well as to the mapping of the human genome (Jeffreys *et al.*, 1985b; Gill *et al.*, 1985).

Anthropological studies

DNA polymorphisms, both mitochondrial (mt) and nuclear, have been applied to the study of evolutionary relationships in man. In one of the first studies mt RFLPs were analysed by Denaro *et al.* (1981), who examined differences between individuals in five ethnic groups and concluded that the origins of man were in Asia. A more comprehensive study (Cann *et al.*, 1987) has traced the origin of man to a single ancestor probably from Africa. RFLP frequency data from the β-globin gene cluster have also provided evidence for a small founder population in Africa (Wainscoat *et al.*, 1986a).

Genetic distance maps between populations can be constructed using the frequencies of polymorphic genetic markers, and DNA polymorphisms have contributed to such studies (Cavalli-Sforza *et al.*, 1986; Wainscoat *et al.*, 1986a).

Hypotheses about the evolution of genes can be formulated using RFLPs. For example, RFLPs associated with the β-globin gene cluster were used to determine the probable evolution of the sickle cell mutation (Kan

and Dozy, 1980), and a single origin has been proposed for the *Pi*^{m} gene responsible for α_1-antitryp in deficiency (Cox *et al.*, 1985).

1.4 Human gene mapping

Gene mapping in prokaryote organisms has been facilitated by the use of conjugation and transduction systems. In man, however, conventional genetic analysis has been hampered by his small family size, long generation time and breeding practices. Since man is a diploid organism, gene mutants are often masked in the heterozygote. Mapping the human genome is further complicated by the presence of introns and so-called 'junk' DNA (Orgel & Crick, 1980; Doolittle & Sapienza, 1980) between genes, and by the fact that, until recently, there were severe restrictions in the number of suitable markers available.

Gene mapping activities were aptly described a decade ago: "The exploration of the human chromosomes now so feverishly active has something of the excitement of geographical cartography, and quite a lot of people deserve a doublet of velvet such as Columbus offered to the first man to see land." (Race and Sanger, 1968). It is now possible to define an almost unlimited number of polymorphic markers and 'feverish' activity is unabated.

1.4.1 Recombination and genetic linkage

The term 'linkage' was originally used to describe genes on the same chromosome which were non-randomly assorted at meiosis, but was later also applied to the assignment of genes to sex chromosomes (X-linkage) or to autosomes (autosomal linkage). Renwick (1969) suggested the use of the term 'syntenic' to describe the relationship between two genes which could be shown to be on the same chromosome; these genes would not necessarily show linkage in terms of meiotic recombination. This convention is now generally accepted.

Recombination

The genes carried on a given chromosome constitute a linkage group, and would be inherited as a block were it not for crossing-over. Since recombination does occur, it is possible to measure the relative position of genes along a chromosome by the relative frequencies of crossing-over between pairs of loci. The distance between two loci is measured as the percentage of recombinant gametes, expressed as centiMorgans (cM) or cross-over units. For short intervals the map length in Morgans is equivalent to the recombination fraction (θ), where θ is defined as the proportion of gametes

that exist due to crossing-over. Since unlinked genes recombine with a a frequency of fifty per cent (indeed, fifty per cent is the maximum recombination frequency that can be directly observed), linkage can be said to exist between two loci when the recombination frequency between them is less than 0.5.

'Hotspots' such as those observed in the *Chi* sites of phage lambda, which are thought to be initiation sites for recombination, may be universal in eukaryotes (reviewed in White and Lalouel, 1987). Their existence has several implications for genetic linkage. The general relationship between recombination frequency and physical distance which holds at large distances would alter for closely linked loci if any such hotspot were present, and it would not be possible to predict the distance between a linked marker and a disease locus. Similarly, a hotspot located near a DNA sequence would have an effect on linkage disequilibrium between it and other linked loci. Furthermore, the possibility arises that a relationship could exist between recombination and the occurrence of mutation which could lead to the expression of disease. The mechanism of recombination is not clear, but construction of the human linkage map could lead to a better understanding of chromosome behaviour at meiosis.

Linkage analysis

Estimation of linkage must rely on probabilities, since for a family to be informative with respect to linkage at least one parent must be doubly heterozygous, and phase is often unknown. Statistical methods have been developed to detect linkage and to estimate recombination from observations of pairs of loci in human pedigrees (Morton, 1955; Ott, 1985). The likelihood of various linkage models is assessed by calculating likelihood ratios and thence lod scores. The lod score, $\log_{10}$ of the likelihood ratio, was devised by Morton (1955). By convention, when the sum of lod scores exceeds +3, linkage can be considered established for any given value of θ between 0.00 and 0.45. Values of +2 to +3 are considered 'suggestive', but if the sum of lod scores at a given value of θ is less than -2 then linkage at that value is excluded. Donis-Keller *et al.* (1987) suggested recently that a score of +3 is not definitive and that a score of +4 should be obtained before linkage can be considered established.

Linkage analysis calculations were performed at first by hand, with the help of tables (Morton, 1955). The advent of computers, appropriate algorithms and programmes such as LIPED (Ott, 1974) and LINKAGE (Lathrop and Lalouel, 1984) has revolutionised linkage analysis. Both programmes use a recursive algorithm for carrying

out likelihood calculations; LIPED uses the Elston-Stewart algorithm (Elston and Stewart, 1971) and LINKAGE is based on the algorithms of Elston-Stewart and Cannings (Cannings *et al.*, 1978). Previously, calculation of results was tedious and error-prone, but computer analysis has made analysis of complex phenotype systems or pedigree data manageable.

