9

The Molecular Genetics of Human Monogenic Diseases

S.P. BALL, S.J. KENWRICK AND K.E. DAVIES

Nuffield Department of Clinical Medicine, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK

Introduction

Until recently, the study of inherited human disease depended on the identification and characterization of a relevant gene product. Recombinant DNA technology now permits the investigation of human disease at the molecular level, even in cases where the basic biochemical defect is unknown. Many advances in the study of the genetic basis for disease are being made as more mutant loci are being mapped to chromosomes and the DNA sequences responsible isolated. This should lead to accurate diagnosis and improvement in the treatment of genetic disease.

Genes and inheritance

During the last century, Mendel studied the inheritance of several contrasting pairs of characteristics in the garden pea, *Pisum sativum*. From his observations he established, first, that these characteristics were determined by pairs of hereditary elements or genes. Secondly, these characteristics were seen to segregate, that is each parent passes on to each offspring one of a pair of these genes. The offspring then had a new combination of pairs of these genes, one from each parent. During gamete formation each pair of characters was seen to assort independently of the other pairs. This is Mendel's law of independent segregation.

Abbreviations: apoE, apoliproprotein E; BMD, Becker muscular dystrophy; bp, base pairs; C3, third complement component; CF, cystic fibrosis; cDNA, complementary DNA; DMD, Duchenne muscular dystrophy; HAT, hypoxanthine, aminopterin, thymidine; HD, Huntingdon's disease; HDL, high-density lipoprotein; HPRT, hypoxanthine phosphoribosyltransferase; kb, kilobases; LDL, low-density lipoprotein; Mbp, mega-base pairs; mRNA, messenger RNA; PH, phenylhydroxylase; PIC, polymorphism information content; PKU, phenylketonuria; RFLP, restriction fragment length polymorphism; TK, thymidine kinase.

Biotechnology and Genetic Engineering Reviews — Vol. 3, September 1985
0264-8725/85/03/275–35$10.00 + $0.00

SEGREGATION OF GENES ON THE SAME CHROMOSOME

At the beginning of this century, Sutton (1903) studied the behaviour of the pairs of chromosomes during reduction division or meiosis, in the development of the sex cells or gametes. He realized that the chromosomes were the carriers of the Mendelian characters, or genes. The chromosomes were observed to segregate independently, which would explain the independent segregation of factors on separate chromosomes. However, because the number of inherited factors exceeds the number of chromosome pairs, then there must be several genes carried on each chromosome. It was De Vries (1910) who pointed out that there must be exchanges of genes between the homologous chromosomes while they were paired at meiosis, in order to explain the independent assortment of genes on the same chromosome.

NON-INDEPENDENT SEGREGATION: LINKAGE

An exception to this rule of independent assortment was discovered by Bateson, Saunders and Punnett (1905a, b). They observed two pairs of contrasting factors in the sweet pea, *Lathyrus odoratus*, which did not segregate independently. Many more of the parental types than recombinant types were counted among the offspring, thus showing non-independent assortment of the genes or 'linkage'.

Another example of linkage was demonstrated in 1911 by Morgan. He studied two segregating characters of *Drosophila*, white eye and miniature wing, which were both known to be carried on the X chromosome. Females who had inherited together from one parent both the white eye allele and the miniature wing allele on one X chromosome, and from the other parent the corresponding wild-type or normal alleles of these loci on the other X chromosome, were mated with normal male flies. Significantly more than the expected half of the offspring showed parental phenotypes having either white eyes together with miniature wings, or were normal for both these characteristics. Considerably less offspring than expected were recombinants between these two loci, having either white eyes and normal wings or normal eyes and miniature wings. This deviation from the expected ratio of phenotypes indicates that assortment between these two loci was not occurring randomly. Because of the location of these two loci on the sex chromosome, Morgan and Cattell (1912) realized that there must be some genetic exchange between the two X chromosomes in order to account for the occurrence of any recombinant offspring as had been suggested by De Vries (1910). The term 'crossing-over' was introduced to describe this exchange of alleles between homologous chromosomes.

LINEAR ARRANGEMENT OF GENES

The concept of the linear arrangement of genes on the chromosomes was proposed by Sturtevant (1913) from his studies on the X-borne markers in *Drosophila*. He also observed that the frequency of crossing-over or recom-

bination was approximately constant between any particular pair of genes, and measures the distance between them. It was also apparent that certain factors known to lie on the X chromosome undergo random assortment at meiosis, from which it was deduced that these factors must be so far apart that at least one crossover was bound to occur between them so that they appear to segregate independently of each other. The term 'recombination fraction' was introduced to describe the proportion of the total number of offspring which were recombinants and was used to measure the genetic distance. The units of measurement were called 'Morgans' where 1M represents 100% recombination between the loci. Fifty per cent recombination or 50 centimorgans (cM) represents independent assortment between the two loci.

MENDELIAN INHERITANCE IN MAN

It was Garrod in 1902 who first applied Mendel's concept to a human character. From studies of patients with alkaptonuria and their families, he established that the symptomatic excretion of homogentisic acid in the urine is a qualitative trait. Furthermore, this abnormality was detected among siblings of the patients but their parents were normal. This is typical of the inheritance of recessive disorders where it is manifested only in individuals who carry the aberrant gene on both chromosomes, that is, they are homozygous. Parents of alkaptonuria patients are heterozygous, that is they carry one copy of the aberrant gene and, on the homologous chromosome, a normal copy of the gene and so do not express the biochemical abnormality. Many more of these 'inborn errors of metabolism' have been discovered, such as albinism, and the majority of them show similar recessive inheritance patterns. On the other hand, certain inherited conditions were seen to be passed on from parent to child without skipping generations. Examples of diseases showing this dominant pattern of inheritance are myotonic dystrophy and Huntington's disease (chorea; HD). In these conditions, affected individuals are heterozygous, meaning that only one copy of the aberrant gene is necessary for the disease to be manifested.

A third pattern of inheritance is observed for diseases carried on the X chromosome. The majority of these are recessive and, consequently, females who are heterozygous carry the disease gene on one chromosome only and do not manifest the condition. However, males who are hemizygous for X chromosome loci, that is, they have only one copy of X-linked genes, will manifest recessive X-linked diseases because they do not have a normal gene to compensate. Examples of such X-linked diseases are Duchenne muscular dystrophy (DMD) and factor-VIII-deficient haemophilia A.

Gene-specific probes

GENE LIBRARIES

If a defect in a particular gene product is known to be responsible for a given phenotype, this information can be employed towards isolating gene-specific

probes. The DNA sequences concerned can be cloned into plasmid or phage vectors and propagated indefinitely in bacterial hosts to allow their detailed characterization. Depending on the level of expression of the gene, different strategies have been employed. In cases where the number of gene transcripts in a particular cell are high, the messenger RNA (mRNA) can be purified and complementary sequences produced with the enzyme reverse transcriptase. The resultant complementary DNA (cDNA) can be cloned into an appropriate vector (Maniatis, Fritsch and Sambrook, 1982). This approach has been used successfully to isolate globin gene sequences from erythropoietic cells where the globin mRNA constitutes approximately 90% of the mRNA population (Wilson *et al.*, 1978). Most mRNA molecules in the cell are present in low abundance, that is at less than 0·1% of the mRNA pool, and cannot be purified easily. Under these circumstances, gene-specific probes may be isolated by screening a library of cloned DNA sequences that has been developed, either using all of the cDNA molecules generated by reverse transcription of total cellular mRNA (a cDNA library), or using total cellular DNA (a genomic library). A cDNA library will contain very few repetitive DNA sequences compared with a genomic library (Crampton, Davies and Knapp, 1981) and is therefore a very useful source of single-copy gene-specific probes. A genomic library, on the other hand, contains all the sequences of the genome, coding and non-coding, and can be used to study DNA sequences that may be involved with gene expression but are not represented in the mRNA population (*see also* Cockburn, Howells and Whitten, 1984).

IDENTIFICATION OF GENE PROBES

Various methods have been adopted for the screening and identification of recombinant clones in DNA libraries. The strategy employed by Brennand *et al.* (1982) to isolate the gene for the enzyme hypozanthine phosphoribosyltransferase (HPRT, EC 2.4.2.8) exploited the difference in HPRT mRNA content observed for two cell lines: one that overproduces and one that underproduces the enzyme. A library of cDNA recombinant molecules was screened using cDNAs from the two cell lines. Clones that hybridized more strongly to cDNA from overproducing cells were considered potential HPRT recombinants. In these experiments, the mRNA preparation used was enriched for HPRT mRNA by sucrose gradient fractionation. Relevant fractions were identified by cell-free translation of the mRNA followed by detection of HPRT using a radioimmune assay.

The amino-acid sequence of a protein can be used to develop synthetic oligonucleotides with which to screen DNA libraries. This approach has been used recently to isolate sequences homologous to part of the human gene coding for factor VIII, the clotting protein that is deficient or abnormal in patients with haemophilia A (Gitschier *et al.*, 1984; Toole *et al.*, 1984). In cases where the appropriate gene has been isolated for another species, this sequence can be used to screen a human cDNA library, provided that sufficient cross-species homology exists. For example, the human gene for

phenylalanine 4-hydroxylase (phenylalanine 4-monooxygenase; EC 1.14.16.1), where a deficiency of the enzyme results in phenylketonuria (PKU), was identified using a rat cDNA clone (Woo *et al.*, 1983).

Alternative methods are possible when the protein sequence is unknown but a protein-specific antibody can be raised. The success of this approach depends on expression of the relevant recombinant DNA sequence in bacteria. This technique is now becoming more important because of the availability of monoclonal antibodies (Kemp *et al.*, 1983; Young and Davis, 1983).

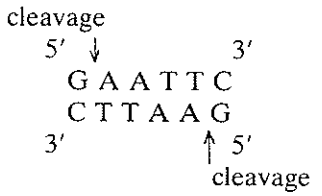
If an appropriate mutant cell line exists, the technique of DNA-mediated gene transfer can be employed to investigate sequences that complement the mutation when integrated into the deficient host genome. This procedure may involve the transfer of human DNA into rodent cells and its subsequent identification using human specific repetitive sequences (Jolly *et al.*, 1982).

Direct analysis of mutations

The reason it is desirable to be able to examine a phenotype at the level of the DNA sequence lies in the mechanism by which proteins are produced. Many eukaryotic genes are split into coding regions (exons) and non-coding regions (introns or intervening sequences). Transcription across these regions generates long precursor mRNA molecules which undergo further processing. This procedure involves capping, the addition of poly(dA) and the removal of intervening sequences by splicing. Mature mRNA molecules are translated into polypeptide chains which may also be modified before a functional protein is produced (for a review, *see* Breathnach and Chambon, 1981). Perturbation of any process in this chain of events may result in the generation of an abnormal protein, or no protein production. Mutations affecting the production of functional protein are not peculiar to the amino-acid-coding sequence of a gene. Base changes, deletions or insertions, in or around the gene, may influence transcription, mRNA processing, or mRNA or protein chain termination. Specific DNA probes can be used to isolate genomic clones that contain flanking control elements in addition to intron and exon sequences, and the molecular environment of the gene can be analysed by DNA sequencing.

SOUTHERN BLOTTING

The technique of Southern blotting is a powerful tool for examining the organization of the human genome (Southern, 1975). Initially, genomic DNA is cut into fragments which are then separated according to size by gel electrophoresis. Cutting the DNA is achieved using restriction endonucleases (now listed as EC 3.1.21.3, 3.1.21.4 and 3.1.21.5), enzymes isolated mainly from bacteria, which recognize specific sequences in double-stranded DNA. This specificity commonly involves a consensus sequence of four to six nucleotides with a two-fold axis of symmetry. For example, the enzyme *Eco*RI recognizes the hexanucleotide sequence:



*Eco*RI cuts double-stranded DNA a few nucleotides apart on either strand, producing fragments with 5′ overhangs (cohesive ends) as illustrated above. Some enzymes (e.g. *Pst* I) generate 3′ cohesive ends whereas others (e.g. *Bal* I) cleave DNA at the axis of symmetry. Following separation of the DNA fragments by agarose or polyacrylamide gel electrophoresis, the fragments are denatured *in situ* by alkali, 'blotted' on to a nitrocellulose paper filter and then fixed by baking at 80°C. This filter replica of the original gel electrophoretic separation is analysed by hybridization of the fixed DNA to radioactive gene-specific probes, and subsequent autoradiography.

