https://dspace.org/

Genes in the aetiology of oculocutaneous albinism in Sub-Saharan Africa and a possible role in tuberculosis susceptibility

Kerr, Robyn

2008-10-01T10:01:11Z
http://hdl.handle.net/10539/5707

Downloaded from a DSpace repository. https://dspace.org/

Genes in the Aetiology of Oculocutaneous Albinism in Sub-Saharan Africa and a Possible Role in Tuberculosis Susceptibility

Robyn Kerr

2007

Genes in the Aetiology of Oculocutaneous Albinism in Sub-Saharan Africa and a Possible Role in Tuberculosis Susceptibility

Robyn Kerr

A thesis submitted to the Faculty of Medicine, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Doctor of Philosophy.

Johannesburg, South Africa, 2007

ABSTRACT

In southern Africa, oculocutaneous albinism (OCA) is the most common autosomal recessive disease amongst black Africans, occurring with a prevalence of approximately 1 in 3 900 individuals. OCA occurs in southern African Caucasoids with a frequency that reflects the European origins of this population, approximately 1 in 20-30 000. OCA type 1 is caused by mutations in the *tyrosinase* gene on chromosome 11q. *Tyrosinase* mutations occur in the Caucasoid population but are extremely rare in black Africans. OCA type 2 is caused by mutations in the *P* gene on chromosome 15q. *P* gene mutations occur in both the black and Caucasoid populations. A sub-type of OCA2 seen in black individuals, brown OCA (BOCA), is also caused by mutations at the *P* gene locus.

A mutation screen was undertaken to identify disease-causing mutations in a group of OCA subjects from Sub-Saharan Africa. A common *P* gene mutation had been identified in the black population, a 2.7 kb intragenic deletion, accounting for 78% of *P* gene mutations in this group. No common *tyrosinase* mutations have been identified to date, in any population. A cohort of OCA subjects from South Africa, Lesotho, Zambia and the Central African Republic (CAR) were available for study in our laboratory. All subjects were screened for the 2.7 kb deletion mutation. Subjects homozygous for this mutation were excluded from further study. Subjects where one or two mutations remained to be identified were included in a mutation screen (63 blacks and 9 Caucasoids). Depending on the clinical categorisation of the type of albinism, subjects were screened for *P* gene mutations only (black OCA2) or were screened for *P* gene mutations and *tyrosinase* mutations (BOCA, unclassified black OCA and unclassified Caucasoid OCA).

All 72 subjects were screened for *P* gene mutations and ten putative pathogenic mutations were identified. In the group of black OCA2 patients, four mutations which are likely to be pathogenic were found: A334V, 614delA, 683insT and 727insG. Mutations were identified in four individuals with an unusual hypopigmentation phenotype: E678K was found in the homozygous state in an individual from the CAR. A second individual was found to be a compound heterozygote for the I370T and the L688F mutations. A third individual was found to be heterozygous for the I370T mutation. Three *P* gene mutations were found in

the Caucasoid sample: IVS 14-2 (a→g), V350M and P743L. No further mutations were identified in the BOCA sample. The *P* gene screen comprised 72 subjects, but 40 were heterozygous for the 2.7 kb deletion, therefore (144 minus 40 alleles) 104 alleles remained to be identified. Identification of 12/104 alleles means that a further 11.5% of the unknown *P* gene mutations are now accounted for. Thirty three of the 72 subjects were included in a further mutation screen – at the *tyrosinase* locus. Four mutations were identified, all in the Caucasoid group. Compound heterozygosity was shown in two individuals, one carrying the mutations, E294K and A490D and the other, P431T and T373K.

Following mutation analysis of the *P* gene, it was apparent that a proportion of mutations did not lie in the coding region of the gene and it was proposed that some of the remaning unidentified mutations might be found in the 5'control or promoter region of the gene. At that time, sequence data for the region upstream of the *P* gene was not known, and so an attempt was made to clone the 5'region of the *P* gene. Two approaches were adopted – a bacterial artificial chromosome (BAC) known to contain this region was subcloned; and secondly, an inverse PCR experiment was undertaken. Neither experiment was successful in generating *P* gene promoter sequence.

Variation at the *P* locus was investigated in a second context. This region of chromosome 15q was implicated as a host susceptibility locus for the infectious disease, tuberculosis (TB). A case-control study was undertaken to compare the frequencies of five intragenic, polymorphic markers in the *P* gene: the 2.7 kb deletion, the R305W polymorphism and the microsatellite markers, D15S1533, D15S1536 and D15S1537, between a group of black South African TB patients from Gauteng and healthy community controls and between a group of mixed-ancestry (Coloured) TB patients and healthy controls from the Western Cape region. Presence or absence of the 2.7 kb deletion mutation does not appear to influence susceptibility to TB in either the black or Coloured population samples studied here. The W allele of the R305W polymorphism is significantly ($p<0.05$) more common in the black patient group than in the black control group, suggesting it may be in linkage disequilibrium with a disease susceptibility allele. Microsatellite marker analysis showed that, in the black population, allele 18 at the D15S1533 locus is significantly ($p<0.05$) associated with susceptibility to TB. In the Cape Coloured population, alleles 20 and 27 at the D15S1533 locus, allele 12 at the D15S1536 locus and allele 16 at the D15S1537 locus are over represented in the patient group suggesting they may be markers for increased

susceptibility to TB. Further, in the Coloured population alleles 12, 13 and 15 at the D15S1537 locus showed significant ($p<0.05$) association with normal controls and may be in linkage disequilibrium with protective or resistance alleles.

The results of this study support the proposal of a TB susceptibility locus on chromosome 15q. OCA-causing mutations have been identified, but many remain elusive. Further characterisation of this region will give us a better understanding of the biological consequences of variation both within and around the *P* locus.

DECLARATION

I declare that this is my own, unaided work, unless otherwise acknowledged. This thesis is being submitted for the degree of Doctor of Philosophy at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in any other university.

I declare that this work has been approved by the Ethics Committee for Research on Human Subjects of the University of the Witwatersrand (clearance certificate numbers M940711 and M980616).

Robyn Kerr

_____ day of July, 2007

Again, to those who made this possible.

This work is dedicated to my family; my husband, Marq, my beautiful and precious children, Nicholas James and Ciara Grace, my parents, Ann and Brian, my sisters Sandy and Jenni and my aunt and uncle, Lois and Ford. Thank you.

ACKNOWLEDGEMENTS

My sincere thanks to:

- All the individuals involved in this study who consented to donating blood and, in certain instances, to using photographs of themselves, for research purposes. Without their participation this study would not have been possible.
- My supervisor, Professor Michele Ramsay, whose understanding and patience was beyond that ordinarily called for.
- Professor Robert Nicholls, who allowed me to visit his laboratory at Case Western Reserve University, Cleveland, Ohio, for a 4 month period in 1997. An opportunity of learning and exposure for which I am extremely grateful. Thank you also for the BAC clone 263 022 and the *P* promoter fragment probe.
- My co-workers in the laboratory, Gwynneth Stevens, Prashiela Manga, Sarah Salm and Robyn Hetem, who all contributed to work done in this study.
- Dr Eileen Hoal-van Helden, University of Stellenbosch, Tygerberg, for donation of Coloured DNA samples used in the TB study.
- Professor Jenny Kromberg, for helpful discussion after reading parts of this manuscript.
- Dr Dries Oelofse for his time and expertise in bioinformatic analysis of data.
- All my colleagues in the laboratory, over a period of (many) years, who made going to work such a pleasure.

Financial support for this study was received from the South African Institute for Medical Research (SAIMR), now the National Health Laboratory Service (NHLS), the Foundation for Research and Development and the University of the Witwatersrand.

I would like to especially thank these people: My sister, Jenni, for her continuous support and many hours spent helping to prepare this manuscript; and my friends, Val and Kim, who, in many ways, are responsible for the completion of this thesis.

PUBLICATONS

Kerr, R., Stevens, G., Manga, P., Salm, S., John, P., Haw, T. and Ramsay, M. (2000) Identification of *P* gene mutations in individuals with oculocutaneous albinism in sub-Saharan Africa. *Hum Mutat* 15, 166-72.

TABLE OF CONTENTS

ABSTRACT ---------- **III**

DECLARATION ---------- **VI**

ACKNOWLEDGEMENTS ---------- **VIII**

PUBLICATONS ---------- **IX**

TABLE OF CONTENTS ---------- **X**

LIST OF FIGURES ---------- **XV**

LIST OF TABLES ---------- **XVI**

LIST OF ABBREVIATIONS ---------- **XVII**

CHAPTER 1 ---------- **1**

INTRODUCTION ---------- **1**

1.1 INTRODUCTION ---------- **2**

1.2 PLAN OF THE THESIS ---------- **5**

1.3 MELANOGENESIS ---------- **5**

1.3.1 EXTRACUTANEOUS MELANOCYTES ---------- 8

1.4 RACIAL DIFFERENCES IN SKIN ARCHITECTURE ---------- **9**

1.5 TYPES OF ALBINISM ---------- **10**

1.5.1 TYROSINASE RELATED ALBINISM (MIM 203100) ---------- 13
1.5.1.1 OCA1A: Tyrosinase-negative OCA ---------- *14*
1.5.1.2 OCA1B: Yellow-mutant OCA ---------- *15*
1.5.1.3 OCA1MP: Minimal-pigment OCA ---------- *15*
1.5.1.4 OCA1TS: Temperature-sensitive OCA ---------- *16*
1.5.2 *P* GENE RELATED ALBINISM (MIM 203200) ---------- 16
1.5.2.1 OCA2 ---------- *16*
1.5.2.2 BOCA: Brown OCA ---------- *18*
1.5.3 OTHER FORMS OF ALBINISM ---------- 19
1.5.3.1 OCA3 (Rufous OCA, MIM 203290) ---------- *19*
1.5.3.2 OCA4 (MIM 606574) ---------- *20*
1.5.3.3 Autosomal dominant OCA ---------- *21*
1.5.3.4 Ocular albinism ---------- *22*

1.6 OTHER HYPOPIGMENTATION DISORDERS ---------- **24**

1.6.1 NEURAL CREST DEFECTS ---------- 24
1.6.1.1 Waardenburg Syndrome (WS1 and WS2: MIM 193500; WS3 MIM 148820) ---------- *24*
1.6.1.2 Hirschsprung Disease (MIM 142623) ---------- *25*
1.6.1.3 Piebaldism (MIM 172800) ---------- *26*
1.6.2 ORGANELLE DEFECTS ---------- 26
1.6.2.1 Hermansky-Pudlak Syndrome (MIM 203300) ---------- *27*
1.6.2.2 Chediak-Higashi Syndrome (MIM 606897) ---------- *27*
1.6.3 ACQUIRED HYPOPIGMENTATION DEFECTS ---------- 27
1.6.3.1 Leprosy (MIM 246300) ---------- *28*
1.6.3.2 Tuberculosis ---------- *29*

1.7 MOUSE PIGMENT LOCI ---------- **30**

1.7.1 THE PINK-EYED DILUTION LOCUS (*P*) ---------- 30
1.7.1.1 The p^{un} allele ---------- *31*
1.7.2 THE AGOUTI (*A*) LOCUS ---------- 31
1.7.3 THE ALBINO (*C*) LOCUS ---------- 32
1.7.4 THE BROWN (*B*) LOCUS ---------- 33

1.7.5 THE SLATY (*SLT*) LOCUS ---------- 34

1.8 HUMAN PIGMENT LOCI ---------- 34

1.8.1 MELANOCORTIN 1 RECEPTOR (*MC1R*) ---------- 35
1.8.2 HUMAN AGOUTI SIGNALLING PROTEIN (*ASIP*) ---------- 36
1.8.3 THE TYROSINASE FAMILY ---------- 36
1.8.3.1 Tyrosinase (TYR) ---------- 37
1.8.3.2 Tyrosinase related protein 1 (TYRP1) ---------- 37
1.8.3.3 Tyrosinase related protein 2 (TYRP2) ---------- 38
1.8.4 THE *P* GENE ---------- 38
1.8.4.1 Cloning the gene at the OCA2 locus ---------- 39
1.8.4.2 P gene mutations cause OCA2 ---------- 40
1.8.4.3 A common P gene mutation ---------- 40
1.8.4.4 Other P gene mutations ---------- 41
1.8.4.5 Structure of the P locus ---------- 41
1.8.4.6 Family of P-like proteins ---------- 42
1.8.4.7 Function of the P protein ---------- 43
1.8.5 ASIP, MSH, MC1R AND THE PHEOMELANIN/EUMELANIN SWITCH ---------- 44
1.8.6 THE ROLE OF PIGMENT GENES IN NORMAL PIGMENT VARIATION IN HUMANS ---------- 44
1.8.6.1 Variants in the P gene influencing normal pigment variation ---------- 45

1.9 THE *P* GENE ENVIRONMENT ---------- 46

1.9.1 PROXIMAL TO *P*: THE PRADER-WILLI SYNDROME/ANGELMAN SYNDROME REGION ---------- 46
1.9.1.1 Deletions in the PWS/AS region ---------- 47
1.9.1.2 Uniparental Disomy in the PWS/AS region ---------- 47
1.9.1.3 Imprinting mutations in the PWS/AS region ---------- 48
1.9.1.4 Clinical presentation of PWS and AS ---------- 48
1.9.1.5 The hypopigmentation phenotype ---------- 49
1.9.1.6 The hyperpigmentation phenotype ---------- 50
1.9.2 DISTAL TO *P*: THE *HERC2* LOCUS ---------- 51
1.9.2.1 The PWS/AS deletion breakpoints ---------- 51
1.9.2.1.1 Other diseases caused by chromosome rearrangements ---------- 51
1.9.2.2. Identification of low-copy repeat elements at PWS/AS breakpoint junctures ---------- 52
1.9.2.3 The mouse p^{6H} allele ---------- 52
1.9.2.4 The gene associated with the PWS/AS breakpoints as well as the jdf-2 mouse phenotype: HERC2 ---------- 53
1.9.2.5 Herc2 may play a role in pigmentation ---------- 53

1.10 PSYCHOSOCIAL ASPECTS OF ALBINISM ---------- 54

1.10.1 SOCIAL ATTITUDES ---------- 55
1.10.2 THE MOTHER-CHILD BOND ---------- 55
1.10.3 MYTHS AND SUPERSTITIONS ---------- 56
1.10.4 THE ISSUE OF WHITENESS IN THE AFRICAN CONTEXT ---------- 56
1.10.5 PRENATAL DIAGNOSIS ---------- 57

1.11 HOST-RELATED GENETIC BACKGROUND AND SUSCEPTIBILITY TO THE INFECTIOUS DISEASE TUBERCULOSIS ---------- 59

1.11.1 HISTORY OF TUBERCULOSIS ---------- 59
1.11.2 THE INTRODUCTION OF TUBERCULOSIS INTO SOUTH AFRICA ---------- 60
1.11.3 CLINICAL TUBERCULOSIS ---------- 61
1.11.4 SELECTION AND INFECTIOUS DISEASE – MALARIA AS THE CLASSIC EXAMPLE ---------- 62
1.11.5 HOST GENETICS AND INFECTIOUS DISEASE ---------- 62
1.11.6 MOLECULAR GENETIC TOOLS TO STUDY HOST SUSCEPTIBILITY TO INFECTIOUS DISEASE ---------- 63
1.11.6.1 Genome wide screens ---------- 64
1.11.6.2 Candidate genes ---------- 65
1.11.6.3 Candidate genes studied in the SA population ---------- 67
1.11.6.4 Mouse genetic susceptibility studies ---------- 69
1.11.7 MENDELIAN SUSCEPTIBILITY TO MYCOBACTERIAL DISEASES ---------- 70
1.11.8 EVIDENCE FOR A TUBERCULOSIS SUSCEPTIBILITY LOCUS ON CHROMOSOME 15Q ---------- 71

AIMS AND OBJECTIVES ---------- 72

CHAPTER 2 ---------- **74**

SUBJECTS, MATERIALS AND METHODS ---------- **74**

2.1 INTRODUCTION ---------- **75**

2.2 SUBJECTS ---------- **76**

2.2.1 SUBJECTS WITH ALBINISM ---------- 76
2.2.1.1 Subjects with unclassified albinism ---------- *76*
2.2.2 SUBJECTS WITH TUBERCULOSIS ---------- 79

2.3 MATERIALS AND METHODS ---------- **82**

2.3.1 COLLECTION AND STORAGE OF BLOOD SAMPLES ---------- **82**

2.3.2 PCR ---------- **83**

2.3.2.1 THE *P* GENE 2.7 KB DELETION PCR ASSAY ---------- 83
2.3.2.2 *P* GENE PCR ---------- 85
2.3.2.3 *TYROSINASE* GENE PCR ---------- 86

2.3.3 PCR-SSCP ANALYSIS ---------- **86**

2.3.3.1 RADIOACTIVE PCR ---------- 87
2.3.3.2 POLYACRYLAMIDE GEL ELECTROPHORESIS ---------- 87
2.3.3.3 SEQUENCING ---------- 87

2.3.4 SOUTHERN BLOTTING ---------- **88**

2.3.4.1 SOUTHERN BLOTTING FOR THE *P* GENE ---------- 89
2.3.4.1.1 DNA digestion ---------- *89*
2.3.4.1.2 Agarose gel electrophoresis ---------- *89*
2.3.4.1.3 Southern blotting ---------- *89*
2.3.4.1.4 Radioactive labelling of the probe: pcDNA3-P ---------- *90*
2.3.4.1.5 Hybridisation and autoradiography ---------- *90*
2.3.4.2 SOUTHERN BLOTTING FOR *HERC2* ---------- 91

2.3.5 CLONING OF THE *P* GENE PROMOTER REGION ---------- **92**

2.3.5.1 IDENTIFICATION OF A BAC CLONE ---------- 92
2.3.5.2 SUBCLONING THE BAC ---------- 92
2.3.5.3 *P* GENE PROMOTER FRAGMENT PROBE ---------- 93
2.3.5.4 PREPARATION OF BAC DNA FOR SUBCLONING ---------- 93
2.3.5.5 PREPARATION OF THE CLONING VECTORS ---------- 94
2.3.5.6 LIGATION OF VECTOR AND INSERT ---------- 94
2.3.5.7 TRANSFORMATION ---------- 95
2.3.5.8 IDENTIFICATION OF POSITIVE TRANSFORMANTS ---------- 95

2.3.6 INVERSE PCR ---------- **96**

2.3.6.1 TEMPLATE DIGESTION ---------- 97
2.3.6.2 LIGATION ---------- 97
2.3.6.3 INVERSE PCR ---------- 97

2.3.7 THE *P* GENE AND TUBERCULOSIS ASSOCIATION ---------- **98**

2.3.7.1 DETECTION OF THE R305W POLYMORPHISM ---------- 99
2.3.7.2 *P* GENE INTRAGENIC MARKER SCREEN ---------- 99
2.3.7.2.1 Typing the microsatellite repeats ---------- *99*
2.3.7.3 STATISTICAL TESTS ---------- 100
2.3.7.3.1 The χ^2 Test ---------- *101*
2.3.7.3.2 Fisher's exact test ---------- *102*

CHAPTER 3 ---------- **103**

RESULTS AND DISCUSSION: OCA MUTATION SCREEN (*P* AND *TYR* GENES) ---------- **103**

3.1 RESULTS: OCA MUTATION SCREEN ---------- **104**

3.1.1 MUTATION DETECTION IN OCA SUBJECTS ---------- **104**

3.1.1.1 DETECTION OF THE 2.7 KB DELETION MUTATION ---------- 104
3.1.1.2 DETECTION OF THE R305W VARIANT ---------- 105
3.1.1.3 PCR-SSCP ANALYSIS ---------- 106
3.1.1.3.1 SSCP analysis for the P gene ---------- *109*
3.1.1.3.2 SSCP analysis for TYR ---------- *113*
3.1.1.4 SOUTHERN BLOTTING ---------- 114
3.1.1.4.1 Southern blotting for the P gene ---------- *114*
3.1.1.4.2 Southern blotting for HERC2 ---------- *117*

3.1.2 SEQUENCING OF THE *P* GENE PROMOTER REGION ---------- 118

3.1.2.1 SUB-CLONING THE BAC ---------- 118
3.1.2.1.1 BAC clone 263 022 appears unstable ---------- *119*
3.1.2.2 INVERSE PCR ---------- 120

3.2 DISCUSSION: OCA MUTATION SCREEN ---------- 122

3.2.1 MUTATION DETECTION IN OCA PATIENTS ---------- 122

3.2.1.1 THE *P* GENE ---------- 122
3.2.1.1.1 The R305W variant ---------- *122*
3.2.1.1.2 Detection of unknown mutations by SSCP analysis ---------- *123*
3.2.1.1.3 Mutation detection by Southern blotting ---------- *124*
3.2.1.2 EPISTASIS ---------- 124

3.2.2 SEQUENCING THE *P* PROMOTER REGION ---------- 125

3.2.2.1 CLONING ---------- 125
3.2.2.2 INVERSE PCR ---------- 127
3.2.2.3 SEQUENCE ANALYSIS OF THE 5' REGION OF THE *P* GENE ---------- 127
3.2.2.3.1 Promoter motifs ---------- *128*
3.2.2.3.2 Ensembl output (www.ensembl.org) ---------- *129*
3.2.2.3.3 MPromDb output (www.bioinformatics.med.ohio-state/MPromDb) ---------- *130*
3.2.2.3.4 Dottup output (www.ebi.ac.uk/emboss) ---------- *131*

3.2.3 POSSIBLE INVOLVEMENT OF *HERC2* IN PIGMENTATION ---------- 135

3.2.4 SOME THOUGHTS ON *TYROSINASE* ---------- 136

3.2.4.1 A CONFOUNDING FACTOR IN *TYR* MUTATION DETECTION – *TYRL* ---------- 136
3.2.4.2 *TYR* MUTATIONS ARE RARE IN BLACK INDIVIDUALS ---------- 136

3.2.5 FUTURE APPROACHES TO MUTATION DETECTION ---------- 137

CHAPTER 4 ---------- 139

RESULTS AND DISCUSSION: TB AND HOST GENETIC SUSCEPTIBILITY ---------- 139

4.1 RESULTS: TB AND HOST GENETIC SUSCEPTIBILITY ---------- 140

4.1.1 THE 2.7 KB DELETION MUTATION ---------- 140
4.1.2 THE R305W POLYMORPHISM ---------- 141
4.1.3 ANALYSIS OF 3 *P* GENE MICROSATELLITE MARKERS ---------- 143
4.1.3.1 D15S1533 (IVS17) ---------- *143*
4.1.3.2 D15S1536 (IVS20) ---------- *145*
4.1.3.3 D15S1537 (IVS24) ---------- *148*
4.1.4 CONSIDERING THE 5 MARKERS AS A HAPLOTYPE ---------- 150
4.1.5 SUMMARY ---------- 150

4.2 DISCUSSION: TB AND HOST GENETIC SUSCEPTIBILITY ---------- 152

4.2.1 OCA2 HETEROZYGOTES AND A SELECTIVE ADVANTAGE FOR TB ---------- 152
4.2.2 THE R305W POLYMORPHISM ---------- 153
4.2.3 ANALYSIS OF 3 *P* GENE MICROSATELLITE MARKERS ---------- 153
4.2.4 THE 5 MARKERS AS A 'HAPLOTYPE' ---------- 154
4.2.5 RECONCILING APPARENTLY CONTRADICTORY FINDINGS IN THE LITERATURE ---------- 155
4.2.6 THIRD WORLD RESOURCES ---------- 156

CHAPTER 5 ---------- 157

CONCLUSIONS ----------157
ELECTRONIC DATABASES ----------162
REFERENCES ----------162
APPENDIX----------180

LIST OF FIGURES

Figure 1.1 Schematic representation of the melanin biosynthesis pathway. 7
Figure 1.2 Diagramatic representation of a dendritic melanocyte.. 10
Figure 1.3 Photograph of the rock star, Bono, visiting a school in Lesotho 12
Figure 1.4 A Caucasiod individual affected with OCA1.. 14
Figure 1.5 The OCA2 phenotype with and without ephelides. 17
Figure 1.6 A black family with three children affected with albinism. 18
Figure 1.7 A black individual (right) affected with ROCA. 20
Figure 1.8. The p^{un} mouse. 31
Figure 1.9 Schematic representation of human chromosome 15q11-q13 47
Figure 1.10 Map of mouse chromosome 7. 54
Figure 2.1 Map of Africa showing countries from which patient samples were collected 78
Figure. 2.2. A schematic representation of the genomic region surrounding the human *P* gene. 81
Figure 2.3 Diagramatic representation showing the position of the PCR primers used to detect the *P* gene 2.7 kb deletion mutation 84
Figure 2.4 Diagram to illustrate the principle of the inverse PCR technique.....96
Figure 2.5 Nucleotide sequence of exon 1 and immediate upstream region of the human *P* gene 98
Figure 3.1 Agarose gel of PCR products detecting the 2.7 kb *P* gene deletion mutation 104
Figure 3.2 Detection of the R305W P gene variant 105
Figure 3.3 Examples of autoradiographs of SSCP gels 107
Figure 3.4 Sequencing of exon 1B of the *TYR* gene 108
Figure 3.5 Diagrammatic representation of the *P* gene cDNA showing positions of variants.....111
Figure 3.6 Autoradiograph of a Southern blot using 3 different enzymes to digest genomic DNA 115
Figure 3.7 Portion of an autoradiograph of a Southern blot used to screen OCA patients for *P* gene mutations 116
Figure 3.8 Autoradiograph of a Southern blot used to screen OCA patients for HERC2 mutations 117
Figure 3.9 Autoradiograph of a Southern blot where different enzymes were used to digest BAC DNA.....118
Figure 3.10 The BAC 263 022 appears to be unstable.....120
Figure 3.11 PCR fragments generated in an inverse PCR experiment 121
Figure 3.12 Ensembl annotation of a 20 kb region on chromosome 15q 129
Figure 3.14a Dot plot analysing a 20 kb region around *P* gene exon 1 132
Figure 3.14b Dot plot analysing the 2 kb region upstream of *P* gene exon 1 133
Figure 3.15 Nucleotide sequence data is now available for the 5'region of the *P* gene.....134
Figure 4.1 Microsatellite allele frequencies at the D15S1533 (IVS17) locus. 144
Figure 4.2. Microsatellite allele frequencies at the D15S1536 (IVS20) locus. 147
Figure 4.3. Microsatellite allele frequencies at the D15S1537 (IVS24) locus. 149
Figure 5.1 Host susceptibility to infectious disease may be considered a complex genetic trait 160

LIST OF TABLES

Table 1.1. Types of albinism.....11
Table 1.2. The estimated prevalence of albinism in different population groups 13
Table 2.1. Albinism subjects involved in a mutation screen for pathogenic variants. 77
Table 2.2. Subjects involved in a screen of the *P* gene for association with TB susceptibility. 80
Table 2.3. Primer sequences used to detect the *P* gene 2.7kb deletion mutation. 84
Table 2.4. Primer sequences for amplification of exon 16 of the human *P* gene.....85
Table 2.5. Primers used to amplify a 1kb fragment from IVS70 of the *HERC2* gene. 91
Table 2.6. Primer sequences used to amplify three repeat regions intragenic to the *P* gene.....100
Table 3.1. Results of the 2.7 kb deletion screen.....104
Table 3.2 Frequencies of the R305W variant in various population samples. 106
Table 3.3 Putative pathogenic *P* gene mutations identified in this study.....110
Table 3.4. Non-pathogenic sequence variants of the human *P* gene in black individuals. 112
Table 3.5. Non-pathogenic sequence variants of the human *P* gene in Caucasoid individuals. 113
Table 3.6 Variants in the *TYR* gene identified in this study in OCA individuals.....114
Table 4.1 Genotype frequencies of the 2.7 kb *P* gene deletion mutation.....140
Table 4.2 Allele frequencies of the 2.7 kb *P* gene deletion mutation 140
Table 4.3 Genotype frequencies for the R305W variant.....141
Table 4.4 Allele frequencies for the R305W variant. 141
Table 4.5 Genotype clusterings for the R305W variant.....142
Table 4.6 p values generated using Fisher's exact test on genotype clustering data.....142
Table 4.7. Allele frequencies D15S1533 (IVS17) 143
Table 4.8. *p* values for allele frequencies at D15S1533.....144
Table 4.9. Allele frequencies D15S1536 (IVS20) 146
Table 4.10. *p* values for allele frequencies at D15S1536.....147
Table 4.11. Allele frequencies D15S1537 (IVS24) 148
Table 4.12 *p* values for allele frequencies at D15S1537.....149
Table 4.13. Exact *p* values generated when performing pairwise comparisons between all 5 marker loci 150
Table 4.14 *p* values genetated using Fisher's exact test for pairwise comparisons for individual *P* gene loci 150

LIST OF ABBREVIATIONS

ά	alpha
β	beta
γ	gamma
λ	lambda
χ	chi
Σ	sigma
µg	microgram
µl	microliter
µM	micromolar
A	adenine
ACD	acid citrate dextrose
ASIP	human agouti signalling protein
ATP	adenosine-5'-triphosphate
BAC	Bacterial Artificial Chromosome
bp	base pair
BM	Boehringer Mannheim
BOCA	Brown Oculocutaneous Albinism
C	cytosine
cDNA	copy DNA
cm	centimetre
cM	centimorgan
°C	degrees centigrade
CpG	Cytosine phosphorylated Guanine
DHI	5,6 dihydroxyindole
DHICA	5,6 dihydroxyindole-2-carboxylic acid
DNA	Deoxyribonucleic acid
dCTP	deoxycytidine triphosphate
dNTP	deoxyribonucleic triphosphate
DOPA	3,4-dihydroxyphenylalanine
dNTP	deoxyribonucleotide-5'-triphosphate
E. coli	*Escherichia coli*
EDTA	Ethylene-diamine-tetra-acetic acid
et al.	*et alii* (and other people)
g	gram
G	guanine
HERC2	HECT domain and RCC-1 motifs
HIV	human immunodeficiency virus
IVS	intervening sequence
kb	kilobase
l	litre
M	molar
mA	milliamps
Mb	megabase
mg	milligram
min	minute

ml	millilitre
mM	millimolar
mm	millimetre
mRNA	messenger Ribonucleic acid
NaCl	sodium chloride
ng	nanogram
NHLS	National Health Laboratory Service
OCA	Oculocutaneous Albinism
P	Pink-eyed dilute
PCR	polymerase chain reaction
pmol	picomole
ROCA	Rufous Oculocutaneous Albinism
rt PCR	reverse transcriptase PCR
sec	second
SDS	sodium dodecyl sulphate
SNP	single nucleotide polymorphism
SSC	sodium citrate solution
SSCP	Single Strand Conformational Polymorphism
T	thymine
TB	tuberculosis
TE	tris-EDTA
TBE	tris borate EDTA
TYR	tyrosinase
TYRP	tyrosinase related protein
UV	ultraviolet
V	volts

CHAPTER 1

INTRODUCTION

1.1 Introduction

The different groups of humanity have long been an issue of significance with biologists, anthropologists, politicians and historians. Human beings have been grouped according to their cultural, linguistic, religious and biological differences. Classification of race has invariably involved one physical attribute – skin colour. Africans, Asians and Caucasians were considered separate entities. Anthropologists spent a great deal of time classifying human skull bones into racial groups by measuring the cephalic index, the ratio of length to breadth of the skull. The word 'Caucasian' derives from the region of the world from which it was claimed the skull which best represents the white-skinned people came, namely, the Caucasus Mountains in eastern Europe.

The idea that humans are divisible into racially distinct groups has had disastrous effects throughout history, perhaps on no continent more obviously so than Africa. To date, only maybe 100 genes out of the estimated 20 000 - 25 000 genes in the human genome have been shown to be associated with skin colour. The study of murine coat colour genes, more easily manipulated in the laboratory, has led to the identification of 127 pigment genes (reviewed by Bennett and Lamoreux 2003) and listed on the Albinism Database website, http://albinismdb.med.umn.edu/genes.htm). An obvious question arises – do genetic trends in skin colour parallel changes in the other tens of thousands of functional genes? DNA and protein studies show that differences in skin colour are not accompanied by consistent patterns of variation in other functional systems. Humans from different parts of the world show distinct patterns of genomic variation, especially at non-coding polymorphic loci, but the overall variation is not as great as early anthropologists may have predicted. At the molecular level, humans exhibit far less variation than might have been anticipated – the homogeneity of the human species is probably due to its relatively recent evolution.

Genetic research on human albinism has been central to many major discoveries in the field of human genetics. Albinism provided the first evidence to support the fact that Mendel's law of genetic segregation (first published in 1903) also applies to humans.

Much more recently, one of the human pigment loci (namely, the *P* gene region on chromosome 15q) has become implicated as a region likely to be involved in the complex

genetics of host susceptibility to infectious disease.

There is good evidence to support the hypothesis that the origin of modern humans, *Homo sapiens*, lies in Africa. This would suggest that our common ancestors were adapted to a hot and relatively dry climate with high levels of ionising radiation from the sun. Early humans were presumably darkly pigmented and those who migrated out of Africa had to adapt to climatic changes contributing to loss of pigmentation and resulting in the spectrum of pigmentary variation found in living peoples today. (This hypothesis has met with controversy - light versus dark skin may simply have resulted from a lifting of selective pressures for darker skin in Africa and random drift.) Several theories have been put forward to explain how and why colder climates were associated with selection for lighter skin colour. However, as reviewed by Jablonski (2004), none of them are entirely satisfactory.

Temperature and associated high levels of UV radiation have been argued against as being the selective agents in skin colour determination; dark skin colour does not protect against high temperatures. On the contrary, black colour absorbs more light than lighter colours. It has been suggested that dark skin colour protects against the damaging effects of UV radiation, the melanin providing a physical block preventing the UV light reaching the dermal layers. This would render individuals with a darker skin colour less susceptible to developing skin cancer. This may explain why skin cancers, including malignant melanoma and squamous and basal cell carcinoma, show an increased incidence with increased exposure to the sun, and are more prevalent in lightly pigmented individuals than darkly pigmented individuals. However, these types of cancers usually kill the individual only after reproductive age, making it an unlikely selective factor for skin colour in the context of climate. Diamond (2005) comments that this may be over-simplistic on the scientists' behalf as it ignores three very important historical human social structures – the long dependence of offspring on parents for learning and social status, the large contribution of hunter-gatherer grandmothers to their grandchildren and the dependence of a clan or village on its elders for knowledge in a preliterate society.

A second hypothesis concerns the ability of skin to synthesize vitamin D (Jablonski and Chaplin, 2000). Vitamin D is synthesized by the body in the presence of UV light. It is possible that the excessive amount of sunlight present in the tropics would lead to an over-

production of vitamin D by the body. The increased amount of melanin present in dark skinned individuals acts as a block to the UV light, reducing the level of radiation that can penetrate the skin. In cloudy climates, the presence of large amounts of melanin in the skin would be disadvantageous, as the pigment would block the already low levels of UV radiation, and inhibit vitamin D synthesis. Lack of vitamin D leads to poor bone development in children and permanent skeletal deformation, a condition known as rickets. In this way, the colder, cloudier climates of Europe would have selected for a lighter skin colour.

A third hypothesis involves the B vitamin folate. This vitamin, essential to health and especially important during early embryonic development, undergoes photolysis in the presence of excessive sunlight. It would then follow that individuals should be darkly pigmented to prevent UV breakdown of folate (Jablonski and Chaplin, 2000).

Jablonski and Chaplin (Jablonski 2004, Chaplin 2004, Jablonski and Chaplin 2000) argue that skin colour is determined by a combination of these factors – skin colour in a certain climatic region will evolve as a compromise between skins light enough to allow sufficient UV penetration for vitamin D synthesis, but dark enough to prevent excessive folate destruction (discussed in Diamond 2005).

While natural variation in skin colour obviously exists, the genes involved in pigment production have been found by looking at the 'unnatural' state - the absence or abnormal presentation of pigmentation. The study of coat colour in mice has provided a wealth of information to the field. Lack of pigmentation is termed albinism. Within the human condition, a range of severity is observed. Individuals affected with oculocutaneous albinism (OCA) type 1 or type 2 have a very severe phenotype, with little or no pigment production. Individuals affected with rufous oculocutaneous albinism (ROCA), classified as OCA type 3, have a very mild phenotype and are often unaware that they have any abnormality. The genes for OCA1 and OCA2, *tyrosinase* and *P* respectively, have been extensively studied due to the fact that mutations at these loci cause such significant pathology.

The genes for ROCA (tyrosinase related protein1) and a fourth type of OCA, OCA4 (membrane-associated transporter protein) have been identified. The genetics of other

single gene disorders e.g. Waardenburg Syndrome, Hirschsprung Disease, Hermansky-Pudlak Syndrome and Chediak-Higashi Syndrome have also been elucidated. These are multi-systemic disorders, where hypopigmentation is just one of the clinical phenotypes.

1.2 Plan of the thesis

Chapter 1 comprises a review of the literature and includes the aims of this project.
Chapter 2 presents the subjects involved in this study and describes the methodologies used. As this project involved two lines of research, the results and discussion for each part will be presented together for the sake of clarity. Chapter 3 comprises the results of the first part of the project, namely the search for pathogenic mutations in albinism subjects, and will include a discussion of these results. Chapter 4 comprises the results of the second part of the project, namely the investigation of the *P* locus as a possible TB susceptibility locus, and will include a discussion of these results. Concluding remarks are made in Chapter 5.

1.3 Melanogenesis

By definition, albinism is distinguished from other hypopigmentation disorders by the presence of eye involvement in affected subjects. Melanin is required during embryonic development for the normal development of the eye. Optic dysfunction is caused by abnormalities in the fovea (the fovea is hypoplastic), misrouting of the optic fibres, nystagmus and strabismus. Foveal hypoplasia leads to a marked reduction in visual acuity. Individuals with albinism typically have a visual acuity in the range of 20/200 to 20/100. Misrouting of the optic tract is one of the most striking characteristics of the albino eye (Creel *et al*. 1978).

In the biological spectrum there are five main pigments: hemoglobins, chlorophylls, carotenoids, flavanoids and melanins. Melanins are insoluable and have essentially defied analysis by classical techniques. Melanins are found throughout the plant and animal kingdoms, and fall into three major categories. Eumelanins are the brown and black pigments, pheomelanins are the red and yellow pigments, and plants contain certain brown and black pigments derived from phenols and catechols.

Pigment formation in mammals takes place in highly specialised cells called melanocytes, which contain unique cytoplasmic organelles, the melanosomes. Melanocytes are large, dendritic cells found in the dermal layer of the skin. Melanogenesis is a metabolic process whereby the substrate tyrosine is converted into the pigment, melanin. Both the amino acid, tyrosine (the precursor) as well as melanin (the end-product) are stable compounds, but the intermediates of melanogenesis are mostly unstable, reactive substances. These melanin intermediates are toxic to the cell and this is why melanin biosynthesis does not take place in the cytoplasm of the cell, but rather must occur in a membrane-bound organelle, the melanosome. If tyrosinase is erroneously compartmentalised by the cell, melanogenesis taking place at an incorrect location in the cell could lead to cell damage and death. Several key steps need to be achieved for normal melanogenesis: tyrosine must be taken up and compartmentalised within the melanosome, the melanosomes must synthesize and store the melanin, and the melanosomes must be successfully transferred through the dendrites from the melanocyte to the surrounding keratinocytes. One melanocyte will feed 36 keratinocytes. The keratinocytes, containing their melanin packages, will gradually move toward the surface of the skin, contributing to skin pigmentation. In hair bulbs, the melanosomes are transferred into the growing hair shaft. Thus, overall skin and hair pigmentation is determined by the size and shape of the melanosome, the type, amount and ratio of eu- to pheomelanin contained in the melanosomes that are transferred from the melanocytes (reviewed by Hearing, 1998).

Melanogenesis is dependant on the activity of the enzyme tyrosinase. Tyrosinase becomes activated through a glycosylation process that occurs in the Golgi-associated endoplasmic reticulum (Mishima and Imokawa 1983). The glycosylated enzyme is then transported, in coated vesicles, to the pre-melanosome. Here, tyrosinase catalyses the initial step in the formation of melanin – the formation of DOPAquinone from tyrosine. A third, downstream, catalytic activity of tyrosinase has been shown (Korner and Pawelek, 1982), namely, the oxidation of dihydroxyindol. Two other regulatory enzymes in the melanogenic pathway have been described, the tyrosinase-related proteins 1 and 2 (TYRP1 and TYRP2).

TYRP1 possesses dihydroxyindol carboxilic acid (DHICA oxidase) activity (Kobayashi *et al*. 1994). TYRP2 was shown to have DOPAchrome tautomerase (DT) activity isomerizing DOPAchrome to DHICA (Tsukamoto *et al.* 1992). TYRP1 has been shown to be

associated with the production of eumelanin rather than pheomelanin in human melanoma cell lines (DelMarmol *et al.* 1993) and is thought to facilitate the specific distal step in the production of brown-black pigments by oxidation of DHICA before the polymerisation takes place. The melanogenic pathway is depicted in Fig.1.1.

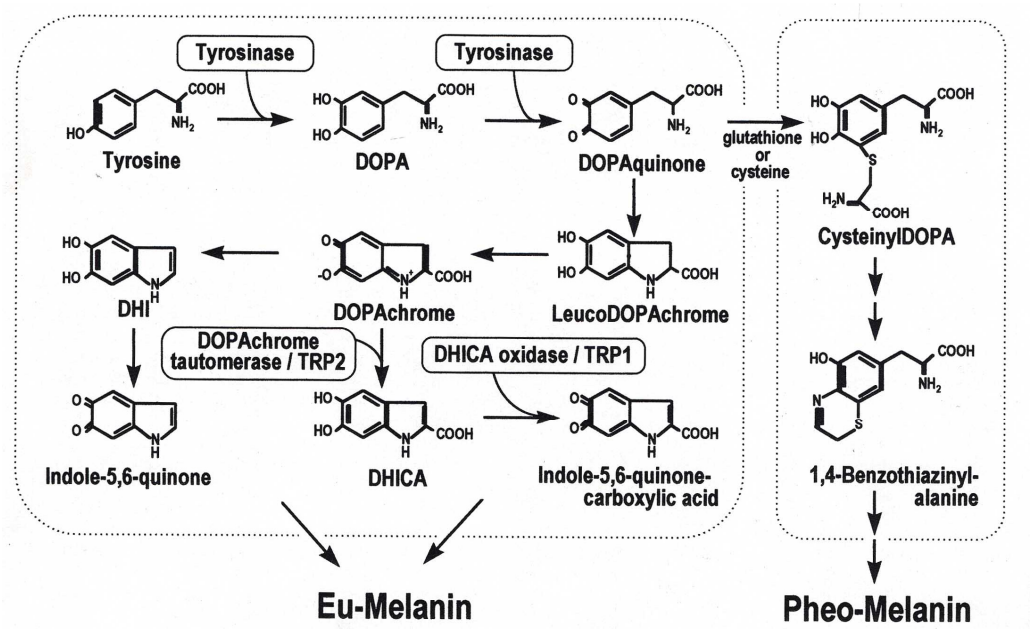


Figure 1.1 Schematic representation of the melanin biosynthesis pathway. Enzymes involved in the pathway for which the genes have been identified are encircled to highlight their point of action (Hearing 1998).

Eumelanogenesis results from the transformation of DOPAquinone to leucodopachrome and subsequent oxidation reduction reactions catalysed by post-DOPA oxidase regulators. Pheomelanogenesis begins with the conjugation of DOPAquinone to cysteine or glutathione which is converted to the final pheomelanin product via a series of chemical reactions. The relative amounts of eu– and pheo-melanin produced in an individual is determined by their genotype as well as the cellular environment.

1.3.1 Extracutaneous melanocytes

Melanocytes in the skin and the bulb region of the hair follicles of higher vertebrates and primates are responsible for making pigmented melanosomes which are transferred to the skin and hair keratinocytes and thereby give colouration to the epidermis and the hair shaft. Melanocytes have been found to occur at certain other sites, namely the eye, ear and leptomeninges of the brain. These populations of melanocytes are not identical to cutaneous melanocytes, but are morphologically and functionally similar (reviewed by Boissy 1998a).

Ocular melanocytes occur in the uvea and the retina of the eye and appear to be essential for the complete embryonic development of the neural retina and the neuronal network to the visual centre of the brain. They function to absorb light as it passes through the photoreceptor cells and also to transfer nutrients and waste to and from the retina. Melanocytes of the eye complete the process of melanisation shortly after birth and then remain melanogenically inactive throughout the life of the individual.

Otic melanocytes are found primarily in the connective tissue of the inner ear and in the cochlea (Meyer zum Gottesberge 1988). These melanocytes appear to function in the development of the endolymph fluid and it is proposed that melanin serves to protect against noise and/or toxin-induced damage by absorbing excess sound waves and toxins, preventing subsequent damage to hair cells (Barrenas and Lindgren 1990, Conlee *et al.* 1986).

The meninges covering the medulla oblongata and the upper cervical cord are pigmented due to the presence of melanocytes in this tissue (Goldgeier *et al*. 1984). As with the melanocytes of the uvea and the retina, these cells are melanogenically inactive in the adult. The function of the melanocytes in the leptomeninges is not clear.

1.4 Racial differences in skin architecture

The difference in the dark skin colour of Negroid and Australian aboriginal individuals, the intermediate skin colouring in Asians and Amerindians, and the lighter skin colouring of Caucasoids, is not attributable to a higher density of melanocytes in the darker skinned races. Rather, melanocyte distribution in the skin is consistent among different racial groups, but the type and distribution of melanosomes in the keratinocytes show distinct variation (Szabo *et al*. 1969). The melanosomes in the keratinocytes of Caucasoids and Asians are usually grouped together - they are seen in clusters which are surrounded by a membrane. Also, a granular substance is seen within the membrane-bound structure, between the melanosomes. In Asian skin, the melanosomes are more densely packaged within the membrane and there is less granular substance observed. In Negroid keratinocytes however, the melanosomes are larger and do not aggregate, but occur individually. Further, there are many more melanosomes in Negroid skin and they are also more melanised (i.e. darker) (See Fig 1.2.). The same situation is seen in aboriginal skin, with more melanosomes, larger in size and individually dispersed (Mitchell 1968). Interestingly, it has been shown that the normal racial distribution of melanosomes in keratinocytes can be altered by non-genetic factors such as ultraviolet light (Toda *et al*. 1972). Following exposure to a single dose of longwave radiation and application of trimethyl psoralen (a photosensitising chemical), skin biopsies taken from the hyperpigmented areas of Caucasoid skin showed non-aggregated melanosomes that were significantly larger than those in control tissue biopsied from the same individual.

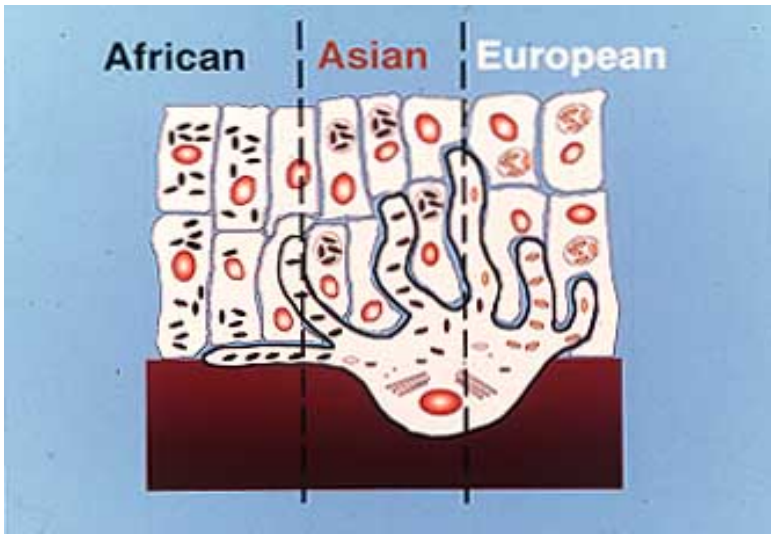


Figure 1.2 Diagramatic representation of a dendritic melanocyte transferring mature melanosomes to surrounding keratinocytes in the dermal layer of the skin (Sturm *et al*. 1998). The figure shows that skin colour differences between the races result from the number, type, level of melanisation and distribution of melanosomes within the keratinocytes.

1.5 Types of albinism

Albinism comprises a number of inherited disorders affecting normal pigmention. Lack of pigmentation may be localised to the eye (ocular albinism, OA) or may be seen in the eye, skin and hair (oculocutaneous albinism, OCA). While albinism was initially thought to be caused by different mutations in one gene, it is now apparent that the genetics of albinism are complex and the disorder shows locus heterogeneity. To date, nine types of OCA and four types of OA are described (summarised in Table 1.1). Type 1 OCA comprises disorders involving mutations in the *tyrosinase* gene (*TYR*) on chromosome 11, rendering *TYR* non-functional and therefore this type of albinism was initially referred to as tyrosinase-negative OCA. Mutations resulting in a complete lack of tyrosinase activity cause OCA type 1A (OCA1A). Mutations resulting in residual tyrosinase activity cause at least two additional types of OCA, OCA1B (yellow OCA) and OCA1MP (minimal pigment OCA). OCA1TS results from mutations which render the tyrosinase enzyme temperature sensitive. Type 2 OCA is a slightly milder phenotype and tyrosinase activity is normal (tyrosinase-positive OCA). OCA2 results from mutations in the *P* gene on

chromosome 15. Brown OCA (BOCA) is also linked to the *P* locus, and is therefore a subtype of OCA2. Red or rufous OCA (ROCA), OCA type 3, is caused by mutations in the *TYRP1* gene on chromosome 9. More recently, a new form of human OCA has been described, OCA type 4. The gene involved is *MATP* (human chromosome 5), the human homologue of the mouse underwhite gene. (The gene symbol for *MATP* is now *SLC45A2*.) An autosomal dominant form of OCA has also described. Two distinct forms of X-linked OA (OA1 and OA2), an autosomal recessive form (OA3) and ocular albinism with sensorineural deafness have been described.

Table 1.1. Types of albinism

TYPE OF ALBINISM	INHERITANCE	LOCUS	CHR	OMIM* No
OCA1		*TYR*	11q14-21	203100
- OCA1A (tyrosinase-negative)	autosomal recessive	*TYR*		
- OCA1B (yellow OCA)	autosomal recessive	*TYR*		
- OCA1MP (minimal pigment)	autosomal recessive	*TYR*		
- OCA1TS (temp sensitive)	autosomal recessive	*TYR*		
OCA2 (tyrosinase-positive)	autosomal recessive	*P*	15q11-13	203200
-BOCA	autosomal recessive	*P*		
OCA3 (ROCA)	autosomal recessive	*TYRP1*	9p23	203290
OCA4	autosomal recessive	*MATP*	5p	606574
Autosomal dominant OCA	autosomal dominant	unknown		none
OA				
- OA1 (Nettleship-Falls)	X-linked	*OA1*	Xp22	300500
- OA2 (Aland Island eye disease)	X-linked	*CACNA1F*	Xp11	300600
AROA	autosomal recessive		6q13-15	203310
OA and deafness	autosomal recessive	*TYR* and *MITF*	11q14-21; 3p14-12	103470

*OMIM reference number = Online Mendelian Inheritance in Man (www.ncbi.nlm.nih.gov)

OCA2 is by far the commonest form of albinism worldwide, due essentially to its high frequency in sub-Saharan Africa. OCA2 is the most common autosomal recessive genetic condition in black populations of sub-Saharan Africa (illustrated in Figure 1.3).

