INTRODUCTION

Single Nucleotide Polymorphism – definition and applications

What are SNPs?

According to the formal definition [Brookes, 1999], Single Nucleotide Polymorphism (SNP, pronounced as "snip") represents a DNA sequence variant of a single base pair, with the minor allele occurring in more than 1% of a given population. SNPs having a minor allele frequency ≥20% are called "common SNPs". Frequently, the term "SNP" is used in a looser sense for short allelic variants - substitutions or small insertions-deletions (indels) without any assumptions about minimum allele frequencies for the polymorphisms. For example, NCBI dbSNP database (http://www.ncbi.nlm.nih.gov/SNP/) uses the SNP term regardless to allelic frequencies.

In principle, SNPs may be two-, three- and four-allelic; however, in practice even three-allelic SNPs are very rare (less than 0.1% of all human SNPs [Lai, 2001]). The overwhelming majority of SNPs are biallelic. There may be four types of SNPs including one transition C<=>T (G<=>A), and three types of transversions: C<=>A (G<=>T), C<=>G (G<=>C) and T<=>A (A<=>T). The most abundant (two thirds of SNPs) are transitions, originating from the deamination of 5'-methylcytosine to thymine [Holliday and Grigg, 1993].

Compared to other genetic markers, e.g. microsatellites or restriction fragment length polymorphisms (RFLPs), SNPs cover genomes with a much higher density. SNPs have a very low mutation rate per generation (10^{-8}) and thus may serve as reliable markers of molecular evolution [Crow et al., 1995; Li et al., 1996]. Finally, from a technological point of view, SNP markers are more convenient for high-throughput analysis.

The stir about SNPs and splash of technology development in the area of SNP genotyping during the last ten years are explained mainly by the hope that SNPs might be used to identify genes and gene variants causative (including those having a minor effect) for multifactorial diseases and certain drug responses [Schork, 1998].

SNPs themselves may be responsible for a certain phenotype (e.g. disease or a susceptibility to a disease in the case of a complex trait) when an SNP is in the coding sequence and leads to a change in the amino acid sequence of the coded protein, or

when an SNP is in the regulatory region and affects the transcription of a certain gene. Most SNPs have no direct influence on phenotype; however, as genetic markers they might be associated with a gene (or genes) determining a disease or drug response, if their frequency is higher among individuals displaying the trait of interest when compared to the rest of the population [Foster and Sharp, 2002]. Associated genes are further investigated to better understand the disease mechanism and finally to develop therapies. Association studies performed for regions already known to contain susceptibility genes for complex diseases (Alzheimer's disease [Martin et al., 2000], migraine [McCarthy et al., 2001], type II diabetes [Horikawa et al., 2000]) have proven that SNPs are suitable for investigating the genetic origin of diseases.

SNPs might play an important role in the following biomedical studies: (i) determination of disease target genes, (ii) creation of diagnostic assays, (iii) revealing genes, which are potential drug targets, (iv) finding correlations between marker variants and drug response, and (v) revealing people with genotypes for whom the drug is effective. These studies would allow for early disease identification and lead to treatments that would suit patients individually (personalized medicine).

Reliable SNP markers are required for forensic medicine and evolution studies.

SNPs are necessary for gene mapping and studying of complex genetic traits in other organisms, as this is required for agriculture and non-human genetics. They are convenient tools for studying recombination, chromosomal dynamics, genome rearrangements, genetic relatedness, and evolutionary relationships between individuals.

Academic institutions and industry undertake great projects and invest a lot in the search for new SNPs, SNP validation, and development of SNP related technologies.

SNP maps

In 1999, eleven large pharmaceutical and technology companies, four major genomic centers, and the Wellcome Trust established The SNP Consortium (TSC). The aim of TSC was to find and map 300000 SNPs in the human genome. Two years later this project was beyond completion, and 1.42 million SNPs were released in public databases [Sachidanandam et al., 2001]. Since then, public SNP databases i. e. NCBI dbSNP database and HGVbase (Human Genome Variation base [Fredman et al., 2004]) are rapidly expanding. Today in the NCBI dbSNP database there are about 5 million validated SNP clusters.

Now the goal has shifted from the discovery of new SNPs to the characterization of frequencies and distribution of known SNPs within and between different human populations [Matise et al., 2003; Thorisson and Stein, 2003]. This information serves as additional criterion for selection of patient populations for clinical trials, specifically in the case that the disease affects different populations unevenly.

Recent research indicates that about 65 to 85 percent of the human genome may be organized into haplotype blocks that are 10000 bases or larger. These blocks remained intact throughout generations, just because there was not enough time to destroy them by meiotic recombination [Nordborg and Tavare, 2002]. To reveal haplotype blocks in human genome, The HapMap project was organized recently [The International HapMap Consortium, 2004]. Most populations share common SNP variants and haplotype patterns, so analysis of only a few SNPs in the block would allow determining the allelic state of other genetic markers. This would reduce considerably the number of markers for association studies.

The data on population frequencies and haplotypes are available at the NCBI dbSNP database (http://www.ncbi.nlm.nih.gov/SNP/) [Sherry et al., 2001; Wheeler et al., 2004].

Dense SNP maps are also being created for model organisms including mouse, rat, fruit fly and *Arabidopsis thaliana*.

Technology development for SNP detection

The broad field of applications for SNPs induced a pressing need for effective instruments for SNP detection. During the last decade a considerable number of methods for SNP discovery (search for new SNPs) and detection (recognition of already known SNPs) were developed. Since SNP discovery has to be performed only once, stronger demands in terms of price and throughput are made for methods detecting already known SNPs, which are supposed to be applied for genotyping of large numbers of individuals.

Current state in the field of SNP detection

Several reviews have been published in the last few years discussing and classifying the existing SNP detection techniques [Shi, 2001; Nedelcheva Kristensen et al., 2001; Kwok, 2001; Syvanen, 2001; Kirk et al., 2002; Breen, 2002; Kwok and Chen, 2003;

Twyman and Primrose, 2003; Twyman, 2004]. P.-Y. Kwok describes principles and characteristics of different SNP discrimination approaches, detection mechanisms and assay formats. G.Breen concentrates on several interesting and novel methods, which from his point of view, have the potential to become widely used. The review of A.-C. Syvanen covers a lot of techniques, including those widely used and those just published, with regard to their suitability for high-throughput analysis. The article has a very thorough reference list, which itself is very informative. R.M. Twyman and S.B. Primrose observe patents concerning SNP technologies and speculate about their influence on commercially valuable SNP genotyping projects and on the further research in the SNP field.

Most SNP detection methods may be subdivided into two large groups according to the principles of SNP discrimination: hybridization-based and enzymatic-based methods. Somewhat apart stands the method of chemical ligation, which exploits the specificity of a chemical reaction.

Hybridization-based methods

This group of methods uses the property of a mismatched nucleotide to destabilize the DNA duplex. Usually two short allele-specific oligonucleotides (ASOs) are designed for each SNP locus with a nucleotide complementary to SNP variants in the middle of the sequence. ASOs are hybridized to the PCR-amplified SNP-containing DNA under stringent conditions, forming either stable (completely matched) or unstable (with a mismatch) duplexes. The stability of duplexes is analyzed using various strategies.

Microarrays

Microarray format permits to perform several hybridization reactions simultaneously. Since the location of elements on the microarray is known, the result of each reaction may be scored separately.

Early SNP-genotyping assays involved immobilized PCR-products and labeled ASOs in the solution – each ASO had to be hybridized to the array separately [Saiki et al., 1986]. The reverse scheme, hybridization on oligonucleotide microarrays has been known since the late 1980’s when a DNA sequencing method based on hybridization of a template with a universal set of oligonucleotides was under development [Lysov et al., 1988; Southern et al., 1992].

Modern microarrays may carry tens of oligonucleotides for each SNP [Hacia et al., 1998; Wang et al., 1998]. For each nucleotide position within an SNP-containing region (5-7 nucleotides), the array has four (or eight – for the analysis of both strands) elements with immobilized probes complementary to the sequence centered at this position. These probes have different nucleotide (A, G, C, T) in the central position. Elements corresponding to nucleotides around the SNP help to reveal unspecific hybridization. Such a microarray is the basis of the GeneChip assay offered by Affymetrix [Pease et al., 1994]. The HuSNP array contains several million elements with 25bp oligonucleotides (40 per SNP) and is capable of genotyping ~100000 human SNPs in one reaction. Hybridization on microarrays is also used for searching for unknown SNPs [Sapolsky et al., 1999; Wang et al., 1998; Hacia et al., 1998; Lindblad-Toh K et al. 2000]). Figure 1 shows that different allelic variants produce different hybridization patterns.

