From CENTER FOR GENOMICS AND
BIOINFORMATICS
Karolinska Institutet, Stockholm, Sweden

TECHNOLOGY DEVELOPMENT FOR GENOME AND POLYMORPHISM ANALYSIS

Magnus Jobs

Stockholm 2003

Published and printed by Karolinska University Press
Box 200, SE-171 77 Stockholm, Sweden

ISBN 91-7349-413-5

Till min bästa vän och hustru Elisabeth…

ABSTRACT

Single nucleotide polymorphism (SNP) is widely expected to provide a multitude of benefits in areas of medical research. There is, however, a growing disparity between the rate of SNP discovery and the development of effective technologies for large scale SNP scoring. This thesis outlines the development of a SNP scoring technology termed dynamic allele-specific hybridization (DASH).

The first publication describes the basic concept of DASH. The strategy involves a short PCR spanning the SNP position, with one PCR primer biotinylated. Via biotin-streptavidine affinity capture the PCR product is bound to the surface of a microtiter plate well. Rinsing in alkali removes the non-biotinylated strand and then hybridization of an allele-specific oligonucleotide probe to the bound target is done. In the presence of a double strand specific fluorescent dye the probe-target duplex is slowly heated whilst fluorescence is monitored. Denaturation of a duplex where the probe is complementary to the SNP allele will cause a drop in fluorescence at a higher temperature than will a probe mismatching at the polymorphic position. Genotypes are obtained by interpretation of the denaturation profiles.

The second publication is an investigation of how truncated oligonucleotide impurities on a surface affect hybridization. It is well known that DNA chips created by photolithography suffer from carrying such impurities, but the effect of these contaminants upon hybridization has not previously been investigated. Impurities were modeled (oligonucleotides of different lengths) in ratios reported from *in situ* array creation by photolithography. Subsequent hybridization experiments where denaturation profiles were generated gave insight in how impurities influence hybridization.

The third publication describes a simple routine for array creation on different surfaces by centrifugation. The centrifugation array concept was later implemented in the second generation of DASH (se below).

The fourth manuscript included in the thesis describes a tool for automated DASH assay design. Secondary structure formation in the hybridization target molecule may compete with the allele-specific probe and can cause an assay to fail. However, by careful assay design it is possible to essentially eliminate secondary structure problems and DASH assay failures Strategies for efficient assay design has been developed and implemented in a DASH-assay design software.

The fifth manuscript describes high throughput implementation of DASH. The new system, which we have termed DASH-2, involves small volume multiplex PCR and multiplex melting temperature analysis on arrays created by centrifugation. A new system of generating hybridization signals, termed iFRET, offers possibilities for multiplexed analysis by separating the assays spectrally. The system is highly flexible (any plate format, PCR multiplexing, serial and parallel array processing, spectral-multiplexing of hybridization probes), thus supporting a wide range of applications, scales and objectives.

LIST OF PUBLICATIONS

I Howell W.M., **Jobs M**., Gyllensten U., Brookes A.J.
Dynamic allele-specific hybridization a new method for scoring single nucleotide polymorphisms
Nature Biotechnology, 1999, 17, 87-88

II **Jobs M***., Fredriksson S*., Brookes A.J., Landegren U.
Effect of oligonucleotide truncation on single-nucleotide distinction by solid phase-hybridization
Analytical Chemistry, 2002, 74, 199-202

III **Jobs M**., Howell W.M., Brookes A.J.
Creating Arrays by Centrifugation
Biotechniques, 2002, 32(6) 1322-1324,1326,1329

IV Fredman D., **Jobs M**., Brookes A.J.
DFold: a Hybridization Assay Design Tool Reducing 2° Structure Formation in Target Molecules
Manuscript

V **Jobs M***., Howell W.M*., Strömqvist L. Brookes A.J.
DASH-2: Flexible, Low-Cost and High-Throughput SNP Genotyping by Dynamic Allele-Specific Hybridization on Membrane Arrays
Manuscript

* These authors contributed equally

CONTENTS

LIST OF ABBREVIATIONS

APEX	Arrayed primer extension
ASH	Allele specific hybridization
ATP	Adenosine triphosphate
bp	Base pair
DASH	Dynamic allele specific hybridization
DNA	Deoxyribonucleic acid
FP	Fluorescence polarisation
FRET	Fluorescence resonance energy transfer
MADGE	Micro plate array diagonal gel electrophoresis
MALDI	Matrix assisted laser desorption/ionisation
MGB	Minor groove binder
MS	Mass spectrometry
NTP	Nucleotide triphosphate
OLA	Oligonucleotide ligation assay
PCR	Polymerase chain reaction
RCA	Rolling circle amplification
RFLP	Restriction fragment length polymorphism
SNP	Single nucleotide polymorphism

1 INTRODUCTION

The science of genetics originated when it was realized that an organism does not pass on a copy of itself to the next generation, but instead passes on information containing material for the creation of a progeny. Gregor Mendel, often referred to as the founder of genetics, did not know anything about the physical or chemical properties of this 'information containing material', but from his studies of pea plants he deduced that the information containing material comprises of distinct units of information, what we now call "genes". He realized that these units are present in pairs in every individual and that variants (alleles) of these units produce variants of the same structure in an organism (e.g. purple or white petals in pea plants). He also realized that only one of these two units is passed on from each parent to the offspring [1].

Today we know that the 'information containing material' is deoxyribonucleic acid (DNA), and that the information is held within the sequential order of the four different nucleotide bases that DNA consists of. We also know that the distinct units of information suggested by Mendel, the genes, are distinct parts of the DNA sequence. It has been revealed how DNA is stored and replicated in the cells and how it is passed on to the next generation. Moreover, the full DNA sequence of several organisms is also known, amongst which lies the human genome [2] [3].

1.1 CLUES IN GENETIC VARIATION

Mendel took an approach where he used phenotypic variation to understand how information is passed on from one generation to the next. He was first in using the clues inherent in genetic variation. We now know that genetic variation holds clues that can lead us to understand mechanisms for genetic disease and other heritable phenotypes.

Studies of the human genome have revealed a high degree of sequence similarity between unrelated individuals. On average, two chromosomes differ at 1/1250 base pairs (bp). However, the genome is very large (~3.3 billion bp in a haploid genome) so in spite of the high degree of similarity the actual number of positions in the genome that vary between individuals is high. Any two randomly chosen chromosomes from the human population have more than 2.5 million variant sites [4]. Some of the variants in a given individual are so called rare variants, meaning that the allele frequency of the most common allele is more than 99%. Polymorphisms are variants that coexist in the population at a frequency of less than 99% for the most common allele [5]. A large fraction of the variants that can be found in the genome do not have any known phenotypic effect while others, in concert with each other and with environmental factors, are responsible for phenotypic variability.

The most abundant form of sequence variation, the single nucleotide polymorphism (SNP), has received much attention in the scientific community over recent years. SNPs have an average frequency of about 3 per 1000 base pairs when looking at the population as a whole [6]. Because of the abundance in the genome they are thought to hold potential for successful genetic studies.

1.2 TECHNOLOGY DEVELOPMENT

When Mendel took the first steps towards modern genetics the only instrument available to him was his pea plant garden. In the 150 years that have lapsed since then we have witnessed incredible advancement in all areas of natural science, and the techniques available today for studying genetics combine the latest advancements from many of the disciplines. However, there are still major requirements on new and better technology. Much of the rapid success in the field of genomics, and indeed the sequencing of the Human Genome Project, was dependent upon technological developments. The human genome project was initiated 1990 and was planned to last for 15 years [7]. One of the projects major goals was to

reveal the complete base pair sequence of the human genome. This part of the project initially proceeded slowly, and for a long while the goal of sequencing the human genome in 15 years seemed unrealistic. However, technological advancements in the form of new sequencing technology (capillary sequencing) and software for assembly and annotation of the generated sequence data suddenly boosted the slowly moving work [8]. On the 26th of June 2000, ahead of schedule, the former US president Bill Clinton and The U.K prime minister Tony Blair announced the completion of the Human Genome Sequence, explaining that mankind's complete genetic heritage had been detailed at base-pair level [2 3].

1.2.1 Extracting information from genetic variation

The sequence of the human genome was made publicly available as the work progressed. In parallel, information about normal variation within the sequence has continuously grown in different public and commercial databases [9 10]. There are now more than 3,000,000 SNP s listed in public databases. Through these resources ther exist great potential for association and linkage-based studies of human disease, population and evolutionary genetics, forensics, clinical diagnostics, and pharmacogenetics at an unprecedented scale.

The clues for understanding early human development and the basis for our present day disease and phenotypic spectra are inherent in SNPs. But extracting the meaningful information from the intrinsic information noise that events in human history have built into the equation is a difficult task [11 12], especially given the preponderance of non-mendelian traits. A non-mendelian (complex) trait results from multiple genes acting in concert with each other and with environmental factors. Identifying loci (positions in the genome) involved in complex traits requires investigation of many polymorphisms in many genes. To deliver the statistical power needed for unambiguous loci detection, large sets of DNA samples from well-defined populations with well-defined

phenotypes of interest are required. With large numbers of SNPs and individuals in the genetic studies, effective SNP genotyping is of absolute necessity. The effective SNP genotyping has turned out to be a major challenge in it self, occupying much research attention [13] [14]. To meet this challenge, a quite substantial list of problems would have to be solved. Properties of the ideal genotyping method would include:

- Robust genotyping reaction even on low quality DNA
- High specificity
- Straight-forward SNP-specific assay design
- Minimal assay optimization
- Easy to automate
- Reliable automated genotype calling (from generated data)
- A scalable system for low to high throughput performance
- Low cost per genotype (reagents, disposables, instrument maintenance and labor included)

A number of strategies for effective analysis have been brought forward to meet the required properties of the ideal genotyping method. Several methods offer satisfactory solutions to parts of the list, but unfortunately while solving one issue another often has to be set aside.

At this point in the in the large project of interpreting the variations in the code that are behind many phenotypic differences, another technological boost, similar to the one that enabled the sequencing of the genetic code, is needed. In the following text I will present the different basic concepts for SNP allelic discrimination that are in use today. I will give examples on how detection of alleles, based on these basic concepts, can be achieved. I will also discuss array technologies and the important benefits offered by having samples laid out in a defined pattern on a surface. Finally, I will discuss the philosophy behind DASH (dynamic allele-specific hybridization) [15] [16], the

genotyping technology that we have created in an attempt to meet the technological challenge of effective SNP genotyping.

2 ALLELE DISCRIMINATION STRATEGIES

Basically all strategies for allele discrimination involve DNA hybridization in one way or another. An exception to this is restriction fragment length polymorphism (RFLP) analysis. RFLP analysis is a classic method for polymorphism analysis relying on DNA sequence specific restriction enzymes for identification and cleavage of the DNA sequence of interest. Alleles are discriminated through allele-specific cleavage [17].

The strategies for allele discrimination I will discuss here are allele-specific hybridization, primer extension, allele-specific ligation, and allele-specific invasive cleavage [18]. All modern SNP analysis methods are based on one or more of these four strategies. With a few exceptions they are more or less dependent on polymerase chain reaction (PCR) prior to (or in some cases during) implementation of the genotyping strategy.

2.1 HYBRIDIZATION

Shortly after the breakthrough discovery of the double helix by Watson and Crick [19] it was shown that the two strands could be separated by heat or alkali. The reversed process, renaturation was also described [20] and it was established that the two strands involved in duplex formation must be complementary to some extent, and that the duplex stability is in direct relation to the degree of complementarity between the strands. Methods for analyzing DNA based on these properties soon emerged [21] [22] [23].

Allele-specific hybridization is the simplest form of the allele discriminatory strategies. It is free from enzymatic steps, which offers the potential for inexpensive genotyping systems. The strategy takes advantage of the difference in DNA duplex stability that a fully matched duplex or a duplex with mismatched base pairs have. A synthetic DNA strand (probe) fully matching one of the alleles is mixed with the sample DNA (target) with the unknown genotype. Under appropriately specific conditions (i.e. temperatures) the

oligonucleotide probe will form a duplex with the fully matching target, while it will not form a duplex with the mismatching target [24] [25] [26].

When genotyping SNPs the allelic stability differences arise from only one base-pair matching or mismatching in the probe-target duplex. However, with a correctly designed PCR and hybridization probe the difference between match and mismatch can still be significant [27].