Figure 1.3 Photograph of the rock star, Bono, visiting a school in Lesotho (TIME April 2, 2007). In this picture, 27 random black high school students are captured on film, one of whom happens to be affected with albinism – an illustration of how common this disorder is amongst southern African black individuals.

(In other parts of Africa, sickle cell anaemia is the most common autosomal recessive disease, carrier numbers having reached such high frequencies due to heterozygote advantage in the presence of endemic malaria.)

The overall frequency of OCA in many population groups is about 1 in 20 000 (Witkop *et al*. 1989). The frequencies of the different types of OCA as they occur in given population groups is given in Table 1.2. OCA2 is far more common in black populations of African origin than in Caucasoid populations.

Table 1.2. The estimated prevalence of albinism in different population groups.

TYPE	POPULATION	PREVALENCE	REFERENCE
OCA2	African		
	Cameroon	1 in 35 000	Aquaron 1990
	- Bamileke, Cameroon	1 in 7 900	Aquaron 1990
	Nigeria	1 in 5 000	Barnicot 1952
	- Ibo, Nigeria	1 in 1 100	Okoro 1975
	South African blacks	1 in 3 900	Kromberg & Jenkins 1982
	Tanzania	1 in 1 400	Luande 1985
	Zimbabwe	1 in 4 200	Lund 1996
	European		
	Northern Ireland	1 in 10 000	Froggatt 1960
	Norway	1 in 9 650	Magnus 1992
	American		
	USA Caucasoid	1 in 36 000	Witkop *et al*. 1989
	USA African-American	1 in 10 000	Witkop *et al*. 1989
	USA Amerindian		
	- Zuni Indians	1 in 240	Witkop *et al.* 1972
	- Hopi Indians	1 in 227	Woolf and Grant 1962
	USA Brandywine isolate	1 in 85	Witkop *et al*. 1972
OCA1	USA Caucasoid	1 in 39 000	Witkop *et al*. 1989
	USA African-American	1 in 28 000	Witkop *et al*. 1989
BOCA	South African blacks	Unknown	Manga *et al*. 2001b
ROCA	South African blacks	1 in 8 500	Kromberg *et al*. 1990b
OA1	General	1 in 50 000	O'Donnell and Green 1989

1.5.1 Tyrosinase related albinism (MIM 203100)

Worldwide, the second most common form of albinism after OCA2 is OCA1. This type of albinism is caused by a loss of function of the enzyme tyrosinase. The condition was first diagnosed using the classic hair-bulb assay (King and Witkop 1976): the tyrosinase enzyme is localised to the melanocytes, thus the hair-bulb of a freshly plucked hair provides the most accessible tissue source. The hair-bulb is incubated in a high concentration tyrosine solution. Tyrosine (the precursor of melanin biosynthesis) will move into the hair-bulb and

if tyrosinase is present, melanin will be synthesised. A positive hair-bulb assay thus indicates the type of albinism is OCA2, while a negative result indicates OCA1. (Some forms of OCA1 with residual TYR function will show pigmented hair bulbs.) Molecular testing is now the method of choice to confirm a clinical diagnosis and to categorise the type of OCA. Individuals with OCA1 present at birth with marked hypopigmentation, the skin is milky-white, the hair white and the eyes light blue. The eyes may become a darker blue over a number of years, the skin may appear to gain colour over time, but never gains the ability to tan and will burn on exposure to the sun. Pigmented lesions like nevi, freckles or lentigines are rarely observed on the skin.

1.5.1.1 OCA1A: Tyrosinase-negative OCA

Individuals with OCA1A have no pigment in the skin, hair or eyes at birth, and pigmentation does not occur over time (see Fig 1.4). The ultrastructure of the melanocyte and the melanosomes is normal, but melanosomes do not reach maturity. Melanocyte distribution in the appropriate tissues is normal.

Figure 1.4 A Caucasiod individual affected with OCA1. Eye abnormalities are a phenotypic presentation, distinguishing albinism from hypopigmentation.

1.5.1.2 OCA1B: Yellow-mutant OCA

This type of albinism is caused by tyrosinase mutations which do not render the enzyme completely non-functional, so-called 'leaky' mutations. Some residual tyrosinase activity (5 - 10%) is apparent in the eye and most obviously, the hair. The hair becomes yellow in colour due to the production of pheomelanin. This type of albinism was first described in the Amish population of Indiana (Nance *et al.* 1970). Affected individuals show the extreme hypopigmentation present at birth characteristic of OCA1, but pigment will begin to accumulate during the first decade of life and the hair colour will become light yellow or blond by adolescence (the phenotype may closely resemble very fair individuals of Scandinavian origin). Affected individuals with near-normal amounts of skin and hair pigmentation may be erroneously diagnosed as having OA. There is a great degree of variability in the phenotype caused by different mutations resulting in more or less residual tyrosinase activity.

1.5.1.3 OCA1MP: Minimal-pigment OCA

This type of OCA1 is very similar to OCA1A with a severe hypopigmentation phenotype, but minimal pigment is seen to accumulate in the iris of the eye in the first decade of life. It is not certain whether or not this phenotype warrants a distinct classification, but evidence of a mouse phenotype would support such a distinction: an allele at the murine c locus (c^{44H}) results in residual tyrosinase activity and the accumulation of pigment in the eye only. This murine phenotype is termed dark-eyed albinism. The other possibility is that this phenotype falls into the range of yellow OCA, but the tyrosinase enzyme has even less activity than in yellow OCA. Affected individuals are often born with pigmented irides (or pigment will accumulate during the first decade) but the skin and hair is completely white. The hair may develop a slight yellow colour with time. So-called 'Platinum OCA' is most likely the same phenotype. From studies of hair-bulb tyrosinase activity in the obligate heterozygote parents of affected individuals, it would appear that OCA1MP results from compound heterozygosity for a severe OCA1 mutation (OCA1A or OCA1B) and a less severe mutation (King and Olds 1985).

1.5.1.4 OCA1TS: Temperature-sensitive OCA

The phenotype of this type of albinism is most obvious in the Siamese cat and the Himalayan rabbit where the body is hypopigmented but the extremities (the nose, ears, tail and paws) are pigmented. The tyrosinase enzyme is functional only at the slightly lower temperature present at the extremities and looses its function at the higher body temperature. A case has been documented in humans, but the condition is rare (King *et al.* 1991). The proband described in this study presented at birth with a phenotype consistent with that of OCA1A. However, at the time of puberty, the hair on her arms became light brown in colour and the leg hair turned dark brown. The hair on her head remained white, as did the skin, and the eyes remained blue. Analysis of tyrosinse activity in this individual showed that the enzyme lost its catalytic activity above 35°C. Ultrastructural analysis of melanocytes in hair-bulbs taken from scalp hair and then from leg hair showed that stage I and II melanosomes were present in melanocytes from the scalp, but stage III and IV melanosomes were present in the melanocytes of leg hair follicles.

1.5.2 *P* gene related albinism (MIM 203200)

1.5.2.1 OCA2

OCA2 is caused by mutations at the *P* locus on chromosome 15q. A somewhat milder phenotype than OCA1 is seen in individuals with OCA2: the eyes are usually pigmented (blue-grey or light brown in colour) and pigment will accumulate in the hair over time (pheomelanin) giving it a straw-coloured appearance. The skin is creamy white and cannot tan (see Fig 1.5). Patches of pigmentation may also develop on the sun exposed regions of the skin (freckles, nevi and lentigines). The ultrastructure and distribution of the melanocytes is normal (Witkop *et al*. 1973).

Black individuals in sub-Saharan Africa affected with OCA2 may present with an interesting phenotype. Approximately 61% of these OCA2 individuals will develop ephelides (Kromberg *et al*. 1989). Ephelides are darkly pigmented patches, similar to large freckles.

They are dendritic in conformation, with a central area of pigmentation and finger-like projections (see Fig 1.5). Ephelides are not present at birth, but develop from the second or third years of life. Ephelides usually only appear on the sun-exposed regions of the skin (the face, neck, forearms and hands), suggesting that they develop in reaction to UV light exposure. Individuals who develop ephelides seem to be protected, to some extent, against radiation damage caused by the sun since they have fewer skin cancers early in life (Kromberg *et al.* 1989).

Figure 1.5 The OCA2 phenotype with and without ephelides. The individual on the left is a black individual affected with OCA2. This individual has not developed ephelides despite age and sun exposure. The boy on the right is a black individual affected with OCA2 who has developed ephelides.

If more than one member of a family is affected with OCA2, they will either all develop ephelides or none of them will develop ephelides (Kromberg *et al*. 1989). This observation has led to the hypothesis that the development of ephelides has a genetic basis and is not a random occurrence. Initially it was proposed that the nature of the mutation in individuals who develop ephelides might be similar to the mutation in the p^{un} mouse (Ramsay *et al*. 1992). By this model, the OCA2 mutation would show a localised reversion to wild-type in the area comprising the ephelus, and normal pigment production could occur. However, with the elucidation of the common African OCA2 mutation (Durham-Pierre *et al.* 1994), the 2.7 kb deletion mutation, and its subsequent description in the homozygous state in

OCA2 individuals with ephelides (Stevens *et al.* 1995), it is now obvious that this is not the molecular mechanism which allows for the development of the ephelides phenotype.

1.5.2.2 BOCA: Brown OCA

It was originally presumed that this type of albinism occurred only in black Africans and African-Americans, but Caucasoid individuals with this form of albinism have been described (King *et al.* 1980, 1985a). Amounts of eumelanin in the skin and eyes are reduced but not absent. The phenotype in affected individuals from the black population is one of light brown skin, darker coloured hair and light brown or grey eyes (see Fig 1.6). Affected individuals present with the same ocular defects as seen in other types of albinism. The Caucasian phenotype was described as being essentially similar to OCA2, with white skin and golden blond hair. At the time the gene defect involved in BOCA was not known. The brown locus has been identified in humans, and BOCA is linked to the *P* locus on chromosome 15q (Manga *et al*. 2001b). Thus OCA2 and BOCA are allelic, with BOCA possibly being caused by milder mutations in the *P* gene. As these conditions are caused by mutations at the same locus, it is possible to have individuals affected with either BOCA or OCA2 in the same family.

Figure 1.6 A black family with three children affected with albinism, one affected with BOCA and two with OCA2. The boy on the left is affected with the slightly milder BOCA phenotype. Two sibs have classic OCA2, while the oldest sister (background) and father (right) are unaffected.

If parents have children some of whom are affected with BOCA and some of whom are affected with OCA2, then one parent at least needs to be affected with BOCA. In the family photographed above, it must be assumed that the mother was affected with BOCA, carrying one severe and one mild *P* gene mutation. The father is normally pigmented, but a carrier of OCA2, therefore carrying one severe *P* gene mutation.

Molecular *P* gene mutation analysis in 10 BOCA individuals showed 9 to be compound heterozygotes for the severe 2.7 kb deletion mutation and another unidentified mutation (Manga *et al*. 2001b). Despite extensive further mutation analysis, a second *P* gene mutation could not be identified in any individual. Haplotype studies showed several BOCA-associated haplotypes exist, making a second common mutation unlikely, although the haplotypes could be divided into two core groups, suggesting a limited number of *P* gene mutations give rise to this phenotype (Manga *et al*. 2001b). It is proposed that the second mutation in these individuals is likely to be a milder, 'leaky', mutation.

Boissy *et al*. (1996) described an African-American patient with the BOCA phenotype. On melanocyte culture, it was found the cells contained no detectable amounts of TYRP1 using various anti-TYRP1 probes. Mutation analysis showed the patient was homozygous for a 1 bp deletion in exon 6 of the *TYRP1* gene. These authors (Boissy *et al*. 1996) concluded that this mutation affects the ability of the *TYRP1* gene to interact with tyrosinase, leading to dysregulated tyrosinase activity which promotes the synthesis of brown melanin as opposed to black melanin. This paper classified BOCA as OCA type 3 (in agreement with the discussion above); however, this patient having a different gene affected, *TYRP1* as opposed to *P*, has led to some controversy.

1.5.3 Other forms of albinism

1.5.3.1 OCA3 (Rufous OCA, MIM 203290)

Rufous oculocutaneous albinism has been described in the native population of Papua New Guinea (Walsh 1971), families in the northwest of Puerto Rico (King *et al*. 1978, 1986) and black Africans (Kromberg *et al*. 1990). Individuals with ROCA have a reddish-bronze skin

colour, ginger coloured hair (which is lighter in colour than the skin) and the eyes are blue or brown (see Fig 1.7).

Figure 1.7 A black individual (right) affected with ROCA standing next to a normally pigmented individual.

The visual anomalies of OCA3 are often mild but nystagmus and misrouting of the optic tract have been described in certain individuals (Kromberg *et al*. 1990), lending support to the delineation of this phenotype as a form of OCA. The prevalence of ROCA in southern Africa is approximately 1 in 8 500 (Kromberg *et al*. 1990). ROCA is caused by mutations at the tyrosinase-related protein 1 (TYRP1) locus, and two common mutations have been described (Manga *et al*. 1997). The nonsense mutation, S166X, and a second mutation, 368delA, account for 45% and 50% respectively, of ROCA chromosomes in southern Africa.

1.5.3.2 OCA4 (MIM 606574)

Due to the observation that certain individuals with OCA were not found to have either *TYR* or *P* gene mutations, it was proposed that other form/s of OCA may exist (Passmore *et al*. 1999). Researchers in the laboratory of Murray Brilliant chose the human homologue of

the mouse gene, underwhite (*uw*), as a good candidate for being involved in another form of human albinism, and cloned the gene at the mouse *uw* locus (Newton *et al*. 2001). The mouse *uw* gene encodes a protein that is predicted to span a lipid bilayer and shows highest homology to sucrose-transporter proteins found in plants. The gene has been renamed *matp* (for membrane-associated transporter protein). The human homologue was subsequently identified through database searches and alignments (Newton *et al.* 2001).

The human gene, *MATP* (now called *SLC45A2*), is located on chromosome 5p. A group of 102 individuals with hypopigmentation were screened for mutations in the *MATP* gene. Mutations were identified in one individual, of Turkish descent. His phenotype most closely resembled that of an OCA2 individual but he had no detectable *P* gene mutations. A study of 75 unrelated albinism patients in Japan revealed that 18 of them (24%) carried *MATP* mutations, making this the most common locus for *tyrosinase*-positive OCA in Japan (Inagaki *et al*. 2004). A study on 176 German patients with albinism identified 5 individuals (3%) with OCA4 (Rundshagen *et al.* 2004). Like the P protein, MATP has been shown to have 12 transmembrane domains and is also predicted to function as a transporter. Work in mice has shown that tyrosinase is synthesised at normal levels in uw/uw melanocytes, but activity levels are only about 20% of that found in wild-type melanocytes (Costin *et al.* 2003). Further, in individuals with OCA4, melanocytes secrete melanosomes which are mostly immature. These authors propose that tyrosinase processing and trafficking to the melanosome is disrupted by the aberrant MATP transporter and that the tyrosinase enzyme is abnormally secreted from the melanocytes in the early stage melanosomes.

1.5.3.3 Autosomal dominant OCA

While families have been described where albinism seems to be segregating as an autosomal dominant trait (Bergsma *et al*. 1974, Fitzpatrick *et al*. 1974), it is not certain whether this is a true form of OCA. The family described by Bergsma *et al*. showed affected parents and children, but normal grandparents. The albinism in this family is possibly exhibiting a pseudo-dominant pattern of inheritance (King *et al*. 1995). The other family (Fitzpatrick *et al.* 1974) does not meet the most definitive criterion for OCA as there is no eye involvement.

1.5.3.4 Ocular albinism

Ocular albinism (OA) is far less prevalent than either OCA1 or OCA2. While the phenotypic manifestation is hypopigmentaion of the eye, ultrastructural changes of the skin and hair are present.

X-linked recessive ocular albinism (OA1, MIM 300500)

This type of albinism is also called Nettleship-Falls ocular albinism. Affected males show all the ocular features of albinism (Charles *et al*. 1993) with the eyes being blue to brown in colour. The skin and hair are normally pigmented. Heterozygous females do present with clinical features in 80% of cases (Charles *et al.* 1992) with pigment changes in the eye being apparent. This is due to X-chromosome inactivation patterns. In certain cases, heterozygous females will present with the ocular albinism phenotype, presumably due to non-random X-inactivation. The melanocytes in OA are normal, but abnormalities in the melanosomes are evident: very large melanosomes are present in the melanocytes, not only of the iris and retina, but also in the skin and hair follicles. The presence of these "giant melanosomes" in multi-systemic organs suggests this disorder may in fact be a type of OCA, with major manifestations in the eye. The gene for X-linked OA was mapped to Xp22.3-22.2 (Schnur *et al*. 1991) and has subsequently been cloned (Bassi *et al.* 1995).

X-linked recessive ocular albinism (OA2, MIM 300600)

A large kindred from the Aland Islands was described (Forsius and Efiksson 1964) with males affected with a type of OA considered distinct from OA1. Patients presented with the severe eye anomalies characteristic of OA1 as well as myopia, astigmatism and colourblindness. Female carriers did not exhibit the complete pigment deficiency characteristic of OA1. Van Dorp *et al.* (1985) showed there was no misrouting of the optics pathways in affected males. A study by Hawksworth *et al.* (1995) on a Welsh pedigree with X-linked congenital nystagmus and manifestation in some females, proposed a clinical diagnosis of either Aland Island eye disease (AIED) or congenital stationary night blindness. The genes for both conditions had been previously mapped to proximal Xp, leading these authors to suggest these might be the same disease. Molecular analysis of a Finnish family with AIED identified mutations in the CACNA1F gene which maps to Xp11.23-11.22 (Wutz *et al.* 2002).

Autosomal recessive ocular albinism (OA3, MIM 203310)

Four families were described (Witkop *et al*. 1978) where males and females were equally affected with ocular albinism. Several isolated cases of affected females were also described. Affected individuals present with ocular albinism and normal cutaneous pigmentation. All the cases described to date have been amongst Caucasoids, the clinical phenotype has not been found in black African or African-American subjects (King *et al.* 1995). It has been proposed that the locus for AROA lies on chromosome 6q13-q15 (Rose *et al.* 1992). These authors studied a dysmorphic boy who had the features of OA and was found to carry a cyotogenetic deletion of chromosome 6q. Ultrastructural analysis of the cutaneous melanocytes showed an absence of giant melanosomes. It is thus debatable whether this individual had OA or not, the hypopigmentation may be simply part of the chromosome abnormality syndrome.

Like autosomal dominant OCA, it is thus still debatable whether this is a true distinct form of albinism, and the clinical manifestations observed may be due to mild or leaky mutations in either tyrosinase or the *P* gene. A 7 year old girl of northern European origin, initially thought to have AROA, was investigated at the molecular level and found to be a compound heterozygote for two *P* gene mutations (Lee *et al*. 1994a), lending support to the latter theory.

Ocular albinism with sensorineural deafness (MIM 103470)

Lewis (1978) described 7 affected males and 5 affected females in a large pedigree of Caucasian origin. Patients presented with classic OA plus deafness. In a family initially presented by Bard (1978) and then later further investigated by Morell *et al.* (1997), sibs with heterochromatic irides were described, as well as an individual with a prominent white forelock, characteristic of Waardenburg syndrome. Morell *et al.* (1997) undertook molecular analysis in this family and found affected individuals to show an apparent ‘digenic’ pattern of inheritance – the phenotype resulting from mutations in 2 distinct genes. Patients were heterozygous for a mutation in the *MITF* gene (which, when mutated, causes Waardenburg syndrome) and either heterozygous or homozygous for a variant in the

tyrosinase gene, R402Q. This *TYR* polymorphism is known to be associated with reduced enzyme activity (Tripathi *et al.* 1991).

1.6 Other hypopigmentation disorders

Melanocytes are embryonically derived from the neural crest. Melanoblast cells will migrate out of the dorsal surface of the neural tube along its entire length, and migrate to several target sites of the developing embryo. Melanoblasts will migrate to areas in four organ systems: the skin, the eye, the ear and the brain. Melanoblasts move to the basal epithelium of the epidermis and the hair bulbs in the skin, to the uveal tract of the eye, to the stria vascularis, the vestibular organ and the endolymphatic sac of the ear, and the leptomeninges of the brain. Once at these sites the melanoblast will differentiate into the melanin-producing cell, the melanocyte. Melanocytes in adult tissue rarely divide or migrate. Further, while melanocytes in the skin and hair bulbs remain melanogenically active, those in the extracutaneous sites become melanogenically inactive for the life of the individual. As many gene products are made and used by the melanocyte as it differentiates and migrates to its target site, there are consequently numerous points at which a mutation in a given gene may disrupt the normal melanogenesis process and lead to a hypopigmentation disorder.

1.6.1 Neural Crest Defects

Since the time that Darwin observed deafness in white, blue-eyed cats a complex picture has emerged of the association between embryonically derived neural crest cells and the presence of melanin pigmentation. Mutations affecting the early stages of melanoblast migration result in failure of these cells to reach their target tissue and the disorders which result present with de-pigmented lesions which contain no melanocytes.

1.6.1.1 Waardenburg Syndrome (WS1 and WS2: MIM 193500; WS3 MIM 148820)

Individuals with Waardenburg syndrome present with patches of cutaneous depigmentation, which are variable in size and number. The irides of the eye are partially

or unilaterally different in colour and deafness, of varying severity, results from the reduction or absence of melanocytes in the cochlea. Mutations in the *PAX3* gene lead to Waardenburg syndrome type 1 and type 3. *PAX3* maps to human chromosome 2q35-37.3 (Tassabehji *et al.* 1992, Hoth *et al.* 1993). These disorders are inherited as autosomal dominant traits. Interestingly, homozygotes present with a more severe phenotype than the heterozygotes, suggesting a dosage effect of the gene product. The *PAX3* gene encodes a transcription factor with three functional domains mediating DNA binding, and a transcription activation site (Walther *et al.* 1991, Wilcox *et al.* 1992). In some cases, but not in most, Waardenburg syndrome type 2 is caused by mutations in the micropthalmia-associated transcription factor gene, *MITF*, (Hughes *et al.* 1994, Tassabehji *et al.* 1994) on human chromosome 3p12. WS2 patients who do not carry *MITF* mutations have mutations in some, as yet undiscovered, gene/s. *MITF* is the human homologue of the mouse *micropthalmia* locus. The MITF protein functions as a transcription factor and has been shown to bind to tyrosinase and TYRP1 (Yasumoto *et al.* 1995), activating these proteins. In this way MITF is thought to upregulate melanin synthesis. In the mouse, mutations affecting different functional domains dictate whether the disease is inherited in a recessive or a dominant fashion. In the human system, all *MITF* mutations show an autosomal dominant mode of inheritance.

1.6.1.2 Hirschsprung Disease (MIM 142623)

Hirschsprung disease presents with congenital megacolon resulting from the absence of the neural crest derived cells in the smooth muscle of the intestinal wall. In a proportion of cases, hypopigmentation occurs together with heterochromatic irises and deafness. The disease shows locus heterogeneity as the phenotype can arise from mutations in several different genes, operating alone or in combination with each other (see OMIM or GeneCards for overview). However it appeared that clinically mutations at one locus were neither necessary nor sufficient to cause disease. Carrasquillo *et al.* (2002) conducted a genome-wide scan looking for areas of linkage disequilibrium due to common ancestry in a cohort of Mennonite families. They identified 3 areas of susceptibility – on chromosomes 10q11, 13q22 and 16q23. They showed the gene at 13q22 is *EDNRB* and the gene at 10q11 is *RET*. Mouse crosses between *ret* null and *ednrb* mice showed non-complementation and joint transmission. Also, joint transmission of *RET* and *EDNRB* alleles in affected

Hirschsprung disease patients lead these authors to suggest an epistatic effect between the two loci leading to disease presentation.

1.6.1.3 Piebaldism (MIM 172800)

Piebaldism is an autosomal dominant disorder characterised by white patches of skin (predominantly on the face, chest and abdomen) and hair, which entirely lack melanocytes. The distinctive phenotype allowed this disorder to be one of the first genetic traits recognized to exhibit autosomal dominant inheritance (Morgan 1786 as cited in Spritz 1992). Pedigrees have been described where a proportion of the affected individuals also present with deafness (Reed *et al.* 1967, Comings and Odland 1966 as cited in Spritz and Beighton 1998) but this is not common and it is generally held that the defects in this condition are essentially cosmetic. Like vitiligo, the patches of affected skin are characterised by a reduction in the number of melanocytes, down to complete absence. Mutations in the *KIT* gene on chromosome 4q12 have been shown to result in piebaldism (Fleischman *et al.* 1991, Giebel and Spritz 1991b, Spritz *et al.* 1992). The murine phenotype of piebaldism is "dominant white spotting". The KIT protein functions as a plasma membrane receptor. After binding with the ligand, the receptor initiates a chain of events which upregulate melanocyte proliferation (Spritz *et al.* 1994). Different types of mutations lead to milder or more severe phenotypes, as do mutations in different positions in the gene (Spritz *et al.* 1992, reviewed in Boissy and Nordlund 1997). Certain heterozygous missense mutations have a "dominant negative" effect in that the mutant dominant allele interferes with the action of the wild-type allele resulting in as much as a 75% loss of normal receptor function (Spritz *et al.* 1992 a and b).

1.6.2 Organelle Defects

After the melanoblasts leave the neural crest and migrate to specific target sites, they will complete their differentiation. Here defects within the melanocyte resulting in an inability to produce pigment result in oculocutaneous albinism (OCA). A further group of disorders involve not only defective melanin production by the melanosome, but also abnormal organellogenesis in other cell types. Disorders of cytoplasmic organelles are exemplified by Hermansky-Pudlak syndrome and Chediak-Higashi syndrome.

1.6.2.1 Hermansky-Pudlak Syndrome (MIM 203300)

The Hermansky-Pudlak syndrome (HPS) is characterised by three distinguishing phenotypes: oculocutaneous albinism, bleeding diathesis due to a deficiency in the platelets and a lysosomal ceroid storage disease. Three different organelles (melanosomes, platelet dense granules and lysosomes), in different cell types, are seen to be affected. Melanocytes in the hair and skin are normal, but lack fully developed melanosomes, only stage I and II melanosomes are found. The disorder was found to be particularly common in two inbreeding population isolates, one in the Swiss Alps and the other in Puerto Rico. (HPS is probably the most common single gene disorder in Puerto Rico with an estimated carrier frequency of 1 in 21 due to founder effect (Wildenberg *et al.* 1995). These populations were used in the mapping of the first locus for HPS to chromosome 10q23 (Fukai *et al.* 1995). The gene at this locus has since been cloned (Oh *et al.* 1996). It has been shown that the syndrome shows locus heterogeneity as it can be caused by mutations in several genes, seven loci have been identified to date (*HPS1* – *HPS7*), all mapping to different chromosomes.

1.6.2.2 Chediak-Higashi Syndrome (MIM 606897)

The Chediak-Higashi syndrome (CHS) is a rare autosomal recessive disorder characterised by incomplete oculocutaneous albinism, recurrent and severe infections, a bleeding tendency and enlarged granules in many cell types. Most patients die in childhood. Giant inclusion bodies in peripheral blood cells are diagnostic of the disease, and melanocytes contain giant hypomelanotic melanosomes (about five times their normal size). These large melanosomes cannot be transferred to the keratinocytes. The gene for CHS has been cloned in human (Barbosa *et al.* 1996, Nagle *et al.* 1996) and mouse (Barbosa *et al.* 1996, Perou *et al.* 1996). The human *CHD* locus is on chromosome 1q43 and is the homologue of the murine *beige* locus.

1.6.3 Acquired hypopigmentation defects

Many agents – chemical, pharmacological, infectious – may induce areas of

hypopigmentation on the skin and/or hair. Two conditions, namely leprosy and tuberculosis, will be discussed due to their relevance to this study.

1.6.3.1 Leprosy (MIM 246300)

In 1882 Robert Koch identified the tubercule bacillus, *Mycobacterium tuberculosis*, and this became established as the 'typical' mycobacterium. In the time following this discovery, other mycobacteria were described and these were consequently classed as 'atypical' mycobacteria. There are 10 species of atypical mycobacteria described (discussed by Meyers 1991) which cause infections of the skin. One of these organisms is the bacterium, *M. leprae*.

M. leprae causes leprosy, a chronic disease affecting the cooler parts of the body: the skin, upper respiratory tract, superficial portions of peripheral nerves and the testes. Skin lesions are the earliest manifestation of the disease, and these lesions are often hypopigmented. In tuberculoid leprosy, a subtype of leprosy, these lesions have sharply demarcated edges and they range in size from 1 cm in diameter to lesions that cover an entire area of the body, e.g. a cheek or thigh. Major clinical symptoms are also observed in the nervous system. In lepromatous leprosy, appropriate immune responses from the body fail, and bacteria spread through the skin to the peripheral nerves. Deeper tissues do not become affected because bacteria reproduce optimally at temperatures between 27°C and 30°C. Pain in the peripheral nerves is experienced with a loss of sensation, usually in the hands and feet. Leprosy affects approximately 15 million people worldwide (World Health Organization), a large number (50%) of whom live in Africa and India. Leprosy is an infectious disease spread most commonly by skin-to-skin contact, nasal secretions, mother's milk and possibly placental transmission from the mother to the fetus. Interestingly, animals have been found affected with leprosy, and the bacteria isolated from three different animal species is indistinguishable from the human pathogen. Naturally acquired leprosy has been described in armadillos in Louisiana and Texas, in a chimpanzee from Sierra Leone and in a mangabey monkey from West Africa. The armadillo population in Louisiana has received special attention, with up to 10% of the wild population being infected with leprosy. Susceptibility of the armadillo is probably due, in large part, to the temperature of the central body portion of the animal which is 32-35°C.

Like tuberculosis, susceptibility to leprosy has been shown to have a genetic component. The disease will only manifest clinically in a small subset of individuals who have been exposed to the pathogen. Familial aggregation is evident, such that distribution of the disease among sibs is not random. A case study undertaken in Papua New Guinea (Shields *et al.* 1987) found that the disease showed family clustering even though the social unit of the population was the community and not the family. A study by Kaur *et al.* (1997) looked at two polymorphic loci in a sample of individuals from New Dehli. These two loci are on human chromosome 2q31-33, a region showing synteny to the area of the mouse genome where the *Bcg* locus is found. The murine *Bcg* locus influences susceptibility to intracellular parasites. An allele of one of these loci, the 104 bp allele of the *CTLA4* locus, was not found in any of the leprosy patients studied (N=25), leading these authors to conclude that the 104 bp allele is correlated with host resistance to developing leprosy, and absence of this allele may be a possible risk factor for the disease. Using a cohort of families with affected sib-pairs from south India, Siddiqui *et al*. (2001) found significant linkage (lod score=4.09, p=0.00002) to chromosome 10q13. Mira *et al*. (2003) described a genome-wide search for susceptibility loci to leprosy in a southern Vietnamese population and found a susceptibility locus on chromosome 6q25.

1.6.3.2 Tuberculosis

While hypopigmented patches in leprosy have been well documented, infection with the mycobacterium M. tuberculosis has, very interestingly, also been shown to alter cutaneous pigmentation. A field doctor working in Nigeria observed that African patients infected with tuberculosis (TB) presented with a 'relative hypochromia', particularly of the face (Pearson 1978, Pearson 1989). Melanocytes in the face are twice as numerous per square millimeter of skin, as compared to any other part of the body, hence it follows that a change in melanin concentration would be more apparent in the face/cheeks. Scientific investigation showed that the hypochromia was significantly different when TB patients were compared to a control group of individuals, matched for ethnicity, socio-economic background, sex and age. Re-pigmentation resulted upon successful treatment of the tuberculosis (within 3-6 weeks of commencement of antibiotic therapy).

Many genes have now been described which may work singly but probably work in combination with each other, to determine susceptibility to TB. Gene variants at loci such as *HLA*, *NRAMP1*, Vitamin D receptor and the mannose-binding protein (*MBL2*) have all been investigated for a possible role in TB susceptibility (discussed further at the end of this chapter).

1.7 Mouse pigment loci

Several mouse coat colour mutations were documented as early as the late 19th century. Initially, mice were bred purely for aesthetic purposes, and the breeder's hobby known as the "mouse fancy" was established. Scientific experiments on the genetics of coat pigmentation began in the first two decades of the 20th century. Early genetic experiments looked at mouse coat colour because of the obvious phenotype available for analysis. Consequently, a wealth of knowledge is now available and the genetics of pigmentation is the best understood complex developmental system in the mouse. To date, at least 127 loci affecting coat colour in the mouse have been described (reviewed by Bennett and Lamoreaux, 2003). Only about 60 of these genes have been cloned and all of them appear to have human orthologues. Six of those most relevant to the present study are reviewed below:

1.7.1 The pink-eyed dilution locus (*p*)

Mutations at the pink-eyed dilution (*p*) locus in mice cause a hypopigmentation phenotype affecting both the coat and the eyes. The pink-eyed dilution phenotype is one of the oldest described mouse variants and is caused by mutations at the *p* locus. The *p* locus is located on mouse chromosome 7 (a region showing synteny to human chromosome 15q). At least 100 alleles are known to occur at this locus but many of these did not arise spontaneously and are radiation-induced deletions (Lyon *et al.* 1992). Mutations at this locus result in a range of phenotypes in the mouse, ranging from slightly lighter coat colour with dark eyes, to the classic white (albino) mouse with pink eyes. The gene at the *p* locus has been cloned, and is termed the *p* gene (Gardner *et al.* 1992). The human homologue, the *P* gene, is responsible for the oculocutaneous albinism type 2 (OCA2) phenotype (Rinchik *et al.* 1993) and will be discussed below in further detail.

1.7.1.1 The p^{un} allele

One of the mutant alleles at the *p* locus is the spontaneously arising pink-eyed unstable (p^{un}) variant (Melvoid 1971, Gondo *et al*. 1993). The p^{un} allele constitutes a 70 kb head-to-tail duplication of genomic DNA within the *p* gene. The duplication includes exons 6-18, with breakpoints in introns 5 and 18 (Oakey *et al*. 1996) and results in a larger mRNA transcript and a non-functional protein. Unlike other *p* alleles, this mutation is associated with a very high reversion frequency. The p^{un} allele exhibits the highest reversion rate reported in mammals (discussed in Brilliant *et al*. 1991). Up to 3.5% of offspring from the cross of two p^{un} homozygous mice will show patches of wild-type coat colour (Melvoid 1971) (See Fig 1.8).

Figure 1.8. The p^{un} mouse. Dark patches of coat colour represent areas of spontaneous reversion to wild type.

This phenotype is due to somatic reversion of the mutation - the duplicated sequence is lost in the cells where pigment can be produced. Tandem head-to-tail duplications (but not inverted duplications) allow for the loss of one of the duplicated copies by homologous, but unequal cross-over during meiosis.

1.7.2 The agouti (*a*) locus

The agouti mouse shows a characteristic hair banding pattern: alternate bands of black and yellow along the hair shaft. The mouse agouti locus (*a*) has been mapped to chromosome 2 and the gene cloned (Bultman *et al.* 1992). The gene spans 18kb of genomic DNA and the

4 exons encode a mRNA transcript of 0.8 kb. The gene is not expressed in melanocytes. The agouti protein acts as a ligand for the melanocyte stimulating hormone receptor (MC1R), resulting in competitive inhibition between the agouti protein and melanocyte stimulating hormone (MSH). Over 18 alleles at this locus have been described, the wild type allele (*A/A*) gives the typical banded pattern while the homozygous mutant mouse (*a/a*), *nonagouti*, has only black hair. There is a *lethal yellow* allele (A^y) that when present in the heterozygous state (A^y/a) results in an all yellow coat, but is lethal in the homozygous state (A^y/A^y). Heterozygosity of the A^y allele is associated with obesity and sterility. The A^y mutation is not within the coding region of the agouti gene. Rather, the phenotype results from a DNA structural alteration that disrupts a gene, *raly*, which is located 5' to the *agouti* locus. On A^y chromosomes, the coding sequence of the *raly* gene is deleted and the *agouti* locus is placed under the control of the *raly* promoter, resulting in deregulated over-expression of the agouti gene (Michaud *et al.* 1993). These authors suggest that the homozygous lethality in A^y/A^y mice may be due to lack of r*aly* expression in the developing embryo as opposed to aberrant agouti expression. It should be noted that the alleles at this locus exert an effect on the phenotype only in the hair follicles. The type of pigment produced at other locations in the body is always eumelanin, regardless of the agouti genotype.

1.7.3 The albino (*c*) locus

The albino locus was thought for a long time to be the structural gene for the enzyme tyrosinase, a key catalytic enzyme in the melanin pathway. In 1987, Kwon *et al.* showed that tyrosinase cDNAs mapped to the *c* locus (Kwon *et al*. 1987), and it was subsequently shown that mutations in the coding region of *Tyr* inactivate the protein (Shibahara *et al.* 1991). Inactivation of tyrosinase leads to complete loss of all pigment production in the melanocyte, and the phenotype is a completely white coat colour.

During the 1980s when researchers were attempting to clone the tyrosinase gene, two proteins with high homology to tyrosinase were isolated as incidental findings. The melanocyte-specific gene products tyrosinase, b-locus protein and dopachrome tautomerase were all cloned by screening melanocyte cDNA expression libraries with antibodies raised in rabbits against hamster tyrosinase (Kwon *et al.* 1987, Jackson 1988, Jackson *et al.*

1992). While tyrosinase is the only enzyme absolutely necessary for melanin production, the other two modify the kind and amount of melanin produced. The genes are located on three different chromosomes and show an overall amino acid sequence identity of approximately 40%. It is proposed that the three genes arose from an ancient triplication event and sufficient time has passed such that the only nucleic acid identity seen is that which is required for amino acid identity (Jackson *et al.* 1992). The sequence homology between these three genes is a useful tool for mutation detection; if a given residue is conserved between the three proteins, it is likely to be functionally important and variation at this point potentially pathogenic.

1.7.4 The brown (*b*) locus

A clone that was initially isolated as a candidate for tyrosinase (Shibahara *et al.* 1986) was subsequently mapped to the brown locus on mouse chromosome 4 (Jackson 1988). The murine b-locus protein is homologous to the human melanocytic glycoprotein, gp75 (Halaban and Moellmann, 1990). The brown locus protein was termed tyrosinase-related protein-1 (*Tyrp1*). Mutations at this locus do not eliminate pigmentation, but result in a mouse with a brown coat colour. The non-mutated *Tyrp1* brown locus is essential for black eumelanin production. The function of tyrp1 remains controversial. It has been proposed to function in the context of hydrogen peroxide elimination (Halaban and Moellmann, 1990). H_2O_2 is probably produced as a by-product of the initial steps of melanogenesis when tyrosine is converted to melanin precursors. If the H_2O_2 is not removed, it can destroy the precursors as well as melanin itself. Subsequently it was shown that the b-protein has dopachrome tautomerase activity (Winder *et al.* 1994), converting dopachrome to DHICA rather than DHI. The product of the slaty locus had already been shown to have dopachrome tautomerase activity (Tsukamoto *et al*. 1992) (see below), and the question immediately arose as to why the melanocyte would require two enzymes with the same activity. Winder *et al*. (1994) suggest that since DHI is far more cytotoxic than DHICA, the cell may require the activity of two proteins capable of converting dopachrome to DHICA. Alternatively, they propose that tyrp1 and tyrp2 may have different functions within the conversion process, one catalysing the conversion of dopachrome to quinone methide, the second converting quinone methide to DHICA.

1.7.5 The slaty (*slt*) locus

The enzyme dopachrome tautomerase was named tyrosinase related protein 2 (tyrp2) and was mapped to the *slaty* locus on mouse chromosome 14 (Jackson *et al.* 1992). The enzyme catalyses the tautomerisation of dopachrome to 5,6-dihydroxyindole-2-carcoxylic acid (DHICA) rather than conversion to 5,6-dihydroxyindole (DHI). The *slaty* variant is a point mutation, resulting in the amino acid substitution, arg194gln (Jackson *et al.* 1992). The slaty phenotype is a slate grey coat colour, lighter than the wild type black coat. Mice homozygous for the slaty mutation show a reduction (3 to 4-fold) in dopachrome tautomerase activity (Tsukamoto *et al.* 1992, Jackson *et al.* 1992), thus the mutation down-regulates enzyme activity but does not abolish it.

1.8 Human pigment loci

The genes directing the formation, transport and distribution of the specialised melanin containing organelles, the melanosomes, can collectively be referred to as the pigmentation genes (Sturm *et al.* 2001). The first of these genes was identified through mapping the loci which resulted in mouse coat colour variants, and subsequently, genes have been identified through molecular characterisation of human inherited hypopigmentaion disorders (e.g. OCA1 and OCA2). The process of melanin production involves stimulation of the melanocyte by the hormone alpha-melanocyte stimulating hormone (α-MSH). The α-MSH will bind to the melanocortin-1 receptor (MC1R), previously called the melanocyte stimulating hormone receptor (MSHR), which lies in the melanocyte membrane, and stimulation of the MC1R will induce melanosome maturation with subsequent transference of pigmented melanosomes to surrounding keratinocytes in the skin. The MC1R also controls switching of the melanin type produced by the melanosomes – red/yellow or brown/black pigments may be produced. The crucial catalytic enzyme, tyrosinase, is active in both pheomelanosomes (producing red/yellow pigments) and eumelanosomes (producing the brown/black pigments), while the latter also require the activity of the tyrosinase-related proteins 1 and 2 (TYRP1 and TYRP2) to produce the darker melanins.

A critical determinant of tyrosinase activity is intramelanosomal pH and this is thought to be controlled by the action of the melanosomal-membrane protein, the P protein. The

interaction of tyrosinase, TYRP1, TYRP2 and P proteins (together forming the melanogenic complex) is not clearly understood.

1.8.1 Melanocortin 1 receptor (*MC1R*)

In mice, the relative amounts of pheomelanin and eumelanin produced is essentially under the control of two loci – *extension* and *agouti*. The *extension* gene is expressed in melanocytes and produces the melanocortin 1 receptor (Mc1r) protein. Mc1r is stimulated by the binding of α-MSH or ACTH (Mountjoy *et al.* 1992), both of which induce melanogenesis (eumelanin production) and darkening of the skin. A number of alleles are described, the dominant allele (*E*) results in black (wild-type) coat colour, the recessive allele (*e*) results in a yellow coat colour when present in the homozygous state. While mutations in this gene lead to a lack of eumelanin production in the skin and hair, this is not the situation in the eye: *e/e* mice are able to produce eumelanin in their retinal melanocytes.

While wild-type Mc1r protein leads to eumelanin production, the murine agouti protein, on the other hand, acts as an antagonist of Mc1r, preventing binding of MSH and thereby preventing eumelanin production. In this situation, the default melanogenic pathway, i.e. pheomelanin production, operates in isolation.

The human homologue of *extension*, *MC1R*, maps to chromosome 16q24 (Magenis *et al.* 1994). As evidence exists that variation at the *MC1R* locus leads to coat colour variation in mice, the human homologue has been investigated for variants that may lead to normal pigmentation variation in humans. It has been shown that sequence variation at the *MC1R* locus is present in red haired, fair skinned individuals of British ancestry (Valverde *et al.* 1995). John and Ramsay (2002), investigating a group of red-haired individuals in South Africa (white, European descent) found all individuals to be either homozygous or compound heterozygotes for *MC1R* mutations, supporting the theory that variation at this locus is recessive with respect to the phenotype and two mutations are necessary (although not sufficient) to result in red hair and pale skin. Other population groups have also been investigated, namely, black Africans, Chinese and Indians (Rana *et al.* 1999), with Africans and Indians showing very little sequence variation at the MC1R locus and a single variant, Arg163Gln, being present at a high frequency (70%) in East and Southeast Asian

population groups. In contrast to the findings of Rana *et al*. (1999), John *et al*. (2003) looked at two black African populations, namely sub-Saharan Negroids and the San bushmen, and found variation in the MC1R gene in both population groups. These authors propose that selection has occurred at this locus in Africans and the MC1R gene has likely played a role in maintaining the dark skin colour of these population groups.

King *et al.* (2003b) have identified 8 Caucasoid patients with OCA2 who have red hair (and not the typical yellow/blond). Molecular analysis revealed that all 8 individuals carry two *P* gene mutations, resulting in the OCA2 phenotype, and 6 out of the 8 individuals carry additional mutations in the *MC1R* gene, resulting in the red hair phenotype. This is thus an example of a second locus modifying the OCA2 phenotype in humans, and is different to rufous albinism in Africans.

1.8.2 Human agouti signalling protein (*ASIP*)

The human homologue of the murine *agouti* locus was cloned using an interspecies DNA-hybridisation approach (Kwon *et al.* 1994). The human gene is 85% identical to the mouse and encodes a 132 amino acid protein. The human *agouti* gene maps to chromosome 20q11, and is closely linked to the *MODY* (maturity onset diabetes of the young) locus. It is interesting to consider the function of human agouti signalling protein, since there is no equivalent human phenotype to the murine wild type banding pattern of the hair shaft. The human protein is not expressed in melanocytes. Rather, the gene is expressed in adipose tissue, testis, ovary and heart, and at lower levels in liver, kidney and foreskin.

1.8.3 The tyrosinase family

Three genes are presently known to form the tyrosinase gene family: *TYR*, *TYRP1* and *TYRP2*. Their mouse homologous loci are *c* (albino), *b* (brown) and *slt* (slaty), respectively. These genes code for proteins with very similar amino acid sequence, all three being recognised by tyrosinase antiserum. It is possible these genes were formed from a common ancestral gene and have diverged over evolutionary time.

1.8.3.1 Tyrosinase (*TYR*)

Tyrosinase is the crucial enzyme of melanin biosynthesis, acting as a catalyst in the formation of DOPAquinone from tyrosine (the rate limiting step in melanin production) and then at least once more in the biosynthetic pathway. Homozygous mutations at the *c* locus in mice produce a white coat colour. Tyrosinase cDNA was first cloned in 1987 (Yamamoto *et al*.) by screening a mouse melanocyte cDNA library using mouse tyrosinase antiserum. Human tyrosinase cDNA was subsequently cloned by screening a human melanocyte cDNA library with hamster tyrosinase antiserum (Kwon *et al*. 1987). Using *in situ* hybridization, the human *TYR* locus was mapped to chromosome 11p14-21 (Barton *et al*. 1988). Both the human and mouse genes span more than 50 kb of genomic DNA, and contain 5 exons (Giebel *et al.* 1991b). Tyrosinase mRNA analysis revealed a high percentage (41%) of transcripts which are alternatively spliced (Porter *et al*. 1991). The original mouse cDNA cloned by Yamamoto *et al*. (1987) was actually missing exon 3. Most of the alternate transcripts exclude important functional domains and do not translate into functional tyrosinase protein. A second locus to which tyrosinase cDNA hybridizes is located at human chromosome 11p11.2-cen (Barton *et al*. 1988, Giebel *et al.* 1991). This locus contains only exons 4 and 5 of the tyrosinase gene and does not produce any detectable transcription product and is assumed to represent a tyrosinase pseudogene.

Deletions, nonsense, missense, frameshift and splice site mutations in *TYR* are all found to be associated with OCA type 1, with these mutations distributed throughout the gene (see the albinism database for a comprehensive listing). A number of missense mutations are clustered around the second copper binding region located toward the 3'region of the gene.

1.8.3.2 Tyrosinase related protein 1 (*TYRP1*)

Mutations at the *b* locus produce a brown coat colour in mice. The brown locus cDNA was initially cloned during a differential hybridization experiment using mouse melanoma cells and mouse neuroblastoma cells, and the cDNA (termed pMT4) was thought to be tyrosinase (Shibahara *et al*. 1986). When the real tyrosinase cDNA was cloned, it was shown that pMT4 was encoded at another pigment locus. This second locus was mapped to mouse chromosome 4 and termed a tyrosinase related protein (*trp*), and later designated *tyrp1* (Jackson 1988, Jackson *et al*. 1990).

The mouse genome contains a *tyrp1* pseudogene, but there is no evidence suggesting the human genome contains a pseudogene (Shibahara *et al*. 1991). Numerous alleles at the brown locus in the mouse have been described (reviewed by King *et al*. 1995). The cDNA of the human homologue, *TYRP1*, was cloned (Cohen *et al.* 1990) and mapped by *in situ* hybridisation to chromosome 9p23 (Murty *et al.* 1992). The brown locus in the mouse, as well as its' human homologue, *TYRP1*, are structurally different from tyrosinase, containing 8 exons and spanning 15-18 kb of genomic DNA (Shibahara *et al*. 1986, Shibata *et a*l. 1992). The gene product is a glycoprotein with a molecular weight of 75 kDa, and is sometimes referred to as gp75. The protein is a transmembrane protein, localized to the melanosome (the same as the P protein). Gp75 is a protein found in abundance in melanoma cell lines, but is not observed in non-pigmented (amelanotic) melanoma cell lines (Tai *et al*. 1983).

1.8.3.3 Tyrosinase related protein 2 (*TYRP2*)

The only known mutant allele at the slaty locus, *slt*, results in a dark gray/brown coat colour in the mouse when present in the homozygous state. The product of the mouse slaty locus and the human homologous locus, *TYRP2*, is the enzyme DOPAchrome tautomerase (DT). The *slt* mutation does not abolish DT activity, but down regulates function such that homozygous mice show a three- to four-fold reduction in DT activity in melanocytes (Jackson *et al.* 1992). The mouse cDNA was first cloned, essentially by accident, during experiments to clone the tyrosinase cDNA using tyrosinase antibody (Jackson 1988, Jackson *et al*. 1990). The human homologue was mapped to chromosome 13q31-q32 by fluorescent *in situ* hybridisation (Sturm *et al*. 1994). A subsequent paper by the same group (Sturm *et al.* 1995) showed that *TYRP1* and *TYRP2* are larger than *TYR* and that *TYRP2* is larger than *TYRP1.* These authors suggest that *TYRP1* was derived from a *TYR* duplication event and then *TYRP1* itself was duplicated to give rise to *TYRP2*.

1.8.4 The *P* gene

After *TYR*, the *P* locus on chromosome 15q is the second major pigment locus in the human genome.

1.8.4.1 Cloning the gene at the OCA2 locus

OCA1 was known to be caused by a defect in the enzyme tyrosinase, and is hence known as tyrosinase-negative (ty-neg) OCA. The gene, *TYR*, was cloned in 1987 (Kwon *et al*.1987). The biochemical defect in OCA2 (or ty-pos OCA: the tyrosinase enzyme functions normally in individuals with this type of albinism) was not known and so localisation of the gene proceded via classic linkage studies and the candidate gene approach. Jenkins *et al*. (1990) reported the exclusion of 57% of the genome after a genome-wide screen using random polymorphic markers. Subsequently, following the candidate gene approach, Ramsay *et al.* (1992) looked at the PWS/AS region on chromosome 15q. PWS and AS are distinct clinical syndromes mapping to 15q11-13, but they both show a hypopigmentation phenotype in a number of cases. Further, the co-existence of PWS and AS with albinism had been documented (Phelan *et al.* 1988, Wallis and Beighton 1989, Weisner *et al.* 1987). Through positive linkage scores to markers at the D15S10 locus, it was shown that the OCA2 locus mapped to chromosome 15q11.2-q12 (Ramsay *et al.* 1992). This region of human chromosome 15 shows synteny to mouse chromosome 7 which harbours two pigment loci, pink-eyed dilute (*p*) and tyrosinase (*c*). The *c* locus was excluded as a candidate for the ty-pos locus (Colman *et al.* 1993) and it was proposed that the human homologue of the *p* gene was the most likely candidate for ty-pos OCA (Ramsay *et al*. 1992).

