Mutation Detection

Introductory article

RGH Cotton, *Genomic Disorders Research Centre, St Vincent's Hospital Melbourne, Fitzroy, Australia*

Mutation detection is essential for the practice of modern biology. Because sequencing deoxyribonucleic acid (DNA) to detect changes in DNA (mutations) is not cost-effective, a range of methods have been developed to detect mutations for the first time and a different set of methods to look for previously defined mutations.

Article contents

doi: 10.1038/npg.els.0005672

Introduction

Changes in the sequence of the deoxyribonucleic acid (DNA) in a living organism can be harmful, harmless or even beneficial to that organism. In humans, harmful mutations can cause disease, for example cystic fibrosis, cancer and other common diseases such as asthma and diabetes. In the case of common diseases such as diabetes, the disease may be caused by mutations which when present alone may not cause disease but which in combination with other genes with similar harmless mutations may cause disease. Mutations that are completely harmless in humans, often called polymorphisms or single nucleotide polymorphisms (SNPs), can be used as aids to finding genes that cause disease and, as such, are used as markers along the genome. Because these polymorphisms may be implicated in common diseases (see above), they have attracted enormous scientific attention. The obvious way to detect these mutational changes is to sequence the DNA by the standard methods, and this is what is often done. However, the standard method in use for sequencing (the Sanger method) is not a trivial exercise and is performed best in skilled hands in a high-throughput environment. Even then the method requires an expensive separation step and complex analysis, especially when used for mutation detection. This situation has led to numerous methods being developed since 1985 to test lengths of DNA for the presence of mutations before it is sequenced, thus potentially eliminating a large amount of sequencing. (***See*** Mutations in Human Genetic Disease; Single-base Mutation.)

Logistically, detecting a mutation for the first time is a very different matter from the situation when one has already detected the mutation in a family, for example, and one wishes to test for this same mutation which might be expected in other members of the family. This has led to two very different sets of methods: those for previously unknown mutations (DNA scanning or mutation discovery methods) and those for previously known mutations (diagnostic, scoring or genotyping methods).

There is a perception that perhaps one day there will be an ideal method for detection of mutations. However, this has clearly not yet been found. This is for two major reasons:

- All the methods to be discussed, particularly the scanning methods, have their disadvantages.
- Some of these less than perfect methods are better in a particular project or laboratory than another; it should be remembered that users of these methods range from laboratories in developing countries to those in the major pharmaceutical companies.

This article attempts to look at the methodology in use at the beginning of the twenty-first century, its impact, strengths and weaknesses and requirements and expectations for the future. (***See*** Robotics and Automation in Molecular Genetics.)

Importance of Mutation Detection

Changes in DNA sequence in organisms (mutations) occur naturally and can be accelerated by the effects of radiation or chemicals. Natural sequence changes (variation) are the main driver of evolution, some DNA variation giving rise to advantageous properties allowing better survival in a particular environment. Conversely, this variation can be harmful and give rise to disease and cancer in humans. (***See*** Genetic Variation: Polymorphisms and Mutations; Genetic Disease: Prevalence.)

A mutation can also be harmless or neutral, but in certain circumstances it can be harmful. Thus, a certain sequence variation may reduce the activity of a drug-clearing enzyme, making the person involved susceptible to side effects from the drug, but a similar gene with such a change may, in combination with

other genes of this type, be responsible for common disorders such as asthma, diabetes, etc. This type of mutation is commonly called a polymorphism but more recently it has been referred to as a single nucleotide polymorphism or SNP (pronounced SNIP). (***See*** Single Nucleotide Polymorphism; Single Nucleotide Polymorphisms (SNPs): Identification and Scoring.)

The study of the evolution of humans (and other organisms) has been aided by the study of SNPs; thus the relationship of races and their origins can be carefully traced by the study of the presence, or absence, of particular SNPs particularly in the Y chromosome (for paternal inheritance). The information derived from these studies can be correlated with DNA from archaeological artifacts such as bone. (***See*** Ancient DNA: Phylogenetic Applications; Ancient DNA: Recovery and Analysis; Kinship Testing; Microarrays and Single Nucleotide Polymorphism (SNP) Genotyping; Polymorphisms: Origins and Maintenance.)

