3

Tools for Molecular Genetic Epidemiology: A Comparison of MADGE Methodology with Other Systems

JOHN W. HOLLOWAY*, SHU YE AND IAN N.M. DAY

Human Genetics Research Division, University of Southampton, Duthie Building (808), Southampton General Hospital, Tremona Road, Southampton S016 6YD, U.K.

Introduction

The analysis of susceptibility loci for complex genetic diseases has become the focus of much activity in recent years. A key to the successful analysis of these disorders is the analysis of many single nucleotide polymorphisms (SNPs) in extensive population samples to identify DNA variants that are risk factors. As a result, efficient cost-effective methods are required for the typing of SNPs. In this review we present an overview of one such method, Microplate-Array Diagonal Gel Electrophoresis (MADGE), and compare it with a number of other methodologies for high throughput SNP typing

In disease gene mapping, an essential role is played by the typing of polymorphism between individuals. In the characterization of etiological genetic sites for polygenic disease traits, due to the nature of the genetic contribution to the disease, and thus to the methods of analysis, the number of polymorphic loci needing to be typed is extremely large. Common polymorphism is likely to underpin many common disease susceptibilities. Either guided by linkage studies or by functional hypotheses concerning specific genes, genetic variation in specific genes is examined by association either in family-based or case-control designs (Weeks and Lathrop, 1995).

There are two main limitations in the analysis of genetic susceptibility to common disease. The first is that for complex diseases, which have multiple disease-causing

*To whom correspondence may be addressed (J.W.Holloway@soton.ac.uk)

Abbreviations: ARMS, amplification refractory mutation detection system; ASO, allele specific oligonucleotide; cM, centimorgan; DGGE, denaturing gradient gel electrophoresis; DNA, deoxyribo-nucleic acid; ELISA, enzyme linked immunosorbent assay; FRET, fluorescence resonance energy transfer; HLA, human leukocyte antigen; Kb, kilobase; MADGE, microplate array diagonal gel electrophoresis; OLA, oligonucleotide ligation assay; PCR, polymerase chain reaction; RFLP, restriction fragment length polymorphism; SNP, single nucleotide polymorphism; SSLP, simple sequence length polymorphism; SSO, sequence specific oligonucleotide; TGGE, temperature gradient gel electrophoresis.

Biotechnology and Genetic Engineering Reviews – Vol. 17, August 2000
0264–8725/00/17/71–88 $20.00 + $0.00

loci each of small effect, the number of subjects needing to be analysed to identify disease-susceptibility loci can be very large. The second main limitation is the speed and cost at which genetic markers can be analysed. It has been shown that, for complex diseases, the number of subjects needing to be collected to identify these genes by association than by linkage can be much smaller, especially for genes of small effect (Risch and Merikangas, 1996). In this respect, association studies are more powerful than linkage studies. However, association studies have one major limitation in that they rely on the polymorphism being tested having direct functional effects or for it to be in linkage disequilibrium with a polymorphism that does. Therefore, as linkage disequilibrium extends over relatively small genetic distances, a polymorphism in (or in strong disequilibrium with) the actual gene or genes involved in the disease must be typed before association can be seen. Thus, while linkage analysis is presently undertaken using panels of several hundred microsatellite markers (Dib *et al.*, 1996), at least several thousand markers would need to be typed to perform genome-wide association analysis. Given that the cost and technical difficulties of typing such a number of markers may well be outweighed by the cost and practical difficulties of recruiting enough subjects to undertake analysis by linkage, genome-wide association analysis may, in the future, be preferable.

For this to be possible, a large number of new genetic markers need to be identified and mapped. SNPs are major contributors to genetic variation and account for approximately 80% of all known polymorphisms in the human genome. They are particularly useful as gene mapping markers as they are more frequent than microsatellite markers, providing markers near to or in the locus of interest, and some SNPs will be located within the gene and can be expected to directly influence protein structure or expression levels giving insights into disease mechanisms. In addition, although they are biallelic and therefore less informative compared with sequence length variants, this also makes them mutationally more stable and the two alleles give a simple plus/minus assay rather than a length measurement, making them more amenable to automation. A recent analysis has shown that a genome scan using biallelic markers at a map density of 1 cM (about 3,000 markers) yields a much higher information content than the standard 10 cM microsatellite marker density currently used in genome scans (Kruglyak, 1997). In these and other contexts, it can be confidently predicted that there will be an extensive need for population studies of SNPs (in gene, genomic region, or genome-wide panels) over the next few years.

For it to be practicably possible to undertake such studies, several things need to be achieved. Firstly, a large number of SNPs need to be identified and mapped. Steps towards this are being made with the establishment of SNP databases such as dbSNP (http://www.ncbi.nlm.nih.gov/SNP/) and the use of sequence information in databases to identify SNPs *in silico* by sequence alignments (Wang *et al.*, 1998; Buetow *et al.*, 1999; Picoult-Newberg *et al.*, 1999). Furthermore, a consortium of 10 pharmaceutical companies led by the Wellcome Trust have entered into a project to identify and map 300,000 SNPs in the next two years (see http://snp.cshl.org/ and also *Nature* (news article) **398**, 545–546, 1999). Secondly, if an SNP panel is to replace microsatellite panels as informative markers in linkage studies, cost effective SNP genotyping methods will need to be developed allowing extremely high throughputs at low cost.

Prior to DNA sequencing, PCR and direct genotyping, many association studies between the HLA gene region and specific diseases were undertaken by serological,

biochemical and cellular definition of allelic variants in the HLA region (Dyer and Middleton, 1993). The recognition of restriction fragment length polymorphisms (RFLPs) represented the earliest phase of examination of single nucleotide polymorphisms (Lander and Botstein, 1986), and initially these could only be identified and typed by the laborious and low throughput procedure of Southern blotting using labelled probes derived from specific gene clones. With the advent of PCR (Mullis *et al.*, 1986), it has become feasible to use sequence specific labelled oligonucleotide hybridization (Wallace *et al.*, 1981) to PCR amplicons to determine genotype at SNP sites and other small insertion/deletion polymorphisms. By appropriate siting of PCR oligonucleotides, it was also possible to determine SNP genotype by allele specific PCR, eg ARMS-PCR (amplification refractory mutation system) (Newton *et al.*, 1989), or to induce a restriction site and hence artificial RFLPs at SNP sites where the natural sequence was refractory to RFLP analysis through lack of a restriction enzyme recognition site.

At least one new principle of SNP analysis has recently emerged, namely direct determination of exact mass of short single-strand DNA fragments by mass spectrometry to make base calls (Nordhoff *et al.*, 1993), and a variety of approaches using old principles have been developed in academic and commercial laboratories. One such method is Microplate-Array Diagonal Gel Electrophoresis (MADGE): an electrophoresis system that allows for high resolution and high throughput analysis of PCR products at low cost.

Microplate-array diagonal gel electrophoresis (MADGE)

MADGE was developed to serve the needs of human molecular genetic epidemiology studies (Day and Humphries, 1994), to provide a simple, high-throughput, low-cost method for SNP genotyping. Traditionally, SNP genotyping has been undertaken by a number of methods such as PCR-RFLP or allele specific PCR that require the resolution of PCR products by electrophoresis. However, this approach has been a major limiting step in improving throughput and cost efficiency of SNP analysis. The problems of gel preparation, long tracks (requiring more cumbersome equipment and longer run times), incompatibility with industry standard 96-well microplates, and the choice between low resolution agarose or high resolution but laborious to prepare vertical polyacrylamide gels, have led to the development of other techniques to type SNPs avoiding electrophoresis (see below).