Collection of linkage data

Possibilities for obtaining linkage data improve when samples can be stored routinely so that new linkage markers can be used retrospectively. The establishment of permanent cell lines, for instance Epstein-Barr virus (EBV) transformed lymphocytes (Neitzel, 1986), is significant in this regard. Other developments leading to increased accuracy and efficiency in data collection include new methods of analysis by computer; new chromosomal morphological markers such as the fragile sites which provide interstitial and distal cytological markers (eighty-nine were reported in Human Gene Mapping 8, 1985); and new methods of finding linkage, such as somatic cell hybridization.

The discovery of new genetic markers is central to the collection of linkage data. An increase of h per cent in the number of polymorphic loci tested gives at least a 2h per cent addition to the linkage information, be-

cause of the extra pairwise linkage tests that are made possible. The quality of data also improves with additional markers, since non-paternity can be detected with greater certainty. In 1977 there were only some 50 marker loci available for family studies on a routine basis (McKusick and Ruddle, 1977). These included a wide range of blood groups, serum proteins, cell enzymes, antibody markers and cell surface antigens and required the involvement of many body tissues and laboratories to carry out the assays. The introduction of the new category of genetic marker, the RFLP, has transformed gene mapping.

1.4.2 Mapping strategies

Strategies employed in the assignment and regional localization of genes or anonymous DNA sequences are, on the one hand, family pedigree analyses based on classical Mendelian genetics, which lead to the construction of the genetic map. On the other hand, strategies which contribute directly to the physical map include parasexual genetics in the form of gene dosage studies, somatic cell hybridization, fluorescence-activated chromosome sorting, and *in situ* hybridization, as well as higher resolution mapping techniques. These are reviewed below.

1.4.2.1 The genetic map

Family studies

Scientists stimulated to devise methods for analysing segregation of genetic markers within human families include Bernstein, Haldane, Fisher, Penrose, Smith, Morton and Renwick (Reviewed in Renwick, 1971).

The value of family studies lies, firstly, in the possibility of quantifying intergenic distance. Secondly, family studies are uniquely useful for establishing the assignment of loci for genetic diseases which may be recognisable, at least initially, only at the clinical level. This is not possible using somatic cell hybrid studies. Lastly, genes can be indirectly assigned to chromosomes if they can be shown to be linked to a gene that is already assigned to a specific chromosome.

First attempts to map the genome exploited family and pedigree analysis, so that gene mapping in man was based on the determination of autosomal or X-linked inheritance of the genes for inherited syndromes and blood group antigens. Gene assignments relied on the segregation of a particular trait with a serological, biochemical, or chromosomal variant over several generations. The relative ease of establishing X-

linkage is illustrated by the fact that the first gene assignment, that of a gene for colour blindness, was to the X-chromosome (Wilson, 1911). Between 1911 and 1967 over 100 genes were assigned to the X-chromosome, but the first assignment to a human autosome was not made until 1968, with the assignment of the Duffy locus to chromosome 1 (Donahue *et al.*, 1968).

1.4.2.2 The physical map

Gene dosage mapping

Gene dosage mapping has used chromosomal deletions and duplications for gene assignments and regional mapping by the quantitative assaying of enzyme activity levels. For example, increased levels of glutathione reductase in trisomy 8 (de la Chapelle *et al.*, 1971) supported the gene's later independent assignment to chromosome 8 using somatic cell hybrids (Kucherlapati *et al.*, 1974). Enzyme activity does not always reflect gene dosage, however. An incorrect assignment by this method was that of galactose-1-phosphate-uridyltransferase to chromosome 3 (Allerdice and Tedesco, 1975). The enzyme was later correctly assigned to chromosome 9 by interspecific somatic cell hybridization (Mohandas *et al.*, 1979).

Somatic cell hybridization

The technique of somatic cell hybridization has been used, to date, to assign at least 85 per cent of all mapped genes, and it continues to be a valuable tool for gene assignments.

The history of the technique began in the early 1960s, with the observation of proliferating cells arising from mammalian cell fusion *in vitro* (Barski *et al.*, 1961). A selection system for the efficient isolation of somatic cell hybrids was then established (Littlefield, 1964) and modified by Davidson and Ephrussi (1965) to allow isolation of hybrids from fusions of drug-resistant permanent cell lines and normal diploid cells. Yerganian and Nel (1966) produced hybrids capable of indefinite proliferation using UV-inactivated Sendai virus, which had been shown by Harris and Watkins (1965) to induce the formation of interspecific heterokaryons. Then, in 1967, Weiss and Green isolated the first proliferating human-rodent hybrids. The key to their use for human gene mapping lay in the preferential and rapid loss of human chromosomes from these hybrids, which had not been observed in intraspecific hybrids.

The initial problem of unreliable chromosome identification was overcome by the introduction of

chromosome banding techniques (Caspersson *et al.*, 1970). This led to the first autosomal assignments of a selectable gene, thymidine kinase (TK) to 17 (Miller *et al.*, 1971) and of a non-selectable gene, lactate dehydrogenase-A (LDH-A) to 11 (Boone *et al.*, 1972). By 1972, fifty human genes had been reliably assigned using this technique (Ruddle, 1972).

The selection systems commonly used today require HGPRT, TK or APRT. Specific chromosomes in the hybrid lines can be identified, and in the absence of chromosome breakage, phenotypes encoded by genes i single chromosome are retained or lost as an intact group. Since the hybrids contain only a few human chromosomes in random combinations, assignments can be made by correlating a particular human genomic sequence, or gene product, with a specific human chromosome or subchromosomal fragment.

