6 Detection Techniques for Single Nucleotide Polymorphisms

W. Mathias Howell, Johan Stenberg, Chatarina Larsson, Mats Nilsson, and Ulf Landegren

CONTENTS

Introduction 140
Single Nucleotide Polymorphism 140
- Discovery, Abundance, and Distribution 142
- Linkage Disequilibrium, Haplotypes, and HapMap Project 142
- SNPs in Inherited Disease 144

SNP Detection Techniques 145
- Allele Distinction Techniques 145
  - Allele-Specific Hybridization 145
  - Restriction Digestion 146
  - Enzymatic Ligation 146
  - Enzymatic Polymerization 147
  - Structure-Specific Cleavage 147
- Signal Detection 147
  - Radioactivity 147
  - Mass Detection 147
  - Fluorescence Detection 148
  - Fluorescence Resonance Energy Transfer 148
  - Fluorescence Polarization 150
  - Chemiluminescence 150
- Assay Formats 150
  - Sorting via Charge, Length, and Mass 150
  - Sorting via Arrays 150
  - Sorting via Optical Spectra 152

SNP Genotyping Platforms 152
- TaqMan® (Hybridization–FRET–Solution) 152
- Pyrosequencing (Polymerization–Chemiluminescence–Solution) 154
- DASH™ (Hybridization–Fluorescence–Plate Array) 154
- Invader® (Cleavage–Fluorescence–Solution) 154
- MassEXTEND® (Polymerization-Mass Spectrometry–Solution) 155
- Masscode™ (Polymerization–Mass Spectrometry–Solution) 155
- SNPstream® (Polymerization–Fluorescence–Microarray) 156
- SNPlex™ (Ligation–Fluorescence–Electrophoresis) 156
- GoldenGate™ (Polymerization/Ligation–Fluorescence–Bead Array) 156

Tecan Ex. 2130, Page 1 of 31
QIAGEN Sciences, LLC v. Tecan Genomics - IPR2025-00028

Padlock Probe and MIP Assays
(Polymerization/Ligation–Fluorescence–Array/Electrophoresis).....157
GeneChip® (Hybridization–Fluorescence–Microarray) 157
Whole Genome Genotyping (Polymerization–Fluorescence–BeadArray™) 158
Caveat 158
Choosing a Method: A Few Clinical Considerations 158
Quantitative Data 159
Quantity of DNA 159
Known versus Unknown Variations 160
Speed to Results.....160
Possibilities and Perspectives 161
References 163

INTRODUCTION

Today's physicians have an expanding repertoire of molecular analytical options to assist in diagnosis and treatment of human disease. Medical research strives to provide tools to optimize patient care according to a variety of molecular measures including variations found in patients' genomes. As more is understood about how subtle genetic variations affect disease states and responses to drug treatment, the more important it becomes to extract genetic information in an easy and inexpensive way. An impressive number of alternative technologies for collecting genotypic data from patient samples currently exists; however, most of the techniques can be clearly understood based on a few basic principles.

This chapter will discuss in some detail methods for detection of single nucleotide polymorphisms (SNPs). For completeness we start with a basic description including SNP abundance and distribution, followed by a brief introduction of concepts such as haplotypes and linkage disequilibrium. The underlying anatomies of common SNP genotyping techniques are presented with emphasis on modes of allele discrimination, methods of signal detection, and the physical arrangements of the assays. We then present a summary of how different commercial platforms combine these three components along with the advantages and disadvantages of each strategy. Finally, we explore the current and potential applications of these technologies.

SINGLE NUCLEOTIDE POLYMORPHISM

A working definition of single nucleotide polymorphism or SNP (Figure 6.1) builds on several closely related terms encountered in the study of genetic variation. *Polymorphism* refers to any position (*locus*) in a genome where an alternative version of the sequence (*allele*) can be observed, on average, when comparing any 100 randomly chosen chromosomes from a given population. Genetic variants found less frequently, that is, loci where the frequency of the less common allele falls below this 1% criterion, are usually referred to as rare variants.[1] This frequency definition is a common source of confusion in discussions of genetics and human disease.

In a classical sense, mutations are aberrant DNA sequences that are necessary and sufficient to cause disease. The Online Mendelian Inheritance in Man (OMIM) database (http://www.ncbi.nlm.nih.gov/omim/) describes almost 1400 diseases with causes that fit this definition. In the context of genetic variation, however, a mutation is considered a DNA variant found infrequently in a population. Millions of recorded point mutations have no known links to diseases. Conversely, many polymorphisms are known to predispose individuals to diseases (reviewed by Zondervan and Cardon[2]). Thus, not all mutations are bad and not all polymorphisms are innocuous.

What distinguishes SNPs from other forms of genetic polymorphism is that the alleles at an SNP locus differ in DNA sequence by a single base. Although it is theoretically possible for an SNP locus to be tetra-allelic, the vast majority of SNPs have only two alleles. Curiously, roughly two thirds of human SNPs involve a single nucleotide substitution of a T for a C (Figure 6.2),

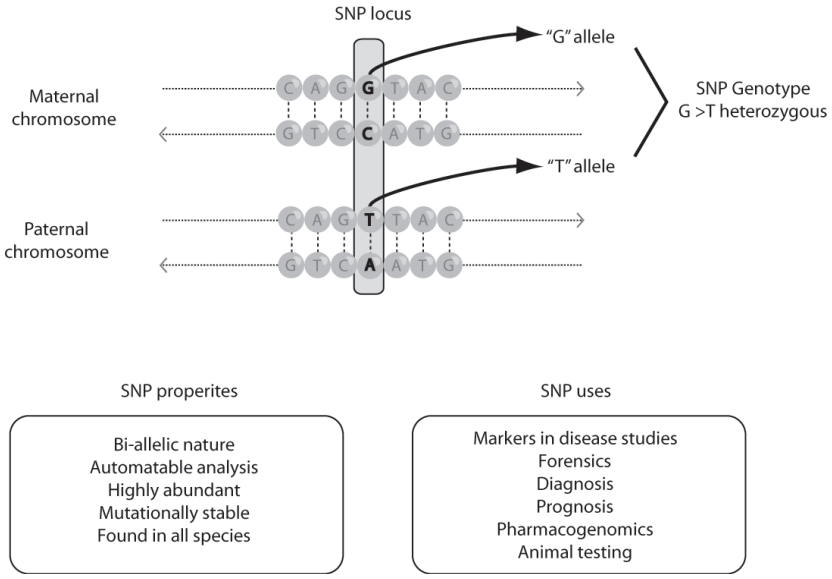


FIGURE 6.1 Fundamentals of single nucleotide polymorphisms (SNPs). A locus is a particular position in the genome. An allele is the particular nucleotide present at the locus. A genotype is the combination of alleles present in an individual.

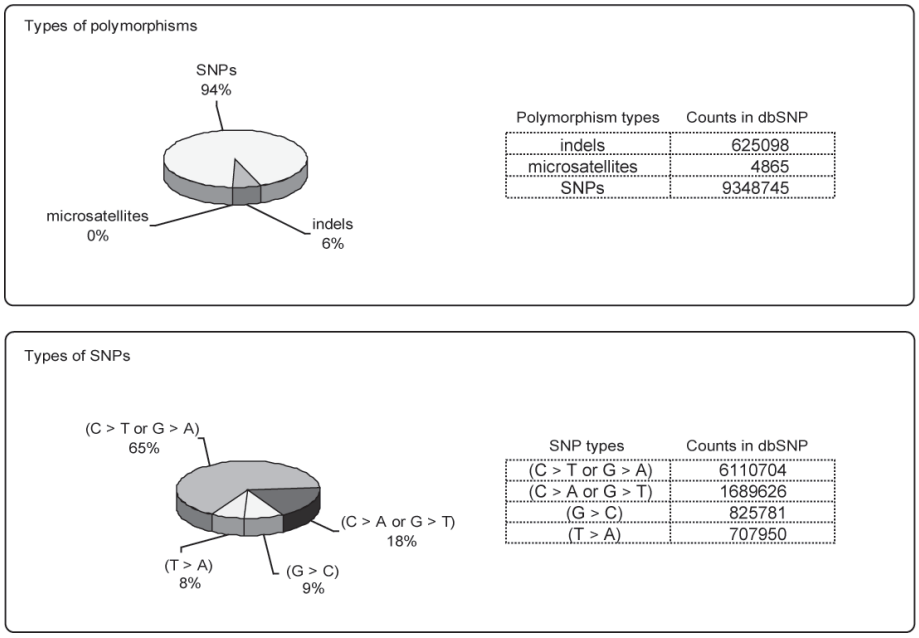


FIGURE 6.2 Summary of polymorphism data as present in dbSNP (build 124) for SNPs, insertions and deletions (indels), and microsatellites. For each section, distributions are represented in a pie chart, and the actual numbers presented in the adjacent table. Under types of SNPs, only four types of substitutions are presented. The reason for combining (C > T) and (G > A) polymorphisms is that in genomic DNA, a substitution of C for a T in one strand results in the G to A substitution in the complementary strand. Thus, the counts for (C > T) and (G > A) are combined to represent type of polymorphism present rather than polarity of DNA strand upon which the SNP was discovered. Similar reasoning is applied to (C > A) and (G > T) substitutions.

Tecan Ex. 2130, Page 3 of 31
QIAGEN Sciences, LLC v. Tecan Genomics - IPR2025-00028

probably arising through a de-amination mechanism known to occur frequently at CpG dinucleotides.[3] The allele with a higher frequency is termed the *major allele*; the other is designated the *minor allele*. Because the human genome is diploid in nature, an individual can be either *homozygous* (having two copies of the same allele at an SNP locus) or *heterozygous* (having one copy of each). The particular combination of alleles that an individual has at a particular locus is called a *genotype*.

Discovery, Abundance, and Distribution

With a complete sequence of the human genome as a reference,[4–6] major efforts have been asserted toward producing detailed maps of genome variation. Many SNPs have been defined by re-analysis of data generated from the human genome sequencing projects. Comparative analysis of DNA sequences from overlapping genomic clones,[5,7,8] shotgun sequencing assemblies,[7,9] and libraries of cDNA sequences[10–14] have been good sources for finding SNPs. A large contributor to the public discovery effort is the SNP Consortium or TSC (http://snp.cshl.org). Using some of the aforementioned strategies, this coalition of industry and academic partners, in a single publication in 2001,[7] released 1.4 million putative SNPs into the public domain. To date, the number has increased to a total of 1.8 million.

These SNPs and additional polymorphisms from many other private and public groups can be found in an online database dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/). This collection of sequence variants is freely accessible and contains information about SNPs and other types of genetic polymorphisms relevant to the study of human disease, for example, microsatellites and insertions/deletions (*indels*).[15] Single nucleotide polymorphisms are by far the most abundant (Figure 6.2). As of the writing of this chapter, there are almost 9.5 million SNP entries in dbSNP (build 124). Roughly half have been validated from at least two independent sources or using two genotyping methods. About a million of these SNP entries include allele frequency information from at least one population. HGVBase (http://hgvbase.cgb.ki.se)[16,17] is another online repository of polymorphisms that is highly curated and focuses on the relationship of phenotype and DNA variation. Additionally, a number of SNP and mutation databases centered on specific genes or diseases can also be found on the Web (Table 6.1).

An estimated 11 million SNPs are thought to reside in the human genome.[18] When averaged across the entire genome, the result is one SNP for every 300 bases. Discovery efforts indicate, however, that SNPs appear to vary widely in distribution among different genomic regions.[5] When SNPs are placed in the context of a chromosomal map, they occur roughly four times more frequently in non-coding sequences than in coding sequences.[19,20]

Within genes, there are fewer SNPs at splice junctions than in exons. Within exons, SNPs are much more likely to reflect "silent" or *synonymous* changes rather than *non-synonymous* changes that when translated would change the amino acid sequence of the encoded protein.[21–24] These trends make sense from a biological perspective, as sequence changes in introns and other non-coding regions are much less likely to adversely affect normal cell function than genetic variants that affect mRNA splicing or protein sequences. Other genes appear to be under diversifying selection, meaning that variation in gene sequence is biologically beneficial, resulting in greater numbers of polymorphisms within such genes. For example, the diversity of polymorphism found in the genes encoding the human leukocyte antigens (HLAs)[25] contributes to the repertoire of proteins and peptides that can be effectively recognized and presented during an immune response.

