Review article

DNA analysis in human disease

A F WRIGHT

From the Medical Research Council Clinical and Population Cytogenetics Unit, Western General Hospital, Edinburgh, Scotland

SUMMARY The analysis of human DNA using recombinant DNA technology is fast becoming an integral part of the diagnosis, assessment, and prevention of inherited and somatic genetic disease. The rationale underlying these methods of analysis is discussed, and the nature and extent of mutational change in heritable disorders and neoplastic development is outlined.

Over 30 years have passed since the genetic material of most organisms was shown to be deoxyribonucleic acid (DNA). While a considerable amount of information on the fine structure and function of bacterial DNA has emerged during this time, the enormously greater complexity of eukaryotic DNA has, until recently, hindered progress in its elucidation.¹ During the past 10 years important methodological and conceptual advances have occurred that are rapidly changing this and at the same time opening up the study of DNA in man and its role in human pathology. Indeed, this has become one of the most exciting and rapidly growing areas of science.

Technological advances

The advances in recombinant DNA technology,^{2,3} which are now so widely used, entail the ability to recombine small sequences of human or other DNA (up to 50 kilobase pairs) into microbiological cloning vectors such as plasmids, bacteriophage, or cosmids; cleave double stranded DNA at specific sites with bacterial restriction endonucleases, separate the resultant small fragments electrophoretically, denature them to single stranded form and transfer to a solid matrix such as nitrocellulose; and identify these immobilised DNA (or RNA) fragments in picogram amounts by allowing them to reanneal with a specific complementary DNA sequence (probe) labelled either with ³²phosphorus or biotin containing nucleotide derivatives. The probe sequence is usually prepared from a recombinant vector, but it can also be chemically synthesised in the form of small (15–100

base pairs) oligonucleotides specifying the required complementary sequence.

CLONING HUMAN DNA SEQUENCES

The major impact of cloning recombinant DNA sequences has been in the ability to follow through advances in the preparative biochemical or immunochemical purification of proteins, nucleic acids, and even whole chromosomes. Relatively abundant proteins of functional interest have been purified and their amino acid sequences identified either in whole or in part. This sequence can be used as the basis for the chemical synthesis of a mixture of possible nucleotide bases coding for a subset of five to 10 amino acids. These are then used as probes to identify recombinant clones containing the true homologous sequence in “libraries” of non-specifically cloned expressed (mRNA) sequences (cDNA library) or of total DNA (genomic library).^{4,5} Genes for less abundant proteins can be isolated by enrichment for the specific mRNA by immunoprecipitation of the polysome bound message, or, more effectively, by screening cDNA libraries cloned into “expression” vectors with specific antibody, permitting direct identification of suitable recombinant clones.^{6,7} Strategies entailing the isolation of a series of overlapping clones have made it possible to map and ultimately sequence large genes such as human factor VIII, which span almost 200 kb of DNA.⁸

DNA obtained from specific human chromosomes separated by flow cytometry or isolated from rodent-human somatic cell hybrid cells can also be digested, inserted into a cloning vector, and amplified to form a chromosomal library.⁹ This can be used to isolate short unique sequences, more or less at random, from the chromosomal library. Short unique segments are

unlikely to be coding for functional products but are potentially useful as markers for gene mapping.¹⁰ There is also increasing interest in the development of strategies that permit larger amounts of DNA to be cloned, which could help in the isolation of disease genes whose function is entirely unknown but whose location has first been identified by mapping with random probes. Potential strategies include the transfection of human DNA into non-human cells in culture and selection by means of cotransfected genes coding for a biochemically selectable product.¹¹ Clones showing stable integration of large human DNA fragments (in the region of 100–10 000 kb) may be isolated in this way.¹² Other potentially useful techniques that may lead to the cloning of disease genes in the absence of information on the gene product include selective cloning of coding, as opposed to non-coding sequences (open reading frames),¹³ and pulse-field gradient electrophoresis, which can separate large DNA fragments of up to 2000 kb in length.¹⁴

RESTRICTION ENDONUCLEASE AND HYBRIDISATION ANALYSIS

DNA analysis in man depends on the isolation of pure DNA from tissues such as blood, tumour, amniotic fluid cells or fibroblasts. The method of DNA extraction can be critical, particularly with regard to minimising degradation and obtaining satisfactory restriction endonuclease digestion.¹⁵ More than 400 bacterial restriction endonucleases are available with different specificities, which cleave double stranded DNA at specific sequences of usually four to six base pairs.¹⁶ The resultant DNA fragment sizes vary from about 200 base pairs to 166 kb using conventional four or six base "cutters", but this can be increased to about 700 kb with seven to 10 base cutters. Enzymes that cut only occasionally may become increasingly useful for the isolation and subsequent analysis of large DNA fragments.

The fundamental step in all molecular genetic analysis is the reannealing or hybridisation of single stranded DNA (immobilised or otherwise) to the denatured and single stranded complementary probe sequence. The length of the cloned probe can vary between a few hundred base pairs and 30–40 kb, although with larger genomic fragments high background labelling due to the inevitable dispersed repetitive sequences becomes a problem.¹⁷ Synthetic oligonucleotide probes are considerably smaller but are capable of forming stable hybrids under certain conditions when greater than about 15 base pairs in length.^{4,18} Probe sequences are generally considerably smaller than the restriction fragments containing their complementary sequence. Most of our understanding of the kinetics and stability of DNA

duplex formation comes from studies of DNA hybridisation in solution rather than from studies of matrix immobilised fragments, although the conclusions are thought to be qualitatively similar.^{1,19,20} Variations in hybridisation conditions are so important in DNA analysis that a brief consideration of the major variables is necessary.

Reannealing of complementary fragments occurs initially by the association of a few pairs of complementary nucleotide bases by hydrogen bonding (nucleation), allowing the remainder of the sequence to reanneal if it is in the correct register.²¹ The rate of duplex formation is influenced by the length and concentration of complementary DNA fragments, the pH and temperature of the reaction, the concentration of formamide or other denaturing agents (if present) and the salt concentration.^{20,22} The rate of formation and stability of the resultant duplex is critically dependent on the hybridisation conditions, degree, and location of base mismatching.²³ The conditions favouring duplex formation and stability include high salt (1 M (Na⁺)), which may serve to neutralise the repulsion of phosphate residues on opposite strands, and a temperature that is about 20–25°C below that at which the duplex melts (T_m; T_m is the temperature at which a given DNA duplex is 50% associated at equilibrium. A T_m value of about 85°C is expected for total human DNA probed with a precisely matched sequence 1 kb in length.) The higher the temperature, the greater the requirement for precise base matching. Hybridised filters are commonly washed either at increased *stringency* (T_m–15°C, 0.015–0.15M (Na⁺)) or reduced stringency (T_m–30°C, 0.9M (Na⁺)), depending on the intention to identify or eliminate signals from poorly matched hybrids. The limits of resolution obtained in this way are such that on the one hand synthetic oligomers, only 17 base pairs in length, can form a stable duplex with human genomic DNA and yet fail to do so if there is a single base mismatch within that sequence.^{24,25} This method of directly recognising a base change within a functionally important sequence is likely to be increasingly important in the diagnosis of inherited disease and in the detection of mutant oncogenes in neoplastic tissues. At the other extreme of resolution, families of related sequences that have diverged by as much as 30% can be detected using appropriately permissive hybridisation conditions.²²

In contrast to DNA from prokaryotic organisms, human DNA is highly heterogeneous, as shown by the fact that it reassociates at rates that vary by about seven orders of magnitude.²⁰ This is because sequences that are repeated throughout the genome are at higher concentration and therefore will reassociate much more rapidly than unique or single copy sequences for a given amount of total DNA. On this

basis human DNA can be arbitrarily classified (depending on hybridisation conditions, fragment size, etc) into four major groups: a zero time reassociation class (6–9%), which probably contains inverted repeated sequences (fold back DNA); a rapidly reassociating class (22%), which contains sequences repeated about 10^3 – 10^6 times per haploid genome; an intermediate reassociating class (13%) containing sequences repeated 10 – 10^3 times; and finally, a slowly reassociating class (5%) representing sequences that are either unique or repeated less than 10 times and so only present at very low concentrations in genomic DNA.²⁰ Some highly repeated sequences occur in blocks of tandem repeats—for example, satellite DNA—others are highly dispersed throughout the genome and can occur as well defined families of related sequences, such as the Alu and Kpn families.²⁶ The dispersed repeats tend to occur in human DNA once every two to 10 kb, which complicates the isolation of unique sequences. Genomic DNA is mostly probed with unique or low repeat sequences.

