BASIC GENETICS FOR THE CLINICAL NEUROLOGIST

M R Placzek, T T Warner

ii5

J Neurol Neurosurg Psychiatry 2002;**73**(Suppl II):ii5–ii11

To the casual observer, the clinical neurologist and molecular geneticist would appear very different species. On closer inspection, however, they actually have a number of similarities: they both use a rather impenetrable language littered with abbreviations, publish profusely without seeming to alter the course of clinical medicine, and make up a small clique regarded as rather esoteric by their peers. In reality, they are both relatively simple creatures who rely on basic sets of rules to work in their specialities. Indeed they have had a productive symbiotic relationship in recent years and the application of molecular biological techniques to clinical neuroscience has had a profound impact on the understanding of the pathophysiology of many neurological diseases. One reason for this is that around one third of recognisable mendelian disease traits demonstrate phenotypic expression in the nervous system.

The purpose of this article is to demystify the basic rules of molecular biology, and allow the clinical neurologist to gain a better understanding of the techniques which have led to the isolation (cloning) of neurological disease genes and the potential uses of this knowledge.

NUCLEIC ACIDS

Deoxyribonucleic acid (DNA) is the macromolecule that stores the blueprint for all the proteins of the body. It is responsible for development and physical appearance, and controls every biological process in the body. DNA is the hereditary material of all organisms with the exception of some viruses which use ribonucleic acid (RNA) and prions, that apparently only contain protein. However, its simple composition meant its great importance in biology was overlooked for many years. It was not until 1952 that Alfred Hershey and Martha Chase, and their experiments on bacteriophage, finally proved that DNA and not the more complex protein was the hereditary material.

DNA is made of two anti-parallel helical polynucleotide chains wrapped around each other and held together with hydrogen bonds to form a double helix. The backbone of the helices are made from alternating phosphate and deoxyribose sugars. Each sugar molecule is joined to one of four nitrogenous bases, adenine, cytosine, guanine or thymine. These bases face into the centre of the helix and hydrogen bond with their partner on the opposite strand. Adenine can only form hydrogen bonds with thymine, and guanine is only able to hydrogen bond with cytosine. The entire genetic code relies upon these four bases and their specificity of binding. The direction of the helices are described as either 5' to 3' or 3' to 5' depending on which carbon atom in the deoxyribose sugar the chain begins and ends with.

GENES AND THE HUMAN GENOME

DNA therefore is the basic substrate for heredity and is divided into functional units known as genes. Recent estimates suggest there are approximately 30–50 000 genes in the human genome. A gene is a sequence of bases that determine the order of monomers—that is, amino acids in a polypeptide, or nucleotides in a nucleic acid molecule. DNA is organised into a three letter code. Each set of three is called a codon, and with four possible bases in each position, there are 64 different combinations, which are more than enough for 21 amino acids.

There is approximately 2 metres of DNA in each of our cells packed into a structure only 0.0002 cm across. (If the DNA from all of our cells were removed and placed end to end the strand would reach to the sun and back several hundred times). This is achieved by packing the DNA into chromosomes.

Humans have 23 pairs of chromosomes in the majority of cells in their body. One of each pair is inherited from each parent, and most cells have diploid status, in that they contain homologous pairs of each chromosome. One of these pairs is the sex chromosomes (XY in males and XX in females) and the remainder are called autosomes. Genes are arranged in linear order on the chromosome, each having a specific position or locus. With the exception of the sex chromosomes each pair of chromosomes carry the same genes as its partner. For any particular character coded for by a gene there may be a number of different forms, which are called alleles. If an individual carries

See end of article for authors' affiliations

Correspondence to:
Dr TT Warner, Department of
Clinical Neurosciences, Royal
Free & University College
Medical School, Rowland Hill
Street, London NW3 2PF, UK;
twarner@rfc.ucl.ac.uk

two different alleles for a particular characteristic, this is termed heterozygous; if both alleles are the same this is called homozygous.

The human genome is a term used to describe all the DNA in human cells and actually comprises two genomes. Firstly there is the nuclear genome which accounts for 3300 Mb of the total genetic makeup of the cell. Secondly there is the much smaller mitochondrial genome. Mitochondria are cytoplasmic organelles that generate energy in the form of ATP by oxidative phosphorylation. They contain 2–15 copies of mitochondrial DNA which comprises a 16539 base pair circle of double stranded DNA. This contains 37 genes specifying 13 polypeptides, 22 transfer RNAs (tRNAs), and two ribosomal RNAs.