The examination of harmless variation in forensic medicine is becoming increasingly important for the identification of individuals who may have left DNA in some form at the scene of a crime. Thus, the pattern of variation in DNA can be measured in such a way that the chance of the pattern being the same for two individuals is of the order of 1 in 10^8.

Harmless variation is also critical in the identification of regions of the huge human genome that contain disease genes so that these regions might be more closely studied for genes causing disease. This is achieved by using such variations as markers or flags of specific regions of the genome. In simple terms, what is done is to use the so far identified SNPs which are equally spaced along the genome and follow the inheritance of each of the tens of thousands of these in a family with a particular disease where the mutant gene causing the disease is sought. The inheritance of each of these is correlated with inheritance of the disease itself where a succession of such SNPs in particular regions are inherited with the disease. This region is then identified for further study to identify the disease gene. In this strategy, the SNPs have to be identified and later they have to be detected in the individuals studied. Such studies are logically called linkage studies. The molecular basis of the side effects of drugs in particular individuals is of particular importance since these side effects can be lethal. As a consequence, and in the interest of identifying the individuals likely to have such side effects, the pharmaceutical industry has embarked on an enormous campaign to identify enzymes or transporter molecules which, when containing certain base changes or SNPs, slow the clearance rate of drugs from the body. This allows a higher than safe level of a drug in the body. This will ensure a more effective use of drugs and the production of genetic tests to identify individuals at risk so that an alternative drug might be used. (***See*** Genetic Linkage Mapping; Linkage Analysis.)

A similar effort is being mounted by the pharmaceutical companies to identify those genes that contribute to common human diseases (asthma, diabetes, cancer, heart disease). In this case, it is hypothesized that several variant genes (except as below) may combine to cause these diseases. Linkage studies as described above are being used, but also association studies. The latter involve first finding an SNP in a particular gene that is suspected of being involved, and then seeing if that particular variety of the gene is always found, for example, in all cases of asthma. If this is the case, or association has been found, then the gene must be under suspicion of being involved. (***See*** Genetic Epidemiology of Complex Traits; Polygenic Disorders; Polygenic Inheritance and Genetic Susceptibility Screening.)

Single-gene disorders are those diseases where a disease and mutant gene is inherited in a simple Mendelian manner. Classical disorders of this type are phenylketonuria, cystic fibrosis, hemophilia, sickle cell disease and thalassemia. The impact of this type of disorder on human health can be seen if one considers the fact that 10% of human disorders are inherited in this manner and thus must be due to a single inherited mutant gene. Since genes can first be localized through linkage studies, the genes in this localized region can be scanned for unknown mutations. These then can be tested to see if they are inherited in the family and thus be incriminated. Once such a mutation in a gene is labeled causative, it needs to be assayed in families for prenatal, prognostic or diagnostic reasons. (***See*** Genetic Disorders; Phenylketonuria; Predictive Genetic Testing.)

Cancer can be a single-gene disorder, for example mutations in *breast cancer 1, early onset* (*BRCA1*) and *breast cancer 2, early onset* (*BRCA2*) can cause breast cancer (accounting for 10–20% of breast cancer cases), the remaining cases being referred to as sporadic without known cause (as with other cancers); it is presumed that this category of cancer is caused by mutations/SNPs in a number of genes.

Methods – General

Many mutations are deletions or insertions ranging in length from a single base to many kilobases. While the smaller deletions and insertions to five bases will be detected by the methods listed below, those above 10–50 bases may be missed and require other methods as follows:

- Standard electrophoresis. Restriction digests of genomic DNA are probed with the gene in question for mutations of almost any size.

- Pulse field electrophoresis. This technique is needed for deletions which may be tens of kilobases long and include several genes. The gels are probed with genes of interest.
- Fluorescence *in situ* hybridization (FISH). With this method, cell or chromosome preparations can be probed with fluorescent copies of the genes (or regions thereof) which will only bind to one allele if the other is deleted.

In this section, only those methods that are most widely used or best known will be covered. (***See*** Fluorescence *In Situ* Hybridization (FISH) Techniques.)

