REVIEW ARTICLE

An appraisal of the application of recombinant DNA techniques to chromosome defects

Kim P. ELLIS* and Kay E. DAVIES†

*Biochemistry Department, St. Mary's Hospital Medical School, London W2 1PG, and

†Nuffield Department of Clinical Medicine, John Radcliffe Hospital, Headington, Oxford OX3 9DU, U.K.

Until the first gene was cloned over 10 years ago, strategies designed to gain insight into human disease relied on the isolation and characterization of gene products. The scientist could only speculate about the molecular basis of a given phenotype. If the biochemical defect is not known, as is still the case for many chromosomal disorders, the task of understanding the disease process is an even more formidable one.

DNA recombinant technology provides the scientist with the necessary tools to examine mutations in man at the level of the DNA sequence and to determine how these influence expression of the gene product. Analysis of genetic diseases, such as the thalassaemia syndromes, has given the molecular biologist great insight into the molecular mechanisms involved in gene expression because these genetic defects involve mutation of consensus sequences needed for transcription, correct processing and translation of the mRNA. More recently, DNA probes have been used as genetic tools for mapping the human genome and recombinant DNA analyses of rearrangements in cancer are revealing some exciting clues to this complex class of human disease. In this short review, we attempt to summarize the impact of some of the most significant advances made in the elucidation of the basis of human chromosomal disorders.

Direct analysis of genetic disease

If a gene product (or lack of gene product) is known to be responsible for a particular phenotype, then it is relatively straightforward to isolate and examine the DNA sequence of the patient. The gene transcript is cloned from normal cells as a cDNA molecule which is then used as a probe in patients suffering from the disease. If the gene is expressed at high levels in a particular cell, such as globin gene in erythropoietic cells, the correspond-

ing cDNA can be cloned by purification of the mRNA (Wilson *et al.*, 1978). Alternatively, if a mutant cell line exists, the gene may be cloned by integrating foreign DNA into the cell and investigating the sequences that complement the mutation (Jolly *et al.*, 1982). Whenever the protein sequence is known, a short corresponding DNA sequence can be predicted and synthesized (Woods *et al.*, 1982). The resulting oligonucleotide is used to screen a cDNA library of a tissue in which the gene is known to be expressed. Often the human protein sequence is unknown but the protein has been characterized in another species. In this case the corresponding sequence is used to probe a human library at an appropriately low stringency (Choo *et al.*, 1982). Finally, if protein sequence data is unavailable, but an antibody to the protein product has been raised, the gene sequence may be identified by direct expression of human cDNA sequences in *Escherichia coli* and their subsequent identification by radioimmunoassay (Young & Davis, 1983).

When the corresponding cDNA clone has been isolated for a gene product, then the normal mechanisms required to produce that protein can be elucidated by obtaining genomic clones in which DNA surrounding the coding region can be sequenced completely to cover flanking control elements and internal intervening sequences. This DNA code can then be interpreted in terms of the transcription and translation of the RNA in normal and mutant alleles. The molecular map of a gene provides the foundation for the understanding of the mutations responsible for a given genetic disorder. Such an analysis of the β -globin cluster on chromosome 11 has demonstrated that there are a large number of different mutations that lead to a restricted number of clinical and biochemical phenotypes (for review see Wood & Weatherall, 1983). Examples are found where transcription does not occur because the gene is completely or partially deleted, or where mutation occurs in the controlling 5' region of the gene, where transcription does occur but the mRNA is not processed,

Abbreviations used: RFLP, restriction fragment length polymorphism; cML, chronic myelocytic leukaemia; BL, Burkitt's lymphoma; DMD, Duchenne muscular dystrophy; CF, cystic fibrosis.

where the mRNA is not processed correctly, where premature termination of translation occurs and where a change of base pair causes the synthesis of an altered protein (Orkin, 1983). All of these defects can be analysed in detail by recombinant techniques such as Southern blotting of DNA, Northern blotting of RNA and transient expression of the mutant sequences in appropriate cell lines (for review see Dahl *et al.*, 1981; Nienhuis *et al.*, 1984).

Over the last few years, the variety of cloned human genes has increased dramatically. Some of those that are important clinically are shown in Table 1. Theoretically, the scientist can determine the molecular basis of all the clinical phenotypes which result from the modification in expression of these gene products. What is apparent, however, is that most of these mutations cannot be detected by a simple restriction enzyme analysis of the patient's DNA. Thus, if diagnosis of these diseases is to become efficient, a different approach has to be adopted.

DNA polymorphism; genotype prediction and diagnosis

In 1978, Kan and Dozy reported that many Blacks suffering from sickle cell anaemia lacked a *HpaI* site at the 3' flanking region of their β -globin gene (Kan & Dozy, 1978). The linkage disequilibrium between a restriction fragment length polymorphism (RFLP) and the sickle cell mutation in this ethnic group has led to the more extensive studies of DNA polymorphisms in the β -globin cluster. RFLP's are phenotypically neutral alter-

ations in DNA sequence that occur throughout the genome in all individuals and which may be detected by changes in the pattern of restriction enzyme fragments. Kazazian *et al.* (1983) defined 12 polymorphic sites within the β -gene cluster and demonstrated that nine of these occur in all racial groups, whereas the other three only occur in Blacks. Analysis of many normal individuals and patients has shown that only a relatively small number of common haplotypes exist with respect to restriction enzyme pattern. (A haplotype is a group of restriction enzyme sites in a given region of DNA.) Furthermore, they demonstrated that the different haplotypes are associated with different β -thalassaemia mutations such that in family studies haplotypes could be used for predictive diagnosis (Orkin *et al.*, 1983).

