

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Polymorphisms

Wasana Sukhumsirichart

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/intechopen.76728>

Abstract

Polymorphism or variation in DNA sequence can affect individual phenotypes such as color of skin or eyes, susceptible to diseases, and respond to drug, vaccine, chemical, and pathogen. It occurs more often than mutations (frequency $\geq 1\%$). The common polymorphism is single nucleotide polymorphism (SNP) which is a single base change in a DNA sequence that occurs most commonly in the human genome. SNPs have been used as molecular markers in a wide range of studies. Genome-wide association studies (GWAS) searches for SNPs that occur more frequently in person with a particular disease than in person without the disease and pinpoint genes or regions that may contribute to a risk of disease. This topic describes about polymorphisms, SNPs, GWAS, linkage disequilibrium (LD), minor allele frequency, haplotype, method for SNP genotyping, and application of SNPs and genome-wide association study in human diseases and drug development.

Keywords: drug development, genome-wide association studies, human diseases, polymorphisms, single nucleotide polymorphism, SNPs

1. Introduction

Phenotype of living organism is controlled by DNA. Variation in DNA sequence or polymorphism may make individual difference such as the differences in phenotype, risk of various diseases, and response to drugs, vaccine, chemical, and pathogen. Polymorphisms commonly occur in nature and related to biodiversity, genetic variation, and adaptation. It helps to maintain variety of forms in a population living in a varied environment [1]. It is preserved by frequency-dependent selection. The polymorphisms in this topic focus on human polymorphisms related to diseases and drug respond. Because the Human Genome Project had been completed, a large number of polymorphisms among the population have been found [2–4]. The most abundant type of the variations is single nucleotide polymorphisms (SNPs), with

more than 9 million reported in public databases [5–7]. In this chapter, the definition of several terms such as polymorphism, minor allele frequency (MAF), allele frequency, haplotype, and linkage disequilibrium (LD) is clarified. Moreover, SNPs, genome-wide association study (GWAS), methods to detect SNPs and application of SNPs in association with diseases and drug development are mainly discussed topics.

2. Polymorphisms

Genetic polymorphism, the definition by Cavalli-Sforza and Bodmer, is the occurrence in the same population of two or more alleles at one locus, each with appreciable frequency, where the minimum frequency is typically taken as 1% [8]. An allele is one of the variant forms of a gene at specific locus on a homologous chromosome. The different forms of the polymorphism (alleles) are observed more often in the general population than mutations. The most common polymorphism in the human genome is the single-nucleotide polymorphism (SNP) [9].

2.1. Single nucleotide polymorphism (SNP or snip)

SNPs are popular molecular genetic markers in disease genetics studies and pharmacogenomic research. It is a single base change in a DNA sequence, with a normal alternative of two possible nucleotides at a given position. This variation occurs at a specific position in the genome and has allele frequency of 1% or greater [10]. Around 325 million, SNPs have been identified in the human genome, 15 million of which are present at frequencies of 1% or higher across different populations worldwide [11]. An example for SNP is shown in **Figure 1**. It demonstrates that at a specific position of human genome when compared between two individuals and two DNA sequences. The DNA sequence of a person 1 has C nucleotide which is similar to most of the other person (majority group), whereas the DNA sequence of a person 2 has T at this position which is minority group of population. It is said that there is an SNP at this specific position between allele C or T.

The majority of SNPs have two alleles, which represent a substitution of one base for another. The SNP occurs at each allele of an individual may be different. If the SNP occurs more frequently in the general population, it is called “major” allele. In contrast, if the frequency of the SNP exist is rare in the population, it is designated the “minor” allele. Since human have two copies of chromosome or diploid, therefore, an individual can have various genotypes such as homozygous of major or minor alleles, or heterozygous of major and minor allele [9]. Many SNPs are correlated with one another, so it is difficult to distinguish the SNP that affects the phenotype from the several SNPs associated with it [12].

SNPs are identified and characterized by sequencing the same genomic region in several populations [13, 14]. The sample size of the population being resequenced is important. In general, larger sample sizes are needed to identify SNPs on the lower end of the minor allele frequency spectrum. The minor allele frequency (MAF) refers to the frequency at which the less common allele occurs in a given population. By using population genetics theory prediction for a SNP detection rate of 99%, a SNP with a minor allele frequency of 5% or greater needs 48

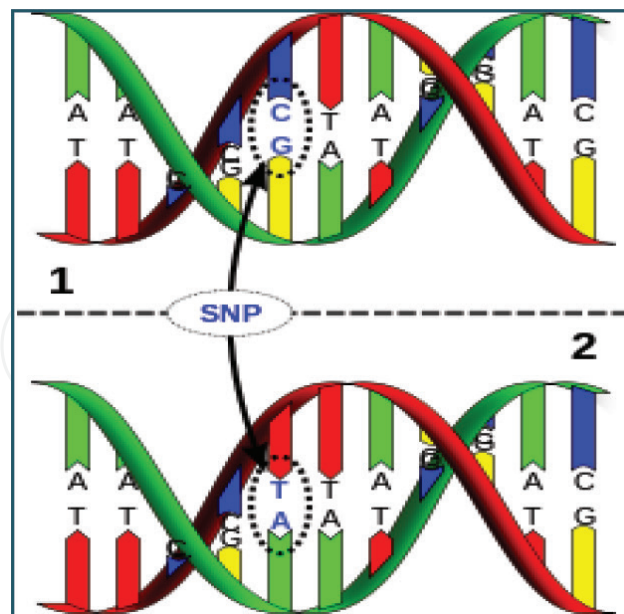


Figure 1. Single nucleotide polymorphism (SNP). (http://en.wikipedia.org/wiki/Single-nucleotide_polymorphism).

chromosomes, whereas a SNP with a minor allele frequency of 1% or greater requires 192 chromosomes for the verification of genotype of SNP [15].

Currently, the genotyping in a large scale of SNPs can be performed by automated machines, which facilitate the genetic association study using DNA-based marker. Human Genome Project rank SNP discovery and characterization as high priorities [16, 17] and encourage public and private sections [4, 18, 19] to push an effort toward these objectives.

2.1.1. Effects of SNPs location

The location of SNPs may affect gene products and others. The SNPs within a gene may alter protein structures. The SNPs in the regulatory region outside a gene may affect when and how the gene is turned on, which affects the quantity of the protein produced. They also affect gene splicing, transcription factor binding, or the sequence of non-coding RNA. The SNPs that are not within the proximity of a gene may be used as genetic markers for locating disease-causing genes (**Figure 2**).

2.1.2. Types of SNPs

As described earlier, the SNPs may fall within coding sequences of gene, or non-coding regions of gene, or in the intergenic regions (regions between genes). The SNPs in the coding region of gene are divided into two types: synonymous and nonsynonymous SNPs. The synonymous SNPs do not change the amino acid sequence of protein or not affect the protein function. The nonsynonymous SNPs are divided into two types: missense and nonsense. A missense SNP, single nucleotide change results in a codon that codes for a different amino acid, resulting in protein nonfunction. For nonsense, a point mutation in a sequence of DNA that changes to a stop codon results in a nonfunctional protein product. SNPs that are in non-coding regions

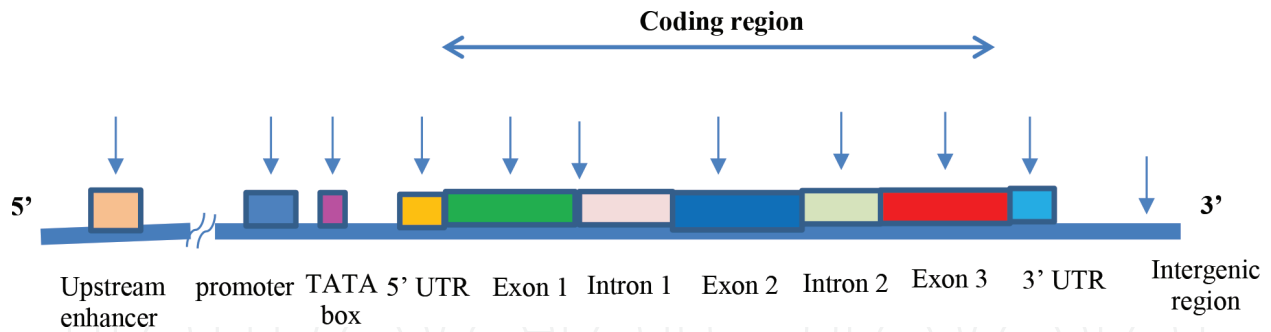


Figure 2. SNP location.

of gene, or in the intergenic regions may affect gene splicing (SNPs at intron region), transcription factor binding (SNPs at 5' untranslated region), messenger RNA degradation, or the sequence of non-coding RNA. The type of SNPs located upstream or downstream from the gene that affect gene expression is referred to an expression SNP (eSNP).

3. Association study

By comparing the patterns with patterns obtained by analyzing the DNA from a group of individuals affected and unaffected by the disease is called as an 'association study'. This study demonstrates the linking between the polymorphism and diseases or drug respond.

3.1. Minor allele frequency

Minor allele frequency (MAF) refers to the frequency at which the less common allele occurs in a given population. MAF is widely employed in GWAS for complex traits [20]. SNPs with a minor allele frequency of 5% or greater were targeted by the HapMap project.

3.2. Allele frequency and genotype frequency

Allele frequency is the relative frequency of an allele at a particular locus in a population.

Genotype frequency in a population is the sum of the individuals with the same genotype divided by the total number of individual in that population. In addition, the genotype frequency, in population genetics, means the proportion or frequency of genotypes in a population ($0 < f < 1$). The genotype frequency may also be used in the genomic profiling to predict someone is having a disease [21] or even a birth [22] defect. It can also be used to determine ethnic diversity.