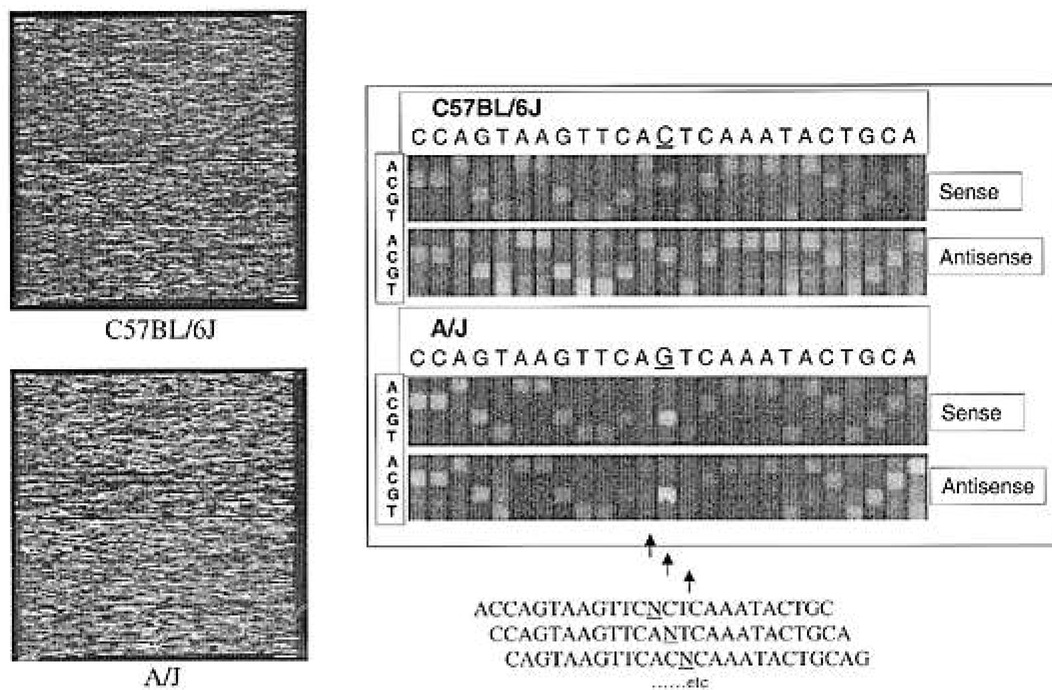

Figure 1. A G/C SNP in A/J and C57BL/6J mouse strains detected by hybridization to a high-density oligonucleotide array [Lindblad-Toh et al., 2000]. Probes: each position of the presented sequence was tested by four 25-nt oligonucleotides having all possible nucleotides (A, G, C, T) in the 13th position. Targets: STSs were PCR amplified, pooled, labeled and hybridized to an array, which was then stained and scanned. In the example shown, the indicated base is C in C57BL/6J and G in A/J.

To genotype known polymorphisms, only two probes corresponding to allelic variants of the SNP may be immobilized on the chip. This approach reduces the costs and increases the throughput of the assay. Such allele-specific sets of probes were used

to genotype a set of known SNPs in the *CFTR* gene [Cronin et al., 1996], the human tyrosinase gene [Guo et al., 1994] and in the cytochrome P450 gene [Cronin et al., 1998].

Although hybridization on microarrays provides a highly parallel genotyping, it exhibits a restricted reliability and accuracy. The specificity of hybridization depends not only on hybridization conditions, but also on the nucleotide sequences that flank SNPs, and on the secondary structures of the target sequences [Mir and Southern, 1999]. Thus, it is practically impossible to select conditions that would be optimal for all loci. In the microarray-based resequencing of the 21st human chromosome (regions without repeats were analyzed, ~22x10^6bp), only 65% of bases yielded data of high enough (97%) quality [Patil et al., 2001]. Mapping studies, in which 400-500 SNPs were analyzed using high-density ASO microarrays, failed to distinguish between heterozygous and homozygous SNP genotypes for about 15% of the SNPs [Wang et al., 1998 Cho et al., 1999]. Other weak points of the microarray approach are its complex target preparation procedure (involving PCR amplification of the regions of interest, pooling and processing of PCR amplicons) and the high expense of chip production.

To increase the reliability of SNP genotyping, the specificity of hybridization may be strengthened by some additional factor. It was reported that LNA (locked nucleic acid) or PNA (peptide nucleic acid) ASOs show higher affinity for complementary DNA and better match/mismatch discrimination than do conventional DNA ASOs. Although these ASO modifications have not yet been tested for highly parallel hybridization in microarray format, they were successfully used for genotyping. An ELISA-like assay (ELISA – enzyme-linked immunosorbent assay) using octamer LNA ASOs was developed and applied for genotyping of the factor V Leiden mutation [Orum et al., 1999] and for SNPs in the apolipoprotein E gene [Jacobsen et al., 2002] in clinical samples. Allele-specific PNA hybridization probes were also used for SNP discrimination, with subsequent detection and identification by matrix assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry [Griffin et al., 1997; Ross et al., 1997; Ren et al., 2004] or fluorescence reader [Gaylord et al., 2004].

DASH

A promising approach called DASH (dynamic allele-specific hybridization) suggests observing the stability of duplexes formed by ASOs over a temperature gradient [Howell et al., 1999]. This solves the problem of adjusting hybridization conditions for individual loci. The scheme of the DASH procedure is shown in Figure 2.

First, a single-stranded template is obtained: the SNP-containing region is amplified with one of the primers bearing biotin, PCR product is bound to streptavidine-coated wells of a microtitre plate and one strand is removed by the addition of alkali. Then an ASO is annealed to the template at a low temperature in the presence of a double strand-specific intercalating dye, such as Sybr Green I (Molecular Probes). During gradual heating the duplex becomes less stable and finally the probe melts away. Perfectly matched duplexes and duplexes with a single mismatch have different melting temperatures.

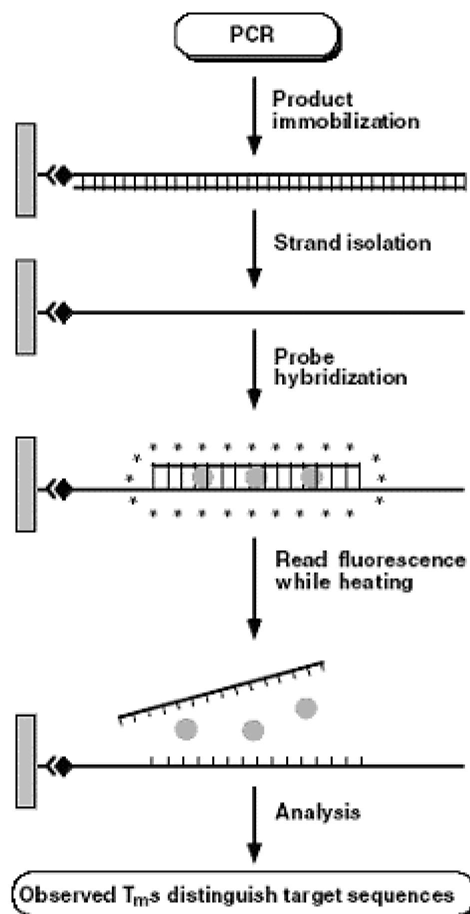

Figure 2. The DASH assay [Howell et al., 1999].

The fluorescence is proportional to the amount of double stranded DNA regions and decreases with denaturation of the duplexes (Figure 3).

Advantages of DASH are its simplicity (no enzymes are used in the procedure) and that it utilizes standard reaction conditions for all loci. The assay is suitable for parallel genotyping and may be easily automated. Disadvantages include the complexity of the template preparation procedure, the considerable costs for streptavidin-coated consumables, and the complexity of the primer design strategies required to overcome potential secondary structure problems. However, the work is going on to overcome

these drawbacks. A computer program was developed that automates the design process [Prince et al., 2001].

A variation of the DASH method uses electric field instead of temperature to denature the ASO-probe/target duplexes [Sosnowski et al., 1997; Heaton et al., 2001].