Linkage disequilibrium observed in the β -globin cluster has not yet been reported for many other gene sequences in the human genome. However, where the exact mutation is unknown, the linkage approach in families has proved to be very useful. Often, a combination of restriction enzymes is required to determine a parental haplotype. This is elegantly shown for the phenylalanine hydroxylase locus in the diagnosis of classical phenylketonuria (Woo *et al.*, 1983). In Fig. 1, both mother and proband are homozygous for the 7 kb *SphI* band, whereas the father is heterozygous. Thus the mutant phenylalanine hydroxylase gene must be associated with the 7 kb allele in the father which he has also transmitted to an unaffected sibling homozygous for the 7 kb band. Since this sibling is also homozygous for the 4 kb *HindIII* band, the

Table 1. *Diseases where DNA probes are available*

Disease	Gene probe
α -Thalassaemia	α -Globin } (Old & Higgs, 1983)
β -Thalassaemia	
Collagen disorders	Collagen (Sykes, 1983; Prockop & Kivirikko, 1984)
Dwarfism	Human growth hormone (Moore <i>et al.</i> , 1982)
Emphysema	α_1 -Antitrypsin (Kidd <i>et al.</i> , 1983)
Lesch-Nyhan syndrome	Hypoxanthine phosphoribosyltransferase (Jolly <i>et al.</i> , 1982; Brennand <i>et al.</i> , 1982)
Phenylketonuria	Phenylalanine hydroxylase (Woo <i>et al.</i> , 1983)
Christmas disease	Factor IX (Choo <i>et al.</i> , 1982)
Haemophilia	Factor VIII
Glucose-6-phosphate dehydrogenase deficiency	Glucose-6-phosphate dehydrogenase (Persico <i>et al.</i> , 1981)
Ornithine transcarbamoylase deficiency	Ornithine transcarbamoylase (Old <i>et al.</i> , 1985)
Antithrombin-3 deficiency	Antithrombin-3 (Bock <i>et al.</i> , 1983)
Cholesterol metabolism/heart disease	Low density lipoprotein receptor (Russell <i>et al.</i> , 1983); 3-hydroxy-3-methylglutaryl-CoA reductase (Chin <i>et al.</i> , 1982); apolipoproteins (Tolleshaug <i>et al.</i> , 1983)
Cancer	Oncogenes (Hamlyn & Sikora, 1983)
Leukaemia, lymphomas	Immunoglobulins (Korsmeyer <i>et al.</i> , 1983; Arnold <i>et al.</i> , 1983)
Immune deficiencies	T-cell receptor (Toyonaga <i>et al.</i> , 1984)

Fig. 1. Haplotype analysis of a Danish phenylketonuria family

DNA samples were digested with either *Hind*III (a) or *Sph*I (b). Lane 1, mother; lane 2, father; lane 3, proband; lane 4, an unaffected sibling; lane 5, a second unaffected sibling. (c) Diagram showing the haplotypes of the phenylalanine hydroxylase gene and the segregation of phenylketonuria alleles (*) in this family. Designations as in (a) and (b). [Reproduced from Woo *et al.* (1983) with kind permission from Macmillan Journals Ltd.]

paternal mutant phenylalanine hydroxylase gene must have the haplotype *Hind*III 4.0:*Sph*I 7.0. The maternal mutant phenylalanine hydroxylase gene must have the haplotype *Hind*III 4.2:*Sph*I 7.0 since both parents and proband are heterozygous for the 4.2kb and 4.0kb *Hind*III bands. Thus the

use of two restriction enzymes enables the informative analysis of this mating.

Analysis of chromosomal deletions, rearrangements and duplications

McKusick's catalogue of mapped genes and

random DNA sequences is now expanding so rapidly that it should soon be possible to analyse the chromosomal rearrangements not visible by conventional cytogenetic analysis (McKusick, 1983). Deletions or duplications of chromosomal segments can be detected by DNA dosage provided a DNA probe is characterized that normally lies in the affected region. Without prior cytogenetic indications of deletion or duplications, this type of analysis is still too time-consuming to be of any practical use. However, if a particular deletion is suspected by cytogenetic analysis and a DNA probe is available, this approach can be very valuable. DNA probes have recently been used to detect Down's syndrome (trisomy 21) (Davies *et al.*, 1984). Here the DNA probes can be particularly helpful as occasionally the extra portion of chromosome 21 is translocated elsewhere in the human genome and is not always easily detectable cytogenetically. Chromosome 21-specific DNA probes which are polymorphic can also be used to determine the parental origin of non-disjunction and, in some instances, the stage of meiosis at which it occurred (Davies *et al.*, 1984). This is important for future genetic counselling and will help in the understanding of the factors involved in the non-disjunction resulting in the chromosomal trisomy.

Chromosome rearrangements in cancer

Studies of chromosomal rearrangements in tumours have shown a notable correlation between mapped cellular oncogenes (a class of genes implicated in the development of tumours), and certain consistently occurring breakpoints (Yunis, 1983; LeBeau & Rowley, 1984). Cytogenetic studies coupled with the use of DNA probes specific for oncogene sequences have made a great contribution towards our understanding of certain types of carcinogenesis.

The first oncogenes were characterized in animal retroviruses and are believed to have derived from normal cellular genes which have become integrated into viral genomes (Duesberg, 1983; Stehelin *et al.*, 1976; Spector *et al.*, 1978; Bishop, 1981; Scolnick *et al.*, 1973; Scolnick & Parks, 1974). The ability of the viral oncogenes to transform may result from inappropriate expression of a normal gene product (Huebner & Todaro, 1969; Vande Woude *et al.*, 1980), or from subtle differences between the viral oncogenes and their cellular counterparts (Schwartz *et al.*, 1983; Takeya & Hanafusa, 1983; Ellis *et al.*, 1982; Dhar *et al.*, 1982; Capon *et al.*, 1983). Regarding the idea of gene activation in carcinogenesis, translocation events represent one apparent means of activation and are the best studied chromosomal rearrangements involved in tumour development.

