Genotyping Techniques

Review

High-Throughput Methods for Detection of Genetic Variation

BioTechniques 30:318-332 (February 2001)

Vessela Nedelcheva Kristensen, Dimitris Kelefiotis, Tom Kristensen, and Anne-Lise Børresen-Dale
Norwegian Radium Hospital, Oslo, Norway, 1 Foundation for Research and Technology—Hellas, Heraklion, Crete, Greece, and 2 University of Oslo, Oslo, Norway

ABSTRACT

Understanding human genetic variation is currently believed to reveal the cause of individual susceptibility to disease and the large variation observed in response to treatment. In this review, we will focus on different approaches to identify and visualize genetic alterations. The various approaches for allele discrimination are formally systematically divided into (i) enzymatic approaches, in which the properties of different enzymes to discriminate between nucleotides are used (restriction enzymes type II, Cleavase and Resolvase, DNA polymerase, and ligase); (ii) electrophoretic methods, in which the allele discrimination is based on the difference in mobility in polymeric gels or capillaries (single- and double-stranded conformation assays, heteroduplex analysis, and DNA sequencing); (iii) solid-phase determination of allelic variants, including high-density oligonucleotide arrays for hybridization analysis, minisequencing primer extension analysis, and fiberoptic DNA sensor array; (iv) chromatographic methods such as denaturing high-performance liquid chromatography (DHPLC); (v) other physical methods of discrimination of allelic variants such as mass spectrometry (mass and charge) or fluorescence exchange-based techniques; and (vi) in silico methods such as high-throughput analysis of expressed sequence tag data. The most frequently used techniques and instrumental settings applied in different combinations are described, and other methods that are less broadly used but have interesting potentials are discussed.

INTRODUCTION

The Human Genome Project will bring the first draft of the full-length sequence of the human genome. The next challenge will be to compare sequences from a large number of individuals to create a map of 100,000 of the most common type of genetic variation, single nucleotide polymorphisms (SNPs). Similar to the government and academic effort, a consortium of 10 pharmaceutical companies and the Wellcome Trust are in the process of identifying an expected number of 300,000 SNPs within the next two years. Efforts from biotechnology companies—most notably Celera Genomics (Rockville, MD, USA), Genset (Paris, France), CuraGene (New Haven, CT, USA) and Incyte Genomics (Palo Alto, CA, USA)—will make available data sets with polymorphic sequences. Non-biotechnology companies such as IBM (http://www.res.ibm.com) also see intellectual and financial interest in utilizing their potential in analysis of biological data. These tremendous developments in both the chemistry and bioinformatics of identifying genetic variants pose pressure on small and middle-size academic units to improve and intensify by orders of magnitude the existing methods for genotyping. In this review, we discuss the most frequently used techniques and some techniques with interesting potential for future development (Table 1). The process of mutation analysis is formally divided into two steps: (i) the identification of mutations according to the physical or enzymatic principle used to reflect the change in the DNA primary structure and (ii) the visualization of the detection products, which involves ways of making this change visible (e.g., labeling and allele separation strategies). This division is formal and arbitrary. The available methods can also be divided according to whether they are based on previous knowledge of the nature of genetic alterations or the need to screen for unknown mutations (Table 2). This division is also not completely precise because a number of screening techniques for unknown mutations do not
give the precise nature of the alteration and thus cannot be directly considered to be alternatives to the screening for known mutations.

**ENZYMATIC APPROACHES FOR DISCRIMINATION OF ALLELIC VARIANTS**

Historically, the first widely used method for the detection of polymorphisms exploited alterations in restriction enzyme sites, leading to the gain or loss of cutting events. A variety of enzymatic approaches for the discrimination of allelic variants has since been described. Even though not all can be considered high throughput in their present format, data from both academic and industrial sources suggest that they have the potential to be developed into fully automated assays.

