Corneal Dystrophies:
Molecular genetic studies and mutational analysis of
candidate genes

Neil D. Ebenezer
A thesis submitted for the Degree of Doctor of Philosophy

- Department of Molecular Genetics
  Institute of Ophthalmology
  University College London
  Bath Street
  London EC1V 9EL

2003
Declaration

I declare that the thesis submitted for the degree of Doctor of Philosophy is composed by myself, and the work herein is my own, or that the author involved is clearly stated.

Neil David Ebenezer B.Sc.
Abstract

Corneal dystrophies are a clinically heterogeneous group of rare inherited ocular disorders that often result in a reduction in visual acuity. These dystrophies can be inherited as either an autosomal dominant or autosomal recessive trait, which can affect one or more layers of the cornea.

Autosomal dominant congenital hereditary endothelial dystrophy (CHED1) is a severe disease, which affects the corneal endothelium. The largest reported family with CHED1 was used to identify the locus for the disease on chromosome 20 using linkage analysis. The CHED1 disease interval was refined to 2.7 cM on chromosome 20p11.2 flanked by the markers D20S48 and D20S471. A physical map was created using a YAC and PAC contig to anchor sequence tagged sites (STSs) and expressed tagged sites (ESTs) to the CHED1 interval. The genes, destrin, VSX1 and the cystatin gene cluster were excluded as candidates for CHED1 (Chapter 3).

The autosomal recessive form of congenital hereditary endothelial dystrophy (CHED2) was previously localised to an 8 cM region on chromosome 20p13. A positional candidate gene approach was adopted to identify the causative gene for CHED2. In collaboration with the Chromosome 20 group at the Sanger Institute, a PAC and BAC contig was created spanning the 3.7 Mb disease interval. Analysis of the genomic sequence predicted 62 genes or transcripts within this region. A small Pakistani family, which had previously been linked to the locus on 20p13, was used to screen CDS2 for mutations (Chapter 4).

Granular corneal dystrophy (CDGG1) is an autosomal dominant disease characterised by non amyloid deposits in the stroma. A British CDGG1 pedigree was linked to chromosome 5q31-33 flanked by the markers D5S421 and D5S399. YACs from the CEPH YAC library were used to create a physical map of the disease interval. βig-H3,
previously identified as the causative gene for a number of stromal dystrophies, was screened in this pedigree and a mutation in exon 12 was identified (Chapter 5).

Cornea plana is characterised by a flattened corneal surface and can be inherited as a mild autosomal dominant disease (CNA1) or a more severe autosomal recessive form (CNA2). A small Hispanic family with CNA2 was ascertained and screened for mutations in the keratocan gene. A homozygous mutation in exon 2 was identified (Chapter 6).

This molecular genetic study of both recessive and dominant disease has contributed to our knowledge of corneal dystrophies.
Dedication

I dedicate this thesis to my family:

To my wife, Anusia, for your encouragement, support and understanding and to my children, Priya, Marisa and Niren who have been a source of inspiration.
Acknowledgements

I owe a great deal of thanks to a number people without whom this would not have been possible. To my supervisors Shomi Bhattacharya and Alison Hardcastle a special thanks. Shomi, for giving me the opportunity to register for a Ph.D., allowing me to pursue my interest in corneal dystrophies, for his enthusiasm and continuous support. Alison, for encouraging, supporting, cajoling and bullying me into finishing the thesis!!!

I would like to thank all those with whom I have collaborated: Namir Toma and Linda Ficker, Steve Tuft, Frank Larkin and Suzanne Cabral at Moorfields. Panos Deloukas and Rhian Gwilliam at the Sanger Institute.

A special thanks to Aunty Ellen and Uncle Paddy for all your support and child minding. Steph Halford for reading a number of drafts of the thesis and turning them ‘flourescent pink’!!! Jude Fitzgibbon whose friendship and frank and honest opinion was always appreciated and who helped read my initial draft. Naheed Kanuga for her support throughout. Subathra Poopalasundaram for her proof reading and editorial skills. Vanita Berry who was one of the first people to suggest that I do a Ph.D.

Thanks also to Tim Jordan, Mohammed El Ashry, Mai Abd El-Aziz, Ordan Lehmann, Louise Ocaka, Aung Tin, Sujitha Tharmarajah, Eranga Vithana, Reshma Patel and Sam Murray.

It has taken many years to finally complete this thesis and I consider myself extremely fortunate to have been part of a department that has had so many wonderful people pass through the years, whose enthusiasm and encouragement, has made it possible. I thank you all !!!!!!

Chris, Binoy, Kamal, Alex, Lesley, Anna, Jimbob, Emma, Jeffrey, Mai, Ilaria, Ro, Marcela, LiLi, Dawn, Tim Stout, Soraya, Sujeewa, Sinthuja, Zara, Bev, Jill, Phillipa, Sue Richard, Rachel, Mani, Myrto, Smaro, Andrew, Angus, Donna, Alan, Karen, Sana, Adam, Clare, Christina, Christianne, Simon, Mark, Quincy, Debbie, Miriam, Katie, Lloyd, and Merve.

