NEW TECHNOLOGIES FOR GENETIC ANALYSIS IN RESEARCH AND DIAGNOSTICS SETTINGS

Thesis submitted for the degree of

Doctor of Philosophy

at the University of Leicester

by

Linda Strömqvist Meuzelaar

Department of Genetics

University of Leicester

November 2007

NEW TECHNOLOGIES FOR GENETIC ANALYSIS IN RESEARCH AND DIAGNOSTICS SETTINGS

LINDA STRÖMQVIST MEUZELAAR

ABSTRACT

This thesis work was designed to provide advanced technologies for the analysis of genetic variation, with emphasis upon complex genome exploration and DNA diagnostics.

For DNA sequence analysis, the work was based on Dynamic Allele-Specific Hybridization (DASH). DASH was originally developed for SNP genotyping via melt-curve analysis of allele-specific oligonucleotides. DASH was adapted for more general use in DNA diagnostics, and was demonstrated for mutation scanning of a 97 bp human genomic sequence and parallel detection of multiple mutations in the *gyrA* gene of *Salmonella*. Mutation scanning studies utilised a series of partially overlapping probes, where base substitutions were detected by alterations in the melting profile of the probes hybridising to each variant base. Three SNPs in the target sequence were correctly detected in 16 samples. The mutation status of 62 *Salmonella* strains was accurately determined, using a combination of wildtype and mutation-specific probes. Work was also initiated to convert this concept to a microarray format, and to establish a platform for Thalassaemia diagnostics.

To facilitate highly parallel analyses, 'MegaPlex PCR' was developed for multiplex amplification of specific DNA fragments. This technology employs solid-phase PCR to capture many different targets and at the same time incorporate common primer sequences. Liquid-phase PCR is then used to amplify all products simultaneously, using a single primer pair. Innovations included the use of partly double-stranded surface-oligonucleotides ('Barrier Oligos') to enhance surface reactions, and the use of complex genomic DNA in native format and after pre-amplification. A final proof-of-principle study targeted 50-plex and 75-plex reactions to genotype human SNPs. Next-generation sequencing revealed that up to 86% of the targeted sequences were recovered and less than 10% of the amplified molecules were primer-dimers.

With further development, methods such as these should have substantial utility for improved DNA diagnostics and personalised medicine.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

LIST OF ABBREVIATIONS

ASO	Allele-specific oligonucleotide hybridisation
bp	Base pair
CGH	Comparative genomic hybridisation
CNV	Copy number variation
DASH	Dynamic allele-specific hybridization
ddNTP	Dideoxynucleoside triphosphate
DGGE	Denaturing gradient gel electrophoresis
DHPLC	Denaturing high performance liquid chromatography
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
FRET	Fluorescence resonance energy transfer
GS20	Genome Sequencer 20 system
gyrA	DNA gyrase subunit A
HBB	Beta globin
LATE-PCR	Linear-after-the-exponential PCR
LD	Linkage disequilibrium
LNA	Locked nucleic acid
MME-PCR	Multiplex microarray-enhanced PCR
MPSS	Massively parallel signature sequencing
OLA	Oligonucleotide ligation assay
PCR	Polymerase chain reaction
RFLP	Restriction fragment length polymorphism
SNP	Single nucleotide polymorphism
SSCP	Single-stranded conformational polymorphism
TGCE	Temperature gradient capillary electrophoresis
TGGE	Temperature gradient gel electrophoresis
Tm	Melting temperature
TTGGE	Temporal temperature gradient gel electrophoresis
wt	Wildtype

ACKNOWLEDGEMENTS

Many people have contributed to this work in different ways. I would especially like to thank:

My supervisor Professor Anthony Brookes, for many years of inspiration and enthusiasm, for scientific guidance and support, and for believing in my abilities.

All my colleagues at the Department of Genetics at the University of Leicester, for creating a friendly working environment. In particular I would like to thank the members of Brookes' group and lab G10:

Karen Monger, for starting up the lab and introducing me to the department. Reshma Bharkhada, for continuous efforts trying to keep the lab in order, for a great time in Heidelberg and for always being positive and helpful. Dr. Pari Datta, for interesting discussions and travel company to Hong Kong. Dr. Owen Lancaster, for bioinformatics support. Dr. Colin Veal, for useful comments on this thesis. Katherine Reekie, for increasing the female influence of the computer room. Gudmundur "Mummi" Thorisson, for being positive. Lisa Åkerman, for Swedish company and interesting collaborations. Dr. Raymond Dalgleish, for always joining our group meals.

Dr. Ingrid Berg, for the best "B&B" in Leicester; my last working-visit to Leicester would have been so much harder without your great food.

My former colleagues at Karolinska Institute, in Sweden. In particular: Dr. Magnus Jobs and Dr. Mathias Howell, for introducing me to method development and DASH. Vivian Daza-Saucedo, for sorting out any paperwork. Fredrik Sjöholm, for excellent food and great company at all those Christmas parties. Anna Höijer, for helping out with the odd experiment.

Collaborators for the different studies in this thesis:

Dr. Katie Hopkins, at the Health Protection Agency Centre for Infections, for providing Salmonella DNA and genotype information on all strains.
Dr. Bruno Frey, Dr. Guido Kopal and Paul Pache, at Roche Diagnostics GmbH, Penzberg, Germany, for sequencing the MegaPlex PCR products on their GS20 instrument.
Dr. George Patrinos, at Erasmus University Medical Center, Rotterdam, The Netherlands, for providing control samples for thalassaemia mutations.

My parents, Elisabet and Staffan, for your support during all my life. My sister Frida, who managed to get a PhD before me, and my brother Gustav, who is aiming for one as well – who could have thought we would end up with so many scientists in the family?!

My husband Gerard, for all your love, support and patience during all the time we were living apart, and when writing this thesis. For great weekends in Leicester and London, and holidays which managed to take my mind off this thesis work. There will be much more time for you & me now, for Dutch lessons and whatever the future brings.

CHAPTER 1.

INTRODUCTION

1.1 GENETIC VARIATION

In human cells the genetic material is stored in 46 chromosomes; 22 pairs of autosomes and two sex chromosomes, either XX in females or XY in males. The euchromatic part of the human genome, covering a total of approximately 3 billion bases, has now been sequenced[1-3] and is publicly available for researchers to use. Sequencing the human genome was the first step towards a detailed understanding of how the genome functions. This has allowed computational analysis to determine the chromosomal location of known genes and also enabled increasingly precise prediction of previously undetected genes[1]. Further efforts are required to identify all functional elements of the genome, not only the genes, but also other regulatory sequences that influence gene expression and genome structure[4,5]. An important aspect of this is the study of genome variation and its implication for human health.

Genetic variation between individuals of the same species (such as humans) includes a spectrum of different types of variation, ranging from large chromosomal differences down to variations at the single base pair level, and include insertion, deletion, substitution and rearrangement of genetic material[6].

Large chromosomal differences are visible if studied under a microscope and include missing or extra chromosomes compared to the normal diploid state (aneuploidy), or large chromosomal rearrangements. Additional or missing whole chromosomes are usually not compatible with life and result in early miscarriage, but there are exceptions such as an extra copy of chromosome 21, leading to Down's syndrome. Additional copies of the sex chromosomes, X and Y, give also only mild or even no symptoms. Genetic material can also be transferred from one chromosome to another, duplicated, inverted or deleted with variable clinical effects[7].

A common type of genetic variation involves the difference in length of repetitive sequences, due to the number of units. Depending on the length of the repeat unit these genetic markers are called minisatellites (for longer units) and microsatellites or short tandem repeats (for shorter units), with a cut-off around 10 bp. The number of repeat units at a specific locus can be highly variable across a population and these markers are therefore ideal for identification of individuals (for example for forensic applications[8,9] and paternity testing[10]).

Single nucleotide polymorphisms (SNPs) are single base pair substitutions, and this type of genetic variation is very common throughout the human genome, with on average one every 200-300 bp in the human population[11]. A distinction is usually made between rare mutations and polymorphisms. An alternative allele is always created from a mutation event, but the position is not regarded polymorphic unless the variant has spread in a population so that the allele frequency of the most common allele is less than 99%. In theory four SNP alleles could be present at any genomic position, but the chances of more than two base alternatives at the same position and that each of them reach high frequency in a population are extremely low and therefore essentially all SNPs have only two alleles[12], for example C and T. Since humans are diploid, with two copies of each chromosome, an individual can have any of three genotypes and either have two copies of the same allele (homozygous), CC or TT in the above example, or have one copy of each allele (heterozygous), CT. Short insertion and deletion polymorphisms are sometimes categorised along with SNPs, since they also affect one or few bases and are stable once formed. They can therefore be used for the same type of applications as SNPs, but insertions and deletions are not as common as base substitutions.

SNPs have become popular genetic markers. This is partly because they are very common throughout the genome, and partly because it is relatively easy to devise genotyping assays and apply automatic analysis, since only the identity of a single base has to be determined and only two alleles are present. Efforts to identify and catalogue all human SNPs has resulted in publicly available information on almost 12 million SNPs (11,883,685 non-redundant variants in build 128, October 2007) in dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/) of which roughly half (6,262,709) have been validated. A SNP is regarded as validated if it has been submitted by multiple

users, or if genotype data is available and the minor allele has been observed in at least two chromosomes. In non-validated SNPs, the allele frequency is either unknown or the minor allele may only have been detected in a single chromosome, which makes it difficult to distinguish between sequencing errors, rare mutations and real SNPs.

The functional impact of SNPs depends on their location. SNPs in coding regions could affect protein function. This is the case if the base substitution alters a codon, so that it codes for a different amino acid (non-synonymous SNPs). The effect on the protein function may vary, depending on the property of the substituted amino acid and the location of the substitution within the protein. One of the most severe substitutions results in a premature stop codon, leading to a truncated and therefore often non-functional protein. SNPs can also affect splice sites and regulatory sequences.

Whilst SNPs are the most common type of genetic variation, they do not account for the majority of variant bases in the genome. Rarer and larger variants, such as insertions, deletions and rearrangements affect a larger proportion of the variant bases in the genome, although the number of these variants is less[13]. During the last few years it has become clear that structural variation in the human genome and the presence of copy number variable segments (CNVs) are much more common than previously thought. CNVs are generally defined as DNA fragments larger than 1 kb that vary in copy number in a population. Efforts are now underway to create a genome map of CNVs[6,14-17]. These large scale studies mainly use techniques such as array-Comparative Genome Hybridisation (array-CGH), where different parts of the genome are represented on the array either by large clones or by multiple oligonucleotides, and differences in the relative hybridisation intensity between two samples at certain genomic regions are indicative of copy number differences at these loci. The detection of CNVs is highly dependant on many parameters, such as the size of the copy number variable region, the genomic coverage and the number of samples that are studied. The full characterisation and detailed analysis of CNVs will require the development of new or adapted methods. But achieving this will not be as straight forward as SNP validation, due to the complexity of CNVs, regarding, for example, size range and variability in copy number.

1.1.1 GENETIC VARIATION AND DISEASE

While some genetic variants have no apparent effect, others influence an individual's appearance[18,19], behaviour or disease risk. Genetic factors can also determine how well an individual responds to drug treatment[20,21]. Some genetic variants directly cause disease by, for example, altering the function of a protein or altering gene dosage, but these are usually rare diseases. Most variants have no or little affect on phenotype, and many common diseases are believed to be caused by a combination of several genetic- and environmental factors[22,23].

While the majority of genes causing monogenic disorders have been identified, much is still unknown about the genetic component of many complex diseases. Great efforts are being made to identify genetic risk factors in order to gain insight in disease mechanism, and in the long term improve diagnostics and lead to new treatment. Since each genetic risk factor only contributes with a small effect, it is, however, much more difficult to identify genetic risk factors for complex disease compared to single-gene disorders[23].

The most common strategy for identifying genetic risk factors of common disease is the use of genetic association studies[24,25], where genetic markers are tested in groups of cases and controls. If the allele frequency of a specific marker is significantly different in the cases compared to the control group, the tested allele is associated with the disease. This may seem straight-forward, but one of the problems is how to choose which markers to test. One approach has been to test markers, usually SNPs, in genes that are thought to be involved in the disease. Such candidate-gene studies have identified some genetic risk factors, for example variants in the APOE gene associated with Alzheimer's disease[26], variants in CTLA4 associated with Type 1 diabetes[27,28] and variants in KCNJ11 and PPARG associated with Type 2 diabetes[29,30], but many other studies produce only negative results[31] and some associations have failed to replicate in subsequent studies[32]. An alternative strategy, particularly if no obvious candidate genes are identified, is to study markers throughout the whole genome in the hope that genetic risk factors can be identified[33,34]. Such large-scale studies have only just become possible due to the development of highly parallel genotyping assays.

For whole-genome studies it is desirable to cover as much as possible of the genome using the minimum number of markers[35,36]. It is possible to reduce the number of markers when the genotype at one locus can predict the genotypes at other loci[37]. Alleles that are correlated in this way are said to be in linkage disequilibrium (LD), and each set of such alleles form a haplotype block. To choose the most informative SNPs for whole-genome genotyping studies it is necessary to have knowledge about the LD patterns and haplotype diversity across the genome[38,39]. The goal of the HapMap project[11,40], was to obtain this information by genotyping SNPs throughout the genome in four different populations. Whole-genome association studies using the information gained in the HapMap project have now started to identify risk alleles for complex diseases, such as childhood asthma[41], colorectal cancer[42,43], prostate cancer[44], type 2 diabetes[45-47] and Crohn's disease[47,48].

However well the markers are chosen, the power to detect an associated allele also depends on the sample size and the effect of the risk allele. When many markers are tested, the risk also increases of finding positive associations by chance[49]. To avoid false positive results the statistics can be adjusted for multiple testing[50], and it is also important to replicate associations in follow-up studies[49]. It is also important to distinguish between statistical significance and biological significance, since even though statistically significant very small effects of a risk allele may not lead to particularly useful biological knowledge.

1.1.2 Methods for DNA Sequence Variation Detection

To study genome variation in an efficient manner it is necessary to have multiple methods to chose from and combine. Those methods can be divided into two main categories, depending on their application. Usually different types of methods are used for detecting previously unknown mutations and sequence variants, for example scanning a specific gene for mutations, than genotyping already characterised mutations and sequence variants[51,52]. Different types of methods are probably also necessary depending of what kind of mutation one is studying, for example a large deletion compared to a single base pair substitution.

Mutation scanning can be done at different resolution depending on how much detailed information is necessary. At the first instance it could be sufficient to know whether there is any sequence difference at all in a specific gene of a specific sample compared to a reference sample, and only when this is confirmed one could proceed and determine the exact nature of that mutation. When a sequence variant is fully characterised a specific genotyping assay could be developed to investigate that particular variant in a larger sample set, and if it has been identified as a risk allele it could be genotyped for diagnostics purposes.

The choice of analysis method might differ depending on the scale of the study and the application. There are partly different criteria when choosing methods for genome scans compared to diagnostic analysis. For efficient high-throughput genotyping the methods need to be scalable to high multiplex levels, preferably in a miniaturised format. Occasional marker dropouts are acceptable, since it is often possible to replace one marker with another. Although multiplexing options will probably become more important also for diagnostic applications, this cannot be done at the cost of genotyping accuracy, which is extremely important. A diagnostic method of choice should be able to accurately score any causative mutations or detect all new mutations in any chosen gene and would therefore need to reliably detect mutations in any sequence context.

The following sections will describe methods for the detection and genotyping of mutations and polymorphisms, with focus on single base pair substitutions. Methods both for mutation scanning, that are able to detect differences down to single base pair substitutions, and methods for SNP genotyping will be reviewed (**Table 1.1**). The principles behind each method are described and developments towards miniaturisation and high-throughput applications are discussed.

Table 1.1 Methods for mutation scanning and genotyping.

Method	Multiplex Platform	Detection	Application
dsDNA stability			
SSCP	CAE	Mobility shift	Mutation scanning
DGGE		Mobility shift	Mutation scanning
TGGE, TGCE	CAE	Mobility shift	Mutation scanning
DHPLC	Capillary columns	Mobility shift	Mutation scanning
High resolution melting analysis		Fluorescence / melting profile	Mutation scanning
Sequencing:			
Sanger sequencing	CAE	Fluorescence	Scanning and genotyping
Sequencing-by-synthesis	Microarray / bead array	Fluorescence	Scanning and genotyping
Pyrosequencing®		Luminescence	Scanning and genotyping
Sequencing-by-ligation		Fluorescence	Scanning and genotyping
Sequencing-by-hybridisation	Microarray	Fluorescence	Scanning and genotyping
DNA cleavage			
RFLP	CAE	Fragment size	Genotyping
TaqMan®		Fluorescence	Genotyping
Invader® assay	Microarray	Fluorescence	Genotyping
DNA polymerisation			
Allele-specific PCR		Electrophoresis, fluorescence, melting profile	Genotyping
Allele-specific primer extension	Microarray	Fluorescence	Genotyping
Minisequencing	Microarray	Fluorescence	Genotyping
DNA ligation			
OLA		Fragment size, fluorescence	Genotyping
Padlock probes	Microarray	Fluorescence	Genotyping
Ligation + extension			
GoldenGate®	Microarray	Fluorescence	Genotyping
Molecular inversion probes	Microarray	Fluorescence	Genotyping
DNA hybridisation			
Molecular beacons		Fluorescence	Genotyping
Homogeneous melt-curve assays		Fluorescence / melting profile	Genotyping
ASO	Microarray	Fluorescence	Genotyping
DASH		Fluorescence / melting profile	Genotyping

CAE: Capillary array electrophoresis

1.2 METHODS FOR MUTATION SCANNING

1.2.1 CONFORMATION AND STABILITY-BASED METHODS

For the detection of previously unknown sequence variants several methods take advantage of sequence-specific mobility patterns in electrophoresis or chromatography separation methods. The target sequence is usually first amplified by PCR and then small sequence differences can be distinguished due to sequence-specific conformation of either single-stranded or double-stranded DNA, that affects the migration of the DNA, under appropriate assay conditions. These methods can therefore detect mutations, but not identify the exact nature and location of the sequence variants.

In single-strand conformation polymorphism (SSCP) analysis[53,54], single stranded DNA fragments are separated due to their difference in mobility in a non-denaturing gel. Even single base changes can alter the conformation of the DNA and thereby change the migration pattern in the gel compared to a reference sample. The two strands of a PCR product can also be detected independently, thereby increasing the specificity, if the primers are labelled with distinct fluorescent dyes. For automated analysis the assay can be run on a capillary electrophoresis system with internal size standards which facilitates the comparison of multiple samples[55,56]. For high-throughput analysis a microfabricated capillary array electrophoresis device, with the possibility of running up to 384 samples in parallel, has been used[57].

Gradient denaturing gels can be used to detect sequence differences in double-stranded DNA. A chemical gradient is used in denaturing gradient gel electrophoresis (DGGE)[58,59] and a temperature gradient is used in temperature gradient gel electrophoresis (TGGE)[60,61] to reveal sequence variants as differences in the DNA mobility pattern. When the DNA duplex is starting to denature, its mobility becomes reduced in the gel. When a mutation alters the duplex stability (melting point) the variant and reference DNA duplexes will denature at different concentrations of the denaturing agent (or at different temperature in TGGE), which will lead to differences in migration. Work towards miniaturisation and high-throughput analysis has resulted in micro-TGGE[62] using much smaller gels and smaller sample volumes. Faster analysis and higher throughput is also possible with temperature gradient capillary

electrophoresis (TGCE), an adapted method using automated capillary array instruments[63,64].

The difference in stability between a fully complementary DNA duplex (homoduplex) and a duplex with one or several mismatch positions (heteroduplex) is larger than the difference between two different homoduplexes. It is therefore easier to detect mutations in form of heteroduplexes with the above mentioned methods. Heteroduplexes can be created by heat denaturation and re-annealing of a PCR product that has amplified a heterozygous locus. When the products re-anneal they will form a mixture of the two original homoduplexes as well as heteroduplexes, when the two strands of different alleles anneal with each other. For improved discrimination between two different homoduplexes, the test samples can also be mixed with a reference homozygous sample and thereby convert samples homozygous for the alternative allele to heterozygous samples.

Heteroduplex formation is also used to detect sequence variants by denaturing high-performance liquid chromatography (DHPLC)[65,66]. The homo- and heteroduplexes will have different mobility patterns in the chromatography column under partially denaturing conditions. The method provides high sensitivity, but typically low throughput. However, higher throughput has been achieved using miniaturised capillary columns[67], and further multiplexing was also demonstrated by labelling samples with different fluorophores.

Of the above mentioned methods the highest sensitivity and specificity is obtained by DHPLC[68]. The equipment cost is, however, very high. A miniaturised format would reduce the reagent cost for all methods and increase throughput and automation, especially compared to slab gels. A drawback of all these methods is that they usually require target-specific optimisations, for example of the assay temperature and the buffer used, or the samples have to be run at multiple conditions[66,69,70].

The difference in duplex stability between homo- and heteroduplexes can also be assessed in a closed-tube system by DNA melting analysis, since mismatched base pairs affect the melting profiles. Technologies that use this principle usually employ fluorescence detection. For example DNA double-strand specific dyes can be used, that fluoresce only when binding to double-stranded DNA as opposed to single-stranded

DNA. The melting profile can therefore be recorded by fluorescence monitoring while heating the sample. Heterozygous variants in PCR products up to 100 base pairs in length have been successfully detected on a real time PCR instrument[71], using the DNA double-strand specific dye SYBR Green I. For larger fragments, the differences in melting profile become very small and a high-resolution melting instrument is required to detect sequence differences. High-resolution amplicon melting analysis was first introduced on products created from a fluorescently labelled amplification primer[72], but has also been implemented with the double-strand specific dye LCGreen I[73]. Using high-resolution melting analysis, heterozygous single base pair variants have been detected in products up to 1kb[74] and distinct homoduplexes can be distinguished in short amplicons, up to 50 bp, with the exception for some A:T or C:G polymorphisms[75].

Melting curve analysis is described in more detail, with regards to genotyping known sequence variants, in section **1.3.4.2**.

1.2.2 DNA SEQUENCING

Direct sequencing of DNA can be used either as a mutation scanning method on its own, or it can be used to confirm and determine the exact location of mutations detected by other scanning methods. Sanger sequencing[76] is still widely used, and modern capillary devices generate on average around 800 base pairs of good quality sequence[77]. Efforts to further increase throughput and reduce costs of the Sanger sequencing method require miniaturisation to reduce reagent costs, and microfluidic devices are being used to achieve this[77].

Miniaturisation is necessary to obtain increased throughput. "Next generation" sequencing platforms are able to do this on high-density array platforms using, for example, sequencing-by-synthesis strategies, which involve real-time detection of incorporated nucleotides[77,78].

One approach has been to use Pyrosequencing®[79,80]. The incorporation of each nucleotide is detected by the release of pyrophosphates, via a series of enzymatic reactions that result in light emission. The different nucleotides are sequentially added to the reaction, each resulting in a positive or negative signal to determine the sequence. Consecutive bases of the same nucleotide result in proportionally higher signals. This

technology has been implemented in high density bead arrays[81] by 454 Life Sciences and is now commercially available through Roche Applied Science. Each target is amplified from a single molecule attached to a bead in emulsion PCR[82]. The beads are then transferred to picolitre-sized reaction vessels for pyrosequencing, capable of read lengths around 200-300 bps.

Another sequencing-by-synthesis strategy involves the incorporation of fluorescently labelled nucleotides. This can be done in a single-colour version where all nucleotides are labelled with the same fluorophore, and in that case the different nucleotides have to be incorporated sequentially[83,84]. If four different fluorophores are used all bases can be added together in each cycle and the incorporated base(s) will be determined by the relative fluorescence detected for each nucleotide[85,86]. In the latter case it is very important that the nucleotides are efficiently blocked for further extension so that only one base is incorporated in each cycle. In either case the fluorescence has to be eliminated between each cycle to avoid signal contamination from the previously incorporated bases. This has been done by bleaching[84], photo cleavage[83,85] or chemical cleavage[86] of the fluorophore. These types of sequencing strategies have a typical read length around 20-30 bases. This sequencing principle is used in the commercial systems developed by Solexa, now available through Illumina, and the soon to be released system from Helicos BioSciences.

Another option is sequencing-by-ligation. For sequencing shorter 'signatures' a method called Massively Parallel Signature Sequencing (MPSS)[87] can be used. Different targets are isolated onto separate beads, where a short sequence can be obtained by cycles of cleavage, adaptor ligation and detection. Adaptors, with all possible combinations of four-base overhangs, are ligated to bead-bound restriction digested DNA. The target overhang sequence is revealed by detecting which adaptors are ligated to which beads. The DNA is then cleaved again by a type II restriction endonuclease, which allows the recognition sequence to be present in the adaptor while the DNA is cleaved to expose the next four bases. Another ligation-based system is the SOLiD system available through Applied Biosystems. This system consists of cycles of probe ligation followed by fluorescence detection and partial cleavage of the probe.

All these “next generation” sequencing methods directly or indirectly sequence single molecules. In the latter case, clonal amplification products are sequenced. Since all reactions are separated there is virtually no upper limit on the degree of multiplexing. It should also be easier to detect heterozygous bases when the two alleles are sequenced independently compared to a mixed sample where heterozygous positions usually produce lower signals with the risk of missing one of the alleles. It is therefore also possible to detect rare alleles in heterogeneous samples, such as cancer tumours[88]. Single molecule sequencing would also be ideal when heterozygous insertions or deletions are present in a sample, since the sequences of mixed samples are shifted after the insertion/deletion break point.

Sequencing by hybridisation on microarrays is also an option; this will be discussed in section **1.3.4.1**.

1.3 GENOTYPING METHODS

To analyse already characterised genetic variants it is most suitable to use a specific assay for each marker. To genotype known single base pair variants, assays can be designed using a wide range of methods. At first glance these methods can seem very diverse, but they are in fact all based on DNA cleavage, DNA polymerisation, DNA ligation or DNA hybridisation.

1.3.1 GENOTYPING BY DNA CLEAVAGE

Sequence-specific DNA cleavage can be used to genotype SNPs, when the DNA is specifically cleaved at the presence of one allele but left intact for the other allele. Highly specific enzymes, that recognise a specific sequence or structure, are used for these reactions. In some assays a larger DNA fragment, such as a PCR product, is cleaved while other assays employ oligonucleotide probes that first must hybridise to the target sequence before being cleaved.

Restriction fragment length polymorphism (RFLP) detects sequence variants with the use of sequence-specific restriction enzymes. To genotype a single base variant, a restriction enzyme is chosen so that its recognition sequence is only present for one of the alleles and not the other, thereby obtaining allele-specific cleavage. After fragment amplification by PCR and restriction enzyme digestion the alleles can be distinguished as fragments of different length (cut or intact DNA) (**Figure 1.1 a**). Those fragments are traditionally identified by gel electrophoresis, but can also be analysed by melting curve analysis[89], since the digested and intact fragments will have different melting temperatures. Melt-curve analysis is further discussed in section **1.3.4.2**.

The 5' nuclease, or TaqMan®, assay is a homogeneous reaction where the amplification and genotyping is done in the same tube and relies on allele-specific cleavage of hybridisation probes[90]. Cleavage of the probe takes place during PCR, when the probes are hybridised to their target sequence, by the 5' nuclease activity of the Taq polymerase during the primer extension reaction (**Figure 1.1 b**). The probe is only cleaved when the probe perfectly matches the target sequence. The probes are dually labelled with a fluorophore on one end and quencher (a molecule that can absorb

fluorescence energy but does not fluoresce itself) at the other end. While the probe is intact the energy of the excited fluorophore is transferred to the quencher and thereby reduce the fluorescence, but as the probe is cleaved the fluorescence increases in the reaction. To detect two SNP alleles simultaneously, two allele-specific probes labelled with different fluorophores must be used in the same reaction.

The Invader® assay[91,92] also employs allele-specific cleavage of an oligonucleotide probe. In this assay the interaction of two probes is required for cleavage. One 'invader probe' and one allele-specific probe overlap at the polymorphic position (**Figure 1.1 c**). This structure is recognised by a specific enzyme, a flap endonuclease, which will cleave the allele-specific probe at the polymorphic site. This structure is only formed when the probe matches the SNP allele, so genotyping can be done with two allele-specific probes. The cleavage event is usually detected by a second, universal, invader reaction where the cleaved probe acts as a second invader probe. This second reaction will cleave a detection probe, so that a fluorophore is released and separated from its adjacent quencher, resulting in increased fluorescence.

Neither of these genotyping principles is ideal for use in high-multiplex homogeneous applications. Multiple targets can, however, be detected in parallel by electrophoresis if the expected fragments (cut and intact) have different size. Capillary electrophoresis then provides higher sensitivity and automation possibilities. The assays that use fluorescence detection offer very low multiplex capabilities as long as the reactions takes place in the same reaction tube and would be limited by the number of fluorescent dyes that could be detected independently. Higher multiplex reactions are possible if the reaction can take place on an array format on solid support, which has been demonstrated for the Invader® assay[93].

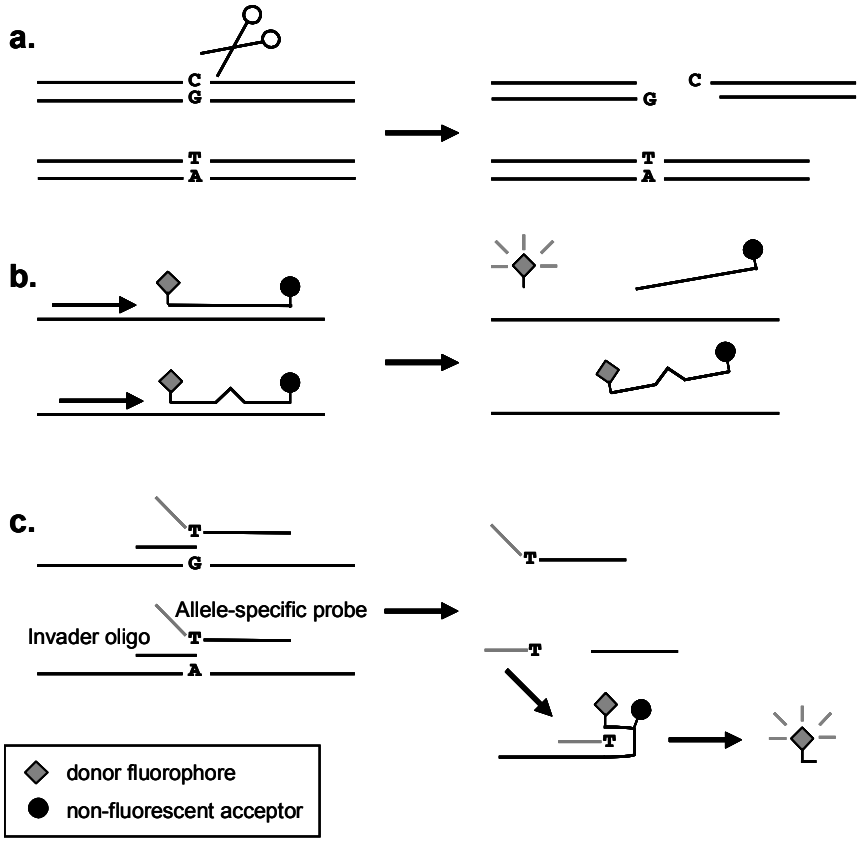


Figure 1.1 SNP genotyping by DNA digestion/cleavage.

a) Genotyping by RFLP. A restriction enzyme is used where one of the SNP alleles disrupts the recognition sequence, leading to allele-specific cleavage.

b) Genotyping by the TaqMan® assay. A hybridisation probe is cleaved when it is perfectly hybridised to the target sequence, by the 5' nuclease activity of Taq polymerase during primer extension. The probe cleavage is detected by the release of a fluorophore, leading to increased fluorescence. For SNP genotyping two allele-specific probes are used.

c) Genotyping by the Invader® assay. The Cleavase® enzyme recognises the invader structure and cleaves the probe at the overlapping position. This structure is formed by the target, the invader oligo and the probe, and only if the probe hybridise to the correct allele. The cleaved, non-complementary part of the probe can then act as an invader oligo in a subsequent detection reaction, where a fluorophore is released.

1.3.2 GENOTYPING BY DNA POLYMERISATION

DNA polymerisation can be used for allele-specific extension of a primer. This can be done in two ways, with the last base of the primer either at, or just before, the polymorphic base.

The first principle is allele-specific PCR or allele-specific primer extension, where the very last base at the 3' end of a primer is complementary to one of the alleles and thus has one base pair mismatched to the other allele. In allele-specific PCR this should result in successful amplification only for the matched allele and prevent amplification of the other allele[94]. To call the correct genotype two allele-specific primers have to be used, one for each allele (**Figure 1.2 a**). To improve the specificity of the primers, additional mismatches can be introduced in the primer sequence[95] or LNA bases can be used at the allele-specific position[96]. The presence or absence of product from each reaction can be detected by gel electrophoresis, real-time PCR[97] or by melt-curve analysis of the product[98].

The primer extension step can also be separated from the amplification step. To genotype by allele-specific primer extension, the targets are first amplified by PCR, and then an extension reaction is used for allelic detection. This strategy has, for example, been implemented in a microarray format, where the allele-specific primers are attached to the surface[99] and the primer extension is detected by incorporation of labelled nucleotides. The same principle is used on the Infinium "whole genome genotyping" assay on bead arrays[100,101] which can genotype up to one million SNPs in parallel in a single experiment.

The second version of allele-specific primer extension is "minisequencing"[102] where one common primer (for each SNP) is hybridised with its 3' end just in front of the variant position. Primer extension is allowed to take place with either of two labelled dideoxy nucleotides, matching each of the alleles, and the genotype can be determined by comparing the relative incorporation of the two bases (**Figure 1.2 b**). Two separate reactions are required if the two nucleotides are labelled in the same way (for example radioactive labels[103] or single fluorophores[104]), but it is possible to use a single reaction if dideoxynucleotides are used that are labelled with different fluorophores[105]. When the

method was first developed one strand of an amplified fragment was bound to a solid support, and a minisequencing primer was hybridised and extended. This strategy allows only one sequence per well. The method has since been implemented in a microarray format with the minisequencing primers for different targets attached to a glass slide array[106,107] or a high density bead array[101,108].

Both strategies are thus suitable for highly parallel assays using both planar arrays and bead arrays with fluorescent detection. The advantage of minisequencing is that only one detection primer per SNP is required, whereas allele-specific extension requires two allele-specific primers. Allele-specific extension, on the other hand, only needs a single label while separate fluorescent dyes are required for minisequencing.

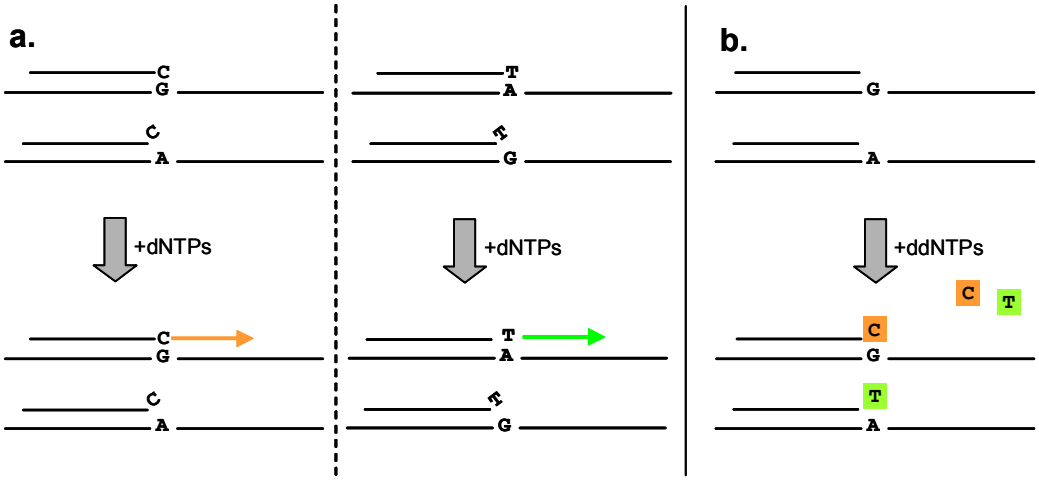


Figure 1.2 SNP genotyping by DNA polymerisation.

a) Allele-specific primer extension. Two allele-specific primers are used separately, in this example the primers have either a 'C' or a 'T' at the 3' ends. The primer is only extended when the 3' end matches the target DNA. Genotyping is done by determining which of the two allele-specific primers are extended. Primer extension can be detected by incorporation of fluorescently labelled nucleotides.

b) Allele-specific incorporation of nucleotides, "minisequencing". The 3' end of the primer is hybridised next to the polymorphic position. Single nucleotide extension is then performed with differently labelled dideoxynucleotides, where only the base matching the SNP allele is incorporated.

1.3.3 GENOTYPING BY DNA LIGATION

Genotyping by DNA ligation takes advantage of the specificity of DNA ligase. This enzyme can ligate the 3' end of one DNA strand to the 5' end of an adjacent strand, if the two ends are correctly hybridised to their complementary strand. To use this principle in genotyping, the SNP should be located at the end of one of the ligating molecules to obtain alleles-specific ligation to the adjacent molecule.

In the oligonucleotide ligation assay (OLA)[109-111] (**Figure 1.3 a**) three probes are used; one common oligonucleotide positioned just next to the variant site and two versions of allele-specific probes, hybridising next to the first probe and thus having the variant position at the end of the probe. The ligation products can be separated in size from the non-ligated oligonucleotides and identified by electrophoresis or the ligation products can be detected, for example, via fluorescent labels on the probes[112].

The same principle of allele-specific ligation is used for padlock probes. Here only one, long, oligonucleotide is used, which has target-specific sequences at both the 3' and 5' ends. These become hybridised next to each other when hybridised to their target sequences, with an internal linker sequence in between[113]. The two ends of the probe can be ligated to form a circular molecule, but only if both ends of the probe are hybridised to their complementary bases (**Figure 1.3 b**). The circularised probes can be specifically amplified by rolling circle amplification[114] or PCR[115] using priming sites in the linker sequence of the padlock probes, after remaining linear molecules have been degraded. Since the amplification takes place after the allelic detection, padlock probes can be used directly on genomic DNA template without prior amplification of the target sequences. An additional benefit of amplifying the targets after the allelic discrimination step is that it provides great possibilities for higher multiplex reactions. This would be achieved if the same common priming sites are introduced in all targets, allowing any number of sequences to be amplified simultaneously.

Two additional reaction principles can be multiplexed in the same way as padlock probes. Molecular inversion probes[116,117] are highly similar to the padlock probes, but employ a combination of primer extension and ligation for genotyping. Instead of having an allele-specific base at the end of the probe the probes hybridise to the target sequence with a single base gap between the 5' and 3' ends, at the SNP position. Using

separate reactions for primer extension with the different nucleotides, the gap will only be filled if the correct, matching, nucleotide is present, after which the probes can be ligated and amplified in the same way as the padlock probes (**Figure 1.3 c**).

The GoldenGate® assay[118] is also based upon a combination of primer extension and ligation (**Figure 1.3 d**). In this method two allele-specific oligonucleotides compete to hybridise with their 3' end to the SNP, while another oligonucleotide is hybridised a few bases away, on the same strand. The gap is filled by extension of the matching allele-specific oligonucleotide, after which the two adjacent molecules can be ligated. Common tail sequences on all oligonucleotides allow subsequent simultaneous amplification of all ligated products.

These three multiplex assays all produce complex mixtures of amplified targets. For genotyping of all targeted SNPs, the products need to be individually detected. Such parallel product detection is typically done on microarrays, by the incorporation of target-specific tag sequences and by fluorescently labelling the products.

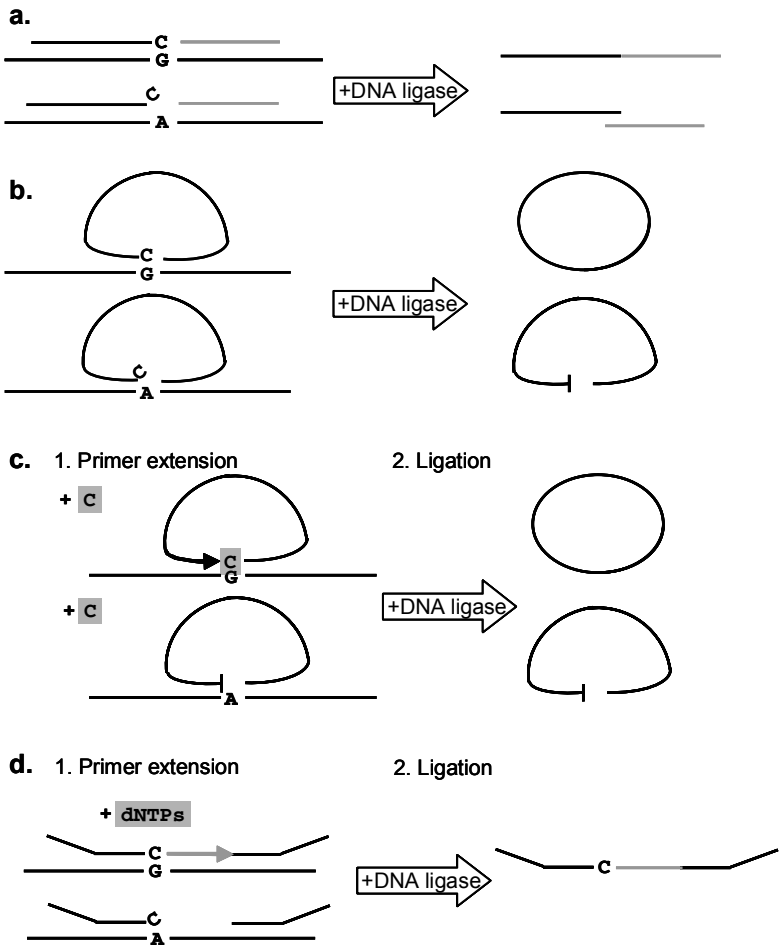


Figure 1.3 SNP genotyping by DNA ligation.

Several genotyping assays rely on the specificity of DNA ligase to ligate two ends of single-stranded DNA only if they are correctly hybridised to adjacent bases in a complementary strand.

a) In OLA two adjacent oligonucleotide probes are ligated over the SNP position.

b) The 5' and 3' ends of padlock probes are ligated over the SNP position.

c) Molecular inversion probes require a single base extension with the correct nucleotide before the probe can be ligated to form a circular molecule.

d) In the GoldenGate® assay two probes can be ligated only when the gap between them has been filled in an allele-specific extension reaction.

1.3.4 GENOTYPING BY DNA HYBRIDISATION

The genotyping methods described so far rely on enzymatic reactions. It is therefore very important that the enzyme used for each method is functioning correctly and efficiently, since an inactive enzyme may lead to incorrect genotype assignment. Internal controls, such as control samples or additional control reactions are therefore important for reliable genotyping. Genotyping is, however, possible without the use of enzymes, by using DNA hybridisation on its own.

DNA hybridisation-based methods take advantage of the difference in duplex stability of fully complementary DNA and DNA with a mismatched base pair somewhere within the duplex. Hybridisation of an allele-specific probe can be used in two alternative ways; either in static assays where the hybridisation efficiency at specific assay conditions are measured, or by recording the melting profile of the probe (**Figure 1.4**). Both kinds of hybridisation assays will be described in more detail in the following sections.

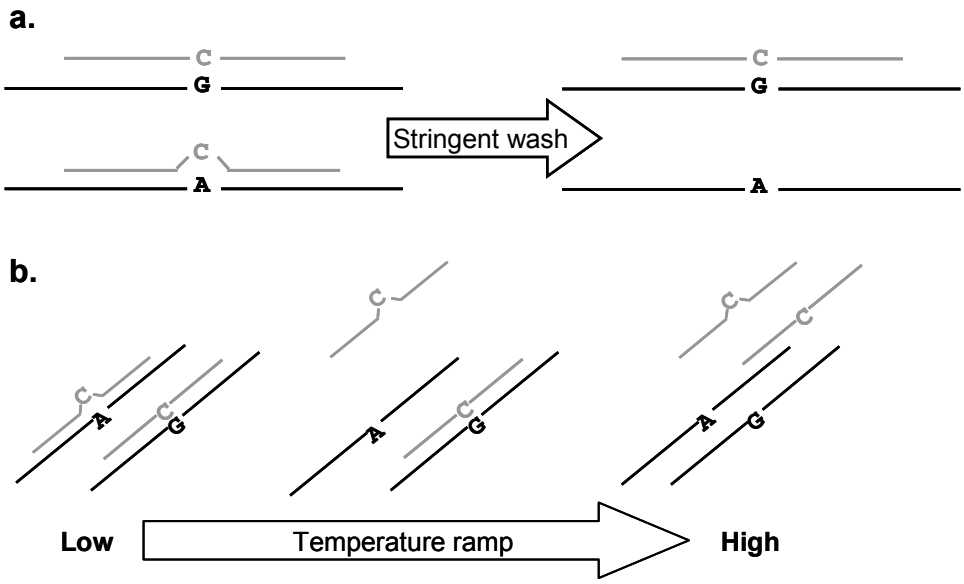


Figure 1.4 SNP genotyping by DNA hybridisation.

Allele-specific oligonucleotide probes are used to genotype SNPs, either by **a**) the specific hybridisation to one allele at stringent conditions or **b**) by recording the melting profile of the probe, where the matching allele will dissociate at a higher temperature relative to the mismatched allele.

1.3.4.1 ALLELE-SPECIFIC OLIGONUCLEOTIDE HYBRIDISATION

In allele-specific oligonucleotide hybridisation (ASO), an oligonucleotide probe is hybridised to the test DNA under stringent conditions, so that it will only hybridise to a fully complementary target, and not to a target with a single base pair mismatch. SNPs can be genotyped with two probes, one for each allele, by determining which of them hybridises best to the target (or equally well for a heterozygous sample).

This strategy was first used in "dot blot" hybridisation[119] in an array format where target DNAs are bound to separate features on a membrane and the two allele-specific probes are hybridised to the array separately. Traditionally radioactive probes have been used to record each hybridisation event, but the probes can also be fluorescently labelled. In "reverse dot blot"[120] the allele-specific oligonucleotides are bound to the membrane array and the target DNA is labelled and hybridised to the array probes under stringent conditions. In this way many sequence variants can be assayed at the same time.

The reverse method is therefore suitable for high-throughput application and is widely used on microarrays. The degree of hybridisation to each feature (fluorescence intensity) reflects the abundance of the matched sequence in the target, provided that stringent conditions are used, so that only matched targets will hybridise. For SNP genotyping and mutation detection, the fluorescence intensity of the separate allele-specific probes is therefore compared to determine the genotype. Besides providing a suitable platform for high-throughput applications, the microarray format also reduces the amount of reagents needed and thereby the cost per result. Applications range from mutation detection in single genes[121] to commercially available arrays for genome-wide SNP genotyping[122,123] where each chip contains up to several million discrete features per array. At present it is possible to genotype almost one million human SNPs on a single Affymetrix array.

Hybridisation to microarrays can also be used as a strategy to sequence DNA. More precisely, sequencing-by-hybridisation can be used for resequencing applications, for example for mutation detection. A target sequence is interrogated by hybridising it to a tiled set of probes, and determining how well a test sample is matching each of these probe sequences. One way of doing this is by a "loss of signal" approach[124], where only wildtype probes, designed from a reference sequence, are used and sequence variants

are detected as a decrease in hybridisation signal compared to the signals of a reference sample. The resolution with which the sequence variant can be located depends on the degree of overlap between probes. For the highest resolution, a complete resequencing of the target, each consecutive base is interrogated with probes for all possible bases (A, C, G, and T). This method has been used for large scale SNP discovery in mitochondrial DNA[125,126] and human genomic DNA[127-129].

Besides the array-based method, allele-specific hybridisation can also be used in homogeneous assays. One technology that uses this strategy is Molecular beacons[130]. The molecular beacon probes contain target-specific sequence in the middle, with self-complementary tails at the 5' and 3' end of the probe (**Figure 1.5 d**), so that the probe forms a hairpin structure unless it is hybridised to a perfectly matched sequence. Both ends of the probe are labelled, with a reporter fluorophore and a quencher respectively, leading to decreased fluorescence of the reporter dye when the probe adopts the hairpin structure when the energy is absorbed by the quencher. When the probe hybridises to the target sequence the reporter and quencher are separated from each other leading to an increase in fluorescence. Allelic discrimination can be done using two allele-specific probes labelled with different fluorophores.

1.3.4.2 MELTING CURVE ANALYSIS

An alternative way to use hybridisation probes for allele discrimination is to use melting profile analysis, where the sequence identity will be determined by the specific melting temperature (Tm) of the probe. In this way, one single probe can potentially genotype both alleles of a SNP. The overall benefit with DNA melting analysis compared to static hybridisation is that any sequence can be analysed under standard run conditions since the assay conditions (the heating profile) can be the same whether the Tm is 50 ºC or 80 ºC. In static hybridisation assays, on the other hand, optimisation of the stringent conditions is required to obtain allelic discrimination, making it difficult to use one set of conditions for multiple assays. Melting curves can either be recorded in a homogeneous (closed tube) reaction or with either the target DNA or the oligonucleotide probes attached to a solid support. The latter will be described in detail for the method Dynamic Allele-Specific Hybridization (DASH)[131,132] in the following section.

In homogeneous assays the oligonucleotide probe is present during the PCR amplification. After completed amplification, the melting profile is recorded by monitoring of fluorescence changes during heating in a real-time PCR machine or in a separate reader, where a rapid change in fluorescence is obtained at a specific probe-target melting temperature. This hybridisation-specific fluorescence is usually obtained by fluorescence resonance energy transfer (FRET)[133,134] between two fluorescent molecules. Energy transfer occurs when the emission spectrum of one fluorophore (the donor) overlaps with the excitation spectrum of the other fluorophore (the acceptor) and the two fluorescent molecules are located close to each other. The donor fluorophore is excited by an external light source and either the change in donor- or acceptor fluorescence (or both) can be monitored to analyse the probe melting profile. The PCR is usually made asymmetric, with higher concentration of one of the primers than the other, to produce more of one strand of the target sequence to facilitate probe binding, since the two strands of a PCR product will easily re-anneal after denaturation and compete with the probe hybridisation. Hybridisation-specific FRET assays can be designed in many different ways, and several assays are schematically described in **Figure 1.5**.

One way of using FRET is to label one of the amplification primers internally, with an acceptor fluorophore (for example Cy5) while the complementary oligonucleotide probe is labelled at the 3' end with a donor fluorophore (for example fluorescein). FRET can take place when the probe is binding to the labelled strand of the PCR product, and during the probe melting the Cy5 fluorescence will decrease and the fluorescein fluorescence increase[135] (**Figure 1.5 a**).

Another alternative is to use two adjacent probes that hybridise to the same strand. One of the probes is longer (more stable) with a donor fluorophore, while the allele-specific probe is shorter and labelled with an acceptor fluorophore[136-138]. FRET will take place as long as both oligonucleotides bind to the target sequence (**Figure 1.5 b**). The probe-target melting profile can be recorded by monitoring the fluorescence of the acceptor during heating. Since the allele-specific probe is shorter than the donor probe it will melt before the longer probe. The fluorescence change is therefore completely determined by the allele-specific probe.

A single allele-specific probe can also be used on its own. It can, for example, be labelled at both ends, such as the MGB Eclipse probes[139,140]. The 3' end is labelled with a fluorophore and the 5' end has an acceptor in form of a dark quencher molecule that does not fluoresce. When the probe is free in solution and folded in structure the fluorescence is quenched, but when it hybridises to its complementary sequence it will consequently linearise, and the fluorescence will increase (due to the increased spacing between the fluorophore and the quencher) (**Figure 1.5 c**). It may also be enough to use just one singly labelled probe. The fluorescence of fluorescein[141,142] or 6-FAM[143] on end-labelled probes is quenched if there is a guanosine nearby in the opposite strand during hybridisation (**Figure 1.5 e**). During melting, as the probe is separated from the target sequence, the fluorescence will increase.

Probe-specific melting profiles can also be obtained from unlabelled probes, if used with a DNA double-strand specific dye, such as LCGreen I[144] and SYBRGreen I[131]. This eliminates the need for expensive fluorophore labelling of the probes. The dye will fluoresce while the DNA is double-stranded and as it denatures the fluorescence will decrease (**Figure 1.5 f**).

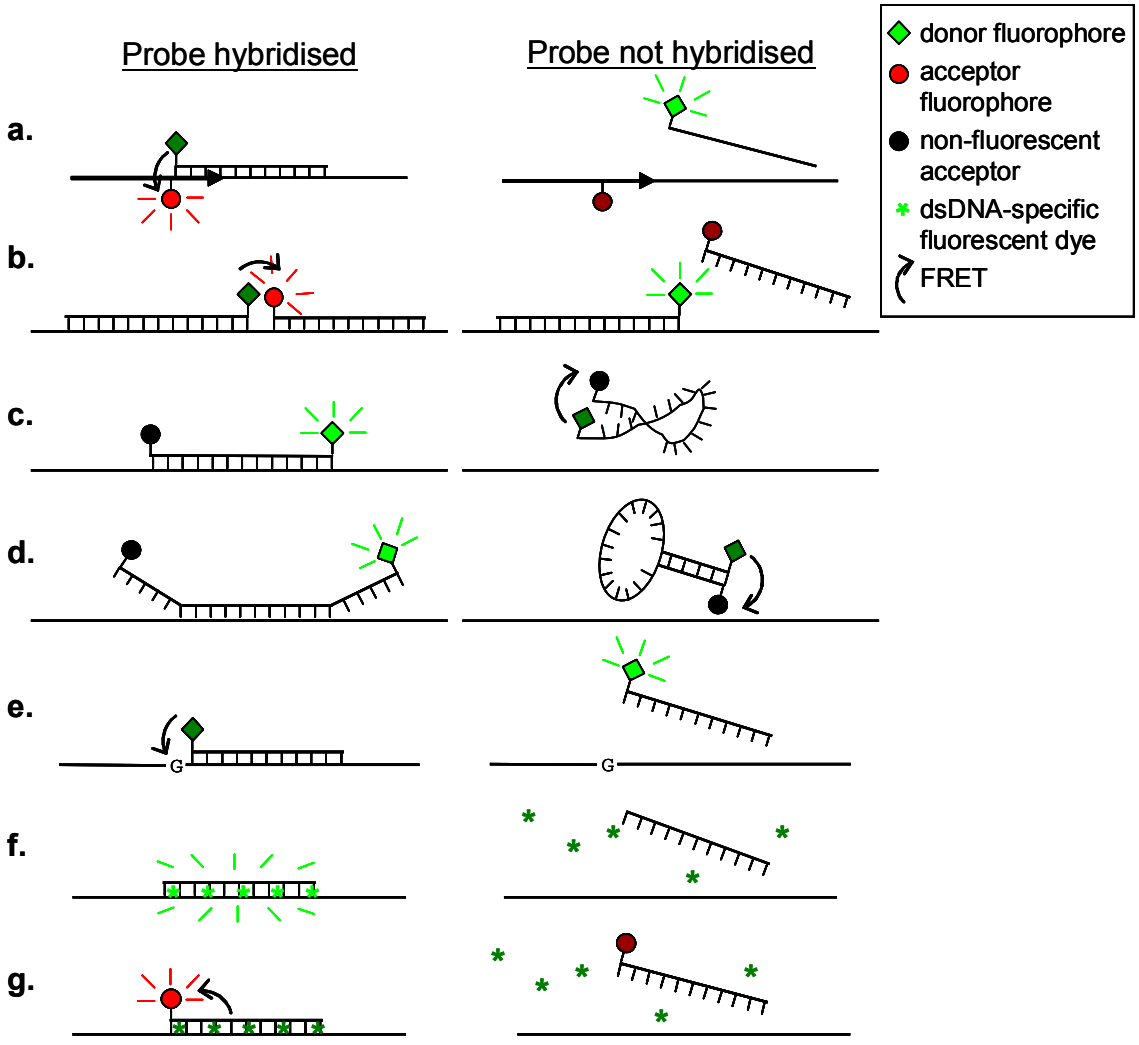


Figure 1.5 Fluorescence detection of hybridisation.

Melting-profiles of hybridisation probes can be recorded by monitoring fluorescence that changes depending on if the probe is hybridised or not. Hybridisation-specific fluorescence (or quenching) can be obtained by;

a) a FRET donor on the probe and an acceptor fluorophore in the target sequence,

b) two adjacent hybridisation probes, one with a donor fluorophore and one with an acceptor,

c) dual-labelled probes that have a donor fluorophore and a dark quencher at either end,

d) molecular beacons are also dual-labelled, and the specific hairpin structure efficiently quenches the fluorescence when the probe is not hybridised (these probes are usually not used for melt-curve analysis),

e) a single labelled probe, where the fluorescence is quenched by closely located deoxyguanosine bases (G) at the opposite strand during hybridisation,

f) DNA double-strand-specific dye,

and **g**) a double-strand-specific dye is used as a donor fluorophore to an acceptor on the probe.

1.3.4.3 DYNAMIC ALLELE-SPECIFIC HYBRIDIZATION

Dynamic Allele-Specific Hybridization (DASH)[131,132] also relies on probe-target melting profiles to determine alleles, but the target DNA is attached to a solid support. The reaction principle is described in **Figure 1.6**. The target sequence is amplified by PCR, where one of the amplification primers is modified with a 5' biotin. The PCR product is then bound to a streptavidin-coated surface by biotin-streptavidin interactions. Only one strand, carrying the biotin, is bound to the solid support, and therefore the other strand can be removed by NaOH. An allele-specific oligonucleotide probe is then hybridised to the single-stranded target at low stringency, so it hybridises to both alleles. The hybridisation is detected by fluorescence, which is recorded while heating the sample to obtain a melting profile.

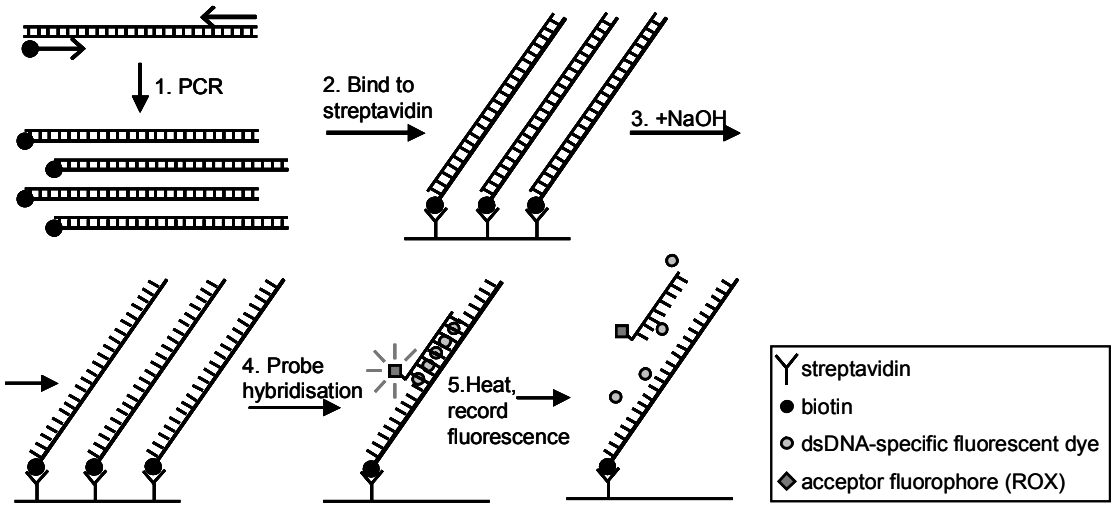


Figure 1.6 Schematic illustration of DASH.