One example is the Philadelphia chromosome, which is the abnormal chromosome 22 found in the haemopoietic cells of 96% of all patients with chronic myelocytic leukaemia (CML) (Groffen *et al.*, 1984). The Philadelphia chromosome (Ph') is the result of a translocation between chromosome 22 and chromosome 9 (Rowley, 1973; Sandberg, 1980). The human *c-abl* gene has been localized to chromosome 9, and shown to translocate to the Philadelphia chromosome in CML (Heisterkamp *et al.*, 1982; De Klein *et al.*, 1982) (Fig. 2a). Only a very small region of chromosome 9 is translocated, about 5000kb at the telomere. Sequencing the translocation join of a Ph' chromosome from a CML patient showed the *c-abl* gene to be just 14kb from the breakpoint (Heisterkamp *et al.*, 1983). Groffen *et al.* (1984) studied precise breakpoint locations from 19 patients with CML and found that all the breakpoints on chromosome 22 fall within a small 5.8kb region referred to as the breakpoint cluster region, 'bcr'. More limited data obtained by the same authors suggest that the breakpoint location on chromosome 9 is not as precise, although still limited. The implication is that both the *bcr* on chromosome 22 and the *c-abl* gene are important in the aetiology of CML, although there is currently no genetic data on their precise roles.

The B-cell line derived tumour Burkitt's lymphoma (BL) provides another well-characterized example of apparent *c-oncogene* activation by re-positioning within the genome. In Burkitt's lymphoma cells the characteristic chromosomal rearrangement is a translocation between the long arm of chromosome 8 (8q24) and either the long arm of chromosome 14 (14q32), the short arm of chromosome 2 (2p12), or the long arm of chromosome 22 (22q11). The 8:14 translocation is the most common. The *c-myc* gene has been mapped to 8q14 (Taub *et al.*, 1982; Dalla-Favera *et al.*, 1982). The other chromosome segments involved encode the immunoglobulin genes 14q32 (IgH) (Croce *et al.*, 1979), 2p12 (κ) (Malcolm *et al.*, 1982) and 22q11 (λ) (Erikson *et al.*, 1981). Taub *et al.* (1982) demonstrated that *c-myc* was translocated in eight of 12 Burkitt's cell lines examined (Fig. 2b). The transcriptional activity of *c-myc* in a BL cell line has been examined directly; some authors report increased transcriptional activity (Erikson *et al.*, 1983; Marcu *et al.*, 1983) whereas others have found no significant change (Maguire *et al.*, 1983; Westin *et al.*, 1982; Shen-Ong *et al.*, 1982). There is also evidence for heterogeneity in the breakpoint on chromosomes 8 and 14 (Dalla-Favera *et al.*, 1982; Taub *et al.*, 1982; Erikson *et al.*, 1983). Hybridization *in situ* with a *c-myc*-specific probe has been used to examine the breakpoints in the 2:8 and 8:14 translocations (Davis *et al.*, 1984);

Fig. 2. Diagram of the t(9;22) Philadelphia chromosome translocation (a) and the t(8;14) translocation (b) of Burkitt's lymphoma (Groffen *et al.*, 1984; Dalla-Favera *et al.*, 1982)

in this powerful technique a radioactive probe sequence is hybridized directly against a metaphase spread. In the two cell lines examined that carried 2:8 translocations, it was found that the breakpoint on chromosome 8 had occurred distal to *c-myc* and that the gene had not been translocated. A similar experiment on an 8:14 translocation line supported the data of Erikson *et al.* (1983), showing the translocation of *c-myc* to chromosome 14 (Dalla-Favera *et al.*, 1982; Taub *et al.*, 1982). The chromosomal translocation point in BL can therefore occur on either side of the *c-myc* gene.

Croce *et al.* (1983) have also reported that *c-myc* remains on chromosome 8 in the 8:22 translocation associated with BL. In view of this, and the debate on whether or not *c-myc* expression is elevated in BL cell lines, it seems unlikely that elevated transcription of the gene is the sole mechanism of oncogenesis. A second transforming gene activated in Burkitt's lymphoma has recently been described. This gene (*Blym* 1) is not homologous with *c-myc* and was localized to the short arm of chromosome 1 by hybridization *in situ*; it has been suggested that *Blym* 1 and *c-myc* may be

involved in different stages of neoplasia (Diamond *et al.*, 1983; Morton *et al.*, 1984). Current research is focusing on how the *c-myc* gene becomes activated: this will provide valuable data for the study of similar systems.

Retinoblastoma is an intraocular childhood cancer. The condition may be sporadic or, more rarely, hereditary. Hereditary retinoblastoma is associated with constitutional deletions of 13q14 in 1–5% of cases (Johnson *et al.*, 1982). Sparkes *et al.* (1980) localized esterase D to 13q14 by deletion studies and demonstrated that the non-deletion form of hereditary retinoblastoma was linked to esterase D, indicating that the same locus was involved whether or not a deletion was apparent.

Until recently, the genetics of retinoblastoma were obscure. In hereditary cases it showed autosomal dominant inheritance but with varied expressivity and penetrance (Vogel, 1979; Matsunaga, 1978; Carlson & Desnick, 1979; Fitzgerald *et al.*, 1983). Additionally, when examined at the cellular level only rare retinal cells develop into tumours. Knudson (1971) proposed a model whereby both the sporadic and hereditary forms of retinoblastoma required two mutational events. Individuals inheriting one mutant allele, as in the familial cases, would require only one more mutational event to manifest the disease. For the sporadic cases to be observed, two mutations must occur within a single cell.

