complete SNP profiles (10 SNPs) were analyzed for 21 tissue-pairs. In eight of those cases, the SNP profiles showed 100% match, and common origin of the tissues was concluded. In the discordant SNP profiles of the other 13 mix-ups, the number of differing SNPs between the two tissues was 3 ($n = 3$), 4 ($n = 1$), 5 ($n = 3$), or ≥ 6 ($n = 6$). All issues of possible mix-up were adequately solved by SNP profiling. The elucidation of six representative cases of sample mix-ups is described below.

Case 1. Mislabeling of Two Mammary Tissues

Two mammary needle biopsies were taken from patients 1 and 2, respectively, and processed on the same day. Both tissue blocks were numbered 1-A. Histological analyses revealed that one of them, assigned 1-A-I, showed the presence of an infiltrative ductal carcinoma (Figure 3A), whereas the other, assigned 1-A-II, was benign (Figure 3B). The archive of paraffin-embedded specimens contained a cervical polyp biopsy (1-B) derived from patient 1. Figure 4 and Table 2 show that 1-A-I and 1-B have an identical SNP profile that differed from 1-A-II. Thus 1-A-I was assigned to patient 1 and 1-A-II was assigned to patient 2. Sixteen days after the needle biopsies were taken, patient 1 underwent a mastectomy (1C). A tumor of 28 mm was found and diagnosed as an infiltrative adenocarcinoma (Figure 3C). The SNP profile of this tumor (1C) was identical to the profile of 1-A-I and 1-B of patient 1 (Table 2). One and a half years after the needle biopsies, two other histological needle biopsies were taken from patient 2. Both were diagnosed as benign lesions. The SNP profiles of both tissues were identical to 1-A-II's profile (Table 2, 1-D and 1-E).

Case 2. Floater in a Lymph Node

The pathologist noticed a floater (2-A) of lung tumor tissue in a lymph node biopsy obtained by mediastinoscopy (2-B) from patient 3. Microscopically the floater was isolated from the surrounding nodal tissue. The question arose whether the floater and the lymph node were from the same individual. SNP analysis revealed a similar SNP profile of the two tissues (Table 2). Theoretically, the chance to have such a profile is lower than 1 in 30,000 assuming $P = 0.5$ for the analyzed SNPs: $1/(0.255 \times 0.55)$. In line with this, a keratin PAN immune peroxidase staining of another gland from patient 3 showed the presence of cells resembling the tumor cells in specimen 2-B (data not shown). The VIC signal of SNP14, although clearly present, was lower in the floater compared with the lymph node (Figure 5). SNP14 is located on 15q21, a region for which, to date, no LOH in lung tumor tissue has been reported.

Case 3. Mix-Up of Four Skin Biopsies

Because of a technical failure of the dissection table, four skin biopsies (3-A, 3-B, 3-C, and 3-D) derived from

Figure 3. Paraffin-embedded mammary tissue. **A:** Tissue assigned 1-A-I, an infiltrative ductal carcinoma. **B:** Tissue assigned 1-A-II, a benign, apocrine metaplasia. **C:** Infiltrative adenocarcinoma. H&E stain. Original magnifications, $\times 100$.

four patients (4, 5, 6, and 7) lost their identification. For three of these patients, archived material was available: a vocal cord of patient 4 (3-E), a skin biopsy of patient 5 (3-F), and curettage of patient 6 (3-G). Table 2 shows that three of the four SNP profiles from the unlabeled skin biopsies matched SNP profiles from the archived tissues.

Figure 4. Fluorescent signals (light gray line, VIC; black line, FAM) of three SNP assays performed on DNA isolated from two mammary tissues (1-A-I and 1-A-II) from patients 1 and 2 and a cervical polyp from patient 1 (1-B).

Thus, similar SNP profiles were observed for 3-B and 3-E (patient 4), 3-C and 3-G (patient 6), and 3-D and 3-F (patient 5). The remaining skin biopsy (3-A) was assigned to patient 7.