CONTROL OF GENES

Controlled synthesis of a gene product is initiated by its promoter. The promoter is the collective name for a number of short sequences, called cis-acting elements, that are usually clustered upstream of the coding sequence of the gene. Transcription factors bind to these sequences and allow the attachment of RNA polymerase. Examples of common cis-acting elements are the GC, TATA, and CAAT boxes. The remainder of the gene can be divided into coding and non-coding regions called exons and introns, respectively. The purpose of introns is not known. However, their presence in all eukaryotes and in most genes means there is either no selective disadvantage to having them, or they have a positive function that is not yet clear. The average exon is between 150–250 nucleotides in length. Genes can have a very large number of exons such as the dystrophin gene (responsible for Duchenne muscular dystrophy) with 79 exons or just one. It is believed that up to 97% of the human genome consists of non-coding sequence.

DNA TO PROTEIN

The process of converting the genetic information in the DNA sequence into a protein product is schematically represented in fig 1. Synthesis of a protein begins with an appropriate signalling molecule binding to the promoter of the gene. This initiates a process called transcription. Transcription creates a single stranded RNA copy of the gene. RNA, like DNA, is composed of a linear sequence of nucleotides, but the sugar-phosphate backbone consists of ribose sugar instead of deoxyribose and the base thymine is replaced by a very similar base uracil. Before the RNA molecule leaves the nucleus it undergoes a process known as splicing to create a messenger RNA molecule, mRNA. Splicing removes all the intron sequences from the RNA, leaving a small molecule containing all the information of the original gene. The expression of the gene can also be modified at this level through a mechanism known as alternative splicing. This is where different forms of mRNA and hence protein are produced by altering which sequences are cut from the original transcribed RNA. An example of alternative splicing is in the production of different isoforms of the D2 dopamine receptor. Once spliced the mRNA can then move into the cytoplasm to direct protein synthesis.

There are two other important molecules that are required for protein synthesis. The first of these are ribosomes. Ribosomes are found free in the cytoplasm and are attached to the surface of the rough endoplasmic reticulum. Once the mRNA has entered the cytoplasm these molecules bind to it and read along the sequence until an AUG is reached. These three bases mark the beginning of translation, the process of

Figure 1 A schematic representation of transcription and translation. The shaded area represents where the gene is transcribed to produce mRNA and spliced. The unshaded area represents cytoplasm where translation and modification of the protein occurs.

reading the sequence and turning it into the appropriate protein molecule. The nucleic acid bases are read in sets of three called codons, where AUG is the start signal and also sets the frame for reading the remaining codons. The second molecule required for protein synthesis is tRNA. For every codon there is a tRNA with a domain of complementary sequence that will selectively bind to it, an anti-codon. Each codon codes for a specific amino acid, and the tRNA with the matching anti-codon is responsible for bringing it to the ribosome where it will bind to the amino acid from the tRNA molecule attached to the previous codon. Any one of three stop or non-sense codons (UAA, UAG, or UGA) signals the termination of protein synthesis. Further information is contained within the protein sequence itself and signal peptides can direct the newly formed protein to particular cell organelles for post-translational modifications (for example, glycosylation, addition of metal ions or other polypeptides), and allow its insertion into membranes.

MITOSIS AND MEIOSIS

An adult human is made up of approximately 50 trillion cells but began life as a single cell. Our bodies are also continuously required to generate new cells for growth and repair. All of these cell divisions take place by mitosis. Mitosis is the simple division of a parent cell into two identical daughter cells. This process is divided into four steps; prophase, metaphase, anaphase, and telophase (fig 2). In this process the chromosomes are condensed and pulled to the equatorial plane at metaphase. The centromere splits in anaphase and the two chromatids of each chromosome are pulled to opposite poles. In telophase the chromosomes reach the poles and start to decondense. The nuclear membrane reforms and the cytoplasm starts to divide yielding the two daughter cells.