The appearance or abolition of a restriction endonuclease site by single base changes, or the insertion or deletion of DNA sequences, will result in a change in fragment size that may be detected by Southern blotting (for review *see* Little, 1981). However, the resolution of this technique is not sufficient to detect size changes less than 50 base pairs (bp), and both resolution of bands and binding to the nitrocellulose is reduced for fragments of very small molecular weight. A modification of this procedure is currently being developed which will allow the detection of small changes in molecular weight. This technique exploits the change in mobility observed for DNA fragments that are partially denatured during electrophoresis in polyacrylamide gels (Fischer and Lerman, 1979).

RESTRICTION ENZYME MAPPING

The relative positions of restriction sites within cloned DNA can be ascertained using restriction enzymes, singly or in combination, followed by accurate sizing of the resultant fragments (for description *see* Maniatis, Fritsch and Sambrook, 1982). These restriction maps can be used to help order genes along a chromosome, to study gene structure, and to provide a basic framework with which to compare abnormal genes or novel fragments observed on Southern blots. The positions of restriction sites for the enzymes *Mst* II and *Eco*RI around the β-globin gene region are illustrated in *Figure 1*.

ANALYSIS OF EXPRESSED GENES

Cloned gene-specific probes can also be used to assay the dysfunction of genes at the RNA level. Transcription levels of relatively abundant mRNAs can be determined and defects in RNA processing may be assessed by a procedure that involves the separation and immobilization of mRNA in a manner similar to Southern blotting of DNA (Alwine, Kemp and Stark, 1977). This technique has become known as Northern blotting. Abnormal translation of

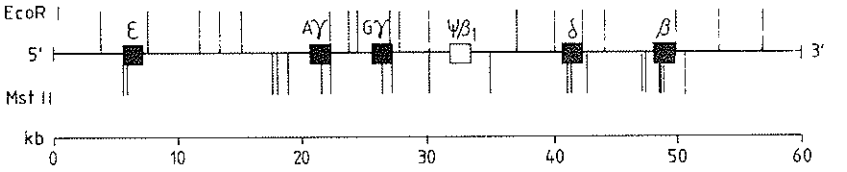


Figure 1. A restriction map of the human β-globin gene region. Restriction sites are shown for the enzymes *Eco*RI and *Mst* II. The black boxes represent functional genes and $\psi\beta_1$ is a pseudogene.

mRNA into protein can be studied using translation assays in cell-free systems following selection of the relevant mRNA with an immobilized gene-specific probe (Pelham and Jackson, 1976). Alternatively, expression systems such as *Xenopus* oocytes can be employed to study the translational efficiency of an mRNA molecule (Lane, Marbaix and Gurdon, 1971). The technology is now becoming available whereby specific mutations can be introduced at specific sites in DNA. This process, termed site-directed mutagenesis, allows the construction of mutations *in vitro* at defined sites in cloned genes. The importance of these sequences in the control of gene expression can then be assessed using an appropriate expression system (for review, *see* Shortle, Di Maio and Nathans, 1981).

MOLECULAR ANALYSIS OF THE GLOBIN GENE CLUSTERS

The globin genes which have been particularly amenable to analysis at the molecular level are situated in two clusters: the α-globin gene complex on chromosome 16 and the β-globin gene complex on chromosome 11. Studies on these genes have provided evidence for the action of many types of mutations, both in structural and in control sequences, that result in disease phenotypes (*see* reviews by Orkin, Antonarakis and Kazazian, 1983; Weatherall, 1984). The most common globin variants are not deleterious and are caused by single nucleotide changes that result in single amino-acid substitutions. Many of these variants have been detected as a result of changes in restriction patterns on Southern blots. Amino-acid substitutions that affect the function of the globin genes have also been identified. For example, a change from glutamic acid to valine is responsible for the abnormal β-globin in patients suffering from sickle cell anaemia and results from a mutation in codon 6 (GAG to GTG) which destroys three restriction enzyme sites in the normal β-globin gene, i.e. those for *Mnl* I, *Dde* I and *Mst* II (Chang and Kan, 1981; Geever *et al.*, 1981; Orkin *et al.*, 1982a; Wilson *et al.*, 1982). This change has formed the basis for direct prenatal diagnosis of sickle cell anaemia using the destruction of the *Mst* II site (Chang and Kan, 1982).

The strategy behind this type of diagnosis is illustrated in *Figure 2*. Digestion with *Mst* II generates 1·1 kb and 1·3 kb fragments for the normal (β^A) and abnormal (β^S) genes respectively. Hybridization of these fragments on a Southern blot with a radioactively labelled probe demonstrates

the presence of one or other fragment in individuals homozygous for the mutant (SS) or normal (AA) genes. Both fragments are detected in DNA from a pregnant heterozygous mother (AS) and in DNA obtained from a biopsy sample of fetal trophoblasts (T). The mutant allele has therefore been inherited by her fetus in this case.

Nucleotide substitutions in flanking control sequences 5′ to the globin gene have been found for some β-thalassaemias (Orkin *et al.*, 1982b; Poncz *et al.*, 1982) and at least one of these has been found to affect mRNA production in a HeLa cell expression system (Treisman, Orkin and Maniatis, 1983). A variety of mutations affecting RNA processing have been identified for the β-thalassaemias, including changes at intervening sequence-coding region junctions, within intervening sequences and even within the coding sequence itself. Mutations affecting globin mRNA function include point mutations that generate stop codons for translation, and frameshift mutations caused by small deletions or insertions that may also lead to premature termination of protein synthesis.

OLIGONUCLEOTIDE DETECTION OF SINGLE BASE CHANGES

The failure of most of these mutations to alter restriction sites in the β-globin gene limits the use of restriction enzyme analysis for detection. A powerful new approach, pioneered by Wallace and his colleagues (Wallace *et al.*, 1981), promises to be of great value in detecting changes in DNA sequences that are not amenable to routine restriction enzyme analysis. This technology relies on the use of short synthetic oligonucleotides to detect single base changes in DNA samples (reviewed by Itakura, Rossi and Wallace, 1984). Under controlled conditions a single nucleotide mismatch between an oligonucleotide and a cloned gene is sufficient to destabilize the DNA hybrid

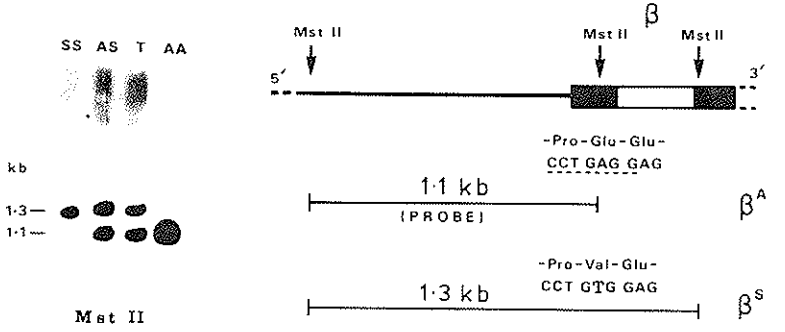


Figure 2. Detection of the abnormal β^S-globin gene. Restriction sites for the enzyme *Mst* II are indicated for the normal β-globin gene, together with DNA fragments generated for normal (β^A) and abnormal (β^S) genes. The autoradiograph (left) demonstrates the presence of the abnormal allele in a pregnant mother and her fetus (lanes AS and T respectively). SS and AA are homozygous controls (reproduced by courtesy of Dr J. Old).

molecule in relation to a perfectly matched one (Wallace *et al.*, 1981; Conner *et al.*, 1983). For example, hybridization of the oligonucleotide with a normal sequence, to a Southern blot of DNA from a normal individual and to one from a fetus at risk, will give a positive signal with the normal DNA and will not give a signal with the fetal DNA if the fetus is affected. Conversely, hybridization of an oligonucleotide corresponding to the mutant sequence, will give a signal with the DNA if the fetus is affected, but will not hybridize to the normal DNA. Heterozygous carriers of a mutation will give a signal with both oligonucleotides.

This technique has recently been employed in the diagnosis of the single base changes that result in a β°-thalassaemia (Piratsu *et al.*, 1983), a β^{+}-thalassaemia (Orkin, Markham and Kazazian, 1983), and α_1-antitrypsin deficiency (Kidd *et al.*, 1983).

Indirect analysis of genetic disorders

The spectrum of mutations which have been demonstrated for the globin diseases may also apply to other genetic disorders. Such a degree of heterogeneity poses problems for the direct identification of mutated sequences giving rise to disease phenotypes. Consequently, indirect approaches have been necessary to tackle these problems, using probes which detect polymorphic DNA sequences adjacent to the mutant gene causing the disease.

RESTRICTION FRAGMENT LENGTH POLYMORPHISMS (RFLPS)

Throughout the genome, normal variations between individuals occur in nucleotide sequences which do not give rise to any change of phenotype. In the case of the β-globin cluster, the frequency of these variations has been estimated to be approximately once in every hundred base pairs (Jeffreys, 1979). If one of these changes causes alterations of a restriction endonuclease recognition site, then digestion of DNA from an individual carrying the altered site would generate larger (loss of a site) or smaller (gain of a site) restriction fragments. These subtle differences can be recognized by a change of fragment mobility on gel electrophoresis. Variants occur at a sufficiently high frequency at many restriction sites for these sites to be described as polymorphic and these restriction fragment length polymorphisms (RFLPs) show simple Mendelian co-dominant inheritance (Wyman and White, 1980).

This phenomenon was first exploited by Kan and Dozy (1978) in analysing the sickle cell anaemia mutation in Americans of West African descent. They isolated a DNA sequence from the 3′ flanking region of the β-globin gene which detected a *Hpa* I polymorphism. They found strong evidence for association of the sickle cell gene with the absence of the *Hpa* I site in this ethnic group. This resulted in the first use of an RFLP in the predictive diagnosis of a genetic disease.

HAPLOTYPE ANALYSIS

Several more RFLPs have been identified in the β-globin gene cluster, of which seven are located in the flanking regions, three are located in the introns of the genes, one has been isolated in the β-pseudogene and one within the β-globin gene coding sequence (Kazazian *et al.*, 1983). All but three of these polymorphisms exist in all racial groups, and may be associated with different mutant β-globin genes which result in various forms of β-thalassaemia. The three less common polymorphisms occur only in West Africans. Population studies of normal individuals and β-thalassaemia patients with respect to these polymorphisms have shown that only a relatively small number of all the possible combinations of restriction site distributions are found in reality (Antonarakis *et al.*, 1982; Orkin *et al.*, 1982b). These particular patterns of restriction sites in a region of DNA have become known as haplotypes.

LINKAGE DISEQUILIBRIUM

It was demonstrated that particular β-thalassaemia mutations are associated in a non-random fashion with particular haplotypes in such a way that predictive diagnosis of affected individuals can be carried out in many populations through the detection of these associated RFLP patterns. Non-random association of this type is classically described as linkage disequilibrium. This type of linkage disequilibrium observed between mutations giving rise to β-globin disorders and certain haplotypes of RFLP patterns has not been described for other genetic diseases. However, even in the absence of linkage disequilibrium and where the exact mutation giving rise to the disease is unknown, DNA probes for the gene in question may be informative for predictive diagnosis by means of family linkage. In cases where a probe exists for the defective gene but where the mutation does not alter the restriction enzyme pattern between normal and mutant chromosomes, the linkage approach may be informative if restriction polymorphisms can be detected in adjacent regions.

HAEMOPHILIA B

Haemophilia B (Christmas disease) is a sex-linked recessive trait caused by clotting factor IX deficiency. Until recently, the only method available for identification of potential carriers was by comparison of factor IX coagulation activity and factor IX antigenicity in potential carriers and normal individuals. Because of the wide variation of these values in the normal population, the results of such tests could have a reliabiity of only about 80%. With the recent application of recombinant DNA and genetic linkage techniques, a new method of accurate carrier diagnosis has become feasible. Molecular cloning of the large human factor IX gene was achieved by using a bovine factor IX cDNA probe to screen a cloned gene library (Choo *et al.*, 1982). Some of these cloned sequences have been used to demonstrate considerable hetero-

geneity at the level of the DNA between patients deficient in factor IX (Gianelli *et al.*, 1983). One of these cloned sequences detects a *Taq* I RFLP in normal individuals with two variable fragments of 1·8 kb and 1·3 kb (Camerino *et al.*, 1984; Gianelli *et al.*, 1984). A large, four-generation family with a severe form of haemophilia B showed complete cosegregation of the haemophilia B mutation with one RFLP allele (1·8 kb) such that it could be used for predictive diagnosis (Grunebaum *et al.*, 1984). However, the frequency of variants of the *Taq I* RFLP is such that only about 40% of haemophilia B families can be used to determine the carrier status of women at risk and provide prenatal diagnosis.