3.3. Haplotypes

Haplotypes are a combination of alleles at different markers along the same chromosome that is inherited as a unit. Each haplotype is a combination of major and minor alleles along

the chromosomes, and each individual is represented twice to account for the maternal and paternal contributions [9]. The fundamental difference between haplotypes and individual genotypes at SNPs is that the alleles are assigned to a chromosome.

Haplotypes inform about the exchange of DNA during meiosis or recombination, which is useful for locating the mutation that are associate with diseases by using linkage method. It has an effect on linkage disequilibrium.

3.4. Linkage disequilibrium (LD)

In population genetics, linkage disequilibrium (LD) is the non-random association of alleles at different loci in a given population that may or may not be on the same chromosome. Loci are said to be in linkage disequilibrium when the frequency of association of their different alleles is higher or lower than what would be expected if the loci were independent and associated randomly [23]. LD can detect differences between the SNP patterns of the two groups and reveal which pattern is most likely associated with the disease-causing gene or response to certain drugs.

LD is an important concept in genetic studies that aims to identify and localize genes related to disease susceptibility. LD is commonly used to indicate that two genes are physically linked. It is defined as the difference between the observed frequency of a particular combination of alleles at two loci and the frequency expected for random formation of haplotypes from alleles. The frequency of a particular allele at a given locus will be independent of alleles at other linked loci. LD plays a crucial role in the current methods for mapping complex disease or trait-associated genes or plays a key role in health and disease. The level of linkage disequilibrium is influenced by a number of factors such as genetic linkage, selection, the rate of recombination, the rate of mutation, genetic drift, non-random mating, and population structure.

3.5. Genome-wide association studies (GWAS)

GWAS identify the common disease-causing variants by using high throughput genotyping equipment to examine hundreds of thousands of common SNPs and compare these common genetic variants in large numbers of affected cases (patient) to those in unaffected controls (non-patient) to determine whether have an association with disease (**Figure 3**) [24, 25].

In most chromosome regions, there is strong association among SNP, therefore, only a few SNPs in each region are selected to be sequenced to predict the alleles of the remaining SNPs in that region. An accurate mapping of LD pattern among SNPs which differ across ancestral groups is required for selecting the best tag SNPs. The precise LD maps are needed to help genetic association studies and stimulated the developing of human haplotype map [26, 27]. GWAS pinpoint genes that may contribute to a risk of developing disease. The data derived from GWAS inform about disease etiology, therapeutic targets, and gene function [28].

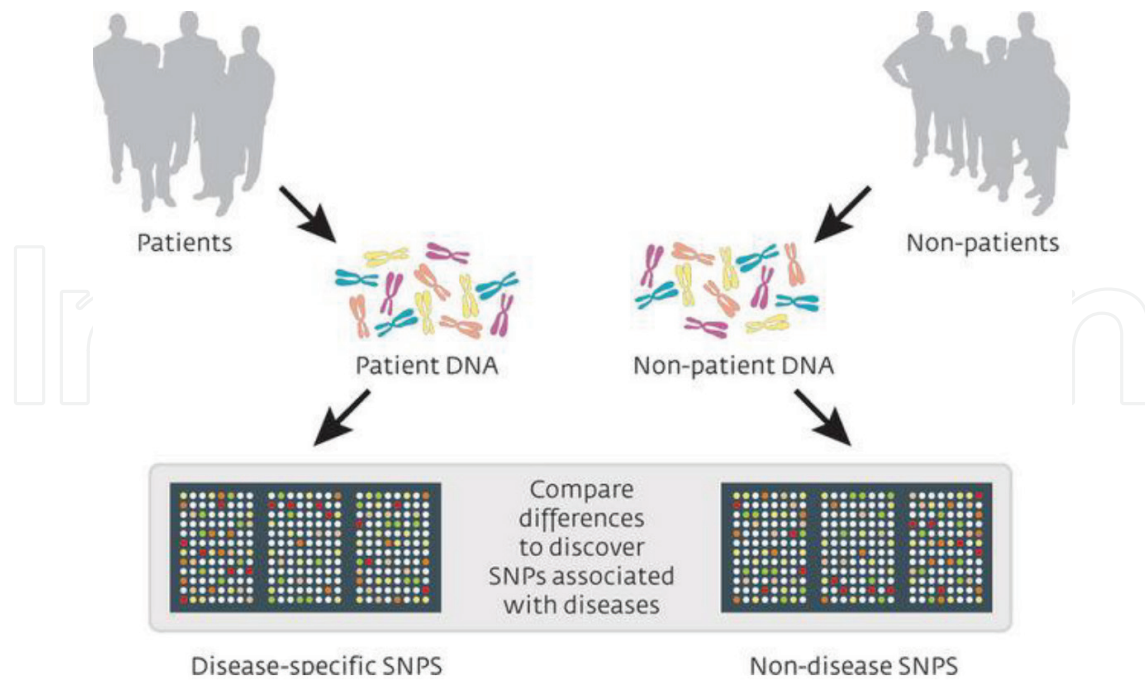


Figure 3. Genome-wide association studies (GWAS). (<http://mmg-233-2014-genetics-genomics.wikia.com/wiki/File:GWAS.jpg>).

4. Method for detection of SNPs

SNP genotyping strategies typically involve allele-discrimination and allele-detection. The other methods based on physical properties of DNA [29].

4.1. Allele-discrimination

4.1.1. Primer extension

These approaches involve allele-specific incorporation of nucleotides in primer extension reaction with a DNA template, utilizing enzyme specificity to accomplish allelic discrimination. In the reaction of a common primer extension (CPE) protocol, a designed primer is annealing with its 3' end near a SNP site and nucleotides are added by polymerase enzyme [30]. The extended nucleotide is examined by either mass or fluorescence to verify SNP genotype. Because of the simplicity in primer and assay design, it can detect multiple SNPs at the same time, therefore, several SNP genotyping use CPE [29]. CPE-based methods that use MALDI-TOF MS (matrix assisted laser desorption/ionization time-of-flight mass spectrometry) for the discrimination of alleles include PinPoint assay [31, 32], MassEXTEND™ [33, 34], SPC-single base extension (SBE) [35], and GOOD assay [36]. In the reaction of these methods, SNP-specific primers are simultaneously extended with numerous nucleotides using PCR products as a template yielding extended products of dissimilar masses. The genotypes of SNPs of the products are examined by mass analysis (**Figure 4a**). The PinPoint assay performs the simplest technique by using dideoxynucleotides (ddNTPs) for single base extension (SBE)

of primer [31, 32]. The CPE approaches that use fluorescence-based detection involve SBE of primer with fluorescently labeled ddNTPs (**Figure 4b**). Specific primer extension (SPE) methods use two primers that have the same nucleotide sequences except specific allele/base at their 3' end. The primers can extend if the nucleotide at their 3' end of the primers perfectly binds with the SNP of complementary template, then allelic discrimination can be examined by mass different using gel electrophoresis (**Figure 4c**).

MassEXTEND™ procedure uses a mixture of dNTPs and ddNTPs for primer extension and discriminate the allele by mass detection. It provides medium-to-high-throughput genotyping