For SNP genotyping by DASH, the target sequence, including a polymorphic base, is first amplified by PCR (1). One of the primers has a 5' biotin group, allowing the PCR product to be bound to a streptavidin-coated surface, such as a microtiter plate well (2). The non-biotinylated strand is removed by a NaOH wash (3) and an allele-specific probe is hybridised at low stringency to the remaining strand, covering the SNP position. Genotyping is done by melt-curve analysis of the probe, by recording the fluorescence while heating the sample (5).

Hybridisation-specific fluorescence can be obtained either by a DNA double-strand specific dye (such as SYBR Green I[131]), or by iFRET[145]. The iFRET system employs SYBR Green I in conjunction with a fluorescently labelled probe. The double-strand specific dye is used as a fluorescence donor that transfers energy to the fluorophore on the probe, and as a result the probe will only fluoresce while it is hybridised to the target sequence (**Figure 1.5 g**). The use of SYBR Green I on its own is obviously a less expensive option, but any double-stranded DNA will cause SYBR Green to fluoresce, which can give rise to high background signals for stable secondary structures in the target sequence. In iFRET, the non-specific double-stranded regions are less problematic, since fluorescence is only recorded specifically from the probe fluorophore. This strategy is still less expensive than the use of dual labelled probes or multiple probes per assay.

Example data for SNP genotyping is seen in **Figure 1.7**. The raw data can be viewed by plotting fluorescence intensity against temperature. The drop in fluorescence indicates probe-target dissociation. The specific target-probe Tm, where half the molecules are double-stranded, can be identified as the temperature where the slope of the curve is the steepest. The standard way to identify the Tm is to plot the negative derivative of the fluorescence (the negative slope), against the temperature. In this way distinct peaks are shown at the Tm of each probe-target duplex, clearly identifying the two alleles of a SNP. The allele that matches the probe sequence will give a high temperature peak, while the other allele will give a peak at a lower temperature, since this duplex is less stable. A heterozygous DNA sample will show two peaks, one at the high- and one at the low temperature.

DASH was initially developed for SNP genotyping in a 96-well microtiter plate format[131,146], but the method has also proven effective for scoring short insertion/deletion polymorphisms[147]. For higher throughput and to reduce assay costs, a second version of the method, DASH-2, was developed. This employs membrane arrays[132], where the sample density is scalable from 384 up to around 6000 features per membrane. For the lowest density, the samples (PCR products) are transferred directly from a 384-well plate by centrifugation transfer[148]. For higher density arrays, products from several plates can be transferred by robotic transfer to the membrane, with the samples from the different plates slightly displaced from each other.

With proper design[146,149], successful DASH assays can be developed for around 95% of attempted SNPs[149], in single-copy genomic regions (based on >3000 validated assay designs in our laboratory). These can all be processed with standard run conditions. A key factor for successful genotyping is the degree of folding of the single stranded target DNA, since a stable secondary structure can impede probe hybridisation. Such troublesome secondary structures in the target can be avoided by engineering of the PCR primers. Specifically, this is done by changing bases that are predicted to take part in stable secondary structures. This strategy is effective, since a typical DASH product is only around 50 bp long and a large proportion of the product therefore consist of primer sequences. This design strategy has been automated in the D-fold software[149].

DASH combines the benefit of qualitative data obtained by melt-curve analysis with the practical advantage of a reaction that takes place on a solid support. The array format should make it possible to adopt DASH for higher multiplex applications and parallel analysis of different mutations in a miniaturised format. These possibilities are further investigated in this thesis.

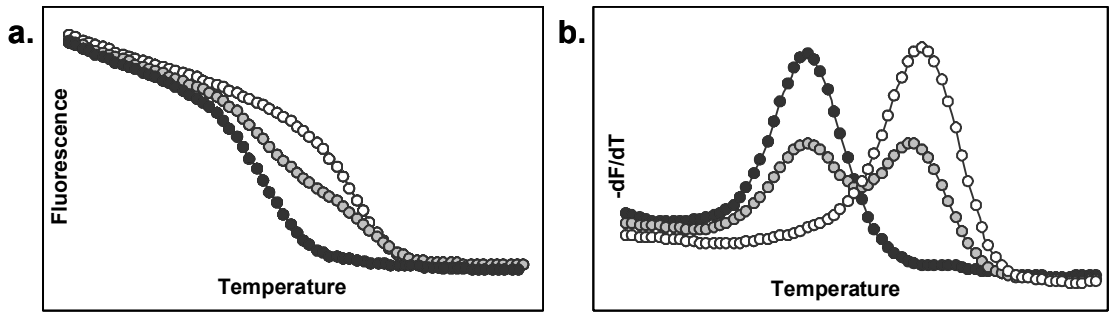


Figure 1.7 DASH example data.

Example data for SNP genotyping. **a**) The raw data consists of fluorescence intensity values at different temperatures for each sample. The three samples clearly have different melting profiles, but the differences are easier seen in **b**), where the negative derivative of fluorescence is plotted versus temperature. A DNA sample homozygous for the allele matching the probe sequence gives a peak at a higher temperature (○), compared to the sample homozygous for the other allele (●) (with a single base pair mismatch to the probe). Heterozygous targets (◍) show peaks at both temperatures.

1.4 MULTIPLEX DNA AMPLIFICATION

For most DNA sequence analysis methods, amplification of the target sequence is required prior to the actual analysis. The amplification has two purposes; to increase the concentrations of the target sequence, and to decrease the sequence complexity compared to complete genomic DNA.

The most common way to amplify DNA is by polymerase chain reaction (PCR)[150] which requires the use of two PCR primers that hybridise specifically to sequences surrounding the desired target sequence. For the parallel analysis of many different target sequences, it is desirable that all of those sequences can be amplified in a single multiplex reaction[151] rather than in many separate reactions.

There are, however, several complications to multiplex PCR. First of all the different primer pairs will have to work under the same conditions. Since the different reactions compete for the same reagents, the primers have to be designed to work equally well to obtain similar levels of all targets in the final product, or alternatively the primer concentrations need to be adjusted for each target[152-154]. The increased number of primers in multiplex reactions also increases the risk of non-specific priming events and the formation of primer-dimers[155,156]. Primer-dimers can easily take over the reaction, since their short length is advantageous when they compete for reagents with the longer target sequences[157], and they therefore present a major problem in multiplex reactions.

Primer design[156-158] as well as reaction conditions and reagent concentrations [153,155,157,159] can be optimised to improve the chances of amplification of all desired target sequences, and to reduce primer-dimer problems in standard single-tube multiplex reaction. Still, most multiplex attempts only succeed in amplifying 10-20 targets simultaneously[157,160,161]. Alternative strategies have also been explored to try and overcome the problems with multiplex PCR. One of these strategies involves the use of common primers for the amplification of all sequences, which should reduce both the risks of primer-dimer formation and the differences in amplification efficiency. Another strategy is to restrict the primers in different primer pairs from interacting with each other, by attaching them to a solid support. Several versions of both these strategies will be reviewed in the next two sections.

1.4.1 MULTIPLEX AMPLIFICATION WITH COMMON PRIMERS

The use of common primers for amplification of multiple sequences usually requires a two-stage process. To start with, the common primer sequences have to be specifically incorporated in all desired target sequences, before they can be amplified with a single primer (pair) in the second stage.

One way to introduce the common primer sequences is by a primary PCR, using primers with common 5' tail sequence (**Figure 1.8 a**). One version of this strategy uses identical tail sequences for both forward- and reverse primers, and includes the common primer (identical to the tail sequence) in the same reaction[162]. The use of just one primer also helps to eliminate primer-dimers that might have been formed in the first PCR cycles, since short products will form very stable hairpin structures that will reduce the priming efficiency. In contrast, for longer products, the effective stability of the hairpin structure is reduced and therefore amplification of the true products is possible with the common primer. An alternative procedure involves introducing two different common sequences for the forward and the reverse primer respectively in a nested multi-step reaction. In the first reaction one of the primers in each pair contains a tail sequence (T1). This sequence is used as a common primer in a second reaction, together with a nested primer containing a second tail sequence (T2). Finally, in a third reaction, the two common primers, T1 and T2, are used for amplification of all sequences[163]. The benefit of using nested reactions is that the specificity of the amplification is increased, but the drawback is that multiple reactions are required.

Another way to introduce common primer sequences is by adaptor ligation. For example, adaptor sequences can be ligated to DNA that has been cut by restriction enzymes (**Figure 1.8 b**). In this way the whole genome will be converted to fragments with common end sequences. Those can therefore be used as priming sites for common primers in multiplex reactions, but it will be difficult to select specific target sequences. This strategy can, however, be used for complexity reduction, by using specific amplification conditions that preferentially amplify fragments in a certain size range[122,123,164]. Target sequences for further analysis can be chosen from the sequences that are predicted to amplify. This strategy has been used for target amplification for up

to several hundred thousand targets for genome-wide SNP genotyping on Affymetix hybridisation arrays.

For selection of specific targets, multiplex PCR can be performed with one specific primer for every sequence and one common primer based on the adaptor sequence[160]. This strategy only reduces the number of primers in the reaction to half compared to standard multiplex PCR, but if each target-specific primer also contains a common tail sequence all sequences can be co-amplified with the same primer pair in a secondary reaction[165].

A third possibility relies on the specific formation of circular molecules, followed by enzymatic selection of the circular DNA and subsequent amplification with a common primer pair. For amplification of just a single base, such as a SNP, molecular inversion probes[116] can be used. The genotyping strategy with these probes has been described in section **1.4.3**. In short, the two ends of the long probe are complementary to the target sequence and hybridise next to each other, with one base pair gap. The probes can therefore only be circularised by ligation if the gap is filled by incorporation of the correct nucleotide. The central sequence of the probe includes common priming sites, allowing amplification of multiple targets with a single primer pair after selection for circular molecules. In this way over 10,000 SNPs have been amplified in a single tube assay[117].

In order to amplify a longer target sequence, a similar concept was used to develop selector probes[166]. The selector probes have a central double-stranded common sequence, including common priming sites, where one of the strands also has target-specific overhang sequences at both the 5' and 3' ends. These overhangs should hybridise to either end of a restriction enzyme digested fragment, allowing the target sequence to be directly ligated to the selector probe and form a circular molecule (**Figure 1.8 c**).

In one final related method, Callow et al[167] used combinations of different universal double-sided adaptors with common priming sites and combinations of 5' overhangs that are ligated to restriction enzyme digested DNA to form circular molecules (**Figure 1.8 d**). Those can be selected for and amplified by a common primer pair. By using Type II restriction enzymes, which cut outside the recognition sequence, most

fragments will have a different combination of overhangs and the adaptors can be chosen to fit the desired fragments. For complex genomes, such as the human genome, two rounds of selection would be required for selection and amplification of specific sequences. This strategy was only demonstrated for a few targets simultaneously and it is uncertain how this strategy would work for highly multiplex reactions.

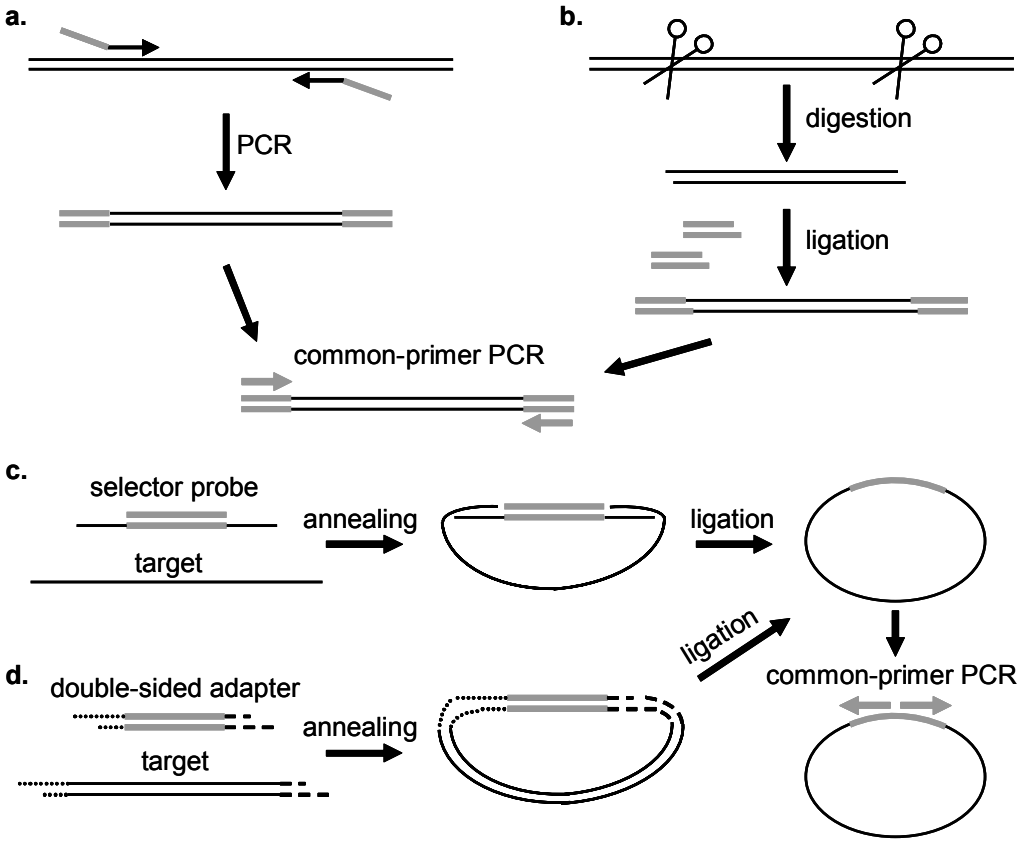


Figure 1.8 Incorporation of common primer sequences.

Introducing common priming sites allows multiple fragments to be amplified simultaneously with a single primer pair. The universal sequences (shown in grey) can be introduced by;

a) A primary PCR, where the primers have a common tail at the 5' end.

b) Digesting the DNA with a restriction enzyme and then ligate common adapters to the cut DNA.

c) Ligation of restriction enzyme digested DNA targets to selector probes, which contain common primer sequences. Correct ligation results in circular molecules, which are enriched by digesting linear DNA.

d) Ligation of Type II restriction enzyme digested DNA targets to double-sided adaptors, which contain common primer sequences. Correct ligation results in circular molecules, which are enriched by digesting linear DNA.

1.4.2 SOLID-PHASE PCR

Interactions between primers in different primer pairs can be avoided if the individual reactions can be physically separated within the same reaction vessel. One way of doing this is by solid-phase PCR, where the primer pairs are attached to distinct locations on some solid support and the amplification reaction takes place directly on the surface. For efficient annealing and extension, the primers should be immobilised at their 5' ends, leaving the 3' ends free for polymerisation. Solid-phase PCR has been described both with one and both of the primers attached to solid support, for slightly different applications and possibilities for multiplexing.

1.4.2.1 ONE PRIMER ATTACHED

Solid-phase primer extension reactions were initially developed to facilitate detection of amplification products. If the PCR product is directly attached to a solid support during amplification it will be ready for further analysis directly after the completion of the reaction. By attaching one of the primers to a solid support, one of the product strands can be directly synthesised on the surface (in addition to having both primers present in the solution phase). The solution-phase reaction will create more and more of the target sequence and thereby promote annealing and extension of the surface-bound primer. To increase the chance of solid-phase primer extension the solution reaction can be asymmetric with more of the primer for the complementary strand[168-171]. For direct fluorescent detection of the solid-phase product, it can be labelled by incorporation of labelled nucleotides[170,172] that can be quantified after the amplification reaction. Alternatively, the products can be detected by hybridisation of labelled oligonucleotide probes[168,171]. For increased specificity and detailed analysis of the product, the solid-phase primers can be nested to anneal internally to the solution-phase PCR product. Several allele-specific probes could also be used in parallel, for multiplex product detection, for applications such as SNP genotyping[173,174] and for the identification of different bacterial species[175].

A simultaneous solution-phase reaction increases the efficiency of the solid-phase reaction, since more targets for the solid-phase primer are constantly produced. This is advantageous if the purpose of the solid-phase reaction is only to attach one or few

products to a solid support for easy detection. However, it is not suitable for higher multiplex applications, since the many primers in solution can easily interact to create primer-dimers and other non-specific products, as already discussed.

1.4.2.2 BOTH PRIMERS ATTACHED

The potentials for high multiplexing would increase if both primers in each pair can be attached to the solid support. In this way each reaction can take place independent from any other reactions, if the different primer pairs are attached to separate areas.

Solid-phase amplification with both primers immobilised to a solid support was first reported about a decade ago[176], when "bridge amplification" was described with primers attached to a range of different kinds of beads. It was called bridge amplification because of the structure created when a double-stranded product is formed between two attached primers, thus creating a bridge between them. Solid-phase amplification was initially demonstrated with purified PCR product as template DNA, and then with human genomic DNA template. Evidence that the reaction took place on the surface was achieved by introducing separate restriction enzyme sites in the two primer sequences and comparing the amount of product released with either or both enzymes after amplification. The majority of the products were only released after treatment with both enzymes, indicating that they were attached to the surface with both ends.

Solid-phase amplification was also demonstrated on glass slides[177], where PCR products were used as DNA template and amplification products were detected by radioactively labelled probes. That particular study also compared different covalent surface chemistries for their stability, and found that a maximum of only 60% of the primers were still attached after thermo-cycling. The release of primers into solution could potentially be a problem for multiplex reactions, since this would mean that part of the amplification takes place in the solution phase. This goes totally against the objective of preventing primer-primer interactions in the solution phase.

Another support for solid-phase primers is polyacrylamide gel, in the form of gel element arrays or gel particles in solution. The use of gels as support material has the benefit that it forms a 3D structure rather than a flat surface and thereby provides a larger reaction area where the primers can interact with target DNA. A 57-plex solid-

phase PCR was demonstrated with the primers attached to separate acrylamide beads, where products between 70 and 1300 base pairs were amplified[178]. The products were separated from the beads by restriction enzyme digestion and visualised by gel electrophoresis. Polyacrylamide can also be formed into nanoliter gel pads on a microarray format, with primer pairs attached to distinct gel pads[179]. After hybridisation of DNA the gel pads were separated from each other by mineral oil, enabling one of the primers in each pair to be detached within the pad to enhance the amplification rate. For direct genotyping within the gel pads, a dormant allele-specific primer was present in each gel pad. This could be activated after the amplification reaction to carry out an allele-specific extension reaction.

Further development of the gel element array resulted in a method called multiplex microarray-enhanced PCR (MME-PCR)[180]. The gel-bound primers were then divided in two segments; the 5' end of the primers contain either of two universal primer sequences, and the 3' end of each primer is target-specific. In order to increase the efficiency of the solid-phase reaction a common primer pair, consisting of the universal primer sequences, is also present in the solution phase. This produces additional template molecules for priming by the surface-bound primers. The surface-bound products were detected with fluorescent-labelled hybridisation probes. With this concept, multiplex amplification with six primer pairs, each replicated in ten gel elements, was demonstrated.

1.5 METHOD CHALLENGES

Despite the efforts in method development in recent years, which have resulted in many different methods for DNA sequence analysis, there is still need for improvement. This thesis is therefore focussed on method development, and two related but distinct areas were chosen for this work. The first area is mutation scanning by resequencing, where an ideal method should be able to detect sequence variants in any sequence context. This sensitivity is particularly important for diagnostic applications. The second area is the simultaneous amplification of many targets, so that one can access specific parts of the genome, for further analysis of these sequences. There is a wide range of highly parallel methods for DNA sequencing and genotyping, but no ideal method for isolating specific targets for these analyses.

1.5.1 METHOD DEVELOPMENT GOALS FOR DNA DIAGNOSTICS

A useful method for DNA diagnostics should be flexible enough to both scan for novel mutations and to detect any combination of known variant positions in the genome. Hybridisation-based assays should be powerful enough to achieve this, if the hybridisation is done dynamically. Since hybridisation-based assays require no enzymatic reactions, they should also be cheaper to run than many other methods.

Since DASH[131,132] combines dynamic hybridisation with the practical advantages of being a surface-based method, it should allow highly parallel analysis on arrays, under standard run conditions. This method should therefore be attractive as the basis for new approaches to DNA diagnostics. Until the start of this project, DASH had only been used to genotype SNPs and short insertion/deletion polymorphisms, in a format where one particular variant was analysed in multiple samples in parallel. In DNA diagnostics, the goal would be to score any different sites in the same sample, thus implying the need to develop the method further to meet these requirements.

To use DASH for wider applications in DNA diagnostics, several aspects of the method have to be considered. It will be necessary to evaluate different assay formats for the feasibility of parallel recording of multiple probe-target interactions, and assess how DASH can be implemented in a microarray format. Several general parameters also

have to be optimised, such as the length of the oligonucleotide probes, the probe spacing for mutation detection, and the possibilities for multiplexing and the limitations of the method have to be considered. Proof-of-principle data will have to be produced to demonstrate how mutation detection and resequencing can be done in practice.

1.5.2 METHOD DEVELOPMENT GOALS FOR MEGAPLEX PCR

To facilitate highly parallel (re)sequencing, a method for highly multiplex amplification of specific DNA fragments need to be developed. MegaPlex PCR should be developed to address this challenge.

As discussed in section **1.4.1**, a powerful way to amplify multiple sequences in a single tube is to add common adapter sequences to all desired targets for amplification with a single primer pair. The challenge, however, is how to add these adaptor sequences to any given set of target sequences in a way that selects all desired fragments and avoids non-specific sequences. The strategy chosen for MegaPlex PCR was to use a solid-phase reaction as a way to introduce these common sequences. The physical separation of different primer pairs in solid-phase PCR should avoid unwanted primer-primer interactions and reduce the risks of amplification of false products, since each isolated location basically promotes a single-plex reaction.

The development of MegaPlex PCR will require optimisation of several parameters. Initially, solid-phase amplification should be demonstrated. Solid-phase primer extension has to be shown both from targets in solution as well as from internal priming from a solid-phase product. Reaction conditions have to be optimised, such as what reaction surfaces to use and the concentrations of different reagents and reaction times. Ideally the amplification should be demonstrated directly from genomic DNA template, and proof-of principle data should be produced for a multiplex application.

CHAPTER 2.

MATERIALS AND METHODS

2.1 DASH DIAGNOSTICS

2.1.1 DNA SAMPLES

Human DNA samples were from 16 Swedish females, and these were previously prepared by standard phenol-chloroform extraction procedures. These samples were part of a larger collection of samples, which were collected from anonymous blood donors. All individuals gave written informed consent to donate their blood anonymously for research.

Salmonella DNA from 62 *Salmonella enterica* isolates was received from the Health Protection Agency's Salmonella Reference Unit. DNA from these strains had been prepared from 24-hour cultures using a Qiagen DNeasy Tissue kit (Qiagen, West Sussex, UK) according to manufacturer's instructions. Additionally, DNA from seven of the strains were also prepared using a simpler method: a single colony was resuspended in 100 µl distilled water with 15% (w/v) Chelex 100 Molecular Biology Grade resin (Biorad, Hertfordshire, UK) and boiled for 10 minutes. The cell suspension was then centrifuged for 5 minutes at 13,000 rpm and the supernatant removed and stored at -20 ºC until required.

2.1.2 OLIGONUCLEOTIDES

PCR primers and probes were obtained from Thermo Electron GmbH and Biomers.net GmbH. Primers were designed either with the Dfold software[149] or using the OLIGO software (Molecular Biology Insights, Inc, USA). Basic rules for primer design were primer length restricted to 20-24 bp, primer Tm difference < 5 ºC for each primer pair, and a maximum of 3 bp complementarity to the 3' end of any primer. All DASH probes were 17 bp. All PCR primers and DASH probes for the human- and Salmonella sequences are listed in **Table 2.1**.

Table 2.1 Oligonucleotides for human and Salmonella sequences.

Name	Modification	Sequence	Function
Human target			
LSCAN-01F		TAACTTACTAGGAGCTTTTAATGG	primer
LSCAN-02F		TCATTGTAGACTGTCTTAATTGC	primer
LSCAN-03F		TCTATCTGTCTTACCTCATCACC	primer
LSCANb04R	5'biotin	AATAGAGTGAAATGTATGATTGG	primer
LSCANb05R	5'biotin	TATTTGCTAATCTCTGAGAAGGC	primer
LSCANb06R	5'biotin	CTGTCAGTTTTACACAATTCATC	primer
LSCANb22R	5'biotin	ATATTAGCCTCATATTTGGGAAG	primer
LSCAN-23F		AGATGGGTAAGAAATGTAGTTAG	primer
LSCAN-24F		TAAAACAAGACTACGTGTCACTG	primer
LSCAN+07P	3'ROX	ATTTATCCGTCTTGCCA	probe
LSCAN+08P	3'ROX	TCCGTCTTGCCATATGA	probe
LSCAN+09P	3'ROX	CTTGCCATATGACTAAG	probe
LSCAN+10P	3'ROX	CATATGACTAAGCAGAT	probe
LSCAN+11P	3'ROX	GACTAAGCAGATCAACA	probe
LSCAN+12P	3'ROX	AGCAGATCAACAAATAC	probe
LSCAN+13P	3'ROX	ATCAACAAATACCTATC	probe
LSCAN+14P	3'ROX	CAAATACCTATCACTCA	probe
LSCAN+15P	3'ROX	ACCTATCACTCACACAC	probe
LSCAN+16P	3'ROX	TCACTCACACACTAGAA	probe
LSCAN+17P	3'ROX	CACACACTAGAAATGAA	probe
LSCAN+18P	3'ROX	ACTAGAAATGAAACAAG	probe
LSCAN+19P	3'ROX	AAATGAAACAAGCTTTA	probe
LSCAN+20P	3'ROX	AAACAAGCTTTAATGCA	probe
LSCAN+21P	3'ROX	AGCTTTAATGCACACAG	probe
Salmonella target			
LGyrAb01R	5'biotin	CAGCATGTAACGCAGCGAGAAT	primer
LGyrA-02F		GCAATGACTGGAACAAAGCCTA	primer
LGyrAb03R	5'biotin	GCTGCGCTATACGAACGATG	primer
LGyrA-04F		CAATGACTGGAACAAAGCCTATAA	primer
LGyrAb05R	5'biotin	CCGTACCGTCATAGTTATCCAC	primer
LGyrA-06F		CAGGGTAACTTCGGTTCTATTG	primer
LGyrAb07R	5'biotin	CCGTACAGTCATAGTTATCCACG	primer
LGyrA-08F		GTTCTATTGAAGGCGACTCCG	primer
GyrA1+01p	3'ROX	AAAAATCTGCCCGTGTC	probe
GyrA1+02p	3'ROX	AAAAATCTCCCCGTGTC	probe
GyrA2+01p	3'ROX	CGTTGGTGACGTAATCG	probe
GyrA2+02p	3'ROX	CGTTGGTGGCGTAATCG	probe
GyrA3+01p	3'ROX	CCACGGCGATTCCGCAG	probe
GyrA3+02p	3'ROX	CCACTGCGATTCCGCAG	probe
GyrA3+03p	3'ROX	CCACAGCGATTCCGCAG	probe
GyrA3+04p	3'ROX	CCACGGCAATTCCGCAG	probe
GyrA3+05p	3'ROX	CCACGGCGGTTCCGCAG	probe
GyrA3+06p	3'ROX	CCACGGCGATTTCGCAG	probe
GyrA3+07p	3'ROX	CCACGGCGATTACGCAG	probe
GyrA3+08p	3'ROX	CCACGGCGATGCCGCAG	probe
GyrA4+01p	3'ROX	CAGTGTATGACACCATC	probe
GyrA4+02p	3'ROX	CAGTGTATGGCACCATC	probe
GyrA4+03p	3'ROX	CAGTGTATAACACCATC	probe
GyrA4+04p	3'ROX	CAGTGTATTACACCATC	probe
GyrA5+01p	3'ROX	CGCGGCGGCAATGCGTT	probe
GyrA5+02p	3'ROX	CGCGGCGGTAATGCGTT	probe
GyrA5+03p	3'ROX	CGCGGCGTCAATGCGTT	probe
GyrA5+04p	3'ROX	CGCGGCGGAAATGCGTT	probe
GyrA6+01p	3'ROX	GAAAATCGCCCACGAAC	probe
GyrA6+02p	3'ROX	GAAAATCGGCCACGAAC	probe
GyrA7+01p	3'ROX	CGATCTCGAAAAAGAGA	probe
GyrA7+02p	3'ROX	CGATCTCGCAAAAGAGA	probe
GyrA8+01p	3'ROX	ACGGTGGATTTCGTGGA	probe
GyrA8+02p	3'ROX	ACGGTGGACTTCGTGGA	probe

2.1.3 PCR

PCR reactions of 5-20 µl volume contained 0.375 µM 5'-biotinylated primer, 0.75 µM of non-labelled primer, 0.03 units/µl AmpliTaq Gold® DNA polymerase (Applied Biosystems), 1x AmpliTaq Gold® Buffer, 1.5 mM MgCl2, 5% Dimethylsulphoxide (DMSO), and 0.2 mM of each dNTP. Human genomic PCRs contained 0.1-0.5 ng DNA per µl. Salmonella PCRs contained typically 0.5-2.5 ng DNA per µl.

Thermal cycling was performed either on a 96-well (for DASH-1), or on a 384-well (for DASH-2) MultiBlock System (Thermo Electron Corporation). Thermal-cycling consisted of an initial activation step of 94°C for 10 minutes, followed by 40 cycles (30 cycles for Salmonella targets) of 94°C for 15 seconds and an assay-specific annealing temperature for 30 seconds (se **Table 2.2**). For longer PCR products (>200 bp), thermal-cycling consisted of an initial 10-minute activation step of 94°C for 10 minutes, followed by 40 cycles of 94°C for 15 seconds and an assay-specific annealing temperature for 30 seconds and 72°C for 1 minute.

Table 2.2 Assay-specific annealing temperatures

Primer F	Primer R	Annealing temperature	Length
human sequence			
LSCAN-01F	LSCANb04R	55 °C	138 bp
LSCAN-23F	LSCANb04R	58 °C	198 bp
LSCAN-01F	LSCANb22R	58 °C	240 bp
LSCAN-23F	LSCANb22R	58 °C	300 bp
LSCAN-24F	LSCANb04R	58 °C	386 bp
LSCAN-24F	LSCANb22R	58 °C	488 bp
LSCAN-02F	LSCANb04R	55 °C	533 bp
LSCAN-01F	LSCANb05R	55 °C	557 bp
LSCAN-02F	LSCANb05R	58 °C	952 bp
LSCAN-01F	LSCANb06R	55 °C	968 bp
LSCAN-03F	LSCANb04R	53 °C	1073 bp
LSCAN-02F	LSCANb06R	53 °C	1363 bp
LSCAN-03F	LSCANb05R	53 °C	1492 bp
LSCAN-03F	LSCANb06R	58 °C	1903 bp
gyrA sequence			
LGyrA-04F	LGyrAb03R	55 °C	116 bp
LGyrA-08F	LGyrAb07R	55 °C	129 bp

2.1.4 DNA SEQUENCING

To sequence the human genomic DNA fragment relevant to our study, 50 µl PCR products were purified using a MinElute PCR Purification Kit (Qiagen) following the manufacturer's protocol. Final elution of the PCR product was in water. These PCRs used primers LSCAN-23F and LSCANb22R to amplify a 300 bp fragment encompassing the 97 bp region of interest. Eluted products were approximately quantified on a 2% agarose gel by reference to a known amount of size standard DNA, and ~25 ng was used as template for cycle sequencing that employed a BigDye 3.1 Terminator Kit (Applied Biosystems). For this, 20 µl volume reactions were heated in a 96-well MultiBlock System (Thermo Electron Corporation) to 96°C for 1 minute followed by 25 cycles of 96°C for 10 seconds, 50°C for 5 seconds and 60°C for 4 minutes. The temperature transition rate was 1°C/second. Two reactions were performed on each PCR product using either one of the initial amplification primers at 0.16 µM concentration, so that both strands of the PCR product were examined. Sequencing products were purified using DyeEx 2.0 spin columns (Qiagen) following the manufacturer's protocol and the samples directly loaded onto a 48-capillary 3730 DNA Analyzer (Applied Biosystems) without drying the sample.

2.1.5 DASH

DASH-1

DASH-1 was conducted as described previously [131]. In brief, PCR products were diluted 1:1 with HEN buffer (0.1 M Hepes, 10 mM EDTA, 50 mM NaCl at pH 7.5) and 20 µl was added to each well of a 96-well streptavidin-coated microtiter plate. Product binding was allowed to proceed for a minimum of 1 hour at room temperature. The solution was then removed and the wells were rinsed once with 25 µl of 0.1 M NaOH to elute the unbound (non-biotinylated) strand of the PCR product. Then a 25-µl solution containing HEN buffer with 15 pmol of the appropriate allele-specific probe was added. The microtiter plate was sealed, heated to 85°C, and air-cooled to 25°C over ~5 minutes enabling the probe to hybridise to the bound PCR product (regardless of which alleles were present). The solution was replaced with HEN buffer containing SYBRGreen I dye (Molecular Probes) at a 1:10,000 dilution. Plates were read in a DASH instrument

(Thermo Hybaid, though any Q-PCR machine would suffice) and fluorescence was recorded while heating from 35°C to 85°C at a rate of 0.3°C/second.

DASH-2

DASH-2 was conducted in a manner similar to that previously described [132], with a few changes that enabled the diagnostic application.

PCR products were first transferred from a 384-well microtiter plate to a streptavidin-coated polypropylene membrane via centrifugation as previously described [148]. To achieve this, the membrane (DynaMetrix Ltd, UK) was pre-moistened in HE buffer (0.05M HEPES, 5mM EDTA, pH 7.5) and placed over the open wells of the microtiter plate. The arrangement was compressed in a clamping device, inverted, and centrifuged at 1500 rpm for 30 seconds in a suitable device (B4i Jouan, Inc, S20 rotor). After sitting at room temperature for 30 minutes (to allow biotinylated PCR products to bind to the membrane) the clamped structure was inverted and briefly centrifuged to return the bulk fluid into the microtiter plate wells. The membranes were then rinsed once in a 0.1 M NaOH bath for 2 minutes to remove non-biotinylated PCR product strands, and once in HE for neutralisation.

To apply different probes to distinct locations on the same membrane, 10 pmol/µl of appropriate probe solution in HE buffer was placed in the matching well of a 384-well plate, and this was transferred to cover the membrane area where PCR product had been bound using the same clamping device and centrifuge as described above. Excess probe solution was immediately spun back to the wells of the plate. The membrane was then recovered from the clamp and placed in a sandwich of two 8x12 cm glass plates (slightly larger than the membrane) so forming a hybridisation chamber. This was heated to 85°C on a flat PCR block (PCR express, Thermo Electron Corporation) and air cooled to room temperature to assist probe annealing. A final rinse was performed in HE buffer to remove excess probe. The plate/array layout of samples and probes for the GyrA genotyping is presented in **Appendix 1**.

To execute the dynamic melt procedure, membranes were soaked for 1 hour in HE buffer containing a 1:20,000 dilution of supplied stock SYBR Green I dye (Molecular Probes), and they were then individually sandwiched between two glass plates and

placed into a DASH-2 genotyping device (DynaMetrix Ltd). Fluorescence images and feature intensity values were collected while heating the membrane assembly from 35–85°C (with a heating rate of 3°C/minute) by imaging every 0.5°C.

Melting-curve analysis

Output data files from the dynamic melt procedures were imported into purpose built software (DynaScore, DynaMetrix Ltd). Using this tool, melt-curves were examined for each microtiter plate well (DASH-1) or array feature (DASH-2) and probe-target denaturation events were visualised by plotting negative derivatives curves of the fluorescence signal versus temperature. A single high-temperature peak indicated the sample was completely matched to the probe sequence. A single low temperature peak indicated a one or two bp mismatch compared to the probe sequence (two base mismatches cause much larger decreases in melting peak temperature, and >2 base mismatched fall below the temperature window examined). If peaks were seen at both temperatures this indicated that two alleles were present in the PCR product, such as would occur with a human heterozygous sample.

2.2 RESEQUENCING BY DYNAMIC HYBRIDISATION

2.2.1 DNA SAMPLES

Thalassaemia reference samples were obtained from George P. Patrinos at the Erasmus University Medical Center, Rotterdam, The Netherlands.

2.2.2 OLIGONUCLEOTIDES

All oligonucleotides were obtained from Biomers.net GmbH. The oligonucleotides used for the *HBB* gene, PCR primers and probes for reverse-DASH, are listed in **Table 2.3**. Primers were designed using the Primer3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3.cgi).

For asymmetric PCRs the primer Tm was calculated separately by the OligoAnalyzer 3.0 software (http://www.idtdna.com/analyzer/ Applications/OligoAnalyzer/Default.aspx) which uses the nearest-neighbour model[181] and salt correction[182]. The calculations

were done for primer concentrations of 1 μM for excess primers and 0.1 μM for limiting primers and the Na+ concentration was set to 50 mM.

Table 2.3 Primers and probes for HBB.

Name	Sequence	Modification	Function
HBB1-01F	GAAGTCCAACTCCTAAGCCAGT		primer
HBB1+01F	GAAGTCCAACTCCTAAGCCAGT	5'Cy3	primer
HBB1-01R	AATCATTCGTCTGTTTCCCATTCT		primer
HBB1-02R	TGCAATCATTCGTCTGTTTCCCATTCT		primer
HBB1-03R	CAGTAACGGCAGACTTCTCCT		primer
HBB1-04F	TGAAGTCCAACTCCTAAGCCAGTGC		primer
HBB1-04R	GGGCAGTAACGGCAGACTTCTCCT		primer
HBB1-05F	TAGAAACTGGGCATGTGGAGAC		primer
HBB1+05F	TAGAAACTGGGCATGTGGAGAC	5'Cy3	primer
HBB1-06F	ACCAATAGAAACTGGGCATGTGGAGAC		primer
HBB1-06R	CCCAAAGGACTCAAAGAACCTC		primer
HBB1-07R	AGATCCCCAAAGGACTCAAAGAACCTC		primer
HBB2-01F	GGGACCCTTGATGTTTTCTTTCC		primer
HBB2-01R	GCCTAGCTTGGACTCAGAATAATCC		primer
HBB2-02R	GGGCCTAGCTTGGACTCAGAATAATCC		primer
HBB3+01F	CATGCCTCTTTGCACCATTCTA	5'Cy3	primer
HBB3-01F	CATGCCTCTTTGCACCATTCTA		primer
HBB3-02R	TGCAGCCTCACCTTCTTTCATGG		primer
HBB_c93-22_A_G_25	GGAAAATAGACCGATAGGCAGAGAG	5'ROX, 3'biotin	probe
HBB_c93-22_A_A_25	GGAAAATAGACCAATAGGCAGAGAG	5'ROX, 3'biotin	probe
HBB_c93-22_A_C_25	GGAAAATAGACCCATAGGCAGAGAG	5'ROX, 3'biotin	probe
HBB_c93-22_A_T_25	GGAAAATAGACCTATAGGCAGAGAG	5'ROX, 3'biotin	probe
HBB_c93-21_A_G_25	GGGAAAATAGACGAATAGGCAGAGA	5'ROX, 3'biotin	probe
HBB_c93-21_A_A_25	GGGAAAATAGACAAATAGGCAGAGA	5'ROX, 3'biotin	probe
HBB_c93-21_A_C_25	GGGAAAATAGACCAATAGGCAGAGA	5'ROX, 3'biotin	probe
HBB_c93-21_A_T_25	GGGAAAATAGACTAATAGGCAGAGA	5'ROX, 3'biotin	probe
HBB_c93-20_A_G_25	TGGGAAAATAGAGCAATAGGCAGAG	5'ROX, 3'biotin	probe
HBB_c93-20_A_A_25	TGGGAAAATAGAACAATAGGCAGAG	5'ROX, 3'biotin	probe
HBB_c93-20_A_C_25	TGGGAAAATAGACCAATAGGCAGAG	5'ROX, 3'biotin	probe
HBB_c93-20_A_T_25	TGGGAAAATAGATCAATAGGCAGAG	5'ROX, 3'biotin	probe
HBB_c117_A_G_25	AAAGAACCTCTGGGTCCAAGGGTAG	5'ROX, 3'biotin	probe
HBB_c117_A_A_25	AAAGAACCTCTGAGTCCAAGGGTAG	5'ROX, 3'biotin	probe
HBB_c117_A_C_25	AAAGAACCTCTGCGTCCAAGGGTAG	5'ROX, 3'biotin	probe
HBB_c117_A_T_25	AAAGAACCTCTGTGTCCAAGGGTAG	5'ROX, 3'biotin	probe
HBB_c118_A_G_25	CAAAGAACCTCTGGGTCCAAGGGTA	5'ROX, 3'biotin	probe
HBB_c118_A_A_25	CAAAGAACCTCTAGGTCCAAGGGTA	5'ROX, 3'biotin	probe
HBB_c118_A_C_25	CAAAGAACCTCTCGGTCCAAGGGTA	5'ROX, 3'biotin	probe
HBB_c118_A_T_25	CAAAGAACCTCTTGGTCCAAGGGTA	5'ROX, 3'biotin	probe
HBB_c119_A_G_25	TCAAAGAACCTCGGGGTCCAAGGGT	5'ROX, 3'biotin	probe
HBB_c119_A_A_25	TCAAAGAACCTCAGGGTCCAAGGGT	5'ROX, 3'biotin	probe
HBB_c119_A_C_25	TCAAAGAACCTCCGGGTCCAAGGGT	5'ROX, 3'biotin	probe
HBB_c119_A_T_25	TCAAAGAACCTCTGGGTCCAAGGGT	5'ROX, 3'biotin	probe

2.2.3 PCR

Standard (symmetric) PCR reactions contained 0.75 µM of each primer, 0.03 units/µl AmpliTaq Gold® DNA polymerase (Applied Biosystems), 1x AmpliTaq Gold® Buffer, 1.5 mM MgCl2, 5% DMSO, and 0.2 mM of each dNTP and approximately 20 ng human genomic DNA. Cycling conditions were 94°C for 10 minutes, followed by 30 cycles of 94°C for 15 seconds, an assay-specific annealing temperature for 30 seconds and extension at 72°C. The extension time was 2 minutes for the whole-gene product, 1 minute for products of 600-800 bp and 20 seconds for shorter products.

Asymmetric PCR reactions contained 1 µM of the excess primer and 0.05 µM of the limiting primer, primary PCR product (HBB1-01F/HBB3-02R) diluted to 10^{-6} of the original concentration, all other reagents were present at the same concentration as in the standard PCR. Cycling conditions were 94°C for 10 minutes, followed by 60 cycles of 94°C for 15 seconds, 60°C for 30 seconds and 72°C for 20 seconds.

2.2.4 ARRAY HYBRIDISATION

Array design

Three regions of the *HBB* gene were represented on the GENIOM array. Each position was interrogated by eight 25-bp probes, four probes each for the forward- and reverse strand, with four alternative alleles (A, C, G, T) at the central position.

Probe Tm:s were calculated using a web based software, dnaMATE: http://protein.bio.puc.cl/cardex/servers/dnaMATE/tm-pred.html[183].

Hybridisation

The PCR products were labelled with Cy3, either with labelled primers or by using Cy3-dCTP in the PCR (**Table 2.4**).

Cy3 labelled versions of HBB1-01F, HBB1-05F and HBB3-01F were used to amplify three short fragments of the HBB gene, both using symmetric and asymmetric PCR. The PCR conditions were the same as described above. Only one fragment was labelled with Cy3-dCTP and here 50% of the dCTPs were replaced by Cy3-dCTP, and the total

concentration of all dNTPs was reduced to 0.1 mM, otherwise the PCR conditions were unchanged. Each product was purified with MinElute spin columns (QIAGEN) and eluted in 10ul H_2O.

Table 2.4 Cy3-labelled PCR products.

Primer ratio	Primer F	Primer R	product	Annealing	Cycles	Labelling
20:1	HBB1+01F	HBB1-04R	264 bp	60°C	60	Cy3 primer
20:1	HBB1+05F	HBB1-07R	145 bp	60°C	60	Cy3 primer
20:1	HBB3+01F	HBB2-02R	209 bp	60°C	60	Cy3 primer
1:1	HBB1+01F	HBB1-03R	261 bp	58°C	30	Cy3 primer
1:1	HBB1+05F	HBB1-06R	140 bp	58°C	30	Cy3 primer
1:1	HBB3+01F	HBB2-01R	207 bp	58°C	30	Cy3 primer
20:1	HBB1-01F	HBB1-04R	264 bp	60°C	60	Cy3-dCTP
1:1	HBB1-01F	HBB1-04R	264 bp	58°C	30	Cy3-dCTP

2.2.5 REVERSE-DASH ON MICROTITER PLATE

Probes for all four possible alleles (A, C, G, T) for positions c.93-22, c.93-21, c.93-20, c.117, c.118 and c.119 in the *HBB* gene were bound to streptavidin-coated microtiter plates. All probes were 25 bp long with biotin on the 3' end and ROX on the 5' end. The probes were diluted to 0.25 µM in HEN buffer and bound to the plates at room temperature for at least one hour. The wells were then washed with HEN buffer to remove any unbound probe.

Two asymmetric unlabelled PCR products were hybridised to these probes; one 145 bp product (HBB1-05F/HBB1-07R) and one 433 bp product (HBB1-05F/HBB1-02R). The PCR products were first denatured at 95°C for 4 minutes and directly cooled on ice. 18 µl of a denatured PCR product was added to each well and hybridised 1 hour at 45°C and then cooled to room temperature. The wells were then washed with SYBRGreen/HEN buffer (1:7,500 dilution) and replaced by 25 µl fresh HEN/SYBRGreen. Fluorescence was recorded in a DASH instrument and fluorescence was recorded while heating from 35°C to 85°C at a rate of 0.3°C/second.

2.3 MEGA-PLEX PCR

2.3.1 OLIGONUCLEOTIDES

All oligonucleoties were obtained from Biomers.net GmbH, Germany. Primers that were used for solid-phase reactions were all synthesised with a 5' biotin.

The target-specific parts of the primers were designed using the D-fold[149] and/or Primer3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3.cgi). The oligo-nucleotides used for MegaPlex PCR are listed under the appropriate section, or in **Appendix 2**.

2.3.2 BINDING PRIMERS TO SOLID SUPPORTS

Primers, with biotin at their 5' end, were attached to solid support by streptavidin-biotin interaction to either streptavidin-coated polypropylene membranes (0.2 µm GHP, Pall Corporation) or to streptavidin-coated magnetic beads (Dynabeads MyOne streptavidin C1, Dynal Biotech/Invitrogen). The beads are either processed as a freely dispersed mixture in solution or, for easier handling, immobilised onto the walls of wells in a microtiter plate.

For binding to membranes, pairs of primers were transferred from a 384-well microtiter plate to equivalent distinct locations on a membrane (via a centrifugation procedure). The primers were diluted as a pair in HEN buffer (0.1 M Hepes, 10 mM EDTA, 50 mM NaCl, pH 7.5) to a concentration of 0.01-0.5 µM of each primer and 5 µl of this mixture was applied to each membrane feature. After incubation for 30 minutes at room temperature the membranes were washed in 0.1 M NaOH followed by a rinse in PCR buffer.

For binding to beads, pairs of primers were bound to Dynabeads following the manufacturer's protocol. Aliquots of 5 µl of beads per target were washed twice in 2x Binding and Washing (B&W) buffer (10 mM Tris-HCl (pH7.5), 1 mM EDTA, 2.0 mM NaCl) and resuspended in double the initial volume. The beads were then mixed with an equal volume of an appropriate pair of biotinylated primers diluted in water (typically 0.01-0.5 µM), and incubated at room temperature for at least 15 minutes whilst gently

agitating. The beads were then washed twice in 1xB&W buffer and finally resuspended in PCR buffer. Mixtures of beads carrying different primer pairs were then made for multiplex experiments.

The beads could also be immobilised on the walls of a streptavidin coated plate (ChiocoCoat Streptavidin plate from Pierce) that was first coated with biotin-BSA (Sigma). Biotin-BSA was diluted to 40 µg/ml in 1xB&W buffer, 40 µl was added to each well of the plate and incubated at room temperature for 2 hours. The wells were the washed with 1xB&W buffer and 30µl of primer-carrying beads in 1xB&W buffer was transferred to each well incubated at room temperature for at least one hour. The wells were then washed in PCR buffer twice, before proceeding with the MegaPlex amplification.

2.3.3 DEMONSTRATION OF PRIMER EXTENSION ON SOLID SUPPORT

Oligonucleotides

Table 2.5 Oligonucleotides for solid-phase primer extension.

Oligo Name	Mod.	Oligo Sequence	Function
JB15-47P		GTTTCATTCCTGTTTGTCAGTTGTACGGTGGGTTGTGCCAAAATGCAGTT	synthetic target
L_dT10_X727_B15b06F	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGCTGCATTTTGGCACAACCC	s-p primer
L_dT9U_X727U_B15b06F	5'biotin	TTTTTTTTTUGAGCGAATTCTAGACTGCAGGCTGCATTTTGGCACAACCC	s-p primer
L_dT10_X727_B15b07R	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGGTTTCATTCCTGTTTGTCAGT	s-p primer
HB15+31P	5'ROX	CAGTTGTACGGTGGGTT	probe
LB15R+01P	5'ROX	AACCCACCGTACAACTG	probe
MPX:X732:+01P	5'ROX	TGCAGTCTAGAATTCGC	probe
MPX:X727:+01P	5'ROX	GAGCGAATTCTAGACTGCAGG	probe

Primer extension

Forward (dT9U_X727U_B15b06F) and reverse (dT10_X727_B15b07R) primers were bound to membranes as described above. An oligonucleotide template (JB15-47P) complementary to the forward primer was hybridised to the surface bound primers: a 100 µl solution of 2 µM template in HEN buffer (0.1 M Hepes, 10 mM EDTA, 50 mM NaCl, pH 7.5) was added to the membrane and heated to 70ºC between two glass slides and then cooled to room temperature. Excess probe was removed by rinsing the membranes in PCR buffer. Primer extension was then achieved by placing the

membrane in 100 µl PCR mix including 1xAmplitaq buffer, 1.5 mM MgCl2, 5% DMSO, 0.2 mM dNTPs and 0.03 U/µl AmpliTaq Gold (Applied Biosystems). The reaction mix was pre-heated to 94ºC for 10 minutes (activating AmpliTaq Gold), cooled to 50ºC before adding the membrane and then incubated for 30 minutes. The membrane was washed sequentially in 0.1 M NaOH and HEN buffer. For the reverse primer extension the membrane was washed in NaOH after the first strand extension (to remove the template DNA) and then neutralised in PCR buffer. The reverse strand extension was performed similar to the first extension reaction, at 50ºC for 1 hour.

Fluorescent detection of extension products

Extension products were detected by iFRET[145], using complementary ROX labelled probes (HB15+31P for the forward strand and LB15R+01P for the reverse strand) and SYBR Green I dye. The membrane soaked in a 2 µM probe solution in HEN buffer was heated between glass slides to 70ºC for 3 minutes and then cooled to room temperature. Excess probe was washed away and the membranes were soaked shortly in a SYBR Green I solution and ROX fluorescence was monitored.

In order to detect the second strand with a complementary probe, the forward strand was first removed to prevent competition with the probe-binding event. The forward primer included a deoxyuridine (dU) base that was cleaved by Uracil DNA Glycosylase (UDG) and sequential NaOH wash. The membrane was placed in 50 µl UDG buffer with 2 U UDG and incubated for 1 hour at 37ºC, and then washed in 0.5 M NaOH for 15 minutes and neutralised in HEN buffer. Then a probe complementary to the reverse strand was hybridised and detected in the same manner as for the forward strand.

2.3.4 PRE-AMPLIFICATION OF TEMPLATE DNA

The DNA was pre-amplified in a standard solution-phase multiplex PCR using primers designed to amplify at least the full stretch of sequences to be captured in the MegaPlex experiment. This reaction contained 50 ng human genomic DNA, 4 mM $MgCl_2$, 200 µM dNTPs, 2 U Amplitaq Gold, and 0.03 µM of each primer in a 20 µl reaction. Thermal cycling consisted of an initial activation step at 94°C for 10 minutes, followed by 27 cycles of 94°C for 15 seconds and 52°C for 30 seconds. From this product single stranded target was isolated using Dynabeads MyOne streptavidin C1 (Dynal

Biotech/Invitrogen). 10 µl beads were washed and resuspended in 2xB&W buffer, mixed with 10 µl PCR product and left to bind for 45 minutes. The beads were then washed three times in 40 µl PCR buffer, then the solution was heated to 90°C for 5 minutes to denature the DNA, the beads quickly separated with a magnet and the supernatant was recovered to a new tube.

2.3.5 MEGAPLEX PCR

Target capture

The primers are attached to solid support as described above. Template DNA (typically 0.1-2 µg of human genomic DNA or pre-amplified DNA) is denatured by placing in a boiling water bath for 5 minutes, followed by direct cooling on wet ice. The denatured template is hybridised to the surface-bound primers via an overnight incubation at 45 ºC in PCR buffer without Taq polymerase or nucleotides (1xAmplitaq Buffer, 1.5 mM MgCl2, 5% DMSO).

To initiate the first extension cycle, additional PCR buffer is added to bring the final concentration of dNTPs to 0.2 mM and AmpliTaq Gold (Applied Biosystems) to 0.04 U/µl. The additional reaction mix is pre-heated at 94ºC for 10 minutes (to activate the AmpliTaq Gold enzyme), and it is then cooled to annealing temperature before being added to the annealing reaction. Primer extension is allowed to proceed for 5 minutes at 45-55°C followed by 5 minutes at 65°C. The reaction is stopped by adding 30 µl of 0.5 M EDTA, and the solid support is washed in 0.1 M NaOH to remove the template DNA, followed by a wash in PCR buffer. For the second cycle extension the procedure is repeated, with the differences that no DNA is added to the reaction and the annealing is only allowed 5 minutes before adding the dNTPs and AmpliTaq Gold.

Blocking of non-extended primers

After the second extension all surface-bound DNA molecules are blocked at their 3' ends by incorporation of a dideoxy base. This was done in a reaction containing 0.05 U/µl Terminal deoxynucleotide Transferase (TdT) (GE Health Care) and 0.5 mM ddATP in 1xTdT buffer (GE Health Care), which was incubated at 37ºC for 1 hour.

PCR with common primers

Finally, a standard solution-based PCR is conducted to bulk-amplify all the surface-bound MegaPlex PCR products, using a single primer pair matched to the common tails of the solid-phase primers. Small sections of membrane or the beads from the experiment were placed directly into a PCR vessel. Alternatively, if the beads were bound to a microtiter plate well, this well is used as the PCR vessel. This PCR is typically of 30 µl volume and it contains 0.4 µM of each of the two common primers, 0.03 U/µl AmpliTaq Gold polymerase (Applied Biosystems), 1x AmpliTaq Gold Buffer, 2.5 mM $MgCl_2$, 5% DMSO, 200µM of each dATP, dCTP and dGTP, 400 µM dUTP and 0.02 U/µl of Uracil DNA glycosylase (GE Health Care). Thermal cycling consisted of an initial 10-minute activation step of 94°C for 10 minutes, followed by 30 cycles of 94°C for 15 seconds 58°C for 30 seconds and 72°C for 1 minute.

2.3.6 MEGAPLEX PCR PRODUCT ANALYSIS

2.3.6.1 TARGET DETECTION BY SECONDARY PCR

Successful MegaPlex amplification was assessed by re-amplifying a 50-plex product in separate sequence specific PCRs. 24 reactions were set up, of which 21 contained primers for sequences included in the MegaPlex and two containing primers for other genomic sequences and one reaction without primers. 0.2 µl of a 50-plex MegaPlex product (amplified with primers X727-01A and Y727-02B) was used as template in a 20 µl reaction containing 0.4µM of two sequence-specific primers, 0.03 U/µl AmpliTaq Gold® polymerase (Applied Biosystems), 1x AmpliTaq Gold® Buffer, 1.5 mM $MgCl_2$, 5% DMSO, and 200 µM of each dATP, dCTP, dGTP, dTTP. Thermal cycling consisted of an initial activation step of 94°C for 10 minutes, followed by 20 cycles of 94°C for 15 seconds and 52°C for 30 seconds. The separate PCR products were analysed by agarose gel electrophoresis.

2.3.6.2 TARGET DETECTION BY MICROARRAY HYBRIDISATION

Control samples preparation: A control sample for hybridising to the microarray was prepared by amplifying each MegaPlex target in a single-plex solution phase reaction containing 0.4 µM of each MegaPlex primer, 0.03 U/µl AmpliTaq Gold® polymerase

(Applied Biosystems), 1x AmpliTaq Gold Buffer, 1.5 mM $MgCl_2$, 5% DMSO, and 200 µM of each dATP, dCTP, dGTP, 400 µM dUTP, and cycling at 94°C for 10 minutes, followed by 10 cycles of 94°C for 15 seconds, 49°C for 30 seconds and 72°C for 1 minute and 30 cycles of 94°C for 15 seconds, 60°C for 30 seconds and 72°C for 1 minute. The individual products were then pooled to a 50-plex and a 90-plex sample.

The MegaPlex products and the control samples were re-amplified with biotin labelled primers (X727b01A and X727b01B). Each 200µl reaction contained 2µl MegaPlex product as template, 0.4µM of each primer, 0.03 U/µl AmpliTaq Gold polymerase (Applied Biosystems), 1x AmpliTaq Gold Buffer, 1.5 mM $MgCl_2$, 5% DMSO, and 200 µM of each dATP, dCTP, dGTP, 400 µM dUTP, and cycling at 94°C for 10 minutes, followed by 25 cycles of 94°C for 15 seconds, 58°C for 30 seconds and 72°C for 1 minute. 125 µl of each biotinylated product was purified with MinElute spin columns. From each column the purified product was eluted in 10 µl H2O.

