Clinical molecular genetics has a relatively brief but spectacular history. In 1970 the first sequence-specific restriction enzyme was discovered and the first gene (yeast alanine transfer ribonucleic acid) was synthesized in vitro. In 1972 the first recombinant DNA molecules were generated and in 1977 the first human gene (human placental lactogen) was cloned. By 1989, 945 human structural genes had been cloned in addition to over 3400 intergenic DNA segments. The first clinical diagnosis using DNA analysis (prenatal diagnosis of sickle cell disease) was made in 1978 and this approach has now become the mainstay of genetic counselling for over 40 important single gene disorders including cystic fibrosis, haemoglobinopathies, thalassaemias, the fragile X syndrome and muscular dystrophies. DNA analysis is also now used for the diagnosis of infectious diseases and is expected to become widely used in predicting risk and prognosis for cancers and other common disorders of adulthood. Molecular genetic techniques are also being used for treatment with either pure protein products or with gene supplementation. In 1977 the first protein product (somatostatin) was made by genetic engineering, followed by insulin in 1979. This approach has already been utilized for a wide variety of hormones and vaccines and in 1990 the first attempts at human supplementation gene therapy for a single gene disorder (adenosine deaminase deficiency) were performed. Gene therapy has also been attempted for treatment of cancer and further therapeutic applications are likely. This article aims to introduce the basic principles of molecular genetics which are common to its now diverse clinical applications and to illustrate some of the applications in modern clinical genetics practice. For more detail on the basic principles and clinical applications the further reading list can be consulted.

Nucleic Acid Structure and Function

There are an estimated 100,000 human structural genes encoded in the DNA (deoxyribonucleic acid) of every nucleated cell. These genes carry information for the direction of protein synthesis and are regulated so that at any one time only about 1% of the total are active in a particular cell type. With the important exception of genes on the sex chromosomes, the genes are paired and
occupy the same location (locus) on a pair of chromosomes. During normal cell division (mitosis) every gene is normally accurately duplicated so that each cell has an identical genetic composition (genotype). In contrast, in the production of gametes (meiosis) each mature sperm or egg receives only one of each pair of genes and children thus receive one half of their genes from each parent.

DNA consists of a double helix with the two strands bound together by hydrogen bonds between projecting nitrogenous bases which are attached to a deoxyribose-phosphate backbone. The nitrogenous bases are attached to the 1' (one prime) position of deoxyribose and phosphate links the 3' and 5' hydroxyl groups of adjacent deoxyribose residues (-5' deoxyribose 3'-phosphate-5' deoxyribose 3'-phosphate-5' deoxyribose 3'-phosphate-). These 5' to 3' links allow DNA molecules to be orientated and within a double helix. The DNA strands run in opposite directions (i.e. one is 5' to 3' whereas the other is 3' to 5'). There are four types of nitrogenous bases: adenine (A), cytosine (C), guanine (G) and thymine (T). 'A' pairs specifically with 'T' and 'G' with 'C' and so the parallel strands of DNA must be complementary to one another. Thus, for example, if one strand reads ATGA the complementary strand must read TACT. This complementarity of base-pairing is the key to understanding normal DNA replication and approaches to DNA diagnosis and manipulation.

The unit of length of DNA is the base pair (bp) with 1000 bp in a kilobase (kb) and 1 000 000 bp in a megabase (Mb). The total length of DNA in a half set of human chromosomes is 3000 Mb and of this one half represents the structural genes which occur singly or in clusters with intervening DNA (intergenic DNA) which is involved in gene regulation or has no known function. Much of the intergenic DNA consists of repetitive DNA, which may be moderately repetitive with several hundred copies or highly repetitive with many thousands of copies, and may be dispersed or occur in clusters. Three families of highly repetitive DNA are particularly frequent as each accounts for about 4% of the total DNA. One of these families (alphoid repeats) is clustered whereas the other two (Alu and LINE 1 or L1 families) are interspersed throughout the genome. The alphoid repeats occur as short tandem repeats near the centromeres of all chromosomes and are especially abundant in chromosomes 1, 9, 16 and the Y chromosome. The Alu family consists of about 500 000 copies of a 300 bp sequence which share a recognition site for the restriction enzyme AluI (see below). The L1 family consists of 50 000–100 000 copies which are found on average every 8 kb. In contrast to the alphoid family, some members of the Alu and L1 families may be transferred (transposed) to different chromosomal locations and if the new location is a functional gene sequence this transposition can cause genetic disease (e.g. some patients with haemophilia A or neurofibromatosis type I).