Method types

Mutation detection is performed in two major modes:

- To detect mutations for the first time in a piece of DNA – this is a search and is referred to as scanning for, or discovery of, mutations.
- To detect on subsequent occasions mutations in DNA once they have been detected initially. This may be for diagnosis of disease in a family, when the methods are referred to as diagnostic methods, or in the case of population studies when assessing SNPs for linkage or other studies, when the methods are called genotyping methods.

Naturally there is overlap and blurring of the boundaries of these two types of methods, but it is useful for broad division.

Principles involved in the methods

Hybridization

Hybridization of complementary strands of DNA is fundamental to all methods from the amplification of DNA for study in the polymerase chain reaction (PCR) to the completed assay. This derives from the fact that base A always hybridizes with base T and base G with base C. Thus any strand of DNA of specific sequence in the living cell is matched and hybridized to a complementary strand, the hybridization being specific and very strong.

Mismatch

A mismatch in such a strand is fundamental to almost all the specific methods. When one hybridizes a piece of duplex DNA with another with a mutation in the central position, say an AT base pair instead of the original GC, the result is two different pairs of mispaired bases AC and TG. Hybridization is achieved by melting the mixture of the two DNAs and then cooling slowly, allowing the two mismatched structures of DNA to form as well as the original strands (see **Figure 1**). The possible mismatched base pairs are AA, TT, GG, CC, AC, AG, TG and TC. The critical point about DNA containing a mismatch (and hence signaling a mutation in one of the original strands) is that it decreases the affinity between the two strands which can be measured in such a way that the original mutation can be detected. This detection is usually achieved by detecting the altered melting point. (***See*** Basepair (bp).)

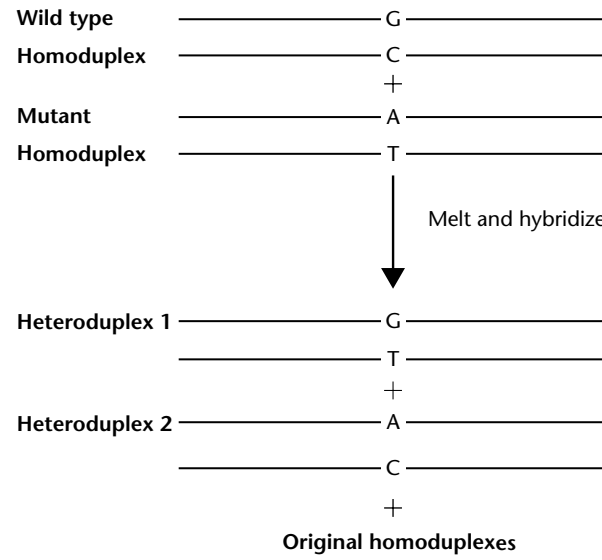

Figure 1 Heteroduplex formation producing mismatched base pairs is essential for many mutation detection methods. Hybridization of mutant and wild-type sequences is shown to produce two complementary heteroduplexes, and the original homoduplexes reform. Full lines represent the DNA sequence.

Mismatch cleavage or reactivity

When bases are mispaired they become more reactive to chemicals and they become the targets of enzymes that recognize distortions in the DNA, such as repair enzymes. Cleavage achieved through enzymes can localize the point of mismatch and hence the mutation.

Platform

The platform on which these methods can be mounted and hence analyzed can vary considerably even for a simple method. These will not be included in this article but they include electrophoresis on gels or in capillaries, separation on high-performance liquid chromatography (HPLC), mass spectroscopy, array or chip fluorescence and spectrophotometry. (***See*** Capillary Electrophoresis; Gel Electrophoresis; Microarrays in Disease Diagnosis and Prognosis; Microarrays and Single Nucleotide Polymorphism (SNP) Genotyping; Microarrays: Use in Mutation Detection.)

Methods for Mutation Discovery (see Figure 2)

Sequencing

All methods rely on sequencing the DNA to define the mutation, but this is also often the actual method of mutation discovery. The achievement of perfect sequence data requires considerable skill and experience. Laboratories that do achieve such excellent data often use this method for mutation discovery. Despite this there are numerous cases where mutations in genes have been missed by this procedure. Discovery of mutations can occur as a result of sequencing the human genome. This is because in some of the sequencing efforts the DNA of five individuals is sequenced. This allows discovery of polymorphisms or SNPs. This activity is further enhanced as still further samples of the same DNA are sequenced in other countries. As much of these data are placed in databases, this has given rise to the term searching for SNPs *in silico*.

