Metabolic Polymorphisms and Susceptibility to Cancer
W. Ryder
IARC Scientific Publications No. 148
International Agency for Research on Cancer, Lyon, 1999

Chapter 13. Laboratory methods for the determination of genetic polymorphisms in humans

B. Blömeke and P. G. Shields

The determination of genetic polymorphisms for susceptibility to human disease has been rapidly increasing since the introduction of the polymerase chain reaction (PCR). In most laboratories the ability exists to conduct studies on more than 10 000 persons, and the prospect of even larger investigations is approaching. Many methods can be used for genotyping individuals but some are more common and less expensive than others. Newer methods will allow for automation. As the number of studies on genetic polymorphisms increases it is to be expected that more pitfalls will be encountered. While larger studies will reduce the importance of misclassification, quality control methods will have to be applied to the processing of large numbers of samples.

Sources of DNA and DNA extraction

The most critical element in the successful analysis of genetic polymorphism is the quality of DNA. When the quality is poor, so is the result. Quality depends on the method of collection, the way in which cells or tissues are stored (e.g. method of freezing, storage temperature, media [paraffin blocks, dried blood dots]), DNA storage, duration of storage, the type of extraction, and the experience of the person performing it. DNA can be extracted from fresh, archival and ancient sources. For each, the number of possible sources has been steadily increasing as new methodologies have emerged (Table 1). For example, forensic analysis using the polymerase chain reaction (PCR) may use a single hair (Han *et al.*, 1992), sperm (Sajantila *et al.*, 1992), dried blood on a blotter (Jinks *et al.*, 1989), urine (Gasparini *et al.*, 1989), and even saliva on cigarette butts (Hochmeister *et al.*, 1991).

One of the most useful tissues for storage is whole blood because it requires little processing time and cost before archiving. Cell separation is avoided and the risk of contamination is reduced when large epidemiological studies are being conducted. Whole blood is typically stored in anticoagulants such as heparin or EDTA, although with heparin some problems in the PCR have occasionally been reported. It is believed that heparin and porphyrin compounds inhibit PCRs, but purified haemoglobin or iron added to a PCR were not inhibitory (Cosma *et al.*, 1993), and there have been many successful studies. Whole blood storage might be particularly useful if Tth polymerases are used (Erlich *et al.*, 1991). This is a more robust enzyme than Taq™ polymerase (Panaccio *et al.*, 1993; Sparkman, 1992), and reduced temperature (FoLT = formamide low temperature) allows direct DNA amplification from blood without extraction. (Note, low temperatures [80°C-85°C for initial melting and annealing temperatures of 40°C or less] could increase non-specific priming and should therefore be tested). A common method for DNA extraction from whole blood is a Chelex™ 100 protocol (Bio-Rad Laboratory, Richmond, CA) (De Lamballerie *et al.*, 1992). Here, the blood is boiled in the presence of Chelex™ 100 resin. DNA recovery is reproducible, and is useful for small volumes (2-10 µl). Chelex™ 100 resin is routinely used for bloodstains, seminal stains, buccal swabs, hair and postcoital samples. A different approach for extracting DNA from small amounts (5 µl) of whole blood is immuno-PCR (Lew & Kelly, 1991). Using anti-histone antibodies in a capture assay or a microtitre tray, chromosomal DNA can be selectively bound and the PCR can then be performed in the same tube. If high molecular weight is needed, as with Southern blot

Table 1. Sources of DNA for genetic polymorphism analysis

DNA source	Method	Reference
Whole blood	FoLT PCR (formamide low temperature) Boiling with Chelex™ 100 Immuno-PCR Rapid chemical extraction of high quality DNA	(Panaccio & Lew, 1991) (Panaccio *et al.*, 1993) (De Lamballerie *et al.*, 1992) (Lew & Kelly, 1991) (Jeanpierre, 1987) (Gustincich *et al.*, 1991)
Leukocytes	Proteinase K/chemical extraction	(Sugimura *et al.*, 1990) (Jeanpierre, 1987)
Hairs	Proteinase K/chemical extraction	(Higuchi *et al.*, 1988) (Han *et al.*, 1992) (Westwood & Werrett, 1990)
Sperm	Proteinase K/chemical extraction	(Sajantila *et al.*, 1992)
DNA from dried blood on blotter and cotton cloth	Chelex™ 100 Proteinase K/chemical extraction	(Jinks *et al.*, 1989) (Walsh *et al.*, 1991) (Lin *et al.*, 1994)
Urine	Proteinase K/chemical extraction	(Gasparini *et al.*, 1989)
Saliva on cigarette butts	Proteinase K/chemical extraction	(Hochmeister *et al.*, 1991)
Fresh tissues and biopsies	Proteinase K/chemical extraction	(Smith *et al.*, 1987) (Shibata, 1992)
DNA swabs (cervical, vulvar, penile, buccal cells, nostril)	Cell lysis DNA adsorption to glass matrices Proteinase K/chemical extraction	(Nickerson *et al.*, 1990) (Nikiforov *et al.*, 1994) (Bertilsson *et al.*, 1989) (Smits *et al.*, 1992)
DNA from ethanol-fixed samples	Proteinase K/chemical extraction	(Broly *et al.*, 1995)
Formaldehyde and paraffin-embedded tissues	Proteinase K/chemical extraction Sonication with glass beads/Proteinase K Boiling	(Goelz *et al.*, 1985) (Greer *et al.*, 1994) (Heller *et al.*, 1992) (Kallio *et al.*, 1991)
Museum specimens	Proteinase K/chemical extraction Centrifugation on Centricon 30 filters	(Higuchi *et al.*, 1984; Paabo, 1985) (Paabo, 1989)

analyses, the method of Gustincich *et al.* (1991) can be used. Typical phenol-based extraction methods can also be employed (Sambrook *et al.*, 1989). Initially, it is optimal to store whole blood by freezing (-70°C). The samples can be stored on ice and shipped, but freezing should occur within 24 hours at the most. However, when whole blood is frozen it is only useful for DNA recovery, as cell membranes are cracked and intact cells are unrecoverable. Plasma is mixed with cell contents.