Two family studies have shown that, despite the autosomal dominant pattern of inheritance, retinoblastoma is recessive. Strong *et al.* (1981) described a family showing inheritance of retinoblastoma over four generations, where transmission was through eight unaffected individuals. The development of retinoblastoma was associated with constitutional chromosome 13 deletion (13q13.1–q14.5). In carrier parents and unaffected sibs the translocation was balanced by insertion into 3p. Therefore, only individuals monosomic for this region manifested retinoblastoma. Additionally, the development of unilateral tumours in this family indicated that the monosomy alone is not sufficient for tumour development, and supported the two-hit model. Benedict *et al.* (1983) reported a 3-year-old female with bilateral tumour. An assay of the esterase D activity (dosage-dependent) showed 50% normal activity; both parents showed normal activity. A karyotype showed no deletion at 13q14 even at high resolution, suggesting a very small deletion not visible by light microscopy. Analysis of retinoblastoma cells revealed two stem cell lines of different karyotype, the only consistent abnormality being the loss of a 13 chromosome. No esterase D activity could be detected in the tumour cells, suggesting that the deleted 13 chromosome had been retained in both cases. Cavanee *et al.*

(1983) supported this conclusion by utilizing DNA probes that detected RFLP's and distinguished between the maternal and paternally-derived chromosome 13. The data strongly indicate that retinoblastoma is due to a loss of genetic material at 13q14; this contrasts with other classes of cancer that are attributable to activation or gain of genetic material as is the case for oncogenes. The chromosomal events that might give rise to retinoblastoma are considered in detail by Cavanee *et al.* (1983).

In the examples considered above, studies have been facilitated by the availability of gene-specific probes, in the case of the oncogenes, and the use of a polymorphic protein marker and probes that detect RFLP's in the case of retinoblastoma.

Indirect analysis of genetic disease

In contrast to the examples discussed previously, the causative gene defect of most genetic disorders remains unknown. In such cases the defective gene cannot be cloned directly. The approach to analysing such disorders, with the ultimate aim of identifying the genetic locus responsible, is necessarily circuitous. Research into the molecular genetics of Duchenne muscular dystrophy (DMD), a lethal X-linked recessive disorder, illustrates the strategies used in tackling this problem.

DMD represents a model system in that, although the biochemical basis of the disease is unknown, the condition is X-linked which immediately narrows the field of search. Additional data concerning the location of the DMD locus came from rare female cases with a Duchenne phenotype. These showed balanced X-autosomal translocations, all with breakpoints at Xp21. It was therefore proposed that the DMD locus lay within this region (Lindenbaum *et al.*, 1979; Zatz *et al.*, 1981; Jacobs *et al.*, 1981; Verellen-Dumoulin *et al.*, 1984).

The most important advances in studying the genetics of DMD have come from the use of RFLPs. If a particular RFLP is genetically linked to a disease, then predictions can be made concerning familial inheritance of the disease, including carrier status, even if no gene-specific probe for the disease itself is available. The accuracy of such predictions depends upon the extent of the linkage between the RFLP and the disease locus.

A source of cloned human DNA is the prime requirement for obtaining probes for linkage studies. Probes may be isolated from total genomic libraries (Maniatis *et al.*, 1978; Dahl *et al.*, 1981), but as the X chromosome only comprises 5.4% of the human genome (Gray *et al.*, 1975) a method of enrichment for chromosome-specific sequences is desirable. Three main approaches have been

employed to obtain chromosome-specific DNA. Schmeckpeper *et al.* (1979) and Olsen *et al.* (1980) used rodent-human hybrid cell lines containing the human X chromosome. DNA from the cell lines was hybridized with excess rodent DNA so that the human chromosomal DNA remains preferentially single stranded, thus providing a means of isolation. This approach has several limitations: a suitable hybrid cell line must be available, there must be some homology between rodent and human DNA, and repetitive human DNA sequences must not remain single stranded under these conditions. An alternative is to clone DNA from a rodent-human somatic cell line directly. The resulting library is then screened with labelled total human DNA to detect the clones that contain human inserts. Cavanee *et al.* (1984) constructed a chromosome 13 enriched library in this manner. Once again a suitable cell hybrid line is required and the nature of the screening is such that only human inserts carrying repetitive sequences will be detected and the low copy regions must then be separated. The most direct way to obtain chromosome-specific libraries is to isolate the chromosome of interest by flow sort cytometry (Carrano *et al.*, 1979). This technique has been used to obtain human X chromosome libraries (Davies *et al.*, 1981; Kunkel *et al.*, 1982), and a chromosome 21,22 library (Krumlauf *et al.*, 1981).

The ability to localize polymorphic probes is a valuable aid to linkage studies. For example, in the case of DMD, probes can be pre-screened for localization at Xp21 before proceeding to pedigree analysis. In theory, 1600 random probes would map the human genome at 20cM intervals, whereas 150 localized probes would suffice (Botstein *et al.*, 1980; Lange & Boehnke, 1982). Probes can be localized directly by hybridization *in situ* to a metaphase spread (Harper & Saunders, 1981; Malcolm *et al.*, 1982; Gerhard *et al.*, 1981; Harper, 1983), or by hybridizing against DNA from rodent-human cell lines that contain whole or partial human chromosomes. Wieacker *et al.* (1984) and Hartley *et al.* (1984) have recently ordered 16 probes along the X chromosome by these approaches.

The first probe demonstrated to show linkage to DMD was described by Murray *et al.* (1982). The probe, RC8, was isolated from a flow-sorted chromosome library (Davies *et al.*, 1981). In 1983, Davies *et al.* (1983b) presented linkage analysis for RC8 and a second probe L.128. The data showed that the two probes flank the DMD locus, each at a distance of about 15cM ($\theta = 0.15$). RC8 was mapped to Xp21-Xp2.23 and L.128 was localized at Xp11.0-Xp11.3. The linkage data were therefore consistent with the assignment of DMD to the

Xp21 region. Although these probes are not close enough for antenatal diagnosis, they have been useful for carrier status prediction in informative families (Harper *et al.*, 1983; Pembrey *et al.*, 1984). The next year should witness the development of more closely linked probes at Xp21 and perhaps the identification of the mutant locus itself.

Becker dystrophy, which manifests as a milder muscular dystrophy than Duchenne, has recently been shown to be localized in the same region of the human X chromosome (Kingston *et al.*, 1983, 1984). The identification of the mutant sequence will enable the investigation of the possibility that these two disorders are allelic, involving different mutations at the same locus.