Oligonucleotide arrays were generated by photo-controlled in situ synthesis as previously described[184] using the GENIOM system and DNA processors (febit biotech, Heidelberg, Germany), which allows hybridisation of eight samples in parallel using individually accessible micro-channels. Each sub-array was washed with 15 µl Pre-hybridisation buffer (100 mM Morpholinoethanesulfonic acid (MES) pH 6.6, 0.9 M NaCl, 20 mM EDTA, 0.01% (v/v) Tween-20, 1% BSA) (febit biotech). 4 µl of each purified product was mixed with 12 µl hybridisation mix 2 (100 mM MES pH 6.6, 0.9 M NaCl, 20 mM EDTA, 0.01% (v/v) Tween-20, 0.1 mg/ml herring sperm DNA, 0.5 mg/ml BSA, control oligo mix 1) (febit biotech), and hybridised to the array for 4 hours at 45°C. Fluorescence-staining was performed with 0.1 mg/ml streptavidin-R-phycoerythrin-conjugate (Molecular Probes, Leiden, The Netherlands) in 6x SSPE at 25°C for 15 minutes within the GENIOM instrument. Subsequently, the arrays were rinsed with 6x SSPE at 25°C. Signal detection was done using the internal CCD-camera system of the GENIOM instrument. For fluorescence detection the Cy3 filter was employed.

2.3.6.3 TARGET DETECTION BY GS20 SEQUENCING

To prepare the MegaPlex PCR products for GS20 sequencing they were re-amplified with fusion primers (GS20A_X13-01A and GS20B_Y13-01B) that contained the necessary sequences for GS20 priming at their 5' ends. For this, 1.5 µl MegaPlex product was used as template in an 80 µl reaction containing 0.03 U/µl AmplitaqGold, 1x Amplitaq Buffer, 1.5 mM $MgCl_2$, 5% DMSO, 0.2 mM of each dATP, dCTP, dGTP, 0.4 mM of dUTP, 0.4 µM of each primer. Thermal amplification consisted of an initial activation step of 94°C for 10 minutes, followed by 5 cycles of 94°C for 15 seconds, 55°C for 30 seconds, 72°C for 1 minute, and then a further 20 cycles of 94°C for 15 seconds, 60°C for 30 seconds, 72°C for 1 minute.

The GS20 Sequencing procedure was carried out at Roche Applied Science, Penzberg, Germany, according to the standard GS20 Operators Manual. MegaPlex PCR samples (MP50 and MP75), now with appropriate 'A' and 'B' sequences at their termini were quantified with Agilent DNA Chip 1000 in order to determine exact concentrations. For each sample, two independent emulsion PCR (emPCR) reactions were set up to ensure sequencing from both ends. After emPCR and bead enrichment the beads of each MegaPlex sample were combined and loaded on a 70x75 Picotiterplate with a two line gasket. A single pyrosequencing run on a GS20 instrument (Roche Applied Science) was performed resulting in 84,483 high quality reads of MP50 and 75,403 reads for MP75.

GS20 reads were aligned by BLAST against the expected sequences of all the MegaPlex targets. High quality reads were taken to be those that matched a target with over 90% identity, and which also perfectly matched at least 2 of the following 4 regions; 15 bp of common primer DNA immediately 5' or 3' to the target-specific sequence; first 15 bp of the target-specific sequence; first 15 bp of the target-specific sequence immediately after the SNP position. From these a few alignments of >90% identity were discarded that only matched the 15 bp stretches of common primer DNA immediately 5' or 3' to the target-specific sequence. Unique and correct alignments were thus established, from which we proceeded to count how many times each target had been recovered in the MP50 and MP75 experiments.

For the SNP genotyping analysis, the number of occurrences of T, C, G, or A alleles were counted at the SNP location in each target. The MP50 and MP75 datasets were then pooled and the overall allelic distribution (% of total counts) was calculated. Targets with less than a total of ten counts were omitted from this genotyping analysis.

Primer-dimers were identified from sequences with lower target similarity that matched more than one primer sequence.

2.3.7 SNP GENOTYPING BY DASH

Genotypes for the 50 SNPs common to the MP50 and MP75 experiments were determined by single-plex genotyping by Dynamic Allele-specific Hybridisation (DASH). Each fragment was amplified in a 20 µl reaction containing 1x Amplitaq buffer, 1.5 mM $MgCl_2$, 5% DMSO, 200 µM of each dNTP, 0.15 µM of a target-specific biotinylated primer, 0.75 µM of a target-specific non-labelled primer, 0.03 U/µl Amplitaq Gold and ~10 ng genomic DNA, and cycled 10 minutes at 94°C followed by 40 cycles of 94°C 15 seconds, 52°C 30 seconds.

The DASH analysis was conducted as described for DASH-1 in section **2.1**.

CHAPTER 3.

DNA DIAGNOSTICS WITH DASH

3.1 INTRODUCTION

DNA diagnostics is an important tool in a range of different applications. It can be used to identify causative mutations for Mendelian diseases, for example Cystic fibrosis and haemoglobin disorders. Diagnostics at the DNA level is particularly important for accurate prenatal diagnostics. It can also be used to test for genetic variants that dramatically increase the risk of certain diseases, such as breast cancer. As more risk factors for complex disease are identified, the need for DNA diagnostics should increase. Additionally, DNA diagnostics can be used for pathogen detection and identification, as an alternative or complement to traditional methods. All these applications need efficient and reliable methods.

Although many different methods are available for DNA sequence variation analysis, there is always a need for faster, cheaper and more flexible methods. It could, for example, be attractive if a single method could be used both for detecting novel mutations and for typing known sequence variants. In this chapter the DASH method is adapted for both mutation scanning and the simultaneous detection of multiple mutations in parallel, to demonstrate the flexibility for several applications in DNA diagnostics[185].

DNA hybridisation is arguably the simplest way to analyse DNA sequence variation. Single base pair differences can be detected by hybridising an oligonucleotide of known sequence to a test sample. The degree of match between the two DNA strands is reflected in the hybridisation efficiency. This can be measured either quantitatively in static hybridisation, where the amount of hybridised probe is measured at stringent conditions, or qualitatively in melt curve analysis, by determining the melting temperature (Tm).

For mutation scanning by oligonucleotide hybridisation, a set of oligonucleotides would at least have to cover the whole target sequence. The desired resolution will determine what probe density has to be used. When it is enough just to know whether a sample is identical or not to a reference sequence, the probes only have to just overlap. For complete resequencing, all bases have to be individually assayed by hybridisation to oligonucleotide probes for all four possible bases (A, C, G, and T). This concept has been developed for static hybridisation assays, where long stretches of sequence can be analysed on high density microarrays by comparing the hybridisation efficiency to the four different probes.

The strategy of using sets of tiled oligonucleotide probes for mutation scanning is less developed for dynamic hybridisation (melt-curve) assays. DASH, for example, was developed for SNP genotyping, where one particular SNP is genotyped in many samples in parallel. In mutation scanning applications the experimental design will be quite different, in that many probes will be used to analyse each sample. The strategy for mutation scanning by dynamic hybridisation would employ a set of tiled probes, which together cover the sequence of interest. Sequence variants would thereby be detected by the probe(s) binding to the variant base, since a base change would alter the melting profile of that particular probe compared to a reference, "wildtype", sequence. A similar concept has been presented before[186], but this involved a homogeneous assay and could therefore only analyse 2-3 probes per reaction. An array-based system would be more practical to scale up for the analysis of more samples and/or the parallel use of an increased number of probes. In this chapter, it was evaluated how DASH can be developed for mutation scanning applications, by testing this concept on a human genomic target sequence.

A related application is the analysis of multiple characterised mutations in one or several genes. For example, if a phenotype can be caused by either of several known mutations, it would be practical to test for all of those mutations in parallel instead of having to sequence the whole gene. In this chapter, it was also evaluated how DASH could be used for this additional application, by detecting multiple mutations in the *gyrA* gene of *Salmonella* that cause fluoroquinolone resistance. Fluoroquinolones are broad-spectrum antibiotics that target the topoisomerase genes, DNA gyrase and topoisomerase IV, resulting in inhibition of DNA replication[187]. Mutations leading to

amino acid substitutions within these genes can therefore lead to quinolone resistance, and point mutations in the gyrA subunit of DNA gyrase is a common cause for quinolone resistance in *Salmonella*[187].

3.1.1 METHODS FOR *GYRA* MUTATION DETECTION

A range of methods have previously been used to develop specific assays for the detection of the most common mutations in the *gyrA* gene. PCR-RFLP has been used to detect mutations in codon 83 and codon 87 in a single assay by identifying mutation-specific digestion patterns[188,189]. Allele-specific PCR has been used with primers specific for the wildtype sequences of codons 83 and 87 respectively, which are both assayed in duplex reactions including primers for a control product[190]. Mutations in these codons suppress the amplification of the wildtype-specific product, and mutations are detected when this product is absent while the control product is amplified. Further, homogeneous real-time PCR methods with mutation-specific probes have been developed both for melt-curve analysis[191-193] and allele-specific cleavage using the Cycling probe technology[194,195] and several mutations have also been typed by microarray hybridisation[196].

For the detection of novel mutations, SSCP[197], DHPLC[198] and high-resolution melting analysis[199] have been used. If the exact identities of the mutations need to be determined, they usually have to be confirmed by sequencing. Recently, Pyrosequencing®[200] has also been used to detect multiple mutations in the *gyrA* gene simultaneously, and this method also has the ability to detect novel mutations.

3.1.2 AIMS

The aims of this chapter are to implement these two applications, mutation scanning and multiple mutation analysis, by further development of DASH. Initially, this will require evaluation and optimisation of different reaction formats. The current options are DASH-1, a microtiter plate format, and DASH-2, a macroarray format. Further focus will be on assay design, both regarding target selection in terms of PCR product length and probe design in terms of probe position and probe spacing for mutation scanning. Data interpretation should also be considered.

3.2 ASSAY FORMATS

DASH was originally developed in a microtiter plate format (DASH-1)[131]. This is a convenient format for low-throughput applications and can easily be automated or semi-automated, for example with the use robotics or multi-channel pipettes. Any combination of probe-target can be analysed in each well, thus providing an excellent platform for proof-of-principle studies for diagnostic applications but probably not ideal for higher throughput applications. To increase genotyping capacity compared to the microtiter plate format, DASH-2 was developed on membrane arrays[132]. For SNP genotyping, the probes are applied to the whole membrane, but this strategy is not ideal when multiple probes need to be used on the same target sequence. In that case the different probes will have to be applied to different locations on the membrane, preventing them from interfering and competing with each other.

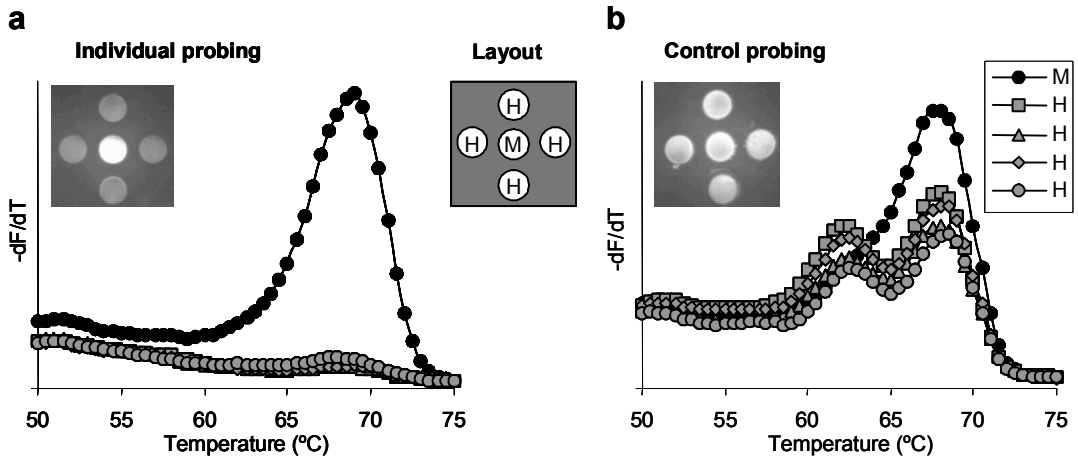


Figure 3.1 Applying probes to individual features.

The sample layout is shown between the two graphs; PCR products are attached to five features on each membrane where 'M' indicates a sample matching the probe and 'H' indicates a heterozygous sample.

a) A DASH probe is only applied to the central spot, by centrifugation transfer from one well in a 384-well plate. This is apparent both from the picture, where the fluorescence is notably stronger in the central spot than in the surrounding spots, and in the graph, where a DASH signal is only seen for the match sample.

b) The same DASH probe is applied to the whole control membrane. The fluorescence intensity is fairly similar for all spots, well above the background level. All features produce DASH curves, homozygous match for the central feature and heterozygous signals for the surrounding features.

The simplest way to transfer the target DNA (PCR product) to the membrane is by centrifugation transfer[148]. A membrane is then placed on top of a PCR plate (usually a 384-well plate) and clamped together in a sandwich-like manner and tightened in a purpose-built clamp. The PCR product is then transferred to the membrane by inverting the clamp and applying centrifugal forces. To avoid cross contamination, the liquid is transferred back into the wells, again by using the centrifuge, after binding to the membrane and before the clamp is removed. This centrifugation transfer method was successfully adapted for probe transfer. Different probes could be transferred from individual wells of the microtiter plate onto separate features on the array when the membrane was placed so that each well was aligned with an area containing DNA template. (The procedure is described in detail in Materials and Methods, **Chapter 2.1.5**.) The clamps are tight enough to prevent leakage of probe into adjacent features (**Figure 3.1**). In this way DASH-2 offers the potential as a platform for medium-throughput diagnostic applications.

3.3 MUTATION SCANNING WITH OVERLAPPING HYBRIDISATION PROBES

A test region on human chromosome 7 was chosen to evaluate the ability of DASH to detect novel mutations. This target fragment was chosen because it contained three SNPs in close proximity. Those SNPs were all present in dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/index.html; rs917188 [C/T], rs917189 [C/G] and rs917190 [C/A]) and validated in different populations. The minor allele frequency in European populations was reported to be between 15% and 35% for those SNPs. At this frequency it should be possible to detect sequence variation at these positions within our small sample set of 16 Swedish DNAs. The mutation scanning was done blind, with regard to the SNP genotypes of the different samples, to test the effectiveness of the method.

A series of fifteen partially overlapping probes were designed to cover a 97 base pair sequence, including the above mentioned SNPs. All probes were seventeen base pairs long, which is a known optimum length for DASH, and were evenly distributed along this sequence so that the position of each probe shifted with five bases from the adjacent probes (**Figure 3.2 a**). In this way each targeted base was covered by up to four separate

probes. The probes were designed for the forward strand of the genome reference sequence and only for one of each SNP allele. The highest frequency allele was chosen for each SNP, which meant that the probes contained a 'C' at each polymorphic position.

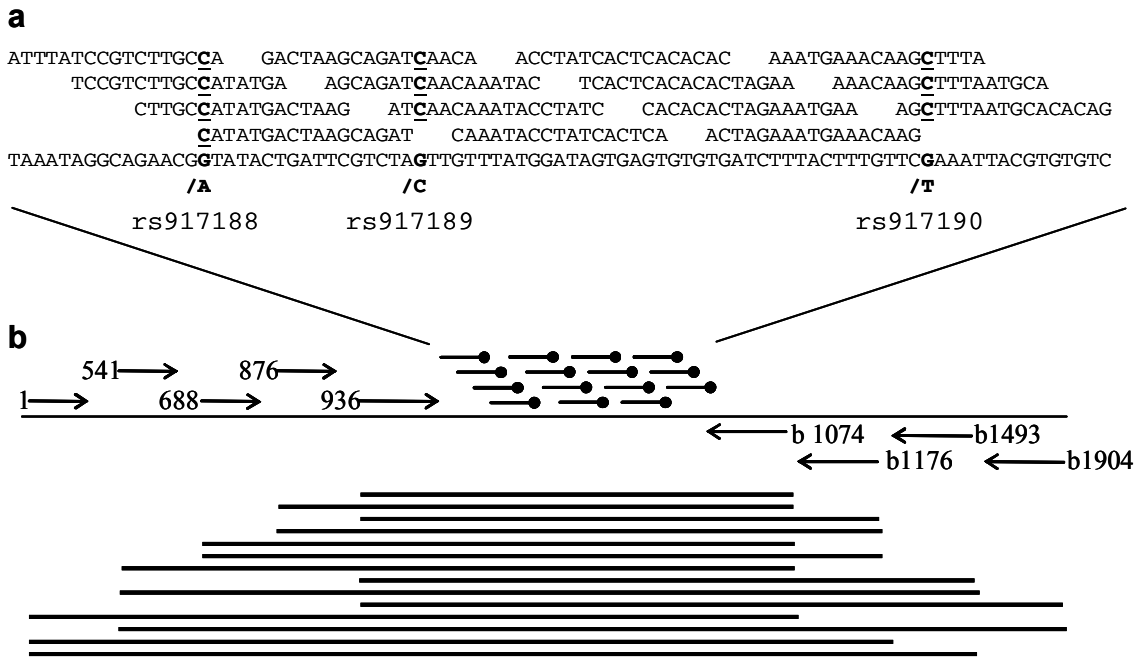


Figure 3.2 Overview of the human genomic sequence scanned for mutations.

a) The DASH probes are shown aligned to the 97 base pair region they are designed to interrogate. SNP positions are marked as bold underlined characters in the probes and the two alleles are shown in the complementary target sequence. **b)** An overview of the 2 kb target region centred upon the 97 base pair target sequence. The DASH probes are drawn with black circles representing the fluorescent 3'-ROX groups. PCR primers are drawn as black arrows (→) indicating 5'-3'orientation, and the number behind each primer indicates the relative position of its 5' base in the target. Primers labelled with 'b' carried 5'-biotin residues to enable surface immobilisation in the DASH procedure. A total of 14 different PCR products, with lengths from 139 to 1904 bp, were amplified using different combinations of the forward- and reverse primers, indicated by black lines.

The first attempt to do mutation scanning with the above described model was done on a microtiter plate (DASH-1) platform. Aliquots of PCR product from the same DNA sample were bound to multiple wells in the microtiter plate and assayed separately with the different probes in different wells in parallel. If the system worked effectively, it should be possible to detect sequence variants from alterations in the melting profile of

the probe(s) hybridising across the variant base; if there is a single base mismatch along the probe sequence, the probe melting will occur at a lower temperature than if it is perfectly matched. In this way heterozygous sequence variants will produce double melting peaks (which will be the most likely result if scanning a diploid genome for rare variants) and samples homozygous for the low frequency allele will produce a single peak at a lower temperature.

Two major factors were evaluated in these mutation scanning experiments with DASH; i) the effect of PCR product length on data quality, and ii) the effect of the position of the variation base within the probe for reliable mutation detection.

3.3.1 PCR PRODUCT LENGTH

In order to evaluate the effect of PCR product length on data quality, several primers were designed around the 97 base pair test sequence. Five forward primers and four reverse primers were used in a total of fourteen combinations to amplify products between 139 and 1904 base pairs (**Figure 3.2 b**). Initially 2-3 of the DNA samples were used for amplification of the different length products, to compare their performance with regard to signal quality of all 15 hybridisation probes. At first a subset of nine different PCR products were used, where the differences in product length were quite large. The shortest product, 138 bp, produced strong hybridisation signals and distinct and strong melting peaks, but the fluorescence signals dropped quickly to the next product at 533 bp and then even more for products above this length (**Figure 3.3 a**). Products over ~500-600 bp would not consistently produce good enough data for reliable sequence variation detection, since the signals were very low. Since there was quite a big gap in signal strength between the 138 bp and the 533 bp products, another five products between 198 and 488 base pairs were then amplified to determine the product length effect at this lower size range. The data show that there is still a substantial difference in signal strength between the shortest and the second shortest product (**Figure 3.3 b**), indicating that product length is very important for signal quality. However, in this range up to 500 bp, all products produce reliable melting profiles that should be good enough for correct sequence variation detection. Therefore, one of the intermediate sized PCR products (387 bp, PCR primers LSCAN-24F and LSCANb04R) was used for analysis of the full set of 16 DNA samples.

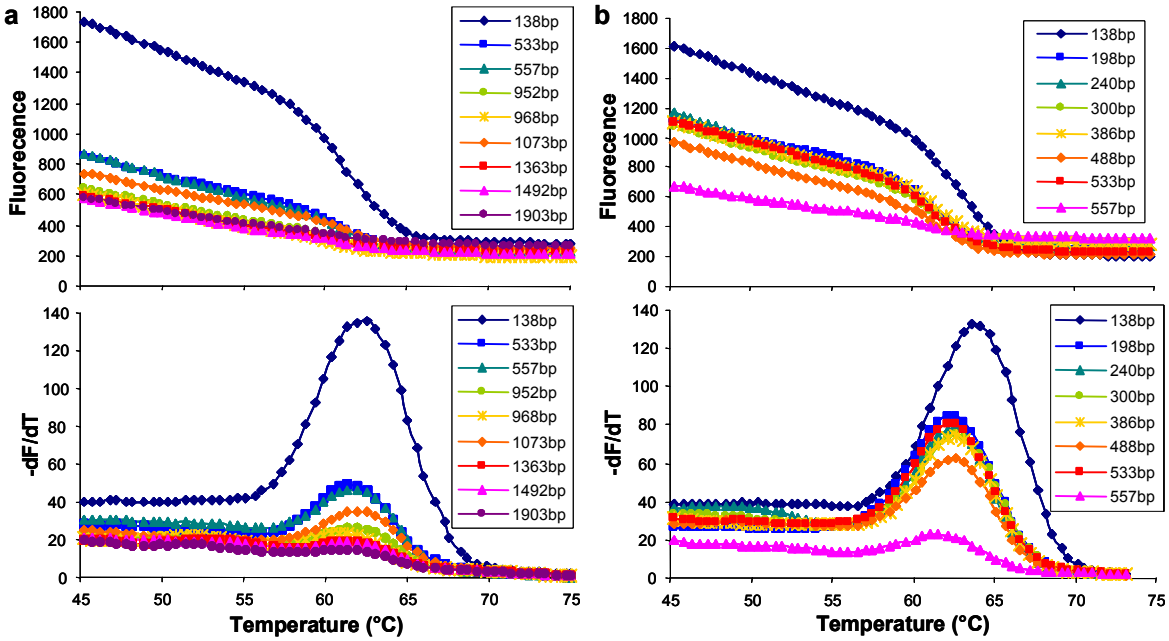


Figure 3.3 The effect of PCR product length on DASH result.

DASH melt-curves are shown for raw data (above) and derivative data (below). **a)** PCR products between 138 and 1903 base pairs and **b)** PCR products between 138 and 557 base pairs, all amplified from the same DNA sample and all spanning the core 97 base pair target sequence. DASH data is shown for one of the probes (LSCAN+18P) to illustrate the effect of increasing product length on data quality.

3.3.2 PROBE POSITION FOR MUTATION DETECTION

The full test set of 16 DNA samples and 15 hybridisation probes was used to determine the consistency of sequence variation detection at the three SNP sites. Ideally, all probes hybridising to the same sample and the same SNP should give identical result. However, the sensitivity of mutation detection might be influenced by the variant position in the probe; if the variant position hybridised to the central part of the probe or towards either end of the probe. The full probe set could be used to determine the importance of the probe position for reliable mutation detection, since three to four probes hybridised to each SNP.

In general, across the 16 DNA samples, all homozygous and heterozygous sequence variants were identified by all probes binding to each variant position. Example data for one of the SNPs, rs917189, is shown in **Figure 3.4**, where the three probes detect the sequence variation equally well. Plotting the data for each of those probes, the 16 samples clearly group into three categories with melting peaks at either high or low temperature, relative to each other, or both. The exact melting temperatures vary for each probe depending on their sequence.

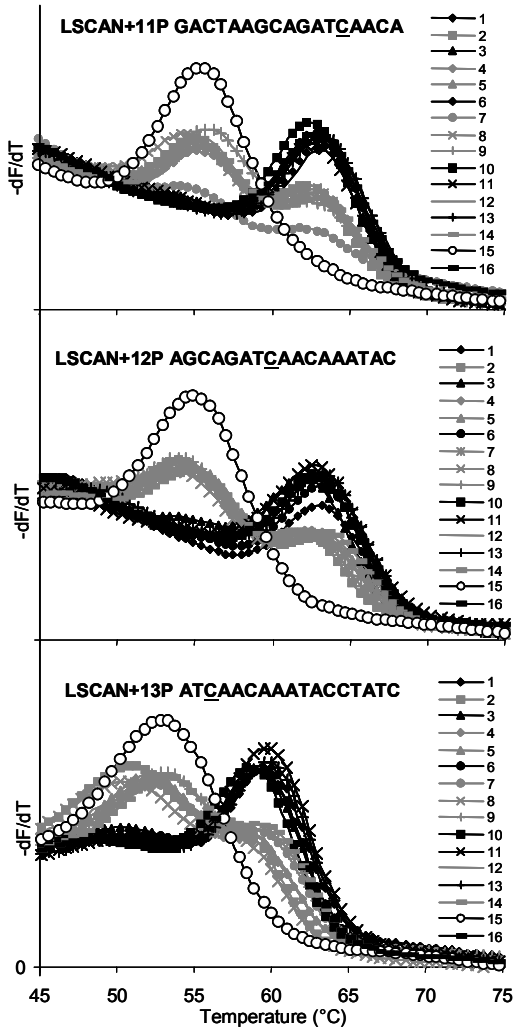


Figure 3.4 Example data from mutation scanning.

One C/G SNP, rs917189, is detected by three overlapping probes that all match the C allele. The three probes detect sequence alterations consistently, as shown by comparing melting temperatures for 16 different human DNA samples; The samples labelled in black are perfectly matched to the probe, giving the highest temperature melting peak, the sample labelled in white is homozygous for the other allele (has one base pair mismatched to the probe) and therefore produces a peak at a lower melting temperature, and the samples labelled in grey are heterozygous and generate peaks for both alleles.

The only probe with poor detection of a sequence variant was one in which the polymorphic site was located at the last (5') base of the probe. This probe position gave rise to only a very small difference in melting temperature between the different genotypes, and would require higher resolution melting data for reliable detection of the sequence variants. Only one probe (LSCAN+10P) for one of the SNP (rs917188) was aligned in this extreme manner and the three other probes that overlapped rs917188 clearly revealed the different genotypes (**Figure 3.5**). Outside the SNP positions no probe showed any variance in melting profile, indicating that the three SNPs were the only variant positions within this assayed region.

To confirm that the genotypes obtained by DASH in this blind study were correct, the target region was also sequenced bidirectionally in all 16 DNA samples. The results are presented in **Table 3.1** and examples of sequencing data is shown in **Figure 3.6**. The DASH results were found to be in agreement with the sequencing data for all three SNPs in all 16 samples.

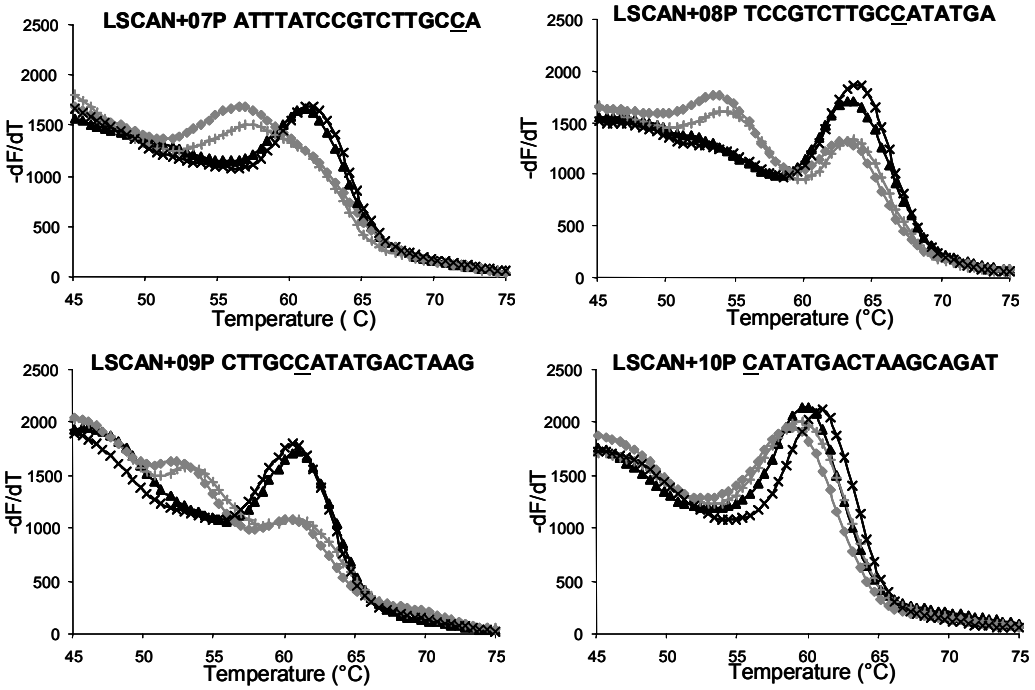


Figure 3.5 DASH data for rs917188.

Four probes are hybridising to a C/T SNP (rs917188). By analysing the melt-curves from three of the probes (LSCAN+07P, LSCAN+08P and LSCAN+09P), homozygous samples matching the probe (labelled in black) can clearly be distinguished form heterozygous samples (labelled in grey). The fourth probe (LSCAN+10P) does not clearly reveal this difference.

Table 3.1 Comparison of DASH analysis and sequencing data

SNP	rs917188		rs917189		rs917190	
Sample	**DASH**	**Seq.**	**DASH**	**Seq.**	**DASH**	**Seq.**
1	Match	C/C	Match	C/C	Match	C/C
2	Match	C/C	Het	C/G	Het	C/A
3	Match	C/C	Match	C/C	Match	C/C
4	Het	C/T	Het	C/G	Het	C/A
5	Match	C/C	Het	C/G	Het	C/A
6	Match	C/C	Match	C/C	Match	C/C
7	Het	C/T	Het	C/G	Het	C/A
8	Het	C/T	Het	C/G	Het	C/A
9	Het	C/T	Het	C/G	Het	C/A
10	Het	C/T	Match	C/C	Match	C/C
11	Het	C/T	Match	C/C	Match	C/C
12	Match	C/C	Het	C/G	Het	C/A
13	Match	C/C	Match	C/C	Match	C/C
14	Match	C/C	Het	C/G	Het	C/A
15	Match	C/C	Mism.	G/G	Mism.	A/A
16	Het	C/T	Match	C/C	Match	C/C

Seq. = Sequencing data for each SNP. The sequence is shown for the forward strand. The DASH result is combined for all probes overlapping each SNP site. All probes had a 'C' at the polymorphic positions, and the 'Match' result indicates a homozygous sample completely complementary to the probe, 'Mism' indicates a homozygous sample with one base pair mismatch compared to the probe and 'Het' indicates a heterozygous sample.

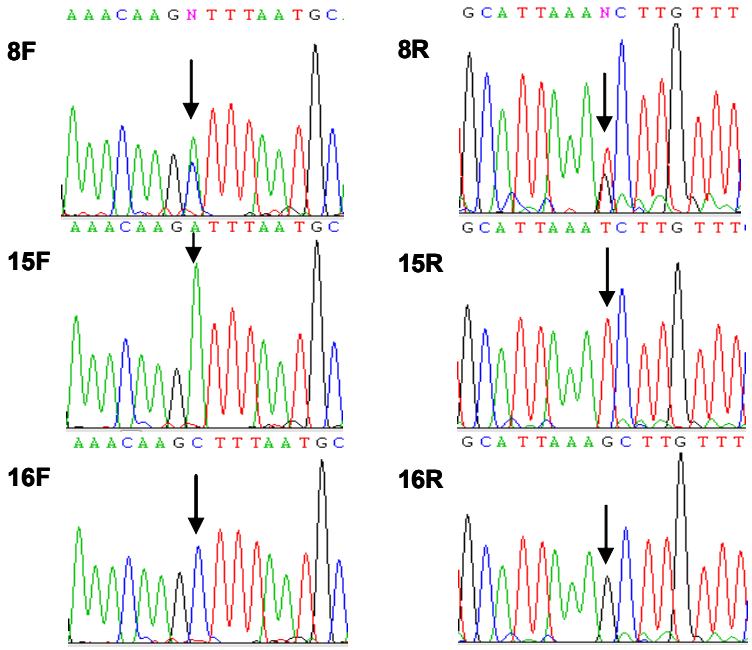


Figure 3.6 Example sequencing data.

Example sequencing data, genotyping rs917190 in three individuals with different genotypes. The samples are sequenced in both the forward (F) and reverse (R) directions. Sample 8 is heterozygous at the SNP position (indicated with an arrow), and the heterozygous base is automatically assigned 'N' both in the forward direction (C/A) and in the reverse direction (G/T). The heterozygous bases have lower peak heights compared to the homozygous samples; sample 15 is homozygous for A (T in the reverse strand) while sample16 is homozygous for C (G in the reverse strand).

3.3.3 DASH-2 FORMAT

Part of the above mutation scanning experiment was replicated on membrane arrays (DASH-2 platform), to demonstrate that the concept could also be used for higher throughput applications. For DASH-2 analysis, PCR products from three samples representing different genotypes of rs917190 were immobilised to membrane arrays by transferring the products from a 384-well plate by centrifugation transfer. The three probes for this SNP (LSCAN+19P, +20P and +21P) were hybridised to separate replicas of each PCR product as described in **Chapter 3.2**. The results of the mutation detection in the DASH-2 format were found to be equivalent to the DASH-1 result (**Figure 3.7**).

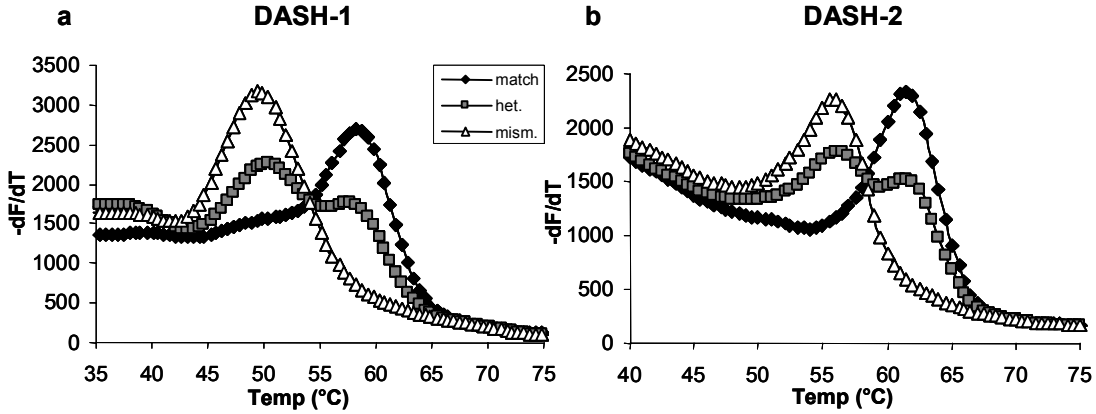


Figure 3.7 Comparison between DASH-1 and DASH-2.

Comparison of DASH-1 and DASH-2 results of probe LSCAN+20P, showing three samples representing the different genotypes.

3.4 SIMULTANEOUS SCORING OF MULTIPLE KNOWN MUTATIONS IN THE *GYRA* GENE OF *SALMONELLA*

To validate DASH for the detection of multiple mutations in the same gene in parallel, mutations in the *gyrA* gene of *Salmonella* were analysed. DASH probes were designed for 18 previously reported single base pair substitutions that cause antibiotic resistance[187] within a 300 bp sequence. These mutations were contained within only ten codons of the gene, with up to four different mutations in any one codon. For the design of DASH probes, mutations at close proximity were covered by a single wildtype probe and up to seven mutation-specific probes, that all spanned the same nucleotide positions of the gene (**Figure 3.8**). The probes were positioned so that as many of the mutation sites as possible were located towards the centre of the probes. This strategy would reduce the number of wildtype control probes, and at the same time also allow direct comparison of the melting profiles from all different probes in the same region. The targeted mutations could therefore be analysed by a total of eight wildtype probes and 18 mutation-specific probes. To evaluate the utility of this design, it was used to determine the mutation status of 62 different *Salmonella* strains.

3.4.1 INITIAL VALIDATION

DNA samples from the different *Salmonella* strains were supplied from collaborators at the Health Protection Agency (HPA) on two different occasions. For a first test of the method, DNA from 14 strains was analysed. The mutation status of each strain was known to HPA from sequencing of the whole gene, but the DASH analysis was undertaken blind to this information.

Since product length was shown to be important for the signal quality of DASH results (as described in section **3.3.1**), the assay was tested, on one sample, with both a single, fullength, product (290 bp) and with two separate products (116 bp and 129 bp respectively). As the shorter products produced much higher signals, the gene was amplified in two separate products in all following experiments, to obtain reliable results. These products were analysed on DASH-2 membrane arrays, under standard run conditions consisting of fluorescence detection during a constant heating rate.

Figure 3.8. Alignment of primers and probes for the GyrA gene.

Primers and probes are aligned to the wildtype sequence of the *Salmonella gyrA* gene (GenBank Accession Number X78977). Primer directions are indicated by arrows (-->) at the 3' end and 5' biotin modifications are indicated by '-b'. There are probes both for the wildtype sequence and for 18 different mutations within the gene. The mutation-specific bases in the probes are indicated by capital bold letters. The resulting amino acid substitution is listed after each probe and wildtype probes are labelled 'wt'.

The (24x16) array was designed with one sample per row, and the probes were applied column by column. To avoid any cross-hybridisations of multiple probes to the same feature, the probes were applied to individual features by the centrifugation transfer method described above. To further reduce the risks of cross-hybridisation the two PCR products were bound to alternate columns on the array or, alternatively, with an empty column between target features (for plate/array layout, see **Appendix 1**). In this way, even if there would be a small leakage during probe hybridisation, the probe would not hybridise to the adjacent feature since that target would contain the other part of the gene.

The mutation status for each strain was determined from the melting patterns produced by the various wildtype- and mutation-specific probes. Since multiple samples were processed in parallel, the results from each probe could be analysed separately by comparing the melting profile of the different samples. Samples containing a mutation could first be identified by the wildtype probes, since those samples would have a lower Tm compared to the majority of samples (assuming that most samples are wildtype at the majority of sites). The exact identity of each mutation was then determined by the mutation-specific probes, by determining which samples matched each probe. If a certain sample matched a mutation-specific probe it should result in a higher Tm compared to the wildtype samples (**Figure 3.9 a**). The difference in Tm (ΔTm) of the probes identifying a mutation ranged from 2.5°C to 8°C compared to the wildtype sequence – consistent with the usual resolution of DASH. When multiple mutations in the same region were assessed, each probe could also act as another (negative) control probe for the samples with a different mutation.

An alternative analysis strategy would be to assess each sample and each region of the gene separately, by comparing the melting profiles of the wildtype and all different mutation-specific probes. The probe with the highest melting temperature would be the one that is perfectly matched to the target sequence and thereby reveal the sequence identity of the sample (**Figure 3.9 b**). All other probes should have at least one base mismatch and therefore produce a lower Tm. Both strategies can determine the mutation status without the use of any reference samples. Five different mutations were identified within this first sample set. One sample was identified as wildtype and the other samples each carried one or two of the mutations.

The genotype data obtained by DASH was then compared to the sequencing data. The mutation status as determined by DASH was in agreement with the reference data in all tested strains.

3.4.2 ADDITIONAL MUTATION DETECTION

When the initial DASH results had been confirmed to be correct by comparison to the sequencing data, the remaining strains were analysed in basically the same way. A few samples were supplied in duplicates, with the DNA prepared both by the standard DNA extraction method and by a simpler method from boiled cells. Seven of those samples were analysed in duplicate with DNA prepared with both methods. Additionally, a sample identified as wildtype in the initial experiment was included as an internal control and for each mutation-specific probe a positive control sample was included where such a mutation had been identified in the first sample set.

The mutation status of all 62 strains was accurately determined, with no false positive or false negative results, as confirmed by the sequencing data obtained from HPA. Three of the samples were identified as wildtype strains and 59 samples were identified as mutant strains, of which 31 had multiple (2-3) mutations within the gene. The results for all strains are presented in **Table 3.2**. The seven duplicated samples gave consistent results, with equally good quality independent of DNA preparation method. In total 101 instances of seven different mutations were identified within amino acids 83, 87 and 144 of the gyrA gene.

Figure 3.9 Example DASH results from mutation scoring in the GyrA gene.

Example data for the mutations at amino acid 83 in the *gyrA* gene is shown for 16 *Salmonella* strains.

a) The presence of a mutation can first be detected by considering the melting curves of the wildtype (wt) probe, in this case for Serine at amino acid 83. Samples shown as solid black lines show high temperature peaks and are therefore perfectly complementary to the wt probe (are non-mutated). The other samples show a lower temperature peak and therefore carry some mutation compared to the wt sequence. The nature of each mutation can then be determined by observing which mutation-specific probe gives a high temperature (matched) peak. In this way we can conclude that the samples labelled with black circles have the Phe83 mutation, the samples labelled with grey circles have the Tyr83 mutation and the samples labelled with white circles have the Ala83 mutation.

b) The mutation status for each strain can also be determined for a single sample at a time. The melting profiles for all probes in the same region are compared, and the probe that gives the highest Tm identifies the genotype. This is demonstrated for four samples, for one wildtype strain, three strains with Phe83, Tyr83 and Ala83 mutations respectively.

Figure 3.9 Example DASH results from mutation scoring in the GyrA gene

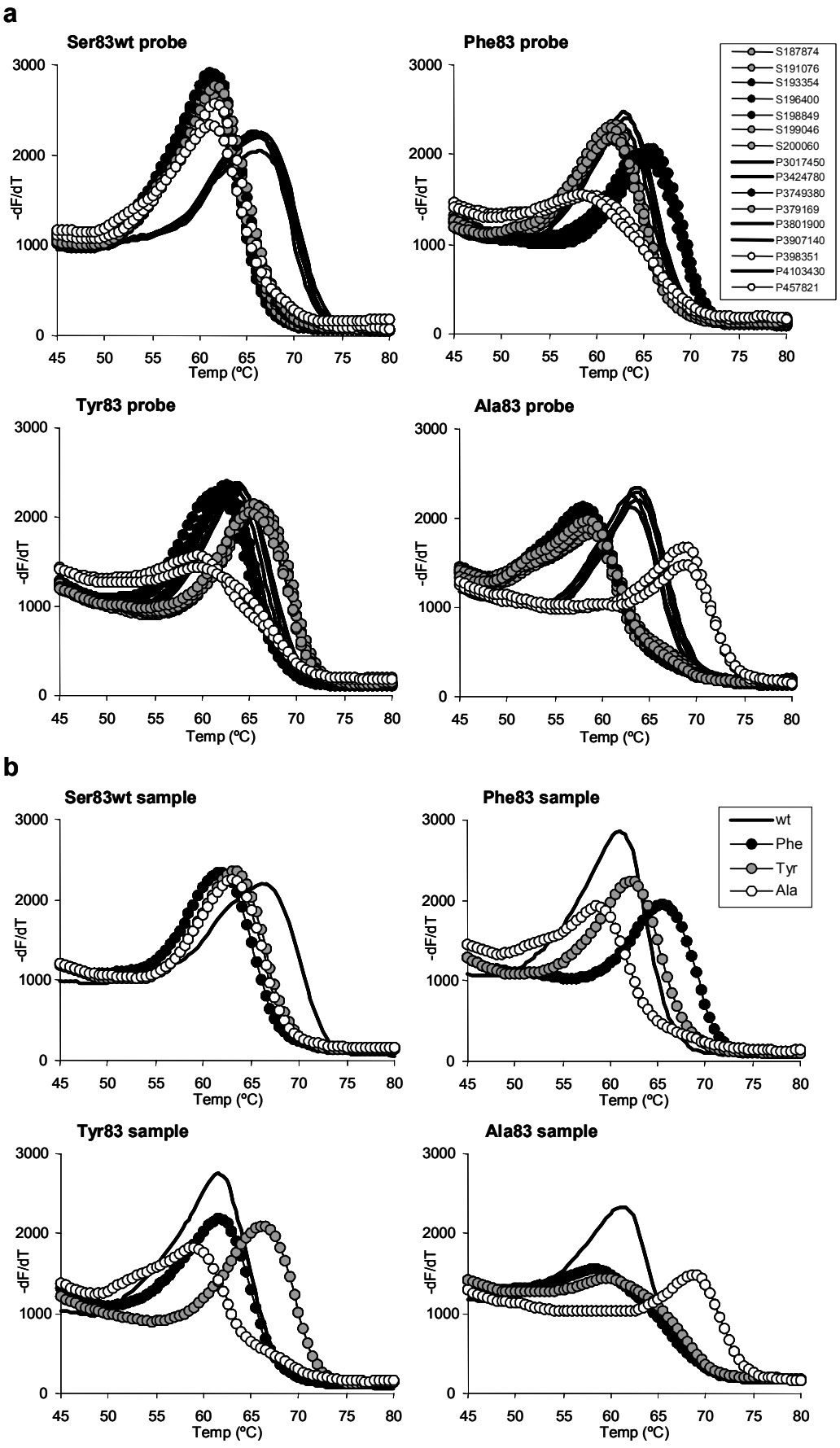
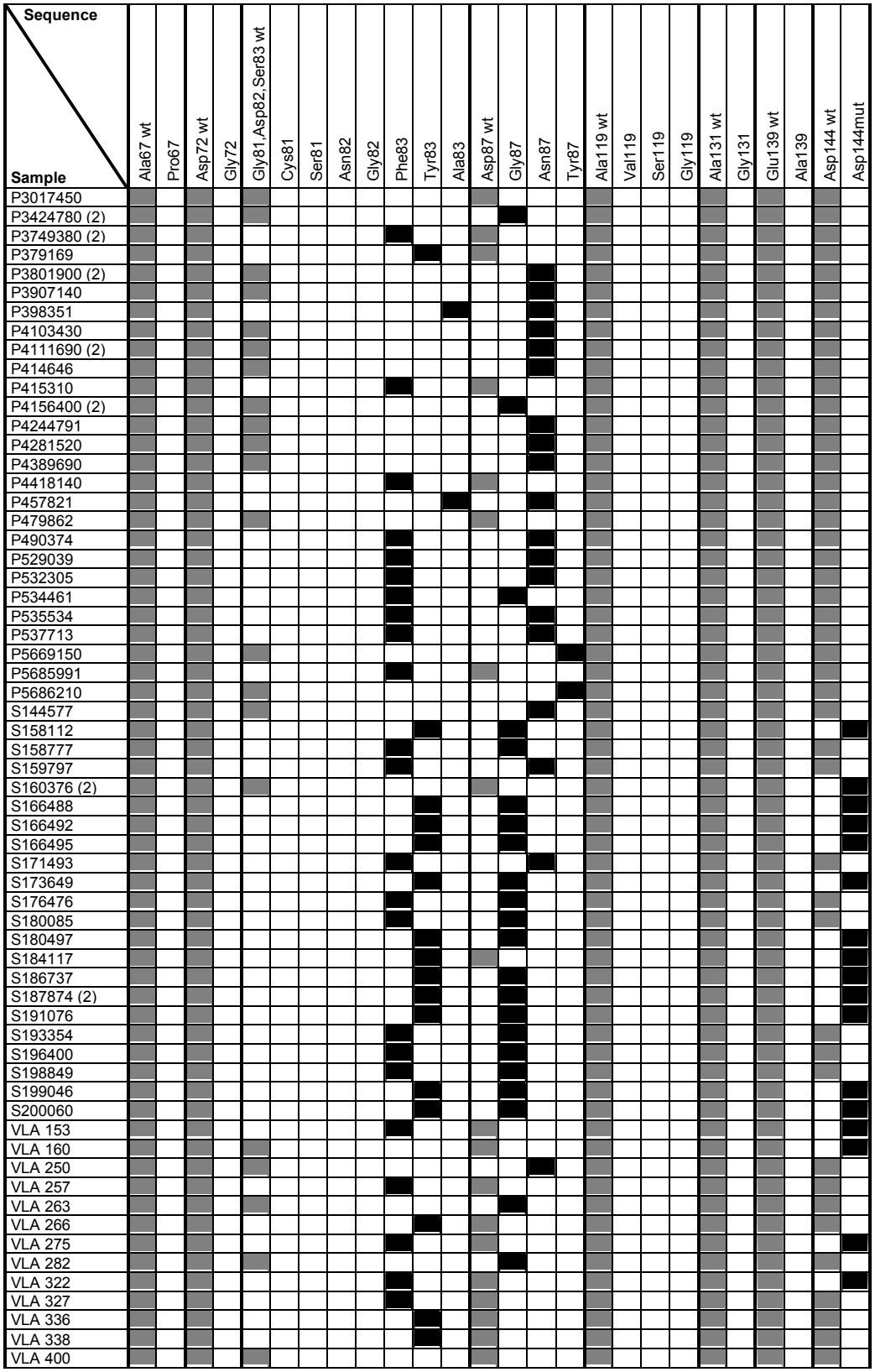


Table 3.2 GyrA mutation status as determined by DASH.



(2) = sample analysed in duplicate, giving identical results.

Matched probes are indicated with ■ (wildtype probes) and ■ (mutation-specific probes).

3.5 DISCUSSION

Dynamic allele-specific hybridization has previously proved to be reliable for SNP genotyping. In this chapter the potential of using DASH for detecting novel mutations was demonstrated, and it was shown that DASH can also be used for simultaneously genotyping of several positions in the same gene, all under standard run conditions. To test the mutation scanning possibilities of DASH, a 97 bp human target sequence was analysed by recording the melting profiles of a tiled set of probes. Three SNPs in this target sequence were detected by the probes binding to each polymorphic position, and it was possible to distinguish between heterozygous and homozygous sequence variants. Multiple antibiotic resistance mutations were tested for in the *gyrA* gene of *Salmonella*. The mutation status of 62 *Salmonella* strains was correctly determined by hybridising wildtype- and mutation-specific probes that tested for 18 previously characterised base substitutions.

The successful results obtained in these experiments would have been difficult to obtain in static hybridisation assays, where all probes have to distinguish between a perfectly matching sequence and a single base mismatch for all sequences at the same hybridisation temperature. The advantage of using a dynamic hybridisation method is that all sequences can be analysed simultaneously under the same standard run conditions independent of their melting temperature. This means that the probe design needs less optimisation since the exact melting temperature doesn't have to be the same for all sequences. The melt-curve analysis is also more informative, since different mutations might have different melting patterns and could be distinguished even when using only wildtype probes. On the other hand, one advantage of static hybridisation is faster data collection, since the fluorescence intensity is recorded only once compared to the many data points required for a melting profile. Data analysis is also simpler in static hybridisation assays since only the signal intensities for different allele probes have to be compared rather than interpreting a melting profile.

The length of the target DNA (PCR product) was found to be an important factor for data quality. The main concern with a long PCR product is the risk of secondary structure formation. Stable secondary structure of the single-stranded DNA target has a major effect on the data quality in that it will interfere with probe binding[146]. Since the

secondary structure is sequence dependent and also influenced by the C+G content, it is impossible to set an absolute limit to the product length. In the initial mutation scanning experiment good signals were obtained from a product of 384 bps, but for the *gyrA* study the target sequence was divided into two separate PCR products, instead of a full-length 300 bp product, to ensure high data quality.

The data in this chapter show that previously unknown base substitutions in a sample could be detected by melt-curve analysis of probes complementary to the reference sequence, since a mutation in the probe-binding sequence would lower the melting temperature of the probe. A heterozygous mutation would therefore result in a melting profile with two peaks, while a homozygous mutation would produce a single peak at a lower temperature compared to a reference sample. Sequence variants are easily detected at any position under the probe, unless the variant position is hybridising to the very last base of the probe. For the detection of a novel mutation (without identifying the exact location and substitution), a set of probes overlapping by at least two base pairs should therefore be able to detect any mutation in the covered sequence. For higher resolution of the mutation location the probes should overlap more. If multiple probes cover the same variant position, redundant information is obtained. This can reduce the risk of false positive results, since a true variation signal should be seen in several adjacent probes and caution therefore will be taken if sequence variation is detected only by one probe. The ultimate resolution would be complete resequencing with four alternative allele probes for each base of the target sequence.

In SNP genotyping by DASH, probes of 11 to 25 bp have been successfully used. The standard length is, however, 17 bp, since this probe length has produced reliable genotyping results in different sequence contexts. Therefore, this probe length was used in this project as well. The probes for mutation detection must be long enough to hybridise efficiently to a sequence with one base mismatch but short enough to provide good separation in Tm of two alleles. The length of the probes, in combination with the probe density, can be changed to either increase the resolution or to reduce costs depending on the needs for a specific project.

For correct identification of novel mutations, when only reference sequence probes are used, a reference sample would be necessary in order to define the matched Tm of each

probe. If a lower temperature peak is observed in the test sample compared to the reference sample (either as a homozygous or heterozygous signal) it would indicate a mutation. The most straight forward design would include an internal wildtype control in each experiment, but there are additional options. For example, in this demonstration of mutation scanning, no reference sample was used for the most common SNP alleles. But since many samples were tested in parallel, genotypes could be determined by comparing the signals of all samples with each other. Another drawback of including a control sample in each run is that it would increase the analysis costs, especially if only one or a few samples are tested on each occasion. A more generic solution would therefore be to compare the data with a recorded reference profile. In that case it should be enough to have one (or few) reference probes to calibrate the temperature between experiments.

When both wildtype and mutation specific probes are used there would be less need for control samples, since the results of the two probes can be internally compared with each other. This strategy was used for the *gyrA* mutation detection, where the samples could also be compared to each other since multiple samples were analysed in parallel. No reference samples were used in the initial analysis of 14 samples, but nevertheless the correct mutation status was determined for each of the strains.

Using a plate format, it is very easy to combine any target-probe combination in separate wells, but a membrane array does not have these natural compartments. To be able to use multiple probes to different copies of the same sequence, the probing procedure had to be adapted, so that individual probes were applied to separate features. In this macroarray format, the probes could be applied in this way by centrifugation transfer from the wells of a microtiter plate, in a similar way that the targets were first transferred to the membrane. For higher multiplex formats, and a miniaturisation of the DASH method the experimental setup will have to be further modified to prevent cross-hybridisation of several probes to the same feature.

In conclusion, an array-based probe melt-curve analysis was found to be a simple but accurate approach to DNA sequence analysis. Multiple mutations can be analysed by different probes simultaneously under standard run conditions, minimising the need for optimisation. The adapted DASH method offers high flexibility of experimental design:

experiments can be designed to either genotype multiple known mutations or to detect any (novel) mutations in a sequence of interest or a combination of both.

CHAPTER 4.

RESEQUENCING BY DYNAMIC HYBRIDISATION

4.1 INTRODUCTION

In the previous chapter, DASH in a macroarray format was used for mutation scanning and simultaneous detection of multiple sequence alterations. Macroarrays could be used for scanning of short sequences or analysis of few mutations, but are not suitable for covering longer sequences and/or a large number of mutations. For such higher throughput applications, it would be desirable to use a microarray format to increase the number of probe-target interactions in a single experiment and to reduce the reaction volumes and thereby the overall reagent costs. For the implementation of DASH on a microarray format, the assay would also have to be turned "upside-down", so that the oligonucleotide probes are attached to the array instead of the target DNA (PCR product). This format is much more suitable when many positions are to be analysed in a single sample, since different copies of the same target sequence can hybridise simultaneously to several probes on different location on the array. In this chapter the feasibility of using dynamic hybridisation on oligonucleotide microarrays to detect thalassaemia mutations is evaluated.

4.1.1 THALASSAEMIA AND HAEMOGLOBIN DISORDERS

Haemoglobin is formed by two α-like and two β-like globin chains. The human globin genes are clustered on chromosome 16 (alpha-like globin gene cluster) and chromosome 11 (beta-like globin gene cluster) (**Figure 4.1**). The genes are arranged in the order they are expressed during embryonic and foetal development. From birth the main form of haemoglobin is Hb A ($\alpha_2\beta_2$) with a small fraction of Hb A_2 ($\alpha_2\delta_2$)[201]. These proteins are coded by the two almost identical α-globin genes (*HBA1* and *HBA2*) and the β-globin- (*HBB*) and δ-globin gene (*HBD*) respectively.

Thalassaemias are a heterogeneous group of disorders that are characterised by globin chain imbalance, due to reduced or absent synthesis of one or more of the globin chains, while other haemoglobin disorders, such as sickle-cell disease, are caused by structurally abnormal proteins[202]. The thalassaemias can be grouped according to which gene is affected; α-thalassaemia and β-thalassaemia being the clinically most important since the main adult form of haemoglobin is formed by alpha- and beta-globin chains[201].

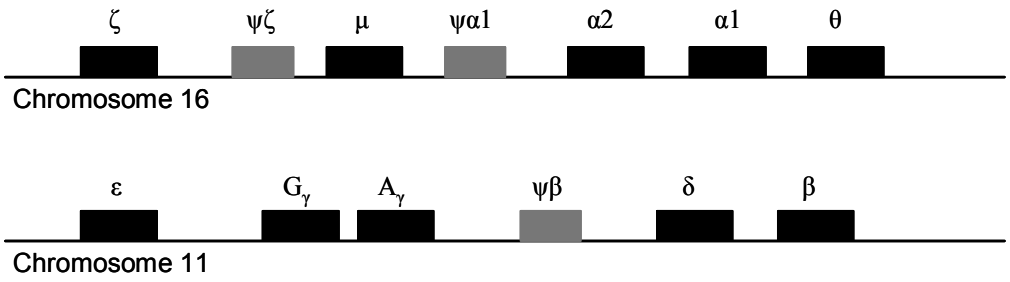


Figure 4.1 Human globin genes.

The alpha-like globin gene cluster is located on chromosome 16 and the beta-like globin gene cluster is located on chromosome 11. Expressed genes are labelled in black and pseudogenes in grey. The genes are arranged in the order they are expressed during embryonic and foetal development. In the embryo Gower 1 (ξ2ε2), Gower 2 (α2ε2) and haemoglobin Portland (ξ2γ2) are expressed, in the foetus haemoglobin F (α2γ2). In adults haemoglobin A (α2β2) is the most common type with low levels of haemoglobin A2 (α2δ2).

Different kinds of mutations can lead to thalassaemia. Large deletions can remove whole genes or large parts of genes that result in complete absence of the gene product (causing α^0-thalassaemia and β^0-thalassaemia respectively[203]). Other causative mutations are point mutations in the form of base substitutions and small insertions or deletions within the coding sequence of the gene, or in its regulatory sequence. Some of these mutations are clinically milder than the large deletion, in that they only reduce the amount of globin produced (causing α^+-thalassaemia and β^+-thalassaemia respectively[203]), but point mutations can be deleterious when they cause premature stop codons, frame shifts, alter the start codon or affect a splice site. Detailed information of all known globin mutations can be found in the HbVar database[204] (http://globin.bx.psu.edu/hbvar). Over 1,300 different mutant alleles are present in the database (as of December 2007), of which 734 affect the *HBB* gene (241 causing β-

thalassaemia) and 385 affect either of the HBA1 and HBA2 genes (72 causing α-thalassaemia.