Mutation and disease

THE EXTENT OF GENETIC DISEASE

It has been estimated that only about 2–5% of human DNA actually contains coding sequences.^{27 28} These estimates are based on analyses of DNA and messenger RNA complexity and on the ratios of coding to non-coding sequences in well characterised genes. This suggests that man has some 50 000–75 000 genes. It is also known that some genes are arranged in clusters of structurally or functionally related units. Unlike the bacterial operon, these are not transcribed together, although they may be coordinately regulated. The proportion of genes arranged in clusters and average cluster size is unknown, but known clusters vary in size from two up to as many as a few hundred genes. McKusick's catalogue of human genetic disorders²⁹ lists more than 2000 genetic syndromes attributed to single genes, which together give rise to major genetic abnormality in about 1% of live births.²⁶ Even more common are the group of disorders in which the mode of inheritance is complex.

In addition to inherited disease, there is increasing direct and indirect evidence that somatic mutations occurring in a smaller number of specific genes confer susceptibility to disorders associated with aging such as cancer and certain types of neuronal and cardiovascular degeneration.^{30 31} Germinal mutation rates in man have been estimated to be in the region of 10^{-6} gametes per locus per generation (10^{-4} – 10^{-8}), while estimates of spontaneous somatic mutation, as measured *in vitro*, are some orders of magnitude higher, even when the number of replicated cycles is taken into account.³² Copy errors occurring during

DNA replication account for most mutations, leading mainly to base pair substitutions, although changed codon reading frames arising from insertion or deletion of bases (frameshifts) and larger structural rearrangements also occur. While it is hazardous to extrapolate from *in vitro* data, it is likely that mutation is occurring at an appreciable rate in somatic cells, more so in rapidly dividing cells, so that mutations that increase growth may contribute to neoplasia in dividing tissues, while in non-dividing tissues mutations causing loss of function and cell death may lead to degenerative changes. DNA repair and immunological or other means of removing mutant somatic cells are likely to have a vital role in maintaining the balance of health and disease.³⁰

The functional consequences of mutation, whether within germinal or somatic cells, depend both on the gene product and on the nature and location of the change within the gene, which is now thought of as including both coding (expressed) and non-coding sequences. The extent of this mutational diversity will now be discussed in the context of the genomic architecture, with particular reference to human disease.

DNA SEQUENCE ORGANISATION AND ITS CHANGE BY MUTATION

The basic functional unit of genetic expression is an arrangement of coding (exon) and non-coding (intron or intervening) sequences with flanking regulatory sequences. Such genes are often arranged in clusters, several of which have been analysed in considerable depth. These vary in size from about 20 kb to 2000 kb in length, with the HLA or immunoglobulin heavy chain clusters encompassing several hundred genes at the upper end and those such as the α - and β -globin clusters encoding only four or five functional genes. Fig 1 summarises the main organisational features of a typical gene. The DNA helix has polarity and at one end of the chain has a phosphate group attached to the 5' position of a deoxyribose sugar while the other end has a 3' hydroxyl residue. The two strands are held together in opposite polarity by hydrogen bonding between complementary bases.

Starting from the 5' end of the gene shown schematically in fig 1, are certain sequences that tend to be conserved in evolution and which are necessary for efficient transcription of the messenger RNA (mRNA). In eukaryotes there are four or five known sequences contributing to promoter function, some or all of which are present in all functional genes. These are thought to be necessary for binding of RNA polymerase II, which synthesises mRNA in a 5'–3' direction. These sequences include the TATA box, CCAAT box, a GC rich region, and finally the transcription initiation (cap) site, which lies close to a sequence presumed to be necessary for binding of

Fig 1 Major structural and regulatory features of DNA sequences determining the expression of a human gene and its transcription into mRNA. Processing of primary mRNA transcript into mature mRNA is shown. Processing of DNA sequences during cell differentiation occurs only in certain genes. AAA, dispersal repetitive sequences.

Table 1 Genetic disorders in which specific mutational changes have been identified either by DNA sequence analysis or by hybridisation analysis using cloned gene probes

Genetic disorder	Mode of inheritance (chromosomal location)	Mutational change	References
α -thalassaemia	AR (16)	Deletions; BPS, FS in exons, splice junction, poly A addition site	60
β -thalassaemia	AR (11)	BPS, FS in exons (chain termination, cryptic splice sites), introns (new splice sites), splice junctions, regulatory sites (ATA, distal promoter), poly A addition site; deletions; insertions	60
Sickle cell anaemia	AR (11)	BPS in exon-I (codon 6)	
Hereditary persistence of fetal haemoglobin	AR (11)	Deletions	
α_1 -antitrypsin deficiency	AR (14)	BPS in exon-V (codon 342) (Z mutation)	25
Familial hypercholesterolaemia	AD (19)	Deletion (5 kb entailing exons at 3' end of LDL receptor)	85
Factor X deficiency	AR (13)	Deletion in some patients	109
Isolated growth hormone deficiency (type IA)	AR (17)	Homozygous deletion of at least 7.5 kb, includes hGH-N gene*	86
Antithrombin III deficiency	AD (1)	Deletion of AT III gene in one of two families	87
Osteogenesis imperfecta (type II)	AR (7)	Heterozygous 500bp deletion in pro α_1 (I) collagen gene*	88
Chorionic somatomammotropin deficiency	AR (17)	Homozygous deletion of at least 18.5 kb (hCS, hCS-2, and hGH-V all deleted)	89
Diabetes mellitus (MODY)	AD (11)	BPS in insulin gene codons 24, 25	104
Immunoglobulin K-chain deficiency	AR (2)	BPS in codons 148, 194 of IGK gene	105
Familial amyloidotic polyneuropathy	AD (18)	BPS in codon 30 of prealbumin gene	103
Congenital adrenal hyperplasia (21-hydroxylase deficiency)	AR (6)	Deletions in a specific cytochrome P-450; non-deletion forms	107
Adenosine deaminase deficiency	AR (20)	BPS	108
Lesch-Nyhan syndrome	XL	Deletions and major rearrangements in 20%, partial gene duplication; BPS in exon-2	81
Factor VIII deficiency	XL	80 kb deletion, BPS in exons	82
Factor IX deficiency	XL	Deletions (> 9 kb; > 18 kb)	83
Norrie's disease	XL	Deletions	101
Duchenne muscular dystrophy	XL	Deletions (250-600 kb)	53, 54
Glycerol kinase deficiency	XL	Deletions	102
Ornithine transcarbamylase deficiency	XL	Deletions (one of 15 patients)	84

AD, autosomal dominant; AR, autosomal recessive; XL, X-linked; BPS, nucleotide base pair substitution; FS, frame shift mutation.

DNA analysis in human disease

Table 2 Genetic disorders in which indirect diagnosis is possible using specific or non-specific cloned DNA probes

Indirect diagnosis using specific gene probes	Indirect diagnosis using non-specific gene probes
Haemoglobinopathies (11, 16)	Huntington's chorea (4)
Phenylketonuria	Myotonic dystrophy (19)
Osteogenesis imperfecta (types I, II, IV) (7)	Adult polycystic kidney disease (16)
α_1 -antitrypsin deficiency (14)	Cystic fibrosis (7)
Isolated growth hormone deficiency (type 1A) (17)	Retinoblastoma (13)
Antithrombin III deficiency (1)	Wilms' tumour (11)
Familial hypercholesterolaemia (type IIa) (19)	Aniridia (11)
Familial apolipoprotein CII deficiency (19)	Adrenoleukodystrophy (X)
Factor X deficiency (13)	Peroneal muscular atrophy (X)
Chorionic somatomammotropin deficiency (17)	Choroideremia (X)
Congenital adrenal hyperplasia (21-hydroxylase deficiency) (6)	Retinitis pigmentosa (X)
Adenosine deaminase deficiency (20)	Fragile X syndrome (X)
Diabetes mellitus (MODY) (11)	Duchenne muscular dystrophy (X)
Immunoglobulin K-chain deficiency (2)	Becker muscular dystrophy (X)
Amyloidotic polyneuropathy (18)	Norrie's disease (X)
Carbamoyl phosphate synthetase 1 deficiency (2)	Ocular albinism (X)
Haemophilia A (factor VIII deficiency) (X)	Haemophilia A (X)
Haemophilia B (factor IX deficiency) (X)	Retinoschisis (X)
Lesch-Nyhan syndrome (X)	Glycerol kinase deficiency (X)
Ornithine transcarbamylase deficiency (X)	

In the case of the diseases diagnosed using specific gene probes, genetic heterogeneity may limit their application to selected families. Genetic diseases that can be diagnosed using non-specific probes located within about 10 centiMorgans are listed. In many cases these methods are restricted to research rather than routine clinical application.

mRNA to the ribosomes. Mutations affecting regulatory sequences have been identified in patients with α positive or β positive thalassaemia, which are associated with reduced transcriptional activity (table 1). After the transcription initiation site there is a 5' untranslated region of variable length that ends in a conserved and presumably functionally important sequence immediately preceding the first codon (ATG), which codes for methionine. The importance of such transcribed but untranslated regions has been emphasised by the finding that such a region regulates the stability of mRNA transcripts coding for the c-myc oncogene.³³ Its deletion in some patients with Burkitt's lymphoma may therefore change the expression of this oncogene and contribute to the genesis of this tumour.