<table>
<thead>
<tr>
<th><strong>Enzymatic Approach for Discrimination of Allelic Variants</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Restriction fragment length polymorphism (RFLP)</td>
</tr>
<tr>
<td>Amplified fragment length polymorphism (AFLP)</td>
</tr>
<tr>
<td>Cleavage fragment length polymorphism (CFLP)</td>
</tr>
<tr>
<td>Ligation-based assays</td>
</tr>
<tr>
<td>Random amplified polymorphic DNA [RAPD (AP-PCR)]</td>
</tr>
<tr>
<td>Direct termination PCR (DT-PCR)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Electrophoretic Discrimination of Allelic Variants</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Single strand conformation polymorphism (SSCP)</td>
</tr>
<tr>
<td>Fragment analysis</td>
</tr>
<tr>
<td>High-performance DNA sequencing</td>
</tr>
<tr>
<td>Doublex sequencing</td>
</tr>
<tr>
<td>Single sequencing reaction (SSR)</td>
</tr>
<tr>
<td>Clipper sequencer</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Solid-Phase Determination of Allelic Variants</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Oligonucleotide arrays</td>
</tr>
<tr>
<td>High-density oligonucleotide arrays for hybridization analysis</td>
</tr>
<tr>
<td>Minisequencing primer extension analysis</td>
</tr>
<tr>
<td>Fiberoptic DNA sensor array</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Chromatographic Methods for Discrimination of Allelic Variants</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Denaturing high performance liquid chromatography (DHPLC)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Physical Methods for Discrimination of Allelic Variants</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Differential sequencing with mass spectrometry</td>
</tr>
<tr>
<td>Fluorescence exchange-based methods</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>In Silico—Analyzing EST Data</strong></th>
</tr>
</thead>
</table>

Restriction fragment length polymorphism (RFLP) is comprised of PCR amplification of a fragment of interest and the subsequent digestion with a restriction enzyme (60). Because of its simplicity and straightforwardness, the method has been extensively used and is still popular, although it confers certain limitations. Only a subset of polymorphisms that happen to reside in an endonuclease restriction site can be studied with the conventional method. The approach has recently been generalized by swapping the amplification and restriction events—the DNA sample is first digested by a combination of restriction enzymes and then amplified. Amplification of the fragments is achieved by ligating adapter sequences on both sides of the fragments, which, together with the restriction target sequence, can be amplified using a specific end-labeled primer [amplified fragment length polymorphism (AFLP)]. A unique triplet of nucleotides following the adapter and restriction enzymes sequences on the primers enables specific amplification of fragments beginning or ending with the complementary triplet. This fingerprinting method allows co-amplification of a high number of restriction fragments without knowledge of the nucleotide sequence (68). A complicating factor may be that the successful implementation is dependent on the completeness of digestion in both studied and reference samples.

**Cleavage Fragment Length Polymorphism Analysis**

A large number of mutation detection strategies are based on the dependence of the secondary structure on the primary sequence of single-stranded fragments or on the kinetics of their formation or renaturing process. One way to treat thermally denatured and renatured strands of DNA is to cleave them with a structure-specific endonuclease, thereby digesting the target heteroduplexes into a series of labeled fragments that can be resolved by gel or capillary electrophoresis (58). The assay has been introduced by Third Wave Technologies (Madison, WI, USA) and can accurately detect sequence changes, including base pair substitutions, insertions, and deletions. A complicating factor may be that the successful implementation is dependent on the high yield of high-grade purified DNA (58). Engineered structure-specific endonucleases such as the Cleavase I enzyme recognize and cut stem loops that form imperfect hybrids. The method has been recently applied to detect a number of mutations in the TP53 gene (53). Another interesting enzyme, Resolvase, cloned from bacteriophage T4, has also been used in a so-called enzymatic mutation detection (EMD), which exploits the same approach to detect all types of mismatches (67). This enzyme, also known as endonuclease VII, cleaves the concatameric DNA intermediates formed dur-
ing phage DNA replication. Labeled primers in the amplification step of the reference and the sample DNA allow the analysis of the cleaved heteroduplexes on systems for automated DNA sequencing systems, such as the ABI 377 or ABI 310 (both from Applied Biosystems, Foster City, CA, USA), with the respective increase of throughput and possibilities for robotization (see below).