Meiosis, however, is a specialised form of cell division and only takes place in cells that are destined to become gametes. There are a number of important differences between mitosis and meiosis. Firstly, in meiosis the daughter cells produced are not identical to the original parent cell. Meiosis consists of two divisions, but the cellular DNA is only replicated once. This

Figure 2 Stages involved in cell division by mitosis and meiosis.

means the daughter cells produced are haploid. Secondly, during prophase I an important event called crossover occurs. Visible manifestations of this event, called chiasmata, can be seen during metaphase I. Crossover is recombination between two non-sister chromatids; there is a precise break, swap, and repair thus exchanging genetic material. This is a very important process that creates genetic diversity within the gametes and therefore the next generation. This phenomenon is also of great significance in the study of genetics as it is the fundamental concept behind genetic mapping and linkage analysis.

PATTERNS OF INHERITANCE

Since Gregor Mendel laid down the first (principle of independent segregation) and second (principle of independent assortment) laws of inheritance it has become clear that there are many cases where a single gene is both necessary and sufficient to express a character. These characters are called mendelian.

A mendelian character is considered dominant if it manifests itself in a heterozygous individual. If the character is masked it is considered recessive. In other words, a dominant allele exerts its effect despite the presence of a corresponding normal allele on the homologous chromosome, whereas in autosomal recessive inheritance, both alleles must be abnormal for the disease trait to be expressed.

There are four common patterns of inheritance seen in neurological disease and these are represented by pedigrees in fig 3. Two other mendelian inheritance patterns (X linked dominant and Y linked) are very rare and are not considered further.

Autosomal dominant inheritance is generally easy to identify in families (fig 3A). An affected individual will always have at least one affected parent, and, depending on whether they are heterozygous or homozygous for the condition, any child will have a 50% or 100% chance of being affected, respectively. Neurological examples of autosomal dominant

conditions include Huntington's and von Hippel Lindau disease, neurofibromatosis types I and II, and more common forms of hereditary spastic paraplegia.

Autosomal recessive conditions are harder to identify in families, as many of the members will be unaffected carriers (fig 3B). Usually the parents of an affected individual are asymptomatic carriers and each subsequent child has a 25% chance of having the condition. Autosomal recessive conditions are often metabolic and examples include Wilson's and Tay Sach's disease, Krabbe and metachromatic leucodystrophy, and the most common form of inherited ataxia, Friedreich's ataxia.

Conditions linked to the sex chromosomes have a different pattern of inheritance (fig 3C). This is because in males the sex chromosomes do not carry the same genes. If a recessive allele is present on the X chromosome in a domain that is not present on the Y, then the male will be affected. This is called X linked recessive inheritance. Females are usually asymptomatic carriers and will have a 50% chance of passing the gene to their children who will either be affected if male or carriers if female. Male to male transmission excludes this pattern. Classical X linked recessive neurological disorders include Duchenne and Becker muscular dystrophy and fragile X syndrome.

Mitochondria are inherited matrilineally (fig 3D). This is because the female gamete is the only one that contributes mitochondria to the fertilised egg. Thus traits resulting from mtDNA mutations show a specific segregation pattern in a pedigree. MtDNA disorders can have tremendous clinical variability within any family because not all mitochondrial genomes may carry the mutation, a condition referred to as heteroplasmy. MtDNA disorders commonly have a neurological phenotype and include Leber's hereditary optic neuropathy, and various encephalomyopathies such as myoclonic epilepsy with ragged red fibres (MERRF).

Establishing a particular mendelian pattern of inheritance enables accurate estimation of recurrence risks but does not predict clinical severity of the disease. Different individuals

Figure 3 Pedigrees demonstrating the three most common mendelian traits (A–C) and mitochondrial inheritance (D). Affected subjects are represented by solid symbols, unaffected as open symbols, and female carriers of an X linked trait as half shaded (squares are males, circles females).

with the same genetic disease can be clinically affected to dramatically different extents, a phenomenon known as variability of expression. Penetrance refers to the fact that not all individuals who inherit the mutant gene develop the clinical phenotype. This is often age related. In autosomal dominant primary torsion dystonia caused by mutation of the *DYT1* gene, for example, although most individuals who develop symptom do so by the age of 30 years, only about 30–40% of people with the disease gene actually develop dystonia at all. Two other features relating to the phenotypic manifestation of disease need also be described: anticipation, which describes the appearance of more severe clinical phenotype in successive generations; and imprinting, which refers to variation in phenotype depending on the parent of origin of the mutant allele.