PHENYLKETONURIA

This same principle was developed for the diagnosis of phenylketonuria caused by phenylhydroxylase (PH) deficiency (Woo *et al.*, 1983). A cloned human *PH* probe was found to detect three different RFLPs using the restriction endonucleases *Msp* I, *Sph* I and *Hin*d III. *Figure 3a* illustrates the use of the *Msp* I RFLP to determine the non-carrier status of the second child. In this family, ascertained via the affected child (proband), both parents are heterozygous for the *Msp* I RFLP (*M23/M19*) and must both be heterozygous carriers of a mutated *PH* gene. The affected child is homozygous for the *M19* RFLP allele, and must have inherited this allele with the *PH* mutation from both parents. Therefore in both parents the mutated *PH* gene (*) must be carried on the same chromosome as the 19 kb fragment, and the normal *PH* gene must be associated with the 23 kb fragment.

Similarly, these authors have shown coinheritance of the *PKU* gene with alleles of the *Sph* I and *Hin*d III RFLPs in informative families. In the family illustrated in *Figure 3b* the *PKU* gene is cosegregating with the *Sph* I, *S9·7* allele from both parents and in *Figure 3c* it is inherited with the *Hin*d III *H4·2* allele.

Haplotype analysis for PKU

However, in many families it is not possible to identify the chromosomes from both parents carrying the *PKU* gene using a single RFLP. For instance, if both parents were heterozygous for an RFLP but carried the *PKU* gene with different alleles, then the proband would be heterozygous. Or, alternatively, if only one parent was heterozygous for that RFLP, it would not be possible to identify the disease gene in the offspring. *Figure 4* illustrates a pedigree of this type. Both parents are heterozygous for the *Hin*d III RFLP (*H4·2/H4·0*), and the proband is also heterozygous. Therefore in one parent the *PKU* mutation must be associated with the *H4·2* allele and in the other parent with the *H4·0* allele. Consequently, this RFLP alone could not be used to identify the disease gene. For the *Sph* I RFLP the father is heterozygous (*S9·7/S7·0*) whereas the mother is homozygous (*S7·0/S7·0*) and so this RFLP also could not be used alone for detecting affected offspring. In cases of this nature the

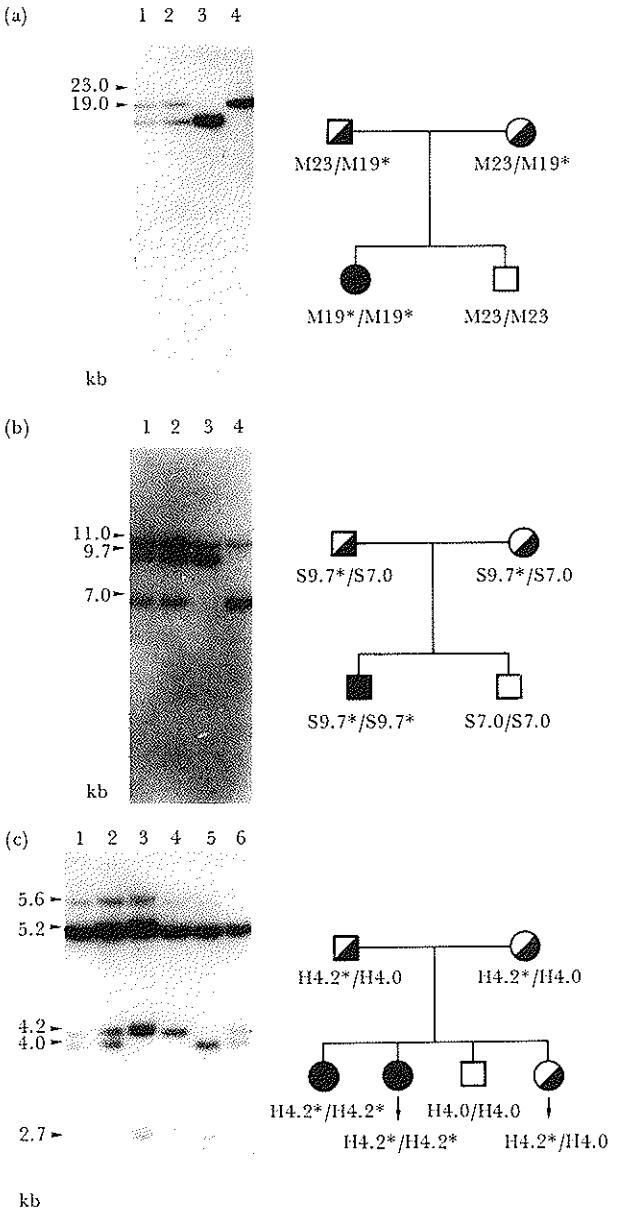


Figure 3. Analyses of three families segregating for PKU and one of the three RFLPs using the enzymes *Msp* I (a), *Sph* I (b), and *Hind* III (c). In Figures 3a and 3b: lane 1, father; lane 2, mother; lane 3, proband and lane 4, unaffected sibling. Figure 3c: lane 1, father; lane 2, mother; lane 3, proband; lane 4, affected sibling and lanes 5 and 6, unaffected siblings. The asterisks identify the RFLP alleles associated with the *PH* mutation in each family (after Woo *et al.* (1983) with kind permission).

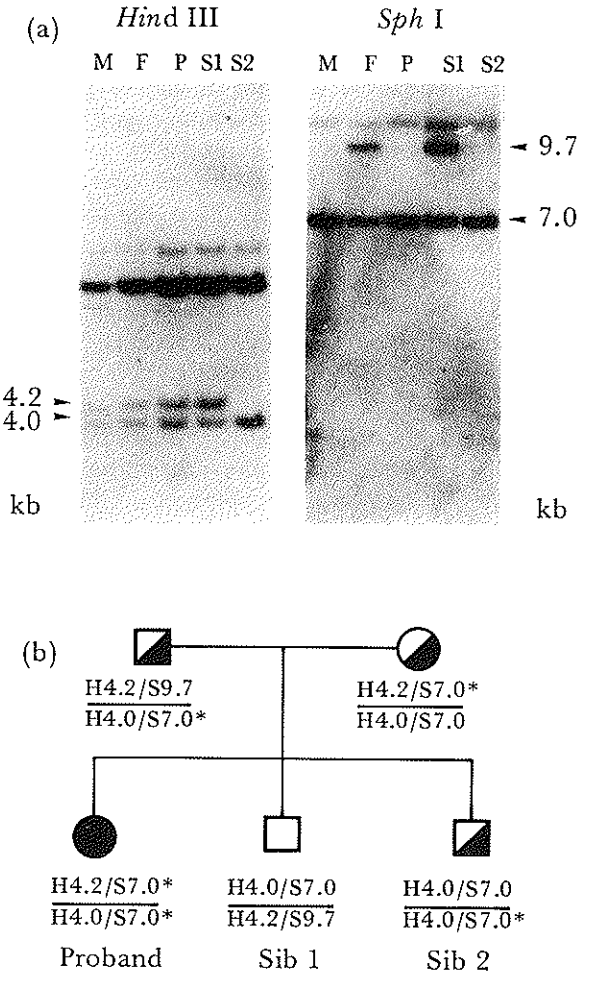


Figure 4. Analysis of a family segregating for PKU using the *Hind* III and *Sph* I RFLPs to determine haplotypes for the *PH* locus. Figure 4a shows Southern blots following digestion of genomic DNA with the enzymes *Hind* III and *Sph* I. The lanes from left to right are: mother, father, proband, sibling 1 and sibling 2. In Figure 4b the asterisk identifies the haplotypes associated with the *PH* mutation (after Woo *et al.* (1983) with kind permission).

extremely powerful method of haplotype analysis can be used. This exploits the combined information from the segregation of more than one RFLP, located within the same gene or its flanking sequences, sufficiently close together that recombination is unlikely to occur between them. This allows the identification of four different parental RFLP patterns or haplotypes. For example in the pedigree shown in *Figure 4* both the mother and proband are homozygous for the *Sph* I 7·0 kb pattern, while the father is heterozygous *S9·7/S7·0* (lanes 1–3). Therefore the mutant *PH* gene must be associated with the *S7·0* allele. The third child (lane 5) is homozygous for the *Sph* I 7·0 kb

band and the *Hin*d III 4·0 kb band and so both parents must have passed on the *PH* haplotype *H4·0:S7·0*. The first child (lane 3) established that the father's defective *PH* gene is carried with the haplotype *H4·0:S7·0* and so by inference the mother must have the mutant gene associated with the *H4·2:S7·0* haplotype. Therefore the normal *PH* genes in the father and mother must be associated with the haplotypes *H4·2:S9·7* and *H4·0:S7·0*, respectively as illustrated on the pedigree. Once the parental haplotypes have been deduced then it is possible to determine the carrier status of the unaffected offspring. The second child (lane 4) has inherited *H4·0:S7·0* from his mother which is associated with the normal *PH* gene, and from his father *H4·2:S9·7* also associated with the normal *PH* gene. Therefore this child is not a carrier. The third child (lane 5) must be a *PKU* carrier, having inherited the *H4·0:S7·0* haplotype associated with the defective gene from his father.

Thus, families that are not informative for predictive diagnosis using a single restriction enzyme may well become informative by using combinations of RFLPs to determine the parental haplotypes. Unlike the situation with β-thalassaemias, where particular mutations show associations at the population level with certain haplotypes of RFLPs in the β-globin gene cluster, the mutant *PH* gene which gives rise to PKU does not show any strong suggestion of disequilibrium with any particular haplotype pattern.

General mapping strategies

The basic defects responsible for many monogenic diseases such as Duchenne muscular dystrophy (DMD) and cystic fibrosis (CF) are not yet understood and consequently the approach of cloning the gene responsible for the disease cannot be used. Indirect methods of investigating such diseases have been employed and the technique of classical family genetic linkage analysis has been of particular value. Genetic linkage depends upon the close physical proximity between two genes such that during the assortative divisions of gametogenesis the randomly distributed recombination events will rarely occur between the two sites in question. Linkage is observed in family studies when two markers appear to segregate together from parents to offspring. Linkage can be detected between two markers, only when they are polymorphic and show variation within individuals.

The first example of human genetic linkage was between two X-borne markers, the gene for haemophilia A and that for colour-blindness (Bell and Haldane, 1937). However, for use as a predictive test, these two genes show variation too rarely in families. The second linkage group discovered in man involves loci that are not sex-linked: they are the loci controlling the secretion of ABH blood group substances in secretory tissues (*Se*) and the Lutheran blood group (*Lu*) (Mohr, 1951). Linkage between the locus causing myotonic dystrophy (*DM*) and these loci was demonstrated soon afterwards (Mohr, 1954; Renwick *et al.*, 1971). The frequency of the polymorphism at the secretor locus is sufficiently high (f(*Se*) = 0·52 and f(*se*) = 0·48) (Race and Sanger, 1975) to suggest that it might provide a useful predictive test. However, estimates of the genetic distance between these two loci indicate a

recombination fraction of about 0·10. This means that, of 100 offspring of parents informative for *DM* and *Se*, about 10 will be recombinants, i.e. they will have a non-parental combination of alleles. Clearly, this is too high an error rate for such a marker to be acceptable for predictive diagnosis. The next line of approach would be to identify common polymorphisms which lie much closer to the disease gene.

MAPPING THE HUMAN GENOME WITH RFLPS

The availability of cloned DNA sequences which detect common RFLPs has significantly increased the value of this strategy as a means of searching for disease mutations where the biochemical nature of the disease is unknown. Botstein *et al.* (1980) further explored this basic principle towards the construction of a genetic map of the human genome. They discuss the value of isolating sufficient numbers of highly polymorphic markers spaced at approximately 20 cM intervals distributed throughout the genome. Then any newly found polymorphism would be within measurable distance of at least one of these markers. The availability of polymorphisms in DNA sequences (RFLPs and variations in numbers of tandemly repeated sequences) ideally suits this strategy because large numbers of such markers could theoretically be isolated at random over the whole genome and, in addition can be physically localized on the chromosomes without prior knowledge of their linkage relationships (*see later*).

Solomon and Bodmer (1979) considered the number of such markers that would be required to cover systematically the entire human genome which has been estimated to be 30 Morgans (3000 cM) in length (Renwick, 1969). They estimated that between 200 and 300 suitably selected probes would provide genetic markers for every 10 cM. Botstein *et al.* (1980) suggested that only about 150 markers would be required if they were more widely spaced (at about 20 cM intervals) as this would mean that any new RFLP or disease gene to be mapped must be less than 10 cM from a known and localized marker. This value of 10 cM represents the range over which linkage can be detected relatively easily.