Invasive Cleavage of Nucleotide Probes

Another application of endonucleases for mutation detection involves Flap endonucleases, enzymes that recognize overlapping nucleotides on a target DNA strand (44). These enzymes are associated with DNA repair and have the function to excise redundant portions (flaps) of the displaced strand occurring during reparative DNA replication. Two oligonucleotides that contain sequences strictly following one another hybridize with a target DNA. The cleavage of the overlapping nucleotide with a Flap endonuclease will be dependent on whether the upstream oligonucleotide (called in the study invasive) overlaps perfectly the 5’ region of the downstream oligonucleotide (signal oligonucleotide), indicating the presence of a target sequence (44). With the use of thermostable Flap endonucleases, the reactions may be performed near melting point temperature of the signal fluorescent probe, allowing multiple copies of it to anneal in succession to a single copy of target molecule and to be cleaved one after another. This allows PCR-free amplification of the cleavage signal. The assay enables simple differentiation of homozygotes and heterozygotes on single-copy genomic sequences. The cleavage signal was shown to be proportional to the amount of target DNA over more than a four log span (44).

Ligase Detection Reaction

Suitably designed allele-specific primers flanking a mutation site on both sides are ligated together if and only if a perfect match of the allele-specific oligonucleotide (ASO) provides all correct nucleotides for the ligation reaction to occur. The resultant ligated fragments can be electrophoresed and analyzed on all types of fluorescence detection systems, as one of the primers used in the reaction is fluorescently labeled. This principle has been recently developed in an array format (16).

Random Amplified Polymorphic DNA (Arbitrarily Primed PCR)

Random amplified polymorphic DNA (RAPD) or arbitrarily primed PCR (AP-PCR) is a technique that allows the detection of qualitative and quantitative differences in the entire genome of two compared systems at the DNA level. The approach is similar to the differential display at the RNA level. RAPD is a PCR-based fingerprinting technique that uses arbitrary primers to amplify DNA or cDNA at random, thus generating a fingerprint specific for a given genome. Because of the arbitrary nature of priming, no preliminary information about the precise sequence is needed. An optimized protocol for the detection of mutations or polymorphic loci in human tissues had been reported recently, showing that the successful implementation is highly dependent on the enzyme chosen for amplification (61).

Directed Termination PCR

A one-step PCR-based method for the detection of SNPs, insertions, and deletions has been proposed that uses limiting amounts of one of the four dNTPs in a conventional PCR protocol (7). Since the nucleotide in limited concentration is a rate-limiting factor in amplification, after a given number of cycles with full-length amplification, DNA synthesis halts at one of the rate-limiting steps. Two ladders are created (each originating from each primer) of differently terminated fragments of diminishing length. The terminated versus full-length signal ratio can be increased by using 35S-labeled dNTP analogues. These are difficult substrates for the polymerase and are therefore utilized after the depletion of the unlabeled dNTP (i.e., in the later cycles in the PCR, when the terminated rather than full-length fragments are amplified). This PCR, which resembles sequencing or sequencing during PCR, creates two families of mutant molecules. These two classes can be visualized as multiple shifts in denaturing gels (DT-SSCP) and are much easier to detect than the single mutant products occurring during conventional single-strand conformation polymorphism (SSCP) (see below). Although the analysis is performed in one step, optimization of both the full-length PCR and the conditions for the terminating reactions are needed for each SNP.

<table>
<thead>
<tr>
<th>Analysis of Known Variants</th>
<th>Screening for Unknown Variants</th>
</tr>
</thead>
<tbody>
<tr>
<td>RFLP</td>
<td>SSCP and heteroduplex analysis</td>
</tr>
<tr>
<td>ASO-PCR</td>
<td>CDGE, DDGE</td>
</tr>
<tr>
<td>Invasive cleavage of nucleotide probes</td>
<td>DNA sequencing</td>
</tr>
<tr>
<td>Ligation-based assays</td>
<td>RAPD (AP-PCR)</td>
</tr>
<tr>
<td>DSCA</td>
<td>DT-PCR</td>
</tr>
<tr>
<td>Molecular beacons</td>
<td>ALFP</td>
</tr>
<tr>
<td>Microarrays “loss of signal”</td>
<td>CFLP, EMD</td>
</tr>
<tr>
<td>FRET</td>
<td>Microarrays “gain of signal”</td>
</tr>
<tr>
<td></td>
<td>Minisequencing</td>
</tr>
</tbody>
</table>

ELECTROPHORETIC DISCRIMINATION OF ALLELIC VARIANTS

Single-Strand Conformation Polymorphism

SSCP and heteroduplex analysis are among the first established methods for the detection of genetic polymorphisms. Because of their simplicity and relative-
ly high sensitivity, these methods are still widely used and further elaborated (recently reviewed in Reference 48). The conventional SSCP analysis involves the denaturing of PCR-amplified fragments with the subsequent formation of sequence-specific secondary and tertiary structures of the single strands during non-denaturing gel electrophoresis. The electrophoretic mobility then depends on the 3-D shape of the single-stranded molecules. The classical “long” gels and $^{32}$P-labeled fragments may nowadays be replaced by fluorescently labeled fragments and automated capillary electrophoresis (ABI PRISM™ 310; Applied Biosystems) (38). The simplicity of the method, combined with automation and short analysis time, contributes to a high-throughput analysis at relatively low cost.