The murine p protein, as well as its human homologue, P, were cloned (simultaneously by two groups: Gardner *et al*. 1992, Rinchik *et al*. 1993). Analysis of human P cDNA sequence revealed a protein of 838 amino acids with 12 putative transmembrane domains. Database searches showed that the P protein sequence showed highest homology (21%) to an *E.coli* tyrosinase-specific transport protein (Rinchik *et al*. 1993) and these authors proposed that the P protein might be a component of the melanosomal membrane and function as a transporter of tyrosine into the melanosome. The mouse p protein constitutes 833 amino acids, with the same 12 hydrophobic domains which could span a lipid bilayer (Gardner *et al*. 1992).

1.8.4.2 *P* gene mutations cause OCA2

At the time of cloning of the mouse p protein, ten different p alleles were studied - in six of these the homozygous mutant showed either an aberrant size mRNA (eg. p^{un} and p^{J}) or undetectable levels of p mRNA (eg. p^{6H}, p^{25H} and *p* - the original mutant allele). Rinchik *et al*. (1993) described a patient with PWS and OCA2 and showed, by Southern blotting, that this patient had homozygous deletions of a region of chromosome 15q which included the PWS region as well as the *P* gene locus, whereas in 15 other PWS/AS patients who did not have OCA homozygous deletions that included both regions (PWS/AS and P) were not observed. It was proposed that defects in a single gene, namely *P*, were likely to result in the OCA2 phenotype in man and mouse.

Lee *et al*. (1994a) investigated four patients, three with OCA2 (one of whom also had PWS) and a fourth with a milder form of albinism, autosomal recessive ocular albinism (AROA). These patients were screened for tyrosinase (*TYR*) as well as *P* gene mutations. *P* gene mutations were identified in all four patients, showing that mutations at the *P* locus give rise to a range of clinical phenotypes (OCA2, OCA2 associated with PWS and AROA). Further evidence for *P* mutations being responsible for the OCA2 phenotype was provided by Durham-Pierre *et al*. (1994) who described an intragenic deletion of the *P* gene in OCA2 individuals from a Brandywine family - an inbreeding population isolate in Maryland, USA (see below).

1.8.4.3 A common *P* gene mutation

In a study by Durham-Pierre *et al*. (1994), OCA2 individuals from three different population groups were studied: the Brandywine Maryland isolate (originally described by Witkop *et al*. 1972), African-Americans and black Africans. An intragenic deletion mutation of the *P* gene was found in a Brandywine family. The mutation is a 2.7 kb deletion which completely removes exon 7 of the *P* gene. The deletion corresponds to nucleotides 699-859 (amino acids 216-270) of the human P cDNA sequence, and results in a premature STOP codon and truncation of the P polypeptide, rendering a non-functional protein. Four unrelated African Americans with OCA2, two Africans (one from Cameroon and one from Zaire), as well as 15 Caucasians with OCA2 from 8 families were screened for the deletion - all four African-Americans and the two Africans were found to be

heterozygous for the deletion while none of the Caucasoids carried the deletion allele (Durham-Pierre *et al*. 1994). The authors proposed an African origin for this mutation, its prevalence in the African-American population resulting from a founder effect in the USA, or alternately the allele may be prevalent in west Africa from where the African-American population derives. The latter scenario has since been shown to be correct.

A screen of 73 unrelated southern African blacks with OCA2 revealed that the 2.7 kb deletion mutation accounted for 78% of OCA2 mutations in this population, confirming that the deletion is in fact common in Africa (Stevens *et al*. 1995, Stevens *et al*. 1997). In a smaller study comprising 13 OCA2 individuals, a similar frequency (77%) was observed for the allele in OCA2 individuals from Tanzania (Spritz *et al*. 1995). The 2.7 kb deletion mutation is also common in the Cameroon, where it accounts for 65% of mutant OCA2 alleles, and the highest frequency reported to date is in Zimbabwe where the deletion comprises 92% of mutant alleles (Puri *et al*. 1997).

1.8.4.4 Other *P* gene mutations

Apart from the 2.7 kb deletion mutation, no other common mutations have been described for the *P* gene. Numerous other mutations have been reported, most occurring within individual patients or families (Lee *et al*. 1994a and b, Spritz *et al*. 1995, Lee *et al.* 1995, Spritz *et al*. 1997, Kerr *et al*. 2000). Recently, a *P* gene mutation, specific to the Native American population group, the Navajo, has been described (Yi *et al*, 2003). The mutation is a 122.5 kb LINE-mediated deletion of the *P* gene and its carrier frequency in this population is approximately 1 in 20.

1.8.4.5 Structure of the *P* locus

The P protein is 838 amino acids with 12 putative transmembrane domains (Rinchik *et al*. 1993). Database homology alignment showed the structural arrangement was similar to that of proteins lying within a membrane and facilitating the transport of small organic molecules across the membrane (Rinchik *et al*. 1993). It was subsequently shown in the mouse that the membrane to which the 110-kDa p protein localises is the melanosomal membrane (Rosemblat *et al*. 1994); the melanosome being the melanocyte-specific organelle in which melanin biosynthesis occurs. Lee *et al*. (1995) characterised the

molecular organisation of the human *P* genomic locus. The locus spans at least 250 kb of genomic DNA (as determined by the group of phage clones that spanned the coding sequences of the locus) but may be as large as 600 kb (this upper limit being determined by the fact that the entire *P* locus is contained within two YAC (yeast artificial chromosome) clones, one containing *P* exons 1 - 22 (this YAC, YAC 93C9, contains an insert of 370 kb) and another containing *P* exons 3 - 25 (this YAC, YAC 149A9, contains an insert of 400 kb and the region of overlap between the two YACs is approximately 170 kb). These two YACs give further information regarding the orientation of the locus - YAC 149A9 (the YAC containing the 3' end of the *P* gene) maps proximal to YAC 93C9, indicating that the *P* gene is orientated 3' - 5' with respect to the centromeric end of 15q, i.e. the human *P* gene is transcribed off the reverse strand. (This is not the situation in the mouse, where the *p* gene is transcribed 5' - 3' with respect to the centromeric end of chromosome 7).

The *P* gene consists of 25 exons, the first of which is non-coding (that is, exon 1 is transcribed into mRNA but is not translated as part of the protein sequence). The DNA sequence of exon 19 contains an in-frame STOP codon (TGA) which, when translated leads to a truncated and non-functional protein. Exon 19 is alternately spliced and is only present in a small percentage of P mRNA (Rinchik *et al*. 1993), ensuring the bulk of P mRNA is translated into full length P protein, which excludes amino acid residues coded for by exon 19. The *P* gene harbours several sequence polymorphisms in both coding and non-coding regions, scattered throughout the gene.

Although the *P* gene is located within the PWS/AS region on chromosome 15q and was therefore initially thought to be imprinted (Gardner *et al*. 1992), neither the human nor the mouse loci are subjected to imprinting phenomena, and individuals heterozygous for *P* mutations are normally pigmented, irrespective of the parent of origin of the mutant allele.

1.8.4.6 Family of P-like proteins

At the time the P protein was cloned and sequenced (Rinchik *et al.* 1993), researchers reported the human P protein showed homology to an *E.coli* tyrosinase transport protein, tyrP. These findings were not supported by subsequent database searches by the same authors (Lee *et al*. 1995). Instead they identified what was termed a novel family of P-like proteins which function as membrane transporters of small molecules. At this time P-like

proteins were identified searching the NBRF-PIR protein sequence database using the BLAST (Altschul 1990) program. Five proteins showing significant homology to P were identified: the Arsb proteins of Staphylococcus aureus and Escherichia coli, the E.coli Nhab Na+/H+ antiporter protein, the *Mycobacterium leprae* 38L protein and the mouse p protein. The 38L protein of *M.leprae* is a cell membrane protein and shares 40% identity (57% similarity) over its 429 amino acid length with the human P protein, suggesting the two proteins may serve a common function.

Homology alignments are extremely useful for defining amino acid residues which are conserved over evolutionary time. If an amino acid at a given position is conserved across species, it is a clue that variation at that point in the protein cannot be tolerated, and indicates the importance of that residue in maintaining normal function of the protein. Thus, if during a screen for mutations in a gene, an amino acid variation is detected, the amino acid change is more likely to be pathogenic if it occurs at a position where the residue is conserved. Looking at amino acid conservation between species is an important tool in assessing the significance of an amino acid substitution mutation.

1.8.4.7 Function of the P protein

The function of the P protein remains unclear. The protein is membrane-bound, being localised to the melanosmal membrane (Rosemblat *et al*.1994). It was initially proposed that the P protein might function as a transporter of tyrosine, the precursor of melanin biosynthesis, into the melanosome (Rinchik *et al*. 1993). It was subsequently shown that the P protein is not a significant transporter of tyrosine (Rosemblat *et al*. 1994). Further database homology searches showed that the human P protein is a member of a family of proteins which function as membrane transporters of small molecules (Lee *et al*. 1995). While it has still not been ascertained what the P protein transports across the melanosomal membrane, this transport function has been shown to be necessary for maintaining an acidic pH within the melanosome (Puri *et al.* 1998). Tissue culture experiments demonstrated that mouse cell lines deficient for p were almost never acidic and did not contain the tyrp1 protein. The cellular pH affects the activity of tyrosinase, the enzyme will only function in a low pH environment. These authors (Puri *et al.* 1998) therefore propose that the p protein mediates ionic transport which serves to maintain an acidic pH within the

melanosome; and further, the incorrect pH may interfere with the assembly of the melanogenic complex as demonstrated by the lack of tyrp1 protein.

1.8.5 ASIP, MSH, MC1R and the pheomelanin/eumelanin switch

Ian Jackson proposed a model as to how the melanocyte undergoes the pheomelanin/eumelanin switch (Jackson 1993): pheomelanin production in the melanocyte is the default pathway in the absence of MSH. If MSH is present and binds to its receptor, then the melanocyte switches to eumelanin production. The agouti protein prevents MSH binding to MC1R, negating the influence of the MSH. In this way the presence of the agouti protein results in pheomelanin production only. The agouti mouse shows alternate black and yellow bands on the hair shaft, the black bands are wild type colour and the yellow bands result from expression of the agouti locus. Mutations at the *agouti* locus would result in a black coat colour. Mutations at the *MC1R* locus would lead to the inability of MSH to bind its receptor; in this scenario the default pheomelanin pathway would be in place and the eumelanin switch could not occur, resulting in the all yellow coat colour. Experimental evidence supports this hypothesis: when human melanocytes in culture were incubated together with high levels of ASIP (human agouti signalling protein) it was found that ASIP blocked the stimulatory effects of MSH on cAMP accumulation, tyrosinase activity, TYRP1 levels and melanocyte proliferation (Suzuki *et al*. 1997). These findings suggest that the role of ASIP in humans is actually similar to that of the agouti protein in mice, the proteins playing a regulatory role in pigmentation.

1.8.6 The role of pigment genes in normal pigment variation in humans

A vast body of knowledge exists regarding coat colour genes in the mouse (with 127 loci now known to affect pigmention), 68 of these have known human homologues. These genes must be seen as candidates for playing a role in normal pigment variation (Hoekstra, 2006).

As discussed (section 1.1), human skin colour was originally dark and has over evolutionarily time become lighter with increased geographical distance from the equator (Relethford, 1997). Most human populations have become fixed for brown hair and eye

colour with phenotypic variation being found amongst individuals of European descent.

Variants associated with normal variation in humans have been described at 9 loci to date. Online Mendelian Inheritance in Man (OMIM) has ascribed the symbols *SHEP1-9* for these loci (*SHEP* for skin/hair/eye pigmentation). *SHEP1* phenotype variation is associated with genotypic variation in the *P* gene, *SHEP2* with variation at the *MC1R* locus, *SHEP3* with variation at the *TYR* locus, *SHEP4* with variation at the *SLC24A5* locus, *SHEP5* with variation at the *SLC45A2* locus, *SHEP6* with variation at the *SLC24A4* locus, *SHEP7* with variation at the *KITLG* locus, *SHEP8* with SNPs on chromosome 6p25.3 and *SHEP9* with polymorphisms in the 3'untranslated region of the *ASIP* gene. Variation at the *P* locus will be discussed in greater detail below.

A recent paper (Sulem *et al*. 2007) describes a huge genome wide association search for variants associated with skin, hair and eye variation in an Icelandic and a Dutch population. The findings of these authors supported previously described associations found in the *P* gene, *MC1R* and *TYR* and found further variants in *SLC24A4*, *KITLG* and on 6p25.3.

1.8.6.1 Variants in the *P* gene influencing normal pigment variation

The first mapping experiments looked for the locus for brown eye colour (BEY2), and found strong linkage of BEY to chromosome 15q11-q12 in 832 families in the Copenhagen area (Eiberg and Mohr, 1996). This study further showed linkage of brown hair colour (HCL3) to markers in the same region, leading these authors to suggest that the *P* gene might be a strong candidate for brown eye and hair colour. These findings were supported by later studies (Frudakis *et al*. 2007, Posthuma *et al*. 2006) also showing brown versus blue-green eye colour and brown hair colour to be linked to the *P* gene locus. The *P* gene plays a role in determining normal variation of eye and hair colour between population groups (Sturm and Frudakis, 2004).

Duffy *et al*. (2007), undertook a twin study in an Australian population (largely therefore of European descent) and found three separate polymorphisms in intron 1 of the *P* gene to be associated with blue eye colour: The three SNPs, located closely and tightly linked, are T/C (rs7495174), G/T (rs4778241) and T/C (rs4778138). The TGT haplotype was found to be strongly associated (present in 90.5% of alleles in individuals) with blue eye colour. The

TGT/TGT diplotype was found in 62.2% of subjects of European descent, as opposed to 7.4% in individuals of African ancestry, a finding consistent with the recessive nature of lighter pigment phenotypes. The authors suggest that this haplotype may be seen as diagnostic for eye colour. They further suggest that the position of these variants - in the 5' regulatory region of the *P* gene – may alter the temporospatial expression of the gene, resulting in the phenotype changes.

An Icelandic study (Sulem *et al*. 2007) also detected these three SNPs, but a fourth variant, rs1667394, showed even stronger association for blue versus brown eyes. The rs1667394 A allele was most frequent in blond haired and blue eyed individuals, being present at a frequency of 80-90% in northern European populations. This SNP is found 200kb downstream of the *P* gene, lying in intron 4 of *HERC2*. The intronic sequence variation within *HERC2* is unlikely to affect *HERC2* expression itself, but may well affect *P* gene expression, thereby producing phenotype changes in pigmentation. Alternatively, the *HERC2* variant may be linkage disequilibrium with a functional variant within the *P* gene itself (Sulem *et al*. 2007).

1.9 The *P* gene environment

Besides the *P* gene itself, neighbouring regions on human chromosome 15q (and the syntenic regions on mouse chromosome 7) are proving interesting candidates in the aetiology of the hypopigmentation phenotype.

1.9.1 Proximal to *P*: The Prader-Willi Syndrome/Angelman Syndrome region

Prader-Willi syndrome (PWS) and Angelman syndrome (AS) are non-allelic, inherited disorders caused by genetic abnormalities of imprinted genes on chromosome 15q11-q13. Imprinting refers to a category of non-Mendelian inheritance where a gene is expressed differently, depending on whether it was inherited from the male or female parent. Following fertilisation, a somatic cell will know which parent the imprinted region came from, leading to parent-of-origin specific gene expression (Sapienza 1990). PWS is caused

by a loss of expression of gene/s of the paternally inherited chromosome 15, while AS is caused by a loss of maternal gene expression.

1.9.1.1 Deletions in the PWS/AS region

Different molecular mechanisms can lead to AS and PWS, but each results in loss of parent-of origin (imprinted) gene expression. In the majority of cases (70-75%), patients have a cytogenetic deletion of chromosome 15 from bands q11-q13 (reviewed in Glenn *et al.* 1997). This represents a region of approximately 4 Mb of DNA which includes the *P* gene at its' telomeric end (see Fig 1.9).

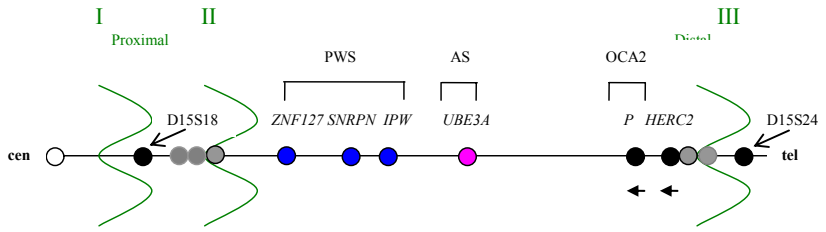


Figure 1.9 Schematic representation of human chromosome 15q11-q13. Region of chromosome 15q showing the centromere (cen), telomere (tel), certain genes and markers the two proximal and one distal breakpoint region associated with PWS and AS (green curvy lines) and the position of duplicated repeat units related to *HERC2* associated with these breakpoint regions (open black circles). The presence of these large, duplicated structures at either end of 15q11-q13 suggests a simple mechanism of chromatid misalignment during meiotic recombination. Imprinting status is indicated in colour: loci showing expression of the paternal allele are shown in blue, maternal allele expressed, pink; nonimprinted genes and markers are shown in black. The black arrows indicate the direction of transcription of the *P* gene and of *HERC2*. (Adapted from Ji *et al*. 1999 and Nicholls 1999.)

1.9.1.2 Uniparental Disomy in the PWS/AS region

Uniparental disomy (UPD) is the next most common cause of PWS (see OMIM reference numbers 176270 [PWS] and 105830 [AS]). In PWS patients with UPD, the individual inherits two maternal alleles, but no paternal allele for the region of chromosome 15. Since the second, totally normal, but also maternally derived copy of chromosome 15 cannot complement the missing paternal chromosome, this shows that the maternally inherited

PWS genes are normally silent and the paternally inherited genes are normally expressed. The converse is true in AS, where patients with UPD have inherited two paternal alleles and no maternal complement of this region of chromosome 15.

1.9.1.3 Imprinting mutations in the PWS/AS region

The remainder of cases are caused by point mutations within a single gene or mutations within a specialised region called the imprinting center (IC). In AS, the involvement of a single gene has been proposed and this gene is UBE3A (Kishino *et al*. 1997, Matsuura *et al*. 1997). A family was described with an isolated deletion of the UBE3A gene (Burger *et al*. 2002). The mutation caused AS when inherited through the female germline but not PWS when paternally inherited. These findings support the hypothesis that loss of UBE3A expression off the maternally inherited chromosome is sufficient to cause AS, and that this gene is not involved in PWS. In PWS it is likely that at least two genes are involved, making this a contiguous gene syndrome. In about 5% of PWS and AS cases, patients have no deletion mutation and no UPD, but they do show abnormal methylation patterns of the imprinted region. Thus, PWS patients have a maternal methylation pattern on their paternally derived chromosome, and AS patients carry a paternal methylation pattern on their maternally inherited chromosome (Nicholls *et al*. 1994). About half of all imprinting cases have a microdeletion involving the 5' region or the first exon of a gene in the region called SNRPN (Dittrich *et al*. 1996, Saitoh *et al*. 1996). The region defined by the microdeletions has been termed the IC, and mutations in this region seem to block the resetting of the correct imprint as a cell passes through the germline of the parent of the affected individual.

1.9.1.4 Clinical presentation of PWS and AS

PWS and AS are clinically distinct syndromes. PWS patients present with neonatal hypotonia and failure to thrive, hyperphagia becomes apparent at 18-36 months and leads to severe obesity. Patients are mildly retarded. Hypogonadism and small hands and feet also characterise this condition (Holm *et al.* 1993). AS patients are severely mentally retarded, with little or no speech development. This syndrome is also known as "happy-puppet syndrome" due to the characteristic happy disposition, protruding tongue, ataxic movement disorder and hyperactivity (Williams *et al*. 1995). Interestingly, about 50% of

both PWS and AS patients are hypopigmented (noticeably more fair in skin and hair colour than their parents). Before the gene for OCA2 was cloned, it had been suggested that a pigment gene might lie on proximal chromosome 15q. In 1986, Butler et al. described a cohort of patients in which those exhibiting hypopigmentation carried deletion mutations as opposed to other (then, as yet undescribed) types of mutations. Also, approximately 1% of PWS and AS patients are affected with oculocutaneous albinism (Wallis and Beighton 1989, Rinchik *et al.* 1993). These observations led to this region being included in a search for the OCA2 locus, using the candidate gene approach (Ramsay *et al.* 1992).

1.9.1.5 The hypopigmentation phenotype

In a study of 28 PWS patients (Spritz *et al*. 1997), 15 (54%) were classified as hypopigmented. Seventeen of the 28 patients carried a deletion which included the *P* gene. Of the deletion patients, 14 (82%) were scored as hypopigmented. Thus, there were 3 patients with a deletion that did not include the *P* gene; one of these patients was hypopigmented. Of the 8 non-deletion patients, none were scored as hypopigmented. These findings show there is a high correlation between hemizygosity for the *P* gene and hypopigmentation in PWS. In contrast, it should be noted that partial deletions of the *P* gene which are associated with OCA2, do not result in hypopigmentation in heterozygotes (Rinchik *et al*. 1993, Durham-Pierre *et al*. 1994, Lee *et al*. 1994a).

Spritz *et al*. (1997) make three suggestions as to how the hypopigmentation phenotype may come about in PWS and AS. Firstly, the *P* gene promoter region may contain functional polymorphisms that down-regulate *P* function; when the intact *P* allele carries these polymorphism/s, reduced transcription occurs off the normal allele. Arguments against this hypothesis include the observation that there is no common polymorphism haplotype of the intact *P* allele amongst hypopigmented individuals (Spritz *et al*. 1997) and further, if the *P* promoter region did contain polymorphisms with functional significance, then OCA2 heterozygotes carrying a *P* mutation on one chromosome and a functionally significant polymorphism on the other chromosome, would also show hypopigmentation, but this is clearly not the case.

Secondly, there may be a second, as yet undescribed pigment gene in this region of chromosome 15 that is being deleted together with *P* in certain PWS and AS patients. This

second gene may lie very close to *P*, or it may even lie within the *P* gene itself, being transcribed as an alternate transcript. (The *P* gene is large and it is feasible that it may contain several other genes.) Mouse data do not support this hypothesis - mice heterozygous for large *P* gene deletions do not show a hypopigmentation phenotype (Johnson *et al*. 1995). This is not to say that a second putative pigment gene could not have been introduced into this region after the mouse-human evolutionary divergence.

Thirdly, it has been noted that the maternal and paternal chromosome 15 homologues associate with each other in the region of 15q11-q13 during late S-phase (LaSalle and Lalande, 1996). These authors suggest that this may help in bringing about allele specific transcription, as occurs in this region. In PWS and AS patients, this physical chromosome 15 homologue association was found to be disrupted, especially in deletion patients. Since there are both imprinted and non-imprinted genes in this region, the control of gene expression is complex and when a deletion disrupts this organisation, the normal expression pattern of both imprinted and non-imprinted genes may be affected. Consequently, the intact *P* allele may show reduced expression.

The question of why certain PWS and AS patients show a hypopigmentation phenotype is an important one in the context of OCA2. Answers to this question will be directly applicable to questions concerning partial/intermediate OCA phenotypes (examples to be presented in chapter 2 and termed 'unclassified black OCA' in this study). Further, it is important always to bear in mind the complexity of the physical genomic region in which the *P* gene lies.

1.9.1.6 The hyperpigmentation phenotype

It is interesting to note that a patient with a duplication of 15q11.2-q14 has been described (Akahoshi *et al*. 2001). The woman presents with clinical manifestations of proximal 15q trisomy and hyperpigmentation. Cytogentic interphase fluorescence *in situ* hybridization (I-FISH) analyses confirmed that the *P* gene is present in three copies in this woman.

1.9.2 Distal to *P*: The *HERC2* locus

1.9.2.1 The PWS/AS deletion breakpoints

As discussed, most cases of PWS and AS result from a deletion on proximal chromosome 15q. The frequency of PWS and AS is similar – approximately 1 in 15 000. This indicates that a deletion mutation is equally as likely to arise during male gametogenesis as it is during female gametogenesis. The frequencies of PWS and AS are also similar across all populations groups, as would be expected for chromosome disorders which arise de novo in each affected patient and where the parents are genotypically normal. By implication then, the genetic make-up of the region is predisposing to the deletion event. The breakpoints indicate an underlying instability, a genetic 'hotspot' for recombination.

It has been shown that most (90%) of the PWS/AS deletions have clustered breakpoint regions. The distal breakpoint is between the *P* locus and the marker D15S24 (Knoll *et al*. 1989), while there are two proximal breakpoint regions, region II lies between D15S18 and ZNF127 and region I lies between D15S18 and the centromere (Christian *et al*. 1995) (see Fig.1.9). About 60% of the deletion patients (class II) show a proximal deletion breakpoint in region II, while 40% (class I patients) show a deletion breakpoint in region I.

1.9.2.1.1 Other diseases caused by chromosome rearrangements

Several other human diseases, e.g. alpha(α)-thalassaemia, Duchenne muscular dystrophy (DMD), steroid sulfatase (STS) deficiency, Charcot-Marie Tooth type 1A (CMT-1A) and haemophilia A (Factor VIII deficiency) often arise from large chromosome rearrangements. e.g. deletions, duplications or inversions. Deletions at the α-globin locus have clustered 3' breakpoints and often involve homologous recombination events between the duplicated α-globin genes (Nicholls *et al.* 1987). Homologous recombination between Alu sequences leads to deletions as well as duplications at the LDL receptor locus (Lehrman *et al*. 1985). It was found that the common 1.9 Mb deletion of the STS locus occurred through homologous recombination between low-copy repeated sequences called S232 elements, which are situated at either end of the deletion on Xp (Yen *et al*. 1990). Also, CMT-1A is associated with a 1.5 Mb region tandem duplication of chromosome

11p11.2-p12 (Pentao *et al*. 1992). This 1.5 Mb region is flanked by a sequence called CMT-1A-REP which is proposed to be a large, complex repeat element of 17 kb (Pentao *et al.* 1992). The molecular aetiology of haemophilia A in approximately 50% of the affected males is a large (~500 kb) inversion on Xq which disrupts the Factor VIII gene. The inversion occurs by homologous recombination between low-copy repeat elements, one in intron 22 of the gene and either of two homologues located 500 kb distal (Lakich *et al.* 1993, Naylor *et al*. 1993).

Clearly, there is evidence that large, low-copy repeat elements predispose to recurring rearrangements (Lupski 2007).

1.9.2.2. Identification of low-copy repeat elements at PWS/AS breakpoint junctures

In order to identify repeat sequences within the chromosome 15q11-q13 interval, Buiting *et al*. (1990) isolated several clones from the region by microdissection. One of these clones (MN7), when used as a probe, detected multiple restriction fragments which had been mapped to 15q11-q13 and 16p11.2 by somatic cell hybrid analysis and in situ hybridisation (Buiting *et al*. 1992). By PCR, it was determined that there were 4 or 5 MN7 loci spread over a large distance within 15q11-q13. Further, the MN7 probe detected several cDNAs that mapped to both 15q and 16p. The authors suggested that MN7 was part of a low-copy repeat element that contains an expressed sequence. By Southern hybridisation, a low-copy repeat element (END repeat) that occurs specifically at the proximal (II) and distal breakpoint regions was identified (Amos-Landgraf *et al*. 1999). A 100bp section of the repeat unit was found to have 98% identity with a database EST, supporting the earlier proposal by Buiting and colleagues that the repeat formed part of an unidentified gene.

1.9.2.3 The mouse p^{6H} allele

The region of the mouse genome syntenic to human chromosome 15q11-q13 is a region on *Mus musculus* chromosome 7 (Mmu7). In the human, the distal PWS/AS breakpoint lies immediately distal of the *P* gene. Therefore, the distal breakpoint homologous region in the mouse must lie close to and proximal to *p* on Mmu7. Many deletions at the mouse *p* locus have been induced by radiation mutagenesis. The radiation-induced mutation, p^{6H}, is a large deletion which includes the *p* gene and extends proximally on Mmu 7 (see Fig 1.9).

Mice homozygous for p^{6H} deletions exhibit a hypopigmentation phenotype (due to loss of the *p* locus) plus three other phenotypes: runtiness, a nervous jerky gait and male sterility with female semi-sterility (Hunt and Johnson 1971, Handel *et al*. 1987, Lyon *et al*. 1992). Using complementation analysis, Rinchik *et al*. (1995) showed that the neurologic and fertility abnormalities were likely to be due to dysfunction at a single locus, *jdf-2* (juvenile development and infertility-2). In the literature this locus is also referred to as the *rjs* (runty, jerky, sterile) locus.

1.9.2.4 The gene associated with the PWS/AS breakpoints as well as the jdf-2 mouse phenotype: *HERC2*

By positional cloning, the gene at the murine *jdf-2/rjs* locus was cloned (Lehman *et al*. 1998). The cDNA is exceptionally large (>15 kb) and the predicted protein contains several motifs including three RCC-1 (Regulator of Chromosome Condensation-1)-like repeats and a HECT (homologous to the E6-AP carboxyl terminus) domain. *RCC-1* is a gene which has been mapped to human chromosome 1p36.1 (Nishimoto *et al.* 1994) and its product is an abundant, nuclear, DNA-binding protein which acts as a guanine nucleotide exchange factor (GEF).

Approaching the issue from a different angle, Ji *et al*. (1999) cloned the gene which contains the transcribed low-copy repeat element that is associated with the breakpoint region of the PWS/AS deletions. The gene is called *HERC2* as it contains a HECT domain and RCC-1 motifs. It is related to *HERC1*, a gene encoding a GEF protein. *HERC2* is the ancestral gene, and the human genome contains duplicated, but truncated, copies of the gene which comprise a family of low-copy repeats. The mouse homologue, *Herc2*, is present only as a single copy gene in the mouse genome and is the gene at the *jdf-2* locus. *Herc2* mutations have been shown to be responsible for the phenotype in three different ethylnitrosuria (ENU)-induced *jdf2* mouse mutants (Ji *et al*. 1999).

1.9.2.5 Herc2 may play a role in pigmentation

Walkowicz and colleagues (1999) have characterised a further 19 *jdf2* mutations (13 were radiation-induced and 6 were ENU-induced mutants). Of special interest here are the alleles p^{12DTR}, p^{39DSD} and p^{103G}. p^{12DTR} is an interstitial deletion which results in loss of

5 kb from the coding sequence of *Herc2*. Both p^{39DSD} and p^{103G} are not deletion alleles, but rather show abnormal size restriction fragments on Southern analysis, indicating a rearrangement event. The rearrangement in p^{39DSD} is confined within *Herc2* itself, lying in the 3' region of the gene. p^{103G} shows a DNA rearrangement extending from the 3' end of *Herc2* into the interval between 3' *Herc2* and the 5' end of the *p* gene (see Fig.1.10).

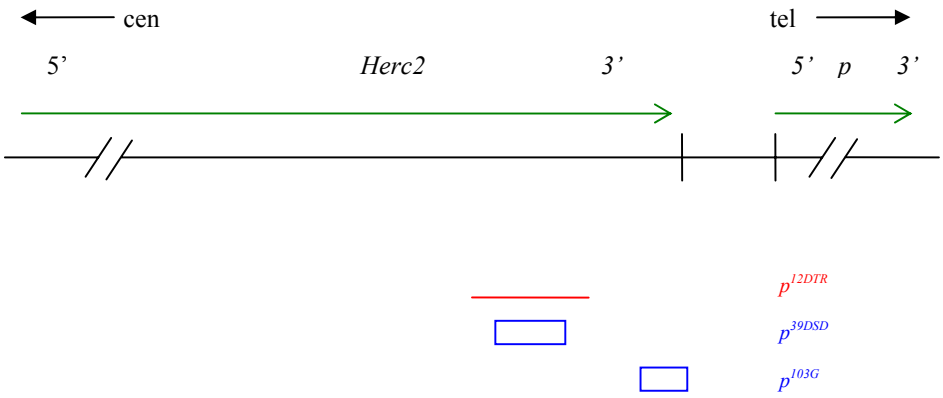


Figure 1.10 Map of mouse chromosome 7 showing the region containing the 3' end of *Herc2* and the 5'region of the *p* gene, plus intervening sequence. Orientation relative to the centromere (cen) and telomere (tel), as well as direction of transcription (green arrows) of the genes is shown. Positions of certain radiation-induced *p* mutations are indicated below the map: the deletion p^{12DTR} is represented by a red line and the intervals in which the rearrangements present in p^{39DSD} and p^{103G} are found, are represented as blue boxes. (Adapted from Walkowicz *et al.* 1999).

Notably, while none of these mutations include the *p* gene, all three alleles lead to a hypopigmentation phenotype in the homozygous state: p^{39DSD} and p^{103G} are mottled mutations (light coloured patches of fur against a black background) and the allele p^{12DTR} shows an intermediate albino phenotype with moderately diluted coat colour and dark pink eyes (Russell *et al.* 1995).

1.10 Psychosocial aspects of albinism

In the black population of South Africa the prevalence of OCA2 is 1 in 3900 (Kromberg and Jenkins, 1982). At the largest black hospital in South Africa, Chris Hani Baragwanath

Hospital in Soweto, just outside Johannesburg, a baby with albinism is expected to be born every two months (Kromberg *et al.* 1987b). Because albinism is so common in South Africa it is essential that the society develops an appreciation of the problems and issues surrounding this condition.

1.10.1 Social attitudes

The striking phenotypic difference between black individuals affected with albinism and their population group, as opposed to white individuals affected with the same condition, has possibly led to the black individual having to deal with greater problems of social adjustment. In the African context this situation is compounded further by ethnic beliefs and myths. A study involving 35 South African black individuals with OCA2 (Kromberg and Jenkins, 1984) found that the subjects, aged between 13 and 21 years, had a positive body image and, contrary to what may have been expected, claimed to be less ashamed of their physical appearance than did the matched control group without albinism. Individuals with albinism were found to suffer from psychosomatic ailments such as sweaty palms and headaches. This group also felt that gaining employment would be a problem, and while individuals with albinism seemed to be accepted in the community, positive attitudes rarely extended to include marriage.

1.10.2 The mother-child bond

Since the mode of inheritance of OCA is autosomal recessive, the birth of an affected baby is usually a shock to the parents as there is often no family history of the condition. Mothers are frequently depressed and unhappy. This manifests in a delayed mother-child bonding and mothers are often reluctant to breast-feed their babies as they look so different to themselves (Kromberg *et al.* 1987b). Babies with albinism were found to exhibit delayed milestones (sitting, crawling, walking). This may be a factor of lack of maternal involvement in the early months of the baby's life, combined with the physical problems an infant with albinism would experience due to poor eye-sight. Vision is very important for the infant's ability in learning to control body movements and developing a perception of the surroundings in which they find themselves (Kromberg *et al.* 1987b). It was found that

maternal attachment tended to normalise by the time the infant was approximately nine months of age.

1.10.3 Myths and superstitions

The occurrence of albinism in Africa has been documented over historical time (reviewed by Kromberg 1990) and myths have often surrounded those affected with albinism. Both positive and negative associations have been documented. Myths surrounding the birth of a baby with albinism include punishment for bad behaviour of the parents, the mother conceived during menstruation or ate undesirable food-types while pregnant (Kromberg 1990). Mythology from central and northern Africa holds that babies with albinism might be the result of a Negress-gorilla or Negress-water spirit mating (Pearson *et al*. 1913 as cited in Kromberg *et al*. 1987b). On the other hand, the birth of a baby with albinism might be due to the mother being touched by a spirit or receiving a good-luck sign from the spirits. Certain communities associate the condition of albinism with the affected individual having increased spiritual or religious powers (Kromberg 1990) although this is probably not a widely held belief and affected individuals themselves do not appear to claim superior spiritual powers (Kromberg 1984). A common belief is that people with albinism do not die, they simply disappear - the death myth (Kromberg 1984). The origin of this myth is not known, but it has been suggested that because infanticide of babies born with albinism was practiced, especially in rural areas in the 19th century mothers who had left their babies in the bush to die would tell people that the baby had ‘disappeared’. Another explanation may be that, since individuals with albinism were conceived by spirits, they were thought not to be fully human and therefore could not die (Kromberg *et al*. 1987a).

1.10.4 The issue of whiteness in the African context

Black individuals with OCA2 are never perceived socially as ‘white’ by their community, they may be ostracised for their physical appearance, but this is more a factor of how different they look from their normally pigmented population group and not because this population attributes any political significance to their being ‘white’. In the Sudan

individuals with albinism were referred to as ‘danedio’ which means ‘white’ and not ‘toulako’ which means ‘white man’ (Kromberg 1990).

1.10.5 Prenatal diagnosis

The option of prenatal diagnosis for albinism was first put forward in 1981 by Haynes and Robertson. These workers looked at scalp samples from 12 aborted fetuses (not affected with OCA) between 16 and 20 weeks gestation. Fetal melanocytes were studied by histochemistry and electron microscopy (EM) in order to ascertain development of the melanosomes. A normal fetus will exhibit fully pigmented melanosomes (stage IV) as early as 16 weeks gestation, the occurrence of OCA in the fetus will be demonstrated by the incomplete development of melanosomes (stage II). Using this technique it would be essential to obtain hair bulb and scalp skin biopsies from the fetus by fetoscopy as melanin is not present in amniotic cells, but only in melanocytes, making amniocentesis unhelpful (Taylor 1987).

The first prenatal diagnosis for albinism was performed in London in 1983 (Eady *et al*. 1983). The procedure was undertaken for a couple from the Middle East who had one normal child and one child affected with albinism. Fetal scalp samples were obtained from the fetus by fetoscopy at 20 weeks gestation and analysed by EM. Stage II melanosomes were observed, but no stage III or IV melanosomes. The fetus was diagnosed as being affected with albinism and the parents elected to terminate the pregnancy.

The techniques involved in prenatal diagnosis were then extended to include the dopa reaction test, useful in diagnosing tyrosinase-negative OCA in a fetus (Gershoni-Baruch *et al.* 1991). To test the feasibility of this test as a tool in prenatal diagnosis, the dopa reaction was performed on skin samples taken from fetuses aborted between 12 and 30 weeks gestation (for reasons not related to albinism). The dopa reaction in hair bulbs of normal fetuses gave a positive result from week 17 (no hair follicles were seen in scalp samples obtained before 17 weeks). This study supported earlier findings that pigmented melanocytes were only present in hair bulbs after 16-17 weeks gestational age. Further, skin biopsies taken from the back and thigh showed that pigmentation was not seen in the epidermis at any gestational age studied. Gershoni-Baruch *et al*. (1991) used EM studies in conjunction with the dopa reaction test to exclude albinism in a fetus by showing stage IV

melanosomes and a positive dopa reaction in hair follicles. In a family where the type of albinism was tyrosinase-negative (OCA1), a dopa reaction test would give a definitive result as to the status of the fetus.

If the type of albinism present in a family is known and the locus causing the albinism has been identified, then the prenatal diagnostic method of choice would be DNA linked marker analysis. This method avoids the more invasive procedure of fetoscopy - fetal DNA may be derived from amniocytes or chorionic villus sample (CVS) material. This option however, requires the availability of DNA from an affected individual in that family. The majority of OCA mutations are rare, usually specific to a given family. Consequently, direct detection of a common mutation in a fetus is not often an option. The notable exception is OCA2 in black African individuals. In this population group, one common mutation (the 2.7 kb deletion) accounts for up to 78% of OCA2 chromosomes (Stevens *et al.* 1995; 1997), making it a feasible scenario that both parents are carriers of this specific mutation. In this case, direct mutation analysis in the fetus is diagnostic.

The laboratory in which this thesis was carried out offers albinism molecular diagnostic services, including prenatal diagnosis in families where this is feasible. Worldwide, laboratories offering prenatal and other diagnostic testing for OCA can be found at GeneTests (www.genetests.org).

1.11 Host-related genetic background and susceptibility to the infectious disease tuberculosis

"We have to inquire whether persons living habitually in the same environment and with practically the same risk of infection have the same chance of developing phthisis whatever be their stock." (Pearson 1912)

1.11.1 History of tuberculosis

It is not known when humans first became afflicted by tuberculosis (TB), but available evidence suggests that the disease predates written records and has been prevalent for several centuries. It is thought that *M. tuberculosis*, the causative agent for TB in humans, derived from *M. bovis*, which causes TB in cattle and other animals, as well as in humans. Cattle first became domesticated in the Neolithic period which is also the same period that shows the earliest evidence of TB in humans. Between 7000 B.C. and 4000 B.C. in the Middle East, it is thought cattle were first domesticated and milking of cattle began. Therefore transmission of the mycobacterium to humans may have been through injestion of infected meat or by drinking milk from infected cows. A skeleton found in a Neolithic cemetery near Heidelberg in Germany (dating to about 5000 B.C.) had deformities characteristic of spinal tuberculosis – collapse and fusion of the thoracic vertebrae leading to a hunchback appearance. Although these skeletal lesions are in keeping with a diagnosis of TB, they are not sufficient evidence to confirm a diagnosis. The earliest conclusive data come from skeletal remains found in Egypt dating to between 3700 B.C. and 1000 B.C. The mummy of a 5 year old child showed evidence of pulmonary and vertebral TB. This was confirmed by the identification of acid-fast bacilli (*M. bovis* or *M. tuberculosis*) on microscopy of material taken from vertebral bone (Zimmerman 1979).

1.11.2 The introduction of tuberculosis into South Africa

While tuberculosis has been present in Europe and Asia for millennia, it is a relatively new disease in southern Africa. Although written records do not exist prior to the late 19th century, it is assumed that the disease did not occur with any significant prevalence in the native populations of southern Africa prior to the era of European colonisation. South Africa was first colonised by the Dutch in 1652, with the purpose of setting up a station in the Cape to provide fresh provisions for trade ships sailing between Europe and Asia. Several waves of immigration followed (including British, French and German) and over the course of the next 300 years, southern Africa found itself completely under European control.

It appears that the spread of TB to the native peoples of southern Africa was initially determined by their degree of contact with Europeans. In the early 19th century TB became problematic amongst the Khoikhoi (Hottentots) peoples, who lived in the south-western areas of South Africa which were colonised early. The San (Bushmen) were ravaged by TB following European contact. These groups may be seen as 'virgin soil' for tuberculosis disease, with individuals having no acquired resistance to the pathogen. Tuberculosis also appears to have become established early among the Coloured peoples living in the south-western Cape. Bantu-speaking black Africans, living in the interior of the country, remained unaffected for longer. It is thought that low levels of endemicity were present amongst the Zulu in Natal and the Xhosa in the Ciskei prior to the 20th century (Packard 1989).

The discovery of diamonds in 1867, and gold in 1886, in the area around Johannesburg (now known as Gauteng), created a demand for cheap labour which led to the development of a migrant labour system, drawing black workers from not only within South Africa's borders, but also from neighbouring countries. Miners worked in close proximity to each other in hot and moist conditions, were housed in over-crowded compounds and generally had a poor diet. High tuberculosis mortality rates in black mine workers ensued. To further aggravate the situation, mines adopted a policy of terminating workers' contracts once they developed the disease and repatriating them to their homelands. This practice played a key role in disseminating the disease to the rural areas. The mining industry also led to rapid

industrialisation, with simultaneous urbanisation and the socio-economic problems which follow.

Anti-tuberculosis drugs were introduced in the 1950s, after which a sharp decrease in mortality rates was noted. However, incidence rates continued to increase. Factors such as over-crowding, unemployment and poverty continue to contribute to the high incidence of tuberculosis in this country.

1.11.3 Clinical tuberculosis

Tuberculosis, an infectious disease, is caused by the human pathogen, *Mycobacterium tuberculosis*. The disease affects approximately 2 billion people worldwide (this is about one third of the world's population) and results in about 2 million deaths per year, making it the infectious agent that claims more lives than any other worldwide (Frieden *et al*. 2003). TB, once thought to be controlled, has reinstated itself as a major health problem, particularly in developing countries. The gravity of the situation led the World Health Organisation (WHO) to declare tuberculosis a global emergency in 1993. Poor TB control measures have led to the emergence of multi-drug resistant disease, the *Mycobacterium bovis* bacille Calmette-Guerin (BCG) vaccination has limited efficacy and the AIDS pandemic have all contributed to increased susceptibility and the dramatic increased incidence of clinical TB.

A two-step process in progression to TB disease takes place. Firstly the individual will become infected with *M. tuberculosis*. Clinically, the reaction to this initial infection usually goes unnoticed. In an individual with a healthy immune system, the bacterium will lie dormant. The second step, clinical TB, will only result following the immunosuppressive effects of factors like malnutrition, another infectious disease, alcoholism or diabetes. The time period between infection and disease can therefore vary from weeks to a lifetime, although the risk of disease is highest in the first year following infection, declining as time proceeds (Comstock 1982).

1.11.4 Selection and infectious disease – Malaria as the classic example

Population studies have shown that infectious diseases can act as selective agents with respect to a populations' genetic structure, strongly influencing allele frequencies at a given locus. The best known example is that of the relationship between the infectious disease, malaria and the human genetic haemoglobinopathies - sickle cell anaemia and thalassemia - where heterozygotes for globin gene mutations have a selective advantage in the presence of malarial disease. Malaria has established itself as 'the strongest known force for evolutionary selection in the recent history of the human genome' (Kwiatkowski 2005).

Sickle haemoglobin (HbS) is a variant of the *HBB* gene (which encodes β-blobin). This variant has arisen independently in different areas and is maintained at a frequency of approximately10% in many malaria endemic areas (Flint *et al*. 1998). That is, 1 in 10 individuals are carriers of the sickle cell trait in these areas, making this the most striking example of heterozygote advantage in human genetics. HbS homozygotes are affected with sickle cell disease, but HbS heterozygotes have a 10-fold reduced risk of suffering severe malaria (Hill *et al*. 1991).

1.11.5 Host genetics and infectious disease

It has long been noted that certain infectious diseases have a 'heritable' component and that different individuals respond very differently following exposure to a given infectious agent. Up until the latter part of the 19th century, diseases such as TB and leprosy were believed to be inherited disorders.

While infectious diseases themselves are obviously not heritable, susceptibility or resistance to certain infectious agents has now clearly been demonstrated for a number of diseases. Early evidence came from incidents like those which occurred in Lubeck, Germany, where infectious *Mycobacterium tuberculosis* was accidentally administered to a group of individuals – some of whom never developed TB disease, others developed the disease but recovered, while in others the infection proved fatal. In the 1980s,

haemophiliacs accidentally given HIV infected blood, showed great variation in their progression to AIDS and AIDS-related illness.

While these examples demonstrate how the individual's genetic background may influence susceptibility to disease, host susceptibility may also be seen at a population level. A West African tribe, the Fulani, appear to be far more resistant to malaria than any of their immediate neighbours (Modiano 1996). In the USA, all the occupants of a nursing home were exposed to the TB pathogen and it was noted that patients of different ethnic backgrounds were differently affected – individuals of African American origin were more susceptible and were at greater risk of developing TB disease than were their white counterparts (Stead *et al*. 1990).

1.11.6 Molecular genetic tools to study host susceptibility to infectious disease

The basic strategies available in studies of genetic association (in decreasing order of confidence) are:

- Twin studies (comparing disease concordance in monozygotic versus dizygotic twins)
- Family linkage studies (looking for association of the occurrence of TB in family members with the inheritance of a particular genetic marker)
- Case-control studies (comparing individuals with TB to control individuals)
- Anecdotal reports (family or ethnic clusters of TB suggesting increased susceptibility)

As the human genome project and the allied pathogen genome mapping projects have come to completion, a new approach to studying host-pathogen interactions has become available to the scientist. Infectious disease can be considered a complex genetic trait with many host genes all having small, but additive or interactive effects on the phenotype. Two main strategies are available when looking for areas of disease susceptibility/resistance in the host genome: genome-wide scanning with micro satellite or SNP markers to identify regions of the genome involved ('positional cloning') and secondly, the candidate gene approach which would look for association between the function-related candidate genes in

the host genome and disease susceptibility.

1.11.6.1 Genome wide screens

Chromosome 17q

In humans, the first example of gene mapping looking for a locus involved in infectious disease susceptibility was undertaken in large Brazilian pedigrees where susceptibility to schistosome infection appeared to be segregating as a Mendelian trait (Marquet *et al*. 1996). Using a genome scan approach and linkage analysis, a single major susceptibility locus on chromosome 5q31-33 was identified. A subsequent study in Brazil (Blackwell *et al*. 1998) looking at multi-case families for TB, leprosy and leishmaniasis, found a significant (P=0.01) association of TB susceptibility to the locus 17q11.2-q12.

Another two-stage study was undertaken using 16 Brazilian multi-case TB families (Miller *et al*. 2004). These authors found association with chromosomal regions 10p26.13, 11q12.3 and 20p12.1. No association was found for the 17q locus earlier described by Blackwell *et al*. (1998), but the small sample size of the Miller study should be kept in mind.

A large case-control study was undertaken looking at 4 candidate genes within the 17q11.2 region using patients from Mexico and Korea (Flores-Villanueva *et al*. 2005.) An association was found for one gene: an A→G polymorphism in the promoter region of MCP-1 (monocyte chemoattractant protein-1) was found to confer increased risk for developing TB in both the heterozygous (AT) and homozygous (GG) state, in both population groups.

Chromosomes 15q and Xq

The second attempt to map a human susceptibility locus for TB, using a large-scale family-based genome scan, was undertaken by Bellamy *et al*. (2000). Using linkage analysis of affected sib-pairs, Bellamy *et al*. (2000) looked at 92 affected sibling pairs collected in the Gambia and South Africa. All individuals were typed for 299 microsatellite markers found throughout the genome. Seven chromosomal regions showed maximum LOD scores >1, and these regions were then fine-mapped using a further 22 markers. Fine mapping

involved typing a second set of 82 affected sib-pairs and their parents. Results showed only two regions of the human genome, namely, chromosome 15q (LOD score=2.00) and chromosome Xq (LOD score=1.77) gave LOD scores suggestive of linkage. While these LOD scores do not achieve the levels of significance recommended for a genome-wide linkage scan (LOD>3), the authors argue that this is probably a reflection of the small sample size used here being insufficient to detect linkage for genes contributing to a complex disease phenotype. When mapping complex disease traits, single gene mapping strategies need to be adapted or reconsidered, as many genes are now involved, having small but additive or interactive effects on the phenotype.

Chromosome 8q

Most recently, a genome-wide scan using 96 multi-case TB families from Morocco (Baghdadi *et al*. 2006) found strong linkage to a single genomic region – on chromosome 8q12-q13. While this locus was not associated in all families, a particularly high association was observed in families where at least one parent was affected, leading the authors to suggest that predisposition at this locus is inherited in an autosomal dominant fashion. Baghdadi *et al*. (2006) suggest that, while TB susceptibility can result from the many-genes-with-small-additive-effects scenario, it can also result from certain susceptibility genes have large and dominant effects.

1.11.6.2 Candidate genes

Several genes that play a role in the human host's immunological response to infection by mycobacteria have been studied, looking for different alleles or polymorphisms that may exist which could cause subtle changes in function of the protein, leading to individual variation in susceptibility to TB. The polymorphic human leukocyte antigens (HLA) were the first proteins to be studied and subsequent work has been done on cytokines and pattern recognition receptors.

Human leukocyte antigens (HLA)

HLA alleles have been found to be associated with susceptibility to several different infectious diseases including malaria, HIV progression and hepatitis B and C persistence

(reviewed by Hill 2006). Two classes of HLA are defined – class I antigens, HLA-A, B and C and class II antigens, HLA-DR, DQ and DP.