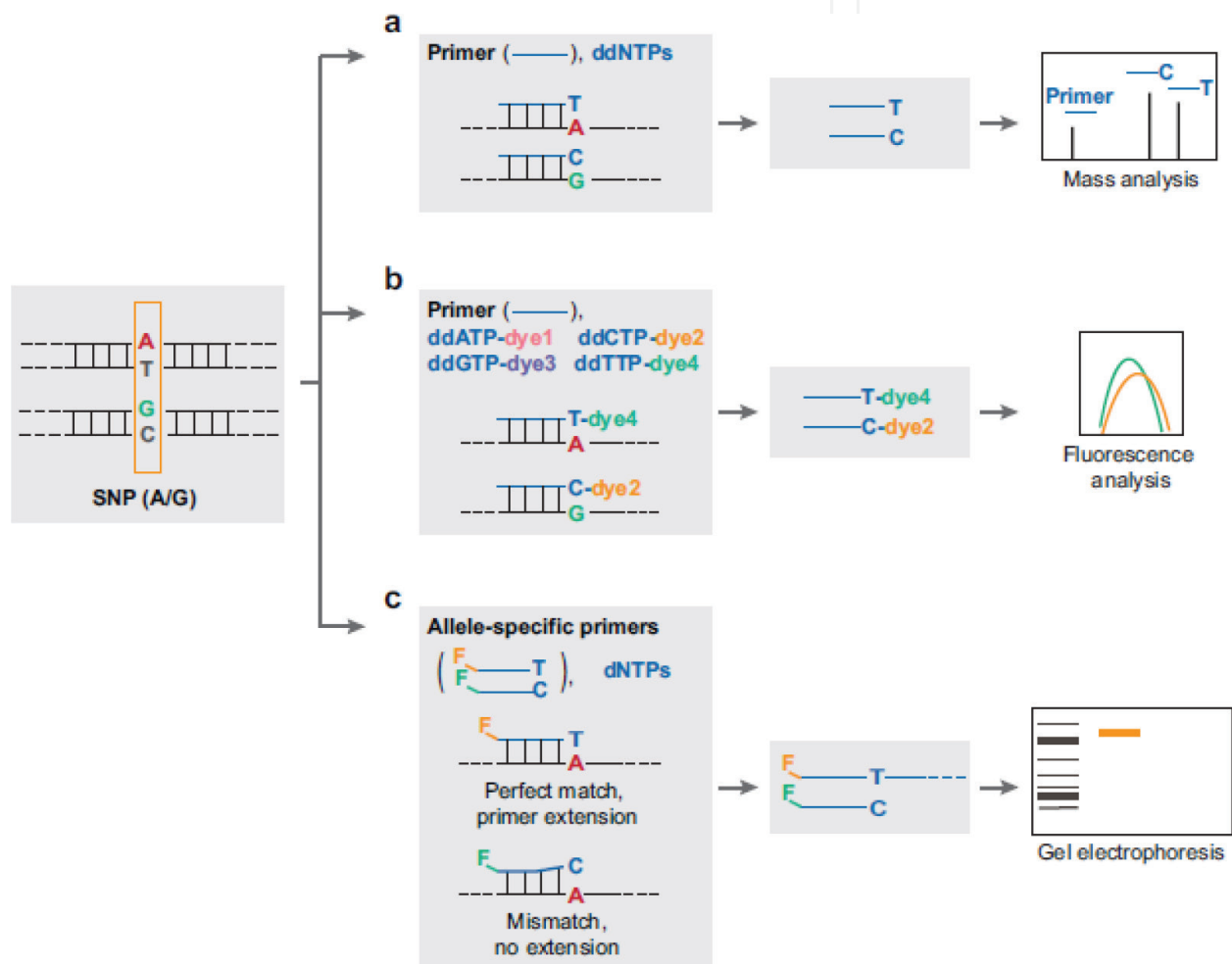


Figure 4. Primer extension approaches for SNP genotyping. (a) Mass spectrometry-based detection. This method uses a primer that anneals one base upstream of the SNP site followed by its extension with ddNTPs. Extension products are detected by mass spectrometry, and the difference between mass of extension product and primer identifies incorporated nucleotide(s) and therefore the SNP genotype. (b) Fluorescence detection using capillary electrophoresis. This technique uses a primer that anneals one base upstream of the SNP site after that its extension with ddNTPs that are labeled with different fluorescent tags. Products are detected by fluorescence after capillary electrophoresis and the color of dye indicates incorporated base(s). (c) Allele-specific primers with detection of PCR products. This method uses two allele-specific primers that anneal with their 3' ends at the SNP site and a common reverse primer (not shown) for PCR reaction. The amplification occurs when the forward primer perfectly matches to the SNP at its 3' end and the genotype can be determined by gel electrophoresis of PCR product [29].

that can be used to genotype a group of SNPs in candidate genes, SNPs for GWAS and SNPs in a large sample set [33, 34].

4.1.2. Hybridization

Hybridization methods use differences in thermal stability of double-stranded DNA to separate between perfectly matched and mismatched target-probe pairs for succeeding allelic discrimination. Hybridization methods have been applied on high-throughput platforms using microarrays (**Figure 5a**). In the GeneChip® array technology (Affymetrix, CA) protocol, probe array is synthesized in an ordered fashion on a solid surface using 25-bases oligonucleotides with specific allele by photolithography [37].

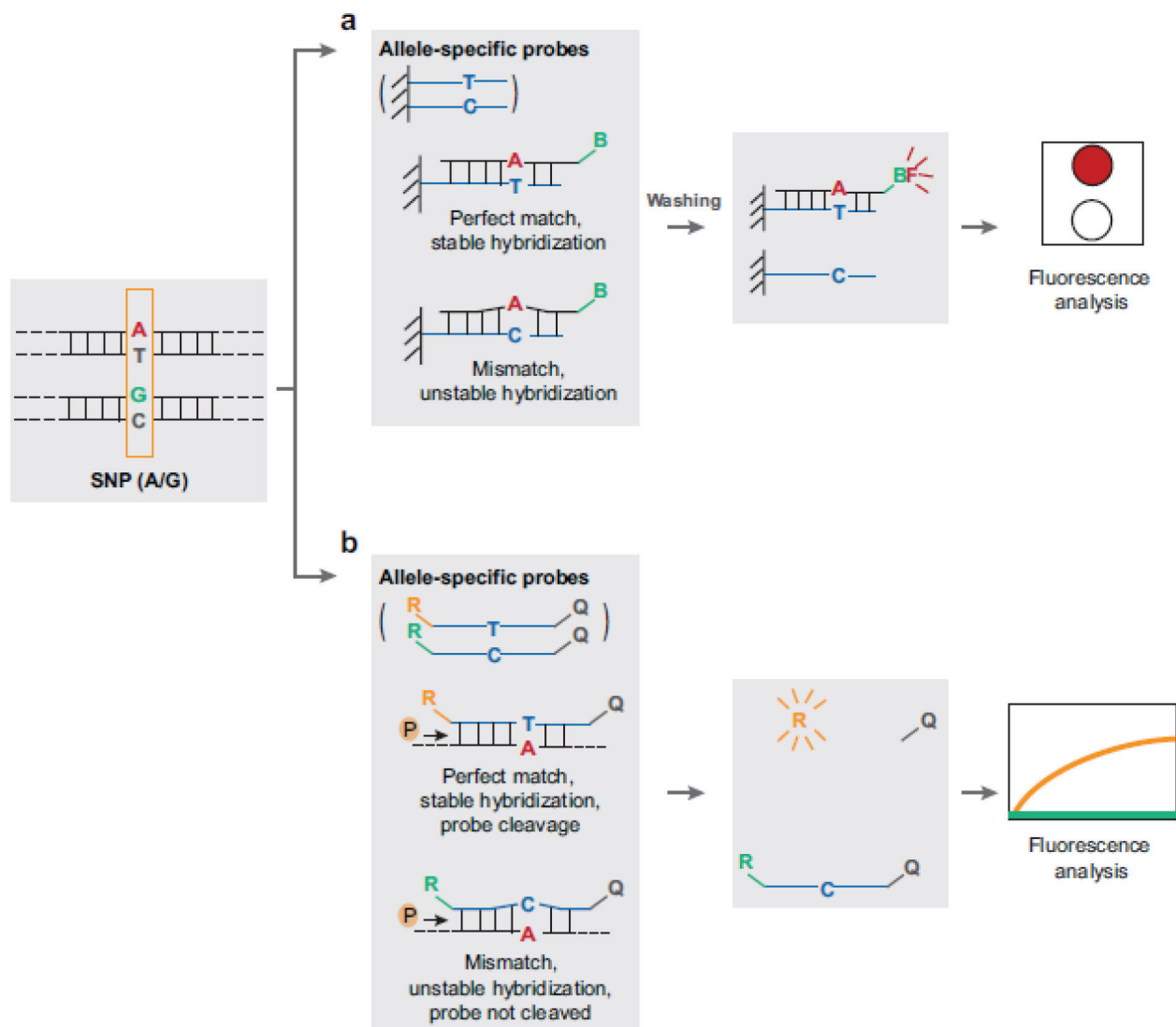


Figure 5. Hybridization-based approaches for SNP genotyping. (a) Target hybridization to a probe array. This method uses allele-specific probes attached to a solid surface for hybridization with tagged-targets containing SNPs. The surface is washed to remove mismatched targets, then the genotype of the perfectly matched target-probe pairs is detected by fluorescence. (b) TaqMan® assay. The assay uses two allele-specific probes with different dyes, reporter (R) and quencher (Q), at either end with a mismatch at the SNP site. The perfectly matched probe is cleaved during PCR amplification of the SNP-containing region, release its reporter, and the SNP genotype indicates by fluorescence analysis [29].

The DNA fragment containing SNPs are amplified from genomic DNA. The PCR products are cleaved, tagged, and hybridized to the probe array under stringent conditions, then wash, and label with fluorescent. The genotypes of SNPs are examined from the fluorescence signal based on probe-target hybridization. Numerous probes that differ at a single base are used to confirm each SNP to increase genotyping accuracy. In this method, a single array contains millions of probes and be used for parallel genotyping of 10^4 to approximately 10^5 SNPs [38]. The TaqMan® genotyping assay (Applied Biosystems, CA) combines the procedure of hybridization and 5' nuclease activity of polymerase coupled with fluorescence detection (**Figure 5b**). It employs four oligonucleotides containing two allele-specific oligonucleotide probes that have a single base mismatch and a pair of PCR primers neighboring the SNP covering region [29, 39].

4.1.3. Ligation

Ligation method discriminates the allele by using specificity of ligase enzymes. When two oligonucleotides (allele-specific probes and ligation probes) hybridize to single-stranded template DNA with perfect complementarity, they are nearby to each other, then ligase enzymes join them together to form a single oligonucleotide. Normally, three oligonucleotide probes are used in ligation assays, two probes are allele specific that bind to the template at the SNP site and ligation probe [40]. Combinatorial fluorescence energy transfer (CFET) tags have been used with ligation for SNP genotyping [41]. CFET tags are composed of fluorescent dyes that can transfer energy when they are in close proximity. For SNP genotyping, two probes of allele specific are labeled with CFET tags and common probe is labeled with biotin (**Figure 6a**). After the ligation reaction, product is separated using the biotin-streptavidin interaction. Genotyping is carried out by using capillary array gel electrophoresis based on tag fluorescence. In Padlock technology, a linear oligonucleotide probe is used with its ends designed to mimic the allele-specific probe and common probe for ligation at the SNP site (**Figure 6b**).

4.1.4. Enzymatic cleavage

The method is based on the capacity of enzymes to cleave DNA by recognition of specific sequences and structures. Difference between alleles can be discriminated when SNPs are located in an enzyme recognition site and affect the enzyme activity.

The ability of restriction enzymes has been used for detection of genetic variation by the method of restriction fragment length polymorphism (RFLP) [42]. These enzymes recognize specific sequences in double-stranded DNA and cleave both strands at a specific site in the sequence or near it to generate smaller DNA fragments (**Figure 7a**). For SNP genotyping, the PCR product containing the SNP is incubated with appropriate restriction enzyme and separated by gel electrophoresis. The SNP genotype is simply determined from sizes of the digested products. This method does not need any probes but it has limited amount and a number of SNPs.