**Heteroduplex Analysis**

If the denatured PCR products are instead allowed to slowly renature, they form DNA duplexes. The duplexes with the same sequence on both strands (homoduplexes) or with a single base pair mismatch on one strand (heteroduplexes) have different electrophoretic mobility in a native gel. In the case of a single base pair substitution, the heteroduplex (called “bubble” type) can easily be separated from a homoduplex. Recently, an assay designated double-strand conformation analysis (DSCA) has been described (2) in which duplexes with fluorescein-labeled reference (FLR) are separated by a non-denaturing gel electrophoresis. The method may be regarded as a multiple heteroduplex analysis in which PCR products with unknown genotype are renatured with Cy5-labeled PCR products of a known genotype. This approach makes it possible to improve conventional heteroduplex analysis in several ways. One can modify both the amount and quality (sequence) of one of the participants in the renaturing process (the labeled reference) to improve resolution before the PCR amplification.

**Fragment Analysis**

Microsatellite repeats are detected by electrophoresis of radiolabeled products in denaturing polyacrylamide gels or by manual or automated sequencing. Multiple 96-capillary array electrophoresis using polyvinylpyrroldone (PVP) sieving matrix has recently been described, which allows accurate determination of the size of the allele by interpolating from the relationship between the fragment size and its migration in the gel (18). Software with complete statistics enables one to determine confidently genotypes of high complexity with high throughput.

Genotyping of microsatellite repeats at lower cost is possible by using chemically modified short acrylamide gels such as Spreadex® (Elchrom Scientific, Jamaica Estates, NY, USA) or modified agarose gels. A polymer, bis-benzi-mide-polyethylene glycol (PEG), which binds preferentially to AT-rich sequence motifs can be used for the detection of purine bases by amplifying the differences in mobility of the fragments when size alone is not sufficient for separation. The polymer is mixed with an agarose gel to retard differentially the AT-rich fragments during conventional horizontal electrophoresis (49).

**High-Performance DNA Sequencing**

DNA sequencing has been proclaimed the “gold standard” used to establish the identity of both known and unknown sequence specific nucleotide variations (14), which in turn precipitated a number of publications questioning this statement (11,15). Four different dyes are used to label each of the DNA bases (A, C, G, and T) that can be electrohoresed together in the ABI sequencing system. The DNA polymerase may discriminate between the differently labeled ddNTPs, resulting in different intensities of terminating signals [i.e., uneven peaks, which may undermine the detection efficiency of heterozygotes, a problem practically solved with the use of stronger fluorescent signals—the BigDyes™ (Applied Biosystems)]. This increased precision is combined with high throughput in the currently available systems (ABI 310 and ABI 3700). On the other hand, the Pharmacia-ALF sequencing technology (Amersham Pharmacia Biotech, Piscataway, NJ, USA) employs a single-dye chemistry format in which Cy5 is used to label DNA fragments for the detection of laser-induced fluorescence. This contributes to obtaining easily interpretable raw data, high accuracy, and no mobility differences between fragments caused by different-size fluorophore labels. However, the disadvantages are lower throughput and higher labor cost. Several attempts to increase the throughput of ALF DNA sequencing have been described. Doublex sequencing involves the combination of an ALF DNA sequencer™ and an ALFexpress™ DNA sequencer, which allows the detection of two different labels, fluorescein isothiocyanate (FITC) and Cy5. With the use of two differently labeled primers, one can sequence both strands in a single reaction and gel run (56). Another approach for the detection of known SNPs is to use single sequencing reactions (SSR) in which termination reactions for only one of the four ddNTPs are generated and analyzed (34). New platforms for rapid automated DNA sequencing have been introduced using ultra-thin gels and thin capillaries (the Clipper sequencer) (72), and protocols for simultaneous PCR amplification and cycle sequencing in a single tube (CAS, DEXAS, or CLIP) have been described (32,57,72).