Proteins, whether structural components, enzymes, carrier molecules, hormones or receptors, are all composed of a series of amino acids whose sequence determines the form and function of the resulting protein. Within a structural gene each set of three DNA base pairs or triplet codes for an amino acid. As each base in the triplet may be any of the four types this results in 4^3 or 64 possible combinations or codons. This exceeds the number (20) of known amino acids and so, except for methionine and tryptophan, all amino acids are coded by more than one codon (Table 1). Three codons designate termination of protein synthesis and one (AUG which also codes for methionine) acts as a start signal for protein synthesis. By convention in Table 1 the codons are shown in terms of the messenger RNA (ribonucleic acid) which is complementary to the coding strand of the structural gene. Messenger RNA has three bases in common with DNA (A, C and G) but uracil (U) replaces thymine and also pairs with A and ribose replaces deoxyribose.

The first stage in protein synthesis is transcription. The two strands of DNA separate in the area of the gene to be transcribed. One strand (the template strand—this strand is consistent for a given gene but varies from one gene to another) functions as a template and is read 3' to 5' and messenger RNA (mRNA) is formed 5' to 3' with a complementary sequence under the influence of the enzyme RNA polymerase II. Transcription proceeds at about 30 bases per second until a transcription terminator is reached. After some processing and modification the mRNA molecule diffuses to the cytoplasm and the DNA strands reassociate. The next stage of protein synthesis occurs in the cytoplasm and is called translation. Each mRNA molecule becomes attached to one or more ribosomes. As the ribosome moves along the mRNA each codon is recognized by a matching transfer RNA which contributes its amino acid to the end of a new growing protein chain (Fig. 1). Many proteins are not in their
Table 1
The genetic code with codons shown as messenger RNA. The corresponding DNA codons are complementary

<table>
<thead>
<tr>
<th>First base</th>
<th>U</th>
<th>C</th>
<th>A</th>
<th>G</th>
<th>Third base</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>UUU Phe</td>
<td>UCU Ser</td>
<td>UAU Tyr</td>
<td>UGU Cys</td>
<td>U</td>
</tr>
<tr>
<td></td>
<td>UUC Phe</td>
<td>UCC Ser</td>
<td>UAC Tyr</td>
<td>UGC Cys</td>
<td>C</td>
</tr>
<tr>
<td>U</td>
<td>UUA Leu</td>
<td>UCA Ser</td>
<td>UAA STOP</td>
<td>UGA STOP</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>UUG Leu</td>
<td>UCG Ser</td>
<td>UAG STOP</td>
<td>UGG Trp</td>
<td>G</td>
</tr>
<tr>
<td>C</td>
<td>CUU Leu</td>
<td>CCU Pro</td>
<td>CAU His</td>
<td>CGU Arg</td>
<td>U</td>
</tr>
<tr>
<td></td>
<td>CUC Leu</td>
<td>CCC Pro</td>
<td>CAC His</td>
<td>CGC Arg</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>CUA Leu</td>
<td>CCA Pro</td>
<td>CAA Gln</td>
<td>CGA Arg</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>CUG Leu</td>
<td>CCG Pro</td>
<td>CAG Gln</td>
<td>CGG Arg</td>
<td>G</td>
</tr>
<tr>
<td>A</td>
<td>AUA Ile</td>
<td>ACA Thr</td>
<td>AUA Asn</td>
<td>AGU Ser</td>
<td>U</td>
</tr>
<tr>
<td></td>
<td>AUG Met</td>
<td>ACC Thr</td>
<td>AAC Asn</td>
<td>AGC Ser</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>GUU Val</td>
<td>GCU Ala</td>
<td>GAU Asp</td>
<td>GGU Gly</td>
<td>U</td>
</tr>
<tr>
<td></td>
<td>GUC Val</td>
<td>GCC Ala</td>
<td>GAC Asp</td>
<td>GGC Gly</td>
<td>C</td>
</tr>
<tr>
<td>G</td>
<td>GUA Val</td>
<td>GCA Ala</td>
<td>GAA Gln</td>
<td>GGA Gly</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>GUG Val</td>
<td>GCG Ala</td>
<td>GAG Gln</td>
<td>GGG Gly</td>
<td>G</td>
</tr>
</tbody>
</table>