Sequencing is needed once a signal is obtained in a study of a length of DNA for mutation discovery by one of the following methods. These methods evolved to save resources so that only stretches of DNA containing a mutation need to be sequenced.

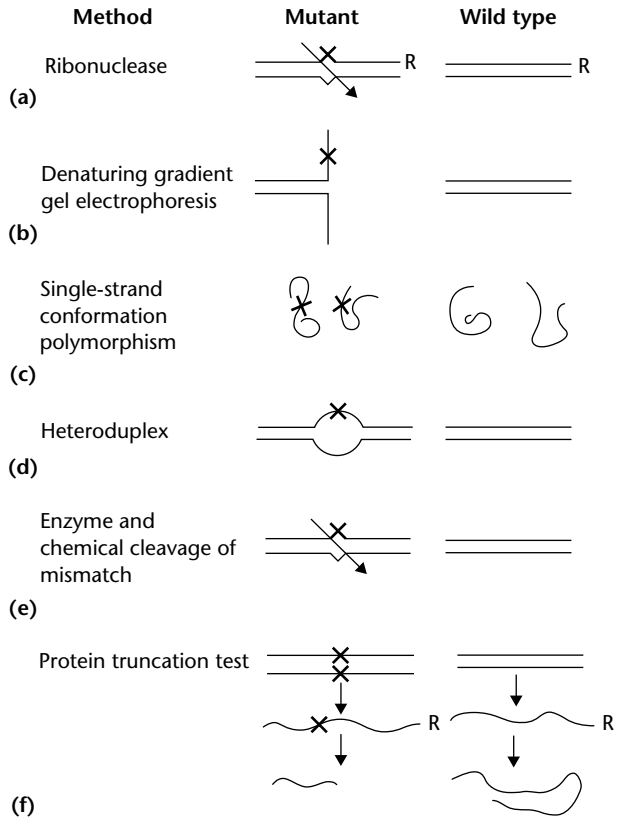

Figure 2 Schematic representations of methods to detect unknown mutations: (a) ribonuclease; (b) denaturing gradient gel electrophoresis; (c) single-strand conformation polymorphism; (d) heteroduplex; (e) enzyme and chemical cleavage of mismatch; (f) protein truncation test. Full lines represent the DNA sequence (except where RNA is indicated by R). Mutations are represented by an X. Arrows represent cleavage.

Single-strand conformation polymorphism

Single-strand conformation polymorphism (SSCP) is the simplest and hence the most widespread of the scanning methods. This method relies on the fact that when DNA is made single-stranded by heating and then cooled quickly, disallowing matching with complementary DNA strands, the strands can only hybridize with parts of themselves forming loops, hybridized stretches, etc. producing a complex shape as complementary sequences can be found within a single strand. Clearly the two complementary strands will fold completely differently, but the main point is that on the majority of occasions a single-base change in these strands produces a different folding pattern. This different folding pattern is usually detected by altered migration in an electrophoretic gel. Because of its low detection rate, over a dozen variations of this method have been produced in an attempt to increase the rate of detection.

Denaturing gradient gel electrophoresis

This method has been in use since the mid-1980s. Denaturing gradient gel electrophoresis (DGGE) relies on the fact that a changed base composition due to mutation in a stretch of DNA changes the stability of that fragment, but more importantly, after a mismatch has been formed, the heteroduplexes show decreased stability to denaturing gradients of temperature or denaturing chemicals. To perform this method, the DNA to be examined is hybridized to ensure that heteroduplexes are formed and then the DNA is electrophoresed into a gel in which there is either an increasing temperature or concentration of chemical denaturant. Duplexes with mismatches will melt first; this retards and almost stops their migration into the gel thus signaling the mismatch and hence a mutation. Later, the homoduplex, mutant or wild type, with the lowest melting point will also melt and stop, closely followed by the other. A very high melting point clamp is often added to the end of the duplex to enhance the detection. This method has been favored by diagnostic laboratories due to its near 100% detection rate and lack of radioactivity, but is likely to be superceded by denaturing high-pressure liquid chromatography (DHPLC) (see below).