Air-dried blood smears are another potential source of DNA, for example from smears on a microscope slide (Fey *et al.*, 1987) or from dots blotted on to cards or cloth (Jinks *et al.*, 1989; Lin *et al.*, 1994; Walsh *et al.*, 1991). These DNA sources allow easy and prolonged conservation of samples, especially for the amplification of PCR fragments of less than 200 bp (Jinks *et al.*, 1989; Lagarde *et al.*, 1995). For microscope slides, samples are air-dried without fixing, or they can be fixed with ethanol or methanol. For bloodstains

on paper (Jinks *et al.*, 1989; Lin *et al.*, 1994) or cotton cloth (Walsh *et al.*, 1991), small areas of the original samples are cut (3 x 3 mm pieces), resuspended, and extracted using Chelex™ 100 or methods involving the use of phenol. In our laboratory we found only a 50% success rate for bloodstains on blotter using fragment sizes greater than 250 bp and a phenol-based extraction method, while Walsh et *al.* (1991) reported that 100% of their samples could be amplified using a Chelex™ 100 extraction method. These samples, while on cards, are optimally stored by freezing at -70°C, although storage at room temperature for periods of less than six months may be acceptable.

Individual blood components frequently serve as DNA sources, the most comonly used being white blood cells. Serum and plasma also provide DNA for genotyping (Lau *et al.*, 1995; Chan *et al.*, 1995; Stein *et al.*, 1995; Martin *et al.*, 1992; Secchiero *et al.*, 1995; Kato *et al.*, 1990; Leon *et al.*, 1977; Shapiro *et al.*, 1983; Vasioukhin *et al.*, 1994; Sorenson *et al.*, 1994), where the DNA comes from lysed cells or is released from lymphocytes (Anker *et al.*, 1975; Leon *et al.*, 1977). For DNA extraction, samples are concentrated by centrifugation (Kato *et al.*, 1990) or are directly digested with proteinase K. Digests are extracted with organic solvents or are used immediately for PCR amplification (Kato *et al.*, 1990). Some investigators isolated DNA using a phenol/chloroform/ether extraction followed by Cs_2SO_4 gradient (Vasioukhin *et al.*, 1994). Recently, we have shown that serum can be routinely used for PCR/RFLP genotyping of cancer susceptibility genes. The isolated DNA was intact to amplify a region of at least 529 base pairs, yields were between 162 and 1,060 ng DNA using 250 µl serum (n = 18). Based on our average usage of 20 ng per PCR, the extractable DNA could be used for 10 to 53 PCR-based genotyping assays. The possibility of genotyping using serum (or plasma) opens the field to many archived settings where conventional DNA sources are unavailable. After separation, the blood components should be frozen at -70°C.

DNA can be obtained non-invasively from the head. Hair bulb DNA can be obtained by washing the bulb in ethanol, digesting with Proteinase K and extracting by means of phenol-based methods (Higuchi *et al.*, 1988; Han *et al.*, 1992), or fixed in methanol and directly used for the PCR (Han *et al.*, 1992). The latter protocol was successfully used for two to six bulbs from freshly plucked hairs for each PCR (Westwood & Werrett, 1990). DNA can be obtained from the oral cavity (buccal membrane) or a nostril by swabbing with a cytobrush or cotton swab (Nikiforov *et al.*, 1994) or taking oral rinses (Nickerson *et al.*, 1990). The cell density of a single swab varies between 5 x 10^3 and 3 x 10^5 cells, and this is sufficient for PCR-based methods. Oral rinses with saline, preceded by swabbing, also yield acceptable DNA.

Tissue from pathology departments is an important source of archived material, in the form of preserved microscope slides, frozen tissues or paraffin-embedded tissue blocks. The type of fixative can greatly influence the quality of extracted DNA. The best fixative is 50% ethanol; formaldehyde and other fixatives are less satisfactory. Formaldehyde (4% aqueous solution) is the most commonly used fixative, and while it is good for proteins, reduced DNA yields are obtained because DNA leaks into the fixative solution (Jackson *et al.*, 1990). Extracting DNA from samples fixed in other agents is also possible, such as neutral-buffered formalin, paraformaldehyde, formol sublimate, and even Bouin's reagent, a fixative based on picric acid (Greer *et al.*, 1994)). As a rule of thumb, the smaller the fragment to be amplified by PCR (<200 bp), the greater the likelihood of success. Most protocols involve complete digestion of a dewaxed sample with Proteinase K (sometimes taking up to five days for larger pieces of tissue). Faster methods such as sonication (Heller *et al.*, 1992) and boiling (Kallio *et al.*, 1991) are available, but may be less efficient than the use of Proteinase K (Forsthoefel *et al.*, 1992). Phenol-based methods can also be used (Sambrook *et al.*, 1989). Using proteinase K digestion, we obtained DNA yields of between 1 µg and 11.7 µg from 5-µm and 20-µm sections.