Earlier studies looked at class I antigens and a meta-analysis, combining many of the results from these studies, by Kettaneh *et al*. (2006) concluded that there was no significant association of HLA class I alleles with TB susceptibility. They did find however that the HLA B13 allele has a protective effect.

The same meta-analysis showed for class II alleles that at the DR locus, DR3 and DR 7 are protective, DR8 is a susceptible allele and DR2 gave borderline results as a susceptibility allele (Kettaneh *et al*. 2006).

The HLA class II antigens, DR, DQ and DP receive more attention – the DR2 antigen has 11 alleles and many reports, some finding association and some not, have been published (reviewed by Takiff 2007). For the DQ locus, a significant association was found for DQB1*0503 with TB susceptibility in a Cambodian population (Goldfeld *et al*. 1998). This allele was found in the TB patient group only and in none of the 107 control individuals, leading the authors to suggest that this allele could have a near Mendelian effect in this population.

Cytokines and cytokine receptors

Cytokines and their receptors are further elements of the immune system, thought to be important in controlling mycobacterial infections. Cytokines, e.g. tumour necrosis factor (TNF), interferon-γ (IFN-γ) and the interleukins (IL-1, IL-8, IL-10 and IL-12) have been investigated. TNF variation has shown no association with either susceptibility or resistance to TB in studies from India (Selvaraj *et al*. 2001), Cambodia (Delgado *et al*. 2002) and Turkey (Oral *et al*. 2006). Published studies looking at INF- γ are inconclusive, with results differing for different populations studied.
Much the same scenario exists with contradictory results from studies on IL-12 and the IL-12 receptor as well as IL-1B. IL-1 gene polymorphisms are not associated with TB in Gambians (Bellamy *et al*. 1998), Gujarati Indians (Wilkinson *et al*. 1999) and Cambodians (Delgado *et al.* 2002). In the Cambodian study, an association for the -1082 polymorphism of the IL-10 promoter with TB susceptibility was reported. A Japanese study found

homozygosity for a certain haplotype of the IL-12 receptor gene to be associated with TB susceptibility (Akahoshi *et al*. 2003).

Vitamin D receptor

Before the availability of antituberculosis drugs, the disease was treated with high doses of vitamin D, leading to the thinking that vitamin D is an important factor in immunity to *M.tb*. As synthesis of vitamin D3 is dependant on the skin's exposure to UV light, it was postulated that darker skin races could synthesise less vitamin D and this may be one reason why more pigmented races are allegedly more susceptible to TB, especially those living in less sunny climates (Wilkinson *et al*. 2000). The effects of vitamin D are exerted by interaction of vitamin D and the vitamin D receptor (VDR). Numerous polymorphisms in the gene for the VDR have been found, but work on TB associations focuses on four of these - polymorphisms that either create or destroy four different restriction enzyme sites: *Fok* I, *Taq* I, *Bsm* I and *Apa* I and the alleles are designated F/f, T/t, B/b and A/a, with the uppercase letters denoting presence of the restriction site (Bornman *et al.* 2004).

Pattern recognition receptors

One of the first lines of defence of the human immune system is the recognition and subsequent uptake of a pathogenic microorganism by macrophages and dendritic cells. On the surface of these phagocytic cells are different types of pattern recognition receptors which bind to different patterns on the pathogens in order to promote phagocytosis and activate the signalling pathway of cytokine production, antigen presentation and the development of the immune response. Pattern recognition receptors of interest in TB susceptibility are toll-like receptors, the macrophage mannose-binding lectin (MBL) and the dendritic-cell-specific intercellular adhesion molecule-3, designated DC-SIGN.

1.11.6.3 Candidate genes studied in the SA population

Mannose binding lectin

Mannose-binding lectin (MBL) is a serum protein which forms part of the immune pathway. MBL, on binding to ligands present on the pathogen, activates the complement cascade. Three mutations in the human MBL gene have been described, all of which cause

low levels of the protein to be produced by the liver. The effect of low levels of MBL on susceptibility has been studied in several population groups, with contradictory results. In the South African Coloured population it was found that polymorphic alleles causing lower levels of MBL were associated, not with TB susceptibility, but with TB resistance, these alleles having a protective effect, especially against TB meningitis (Hoal-van Helden *et al.* 1999).

DC-SIGN

Dendritic-cell-specific intercellular adhesion molecule is a protein present on macrophages that recognises pathogens and is a major receptor for *M.tuberculosis* on dendritic cells (Tailleux *et al*. 2003). Two promoter region variants (-871G and -336A) have been identified in the gene encoding DC-SIGN, CD209. These two variants have been found to be associated with a lower risk of developing TB in the South African Coloured population (Barreiro *et al.* 2006). These authors investigated the frequency of the -871G protective allele in three other populations, Asians, Europeans and black Africans. Presence of the G allele was found in 21% of Asians, 38% of Europeans but was absent in Africans. Barreiro *et al*. (2006) propose that this could be a contributing factor to the increased susceptibility of this ethnic group.

INF-γ

A SNP in position +874 (T→A) of the INF-γ gene was studied and it was found that the 'A' allele, thought to down-regulate protein transcription, was more common in SA Coloured TB patients and the 'T' allele was more common in controls (Rossouw *et al.* 2003).

SLC11A1 (NRAMP1)

Both SLC11A1 (NRAMP1) and SLC11A2 (NRAMP2) were investigated in the SA Coloured population (Hoal *et al*. 2004). The $(GT)_9$ allele of the $(GT)_n$ dinucleotide repeat in the 5' region of SLC11A1 was found to be significantly more common in controls than in patients, suggesting this is a protective allele. No associations between TB susceptibility and variation at the SLC11A2 were found.

HLA class II

Lombard *et al*. (2006) looked at HLA-DRB1 and HLA-DQB1 polymorphisms in the Venda population. DRB1*1302 and DQB1*0301were found to be susceptibility alleles in this black population group. Interestingly, these particular alleles have been found to be protective with respect to malarial and hepatitis B and C persistence.

VDR

The vitamin D receptor restriction enzyme alleles have been studied in the Venda (Lombard *et al*. 2006) and the Cape Coloured population (Babb *et al*. 2007a). In the Venda, the haplotype F-b-A-T significantly protected against TB. In the Coloured population, the *Apa* I 'AA' and the *Taq* I 'TT' or 'Tt' genotypes were found to be predictive of a faster response to antituberculosis chemotherapy.

SP110

The human homologue of the mouse *Ipr1* gene, SP110, was investigated. No significant association was found between TB susceptibility and any of the 8 polymorphisms studied (Babb *et al*. 2007b).

1.11.6.4 Mouse genetic susceptibility studies

Nramp-1 (SLC11A1)

The first example of gene identification of a host susceptibility gene by positional cloning was the cloning of the murine natural resistance-associated macrophage protein 1 gene (Nramp-1) (Vidal *et al*. 1993). It was known that a locus in the mouse conferred resistance to intracellular infections such as Leishmania, Salmonella and the BCG strain of *Mycobacterium bovis*. The Nramp-1 gene was amenable to a positional cloning strategy given that the phenotype of the mice (either resistant or susceptible to infection) was clearly distinguishable in the F2 and backcross progeny. Resistant and susceptible inbred mouse strains showed Mendelian inheritance patterns, with resistance being almost completely dominant. The human homologue of Nramp-1, SLC11A1, was identified

(Cellier *et al.* 1994) and currently there are four polymorphisms used in TB association studies. A meta-analysis (Li *et al.* 2006) analysing data from 17 different studies concluded that, although the association of SLC11A1 with TB was not found in all studies, there seems to be sufficient evidence that some of the variants exert a mild effect on susceptibility.

Ipr1 (SP110)

A specific mouse strain, C3HeB/FeJ, shows extreme sensitivity to *M. tuberculosis.* A second murine susceptibility locus was identified (Kramnik *et al.* 2000) and termed *sst1*, for super susceptibility to TB. The gene was subsequently cloned and called *Ipr1* (intracellular pathogen resistance 1). In searching for the human homologue, the closest protein identified was SP110, which is only 41% identical to *Ipr1*. Tosh *et al*. (2006) looked at variation in the SP110 gene in three African countries, the Gambia, Guinea-Bissau and the Republic of Guinea. Results showed that 3 of the 20 SNPs tested were associated with TB in one or the other of these populations, but results were not consistent. A study from Ghana (Thye *et al.* 2006) found no association between SP110 variation and TB susceptibility as did a study by Babb *et al.* (2007) in the South African population.

1.11.7 Mendelian susceptibility to mycobacterial diseases

A rare but very important group of patients comprises individuals who are hyper susceptible to mycobacteria, contracting infections in childhood or early adolescence. Disease in these patients is caused not by *M. tuberculosis*, but by BCG following vaccination (Petrini 2006). Presumably these individuals would be susceptible to *M. tuberculosis* as well, but were simply exposed to BCG first.

Mutations may either cause complete absence of gene product or may cause reduced levels of protein product. Infection in children with complete deficiency is generally fatal, being resistant to all forms of antibiotic therapy. Prognosis for patients with partial deficiencies is good, with affected individuals surviving into adulthood (Casanova *et al.* 2002, Ottenhoff *et al.* 2005, Fernando *et al.* 2006).

The genetic defects in many of the affected patients/families have been identified and are defining the human immunological response essential for controlling mycobacterial infections. The mutations are heterogeneous, often occurring in the children of consanguineous parents. Mutations are inherited in a classic Mendelian fashion, most commonly autosomal recessive, but autosomal dominant and X-linked inheritance has been described in certain families. The syndrome has been termed Mendelian susceptibility to mycobacterial diseases (MSMD).

Mutations in 5 different genes have been identified: IFN-γR1 and IFN-γR2 (the two chains of the IFN-γ receptor), IL-12B, IL-12RB1 (the β1 subunit of the IL-12 receptor) and STAT1 (signal transducer and activator of transcription 1), reviewed by Takiff (2007).

1.11.8 Evidence for a tuberculosis susceptibility locus on chromosome 15q

Much of the molecular biology addressing the relationship between infectious disease and host genetics started during the course of this PhD project. A possible connection between the *P* gene and TB was suggested (1997) when database searching showed that the genome of *Mycobacterium tuberculosis* contains, not only one, but two *P*-like genes (Rob Nicholls, personal communication). As *M.tb* is a unicellular organism and does not produce melanin, this was an unexpected finding. At the time it led to the question of whether variation at the *P* locus (more particularly, pathogenic variation) might lead to a selective advantage for OCA2 heterozygotes. OCA2 is the most common autosomal recessive disorder amongst blacks in southern Africa and the carrier frequency has been estimated as 1 in 35. Given the climate - hot, with high levels of ionising radiation - it would be easier to appreciate if this phenotype had been selected against (in the homozygous state). It has now been shown that the P protein is expressed in certain immune tissues (Prof Robert Nicholls, personal communication). In these tissues, the P protein does not function in its traditional role as part of the melanin biosynthesis pathway, and while its function is unclear, we presume its role as a membrane transport channel is required in these tissues. Given this unexpected finding, it may be suggested that variation at the *P* locus might influence an individuals' immune state. Further, if a known mutation caused loss of function of the *P* gene, might carriers of that mutation be either at an advantage (or disadvantage) with respect to an infectious disease?

The findings of Bellamy *et al.* (2000), as discussed above, lent strong support to the hypothesis that a locus on chromosome 15q11-13 was associated with susceptibility to TB. Fine mapping of this region (Cervino *et al.* 2002), using 10 microsatellites and 5 polymorphic markers in this 14cM region, showed that a polymorphism in UBE3A is significantly associated (p=0.002) with susceptibility to TB. This study did include a P gene marker (in IVS13), but it did not show association with TB susceptibility.

Given the fact that work in our laboratory was already being undertaken on the *P* gene (which maps to chromosome 15q12-13) and in light of evidence in the literature that this region is likely to be associated with TB susceptibility, it immediately begged the question, 'Is the *P* gene itself involved in host susceptibility to TB?' Since TB is such a significant health burden in this country, this question is not only interesting but also highly relevant.

Aims and Objectives

The initial aim of this project was to screen a cohort of albinism patients for the mutations causing their condition. As mentioned, OCA2 is the most common autosomal recessive condition in the black population of South Africa. OCA2 is caused by mutations in the *P* gene on chromosome 15q. All patients had previously been screened for the common black OCA2 mutation, but 22% of OCA2 gene mutations remained unaccounted for. Given the disease prevalence, it was justified that the search for the remaining mutations continues. It may have been that a second common mutation could be found or, at least, if family-specific rare mutations could be identified, this information would be useful to the given families. As our patient sample consisted mainly of black individuals, a full *P* gene exon screen was undertaken. The patient sample also included 9 white families – for this population group both the *P* gene and the *TYR* gene were investigated.

As the mutation screen drew to a close, and very few additional mutations were identified, we proposed that some OCA2 and BOCA mutations might lie in the promoter region of the *P* gene and act as down-regulators of *P* gene expression, rather than severe mutations which stop *P* gene expression altogether. At the time (1997), sequence data for the 5' region of the *P* gene was not available and so it became another aim of this project to clone

and sequence the *P* gene promoter region.

At about this time the *P* gene became interesting in a different and somewhat unexpected context. A possible link between the *P* gene region on chromosome 15q and genetic host resistance/susceptibility to the infectious disease TB, was suggested in the literature. As our research group was already interested in and working on the *P* gene and secondly, given the huge health burden of TB in South Africa, this new angle seemed relevant and important to pursue. Consequently, the second major aim of this project became to undertake an association study, looking at whether genotypic variation in the *P* gene region on chromosome 15q might be associated with host susceptibility to TB. TB is highly prevalent in two South African population groups, namely the black population and the Coloured population. The association study involved typing TB patients and unaffected controls for DNA variants at 5 loci within the *P* gene. Statistical analysis of the data – looking at whether a certain variant at a given locus may be more or less common in the patient versus control group – would lend support to the hypothesis that variation in this region of chromosome 15q is associated with host resistance (or susceptibility) to the infectious disease, TB.

The aims of this project could therefore be stated as follows:

1. Oculocutaneous albinism mutation screening

- (a). To screen the *P* gene coding regions in a cohort of individuals with albinism from the black population, who do not carry the common OCA2 mutation (the 2.7 kb deletion), for the mutation/s causing their condition.

- (b). To screen the *P* gene and the *TYR* gene coding regions in a cohort of individuals with albinism from the white population for mutations causing their condition.

- (c). To clone and sequence the *P* gene promoter region.

2. To determine whether genotypic variation in the *P* gene region on chromosome 15q is associated with host susceptibility to TB.

CHAPTER 2

SUBJECTS, MATERIALS AND METHODS

2.1 Introduction

Albinism is the most common autosomal recessive disorder amongst southern African blacks and has been a focus of research interest in the Department of Human Genetics, University of the Witwatersrand, for many years. Consequently, a large patient base has been identified in South Africa, and DNA samples from affected individuals and their families are available in the laboratory. Following the identification of a common black OCA2 mutation (the 2.7 kb *P* gene deletion), it became possible to screen individuals for this mutation directly by the polymerase chain reaction (PCR). This mutation however accounts for only approximately 78% of black OCA2 mutations, and is very rare in white individuals. No other common mutations have been described for either the *P* gene or the *TYR* gene, in any population group. Therefore, in families where the 2.7 kb *P* gene deletion was not the sole cause of the pathogenicity observed, the unknown mutation/s had to be identified through various other methodologies. In this study, mutation detection involved two techniques, firstly, single strand conformational polymorphism (SSCP) with sequencing of exons showing variation and secondly, Southern blotting.

As an additional point of interest in the study of the genetics of hypopigmentation, the gene *HERC2* was investigated. This gene, a close neighbour of the *P* gene, may play a role in the aetiology of the hypopigmentation phenotype. A subgroup of the OCA subjects and normal controls were investigated at the DNA level by Southern blot analysis.

In order to investigate a possible association between the *P* gene and the infectious disease, tuberculosis (TB), the second part of the project included a group of black and a group of mixed-ancestry (so-called 'coloured') South African patients with tuberculosis who were screened (by PCR) for the 2.7 kb deletion mutation of the *P* gene. These same TB patients, plus random control individuals from the same ethnic groups, were genotyped for alleles at an additional four polymorphic intragenic *P* gene marker loci. The aim was to establish whether the *P* gene may play a role in resistance/susceptibility to infection and disease development in tuberculosis.

2.2 Subjects

2.2.1 Subjects with albinism

Subjects involved in the screen for OCA-causing mutations had been collected during previous research projects conducted at the South African Institute for Medical Research (now called the National Health Laboratory Service), and on field trips undertaken to Lesotho, Zambia and the Central African Republic (CAR). All subjects with albinism were tested for the 2.7 kb *P* gene deletion mutation prior to the present study. Subjects who were shown to be homozygous for the 2.7 kb deletion mutation were not included in the present study. Subjects who had one or two undetected mutations (there were 72 of them) were then selected for further investigation.

This subject cohort affected with albinism comprised two groups of individuals: 1) subjects of black African origin and 2) subjects of white European origin (most South African whites are of Dutch, French, German or English descent). These subjects were screened for mutations in the two pigment genes most likely to carry OCA mutations in the given populations, namely the *P* gene and the *TYR* gene. *P* gene mutations (responsible for OCA2) have been found to occur in both black and white populations. *TYR* mutations (responsible for OCA1) have, to date, essentially only been found to occur in white populations or populations of mixed ancestry.

2.2.1.1 Subjects with unclassified albinism

During collection of subjects on the various field trips, certain hypopigmented individuals were observed. These individuals were all from the black African population, but their phenotype was not typical of OCA2, ROCA or BOCA (and could not be classified as OCA1 as some pigment production was evident in skin, hair and eyes). These individuals were grouped together and termed ‘unclassified black’ for the purpose of this study and as published in Kerr *et al.* (2000).

As the type of OCA can not be determined from phenotypic presentation in the white

population (both OCA1 and OCA2 may look similar in this population group), Caucasoid individuals with OCA were also termed 'unclassified' (unclassified white) in this study and as published in Kerr *et al*. (2000).

A total of 72 individuals with OCA were included in a screen for OCA-causing mutations (see Table 2.1). The majority of these subjects (63/72) were of black African origin, as would be expected given the relative size of the population and the population frequency of OCA2, but a number of white families were also included. Black individuals with the classic OCA2 phenotype were only screened for *P* gene mutations. The albinism phenotype shows locus homogeneity in the southern African black population, with all cases showing linkage to the *P* locus (Ramsay *et al*. 1992, Kedda *et al*. 1994). While BOCA has also been shown to be linked to the *P* locus in the same sample studied (Manga *et al*. 2001), black individuals with BOCA, as well as those with an unclassified form of albinism, were screened for *TYR* mutations as well as *P* gene mutations. The intermediate hypopigmentation phenotype observed in these two groups of individuals may have been caused by epistatic effects, with mutations at two different loci (*P* and *TYR*) contributing to the phenotype. White individuals were screened for both *P* and *TYR* mutations, in an attempt to determine which type of albinism was present in a given family.

Table 2.1. Albinism subjects involved in a mutation screen for pathogenic variants.

SUBJECT GROUP	TYPE OF OCA	N	%	GENE SCREENED FOR MUTATIONS
Black	OCA2	39	54%	*P*
	BOCA	9*	13%	*P, TYR*
	Unclassified	15	20%	*P, TYR*
White	Unclassified	9	13%	*P, TYR*
Total		72	100	

N= number of individuals.

*= In the original BOCA sample collected by P. Manga and in her subsequent publications (Manga *et al*. 1995, 1997), 10 BOCA individuals were available for study. DNA on one sample had run out and consequently only 9 BOCA individuals were included in the present study.

The majority of subjects were from South Africa, unless otherwise stated: of the 39 black OCA2 individuals, 31 were South Africans, 6 were from the CAR and 2 from Zambia; the

9 BOCA individuals were all collected by P. Manga and Professor J. Kromberg on a field trip to Lesotho; the 15 black individuals with an unclassified form of albinism included 7 from the CAR and 8 from Lesotho; and there were 9 white South African OCA individuals. The countries of origin of the subjects are shown on the map in Fig. 2.1. All subjects were unrelated with the exception of one of the CAR individuals with an unknown form of albinism who is the half-sib of an individual with classic OCA2 (both included in this study).

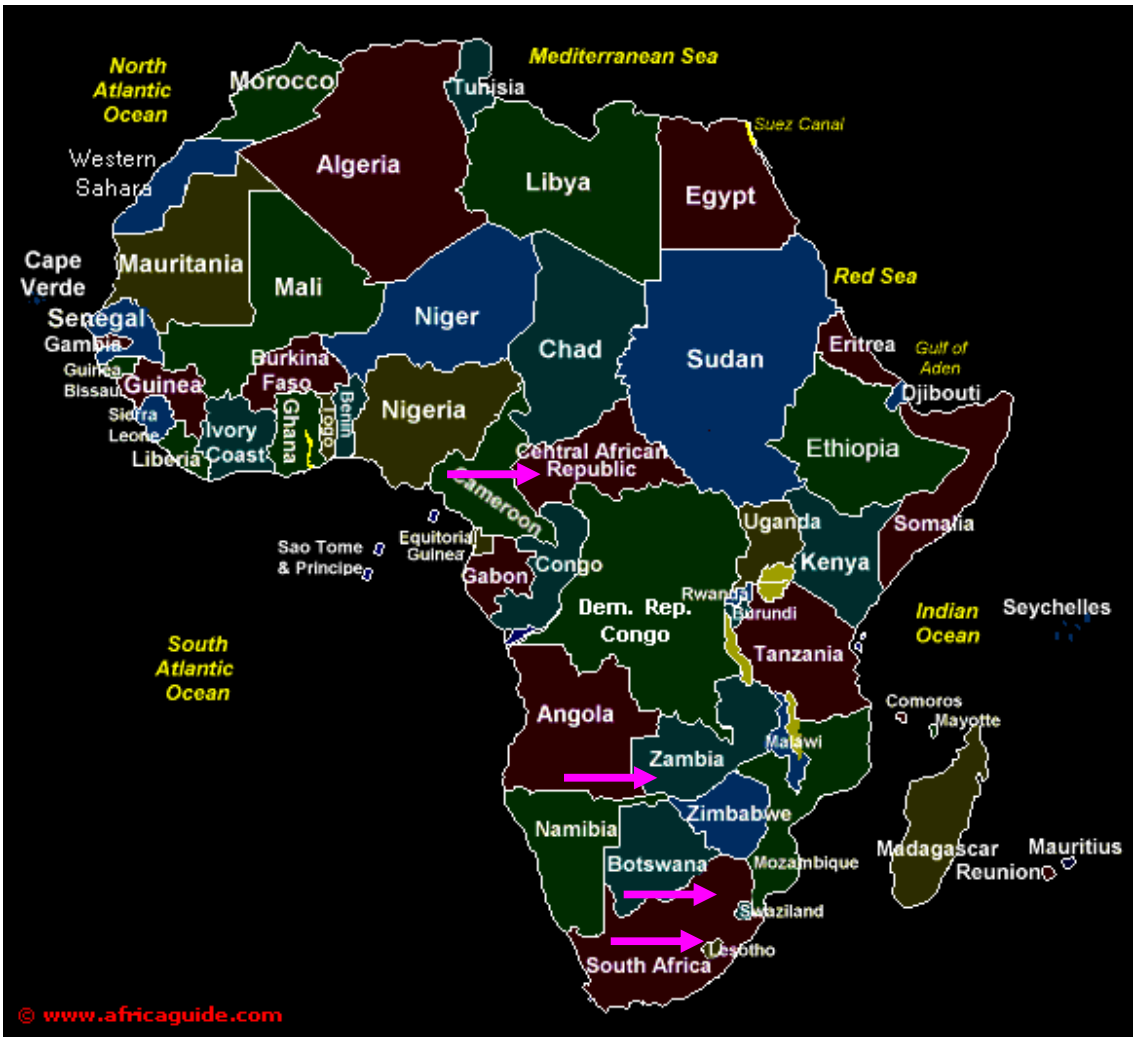


Figure 2.1 Map of Africa showing countries from which patient samples were collected. The four areas of sampling (Gauteng, South Africa; Lesotho; Zambia and the CAR) are highlighted using pink arrows. Lesotho is a land-locked country within South Africa's boarders; the Central African Republic is further north, lying just above the Congo.

2.2.2 Subjects with tuberculosis

While OCA is a single gene disorder, the study of host genetic response to an infectious disease can be considered a complex trait. When investigating complex diseases, large sample sizes are needed to see small effects. Due to the difficulties encountered in collecting the black TB patient sample for the present study, smaller numbers than hoped for were included, and this work should therefore be considered a pilot study.

In order to address the issue of a possible role for the *P* gene as a potential resistance/susceptibility locus in the aetiology of tuberculosis, blood samples from (normally pigmented) black patients with TB were collected in order to determine an observed carrier frequency of the 2.7 kb deletion mutation in this sample group and to establish whether it differed significantly from the general black population. The black population was targeted because this is the only population group in which a common OCA-causing mutation has been identified (i.e. the 2.7 kb *P* gene deletion) and its carrier frequency calculated in this population (1 in 40). Therefore, in the South African black population, we have figures for an expected carrier frequency to use for comparative purposes. Patients were collected from Sizwe Hospital, Charles Hurwitz Hospital and Lifemed Hospital (all dedicated TB hospitals in the Johannesburg area), and from the Tokoza Clinic (a general out-patient clinic also in the Johannesburg area). Visits to these hospitals were made by the writer and a Genetics nurse, Sr Merlyn Glass. The project was explained to the patients in as simple a way as possible, questions were encouraged and then volunteers were asked to donate blood. Informed consent forms were signed by all volunteers. Non-genetic TB susceptibility factors (e.g. HIV infection, diabetes, alcoholism) had to be taken into account, and any individual in which active TB disease could be attributable to any of these identifiable factors was excluded from the study.

Approval for the collection of samples was obtained from the Committee for Research on Human Subjects, University of the Witwatersrand (see Appendix). All samples were coded prior to DNA extraction, and research results were not conveyed back to the individual. In cases where the HIV status of the patient was not known, an additional blood sample was taken and the HIV status established. Only HIV negative subjects were included in the study. Due to the extremely high prevalence of HIV in TB patients, over a period of 18 months, a sample of only 72 black TB patients who were HIV negative was collected,

despite seeing over 600 patients in this time. Black controls were random individuals referred to the NHLS, Department of Human Genetics, for paternity testing.

In order to extend the study on black patients with TB, a second group was selected. These patients were of South African mixed ancestry (so-called 'coloured' individuals of Khoisan, black and white ancestry) and they were ascertained in the Western Cape province. Blood samples were collected from Coloured patients and DNA extracted from these samples by the staff of Dr Eileen Hoal-van Helden at Tygerberg Hospital, University of Stellenbosch Medical School. The mixed ancestry group of TB patients was not subject to the strict selection criteria of the black sample with respect to HIV status – at the time of sampling the HIV prevalence in this group was low and HIV status was not used as an exclusion criterion. HIV status of the coloured TB patients in this study was unknown. The controls were random blood donors, matched for geographic and ethnic origin. DNA aliquots were sent to our laboratory for DNA genotyping for the purposes of the present study. The subject group then totalled 350 (see Table 2.2). (Note – for the different marker assays, results could not always be obtained for all individuals typed due to technical difficulties. Therefore, for the different assays, slightly smaller sample sizes are reported on.)

Table 2.2. Subjects involved in a screen of the *P* gene for association with TB susceptibility.

SUBJECT GROUP	N*	%
Coloured controls	87	25%
Coloured TB patients	92	26%
Black controls	99	28%
Black TB patients	72	21%
Total	350	100

*N= number of individuals.

In addition to the 2.7 kb deletion mutation, four other *P* gene polymorphisms were typed, both in the cohort of coloured TB patients (plus matched controls) and the cohort of black TB patients (plus their matched controls). Three of these were $(AC)_n$ repeat microsatellite markers and the fourth was a single nucleotide polymorphism (SNP). Two of the AC

repeats were identified by searching genomic *P* gene sequence (downloaded from Celera) for stretches of 10 AC repeat units or more. A minimum of 10 repeats was chosen to give a greater chance that the repeat would be polymorphic. (Note – The sequence data used for this experiment, downloaded from the Celera website (www.celera.com), stretched from exon 52 of *HERC2* through *P* and included some 3' sequence. This work was done in 2001 before a more complete version of the human genome was publicly available. There were gaps in the sequence and exon 1 and 2 of the *P* gene were not present in the sequence download.) Three useful AC repeats were identified – in introns 17, 20 and 24. Information on the 3 novel markers and their polymorphic alleles was submitted to the Genome Database (GDB) (www.gdb.org) and they were assigned the following 'D' numbers: D15S1533 (IVS17), D15S1536 (IVS20) and D15S1537 (IVS24). The SNP is a published variant, occurring in the coding region (exon 9) of the *P* gene (Lee *et al*. 1995). The variant is a C to T base change, leading to an amino acid substitution in the P protein (R305W). See Figure 2.2 for the relative positions of the 5 markers typed.

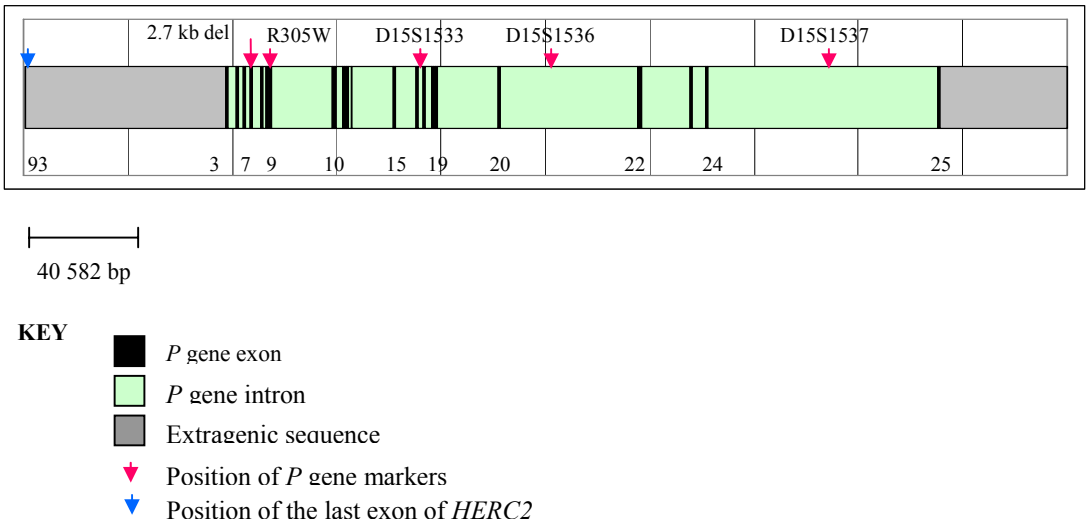


Figure. 2.2. A schematic representation of the genomic region surrounding the human *P* gene, showing the positions of the five intragenic markers used to screen TB patients and their appropriate controls. The 2.7 kb deletion mutation spans exon 7, the variant R305W is found in exon 9, the three microsatellite markers, D15S1533, D15S1536 and D15S1537, are found in introns 17, 20 and 24 respectively.

2.3 Materials and Methods

All chemical solutions used in this study were made up as described in Sambrook *et al*. (1989).

Methodologies used to screen OCA patients for disease-causing mutations will be described in this section (PCR-SSCP analysis and Southern blotting with DNA sequencing of variants). Two genes were included in the mutation screen: tyrosinase and the *P* gene.

Normal sequence data for the region 5' of *P* were not available at the commencement of this project; hence an attempt was made to clone and sequence the *P* gene promoter region. Two methods were used to try and obtain *P* gene promoter sequence – subcloning of a bacterial artificial chromosome (BAC) known to contain sequence 5' of *P*, and inverse PCR using primers located immediately 5' to *P* gene exon 1.

Methodologies used to type polymorphic markers within the *P* gene in the subjects with TB and their controls will be described below (specific mutation detection as well as microsatellite analysis on the ABI 377 sequencer). Five intragenic markers were examined to establish whether the *P* gene may be associated with tuberculosis susceptibility/resistance.

2.3.1 Collection and storage of blood samples

Blood samples were collected from all the subjects in either EDTA or ACD tubes, to prevent coagulation of the blood. Blood was spun down at 2300rpm for 10 min in order to separate the plasma, white blood cells (buffy coat), and red blood cells. The buffy coat was aspirated into a 50ml polypropanol tube and the DNA extracted immediately, or the buffy coat was stored at -20°C.

DNA was extracted by the salting out procedure, essentially as described by Miller *et al*. (1988). The method involves three basic steps. Firstly, Triton X washes remove unwanted red blood cells, secondly, overnight incubation with the enzyme Proteinase K which digests

proteins (which are subsequently pelleted and discarded after addition of saturated NaCl) and thirdly, precipitation of genomic DNA after addition of 100% ethanol.

If the blood sample volume was very small (< 1 ml), the DNA was extracted using the High Pure PCR Template Preparation Kit (Roche Diagnostics).

Following extraction, DNA was dissolved in 1X TE buffer and stored at 4°C or at -80°C for long term storage.

2.3.2 PCR

The polymerase chain reaction (PCR) is a technique used to amplify a specific short stretch of DNA. Specificity is determined by a set of primers which are designed to flank the region of DNA that is to be amplified. Different amplification cycling conditions will be optimal for different primer sets.

2.3.2.1 The *P* gene 2.7 kb deletion PCR assay

Firstly, all the subjects with albinism were tested for the 2.7 kb deletion mutation. A PCR assay was designed by Durham-Pierre *et al*. (1994) which allows rapid detection of the mutation. One of the primers (MHB51) has since been redesigned by the same authors and was used in this study. Three primers are added to the reaction (see Figure 2.3), two forward primers, MHB107 and MHB72, and a common reverse primer, MHB71.

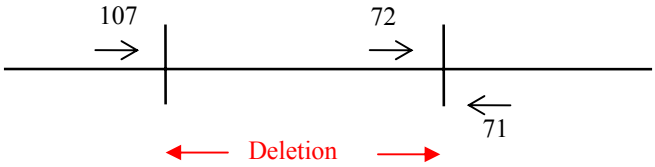


Figure 2.3 Diagramatic representation showing the position of the PCR primers used to detect the *P* gene 2.7 kb deletion mutation (Durham-Pierre *et al*. 1994).

If the deletion is not present, MHB107 and MHB71 are too far apart for a PCR product to be generated. In this situation (i.e. on a normal chromosome), a product of 240 bp will be generated between MHB72 and MHB71. If the deletion is present, then MHB72 would not be able to bind and MHB107 and MHB71 would be significantly (2.7 kb) closer and a product of 398 bp would be generated between these two primers. The primer sequences are given in Table 2.3.

Table 2.3. Primer sequences used to detect the *P* gene 2.7kb deletion mutation.

DIRECTION	NAME	SEQUENCE
F	MHB107	5'- CAT AGT CTT GGT TTT TGT AGT CCT -3'
F	MHB72	5'- GCG GTG GCT GTC ATG GC -3'
R	MHB71	5'- GGA GGG TGC ATT CAT TCT TCA G -3'

The PCR reaction was carried out in a total reaction volume of 25µl containing approximately 100ng DNA, 10pmol each primer, 1 unit Taq DNA polymerase (Boehringer Mannheim), 1.25µM each dNTP and 2.5µl 10X PCR buffer. Cycle conditions were 94°C 1 min, 60°C 1 min, 72°C 1 min for 30 cycles. PCR product was visualised on a 2% or 3% agarose gel stained with ethidium bromide.

If an individual affected with OCA2 did not carry the 2.7 kb deletion mutation (i.e. carried two unknown mutations), or if they were heterozygotes for the 2.7 kb deletion (i.e. carried one unknown mutation), the aim of this study was to identify the unknown mutation/s.

Further, all of the 9 BOCA patients included in this study which had been collected by P. Manga (PhD thesis, 1997) were found to be heterozygous for the 2.7 kb deletion. The second unknown mutation remained to be identified.

2.3.2.2 *P* gene PCR

The aim of this part of the study was to identify unknown *P* gene mutations. Individuals affected with OCA2 were selected to include those who did not carry the 2.7 kb deletion mutation (i.e. carried two unknown mutations), or were heterozygous for the 2.7 kb deletion (i.e. carried one unknown mutation). Further, out of a group of BOCA patients collected by P. Manga (Manga *et al*. 2001), 9/10 were found (by her) to be heterozygous for the 2.7 kb deletion and the second unknown mutation remained to be identified. The remaining individual did not carry the 2.7 kb deletion on either chromosome.

The *P* gene has 25 exons which are amplified individually using primers situated in the flanking intronic regions. PCR primers and cycling conditions were as described (Lee *et al.* 1995) for all exons except exon 16. No amplification was obtained for this exon using the published primer sequences and various PCR conditions, thus the primers were redesigned. The sequences of these primers are given in Table 2.4.

Table 2.4. Primer sequences for amplification of exon 16 of the human *P* gene.

DIRECTION	NAME	SEQUENCE
F	Primer 16F	5'- CTG CAG AAC TGA AGC ACG AG -3'
R	Primer 16R	5'- ATG TTC TGC TGC ACA CCA AG -3'

For exon 16, the PCR mix was made up as described, but only 1.25pmol of each primer was added in a 12.5µl reaction. Thirty five cycles of amplification were performed with denaturation at 95°C for 1 min, annealing at 54°C for 1 min and primer extension at 72°C for 1 min.

The exon 25 PCR product was too large for SSCP analysis (574 bp), and was therefore

digested with either *Hinf* I or *Bst* NI before running on a gel.

2.3.2.3 *Tyrosinase* gene PCR

To date, no common mutation/s have been identified at the *TYR* locus in any population group. Consequently, described *TYR* mutations are rare, and are largely specific to the family in which they occur. Black individuals with classic OCA2 were not included in a screen for *TYR* mutations. Caucasoid individuals with OCA, black individuals with an unclassified form of hypopigmentation and black patients with BOCA were screened for *TYR* mutations by SSCP analysis. The *TYR* gene comprises 5 exons. Exon 1 is too large (819 bp) to amplify as a single fragment for PCR-SSCP analysis and is therefore amplified in two smaller fragments, termed 1a and 1b (Oetting *et al*. 1994). The primer sequences used to amplify the exons of the *TYR* gene were all as published (Oetting *et al*. 1994).

2.3.3 PCR-SSCP analysis

Single strand conformation polymorphism is a method of mutation screening. The technique exploits the fact that single stranded DNA will take on a secondary structure which is directly determined by the sequence of the DNA strand. Consequently, if a DNA sequence variation has arisen, the single stranded DNA fragment containing that variant will take on a different secondary structure to that of the normal fragment, resulting in a mobility shift when the DNA fragment is run on a gel. Thus, when performing SSCP analysis, band patterns (and not band sizes) are compared between individuals. The method was first described by Orita *et al*. (1989a and b). Sensitivity of the technique is not 100%, but reportedly has a pick-up rate of 80-90% (Hayashi and Yandell 1993, Sheffield *et al*. 1993). Detection rate is improved by varying gel running conditions (e.g. adding glycerol to the gel mix, or changing the temperature at which the gel is run) and by keeping the size of the DNA fragments run on the gel small, optimally between 100-250 bp. The DNA single strands are run on a polyacrylamide gel, which gives better resolution than an agarose gel. In the present study two gel conditions were used to run the same sample under two different conditions in order to improve the detection rate – gels were made up with or without glycerol.

2.3.3.1 Radioactive PCR

A given exon was amplified by PCR in a total volume of 12.5µl containing approximately 100ng of genomic DNA, 5pmol of each primer, 125µM of each dNTP, 1.25µl 10X PCR buffer, 0.625µl (^{32}P) dCTP and 0.5 units Taq polymerase (Boehringer Mannheim).

2.3.3.2 Polyacrylamide gel electrophoresis

The radioactive PCR product was denatured by heating at 95°C for 3 minutes, and then loaded onto a 6% MDE (Hydrolink) vertical polyacrylamide gel. Each DNA sample was run under two different conditions (firstly on a gel containing 10% glycerol and secondly on a gel without glycerol) to increase the probability of detecting band shifts reflecting DNA variants. Gels without glycerol were run at 28 V/cm for 5-7 hours and gels with glycerol were run overnight at 10 V/cm. A fan was directed at the gel to ensure the gel did not overheat. This is obviously not a precise method of regulating temperature and consistency between gels could not be guaranteed. DNA fragments were visualised following autoradiography. If a SSCP variant was detected, the given exon was re-amplified in the absence of radioactive label, and the PCR product sequenced.

2.3.3.3 Sequencing

Several methods of DNA sequencing are available. In this project, two methods were used: manual sequencing and then later, when an automated DNA sequencer became available in the laboratory, automated DNA sequencing. Both methods used are based on the Sanger method of DNA sequencing where double stranded PCR product is the template DNA for the sequencing reaction.

PCR product (non-radioactive) was generated and then run on a 0.8% low melting temperature agarose gel. The DNA was extracted from the gel using the QIAGEN Gel Extraction Kit. This ensured the DNA fragment to be sequenced was separated from the

other PCR reagents (dNTPs and primers) and other possible spurious PCR bands. For manual sequencing, the USB sequencing kit for double stranded DNA sequencing (Perkin-Elmer) was used. Sequencing product generated through manual sequencing incorporated the radioactive isotope ^{35}S. This product was run on a 6% denaturing polyacrylamide gel at 35 V/cm to generate high temperatures, which ensures DNA remains denatured. DNA was visualised through autoradiography. For automated sequencing, sequencing product generated using the Taq FS dye terminator kit was run through a Centricep column to remove unincorporated, fluorescently labeled dNTPs. Samples were dried using an Eppendorf 5301 Concentrator. The dried pellet was mixed with 3µl dextran-formamide loading dye, denatured for 2 minutes at 95°C, then electrophoresed on a 4.3% denaturing polyacrylamide gel (36cm WTR plates at a constant 200W) for 7 hours on an ABI 377 Automated DNA Sequencer. Data were analysed using Version 3.4.1 Sequencer Analysis Software (ABI).

2.3.4 Southern blotting

This technique was developed in the 1970s by Ed Southern (Southern 1975) and essentially ear-marked the beginning of molecular genetic technology. It is a time-consuming procedure involving multiple steps and is therefore difficult to trouble-shoot if technical difficulties are encountered. Further, many laboratories still make use of radioactivity to label probes. For these reasons it is not a method of choice if PCR is a viable alternative. However, PCR is limited in the size of DNA fragment that can be analysed, making Southern blotting a necessary method for detecting large chromosomal rearrangements, deletions or duplications. The process involves restriction of genomic DNA using a specific restriction enzyme, separation of the digested fragments by agarose gel electrophoresis, denaturation of the DNA *in situ*, transfer of the single stranded DNA from the gel to a nylon membrane, hybridisation of the membrane with a radioactively (or biotin) labelled single stranded probe and finally exposure of the membrane to autoradiograph film.

While PCR-SSCP analysis was used to detect single base changes or small sequence variants, Southern blotting was used to look for large rearrangements firstly within the *P* gene and secondly, in the region immediately 5' of the *P* gene (i.e. the *HERC2* region).

2.3.4.1 Southern blotting for the *P* gene

2.3.4.1.1 DNA digestion

As an initial trial, human genomic DNA was digested with three enzymes (*Bam* HI, *Eco* RI and *Hind* III) in order to look for an enzyme digest suitable for the probe. Results of this blot showed *Eco* RI to be most suitable. This enzyme was chosen as the 2.7 kb deletion is contained within an *Eco* RI fragment, resulting in the presence of a smaller fragment, rather than a missing fragment when using a 2.8 kb *P* gene cDNA fragment as a probe (Durham-Pierre *et al*. 1994). Therefore 2.7 kb deletion heterozygotes and homozygotes can act as controls on a blot.

All subjects in this study (N=72) were included in the *P* gene mutation screen by Southern blotting. Genomic DNA (10µl) was run on a 0.8% agarose gel in order to determine DNA quality. Some DNA samples had been stored in the laboratory for many years and DNA degradation was possible. If DNA was degraded, the sample could not be used for Southern blotting. DNA concentration was determined using spectrophotometry. For Southern blotting, 5µg genomic DNA was digested with 30 units of *Eco* RI overnight at 37°C in a total digest volume of 50µl. Five microlitres of the digest was run on a 0.8% agarose gel to check that the digest has been successful.

2.3.4.1.2 Agarose gel electrophoresis

Agarose was dissolved in 1X TBE buffer by boiling. The gel mix was allowed to cool before ethidium bromide was added to a final concentration of 0.3µg/ml, for visualisation of DNA. Digested genomic DNA was separated on a 0.8% agarose gel run overnight (approximately 20 hours).

2.3.4.1.3 Southern blotting

Following alkali denaturation and neutralisation of the DNA fragments in the gel, DNA was transferred from the gel onto a nylon membrane (HybondN, Amersham) by Southern

transfer (Southern 1975). DNA was fixed onto the membrane by UV cross-linking or by baking at 80°C for 2 hours.

2.3.4.1.4 Radioactive labelling of the probe: pcDNA3-P

The plasmid, pcDNA3-P, contains the complete cDNA sequence of the *P* gene. This clone was originally generated in the laboratory of Professor Richard Spritz (University of Wisconsin, Madison, Wisconsin, USA), and was a gift to us from his collaborator, Professor Robert Nicholls. The 5.4 kb plasmid vector contains a 2.8 kb *Bam* HI/*Xba* I fragment insert comprising the full length *P* gene cDNA. Approximately 10µg of plasmid DNA was digested overnight at 37°C with 20 units *Bam* HI and 20 units *Xba* I in a total volume of 150µl. The digest was run on a 0.8% low gelling temperature agarose gel (FMC BioProducts) to separate the vector and insert bands. The 2.8 kb insert band was cut out of the gel, and the DNA extracted from the gel using the Gel Extraction kit (QIAGEN). Five microlitres of DNA was run on a 0.8% agarose gel to ascertain the concentration of DNA recovered. Approximately 50ng of insert DNA was radioactively labeled with ^{32}P dCTP using the Megaprime Labeling kit (Amersham).

2.3.4.1.5 Hybridisation and autoradiography

The HybondN membrane, to which the DNA had been fixed, was placed in a glass hybridisation tube. All prehybridisation, hybridisation and washing steps were undertaken in a Hybaid oven. The membrane was prehybridised with 10ml hybridisation solution at 65°C for at least 2 hours. The radioactively labelled probe was denatured at 95°C for 5 minutes and then added to the hybridisation solution. Hybridisation was at 65°C, overnight. Following hybridisation, the blot was washed to remove unbound and non-specifically bound probe. The blot was washed twice with 2X SSC/0.1% SDS for 15 minutes at room temperature, once with 1X SSC/0.1% SDS for 20 minutes at 65°C, and finally, once with 0.2X SSC/0.1% SDS for 20 minutes at 65°C. The membrane was sealed in a plastic bag and then placed in an X-ray cassette (Okomoto) together with Hyperfilm (Amersham) X-ray film. Following exposure at room temperature for 2-6 days, the film was developed in an automatic X-ray developer (Pako 14X).

2.3.4.2 Southern blotting for *HERC2*

The human gene, *HERC2*, lies immediately 5' to the *P* gene on chromosome 15q. Mouse studies have shown that variation in this gene can affect coat colour, leading to a hypopigmented phenotype. To investigate whether mutations in this gene may influence human pigmentation, Southern blot analysis was undertaken to look for large structural rearrangements in *HERC2* in a subset of the patients involved in this study. Individuals with BOCA, unclassified OCA and OCA2 were included in this investigation. The multiple enzyme blot used above was stripped by boiling in water for 30 minutes and then hybridised with a labelled single stranded probe as described above. The *HERC2* probe was a 1002 bp PCR product amplified from intron 70. The primer sequences are given in Table 2.5:

Table 2.5. Primers used to amplify a 1kb fragment from IVS70 of the *HERC2* gene.

DIRECTION	NAME	SEQUENCE
F	HERC2-E	5'-CGT CTC CGT GCG GTC AGG TAA G-3'
R	HERC2-F	5'-GTT CAG CAC GGC CCA GAG TAG G-3'

The PCR reaction was carried out in a total reaction volume of 25µl containing approximately 100ng DNA, 10pmol each primer, 1 unit HiFi Taq DNA polymerase (Roche Diagnostics), 1.25µM each dNTP, 2.5µl DMSO and 2.5µl 10X PCR buffer (number 2). Cycle conditions were 95°C 5 min and then 95°C 1 min 52°C 1 min, 72°C 2.5 min for 30 cycles. PCR product was run on a 0.8% agarose gel stained with ethidium bromide. The 1 kb fragment was excised from the gel and purified using the QIAGEN gel purification kit. DNA was eluted in 10mM Tris buffer. Five microlitres of recovered fragment was run on a 0.8% agarose gel to check recovery and to estimate concentration.

Probe labelling and blot hybridisation was as described above. Of the three enzymes tested (*Bam* HI, *Eco* RI and *Hind* III), *Hind* III gave the most clear banding pattern and multiple bands were detected per individual. Consequently this enzyme was chosen to digest patient DNA samples.

2.3.5 Cloning of the *P* gene promoter region

Following exon by exon mutation analysis of the *P* gene by SSCP analysis, it became apparent that a large proportion of unidentified mutations did not lie within the coding region of the gene. Taking into consideration the BOCA phenotype, which presents with intermediate pigmentation, as well as the observation of ''mild'' OCA2 phenotypes, it was proposed that *P* gene mutations may lie in the 5' control or promoter region of the gene (Kerr *et al*. 2000). These mutations may influence the level of *P* gene expression, without rendering the gene completely non-functional. It was therefore necessary to clone and sequence the *P* promoter region to establish the normal sequence, before variation in this region could be detected. Some of the following work was undertaken by the writer in the laboratory of Professor Robert Nicholls, Case Western Reserve University, Cleveland, Ohio between May and September 1997. The completion of the research was undertaken by the writer subsequent to this visit at the University of the Witwatersrand.

2.3.5.1 Identification of a BAC clone

Prior to the writer's arrival in the laboratory of Professor Robert Nicholls, a Bacterial Artificial Chromosome (BAC) clone had been identified (through library screening using PCR), which contained the 5' exons of the *P* gene, as well as the 3' region of the gene *HERC2* (which was then known as *ERY-1*). Thus, this BAC contained the 5' upstream region of the *P* gene (see Fig 1.9 for a map of the region). The BAC clone, plate number 263 022 (Research Genetics), contained at least exons 1-3 of the *P* gene; *P* exons 12-25 were definitely not present in this BAC clone (verified through exon by exon PCR amplification using the BAC as a source of template DNA). PCR amplification of exons 4 – 11 gave inconsistent results.

2.3.5.2 Subcloning the BAC

A number of restriction enzymes that have a 6 bp recognition site were chosen by the writer in order to identify a restriction fragment that would be of an appropriate length to clone into a plasmid and sequence (about 2-5 kb). For each reaction, 500ng of BAC DNA

was digested for 2 hours with 10 units of *Bam* HI, *Dra* I, *Eco* RI, *Hind* III, *Hinf* I, *Nco* I, *Xba* I or *Xho* I according the manufacturer's instructions. Digest fragments were separated on a 0.8% agarose gel which was run at 45V for 17.5 hours. The DNA was transferred to HybondN membrane as described. The blot was hybridised overnight with the *P* promoter fragment probe (see next section) and the final wash was in a 0.1X SSC solution at 56°C. Autoradiography was at -80°C for 3-6 hours. A second Southern blot was generated using a range of 4 bp restriction enzyme cutters, which generated smaller fragments in the range of 1-3 kb: *Apo* I, *Bst* YI, *Msp* I, *Pst* I or *Tsp* 509I.