Heteroduplex

Not only are DNA duplexes containing mismatches less stable than the parent duplexes, but they also contain bends or bulges at the point of the mismatch. This property has been exploited in a method almost as simple as SSCP. However, it also has a low detection rate in its simplest form where the heteroduplexes are simply analyzed for altered mobility on the gel due to the bulge or bend.

Two important modifications have increased this rate and hence the utility of this method. The first of these is conformation-sensitive gel electrophoresis (CSGE) where the bend or bubble is accentuated by separation of molecules in a gel matrix which contains a denaturant, so the duplex is closer to the denaturation and presumably the bend or bulge is bigger. More dramatic has been a principle applied to the duplex in which separation of the duplexes is performed on a high-pressure liquid chromatograph (HPLC). This method, termed denaturing high-pressure liquid chromatography (DHPLC), has been popular despite the expensive equipment required; the popularity is due to the lack of a need for a label and the fact that 96 samples can be loaded in a plate and the operator can walk away after samples are loaded.

Chemical and enzyme cleavage of mismatch

Mismatched bases are more reactive than matched bases allowing the chain to be cleaved at the point of the mismatch by a succession of reactions. Separation of the fragments allows positioning of the mismatch and hence the mutation. Similarly, the mismatches can be recognized by enzymes. One of these, a resolvase, can detect all mismatches and can detect a high percentage of mutations by cleavage. The use of ribonuclease with a ribonucleic acid (RNA) probe has the advantages of cheapness and availability as a commercial kit. The chemical method, while very effective, has the disadvantage of multiple manipulations and toxic chemicals, features which have been eliminated recently.

All of the above methods have their disadvantages and advantages and projects to which they are best suited. However, there is room for improvement (see below).

Protein truncation test

This method essentially takes the DNA of the stretch of DNA to be tested and places it in an expression system to produce protein; the protein is then assessed for size on a protein gel. The DNA may be whole complementary DNA (cDNA) from mRNA or large exons. In any event the code needs to be in-frame with the start codon. The method is essentially an assay for mutations causing stop codons. Thus the protein will stop being translated when the machinery gets to a stop codon, whereas control normal DNA will produce full-length product. This method is popular in those diseases where a large proportion of mutations causing disease produce stop codons. This occurs, for example, in a form of colon cancer and is heavily used in diagnostic laboratories for this disease.

Methods for Detecting Known Mutation (Figure 3)

Sequencing

This method has been used to detect known mutations but it is generally thought to be too time-consuming. However, there are two forms of sequencing that are less time-consuming than usual full Sanger sequencing that are used for this purpose. One is 'sequencing by synthesis' or pyrosequencing for which a machine is

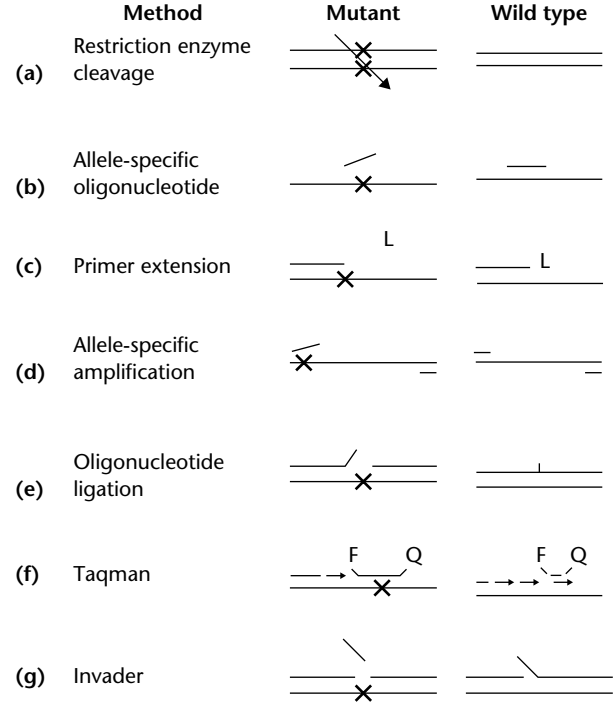

Figure 3 Schematic representation of methods to detect known mutations: (a) restriction enzyme cleavage; (b) allele-specific oligonucleotide; (c) primer extension; (d) allele-specific amplification; (e) oligonucleotide ligation; (f) Taqman. Full lines represent the DNA sequence. Mutations are represented by an X. L and F represent labels, usually fluorescent, and Q, a quencher. Diagonal arrow represent cleavage.

available. The other method, which is also commercially available, is half sequencing by a commercial kit called the Bess scan which essentially is an enzyme method cutting at specific bases.