Various methods have been described for DNA extraction from fresh tissues and biopsies. For cryopreserved material it is usual to begin by grinding up to 100 mg of tissue in liquid nitrogen to produce a fine powder. The resulting homogenate is used for the DNA extraction procedure. In the case of small fresh and frozen biop-

sies even crude lysates have been used for PCR amplifications (Shibata, 1992).

DNA from ancient organic remains has been used in several anthropological and evolutionary studies (Kato *et al.*, 1994; Paabo, 1989) and in forensic science (Higuchi *et al.*, 1984). The tissue can either be crushed or minced before Proteinase K digestion (for up to five days). The digest is then extracted with phenol/chloroform, or the DNA is recovered by centrifugation on Centricon 30 filters (Amicon, Beverly, MA) (Paabo, 1989). DNA yields vary between 200 ng and 20 μg per gram of dried tissue, approaching 0.005% and 5% respectively of the amount expected from fresh tissue. Because nuclear DNA is degraded after death by endogenous hydrolytic processes (Rebrov *et al.*, 1983) it seems likely that the rapidity with which the body is desiccated immediately after death is the major factor that determines the yield and suitability of DNA. Oxidation products of pyrimidines, AP sites, damaged sugar residues, as well as intra- and intermolecule cross-links are potential problems. Because of post-mortem changes in DNA, caution has to be exercised in the use of these samples. Several microsatellite repeats can be feasibly analysed in DNA from bones and teeth up to 5000 years of age. However, Ramos *et al.* (1995) found that the repeated analysis of each marker produced different genotypes in as many as 97% of samples. Alleles differing from the originals consisted of additions or deletions of 1-39 dinucleotides.

Once extracted, DNA should be stored at temperatures appropriate to its intended use. For long-term storage, -70°C is optimal. However, multiple aliquots are preferable in order to avoid repetitive freezing and thawing, which could result in DNA degradation. In our laboratory we try to store samples in aliquots useful for several months of assays, so that once thawing has occurred the samples can be left at 4°C. However, if the samples are not going to be used for at least four months we freeze them again. We have also stored samples at room temperature for over six months in aliquots useful for one PCR, without problems. The samples typically become dry but amplification is successful. Whenever working with stored DNA it is important to use a sterile technique so as to avoid microbial contamination. Furthermore, we try to manipulate samples as little as possible. Multiple pipetting seems to be a greater factor in DNA degradation than length of storage.

Genetic polymorphism analysis by the polymerase chain reaction (PCR)

The PCR is an *in vitro* method for enzymatically synthesizing defined sequences of DNA. The reaction uses two oligonucleotide primers that hybridize to opposite strands and flank the target DNA sequence of interest. The elongation of the primers to amplify the gene is catalysed by a thermostable polymerase, the one most often used being Taq™ DNA polymerase, which is derived from the thermophilic eubacterium *Thermus aquaticus*. Usually, all the reagents (deoxynucleotide triphosphates (dNTPs), primers, reaction buffer, target DNA and Taq™ polymerase) are mixed and a repetitive series of cycles involving template denaturation, primer annealing, and extension of the primers by the enzyme results in exponential accumulation of a specific DNA fragment. Special procedures to improve the PCR include the hot-start procedure, which is a prolonged step of denaturation administered before Taq™ polymerase addition, and a two-round amplification with nested priming (using two separate PCR assays, where the product of the first serves as the template for the second). Recently, the long PCR has been used, in which the thermostable DNA polymerases are employed (e.g. Vent, Deep Vent, Pwo or Pfu polymerase) to amplify fragments as long as 42 000 bp (see below). This assay, while useful, requires better quality DNA and success is more difficult to achieve.

An important component of a successful PCR lies in the selection of primers (Table 2). Ideally, the primers should have the same annealing temperature, resulting in high amplification efficiency. In the case of multiplex PCR amplification (amplifying more than one gene locus in a single reaction using more than one primer pair), the T_m values for all primer sets should be within the range of ±5°C. Similar annealing temperatures can be achieved by designing primers that are the same length and display not only a similar G/C base content but also similar percentage of G and C bases within the primer. A G plus C base percentage near 50% is optimal, allowing for effi-

Parameter

Oligonucleotide sequenc

No self-complementarity

No complementarity to a

Match primer melting ter
(T_m [°C] = 2AT+4GC)

Base composition

Base distribution

Primer length

Distance of intraprimer se
1. Fresh tissue, blood, ce
2. Plasma, serum, formal
embedded tissues
3. Museum specimens

cient melting during
anneal with sequences
tents can be compensa
length, and primers of
ommended as being st
human genome (Thein
end of a primer is impo
annealing temperature.
rules for primer design
no guarantee of success.
available to help in t
primers (Hillier & Green

DNA polymerases ca
rate, which varies depen
merase, the substrate an
During the PCR, amplifi
template for subsequent
are consequently inherit
accumulate. The errors
merase, the most widel
amplification methods,
per nucleotide at high d
centrations, and as low a
tions. Other enzymes, su
Pyrococcus furiosus (Pfu™
have an associated 3'→
which selectively rem
nucleotides to generate

Table 2. Optimum PCR primer designs

Parameter	Optimum values
Oligonucleotide sequence	Unique
No self-complementarity	≤3 contiguous bases
No complementarity to another (especially at the 3' end)	≤3 contiguous bases
Match primer melting temperatures (T_m [°C] = 2AT+4GC)	Less than 5°C if possible
Base composition	G/C content near 50%
Base distribution	Random or 1-2 G/C nucleotides at the 3' end
Primer length	20-25 bases
Distance of intraprimer sequence	
1. Fresh tissue, blood, cells, alcohol-fixed samples	100-2000 bases apart
2. Plasma, serum, formaldehyde-fixed and paraffin-embedded tissues	100-600 bases apart
3. Museum specimens	100-150 bases apart

cient melting during the PCR. Primers that anneal with sequences displaying lower G/C contents can be compensated for by an increase in length, and primers of at least 17 bases are recommended as being statistically unique in the human genome (Thein & Wallace, 1986). The 3'-end of a primer is important for determining the annealing temperature. However, most of the rules for primer design are empirical and there is no guarantee of success. Computer programs are available to help in the selection of unique primers (Hillier & Green, 1991).