BASIC TOOLS OF MOLECULAR BIOLOGY

The success of molecular biology in identifying disease genes has come from the ability to manipulate DNA. Nowadays there are a myriad of molecular techniques. However, they all rely to a greater or lesser extent on a few basic techniques, namely the ability to cut, copy, and deduce the sequence of DNA, to identify specific sequences of DNA by hybridisation to labelled DNA probes, and to introduce DNA into vectors.

The identification of restriction endonucleases in bacteria that recognise and cut a specific DNA sequence (usually 4–8 bp in length), enabled the manipulation of DNA. The cut sequences are left with so called “sticky ends” which means that they can join with other sequences that have been cut with the same enzyme under appropriate conditions with a DNA ligase. This feature is critical to allow joining of a chosen DNA segment into the DNA backbone of a vector. The vector can then be transferred into a host cell to allow propagation or cloning of many copies of the original piece of DNA. Vectors for cloning usually derive from bacteria or bacteriophages. They can be used to express a gene in host cells, to perform DNA sequencing, and large insert vectors (such as yeast artificial chromosomes) can be used to make a library of an entire genome.

A library contains a collection of many different clones. The type of library most frequently used in gene identification is one for expressed genes only, known as a cDNA library. Total mRNA from an organism or specific tissue is extracted and is reverse transcribed (using a reverse transcriptase) to generate the complementary DNA (cDNA) strand. The original RNA is then degraded and replaced with a second DNA strand. The final double stranded cDNA is cloned into a vector to make the library, which represents all mRNAs from the original organism or tissue. These libraries can then be probed with a specific short sequence of DNA or synthetic oligonucleotide to identify a clone of interest.

Another important technique is the use of gel electrophoresis to separate fragments of DNA, RNA or proteins by their size. The polar molecules migrate through agarose or acrylamide gels according to their size when a voltage is placed across the gel. The separated product can be blotted onto membranes and then probed with labelled DNA or antibody to detect specific targets. Southern blotting is used to detect DNA fragments, Northern blots for RNA, and Western blotting for proteins.

Polymerase chain reaction (PCR) is one of the first steps employed when investigating a specific gene, and is the method used to amplify a specific DNA sequence from the entire genome. The only knowledge required is of a small amount of sequence surrounding the gene to be studied. This provides the information to construct short single stranded DNA molecules (primers) that are used as binding domains for DNA polymerase. PCR specifically amplifies the gene of interest by thermocycling the genomic DNA mixed with thermostable DNA polymerase, primers and dNTPs, the building blocks of DNA. The PCR product can then be separated, identified, and isolated by gel electrophoresis.

The final important technique is DNA sequencing which determines the exact linear sequence of nucleotides in a DNA fragment. Sequencing, as the name suggests, is a technique that can be used to determine the sequence of bases in a length of DNA. The process is similar to PCR; however, there is one additional ingredient added to the thermocycling

reaction. A proportion of the dNTPs are dideoxyribonucleotides (ddNTPs). These prevent further elongation of the DNA sequence. If appropriate concentrations of each of the four dNTPs and ddNTPs are included, a series of incompletely elongated chains is produced, each differing by one base pair. When separated on a polyacrylamide gel with four lanes (each one representing a different ddNTP) the sequence can be determined. Nowadays this process has been simplified by use of nucleotides labelled with different coloured fluorescent dyes, which can be distinguished when run in the same lane of a gel on an automated sequencer.

IDENTIFYING DISEASE GENES

Using molecular biological techniques based on those described in the previous section, numerous neurological disease genes have been mapped, cloned, and characterised over the last 20 years. The majority of these are genes with simple mendelian inheritance patterns, where the disease phenotype can be correlated to a particular genotype. Initially genes were identified because there was pre-existing knowledge of the basic biochemical defect without knowing the chromosomal map position. This method was called functional cloning and examples of disease genes isolated in this way include phenylketonuria and sickle cell anaemia. However, for most single gene disorders such biological clues as to the gene function do not exist, which led to the development of positional cloning, or isolation of the gene purely on the basis of their chromosomal map position.