**FROM GELS TO CHIPS—SOLID-PHASE DETERMINATION OF ALLELIC VARIANTS (OLIGONUCLEOTIDE ARRAYS)**

Short oligonucleotides (8–80 nucleotides) or longer cDNA (more than 100 nucleotides) attached to a solid support (microscope slide or microplate) are referred to as microarrays. High-density oligonucleotide arrays on miniature glass supports are also designated as DNA chips. The attachment of one of the components of a reaction to a solid support allows fast separation of the different components of the reaction by simple washing. The immobilized nucleic acid may be either the known sequence, referred to as the probe, or the unknown (questioned or interrogated) sequence, also referred to as the target. Such microarrays may be used in mutational and comparative sequence analysis (reviewed in Reference 25). The probes, attached to a solid support, may be hybridized to the unknown

---

**Genotyping Techniques**
sequence, and either the completeness of hybridization is used as a signal for the presence or the absence of a mutation or the target is used as a template on which the probe is further extended as a sequencing primer with one or more nucleotides to identify the first proximal base(s).

**High-Density Oligonucleotide Arrays for Hybridization Analysis**

Hybridization analysis on microarrays extends an attempt made in the 1980s to sequence DNA by hybridization to universal combinatorial oligonucleotides of certain length (usually 8mers) (45,63). As the length of reads was an important survival factor in the evolution of sequencing techniques at that time, the method could have remained a dead end because base calling accuracy decreased rapidly for larger targets. The growing need to compare known sequences of moderate length to each other, rather than continuously read through new sequences in steps as long as possible, brought the comeback of hybridization techniques. Several variations on this approach have been described to date. The highest complexity setup is needed to search for unknown changes, in which four possible nucleotides are present at each position of a nucleotide to form a variant detector array (VDA) (59). The perfect match probe presents a stronger fluorescence signal relative to the intensity of the three mismatched probes. Using this technique, a large screen for SNPs using 2 Mb genomic sequence, comprising about 16 000 human sequence tagged sites (STs) has recently been reported. There was very good agreement (90%) between findings with high-density detection of DNA variation and conventional gel-based sequencing analysis (71). This approach is also referred to as the “gain-of-signal” assay because the mutant allele should produce a stronger signal when hybridizing to its perfectly matching counterpart (25,59). Oligonucleotides containing all four mismatches, all four possible insertions, or deletions of one, two, or three nucleotides may be synthesized for each examined position on the target DNA. Simpler experimental design is applied for known SNPs or sets of mutations when only oligonucleotides for each variant allele are used in the hybridization, instead of all four possible probes, as described above. Such allele-specific oligonucleotide blocks have been developed to genotype a full set of known single-base substitutions in CFTR (12), human tyrosinase gene (23), and cytochrome P450 (13). The gain-of-signal approach has also been applied for genes of the HIV reverse transcriptase and protease (33), human mitochondrial DNA (6), Mycobacterium tuberculosis rpoB gene (19), and BRCA1 (24). An even simpler setup for genotyping is achievable if an oligonucleotide array consists of oligonucleotides representing only the wild-type allele sequences (i.e., only one oligonucleotide per variant base, instead of two with ASO or four with VDA (multiplied by two for both strands). Here, all wild-type oligonucleotides will match perfectly, and all deviations from the wild-type will be detected with a weaker signal; thus, the approach is designated “loss of signal”. Although the simplicity in the set of probes keeps the probes to a reasonable number and is particularly suitable for screening for low-frequency SNPs, it unfortunately does not allow closer characterization of the change in the DNA, and additional conventional sequencing analysis is needed. In reality, multiple probes (n probes for n nucleotides in an oligonucleotide) are used, thus minimizing random sources of errors (6). References of known sequence tagged with a second label have been added to compete with the labeled test targets in two-colored assays to increase the specificity and sensitivity of the method. This approach using fluorescein (green)-labeled reference and biotinylated (red) targets has been used to characterize sequence changes in the coding region of ATM (26), BRCA1 (24), and large segments of mitochondrial DNA (6). By simple gain of signal analysis, 99% of the entire 16.6-kb human mitochondrial genome was read correctly, and only 2% of the 2.5-kb mitochondrial sequence interrogated for polymorphisms needed to be sequenced by conventional methods (6). Gain-of-signal and loss-of-signal approaches have been directly compared by analyzing the 3.43-kb exon 11 of BRCA1 for heterozygous sequence variations (24). The loss-of-signal assay had higher sensitivity and specificity compared to the gain-of-signal assay, especially for large deletions or single base pair substitutions.