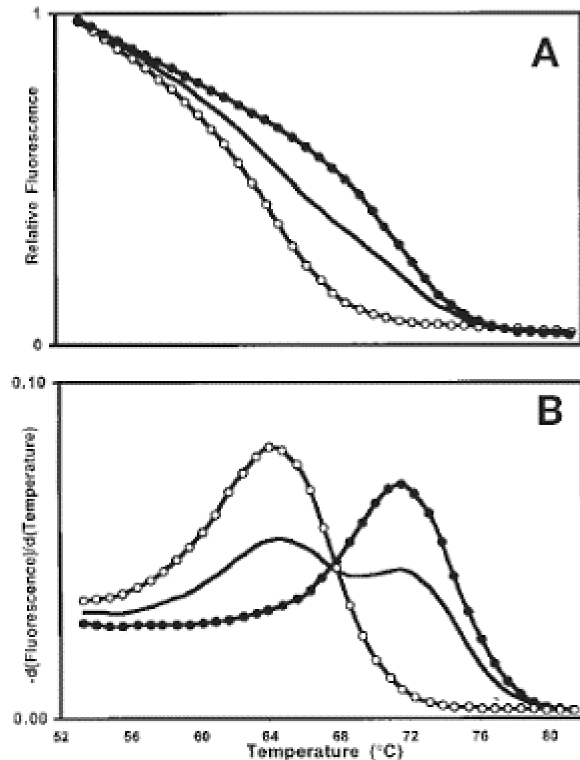

Figure 3. DASH was performed using two biotinilated oligonucleotide templates: 5'-ATTAGGG(T/C)GGGGAGA-3' (alternative nucleotides shown in parentheses) and the hybridization probe: 5'-TCTCCCC(G)CCCTAAT-3' [Howell et al., 1999]. (A) The resulting fluorescence changes against temperature (melting curve) and (B) the negative first derivatives are shown for the template perfectly matched to the probe (line with filled circles), for the template with a single-base mismatch (line with empty circles) and for 50/50 template mix (plain line).

Homogenous assays

The most widely used ASO hybridization-based SNP detection methods involve TaqMan (Applied Biosystems) [Livak, 1999] or Molecular Beacon probes [Tyagi et al., 1998]. Both methods exploit the principle of fluorescence resonance energy transfer (FRET) between two fluorophores (reporter and quencher). The quencher absorbs light emitted by the excitated reporter. The quenching effect depends on the distance between the fluorophores. If they are brought near each other – being located at the ends of a short oligonucleotide (as in TaqMan probes), or brought together during the formation of a stem-loop structure (Molecular beacons), – low fluorescence is observed at the donor emission wavelength. If they are separated, fluorescence increases.

In TaqMan assay, separation of fluorophores is performed during PCR amplification (Figure 4). TaqMan probes have blocked 3'-ends and anneal to the target sequence between PCR primers. When the extended PCR primer reaches the 5'-end of a TaqMan probe, *Taq* polymerase destroys it due to 5'-nuclease activity (TaqMan assay is also called the 5'-nuclease assay), and the fluorophores are separated from each other. If a

TaqMan probe is complementary to the target, the increase of fluorescence is monitored during amplification. A single-base mismatch in a TaqMan probe/target duplex makes the duplex unstable and prevents nuclease digestion of the probe. The use of two probes, each labeled with a different reporter fluorophore, allows both SNP alleles to be analyzed in a single tube.

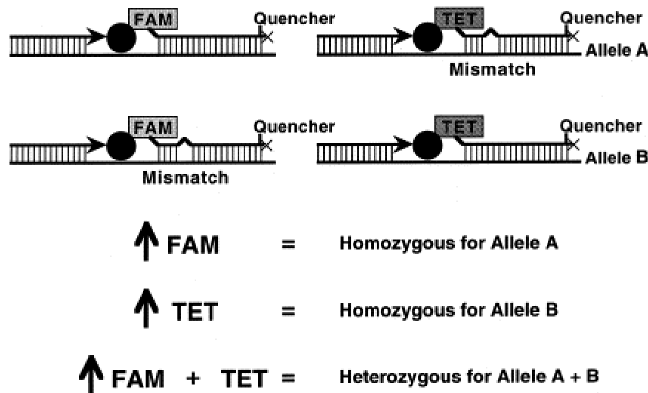

Figure 4. Allelic discrimination by 5'-nuclease assay (TaqMan). Mismatches influence the efficiency of probe destruction [Livak, 1999]. An increase of only Fam signal indicates that the analyzed sample is homozygous for allele A.

Molecular beacons are oligonucleotides with short sequences on the ends, which form stem-loop structures that hold a fluorescent reporter in close proximity to a quencher (Figure 5). When the loop portion of the molecular beacon is hybridized to a perfectly complementary target, the stem is opened, and the fluorescence is observed [Tyagi et al., 1998; Mhlanga and Malmberg, 2001].

The tendency to adopt a stem-loop structure, which additionally destabilizes mismatched hybrids, explains the higher discrimination power of Molecular Beacon probes when compared to TaqMan probes.

Molecular Beacon probes may be immobilized on microarray surfaces for the detection of unlabeled DNA targets [Steemers et al., 2000]

Both TaqMan and Molecular Beacons assays may be multiplexed using probes labeled with different fluorophores [Lee et al., 1999; Tyagi et al., 2000].

TaqMan and Molecular Beacon assays may be performed in 96-well or 384-well microtitre plates, and the fluorescence may be monitored either in real time or by end point detection after completion of the PCR. In the latter case, the results are expressed as a signal ratio that reflects the hybridization of the two oligonucleotides to the target sequence (Figure 6), so differences in amplification efficiency between samples do not affect interpretation of the genotyping results.

The main drawback of these methods is that two expensive fluorescently labeled probes are required for each SNP. In addition, assay conditions should be optimized for

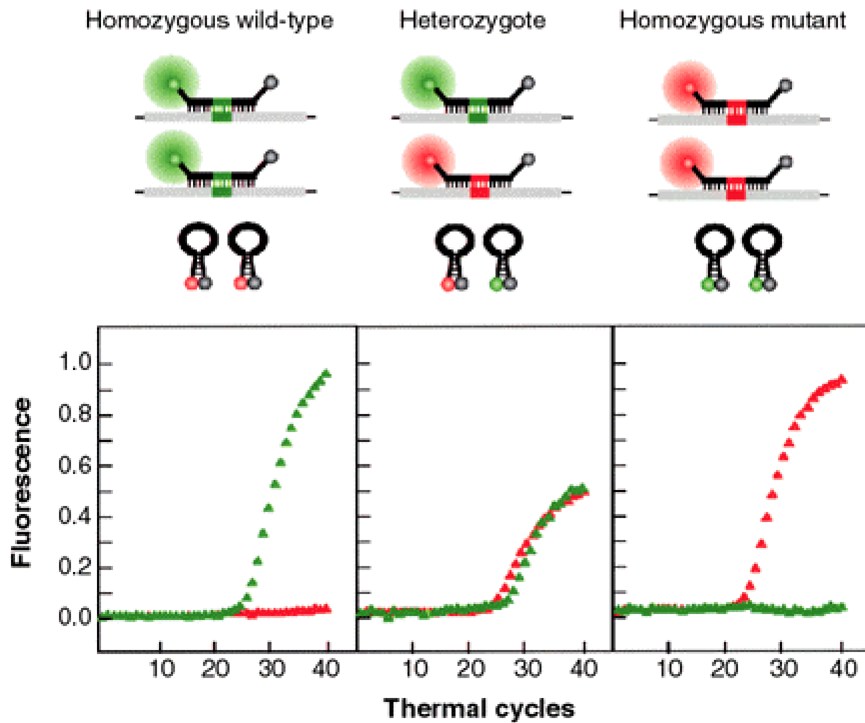

Figure 5. Principle of molecular beacons genotyping [Kostrikis et al., 1998]. In the case of wild-type homozygous DNA, only fluorescein (green)-labeled molecular beacons hybridize to the amplicons, generating green fluorescence, whereas the tetramethylrhodamine (red)-labeled molecular beacons retain their stem-and-loop structure and cannot produce a red fluorescent signal. In the case of heterozygous DNA, both molecular beacons hybridize to the amplicons and generate both green and red fluorescence. In the case of homozygous mutant DNA, only the tetramethylrhodamine-labeled molecular beacons hybridize to the amplicons, generating red fluorescence, whereas the fluorescein-labeled molecular beacons remain dark.

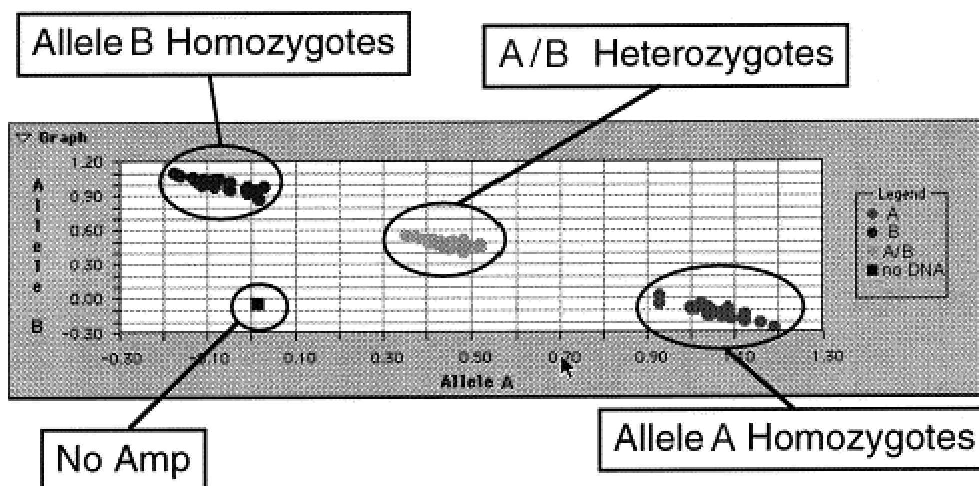

Figure 6. Allelic discrimination by 5'-nuclease assay (TaqMan). End point signal distribution for 43 independent genotypings [Livak, 1999].

each locus. However, both TaqMan and Molecular Beacon assays are simple to perform and need no post-PCR processing. The closed-tube format of these assays minimizes the cross-contamination. These features make them methods of choice for clinical genotyping assays.