Abbreviations as for amino acids (short code):
- Ala Alanine (A)
- Leu Leucine (L)
- Arg Arginine (R)
- Lys Lysine (K)
- Asn Asparagine (N)
- Met Methionine (M)
- Asp Aspartic acid (D)
- Phe Phenylalanine (F)
- Cys Cysteine (C)
- Pro Proline (P)
- Gin Glutamine (Q)
- Ser Serine (S)
- Glu Glutamic acid (E)
- Thr Threonine (T)
- Gly Glycine (G)
- Trp Tryptophan (W)
- His Histidine (H)
- Tyr Tyrosine (Y)
- Ile Isoleucine (I)
- Val Valine (V)

Other abbreviation:
- STOP chain terminators (X)

*Start codon for protein synthesis.

Final form after ribosomal translation and post-translational processing (e.g. cleavage, formation of disulphide bonds or glycosylation) may be required to allow for protein to achieve its final functional form.

The average protein contains about 300 amino acids which could be coded by 900 base pairs. It was thus a surprise to discover that human genes are much larger than expected. Some of this excess is due to regulatory sequences and to processing of the initial protein product but the majority is due to the presence of intervening sequences (or introns). Nearly all structural genes have been found to consist of alternating protein-coding segments (or exons) and non-protein coding intervening sequences of 50 to over 10,000 bp. The initial messenger RNA transcript is a complete complementary copy of the coding strand of the gene (including exons, intervening sequences and flanking sequences) but prior to its entry to the cytoplasm the segments corresponding to the intervening sequences are removed by splicing. The sequences at the boundaries of each intervening sequence (commonly GT at the beginning and AG at the end) are important for such splicing and mutations in these sequences can result in defective splicing and consequent disease.

All nucleated cells of an individual have an identical genome, yet at any one time in a cell only about 1% of the total is being expressed and the relative pattern of expression varies widely not only for the differentiation of cells and tissues, but also to meet fluctuating demands for synthesis of different proteins in each cell. In addition to the start and chain terminator codons, areas of each gene and of the neighbouring DNA seem to play an important role in regulating transcription and hence synthesis of each protein. Mutations within a gene’s regulatory sequences can occur and may result in no gene product (e.g. some patients with β-thalassaemia), abnormal persistence of a fetal gene product (e.g. hereditary persistence of α-fetoprotein or
haemoglobin F) or anomalous patterns of gene expression (e.g. ectopic expression of creatine kinase).

**Nucleic Acid Pathology**

Normally DNA replication at mitosis or meiosis is completely accurate but errors or mutations can occur. These are divisible into length mutations with gain or loss of DNA and point mutations with alteration of the codon but no change in the amount of DNA.