2.3.5.3 *P* gene promoter fragment probe

The fragment used to probe the human *P* promoter region was a 270 bp *Eco* RI/*Bam* HI fragment comprising exon 1 of the *P* gene and 180 bp 5' of exon 1. The fragment was cloned into pBluescriptKSI. The insert was excised from the plasmid by digesting approximately 10µg of probe DNA with 25 units *Eco* RI and 25 units *Xba* I in a total digest volume of 200µl for 2-3 hours. The digest was run on a 3% agarose gel, the 270 bp band was excised from the gel, and the DNA recovered using the QIAGEN Gel Extraction kit. Approximately 50ng of insert DNA was radioactively labelled with ^{32}P dCTP using the Rediprime Labeling kit (Amersham).

2.3.5.4 Preparation of BAC DNA for subcloning

Two restriction fragments were chosen as possible candidates for subcloning: the 2.3 kb *Tsp* 509I fragment, and the 7 kb *Xba* I fragment. Digestion with *Tsp* 509I (/AATT) produces sticky ends which are clonable into *Eco* RI (G/AATTC) sites. Since most commercially available vectors will contain an *Eco* RI site in the polylinker, *Tsp* 509I fragments will be amenable to cloning into a number of vectors. Further, the size of the fragment made it an ideal candidate for sequencing as it was not too large. However, taking into consideration the fact that the *Tsp* 509I fragment to which the *P* promoter probe hybridised, may equally as likely have comprised sequence which lay 3' to *P* exon 1, rather than 5' promoter sequence, it was decided to clone a larger fragment, like *Xba* I, simultaneously. Approximately 10µg BAC DNA was digested with 40 units of the

appropriate enzyme in a 100µl digest volume. Following digestion, the DNA was ethanol precipitated (as described, Sambrook *et al*. 1989) and resuspended in 20µl of dH_2O. The digests were loaded onto a 1% agarose gel (made up without the addition of ethidium bromide) and run at 100V for 5 hours. The 2.3 kb *Tsp* 509I band and the 7 kb *Xb*a I band were cut out of the gel using molecular weight markers as a guide to where these fragments would be, and the DNA recovered using the QIAquick gel extraction kit (QIAGEN).

2.3.5.5 Preparation of the cloning vectors

Two different plasmid vectors were used, pZERO and pBluescript. pZERO was the initial vector of choice, but when a positive clone failed to be recovered after numerous attempts, pBluescript was used as an alternate vector. pZERO is a so-called 'suicide' vector where the cloning site interrupts a lethal gene so that only vectors which contain an insert will survive, vectors which simply re-circularize will necessarily commit 'suicide'. A few micrograms of each plasmid vector were digested with either *Eco* RI or *Xba* I. To prevent recircularization of pBluescript, the linearised plasmid was treated with Calf Intestinal Phosphatase (CIP) according to the manufacturer's instructions. This enzyme phosphorylates the 5' overhang, thereby preventing vector re-ligation. Following digestion, the plasmids were run on a 0.8 % agarose gel, and the DNA purified from the gel using the QIAquick gel purification kit (QIAGEN).

2.3.5.6 Ligation of vector and insert

Ligation reactions were undertaken in a reaction volume of 20µl, at 14°C overnight. Sticky end ligation reactions require an optimal DNA concentration of 20-50µg of DNA (in total) per ml of reaction volume, using 2-3 times more insert DNA than vector DNA. Four hundred units of T4 DNA ligase (NEB) were added to each ligation reaction. T4 DNA ligase requires ATP as an energy source. While ATP is present in the enzyme buffer, it is easily destroyed by high temperatures and repeated freeze-thaw episodes. Therefore, 20mM ATP was also added to the ligation reaction.

2.3.5.7 Transformation

Following ligation of the insert DNA into a vector, the DNA was transformed into 50µl competent DH5α *E. coli* cells (GibcoBRL) according to the manufacturer's instructions, and the cells plated onto agar plates containing the appropriate antibiotic. The cells transformed with the pZERO vector were plated onto agar plates made up to contain kanamycin and IPTG. IPTG induces expression of the 'suicide' gene. The cells transformed with the pBS vector were plated onto agar plates containing ampicillin and Xgal. All antibiotic stock solutions, IPTG and Xgal were made up as described (Sambrook *et al*. 1989). The cloning vector, pBluescript, contains part of the coding sequence of the β-galactosidase gene (lacZ). The polycloning site of the vector lies within this region. In the presence of the chromogenic substrate, Xgal, *E.coli* transformed by such a vector form blue colonies. However, if the vector contains an insert, the foreign DNA will have inserted into the polycloning site, and the bacteria carrying recombinant plasmids will therefore not be able to form blue colonies, and will appear white. The plates were incubated at 37°C overnight.

2.3.5.8 Identification of positive transformants

Two methods were used to check if the transformants contained the correct insert. Firstly, individual colonies were picked and grown up overnight in liquid Luria broth (LB) containing the appropriate antibiotic. The plasmid DNA was extracted from the host cells using the QIAGEN plasmid miniprep kit and then digested with either *Eco* RI or *Xba* I to check for the presence of, and size of, an insert fragment. Alternatively, colony filter lifts were performed, whereby the colonies were transferred to a HybondN membrane by gently placing a filter on top of a pre-cooled agar plate, the DNA denatured by floating the filter, DNA side up, on denaturing solution for 2 minutes and then neutralised by floating the filter on neutralising solution. The DNA was fixed to the membrane by UV cross-linking, and the membrane hybridised with the *P* promoter probe as described. The original plate used for the filter lift was either left at room temperature over-night, or at 37°C for a few hours to allow regrowth of the colonies. If a positive transformant was identified following hybridisation with the promoter probe, that colony was picked, grown up in LB containing the appropriate antibiotic, and the DNA extracted using the QIAGEN plasmid miniprep kit.

2.3.6 Inverse PCR

In order to determine the normal sequence of the *P* gene promoter region, the traditional method of cloning the area and then sequencing was adopted (see above). As an alternate strategy, the technique of inverse PCR (Ochman *et al.* 1988, Triglia *et al.* 1988) was attempted. This methodology does not involve cloning of the DNA to be sequenced, hereby circumventing problems encountered with regions of the genome which prove difficult to clone. Fig 2.4 shows a schematic outline of the technique. PCR is used to amplify a region of unknown DNA sequence that is adjacent to a region of known DNA sequence. Two PCR primers are designed within the region of known sequence, but here they are orientated such that chain elongation will proceed outwards, away from each other. The DNA fragment containing this region of known sequence is digested with an appropriate enzyme to produce a fragment of suitable size for PCR amplification. The ends of the fragment are then ligated to form a circle. The primers, when binding to their homologous sequence, will now be orientated such that chain elongation will proceed across the region of uncharacterised sequence. PCR product generated in this way may then be sequenced.

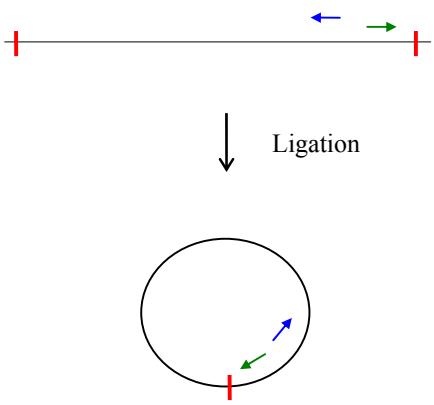


Figure 2.4 Diagram to illustrate the principle of the inverse PCR technique. DNA is digested with a given enzyme (digest site in red). DNA fragments are ligated to form monomeric circles and this circularised DNA is used as template in a PCR reaction.

2.3.6.1 Template digestion

Human genomic as well as cloned (BAC) DNA was used concurrently as template DNA in this experiment. BAC DNA was used as this would give a better chance of success due to the presence of exponentially more copies of the target for PCR being present.

Bgl II was the enzyme of choice in this inverse PCR reaction: There is a *Bgl* II site within exon 1 on the *P* gene (see Fig 2.5). The appropriate amount of genomic or BAC (263 022) DNA was digested overnight with 30 units (genomic) or 10 units (BAC) of *Bgl* II in a total digest volume of 50µl. Following digestion, the DNA was extracted using phenol/chloroform to remove the restriction enzyme (*Bgl* II cannot be heat inactivated). DNA was resuspended in 50 µl dH_2O.

2.3.6.2 Ligation

The ligation step requires the DNA to be in a dilute concentration so as to favour the formation of monomeric circles. The ligation reaction was set up in a volume of 50µl and included 100ng of genomic DNA or 1-5ng of BAC DNA and 400 units of T4 DNA ligase (NEB). The ligation reaction took place at 16°C, overnight. The ligase enzyme was inactivated by heating at 68°C for 15 minutes.

2.3.6.3 Inverse PCR

The *P* gene exon 1 forward primer was suitable for use here and the second primer (in the reverse orientation) was positioned as shown in Fig 2.5. The sequence of this primer was 5'-CCACAAAGCTGCCCACATGG-3'.

The concentration of template DNA optimal for inverse PCR is approximately 2 µg/ml of genomic DNA or 20 ng/ml of cloned DNA. The inverse PCR reaction was set up on ice and included 1-10 µl of the ligation reaction as template DNA, 15 pmoles of each primer and units of Expand HiFi DNA polymerase (Boehringer Mannheim) in a total volume of 50 µl. The PCR conditions were 94°C for 2 min, then 94°C for 15 sec, 56°C for 30 sec and

68°C for 4 min for 10 cycles; followed by 94°C for 15 sec, 56°C for 30 sec and 68°C for 4 min + 20 sec each cycle, for 20 cycles. The final elongation step was 72°C for 7 minutes.

The PCR products were sequenced using the ABI 377 automated sequencer, as described above.

gi|773301|gb|U19152|HSPPROT01 Human P protein (P) gene, exon 1

gaattcatgccgggggtgcggaacgggcgagctgtgggcggggccatgtgggcagctttgtgggcgtggccgggaggg

gcgtgctacggcgtggccgcactgtgggcggggccgagtggggagtgctgtgggcggggccgggccgggggggtgcc

tgggtggggccgctgcgggagcgccgagctcccTCTGAGTTCTTACTTCGAAGGCTGTGCTCC

GCTCACCATCCAGAGCGGAGGTGCGGACCTTAAACTCACTCCTGGAGAAA/G

ATCTGCAAGTGCGCAGgtaaagtgcacg

Figure 2.5 Nucleotide sequence of exon 1 and immediate upstream region of the human *P* gene (Lee *et al*. 1995). Uppercase letters denote the exon, primer positions and orientations are indicated by arrows and the *Bgl* II site is highlighted in red.

2.3.7 The *P* gene and tuberculosis association

This part of the study was aimed at typing five intragenic *P* gene markers in a group of South African black TB patients and a group of Cape Coloured TB patients, plus appropriate controls for each group. The five markers typed were the SNP, R305W, the common black OCA2 mutation, the 2.7 kb deletion, and three microsatellite markers – D15S1533 (IVS17), D15S1536 (IVS20) and D15S1537 (IVS24).

The 2.7 kb deletion was screened for as described above.

2.3.7.1 Detection of the R305W polymorphism

The C to T base change in exon 9 of the *P* gene alters the recognition sequence of the enzyme *Msp* I. *Msp* I cuts the sequence CCGG and will not cut the sequence CTGG. In order to test for this variant, exon 9 of the P gene was amplified by PCR under the following conditions: the reaction was carried out in a total volume of 20µl containing approximately 100ng DNA, 2pmol each primer, 1 unit *Taq* DNA polymerase (Roche Diagnostics), 1.25µM each dNTP and 2µl 10X PCR buffer. Primer sequences were as published (Lee *et al*. 1995). Cycle conditions comprised an initial denaturation step of 94°C for 2 min, followed by 35 cycles of 94°C 40 sec, 55°C 1 min, 72°C 1 min. Eight to ten microliters of PCR product was run on a 3% HGT agarose gel in order to check that the PCR reaction had worked. The remaining 10µl of PCR product was digested with 10 units of *Msp* I (Roche Diagnostics) in a total digest volume of 20µl, for approximately 3 hours. The entire digest volume was run on a 4% HGT agarose gel in order to visualise digest products. The PCR product was 272 bp and digestion results in two fragments of 190 bp and 82 bp. Therefore, following gel electrophoresis of digest products, presence of the R allele will result in the two digest products being visible, while presence of the W allele will result in the uncut 272 bp band being present.

2.3.7.2 *P* gene intragenic marker screen

Five marker systems, internal to the *P* gene, were studied in an attempt to establish whether or not the *P* gene plays a role in susceptibility/resistance to the infectious disease tuberculosis. Patients plus ethnically matched controls were typed for alleles at 5 loci: the 2.7 kb deletion, the R305W polymorphism and the microsatellite $(AC)_n$ repeats D15S1533, D15S1536 and D15S1537.

2.3.7.2.1 Typing the microsatellite repeats

For markers D15S1533, D15S1536 and D15S1537, DNA was PCR amplified in 25µl final reaction volume with 125ng genomic DNA, 1xPCR buffer with $MgCl_2$ (Roche), 5pm of each sense and antisense primer, 1.25mM dNTPs, 1unit Taq polymerase (Roche). Primer

sequences are given in Table 2.6:

Table 2.6. Primer sequences used to amplify three repeat regions intragenic to the *P* gene. The D15-SEG11 primer pair amplify the D15S1533 repeat segment; the P-IVS20 primer pair amplify the D15S1536 repeat segment; and the P-IVS24 primer pair amplify the D15S1537 repeat segment.

PRIMER NAME	FLOURESCENT LABEL	SEQUENCE
D15-SEG11 F	FAM (blue)	5'-/5FAM/CTT GGC AAC ATC CCT GTA TCA -3'
D15-SEG11 R		5'-TGA ATG CCA TTA TTT CAT TCC TT -3'
P-IVS20 F	HEX (green)	5'-/5HEX/TAA TAC CTG GGT GAC AAA ATA AT -3'
P-IVS20 R		5'-ATA GGA GTA AAT AAG ATA GAC -3'
P-IVS24 F	FAM (blue)	5'-/5FAM/GCC GCA CTC AGC CAG CAT A -3'
P-IVS24 R		5'-CAG CAA TCC AGC CCT CTT TAG TTT -3'

A small volume (5µl) of PCR product was run on a 3% agarose gel to check if the PCR amplification was successful. The sample was diluted depending on the concentration of the PCR product (estimated from the agarose gel) and 1µl of the diluted PCR product was mixed with 2.3µl dextran-formamide loading buffer and 0.2µl internal lane size standard, GENESCAN-500 ROX (ABI). The PCR products, formamide loading dye and internal size standard were denatured for 2 minutes at 95°C, then electrophoresed on a 4.3% denaturing polyacrylamide gel (36cm WTR) at a constant 200W for 2 hours using a ABI model 377 Automated DNA Sequencer. The two markers labelled with FAM (SEG 11/IVS 17 and IVS 24) could be run on the same gel as the PCR products were of sufficiently different sizes. Data were analysed automatically using Version 3.1.2 Genescan and Version 2.5 Genotyper Analysis Software (ABI) to produce an electropherogram for each sample showing the size (bp) of each amplification product.

2.3.7.3 Statistical tests

Data generated in the TB host susceptibility study was analysed using statistical tests in order to establish significance of the results obtained.

The null hypothesis

The null hypothesis (designated the symbol H_0) is a hypothesis of no differences, i.e. that there are no differences between, for example, the two sample groups under investigation. The null hypothesis is usually formatted for the sole purpose of being rejected but it acts as the statement against which the experimenter will test his/her research hypothesis (designated the symbol H_1). Thus when we want to make a decision about a theory we test H_0 (the null hypotheses) against H_1 (the alternative or test hypothesis).

Level of significance

When can H_1 be accepted and H_0 be rejected?
A researcher can reject H_0 in favour of H_1 if a statistical test yields a valuable whose corresponding probability is so small so as to make the probability of H_0 being true correspondingly small. The probability associated with the occurrence of a particular value is designated the symbol *p*. This probability is referred to as the level of significance. Numerical values for *p* that are considered statistically significant are arbitrarily chosen but common values used are 0.05 and 0.01. If a significant *p* valu*e* is chosen at $p \leq 0.05$ then the researcher is acknowledging that 5% of the time there will be calculations showing differences (just by chance), so the number of differences found must be greater than or equal to 5 in 100 in order to be able to reject H_0 and accept H_1.

2.3.7.3.1 The χ^2 Test

The Chi-squared (χ^2) test can be used to test whether a significant difference exists between an observed number of objects or subjects falling into a category (H_1) and an expected number deduced from the null hypothesis (H_0). The χ^2 methodology is used to test whether the observed frequencies are sufficiently close to the experimental ones, in which case the H_0 holds. In order to reject the null hypothesis, the observed frequencies need to be sufficiently different from the experimental ones.

A χ^2 value is generated using the following formula:

$$\chi^2 = \sum_{i=1}^{k} \frac{(Oi - Ei)^2}{Ei}$$

Where O*i* = observed number of individuals (specimens/cases)

E = expected number of individuals (specimens/cases)

Σ = sum of values over all (k) categories

If the observed values (O) are close to the expected values (E) then (Oi – E) will be small and consequently χ^2 will be small. Therefore, the larger χ^2 is, the more likely it is that the observed frequencies are significantly different from the population on which the null hypothesis is based.

2.3.7.3.2 Fisher's exact test

Fisher's exact test is used as an alternative to the χ^2 test and generates a *p* value. The exact test is a statistical significance test used to analyse data when the sample sizes are small. When large sample sizes are available for analysis, a χ^2 test can be used. However, this test is not suitable when the 'expected' values in any of the cells in a 2x2 contingency table is below 10 and there is only one degree of freedom.

CHAPTER 3

Results and Discussion: OCA mutation screen (*P* and *TYR* genes)

3.1 Results: OCA mutation screen

3.1.1 Mutation detection in OCA subjects

3.1.1.1 Detection of the 2.7 kb deletion mutation

All subjects collected for albinism-related research by workers in our laboratory were routinely tested for the 2.7 kb deletion mutation. For the present study, a subset of these was chosen - that is, subjects who were heterozygous for the deletion or subjects who did not carry the deletion on either chromosome. There were 72 such subjects (see Table 2.1). The assay was a PCR reaction, the products of which were run on a 2% agarose gel (see Fig 3.1).

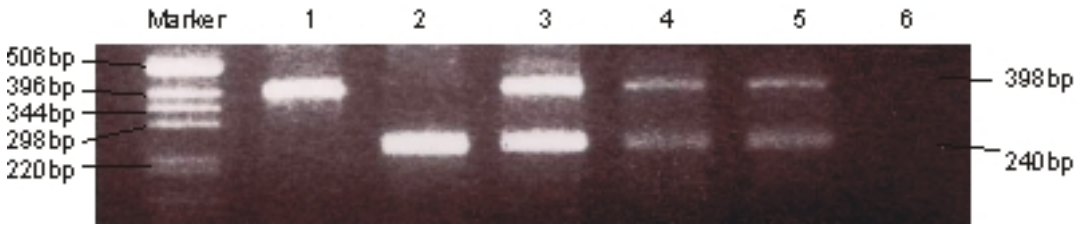


Figure 3.1 Agarose gel of PCR products detecting the 2.7 kb *P* gene deletion mutation. The smaller 240 bp fragment represents absence of the deletion (normal allele) and the larger 398 bp fragment represents presence of the deletion (mutant allele). Marker lane: 1kb marker (Roche Diagnostics), Lane 1, individual homozygous for the deletion; lane 2, individual does not carry the deletion; lanes 3, 4, 5, individuals are heterozygous for the deletion; lane 6, blank PCR control.

Table 3.1. Results of the 2.7 kb deletion screen.

SUBJECT GROUP	TOTAL	2.7kb DELETION STATUS	
		N/N*	N/del*
Black OCA	39	13	26
- BOCA	9	0	9
- Unclassified	15	10	5
White Unclassified	9	8	1
Total	72	31	41

*N= Normal (absence of the 2.7kb deletion mutation)

del= Deletion present

3.1.1.2 Detection of the R305W variant

The black OCA2 individuals (N=39) and the BOCA individuals (N=9) involved in the OCA mutation screen were tested for the R305W polymorphism. Controls for the black population were randomly selected black individuals sent into the laboratory for paternity testing (N=45). Control figures for the white population had been published (Lee *et al*. 1995). The assay was a PCR reaction followed by restriction digestion with *Msp* I (see Fig 3.2.).

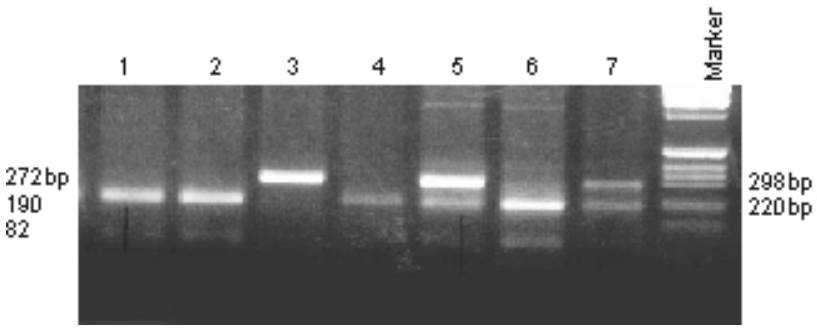


Figure 3.2 Detection of the R305W P gene variant. The uncut PCR product is 272 bp in length. Following digestion with *Msp* I, the normal sequence is cut and two fragments of 190 bp and 82 bp are generated. The small 82 bp fragment is often not visible on the gel. Presence of the W variant abolishes the restriction site. Digest products were run on a 3 or 4% HGT agarose gel. Lanes 1, 2, 4, 6, individuals are homozygous for the R allele; lane 3, individual is homozygous for the W allele; lanes 5, 7, individuals are heterozygous carrying one R and one W allele. Marker: 1 kb marker (Roche Diagnostics).

Results of the R305W screen on the subjects in the present study in comparison with those presented by Lee *et al*. (1995) on a cohort of African Americans are given in Table 3.2.

Table 3.2 Frequencies of the R305W variant in various population samples.

POPULATION GROUP	n	R	W	REFERENCE
Black OCA2	78	0.42	0.58	This study
Normal African American	100	0.10	0.90	Lee *et al*. 1995
Normal SA blacks	90	0.91	0.09	This study
Normal Caucasoid	35	0.83	0.17	Lee *et al*. 1995
BOCA	18	0.66	0.33	This study

n=number of alleles

R= argenine allele frequency

W= tryptophan allele frequency

The W variant occurs at a frequency of 0.58 (45/78 chromosomes) in the black OCA2 sample in the present study. This is significantly different from the frequency in normally pigmented African-American individuals (0.90) as reported by Lee *et al*. (1995). In order to determine whether the frequency of this variant is different to normally pigmented southern African blacks, 45 such individuals were screened for the R305W variant. It was found that 8/90 chromosomes (0.09) carried the tryptophan (W) variant. The frequency of this amino acid substitution in a normal Caucasoid population sampled by Lee *et al*. (1995) is 0.17. Six out of the nine BOCA patients were heterozygous for the R305W variant. Out of a sample of eight previously studied individuals homozygous for the 2.7 kb deletion (Stevens *et al.* 1995), six were found to be homozygous for the R305W variant, one was heterozygous and one did not carry the W allele at all (i.e. 13/16 chromosomes carrying the 2.7 kb deletion also carried the R305W variant). It would appear that the 2.7 kb deletion mutation and the R305W tryptophan (W) allele are co-segregating much of the time, i.e. they are in linkage disequilibrium.

3.1.1.3 PCR-SSCP analysis

Denatured, radioactively labelled PCR product was run on MDE polyacrylamide gels, made up with or without glycerol, in order to identify possible sequence variants between PCR fragments amplified from different individuals. All subjects involved in the OCA mutation screen were screened for *P* gene mutations. A subset of these subjects, namely the Caucasoid subjects (N=9), the black subjects with an unclassified form of OCA (N=15) and the BOCA subjects (N=9) were screened for *TYR* mutations in addition to *P* gene

mutations. Examples of SSCP autoradiographs are given in Fig 3.3.

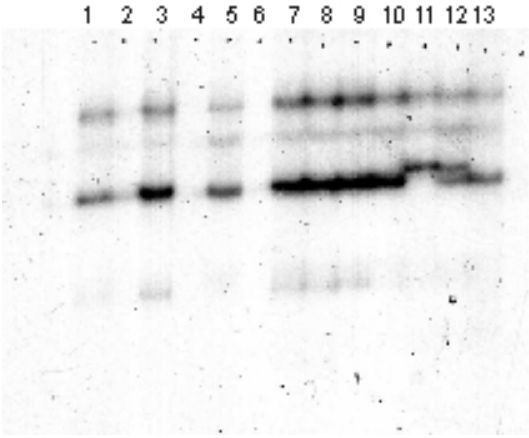


Figure (a)

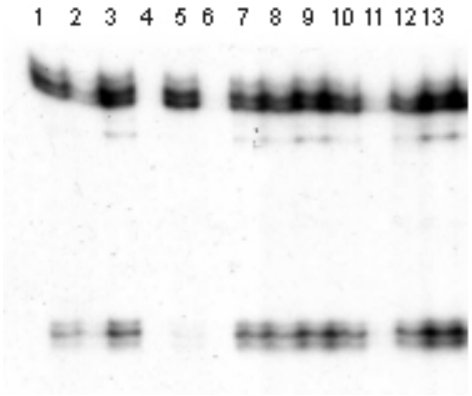
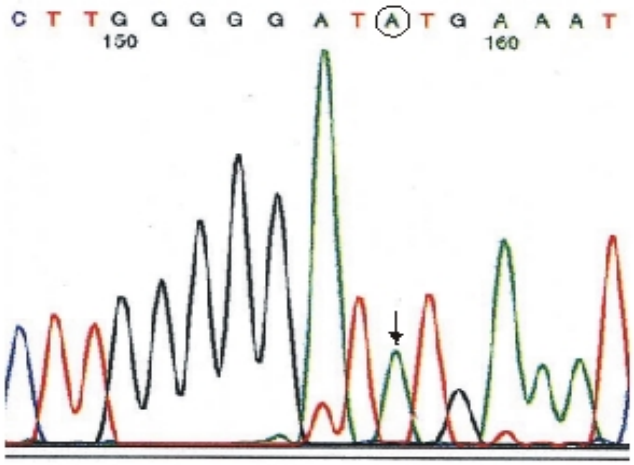


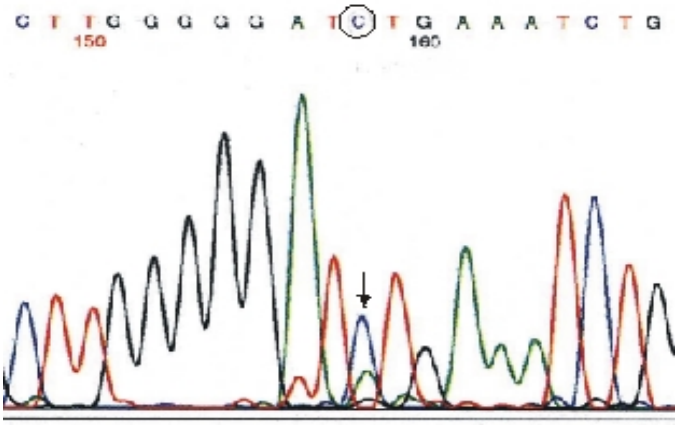
Figure (b)

Figure 3.3 Examples of autoradiographs of SSCP gels. PCR product was radioactively amplified from *TYR* exon 1B. The same products were run on MDE gels with (Fig a) or without (Fig b) glycerol. Note that the two different gel conditions produce different band patterns. For this particular PCR product, band shifts are seen only in the gel made up with glycerol. In Fig (a), lanes 1, 2, 3, 5, 7, 8, 9, 10 and 13 show SSCP variant type 1, lane 11 shows SSCP variant type 2 and lane 12 shows SSCP variant type 3.

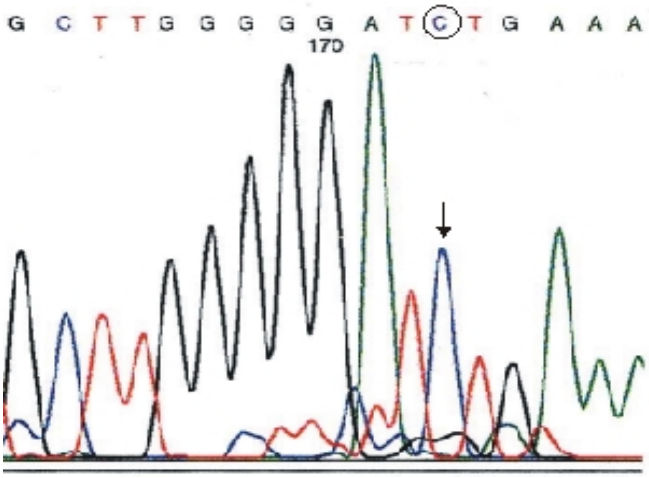
DNA from individuals in whom band shifts were detected was then sequenced in order to determine if one or more sequence variants were present in this exon. It is possible that band shifts are seen on SSCP gels, with no corresponding sequence variation. Using the example above, sequencing of variants using the ABI 377 sequencer gave the following results:



(a)



(b)



(c)

Figure 3.4 Sequencing of exon 1B of the *TYR* gene shows that SSCP variant type 3 represents the normal published sequence (a), variant type 2 is heterozygous for the Y192S (TAT→TCT) polymorphism (b) and variant type 1 represents an individual homozygous for the Y192S polymorphism (c). (The database reference sequence is TAT. The TCT variant is, however, more common in this population and is actually the wild-type sequence.)

3.1.1.3.1 SSCP analysis for the *P* gene

A total of 72 individuals were screened for mutations using the method of SSCP analysis following PCR amplification of each of the 25 *P* gene exons. Ten mutations (not counting the 2.7 kb deletion mutation) which are likely to be pathogenic were found, nine of which are novel.

In the group of black OCA2 patients, four mutations which are likely to be pathogenic were found: A334V, 614delA (in a CAR individual), 683insT and 727insG (Table 3.3, Fig 3.5). 614delA, 683insT and 727insG are all frame-shift mutations. The 614delA mutation causes a premature STOP codon in exon 20 and both the 683insT and the 727insG mutations result in premature STOP codons in exon 22, leading to a truncated P protein. Mutations were identified in four individuals with an unusual hypopigmentation phenotype. E678K was found in the homozygous state in an individual from the CAR with very light coloured skin and hair (a light tan colour). A second individual was found to be a compound heterozygote for the I370T and the L688F mutations. Phenotypically, this individual has light brown eyes and very light coloured hair (yellowish-white). A third individual was found to be heterozygous for the I370T mutation. His skin was a reddish-brown colour, like that of an individual with ROCA. He had tested negative for *TYRP1* mutations (Manga, 1997). The fourth individual is the half-sib of an OCA2 individual. The OCA2 patient is a compound heterozygote for the 2.7kb deletion and the mutation, 614delA. The half-sib with an unusual phenotype has green-blue eyes and brown skin, much lighter than that of the parents or the unaffected sibs. This individual is heterozygous for the 614delA mutation. In the case of patients 3 and 4, it is assumed that a second mutation is present but could not be identified. This assumption is based on the fact that OCA is an autosomal recessive condition and it has been shown that carriers do not exhibit a significant reduction in skin pigmentation (Roberts *et al*., 1986).

Three *P* gene mutations were found in the Caucasoid sample (Table 3.3; Fig 3.5): IVS 14-2 (a→g), a splice site mutation, V350M and P743L (the latter previously reported as a pathogenic mutation by Lee *et al*. (1994), in an African American individual). The A334V, V350M, I370T, E678K, L688F and P743L mutations are considered to have a high

likelihood of being pathogenic, as the normal amino acid residue shows high cross-species conservation (Table 3.3) and the variant alleles were not detected by Lee *et al.* (1995) in a control group of normally pigmented individuals. (Local controls were not screened for any of these mutations, but this would be something to consider including in a future study.)

No further BOCA mutations (besides the deletion mutation) were identified.

Table 3.3 Putative pathogenic *P* gene mutations identified in this study in OCA individuals from Sub-Saharan Africa.

Pathogenic variant	Position in protein	Cross-species conservation[1]	Population group (present study)	Geographic origin	Reference where first described
2.7 kb deletion	Truncated after 2nd TM domain		Black OCA2, BOCA, Black unclassified OCA	South Africa, CAR, Zambia	Durham-Pierre *et al.* (1994)
A334V	2nd TM domain	H, M, *Mtb*A, *Mtb*B, *Mlep*	Black OCA2	Zambia	This study
V350M	Cytoplasmic loop	H, M, *Mtb*A, *Mtb*B, *Mlep*	Caucasoid unclassified OCA	South Africa	This study
I370T	3rd TM domain	H, *Mtb*A, *Mlep*, *Syn*	Black unclassified OCA	CAR	This study
IVS 14-2 (a→g)	Splice mutation		Caucasoid unclassified OCA	South Africa	This study
614delA	Truncated after 8th TM		Black OCA2, Black unclassified OCA	CAR	This study
E678K	Cytoplasmic loop	H, M, *Mtb*A, *Mtb*B, *Mlep*	Black unclassified OCA	CAR	This study
L688F	9th TM domain	H, M, *Mtb*A, *Mtb*B	Black unclassified OCA	CAR	This study
683insT	Truncated after 9th TM		Black OCA2	South Africa	This study
727insG	Truncated after 9th TM		Black OCA2	Zambia	This study
P743L	Cytoplasmic loop	H, M, *Mtb*A, *Mtb*B, *Mlep*	Caucasoid unclassified OCA	South Africa	Lee *et al.* (1994)

[1]Human (H), Mouse (M), *Mycobacterium tuberculosis* (has two homologues of the *P* gene, designated *Mtb*A and *Mtb*B), *Mycobacterium leprae* (*Mlep*) and *Synechosystis* (*Syn*).

The 2.7 kb deletion was found in the heterozygous state in 27 Black OCA2 individuals, in 8 BOCA individuals, in 5 Black Unclassified individuals and in one Caucasoid unclassified

individual. All the other mutations listed in Table 3.3 were found in the heterozygous state in one individual only, except for: I370T which was found in two CAR individuals; 614delA which was found in two black individuals, one Black OCA2 and one Black unclassified OCA; E678K which was found in the homozygous state in a CAR individual; and 683insT which was found in two Black OCA2 individuals.

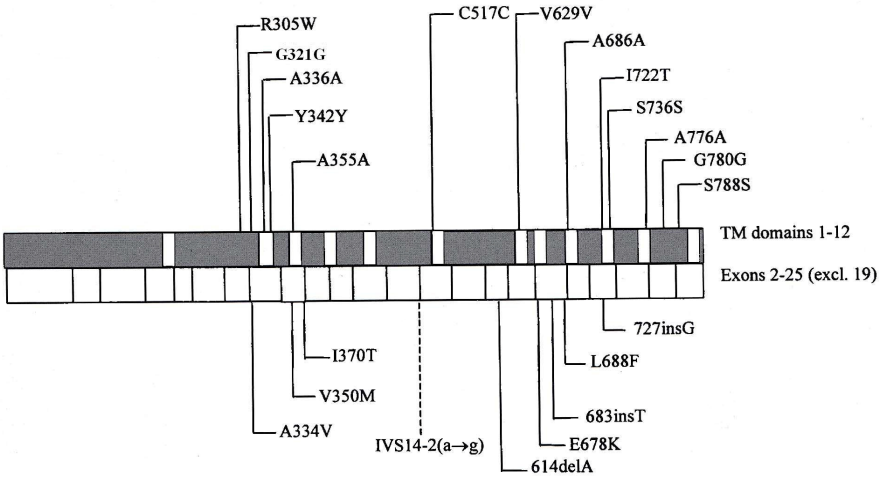


Figure 3.5 Diagrammatic representation of the *P* gene cDNA showing positions of variants. Non-pathogenic variants are labelled above the diagram and pathogenic mutations below the diagram. TM=transmembrane domain.

A large number of non-pathogenic variants were identified in both the black (data summarized in Table 3.4) and Caucasoid (Table 3.5) groups. In this study, 27 non-pathogenic sequence variants were identified, of which 7 are novel (the remainder had been previously reported by Lee *et al*., 1995). Most of these variants are silent, or occur within intronic sequence, and so are unlikely to affect protein structure or function.

Besides the R305W variant, the other amino acid substitution mutation which was classified as non-pathogenic is I722T. This amino acid residue is not conserved evolutionary, and the variant allele occurs in both the Caucasoid (0.26) and American black (0.07) populations (Lee *et al*. 1995).

Table 3.4. Non-pathogenic sequence variants of the human *P* gene in black individuals.

NON-PATHOGENIC VARIANTS	OCA2 (n=78)	BOCA (n=18)	UNKNOWN OCA (n=30)	NORMAL[1]	REFERENCE
IVS 5-18 (c→t)	1 (0.01)	0	0	(0.02)	Lee *et al*. 1995
IVS 5-39 (t→c)	1 (0.01)	0	0		This study
R305W (CGG→TGG)	45 (0.58)	6 (0.33)	7 (0.23)	(0.9)[2]	Lee *et al*. 1995
G321G (CAG→CAA)	0	1 (0.05)	0		This study
A336A (GCC→GTC)	3 (0.04)	1 (0.05)	0		This study
Y342Y (TAC→TAT)	3 (0.04)	1 (0.05)	0	(0.01)	Lee *et al*. 1995
A355A (GCG→GCA)	7 (0.09)	7 (0.39)	5 (0.17)	(0)	Lee *et al*. 1995
IVS 12+19 (g→a)	1 (0.01)	0	0		This study
IVS 11-4 (a→g)	39 (0.50)	11(0.61)	12 (0.40)	(0.38)	Lee *et al*. 1995
IVS 13+4 (c→a)	1 (0.01)	0	0	(<0.01)	Lee *et al*. 1995
IVS 13+26 (a→g)	11(0.141)	0	1 (0.03)	(0.33)	Lee *et al*. 1995
IVS 13-15 (t→c)	30(0.385)	0	3 (0.10)	(1.0)	Lee *et al*. 1995
IVS 14+5 (g→a)	14(0.179)	0	0		This study
C517C (TGC→TGT)	0	0	2 (0.07)	(0.20)	Lee *et al*. 1995
IVS 16-49 (c→a)	1 (0.01)	0	0		This study
IVS 16-47 (g→a)	24 (0.31)	5 (0.28)	5 (0.17)	(0.06)	Lee *et al*. 1995
V629V (GTG→GTT)	3 (0.04)	1 (0.05)	1 (0.03)	(0.07)	Lee *et al*. 1995
IVS 18+24 (c→g)	1 (0.03)	1 (0.05)	1 (0.03)		This study
A686A (GCA→GCC)	5 (0.06)	6 (0.34)	3 (0.10)	(0.29)	Lee *et al*. 1995
IVS 21+22 (a→t)	1 (0.01)	4 (0.22)	2 (0.07)	(0.03)	Lee *et al*. 1995
S736S (TCG→TCA)	2 (0.03)	2 (0.11)	0	(0.07)	Lee *et al*. 1995
I722T (ATA-ACA)	4 (0.05)	2 (0.11)	0	(0.07)	Lee *et al*. 1995
IVS 22+18 (a→g)	0	0	1 (0.03)	(0.07)	Lee *et al*. 1995
IVS 22+25 (g→c)	1 (0.01)	0	0	(0)	Lee *et al*. 1995
A776A (GCT→GCC)	27 (0.35)	8 (0.44)	3 (0.10)	(0.69)	Lee *et al*. 1995
G780G (GGC→GGT)	8 (0.10)	10(0.56)	4 (0.13)	(0.07)	Lee *et al*. 1995
S788S (TCG→TCA)	4 (0.05)	5 (0.28)	2 (0.07)	(0)	Lee *et al*. 1995

n=number of chromosomes

[1]Frequencies in unaffected individuals reported by Lee *et al*. 1995.

[2]This is the figure for unaffected African-Americans. The writer found the frequency of the variant allele to be (0.09) in normally pigmented black Africans.

Table 3.5. Non-pathogenic sequence variants of the human *P* gene in Caucasoid individuals.

NON-PATHOGENIC VARIANTS	UNKNOWN OCA (n=18)	NORMAL[1]	REFERENCE
R305W (CGG→TGG)	3 (0.17)	(0.17)	Lee *et al*. 1995
A355A (GCG→GCA)	20 (0.66)	(0.93)	Lee *et al*. 1995
IVS 11-4 (a→g)	16 (0.88)	(0.91)	Lee *et al*. 1995
C517C (TGC→TGT)	4 (0.22)	(0.63)	Lee *et al*. 1995
IVS 16-47 (g→a)	11 (0.61)	(0.13)	Lee *et al*. 1995
V629V (GTG→GTT)	2 (0.11)	(0)	Lee *et al*. 1995
IVS 22+25 (g→c)	4 (0.22)	(0.28)	Lee *et al*. 1995
A776A (GCT→GCC)	3 (0.17)	(0.36)	Lee *et al*. 1995
G780G (GGC→GGT)	8 (0.44)	(0.27)	Lee *et al*. 1995
S788S (TCG→TCA)	4 (0.22)	(0.09)	Lee *et al*. 1995

n=number of chromosomes

[1] Figures from unsffected Caucasoid individuals as reported in Lee *et al*. 1995

3.1.1.3.2 SSCP analysis for *TYR*

A total of 33 individuals with albinism were screened for *TYR* mutations, (15 black individuals with an unclassified form of albinism, 9 BOCA individuals and 9 Caucasoid individuals). All 5 *TYR* exons (the first exon in two pieces), were analysed by PCR-SSCP analysis. Four mutations were identified, all in the Caucasoid group. One individual was found to be a compound heterozygote for the mutations, E294K and A490D. Compound heterozygosity in a second individual was also shown upon identification of the P431T mutation and the T373K mutation. Three of these variants have been previously reported in the literature and are classified as pathogenic (see Table 3.6). The variant A490D has not been previously reported. However, variation at this specific codon has been shown to be pathogenic – the mutation A490G was reported by King *et al*. (2003). Three polymorphisms were identified, Y192S, L279L and R402Q, one of which (L279L) is novel (see Table 3.6).

Table 3.6 Variants in the *TYR* gene identified in this study in OCA individuals.

VARIANT	POSITION	POPULATION GROUP	CLASSIFICATION	REFERENCE
E294K	Exon 2	Caucasoid	Pathogenic	Gershoni-Baruch *et al.* 1994
T373K	Exon 3	Caucasoid	Pathogenic	Spritz *et al.* 1990
P431T	Exon 4	Caucasoid	Pathogenic	Spritz *et al.* 1993
A490D	Exon 5	Caucasoid	Pathogenic	This study
Y192S	Exon 1	Caucasoid	Non-pathogenic	Giebel and Spritz 1990
L279L	Exon 2	BOCA, Black unclassified OCA	Non-pathogenic	This study
R402Q	Exon 4	Black unclassified OCA, Caucasoid	Non-pathogenic	Tripathi *et al.* 1991

All the pathogenic mutations listed in Table 3.6 were found only once and were present in the heterozygous state in each individual.

A listing of all published variants (both pathogenic and non-pathogenic) for both the *P* gene and *TYR* is available on the Albinism Database website (http//albinismdb.med.umn.edu).

3.1.1.4 Southern blotting

3.1.1.4.1 Southern blotting for the *P* gene

Three different genomic DNA enzyme digests were initially undertaken as a trial to ascertain which enzyme might be most suitable to use to screen the cohort of patients for mutations in the *P* gene (see Fig 3.6).

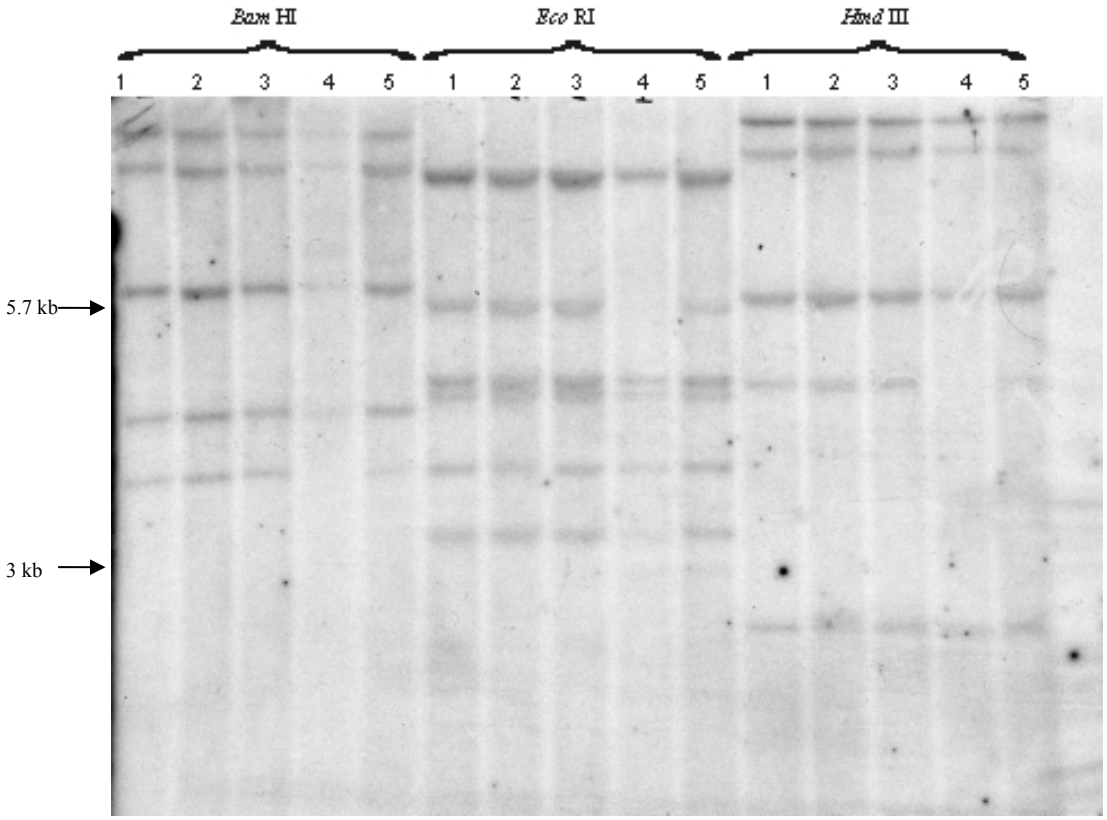


Figure 3.6 Autoradiograph of a Southern blot using 3 different enzymes to digest genomic DNA which was probed with a 2.8 kb fragment comprising the full length P gene cDNA. Lanes 1-5 Bam HI digests; lanes 6-10 *Eco* RI digests, lanes 11-15 *Hind* III digests. In each set of digests, the first three lanes contain DNA from normally pigmented control individuals, the fourth lane contains DNA from an individual homozygous for the 2.7 kb *P* gene deletion mutation and the fifth lane contains DNA from an individual heterozygous for the 2.7 kb *P* gene deletion mutation.

It can be seen that the 2.7 kb deletion is contained within an Eco RI fragment – the 5.7kb band, present in normal controls, is missing in the homozygous deletion patient. This patient carries an additional 3 kb band when compared to normal controls. (This smaller band is seen more clearly on Figure 3.7 which has a darker exposure.) In heterozygous individuals, both bands will be seen – the 5.7 kb fragment representing the normal allele and the 3 kb fragment representing the deleted allele.

For reasons discussed in the methods section, *Eco* RI was selected as the enzyme of choice for the subject screen. See Fig 3.7.

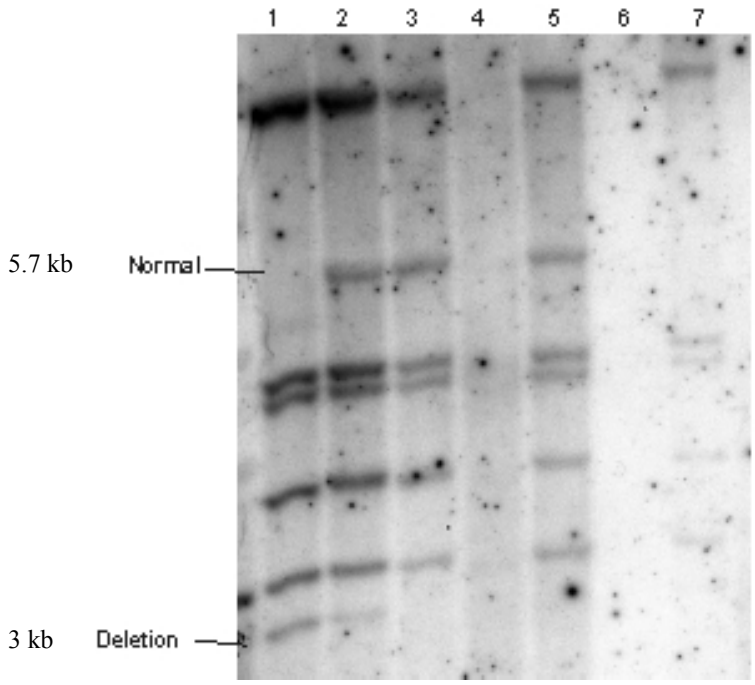


Figure 3.7 Portion of an autoradiograph of a Southern blot used to screen OCA patients for *P* gene mutations. Lane 1, 2.7 kb deletion homozygote control; lane 2, 2.7 kb deletion heterozygote control, lanes 3-7, subject DNA.

Using the method of Southern blot analysis, where subject genomic DNA had been digested with *Eco*RI, no major mutations or rearrangements within the *P* gene were found in any of the population groups tested.