Gene mapping by somatic cell genetics today does not depend only on human-rodent hybrids, but also on hybrid cells created by gene transfer methods: microcell-mediated (Fournier and Ruddle, 1977), chromosome-mediated (McBride and Ozer, 1973; Klobutcher and Ruddle, 1981), or DNA-mediated (Wigler *et al.*, 1977).

As the methods of characterising hybrid clones were refined, it became clear that the use of a small number

of well-characterised hybrid clones could lead to more efficient mapping and thereby obviate the use of phenotype assays that were difficult or expensive. Such a clone panel was proposed by Creagan and Ruddle (1975). The minimum number of ideal hybrid lines necessary for a unique bimodal signature for every chromosome is surprisingly small. Twenty-four patterns can be provided with a minimum of five clones (2^5 = 32 patterns). However, redundancy in a panel is now thought to actually increase its accuracy. According to Kamarck *et al.* (1984), approximately ten hybrid clones, each containing about twelve randomly permuted human chromosomes, should be used for human mapping studies. Incorrect assignments would then be made only if three or four false-negative results were generated.

Whereas previously only biochemically defined gene products could be studied, the cloning of genes and DNA sequences and their use as probes has permitted rapid detection of specific sequences without the need for their expression in hybrid cells. The human α- and β-globin genes were the first to be mapped using a combination of somatic cell hybridization and molecular genetic techniques (Deisseroth *et al.*, 1977, 1978). Random single-copy DNA fragments as well as structural gene sequences can now be mapped, using hybrid panels.

Fine mapping using somatic cell hybrids is possible when the human parental cells contain deleted or translocated chromosomes; they could be spontaneous *in vitro* chromosomal rearrangements such as those which occur during hybrid stabilization; or they could be rearrangements induced with physical or chemical mutagens, as was first shown by Goss and Harris (1975). The Human Genetic Mutant Cell Repository is an important source of such chromosome abnormalities. The establishment of a large library of hybrid lines containing well-characterised mutant chromosomes will certainly complement the *in situ* hybridization technique and may even supplant it.

Fluorescence-activated chromosome sorting

Fluorescence-activated chromosome sorting could be considered the "mechanical" counterpart of somatic cell hybridization. Human mitotic chromosomes are resolved and sorted into size groups, and this is useful for regional mapping: translocations alter the size of chromosomes in a predictable way and their respective size classes are readily distinguishable from the normal sorting pattern. In this way the human β-, γ- and δ-globin genes were assigned to the distal portion of chromosome 11p (Lebo *et al.*, 1979). The full potential of this technology is not yet realised, partly because of problems relating to cost, the length

of time required to obtain sufficient material, and the difficulties in resolving chromosomes of the same size, for example the D-group chromosomes.

In situ hybridization

In situ hybridization was initially used to locate amplified single-copy genes in *Drosophila* polytene chromosomes (Gall and Pardue, 1969; Steffenson and Wimber, 1971). The technique has been more recently refined to detect unamplified genes (Malcolm *et al.*, 1981; Harper *et al.*, 1981; Harper and Saunders, 1981; Harper *et al.*, 1982), exploiting the discovery that ten per cent dextran sulfate accelerates the rate of hybridization of DNA probes to nucleic acid (Wahl *et al.*, 1979). DNA probes are radio-labelled (usually with ^{125}I-dNTPs or ^{3}H-dNTPs) and hybridized to mitotic chromosome preparations. Statistical analysis of grain distribution after autoradiography then provides a regional assignment.

The non-specific hybridization of probes limits the resolution of *in situ* hybridization to approximately 10 000 kb (10 cM). It is believed that chromosome deletions and translocations have a potential resolution of between one to two orders of magnitude less (Ruddle, 1981), making subchromosomal somatic cell

hybrid mapping panels more definitive and as accurate as regional mapping tools.

The *in situ* hybridization technique has made a substantial contribution to the mapping of human genes and single-copy DNA sequences. The technique has a variable success rate, however (Donlon, 1986), and requires a high degree of technical expertise. There are several sources of error inherent in the technique: for example, Goode *et al.* (1986) used somatic cell hybridization to show that the human DNA segment D1S1, mapped by two other laboratories to chromosome 1p36 by *in situ* hybridization, was in fact from chromosome 3 and duplicated in part on chromosome 1. In this case a probe believed to be single-copy was found not to be such. It seems that two mapping methods should be used to verify the localization of a gene or DNA segment, particularly a polymorphic one which may itself be used in mapping studies.

Refinement of the physical map

Current physical maps generated by the above methods have a resolution of about 10 000 kb. They can be extended to several hundred kilobases using chromosome-walking techniques (Steinmetz *et al.*, 1982), but these become tedious over long distances and if highly repeated DNA sequences are present. New techniques

allow the bridging of molecular and cytogenetic or somatic data in order to construct a physical map with a resolution of 100 to 1 000 kb. A strategy proposed by Smith and Cantor (1986a) requires the isolation of unbroken genomic DNA, its size fractionation into large fragments using rare-cutting restriction enzymes (sites for these enzymes seem to occur preferentially in HTF islands (Bird *et al.*, 1985), separation of the large fragments using the technique of pulsed field gel electrophoresis (Schwartz and Cantor, 1984; Smith and Cantor, 1986b; Carle *et al.*, 1986) and "macrorestriction mapping". Such mapping entails Southern analysis, preferably using junction probes which either span rare cutting sites or contain the ends of a single large fragment. A library of jumping probes could be used for this purpose (Collins and Weissman, 1984).