DNA polymerases can make errors at a low rate, which varies depending on the type of polymerase, the substrate and the reaction conditions. During the PCR, amplification products serve as a template for subsequent cycles. Polymerase errors are consequently inherited and these changes can accumulate. The errors produced by Taq™ polymerase, the most widely used for *in vitro* DNA amplification methods, can be higher than 10^{-3} per nucleotide at high dNTP and Mg^{2+} ion concentrations, and as low as 10^{-6} under other conditions. Other enzymes, such as polymerases from *Pyrococcus furiosus* (Pfu™) and *P. woesei*, (Pwo™) have an associated 3'→5' exonuclease activity, which selectively removes misincorporated nucleotides to generate a correctly base-paired primer terminus (proofreading), resulting in much higher fidelity. When the errors during PCR are distributed evenly throughout the fragment, no particular mutated sequence constitutes a major subpopulation. Here, the predominant species at each nucleotide position is that of the initial sequence. Therefore, for the detection of genetic polymorphisms where a homogeneous DNA population is analysed, PCR mistakes are of little concern. However, parameters increasing the error rate and the known context-dependent effects on the fidelity of polymerases (Goodman *et al.*, 1993) are to be avoided.

The ability to amplify as much as 25 kilobases of DNA using PCR (long PCR) has recently provided new opportunities. The combination of thermostable DNA polymerases with proofreading activity (e.g. Vent, Deep Vent, Pwo or Pfu polymerase) and ones lacking such activity (e.g. rTth polymerase) together with improved buffer and cycling conditions (longer extension times) (Cheng *et al.*, 1994a; Barnes, 1994; Cheng *et al.*, 1994b) can achieve longer amplification. Typical primers for long PCR amplifications have been 21-34 bp with melting temperatures near 65-70°C. The specificity of the primers is critical, as the amplification of long targets is compromised by preferential amplification of shorter

nonspecific products. The product must be verified as the full-length target. Misamplification can be avoided by repeated reactions, use of different primer pairs or product digestion with specific restriction enzymes. This method allows the amplification of genes long enough to include several polymorphic sites and multiple adjacent genes (e.g. GSTM1 and GSTM3), and can be used to amplify across polymorphically deleted genes. The deletion is thereby seen conclusively. It also reduces the risk of amplifying pseudogenes (Steen *et al.*, 1995; Broly *et al.*, 1995).

Detection of genetic polymorphisms

The study of genetic polymorphisms in humans is playing an important role in the diagnosis of genetic and malignant diseases (Kerem *et al.*, 1989; Fearon & Vogelstein, 1990), forensic sciences (Nakamura *et al.*, 1987; Lander, 1989), and in gene mapping projects (Donis-Keller *et al.*, 1987). Each of these applications involves the analysis of many samples and requires rapid, inexpensive, automated methods. Usually, methods that can be used to discover new mutations can be applied to typing those that are already known, as with Southern blot analysis (Southern, 1975), restriction fragment length polymorphism analysis (Botstein *et al.*, 1980), allele-specific oligonucleotide hybridization (Conner *et al.*, 1983), denaturing gradient gel electrophoresis (Myers *et al.*, 1987), heteroduplex analysis (Keen *et al.*, 1991), single-strand conformation polymorphism (Hayashi, 1991), allele-specific PCR (Chehab & Kan, 1989; Wu *et al.*, 1989; Newton *et al.*, 1989), allele-specific oligonucleotide probes (Saiki *et al.*, 1986), and the detection of amplified products by oligomer hybridization (Kwok *et al.* 1989). Some detection methods do not rely on electrophoresis, for instance the oligonucleotide ligation assay (Landegren *et al.*, 1988; Nickerson *et al.*, 1990) and the primer-guided incorporation techniques, e.g. genetic bit analysis (Nikiforov *et al.*, 1994).

Restriction length fragment polymorphisms (RFLP) have been frequently used (Goodfellow, 1992) for genetic linkage maps (Botstein *et al.*, 1980), to identify individuals in forensic science (Nakamura et al., 1987), and for the detection of diseases (Chehab *et al.*, 1987). RFLPs can be used for genetic polymorphism analysis where the polymorphic site is within a palindromic run of nucleotides susceptible to cleavage by a known restriction endonuclease. The resulting fragment size pattern can be detected by Southern analysis (Budowle & Baechtel, 1990) or PCR followed by restriction digestion, and gel electrophoresis with ethidium bromide staining (Day & Humphries, 1994). It should be noted that restriction enzyme cleavage may be inefficient when the palindromic site is found near the end of the DNA fragment (for specifics see manufacturer's instructions). This is important if misclassification by incomplete digestion is to be avoided. Primers may be designed to include a non-polymorphic site susceptible to cleavage within the PCR fragment, in order to show completeness of digestion and enhance quality control. When the polymorphic site of interest is not contained within a restriction site it may be possible to introduce a palindrome artificially into a primer by modifying one or two bases so that a restriction site is created and the restriction enzyme cleaves off the primer.