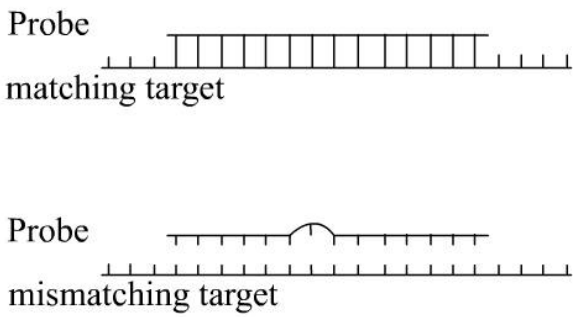

Figure 1 SNP allele discrimination based on allele specific hybridization relies on the duplex stability difference between a fully matched and a single base mismatch probe-target duplex.

2.1.1 Hybridization based methods

There are a number of SNP genotyping methods based on allele-specific hybridization. The different methods are separated by how hybridization is detected and reaction format. Detection may involve radioactive labels, fluorescent labels, intercalating fluorescent dyes or electrical conductivity. The reaction format may be in solution or solid phase.

2.1.1.1 Dot-blots and micro arrays

An early technique used for SNP genotyping based on allele-specific hybridization is the dot-blot technique [28] [26]. The procedure of dot-blotting involves taking an aqueous solution of target DNA (i.e. PCR fragments amplified over the SNP position) and simply spotting it on to a nitrocellulose or nylon membrane and then allowing it to dry. The target DNA is denatured either by exposing it to heat or alkali. The membrane has to be blocked to avoid further capture of DNA and then the allele-specific probe, typically 13-19 nucleotides long, can be hybridized to the bound targets. Hybridization takes place under conditions at which the DNA duplex between probe and target is stable only if there is a perfect base complementarity between them. A single mismatch at the SNP position will prevent the probe from hybridizing to the

target. The probe is either carrying a radioactive label that can be detected by autoradiography [28],or a fluorescent label that can be detected by fluorescence imaging [29]. The appearance of a spot on such an image indicates the presence of the probe-matching allele.

In a reverse dot-blot [30] unlabeled allele-specific probes are bound to the membrane in different array features and labeled targets are provided in solution. Hybridization to a specific probe on the membrane indicates the presence of that specific sequence in the target DNA.

The micro-array technologies are closely related to the dot-blot method. The difference lies in the array feature density and the more advanced equipment needed for analysis. Analyzing many different SNP positions in one array (reverse dot-blot equivalent) means designing all of the probes for the individual hybridization assays to be allele discriminatory at a common stringency[31 32] . This is almost impossible since hybridization stability between probe and target depends on the sequence context. Designing all the probes for the different assays to be of the same length does not mean that they will share the same denaturation temperature. Moreover, different DNA bases mismatch each other with different stability [33 34 35 36], which makes it even harder to design an array where only the matched targets hybridizes at a single hybridization temperature. However, it is possible to have several array features with probes of different lengths for each of the interrogated targets. With such sets of tiling probes for each of the targets, it is likely that at least one probe length in each set is allele discriminatory at the chosen hybridization temperature. [37 38 39 40].

2.1.1.2 DASH

Dynamic allele-specific hybridization (DASH), the central method to this thesis, is also based on the hybridization strategy [15 16]. DASH will be discussed further in chapter 4.

2.1.1.3 TaqMan assay

The TaqMan assay exploits the 5’ exonuclease activity of Taq DNA polymerase to degrade an allele-specific probe hybridized during PCR [41 42].

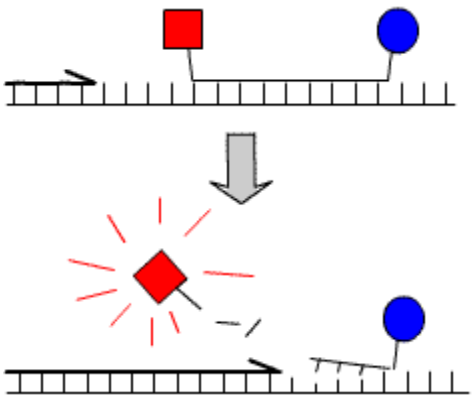

Figure 2 The TaqMan probe only hybridizes to the fully matching allele. The exonuclease activity of the DNA polymerase fragments the probe and the reporter molecule is released. Increase in fluorescence indicates the presence of the matching allele.

The probe is designed to hybridize, at PCR annealing temperature, only when perfect complementarity between the target and the probe exists. Thus, degradation of the probe can only occur when the probe-matching allele is present in the template DNA. The TaqMan-probe carries two fluorescent labels that interact with each other via fluorescence resonance energy transfer (FRET). One of the molecules, the reporter, is excited by UV light. As long as the TaqMan probe is intact the other molecule, the quencher, remains in the reporter's proximity, and the energy absorbed by the reporter is transferred to the quencher. If the probe hybridizes to the target the PCR taq polymerase degrades the probe and the reporter escapes the proximity of the quencher. When the reporter is no longer in the quenchers proximity it will fluoresce. Increasing fluorescence from the reporter, as thermo-cycling (real-time PCR) proceeds, indicates the presence of the matching allele.

Probe cleavage can also be detected by fluorescence polarization (FP) [43]. When fluorescent molecules are excited with plane polarized light, they emit light in the same polarized plane, provided that the molecule remains stationary throughout the excited state. However, if the excited molecule rotates or tumbles out of the plane of polarized light during the excited state, light is emitted in a different plane from that of the initial excitation. If vertically polarized light is used to excite the fluorophore, the emission light intensity can be monitored in both the original vertical plane and also the horizontal plane. The degree to which the emission intensity moves from the vertical to horizontal plane is related to the mobility of the fluorescently labeled molecule. If fluorescently labeled molecules are very large, they move very little during the excited state interval, and the emitted light remains highly polarized with respect to the excitation plane. If fluorescently labeled molecules are small, they rotate or tumble faster, and the resulting emitted light is depolarized

relative to the excitation plane. Thus, when the TaqMan probe is cleaved by the polymerase it becomes smaller and will tumble faster. The FP value is altered.

The TaqMan assay is attractive because of its closed-tube homogenous format [41]. Once PCR has been set up no further sample processing is required. However, allele discrimination is solely dependent on that the mismatching probe does not hybridize to the target at PCR annealing temperature while the matching probe does. So, for unambiguous SNP genotyping both allele specific probes have to be used. The reporter dyes on the two probes are well separated in emission spectra and can thus be detected separately. Moreover, in order to increase the Tm difference between the matching and mismatching probe-target duplexes minor grove binding (MGB) proteins are used . The MGBs stabilizes the fully complementary duplexes but have less effect on duplexes mismatching at the SNP position [44]. The success rate of designing functioning TaqMan assays has been reported to be limited and the probes are quite expensive. The MGBs also increase the cost.

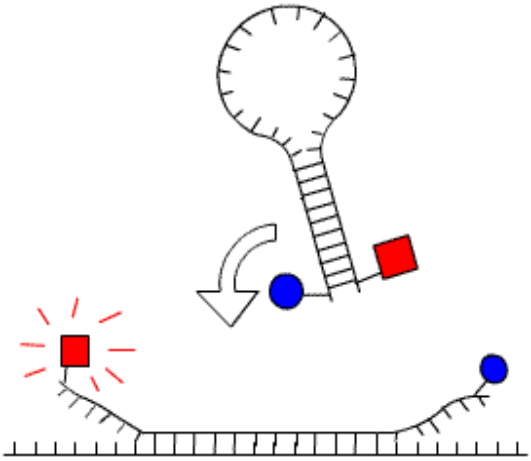

Figure 3 The molecular beacon will hybridize only to the fully complementary SNP allele. Hybridization separates the quencher and reporter molecules and the reporter can emit light.

2.1.1.4 Molecular beacons

Another method for SNP genotyping based on hybridization utilizes the molecular beacon [45 46 47 48]. The molecular beacon is a stem-loop oligonucleotide probe with a FRET pair attached to the stem in such a way that the quencher and the reporter only come close to each other when the stem loop structure is formed. The loop-part of the molecular beacon is complementary to one of the SNP alleles and surrounding sequence. When a sample containing the complementary allele is amplified in PCR the molecular beacon can hybridize and disrupt the stem structure of the molecular beacon. This separates the reporter and the quencher, thus the reporter can give of its energy in the form of light. In a real-time PCR the signals are detected as the PCR advances. The molecular beacon can also be used to detect hybridization in post PCR hybridization [49].

The allele-discrimination method based on the molecular beacons can be set up in a closed-tube system, and is thus attractive for automation reasons. Using fluorescent reporters with different emission spectra opens up some possibilities for multiplexing. However, like the TaqMan assay the molecular beacon system suffers from difficult assay design with limited assay success rate. The oligonucleotides also have to be doubly labeled which makes it expensive.

2.2 PRIMER EXTENSION

The primer extension strategy for genotyping makes use of DNA polymerase for the incorporation of deoxyribonucleotides complementary to the DNA template sequence of interest. Primer extension genotyping can be accomplished by utilizing the fact that DNA polymerase is incapable of extending a priming sequence that, in its 3' end, is not perfectly complementary to the template sequence [50].

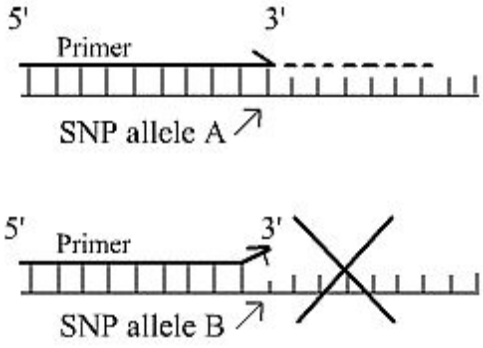

Figure 4 Allele specific primer extension rely on the inability of DNA polymerase to extend a primer that is not complementary to the template sequence in its 3' end.

By designing the oligonucleotide primer so that the 3' end overlaps the SNP position the primer will be extended only when the complementary SNP allele is present. Another approach is to design the oligonucleotide primer to hybridize a single base upstream of the SNP position and then detect the type of nucleotide the primer is extended with [51]. The dideoxyterminators used in the reaction mix prevents the polymerase from elongating by more than one base.

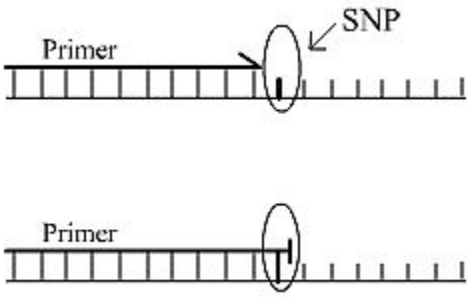

Figure 5 In single base primer extension based methods the DNA polymerase extend the primer with the complementary dideoxy terminator

2.2.1 Primer extension based methods

There are a wide range of SNP genotyping methods based on the primer extension strategy. Systems for detection of primer extension show a wealth of diversity. Gel electrophoresis, FRET, fluorescence labeled nucleotides, fluorescence polarisation (FP), mass spectrometry (MS) and detection of pyrophosphate formation are some of the detection strategies that have been used. Primer extension strategies for SNP analysis have much to offer SNP genotyping in terms of cost and specificity, but are not free from problems.

2.2.1.1 Allele-specific PCR

One of the classic genotyping methods is allele-specific PCR. By designing one of the two PCR primers to overlap the SNP position in its 3' end, PCR can only occur if the if the complementary allele is present in the DNA sample being investigated [52]. The strategy involves PCR primer design for allele-specific amplification of both alleles. By running two separate PCRs on the investigated DNA sample with the two sets of PCR primers, and then looking for amplicons by gel electrophoresis, it is possible to call the genotype. The system is impractical for larger studies because of the two separate PCRs per genotype, the detection by gel electrophoresis and the difficulties of controlling for false negative results.

A number of innovations coupled to allele-specific PCR have increased the fidelity of the method. Real time fluorescence detection of intercalating dye to monitor the formation of PCR products removes the gel electrophoresis step from the procedure [53]. Adding a GC-rich tail to one of the allele-specific primers and, at the end of the thermo cycling, monitoring heat denaturation of the products formed, makes it possible to perform the genotyping in a single closed tube. A high temperature denaturation profile indicates the presence of the allele that was amplified by the GC-tailed primer, and a low temperature profile indicates presence of the allele amplified by the regular primer. Additional ways to detect formation of PCR product in real time have been developed, as well as polymerases with better specificity in order to avoid false positives in the allele-specific PCR.