The average resolution of such a physical map will be approximately ten times that of the linkage map. The physical map would be useful for studying inter- and intra-population variation; and could be used to calibrate the genetic map, thereby revealing whether the linkage map is distorted by recombination hotspots, and where these are located.

1.4.3 The human linkage map

The rate of gene mapping has been significantly affected by technical developments, the first major one being the introduction of somatic cell hybrid analysis and the second that of recombinant DNA technology. Each technical epoch is characterised by a speeding up of the rate of gene mapping followed by a slowing down as the loci which are more directly amenable to a particular technique have been placed on a map (Pearson, 1985).

Calculations of the number of genetic markers needed to give an adequate coverage of the human genome for linkage studies suggest that approximately 1 600 random or 750 chromosome-specific markers would be required to give an average genetic distance between markers of approximately 20 cM (Lange and Boehnke, 1982; Bishop *et al.*, 1983). These figures are based on Renwick's (1971) estimate of 33 Morgans for the length of the human genome. The above two studies on number of markers emphasise that linkage cover of about 50 per cent of the genome could be achieved with 130 markers and an average distance between markers of 20 cM, and that filling in the spaces between markers becomes progressively more difficult as the spaces become smaller. So, for example, a 90 per cent cover would

require about 400 markers and a further 1 200 markers would be needed to cover the last 10 per cent.

Mapping the human genome is now feasible, given the potential for sufficient polymorphic markers, but optimal strategies should be used for data collection, management and analysis. The ordering of loci belonging to a linkage group requires multilocus analysis of their joint segregation; this is best inferred using three-point rather than two-point tests (Lathrop *et al.*, 1984, 1985). A mapping approach that relies on a common panel of test loci is thus sound.

The *Centre d'Etude du Polymorphisme Humain* (C.E.P.H.), under the directorship of Nobel Laureate Jean Dausset in Paris, has undertaken to make available to the scientific community DNA samples from a reference panel of families. International collaborators are presently characterizing DNA sequence polymorphisms on this common panel and a genetic map is being constructed using multilocus linkage information. C.E.P.H. has undertaken to manage the common data base. This collaboration should maximise the efficiency of linkage studies, since RFLP characterization is achieved using standard techniques, samples can be maintained through transformed cell lines, and some order can be brought to the 'shotgun' approach to linkage analysis.

Information about human gene loci (obtained by linkage analysis, and somatic cell genetics) is collected in McKusick's Catalogues and regularly updated. Every two years, a human gene mapping workshop is held and the proceedings are published in *Cytogenetics and Cell Genetics*. At the 8th workshop, held in 1985, the total number of mapped genes and markers reported was 1479, of which 559 were arbitrary DNA segments and 89 were fragile sites (de la Chapelle, 1985).

Preliminary genetic linkage maps of human chromosomes have been constructed, notably for the X-chromosome (Drayna and White, 1985); and for chromosomes 6p (Leach *et al.*, 1986; HGM8, 1985 pp 132,136,139,142), 11p (Kittur *et al.*,1985; Holm *et al.*, 1985), 12 (O'Connell *et al.*, 1985), and 13q (Leppert *et al.*, 1986).

A feature of maps already constructed is the difference in recombination frequencies between male and female meioses. It is not clear whether this phenomenon is generalised throughout the genome or localized. The non-independence of estimates of recombination frequencies, their non-uniform distribution, and their variation in occurrence between sexes will no doubt continue to provide a challenge to statisticians and geneticists in attempts to create linkage maps.

1.5 Albinism

Albinism is a condition known of and described since antiquity, although it has not been studied intensively until this century.

According to Sorsby (1958), the first written descriptions of the albinism phenotype probably appeared in the Bible, in the Pseudoepigrapha in the book of Enoch, and in Revelations 1:14. The term *albino* was first used in about 1600 AD by Balthazar Tellez, a Portuguese explorer and historian, and was derived from the Latin word *albus*, meaning *white*, to describe the white Negroes he had seen on the west coast of Africa (Pearson *et al.*, 1911-1913).

Garrod included albinism among his original inborn errors of metabolism (Garrod, 1908), and the disorder has since been accepted as a classic example of an autosomal recessive trait in man. Pearson *et al.* (1911-1913) collated what was then known about the incidence, nature, distribution and inheritance of albinism in a monumental monograph, and by 1913 the general outlines of what is now known about albinism had either been established or predicted.

During the present century animal models, rodent in particular, have been used to study the genetics and

biochemistry of pigmentation and albinism. In man, albinism phenotypes have been classified and terms defined. Witkop and his group have made a large contribution to this work. Pigmentation is better understood in the mouse than than in man (Silvers, 1979), but very little is known about the molecular genetics of albinism.

In the sections below, phenotypic classification and the problems in differentiating between albinism phenotypes will be described, followed by a discussion of possible metabolic defects in albinism. The tyrosinase-positive form of oculocutaneous albinism will be considered in more detail, since this is the phenotype most common in Africa, and was therefore suitable for a linkage study on this continent. Lastly, genetic aspects of albinism will be discussed, together with linkage studies, past and future.