The Invader® assay (Third Wave™ Technologies, WI) utilizes structure-specific cleavage by a flap endonuclease enzyme (**Figure 7b**). It uses three probes for genotyping a SNP, two allele-specific probes and a third common probe (invader) [43].

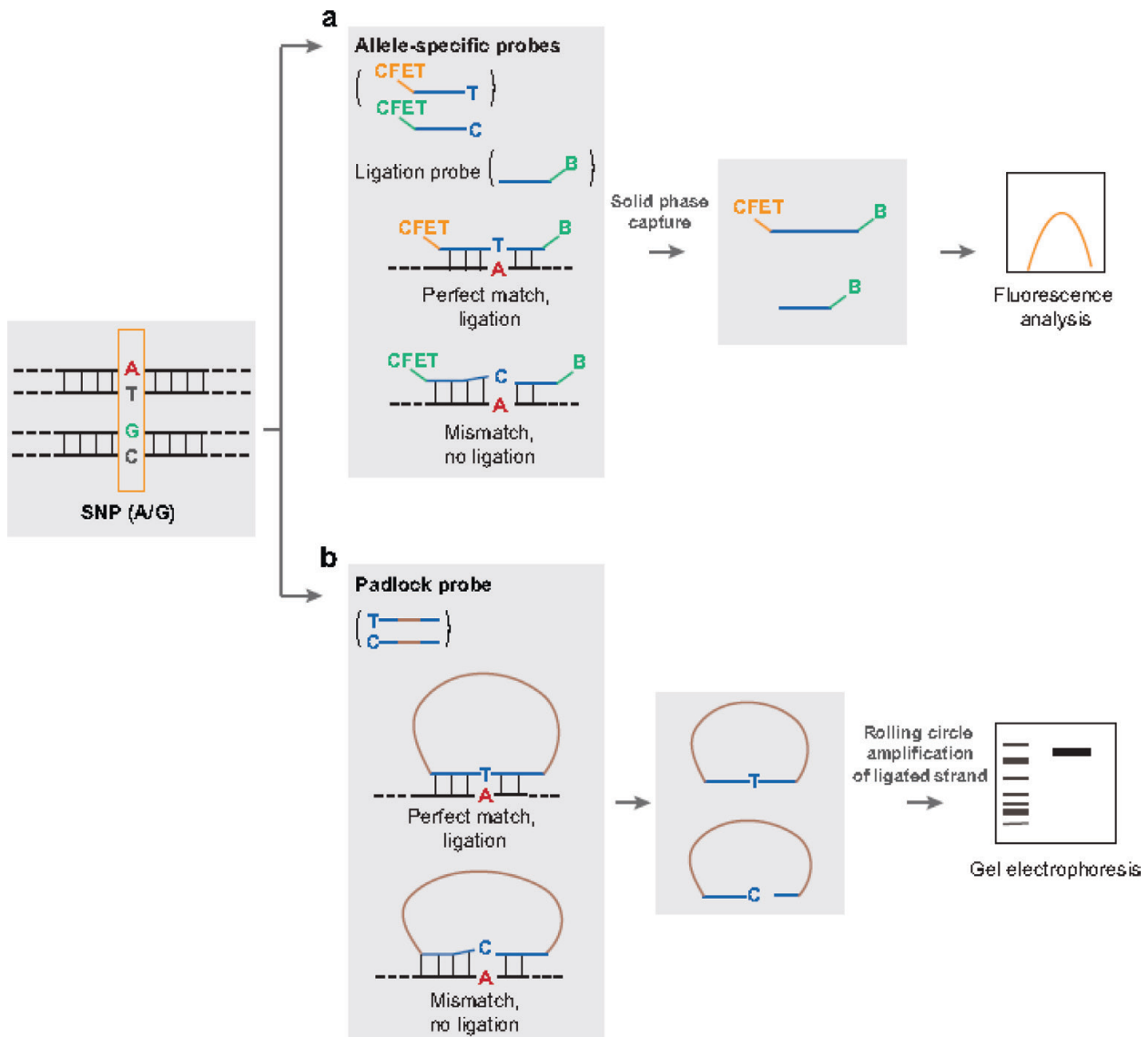


Figure 6. Ligation-based SNP genotyping approaches using ligase enzymes. (a) Detection using CFET labels. This protocol uses two allele-specific probes that have different CFET labels and a biotin-tagged common probe that hybridizes adjacent to them. Allele-specific probe and common probe bind to the template containing SNP. If the sequences are perfectly matches at SNP site, the two probes are ligated. The ligated product is proved by isolation using the biotin tag on the common probe and the fluorescence signal from the CFET label to indicate SNP genotype. (b) Ligation of padlock probe. This method uses two allele-specific probes that hybridize to target DNA with their ends aligned end-to-end at the SNP site. One end of the probe is allele-specific that is ligated to the other end only when there is a perfect match at the SNP site, leading to circularization of the probe. After rolling circle amplification using a specific prime, the product is analyzed by gel electrophoresis [29].

4.2. Allele detection methods

4.2.1. Mass-based detection

For mass analysis of oligonucleotides, MALDI-TOF MS is a widely used method. It involves the use of a small organic molecule termed matrix that absorbs energy from a laser source of certain wavelength for ionization. When analytes are mixed and cocrystallized with matrix, they are ionized in the form of intact molecules owing to transfer of energy from the matrix molecules [44, 45]. **Figure 8** demonstrates a multiplex SNP genotyping by primer extension and MALDI-TOF MS detection.

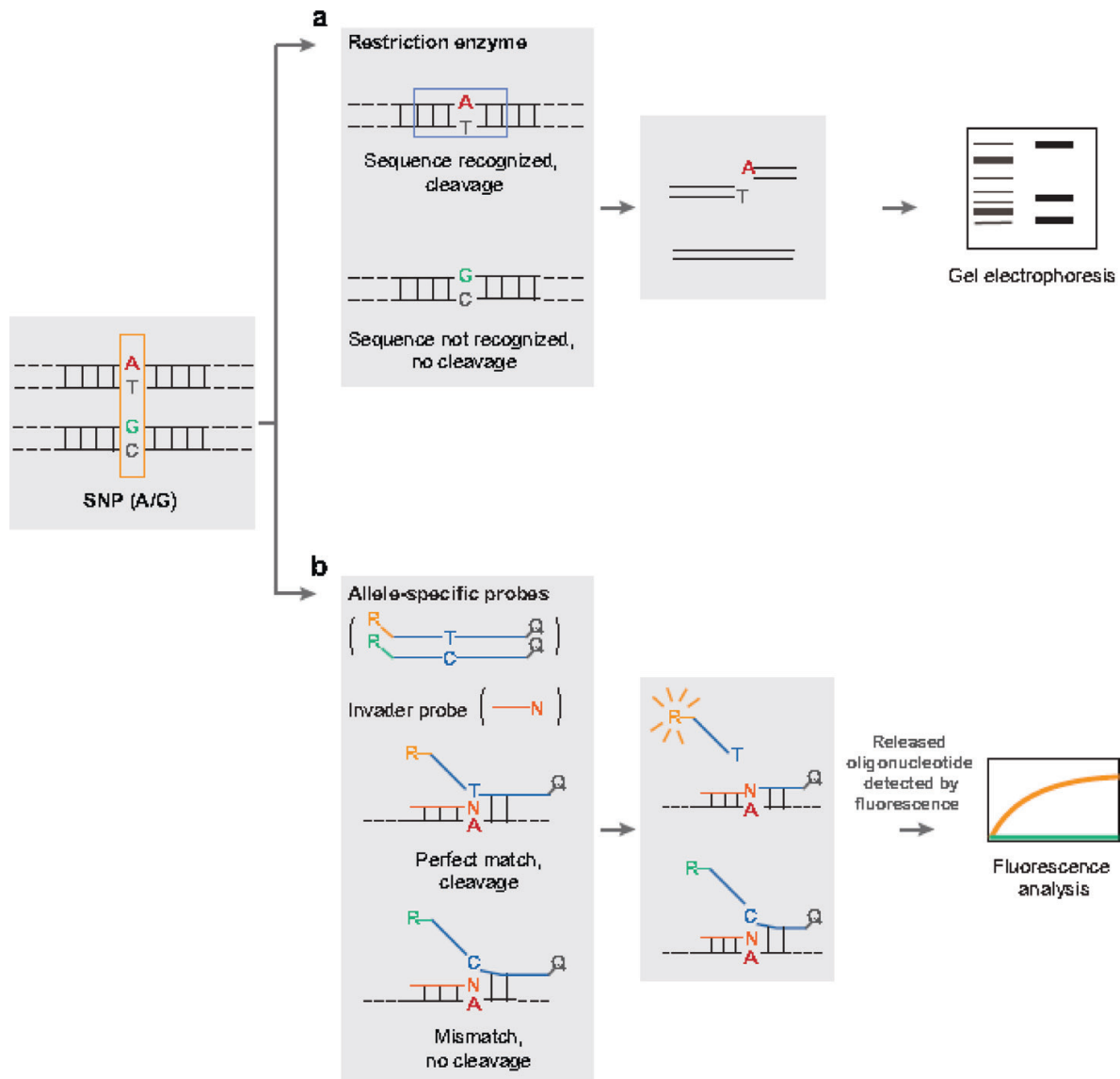


Figure 7. Enzymatic cleavage for SNP genotyping. (a) Digested with specific restriction enzyme. This method uses a restriction enzyme that cleaves only one of the alleles. The digested products are run on agarose gel electrophoresis, and the SNP genotype is determined based on the size and number of DNA fragments. (b) Invader® assay. This assay uses two allele-specific probes with different dyes, reporter (R) and quencher (Q), at either end or one common invader probe. The allele-specific probe and invader probe hybridize with target DNA to form a three-dimensional structure at the SNP site which recognized by cleavase enzyme. The allele-specific probe that complementary to the SNP is cleaved by the enzyme and releases its reporter dye in which SNP genotype can be discriminated by fluorescence analysis [29].