Enzymatic methods

SNP discrimination by polymerase

Allele-specific PCR

Allele-specific PCR exploits the specificity of polymerases: PCR amplification fails if the 3'-nucleotide of a PCR primer is not complimentary to the template. Two allele-specific primers (with different 3'-ends) and one common PCR primer are required per biallelic SNP. Genotyping involves two PCR reactions with different allele-specific primers [Newton et al., 1989]. The discrimination power of this approach may be improved by introducing an additional mismatched nucleotide in the second or third position of the allele-specific PCR primers or by performing a competitive PCR reaction in each tube [Zhu and Clark, 1996]. Another possibility is to perform bi-directional analysis in the same tube [Sasvari-Szekely et al., 2000].

Early approaches were coupled with gel electrophoresis for product analysis. Introduction of a DNA-intercalating dye and discrimination of allelic variants by melting curve analysis in a single tube eliminated this laborious and contamination-dangerous step [Germer and Higuchi, 1999]. To facilitate the discrimination of allelic variants, a GC-rich tag is added to one of the allele-specific primers to increase the difference between melting temperatures of PCR products.

Kbiosciences suggested an Amplifluor genotyping assay based on allele-specific PCR (Figure 7). This method involves allele-specific PCR primers with different 5'-tails. During the first several cycles of PCR, allele-specific products appear, and annealing sites for universal energy-transfer-labeled primers are formed. Then energy-transfer-labeled primers begin to participate in the reaction. During PCR, molecular beacon-like stem-loop structures on the 5'-ends of the energy-transfer-labeled primers are opened and fluorescence corresponding to the allelic variant is monitored. The nice feature of this assay is that it uses universal fluorophore-bearing primers that are suitable for any locus.

The same principle (formation of allele discrimination structures which are not present in the reaction at the beginning) is used in the DzyNA-PCR approach [Todd et

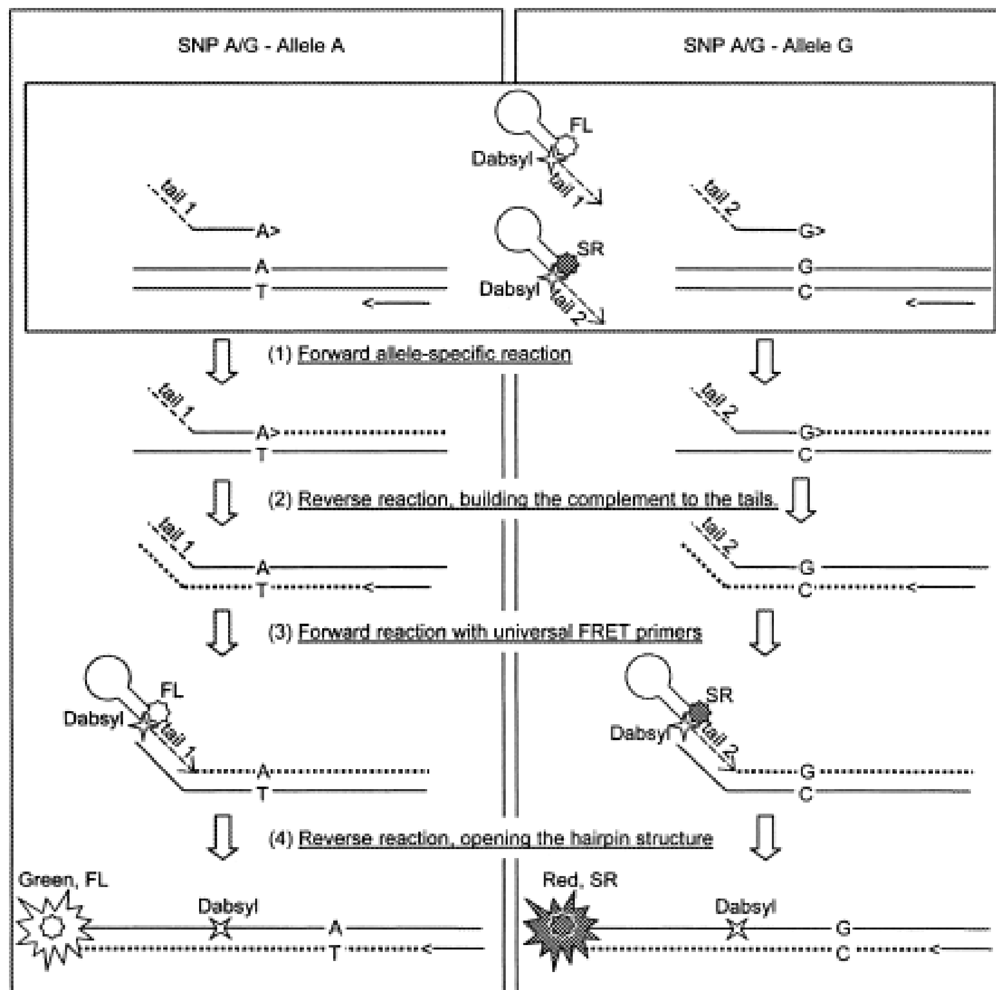

Figure 7. Scheme of the Amplifluor genotyping assay for genotyping of A/G SNP. FL is fluorescein. SR is sulforhodamine [Myakishev et al., 2001].

al., 2000]. The method is based on the formation of a DNAzyme during allele-specific PCR (the antisense sequence of a DNAzyme is included in the tail of the allele-specific primer). DNAzyme destroys the reporter substrate (containing reporter and quencher fluorophores) present in the reaction, and the released fluorescence is monitored. The limitation of this approach is that two parallel reactions have to be performed for each SNP.

Detection of allele-specific PCR products for many loci may be performed in parallel on a Homogeneous AlphaScreen (Packard Bioscience) proximity assays gel by introducing tails of different length in allele-specific primers [Beaudet et al., 2001]. Size separation in a high throughput mode might be accomplished by using 96-channel

capillary sequencing instruments, microplate array diagonal electrophoresis [Ye et al., 2001], or capillary array electrophoresis microplates [Medintz et al., 2001].

Using of proofreading (possessing 3'→5' exonuclease activity) DNA polymerases for SNP discrimination by allele-specific PCR was suggested recently [Zhang and Li, 2003]. Proofreading polymerases are much more sensitive to mismatches in the 3'-region of the PCR primer if compared to (3'→5' exo-) polymerases. Exonuclease-resistant (3'-phosphorothioate) primer cannot be extended by a proofreading *Pfu* polymerase if it contains a mismatch within the 8nt region on its 3'-end.

Although allele-specific PCR methods are simple to perform, all of them (including Amplifluor genotyping assay, which is now one of the cheapest commercially available methods) require optimization of reaction conditions for each SNP. Multiplexing possibilities are also very limited [Kwok, 2001].

Primer extension

Like allele-specific PCR, primer extension is based on the accuracy of DNA polymerases. However, separation of the amplification step from allelic discrimination results in a better specificity and accuracy of the reaction. In general, primer extension SNP detection techniques include amplification of an SNP-containing region and a primer extension reaction, where the amplicon is used as a template. Primers may be designed to anneal immediately upstream of the SNP locus (for single nucleotide primer extension or "minisequencing") or allele-specific primers with 3'-nucleotide corresponding to the SNP may be used for each locus (allele-specific primer extension). Nucleotide incorporation may be performed on a solid support or in solution. Primer-extension reactions may be highly multiplexed.