However, it remains that electrophoresis still has many advantages over other systems, both in analysis of size, shape and charge of molecules. Electrophoresis is also well known and widely available and requires no expensive hardware. In addition, a range of well-established methodologies use electrophoresis (allele specific PCR, RFLP, heteroduplex analysis, denaturing gradient gel electrophoresis (DGGE), single-strand conformational polymorphism (SSCP), etc.). Therefore, MADGE was designed to overcome the difficulties of electrophoretic analysis by reducing it to short track lengths, coupled with a high density of wells in an industry standard format for ease of loading. Additionally, MADGE simplifies gel preparation and data acquisition using software. Thus, MADGE acts as a simple interface between liquid phase microplate procedures and electrophoresis.

In its simplest format (*Figure 3.1*), MADGE utilizes a two piece gel former

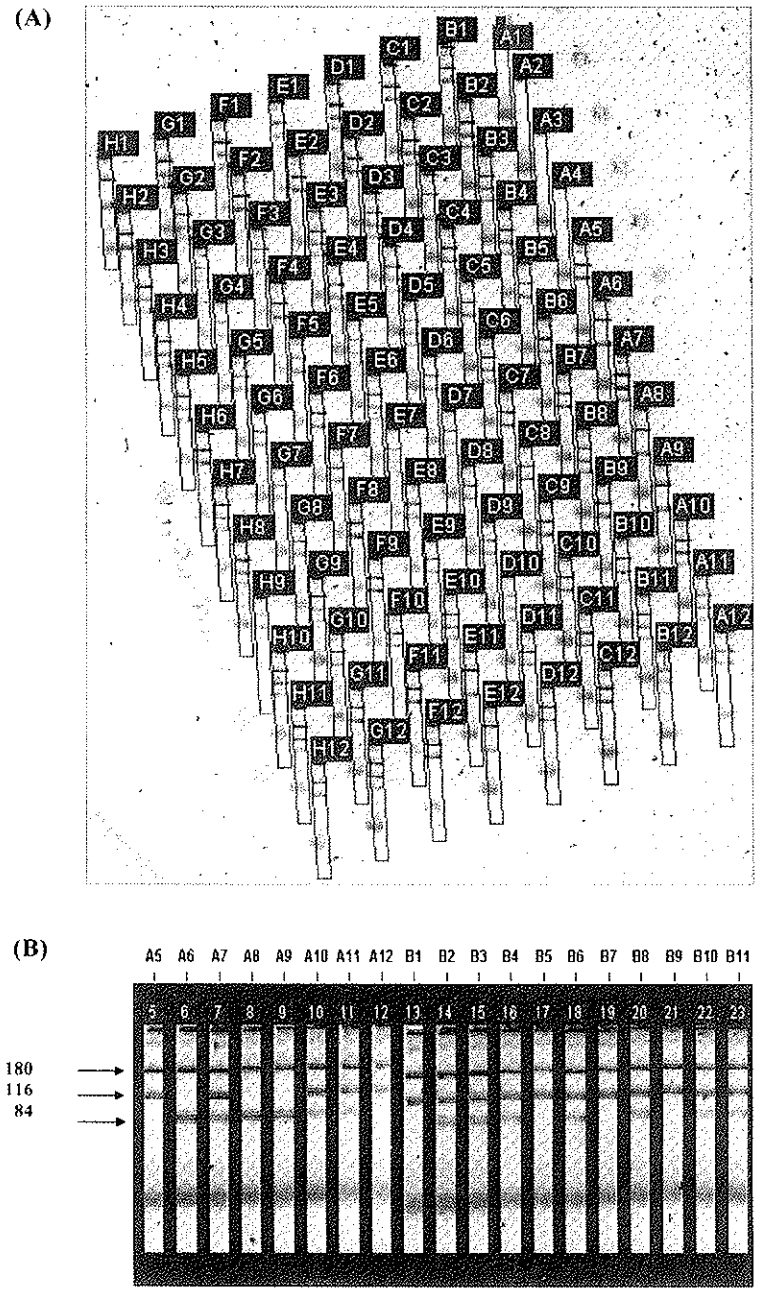

Figure 3.1. RFLP-MADGE Analysis using Phoretix 1D. Panel (A) illustrates a 96-well RFLP-MADGE gel stained with vistra green and analysed using the Fluorimager 595. Phoretix 1D™ software was used to enhance the image, track lanes and genotype the polymorphism. Panel (B) shows a transformed image of the gel lanes to enable easy allele discrimination. Examples of each genotype of a polymorphism in the Leukotriene C_4 synthase gene can be seen, homozygous A/A (lane A5), homozygous C/C (lane A6) and heterozygous A/C (lane A7) corresponding to the SNP generating a MspI site additional to an endogenous MspI site. DNA fragment sizes are shown for reference.

consisting of a plastic former and a glass plate coated with 'sticky saline' (γ-methacryloxypropyltrimethoxy-saline). The plastic former is a piece of plastic containing a 2 mm deep 100 × 500 mm rectangular 'swimming pool' within which there are 96 2 mm cubic teeth (well formers) in an 8 × 12 array (with a 9 mm pitch) directly compatible with 96-well microplates. The array is set on an angle of 71.6° to the long side of the 'pool', which is parallel to the direction of electrophoresis, giving tracks that can pass between two succeeding rows of wells (2 × 4 diagonal). In effect, this is a three-dimensional gel comb in which an acrylamide mix can be poured and overlaid with the silanized glass plate. Upon polymerization, the silanized glass plate is lifted with the set gel adhered. This gives an open-faced gel with ninety-six 2 mm cubic wells. Electrophoresis on the 2 × 4 diagonal gives a track length of 26.5 mm, with longer track lengths being achieved by increasing the angle at a cost of narrower wells (Day and Humphries, 1994).

Generally, 1–5 μl of PCR product is loaded into the wells with the 96-well microplate format of the MADGE gel facilitating sample loading using multi-channel (8,12 or 96) pipettors. Generally, MADGE gels are run in standard agarose electrophoresis tanks submerged in buffer with electrophoresis running time being less than 20 minutes. For visualization, MADGE gels are pre- or post-stained with any nucleic acid intercalating dye such as Ethidium Bromide or Vistra Green and imaged on a UV transilluminator using a CCD camera or a fluorimager such as the Molecular Dynamics Fluorimager 595. These gel images can easily be read by eye, although software is available to automate this process (http://www.phoretix.com/menu.htm) (see *Figure 3.1*).

Thus, the advantages of MADGE include being robust and easy to handle due to the glass backing and palm size; the good thermal transfer properties of the glass; the gels are directly stackable for storage and electrophoresis; and the gels can be re-used. In addition, the adherence to the 96-well format ensures compatibility with other laboratory equipment from PCR machines to pipettors, allowing fast and accurate transfer of samples from 96- (or 384-) well PCR microplate directly to the gel without sample re-coding.