The allele-specific PCR is commonly used when a polymorphic site of interest is not found within a palindromic sequence susceptible to restriction enzymatic digestion. This method is based on the principle that mismatches of the base at the 3' end of a primer do not amplify by PCR, or amplify at very low rates. The Taq™ polymerase is especially useful for this because it lacks 3'-5' exonuclease activity. In its simplest form, two PCR assays are performed in parallel, using a common primer in both and unique allele-specific primers in each, the latter primers being matched to one polymorphic variant or the other. Perfect matches can then amplify, revealing which sequence is present. However, this method has the limitation that the lack of amplification may suggest a failed PCR attempt (multiplex PCR, amplifying an additional gene as an internal control, eliminates this possibility). Alternatively, the presence of a band might indicate false amplification, as occurs under suboptimal conditions (low annealing temperatures or magnesium concentrations). In fact this method is qualitative, not quantitative, and false amplification can therefore easily occur under the wrong conditions. It is thus important to optimize the conditions and to compare the intensity of bands

for the matched
same gel. This
two-step assay, t
of interest by P
oligospecific pro
tions. The adva
the specificity by
plates. However,
one; a lack of su
sistent results in
due to a suboptir

PCR coupled v
assay (OLA) uses
oligonucleotides
plate under opt
occur only when
ly complementar
al., 1988; Nicker
reactions are run
oligonucleotides
the polymorphic
For this method
biotinylated and
at its 3' end wi
primer is ligated
matched to the D
5' biotinylated
vidin-coated mic
occurred the pres
is detected in an
tage of this met
dure can be pe
allowing the au
are available) and
disadvantage is t
cessful amplificat
each reaction wel

Genetic bit a
group of primer-
assays (Syvanen
et al., 1991). In
1994), specific fr
taining the polyr
PCR with one reg
modified primer.
rendered single
microtitre plates
lized oligonucleot
hybridize to the D
cent to the polym

for the matched and mismatched primers on the same gel. This method is frequently used as a two-step assay, the first step amplifying the gene of interest by PCR, and the second PCR using oligospecific probes in two simultaneous reactions. The advantage of this is that it increases the specificity by increasing the number of templates. However, the most critical step is the first one; a lack of successful amplification or inconsistent results in the second step are frequently due to a suboptimal first step.

PCR coupled with the oligonucleotide ligation assay (OLA) uses DNA ligases to join two adjacent oligonucleotides when hybridized to a DNA template under optimized conditions. Ligations occur only when the oligonucleotides are perfectly complementary to the template (Landegren *et al.*, 1988; Nickerson *et al.*, 1990). Two ligation reactions are run in parallel, corresponding to the oligonucleotides matching in sequence either of the polymorphic variants (allele-specific probes). For this method the allele-specific probes are biotinylated and the common primer is labelled at its 3′ end with digoxigenin. The common primer is ligated to the allele-specific probe when matched to the DNA sequence. After ligation, the 5′ biotinylated probes are captured on strepavidin-coated microtitre plates. If ligation has occurred the presence of the digoxigenin reporter is detected in an ELISA-based format. The advantage of this method is that the complete procedure can be performed on microtitre plates, allowing the automation (robotic workstations are available) and rapid processing of samples. Its disadvantage is that there is no control for successful amplification and no internal control in each reaction well.

Genetic bit analysis (GBA) belongs to the group of primer-guided nucleotide incorporation assays (Syvanen *et al.*, 1990, 1993; Kuppuswamy *et al.*, 1991). In this method (Nikiforov *et al.*, 1994), specific fragments of genomic DNA containing the polymorphic site(s) are amplified by PCR with one regular and one phosphorothioate-modified primer. The double-stranded DNA is rendered single-stranded and captured on microtitre plates by hybridization to an immobilized oligonucleotide primer, which is designed to hybridize to the DNA sequence immediately adjacent to the polymorphic site. A DNA polymerase extends the 3′ end of the oligonucleotide by just one base, using a mixture of one biotin-labelled and one fluorescein-labelled dideoxynucleotide triphosphate (ddNTP), corresponding to the bases of the two polymorphic alleles. Antibody conjugates of alkaline phosphatase and horseradish peroxidase are used to determine the nature of the extended base in an ELISA format. The method is highly flexible but has no internal control for successful PCR amplifications. The use of two labelled ddNTPs, allowing the determination of both alleles in the same well, increases throughput, reduces the amount of PCR targets required and, more importantly, serves as a powerful internal control for all post-PCR steps.

Dot or slot blot and reverse dot blot hybridizations are non-isotopic methods that use allele-specific oligonucleotide probes (Saiki *et al.*, 1986; Saiki *et al.*, 1989). Dot blots are performed following PCR in the region of interest, and the PCR product is blotted on to a membrane. The allele-specific probes are then hybridized to the blots, and under optimal conditions only the matched probes bind. The allele-specific probes are approximately 19 bp in length and the polymorphic base is located at the centre of the probe. For the reverse dot blot, allele-specific probes are fixed to a membrane strip and the amplified PCR fragments are hybridized to the immobilized probes. Binding to the complementary polymorphic site is detected via a biotin tag on the 5′ end of the primer, followed by addition of a strepavidin-horseradish peroxidase complex that oxidizes tetramethylbenzidine, yielding a blue dot (Garcia-Pacheco *et al.*, 1995). The advantage of this method is that a panel of oligonucleotide probes can be used to type multiple alleles. The sensitivity is increased by the reverse dot blot because the amplified material is hybridized against several allelic probes in a single hybridization assay. Dot blots are ideal for polymorphisms that are not included in a palindromic restriction site, especially when only small amounts of starting material or degraded DNA are available for typing. The technique is at least ten times more sensitive then PCR-RFLP. Maekawa *et al.* (1995) compared the feasibility of this method with PCR-RFLP for the three common apolipoprotein E (apo E) genotypes using DNA from leukocytes. They found complete agreement for each sample

(n = 93) but concluded that the RFLP was simpler to perform and more suitable for large-scale studies.