Last but not least thanks to Leen and Naheed R, who have been very supportive, knowing it won’t take them as long to finish!!!
# Table of Contents

## CHAPTER 1
### Introduction 1
#### 1.1 The structure of the eye 1
##### 1.1.1 Basic anatomy 1
##### 1.1.2 Function of the eye 2
#### 1.2 The cornea 2
##### 1.2.1 Anatomy 4
###### 1.2.1.1 Epithelium 4
###### 1.2.1.2 Bowman's membrane 5
###### 1.2.1.3 Stromal layer 5
###### 1.2.1.4 Descemet's membrane 6
###### 1.2.1.5 Endothelial layer 7
##### 1.2.2 Structural proteins of the cornea 8
###### 1.2.2.1 Collagen 8
###### 1.2.2.2 Proteoglycans 8
##### 1.2.3 Corneal metabolism 10
##### 1.2.4 Corneal physiology 10
##### 1.2.5 Corneal transparency 11
##### 1.2.6 Development of the cornea 13
#### 1.3 Inherited corneal dystrophies 14
##### 1.3.1 Epithelial dystrophies 15
###### 1.3.1.1 Epithelial basement membrane dystrophy 15
###### 1.3.1.2 Meesmann's corneal dystrophy (MCDI) 16
##### 1.3.2 Stromal dystrophies 17
###### 1.3.2.1 Granular corneal dystrophy type I (CDGG1) 17
###### 1.3.2.2 Reis-Bucklers' dystrophy (CDRBI) 18
###### 1.3.2.3 Lattice corneal dystrophy (CDL) 18
###### 1.3.2.4 Avellino corneal dystrophy 20
###### 1.3.2.5 Fleck dystrophy 21
###### 1.3.2.6 Cornea plana 22
##### 1.3.3 Endothelial dystrophies 22
###### 1.3.3.1 Fuch's endothelial dystrophy (FECD) 22
###### 1.3.3.2 Congenital hereditary endothelial dystrophy (CHED) 23
###### 1.3.3.3 Posterior polymorphous corneal dystrophy (PPCD) 24
#### 1.4 Strategies to identify genes implicated in human disease 25
#### 1.5 Genetic mapping 26
##### 1.5.1 Meiotic recombination 26
##### 1.5.2 Recombination frequency 26
##### 1.5.3 Genetic map distance 27
##### 1.5.4 Polymorphic markers used in linkage analysis 28
##### 1.5.5 RFLPs and mini-satellite DNA 28
##### 1.5.6 Microsatellite markers (Short Tandem Repeat Polymorphisms) 28
##### 1.5.7 Single nucleotide polymorphisms (SNPs) 29
2.5.5 Polymerase Chain Reaction (PCR)
  2.5.5.1 Standard parameters for a typical PCR
  2.5.5.2 Primer design
  2.5.5.3 PCR optimisation
  2.5.5.4 Microsatellite analysis
    2.5.5.4.1 Incorporation labelling of PCR product
    2.5.5.4.2 End labelling of primer
    2.5.5.4.3 Fluorescent automated genotyping
  2.5.5.5 Reverse Transcriptase (RT) PCR
  2.5.5.6 Inter-ALU PCR
2.5.6 YAC/PAC library screening by PCR
  2.5.6.1 Colony PCR
  2.5.6.2 PAC library screening by PCR
2.5.7 Purification of PCR products
  2.5.7.1 Sephacryl(r) HR columns (S200/400)
  2.5.7.2. Centricron 100 spin columns (Centricron, USA)
  2.5.7.3. QIAquick spin columns (QIAGEN)
  2.5.7.4. QIAquick gel extraction (Qiagen)
2.5.8 Automated fluorescent DNA sequencing
2.5.9 DNA Hybridisation
  2.5.9.1 Random priming labelling of probes
  2.5.9.2 Hybridisation
  2.5.9.3 PAC Library Screening
  2.5.9.4 Post hybridisation washes
  2.5.9.5 Fluorescent in situ hybridisation
2.5.10 Microsatellite markers
2.5.11 Mutation Detection
  2.5.11.1 Hetereoduplex Analysis on DHPLC
  2.5.11.2 Direct sequencing
2.5.12 Tissue culture techniques
  2.5.12.1. Harvesting of cells
  2.5.12.2 Freezing of cultured cells
  2.5.12.3 Reviving frozen cells
2.5.13 Computational tools and databases
  2.5.13.1 DNA analysis
  2.5.13.2 Computer packages for Linkage
  2.5.13.3 Bioinformatics
  2.5.13.4 Web sites and databases used

CHAPTER 3
Congenital Hereditary Endothelial Dystrophy CHED1
3.1 Introduction
3.2 CHED1 family pedigree
3.3 Linkage analysis of the CHED1 pedigree
  3.3.1 Microsatellite analysis of candidate loci
  3.3.1.1 Calculation of lod scores for microsatellite markers
3.3.1.2 Exclusion of candidate loci for CHED1