Microarrays have evolved by increasing the density of spotting and the number of oligonucleotides, thereby becoming a powerful tool for the detection of candidate SNPs (52). Further development is needed to increase the specificity and sensitivity of the high-density hybridization arrays. The correct identification of a sequence change may be compromised by cross-hybridization between single-nucleotide deletions and single-nucleotide substitutions, as well as between mismatched probes masking the other allele in heterozygotes (25). The hybridization is dependent on secondary structures and the sequence context because more complex repeat sequences or single-nucleotide stretches, insertions, and deletions may have several hybridization patterns complicating the evaluation of results. Despite the impressive results, this leading method for high-throughput mutation detection in its present form is still dependent on a large amount of material and has limited potential to detect small copy number of mutated alleles in high wild-type background.

**Minisequencing Primer Extension Analysis**

Some of the technical problems of the analysis based on hybridization described above may be solved by minisequencing assays because the thermal stability of the hybrids is no longer of detection value. The hybridized target is instead used as a template on which DNA polymerase further extends the immobilized probe without proofreading activity, using one or more labeled nucleotides to identify the first proximal base(s). Labeled ddNTP may be used instead of dNTPs to both label and terminate the chain elongation, and, thus, different dyes may simultaneously be used to label the separate chains (50). In contrast to the chips used in the hybridization analysis (the Affymetrix® chip; Affymetrix, Santa Clara, CA, USA), the oligonucleotide probes are attached to the surface of a glass or
microplate via the 5’ end, leaving the 3’ end free for extension. The concept has been developed based on primer extension (65) or primer-guided nucleotide incorporation (62) for the detection of nucleotides and has many variations and abbreviations: single-nucleotide primer extension (SnuPE), allele-specific primer extension (AS-PE capture), genetic bit analysis (GBA), arrayed primer oligonucleotide base extension (PROBE), and the Pint-Point assay (reviewed in Reference 65) (Table 1). Although the solid-phase minisequencing has attracted attention, primarily due to its possibilities to be extended in a microarray format, there are several other commonly used detection methods for this assay, including gel-based detection, homogenous detection, and ELISA-based formats (65) or a simple gel electrophoresis (55,62).

Either PCR products or primers synthesized *de novo* may be used as probes in the minisequencing reaction (50,66). The immobilization of the single-stranded primers on the solid support may be through biotin-avidin/streptavidin reaction (66) or covalently via 5’-disulfide groups (50,54). Double-stranded templates may also be used with reduced efficiency of the reaction or in cycle sequencing protocols (9,36). In addition to fluorescent tags, ddNTPs may be labeled with biotin (if it has not already been used for immobilizing the probe) or haptons, detected indirectly through antibodies conjugated to alkaline phosphatase or peroxidase using colorimetric or hemiluminiscent markers in ELISA formats (65). All possible sequence changes in a region of the *TP53* gene have been determined with this assay (27). Fast and reliable multiplex PCR for the preparation of both probes and targets, as well as the availability of the inverted (compared to the Affymetrix 3’-5’chip) high-density 5’-3’ microarray, is needed to validate the primer extension approach for large-scale genotyping.

**Fiberoptic DNA Sensor Array**

Replacement of microscope slides as solid supports in microarrays by optical fibers offers easier handling of both excitation and emission radiations (70). An integrated fiberoptic sensor array capable of simultaneously monitoring multiple hybridization events has been reported and used to genotype the H-ras oncogene (28). Field excitation of fluorophores at the side of an optical fiber and the collection of the emitted light on the same fiber have been used to construct a single fiber evanescent wave hybridization sensor. Oligonucleotides covalently attached to the evanescent field surface attract their complementary target. Fluorescence signal is generated by labeled target or competitive DNA molecules, intercalation reagents, or anchored molecular beacons that become fluorescent upon hybridization to the target sequence.