4.2.2. Fluorescence signal-based detection

Monitoring fluorescence signal is widely used in genotyping technologies because its operation is simple and detection is fast with high sensitivity. Fluorescence detection is used for direct sequencing (DS) using capillary array electrophoresis. Fluorescence polarization (FP)-based detection uses the change in polarization of plane polarized light by a fluorescent dye molecule owing to change in its molecular weight under conditions of constant temperature and solvent viscosity [46]. FP has been coupled with other SNP genotyping techniques including TaqMan® and Invader® [47]. TaqMan® assay is a single-step assay that uses fluorescence-based detection.

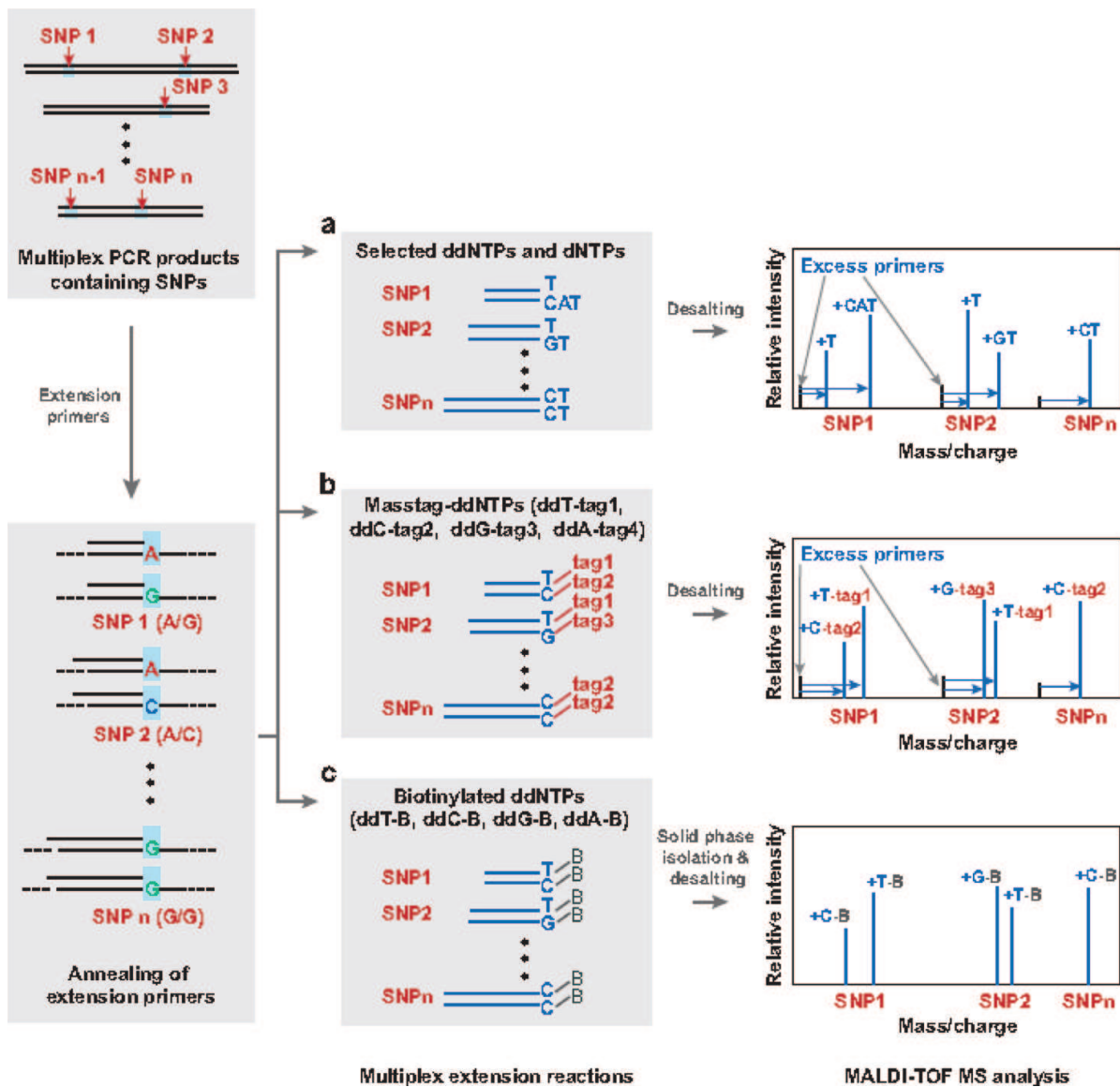


Figure 8. Multiplex SNP genotyping by primer extension and MALDI-TOF MS [29]. First, genomic DNA is amplified in multiplex PCR reaction, and the PCR products are used for a primer extension reaction with a collection of primers specific to each SNP site. The mass of the primer indicates SNP, and the mass difference between primer and its extension product(s) tells identity of incorporated nucleotide(s) and the genotype. (a) MassEXTEND[™] method in which primers are prolonged with a specific mixture of dNTPs and ddNTPs by one or more bases, specific to each allele. (b) Modified PinPoint method using mass-tagged ddNTPs for primer extension. This increases mass difference between extension products of alleles of a SNP, when compared with regular ddNTPs, increasing the accuracy of genotyping. (c) SPC-SBE method relating the production of primer extension products with biotinylated ddNTPs for isolation on a solid surface. This excludes excess primers from the mass spectrum, permitting the analysis of more products at the same time [29].

It is well suited for low-to-medium throughput genotyping applications but is currently limited to genotyping of one SNP per assay.

4.2.3. Chemiluminescence

Chemiluminescence has several advantages as a detection technique, such as high signal-to-noise ratio, rapid detection, and feasibility for automation. Pyrosequencing[™] (Biotage, Sweden) employs chemiluminescence-based detection for SNP genotyping using a cascade

of enzyme reactions [48]. Pyrosequencing™ is an approach that combines sequencing-by-synthesis with chemiluminescence detection. In SNP genotype, it provides sequence information on the region surrounding the SNP site.

4.3. The other methods

4.3.1. Single-strand conformation polymorphism (SSCP)

SSCP discriminates the allele by using secondary structure of single-stranded DNA. The single-stranded DNA molecules which differ at a single base run on a non-denaturing gel electrophoresis display different mobility based on their native conformations (**Figure 9**) [49].

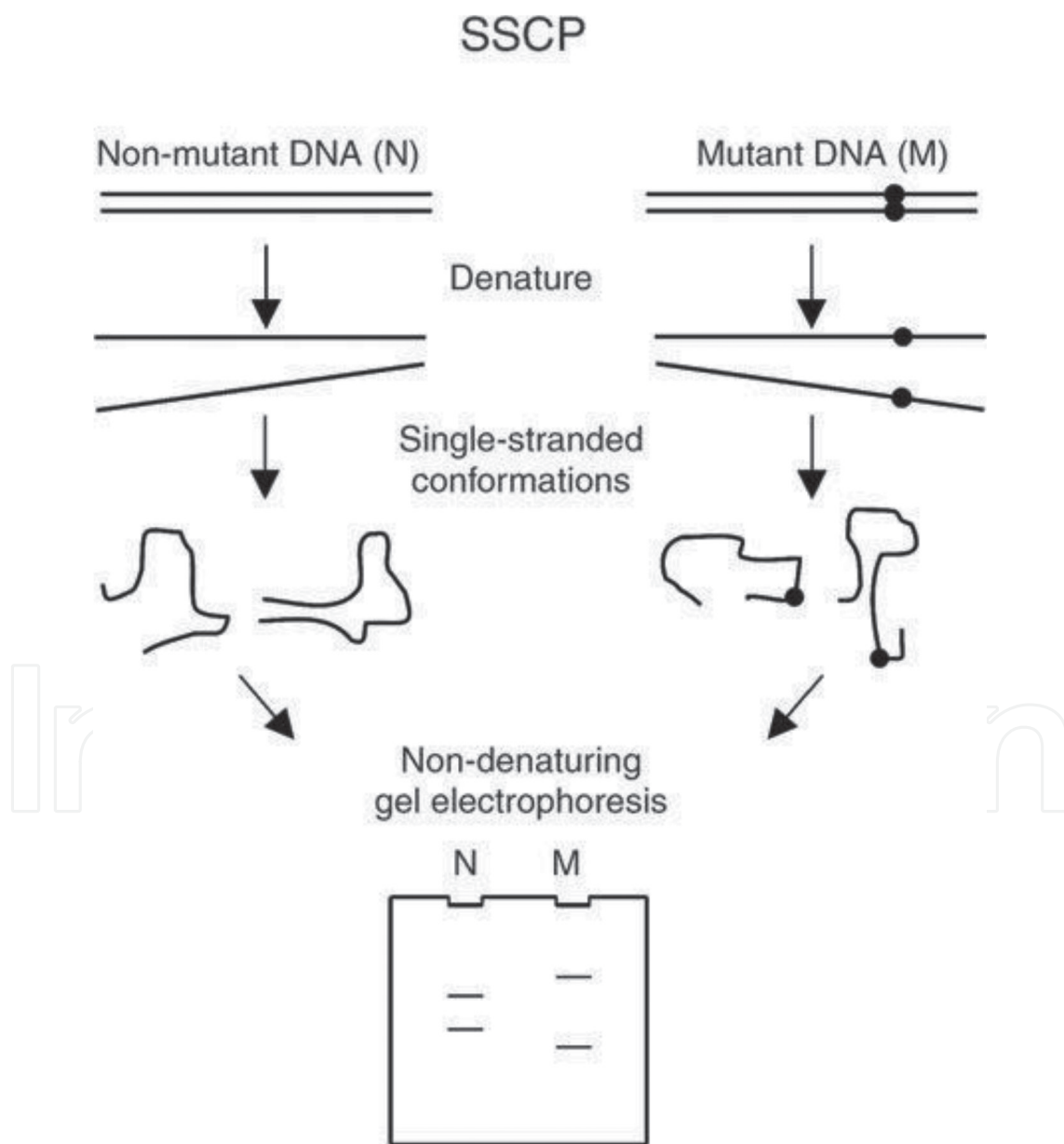


Figure 9. Single-strand conformation polymorphism (SSCP) (<https://media.nature.com/m685/nature-assets/nprot/journal/v1/n6/images/nprot.2006.485-F1.jpg>).