The probe L.128 also shows linkage with (0-15cM) retinitis pigmentosa, another serious sex-linked disorder, further demonstrating the usefulness of these techniques (Bhattacharya *et al.*, 1984). More X chromosome specific probes are now available and work on completing an RFLP-defined genetic map of this chromosome is well underway (Drayna *et al.*, 1984). The availability of such a map will make it possible to establish linkage for any X-linked disorder.

Another successful application of recombinant DNA techniques is illustrated by the case of Huntington's disease (Huntington's chorea). As an autosomal dominant, this disease presents no problem regarding the identification of carriers. However, the late age of onset (generally 30-50 years), is troublesome in pedigree studies. Until very recently, there was no knowledge concerning the chromosomal location of the disease. Gusella *et al.* (1983) reported a linkage between G8, a random probe obtained from a total genomic library, and Huntington's chorea. The G8 probe was then localized to the short arm of chromosome 4 by hybridization to rodent-human cell hybrid DNA. Work is now in progress to obtain a flanking marker for the locus.

The successes achieved with DMD and Huntington's chorea indicate that the use of these methods to locate the cystic fibrosis (CF) gene should also prove successful given time. CF is perhaps the most common and serious genetic disorder in Caucasian populations. As in DMD and Huntington's disease, the underlying defect is unknown; unlike DMD, cystic fibrosis is inherited as an autosomal recessive, and there are no data available concerning the chromosomal location of the defect. Present methods for carrier status and antenatal diagnosis are still experimental and depend on the detection of abnormal levels of protein probably not related to the basic defect (Brock, 1984). The lack of a reliable carrier test complicates the analysis of linkage data since either a large number of families, or very large

Fig. 3. Pedigree of the Venezuelan Huntington's disease family

A, B, C and D refer to the *Hind*III haplotypes observed for the linked G8 probe. [Reproduced from Gusella *et al.* (1983) with kind permission from Macmillan Journals Ltd.]

pedigrees, are needed for linkage to be established or excluded. Gusella *et al.* (1983) had access to an extensive Venezuelan family in which Huntington's chorea could be traced back to a common ancestor. A small part of this pedigree is reproduced in Fig. 3. Large pedigrees of this type greatly increase the probability of finding a significant linkage of the disease with a DNA marker. Large CF families among the Amish and Mennonite populations in the U.S.A. have also been described (Klinger, 1983; Schwartz & Doherty, 1984), and will no doubt prove just as valuable in CF research.

In summary, the power of recombinant DNA techniques in permitting the analysis of loci with totally unknown function is confirmed by the research on DMD and Huntington's disease. Cystic fibrosis represents a case for the future, but there can be no doubt that a linked probe will be isolated. The next major step in this field will be the development of techniques that will make it possible to move from a linked DNA sequence to the disease locus.

Gene therapy

The identification of genes involved in the development of clinical phenotypes immediately presents the scientist with the possibility of curing or treating the particular disease. The most obvious treatment is replacement therapy. This is often performed by administration of the missing protein product but means that patients need treatment throughout their lives. Alternatively, it should be theoretically possible to replace non-functional genes with 'good' ones. This is a very difficult process because of the complexity of eukaryotic gene expression. The new gene needs to be introduced such that it is expressed in a tissue-specific manner and regulated correctly. In addition, the final manipulated cell must have a selective advantage. Gene manipulation has

shown some promise in experiments with mice by the injection of immunoglobulin and growth hormone genes into fertilized eggs (Palmiter *et al.*, 1982; Brinster *et al.*, 1983). In the latter case, tissue-specific expression of growth hormone was not observed, but the production of growth hormone in different cells was not found to be detrimental to the mice. However, this approach is not universally successful and has not yet been reported to work for other genes such as globin (Lacy *et al.*, 1983).

The most exciting advance in gene therapy was recently reported by Williams *et al.* (1984). They demonstrated that, using retroviral vectors, it is possible to introduce efficiently exogenous DNA sequences into pluripotent haematopoietic stem cells, and to use bone marrow transplantation to engraft mice with the genetically modified haematopoietic cells. Although the efficiency of this process in man and the safety of retroviral-mediated transformation remain to be evaluated, these experiments represent a significant advance in the development of gene replacement therapy.

A different strategy from replacement therapy can be adopted if the patient has normal fetal genes which can be reactivated in adult life to substitute for the non-functional adult genes. This has met with some success for the treatment of patients with β -thalassaemia and homozygous sickle cell disease, where the administration of the drug 5'-aza-cytidine has been shown to reactivate the fetal γ -globin genes (Ley *et al.*, 1982). The molecular basis of this response is not yet entirely clear.

Finally, scientists are moving into an era where human eggs can be isolated and fertilized. The next few years promise to be no less exciting as DNA recombinant technology begins to unravel the secrets of early embryonic development. The fact that the first gene was cloned within the last decade bears witness to the rapid advances being made in the elucidation of chromosomal disorders in man.

We are grateful to Professor Bob Williamson for his enthusiasm and support and Dr. Doug Higgs for critical reading of the manuscript. We also thank Rachel Kitt (Oxford) for patient typing of the manuscript. This work was supported by the Medical Research Council, the Muscular Dystrophy Group of Great Britain and the Muscular Dystrophy Association of America.