Case 4. Postmortem Identification

Dutch parents received information that their son had died in a foreign country. A high degree of degradation complicated the identification of the corpse. The parents asked the Jeroen Bosch Hospital to analyze whether a skin biopsy taken postmortem could be their son's. The

Figure 5. Diminished fluorescent signal of SNP14 probably attributable to LOH. Fluorescent signal (light gray line, VIC; black line, FAM) of the SNP assays performed on DNA isolated from the lymph node biopsy (2-B), the floater of lung tumor tissue (2-A), and a skin sample from another patient (for SNP14 homozygous, genotype T/T). The near absence of the VIC signal for SNP14 in the homozygous individuals demonstrates that the VIC signals in the floater are not because of cross-reaction of the VIC probe with the other allele. Thus, the diminished VIC signal for SNP14 in the floater is likely attributable to LOH. The fluorescent signals of SNP13 show no indication of LOH.

SNP profile of the postmortem skin biopsy (4-A) was identical to the SNP profile of an archived sample (4-B) of the son. Theoretically, the chance for the observed SNP profile was ~ 1 in 130,000. Another laboratory in The Netherlands later analyzed 10 tetranucleotide STR loci and Amelogenin using the AmpFLSTR SGM Plus PCR amplification kit (Applied Biosystems) of muscle and bone tissue of the deceased and mouth epithelial cells of the supposed mother. The DNA profiles very strongly suggested a mother-child relationship between the two individuals involved.

Case 5. Mislabeling of Prostate Gland Biopsies

Two prostate glands (5-1 from patient 8 and 5-2 from patient 9) were processed sequentially. According to our routine protocol, from each prostate, 17 biopsies were sectioned and labeled. The first sections from both prostates were labeled correctly (5-1-1 and 5-2-1), whereas all of the other sections were assigned to prostate gland 5-1 (two series labeled 5-1-2 to 5-1-17). From patient 8, archived material (5A) was present. An SNP profile was generated from 5-1-1 and 5-2-1 and 5A to determine which SNPs were discriminatory (Table 2). SNP1 and SNP5 (CC and AA for patient 8, AA and AC for patient 9, respectively) were selected to analyze the remaining 32 samples. One section (5-1-12) yielded no results. Because the other section labeled 5-1-12 showed SNP1 and SNP5 alleles of patient 8, this section was assigned to patient 9. Thus, 16 sections could be assigned to patient 8 and 16 sections to patient 9 (Table 2).

Case 6. Tracing of Mammary Tumor

A mammary biopsy, 6-A, was sent in for analysis and found to be malignant. Mammary tissue was removed by mastectomy and analyzed (6-B and 6-C), but no tumor cells were found. The identical SNP profile of the biopsy and two other tissues (Table 2) rendered the possibility of sample mix-up extremely unlikely (1 in $30,000$), which led us to further analyze 6-B and 6-C. Subsequent thorough analysis of the removed mammary tissue revealed a 6-mm medium differentiated malignant tumor in 6-C, thus confirming the exclusion of a sample mix-up.

Discussion

SNP profiling provides a quick, easy, and reliable way to confirm the source of human tissues. The SNP assays to be used to generate an SNP profile were carefully selected to meet certain criteria. We chose a symmetrical distribution of allele frequency (SNPs with minor, and major, allele frequency of ~ 0.5) to maximize the chance that individuals can be distinguished on their SNP profile while analyzing a limited number of SNPs. To minimize the chance of linkage, the 10 SNPs were chosen on 10 different chromosomes. If an SNP consists of two variants (ie, A and C) an individual has three possibilities (AA, AC, and CC). Thus, for 10 SNPs, 3^{10} (59,049) different profiles

exist. Assuming the genotype frequencies to be in Hardy-Weinberg equilibrium, when 10 independent SNPs with $P = 0.5$ are analyzed, the chance for two randomly selected individuals to have the same genotype for one SNP is $0.252 + 0.52 + 0.252 = 0.375$. For 10 SNPs, this chance is $0.375^{10} = 1$ in 18,000. If the minor allele frequency is 0.4, this chance is $0.38565^{10} = 1$ in 14,000, with minor allele frequency of 0.2 this chance is 1 in 900. Thus, the resolution of the test is acceptable in the range of $0.4 < P < 0.5$ but diminishes with more asymmetrical allele frequencies. Three SNP assays (SNP4, SNP10, and SNP12) with observed allele frequencies that differed significantly (χ^2 test; P value of 0.05) from 0.5 were excluded from the recommended SNP panel.