Primer extension on microarrays

Arrayed primer extension (APEX) on an oligonucleotide microchip uses the principle of Sanger dideoxy sequencing in a parallel and high-throughput microarray format (Figure 8a) [Pastinen et al., 1997; Raitio et al., 2001; Tonisson et al., 2002]. An amplified DNA sample is fragmented enzymatically and annealed to locus-specific primers that are immobilized on a glass slide and have 3'-ends free for extension [Lindroos et al., 2001]. One [Shumaker et al., 1996] or two (one for the sense and another for the antisense strand [Kurg et al., 2000]) primers adjoining the SNP position are used per locus. Extension reactions are performed by *Taq* polymerase with fluorescently labeled dideoxynucleotides. Compared to the ASO-hybridization

approach on microarrays, APEX is more precise and provides ~10 fold better

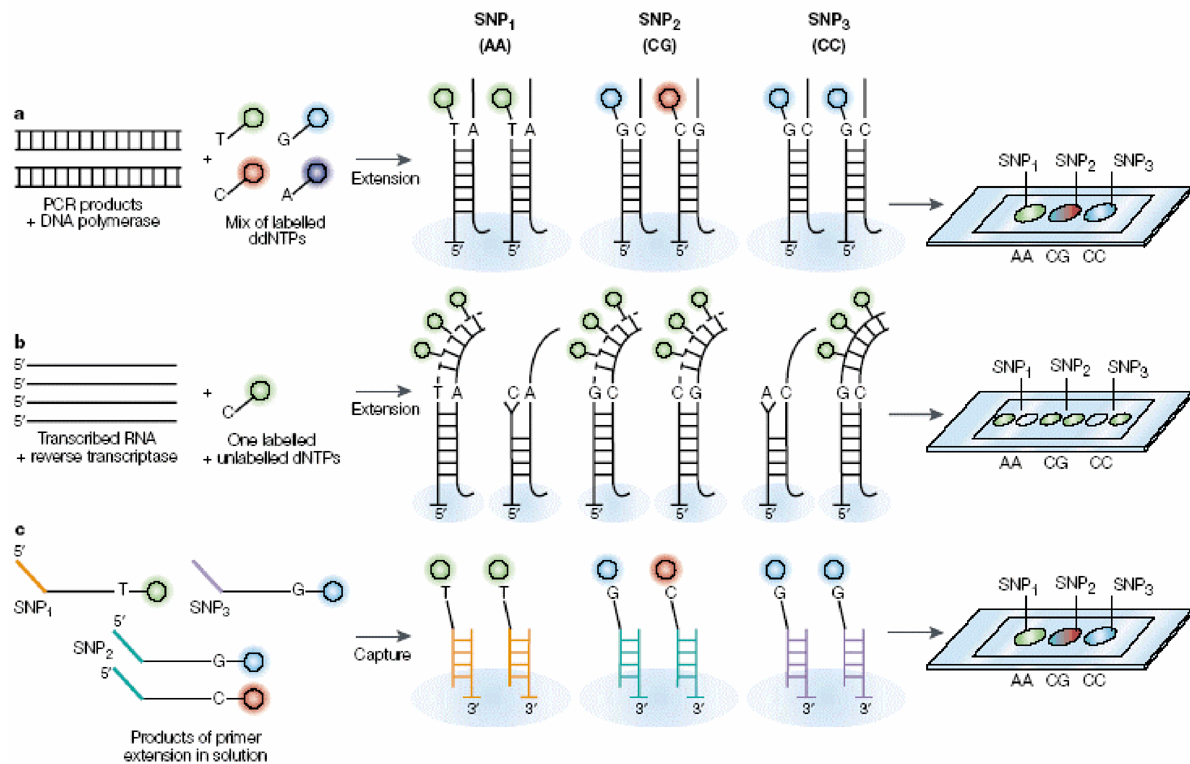

Figure 8. SNP genotyping on microarrays [Syvanen, 2001]. a) Arrayed primer extension: locus-specific oligonucleotides immobilized on the microarray are hybridized to the SNP-containing targets upstream of the SNP position and are terminated by fluorescently labeled dideoxynucleotides complementary to the nucleotide in the SNP position. b) allele-specific primer extension: allele-specific oligonucleotides immobilized on the microarray are hybridized to the RNA molecules transcribed from SNP-containing PCR products; depending on the SNP variant, they are either extended or not by reverse transcriptase. c) Primer extension of locus-specific oligonucleotides bearing universal tags is performed in solution by fluorescently labeled dideoxynucleotides; terminated primers are then hybridized to the universal microarray.

discrimination power. Using the principle of "array of arrays" allows for analyzing several samples on one microarray simultaneously, which increases the throughput of the method [Pastinen et al., 2000].

Array-based primer extension reactions are relatively cheap because a universal extension mix containing dye-labeled nucleotides and polymerase is used for all SNPs. However, the thermal cycling is not as easily performed on the solid phase as it is in solution. Another weak point is preparation of single-stranded templates, which requires a lot of amplified material. The latter inconvenient and laborious step can be avoided by introducing the sequence of the T7 RNA polymerase promoter into one of the PCR primers [Pastinen et al., 2000]. It becomes then possible to generate single-

stranded RNA molecules from PCR-products. Allele-specific primers (with 3'-nucleotides corresponding to the SNP) are extended on array by reverse transcriptase (Figure 8b).

Another array-based strategy that is less dependent on the hybridization peculiarities of the locus-specific sequences is the use of universal tags – artificial sequences on the 5'-end of the extension primer. In this approach, the extension is performed in solution and extended primers are hybridized to an array of oligonucleotides complementary to the tags (Figure 8c).

The "coding" of analyzed loci results in a better specificity of hybridization since tag sequences are carefully designed to provide optimal annealing for all targets. Besides, the same set of tag sequences and the same microarray may be used for any set of SNP loci, which makes the technology flexible and cheap [Hirschhorn et al., 2000; Fan et al., 2000]. Another important aspect of using tags is that hybridization and extension of locus-specific oligonucleotides are performed in solution. Homogenous enzymatic reactions are in general more effective. Tag-coding strategy is realized in the GeneFlex microarrays of Affymetrix and in the xMAP technology of Luminex. The latter suggests polystyrene microspheres as a solid phase, which are filled with fluorescent dye. Mixing different ratios of red and infrared fluorophores makes it possible to obtain batches of microspheres with different spectral signatures [Taylor et al., 2001]. Each microsphere carries a unique tag for capturing the products of primer extension. After capturing, microspheres are analyzed in a flow cytometer, which determines the type of microsphere and the genotype of the corresponding locus [Cai et al., 2000; Chen et al., 2000]. The important advantage of microspheres is that they approximate the hybridization conditions of those in the solution.

Homogenous primer extension assays

Non-microarray primer extension methods use different detection approaches.

Genetic bit analysis [Nikiforov et al., 1994], introduced to the market by Orchid Biosciences, is a colorimetric assay. Immobilized primers are extended with biotin- or fluorescein- labeled dye-terminators. Antibodies recognize the incorporated haptens, and enzymes that are conjugated to the antibodies catalyze the formation of a colored product. This detection strategy in a 384-well microtitre plate format has been automated for production-scale SNP genotyping (SNPstream, Orchid Biosciences).

SNaPshot assay (Applied Biosystems) is a simple and robust method: primers allele-specifically terminated with fluorescently labeled dideoxynucleotides are

detected in a gel. Modified primers with 5'-tails of different length allow moderate multiplexing in the range of tens of SNPs per reaction [Kuppuswamy et al., 1991; Lindblad-Toh et al., 2000]. By using 96-channel capillary sequencers this assay can be automated for high-throughput genotyping.

The template-directed dye-terminator incorporation assay (TDI) is based on FRET detection [Chen and Kwok, 1997]. In this method the primer is 5'-labeled with the donor dye (fluorescein) and one of the ddNTPs are labeled with an acceptor dye (6-carboxy-X-rhodamine, ROX). FRET occurs when the dye-labeled ddNTP is incorporated onto the sequencing primer in the presence of DNA polymerase and target DNA. The genotype of the target DNA molecule is determined by exciting the fluorescein dye on the sequencing primer and seeing if the acceptor dye exhibits FRET.

In the FP-TDI assay the FRET detection is substituted for fluorescence polarization (FP) detection [Chen et al., 1999]. When a fluorescent molecule is excited by plane-polarized light, it emits polarized fluorescent light into the same plane if the molecules remain stationary between excitation and emission. In the FP-TDI assay ddNTPs labeled with different fluorophores are incorporated onto the unlabeled primer. Unincorporated ddNTPs are small molecules, they rotate and tumble and FP is largely lost (depolarized). Incorporated ddNTPs acquire an additional weight of the primer and FP is preserved.

Primer extension with MALDI-TOF detection

Several primer extension techniques have been successfully combined with mass spectrometry (MALDI-TOF, matrix-associated laser desorption time-of-flight mass spectrometry) as a detection platform. In general, in the MALDI-TOF procedure products of the primer extension are mixed with a special matrix and are transferred into gas phase by a short laser pulse and ionized by applying a strong potential difference. Ions are accelerated in an electric field and the time of their flight (till they reach the detector) is measured. The time depends on the mass and charge of the ions, which are allele-specific. One of the first procedures was the PROBE (primer oligo base extension) assay [Braun et al., 1997; Little et al., 1997; Little et al., 1997a] marketed by Sequenom, Inc. as the MassEXTEND technology. The method involves primer extension with one of four nucleotides in the dideoxy form, so that the primer is either terminated by ddNTP at the SNP position, or several dNTPs are incorporated, depending on the template sequence. Thus allele-specific products differ distinctly in length and mass.