References

- Arnold, A., Cossman, J., Bakshi, A., Jaffe, E. S., Waldmann, T. A. & Korsmeyer, S. J. (1983) *New Engl. J. Med.* **309**, 1593–1599
- Benedict, W. F., Murphree, A. L., Banerjee, A., Spina, C. A., Sparkes, M. C. & Sparkes, R. S. (1983) *Science* **219**, 973–979
- Bhattacharya, S. S., Wright, A. F., Clayton, J. F., Price, W. H., Phillips, C. I., McKeown, C. M. E., Jay, M., Bird, A. C., Pearson, P. L., Southern, E. M. & Evans, H. J. (1984) *Nature (London)* **309**, 253–255
- Bishop, J. M. (1981) *Cell* **23**, 5–6
- Bock, S. C. & Levitan, D. J. (1983) *Nucleic Acids Res.* **11**, 8569–8582
- Botstein, D., White, R. L., Skolnick, M. H. & Davis, R. W. (1980) *Am. J. Hum. Genet.* **32**, 314–331
- Brennand, J., Chinault, A. C., Konecki, D. W., Melton, D. W. & Caskey, C. T. (1982) *Proc. Natl. Acad. Sci. U.S.A.* **79**, 1950–1954
- Brinster, R. L., Ritchie, K. A., Hammer, R. E., O'Brien, R. L., Arp, B. & Storb, U. (1983) *Nature (London)* **306**, 332–336
- Brock, D. J. H. (1984) in *Cystic Fibrosis: Horizons* (Lawson, D., ed.), pp. 1–12, John Wiley and Sons, London and New York
- Capon, D. J., Chen, E. Y., Levinson, A. D., Seeburg, P. H. & Goeddel, D. V. (1983) *Nature (London)* **302**, 33–37
- Carlson, E. A., & Desnick, R. J. (1979) *Am. J. Med. Genet.* **4**, 365–381
- Carrano, A. V., Gray, J. W., Langlois, R. G., Burkhart-Schultz, K. J. & Van Dilla, M. A. (1979) *Proc. Natl. Acad. Sci. U.S.A.* **76**, 1382–1384
- Cavane, W. K., Dryja, T. P., Phillips, R. A., Benedict, W. F., Godbout, R., Gallie, B. L., Murphree, A. L., Strong, L. C. & White, R. L. (1983) *Nature (London)* **305**, 779–784
- Cavane, W. K., Leach, R., Mohandas, T., Pearson, P. & White, R. L. (1984) *Am. J. Hum. Genet.* **36**, 10–24
- Chin, D. J., Luskey, K. L., Faust, J. R., MacDonald, R. J., Brown, M. S. & Goldstein, J. L. (1982) *Proc. Natl. Acad. Sci. U.S.A.* **79**, 7704–7708
- Choo, K. H., Goulet, K. G., Rees, D. J. G. & Brownlee, G. G. (1982) *Nature (London)* **299**, 178–180
- Croce, C. M., Shandair, M., Martinis, J., Cicurel, L., D'Ancona, G. G., Dolby, T. W. & Koprowski, H. (1979) *Proc. Natl. Acad. Sci. U.S.A.* **76**, 3416–3419
- Croce, C. M., Thierfelder, W., Erikson, J., Nishikura, K., Finan, J., Lenoir, G. M. & Nowell, P. C. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 6922–6926
- Dahl, H. H., Flavell, R. A. & Grosveld, F. G. (1981) in *Genetic Engineering 2* (Williamson, R., ed.), pp. 50–118, Academic Press, New York
- Dalla-Favera, R., Bregni, M., Erikson, J., Patterson, D., Gallo, R. C. & Croce, C. M. (1982) *Proc. Natl. Acad. Sci. U.S.A.* **79**, 7824–7827
- Davies, K. E., Young, B. D., Elles, R. G., Hill, M. E. & Williamson, R. (1981) *Nature (London)* **293**, 374–376
- Davies, K. E., Gilliam, C. & Williamson, R. (1983a) *Mol. Biol. Med.* **1**, 185–190
- Davies, K. E., Pearson, P. L., Harper, P. S., Murray, J. M., O'Brien, T., Sarfarazi, M. & Williamson, R. (1983b) *Nucleic Acids Res.* **11**, 2302–2312
- Davies, K. E., Harper, K., Bonthron, D., Krumlauf, R., Polkey, A., Pembrey, M. E. & Williamson, R. (1984) *Hum. Genet.* **66**, 54–66
- Davis, M., Malcolm, S. & Rabbitts, T. H. (1984) *Nature (London)* **308**, 286–288
- De Klein, A., Geurts van Kessel, A., Grosveld, G., Bartram, C. R., Hagemeijer, A., Bootsma, D., Spurr, N. K., Heisterkamp, N., Groffen, J. & Stephenson, J. R. (1982) *Nature (London)* **300**, 765–767
- Dhar, R., Ellis, R. W., Shih, T. Y., Oroszlan, S., Shapiro, B., Maizel, J., Lowy, D. & Scolnick, E. (1982) *Science* **217**, 934–936
- Diamond, A., Cooper, G. M., Ritz, J. & Lane, M. A. (1983) *Nature (London)* **305**, 112–116
- Drayna, D., Davies, K. E., Hartley, D. A., Williamson, R. & White, R. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 2836–2839
- Duesberg, P. H. (1983) *Nature (London)* **304**, 219–226
- Ellis, R. W., Lowy, D. R. & Scolnick, E. M. (1982) *Adv. Viral Oncol.* **1**, 107–126
- Erikson, J., Martinis, J. & Croce, M. (1981) *Nature (London)* **294**, 173–175
- Erikson, J., ar-Rushdi, A., Drwinga, H. L., Howell, P. C. & Croce, C. M. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 820–824
- Fitzgerald, P. H., Stewart, J. & Suckling, R. D. (1983) *Hum. Genet.* **64**, 128–130
- Gerhard, T. S., Kawasaki, E. S., Bancroft, F. C. & Szabo, P. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78**, 3755–3759
- Gray, J. W., Carrano, A. V., Steinmetz, L. L., Van Dilla, M. A., Moore, D. H., Mayall, B. H. & Mendelsohn, M. L. (1975) *Proc. Natl. Acad. Sci. U.S.A.* **72**, 1231–1234
- Groffen, J., Stephenson, J. R., Heisterkamp, N., de Klein, A., Bartram, C. R. & Grosveld, G. (1984) *Cell* **36**, 93–99
- 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. & Martin, J. B. (1983) *Nature (London)* **306**, 234–239
- Hamlyn, P. & Sikora, K. (1983) *Lancet* **ii**, 326–330
- Harper, M. E. (1983) *Banbury Rep.* **14**, 159–165
- Harper, M. E. & Saunderson, G. F. (1981) *Chromosoma* **83**, 431–439
- Harper, P. S., O'Brien, T., Murray, J. M., Davies, K. E., Pearson, P. L. & Williamson, R. (1983) *J. Med. Genet.* **20**, 252–254
- Hartley, D. A., Davies, K. E., Drayna, D., White, R. L. & Williamson, R. (1984) *Nucleic Acids Res.* **12**, 5277–5285