Most thalassaemia mutations are recessive, usually heterozygous carriers have no or very mild symptoms if one of the two beta-globin genes or at least two of the four alpha-globin genes are functional. Correct molecular diagnosis of carriers is however important for future disease prevention in populations with a high frequency of thalassaemia, such as the populations around the Mediterranean, in Africa and South East Asia [205].

4.1.2 THALASSAEMIA DIAGNOSTICS

The diagnosis of thalassaemia usually involves a combination of haematological and molecular tests. Results from haematological tests, such as red blood cell counts and the quantitation of Hb A_2 and Hb F[201], can indicate which gene is affected, and that gene can then be examined in more detail to determine the carrier allele. The gene is usually first scanned for common mutations, and only if none of those are present is the gene analysed in greater detail. Common mutations in the α- and β-globin genes have been diagnosed with a set of mutation-specific assays, using a great variety of methods[201-203,206] and commercial kits are also available. Which methods are most suitable depends on the type of mutation being assayed. While α-thalassaemia is often caused by large deletions of one or both α-globin genes on the same chromosome, most cases of β-thalassaemia are caused by different point mutations[206].

Single base pair substitutions in the globin genes can, for example, be detected with restriction enzyme digestion, allele-specific PCR and allele-specific oligonucleotide hybridisation. Real-time PCR assays have also been used for mutation detection in the β-globin gene[138,207], using solution-based melt-curve analysis of probes hybridising to common mutation sites. A homogeneous version of allele-specific PCR has also been demonstrated for β-thalassaemia mutations, where the mutation-specific products are identified by their specific melting temperature, which is recorded by SYBR Green I fluorescence detection while heating the sample directly after PCR[98]. These homogeneous assays are fast and convenient, but although a few mutations located at

close proximity can be distinguished by the same probe, different mutations usually have to be analysed in separate reactions.

For the simultaneous detection of multiple mutations, microarray assays have been developed for sets of common mutations. Mutations have been detected by allele-specific arrayed primer extension[208] and in combination with minisequencing with differently labelled ddNTPs[209]. Multiplex minisequencing of 15 common β-thalassaemia mutations has also been done in solution with detection by capillary electrophoresis, where the different loci were identified due to their specific primer length[210].

For large deletions other methods are more suitable. The most common method, if the deletion boundaries are known is 'gap-PCR'. In this approach amplification primers are designed so that one sits on either side of the deletion break-point. These primers therefore amplify a shorter product if the deletion is present, compared to the wildtype sequence. If the deletion is large, the wildtype would not amplify successfully and then an extra primer, annealing within the deleted sequence, could be included as an internal control to amplify a product only in the wildtype. The control product is usually designed to be longer than the mutation-specific product and can therefore be distinguished by electrophoresis methods. Up to seven common α-thalassaemia mutations can be detected simultaneously in a multiplex deletion-specific assay[211,212]. Instead of detection by gel electrophoresis, deletion products have also been detected by hybridisation to microarrays[213]. If the deletion break points are unknown, large deletions can be detected by Southern blot analysis.

For the diagnosis of rare mutations, the full extent of each globin gene can be scanned for mutations by methods such as DHPLC[214,215], DGGE and CDCE[216], temporal temperature gradient gel electrophoresis (TTGGE)[217] or SSCP. Mutations can also be revealed by DNA sequencing, which is usually also used to confirm the exact location and nature of mutations initially detected by another method.

To summarise, many different methods have been used for mutation detection in the globin genes. The microarray format provides good possibilities for highly parallel analysis of many mutations, and also for complete resequencing of the globin genes. If dynamic detection can be used in this format, it opens the possibility to do both

mutation scanning to detect novel mutations, and to score common mutations, reliably on the same platform.

4.2 REVERSE-DASH

The goal in this chapter was to alter the reaction principle of DASH, so that the probes are attached to a solid support and the target DNA is applied in solution to hybridise to the probes, rather than the other way around. This format would enable microarray applications of DASH. This "reverse-DASH" strategy was first evaluated on a macro-scale. This was done on the standard microtiter plate format used for DASH-1, in collaboration with another member of the lab, Lisa Åkerman.

4.2.1 MODEL EXPERIMENTS ON MICROTITER PLATE FORMAT

Previous attempts to turn the DASH concept "upside-down", i.e. hybridising PCR product to oligonucleotides attached to a solid support followed by melt-curve analysis, had not been very successful. It had been difficult to obtain efficient hybridisation of the PCR product, probably because the two strands of the PCR product could easily re-anneal to each other instead of hybridising to the oligonucleotide probes.

The reaction conditions were therefore thoroughly optimised on a model system, using only synthetic DNA, where either single-stranded or double-stranded (a mixture of two complementary oligonucleotides) target DNA was hybridised to target-specific oligonucleotide probes that were bound to the wells of a microtiter plate. In order to obtain strong and specific signals the iFRET system was used, where the surface-bound oligonucleotides were labelled with a fluorescent ROX group at the 5' end, while they were specifically bound to the streptavidin-coated wells by a biotin group at the 3' end.

Several parameters were investigated using this model system. The oligonucleotide concentration on the surface was optimised by comparing the signal strengths of probe concentrations between 1.5 and 24 pmols of probe per well. The strongest melting peaks were observed when 3 to 6 pmols of the oligonucleotide was allowed to bind to the streptavidin-coated surface. Both higher and lower concentrations gave lower signals, where steric hindrance probably reduced the accessibility of the target DNA when

higher concentrations were used. The length of the surface-bound probe was also investigated by comparing probes of 17, 25, 31 and 37 nucleotides in length. 25 bp was found to be a good length, producing the strongest and sharpest melting peaks. It was clear from these experiments that the signal strength was reduced when the degree of double-stranded DNA was increased as the complementary strand was titrated from 0 to 100% of the target strand. This shows the importance of having as much single-stranded target DNA as possible, since that will be available for hybridisation to the surface-bound oligonucleotides. The melting peaks were also stronger if partly double-stranded target DNA was first heat denatured and quickly cooled on ice to keep the target in the single-stranded form before it was incubated to hybridise to the oligonucleotides. Additionally, longer hybridisation times were beneficial, at least one hour was required, as shorter hybridisation times resulted in dramatically weaker signals.

4.2.2 ASYMMETRIC PCR

For the analysis of a genomic target, the sequence would have to be amplified before hybridisation to the oligonucleotide probes. At the same time it is important that the target is prepared in a way that produces single-stranded DNA rather than double-stranded DNA. PCR is the easiest option for amplification, and single-stranded product could, for example, be obtained by post-PCR treatment where one of the product strands is specifically degraded.

An alternative way to produce a large proportion of single-stranded product, without the need for post-PCR treatment is to perform asymmetric PCR. If the two primers are present at different concentration, the excess primer, at the higher concentration, will continue to amplify one of the target strands when the other primer has been used up. Asymmetric PCR can be difficult to optimise if the primers used are designed for normal symmetric PCR. The reason for this is that the Tm of each primer is affected by its concentration, leading to a decrease in Tm at lower concentrations. This will cause inefficient amplification or, alternatively, increased risk of non-specific amplification if the annealing temperature is reduced to compensate for the change in Tm.

This problem can be reduced or eliminated if the concentration effect of primer Tm is taken into account when designing the primers. This is done in the amplification

strategy called Linear-After-The-Exponential (LATE)-PCR[218,219]. Asymmetric reactions with the concentration of limiting primer being 1-10% of the excess primer was shown to be both more efficient and more specific when the limiting primer was designed to have a concentration-adjusted Tm higher than the Tm of the excess primer. For continued efficiency of the linear phase, when the single-stranded product is created, the primer Tm should also be as close as possible to the amplicon Tm[219], since re-annealing of the two product stands compete with primer annealing. This amplification strategy has previously been used for real-time PCR with fluorescent detection of the amplification product by molecular beacons[218,219] and template preparation for pyrosequencing[220].

The LATE-PCR amplification strategy was tested for asymmetric target amplification for reverse-DASH analysis where the limiting primer was used at 5% of the excess primer concentration. The pilot study amplified three different targets between 125 and 218 bp for the genotyping of internal SNPs, and all assays produced strong signals for both SNP alleles with clear separation of the different genotypes with little optimisation needed. Asymmetric PCR with this primer design strategy could therefore be a good option for target preparation for reverse-DASH applications.

4.3 MUTATION DETECTION IN THE β-GLOBIN GENE

To test the resequencing by dynamic hybridisation strategy for detection of thalassaemia mutations, the beta-globin gene was chosen for this first study. Since most of the mutations leading to beta thalassaemia are single base pair mutations in the *HBB* gene most of these mutations should be possible to detect by a resequencing strategy.

An important consideration when studying any of the globin genes at the genomic- or mRNA level is to be observant about the sequence similarity to other globin genes. Regarding the beta-globin gene, in particular the exon sequences are similar to those of the other genes in the beta-globin gene cluster (*HBE*, *HBG2*, *HBG1* and *HBD*), and of those the highest similarity is to *HBD* where the sequence identity is over 90%. This sequence similarity must be taken into account when designing amplification primers, so that one gene is specifically amplified.

4.3.1 ASYMMETRIC PCR IN THE *HBB* GENE

For specific amplification of the *HBB* gene, it would be desirable to amplify the gene either in a single fragment, or in only a few separate fragments, to avoid placing primers where the sequence similarity is high to other globin genes. Primers were therefore designed to amplify the β-globin gene in a single fragment (1.9 kb), as well as in three separate fragments (650-800 bp). All these fragments were successfully amplified in standard, symmetric, PCRs, but it was difficult to obtain specific products with asymmetric PCR, despite designing primers according to the LATE-PCR criteria.

Asymmetric PCR is typically used for amplifying shorter product, up to around 200 bp. Since product length was shown to be important for data quality in DASH, shorter products will probably also be beneficial in reverse-DASH. Primers were therefore designed, using the LATE-PCR strategy, to amplify shorter, asymmetric, products within the *HBB* gene. These asymmetric products could be amplified directly from genomic DNA, or in a secondary reaction using an aliquot of the whole-gene product as template. There are two main benefits of using a secondary reaction; it would require less genomic DNA, since only one reaction is done with genomic template DNA, and it would increase the specificity, since the primary reaction assures that fragments from only the *HBB* gene are amplified.

Seven different products between 145 and 435 bp (**Table 4.1**) were successfully amplified with a primer ratio of 20:1, using the 1.9 kb primary product as template DNA at a concentration of $1{:}10^6$, during 60 cycles of PCR. This primer design strategy was successful, since the majority of asymmetric reactions worked well with a range of different annealing temperatures, and therefore did not require extensive optimisation. The single-stranded products, produced by the excess primer, were seen as shadow bands below their double-stranded product (which all produce bands of the expected sizes) when analysed by agarose gel electrophoresis (**Figure 4.2 a**). By just analysing the products with gel electrophoresis it is impossible to know whether these bands were actually single-stranded products and not other non-specific products. To confirm that the lower bands were the single-stranded products all asymmetric products were also run on an alkaline gel, where all fragments run single-stranded. If the second band was truly a single-stranded version of the targeted product they should now run at the same

speed, while if the second band was another double-stranded product it should run faster also on the alkali gel. Only one main band was seen for each product at the correct product size (**Figure 4.2 b**), indicating that all extra bands were indeed the single-stranded version of each product.

4.3.2 MICROARRAY HYBRIDISATION

The original aim was to develop dynamic hybridisation detection on microarrays in collaboration with Febit biomed GmbH (Heidelberg, Germany), on their GENIOM microarray platform. In this system, oligonucleotide probes are synthesised directly on a chip in eight sub-arrays, providing a flexible platform for any chosen sequence. The main application on this instrument is expression studies, for which the standard protocols were optimised, and alterations to these protocols were therefore necessary for melt-curve analysis. While our laboratory was responsible for developing and testing the reaction chemistry, the company was responsible for adjusting the hardware and software to allow dynamic hybridisation on the instrument. At the time for this thesis work, it was still not possible to heat the sample and detect fluorescence at the same time on the GENIOM instrument. Therefore it was not possible to perform melt-curve analysis, so instead an initial pilot study was undertaken using static hybridisation. This type of analysis would still be useful to determine the hybridisation specificity to the array, and to compare different target preparation methods and labelling strategies.

An oligonucleotide microarray was therefore designed on the GENIOM platform, where sets of probes for three different regions of the *HBB* gene were present (**Figure 4.3**). For each position in these three regions eight 25-bp probes were present, four oligonucleotides with either A, C, G or T in the central position, targeting each of the forward- and reverse strands. The target sequences for the three regions on the array were amplified in three short PCRs (140-264 bp), and pooled before hybridising to the array. These products were amplified with both symmetric and asymmetric PCRs. The products were also labelled with different strategies, and the targets prepared in different ways were hybridised to identical sub-arrays on the same chip.

Table 4.1 Asymmetric PCRs in the *HBB* gene.

No.	Primer Name	Primer Sequence	Conc.	Tm(°C)	Length
1	HBB1-01F	GAAGTCCAACTCCTAAGCCAGT	1 µM	58.3	264 bp
	HBB1-04R	GGGCAGTAACGGCAGACTTCTCCT	0.05 µM	60.7	
2	HBB1-05F	TAGAAACTGGGCATGTGGAGAC	1 µM	58.6	145 bp
	HBB1-07R	AGATCCCCAAAGGACTCAAAGAACCTC	0.05 µM	58.3	
3	HBB3-01F	CATGCCTCTTTGCACCATTCTA	1 µM	57.2	209 bp
	HBB2-02R	GGGCCTAGCTTGGACTCAGAATAATCC	0.05 µM	58.7	
4	HBB1-06R	CCCAAAGGACTCAAAGAACCTC	1 µM	57.4	145 bp
	HBB1-06F	ACCAATAGAAACTGGGCATGTGGAGAC	0.05 µM	58.8	
5	HBB1-03R	CAGTAACGGCAGACTTCTCCT	1 µM	58.3	262 bp
	HBB1-04F	TGAAGTCCAACTCCTAAGCCAGTGC	0.05 µM	58.9	
6	HBB1-05F	TAGAAACTGGGCATGTGGAGAC	1 µM	58.6	433 bp
	HBB1-02R	TGCAATCATTCGTCTGTTTCCCATTCT	0.05 µM	57.2	
7	HBB1-01R	AATCATTCGTCTGTTTCCCATTCT	1 µM	56.3	435 bp
	HBB1-06F	ACCAATAGAAACTGGGCATGTGGAGAC	0.05 µM	58.8	

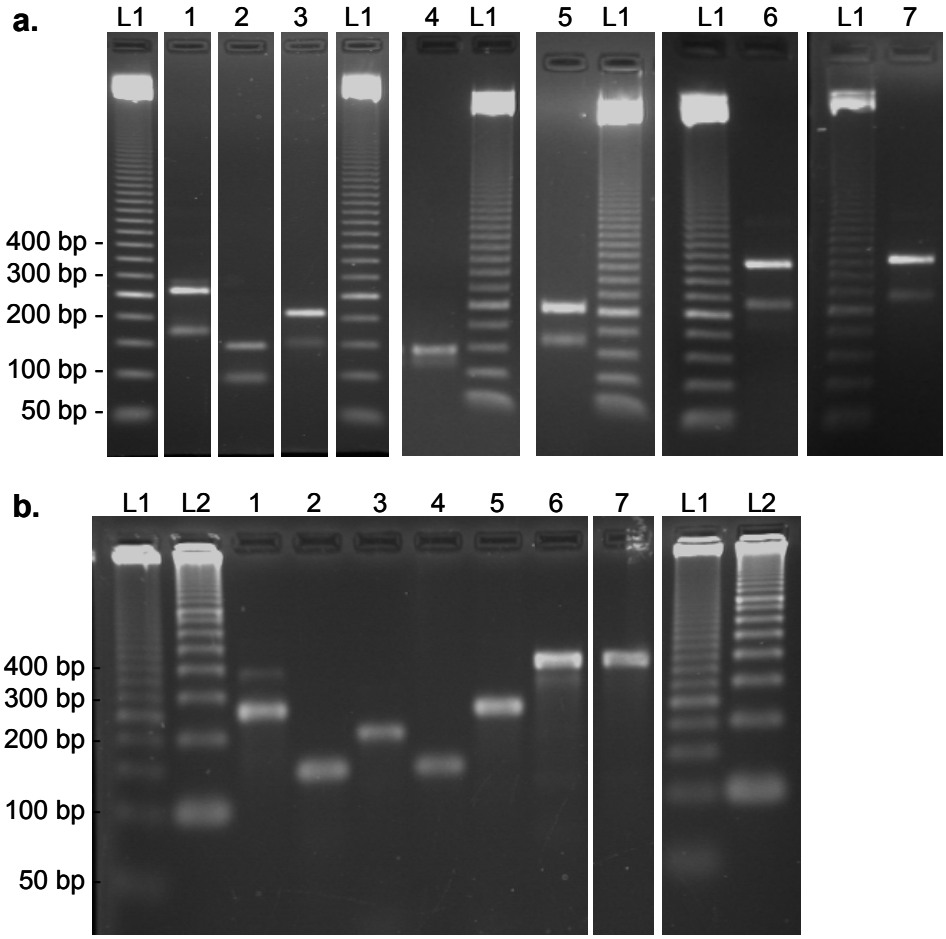


Figure 4.2 Asymmetric PCRs in the *HBB* gene.

Seven asymmetric PCR products (1-7, as numbered in **Table 4.1.**) were analysed by agarose gel electrophoresis, and compared to size standards, L1: 50 bp ladder and L2: 100 bp ladder.

a) Neutral agarose electrophoresis. The gel was pre-stained with ethidium bromide and run in neutral TBE buffer containing ethidium bromide.

b) Alkaline gel electrophoresis. The products were run single-stranded in alkaline buffer and the gel was post-stained with ethidium bromide after neutralisation.

HBB genomic sequence

>hg18_refGene_NM_000518 range=chr11:5203178-5205077
revComp=TRUE strand=-

atatatcttagagggagggctgagggttt**gaagtccaactcctaagccag** c.-201
tgccagaagagccaaggacaggtacggctgtcatcacttagacctcaccc c.-151
tgtggagccacaccctagggttggccaatctactcccaggagcagggagg c.-101
gcaggagccagggctgggcataaaagtcagggcagagccatctattgctt c.-51
ACATTTGCTTCTGACACAACTGTGTTCACTAGCAACCTCAAACAGACACC c.-1
*ATG*GTGCATCTGACTCCTG**AGGAGAAGTCTGCCGTTACTGCCC**TGTGGGG c.50
CAAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCTGGGCAGgttggtat c.92+8
caaggttacaagacaggtttaaggag**accaatagaaactgggcatgtgga** c.92+58
gacagagaagactcttgggtttctgataggcactgactctctctgcctat c.93-23
tggtctattttcccacccttagGCTGCTGGTGGTCTACCCTTGGACCCA**G** c.120
AGGTTCTTTGAGTCCTTTGGGGATCTGTCCACTCCTGATGCTGTTATGGG c.170
CAACCCTAAGGTGAAGGCTCATGGCAAGAAAGTGCTCGGTGCCTTTAGTG c.220
ATGGCCTGGCTCACCTGGACAACCTCAAGGGCACCTTTGCCACACTGAGT c.270
GAGCTGCACTGTGACAAGCTGCACGTGGATCCTGAGAACTTCAGGgtgag c.315+5
tctatgggacgcttgatgttttctttccccttcttttctatggttaagtt c.315+55
catgtcataggaaggggataagtaacagggtacagtttagaatgggaaac c.315+105
agacgaatgattgcatcagtgtggaagtctcaggatcgttttagtttctt c.315+155
ttatttgctgttcataacaattgttttcttttgtttaattcttgctttct c.315+205
ttttttttcttctccgcaatttttactattatacttaatgccttaacatt c.315+255
gtgtataacaaaaggaaatatctctgagatacattaagtaacttaaaaaa c.315+305
aaactttacacagtctgcctagtacattactatttggaatatatgtgtgc c.315+355
ttatttgcatattcataatctccctactttatttttcttttatttttaatt c.315+405
gatacataatcattatacatatttatgggttaaagtgtaatgttttaata c.316-396
tgtgtacacatattgaccaaatcagggtaattttgcatttgtaattttaa c.316-346
aaaatgctttcttcttttaatatacttttttgtttatcttatttctaata c.316-296
ctttccctaatctctttctttcagggcaataatgatacaatgtat**catgc** c.316-246
ctctttgcaccattctaaagaataacagtgataatttctgggttaaggca c.316-196
atagcaatatctctgcatataaatatttctgcatataaattgtaactgat c.316-146
gtaagaggtttcatattgctaatagcagctacaatccagctaccattctg c.316-96
cttttattttatggttgggataaggct**ggattattctgagtccaagctag** c.316-46
gcccttttgctaatcatgttcatacctcttatcttcctcccacagCTCCT c.320
GGGCAACGTGCTGGTCTGTGTGCTGGCCCATCACTTTGGCAAAGAATTCA c.370
CCCCACCAGTGCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAAT c.420
GCCCTGGCCCACAAGTATCAC*TAA*GCTCGCTTTCTTGCTGTCCAATTTCT c.*26
ATTAAAGGTTCCTTTGTTCCCTAAGTCCAACTACTAAACTGGGGGATATT c.*76
ATGAAGGGCCTTGAGCATCTGGATTCTGCCTAATAAAAAACATTTATTTT c.*126
CATTGCaatgatgtatttaaattatttctgaatattttactaaaaaggga c.*176
atgtgggaggtcagtgcatttaaaacataaagaaatgaagagctagttca c.*226

Figure 4.3 *HBB* sequence.

The *HBB* genomic reference sequence, according to genome build 36.2. Exons are in upper case and the start and stop codons in italics. The three regions represented on the microarray are underlined and priming sites for the asymmetric PCR products are in bold. Gene positions for the last nucleotide of each row are indicated at the end of each row, according to the Human Genome Variation Society (HVGS) nomenclature[221] (http://www.hgvs.org).

To allow fluorescence detection of the hybridised targets on the array, the PCR products were labelled with Cy3. This was done either by having one of the PCR primers labelled at the 5' end, or by using Cy3 nucleotides in the PCR. An overview of the hybridisation signals across the array shows that the hybridisation signals were much higher for the end-labelled product, compared to the product internally labelled with Cy3 nucleotides (**Figure 4.4 a**). This data also shows that the asymmetric PCR product is clearly more efficiently hybridised to the array, giving around ten times stronger hybridisation signals compared to the symmetric PCR product. It is noticeable that the hybridisation signals vary across the different probes. This could be partly explained by preferential hybridisation of the target to probes with higher Tm, although this is not a perfect correlation.

Considering the effect of probe Tm, the hybridisation efficiency should also be influenced by the probe length. Three short regions in the gene were interrogated by probes of 21, 23, 25, 27 and 29 bp. Five consecutive bases in each region were interrogated by probes of these different lengths. These three regions were chosen due to their difference in C+G content, which ranged from 32% to 80% in the probes, leading to probe sets with different range in Tm. The hybridisation intensities were generally higher for the probes with higher C+G content (and thereby also having a higher Tm), and for all probe sets the hybridisation intensities increase for the longer probes (**Figure 4.4 d**). The effect of the probe length was larger for the regions with low (position c.-47 to c.-51) or medium (positions c.-173 to c.-177) C+G content, compared to the region with high C+G content (position c.-92 to c.-96), where the Tm is already high for the shorter probes.

The hybridisation specificity could be assessed by comparing the hybridisation to the sense- and antisense probes. This is illustrated with the hybridisation data for an asymmetric PCR product, which is labelled only on the excess strand (by one of the PCR primers). As the excess primer was designed for the sense strand, the product should specifically hybridise to the antisense probes. With only a few exceptions, the hybridisation signals are much higher for the antisense probes (**Figure 4.4 b**), indicating that overall the target is specifically hybridised to the array.

Further, as the hybridisation detection was done at a single temperature, the hybridisation specificity is important for allelic discrimination. In order to determine the sequence, the target needs to preferentially hybridise to the probes for the matching allele at each position. In this array, the sample is usually preferentially hybridised to the probe matching the wildtype allele, but looking across several positions this is not always consistent (**Figure 4.4 c**). Although the hybridisation conditions were not thoroughly optimised in this experiment, it shows the limitations of static hybridisation assays. It can be difficult to obtain preferential hybridisation to the matched allele, especially when many positions are interrogated on the same array, and the hybridisation condition for each probe set therefore cannot be optimised independently.

This problem is irrelevant in dynamic hybridisation assays, since the hybridisation is recorded at a range of different temperatures, and the exact melting temperature therefore can be determined for each probe-target complex. Examples of melt-curve analysis are demonstrated in the next section.

Figure 4.4 Microarray hybridisation data.

a) Overview of the hybridisation to the HBB microarray. The Cy3 hybridisation intensities are shown for the anti-sense probes, all 25 bp long, covering positions c.-212 to c.70 (using the HGVS nomenclature) of the *HBB* gene on three identical sub-arrays. The fluorescence values are shown consecutively for probes with G, A, C and T alleles for each nucleotide position. Array 1: Asymmetric PCR products with Cy3-labelled primers. Array 3: Symmetric PCR product with Cy3-labelled primers. Array 5: Asymmetric PCR product labelled with Cy3-dCTPs. The estimated probe Tm:s are also plotted in the graph.

b) Comparison of hybridisation intensities for the antisense and sense probes in Array 1. The PCR product is asymmetric with excess of the sense strand. The sense strand is also the only labelled strand.

c) Example hybridisation data for eleven consecutive positions, c.93-26 to c.93-16. The target DNA is amplified by asymmetric PCR with Cy3-labelled primers. The hybridisation intensity is plotted for 25-bp probes for the anti-sense strand. For many positions preferential hybridisation is seen for the wildtype allele, but this is not always the case. The reference sequence is shown above each position.

d) Hybridisation efficiencies to probes of different lengths, between 21 and 29 bp, are shown in duplicates. Cy3 hybridisation intensities are shown for three different gene regions with low (positions c.-47 to c.-51), medium (positions c.-173 to c.-177) and high (positions c.-92 to c.-96) C+G content. The fluorescence intensity is plotted for the probes of each allele and each length. The average Tm for each probe length is plotted in the same graph.

Figure 4.4 Microarray hybridisation data.

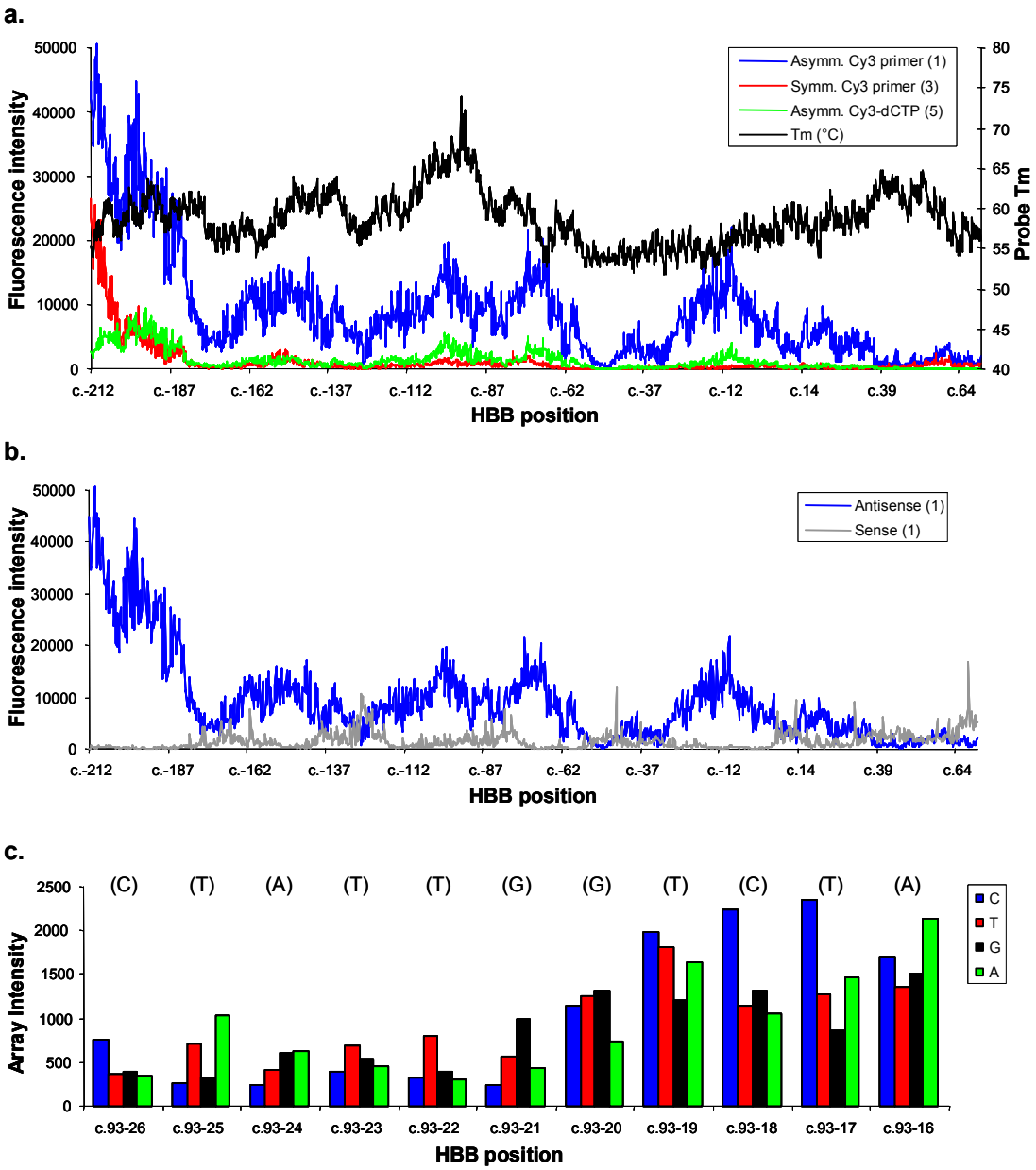
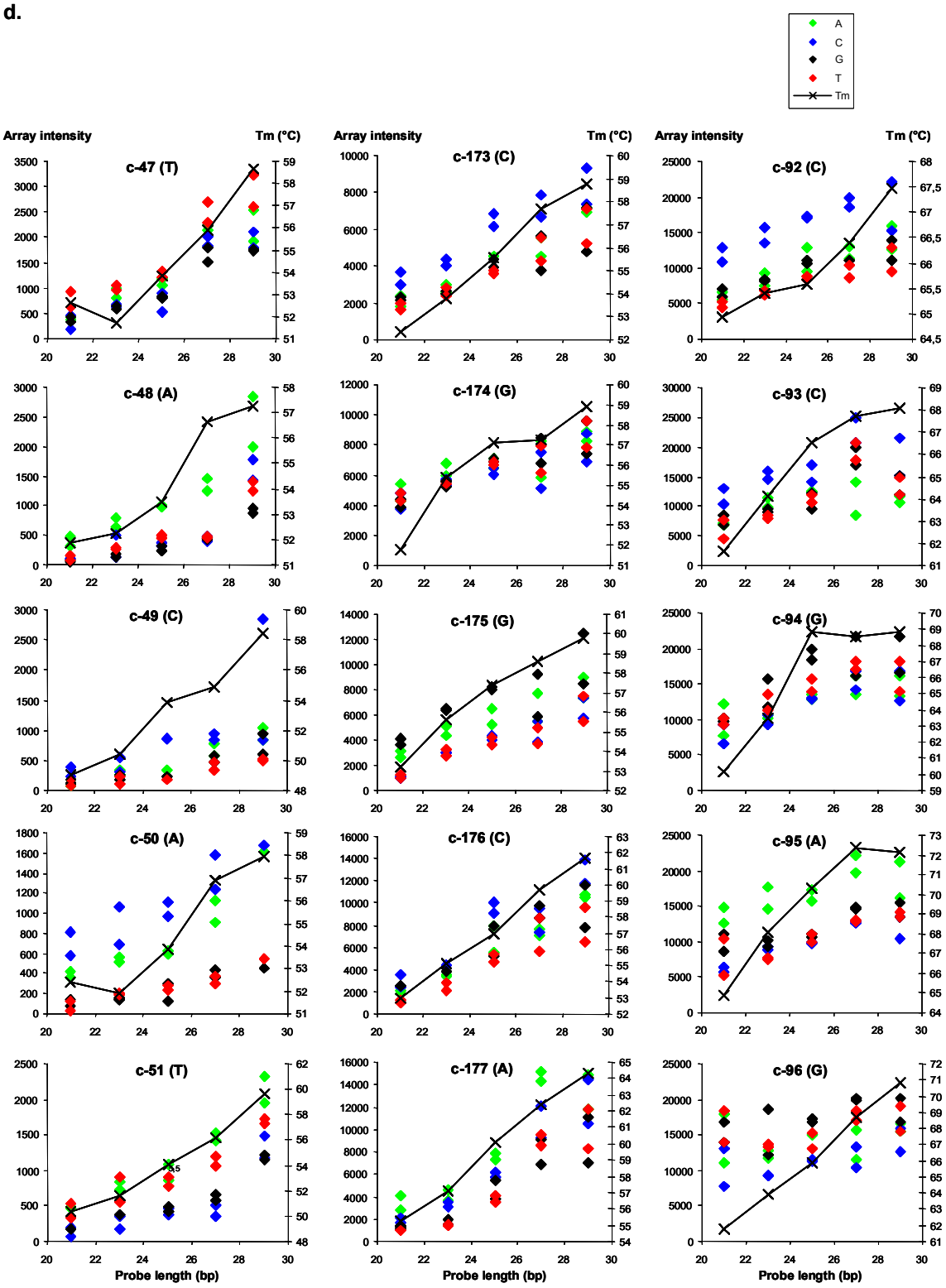


Figure 4.4 continued.

d.



4.3.3 VALIDATION OF MUTATION DETECTION BY MELT-CURVES

The mutation detection in the *HBB* gene was further validated on the microtiter-plate format, since melt-curve analysis on the microarray instrument was currently not possible. Probes were designed to sequence two sets of three bases, centred on known Thalassaemia mutations, for which reference samples were available. These mutations are located at position c.93-21G>A (HGVS mutation nomenclature[221], common name is IVS I-110 G>A) and c.118C>T (common name Cd39 C>T). Each interrogated base was represented by four 25-bp probes, testing for all four possible bases (A, C, G and T) at the central position (**Figure 4.5**).

Probe alignment:

```
                                  c93-21                                                    c118
5'-ctgataggcactgactctctctgcctattggtctattttcccacccttagGCTGCTGGTGGTCTACCCTTGGACCCAGAGGTTCTTTGA-3'
 HBB_c93-22_A_G 3'-GAGAGACGGATAGCCAGATAAAAGG-5'  HBB_c117_A_G 3'-GATGGGAACCTGGGTCTCCAAGAAA-5'
 HBB_c93-22_A_A 3'-GAGAGACGGATAACCAGATAAAAGG-5'  HBB_c117_A_A 3'-GATGGGAACCTGAGTCTCCAAGAAA-5'
 HBB_c93-22_A_C 3'-GAGAGACGGATACCCAGATAAAAGG-5'  HBB_c117_A_C 3'-GATGGGAACCTGCGTCTCCAAGAAA-5'
 HBB_c93-22_A_T 3'-GAGAGACGGATATCCAGATAAAAGG-5'  HBB_c117_A_T 3'-GATGGGAACCTGTGTCTCCAAGAAA-5'
  HBB_c93-21_A_G 3'-AGAGACGGATAAGCAGATAAAAGGG-5'  HBB_c118_A_G 3'-ATGGGAACCTGGGTCTCCAAGAAAC-5'
  HBB_c93-21_A_A 3'-AGAGACGGATAAACAGATAAAAGGG-5'  HBB_c118_A_A 3'-ATGGGAACCTGGATCTCCAAGAAAC-5'
  HBB_c93-21_A_C 3'-AGAGACGGATAACCAGATAAAAGGG-5'  HBB_c118_A_C 3'-ATGGGAACCTGGCTCTCCAAGAAAC-5'
  HBB_c93-21_A_T 3'-AGAGACGGATAATCAGATAAAAGGG-5'  HBB_c118_A_T 3'-ATGGGAACCTGGTTCTCCAAGAAAC-5'
   HBB_c93-20_A_G 3'-GAGACGGATAACGAGATAAAAGGGT-5'  HBB_c119_A_G 3'-TGGGAACCTGGGGCTCCAAGAAACT-5'
   HBB_c93-20_A_A 3'-GAGACGGATAACAAGATAAAAGGGT-5'  HBB_c119_A_A 3'-TGGGAACCTGGGACTCCAAGAAACT-5'
   HBB_c93-20_A_C 3'-GAGACGGATAACCAGATAAAAGGGT-5'  HBB_c119_A_C 3'-TGGGAACCTGGGCCTCCAAGAAACT-5'
   HBB_c93-20_A_T 3'-GAGACGGATAACTAGATAAAAGGGT-5'  HBB_c119_A_T 3'-TGGGAACCTGGGTCTCCAAGAAACT-5'
```

Figure 4.5 HBB probe/target alignment.

Probes used in reverse-DASH analysis in a microtiter plate format. The probe sequences, based on the anti-sense sequence, are shown in the 3' to 5' direction below the target sequence. The interrogated bases in the probes are marked in bold and the mutation positions are underlined in the target sequence.

Two of the asymmetric PCR products contained both these mutation sites, one 145 bp product (primers HBB1-05F and HBB1-07R) as well as a longer 433 bp product (primers HBB1-05F and HBB1-02R). These products were first hybridised to the wildtype probe for each of the two mutations, to compare the results of the two products and to optimise reaction conditions. Both PCR products produced strong hybridisation signals and distinct melting profiles. The shorter PCR product produced higher signals, but the longer PCR product also produced reliable melt curves (**Figure 4.6**). There was also a difference in signal intensity between the two probes. The signals for HBB_c118_A_G_25 were lower, and this probe also produced a more pronounced background signal than the other probe.

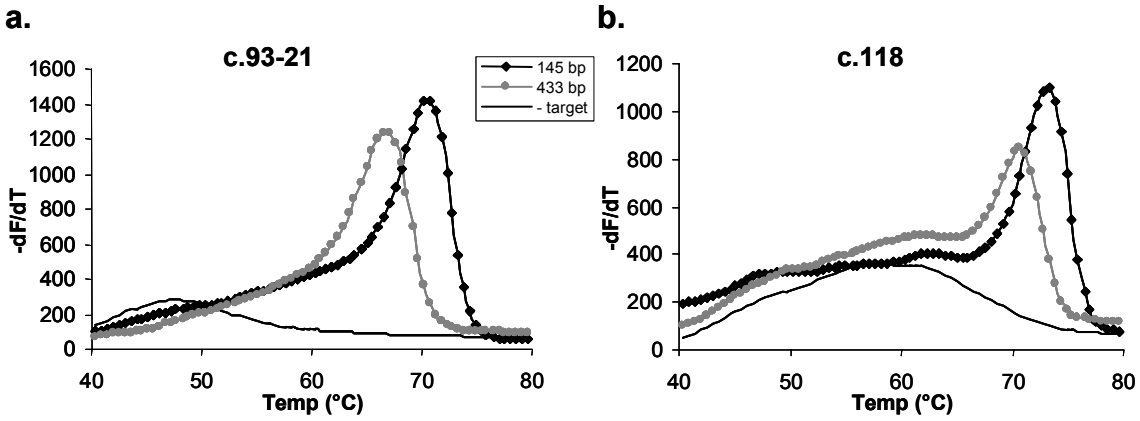


Figure 4.6 PCR product length.

Two asymmetric PCR products, 145 bp (HBB1-05F/HBB1-07R) and 433 bp (HBB1-05F/HBB1-02R) are hybridised to two probes:

a) HBB_c93-21_A_C: GGGAAAATAGAC**C**AATAGGCAGAGA and

b) HBB_c118_A_G: CAAAGAACCTCT**G**GGTCCAAGGGTA.

Their respective melting profiles are compared to the signals obtained from the probes alone in the no-target controls.

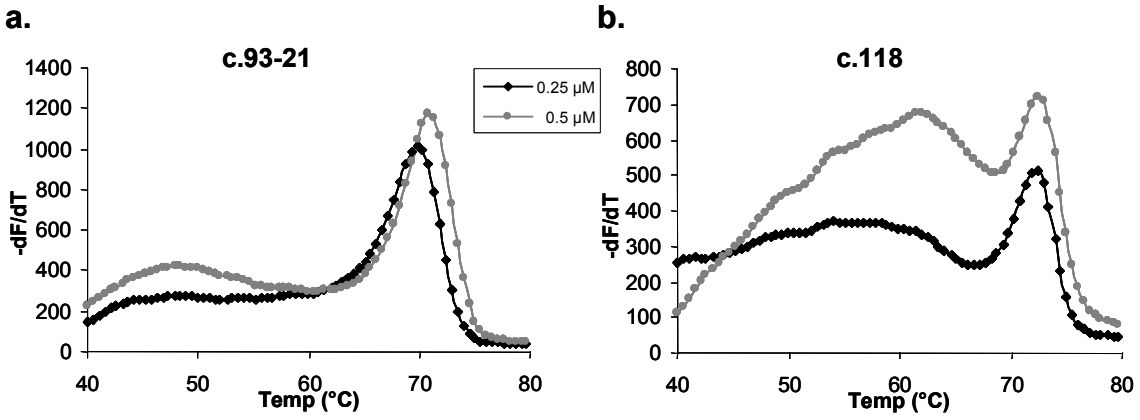


Figure 4.7 Probe concentration on solid support.

Comparison of different probe concentrations of **a)** HBB_c93-21_A_C_25 and **b)** HBB_c118_A_G_25. Increasing the amount of probe bound to the microtiter plate wells from 5 pmol (0.25 µM) to 10 pmol (0.5 µM) only slightly increases the specific probe-target signals while the background signals are much increased in (**b**).

This background was also apparent in the no-product control, indicating that this signal came from secondary structure in the probe and not from any non-specific hybridisation of other parts of the PCR product. The background signal also increased with higher probe concentration, while the specific probe-target signal was only little improved (**Figure 4.7**). The lower probe concentration was therefore preferable.

One heterozygous control sample for each the c.93-21G>A mutation and the c.118C>T mutation were obtained from George Patrinos, Erasmus University Medical Center, Rotterdam, the Netherlands. Together with these samples, two control samples, without any mutations in this region of the gene, were analysed with reverse-DASH using the whole probe-set. The sequence of each interrogated base could be determined by comparing the melting profile of the four probes representing each position (**Figures 4.8 and 4.9**). The probe with the highest Tm indicated which base was present at each position of the gene. Heterozygous mutations were revealed when the two matching allele-probes showed double-peaks, at high and low temperatures, at the same position while the other two probes showed single low temperature peaks. These heterozygous signals are seen in sample B235 at position c.93-21 for the 'A' and 'G' probes. The 'A' signal is not the characteristic double-peak, but rather a fused broad peak, but still clearly different from the mismatch signal in the wildtype control sample. The heterozygous C/T mutation was also correctly detected in sample 1709 at position c.118. Heterozygous mutations could also be revealed by the probes for the surrounding bases, since they will also hybridise to the heterozygous position. This should also result in heterozygous patterns (double-peaks), but if the probes are not testing for the mutation site only one probe will have a signal at the high temperature. Such heterozygous signal is clearly seen for sample B235 at position c.93-22 and, although less pronounced, at position c.93-20 (**Figure 4.8**). Similarly heterozygous signals are seen for sample 1709 at positions c.117 and c.119 (**Figure 4.9**).

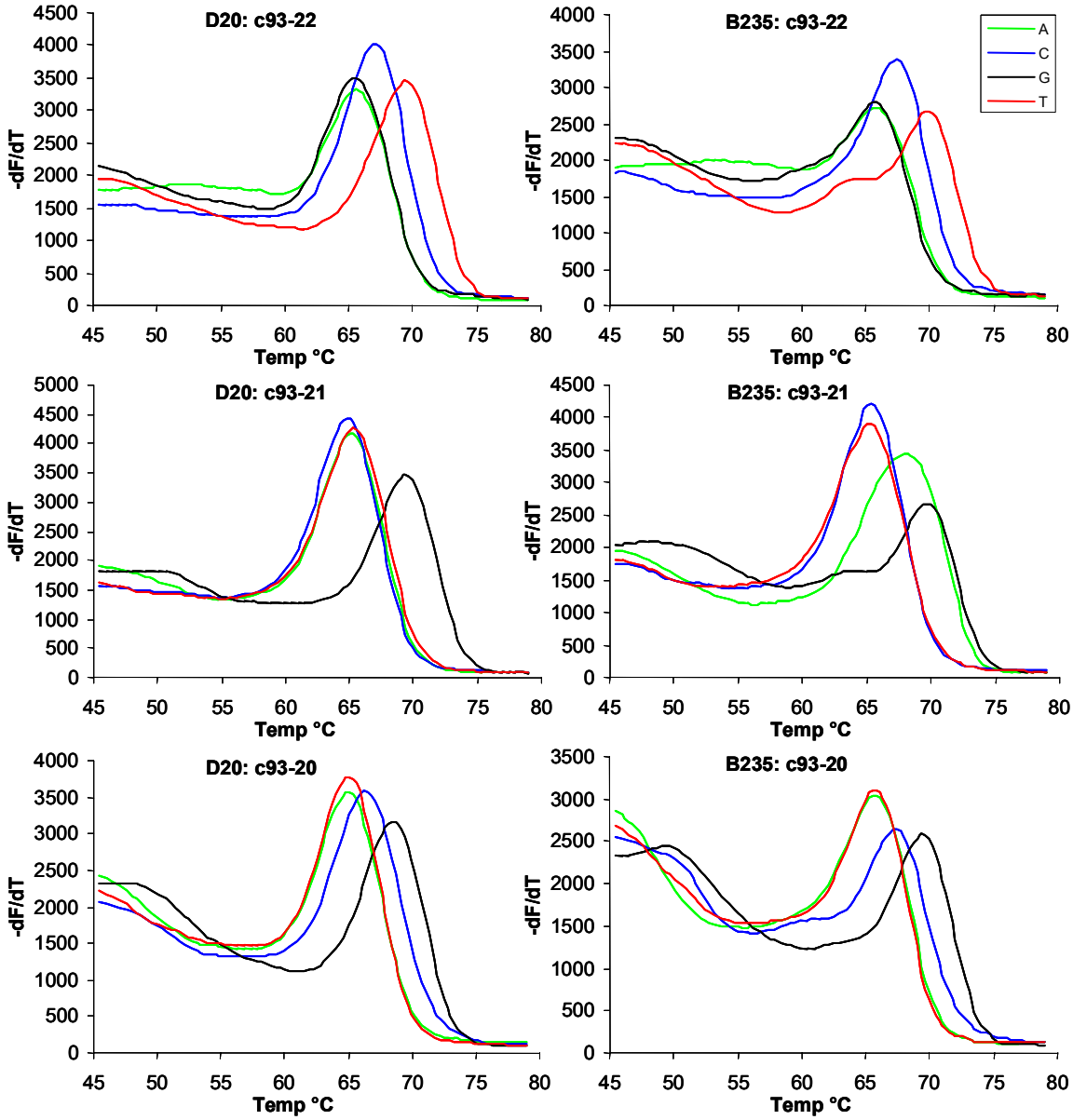


Figure 4.8 Mutation detection at position c.93-21.

Example data from sample D20 (control) and sample B235 (heterozygous for c.93-21G>A). The probe alleles are indicated for the sense strand of the *HBB* gene. The probe with the highest Tm indicates which base is present at each position. The D20 sequence for positions c.93-22 to c.93-20 is therefore 'TGG'. The heterozygous mutation in sample B235 is revealed at position c.93-21, with heterozygous patterns for both 'A' and 'G'. The probe for 'G' produces a double-peak, while the heterozygous 'A' signal has fused into a broad single peak. This is still clearly distinct from the 'A' peak of the control sample.

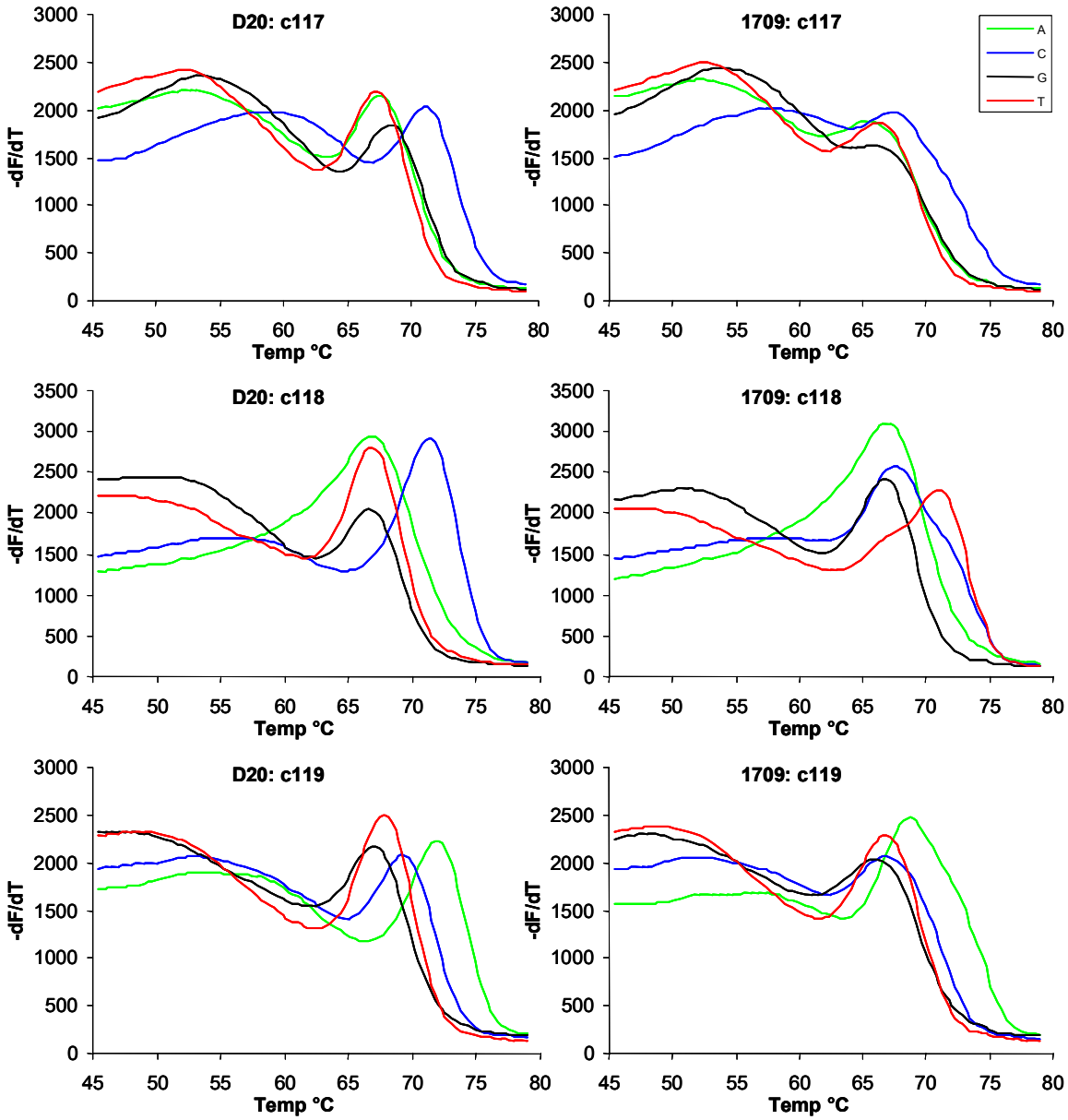


Figure 4.9 Mutation detection at position c.118.

Example data from sample D20 (control) and sample 1709 (heterozygous for c.118C>T). The D20 sequence for positions c.117 to c.119 is interpreted as 'CCA'. Position c.118 clearly reveals a C/T heterozygous signal in sample 1709.

4.4 DISCUSSION

This chapter describes work towards a microarray version of DASH, which when fully developed would provide a platform for resequencing and highly parallel mutation detection by dynamic hybridisation to oligonucleotide probes. Due to technical limitations the dynamic hybridisation on microarrays could not be evaluated properly. Therefore the microarray hybridisation and the mutation detection by melt-curve analysis in the beta-globin gene were evaluated separately.

Compared to the original microtiter plate version of DASH, where the probes are applied in solution-phase to hybridised to PCR products that are attached to solid support (the microtiter plate wells), the experimental setup was turned "upside-down" in a microarray version, to allow different copies of the same target sequence to hybridise to multiple probes. For the most efficient hybridisation of the target sequence to the array it had to be made single-stranded, to remove any risks of the two strands of a double-stranded product to re-anneal. Single-stranded target preparation was achieved by using asymmetric PCR. Using the primer design rules of LATE-PCR[219] the reactions were also easy to optimise. For an efficient asymmetric reaction to produce as much single-stranded product as possible, the primer ratio and concentration are important and had to be optimised. The concentration of the limiting primer was made low enough to promote the single-strand synthesis of the other strand, but high enough to efficiently produce enough template in the first few PCR cycles.

Resequencing and mutation detection in the *HBB* gene was demonstrated in a microtiter plate format. Although the mutation detection was only validated for two mutations, and a total of six bases were sequenced, it illustrates the mutation detection principle. Robust detection of all sequence variants could be done under standard run conditions. In principle, this sequencing could be done without the inclusion of a reference sample, although reference data would be useful for comparison. The base identity for each position was in all cases resolved by identifying the probe with the highest Tm, and heterozygous base substitutions were identified when two probes show heterozygous signals for the same position.

The presence of a heterozygous mutation was also indicated, but not identified, by the signals from surrounding probes that also hybridised to the mutation position. These probes had one base mismatched to the mutated target, and in addition three out of four probes will have an additional mismatch at the tested position. In theory all probes should produce double-peaks, but the signal from probes with two mismatch positions may not be seen due to inefficient hybridisation. Only one probe will have a high-temperature component, identifying the base at the tested position. In a similar way insertion or deletion mutations could be detected, although not identified, since they would also destabilise the probe-target duplex and cause a shift in Tm. In a high density microarray format control probes could be included for known insertion- and deletion mutations to enable the detection of those.

In contrast to the unambiguous sequencing data obtained from the melt-curve analysis, the static hybridisation to the HBB microarray could not correctly reveal the base identity of all tested positions in the gene. Although the hybridisation conditions were not thoroughly optimised, it illustrates the problems with static hybridisation assays. The hybridisation intensities also varied substantially across the different probes of the array. The hybridisation efficiency is partly related to the probe Tm, but probes with similar Tm also showed different levels of hybridisation signals. Since low hybridisation signals may affect the quality of future melting analysis, it could be necessary to adjust probe lengths to obtain more similar Tm, since many different probes compete for the same target. Other possible reasons for differences in hybridisation efficiency could be stable secondary structure in the target or the probe, which could prevent part of the target from hybridising efficiently to the array. Allele-specific differences in secondary structure could even, in some cases, lead to higher hybridisation signals for mismatched probes compared to perfectly matched probes.

In dynamic hybridisation assays it is preferable that the target hybridises equally well to both matched and mismatched allelic probes, as opposed to genotyping by static hybridisation assays where the allelic discrimination is done based on the hybridisation efficiency. While a target has to hybridise efficiently to the wrong allele probe to produce a reliable melting profile, the hybridisation conditions should not allow hybridisation of unrelated sequences since this would cause high background and it might be difficult to identify the melting profile of the true target. The hybridisation

conditions may therefore still have to be carefully optimised even for dynamic hybridisation on microarrays, especially in highly multiplex applications where many different targets should hybridise to different probe sets on the same array.

CHAPTER 5.

MEGAPLEX PCR – OPTIMISATION OF REACTION CONDITIONS

5.1 INTRODUCTION

When it is desirable to access many target sequences, it would be attractive to amplify all sequences simultaneously in a multiplex amplification reaction. For SNP genotyping, highly multiplex analysis is possible on some platforms with complexity-reduced DNA template or whole-genome amplified genomic DNA. For the analysis of longer sequences, for applications such as sequencing, it is still necessary to isolate the target sequences. Based on these needs, MegaPlex PCR was developed to provide a general strategy for highly multiplex amplification of specific DNA sequences.

Many strategies for multiplex amplification have employed common primers to amplify all desired targets in a single reaction using a single primer pair, thereby practically eliminating the risk of primer-dimer formation, which is a major problem in standard multiplex PCR. The challenge is, however, how to specifically introduce these common primer sequences in any set of chosen target sequences.

The strategy we chose for MegaPlex PCR employed solid-phase PCR, to specifically capture the desired target sequences and at the same time build in common adaptor sequences, which would allow amplification of all target sequences with a common primer pair. The method therefore uses primers attached to solid support, where each primer pair is physically separated from any other primer pairs to avoid unwanted interactions between primers in different pairs. The solid-phase primers were attached to the surface at the 5' end, where a common primer sequence was followed by a target specific primer sequence at the 3' end. The general strategy for MegaPlex PCR is illustrated schematically in **Figure 5.1**.

The next section discusses some general issues of the processes used in MegaPlex PCR, before presenting the results on the development of the method.

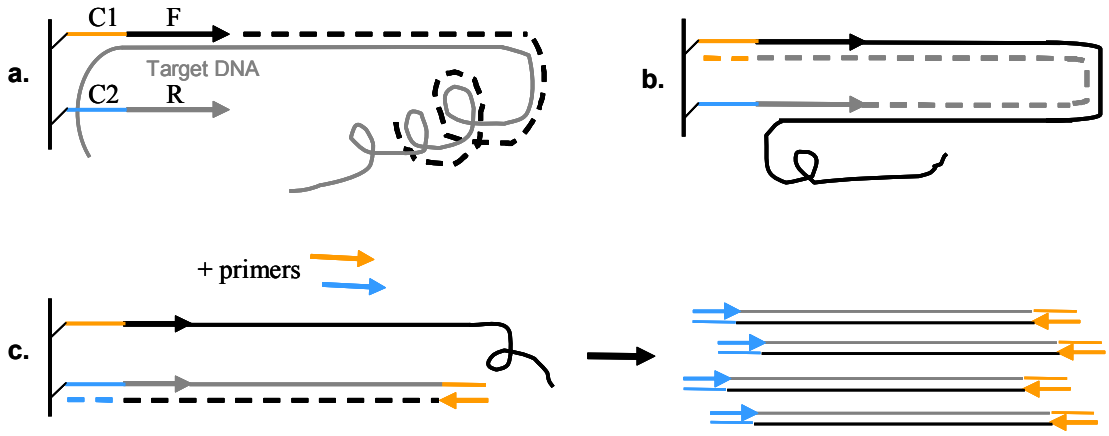


Figure 5.1 Working model for MegaPlex PCR.