Conformational changes in single-stranded DNA caused by mutation can be exploited to detect polymorphic sites using single-stranded conformation polymorphism (SSCP) analysis (Orita *et al.*, 1989). Since its inception, SSCP has been used very extensively in connection with PCR amplification (PCR-SSCP) for the detection of point mutations (for example p53, reviewed by Hayashi, 1991). It is a highly sensitive method, but its specificity and sensitivity vary with the technique employed. PCR-SSCP is a simple alternative to other genotyping procedures and allows for the simultaneous detection of both known and unknown mutations. In this analysis the target sequence is amplified by PCR, at which time the product is end-labelled by radioactive primers (Hayashi *et al.*, 1989) or fluorescent primers (Broly *et al.*, 1995), or the product is labelled during PCR by the addition of radioactive deoxynucleotide triphosphates to the PCR mixture (Petersen *et al.*, 1990). The amplified product is then heated to dissociate the strands and is subjected to non-denaturing polyacrylamide gel electrophoresis. Polymorphism variants are detected as shifts in the mobility of bands using autoradiography or flourography. The latter method is suitable for an automated analysis. Although detection of alterations in electrophoretic mobility is strongly dependent on the size of the fragment (maximum 300-400 nucleotides per gel), this limitation can be avoided by enzymatic digestion of a large PCR fragment to smaller fragments (Peinado *et al.*, 1993) or by SSCP analysis on smaller secondary fragments generated with a set of nested primers (Broly *et al.*, 1995).

Quality control

The analysis of samples for epidemiology or the diagnosis of disease in the individual requires specific and consistent quality control measures. In clinical pathology laboratories, standards and protocols have been established by organizations such as the College of American Pathologists and The National Committee for Clinical Laboratory Standards. Their interests include forensic and paternity testing, beyond diagnostic assays, so that most of their criteria apply to the research laboratory performing genetic polymorphism analysis. There are standards for proficiency testing, quality improvement and control, use of standards, methods of interpretation, specimen handling, labelling and processing, and reporting of results. There are also criteria for facilities and the maintenance of equipment.

While studies in research settings may not demand the rigour of those in commercial laboratories it is nevertheless important to check continuously that results are reliable and that misclassification is avoided. Quality control methods are therefore needed as from initial sampling to the reporting of results.

The verification of sample collection obviously depends on the method of collection. New methods should be pilot-tested; those requiring a procedure to be effected by the study subject are difficult to verify. Protocols should be in place to ensure that materials (i.e. preservatives) have not been held beyond their expiry date. The times of sample collection, shipment and processing should be recorded in a logbook. A formal chain of custody may be helpful. One potential difficulty during sample collection is that of improper labelling. A unique identification number should be given to each subject at the time of enrolment, identifying both the subject and the study. Preprinted labels can be placed on the sample containers and the questionnaire to ensure that there are no errors. Labels should be tested before use to ensure that they remain fastened at freezer temperatures.

The enhanced sensitivity provided by a method such as PCR amplification demands rigorous attention to possible sample-to-sample contamination, contamination with inhibitors or organisms that can degrade DNA, and cross-contamination of DNA with PCR products or control DNA. During DNA extraction by phenol-based methods the organic and aqueous phases must be adequately separated to reduce contamination with RNA, haemoglobin and proteins. For Chelex™ 100 and other commercially distributed extraction methods the manufacturer's instructions should be followed closely and the methods should be tested before use in studies. The probability of contamination from organisms and of the cross-contamination of samples is reduced if gloves are changed frequently and if unnecessary manipulation in avoided. It is preferable to use

screw-cap tubes rather than flip-top tubes to avoid contaminating the fingers when tubes are opened (openers, similar in style to can-openers, greatly reduce this problem). Absolute sterile technique is required because moulds can easily grow in samples and either inhibit the PCR or degrade the DNA. When an inhibitor is suspected, dilution of the sample (i.e. using less DNA) may result in successful amplification. If degradation is suspected, quantitation by UV absorbance may not change, but the performance of another assay in which a larger fragment is successfully amplified helps to rule out degradation. When extracting DNA from paraffin blocks, microtome blades and gloves should be changed after each sample to avoid contamination with wild-type DNA (i.e. from another sample or from desquamated skin cells of the technician, which might still have better quality DNA that is more easily amplified than the fixed DNA). No matter what the source of tissue or DNA, specimens should be prepared and stored at sites physically separated from locations where PCR amplification is performed or PCR products (i.e. running gels) are otherwise used, so as to avoid contaminating the sample with amplified PCR products. Nonetheless, to confirm that contamination has not occurred, a negative control (i.e. PCR of a PCR reaction mixture, set up during the time that samples are to be analysed, but without the addition of template DNA) must be included with every set of samples (we use one per row of gel - 20 samples) to monitor and identify sample-to-sample contamination. These controls must be carried through all phases of sample processing and amplification. Positive controls carrying a known sequence are also useful but should be included only as the final sample to reduce cross-contamination.

After the extraction of DNA, some laboratories place small aliquots of DNA in several tubes so as to avoid repeated freezing and thawing of samples. Furthermore, it may be advantageous to put aliquots of DNA into tubes or on plates used for PCR at the time of extraction, and then to store the DNA at -20°C. In this way, when an assay is performed the samples are already in aliquots and the chances of contamination are reduced. Other methods of standardization may include making enough buffer, NTP mixes, primers, etc., for an entire study, so that consistency is achieved in the reagents. However, these too should be divided into aliquots.

The validation of an assay requires more than identifying the predicted base-pair length: a pseudogene might be identical in length. Validation can be accomplished through direct sequencing of the PCR product for homozygous individuals and comparison with published sequences. However, if the PCR product contains another polymorphic site that causes a shift in the sequence (i.e. a nucleotide deletion), the sequencing methods cannot be used unless the PCR fragment alleles are cloned first. An alternative approach would be to perform the assay for members of families and prove Mendelian inheritance. Large families are needed, with multiple generations, and these are available commercially (Coriell Institute, Camden, NJ). This has the advantage of allowing the investigator to practise on a large number of samples and learn to interpret the gels before valuable field samples are used.

Positive controls are important. A known standard for each informative genotype should be included in each experiment to verify that the assay is working properly. Further, possible inhibitors in the sample itself or highly degraded sample DNA may lead to false negative results, which would be suggested if a positive control amplifies but the samples do not.

Problems for the PCR include false priming, which results in the amplification of a non-specific product. One cause is a low annealing temperature, so that the primers anneal to genes with similar homology. Hot-start methods or raising the annealing temperature helps to exclude this problem. The simplest approach for hot starts is the addition of the polymerase after template denaturation. This method is not recommended, however, for large numbers of samples, because the chance of cross-contamination is increased by handling hot tubes (condensation at the walls and cap). Another approach for hot starts, which is time-consuming, is that of separating the polymerase and reaction mixture by a wax barrier, which liquefies after template denaturation and allows the reagents to mix (Sparkman, 1992). Alternativly, it is possible to use a Taq™ polymerase bound to an antibody so as to release the active enzyme after completion of template denaturation.

PCR mistakes can be based on polymerase errors. Further, inter- and intramolecular homology of primers can result in a primer-dimer product of similar size. Primer-dimers are template-independent, duplex PCR products (50-60 bp) composed of the extension added to one primer using the second primer as template. Once initiated, primer-dimer products are amplified very efficiently and limit the amplification of the target DNA. The complementarity between the two 3′ ends of the primer set has been shown to enhance primer-dimer formation products. On the other hand, the presence or absence of primer-dimers can be used as indicator of Taq™ polymerase function and whether the source DNA contains inhibitory substances.

Independently of the above sources of misclassification, another problem can arise when attempts are made to amplify sequences that are members of a related gene family or for which one or more pseudogenes are present in the genome. The latter is often the case in studies on metabolic polymorphisms. It is possible for such artefacts to be generated when primers are not unique. Statistically, primers need to have at least 17 bases to be unique in the human genome (Thein & Wallace, 1986) but they must also be designed to anneal with regions that are unique among the related sequences. If such primers cannot be found, specific amplification can often be achieved by using primers that have at least three unique bases at their 3′ end. Furthermore, products of similar genes or pseudogenes can frequently be distinguished by means of restriction enzyme digests. The specificity of different primer pairs can therefore be easily controlled by digestion of the products with restriction enzymes. An alternative approach to excluding the amplification of homologous genes is to predigest the template DNA before the PCR and thus eliminate amplification from the allele retaining the restriction site.

The correct choice of gel for the detection of amplified or enzymatically digested fragments is important for increasing the accuracy of the assay. The matrix must be adequate to achieve sufficient resolution of PCR products. We have found that polyacrylamide gels are among the best for resolution but the most cumbersome with which to work. NuSieve (FMC Products, Rockland, ME) or Metaphor (FMC Products, Rockland, ME) gels are useful for the detection of small fragments and fragments that need to be resolved and have similar base-pair lengths (i.e. a VNTR of 4-16 bases).

The interpretation of assay results and the reading of gels requires experience. Typically, amplified fragments have different intensities after ethidium bromide staining, and shorter fragments are less intense than longer ones because less dye is taken up. Also, heterozygote bands are half as intense as homozygote bands. Thus, each sample needs to be interpreted in the context of the other samples on the gel and the positive controls. For example, we do not report results for a sample where there is an upper band and no lower band if there is not another sample on the same row which has a less intense upper band and a lower band. When reading gels, our laboratory uses two independent blind reviewers. The first reads the results and enters the data in a database. The second then reads the gels, and the first cross-checks the database. Repeats are done for discrepant results, and if there are not two results that are clearly interpretable and agree the sample is eliminated from the analysis. We do repeats on either 20% or 100% of the samples, depending on the difficulty of the assay or the quality of the DNA; 100% is used for the more difficult assays.