Allele-specific PCR is attractive for cost reasons, but a remaining problem with allele-specific PCR is the limited flexibility in PCR design. The allele recognizing primer must always be designed to overlap the polymorphic site in its 3' end, which may make it difficult to design assays for SNPs in problematic sequence contexts.

2.2.1.2 Allele specific primer extension

The allele-specific primer extension strategy used in allele-specific PCR can also be used in single base extension assays. Here the target sequence is amplified in regular (not allele-specific) PCR. Then to decide presence or absence of SNP-alleles an allele-specific primer is used on the PCR product in a single base extension reaction [54]. Radioactively or fluorescent labeled dideoxyterminators are used to detect primer extension.

A problem for primer extension performed on solid support is that the PCR product serving as template in the primer extension reaction has to be concentrated and rendered single-stranded before hybridizing to the solid support bound primers. If this is not done, too few targets hybridize and the primer extension with fluorescent labeled nucleotides cannot be detected. A recent protocol for solid support allele-specific primer extension overcomes this problem. The new protocol makes use of T7 RNA polymerase to generate large amounts of hybridization targets. By tailing one of the PCR primers with aT7 RNA polymerase promoter sequence, RNA templates can be generated from the PCR product. The RNA templates are allowed to hybridize to the allele-specific solid support bound primers and reverse transcriptase extends the primers if the 3' end matches the template. Dye-labeled rNTPs are used in the extension reaction, and as opposed to single base extension, a long stretch of labeled nucleotides is incorporated. In this way, strong signals can be produced from small amounts of product [50].

The drawback of this method is the complex protocol. Several enzymes and the special rNTPs add to cost and complexity.

2.2.1.3 Single base primer extension

The previously described allele-specific primer extension strategy gives a yes or no answer to whether or not an allele is present in a sample. Either the primer is extended or it is not. In the single base primer extension strategy the primer is always extended (if there is a template present). The primer is placed with its 3' end immediately upstream to the SNP position and the polymerase extends the primer over the SNP position with a complementary dideoxyterminator [51]. The four different dideoxyterminators labeled with different dyes of different fluorescence emission spectra makes it possible to deduce the genotype of the sample by analysing the spectra of the emitted light from the incorporated dideoxyterminator [55].

Single base extension is used in a wide variety of systems. In the arrayed primer extension (APEX) approach the single base extension primers are attached to a surface via their 5' end [56]. The primers are then allowed to hybridize to PCR products amplified over the SNP positions. In presence of the four differently labeled dideoxyterminators DNA polymerase extend the primers. Color analysis of the emitted light from the spots on the array indicates the genotypes [55].

As mentioned in the previous section, this approach requires high concentrations of single-stranded PCR products to be robust. The generation of single-stranded PCR product is an extra procedural step with its own required reagents that adds to cost.

Single base primer extension can be performed in solution without the use of costly FRET-pairs (donor and acceptor dyes) by monitoring fluorescence polarisation (FP) [57]. When the dye-labeled dideoxyterminators are incorporated onto a primer the molecular weight of the formed molecule is much greater than the molecular weight of just the terminator-dye molecule. The FP value decreases dramatically as the deoxyterminators become incorporated. The assay cost for this system is small and for that reason attractive.

Single base primer extension can also be performed in solution. A FRET based system can detect incorporation of dideoxyterminators by having the

FRET donor attached to the primer and the acceptor attached to the didexyterminator [58]. The acceptor dye will quench the donor dye as soon as it is incorporated with the primer and the acceptor dye starts to emit light. Again, different emission spectra of the acceptor dyes make it possible to distinguish incorporated dideoxyterminators for genotype calling.

It is possible to perform multiplexed primer extension reactions in solution and separate the assays by assay specific capturing, to a solid support, of the extension products. This is done by utilizing primers in the primer extension reaction that in the 3' region are complementary to the SNP loci, and in the 5' region are complementary to specific capture probes on the solid support. The solid support can be an array [59] or color-coded microspheres [60] [61]. The emission spectra of the incorporated dye-labeled terminator reveal the genotype, whereas the specific capture probe sequence reveals the identity of the assay. Color-coded beads can be sorted by flow cytometry.

For any in-solution based single base primer extension approach it is necessary to degrade excess of primers and dNTPs after the PCR, otherwise these will interfere with the single base extension reaction. The enzyme Exonuclease I recognizes single stranded DNA for degradation and is used for purification of the reaction from excess of single stranded primers. The enzyme shrimp alkaline phosphatase degrades excess of dNTPs. The enzymes used to purify the reaction mix are of course also harm-full for the primer extension reaction and have to be inactivated before the primer extension reaction can be performed.

2.2.1.4 MALDI-TOF MS

Mass spectrometry (MS) has been used to resolve both single base primer extension and allele-specific primer extension products. The matrix assisted laser desorption/ionisation (MALDI) [62] is a method to ionize macromolecules (such as DNA or proteins) without breaking the molecules apart. The ionized molecules can then be weighted in time of flight mass spectrometry. The matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF MS) has successfully been used in genotyping experiments. The small difference in mass between the two allelic extension products is enough

to be distinguished. It is also possible to perform multiplexed analysis by designing the primers for the different targets with different lengths so that the different assays are well separated in mass [63] [64].

An attractive feature of using MS to detect primer extension is the complete elimination of fluorescence labels in the assay. Increasing the capacity of the MS instrument, and optimization of the necessary purification steps involved will make the system even more attractive.

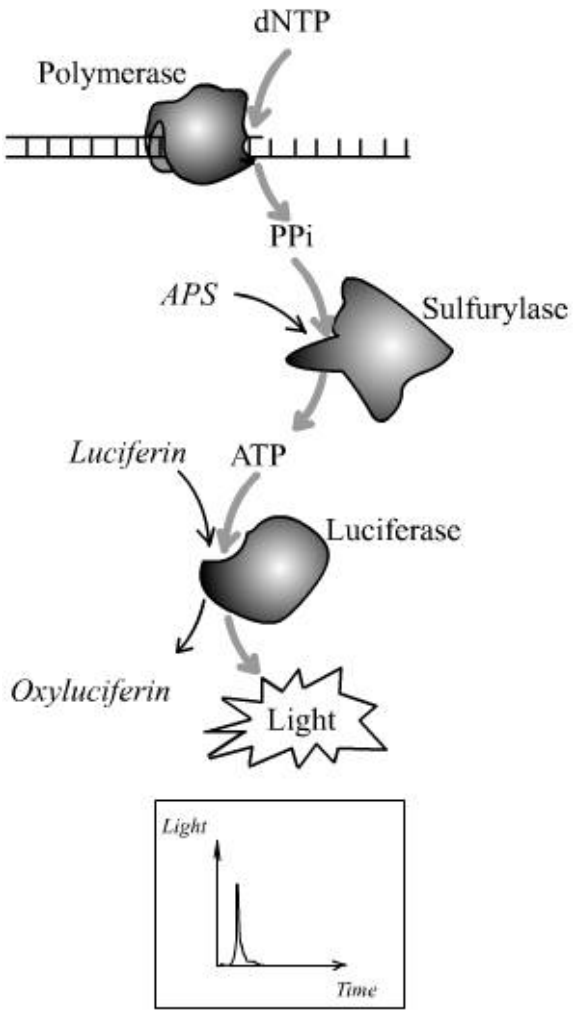

Figure 6 In the pyrosequencing reaction polymerase incorporates the deoxy nucleoside triphosphate.(dNTP) complementary to the template. Pyrophsphate (PPi) is released in the reaction. APS and PPi are then turned into ATP, which is used by luciferase to produce light. By adding the different dNTPs, one at a time, and monitor light it is possible to sequence the template

2.2.1.5 Pyrosequencing

Pyrosequencing belongs to the single base extension category, even though the method is actually able to reveal the identity of more than just a single base downstream the primer as the reaction proceeds [65] [67] [68]. The technique is based polymerase primer extension. Detection of primer extension is based on the released pyrophosphate from the primer extension reaction. When DNA polymerase incorporates a deoxynucleoside triphosphate onto the growing strand, pyrophosphate is formed. Enzymes present in the reaction mix convert the pyrophosphate into ATP. Luciferase (another enzyme in the reaction mix) use the formed ATP to produce light, and the emitted light is taken as evidence for incorporation of a nucleotide. By adding the different nucleotides stepwise and monitoring light it is possible to sequence up to 100 bases [69].

The architecture of pyrosequencing is highly complex. In the homogenous assay format seven different enzymes are required. It is also necessary to remove excess of nucleotides and primers form the PCR. Between each addition of nucleotides, unincorporated nucleotides must be degraded. These factors make pyrosequencing a method difficult to bring down to the low costs necessary for large-scale genotyping studies.

2.3 LIGATION

DNA ligase catalyzes the covalent joining of two segments to one uninterrupted strand in a DNA duplex, provided that no nucleotides are missing at the junction. The repair (ligation) mechanism of the enzyme is highly specific and takes place if the DNA bases on both sides of the break point (the nick) are complementary to the hybridized DNA strand [70]. SNP genotyping methods based on DNA nick ligation makes use of the high specificity of the DNA ligase to ligate only when the bases on either side of the nick are complementary to the interrogated target strand. A single base mismatch next to the nick prevents ligation. Two oligonucleotides, of which one is allele-specific, are allowed to hybridize to the target strand of interest, and if the allele-specific oligonucleotide is complementary to the SNP allele, ligation of the two oligonucleotides will occur. Detection of the ligation product is taken as evidence for the presence of the matching allele in the interrogated sample.

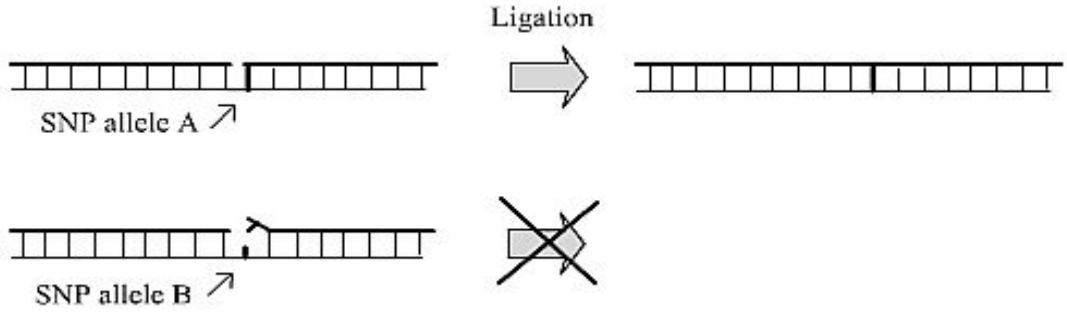

Figure 7 DNA nick ligation can only occur if the strand with the nick is perfectly complementary, on both sides of the nick, to the hybridized strand. This mechanism can be used in SNP genotyping. Two oligonucleotides, of which one is allele-specific, are allowed to hybridize to the interrogated target strand. If the allele matching the allele-specific oligonucleotide is present in the sample the two oligonucleotides will be ligated when treated by DNA ligase.

2.3.1 Ligation based methods

2.3.1.1 Oligonucleotide ligation assay (OLA)

The original format of ligation-based genotyping, the oligonucleotide ligation assay (OLA), involves a 5' biotin labeled oligonucleotide and two differently fluorescence labeled allele-specific probes [71]. After interrogation of the sample sequence the ligation products are affinity captured to a streptavidin coated solid support and washed to remove non-ligated probes. By monitoring the fluorescence from the solid support bound ligation products it is possible to

decide which of the allele-specific probes that ligated, thus the genotype of the sample. It is of course also possible to detect ligation by gel electrophoresis separation. The ligated products are significantly larger that the non-ligated and genotypes are inferred by band patterns on electrophoresis gels.

2.3.1.2 Ligation and rolling circle amplification (RCA)

Circular DNA can be used as a template for producing many copies of the circular DNA sequence repeated in one long strand. This is accomplished by using a primer, complementary to one part of the circular DNA, as the starting point for DNA polymerase extension. The DNA polymerase must be of a type that does not exhibit exonuclease activity but is capable of strand displacement. The technique, termed rolling circle amplification (RCA) [72 73], can be used to detect a ligation event at a SNP position directly on genomic DNA without PCR amplification [74 75].

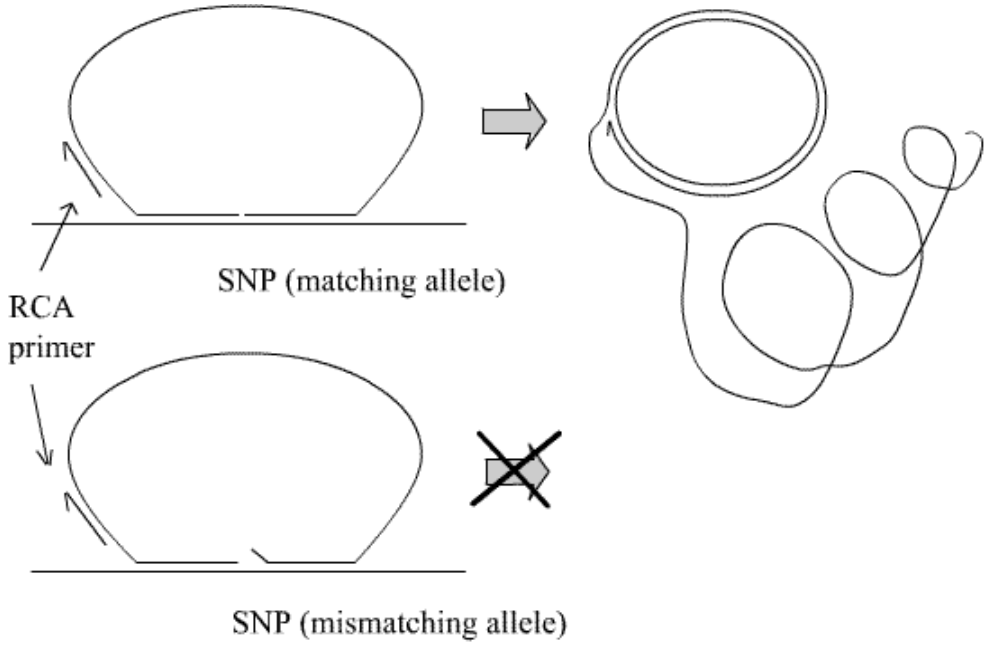

Figure 8 The ~90 bases long probe, termed padlock probe, hybridizes with its ends on both sides of the SNP position with one end overlapping the polymorphic site. When the overlapping end is complementary to the SNP the small nick between the ends can be ligated by DNA ligase and the probe becomes circularized. DNA polymerase then synthesizes a long strand with the circularized probe for template (rolling circle anplification (RCA)). If the mismatching allele is present in the interrogated sample ligation will not occur.

By having the two ends of an oligonucleotide probe (~90 bases) hybridize on opposite sides of an SNP position with one end allele-specifically overlapping the SNP, treatment with DNA ligase will form circular single stranded DNA only

when the matching allele is present. Running RCA reveals whether or not ligation took place.

Additional primers, called branching primers, complementary to the repetitive sequence produced during the primary RCA will boost the amplification significantly. The primers hybridize to each of the sequence repetitions and prime DNA synthesis. As these multiply hybridized primers elongate, strand-displacing activity of the polymerase displaces the growing strands and a multitude of new recognition sites for the first RCA primer are exposed. With the 5' end of the branching primers tagged with a FRET-pair labeled stem loop structure (similarly to the previously discussed molecular beacon (2.1.1.4)) it is possible to view amplification as increased fluorescence from the donor molecule in the FRET pair. As the complementary strand to the stem loop primer is synthesized the polymerase reads through the stem loop and opens it up. This shifts the FRET pair molecules out of each other's proximity and quenching is ceased [76]. Amplification is taken as evidence for the presence of the matching allele.

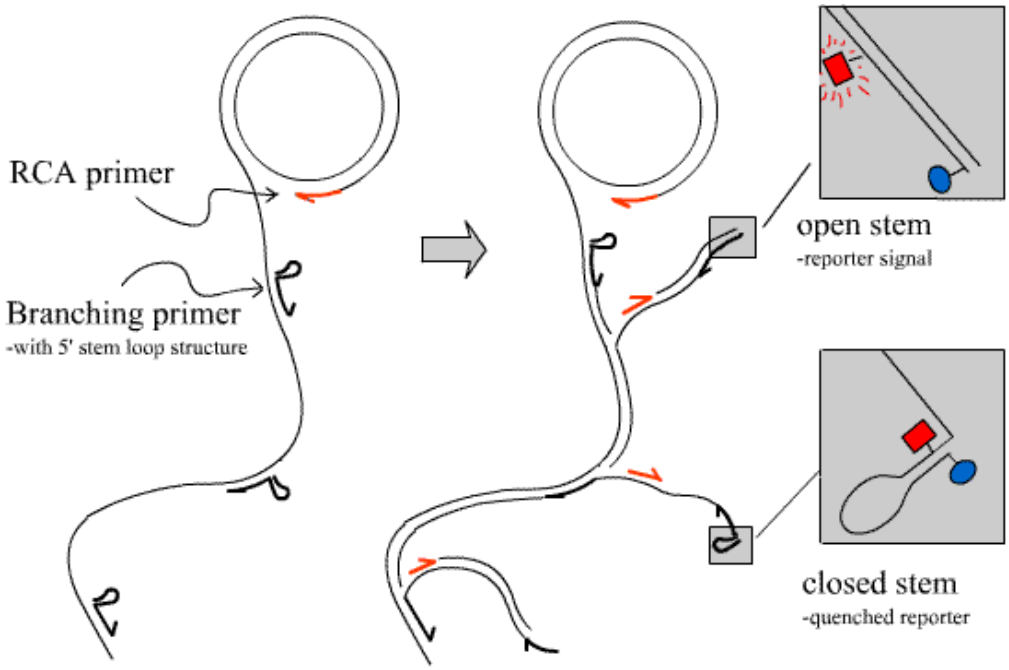

Figure 9 Primers complementary to the repetitive RCA produced sequence hybridize and prime DNA synthesis. As these primers elongate, strand-displacing activity of the polymerase displaces the growing strands and a multitude of new recognition sites for the first RCA primer are exposed for priming and polymerization. The branching primer contains a 5' stem loop structure labeled with a FRET-pair. When the stem loop structure is opened up by the displacing activity of the DNA polymerase the FRET-pair is separated and a fluorescent signal emitted.

This technique involves very long oligonucleotides, unlabeled as well as doubly labeled, and non-standard enzymes are used. Nevertheless, it is attractive with a method able to retrieve genotypes directly from genomic DNA eliminating the otherwise more or less mandatory PCR. This method has great potential for high throughput genotyping.

2.3.1.3 FRET detection of ligation

An oligonucleotide ligation event can be detected by FRET. Having one of the two oligonucleotides labeled with a donor dye and the other with an acceptor dye will cause FRET (the acceptor dye will fluoresce as the donor molecule is excited) between the dyes if the oligonucleotides are ligated. A closed-tube in-solution SNP analysis method is built on this fundamental principle [77]. In this system the ligation oligonucleotides have been designed to hybridize at a significantly lower temperature than the annealing temperature of the PCR primers. In this way the PCR can be performed with the ligation oligonucleotides and the DNA ligase enzyme present in the reaction mix without the risk that the ligation oligonucleotides will interfere with the PCR priming. When thermal cycling is near completion the thermo cycling conditions are altered so that the ligation oligonucleotides can hybridize to the newly-produced PCR product and allele specific ligation can take place. FRET is monitored in real time, and genotypes can be determined through the light intensity emitted from the two differently labeled allele-specific ligation probes.

Ligation is specific and in conjunction with PCR an additional level of specificity is added to the genotype call. The drawback is the cost of all the labeled probes and the ligase.

2.4 DNA CLEAVAGE

DNA cleavage has long been used in polymorphism analysis. The first generation of DNA markers was restriction fragment length polymorphisms (RFLPs). Genotyping these markers involved cleavage of genomic DNA with restriction endonucleases, separation of the cleaved DNA by gel-

electrophoresis and hybridization of radioactively labeled probes in a southern blot. This technique could be used in crude linkage studies but required plenty of time, money and DNA.

In conjunction with PCR, it is possible to make polymorphism analysis based on DNA cleavage more efficient and precise.

2.4.1 DNA cleavage based methods

2.4.1.1 RFLP genotyping of SNPs

There are many restriction enzymes that recognize different sequences for cleavage. RFLP genotyping utilizes the restriction enzyme specificity in cleaving certain DNA sequences. The trick is to find enzymes cleaving one of the alleles but not the other, and then separating the cleavage products by electrophoresis. If such an enzyme cannot be found it is still possible to create an allele-specific cleavage site close to the SNP position by altering the sequence of a PCR primer close to the polymorphic position. There are many impracticalities related to this technique. Purification of the PCR product prior to cleavage and then the cleavage reaction it-self are time-consuming steps. Gel-electrophoresis separation of the cleavage products is another procedure that makes this technique less suitable for high throughput applications. However, a few fairly recent attempts to bring throughput and economy into this concept have been made. The detection of cleavage products is one of the obstacles, and the classic way of running the enzyme-treated PCR products on electrophoresis gels stained in ethidiumbromide is slow and inefficient. Modern electrophoresis techniques such as microplate array diagonal gel electrophoresis (MADGE) [78] and capillary electrophoresis [79] improve this somewhat.

2.4.1.2 McSNP

McSNP combines RFLP with melting curve analysis. The technique involves allele-specific restriction enzyme cleavage of non-purified PCR products. After cleavage the enzyme-treated products are analyzed by slowly heating them in the presence of the double-strand DNA-specific fluorescent dye, SYBR Green

I. As the samples are heated, fluorescence rapidly decreases when the melting temperatures of different fragments are reached. In this way it is possible to determine the composition of simple mixtures of DNA fragments in the analyzed sample. Cleaved products will cause denaturation at low temperature while non-cleaved products will cause high temperature denaturation [80].

Impracticalities and high costs are inherently high when restriction endonucleases are used in SNP analysis. Restriction enzymes are often expensive and digestion of unpurified PCR product requires long incubation times (up to 24h).

2.4.1.3 Invasive cleavage (Invader assay)

The invasive cleavage strategy exploits a thermo stable structure-specific flap endonuclease to cleave the DNA at specific sites based on structure rather than sequence. Even though the enzyme is not sequence specific it can be guided to a specific sequence. The enzyme is used in conjunction with structure-forming probes for the sequence of interest so that the enzyme cleaves in a structure, and target, specific manner.

The invader assay uses two target-specific oligonucleotides to form the structure recognized by the enzyme [81]. The structure is formed when an upstream invader oligonucleotide and a downstream signal probe hybridize on either side of the SNP position, both allele-specifically overlapping the SNP. The 3' end of the invader oligonucleotide overlaps the SNP with one base. The 5' region of the signal probe overlaps the SNP plus is tailed by unpaired bases. When the invader oligonucleotide and the signal probe match the SNP allele the flap endonuclease cleaves the signal probe at the overlapping position and the 5' tail of the signal probe is released. If the probes mismatch the SNP the enzyme is prevented from cleaving the signal probe.

The detectable signal is generated in a second cleavage event. The 5' tail released in the first reaction participate in a second invasive cleavage reaction. The released tail functions as an invader oligonucleotide that hybridizes to a supplied FRET probe. The FRET probe is labeled with a FRET-pair that is separated when the probe is cut. The separation of the donor and acceptor

dyes prevents energy transfer between the dyes and the donor dye starts to emit light. Enhanced fluorescence signal from the donor is detected and taken as evidence for the presence of the matching allele.