The PINPOINT assay [Haff and Smirnov, 1997; Ross et al., 1998] uses four ddNTPs for primer extension. SNP is defined according to the mass of the incorporated nucleotide. This method is suitable for multiplexing if the extension products for different loci have distinguishable masses. The main problem of this method is that ddA and ddT have a mass difference of only 9Da, which results in poor discrimination of this pair. The mass difference may be increased by modified mass-tagged ddNTPs [Fei et al., 1998; Mengel-Jorgensen et al., 2004]. Another way is to cleave the common part of the extended primer so that smaller allele-specific products, which could be analyzed more specifically, would remain [Li et al., 1999; Vallone et al., 2004].

A compromise between PROBE and PINPOINT assays is the VSET (very short extension) method [Sun et al., 2000]. Here extension mixture contains three ddNTPs and only one nucleotide in deoxy form, so that primer is terminated either at the SNP site or the site next to the SNP. This results in short and distinguishable products.

MALDI-TOF requires thoroughly purified samples free of ions and other impurities. That means introducing time-consuming and expensive purifying steps. In the 'GOOD' assay [Sauer et al., 2000] primer extension products (3'-end thio-modified oligonucleotide is extended with α-S-dNTPs or α-S-ddNTPs) are modified (thiol groups are alkylated) to carry either one positive or one negative charge, which increases the sensitivity of the analysis. Less analyte is taken for the assay and no purification before MALDI-TOF analysis is required [Sauer et al., 2000; Li et al., 1999].

Pyrosequencing

Pyrosequencing is a DNA sequencing method based on real-time monitoring of the pyrophosphate (PPi) formation during nucleotide incorporation [Ronaghi et al., 1996]. SNP-containing sequence is amplified, one strand is removed and the primer is annealed upstream of the SNP position. Primer extension is performed with stepwise addition of four different deoxynucleotides. In the case that the nucleotide is incorporated by polymerase, the released pyrophosphate is enzymatically converted into ATP, which is utilized in a luciferase-based light-producing reaction. Based on the light, the specific nucleotide incorporated at the SNP position can be determined. Unused ATP and deoxynucleotides are degraded by the apyrase, and the subsequent nucleotide may be added to the reaction (Figure 9).

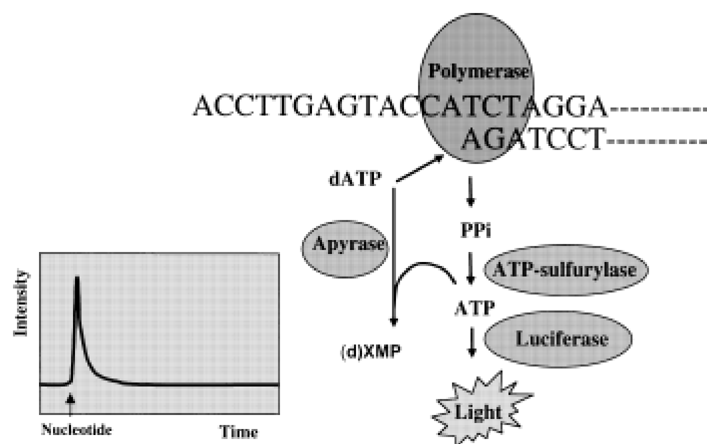

Figure 9. The chain of enzymatic reactions in liquid-phase pyrosequencing [Ronaghi, 2001]. The graph shows the increase in light intensity upon addition of the nucleotide. (d)XMP indicates one of the four degraded nucleotides or AMP.

Generated light is proportional to the number of incorporated nucleotides. The relationship is very accurate; therefore, pyrosequencing is suitable for SNP analysis not only in diploid but also in polyploid genomes [Rickert et al., 2002].

The main advantage of this method is that comparatively long sequences (~50bp and longer) may be determined. This is convenient for the genotyping of a number of closely spaced SNPs (HLA typing) and for detection of insertions/deletions [Guo et al., 2003]. The method also gives freedom in selecting the position of the sequencing primer and has proved to be very accurate [Ronaghi, 2001; Ahmadian et al., 2000; Chen et al., 2003; Alderborn et al., 2000]. To increase specificity and obtain longer reads in pyrosequencing, several improvements were recently suggested: increasing working temperature of the reaction [Eriksson et al., 2004], blocking the 3'-ends of the template DNA to prevent unspecific nucleotide incorporation [Utting et al., 2004], and using Sequenase instead of Klenow fragment of DNA polymerase I [Gharizadeh et al., 2004]. The disadvantage of pyrosequencing is the high price of genotyping given that the reaction requires several enzymes.

Instrumentation for pyrosequencing in 96-well or 384-well microtitre plate formats and kits are available at Pyrosequencing AB. Pyrosequencing on miniaturized arrays is also presently being developed [Ehn et al., 2004].

In general, primer extension is a simple, robust and accurate SNP discrimination step, which may be combined with different detection techniques and adjusted to different genotyping scales. However, it requires PCR amplification of SNP-containing regions and preparation of a single-stranded template for primer annealing. In homogenous procedures, PCR-primers and nucleotides need to be removed so that they do not interfere with primer extension.

SNP discrimination by ligase

The adaptability of ligation for SNP detection was first described in 1988 by U.Landegren [Landegren et al., 1988]. The author demonstrated that single-nucleotide mispairing at the junction of annealed oligonucleotides prevents ligation.

It was shown that ligase is more sensitive to the mismatches occurring at the 3'-side rather than at the 5'-side of the junction [Wu and Wallace, 1989], so usually two oligonucleotides with different 3'-nucleotides corresponding to the SNP position and one oligonucleotide bearing a phosphate on the 5'-end are used to determine the state of each biallelic SNP. The principle of ligation-based discrimination of SNPs is shown in Figure 10.

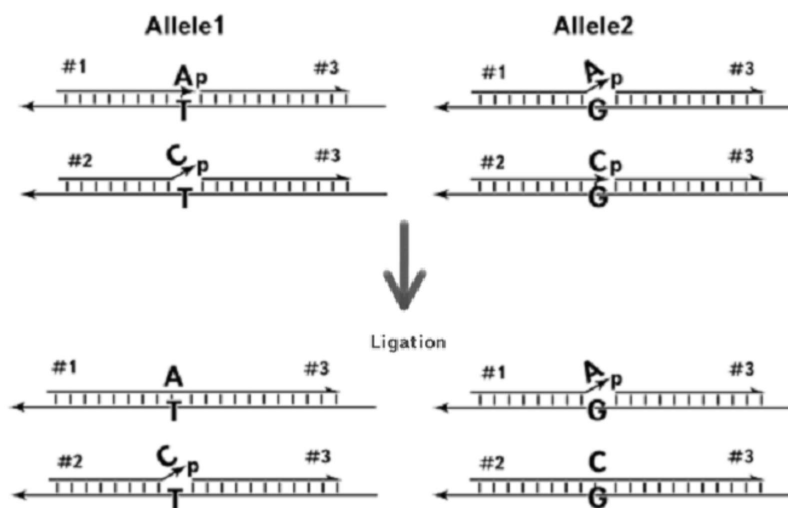

Figure 10. Principle of the ligation detection. Two allele-specific oligonucleotides #1 and #2, differing by 3'-nucleotide corresponding to the SNP and oligonucleotide #3, are annealed to the template. During ligation only perfect nicks are closed. Mismatched oligonucleotides remain unligated.

One of the first oligonucleotide ligation assays (OLAs) [Nickerson et al., 1990] used two biotinilated allele-specific probes and a reporter probe with digoxigenin on the 3'-end, which were annealed to the amplified target DNA and ligated. Biotinilated probes were then captured on the streptavidine-coated wells of the microtitre plate and ELISA (enzyme-linked immunosorbent assay) was performed to determine whether they were coupled to digoxigenin-labeled probes. Since only one type of label was used, allele-specific probes were added to the target DNA separately. If allele-specific probes are differently labeled, then they can be combined in one tube. In the dual-color ligation assay [Samiotaki et al., 1993], a set of three oligonucleotides is added to the amplified target DNA: allele-specific oligonucleotides labeled with europium or therbium chelates and common oligonucleotide having biotin at the 3'-end. Biotinilated

ligation products are captured on a streptavidine coated pin capture manifold. After washing the manifold, the fluorescence is measured in each reaction tube. Allele-specific ligation products may also be distinguished by gel electrophoresis if allele-specific probes carry adapter sequences of different sizes on the 5'-ends [Barany, 1991].