For the Hardy-Weinberg equilibrium, the following assumptions are made: 1) individuals from the population are diploid and reproduce sexually; 2) the population size is infinite; 3) there is no movement from one population to the other; 4) there is no mutation (no biochemical changes in DNA that produce new alleles); 5) mating is random (this means that individuals do not select mates based on the genotype of the SNPs); and 6) the different genotypes of the SNPs have equal fitness. The 10 selected SNPs were located on 10 autosomes, ensuring that two copies of each selected locus are present (diploid) in each individual.

Because most people in The Netherlands are of Caucasian origin, a population that is reasonably assumed to be infinite, we selected SNPs with P values of ~ 0.5 within the Caucasian population. Because allele frequencies differ from one population to the other, the resolution of the test within a non-Caucasian population is lower because of more asymmetrical distribution of alleles. Assuming allele frequencies indicated by Applied Biosystems, for African Americans the discriminatory power of the test using the 10 recommended SNP assays is in the order of 4×10^3 , for Chinese and Japanese even lower, 10^3 . Thus for an SNP profiling test, it is important to select SNPs on the basis of allele frequencies within the population studied. With a limited number of people from other populations present in a Caucasian population, however, the test as presented here will maintain its resolution. Mutation rate in mammalian germ cells is low, estimated to be in the order of 1×10^5 per locus per generation.^{24,25} By choosing SNPs not crucial for gene function and expression, we have reduced the risk of analyzing SNPs related to health issues. When analyzing genetically related individuals, the Hardy-Weinberg equilibrium cannot be assumed, and the resolution of the SNP profiling test is reduced. The resolution of the test is also reduced when individuals from small isolated populations are analyzed.

Another criterion for selection of SNPs for SNP profiling was a low probability of LOH of the region where the SNP is located. In 1 of the 111 tissues analyzed (case 2, a floater of lung tumor tissue in a lymph node), we noticed decreased signals in one of the two alleles in the SNP14 assay (Figure 5). LOH is a common occurrence in cancer tissues.²⁶⁻³² For rs690054 (SNP14) located at 15q21, however, no LOH in lung tumor tissue has been described. Only few reports on LOH of this locus in other

tumor tissue exist.³³⁻³⁷ Circumstantial evidence that the floater indeed belonged to patient 3 was that 1) adjacent to the floater tumor cells were present in the lymph node, 2) another lymph node of the same patient contained tumor cells, and 3) the other SNPs (total of 12) were identical. This leaves LOH as the most likely explanation for the diminished fluorescent signal. The remaining signal, probably attributable to the presence of normal infiltrating cells in the tumor tissue, rendered the SNP profile suitable for interpretation.

With regard to the quality of the assays, the fluorescent signals from some assays were easier to interpret than others. Ideally, the MGB probe complementary to one allele is not hydrolyzed in the presence of the other allele and vice versa. However, hydrolysis of the probes was not always 100% specific, and some cross-reaction occurred. In the SNP10 assay, for instance, the VIC (A) probe cross-reacted slightly with the C allele (Figure 1B, C/C genotype). In the SNP3, SNP8, and SNP15 assays, the FAM probe cross-reacted slightly with the allele recognized by the VIC probe and in the SNP1 assay, the VIC probe showed some cross-reaction (Figures 2 and 4). Because SNP assays 6 and 9 did not perform adequately (data not shown), they were excluded from the panel of SNP assays that we recommend to be used for the SNP profiling assay. The SNP test could successfully be applied to paraffin-embedded human tissues, blood, swabs, and serum samples (data not shown). As expected, the Ct values for paraffin-embedded tissues were higher than for blood samples (~ 28 versus 23, respectively), because of the presence of lower quantities of DNA.