Restriction enzyme digestion

Restriction enzymes digest DNA at specific sites that can usually be predicted by their sequence, for example CCGG. When a mutation is found to be in such a site or creates such a site, the mutation can clearly be detected by all or nothing cutting of the strands indicating the presence or absence of mutations. The detection of cutting is usually by separation of fragments by electrophoresis. If no restriction site is created or destroyed by the mutation it is possible to engineer a site by use of the PCR process and an oligonucleotide slightly different from the usual specified complementary sequence, but specifying in the PCR product a restriction site that is altered or created by the mutation. (***See*** Restriction Enzymes; Restriction Fragment Length Polymorphism (RFLP).)

Allele-specific oligonucleotide (or sequence-specific nucleotide) hybridization

This is the most popular method for detecting known mutations. The sample to be assayed is usually fixed in duplicate to a solid phase such as a well or membrane. Two complementary oligonucleotides are synthesized to one of the strands, one complementary to the wild-type sequence and one complementary to the mutant sequence. These are labeled in some way and the mutant probe hybridizes to one of the sample duplicates and the wild type to the other. The method depends on the fact that a mismatched base pair in oligonucleotides causes a lower melting temperature. Thus a temperature for the assay, that is, hybridization, is chosen so that where there is a perfect match the sample is labeled and where there is a mismatch labeling is not achieved. If only the mutant probe lights up its sample and not the wild type, the sample is a 'homozygous' mutant; if only the wild type probe lights up, it is homozygous wild type. If both probes light both samples, the sample is heterozygous. This method is used on strips of membrane to test 30 or so cystic fibrosis mutations and the same principle is used to detect SNPs on chips. The main problem is to find conditions in which all samples are assayed under the same conditions. (***See*** Nucleic Acid Hybridization.)

Primer extension or minisequencing

This method relies on synthesizing a primer that ends at the 5′ base before the site to be assayed. If the site to be assayed is a G or T base in mutant and wild type respectively, labeled C or A nucleotide is provided in a DNA polymerase reaction mixture containing the sample, usually PCR amplified. After the reaction, identification of whether the C or A or both have been incorporated allows deduction of the genetic status of the sample. This assay is thought to be one of the superior ones because the reaction is covalent and the reaction conditions are the same regardless of the sample. This method has also been adapted to chips.

Allele-specific amplification

One of the primers of a PCR reaction is placed at the position of the mutation to be detected. This is because the mutation is in the center or at the 3′ end of the primer. The PCR hybridization conditions are chosen so that hybridization occurs if there is a match with the primer and does not occur if there is a mismatch; thus the production of a PCR product is indicative of a match depending on whether mutant or wild-type oligonucleotide is used. In the latter case, the hybridization temperature is not so important as a match and mismatch at the 3′ end will allow and disallow formation of a PCR product respectively.

Oligonucleotide ligation assay

In this assay, a 5′ primer is placed so that its 3′ base is on the position of the base to be assayed. Abutting this is a 3′ probe. Thus depending on whether there is a match or mismatch at the 3′ end of the 5′ primer a ligase will or will not join the two oligonucleotides. Thus variation can be tested at the base. This assay does have the disadvantage that two primers need to be synthesized, but this assay can be developed into an amplification reaction.

Taqman

This assay depends on the 5′ nuclease activity of *Taq* DNA polymerase and the inhibition of the 5′ nuclease activity by a mismatch. Thus a mutation-specific oligonucleotide is placed over the site of the mutation to be analyzed. This oligonucleotide has at its 5′ end a fluorescent dye and at the other end a fluorescence quencher which stops the first dye fluorescing. The mismatch is between these. A further oligonucleotide is placed 5′ to the first and a polymerase reaction is started with *Taq* DNA polymerase. When the growing DNA chain and the enzyme get to the labeled primer, it is displaced and the 5′ nuclease activity digests the labeled probe if there is no mismatch but not if there is a mismatch. Thus a mutation may be detected as the fluorescent dye is released from the quenching influence of the quencher; this does not occur if there is no mismatch.