Multiplexing of the ligation reaction is achieved more easily than multiplexing of PCR because normally, DOs don't interfere with each other during ligation; therefore, a number of SNP-loci may be analyzed simultaneously in one reaction mixture. Ligated products may be further analyzed in a sequencing gel [Grossman et al., 1994; Schouten et al., 2002] or captured by hybridization to generic tags immobilized on arrays [Gerry et al., 1999; Favis et al., 2000; Consolandi et al., 2004] or microparticles [Iannone et al., 2000]. Ligation itself may be performed on a solid phase [Lizardi et al., 1998].

Early assays involved T4 DNA Ligase as a key enzyme. F.Barany suggested using thermostable *Tth* ligase for ligation-based SNP discrimination and demonstrated the sensitivity and specificity of this enzyme [Barany, 1991]. Thermostable ligases allow for performing ligation at high temperatures, which increases the specificity of oligonucleotide hybridization. Besides, multiple ligation cycles may be performed, resulting in higher assay sensitivity.

Some kind of signal amplification is used in all ligation-based methods because of the small content of any particular locus in genomic DNA (e.g. there is less than 1amol of a unique sequence per microgram of human genomic DNA). One approach is based on pre-amplification of regions of interest and on subsequent direct detection of ligated DOs (Chen et al., 1998; Samiotaki et al. 1994; Favis et al. 2000; Khanna et al., 1999; Pickering et al., 2002). Another is based on performing the ligation detection reaction directly on genomic DNA and on subsequent amplification of ligated DOs. Amplification may be performed by PCR [Barany et al., 2001], rolling circle amplification (RCA) [Faruqi et al., 2001; Lizardi et al., 1998; Hardenbol et al., 2003], and ligation chain reaction (LCR) [Barany, 1991].

Both pre- and post-amplification approaches have their own advantages. In the first case it is possible to use labeled (by fluorophore or hapten) "left" DOs for the discrimination of SNP loci. Disadvantages of this approach are: (i) preliminary PCR-amplification should be performed for each locus; (ii) a relatively large amount of labeled site-specific oligonucleotides is used for each reaction, which may be comparatively expensive; and (iii) working with amplified material is dangerous because of possible contamination. An interesting approach to overcome the last

problem is to perform PCR amplification in the presence of ligation probes and ligase [Chen et al., 1998]. The lower melting temperature of ligation probes compared to PCR primers prevents their ligation during PCR. After PCR, ligation is performed at lower temperatures.

In the case of "post-amplification", only nucleotide sequences may be used as detector markers for "left" oligonucleotides (fluorophores and haptens can't be copied during amplification). This means that ligation probes should be longer. On the other hand, a relatively small amount of each DO may be used in the detection reaction, thus compensating for their price.

Alternatives to linear DOs are padlocks – oligonucleotides forming circles upon ligation. Padlocks carry locus-specific sequences on both ends and a spacer sequence in the middle. During ligation, the padlock probe anneals to the template so that its 3'- and 5'-ends are brought together at the SNP site. If the base pairing is perfect, the ligase closes the nick and a circular oligonucleotide is formed [Nilsson et al., 1994; Baner et al., 1998]. Using padlocks increases the efficiency of ligation since the local concentration of ligating ends is high – after hybridization of one end this is practically an intramolecular reaction. Two allele-specific padlocks are required for each SNP locus. Alternatively, one padlock is used, which leaves a small gap around the SNP. The gap is filled either by polymerase or by ligating a small discriminating oligonucleotide. Such a step not only reduces the price but also increases the specificity of the circle formation [Lizardi et al., 1998].

Circular padlocks may be amplified with primers located in the middle spacer. PCR amplification may be replaced by rolling circle amplification (=rolling circle replication, RCA or RCR) performed at a constant temperature by a strand displacing DNA polymerase without dissociation from the template [Nilsson et al., 1997; Lizardi et al., 1998; Nilsson et al., 2002]. As a result, a long single-stranded DNA molecule appears, consisting of repeated copies of the padlock (Figure 11).

Introducing a second primer to RCA increases the yield of the reaction and leads to formation of branched structures [Zhang et al., 1998]. An elegant SNiPer system (Amersham Bioscience) involves allele-specific second primers with stem-loop structures with a fluorophore and quencher on 5'-ends. During branching, the stem-loop structure is destroyed. The fluorophore is thus separated from the quencher and the fluorescence is registered.

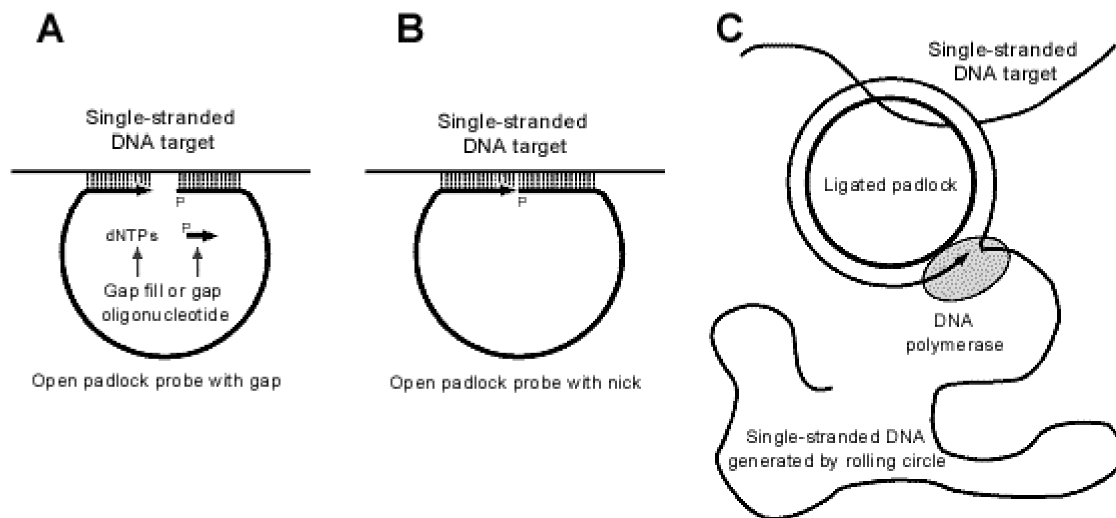

Figure 11. Padlocks (a modification of the published figure [Lizardi et al., 1998]). Padlocks may be circularized either by (A) ligation after gap filling (by a polymerase or by a gap oligonucleotide) or (B) ligation of adjacent ends of a padlock probe. (C) Rolling circle amplification (RCA).

To increase the specificity of the padlock circularization and prevent misamplification of unligated probes, a 3'-hairpin may be introduced into the padlock [Alsmadi et al., 2003]. During RCA, this hairpin is extended, resulting in an inert double-stranded structure, which is unlikely to participate in faulty priming reactions.

Highly multiplexed genotyping (1500 loci) has been reported using molecular inversion probes (MIP) [Hardenbol et al., 2003]. Padlock probes with locus-specific generic tags in the linker part (one probe per locus) are annealed to genomic DNA, leaving a single-nucleotide gap at the SNP position. Annealing is performed in four tubes in the presence of a thermostable ligase, polymerase, and one of four nucleotides. If a nucleotide corresponding to the SNP is present in the solution, the gap is filled, and the ends of the padlock are ligated. Unligated padlocks are destroyed by exonucleases. Ligated padlocks are released from genomic DNA by destroying the uracil-residue in the linker part of the probe by uracil-N-glycosylase. The padlocks are thus inverted: now locus-specific ligated parts are surrounded by parts of the linker sequence. The latter correspond to the amplification primers, which are common for all probes. During amplification, a fluorescent label is introduced into the PCR products and the SNP status of each locus is then determined by hybridization on a universal microarray.

In an alternative protocol suggested by the same laboratory, a single ligation reaction is used to determine genotypes [Baner et al., 2003]. The protocol requires two allele-specific padlock probes per locus. Padlocks don't contain uracil residues: topological inhibition of PCR is resolved by fragmenting the genomic DNA before padlock probes hybridization and ligation. This strategy allows performing cyclic

ligation, which increases the circularization yield. The detection scheme is the same as for the MIP technology.

In SNPlex assay, commercially available from Applied Biosystems, linear probes (three probes per locus) are ligated. Unligated primers are removed enzymatically. The ligated product is amplified by PCR with common primers, one of which is biotinylated. PCR products are attached to the streptavidine-coated surface. One strand is removed, and a mix of ZipChute Mobility Modifier probes is hybridized to the other strand. Unbound probes are washed away and hybridized probes are eluted and analyzed by electrophoresis: allele-specific probes differ by mobility and bear a locus-specific fluorophore. The company provides kits capable of multiplexing 48 SNPs in a single reaction.