**Length mutations**

Length mutations include deletions, trinucleotide amplifications, duplications and insertions of DNA.
Deletions can arise from chromosomal breakage, as a result of a parental translocation or inversion (which is itself caused by chromosomal breakage), as a result of slipped mispairing with excision of the single-stranded loop, or by unequal crossing-over. The spontaneous rate of chromosomal breakage is markedly increased by ionizing radiation and by some mutagenic chemicals. Deletions can vary in size from one base pair to many megabases. Very large deletions (over 4 Mb) will be visible at chromosomal analysis but smaller deletions are submicroscopic and need DNA analysis for their diagnosis. Removal of all of a gene directly prevents transcription but smaller deletions of more or less than three (or a multiple of three) base pairs can be equally serious by altering the reading frame of the messenger RNA (frame shift/nonsense mutations, Table 2). Duplications can also affect the reading frame and in some patients the structural gene is disrupted by insertion of an L1 or Alu sequence.

These length mutations are normally stably inherited but unstable length mutations have also been described due to trinucleotide amplifications. For example, mental handicap due to the fragile X syndrome occurs when the normal number of 6-54 CGG repeats in the first exon of this gene is exceeded and the repeat number tends to increase at subsequent cell divisions with a parallel increase in disease severity.

**Point mutations**

Alteration of a single base may (25%) lead to no change in the amino acid coded by that triplet due to overlap of coding, may result in the substitution of a different amino acid (70%, missense mutations) or alteration to a chain terminator (5%, nonsense mutation). These mutations can be described using a standardized nomenclature which is based upon the type of mutation, the amino acid short code (Table 1) and the amino acid position within the protein (Table 3). Most point mutations are spontaneous and unexplained but certain factors particularly mutagenic chemicals can increase the spontaneous mutation rate.

Substitution of T for C is a particularly common point mutation and accounts for 35–50% of all point mutations. In contrast to other types of deamination which can be identified and repaired, deamination of methylated cytosine produces thymidine (with substitution of adenine for guanine on the complementary strand): changes which are not recognized by the DNA repair mechanisms.

**Molecular pathology of single gene disorders**

Determination of the molecular lesion in a single gene disorder is not just of academic interest since it allows the mutation to be tracked within a family in order to provide accurate genetic counselling. As more conditions are studied two generalizations are possible. Lesions can occur at any stage in the protein biosynthetic pathway and most conditions show a diversity of molecular pathology (molecular heterogeneity).

The β-globin gene is currently the best documented gene in respect of molecular pathology. In patients with β-thalassaemia over 100 different molecular defects have now been described and can result in reduced output of the normal gene product (β⁰, i.e. reduction in protein function proportionate to level) or synthesis of an abnormal gene product (β⁺, i.e. reduction in protein function disproportionate to level). In general, mutations affecting transcription (16% of the total), mRNA processing (33%) or translation (33%) result in β⁰-thalassaemia whereas missense mutations, fusion genes and in-frame small deletions or
insertions result in $\beta^+\$-thalassaemia. Both length and point mutations are observed with the latter predominating.

In contrast in conditions such as $\alpha$-thalassaemia and Duchenne muscular dystrophy length mutations, particularly deletions, are far more frequent than point mutations and in sickle cell anaemia all patients have the same point mutation ($E6V$ i.e. the sixth amino acid of the $\beta$-chain changed from glutamic acid to valine, see Table 1).

**DNA polymorphisms**

Most, but not all, mutations in protein-coding DNA result in an altered protein which may or may not cause disease. In contrast mutations in non-protein coding DNA (intervening sequences and intergenic DNA) usually do not cause disease. Many of these mutations which are not associated with disease are found at relatively high frequencies within the population and by definition if one in 50 or more have the variant it is called a DNA polymorphism.

Length or point mutations can be involved in DNA polymorphisms. The length polymorphisms are usually associated with multiple repeats of a dinucleotide or tetranucleotide (microsatellite repeats), or larger repeat unit (commonly 10–15 base pairs, minisatellite repeats). Dinucleotide repeats especially CA and CT are very frequent with an estimated total of 50,000 dispersed throughout the genome. This abundance coupled with the fact that most (about 70%) individuals generate different sized fragments (i.e. heterozygous) from each member of a chromosome pair has meant that this type of polymorphism is very useful for tracking mutant genes within affected families and for gene mapping studies. The minisatellite repeats also show marked variability in the number of repeats and again most (about 70%) individuals are heterozygous. This type of polymorphism is also useful for gene mapping studies but this has been limited by their uneven distribution with many close to the ends of the chromosomes.