Further upstream (towards the 5' end) from promoter sequences or sometimes in intervening sequences or remote 3' sites there may be enhancer sequences important for increased or tissue specific

expression of genes. A role for enhancers in the genesis of certain tumours has been suggested by the translocation of c-myc oncogenes (on chromosome 8) to the vicinity of strong immunoglobulin enhancers (on chromosomes 14, 2, or 22) in some patients with Burkitt's lymphoma, although in most cases the break point lies proximal to the enhancer.³³ It has been suggested that a variable tandem repeat (VTR) sequence located 3' to the c-Ha-ras oncogene may also be an enhancer, perhaps explaining the presence of unusual VTR alleles in patients with various solid and leukaemic tumours.³⁴

Coding sequences are separated by intervening sequences of varying length, which are transcribed but not translated and are removed by being "spliced out" in the nucleus. Splicing of messenger RNA seems to depend on the precise recognition of specific sequences of nine to 12 bases at the 5' and 3' exon-intron junctions, which can be changed in diseases such as thalassaemia (table 1). Splicing is, in many cases, necessary for stabilisation of the mRNA and for export to the cytoplasm. Other sequences within the intron have also been shown to be necessary for recognition by the splicing enzymes. In procollagen genes there are about 50 introns, each of which requires correct splicing³⁵ so that mutations affecting this process are likely to be identified with the availability of cloned collagen sequences in heritable disorders affecting these molecules (table 1).

Most mutations affecting gene expression are base pair substitutions within the coding sequences (table 1). Our current knowledge of specific mutations in most diseases is heavily biased towards the more readily identified changes such as deletions and gross rearrangements. Moving towards the 3' end of the gene, the last exon is terminated by a stop codon (UAA, UAG, or UGA), which terminates translation, although transcription is continued for a variable distance towards the 3' end of the gene where it stops several hundred base pairs 3' to the polyadenylation signal sequence (AATAAA), which is the site of mRNA cleavage. Mutations resulting in loss of the polyA signal have been found in α -chain positive and β -chain positive thalassaemia, resulting in an abnormally long globin message (table 1). In addition to nuclear splicing, most eukaryotic mRNAs are modified before export through the nuclear envelope by capping the 5' end by addition and methylation of a Gppp residue to a pyrimidine and also by tailing the 3' end by addition of 50–150 adenylic acid residues. Differential processing of mRNA in the cytoplasm is another means of regulating gene expression and is therefore a potential site of mutational change (fig 1).

One final way in which DNA organisation is relevant to human disease is in the rearrangement or irreversible loss of large segments of DNA during the

course of cellular differentiation.³⁶ This phenomenon occurs, but may not be unique to, the immunoglobulin (Ig) and T cell receptor loci and serves to generate diversity and activate genes at appropriate developmental stages. This type of mechanism is already important for diagnostic pathology. The arrangement of the Ig heavy chain cluster in germ line—that is, undifferentiated DNA—is different from that in a cell that has progressed along the path of B cell differentiation as a result of local rearrangements of V, D, and J segment genes. Thus a histologically undifferentiated lymphoma cell can be classified as B or non-B lymphocyte in origin.³⁷ A second class of rearrangement occurs in B cell differentiation entailing the orderly switching of expressed heavy chain genes (class switch)³⁸ and seems to be initiated by deletion of several genes located 5' to the expressed heavy chain by a recombinational process. Both of these types of rearrangement may be important in the pathogenesis of certain types of lymphoid tumour, as they appear to be relatively prone to the occurrence of rearrangements, some of which can change the expression of cellular oncogenes. This type of DNA processing as a means of bringing about differential expression of a single gene in different tissues or at different stages of development may be relatively common. It has recently been pointed out that the c-abl oncogene codes for at least four different proteins by differential splicing of the first coding exon, a process that is inevitably complex and perhaps prone to the occurrence of genetic rearrangements and fusion products (cf, abl, trk mutations; table 2).³⁹

Direct and indirect diagnosis of genetic disease

The idea that a cloned gene can be used to make a direct diagnosis of genetic abnormality at the DNA level is a simple one that has proved remarkably elusive in practice despite the enormous growth in the number of cloned genes in recent years. In many cases it is simpler to make a direct diagnosis—that is, by demonstrating a specific abnormality in the patient—at the level of the gene product (mRNA, protein). Only recently has a general strategy been suggested that might make it possible to make a direct diagnosis at the DNA level in a variety of genetic disorders,⁴⁰ but this is far from being in routine use in diagnostic laboratories. Direct diagnosis at the level of the gene product is not, however, usually a feasible proposition in the case of prenatal diagnosis, where the ability to show abnormality in the DNA would circumvent several problems such as age and tissue specific expression. Despite this it is often only possible to make an indirect diagnosis at the DNA level using a cloned gene specific to the disease. Indirect

diagnosis here implies diagnosis by association within the context of a family study. The need to resort to indirect diagnosis is an irritating necessity in the context of diagnosis using specific cloned genes but, in the context of Mendelian disorders whose molecular nature is unknown, it is an extremely powerful approach that looks set to make a greater impact both on our understanding and on the diagnosis of genetic disease. The rationale for applying such an indirect diagnosis is best discussed in the context of the restriction fragment length polymorphism (RFLP)

RESTRICTION FRAGMENT LENGTH POLYMORPHISM

The idea of using RFLPs to map indirectly Mendelian disorders whose molecular basis is unknown was first suggested by Solomon and Bodmer⁴¹ and later discussed at length by Botstein *et al.*⁴² It exploits the presence of an unexpectedly high prevalence of DNA sequence variation throughout the genome, much of which is potentially detectable when it leads to the loss or gain of a restriction endonuclease cleavage site, with resultant variation in restriction fragment length. Fig 2 shows the detection of RFLPs. It has been estimated that on average 1 in 300 base pairs shows common or polymorphic variation in the general population.⁴³ By screening DNA transfers from a panel of normal subjects cut with various restriction enzymes, one or two enzymes are usually found that identify normal and variant fragments due to loss or gain of a cleavage site. These variants are commonly located some 2–20 kb outside the sequence that shows homology with the probe and are generally found in non-coding DNA sequences. As they do not produce a change in phenotype they are presumably unaffected by natural selection and therefore have accumulated by random mutation over enormous lengths of evolutionary time.

The requirements for indirectly identifying the location or presence of a gene using RFLPs in a family study can be summarised as follows:

- 1 A clinically and genetically homogeneous group of families should be available, in which members are unambiguously assigned to affected and unaffected groups. A limited amount of genetic heterogeneity between families can be tolerated in the analysis, but too much will obscure the result.
- 2 The probe or RFLP site must be situated within a genetic distance that allows one to distinguish linked from unlinked genes. In practice a genetic distance less than 30 centiMorgans (one centiMorgan (cM) is a unit of genetic distance corresponding to 1% recombination, roughly equivalent to a physical length of 10^6 nucleotide base pairs of DNA) is

Fig 2 Detection of restriction fragment length polymorphisms (RFLP) in genomic DNA by restriction endonuclease digestion, agarose gel electrophoresis, and Southern transfer methods. Hybridisation of a ^{32}p -labelled cloned DNA probe that is homologous with sequences on both the 5 kb and 15 kb genomic fragments shown, will result in detection of corresponding bands after autoradiography. RFLP commonly occurs where base change within endonuclease cleavage site leads to loss of that site and change in detected fragment length. Segregation of RFLP within family with a genetic disorder (●, ■) is shown, in which recombination has occurred between probe and disease locus.

required, which corresponds to about 30 million base pairs of DNA.

3 The affected or carrier parent in a family should be informative for linkage in that he or she has one detectable RFLP allele on one chromosomal homologue and another allele on the other homologue, so that each restriction fragment acts as a marker for the segregation of the adjacent chromosomal region in the offspring.

4 It must be known which RFLP allele is associated with the normal or disease gene on each chromosomal homologue—that is, the linkage haplotype or phase. This is most easily identified by looking at the RFLP state of the grandparental generation.

5 The RFLP alleles are then typed in all offspring whose genetic state can be inferred from their clinical state. The association of RFLP and disease phenotypes in the children is either of the parental type (non-recombinant) or shows new genetic combinations (recombinant).

6 The proportion of recombinant to total offspring (recombination fraction) provides an estimate of the

genetic distance between marker and disease loci. The genetic distance measured in terms of recombination shows a high correlation with physical separation up to about 30% recombination, but thereafter the correlation rapidly decreases towards zero.