In reality, the number of random polymorphic markers required is probably larger, in the region of 500–1000. This is because their distribution would not be completely random and the degree of polymorphism of RFLP markers is seldom so high that every individual yields information on its segregation. The term polymorphism information content (PIC) has been introduced to evaluate the informativeness of markers. For instance, a highly polymorphic marker (PIC $> 0{\cdot}5$) will segregate informatively in more than 50% of families and so it will be relatively easy to demonstrate linkage (or lack of it) with new markers. On the other hand, a slightly informative marker (PIC $< 0{\cdot}25$) will segregate only rarely in families of interest and consequently will generate less genetic linkage information.

In addition to these factors, account must be taken of the non-random distribution of crossing-over events. When physical maps of chromosomes are compared against genetical maps, there appears to be considerably more

recombination at the distal ends of the chromosomes than near their centromeres (Drayna *et al.*, 1984; Hartley *et al.*, 1984). These observations confirm the findings of Hulten (1974) and Laurie, Hulten and Jones (1981) that the distribution of chiasmata during spermatogenesis (thought to be cytologically visible sites of cross-over events) is not random, and indeed are seen most frequently at the telomeres (distal regions) of the chromosomes. For this reason, a denser distribution of probes is required around the terminal regions of the chromosomes to provide RFLP markers at 20 cM intervals.

The situation of having a set of polymorphic markers spanning the length of a chromosome is approached with the linkage data on one blood group and eight RFLP loci mapped to the human X-chromosome (Drayna *et al.*, 1984). A few of the intervals between these nine markers are estimated to be greater than 20 cM and so intermediate markers will be required to construct a total X-chromosome genetic map, although the existing markers should allow access by linkage analysis to most X-linked disease loci.

Isolation and localization of chromosome-specific sequences

For indirect analysis by linkage studies, appropriate DNA probes are required, localized along a chromosome. To obtain these, cloned human sequences may be isolated from libraries of DNA sequences developed from total genomic DNA (Maniatis *et al.*, 1978), or from libraries of cDNA sequences. However, mapping of the human genome in general would also be made easier if approached one chromosome at a time, using sequences originating from that chromosome. Moderate enrichment of chromosome-specific sequences has been achieved using cell lines containing multiple copies of a chromosome of interest. X-chromosome sequences were isolated from a 4X cell line (Wolf, Mareni and Migeon, 1980) and more recently this method has been used for the isolation of clones containing sequences of the X-linked factor VIII gene (Wood *et al.*, 1984).

The construction of libraries in which the majority of clones originate from a single human chromosome has been achieved using several strategies. One of these relies on the availability of two cell lines that are identical in every respect except that one contains the sequences of interest and the other does not. DNA from both sources is cut into fragments of suitable size. Selection then involves exhaustive hybridization of a trace amount of DNA from the cells possessing the relevant sequences with an excess of DNA from the deficient line. Unhybridized DNA containing desirable sequences can be isolated by virtue of its differential retention on a column of hydroxylapatite. Suitable pairs of cell lines for this purpose may be male and female cells that differ with respect to the presence of the Y chromosome, or rodent–human hybrid cell lines containing only one human chromosome in conjunction with the appropriate rodent cells. Schmeckpeper *et al.* (1979) have used this approach to isolate DNA probes from the human X chromosome. Kunkel, Smith and Boyer (1976) first used this technique to select Y-chromosome-specific sequences by the hybridization of excess female DNA to male DNA.

An interesting variation on this technique was devised by Lamar and Palmer (1984) for the isolation of Y sequences. This procedure, termed 'deletion enrichment', involves hybridization of an excess of sonicated female DNA to male DNA that has been cut with the restriction endonuclease *Mbo* I. Y-specific sequences not possessing homology to autosomal or X-chromosome DNA can only self-hybridize under the conditions used. These hybrid molecules will possess GATC cohesive ends determined by the sequence specificity of the enzyme *Mbo* I and can easily be cloned into a plasmid that has been cut with *Bam*H I, an enzyme that also generates GATC ends. This method could, in principle, be extended for use with other chromosomes, provided that an appropriate rodent–human hybrid was available.

Another example of the use of somatic-cell hybrids for the isolation of the X-chromosome-specific sequences was provided by Olsen, McBride and Otey (1980). In this case, human DNA was initially self-hybridized under conditions that would allow only repetitive sequences to re-anneal. Homopolymer tails of poly (dA) were added to the remaining single-stranded radioactively labelled unique sequence DNA, which was hybridized with excess labelled mouse DNA to remove any cross-species homology. The labelled DNA without homology to mouse was then hybridized with an excess of DNA from a mouse–human hybrid containing only the X chromosome. The only successful hybridizations in this mixture would be those between X-chromosome sequences. Removal of human labelled sequences from the excess of rodent–human hybrid was conducted by passing the mixture through a column of oligo (dT) which removes the (dA) tailed molecules.

Despite the novel variations that have been introduced into the scheme of isolating unhybridized chromosome-specific sequences, a number of limitations apply. Suitable cells must be available that differ by only one chromosome. Stable somatic-cell hybrids containing only one human chromosome are difficult to obtain, especially if no selection exists for the chromosome of interest. Complications arise because of sequence homologies between rodent and human cells or between certain human chromosomes and it is difficult to ascertain if rearrangements have occurred that are undetectable by conventional methods. DNA libraries cloned directly from hybrid lines have been screened with total human DNA to detect clones containing human inserts (Gusella *et al.*, 1980; Cavanee *et al.*, 1984), although this approach will isolate selectively clones containing repetitive sequences which therefore need to be further subcloned to identify single-copy sequences suitable for use as genetic linkage markers.

FLOW CYTOMETRY

Recently, the technology has become available to isolate large numbers of specific human chromosomes. High-resolution flow cytometry separates metaphase chromosomes on the basis of their reaction with a suitable fluorochrome (for review *see* Young and Davies, 1983). Chromosomes are stained in suspension and then passed at high speed in single file through the

path of a laser beam. The fluorescence emitted for each chromosome upon excitation of the stain is measured photometrically and can be used to sort the chromosomes and to record their numbers in the form of a histogram. This technique was first applied for the analysis of Chinese hamster chromosomes (Gray *et al.*, 1975) and has since been extended for use with human cells (Carrano *et al.*, 1979). Human chromosomes can be sorted by flow cytometry into 12 main human groups. An example of such a flow karyotype is shown in *Figure 5*. The assignment of each peak to a particular chromosome is based on published DNA contents and a computer analysis of the area under each peak. It has become clear that certain chromosomes can be sorted with a fairly high level of purity. This technique was used by Davies *et al.* (1981) to provide a population highly enriched for X chromosomes. The assignment of the X chromosome to the peak illustrated in *Figure 5* was confirmed by comparing the flow karyotype obtained from a cell line containing four X chromosomes (*Figure 5b*) with that containing only one X chromosome (*Figure 5a*): in the former case the area of the peak was four times

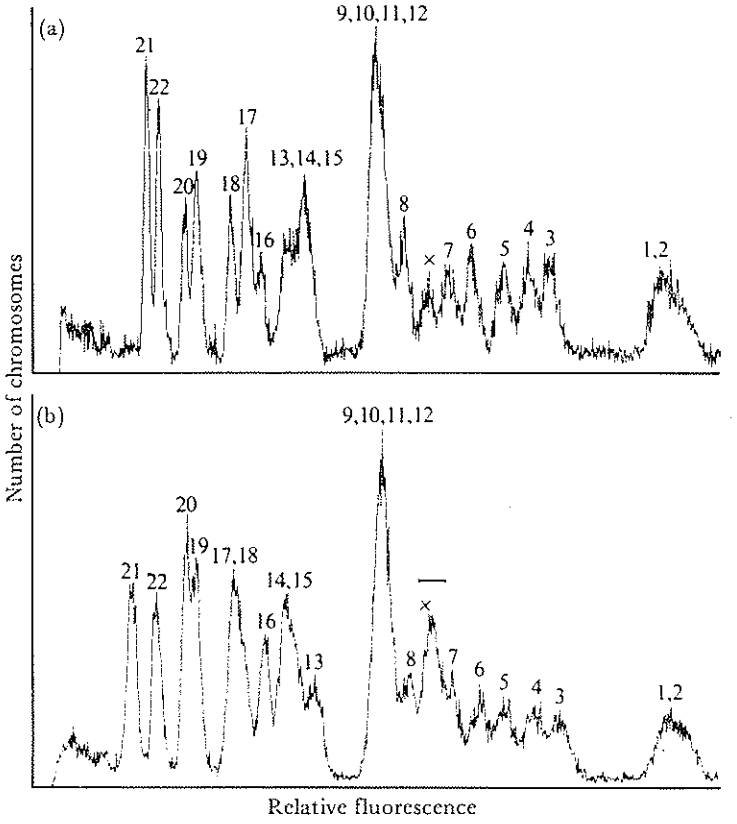


Figure 5. Flow karyotypes obtained from cells containing 46(XY) (a) and 48(XXXX) (b) chromosomes. The chromosomes assigned to each peak are indicated (Davies *et al.* 1981; reproduced by courtesy of *Nature*).

greater. The horizontal bar defines the sorting window used for purification of the X chromosome (Davies *et al.*, 1981). About 2×10^6 copies of the X chromosome were sorted from the cell line GM1416 (48,XXXX), the DNA was digested with the restriction enzyme *Eco*RI and cloned into a phage vector: a total of 50 000 recombinants were obtained and characterized. Similar procedures have been used by Krumlauf, Jeanpierre and Young (1982) to clone libraries of chromosomes 22 and 21, and by Kunkel *et al.* (1982) to prepare a library of the X chromosome.

A major advantage of creating chromosome-enriched libraries after flow cytometry is that a large proportion of the chromosome is represented. Both single-copy and repetitive sequences are isolated, as are sequences that possess some degree of homology with other chromosomes or other species. The specificity and nature of the sequences in these libraries can be investigated using rodent–human hybrids, *in situ* hybridization, and dosage analysis (*see* pages 293–298), and single-copy sequences can be subcloned from clones containing repetitive sequences.

PHYSICAL LOCALIZATION OF DNA PROBES

Physical localization of single-copy DNA probes in the human genome has been accomplished using four main strategies.

Dosage hybridization

The first of these strategies relies on a quantitative difference in the hybridization of a probe to restriction fragments of DNA isolated from cells containing different numbers or sections of a particular chromosome. Cells possessing a deletion on one homologue of a chromosome pair can be used in this way for subchromosomal localization (Kunkel *et al.*, 1982). Assignments of this nature are not unequivocal, as the intensity of a band on an autoradiograph of a Southern blot may vary for a number of reasons other than gene dosage. This technique is often used to establish the origin of clones in X-chromosome-specific libraries, such as those described earlier (Kunkel *et al.*, 1982; Davies *et al.*, 1981, 1983a) as cell lines with multiple copies of this chromosome are readily available.

Somatic-cell genetics

A more widely used method of assignment utilizes somatic-cell hybrids. Fusion between somatic cells (not of germ-line origin) is achieved *in vitro* using either Sendai virus or polyethylene glycol as a fusing agent. Hybrids formed between genetically distinct cells will lose chromosomes predominantly from one or other of the parental cells upon growth in culture. Fusion between rodent and human cells usually results in the loss of human chromosomes, generating a population of hybrids with different human chromosome complements (Weiss and Green, 1967). For efficient isolation of fusion products, a selection pressure is exerted that discriminates against the

growth of parental cells, in favour of hybrid-cell survival. Often this procedure will involve selection for the presence of a particular human gene product that will complement a deficiency in the rodent parent and vice versa. The commonly used HAT (hypoxanthine, aminopterin, thymidine) system (Littlefield, 1964), is used to select for hybrid cells containing the enzymes HPRT (hypoxanthine phosphoribosyltransferase) and thymidine kinase (TK) and against cells deficient in either enzyme. The human genes for HPRT and TK have been mapped to chromosomes X and 17 respectively and, therefore, using an appropriate enzyme-deficient rodent parent, positive selection can be applied for the retention of one or other of these human chromosomes.

Once a hybrid-cell line or panel of hybrid cells containing different human chromosomes is available, assignments can be made by correlating the presence of a particular chromosome with that of a fragment on a Southern blot. The advantage of using restriction fragments for this analysis is that even if some interspecies homology exists for the DNA probe being used, restriction-fragment site differences surrounding a gene occur frequently between species and therefore a distinction is possible. This sort of analysis is obviously facilitated when hybrids containing only one human chromosome are available.