3.1.1.4.2 Southern blotting for *HERC2*

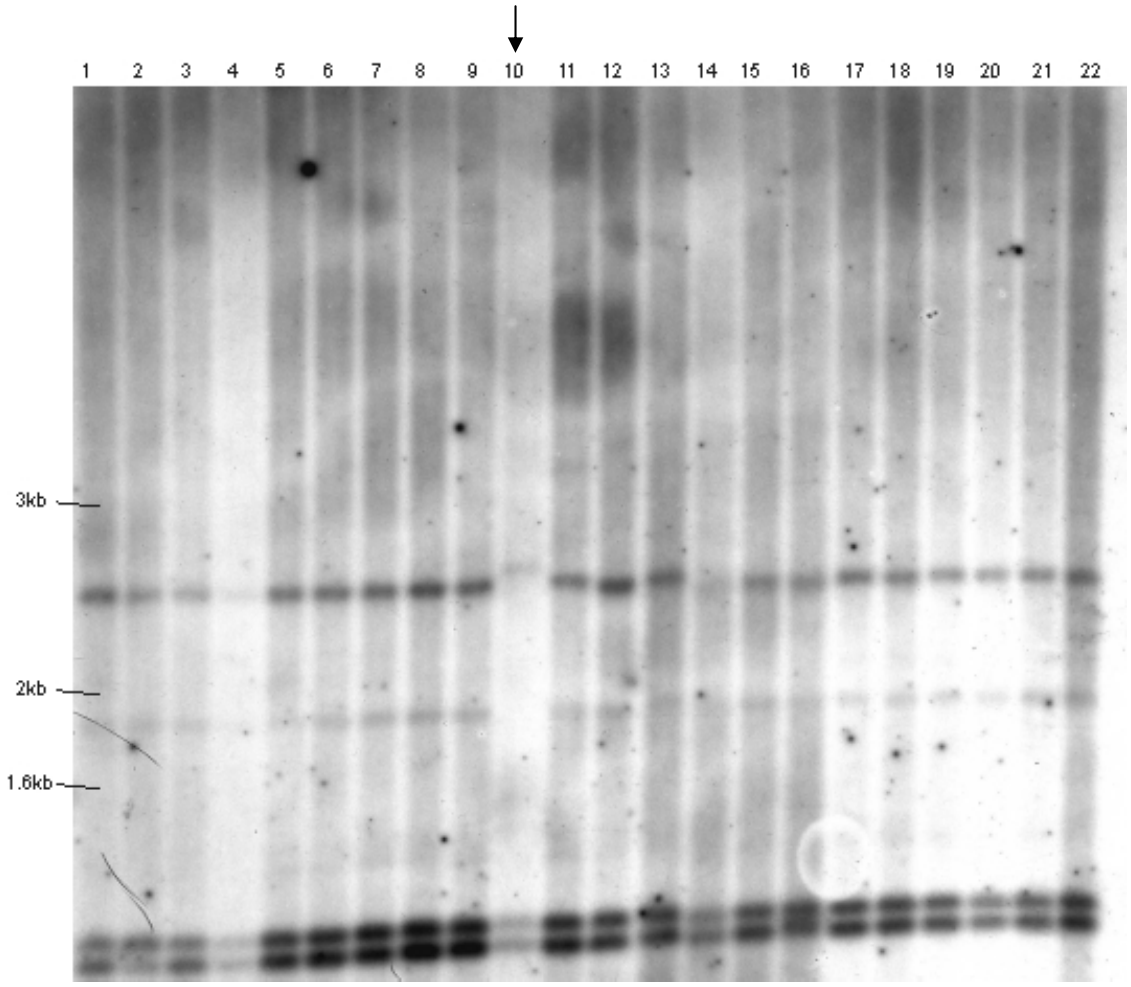


Figure 3.8 Autoradiograph of a Southern blot used to screen OCA patients for HERC2 mutations. Lane 1, 2 control individuals for the *P* gene blot; lanes 3-22, patient DNA. Lane 10 shows possible variation, with absence of the 1.8 kb and 2.6 kb fragments but presence of a larger fragment of approximately 2.7 kb. DNA in lane 10 was from an individual affected with BOCA.

Using the method of Southern blot analysis, where subject genomic DNA (from all BOCA, unclassified black and black OCA individuals) had been digested with *Hind* III and probed with a DNA fragment which hybridised to the 3' end of the gene, a possible rearrangement within *HERC2* was found in an individual with brown OCA (see Fig. 3.8, lane 10).

3.1.2 Sequencing of the *P* gene promoter region

3.1.2.1 Sub-cloning the BAC

Two initial Southern blots were undertaken in order to identify a restriction fragment of the BAC insert that would be of an appropriate length to subclone into a plasmid and sequence. A blot of multiple enzymes which are 6 bp cutters and a blot of multiple enzymes which are 4 bp cutters were generated (results not shown). Enzymes with 6 bp recognition sequences were *Bam* HI, *Dra* I, *Eco* RI, *Egl* II, *Hind* III, *Hinf* I, *Nco* I, *Xba* I and *Xho* I. Enzymes with 4 bp recognition sequences, producing smaller digest fragments, were *Apo* I, *Bst* YI, *Msp* I, *Pst* I, *Sac* I, *Tsp* 509I. A subset of these enzymes was chosen and a second blot subsequently undertaken. This blot is shown in Fig 3.9.

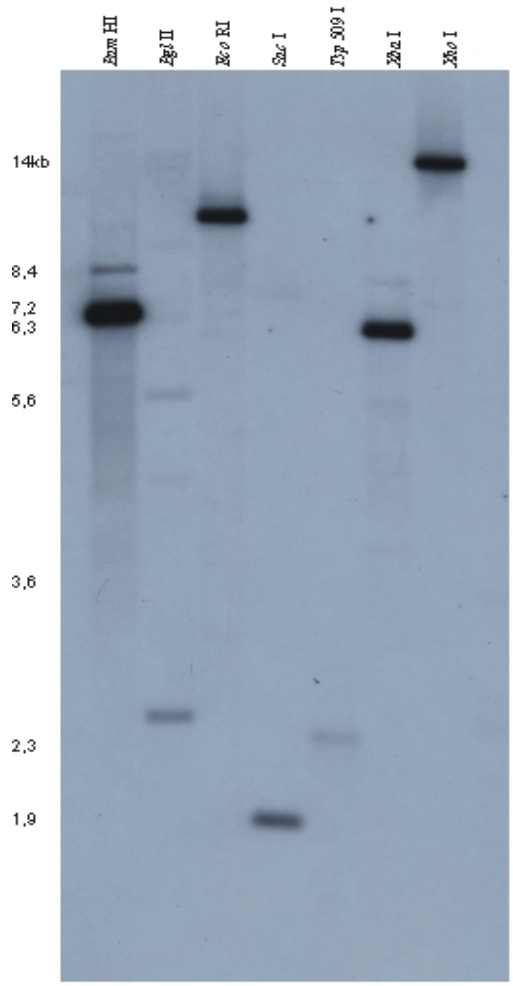


Figure 3.9 Autoradiograph of a Southern blot where different enzymes were used to digest BAC DNA. Digest fragments were probed with the *P* promoter fragment probe.

Due to its' appropriate size (2.4 kb), the *Tsp* 509I fragment was chosen as a suitable fragment for subcloning. A larger fragment, namely the *Xba* I fragment (6.1 kb), was also included in the subcloning experiment as a back-up (discussed in Methods section).

No positive results were obtained from these experiments, despite several different cloning vectors being used. No positive transformants were recovered from the initial plasmid of choice, pZERO, and so an alternate cloning vector was attempted (namely, pBluescript). No positive transformants were recovered when using this plasmid. It seemed that the fragment we were trying to clone could not be contained within a plasmid vector and produce viable plasmids. The fragment appeared deleterious to the plasmids used and therefore unclonable in these vectors. The use of another type of vector, namely lambda phage, produced similar results with no positive transformants recovered.

3.1.2.1.1 BAC clone 263 022 appears unstable

As work with the BAC progressed, it appeared that different preparations of BAC DNA were giving inconsistent results. In order to establish whether this was the case, three different DNA preparations were used in an experiment. Sample 1 was BAC DNA extracted in Rob Nicholls' lab at CWRU and brought back to South Africa by the writer. Sample 2 was BAC DNA extracted in Rob Nicholls' (RN) lab by Nancy Rebert (a staff member in RN lab) and sent to South Africa. Sample 3 was BAC DNA extracted in South Africa from bacteria containing clone 263 022 sent to South Africa in a stab from RN lab. The three different preparations were digested with various enzymes and the digest products run on a 0.8% agarose gel. The agarose gel as well as the Southern blot showed that all three DNA samples were giving different digest patterns (see Fig 3.10).

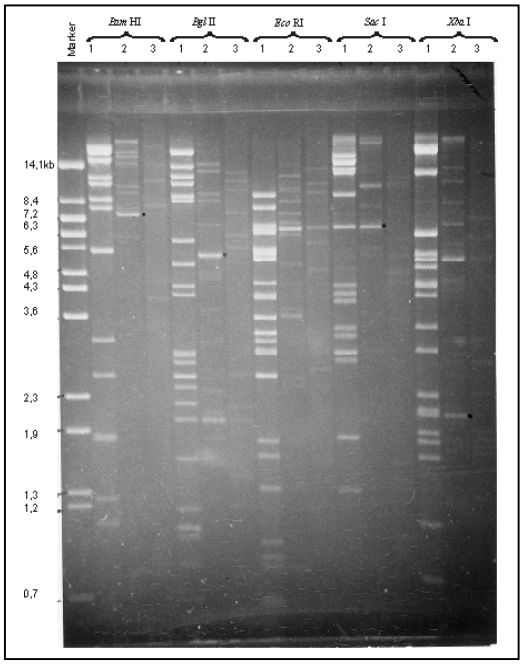


Figure (a)

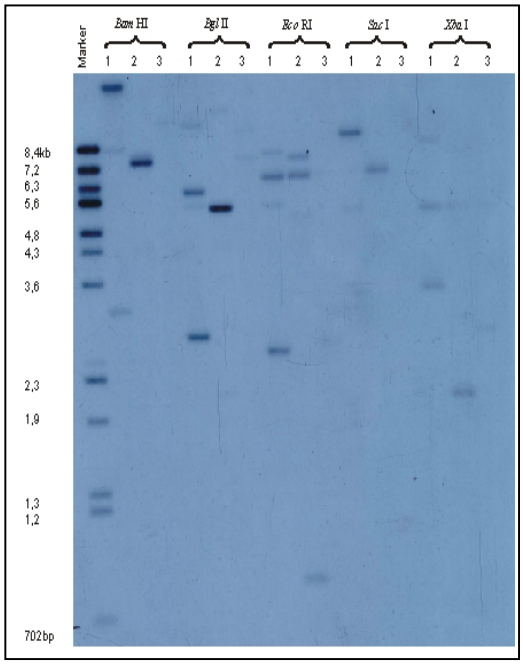


Figure (b)

Figure 3.10 The BAC 263 022 appears to be unstable. Three different BAC DNA preparations were digested with *Bam* HI, *Bgl* II, *Eco* RI, *Sac* I or *Xba* I. The agarose gel (Fig a) was blotted and probed with the human *P* gene promoter fragment. The autoradiograph is shown (Fig b). Theoretically, all three preparations should give identical results.

3.1.2.2 Inverse PCR

As it was not possible to subclone the 5' region of the *P* gene into a plasmid vector, possibly due to instability of the region, another approach to gaining sequence data for the region was undertaken. Both BAC DNA and human genomic DNA were used as template DNA in an inverse PCR experiment. See Fig 3.11.

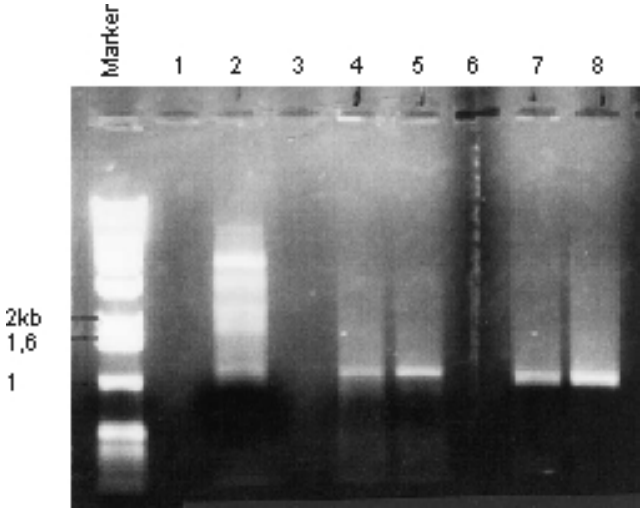


Figure 3.11 PCR fragments generated in an inverse PCR experiment. Marker lane (1 kb marker), Lane 1, 3, 6 empty lanes; Lane 2, genomic DNA used as template, Lanes 4, 5, 7 and 8, BAC DNA used as template. The 1.2 kb band in all lanes was cut out as well as the larger brightest band in the genomic DNA sample (approximately 4 kb) and the DNA purified from the gel. Purified DNA was used as template for sequencing. The sequencing primer was PE1F.

Genomic DNA (both the 1.2 kb and 4 kb templates) gave no results on sequencing, probably due to low recovery after gel extraction and subsequent weak concentration for sequencing reaction.

BAC DNA gave good sequencing results and clear sequence could be read. Analysis of this sequence data using BLAT (part of the University of California at Santa Cruz, UCSC, suite) shows the fragment generated in the inverse PCR experiment maps within intron 2 of the *P* gene.

3.2 Discussion: OCA mutation screen

3.2.1 Mutation detection in OCA patients

OCA is the most common single gene disorder in southern Africa. OCA1 is caused by mutations at the *TYR* locus and OCA2 is caused by mutations at the *P* locus. In this study we investigated the *P* gene and the *TYR* gene in an attempt to define further variation at these loci in our local population groups.

3.2.1.1 The *P* gene

3.2.1.1.1 The R305W variant

The R305W variant has been previously described and reported to occur at very high frequencies in African-Americans (0.9) (Lee *et al*. 1995). It was suggested that this variant may play a role in normal pigment variation since its frequency was dramatically lower in Caucasoids (0.17). In the present study, the frequency of the W allele in normally pigmented black Africans was found to be 0.09. The contrary results of the Lee *et al.* (1995) study might be due to sample bias. Alternatively, the allele may have reached such high frequencies by random drift or the bottle-neck phenomenon created by the bringing of the early African slaves to America. The frequency of this variant in OCA2 individuals (0.58) is significantly higher ($p<0.001$) than the frequency observed in a sample of normally pigmented black Africans (0.09). Since the R305W variant is much more common in black African OCA2 individuals compared with normally pigmented individuals from the same population, it may be postulated that the R305W variant is in linkage disequilibrium with decreased skin pigmentation. This is in keeping with the observation that the R305W variant and the 2.7 kb deletion mutation are co-segregating much of the time. The R305W mutation may be an ancient mutation and the original 2.7 kb deletion mutation may have arisen on a chromosome carrying the W variant. As the association between R305W and the 2.7 kb deletion mutation is not absolute, it may be postulated that discrepancies have arisen over evolutionary time due to mechanisms such as cross-over events or the possibility of more than one origin for the deletion mutation.

3.2.1.1.2 Detection of unknown mutations by SSCP analysis

Numerous synonymous codon mutations were detected and all scored as neutral, non-pathogenic changes (G321G, A336A, Y342Y, A355A, C517C, V629V, A686A, S736S, A776A, G780G and S788S). It is interesting to bear in mind that specific examples have been cited in which apparently 'neutral' variants have been shown to be responsible for disease phenotype (Akli *et al*. 1990, Li *et al*. 1995, Richard and Beckmann 1995, Liu *et al*. 1997). These mutations all result in either the creation or abolition of a splice site and lead to aberrant RNA processing. These types of mutations are obviously difficult to find, requiring appropriate tissue from each patient and subsequent testing for illegitimate RNA transcription products. The synonymous amino acid mutations described here do not appear to lead to the gain or loss of putative splicing consensus sites.

The difficulty of assigning non-pathogenic status to a variant just because it is polymorphic in a population is again exemplified by the findings of Fukai *et al*. (1995) who described two patients with autosomal recessive ocular albinism – each was a compound heterozygote for a different pathogenic mutation and the 'normal' tyrosinase polymorphism arg402gln. Here, ocular albinism appeared to be a clinically mild form of OCA1.

Relatively few mutations were found during this screen. The technique of SSCP analysis does not detect all small mutations, but reportedly has a sensitivity rate of 80-90% (Hayashi and Yandell 1993, Sheffield *et al*. 1993). Of the 39 OCA2 patients, 26 were heterozygous for the 2.7kb deletion and 13 had no identifiable mutation. Therefore of the 78 OCA2 alleles, 52 could not be accounted for by the deletion mutation. Mutations could be identified in four (0.077) of these. As non-2.7 kb deletion genes account for 23% of black OCA2 genes in southern Africa, these additional four mutations account for 1.7% (7.7% of 23%) of OCA2 genes in the overall population group. However, many nonpathogenic variants and polymorphisms were detected, making it unlikely that the lack of mutation detection was a technical problem.

It is proposed that some of these unidentified mutations may lie in the promoter or control regions of the gene or in intronic regions. Further, mutations which affect the regulation of the *P* gene may play an important role in the aetiology of BOCA, and the promoter region may well house mutations which down-regulate the *P* gene without rendering it non-

functional, hereby causing the intermediate phenotype seen in individuals with BOCA, and possibly other milder hypopigmentation phenotypes as exemplified by the CAR family described in this study.

It is worth bearing in mind that an individual with OCA type 1A has been reported (Oetting and King 1991) who was found to be a compound heterozygote for a substitution mutation in exon 1 of tyrosinase and a promoter mutation at –199 that altered the CCAAT box (CCAATTC→CCAATTA). This promoter mutation obviously rendered the tyrosinase allele non-functional and did not merely down-regulate transcriptional activity. Therefore, it may well be that certain elusive OCA2 mutations are to be found in the *P* promoter region, and not only possible BOCA mutations.

3.2.1.1.3 Mutation detection by Southern blotting

While SSCP methodology is useful in detecting small DNA variation of the order of a few base pairs, large scale deletions, insertions or rearrangements, of the order of several hundred or thousand base pairs, will only be detected using Southern blot analysis. When looking for a known mutation, Southern blotting is obviously an extremely useful tool, but when screening for unknown mutations the technique possesses something of a 'shot in the dark' quality. If the appropriate enzyme for genomic DNA digestion is not chosen, in conjunction with the appropriate probe, the mutation may well be missed. Also, Southern blotting will miss heterozygote deletions if there is no fragment size change. In this study, only one restriction enzyme (*Eco* RI) was used in Southern blot analysis due to a restriction on material (DNA) and economic resources. If further enzymes had been used, a mutation/s may well have been identified using this technique. Due to the fact that Southern blotting is labour intensive, slow to yield results, costly and usually involves the need to work with radioactivity, it is now seldom used as a method of mutation screening. SSCP analysis, protein truncation tests or, more recently, direct sequencing (Oetting *et al*. 2005) are the methods of choice for mutation detection.

3.2.1.2 Epistasis

While it is immediately obvious in a mutation detection experiment to look within or near

the disease locus for pathogenic variation, the possible epistatic effects of other genes should not be ignored. Epistasis is a situation where the alleles in one gene cover up or alter the expression of alleles at another gene. As discussed in this thesis, variation at the *HERC2* locus may possibly influence *P* gene expression.

3.2.2 Sequencing the *P* promoter region

3.2.2.1 Cloning

While areas such as promoters or enhancers are immediately recognisable as control elements, it is becoming increasingly more apparent that the regulatory organisation leading to correct gene expression involves these plus more complex factors. The assumption for this section of the project was that the *P* gene promoter lies in the immediate 5' region of the first exon of *P*. As will be discussed, this is not always the case, but where no other information is available, this is the only premise with which to start. In this study, traditional cloning techniques failed to produce results, and a viable clone containing the area 5' of *P* could not be generated, despite several different approaches. It seems likely that this region contains a DNA element (or elements) that, when cloned into a vector, interferes with the vectors' functioning to such an extent so as to render the vector non-viable and the region of DNA to be cloned essentially 'unclonable'.

This phenomenon, whereby short stretches of eukaryotic DNA show instability when propagated in prokaryotic hosts, is not uncommon. There are no obvious 'rules' allowing recognition of a DNA sequence as unstable (prone to rearrangement). However, high AT-content or the presence of multiple repeats in a DNA segment may result in non-clonability of a sequence (Razin *et al*. 2001). Both of these phenomena are exemplified in the recent (January 2007) cloning of a portion of chromosome 22q11 – a region located within one of the remaining unclonable gaps from the Human Genome Project (Kurahashi *et al*. 2007). This region is notoriously rearrangement-prone with breakpoints of several constitutional translocations clustering here. Kurahashi *et al*. (2007) identified a 595-bp palindromic AT-rich repeat (PATRR) within the region and propose these repeat units lead to unusual DNA secondary structure with subsequent double-strand-breakage within the palindrome followed by nonhomologous end-joining rather than homologous recombination.

A second problem in trying to clone the promoter region of a gene is being able to identify and locate the actual promoter. The simplistic assumption that each gene has its own promoter/enhancer and that these lie immediately 5' of the gene does not always hold. Cis-regulatory regions can be much larger than expected, extending over hundreds of kb of DNA. Certain genes, for example the chicken lysozyme locus, have more than one enhancer region (namely, three) and all three need to be in their correct spatial orientation for appropriate gene expression (Bonifer *et al*. 1994; Jägle *et al.* 1997). Correct gene expression implies appropriate tissue specification, appropriate developmental regulation and appropriate expression levels. Other genes, for example genes in the β-globin gene cluster, may share a control region – each gene in the locus has its own promoter, but all genes share a major regulatory element, the LCR (Milot *et al.* 1996, Bungert *et al*. 1995). Another interesting phenomenon illustrated by the β-globin cluster - and also seen in the *Hox* gene cluster (Kmita *et al.* 2000) – is that spatial orientation and gene order of genes is important in determining their regulation. Here the most 5' gene in the cluster is expressed first during development. Further, when looking at DNA sequence it is important to remember that DNA *in vivo* is wrapped around nucleosomes, which brings transcription factor binding sites together, that on 2-D sequence information will seem further apart. Besides chromatin architecture, chromosomal environment will also affect gene expression – genes may be localised in heterochromatic or euchromatic regions and epigenetic phenomena such as methylation, phosphorylation or acetylation will also affect their regulatory status.

All these phenomena contributing to gene expression make characterisation of the *P* gene regulatory machinery extremely difficult. Some of the issues discussed above may influence expression at the *P* locus, but this can only be tested by performing transfection experiments, generating transgenic mice where the p gene plus differing amounts of flanking sequence are subject to gene transfer experiments in mice. The influence of different cis-DNA regions co-transfected with *p* can then be assessed with respect to their influence on *p* expression.

3.2.2.2 Inverse PCR

The inverse PCR experiment, using BAC DNA as template, generated a clear band of 1.2 kb. Sequencing this 1.2 kb fragment showed it mapped to intron 2 of the *P* gene. The sequencing primer used was the *P* gene exon 1 forward primer. Thus, even though clean PCR product was generated, this product was produced by illegitimate primer binding – the PCR primers bound to a fragment from intron 2 and not to the ligated fragment containing *P* exon 1. Although initially exciting, it is however not surprising that the PCR product maps to the 5' region of the *P* gene, given that the template DNA was not genomic, but a BAC containing the DNA region stretching from the 3'end of HERC2 to the 5'region of the *P* gene.

3.2.2.3 Sequence analysis of the 5' region of the *P* gene

One of the main aims of this project was to identify pathogenic mutations in a cohort of subjects with OCA2. As relatively few mutations were detected (screening all 25 exons, plus exon/intron boundaries), it was postulated that some of the unidentified mutations might lie in the *P* gene promoter region. However, little upstream sequence of *P* exon 1 was available at that time (1996) - only approximately 200 bp - and consequently another aim of this project became the cloning and sequencing of the region immediately 5'of *P*. Using various methodologies, cloning of this region was unsuccessful. Now (2007), the Human Genome Project being essentially complete, DNA sequence of the 5'region of *P* is available. Bioinformatic analysis was undertaken to investigate this region and to attempt to identify the important regulatory elements - which would then be a target region in future mutation detection studies, and/or any motifs that might have rendered this region 'unclonable' by the methods used in this study.

The sequence 5'of *P* was identified and analysed using Ensembl, looking for CpG islands, repeats and regulatory features. Analysis using the Mammalian Promoter Database site, looked for *Hpa* II restriction sites which predict methylation status, CpG islands (see below) and predicted transcription factor binding sites. Dot plots were generated using the dottup programme, part of the EMBOSS suite. Note: The *P* gene is transcribed on the

reverse strand (i.e. in the direction chr 15q ter → 15q cen).

3.2.2.3.1 Promoter motifs

A promoter is the region of an operon or gene that acts as the initial binding site for RNA polymerase. Proteins known as transcription factors (TFs) bind to the promoter DNA sequences and facilitate the recruitment of RNA polymerase, the enzyme that synthesises RNA from the coding region of the gene. Promoter motifs like a CCAAT-box, an E-box and a TATA-box present near the transcription start site and are necessary for full promoter activity and are highly conserved. Promoters work together with enhancers, silencers and other regulatory elements to control the level of transcription of a gene. While prokaryotic promoters are easily identifiable, eukaryotic promoters are more complex, more diverse and often difficult to characterise.

CpG islands

CpG islands are regions of the genome where a large number of cytosine and guanine residues are found. In general, the CpG dinucleotide is generally under-represented in the vertebrate genome. DNA methylation occurs only at CpG sites. However, methyl cytosines are not stable and usually mutate to thymine over evolutionary time (CpG→ TpG) leading to the situation where CpGs are approximately five times less frequent than expected (Jones *et al.* 1992).

The 5' regions of active genes – house-keeping genes and regulated genes which are being expressed – are not methylated and the absence of methylation slows CpG decay. DNA methylation represses transcription by physically inhibiting the binding on transcription factors and also by recruiting methyl-CpG-binding proteins which alter chromatin structure (Jirtle and Skinner 2007). Approximately 56% of human genes are associated with CpG islands (Antequera and Bird 1993) and the CpG island often includes the promoter and extends approximately 1 kb downstream into the transcription unit. Therefore, identification of a CpG island helps to define the 5' limit of a gene and would help define the region in which the promoter is likely to be found (and therefore the target region for future mutation detection experiments).

CCAAT-box, TATA-box and E-box

A CCAAT-box is a eukaryotic promoter element, located approximately 75 – 80 bp upstream of the start site of transcription. CCAAT-boxes are generally associated with non-house-keeping genes, i.e. with genes under regulatory control and this element seems to be necessary for the transcription of sufficient protein product. Deletions or mutations of the CCAAT-box lead to a significant decrease in gene expression – the factor that binds to the CCAAT-box confers substantial transcriptional activity to the promoter. Also, any alteration in methylation status of the CCAAT-box abolishes factor binding.

The consensus recognition sequence of the TATA-box, TATAAA, binds a TATA binding protein, which assists in the formation of the RNA polymerase transcription complex. The TATA-box typically lies very close to the transcription start site (often within 50 bases).

A third motif is the so-called E-box, having the consensus sequence CANNTG, where the 'N' bases are specified according to the transcription factor. An E-box binds transcription factors which contain a basic-helix-loop-helix motif (bHLH).

3.2.2.3.2 Ensembl output (www.ensembl.org)

Input sequence was 20 kb (*P* gene exon 1 plus 10 kb upstream and 10 kb downstream). Sequence release: 44, April 2007.

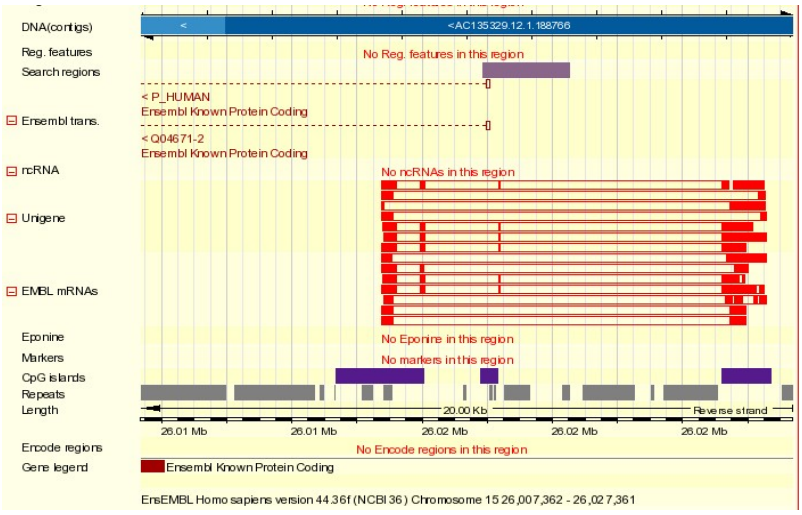


Figure 3.12 Ensembl annotation of a 20 kb region on chromosome 15q which includes the first exon of the *P* gene (nucleotide coordinates 26 007 362 – 26 027 361).

Ensembl annotation of this region (Fig 3.12) shows: 3 CpG islands in close proximity to exon 1; a large number of repeats in the exon 1 region; and no prediction of any obvious regulatory features.

3.2.2.3.3 MPromDb output (www.bioinformatics.med.ohio-state/MPromDb)

When using an annotation programme which looks more specifically at promoter motifs (MPromDb), the following output was generated (Fig 3.13):

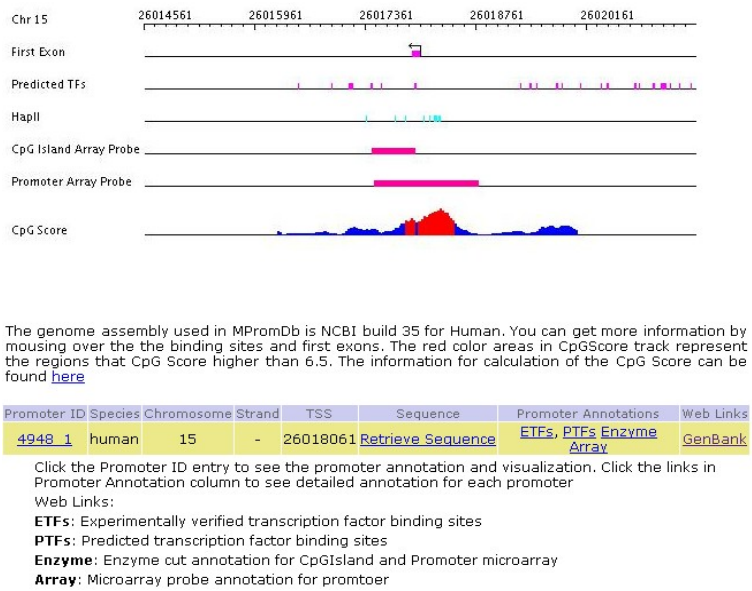


Figure 3.13 MPromDb annotation of a 20 kb region which includes the first exon of the *P* gene (the same sequence data input as for the Ensembl analysis).

The results from the mammalian promoter database are in agreement and validate the Ensembl results - the first exon of the *P* gene is predicted to be CpG related. Additional information gained indicates the presence of multiple transcription factor binding sites and several *Hpa* II restriction enzyme sites, used to determine methylation status of the region.

In order to look in more detail at the repeat areas detected by the Ensembl annotation, a dot plot of this 20 kb region was generated.

3.2.2.3.4 Dottup output (www.ebi.ac.uk/emboss)

Dot plots are a means of nucleic acid sequence analysis, which search out regions of similarity between two sequences. These two sequences may be generated in one of two ways –

- Two different sequences may be compared to each other in order to identify regions of potential homology, or
- A sequence may be compared to itself in order to find regions of similarity within the molecule, i.e. repeats.

A dot plot comparing two sequences (or one with itself) is generated by sliding a window of a size (defined by the user) along both sequences. If the sequences within that window match with sufficient precision (again, set by the user), a dot is placed in the middle of the window signifying a match. If a window size of 10 is defined, then the programme will read the sequence in blocks of 10 nucleotides; if the mismatch limit is set at 2, then up to 2 out of the 10 bases may mismatch and the window will still be classified as a match. Therefore, the larger the window of comparison, and the lower the mismatch limit, the more stringent the comparison.

The 20 kb region around *P* gene exon 1 was analysed using dottup (part of the EMBOSS suite). The window size was set at 15 and mismatch limit was set at 0. See Fig 3.14a and b.

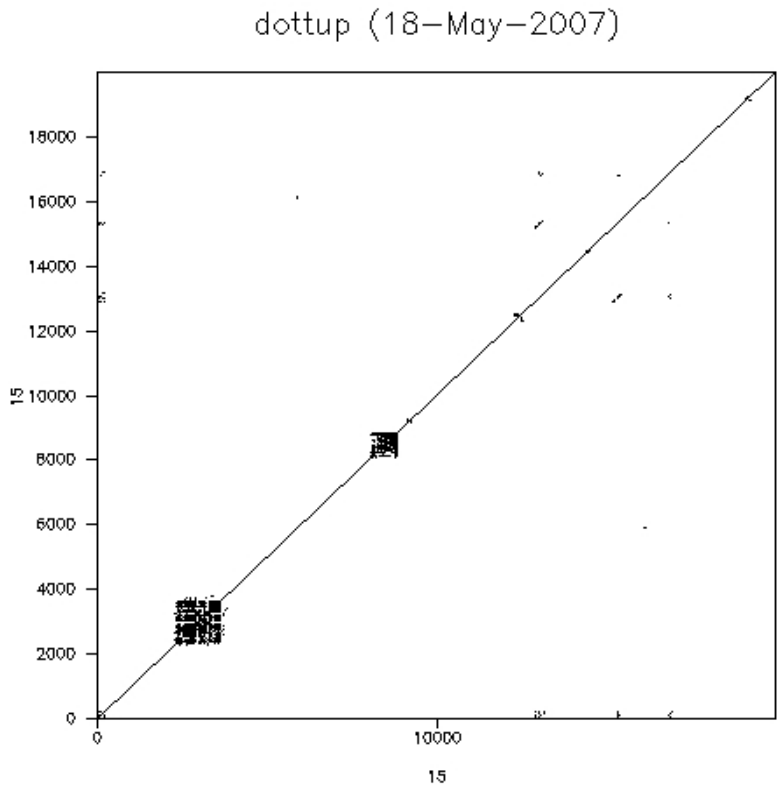


Figure 3.14a Dot plot analysing a 20 kb region around *P* gene exon 1, comparing this sequence to itself in order to identify areas of repetitive DNA.

This dot plot output shows two regions containing repetitive DNA, the first just upstream of exon 1 and the second approximately 6 kb upstream of exon 1. (Exon 1 is located approximately mid-way through this sequence input, i.e. at about position 10 000 bp.) Zooming in on the region immediately around exon 1 (input 2 kb only), clearly shows a highly repetitive region of about 800 bp in length, approximately 1 kb upstream of exon 1 (see Fig 3.14b).

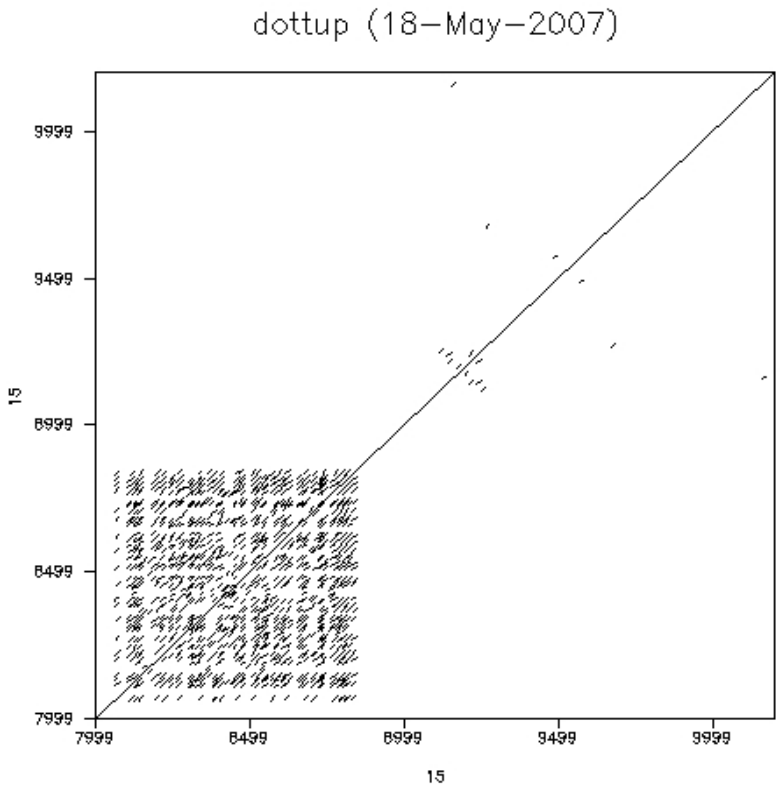


Figure 3.14b Dot plot analysing the 2 kb region upstream of *P* gene exon 1.

Actual sequence data for this region was downloaded from the University of California at Santa Cruz website (http://genome.ucsc.edu):

```
>hg17_knownGene_NM_000275 range=chr15:25673628-26020061 5'pad=0 3'pad=0
revComp=TRUE strand=- repeatMasking=none
ggacacatagatttggctcacccgagccagtccctggcatggaagttggg
tttcctccattactgtggccactcagggtctatccctggagctggggtag
ggttgctccttccccatacattcagcagttacacaaaaacaactgtgacc
tgcaaggcacagagagccacacgcttggccgcagctgtcctgctggcgtt
tcaaagcccgcatttcgacagatctgcgtgcacaccaagagccttctgcc
cactcagggagcctccccgggctctgtgccccagccccttcccgccctg
taacagtgtccctgaggcactgggaggggagcaggtgcagtgatgagcga
ggaggagacagtatttctgggacgggaactggtgcatcggtggtcacagg
ctccctggagccgccatcccacaaggatgggggtcaccatgcaggctgct
tgccagccatatccacacctaacgcacacggcggtctgaactgaggggtc
tgcacatgcgacagtgaatgtggtctcagagatctcgggctctcctgctg      - 1500 bp
tcagggccctggctgggaacttcagtgtcctggagcctgccatgccgagg
acatcctcggtccctgctaccagggcaacccctggcccaccctgtggctg
tgcaatgtgctctgttataaggcatgagatctctttcagctctgctgtta
cactgtggatggagcctcttacttcctgccaattcctgtgcagggcctgg
ccactgcagctgtgtgtgtgtctgtagctgtgttgtctccatgtgtttgt
gcgtgtatctgtgtgtgtgtctgtatctgtgtgtgtatctgtgtgtgtct
gtgcgtatatctgtgtttgtatatctgtatatgtgtatctgcgtgtgtgt
ctgtgcatgtatctgtgtgtgtgtctgtatgtgtctatgtatctgtgtat
ctctgtgtctgcgtatctgtgtgtctgtgtatctctgtgtgtagatctct
gtatctgtgtgtatctgtgcatgtgtgtatctgtgtgtctgtatctgtgt      - 1000 bp
gcctgtgtatatctgtgtctgtgtatctctgtgtgtgtccgtgtgtgtat
ctgtgtgtgtctctgtatctctgtatctgtatgtatctctgtatctgtat
ctgtgtgtctgtacctgtgtctgtgtatctgtgtctgagtatctgtatat       Highly
atttgtgtgtgtctgtgtatctgtgtctgtgtgtatatctgtgtatctgt       Repetitive
gtgtatctctgtgcctgtgtgtgtgtctctgtgtatctgtgcatgtgtct       Region
gtgtgtatctgtgtatgtgtgtctgtgtatctgcatgtgtccttgtgtgt
catctgtgtatctgtgtcctgtatctgtgtgtctgtgtattgctgtgtct
ctgtgtgtgtctctgtgtatgtgtatctgtgtatctgtgtctgtgtatat
cgtgtctatctctgtatctgtgtgtgtgtctgtgtgtatctgtgtgtgtc
tgtgtatctgtgtgtgtctgtgtatctcgtgtgtgtgtctgtgctgcagg      - 500 bp
aggggctgtcggcaagtgtaaccttggcgtcctagctggagggcctctca
gtgcccgctgggagagctctgaagggaacagaatgttcactgagcatttc
cccgttcttgttcatagggaaaagcctgtccgcagtcctgggccgcctct
ttagaggtaaagctggaaactatccccgcggagagcgctgtcggggacgc
cggggcggcgtggccttagacggagctgctctcagcccgcgccgccgggg
gctggggagctgaggggccgggggtgcggagcgggcgagctgtgggcggg
gccatgtgggcagctttgtgggcgtggccgggaggggcgtgctacgggcg
tggccgcactgtgggcggggccgagtggggagtgctgtgggcggggccgg
gccgggggcgtgcctgggtgggggacccgctgcgggagcgccgagctccc
TCTGAGTTCTTACTTCGAAGGCTGTGCTCCGCTCACCATCCAGAGCGGAG      Exon 1
GTGCGGACCTTAAACTCACTCCTGGAGAAAGATCTGCAAGTGCGCAGgta
aagtgcacgtgctccgcggtcgggaggaaggaggcgaggagccagactag
```

Figure 3.15 Nucleotide sequence data is now available for the 5'region of the *P* gene. (Sequence downloaded from UCSC site.) An area of repetitive G/C-rich DNA is found upstream from exon 1 (approximately -500 bp to -1 250 bp upstream). Putative promoter elements, CCAAT-box and E-box, are highlighted in green.

The block of repetitive DNA in close proximity to exon 1 may certainly have been a confounding factor in attempts to clone this DNA segment into plasmid vectors. A CCAAT-box sequence and an E-box sequence (CACGTG) were identified in the exon 1

region, at -1272 bp and +9 bp, respectively. These motifs are not in recognised positions, but are still worth taking into account in future mutation detection exercises. They essentially define a region of approximately 1.3 kb around exon 1 which may putatively be described as the *P* gene promoter region.

3.2.3 Possible involvement of *HERC2* in pigmentation

In mice it has been found that mutations in the gene which lies immediately adjacent to the *p* gene, *Herc2*, can produce a partial pigment phenotype (Walkowicz *et al*., 1999; Lehman *et al*., 1998). Three hypotheses that may explain the hypopigmentation phenotype seen in these mice are: hypopigmentation could result from loss of a DNA element controlling *p* gene expression; *Herc2* itself could play a role in pigmentation; or a second intronic gene which affects pigmentation may be present within the *Herc2* locus. If conserved in humans, this region could also play a role in unknown OCA2 or BOCA phenotypes. Mutations that affect the regulation of the *P* gene may play an important role in the aetiology of BOCA, and the promoter or other regions may well house mutations that down-regulate *P* gene expression without rendering it non-functional. This could result in the intermediate phenotype seen in individuals with BOCA. In support of this theory is identification of DNA variation in this study at the *HERC2* locus, as detected by Southern blot analysis, in a BOCA individual. Down-regulation of normal expression may possibly lead to other milder hypopigmentation phenotypes, as exemplified by the CAR family described in this study.

HERC2 mutation screening in this study was very limited, comprising only Southern blotting with one enzyme and one probe. The investigation as to the possible involvement of *HERC2* in pigmention is an important and interesting issue, and could certainly be expanded in future studies.

3.2.4 Some thoughts on *tyrosinase*

3.2.4.1 A confounding factor in *TYR* mutation detection – *TYRL*

At the time of cloning the human tyrosinase gene (chromosome 11q14-21), a tyrosinase-related 'gene' containing only the last two exons, exons 4 and 5, of the primary locus was also identified in the region 11p11.2-cen (Barton *et al*. 1998, Giebel *et al*. 1991). Giebel *et al.* (1991) suggest the tyrosinase-like segment was duplicated from the original tyrosinase gene approximately 24 million years ago. When undertaking mutation detection for the *TYR* gene it is obviously important to consider the presence of this pseudogene. As it shares 98% identity with the 3' region of *TYR*, variants identified in this region could actually be in the *TYR*-like gene, *TYRL* (Ray *et al*. 2007). Also, if a patient had a deletion in this region of *TYR*, this could be masked by presence of normal sequence at the *TYR*-like locus.

3.2.4.2 *TYR* mutations are rare in black individuals

In the literature, only four cases of black individuals with classic OCA1 have been described (Spritz *et al*. 1991; Oetting *et al.* 1993; King *et al.* 2003; Badens *et al*. 2006). In the experience of our laboratory, no South African black person with OCA1 has ever been seen, either in the field or in the Genetic Counselling Clinic. The individuals presented by Spritz *et al*., Oetting *et al.*, and King *et al.,* were all African-American. African-Americans carry a large percentage of white admixture and it may be hypothesized that *TYR* mutations were introduced into this population group via this route. The individual described by Badens *et al*. (2006) is the first report of a black African with classic OCA type 1A. The proband, a 34 year old Cameroonian black male of the Bamileke ethnic group, presented with a very severe phenotype, including snow white hair, milky white skin and grey-blue irides. Interestingly, the subject also presented with clearly demarcated darkly pigmented patches (ephelides) on his forehead, ear and cheek. To date, ephelides have only ever been reported in association with OCA type 2 (*P* gene mutations). Mutation analysis showed the patient was a compound heterozygote for two exon 1 mutations: a 4 bp deletion (del AACT) at nucleotide 781, and a transition (T→C) resulting in the substitution mutation, C89R.

3.2.5 Future approaches to mutation detection

Technical difficulties and limitations must be seen as the major limitation of this study. Methodologies used for mutation screening (Southern blotting and SSCP analysis) are now out-dated in most circumstances and newer techniques are available for the detection of larger deletions, chromosomal rearrangement events or alterations in expression patterns. These techniques include DNA microarray technology and MLPA (Multiplex Ligation-dependant Probe Amplification).

A DNA microarray comprises a solid support, usually a glass slide, onto which hundreds or thousands of known DNA gene sequences have been dotted by robotic machinery. mRNA is extracted from tissue to be tested, subject to rtPCR (using one fluorescently labelled primer) in order to produce cDNA. The cDNA is then hybridised to the microarray and will bind to complementary DNA sequences on the slide. Computer analysis detects specific spots to which test cDNA has bound. Microarrays can be used to determine which genes are being expressed in a single test sample or can be used to compare gene expression in two different cell types or tissue samples. Fluorescence intensity can also be measured in order to determine gene copy number. Direct mutation detection of known mutations or SNPs can also be carried out – here the microarray will be specifically designed to carry target DNA from only the gene of interest, each spot containing a different SNP or known mutation.

MLPA is a variation of the polymerase chain reaction. The two probes (equivalent to primers) are designed such that each contains a portion complementary and specific to the target sequence, called the hybridisation sequence. The two probes are designed such that they will bind to target DNA immediately next to each other. Each probe consists of a second sequence, the primer sequence. In addition, one probe will contain a stuffer sequence – an arbitary piece of DNA but of specific length. The stuffer sequence and the primer sequences do not bind to the target. When the probes correctly hybridise to the target, they are ligated by a thermo stable enzyme. The PCR primers then exponentially amplify the ligated probes. Separation of the amplification products is achieved by running on a sequence type gel. MPLA involves designing different length stuffer sequences for each different hybridisation sequence. In this way the technique could be used for fast,

cheap and easy screening of known mutations – up to 40 in a single reaction.

CHAPTER 4

Results and Discussion: TB and host genetic susceptibility

4.1 Results: TB and host genetic susceptibility

4.1.1 The 2.7 kb deletion mutation

The genotype frequencies and the allele frequencies of the 2.7 kb deletion mutation in the black TB patient group and the coloured TB patient group together with their controls are given in Table 4.1 and Table 4.2 respectively.

Table 4.1 Genotype frequencies of the 2.7 kb *P* gene deletion mutation.

SUBJECT GROUP	N (ind)	N/N	N/D	D/D
Black TB	70	0.943 (66)	0.057 (4)	0
Black Controls	99	0.960 (95)	0.040 (4)	0
Coloured TB	92	0.989 (91)	0.011 (1)	0
Coloured Controls	87	0.977 (85)	0.023 (2)	0

N (ind) = number of individuals

N/N = number of individuals with normal/normal genotype w.r.t. 2.7 kb deletion

N/D= number of individuals heterozygous for 2.7 kb deletion

D/D = number of individuals homozygous for 2.7 kb deletion

Table 4.2 Allele frequencies of the 2.7 kb *P* gene deletion mutation

SUBJECT GROUP	n (chr)	Del	Non-del
Black TB	140	0.0285 (4)	0.9710 (136)
Black Controls	198	0.0200 (4)	0.9800 (194)
Coloured TB	184	0.0050 (1)	0.9950 (183)
Coloured Controls	174	0.0115 (2)	0.9885 (172)

n (chr) = number of chromosomes

Del = proportion carrying the 2.7 kb deletion (number of chromosomes)

Non-del = proportion of non-deletion subjects (number of chromosomes)

Using the χ^2 test, no significant differences were seen between controls and patients for either population group. Presence or absence of the 2.7 kb deletion mutation does not

appear to influence susceptibility to TB in either the black or Coloured population samples studied here.

4.1.2 The R305W polymorphism

Genotype frequencies and allele frequencies of the R305W variant in the black and coloured TB patient group together with their control groups are given in Table 4.3 and Table 4.4.

Throughout this section, significant *p* values (for $p \leq 0.05$) are highlighted in red.

Table 4.3 Genotype frequencies for the R305W variant.

SUBJECT GROUP	N (ind)	C/C	C/T	T/T	Hardy-Weinberg (p values)
Black TB	68	0.676 (46)	0.221 (15)	0.103 (7)	0.0099
Black Controls	97	0.866 (84)	0.124 (12)	0.010 (1)	0.4496
Coloured TB	78	0.847 (66)	0.141 (11)	0.013 (1)	0.4136
Coloured Controls	75	0.854 (64)	0.133 (10)	0.013 (1)	0.3716

N = number of individuals

C/C = homozygous for cytosine allele at codon 305 (i.e. R allele)

C/T =heterozygous for cytosine and thymidine at codon 305 (i.e. R and W alleles)

T/T = homozygous for thymidine allele at codon 305 (i.e. W allele)

Table 4.4 Allele frequencies for the R305W variant.

SUBJECT GROUP	n (chr)	C	T
Black TB	136	0.816 (111)	0.184 (25)
Black Controls	194	0.919 (178)	0.081 (16)
Coloured TB	156	0.918 (143)	0.082 (13)
Coloured Controls	150	0.920 (138)	0.080 (12)

n = number of chromosomes

C = cytosine (i.e. proportion carrying the R (argenine) allele)

T = thymidine (i.e. proportion carrying the W (tryptophan) allele).

The genotype distribution in the black control, coloured control and coloured TB samples was found to be in accordance with Hardy-Weinberg equilibrium. The genotype distribution in the first subject group (black TB patients) shows that this group is not in Hardy-Weinberg equilibrium (p=0.0099), suggesting that selection for a certain allele has taken place.

Genotypes were further analysed by comparing individuals with C in their genotype (C/C plus C/T individuals) and without C in their genotype (T/T individuals) and then individuals with T in their genotype (C/T plus T/T individuals) and without T in their genotype (Table 4.5).

Table 4.5 Genotype clusterings for the R305W variant.

SUBJECT GROUP	N (ind)	+ C	- C		+ T	- T
		C/C + C/T	T/T		C/T + T/T	C/C
Black TB	68	0.897 (51)	0.103 (7)		0.324 (22)	0.676 (46)
Black Controls	97	0.990 (96)	0.010 (1)		0.134 (13)	0.866 (84)
Coloured TB	78	0.988 (77)	0.013 (1)		0.154 (12)	0.846 (66)
Coloured Controls	75	0.987 (74)	0.013 (1)		0.146 (11)	0.854 (64)

N= number of individuals

In order to test whether there was a significant difference in the genotype clusterings between patients and controls, the exact test was performed on the above data. (Table 4.6)

Table 4.6 p values generated using Fisher's exact test on genotype clustering data.

	BLACK TB vs BLACK CONTROLS	COLOURED TB vs C CONTROLS
+ C vs - C	0.009	0.742
+ T vs - T	0.003	0.541

The 'T' allele is significantly more common in black TB patients when compared to the

black control group. There is no significant difference between the Coloured TB group and their control group.

4.1.3 Analysis of 3 *P* gene microsatellite markers

4.1.3.1 D15S1533 (IVS17)

The frequencies of the different alleles at the D15S1533 locus are given in Table 4.7 and graphically depicted in Figure 4.1.