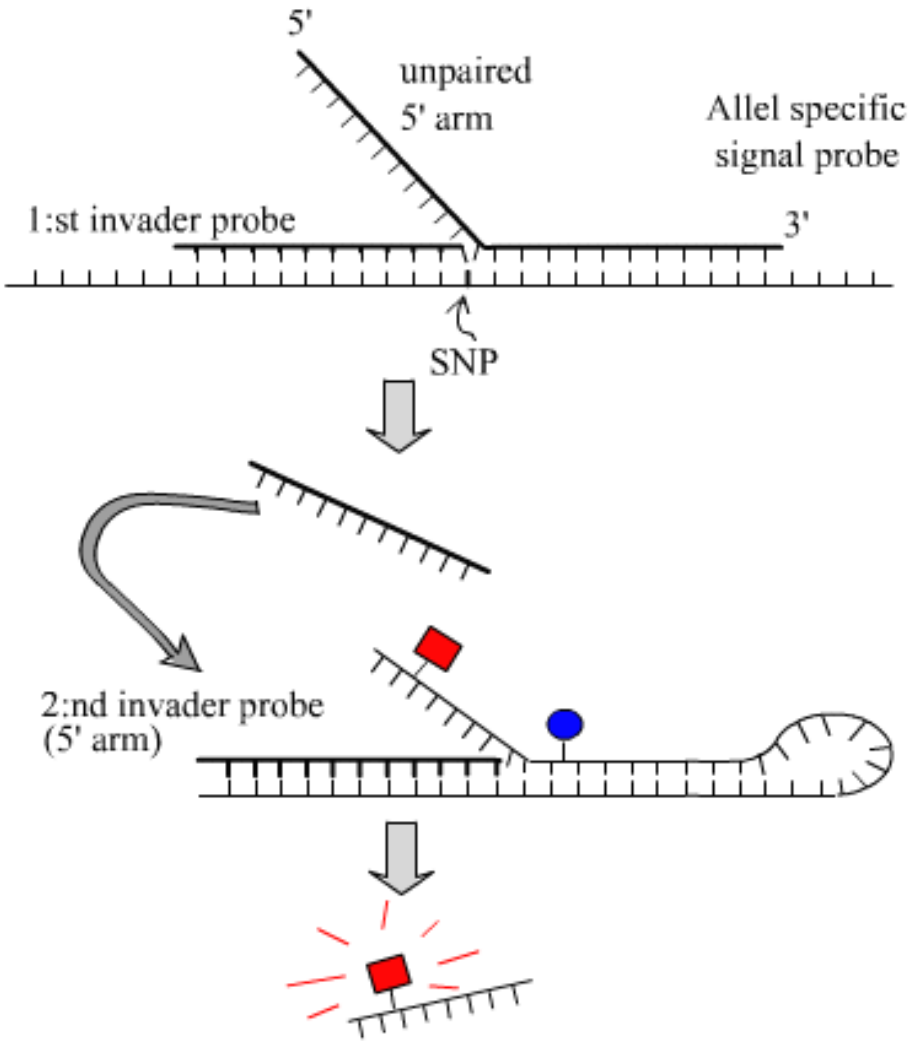

Figure 10 The structure recognized for cleavage by the flap endonuclease enzyme is formed when an upstream invader oligonucleotide and a downstream signal probe hybridize on either side of the SNP position, both allele-specifically overlapping the SNP. The 3' end of the invader oligonucleotide overlaps the SNP. The 5' region of the signal probe overlaps the SNP, and is also tailed by unpaired bases. When the invader oligonucleotide and the signal probe match the SNP allele the clevase cut the signal probe at the overlapping position and the 5' tail of the signal probe is released. If the probes mismatch the SNP the enzyme does not cleave the signal probe. The detectable signal is generated in a second cleavage event. The 5' tail released in the first reaction participates in a second invasive cleavage reaction. The released tail functions as an invader oligonucleotide that hybridizes to a supplied FRET probe. The FRET probe is labeled with a FRET-pair that is separated when the probe is cleaved. The separation of the donor and acceptor prevents energy transfer and thus the donor starts to emit light. Increased fluorescence is taken as evidence for the presence of the matching allele.

The reaction is conducted at a single temperature close to the melting point of the probes. This allows new probes to hybridize to the target after cleavage, and signal amplification without thermo cycling. The method is very sensitive and if more than 100ng of genomic DNA is supplied in the reaction mix it can

be performed without prior PCR amplification [82]. However, the amount of DNA needed to reliably run this technique without the use of PCR is too large (>100ng/genotype) to be practical. It is almost always the case that the materials used in genetic studies are limited. Saving DNA will continue to be important until whole genome amplification becomes routine.

3 ARRAYS

Taking SNP scoring technology towards the high throughput capacity required in future genetic studies involves the elimination of labor-intensive and slow procedures. Moving to systems based on sample analysis upon planar surfaces rater than in sample tubes or wells allows massive parallel sample processing with minimal manual liquid handling.

To put my own work on arrays in context, I will discuss current array technologies.

3.1 SOLID SUPPORTS

Edwin Southern realized in the early-1970s that labeled DNA of known sequence could be used to interrogate solid support bound DNA of unknown sequence [23]. The southern blot was the first step towards array technology. The technique of analyzing multiple target molecules in parallel by binding them to a solid planar support in a defined pattern and then hybridizing a labeled probe to them arrived with the dot blot technique [28]. The advantages were tremendous. Not only could the hybridizations be carried out in one common step by immersing the array in probe solution, but also processing and signal measurements could be conducted in a parallel fashion. The explosion of array technologies over recent years originated in these early techniques, and array technologies now exists in numerous formats and attracts a lot of research attention.

3.1.1 Membranes

The dot blot and the reverse dot blot techniques [28 30] were the true forerunners of today's array technology. These techniques made use of porous membranes to capture DNA. DNA is captured to the membrane either through electrostatic, interaction, hydrophobic interaction, or covalent binding to active groups on the membrane. The advantage of a porous support is the extended

surface that the pores provide. Molecules are bound in three dimensions (as opposed to mono-layers on non-porous supports) and large amounts of DNA can be attached in relatively small array features. Further, it is an advantage that the molecules can be applied on to the membrane in relatively large volumes. The structure of the membrane absorbs the large liquid volume and brings the molecules into the pores where they can attach without excessive lateral spreading. However, membrane spotted features tends to have poorly defined shapes and boundaries, and with increasing array feature density (smaller features) this becomes a problem. It is difficult to reduce the size of the spots smaller than 200µM while maintaining control over how the features form on the membrane. Membranes and porous supports are preferable in macro-array applications, where array feature density is <100 features per cm^2.

3.1.2 Impermeable supports

For automated analysis of hybridization signals it is important that position and shape of array features are consistent and precise. When feature size and distance between features shrink precision becomes a difficult challenge. Permeable porous supports tend to expand when immersed in buffer and distort and shrink as they dry. This is one reason why rigid solid supports are preferable for high-density array applications. Moreover, when oligonucleotide probes are spotted on a solid planar support a mono-layer of probes is formed. If the surface has been evenly activated for molecule capture and binding the surface becomes saturated and the amount attached is consistent from one region of the surface to another. This is particularly important in expression analysis applications where the quantity of hybridized molecules is compared. It is also favorable to have all the nucleic acid molecules on the surface, freely available for hybridization targets. With porous supports it takes more time for the hybridization targets to diffuse into the pores and find their complementary nucleic acid strands.

There are alternative materials for making arrays on solid impermeable supports [83] [84]but glass is the most common. Glass has the advantage of being sheap, is easily modified for covalent attachment of molecules [85] and

possesses preferable physical characteristics such as rigidity and low intrinsic fluorescence.

3.2 ARRAY CREATION

The basic strategies to create arrays are either to attach pre-synthesized nucleic acids to the surface or to synthesize the molecules directly on the surface (*in situ*). Both strategies have advantages and disadvantages and offer different technological challenges.

3.2.1 Pre-synthesized probes

3.2.1.1 Contact based arraying

Array creation by spotting oligonucleotides, PCR product or cloned sequences is in theory straightforward. Most broadly, arraying technologies falls into two categories; the contact based and the non-contact based. The contact based arraying systems employ pins, adapted from designs used in ink pens (metal capillaries or quills), controlled by a computer to transfer nucleic acid solution from multi-well plates to the array. By letting the pins touch the array surface the solution deposited. Numerous designs of arraying robots exist [86] and many of them can be adapted to the users needs. There are single pin systems for transfer of only one sample at a time and there are devices with up to 1536 prongs for instant transfer of 1536 samples to macro arrays in one action. In contact based arraying systems the actual sample volume transferred is dependent on both the properties of the transferred liquid and the wetting properties of the array surface.

3.2.1.2 Non-contact based arraying

The non-contact arraying systems are built on the inkjet-printing technology developed for high-resolution computer printers. Even though the technology generally is more associated with *in-situ* synthesis of oligonucleoides directly on the array surface (discussed below) it has also been used for arraying of

pre-synthesized molecules [87] [88] [89]. The array feature resolution can be set higher than for the contact based systems, and the volumetric accuracy of ink-jet dispensing is not affected by how the fluid wets the surface. However, handling many different samples requires many different dispensing channels and cleaning protocols for the print heads, which is technically complicated to construct. The ink-jet technology is more suitable for handling of only a few reagents.

3.2.2 In situ synthesis of probes

The synthesis of oligonucleotides involves stepwise sequential addition of the four different nucleosides and de-protection of functional binding groups. The building blocks used for synthesis are commonly referred to as monomers, which are activated DNA phosphoramidite nucleosides modified with a leaving group on the 5'-end and a 3'-phosphite group. By cleaving the 5' leaving group from the last residue on the growing oligonucleotide and adding the next activated monomer to couple to the available 5' hydroxyl group oligonucleotides up to 100 bases can be synthesized (limitation in length is dependent on the stepwise yield during synthesis).

3.2.2.1 Ink-Jet technology

The ink-jet technology is suitable for *in situ* synthesis of oligonucloetides on an array surface [90]. Only a few reagents are required and can all be loaded in separate ink-jet tanks (channels), thus the system has much in common with an ink-jet computer printer where the four colors are kept in separate tanks. The process of moving the printer head and the substrate (the array surface) and correct dispensing has easily been adapted from color printing, and allows for quick and easy customization of high-density (100µm features) oligonucleotide arrays. Deprotection reagents can be flushed over the surface between each addition of monomers to the growing oligonucleotides.

3.2.2.2 Photolithography

In the case of photo cleavable protecting groups the deprotection of functional groups on the growing oligonucleotides during *in situ* synthesis can

be mediated trough UV light illumination. This concept has been used in fabrication of extremely high-resolution oligonucleotide arrays (array feature sizes in the 25µm range) [91 92 93]. Photolithographic masks are used for site-specific deprotection and activation of functional groups for coupling of the next set of monomers. By applying different photolithographic masks in each cycle of deprotection, it is possible to have full control over the oligonucleotide sequence in each array feature.

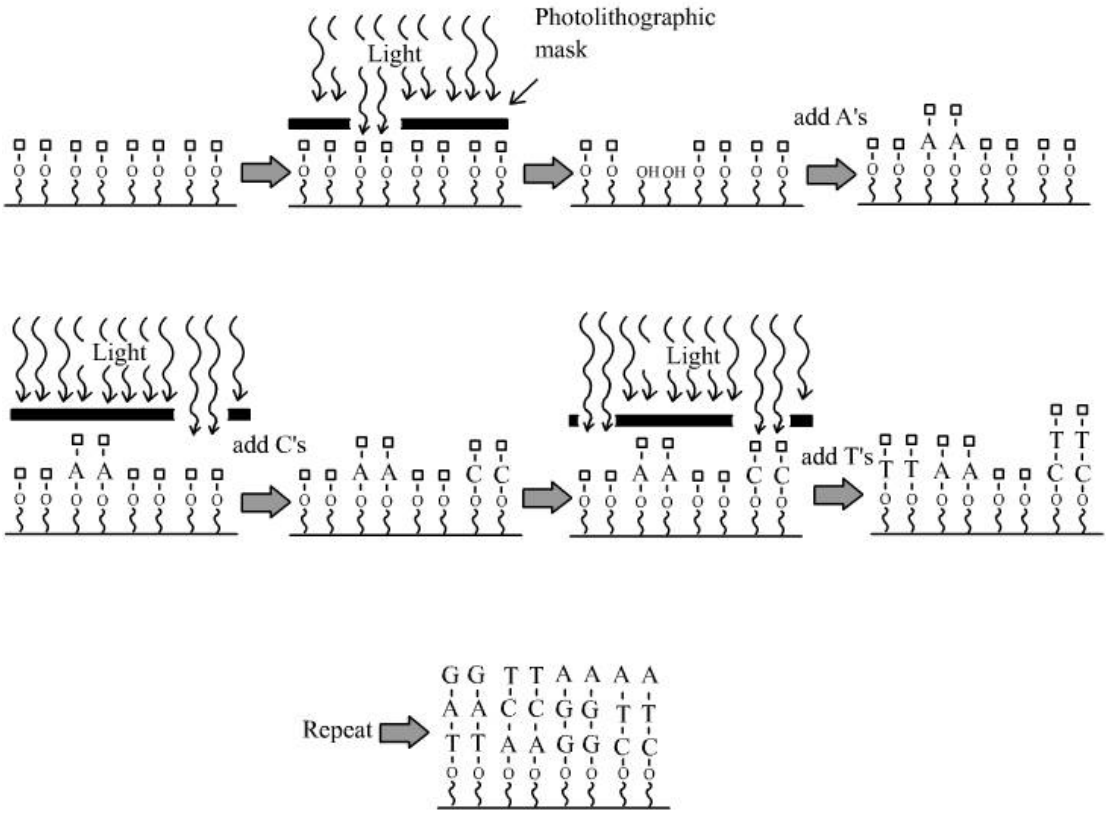

Figure 11 Oligonucleotides are synthesized by the stepwise addition of 5' protected nucleotides (monomers) and deprotection of functional groups on the growing oligonucleotide. In the case of photo cleavable protecting groups deprotection can be mediated through UV illumination. By using fine photolithographic masks precise illumination of predefined areas of a substrate can be carried out. When adding monomers to the substrate only the oligonucleotides in the deprotected areas will react and be extended. By stepwise illumination through new photolithographic masks and the addition of monomers arrays with features in the 25µm range can be created.