Sub-chromosomal localization of probes has also been achieved using rodent–human hybrid cells. This has been made possible by the high frequency of reciprocal translocation that has been detected in human families. Panels of hybrid cells containing a series of translocations involving different sites along a chromosome can be used to define the location of the probe of interest within certain limits. The term 'smallest region of overlap' has been coined to describe this position.

Figure 6 illustrates the localization of a probe isolated from a flow-sorted X-specific library to the region Xp21 to Xp22·3 on the X chromosome (p and q refer to the short and long arms of chromosomes respectively; the numbers refer to subdivisions of p and q). This clone, RC8, detects a 6·1 kb (*Eco*RI) restriction fragment only from hybrids containing this region and this is independent of the presence of the portion of the X chromosome distal to Xp22·3. Similar analyses have been used to localize DNA sequences belonging to genes with characterized products including, for example, genes encoding human β-globin to chromosome 11 (Gusella *et al.*, 1979) and the human factor IX blood-clotting protein (Chance *et al.*, 1983; Boyd *et al.*, 1984) to the X chromosome. Several DNA sequences of unknown function have been located to the X chromosome in this manner (Wieacker *et al.*, 1984).

Somatic-cell hybrids containing a single human chromosome with an internal deletion of a region of interest are extremely useful for mapping purposes. A human–rodent hybrid containing an X chromosome with a deletion of the Xp21 band has been used to assign several DNA sequences to the missing region by comparison with cells containing an undeleted X chromosome (Ingle *et al.*, 1985). Hybrid lines such as these are also appropriate starting materials for the development of libraries of DNA

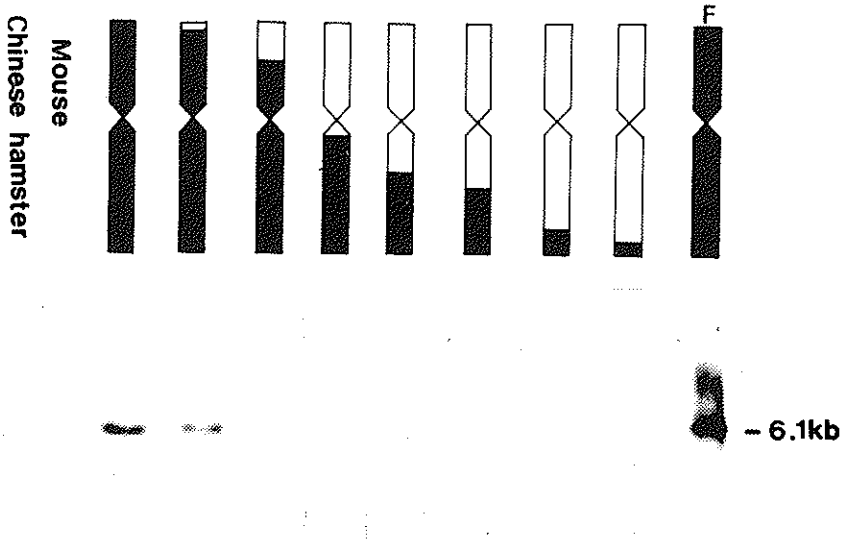


Figure 6. Mapping of cloned sequence RC8 to region Xp21 to Xp22.3 on the human X chromosome using somatic-cell hybrids. The presence or absence of a 6·1 kb fragment detected by RC8 in rodent parent, hybrid or human fibroblast (F) cell lines, is indicated. The portions of the human X chromosome contained in each line are shown above each lane (shaded areas) (Murray *et al.*, 1982; reproduced by courtesy of *Nature*).

probes from specific chromosomal regions, using the differential screening technique described earlier.

Spontaneous rearrangement occurring in established hybrid lines can be exploited for mapping purposes. Where the appropriate translocations or deletions have not been found, random chromosome breaks can be induced in the human parent cell by γ-irradiation before fusion (Goss and Harris, 1977). A limitation of this procedure is that a selection system is desirable for the region of interest in order to maintain it in the hybrid cell.

Chromosome sorting

The technique of chromosome sorting can also be used to define the subchromosomal site of a DNA sequence, provided that the position of an abnormal chromosome in a flow karyotype is sufficiently different from that of the normal homologue (Lebo *et al.*, 1979).

In situ *hybridization*

Although somatic-cell hybridization has proved to be a powerful tool for localization of markers within the genome, the resolution of this technique relies heavily on the availability of suitable chromosome abnormalities, and no account can be taken of any undetectable rearrangements that may have

occurred. More precise localization of DNA sequences is now possible using the technique of *in situ* hybridization. In this procedure the hybridization of radioactively labelled DNA to a fixed preparation of metaphase chromosomes is visualized after exposure to autoradiographic emulsion. Dark spots or grains are produced on the film indicating the position of DNA hybridization (approximately 0·12 grains per beta disintegration for a tritiated DNA probe (Hartley, 1984). The grain distribution over the chromosomes is determined and interpreted with respect to the banding patterns of the same preparations. Pioneered by Gall and Pardue (1969) and John, Birnstiel and Jones (1969), for the localization of *Xenopus* ribosomal DNA, and later used by Jones (1970) and Pardue and Gall (1970) for the localization of mouse repetitive sequences, *in situ* hybridization has been extended for use with repetitive and single-copy human sequences (Harper, Ullrich and Saunders, 1981; Malcolm *et al.*, 1981).

An example of mapping using this technique is shown in *Figure 7*. The chromosomes are schematically represented (an ideogram) with the grain (•) distribution obtained after hybridization with an X-specific cloned DNA sequence DX13 (Hartley, 1984; Hartley *et al.*, 1984). It can be seen that the majority of grains lie in the region Xq28.

Markers of known and unknown function have been mapped using this technique. Compared with mapping using a pre-existing panel of somatic cell hybrids, *in situ* hybridization is more time-consuming and unsuitable for screening a large number of probes. As yet, this technique does not allow assignment of sequences to a single chromosome band, because of the scatter of radioactive disintegrations in three dimensions. Lindgren *et al.* (1984) have overcome this problem to a certain extent by exploiting the availability of cell lines containing X translocations or X-chromosome deletions.

Improvements in chromosome-banding techniques are increasing the efficiency of localization by *in situ* hybridization as well as the karyotypic analysis of the chromosomal abnormalities utilized for mapping using somatic-cell hybrids. Recent developments in this area have permitted the resolution of 2000 bands per haploid genome (Yunis, 1981). These modifications include the synchronization of cells within the cell cycle in order to gain a high yield of mitotic cells, inhibition of chromosome condensation and improved spreading. If the total length of the human genome is assumed to be about 3000 centimorgans (units of recombination) (Renwick, 1969), then the number of bands is less than one half of the number of centimorgans. As there are about 3×10^9 nucleotide base pairs in the human haploid genome, one centimorgan approximates to 10^6 base pairs (1Mbp) on average. Utilizing the new banding procedures it may be possible, therefore, to localize genes to within 2–5 Mbp.

Recombinant DNA techniques have been used to provide molecular maps at the level of single nucleotide base pairs over distances of 10^5 base pairs. The technique of chromosome 'walking' has enabled these distances to be covered. In this method, a genomic clone is used to identify other recombinant molecules from a genomic library that contain a region of overlap with respect to DNA sequence (*see* Cockburn, Howells and Whitten, 1984). The

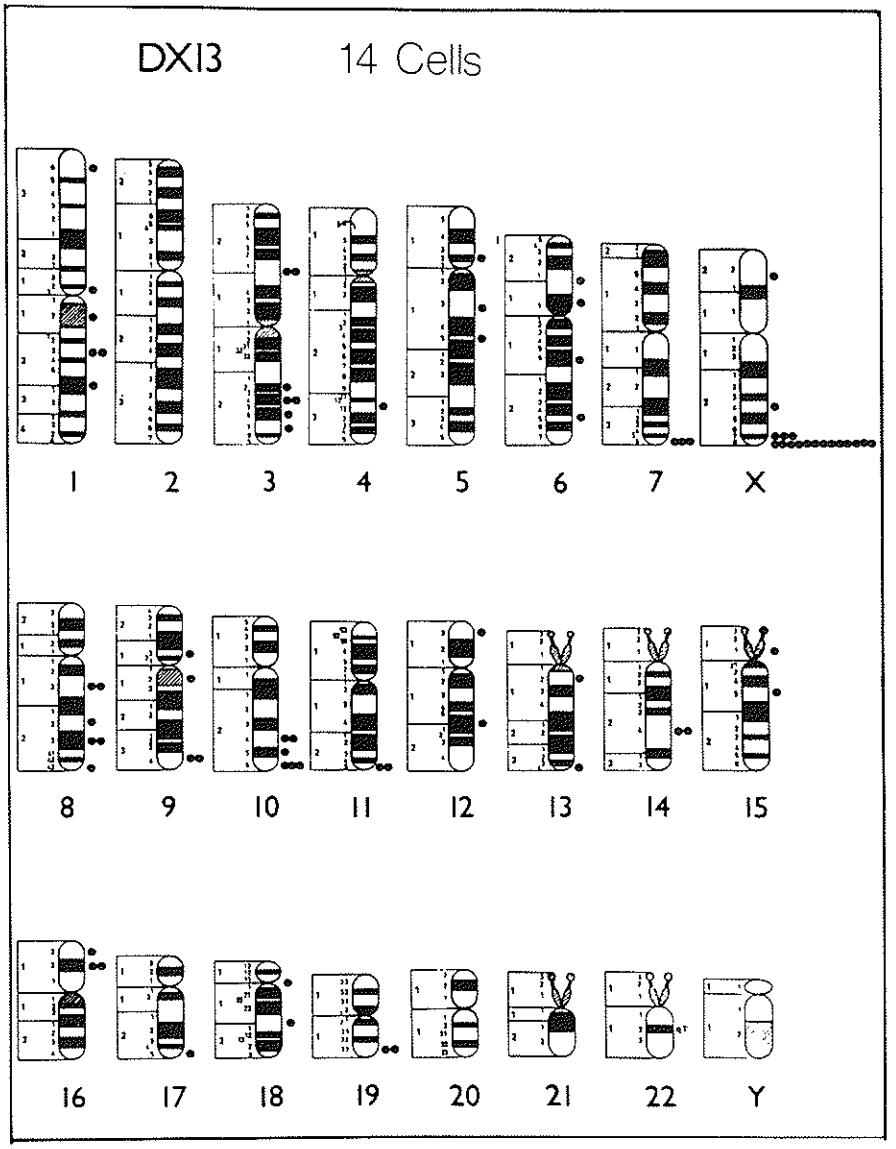


Figure 7. *In situ* hybridization of the cloned DNA sequence DX13 to human chromosomes. A composite grain distribution over 14 cells is illustrated where each dot represents the chromosomal position of a single grain. (Hartley, 1984; reproduced by courtesy of D. Hartley).

map is then extended using these clones to probe the library. Thus, a gap of about 0·1–0·5 Mbp in mapping resolution exists between that obtainable using conventional somatic-cell and cytogenetic techniques and the resolution afforded by recombinant DNA techniques. Two strategies that are being developed to bridge this gap involve directional cloning in 0·25 Mbp jumps (chromosome hopping; Collins and Weissman, 1984 and *see also* Chapter 10

of this volume) and chromosome-mediated gene transfer. In the latter method, small regions of a human chromosome are integrated into a rodent-host-cell genome (Miller and Ruddle, 1978).

Mapping monogenic diseases

Most monogenic diseases are rare, apart from a few notable exceptions, such as cystic fibrosis which affects about 1 in 2000 total live births in the UK, and Duchenne muscular dystrophy which affects about 1 in 3000 male live births. In only a few of these diseases is the basic biochemical defect known. However, advances in genetics and recombinant DNA technology make it possible to map and isolate the gene responsible without the need to identify the primary gene product. The phrase 'reverse genetics' has been aptly used to describe this system of mapping and ultimately isolating genes responsible for disease phenotypes without prior knowledge of the mechanisms by which they act.