Different primer pairs are attached to surface areas. Each solid-phase primer consists of a common sequence at their 5' end (C1 or C2) and a target-specific primer sequence at their 3' end (F or R).

a) Solid-phase PCR, cycle 1: The target DNA is hybridised to one of the surface-bound primers, which is then extended to copy the target sequence. The template DNA is then removed.

b) Solid-phase PCR, cycle 2: The first extension product strand will act as template for the other primer of the primer-pair, which can be extended to produce a double-stranded product. At the end of this extension the complement of the common sequence will be synthesised, creating a copy of the target sequence, flanked by common primer regions at both ends.

c) In the final step of the MegaPlex PCR procedure, all different products are co-amplified simultaneously by a common primer pair in a standard solution-phase PCR.

5.1.1 SOLID-PHASE PCR

Fundamentally, solid-phase PCR involves the annealing to, and extension of, a solid-phase primer. For both of these processes steric hindrance can be a problem. If primers are attached too close together, the efficiency of DNA hybridisation will decrease due to low accessibility for the target DNA. Therefore optimisation of primer density could be important to obtain an efficient reaction. Another option is to introduce spacer molecules between the support and the 5' end of the primer sequence to increase the volume in which the primer and the target DNA can interact.

Previous studies have shown that spacer molecules can improve the efficiency of both the hybridisation[222] and primer extension[171]. However, other studies showed no effect of

spacer length on hybridisation[223,224] while yet another study showed that long spacers had a negative effect on the hybridisation efficiency, but on the other hand increased the primer extension efficiency[225]. It is possible that the spacers only have an effect on hybridisation efficiency at high primer densities and the studies that showed no effect of spacers may have used a lower primer density where the oligonucleotides were easily accessible even without the spacers. It is also possible that different surfaces affect the hybridisation in different ways. Independent of the primer density the distance to the surface is probably important for solid-phase PCR, since primer extension was shown to be more efficient with increased distance to the surface[225], probably due to the accessibility of the DNA polymerase.

5.1.2 COMMON-PRIMER PCR

The final amplification step in the MegaPlex PCR strategy is a standard solution-phase PCR using a single, common, primer pair. This is a convenient amplification step where any combination of MegaPlex products could be amplified simultaneously and with little risk of primer-dimer formation since only one primer pair is used. There are, however, increased risks of contamination from previous PCR products when the same PCR primers are used for many different reactions in the same lab.

One way to prevent amplification of previous PCR products is to perform all common-primer PCRs with dTTP replaced by dUTP and include Uracil DNA Glycosylase (UDG) in the reaction[226]. All PCR products from such reactions will contain uracil instead of thymine and are therefore degradable by UDG. Incubation with UDG before the PCR cycling will degrade the uracil bases present in DNA and thereby prevent contaminating PCR products from being amplified. Normal DNA template containing only A, C, G and T bases, on the other hand, is not degraded. The UDG is heat inactivated at the initial PCR denaturation step in order to enable successful synthesis of new dU-containing PCR products.

This contamination-prevention strategy was used for all common-primer PCRs of MegaPlex products. The initial solid-phase reactions were performed with normal DNA nucleotides, so that only contaminated products and not the desired templates are degraded.

5.2 SOLID-PHASE PRIMER EXTENSION

In MegaPlex PCR, the solid-phase PCR is used to selectively capture specific target sequences. It was therefore important to establish basic reaction conditions for solid-phase PCR to confirm i) that surface-bound primer extension was successful from a target in solution and ii) that the extension product itself could act as a template for other surface-bound primers to create double-stranded solid-phase products with common sequences at both ends.

These reactions were studied using membrane-bound primers and single-stranded DNA template, and each solid-phase extension product was detected by hybridisation of target-specific fluorescent oligonucleotide probes. The experimental setup is illustrated in **Figure 5.2**. A short (50 bp) target sequence was chosen, for which primers had already been designed and were confirmed to efficiently amplify the fragment in normal solution-phase PCR. The forward- and reverse primer for this target sequence, both with a dT(10) tail and common primer sequence at the 5' end (dT10_X727U_B15b06F or dT9U_X727U_B15b06F and dT10_X727_B15b07R) were attached by biotin-streptavidin interactions to separate features on a membrane array as well as to the same feature. A single-stranded synthetic version of the "reverse" strand of the target sequence (JB15-47P) was hybridised to the array and excess target was removed. Primer extension was then achieved by placing the membrane in 100 µl PCR buffer including DNA polymerase and nucleotides.

After the primer-extension reaction the template DNA was removed and extension products were detected by hybridisation of a fluorescent oligonucleotide probe, complementary to the forward strand (HB15+31P). Fluorescence from this probe was detected only from features containing the forward primer (**Figure 5.2 b**). Features containing only the reverse primer produced only low background fluorescence at the same level as no-probe or non-specific probe controls, showing that the specific extension product is only created from the forward target-specific primer.

To demonstrate extension of the reverse primer by priming the first extension product, the above membrane was alkali rinsed to remove the fluorescent probe, and a second round of annealing and extension was conducted, now without target DNA in the

solution phase. In features that had both forward- and reverse primers attached, the reverse primer should now be able to anneal to the forward primer extension product and promote primer extension of the reverse primer to create the reverse strand of the product. Before detection of this newly synthesised strand by probe hybridisation, the initially synthesised strand was removed from the array surface. This was done in order to prevent re-hybridisation of the two product strands, which would compete with the hybridisation of the detection probes. Specific removal of one strand was achieved by including a deoxyuridine (dU) base in the "forward" surface-bound primer, providing a residue that could be cleaved by Uracil DNA glycosylase. After conducting this cleavage and removing the first synthesized strand by an alkali rinse, a fluorescent probe (LB15R+01P) for the second synthesized strand was hybridised to the membrane. As illustrated in **Figure 5.2 e**, only the spot that carried both the forward and the reverse primers produced a fluorescence signal with this probe, indicating that a specific reverse extension reaction had taken place. The fluorescence intensity of this feature is comparable to the fluorescence from the probe binding to the initially extended forward strand, suggesting that the reverse extension reaction is efficient.

However, one of the other probes used in this experiment was also complementary to the reverse strand. This probe (X732+01P) should hybridise to the reverse complement of the common primer sequence at the very 3' end of the extension product, but this probe was not efficiently hybridised to its target. This might indicate that the reverse extension reaction was not complete, but in this experiment the common sequences were identical for the forward- and reverse primers, meaning that if the extension was complete the 5'- and 3' ends of this DNA strand would be complementary to each other and would be able to create a stable secondary structure that would easily compete with the probe hybridisation. Due to this possible "pan-handle" formation the results are not fully conclusive on whether the extension reaction is complete.

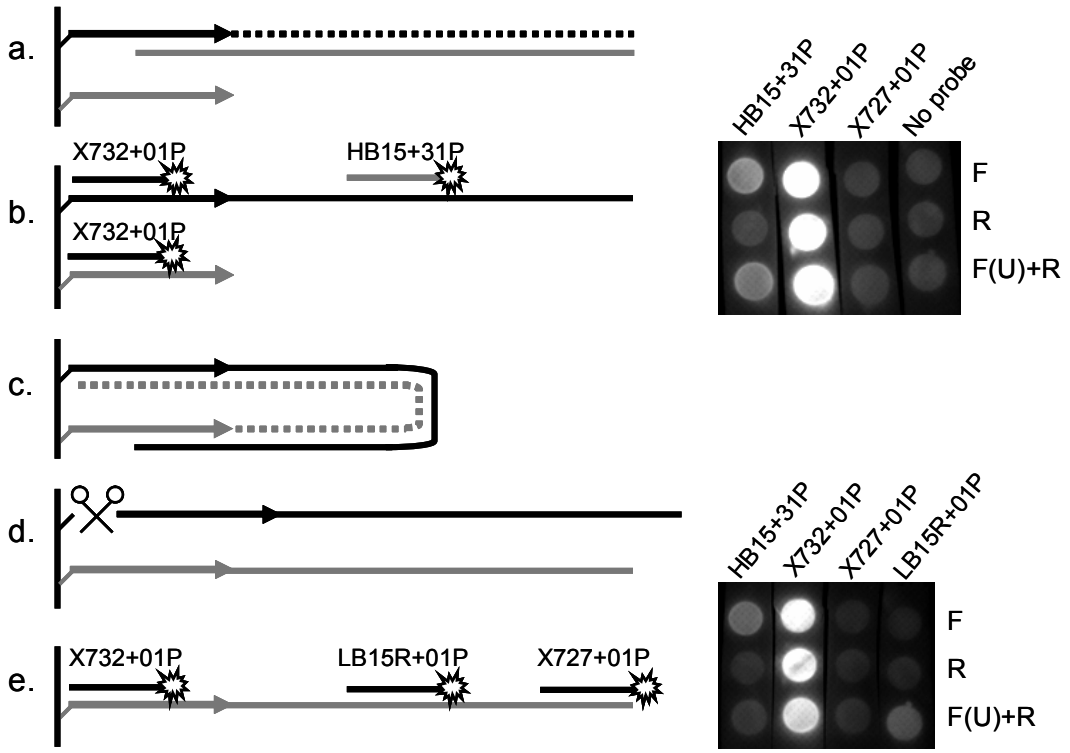


Figure 5.2 Detection of solid-phase primer extension products.

Solid-phase primer extension is demonstrated on membrane arrays, where the forward, F, and reverse, R, primers are bound to separate features as well as to the same feature (where the forward primer, F(U), contains a uracil residue).

a) Solid-phase primer extension of the forward primer, by annealing to target DNA present in solution.

b) The template DNA is removed and three fluorescent detection probes are hybridised to identical membranes. A fourth membrane is used as a no-probe control. The HB15+31P probe is complementary to the extended forward strand and should fluoresce where a forward primer (F) is present. The X732+01P probe is complementary to the common sequence in all surface bound primers and is a positive control. The X727+01P probe is identical to the common primer sequence and could only hybridise to the reverse complement of this sequence (see figure **e.**).

c) The probes are removed and the reverse primer is extended along the first strand.

d) The forward strand containing uracil is cleaved off by UDG and NaOH.

e) Four different fluorescent detection probes are hybridised. The first three probes are identical to the probes in the first hybridisation. The HB15+31P probe should fluorescence only where the intact forward primer F is present (the F(U) primer is cleaved and removed from the surface). The X732+01P probe is still a positive control probe. The X727+01P should now hybridise to the 3' end of the reverse strand (but no fluorescence is seen). The fourth probe, LB15R+01P, is binding internally to the extended reverse strand.

To summarise, these experiments show that solid-phase primer extension was successful both from target DNA in solution and from surface-bound target. It is, however, important that the extension reaction is complete since the reverse complement of the common primer sequence is required for successful MegaPlex amplification. Unfortunately this could not be shown with these particular experiments. In case the extension was actually incomplete, it could be caused by steric hindrance of Taq Polymerase or by restrictions of the flexibility of DNA to bend. Both these factors could be reduced by using longer spacers in the primers between the solid support and the primer sequence. For future experiments it was decided that longer spacers should be used (50 or 60 units of dT) and that the forward- and reverse common tails should have different sequences to avoid strong secondary structures, since these self-binding structures would also compete with primer annealing of a single common primer.

5.3 DEMONSTRATION OF MEGAPLEX PCR

Having confirmed successful primer extension on the solid support, the whole MegaPlex PCR process was then tested. To detect solid-phase products after only two cycles, high concentration of primers and template were required to produce enough product for detection by hybridisation probes. Since MegaPlex PCR should ideally work from genomic DNA template at much lower concentrations, the reaction had to be tested with lower concentrations of target DNA.

5.3.1 MEGAPLEX AMPLIFICATION OF PURIFIED TEMPLATE DNA

As an intermediate before using genomic DNA, an approximately 2 kb long purified PCR product (primers LSCAN-03F, LSCAN-06R) was used as template DNA. This template PCR product was purified by gel electrophoresis and then heat denatured and directly cooled on ice to provide single-stranded template. A set of primers were designed to amplify four internal products between 138 and 488 base pairs within this template DNA. All primers used in this experiment were synthesised with a 5' biotin, followed by a dT(60) tail, common primer sequences (identical for all forward primers and for all reverse primers, but different from each other) and each target-specific sequence at the 3' end. The long dT tail was chosen since it was assumed to be advantageous for the solid-phase PCR to take place further away from the surface and the common primer sequences were designed with two parts (**Figure 5.3**) to allow nested amplification of the MegaPlex products.

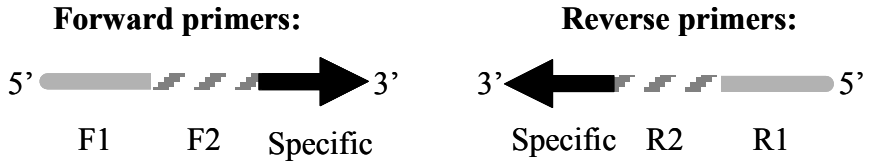


Figure 5.3 MegaPlex primer design.

Each surface-bound primer has three parts: a specific primer sequence at the 3' end and two consecutive common sequences (F1, F2 and R1, R2) at the 5' end.

Since the template concentration of this purified PCR product template was lower than the oligonucleotide template used in previous experiments, the first annealing reaction was carried out for a longer time. Annealing was allowed to take place overnight in PCR buffer, but without DNA polymerase or nucleotides to avoid non-specific primer extension. For the primer extension reaction additional PCR buffer including the polymerase and nucleotides were added to the reaction, resulting in standard concentration of all reagents. After the first extension reaction the template DNA was removed. A second annealing and extension reaction was performed to create the reverse complement of the first synthesised strand. This annealing reaction was shorter (10 minutes), since the target was now linked to the solid support, providing a high local concentration of the targets. When the second extension reaction was complete, the common primer pair was used for further amplification in solution.

Reaction conditions, such as primer concentration, reaction time and temperature were optimised mainly with the shortest of the products (138 bp). MegaPlex amplification product of the 138 bp target (180 bp including common primer sequences) was clearly detectable by agarose gel electrophoresis using down to ~10^{-14} g of the 2 kb template per reaction, after two rounds of (nested) common primer PCR (**Figure 5.4**), although with co-amplification of a shorter non-specific product when ~10^{-13} g or less of the template DNA was employed. This detection limit is equivalent to ~5000 copies of the target, corresponding to ~15 ng human genomic DNA.

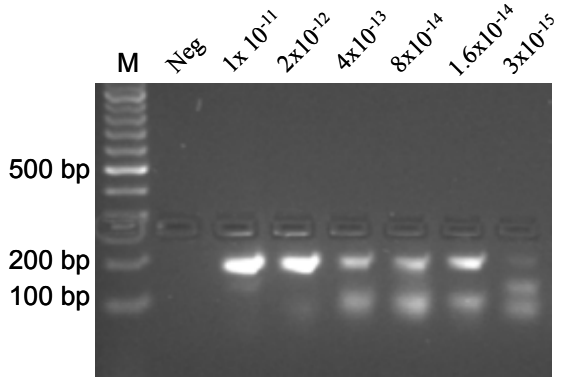


Figure 5.4 Template concentration for MegaPlex PCR.

MegaPlex amplification of a 138 bp target sequence from a 2 kb template, after two rounds of common-primer PCR (30 + 30 cycles). The correct product is seen as a 180 bp fragment due to the added common primer sequences. M: 100 bp ladder. Titration of the template was done in 5-fold dilutions from 10^{-11} (in lane 2) down to $3x10^{-15}$ g (in lane 7). The negative control reaction contained no template DNA.

5.3.2 MEGAPLEX AMPLIFICATION FROM HUMAN GENOMIC DNA TEMPLATE

As MegaPlex amplification was shown successful with low concentrations of the 2 kb DNA template, amplification directly from human genomic DNA should also be possible. Using the initial primer set, all four products within the 2 kb target sequence were successfully amplified from ~1 µg human genomic DNA template. To illustrate the importance of high quality intact DNA for successful amplification of longer products, MegaPlex amplification was performed with both high molecular weight genomic DNA as well as with sonicated DNA, where the majority of fragments were shorter than 500 bp (**Figure 5.5**). The intact DNA could successfully amplify four different targets between 138 and 488 bp, but only the two shortest products were efficiently amplified when the sonicated DNA template was used. It is therefore important to use high molecular DNA template for the amplification of long targets.

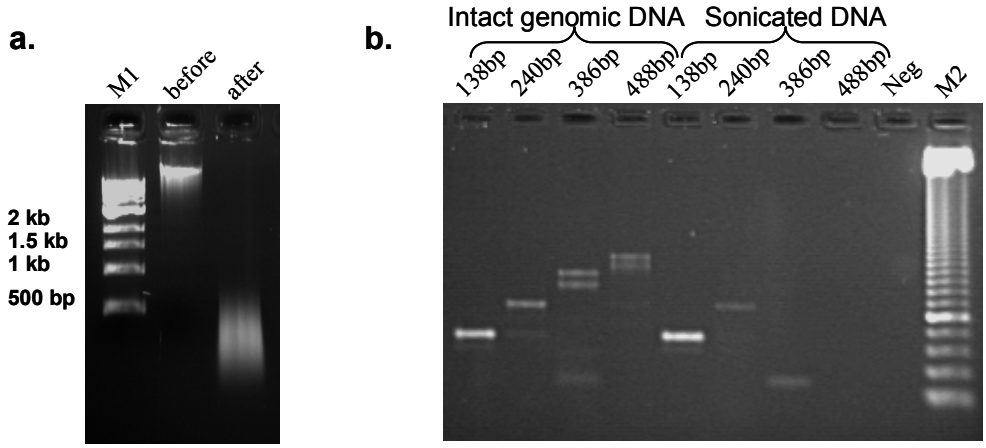


Figure 5.5 Effect of DNA quality on MegaPlex PCR.

a) Human genomic DNA before and after sonication. M1: 1kb ladder, followed by DNA before and after sonication (resulting in 200-500 bp fragments).

b) MegaPlex products visualised by agarose gel electrophoresis. For products in lanes 1-4 the template DNA was intact human genomic DNA and in lanes 5-8 the template was sonicated human genomic DNA. Four sequences were targeted with each template DNA; 138 bp, 240 bp, 386 bp and 488 bp. Lane 9 contains a negative control with no template DNA, and M2 is a 50bp DNA ladder.

In order to evaluate the MegaPlex PCR amplification reaction with a larger target set, new MegaPlex primers were designed for 15 unrelated sequences, all located on different human chromosomes. This primer set was designed to capture three different products each of ~100, ~200, ~300, ~400 and ~500 base pairs (excluding common

primer sequences). This set of primers allowed demonstration of multiplex amplification of independent targets as well as to study any effects of product length on the method. For these experiments each primer pair was separately bound to magnetic beads, which were then pooled in desired combinations for multiplex reactions. Five different 3-plex reactions, with target sequences of the same length, as well as three different combinations of 5-plex reactions for different sized targets were conducted. Additionally, all 15 primer pairs were combined in a single 15-plex reaction. Each reaction used ~2 µg denatured human genomic DNA as template. After the final solution-phase amplification with a common primer pair, products of all tested lengths were successfully detected by agarose gel electrophoresis (**Figure 5.6**).

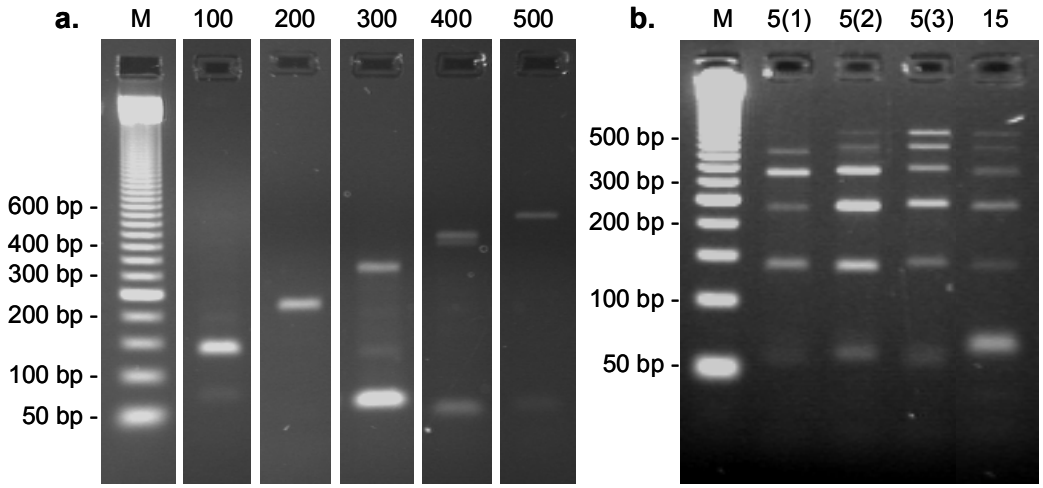


Figure 5.6 Demonstration of multiplex MegaPlex amplification.

MegaPlex amplification of targets of different length, using primers with dT(50) spacers. M=50 bp ladder.

a) 3-plex amplification reactions of targets of the same length; 100, 200, 300, 400 and 500 bp respectively.

b) Three 5-plex amplification of targets of different length: 100+200+300+400+500 bp in each multiplex, and one 15-plex reaction (a combination of the three 5-plex reactions).

Two of the 5-plex assays produced all five distinct products, clearly visible and separated on the agarose gel. The third 5-plex assay showed four visible bands on the gel, lacking only the largest product. The 15-plex reaction showed bands of all target sizes of fairly equal intensities, although it would be impossible to resolve/identify the products of equal size. The overall success in amplification of this first set of 15 random

human genomic targets, in particular the recovering of molecules over a five-fold size range in a single reaction, indicated that most chosen targets, in the tested size range, should be amplifiable with the MegaPlex method. This was encouraging for further improvement of the method and the work towards higher multiplex reactions.

5.3.3 5' SPACERS

As MegaPlex PCR was shown to successfully amplify targets directly from genomic DNA template it was interesting to test the effect of spacer length, between the surface and the primer sequence, once more. Three primer pairs had therefore been ordered with dT(10) tails as well as with the dT(50)-tails that were used in all multiplex experiments described above. A direct comparison of the different spacer lengths was done by performing parallel 3-plex amplifications of the same targets, using the two sets of primers with different dT length. Equal amounts of the products were analysed by agarose gel electrophoresis (**Figure 5.7**). The results show that larger quantities were produced with the shorter, dT(10), spacers, indicating that longer spacers are not necessary.

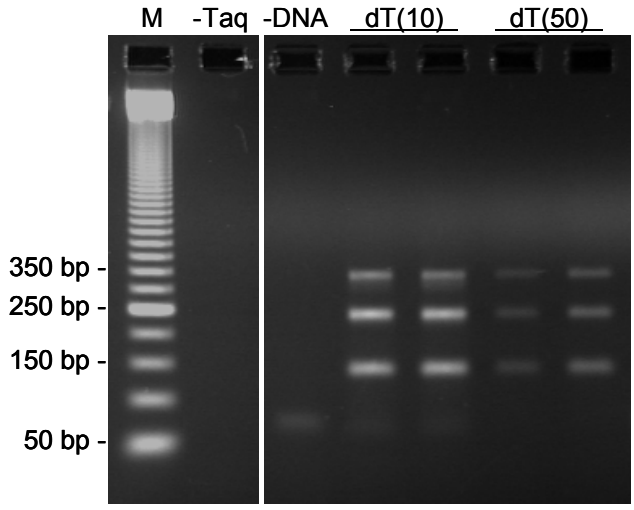


Figure 5.7 dT spacer length.

MegaPlex amplification using solid-phase primers with different length 5' dT tails. Products are detected by agarose gel electrophoresis, after 30 cycles common-primer PCR. M = 50 bp DNA ladder, lane 2 and 3 are control reaction without DNA polymerase or without DNA, lanes 4-5: 3-plex reaction with dT(10) tails and lanes 6-7: 3-plex reaction with dT(50) tails.

5.4 PRIMER-DIMER PREVENTION

In addition to the desired MegaPlex products, short non-specific products were occasionally also detected by the gel electrophoresis. Those products, since they were short, were probably primer-dimers. Primer-dimers can be a serious problem to any PCR, and especially to multiplex reactions, since they are usually very efficiently amplified due to their small size, which means that they can out-compete the desired products. In case of MegaPlex PCR, since the primer pairs are physically separated during the initial cycles, the primer-dimers seen should mainly be generated by primer interactions within the same primer pair and not between primers in different primer pairs. In this case, the amount of primer-dimer should scale linearly with the degree of multiplexing, as would the target amplicons. Therefore the amount of primer-dimers should not dramatically increase by higher multiplex reactions, unlike standard multiplex reactions where all primers are free to interact in solution resulting in $2n^2 + n$ different possible primer interactions when n primer pairs are used[227].

However, for the best performance of MegaPlex PCR, any primer-dimer formation should be avoided if possible. By identifying the main factors that affect primer-dimer formation, solutions to minimise the primer-dimer problem could be developed. One of the main factors for primer-dimer formation was found to be the primer concentration on the solid support, since the distance between primers should logically influence the risk of primers to anneal to each other. Other attempts to reduce primer-dimer formation are also described in the following sections. The primer-dimer artefacts were also seen to vary in intensity between different multiplex combinations, indicating that some primer sequences are more prone to primer-dimer formation, even though no apparent complementarity was seen within those primer pairs.

5.4.1 PRIMER CONCENTRATION ON SOLID SUPPORT

The effects of primer concentration on solid support was evaluated by diluting the primers before binding to the solid support, and then comparing the MegaPlex results from the experiments using different primer concentrations. The amount of biotinylated primers present in the solution that reacts with the streptavidin on the solid support

should indirectly reflect the primer density on the surface, although the exact primer densities on the solid support were never determined.

The results presented in **Figure 5.8** include data from two experiments, with primers attached to membrane and beads respectively. As shown in **Figure 5.8 a**, at the highest primer concentration on the membrane support, the ~200 bp desired product was amplified in conjunction with a ~60 bp primer-dimer. At all lower primer concentrations the primer-dimer was no longer visible on the gel, indicating that the distance between primers had become too long for the un-wanted primer-primer interactions. Regarding the desired MegaPlex product, the strongest band was detected with the second tested primer concentration, where there was no longer competition from the primer-dimer. With further dilutions of the primer, however, the correct product was seen at reduced strength and finally is not visible at all, indicating that the amount of primer was too low and the distance between primers is too long for efficient target amplification. The last sample on the gel, showing only primer-dimer, used the highest tested primer concentration but included no template DNA, thus proving that the primer-dimers are derived from primer sequences alone.

The experiment on bead support (**Figure 5.8 b**) used a lower range of primer concentrations. The first product (in lane 2-4) shows a similar effect as the membrane experiment; at the highest primer concentration the reaction is completely taken over by the primer-dimers, whereas the lower primer concentrations improve the amplification of the desired product. However, the second product (in lanes 5-7) has no primer-dimer problem, and there the target amplification efficiency is reduced when lower primer concentrations are used.

Overall, these experiments show that optimisation of primer concentration on the solid support is important for the performance of MegaPlex PCR. The risk of primer-dimer formation clearly increases with higher primer concentrations, but the efficiency of the target amplification also decreases if the primer concentration is too low. The primer concentration should therefore be low enough to reduce the risks of primer-dimer formation, but not lower since that will reduce the efficiency of the target amplification.

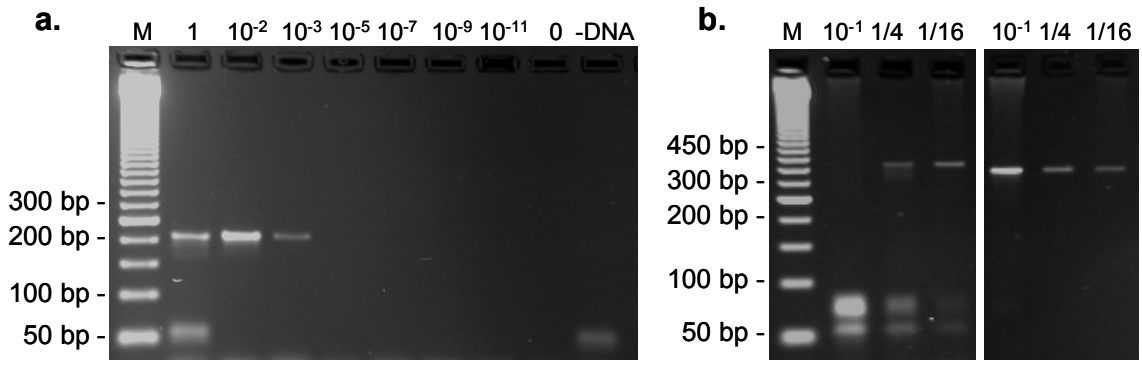


Figure 5.8 The effect of primer concentration on solid support.

MegaPlex amplifications from human genomic DNA was conducted both for a ~200bp target with membrane-bound primers (**a**) and for ~350 bp targets with bead-bound primers (**b**). For both experiments the products were detected on a 3% agarose gel after 30 cycles PCR with a common primer pair.

a) M: 50 bp DNA ladder. The reaction that contained the highest concentration of membrane-bound primers were run in lane 2 (1μM), and decreasing amounts of primer generated products shown in lanes 3-9 (primer concentrations of 10^{-2}, 10^{-3}, 10^{-5}, 10^{-7}, 10^{-9}, 10^{-11}, and 0 μM respectively). In lane 10, where only a primer-dimer band is seen, 1μM primer concentration was used, but without any template DNA.

b) Two MegaPlex products are shown, one in lanes 2-4 and the other in lanes 5-7. M: 50 bp DNA ladder. Bead-bound primer concentrations were 0.1 μM in lanes 2 and 5, and then diluted 1/4 to 25 nM in lanes 3 and 6 and to 6.25 nM in lanes 4 and 7.

5.4.2 DNA TEMPLATE CONCENTRATION

The amount of template DNA was also shown to effect the generation of primer-dimers. An example is shown below, where four different targets were amplified individually as well as combined in a 4-plex reaction. The experiment was performed in three replicates with different DNA concentrations. The effect of the template concentration is presented in **Figure 5.9**. The products obtained with 200 ng genomic DNA were slightly weaker than if more DNA was used, but all products are still visible on the gel, indicating that this amount of DNA was above the detection limit. It is, however, clear that the amount of primer-dimers is more pronounced at the lower target concentration as well as in the control reaction without any DNA. The most likely explanation for this effect would be that primers that do not hybridise to their true target DNA are more likely to interact with neighbouring primers. Additionally, a lower template concentration would simply produce less of the desired product and thus change the ratio between true product and primer-dimer and in that way favour primer-dimer amplification.

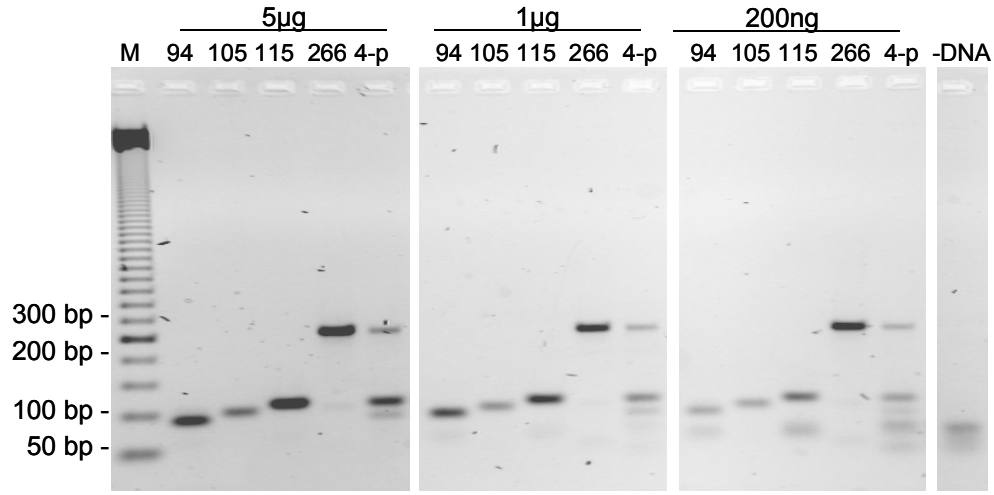


Figure 5.9 Genomic DNA concentration.

MegaPlex PCR products are visualised by gel electrophoresis. M: 50 bp ladder. Five reactions were performed in triplicate, using 5 µg, 1 µg and 200 ng human genomic DNA. The five reactions amplify one 94 bp product, one 105 bp, one 115 bp product, one 266 bp product a 4-plex reaction for the four combined targets. The control reaction without DNA contains primers for the four targets. The amount of input DNA is reflected in the final concentration of the product. More importantly, the amount of short non-specific products increase as the amount of DNA is decreased. The non-specific product is most apparent in the -DNA control.

5.4.3 BARRIER OLIGOS

Another strategy that was tested to reduce the generation of primer-dimers was a concept we called "Barrier Oligos". These oligos are complementary to the common sequence of the MegaPlex primer sequences and will, when hybridising to those, create partly double-stranded surface-bound primers. The reason behind this design is that double-stranded DNA is less flexible than single-stranded DNA, and these "straightened" primers may have less chance to interact with adjacent primers. The Barrier Oligos are included in excess to the surface-bound primers during the annealing reaction, together with the template DNA. To prevent the Barrier Oligos taking part in any extension reaction they carried a phosphate group at the 3' end.

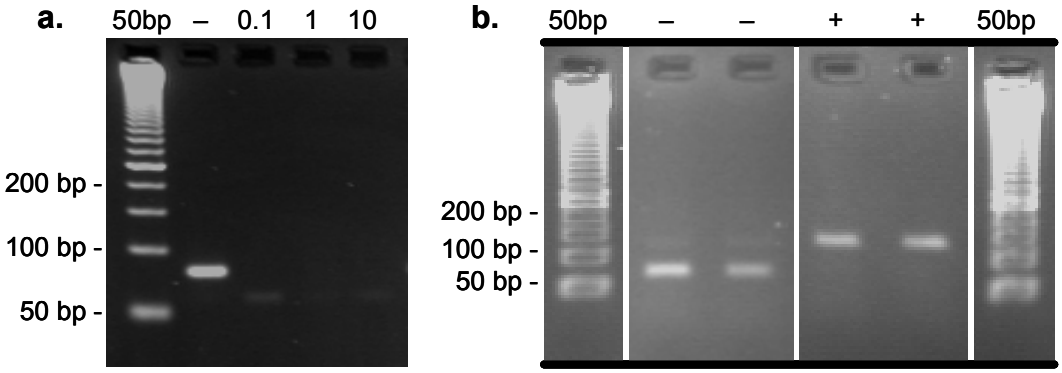


Figure 5.10 Barrier Oligos.

The use of 'Barrier Oligos', complementary to the common primer sequences, can reduce the primer-dimer formation in MegaPlex PCR. This example employs a primer pair for a 140 bp target sequence that easily produces a ~70 bp primer-dimer.

a) Comparison of reactions without template DNA containing no Barrier Oligos (-), 0.1 µM, 1 µM, and 10 µM Barrier Oligos during annealing and extension. Primer-dimers are visualised by gel electrophoresis after 30 cycles PCR amplification with common primers.

b) MegaPlex amplification of the same target from 1 µg human genomic DNA, comparing products produced with (+) and without (-) the use of 0.3 µM Barrier Oligos during annealing and extension. The products are analysed by gel electrophoresis after 30 cycles PCR amplification with common primers.

Since primer-dimers were more easily produced without template DNA present in the reaction, the effect of Barrier Oligos was first tested in this worst case scenario to see if they could have any effect on reducing primer-dimer formation. The results are shown in **Figure 5.10 a**, where even the lowest tested concentration of Barrier Oligos eliminates the strong primer-dimer otherwise produced. To see how this improved the amplification of the desired target sequence, MegaPlex reactions were again set up, now with genomic template DNA. These reactions were set up with bead-bound primers, both with and without the use of Barrier Oligos. When Barrier Oligos were employed, equal amounts of the two oligonucleotides (for the forward- and reverse common primer sequences respectively) were included in the annealing solution, both in the first cycle (with DNA present) and in the second cycle. As shown in **Figure 5.10 b**, the amount of the longer, specific, target amplicon is increased and the amount of the shorter primer-dimer artefact is dramatically decreased, when Barrier Oligos are present in the MegaPlex PCR.

5.4.4 BLOCKING NON-EXTENDED PRIMERS

For the development of MegaPlex PCR streptavidin-biotin interactions were used to bind the primers to the surface, either to membranes or to beads. Although the biotin-streptavidin bond is strong it is not stable enough to withstand the high temperatures necessary for PCR cycling. The reactions were therefore treated with NaOH to remove the target DNA between the two solid-phase PCR cycles, instead of using thermal denaturation. A covalent chemistry could be more practical, but even those have been shown to release primers during PCR cycling[177].

To prevent unwanted primer-primer interactions during the common-primer PCR, the non-extended solid-phase primers would have to be either removed or functionally inactivated, since it would be impossible to prevent detachment of the biotin-labelled primers during this final amplification step. It should be enough to block the primers from further extension after the second solid-phase extension, and this was successfully done by incorporation of ddATP nucleotides by Terminal Deoxynucleotidyl Transferase (TdT).

5.4.5 INDUCIBLE PRIMERS

Another strategy for primer-dimer reduction that was tested was to block one of the primers in a primer pair in the first extension cycle. These primers would be activated only before the second extension reaction. Primers could be blocked for extension by incorporation of a phosphate at the 3' end. The phosphate can later be removed by (Calf Intestine) Alkaline Phosphatase to create a free 3' end for primer extension. However, no advantage could be seen using this strategy, probably due to the fact that the active primer in the first cycle could still create primer-dimers, and the other primer would be free to do so in the second reaction cycle.

5.5 DISCUSSION

In this chapter the basic reaction conditions for MegaPlex PCR, a strategy for multiplex amplification of specific DNA sequences, have been developed and optimised. The strategy behind MegaPlex PCR was to combine solid-phase PCR with normal liquid-phase PCR and take advantage of their respective benefits. The solid-phase reaction is used to capture different targets at different locations and to build in common primer sequences in all selected targets. The common priming sites then make it possible to co-amplify all captured targets with a single primer pair in an efficient solution-phase amplification reaction.

The use of physically separated primer pairs in the solid-phase reaction should eliminate any troublesome cross-reactions between different pairs of primers, and thereby provide a multiplex potential that is, in principle, unlimited. A MegaPlex primer, attached to solid support, consists of a suitable spacer and common primer sequence added directly to the 5' end of a target-specific primer sequences. No special software is therefore required for the primer design, but primers can be chosen relatively freely for each target. The primer location is also not dependent on restriction enzyme sites as for some other multiplex methods[160,164-167].

In the initial design, a single common primer sequence was used, the same for forward- and reverse primers. This would only require a single primer for amplification of any number of products. Using only a single primer could prevent primer-dimer

amplification[162] since single stranded products form a "pan handle" structure when the two ends of the molecule hybridise to each other, which would reduce the priming efficiency of these sequences. It would be excellent to eliminate the primer-dimers, but the amplification efficiency is also reduced for targets up to a few hundred base pairs. To be able to amplify also shorter targets, two separate adaptor sequences, one for each of the two primers in each primer pair, were used in the remainder of the experiments. The new adaptor sequences were also designed with two consecutive parts to allow nested amplification with two different common primer pairs.

The use of a single common primer pair for the amplification of all sequences simultaneously allows the amplification of any number of sequences without increased risk of primer-dimer formation. The amplification bias between different targets is also minimised in this way. However, factors like secondary structure formation could affect the priming efficiency between sequences and shorter products could have an overall advantage over long products. Therefore designing products in the same size range might increase the chances of even amplification of all desired sequences, although little bias with regard to product length was seen in experiments with products between 100 and 500 bp.

The distance between the primer sequence and the surface has previously been shown to be important for efficient primer extension[171,225]. In MegaPlex PCR, a dT spacer was used to move the reaction away from the surface. It was initially assumed that the spacer should be as long as possible to obtain the best result, but by direct comparison of dT spacers of ten and 50 units, the experiments using shorter spacers produced more MegaPlex product than the ones with the longer spacers. This would indicate that longer spacers are not necessary, although the intermediate spacer lengths were not tested. The fact that short spacers work sufficiently well, might also be explained by the common primer sequences that act as an additional spacer, moving the primer annealing reactions further away from the surface for interaction with the template DNA.

Primer-dimer formation is the main obstacle to most multiplex PCR methods. When multiple primer pairs are present in solution, the risk of primer-dimer formation increases, both because of the increased number of possible primer-primer interactions and the increased total primer concentration in the reaction. In MegaPlex PCR the

overall risk of primer-dimer formation is much reduced compared to standard multiplex PCR, since primer-dimers can only be generated by low-level inter-primer annealing and extension within each primer pair during the solid-phase reaction, due to the physical distance between the different primer pairs. More specifically, the primer-dimers that are amplified in the solution-phase PCR must have been generated by cross-priming between the two different primers in a pair, since primer-dimers from a single primer would form the stable pan handle structure that should not easily be amplified.

Several factors that affect the primer-dimer formation in MegaPlex PCR were identified, and the protocols were adjusted to further reduce the risks of primer-dimer formation. One such factor was the solid-phase primer concentration. A high primer concentration means shorter distance between the primers, which would increase the risks of primer-primer annealing on the solid support, which leads to primer-dimer formation. It is also possible that a too high primer concentration could create steric hindrance for template DNA to access the primers and thereby reduce the priming efficiency of the true targets. The primer concentration has to be reduced to a level that restricts these problems. But a lower primer concentration can also reduce the amplification efficiency, both because fewer primers are available to capture the target sequences and since an increased distance between primers can reduce the possibilities of internal solid-phase priming.

The use of 'Barrier Oligos', that create partly double-stranded solid-phase primers, was also shown to reduce the primer-dimer formation. The 'Barrier Oligos' are complementary to the common tail sequences to leave the target-specific part of the primer free to interact with the template DNA. The Barrier Oligos can potentially interfere with primer-dimer formation in several ways; they will block part of the primer sequence for other primer to anneal to, they will lock the primers in a more rigid conformation that might restrict the movement of each primer, and it is also possible that they will even enhance the primer interaction with template DNA by moving the primers further away from the surface.

When titrating the amount of template DNA, it was clear that more non-specific products were amplified when less DNA template was used. Low concentration of template DNA could increase the risks of primer-dimer amplification compared to

higher concentrations in two ways. The amount of desired products would decrease with a lower template concentration, so that the relative amount of primer-dimers would increase. Primer-primer interactions would also be further favoured when fewer primers are hybridised to any target DNA and therefore are free to interact with other primers.

By comparing different targets, it was also apparent that certain primer sequences were more prone to primer-dimer formation than others, even though no apparent 3' complementarity could be detected in these primer pairs. An apparent increase in primer-dimer concentration was also detected in the higher multiplexed reactions. This could partly be explained by the increased risk of including such 'bad' primer pairs when multiple targets are chosen, but also because the total reaction area (number of beads or area on membrane array) per sequence were reduced for each individual sequence compared to single-plex reactions, thus providing fewer primers for each target.

An absolute limit for how long targets can be successfully amplified with MegaPlex PCR and the lowest detection limit for genomic DNA has not been determined. The size range tested, 100 to 500 bp, is however a range that would allow amplification of most exons in the human genome. The upper limit for target size is probably depending on the quality of the template DNA, and the internal solid-phase priming efficiency is probably reduced for long targets due to secondary structures in the solid-phase product.

Overall the results were encouraging for the use of MegaPlex PCR for higher multiplex reactions and for utility in many applications.

CHAPTER 6.

MEGAPLEX PCR – MULTIPLEX APPLICATION

6.1 INTRODUCTION

In the previous chapter, the general strategy for MegaPlex PCR was tested and the reaction conditions were optimised. All in all the results were promising, but in those experiments only single-plex and low-plex reactions were tested. MegaPlex PCR would be attractive to use mainly for higher multiplex reactions, where standard multiplex PCR would be less likely to work reliably. The next step was therefore to demonstrate a higher multiplex application with MegaPlex PCR[228]. It would also be important to determine the relative amplification efficiency of the different targets in the multiplex reactions and determine the reproducibility of the method.

6.2 PRE-AMPLIFICATION OF TEMPLATE DNA

Based on the MegaPlex PCR data obtained so far, the risks of amplifying primer-dimers, in addition to the desired targets, seemed to increase with higher multiplex reactions. This could partly be due to the increased chance of including a primer pair that is prone to primer-dimer formation when more primer pairs are used, and it was shown that the primer-dimer formation can be reduced by careful optimisation of the solid-phase primer concentration and the use of "Barrier Oligos", as described in **Chapter 5.4**. Increasing the DNA template concentration also has a positive effect in decreasing the amounts of primer-dimers formed. It is, however, not desirable to use too much genomic DNA for each experiment, since there is usually a limited resource of DNA samples. One way to increase the template concentration without using too much of the DNA stock would be to start with a pre-amplification step that enriches for the desired sequences. Such a pre-amplification step could potentially increase the efficiency of highly multiplex MegaPlex reactions and reduce the required amount of starting DNA template.

To evaluate this idea in practice, MegaPlex targets were pre-amplified in a standard multiplex PCR, containing primer pairs for all targeted sequences, but without the common-primer sequences used in the solid-phase reaction. This enrichment strategy has previously been used when multiple sequences should be analysed from the single molecule level, such as digital molecular counting applications[229], genome mapping[230,231] and molecular haplotyping[232], to enable subsequent detection of multiple loci and has been used up to 1200-plex applications[233]. This pre-amplification cannot be expected to work perfectly for un-selected complex target sets, since primer-dimers are easily amplified (**Figure 6.1**) when many primers are present in solution, unless extensive primer selection is used[158]. The reaction should nevertheless enrich for the target sequences, whilst any false products or primer-dimers would not be efficiently captured by the solid-phase primers in the downstream MegaPlex reaction.

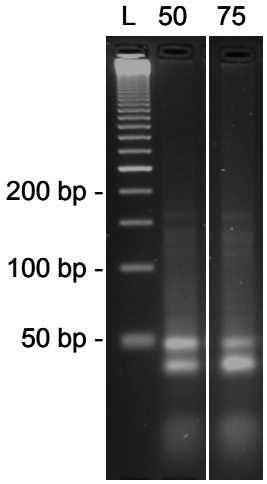


Figure 6.1 Pre-amplified template DNA.

MegaPlex targets are pre-amplified by standard solution-phase multiplex PCR, using primers without the added common primer sequences. L = 50 bp DNA ladder, 50 = 50-plex reaction and 75 = 75-plex reaction. The targeted products are between 47 bp and 219 or 246 bp respectively for the 50-plex and 75-plex reactions. Outside this size range strong primer-dimers are apparent (<45 bp).

6.3 MULTIPLEX GENOTYPING OF MEGAPLEX PCR AMPLIFIED DNA

For a multiplex proof-of-principle experiment of MegaPlex PCR, 75 human genomic targets were chosen. The targets were chosen to each include a validated SNP, which would also allow the demonstration of a multiplex genotyping application. Genotyping assays were also available in the lab for these SNPs, which provided a possibility to validate the genotyping results obtained in the multiplex experiment. This also meant that validated target-specific primers were available, and these sequences could be used for the target-specific part of the MegaPlex primers. These primers were known to successfully amplify their specific target in single-plex reactions, but had never been used in multiplex reactions before. Therefore, the MegaPlex amplification of these 75 targets would be the first attempt to amplify these targets in a single reaction. In addition to the 75-plex reaction (MP75) a subset of 50 targets (MP50) were also amplified by MegaPlex PCR in a separate reaction in parallel, and could be used in order to evaluate the reproducibility of the method.

6.3.1 METHOD OVERVIEW

The experimental setup for the MP50 and MP75 is illustrated schematically in **Figure 6.2**.

The MegaPlex primers were first attached specifically at their 5' ends to beads by biotin-streptavidin interaction (**Figure 6.2 a**). Each primer pair was bound to separate sets of beads (represented by different colours), which were then washed and combined in the desired multiplex combinations. The beads are then attached to the surface of a microtiter plate well (**6.2 b**). This provided an accessible reaction surface for the DNA and enzymes present in the reaction mix.

The target sequences were pre-amplified from genomic DNA template in a multiplex PCR containing primers for all 50 or 75 targets respectively (**6.2 c**), before they were hybridised to the surface-bound primers in the first solid-phase extension reaction (**6.2 d**). The primers were then extended to copy their respective target sequence. Primer-dimer formation within each primer pair was prevented by using an optimal surface primer concentration and by having 'Barrier Oligos' duplexed with the common primer sequence of each surface primer. After this first extension cycle, the template DNA was removed.

In the second solid-phase extension reaction (**6.2 e**), the product strand from the first extension reaction was hybridised to the other primer of the primer-pair, which was then extended to produce a copy of the target sequence with additional common primer regions at both ends. In this way different target fragments were captured on the different beads, and they would all have the same flanking common sequences.

After the two solid-phase extension reactions, all surface-bound molecules were blocked at their 3' ends by incorporation of a dideoxy nucleotide (**6.2 f**). This should prevent the 3' ends of any of the primers or DNA fragments from taking part in any priming events. Finally, common primers were added to the reaction to co-amplify all captured MegaPlex targets in a regular solution-phase PCR (**6.2 g**). The products could then be analysed with a chosen method.

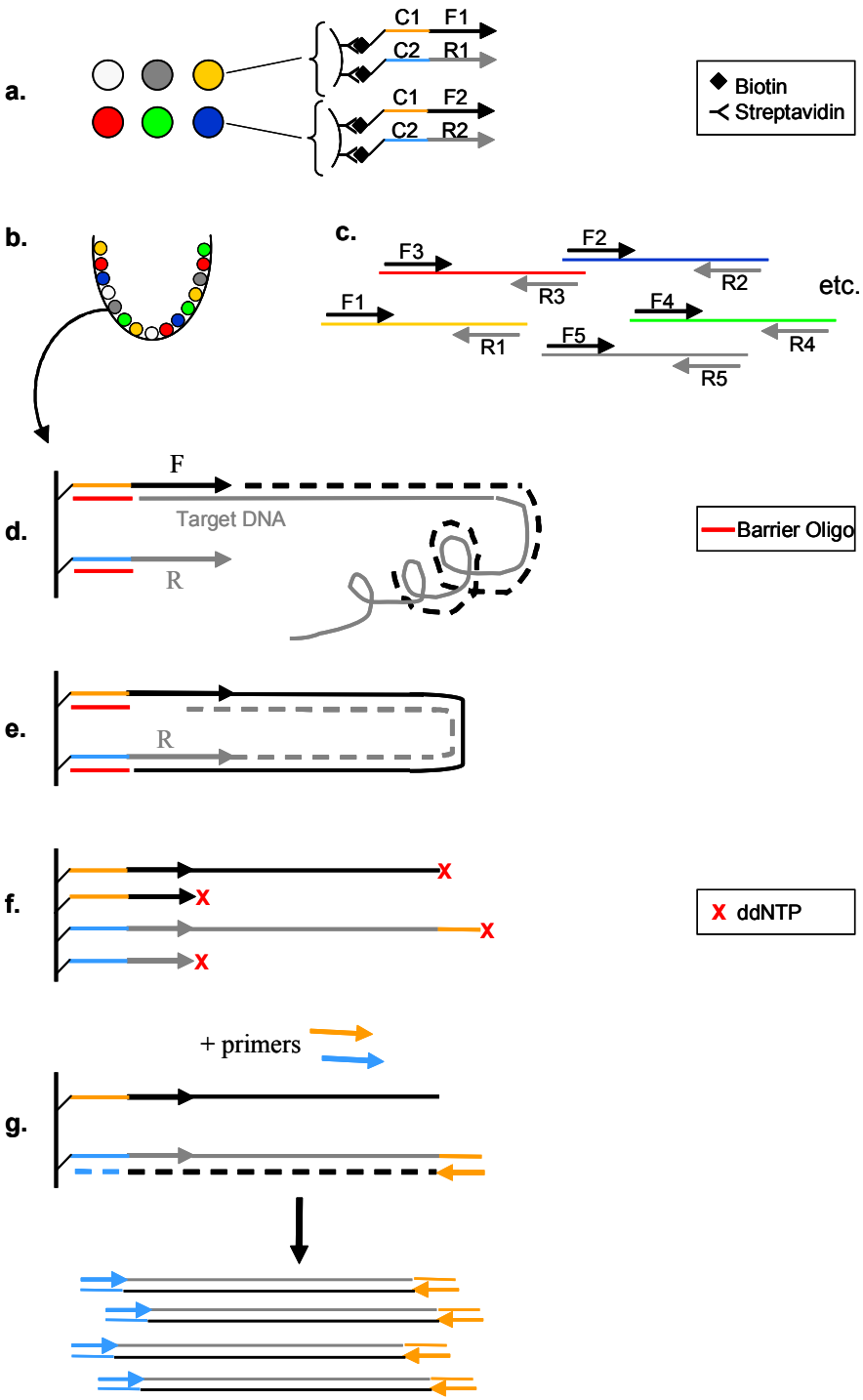


Figure 6.2 Schematic overview of MegaPlex PCR.

The MegaPlex PCR procedure is schematically illustrated, as described in the text.

6.4 ANALYSIS OF MEGAPLEX PCR PRODUCTS

The MegaPlex PCR products were analysed in several different ways, to determine which products were successfully amplified and to estimate their relative recovery rate. Initial analysis was done by agarose gel electrophoresis and then individual targets were detected by microarray hybridisations on the GENIOM platform[184] and by representational sequencing of single molecules on the Roche Genome Sequencer 20 system (GS20)[81]. The target recovery results are presented graphically for each analysis method, and in a tabulated format in **Appendix 3**.

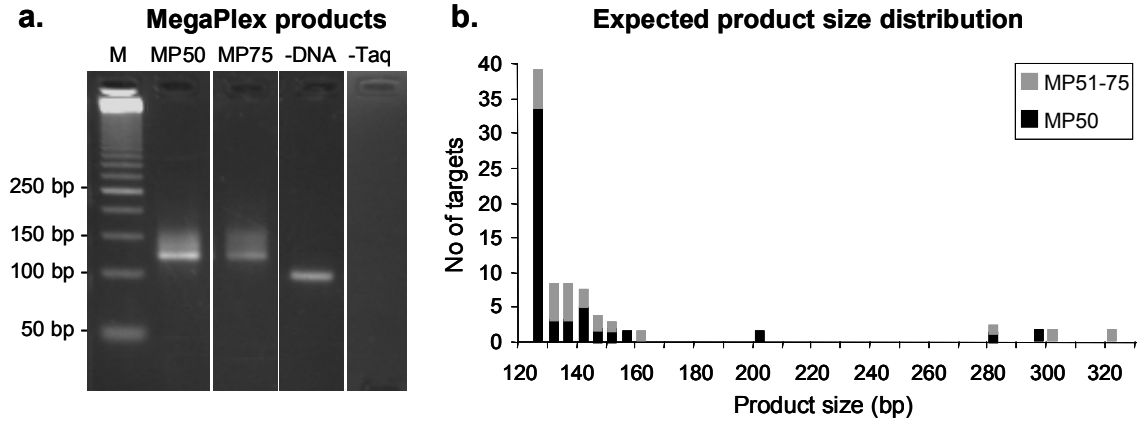


Figure 6.3 Product size distribution.

a) MegaPlex products detected on a 3% agarose gel after 30 cycles PCR amplification with a common primer pair. M: 50 bp size ladder, followed by two MegaPlex products, MP50 and MP75, and two control reactions, the first without template DNA (showing only a short non-specific primer-dimer) and the second without either template DNA or DNA polymerase during the solid-phase reaction.

b) The expected size distribution of the full set of MP50 and MP75 targets. The given sizes include the common primer sequences and are shown in bins of 5 bp.

6.4.1 GEL ELECTROPHORESIS

The MegaPlex products were first analysed by agarose gel electrophoresis. This provided an overview of the size-distribution of the amplified products and could thereby also give information on whether any large amounts of primer-dimers were present. Gel images are shown in **Figure 6.3 a**, and the size distribution can be compared to the expected product sizes in **Figure 6.3 b**. Most targets were between 125 and 130 bp, which is clearly reflected in the gel picture. No products larger than ~160 bp were seen on the gel, but since the products above this size only represent a small fraction of the total number of targets it is possible that they are present at a concentration too low to detect on the gel. Encouragingly, no products smaller than the expected size range were seen in the two MegaPlex products, indicating that no or only negligible amounts of primer-dimers were amplified. The only reaction where primer-dimers were visibly detected by gel electrophoresis was the control reaction without DNA template.

6.4.2 SECONDARY PCRS

To get a better idea of the proportion of sequences that were actually amplified in the MegaPlex PCR, secondary PCRs were run with a diluted MP50 product as template DNA. 21 reactions contained a primer pair targeting a random MegaPlex target while two reactions contained primer pairs targeting sequences that were not present in the MegaPlex set and an additional control contained no primers at all. Of the 21 MegaPlex target-specific reactions 20 amplified products of the expected product length, the only missing product being a 206 bp fragment, while none of the negative control reactions produced any visible band (**Figure 6.4**). The positive results for all but one of the tested MegaPlex targets, in combination with the negative results for the controls, indicate that most products should have been amplified in the MegaPlex reactions. For further analysis of the MegaPlex products, in order to obtain quantitative data for all MegaPlex targets, both the MP50 and MP75 products were analysed by array hybridisation as well as with single molecule counting by a next generation sequencing method.

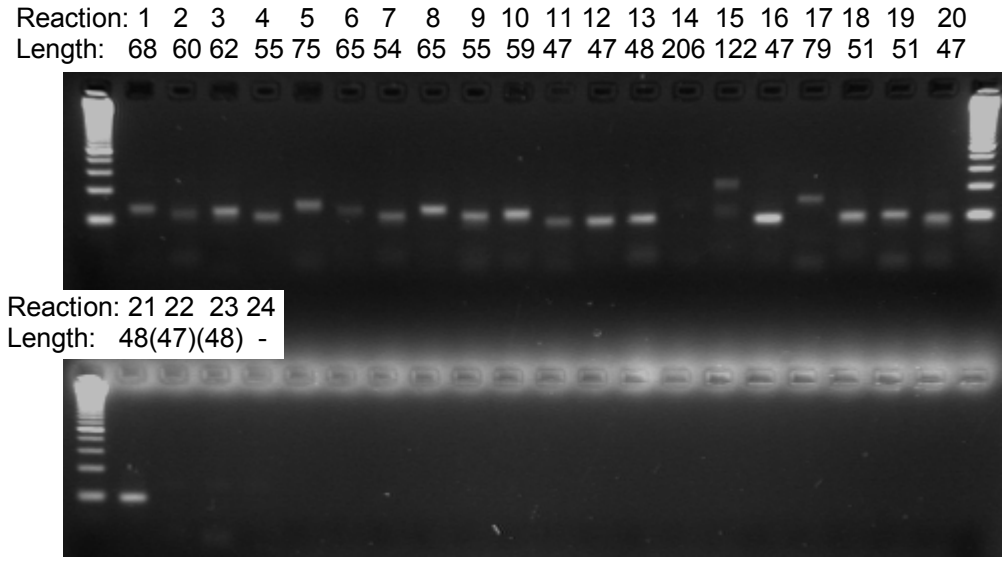


Figure 6.4 Secondary sequence-specific PCRs of MegaPlex products.

Secondary single-plex PCR reactions were used to test a MPX50 product for the presence of a random set of the MegaPlex targets. The products are analysed by agarose gel electrophoresis, where the products are compared to 50 bp DNA ladders. The reaction and the expected product length are indicated above each track of the gel. Reactions 1-21 contained amplification primers for targeted MegaPlex products, reactions 22 and 23 contained primers for human targets not included in the MP50 set and reaction 24 contained no primers. All of the intended targets except one were successfully amplified as clean bands of the expected size, whilst the negative controls did not produce any visible products.

6.4.3 MICROARRAY ANALYSIS

The MegaPlex products were analysed on custom arrays on the GENIOM microarray platform[184]. GENIOM arrays are synthesised *in situ* during an overnight process where eight independent sub-arrays are synthesised on the same chip, each containing 6776 oligonucleotide probes. For the detection of MegaPlex products, several 21 bp probes were designed for each target sequence and for both strands of the product, with four different alleles for each SNP position. Such probe sets were designed for all 75 targets included in the MegaPlex test set as well as for another 42 randomly chosen human single-copy fragments to use as controls. The probes for all targets were synthesised in duplicates on each array.

Hybridisation to oligonucleotide arrays is a semi-quantitative detection method. While the hybridisation intensity to a target-specific oligonucleotide is related to the concentration of a particular target, different targets can hybridise with different efficiency to their respective oligonucleotides. Therefore hybridisation intensity alone could not be used to compare the concentration of the different MegaPlex targets. A more accurate estimation would be possible if the hybridisation intensities of each product could be compared to a reference sample.

The MegaPlex products were labelled with biotin to allow fluorescence detection with streptavidin-phycoerythrin after hybridisation. The labelling was done either by a random priming reaction with biotin-labelled nucleotides or by PCR with 5' biotin-labelled versions of the common primers. A reference sample was prepared containing 90 target sequences, all targeted on the array and including the 75 MegaPlex targets. The 90 targets were amplified individually in separate PCRs and then pooled in equal amounts to one 90-plex control sample (C90). This control was biotin labelled in the same way as the MegaPlex products, and hybridised to identical arrays. Since each GENIOM chip contains eight separate sub-arrays, the control- and MegaPlex samples could be analysed at the same time on the same chip and thereby avoid any run-to-run differences affecting the result.

The hybridisation intensity was compared for the different targets and the different samples. The good correlation of target intensity of the targets shared between the two

MegaPlex experiments indicates good reproducibility of the MegaPlex amplification (**Figure 6.5 a**). The overall spread in signal intensities for the different targets is quite large, not only in the MegaPlex samples but also in the C90 control sample. Part of the variation in hybridisation signal is therefore most likely due to differences in hybridisation efficiency to different probes, and not to difference in concentration between targets. The hybridisation signals for the two MegaPlex reactions were compared to those of the C90 control. The data is presented in **Figures 6.5 b-c** and a reasonable correlation is seen for the majority of the MegaPlex targets. In these graphs are also plotted the signal intensities for the control probe sets, targeting sequences that are not present in either the MegaPlex samples or the C90 control. These signals cluster at low hybridisation intensities, indicative of the levels that can be expected for non-specific hybridisation. Unfortunately, some of the MegaPlex targets were not reliably detected on this array, since their hybridisation signals were as low as the negative control signals (for both the control samples and the MegaPlex samples). MegaPlex targets that produce signals below this background level in the C90 control sample (six targets in MP50 additionally eleven in MP75) were omitted from further analysis, since signals below background levels are not informative. Among the remaining targets, only four gave signals weaker than background in the MegaPlex samples even though they were detectable in the control hybridisation, indicating that they were probably not recovered in the MegaPlex amplification.

For an estimation of the relative recovery of each target, the MegaPlex product signals were divided by their paired C90 signals, after subtraction of the non-specific background. This analysis indicates that the majority of targets that were detectable on the array were recovered at final concentrations that varied little over a ten-fold range, as shown in **Figure 6.6** where the array signal ratios are plotted against the results from the GS sequencing (described below).

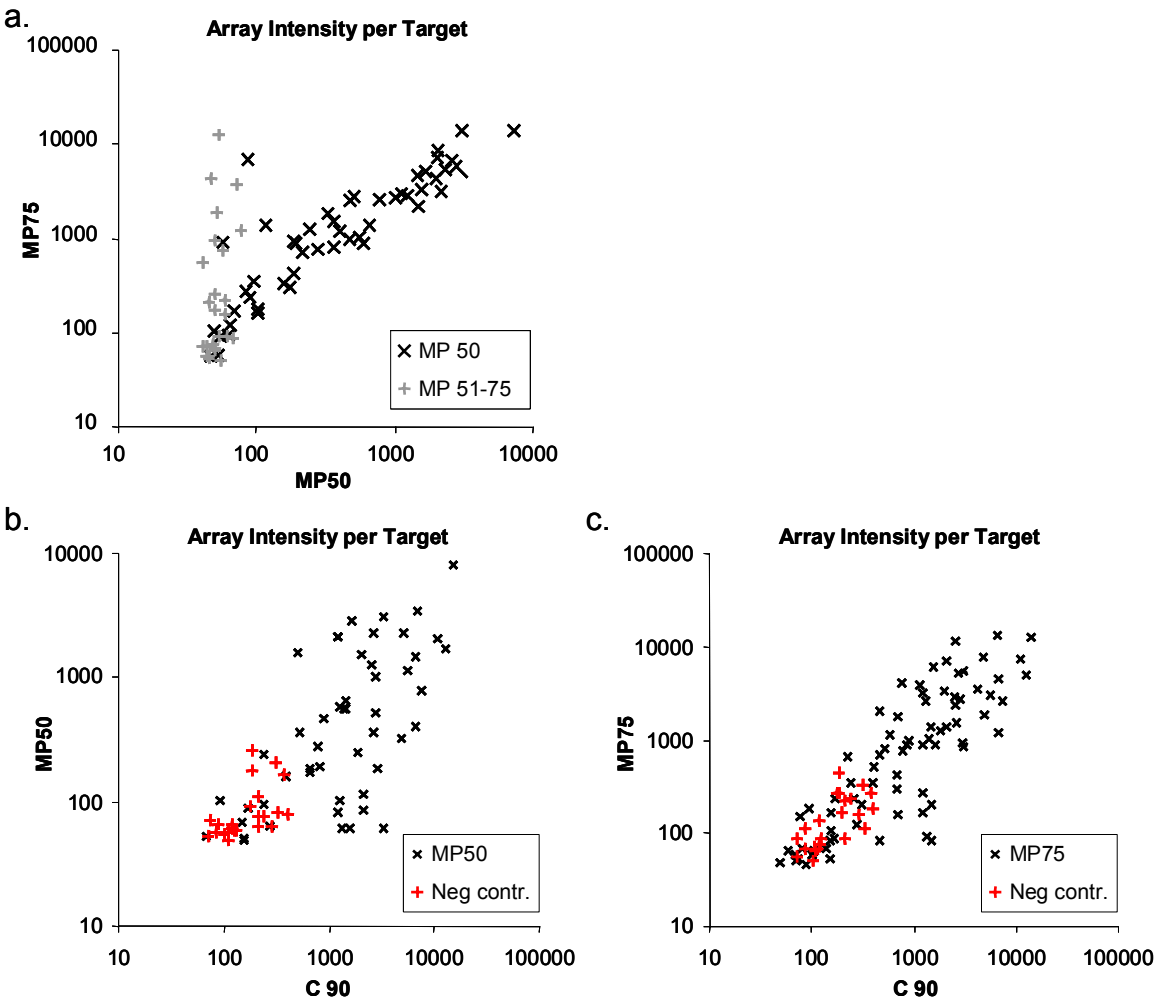


Figure 6.5 Microarray analysis of MegaPlex PCR products.

The average array signal intensity is plotted for each target.

In graph **a)** the MP75 signals are plotted against the data for MP50. The close matching for the 50 targets common to MP50 and MP75 (x) (correlation coefficient 0.90) indicates highly reproducible recovery of these fragments, whereas sequences targeted by only the MP75 experiment (+) gave signals above background only in the MP75 amplification.

In graphs **b)** and **c)** the signals from the MP50 and MP75 samples respectively are plotted against the signals from the C90 control sample. In these graphs the signals are plotted for the targeted sequences in each MegaPlex experiment (x) as well as for negative control probe sets (+) that were not targeted in either the MegaPlex or the control samples.

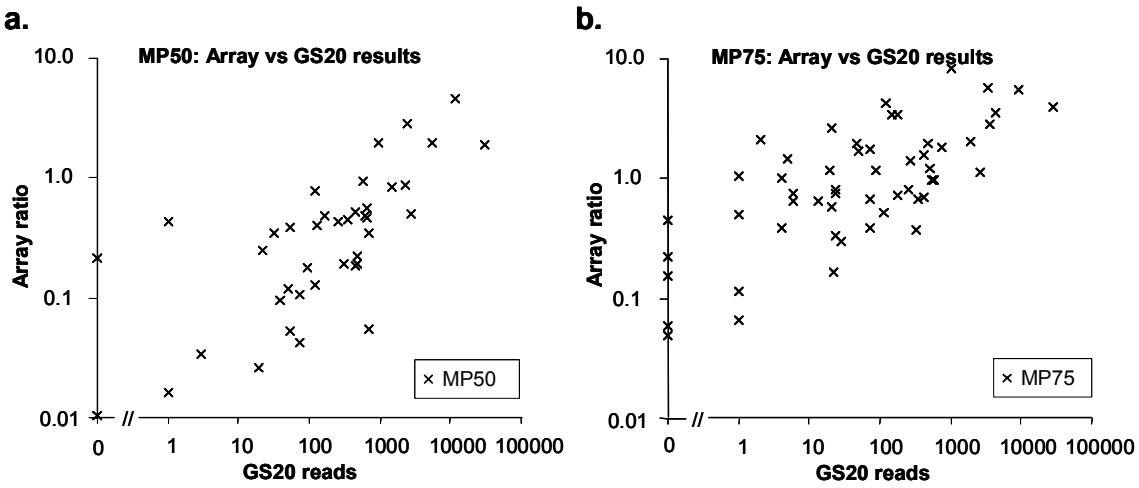


Figure 6.6 Comparison of Microarray results and GS20 results.

The relative recovery rates for the MegaPlex PCR targets estimated by array hybridisation are compared to those of GS20 results. Graph **a)** shows the results for MP50 and **b)** the results for MP75. The recovery rate estimation from the microarray data were taken as the signal ratio between MegaPlex product signals and their paired C90 signals after subtraction of the non-specific background signal. The GS20 data is a simple count of how many times each target was observed in the GS20 sequencing.

6.4.4 GS20 SEQUENCING

The MegaPlex products were further analysed by collaborators at Roche on their Genome Sequencer 20 (GS20) system[81]. With this sequencing technology, developed by 454 Life Sciences, a representation of all molecules in a sample can be individually sequenced. Each DNA molecule can be sequenced separately by capturing and amplifying individual molecules on beads in an emulsion PCR[82] reaction. After this, each bead is placed in a separate picotiter plate well to be sequenced by Pyrosequencing®. By analysing MegaPlex products in this way, it was possible to count the relative abundance of each target to determine the uniformity of the MegaPlex amplification. The detailed sequence information could also be used to determine the genotype for each SNP.

For the GS20 sequencing of the MegaPlex products, new flanking adaptor sequences had to be added to the ends of all products. For this, the MP50 and MP75 products were re-amplified with fusion primers, which contained MegaPlex common sequences at the

3’ end and GS20-specific primer sequences at the 5’ end. Each sample was then split in two aliquots and processed separately in parallel to obtain sequencing data from both directions. The output sequences, 84,483 reads of MP50 and 75,403 reads for MP75, were aligned by BLAST to the predicted MegaPlex target sequence, where each target was represented twice for both alleles of each SNP. The most similar target sequence was identified and counted as a hit only if the sequence similarity was >90%.

The most abundant target was sequenced over 30,000 times, while seven of the MP50 targets and 15 of the MP75 targets were not represented at all, but the majority of targets were detected within a 100-fold range (for tabulated results, see **Appendix 3**). Again, good reproducibility is seen between the two MegaPlex experiments for the targets in common (**Figure 6.7**). The one striking outlier seen is a target that was pre-amplified, but not targeted by the solid-phase reaction in the MP50 reaction. The GS20 counts were also compared to the hybridisation array signal ratio for all targets where data was obtained by both methods (**Figure 6.6**) and the results of the two analysis methods are generally in agreement with each other, thereby cross-validating the two methods. However, the spread is considerably greater in the GS20 analysis, suggesting that the additional amplification with fusion primers might have distorted and exaggerated the recovery distribution compared to their relative levels in the primary MegaPlex product.

Further, the GS20 data also allowed analysis of primer-dimers formed during the MegaPlex amplification, since those would also have been sequenced. Identified primer-dimers represented less than 10% of the total number of sequenced products. Six different primer-dimers were identified, involving a total of nine primers. These primer-dimers were in each case formed between primers from different pairs. Closer inspection of these primer-dimers revealed that most of these primer-dimers were formed due to several bases complementary at the 3’ ends of the primers.

Finally, the GS20 sequencing data enabled genotyping of the SNPs in the input DNA sample. Each sequence read was assigned to either allele of a particular SNP depending on the highest alignment score to the two versions of that target, and in this way the hits for each allele were counted for all SNPs (**Table 6.1**). If only one allele was seen repeatedly the sample was taken to be homozygous for that allele, whereas if many

examples of both alleles were observed it was scored as heterozygous. For the targets present in both experiments (MP50 and MP75) the genotyping results were consistent, and were validated by independent genotyping by DASH[131]. Providing that the total number of alignments exceeded ten, the heterozygous SNPs were correctly identified by the MegaPlex PCR–GS20 sequencing approach. The heterozygous SNPs showed fairly equal recovery of the two alleles, among the twelve detected heterozygous SNPs the median percentage of the least represented allele was 34.8% (with a minimum of 15.3%). Similarly for the homozygous SNPs the median proportion of the correct allele was 99.4% (with a minimum of 95.9%). For both the heterozygous and homozygous SNPs the largest deviation from the ideal result (50% allelic representation in heterozygous SNP and 100% for homozygous SNPs) was seen for targets with relatively low recovery.

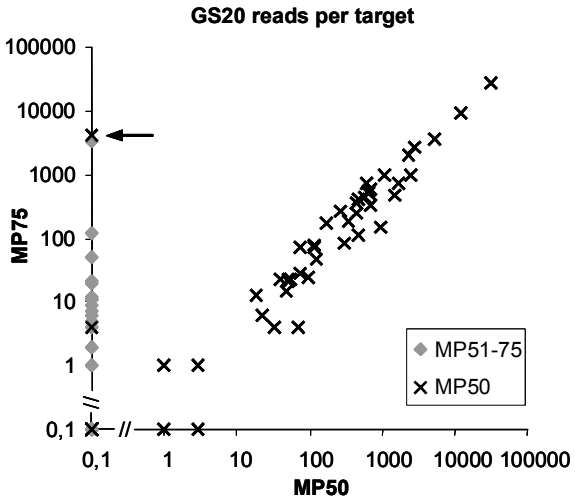


Figure 6.7 GS20 analysis of MegaPlex PCR products.

GS20 analysis of MegaPlex PCR products: Highly-parallel sequencing of individual amplicon molecules from the MegaPlex PCR products enabling direct counting and relative recovery rate determination for each MP50 and MP75 target. The counts per target are plotted for MP75 against those for MP50. The excellent matching for targets common to MP50 and MP75 (×) (correlation coefficient 0.99) indicates highly reproducible recovery of these fragments, whereas sequences targeted by only the MP75 experiment were not recovered at all in the MP50 amplification (♦). One striking outlier MP50 data point can be seen sitting uppermost on the y-axis, indicated with an arrow. This represents one target that was pre-amplified for both MP50 and MP75, but only targeted in the MP75 solid-phase reaction and not in the MP50.

Table 6.1 Genotyping via GS20 sequencing.

Target		MP50, GS20 Counts[1]				MP75, GS20 Counts[1]				MP50+MP75 (%)[2]				DASH[3]
No	SNP	A	C	G	T	A	C	G	T	A	C	G	T	
1	rs4882913[A/T]	0	0	0	1	1	0	0	0					A/T
2	rs6486847[A/G]	686	0	1	0	326	0	0	0	99.9		0.1		A
3	rs10772596[C/T]	0	168	0	0	0	178	0	0		100			C
4	rs10840759[C/T]	0	1	0	437	0	1	0	359		0.3		99.7	T
5	rs2377422[C/T]	0	3	0	0	0	0	0	0					C
6	rs1894824[A/G]	0	0	1	0	0	0	0	0					A/G
7	rs12819884[A/G]	0	0	75	0	0	0	71	0			100		G
8	rs1805721[C/T]	0	53	0	0	0	22	0	1		98.7		1.3	C
9	rs4304840[A/G]	551	0	983	0	162	0	320	0	35.4		64.6		A/G
10	rs1561560[A/G]	49	0	0	0	15	0	0	0	100				A
11	rs1805731[A/G]	9	0	42	0	2	0	19	0	15.3		84.7		A/G
12	rs7300836[C/T]	0	1694	0	10	0	725	0	5		99.4		0.6	ND
13	rs2580874[A/G]	13	0	26	0	14	0	9	0	43.5		56.5		A/G
14	rs919209[C/T]	0	0	0	0	0	0	0	0					ND
15	rs4620776[A/G]	6	0	13	0	6	0	7	0	37.5		62.5		A/G
16	rs7300097[A/G]	0	0	0	0	4	0	0	0					A
17	rs11046892[C/T]	0	0	0	0	0	13	0	4277		0.3		99.7	ND
18	rs2193005[C/G]	0	0	0	0	0	0	0	0					C
19	rs1805673[C/T]	0	32	0	0	0	4	0	0		100			C
20	rs11045985[A/C]	31	38	0	0	1	3	0	0	43.8	56.2			A/C
21	rs7310161[A/T]	686	0	0	3	585	0	0	10	99.0			1.0	A
22	rs1805664[C/T]	0	22	0	0	0	6	0	0		100			C
23	rs2075395[A/G]	117	0	189	0	27	0	59	0	36.7		63.3		A/G
24	rs11046589[C/T]	0	0	0	3	0	0	0	1					C/T
25	rs2231754[A/G]	347	0	0	0	179	0	0	0	100				A
26	rs4883146[C/T]	0	437	0	236	0	307	0	218		62.1		37.9	C/T
27	rs226380[G/T]	4	0	2755	37	3	0	2657	37	0.1		98.5	1.3	G
28	rs11057065[A/G]	3	0	465	0	1	0	416	0	0.5		99.5		G
29	rs1894814[A/G]	0	0	52	0	0	0	22	0			100		G
30	rs4604965[A/G]	117	0	0	0	73	0	0	0	100				A
31	rs10492115[A/G]	659	0	16	0	549	0	4	0	98.4		1.6		A
32	rs7973072[A/G]	19	0	54	0	5	0	24	0	23.5		76.5		A/G
33	rs7307991[C/T]	0	16	1	2345	0	19	0	1922		0.8	0.02	99.2	T
34	rs3026251[A/G]	5440	1	15		3590	1	13		99.7	0.02	0.3		A
35	rs3026252[A/G]	0	0	478	0	0	0	114	0			100		G
36	rs4882965[C/G]	0	620	442	0	0	573	391	0		58.9	41.1		C/G
37	rs7299659[A/G]	5	0	116	0	3	0	70	0	4.1		95.9		G
38	rs226389[C/T]	0	560	0	5	0	425	0	2		99.3		0.7	C
39	rs7134202[G/T]	0	0	261	3	0	0	268	5			98.5	1.5	G
40	rs1805750[A/G]	700	0	1838	0	315	0	692	0	28.6		71.4		A/G
41	rs2024301[A/T]	126		2	0	46	0	0	0	98.9		1.1		A
42	rs4883475[C/T]	2	155	0	11936	0	129	0	9215	0.01	1.3		98.7	T
43	rs2302516[C/G]	22	3	30860	4	25	13	28042	2	0.1	0.03	99.9	0.01	G
44	rs759052[C/T]	0	175	0	272	0	100	0	153		39.3		60.7	C/T
45	rs1133104[G/T]	0	0	613	1	0	0	744	0			99.9	0.1	G
46	rs2110072[C/T]	0	960	0	2	0	147	0	1		99.7		0.3	C
47	rs2071079[C/T]	0	51	0	43	0	8	0	16		50.0		50.0	C/T
48	rs226406[A/T]	0	0	0	0	0	0	0	0					T
49	rs3741854[G/T]	0	0	0	0	0	0	0	0					T
50	rs3809218[A/G]	0	0	0	0	0	0	0	0					A

(1) GS20 counts = number of reads with >90% sequence similarity to each target.

(2) Allele percentage is only calculated for targets where (MP50+MP75) > 10 counts.

(3) Genotype results determined by independent genotyping by DASH.

ND = Not Determined.

6.5 DISCUSSION

In this chapter, MegaPlex PCR was tested in a proof-of-principle experiment to evaluate the potential use of this amplification method for multiplex applications. A 50-plex or 75-plex amplification reaction is not likely to work efficiently in a standard solution-phase multiplex PCR[157,160,161], unless very sophisticated primer design[156-158] is used accompanied by careful optimisation of the reaction conditions[153,155,157,159]. The 75 targets chosen for the MegaPlex experiments were randomly chosen from sequences that had previously been amplified in our laboratory. The target-specific primer sequences had thus previously been validated in single-plex applications, but had never been used for multiplex reactions before.

Most of the target sequences were recovered by MegaPlex PCR, as shown by the different analysis methods. All targets were analysed by microarray hybridisation and GS20 sequencing, and the results were largely in agreement. In the GS20 analysis, seven and 15 expected SNP loci were not observed at all in the sequencing reads for MP50 and MP75 respectively, and these targets either produced hybridisation signals below the non-specific background level, or low signals, in the microarray analysis. Unfortunately some targets were not reliably detectable on the array, since their hybridisation signals were below the background even for the C90 control product. This is why those targets had to be omitted for further analysis.

Pre-amplification of the target DNA can be used as a means to increase the target concentration without sacrificing precious samples, since the MegaPlex amplification works better with higher template DNA concentration. The pre-amplification method used here was a standard solution-phase multiplex PCR. Even though this amplification method is not likely to work equally well for all targets[152-154], and there is the risk of substantial primer-dimer formation[155,156,158] (which is seen for these particular targets in **Figure 6.1**), most targets should be enriched over most other sequences in the genome in this process, and thereby improve the yield and effectiveness of the subsequent MegaPlex PCR steps. Non-specific products or primer-dimers should not be efficiently captured by the solid-phase primers, and for further increased specificity nested primers could be used for the solid-phase reaction. It is, however, likely that this pre-amplification step is one of the main factors causing the concentration differences

between the different targets, and the complete absence of some of the targets, in the final amplification product. For further use and development of MegaPlex PCR it could therefore be necessary to explore other methods for pre-amplification, to obtain a more even distribution of targets.

Importantly, only low levels of primer-dimers were detected in the MegaPlex product. Primer-dimers can otherwise be problematic for multiplex reactions, firstly since they will compete for reagents with the target sequences, but further they can also interfere with the analysis of the targets. The presence of non-specific sequences does not have to be a problem for all analysis methods, for example if internal bases are specifically detected by hybridisation or a separate primer extension reaction. On the other hand, if the application requires sequencing of all targets, for example via the GS20 method, a representation of all molecules present in the amplified sample will be sequenced and the presence of nonsense sequences will reduce the amount of true targets being sequenced.

In MegaPlex PCR, primer-dimers were only expected within each primer pair, since the different primer pairs were separated during the solid-phase reaction. The primer-dimers identified by GS20 sequencing were, nevertheless, formed between primers in different pairs. There are two possibilities of how these primer-dimers could be formed. The first possibility is that primer-primer interactions could take place between different beads during the initial primer extension reactions if some beads were located sufficiently close to allow cross-bead reactions. The other possibility is that some primers were not efficiently blocked at their 3' end and therefore could interact with other primers during the final solution-phase PCR. It is also worth noting that all primers involved in these primer-dimers also took part in amplifying their intended genomic targets, for all of which several representations were sequenced in the GS20 analysis.

Most target sequences in the MP50 and MP75 experiments were short products, the majority between 125 and 150 bp, including the common primer sequences. Only three (in MP50) and six (in MP75) of those products respectively were longer than 200 bp, so there is not enough data to draw any definite conclusions on the importance of target length. However, it seems that longer products are disadvantaged in the procedure. None of the products >200 bp are detected by the GS20 sequencing, and in the array

hybridisation their signals were either below the background level or low (see **Appendix 3**). In the initial development work with low multiplex reactions there were no problems to co-amplify targets over a 5-fold size range, but perhaps the shorter targets have an advantage in the pre-amplification at this higher multiplex level.

SNP genotypes were determined by recording the allelic representation in the GS20 sequencing, and the results were in agreement with the independent genotyping by DASH. For the SNPs validated as homozygous, additional allele(s) were detected at a low level in 19 of the 28 homozygous targets. The alternative SNP allele was always detected in these targets. In five of the targets additional alleles were also detected, but only in one of the targets (rs2302516, number 43) was an additional allele seen at a higher level than the alternative SNP allele. The overall higher prevalence of alternative SNP alleles in homozygous positions would indicate low level contamination during the original sample preparation, but the presence of additional alleles indicate that a fraction of these errors are caused by misincorporation of bases during amplification or sequencing errors.

CHAPTER 7.

DISCUSSION

The work in this thesis was focussed upon technology development, targeting current limitations in genetic analysis for research and diagnostic applications. Specifically, method development was done in two areas; i) DNA variation detection for resequencing applications, and ii) novel approaches to target preparation for multiplex analysis of many targets. Successful results were achieved in both these areas. The proof-of-principle data obtained imply that these methods can be useful in research and diagnostic applications in the future.

7.1 HUMAN GENOME ANALYSIS

Despite recent technology development for complex genome analysis, we are still a long way from having fully characterised the human genome. Ideally one would like to gain knowledge and understanding about all the functional elements at the DNA-, RNA- as well as the protein level, and the interaction between them. Regarding DNA analysis, the efforts to sequence the human genome were only the start of genomics research. The publicly available sequences from the Human Genome Sequencing Consortium[1,3] and Celera[2] have only provided an initial map of the human genome, which is still incomplete. Further efforts are therefore being made to identify genes, regulatory sequences and other functional elements. Genetic variation is also being mapped in the human population, and efforts are being made to determine its implication in health and disease.

7.1.1 METHODS FOR STUDYING GENETIC VARIATION

There are many forms of genetic variation, from single base differences to large insertions, deletions and copy number variable regions. Until now, most studies have focussed on SNPs, leading to extensive knowledge of genome-wide distribution of SNPs in the human populations. Many high-throughput methods for SNP genotyping

have therefore been developed during the last few years. The continuously increasing capacity of many genotyping platforms have made it possible to now analyse over one million positions in the genome in a single experiment with Affymetrix and Illumina platforms. These technological advances have made whole-genome genotyping possible, and the platforms that have demonstrated the highest multiplex levels are using simple assay formats such as probe hybridisation and single base extension reactions, rather than complex reaction schema.

Other forms of genetic variation, such as structural variation and CNVs are likely to be biologically important, since they account for more variant bases in the genome than SNPs[13]. Methods for the analysis of this type of variation are not as developed as for SNP genotyping, and method development in this area will be a major challenge for the future. To an extent, some high-throughput SNP genotyping methods have been modified to detect copy number variable regions, but they are far from ideal for this application. The resolution with which copy number regions are detected depends primarily on the genomic distance between the markers in relation to the size of the CNV. These methods can be used as an alternative or complement to array-CGH, where copy number differences are detected by recording the hybridisation intensity to different part of the genome, represented by oligonucleotides or large clones on a microarray, and then comparing the relative hybridisation intensities between different samples.

In recent years, several new "next generation" sequencing technologies have been developed, which are capable of highly parallel sequencing. These technologies usually have limitations in read length, which are substantially shorter than Sanger sequencing reads, but since so many targets can be sequenced in parallel the cost per sequenced nucleotide is lower. Small genomes, such as bacterial genomes, can be sequenced in a single experiment. Selected targets can also be sequenced for mutation detection purposes. When each sequencing read is derived from a single molecule it also allows detection of low level mutations in heterogeneous samples, such as cancer[88] .

When this thesis work was begun, these high-throughput methods were not available, or at least not fully developed. SNP genotyping had been used for a numbers of years in the hope to identify genetic risk factors for disease by using association studies. There

were an increasing number of different SNP genotyping methods to choose from, but most of them were mainly used for single-plex analysis or at a low multiplex level. Although some genetic risk factors had been identified, many studies had given negative results and new hope was given to whole-genome association studies, which would require highly parallel genotyping assays. Development was going towards miniaturisation, which would be necessary to cut costs, and many assays were being converted to a microarray format.

7.1.2 PROJECT BACKGROUND

As mentioned above, the areas of method development explored in this thesis span sequence variation detection for research and diagnostics, and multiplex amplification of specific targets. These areas were chosen based on previous activities and experience in the research group, and on the expected future needs of the field.

7.1.2.1 DNA DIAGNOSTICS BY DASH

At the start of this thesis work, SNP genotyping and genetic association studies were becoming increasingly popular to study the genetic contribution to complex diseases. With the increased needs for highly parallel analysis, this was clearly an area where further method development was needed. Clearly, as more genetic risk factors would be identified the need for diagnostic tests should also increase. Therefore highly robust methods would be needed to meet the requirements for diagnostic testing. Against this background we reasoned that reliable and flexible sequence variation detection methods, with good capability for multiplex analysis would be attractive for both research and diagnostic applications. Methods for mutation scanning and resequencing would be useful, for example to map genetic variation in genes associated with disease, and methods for parallel mutation detection would be useful for testing multiple risk alleles or causative variants in parallel.

Since our laboratory had extensive experience in DNA hybridisation melt-curve analysis for SNP genotyping using DASH, we felt confident about the robustness of this method and we were convinced that it could be adapted for further applications and in

higher throughput formats. There are several key factors supporting the utility of DASH.

Firstly, the reaction mechanism of DASH is simple; it involves nothing more than hybridising an oligonucleotide probe of known sequence to the target, and determining how well they match by recording a melting profile. A mismatch position will decrease the duplex stability and thereby lead to dissociation at a lower temperature, compared to a fully matched sequence. Since the reaction is only based on DNA hybridisation it is not dependant on enzyme activity and no speciality reagents are needed except the oligonucleotide probes. It should therefore be more cost effective compared to many other methods.

Secondly, melt-curve analysis provides a qualitative measurement of the hybridisation; it does not only measure if the probe hybridises to the target, but also how well it hybridises. In static hybridisation assays this is done by probe-target hybridisation at stringent conditions and comparing the hybridisation efficiency to different (allelic) probes. The problem with this static strategy is that it can be difficult to optimise the hybridisation conditions to obtain allelic discrimination, and for multiplex applications all assays would have to work well under a single reaction condition. Dynamic hybridisation assays are therefore more flexible regarding the probe design, and offer an extra dimension to hybridisation which is bound to improve the reliability of the results to meet the requirements for diagnostic applications.

Thirdly, melt-curve analysis has previously been used for DNA diagnostics. However, this has been done in homogeneous assay formats, by recording a probe-target melting profile during or after PCR amplification. Although homogeneous assays are very practical, both because they require little hands-on time in the laboratory and because they reduce the risk of contamination since a closed tube system is used, they offer limited multiplexing potential.

Fourthly, the DASH reaction takes place with one of the DNA strands attached to solid support, making the system compatible for further development towards microarray implementation.

7.1.2.2 MULTIPLEX DNA AMPLIFICATION

At the start of this thesis work, it was clear that there was an increasing need for high-throughput DNA analysis methods in the field of genetics. Several highly parallel analysis platforms were being developed, to use with a range of different analysis methods. Consequently, effective systems for multiplex target amplification would also be needed, since most DNA analysis methods require amplification of the target sequence prior to analysis. The main DNA amplification method is PCR, which has been an invaluable tool in genetic analysis during the past twenty years. Unfortunately standard PCR protocols do not work very well with several primer pairs in the same reaction and even with extensive optimisation it is difficult to amplify more than 10-20 targets in this way[160,161] due to the increased risk of primer-dimer formation. The short primer-dimers can easily take over the reaction and prevent efficient amplification of the true targets since they compete for the same reagents. If targets have to be amplified individually or in low-plex reactions it would remove some of the benefits with the multiplex analysis and increase both reagent costs and perhaps also the amount of DNA needed. Therefore better alternatives for multiplex amplification of specific target sequences were clearly needed.

While some methods for whole-genome scans can use a combined amplification and complexity reduction step, it is difficult to select exactly which targets are amplified with this strategy. For detailed sequence analysis of a selected genomic region, particular candidate genes or genes in specific pathways other amplification methods are therefore required. Some of the previous attempts to develop multiplex amplification strategies had focused on preventing primer-dimer amplification by using a single common primer for amplification and primer-dimers would form stable hairpin structures that cannot easily be primed[162], but this strategy has never been shown for high-multiplex reactions.

Some innovative methods had been used for multiplex amplification of single base pairs for SNP genotyping, such as the molecular inversion probes[117]. With this strategy the SNPs are captured directly from genomic DNA while introducing common priming sites, thus allowing co-amplification of thousands of targets. This strategy cannot, however, be easily adapted for amplification of longer targets.

Another promising development was solid-phase PCR, where the primers are attached to solid support. In this way, many different targets could be amplified on physically separated locations, as independent, separated, reactions. The main limitation of solid-phase PCR is that the reaction is very inefficient compared to standard PCR, due the restricted mobility of the primers which limits interaction between primers and target.

As no ideal method for the co-amplification of many specific targets yet existed, a new strategy was developed in the frame of this thesis, which we called MegaPlex PCR. This idea involved a combination of solid-phase and liquid-phase amplification. Solid-phase PCR was used initially to capture the different targets independently, and at the same time build in common primer sequences at both ends of all targets. When all targets have the same common priming sites they can be efficiently amplified in a standard solution-phase PCR. Since only one primer pair is used in the solution-phase reaction there is no increased risk of primer-dimer formation compared to single-plex reactions.

The efforts to develop both DASH, for further applications in DNA diagnostics, and MegaPlex PCR, for multiplex amplification of many DNA targets, will be discussed in turn.

7.2 DASH DIAGNOSTICS

7.2.1 SUMMARY OF ACHIEVEMENTS

Instead of using DASH for analysing one SNP in many samples, the method was adapted for mutation scanning, resequencing and simultaneous testing of several mutations, i.e. analysing many positions in the same sample. This change required some adjustment in experimental setup and data analysis.

DASH had previously been taken from a microtiter plate format to a higher density format on membrane macroarrays, where the target sequence from each sample is bound to a distinct feature on the array and then investigated by hybridising an oligonucleotide probe. This format was first adapted so that different copies of the same target could be hybridised with several different probes in parallel without cross

hybridisation of many probes to the same feature, by specifically applying different probes to different areas of the array. Since this strategy would not be practical in a microarray format the DASH concept was then turned "upside-down", hybridising the target to oligonucleotide probes on the array surface.

For mutation scanning, sequence variants are detected in a similar way as in SNP genotyping. Melt-curve analysis of an oligonucleotide probe hybridised to the target sequence will determine if the test sequence is matching the probe or not, based on the melting temperature of the duplex. Any mismatched base in the target will result in a lower Tm than the matching reference sequence. A heterozygous mutation would therefore result in a melting profile with two peaks, while a homozygous mutation would produce a single peak at a lower temperature compared to a reference sample. By using mutation-specific probes, or using multiple probes for each position of the target sequence for resequencing, the identity of the base substitution can be identified.

The idea of using DASH for both mutation scanning and multiple mutation detection was successfully demonstrated in the macroarray format. For mutation scanning, a series of tiled probes were designed for a 97 bp human genomic sequence. The fifteen probes were evenly spaced, so that each position was covered by up to four hybridisation probes. All probes were designed to match the reference sequence and therefore sequence variants would be detected without identifying their exact location or nature. Without prior genotype information of the sixteen test samples, three SNPs were detected and the different samples could easily be assigned to different genotypes.

To assess multiple mutations in parallel, the *gyrA* gene of *Salmonella* was chosen. Mutations in this gene can cause antibiotic resistance, and the presence of eighteen previously identified mutations was tested for in 62 different *Salmonella* strains. Wildtype as well as mutation-specific probes were used to distinguish between the different mutations. Seven of the 18 mutations that we scanned for were detected in the sample set. While two strains were identified as wildtype, the others had up to three different mutations each. All results, both in the mutation scanning of the human test sequence and the *gyrA* mutation detection were confirmed by sequencing of the tested samples, with neither false positive or false negative results.

Work was also initiated to implement DASH in a microarray format, which would allow simultaneous analysis of hundreds or thousands of different positions in the genome and also complete resequencing applications. The first steps were taken to create a resequencing array for the beta-globin gene. Due to technical limitations of the current microarray system, it was not possible to test the melting profile analysis on the microarray instrument. However, a proof-of-principle experiment was performed on the microtiter plate format, and this conclusively detected two thalassaemia mutations.

7.2.2 MACROARRAY DASH

For DASH diagnostics on macroarrays, the experimental layout was similar to that of SNP genotyping in that the amplified target sequence was arrayed out on the membrane. Since there was more than one probe per target, several copies of each target were present on the array and different probes were hybridised to each copy. To prevent probes from hybridising to more than one feature, there had to be some kind of barrier between the features. This could perhaps have been obtained by using a gasket that would isolate each spot on the array during hybridisation. Instead, the different probes were transferred from separate wells in a microtiter plate onto the membrane by centrifugal forces while being clamped together to prevent leakage to surrounding features. The same procedure can also be used to first apply the target to separate locations on the array. With this strategy no cross contamination of the probe was seen to surrounding features. To be extra cautious, the two different PCR products in the *gyrA* experiment were either alternated on the array (since the probe would then not match the target in surrounding features), or otherwise one row was left blank in between features of the same target.

Mutation scanning applications can range from low resolution mutation detection, where it is enough to determine whether a test sample is similar or different to a reference, to complete resequencing, where the identity of every base in the target sequence is assessed. In oligonucleotide hybridisation assays, the resolution of the mutation detection is determined by the probe density and whether only wildtype probes are used or if probes with alternative alleles are also employed. Each base in the target sequence must be covered by at least one probe to be able to detect sequence variants at any position. Higher probe density improves both the precision to identify the location

of the sequence variant and the specificity of the mutation detection since redundant results are obtained if more than one probe covers each variant base and the results of these probes will act as internal controls for each other.

In the mutation scanning of the human target sequence, most bases were covered with three or four probes and each of the three SNPs were therefore detected by multiple adjacent probes, thus validating the result of each other. The use of partly overlapping probes also allowed evaluation of the effect of the mutation position in the probe for mutation detection. In SNP genotyping by DASH, the probes are usually designed to be centred on the variant position, but in a mutation scanning application the variant base could hybridise to any position in the probe sequence. All of the probes that hybridised to the same SNP detected the sequence variation equally well, except when the variant position was at the very last position of the probe. In that case the difference in Tm between the two alleles was very small, and therefore reliable mutation detection would not be possible.

For mutation detection in the *gyrA* gene, the probes were designed so that the mutation position was located towards the centre of the probe, but when many mutations were located close to each other the probe positions were adjusted so that many mutation-specific probes covered the same nucleotide positions. In this way, the same wildtype control probe could be used for all those mutations, and the melting profiles of the different mutation-specific probes could also be directly compared to determine the sample genotype.

Another important factor for data quality was the length of the PCR product. The signal strength would generally decrease for longer products, probably due to reduced accessibility of the probe to its target sequence. Stable secondary structure of the single-stranded target has previously been shown to reduce data quality[146], so the effect of product length is probably related to increased secondary structure of the target and therefore highly sequence dependent. The maximum target length would therefore depend on the sequence context. Different product lengths were also used in the human scanning experiment and the *Salmonella* mutation detection.

For diagnostic applications it would probably be necessary to use reference samples. In mutation scanning applications, the signals would have to be compared to a reference

sample matching the probe sequences, in particular to determine whether a homozygous result (single peak) corresponds to a perfect match or to a mismatch compared to the probe sequence. Alternatively, if many samples are tested in parallel, the signals of all samples can be compared to identify the match- and mismatch peaks. This was also done for the mutation scanning experiment, since no separate control sample was available. It would also be possible to compare the results to previously recorded reference data.

Similarly, for diagnostics of specific mutations the assays would have to be validated with control samples for each mutation and such controls may have to be included in some routine tests. Quite the opposite was done in the gyrA study. This was instead undertaken blind to the mutation status of each sample, and not even a wildtype control was included. It was still possible to determine the mutation status of each strain by comparing the results from wildtype- and mutation-specific probes for each tested site, and by comparing the results of different samples. Unfortunately, in this study only seven of the eighteen tested mutations were present in the sample set. It would have been ideal to be able to validate all mutations, but it was more difficult than expected to get hold of such reference samples.

7.2.3 MICROARRAY DASH

Development of a microarray version of DASH required the general assay design and the target preparation method to be adapted. To be able to conveniently hybridise many different probes to the same target, amplified targets should be hybridised to oligonucleotide probes that are arrayed on the chip, rather than the other way around. This is also the normal array design for static hybridisation microarrays and it would also allow the simultaneous hybridisation of many different targets to the array. There are, however, also drawbacks to this design. When the target (PCR product) was bound to the solid support it was easily made single stranded, by binding only one of the strands to the surface and removing the other. Since it is much more efficient to hybridise single-stranded rather than double-stranded DNA, it was necessary to produce single-stranded template with alternative methods.

To avoid any post-PCR treatment, asymmetric PCR was chosen as a way to produce a relatively high fraction single stranded target. Asymmetric PCR can be difficult to optimise, but by using the primer design rules for LATE-PCR[219], most tested reactions worked very well. Several asymmetric PCRs were performed to amplify parts of the *HBB* gene. These amplifications could be done either directly from genomic DNA, or from a primary PCR product, covering the whole gene (1.9 kb). Although the primary amplification reaction adds an extra step to the analysis it could reduce the amount of DNA needed compared to if several amplifications were done directly from genomic DNA. It will also increase the amplification specificity, particularly in coding parts of the gene where there is high sequence similarity between different globin genes.

The microarray hybridisation results clearly showed the benefit of asymmetric PCR compared to standard PCR, where the hybridisation intensities were much higher for the asymmetric products. Although only static hybridisation was possible, due to technical limitations of the microarray instrument, the experiment still showed the importance of single stranded target for efficient hybridisation to the array. The hybridisation was also highly specific, which could first be estimated by comparing the hybridisation intensities to the sense- and antisense probes. As only the forward strand of the product was labelled (and amplified in excess of the other strand) it should preferably hybridise to the antisense probes, and this was generally the case on the microarray. Another measure of hybridisation specificity is the preferential hybridisation to the reference allele probes compared to the other allele probes. This allelic discrimination was, however, not seen throughout the array. Although the hybridisation conditions were not optimised for such allelic discrimination it shows the difficulties with static hybridisation arrays to obtain allele-specific hybridisation for many probe sets using a single hybridisation condition.

The hybridisation intensity also differed between different regions of the gene. This difference was partly related to the difference in Tm between the probes. The probes were only designed to be 25 base pairs long with the questioned base in the middle, so it is no surprise that the probe Tm should vary due to different C+G content in the probes. Since all probes compete for the same target sequence it is likely that more copies of the target are hybridised to the probes with higher Tm. There could also be other factors

affecting the hybridisation, such as secondary structure in the target or in the probe, which would prevent efficient hybridisation.

Since dynamic detection was not possible on the microarray, the reverse-DASH concept was fully demonstrated on the microtiter plate format. This was done by designing probes for sequencing a total of six base pairs (three adjacent base pairs at two different locations) of the *HBB* gene, covering two mutation sites. Sequencing was done by comparing the melting profiles of the four different probes (A, C, G, and T) at each position, and identifying the probe with the highest Tm. Heterozygous positions were identified when two of the allelic probes show double peaks, with both match and mismatch signals.

7.2.4 FUTURE DEVELOPMENTS

The macroarray results showed that dynamic hybridisation of oligonucleotide probes can be used to reliably detect sequence variation using either only wildtype probes or a combination of wildtype and mutation-specific probes. The macroarray format is however only suitable for low- or medium throughput applications, and the benefit of array-based analysis compared to homogeneous assays is best seen in the miniaturised format, so the development of a microarray implementation of DASH must continue.

For a fully implemented version of DASH on the microarray format a few things still remain to be done. One important issue is the temperature control on the array. The temperature needs to be uniform across all features on the array in order to accurately compare the melting profiles of different probes and to allow run to run comparisons.

Furthermore a hybridisation-specific fluorescence detection system, such as iFRET, is necessary to follow the dissociation of the probe-target duplex in real time. It is important that the fluorescence intensity is strong and highly specific to the probe-target duplex with low background levels, so that reliable melting profiles are obtained.

It will also be necessary with automatic data analysis for efficient sequence interpretation, especially for microarray data where thousands of data points have to be addressed. It will therefore be necessary to develop software for this type of analysis.

If the same accuracy in mutation detection can be obtained in the microarray format as have already been achieved on macroarrays, the melt-curve analysis could offer an extra dimension to static hybridisation microarrays, which would increase the specificity of hybridisation assays. The qualitative measurement of the hybridisation efficiency should facilitate sequence variation detection in any sequence context, which is especially important for diagnostic applications and could potentially allow a more flexible probe design. If a flexible microarray system is used, such as the GENIOM, custom microarrays could easily be designed for any chosen target sequence for mutation scanning, with or without probes for known variants and parallel mutation detection or complete resequencing.

7.2.5 LIMITATIONS

One of the general limitations of DASH, as for all methods using oligonucleotide hybridisation probes, is that it can mainly be used to detect single base pair variants. In a mutation scanning application short insertions and deletions could probably also be detected, since such sequence alterations would also affect the melting profile of the probe-target duplex, but it would not be possible to identify the exact nature of the variant. Theoretically large insertions or deletions could be detected in the same way, but such sequence variants probably result in inefficient amplification of the targets, and then the variation would not be detected anyway. For known insertions and deletions mutation-specific probes could be designed if the exact locations of the break points are known, and if all variants have been efficiently amplified.

The asymmetric amplification strategy for the microarray version of DASH will work best for relatively short targets, since one of the important factors in asymmetric amplification is low Tm difference between primer and product and this is difficult to obtain for long products. This is perhaps not such a big problem, since short products are hybridised more easily to the array. Long products could therefore be fragmented to increase the hybridisation efficiency. This is, however, not an ideal option for melt-curve analysis, since some of the fragmented molecules might hybridise to only part of the probe, and thereby affect the melting profile.

7.3 MEGAPLEX PCR

7.3.1 SUMMARY OF ACHIEVEMENTS

MegaPlex PCR was developed to improve multiplex amplification of specific genomic sequences. The combination of solid-phase target capture with co-amplification of all targets using a common primer pair was demonstrated from genomic DNA template for targets up to 500 bp. A 15-plex reaction amplified targets between 100-500 bp with little bias regarding target length. Higher multiplex reactions used pre-amplified DNA, enriched for the target sequences by a standard multiplex PCR.

In a proof-of-principle experiment, 75 targets containing SNPs were simultaneously amplified in one reaction, and 50 of those targets were amplified in a second reaction. Standard multiplex PCR, with primers for all targets in the same tube, is not likely to work well for randomly selected targets at this level of multiplex due to an increased risk of primer-dimer formation. This was the case for the 50-plex and 75-plex reactions in this study.

The MegaPlex products were analysed by microarray hybridisation and by parallel sequencing-by-synthesis using the GS20 platform. Excellent reproducibility was seen between the two MegaPlex reactions for the targets they had in common. 86% of the targets in the 50-plex reaction and 80% of the targets in the 75-plex reaction were recovered, and the primer-dimer level was less than 10% of the total number of amplified molecules. These recovery rates were determined by the GS20 sequencing by counting the presence of each target. From those sequences correct genotypes could also be extracted for the recovered targets, as validated by independent genotyping with DASH.

Before conducting these multiplex experiments the reaction conditions were optimised in singe-plex and low-plex reactions. The solid-phase reaction primers were attached to the solid support by biotin-streptavidin interactions. The biotin at the 5' end of each primer was followed by a dT spacer, then a common primer sequence and finally the specific primer sequence in the 3' end of the primer. Two types of streptavidin coated solid support were used in the development work; a membrane support where different primer pairs could be arrayed to different locations, and micro-beads where different

primer pairs were bound to beads separately and the beads could then be pooled in different combinations for multiplex reactions.

7.3.2 MEGAPLEX PCR STRATEGY

In the initial solid-phase PCR, different primer pairs are physically separated from each other. This separation prevents interactions between primers in different pairs and each reaction should take place independent from each other which should allow a multiplex potential that is, in principle, unlimited. Once each target has been captured and the common primer sequences have been copied into both ends of the target any number of targets can be amplified in a solution-phase reaction with a single primer pair.

No special primer design rules were applied in these studies. Only general primer design criteria were used, such as avoiding self-complementarity of the individual primers or complementarity between the two primers in each pair, and avoiding stable hairpin structures. In contrast to some other multiplex methods, the primer positions for MegaPlex PCR are not dependant on restriction enzyme sites[160,164-167].

The primers were attached with biotin-streptavidin interactions to membrane arrays or to beads. The attachment of beads to the walls of the microtiter plate wells facilitated the different washes compared to the handling of beads in solution or membranes. Since the biotin-streptavidin bond cannot withstand PCR cycling conditions, the denaturation after the first hybridisation and extension reactions was done by rinsing the membrane/beads in NaOH and neutralising before the second extension reaction. Covalent binding chemistries could potentially allow multiple cycles of solid-phase PCR and require less hands-on time, if denaturation and annealing instead could be controlled by temperature. Perhaps multiple solid-phase cycles could also increase the efficiency of target recovery, but there is a risk that additional amplification cycles would increase the primer-dimer formation as well. It has also previously been shown that a fraction of surface-bound primers are detached from the solid support even when using covalent chemistries[177]. For much higher multiplex levels than was demonstrated in this thesis it would probably be necessary to increase the total surface area to maintain efficient capture of all targets. Other reaction formats, such as microarray surfaces and other types if beads might be worth testing.

7.3.3 PRIMER-DIMER PREVENTION

One of the main objectives when developing MegaPlex PCR was to eliminate the primer-dimer formation, since those can out-compete the amplification of the true targets. While the solid-phase reaction prevents interaction between primers in different pairs, primer-dimers could still be formed within the individual primer pair. An important factor for preventing such internal primer-dimer formation was the concentration of primers on the solid support. Higher primer concentration, and thereby shorter distance between the primers on the surface, increases the chance of primer-primer interactions and it was therefore necessary to optimise the primer concentration so that primer-primer interactions are minimised while the target capture efficiency is maintained at a reasonable level.

The primer-dimer formation was also reduced by using Barrier Oligos. Those are oligonucleotides that are complementary to the common primer sequences. The solid-phase primers will therefore become double-stranded in the 5' end when those oligonucleotides hybridise to their target sequences. The exact mechanism by which the Barrier Oligos prevent primer-dimer formation is not completely established, but by making the primers partly double-stranded they should be more rigid and have restricted movement which should reduce the interaction with other primers. It is also possible that they can even enhance the primer interaction with template DNA by moving the primers further away from the surface.

Since only one primer pair is used in the final solution-phase PCR there should be no increased risk of primer-dimer formation compared to single-plex PCRs. But as the solid-phase primers could easily fall off during temperature cycling, there was still a risk that those primers could form primer-dimers. Since we could not come up with any strategy to specifically remove non-extended primers from the solid support it was decided that they should be functionally inactivated. This was done by blocking the 3' end of all surface-bound DNA molecules by incorporation of a dideoxy nucleotide that cannot be further extended, before the solution-phase reaction.

The final factor that was identified to influence primer-dimer formation was the target concentration. By using a higher concentration of the target primer-dimers could be out-competed. A higher target concentration should increase the efficiency of the specific

hybridisation reactions, leading to a higher concentration of captured targets and an increased target-to-primer ratio after the solid-phase reaction. It is also possible that a higher target concentration could reduce the primer-primer interactions since more primers will be hybridised to their specific target and then have less chance to interact with neighbouring primers. It is however desirable to use as little genomic DNA as possible, since resources are usually limited. To avoid using masses of genomic DNA the concentration of the target sequences could be increased by a pre-amplification step before the MegaPlex PCR.

7.3.4 TARGET PREPARATION

The target pre-amplification in the final MegaPlex study was done by standard solution-phase multiplex PCR. Such reactions would most certainly produce primer-dimers and other non-specific products, but from this mix of sequences the subsequent MegaPlex PCR would only efficiently amplify the specific products, since only those can be efficiently captured in two annealing and extension cycles by the solid-phase primers. This type of amplification followed by a target-selection step has also been used in other multiplex amplification strategies[234].

Nevertheless, it is likely that the different targets are pre-amplified with variable efficiency, thereby leading to the observed spread in final concentration of the different targets. To obtain a more even target distribution in the final product the pre-amplification would have to be further optimised and alternative methods could also be explored. For multiplex mutation detection applications with methods such as the GS20 sequencing it would be important with as even recovery as possible between targets for a cost-effective use of the method. More clones have to be sequenced to obtain a certain sequencing depth if the spread in recovery rates is big compared to if the recovery is even.

As an additional way to save DNA resources, the MegaPlex procedure offers possibilities to recover and reuse DNA for multiple experiments. It should be possible to recover the DNA after the initial extension reaction as the target DNA is in any case removed from the reaction, although this has not been properly demonstrated.

7.3.5 RECENT DEVELOPMENTS, NEW METHODS

Since the start of the MegaPlex PCR project several other groups have also addressed the problem of co-amplification of specific genomic targets.

While many previous methods have focused on the introduction of common primer sites and only use one primer pair in a solution-phase PCR to avoid primer-dimer formation, Wang et al could amplify over 1000 targets for SNP genotyping in a standard multiplex PCR by employing strict primer design rules[158]. Before the genotype detection step they produced single-stranded targets by internal priming reactions, which should improve the specificity. It is difficult to predict an upper limit of how many targets that can be co-amplified with this method, but it seems that the strict primer design rules will restrict the target selection.

Two new amplification strategies have recently been developed based on the specific generation and selection of circular molecules. Multiplex amplification using the selector probe technology[166] and gene-collector probes[234] have demonstrated successful amplification of around 90% of the targeted sequences. Selector probes capture target sequences directly from genomic DNA, while the gene-collectors use an initial solution-phase multiplex PCR. A recent study used the selector probe technology for multiplex amplification of cancer genes for subsequent mutation discovery by parallel sequencing[235]. The selector probe technology is partly dependant on restriction enzyme sites, which could potentially be a limitation to which targets can be amplified in the same reaction.

A strategy that is quite similar to MegaPlex PCR is the multiplex microarray-enhanced PCR[180], since it uses solid-phase primers with common primer sequences at the 5' end and common primers in the solution phase. This strategy is however focused on improving the efficiency of the solid-phase reaction and the solution-phase product contained high concentrations of non-specific primer-dimers even though the demonstrated multiplex level was low.

Compared to these methods MegaPlex PCR amplifies at least as low or lower levels of primer-dimers. It still remains to see how well the MegaPlex amplification will work at higher multiplex levels, and although some of the other methods have demonstrated

higher multiplex reactions than MegaPlex PCR they have not demonstrated levels beyond 1000-plex.

7.3.6 FUTURE DEVELOPMENT

Due to the low levels of primer-dimers observed in MegaPlex PCR it should be possible to significantly increase the multiplex level. The low levels of non-specific product make MegaPlex PCR a good amplification method for parallel sequencing applications, especially where a representation of all molecules in the final product are analysed, since additional products would reduce the sequencing efficiency.

Since the sequencing efficiency is also reduced by a large spread in the recovery rate between different targets it would be necessary to improve this aspect of MegaPlex PCR. The large differences in target recovery rate seen in the MegaPlex experiments were probably largely due to the pre-amplification of the targets, and therefore this step will have to be improved. Either the current method will have to be modified or other methods such as whole genome amplification might be suitable. Since the main reason for the pre-amplification step is to reduce primer-dimer formation the pre-amplification could perhaps be omitted if the primers were better selected to avoid primer-dimer formation. If the pre-amplification is still done by multiplex PCR the specificity of the MegaPlex PCR could perhaps be further increased by using nested primers in the solid-phase reaction.

The limitations of the method, primarily in terms of the multiplex level and the target length will have to be further investigated and, if necessary improved. By performing higher multiplex amplifications it will be easier to identify critical parameters for successfully amplified targets, perhaps by intentionally choosing targets with different properties such as different length, C+G content etc.

For higher multiplex reactions, it would probably be necessary to evaluate alternative reaction surfaces and binding chemistries. Perhaps microarrays could be used where different features contain different primer pairs. An important aspect would be to increase the reaction area between the solid-phase primers and their target DNA to improve their interaction.

7.4 FUTURE PERSPECTIVE

There is still some way to go before we have fully characterised the organisation of all functional elements in the human genome. A significant part of this effort is directed towards mapping and characterising genetic variation throughout the genome. While the majority of the predicted number of SNPs in the human genome have been identified and information of those is stored in public data bases, more efforts are needed to map structural variation, such as copy number variable regions in the genome. Although larger deletions, insertions and copy number variable regions are less common than single base pair variants they affect a larger proportion of the genome[13]. New innovative methods will be needed to study this type of genetic variation in detail.

The HapMap project and the development of highly parallel genotyping assays have made whole-genome association studies possible, and some susceptibility loci for complex diseases have been identified with this strategy. Such loci will have to be further characterised to identify the causative alleles and gain knowledge about their functional effects.

Continued efforts will be made to develop microarray-DASH for mutation scanning and resequencing applications. At the first instance the project to devise a resequencing array for *HBB* will be taken forward. If the results are satisfactory the project might be extended to other globin genes, to be used in research and diagnostics for thalassaemia and other haemoglobin disorders. Arrays could either be designed for detailed resequencing of the whole genes, or combine lower resolution mutation scanning with control probes for known causative variants.

Multiplex amplification methods will continue to be useful to get hold of selected targets for parallel analysis. Unlimited multiplex analysis would be possible if single molecules could be independently detected, and then there would be no need for target amplification, but these technologies are not here yet. MegaPlex PCR will therefore be further developed. It is possible that a continued collaboration with Roche will further improve MegaPlex PCR as a method for amplifying targets for parallel sequencing with the GS20 instrument, for mutation detection in selected genes. MegaPlex PCR could also be used to amplify targets for other type of parallel analysis. It should be possible

to adapt MegaPlex PCR for asymmetric amplification, for example for use in the microarray version of DASH. MegaPlex PCR, as well as other amplification strategies using common primer pairs, could modify the primers to be suitable for asymmetric PCR (by using the LATE-PCR[219] criteria).

Genetic diagnostics for dominant or recessive disorders is important to give the correct diagnosis to affected individuals. For severe genetic disorders with recessive inheritance pattern, genetic screening in risk populations could be an important tool to prevent the disease in future generations. It is difficult to predict to what extent genetic diagnostics will ever be used for complex disease. It will probably depend on the effect of the increased risk and if it is possible to prevent the disease. It is then perhaps more likely to use genetic diagnostics to discriminate between different forms of disease where this information is useful to give the appropriate treatment.

APPENDIX 1.

SAMPLE AND PROBE LAYOUT FOR GYRA GENOTYPING

Salmonella strains

No	Strain	No	Strain	No	Strain	No	Strain	No	Strain	Comment
1	S144577	14	S184117	27	P3801900	40	P479862	53	VLA338	
2	S158112	15	S186737	28	P3907140	41	P490374	54	VLA400	control wt
3	S158777	16	S187874	29	P398351	42	P529039	55	VLA275	control Phe83, Asp144
4	S159797	17	S191076	30	P4103430	43	P532305	56	VLA250	control Asn87
5	S160376	18	S193354	31	P4111690	44	P534461	57	VLA266	control Tyr83
6	S166488	19	S196400	32	P414646	45	P535534	58	VLA263	control Gly87
7	S166492	20	S198849	33	P415310	46	P537713	59	S160376	boiled cells
8	S166495	21	S199046	34	P4156400	47	P5669150	60	S187874	boiled cells
9	S171493	22	S200060	35	P4244791	48	P5685991	61	P3424780	boiled cells
10	S173649	23	P3017450	36	P4281520	49	P5686210	62	P3749380	boiled cells
11	S176476	24	P3424780	37	P4389690	50	VLA153	63	P3801900	boiled cells
12	S180085	25	P3749380	38	P4418140	51	VLA322	64	P4111690	boiled cells
13	S180497	26	P379169	39	P457821	52	VLA263	65	P4156400	boiled cells

Plate/Array 1

Probe:	GyrA1+01P		GyrA5+01P		GyrA1+02P		GyrA5+02P		GyrA2+01P		GyrA5+03P		GyrA2+02P		GyrA5+04P		GyrA3+01P		GyrA6+01P		GyrA3+02P		GyrA6+02P	
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
A	1	16	1	16	1	16	1	16	1	16	1	16	1	16	1	16	1	16	1	16	1	16	1	16
B	2	17	2	17	2	17	2	17	2	17	2	17	2	17	2	17	2	17	2	17	2	17	2	17
C	3	18	3	18	3	18	3	18	3	18	3	18	3	18	3	18	3	18	3	18	3	18	3	18
D	4	19	4	19	4	19	4	19	4	19	4	19	4	19	4	19	4	19	4	19	4	19	4	19
E	5	20	5	20	5	20	5	20	5	20	5	20	5	20	5	20	5	20	5	20	5	20	5	20
F	6	21	6	21	6	21	6	21	6	21	6	21	6	21	6	21	6	21	6	21	6	21	6	21
G	7	22	7	22	7	22	7	22	7	22	7	22	7	22	7	22	7	22	7	22	7	22	7	22
H	8	23	8	23	8	23	8	23	8	23	8	23	8	23	8	23	8	23	8	23	8	23	8	23
I	9	24	9	24	9	24	9	24	9	24	9	24	9	24	9	24	9	24	9	24	9	24	9	24
J	10	25	10	25	10	25	10	25	10	25	10	25	10	25	10	25	10	25	10	25	10	25	10	25
K	11	26	11	26	11	26	11	26	11	26	11	26	11	26	11	26	11	26	11	26	11	26	11	26
L	12	27	12	27	12	27	12	27	12	27	12	27	12	27	12	27	12	27	12	27	12	27	12	27
M	13	28	13	28	13	28	13	28	13	28	13	28	13	28	13	28	13	28	13	28	13	28	13	28
N	14	29	14	29	14	29	14	29	14	29	14	29	14	29	14	29	14	29	14	29	14	29	14	29
O	15	30	15	30	15	30	15	30	15	30	15	30	15	30	15	30	15	30	15	30	15	30	15	30
P	**54**		**54**		**54**		**54**		**54**		**54**		**54**		**54**		**54**		**54**		**54**		**54**	

Plate/Array 2

Probe:	GyrA1+01P		GyrA5+01P		GyrA1+02P		GyrA5+02P		GyrA2+01P		GyrA5+03P		GyrA2+02P		GyrA5+04P		GyrA3+01P		GyrA6+01P		GyrA3+02P		GyrA6+02P	
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
A	31	46	31	46	31	46	31	46	31	46	31	46	31	46	31	46	31	46	31	46	31	46	31	46
B	32	47	32	47	32	47	32	47	32	47	32	47	32	47	32	47	32	47	32	47	32	47	32	47
C	33	48	33	48	33	48	33	48	33	48	33	48	33	48	33	48	33	48	33	48	33	48	33	48
D	34	49	34	49	34	49	34	49	34	49	34	49	34	49	34	49	34	49	34	49	34	49	34	49
E	35	50	35	50	35	50	35	50	35	50	35	50	35	50	35	50	35	50	35	50	35	50	35	50
F	36	51	36	51	36	51	36	51	36	51	36	51	36	51	36	51	36	51	36	51	36	51	36	51
G	37	52	37	52	37	52	37	52	37	52	37	52	37	52	37	52	37	52	37	52	37	52	37	52
H	38	53	38	53	38	53	38	53	38	53	38	53	38	53	38	53	38	53	38	53	38	53	38	53
I	39	59	39	59	39	59	39	59	39	59	39	59	39	59	39	59	39	59	39	59	39	59	39	59
J	40	60	40	60	40	60	40	60	40	60	40	60	40	60	40	60	40	60	40	60	40	60	40	60
K	41	61	41	61	41	61	41	61	41	61	41	61	41	61	41	61	41	61	41	61	41	61	41	61
L	42	62	42	62	42	62	42	62	42	62	42	62	42	62	42	62	42	62	42	62	42	62	42	62
M	43	63	43	63	43	63	43	63	43	63	43	63	43	63	43	63	43	63	43	63	43	63	43	63
N	44	64	44	64	44	64	44	64	44	64	44	64	44	64	44	64	44	64	44	64	44	64	44	64
O	45	65	45	65	45	65	45	65	45	65	45	65	45	65	45	65	45	65	45	65	45	65	45	65
P	**54**		**54**		**54**		**54**		**54**		**54**		**54**		**54**		**54**		**54**		**54**		**54**	

PCR products: LGyrA-04F/LGyrAb03R (in black) and LGyrA-08F/LGyrAb07R (in grey).

Plate/Array 3

Probe:	GyrA3+03P		GyrA7+01P		GyrA3+04P		GyrA7+02P		GyrA3+05P		GyrA8+01P		GyrA3+06P		GyrA8+02P		GyrA3+07P		no probe		GyrA3+08P		no probe	
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
A	1	16	1	16	1	16	1	16	1	16	1	16	1	16	1	16	1	16	1	16	1	16	1	16
B	2	17	2	17	2	17	2	17	2	17	2	17	2	17	2	17	2	17	2	17	2	17	2	17
C	3	18	3	18	3	18	3	18	3	18	3	18	3	18	3	18	3	18	3	18	3	18	3	18
D	4	19	4	19	4	19	4	19	4	19	4	19	4	19	4	19	4	19	4	19	4	19	4	19
E	5	20	5	20	5	20	5	20	5	20	5	20	5	20	5	20	5	20	5	20	5	20	5	20
F	6	21	6	21	6	21	6	21	6	21	6	21	6	21	6	21	6	21	6	21	6	21	6	21
G	7	22	7	22	7	22	7	22	7	22	7	22	7	22	7	22	7	22	7	22	7	22	7	22
H	8	23	8	23	8	23	8	23	8	23	8	23	8	23	8	23	8	23	8	23	8	23	8	23
I	9	24	9	24	9	24	9	24	9	24	9	24	9	24	9	24	9	24	9	24	9	24	9	24
J	10	25	10	25	10	25	10	25	10	25	10	25	10	25	10	25	10	25	10	25	10	25	10	25
K	11	26	11	26	11	26	11	26	11	26	11	26	11	26	11	26	11	26	11	26	11	26	11	26
L	12	27	12	27	12	27	12	27	12	27	12	27	12	27	12	27	12	27	12	27	12	27	12	27
M	13	28	13	28	13	28	13	28	13	28	13	28	13	28	13	28	13	28	13	28	13	28	13	28
N	14	29	14	29	14	29	14	29	14	29	14	29	14	29	14	29	14	29	14	29	14	29	14	29
O	15	30	15	30	15	30	15	30	15	30	15	30	15	30	15	30	15	30	15	30	15	30	15	30
P	**54**		**54**		**54**		**54**		**54**		**54**		**54**	**55**	**54**	**55**	**54**	**57**	**54**		**54**		54	

Plate/Array 4

Probe:	GyrA3+03P		GyrA7+01P		GyrA3+04P		GyrA7+02P		GyrA3+05P		GyrA8+01P		GyrA3+06P		GyrA8+02P		GyrA3+07P		no probe		GyrA3+08P		no probe	
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
A	31	46	31	46	31	46	31	46	31	46	31	46	31	46	31	46	31	46	31	46	31	46	31	46
B	32	47	32	47	32	47	32	47	32	47	32	47	32	47	32	47	32	47	32	47	32	47	32	47
C	33	48	33	48	33	48	33	48	33	48	33	48	33	48	33	48	33	48	33	48	33	48	33	48
D	34	49	34	49	34	49	34	49	34	49	34	49	34	49	34	49	34	49	34	49	34	49	34	49
E	35	50	35	50	35	50	35	50	35	50	35	50	35	50	35	50	35	50	35	50	35	50	35	50
F	36	51	36	51	36	51	36	51	36	51	36	51	36	51	36	51	36	51	36	51	36	51	36	51
G	37	52	37	52	37	52	37	52	37	52	37	52	37	52	37	52	37	52	37	52	37	52	37	52
H	38	53	38	53	38	53	38	53	38	53	38	53	38	53	38	53	38	53	38	53	38	53	38	53
I	39	59	39	59	39	59	39	59	39	59	39	59	39	59	39	59	39	59	39	59	39	59	39	59
J	40	60	40	60	40	60	40	60	40	60	40	60	40	60	40	60	40	60	40	60	40	60	40	60
K	41	61	41	61	41	61	41	61	41	61	41	61	41	61	41	61	41	61	41	61	41	61	41	61
L	42	62	42	62	42	62	42	62	42	62	42	62	42	62	42	62	42	62	42	62	42	62	42	62
M	43	63	43	63	43	63	43	63	43	63	43	63	43	63	43	63	43	63	43	63	43	63	43	63
N	44	64	44	64	44	64	44	64	44	64	44	64	44	64	44	64	44	64	44	64	44	64	44	64
O	45	65	45	65	45	65	45	65	45	65	45	65	45	65	45	65	45	65	45	65	45	65	45	65
P	**54**		**54**		**54**		**54**		**54**		**54**		**54**	**55**	**54**	**55**	**54**	**57**	**54**		**54**		54	

Plate/Array 5

Probe:	GyrA4+01P		No probe	GyrA4+02P		No probe	GyrA4+03P		No probe	GyrA4+04P	
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
A	1	16		1	16		1	16		1	16
B	2	17		2	17		2	17		2	17
C	3	18		3	18		3	18		3	18
D	4	19		4	19		4	19		4	19
E	5	20		5	20		5	20		5	20
F	6	21		6	21		6	21		6	21
G	7	22		7	22		7	22		7	22
H	8	23		8	23		8	23		8	23
I	9	24		9	24		9	24		9	24
J	10	25		10	25		10	25		10	25
K	11	26		11	26		11	26		11	26
L	12	27		12	27		12	27		12	27
M	13	28		13	28		13	28		13	28
N	14	29		14	29		14	29		14	29
O	15	30		15	30		15	30		15	30
P	**54**			**54**	**58**		**54**	**56**		**54**	

Plate/Array 6

	GyrA4+01P		No probe	GyrA4+02P		No probe	GyrA4+03P		No probe	GyrA4+04P	
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
A	31	46		31	46		31	46		31	46
B	32	47		32	47		32	47		32	47
C	33	48		33	48		33	48		33	48
D	34	49		34	49		34	49		34	49
E	35	50		35	50		35	50		35	50
F	36	51		36	51		36	51		36	51
G	37	52		37	52		37	52		37	52
H	38	53		38	53		38	53		38	53
I	39	59		39	59		39	59		39	59
J	40	60		40	60		40	60		40	60
K	41	61		41	61		41	61		41	61
L	42	62		42	62		42	62		42	62
M	43	63		43	63		43	63		43	63
N	44	64		44	64		44	64		44	64
O	45	65		45	65		45	65		45	65
P	**54**			**54**	**58**		**54**	**56**		**54**	

PCR products: LGyrA-04F/LGyrAb03R (in black) and LGyrA-08F/LGyrAb07R (in grey).

APPENDIX 2.

OLIGONUCLEOTIDES USED IN MEGAPLEX PCR

Common oligonucleotides for MegaPlex PCR and analysis

Oligo Name	Modification	Oligo Sequence	Function
MPX:X2:comp01A	3'phosphate	ACTGGCCGTCGTTTTACAACCTGCAGTCTAGAATTCGCTC	barrier oligo
MPX:Y2:comp01B	3'phosphate	GGTCATAGCTGTTTCCTGTGGCAGTCTAGAATTCGGAC	barrier oligo
X727-01A		GAGCGAATTCTAGACTGCAGG	Common primer
Y727-02B		GTCCGAATTCTAGACTGCCAC	Common primer
X727b01A	5'biotin	GAGCGAATTCTAGACTGCAGG	Common primer
Y727b01B	5'biotin	GTCCGAATTCTAGACTGCCAC	Common primer
GS20A_X13-01A		GCCTCCCTCGCGCCATCAGGTTGTAAAACGACGGCCAGT	fusion primer
GS20B_Y13-01B		GCCTTGCCAGCCCGCTCAGCACAGGAAACAGCTATGACC	fusion primer

Oligonucleotides used for MegaPlex PCR development work

Oligo Name	Modification	Oligo Sequence	Target
LSCAN-03F	None	TCTATCTGTCTTACCTCATCACC	primer for 2kb target
LSCAN-06R	None	CTGTCAGTTTTACACAATTCATC	primer for 2kb target
Solid-phase primers			
MPX:dT60_X596-X727_LSCAN:b01A	5'biotin, X=Spacer18	dT(60)XTAGACGGGTCGACACGCGAGCGAATTCT AGACTGCAGGTAACTTACTAGGAGCTTTTAATGG	138 bp
MPX:dT60_Y596-Y727_LSCAN:b04B	5'biotin, X=Spacer18	dT(60)XTACGTTCCGGTAGCACGGTCCGAATTCT AGACTGCCACAATAGAGTGAAATGTATGATTGG	138 bp
MPX:dT60_X596-X727_LSCAN:b01A	5'biotin, X=Spacer18	dT(60)XTAGACGGGTCGACACGCGAGCGAATTCT AGACTGCAGGTAACTTACTAGGAGCTTTTAATGG	240 bp
MPX:dT60_Y596-Y727_LSCAN:b22B	5'biotin, X=Spacer18	dT(60)XTACGTTCCGGTAGCACGGTCCGAATTCT AGACTGCCACATATTAGCCTCATATTTGGGAAG	240 bp
MPX:dT60_X596-X727_LSCAN:b24A	5'biotin, X=Spacer18	dT(60)XTAGACGGGTCGACACGCGAGCGAATTCT AGACTGCAGGTAAAACAAGACTACGTGTCACTG	386 bp
MPX:dT60_Y596-Y727_LSCAN:b04B	5'biotin, X=Spacer18	dT(60)XTACGTTCCGGTAGCACGGTCCGAATTCT AGACTGCCACAATAGAGTGAAATGTATGATTGG	386 bp
MPX:dT60_X596-X727_LSCAN:b24A	5'biotin, X=Spacer18	dT(60)XTAGACGGGTCGACACGCGAGCGAATTCT AGACTGCAGGTAAAACAAGACTACGTGTCACTG	488 bp
MPX:dT60_Y596-Y727_LSCAN:b22B	5'biotin, X=Spacer18	dT(60)XTACGTTCCGGTAGCACGGTCCGAATTCT AGACTGCCACATATTAGCCTCATATTTGGGAAG	488 bp
rs1544396:dT50_X596_X727_b01A	5'biotin, X=Spacer18	dT(50)XTAGACGGGTCGACACGCGAGCGAATTCT AGACTGCAGGAGACTGGCGGGTAGAAAGG	100 bp (1)
rs1544396:dT50_Y596_Y727_b02B	5'biotin, X=Spacer18	dT(50)XTACGTTCCGGTAGCACGGTCCGAATTCT AGACTGCCACTTCCTTCGCACTCTTAGGAT	100 bp (1)
rs2614166:dT50_X596_X727_b01A	5'biotin, X=Spacer18	dT(50)XTAGACGGGTCGACACGCGAGCGAATTCT AGACTGCAGGCATGGTAGTTTAGCTGGTTGAC	100 bp (2)
rs2614166:dT50_Y596_Y727_b02B	5'biotin, X=Spacer18	dT(50)XTACGTTCCGGTAGCACGGTCCGAATTCT AGACTGCCACGCTCTATTTAAAGATGGCAAGA	100 bp (2)
rs627839:dT50_X596_X727_b01A	5'biotin, X=Spacer18	dT(50)XTAGACGGGTCGACACGCGAGCGAATTCT AGACTGCAGGTTCCCAGGGTCTCCTCCA	100 bp (3)
rs627839:dT50_Y596_Y727_b02B	5'biotin, X=Spacer18	dT(50)XTACGTTCCGGTAGCACGGTCCGAATTCT AGACTGCCACGGCAGCTCTTCCGTTCTCT	100 bp (3)
rs3846382:dT50_X596_X727_b01A	5'biotin, X=Spacer18	dT(50)XTAGACGGGTCGACACGCGAGCGAATTCT AGACTGCAGGTTTATCAATTCCTGCTCAAACA	200 bp (1)
rs3846382:dT50_Y596_Y727_b02B	5'biotin, X=Spacer18	dT(50)XTACGTTCCGGTAGCACGGTCCGAATTCT AGACTGCCACCAAATCAGCAATGATGGGATA	200 bp (1)
rs2073454:dT50_X596_X727_b01A	5'biotin, X=Spacer18	dT(50)XTAGACGGGTCGACACGCGAGCGAATTCT AGACTGCAGGAGAGCTGGCCTGATTCACTT	200 bp (2)
rs2073454:dT50_Y596_Y727_b02B	5'biotin, X=Spacer18	dT(50)XTACGTTCCGGTAGCACGGTCCGAATTCT AGACTGCCACACACTCACAAGACCGATTTCC	200 bp (2)
rs1041823:dT50_X596_X727_b01A	5'biotin, X=Spacer18	dT(50)XTAGACGGGTCGACACGCGAGCGAATTCT AGACTGCAGGGCGGCCCTTGTTCAAGAT	200 bp (3)

rs1041823:dT50_Y596_Y727_b02B	5'biotin, X=Spacer18	dT(50)XTACGTTCCGGTAGCACGGTCCGAATTCT AGACTGCCACACGATAGTCATGGGGCTGT	200 bp (3)
rs1468260:dT50_X596_X727_b01A	5'biotin, X=Spacer18	dT(50)XTAGACGGGTCGACACGCGAGCGAATTCT AGACTGCAGGAGTAAGTCCCCAGCAAGGTG	300 bp (1)
rs1468260:dT50_Y596_Y727_b02B	5'biotin, X=Spacer18	dT(50)XTACGTTCCGGTAGCACGGTCCGAATTCT AGACTGCCACGGGGCCAGGTGATATGTTC	300 bp (1)
rs1898606:dT50_X596_X727_b01A	5'biotin, X=Spacer18	dT(50)XTAGACGGGTCGACACGCGAGCGAATTCT AGACTGCAGGGGCAGTGTGTAAGGAAGACAGA	300 bp (2)
rs1898606:dT50_Y596_Y727_b02B	5'biotin, X=Spacer18	dT(50)XTACGTTCCGGTAGCACGGTCCGAATTCT AGACTGCCACAAGGGACTCCGTGTATGAAGAA	300 bp (2)
rs721689:dT50_X596_X727_b01A	5'biotin, X=Spacer18	dT(50)XTAGACGGGTCGACACGCGAGCGAATTCT AGACTGCAGGCACCCTCCCATGAACATTAT	300 bp (3)
rs721689:dT50_Y596_Y727_b02B	5'biotin, X=Spacer18	dT(50)XTACGTTCCGGTAGCACGGTCCGAATTCT AGACTGCCACTGTGTGTTGTGTTGGATGC	300 bp (3)
rs1550538:dT50_X596_X727_b01A	5'biotin, X=Spacer18	dT(50)XTAGACGGGTCGACACGCGAGCGAATTCT AGACTGCAGGGAGCAGGAAGAACGGAATG	400 bp (1)
rs1550538:dT50_Y596_Y727_b02B	5'biotin, X=Spacer18	dT(50)XTACGTTCCGGTAGCACGGTCCGAATTCT AGACTGCCACATTCTCGACGGCTCACAAG	400 bp (1)
rs1019982:dT50_X596_X727_b01A	5'biotin, X=Spacer18	dT(50)XTAGACGGGTCGACACGCGAGCGAATTCT AGACTGCAGGCTGGATTCTGAGGTCACTTCTAC	400 bp (2)
rs1019982:dT50_Y596_Y727_b02B	5'biotin, X=Spacer18	dT(50)XTACGTTCCGGTAGCACGGTCCGAATTCT AGACTGCCACCAACACAAGGCTTTTGCACTT	400 bp (2)
rs2039078:dT50_X596_X727_b01A	5'biotin, X=Spacer18	dT(50)XTAGACGGGTCGACACGCGAGCGAATTCT AGACTGCAGGCCCATCTTAGGTTCTGGGTCT	400 bp (3)
rs2039078:dT50_Y596_Y727_b02B	5'biotin, X=Spacer18	dT(50)XTACGTTCCGGTAGCACGGTCCGAATTCT AGACTGCCACTGAGATATTCCTTTGGGCTCT	400 bp (3)
rs2223114:dT50_X596_X727_b01A	5'biotin, X=Spacer18	dT(50)XTAGACGGGTCGACACGCGAGCGAATTCT AGACTGCAGGTTGGCAATGTTGCAAGAAC	500 bp (1)
rs2223114:dT50_Y596_Y727_b02B	5'biotin, X=Spacer18	dT(50)XTACGTTCCGGTAGCACGGTCCGAATTCT AGACTGCCACCAAGCCTCGTCGTACTAATGA	500 bp (1)
rs1478461:dT50_X596_X727_b01A	5'biotin, X=Spacer18	dT(50)XTAGACGGGTCGACACGCGAGCGAATTCT AGACTGCAGGCCTGCACCAGTTTCCATCC	500 bp (2)
rs1478461:dT50_Y596_Y727_b02B	5'biotin, X=Spacer18	dT(50)XTACGTTCCGGTAGCACGGTCCGAATTCT AGACTGCCACAGTCTGGCAGGTCGGTTCT	500 bp (2)
rs1871113:dT50_X596_X727_b01A	5'biotin, X=Spacer18	dT(50)XTAGACGGGTCGACACGCGAGCGAATTCT AGACTGCAGGGCCATCGATGTTGACTTTAGA	500 bp (3)
rs1871113:dT50_Y596_Y727_b02B	5'biotin, X=Spacer18	dT(50)XTACGTTCCGGTAGCACGGTCCGAATTCT AGACTGCCACAAACAGCTGACAAACATTGGA	500 bp (3)
rs2614166:dT10_X596_X727_b01A	5'biotin	TTTTTTTTTTTAGACGGGTCGACACGCGAGCGAAT TCTAGACTGCAGGCATGGTAGTTTAGCTGGTTGAC	100 bp (2)
rs2614166:dT10_Y596_Y727_b02B	5'biotin	TTTTTTTTTTTACGTTCCGGTAGCACGGTCCGAAT TCTAGACTGCCACGCTCTATTTAAAGATGGCAAGA	100 bp (2)
rs2073454:dT10_X596_X727_b01A	5'biotin	TTTTTTTTTTTAGACGGGTCGACACGCGAGCGAAT TCTAGACTGCAGGAGAGCTGGCCTGATTCACTT	200 bp (2)
rs2073454:dT10_Y596_Y727_b02B	5'biotin	TTTTTTTTTTTACGTTCCGGTAGCACGGTCCGAAT TCTAGACTGCCACACACTCACAAGACCGATTTCC	200 bp (2)
rs1898606:dT10_X596_X727_b01A	5'biotin	TTTTTTTTTTTAGACGGGTCGACACGCGAGCGAAT TCTAGACTGCAGGGGCAGTGTGTAAGGAAGACAGA	300 bp (2)
rs1898606:dT10_Y596_Y727_b02B	5'biotin	TTTTTTTTTTTACGTTCCGGTAGCACGGTCCGAAT TCTAGACTGCCACAAGGGACTCCGTGTATGAAGAA	300 bp (2)

Solid-phase primers for 50-plex and 75-plex MegaPlex PCR

Oligo Name	Modification.	Oligo Sequence	SNP ID	Set
T10:X2:rs4882913:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTGATGAATGCAACATAAGTCTAT	rs4882913	MP50
T10:Y2:rs4882913:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCTGCTTCCACACACAAATGTAAT	rs4882913	MP50
T10:X2:rs6486847:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTAGTATTTTCCTGGTATTAGGGG	rs6486847	MP50
T10:Y2:rs6486847:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCGAACTCTGTTACATGCCTCATT	rs6486847	MP50
T10:X2:rs10772596:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTGTCACAAGCTTATTACATCCTA	rs10772596	MP50
T10:Y2:rs10772596:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCTACTTTAGGGATGAGTGGGAAC	rs10772596	MP50
T10:X2:rs10840759:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTAATTATATGATGTGGTGTCTCC	rs10840759	MP50
T10:Y2:rs10840759:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCAGGGATGAAGAGAAAACCAGAC	rs10840759	MP50
T10:X2:rs2377422:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTAACCCTAGGATATTACTGAGGA	rs2377422	MP50
T10:Y2:rs2377422:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCATTTCACTAAAACCATCCCTAA	rs2377422	MP50

T10:X2:rs1894824:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTTAACAATTTCTGTCCTTCAGAT	rs1894824	MP50
T10:Y2:rs1894824:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCATAGGCATACAACTTTTCTGAG	rs1894824	MP50
T10:X2:rs12819884:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTCTGGTCAAAGAAGATAGAGACT	rs12819884	MP50
T10:Y2:rs12819884:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCAGTCTGGCACAGGTGTCTTCAG	rs12819884	MP50
T10:X2:rs1805721:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTGGGTTAGAGAAATGTGTAACAA	rs1805721	MP50
T10:Y2:rs1805721:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCAGGACATTATGAGGTATTCAAA	rs1805721	MP50
T10:X2:rs4304840:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTGCTGTAGTTTTCATCTTACTTCT	rs4304840	MP50
T10:Y2:rs4304840:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCACAACTTGCAATAAAACAGACA	rs4304840	MP50
T10:X2:rs1561560:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTCCTAAAGTTTAAAGTGCTTTCC	rs1561560	MP50
T10:Y2:rs1561560:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCTAAATGTCTCCCCTCACGTACA	rs1561560	MP50
T10:X2:rs1805731:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTTTTTCAGGGAATGTATCTTAGG	rs1805731	MP50
T10:Y2:rs1805731:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCAACCTCTGCATTTATTGTCAGC	rs1805731	MP50
T10:X2:rs7300836:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTTATCACAGGTCTTTTGGTTTTT	rs7300836	MP50
T10:Y2:rs7300836:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCCAGTTACAGCCATCAGAAAACT	rs7300836	MP50
T10:X2:rs2580874:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTGTGTTAGGAGAGAGGAGATACC	rs2580874	MP50
T10:Y2:rs2580874:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCGGGTTTACATAAACTTCTGCCC	rs2580874	MP50
T10:X2:rs919209:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTGAAGTAGGGAAAACACTGGTTG	rs919209	MP50
T10:Y2:rs919209:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCCACCTGTGGGAAATGAAGGAAA	rs919209	MP50
T10:X2:rs4620776:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTTGTTCCAAAAAGAGATCTATGG	rs4620776	MP50
T10:Y2:rs4620776:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCGTAAAATCTGCAACTCTTCCTT	rs4620776	MP50
T10:X2:rs7300097:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTATCAAACACCATACAAAAACCA	rs7300097	MP50
T10:Y2:rs7300097:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCGTTTGATTCATTCGAGACAACA	rs7300097	MP50
T10:X2:rs7138828:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTGTTCATACAGCGTTTTTAGTCA	rs11046892	MP50
T10:Y2:rs7138828:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCCGAATGAATCAAACAAAATAAAA	rs11046892	MP50
T10:X2:rs2193005:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTAAAGTCATTAGGTGAGCAAAAA	rs2193005	MP50
T10:Y2:rs2193005:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCGCCATATACAATCTTTGAGTAG	rs2193005	MP50
T10:X2:rs1805673:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTAAGATAATTTGGTGATCCAACC	rs1805673	MP50
T10:Y2:rs1805673:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCGAGGAAAGTTTACAGAAACAGT	rs1805673	MP50
T10:X2:rs11045985:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTACGTGAGTACTTTCTCTCCTTT	rs11045985	MP50
T10:Y2:rs11045985:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCTAATTCCATCTGCTTCTTTGAA	rs11045985	MP50
T10:X2:rs7310161:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTCTTTCAACATCCTAGCTCCAAC	rs7310161	MP50
T10:Y2:rs7310161:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCGCAGGGCCACAGCAGGTTAGAC	rs7310161	MP50
T10:X2:rs1805664:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTATGCATTTACCTTCCCAGATGT	rs1805664	MP50
T10:Y2:rs1805664:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCAGAGAACTTCCAGTCTATTTGC	rs1805664	MP50
T10:X2:rs2075395:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTCTTAGCTTCTCACCAAAATGAA	rs2075395	MP50
T10:Y2:rs2075395:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCAGAGTGGAAGACACCGTTGTAA	rs2075395	MP50
T10:X2:rs11046589:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTTTAACAAGAACCATGCCATTTT	rs11046589	MP50
T10:Y2:rs11046589:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCCCCGCTACAGATGAAACAGGTA	rs11046589	MP50

T10:X2:rs2231754:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTTCCATTGCTTCTCCAGATCAAA	rs2231754	MP50
T10:Y2:rs2231754:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCGGTGTAGGAGATATGTACAGTCAAT	rs2231754	MP50
T10:X2:rs4883146:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTAGATGAGCTGCTGATAAGTTCT	rs4883146	MP50
T10:Y2:rs4883146:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCAGCCTAAGTTAGTGTCCCCAAG	rs4883146	MP50
T10:X2:rs226380:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTGGGAGTAGGGTACAATACAGTCT	rs226380	MP50
T10:Y2:rs226380:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCCAGAAAGAAGGAGCTGGAGGAG	rs226380	MP50
T10:X2:rs11057065:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTCTTTCCCACCAAGGTCACAAAA	rs11057065	MP50
T10:Y2:rs11057065:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCGGTTGAATTAGTTGACCCTGAC	rs11057065	MP50
T10:X2:rs1894814:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTCCTCCTTCATTAGATCCATTAATTT	rs1894814	MP50
T10:Y2:rs1894814:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCCAGACAGGCAGTGAGCAGAGGG	rs1894814	MP50
T10:X2:rs4604965:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTTTTTGGTTCTGAACTCATCCAT	rs4604965	MP50
T10:Y2:rs4604965:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCGTAGTAGCAATTTGTTTGGTGA	rs4604965	MP50
T10:X2:rs10492115:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTCATGCATGTCCCACCACAATTA	rs10492115	MP50
T10:Y2:rs10492115:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCTCTTTGATACACGGCAAGGTGC	rs10492115	MP50
T10:X2:rs7973072:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTTTTCCTACTATGCCCAAACCCA	rs7973072	MP50
T10:Y2:rs7973072:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCCCAAAAAGGGTTTCTATTTTAG	rs7973072	MP50
T10:X2:rs7307991:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTAAAGAGAGAGAGAGAATTTTAGCAT	rs7307991	MP50
T10:Y2:rs7307991:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCTAACAAGTCCGTAGGGTTTCCAT	rs7307991	MP50
T10:X2:rs3026251:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTTTACCTTGAGAAGTTTAGAGGTGAT	rs3026251	MP50
T10:Y2:rs3026251:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCAATTCTGGCTGGAGATGGGAAG	rs3026251	MP50
T10:X2:rs3026252:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTGGGTCCAAGTAGGCAAAGAGAC	rs3026252	MP50
T10:Y2:rs3026252:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCACATTTATGGGCTCTGCTCTTA	rs3026252	MP50
T10:X2:rs4882965:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTCTGGAATCACACCTCCCCCATA	rs4882965	MP50
T10:Y2:rs4882965:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCGATGCTTCCAATTTAGAGACAA	rs4882965	MP50
T10:X2:rs7299659:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTTAAGCCACACTGATGCCTTGAC	rs7299659	MP50
T10:Y2:rs7299659:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCGAATAGGCCCCTCTCTCTGGTC	rs7299659	MP50
T10:X2:rs226389:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTCATGGAGGCTTTAGATGGCTCA	rs226389	MP50
T10:Y2:rs226389:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCGTGAGTGTTTCAGAACGATAGA	rs226389	MP50
T10:X2:rs7134202:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTACGGTCTCTTTTGTTCACTGGC	rs7134202	MP50
T10:Y2:rs7134202:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCAAGTCAGATCTTGCTCTATGTG	rs7134202	MP50
T10:X2:rs1805750:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTTCTCAGTTCCCACTAGGCCAAA	rs1805750	MP50
T10:Y2:rs1805750:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCGGGAAGGATATCGCATCTTAAA	rs1805750	MP50
T10:X2:rs2024301:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTCAGCTTCCAAGGAGAAGACTGC	rs2024301	MP50
T10:Y2:rs2024301:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCTGGGGAATCCGGTATTACTTTT	rs2024301	MP50
T10:X2:rs4883475:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTCCGCAGAAATGCTTCTTCCGTT	rs4883475	MP50
T10:Y2:rs4883475:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCTCTCATCTGGGGAAAGACTGTG	rs4883475	MP50
T10:X2:rs2302516:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTGTTGACCCAGAGATCCCCCAGG	rs2302516	MP50
T10:Y2:rs2302516:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCAGACAGCTTGGTGGGGATACAC	rs2302516	MP50

T10:X2:rs759052:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTCAAGTGAGTTGCCCTGACTGCC	rs759052	MP50
T10:Y2:rs759052:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCGGACAAAGGGGTGGATGCCGGG	rs759052	MP50
T10:X2:rs1133104:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTGTACTGGAGGCCCCCATTGTGC	rs1133104	MP50
T10:Y2:rs1133104:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCCCGGATAAAAATTAAGAGAGACTCA	rs1133104	MP50
T10:X2:rs2110072:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTAAGACCTACATCGCCAGCCAGG	rs2110072	MP50
T10:Y2:rs2110072:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCTCATTGACCGTGGCCTCAAGAC	rs2110072	MP50
T10:X2:rs2071079:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTCAGTCCTACTGTGGGTGCCTGC	rs2071079	MP50
T10:Y2:rs2071079:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCGCCCAAGCTAGTCAGGTCAAGG	rs2071079	MP50
T10:X2:rs226406:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTCCATCTCACTCCTAGCTTATTC	rs226406	MP50
T10:Y2:rs226406:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCTCTTTGAGGAGGAAATATGTTG	rs226406	MP50
T10:X2:rs3741854:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTCATCAGCCTCTCTTTAAAATGT	rs3741854	MP50
T10:Y2:rs3741854:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCGTAAGGGGCTTCAAAGGAATGC	rs3741854	MP50
T10:X2:rs3809218:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTCCCCTAGAGAAACGATAGACTG	rs3809218	MP50
T10:Y2:rs3809218:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCTTGAGGTGAGGGATTAAACAAT	rs3809218	MP50
T10:X2:rs7298114:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTAATGAATAACAGTCATGTCTTC	rs7298114	MP75
T10:Y2:rs7298114:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCACTACGAGAAAATCACAAAATA	rs7298114	MP75
T10:X2:rs1805677:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTCTGACAGGTTACAAATTACTAGA	rs1805677	MP75
T10:Y2:rs1805677:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCAGGGAAAATTCTTATTGATATT	rs1805677	MP75
T10:X2:rs1805671:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTGCAGGAAGTATCCCTTAGTTAT	rs1805671	MP75
T10:Y2:rs1805671:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCATTGAGAGTCACTACCAAAATA	rs1805671	MP75
T10:X2:rs10770736:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTATAGAATACACCAATCTTGAGC	rs10770736	MP75
T10:Y2:rs10770736:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCTTAAAATACTGAAATAGACTTTCTT	rs10770736	MP75
T10:X2:rs1805732:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTTTCTCATCTTATGATGTAGCTG	rs1805732	MP75
T10:Y2:rs1805732:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCACACAAAAACATTTTATCCAAA	rs1805732	MP75
T10:X2:rs7954916:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTAAAAATTATTACTAAGTTGACCAG	rs7954916	MP75
T10:Y2:rs7954916:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCGGAATTTTAGTTTCAGGATATT	rs7954916	MP75
T10:X2:rs7976134:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTCGACTGGATAAAATTTCTTAAT	rs7976134	MP75
T10:Y2:rs7976134:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCCACGAATACAATGTTACTTCAA	rs7976134	MP75
T10:X2:rs7307228:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTAAAATTAAAAATGCTATGCTCT	rs7307228	MP75
T10:Y2:rs7307228:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCATTTGTTAGGGAGTGAGTTTTT	rs7307228	MP75
T10:X2:rs6488610:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTAGTAAATCAGACCTCATATAGATT	rs6488610	MP75
T10:Y2:rs6488610:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCATATCCAAATGATTCAAAGCCA	rs6488610	MP75
T10:X2:rs1805651:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTATGTTATGTTAGGCTTGTTTCC	rs1805651	MP75
T10:Y2:rs1805651:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCAGAAAACCTTGGTACATTATTA	rs1805651	MP75
T10:X2:rs17801827:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTTTGATGTTCACAAGAAAACAGA	rs17801827	MP75
T10:Y2:rs17801827:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCTTCCATTTTAGCAGCATTTTAT	rs17801827	MP75
T10:X2:rs6487220:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTCATCAAACAATAAAAATGTGGT	rs6487220	MP75
T10:Y2:rs6487220:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCCCCAAATATTGCAAGCACTTTA	rs6487220	MP75

T10:X2:rs14541:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTTTTCTTAAACCAATCCTTTTGC	rs14541	MP75
T10:Y2:rs14541:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCTAGAGAGTAGGGGTAAAAGCTG	rs14541	MP75
T10:X2:rs11046892:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTAACTACACATCTTGATCAGCTT	rs7138828	MP75
T10:Y2:rs11046892:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCAGAGGAAATAATCCAGGCAAGG	rs7138828	MP75
T10:X2:rs1805740:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTTATTTCTTTGACGTGTGACTCT	rs1805740	MP75
T10:Y2:rs1805740:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCCATAGAAAACTACAGATTGATGT	rs1805740	MP75
T10:X2:rs7307734:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTTTAAGATTAAAACTGGCATCAT	rs7307734	MP75
T10:Y2:rs7307734:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCGGATTGGAGAGAATTGACATTG	rs7307734	MP75
T10:X2:rs11046349:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTTCCTTCAAGTCTTCTCTGTTTT	rs11046349	MP75
T10:Y2:rs11046349:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCTCTGTAAACTCTAAGAAAGTGAG	rs11046349	MP75
T10:X2:rs6488608:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTGGAATCTTTTTGGACATCTGCC	rs6488608	MP75
T10:Y2:rs6488608:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCTCATCTTAATCAACTCAAGCATA	rs6488608	MP75
T10:X2:rs6486850:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTGTCTTGGGTTTTTCTTTACGTT	rs6486850	MP75
T10:Y2:rs6486850:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCCAATAGTGAAAAATCTTAAAAGTAG	rs6486850	MP75
T10:X2:rs5014225:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTCTCTCTCTCTCTTCCTCTGT	rs5014225	MP75
T10:Y2:rs5014225:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCAATAGAAAAGTATTCACTGAAGTT	rs5014225	MP75
T10:X2:rs2889626:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTCTCAAGTCTGAGGGAATCTGCT	rs2889626	MP75
T10:Y2:rs2889626:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCTTAACAAATTCTGTAAGTAAATGGA	rs2889626	MP75
T10:X2:rs7842:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTGGGATAGAAATAGAACTAATGTTT	rs7842	MP75
T10:Y2:rs7842:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCAAAAACTGTCAAAGAATCAATC	rs7842	MP75
T10:X2:rs10840926:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTACTTAACGAATATTTTCAATGTA	rs10840926	MP75
T10:Y2:rs10840926:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCCATAGACTAAAGCAGAAGAGTT	rs10840926	MP75
T10:X2:rs1062836:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTTTTGTCTCATCTTTTAATTTCTTTA	rs1062836	MP75
T10:Y2:rs1062836:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCCAAAGAGGCACACAAGGGTCAT	rs1062836	MP75
T10:X2:rs12812856:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTTGCTATTATTTAAATTTTCCTG	rs12812856	MP75
T10:Y2:rs12812856:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCCTGATGGAATGTTTGCTTAATA	rs12812856	MP75
T10:X2:rs1805659:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTAAAAAAGATTTTTATTTCTCTCT	rs1805659	Contr.
T10:Y2:rs1805659:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCAAATGTTTCTATTGTTTCTTGA	rs1805659	Contr.
T10:X2:rs2889504:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTGAAATCTAAACATGCCATTACT	rs2889504	Contr.
T10:Y2:rs2889504:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCACTGGTTGAGGAGAACAGGCAG	rs2889504	Contr.
T10:X2:rs4573763:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTAATACCTTAGAGAAACAAGAAGA	rs4573763	Contr.
T10:Y2:rs4573763:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCGAGCTCTAATGCAGCTGTTGCT	rs4573763	Contr.
T10:X2:rs1476910:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTAATAGGAAAAGCTAGATGTGAG	rs1476910	Contr.
T10:Y2:rs1476910:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCAGTCAAGGATCGAGGTGAAGTC	rs1476910	Contr.
T10:X2:rs226386:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTACCCTTTACTACTACCCCATTT	rs226386	Contr.
T10:Y2:rs226386:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCAACCCTAAGTTATCTACAAAAAG	rs226386	Contr.
T10:X2:rs3759274:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTCTAGAGTATTATCAGAACCTCCTA	rs3759274	Contr.
T10:Y2:rs3759274:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCTAAGGAAATTAGTGCTTTTAGA	rs3759274	Contr.

T10:X2:rs12833793:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTCATCTCTTTTCTCTTCCTTTGC	rs12833793	Contr.
T10:Y2:rs12833793:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCGACCCCCAGGGGTATCCAGTCT	rs12833793	Contr.
T10:X2:rs3213831:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTTCACTGTCCAATACAAGGTTTG	rs3213831	Contr.
T10:Y2:rs3213831:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCAGAATCCTATTGTATCTGTGGA	rs3213831	Contr.
T10:X2:rs4534636:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTGATATATAGAACAAGTGTGGAAAT	rs4534636	Contr.
T10:Y2:rs4534636:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCGGGTGATGAGATAAAGTAAAACT	rs4534636	Contr.
T10:X2:rs1805761:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTCACCAGTCTAGTAAGTTTGGGA	rs1805761	Contr.
T10:Y2:rs1805761:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCAATGAAAACACACAGACAGCTC	rs1805761	Contr.
T10:X2:rs3782681:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTTCATTGGAAACCTGCTCATTGG	rs3782681	Contr.
T10:Y2:rs3782681:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCGGTCATCTGAGCTGAAGAGTGG	rs3782681	Contr.
T10:X2:rs7302963:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTCAGCTTCTCCTCCATCTGCATT	rs7302963	Contr.
T10:Y2:rs7302963:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCAGGCACAGTTTTCAAAGGAAAT	rs7302963	Contr.
T10:X2:rs2953818:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTCTGTGTTAACCGAACCCCCAGA	rs2953818	Contr.
T10:Y2:rs2953818:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCAGTAGAAGGCCACACTTGAGAT	rs2953818	Contr.
T10:X2:rs2192136:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTTCACCAAGTCTGCCCAGAAAGC	rs2192136	Contr.
T10:Y2:rs2192136:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCGGGATGATATTCATCTAGCCTT	rs2192136	Contr.
T10:X2:rs7294364:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTAAGAATACTTGCCAAGCCCCCG	rs7294364	Contr.
T10:Y2:rs7294364:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCCCATGACTTCTGAGAAAGGTAAAAT	rs7294364	Contr.
T10:X2:rs4619218:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTGGGTCTTGGTCCTGAGCCTGCC	rs4619218	Contr.
T10:Y2:rs4619218:b01B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCCGGGGTTCTGGTGTAGGTGGAT	rs4619218	Contr.
T10:X2:rs2073454:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTAGAGCTGGCCTGATTCACTT	rs2073454	Contr.
T10:Y2:rs2073454:b02B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCACACTCACAAGACCGATTTCC	rs2073454	Contr.
T10:X2:rs1898606:b01A	5'biotin	TTTTTTTTTTGAGCGAATTCTAGACTGCAGGTTGTAAAA CGACGGCCAGTGGCAGTGTGTAAGGAAGACAGA	rs1898606	Contr.
T10:Y2:rs1898606:b02B	5'biotin	TTTTTTTTTTGTCCGAATTCTAGACTGCCACAGGAAACA GCTATGACCAAGGGACTCCGTGTATGAAGAA	rs1898606	Contr.

Primers for pre-amplification and amplification for DASH genotyping

Oligo Name	Modification	Oligo Sequence	SNP ID	Set
rs4882913:b01A	5'biotin	GATGAATGCAACATAAGTCTAT	rs4882913	MP50
rs4882913:-01B		TGCTTCCACACACAAATGTAAT	rs4882913	MP50
rs6486847:b01A	5'biotin	AGTATTTTCCTGGTATTAGGGG	rs6486847	MP50
rs6486847:-01B		GAACTCTGTTACATGCCTCATT	rs6486847	MP50
rs10772596:b01A	5'biotin	GTCACAAGCTTATTACATCCTA	rs10772596	MP50
rs10772596:-01B		TACTTTAGGGATGAGTGGGAAC	rs10772596	MP50
nt8182473:b01A	5'biotin	AATTATATGATGTGGTGTCTCC	rs10840759	MP50
nt8182473:-01B		AGGGATGAAGAGAAAACCAGAC	rs10840759	MP50
rs2377422:b01A	5'biotin	AACCCTAGGATATTACTGAGGA	rs2377422	MP50
rs2377422:-01B		ATTTCACTAAAACCATCCCTAA	rs2377422	MP50
rs1894824:b01A	5'biotin	TAACAATTTCTGTCCTTCAGAT	rs1894824	MP50
rs1894824:-01B		ATAGGCATACAACTTTTCTGAG	rs1894824	MP50
rs12819884:b01A	5'biotin	CTGGTCAAAGAAGATAGAGACT	rs12819884	MP50
rs12819884:-01B		AGTCTGGCACAGGTGTCTTCAG	rs12819884	MP50
rs1805721:b01A	5'biotin	GGGTTAGAGAAATGTGTAACAA	rs1805721	MP50
rs1805721:-01B		AGGACATTATGAGGTATTCAAA	rs1805721	MP50
rs4304840:b01A	5'biotin	GCTGTAGTTTTCATCTTACTTCT	rs4304840	MP50
rs4304840:-01B		ACAACTTGCAATAAAACAGACA	rs4304840	MP50
rs1561560:b01A	5'biotin	CCTAAAGTTTAAAGTGCTTTCC	rs1561560	MP50

rs1561560:-01B		TAAATGTCTCCCCTCACGTACA	rs1561560	MP50
rs1805731:b01A	5'biotin	TTTTCAGGGAATGTATCTTAGG	rs1805731	MP50
rs1805731:-01B		AACCTCTGCATTTATTGTCAGC	rs1805731	MP50
rs7300836:b01A	5'biotin	TATCACAGGTCTTTTGGTTTTT	rs7300836	MP50
rs7300836:-01B		CAGTTACAGCCATCAGAAAACT	rs7300836	MP50
rs2580874:b01A	5'biotin	GTGTTAGGAGAGAGGAGATACC	rs2580874	MP50
rs2580874:-01B		GGGTTTACATAAACTTCTGCCC	rs2580874	MP50
rs919209:b01A	5'biotin	GAAGTAGGGAAAACACTGGTTG	rs919209	MP50
rs919209:-01B		CACCTGTGGGAAATGAAGGAAA	rs919209	MP50
rs4620776:b01A	5'biotin	TGTTCCAAAAAGAGATCTATGG	rs4620776	MP50
rs4620776:-01B		GTAAAATCTGCAACTCTTCCTT	rs4620776	MP50
rs7300097:b01A	5'biotin	ATCAAACACCATACAAAAACCA	rs7300097	MP50
rs7300097:-01B		GTTTGATTCATTCGAGACAACA	rs7300097	MP50
rs11046892:b01A	5'biotin	AACTACACATCTTGATCAGCTT	rs11046892	MP50
rs11046892:-01B		AGAGGAAATAATCCAGGCAAGG	rs11046892	MP50
rs2193005:b01A	5'biotin	AAAGTCATTAGGTGAGCAAAAA	rs2193005	MP50
rs2193005:-01B		GCCATATACAATCTTTGAGTAG	rs2193005	MP50
rs1805673:b01A	5'biotin	AAGATAATTTGGTGATCCAACC	rs1805673	MP50
rs1805673:-01B		GAGGAAAGTTTACAGAAACAGT	rs1805673	MP50
rs11045985:b01A	5'biotin	ACGTGAGTACTTTCTCTCCTTT	rs11045985	MP50
rs11045985:-01B		TAATTCCATCTGCTTCTTTGAA	rs11045985	MP50
rs7310161:b01A	5'biotin	CTTTCAACATCCTAGCTCCAAC	rs7310161	MP50
rs7310161:-01B		GCAGGGCCACAGCAGGTTAGAC	rs7310161	MP50
rs1805664:b01A	5'biotin	ATGCATTTACCTTCCCAGATGT	rs1805664	MP50
rs1805664:-01B		AGAGAACTTCCAGTCTATTTGC	rs1805664	MP50
rs2075395:b01A	5'biotin	CTTAGCTTCTCACCAAAATGAA	rs2075395	MP50
rs2075395:-01B		AGAGTGGAAGACACCGTTGTAA	rs2075395	MP50
rs11046589:b01A	5'biotin	TTAACAAGAACCATGCCATTTT	rs11046589	MP50
rs11046589:-01B		CCCGCTACAGATGAAACAGGTA	rs11046589	MP50
rs2231754:b01A	5'biotin	TCCATTGCTTCTCCAGATCAAA	rs2231754	MP50
rs2231754:-01B		GGTGTAGGAGATATGTACAGTCAAT	rs2231754	MP50
rs4883146:b01A	5'biotin	AGATGAGCTGCTGATAAGTTCT	rs4883146	MP50
rs4883146:-01B		AGCCTAAGTTAGTGTCCCCAAG	rs4883146	MP50
rs226380:b01A	5'biotin	GGGAGTAGGGTACAATACAGTCT	rs226380	MP50
rs226380:-01B		CAGAAAGAAGGAGCTGGAGGAG	rs226380	MP50
rs11057065:b01A	5'biotin	CTTTCCCACCAAGGTCACAAAA	rs11057065	MP50
rs11057065:-01B		GGTTGAATTAGTTGACCCTGAC	rs11057065	MP50
rs1894814:b01A	5'biotin	CCTCCTTCATTAGATCCATTAATTT	rs1894814	MP50
rs1894814:-01B		CAGACAGGCAGTGAGCAGAGGG	rs1894814	MP50
rs4604965:b01A	5'biotin	TTTTGGTTCTGAACTCATCCAT	rs4604965	MP50
rs4604965:-01B		GTAGTAGCAATTTGTTTGGTGA	rs4604965	MP50
rs10492115:b01A	5'biotin	CATGCATGTCCCACCACAATTA	rs10492115	MP50
rs10492115:-01B		TCTTTGATACACGGCAAGGTGC	rs10492115	MP50
rs7973072:b01A	5'biotin	TTTCCTACTATGCCCAAACCCA	rs7973072	MP50
rs7973072:-01B		CCAAAAAGGGTTTCTATTTTAG	rs7973072	MP50
rs7307991:b01A	5'biotin	AAAGAGAGAGAGAGAATTTTAGCAT	rs7307991	MP50
rs7307991:-01B		TAACAAGTCCGTAGGGTTTCCAT	rs7307991	MP50
rs3026251:b01A	5'biotin	TTACCTTGAGAAGTTTAGAGGTGAT	rs3026251	MP50
rs3026251:-01B		AATTCTGGCTGGAGATGGGAAG	rs3026251	MP50
rs3026252:b01A	5'biotin	GGGTCCAAGTAGGCAAAGAGAC	rs3026252	MP50
rs3026252:-01B		ACATTTATGGGCTCTGCTCTTA	rs3026252	MP50
rs4882965:b01A	5'biotin	CTGGAATCACACCTCCCCCATA	rs4882965	MP50
rs4882965:-01B		GATGCTTCCAATTTAGAGACAA	rs4882965	MP50
rs7299659:b01A	5'biotin	TAAGCCACACTGATGCCTTGAC	rs7299659	MP50
rs7299659:-01B		GAATAGGCCCCTCTCTCTGGTC	rs7299659	MP50
rs226389:b01A	5'biotin	CATGGAGGCTTTAGATGGCTCA	rs226389	MP50
rs226389:-01B		GTGAGTGTTTCAGAACGATAGA	rs226389	MP50
rs7134202:b01A	5'biotin	ACGGTCTCTTTTGTTCACTGGC	rs7134202	MP50
rs7134202:-01B		AAGTCAGATCTTGCTCTATGTG	rs7134202	MP50
rs1805750:b01A	5'biotin	TCTCAGTTCCCACTAGGCCAAA	rs1805750	MP50
rs1805750:-01B		GGGAAGGATATCGCATCTTAAA	rs1805750	MP50
rs2024301:b01A	5'biotin	CAGCTTCCAAGGAGAAGACTGC	rs2024301	MP50
rs2024301:-01B		TGGGGAATCCGGTATTACTTTT	rs2024301	MP50
rs4883475:b01A	5'biotin	CCGCAGAAATGCTTCTTCCGTT	rs4883475	MP50
rs4883475:-01B		TCTCATCTGGGGAAAGACTGTG	rs4883475	MP50
rs2302516:b01A	5'biotin	GTTGACCCAGAGATCCCCCAGG	rs2302516	MP50

rs2302516:-01B		AGACAGCTTGGTGGGGATACAC	rs2302516	MP50
rs759052:b01A	5'biotin	CAAGTGAGTTGCCCTGACTGCC	rs759052	MP50
rs759052:-01B		GGACAAAGGGGTGGATGCCGGG	rs759052	MP50
rs1133104:b01A	5'biotin	GTACTGGAGGCCCCCATTGTGC	rs1133104	MP50
rs1133104:-01B		CCGGATAAAAATTAAGAGAGACTCA	rs1133104	MP50
rs2110072:b01A	5'biotin	AAGACCTACATCGCCAGCCAGG	rs2110072	MP50
rs2110072:-01B		TCATTGACCGTGGCCTCAAGAC	rs2110072	MP50
rs2071079:b01A	5'biotin	CAGTCCTACTGTGGGTGCCTGC	rs2071079	MP50
rs2071079:-01B		GCCCAAGCTAGTCAGGTCAAGG	rs2071079	MP50
rs226406:b01A	5'biotin	CCATCTCACTCCTAGCTTATTC	rs226406	MP50
rs226406:-01B		TCTTTGAGGAGGAAATATGTTG	rs226406	MP50
rs3741854:b01A	5'biotin	CATCAGCCTCTCTTTAAAATGT	rs3741854	MP50
rs3741854:-01B		GTAAGGGGCTTCAAAGGAATGC	rs3741854	MP50
rs3809218:b01A	5'biotin	CCCCTAGAGAAACGATAGACTG	rs3809218	MP50
rs3809218:-01B		TTGAGGTGAGGGATTAAACAAT	rs3809218	MP50
rs7298114:b01A	5'biotin	AATGAATAACAGTCATGTCTTC	rs7298114	MP75
rs7298114:-01B		ACTACGAGAAAATCACAAAATA	rs7298114	MP75
rs1805677:b01A	5'biotin	CTGACAGGTTACAAATTACTAGA	rs1805677	MP75
rs1805677:-01B		AGGGAAAATTCTTATTGATATT	rs1805677	MP75
rs1805671:b01A	5'biotin	GCAGGAAGTATCCCTTAGTTAT	rs1805671	MP75
rs1805671:-01B		ATTGAGAGTCACTACCAAAATA	rs1805671	MP75
rs10840926:b01A	5'biotin	ACTTAACGAATATTTTCAATGTA	rs10840926	MP75
rs10840926:-01B		CATAGACTAAAGCAGAAGAGTT	rs10840926	MP75
rs10770736:b01A	5'biotin	ATAGAATACACCAATCTTGAGC	rs10770736	MP75
rs10770736:-01B		TTAAAATACTGAAATAGACTTTCTT	rs10770736	MP75
rs1805732:b01A	5'biotin	TTCTCATCTTATGATGTAGCTG	rs1805732	MP75
rs1805732:-01B		ACACAAAAACATTTTATCCAAA	rs1805732	MP75
rs7954916:b01A	5'biotin	AAAAATTATTACTAAGTTGACCAG	rs7954916	MP75
rs7954916:-01B		GGAATTTTAGTTTCAGGATATT	rs7954916	MP75
rs7976134:b01A	5'biotin	CGACTGGATAAAATTTCTTAAT	rs7976134	MP75
rs7976134:-01B		CACGAATACAATGTTACTTCAA	rs7976134	MP75
rs7307228:b01A	5'biotin	AAAATTAAAAATGCTATGCTCT	rs7307228	MP75
rs7307228:-01B		ATTTGTTAGGGAGTGAGTTTTT	rs7307228	MP75
rs6488610:b01A	5'biotin	AGTAAATCAGACCTCATATAGATT	rs6488610	MP75
rs6488610:-01B		ATATCCAAATGATTCAAAGCCA	rs6488610	MP75
rs1805651:b01A	5'biotin	ATGTTATGTTAGGCTTGTTTCC	rs1805651	MP75
rs1805651:-01B		AGAAAACCTTGGTACATTATTA	rs1805651	MP75
rs17801827:b01A	5'biotin	TTGATGTTCACAAGAAAACAGA	rs17801827	MP75
rs17801827:-01B		TTCCATTTTAGCAGCATTTTAT	rs17801827	MP75
rs6487220:b01A	5'biotin	CATCAAACAATAAAAATGTGGT	rs6487220	MP75
rs6487220:-01B		CCCAAATATTGCAAGCACTTTA	rs6487220	MP75
rs14541:b01A	5'biotin	TTTCTTAAACCAATCCTTTTGC	rs14541	MP75
rs14541:-01B		TAGAGAGTAGGGGTAAAAGCTG	rs14541	MP75
rs7138828:b01A	5'biotin	GTTCATACAGCGTTTTTAGTCA	rs7138828	MP75
rs7138828:-01B		CGAATGAATCAAACAAAATAAAA	rs7138828	MP75
rs1805740:b01A	5'biotin	TATTTCTTTGACGTGTGACTCT	rs1805740	MP75
rs1805740:-01B		CATAGAAAACTACAGATTGATGT	rs1805740	MP75
rs7307734:b01A	5'biotin	TTAAGATTAAAACTGGCATCAT	rs7307734	MP75
rs7307734:-01B		GGATTGGAGAGAATTGACATTG	rs7307734	MP75
rs11046349:b01A	5'biotin	TCCTTCAAGTCTTCTCTGTTTT	rs11046349	MP75
rs11046349:-01B		TCTGTAAACTCTAAGAAAGTGAG	rs11046349	MP75
rs6488608:b01A	5'biotin	GGAATCTTTTTGGACATCTGCC	rs6488608	MP75
rs6488608:-01B		TCATCTTAATCAACTCAAGCATA	rs6488608	MP75
rs6486850:b01A	5'biotin	GTCTTGGGTTTTTCTTTACGTT	rs6486850	MP75
rs6486850:-01B		CAATAGTGAAAAATCTTAAAAGTAG	rs6486850	MP75
rs1062836:b01A	5'biotin	TTTGTCTCATCTTTTAATTTCTTTA	rs1062836	MP75
rs1062836:-01B		CAAAGAGGCACACAAGGGTCAT	rs1062836	MP75
rs5014225:b01A	5'biotin	CTCTCTCTCTCTCTTCCTCTGT	rs5014225	MP75
rs5014225:-01B		AATAGAAAAGTATTCACTGAAGTT	rs5014225	MP75
rs2889626:b01A	5'biotin	CTCAAGTCTGAGGGAATCTGCT	rs2889626	MP75
rs2889626:-01B		TTAACAAATTCTGTAAGTAAATGGA	rs2889626	MP75
rs12812856:b01A	5'biotin	TGCTATTATTTAAATTTTCCTG	rs12812856	MP75
rs12812856:-01B		CTGATGGAATGTTTGCTTAATA	rs12812856	MP75
rs7842:b01A	5'biotin	GGGATAGAAATAGAACTAATGTTT	rs7842	MP75
rs7842:-01B		AAAAACTGTCAAAGAATCAATC	rs7842	MP75