FROM AFFECTED FAMILIES TO DISEASE GENE (POSITIONAL CLONING)

Gene identification by positional cloning starts with the collection of families in which the disease is segregating. Genetic linkage analysis is then used to localise the disease gene to a particular chromosome between two defined markers. The candidate interval is then refined further by study of other genetic markers in this interval. This region of DNA is then cloned usually in the form of a series of overlapping fragments inserted in vectors (known as a contig). Finally, candidate genes are isolated from within the cloned contig and mutations in these genes sought in affected individuals.

The basic premise underlying linkage analysis relies on the fact that if two genetic loci occur close to each other on a chromosome, they do not segregate independently during meiosis, and the degree to which this happens is a reflection of their physical proximity. As stated above, during meiosis recombination can occur and is the process whereby genetic information is exchanged at chiasmata when homologous chromosomes pair. For loci close together, the chance of recombination between them is small, and they will tend to be inherited together. Loci further apart are more likely to have a recombination event occurring between them. The probability of a recombination event occurring between two loci during meiosis is termed the recombination fraction (θ) which is taken as a measure of the genetic distance between loci. The recombination fraction can vary from $\theta = 0$ for loci right next to each other to $\theta = 0.5$ for loci far apart (or on different chromosomes). The function which relates genetic to physical distance is called a mapping function, and translates recombination frequency (in per cent) into mapping distance measured in centimorgans (cM).

To map a disease gene, therefore, the segregation of the disease locus and a known genetic marker through one large family, or a number of pedigrees, is analysed to determine if

the loci are linked and then the level of recombination between them assessed. Using the likelihood method, LOD (likelihood of odds ratio) scores are generated over a range of θ . An LOD score is defined as $\log_{10}$ of the odds ratio for cosegregation of the loci versus independent assortment. The value of θ at which the LOD score is largest represents the best estimate of genetic distance between the two loci under study (referred to as two point LOD scores). Linkage is considered significant when the LOD score is > 3.0 , corresponding to the odds for linkage of at least 1000:1. In practice this correlates with a probability for linkage of 20:1, due to the prior probability that two autosomal loci are linked because they must be on one of the 22 pairs of chromosomes. An LOD score of < -2.0 is taken as rejection of linkage.

To undertake linkage analysis there are two basic requirements: (1) families that segregate the disease as a mendelian trait; and (2) polymorphic genetic markers distributed throughout the human genome. Polymorphisms are natural variations in the human genome and enable distinction between the two chromosomes of a homologous pair. The more polymorphic a locus, the more suitable it is for linkage analysis, as it is more likely that a given meiosis will be informative for that marker. Nowadays, microsatellite markers are almost exclusively used for genetic mapping. These consist of repeated dinucleotides, such as (CA)_n. The polymorphism results from a difference in the number of repeat units. They can be readily genotyped by PCR and are localised throughout the genome at small genetic intervals. Microsatellites are available in the form of panels of markers which can be amplified by PCR in a large number of samples. The resulting PCR products are then electrophoresed and analysed on automated DNA sequencers which use fluorescent labelling to genotype each sample. The genotypes can then be used to generate LOD scores using various computer programs. Figure 4 shows examples of an autosomal dominant pedigree with

Figure 4 Pedigree of family with autosomal dominant hereditary spastic paraplegia. Marker genotypes are shown beneath each individual and shaded boxes indicate chromosome carrying the disease allele. Recombination events indicated by arrows which allow definition of the candidate interval. The table shows two point LOD scores generated for these markers demonstrating significant linkage of the disease gene to this region (scores > 3.0). Reproduced from Heinzl et al. *J Med Genet* 1998;**35**:89–93, with permission.

Table 1 Main classes of mutation

Mutation type	
Chromosomal abnormalities	Alterations in number or structure of chromosomes
Deletions	Deletions of DNA from 1 base to megabase
Insertions (including duplications)	Insertion of DNA
Single base substitutions	
Missense mutations	One amino acid replaced with another in gene product
Nonsense mutations	Amino acid codon replaced with a Stop codon
Splice site mutations	Creation or destruction of signals for exon-intron splicing
Frame shifts	Disruption of Open reading frame, can be produced by deletion, insertions or splicing errors
Dynamic mutations	Repeat sequences that can change size on transmission—for example, triplet repeat disorders

hereditary spastic paraplegia, the genotypes for a number of markers on chromosome 2, and the resulting LOD scores generated, indicating evidence for linkage to this region.