1.5.1 Classification, incidence and differential diagnosis

Classification

Albinism in man refers to a genetically heterogeneous group of disorders, each presumably the result of different biochemical blockages in pigment formation. Two clinical categories are recognised: ocular albinism

(OA) and oculocutaneous albinism (OCA). The phenotypes are usually described as having in common decreased or absent pigmentation in the hair, skin and irides (OCA) or in the ocular fundus only (OA); impaired visual acuity; hypoplasia of the fovea and macula with varying degrees of nystagmus, strabismus and photophobia (Witkop *et al.*,1983); and anomalous decussation of central optic fibres (Creel *et al.*, 1974). Misrouting of brainstem auditory pathways has also been reported (Creel *et al.*, 1980).

Eleven disorders with features of OCA and four with features of OA have been described on the basis of clinical, biochemical, ultrastructural and genetic characteristics (Witkop *et al.*, 1973, Witkop *et al.*, 1983, Witkop 1985). The present study is not concerned with the OA phenotypes, which are inherited as X-linked or autosomal recessive traits. The OCA phenotypes are inherited in an autosomal recessive fashion, with the exception of one, which is an autosomal dominant.

In most populations, the tyrosinase-negative OCA (TNOCA) and tyrosinase-positive OCA (TPOCA) forms are the most frequently occurring phenotypes (Witkop *et al.*, 1983), but the TNOCA form appears to be rare or absent in Negroid populations (Dogliotti, 1973; King *et al.*, 1980; Kromberg, 1985). The only phenotypes which have been found in Africa are TNOCA, TPOCA, Brown

(BrOCA) and Rufous (RuOCA) (Barnicot, 1953; King *et al.*, 1980; Witkop *et al.*, 1983; Kromberg, 1985).

Incidence

The incidence of albinism varies widely in different population groups. In Europe the overall frequency is at least 1 in 15 000 (van Dorp, 1987). Among the highest reported rates are 1 in 227 for the Hopi Indians of Arizona (Woolf and Grant, 1962) and 1 in 143 for the Tele Cuna of Panama (Witkop *et al.*, 1983). In Africa estimates are generally higher than 1 in 5 000 (Barnicot, 1952; Kromberg, 1985), and in South Africa the Sotho-Tswana and Nguni have rates exceeding 1 in 2 000 (Kromberg and Jenkins, 1982). Among urban Johannesburg Bantu-speaking Negroes a prevalence of 1 in 3 900 has been estimated (Kromberg and Jenkins, 1982), giving a carrier rate of about 1 in 32.

Factors leading to a high prevalence of OCA in tropical populations have not been identified. It is difficult to choose between heterozygote advantage and random drift, but the fact that the phenotype occurs at a frequency too high for the gene or genes to be maintained by recurrent mutation possibly favours the influence of heterozygote advantage. Oettlè (1963) suggested that the carrier may have a paler skin colour than the homozygote non-albino individual and so derive some social

advantage that influences mating preference, and Woolf and Grant (1962) calculated that only a slight advantage of the carrier would maintain the albino gene in the population. Although heterozygotes for all forms of OCA are fully pigmented (King and Witkop, 1977), skin reflectance measurements in South African Negroids have shown a difference in skin colour that may influence prevalence rates (Roberts *et al.*, 1986; Kromberg, 1985).

Differential diagnosis

Difficulties inherent in the differential diagnosis of the albinism phenotype are emphasised by the finding that apparently normally-pigmented individuals can be albinos (van Dorp, 1987). This finding is based on the neurological data of Apkarian *et al.* (1983), who have shown that aberrant visual pathway decussation can be accurately detected, enabling the diagnosis of albinism with certainty. There is some doubt whether this is true for the Rufous type of albinism however (D. Castle and J.G.R. Kromberg, personal communication 1987), a type which was not included in the Apkarian study. Further, such electrophysiological examination has not distinguished between the TNOCA and TPOCA phenotypes (Apkarian *et al.*, 1983).

TNOCA and TPOCA were first distinguished biochemically using the hairbulb incubation test (Kugelman and van Scott, 1961; Witkop *et al.*, 1970), in which tyrosinase activity was estimated qualitatively under a dissecting microscope in anagen hairbulbs. A later test for detecting different phenotypes semi-quantitatively was the hairbulb tyrosinase assay described by King and Witkop (1976, 1977) and King and Olds (1985).

Clinical criteria for distinguishing albinism phenotypes have been well described (Witkop et al, 1983): in TNOCA the melanocytes are tyrosinase negative, there is a complete lack of pigment, and the colour of the hair is white, whereas in TPOCA the melanocytes are tyrosinase positive and there is some visible pigment, the skin contains lightly pigmented moles and freckles, and the colour of the hair varies from white to yellow-red to brown and darkens with advancing age. Intensity of pigment accumulation depends on racial background. There is overlap of clinical criteria, however, and individuals with a negative hairbulb test may have significant tyrosinase activity (Van Dorp, 1987).

A comprehensive study of over 100 albinos and their heterozygote family members has shown that Witkop's phenotypic classification based on clinical and hairbulb tests is not useful in distinguishing TNOCA

and TPOCA (van Dorp, 1987). These are probably not genetically distinct forms: although albino matings with normally pigmented offspring are well documented (Witkop *et al.*, 1983), the inference that the hairbulb test can distinguish between genotypes is perhaps unwarranted. Without a better understanding of the metabolic basis of all forms of albinism it is perhaps not possible to correlate phenotype with genotype. Current classifications are clearly not adequate.