Reverse genetics: Duchenne muscular dystrophy (DMD)

The strategy of reverse genetics is currently being used to identify the molecular basis for the lethal sex-linked recessive disorder Duchenne muscular dystrophy (DMD), the biochemical defect of which is unknown. Clues to the location of the *DMD* gene on the X-chromosome involve rare cases of females who apparently have Duchenne muscular dystrophy (for review, *see* Moser, 1984). Cytogenetic investigation revealed that all of these females had balanced translocations, involving breakage of the X chromosome at band Xp21 and the reciprocal exchange of the chromosomal material distal to this breakpoint with material from another autosome. The autosome breakpoints were all different. The deduction from these observations is that the Duchenne muscular dystrophy locus lies within this region (Canki, Dutrillaux and Tivadar, 1979; Lindenbaum *et al.*, 1979; Jacobs *et al.*, 1981; Zatz *et al.*, 1981; Verellen-Dumoulin *et al.*, 1984).

In view of this proposal, the field of search for the *DMD* gene has been narrowed down to the relatively small region around Xp21 in the middle of the short arm of the X chromosome. The most significant advances towards identifying the *DMD* gene by linkage analysis have come from the use of cloned DNA probes which detect unique sequences mapped by somatic-cell hybrid analysis or *in situ* hybridization methods to the short arm of the human X chromosome and which, in addition, reveal RFLPs. By studying the inheritance of alleles of these RFLPs through families with Duchenne muscular dystrophy, estimates can be made of the genetic distance between the RFLP locus and the *DMD* gene by assessing the frequency of recombination in the intervening region.

Linkage between the probe RC8 and *DMD* was demonstrated by Murray *et al.* (1982). Subsequently, Davies *et al.* (1983b) confirmed this linkage and that of another probe, L1.28, with *DMD*. These markers were found to lie on opposite sides of the *DMD* locus: *RC8* was localized to the region Xp22·3–

Xp21, and *L1.28* mapped to Xp11·3–Xp11·0 (Wieacker *et al.*, 1984). The estimated recombination fractions between each probe and *DMD* were 0·15 in both cases and these results are consistent with the assignment of *DMD* to band Xp21. Using these two bridging markers, *RC8* and *L1.28*, Kingston *et al.*, (1983, 1984) were able to demonstrate that the locus for a second, milder X-linked muscular dystrophy, Becker muscular dystrophy (*BMD*), lies between them in the same region of the X chromosome as the *DMD* locus and thus may be allelic.

Currently, considerable effort is being invested in isolating more RFLPs on the short arm of the X chromosome (Davies *et al.*, 1983b, 1985; Aldridge *et al.*, 1984; Drayna *et al.*, 1984; Hartley *et al.*, 1984). The bridging markers *RC8* and *L1.28* are extremely useful in making genetic assignments of these new probes, particularly with locating markers closer to the *DMD* locus. With the availability of RFLPs lying very close to the *DMD* locus on both sides, antenatal diagnosis of *DMD* by linkage analysis will become feasible (Pearson and Van Ommen, 1984). The probes isolated so far are already in use for predicting the carrier status of women in families at risk (Harper *et al.*, 1983; Pembrey *et al.*, 1984; Davies *et al.*, 1985).

Thus the *DMD* gene has been localized to within 2000 kb of the short arm of the X chromosome by linkage analysis. With the aid of directional cloning techniques (Collins and Weissman, 1984), library–library screening (Davies, Taylor and Muller, 1983) and cloning of the translocation breakpoints (Worton *et al.*, 1984), the mutant sequences responsible should be identified within the next two years.

Once probes have been localized by both physical and genetical methods along the X chromosome, they will provide a substantial map within which the position of disease loci can be defined. The X-chromosome map drawn up by Drayna *et al.* (1984) approaches this situation although filling of some of the intervals will be necessary to complete the map. From these data, the genetic length of the X chromosome from Xp22·3 to Xqter can be estimated to be of the order of 215 cM. These studies also demonstrate the increase in frequency of recombination towards the chromosome ends.

HUNTINGTON'S CHOREA

The reverse genetics approach using polymorphisms in DNA sequences has been very successfully applied to Huntington's chorea (Huntington's disease). This is a particularly distressing progressive neurodegenerative disease which is inherited in an autosomal dominant fashion. It shows complete penetrance, meaning that everybody who carries the gene will eventually develop the symptoms, although the age of onset may be quite late in life, after the reproductive age. It is characterized by deterioration of the nervous system with loss of intellectual ability and psychiatric disturbance. The primary defect remains unknown (Martin, 1982). For many years the Huntington's chorea locus (*HD*) had evaded genetic mapping by linkage studies (Volkers *et al.*, 1980; Keats, 1981). However, a systematic search for linkage of the *HD* gene with a series of random RFLPs yielded strong evidence for close linkage

in two large families with an anonymous segment of unique sequence genomic DNA located by physical methods to the short arm of chromosome 4 (Gusella *et al.*, 1983).

Cystic fibrosis

The achievements using recombinant DNA technology in the genetic mapping of Duchenne and Becker muscular dystrophies, Huntington's chorea and several other inherited diseases should encourage researchers investigating the more elusive class of autosomal recessive genetic disorders. Probably the most important of these in the UK is cystic fibrosis (*CF*) affecting about 1 in 2000 live births. Because of its recessive mode of inheritance, heterozygous carriers of cystic fibrosis cannot be reliably detected by the current methods, which depend upon abnormal levels of proteins probably unrelated to the primary defect (Brock, 1984). Analysis of families for information on the linkage relationships of the cystic fibrosis locus (*CF*) is an arduous task because informative heterozygotes cannot be unequivocally identified until they produce homozygous affected children.

Exclusion mapping

The method of exclusion mapping is currently being employed to tackle the problem of locating the *CF* locus. This involves excluding *CF* from chromosomal regions by demonstrating its lack of linkage with polymorphic markers known to lie within that region. This method was illustrated in detail by Cook *et al.* (1980) using the example of mapping the *MNSs* blood group loci. Extensive resources of negative information were available on the linkage relationships of this highly polymorphic complex locus with other conventional polymorphic loci and breakpoint sites in chromosome rearrangements. In conjunction with convincing deletion evidence they were able to exclude *MNSs* from almost all of the genome with the exception of bands 4q28 and 4q31. Suggestions from cytogenetic and linkage analyses support the assignment of *MNSs* to this region of chromosome 4.

To date, exclusion mapping of *CF* has merely succeeded in ruling out the possibility of *CF* being located in certain regions of chromosome 4 (Scambler *et al.*, 1985). However, with the recent surge of interest in linkage analysis, particularly with respect to RFLPs, this method could prove successful in mapping *CF* in the foreseeable future.

Another approach to the investigation of the genetic defect in diseases where the biochemical defect is unknown is to exclude 'candidate genes' by use of specific cloned sequences. For example, in the case of cystic fibrosis there were indications that mutations at the locus for the third complement component (*C3*) may have been the cause of this disease. However, using a cloned *C3* gene probe, Davies, Gilliam and Williamson (1983) demonstrated that the segregation of alleles at the *C3* locus was independent of the segregation of the cystic fibrosis gene in two families. If cystic fibrosis were caused by mutation at the *C3* locus, then coinheritance of the alleles of the *C3*

locus would be expected with the cystic fibrosis gene; *C3* therefore can be excluded as a candidate gene.

Multifactorial diseases

Many common diseases which are seen to run in families are not unifactorial and do not show the simple Mendelian inheritance of monogenic disorders. Indeed, they may be described as multifactorial, having a genetic component in their aetiology in addition to susceptibilities to particular environmental factors. The genetic components of such disorders are often complex and involve interactions between several genes. Examples of these polygenic diseases include many neuropsychiatric disorders such as schizophrenia and manic depressive disease, atherosclerosis and coronary artery disease, hypertension, diabetes mellitus and some commoner forms of cancer. The approach to identifying some of the genes responsible for a polygenic condition is much more circuitous. Because of the very nature of polygenic diseases, identification of a single gene defect which gives rise to the phenotype in some cases will not be the same for other individuals.

Investigations into the hereditary components of atherosclerosis illustrate the strategies utilized. Atherosclerosis is a hardening of the arterial walls due to a deposition of lipids; the walls then become fibrotic and eventually calcify. This results in narrowing of the bore of the vessels, causing ischaemia of the tissues they supply, which ultimately leads to coronary artery disease and cerebrovascular disease. Atherosclerosis is a very variable disorder and is usual in the mild stages in later age, whereas some patients develop coronary artery disease during teenage years. Association is seen between premature vascular disease and certain unifactorial disorders, familial hypercholesterolaemia in particular. This is an autosomal dominant disorder with a frequency of about 1 in 500 individuals in the UK population. About 1 in 20 patients with ischaemic heart disease are heterozygous for this gene, which has now been identified to code for a defective low-density-lipoprotein cell-surface receptor which reduces the efficiency of cholesterol uptake from plasma, resulting in elevated plasma cholesterol levels.

Such a simple genetic model cannot yet explain the majority of cases of coronary artery disease, although other genes which may be candidates for predisposing their hosts can be identified. For example, the genetic control of lipoproteins must be relevant because elevated levels of low-density lipoprotein (LDL) and depressed levels of high-density lipoprotein (HDL) are each considered to indicate individuals at high risk of premature atherosclerosis. The major protein constituent of LDL is apolipoprotein E (apo E) which is known to show phenotypic polymorphism (Cumming and Robertson, 1982). Homozygosity for one of the apoliproprotein alleles, *EII*, is recognized as a necessary prerequisite for the development of type III hyperlipidaemia (Utermann *et al.*, 1979). Many other genes can be suggested as possible culprits in the aetiology of vascular disease, such as other apoproteins, fibrinogens and cell-surface receptors, in addition to conventional risk-factors such as obesity and hypertension. The application of recombinant DNA

technology to clone these candidate genes may well soon elucidate the complexity of the genetics of these polygenic conditions.

Conclusion

Recombinant DNA is revolutionizing human genetics. To date, several human chromosomes (X, 4, 19, 21, 1p, 6p, 11p, 13q) are quite extensively genetically mapped using RFLP loci. One important autosomal dominant disorder and several X-linked disorders have been localized. The next few years should reveal the assignments of more monogenic diseases which will eventually lead to the identification of the mutant loci responsible. The molecular geneticist now has the tools to begin to understand the genetics of all monogenic disorders and the diversity of their causation at the DNA level.

Acknowledgements

We would like to thank R. Kitt for patient typing of the manuscript. We are grateful to the Muscular Dystrophy Group of Great Britain, the Muscular Dystrophy Association of America and The Medical Research Council for financial support.

References

Aldridge, J., Kunkel, L., Bruns, G., Lalande, M., Tantravahi, U. and Latt, S. (1984). A strategy for construction of highly polymorphic DNA haplotypes in specific human chromosomal regions: application to linkage analysis of X chromosome specific diseases such as Duchenne muscular dystrophy. *Italian Journal of Neurological Science* **3**, 39–46.

Alwine, J.C., Kemp, D.J. and Stark, G.R. (1977). Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl paper and hybridisation with DNA probes. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 5350–5354.

Antonarakis, S.E., Boehm, C.D., Giardina, P.J.V. and Kazazian, H.H. (1982). Non random association of polymorphic restriction sites in the β-globin gene complex. *Proceedings of the National Academy of Sciences of the United States of America* **79**, 137–141.

Bateson, W., Saunders, E.R. and Punnett, R.C. (1905a). Experimental studies in the physiology of heredity. *Evolutionary Communications of the Royal Society of London* **2**, 1–55.

Bateson, W., Saunders, E.R. and Punnett, R.C. (1905b). Experimental studies in the physiology of heredity (continued). *Evolutionary Communications of the Royal Society of London* **2**, 80–99.

Bell, J. and Haldane, J.B.S. (1937). The linkage between the genes for colourblindness and haemophilia in man. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **123**, 119–150.

Botstein, D., White, R.L., Scolnick, M.H. and Davis, R.W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* **32**, 314–331.

Boyd, Y., Buckle, V.J., Munro, E.A., Choo, K.H., Migeon, B.R. and Craig, I.W. (1984). Assignment of the haemophilia B (factor IX) locus to the q26-qter region of the X chromosome. *Annals of Human Genetics* **48**, 145–152.

BREATHNACH, R. AND CHAMBON, P. (1981). Organisation and expression of eukaryotic split genes coding for proteins. *Annual Review of Biochemistry* **50**, 349–383.

BRENNAND, J., CHINAULT, A.C., KONECKI, D.W., MELTON, D.W. AND CASKEY, C.T. (1982). Cloned cDNA sequences of the hypoxanthine/guanine phosphoribosyltransferase gene from a mouse neuroblastoma cell line found to have amplified genomic sequences. *Proceedings of the National Academy of Sciences of the United States of America* **79**, 1950–1954.

BROCK, D.J.H. (1984). Prenatal diagnosis and heterozygote detection. In *Cystic Fibrosis: Horizons* (D. Lawson, Ed.), pp. 1–12. John Wiley and Sons, London and New York.