The authors of the amplified fragment length polymorphism (AFLP) technique suggested the SNPWave technology, now commercially available at Keygene NV [Michiel et al., 2004]. The technology employs multiplex ligation of allele-specific padlock probes on genomic DNA. Subsets of ligated products may be selectively amplified with PCR primer pairs differing by two 3'-end nucleotides (the same principle as the selective amplification utilized in AFLP). PCR products of each subset are labeled with the same fluorophore. To discriminate loci and alleles, length stuffers are incorporated in all padlocks to provide spacing of 2bp between alleles and 3bp between loci. Analysis is performed in a sequencing gel.

The ligation detection reaction is (i) specific, (ii) allows high level of multiplexing, and (iii) is appropriate for alternative amplification techniques: rolling circle amplification, branched rolling circle amplification, ligation chain reaction. Unlike primer extension- and hybridization-based techniques, ligation detection may be performed directly on genomic DNA.

Invasive cleavage

Allele-specific invasive cleavage is based on the ability of flap endonucleases to cleave a flap structure formed by two oligonucleotides overlapping at the SNP site [Lyamichev et al., 1999].

Two oligonucleotides - Invader Oligo and signal oligonucleotide (Probe 1) are hybridized to the SNP-containing target nucleic acid. Probe 1 has a non-complementary allele-specific 5'-tail. Cleavage occurs if the nucleotide in Probe 1 (marked red in the

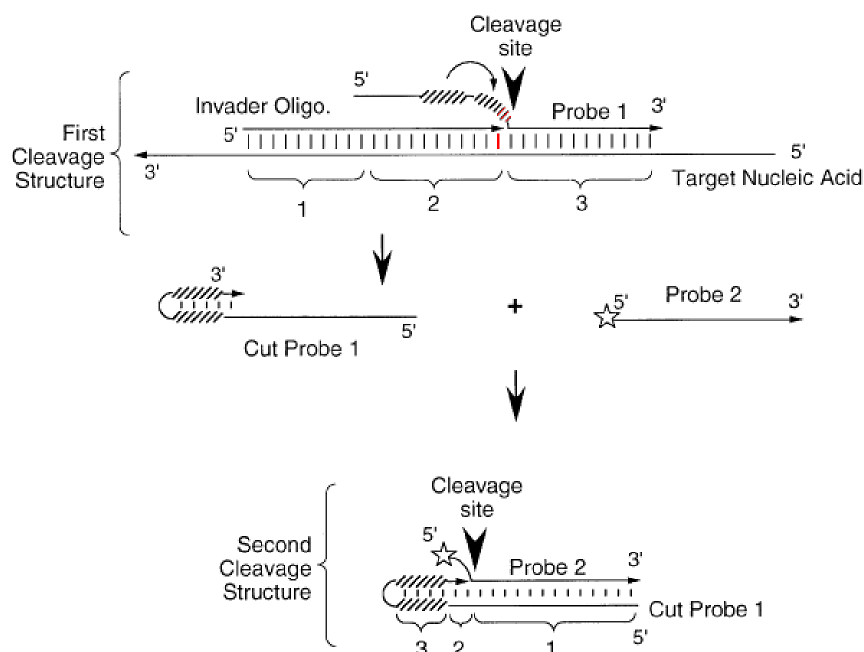

Figure 12. The scheme of the Invader assay [Hall et al., 1999].

picture), which overlaps with the Invader Oligo, is complementary to the template. Interestingly, the nucleotide in the 3'-terminal position of the Invader oligo does not influence the cleavage. The cleaved oligonucleotide (Cut Probe 1 on Figure 12) may be detected by ELISA [Lyamichev et al., 1999], electrophoresis [O'Connell et al., 1999; Sander et al., 1999; Oldenburg and Siebert, 2000] or MALDI-TOF [Griffin et al., 1999]. It may also serve as an invasive nucleotide and the template for the fluorescently labeled universal reporter Probe 2 (Figure 12) in a serial invasive signal amplification reaction [Hall et al., 2000].

Using thermostable flap-endonucleases allows for performing the reaction at a constant temperature close to the melting temperature of the oligos, enabling several rounds of cleavage of the probe for each locus. High sensitivity and the isothermal mode of the invasive cleavage make this approach very attractive.

The feasibility of the Invader assay on a solid support (microbeads) was recently showed [Rao et al., 2003]. The assay involves invasive oligonucleotides and allele-specific signal probes bearing a fluorophore and a quencher on both ends immobilized on microbeads. Microbeads are incubated with genomic DNA (a comparatively large amount of ~25μg). Flap endonuclease cuts the signal probe and fluorescence is detected by flow cytometry. This work is regarded as a step for the multiplex Invader assay.

The Invader assay is commercially available from Third Wave Technologies.

Site-specific cleavage

The first method to detect SNPs was restriction fragment length polymorphism (RFLP) analysis, which exploits the specificity of restriction endonucleases. The SNP containing region is amplified and the PCR product is cut by a corresponding restriction endonuclease. Digested fragments are analyzed by gel electrophoresis.

This method is suited for and also restricted to SNPs which alter the recognition site so that the restriction enzyme cuts one of the SNP alleles without cutting the other [Schumm et al., 1988]. The restriction site may be created artificially by using mutagenic primers.

A modification of RFLP, called melting curve SNP (McSNP) genotyping, utilizes the difference between melting temperatures of intact and digested PCR-products. With this method, the time- and labor- consuming gel electrophoresis step is substituted by closed-tube melting curve analysis [Akey et al., 2001].

Chemical ligation

An attempt to exclude enzymes for SNP detection was described in [Xu and Kool, 1999]. More than100-fold selectivity against a single mismatch at either side of the junction was observed in the phosphorothioate-iodide DNA autoligation reaction, both for linear ligation probes and for a padlock. Autoligation occurs between phosphorothioate at the 3'-end of the junction and the iodide group on thymidine at the 5'-end of the adjacent oligonucleotide.

* * *

The spectrum of SNP genotyping methods is extremely broad, because basic SNP discrimination mechanisms (hybridization-, enzymatic- and chemical ligation-based) may be combined with different detection strategies and assay formats. Detection approaches include colorimetry, mass spectrometry, gel electrophoresis analysis, measurement of radioactivity, fluorescence, or chemiluminescence. Genotyping may be performed in homogenous (in solution) or heterogeneous (involving both a liquid and a solid phase) formats. The solid phase may be a glass slide, a plastic chip, or latex microbeads. Single locus or a number of loci may be analyzed in one reaction. Some of

the assays require prior amplification of the SNP-containing genomic region, while others were modified to work directly on genomic DNA.

Despite this variety, there is no "golden standard" method in this area, as the Sanger method is for sequencing. Each method has its own advantages and limitations. The choice of genotyping approach depends on the aim of a particular project. The following characteristics of SNP genotyping methods must be considered for the selection of the optimal approach:

- setup price (locus/allele-specific oligonucleotides; locus-specific chips, beads, probes, etc.);
- price of the genotyping reaction itself;
- accuracy (percent of correctly determined SNPs);
- reproducibility;
- sensitivity (amount of DNA sufficient for the analysis);
- throughput (number of genotypings per day per person);
- complexity of handling;
- resistance against contamination;
- equipment requirements;
- amenability for automation;
- need for optimization of the assay for each locus.

When hundreds of thousands of SNPs are to be screened in many individuals (e.g. in association studies), the throughput capabilities without individual optimization and price of the genotyping technology are critical. Therefore, methods providing parallel genotyping and/or parallel detection on microarray or microbeads should be selected (Affymetrix chips, xMAP technology of Luminex; MIP approach [Hardenbol et al., 2003]).

For small- and middle-scale research applications, the choice criteria depend on the number of SNPs and number of DNA samples to be genotyped. If a large number of loci are to be analyzed in a restricted number of samples (e.g. for correlation studies), the procedure should be cheap (both in terms of setup and genotyping) and require no locus-specific optimization. The possibility of parallel genotyping may also be desirable. If a small or moderate number of SNPs must be genotyped on a considerable number of samples (e.g. for determining allelic frequencies in populations or for molecular evolution studies), higher setup costs may be relevant; however, the genotyping price should be kept low. In both cases, equipment available in the laboratory would influence the choice, e.g. if there is no sequencer or a mass-

spectrometer in the laboratory, then methods involving separation in a sequencing gel or MALDI-TOF analysis would be very expensive. For non-specialized laboratories, the simplicity of the procedure (minimal handwork) and resistance against contamination are important.

Methods for routine clinical trials (e.g. carrier screening and diagnostics), paternity analysis, and forensic applications should be highly accurate and reliable. Since the same SNPs must be analyzed in a large number of individuals, high setup cost is acceptable. The chosen method should be contamination-free. Usually the preference is given to closed-tube procedures (TaqMan, Invader-assay, Amplifluor).