1.5.2 The metabolic basis of albinism

Heritable metabolic defects in the melanocyte sy[stem] the skin and eye are responsible for the vary[ing] degrees of hypomelanosis in albinism. In human albinos, with the exception of certain syndromes in which albinism is a feature, melanocytes appear to be normally distributed (Witkop *et al.*, 1983). It seems, therefore, that aberrations lea[ding] to the albino phenotype probably occur during [the] synthesis of melanin and/or during its transport.

The melanin pathway is shown in Figure 1.1. Tyrosinase, an aromatic, copper-containing enzyme was previously thought to be the single enzyme controlling melanin synthesis. The enzyme catalyzes the hydroxylation of tyrosine to dopa and the oxidation of dopa to dopaquinone (Lerner and Fitzpatrick, 1950)

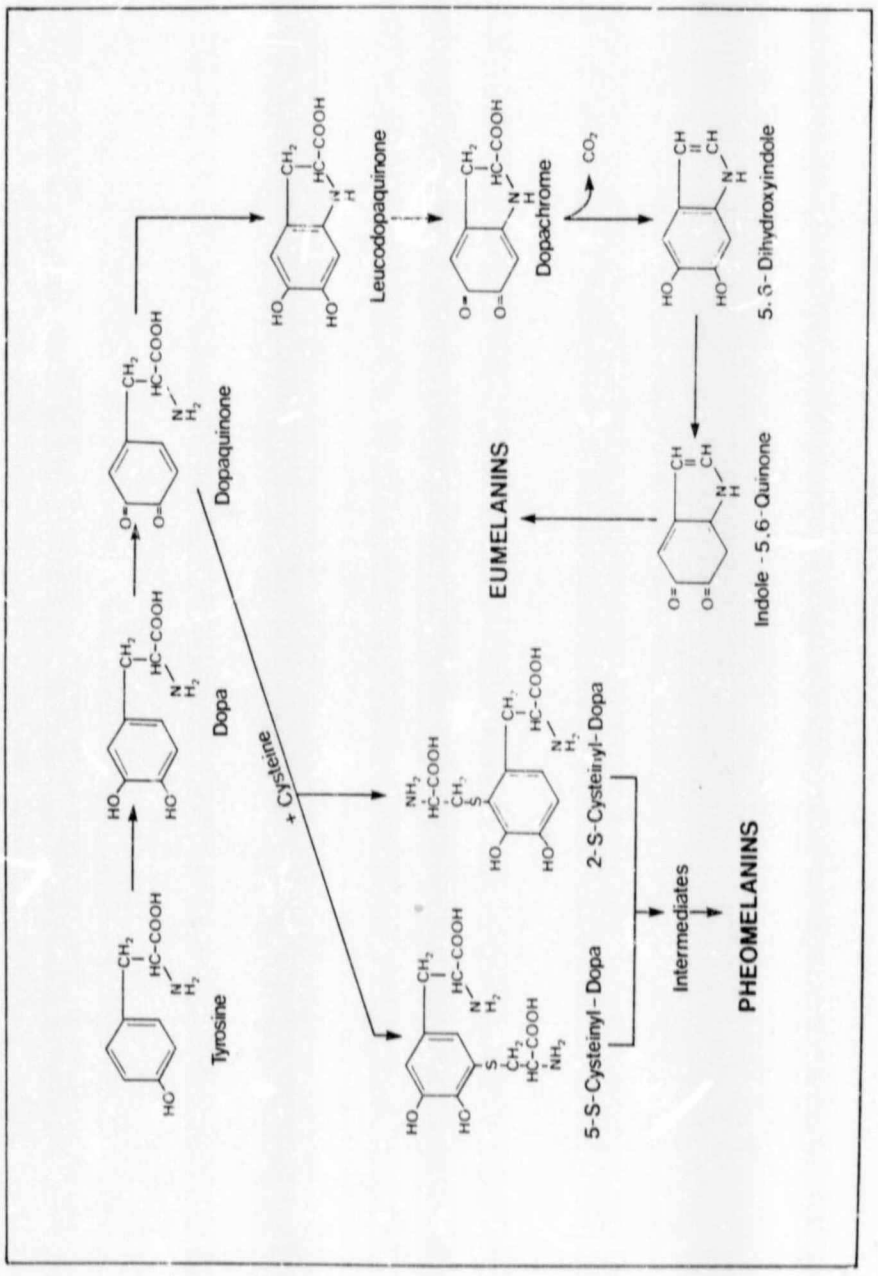

Figure 1.1 The melanin pathway, showing pheomelanin and eumelanin synthesis from tyrosinase (after King and Olds, 1985).

as well as the conversion of 5,6-dihydroxyindole to melanochrome (Korner and Pawelek, 1982).

A second enzyme, dopachrome oxidoreductase (DCOR), is now known to be required in the synthesis of melanin. The enzyme catalyzes the conversion of dopachrome to 5,6-dihydroxyindole and also apparently blocks the pigment pathway at the latter compound in the absence of tyrosinase (Barber *et al.*, 1984).

The eumelanins and pheomelanins follow different biosynthetic pathways. The eumelanins are brown-black pigments formed in the presence of oxygen and the pheomelanins are red-yellow pigments which require cysteine in their formation. In OCA individuals (with the exception of TNOCA), hypopigmentation at birth is followed by development of white or light yellow hair, dark yellow and then blond or light brown hair. This pattern of pigment development suggests that pheomelanin is the first to form when an individual with OCA forms pigment, followed after a long period by eumelanin formation (King and Olds, 1985).