The stepwise yield in each cycle of deprotection and oligonucleotide extension using photolithographic oligonucleotide synthesis has been reported to be in the range of 82% - 92% [94]. This makes it difficult to create oligonucleotides longer then 25 bases and truncated versions of the desired sequence contaminate each feature.

4 PRESENT INVESTIGATION

During the past five years (the period over which this project spans) the interest in SNPs has grown immensely within the scientific community. SNP research is expected to reveal many of the missing parts in the gigantic puzzle of understanding human disease and phenotypic variation. At an early stage a major obstacle for success in this task was identified, namely the enormous amounts of data needed to make any sense out of the clues that SNPs may provide. The few methods for SNP scoring available were slow, labor intensive and costly. This is the context within which the present work should be viewed.

4.1 PAPER I

DASH (dynamic allele specific hybridization) was developed as a direct response to increasing requirements in genotyping capacity in our laboratory. An ambitious study at that time was undertaken in our laboratory to identify a large number of SNPs in candidate genes for involvement in Alzheimer's disease [95]. The SNPs were to be used as markers in subsequent association studies on a large patient and control material [11 96 97]. However, the cost and time that it would take to perform association studies for all the identified SNPs on the large sample set using existing SNP genotyping techniques was unacceptable. It was also evident that other research groups doing SNP genotyping encountered the same problems.

Paper I is the first presentation of DASH, a method for calling SNP genotypes in a quick and low cost manner. The paper describes how SNP alleles unambiguously can be discriminated using the principles of ASH but without the optimizations normally associated with this technique. The dynamic heating and real-time monitoring of hybridization status is the key concept that removes the difficulty of finding stringencies at which the allele-specific probe only hybridizes to the perfectly complementary target and not to the single base mismatched.

The DASH technique involves a short PCR spanning the SNP position. One of the PCR primers is biotinylated and via streptavidin-biotin affinity capture, the PCR product is bound to a streptavidin coated solid support. The non-biotinylated strand of the PCR product is then removed by an alkali rinse. An allele-specific probe is annealed to the bound target molecule along with a double-strand specific fluorescent dye. By slowly increasing the sample temperature and simultaneously monitoring fluorescence from the double-strand specific dye, denaturation of the probe-target duplex can be plotted as a function of fluorescence intensity over temperature. A homozygous sample of the mismatching allele generates a drop in fluorescence intensity at a low temperature. A homozygous sample of the matching allele generates a high temperature fluorescence intensity drop, and a heterozygous sample generates two distinct drops in fluorescence intensity. The first negative derivative of these denaturation curves displays easily interpretable peaks at the melting temperatures of the probe-target duplexes. A single low or a single high temperature peak represents homozygous mismatching or matching samples respectively. A double peak represents a heterozygous sample.

Assay conditions, such as what buffers to use, pH level, salt concentrations, what dyes to use etc. were optimized on model systems of biotinylated oligonucleotides, representing alleles of target sequences, with their corresponding probes. It was found that a buffer at pH 8.0 comprising 0.1 M HEPES, 50mM NaCl, 1 mM EDTA with Sybr Green I dye at 1:10 000 had the desirable properties for DASH.

Using the assay conditions found to be preferable for a high-quality DASH, experiments were performed on PCR product amplified over known SNP positions. To show the applicability of DASH, assays were designed for a G-A, C-A, T-A and a C-G polymorphism. DASH performed well on the PCR products, and scored the genotypes in agreement with genotype data produced, using other genotyping methods.

4.2 PAPER II

This project is a joint effort with a group at Uppsala University. In my research I looked at the possibilities of inverting DASH. That is, binding the probe to the solid support and then providing the PCR product in solution, as opposed to regular DASH where the PCR product is bound and the probe is provided in solution. Simultaneously, the Uppsala group was working on a means to produce pure oligonucleotides *in situ* on a solid support. A project initiated by the known fact that *in situ* photolithographic synthesis of oligonucleotides on a solid support tends to leave substantial amounts of truncated oligonucleotide impurities on the surface [94]. A fair hypothesis was that these impurities deteriorate the performance of DNA chips, but no one had fully studied the problem before. We used the concept of inverted DASH to reveal the effect of truncated oligonucleotide impurities in hybridization experiments distinguishing SNP alleles.

The inverted version of DASH was used to study how the truncated molecules affect hybridization. 9 to 20 mer oligonucleotides were bound to a solid support in ratios as could be expected from the *in situe* photolithographic oligonucleotide synthesis of 20 mers. Three series of truncated probes, with relative ratios as expected from an average yield of 82%, 88% and 92% per synthesis cycle in the photolithographic oligonucleotide synthesis process were analyzed. Position number ten of the bound molecules was designed to represent a SNP position, and two versions of each molecule were made representing the two alleles. Target oligonucleotides matching, or mismatching at the SNP position, were hybridized to the bound molecules and denatured using a dynamic heat sweep. The difference between the matched and the mismatched denaturing data was calculated and compared to the denaturing data of matched and mismatched pure 20 mer duplexes.

It was concluded from the experimental data that the truncated contaminants had a deteriorating effect upon SNP allele discrimination at the optimal discriminatory temperature.

4.3 PAPER III

The Increasingly complex technologies flooding biology research puts a heavy pressure on research budgets for groups wanting to keep up with projects at the forefront. Advanced tends to spell expensive. The philosophy behind the whole DASH development project has been to stay away from highly advanced concepts, without compromising competitiveness, in order to keep costs at minimal level. This paper is in line with that philosophy.

Many analysis methods in molecular biology include solid phase reactions. Samples are transferred and chemically bound to the solid phase in order to facilitate removal of chemicals harmful to the analysis procedure without risk of loosing the sample. The solid phase can be the inner surface of a sample tube, micro beads, a planar solid surface (i.e. glass or polymer) or a membrane. The planar solid surface and the membrane have the advantage over the other supports in that they easily can be immersed in reagents during subsequent hybridization- and washing steps. In the case of few- large volume samples the transfer from reaction vessels to the solid surface can be mediated through pipeting or by vacuum aspiration. However, when the numbers of samples to be transferred is many hundreds and thousands and the sample volumes are in the range of only a few micro liters and smaller, pipeting or vacuum aspiration are no longer options if costs are to be kept low and accuracy acceptable.

Here we describe a simple alternative to complex liquid handling systems for sample transfer and creation of macro arrays. Centrifugation is used to create arrays on both glass surfaces and membranes directly from 1536 sample plates. Tools and strategies have been designed to avoid cross contamination between features in the action of centrifuging samples on to both solid surfaces and to membranes.

4.4 PAPER IV

At an early stage in the development of DASH we encountered problematic assays where the PCR seemed to work fine but the produced DASH data were of too low quality to be reliably translated into genotypes. Careful analysis of the PCR product sequences revealed that the target molecules in the poorly performing DASH assays contained harsh secondary structures. It was also found that by altering the target sequence at a few key positions involved in secondary structure formation could overcome these problems. DASH assay PCR products are always short fragments, so the majority of the product is comprised of PCR primer sequence. Therefore it is possible to mutate the DASH target sequence by substituting bases in the PCR primers [16].

A strategy for identifying key bases involved in secondary structure was worked out. The strategy made use of a web-based secondary structure prediction tool (M-fold) and a regular PCR design tool (Oligo 5.0). PCR primers were first designed to fill criteria for good PCR reactions. Then, the secondary structure formation of the product was predicted in the mfold [98] software (http://www.bioinfo.rpi.edu/applications/mfold/old/dna/). If troublesome secondary structures were predicted the primers were altered at key positions involved in secondary structure formation. Folding the sequence again determined whether the change had removed the secondary structure. This was repeated over and over again until a PCR product with minimal predicted secondary structure had been identified. Often more than one base per PCR primer has to be altered to produce a PCR product with satisfactory level of secondary structure. This makes the number of combinations in which the primers can be altered very high. Because of this the manual DASH assay design is a slow and difficult work.

Paper IV is a presentation of software developed in our lab that automatically searches for good PCR primers and minimizes secondary structure formation in the PCR product target by altering bases in the PCR primers. The program has been validated and is now in use by our laboratory staff as well as by other laboratories in the world using DASH.

4.5 PAPER V

The overall goal of my research project following the first publication of DASH has been to advance the concept of dynamic allele specific hybridization. Even though the first version of DASH (se section 4.1) dramatically improved genotyping capacity in our laboratory, and ever since has generated high quality genotyping data in several biology studies, [99] [11] [100] [101] [102] [103] it has always been our ambition to reach further with this technology. Our goal has been to provide a flexible technology where the cost per genotype is minimal and the genotyping throughput can be scaled from low to very high, still using the same system.

The system described in paper V is termed DASH-2, and is a highly flexible system for medium to high throughput genotyping of SNPs built on the inherent simplicity of DASH. The DASH-2 system includes macro array creation by centrifugation (se section 4.3) on streptavidin coated polypropylene membranes. A concept developed in our laboratory for fluorescence signal generation termed iFRET [104] has also been implemented in the new system. iFRET is based on FRET between intercalating dye and a reporter dye attached to the DASH probe. When iFRET is used only the signal from the reporter is monitored and unspecific hybridization is not. iFRET also open up possibilities for multiplexing. By labeling DASH probes for different assays with dyes having separated emission spectra the melting profiles for the different assays can be generated by monitoring the different dyes separately using optical filters specific for the different dyes.

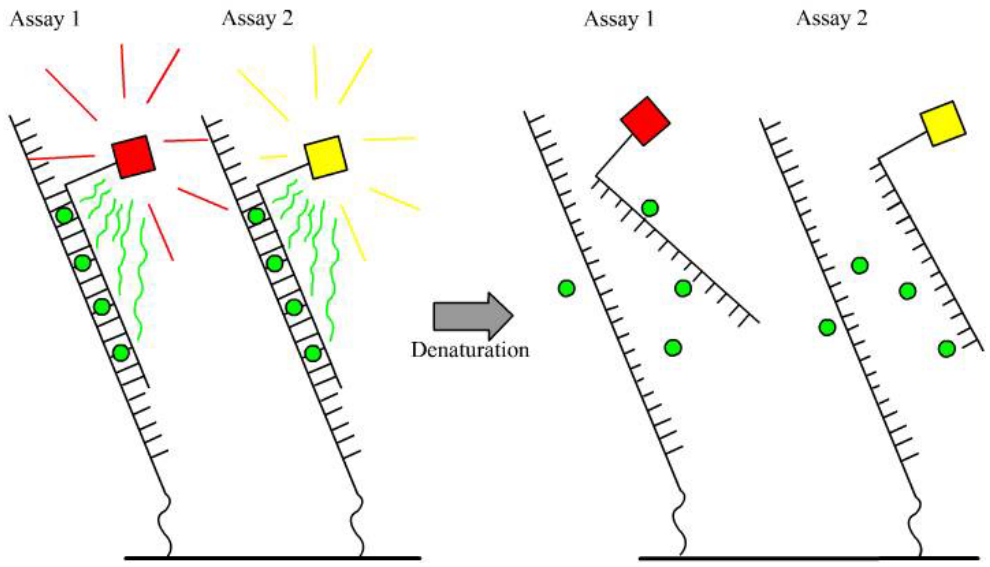

Figure 12 iFRET is based on fluorescence resonance energy transfer between intercalating dye and probe bound dye. When the intercalating dye is excited by UV illumination it normally emits light, but dye intercalated in the probe-target duplex loses the energy to the probe-bound dye, which in turn emits light at wavelengths well separated from the emission spectra of the intercalating dye. By monitoring the emitted light from the probe bound reporter dye (filtering out the light from the intercalating dye) high signal specificity from the probe-target duplex is achieved. By using different reporter dyes, with separated emission spectra, for different assays multiplexing is possible.