Applications of MADGE

MADGE has many applications, including the typing of simple sequence length polymorphisms (SSLPs), the typing of SNPs using PCR-RFLP or allele specific (ARMS) PCR or heteroduplex analysis, and the identification typing of novel polymorphism through the use of temporal-thermal-ramp electrophoresis MADGE (melt-MADGE). Using 7.5% polyacrylamide for the gel matrix, 5% mobility differences can be resolved. This opens the way to the analysis of small deletions or insertion polymorphisms in relatively small amplicons. This approach has been taken in typing an insertion deletion polymorphism in intron 16 of the ACE gene (O'Dell *et al.*, 1995). The ACE gene sequence encompassing the intron 16 insertion/deletion polymorphism is amplified by PCR using two oligonucleotide primers, (ACE1 and ACE3) outside the insertion sequence, and one primer (ACE2) inside the sequence (Evans *et al.*, 1994). The 84 and 65 bp products indicating the presence of deletion (D) and insertion (I) alleles respectively are clearly resolved after electrophoresis on a 7.5% polyacrylamide MADGE gel. The use of alternative gel matrices such as

hydrolink (FMC bioproducts) and an increased angle of electrophoresis giving a longer track length enables the separation of mobility differences as small as 2%, allowing the analysis of minisatellite and some microsatellite polymorphisms (O'Dell *et al.*, unpublished observations).

In combination with PCR amplification, restriction enzyme analysis forms a rapid and general technique to identify gene lesions that change a restriction enzyme site. In general, the combined PCR-restriction enzyme fragment length polymorphism (PCR-RFLP) method is a two-step process. Firstly, a set of specific primers is annealed to the target DNA and amplified to give a specific PCR product, ie one single band in analytical gel electrophoresis. Secondly, after the PCR product has been digested with a known restriction enzyme and subjected to electrophoresis, a specific pattern of bands appears on the gel. The specificity of DNA cleavage by restriction endonucleases provides an easily scored difference in the pattern of fragments if a recognition site for a specific endonuclease is created or destroyed due to the presence or absence of a particular allele. However, approximately 50% of SNPs do not alter restriction sites, this may be overcome using mismatched oligonucleotides which introduce a restriction site into the PCR product utilizing a specific mismatch (Haliassos *et al.*, 1989). RFLP-MADGE has been successfully used to undertake the typing of several SNPs in large population samples, an example is the analysis of an Apa I polymorphism of the IGF1 gene (O'Dell *et al.*, 1997).

The Amplification Refractory Mutation System (ARMS) or allele specific PCR is a simple, rapid and reliable method for detection of point mutations or small deletions in a DNA test sample (Newton *et al.*, 1989). Following optimization, the ARMS technique has been shown to be both specific and sensitive and is often used in situations where high quality and accurate results are essential, such as in the clinical diagnosis of inherited mutations and DNA-based tissue typing. ARMS, although requiring attention to detail in initial set-up, can be used for multiplex reactions and in determining haplotype if the SNPs are in close proximity (Ferrie *et al.*, 1992).

An ARMS test for a DNA polymorphism consists of two complementary PCR reactions. The first contains an ARMS primer whose 3' terminal base is complementary to the normal variant and will not amplify the target containing the mutant variant. The second reaction contains the mutant specific primer and will not amplify normal DNA. Both reactions contain a common reverse primer, yielding the same size amplicon. Both reactions are performed on the same test DNA and the presence or absence of PCR products in the normal and mutant reactions indicates the genotype of the test sample. The use of an extra destabilizing mismatch at the penultimate base improves specificity by preventing mis-priming from the non-target sequence, while allowing efficient priming from the target. As the sample genotype is inferred from the presence or absence of a reaction product, it is essential to include a PCR control reaction, which amplifies a separate section of the test sample and gives rise to an amplicon of larger size than the ARMS product. Allele typing by ARMS lends itself readily to the MADGE format with allele scoring simply relying on a short electrophoresis to resolve the control amplicon from the allele specific amplicon. Throughput of ARMS typing on the MADGE system has also recently been improved by the availability of gel formers containing 192 wells comprised of two overlaid sets of 96 wells. This allows the electrophoresis of both allele specific reactions of the ARMS test end by end, improving not only throughput but also ease of analysis (*Figure 3.2*).

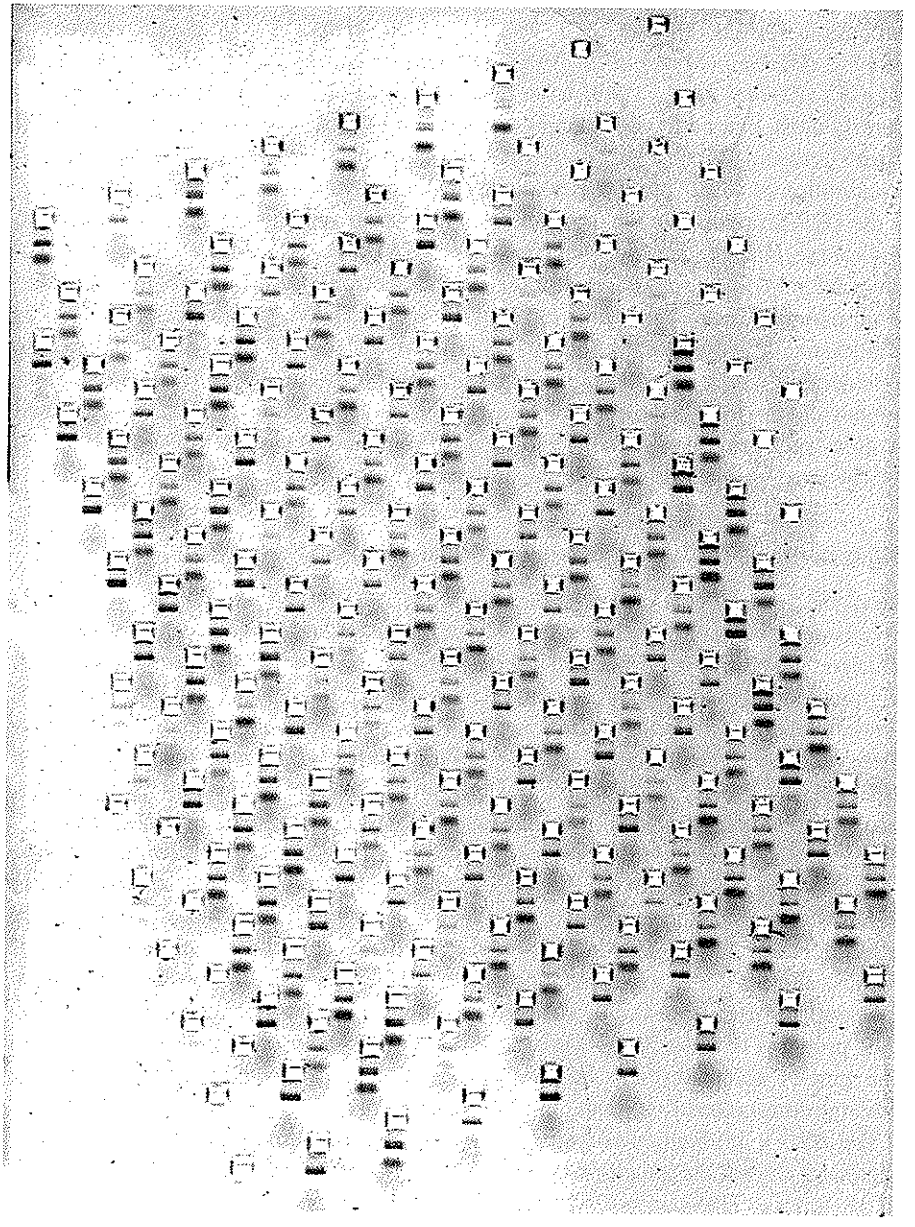

Figure 3.2. A 192-well ARMS-MADGE gel stained with ethidium bromide and analysed using the Fluorimager 595. The 192-well ARMS-MADGE allows end-by-end electrophoresis of both allele specific reactions of an ARMS test. Shown here is an ARMS test for a single nucleotide polymorphism of the Interleukin 4 receptor alpha chain (Q576R). As illustrated in *Figure 3.1*, Phoretix 1D™ software can be used to track lanes and transform the image of the gel lanes allowing easy comparison of the two allele specific reactions for genotyping.