Polymorphisms due to point mutations occur every 200–500 bp and as only two alternative sequences (normal and altered) are found most individuals have identical sequences (homozygous) and relatively few (up to 50%) are heterozygous. This type of polymorphism is usually demonstrated by loss (or gain) of a cutting site for a DNA cleavage enzyme (restriction enzyme) and thus they are usually called restriction fragment length polymorphisms (RFLPs).

All of these DNA polymorphisms are normally stably inherited and will cosegregate with neighbouring genes. This cosegregation is known as linkage and allows mutant genes to be tracked within families.

**Nucleic Acid Analysis**

DNA can be extracted from any nucleated tissue and lymphocytes from a 10 ml venous blood sample yield about 300 $\mu$g which is sufficient for many DNA analyses. At postmortem a sample of spleen or liver can be taken into a dry sterile tube or snap-frozen in liquid nitrogen and stored at $-20^\circ C$ pending DNA extraction.

DNA probes are labelled (radioactive or non-radioactive) sections of DNA from tens of base pairs to several kilobases in size which are used to identify fragments with a complementary sequence amongst a mixture of DNA fragments. The probe and target DNA are first rendered single-stranded (by heating or exposure to alkali) and complementary fragments hybridize to form labelled double-stranded DNA fragments which can be visualized. For visualization a large number of labelled hybrid DNA fragments need to be formed and this is achieved by either starting with a relatively large amount of target DNA (Southern analysis) or by selectively amplifying the target sequence(s) from an initial tiny sample of DNA (polymerase chain reaction analysis, PCR analysis). In general, Southern analysis utilizes 5–10 $\mu$g of DNA for each analysis and takes 5–7 days to produce a result. Each Southern analysis needs target DNA which is fragmented with the appropriate restriction enzyme, separated according to fragment size by gel electrophoresis and then transferred to a DNA binding filter. The appropriate DNA probe will then identify the complementary sequence(s) and
the labelled hybrid molecules are identified by autoradiography (for radiolabelled probes, e.g. Fig. 2). Each PCR reaction needs target DNA, a pair of primers which are complementary to DNA sequences flanking the target sequence and an enzyme (Taq polymerase) which directs repeated rounds of localized DNA replication. Theoretically, amplification will increase exponentially to $2^n$ (where $n$ is the number of cycles) and amplification of more than one million fold can be routinely obtained from 30 or so cycles which take a couple of hours in an automated procedure. The amplified DNA segment can then be digested if required with the appropriate restriction enzyme and the resulting fragment(s) separated according to size by gel electrophoresis and visualized directly when stained with ethidium bromide and viewed under ultraviolet light (Fig. 3). Direct visualization in this fashion gives white target DNA bands on a dark background as compared with autoradiography in Southern analysis which gives black target DNA bands on a light background. Each PCR analysis can start with considerably less than 1 µg of target DNA and produces a result in a matter of hours. Its speed and sensitivity are key advantages over Southern analysis but it can only be applied for DNA sequences where flanking primer sequence is available and its extreme sensitivity means that meticulous care to avoid exogenous DNA contamination (and hence a false result) is required. As PCR analysis can use tiny starting quantities of impure genomic DNA it has proved possible to use PCR amplification to provide a DNA analysis on buccal mucosal cells from a mouth rinse, single hairs, single cells (sperm, ova, preimplantation embryo biopsies etc.), fixed pathological specimens and dried blood spots including stored Guthrie neonatal screening cards.