CAMERINO, G., GRZESCHIK, K.H., JAYE, M., DE LA SALLE, H., TOLSTOSHEV, P. AND LECOCQ, J.P. (1984). Regional localization on the human X chromosome and polymorphism of the coagulation factor IX gene (hemophilia B locus). *Proceedings of the National Academy of Sciences of the United States of America* **81**, 498–502.

CANKI, N., DUTRILLAUX, B. AND TIVADAR, F. (1979). Dystrophie musculaire de Duchenne chez une petite fille porteuse d'une translocation t(X;3) (p21; q13) *de novo*. *Annales de Genetique* **22**, 35–39.

CARRANO, A.V., GRAY, J.W., LANGLOIS, R.G., BURKHARDT-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.

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* **30**, 10–24.

CHANCE, P.F., DYER, K.A., YOSHITAKE, S., ROPERS, H.-H., WIEACKER, P. AND GARTLER, S.M. (1983). Regional localisation of the human factor IX gene by molecular hybridisation. *Human Genetics* **65**, 207–208.

CHANG, J.C. AND KAN, Y.W. (1981). Antenatal diagnosis of sickle-cell anaemia by direct analysis of the sickle mutation. *Lancet* **ii**, 1127–1129.

CHANG, J.C. AND KAN, Y.W. (1982). A sensitive new prenatal test for sickle-cell anaemia. *New England Journal of Medicine* **307**, 30–32.

CHOO, K.H., GOULD, K.G., REES, D.J.G. AND BROWNLEE, G.G. (1982). Molecular cloning of the gene for human anti-haemophilic factor IX. *Nature* **299**, 178–180.

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.

COLLINS, F.S. AND WEISSMAN, S.M. (1984). Directional cloning of DNA fragments at a large distance from an initial probe: a circularization 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 β^s-globin allele by hybridization with synthetic oligonucleotides. *Proceedings of the National Academy of Sciences of the United States of America* **80**, 278–282.

COOK, P.J.L., NOADES, J.E., LOMAS, C.G., BUCKTON, K.E. AND ROBSON, E.B. (1980). Exclusion mapping illustrated by the MNSs blood group. *Annals of Human Genetics* **44**, 61–73.

CRAMPTON, J.M., DAVIES, K.E. AND KNAPP, T.F. (1981). The occurrence of families of repeated sequences in a library of cloned cDNA from human lymphocytes. *Nucleic Acids Research* **9**, 3821–3824.

CUMMING, A.M. AND ROBERTSON, F.W. (1982). Genetics of the apolipoprotein-E isoprotein system in man. *Journal of Medical Genetics* **19**, 417–423.

DAVIES, K.E., GILLIAM, T.C. AND WILLIAMSON, R. (1983). Cystic fibrosis is not caused by a defect in the gene coding for human complement C3. *Molecular Biology and Medicine* **1**, 185–190.

DAVIES, K.E., TAYLOR, P. AND MULLER, C.R. (1983). Sex chromosome-specific DNA sequences. *Differentiation* **23**, 44–47.

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., HARTLEY, D.A., MURRAY, J.M., HARPER, P.S., HILL, M.E.E., CASEY, G., TAYLOR, P. AND WILIAMSON, R. (1983a). The characterisation of sequences from a human X chromosome library for the study of X-linked diseases. In *Banbury Report: Recombinant DNA Applications to Human Disease,* 14th Edn (C.T. Caskey and R.L. White, Eds), pp. 279–290. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

DAVIES, K.E., PEARSON, P.L., HARPER, P.S., MURRAY, J.M., O'BRIEN, T., SARFARAZI, M. AND WILLIAMSON, R. (1983b). Linkage analysis of the two cloned DNA sequences flanking the Duchenne muscular dystrophy locus on the short arm of the human X chromosome. *Nucleic Acids Research* **11**, 2303–2312.

DAVIES, K.E., BRIAND, P., IONASESCU, V., IONASESCU, G., WILLIAMSON, R., BROWN, C., CAVARD, C. AND CATHELINEAU, L. (1985). Gene for OTC: characterisation and linkage to Duchenne muscular dystrophy. *Nucleic Acids Research*, in press.

DE VRIES, H. (1910). Befruchtung und bastardiening. In *Intracellular Pangenesis* (C.S. Gager, Ed.), pp. 217–263. Chicago Open Court Publishing Co, Chicago.

DRAYNA, D., DAVIES, K.E., HARTLEY, D.A., WILLIAMSON, R. AND WHITE, R. (1984). Genetic mapping of the human X chromosome using restriction fragment length polymorphisms. *Proceedings of the National Academy of Sciences of the United States of America* **81**, 2836–2839.

FISCHER, S.G. AND LERMAN, L.S. (1979). Length-independent separation of DNA restriction fragments in two-dimensional gel electrophoresis. *Cell* **16**, 191–200.

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.

GARROD, E. (1902). The incidence of alcaptonuria: a study in chemical individuality. *Lancet* **ii**, 1616–1620.

GEEVER, R.F., WILSON, L.B., NALLASETH, F.S., MILTHER, P.F., BITTNER, M. AND WILSON, J.T. (1981). Direct identification of sickle cell anemia by blot hybridisation. *Proceedings of the National Academy of Sciences of the United States of America* **78**, 5081–5085.

GIANELLI, F., CHOO, K.H., REES, D.J.G., BOYD, Y., RIZZA, C.R. AND BROWNLEE, G.G. (1983). Gene deletions in patients with haemophilia B and anti-factor IX antibodies. *Nature* **303**, 181–182.

GIANELLI, F., CHOO, K.H., WINSHIP, P.R., ANSON, D.S., REES, D.J.G., FERRARI, N., RIZZA, C.R. AND BROWNLEE, G.G. (1984). Characterisation and use of an intragenic polymorphic marker for detection of carriers of haemophilia B (factor IX deficiency). *Lancet* **i**, 239–241.

GITSCHIER, J., WOOD, W.I., GORALKA, T.M., WION, K.L., CHEN, E.Y., EATON, D.H., VEHAR, G.A., CAPON, D.J. AND LAWN, R.M. (1984). Characterisation of the human factor VIII gene. *Nature* **312**, 321–326.

GOSS, S. AND HARRIS, H. (1977). Gene transfer by means of cell fusion: radiation induced gene segregation. *Journal of Cell Science* **25**, 17–37.

GRAY, J.W., CARRANO, A.V., STEINMETZ, L.L., VAN DILLA, M.A., MOORE, D.H., MAYALL, B.H. AND MENDELSOHN, M.L. (1975). Chromosome measurement and sorting by flow systems. *Proceedings of the National Academy of Sciences of the United States of America* **72**, 1231–1234.

GRUNEBAUM, L., CAZENAVE, J., CAMERINO, G., KLOEPFER, C., MANDEL, J., TOLSTOSHEV, P., JAYE, M., DE LA SALLE, M. AND LECOCQ, J. (1984). Carrier detection of hemophilia B by using a restriction site polymorphism associated with the coagulation factor IX gene. *Journal of Clinical Investigation* **73**, 1491–1495.

GUSELLA, J.F., VARSANYI-BREINER, A., KAO, F.T., JONES, C., PUCK, T.T., KEYS, C.,

ORKIN, S. AND HOUSEMAN, D. (1979). Precise localization of human β-globin gene complex on chromosome 11. *Proceedings of the National Academy of Sciences of the United States of America* **76**, 5239–5243.

GUSELLA, J.F., KEYS, C., VARSANYI-BREINER, A., KAO, F.T., JONES, C., PUCK, T.T. AND HOUSEMAN, D. (1980). Isolation and localization 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., WEXLER, M.S., CONNEALLY, P.M., NAYLOR, S.L., ANDERSON, M.A., TANZI, R.E., WATKINS, P.C., OTTINA, K., WALLACE, M.R., SAKAGUCHI, A.Y., YOUNG, A.B., SHOULSON, I., BONILLA, E. AND MARTIN, J.B. (1983). A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* **306**, 234–239.

HARPER, M.E., ULLRICH, A. AND SAUNDERS, G.F. (1981). Localization 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., O'BRIEN, T., MURRAY, J.M., DAVIES, K.E., PEARSON, P.L. AND WILLIAMSON, R. (1983). The use of linked DNA polymorphisms for genotype prediction in families with Duchenne muscular dystrophy. *Journal of Medical Genetics* **20**, 252–254.

HARTLEY, D.A. (1984). *Molecular Genetics of the Human X Chromosome.* DPhil. Thesis, University of London [3442].

HARTLEY, D.A., DAVIES, K.E., DRAYNA, D., WHITE, R.L. AND WILLIAMSON, R. (1984). A cytological map of the human X chromosome – localisation of genetic markers by *in situ* hybridisation. *Nucleic Acids Research* **12**, 5277–5285.

HULTEN, M. (1974). Chiasma distribution at diakinesis in the normal human male. *Hereditas* **76**, 55–78.

INGLE, C., WILLIAMSON, R., DE LA CHAPELLE, A., HERVA, R.R., HAAPALA, K., BATES, G., WILLARD, H.F., PEARSON, P. AND DAVIES, K.E. (1985). Mapping DMD sequences in a human X chromosome deletion which extends across the region of the DMD mutation. *American Journal of Human Genetics*, in press.

ITAKURA, K., ROSSI, J.J. AND WALLACE, R.B. (1984). Synthesis and use of synthetic oligonucleotides. *Annual Review of Biochemistry* **53**, 323–356.

JACOBS, P.A., HUNT, P.A., MAYER, M. AND BART, R.D. (1981). Duchenne muscular dystrophy (DMD) in a female with an X/autosome translocation: further evidence that the DMD locus is at Xp21. *American Journal of Human Genetics* **33**, 513–518.

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

JOLLY, D.J., ESTY, A.C., BERNARD, H.U. AND FRIEDMAN, T. (1982). Isolation of a genomic clone partially encoding human hypoxanthine phosphoribosyltransferase. *Proceedings of the National Academy of Sciences of the United States of America* **79**, 5038–5041.

JOHN, H.A., BIRNSTIEL, M.L. AND JONES, K.W. (1969). RNA–DNA hybrids at the cytological level. *Nature* **223**, 582–587.

JONES, K.W. (1970). Chromosomal and nuclear location of mouse satellite DNA in individual cells. *Nature* **225**, 912–915.

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.

KAZAZIAN, H.H., ANTONARAKIS, S.E., CHENG, J.-C., BOEHM, C.D. AND WABER, P.G. (1983). DNA polymorphisms in the β-globin gene cluster: use in the discovery of mutations and prenatal diagnosis. In *Banbury Report: Recombinant DNA Applications to Human Disease* 14th Edn (C.T. Caskey and R.L. White, Eds), pp. 29–39. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

KEATS, B.J.B. (1981). *Linkage and Chromosome Mapping in Man.* Population Genetics Laboratory, University of Hawaii, Hawaii.

KEMP, D.J., COPPEL, R.L., COWMAN, A.F., SAINT, R.B., BROWN, G.V. AND ANDERS,

R.F. (1983). Expression of plasmodium falciparum blood-stage antigens in *Escherichia coli:* Detection with antibodies from immune humans. *Proceedings of the National Academy of Sciences of the United States of America* **80**, 3787–3791.

Kidd, V.J., Wallace, R.B., Itakura, K. and Woo, S.L.C. (1983). α_1-Antitrypsin deficiency detection by direct analysis of the mutation in the gene. *Nature* **304**, 230–234.

Kingston, H.M., Thomas, M.F.T., Pearson, P.L., Sarfarazi, M. and Harper, P.S. (1983). Genetic linkage between Becker muscular dystrophy and a polymorphic DNA sequence on the short arm of the X chromosome. *Journal of Medical Genetics* **20**, 255–258.

Kingston, H.M., Sarfarazi, M., Thomas, N.S.T. and Harper, P.S. (1984). Localisation of the BMD gene on the short arm of the X chromosome by linkage to cloned DNA sequences. *Human Genetics* **67**, 6–17.

Krumlauf, R., Jeanpierre, M. and Young, B.D. (1982). Construction and characterisation of genomic libraries from specific human chromosomes. *Proceedings of the National Academy of Sciences of the United States of America* **79**, 2971–2975.

Kunkel, L.M., Smith, K.D. and Boyer, S.H. (1976). Human Y chromosome-specific reiterated DNA. *Science* **191**, 1189–1190.

Kunkel, L.M., Tantravahi, V., Eisenhard, M. and Latt, S.A. (1982). Regional localisation on the human X of DNA segments cloned from flow sorted chromosomes. *Nucleic Acids Research* **10**, 1557–1578.

Lamar, E.E. and Palmer, E. (1984). Y-encoded, species-specific DNA in mice: evidence that the Y chromosome exists in two polymorphic forms in inbred strains. *Cell* **37**, 171–177.

Lane, C.D., Marbaix, G. and Gurdon, J.B. (1971). Duck haemoglobin synthesis in frog cells: the translation of reticulocyte 95 RNA in frog oocytes. *Journal of Molecular Biology* **61**, 73–92.

Laurie, D.A., Hulten, M. and Jones, G.H. (1981). Chiasma frequency and distribution in a sample of human males. *Cytogenetics and Cell Genetics (Basel)* **31**, 153–166.

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

Lindenbaum, R.H., Clarke, G., Patel, C., Moncrieff, M. and Hughes, J.T. (1979). Muscular dystrophy in an X:1 translocation female suggests that Duchenne locus is on X chromosome short arm. *Journal of Medical Genetics* **16**, 389–392.

Lindgren, V., De Martinville, B., Horwich, A.L., Rosenberg, L.E. and Francke, U. (1984). Human ornithine transcarbamylase locus mapped to band Xp21.1 near the Duchenne muscular dystrophy locus. *Science* **226**, 698–700.

Little, P.F.R. (1981). Antenatal diagnosis of haemoglobinopathies. In *Genetic Engineering* (R. Williamson, Ed.), pp. 61–102. Academic Press, London.

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

Malcolm, S., Barton, P., Murphy, C. and Ferguson-Smith, M.A. (1981). Chromosomal localisation of a single-copy gene by *in situ* hybridisation – human β-globin genes on the short arm of chromosome 11. *Annals of Human Genetics* **45**, 135–141.

Maniatis, T., Frisch, E.F. and Sambrook, J. (1982). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, New York.

Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Connel, C., Quon, D., Sim, G.K. and Efstratiadis, A. (1978). The isolation of structural genes from libraries of eukaryotic DNA. *Cell* **15**, 687–701.

Martin, J.B. (1982). Huntington's disease: genetically programmed cell death in the

human central nervous system. *Nature* **299**, 205–206.

MILLER, C.L. AND RUDDLE, F.H. (1978). Co-transfer of human X-linked markers into murine somatic cells via isolated metaphase chromosomes. *Proceedings of the National Academy of Sciences of the United States of America* **75**, 3346–3350.

MOHR, J. (1951). A search for linkage between the Lutheran blood group and other hereditary characters. *Acta pathologica et microbiologica scandinavica. Section A: Pathology* **28**, 207–210.

MOHR, J. (1954). *Opera ex Domo Biologiae Hereditariae Humanae Universitatis Hafniensis*, 33rd Edn. Munksgaard, Copenhagen.

MORGAN, T.H. (1911). An attempt to analyse the constitution of the chromosomes on the basis of sex-linked inheritance in *Drosophila*. *Journal of Experimental Zoology* **11**, 365–414.

MORGAN, T.H. AND CATTELL, E. (1912). Data for the study of sex-linked inheritance in *Drosophila*. *Journal of Experimental Zoology* **13**, 79–101.

MOSER, H. (1984). Duchenne muscular dystrophy: pathogenic aspects and genetic prevention. *Human Genetics* **66**, 17–40.

MURRAY, J.M., DAVIES, K.E., HARPER, P.S., MEREDITH, L., MULLER, C. 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.

OLSEN, A.S., MCBRIDE, O.W. AND OTEY, M.C. (1980). Isolation of unique sequence human chromosomal deoxyribonucleic acid. *Biochemistry* **19**, 2419–2428.

ORKIN, S.H., ANTONARAKIS, S.E. AND KAZAZIAN, H.H. (1983). Polymorphism and molecular pathology of the human β-globin gene. *Progress in Hematology* **13**, 49–73.

ORKIN, S.H., MARKHAM, A.F., AND KAZAZIAN, H.H. (1983). Direct detection of the common Mediterranean β-thalassaemia gene with synthetic DNA probes: an alternative approach for prenatal diagnosis. *Journal of Clinical Investigation* **71**, 775–779.

ORKIN, S.H., LITTLE, P.F.R., KAZAZIAN, H.H. AND BOEHM, C.D. (1982a). Improved detection of the sickle mutation by DNA analysis. *New England Journal of Medicine* **307**, 32–36.

ORKIN, S., KAZAZIAN, H.H., ANTONARAKIS, S.E., GOFF, S.C., BOEHM, C.D., SEXTON, J.P., WABER, P.G. AND GIARDINA, P.J.V. (1982b). Linkage of β-thalassaemia mutations and β-globin gene polymorphisms with DNA polymorphisms in human β-globin gene cluster. *Nature* **296**, 627–631.

PARDUE, M.L. AND GALL, J.G. (1970). Chromosomal localisation of mouse satellite DNA. *Science* **168**, 1356–1358.

PEARSON, P.L. AND VAN OMMEN, G.B. (1984). Recent developments in DNA research of Duchenne muscular dystrophy. In *Research into the Origin and Treatment of Muscular Dystrophy* (L.P. Ten Kate, P.L. Pearson and A.M. Stadhouders, Eds), pp. 91–100. Excerpta Medica, Amsterdam.

PELHAM, H.R.B. AND JACKSON, R.J. (1976). An efficient mRNA-dependent translation system from reticulocyte lysates. *European Journal of Biochemistry* **67**, 247–256.

PEMBREY, M.E., DAVIES, K.E., WINTER, R.M., ELLES, R.G., WILLIAMSON, R., FAZZONI, T.A. AND WALKER, C. (1984). The clinical use of DNA markers linked to the gene for Duchenne muscular dystrophy. *Archives of Disease in Childhood* **59**, 208–216.

PIRATSU, M., KAN, Y.W., CAO, A., CONNER, B.J., TEPLITZ, R.L. AND WALLACE, R.B. (1983). Prenatal diagnosis of β-thalassaemia. Detection of a single nucleotide mutation in man. *New England Journal of Medicine* **309**, 284–287.

PONCZ, M., BALLANTINE, M., SOLOWIEJCZYK, D., BARAK, I., SCHWARTZ, E. AND SURREY, S. (1982). β-Thalassaemia in a Kurdish Jew. *Journal of Biological Chemistry* **257**, 5994–5996.

RACE, R.R. AND SANGER, R. (1975). *Blood Groups in Man*, 6th Edn. Blackwells, Oxford, London.

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

RENWICK, J.H., BUNDEY, S.E., FERGUSON-SMITH, M.A. AND FZATT, M.M. (1971). Confirmation of linkage of the loci for myotonic dystrophy and ABH sec. *Journal of Medical Genetics* **8**, 407–416.

SCAMBLER, P., ROBBINS, T., GILLIAM, T.C., BOYLSTON, A., TIPPETT, P., WILLIAMSON, R. AND DAVIES, K.E. (1985). Linkage studies between polymorphic markers on chromosome 4 and cystic fibrosis. *Human Genetics*, in press.

SCHMECKPEPER, B.J., SMITH, K.D., DORMAN, B.P., RUDDLE, F.H. AND TALBOT, C.C. (1979). Partial purification and characterisation of DNA from the human X chromosome. *Proceedings of the National Academy of Sciences of the United States of America* **76**, 6525–6528.

SHORTLE, D., DI MAIO, D. AND NATHANS, D. (1981). Directed mutagenesis. *Annual Review of Genetics* **15**, 265–294.

SOLOMON, E. AND BODMER, W.F. (1979). Evolution of sickle variant gene. *Lancet* **i**, 923.

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

STURTEVANT, A.H. (1913). The linear arrangement of six sex-linked factors in Drosophila as shown by their mode of association. *Journal of Experimental Zoology* **14**, 43–59.

SUTTON, W.S. (1903). The chromosomes in heredity. *Proceedings of the National Academy of Sciences of the United States of America* **4**, 231–251.

TOOLE, J.J., KNOPF, J.L., WOZNEY, J.M., SULTZMAN, L.A., BUECKER, J.L., PITTMAN, D.D., KAUFMAN, R.J., BROWN, E., SHOEMAKER, C., ORR, E.C., AMPHLETT, G.N., FOSTER, W.B., COE, M.L., KNUTSON, G.J., FASS, D.N. AND HEWICK, R.M. (1984). Molecular cloning of a cDNA encoding human antihaemophilic factor. *Nature* **312**, 342–347.

TREISMAN, R., ORKIN, S.H. AND MANIATIS, T. (1983). Specific transcription and RNA splicing defects in five cloned β-thalassaemia genes. *Nature* **302**, 591–596.

UTERMANN, G., VOGELBER, K.H., STEINMETZ, A., SCHOENBORN, W., PRUIN, N., JAESCHKE, M., HESS, M. AND CANZLER, H. (1979). Polymorphism of apolipoprotein E. II. Genetics of hyperlipoprotein type III. *Clinical Genetics* **15**, 37–62.

VERELLEN-DUMOULIN, C., FREUND, M., DE MEYER, R., LATERRE, C., FREDERIC, J., THOMPSON, M.W., MARKOVIC, V.D. AND WORTON, R.G. (1984). Expression of an X-linked muscular dystrophy in a female due to translocation involving Xp21 and non-random inactivation of the normal X chromosome. *Human Genetics* **67**, 115–119.

VOLKERS, W.S., WENT, L.N., VEGTER-VANDER VLIS, M., HARPER, P.S. AND CARO, A. (1980). Genetic linkage studies in Huntington's chorea. *Annals of Human Genetics* **44**, 75–79.

WALLACE, R.B., JOHNSON, M.J., HIROSE, T., MIYAKE, T., KAWASHIMA, E.H. AND ITAKURA, K. (1981). The use of synthetic oligonucleotides as hybridization probes. II. Hybridization of oligonucleotides of mixed sequence to rabbit β-globin DNA. *Nucleic Acids Research* **9**, 879–894.

WEATHERALL, D.J. (1984). The Thalassaemias. In *Molecular Biology and Human Disease* (A. Macleod and K. Sikura, Eds), pp. 105–130. Blackwell Scientific Publications, Oxford.

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

WIEACKER, P., DAVIES, K.E., COOKE, H.J., PEARSON, P.L., WILLIAMSON, R., BHATTACHARYA, S., ZIMMER, J. AND ROPERS, H-H. (1983). Towards a complete linkage

map of the human X chromosome: regional assignment of 16 cloned single-copy DNA sequences employing a panel of somatic cell hybrids. *American Journal of Human Genetics* **36**, 265–276.

WILSON, J.T., MILNER, P.F., SUMMER, M.E., NALLASETH, F.S., FADEL, H.E., REINDOLLAR, R.H., MCDONOUGH, P.G. AND WILSON, L.B. (1982). Use of restriction endonucleases for mapping the allele for β^s-globin. *Proceedings of the National Academy of Sciences of the United States of America* **79**, 3628–3631.

WILSON, J.T., WILSON, L.B., DERIEL, J.K., VILLA-KOMAROFF, L., EFSTRATIADIS, A., FORGET, B.G. AND WEISSMAN, S.M. (1978). Insertion of synthetic copies of human globin genes into bacterial plasmids. *Nucleic Acids Research* **5**, 563–581.

WOLF, S.F., MARENI, C. AND MIGEON, B. (1980). Isolation and characterisation of cloned DNA sequences that hybridise to the human X chromosome. *Cell* **21**, 95–102.

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.

WOOD, W.I., CAPON, D.J., SIMONSEN, C.C., EATON, D.L., GITSCHIER, J., KEYT, B., SEEBURG, P.H., SMITH, D.H., HOLLINGSHEAD, P., WION, K.L., DELWART, E., TUDDENHAM, E.G.D., VEHAR, G.A. AND LAWN, R.M. (1984). Expression of active human factor VIII from recombinant DNA clones. *Nature* **312**, 330–337.

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.

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. AND DAVIES, K.E. (1983). Construction of a DNA library from the human X chromosome. In *Cytogenetics of the Mammalian X Chromosome* (A.A. Sandberg, Ed.), pp. 479–491. A.R. Liss, New York.

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.

YUNIS, J.J. (1981). Mid-prophase human chromosomes. The attainment of 2000 bands. *Human Genetics* **56**, 293–298.

ZATZ, M., VIANNA-MORGANTE, A.M., CAMPOS, P. AND DIAMENT, A.J. (1981). Translocation (X;6) in a female with Duchenne muscular dystrophy: implications for the localisation of the Duchenne muscular dystrophy locus. *Journal of Medical Genetics* **18**, 442–447.