4.3.2. High resolution melting analysis (HRM)

This method, the fragment cover SNP is amplified by real-time PCR and followed by HRM. The HRM is a technique for the detection mutations and SNPs, which based on analysis of melting curve when double-strand DNA (dsDNA) separate into single-strand DNA (ssDNA) during increased temperature from around 50°C up to around 95°C. At some point during this process, the melting temperature of the amplicon is reached and the two strands of DNA separate. This can visualize the melting behavior of the product through a fluorescent dye [50]. The fluorescent dye binds to double-strand DNA during the amplification resulting in an increase of fluorescence (**Figure 10**). This method is simple, cost-effective, fast and able to accurately genotype many samples rapidly. It also reduces the need to design multiple pairs of primers or purchase expensive probes.

4.3.3. Denaturing high performance liquid chromatography (DHPLC)

It is a method for screening DNA samples for SNPs and inherited mutations. The analysis begins with a PCR amplification, followed by a step of denaturation-renaturation to create hetero- and homoduplexes from the two populations in the PCR. The heteroduplexes with mismatch pairing and homoduplex can be detected on reversed-phase chromatography of denaturing high performance liquid chromatography (DHPLC). The heteroduplexes thermally that less stable than their corresponding homoduplexes will be resolved by chromatography when subjected to a sufficiently high temperature. This mismatch will decrease the interaction with the column and a reduced retention time compared to the homoduplexes (**Figure 11**).

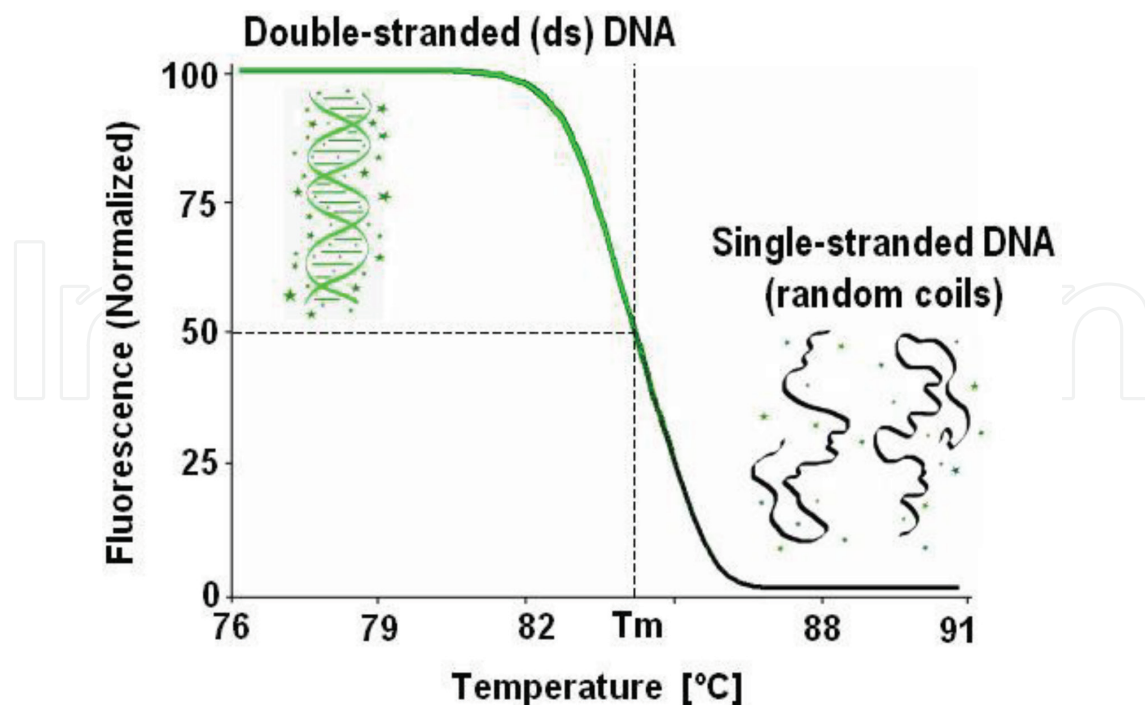


Figure 10. High resolution melting (HRM) analysis (https://www.dna.utah.edu/Image/Hi_Res%20Melting_Normalized.JPG).

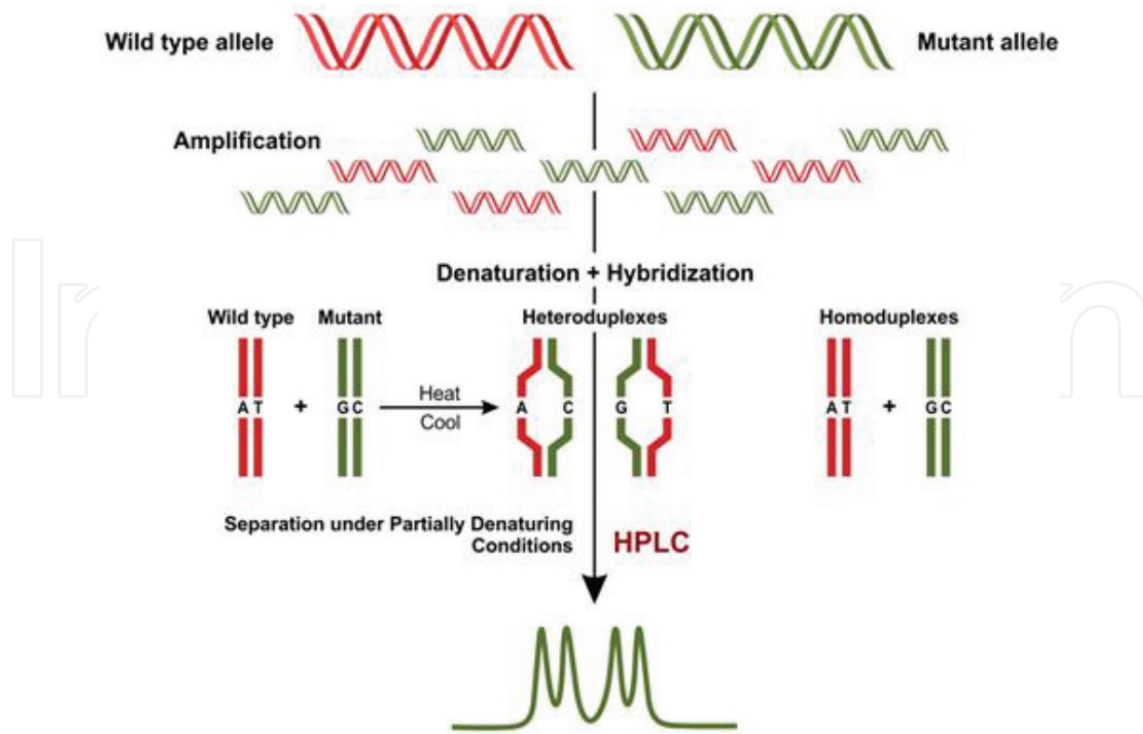


Figure 11. Denaturing high performance liquid chromatography (DHPLC) (http://www.tankonyvtar.hu/en/tartalom/tamop425/0011_1A_Molekularis_diagnosztika_en_book/images/image047.png).

5. Application of SNPs

5.1. SNPs as biological markers of human diseases

Most SNPs are not responsible for a disease state. They serve as biological markers for pinpointing a disease on the human genome map. SNPs occur on average once every 200 base pairs [18, 51–53] in the human genome. Common SNPs (a minor allele frequency range from 5% to >20%) can explain a proportion of common human disease. Most SNPs do not occur in the coding region of genes or even in genes [4]. Nonsynonymous SNPs that suspect in causing a human disease do not account for all SNPs that can cause disease or susceptibility to disease. Other functional SNPs associated with human disease or susceptibility to disease includes SNPs located in promoters [54, 55], introns [56], splice sites [57], and intragenic region [56, 58]. Even synonymous SNPs have been involved in functional consequences via unknown mechanism [59].

5.2. SNPs and drug development

Variants of genes encoding drug metabolizing enzymes or drug targets have been studied in association with personal drug responses. SNPs are popular molecular markers in such pharmacogenomics studies. Using SNPs to study the genetics of drug response will help in the creation of personalized medicine or the most appropriate drug for an individual and could be determined in advance of treatment by analyzing a patient's SNP profile. SNPs may

be associated with the absorbance and clearance of therapeutic agents. The association of different SNPs with a wide range of human diseases such as cancer, infectious diseases autoimmune, neuropsychiatric and others can be used as targets for drug therapy [60].

6. Studies of SNPs in association with human diseases

6.1. Cancer

Cancer is a disease that involves abnormal cell growth. There are several kinds of cancers. Gemignani et al. studied polymorphisms in dopamine receptor gene, *DRD2*, in association with colorectal cancer risk utilizing the APEX system which is primer extension assays that use fluorescence and mass detection [61]. They genotyped seven SNPs of *DRD2* in 370 case and 327 control samples and found three of the seven SNPs to be highly associated with colorectal cancer, related to reduced levels of D2 dopamine receptor. Hartikainen et al. investigated genes associated with breast cancer in an eastern Finnish population [62]. Ten SNPs were genotyped in the 22q12-q13 region from DNA of 497 patients and 458 control subjects using the TaqMan® assay. The SNP, rs733655, in matriptase-2 gene (*TMPRSS6*) was found to have a strong association with risk of breast cancer. Ribas et al. genotyped 899 SNPs in 175 candidate cancer genes from Spanish population using BeadARRAY™ and SNPlex™ methods to evaluate applicability of HapMap data (for subjects of European ancestry) to cancer research. They found that allele frequencies and haplotype distributions obtained in the study were consistent with the HapMap data [63]. Shatalova et al. examined SNPs in sulfotransferase 1A1 (*SULT1A1*) and UDP-glucuronosyltransferase 1A1 (*UGT1A1*) genes in 119 patients with breast cancer and 121 controls using Pyrosequencing™ [64]. They found only *UGT1A1* gene to be associated with risk of breast cancer. Copson et al. genotyped a T/G SNP (SNP309) in the *MDM2* gene in 116 patients with pathogenic mutations in *BRCA1* gene and 102 healthy controls [65]. Results suggested that the *MDM2* SNP309 locus does not have a significant association with accelerated cancer development in carriers of known pathogenic mutations of *BRCA1*.

6.2. Schizophrenia

Schizophrenia is a severe psychiatric disorder characterized by hallucinations, delusions, cognitive deficits, and apathy, with a lifetime prevalence of ~1%. Epidemiologic studies on twins indicate that schizophrenia has a complex genetic background with heritability estimated at 73–90%.

Arinami et al. genotyped 5861 SNPs of 602 individuals from 236 Japanese families using the BeadARRAY™ Linkage Panel (IV) for the genome-wide linkage analysis [66]. They found a strong association of schizophrenia to the region 1p21-p13 and implied that schizophrenia might have common susceptibility loci across populations with different ethnicity-specific effects.

Panichareon B et al. used GWAS-discovered SNPs of Europeans ancestry in *OPCML* gene and investigated SNPs in Thai schizophrenia patient by using polymerase chain reaction (PCR) and high-resolution melting (HRM) analysis. The results of this study found a strong

association between an intronic SNP (rs1784519) and the risk of schizophrenia in a Thai population [$p = 0.00036$, odds ratio for the minor A allele: 2.11(1.57–2.84)] [67].

6.3. Dyslipidemia

Dyslipidemia is an abnormal of lipid and/or lipoproteins in the blood. It is a major risk factor of coronary heart disease and atherosclerosis. A genome-wide association study (GWAS) examined the concentrations of HDL-C and triglyceride in European ethnic and identified the SNP at 15 loci which associated with HDL-C levels (such as, APOA1/C3/A4/A5 gene cluster) and SNPs at 12 loci associated with triglycerides (such as APOB, APOE gene) [68]. Thongket et al. examined SNP in apolipoprotein E receptor 2 gene using real-time PCR and HRM analysis and found that the rs2297660 showed strong association with risk of dyslipidemia in Thai population [69].

6.4. Diabetes mellitus

Diabetes mellitus (DM) is a group of metabolic disorders. Untreated diabetes patient can cause many complications. Acute complications can include diabetic ketoacidosis, hyperosmolar hyperglycemic state, or death [70]. SNPs in the gene encoding aldose reductase (*AKR1B1*) were studied for their association with diabetic nephropathy by Wolford et al. They genotyped eight SNPs of *AKR1B1* gene in two different case control sets by Pyrosequencing™ method and found that common *AKR1B1* SNPs were unlikely to be major determinants of diabetic nephropathy [71]. Altshuler et al. genotyped SNPs from 3000 individuals to investigated the association of two polymorphisms, a missense variant in *PPARG* (Pro12Ala) and a silent C/T polymorphism in exon 22 of *ABCC8*, with type 2 diabetes using FRET and FP methods. It was found that only Pro12Ala showed a significant association with decreased risk of type 2 diabetes [72].

7. Conclusion

Polymorphism is a variation in DNA sequence that may affect individual phenotypes. It occurs more often in the general population than mutations (frequency $\geq 1\%$). The majority of variation is single nucleotide polymorphism (SNP) which is a single base change in a DNA sequence that occurs at a specific position in the genome. SNP may locate within coding, or non-coding, or intergenic regions of genes. Most of SNPs have two alleles, for an individual SNP, one is major allele and the other is the minor allele based on their observed frequency in the general population. Genome-wide association studies (GWAS) search for SNPs that occur more frequently in person with a particular disease than in person without the disease and pinpoint genes that may contribute to risk of disease. Linkage disequilibrium (LD) is commonly used to indicate that two or more genes are physically linked. It plays an important role in health and disease. Most SNPs are not responsible for a disease state but serve as biological markers for various complex diseases such as cancer, diabetes, dyslipidemia, schizophrenia, and so on. There are several methods for analyzing SNPs such as MALDI-TOF MS, GeneChip® array, pyrosequencing, DHPLC, HRM, RFLP, and so on.

Author details

Wasana Sukhumsirichart

Address all correspondence to: wasanas@g.swu.ac.th

Department of Biochemistry, Faculty of Medicine, Srinakharinwirot University, Bangkok, Thailand

References

- [1] Dobzhansky T. *Genetics of the Evolutionary Process*. New York: Columbia University Press; 1970
- [2] Collins FS, Brooks LD, Chakravarti A. A DNA polymorphism discovery resource for research on human genetic variation. *Genome Research*. 1998;**8**:1229-1231
- [3] Collins FS, Morgan M, Patrinos A. The human genome project: Lessons from large-scale biology. *Science*. 2003;**300**:286-290
- [4] Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*. 2001;**409**:928-933
- [5] Kwok PY. Methods for genotyping single nucleotide polymorphisms. *Annual Review of Genomics and Human Genetics*. 2001;**2**:235-258
- [6] Altshuler D, Brooks L, Chakravarti A, Collins F, Daly M, Donnelly P. A haplotype map of the human genome. *Nature*. 2005;**437**:1299-1320
- [7] Rocha D, Gut I, Jeffreys AJ, Kwok PY, Brookes AJ, Chanock SJ. Seventh international meeting on single nucleotide polymorphism and complex genome analysis: "Ever bigger scans and an increasingly variable genome". *Human Genetics*. 2006;**119**:451-456
- [8] Philip H. *Genetics of Populations*. Jones & Bartlett Learning. p. 104. ISBN 978-0-7637-5737-3. Retrieved 8 July 2013
- [9] Crawford DC, Nickerson DA. Definition and clinical importance of haplotypes. *Annual Review of Medicine*. 2005;**56**:303-320
- [10] Rieder MJ, Taylor SL, Clark AG, Nickerson DA. Sequence variation in the human angiotensin converting enzyme. *Nature Genetics*. 1999;**22**:59-62
- [11] Durbin RM, Altshuler DL, Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, et al. A map of human genome variation from population-scale sequencing. *Nature*. 2010;**467**(7319):1061-1073
- [12] Zhu X, McKenzie CA, Forrester T, Nickerson DA, Broeckel U, Schunkert H, et al. Localization of a small genomic region associated with elevated ACE. *American Journal of Human Genetics*. 2000;**67**:1144-1153

- [13] Schlotterer C. The evolution of molecular markers—Just a matter of fashion? *Nature Reviews Genetics*. 2004;**5**:63-69
- [14] Carlson CS, Newman TL, Nickerson DA. SNPing in the human genome. *Current Opinion in Chemical Biology*. 2001;**5**:78-85
- [15] Kruglyak L, Nickerson DA. Variation is the spice of life. *Nature Genetics*. 2001;**27**:234-236
- [16] Collins FS, Patrinos A, Jordan E, Chakravarti A, Gesteland R, Walters L. New goals for the U.S. human genome project: 1998-2003. *Science*. 1998;**282**:682-689
- [17] Collins FS, Green ED, Guttmacher AE, et al. A vision for the future of genomics research. *Nature*. 2003;**422**:835-847
- [18] Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. *Science*. 2001;**291**:1304-1351
- [19] Stephens JC, Schneider JA, Tanguay DA, Choi J, Acharya T, Stanley SE, et al. Haplotype variation and linkage disequilibrium in 313 human genes. *Science*. 2001;**293**:489-493
- [20] The International HapMap Consortium. A haplotype map of the human genome. *Nature*. 2005;**437**(7063):1299-1320
- [21] Janssens AC, Moonesinghe R, Yang Q, Steyerberg EW, van Duijn CM, Khoury MJ. The impact of genotype frequencies on the clinical validity of genomic profiling for predicting common chronic diseases. *Genetics in Medicine*. 2007;**9**(8):528-535
- [22] Shields DC, Kirke PN, Mills JL, Ramsbottom D, Molloy AM, Burke H, et al. The “Thermolabile” variant of methylenetetrahydrofolate reductase and neural tube defects: An evaluation of genetic risk and the relative importance of the genotypes of the embryo and the mother. *American Journal of Human Genetics*. 1999;**64**:1045-1055
- [23] Slatkin M. Linkage disequilibrium — Understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics*. 2008;**9**(6):477-485
- [24] Lander ES. The new genomics: Global views of biology. *Science*. 1996;**274**:536-539
- [25] Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science*. 1996;**273**:1516-1517
- [26] International HapMap Consortium. The international HapMap project. *Nature*. 2003;**426**:789-794
- [27] Eberle MA, Ng PC, Kuhn K, et al. Power to detect risk alleles using genome-wide tag SNP panels. *PLoS Genetics*. 2007;**3**:1827-1837
- [28] Manolio TA, Collins FS. The HapMap and genome-wide association studies in diagnosis and therapy. *Annual Review of Medicine*. 2009;**60**:443-456
- [29] Kim S, Misra A. SNP genotyping: Technologies and biomedical applications. *Annual Review of Biomedical Engineering*. 2007;**9**:289-320
- [30] Sokolov BP. Primer extension technique for the detection of single nucleotide in genomic DNA. *Nucleic Acids Research*. 1990;**18**:3671

- [31] Ross P, Hall L, Smirnov I, Haff L. High level multiplex genotyping by MALDI-TOF mass spectrometry. *Nature Biotechnology*. 1998;**16**:1347-1351
- [32] Haff LA, Smirnov IP. Single-nucleotide polymorphism identification assays using a thermostable DNA polymerase and delayed extraction MALDI-TOF mass spectrometry. *Genome Research*. 1997;**7**:378-388
- [33] Braun A, Little DP, Koster H. Detecting CFTR gene mutations by using primer oligo base extension and mass spectrometry. *Clinical Chemistry*. 1997;**43**:1151-1158
- [34] Cashman J, Zhang J, Leushner J, Braun A. Population distribution of human flavin-containing monooxygenase form 3: Gene polymorphisms. *Drug Metabolism and Disposition*. 2001;**29**:1629-1637
- [35] Kim S, Edwards JR, Deng L, Chung W, Ju J. Solid phase capturable dideoxynucleotides for multiplex genotyping using mass spectrometry. *Nucleic Acids Research*. 2002;**30**:e85
- [36] Sauer S, Gelfand DH, Boussicault F, Bauer K, Reichert F, Gut IG. Facile method for automated genotyping of single nucleotide polymorphisms by mass spectrometry. *Nucleic Acids Research*. 2002;**30**:e22
- [37] Kennedy GC, Matsuzaki H, Dong S, Liu WM, Huang J, et al. Large-scale genotyping of complex DNA. *Nature Biotechnology*. 2003;**21**:1233-1237
- [38] Matsuzaki H, Loi H, Dong S, Tsai YY, Fang J, et al. Parallel genotyping of over 10,000 SNPs using a one-primer assay on a high-density oligonucleotide array. *Genome Research*. 2004;**14**:414-425
- [39] Holland PM, Abramson RD, Watson R, Gelfand DH. Detection of specific polymerase chain reaction product by utilizing the 5'-3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proceedings of the National Academy of Sciences*. 1991;**88**:7276-7280
- [40] Landegren U, Kaiser R, Sanders J, Hood L. A ligase-mediated gene detection technique. *Science*. 1988;**241**:1077-1080
- [41] Tong AK, Li Z, Jones GS, Russo JJ, Ju J. Combinatorial fluorescence energy transfer tags for multiplex biological assays. *Nature Biotechnology*. 2001;**19**:756-759
- [42] Botstein D, White RL, Skolnick M, Davis RW. Construction of a genetic linkage map in man using restriction length polymorphisms. *American Journal of Human Genetics*. 1980;**32**:314-331
- [43] Lyamichev V, Mast AL, Hall JG, Prudent JR, Kaiser MW, Takova T, et al. Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes. *Nature Biotechnology*. 1999;**17**:292-296
- [44] Karas M, Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Analytical Chemistry*. 1988;**60**:2299-2301
- [45] Tanaka K, Waki H, Ido Y, Akita S, Yoshida Y, Yoshida T. Protein and polymer analyses up to m/z 100,000 by laser ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry*. 1988;**2**:151-153

- [46] Chen X, Levine L, Kwok PY. Fluorescence polarization in homogeneous nucleic acid analysis. *Genome Research*. 1999;**9**:492-498
- [47] Kwok PY. SNP genotyping with fluorescence polarization detection. *Human Mutation*. 2002;**19**:315-323
- [48] Ronaghi M, Karamohamed S, Pettersson B, Uhlen M, Nyren P. Realtime DNA sequencing using detection of pyrophosphate release. *Analytical Biochemistry*. 1996;**242**:84-89
- [49] Orita M, Suzuki Y, Sekiya T, Hayashi K. Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics*. 1989;**5**:874-879
- [50] Pasay C, Arlian L, Morgan M, Vyszenski-Moher D, Rose A, Holt D, et al. High-resolution melt analysis for the detection of a mutation associated with permethrin resistance in a population of scabies mites. *Medical and Veterinary Entomology*. 2008;**22**(1):82-88
- [51] Halushka MK, Fan JB, Bentley K, Hsie L, Shen N, Weder A, et al. Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. *Nature Genetics*. 1999;**22**:239-247
- [52] Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *American Journal of Human Genetics*. 2004;**74**:106-120
- [53] Livingston RJ, von Niederhausern A, Jegga AG, Crawford DC, Carlson CS, Rieder MJ, et al. Pattern of sequence variation across 213 environmental response genes. *Genome Research*. 2004;**14**:1821-1831
- [54] Lin MT, Storer B, Martin PJ, Tseng LH, Gooley T, Chen PJ, et al. Relation of an interleukin-10 promoter polymorphism to graft-versus-host disease and survival after hematopoietic-cell transplantation. *The New England Journal of Medicine*. 2003;**349**:2201-2210
- [55] Klerkx AH, Tanck MW, Kastelein JJ, Molhuizen HO, Jukema JW, Zwinderman AH, et al. Haplotype analysis of the CETP gene: Not TaqIB, but the closely linked-629C→A polymorphism and a novel promoter variant are independently associated with CETP concentration. *Human Molecular Genetics*. 2003;**12**:111-123
- [56] Tokuhira S, Yamada R, Chang X, Suzuki A, Kochi Y, Sawada T, et al. An intronic SNP in a RUNX1 binding site of SLC22A4, encoding an organic cation transporter, is associated with rheumatoid arthritis. *Nature Genetics*. 2003;**35**:341-348
- [57] Betticher DC, Thatcher N, Altermatt HJ, Hoban P, Ryder WD, Heighway J. Alternate splicing produces a novel cyclin D1 transcript. *Oncogene*. 1995;**11**:1005-1011
- [58] Helms C, Cao L, Krueger JG, Wijsman EM, Chamian F, Gordon D, et al. A putative RUNX1 binding site variant between SLC9A3R1 and NAT9 is associated with susceptibility to psoriasis. *Nature Genetics*. 2003;**35**:349-356
- [59] Duan J, Wainwright MS, Comeron JM, Saitou N, Sanders AR, Gelernter J, et al. Synonymous mutations in the human dopamine receptor D2 (DRD2) affect mRNA stability and synthesis of the receptor. *Human Molecular Genetics*. 2003;**12**:205-216

- [60] Fareed M, Afzal M. Single-nucleotide polymorphism in genome-wide association of human population: A tool for broad spectrum service. *Egyptian Journal of Medical Human Genetics*. 2013;**14**(2):123-134
- [61] Gemignani F, Landi S, Moreno V, Gioia-Patricola L, Chabrier A, et al. Polymorphisms of the dopamine receptor gene DRD2 and colorectal cancer risk. *Cancer Epidemiology, Biomarkers & Prevention*. 2005;**14**:1633-1638
- [62] Hartikainen JM, Tuhkanen H, Kataja V, Eskelinen M, Uusitupa M, et al. Refinement of the 22q12-q13 breast cancer-associated region: Evidence of TMPRSS6 as a candidate gene in an eastern Finnish population. *Clinical Cancer Research*. 2006;**12**:1454-1462
- [63] Ribas G, Gonzalez-Neira A, Salas A, Milne RL, Vega A, et al. Evaluating HapMap SNP data transferability in a large-scale genotyping project involving 175 cancer-associated genes. *Human Genetics*. 2006;**118**:669-679
- [64] Shatalova E, Loginov V, Braga E, Kazubskaja T, Sudomoina M, et al. Association of SULT1A1 and UGT1A1 polymorphisms with breast cancer risk and phenotypes in Russian women. *Molecular Biology*. 2006;**40**:228-234
- [65] Copson ER, White HE, Blaydes JP, Robinson DO, Johnson PW, Eccles DM. Influence of the MDM2 single nucleotide polymorphism SNP309 on tumor development in BRCA1 mutation carriers. *BMC Cancer*. 2006;**6**:80
- [66] Arinami T, Ohtsuki T, Ishiguro H, Ujike H, Tanaka Y, Morita Y, et al. Genomewide high-density SNP linkage analysis of 236 Japanese families supports the existence of schizophrenia susceptibility loci on chromosomes 1p, 14q, and 20p. *American Journal of Human Genetics*. 2005;**77**:937-944
- [67] Panichareon B, Nakayama K, Thurakitwannakarn W, Iwamoto S, Sukhumsirichart W. *OPCML* gene as a schizophrenia susceptibility in Thai population. *Journal of Molecular Neuroscience*. 2012;**46**(2):373-377
- [68] Kathiresan S, Willer CJ, Peloso GM, Demissie S, Musunuru K, Schadt EE, et al. Common variants at 30 loci contribute to polygenic dyslipidemia. *Nature Genetics*. 2009;**41**:56-65
- [69] Thongket P, Rattanathanawan K, Seesom W, Sukhumsirichart W. Apolipoprotein E receptor 2 gene polymorphism associated with dyslipidemia among Thai population. *Journal of the Medical Association of Thailand*. 2015;**98**:85-89
- [70] Kitabchi AE, Umpierrez GE, Miles JM, Fisher JN. Hyperglycemic crises in adult patients with diabetes. *Diabetes Care*. 2009;**32**(7):1335-1343
- [71] Wolford JK, Yeatts KA, Red Eagle AR, Nelson RG, Knowler WC, Hanson RL. Variants in the gene encoding aldose reductase (AKR1B1) and diabetic nephropathy in American Indians. *Diabetic Medicine*. 2006;**23**:367-376
- [72] Altshuler D, Hirschhorn JN, Klannemark M, Lindgren CM, Vohl MC, et al. The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nature Genetics*. 2000;**26**:76-80