For these experiments DNA is cleaved using restriction enzymes. Restriction enzymes (restriction endonucleases) are widespread in bacteria where they function as a defence mechanism against the incorporation of foreign DNA. More than 400 different restriction enzymes have been described which have over 100 different recognition sites. Each is named after the organism from which it was first isolated, and each will only cleave at a specific DNA sequence—the recognition site which is commonly 4 or 6 bases in length to produce fragments with flush (blunt) or staggered (sticky or cohesive) ends (Fig. 4). The enzyme TaqI will cut DNA at each point where the sequence

![Figure 2. Autoradiograph of Southern DNA analysis with detection of a target sequence in serial dilutions of a DNA sample.](image)

![Figure 3. Direct visualization of a PCR amplified specific DNA fragment in four DNA samples (with size controls in the right-hand lane).](image)

<table>
<thead>
<tr>
<th>Name</th>
<th>Source</th>
<th>Recognition site</th>
<th>Result of cleavage</th>
</tr>
</thead>
<tbody>
<tr>
<td>EcoRI</td>
<td><em>Escherichia coli</em></td>
<td>$\underbrace{G\text{-}A\text{-}A\text{-}T\text{-}T\text{-}C\text{-}}<em>{\downarrow}$  $\underbrace{C\text{-}T\text{-}T\text{-}A\text{-}A\text{-}G\text{-}}</em>{\uparrow}$</td>
<td>$A\text{-}A\text{-}T\text{-}T\text{-}C\text{-}$ $G\text{-}$ $C\text{-}T\text{-}T\text{-}A\text{-}A$</td>
</tr>
<tr>
<td>TaqI</td>
<td><em>Thermus aquaticus</em></td>
<td>$\underbrace{T\text{-}C\text{-}G\text{-}A\text{-}}<em>{\downarrow}$  $\underbrace{A\text{-}G\text{-}C\text{-}T\text{-}}</em>{\uparrow}$</td>
<td>$C\text{-}G\text{-}A\text{-}$ $T\text{-}$ $A\text{-}G\text{-}C$</td>
</tr>
<tr>
<td>SmaI</td>
<td><em>Serratia marcescens</em></td>
<td>$\underbrace{C\text{-}C\text{-}C\text{-}G\text{-}G\text{-}G\text{-}}<em>{\downarrow}$  $\underbrace{G\text{-}G\text{-}G\text{-}C\text{-}C\text{-}C\text{-}}</em>{\uparrow}$</td>
<td>$C\text{-}C\text{-}C\text{-}G\text{-}G\text{-}G$ $G\text{-}G\text{-}G\text{-}C\text{-}C\text{-}C$</td>
</tr>
</tbody>
</table>

*Figure 4. Examples of restriction enzymes with their recognition sites (shown 5' to 3' on the upper DNA strand) and their cleavage products.*
TCGA occurs. Human DNA contains about one million TaqI recognition sites and so cleavage (digestion) with this enzyme yields about one million fragments of DNA. These fragments would be of variable length but each would have the same base order at their staggered ends. As these fragments differ in length they can be separated by gel electrophoresis. These DNA fragments can be visualized by viewing the gel under ultraviolet light and at this stage appear as a smear of overlapping fragments (Fig. 5). Subsequently, a DNA probe would be used to identify specific fragments from this complex mixture.

**Indirect mutant gene tracking**

DNA polymorphisms can be used to follow or track the inheritance of a part of a chromosome through a family. Figure 6 shows a family with X-linked Duchenne muscular dystrophy. The mother is a carrier as she had an affected brother and now has an affected child. The mother is heterozygous for a DNA polymorphism and the affected son has inherited the lower maternal DNA band with the disease. The eldest daughter has inherited the same band from her mother (with the larger DNA fragment from her father) and is thus predicted to be a carrier. Her sister has inherited the upper band from her mother and is thus predicted to be not a carrier. This DNA marker is outside the gene for Duchenne muscular dystrophy (i.e. extragenic) and there is thus a small possibility that at meiosis a recombination could occur between the mutant gene and the DNA polymorphism. This error rate due to recombination increases as more distant polymorphic marker and hence intragenic or very tightly linked extragenic polymorphisms should be used wherever possible.