Temporal-thermal-ramp electrophoresis MADGE (melt-MADGE) is a technique that allows both *de novo* mutation scanning coupled with allele typing in a high throughput format (Day *et al.*, 1998). Previously, denaturing gradient gel electrophoresis (DGGE) (Myers *et al.*, 1987) and temperature gradient gel electrophoresis (TGGE) (Riesner *et al.*, 1989) have been shown to be sensitive techniques able to identify most, if not all, base changes. These techniques use a temperature or chemical spatial gradient down a gel to resolve PCR products of differing sequence. The denaturation of a PCR product is sequence dependent and denaturation alters mobility of the PCR product, causing a reduction in migration velocity. Thus, amplicons with single-base changes can be identified. Melt-MADGE replaces the spatial gradient down the gel with a temporal thermal ramp, the electrophoresis buffer being increased in temperature during the course of electrophoresis. This gives several advantages in addition to allowing the use of the spatial array of wells in MADGE gels and therefore considerably improving throughput; programmable thermal ramps are more convenient and flexible than the use of gradient gels and also enable fast run times, as well as short track lengths. In addition to *de novo* mutation scanning, melt-MADGE can also be used to type known SNPs that give migration changes in PCR amplicons.

Other methods for higher-throughput analysis of SNP markers

Given the need for cost-effective, high-throughput SNP typing, a number of methods have been developed to achieve this. However, most laboratories still rely on three effective, generally accessible, and in-house configurable, principles: sequence specific oligonucleotide binding to one strand of a denatured duplex such as sequence specific oligonucleotide (SSO) hybridization, TaqMan™ or molecular beacons; the recognition of specific sequence directly within the duplex by a restriction enzyme as in RFLP-MADGE; or the exploitation of a DNA polymerase or ligase with a priming, incorporation or ligation discrimination as in ARMS, minisequencing (Syvanen *et al.*, 1990) or oligonucleotide ligation assay (eg Baron *et al.*, 1996) (see *Table 3.1*).

SSO hybridization utilizes the stability differences of short oligonucleotides to perfectly matched and mismatched target DNA sequences (Wallace *et al.*, 1979; Wallace *et al.*, 1981). The PCR products are denatured to separate the two DNA strands, blotted onto nylon membranes and hybridized with the labelled oligonucleotide probes, followed by stringent washing so that a probe mismatched with the PCR products is washed off, while a perfectly matched probe remains on the filter. The probes are commonly labelled with $\gamma^{32}P$ and detected by autoradiography, although non-radioactive procedures can also be used.

The SSO hybridization technique has now been revolutionized by the advent of microarray hybridization chips, which use DNA samples or oligonucleotides arrayed on the surface. These arrays have obvious applications in high throughput screening. Multiple DNA samples can be arrayed allowing the rapid screening of many patient samples for the presence of a few disease-causing alleles. For example, the multiplexed allele specific diagnostic assay (MASDA) uses multiplex DNA samples immobilized to the array, and a single hybridization is performed with a pool of allele-specific oligonucleotide probes. Any probes complementary to specific mutations present in a given sample are, in effect, affinity purified from the pool by the target DNA. Sequence-specific band patterns (fingerprints), generated by chemical or enzymatic

Table 3.1. Main features of SNP typing systems

Technique	Basis	Features	Applicability
ARMS-MADGE or RFLP-MADGE	RFLP, ARMS	Low capital and consumable costs, medium throughput, highly adaptable system	Single SNP or gene project small to medium sized studies (= 5,000 samples) MADGE is applicable to many other categories of electrophoretic analysis
TaqMan™	ASO	High throughput, high capital and consumable costs	Single SNPs in large populations (eg 10,000 samples) Non cost conscious projects
PCR-SSOP	ASO	Low consumable cost, medium capital costs Potentially high throughput	Single gene or gene region project (= 10 SNPs per sample) Reliant on effective PCR multiplexing or multiple SNPs per amplicon (eg MHC)
DNA Chips	ASO	Very high throughput. Application limited by availability of chips. High capital cost but low consumable cost (when used for very many (eg > 1,000) SNPs per sample)	Very large numbers of SNPs in larger populations Potential to replace microsatellite markers for genome scans Not suitably configured at present for single SNP or gene/gene region projects in cohort studies
Molecular Beacons	ASO	Closed tube like TaqMan™. Potentially high throughput but with high equipment and consumable costs. Potentially amenable to limited multiplexing	Single SNPs in large populations
DASH	ASO	Potentially high throughput but with high capital costs, not able to be multiplexed	
OLA	Ligation	Microtitre readout increases throughout through removal of electrophoresis step. Multiplexing difficult except for multiple SNPs in the same amplicon	Single SNPs in large populations
Invader Oligos	ASO	Potentially high throughput and very cost effective through avoidance of PCR	
Padlock Probes	Ligation	As above	
Minisequencing	Nucleotide Incorporation	Multiple forms with several readouts, by ELISA, by Mass Spec or by FRET. High throughput but limited capability for multiplexing	
Minisequencing Multiplex gel	Nucleotide Incorporation	Solid Phase + Gel	Automated Sequencer

sequencing of the bound probe(s), easily identify the specific mutation(s) (Shuber *et al.*, 1997). Oligonucleotide arrays (Yershov *et al.*, 1996) are applicable to the genotyping of many SNPs; the Whitehead Institute/Affymetrix microarray has been designed to genotype 2,000 SNPs. This, combined with improvements in the robustness of multiplex PCR involving 1,000s of amplicons, will lead toward the generation of genome-wide genotypic information. The hybridization of fluorescent PCR products is visualized by confocal microscopy; comparison with a control sample is used to determine genotype. For a detailed description of microarrays, see *Chapter 8* and multiple reviews in *Nature Genetics Supplement* 21, 1999.

TaqMan™ or the 5' nuclease assay is a method of allele discrimination that utilizes the 5' to 3' exonuclease activity of *Taq* polymerase and has been developed commercially by PE Applied Biosystems. In the 5' nuclease assay, a hybridization probe included in the PCR is cleaved by the inherent 5' nuclease activity of *Taq* polymerase (Holland *et al.*, 1991), only if the probe's target sequence is being amplified. The hybridization probe is double-labelled with fluorescent reporter and quencher dyes, and probe cleavage results in an increase in fluorescence emission intensity, due to fluorescence resonance energy transfer (FRET), from the reporter dye, which can be measured with a fluorimeter. Thus, the 5' nuclease assay achieves amplification and detection of specific PCR products in a single step.

Lee *et al.* (1993) first demonstrated that the 5' nuclease assay could be used for allelic discrimination. In an allelic discrimination assay, two double-labelled fluorescent probes are included in the PCR reaction, one probe specific for each allele. The probes can be distinguished through the use of different fluorescent reporter dyes (usually FAM and TET). A mismatch between probe and target greatly reduces the efficiency of probe hybridization and cleavage. Following PCR, an increase in the level of an FAM fluorescent signal without a rise in the TET specific signal indicates that only the FAM specific sequence (allele) was present and that the sample is homozygous, and vice versa. An increase in both reporter signals indicates homozygosity.