Linkage is analysed statistically by calculating the probability of observing the given allelic combinations, pooled over all the families, given the assumption of any specific value of the true recombination fraction (θ).⁴⁴ As the probabilities for all values of θ are not normalised they are referred to as likelihoods ($L(\theta)$). The likelihood of observing the family data is then calculated for the null hypothesis, where the two loci are not physically linked so that they assort independently ($L(\theta) = 0.5$). Finally, the likelihood ratio is obtained:

$$L(\theta) / L(\theta = 0.5)$$

and the log to the base 10 of this value is called the lod score. The conventional summarising statistic is the maximum likelihood value of θ , which reaches formal significance when the corresponding lod score exceeds

3—that is, 1000:1 in favour of that value of θ over non-linkage. Computer programs are available that estimate these likelihoods both for a single pair of loci⁴⁵ and for multipoint data using several RFLP markers.^{46 47}

INDIRECT DIAGNOSIS OF GENETIC DISEASE

In 1980 Botstein *et al*⁴² suggested that it would be possible to identify indirectly the location of any disease gene by linkage analysis using RFLPs, without having any knowledge of the specific gene responsible, or of its products. It was suggested that a limited number (150–200) of RFLP markers spaced at about 20cM intervals would be sufficient to do this as the total haploid genetic length is about 3300 cM. The average size of a human chromosome is about 150 cM so that five to 10 suitably spaced markers per chromosome should be sufficient. Table 2 gives examples of Mendelian disorders whose molecular basis may or may not be known, but which have been mapped indirectly using recombinant DNA probes and RFLPs. The number of families required to establish linkage using random probes varies with the genetic distances entailed, the mode of inheritance, and the structure of the families.⁴⁸ Almost invariably the required number of families is larger than that predicted from theoretical considerations because of uncertainty about the linkage phase, more so in autosomal recessive disease. Highly polymorphic multiple allele markers are considerably better than the conventional two allele RFLP systems, as correspondingly more families are informative for linkage.⁴⁹ Large families are preferable for establishing a new linked marker as they tend to give more information on the linkage phase and because genetic heterogeneity becomes less of a problem. In practice mapping using RFLPs is usually complemented by the identification of patients with chromosomal rearrangements or deletions that further localise the gene.^{50–54} The identification of a gene before rather than after the identification of its product (“reverse genetics”) has been achieved with several oncogenes and may soon become a reality with genes for inherited disease. In Duchenne muscular dystrophy (DMD) the gene has been localised to within 2 cM by RFLP mapping,⁵⁵ but to within about 600 kb by the identification of small deletions and an X to autosomal translocation breakpoint situated close to the DMD locus.^{53 54}

The application of such RFLP mapping techniques to routine clinical diagnosis is likely to be a process of slow evolution rather than immediate transition. The initial goal is to obtain RFLPs within 10 cM on each side of the gene (bridging markers), thus reducing the predictive error to about 1%. This amount of error is probably acceptable in the diagnosis of serious gen-

etic disorders and is likely to lead to wide application, provided that genetic heterogeneity has been shown to be minimal. The ability to isolate DNA from the fetal chorionic trophoblast during the first trimester of pregnancy potentially provides a real advance over second trimester diagnosis when coupled with such DNA probing techniques, although the safety of chorion biopsies has yet to be fully shown.⁵⁶

DIRECT DIAGNOSIS OF GENETIC DISEASE

There are several problems associated with the use of specific gene probes in the diagnosis of genetic disorder. These problems arise in two different contexts: first, where gene structure or expression is changed in a disease and there is little doubt that the cloned gene is the site of primary abnormality; secondly, where the available understanding of the disease suggests that a particular cloned gene is a good candidate for the site of primary abnormality—but this has been hard to confirm at the level of gene expression for any of a variety of reasons. In this case the problem is to screen a “candidate” gene for its likely role in the disease process, without initially resorting to a comparison of normal and presumed abnormal DNA sequences. In the former case the cloning and sequencing of the genes are usually necessary at some stage.

In screening a “candidate” gene, a comparatively quick answer can be obtained by means of a family linkage study, assuming that it is possible to identify flanking RFLPs. If the “candidate” gene or any gene within about 1000 kb of it (which therefore includes other members of a gene cluster), is the site of primary abnormality, there should be little or no recombination between this site and the disease locus during meiosis (assuming a limited sample size). The principle is the same as with indirect mapping using random probes, except that one is comparing the relatively easily distinguished hypotheses of no recombination v 50% recombination (random assortment). A comparatively small set of well characterised families is required for this to be done, which will tend either to confirm or exclude the “candidate” gene. For example, the role of the various cloned collagen genes in complex disorders such as Marfan’s syndrome and osteogenesis imperfecta is beginning to be elucidated by this approach.^{57 58} Heritable abnormalities in the c-Ha-ras and c-Ki-ras oncogenes have been excluded as the primary cause of Gardner’s syndrome by the same method.⁵⁹ Linkage studies are unlikely, however, to distinguish between genes that are separate but very closely linked (up to 1000 kb apart), unless a very large number of informative families are available. The finding of no recombination between the cloned gene and the disease gene cannot therefore justify the conclusion that

this gene is the site of the primary mutation, but it does suggest that the responsible gene is within that region and that the cloned probe could provide a suitable predictive test for genetic counselling purposes. This type of study might focus further attention on the expression of mRNA or protein products and serve to initiate attempts to clone and sequence the gene from diseased and normal subjects.

In the case of genetic disorders where an abnormal DNA sequence or product has established that the cloned gene is the likely site of primary abnormality there are still considerable problems about making a diagnosis at the DNA level in a routine clinical context. The simplest approach is the indirect one, which utilises flanking RFLPs and provides a diagnosis in the context of a family study. This is still the only available means of prenatal diagnosis in many types of thalassaemia^{60 61} despite a considerable knowledge of the molecular nature of the disease. A direct diagnosis has been achieved in sickle cell disorder, where the Hb^S mutation is associated with the loss of a Mst II restriction site in codon 6 of the β globin gene, leading to a detectable change in restriction fragment pattern, although this is not completely specific to the Hb^S mutation.⁶² A similar type of direct diagnosis can be made in some types of thalassaemia^{60 61} and in genetic disorders where there is deletion of part or all of the gene (table 1), leading to a changed restriction fragment pattern. More recently, an understanding of the molecular nature of particular mutations has made it possible to synthesise small oligonucleotide probes, specific either to the normal or abnormal gene sequence, which can distinguish between these two alternatives under appropriate hybridisation conditions. This method has been applied to the direct diagnosis of the Pi^z allele responsible for α_1 -antitrypsin deficiency²⁵ and to certain haemoglobinopathies.^{26 60 61} A third approach to the direct diagnosis of molecular disease has not yet been fully exploited but arises from the ability to detect small mismatches in RNA-DNA hybrids using high specific activity RNA probes specific for the normal DNA sequence and ribonuclease A which will recognise and cleave mismatched hybrids.⁴⁰ Mismatched hybrids will occur when there is a base substitution in the DNA of the affected patient and can lead to the identification of abnormal RNase cleavage products following electrophoresis. If this technique is found to be generally applicable it would provide an important advance in the use of DNA in the diagnosis both of genetic disorders and of somatic mutation in oncogenesis. In general, a group of unrelated patients with a particular genetic disorder are likely to have mutations affecting any of several different sites within the gene, each of which produces abnormal function. This will be the

general expectation unless there is either a very low mutation rate within that gene or there has been strong selection in favour of a particular mutation in the past, so that it is found in a sizeable proportion of disease bearing patients. Huntington's chorea may be an example of the case where the mutation rate is low, with the result that most of the cases in north America are probably descended from a few original mutations introduced in the seventeenth century.⁶³ Sickle cell anaemia and some of the thalassaemia mutations may be examples of the second mechanism, in which natural selection has amplified a few ancient mutations in certain populations.⁶¹ Most genetic disorders may not come into either category, so that it will not be feasible either to identify specific restriction fragment variations or develop the appropriate oligomers for each mutational type. A general method for detecting mismatches within a relatively large cloned sequence⁴⁰ will perhaps help to fill this diagnostic gap, although it is still likely to be applied in the context of a family study wherever possible.

DNA probes in oncogenesis

Although the use of DNA probes in the diagnosis of microbiological disease is already of considerable importance,⁶⁴ besides the area of genetic pathology, the biggest impact of recombinant DNA technology in medicine is likely to be in the study of oncogenesis. Several excellent reviews of this area have been published,^{31 65-68} but it is worth emphasising some specific areas in which DNA analysis is increasingly relevant to diagnosis.

It is now beyond reasonable doubt that a variety of specific somatic mutational changes accompany the progression of individual tumours, although it is hard to analyse the clinical importance of any one of these outside the context of a complex and evolving cellular phenotype. The analysis of tumour pathology has gradually progressed from the cellular to the chromosomal and now to the DNA level, where the identification of somatic genetic changes are beginning to elucidate phenotypic correlates of malignancy at other levels of analysis. The implications for typing the nature and progression of individual tumours are clear. There are three areas for which DNA analysis is particularly relevant to the study of oncogenesis—namely, inherited predisposition to cancer; DNA changes associated with the initiation of tumour growth; and DNA changes associated with tumour progression.

INHERITED PREDISPOSITION

Inherited preredisposition to cancer is well established in a small number of Mendelian disorders, some of which have been mapped to chromosomal

regions as a result of visible chromosomal abnormalities. For example, a small proportion of patients with retinoblastoma and Wilms' tumour (nephroblastoma) have deletions in chromosomes 13 and 11, respectively.^{69 70} In some of these disorders there seems to be an interaction between the inherited mutation that predisposes to a particular type of cancer and secondary somatic mutational events. The analysis of retinoblastoma (Rb) tumours suggests that a secondary mutation is required for tumour development.⁶⁹ This results in loss of the normal allele(s) leading to hemizyosity or homozygosity of the Rb allele. The secondary mutation seems to occur either by mitotic recombination, non-dysjunction, or gene conversion. Similar mechanisms have been postulated for Wilms' tumour.⁷⁰ For this reason comparisons between germline DNA and tumour DNA are potentially useful in elucidating disease mechanisms. Only a few of these inherited cancers are associated with visible deletion, so that others may entail point mutations or smaller deletions at the same loci.

Inherited predisposition to neoplasia is present in another group of Mendelian disorders associated with chromosomal instability.⁷¹ Conditions such as ataxia telangiectasia and xeroderma pigmentosum show a greatly increased risk of certain types of neoplasia. Specific defects in DNA repair have been identified both in AT and in XP, so that it seems that there is a consequent tendency to somatic mutation in proliferating tissues leading to an increased risk of neoplasia. There is a concomitant tendency to degenerative change in end differentiated tissues such as neurons. The search for specific DNA repair defects in other disorders associated with increased spontaneous or inducible chromosomal aberrations and with increased neoplasia has so far been negative. Other types of neoplasia, such as carcinoma of the breast, may entail a heritable susceptibility that is polygenic in nature. As with other disorders showing complex inheritance, the practical difficulties of using DNA analysis to locate such genes are considerable.

Cellular oncogenes are obvious candidate genes for some of the inherited types of cancer, so that conventional linkage or RFLP studies using cloned oncogene sequences provide another approach to these disorders.⁵⁹ It may also be possible to detect subtle differences within or close to certain proto-oncogenes, such as c-Ha-ras I, which influence susceptibility to some tumours, perhaps in concert with other genetic and environmental influences.³⁴

DNA CHANGES ASSOCIATED WITH THE INITIATION OF TUMOUR GROWTH

The most important factors concerned with the initiation and promotion of tumour growth that have emerged in recent years concern the role of viral and

cellular oncogenes. This subject has been extensively reviewed⁶⁵⁻⁶⁸ and is one of the most active research areas in biological science. It is now clear that the functions of some 20-30 cellular or proto-oncogenes (c-onc genes) are capable of being distorted: firstly, by incorporation into exogenous retroviruses as viral or v-onc genes and subsequent reinsertion into the host genome; secondly, by changed regulation following the insertion of retrovirus into adjacent sequences; thirdly, by somatic mutation within the c-onc itself. These proto-oncogenes are thought to have a normal role in such diverse activities as the control of differentiation, cell division, growth factor production and response.⁶⁸ Mutational disturbance of c-onc function is thought to be associated with the initiation or progression of the sequence of genetic and epigenetic changes associated with neoplasia.

Retroviral oncogenes generally show mutational differences from their proto-oncogene equivalents at several sites, only some of which may be crucial to neoplastic transformation. The retrovirus has a single stranded RNA genome, which can infect the host and become copied into DNA by the action of reverse transcriptase. It may then either remain as an integrated provirus or be transcribed into RNA allowing expression of its own gene products, which permit the repackaging and release of infective virus. Mutational damage can occur as a consequence of insertion into the host genome.⁶⁵ A surprisingly high proportion (0.5%) of the DNA of some mammalian species contains proviral sequences,⁷² which seem to be contained by cellular mechanisms that prevent expression or transmission between cells.

Most proto-oncogenes have been discovered through their homology with retroviral v-onc derivatives, so that our current knowledge of proto-oncogenes is biased by this perspective. A dozen or so c-onc genes that have no retroviral equivalents have now been identified from tumour cells by their ability to transform cells in culture. While it is clear that only a minority of retroviruses carry oncogenes, those that do can subvert the function of their cellular c-onc equivalents and give rise to neoplasia. Retroviruses that do not contain v-oncs are also potentially oncogenic, although less dramatically so. They seem to act by random insertion into the host genome, occasionally disrupting the regulation of cellular proto-oncogenes and initiating tumour development.⁶⁵ The best example in man is the group of human T cell leukaemia viruses (HTLV I-III).⁷⁴ HTLV I expression is strongly associated with adult T cell leukaemias and lymphomas that show a tendency both to geographical and to familial clustering. Some insight into the mechanism of action of HTLV I is beginning to emerge with the identification of a sequence that is capable of activating the expression

of remote but as yet unidentified host genes, possibly including certain growth factor receptors, which are expressed at a high level in these tumours.⁷³ Although the role of retroviruses in oncogenesis has attracted most attention to date, their association with neurological, cardiovascular, and other disorders has been observed in animals and man,⁷⁴ which will undoubtedly be elucidated by the use of cloned sequences to probe diseased tissues for expression or integration of viral sequences.

DNA CHANGES ASSOCIATED WITH TUMOUR PROGRESSION

It seems likely that only a few human cancers are pre-

cipitated by retroviral insertion, but this work has highlighted the changes occurring in their normal cellular counterparts during tumour progression. Two major types of change have been found to occur in proto-oncogenes in neoplastic tissue—namely, increased expression of a normal sequence and normal expression of a mutant sequence.

Increased expression of certain gene products has been known to occur in viral oncogenesis for a long time, but amplification or overexpression of some cellular oncogenes has now been shown in a wide variety of human tumours and tumour cell lines (table 3). The most striking example in man is the amplification of apparently normal N-myc genes in stage III and IV

Table 3 Human tumours associated with mutations in oncogene sequences*

Somatic disorder	Genetic locus (chromosome)	Mutational change	References
Bladder carcinoma (CL)	c-Ha-ras (11)	BPS in exon-1 (codon 12)	90, 67
Malignant melanoma (CL)			
Lung carcinoma (CL)	c-Ki-ras (6)	BPS in exon-2 (codon 61)	91, 67
Lung carcinoma (CL, T)		BPS in exon-1 (codon 12)	
Bladder carcinoma (CL)		BPS in exon-2 (codon 61)	
Neuroendocrine tumour (CL)			
Colon carcinoma (CL, T)			
Carcinoma of lung, bladder, colon, rectum (T)	N-ras (1)	Amplification	92, 67
Lung carcinoma (CL)		BPS in exon-1 (codons 13, 12)	
Acute myeloid leukaemia (T)		BPS in exon-2 (codon 61)	
Neuroblastoma (CL)		Amplification	
Teratocarcinoma (CL)			
Fibrosarcoma (CL)			
Mammary carcinoma (CL)			
Malignant melanoma (CL)	c-myc (8)	Deletion of 5' regulatory sequences associated with t(8;14) translocation: amplification	78, 67
Burkitt's lymphoma (CL)			
Colon carcinoma (CL)			
Small cell lung carcinoma (CL)			
Promyelocytic leukaemia (CL, T)	N-myc (2)	Amplification	94, 67, 97
Neuroblastoma (CL, T)			
Retinoblastoma (CL, T)			
Teratocarcinoma (CL)			
Small cell lung carcinoma (CL, T)	L-myc	Amplification	95
Small cell lung carcinoma (CL, T)	c-abl (9)	Fused bcr-abl loci associated with t(9;22) translocation and Ph ¹ chromosome: amplification	96, 67, 93
Chronic myeloid leukaemia (CL)			
Erythroleukaemia (CL)			
Chronic myeloid leukaemia (CL)	c-erb B-1 (7)	Truncated amplified EGF receptor in association with t(7;11) translocation	98
Epidermoid carcinoma (CL)			
Gliomas (T)			
Salivary and mammary carcinomas (T)	c-erb B-2 (17)	Truncated mRNA product and amplification	99
Gastric carcinoma (CL)	trk	Truncation and fusion of tropomyosin and a tyrosine-specific protein kinase	39
Colon carcinoma (T)			
Acute myeloid leukaemia (T)	c-myb (6)	Amplification and gene rearrangements, deletions	93, 67
Colon carcinoma (CL)			
T cell leukaemia			
Acute lymphocytic leukaemia			
Chronic lymphocytic leukaemia (CL)	bcl (11)	Gene rearrangements, associated with t(11;14) translocation	100
Large cell lymphoma (CL)			

*It is possible that some of the changes shown in cell lines (CL) derived from tumours have arisen in vitro, although similar changes have generally been found in samples of tumour (T) DNA.

neuroblastomas, although this has also been observed with *ras*, *abl*, *mos* and *erb-b* genes.⁷⁵ In transfection assays it has been shown that increased expression of normal *ras* oncogenes, as well as mutant expression of such genes, can transform mouse cells although this may only reflect the sensitivity of mouse 3T3 cells to transformation.⁷⁶ The relative importance of increased expression *v* mutant expression of oncogenes in human tumours is still unclear as it is hard to exclude the presence of subtle mutational changes. At the molecular level increased expression and changed expression may have similar consequences with regard to loss of normal regulatory controls affecting functions such as phosphorylation and responsiveness to growth factors.

Increased expression of proto-oncogenes may also be associated with genetic rearrangement at the chromosomal level. The *c-myc* gene seems to be activated as a consequence of translocation from its chromosomal location on 8q24 to sites either on chromosome 14 or 22 where immunoglobulin heavy or light chain sequences are being actively transcribed.⁷⁷ The increased proximity of immunoglobulin enhancers may be sufficient to activate the translocated *ras* gene in some as yet uncertain manner. Somatic mutation may also be entailed,⁷⁸ perhaps induced by similar mechanisms to the immunoglobulin variable region genes, which normally occupy the same sites. It is still not clear to what extent events of this sort, which result in changed expression of proto-oncogenes, occur early in oncogenesis and are therefore primary, rather than late as a result of selection for increased growth or metastasis.

Proto-oncogenes can be activated by mutation without greatly increased expression, at least initially. The activated *ras* oncogene product, p21, shows evidence of mutation at only three specific sites in all instances studied to date (table 3), implying that a relatively specific functional change is necessary to promote tumour growth. Similar mutations may occur in other proto-oncogenes, although directly identifying a mutation in any one of many nucleotide sites in the gene, as with the direct identification of inherited genetic disease, is difficult.

Evidence that neither mutation nor changed expression of a single proto-oncogene is normally sufficient to initiate neoplasia *de novo* was obtained by Weinberg *et al.*⁷⁹ They cultured untransformed cells directly from animals, rather than studying the transformation of cells that had already been in culture for long periods, and showed that more than one oncogene was required—for example, *ras* and *myc*—to transform cells. The multistep nature of carcinogenesis has been known for a long time, but only now is its molecular basis beginning to emerge.^{66 68} Most proto-oncogenes identified so far have functions

closely related to growth control, suggesting that there may be only a limited number of such genes (less than 50) with major effects on oncogenesis. There will almost certainly be many more with minor effects on neoplastic progression. This might lead to a useful distinction between major and minor oncogenes on some empirical basis, such as their ability to transform NIH-3T3 cells.

A progression of genetic and epigenetic changes, combined with hormonal or other environmental factors, are likely to be necessary for the gradual transition from the preneoplastic state to the undifferentiated invasive neoplasm with metastatic potential. At clinical presentation about 15% of all human tumours contain transformationally active *ras* oncogenes.³¹ Other transforming genes that lack *v-onc* counterparts have also been found in such tumours. Chromosomal rearrangements such as amplification and translocation often occur, some of which seem to confer a growth advantage (table 3). The sometimes rapid progression of tumours *in vivo*, coupled with the *in vitro* finding of an increased rate of mutational change with increased metastatic potential, suggests the further possibility that the packaging or repair of DNA in chromatin may be changed so as to make it more sensitive to mutation in tumour cells.⁸⁰

Already certain tumours can be typed and staged by means of DNA analysis. Lymphomas can be classified as T or B cell in origin according to whether they show germ line or differentiated arrangements of immunoglobulin genes.³⁷ Neuroblastomas can be classified using DNA-RNA hybridisation according to the level of *N-myc* expression, which correlates with increasing malignancy,⁷⁵ as does *ras* expression in other types of tumours.³¹ As our knowledge of the functions of proto-oncogenes increases new opportunities for treatment may emerge so that the analysis of major and minor oncogene structure and expression in tumours and preneoplastic tissue looks set to become an integral part of diagnosis and treatment.

Future prospects

It is salutary to try and envisage the extent to which DNA analysis will become an established part of clinical investigation in the future. Future applications in the diagnosis of Mendelian disease and the typing and staging of neoplasms have been discussed at length. Less obvious areas of impact include those conditions showing complex inheritance—heart disease, mental disorder, congenital abnormalities and disorders of aging, in which the increased ability to distinguish single gene effects on a multifactorial background will be exploited. Our present restricted view of the scope of mutation in disease (tables 1 and 3), determined

largely by our limited ability to detect all but the grosser deletions and rearrangements, will certainly change to show enormously greater complexity and variety.

Will these advances bring about an important improvement in health care? The cynical retort is to say only at a considerable cost, because as they are currently used the techniques are undoubtedly expensive. Weatherall⁶¹ has argued on the basis of the experience with haemoglobinopathies that such measures can be very effective in financial and health terms. Although this conclusion may not apply to individually rare disorders, already relatively common conditions such as Duchenne muscular dystrophy, cystic fibrosis, and Huntington's chorea have become amenable to preventive measures using DNA analysis, so that together they constitute a group of chronically debilitating conditions that are extremely costly on any terms. In the diagnosis of tumour pathology these techniques are likely to determine treatment options at which point their benefits will become self evident. In view of the high stakes it is important that the process of "selling" these methods both to the medical profession and to the general public should be characterised by a concern for patient welfare as these advances will inevitably give rise to some awesome ethical dilemmas.

References

- 1 Southern EM. *Eukaryotic DNA. Biochemistry of nucleic acids*. MTP International Review of Science, Vol 6. London: Butterworths, 101-40.
- 2 Williamson R. *Genetic engineering*. London: Academic Press, 1982.
- 3 Old RW, Primrose SB. *Principles of gene manipulation*. 3rd ed. Oxford: Blackwell Scientific Publications, 1985.
- 4 Wallace RB, Johnson MJ, Hirose T, Miyake T, Kawashima EH, Itakura K. The use of synthetic oligonucleotides as hybridisation probes. II Hybridisation of oligonucleotides of mixed sequence to rabbit, β -globin DNA. *Nucleic Acids Res* 1981;9:879-94.
- 5 Choo KH, Gould KG, Rhys EJJ, Brownlee GG. Molecular cloning of the gene for human anti-haemophilic factor IX. *Nature* 1982;299:178-80.
- 6 Michaelson AM, Markham AF, Orkin SH. Isolation and DNA sequence of a full-length cDNA clone for human X chromosome-encoded phosphoglycerate kinase. *Proc Natl Acad Sci USA* 1983;80:472-6.
- 7 Young RA, Davis RW. Efficient isolation of genes by using antibody probes. *Proc Natl Acad Sci USA* 1983;80:1194-8.
- 8 Gitschier J, Wood WI, Goralka TM, et al. Characterisation of the human factor VIII gene. *Nature* 1984;312:326-30.
- 9 Kunkel LM, Tantravahi U, Eisenhard M, Latt SA. Regional localisation on the human X of DNA segments cloned from flow sorted chromosomes. *Nucleic Acids Res* 1982;10:1557-78.
- 10 Wieacker P, Davies K, Cooke HJ, et al. Towards a linkage map of the human X chromosome: regional assignment of 16 cloned single copy DNA sequences employing a panel of somatic cell hybrids. *Am J Hum Genet* 1984;36:265-76.
- 11 Mulligan RC, Berg P. Selection for animal cells that express the *Escherichia coli* gene coding for xanthine-guanine phosphoribosyltransferase. *Proc Natl Acad Sci USA* 1981;78:2072-6.
- 12 Willecke K, Ruddle FH. Transfer of the gene for hypoxanthine-guanine phosphoribosyltransferase via isolated human metaphase chromosomes into mouse L-cells. *Proc Natl Acad Sci USA* 1975;72:1792-6.
- 13 Weinstock GM, Rhys C, Berman ML, et al. Open reading frame expression vectors; a general method for antigen production in *Escherichia coli* using small fusions to β -galactosidase. *Proc Natl Acad Sci USA* 1983;80:4432-6.
- 14 Carle GF, Olson MV. Separation of chromosomal DNA molecules from yeast by orthogonal-field alternation gel electrophoresis. *Nucleic Acids Res* 1984;12:5647-64.
- 15 Vandenplas S, Weid I, Grobler-Rabie A, et al. Blot hybridisation analysis of genomic DNA. *J Med Genet* 1984;21:164-72.
- 16 Roberts RJ. Restriction and modification enzymes and their recognition sequences. *Nucleic Acids Res* 1984;12(supplement):r167-r204.
- 17 Sealey PG, Whittaker PA, Southern EM. Removal of repeated sequences from hybridisation probes. *Nucleic Acids Res* 1985;13:1905-22.
- 18 Wallace RB, Shaffer J, Murphy RF, Bonner J, Hirose T, Itakura K. Hybridisation of synthetic oligodeoxyribonucleotides to ϕ X174 DNA: the effect of single base pair mismatch. *Nucleic Acids Res* 1979;6:3543-57.
- 19 Marmur J, Doty P. Thermal renaturation of deoxyribonucleic acids. *J Mol Biol* 1961;3:585-94.
- 20 Schmid CW, Deininger PL. Sequence organisation of human genome. *Cell* 1975;6:345-58.
- 21 Wetmur JG, Davidson N. Kinetics of renaturation of DNA. *J Mol Biol* 1968;31:349-70.
- 22 Beltz GA, Jacobs KA, Eickbush TH, Cherbas PT, Kafatos FC. Isolation of multigene families and determination of homologies by filter hybridisation methods. *Methods Enzymol* 1983;100:266-85.
- 23 Bonner TI, Brenner DJ, Neufeld BR, Britten RJ. Reduction in the rate of DNA reassociation by sequence divergence. *J Mol Biol* 1973;81:123-35.
- 24 Kidd VJ, Golbus MS, Wallace RB, Itakura K, Woo SLC. α_1 -antitrypsin deficiency detection by direct analysis of the mutation in the gene. *N Engl J Med* 1984;310:639-42.
- 25 Thein SL, Wainscoat JS, Old JM, et al. Feasibility of prenatal diagnosis of β -thalassaemia with synthetic DNA probes in two Mediterranean populations. *Lancet* 1985;ii:345-7.
- 26 Evans HJ. Structure and organisation of the human genome. *Mutations in man*. Obe G, ed. Berlin: Springer-Verlag, 1984:58-100.
- 27 Bishop JO. The gene numbers game. *Cell* 1974;2:81-6.
- 28 Bodmer WF. Gene clusters, gene organisation and complex phenotypes. *Am J Hum Genet* 1981;33:664-82.
- 29 McKusick VA. *Mendelian inheritance in man*. 6th ed. Baltimore: John Hopkins University Press, 1983.
- 30 Burnet M. *Intrinsic mutagenesis: a genetic approach to ageing*. Lancaster: MTP, 1974.
- 31 Weiss RA, Marshall CJ. Oncogenes. *Lancet* 1984;ii:1138-42.
- 32 Vogel F, Rathenberg R. Spontaneous mutation in man. *Adv Hum Genet* 1975;5:223-318.
- 33 Cole MD. Regulation and activation of c-myc. *Nature* 1985;318:510-1.
- 34 Krontiris TG, Di Martino MA, Colb M, Parkinson DR. Unique allelic restriction fragments of the human Ha-ras locus in leukocyte and tumour DNAs of cancer patients. *Nature* 1985;313:369-73.
- 35 Prockop DJ, Kivirikko KI. Heritable diseases of collagens. *N Engl J Med* 1984;311:376-86.
- 36 Ellison JW, Hood LE. Human antibody genes. *Adv Hum Genet* 1983;13:113-47.
- 37 Arnold A, Cossman J, Bakhshi A, Jaffe ES, Waldmann TA, Korsmeyer SJ. Immunoglobulin-gene rearrangements as

- unique clonal markers in human lymphoid neoplasms. *N Engl J Med* 1983;309:1593-9.
- 38 Shimizu A, Honjo T. Immunoglobulin class switching. *Cell* 1984;36:801-3.
 - 39 Martin-Zanca D, Hughes SH, Barbacid M. A human oncogene formed by the fusion of truncated tropomyosin and protein tyrosine kinase sequences. *Nature* 1986;319:743-8.
 - 40 Myers RM, Larin Z, Maniatis T. Detection of single base substitutions by rinonuclease cleavage at mismatches in RNA:DNA duplexes. *Science* 1985;230:1242-8.
 - 41 Solomon E, Bodmer WF. Evolution of the sickle variant gene. *Lancet* 1979;i:923.
 - 42 Botstein D, White RL, Skonick M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 1980;32:314-31.
 - 43 Cooper DN, Schmidtke J. DNA restriction fragment length polymorphisms and heterozygosity in the human genome. *Hum Genet* 1984;66:1-16.
 - 44 Conneally PM, Rivas ML. Linkage analysis in man. *Adv Hum Genet* 1980;10:209-66.
 - 45 Ott J. Estimation of the recombination fraction in human pedigrees; efficient computation of the likelihood for human linkage studies. *Am J Hum Genet* 1974;26:588-97.
 - 46 Lathrop GM, Lalouel JM, Julier C, Ott J. Strategies for multi-locus linkage analysis in humans. *Proc Natl Acad Sci USA* 1984;81:3443-6.
 - 47 Clayton JF. A multipoint linkage analysis program for X-linked disorders, with the example of Duchenne muscular dystrophy and seven DNA probes. *Hum Genet* 1986;73:68-72.
 - 48 Thompson EA, Kravitz K, Hill J, Skolnick MH. Linkage and the power of a pedigree structure. *Genetic epidemiology*. Morton, NE, Chung CS, eds. New York: Academic Press, 1978:247-53.
 - 49 Reeders ST, Breuning M, Davies KE, et al. A highly polymorphic marker linked to adult polycystic kidney disease on chromosome 16. *Nature* 1985;317:542-4.
 - 50 Verellen-Dumulin Ch, Freunde M, De Meyer R, et al. Expression of an X-linked muscular dystrophy in a female due to translocation involving Xp21 and non-random inactivation of the normal X chromosome. *Hum Genet* 1984;67:115-9.
 - 51 Francke U, Ochs HD, De Martinville B, et al. Minor Xp21 chromosome deletion in a male associated with expression of Duchenne muscular dystrophy, chronic granulomatous disease, retinitis pigmentosa, and McLeod syndrome. *Am J Hum Genet* 1985;37:250-67.
 - 52 Van Heyningen V, Boyd PA, Seawright A, et al. Molecular analysis of chromosome 11 deletions in aniridia/Wilms' tumour syndrome. *Proc Natl Acad Sci USA* 1985;82:8592-6.
 - 53 Monaco AP, Bertelson CJ, Middlesworth W, et al. Detection of deletions spanning the Duchenne muscular dystrophy locus using a tightly linked DNA segment. *Nature* 1985;316:842-5.
 - 54 Ray PN, Belfall B, Duff C, et al. Cloning of the breakpoint of an X;21 translocation associated with Duchenne muscular dystrophy. *Nature* 1985;318:672-5.
 - 55 Bakker E, Hofker MH, Goor N, et al. Prenatal diagnosis and carrier detection of Duchenne muscular dystrophy with closely linked restriction fragment length polymorphisms. *Lancet* 1985;i:655-8.
 - 56 Rodeck CH, Morsman JM. First trimester chorion biopsy. *Br Med Bull* 1983;39:338-42.
 - 57 Børresen AL, Banforth S, Tsiouras P, Berg K. Studies of restriction fragment length polymorphisms at the pro- α_2 (I) collagen locus excludes close linkage to Marfan's syndrome. *Cytogenet Cell Genet* 1985;40:585.
 - 58 Tsiouras P, Børresen AL, Dickson LA, Berg K, Prockop DJ, Ramirez F. Molecular heterogeneity in the mild autosomal dominant forms of osteogenesis imperfecta. *Am J Hum Genet* 1984;36:1172-9.
 - 59 Barker D, McCoy M, Weinberg R, et al. A test of the role of two oncogenes in inherited predisposition to colon cancer. *Mol Biol Med* 1983;1:199-206.
 - 60 Collins FS, Weissman SM. The molecular genetics of human hemoglobin. *Prog Nucleic Acid Res Mol Biol* 1984;31:317-462.
 - 61 Weatherall DJ. *The new genetics and clinical practice*. 2nd ed. Oxford: Oxford University Press, 1985.
 - 62 Orkin SH, Little PFR, Kazazian HH, Jr, Boehm CD. Improved detection of the sickle mutation by DNA analysis. *N Engl J Med* 1982;307:32-6.
 - 63 Hayden MR. *Huntington's chorea*. Berlin: Springer-Verlag, 1981.
 - 64 Engleberg NC, Eisenstein BI. The impact of new cloning techniques on the diagnosis and treatment of infectious diseases. *N Engl J Med* 1984;311:892-901.
 - 65 Bishop JM. Cellular oncogenes and retroviruses. *Ann Rev Biochem* 1983;52:301-54.
 - 66 Land H, Parada LF, Weinberg RA. Cellular oncogenes and multistep carcinogenesis. *Science* 1983;222:771-8.
 - 67 Varmus HE. The molecular genetics of cellular oncogenes. *Ann Rev Genet* 1984;18:553-612.
 - 68 Weinberg RA. The action of oncogenes in the cytoplasm and nucleus. *Science* 1985;230:770-6.
 - 69 Cavenee WK, Dryja TP, Phillips RA, et al. Expression of recessive alleles by chromosomal mechanisms in retinoblastomas. *Nature* 1983;305:779-84.
 - 70 Solomon E. Recessive mutation in aetiology of Wilms' tumour. *Nature* 1984;309:111-2.
 - 71 Friedberg EC, Ehmann UK, Williams JI. Human diseases associated with defective DNA repair. *Advances in Radiation Biology* 1979;8:85-174.
 - 72 Kozak C, Silver J. The transmission and activation of endogenous mouse retroviral genomes. *Trends in Genetics* 1985;1:331-4.
 - 73 Weiss RA, Teich N, Varmus HE, Coffin JM. *The molecular basis of tumour viruses, part III. RNA tumour viruses*. New York: Cold Spring Harbour Press, 1982.
 - 74 Sugamura K, Fujii M, Kobayashi N, Sakitani M, Hatanaka M, Hinuma Y. Retrovirus-induced expression of interleukin 2 receptors on cells of human B-cell lineage. *Proc Natl Acad Sci USA* 1984;81:7441-5.
 - 75 Schwab M, Ellison J, Busch M, Rosenau W, Varmus HE, Bishop JM. Enhanced expression of the human gene N-myc consequent to amplification of DNA may contribute to malignant progression of neuroblastoma. *Proc Natl Acad Sci USA* 1984;81:4940-4.
 - 76 Chang EH, Furth ME, Scolnick EM, Lowy DR. Tumorigenic transformation of mammalian cells induced by a normal human gene homologous to the oncogene of Harvey murine sarcoma virus. *Nature* 1982;297:479-83.
 - 77 Leder P, Battey J, Lenoir G, et al. Translocation among antibody genes in human cancer. *Science* 1983;222:765-71.
 - 78 Rabbitts TH, Foster A, Hamlin P, Baer R. Effect of somatic mutation within translocated c-myc genes in Burkitt's lymphoma. *Nature* 1984;309:592-7.
 - 79 Land H, Parada LF, Weinberg RA. Tumorigenic conversion of primary embryo fibroblasts require at least two cooperating oncogenes. *Nature* 1983;304:596-601.
 - 80 Cifone MA, Fidler IJ. Increasing metastatic potential is associated with increasing genetic instability of clones isolated from murine neoplasm. *Proc Natl Acad Sci USA* 1981;78:6949-52.
 - 81 Yang TP, Patel PI, Chinault AC, et al. Molecular evidence for new mutation at the hprt locus in Lesch-Nyhan patients. *Nature* 1984;310:412-4.
 - 82 Antonarakis SE, Waber PG, Kittur SD, et al. Haemophilia A: molecular defects and carrier detection by DNA analysis. *N Engl J Med* 1985;313:842-8.
 - 83 Gianelli F, Choo KH, Rees DJG, Boyd Y, Rizza CR, Brownlee GG. Gene deletions in patients with haemophilia B and anti-factor IX antibodies. *Nature* 1983;303:181-2.
 - 84 Rozen R, Fox J, Fenton WA, Horwich AL, Rosenberg LE. Gene deletion and restriction fragment length polymorphisms