In the identity test as presented here, SNP profiles from recent tissues are compared with SNP profiles from archived samples. As an alternative for archived tissues, blood, or buccal swab can be obtained from the patients involved. We prefer to solve potential error issues, if available, with archived materials. The main reason for this is that the patient needs to be informed of the potential mix-up to obtain a new DNA sample. This may raise doubt with regard to clinical diagnosis and cause the patient unnecessary stress. We realize that when using archived materials, correct labeling of the archived tissue(s) is assumed. The potential mix-up percentage was 0.06% in our laboratory throughout the last 3 years. Although some errors may have remained unnoticed, the actual error rate is supposedly less and in range with error rates reported to be normal (0.08 to 1.2%).¹⁻⁶

Other methods currently used to exclude sample labeling errors include characterization of DNA repeats, ie, short tandem repeats (STRs),⁷⁻⁹ HLA,¹⁰⁻¹⁵ and other polymorphic genetic loci.¹⁴ In addition, commercial identification kits allow identification by use of DNA sequencing equipment.^{16-18,38-40} With STR analysis, when using a combination of many STR systems, the power of discrimination may exceed 10^{11} .⁴¹⁻⁴³ With the AmpFISTR SGM Plus PCR amplification kit up to 5×10^{12} average probability of identity can be obtained. Clearly, such strong evidence can have a powerful effect in courtroom proceedings. However, to solve potential issues of sample errors, the accuracy of our method (18×10^3 average

probability of identity) seemed sufficient. Of course, analysis of more SNPs will increase discriminatory power of the SNP profiling test.

Because the size of DNA fragments is limited in paraffin-embedded tissues, the SNP profiling assay has the significant advantage of requiring the presence of very small DNA fragments (up to 80 bp), whereas STR-based kits require up to 360-bp fragments. In addition, the SNP profiling assay can be performed in any real-time PCR machine and does not require DNA sequencing equipment. A third advantage of the SNP profiling test is that, in contrast to all other methods described so far, processing after PCR is not required in real-time PCR analysis. This substantially reduces the risk of contamination with PCR products. Each SNP profile, as described here, costs approximately €40 (US \$50) per sample. Purchasing large-scale assays and reducing reaction volumes (5- μ l reactions in 384-well plates) can reduce the costs to approximately one fifth. Thus, the costs of SNP profiling are relatively low as compared with the costs of biopsy needles or other assays.

We conclude that we have developed a simple, quick, and reliable SNP profiling test that can be used to confirm the identity of paraffin-embedded tissues to detect potential sample mix-ups. Besides this direct application, the assay can be used to survey the quality of tissue block handling and storage in medical laboratories.

Acknowledgments

We thank Dries Budding for statistical advice and Colin Ingham and Peet Nooijen for critical reading of this manuscript.