We demonstrate the flexibility of the system by creating arrays by centrifugation [105] from different microtiter plate formats. Cost effectiveness is demonstrated in the small volume PCRs required and the possibilities for several levels of multiplexing.

5 DISCUSSION

The overall aim of the research underlying this thesis has been to provide the genetic research community with tools for effective and inexpensive analysis of SNPs. A guiding philosophy for this research, that has proven fruitful, has been to always look for the simple solutions. The simplicity of DASH is in line with this philosophy and very much because of its simple nature has it been possible to miniaturize and transform the original version of DASH into the highly effective system that DASH-2 constitutes. The array creation by centrifugation is another example where simplicity allowed for the development of an efficient and cheap routine, a routine that also was implemented in the DASH-2 system. It is very much because of the inherent simplicity that the DASH-2 system can turn out genotypes at costs of only a few cents, and with virtually 100% accuracy. The applicability of DASH is high. Only occasionally is the sequence context surrounding an SNP such that a functioning assay cannot be designed. DASH reaction conditions do not alter between assays and thus the only required optimization for individual assays is for the PCR.

The development of DASH continues. Currently the main objective in the DASH development project is the manual processing steps involved in the procedure. There are a few methods for SNP analysis that do not require manual attention once the analysis reaction mix has been added to the DNA sample, and of course such methods are attractive in high throughput settings. However, the allele discriminating chemistry in such methods tend to be of complex nature and thus expensive. We are currently developing the processing procedures of the DASH-2 system to minimize manual handling and provide an arrangement that come as close as possible to the automatic (no-attention) systems but with maintained low cost. Processing platforms for automatic array washing and probing are being constructed and software for analysis of DASH data (automatic scoring) is under development. In order to minimize costs further we are also looking into alternative array surfaces and DNA binding chemistries. The streptavidin coated polypropylene membranes

currently in use in the DASH-2 system are expensive mainly due to reagent costs for the coating.

In conclusion, DASH provides a powerful alternative for genotyping. The method is flexible and inexpensive enough for all to use, and it does not require the purchasing of expensive robotics. These benefits stem directly from the elemental simplicity of the underlying DASH reaction concept, which perhaps also has potential for expression array analyses, (re)sequencing, and DNA fingerprinting. We additionally envisage further implementation improvements towards DASH-3 (e.g., nano-scale and microbead versions) that could bring even faster and cheaper genotyping possibilities in the future.

6 ACKNOWLEDGEMENTS

There are many people that in different ways have influenced and inspired me during my PhD. Without some of them this work would not bee. Others have been there sharing times of despair as well as times of joy with me. As I close this chapter in my life I want to express my sincere appreciation to these people. In particular I want to direct gratitude towards…

…my supervisor Professor **Dr. Anthony Brookes** who gave me room in his forefront research group. It has all been exciting and I have learnt many things from you. Thank you Tony!

… **Dr. Claes Wahlestedt** for building up a fantastic department.

…all my dear present and former PhD student colleagues in the research group at the Center for Genomics and Bioinformatics . **Mathias Howell** for deep friendship and generosity, and for a sharp mind that helped me view difficulties of various kinds in new light. I could not have done this without you Math. **Dr. Lars Feuk** I admire you for your razor-sharp intellect that you with brilliance use not only in science but also to cheer people up and to bring out joy and laughter. **David Fredman** I have much enjoyed listening to your music and your 'fishy' humour. I admire you for your wizard-like skills behind the keyboard. David and Lars, I share precious memories with you from the trip through New Mexico and Arizona to the Grand Canyon. **Sarah Sawyer** for your sharpness and your ability to deliver sweet compliments, and for your drive in educating the group through Journal Club. **Dr.Tesfai Emahazion** for taking care of me when I initiated my work back in Uppsala and for being a good friend. I have missed you.

...all the post Docs that have brought in fresh ideas to the group. In particular **Dr.Jonathan Prince** for being so determined to do good science and for all your stories about strange people that seems to flock around you. **Dr.Torsten Mayr** for adding new fuel to the DASH development project and for being a very nice person, **Dr.Harvest Gu** for smiling a lot, **Dr.Salim Mottagui-Tabar** for sharpness.

...the highly skilled technical and administrational staff in the group (former and present). **Linda Strömqvist** thank you for your patience and your bright input in the DASH project, **Daniel Pederson** for being clever and friendly, **Susanna Potter** for the sunshine that always follows you, **Camilla Lagerberg** for laughing a lot, **Sandra Tammer** for your fighting spirit, **Linda Berglind** for being so cool, **Vivianne Daza-Saucedo** thank you for being so tough, **Alex Covacs, Per Ewing, Isabelle Dussauge, Nina Juntilla, Hagit Katzov, Astrid Lindstedt, Gill Munns, Daniel Rios, Marianne Siegfried and Satu Valtonen** you are all fantastic people.

… all the people at the department of genetics and pathology in Uppsala, especially **Anna Beskow, Dr.Patrik Magnusson** and **Inger Jonasson**.

… my parents **Per** and **Birgitta Jobs**, and my parents in law **Torsten** and **Marianne Baier**. Thank you for always being there for my family and me! I am lucky to have you. I also want to acknowledge my siblings **Daniel Jobs** and **Karin Jobs Munters** and their families, as well as my aunt **Kerstin Trygg.**

…my wife **Elisabeth** and my children **Andrea, Daniela, Jonathan** and **Benjamin** words can not express my feelings for you. I love you so much. Without you I would be nothing.

…**Thermo Hybaid** for financial support.

7 REFERENCES

1. Mendel Versuche über Plflanzenhybride. *Verhandlungen des naturforschenden Vereines in Brünn* **4** (1865).
2. Lander, E.S. et al. Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921. (2001).
3. Venter, J.C. et al. The sequence of the human genome. *Science* **291**, 1304-1351. (2001).
4. Reich, D.E. et al. Human genome sequence variation and the influence of gene history, mutation and recombination. *Nat Genet* **32**, 135-142. (2002).
5. Brookes, A.J. The essence of SNPs. *Gene* **234**, 177-186. (1999).
6. Kruglyak, L. & Nickerson, D.A. Variation is the spice of life. *Nat Genet* **27**, 234-236. (2001).
7. DOE/NIH, D.o.E.a.N.I.o.H. Understanding our genetic inheritance the U.S. human genome project The First Five Years: Fiscal Years 1991-1995., (1990).
8. Collins, F.S. et al. New goals for the U.S. Human Genome Project: 1998-2003. *Science* **282**, 682-689. (1998).
9. Fredman, D. et al. HGVbase: a human sequence variation database emphasizing data quality and a broad spectrum of data sources. *Nucleic Acids Res* **30**, 387-391. (2002).
10. Sherry, S.T. et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* **29**, 308-311. (2001).
11. Emahazion, T. et al. SNP association studies in Alzheimer's disease highlight problems for complex disease analysis. *Trends Genet* **17**, 407-413. (2001).
12. Weiss, K.M. & Terwilliger, J.D. How many diseases does it take to map a gene with SNPs? *Nat Genet* **26**, 151-157. (2000).
13. Brennan, M.D. High throughput genotyping technologies for pharmacogenomics. *Am J Pharmacogenomics* **1**, 295-302 (2001).
14. Kwok, P.Y. Methods for genotyping single nucleotide polymorphisms. *Annu Rev Genomics Hum Genet* **2**, 235-258 (2001).
15. Howell, W.M., Jobs, M., Gyllensten, U. & Brookes, A.J. Dynamic allele-specific hybridization. A new method for scoring single nucleotide polymorphisms. *Nat Biotechnol* **17**, 87-88. (1999).
16. 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).
17. Jeffreys, A.J. DNA sequence variants in the G gamma-, A gamma-, delta- and beta-globin genes of man. *Cell* **18**, 1-10. (1979).
18. Kwok, P.Y. High-throughput genotyping assay approaches. *Pharmacogenomics* **1**, 95-100. (2000).
19. Watson, J.D. & Crick, F.H.C. A structure for deoxyribose nucleic acid. *Nature* **171**, 737-738 (1953).

20. Marmur, J. & Doty, P. Thermal renaturation of deoxyribonucleic acids. *J Mol Biol* **3**, 583-594 (1961).
21. Nygaard, A.P. & Hall, B.D. Formation and properties of RNA-DNA complexes. *J Mol Biol* **9**, 125-142 (1964).
22. Gillespie, D. & Spiegelman, S. A quantitative assay for DNA-RNA hybrids with DNA immobilized on a membrane. *J Mol Biol* **12**, 829-842. (1965).
23. Southern, E.M. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* **98**, 503-517. (1975).
24. Wallace, R.B. et al. Hybridization of synthetic oligodeoxyribonucleotides to phi chi 174 DNA: the effect of single base pair mismatch. *Nucleic Acids Res* **6**, 3543-3557. (1979).
25. Conner, B.J. et al. Detection of sickle cell beta S-globin allele by hybridization with synthetic oligonucleotides. *Proc Natl Acad Sci U S A* **80**, 278-282. (1983).
26. Ylitalo, N., Bergstrom, T. & Gyllensten, U. Detection of genital human papillomavirus by single-tube nested PCR and type-specific oligonucleotide hybridization. *J Clin Microbiol* **33**, 1822-1828. (1995).
27. Sapolsky, R.J. et al. High-throughput polymorphism screening and genotyping with high-density oligonucleotide arrays. *Genet Anal* **14**, 187-192. (1999).
28. 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).
29. Templeton, E.F., Wong, H.E., Evangelista, R.A., Granger, T. & Pollak, A. Time-resolved fluorescence detection of enzyme-amplified lanthanide luminescence for nucleic acid hybridization assays. *Clin Chem* **37**, 1506-1512. (1991).
30. Saiki, R.K., Walsh, P.S., Levenson, 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).
31. Chee, M. et al. Accessing genetic information with high-density DNA arrays. *Science* **274**, 610-614. (1996).
32. Cutler, D.J. et al. High-throughput variation detection and genotyping using microarrays. *Genome Res* **11**, 1913-1925. (2001).
33. Allawi, H.T. & SantaLucia, J., Jr. Thermodynamics and NMR of internal G.T mismatches in DNA. *Biochemistry* **36**, 10581-10594. (1997).
34. Allawi, H.T. & SantaLucia, J., Jr. Thermodynamics of internal C.T mismatches in DNA. *Nucleic Acids Res* **26**, 2694-2701. (1998).
35. Allawi, H.T. & SantaLucia, J., Jr. Nearest neighbor thermodynamic parameters for internal G.A mismatches in DNA. *Biochemistry* **37**, 2170-2179. (1998).