Molecular Beacons (Tyagi and Kramer, 1996; Tyagi *et al.*, 1998; Marras *et al.*, 1999) are oligonucleotide hybridization probes that have two complementary DNA sequences flanking the target sequence and are labelled at one end with a donor dye and at the other end with an acceptor dye. In the absence of hybridization to the target sequence, the donor/acceptor dyes are in close proximity due to the adoption of a hairpin loop conformation by the probe, which limits the fluorescent signal. When hybridized, the donor and acceptor dyes are separated sufficiently to increase the reporter fluorescence signal by 900-fold, thus alleles can be discriminated. Compared with methods based on FRET, such as TaqMan™, these donor and acceptor dyes do not need to have overlapping spectra due to their close proximity, therefore up to four different molecular beacons can be assayed in the same reaction.

Another technique that relies on oligonucleotide hybridization is dynamic allele-specific hybridization (DASH) (Howell *et al.*, 1999). In DASH, a target sequence is amplified by PCR in which one primer is biotinylated. The biotinylated product strand is bound to a streptavidin-coated microtitre plate well, and the non-biotinylated strand is removed. An oligonucleotide probe, specific for one allele, is hybridized to the target at low temperature. This forms a duplex DNA region that interacts with a double strand-specific intercalating dye. Upon excitation, the dye emits fluorescence

proportional to the amount of double-stranded DNA (probe-target duplex) present. The sample is then steadily heated while fluorescence is continually monitored. A rapid fall in fluorescence indicates the denaturing (or 'melting') temperature of the probe-target duplex. When performed under appropriate buffer and dye conditions, a single-base mismatch between the probe and the target results in a dramatic lowering of melting temperature (tm) that can be easily detected.

The oligonucleotide ligation assay (OLA) can be utilized for the genotyping of SNPs and insertions/deletions (Landegren *et al.*, 1988). The ligation assay is simple and determines whether or not two adjacent primers (usually ~ 20 mers) become covalently joined by a DNA ligase when hybridized to a complementary target DNA sequence (PCR amplicon or genomic DNA). Ligation is dependent on the hybridization of the primers within the target, the juxtaposition of the primers on the target, ie must lie directly next to each other in a 5' to 3' orientation with no interval and primers must have perfect complementarity with the target sequence. Even a single nucleotide mismatch between the primer and target will inhibit primer joining discriminating the SNP allele. Several methods have been used for visualization including hapten-labelled primers, eg digoxigenin and fluorescein followed by detection using hapten-specific antibodies labelled with alkaline phosphatase or horseradish peroxidase. This system allows multiplexing and enzyme-linked immunosorbent assays (ELISA) detection for high throughput genotyping (Tobe *et al.*, 1996). Alternatively, multiplexed ligation reactions can be visualized using electrophoresis on an automated DNA sequencer using fluorescent labelling (Day *et al.*, 1995). FRET has also been used for visualization with the dye-labelled oligonucleotide ligation (DOL) assay (Chen and Kwok, 1999).

Minisequencing involves the target dependent addition of a specific nucleotide to a single primer (Syvanen *et al.*, 1990) and can distinguish the nucleotides located immediately downstream of the primer. Multiple SNPs can be analysed by separating sequence specific primers by size using electrophoresis (Pastinen *et al.*, 1996). Minisequencing of SNPs can also be carried out using immobilized amplification products (Syvanen *et al.*, 1990) and using FRET detection (Chen and Kwok, 1999). An extension of minisequencing is pyrosequencing; the incorporation of a nucleotide is monitored by the release of inorganic pyrophosphate that fuels a luciferase reaction (Nyren *et al.*, 1993).

All of the SNP typing methods described above require the initial generation of a PCR product that is then interrogated to determine genotype. Therefore, one major limiting factor in throughput for SNP typing will be the rate at which PCR amplicons can be generated. A number of SNP typing methods have been developed recently which do not require pre-amplification of DNA target by PCR before interrogation, although these are not in routine use at present. One of these techniques is the use of 'padlock probes'. Padlock probes are oligonucleotides consisting of two target-complementary segments connected by a linker sequence. When the two target-specific segments bind to the target DNA region, the ends of the probe can be joined through the action of ligases creating circular DNA (Nilsson *et al.*, 1994), which is catenated to the genomic template. This method is attractive in its potential for the analysis of genomic DNA without the need for pre-amplification. Detection of these ligated probes has been problematic, although the amplification of the signal can be achieved by rolling circle replication (Baner *et al.*, 1998).

A second method for SNP typing without PCR amplification relies on the use of 'Invader' oligonucleotides (Lyamichev *et al.*, 1999). Invader and signal probes are designed to hybridize to single base pair overlapping sites on the target strand. The invader probe can displace part of the signal probe, which forms a structure that is susceptible to flap endonucleases (FENs) isolated from archaea, resulting in the generation of detectable products. By positioning the cleavage site at an SNP, it is possible to discriminate the allele depending on the sequence of the Invader probe. This technique has been used to develop an assay for the Factor V Leiden mutation (Ryan *et al.*, 1999). However, large template quantities are needed for the Invader assay, so usage at present often requires PCR amplification as a first step.

Comparison of systems for higher-throughput genotyping

This overview of current methods used for SNP analysis highlights the rapid progress that is being made in the development of cost-effective, high-throughput techniques. Increasingly sophisticated approaches to old principles of SNP typing are being developed in academic and commercial laboratories. When considering which methodology to use for high-throughput SNP analysis, several factors must be considered including; (a) is the particular method applicable to simultaneous analysis of multiple SNPs? (b) time considerations, how long will the assay require to optimize and then complete for the given sample set? (c) cost considerations, what is the most cost-effective method for genotyping a particular SNP or set of SNPs? (d) equipment considerations, is a large financial investment required in order to establish the technique?

Most laboratories settle with one methodology, with choice being determined by availability of reagents and equipment and by suitability of the assay for a particular polymorphism. However, as discussed above, it can be confidently predicted that there will be an extensive need for population studies of SNPs (in gene, genomic region, or genome-wide panels) over the next few years. This marked increase in the numbers of SNPs needing to be typed forces consideration of factors such as throughput and cost per SNP called.

Unfortunately, few comparisons have been made between the advantages and disadvantages of these different methods. Recently, we have analysed the throughput and cost-effectiveness of several commonly used methods for SNP typing (Holloway *et al.*, 1999). In this analysis, ARMS-MADGE, TaqMan™ and SSO typing were compared for both consumable and staff costs and throughput, with each system being in a format that would be considered 'typical' for that technique in laboratories with moderately high throughput requirements. The results of this analysis showed that these three methods differed very dramatically for all of the factors critical to choice of a particular SNP typing system.

MADGE is considered a moderately high throughput method (*Table 3.2*) and is applicable to limited multiplexing of reactions, eg ARMS-MADGE which is determined by the number of DNA bands that can be resolved on a short electrophoresis track (approximately 20–50). RFLP-MADGE, is not applicable to multiplexing, although throughput has been improved by a 384 format MADGE resulting in time reductions for the procedure (Hinks and Day, unpublished observations). However, it is very flexible, with many types of SNP assay adaptable to the system. In addition, the equipment and consumable costs associated with MADGE are low.