References

- Lind AC, Bewtra C, Healy JC, Sims KL: Prospective peer review in surgical pathology. *Am J Clin Pathol* 1995, 104:560-566
- Renshaw AA, Young ML, Jirouek MR: How many cases need to be reviewed to compare performance in surgical pathology? *Am J Clin Pathol* 2003, 119:388-391
- Renshaw AA, Cartagena N, Granter SR, Gould EW: Agreement and error rates using blinded review to evaluate surgical pathology of biopsy material. *Am J Clin Pathol* 2003, 119:797-800
- Ramsay AD, Gallagher PJ: Local audit of surgical pathology. 18 month's experience of peer review-based quality assessment in an English teaching hospital. *Am J Surg Pathol* 1992, 16:476-482
- Safran RE, Bark CJ: Surgical pathology sign-out. Routine review of every case by a second pathologist. *Am J Surg Pathol* 1993, 17:1190-1192
- Wakely SL, Baxendine-Jones JA, Gallagher PJ, Mullee M, Pickering R: Aberrant diagnoses by individual surgical pathologists. *Am J Surg Pathol* 1998, 22:77-82
- Koyama H, Iwasa M, Tsuchimochi T, Maeno Y, Isobe I, Matsumoto T, Nagao M: Utility of Y-STR haplotype and mtDNA sequence in personal identification of human remains. *Am J Forensic Med Pathol* 2002, 23:181-185
- Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM, Smerick JB, Budowle B: Validation of short tandem repeats (STRs) for forensic usage: performance testing of fluorescent multiplex STR systems and analysis of authentic and simulated forensic samples. *J Forensic Sci* 2001, 46:647-660
- Romero RL, Juston AC, Ballantyne J, Henry BE: The applicability of formalin-fixed and formalin fixed paraffin embedded tissues in forensic DNA analysis. *J Forensic Sci* 1997, 42:708-714
- Bateman AC, Sage DA, Al-Talib RK, Theaker JM, Jones DB, Howell WM: Investigation of specimen mislabelling in paraffin-embedded tissue using a rapid, allele-specific, PCR-based HLA class II typing method. *Histopathology* 1996, 28:169-174
- Bateman AC, Hemmatpour SK, Theaker JM, Howell WM: Genetic analysis of hydatidiform moles in paraffin wax embedded tissue using rapid, sequence specific PCR-based HLA class II typing. *J Clin Pathol* 1997, 50:288-293
- Giroti R, Kashyap VK: Detection of the source of mislabeled biopsy tissue paraffin block and histopathological section on glass slide. *Diagn Mol Pathol* 1998, 7:331-334
- Shibata D, Namiki T, Higuchi R: Identification of a mislabeled fixed specimen by DNA analysis. *Am J Surg Pathol* 1990, 14:1076-1078
- Shibata D, Kurosu M, Noguchi TT: Fixed human tissues: a resource for the identification of individuals. *J Forensic Sci* 1991, 36:1204-1212
- Tsongalis GJ, Berman MM: Application of forensic identity testing in a clinical setting. Specimen identification. *Diagn Mol Pathol* 1997, 6:111-114
- Baird ML: Use of the AmpliType PM + HLA DQA1 PCR amplification and typing kits for identity testing. *Methods Mol Biol* 1998, 98:261-277
- Tsongalis GJ, Wu AH, Silver H, Ricci Jr A: Applications of forensic identity testing in the clinical laboratory. *Am J Clin Pathol* 1999, 112:S93-S103
- Walsh PS, Fildes N, Louie AS, Higuchi R: Report of the blind trial of the Cetus AmpliType HLA DQ alpha forensic deoxyribonucleic acid (DNA) amplification and typing kit. *J Forensic Sci* 1991, 36:1551-1556
- Casey D: Scientists Hunt SNPs to Uncover Variation, Disease. *Human Genome Program, U.S. Department of Energy. Washington, DC, Human Genome News*, 1999, p 10
- Afonina I, Kutayin I, Lukhtanov E, Meyer RB, Gamper H: Sequence-specific arrest of primer extension on single-stranded DNA by an oligonucleotide-minor groove binder conjugate. *Proc Natl Acad Sci USA* 1996, 93:3199-3204
- Hardy GH: Mendelian proportions in a mixed population. *Science* 1908, 28:49-50
- Weinberg W: Über den nachweis der verebung beim menschen. *Nat Württemberg* 1908, 64:368-382
- Lo YM, Tein MS, Lau TK, Haines CJ, Leung TN, Poon PM, Wainscoat JS, Johnson PJ, Chang AM, Hjelm NM: Quantitative analysis of fetal DNA in maternal plasma and serum: implications for noninvasive prenatal diagnosis. *Am J Hum Genet* 1998, 62:768-775
- Russell LB, Russell WL: Frequency and nature of specific-locus mutations induced in female mice by radiations and chemicals: a review. *Mutat Res* 1992, 296:107-127
- Kuick RD, Neel JV, Strahler JR, Chu EH, Bargal R, Fox DA, Hanash SM: Similarity of spontaneous germinal and in vitro somatic cell mutation rates in humans: implications for carcinogenesis and for the role of exogenous factors in "spontaneous" germinal mutagenesis. *Proc Natl Acad Sci USA* 1992, 89:7036-7040
- Girard L, Zochbauer-Muller S, Virmani AK, Gazdar AF, Minna JD: Genome-wide allelotyping of lung cancer identifies new regions of allelic loss, differences between small cell lung cancer and non-small cell lung cancer, and loci clustering. *Cancer Res* 2000, 60:4894-4906
- Arvanitis DA, Papadakis E, Zafiropoulos A, Spandidos DA: Fractional allele loss is a valuable marker for human lung cancer detection in sputum. *Lung Cancer* 2003, 40:55-66
- Hughes C, Murphy A, Martin C, Sheils O, O'Leary J: Molecular pathology of prostate cancer. *J Clin Pathol* 2005, 58:673-684
- Laigle-Donadey F, Martin-Duverneuil N, Lejeune J, Criniere E, Capelle L, Duffau H, Cornu P, Broet P, Kujas M, Mokhtari K, Carpentier A, Sanson M, Hoang-Xuan K, Thillet J, Delattre JY: Correlations between molecular profile and radiologic pattern in oligodendroglioma tumors. *Neurology* 2004, 63:2360-2362
- Hussein MR: Genetic pathways to melanoma tumorigenesis. *J Clin Pathol* 2004, 57:797-801
- Dacic S, Sasatomi E, Swalsky PA, Kim DW, Finkelstein SD, Yousem SA: Loss of heterozygosity patterns of sclerosing hemangioma of the