Probes for DASH genotyping

Oligo Name	Modification	Oligo Sequence	SNP ID	Set
KSG:rs4882913:+01P	3'ROX	ACCAAACATCCTGAACA	rs4882913	MP50
AJB:rs6486847:+01P	3'ROX	CCTCATTGCTCCCCTAA	rs6486847	MP50
KSG:rs10772596:+01P	3'ROX	TGGGAACTGTTAGGATG	rs10772596	MP50
KSG:nt8182473:+01P	3'ROX	CCAGACAGGAGGAGACA	rs10840759	MP50
KSG:rs2377422:+01P	3'ROX	TCCCTAAAGTGGATCCT	rs2377422	MP50
KSG:rs1894824:+01P	3'ROX	AAGAGTTTCCATCTGAA	rs1894824	MP50
KSG:rs12819884:+01P	3'ROX	ATTCACCCCTGAGTCTC	rs12819884	MP50
KSG:rs1805721:+01P	3'ROX	GACACCACGCATTGTTA	rs1805721	MP50
KSG:rs4304840:+01P	3'ROX	ACAGACACCGAGAAGTA	rs4304840	MP50
KSG:rs1561560:+01P	3'ROX	GAAGCTTACTGGAAAGC	rs1561560	MP50
KSG:rs1805731:+01P	3'ROX	TCAGCTGTCTCCTAAGA	rs1805731	MP50
KSG:rs7300836:+01P	3'ROX	AAAACTATGGAAAAACC	rs7300836	MP50
KSG:rs2580874:+01P	3'ROX	TCTGCCCACGGGTATCT	rs2580874	MP50
KSG:rs919209:+01P	3'ROX	AAGGAAACGTCAACCAG	rs919209	MP50
KSG:rs4620776:+01P	3'ROX	CTTCCTTTCCTTCCATA	rs4620776	MP50
KSG:rs7300097:+01P	3'ROX	GTAGATAACGTGGTTTT	rs7300097	MP50
KSG:rs11046892:+01P	3'ROX	GGCAAGGAGGCACAAGC	rs11046892	MP50
KSG:rs2193005:+01P	3'ROX	TGAGTAGAGAATTATCT	rs2193005	MP50
KSG:rs1805673:+01P	3'ROX	CAGTTTATGTGGTTGGA	rs1805673	MP50
KSG:rs11045985:+01P	3'ROX	TTTGAAATGGAAAGGAG	rs11045985	MP50
KSG:rs7310161:+01P	3'ROX	GTTAGACCTTGTTGGAG	rs7310161	MP50
KSG:rs1805664:+01P	3'ROX	CTCTAAGGGAACATCTG	rs1805664	MP50
KSG:rs2075395:+01P	3'ROX	CGTTGTAACCATTCATTTT	rs2075395	MP50
KSG:rs11046589:+01P	3'ROX	GATGAAACAGGTAAGTTTT	rs11046589	MP50
KSG:rs2231754:+01P	3'ROX	AATTGAAACCTTTGATC	rs2231754	MP50
KSG:rs4883146:+01P	3'ROX	CCCCAAGTGAAGAACTT	rs4883146	MP50
KSG:rs226380:+01P	3'ROX	GGAGGAGACCAGACTGT	rs226380	MP50
KSG:rs11057065:+01P	3'ROX	CCCTGACACGTTTTGTG	rs11057065	MP50
KSG:rs1894814:+01P	3'ROX	CAGAGGGGCATAAACAG	rs1894814	MP50
AJB:rs4604965:+01P	3'ROX	GTACCACACGCATGGAT	rs4604965	MP50
KSG:rs10492115:+01P	3'ROX	AAGGTGCTCATTAATTG	rs10492115	MP50
AJB:rs7973072:+01P	3'ROX	ATTTTAGACTTGGGTTT	rs7973072	MP50
KSG:rs7307991:+01P	3'ROX	CCATGGTTGGAGCGCAT	rs7307991	MP50
KSG:rs3026251:+01P	3'ROX	TGGGAAGACGGATCACC	rs3026251	MP50
KSG:rs3026252:+01P	3'ROX	GCTCTTAACTGTCTCTT	rs3026252	MP50
KSG:rs4882965:+01P	3'ROX	GAGACAAAGTTATGGGG	rs4882965	MP50
KSG:rs7299659:+01P	3'ROX	CTGGTCAGCTGTCAAGG	rs7299659	MP50
KSG:rs226389:+01P	3'ROX	CGATAGACGCTGAGCCA	rs226389	MP50
KSG:rs7134202:+01P	3'ROX	CTATGTGCCTGCCAGTG	rs7134202	MP50
KSG:rs1805750:+01P	3'ROX	TCTTAAAGCAATTTGGC	rs1805750	MP50
KSG:rs2024301:+01P	3'ROX	TACTTTTGAGAGGGGCA	rs2024301	MP50
KSG:rs4883475:+01P	3'ROX	GACTGTGGGTAACGGAA	rs4883475	MP50
KSG:rs2302516:+01P	3'ROX	GATACACAGAGCCTGGG	rs2302516	MP50
KSG:rs759052:+01P	3'ROX	TGCCGGGCGGTGGCAGT	rs759052	MP50
KSG:rs1133104:+01P	3'ROX	TTCTCTCCCTGTGTGCA	rs1133104	MP50
KSG:rs2110072:+01P	3'ROX	TCAAGACAGCCCTGGCT	rs2110072	MP50
KSG:rs2071079:+01P	3'ROX	TCAAGGCCGGGCAGGCA	rs2071079	MP50
KSG:rs226406:+01P	3'ROX	TATGTTGGTTGACTTAA	rs226406	MP50
KSG:rs3741854:+01P	3'ROX	ATAGGGAACAAGGAAGG	rs3741854	MP50
KSG:rs3809218:+01P	3'ROX	GCCTTTTCCTTCCTTTC	rs3809218	MP50

APPENDIX 3.

TABULATED RESULTS OF MEGAPLEX PCR TARGET RECOVERY

Target			MP50			MP75	
No	SNP(rsID)	Product length (bp)	PCR detection[1]	Array intensity[2]	GS20[3]	Array intensity[2]	GS20[3]
1	rs4882913	146	+	94	1	231	1
2	rs6486847	125	+	259	689	1756	327
3	rs10772596	125	nd	524	168	780	178
4	rs10840759	126	nd	1967	438	7128	360
5	rs2377422	128	nd	37	3	53	0
6	rs1894824	138	+	17	1	162	0
7	rs12819884	140	+	180	75	1139	71
8	rs1805721	135	nd	488	53	926	23
9	rs4304840	126	nd	289	1534	688	482
10	rs1561560	133	+	23	49	125	15
11	rs1805731	127	nd	290	51	1434	21
12	rs7300836	126	nd	-17	1704	-7	730
13	rs2580874	125	nd	703	39	2469	23
14	rs919209	125	nd	-5	0	-18	0
15	rs4620776	127	nd	50	19	1248	13
16	rs7300097	157	+	105	0	191	4
17	rs11046892	128	nd	20	0	6906	4290
18	rs2193005	140	nd	-14	0	-52	0
19	rs1805673	128	nd	211	32	636	4
20	rs11045985	126	nd	2	69	58	4
21	rs7310161	125	nd	930	689	2570	595
22	rs1805664	153	+	121	22	317	6
23	rs2075395	125	+	126	306	773	86
24	rs11046589	143	+	-3	3	12	1
25	rs2231754	132	+	31	347	235	180
26	rs4883146	125	nd	399	673	878	525
27	rs226380	126	nd	1184	2796	2708	2697
28	rs11057065	125	+	1415	468	4470	417
29	rs1894814	143	+	335	52	1081	22
30	rs4604965	133	+	1475	118	3217	73
31	rs10492115	126	nd	586	675	1249	553
32	rs7973072	125	nd	118	73	829	29
33	rs7307991	137	+	2232	2362	5161	1941
34	rs3026251	129	+	2069	5456	2997	3604
35	rs3026252	125	nd	1064	478	2874	114
36	rs4882965	125	nd	37	1062	71	964
37	rs7299659	126	+	1602	121	4939	75
38	rs226389	125	nd	2966	566	5093	427
39	rs7134202	125	nd	2179	264	7296	273
40	rs1805750	126	nd	172	2538	510	1007
41	rs2024301	129	+	491	129	2323	46
42	rs4883475	125	+	1524	12093	1859	9344
43	rs2302516	126	+	2781	30891	5778	28083
44	rs759052	126	nd	7883	447	12029	253
45	rs1133104	141	nd	3320	614	12452	745
46	rs2110072	125	nd	2079	962	3640	148
47	rs2071079	126	nd	460	94	2143	24
48	rs226406	297	nd	-15	0	-61	0
49	rs3741854	284	-	-5	0	-30	0
50	rs3809218	200	+	-5	0	696	0

continued

Target			MP50			MP75	
No	**SNP(rsID)**	**Product length (bp)**	**PCR detection[1]**	**Array intensity[2]**	**GS20[3]**	**Array intensity[2]**	**GS20[3]**
51	rs7298114	130				-61	0
52	rs1805677	127				-46	4
53	rs1805671	125				-64	11
54	rs10840926	145				-66	0
55	rs10770736	151				715	0
56	rs1805732	138				373	5
57	rs7954916	143				-30	0
58	rs7976134	125				-48	7
59	rs7307228	282				-33	0
60	rs6488610	130				560	51
61	rs1805651	136				117	19
62	rs17801827	125				-54	9
63	rs6487220	132				-29	12
64	rs14541	128				3750	3443
65	rs7138828	133				-49	2
66	rs1805740	126				1559	21
67	rs7307734	132				78	1
68	rs11046349	135				31	11
69	rs6488608	139				3222	6
70	rs6486850	160				39	1
71	rs1062836	145				-45	0
72	rs5014225	142				945	2
73	rs2889626	135				10859	117
74	rs12812856	324				-61	0
75	rs7842	304				86	0

(1) [+] product detected with secondary PCR, [-] product not detected, [nd] not determined.
(2) Average array hybridisation intensity for each target, after background subtraction.
(3) Number of GS20 sequence reads matching each target.

REFERENCES

1. Lander, E. S. *et al*. Initial sequencing and analysis of the human genome. *Nature.* **409**, 860-921 (2001).
2. Venter, J. C. *et al*. The Sequence of the Human Genome. *Science.* **291**, 1304-1351 (2001).
3. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature.* **431**, 931-945 (2004).
4. The ENCODE Project Consortium,. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science.* **306**, 636-640 (2004).
5. The ENCODE Project Consortium. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature.* **447**, 799-816 (2007).
6. Redon, R. *et al*. Global variation in copy number in the human genome. *Nature.* **444**, 444-454 (2006).
7. Stankiewicz, P. & Lupski, J. R. Genome architecture, rearrangements and genomic disorders. *Trends Genet.* **18**, 74-82 (2002).
8. Gill, P., Jeffreys, A. J. & Werrett, D. J. Forensic application of DNA 'fingerprints'. *Nature.* **318**, 577-579 (1985).
9. Jobling, M. A. & Gill, P. Encoded evidence: DNA in forensic analysis. *Nat Rev Genet.* **5**, 739-751 (2004).
10. Jeffreys, A. J., Wilson, V. & Thein, S. L. Individual-specific 'fingerprints'. *Nature.* **316**, 76-79 (1985).
11. The International HapMap Consortium. The International HapMap Project. *Nature.* **426**, 789-796 (2003).
12. Brookes, A. J. The essence of SNPs. *Gene.* **234**, 177-186 (1999).
13. Levy, S. *et al*. The Diploid Genome Sequence of an Individual Human. *PLoS Biol.* **5**, e254 (2007).
14. Sharp, A. J. *et al*. Segmental duplications and copy-number variation in the human genome. *Am J Hum Genet.* **77**, 78-88 (2005).
15. Tuzun, E. *et al*. Fine-scale structural variation of the human genome. *Nat Genet.* **37**, 727-732 (2005).
16. Sebat, J. *et al*. Large-Scale Copy Number Polymorphism in the Human Genome. *Science.* **305**, 525-528 (2004).
17. Iafrate, A. J. *et al*. Detection of large-scale variation in the human genome. *Nat Genet.* **36**, 949-951 (2004).
18. Sulem, P. *et al*. Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nat Genet.* **39**, 1443-1452 (2007).
19. Shriver, M. D. *et al*. Skin pigmentation, biogeographical ancestry and admixture mapping. *Hum Genet.* **112**, 387-399 (2003).
20. Evans, W. E. & Relling, M. V. Moving towards individualized medicine with pharmacogenomics. *Nature.* **429**, 464-468 (2004).
21. Weinshilboum, R. Inheritance and Drug Response. *N Engl J Med.* **348**, 529-537 (2003).
22. Mathew, C. Science medicine and the future: Postgenomic technologies: hunting the genes for common disorders. *BMJ.* **322**, 1031-1034 (2001).

23. Dean, M. Approaches to identify genes for complex human diseases: Lessons from Mendelian disorders. *Hum Mutat.* **22**, 261-274 (2003).

24. Cardon, L. R. & Bell, J. I. Association study designs for complex diseases. *Nat Rev Genet.* **2**, 91-99 (2001).

25. Hattersley, A. T. & McCarthy, M. I. What makes a good genetic association study? *Lancet.* **366**, 1315-1323 (2005).

26. Roses, A. D. Apolipoprotein E affects the rate of Alzheimer disease expression: beta-amyloid burden is a secondary consequence dependent on APOE genotype and duration of disease. *J Neuropathol Exp Neurol.* **53**, 429-437 (1994).

27. Marron, M. P. *et al*. Insulin-dependent diabetes mellitus (IDDM) is associated with CTLA4 polymorphisms in multiple ethnic groups. *Hum Mol Genet.* **6**, 1275-1282 (1997).

28. Ueda, H. *et al*. Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. *Nature.* **423**, 506-511 (2003).

29. Altshuler, D. *et al*. The common PPAR[gamma] Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat Genet.* **26**, 76-80 (2000).

30. Gloyn, A. L. *et al*. Large-Scale Association Studies of Variants in Genes Encoding the Pancreatic {beta}-Cell KATP Channel Subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) Confirm That the KCNJ11 E23K Variant Is Associated With Type 2 Diabetes. *Diabetes.* **52**, 568-572 (2003).

31. Blomqvist, M. *et al*. Towards compendia of negative genetic association studies: an example for Alzheimer disease. *Hum Genet.* **119**, 29-37 (2006).

32. Prince, J. A. *et al*. Lack of replication of association findings in complex disease: an analysis of 15 polymorphisms in prior candidate genes for sporadic Alzheimer's disease. *Eur J Hum Genet.* **9**, 437-444 (2001).

33. Risch, N. & Merikangas, K. The future of genetic studies of complex human diseases. *Science.* **273**, 1516-1517 (1996).

34. Risch, N. J. Searching for genetic determinants in the new millennium. *Nature.* **405**, 847-856 (2000).

35. Meng, Z., Zaykin, D., Xu, C., Wagner, M. & Ehm, M. G. Selection of Genetic Markers for Association Analyses, Using Linkage Disequilibrium and Haplotypes. *Am J Hum Genet.* **73**, 115-130 (2003).

36. Carlson, C. S. *et al*. Selecting a Maximally Informative Set of Single-Nucleotide Polymorphisms for Association Analyses Using Linkage Disequilibrium. *Am J Hum Genet.* **74**, 106-120 (2004).

37. Johnson, G. C. L. *et al*. Haplotype tagging for the identification of common disease genes. *Nat Genet.* **29**, 233-237 (2001).

38. Kruglyak, L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat Genet.* **22**, 139-144 (1999).

39. Reich, D. E. *et al*. Linkage disequilibrium in the human genome. *Nature.* **411**, 199-204 (2001).

40. The International HapMap Consortium. A haplotype map of the human genome. *Nature.* **437**, 1299-1320 (2005).

41. Moffatt, M. F. *et al*. Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature.* **448**, 470-473 (2007).

42. Broderick, P. *et al*. A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. *Nat Genet.* **39**, 1317 (2007).

43. Zanke, B. W. *et al*. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat Genet.* **39**, 989-994 (2007).

44. Gudmundsson, J. *et al*. Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. *Nat Genet.* **39**, 977-983 (2007).

45. Scott, L. J. *et al*. A Genome-Wide Association Study of Type 2 Diabetes in Finns Detects Multiple Susceptibility Variants. *Science.* **316**, 1341-1345 (2007).

46. Sladek, R. *et al*. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature.* **445**, 881-885 (2007).

47. The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature.* **447**, 661-678 (2007).

48. Libioulle, C. *et al*. Novel Crohn Disease Locus Identified by Genome-Wide Association Maps to a Gene Desert on 5p13.1 and Modulates Expression of PTGER4. *PLoS Genet.* **3**, e58 (2007).

49. NCI-NHGRI Working Group on Replication in Association Studies *et al*. Replicating genotype-phenotype associations. *Nature.* **447**, 655-660 (2007).

50. Marchini, J., Donnelly, P. & Cardon, L. R. Genome-wide strategies for detecting multiple loci that influence complex diseases. *Nat Genet.* **37**, 413-417 (2005).

51. Shi, M. M. Enabling Large-Scale Pharmacogenetic Studies by High-Throughput Mutation Detection and Genotyping Technologies. *Clin Chem.* **47**, 164-172 (2001).

52. Mathew, C. G. DNA diagnostics: goals and Challenges. *Br Med Bull.* **55**, 325-339 (1999).

53. Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K. & Sekiya, T. Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc Natl Acad Sci U S A.* **86**, 2766-2770 (1989).

54. Orita, M., Suzuki, Y., Sekiya, T. & Hayashi, K. Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics.* **5**, 874-879 (1989).

55. Larsen, L. A., Christiansen, M., Vuust, J. & Skytt Andersen, P. High-throughput single-strand conformation polymorphism analysis by automated capillary electrophoresis: Robust multiplex analysis and pattern-based identification of allelic variants. *Hum Mutat.* **13**, 318-327 (1999).

56. Andersen, P. S., Jespersgaard, C., Vuust, J., Christiansen, M. & Larsen, L. A. High-throughput single strand conformation polymorphism mutation detection by automated capillary array electrophoresis: validation of the method. *Hum Mutat.* **21**, 116-122 (2003).

57. Tian, H. *et al*. High-throughput single-strand conformation polymorphism analysis on a microfabricated capillary array electrophoresis device. *Electrophoresis.* **26**, 1834-1842 (2005).

58. Fischer, S. G. & Lerman, L. S. DNA fragments differing by single base-pair substitutions are separated in denaturing gradient gels: correspondence with melting theory. *Proc Natl Acad Sci U S A.* **80**, 1579-1583 (1983).

59. Myers, R. M., Maniatis, T. & Lerman, L. S. Detection and localization of single base changes by denaturing gradient gel electrophoresis. *Methods Enzymol.* **155**, 501-527 (1987).

60. Rosenbaum, V. & Riesner, D. Temperature-gradient gel electrophoresis. Thermodynamic analysis of nucleic acids and proteins in purified form and in cellular extract. *Biophys Chem.* **26**, 235-246 (1987).

61. Wartell, R. M., Hosseini, S. H. & Moran, C. P. Detecting base pair substitutions in DNA fragments by temperature-gradient gel electrophoresis. *Nucleic Acids Res.* **18**, 2699-2705 (1990).

62. Biyani, M. & Nishigaki, K. Hundredfold productivity of genome analysis by introduction of microtemperature-gradient gel electrophoresis. *Electrophoresis.* **22**, 23-28 (2001).

63. Li, Q., Liu, Z., Monroe, H. & Culiat, C. T. Integrated platform for detection of DNA sequence variants using capillary array electrophoresis. *Electrophoresis.* **23**, 1499-1511 (2002).

64. Girald-Rosa, W., Vleugels, R. A., Musiek, A. C. & Sligh, J. E. High-Throughput Mitochondrial Genome Screening Method for Nonmelanoma Skin Cancer Using Multiplexed Temperature Gradient Capillary Electrophoresis. *Clin Chem.* **51**, 305-311 (2005).

65. Underhill, P. A. *et al*. Detection of Numerous Y Chromosome Biallelic Polymorphisms by Denaturing High-Performance Liquid Chromatography. *Genome Res.* **7**, 996-1005 (1997).

66. O'Donovan, M. C. *et al*. Blind Analysis of Denaturing High-Performance Liquid Chromatography as a Tool for Mutation Detection. *Genomics.* **52**, 44-49 (1998).

67. Premstaller, A. *et al*. Temperature-Modulated Array High-Performance Liquid Chromatography. *Genome Res.* **11**, 1944-1951 (2001).

68. Hestekin, C. N. & Barron, A. E. The potential of electrophoretic mobility shift assays for clinical mutation detection. *Electrophoresis.* **27**, 3805-3815 (2006).

69. Kuhn, D. N. *et al*. Single-strand conformation polymorphism analysis of candidate genes for reliable identification of alleles by capillary array electrophoresis. *Electrophoresis.* **26**, 112-125 (2005).

70. Chou, L., Gedge, F. & Lyon, E. Complete Gene Scanning by Temperature Gradient Capillary Electrophoresis Using the Cystic Fibrosis Transmembrane Conductance Regulator Gene as a Model. *J Mol Diagn.* **7**, 111-120 (2005).

71. Lipsky, R. H. *et al*. DNA Melting Analysis for Detection of Single Nucleotide Polymorphisms. *Clin Chem.* **47**, 635-644 (2001).

72. Gundry, C. N. *et al*. Amplicon Melting Analysis with Labeled Primers: A Closed-Tube Method for Differentiating Homozygotes and Heterozygotes. *Clin Chem.* **49**, 396-406 (2003).

73. Wittwer, C. T., Reed, G. H., Gundry, C. N., Vandersteen, J. G. & Pryor, R. J. High-Resolution Genotyping by Amplicon Melting Analysis Using LCGreen. *Clin Chem.* **49**, 853-860 (2003).

74. Reed, G. H. & Wittwer, C. T. Sensitivity and Specificity of Single-Nucleotide Polymorphism Scanning by High-Resolution Melting Analysis. *Clin Chem.* **50**, 1748-1754 (2004).

75. Liew, M. *et al*. Genotyping of Single-Nucleotide Polymorphisms by High-Resolution Melting of Small Amplicons. *Clin Chem.* **50**, 1156-1164 (2004).

76. Sanger, F., Nicklen, S. & Coulson, A. R. DNA Sequencing with Chain-Terminating Inhibitors. *Proc Natl Acad Sci U S A.* **74**, 5463-5467 (1977).

77. Metzker, M. L. Emerging technologies in DNA sequencing. *Genome Res.* **15**, 1767-1776 (2005).

78. Shendure, J., Mitra, R. D., Varma, C. & Church, G. M. Advanced sequencing technologies: methods and goals. *Nat Rev Genet.* **5**, 335-344 (2004).

79. Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlén, M. & Nyrén, P. Real-Time DNA Sequencing Using Detection of Pyrophosphate Release. *Anal Biochem.* **242**, 84-89 (1996).

80. Ronaghi, M., Uhlen, M. & Nyren, P. A Sequencing Method Based on Real-Time Pyrophosphate. *Science.* **281**, 363-365 (1998).

81. Margulies, M. *et al*. Genome sequencing in microfabricated high-density picolitre reactors. *Nature.* **437**, 376-380 (2005).

82. Dressman, D., Yan, H., Traverso, G., Kinzler, K. W. & Vogelstein, B. Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proc Natl Acad Sci U S A.* **100**, 8817-8822 (2003).

83. Mitra, R. D., Shendure, J., Olejnik, J., Krzymanska-Olejnik, E. & Church, G. M. Fluorescent in situ sequencing on polymerase colonies. *Anal Biochem.* **320**, 55-65 (2003).

84. Aksyonov, S. A. *et al*. Multiplexed DNA sequencing-by-synthesis. *Anal Biochem.* **348**, 127-138 (2006).

85. Seo, T. S. *et al*. Four-color DNA sequencing by synthesis on a chip using photocleavable fluorescent nucleotides. *Proc Natl Acad Sci U S A.* **102**, 5926-5931 (2005).

86. Ju, J. *et al*. Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators. *Proc Natl Acad Sci U S A.* **103**, 19635-19640 (2006).

87. Brenner, S. *et al*. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat Biotechnol.* **18**, 630-634 (2000).

88. Thomas, R. K. *et al*. Sensitive mutation detection in heterogeneous cancer specimens by massively parallel picoliter reactor sequencing. *Nat Med.* **12**, 852-855 (2006).

89. Akey, J. M. *et al*. Melting curve analysis of SNPs (McSNP): a gel-free and inexpensive approach for SNP genotyping. *Biotechniques.* **30**, 358-367 (2001).

90. Livak, K. J. Allelic discrimination using fluorogenic probes and the 5′ nuclease assay. *Genet Anal.* **14**, 143-149 (1999).

91. Lyamichev, V. *et al*. Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes. *Nat Biotechnol.* **17**, 292-296 (1999).

92. Olivier, M. The Invader assay for SNP genotyping. *Mutat Res.* **573**, 103-110 (2005).

93. Lu, M. *et al*. A surface invasive cleavage assay for highly parallel SNP analysis. *Hum Mutat.* **19**, 416-422 (2002).

94. Wu, D. Y., Ugozzoli, L., Pal, B. K. & Wallace, R. B. Allele-Specific Enzymatic Amplification of beta-globin Genomic DNA for Diagnosis of Sickle Cell Anemia. *Proc Natl Acad Sci U S A.* **86**, 2757-2760 (1989).

95. Newton, C. R. *et al*. Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res.* **17**, 2503-2516 (1989).

96. Latorra, D., Campbell, K., Wolter, A. & Hurley, J. M. Enhanced allele-specific PCR discrimination in SNP genotyping using 3' locked nucleic acid (LNA) primers. *Hum Mutat.* **22**, 79-85 (2003).

97. Whitcombe, D. *et al*. A homogeneous fluorescence assay for PCR amplicons: its application to real-time, single-tube genotyping. *Clin Chem.* **44**, 918-923 (1998).

98. Quek, D. L. *et al*. Rapid carrier screening for β-thalassemia by single-step allele-specific PCR and detection. *Clin Biochem.* **40**, 427-430 (2007).

99. Pastinen, T. *et al*. A System for Specific, High-throughput Genotyping by Allele-specific Primer Extension on Microarrays. *Genome Res.* **10**, 1031-1042 (2000).

100. Gunderson, K. L., Steemers, F. J., Lee, G., Mendoza, L. G. & Chee, M. S. A genome-wide scalable SNP genotyping assay using microarray technology. *Nat Genet.* **37**, 549-554 (2005).

101. Gunderson, K. L. *et al*. in *Methods in Enzymology* (ed Alan Kimmel,and Brian Oliver) 359-376 (Academic Press, 2006).

102. Syvanen, A., Aalto-Setälä, K., Harju, L., Kontula, K. & Söderlund, H. A primer guided nucleotide incorporation assay in the genotyping of apolipoprotein E. *Genomics.* **8**, 684-692 (1990).

103. Syvanen, A., Sajantila, A. & Lukka, M. Identification of individuals by analysis of biallelic DNA markers, using PCR and solid-phase minisequencing. *Am J Hum Genet.* **52**, 46-59 (1993).

104. Pastinen, T., Partanen, J. & Syvanen, A. C. Multiplex, fluorescent, solid-phase minisequencing for efficient screening of DNA sequence variation. *Clin Chem.* **42**, 1391-1397 (1996).

105. Lindroos, K., Sigurdsson, S., Johansson, K., Ronnblom, L. & Syvanen, A. Multiplex SNP genotyping in pooled DNA samples by a four-colour microarray system. *Nucleic Acids Res.* **30**, e70 (2002).

106. Pastinen, T., Kurg, A., Metspalu, A., Peltonen, L. & Syvanen, A. Minisequencing: A Specific Tool for DNA Analysis and Diagnostics on Oligonucleotide Arrays. *Genome Res.* **7**, 606-614 (1997).

107. Raitio, M. *et al*. Y-Chromosomal SNPs in Finno-Ugric-Speaking Populations Analyzed by Minisequencing on Microarrays. *Genome Res.* **11**, 471-482 (2001).

108. Steemers, F. J. *et al*. Whole-genome genotyping with the single-base extension assay. *Nat Methods.* **3**, 31-33 (2006).

109. Landegren, U., Kaiser, R., Sanders, J. & Hood, L. A ligase-mediated gene detection technique. *Science.* **241**, 1077-1080 (1988).

110. Grossman, P. D. *et al*. High-density multiplex detection of nucleic acid sequences: oligonucleotide ligation assay and sequence-coded separation. *Nucleic Acids Res.* **22**, 4527-4534 (1994).

111. Baron, H. *et al*. Oligonucleotide ligation assay (OLA) for the diagnosis of familial hypercholesterolemia. *Nat Biotechnol.* **14**, 1279-1282 (1996).

112. Chen, X., Livak, K. J. & Kwok, P. A Homogeneous, Ligase-Mediated DNA Diagnostic Test. *Genome Res.* **8**, 549-556 (1998).

113. Nilsson, M. *et al*. Padlock probes: circularizing oligonucleotides for localized DNA detection. *Science.* **265**, 2085-2088 (1994).

114. Baner, J., Nilsson, M., Mendel-Hartvig, M. & Landegren, U. Signal amplification of padlock probes by rolling circle replication. *Nucleic Acids Res.* **26**, 5073-5078 (1998).

115. Baner, J. *et al*. Parallel gene analysis with allele-specific padlock probes and tag microarrays. *Nucleic Acids Res.* **31**, e103 (2003).

116. Hardenbol, P. *et al*. Multiplexed genotyping with sequence-tagged molecular inversion probes. *Nat Biotechnol.* **21**, 673-678 (2003).

117. Hardenbol, P. *et al*. Highly multiplexed molecular inversion probe genotyping: Over 10,000 targeted SNPs genotyped in a single tube assay. *Genome Res.* **15**, 269-275 (2005).

118. Shen, R. *et al*. High-throughput SNP genotyping on universal bead arrays. *Mutat Res.* **573**, 70-82 (2005).

119. Kafatos, F. C., Jones, C. W. & Efstratiadis, A. Determination of nucleic acid sequence homologies and relative concentrations by a dot hybridization procedure. *Nucleic Acids Res.* **7**, 1541-1552 (1979).

120. Saiki, R. K., Walsh, P. S., Lewenson, C. H. & Erlich, H. A. Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes. *Proc Natl Acad Sci U S A.* **86**, 6230-6234 (1989).

121. Yim, S., Park, H. G., Chang, H. N. & Cho, D. Array-based mutation detection of *BRCA1* using direct probe/target hybridization. *Anal Biochem.* **337**, 332-337 (2005).

122. Matsuzaki, H. *et al*. Parallel Genotyping of Over 10,000 SNPs Using a One-Primer Assay on a High-Density Oligonucleotide Array. *Genome Res.* **14**, 414-425 (2004).

123. Matsuzaki, H. *et al*. Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays. *Nat Methods.* **1**, 109-111 (2004).

124. Hacia, J. G. Resequencing and mutational analysis using oligonucleotide microarrays. *Nat Genet.* **21**, 42-47 (1999).

125. Chee, M. *et al*. Accessing genetic information with high-density DNA arrays. *Science.* **274**, 610-614 (1996).

126. Maitra, A. *et al*. The Human MitoChip: A High-Throughput Sequencing Microarray for Mitochondrial Mutation Detection. *Genome Res.* **14**, 812-819 (2004).

127. Wang, D. G. & Lipshutz, R. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science.* **280**, 1077-1082 (1998).

128. Warrington, J. A. *et al*. New developments in high-throughput resequencing and variation detection using high density microarrays. *Hum Mutat.* **19**, 402-409 (2002).

129. Patil, N. *et al*. Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science.* **294**, 1719(5)-1723 (2001).

130. Tyagi, S., Bratu, D. P. & Kramer, F. R. Multicolor molecular beacons for allele discrimination. *Nat Biotechnol.* **16**, 49-53 (1998).

131. Howell, W. M., Jobs, M., Gyllensten, U. & Brookes, A. J. Dynamic allele-specific hybridization. *Nat Biotechnol.* **17**, 87-88 (1999).

132. Jobs, M., Howell, W. M., Strömqvist, L., Mayr, T. & Brookes, A. J. DASH-2: Flexible, Low-Cost, and High-Throughput SNP Genotyping by Dynamic Allele-Specific Hybridization on Membrane Arrays. *Genome Res.* **13**, 916-924 (2003).

133. Stryer, L. & Haugland, R. P. Energy transfer: a spectroscopic ruler. *Proc Natl Acad Sci U S A.* **58**, 719-726 (1967).

134. Wu, P. G. & Brand, L. Resonance Energy Transfer: Methods and Applications. *Anal Biochem.* **218**, 1-13 (1994).

135. Lay, M. J. & Wittwer, C. T. Real-time fluorescence genotyping of factor V Leiden during rapid-cycle PCR. *Clin Chem.* **43**, 2262-2267 (1997).

136. Bernard, P. S., Ajioka, R. S., Kushner, J. P. & Wittwer, C. T. Homogeneous Multiplex Genotyping of Hemochromatosis Mutations with Fluorescent Hybridization Probes. *Am J Pathol.* **153**, 1055-1061 (1998).

137. Herrmann, M. G., Dobrowolski, S. F. & Wittwer, C. T. Rapid {beta}-Globin Genotyping by Multiplexing Probe Melting Temperature and Color. *Clin Chem.* **46**, 425-428 (2000).

138. Vrettou, C., Traeger-Synodinos, J., Tzetis, M., Malamis, G. & Kanavakis, E. Rapid Screening of Multiple β-Globin Gene Mutations by Real-Time PCR on the LightCycler: Application to Carrier Screening and Prenatal Diagnosis of Thalassemia Syndromes. *Clin Chem.* **49**, 769-776 (2003).

139. Afonina, I. A., Reed, M. W., Lusby, E., Shishkina, I. G. & Belousov, Y. S. Minor groove binder-conjugated DNA probes for quantitative DNA detection by hybridization-triggered fluorescence. *Biotechniques.* **32**, 940-949 (2002).

140. Belousov, Y. S. *et al*. Single nucleotide polymorphism genotyping by two colour melting curve analysis using the MGB Eclipse™ Probe System in challenging sequence environment. *Hum Genomics.* **1**, 209-217 (2004).

141. Crockett, A. O. & Wittwer, C. T. Fluorescein-Labeled Oligonucleotides for Real-Time PCR: Using the Inherent Quenching of Deoxyguanosine Nucleotides. *Anal Biochem.* **290**, 89-97 (2001).

142. Leman, A. R., Polochock, S., Mole, S. E., Pearce, D. A. & Rothberg, P. G. Homogeneous PCR nucleobase quenching assays to detect four mutations that cause neuronal ceroid lipofuscinosis: T75P and R151X in CLN1, and IVS5-1G > C and R208X in CLN2. *J Neurosci Methods.* **157**, 124-131 (2006).

143. Rothberg, P. G., Ramirez-Montealegre, D., Frazier, S. D. & Pearce, D. A. Homogeneous Polymerase Chain Reaction Nucleobase Quenching Assay to Detect the 1-kbp Deletion in CLN3 That Causes Batten Disease. *J Mol Diagn.* **6**, 260-263 (2004).

144. Zhou, L., Myers, A. N., Vandersteen, J. G., Wang, L. & Wittwer, C. T. Closed-Tube Genotyping with Unlabeled Oligonucleotide Probes and a Saturating DNA Dye. *Clin Chem.* **50**, 1328-1335 (2004).

145. Howell, W. M., Jobs, M. & Brookes, A. J. iFRET: An Improved Fluorescence System for DNA-Melting Analysis. *Genome Res.* **12**, 1401-1407 (2002).

146. Prince, J. A. *et al*. Robust and Accurate Single Nucleotide Polymorphism Genotyping by Dynamic Allele-Specific Hybridization (DASH): Design Criteria and Assay Validation. *Genome Res.* **11**, 152-162 (2001).

147. Sawyer, S. L., Howell, W. M. & Brookes, A. J. Scoring insertion-deletion polymorphisms by dynamic allele-specific hybridization. *Biotechniques.* **35**, 292-298 (2003).

148. Jobs, M., Howell, W. M. & Brookes, A. J. Creating arrays by centrifugation. *Biotechniques.* **32**, 1322-1329 (2002).

149. Fredman, D., Jobs, M., Strömqvist, L. & Brookes, A. J. DFold: PCR design that minimizes secondary structure and optimizes downstream genotyping applications. *Hum Mutat.* **24**, 1-8 (2004).

150. Saiki, R. K. *et al*. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science.* **230**, 1350-54 (1985).

151. Chamberlain, J. S., Gibbs, R. A., Rainer, J. E., Nguyen, P. N. & Thomas, C. Deletion screening of the Duchenne muscular dystrophy locus via multiplex DNA amplification. *Nucleic Acids Res.* **16**, 11141-11156 (1988).

152. Ferrie, R. M. *et al*. Development, multiplexing, and application of ARMS tests for common mutations in the CFTR gene. *Am J Hum Genet.* **51**, 251-262 (1992).

153. Markoulatos, P., Siafakas, N. & Moncany, M. Multiplex polymerase chain reaction: A practical approach. *J Clin Lab Anal.* **16**, 47-51 (2002).

154. Schoske, R., Vallone, P. M., Ruitberg, C. M. & Butler, J. M. Multiplex PCR design strategy used for the simultaneous amplification of 10 Y chromosome short tandem repeat (STR) loci. *Anal Bioanal Chem.* **375**, 333-343 (2003).

155. Vandenvelde, C., Verstraete, M. & Van Beers, D. Fast multiplex polymerase chain reaction on boiled clinical samples for rapid viral diagnosis. *J Virol Methods.* **30**, 215-227 (1990).

156. Vallone, P. M. & Butler, J. M. AutoDimer: a screening tool for primer-dimer and hairpin structures. *Biotechniques.* **37**, 226-231 (2004).

157. Edwards, M. C. & Gibbs, R. A. Multiplex PCR: advantages, development, and applications. *PCR Methods Appl.* **3**, S65-75 (1994).

158. Wang, H. *et al*. A genotyping system capable of simultaneously analyzing >1000 single nucleotide polymorphisms in a haploid genome. *Genome Res.* **15**, 276-283 (2005).

159. Henegariu, O., Heerema, N. A., Dlouhy, S. R., Vance, G. H. & Vogt, P. H. Multiplex PCR: Critical Parameters and Step-by-Step Protocol. *Biotechniques.* **23**, 504-511 (1997).

160. Broude, N. E., Zhang, L., Woodward, K., Englert, D. & Cantor, C. R. Multiplex allele-specific target amplification based on PCR suppression. *Proc Natl Acad Sci U S A* **98**, 206-211 (2001).

161. Syvanen, A. Toward genome-wide SNP genotyping. *Nat Genet.* **37 Supplement**, S5-S10 (2005).

162. Brownie, J. *et al*. The elimination of primer-dimer accumulation in PCR. *Nucleic Acids Res.* **25**, 3235-3241 (1997).

163. Lin, Z., Cui, X. & Li, H. Multiplex genotype determination at a large number of gene loci. *Proc Natl Acad Sci U S A.* **93**, 2582-2587 (1996).

164. Kennedy, G. C. *et al*. Large-scale genotyping of complex DNA. *Nat Biotechnol.* **21**, 1233-1237 (2003).

165. Shapero, M. H. *et al*. MARA: a novel approach for highly multiplexed locus-specific SNP genotyping using high-density DNA oligonucleotide arrays. *Nucleic Acids Res.* **32**, e181 (2004).

166. Dahl, F., Gullberg, M., Stenberg, J., Landegren, U. & Nilsson, M. Multiplex amplification enabled by selective circularization of large sets of genomic DNA fragments. *Nucleic Acids Res.* **33**, e71 (2005).

167. Callow, M. J., Drmanac, S. & Drmanac, R. Selective DNA amplification from complex genomes using universal double-sided adapters. *Nucleic Acids Res.* **32**, e21 (2004).

168. Oroskar, A. A. *et al*. Detection of immobilized amplicons by ELISA-like techniques. *Clin Chem.* **42**, 1547-1555 (1996).

169. Andreadis, J. D. & Chrisey, L. A. Use of immobilized PCR primers to generate covalently immobilized DNAs for in vitro transcription/translation reactions. *Nucleic Acids Res.* **28**, e5 (2000).

170. Turner, M. S. *et al*. Solid-Phase Amplification for Detection of C282Y and H63D Hemochromatosis (HFE) Gene Mutations. *Clin Chem.* **47**, 1384-1389 (2001).

171. Sjöroos, M., Ilonen, J. & Lövgren, T. Solid-Phase PCR with Hybridization and Time-resolved Fluorometry for Detection of HLA-B27. *Clin Chem.* **47**, 498-504 (2001).

172. Lockley, A. K., Jones, C. G., Bruce, J. S., Franklin, S. J. & Bardsley, R. G. Colorimetric detection of immobilised PCR products generated on a solid support. *Nucleic Acids Res.* **25**, 1313-1314 (1997).

173. Huber, M. *et al*. Detection of Single Base Alterations in Genomic DNA by Solid Phase Polymerase Chain Reaction on Oligonucleotide Microarrays. *Anal Biochem.* **299**, 24-30 (2001).

174. Huber, M. *et al*. Accessing Single Nucleotide Polymorphisms in Genomic DNA by Direct Multiplex Polymerase Chain Reaction Amplification on Oligonucleotide Microarrays. *Anal Biochem.* **303**, 25-33 (2002).

175. Mitterer, G. *et al*. Microarray-Based Identification of Bacteria in Clinical Samples by Solid-Phase PCR Amplification of 23S Ribosomal DNA Sequences. *J Clin Microbiol.* **42**, 1048-1057 (2004).

176. Bing, D. H. et al. *Bridge Amplification: A Solid Phase PCR System for the Amplification and Detection of Allelic Differences in Single Copy Genes* (Seventh International Symposium on Human Identification, 1996).

177. Adessi, C. *et al*. Solid phase DNA amplification: characterisation of primer attachment and amplification mechanisms. *Nucleic Acids Res.* **28**, e87 (2000).

178. Shapero, M. H., Leuther, K. K., Nguyen, A., Scott, M. & Jones, K. W. SNP Genotyping by Multiplexed Solid-Phase Amplification and Fluorescent Minisequencing. *Genome Res.* **11**, 1926-1934 (2001).

179. Tillib, S. V., Strizhkov, B. N. & Mirzabekov, A. D. Integration of Multiple PCR Amplifications and DNA Mutation Analyses by Using Oligonucleotide Microchip. *Anal Biochem.* **292**, 155-160 (2001).

180. Pemov, A., Modi, H., Chandler, D. P. & Bavykin, S. DNA analysis with multiplex microarray-enhanced PCR. *Nucleic Acids Res.* **33**, e11 (2005).

181. Allawi, H. T. & SantaLucia, J. Thermodynamics and NMR of Internal G·T Mismatches in DNA. *Biochemistry.* **36**, 10581-10594 (1997).

182. Owczarzy, R. *et al*. Effects of Sodium Ions on DNA Duplex Oligomers: Improved Predictions of Melting Temperatures. *Biochemistry.* **43**, 3537-3554 (2004).

183. Panjkovich, A., Norambuena, T. & Melo, F. dnaMATE: a consensus melting temperature prediction server for short DNA sequences. *Nucleic Acids Res.* **33**, W570-572 (2005).

184. Baum, M. *et al*. Validation of a novel, fully integrated and flexible microarray benchtop facility for gene expression profiling. *Nucleic Acids Res.* **31**, e151 (2003).

185. Strömqvist Meuzelaar, L., Hopkins, K., Liebana, E. & Brookes, A. J. DNA Diagnostics by Surface-Bound Melt-Curve Reactions. *J Mol Diagn.* **9**, 30-41 (2007).

186. Millward, H., Samowitz, W., Wittwer, C. T. & Bernard, P. S. Homogeneous Amplification and Mutation Scanning of the p53 Gene Using Fluorescent Melting Curves. *Clin Chem.* **48**, 1321-1328 (2002).

187. Hopkins, K. L., Davies, R. H. & Threlfall, E. J. Mechanisms of quinolone resistance in Escherichia coli and Salmonella: Recent developments. *Int J Antimicrob Agents.* **25**, 358-373 (2005).

188. Brown, J. C., Shanahan, P. M. A., Jesudason, M. V., Thomson, C. J. & Amyes, S. G. B. Mutations responsible for reduced susceptibility to 4-quinolones in clinical isolates of multi-resistant Salmonella typhi in India. *J Antimicrob Chemother.* **37**, 891-900 (1996).

189. Hirose, K., Tamura, K. & Watanabe, H. Screening Method for *Salmonella enterica* Serovar Typhi and Serovar Paratyphi A with Reduced Susceptibility to Fluoroquinolones by PCR-Restriction Fragment Length Polymorphism. *Microbiol Immunol.* **47**, 161-165 (2003).

190. Qiang, Y. Z. *et al*. Use of a rapid mismatch PCR method to detect gyrA and parC mutations in ciprofloxacin-resistant clinical isolates of Escherichia coli. *J Antimicrob Chemother.* **49**, 549-552 (2002).

191. Walker, R. A. *et al*. Use of a LightCycler *gyrA* Mutation Assay for Rapid Identification of Mutations Conferring Decreased Susceptibility to Ciprofloxacin in Multiresistant *Salmonella enterica* Serotype Typhimurium DT104 Isolates. *J Clin Microbiol.* **39**, 1443-1448 (2001).

192. Liebana, E. *et al*. Comparison of *gyrA* Mutations, Cyclohexane Resistance, and the Presence of Class I Integrons in *Salmonella enterica* from Farm Animals in England and Wales. *J Clin Microbiol.* **40**, 1481-1486 (2002).

193. Walker, R. A., Skinner, J. A., Ward, L. R. & Threlfall, E. J. LightCycler gyrA mutation assay (GAMA) identifies heterogeneity in GyrA in Salmonella enterica serotypes Typhi and Paratyphi A with decreased susceptibility to ciprofloxacin. *Int J Antimicrob Agents.* **22**, 622-625 (2003).

194. Esaki, H. *et al*. Rapid detection of quinolone-resistant Salmonella by real time SNP genotyping. *J Microbiol Methods.* **58**, 131-134 (2004).

195. Morita, M. *et al*. Development of a Real-Time PCR Assay for Detection of *gyrA* Mutations Associated with Reduced Susceptibility to Ciprofloxacin in *Salmonella enterica* Serovar Typhi and Paratyphi A. *Microbiol Immunol.* **50**, 707-711 (2006).

196. Yu, X., Susa, M., Knabbe, C., Schmid, R. D. & Bachmann, T. T. Development and Validation of a Diagnostic DNA Microarray To Detect Quinolone-Resistant *Escherichia coli* among Clinical Isolates. *J Clin Microbiol.* **42**, 4083-4091 (2004).

197. Piddock, L. J., Ricci, V., McLaren, I. & Griggs, D. J. Role of mutation in the gyrA and parC genes of nalidixic-acid-resistant salmonella serotypes isolated from animals in the United Kingdom. *J Antimicrob Chemother.* **41**, 635-641 (1998).

198. Eaves, D. J., Liebana, E., Woodward, M. J. & Piddock, L. J. V. Detection of *gyrA* Mutations in Quinolone-Resistant *Salmonella enterica* by Denaturing High-Performance Liquid Chromatography. *J Clin Microbiol.* **40**, 4121-4125 (2002).

199. Slinger, R., Bellfoy, D., Desjardins, M. & Chan, F. High-resolution melting assay for the detection of gyrA mutations causing quinolone resistance in Salmonella enterica serovars Typhi and Paratyphi. *Diagn Microbiol Infect Dis.* **57**, 455-458 (2007).

200. Hopkins, K. L., Arnold, C. & Threlfall, E. J. Rapid detection of gyrA and parC mutations in quinolone-resistant Salmonella enterica using Pyrosequencing® technology. *J Microbiol Methods.* **68**, 163-171 (2007).

201. Clark, B. E. & Thein, S. L. Molecular diagnosis of haemoglobin disorders. *Clin Lab Haematol.* **26**, 159-176 (2004).

202. Patrinos, G. P., Kollia, P. & Papadakis, M. N. Molecular diagnosis of inherited disorders: lessons from hemoglobinopathies. *Hum Mutat.* **26**, 399-412 (2005).

203. Gu, X. & Zeng, Y. A Review of the Molecular Diagnosis of Thalassemia. *Hematology.* **7**, 203-209 (2002).

204. Giardine, B. *et al*. HbVar database of human hemoglobin variants and thalassemia mutations: 2007 update. *Hum Mutat.* **28**, 206-206 (2007).

205. Angastiniotis, M. & Modell, B. Global Epidemiology of Hemoglobin Disorders. *Ann. N. Y. Acad. Sci.* **850**, 251-269 (1998).

206. Old, J. M. Screening and genetic diagnosis of haemoglobin disorders. *Blood Rev.* **17**, 43-53 (2003).

207. Vrettou, C. *et al*. Real-time PCR for single-cell genotyping in sickle cell and thalassemia syndromes as a rapid, accurate, reliable, and widely applicable protocol for preimplantation genetic diagnosis. *Hum Mutat.* **23**, 513-521 (2004).

208. Chan, K., Wong, M. S., Chan, T. K. & V, C. A thalassaemia array for Southeast Asia. *Br J Haematol.* **124**, 232-239 (2004).

209. Gemignani, F. *et al*. Reliable Detection of β-Thalassemia and G6PD Mutations by a DNA Microarray. *Clin Chem.* **48**, 2051-2054 (2002).

210. Wang, W., Kham, S. K. Y., Yeo, G., Quah, T. & Chong, S. S. Multiplex Minisequencing Screen for Common Southeast Asian and Indian β-Thalassemia Mutations. *Clin Chem.* **49**, 209-218 (2003).

211. Chong, S. S., Boehm, C. D., Higgs, D. R. & Cutting, G. R. Single-tube multiplex-PCR screen for common deletional determinants of alpha -thalassemia. *Blood.* **95**, 360-362 (2000).

212. Tan, A. S. C., Quah, T. C., Low, P. S. & Chong, S. S. A rapid and reliable 7-deletion multiplex polymerase chain reaction assay for alpha -thalassemia. *Blood.* **98**, 250-251 (2001).

213. Zesong, L., Ruijun, G. & Wen, Z. Rapid detection of deletional alpha-thalassemia by an oligonucleotide microarray. *Am J Hematol.* **80**, 306-308 (2005).

214. Bournazos, S. N., Tserga, A., Patrinos, G. P. & Papadakis, M. N. A versatile denaturing HPLC approach for human beta-globin gene mutation screening. *Am J Hematol.* **82**, 168-170 (2007).

215. Colosimo, A. *et al*. Reliability of DHPLC in mutational screening of beta-globin (HBB) alleles. *Hum Mutat.* **19**, 287-295 (2002).

216. 216.Li-Sucholeiki, X., Hu, G., Perls, T., Tomita-Mitchell, A. & Thilly, W. G. Scanning the beta-globin gene for mutations in large populations by denaturing capillary and gel electrophoresis. *Electrophoresis.* **26**, 2531-2538 (2005).

217. Shaji, R. V., Edison, E. S., Poonkuzhali, B., Srivastava, A. & Chandy, M. Rapid Detection of β-Globin Gene Mutations and Polymorphisms by Temporal Temperature Gradient Gel Electrophoresis. *Clin Chem.* **49**, 777-781 (2003).

218. Sanchez, J. A., Pierce, K. E., Rice, J. E. & Wangh, L. J. Linear-After-The-Exponential (LATE)-PCR: An advanced method of asymmetric PCR and its uses in quantitative real-time analysis. *Proc Natl Acad Sci U S A.* **101**, 1933-1938 (2004).

219. Pierce, K. E., Sanchez, J. A., Rice, J. E. & Wangh, L. J. Linear-After-The-Exponential (LATE)-PCR: Primer design criteria for high yields of specific single-stranded DNA and improved real-time detection. *Proc Natl Acad Sci U S A.* **102**, 8609-8614 (2005).

220. Salk, J. J. *et al*. Direct amplification of single-stranded DNA for pyrosequencing using linear-after-the-exponential (LATE)–PCR. *Anal Biochem.* **353**, 124-132 (2006).

221. den Dunnen, J. T. & Antonarakis, S. E. Mutation nomenclature extensions and suggestions to describe complex mutations: A discussion. *Hum Mutat.* **15**, 7-12 (2000).

222. Shchepinov, M. S., Case-Green, S. C. & Southern, E. M. Steric factors influencing hybridisation of nucleic acids to oligonucleotide arrays. *Nucleic Acids Res.* **25**, 1155-1161 (1997).

223. Steinberg, G., Stromsborg, K., Thomas, L., Barker, D. & Zhao, C. Strategies for covalent attachment of DNA to beads. *Biopolymers.* **73**, 597-605 (2004).

224. O'Donnell, M. J., Tang, K., Koster, H., Smith, C. L. & Cantor, C. R. High-Density, Covalent Attachment of DNA to Silicon Wafers for Analysis by MALDI-TOF Mass Spectrometry. *Anal Chem.* **69**, 2438-2443 (1997).

225. Carmon, A. *et al*. Solid-phase PCR in microwells: effects of linker length and composition on tethering, hybridization, and extension. *Biotechniques.* **32**, 410-420 (2002).

226. Longo, M. C., Berninger, M. S. & Hartley, J. L. Use of uracil DNA glycosylase to control carry-over contamination in polymerase chain reactions. *Gene.* **93**, 125-128 (1990).

227. Nilsson, M. *et al*. Making ends meet in genetic analysis using padlock probes. *Hum Mutat.* **19**, 410-415 (2002).

228. Strömqvist Meuzelaar, L., Lancaster, O., Pasche, J. P., Kopal, G. & Brookes, A. J. MegaPlex PCR: a strategy for multiplex amplification. *Nat Methods.* **4**, 835-837 (2007).

229. Daser, A. *et al*. Interrogation of genomes by molecular copy-number counting (MCC). *Nat Methods.* **3**, 447-453 (2006).

230. Hamilton, E. P. *et al*. Use of HAPPY mapping for the higher order assembly of the Tetrahymena genome. *Genomics.* **88**, 443-451 (2006).

231. Krause, J. *et al*. Multiplex amplification of the mammoth mitochondrial genome and the evolution of Elephantidae. *Nature.* **439**, 724-727 (2006).

232. Konfortov, B. A., Bankier, A. T. & Dear, P. H. An efficient method for multi-locus molecular haplotyping. *Nucleic Acids Res.* **35**, e6 (2007).

233. Eichinger, L. *et al*. The genome of the social amoeba Dictyostelium discoideum. *Nature.* **435**, 43-57 (2005).

234. Fredriksson, S. *et al*. Multiplex amplification of all coding sequences within 10 cancer genes by Gene-Collector. *Nucleic Acids Res.* **35**, e47 (2007).

235. Dahl, F. *et al*. Multigene amplification and massively parallel sequencing for cancer mutation discovery. *Proc Natl Acad Sci U S A.* **104**, 9387-9392 (2007).