Table 3.2. Maximum throughput of three commonly used SNP typing systems

Method	Runs per day	Samples per week
ARMS-MADGE	in 1 × 96 or 4 × 96 format	2,700 samples per week
TaqMan™	5 runs per day (in 96 or 4 × 96-well format). 90 samples per plate	9,000 samples per week
PCR-SSOP HLA DRB1 typing	2 runs per week (in 96-well or 192-well format)	1,000 samples per week

Comparison of the maximum throughput of ARMS-MADGE with the TaqMan™ system and the PCR-SSOP systems. It is important to note that with PCR-SSOP typing it is possible to type 10 SNPs per sample in each run, assuming that there are multiple SNPs in each amplicon or PCR multiplexing has been used to give multiple templates for probe hybridization. This throughput analysis was based on one researcher working full-time in a standard molecular genetics laboratory with limited automation (Holloway *et al.*, 1999).

The TaqMan™ system was shown to have a very high throughput, limited only by the rate of PCR, this being due to this system's ability to determine genotype directly with the amplification and detection of specific PCR products occurring in a single step. However, the savings achieved with the reduction in staff time are outweighed by the high consumable costs associated with this technique. TaqMan™ was found to have a total cost per SNP called of £0.84–£0.92 for a 10,000 DNA sample typing project. This compares with £0.28 for an ARMS-MADGE call. The cost per sample for HLA DRB1 SSO typing was £3.17. It is important to note, however, that this method used 18 probes to call 15 alleles, therefore this cost is equivalent to calling 7.5 biallelic SNPs at a cost of £0.42 per SNP. With the development of suitable high-throughput multiplexing of PCR reactions allowing the co-amplification of 9 amplicons, 18 SSOPs would allow the typing of 9 biallelic SNPs at a cost of £0.35 per SNP called. Further multiplexing would allow the reduction of cost per SNP call for SSOP to be reduced still more, although with some increase in PCR costs. The TaqMan™ system in its current format is not applicable to multiplexing due to restrictions on fluorescent dye usage in FRET detection (see introduction). This limits the procedure to the examination of one SNP per analysis.

SSO hybridization has a much lower throughput in terms of samples (1,000 samples per week) than TaqMan™ and MADGE (9,000 and 2,700 samples per week respectively). In addition, it requires a large amount of technician's time. This makes it extremely unsuitable for high-throughput typing if it is used for the examination of a single SNP per assay, as it has been predominantly in the literature. However, as is shown by the case of HLA-DRB1 typing (Holloway *et al.*, 1999), where 18 probes are used to call 15 alleles, the cost per SNP typed drops from £3.17 to £0.42 per SNP. With the development of suitable high-throughput multiplexing of PCR reactions allowing the co-amplification of 9 amplicons, 18 SSO probes would allow the typing of 9 biallelic SNPs at a cost of £0.35 per SNP called. Further multiplexing would allow the reduction of cost per SNP call for SSOP to be reduced still more, although with some increase in PCR costs. In addition, if multiple SNPs are being typed per sample, then the achievable throughput for SSO probes using multiplex PCR is similar to, if not greater than, the TaqMan™ system.

Another important factor to be considered when choosing an allele typing system is accuracy of allele calling. Unfortunately, very few comparisons of genotyping

accuracy have been undertaken, although recent comparison of an allele specific PCR-MADGE with SSO and RFLP-MADGE methods for a polymorphism of the angiotensin II type-1 receptor showed not only that the allele specific PCR method was considerably higher throughput than SSO, but also had a greater accuracy (Hunt *et al.*, 1999). However, another comparison of a PCR ELISA SSO method and ARMS typing for Tumour Necrosis Factor polymorphisms found little difference in accuracy of genotyping between the methods, although ARMS PCR was still considerably higher throughput when only a single SNP was being assayed (Knight *et al.*, 1999).

Other methods described in this chapter are also particularly applicable to the analysis of multiple SNPs including molecular beacons and OLA. However, at the present time, the reliable multiplexing of many PCR amplicons together is still difficult. Further increments in throughput and reductions in cost are readily feasible by reduction in reagent/reaction volumes and increase in array density (eg to 384 and higher). In our view, a critical breakpoint arises at 384-well density. This is about the maximum density that human eye and hand have the precision to use. Initially, throughput can be increased by the use of 'semi-automation' with human interventions (eg manual 96-channel pipetting), permitting rapid adaptability. However, at higher densities and with sub-microlitre reaction volumes, more radical engineering solutions may be needed (Kricka, 1998). These will incur greater equipment costs, which will be increased by deviation from 96 and n × 96 industry standards.

It is evident, therefore, that the analysis of multiple SNPs within a single gene or gene region can be handled readily within the standard laboratory by simply improving the throughput of commonly used assays, whether this be achieved by using MADGE to speed electrophoresis throughput of ARMS or RFLP assays or the use of techniques such as TaqMan™ or OLA to avoid electrophoresis all together. However, in the future (as seems likely), whole genome scans are to be undertaken using SNPs and linkage disequilibrium mapping, an increase of one or two orders of magnitude and similar cost reduction will be necessary to achieve this within a practicable time frame and acceptable costs.

In a hypothetical project of typing 3,000 SNP markers in a DNA bank of 3,000 samples, we have compared the resources needed to complete this for several techniques (Holloway *et al.*, 1999). This project is approximately what will be required to undertake genome-wide screens for susceptibility loci for complex diseases using linkage (*Table 3.3*). In this analysis we also included a comparison with an allele typing technique not yet in widespread use, namely DNA chips. The

Table 3.3. Costs and time scale of 3,000 SNPs by 3,000 samples, 'one researcher' project*

Method	Time	Cost
TaqMan™	19.2 years	£7.56– £8.28 million
PCR-SSOP†	19.2 years	£3.15 million
ARMS-MADGE	64.1 years	£2.52 million
DNA Chips	50 days	£3.06 million

*Figures refer to the cost and time of completing 3,000 SNPs typed in 3,000 samples by one researcher project given the maximum throughput and costs of current SNP typing techniques.
†PCR-SSOP typing based on 9 biallelic SNPs called per SSOP run, assuming multiple SNPs per amplicon or multiplexing.
A comparison of the time scale needed and the costs of a hypothetical project needed to complete a genome scan for disease susceptibility genes using linkage disequilibrium mapping with SNP markers (Holloway *et al.*, 1999).

analysis of cost and throughput for DNA chips is based on the current throughput and cost of the Affymetrix GeneChip® HuSNP™ Mapping Assay containing 1,500 SNP markers using multiplex amplification of approximately 100 different SNP markers in a single PCR reaction (http://www.affymetrix.com/products/HuSNP.html). Costs were extrapolated from 1,500 to 3,000 SNPs by estimating a cost per SNP call and multiplying by 3,000. The list price for the HuSNP2K chip in Europe will be US$864 (£510.56) for the array and reagents, including all the pooled PCR primers; this equates to a cost of £0.34 per genotype called. For the purpose of this 'one researcher' analysis, it was assumed that one technician working a standard 8-hour day can run the system at maximum throughput, including undertaking the PCR reactions. While there are currently only 1,500 SNP markers on the available SNP chips, other DNA chip applications, such as cDNA expression analysis, already use chips of greater than 3,000 marker density, so throughput was calculated assuming a 3,000 SNP chip, giving a total throughput of 60 samples per day. The maximum throughput for the chip scanner is 180 samples in 24 continuous hours. However, the fluidics station (for chip hybridization) only supports approximately 50 samples per 24 hours, but throughput can be scaled by adding additional stations.

From this analysis, it is clear that the only technique that can currently achieve this type of project within a realistic time frame is microarrays, although the cost of this approach is still similar to more traditional methods. In addition, using these 'SNPs on chips' is restricted to defined SNPs which, at this time, are chosen by the manufacturer. Therefore, for a random genome analysis of SNPs, microarrays are the method of choice but, for more defined projects (eg a set of SNPs representing diversity in a gene, genomic region or functional pathway), MADGE, TaqMan™ and SSO hybridization are more applicable.

Despite the development of all of these techniques, there is still a need for improved throughput, although high specificity of allele discrimination must be maintained. On one in which this could be achieved would be through the use of methods that do not require pre-amplification of the target regions, such as Padlock probes or Invader probes. Although still being developed, these techniques show great promise and are likely to be very cost effective due to the absence of a PCR stage.

All of these considerations make the choice of SNP analysis method a difficult one. For moderately high throughput with low cost, genomic context and ease of set-up may determine the choice between MADGE and SSO hybridization for defined studies involving limited numbers of SNPs. For large genome-wide studies, the microarray hybridization chips are likely to be the method of the future. At present, it appears to be a feasible goal to complete comprehensively one polygene analysis, or possibly one genomic region association analysis as a 'one researcher' project (eg 30–300 SNPs in 3,000 samples). The choice of typing system to complete such a project will depend on the resources in terms of labour, equipment and consumable costs available to the individual. It may be that an order of magnitude improvement to relatively simple systems such as these described here will prove sufficient for the analysis of selected genomic regions of interest by association, a model that for linkage has proven quite successful in the analysis of single gene disorders.

Acknowledgements

John Holloway is supported by grants from Genome Therapeutics Corporation to Professor Stephen Holgate. Shu Ye acknowledges British Heart Foundation grants PG/98192 and PG/98183. Ian Day is a Lister Institute Research Professor and thanks the MRC for grants G9516890MB, G9605150MB and G9828424. The Wessex Medical Trust is thanked for its support of the Human Genetics Research Division.

References

BANER, J., NILSSON, M., MENDEL-HARTVIG, M. AND LANDEGREN, U. (1998). Signal amplification of padlock probes by rolling circle replication. *Nucleic Acids Research* **26**, 5073–5078.

BARON, H., FUNG, S., AYDIN, A., BAHRING, S., LUFT, F.C. AND SCHUSTER, H. (1996). Oligonucleotide ligation assay (OLA) for the diagnosis of familial hypercholesterolemia. *Nature Biotechnology* **14**, 1279–1282.

BUETOW, K.H., EDMONSON, M.N. AND CASSIDY, A.B. (1999). Reliable identification of large numbers of candidate SNPs from public EST data. *Nature Genetics* **21**, 323–325.

CHEN, X. AND KWOK, P.Y. (1999). Homogeneous genotyping assays for single nucleotide polymorphisms with fluorescence resonance energy transfer detection. *Genetic Analysis* **14**, 157–163.

DAY, D.J., SPEISER, P.W., WHITE, P.C. AND BARANY, F. (1995). Detection of steroid 21-hydroxylase alleles using gene-specific PCR and a multiplexed ligation detection reaction. *Genomics* **29**, 152–162.

DAY, I.N., SPANAKIS, E., PALAMAND, D., WEAVIND, G.P. AND O'DELL, S.D. (1998). Microplate-array diagonal-gel electrophoresis (MADGE) and melt-MADGE: tools for molecular-genetic epidemiology. *Trends in Biotechnology* **16**, 287–290.

DAY, I.N.M. AND HUMPHRIES, S.E. (1994). Electrophoresis For Genotyping - Microtiter Array Diagonal Gel- Electrophoresis On Horizontal Polyacrylamide Gels, Hydrolink, or Agarose. *Analytical Biochemistry* **222**, 389–395.

DIB, C., FAURE, S., FIZAMES, C., SAMSON, D., DROUOT, N., VIGNAL, A., MILLASSEAU, P., MARC, S., HAZAN, J., SEBOUN, E., LATHROP, M., GYAPAY, G., MORISSETTE, J. AND WEISSENBACH, J. (1996). A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* **380** (6570), 152–154.

DYER, P. AND MIDDLETON, D., Eds. (1993). *Histocompatibility Testing: A Practical Approach.* Oxford: IRL Press at Oxford University Press.

EVANS, A.E., POIRIER, O., KEE, F., LECERF, L., MCCRUM, E., FALCONER, T., CRANE, J., O'ROURKE, D.F. AND CAMBIEN, F. (1994). Polymorphisms of the angiotensin-converting-enzyme gene in subjects who die from coronary heart-disease. *Quarterly Journal of Medicine* **87** (4), 211–214.

FERRIE, R.M., SCHWARZ, M.J., ROBERTSON, N.H., VAUDIN, S., SUPER, M., MALONE, G. AND LITTLE, S. (1992). Development, multiplexing, and application of ARMS tests for common mutations in the CFTR gene. *American Journal of Human Genetics* **51**, 251–262.

HALIASSOS, A., CHOMEL, J.C., GRANDJOUAN, S., KRUH, J., KAPLAN, J.C. AND KITZIS, A. (1989). Detection of minority point mutations by modified PCR technique: a new approach for a sensitive diagnosis of tumor-progression markers. *Nucleic Acids Research* **17**, 8093–8099.

HOLLAND, P.M., ABRAMSON, R.D., WATSON, R. AND GELFAND, D.H. (1991). Detection of specific polymerase chain reaction product by utilizing the 5'-3' exonuclease activity of Thermus aquaticus DNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 7276–7280.

HOLLOWAY, J.W., BEGHÉ, B., TURNER, S., HINKS, L.J., DAY, I.N.M. AND HOWELL, W.M. (1999). Comparison of three methods for single nucleotide polymorphism typing for DNA bank studies: Sequence-specific oligonucleotide probe hybridisation, TaqMan liquid phase hybridisation, and microplate array diagonal gel electrophoresis (MADGE). *Human Mutation* **14**, 340–347.

HOWELL, W.M., JOBS, M., GYLLENSTEN, U. AND BROOKES, A.J. (1999). Dynamic allele-specific hybridization. A new method for scoring single nucleotide polymorphisms. *Nature Biotechnology* **17**, 87–88.

HUNT, C.C.J., BURLEY, J.E., CHAPMAN, C.M.L. AND BEILBY, J.P. (1999). A high throughput MS-PCR methods on MADGE gels for ANG II type-1 receptor A1166C polymorphism. *Physiological Genomics* **1**, 71–73.

KNIGHT, J.C., MCGUIRE, W., KORTOK, M.M. AND KWIATKOWSKI, D. (1999). Accuracy of genotyping of single-nucleotide polymorphisms by PCR-ELISA allele-specific oligonucleotide hybridization typing and by amplification refractory mutation system. *Clinical Chemistry* **45**, 1860–1863.

KRICKA, L.J. (1998). Miniaturization of analytical systems. *Clinical Chemistry* **44** (9), 2008–2014.

KRUGLYAK, L. (1997). The use of a genetic map of biallelic markers in linkage studies. *Nature Genetics* **17**, 21–24.

LANDEGREN, U., KAISER, R., SANDERS, J. AND HOOD, L. (1988). A ligase-mediated gene detection technique. *Science* **241** (4869), 1077–1080.

LANDER, E.S. AND BOTSTEIN, D. (1986). Strategies for studying heterogeneous genetic traits in humans by using a linkage map of restriction fragment length polymorphisms. *Proceedings of the National Academy of Sciences of the United States of America* **83**, 7353–7357.

LEE, L.G., CONNELL, C.R. AND BLOCH, W. (1993). Allelic discrimination by nick-translation PCR with fluorogenic probes. *Nucleic Acids Research* **21**, 3761–3766.

LYAMICHEV, V., MAST, A.L., HALL, J.G., PRUDENT, J.R., KAISER, M.W., TAKOVA, T., KWIATKOWSKI, R.W., SANDER, T.J., DE ARRUDA, M., ARCO, D.A., NERI, B.P. AND BROW, M.A. (1999). Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes. *Nature Biotechnology* **17**, 292–296.

MARRAS, S.A., KRAMER, F.R. AND TYAGI, S. (1999). Multiplex detection of single-nucleotide variations using molecular beacons. *Genetic Analysis* **14**, 151–156.

MULLIS, K., FALOONA, F., SCHARF, S., SAIKI, R., HORN, G. AND ERLICH, H. (1986). Specific enzymatic amplification of DNA *in vitro*: the polymerase chain reaction. *Cold Spring Harbour Symposia on Quantitative Biology* **51 Pt 1**, 263–273.

MYERS, R.M., MANIATIS, T. AND LERMAN, L.S. (1987). Detection and localization of single base changes by denaturing gradient gel electrophoresis. *Methods in Enzymology* **155**, 501–527.

NEWTON, C.R., GRAHAM, A., HEPTINSTALL, L.E., POWELL, S.J., SUMMERS, C., KALSHEKER, N., SMITH, J.C. AND MARKHAM, A.F. (1989). Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Research* **17**, 2503–2516.

NILSSON, M., MALMGREN, H., SAMIOTAKI, M., KWIATKOWSKI, M., CHOWDHARY, B.P. AND LANDEGREN, U. (1994). Padlock probes: circularizing oligonucleotides for localized DNA detection. *Science* **265** (5181), 2085–2088.

NORDHOFF, E., CRAMER, R., KARAS, M., HILLENKAMP, F., KIRPEKAR, F., KRISTIANSEN, K. AND ROEPSTORFF, P. (1993). Ion stability of nucleic acids in infrared matrix-assisted laser desorption/ionization mass spectrometry. *Nucleic Acids Research* **21**, 3347–3357.

NYREN, P., PETTERSSON, B. AND UHLEN, M. (1993). Solid phase DNA minisequencing by an enzymatic luminometric inorganic pyrophosphate detection assay. *Analytical Biochemistry* **208**, 171–175.

O'DELL, S.D., HUMPHRIES, S.E. AND DAY, I.N. (1995). Rapid methods for population-scale analysis for gene polymorphisms: the ACE gene as an example. *British Heart Journal* **73**, 368–371.

O'DELL, S.D., MILLER, G.J., COOPER, J.A., HINDMARSH, P.C., PRINGLE, P.J., FORD, H., HUMPHRIES, S.E. AND DAY, I.N. (1997). Apal polymorphism in insulin-like growth factor II (IGF2) gene and weight in middle-aged males. *International Journal of Obesity and Related Metabolic Disorders* **21**, 822–825.

PASTINEN, T., PARTANEN, J. AND SYVANEN, A.C. (1996). Multiplex, fluorescent, solid-phase minisequencing for efficient screening of DNA sequence variation. *Clinical Chemistry* **42**, 1391–1397.

PICOULT-NEWBERG, L., IDEKER, T.E., POHL, M.G., TAYLOR, S.L., DONALDSON, M.A., NICKERSON, D.A. AND BOYCE-JACINO, M. (1999). Mining SNPs from EST databases. *Genome Research* **9**, 167–174.

RIESNER, D., STEGER, G., ZIMMAT, R., OWENS, R.A., WAGENHOFER, M., HILLEN, W., VOLLBACH, S. AND HENCO, K. (1989). Temperature-gradient gel electrophoresis of nucleic acids: analysis of conformational transitions, sequence variations, and protein-nucleic acid interactions. *Electrophoresis* **10**, 377–389.

RISCH, N. AND MERIKANGAS, K. (1996). The future of genetic studies of complex human diseases. *Science* **273** (5281), 1516–1517.

RYAN, D., NUCCIE, B. AND ARVAN, D. (1999). Non-PCR-dependent detection of the factor V Leiden mutation from genomic DNA using a homogeneous invader microtiter plate assay. *Molecular Diagnostics* **4**, 135–144.

SHUBER, A.P., MICHALOWSKY, L.A., NASS, G.S., SKOLETSKY, J., HIRE, L.M., KOTSOPOULOS, S.K., PHIPPS, M.F., BARBERIO, D.M. AND KLINGER, K.W. (1997). High throughput parallel analysis of hundreds of patient samples for more than 100 mutations in multiple disease genes. *Human Molecular Genetics* **6**, 337–347.

SYVANEN, A.C., AALTO-SETALA, K., HARJU, L., KONTULA, K. AND SODERLUND, H. (1990). A primer-guided nucleotide incorporation assay in the genotyping of apolipoprotein E. *Genomics* **8**, 684–692.

TOBE, V.O., TAYLOR, S.L. AND NICKERSON, D.A. (1996). Single-well genotyping of diallelic sequence variations by a two-color ELISA-based oligonucleotide ligation assay. *Nucleic Acids Research* **24**, 3728–3732.

TYAGI, S. AND KRAMER, F.R. (1996). Molecular beacons: probes that fluoresce upon hybridization. *Nature Biotechnology* **14**, 303–308.

TYAGI, S., BRATU, D.P. AND KRAMER, F.R. (1998). Multicolor molecular beacons for allele discrimination. *Nature Biotechnology* **16**, 49–53.

WALLACE, R.B., SHAFFER, J., MURPHY, R.F., BONNER, J., HIROSE, T. AND ITAKURA, K. (1979). Hybridization of synthetic oligodeoxyribonucleotides to phi chi 174 DNA: the effect of single base pair mismatch. *Nucleic Acids Research* **6**, 3543–3557.

WALLACE, R.B., JOHNSON, M.J., HIROSE, T., MIYAKE, T., KAWASHIMA, E.H. AND ITAKURA, K. (1981). The use of synthetic oligonucleotides as hybridization probes. II. Hybridization of oligonucleotides of mixed sequence to rabbit beta- globin DNA. *Nucleic Acids Research* **9**, 879–894.

WANG, D.G., FAN, J.B., SIAO, C.J., BERNO, A., YOUNG, P., SAPOLSKY, R., GHANDOUR, G., PERKINS, N., WINCHESTER, E., SPENCER, J., KRUGLYAK, L., STEIN, L., HSIE, L., TOPALOGLOU, T., HUBBELL, E., ROBINSON, E., MITTMANN, M., MORRIS, M.S., SHEN, N., KILBURN, D., RIOUX, J., NUSBAUM, C., ROZEN, S., HUDSON, T.J., LIPSHUTZ, R., CHEE, M. AND LANDER, E.S. (1998). Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* **280** (5366), 1077–1082.

WEEKS, D.E. AND LATHROP, G.M. (1995). Polygenic disease: methods for mapping complex disease traits. *Trends in Genetics* **11**, 513–519.

YERSHOV, G., BARSKY, V., BELGOVSKIY, A., KIRILLOV, E., KREINDLIN, E., IVANOV, I., PARINOV, S., GUSCHIN, D., DROBISHEV, A., DUBILEY, S. AND MIRZABEKOV, A. (1996). DNA analysis and diagnostics on oligonucleotide microchips. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 4913–4918.

PART 2

Molecular Biology & Genetics