- Heisterkamp, N., Groffen, J., Stephenson, J. R., Spurr, N. K., Goodfellow, P. N., Solomon, E., Carritt, B. & Bodmer, W. F. (1982) *Nature (London)* **299**, 747-749
- Heisterkamp, N., Stephenson, J. R., Groffen, J., Hansen, P. F., de Klein, A., Bartram, C. R. & Grosveld, G. (1983) *Nature (London)* **306**, 239-242
- Huebner, R. J. & Todaro, G. J. (1969) *Proc. Natl. Acad. Sci. U.S.A.* **64**, 1087-1094
- Jacobs, P. A., Hunt, P. A., Mayer, M. & Bart, R. D. (1981) *Am. J. Hum. Genet.* **33**, 513-518
- Johnson, M. P., Ramsay, N., Cervenka, J. & Wang, N. (1982) *Cancer Genet. Cytogenet.* **6**, 29-37
- Jolly, D. J., Esty, A. C., Bernard, H. U. & Friedman, T. (1982) *Proc. Natl. Acad. Sci. U.S.A.* **79**, 5038-5041
- Jolly, D., Okayama, H., Berg, P., Esty, A. C., Filpula, D., Bohleri, P., Johnson, G. G., Shively, J. E., Hunkapillar, T. & Friedman, T. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 477-481
- Kan, Y. W. & Dozy, A. M. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75**, 5631-5635
- Kazazian, H. H., Antonarakis, S. E., Cheng, J.-C., Boehm, C. D. & Waber, P. G. (1983) *Banbury Rep.* **14**, 29-39
- Kidd, V. J., Wallace, R. B., Itakura, K. & Woo, S. L. C. (1983) *Nature (London)* **304**, 230-234
- Kingston, H. M., Thomas, M. F. T., Pearson, P. L., Sarfarazi, M. & Harper, P. S. (1983) *J. Med. Genet.* **20**, 255-258
- Kingston, H. M., Sarfarazi, M., Thomas, N. S. T. & Harper, P. S. (1984) *Hum. Genet.* **67**, 6-17
- Klinger, K. (1983) *Hum. Genet.* **65**, 94-98
- Knudson, A. G. (1971) *Proc. Natl. Acad. Sci. U.S.A.* **68**, 820-823
- Korsmeyer, S. J., Arnold, A., Bakshi, A., Ravetch, J. V., Siebenlist, U., Hieter, P. A., Sharrow, S. O., Leibin, T. W., Kersey, J. H., Poplack, D. G., Leder, P. & Waldmann, T. A. (1983) *J. Clin. Invest.* **71**, 301-313
- Krumlauf, R., Jeanpierre, M. & Young, B. D. (1982) *Proc. Natl. Acad. Sci. U.S.A.* **79**, 2971-2975
- Kunkel, L. M., Tantravahi, V., Eisenhard, M. & Latt, S. A. (1982) *Nucleic Acids Res.* **10**, 1557-1578
- Lacy, E., Roberts, S., Evans, E. P., Burtenshaw, M. D. & Constantini, F. (1983) *Cell* **34**, 343-358
- Lange, K. & Boehnke, M. (1982) *Am. J. Hum. Genet.* **35**, 177-192
- LeBeau, M. M. & Rowley, J. (1984) *Nature (London)* **308**, 607-608
- Ley, T. J., DeSimone, J., Anagou, P., Keller, G. H., Humphries, K., Turner, P. H., Young, N. S., Heller, P. & Nienhuis, A. W. (1982) *New Engl. J. Med.* **307**, 1469-1475
- Lindenbaum, R. H., Clarke, G., Patel, C., Moncrieff, M. & Hughes, J. T. (1979) *J. Med. Genet.* **16**, 389-392
- Maguire, R. T., Robins, T. S., Thorgeirsson, S. S. & Heilman, C. A. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 1947-1950
- Malcolm, S., Barton, P., Murphy, C., Ferguson-Smith, M. A., Bentley, D. L. & Rabbitts, T. H. (1982) *Proc. Natl. Acad. Sci. U.S.A.* **79**, 4957-4961
- Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G. K. & Estradiadis, A. (1978) *Cell* **15**, 687-691
- Marcu, K. B., Harris, L. J., Stanton, L. W., Erikson, J., Watt, R. & Croce, C. M. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 519-523
- Matsunaga, E. (1978) *Am. J. Hum. Genet.* **30**, 406-424
- McKusick, V. A. (1983) *Mendelian Inheritance in Man*, John Hopkins University Press, Baltimore and London
- Moore, D. D., Conkling, M. E. & Goodman, H. M. (1982) *Cell* **29**, 285-286
- Morton, C. C., Taub, R., Diamond, A., Lane, M. A., Cooper, G. M. & Leder, P. (1984) *Science* **223**, 173-175
- Murray, J. M., Davies, K. E., Harper, P. S., Meredith, L., Muller, C. & Williamson, R. (1982) *Nature (London)* **300**, 69-71
- Nienhuis, A. W., Anagnou, N. P. & Ley, T. J. (1984) *Blood* **63**, 738-758
- Old, J. M. & Higgs, D. R. (1983) in *The Thalassaemias: Methods in Haematology* (Weatherall, D. J., ed.), pp. 74-102, Churchill Livingstone, Edinburgh
- Old, J. M., Briand, P. L., Purvis-Smith, S., Howard, N. J., Wilcken, B., Hammond, J., Pearson, P., Cathelineau, L., Williamson, R. & Davies, K. E. (1985) *Lancet*, in the press
- Olsen, A. S., McBride, O. W. & Otey, M. C. (1980) *Biochemistry* **19**, 2419-2428
- Orkin, S. H. (1983) *Banbury Rep.* **14**, 19-28
- Palmiter, R. D., Brinster, R. L., Hammer, R. E., Trumbauer, M. E., Rosenfeld, M. G., Birnberg, N. C. & Evans, R. M. (1982) *Nature (London)* **300**, 611-615
- Pembrey, M. E., Davies, K. E., Winter, R. M., Elles, R. G., Williamson, R., Fazzoni, T. A. & Walker, C. (1984) *Arch. Dis. Child.* **59**, 208-216
- Persico, M. G., Toniolo, D., Nobile, C., D'Urso, M. & Luzzato, L. (1981) *Nature (London)* **294**, 778-780
- Prockop, D. J. & Kivirikko, K. I. (1984) *New Engl. J. Med.* **311**, 376-386
- Rowley, J. D. (1973) *Nature (London)* **243**, 290-293
- Russell, D. W., Yamamoto, T., Schneider, W. J., Slaughter, C. J., Brown, M. S. & Goldstein, J. L. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 7501-7505
- Sandberg, A. A. (1980) *The Chromosomes in Human Cancer and Leukaemia*, Elsevier, New York
- Schmeckpeper, B. J., Smith, K. D., Dorman, B. P., Ruddle, F. H. & Conover-Talbot, C. (1979) *Proc. Natl. Acad. Sci. U.S.A.* **76**, 6525-6528
- Schwartz, D. E., Tizard, R. & Gilbert, W. (1983) *Cell* **32**, 853-869
- Schwartz, R. H. & Doherty, R. A. (1984) in *Cystic Fibrosis: Horizons* (Lawson, D., ed.), p. 187. John Wiley and Sons, London and New York
- Scolnick, E. M. & Parks, W. P. (1974) *J. Virol.* **13**, 1211-1219
- Scolnick, E. M., Rands, E., Williams, D. & Parks, W. P. (1973) *J. Virol.* **12**, 458-463
- Shen-Ong, G. L. C., Keath, E. J., Piccoli, S. P. & Cole, M. D. (1982) *Cell* **31**, 443-452
- Sparkes, R. S., Sparkes, M. C., Wilson, M. G., Towner, J. W., Benedict, W., Murphree, A. L. & Yunis, J. J. (1980) *Science* **208**, 1042-1044
- Sparkes, R. S., Murphree, A. L., Lingua, R. W., Sparkes, M. C., Field, L. L., Funderburk, S. J. & Benedict, W. F. (1983) *Science* **219**, 971-973
- Spector, D. H., Smith, K., Padgett, T., McCombe, P., Roulland-Dussoix, D., Muscovici, C., Varmus, H. E. & Bishop, J. M. (1978) *Cell* **13**, 371-379
- Stehelin, D., Varmus, H. E., Bishop, J. M. & Vogt, P. T. (1976) *Nature (London)* **260**, 170-173