36. Peyret, N., Seneviratne, P.A., Allawi, H.T. & SantaLucia, J., Jr. Nearest-neighbor thermodynamics and NMR of DNA sequences with internal A.A, C.C, G.G, and T.T mismatches. *Biochemistry* **38**, 3468-3477. (1999).
37. Drmanac, R. et al. Sequencing by hybridization: towards an automated sequencing of one million M13 clones arrayed on membranes. *Electrophoresis* **13**, 566-573. (1992).
38. Hacia, J.G., Brody, L.C., Chee, M.S., Fodor, S.P. & Collins, F.S. Detection of heterozygous mutations in BRCA1 using high density oligonucleotide arrays and two-colour fluorescence analysis. *Nat Genet* **14**, 441-447. (1996).
39. Lipshutz, R.J. et al. Using oligonucleotide probe arrays to access genetic diversity. *Biotechniques* **19**, 442-447. (1995).
40. Guo, Z., Guilfoyle, R.A., Thiel, A.J., Wang, R. & Smith, L.M. Direct fluorescence analysis of genetic polymorphisms by hybridization with oligonucleotide arrays on glass supports. *Nucleic Acids Res* **22**, 5456-5465. (1994).
41. Livak, K.J. Allelic discrimination using fluorogenic probes and the 5' nuclease assay. *Genet Anal* **14**, 143-149. (1999).
42. Livak, K.J., Flood, S.J., Marmaro, J., Giusti, W. & Deetz, K. Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. *PCR Methods Appl* **4**, 357-362. (1995).
43. Latif, S., Bauer-Sardina, I., Ranade, K., Livak, K.J. & Kwok, P.Y. Fluorescence polarization in homogeneous nucleic acid analysis II: 5'- nuclease assay. *Genome Res* **11**, 436-440. (2001).
44. Kutyavin, I.V. et al. 3'-minor groove binder-DNA probes increase sequence specificity at PCR extension temperatures. *Nucleic Acids Res* **28**, 655-661. (2000).
45. Kostrikis, L.G., Tyagi, S., Mhlanga, M.M., Ho, D.D. & Kramer, F.R. Spectral genotyping of human alleles. *Science* **279**, 1228-1229. (1998).
46. Tyagi, S., Bratu, D.P. & Kramer, F.R. Multicolor molecular beacons for allele discrimination. *Nat Biotechnol* **16**, 49-53. (1998).
47. Marras, S.A., Kramer, F.R. & Tyagi, S. Multiplex detection of single-nucleotide variations using molecular beacons. *Genet Anal* **14**, 151-156. (1999).
48. Bonnet, G., Tyagi, S., Libchaber, A. & Kramer, F.R. Thermodynamic basis of the enhanced specificity of structured DNA probes. *Proc Natl Acad Sci U S A* **96**, 6171-6176. (1999).
49. Poddar, S.K. Detection of adenovirus using PCR and molecular beacon. *J Virol Methods* **82**, 19-26. (1999).
50. Pastinen, T. et al. A system for specific, high-throughput genotyping by allele-specific primer extension on microarrays. *Genome Res* **10**, 1031-1042. (2000).
51. Syvanen, A.C., Aalto-Setala, K., Harju, L., Kontula, K. & Soderlund, H. A primer-guided nucleotide incorporation assay in the genotyping of apolipoprotein E. *Genomics* **8**, 684-692. (1990).

52. 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).
53. Germer, S. & Higuchi, R. Single-tube genotyping without oligonucleotide probes. *Genome Res* **9**, 72-78. (1999).
54. Ugozzoli, L., Wahlqvist, J.M., Ehsani, A., Kaplan, B.E. & Wallace, R.B. Detection of specific alleles by using allele-specific primer extension followed by capture on solid support. *Genet Anal Tech Appl* **9**, 107-112. (1992).
55. Pastinen, T., Kurg, A., Metspalu, A., Peltonen, L. & Syvanen, A.C. Minisequencing: a specific tool for DNA analysis and diagnostics on oligonucleotide arrays. *Genome Res* **7**, 606-614. (1997).
56. Shumaker, J.M., Metspalu, A. & Caskey, C.T. Mutation detection by solid phase primer extension. *Hum Mutat* **7**, 346-354 (1996).
57. Chen, X., Levine, L. & Kwok, P.Y. Fluorescence polarization in homogeneous nucleic acid analysis. *Genome Res* **9**, 492-498. (1999).
58. Chen, X. & Kwok, P.Y. Template-directed dye-terminator incorporation (TDI) assay: a homogeneous DNA diagnostic method based on fluorescence resonance energy transfer. *Nucleic Acids Res* **25**, 347-353. (1997).
59. Fan, J.B. et al. Parallel genotyping of human SNPs using generic high-density oligonucleotide tag arrays. *Genome Res* **10**, 853-860. (2000).
60. Cai, H. et al. Flow cytometry-based minisequencing: a new platform for high-throughput single-nucleotide polymorphism scoring. *Genomics* **66**, 135-143. (2000).
61. Chen, J. et al. A microsphere-based assay for multiplexed single nucleotide polymorphism analysis using single base chain extension. *Genome Res* **10**, 549-557. (2000).
62. Karas, M. & Hillenkamp, F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* **60**, 2299-2301. (1988).
63. Ross, P., Hall, L., Smirnov, I. & Haff, L. High level multiplex genotyping by MALDI-TOF mass spectrometry. *Nat Biotechnol* **16**, 1347-1351. (1998).
64. Buetow, K.H. et al. High-throughput development and characterization of a genomewide collection of gene-based single nucleotide polymorphism markers by chip- based matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Proc Natl Acad Sci U S A* **98**, 581-584. (2001).
65. Ahmadian, A. et al. Single-nucleotide polymorphism analysis by pyrosequencing. *Anal Biochem* **280**, 103-110. (2000).
66. Ahmadian, A., Gharizadeh, B., O'Meara, D., Odeberg, J. & Lundeberg, J. Genotyping by apyrase-mediated allele-specific extension. *Nucleic Acids Res* **29**, E121. (2001).

67. Ronaghi, M., Nygren, M., Lundeberg, J. & Nyren, P. Analyses of secondary structures in DNA by pyrosequencing. *Anal Biochem* **267**, 65-71. (1999).
68. Nordfors, L. et al. Large-scale genotyping of single nucleotide polymorphisms by Pyrosequencingtrade mark and validation against the 5'nuclease (Taqman((R))) assay. *Hum Mutat* **19**, 395-401. (2002).
69. Gharizadeh, B., Nordstrom, T., Ahmadian, A., Ronaghi, M. & Nyren, P. Long-read pyrosequencing using pure 2'-deoxyadenosine-5'-O'-(1- thiotriphosphate) Sp-isomer. *Anal Biochem* **301**, 82-90. (2002).
70. Wu, D.Y. & Wallace, R.B. Specificity of the nick-closing activity of bacteriophage T4 DNA ligase. *Gene* **76**, 245-254 (1989).
71. Landegren, U., Kaiser, R., Sanders, J. & Hood, L. A ligase-mediated gene detection technique. *Science* **241**, 1077-1080. (1988).
72. Nilsson, M. et al. Padlock probes: circularizing oligonucleotides for localized DNA detection. *Science* **265**, 2085-2088. (1994).
73. 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).
74. Lizardi, P.M. et al. Mutation detection and single-molecule counting using isothermal rolling-circle amplification. *Nat Genet* **19**, 225-232. (1998).
75. Qi, X., Bakht, S., Devos, K.M., Gale, M.D. & Osbourn, A. L-RCA (ligation-rolling circle amplification): a general method for genotyping of single nucleotide polymorphisms (SNPs). *Nucleic Acids Res* **29**, E116. (2001).
76. Faruqi, A.F. et al. High-throughput genotyping of single nucleotide polymorphisms with rolling circle amplification. *BMC Genomics* **2**, 4 (2001).
77. Chen, X., Livak, K.J. & Kwok, P.Y. A homogeneous, ligase-mediated DNA diagnostic test. *Genome Res* **8**, 549-556. (1998).
78. Day, I.N., Humphries, S.E., Richards, S., Norton, D. & Reid, M. High-throughput genotyping using horizontal polyacrylamide gels with wells arranged for microplate array diagonal gel electrophoresis (MADGE). *Biotechniques* **19**, 830-835. (1995).
79. Butler, J.M. & Reeder, D.J. Detection of DNA polymorphisms using PCR-RFLP and capillary electrophoresis. *Methods Mol Biol* **163**, 49-56 (2001).
80. Akey, J.M. et al. Melting curve analysis of SNPs (McSNP): a gel-free and inexpensive approach for SNP genotyping. *Biotechniques* **30**, 358-362, 364, 366-357. (2001).
81. Lyamichev, V. et al. Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes. *Nat Biotechnol* **17**, 292-296. (1999).

82. Hall, J.G. et al. Sensitive detection of DNA polymorphisms by the serial invasive signal amplification reaction. *Proc Natl Acad Sci U S A* **97**, 8272-8277. (2000).
83. Matson, R.S., Rampal, J., Pentoney, S.L., Jr., Anderson, P.D. & Coassin, P. Biopolymer synthesis on polypropylene supports: oligonucleotide arrays. *Anal Biochem* **224**, 110-116. (1995).
84. Matson, R.S., Rampal, J.B. & Coassin, P.J. Biopolymer synthesis on polypropylene supports. I. Oligonucleotides. *Anal Biochem* **217**, 306-310. (1994).
85. Lindroos, K., Liljedahl, U., Raitio, M. & Syvanen, A.C. Minisequencing on oligonucleotide microarrays: comparison of immobilisation chemistries. *Nucleic Acids Res* **29**, E69-69. (2001).
86. Fitzgerald, D.A. Microarrayers on the Spot. *The Scientist* **16** (2002).
87. Cooley, P., Hinson, D., Trost, H.J., Antohe, B. & Wallace, D. Ink-jet-deposited microspot arrays of DNA and other bioactive molecules. *Methods Mol Biol* **170**, 117-129 (2001).
88. Okamoto, T., Suzuki, T. & Yamamoto, N. Microarray fabrication with covalent attachment of DNA using bubble jet technology. *Nat Biotechnol* **18**, 438-441. (2000).
89. Allain, L.R., Askari, M., Stokes, D.L. & Vo-Dinh, T. Microarray sampling-platform fabrication using bubble-jet technology for a biochip system. *Fresenius J Anal Chem* **371**, 146-150. (2001).
90. Hughes, T.R. et al. Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat Biotechnol* **19**, 342-347. (2001).
91. Fodor, S.P. et al. Light-directed, spatially addressable parallel chemical synthesis. *Science* **251**, 767-773. (1991).
92. Pease, A.C. et al. Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proc Natl Acad Sci U S A* **91**, 5022-5026. (1994).
93. Singh-Gasson, S. et al. Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. *Nat Biotechnol* **17**, 974-978. (1999).
94. Pirrung, M.C., Fallon, L., Zhu, J. & Lee, Y.R. Photochemically removable silyl protecting groups. *J Am Chem Soc* **123**, 3638-3643. (2001).
95. Emahazion, T. et al. Identification of 167 polymorphisms in 88 genes from candidate neurodegeneration pathways. *Gene* **238**, 315-324. (1999).
96. Brookes, A.J., Howell, W.M., Woodburn, K., Johnstone, E.C. & Carothers, A. Presenilin-I, presenilin-II, and VLDL-R associations in early onset Alzheimer's disease. *Lancet* **350**, 336-337. (1997).
97. Howell, W.M. & Brookes, A.J. Evaluation of multiple presenilin 2 SNPs for association with early- onset sporadic Alzheimer disease. *Am J Med Genet* **111**, 157-163. (2002).

98. SantaLucia, J., Jr. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest- neighbor thermodynamics. *Proc Natl Acad Sci U S A* **95**, 1460-1465. (1998).
99. Feuk, L. et al. apolipoprotein-E dependent role for the FAS receptor in early onset Alzheimer's disease: finding of a positive association for a polymorphism in the TNFRSF6 gene. *Hum Genet* **107**, 391-396. (2000).
100. 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).
101. Magnusson, V. et al. Fine mapping of the SLEB2 locus involved in susceptibility to systemic lupus erythematosus. *Genomics* **70**, 307-314. (2000).
102. Johansson, C. et al. Association analysis with microsatellite and SNP markers does not support the involvement of BCL-2 in systemic lupus erythematosus in Mexican and Swedish patients and their families. *Genes Immun* **1**, 380-385. (2000).
103. Pitarque, M. et al. Identification of a single nucleotide polymorphism in the TATA box of the CYP2A6 gene: impairment of its promoter activity. *Biochem Biophys Res Commun* **284**, 455-460. (2001).
104. Howell, W.M., Jobs, M. & Brookes, A.J. iFRET: An Improved Fluorescence System for DNA-Melting Analysis. *Genome Res* **12**, 1401-1407. (2002).
105. Jobs, M., Howell, W.M. & Brookes, A.J. Creating arrays by centrifugation. *Biotechniques* **32**, 1322-1324, 1326, 1329. (2002).