TNOCA is notable in that there is no detectable tyrosinase activity and no melanin formation, unlike all other types of OCA described (King and Witkop, 1976). It is currently thought that the metabolic block is in the initial step in the pathway and is most

likely to represent an inactive or absent tyrosinase (King and Witkop, 1976). Yellow mutant OCA, now thought to be allelic to TNOCA (Hu *et al.*, 1980), could be the result of a partial deficiency of tyrosinase activity. The residual activity would produce small amounts of dopaquinone, which would rapidly enter the high affinity pheomelanin pathway, and only as the pathway approached saturation, would eumelanin form (King and Olds, 1985).

The biochemical defect responsible for the TPOCA phenotype is not known, although several possibilities have been considered. The TPOCA phenotype and the tyrosinase detectable in hairbulb assays are evidence for the metabolic defect in TPOCA individuals occurring after the formation of dopaquinone. Recent studies have shown the existence of enzyme control in the distal eumelanin pathway, suggesting sites where a block could develop (Pawelek *et al.*, 1980; Korner and Pawelek, 1982; Barber *et al.*, 1984). Since eumelanin clearly forms in some individuals with TPOCA, any postulated block cannot be complete.

A substrate transport defect is not responsible for the TPOCA phenotype, although tyrosine levels in TPOCA individuals vary from below normal to normal (Witkop *et al.*, 1972). The defect is possibly due to a melanin transport block: TPOCA melanocytes show dense

accumulations of pre-melanosomes and melanosome complexes, which may indicate a defective ability to pass their products to keratinocytes (Witkop *et al.*, 1973). A further possibility is that the genetic defect is regulatory, rather than structural, suggested by the finding of pigmented moles in non-pigmented epidermis which indicated the formation of tyrosinase (van Dorp *et al.*, 1983).

Tyrosinase in TPOCA individuals is kinetically normal (King *et al.*, 1978) and activity values fall into a wide range, sometimes exceeding that for normally pigmented individuals (King and Witkop, 1976). Some investigators have considered this to be evidence for genetic heterogeneity in the TPOCA phenotype (Witkop, 1985), but even if this is true the classes cannot be clearly separated using known methodologies.

In summary, the gene or genes responsible for the TPOCA phenotype could be structural or regulatory, and may be involved in either the metabolism or transport of melanin. The establishment of genetic linkage might be a first step in resolving these conflicting ideas.

1.5.3 Genetic and linkage studies

It has been estimated that the additive interaction of genes at 3 or 4 loci is sufficient to account for the

variation in skin colour between Negro and Caucasoid Americans (Stern, 1970), and that this number of loci could account for the range of skin colour found throughout human populations. Harrison (1973) suggests that pigmentation may be considerably more complicated, and this view seems more likely. In mice, more than 147 genes at about 62 loci affect skin and hair colour (Silvers, 1979). Human and other albino animals have melanocytes and melanosomes that are structurally similar, leading to the view that the genetic mechanisms regulating melanin production are comparable (Witkop *et al.*, 1983) and may approach a similar complexity.

In man, at least two genes are responsible for the albinism phenotypes. The existence of two different non-allelic genes for OCA phenotypes has been clearly established (Trevor-Roper 1952; Witkop *et al.*, 1970). Further, genes for at least four phenotypes, wild-type, TNOCA, Ym and Pt, occupy the same locus in man (Hu *et al.*, 1980; Witkop, 1985).

The human OCA loci have not yet been assigned chromosomally, although the establishment of genetic linkage would be of considerable value. Once an OCA gene has been localized, it can potentially be cloned and will be useful in unraveling the molecular basis of pigmentation or part of it. Polymorphic markers which may be

variation in skin colour between Negro and Caucasoid Americans (Stern, 1970), and that this number of loci could account for the range of skin colour found throughout human populations. Harrison (1973) suggests that pigmentation may be considerably more complicated, and this view seems more likely. In mice, more than 147 genes at about 62 loci affect skin and hair colour (Silvers, 1979). Human and other albino animals have melanocytes and melanosomes that are structurally similar, leading to the view that the genetic mechanisms regulating melanin production are comparable (Witkop *et al.*, 1983) and may approach a similar complexity.

In man, at least two genes are responsible for the albinism phenotypes. The existence of two different non-allelic genes for OCA phenotypes has been clearly established (Trevor-Roper 1952; Witkop *et al.*, 1970). Further, genes for at least four phenotypes, wild-type, TNOCA, Ym and Pt, occupy the same locus in man (Hu *et al.*, 1980; Witkop, 1985).

The human OCA loci have not yet been assigned chromosomally, although the establishment of genetic linkage would be of considerable value. Once an OCA gene has been localized, it can potentially be cloned and will be useful in unraveling the molecular basis of pigmentation or part of it. Polymorphic markers which may be

found in and around the loci will enable counselling for 'at-risk' couples, and prenatal diagnosis. Inbreeding has been recognised as a major factor where there is a high incidence of albinism (Okoro, 1975), and in some communities this factor may be altered by counselling. Prenatal diagnosis and therapeutic abortion could be offered if it were considered morally justifiable; the ethical issues are not dissimilar, perhaps, from those pertaining to other genetic disorders associated with a reasonable quality of life.

On reviewing the literature, the author identified only two reports of linkage studies in albinos. The first, the proposal of linkage between TNOCA and cerebellar ataxia, is interesting. In a highly inbred kindred with a sibship of nine, a lod score of 2.437 at $\theta = 0$ was obtained (Skre and Berg, 1974). No further such linkage has been reported, but a family has recently been described in which cerebellar ataxia occurred with TNOCA (Bamezai *et al.*, 1987).