- Strong, L. C., Riccardi, V. M., Ferrell, R. E. & Sparkes, R. S. (1981) *Science* **213**, 1501-1503
- Sykes, B. (1983) *Dis. Markers* **1**, 141-146
- Takeya, T. & Hanafusa, H. (1983) *Cell* **32**, 881-890
- Taub, R., Kirsch, I., Morton, C., Lenoir, G., Swan, D., Tronick, S., Aaronson, S. & Leder, P. (1982) *Proc. Natl. Acad. Sci. U.S.A.* **79**, 7837-7841
- Tolleshaug, H., Hobgood, K. K., Brown, M. S. & Goldstein, J. L. (1983) *Cell* **32**, 941-951
- Toyonaga, B., Yanagi, Y., Suciu-Foca, N., Minden, M. & Mak, T. W. (1984) *Nature (London)* **311**, 385-387
- Vande Woude, G. S., Maizel, J. B., Blair, D. G., Clemens, W. L. & Oskarsson, M. (1980) *Science* **207**, 1222-1224
- Verellen-Dumoulin, Ch., Freund, M., De Meyer, R., Laterre, Ch., Frederic, J., Thompson, M. W., Markovic, V. D. & Worton, R. G. (1984) *Hum. Genet.* **67**, 115-119
- Vogel, F. (1979) *Hum. Genet.* **52**, 1-54
- Westin, E. H., Wong-Staal, F., Gelmann, E. P., Dalla-Favera, R., Papas, T. S., Lautenberger, J. A., Eva, A., Reddy, E. P., Tronick, S. R., Aaronson, S. A. & Gallo, R. C. (1982) *Proc. Natl. Acad. Sci. U.S.A.* **79**, 2490-2494
- Wieacker, P., Davies, K. E., Cooke, H. J., Pearson, P. L., Williamson, R., Southern, E., Zimmer, J. & Ropers, H.-H. (1984) *Am. J. Hum. Genet.* **36**, 265-276
- Williams, D. A., Lenischka, I. R., Nathan, D. G. & Mulligan, R. C. (1984) *Nature (London)* **310**, 476-480
- Wilson, J. T., Wilson, L. B., de Riel, J. K., Villa-Komaroff, L., Efstratiadis, A., Forget, B. G. & Weissman, S. M. (1978) *Nucleic Acids Res.* **5**, 563-581
- Woo, S. L. C., Lidsky, A. S., Guttler, F., Chandra, T. & Robson, K. J. H. (1983) *Nature (London)* **306**, 151-155
- Wood, W. G. & Weatherall, D. J. (1983) *Biochem. J.* **215**, 1-10
- Woods, D. E., Markham, A. F., Ricker, A. T., Goldberger, G. & Colten, H. R. (1982) *Proc. Natl. Acad. Sci. U.S.A.* **79**, 5661-5665
- Young, R. A. & Davis, R. W. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 1194-1198
- Yunis, J. J. (1983) *Science* **221**, 227-236
- Zatz, M., Vianna-Morgante, A. M., Campos, P. & Diamant, A. J. (1981) *J. Med. Genet.* **18**, 442-447