- lung and bronchioloalveolar carcinoma indicate a similar molecular pathogenesis. *Arch Pathol Lab Med* 2004, 128:880–884
32. Feitelson MA, Pan J, Lian Z: Early molecular and genetic determinants of primary liver malignancy. *Surg Clin North Am* 2004, 84:339–354
33. Poetsch M, Kleist B: Loss of heterozygosity at 15q21.3 correlates with occurrence of metastases in head and neck cancer. *Mod Pathol* 2006, 19:1462–1469
34. Jou YS, Lee CS, Chang YH, Hsiao CF, Chen CF, Chao CC, Wu LS, Yeh SH, Chen DS, Chen PJ: Clustering of minimal deleted regions reveals distinct genetic pathways of human hepatocellular carcinoma. *Cancer Res* 2004, 64:3030–3036
35. Natrajan R, Louhelainen J, Williams S, Laye J, Knowles MA: High-resolution deletion mapping of 15q13.2-q21.1 in transitional cell carcinoma of the bladder. *Cancer Res* 2003, 63:7657–7662
36. Beder LB, Gunduz M, Ouchida M, Fukushima K, Gunduz E, Ito S, Sakai A, Nagai N, Nishizaki K, Shimizu K: Genome-wide analyses on loss of heterozygosity in head and neck squamous cell carcinomas. *Lab Invest* 2003, 83:99–105
37. Kersemaekers AM, Kenter GG, Hermans J, Fleuren GJ, van de Vijver MJ: Allelic loss and prognosis in carcinoma of the uterine cervix. *Int J Cancer* 1998, 79:411–417
38. Pontes ML, Caine L, Abrantes D, Lima G, Pinheiro MF: Allele frequencies and population data for 17 Y-STR loci (AmpFISTR(R)) Y-filer-trade mark) in a Northern Portuguese population sample. *Forensic Sci Int* 2007, [Epub ahead of print]
39. Mulero JJ, Chang CW, Calandro LM, Green RL, Li Y, Johnson CL, Hennessy LK: Development and validation of the AmpFISTR Yfiler PCR amplification kit: a male specific, single amplification 17 Y-STR multiplex system. *J Forensic Sci* 2006, 51:64–75
40. Collins PJ, Hennessy LK, Leibelt CS, Roby RK, Reeder DJ, Foxall PA: Developmental validation of a single-tube amplification of the 13 CODIS STR loci, D2S1338, D19S433, and amelogenin: the AmpFISTR Identifiler PCR amplification kit. *J Forensic Sci* 2004, 49:1265–1277
41. Lander ES, Budowle B: DNA fingerprinting dispute laid to rest. *Nature* 1994, 371:735–738
42. McKusick VA, Ferrara PB, Kazazian HH, King MC, Lander ES, Lee HC, Lempert RO, Macklin R, Marr TG, Reilly PR, Senbaugh Jr GF, Weinstein JB: DNA technology in forensic science. Committee on DNA Technology in Forensic Science, Board on Biology, Commission on Life Sciences, National Research Council. Washington, DC, National Academy Press, 1992
43. Crow JF, Berger MA, Diamond SS, Kaye DH, Kazazian HH, Motulsky AG, Nagylaki TA, Nei M, Sensabaugh GF, Siegmund DO, Stigler SM: The evaluation of forensic DNA evidence. Committee on DNA Forensic Science: An Update, Commission on DNA Forensic Science: An Update, National Research Council. Washington, DC, National Academy Press, 1996