Table 4.7. Allele frequencies D15S1533 (IVS17)

	B TB PATIENTS	B CONTROLS		C TB PATIENTS	C CONTROLS
Allele	(n=72)	(n=99)		(n=92)	(n=87)
16		0.005			
17					
18	0.097	0.091		0.321	0.333
19		0.015 (3)		0.043 (8)	0.017 (3)
20		0.005 (1)		0.071 (13)	0.017 (3)
21	0.104 (15)	0.061 (12)		0.038 (7)	0.052 (9)
22	0.111	0.111		0.054	0.063
23	0.028 (4)	0.045 (9)		0.071 (13)	0.115 (20)
24	0.257	0.278		0.185	0.201
25	0.097	0.101		0.098	0.080
26	0.139	0.126		0.071	0.075
27	0.104 (15)	0.091 (18)		0.078 (14)	0.029 (5)
28	0.028	0.035		0.016	
29	0.014	0.010			0.006
30					0.006
31	0.014	0.020		0.005	0.006
32	0.007	0.005			

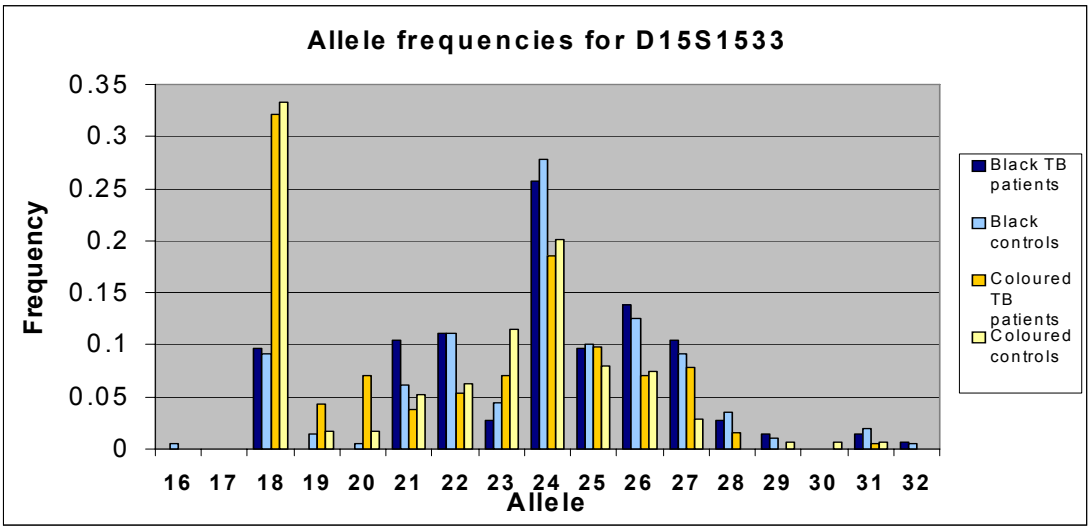


Figure 4.1 Microsatellite allele frequencies at the D15S1533 (IVS17) locus.

Tests of significance (Fisher's exact test) were performed for each allele size where there appeared to be an interesting difference in allele frequencies between the control and the patient group for each population.

For the D15S1533 marker, alleles 19, 20, 21, 23 and 27 were investigated. *p* values were generated using Fisher's exact test (see Table 4.8):

Table 4.8. *p* values for allele frequencies at D15S1533.

MARKER	ALLELE	BLACK	COLOURED
		p values (B controls vs B TB patients)	*p* values (C control vs C TB patients)
D15S1533	19	0.192	0.125
	20	0.579	0.011
	21	0.092	0.352
	23	0.289	0.091
	27	0.404	0.033

Alleles 20 and 27 show significantly different allele frequencies between patients and

controls in the Coloured sample. (Differences between allele frequencies should also be noted for allele 21 in the black population and allele 23 in the Coloured population – while these differences are not significant at the $p \leq 0.05$ level, they are however significant at $p \leq 0.10$.)

Also note – for certain alleles, clear population differences for allele frequencies can be seen. For example, allele 18 is more common in the Coloured population than it is in the black population (0.333 vs 0.091) and allele 24 is more common in the black population than it is in the Coloured population (0.278 vs 0.201).

4.1.3.2 D15S1536 (IVS20)

The frequencies of the different alleles at the D15S1536 locus are given in Table 4.9 and graphically depicted in Figure 4.2.

Table 4.9. Allele frequencies D15S1536 (IVS20)

	B TB PATIENTS	B CONTROLS		C TB PATIENTS	C CONTROLS
Allele	(n=72)	(n=99)		(n=92)	(n=87)
11				0.005	0.006
12	0.236	0.212 (42)		0.337 (62)	0.230 (40)
13	0.014	0.015		0.033	0.029
14	0.014	0.010			
15	0.132	0.152		0.082	0.092
16	0.125 (18)	0.126 (25)		0.098 (18)	0.138 (24)
17	0.104 (15)	0.086 (17)		0.071 (13)	0.115 (20)
18	0.201 (29)	0.086 (17)		0.071 (13)	0.098 (17)
19	0.014 (2)	0.030 (6)		0.016	0.023
20	0.076	0.086		0.120	0.149
21		0.050		0.033	0.057
22	0.014 (2)	0.045 (9)		0.043	0.029
23		0.015		0.005	
24	0.021	0.015		0.022	0.011
25		0.005		0.011	
26	0.021	0.035		0.033	0.011
27	0.028	0.015		0.022	0.011
28		0.010			
29					
30					
31		0.005			

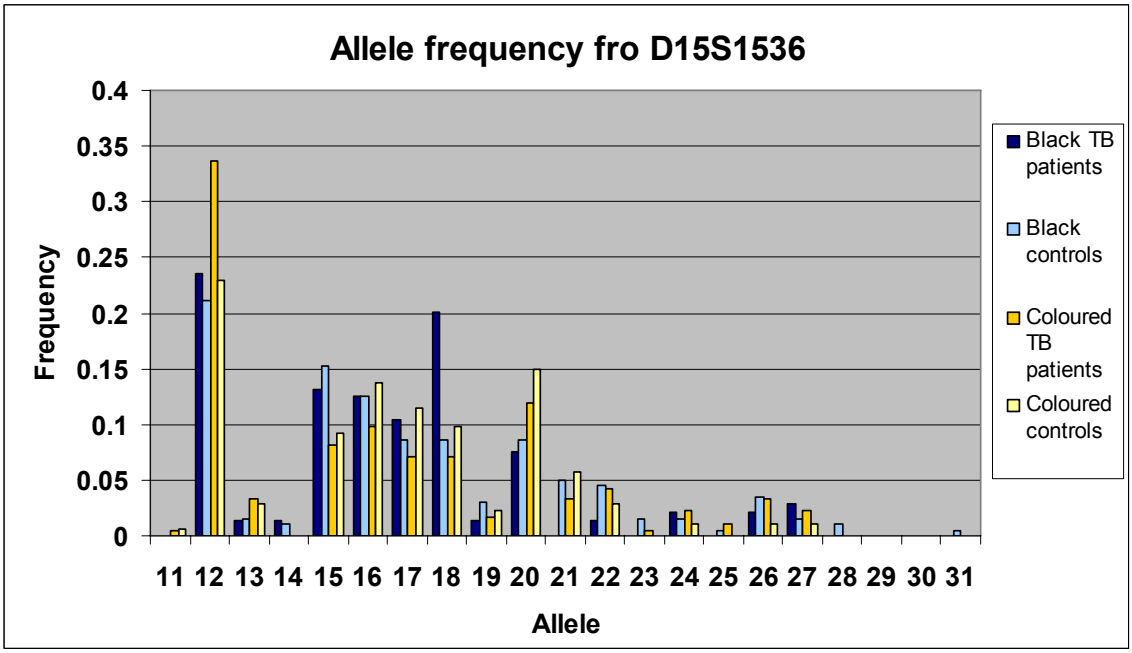


Figure 4.2. Microsatellite allele frequencies at the D15S1536 (IVS20) locus.

Tests of significance were performed for alleles 12, 16, 17 and 18 and the results are shown in Table 4.10.

Table 4.10. *p* values for allele frequencies at D15S1536.

MARKER	ALLELE	BLACK *p* VALUES (B TB PATIENTS vs B CONTROLS)	COLOURED *p* VALUES (C TB PATIENTS vs C CONTROLS)
D15S1536	12	0.320	0.003
	16	0.558	0.138
	17	0.340	0.091
	18	0.001	0.221

Allele 12 shows significantly different allele frequencies between the patient and control group in the Coloured population and allele 18 shows significantly different allele frequencies between the patient and control group in the Black population for p values ≤ 0.05. (When using the less stringent critical value for p, where p ≤ 0.10, allele 17 is then significantly more common in the Coloured control group than in the TB patient group.)

4.1.3.3 D15S1537 (IVS24)

The frequencies of the different alleles at the D15S1537 locus are given in Table 4.11 and graphically depicted in Figure 4.3.

Table 4.11. Allele frequencies D15S1537 (IVS24)

	B TB PATIENTS	B CONTROLS		C TB PATIENTS	C CONTROLS
Allele	(n=72)	(n=99)		(n=92)	(n=87)
7				0.005	
8					0.006
12	0.014 (2)	0.020 (4)		0.005 (1)	0.040 (7)
13	0.097 (14)	0.101 (20)		0.049 (9)	0.161 (28)
14	0.111 (16)	0.136 (27)		0.098 (18)	0.057 (10)
15	0.056 (8)	0.066 (13)		0.103 (19)	0.201 (35)
16	0.070 (10)	0.045 (9)		0.228 (42)	0.098 (17)
17	0.021	0.020		0.027	0.046
18	0.049	0.051		0.033	0.011
19	0.007	0.010		0.027	0.006
20	0.083 (12)	0.071 (14)		0.071 (13)	0.029 (5)
21	0.034 (5)	0.384 (76)		0.179 (33)	0.132 (23)
22	0.069 (10)	0.045 (9)		0.071	0.063
23	0.028	0.020		0.049	0.052
24	0.028 (4)	0.020 (4)		0.022(4)	0.086 (15)
25		0.010		0.016	0.006
26	0.014			0.011	0.006
27				0.005	

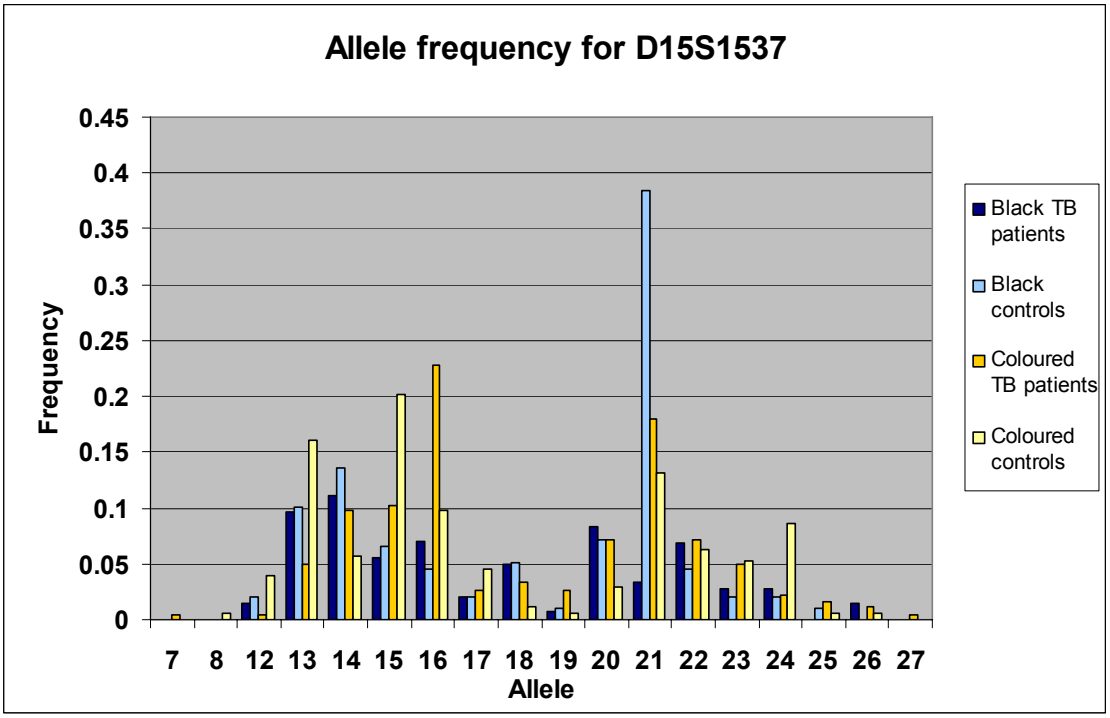


Figure 4.3. Microsatellite allele frequencies at the D15S1537 (IVS24) locus.

Tests of significance were performed for alleles 12, 13, 14, 15, 16, 20, 21 and 24 (for which frequencies appeared to differ between the subject and control groups), and the results are shown in Table 4.12.

Table 4.12 *p* values for allele frequencies at D15S1537.

MARKER	ALLELE	BLACK *p* VALUES (B CONTROLS vs B TB PATIENTS)	COLOURED *p* VALUES (C CONTROL vs C TB PATIENTS)
D15S1537	12	0.500	0.027
	13	0.531	<0.001
	14	0.285	0.100
	15	0.440	0.003
	16	0.229	<0.001
	20	0.403	0.052
	21	<0.001	0.115
	24	0.455	0.005

Alleles 12, 13, 15, 16 and 24 show significantly different allele frequencies between the patient and control group in the Coloured population sample. For allele 12, fewer patients have this allele than do controls; for allele 13, fewer patients have this allele than do

controls; for allele 15, fewer patients have this allele than do controls; for allele 16 fewer controls have this allele than do patients and for allele 24, fewer patients have this allele than do controls.

Allele 21 shows significantly different allele frequencies between the patient and control group in the black population sample. Far fewer patients carry this allele as compared to the control group.

4.1.4 Considering the 5 markers as a haplotype

Although haplotype data could not be unambiguously established, as the sample groups comprised individuals and not families, haplotypes were estimated and data analysed using the software package, Arlequin.

Table 4.13. Exact *p* values generated when performing pairwise comparisons between all 5 marker loci simultaneously (Arlequin), i.e. considering the 5 markers as a haplotype.

	C CONTROLS	C TB PATIENTS	B CONTROLS
C TB patients	0.533		
B Controls	0.077	0.118	
B TB patients	0.223	0.498	0.479

No significant differences were seen between either patient and their respective control group. However, as might be expected, the p value for comparison of black controls vs Coloured controls (0.077) is almost significant, a reflection of population differences.

4.1.5 Summary

In order to summarise the results in this chapter, probability values (Fisher's exact test) were generated for pairwise comparisons for individual loci (Table 4.14).

Table 4.14 *p* values genetated using Fisher's exact test for pairwise comparisons for individual *P* gene loci.

	2.7 kb del	R305W	D15S1533	D15S1536	D15S1537
C C vs C TB	1	1	0.239	0.419	<0.001
B C vs B TB	1	0.0019	0.962	0.036	0.952
C C vs B C	1	1	<0.001	0.360	<0.001
C TB vs B TB	1	0.0119	<0.001	0.003	<0.001

Highly significant differences are seen between the Coloured control and patient group at the D15S1537 locus.

Highly significant differences are seen between the black control and patient group at the R305W locus and at the D15S1536 locus.

When comparing the black and Coloured control groups to each other and when comparing the black and Coloured patient groups to each other, several loci show significant differences in allele frequencies (R305W, D15S1533 and D15S1537). These differences are population differences, and they are not considered relevant to the TB susceptibility study.

4.2 Discussion: TB and Host Genetic Susceptibility

While infectious disease genetics is not a new area of study – some of the largest twin studies were performed more than 50 years ago (Puffer 1946, Stocks 1928, Frost 1933) – the current attention being afforded host genetics and susceptibility to infectious disease reflects the recent increase in power of new molecular approaches being used to identify susceptibility and resistance genes. The number of candidate genes being proposed as potential susceptibility genes is rapidly increasing. Moreover, genome-wide linkage analysis is beginning to reveal regions of the human genome which harbour further susceptibility or resistance loci. In this study the *P* gene was investigated as a possible candidate falling within a region on human chromosome 15q previously identified as being associated with susceptibility to TB.

4.2.1 OCA2 heterozygotes and a selective advantage for TB

At the outset of this study it was tempting to postulate that the *P* gene, shown unexpectedly to be expressed in certain immune tissues, might have a function other than its role in melanin production. If the P protein functions as a part of the immune system then it would be feasible to postulate that this locus would find itself subject to selective pressure in the presence of infectious disease and it would therefore be expected that heterozygote *P* mutation carriers would have a selective advantage over non-mutation carriers. To test this theory, the only known common *P* gene mutation, the 2.7 kb deletion mutation, was screened for in a group of black TB patients and a group of Coloured TB patients, plus their respective ethnically and geographically matched controls.

The frequency of the 2.7 kb deletion was not significantly different between patients and controls in both the Black and Coloured groups tested. It can be concluded that this mutation is unlikely to play a major role in either resistance or susceptibility to TB disease. It should be noted that the sample sizes are small. Given the low frequency of the mutant relative to the normal allele, significant differences in allele frequencies may be detected only in a much larger sample.

4.2.2 The R305W polymorphism

This SNP is considered a non-pathogenic variant as it is polymorphic in many populations (Lee *et al*. 1995). It could, however, have some functional significance given that it leads to an amino acid change in the P protein.

Black TB patients have a significantly higher frequency of the T (W) allele when compared to the black control group. It may be suggested that the T allele is a susceptibility allele (or is in linkage disequilibrium with a susceptibility allele) in this population group and individuals with the genotypes that include a 'T' allele are more likely to develop TB disease. Grouped genotype comparisons support this, i.e. genotype frequencies with T (C/T and T/T) or without T (C/C) are significantly different between patients and controls in the black population samples tested here.

The frequency of the C and T alleles is not significantly different in the Coloured patient versus control group.

4.2.3 Analysis of 3 *P* gene microsatellite markers

For the marker, D15S1533 (IVS 17), allele 20 and allele 27 are significantly more common in Coloured TB patients than in Coloured controls. These may be possible susceptibility alleles for this population group. No significant differences in allele frequencies in the black sample are seen.
Allele 18 is more common in the Coloured group as compared to the black group of individuals, suggesting this may be a Caucasoid allele introduced through population admixture.

For the marker, D15S1536 (IVS 20), allele 12 is significantly more common in Coloured TB patients than in Coloured controls. Allele 18 is significantly more common in Black TB patients than in the Black control group. These may be susceptibility alleles in their respective population groups.

For the marker, D15S1537 (IVS 24), alleles 12, 13, 15 are significantly more common in

the Coloured control group than in the Coloured patients. These may be protective alleles in this population. Allele 16 is more common in Coloured patients than in Coloured controls, suggesting it may be a susceptibility allele.

4.2.4 The 5 markers as a 'haplotype'

Since families were not tested, it was not possible to deduce haplotypes. Some attempt was, however, made to infer haplotypes using the Arlequin package. Based on inferred haplotypes, there is no significant difference between the TB patient and control groups in either population. This would indicate that the region covered by the 'haplotype' is not associated with host resistance/susceptibility to TB. However, given that several loci within this region individually show significant association to either resistance or susceptibility to TB, we can conclude that, while the *P* gene itself may not be directly involved in TB susceptibility, genetic variants close to *P*, could still be significant in host susceptibility to this infectious disease. Further, the markers D15S1533, D15S1536 and D15S1537 used in this study are microsatellite markers and it should be borne in mind that microsatellites are less stable than SNPs (due to slippage). Consequently, given sufficient time, microsatellite haplotypes may decay, making the interpretation of haplotype experiments more complicated.

Since the original paper by Bellamy *et al*. (2000) describing linkage of a region of chromosome 15q to TB susceptibility, and subsequent fine mapping of the region (Cervino et al. 2002), no further literature investigating this genomic region has been published. The study undertaken here supports these earlier findings, with significant associations being found in both the South African black and Coloured population groups with alleles in the *P* gene. In the black population, the T allele of the R305W polymorphism and allele 18 at the D15S1533 locus are significantly associated with susceptibility to TB. In the Cape Coloured population, alleles 20 and 27 at the D15S1533 locus, allele 12 at the D15S1536 locus and allele 16 at the D15S1537 locus are associated with susceptibility to TB. Further, in the Coloured population alleles 12, 13 and 15 at the D15S1537 locus showed significant association with normal controls and may be classified as protective or resistance alleles. Unpublished data from the laboratory of Professor Rob Nicholls shows that the P protein is expressed in lymph tissue in the chicken. Expression of this protein in a tissue involved in

immunity is suggestive of another role for P, besides its classic function in the pigment pathway.

4.2.5 Reconciling apparently contradictory findings in the literature

There is now a substantial body of work looking for the genes responsible for susceptibility to mycobacterial infections. Many genes have been tested for association with susceptibility to TB, in many diverse populations and using a variety of study designs.

Besides the clear findings of studies on the genes involved in extreme susceptibility to mycobacterial disease, inherited in a Mendelian fashion, other studies have left us with results that are sometimes contradictory. Takiff (2007) argues that the accumulated literature on host genetic susceptibility to TB is "often contradictory" and can be "inconclusive". While a given polymorphism in a certain gene may show association with TB susceptibility in one population, a study undertaken in another population group may show no association. How is this possible? One explanation might be: Given that numerous genes are necessarily involved in determining susceptibility, the subset of these genes exerting their cumulative effect may be different in different populations (Hill 2006). Another explanation might be: If a major determinant of susceptibility is present in a particular ethnic group, another ethnic group may well have a different major determinant of susceptibility (for example the HLA allele DQB1*0503 is a major determinant of susceptibility in Cambodians (Goldfeld *et al*. 1998), but has not been reported in any other population). A third argument, put forward by Casanova and Abel (2007), proposes that much of TB susceptibility may be determined, not by the small effects of many genes, but by the large effects of fewer, as yet unidentified, dominant genes – a scenario exemplified by the results of the Baghdadi study of Moroccan families (Baghdadi *et al*. 2006).

It is also essential to take into account the important differences of intensity of exposure within a population, strain virulence, length of time the individual has been exposed to the pathogen and environmental factors when interpreting molecular studies. Rigorous criteria for the selection of both cases and controls helps to minimize these variables which could easily modify effects and lead to confounding results. Also, a study (such as this one) which analyses small individual allele effects in what is most likely a complex multi-locus

interaction will always be limited. When investigating the genetics of complex diseases there is a need for large multi-locus studies which are sufficiently powered.

4.2.6 Third world resources

An important issue is brought into question by Takiff (2007): "Will identifying genetic determinants of susceptibility contribute to the control of TB?" Given that TB is endemic in poorer countries with limited health care resources, would it not be more practical (even more ethical?) to concentrate rather on diagnosing, effectively treating and improving local control programs? Certainly socioeconomic conditions and TB control programs can modify, and in extreme circumstances, override, the effects of host genetics in determining susceptibility to this infectious disease.

Given the emergence of increasingly virulent, multi-drug resistant strains of *M. tuberculosis* and the severity of the problem this presents (Raviglione 2006), it may well become necessary to target chemotherapy and possibly vaccines, prophylaxis and control programs to the genetic composition of the individual, family or ethnic group. In populations where certain loci have been found to be highly associated with TB (e.g. in Cambodians [Goldfeld *et al*. 1998] and in Moroccans [Baghdadi *et al*. 2006]), a cost-benefit analysis would certainly be feasible. The identification of host susceptibility alleles responsible for cases of MSMD has, and will continue to, contribute to knowledge regarding the functional elements of the human immune system with respect to mycobacterial infection. Given that TB, together with AIDS and malaria, are the three major killers among infectious diseases, new approaches to preventing, diagnosing and curing TB are needed – which in turn depend on a better understanding of the mycobacterium as well as the host.

Chapter 5

Conclusions

Conclusions

OCA mutation detection

The mutation screen in subjects with albinism identified several mutations, but many remain elusive. Limitations of this study are due to the approach used, which could not detect rearrangements, mutations hidden in the regulatory region/s, intron mutations, or single-exon, multi-exon or whole gene deletions on one chromosome. In OCA1, partial deletions of the 3'region of the gene may have gone undetected due to the presence of the *TYR*-like pseudogene.

Further, we should bear in mind that mutations influencing the OCA2 phenotype may not necessarily be restricted to the *P* gene, and similarly, the OCA1 phenotype may be modified by loci other than *TYR*. A mutation may be present in another gene whose product participates in the melanin metabolic pathway. The interaction of pigment proteins is clearly illustrated by a family described by Manga *et al.* (1997). Two sibs from the black African population were described with an atypical form of OCA – the phenotype was essentially an intermediate form between OCA2 and ROCA. The individuals presented with hair similar to that of OCA2 individuals, but reddish and skin much lighter but still red-yellow in colour which is typical of the ROCA phenotype. Mutation analysis revealed the individuals were compound hetrozygotes for two *TYRP1* mutations (i.e. they were affected with ROCA), but in addition were heterozygous for the 2.7 kb deletion mutation at the *P* locus. This scenario of phenotypic variation of the ROCA phenotype by mutation at the *P* locus highlights the importance of considering the interaction of gene products in the pigment biosynthesis pathway. Further, (although the presence of a second *P* mutation could not be ruled out in these individuals) it seems that heterozygosity for a *P* mutation is sufficient to influence phenotype when occurring together with mutation/s at another related locus.

Epigenetic phenomena in the *P* gene region revisited

Epigenetic imprinting characterises the genes which cause Prader Willi Syndrome (PWS) and Angelman Syndrome (AS). Interestingly patients with PWS and AS often present with hypopigmentation relative to other family members. PWS and AS are developmental

disorders associated with specific, but different, imprinted regions on chromosome 15q. In both of these syndromes, dysregulation of the methylation status of the PWS or AS loci, close to *P*, alters (down-regulates) *P* gene expression. Although isolated OCA2 carrier status does not usually lead to hypopigmentation, the heterozygous loss of *P* gene expression in PWS and AS patients more often translates into a phenotype. It may be possible that methylation changes at the PWS or AS loci somehow affect the surrounding DNA, by altering DNA methylation status or by altering chromatin structure through histone acetylation or methylation, and the syntenic CpG island associated with the *P* gene may become dysregulated. Why heterozygous loss of function of *P* is sufficient to cause a hypopigmentation phenotype for this classic autosomal recessive locus in these syndromes remains an intriguing question.

The function of the P protein

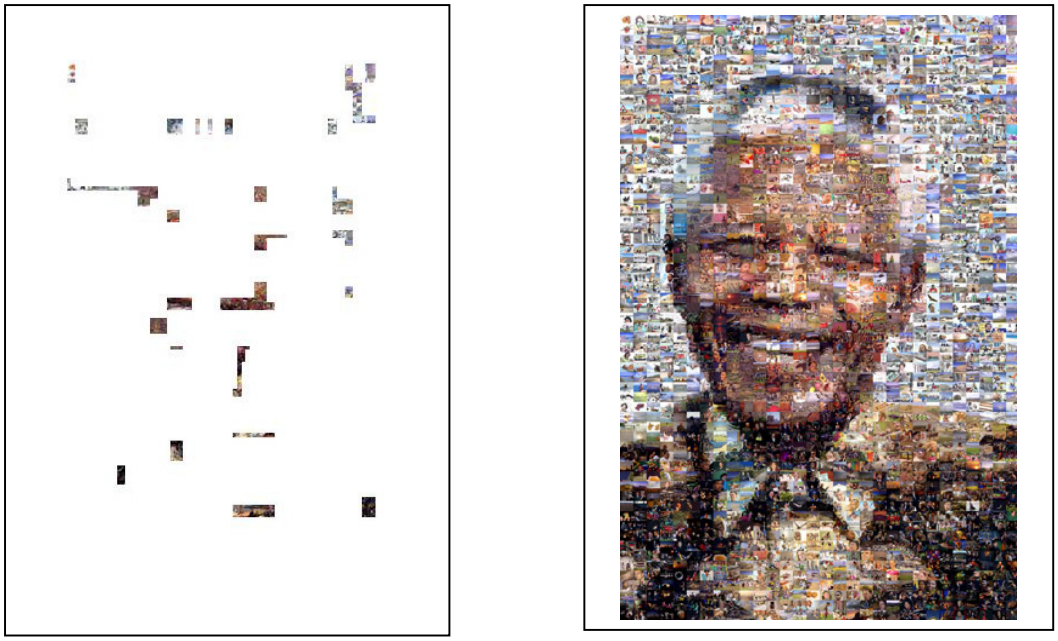
The function of the P protein remains unclear – its role in pigment is undisputed, although its precise mode of action has yet to be elucidated. Mouse studies show that in cells null for the p protein, melanosomal ultrastructure is abnormal, the localisation and trafficking of tyrosinase is altered and the transport of other melanosomal proteins into the melanosome fails (Manga *et al.* 2001a). This function of the P protein as a transport channel may be important in a context other than pigment production. As discussed in chapter 1, the genome of *M. tuberculosis* contains two *P*-like genes. In this unicellular organism which does not produce pigment, the requirement for the P protein is presumably in context of its function as a trans-membrane channel, or it may have another, as yet, unidentified function.

It is proposed that any selection taking place at the *P* locus is associated with another function of the *P* protein, perhaps related to host immunity. This “other” function of the P protein may also be related to the high carrier status of OCA2 heterozygotes in Africa.

Further, it is unlikely that the selective agent is only the TB pathogen. It could be other pathogens, since TB infection in the black African population is a relatively recent phenomenon which occurred less than 200 years ago. TB is unlikely to have been acting long enough or to have had a strong enough selective pressure to have attained the high *P* gene mutation frequency observed in sub-Saharan Africans.

TB and host susceptibility: The bigger picture will emerge

Host susceptibility is a complex genetic trait involving many host genes (some having greater and some lesser effects), many pathogen genes (conferring greater or lesser virulence) and these are all influenced by other environmental factors. The picture is a detailed puzzle of which small pieces are now becoming apparent (see Fig 5.1).



A B

Figure 5.1 Host susceptibility to infectious disease may be considered a complex genetic trait, with multiple genes involved and environmental conditions having significant influence. TB susceptibility is a sum of many different parts. In the last decade, small pieces of the puzzle have become visible, as depicted in A, but the bigger picture, B, has yet to emerge.

The tremendous public health problem presented by TB persists, despite the widespread use of the BCG vaccine and antibiotic therapy. The vaccine is variable in the protection it provides and the antibiotics in current use are losing their efficacy in the face of emerging drug resistance. To develop new prevention and treatment strategies the field of molecular genetics must address the following questions: Why is it that not all individuals exposed to *M. tuberculosis* will become infected by the pathogen?; Once they infect an individual, how do these bacteria evade the host defence system to survive, sometimes for years? Why is it that only 10% of those infected will go on to develop TB disease?; Why are some

strains of *M. tuberculosis* more virulent than others?; and How does *M. tuberculosis* react to the host at different stages of infection? Successful research in these areas will aid in the design of novel TB therapies, a global health issue whose requirement is becoming increasingly urgent.

ELECTRONIC DATABASES

Albinism Database	http://albinismdb.med.umn.edu
Dottup	http://www.ebi.ac.uk/emboss
Ensembl	http://www.ensembl.org
GeneCards	http://www.genecards.org
MPromDb	http://www.bioinformatics.med.ohio-state/MPromDb
Online Mendelian Inheritance in Man	http://www.ncbi.nlm.nih.gov/omim
University of California at Santa Cruz	http://genome.ucsc.edu

REFERENCES

Akahoshi, K., Fukai, K., Kato, A., Kimiya, S., Kubota, T., Spritz, R.A. (2001) Duplication of 15q11.2-q14, including the *P* gene, in a woman with generalized skin hyperpigmentation. *Am J Med Genet* 2001 104, 299-302.

Akahoshi, M., Nakashima, H., Miyake, K., Inoue, Y., Shimizu, S., Tanaka, Y., Okada, K., Otsuka, T. and Harada, M. (2003) Influence of interleukin-12 receptor beta1 polymorphisms on tuberculosis. *Hum Genet* 112, 237-43.

Akli, S., Chelly, J., Mezard, C., Gandy, S., Kahn, A. and Poenaru, L. (1990) A "G" to "A" mutation at position -1 of a 5' splice site in a late infantile form of Tay-Sachs disease. *J Biol Chem* 265, 7324-30.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* 215, 403-10.

Amos-Landgraf, J.M., Ji, Y., Gottlieb, W., Depinet, T., Wandstrat, A.E., Cassidy, S.B., Driscoll, D.J., Rogan, P.K., Schwartz, S. and Nicholls, R.D. (1999) Chromosome breakage in the Prader-Willi and Angelman syndromes involves recombination between large, transcribed repeats at proximal and distal breakpoints. *Am J Hum Genet* 65, 370-86.

Antequera, F. and Bird, A. (1993) Number of CpG islands and genes in human and mouse. *Proc Natl Acad Sci U S A* 90, 11995-9.

Aquaron, R. (1990) Oculocutaneous albinism in Cameroon. A 15-year follow-up study. *Ophthalmic Paediatr Genet* 11, 255-63.

Babb, C., Keet, E.H., van Helden, P.D. and Hoal, E.G. (2007a) SP110 polymorphisms are not associated with pulmonary tuberculosis in a South African population. *Hum Genet* 121, 521-2.

Babb, C., van der Merwe, L., Beyers, N., Pheiffer, C., Walzl, G., Duncan, K., van Helden, P. and Hoal, E.G. (2007b) Vitamin D receptor gene polymorphisms and sputum conversion time in pulmonary tuberculosis patients. *Tuberculosis (Edinb)* 87, 295-302.

Badens, C., Courrier, S. and Aquaron, R. (2006) A novel mutation (delAACT) in the tyrosinase gene in a Cameroonian black with type 1A oculocutaneous albinism. *J Dermatol Sci* 42, 121-4.

Baghdadi, J.E., Orlova, M., Alter, A., Ranque, B., Chentoufi, M., Lazrak, F., Archane, M.I., Casanova, J.L., Benslimane, A., Schurr, E. and Abel, L. (2006) An autosomal dominant major gene confers predisposition to pulmonary tuberculosis in adults. *J Exp Med* 203, 1679-84.

Barbosa, M.D., Nguyen, Q.A., Tchernev, V.T., Ashley, J.A., Detter, J.C., Blaydes, S.M., Brandt, S.J., Chotai, D., Hodgman, C., Solari, R.C., Lovett, M. and Kingsmore, S.F. (1996) Identification of the homologous beige and Chediak-Higashi syndrome genes. *Nature* 382, 262-5.

Bard, L.A. (1978) Heterogeneity in Waardenburg's syndrome. Report of a family with ocular albinism. *Arch Ophthalmol* 96, 1193-8.

Barnicot, N.A. (1952) Albinism in south-western Nigeria. *Ann Eugen* 17, 38-73.

Barreiro, L.B., Neyrolles, O., Babb, C.L., Tailleux, L., Quach, H., McElreavey, K., Helden, P.D., Hoal, E.G., Gicquel, B. and Quintana-Murci, L. (2006) Promoter variation in the DC-SIGN-encoding gene CD209 is associated with tuberculosis. *PLoS Med* 3, e20

Barrenas, M.L. and Lindgren, F. (1990) The influence of inner ear melanin on susceptibility to TTS in humans. *Scand Audiol* 19, 97-102.

Barton, D.E., Kwon, B.S. and Francke, U. (1988) Human tyrosinase gene, mapped to chromosome 11 (q14----q21), defines second region of homology with mouse chromosome 7. *Genomics* 3, 17-24.

Bassi, M.T., Schiaffino, M.V., Renieri, A., De Nigris, F., Galli, L., Bruttini, M., Gebbia, M., Bergen, A.A., Lewis, R.A. and Ballabio, A. (1995) Cloning of the gene for ocular albinism type 1 from the distal short arm of the X chromosome. *Nat Genet* 10, 13-9.

Bellamy, R., Beyers, N., McAdam, K.P., Ruwende, C., Gie, R., Samaai, P., Bester, D., Meyer, M., Corrah, T., Collin, M., Camidge, D.R., Wilkinson, D., Hoal-Van Helden, E., Whittle, H.C., Amos, W., van Helden, P. and Hill, A.V. (2000) Genetic susceptibility to tuberculosis in Africans: a genome-wide scan. *Proc Natl Acad Sci U S A* 97, 8005-9.

Bellamy, R., Ruwende, C., Corrah, T., McAdam, K.P., Whittle, H.C. and Hill, A.V. (1998) Variations in the NRAMP1 gene and susceptibility to tuberculosis in West Africans. *N Engl J Med* 338, 640-4.

Bennett, D.C. and Lamoreux, M.L. (2003) The color loci of mice--a genetic century. *Pigment Cell Res* 16, 333-44.

Bergsma, D.R. and Kaiser-Kupfer, M. (1974) A new form of albinism. *Am J Ophthalmol* 77, 837-44.

Blackwell, J.M. (1998) Genetics of host resistance and susceptibility to intramacrophage pathogens: a study of multicase families of tuberculosis, leprosy and leishmaniasis in north-eastern Brazil. *Int J Parasitol* 28, 21-8.

Boissy, R.E. and Nordlund, J.J. (1997) Molecular basis of congenital hypopigmentary disorders in humans: a review. *Pigment Cell Res* 10, 12-24.

Boissy, R. (1998a) In: Nordlund, J.J., Boissy, R., Hearing, V.J., and Ortonne, J-P (Eds). The Pigmentary System – Physiology and Pathophysiology. Oxford University Press, New York, pp 59-72

Boissy, R.E., Sakai, C., Zhao, H., Kobayashi, T. and Hearing, V.J. (1998b) Human tyrosinase related protein-1 (TRP-1) does not function as a DHICA oxidase activity in contrast to murine TRP-1. *Exp Dermatol* 7, 198-204.

Bonifer, C., Bosch, F.X., Faust, N., Schuhmann, A. and Sippel, A.E. (1994) Evolution of gene regulation as revealed by differential regulation of the chicken lysozyme transgene and the endogenous mouse lysozyme gene in mouse macrophages. *Eur J Biochem* 226, 227-35.

Bornman, L., Campbell, S.J., Fielding, K., Bah, B., Sillah, J., Gustafson, P., Manneh, K., Lisse, I., Allen, A., Sirugo, G., Sylla, A., Aaby, P., McAdam, K.P., Bah-Sow, O., Bennett, S., Lienhardt, C. and Hill, A.V. (2004) Vitamin D receptor polymorphisms and susceptibility to tuberculosis in West Africa: a case-control and family study. *J Infect Dis* 190, 1631-41.

Brilliant, M.H., Gondo, Y. and Eicher, E.M. (1991) Direct molecular identification of the mouse pink-eyed unstable mutation by genome scanning. *Science* 252, 566-9.

Buiting, K., Greger, V., Brownstein, B.H., Mohr, R.M., Voiculescu, I., Winterpacht, A., Zabel, B. and Horsthemke, B. (1992) A putative gene family in 15q11-13 and 16p11.2: possible implications for Prader-Willi and Angelman syndromes. *Proc Natl Acad Sci U S A* 89, 5457-61.

Bultman, S.J., Michaud, E.J. and Woychik, R.P. (1992) Molecular characterization of the mouse agouti locus. *Cell* 71, 1195-204.

Bungert, J., Dave, U., Lim, K.C., Lieuw, K.H., Shavit, J.A., Liu, Q. and Engel, J.D. (1995) Synergistic regulation of human beta-globin gene switching by locus control region elements HS3 and HS4. *Genes Dev* 9, 3083-96.

Burger, J.; Horn, D.; Tonnies, H.; Neitzel, H.; Reis, A. (2002) Familial interstitial 570 kbp deletion of the UBE3A gene region causing Angelman syndrome but not Prader-Willi syndrome. *Am. J. Med. Genet.* 111: 233-237.

Carrasquillo, M.M., McCallion, A.S., Puffenberger, E.G., Kashuk, C.S., Nouri, N. and Chakravarti, A. (2002) Genome-wide association study and mouse model identify interaction between RET and EDNRB pathways in Hirschsprung disease. *Nat Genet* 32, 237-44.

Casanova, J.L. and Abel, L. (2002) Genetic dissection of immunity to mycobacteria: the human model. *Annu Rev Immunol* 20, 581-620.

Casanova, J.L. and Abel, L. (2007) Human genetics of infectious diseases: a unified theory. *EMBO J* 26, 915-22.

Cellier, M., Govoni, G., Vidal, S., Kwan, T., Groulx, N., Liu, J., Sanchez, F., Skamene, E., Schurr, E. and Gros, P. (1994) Human natural resistance-associated macrophage protein: cDNA cloning, chromosomal mapping, genomic organization, and tissue-specific expression. *J Exp Med* 180, 1741-52.

Cervino, A.C., Lakiss, S., Sow, O., Bellamy, R., Beyers, N., Hoal-van Helden, E., van Helden, P., McAdam, K.P. and Hill, A.V. (2002) Fine mapping of a putative tuberculosis-susceptibility locus on chromosome 15q11-13 in African families. *Hum Mol Genet* 11, 1599-603.

Chaplin, G. (2004) Geographic distribution of environmental factors influencing human skin coloration. *Am J Phys Anthropol* 125, 292-302.

Charles, S.J., Green, J.S., Grant, J.W., Yates, J.R. and Moore, A.T. (1993) Clinical features of affected males with X linked ocular albinism. *Br J Ophthalmol* 77, 222-7.

Charles, S.J., Moore, A.T., Grant, J.W. and Yates, J.R. (1992) Genetic counselling in X-linked ocular albinism: clinical features of the carrier state. *Eye* 6 (Pt 1), 75-9.

Christian, S.L., Robinson, W.P., Huang, B., Mutirangura, A., Line, M.R., Nakao, M., Surti, U., Chakravarti, A. and Ledbetter, D.H. (1995) Molecular characterization of two proximal deletion breakpoint regions in both Prader-Willi and Angelman syndrome patients. *Am J Hum Genet* 57, 40-8.

Cohen, T., Muller, R.M., Tomita, Y. and Shibahara, S. (1990) Nucleotide sequence of the cDNA encoding human tyrosinase-related protein. *Nucleic Acids Res* 18, 2807-8.

Colman, M.A., Stevens, G., Ramsay, M., Kwon, B. and Jenkins, T. (1993) Exclusion of two candidate pigment loci, c and b, part of chromosome 11p, and 33 random polymorphic markers as the locus for tyrosinase-positive oculocutaneous albinism. *Hum Genet* 90, 556-60.

Comings, D.E. and Odland, G.F. (1966) Partial albinism. *JAMA* 195, 519-23.

Comstock, G.W. (1982) Epidemiology of tuberculosis. *Am Rev Respir Dis* 125, 8-15.

Comstock, G.W. (1978) Tuberculosis in twins: a re-analysis of the Prophit survey. *Am Rev Respir Dis* 117, 621-4.

Conlee, J.W., Abdul-Baqi, K.J., McCandless, G.A. and Creel, D.J. (1986) Differential susceptibility to noise-induced permanent threshold shift between albino and pigmented guinea pigs. *Hear Res* 23, 81-91.

Costin, G.E., Valencia, J.C., Vieira, W.D., Lamoreux, M.L. and Hearing, V.J. (2003) Tyrosinase processing and intracellular trafficking is disrupted in mouse primary melanocytes carrying the underwhite (uw) mutation. A model for oculocutaneous albinism (OCA) type 4. *J Cell Sci* 116, 3203-12.

Creel, D., O'Donnell, F.E. Jr and Witkop, C.J. Jr (1978) Visual system anomalies in human ocular albinos. *Science* 201, 931-3.

del Marmol, V., Ito, S., Jackson, I., Vachtenheim, J., Berr, P., Ghanem, G., Morandini, R., Wakamatsu, K. and Huez, G. (1993) TRP-1 expression correlates with eumelanogenesis in human pigment cells in culture. *FEBS Lett* 327, 307-10.

Delgado, J.C., Tsai, E.Y., Thim, S., Baena, A., Boussiotis, V.A., Reynes, J.M., Sath, S., Grosjean, P., Yunis, E.J. and Goldfeld, A.E. (2002) Antigen-specific and persistent tuberculin anergy in a cohort of pulmonary tuberculosis patients from rural Cambodia. *Proc Natl Acad Sci U S A* 99, 7576-81.

Diamond, J. (2005) Evolutionary biology: geography and skin colour. *Nature* 435, 283-4.

Dittrich, B., Buiting, K., Korn, B., Rickard, S., Buxton, J., Saitoh, S., Nicholls, R.D., Poustka, A., Winterpacht, A., Zabel, B. and Horsthemke, B. (1996) Imprint switching on human chromosome 15 may involve alternative transcripts of the SNRPN gene. *Nat Genet* 14, 163-70.

Duffy, D.L., Montgomery, G.W., Chen, W., Zhao, Z.Z., Le, L., James, M.R., Hayward, N.K., Martin, N.G., Sturm, R.A. (2007) A three-single-nucleotide polymorphism haplotype in intron 1 of OCA2 explains most human eye-color variation. *Am J Hum Genet* 80, 241-52.

Durham-Pierre, D., Gardner, J.M., Nakatsu, Y., King, R.A., Francke, U., Ching, A., Aquaron, R., del Marmol, V. and Brilliant, M.H. (1994) African origin of an intragenic deletion of the human P gene in tyrosinase positive oculocutaneous albinism. *Nat Genet* 7, 176-9.

Eady, R.A., Gunner, D.B., Garner, A. and Rodeck, C.H. (1983) Prenatal diagnosis of oculocutaneous albinism by electron microscopy of fetal skin. *J Invest Dermatol* 80, 210-2.

Eiberg, H. and Mohr, J. (1996) Assignment of genes coding for brown eye colour (BEY2) and brown hair colour (HCL3) on chromosome 15q. *Eur J Hum Genet* 4, 237-41.

Fernando, S.L. and Britton, W.J. (2006) Genetic susceptibility to mycobacterial disease in humans. *Immunol Cell Biol* 84, 125-37.

Fleischman, R.A., Saltman, D.L., Stastny, V. and Zneimer, S. (1991) Deletion of the c-kit protooncogene in the human developmental defect piebald trait. *Proc Natl Acad Sci U S A* 88, 10885-9.

Flint, J., Harding, R.M., Boyce, A.J. and Clegg, J.B. (1998) The population genetics of the haemoglobinopathies. *Baillieres Clin Haematol* 11, 1-51.

Flores-Villanueva, P.O., Ruiz-Morales, J.A., Song, C.H., Flores, L.M., Jo, E.K., Montano, M., Barnes, P.F., Selman, M. and Granados, J. (2005) A functional promoter polymorphism in monocyte

chemoattractant protein-1 is associated with increased susceptibility to pulmonary tuberculosis. *J Exp Med* 202, 1649-58.

Forsius, H. and Eriksson, A.W. (1964) [A new eye syndrome with X-chromasomal transmission.]. *Klin Monatsbl Augenheilkd* 144, 447-57.

Frieden, T.R., Sterling, T.R., Munsiff, S.S., Watt, C.J. and Dye, C. (2003) Tuberculosis. *Lancet* 362, 887-99.

Friedlaender, J.S., Sgaramella-Zonta, L.A., Kidd, K.K., Lai, L.Y., Clark, P. and Walsh, R.J. (1971) Biological divergences in south-central Bougainville: an analysis of blood polymorphism gene frequencies and anthropometric measurements utilizing tree models, and a comparison of these variables with linguistic, geographic, and migrational "distances". *Am J Hum Genet* 23, 253-70.

Froggatt, P. (1960) Albinism in Northern Ireland. *Ann Hum Genet* 24, 213-38.

Frudakis, T., Terravainen, T., Thomas, M. (2007) Multilocus OCA2 genotypes specify human iris colors. *Hum Genet* 122, 311-26.

Fryburg, J.S., Breg, W.R. and Lindgren, V. (1991) Diagnosis of Angelman syndrome in infants. *Am J Med Genet* 38, 58-64.

Fryer, J.P., Oetting, W.S. and King, R.A. (2003) Identification and characterization of a DNase hypersensitive region of the human tyrosinase gene. *Pigment Cell Res* 16, 679-84.

Fukai, K., Oh, J., Frenk, E., Almodovar, C. and Spritz, R.A. (1995) Linkage disequilibrium mapping of the gene for Hermansky-Pudlak syndrome to chromosome 10q23.1-q23.3. *Hum Mol Genet* 4, 1665-9.

Gardner, J.M., Nakatsu, Y., Gondo, Y., Lee, S., Lyon, M.F., King, R.A. and Brilliant, M.H. (1992) The mouse pink-eyed dilution gene: association with human Prader-Willi and Angelman syndromes. *Science* 257, 1121-4.

Gershoni-Baruch, R., Benderly, A., Brandes, J.M. and Gilhar, A. (1991) Dopa reaction test in hair bulbs of fetuses and its application to the prenatal diagnosis of albinism. *J Am Acad Dermatol* 24, 220-2.

Gershoni-Baruch, R., Rosenmann, A., Droetto, S., Holmes, S., Tripathi, R.K. and Spritz, R.A. (1994) Mutations of the tyrosinase gene in patients with oculocutaneous albinism from various ethnic groups in Israel. *Am J Hum Genet* 54, 586-94.

Giebel, L.B. and Spritz, R.A. (1991a) Mutation of the KIT (mast/stem cell growth factor receptor) protooncogene in human piebaldism. *Proc Natl Acad Sci U S A* 88, 8696-9.

Giebel, L.B., Strunk, K.M. and Spritz, R.A. (1991b) Organization and nucleotide sequences of the human tyrosinase gene and a truncated tyrosinase-related segment. *Genomics* 9, 435-45.

Glenn, C.C., Driscoll, D.J., Yang, T.P. and Nicholls, R.D. (1997) Genomic imprinting: potential function and mechanisms revealed by the Prader-Willi and Angelman syndromes. *Mol Hum Reprod* 3, 321-32.

Goldfeld, A.E., Delgado, J.C., Thim, S., Bozon, M.V., Uglialoro, A.M., Turbay, D., Cohen, C. and Yunis, E.J. (1998) Association of an HLA-DQ allele with clinical tuberculosis. *JAMA* 279, 226-8.

Goldgeier, M.H., Klein, L.E., Klein-Angerer, S., Moellmann, G. and Nordlund, J.J. (1984) The distribution of melanocytes in the leptomeninges of the human brain. *J Invest Dermatol* 82, 235-8.

Gondo, Y., Gardner, J.M., Nakatsu, Y., Durham-Pierre, D., Deveau, S.A., Kuper, C. and Brilliant, M.H. (1993) High-frequency genetic reversion mediated by a DNA duplication: the mouse pink-eyed unstable mutation. *Proc Natl Acad Sci U S A* 90, 297-301.

Halaban, R. and Moellmann, G. (1990) Murine and human b locus pigmentation genes encode a glycoprotein

(gp75) with catalase activity. *Proc Natl Acad Sci U S A* 87, 4809-13.

Handel, M.A., Washburn, L.L., Rosenberg, M.P. and Eicher, E.M. (1987) Male sterility caused by p6H and qk mutations is not corrected in chimeric mice. *J Exp Zool* 243, 81-92.

Hawksworth, N.R., Headland, S., Good, P., Thomas, N.S. and Clarke, A. (1995) Aland island eye disease: clinical and electrophysiological studies of a Welsh family. *Br J Ophthalmol* 79, 424-30.

Hayashi, K. and Yandell, D.W. (1993) How sensitive is PCR-SSCP? *Hum Mutat* 2, 338-46.

Hearing, V.J. (1998) In: Nordlund, J.J., Boissy, R., Hearing, V.J., and Ortonne, J-P (Eds). The Pigmentary System – Physiology and Pathophysiology. Oxford University Press, New York, pp 423-438

Herndon, C.N. and Jennings, R.G. (1951) A twin-family study of susceptibility to poliomyelitis. *Am J Hum Genet* 3, 17-46.

Hill, A.V. (2006) Aspects of genetic susceptibility to human infectious diseases. *Annu Rev Genet* 40, 469-86.

Hill, A.V., Allsopp, C.E., Kwiatkowski, D., Anstey, N.M., Twumasi, P., Rowe, P.A., Bennett, S., Brewster, D., McMichael, A.J. and Greenwood, B.M. (1991) Common west African HLA antigens are associated with protection from severe malaria. *Nature* 352, 595-600.