Linkage Disequilibrium, Haplotypes, and HapMap Project

The phenomenon of linkage disequilibrium (LD) has been widely discussed in the SNP literature.[26] Basically, LD is a statistical measure of relationships between alleles at neighboring polymorphic loci along a chromosome. Although alleles located close to one another tend to be co-inherited as

TABLE 6.1
Online Databases Containing Entries on Human Genetic Variation

Type of Database	Name	Description	Location
Public	dbSNP — a database of single nucleotide polymorphisms	Major online repository for DNA polymorphisms including SNPs, microsatellites, and small insertion/deletions	http://www.ncbi.nlm.nih.gov/SNP/
	HGVbase — human genome variation database	Large database of polymorphisms with emphasis on genotype/phenotype relationships	http://hgvbase.cgb.ki.se/
	GBrowse — SNP resource	Database of SNPs discovered by SNP Consortium, now hosted on International HapMap project website	http://www.hapmap.org/cgi-perl/gbrowse/gbrowse
	ALFRED — allele frequency database	Resource of gene frequency data on human populations supported by U.S. National Science Foundation	http://alfred.med.yale.edu/alfred/index.asp
	CEPH — genotype database	Database of genotypes for genetic markers typed on CEPH reference family resource for linkage mapping	http://www.cephb.fr/cephdb
	GeneSNPs	Database sponsored by Environmental Genome Project that integrates gene, sequence. and polymorphism data into individually annotated gene models	http://www.genome.utah.edu/genesnps/
Commercial	Perlegen genotype browser	Database of SNPs, linkage disequilibrium bins, and haplotype blocks across all three populations examined in the study by Hinds et al. [34]	http://genome.perlegen.com/browser/index.html
	Human SNP database	Online resource of SNPs created in collaboration with Affymetrix, Inc., sponsored and supported in part by pharmaceutical companies and the National Institute of Standards and Technology	http://www.broad.mit.edu/snp/human/
	SNPBrowser	Applied Biosystems software and database over SNPs integrated with assay information	www.allsnps.com/snpbrowser.
Mutation	HGMD — human gene mutation database	Database of various types of mutation within coding regions of human nuclear genes causing inherited disease	http://archive.uwcm.ac.uk/uwcm/mg/hgmd0.html
	Mutation database website	Database sponsored by HUGO Mutation Database Initiative (MDI) that hosts links to numerous mutation and DNA variation databases	http://www2.ebi.ac.uk/mutations/cotton/
	Genome Web — human mutation database	Extensive collection of links to human mutation and SNP databases	http://www.hgmp.mrc.ac.uk/GenomeWeb/human-gen-db-mutation.html
Clinical resource	EDDNAL — European Directory of DNA Diagnostic Laboratories	Directory of non-profit European laboratories offering diagnostic services for numerous diseases	http://www.eddnal.com/
	PharmGKB — pharmacogenetics and pharmacogenomics knowledge base	Integrated resource about how variation in human genes leads to variation in drug response	http://www.pharmgkb.org/

QIAGEN Sciences, LLC v. Tecan Genomics - IPR2025-00028
Tecan Ex. 2130, Page 5 of 31

a unit, alleles at loci that are far apart on a chromosome tend to be shuffled as an effect of recombination during meiosis. LD can be thought of as a measurement of the probability that two or more alleles occur together because they are linked as compared to co-occurrence because of chance. There are several alternative methods for calculating LD based on different statistical models of inference.[27–33]

The existence of LD is the underlying reason that it is possible to identify discrete collections of alleles from nearby loci that tend to occur together. These clusters or combinations of alleles are termed *haplotypes*.[34,35] Since recombination is spotty and occurs with varying frequency along the chromosomes,[36–38] it creates many islands of chromosomal segments where LD is strong and clusters of common haplotypes can be identified.[39–43] These conserved regions are called haplotype blocks.

In addition to recombination, the sizes and compositions of haplotype blocks are also influenced by other forces such as natural selection or genetic drift.[2] A current international effort called the HapMap Project is underway to identify common haplotype blocks across genomes in four human populations.[44] Once completed, haplotype frequency and distribution information can be used to guide optimal selection of SNPs for studies designed to identify genetic factors influencing human disease.

SNPs in Inherited Disease

The demands on SNP genotyping techniques vary greatly, depending on intended applications, for example, studies designed to seek the genetic determinants of disease where hereditary components are implicated. A classic approach is to perform an association study (reviewed in Reference 45). In this design, a group of individuals affected with a disease of interest (cases), and a group of people from the general population (controls) are selected. The cases and controls are chosen to be as closely matched as possible to avoid stratification due to genetic and environmental differences that are irrelevant for the investigated trait.

One or more SNPs located in genes suspected to be biologically relevant for the disease are chosen and genotyping is performed in the two groups. Statistically significant differences in allele or genotype frequencies of the two groups are taken as evidence of disease association. A positive association can be attributed to either the tested SNP allele directly influencing the risk of disease or indirect action via linkage disequilibrium with a nearby disease-influencing genetic variant.[46] Typically, a limited number of SNPs are genotyped in a relatively small number of individuals.

Many common diseases have been studied with this traditional approach[47] — with precious little success.[48–50] One possible explanation is that many common diseases are caused by combinations of alleles at a number of loci that act in concert to influence disease. Under such conditions, analysis of individual loci would be insufficient to identify genetic risk factors of that disease. The common disease/common variant (CD/CV) theory[51] proposes that some or all common diseases are due to relatively few high frequency risk alleles present at a few major disease loci. Since common alleles are traceable through haplotypes, the CD/CV theory has inspired much of the work of the HapMap Project and fueled interest in the possibility of performing whole genome association studies.[52,53] Rather than relying on biological guesswork, this "hypothesis-free" approach involves performing a large number of association studies on SNPs covering the entire human genome. Information from the HapMap Project will be useful for such an approach; SNPs to include in the study could be chosen to scan most of the common haplotypes using the most informative SNPs.[54]

The sensibility of performing whole genome association studies is a source of great controversy,[55–58] but the demand for higher throughput–lower cost genotyping has made a dramatic impact on SNP profiling techniques. At least four companies have responded and are developing assays specifically for whole genome analysis: Affymetrix (Santa Clara, California, USA, http://www.affymetrix.com/), Centurion DNA Array; Illumina (San Diego, California, USA, http://www.illumina.com/), Linkage IV Panel; Perlegen (Santa Clara, California, USA, http://www.perlegen.com/), Whole Genome Scanning Collection; and ParAllele BioScience (South San Francisco, California, USA,

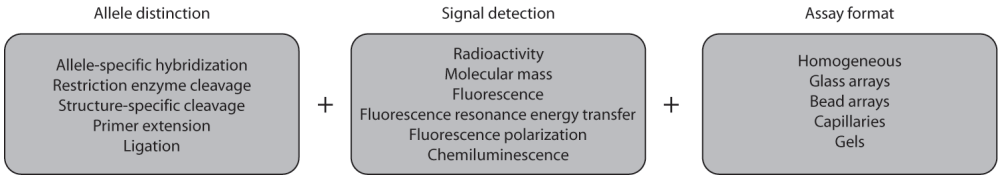


FIGURE 6.3 The anatomy of SNP genotyping protocols. Genotyping platforms often involve combinations of one or more of these allele distinction techniques, along with a signal detection strategy using one of these assay formats.

http://www.parallelebio.com), MegAllele™ Mapping Set. These systems allow the prospect of genotyping many thousands to possibly millions of SNPs simultaneously and in multiplex at a fraction of the previous cost.

SNP DETECTION TECHNIQUES

As noted earlier, genetic studies vary greatly in design and consequently impose very different requirements on SNP typing methods. Fortunately, researchers have a generous variety of innovative SNP genotyping solutions to choose from, each with its own advantages and disadvantages. All these techniques can be thought of as particular assemblies of three fundamental components (Figure 6.3):

1. Allele distinction
2. Signal detection
3. Assay format

In brief, allele distinction refers to the reaction principle by which alleles are distinguished. Signal detection is the way in which allele-specific products are visualized, and the assay format is simply the milieu in which the reactions take place. In the following sections, these key building blocks are described individually, setting the stage for an understanding of the architectures of commercial SNP genotyping systems.

In addition, the reaction format of most SNP detection techniques relies on amplification of the target molecule prior to the allele discrimination step, while other techniques amplify signals solely during the detection step. Target amplification serves to both reduce complexity of the genetic sample to be interrogated and to enhance subsequent detection signals via the increased quantity of targets. If allele distinction is performed directly on the genomic material, the intent of a signal amplification step is to increase the specific genotyping signal relative to background. In both cases, the use of the polymerase chain reaction (PCR) is by far the most common means of enhancing target concentration and/or signal intensity.

Allele Distinction Techniques

A central aspect of any genotyping method is the technique used to distinguish different allelic variants. In the case of SNP genotyping, this requires that single base differences in a certain position can be identified and scored. The five fundamentally different techniques currently in use for this purpose are (1) allele-specific hybridization, (2) restriction enzyme cleavage, (3) enzymatic ligation, (4) primer extension by polymerases, and (5) structure-specific cleavage (Figure 6.4).

Allele-Specific Hybridization

Allele-specific hybridization techniques take advantage of the fact that the stability of the duplex formed between two nucleic acid molecules is sensitive to mismatches in the duplex where the two

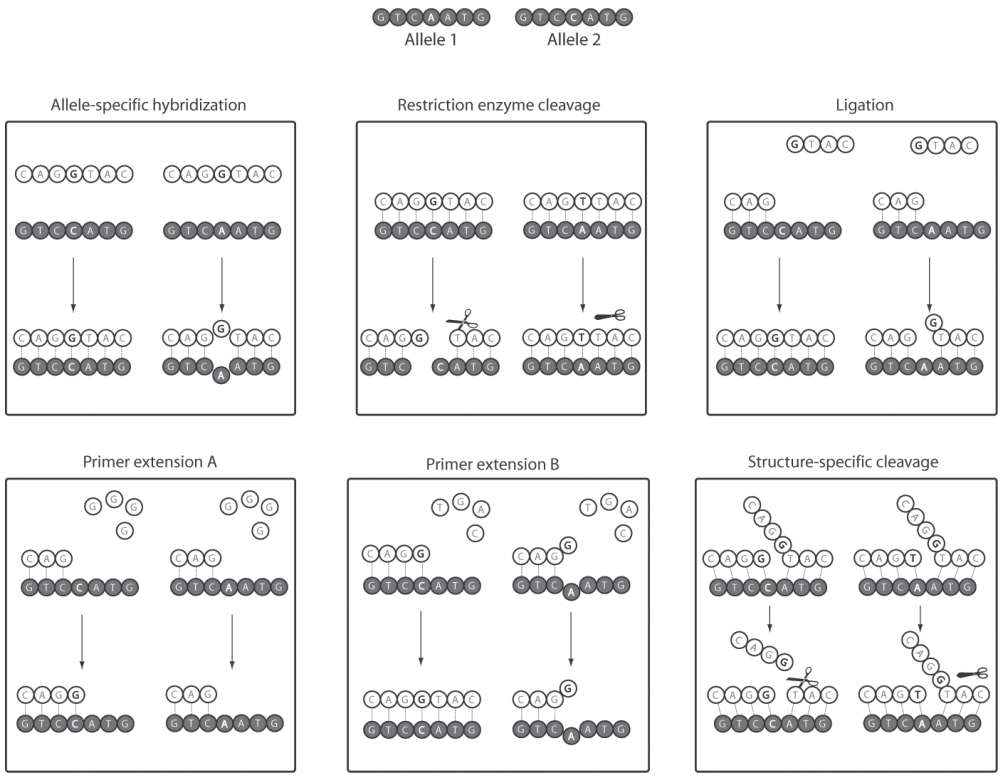


FIGURE 6.4 Principles of allele distinction. Target molecules are represented by dark circles, probe sequences by grey circles. Two different mechanisms for allele distinction by primer extension are presented. Primer extension A depicts the single base pair extension approach where polymerization results in incorporation of an allele-specific nucleotide. Primer extension B refers to allele-specific primer extension in which polymerization occurs only with primers that are complementary at the polymorphic position.

opposing bases do not form a Watson–Crick base-pair. This is utilized by having oligonucleotide probes form duplexes with the target molecules. Perfectly matched duplexes exhibit greater stability, and thus require higher temperatures to separate, than duplexes with mismatched positions. The existence of duplexes can be monitored by different means to reveal the alleles present in a sample.

Hybridization modifiers such as minor-groove binders (MGBs)[59] and locked nucleic acids (LNAs)[60] have been incorporated into a variety of genotyping platforms. These modifications increase the stability of DNA duplexes, thereby enabling the use of shorter oligonucleotides in analyses which in turn leads to greater differences in allele-specific hybridization reactions.

Restriction Digestion

Restriction enzymes cleave DNA molecules at specific sequences. The substitution of a single nucleotide for another frequently abolishes or creates restriction enzyme recognition sequences, thereby altering the ability of the restriction enzyme to cleave. Thus, single nucleotide differences may be detected as a change in the set of fragments resulting from a restriction enzyme cleavage reaction.[61]

Enzymatic Ligation

Ligation-based distinction techniques utilize the ability of DNA ligase enzymes to discriminate between matched and mismatched ligation substrates. In these methods, oligonucleotide probes are

Tecan Ex. 2130, Page 8 of 31
QIAGEN Sciences, LLC v. Tecan Genomics - IPR2025-00028

hybridized to the target molecule so that a 5′ end and a 3′ end are positioned immediately next to each other. If the ends are perfectly matched to the target, a ligase is able to join the two ends. If either end is mismatched (in particular the 3′ end), ligation is greatly inhibited.[62]

Enzymatic Polymerization

Primer extension is used for allele distinction in several SNP genotyping strategies. A DNA polymerase is used to interrogate the base at the variable position. Two different approaches are possible here. In the first approach (Figure 6.4, Primer Extension A), a primer is designed to hybridize to the target with its 3′ end complementary to a position immediately 3′ of the nucleotide to be interrogated, so that the nucleotide incorporated in the primer by the polymerase will reflect the identity of the nucleotide present at the variable position.[63]

The nucleotide incorporated can then be determined by adding only one type of nucleotide at a time,[64,65] by using differentially labeled nucleotides[66.67] or by using a mix of nucleotides (some terminating or absent) so that the length of the extension product depends on what nucleotide can be incorporated.[68,69] In the second approach, allele-specific primers are designed to be extended only if they are correctly matched to the target at the variable position[70,71] or, in an alternative approach, only if mismatched.[72]

Structure-Specific Cleavage

The structure-specific cleavage or "invader" technique uses the capacities of certain classes of enzymes to cleave nucleic acid molecules at the site of a specific structure formed when a DNA strand hybridizes to a target nucleic acid molecule in such a way that its 3′ end partially displaces another nucleic acid. The displaced strand is then cleaved at the position corresponding to the 3′ end of the displacing or invading strand. The target strand and the displaced strand must be matched at the invasion site for this cleavage to occur, thus allowing allele discrimination.[73]

Signal Detection

Rather than detecting hybridizing or enzymatically modified DNA directly, most SNP screening techniques measure a change in a surrogate molecule that is modified during or as a result of the allele distinction process. This is typically accomplished through addition of labels or tags to DNA probes, which can be followed through changes in one of many physical properties. Physical properties used in SNP detection strategies include radioactivity, molecular mass, and a variety of luminescence phenomena such as fluorescence, fluorescence resonance energy transfer (FRET), fluorescence polarization (FP), and chemiluminescence.

Radioactivity

One of the earliest approaches to detecting DNA molecules was through attachment of radioactive tags. Different isotopes such as ^{3}H, ^{32}P, or ^{35}S can be readily incorporated into DNA via polymerase or kinase reactions and the labeled products can be detected by autoradiography or other radiation measuring devices such as scintillation counters or phosphoimagers. Although radiolabeling is very sensitive, commercial platforms have adopted more benign alternatives for DNA detection.

Mass Detection

Mass spectrometry (MS) can be used to measure small differences in molecular weights of allele-specific reaction products. MS is also the only signal detection technique that does not require labeling of DNA prior to detection. The MS technique termed matrix-assisted laser desorption/ion-

ization time-of-flight mass spectrometry (MALDI-TOF) has been successfully implemented in a number of SNP detection schemes (reviewed in Reference 74). In these approaches, small amounts of allele-specific reaction product and matrix are co-deposited onto a target plate. A short laser pulse serves to desorb the DNA sample/matrix into the gas phase and the ionized products are accelerated toward the detector.

The mass of the product can be inferred from the relative time it takes for the product to travel from the point of ionization to the detector. The resolution of this type of mass measure allows for even single base differences to be distinguished with ease in short DNA fragments. In addition, the dynamic range of MALDI-TOF has been extended by incorporation of mass tags into the allele-specific products.[75] Mass tags are small molecules with defined masses that can be designed to differ by only a few Daltons. Libraries of mass tags can be used to sort allele-specific products from several SNP loci in parallel.

Fluorescence Detection

Fluorescence detection involves the measurement of emitted light. In principle, a fluorescent dye or *fluorophore* is excited with an appropriate wavelength of light and induces an electron to enter an excited state. As the electron returns to the more stable ground state, excess energy is dissipated as fluorescence or longer-lived phosphorescence, depending on the type of fluorophore used. The emitted light is always of a lower energy/longer wavelength than that of the excitation source. Fluorophores having specific excitation and emission spectra that span a wide range of wavelengths are commonly available. By selecting fluorophores with dissimilar emission spectra, differently labeled DNA molecules can be resolved on the bases of their fluorescent signatures.

Many fluorescent detection reagents have been developed to stain or tag DNA molecules for SNP genotyping. For example, SYBR® Green I from Molecular Probes (Eugene, Oregon, USA, http://probes.invitrogen.com/) is an intercalating dye that undergoes a 10,000-fold increase in fluorescence when bound to double-stranded DNA. This dye has proven useful in assays that detect the presence or absence of DNA duplexes as part of a genotyping procedure.[76] Dyes with other desirable characteristics such as a large *Stokes shift* (difference between peak excitation and emission wavelengths), a high *extinction coefficient* (ability to capture excitation photons), and a high *quantum yield* (efficiency of converting between incoming and outgoing photons) are under continual development, exemplified by the Bodipy® and Alexa Flour® dye series also from Molecular Probes. Quantum dots or qdots represent a relatively new class of fluorescent labels composed of Zn or Cd nanoparticles encapsulated in a biocompatible, hydrophilic shell. Among the advantages of qdots over other fluorophores is that they are highly resistant to *photobleaching* (fluorophore destruction due to over-excitation), have relatively narrow emission spectra, and can all be excited at a fixed wavelength of light.

Fluorescence Resonance Energy Transfer

Fluorescence resonance energy transfer (FRET) is a phenomenon that occurs between appropriately matched pairs of fluorophores. The process involves the transfer of excitation energy from a *donor* fluorophore to an *acceptor* fluorophore located in close proximity. The donor does not emit fluorescence; rather it gives its energy to the acceptor fluorophore via dipole–dipole interaction. Upon energy transfer, the acceptor enters an excited state and can then dissipate the energy according to its own spectral properties. For FRET to occur, the emission spectrum of the donor must overlap the excitation spectrum of the acceptor.

In addition, the efficiency of energy transfer is greatly influenced by the spatial relationship between the donor and acceptor. In general, the closer the donor is to the acceptor, the more efficient is the energy transfer. Many SNP detection strategies have been developed in which a FRET reaction changes as a response to the allele discrimination event (reviewed in Reference 77; see Figure 6.5).

Tecan Ex. 2130, Page 10 of 31
QIAGEN Sciences, LLC v. Tecan Genomics - IPR2025-00028

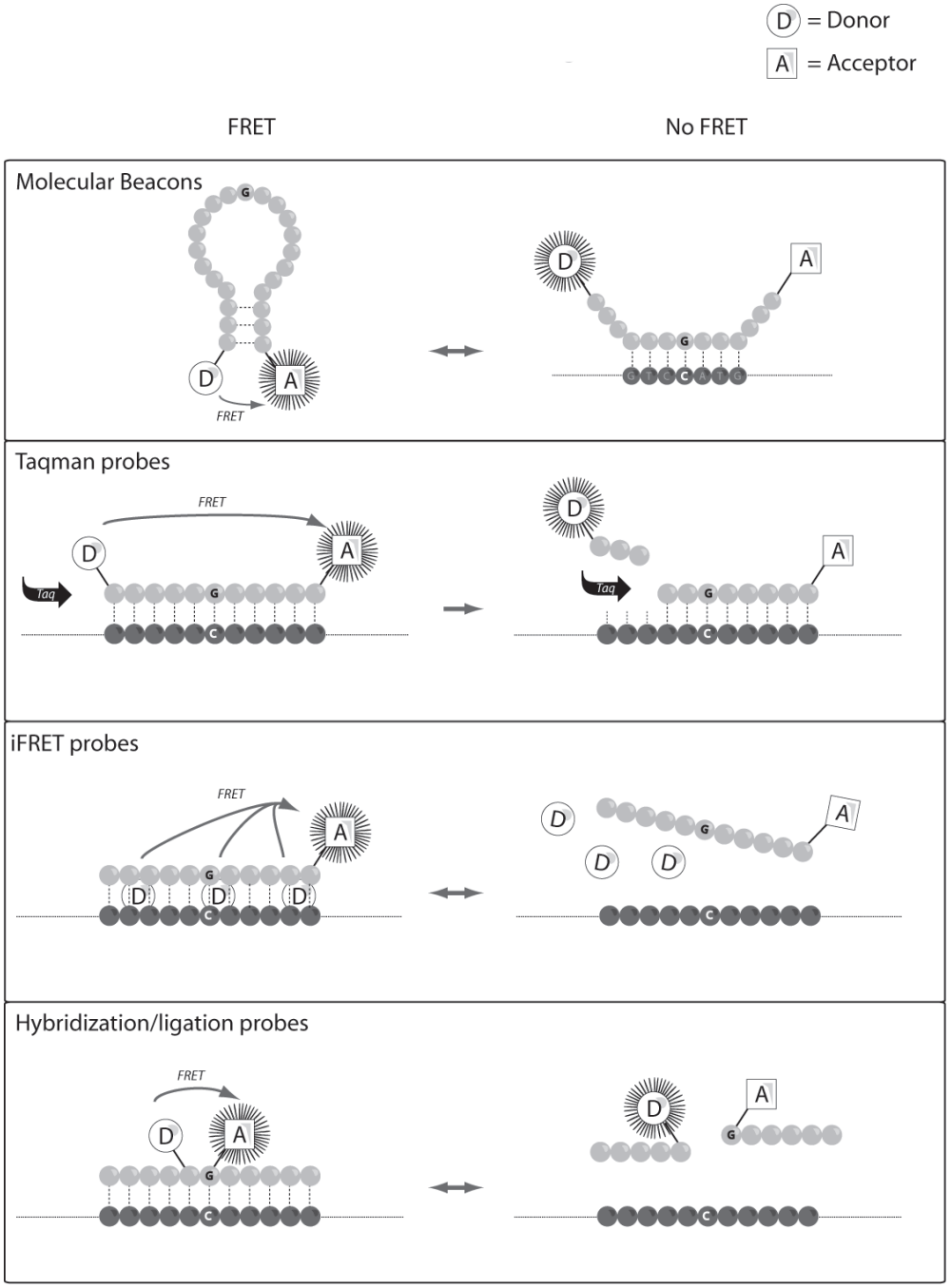


FIGURE 6.5 Alternative FRET-based detection strategies. A depiction of how FRET donor (D) and acceptor (A) molecules (sometimes referred to as fluorophore and quencher moieties) are separated as part of the allele discrimination event for a number of different platforms. Radiating lines depict fluorescent emission.

Most methods rely on covalent attachment or linkage of donor and acceptor fluorophores to a probe DNA molecule. One variation known as induced fluorescent energy transfer (iFRET™)[78] circumvents the necessity for physical attachment of the donor by employing the intercalating SYBR Green I dye as a donor.

Tecan Ex. 2130, Page 11 of 31
QIAGEN Sciences, LLC v. Tecan Genomics - IPR2025-00028

Fluorescence Polarization

Another aspect of luminescence that has been used successfully for SNP detection is measuring the polarization of light emitted by fluorophores excited with polarized light. A stationary fluorophore, when excited with plane-polarized light, will emit fluorescence with the same plane of polarization. If, however, the fluorophore is free to tumble in solution before light is emitted, then emission polarization is reduced or lost. Reduction or increase of size of a fluorophore-labeled DNA molecule as part of an allele-specific reaction will influence the rate of tumbling, and thereby result in loss or gain of fluorescence polarization (FP) that is detectable with appropriate instrumentation (reviewed in Reference 79).

Chemiluminescence

Light produced via a chemical or biochemical reaction is referred to as chemiluminescence. The process often involves the enzymatic conversion of a substrate from one form to another, with part of the catalysis involving emission of light. A prominent example in SNP typing is the activity of the luciferase enzyme on the luciferin substrate that produces a greenish yellow fluorescent emission.

Assay Formats

Genetic material can be organized in a number of ways prior to analysis. Solution-based analyses can take place in tubes, microtiter plates, or capillaries. For assays involving immobilized genetic material, a wide variety of solid supports including various types of beads, glass slides, membranes, and gel matrices are available. Active movement of reactants between different formats often involves pipetting, electrophoresis, magnetism, or fluidic manipulation.

Since target amplification-based and signal amplification-based techniques, microarrays, and bead flow systems are discussed in depth in other chapters of this book, the following section will discuss only a few examples of the integration of these principles and components into the sorting procedures of various SNP genotyping techniques.

Sorting via Charge, Length, and Mass

DNA has an inherent negative charge due to its chemical composition. Thus, when placed in an electrical field, DNA tends to migrate toward the positive pole. This property is used in electrophoresis in which DNA molecules are electrically drawn through a gel matrix. The matrix, usually composed of agarose or polyacrylamide, serves to separate DNA fragments based on length and strand conformation. Slab-gel and capillary electrophoresis are common formats for many molecular biological techniques including SNP genotyping.

Similar to electrophoresis, mass spectrometry uses the charges on ionized DNA molecules to accelerate the molecules within an electric field. DNA fragments are further sorted and identified according to differences in molecular mass. In addition to relying on the physical properties inherent to DNA, tags influencing mass, length, or electrophoretic mobility can be attached to probes to further enhance separation of allele-specific signals.

Sorting via Arrays

Arranging clusters of DNA molecules in a two-dimensional space is a highly effective means of separating reactions and signals. In the case of microarrays, DNA can be synthesized separately and spotted to particular locations on a glass slide or synthesis can take place directly on a surface. The spots or *features* can be made to contain a series of probes that correspond to a number of different SNP loci and subsequent allele distinction occurs in these features. Alternatively, generic arrays can be created.

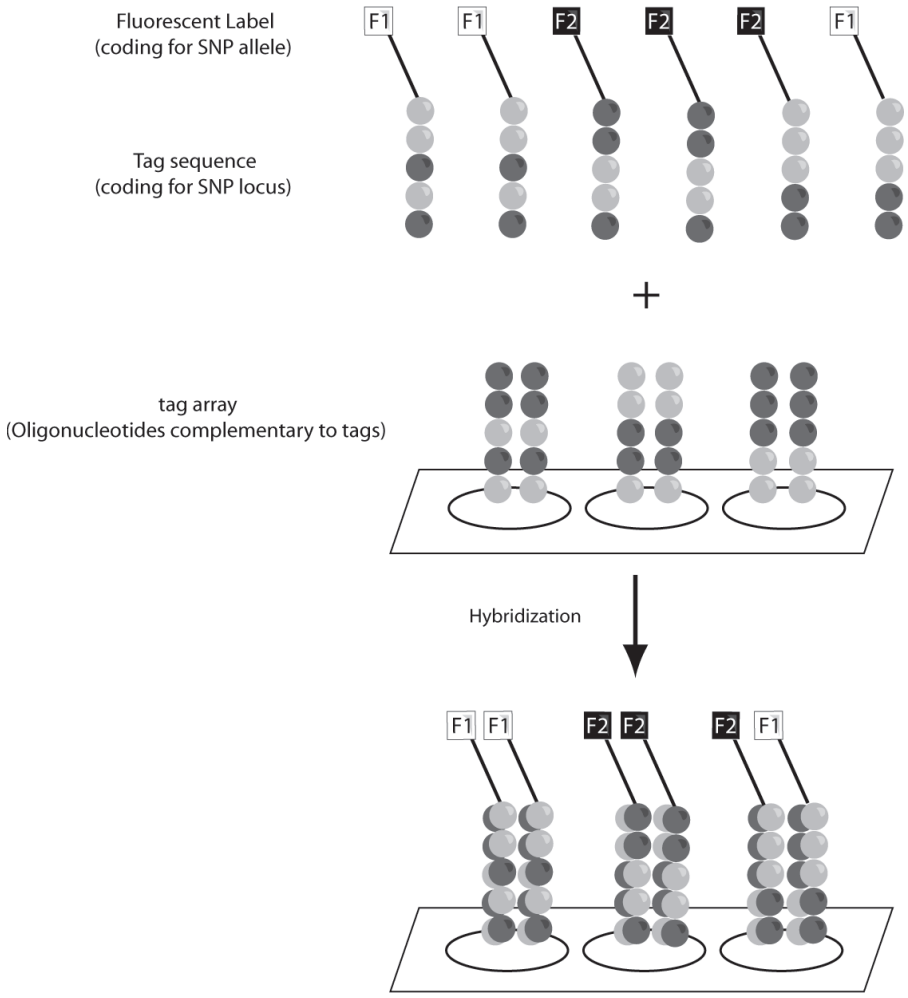


FIGURE 6.6 Principles of a tag array. Multiplex SNP detection techniques such as the SNPStream, GoldenGate, and MIP assays result in production of DNA molecules bearing a tag sequence coding for the SNP locus, and a fluorescent label coding for the allele detected. These products are then hybridized to a complementary tag array and genotypes are scored by comparison of fluorescence at each feature of the tag array.

Instead of representations of target or probe sequences, a generic or tag array involves the immobilization of a series of unique sequences to the surface. Complementary sequences to the immobilized tags can be built into SNP detection probes. After an allele discrimination reaction, probes can be sorted on the tag array via selective hybridization of tail sequences to corresponding array tags (Figure 6.6). Thus, the same array can be used to assay many different SNPs as the specificity lies in design of the probe rather than the array.

Other types of supports for arrays are microparticles or beads. They are often small silica or plastic particles that have been surface functionalized for the attachment of DNA. Once DNA is stuck to a bead, it is possible to interrogate the beads in fluidic suspensions for determination of the identity of the genetic material on the bead. It is also possible to create a two-dimensional structure of beads by capture of the beads at appropriate locations on a surface. Both specific and tag array formats have been successfully implemented with beads.

Tecan Ex. 2130, Page 13 of 31
QIAGEN Sciences, LLC v. Tecan Genomics - IPR2025-00028

Beads have a number of useful properties. They have large surface-to-mass ratios, facilitating attachment of large numbers of DNA molecules. Furthermore, reactions involving beads with immobilized DNA behave more like solution-phase assays than the linear arrays because the beads float around and come into contact with more solution. The beads can be labeled and aid in decoding the signals from the genotyping procedure.

Sorting via Optical Spectra

Sorting of mixtures of DNA can also be accomplished through spectral analysis. By labeling DNA with different fluorescent dyes, examination of a mixed sample over different portions of the emission spectra will allow inference as the sum totality of the contents of the tube. Otherwise referred to as *spectral multiplexing*, this type of signal separation permits simultaneous detection of several different allele-specific products. The use of optical spectral separation can be combined with microarrays or bead arrays to increase the information content of assays performed in those formats.

SNP GENOTYPING PLATFORMS

Drawing on the common elements of allele distinction, signal detection, and reaction format, a diversified assortment of SNP typing platforms have been commercialized and put into use in clinical labs around the world. A number of commercially available genotyping platforms are listed in Table 6.2. In the following section, a number of SNP typing strategies are described along with details regarding nuances of each technique. Variations of these techniques are also described to illustrate modifications commonly encountered in applying the various methods.

TaqMan® (Hybridization–FRET–Solution)

This assay from Applied Biosystems (Foster City, California, USA, http://www.appliedbiosystems.com/) involves a pair of PCR primers and two detection probes.[80,81] The PCR primers are designed to amplify a short DNA fragment containing the polymorphic locus. Two probes, one for each allele and labeled with different colored FRET pairs of fluorophores, are designed to hybridize across the polymorphic position. PCR conditions are chosen so that allele-specific hybridization occurs between probe and PCR target sequences. During the extension step of the PCR cycle, *Taq* polymerase adds nucleotides as it reads from the target strand. Upon encountering a correctly annealed probe, the nuclease activity of *Taq* polymerase cleaves the probe.

Digestion of the probe causes the separation of the FRET donor and acceptor, resulting in an increase of fluorescence from the donor. During PCR, this process is cycled, resulting in accumulation of donor fluorescence throughout the PCR process. The fractional point (cycle number) at which the fluorescence signal from the digested probe exceeds the assay background is termed the *Ct value*. Collection of data in real time allows for quantitative analysis over a broad range of target DNA concentrations, but for simple scoring of SNP genotypes, endpoint analysis is sufficient since the presence or absence of each allele can be inferred via comparison of total fluorescence from each probe.[82]

The major advantage of the TaqMan assay is its homogeneous format. After patient DNA and reaction mix are added to a reaction vessel, the tube is sealed and never has to be opened again, reducing the chances of cross-contamination. Also, Applied Biosystems has provided a number of conveniences for users. Hundreds of thousands of analytically pre-validated assays along with searchable mapping software over the chromosomal locations of the SNPs are available online.

A related strategy for homogeneous SNP detection using fluorescence involves molecular beacon (MB) probes.[83,84] An MB probe consists of an oligonucleotide bearing a member of a FRET pair at each end. The middle of the MB probe encodes the allele recognition sequence. The ends of the oligonucleotide are designed to be weakly self-complementary. In the absence of target, the

Tecan Ex. 2130, Page 14 of 31
QIAGEN Sciences, LLC v. Tecan Genomics - IPR2025-00028

TABLE 6.2
Classification of Common SNP Genotyping Platforms

Allele Distinction	Detection Mechanism	Reaction Format	Platform and Provider	Website
Hybridization	FRET	Solution	TaqMan, Applied Biosystems	http://www.appliedbiosystems.com/
Enzymatic polymerization	Chemiluminescence	Solution	Pyrosequencing, Biotage	http://www.pyrosequencing.com/
Hybridization	iFRET	Array	DASH, DynaMetrix	http://www.dynametrix-ltd.com/
Structure-specific cleavage	Fluorescence	Solution	Invader, Third Wave Technologies	http://www.twt.com/
Enzymatic polymerization	Mass spectrosopy	Solution	Homogenous MassExtend, Sequenom	www.sequenom.com/
Enzymatic polymerization	Mass spectroscopy	Solution	MASScode, Qiagen	http://www1.qiagen.com/
Enzymatic polymerization	Fluorescence	Microarray	SNPstream, Beckman Coulter	http://www.beckman.com/
Enzymatic polymerization	Fluorescence	Microarray	APEX, Asper Biotech	http://www.asperbio.com/
Ligation	Fluorescence	Electrophoresis	SNPlex, Applied Biosystems	http://www.appliedbiosystems.com/
Enzymatic polymerization	Fluorescence	Bead array	GoldenGate, Illumina	http://www.illumina.com/
Ligation	Fluorescence	Microarray	MIP, ParAllele	http://www.parallelebio.com/
Ligation	Fluorescence	Electrophoresis	SNPWave, Keygene	http://www.keygene.com/
Hybridization	Fluorescence	Microarray	GeneChip, Affymetrix	http://www.affymetrix.com/
Enzymatic polymerization	Fluorescence	Beadarray	WGG, Illumina	http://www.illumina.com/

Tecan Ex. 2130, Page 15 of 31
QIAGEN Sciences, LLC v. Tecan Genomics - IPR2025-00028

MB probe forms a stem–loop structure bringing the FRET pair close together and thus enabling energy transfer. Upon hybridization of the MB probe to a target, the stem structure is opened, separating the FRET pair. An increase in donor fluorescence is an indication of hybridization of the MB probe to its target.

Pyrosequencing (Polymerization–Chemiluminescence–Solution)

When a polymerase enzyme extends a primer by the incorporation of a nucleotide added as a nucleotide triphosphate, a pyrophosphate (PPi) molecule is released. By adding a single kind of deoxyribonucleotide triphosphate at a time to a primer extension reaction and monitoring the release of PPi once incorporation occurs, it is possible to extract sequence information from a sample. This is utilized in the pyrosequencing assay[85] currently marketed by Biotage (Uppsala, Sweden, http://www.biotage.com/) in which several cycles of nucleotide addition are performed and the PPi release is monitored in real time. The detection is based on the conversion of PPi to ATP and ATP to emitted light in an enzymatic chain reaction.

Different chemistries and ways of removing unincorporated deoxyribonucleotide triphosphates are in use.[86] Like many other assays, pyrosequencing requires amplification of the genomic target regions prior to analysis, and the assay has a very limited capacity for multiplexing. However, its ability to extract stretches of DNA sequence information from a sample makes it useful also for purposes other than SNP genotyping.

DASH™ (Hybridization–Fluorescence–Plate Array)

Dynamic allele-specific hybridization (DASH), supported by DynaMetrix Ltd. (Leicester, United Kingdom, www.dynametrix-ltd.com/) is based on analysis of DNA melting curves.[76,87] The assay involves PCR amplification of the SNP locus and subsequent immobilization to a well of a streptavidin-coated microtiter plate via a biotin attached to one of the PCR primers.

Following removal of excess PCR reagents and the non-biotinylated strand, a detection probe is hybridized over the polymorphic positions on the solid phase targets. The SYBR Green I dye is added in solution and allowed to bind to the DNA duplexes. The samples are heated through a temperature range while fluorescence is continually monitored. As the melting temperature (Tm) of each allele-specific duplex is surpassed, the DNA strands separate, releasing the dye and causing a dramatic drop in fluorescence. The DNA duplexes formed with targets containing the 100% complementary or "matched" allele are more stable than duplexes formed with the alternate or "mismatched" allele.

A plot of fluorescence versus temperature thus reveals the Tms of the duplexes in the sample, and genotypes are inferred by comparison of their relative Tms. The most recent rendition of the assay termed DASH-2[88] involves fluorescence detection via iFRET™ detection[78] as facilitated by attachment of a FRET acceptor to a detection probe and the use of a macroarray format created by robotic spotting of samples on a streptavidin-coated membrane.

Invader® (Cleavage–Fluorescence–Solution)

The Invader assay from Third Wave Molecular Diagnostics (Madison, Wisconsin, USA, http://www.twt.com/) utilizes structure-specific endonucleolytic cleavage to achieve SNP genotyping.[89,90] A locus-specific invader oligonucleotide is used to partially displace or invade one of two allele-specific probes hybridized to a PCR-amplified target sequence. These allele-specific probes differ in the base positions opposite the SNP to be analyzed. The 3′ end of the locus-specific oligonucleotide displaces the allele-specific probe at the SNP position, and the specificity of the endonuclease results in the cleavage of only those probes that are correctly matched to the target.

The cleavage that occurs thus reflects the alleles of the SNP and is commonly detected by fluorescence. In the original set-up, the allele-specific probes are differentially labeled with fluorophores at their 5′ ends and they have quenchers 3′ of the cleavage site. Thus, cleavage results in a pattern of fluorescence that reveals the alleles of the SNP. Reaction conditions can be chosen so that probes will repeatedly hybridize to and dissociate from the target molecule, resulting in multiple cleavage events per target in an isothermal reaction.

In another set-up known as the serial invasive signal amplification reaction (SISAR), the allele-specific probes are unlabeled, and secondary detection probes are used instead.[91] The fragments cleaved off the primary detection probes are used to invade corresponding secondary probes that are labeled as previously. This procedure results in increased signal amplification and allows the detection of sequences directly from limited amounts of genomic DNA. Also, the expensive dually labeled probes are generic and a single pair of such probes may be used to genotype many SNPs, using many different pairs of primary probes. Use of one pair of detection probes, however, does limit the number of SNPs that can be scored in each genotyping reaction.

MassEXTEND® (Polymerization–Mass Spectrometry–Solution)

The homogeneous MassEXTEND (hME) assay[68] is an SNP genotyping strategy implemented on the MassARRAY™ system[92] of Sequenom (San Diego, California, USA, http://www.sequenom.com/). The hME assay involves initial PCR amplification of target DNA containing the polymorphic site. Following an enzymatic step to deactivate excess PCR reagents, a locus-specific oligonucleotide is added and annealed adjacent to the SNP position. DNA polymerase and a cocktail of three dNTPs (nucleotides) and one selected ddNTP (terminating nucleotide) are added to the sample. The annealed oligonucleotide acts as a primer and is extended by the DNA polymerase until a terminating nucleotide is incorporated. Allelic differences in the targets give rise to extension products of different lengths. After the extension reactions, the samples are purified with a resin and arrayed onto a silicon chip (384 SpectroCHIP™ bioarray) preloaded with matrix. The SpectroCHIP is then read by the MALDI-TOF instrument and genotype calls are recorded by measuring the masses of the extension products observed in the sample compared to a theoretical mass based on the allele-specific differences in length expected from the extension reaction.

The hME assay is used to illustrate SNP genotyping by multiple-base extension and MS detection as Sequenom offers a turn-key solution for performing SNP genotyping including chemistries, sample preparation robotics, MALDI-TOF instrumentation, and analysis software. The company also has validated assays for several hundred thousand SNPs and the current system offers up to 12-fold multiplexing.

The predecessor of the MassEXTEND is called the primer oligo base extension (PROBE) assay.[93] The systems differ in several ways. The PROBE procedure involves immobilization of PCR products to streptavidin-coated beads via a biotin label present in one PCR primer. The extension reaction occurs on the solid phase target and sample purification is via ethanol precipitation.

Another MS-based method called the GOOD assay[94,95] avoids the need for column purification or ethanol precipitation. A chemical modification and enzymatic digestion strategy enables preparation of samples for analysis in as few as three addition reactions.[96] The PROBE and GOOD assays and several other SNP genotyping platforms based on MALDI-TOF are extensively described in a recent review.[97]

Masscode™ (Polymerization–Mass Spectrometry–Solution)

The Masscode SNP allele discrimination assay[98] marketed by Qiagen (Venlo, Netherlands, http://www1.qiagen.com/) is based on allele-specific PCR and a series of Masscode tags for allele detection using MS. In principle, two rounds of PCR are performed; first a locus-specific PCR is

used to amplify the SNP-containing fragment, then an allele-specific PCR incorporates a unique Masscode tag to each allele-specific product. The mass tags are photocleavable. After purification of the amplicons, the tags are released by treating with light and injected into a mass spectrometer. Alleles are scored via analysis of mass signatures. As of 2001, 30 unique Masscode tags had been characterized, enabling multiplex analysis of up to 15 SNPs.

SNPstream® (Polymerization–Fluorescence–Microarray)

The GenomeLab™ SNPstream genotyping system of Beckman Coulter (Fullerton, California, USA, www.beckman.com/) is based on the SNP-IT™ assay[99,100] originally marketed by Orchid Cellmark (Princeton, New Jersey, http://www.orchid.com/). Essentially, the method entails single-base extension (SBE) or "minisequencing" followed by fluorescence detection on tag arrays. The assay procedure starts by multiplex PCR amplification of up to 12 different SNP loci. After purification, locus-specific primers are added and annealed adjacent to the SNP position. Each primer contains a unique tag sequence. A cyclic, allele-specific extension reaction ensues in which fluorescence-labeled ddNTPs are added to the extension primers.

The extension products are sorted via their tag sequences on a complementary tag array, followed by placement into an array scanner. Alleles are called by measuring the color and intensity of fluorescence at each position on the tag array. Currently the 4 × 4 tag arrays are printed in a 384-well format, increasing the potential throughput of SNP genotypes called per scan.

Another assay involving minisequencing on an array is the arrayed primer extension (APEX) assay[101] of Asper Biotech (Tartu, Estonia, http://www.asperbio.com/). Rather than using a tag array as described above, extension reactions are performed directly on probes immobilized to the microarray surface.

SNPlex™ (Ligation–Fluorescence–Electrophoresis)

The SNPlex assay[102] recently released by Applied Biosystems is a multiplexed ligation-based assay performed on fragmented genomic DNA. The allele discrimination procedure basically involves differential ligation of one locus-specific oligonucleotide with a pair of allele-specific oligonucleotides. Each allele-specific oligonucleotide encodes a unique tag termed a ZipCode™. Following a clean-up step, successfully ligated products are amplified by PCR and immobilized to a streptavidin-coated plate via a biotin label on one of the PCR primers. A second clean-up step serves to remove excess PCR reagents and the non-biotinylated strand of the PCR product. A set of ZipChute™ detection probes is added to the plate and allowed to hybridize.

Each ZipChute probe contains an oligonucleotide that is complementary to one of the ZipCode sequences, a fluorescent label, and a unique mobility shift tag. Non-hybridized probes are washed away. The hybridized ZipChute probes are then eluted and run on a capillary electrophoresis instrument. Since each probe migrates at a different rate, alleles are scored by measuring the sizes of the fluorescent bands during electrophoresis. The assay reagents have been optimized for 48-plex reactions.

GoldenGate™ (Polymerization/Ligation–Fluorescence–Bead Array)

In the GoldenGate™ assay developed by Illumina, allele-specific extension and ligation are combined to achieve a high performance multiplexed genotyping method.[103] Sample genomic DNA is bound to paramagnetic particles and locus- and allele-specific oligonucleotide probes are hybridized to the genomic DNA. For each SNP analyzed, there are two allele-specific probes and one locus-specific probe.

The allele-specific probes are designed to hybridize to the target so that their 3′ terminal nucleotides are complementary to either variant of the SNP. The locus-specific probe is designed to hybridize some distance downstream of the allele-specific probe. By adding a polymerase and

a ligase, only allele-specific probes that are matched at their 3′ ends are extended and can be ligated to a locus-specific probe hybridized downstream of the primer. Undesired reaction byproducts and remaining non-hybridized probes are removed by washing.

All ligation products can be amplified with a standard set of amplification primers, while each locus-specific probe is also equipped with a unique address sequence. Allele information is thus encoded in the reaction product as one of two primer sequences, while locus information is encoded as an address sequence. Following PCR amplification of the reacted probes with differentially labeled primers, allele information is represented by one of two fluorescent molecules. The amplification products are then rendered single-stranded by a solid phase extraction and hybridized to a generic oligonucleotide array, the BeadArray™ platform, containing oligonucleotides complementary to each of the address sequences used. This way, locus information is converted into position information, and reading the fluorescence of each position of the array reveals the genotype at the corresponding loci.

Padlock Probe and MIP Assays (Polymerization/Ligation–Fluorescence–Array/Electrophoresis)

The oligonucleotide ligation assay mentioned previously has been further developed in the padlock probe assay.[104] Allele-specific and locus-specific probes are designed so that they are connected at their distal ends to form a single probe that becomes circularized upon ligation. Its shape resembles a padlock locked around its target molecule. Two such allele-specific padlock probes are used for each SNP to be genotyped.

The circular nature of reacted probes allows them to be distinguished from unreacted or dimerized probes, and the padlock probe method requires no washes. Ligation is performed directly on genomic DNA and the reacted probes are amplified to generate sufficient signal for detection. Locus- and allele-specific information can be encoded in the parts of the padlock probes that are not target-complementary via tag sequences for array hybridization and for detection of allele information using differentially labeled fluorescent probes.[105] The circular probes may also be amplified using the circle-to-circle amplification technique.[106] In the SNPWave™ assay (Keygene, Wageningen, Netherlands, http://www.keygene.com/), allele and locus information is instead encoded in the length of the padlock probe and decoding is achieved through physical separation by capillary gel electrophoresis.[107]

A variation of the padlock probe assay is the molecular inversion probe (MIP) assay in which each SNP is interrogated by a single probe and allele distinction is achieved by polymerization instead of ligation.[65] The probes are designed so that the ends hybridize to the target, leaving a single nucleotide gap at the position of the SNP. In separate reactions for each nucleotide, this gap is filled with an allele-specific nucleotide by a polymerase, allowing the probe to be circularized by ligation. Next, cleavage of the probes between the PCR primer sequences results in "inversion" of the ligation product, releasing the probes from the target sequence for PCR amplification. Following amplification, the PCR products are hybridized to tag oligonucleotide arrays as described above. Molecular inversion probes have been used by ParAllele BioScience to achieve 10,000-plex SNP genotyping.[108]

GeneChip® (Hybridization–Fluorescence–Microarray)

A primary example of the use of microarrays for SNP detection is the GeneChip® probe assay from Affymetrix. Pre-selected sets of probe sequences for 10,000[109] or 100,000[110] SNPs are synthesized directly on glass slides using a photolithographic technique.[111] SNPs for inclusion on the arrays were chosen from information from the human genome sequencing project. Thus the precise locations of the SNPs and the surrounding DNA sequences are known.

The method involves a complexity reduction step to prepare patient DNA samples for hybridization to the array. Whole genomic DNA from the patient is digested with a chosen set of restriction enzymes to produce 400- to 800-bp fragments of DNA. The restriction fragments containing the SNP target loci are of optimal length for PCR. Adapter cassettes encoding a universal PCR primer sequence are ligated to both ends of each fragment. A single PCR primer sequence is then used for multiplex amplification of the set of SNP targets. The PCR products are digested with an endonuclease to produce shorter fragments and are end-labeled with biotin. Finally, the digestion mix is hybridized to the array and labeled using a fluorescence staining procedure. The arrays are placed into a scanner and sequence variants are determined from examination of the fluorescence patterns across the surface of the array.

This approach provides several advantages. From a relatively small amount of DNA (250 ng), it is possible to retrieve genotypes from up to 100,000 SNPs.[110] Also, a single PCR primer is used for amplifying all SNPs, reducing the number of assay-specific reagents needed per SNP studied. The investment per study thus resides mainly in the production and analysis of the array. This assay is most suitable for large scale genotyping. Also, custom SNPs to be added to the chip must conform to the entire strategy which may or may not be a trivial consideration.

Whole Genome Genotyping (Polymerization–Fluorescence–BeadArray™)

Illumina recently published a radically new genotyping strategy called the whole genome genotyping (WWG) assay.[112] In this approach an initial whole genome amplification reaction similar to the ϕ29 method described in the "Quantity of DNA" section below yields a >1000-fold amplification of 100- to 200-bp DNA fragments covering the entire genome. Following purification, the amplified products are hybridized to an array. The oligonucleotides on the array contain a complementary sequence to one SNP loci with a 3′ allele-specific nucleotide. DNA polymerase and a mixture of labeled nucleotides are added to perform allele-specific primer extension reactions.

After a clean-up step and staining procedure, genotypes are scored by measuring fluorescence in an array scanner. A major advantage of the WGG is that the assay performs well in the full complexity of genomic DNA. This allows great flexibility in selection of SNPs to be analyzed, while the array format maintains the ability to perform highly multiplexed reactions.

CAVEAT

Admittedly, this list of techniques is far from exhaustive. It is meant to give readers a foundation for understanding how common genotyping platforms work. Further descriptions of the presented techniques can be found in the primary literature. Expanded coverage over the breadth of SNP genotyping methodologies can be found in one of many high-quality reviews.[113–122]

Choosing a Method: A Few Clinical Considerations

A number of factors are important in selecting a method for scoring SNP genotypes. Non-trivial matters, especially in clinical settings, include approval of methods and tests by regulatory agencies and accessibility to equipment. Although technology is changing, the basics of PCR machines, electrophoresis equipment, and DNA sequencers form the core activities of many clinical laboratories. It is no surprise that methods such as restriction fragment length polymorphism (RFLP) and sequencing account for a large proportion of techniques used to genotype SNPs in patients today.

The cost of importing a new genotyping platform into an existing lab involves several different sources. The initial investment for instrumentation, potential robotic systems, software, and maintenance is one source. A second is the running cost of assays including reagents, plastics, enzymes, oligonucleotides, and other consumables involved in performing individual genotyping reactions. The third and often underestimated cost is for labor. Beyond the hours involved in performing the

Tecan Ex. 2130, Page 20 of 31
QIAGEN Sciences, LLC v. Tecan Genomics - IPR2025-00028

assay, the expense involved in training staff to design, perform, and interpret assay data as well as repeat inconclusive assays is often omitted from expense calculations. Meaningful comparisons of cost are thus extremely difficult; few publications report comprehensive analyses of the expenses involved.

Aside from these practical matters, the choice of SNP genotyping technology is highly application-dependent. For example, characteristics such as sensitivity and specificity may be prioritized over cost and throughput in diagnostic tests. For other applications such as whole genome association studies, somewhat lower accuracy rates may be tolerated but throughput must be high to finish a project within a reasonable time frame and cost per genotype must be kept to a minimum. Several other criteria worth considering are noted below.

Quantitative Data

For some applications determining the presence or absence of alleles is insufficient. The number of copies of an SNP locus is the clinically relevant fact. For example, gene copy number changes are important elements of malignancy in diseases such as breast cancer[123] and they can influence susceptibility to infectious diseases such as HIV/AIDS.[124] In such cases, both genotypic variation and gene copy number can impact the risk or severity of disease. Another clinically relevant example of the importance of quantification is in monitoring the success of bone marrow transplantation. This can be done by analyzing polymorphisms in different fractions of patient blood.

Genetic markers in which the donor and recipient differ in genotype are chosen and blood samples are taken repeatedly after transplantation. The stable presence of both donor and recipient genotypes (*chimerism*) is taken as a sign of successful transplantation. Reversion of genotype toward that of the recipient is an indication of graft rejection. Traditionally, various repeat polymorphisms have been used for this type of analysis (reviewed in Reference 125), but recent publications discuss the use of sets of SNPs.[126] Examples of techniques that are better suited for quantitative measures are pyrosequencing and the Sequenom MassARRAY™ system.

Quantity of DNA

All SNP genotyping methods are highly affected by the quality and quantity of DNA used in the analysis. Although genetic material can be extracted from a variety of sources including biopsies and tissue cultures, one very common source is patient blood. From 1 ml of blood, roughly 25 μg of DNA can be retrieved. For many SNP genotyping platforms, a minimum of 1 to 10 ng of genomic DNA (corresponding to 150 to 1500 copies of any single copy gene) is required. Considering possible loss of DNA during pipetting, storage, dilution, and assay optimization, DNA extracted from 1 ml of blood would thus be sufficient for roughly 2000 genotyping assays on typical genotyping platforms. This is more than enough for genotyping a small number of loci; however, it may prove insufficient for large scale projects like whole genome association studies.

For some purposes, generation of immortalized cell lines from patients can be a suitable means of ensuring access to biological material. A number of *in vitro* approaches can extend the amount of genomic DNA starting from a minimal amount of genetic material. The first is the whole genome amplification (WGA) method. Although several different strategies have been proposed (reviewed in Reference 127), one method that has been successfully implemented into SNP genotyping is multiple displacement amplification (MDA).[128,129] In this isothermal technique, a series of random hexamers (six-base oligonucleotides) are mixed with genomic DNA along with ϕ29 DNA polymerase. When the hexamers anneal to the genome, they serve as primers for replication by the polymerase.

Primer extension progresses along the target strand, displacing any DNA molecules encountered along the way. Displaced strands then serve as templates for hexamer annealing and extension, and the process of DNA duplication continues. Starting from as few as 1 to 10 copies of genomic DNA,

up to 20 to 30 μg of amplified product may be produced.[129] The fidelity of genotyping results from MDA-amplified samples has been demonstrated in a number of publications.[130–132]

Instead of amplifying the entire genome, directed amplification of selected portions of the genome may be more appropriate for some applications. PCR is commonly used for amplification of a single target sequence or a few sequences in parallel. However, for high levels of parallelism, a large number of separate amplification reactions must be performed.

As briefly noted previously, the so-called adaptor ligation method uses restriction endonucleases to digest a genomic sample into fragments of a known distribution. Common adaptor oligonucleotides are then ligated to the ends of the generated fragments. A subsequent PCR amplification with a single pair of primers can be adjusted to preferentially amplify fragments within a certain range of sizes. This way, a subset of the genome is amplified and a reduction of sequence complexity is achieved.[109,133] A variant of the method uses adaptors, one common primer, and a single specific primer per target sequence to achieve amplification of specific genomic sequences.[134,135]

If restriction digestion is carried out with type II S restriction enzymes, different end sequences are produced for different fragments. This way, adaptor sequences can be selected to circularize a certain population of fragments. These fragments can then be amplified and the procedure repeated, allowing the selective amplification of a single fragment from a genomic DNA sample.[136] This method may also be multiplexed by using several adaptor sequences, but incompatibilities of end sequences limit the extent to which this can be performed.

None of the methods above achieves the ideal goal of amplifying a large set of arbitrary sequences without the production of undesired amplification products. In the recently described selector amplification method, restriction fragments are made single-stranded by heat denaturation, circularized by hybridization to specific selector probes, and ligated to a common oligonucleotide, thus incorporating a common sequence motif in the single-stranded circles formed from the selected fragments. The selector probe end sequences are 10 to 20 nt in length to achieve specificity. Furthermore, structure-specific cleavage, using the same mechanism as in the Invader assay, can be utilized to remove a 5′ portion of the fragment prior to ligation, allowing the creation of circles of similar lengths. All circles can then be amplified by PCR using a single primer pair for multiplex amplification for hundreds to thousands of selected fragments of interest.[137]

Known versus Unknown Variations

The vast majority of techniques described here are applied to SNPs where sequence and mapping information are known. Other methods are more adept at looking for sequence variants where exact information is unavailable. For example, if a particular gene has been implicated in a given disease, scanning the exons and promoters for variation could be desirable from a diagnostic point of view.

A number of methods can speed the search for such unknown sequence variants. Single-strand conformation polymorphism (SSCP),[138] denaturing gradient gel electrophoresis (DGGE),[139] and denaturing high performance liquid chromatography (dHPLC)[140] are examples of such methods. Although these three techniques can identify DNA fragments bearing sequence differences, the position or nature of genetic change is difficult to characterize based on results of these methods alone. Therefore, these methods are combined with or in fact replaced by sequencing techniques to derive exact sequence information.

Speed to Results

Another aspect to consider when selecting a genotyping platform is amount of time required to obtain the results. For point-of-care applications, the homogeneous PCR-based methods are quickest for typing a limited number of genetic variants in one or a few patients. In addition to the

homogenous TaqMan and molecular beacon assays mentioned previously, methods based on allele-specific PCR have been developed for SNP genotyping.

The principles of the amplification refractory mutation system[70] have been adapted for fluorescence detection using intercalating dyes as in the Tm shifting assay[141] or through strategies involving FRET reactions.[142–146] These methods require both thermocycling and fluorescent detection, and many commercially available devices combine these two functions. The LightCycler® marketed by Roche Diagnostics (division of F. Hoffmann–La Roche Ltd, Basel, Switzerland, http://www.roche-diagnostics.com/) and the SmartCycler® by Cepheid (Sunnyvale, California, USA, http://www.cepheid.com/) perform PCR with real time fluorescence detection in as few as 20 minutes.

POSSIBILITIES AND PERSPECTIVES

Poised with a broad range of tools and technologies, biomedical research is entering a phase of exploring the information embodied within the human genome. Starting from extensive knowledge about the basic architecture and sequence of our genetic blueprint, major efforts are underway to determine the impact of genetic variation on risk for developing disease or responding to medical therapy.[147] For the field of pharmacogenomics, analysis of SNPs will probably not become the sole basis for personalized medicine, but rather an important component to the sum of molecular measurements that together reveal an individual's medical condition. A physician's evaluation will involve a totality of molecular information together with other information to determine the cause and identity of the disease and the best course of treatment.

The continued advancement of SNP genotyping technologies opens new and exciting possibilities in molecular medicine. One such opportunity is investigation of polymorphic differences directly within cells or tissue samples. SNP genotyping *in situ* can be important if the material to be analyzed is characterized by significant cell-to-cell variation as in tumor-derived samples. When such heterogeneity is encountered, conventional analysis in solution may fail to reveal heterogeneity. Using solution-phase methods for detection of the distribution of cells with specific sequence variants cannot be evaluated unless samples are micro-dissected.

Average measurements over many cells also fail to show the presence of distinct populations of cells that differ in target content. To access this kind of information, single cells must be defined and analyzed. The most widespread technique for analyzing the distribution of DNA and RNA sequences is *in situ* hybridization (ISH), which can be combined with fluorescent detection and is then known as fluorescence *in situ* hybridization (FISH). Until now, no reliable method has been available for *in situ* analysis of single copy molecules for SNP discrimination. So-called padlock probes have been applied recently for *in situ* analysis of mitochondrial DNA,[148,149] offering new possibilities for diagnostics.

As described earlier, padlock probes are linear oligonucleotides whose ends, upon hybridization next to each other on a target strand, can be joined by ligation. This creates a circularized DNA strand, wound around a target.[104] Circularized probes can then be used for template synthesis of extensive rolling-circle replication products anchored at targets and easily detected using fluorescent probes hybridized to the repeated sequence.

The robust target-primed rolling circle amplification (RCA) approach has been used on mitochondrial DNA targets in both cultured cells (Figure 6.7) and fresh frozen tissue sections.[148] Target-primed RCA should also be applicable for studies of gene expression by enabling *in situ* studies of allele-specific transcripts and the expression of highly similar members of gene families and splice variants. Another possible application is localized detection of infectious microbial agents. The procedure is suitable for use in pathological examinations by providing short analysis time without a need for expensive laboratory equipment other than a fluorescence microscope.

Adaptation of SNP genotyping platforms is also used to analyze other molecular genomic variables such as gene copy number[150] and allele-specific expression of transcripts[92] due, for example, to promoter polymorphisms. Epigenetic factors modulating gene expression such as

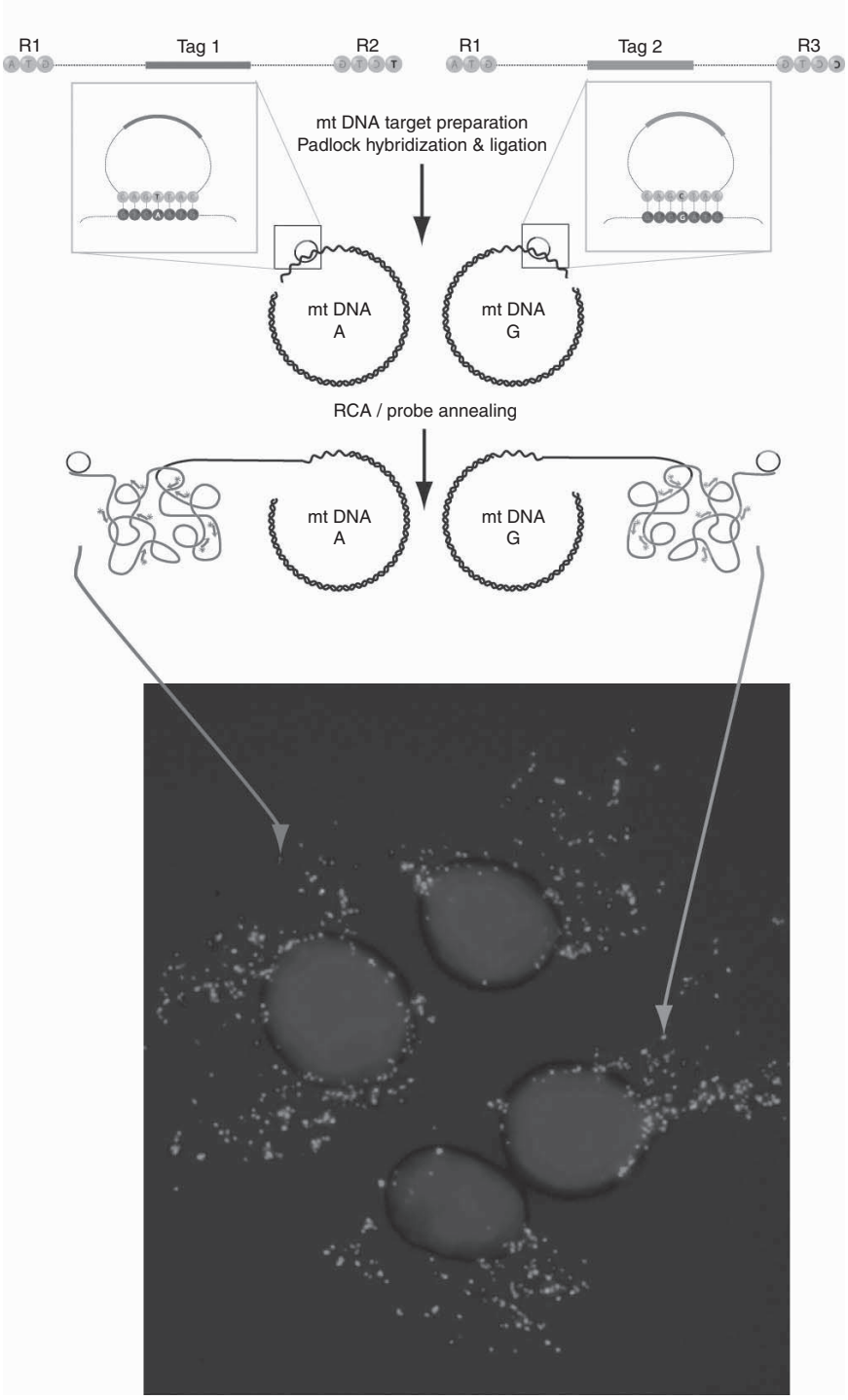


FIGURE 6.7 ***(A color version follows page 270)*** *In situ* genotyping using padlock probes. R1 are the locus-specific recognition sequences of the padlock probe; R2 and R3 are the allele-specific recognition sequences. The drawing (top) depicts the basic steps of the procedure, and the picture (bottom) presents an example of genotyping of a heteroplasmic cell line called G55 1.1. In these cells, roughly 65% of mitochondrial genomes contain the A3243G (MELAS) mutation.

Tecan Ex. 2130, Page 24 of 31
QIAGEN Sciences, LLC v. Tecan Genomics - IPR2025-00028

methylation and histone hyperacetylation[151] have been studied by technologies similar to those used for SNP genotyping.

Infectious disease diagnostics are also of potentially great value; however, this type of application introduces new problems. Many pathogenic organisms mutate rapidly as part of their biological strategies to maintain virulence. Methods for SNP genotyping based on recognition of specific genetic sequences encounter difficulties when applied to viruses and bacteria that evolve rapidly. Assay specificity and selectivity must be balanced with the ability to detect slight variations in sequence in order to successfully detect or distinguish variants of the organism if present.

One further technological development that could drastically affect the way we approach analysis of the genome is the emergence of powerful and economical sequencing techniques. Rather than looking at single base positions, the strategy would involve decoding the entire genome or extensive selected regions thereof. Complete information of the genetic sequence of an individual would thus become available, obviating the need for genotyping specific single base positions. Research activities aimed at achieving "$1000" genome sequencing[152] are under way, making cataloguing entire genome sequence information in populations a foreseeable possibility. In the meantime, many of the technologies described here will certainly assist in harvesting the fruits of the human genome initiative and play key roles in biomedical research for years to come.

REFERENCES

1. Brookes AJ. The essence of SNPs. *Gene* 1999; 234 (2): 177–186.
2. Zondervan KT, Cardon LR. The complex interplay among factors that influence allelic association. *Nat Rev Genet* 2004; 5 (2): 89–100.
3. Holliday R, Grigg GW. DNA methylation and mutation. *Mutat Res* 1993; 285 (1): 61–67.
4. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* 2001; 409 (6822): 860–921.
5. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA et al. Sequence of the human genome. *Science* 2001; 291 (5507): 1304–1351.
6. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* 2004; 431 (7011): 931–945.
7. International SNP Map Working Group. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 2001; 409 (6822): 928–933.
8. Taillon-Miller P, Gu Z, Li Q, Hillier L, Kwok PY. Overlapping genomic sequences: a treasure trove of single-nucleotide polymorphisms. *Genome Res* 1998; 8 (7): 748–754.
9. Altshuler D, Pollara VJ, Cowles CR, Van Etten WJ, Baldwin J, Linton L, Lander ES. An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature* 2000; 407 (6803): 513–516.
10. Picoult-Newberg L, Ideker TE, Pohl MG, Taylor SL, Donaldson MA, Nickerson DA, Boyce-Jacino M. Mining SNPs from EST databases. *Genome Res* 1999; 9 (2): 167–174.
11. Irizarry K, Kustanovich V, Li C, Brown N, Nelson S, Wong W, Lee CJ. Genome-wide analysis of single-nucleotide polymorphisms in human expressed sequences. *Nat Genet* 2000; 26 (2): 233–236.
12. Garg K, Green P, Nickerson DA. Identification of candidate coding region single nucleotide polymorphisms in 165 human genes using assembled expressed sequence tags. *Genome Res* 1999; 9 (11): 1087–1092.
13. Marth GT, Korf I, Yandell MD, Yeh RT, Gu Z, Zakeri H, Stitziel NO, Hillier L, Kwok PY, Gish WR. A general approach to single-nucleotide polymorphism discovery. *Nat Genet* 1999; 23 (4): 452–456.
14. Buetow KH, Edmonson MN, Cassidy AB. Reliable identification of large numbers of candidate SNPs from public EST data. *Nat Genet* 1999; 21 (3): 323–325.
15. Schlotterer C. The evolution of molecular markers: just a matter of fashion? *Nat Rev Genet* 2004; 5 (1): 63–69.
16. Fredman D, Siegfried M, Yuan YP, Bork P, Lehvaslaiho H, Brookes AJ. HGVbase: a human sequence variation database emphasizing data quality and a broad spectrum of data sources. *Nucleic Acids Res* 2002; 30 (1): 387–391.

17. Fredman D, Munns G, Rios D, Sjoholm F, Siegfried M, Lenhard B, Lehvaslaiho H, Brookes AJ. HGVbase: a curated resource describing human DNA variation and phenotype relationships. *Nucleic Acids Res* 2004; 32: D516–D519.
18. Kruglyak L, Nickerson DA. Variation is the spice of life. *Nat Genet* 2001; 27 (3): 234–236.
19. Nickerson DA, Taylor SL, Weiss KM, Clark AG, Hutchinson RG, Stengard J, Salomaa V, Vartiainen E, Boerwinkle E, Sing CF. DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. *Nat Genet* 1998; 19 (3): 233–240.
20. Li WH, Sadler LA. Low nucleotide diversity in man. *Genetics* 1991; 129 (2): 513–523.
21. Halushka MK, Fan JB, Bentley K, Hsie L, Shen N, Weder A, Cooper R, Lipshutz R, Chakravarti A. Patterns of single-nucleotide polymorphisms in candidate genes for blood pressure homeostasis. *Nat Genet* 1999; 22 (3): 239–247.
22. Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, Shaw N, Lane CR, Lim EP, Kalyanaraman N et al. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat Genet* 1999; 22 (3): 231–238.
23. Livingston RJ, von Niederhausern A, Jegga AG, Crawford DC, Carlson CS, Rieder MJ, Gowrisankar S, Aronow BJ, Weiss RB, Nickerson DA. Pattern of sequence variation across 213 environmental response genes. *Genome Res* 2004; 14 (10A): 1821–1831.
24. Costas J, Salas A, Phillips C, Carracedo A. Human genome-wide screen of haplotype-like blocks of reduced diversity. *Gene* 2005; 349: 219–225.
25. Horton R, Niblett D, Milne S, Palmer S, Tubby B, Trowsdale J, Beck S. Large-scale sequence comparisons reveal unusually high levels of variation in the HLA-DQB1 locus in the class II region of the human MHC. *J Mol Biol* 1998; 282 (1): 71–97.
26. Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R et al. Linkage disequilibrium in the human genome. *Nature* 2001; 411 (6834): 199–204.
27. Clark AG. Inference of haplotypes from PCR-amplified samples of diploid populations. *Mol Biol Evol* 1990; 7 (2): 111–122.
28. Excoffier L, Slatkin M. Maximum likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol Biol Evol* 1995; 12 (5): 921–927.
29. Hawley ME, Kidd KK. HAPLO: a program using the EM algorithm to estimate the frequencies of multi-site haplotypes. *J Hered* 1995; 86 (5): 409–411.
30. Long JC, Williams RC, Urbanek M. An EM algorithm and testing strategy for multiple-locus haplotypes. *Am J Hum Genet* 1995; 56 (3): 799–810.
31. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 2001; 68 (4): 978–989.
32. Stephens M, Donnelly P. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet* 2003; 73 (5): 1162–1169.
33. Weale ME. A survey of current software for haplotype phase inference. *Hum Genomics* 2004; 1 (2): 141–144.
34. Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E, Ballinger DG, Frazer KA, Cox DR. Whole-genome patterns of common DNA variation in three human populations. *Science* 2005; 307 (5712): 1072–1079.
35. Crawford DC, Nickerson DA. Definition and clinical importance of haplotypes. *Annu Rev Med* 2005; 56: 303–320.
36. Winckler W, Myers SR, Richter DJ, Onofrio RC, McDonald GJ, Bontrop RE, McVean GA, Gabriel SB, Reich D, Donnelly P et al. Comparison of fine-scale recombination rates in humans and chimpanzees. *Science* 2005; 308 (5718): 107–111.
37. Zhang K, Akey JM, Wang N, Xiong M, Chakraborty R, Jin L. Randomly distributed crossovers may generate block-like patterns of linkage disequilibrium: an act of genetic drift. *Hum Genet* 2003; 113 (1): 51–59.
38. Crawford DC, Bhangale T, Li N, Hellenthal G, Rieder MJ, Nickerson DA, Stephens M. Evidence for substantial fine-scale variation in recombination rates across the human genome. *Nat Genet* 2004; 36 (7): 700–706.
39. Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES. High-resolution haplotype structure in the human genome. *Nat Genet* 2001; 29 (2): 229–232.

40. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M et al. The structure of haplotype blocks in the human genome. *Science* 2002; 296 (5576): 2225–2229.
41. Phillips MS, Lawrence R, Sachidanandam R, Morris AP, Balding DJ, Donaldson MA, Studebaker JF, Ankener WM, Alfisi SV, Kuo FS et al. Chromosome-wide distribution of haplotype blocks and the role of recombination hot spots. *Nat Genet* 2003; 33 (3): 382–387.
42. Wall JD, Pritchard JK. Assessing the performance of the haplotype block model of linkage disequilibrium. *Am J Hum Genet* 2003; 73 (3): 502–515.
43. Patil N, Berno AJ, Hinds DA, Barrett WA, Doshi JM, Hacker CR, Kautzer CR, Lee DH, Marjoribanks C, McDonough DP et al. Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science* 2001; 294 (5547): 1719–1723.
44. International HapMap Consortium. The International HapMap Project. *Nature* 2003; 426 (6968): 789–796.
45. Newton–Cheh C, Hirschhorn JN. Genetic association studies of complex traits: design and analysis issues. *Mutat Res* 2005; 573 (1–2): 54–69.
46. Cardon LR, Bell JI. Association study designs for complex diseases. *Nat Rev Genet* 2001; 2 (2): 91–99.
47. Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for Mendelian disease, future approaches for complex disease. *Nat Genet* 2003; 33 Suppl: 228–237.
48. Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet* 2003; 33 (2): 177–182.
49. Tabor HK, Risch NJ, Myers RM. Opinion: candidate-gene approaches for studying complex genetic traits: practical considerations. *Nat Rev Genet* 2002; 3 (5): 391–397.
50. Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K. A comprehensive review of genetic association studies. *Genet Med* 2002; 4 (2): 45–61.
51. Reich DE, Lander ES. On the allelic spectrum of human disease. *Trends Genet* 2001; 17 (9): 502–510.
52. Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science* 1996; 273 (5281): 1516–1517.
53. Carlson CS, Eberle MA, Kruglyak L, Nickerson DA. Mapping complex disease loci in whole-genome association studies. *Nature* 2004; 429 (6990): 446–452.
54. Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, Ueda H, Cordell HJ, Eaves IA, Dudbridge F et al. Haplotype tagging for the identification of common disease genes. *Nat Genet* 2001; 29 (2): 233–237.
55. Kwok PY. Genomics: genetic association by whole-genome analysis? *Science* 2001; 294 (5547): 1669–1670.
56. Altshuler D, Daly M, Kruglyak L. Guilt by association. *Nat Genet* 2000; 26 (2): 135–137.
57. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* 2005; 6 (2): 95–108.
58. Wang WY, Barratt BJ, Clayton DG, Todd JA. Genome-wide association studies: theoretical and practical concerns. *Nat Rev Genet* 2005; 6 (2): 109–118.
59. Kutyavin IV, Afonina IA, Mills A, Gorn VV, Lukhtanov EA, Belousov ES, Singer MJ, Walburger DK, Lokhov SG, Gall AA et al. 3′ minor groove binder DNA probes increase sequence specificity at PCR extension temperatures. *Nucleic Acids Res* 2000; 28 (2): 655–661.
60. Vester B, Wengel J. LNA (locked nucleic acid): high-affinity targeting of complementary RNA and DNA. *Biochemistry* 2004; 43 (42): 13233–13241.
61. Kan YW, Dozy AM. Polymorphism of DNA sequence adjacent to human -globin structural gene: relationship to sickle mutation. *Proc Natl Acad Sci USA* 1978; 75 (11): 5631–5635.
62. Landegren U, Kaiser R, Sanders J, Hood L. A ligase-mediated gene detection technique. *Science* 1988; 241 (4869): 1077–1080.
63. Syvanen AC, Aalto-Setala K, Harju L, Kontula K, Soderlund H. A primer-guided nucleotide incorporation assay in the genotyping of apolipoprotein E. *Genomics* 1990; 8 (4): 684–692.
64. Nyren P., Pettersson B., Uhlen M. Solid Phase DNA Minisequencing by an enzymatic luminometric inorganic pyrophosphate detection assay. *Anal Biochem* 1993; 208 (1): 171–175.

65. Hardenbol P, Baner J, Jain M, Nilsson M, Namsaraev EA, Karlin-Neumann GA, Fakhrai-Rad H, Ronaghi M, Willis TD, Landegren U et al. Multiplexed genotyping with sequence-tagged molecular inversion probes. *Nat Biotechnol* 2003; 21 (6): 673–678.
66. Chen X, Zehnbauer B, Gnirke A, Kwok PY. Fluorescence energy transfer detection as a homogeneous DNA diagnostic method. *Proc Natl Acad Sci USA* 1997; 94 (20): 10756–10761.
67. Fan JB, Chen X, Halushka MK, Berno A, Huang X, Ryder T, Lipshutz RJ, Lockhart DJ, Chakravarti A. Parallel genotyping of human SNPs using generic high-density oligonucleotide tag arrays. *Genome Res* 2000; 10 (6): 853–860.
68. Rodi CP, Darnhofer-Patel B, Stanssens P, Zabeau M, van den Boom D. A strategy for the rapid discovery of disease markers using the MassARRAY™ system. *Biotechniques* 2002; 32 (Suppl 2): S62–S69.
69. Blondal T, Waage BG, Smarason SV, Jonsson F, Fjalldal SB, Stefansson K, Gulcher J, Smith AV. A novel MALDI-TOF-based methodology for genotyping single nucleotide polymorphisms. *Nucleic Acids Res* 2003; 31 (24): e155.
70. Newton CR, Heptinstall LE, Summers C, Super M, Schwarz M, Anwar R, Graham A, Smith JC, Markham AF. Amplification refractory mutation system for prenatal diagnosis and carrier assessment in cystic fibrosis. *Lancet* 1989; 2 (8678–8679): 1481–1483.
71. Pastinen T, Raitio M, Lindroos K, Tainola P, Peltonen L, Syvanen AC. A system for specific, high-throughput genotyping by allele-specific primer extension on microarrays. *Genome Res* 2000; 10 (7): 1031–1042.
72. Cahill P, Bakis M, Hurley J, Kamath V, Nielsen W, Weymouth D, Dupuis J, Doucette–Stamm L, Smith DR. Exo-proofreading, a versatile SNP scoring technology. *Genome Res* 2003; 13 (5): 925–931.
73. Lyamichev V, Brow MA, Dahlberg JE. Structure-specific endonucleolytic cleavage of nucleic acids by eubacterial DNA polymerases. *Science* 1993; 260 (5109): 778–783.
74. Tost J, Gut IG. Genotyping single nucleotide polymorphisms by MALDI mass spectrometry in clinical applications. *Clin Biochem* 2005; 38 (4): 335–350.
75. Fei Z, Ono T, Smith LM. MALDI-TOF mass spectrometric typing of single nucleotide polymorphisms with mass-tagged ddNTPs. *Nucleic Acids Res* 1998; 26 (11): 2827–2828.
76. Howell WM, Jobs M, Gyllensten U, Brookes AJ. Dynamic allele-specific hybridization: a new method for scoring single nucleotide polymorphisms. *Nat Biotechnol* 1999; 17 (1): 87–88.
77. De Angelis DA. Why FRET over genomics? *Physiol Genomics* 1999; 1 (2): 93–99.
78. Howell WM, Jobs M, Brookes AJ. iFRET: an improved fluorescence system for DNA-melting analysis. *Genome Res* 2002; 12 (9): 1401–1407.
79. Kwok PY. SNP genotyping with fluorescence polarization detection. *Hum Mutat* 2002; 19 (4): 315–323.
80. Livak KJ, Flood SJ, Marmaro J, Giusti W, Deetz K. Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. *PCR Methods Appl* 1995; 4 (6): 357–362.
81. Livak KJ. Allelic discrimination using fluorogenic probes and the 5′ nuclease assay. *Genet Anal* 1999; 14 (5–6): 143–149.
82. Lee LG, Livak KJ, Mullah B, Graham RJ, Vinayak RS, Woudenberg TM. Seven-color, homogeneous detection of six PCR products. *Biotechniques* 1999; 27 (2): 342–349.
83. Tyagi S, Kramer FR. Molecular beacons: probes that fluoresce upon hybridization. *Nat Biotechnol* 1996; 14 (3): 303–308.
84. Tyagi S, Bratu DP, Kramer FR. Multicolor molecular beacons for allele discrimination. *Nat Biotechnol* 1998; 16 (1): 49–53.
85. Ronaghi M, Karamohamed S, Pettersson B, Uhlen M, Nyren P. Real-time DNA sequencing using detection of pyrophosphate release. *Anal Biochem* 1996; 242 (1): 84–89.
86. Langaee T, Ronaghi M. Genetic variation analyses by pyrosequencing. *Mutat Res* 2005; 573 (1–2): 96–102.
87. Prince JA, Feuk L, Howell WM, Jobs M, Emahazion T, Blennow K, Brookes AJ. Robust and accurate single nucleotide polymorphism genotyping by dynamic allele-specific hybridization (DASH): design criteria and assay validation. *Genome Res* 2001; 11 (1): 152–162.

88. Jobs M, Howell WM, Stromqvist L, Mayr T, Brookes AJ. DASH-2: flexible, low-cost, and high-throughput SNP genotyping by dynamic allele-specific hybridization on membrane arrays. *Genome Res* 2003; 13 (5): 916–924.
89. Lyamichev V, Mast AL, Hall JG, Prudent JR, Kaiser MW, Takova T, Kwiatkowski RW, Sander TJ, de Arruda M, Arco DA et al. Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes. *Nat Biotechnol* 1999; 17 (3): 292–296.
90. Olivier M. The invader assay for SNP genotyping. *Mutat Res* 2005; 573 (1–2): 103–110.
91. Olivier M, Chuang LM, Chang MS, Chen YT, Pei D, Ranade K, de Witte A, Allen J, Tran N, Curb D et al. High-throughput genotyping of single nucleotide polymorphisms using new biplex invader technology. *Nucleic Acids Res* 2002; 30 (12): e53.
92. Jurinke C, Denissenko MF, Oeth P, Ehrich M, van den Boom D, Cantor CR. A single nucleotide polymorphism based approach for the identification and characterization of gene expression modulation using MassARRAY. *Mutat Res* 2005; 573 (1–2): 83–95.
93. Braun A, Little DP, Koster H. Detecting *CFTR* gene mutations by using primer oligo base extension and mass spectrometry. *Clin Chem* 1997; 43 (7): 1151–1158.
94. Sauer S, Lechner D, Berlin K, Lehrach H, Escary JL, Fox N, Gut IG. A novel procedure for efficient genotyping of single nucleotide polymorphisms. *Nucleic Acids Res* 2000; 28 (5): e13.
95. Sauer S, Lechner D, Berlin K, Plancon C, Heuermann A, Lehrach H, Gut IG. Full flexibility genotyping of single nucleotide polymorphisms by the GOOD assay. *Nucleic Acids Res* 2000; 28 (23): e100.
96. 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 Res* 2002; 30 (5): e22.
97. Tost J, Gut IG. Genotyping single nucleotide polymorphisms by mass spectrometry. *Mass Spectrom Rev* 2002; 21 (6): 388–418.
98. Kokoris M, Dix K, Moynihan K, Mathis J, Erwin B, Grass P, Hines B, Duesterhoeft A. High-throughput SNP genotyping with the MASScode system. *Mol Diagn* 2000; 5 (4): 329–340.
99. Nikiforov TT, Rendle RB, Goelet P, Rogers YH, Kotewicz ML, Anderson S, Trainor GL, Knapp MR. Genetic bit analysis: a solid phase method for typing single nucleotide polymorphisms. *Nucleic Acids Res* 1994; 22 (20): 4167–4175.
100. Bell PA, Chaturvedi S, Gelfand CA, Huang CY, Kochersperger M, Kopla R, Modica F, Pohl M, Varde S, Zhao R et al. SNPstream® UHT: ultra-high throughput SNP genotyping for pharmacogenomics and drug discovery. *Biotechniques* 2002; 32 (Suppl 2): S70–S77.
101. Shumaker JM, Metspalu A, Caskey CT. Mutation detection by solid phase primer extension. *Hum Mutat* 1996; 7 (4): 346–354.
102. de la Vega FM, Lazaruk KD, Rhodes MD, Wenz MH. Assessment of two flexible and compatible SNP genotyping platforms: TaqMan® SNP genotyping assays and the SNPlex genotyping system. *Mutat Res* 2005; 573 (1–2): 111–135.
103. Oliphant A, Barker DL, Stuelpnagel JR, Chee MS. BeadArray™ technology: enabling an accurate, cost-effective approach to high-throughput genotyping. *Biotechniques* 2002; 32 (Suppl 2): S56–S61.
104. Nilsson M, Malmgren H, Samiotaki M, Kwiatkowski M, Chowdhary BP, Landegren U. Padlock probes: circularizing oligonucleotides for localized DNA detection. *Science* 1994; 265 (5181): 2085–2088.
105. Banér J, Isaksson A, Waldenstrom E, Jarvius J, Landegren U, Nilsson M. Parallel gene analysis with allele-specific padlock probes and tag microarrays. *Nucleic Acids Res* 2003; 31 (17): e103.
106. Dahl F, Baner J, Gullberg M, Mendel-Hartvig M, Landegren U, Nilsson M. Circle-to-circle amplification for precise and sensitive DNA analysis. *Proc Natl Acad Sci USA* 2004; 101 (13): 4548–4553.
107. van Eijk MJ, Broekhof JL, van der Poel HJ, Hogers RC, Schneiders H, Kamerbeek J, Verstege E, van Aart JW, Geerlings H, Buntjer JB et al. SNPWave™: a flexible multiplexed SNP genotyping technology. *Nucleic Acids Res* 2004; 32 (4): e47.
108. Hardenbol P, Yu F, Belmont J, Mackenzie J, Bruckner C, Brundage T, Boudreau A, Chow S, Eberle J, Erbilgin A et al. Highly multiplexed molecular inversion probe genotyping: over 10,000 targeted SNPs genotyped in a single tube assay. *Genome Res* 2005; 15 (2): 269–275.
109. Kennedy GC, Matsuzaki H, Dong S, Liu WM, Huang J, Liu G, Su X, Cao M, Chen W, Zhang J et al. Large-scale genotyping of complex DNA. *Nat Biotechnol* 2003; 21 (10): 1233–1237.

110. Matsuzaki H, Dong S, Loi H, Di X, Liu G, Hubbell E, Law J, Berntsen T, Chadha M, Hui H et al. Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays. *Nat Methods* 2004; 1 (2): 109–111.
111. Fodor SP, Read JL, Pirrung MC, Stryer L, Lu AT, Solas D. Light-directed, spatially addressable parallel chemical synthesis. *Science* 1991; 251 (4995): 767–773.
112. Gunderson KL, Steemers FJ, Lee G, Mendoza LG, Chee MS. A genome-wide scalable SNP genotyping assay using microarray technology. *Nat Genet* 2005; 37 (5): 549–554.
113. Nakatani K. Chemistry challenges in SNP typing. *ChemBioChem* 2004; 5 (12): 1623–1633.
114. Chen X, Sullivan PF. Single nucleotide polymorphism genotyping: biochemistry, protocol, cost and throughput. *Pharmacogenomics J* 2003; 3 (2): 77–96.
115. Kwok PY. High-throughput genotyping assay approaches. *Pharmacogenomics* 2000; 1 (1): 95–100.
116. Landegren U, Nilsson M, Kwok PY. Reading bits of genetic information: methods for single-nucleotide polymorphism analysis. *Genome Res* 1998; 8 (8): 769–776.
117. Tsuchihashi Z, Dracopoli NC. Progress in high throughput SNP genotyping methods. *Pharmacogenomics J* 2002; 2 (2): 103–110.
118. Brennan MD. High throughput genotyping technologies for pharmacogenomics. *Am J Pharmacogenomics* 2001; 1 (4): 295–302.
119. Gut IG. Automation in genotyping of single nucleotide polymorphisms. *Hum Mutat* 2001; 17 (6): 475–492.
120. Syvanen AC. Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nat Rev Genet* 2001; 2 (12): 930–942.
121. Shi MM. Technologies for individual genotyping: detection of genetic polymorphisms in drug targets and disease genes. *Am J Pharmacogenomics* 2002; 2 (3): 197–205.
122. Drabek J. A commented dictionary of techniques for genotyping. *Electrophoresis* 2001; 22 (6): 1024–1045.
123. Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* 1987; 235 (4785): 177–182.
124. Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, Catano G, Nibbs RJ, Freedman BI, Quinones MP, Bamshad MJ et al. The influence of *CCL3L1* gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* 2005; 307 (5714): 1434–1440.
125. Leclair B, Fregeau CJ, Aye MT, Fourney RM. DNA typing for bone marrow engraftment follow-up after allogeneic transplant: a comparative study of current technologies. *Bone Marrow Transplant* 1995; 16 (1): 43–55.
126. Oliver DH, Thompson RE, Griffin CA, Eshleman JR. Use of single nucleotide polymorphisms (SNP) and real-time polymerase chain reaction for bone marrow engraftment analysis. *J Mol Diagn* 2000; 2 (4): 202–208.
127. Lasken RS, Egholm M. Whole genome amplification: abundant supplies of DNA from precious samples or clinical specimens. *Trends Biotechnol* 2003; 21 (12): 531–535.
128. Lizardi PM, Huang X, Zhu Z, Bray-Ward P, Thomas DC, Ward DC. Mutation detection and single-molecule counting using isothermal rolling-circle amplification. *Nat Genet* 1998; 19 (3): 225–232.
129. Dean FB, Hosono S, Fang L, Wu X, Faruqi AF, Bray-Ward P, Sun Z, Zong Q, Du Y, Du J et al. Comprehensive human genome amplification using multiple displacement amplification. *Proc Natl Acad Sci USA* 2002; 99 (8): 5261–5266.
130. Lovmar L, Fredriksson M, Liljedahl U, Sigurdsson S, Syvanen AC. Quantitative evaluation by minisequencing and microarrays reveals accurate multiplexed SNP genotyping of whole genome amplified DNA. *Nucleic Acids Res* 2003; 31 (21): e129.
131. Paez JG, Lin M, Beroukhim R, Lee JC, Zhao X, Richter DJ, Gabriel S, Herman P, Sasaki H, Altshuler D et al. Genome coverage and sequence fidelity of 29 polymerase-based multiple strand displacement whole genome amplification. *Nucleic Acids Res* 2004; 32 (9): e71.
132. Pask R, Rance HE, Barratt BJ, Nutland S, Smyth DJ, Sebastian M, Twells RC, Smith A, Lam AC, Smink LJ et al. Investigating the utility of combining 29 whole genome amplification and highly multiplexed single nucleotide polymorphism BeadArray™ genotyping. *BMC Biotechnol* 2004; 4 (1): e15.

133. Matsuzaki H, Loi H, Dong S, Tsai YY, Fang J, Law J, Di X, Liu WM, Yang G, Liu G et al. Parallel genotyping of over 10,000 SNPs using a one-primer assay on a high-density oligonucleotide array. *Genome Res* 2004; 14 (3): 414–425.
134. Broude NE, Zhang L, Woodward K, Englert D, Cantor CR. Multiplex allele-specific target amplification based on PCR suppression. *Proc Natl Acad Sci USA* 2001; 98 (1): 206–211.
135. Shapero MH, Zhang J, Loraine A, Liu W, Di X, Liu G, Jones KW. MARA: a novel approach for highly multiplexed locus-specific SNP genotyping using high-density DNA oligonucleotide arrays. *Nucleic Acids Res* 2004; 32 (22): e181.
136. Callow MJ, Drmanac S, Drmanac R. Selective DNA amplification from complex genomes using universal double-sided adapters. *Nucleic Acids Res* 2004; 32 (2): e21.
137. Dahl F, Gullberg M, Stenberg J, Landegren U, Nilsson M. Multiplex amplification enabled by selective circularization of large sets of genomic DNA fragments. *Nucleic Acids Res* 2005; 33 (8): e71.
138. 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 (4): 874–879.
139. Myers RM, Lumelsky N, Lerman LS, Maniatis T. Detection of single base substitutions in total genomic DNA. *Nature* 1985; 313 (6002): 495–498.
140. Underhill PA, Jin L, Lin AA, Mehdi SQ, Jenkins T, Vollrath D, Davis RW, Cavalli-Sforza LL, Oefner PJ. Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res* 1997; 7 (10): 996–1005.
141. Germer S, Higuchi R. Single-tube genotyping without oligonucleotide probes. *Genome Res* 1999; 9 (1): 72–78.
142. Bernard PS, Wittwer CT. Homogeneous amplification and variant detection by fluorescent hybridization probes. *Clin Chem* 2000; 46 (2): 147–148.
143. Solinas A, Brown LJ, McKeen C, Mellor JM, Nicol J, Thelwell N, Brown T. Duplex scorpion primers in SNP analysis and FRET applications. *Nucleic Acids Res* 2001; 29 (20): e96.
144. Nazarenko IA, Bhatnagar SK, Hohman RJ. A closed tube format for amplification and detection of DNA based on energy transfer. *Nucleic Acids Res* 1997; 25 (12): 2516–2521.
145. Myakishev MV, Khripin Y, Hu S, Hamer DH. High-throughput SNP genotyping by allele-specific PCR with universal energy-transfer-labeled primers. *Genome Res* 2001; 11 (1): 163–169.
146. Whitcombe D, Theaker J, Guy SP, Brown T, Little S. Detection of PCR products using self-probing amplicons and fluorescence. *Nat Biotechnol* 1999; 17 (8): 804–807.
147. Collins FS, Guttmacher AE. Genetics moves into the medical mainstream. *JAMA* 2001; 286 (18): 2322–2324.
148. Larsson C, Koch J, Nygren A, Janssen G, Raap AK, Landegren U, Nilsson M. *In situ* genotyping individual DNA molecules by target-primed rolling-circle amplification of padlock probes. *Nat Methods* 2004; 1 (3): 227–232.
149. Christian AT, Pattee MS, Attix CM, Reed BE, Sorensen KJ, Tucker JD. Detection of DNA point mutations and mRNA expression levels by rolling circle amplification in individual cells. *Proc Natl Acad Sci USA* 2001; 98 (25): 14238–14243.
150. Bignell GR, Huang J, Greshock J, Watt S, Butler A, West S, Grigorova M, Jones KW, Wei W, Stratton MR et al. High-resolution analysis of DNA copy number using oligonucleotide microarrays. *Genome Res* 2004; 14 (2): 287–295.
151. Li L, Shi H, Yiannoutsos C, Huang TH, Nephew KP. Epigenetic hypothesis tests for methylation and acetylation in a triple microarray system. *J Comput Biol* 2005; 12 (3): 370–390.
152. Collins FS, Green ED, Guttmacher AE, Guyer MS. A vision for the future of genomics research. *Nature* 2003; 422 (6934): 835–847.