References

Anker, P., Stroun, M. & Maurice, P.A. (1975) Spontaneous release of DNA by human blood lymphoctyes as shown in an *in vitro* system. *Cancer Res.*, 35, 2375-2382

Barnes, W.M. (1994) PCR amplification of up to 35-kb DNA with high fidelity and high yield from lambda bacteriophage templates. *Proc. Natl. Acad.. Sci. USA*, 91, 2216-2220

Bertilsson, L., Alm, C., De Las Carreras, C., Widen, J., Edman, G. & Schalling, D. (1989) Debrisoquine hydroxylation polymorphism and personality. *Lancet*,1, 555

Botstein, D., White, R.L., Skolnick, M. & Davis, R.W. (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.*, 32, 314-331

Broly, F., Marez, Millecamps, S., I Meyer, U.A. (19 detection of kno CYP2D6 gene u tion polymorphi 5, 373-384

Budowle, B. & Ba to improve the e ment length poly *Electrophor.*, 1, 181

Chan, C.Y., Lee, S C.L. & Lo, K.J. (DNA assay and virus RNA in Chi chronic hepatiti 443-446

Chehab, F.F., Dol Cooper, S. & Ru sickle cell anaem 329, 293-294 [er

Chehab, F.F. & K specific DNA seq fication: a colour *Natl. Acad. Sci. U*

Cheng, S., Chan (1994a) Long PC

Cheng, S., Fockle R. (1994b) Effect gets from cloned DNA. *Proc. Natl.*

Conner, B.J., Re K., Teplitz, R. Detection of sick hybridization wi *Proc. Natl. Acad.*

Cosma, G., Crol Toniolo, P. & G. ences in restricti phisms and mess human CYP1A *Biomarkers Prev.*,

Broly, F., Marez, D., Sabbagh, N., Legrand, M., Millecamps, S., Lo Guidice, J.M., Boone, P. & Meyer, U.A. (1995) An efficient strategy for detection of known and new mutations of the CYP2D6 gene using single strand conformation polymorphism analysis. *Pharmacogenetics*, 5, 373-384

Budowle, B. & Baechtel, F.S. (1990) Modifications to improve the effectiveness of restriction fragment length polymorphism typing. *Appl. Theor. Electrophor.*, 1, 181-187

Chan, C.Y., Lee, S.D., Hwang, S.J., Lu, R.H., Lu, C.L. & Lo, K.J. (1995) Quantitative branched DNA assay and genotyping for hepatitis C virus RNA in Chinese patients with acute and chronic hepatitis. *Chin. J. Infect. Dis.*, 171, 443-446

Chehab, F.F., Doherty, M., Cai, S.P., Kan, Y.W., Cooper, S. & Rubin, E.M. (1987) Detection of sickle cell anaemia and thalassaemias. *Nature*, 329, 293-294 [erratum in *Nature*, 329, 678]

Chehab, F.F. & Kan, Y.W. (1989) Detection of specific DNA sequences by fluorescence amplification: a colour complementation assay. *Proc. Natl. Acad. Sci. USA*, 86, 9178-9182

Cheng, S., Chang, S.Y., Gravitt, P. & Respess, R. (1994a) Long PCR. *Nature*, 369, 684-685

Cheng, S., Fockler, C., Barnes, W.M. & Higuchi, R. (1994b) Effective amplification of long targets from cloned inserts and human genomic DNA. *Proc. Natl. Acad. Sci. USA*, 91, 5695-5699

Conner, B.J., Reyes, A.A., Morin, C., Itakura, K., Teplitz, R.L & Wallace, R.B. (1983) Detection of sickle cell beta S-globin allele by hybridization with synthetic oligonucleotides. *Proc. Natl. Acad. Sci. USA*, 80, 278-282

Cosma, G., Crofts, F., Currie, D., Wirgin, I., Toniolo, P. & Garte, S.J. (1993) Racial differences in restriction fragment length polymorphisms and messenger RNA inducibility of the human CYP1A1 gene. *Cancer Epidemiol. Biomarkers Prev.*, 2, 53-57

Day, I.N. & Humphries, S.E. (1994) Electrophoresis for genotyping: microtiter array diagonal gel electrophoresis on horizontal polyacrylamide gels, hydrolink, or agarose. *Anal. Biochem.*, 222 (2), 389-395

De Lamballerie, X., Zandotti, C., Vignoli, C., Bollet, C. & de Micco, P. (1992) A one-step microbial DNA extraction method using "Chelex 100" suitable for gene amplification. *Res. Microbiol.*, 143, 785-790

Donis-Keller, H., Green, P., Helms, C., Cartinhour, S., Weiffenbach, B., Stephens, K., Keith, T.P., Bowden, D.W., Smith, D.R., Lander, E.S., Botstein, D., Akots, G., Rediker, K.S., Gravius, T., Brown, V.A., Rising, M.B., Parker, C., Powers, J.A., Watt, D.E., Kauffman, E.R., Bricker, A., Phipps, P., Muller-Kahle, H., Fulton, T.R., Ng, S., Schumm, J.W., Braman, J.C., Knowlton, R.G., Barker, D.F., Crooks, S.M., Lincoln, S.E., Daly, M.J. & Abrahamson, J. (1987) A genetic linkage map of the human genome. *Cell*, 51, 319-337

Erlich, H.A., Gelfand, D. & Sninsky, J.J. (1991) Recent advances in the polymerase chain reaction. *Science*, 252, 1643-1651

Fearon, E.R. & Vogelstein, B. (1990) A genetic model for colorectal tumorigenesis. *Cell*, 61, 759-767

Fey, M.F., Pilkington, S.P., Summers, C. & Wainscoat, J.S. (1987) Molecular diagnosis of haematological disorders using DNA from stored bone marrow slides. *Br. J. Haematol.*, 67, 489-492

Forsthoefel, K.F., Papp, A.C., Snyder, P.J. & Prior, T.W. (1992) Optimization of DNA extraction from formalin-fixed tissue and its clinical application in Duchenne muscular dystrophy. *Am. J. Clin. Pathol.*, 98, 98-104

Garcia-Pacheco, J.M., Mantilla, P., Garcia-Olivares, E. & Manzano-Fernandez, M.N. (1995) Routine HLA DRB/DQB oligonucleotide typing by a non-radioactive dot-blot micromethod. *J. Immunol. Methods*, 180, 35-43