The second study was in a three-generation family of Yemenite-Jewish descent, where an oculocutaneous type of albinism was associated with schizophreniform psychosis (Baron, 1976). Evidence for linkage was suggestive, with a maximum lod score of 1.55 at $\theta = 0$. The type of OCA was not indicated and skin tyrosinase was not tested, although it had been possible to

distinguish OCA phenotypes for some years by the hair-bulb incubation test (Kugelman and van Scott, 1961; Witkop *et al.*, 1970).

In 1965, Greiner and Nicolson observed hypomelanosis in schizophrenic patients, and hypothesised that schizophrenics are abnormally pigmented. An extensive study by Robins (1970), however, could not show an association between schizophrenia and melanogenesis. Few other individuals with both schizophrenia and albinism have been reported, and in no case has there been evidence for a genetic relationship (Leibowitz *et al.*, 1978).

Investigation of linkage to a recessive disorder such as albinism requires a large number of informative families and the co-operative efforts of clinicians and geneticists (Wong *et al.*, 1986a). These factors may have inhibited the initiation of such a project to date.

Once such a project has been initiated, the first genetic markers to be investigated for linkage will be those routinely used in the laboratory. However, first approaches towards establishing linkage to an OCA phenotype ought to include the comparison of other mammalian linkage groups with those in man. The diverse groups of extant mammals arose from closely

related ancestral stocks and therefore share many genes in common. Comparison of the linkage maps of cat and man reveals a striking degree of homology (O'Brien and Nash, 1982), and there is evidence for a conserved chromosome group which contains albino loci in the mouse, cat, rat and deermouse, and is homologous to chromosome 11 in man (O'Brien *et al.*, 1986).

The pink-eyed dilution (p/p) mutant variety of coat colour in the mouse closely approximates TPOCA in man, and human TNOCA is strikingly similar to the type of albinism in mice regulated by the *c* locus (Witkop, 1985). The *p*, *c* and haemoglobin β chain loci are included in the conserved linkage group, and this constitutes indirect evidence for the human albinism loci equivalent to *p* and *c* being linked to the human β-globin gene cluster.

1.6 A note on nomenclature

Terms used to identify population groups are often emotive and ill-defined. The term Black is currently socially and politically more acceptable than is Negro, for example. Indeed, 'Negro' has been included in a UNESCO report as being derogatory and unacceptable (Baird, 1970). The term Black is not however scientifically defined, and in some contexts might refer to all people who are not white Caucasoids, including

populations of mixed racial origin. An alternative term 'Bantu' is sometimes used, but studies on a variety of genetic markers, including blood-groups, haemoglobins, serum proteins and enzyme polymorphisms, have shown that there is no biologically valid basis for the term Bantu (Tobias, 1971). 'Bantu' refers, in fact, to linguistic affiliation and not physical type.

Participants at a conference on The Peoples of Southern Africa in 1971 attempted to define nomenclature in order to avoid confusion and misleading assumptions. Terms were agreed upon with which to refer to the biological entities (race or physical type), the languages, and the economies or modes of subsistence, of the diverse southern African populations (Jenkins and Tobias, 1977). It was agreed that for three of the population groups the terms San, Khoikhoi and Negro would be used; their respective languages would be referred to as Bushman, Hottentot and Bantu; and their mode of subsistence could be any of hunter/hunter-gatherer, pastoralist (or herder) and cultivator, peasant-villager or townsman.

Terms used in the present study will be those adopted at the 1971 conference. In particular, 'San' will be used to refer to the group sometimes referred to as 'Bushman'; and 'Southern African Negro' (or, simply,

Negro) will be used to refer to black Bantu-speaking Africans from the sub-Continent.

1.7 Aims

The aims and objectives of the study can be summarised as follows:

1. To isolate recombinant bacteriophage containing unique sequence DNA inserts ('single-copy') from *Eco*RI and *Hae*III/*Alu*I human genomic libraries, and to use these as RFLP-detecting probes in the South African Negroid population.

2. To characterise any polymorphisms by a) determining whether the RFLP 'alleles' (morphs) are inherited in a Mendelian fashion; b) determining their frequencies in the Negroid and other southern African populations; and c) localizing the RFLP-detecting probes chromosomally.

3. To contribute the new polymorphisms to the international effort of creating a human linkage map via the C.E.P.H. programme.

4. To use the polymorphisms in a search for linkage to tyrosinase-positive oculocutaneous

albinism, and to co-ordinate the use of other genetic markers in this search.

The strategy adopted for searching for RFLP was that pioneered by R. White (Wyman and White, 1980). Fourteen enzymes were used which had been shown in the literature to detect RFLP most often (Cooper and Schmidtke, 1984). Only six or seven individuals were screened with each enzyme, the rationale for this being that polymorphisms for which the frequency of the rarer allele is very low would be missed and that these would not in fact be useful for linkage studies.

1.8 Plan of the thesis

In the first chapter, the study has been broadly motivated, the literature reviewed so that the present study can be placed in context, the terms used for the populations studied clarified, and the plan of the thesis outlined. Specific aims of the study are defined at the end of this chapter.

The subjects of the study are described in the second chapter, as are the methodologies used in pursuing the study aims. Results obtained are presented and discussed in the next three chapters, and compared with any similar findings in the literature. Lastly, there

is a general discussion, conclusions are drawn, and suggestions for future research are presented.

CHAPTER 2