Linkage analysis in autosomal recessive disorders can be more difficult as affected individuals are usually confined to a single sibship, and families containing three or more patients are rare. Alternative strategies have been devised for these situations including use of inbred families for homozygosity mapping. Alternatively screening the genome for regions of identity-by-descent can be performed by pooling samples from affected individuals and comparing the distribution of alleles in the pooled sample to a control pool from unaffected family members.

The ideal situation is to have one single family large enough to allow significant LOD scores to be generated. Frequently this is not the case, and linkage analysis is performed on a number of families and LOD scores are added. This is based on the assumption (sometimes erroneous) that the disorder is genetically homogeneous—that is, the same gene is responsible for the disorder in all the families.

Once two point analysis has detected linkage between a marker and a disease locus, multipoint analysis can be performed to determine which two loci most closely flank the disease gene. The genetic markers that flank the candidate region delineate the genetic distance. One centimorgan genetic distance is roughly equivalent to one million base pairs (1 Mb) of physical distance. This candidate region must then be physically linked by cloning the intervening DNA segment into a series of overlapping vectors (such as yeast artificial chromosomes) to form a contig, or by long range mapping using a technique called pulse field gel electrophoresis. This region is then searched for candidate genes by a number of techniques which either identify familial sequences found at the 5' end of genes or evolutionarily conserved sequences, or isolation of expressed coding sequences (so called exon trapping).

Proof that a candidate gene is actually the disease gene requires demonstration that a mutation found in affected individuals is not present in unaffected ones, and also should not occur as a benign polymorphism in the normal population. In addition, mutations should be able to cause a pathogenic effect—for example, disrupt the coding sequence of the gene. This process of positional cloning can be accelerated by initially examining candidate genes already known to map to the critical region. This method, sometimes referred to as the positional candidate approach, is becoming increasingly used with the completion of the Human Genome Project and the availability of cDNA libraries.

TYPES OF MUTATION

The human genome is not a static entity and is subject to a variety of different types of heritable change (mutation). The

main types of mutation are described in table 1. In general mutations can produce their effect in one of two ways.

The gene product has reduced or no function (loss of function mutation)—Loss of function mutations usually produce recessive phenotypes. For most gene products the precise quantity is not critical and a reduction by half can be coped with. Thus most inborn errors of metabolism are recessive. However, for some gene products 50% of the normal level is not sufficient for normal function, and haplo-insufficiency produces an abnormal phenotype which is therefore inherited in a dominant manner. An example of this is seen in dopa responsive dystonia where a dominant mutation in the gene for GTP-cyclohydrolase I, the rate limiting step in producing tetrahydrobiopterin (key co-factor in monoamine synthesis) produces the phenotype. Sometimes, the mutant non-functional polypeptide interferes with the function of the normal protein in heterozygous individuals leading to a dominant negative effect. There has been debate as to whether this truly represents a loss of function or is due to gain of function, as described below.

The gene product gains a new abnormal function (gain of function mutation)—Gain of function mutations usually cause dominant phenotypes, because the presence of the normal allele does not prevent the mutant allele from behaving abnormally. Sometimes the gain of function involves the product having a novel function—for example, if the protein product contains expanded polyglutamine tracts which appear to cause it to form aggregates (see below).

TRINUCLEOTIDE REPEAT DISEASES

Unstable trinucleotide repeats were an entirely novel disease mechanism when first discovered in 1991. These subsequently were found to be an important cause of neurogenetic disease. A hallmark of these conditions is anticipation, with earlier onset and more severe phenotype in successive generations, which appears to correlate with increasing repeat size during gamete formation. Two different classes of expansion have been identified:

- ▶ Highly expanded repeats outside the gene coding region causing loss of function by reducing or abolishing transcription. Examples in this group include fragile X syndrome, Friedreich's ataxia, and myotonic dystrophy.
- ▶ CAG repeats within the coding region that encode polyglutamine tracts which lead to protein aggregation and cell death. The typical example in this group is Huntington's disease, but it also includes the dominant spinocerebellar ataxias and spinobulbar muscular atrophy (Kennedy's disease).

MUTATION DETECTION

Once a disease gene is isolated, detection of mutations can be an important diagnostic tool. The ease with which this can be done depends on the type of mutation (for example, deletion versus point mutation) and the structure and size of the gene to be examined. Large DNA rearrangements such as the deletions seen in the dystrophin gene in Duchenne muscular dystrophy can be detected by restriction digest and Southern blotting to identify size differences. As most of the deletions occur in specific regions of this large gene, a PCR based technique has been devised which simultaneously amplifies the different regions of interest with products that differ in size. The absence of one of these products implies that there is a deletion involving the region where one of the primers would have annealed.

Trinucleotide CAG repeat disorders can be easily tested for with a PCR assay using primers that flank the repeat sequence. For the conditions with a very large expansion, such as myotonic dystrophy, Southern blotting analysis using a probe from a region flanking the trinucleotide is used to identify a larger size fragment containing the expanded repeat.

Point mutations within a gene can be more difficult to detect, particularly in large genes. There are a number of methods that rely on the fact that fragments of gene amplified by PCR containing a point mutation migrate with different mobility on a gel compared to the wild type fragment. The techniques use different conditions and include heteroduplex analysis, single strand conformational polymorphism (SSCP) analysis, and denaturing gradient gel electrophoresis (DGGE). For mutations that lead to premature termination of translation and a smaller mutant protein, the mutation detection system called protein truncation test can be used.

Many laboratories now rely on large scale DNA sequencing to detect mutations rather than on the methods above, and this is likely to be increasingly used in neurogenetic service testing in the future.

CLINICAL APPLICATION OF DISEASE GENE MAPPING AND CLONING

Knowledge of the map position of a gene can have clinical use, even if the causal gene itself has not been isolated. A good example of this is Huntington's disease. Before the discovery of an expanded CAG repeat within the first exon of the gene encoding huntingtin, it was known that almost all cases were linked to chromosome 4p. Predictive and prenatal tests were available using markers closely linked to the Huntington's disease locus providing DNA was available from other family members, including at least one affected. Genotyping of markers would then allow identification of which marker allele was segregating with the disease. The limitation of this

type of test was the small, but finite, possibility of recombination occurring between the markers used and the disease locus, which meant that 100% accuracy could not be guaranteed. Linked markers can still be useful even if the disease gene has been identified. If the gene is large with many exons and there are numerous point mutations known to cause the phenotype, it is impossible to screen for all of these. The use of markers within the gene can be used again to determine if the disease chromosome(s) has been inherited.

Once a gene has been isolated and mutation(s) identified, direct testing for mutations can be performed. The ideal situation is where a single mutation type is readily detectable and is responsible for the majority of cases of the disease. Examples of this include the CAG repeat expansion for Huntington's disease, the GAG deletion in the DYT1 gene causing childhood onset autosomal dominant primary torsion dystonia, and the 1.5 Mb duplication including the PMP22 gene in 70% of patients with HMSN1A.

CONCLUSIONS

The advances in molecular biology in recent years have led to the isolation of numerous genes responsible for neurogenetic conditions and characterisation of the pathogenic mutations. This has had a number of practical benefits as identification of a specific molecular lesion in a patient establishes a definite diagnosis, prevents further investigations, and may provide prognostic information. It also makes it possible to diagnose or exclude a diagnosis with a simple blood test in other family members at risk of developing the condition, and in some cases allows prenatal testing to be offered. Finally, the identification of disease genes is an important step in establishing the pathogenetic mechanisms underlying many neurological diseases.

ACKNOWLEDGEMENTS

The authors would like to thank Dr Sarah Tabrizi and Dr Christos Proukakis for helpful comments during the preparation of this manuscript.

Authors' affiliations

M R Placzek, T T Warner, Department of Clinical Neurosciences, Royal Free & University College Medical School, London, UK

KEY REFERENCES

- 1 **Strachan T**, Read AP. *Human molecular genetics*. Oxford: Bios Scientific, 1999.
- 2 **Rosenberg R**, Prusiner S, DiMauro S, Barchi R, eds *The molecular genetic basis of neurological disease*. Butterworth-Heinemann, 1997.
- 3 **A comprehensive review of neurogenetic disorders.**
- 3 **Mendelian Inheritance in Man (OMIM)**. URL: <http://www.ncbi.nlm.nih.gov/omim/>
- **A useful updated website for both genetic and clinical information.**