**CHROMATOGRAPHIC METHODS FOR DISCRIMINATING ALLELIC VARIANTS**

Several techniques for the detection of mutations [e.g., constant gradient gel electrophoresis (CGGE), denaturing gradient gel electrophoresis (DGGE), and temperature gradient gel electrophoresis (TGGGE)] are based on the subtle differences of melting points of DNA fragments dependent on base pair composition and the resultant difference of mobility of the mutant fragment in gels. PCR products are obtained and are partially denatured by formamide, urea (CDGE, DGGE), and temperature (TTGE) and are run in constant (CDGE) or gradient denaturing gels (DGGE). Clamped primers are often used to amplify the difference in melting points. Instead of electrophoresis as a detection step, the partially heat-denatured PCR products may be analyzed in a column containing alkylated nonporous particles by denaturing high-performance liquid chromatography (DHPLC) (29). This detection method has gained publicity as an HPLC method analogous to those used in analytic chemistry; however, some authors suggest temperature-modulated heteroduplex chromatography (TMHC) as a term to more accurately describe the process (35). The method has recently been demonstrated to rapidly detect homozygous and heterozygous sequence alterations in the *CFTR*, *RET*, and *PTEN* genes (43), *BRCA1* (3), and genetic polymorphisms (10,22). The mutation analysis with DHPLC can be almost totally automated with an autosampler on one end and a fraction collector on the other. Analysis is rapid (about 5 min per sample), with a simple evaluation of data, distinguishing between simple and multiple peaks in the elution profiles, and allows lengths of up to 1.5 kb DNA to be analyzed. A disadvantage may be the recommended use of *Pfu* DNA polymerase (35), which, as a high-fidelity enzyme, allows sharper peaks but may be less successful in amplifying some regions. The use of *Pfu* DNA polymerase may be avoided, especially for analyzing shorter fragments, in which the probability of mismatch caused by the polymerase is low, as several validation studies have evaluated the accuracy of mutation/polymorphism detection with DHPLC using *Tag* DNA polymerase or AmpliTaq Gold® (Applied Biosystems) and registered no false positives (3,43).

**OTHER PHYSICAL METHODS FOR DISCRIMINATING ALLELIC VARIANTS**

**Differential Sequencing with Mass Spectrometry**

Differential sequencing with mass spectrometry is an attractive alternative for high-throughput detection of polymorphisms because it allows the generation of accurate data in seconds per sample without any probes or labeling of targets (20). Matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) (21) and electrospray ionization mass spectrometry (ESI-MS) (1) are techniques for creating ionized gas-phase DNA whose ionized molecules are accelerated in an electric field, followed by a flight through a vacuum chamber to a detector. This detection method, also called time-of-flight measurement (TOF-MS), is based on the time each particle needs to fly to the detector, inverse to the velocity, which in turn is proportional to the mass/charge ratio of a flying particle, a specific characteristic for the
Genotyping Techniques

Figure 1: Schematic view of the main principles of interaction detection and visualization.
samples under examination (20). Since the mass/charge ratio is an intrinsic property of the molecule (fragment DNA), dependent only on base composition, this detection method is able to address most of the limitations of the hybridization and fluorescence-based approaches discussed above. The method has been developed in an array format that allows the analysis of very small amounts (nanoliter scale) of product (41). Since the kinetics of flight to the detector is precisely mathematically describable as a function of mass, it is possible to create virtual libraries for SNPs to predict the behavior of the different genotypes before experiment or a library of wild-type references (20).

Mass spectrometry analysis has been successfully applied as a detection step in the cleavage assay or in the primer extension minisequencing techniques (both described above) as PROBE or Pint-Point. Mutation analysis in several different types of cancer (42), diseases such as cystic fibrosis (4), and hypertension (40) has been performed by PROBE. The method has been particularly useful in analyzing “difficult” sequences [e.g., tri-nucleotide repeats (69)]. MALDI-MS has been used to visualize direct reads from sequencing TP53 (exons 5–8) and was shown to be able to detect reliably all mutations, heterozygous and homozygous and false stops of the sequencing reaction (17), as well as a number of SNPs (21,31,39).