Indirect mutant gene tracking has the advantage of being applicable in the absence of knowledge of the precise molecular defect but has the disadvantages of potential error if the polymorphic site is at some distance from the mutant gene or if genetic heterogeneity exists with different genes causing a similar clinical picture in different families. Mutant gene tracking also requires DNA samples from several family members in order to distinguish the mutant and normal genes and may not be applicable if few samples are available or if key individuals are homozygous for the polymorphic site. Direct mutation detection circumvents many of these difficulties as demonstration of the molecular lesion confirms the diagnosis and allows direct demonstration of other family members who carry each mutant gene. The disadvantage of this approach is the need to define the mutation(s) for each family.

**Direct mutation detection**

Large length mutations may be visible at cytogenetic analysis but if less than 4 Mb in size then DNA analysis is required. In contrast, even tiny deletions can be revealed by DNA analysis. The single commonest mutation in cystic fibrosis is a three base pair deletion which removes the codon for phenylalanine at position 508 (ΔF508). Figure 7 demonstrates this mutation with normal homozygotes in lanes 2 and 3, heterozygotes in lanes 1, 4 and 6 and a homozygote for ΔF508 in lane 5. The PCR product from the mutant
Figure 7. PCR amplification of a segment of the cystic fibrosis gene containing the common 3 bp deletion mutation. Lanes 2 and 3 homozygous normal, lanes 1, 4 and 6 heterozygous and lane 5 homozygous mutant.

gene is 3 bp smaller and hence migrates further in the gel than the normal fragment (the additional pair of slowly migrating bands in the heterozygotes represent heteroduplexes between the normal and mutant PCR products).

Point mutations can be detected by several approaches including loss or gain of a cutting site for a restriction enzyme and by allele specific oligonucleotide probes. Figure 8 shows an example of a point mutation which can be detected by loss of a cutting site for a restriction enzyme. The mutation G551D in the gene for cystic fibrosis results in loss of a site for the restriction enzyme HincII and hence a larger sized fragment than normal. Lanes 2 and 4 show heterozygotes and lanes 1 and 3 are homozygous mutant and normal respectively.

Allele specific oligonucleotide (ASO) probes are short (17–30 nucleotide) probes which have the complementary sequence to either the normal DNA or mutant DNA sequence at the point of interest. Under appropriate experimental conditions the presence or absence of hybridization with these probes will distinguish normal homozygotes from heterozygotes and homozygous affected individuals.

If the mutation is unknown in a family then a variety of approaches can be used to delineate the molecular lesion. Screening for length mutations can be performed by Southern analysis or PCR analysis with absence of specific hybridization or an altered length of the specific DNA fragment. Screening for point mutations is more difficult and several approaches are under active development. Heteroduplexes between PCR-amplified single-stranded normal and mutant DNA may show altered electrophoretic mobility in special gels (e.g. hydrolink gel electrophoresis and denaturing gradient gel electrophoresis) or the point mismatches may be identified by susceptibility to chemical cleavage (amplification and mismatch detection). The mobility of PCR-amplified

Figure 8. PCR amplification of a segment of the gene for cystic fibrosis which contains the point mutation G551D. This causes loss of a restriction site for HincII and thus lane 1 is homozygous mutant, lanes 2 and 4 are heterozygous and lane 3 is homozygous normal.

Figure 9. DNA sequencing with normal sequence in the left-hand set of lanes (TCGA) and mutant sequence in the right-hand set of lanes showing replacement of a C with an A in a patient with acute intermittent porphyria.
single-stranded DNA depends upon both its size and sequence and single base changes may be detected as mobility shifts (single-stranded conformation polymorphism analysis). Confirmation of the suspected point mutation then requires DNA sequencing. Figure 9 illustrates normal DNA sequence and mutant sequence in a patient with autosomal dominant acute intermittent porphyria. The disease in this patient is due to a point mutation with substitution of A for C in the gene for porphobilinogen deaminase. This information can then be used to identify other family members carrying the mutation by ASOs or gain or loss of an RFLP.

References