Invader

This method is essentially novel as it detects the difference between a hybridized and matching and nonmatching probe on the basis of structural difference. A probe is placed with its 3′ end over the mutation to be analyzed and overlaps by one base with a 3′ probe with a 5′ extension arm of unrelated sequence. A mismatch in this position alters the structure so cleavage occurs, whereas none occurs when there is no mismatch. The enzyme involved is cleavase. The method can amplify the signal from genomic DNA without temperature cycling and the signal can be amplified again by a second unrelated invader assay.

Recent advances

In the last several years there have only been a few notable advances. These include the development of DHPLC, the application of the high-throughput mass spectrometer system to primer extension, the application of capillary electrophoresis to the DGGE system, the application of capillary electrophoresis to CSGE and SSCP, and the surprising Invader system, as well as the very simple pyrosequencing system. Pyrosequencing is a recent development, and due to its short read length it is mainly used for genotyping. Its major advantage is that no separation step is involved.

Remaining problems

In the methods for discovery of new mutations it is clear that the expensive separation step is the main barrier to a very cheap test after the PCR step. Labeling, enzymes and sample processing are also significant costs in some of the tests. Clearly it would be ideal if PCR costs could be avoided as this is a considerable expense.

Many of the methods for detecting known mutations involve expensive machines, labels and enzymes. Almost all methods require sample amplification except for the Invader assay, which offers a major saving.

In all methods it needs to be remembered that there are applications for very low throughput, moderate throughput, for example, 96 well plates, and huge throughput, for example, mass spectrometry. These divisions are not always considered by the range of users from laboratories in developing countries to large pharmaceutical companies.

Trends and Expectations

It is quite clear that we are heading for at least two quite dramatic outcomes which are already being planned by those looking well ahead, and which are extremely challenging.

The first of these is the need for healthcare workers to have a hand-held portable apparatus for genetic testing into which can be put saliva, blood, buccal cells or hair roots. This is a tall order; whether it will be achieved is another matter as notionally this would mainly be applicable to high-risk alleles for common disorders and multiple alleles in many (probably thousands at least) of the genes causing single-gene disorders out of the 10% of all human disorders that are inherited. Thus, each machine/surgery will have to have reagents in stock (currently PCR primers) to localize the study of the gene of interest and the region in it. This quest clearly requires miniaturization and perhaps completely new mutation detection technology.

The second is the objective of placing a device in the bloodstream so that harmful mutations, such as those causing cancer, can be detected in circulating DNA as soon as they appear. This clearly needs considerable miniaturization and perhaps electronic mutation detection; however, there is a laboratory in Baltimore in the United States working toward this objective.

See also

Microarrays: Use in Mutation Detection

Further Reading

Antonarakis SE and the Nomenclature Working Group (1998) Recommendations for a nomenclature system for human gene mutations. *Human Mutation* **11**: 1–3.

Cotton RGH (1997) *Mutation Detection*. Oxford, UK: Oxford University Press.

Dracopoli NC, Haines JL, Korf BR, *et al*. (1996) (eds.) *Current Protocols in Human Genetics*. New York, NY: Wiley Liss.

Eng C and Vijg J (1997) Genetic testing: the problems and promise. *Nature Biotechnology* **15**: 422–426.

Kwok P-Y (2000) High-throughput genotyping assay approaches. *Pharmacogenomics* **1**: 1–6.

Kwok P-Y (ed.) (2002) *The Nucleic Acid Protocols Handbook*. Totowa, NJ: Humana Press.

Semino O, Passarino G, Oefner PJ, *et al*. (2000) The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: a Y chromosome perspective. *Science* **290**: 1155–1158.

Syvanen A-C (2001) *Nature Reviews Genetics* **2**: 930–942.

Weaver TA (2000) High-throughput SNP discovery and typing for genome-wide genetic analysis. *New Technologies for Life Sciences: A Trends Guide*, pp. 36–42. Amsterdam: Elsevier Science Limited.