Fluorescence Exchange-Based Method

Single-phase minisequencing tries to circumvent all washing steps by enzymatically degrading the excess of PCR primers and dNTPs and measuring bound \(^3\)H-dNTPs (30) or pyrophosphate released from incorporated dNTPs by the enzymatic luminometric method (51) (Figure 1). A fully homogenous minisequencing assay has been reported to use alkaline phosphatase to inactivate the unused primers and ddNTPs and fluorescence resonance energy transfer (FRET) to detect the quenching caused by overlap of emission spectra between the labeled ddNTP and the 5’ FITC-labeled probe using the TaqMan\textsuperscript{®} system (ABI 700) (8) (Figure 1). Another way to visualize point mutations in a solution in single tubes is by hybridizing the fragment carrying the mutation to single-stranded oligonucleotide probes that become fluorescent when they bind to perfectly complementary nucleic acids (46). Such probes (molecular beacons) possess a stem and loop structure that keeps the fluorophore and quencher covalently bound to each side of the probe so close to each other that fluorescence does not occur unless the beacon is stretched by hybridizing to its target (Figure 1). The number of SNPs simultaneously detected with this method is dependent on the number of fluorophores used and the quality of the beacons (efficiency of quenching in dormant state). Future reports will verify the universality of this approach for different sequence targets.

FUTURE PERSPECTIVES

DNA chip-based strategies, mass spectrometry-based hybridization detection, or new, radical, and revolutionary techniques may influence the future of genetic analysis. One such method, radically different from the existing approaches, is to use ion channels as sensors for the detection of base pair composition of DNA. If a single strand of DNA is pulled through such an ion channel, the bases on the DNA string will periodically block these pores that are slightly more than 1 nm wide, causing a reduction in the ion flow. This reduction may be specific for each of the four nucleotides, assuming that each of them will block the pore to different extents, and one might theoretically be able to measure these tiny changes in the thin current of ions. This may allow DNA sequencing at extremely high speed by simply pulling an ssDNA through an ion channel (47). Ironically, the high speed is a technical problem at present. The molecular chain ploughs through the pore so quickly that it is very close to the detection limits of measurements at each base, and attempts are made to slow it down by using both strands, the one remaining outside the pore and breaking the passage of the other through it. At present, this approach appears to be more intellectually inspiring than technically achievable, but it may help to make some brave predictions about what the future perspectives might be.

The rapidly accumulating sequence data available at various genetic databases on the Internet allow the analysis of sequences and screening for polymorphic sites without actually sequencing (37). A total of 10435 possible new variations in human genes with various levels of confidence (available at http://lpg.nci.nih.gov/GAI) has been deduced by the alignment of multiple entries of expressed sequence tags (ESTs) or full-length sequences of various genes (5). A detection system for screening this public-domain sequence data called Single Nucleotide Polymorphism Pipeline (SNPpipeline) has been developed, and SNP assemblies are available at http://www.chlc.org/cgi (5). This bioinformatics tool identified 3089 “candidate SNP” loci of high confidence. Validation experiments using DNA sequencing and RFLP for those loci affecting the restriction enzyme site showed an overall yield of correctly predicted sites of 56%.

CONCLUSION

Further analysis of the growing volume of sequence data will enable the detection of rare allelic variants and the amount of noise in the databases by confirming more and more candidate SNPs. A larger number than so far expected may be functional (64). The increase of throughput and sensitivity of mutation analysis reveals the genetic basis of a great number of malignant, neurological, immunological, and other diseases, and an increasing number of new genetic guides and databases are being created. There is, however, an increasing gap in throughput capabilities between industrial laboratories or pharmaceutical companies and small academic laboratories. As long as even relatively small scientific units should have the flexibility to create and test their hypothesis by themselves, it is critical to adapt the existing large-scale genotyping methods for small-budget academic units. The present review summarizes the existing strategies for genotyping and may be helpful for making the correct choices.
Genotyping Techniques

ACKNOWLEDGMENTS

This work was supported by research grant no. D99061/004 from the Norwegian Cancer Society and grant no. 122772/310 from the Research Council of Norway and Det Norske Videnskaps Akademi.

REFERENCES

Genotyping Techniques


Address correspondence to Dr. Vessela Nedelcheva Kristensen
Institute for Cancer Research
The Norwegian Radium Hospital
Montebello 0310
Oslo, Norway
e-mail: nedelcheva.vessela@dnr.uio.no