Hoal, E.G., Lewis, L.A., Jamieson, S.E., Tanzer, F., Rossouw, M., Victor, T., Hillerman, R., Beyers, N., Blackwell, J.M. and Van Helden, P.D. (2004) SLC11A1 (NRAMP1) but not SLC11A2 (NRAMP2) polymorphisms are associated with susceptibility to tuberculosis in a high-incidence community in South Africa. *Int J Tuberc Lung Dis* 8, 1464-71.

Hoekstra, H.E. (2006) Genetics, development and evolution of adaptive pigmentation in vertebrates. *Heredity* 97, 222-34.

Holm, V.A., Cassidy, S.B., Butler, M.G., Hanchett, J.M., Greenswag, L.R., Whitman, B.Y. and Greenberg, F. (1993) Prader-Willi syndrome: consensus diagnostic criteria. *Pediatrics* 91, 398-402.

Hoth, C.F., Milunsky, A., Lipsky, N., Sheffer, R., Clarren, S.K. and Baldwin, C.T. (1993) Mutations in the paired domain of the human PAX3 gene cause Klein-Waardenburg syndrome (WS-III) as well as Waardenburg syndrome type I (WS-I). *Am J Hum Genet* 52, 455-62.

Hughes, A.E., Newton, V.E., Liu, X.Z. and Read, A.P. (1994) A gene for Waardenburg syndrome type 2 maps close to the human homologue of the microphthalmia gene at chromosome 3p12-p14.1. *Nat Genet* 7, 509-12.

Hunt, D.M. and Johnson, D.R. (1971) Abnormal spermiogenesis in two pink-eyed sterile mutants in the mouse. *J Embryol Exp Morphol* 26, 111-21.

Inagaki, K., Suzuki, T., Shimizu, H., Ishii, N., Umezawa, Y., Tada, J., Kikuchi, N., Takata, M., Takamori, K., Kishibe, M., Tanaka, M., Miyamura, Y., Ito, S. and Tomita, Y. (2004) Oculocutaneous albinism type 4 is one of the most common types of albinism in Japan. *Am J Hum Genet* 74, 466-71.

Jablonski, D. (2004) Extinction: past and present. *Nature* 427 , 589

Jablonski, N.G. and Chaplin, G. (2000) The evolution of human skin coloration. *J Hum Evol* 39, 57-106.

Jackson, I.J. (1988) A cDNA encoding tyrosinase-related protein maps to the brown locus in mouse. *Proc Natl Acad Sci U S A* 85, 4392-6.

Jackson, I.J. (1993) Molecular genetics. Colour-coded switches. *Nature* 362, 587-8.

Jackson, I.J., Chambers, D., Rinchik, E.M. and Bennett, D.C. (1990) Characterization of TRP-1 mRNA

levels in dominant and recessive mutations at the mouse brown (b) locus. *Genetics* 126, 451-9.

Jackson, I.J., Chambers, D.M., Tsukamoto, K., Copeland, N.G., Gilbert, D.J., Jenkins, N.A. and Hearing, V. (1992) A second tyrosinase-related protein, TRP-2, maps to and is mutated at the mouse slaty locus. *EMBO J* 11, 527-35.

Jagle, U., Muller, A.M., Kohler, H. and Bonifer, C. (1997) Role of positive and negative cis-regulatory elements in the transcriptional activation of the lysozyme locus in developing macrophages of transgenic mice. *J Biol Chem* 272, 5871-9.

Jenkins, T., Heim, R.A., Dunn, D.S., Zwane, E., Colman, M.A., Ramsay, M. and Kromberg, J.G. (1990) In quest of the tyrosinase-positive oculocutaneous albinism gene. *Ophthalmic Paediatr Genet* 11, 251-4.

Ji, Y., Walkowicz, M.J., Buiting, K., Johnson, D.K., Tarvin, R.E., Rinchik, E.M., Horsthemke, B., Stubbs, L. and Nicholls, R.D. (1999) The ancestral gene for transcribed, low-copy repeats in the Prader-Willi/Angelman region encodes a large protein implicated in protein trafficking, which is deficient in mice with neuromuscular and spermiogenic abnormalities. *Hum Mol Genet* 8, 533-42.

John, P.R., Makova, K., Li, W.H., Jenkins, T. and Ramsay, M. (2003) DNA polymorphism and selection at the melanocortin-1 receptor gene in normally pigmented southern African individuals. *Ann N Y Acad Sci* 994, 299-306.

John, P.R. and Ramsay, M. (2002) Four novel variants in MC1R in red-haired South African individuals of European descent: S83P, Y152X, A171D, P256S. *Hum Mutat* 19, 461-2.

Johnson, L.R., Pilder, S.H. and Olds-Clarke, P. (1995) The cellular basis for interaction of sterility factors in the mouse t haplotype. *Genet Res* 66, 189-93.

Jones, M.E., Ponsonby, A.L., Dwyer, T. and Gilbert, N. (1994) The relation between climatic temperature and sudden infant death syndrome differs among communities: results from an ecologic analysis. *Epidemiology* 5, 332-6.

Jones, P.A., Rideout, W.M. 3rd, Shen, J.C., Spruck, C.H. and Tsai, Y.C. (1992) Methylation, mutation and cancer. *Bioessays* 14, 33-6.

Kaur, G., Sachdeva, G., Bhutani, L.K. and Bamezai, R. (1997) Association of polymorphism at COL3A and CTLA4 loci on chromosome 2q31-33 with the clinical phenotype and in-vitro CMI status in healthy and leprosy subjects: a preliminary study. *Hum Genet* 100, 43-50.

Kedda, M.A., Stevens, G., Manga, P., Viljoen, C., Jenkins, T. and Ramsay, M. (1994) The tyrosinase-positive oculocutaneous albinism gene shows locus homogeneity on chromosome 15q11-q13 and evidence of multiple mutations in southern African negroids. *Am J Hum Genet* 54, 1078-84.

Kerr, R., Stevens, G., Manga, P., Salm, S., John, P., Haw, T. and Ramsay, M. (2000) Identification of P gene mutations in individuals with oculocutaneous albinism in sub-Saharan Africa. *Hum Mutat* 15, 166-72.

Kettaneh, A., Seng, L., Tiev, K.P., Toledano, C., Fabre, B. and Cabane, J. (2006) Human leukocyte antigens and susceptibility to tuberculosis: a meta-analysis of case-control studies. *Int J Tuberc Lung Dis* 10, 717-25.

King, R.A., Creel, D., Cervenka, J., Okoro, A.N. and Witkop, C.J. (1980) Albinism in Nigeria with delineation of new recessive oculocutaneous type. *Clin Genet* 17, 259-70.

King, R.A., Lewis, R.A., Townsend, D., Zelickson, A., Olds, D.P. and Brumbaugh, J. (1985a) Brown oculocutaneous albinism. Clinical, ophthalmological, and biochemical characterization. *Ophthalmology* 92, 1496-505.

King, R.A. and Olds, D.P. (1985b) Hairbulb tyrosinase activity in oculocutaneous albinism: suggestions for

pathway control and block location. *Am J Med Genet* 20, 49-55.

King, R.A., Olds, D.P. and Witkop, C.J. (1978) Characterization of human hairbulb tyrosinase: properties of normal and albino enzyme. *J Invest Dermatol* 71, 136-9.

King, R.A., Pietsch, J., Fryer, J.P., Savage, S., Brott, M.J., Russell-Eggitt, I., Summers, C.G. and Oetting, W.S. (2003a) Tyrosinase gene mutations in oculocutaneous albinism 1 (OCA1): definition of the phenotype. *Hum Genet* 113, 502-13.

King, R.A., Townsend, D., Oetting, W., Summers, C.G., Olds, D.P., White, J.G. and Spritz, R.A. (1991) Temperature-sensitive tyrosinase associated with peripheral pigmentation in oculocutaneous albinism. *J Clin Invest* 87, 1046-53.

King, R.A., Willaert, R.K., Schmidt, R.M., Pietsch, J., Savage, S., Brott, M.J., Fryer, J.P., Summers, C.G. and Oetting, W.S. (2003b) MC1R mutations modify the classic phenotype of oculocutaneous albinism type 2 (OCA2). *Am J Hum Genet* 73, 638-45.

King, R.A., Wirtschafter, J.D., Olds, D.P. and Brumbaugh, J. (1986) Minimal pigment: a new type of oculocutaneous albinism. *Clin Genet* 29, 42-50.

King, R.A. and Witkop, C.J. Jr (1976) Hairbulb tyrosinase activity in oculocutaneous albinism. *Nature* 263, 69-71.

Kishino, T., Lalande, M. and Wagstaff, J. (1997) UBE3A/E6-AP mutations cause Angelman syndrome. *Nat Genet* 15, 70-3.

Kmita, M., van Der Hoeven, F., Zakany, J., Krumlauf, R. and Duboule, D. (2000) Mechanisms of Hox gene colinearity: transposition of the anterior Hoxb1 gene into the posterior HoxD complex. *Genes Dev* 14, 198-211.

Knoll, J.H., Nicholls, R.D., Magenis, R.E., Graham, J.M. Jr, Lalande, M. and Latt, S.A. (1989) Angelman and Prader-Willi syndromes share a common chromosome 15 deletion but differ in parental origin of the deletion. *Am J Med Genet* 32, 285-90.

Kobayashi, T., Urabe, K., Winder, A., Tsukamoto, K., Brewington, T., Imokawa, G., Potterf, B. and Hearing, V.J. (1994) DHICA oxidase activity of TRP1 and interactions with other melanogenic enzymes. *Pigment Cell Res* 7, 227-34.

Korner, A. and Pawelek, J. (1982) Mammalian tyrosinase catalyzes three reactions in the biosynthesis of melanin. *Science* 217, 1163-5.

Kramnik, I., Dietrich, W.F., Demant, P. and Bloom, B.R. (2000) Genetic control of resistance to experimental infection with virulent Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A* 97, 8560-5.

Kromberg, J.G., Castle, D., Zwane, E.M. and Jenkins, T. (1989) Albinism and skin cancer in Southern Africa. *Clin Genet* 36, 43-52.

Kromberg, J.G., Castle, D.J., Zwane, E.M., Bothwell, J., Kidson, S., Bartel, P., Phillips, J.I. and Jenkins, T. (1990) Red or rufous albinism in southern Africa. *Ophthalmic Paediatr Genet* 11, 229-35.

Kromberg, J.G. and Jenkins, T. (1984) Albinism in the South African Negro. III. Genetic counselling issues. *J Biosoc Sci* 16, 99-108.

Kromberg, J.G., Zwane, E., Castle, D. and Jenkins, T. (1987a) Albinism in South African blacks. *Lancet* 2, 388-9.

Kromberg, J.G., Zwane, E.M. and Jenkins, T. (1987b) The response of black mothers to the birth of an albino infant. *Am J Dis Child* 141, 911-6.

Kurahashi, H., Inagaki, H., Hosoba, E., Kato, T., Ohye, T., Kogo, H. and Emanuel, B.S. (2007) Molecular cloning of a translocation breakpoint hotspot in 22q11. *Genome Res* 17, 461-9.

Kwiatkowski, D.P. (2005) How malaria has affected the human genome and what human genetics can teach us about malaria. *Am J Hum Genet* 77, 171-92.

Kwon, B.S., Haq, A.K., Pomerantz, S.H. and Halaban, R. (1987) Isolation and sequence of a cDNA clone for human tyrosinase that maps at the mouse c-albino locus. *Proc Natl Acad Sci U S A* 84, 7473-7.

Kwon, H.Y., Bultman, S.J., Loffler, C., Chen, W.J., Furdon, P.J., Powell, J.G., Usala, A.L., Wilkison, W., Hansmann, I. and Woychik, R.P. (1994) Molecular structure and chromosomal mapping of the human homolog of the agouti gene. *Proc Natl Acad Sci U S A* 91, 9760-4.

Lakich, D., Kazazian, H.H. Jr, Antonarakis, S.E. and Gitschier, J. (1993) Inversions disrupting the factor VIII gene are a common cause of severe haemophilia A. *Nat Genet* 5, 236-41.

Lamprecht, J. and Meyer zum Gottesberge, A.M. (1988) The presence and localization of receptors for atrial natriuretic peptide in the inner ear of the guinea pig. *Arch Otorhinolaryngol* 245, 300-1.

LaSalle, J.M. and Lalande, M. (1996) Homologous association of oppositely imprinted chromosomal domains. *Science* 272, 725-8.

Lee, S.T., Nicholls, R.D., Bundey, S., Laxova, R., Musarella, M. and Spritz, R.A. (1994a) Mutations of the P gene in oculocutaneous albinism, ocular albinism, and Prader-Willi syndrome plus albinism. *N Engl J Med* 330, 529-34.

Lee, S.T., Nicholls, R.D., Jong, M.T., Fukai, K. and Spritz, R.A. (1995) Organization and sequence of the human P gene and identification of a new family of transport proteins. *Genomics* 26, 354-63.

Lee, S.T., Nicholls, R.D., Schnur, R.E., Guida, L.C., Lu-Kuo, J., Spinner, N.B., Zackai, E.H. and Spritz, R.A. (1994b) Diverse mutations of the P gene among African-Americans with type II (tyrosinase-positive) oculocutaneous albinism (OCA2). *Hum Mol Genet* 3, 2047-51.

Lehman, A.L., Nakatsu, Y., Ching, A., Bronson, R.T., Oakey, R.J., Keiper-Hrynko, N., Finger, J.N., Durham-Pierre, D., Horton, D.B., Newton, J.M., Lyon, M.F. and Brilliant, M.H. (1998) A very large protein with diverse functional motifs is deficient in rjs (runty, jerky, sterile) mice. *Proc Natl Acad Sci U S A* 95, 9436-41.

Lehrman, M.A., Schneider, W.J., Sudhof, T.C., Brown, M.S., Goldstein, J.L. and Russell, D.W. (1985) Mutation in LDL receptor: Alu-Alu recombination deletes exons encoding transmembrane and cytoplasmic domains. *Science* 227, 140-6.

Lewis, S.E., Turchin, H.A. and Wojtowicz, T.E. (1978) Fertility studies of complementing genotypes at the albino locus of the mouse. *J Reprod Fertil* 53, 197-202.

Lin, T.M., Chen, C.J., Wu, M.M., Yang, C.S., Chen, J.S., Lin, C.C., Kwang, T.Y., Hsu, S.T., Lin, S.Y. and Hsu, L.C. (1989) Hepatitis B virus markers in Chinese twins. *Anticancer Res* 9, 737-41.

Liu, B. and Hannun, Y.A. (1997) Inhibition of the neutral magnesium-dependent sphingomyelinase by glutathione. *J Biol Chem* 272, 16281-7.

Lombard, Z., Dalton, D.L., Venter, P.A., Williams, R.C. and Bornman, L. (2006) Association of HLA-DR, -DQ, and vitamin D receptor alleles and haplotypes with tuberculosis in the Venda of South Africa. *Hum Immunol* 67, 643-54.

Luande, J., Henschke, C.I. and Mohammed, N. (1985) The Tanzanian human albino skin. Natural history. *Cancer* 55, 1823-8.

Lund, P.M. (1996) Distribution of oculocutaneous albinism in Zimbabwe. *J Med Genet* 33, 641-4.

Lupski, J.R. (2007) Genomic rearrangements and sporadic disease. *Nat Genet* 39, (7 Suppl):S43-7.

Lyon, M.F., King, T.R., Gondo, Y., Gardner, J.M., Nakatsu, Y., Eicher, E.M. and Brilliant, M.H. (1992) Genetic and molecular analysis of recessive alleles at the pink-eyed dilution (p) locus of the mouse. *Proc Natl Acad Sci U S A* 89, 6968-72.

Magenis, R.E., Smith, L., Nadeau, J.H., Johnson, K.R., Mountjoy, K.G. and Cone, R.D. (1994) Mapping of the ACTH, MSH, and neural (MC3 and MC4) melanocortin receptors in the mouse and human. *Mamm Genome* 5, 503-8.

Magnus, P. (1992) Genetic epidemiology--possibilities and problems. *Scand J Soc Med* 20, 193-5.

Manga, P., Boissy, R.E., Pifko-Hirst, S., Zhou, B.K. and Orlow, S.J. (2001a) Mislocalization of melanosomal proteins in melanocytes from mice with oculocutaneous albinism type 2. *Exp Eye Res* 72, 695-710.

Manga, P., Kromberg, J., Turner, A., Jenkins, T. and Ramsay, M. (2001b) In Southern Africa, brown oculocutaneous albinism (BOCA) maps to the OCA2 locus on chromosome 15q: P-gene mutations identified. *Am J Hum Genet* 68, 782-7.

Manga, P., Kromberg, J.G., Box, N.F., Sturm, R.A., Jenkins, T. and Ramsay, M. (1997) Rufous oculocutaneous albinism in southern African Blacks is caused by mutations in the TYRP1 gene. *Am J Hum Genet* 61, 1095-101.

Manga, P. and Orlow, S.J. (2001) Inverse correlation between pink-eyed dilution protein expression and induction of melanogenesis by bafilomycin A1. *Pigment Cell Res* 14, 362-7.

Marquet, S., Abel, L., Hillaire, D., Dessein, H., Kalil, J., Feingold, J., Weissenbach, J. and Dessein, A.J. (1996) Genetic localization of a locus controlling the intensity of infection by Schistosoma mansoni on chromosome 5q31-q33. *Nat Genet* 14, 181-4.

Matsuura, T., Sutcliffe, J.S., Fang, P., Galjaard, R.J., Jiang, Y.H., Benton, C.S., Rommens, J.M. and Beaudet, A.L. (1997) De novo truncating mutations in E6-AP ubiquitin-protein ligase gene (UBE3A) in Angelman syndrome. *Nat Genet* 15, 74-7.

Meyer zum Gottesberge, A.M. (1988) Physiology and pathophysiology of inner ear melanin. *Pigment Cell Res* 1, 238-49.

Meyers, W.M. and Marty, A.M. (1991) Current concepts in the pathogenesis of leprosy. Clinical, pathological, immunological and chemotherapeutic aspects. *Drugs* 41, 832-56.

Michaud, E.J., Bultman, S.J., Stubbs, L.J. and Woychik, R.P. (1993) The embryonic lethality of homozygous lethal yellow mice (Ay/Ay) is associated with the disruption of a novel RNA-binding protein. *Genes Dev* 7, 1203-13.

Miller, E.N., Jamieson, S.E., Joberty, C., Fakiola, M., Hudson, D., Peacock, C.S., Cordell, H.J., Shaw, M.A., Lins-Lainson, Z., Shaw, J.J., Ramos, F., Silveira, F. and Blackwell, J.M. (2004) Genome-wide scans for leprosy and tuberculosis susceptibility genes in Brazilians. *Genes Immun* 5, 63-7.

Miller, S.A., Dykes, D.D. and Polesky, H.F. (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 16, 1215

Milot, E., Strouboulis, J., Trimborn, T., Wijgerde, M., de Boer, E., Langeveld, A., Tan-Un, K., Vergeer, W., Yannoutsos, N., Grosveld, F. and Fraser, P. (1996) Heterochromatin effects on the frequency and duration of LCR-mediated gene transcription. *Cell* 87, 105-14.

Mira, M.T., Alcais, A., Van Thuc, N., Thai, V.H., Huong, N.T., Ba, N.N., Verner, A., Hudson, T.J., Abel, L. and Schurr, E. (2003) Chromosome 6q25 is linked to susceptibility to leprosy in a Vietnamese population. *Nat Genet* 33, 412-5.

Mishima, Y. and Imokawa, G. (1983) Selective aberration and pigment loss in melanosomes of malignant melanoma cells in vitro by glycosylation inhibitors: premelanosomes as glycoprotein. *J Invest Dermatol* 81, 106-14.

Mitchell, R.E. (1968) The skin of the Australian Aborigine; a light and electronmicroscopical study. *Australas J Dermatol* 9, 314-28.

Modiano, D., Petrarca, V., Sirima, B.S., Nebie, I., Diallo, D., Esposito, F. and Coluzzi, M. (1996) Different response to Plasmodium falciparum malaria in west African sympatric ethnic groups. *Proc Natl Acad Sci U S A* 93, 13206-11.

Morell, R., Spritz, R.A., Ho, L., Pierpont, J., Guo, W., Friedman, T.B. and Asher, J.H. Jr (1997) Apparent digenic inheritance of Waardenburg syndrome type 2 (WS2) and autosomal recessive ocular albinism (AROA). *Hum Mol Genet* 6, 659-64.

Mountjoy, K.G., Robbins, L.S., Mortrud, M.T. and Cone, R.D. (1992) The cloning of a family of genes that encode the melanocortin receptors. *Science* 257, 1248-51.

Murty, V.V., Bouchard, B., Mathew, S., Vijayasaradhi, S. and Houghton, A.N. (1992) Assignment of the human TYRP (brown) locus to chromosome region 9p23 by nonradioactive in situ hybridization. *Genomics* 13, 227-9.

Nagle, D.L., Karim, M.A., Woolf, E.A., Holmgren, L., Bork, P., Misumi, D.J., McGrail, S.H., Dussault, B.J. Jr, Perou, C.M., Boissy, R.E., Duyk, G.M., Spritz, R.A. and Moore, K.J. (1996) Identification and mutation analysis of the complete gene for Chediak-Higashi syndrome. *Nat Genet* 14, 307-11.

Nance, W.E., Jackson, C.E. and Witkop, C.J. Jr (1970) Amish albinism: a distinctive autosomal recessive phenotype. *Am J Hum Genet* 22, 579-86.

Naylor, J., Brinke, A., Hassock, S., Green, P.M. and Giannelli, F. (1993) Characteristic mRNA abnormality found in half the patients with severe haemophilia A is due to large DNA inversions. *Hum Mol Genet* 2, 1773-8.

Newton, J.M., Cohen-Barak, O., Hagiwara, N., Gardner, J.M., Davisson, M.T., King, R.A. and Brilliant, M.H. (2001) Mutations in the human orthologue of the mouse underwhite gene (uw) underlie a new form of oculocutaneous albinism, OCA4. *Am J Hum Genet* 69, 981-8.

Nicholls, R.D. (1994) New insights reveal complex mechanisms involved in genomic imprinting. *Am J Hum Genet* 54, 733-40.

Nicholls, R.D., Fischel-Ghodsian, N. and Higgs, D.R. (1987) Recombination at the human alpha-globin gene cluster: sequence features and topological constraints. *Cell* 49, 369-78.

Nicholls, R.D., Ohta, T. and Gray, T.A. (1999) Genetic abnormalities in Prader-Willi syndrome and lessons from mouse models. *Acta Paediatr Suppl* 88, 99-104.

Nishimoto, T., Seino, H., Seki, N. and Hori, T.A. (1994) The human CHC1 gene encoding RCC1 (regulator of chromosome condensation) (CHC1) is localized to human chromosome 1p36.1. *Genomics* 23, 719-21.

O'Donnell, F.E. Jr, King, R.A., Green, W.R. and Witkop, C.J. Jr (1978) Autosomal recessively inherited ocular albinism. A new form of ocular albinism affecting females as severely as males. *Arch Ophthalmol* 96, 1621-5.

Oakey, R.J., Keiper, N.M., Ching, A.S. and Brilliant, M.H. (1996) Molecular analysis of the cDNAs encoded by the pun and pJ alleles of the pink-eyed dilution locus. *Mamm Genome* 7, 315-6.

Ochman, H., Gerber, A.S. and Hartl, D.L. (1988) Genetic applications of an inverse polymerase chain reaction. *Genetics* 120, 621-3.

Oetting, W.S., Fryer, J.P., Oofuji, Y., Middendorf, L.R., Brumbaugh, J.A., Summers, C.G. and King, R.A. (1994) Analysis of tyrosinase gene mutations using direct automated infrared fluorescence DNA sequencing of amplified exons. *Electrophoresis* 15, 159-64.

Oetting, W.S., Garrett, S.S., Brott, M. and King, R.A. (2005) P gene mutations associated with oculocutaneous albinism type II (OCA2). *Hum Mutat* 25, 323

Oh, J., Bailin, T., Fukai, K., Feng, G.H., Ho, L., Mao, J.I., Frenk, E., Tamura, N. and Spritz, R.A. (1996) Positional cloning of a gene for Hermansky-Pudlak syndrome, a disorder of cytoplasmic organelles. *Nat Genet* 14, 300-6.

Okoro, A.N. (1975) Albinism in Nigeria. A clinical and social study. *Br J Dermatol* 92, 485-92.

Oral, H.B., Budak, F., Uzaslan, E.K., Basturk, B., Bekar, A., Akalin, H., Ege, E., Ener, B. and Goral, G. (2006) Interleukin-10 (IL-10) gene polymorphism as a potential host susceptibility factor in tuberculosis. *Cytokine* 35, 143-7.

Orita, M., Suzuki, Y., Sekiya, T. and Hayashi, K. (1989) Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics* 5, 874-9.

Ottenhoff, T.H., Verreck, F.A., Hoeve, M.A. and van de Vosse, E. (2005) Control of human host immunity to mycobacteria. *Tuberculosis (Edinb)* 85, 53-64.

Packard, R.M. (1989) Industrial production, health and disease in sub-Saharan Africa. *Soc Sci Med* 28, 475-96.

Passmore, L.A., Kaesmann-Kellner, B. and Weber, B.H. (1999) Novel and recurrent mutations in the tyrosinase gene and the P gene in the German albino population. *Hum Genet* 105, 200-10.

Pearson, C.A. (1978) Hypochromia as a clinical sign of tuberculosis in the tropics. *Tubercle* 59, 111-9.

Pearson, C.A. (1989) Hypomelanosis as an aid to early recognition of tuberculosis in the African. *Trop Doct* 19, 188

Pentao, L., Wise, C.A., Chinault, A.C., Patel, P.I. and Lupski, J.R. (1992) Charcot-Marie-Tooth type 1A duplication appears to arise from recombination at repeat sequences flanking the 1.5 Mb monomer unit. *Nat Genet* 2, 292-300.

Perou, C.M., Moore, K.J., Nagle, D.L., Misumi, D.J., Woolf, E.A., McGrail, S.H., Holmgren, L., Brody, T.H., Dussault, B.J. Jr, Monroe, C.A., Duyk, G.M., Pryor, R.J., Li, L., Justice, M.J. and Kaplan, J. (1996) Identification of the murine beige gene by YAC complementation and positional cloning. *Nat Genet* 13, 303-8.

Petrini, B. (2006) Mycobacterium abscessus: an emerging rapid-growing potential pathogen. *APMIS* 114, 319-28.

Phelan, M.C., Morton, C.C., Stevenson, R.E., Tanzi, R.E., Stewart, G.D., Watkins, P.C., Gusella, J.F. and Amos, J.A. (1988) Molecular and cytogenetic characterization of a de novo t(5p;21q) in a patient previously diagnosed as monosomy 21. *Am J Hum Genet* 43, 511-9.

Porter, S. and Mintz, B. (1991) Multiple alternatively spliced transcripts of the mouse tyrosinase-encoding gene. *Gene* 97, 277-82.

Posthuma, D., Visscher, P.M., Willemsen, G., Zhu, G., Martin, N.G., Slagboom, P.E., de Geus, E.J., Boomsma, D.I. (2006) Replicated linkage for eye color on 15q using comparative ratings of sibling pairs. *Behav Genet* 36,12-7.

Puri, N., Durbam-Pierre, D., Aquaron, R., Lund, P.M., King, R.A. and Brilliant, M.H. (1997) Type 2 oculocutaneous albinism (OCA2) in Zimbabwe and Cameroon: distribution of the 2.7-kb deletion allele of the P gene. *Hum Genet* 100, 651-6.

Ramsay, M., Colman, M.A., Stevens, G., Zwane, E., Kromberg, J., Farrall, M. and Jenkins, T. (1992) The tyrosinase-positive oculocutaneous albinism locus maps to chromosome 15q11.2-q12. *Am J Hum Genet* 51, 879-84.

Rana, B.K., Hewett-Emmett, D., Jin, L., Chang, B.H., Sambuughin, N., Lin, M., Watkins, S., Bamshad, M., Jorde, L.B., Ramsay, M., Jenkins, T. and Li, W.H. (1999) High polymorphism at the human melanocortin 1 receptor locus. *Genetics* 151, 1547-57.

Raviglione, M. (2006) XDR-TB: entering the post-antibiotic era? *Int J Tuberc Lung Dis* 10, 1185-7.

Ray, K., Chaki, M. and Sengupta, M. (2007) Tyrosinase and ocular diseases: Some novel thoughts on the molecular basis of oculocutaneous albinism type 1. *Prog Retin Eye Res* 26, 323-58.

Razin, S.V., Ioudinkova, E.S., Trifonov, E.N. and Scherrer, K. (2001) Non-clonability correlates with genomic instability: a case study of a unique DNA region. *J Mol Biol* 307, 481-6.

Reed, W.B., Stone, V.M., Boder, E. and Ziprkowski, L. (1967) Pigmentary disorders in association with congenital deafness. *Arch Dermatol* 95, 176-86.

Relethford, J.H. (1997) Hemispheric difference in human skin color. *Am J Phys Anthropol* 104, 449-57.

Richard, I. and Beckmann, J.S. (1995) How neutral are synonymous codon mutations? *Nat Genet* 10, 259.

Richard, I., Broux, O., Allamand, V., Fougerousse, F., Chiannilkulchai, N., Bourg, N., Brenguier, L., Devaud, C., Pasturaud, P., Roudaut, C. and et, a.l. (1995) Mutations in the proteolytic enzyme calpain 3 cause limb-girdle muscular dystrophy type 2A. *Cell* 81, 27-40.

Rinchik, E.M., Bultman, S.J., Horsthemke, B., Lee, S.T., Strunk, K.M., Spritz, R.A., Avidano, K.M., Jong, M.T. and Nicholls, R.D. (1993) A gene for the mouse pink-eyed dilution locus and for human type II oculocutaneous albinism. *Nature* 361, 72-6.

Rinchik, E.M., Carpenter, D.A. and Handel, M.A. (1995) Pleiotropy in microdeletion syndromes: neurologic and spermatogenic abnormalities in mice homozygous for the p6H deletion are likely due to dysfunction of a single gene. *Proc Natl Acad Sci U S A* 92, 6394-8.

Roberts, D.F., Kromberg, J.G. and Jenkins, T. (1986) Differentiation of heterozygotes in recessive albinism. *J Med Genet* 23, 323-7.

Rose, N.C., Menacker, S.J., Schnur, R.E., Jackson, L., McDonald-McGinn, D.M., Stump, T., Emanuel, B.S. and Zackai, E.H. (1992) Ocular albinism in a male with del (6)(q13-q15): candidate region for autosomal recessive ocular albinism? *Am J Med Genet* 42, 700-5.

Rosemblat, S., Durham-Pierre, D., Gardner, J.M., Nakatsu, Y., Brilliant, M.H. and Orlow, S.J. (1994) Identification of a melanosomal membrane protein encoded by the pink-eyed dilution (type II oculocutaneous albinism) gene. *Proc Natl Acad Sci U S A* 91, 12071-75.

Rossouw, M., Nel, H.J., Cooke, G.S., van Helden, P.D. and Hoal, E.G. (2003) Association between tuberculosis and a polymorphic NFkappaB binding site in the interferon gamma gene. *Lancet* 361, 1871-2.

Rundshagen, U., Zuhlke, C., Opitz, S., Schwinger, E. and Kasmann-Kellner, B. (2004) Mutations in the MATP gene in five German patients affected by oculocutaneous albinism type 4. *Hum Mutat* 23, 106-10.

Russell, L.B., Montgomery, C.S., Cacheiro, N.L. and Johnson, D.K. (1995) Complementation analyses for 45 mutations encompassing the pink-eyed dilution (p) locus of the mouse. *Genetics* 141, 1547-62.

Saitoh, S., Buiting, K., Rogan, P.K., Buxton, J.L., Driscoll, D.J., Arnemann, J., Konig, R., Malcolm, S., Horsthemke, B. and Nicholls, R.D. (1996) Minimal definition of the imprinting center and fixation of chromosome 15q11-q13 epigenotype by imprinting mutations. *Proc Natl Acad Sci U S A* 93, 7811-5.

Sambrook, J., Fritsch, E.F., Maniatis I. (1989) Molecular Cloning: A laboratory manual, second edition. Cold Spring Harbour Press.

Sapienza, C. (1990) Parental imprinting of genes. *Sci Am* 263 , 52-60.

Schnur, R.E., Nussbaum, R.L., Anson-Cartwright, L., McDowell, C., Worton, R.G. and Musarella, M.A. (1991) Linkage analysis in X-linked ocular albinism. *Genomics* 9, 605-13.

Selvaraj, P., Sriram, U., Mathan Kurian, S., Reetha, A.M. and Narayanan, P.R. (2001) Tumour necrosis factor alpha (-238 and -308) and beta gene polymorphisms in pulmonary tuberculosis: haplotype analysis with HLA-A, B and DR genes. *Tuberculosis (Edinb)* 81, 335-41.

Shaw, M.A., Donaldson, I.J., Collins, A., Peacock, C.S., Lins-Lainson, Z., Shaw, J.J., Ramos, F., Silveira, F. and Blackwell, J.M. (2001) Association and linkage of leprosy phenotypes with HLA class II and tumour necrosis factor genes. *Genes Immun* 2, 196-204.

Sheffield, V.C., Beck, J.S., Kwitek, A.E., Sandstrom, D.W. and Stone, E.M. (1993) The sensitivity of single-strand conformation polymorphism analysis for the detection of single base substitutions. *Genomics* 16, 325-32.

Shibahara, S., Taguchi, H., Muller, R.M., Shibata, K., Cohen, T., Tomita, Y. and Tagami, H. (1991) Structural organization of the pigment cell-specific gene located at the brown locus in mouse. Its promoter activity and alternatively spliced transcript. *J Biol Chem* 266, 15895-901.

Shibahara, S., Tomita, Y., Sakakura, T., Nager, C., Chaudhuri, B. and Muller, R. (1986) Cloning and expression of cDNA encoding mouse tyrosinase. *Nucleic Acids Res* 14, 2413-27.

Shibata, K., Takeda, K., Tomita, Y., Tagami, H. and Shibahara, S. (1992) Downstream region of the human tyrosinase-related protein gene enhances its promoter activity. *Biochem Biophys Res Commun* 184, 568-75.

Shields, E.D., Russell, D.A. and Pericak-Vance, M.A. (1987) Genetic epidemiology of the susceptibility to leprosy. *J Clin Invest* 79, 1139-43.

Siddiqui, M.R., Meisner, S., Tosh, K., Balakrishnan, K., Ghei, S., Fisher, S.E., Golding, M., Shanker Narayan, N.P., Sitaraman, T., Sengupta, U., Pitchappan, R. and Hill, A.V. (2001) A major susceptibility locus for leprosy in India maps to chromosome 10p13. *Nat Genet* 27, 439-41.

Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* 98, 503-17.

Spritz, R.A. (1992a) The molecular basis of human piebaldism. *Pigment Cell Res* 5, 340-3.

Spritz, R.A. (1993) Molecular genetics of oculocutaneous albinism. *Semin Dermatol* 12, 167-72.

Spritz, R.A. (1994) Molecular basis of human piebaldism. *J Invest Dermatol* 103, 137S-140S.

Spritz, R.A. and Beighton, P. (1998) Piebaldism with deafness: molecular evidence for an expanded syndrome. *Am J Med Genet* 75, 101-3.

Spritz, R.A., Fukai, K., Holmes, S.A. and Luande, J. (1995) Frequent intragenic deletion of the P gene in Tanzanian patients with type II oculocutaneous albinism (OCA2). *Am J Hum Genet* 56, 1320-3.

Spritz, R.A., Giebel, L.B. and Holmes, S.A. (1992b) Dominant negative and loss of function mutations of the c-kit (mast/stem cell growth factor receptor) proto-oncogene in human piebaldism. *Am J Hum Genet* 50, 261-9.

Spritz, R.A., Lee, S.T., Fukai, K., Brondum-Nielsen, K., Chitayat, D., Lipson, M.H., Musarella, M.A., Rosenmann, A. and Weleber, R.G. (1997) Novel mutations of the P gene in type II oculocutaneous albinism (OCA2). *Hum Mutat* 10, 175-7.

Spritz, R.A., Strunk, K.M., Giebel, L.B. and King, R.A. (1990) Detection of mutations in the tyrosinase gene in a patient with type IA oculocutaneous albinism. *N Engl J Med* 322, 1724-8.

Spritz, R.A., Strunk, K.M., Hsieh, C.L., Sekhon, G.S. and Francke, U. (1991) Homozygous tyrosinase gene mutation in an American black with tyrosinase-negative (type IA) oculocutaneous albinism. *Am J Hum Genet* 48, 318-24.

Stead, W.W., Senner, J.W., Reddick, W.T. and Lofgren, J.P. (1990) Racial differences in susceptibility to infection by Mycobacterium tuberculosis. *N Engl J Med* 322, 422-7.

Stevens, G., Ramsay, M. and Jenkins, T. (1997) Oculocutaneous albinism (OCA2) in sub-Saharan Africa: distribution of the common 2.7-kb P gene deletion mutation. *Hum Genet* 99, 523-7.

Stevens, G., van Beukering, J., Jenkins, T. and Ramsay, M. (1995) An intragenic deletion of the P gene is the common mutation causing tyrosinase-positive oculocutaneous albinism in southern African Negroids. *Am J Hum Genet* 56, 586-91.

Sturm, R. A.; Baker, E.; Sutherland, G. R. (1994) Assignment of the tyrosinase-related protein-2 gene (TYRP2) to human chromosome 13q31-q32 by fluorescence in situ hybridization: extended synteny with mouse chromosome 14. *Genomics* 21: 293-296.

Sturm, R. A.; O'Sullivan B. J.; Box, N. F.; Smith, A. G.; Smit, S. E.; Puttick, E. R. J.; Parsons, P. G.; Dunn, I. S. (1995) Chromosomal structure of the human TYRP1 and TYRP2 loci and comparison of the tyrosinase-related protein gene family. *Genomics* 29: 24-34.

Sturm, R.A., Box, N.F. and Ramsay, M. (1998) Human pigmentation genetics: the difference is only skin deep. *Bioessays* 20, 712-21.

Sturm, R.A., Teasdale, R.D. and Box, N.F. (2001) Human pigmentation genes: identification, structure and consequences of polymorphic variation. *Gene* 277, 49-62.

Sturm, R.A. and Frudakis, T.N. (2004) Eye colour: portals into pigmentation genes and ancestry. *Trends Genet* 20, 327-32.

Sulem, P., Gudbjartsson, D.F., Stacey, S.N., Helgason, A., Rafnar, T., Magnusson, K.P., Manolescu, A., Karason, A., Palsson, A., Thorleifsson, G., Jakobsdottir, M., Steinberg, S., Pálsson, S., Jonasson, F., Sigurgeirsson, B., Thorisdottir, K., Ragnarsson, R., Benediktsdottir, K.R., Aben, K.K., Kiemeney, L.A., Olafsson, J.H., Gulcher, J., Kong, A., Thorsteinsdottir, U., Stefansson, K. (2007) Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nat Genet* 39, 1443-52.

Suzuki, I., Tada, A., Ollmann, M.M., Barsh, G.S., Im, S., Lamoreux, M.L., Hearing, V.J., Nordlund, J.J. and Abdel-Malek, Z.A. (1997) Agouti signaling protein inhibits melanogenesis and the response of human melanocytes to alpha-melanotropin. *J Invest Dermatol* 108, 838-42.

Szabo, G., Gerald, A.B., Pathak, M.A. and Fitzpatrick, T.B. (1969) Racial differences in the fate of melanosomes in human epidermis. *Nature* 222 , 1081-2.

Tai, T., Eisinger, M., Ogata, S. and Lloyd, K.O. (1983) Glycoproteins as differentiation markers in human malignant melanoma and melanocytes. *Cancer Res* 43, 2773-9.

Tailleux, L., Gicquel, B. and Neyrolles, O. (2003) [DC-SIGN, a key receptor of Mycobacterium tuberculosis?]. *Med Sci (Paris)* 19, 658-60.

Takiff, H.E. (2007) In: Tuberculosis 2007. From Basic Science to Patient Care. E-published www.tuberculosistextbook.com pp 207-262.

Tassabehji, M., Newton, V.E. and Read, A.P. (1994) Waardenburg syndrome type 2 caused by mutations in the human microphthalmia (MITF) gene. *Nat Genet* 8, 251-5.

Tassabehji, M., Read, A.P., Newton, V.E., Harris, R., Balling, R., Gruss, P. and Strachan, T. (1992) Waardenburg's syndrome patients have mutations in the human homologue of the Pax-3 paired box gene. *Nature* 355, 635-6.

Taylor, W.O. (1987) Prenatal diagnosis in albinism. *Lancet* 1, 1307-8.

Thye, T., Browne, E.N., Chinbuah, M.A., Gyapong, J., Osei, I., Owusu-Dabo, E., Niemann, S., Rusch-Gerdes, S., Horstmann, R.D. and Meyer, C.G. (2006) No associations of human pulmonary tuberculosis with Sp110 variants. *J Med Genet* 43, e32

Toda, K., Pathak, M.A., Parrish, J.A., Fitzpatrick, T.B. and Quevedo, W.C. Jr (1972) Alteration of racial differences in melanosome distribution in human epidermis after exposure to ultraviolet light. *Nat New Biol* 236, 143-5.

Tosh, K., Campbell, S.J., Fielding, K., Sillah, J., Bah, B., Gustafson, P., Manneh, K., Lisse, I., Sirugo, G., Bennett, S., Aaby, P., McAdam, K.P., Bah-Sow, O., Lienhardt, C., Kramnik, I. and Hill, A.V. (2006) Variants in the SP110 gene are associated with genetic susceptibility to tuberculosis in West Africa. *Proc Natl Acad Sci U S A* 103, 10364-8.

Triglia, T., Peterson, M.G. and Kemp, D.J. (1988) A procedure for in vitro amplification of DNA segments that lie outside the boundaries of known sequences. *Nucleic Acids Res* 16, 8186

Tripathi, R.K., Giebel, L.B., Strunk, K.M. and Spritz, R.A. (1991) A polymorphism of the human tyrosinase gene is associated with temperature-sensitive enzymatic activity. *Gene Expr* 1, 103-10.

Tsukamoto, K., Jackson, I.J., Urabe, K., Montague, P.M. and Hearing, V.J. (1992) A second tyrosinase-related protein, TRP-2, is a melanogenic enzyme termed DOPAchrome tautomerase. *EMBO J* 11, 519-26.

Valverde, P., Healy, E., Jackson, I., Rees, J.L. and Thody, A.J. (1995) Variants of the melanocyte-stimulating hormone receptor gene are associated with red hair and fair skin in humans. *Nat Genet* 11, 328-30.

van Dorp, D.B., Eriksson, A.W., Delleman, J.W., van Vliet, A.G., Collewijn, H., van Balen, A.T. and Forsius, H.R. (1985) Aland eye disease: no albino misrouting. *Clin Genet* 28, 526-31.

van Helden, P. and Hoal-van Helden, E. (1999) Mannose-binding lectin and meningococcal disease. *Lancet* 354, 337-8.

Vidal, S.M., Malo, D., Vogan, K., Skamene, E. and Gros, P. (1993) Natural resistance to infection with intracellular parasites: isolation of a candidate for Bcg. *Cell* 73, 469-85.

Walkowicz, M., Ji, Y., Ren, X., Horsthemke, B., Russell, L.B., Johnson, D., Rinchik, E.M., Nicholls, R.D. and Stubbs, L. (1999) Molecular characterization of radiation- and chemically induced mutations associated with neuromuscular tremors, runting, juvenile lethality, and sperm defects in jdf2 mice. *Mamm Genome* 10, 870-8.

Wallis, C.E. and Beighton, P.H. (1989) Synchrony of oculocutaneous albinism, the Prader-Willi syndrome, and a normal karyotype. *J Med Genet* 26, 337-9.

Walsh, R.J. (1971) A distinctive pigment of the skin in New Guinea indigenes. *Ann Hum Genet* 34, 379-88.

Walsh, R.J. (1971) A project of the International Biological Programme in New Guinea. *Med J Aust* 1, 1235-9.

Walther, C., Guenet, J.L., Simon, D., Deutsch, U., Jostes, B., Goulding, M.D., Plachov, D., Balling, R. and Gruss, P. (1991) Pax: a murine multigene family of paired box-containing genes. *Genomics* 11, 424-34.

Weisner, B., Hartard, C. and Dieu, C. (1987) CSF copper concentration: a new parameter for diagnosis and monitoring therapy of Wilson's disease with cerebral manifestation. *J Neurol Sci* 79, 229-37.

Wilcox, E.R., Rivolta, M.N., Ploplis, B., Potterf, S.B. and Fex, J. (1992) The PAX3 gene is mapped to human chromosome 2 together with a highly informative CA dinucleotide repeat. *Hum Mol Genet* 1, 215

Wildenberg, S.C., Oetting, W.S., Almodovar, C., Krumwiede, M., White, J.G. and King, R.A. (1995) A gene causing Hermansky-Pudlak syndrome in a Puerto Rican population maps to chromosome 10q2. *Am J Hum Genet* 57, 755-65.

Wilkinson, R.J., Llewelyn, M., Toossi, Z., Patel, P., Pasvol, G., Lalvani, A., Wright, D., Latif, M. and Davidson, R.N. (2000) Influence of vitamin D deficiency and vitamin D receptor polymorphisms on tuberculosis among Gujarati Asians in west London: a case-control study. *Lancet* 355, 618-21.

Wilkinson, R.J., Patel, P., Llewelyn, M., Hirsch, C.S., Pasvol, G., Snounou, G., Davidson, R.N. and Toossi, Z. (1999) Influence of polymorphism in the genes for the interleukin (IL)-1 receptor antagonist and IL-1beta on tuberculosis. *J Exp Med* 189, 1863-74.

Williams, C.A., Angelman, H., Clayton-Smith, J., Driscoll, D.J., Hendrickson, J.E., Knoll, J.H., Magenis, R.E., Schinzel, A., Wagstaff, J., Whidden, E.M. and et, a.l. (1995) Angelman syndrome: consensus for diagnostic criteria.Angelman Syndrome Foundation. *Am J Med Genet* 56, 237-8.

Winder, A.J., Wittbjer, A., Odh, G., Rosengren, E. and Rorsman, H. (1994) The mouse brown (b) locus protein functions as a dopachrome tautomerase. *Pigment Cell Res* 7, 305-10.

Witkop, C.J. Jr (1989) Albinism. *Clin Dermatol* 7, 80-91.

Witkop, C.J. Jr, Hill, C.W., Desnick, S., Thies, J.K., Thorn, H.L., Jenkins, M. and White, J.G. (1973) Ophthalmologic, biochemical, platelet, and ultrastructural defects in the various types of oculocutaneous albinism. *J Invest Dermatol* 60, 443-56.

Witkop, C.J. Jr, Niswander, J.D., Bergsma, D.R., Workman, P.L. and White, J.G. (1972) Tyrosinase positive oculocutaneous albinism among the Zuni and the Brandywine triracial isolate: biochemical and clinical characteristics and fertility. *Am J Phys Anthropol* 36, 397-405.

Woolf, C.M. and Grant, R.B. (1962) Albinism among the Hopi Indians in Arizona. *Am J Hum Genet* 14, 391-400.

Wutz, K., Sauer, C., Zrenner, E., Lorenz, B., Alitalo, T., Broghammer, M., Hergersberg, M., de la Chapelle, A., Weber, B.H., Wissinger, B., Meindl, A. and Pusch, C.M. (2002) Thirty distinct CACNA1F mutations in 33 families with incomplete type of XLCSNB and Cacna1f expression profiling in mouse retina. *Eur J Hum Genet* 10, 449-56.

Yamamoto, H., Ito, K., Ishiguro, S. and Takeuchi, T. (1987) Gene controlling a differentiation step in the quail melanocyte. *Dev Genet* 8, 179-85.

Yasumoto, K., Mahalingam, H., Suzuki, H., Yoshizawa, M. and Yokoyama, K. (1995) Transcriptional activation of the melanocyte-specific genes by the human homolog of the mouse Microphthalmia protein. *J Biochem (Tokyo)* 118, 874-81.

Yen, P.H., Li, X.M., Tsai, S.P., Johnson, C., Mohandas, T. and Shapiro, L.J. (1990) Frequent deletions of the human X chromosome distal short arm result from recombination between low copy repetitive elements. *Cell* 61, 603-10.

Yi, Z., Garrison, N., Cohen-Barak, O., Karafet, T.M., King, R.A., Erickson, R.P., Hammer, M.F. and Brilliant, M.H. (2003) A 122.5-kilobase deletion of the P gene underlies the high prevalence of oculocutaneous albinism type 2 in the Navajo population. *Am J Hum Genet* 72, 62-72.

Zimmerman, M.R. (1979) Pulmonary and osseous tuberculosis in an Egyptian mummy. *Bull N Y Acad Med* 55, 604-8.

APPENDIX

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

Division of the Deputy Registrar (Research)

COMMITTEE FOR RESEARCH ON HUMAN SUBJECTS (MEDICAL)
Ref: R14/49 Ramsay

CLEARANCE CERTIFICATE PROTOCOL NUMBER M 940711

PROJECT Detection of mutations within the genes for brown, rufous & tyrosinase-positive oculoutaneous albinism & structural ....

INVESTIGATORS Dr M Ramsay

DEPARTMENT Human Genetics, SAIMR

DATE CONSIDERED 940729

DECISION OF THE COMMITTEE *

Unconditionally approved

DATE 940818

CHAIRMAN. .....P Bloom pp (Professor P E Cleaton-Jones)

* Guidelines for written "informed consent" attached where applicable.

c c Supervisor: Professor T Jenkins
Dept of Human Genetics, SAIMR

Works2\other\heclear\ 940711

==========

DECLARATION OF INVESTIGATOR(S)

To be completed in duplicate and ONE COPY returned to the Secretary at Room 10001, 10th Floor, Senate House, University.

I/we fully understand the conditions under which I am/we are authorized to carry out the abovementioned research and I/we guarantee to ensure compliance with these conditions. Should any departure to be contemplated from the research procedure as approved I/we undertake to resubmit the protocol to the Committee.

DATE.....SIGNATURE .....M. Ramsay.....

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

Division of the Deputy Registrar (Research)

COMMITTEE FOR RESEARCH ON HUMAN SUBJECTS (MEDICAL)
Ref: R14/49 Ramsay/Kerr

CLEARANCE CERTIFICATE **PROTOCOL NUMBER** M980616

PROJECT A Search For Host Susceptibility Loci For M Tuberculosis

INVESTIGATORS Prof/Dr M/R Ramsay/Kerr

DEPARTMENT Dept of Human Genetics, SAIMR

DATE CONSIDERED 980626

DECISION OF THE COMMITTEE *

Approved unconditionally

DATE 980730 **CHAIRMAN**.....(Professor P E Cleaton-Jones)

* Guidelines for written "informed consent" attached where applicable.

c c Supervisor: Prof M Ramsay
Dept of Dept of Human Genetics, SAIMR

Works2\lain0015\HumEth97.wdb\M 980616
==========

DECLARATION OF INVESTIGATOR(S)

To be completed in duplicate and **ONE COPY** returned to the Secretary at Room 10001, 10th Floor, Senate House, University.

I/we fully understand the conditions under which I am/we are authorized to carry out the abovementioned research and I/we guarantee to ensure compliance with these conditions. Should any departure to be contemplated from the research procedure as approved I/we undertake to resubmit the protocol to the Committee.

DATE 11-8-98 SIGNATURE M. Ramsay Robyn Kerr

PROTOCOL NO.: M 980616

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES