10

# Myotonic Dystrophy and Gene Mapping on Human Chromosome 19

J.D. BROOK, D.J. SHAW AND A.L. MEREDITH

*Section of Medical Genetics, University of Wales College of Medicine, Heath Park, Cardiff CF4 4XN, UK*

## Introduction

### BACKGROUND GENETICS

The laws of inheritance, as discovered and described by Mendel in 1865, are as applicable to humans as they are to the original subject of Mendel's study, the garden pea. The ease with which the inheritance of genetic traits can be studied in these two species, however, differs considerably.

In humans, genes are arranged along 22 pairs of autosomes and two sex chromosomes, all of which can be distinguished cytogenetically (*see Figure 1*). Any locus (the site of a gene on a chromosome) may be occupied by one of several different forms of the gene which are called alleles. Two loci are said to be linked when they are carried within a measurable distance of each other on the same chromosome. The closer they are together, the closer the linkage. Alleles at loci which are some distance from each other will often be separated by recombination at meiosis (*see Figure 2*). In most cases loci are so far apart that it is not possible to demonstrate linkage directly, even though they are known to be located on the same chromosome. Such loci are described as syntenic (Renwick, 1969).

Distances between genes are measured in centimorgans (cM), where one cM represents 1% recombination between two loci. These distances can also be expressed in terms of a recombination fraction ($\theta$). Thus if 5% of the progeny from matings informative at two loci are of the recombinant type,

Abbreviations: ApoCII, apolipoprotein CII; ApoE, apolipoprotein E; 8-AZA, 8–azaguanine; bp, base pairs; BUDR, 5-bromodeoxyuridine; C3, complement component 3; cM, centimorgans; DM, myotonic dystrophy; FACS, fluorescence-activated cell sorting; FTH, ferritin heavy chain; FTL, ferritin light chain; GPI, glucose phosphate isomerase; HAT, hypoxanthine, aminopterin and thymidine; HD, Huntington's disease; HLA, human leucocyte antigen; HPRT, hypoxanthine phosphoribosyl transferase; kb, kilobases; LDL, low-density lipoprotein; Mb, megabase ($10^6$ base pairs); PEG, polyethylene glycol; RFLP, restriction fragment length polymorphism.

*Biotechnology and Genetic Engineering Reviews* — Vol. 3, September 1985
0264–8725/85/03/311–37$10.00 + $0.00 

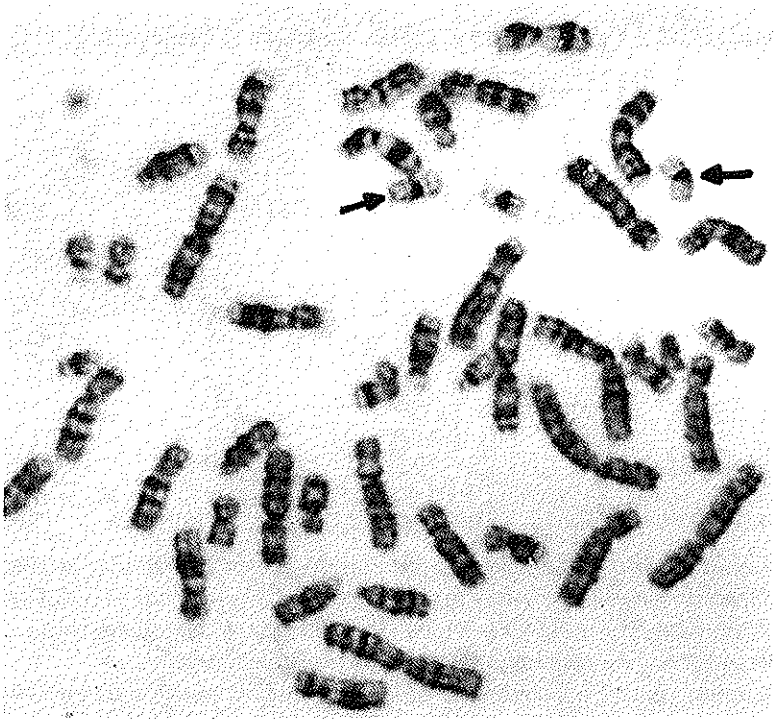

**Figure 1**. Human chromosomes. A G-banded mitotic metaphase chromosome spread, produced by treatment with trypsin followed by Giemsa staining (Seabright, 1971). Bands are numbered according to a standard nomenclature (*see* ISCN, 1981). Chromosomes 19 are arrowed.

these loci are said to have a recombination fraction of $\theta = 0{\cdot}05$. Over short distances the recombination fraction is equivalent to the distance in cM between the two genes. This relationship does not hold, however, for larger distances because of the probability of double cross-overs. The maximum recombination fraction between two genes, therefore, is 0·5, equivalent to random assortment, whereas distances in cM are additive along the length of the chromosome.

Because of the structure and small size of most human families it is sometimes difficult to demonstrate whether or not two loci are linked and, if linked, to establish precisely the distance between them. The detection and measurement of linkage in man is the subject of considerable mathematical method (*see* pages 332–335).

Until recently a shortage of useful polymorphic markers has limited the extent to which the inheritance of genetic diseases could be studied. The most useful polymorphic markers available include red cell antigens (Race and Sanger, 1975), isoenzymes (Harris, 1980) and human leucocyte antigen (HLA) types (Bodmer, 1980). No systematic method for mapping human inherited disease existed, due mainly to the lack of marker loci (Botstein *et*

*al.*, 1980), and hence only 2–3% of Mendelian disorders had been mapped.

Botstein *et al.* (1980) have proposed that, using recombinant DNA techniques, it should be possible to construct a genetic linkage map of the entire human genome. Fragments of different sizes, generated when human DNA is digested by restriction endonucleases, separated by electrophoresis, and blotted on to nitrocellulose filters (Southern, 1975), can be identified using random single-copy DNA sequences which have been radiolabelled (Wyman and White, 1980; *see also* Chapter 9 of this volume). Differences in the lengths of particular restriction fragments between individuals are termed restriction fragment length polymorphisms (RFLPs). These are inherited as simple codominant Mendelian markers and can be used in the same way as other polymorphic markers to study the inheritance of genetic diseases (*see Figure 3*). Botstein *et al.* (1980) propose that 150 highly polymorphic RFLPs, evenly distributed throughout the genome, would be sufficient to map every human gene.

This approach to gene mapping has already produced one notable success, that of Gusella *et al.* (1983). In order to locate the previously unassigned gene for Huntington's disease (HD), this group found that one of the first few probes they used, G8, showed close linkage to the HD gene and mapped to chromosome 4. Although such rapid success is generally unlikely — Gusella

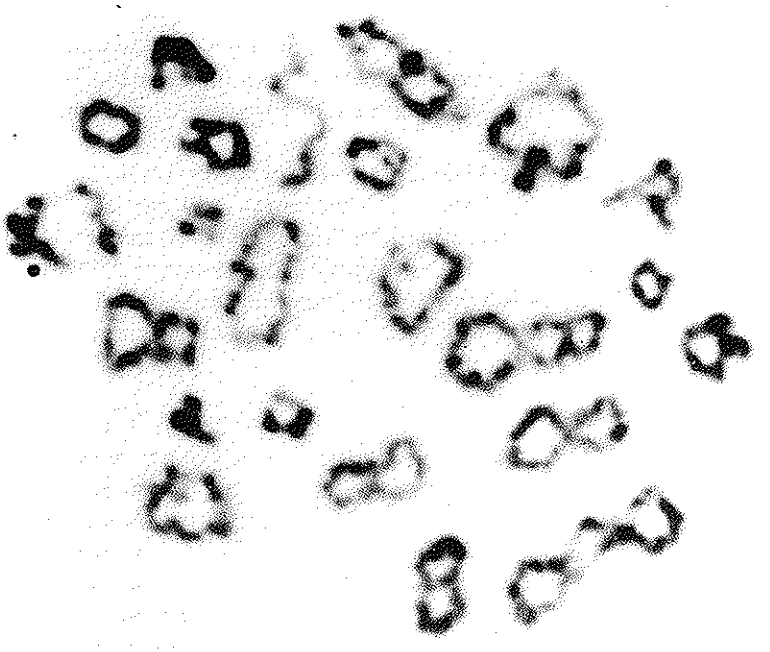

**Figure 2.** Meiosis. Human male meiotic metaphase I chromosome spread, showing paired homologous chromosomes which, having undergone recombination, are about to separate. For a detailed explanation of meiosis, *see* Whitehouse (1973).

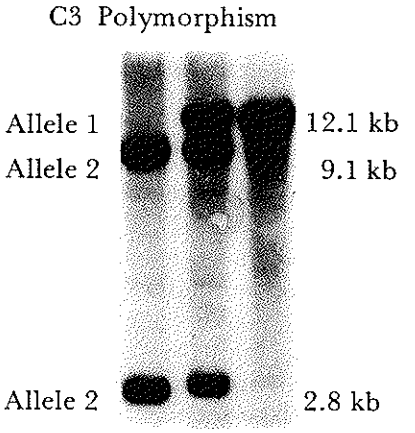


**Figure 3.** Restriction fragment length polymorphism detected by complement component 3 probe, using the enzyme *Sst* I. The probe detects either one fragment of genomic DNA of length 12 kb, or two fragments of lengths 9 and 3 kb, depending on the presence or absence of an *Sst* I restriction site The centre track represents a heterozygous individual and, therefore, shows all three bands.

*et al.* (1984) had expected to screen several hundred DNA markers before finding one linked to the HD gene — once a disease gene has been assigned to a particular chromosome the approach becomes less 'random' and more 'directed' towards that particular chromosome. For chromosome 19, for example, with an estimated maximum genetic length of 100 cM, five or six evenly spaced RFLPs should allow any genetic disease locus present on this chromosome to be mapped.

In this article we shall address ourselves to the problems of, and approaches to, gene mapping on specific human chromosomes, and particularly to our attempts to identify closely linked markers for myotonic dystrophy (DM) on chromosome 19, which can be used for genetic prediction in this disorder. We shall also consider the question of the identification of the actual gene defect in an inherited disorder, the molecular basis of which is unknown.

## MYOTONIC DYSTROPHY

Myotonic dystrophy (DM) was first recognized as a distinct disorder at the beginning of this century (*see* Harper, 1979). It is the commonest inherited neuromuscular disease of adult life and shows autosomal dominant inheritance (Bell, 1948) with an incidence of approximately 1 in 8000 (Harper, 1979). Unlike Duchenne muscular dystrophy, where about one-third of cases are due to new mutations, these are rare in myotonic dystrophy. The severity of the disorder is extremely variable, although affected individuals have, on average, a shortened life expectancy. It is characterized by facial muscle weakness, ptosis of the eyes, and myotonia, best demonstrated by a delayed ability to relax the hand following a forceful grip. There are, however, a

variety of other abnormalities including retinal degeneration and cataracts, cardiac problems and mild mental deterioration frequently found in adults, as well as frontal balding and endocrine disturbances.

The disease is particularly variable in respect of severity and age of onset, but it can be divided into congenital and adult forms. In the congenital form, hypotonia is one of the most common features, but among children who survive there is a steady improvement and hypotonia is rarely present beyond the age of 3 or 4 years. Myotonia and other symptoms develop in later life. In the 'adult' forms the characteristics described above are usually apparent between 20 and 40 years of age.

The congenital form is seen almost exclusively when the affected parent is the mother, although both forms can occur within the same family, and appear to involve the same abnormal gene. Thus, despite the mode of inheritance being basically autosomal dominant, it would seem that some intrauterine factor may also be involved in some cases. The basic gene defect in DM remains unknown (for a review of the disease *see* Harper, 1979).

## THE CURRENT GENETIC MAP OF CHROMOSOME 19

The first autosomal linkage group discovered in man by Mohr (1951) was between the Lutheran blood group (Lu), and what was later interpreted as the Secretor (Se) locus (Grubb, 1953; Mohr, 1954). The locus for myotonic dystrophy was subsequently added to this linkage group by Renwick and Bolling (1971) and Harper *et al.*, (1972), thus confirming the earlier suggestion of linkage found by Mohr (1954).

Linkage between complement component 3 (C3) and the DM-Se-Lu linkage group was demonstrated by Eiberg *et al.*, (1983). It was not until Whitehead *et al.*, (1982) mapped the human C3 locus, however, that the entire linkage group could be assigned to chromosome 19. Peptidase D (PepD) which was assigned to chromosome 19 by somatic cell hybrid studies (McAlpine *et al.*, 1976), has also shown very close linkage to the myotonic dystrophy locus with a Lod score of 3·51 at $\theta$ of 0 (O'Brien *et al.*, 1983). For a full explanation of Lod scores *see* page 333.

On the basis of combined protein and DNA polymorphism data, Davies *et al.* (1983) found linkage between C3 and the myotonic dystrophy gene with a Lod score of +3·36 at $\theta$ of 0·05 in males. This was not the case in females, however, with a maximum Lod of +0·15 occurring at $\theta = 0{\cdot}40$. Differences in the recombination frequency between the sexes are not uncommon (Renwick, 1969).

Ball *et al.*, (1984) have localized the C3 gene to the short arm of chromosome 19 in the region 19pter-19p13·2, using somatic-cell hybrids (*see* pages 325–330). This was subsequently confirmed in our laboratory, using a more extensive hybrid-cell panel, and we also localized PepD to the p13-q13 region of chromosome 19 using the same panel, thus permitting the orientation of the linkage group as shown in *Figure 4*. The genes for lysosomal mannosidase, lysosomal DNAase, and glucose phosphate isomerase (GPI) were also localized to the region 19p13-19q13 (Brook *et al.*, 1984).

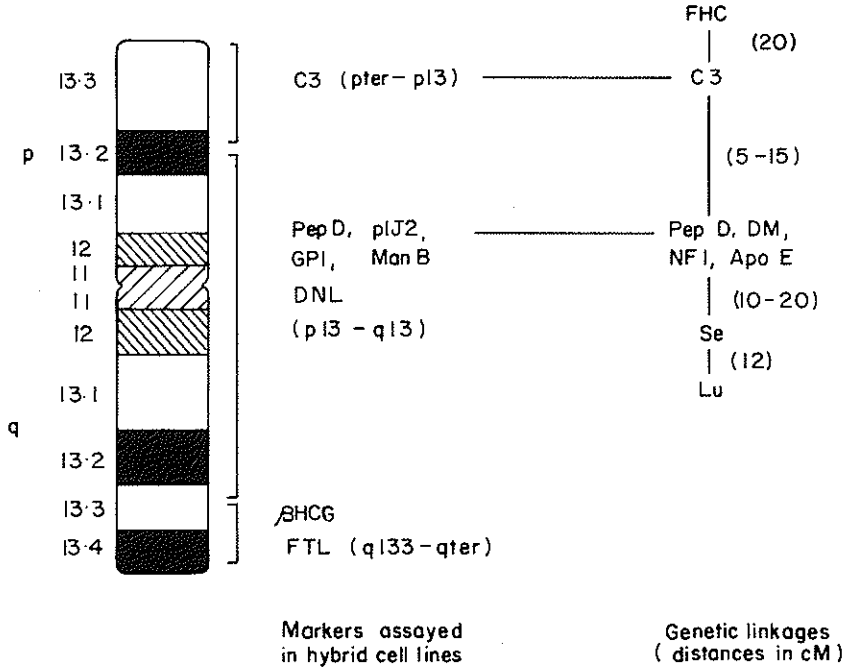


**Figure 4.** Chromosome 19 markers. Genetic and physical maps of chromosome 19, showing the relative positions of the genetic markers along the chromosome. The numbering of bands is shown on the left.

On the basis of linkage data, familial hypercholesterolaemia, caused by mutant alleles at the low-density lipoprotein (LDL) receptor locus, appears to be located 30 cM distal to the C3 gene (Berg *et al.*, 1984) in the region 19pter-19p13. The LDL receptor has been directly localized to chromosome 19 using somatic-cell hybrids (Francke, Brown and Goldstein, 1984).

Caskey *et al.* (1983) assigned the genes for the ferritin light and heavy chains (FTL and FTH respectively) to chromosome 19. More recently, Worwood *et al.* (1985) have confirmed the FTL result and further localized the gene to the region 19q13·3–19qter. The active FTH gene, however, has been assigned to chromosome 11 using a more specific assay (Worwood *et al.*, 1985); there are pseudogenes for both FTL and FTH distributed among several other chromosomes. The human chorionic gonadotrophin B sub-unit localized to chromosome 19 by Naylor *et al.*, (1984) has also been regionally localized to 19q13·3–19qter (Brook *et al.*, 1984).

Other genes assigned to chromosome 19 include one for branched-chain amino-acid transferase (Naylor and Shows, 1979), elongation factor 2 (Kaneda *et al.*, 1984) and one of the cytochrome P–450 genes (Phillips *et al.*, 1985), as well as a number of virus receptor sites, such as the polio virus receptor (Miller *et al.*, 1974), Echo II virus receptor (Gerald and Bruns, 1978), and the RD114 receptor (Schnitzer *et al.*, 1980). Recently Rettig *et al.* (1984) have

raised monoclonal antibodies to two cell-surface antigens, the genes for which map to chromosome 19.

A group of genes coding for apolipoproteins has been mapped to chromosome 19 using both linkage and somatic-cell hybrid studies. Gedde-Dahl *et al.* (1984) have found close linkage between apolipoprotein E (ApoE) and the Lu locus. Furthermore, Humphries *et al.* (1984) and Myklebost *et al.* (1984a, b) have both reported close linkage of apolipoprotein CII (ApoCII) to ApoE. We have found that ApoCII is closely linked to myotonic dystrophy with a maximum Lod score of 7·8 at 4% recombination (Shaw *et al.*, 1985) and have localized this and the apolipoprotein genes E and CI to the 19p13–19q13 region.

The use of RFLPs identified by cloned genes, such as C3 (Davies *et al.*, 1983) and ApoCII, in linkage studies as outlined above is becoming increasingly common. These studies complement those using random DNA sequences which identify RFLPs spanning the chromosome. This approach to localizing human disease genes, together with a consideration of their ultimate isolation, is outlined below (and *see also* Chapter 9 of this volume).

## DNA cloning and library construction

Genetic variation is attributable to differences in DNA base sequence. Occasionally a sequence change results in a phenotypic alteration such as a difference in a biochemical property of a known protein, and this class of variant forms the basis of 'traditional' genetic markers. It is generally accepted that approximately 1% of bases in the human genome are variable and, as the overall length of the genome is $3 \times 10^9$ bp, there are about $10^7$ potential genetic markers available. In principle, it is possible to detect all of this variation using recombinant DNA probes, and we will now consider the ways in which these probes are produced.

The techniques of cloning fragments of DNA in the bacterium *Escherichia coli* are by now well known, but readers who are still unfamiliar with this technology are referred to texts such as that by Old and Primrose (1981). The series edited by Williamson (1981–1983) contains a number of reviews on more specialized topics. For our purposes, it is sufficient to note that human DNA can be cloned as easily as that of any other organism, and that the main problems are a consequence of the high degree of complexity ($3 \times 10^9$ bp) of the human genome.

### GENOMIC LIBRARIES

A collection of DNA fragments that is intended to represent the entire genome of an organism is called a total genomic library. The best-known example of a human genomic library, and one that is still widely used by laboratories around the world, is that constructed by Maniatis and his colleagues (Lawn *et al.*, 1978).

The problem in constructing a human genomic library is to ensure complete representation of the genome. It can be shown by calculation that in a library

constructed from randomly generated 20 kb genomic fragments, about $8 \times 10^5$ clones are required to represent 99% of the genome. The number of clones required decreases with increasing length of the genomic fragments.

If human DNA is digested to completion with an enzyme such as *Eco*RI, a variety of fragments of widely different sizes, with an average length of about 4 kb, is produced. As most of the commonly used phage and cosmid vectors are only able to accept fragments within a narrow size range (*see below*), a library made using completely digested genomic DNA cannot be fully representative.

It is, therefore, desirable to be able to generate reasonably large overlapping genomic fragments in a random or pseudo-random fashion. Two ways in which this has been achieved, are (1) to fragment DNA mechanically to the required size range, e.g. by sonication, and then to add synthetic adaptors to the ends for cloning; or (2) to partially digest the DNA to the required average length with an enzyme such as *Mbo*I that cuts very frequently. Both of these procedures can achieve pseudo-random fragmentation of DNA and are therefore suitable for constructing representative genomic libraries. In practice, the second option has found the widest use as it is technically more convenient. Readers are referred to Seed, Parker and Davidson (1982) for a full theoretical treatment of the problems of representation.

To keep to a minimum the large numbers of clones that must be generated, it is desirable to maximize the size of the cloned DNA fragments. The most widely used vectors are derivatives of bacteriophage lambda (Murray, 1983). By deleting non-essential genes, lambda can be made to accept up to 22 kb of foreign DNA. This limit is set by the capacity of the lambda phage particle for DNA (52 kb) and the total length of the essential lambda genes (30 kb).

Cosmid vectors can accept up to 45 kb of exogenous DNA, and therefore a cosmid library need contain only half the number of clones of an equivalent lambda library. A cosmid has four essential features: (1) The *cos* sequences of phage lambda which enable it to be packaged *in vitro* to form infectious phage particles, allowing efficient recovery of cloned DNA sequences; (2) an origin of replication, and (3) an antibiotic-resistance (or other selectable) gene, which allow the cosmids to be selected for and stably maintained as plasmids in *E. coli*; and (4) suitable restriction enzyme sites for cloning. As all of these requirements can be met by about 5 kb of DNA, the remainder of the 52 kb capacity of the lambda particle can be made up of foreign DNA. Ish-Horowicz and Burke (1981) have described the use of cosmids for constructing genomic libraries.

Despite the greater capacity of cosmids for DNA, phage lambda is still used preferentially by many investigators; this is because it is considerably easier to construct, maintain and screen lambda libraries.

### cDNA LIBRARIES

Only a small fraction of the human genome actually codes for proteins, and different sets of genes are expressed in different tissues. A cDNA library is constructed using transcribed sequences only, and is intended to represent all the active genes in a particular tissue or cell type at a specific developmental

stage. The clones in a cDNA library represent the mature form of the mRNA, which has been processed to remove its introns.

Many variations on the following basic procedure for the construction of cDNA libraries have been described (*see*, for example, Okayama and Berg, 1982; Gubler and Hoffman, 1983). First, total RNA is prepared from the tissue of interest. Messenger RNA can then be selectively purified from the bulk of ribosomal and transfer RNA, by virtue of the fact that most eukaryotic mRNA species have a 3′ polyA 'tail' and can be absorbed on to a poly-dT affinity column. The mRNA is copied into single-stranded DNA using the enzyme reverse transcriptase and the four deoxynucleoside triphosphates; the mRNA is then removed and the DNA made double-stranded using DNA polymerase. To the double-stranded cDNA are added single-stranded polynucleotide 'tails' or synthetic adaptors, and the DNA is then cloned into complementary sites in a suitable vector. Plasmids such as pBR322 have generally been used for this purpose; the relatively low capacity of plasmid vectors is not a disadvantage because most processed mRNAs are fairly short sequences. More recently, specialized lambda derivatives, such as λgt11 (Young and Davis, 1983) and expression plasmids (Helfman *et al.*, 1983) have been used as vectors. These have the advantage that inserted DNA may be placed under the control of lambda or plasmid promoter sequences and expressed in the cloning host, which allows immunological screening strategies to be used (pages 322–324).

The problem of representation differs between cDNA and genomic libraries. Differences in the rate of expression of individual genes result in variation in the abundance of mRNA species, and this is reflected in the composition of the resulting cDNA library. The number of clones that must be generated is therefore a function of the abundance of the rarest sequence one wishes to detect. Consequently, it can be very difficult to isolate cDNA clones corresponding to rare mRNA species. cDNA libraries have been constructed from a number of human tissues, such as muscle (adult and fetal), leucocytes and liver. The choice of tissue is governed by the pattern of expression of the gene one is interested in. In the case of myotonic dystrophy and the majority of other inherited diseases, the choice of tissue for study must be an informed guess based on the clinical characteristics of the disease. For example, one might use a muscle cDNA library to search for muscular dystrophy genes or a brain library as a source of candidates for the Huntington's disease gene. Another variable that has to be considered, is the developmental age of the tissue used, as many genes show temporal variations in expression as part of the developmental process. Therefore, the cDNA approach to looking for genes causing inherited diseases such as myotonic dystrophy is at present fraught with difficulties. cDNA libraries have generally been most useful when some knowledge of the nature of the gene product has been available, enabling the isolation of a wide variety of human genes.

### CHROMOSOME-SPECIFIC LIBRARIES

In many cases the chromosomal localization of a disease gene will have been established by family studies or by the occurrence of specific chromosomal

rearrangements associated with the disease. It is then worth while to restrict one's attention to a particular chromosome. Methods have become available for the construction of libraries specific for one or a small number of human chromosomes, and these are of two types, outlined below.

*Physical methods*

The chromosome of interest may be physically removed from the remainder of the genome. This can be achieved either by sedimentation through a sucrose gradient (Collard *et al.*, 1981) or by fluorescence-activated chromosome sorting, in which mitotic chromosomes are stained with a dye such as ethidium bromide, excited by laser light, and sorted into fractions on the basis of their fluorescence intensity (Carrano *et al.*, 1979; Young *et al.*, 1981). It is also possible to combine these two procedures, the sedimentation step preceding the chromosome sort. Both of these techniques are based on differences in the size of the various chromosomes, and have the disadvantage that similar-sized chromosomes such as numbers 9, 10, 11 and 12 cannot be completely purified. Recent refinements of the fluorescence-activated sorting procedures include the use of dual-wavelength, two-dimensional sorting, which introduces a second sorting parameter based on a different fluorescent property of the chromosomes. This has overcome many of the original limitations but the equipment required is highly complex and beyond the reach of most laboratories (Langlois *et al.*, 1982).

A noteworthy example of a human chromosome-specific library constructed by sorting is that of Davies *et al.* (1981); this library is relatively specific for the human X chromosome and provided the first random genomic DNA polymorphism linked to a disease locus (Duchenne muscular dystrophy; Murray *et al.*, 1982).

At present it is not a straightforward exercise to construct a specific chromosome library using sorting, for three reasons: first, the availability of the apparatus required; second, the necessity for very long sorting times to obtain enough DNA to clone (typically of the order of days); and third, the problem of contamination of the sorted material with similarly sized chromosomes or fragments of longer ones. It is likely however, that some of these problems will soon be overcome, and one welcome development is the undertaking by the Los Alamos laboratory in the USA, to provide chromosome-specific libraries, constructed using their highly sophisticated apparatus, as a service to the research community (Deaven, 1984).

*Somatic-cell hybrids*

Somatic-cell hybrids with a restricted human chromosome content can be used as a source of DNA for isolating sequences specific for the human chromosomes they contain (*see* pages 325–328). The major drawback of this approach is the difficulty in obtaining suitable hybrids. In the case of human chromosomes such as 17 and X, where a mutant rodent cell line and selective system is available, it is possible to retain selectively the individual chromo-

some. Otherwise, one has to rely on the chromosome of interest being retained fortuitously. Gusella *et al.* (1980) have successfully applied this approach to the isolation of sequences from chromosome 11, and Cavenee *et al.* (1984) have achieved similar results with chromosome 13. Apart from the lack of suitable selective systems for most human chromosomes, the other disadvantages of somatic-cell hybrids are (1) contamination of the library with other human material; (2) missing sequences due to rearrangement of the human chromosome during cell culture; and (3) the need to handle a large number of clones because a library obtained in this way represents the entire rodent genome as well as the human material. However, if a suitable cell line is available, this approach is technically easier than the use of chromosome sorting. Improvements to the somatic-cell hybrid methodology are likely in the near future, for example, the discovery of new selective systems, or the deliberate introduction of selectable markers into the chromosome of interest and the development of microcell-mediated chromosome transfer for the introduction of single human chromosomes into a rodent cell.

There are a number of libraries in existence that are enriched for chromosome 19 sequences. Davies *et al.* (1984) reported a library constructed from flow-sorted chromosomes, and Brook *et al.* (1984) described the isolation and mapping of a probe from this library that detects a chromosome 19 RFLP. Roses and his colleagues have also constructed a flow-sorted library and isolated several chromosome 19 sequences from it (Allen Roses, personal communication) and a third flow-sorted library is being constructed at the Los Alamos laboratory (Deaven, 1984). In our own laboratory we have constructed a library using DNA from a mouse–human-hybrid cell line that has chromosome 19 as its major human component. More than 50% of the human sequences in this collection are specific for chromosome 19, and several have been mapped to specific regions of the chromosome and used to identify new RFLPs (unpublished results).

## SPECIALIZED CLONING VECTORS

Some of the commonly used vectors for cloning human DNA in *E. coli* are described above; however, much effort has been expended recently in the design and construction of vectors that will enable cloned DNA to be introduced into, maintained stably in and expressed in, eukaryotic cells. This will enable the function of cloned sequences to be investigated, and the gene products to be synthesized, in the appropriate eukaryotic cellular environment.

In order to ensure efficient recovery of the cloned DNA, it is usually necesary to use a selection system. This may be based on the complementation of a mutation in the host cells by a gene in the vector: for example, mouse cells deficient in thymidine kinase can be used in conjunction with a vector containing an active thymidine kinase gene; in the appropriate medium only those cells that have taken up a recombinant DNA molecule can grow. Alternatively, one can use a vector containing a gene coding for resistance to

a toxic drug. Cells transformed with this DNA will be able to grow in a medium containing the drug.

Because of the large size (up to 100 kb) of many eukaryotic genes, a vector with a large DNA capacity is required if expression of the cloned gene is the aim. Grosveld *et al.* (1982) and Lau and Kan (1983) have reported the construction of cosmid vectors that can be used to introduce cloned DNA up to 45 kb in length into mouse cells, and have shown that the globin genes are accurately expressed in this system. More recently, Cepko, Roberts and Mulligan (1984) developed a vector based on a mouse leukaemia virus which can be efficiently introduced into and recovered from mammalian cells, and also used to generate cDNA (processed) versions of genomic sequences.

It is likely that mammalian host–vector systems such as those described above will rapidly become more refined and easier to use, and will be the method of choice when functional aspects of cloned DNA sequences are being studied.

## Screening libraries and isolating probes

Libraries of mammalian DNA generally contain large numbers of individual clones and therefore efficient screening strategies are required. Some of these are described below.

### Screening a cDNA library

In many cases, the project requires the cloning of a particular gene sequence from a cDNA library, and to ensure a reasonable chance of success the library must be representative (*see* previous section). Screening is usually carried out using nucleic acid hybridization. When some DNA or protein sequence information is available (it can be as little as 15 bases or five amino acids), a complementary oligonucleotide can be synthesized and used as a hybridization probe. Clones from the library are plated out on to Petri dishes and replicas are made on to nitrocellulose filters, which are then hybridized with the radioactive oligonucleotide probe. The colonies that hybridize are picked from the original Petri dish and grown up for further study.

Alternatively, if an antibody specific for the product of the gene of interest is available, an immunological screening procedure can be used. The cDNA library is constructed using one of the 'expression vectors' that allow the synthesis in *E. coli* of at least part of the product of the cloned gene, under the control of a bacterial promoter sequence (Helfman *et al.*, 1983; Young and Davis, 1983). Bacterial colonies that contain the gene of interest are identified in much the same way as described above, by making a replica of the library on to a filter and reacting this with labelled antibody. It is not possible to screen genomic libraries in this way because *E. coli* lacks the enzymes necessary for the removal of introns and further processing of the primary transcript of a eukaryotic gene. In addition to their usefulness in the study of gene structure and function, cDNA clones can be used to detect RFLPs for use in linkage studies: for example, the cloned complement C3 and apoli-

poprotein CII genes have been used for genetic studies on chromosome 19 (Davies *et al.*, 1983; Myklebost *et al.*, 1984a; Shaw *et al.*, 1985).

### SCREENING GENOMIC LIBRARIES

Genomic libraries are usually screened using nucleic acid hybridization. The probe used depends on the type of clone one is trying to isolate. In some cases one will be seeking a specific sequence, such as a particular gene; the probe used could be either a synthetic oligonucleotide, as described above, or a cloned DNA fragment such as a sequence isolated from a cDNA library. Isolation of genomic clones allows the structure of flanking regions, introns and other features of mammalian genes to be studied.

Alternatively, the library can be screened with labelled cDNA made from mRNA of a particular tissue, so that clones containing transcribed sequences can be identified. It can be difficult to identify clones representing genes expressed at low levels, because the concentration of their particular mRNA species will be correspondingly low. This problem can be partially overcome by using as probe DNA from pools of clones from a cDNA library, having first grouped the cDNA clones according to their abundance.

For genetic studies on a particular chromosome, probes are required that can be used to detect DNA sequence polymorphisms. These are isolated from a flow-sorted or somatic-cell hybrid library, and their chromosomal origin is confirmed before use. In a somatic-cell hybrid library the majority of the clones are derived from rodent DNA, and the human sequences are first identified by probing with a labelled total human DNA. The hybridization is due to highly repeated sequence elements, of which there are several classes distributed throughout mammalian genomes; because these usually do not cross-hybridize between species as different as mouse or hamster and man, they may be used to distinguish between clones of different origin (Gusella *et al.*, 1980). In order to use clones isolated in this way for genetic studies, the repetitive sequence elements have first to be removed from them. This is commonly done by isolating single-copy DNA fragments (identified by their lack of hybridization with a total human DNA probe) from a restriction enzyme digest of the clone, and using these as probes of the genomic DNA. Alternatively, the repetitive elements can be prevented from hybridizing by preincubating the probe with a large excess of unlabelled human DNA, so that unlabelled repetitive sequences will out-compete the labelled probe repetitive DNA in the subsequent hybridization. This procedure allows the genomic clones to be used as probes intact, and so has the advantage over small fragments of allowing a larger region of genomic DNA to be examined. Clones from a flow-sorted library should all be of human origin and do not need to be preselected; however, a majority of them will contain repetitive sequences and will need to be treated as described above to be useful as probes.

Repetitive sequences that have low copy numbers (1000 or less) in the human genome, can however be useful for chromosome mapping (Gusella *et al.*, 1982; Law, Davidson and Kao, 1982). When these sequences are

hybridized to a Southern blot of digested genomic DNA, a 'ladder' pattern of bands is produced, each band representing a genomic copy of the repetitive sequence. In principle this method should allow the mapping of large numbers of loci simultaneously and is being used in our laboratory for identifying different fragments of human chromosome 19 present in somatic-cell hybrids (unpublished results).

### CHROMOSOME WALKING

Having obtained a cloned DNA sequence representing part of a region of interest it is often useful to isolate a set of overlapping genomic clones from the same region. Many mammalian genes are too large to be cloned intact and so have to be isolated as a series of clones. For genetic linkage studies the informativeness of a marker can be increased by isolating new polymorphic sequences obtained by 'walking' away from the original probe. The basic technique of chromosome walking is as follows (*and see* Cockburn, Howells and Whitten, 1984). A single-copy probe is used as a starting point to screen a genomic library and to obtain new overlapping clones; the relationship of these to each other and to the original probe is established by restriction mapping, and new single-copy walking probes are isolated from the distal regions of the new clones. The walking process is then repeated, as many times as is required, in either direction. Despite the laborious nature of this process, successful walks have been achieved, for example, to isolate more than 500 kb of DNA from the major histocompatibility region of mouse chromosome 17 (Steinmetz *et al.*, 1982; Chaplin *et al.*, 1983).

An elegant and potentially labour-saving alternative to library screening by DNA hybridization has been developed for lambda or cosmid libraries (Seed, 1983; Poustka *et al.*, 1984). The probe is cloned into a small plasmid containing a selectable marker gene, and a culture of bacteria carrying the plasmid is infected with an aliquot of the genomic library. If a bacterial cell becomes infected with a library clone that has sequence homology with the original probe, then recombination between the two molecules can occur, resulting in a lambda or cosmid clone in which the probe plasmid together with its selectable gene is integrated. To recover these particular clones the library is repropagated on a new host strain, under selective conditions so that only those clones that have integrated the probe plasmid can grow. It is also possible to recover selectively human clones from a somatic-cell hybrid library, by using as the probe a cloned human repetitive sequence (Neve *et al.*, 1983). *In vivo* recombination is a particularly promising technique as it avoids much of the tedious work involved in screening by conventional methods, and has also allowed the isolation of previously undetected clones from libraries that had already been extensively screened.

'Chromosome hopping' is another recent development which should allow the isolation of new DNA fragments at distances of 100 kb or more from an initial probe (Collins and Weissman, 1984). Genomic DNA is fragmented to a length of $n$ kb, corresponding to the distance of the 'hop', and ligated to a small marker DNA fragment, to form circular molecules. These are then

digested with a restriction enzyme that cuts within the genomic DNA but not in the marker sequence, and the digested fragments are re-circularized. By selection for the marker gene, plasmids containing a 'junction fragment' (representing two sequences that were originally separated in the genome by *n* kb) can be isolated. Screening a library of such fragments should allow one to isolate sequences at the hop distance from the original probe. This technique is potentially useful for mapping complex loci or moving between linked marker loci.

## Localization of probes

Single-copy sequences, irrespective of their source, have to be localized to particular chromosomes, and if possible, to subchromosomal regions, before polymorphism searches and time-consuming linkage studies are undertaken. For the X chromosome, for example, where the estimated genetic length is at least 200 cM (Drayna *et al.*, 1984), any probe mapping to the long arm would be of little or no use in linkage studies with Duchenne muscular dystrophy (located at Xp21), as it would be so far away that it would appear unlinked. Even for a small human chromosome such as 19, the genetic length may well be in excess of 100 cM (Davies *et al.*, 1983), so that markers close to the centromere may not show linkage to those at the distal end of either arm. Consequently, the more precise the localization of any particular probe, the more readily can its usefulness be predicted.

In order to localize single-copy sequences to particular chromosomes, two basic techniques have been employed extensively: hybridization to somatic-cell hybrids and *in situ* hybridization (*see also* Chapter 9 of this volume). A third technique, which is being used increasingly in such studies, involves hybridization to DNA obtained from chromosomes sorted on the fluorescence-activated cell sorter (FACS).

### HUMAN–RODENT SOMATIC-CELL HYBRIDS

The concept of selective systems in somatic-cell genetics was introduced by Littlefield (1964). These systems exploit drug-resistant cells which have arisen when cultured cells were exposed to high concentrations of chemicals such as 8-azaguanine (8-AZA) or 5-bromodeoxyuridine (BUDR). Littlefield (1964) demonstrated that it was possible to select for hybrid cells which occurred spontaneously when 8-AZA-resistant and BUDR-resistant cells were co-cultured, using selective medium containing hypoxanthine, aminopterin and thymidine (HAT).

The fusion of mouse and human cells by means of Sendai virus (Harris and Watkins, 1965), or polyethylene glycol with selection of hybrids in HAT medium, led to the observation that human chromosomes are subsequently lost from resulting hybrid cells. This is summarized in *Figure 5*. Hybrid cells resulting from the fusion of a human cell line with an 8-AZA-resistant rodent-cell line (which is deficient therefore in hypoxanthine phosphoribosyl transferase ($HPRT^-$)) and selected in HAT medium, will retain the human X

chromosome as the human HPRT gene is located on this chromosome. Similarly, human chromosome 17 will be retained in hybrid cells resulting from the fusion of BUDR-resistant mouse cells (thymidine kinase-negative ($TK^-$)) and human cells as the human TK gene is located on chromosome 17 in man.

The retention of human chromosomes, other than those for which it is possible to select, would appear to be completely fortuitous. The loss of human chromosomes seems to occur during the early divisions following fusion (Weiss and Green, 1967). Generally, once it has stabilized, a hybrid cell will grow and retain a certain combination of human chromosomes, although further loss of individual chromosomes is possible at any stage. Hybrid cells can be characterized using cytogenetic and enzymatic techniques, as well as chromosome-specific DNA probes. Different hybrids retain different combinations of human chromosomes: thus, with a panel of characterized hybrid cells (*see Figure 6*) it is possible to localize unique-sequence DNA probes to particular chromosomes. Furthermore, by using cells from individuals who carry balanced reciprocal translocations of certain chromosomes it is possible to construct hybrids containing particular sub-chromosomal regions. If the translocation involves chromosomes 19 and X, for example, the part of chromosome 19 joined on to the part of the X

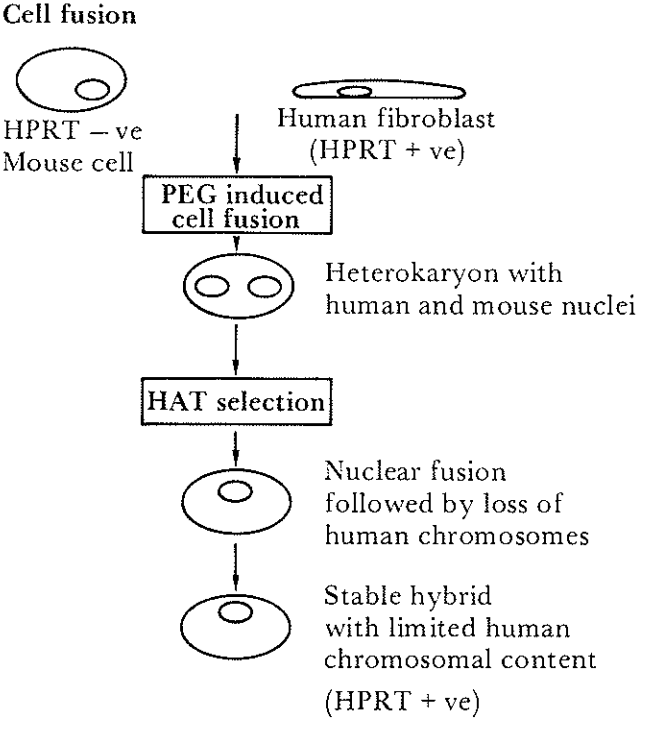


**Figure 5.** Cell fusion. Production of mouse–human somatic-cell hybrids, using polyethylene glycol (PEG) with hypoxanthine–aminopterin–thymidine (HAT) selection.

chromosome carrying the HPRT gene will be selectively retained (*see Figure 7*). Using such human–rodent somatic-cell hybrids in this way, unique-sequence DNA probes can be localized to specific parts of a chromosome. A number of individuals carrying balanced reciprocal translocations of chromosome 19 are to be found in the general population and fibroblast cultures have been established from several of them. Using these fibroblasts it has been possible to construct a series of hybrid cells, with which probes on chromosome 19 can be localized to any one of nine different regions (*see Figure 8*).

Somatic-cell hybrid panels for regional localizations on the X chromosome have been produced by a number of groups (*see* Murray *et al.*, 1982, for example). This has been facilitated, however, by the ability to select for the X chromosome. Except in the case of 17, selective systems for other human chromosomes have not been widely used. A number of such systems have

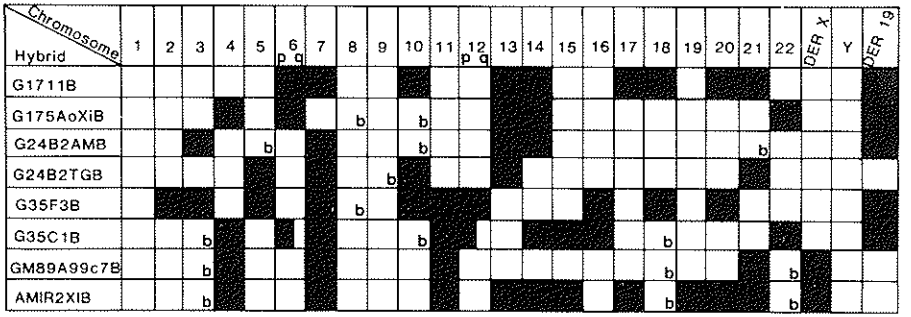


**Figure 6.** Hybrid panel. Panel of rodent–human hybrid cell lines each with a different human chromosome content. Black squares indicate where a chromosome was present; white squares, absent by both cytogenetic and biochemical criteria; b, absent cytogenetically but biochemical data not available. DER X and DER 19 represent parts of chromosomes X and 19 respectively and are fully described in Brook *et al.* (1984).

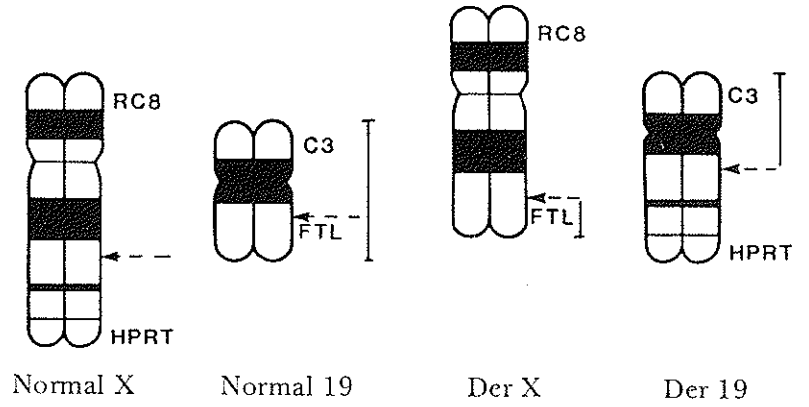


**Figure 7.** An example of a useful balanced reciprocal translocation in which part of chromosome 19 is joined on to the bottom part of the X chromosome giving a derivative 19 (DER 19) which can then be selectively retained in somatic-cell hybrids.

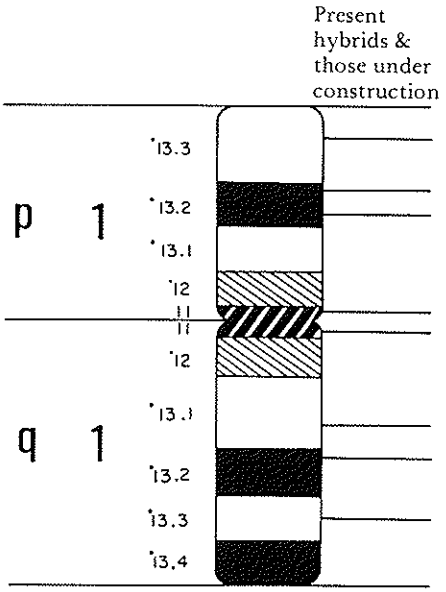


**Figure 8.** The break-points on chromosome 19 which are of use in the construction of somatic-cell hybrids. These provide nine different regions (shown by the lines on the right) to which probes can be mapped.

been proposed, for example the use of uridine monophosphate kinase mutants to select for chromosome 1 (Medrano and Green, 1974) and a system based on branched-chain amino-acid transferase to select for chromosome 19. In both cases the selective systems were unstable and therefore of little practical use. Two other potentially useful selections for chromosome 19 include one based on diphtheria-toxin resistance (Kaneda *et al.*, 1984) and another which exploits the glucose phosphate isomerase gene on this chromosome (Michael Morgan, personal communication). Neither of these selective systems are currently in general use. Studies involving chromosomes other than 17 or X, therefore, still have to rely on their fortuitous retention in somatic-cell hybrids. Furthermore, there are two other drawbacks in the use of hybrid cells in gene-localization experiments. Each time a cell line is grown up, complete re-characterization is necessary as further human chromosome loss can occur, and minor chromosomal re-arrangements can be occurring constantly (Engel, McGee and Harris, 1969). Consequently, in some situations only part of a chromosome may be present in a hybrid for which it is assumed that the complete chromosome is present. If sufficient hybrids are used in a localization panel, however, 20 or so being used in many cases (*see* Gusella *et al.*, 1984, for example), the association of a unique DNA sequence with any particular chromosome is usually apparent.

### *IN SITU* HYBRIDIZATION

The technique of *in situ* hybridization (Gall and Pardue, 1969; John, Birnstiel and Jones, 1969), initially used to map repeated sequences (Evans, Buckland

and Pardue, 1974) is now used to localize unique sequences as short as 900 bp (Harper, Ullrich and Saunders, 1981). The technique involves radiolabelling the sequence of interest to a high specific activity by either nick translation (Rigby *et al.*, 1977) or primed synthesis (Feinberg and Vogelstein, 1983), and its subsequent hybridization to metaphase chromosome spreads on a glass slide. The slides are dipped in photographic emulsion and left to expose for a time which depends on whether the sequence is unique or repetitive, its size and the specific activity of the label. The film is developed and the metaphase cells re-examined to determine the precise location of the grains over the chromosomes (*see Figure 9* and *see also Figure 7* of Chapter 9 of this volume).

Originally, procedures involved chromosome identification prior to *in situ* hybridization with subsequent identification of metaphase spreads after developing the film. More recent staining techniques permit chromosome identification following *in situ* hybridization (Harper, Ullrich and Saunders, 1981), thus shortening the procedure. Prometaphase banding — the technique whereby extremely elongated chromosome spreads can be obtained — used in conjunction with *in situ* hybridization, permits very accurate probe localization (Chandler and Yunis, 1978). The precision with which unique

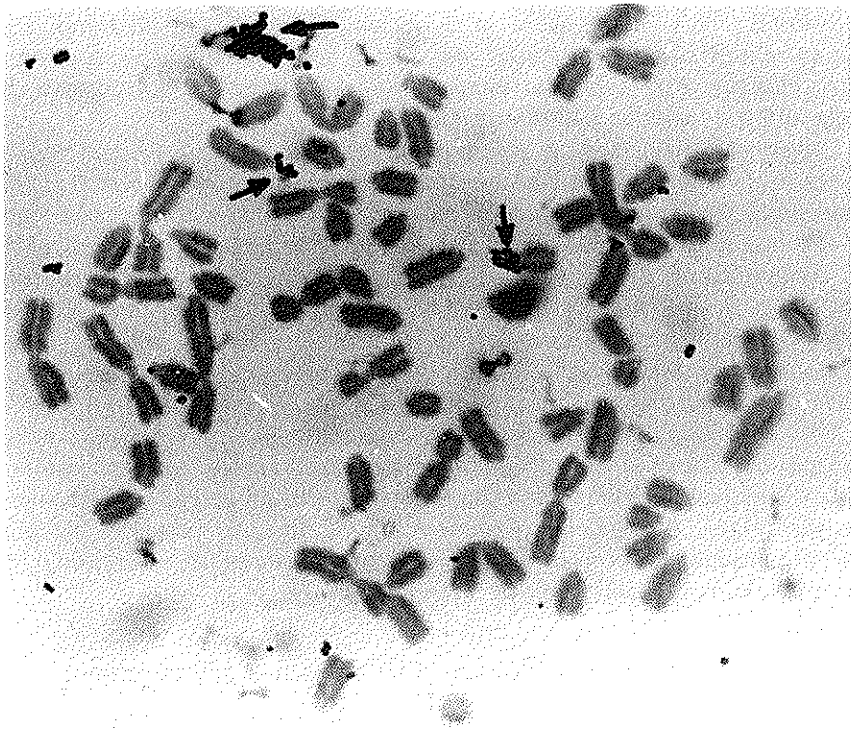

**Figure 9.** The result of *in situ* hybridization with labelled total human DNA on chromosomes of a mouse–human somatic-cell hybrid. Arrows indicate grains covering human chromosomal elements.

DNA sequences are localized, can be further improved by using metaphase spreads containing translocations with defined breakpoints (Fennel *et al.*, 1979; Davies *et al.*, 1984).

The technique of *in situ* hybridization has achieved a high degree of resolution; however, it is not faultless (Bishop and Jones, 1972; Goode, Ledbetter and Daiger, 1984). It is probably best used for localizing specific sequences of known function (Harper, Ullrich and Saunders, 1981; Davis, Malcolm and Rabbitts, 1984) rather than random unique sequences, as a fairly precise localization for one specific probe can be obtained in one experiment. Somatic-cell hybrid techniques, where a well-characterized panel is available, provide a better means of localizing a whole series of unique sequences to particular chromosomes, in a repeatable and relatively rapid procedure.

### DOT BLOTS OF SORTED CHROMOSOMES

Fluorescence-activated cell sorting (FACS) has been used for karyotype analysis, and also to provide a source of material from which chromosome-specific libraries can be constructed (Davies *et al.*, 1981) (*see* page 320). It has also been used to localize genes to specific chromosomes. The technique involves the preparation of single chromosomes in suspension, staining with a fluorescent dye and passing them through a laser beam in the cell sorter. This causes them to fluoresce in such a way that individual chromosomes can be identified and sorted independently, as dots on to nitrocellulose. Any particular unique sequence can be hybridized to such a filter and thus provide a chromosomal localization.

Lebo *et al.* (1979) used this technique to localize human α-, β-, and γ-globin genes to the short arm of chromosome 11. Subsequently, Lebo *et al.* (1982) used similar techniques to localize the human insulin gene to the same region of chromosome 11. These studies used single-laser analysis, which is capable of distinguishing only 14 individual chromosome types. More recently, however, Lebo *et al.* (1984) have employed dual-laser and a DIPI-chromomycin staining technique to identify 22 fractions with 21 unique chromosome types. Using this procedure they have assigned the gene causing McArdle's syndrome to chromosome 11 (Lebo *et al.*, 1984).

This technique has not been used extensively by many groups, mainly because of the cost of the machines involved, the long sort times necessary to obtain adequate numbers of chromosomes, and (until recently) the inadequate resolution. As procedures improve and sufficiently powerful machines become available, this technique will probably be employed increasingly for the chromosomal localization of DNA sequences.

## Identification and genetic applications of DNA sequence polymorphisms

A single-copy DNA sequence, obtained from a library and localized to a specific chromosomal region, can be employed to identify RFLPs (*see* pages 313–314) for use in studies on the inheritance of a disease locus. This practical

application of RFLPs was recognized in 1980 by Wyman and White. Five years on, RFLPs are becoming increasingly employed in the prenatal diagnosis of genetic disease.

In developing a strategy for revealing RFLPs, the number and nature of individuals and enzymes used, and the characteristics of the probes, have to be considered.

## IDENTIFYING POLYMORPHISMS

### *Probes*

Use of large (15–20 kb) fragments such as intact clones from a genomic library is advantageous because this permits the screening of a large number of restriction sites at one time. In addition, large probes are more efficient for locating insertion–deletion polymorphisms since on average each probe screens its own length plus twice the length of the average size of fragment produced by the chosen enzyme (Skolnick and White, 1982). Smaller genomic probes usually cloned into plasmids (0·5 kb–5 kb) permit rapid screening and characterization of point mutation polymorphisms (such as CG–TG transitions) where these affect a restriction enzyme recognition sequence.

### *Panel of normal DNAs*

In order to identify RFLPs the relevant probe is hybridized to DNA from a sample of unrelated individuals. As one would expect, the greater the number of individuals screened, the more RFLPs will be detected. However, this can be counterproductive because many of the RFLPs will be of low frequency in the population and will therefore be of little use in linkage studies. It is more efficient in this case to sacrifice the detection of rarer polymorphisms, particularly if the number of new probes for study is not limited. Aldridge *et al.* (1984) used a panel of three individuals (one male and two females) in their search for RFLPs associated with the X chromosome. They estimate that polymorphisms have been found by at least one-third of the probes tested, the frequency of the rarer allele usually being above 20%. Skolnick and White (1982) have studied the efficacy of screening a panel of four, nine and 19 individuals. They concluded that, using four individuals, the probability of missing potentially useful polymorphisms (rarer allele frequency = 0·1) was nearly 50% and using 19 individuals over 30% of very rare polymorphisms (rarer allele frequency = 0.01) were picked up. They concluded that, using nine individuals, about 15% of the marginally useful polymorphisms (rarer allele frequency = 0·1) were missed but most of the useful polymorphisms would be detected.

The nature of the selected individuals will depend on the genetics of the disease which is under study: for instance, screening of probes localized to the X chromosome is usually performed on a panel of female DNAs because to use male DNA is effectively halving the number of chromosomes under

study. For autosomal diseases, however, this need not be considered. It is important to confirm the allele frequency of an RFLP in a population in which the disease commonly occurs. We routinely screen ten individual samples of DNA extracted from placental tissue. Placenta was selected because large amounts of DNA can be easily extracted from a few grams of this readily obtained tissue.

*Restriction enzymes*

The selection of restriction enzymes is the last factor to consider in screening for DNA polymorphisms. Some enzymes remain poorly studied because their cost is prohibitively high.

It has been estimated that variations in DNA sequence are found at a frequency of at least 1 in 100 base pairs (bp) (Jeffreys, 1979). In theory, restriction sites for enzymes which recognize sequences 4, 5 and 6 bp long, should occur at intervals of 250 bp, 1000 bp and 4000 bp respectively. In reality, the nucleotide distribution is far from random, some dinucleotides being represented more often than others. Thus recognition sites have a distribution frequency which varies from enzyme to enzyme. In man the ratio A+T/C+G is 1·45 which corresponds to 59% A+T and 41% C+G, and means that the sequence AATT occurs four times as frequently as GGCC (Bastie-Sigeac and Lucotte, 1983). Of the more commonly used enzymes there is some evidence that more RFLPs occur at restriction enzyme recognition sites which contain a CG dimer (Barker, Schafer and White, 1984). In human DNA this dinucleotide is often methylated at the C position and is thus subject to frequent replacement by thymidine, because of deamination of the methylated base (Coulondre *et al.*, 1978). This explains the relative rareness of the CG dimers in mammals, as shown by the large average fragment size seen when total human DNA is digested with restriction enzymes whose recognition sequences contain CG. This substitution would result, in some cases, in the addition or removal of a restriction site and thus give rise to polymorphism. *Taq*I and *Msp*I are enzymes to which this phenomenon applies, the recognition sites being TCGA and CCGG respectively, and although no formal study has been undertaken, initial results indicate that a large number of polymorphisms are attributable to changes in *Taq*I and *Msp*I sites. Other enzymes which possess four base-pair recognition sites, e.g. *Alu*I, *Rsa*I, produce many fragments of below 200 bp in length, which are generally not revealed during molecular hybridization. This results in many polymorphisms remaining undetected and the resulting screen is therefore much less efficient.

The optimum enzyme panel is constructed by considering both cost and efficiency of screening. The enzymes used routinely in our laboratory, together with the number of RFLPs associated with each, are shown in *Table 1*.

## LINKAGE AND PREDICTION USING LINKED MARKERS

When probes have been isolated that show DNA sequence polymorphisms and are located in the relevant chromosomal region, linkage studies are

**Table 1.** Restriction enzymes used routinely in screening for DNA polymorphisms

| Enzyme | Recognition sequence | Average fragment size (bp) | Number of RFLPs* |
|---|---|---|---|
| *Taq*I | TCGA | 1179 | 8 |
| *Msp*I | CCGG | 1747 | 3 |
| *Hin*dIII | AAGCTT | 1873 | 1 |
| *Bgl*II | AGATCT | 2699 | 3 |
| *Eco*RI | GAATTC | 3013 | 3 |
| *Xba*I | TCTAGA | 3183 | 0 |
| *Pst*I | CTGCAG | 3213 | 2 |
| *Pvu*II | CAGCTG | 3213 | 2 |
| *Sst*I | GAGCTC | 5082 | 2 |
| *Bam*HI | GGATCC | 5534 | 1 |

*Combined data from probes isolated in our own laboratory and obtained from elsewhere.

undertaken in order to define their genetic distance from the disease locus and their position relative to other markers. Both of these factors are important in establishing a precise chromosomal localization for the disease; this knowledge can then be used in a search for the disease gene itself (pages 336–341), and more immediately for clinical use in establishing the status of at-risk individuals or pregnancies. One of the advantages of using DNA markers for this purpose is the ease with which suitable samples are obtained. From individuals a small venous blood sample yields enough DNA for several hundred restriction enzyme digests and, although amniocentesis can provide fetal material, the recent developments in chorionic villus biopsy (which is performed transcervically) allow a sample of fetal DNA to be obtained at 8–12 weeks of pregnancy (Rodeck and Morsman, 1983).

Linkage data from humans is usually evaluated by likelihood analysis (Morton, 1955) in which the likelihood that two loci are linked at a given distance is expressed by a Lod score (where Lod represents the 'logarithm of the odds' favouring linkage). The Lod score for a particular family represents the weight of evidence for or against linkage at various values of the recombination fraction ($\theta$), usually from 0·00 to 0·45 (*see* page 315). The Lod scores can be summed for a series of families. Linkage is considered to be established where the Lod score reaches +3 at any value of $\theta$. Linkage is ruled out when for any given value of $\theta$ the Lod score is −2 (Race and Sanger, 1975).

When RFLPs closely linked to the myotonic dystrophy gene have been identified, their practical application in estimating the risk of a fetus being affected with this disorder can be assessed. The reliability of a risk estimate depends on the distance between the marker and the disease loci. A distance of 1 cM corresponds to one recombination event in 100 meioses (*see also* pages 336–341), and therefore use of a marker at 1 cM from the disease locus would result in an incorrect risk prediction in 1% of cases. This situation would be greatly improved by the isolation of a second RFLP on the other side of the disease gene. This would minimize the chance of erroneous risk

predictions where both RFLPs are informative, because only a case where a double crossover occurs between the flanking RFLP loci will cause an incorrect prediction to be made. In the case of two markers at, for example, 2 cM and 3 cM on either side of the myotonic gene, the risk of a double crossover would be 2% of 3%, i.e. 0.06% of cases, or less than 1 in 1000. In practice, double crossing over does not occur at these small distances, because of interference between chiasmata.

One limitation of this technique lies in the informativeness of the markers. An RFLP of low frequency will not be useful in many families and it would therefore be necessary to increase the usefulness of such a marker by isolating adjacent single-copy sequences and screening for additional polymorphisms. Chromosome walking, described on pages 324–325, is a method of achieving this.

Even with fully informative RFLPs, genotype prediction for a pregnancy depends on the availability of a basic family structure. A three-generation family usually permits one to establish phase, i.e. to decide which of the marker alleles is associated with the disease gene. A prediction for the fetus can then be made allowing for the limitation of the recombination fraction for a particular RFLP. If the affected grandparent is not available then it is possible to deduce phase from the results obtained from sibs of the affected parent, together with the unaffected grandparent. *Figure 10* illustrates both possible family structures. The initial screen is done using key members of these families i.e. I.1 and II.2. These individuals must be heterozygous for the RFLP in order to establish which one of the parental chromosomes has been passed to the fetus. In the example, I.1 and II.2 are heterozygous and II.2 has inherited the disease together with allele 1. Individuals III.1 and III.2 are both homozygous for allele 2 and are clinically normal. The fetus III.3 has inherited a paternal allele 2 and maternal allele 1, which suggests that, unless a recombination event has occurred, the fetus carries the disease gene.

The situation where DNA from I.1 is not available can be overcome in this family by typing I.2, II.2, II.3, II.4 and II.5. From the genotypes of these individuals the alleles of I.1 can be inferred. All the sibs have allele 2 from I.2; two have allele 2 from I.1 and two have allele 1 from I.1, thus he must be heterozygous for the RFLP.

Until recently, there have been no RFLPs close enough to the myotonic dystrophy locus to be of practical use for prediction. We have recently found that a *Taq* I RFLP at the apolipoprotein CII (apoCII) locus is closely linked (Lod score 7·8 at recombination fraction of 0·04) and should be of clinical use in many cases. The isolation of further polymorphic markers in the region of apoCII, and of suitable markers on the other side of the disease gene, will enable a prediction of genotype to be made for the majority of pregnancies or for at-risk individuals; until the myotonic dystrophy gene itself is identified, the linkage-based approach will continue to be used. However, the major limitation of using linkage, even with ideally informative markers, is the availability of a family structure as described in the examples above: in practice, key members of the family may not be available, and the prediction

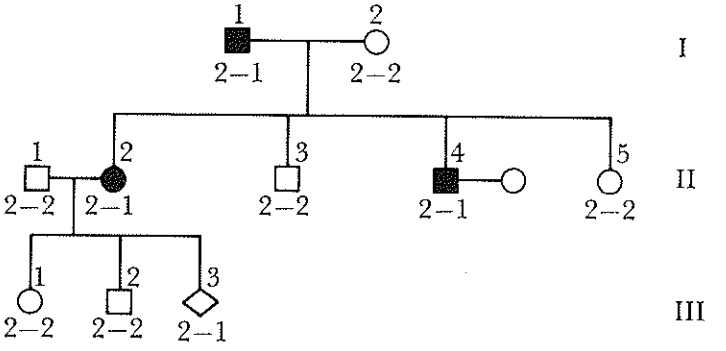


**Figure 10.** The pedigree of a hypothetical family illustrating how linkage between a disease locus and a closely linked probe can aid prenatal diagnosis.

of genotype becomes a matter of probability rather than a definitive judgement.

### PREDICTION USING GENE-SPECIFIC PROBES

Many of the limitations of the linkage approach will be overcome when the gene causing the disease is identified and gene-specific probes become available. This is already the case for the haemoglobinopathies, phenylketonuria, familial growth hormone deficiency, and several other disorders. Isolation of the gene itself permits three different approaches to prediction.

#### *1. RFLPs showing complete linkage*

Any sequence polymorphisms that are detected by a probe specific for the disease gene, even if the polymorphic site is outside the gene itself, will be so closely linked to the gene defect causing the disease that the recombination frequency will be negligible. Such an approach may be useful for the diagnosis of phenylketonuria using a cDNA probe for the phenylalanine hydroxylase gene (Woo *et al.*, 1983) but will still require the investigation of the family as a whole.

#### *2. RFLPs in linkage disequilibrium*

When a mutation giving rise to a new polymorphism occurs in an individual within a population, it is initially associated with a particular set of alleles on the chromosome where it has arisen. Over a period of time recombinations will occur between the markers until eventually the new polymorphism will reach a state of equilibrium, such that the probability of an individual having a particular genotype at the new locus is not influenced by the genotypes at other, linked loci. The closer two loci are together, the longer it takes for

equilibrium to be reached, and in the case of loci that are extremely close it is sometimes possible to observe linkage disequilibrium in the population. Disequilibrium is said to exist when the frequency of each combination of alleles at the two loci differs significantly from what would be expected on the basis of random assortment and, if established in relation to a disease gene, can be of predictive value in the absence of the rest of the family. The feasibility of this approach has been shown in relation to sickle-cell anaemia and one form of β-thalassemia, using a β-globin gene probe (Kan and Dozy, 1978; Kan *et al.*, 1980).

*3. Detection of disease-specific mutations*

Ultimately, normal and defective forms of the disease gene will be isolated and their DNA sequences compared, and the alterations that cause the disease identified. In the case of mutations that affect a restriction site, detection using the appropriate restriction enzyme will be possible. This approach is already possible for sickle-cell anaemia, in which the common mutation causes the loss of a site for the enzyme *Mst*II (Chang and Kan, 1982). Other changes can be detected using oligonucleotide probes, which are chemically synthesized sequences a few tens of base pairs long. By choosing a suitable hybridization temperature, an oligonucleotide probe can discriminate between a perfectly matched genomic sequence and one that is mismatched by only a single base change. The technique has been used to detect mutations in the β-globin gene resulting in sickle-cell anaemia or β-thalassemia (Connor *et al.*, 1983; Orkin, Markin and Kazazian, 1983) and although technically difficult at present, can be used to identify any known gene mutation directly as well as to identify random DNA sequence polymorphisms not accessible by other methods. However, it should be noted that this approach is mutation specific, and where different mutations have been shown to cause the same disease, a different oligonucleotide or restriction enzyme will be required for each individual mutation. It will therefore be necessary to define which of the possible mutations is occurring in the family under study.

The use of linkage for prediction assumes that heterogeneity in the disease does not exist or, in other words, that linkage information obtained from a reference panel of families applies equally to any new families presenting for counselling. The availability of gene-specific probes for the disease should allow one to test much more rigorously for heterogeneity, and also to investigate new mutations that arise (as is common for Duchenne muscular dystrophy) where the linkage approach is by definition not applicable.

**Approaches to the isolation of human disease genes**

We have seen how the techniques of recombinant DNA and somatic-cell genetics can be applied to mapping human chromosomes and providing genetic markers closely linked to a disease locus. In the final section of this

article we shall consider some of the possibilities for identifying and isolating genes responsible for the many inherited human diseases, the molecular basis of which is not understood.

## LIMITS OF THE LINKAGE APPROACH

The material available for studying the linkage of markers to genetic diseases in man is limited by the difficulty of obtaining suitably structured families. Because many of these diseases are quite rare, one cannot generally sample more than one or two hundred informative meiotic events, and this sets the limit on the accuracy of the estimate of linkage distance. This situation is totally different from that in the study of bacterial, *Drosophila* or even mouse genetics, where it is possible to set up specific matings and analyse as many progeny as necessary. In human genetics the confidence limits on the estimates of genetic distance are generally quite wide and in practice one will not usually be able to define a linkage closer than about 1–2 cM.

A consideration of the average recombination frequency and overall length of the human genome shows that 1 cM corresponds to about 1 million base-pairs (1 Mb). It will not therefore be possible to find a disease gene by sequencing the DNA starting at a linked marker; new approaches to isolating and analysing DNA fragments a million or more base pairs long will be required.

## ISOLATING LARGE DNA FRAGMENTS

When the application of genetic linkage has provided closely linked markers on each side of a disease gene, a region of the chromosome within which the gene must lie will be defined. The length of this region will be uncertain (for the reasons described above) but could be up to 10 Mb.

This is obviously much too large to be cloned intact using conventional techniques: covering the region by chromosome walking (*see* pages 324–325) would involve a prohibitively large number of steps.

Three types of approach for isolating DNA fragments in this size range have been suggested. The first is a development of the somatic-cell hybridization technique and aims to isolate specific sub-chromosomal fragments, rather than intact human chromosomes, in a rodent-cell background. In the first stage of the procedure, a selectable marker (such as a gene coding for a drug resistance) is introduced into a human-cell line, or into a hybrid line containing limited human material, by means of a retrovirus vector or by DNA transformation. The marker should integrate at random into one of the chromosomes of the cell. Fragmentation of the chromosomes can then be induced by X-ray treatment, and the fragments transferred to a rodent host cell by fusion; alternatively, metaphase chromosomes can be prepared from the donor cell line and precipitated on to the host cells, which will then take up fragments of the donor chromosome. By applying selection for the marker

gene that was introduced at the first stage, the marked donor chromosome fragment can be maintained in the rodent host cell. Because it is not possible at present to direct the selectable marker to a specific chromosomal site, the hybrid cell lines produced by this technique will contain random fragments of donor DNA. A particular line that carries the chromosomal fragment of interest can be identified by screening a large number of lines, for example by DNA hybridization using as probe a cloned marker that is closely linked to the disease gene. Weis *et al.* (1984) have recently applied this technique to the isolation of the mouse major histocompatibility region, using a monoclonal antibody specific for one of the cell-surface antigens coded by this region to identify the required hybrid cells. It has been estimated that these procedures result in the transfer of DNA fragments with lengths of approximately 3–30 Mb; this is of the right order of magnitude for our purposes.

The second procedure for isolating large DNA fragments is a biochemical one, and uses a recently developed form of agarose gel electrophoresis to achieve separations of fragments longer than 100 kb (which are not resolved by conventional electrophoresis). The method is known as pulsed-field or orthogonal-field alteration gel electrophoresis, and the apparatus used contains two pairs of electrodes in a square arrangement; electrical current is switched back and forth between the two pairs of electrodes to create an electrical field, the direction of which is constantly altering. The migration of the DNA molecules depends upon the rate at which they can alter their configuration in the gel in response to the changes in the electrical field; this parameter is critically dependent on the length of the DNA molecule. Pulsed-field electrophoresis has been used by Schwartz and Cantor (1984) and by Carle and Olsen (1984) to separate yeast chromosomes, which range in size from 30 to 2000 kb. In principle at least, DNA from mammalian cells could be cut into fragments of this size, separated, the fragments of interest identified by molecular hybridization and isolated from the gel. Use of a somatic-cell hybrid containing only one human chromosome would simplify the problem of resolving the required human DNA fragments from the many others that would be produced.

Thirdly, recent reports have shown that it is feasible to isolate a small region of a chromosome by microdissection of a large number of metaphase spreads. This has been achieved using both *Drosophila* and mouse chromosomes (Scalenghe *et al.*, 1981; Rohme *et al.*, 1984) and the method could, in principle, be applied also to human material. The procedures used for microdissection and subsequent 'micro-cloning' of the isolated fragments are technically very demanding.

Having isolated by some means a chromosomal fragment a few Mb long, within which the disease gene is thought to be located, it would be a relatively simple matter to construct a complete library of this region in cosmid or phage vectors. For example a 2 Mb fragment could be completely covered by about 100 overlapping cosmid clones. Using this 'mini-library', a complete map of the region could be constructed using restriction enzyme sites as markers and probes could be selected from any particular site for genetic analysis.

CLUES TO FINDING THE DISEASE GENE

*Linkage disequilibrium*

This phenomenon has been described on pages 335–336. In the case of a disease where the frequency of new mutations is very low (such as myotonic dystrophy or Huntington's chorea) it may be possible to use the phenomenon of disequilibrium to find genetic markers that are extremely close to the disease locus — closer than could be defined using conventional linkage analysis. If in the population under study, all (or the majority) of affected individuals are descendants of the person in whom the disease mutation first occurred, then polymorphic loci that are extremely close to the disease gene may be in linkage disequilibrium. The alleles at such a locus will show markedly different frequencies between the population as a whole and the population of affected individuals.

*Chromosomal rearrangements*

In some genetic diseases, *de novo* chromosomal translocations are associated with the appearance of the disease in an individual. An example is the occurrence of X–autosome translocations in sporadic cases of girls with Duchenne muscular dystrophy (Worton *et al.*, 1984). However, no such rearrangements associated with myotonic dystrophy have been reported. In such a case, a mini-library of cloned DNA segments could be prepared from the relevant chromosomal region of the affected individual, as described above, and the structure of clones derived from it compared with those from a similar library made using normal DNA. The position of the translocation should be revealed as a disruption of the normal map, and will then pinpoint the disease gene. This type of approach has been used to study oncogenes, the expression of which is altered by specific translocations that are observed in tumour cells (Davies, Malcolm and Rabbitts, 1984).

*Expressed sequences*

If a tissue in which the disease gene is expressed can be chosen with a reasonable degree of confidence, then mRNA from it could be used to screen the clones in the mini-library to identify expressed sequences. The number of genes that are active in a given tissue and map within our defined region of the chromosome, is likely to be quite small. The chances of success depend on choosing the right tissue for isolation of mRNA, and on the disease gene being sufficiently highly expressed to be detectable.

Another method for detecting expressed sequences is to use open reading frame vectors (Weinstock *et al.*, 1983) in which the incorporation of a DNA fragment containing a protein-coding region (lacking translational stop codons) leads to the production of an enzymatically detectable hybrid protein. The technique should enable one to identify all the potential genes within a region of the chromosome, regardless of which tissue they are

normally expressed in; in addition, the production of a fusion protein should allow the isolation of monoclonal antibodies specific for the normal gene product.

## TESTING CANDIDATE GENES

Once a number of candidate genes have been identified and isolated, their structure and function will need to be investigated in order to decide which of them is the disease gene itself. Some of the experiments one could do would be as follows.

### *Patterns of expression*

The cloned genes could be used as probes to study the production of the corresponding mRNA in tissues from normal and affected individuals. If the disease mutation causes a reduction or complete absence of the mRNA then this approach should be profitable. In the case of thalassaemias, several different molecular pathologies have been found, many of which would be detectable in this way. Similar heterogeneity may exist for other diseases. As is the case with all techniques based on mRNA, correct choice of tissue is very important; however, the variation in expression of a candidate gene between tissues may give clues as to its normal function.

### *Gene sequencing and prediction of function*

The candidate gene could be cloned from both normal and affected individuals and the structures compared by sequencing. The base sequence can be translated into a protein sequence, and an attempt made to predict its function, based on comparisons with other proteins of known function. Although many sequence differences would be found that are part of the normal individual variation, with no significance for the disease, some changes that are correlated with the disease may be apparent. Even so, correlations may not be functionally significant as they could be the result of linkage disequilibrium (see above).

### *Functional testing and isolation of gene products*

A cloned gene can be made to express its product *in vitro*, by introduction into cultured cells, or in animals by introduction into stem cells (Williams *et al.*, 1984). Expression in an *in vitro* system would allow antibodies to the gene product to be raised, and thus open up the field for investigation at the protein level in a way that has not previously been possible.

Although animal or cultured-cell models for human diseases are not generally available, it is possible that interesting phenotypes could arise if the candidate genes were introduced into, and made to express in, normal cells. Increasingly more sophisticated procedures are being developed to do this.

There are, therefore, many possibilities for tackling the problem of how to locate, identify and isolate a disease gene with unknown biochemical properties. Some of these techniques are already in use; others are still being developed. As investigators in this field ourselves, we are excited and encouraged by the rapid progress that has recently been made, as a result of the rational application of new technologies to previously intractable problems.

**Acknowledgements**

Many thanks to Professor Peter Harper and Dr Nick Thomas for reading the manuscripts and for helpful comments, to Dr A. C. Chandley for Figure 2 and to Mrs G. Gulliford for excellent typing.

**References**

ALDRIDGE, J., KUNKEL, L., BRUNS, G., TANTRAVAHI, U., LALANDE, M., BREWSTER, T., MOREAU, E., WILSON, M., BROMLEY, W., RODERICK, T. AND LATT, S.A. (1984). A strategy to reveal high frequency RFLPs along the human X chromosome. *American Journal of Human Genetics* **36**, 546–564.

BALL, S., BUCKTON, K.E., CORNEY, G., FEY, G., MONTEIRO, M., NOADES, J.E., PYM, B., ROBSON, E.B. AND TIPPETT, P. (1984). Mapping studies with peptidase D (Pep D). *Cytogenetics and Cell Genetics* **37**, 411.

BARKER, D., SCHAFER, M. AND WHITE, R. (1984). Restriction sites containing CpG show a higher frequency of polymorphism in human DNA. *Cell* **36**, 131–138.

BASTIE-SIGEAC, F. AND LUCOTTE, G. (1983). Optimal use of restriction enzymes in the analysis of human DNA polymorphism. *Human Genetics* **63**, 162–165.

BELL, J. (1948). Dystrophia myotonica and allied diseases. In *Treasury of Human Inheritance*, Volume 9, pp. 343–410. Cambridge University Press.

BERG, K., BORRESEN, A-L., HEIBERG, A. AND MAARTMANN-MOE, K. (1984). Study of the linkage relationship between *ApoE* and familial hypercholesterolemia with xanthomatosis. *Cytogenetics and Cell Genetics* **37**, 418.

BISHOP, J.O. AND JONES, K.W. (1982). Chromosomal localisation of human haemoglobin structural genes. *Nature* **240**, 149–150.

BODMER, W.F. (1980). The HLA system and disease. *Journal of the Royal College of Physicians of London* **14**, 43–50.

BOTSTEIN, D., WHITE, R.L., SKOLNICK, M. AND DAVIES, R.W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* **32**, 314–331.

BROOK, J.D., SHAW, D.J., MEREDITH, L., BRUNS, G.A.P. AND HARPER, P.S. (1984). Localisation of genetic markers and orientation of the linkage group on chromosome 19. *Human Genetics* **68**, 282–285.

CARLE, G.F. AND OLSON, M.V. (1984). Separation of chromosomal DNA molecules from yeast by orthogonal-field-alteration gel electrophoresis. *Nucleic Acids Research* **12**, 5647–5664.

CARRANO, A.V., GRAY, J.W., LANGLOIS, R.G., BURKHART-SCHULTZ, K.J. AND VAN DILLA, M.A. (1979). Measurement and purification of human chromosomes by flow cytometry and sorting. *Proceedings of the National Academy of Sciences of the United States of America* **76**, 1382–1384.

CASKEY, J.H., JONES, C., MILLER, Y.E. AND SELIGMAN, P.A. (1983). Human ferritin gene is assigned to chromosome 19. *Proceedings of the National Academy of Sciences of the United States of America* **80**, 482–486.

Cavenee, W., Leach, R., Mohandas, T., Pearson, P. and White, R. (1984). Isolation and regional localisation of DNA segments revealing polymorphic loci from human chromosome 13. *American Journal of Human Genetics* **36**, 10–24.

Cepko, C.L., Roberts, B.E. and Mulligan, R.C. (1984). Construction and applications of a highly transmissible murine retrovirus shuttle vector. *Cell* **37**, 1053–1062.

Chandler, M.E. and Yunis, J.J. (1978). A high resolution *in situ* hybridisation technique for the direct visualisation of labelled G-banded early metaphase and prophase chromosomes. *Cytogenetics and Cell Genetics* **22**, 352–356.

Chang, J.C. and Kan, Y.W. (1982). A sensitive new prenatal test for sickle-cell anemia. *New England Journal of Medicine* **307**, 30–32.

Chaplin, D.D., Woods, D.E., Whitehead, A.S., Goldberger, G., Colten, H.R. and Seidman, J.G. (1983). Molecular map of the murine S region. *Proceedings of the National Academy of Sciences of the United States of America* **80**, 6947–6951.

Cockburn, A.F., Howells, A.J. and Whitten, M.J. (1984). Recombinant DNA technology and genetic control of pest insects. In *Biotechnology and Genetic Engineering Reviews* (G.E. Russell, Ed.), volume 2, pp. 69–99. Intercept, Ponteland, Newcastle upon Tyne.

Collard, J.G., Tulp, A., Stegeman, J. and Boezeman, J. (1981). Separation of human metaphase chromosomes at neutral pH by velocity sedimentation at 20 times gravity. *Experimental Cell Research* **133**, 341–346.

Collins, F.S. and Weissman, S.M. (1984). Directional cloning of DNA fragments at a large distance from an initial probe: a circularisation method. *Proceedings of the National Academy of Sciences of the United States of America* **81**, 6812–6816.

Conner, B.J., Reyes, A.A., Morin, C., Itakura, K., Teplitz, R.L. and Wallace, R.B. (1983). Detection of sickle cell $\beta^s$-globin allele by hybridisation with synthetic oligonucleotides. *Proceedings of the National Academy of Sciences of the United States of America* **80**, 278–282.

Coulondre, C., Miller, J.H., Farabough, P.J. and Gilbert, W. (1978). Molecular basis of base substitution hotspots in *Escherichia coli*. *Nature* **274**, 775–780.

Davies, K.E., Young, B.D., Elles, R.G., Hill, M.E. and Williamson, R. (1981). Cloning of a representative genomic library of the human X chromosome after sorting by flow cytometry. *Nature* **293**, 374–376.

Davies, K.E., Jackson, J., Williamson, R., Harper, P.S., Ball, S., Sarfarazi, M., Meredith, L. and Fey, G. (1983). Linkage analysis of myotonic dystrophy and sequences on chromosome 19 using a cloned complement 3 gene probe. *Journal of Medical Genetics* **20**, 259–263.

Davies, K.E., Shull, S., Donald, J.A., McFadden, J.J., Young, B.D., Carritt, B., Boylston, A., Williamson, R. and Humphries, S.E. (1984). The isolation of DNA linkage markers for chromosome 19. *Cytogenetics and Cell Genetics* **37**, 448.

Davis, M., Malcolm, S. and Rabbitts, T.H. (1984). Chromosomal translocation can occur on either side of the c-myc oncogene in Burkitt lymphoma cells. *Nature* **308**, 286–288.

Deaven, L. (1984). Construction of human chromosome-specific DNA libraries from flow-sorted chromosomes. *American Journal of Human Genetics* **36**, 1345.

Drayna, D., Davies, K., Hartley, D., Mandel, J-L., Camerino, G., Williamson, R. and White, R. (1984). Genetic mapping of the human X chromosome by using restriction fragment length polymorphisms. *Proceedings of the National Academy of Sciences of the United States of America* **81**, 2836–2839.

Eiberg, H., Mohr, J., Nielsen, L.S. and Simonsen, N. (1983). Genetics and linkage relationships of the C3 polymorphism: discovery of the C3-Se linkage and assignment of LES-C3-DM-Se-PEPD-Lu synteny to chromosome 19. *Clinical Genetics* **24**, 159–170.

Engel, E., McGee, B.J. and Harris, H. (1969). Recombination and segregation in somatic cell hybrids. *Nature* **223**, 152–155.

EVANS, H.J., BUCKLAND, R.A. AND PARDUE, M.L. (1974). Location of the genes coding for 18s and 28s ribosomal RNA in the human genome. *Chromosoma* **48**, 405–426.

FEINBERG, A.P. AND VOGELSTEIN, B. (1983). A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Analytical Biochemistry* **132**, 6–13.

FENNELL, S.J., MALCOLM, S., WILLIAMSON, R. AND FERGUSON-SMITH, M.A. (1979). Use of chromosomal translocation with in situ DNA hybridisation to confirm localisation of human 5S ribosomal RNA genes. *Journal of Medical Genetics* **16**, 246–253.

FRANCKE, U., BROWN, M.S. AND GOLDSTEIN, J.L. (1984). Assignment of the human gene for the low density lipoprotein receptor to chromosome 19: Synteny of a receptor, a ligand, and a genetic disease. *Proceedings of the National Academy of Sciences of the United States of America* **81**, 2826–2830.

GALL, J.G. AND PARDUE, M.L. (1969). Formation and detection of RNA-DNA hybrid molecules in cytological preparations. *Proceedings of the National Academy of Sciences of the United States of America* **63**, 378–383.

GEDDE-DAHL, T., OLAISEN, B., TEISBERG, P., CONSTANT-WILHELMY, M., MEVAG, B. AND HELLAND, R. (1984). The locus for apolipoprotein E (apoE) is close to the Lutheran (Lu) blood group locus on chromosome 19. *Human Genetics* **67**, 178–182.

GERALD, P.S. AND BRUNS, G.A. (1978). Genetic determinants of viral susceptibility. *Birth Defects* **14**, (6A), 1–7.

GOODE, M.E., LEDBETTER, D.H. AND DAIGER, S.P. (1984). Failure of *in situ* hybridisation to detect a primary copy of λH3. *American Journal of Human Genetics* **36**, 138s.

GROSVELD, F.G., LUND, T., MURRAY, E.J., MELLOR, A.L., DAHL, H.H.M. AND FLAVELL, R.A. (1982). The construction of cosmid libraries which can be used to transform eukaryotic cells. *Nucleic Acids Research* **10**, 6715–6732.

GRUBB, R. (1953). Zur Genetik des Lewis-systems. *Naturwissenschaften* **21**, 560–561.

GUBLER, U. AND HOFFMAN, B.J. (1983). A simple and very efficient method for generating cDNA libraries. *Gene* **25**, 263–269.

GUSELLA, J.F., KEYS, K., VARSANYI-BREINER, A., KAO, F-T., JONES, C., PUCK, T.T. AND HOUSMAN, D. (1980). Isolation and localisation of DNA segments from specific human chromosomes. *Proceedings of the National Academy of Sciences of the United States of America* **77**, 2829–2833.

GUSELLA, J.F., JONES, C., KAO, F.T., HOUSMAN, D. AND PUCK, T.T. (1982). Genetic fine-structure mapping in human chromosome 11 by use of repetitive DNA sequences. *Proceedings of the National Academy of the United States of America* **79**, 7804–7808.

GUSELLA, J.F., WEXLER, N.S., CONNEALLY, P.M., NAYLOR, S.L., ANDERSON, M.A., TANZI, R.E., WATKINS, P.C., OTTINA, K., WALLACE, M.R., SAKAGUCHI, A.I., YOUNG, A.B., SHOULSON, I., BONILLA, E. AND MARTIN, B. (1983). A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* **306**, 234–238.

GUSELLA, J.F., TANZI, R.E., ANDERSON, M.A., HOBBS, W., GIBBONS, K., RASCHTCHIAN, R., GILLIAM, T.C., WALLACE, M.R., WEXLER, N.S. AND CONNEALLY, P.M. (1984). DNA markers for nervous system diseases. *Science* **225**, 1320–1326.

HARPER, M.E., ULLRICH, A. AND SAUNDERS, G.F. (1981). Localisation of the human insulin gene to the distal end of the short arm of chromosome 11. *Proceedings of the National Academy of Sciences of the United States of America* **78**, 4458–4460.

HARPER, P.S. (1979). *Myotonic Dystrophy*. W.B. Saunders Company, London.

HARPER, P.S., RIVAS, M.L., BIAS, W.B., HUTCHINSON, J.R., DYKEN, P.R. AND MCKUSICK, V.A. (1972). Genetic linkage confirmed between the locus for myotonic dystrophy and the ABH-secretion and the Lutheran blood group loci. *American Journal of Human Genetics*, **24**, 310–316.

HARRIS, H. (1980). *The Principles of Human Biochemical Genetics*. 3rd edn.

Elsevier/North-Holland Publishing Company, Amsterdam.

Harris, H. and Watkins, J.F. (1965). Hybrid cells derived from mouse and man: artificial heterokaryons of mammalian cells from different species. *Nature* **205**, 640–646.

Helfman, D.M., Feramisco, J.R., Fiddes, J.C., Thomas, G.P. and Hughes, S.H. (1983). Identification of clones that encode chicken tropomyosin by direct immunological screening of a cDNA expression library. *Proceedings of the National Academy of Sciences of the United States of America* **80**, 31–35.

Humphries, S.E., Berg, K., Gill, L., Cumming, A.M., Robertson, F.W., Stalenhoef, A.F.H., Williamson, R. and Borrensen, A.L. (1984). The gene for apoliprotein C-II is closely linked to the gene for apolipoprotein E on chromosome 19. *Clinical Genetics* **26**, 389–396.

ISCN (1981). An international system for human cytogenetic nomenclature. *Birth Defects* **14**, No. 8.

Ish-Horowicz, D. and Burke, J.F. (1981). Rapid and efficient cosmid cloning. *Nucleic Acids Research* **9**, 2989–2998.

Jeffreys, A.J. (1979). DNA sequence variants in the $^{G}\gamma$-, $^{A}\gamma$-, δ and β globin genes of Man. *Cell* **18**, 1–10.

John, H.A., Birnstiel, M.L. and Jones, K.W. (1969). RNA–DNA hybrids at the cytological level. *Nature* **223**, 582–587.

Kan, Y.W. and Dozy, A.M. (1978). Polymorphism of DNA sequence adjacent to human β-globin structural gene: Relationship to sickle mutation. *Proceedings of the National Academy of Sciences of the United States of America* **75**, 5631–5635.

Kan, Y.W., Lee, K.Y., Furbetta, M., Angius, A. and Cao, A. (1980). Polymorphism of DNA sequence in the β-globin gene region. *New England Journal of Medicine* **302**, 185–188.

Kaneda, Y., Yoshida, M.C., Kohno, K., Uchida, T. and Okada, Y. (1984). Chromosomal assignment of the gene for human elongation factor 2. *Proceedings of the National Academy of Sciences of the United States of America* **81**, 3158–3162.

Langlois, R.G., Yu, R.C., Gray, J.W. and Carrano, A.V. (1982). Quantitative karyotyping of human chromosomes by dual beam flow cytometry. *Proceedings of the National Academy of Sciences of the United States of America* **79**, 7876–7880.

Lau, Y.F. and Kan, Y.W. (1983). Versatile cosmid vectors for the isolation, expression and rescue of gene sequences: studies with the human α-globin gene clusters. *Proceedings of the National Academy of Sciences of the United States of America* **80**, 5225–5229.

Law, M.L., Davidson, J.N. and Kao, F.T. (1982). Isolation of a human repetitive sequence and its application to regional chromosome mapping. *Proceedings of the National Academy of Sciences of the United States of America* **79**, 7390–7394.

Lawn, M.M., Fritsch, E.F., Parker, R.C., Blake, G. and Maniatis, T. (1978). The isolation and characterisation of linked δ and β-globin genes from a cloned library of human DNA. *Cell* **15**, 1157–1174.

Lebo, R.V., Carrano, A.V., Bukhart-Schult, K., Dozy, A.M., Yu, L.C. and Kan, Y.W. (1979). Assignment of human β-, γ- and δ-globin genes to the short arm of chromosome 11 by chromosome sorting and DNA restriction enzyme analysis. *Proceedings of the National Academy of Sciences of the United States of America* **76**, 5804–5808.

Lebo, R.V., Kan, Y.W., Cheung, M.C., Carrano, A.V., Yu, L.C., Chang, J.C., Cordell, B. and Goodman, H.M. (1982). Assigning the polymorphic human insulin gene to short arm of chromosome 11 by chromosome sorting. *Human Genetics* **60**, 10–15.

Lebo, R.V., Gorin, F., Fletterick, R.J., Kao, F.T., Cheung, M.C., Bruce, B.D. and Kan, Y.W. (1984). High-resolution chromosome sorting and DNA spot-blot analysis assign McArdles syndrome to chromosome 11. *Science* **225**, 57–59.

Littlefield, J.W. (1964). Selection of hybrids from matings of fibroblasts *in vitro* and their presumed recombinants. *Science* **145**, 709–710.

McAlpine, P.J., Mohandas, T., Ray, M., Wang, H. and Hamerton, J.L. (1976). Assignment of the peptidase D gene locus (PEPD) to chromosome 19 in man. *Cytogenetics and Cell Genetics* **16**, 204–205.

Medrano, L. and Green, H. (1974). A uridine-kinase-deficient mutant of 3T3 and a selective method for cells containing the enzyme. *Cell* **1**, 23–26.

Miller, D.A., Miller, O.J., Dev, V.G., Hashmi, S., Tantravahi, R., Medrano, L. and Green, H. (1974). Human chromosome 19 carries a polio virus receptor gene. *Cell* **1**, 167–173.

Mohr, J. (1951). Estimation of linkage between the Lutheran and the Lewis blood groups. *Acta pathologica et microbiologica scandinavica*, **29**, 339–344.

Mohr, J. (1954). *A Study of Linkage in Man*. Munksgaard, Copenhagen.

Morton, N.E. (1955). Sequential tests for the detection of linkage. *American Journal of Human Genetics* **7**, 277–318.

Murray, J.M., Davies, K.E., Harper, P.S., Meredith, L., Mueller, C.R. and Williamson, R. (1982). Linkage relationship of a cloned DNA sequence on the short arm of the X chromosome to Duchenne muscular dystrophy. *Nature* **300**, 69–71.

Murray, N.E. (1983). Phage lambda and molecular cloning. In *Lambda II* (R. Hendrix, J.W. Roberts, F.W. Stahl and R.A. Weisberg, Eds), pp. 395–432. Cold Spring Harbor Laboratory, New York.

Myklebost, O., Rogne, S., Olaisen, B., Gedde-Dahl, T. and Prydz, H. (1984a). The locus for apolipoprotein CII is closely linked to the apolipoprotein E locus on chromosome 19 in man. *Human Genetics* **67**, 309–312.

Myklebost, O., Williamson, R., Markham, A.F., Myklebost, S.R., Rogers, J., Woods, D.E. and Humphries, S.E. (1984b). The isolation and characterisation of cDNA clones for human apolipoprotein CII. *Journal of Biological Chemistry* **259**, 4401–4404.

Naylor, S.L. and Shows, T.B. (1979). Branched chain aminotransferase genes (BCT1 and BCT2) assigned to human chromosomes 12 and 19 using α-keto acid selection media. *Cytogenetics and Cell Genetics* **25**, 191.

Naylor, S.L., Chin, W.W., Goodman, H.M. and Sakaguchi, A.Y. (1984). Chromosomal locations of human glycoprotein hormone genes; Alpha subunit gene on 6 and chorionic gonadotrophin beta-leuteinizing hormone beta gene complex on 19. *Cytogenetics and Cell Genetics* **37**, 552.

Neve, R.L., Bruns, G.A.P., Dryja, T.P. and Kurnit, D.M. (1983). Retrieval of human DNA from rodent-human genomic libraries by a recombination process. *Gene* **23**, 343–354.

O'Brien, T., Ball, S., Sarfarazi, M., Harper, P.S. and Robson, E.B. (1983). Genetic linkage between the loci for myotonic dystrophy and peptidase D. *Annals of Human Genetics* **47**, 117–121.

Okayama, H. and Berg, P. (1982). High-efficiency cloning of full-length cDNA. *Molecular and Cellular Biology* **2**, 161–170.

Old, R.W. and Primrose, S.B. (1981). *Principles of Gene Manipulation*. Blackwell Scientific Publications, Oxford.

Orkin, S.H., Markham, A.F. and Kazazian, H.H. (1983). Direct detection of the common Mediterranean B thalassemia gene with synthetic probes: An alternative approach for prenatal diagnosis. *Journal of Clinical Investigation* **71**, 775.

Phillips, I.R., Shephard, E.A., Rabin, B.R., Ashworth, A., Pike, S.F. and Povey, S. (1985). Structure, evolution and expression of cytochrome P-450 genes. *Heredity*, in press.

Poustka, A., Rackwitz, H.R., Frischauf, A.M., Hohn, B. and Lehrach, H. (1984). Selective isolation of cosmid clones by homologous recombination in *E. coli*. *Proceedings of the National Academy of Sciences of the United States of America* **81**, 4129–4133.

Race, R.R. and Sanger, R. (1975). *Blood Groups in Man*, 6th edn. Blackwell Scientific Publications, London.

Renwick, J.H. (1969). Progress in mapping human autosomes. *British Medical Bulletin* **25**, 65–73.

Renwick, J.H. and Bolling, D.R. (1971). An analysis procedure illustrated on triple linkage of use for prenatal diagnosis of myotonic dystrophy. *Journal of Medical Genetics* **8**, 399–406.

Rettig, W.J., Dracopoli, N.C., Goetzger, T.A., Spengler, B.A., Biedler, J.L., Oettgen, H.F. and Old, L.J. (1984). Somatic cell genetic analysis of human cell surface antigens: Chromosomal assignments and regulation of expression in rodent-human hybrid cells. *Proceedings of the National Academy of Sciences of the United States of America* **81**, 6437–6441.

Rigby, P.W.J., Dieckman, M., Rhodes, C. and Berg, P. (1977). Labelling deoxyribonucleic acid to high specific activity *in vitro* by nick translocation with polymerase. *Journal of Molecular Biology* **113**, 237–251.

Rodeck, C.H. and Morsman, J.M. (1983). First trimester chorion biopsy. *British Medical Bulletin* **39**, 338–342.

Rohme, D., Fox, H., Herrman, B., Frischauf, A.M., Edstrom, J.E., Mains, P., Silver, L.M. and Lehrach, H. (1984). Molecular clones of the mouse t complex derived from microdissected metaphase chromosomes. *Cell* **36**, 783–788.

Scalenghe, F., Turco, E., Edstrom, J.E., Pirotta, V. and Melli, M. (1981). Microdissection and cloning of DNA from a specific region of *Drosophila melanogaster* polytene chromosomes. *Chromosoma* **82**, 205–216.

Schnitzer, T.J., Weiss, R.A., Juricek, D.K. and Ruddle, F.H. (1980). Use of vesicular stomatitis virus pseudotypes to map viral receptor genes: Assignment of RD114 virus receptor gene to human chromosome 19. *Journal of Virology* **35**, 575–580.

Schwartz, D.C. and Cantor, C.R. (1984). Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell* **37**, 67–75.

Seabright, M. (1971). A rapid banding technique for human chromosomes. *Lancet* **ii**, 971–972.

Seed, B. (1983). Purification of genomic sequences from bacteriophage libraries by recombination and selection *in vivo*. *Nucleic Acids Research* **11**, 2427–2445.

Seed, B., Parker, R.C. and Davidson, N. (1982). Representation of DNA sequence in recombinant DNA libraries prepared by restriction enzyme partial digestion. *Gene* **19**, 201–209.

Shaw, D.J., Meredith, A.L., Sarfarazi, M., Huson, S.M., Brook, J.D., Myklebost, O. and Harper, P.S. (1985). The apolipoprotein CII gene: subchromosomal localisation and linkage to the myotonic dystrophy locus. *Human Genetics*, in press.

Skolnick, M.H. and White, R. (1982). Strategies for detecting and characterising restriction fragment length polymorphisms (RFLPs). *Cytogenetics and Cell Genetics* **32**, 58–67.

Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology* **98**, 503–517.

Steinmetz, M., Minard, K., Horvath, S., McNicholas, J., Spelinger, J., Wake, C., Long, E., Mach, B. and Hood, L. (1982). A molecular map of the immune response region from the major histocompatibility complex of the mouse. *Nature* **300**, 35–42.

Weinstock, G.M., Ap Rhys, C., Berman, M.L., Hampar, B., Jackson, D., Silhavy, T.J., Weiseman, J. and Zweig, M. (1983). Open reading frame expression vectors: A general method for antigen production in *Escherichia coli* using protein fusions to β-galactosidase. *Proceedings of the National Academy of Sciences of the United States of America* **80**, 4432–4436.

Weis, J.H., Nelson, D.L., Przyborski, M.J., Chaplin, D.D., Mulligan, R.C., Housman, D.E. and Seidman, J.G. (1984). Eukaryotic chromosome transfer: linkage of the murine major histocompatibility complex to an inserted dominant selectable marker. *Proceedings of the National Academy of Sciences of the United States of America* **81**, 4879–4883.

WEISS, M.C. AND GREEN, H. (1967). Human-mouse cell lines containing partial complements of human chromosomes and functional human genes. *Proceedings of the National Academy of Sciences of the United States of America* **58**, 1104–1111.

WHITEHEAD, A.S., SOLOMON, E., CHAMBERS, S., BODMER, W.F., POVEY, S. AND FEY, G. (1982). Assignment of the structural gene for the third component of human complement to chromosome 19. *Proceedings of the National Academy of Sciences of the United States of America* **79**, 5021–5025.

WHITEHOUSE, H.L.K. (1973). *Towards an Understanding of the Mechanism of Heredity*, 3rd edn, pp. 62–74. Edward Arnold (Publishers) Ltd, London.

WILLIAMS, D.A., LEMISCHKA, I.R., NATHAN, D.G. AND MULLIGAN, R.C. (1984). Introduction of new genetic material into pluripotent haematopoietic stem cells of the mouse. *Nature* **310**, 476–480.

WILLIAMSON, R. (ED.) (1981–3). *Genetic Engineering*, Volumes 1–4. Academic Press, London.

WOO, S.L.C., LIDSKY, A.S., GUTTLER, F., CHANDRA, T. AND ROBSON, K.J.H. (1983). Cloned human phenylalanine hydroxylase gene allows prenatal diagnosis and carrier detection of classical phenylketonuria. *Nature* **306**, 151–155.

WORTON, R.G., DUFF, C., SYLVESTER, J.E., SCHMICKEL, R.D. AND WILLARD, H.F. (1984). Duchenne muscular dystrophy involving translocation of the DMD gene next to ribosomal RNA genes. *Science* **224**, 1447–1449.

WORWOOD, M., BROOK, J.D., CRAGG, S.J., HELLKUHL, B., JONES, B.M., PERERA, P., ROBERTS, S.H. AND SHAW, D.J. (1985). Assignment of human ferritin genes to chromosomes 11 and 19q13.3→19qter. *Human Genetics* **69**, 371–374.

WYMAN, A.R. AND WHITE, R. (1980). A highly polymorphic locus in human DNA. *Proceedings of the National Academy of Sciences of the United States of America* **77**, 6754–6758.

YOUNG, B.D., FERGUSON-SMITH, M.A., SILLAR, R. AND BOYD, E. (1981). High-resolution analysis of human peripheral lymphocyte chromosomes by flow cytometry. *Proceedings of the National Academy of Sciences of the United States of America* **78**, 7727–7731.

YOUNG, R.A. AND DAVIS, R.W. (1983). Efficient isolation of genes by using antibody probes. *Proceedings of the National Academy of Sciences of the United States of America* **80**, 1194–1198.