- at the human ornithine transcarbamylase locus. *Nature* 1985;313:815-7.
- 85 Sudhof TC, Goldstein JL, Brown MS, Russell DW. The LDL receptor gene: a mosaic of exons shared with different proteins. *Science* 1985;228:815-20.
 - 86 Phillips JA III, Hjelle BL, Seeburg PH, Zachman M. Molecular basis for familial isolated growth hormone deficiency. *Proc Natl Acad Sci USA* 1981;78:6372-5.
 - 87 Prochovnik EV, Antonarakis S, Bauer KA, Rosenberg RD, Fearon ER, Orkin SH. Molecular heterogeneity of inherited antithrombin III deficiency. *N Engl J Med* 1983;308:1549-52.
 - 88 Chu M-L, Williams CJ, Pepe G, Hirsch JL, Prockop DJ, Ramirez F. Internal deletion in a collagen gene in a perinatal lethal form of Osteogenesis Imperfecta. *Nature* 1983;304:78-80.
 - 89 Wurzel JM, Parks JS, Herd JE, Nielsen PV. A gene deletion is responsible for absence of human chorionic somatomammotropin. *DNA* 1982;1:251-7.
 - 90 Yuasa Y, Srivastava SK, Dunn CY, Rhim JS, Reddy EP, Aaronson SA. Acquisition of transforming properties by alternative point mutations within c-bas/has human proto-oncogene. *Nature* 1983;303:775-9.
 - 91 Nakano H, Yamamoto F, Neville C, Evans D, Mizuno T, Perucho M. Isolation of transforming sequences of two lung carcinomas: structural and functional analysis of the activated c-K-ras oncogenes. *Proc Natl Acad Sci USA* 1984;81:71-5.
 - 92 Bos JL, Toksoz D, Marshall CJ, et al. Amino acid substitution at codon 13 of the N-ras oncogene in human acute myeloid leukaemia. *Nature* 1985;315:726-30.
 - 93 Pelica P-G, Lanfrancione L, Brathwaite MD, Wolman SR, Dalla-Favera R. Amplification of the c-myc oncogene in a case of human acute myelogenous leukemia. *Science* 1984;224:1117-21.
 - 94 Brodeur GM, Seeger RC, Schwab M, Varmus HE, Bishop JM. Amplification of N-myc in untreated neuroblastomas correlates with advanced disease stage. *Science* 1984;224:1121-4.
 - 95 Nau MM, Brooks BJ, Battey J, et al. L-myc, a new myc-related gene amplified and expressed in human small cell lung cancer. *Nature* 1985;318:69-73.
 - 96 Shtivelman E, Lifshitz B, Gale RP, Canaani E. Fused transcript of abl and bcr genes in chronic myelogenous leukemia. *Nature* 1985;315:550-4.
 - 97 Jakobovits A, Schwab M, Bishop JM, Martin GR. Expression of N-myc in teratocarcinoma stem cells and mouse embryos. *Nature* 1985;318:188-91.
 - 98 Ullrich A, Coussens L, Hayflick JS, et al. Human epidermal growth factor receptor cDNA sequence and aberrant expression of the amplified gene in A431 epidermoid carcinoma cells. *Nature* 1984;309:418-25.
 - 99 Yamamoto T, Ikawa S, Akiyama T, et al. Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth factor receptor. *Nature* 1986;319:230-4.
 - 100 Tsujimoto Y, Yunis J, Onorato-Showe L, Erikson J, Noell PC, Croce CM. Molecular cloning of the chromosomal breakpoint of B-cell lymphomas and leukemias with the t(11;14) chromosome translocation. *Science* 1984;224:1403-6.
 - 101 Chapelle A de La, Sankila E-M, Lindlöf M, Aula P, Norio R. Norrie disease caused by a gene deletion allowing carrier detection and prenatal diagnosis. *Clin Genet* 1985;28:317-20.
 - 102 Wieringa B, Hustinx T, Scheres J, et al. Glycerol kinase deficiency syndrome explained as X-chromosomal deletion. *Cytogenet Cell Genet* 1985;40:777.
 - 103 Sasaki H, Sakaki Y, Takagi Y, et al. Presymptomatic diagnosis of heterozygosity for familial amyloidotic polyneuropathy by recombinant DNA techniques. *Lancet* 1985;1:100.
 - 104 Haneda M, Chan SJ, Kwok SCM, Rubenstein AH, Steiner DF. Studies on mutant human insulin genes: identification and sequence analysis of a gene encoding [Ser^{B24}] insulin. *Proc Natl Acad Sci USA* 1983;80:6366-70.
 - 105 Stavnezer-Nordgren J, Kekish O, Zegers BJM. Molecular defects in a human immunoglobulin K chain deficiency. *Science* 1985;230:458-61.
 - 106 Scambler PJ, Williamson R. The structural gene for human coagulation factor X is located on chromosome 13q34. *Cytogenet Cell Genet* 1985;39:231-3.
 - 107 White PC, New MI, Dupont B. HLA-linked congenital adrenal hyperplasia results from a defective gene encoding a cytochrome P-450 specific for steroid 21-hydroxylation. *Proc Natl Acad Sci USA* 1984;81:7505-9.
 - 108 Bonthron DT, Markham AF, Ginsburg D, Orkin SH. Identification of a point mutation in the adenosine deaminase gene responsible for immunodeficiency. *J Clin Invest* 1985;76:894-7.
 - 109 Scambler PJ, Williamson R. The structural gene for human coagulation factor X is located on chromosome 13q34. *Cytogenet Cell Genet* 1985;39:231-3.
- Requests for reprints to: Dr AF Wright, Department of Psychiatry, Washington University School of Medicine, 4940 Audobon Avenue, St Louis, Missouri 63110, USA.