3.3.2 Linkage of the CHED1 disease phenotype to chromosome 20
  3.3.2.1 Chromosome 20 microsatellite and haplotype analysis
  3.3.2.3 Multipoint Linkage Analysis
  3.3.2.4 Further Analysis of the CHED1 pedigree with new polymorphic markers

3.4 Physical mapping of the CHED1 disease interval
  3.4.1 Isolation of YAC clones
  3.4.2 FISH mapping of critical YACs
  3.4.3 Isolation and characterisation of PAC clones within the CHED1 interval

3.5 Integration of physical and genetic maps for the CHED1 interval

3.6 Genetic refinement

3.7 Identification of candidate genes
  3.7.1 Cystatin Genes
  3.7.2 VSX1
  3.7.3 Destrin
    3.7.3.1 Destrin expression in the cornea
    3.7.3.2 Amplification of destrin from patient RNA and mutation screening
    3.7.3.3 Genomic structure of destrin

3.8 Discussion
  3.8.1 Candidate genes
    3.8.1.1 Filensin or beaded filament structural protein 1 (BFSP1)
    3.8.1.2 Solute carrier 24A3 (SLC 24A3)
    3.8.1.3 CRP2BP
  3.8.2 Synteny between human chromosome 20 and mouse chromosome 2

CHAPTER 4
Congenital Hereditary Endothelial Dystrophy CHED2
  4.1 Introduction
  4.2 Pedigree and clinical data
  4.3 Linkage analysis of the CHED2 pedigree
  4.4 Positional candidate gene analysis within the CHED2 disease interval
    4.4.1 CDP-diacylglycerol synthase 2 (CDS2)
      4.4.1.1 Genomic organisation of CDS2
    4.5 Mutation screening of CDS2
  4.6 Discussion
    4.6.1 Candidate genes
      4.6.1.1 CDS2
      4.6.1.2 Transglutaminase
      4.6.1.3 Solute carriers

CHAPTER 5
Granular Corneal Dystrophy

x
5.1 Introduction
5.2 Granular corneal dystrophy type I pedigree and phenotype
5.3 Linkage study for CDGG1
   5.3.1 Refinement of the critical interval
5.4 Physical Mapping of 5q31
   5.4.1 FISH analysis of YAC clones on 5q31
5.5 Candidate Genes mapping to the critical interval on 5q31
   5.5.1 LOX1
   5.5.2 Big-H3
   5.5.3 Genomic structure of Big-H3
   5.5.4 Big-H3 mutation detection
5.6 Discussion
   5.6.1 Stromal Dystrophies mapping to chromosome 5q
   5.6.2 TGFB1 (BigH3) gene and keratoepithelin protein
   5.6.3 The role of keratoepithelin (KE) in the cornea
   5.6.4 Future Work and Conclusion

CHAPTER 6
Cornea Plana
6.1 Introduction
   6.1.1 Cornea plana
6.1.2 Linkage analysis and mutation identification in CNA2 families
6.1.3 Keratocan
6.2 CNA2 pedigree and phenotype
   6.2.1 Corneal Topography
6.3 Keratocan gene screening
6.4 Alignment of conserved residues
6.5 KERA protein modeling
6.4 Discussion

CHAPTER 7
Discussion
7.1 Overview of the work presented
7.2 Genetic heterogeneity
7.3 Genotype / Phenotype correlation
7.4 Advances in the human genome project
7.5 Animal models
7.6 Future perspectives for therapy

References

Publication
Figures

Figure 1.1 Cross-section of the human eye
Figure 1.2 Light micrograph cross-section of a H&E stained normal cornea
Figure 1.3 Light micrograph section of the anterior stroma
Figure 1.4 Schematic diagram showing corneal collagen lamellae
Figure 1.5 Light micrograph of a H&E stained section of posterior cornea
Figure 1.6 Specular microscopy showing endothelial cells at x133 magnification
Figure 1.7 Stromal collagen stained with cupromeronic blue
Figure 1.8 Schematic representation of the endothelial pump mechanism
Figure 1.9 Lattice arrangement of collagen fibrils in the stroma
Figure 1.10 Schematic showing the development of the human cornea
Figure 1.11 Epithelial basement membrane or Cogan's microcystic dystrophy
Figure 1.12 Meesmann's corneal dystrophy shows cysts
Figure 1.13 Granular corneal dystrophy showing granular deposits in the stroma
Figure 1.14 Slit lamp photograph and light micrograph of CDRBl
Figure 1.15 Lattice corneal dystrophy with refractile lines and amyloid deposits
Figure 1.16 Slit lamp photograph of Avellino corneal dystrophy
Figure 1.17 Slit lamp photograph of Fleck dystrophy with mid-stromal deposits
Figure 1.18 Slit lamp and light micrograph of a patient with Fuchs' dystrophy
Figure 1.19 Slit lamp photograph of a CHED patient with corneal haziness
Figure 1.20 Cornea of patients with PPCD
Figure 3.1 Slit lamp photograph of a cornea from a CHED1 patient
Figure 3.2 Light micrograph sections of corneal tissue
Figure 3.3 CHED1 pedigree including new generation
Figure 3.4 Microsatellite analysis of the CHED1 family for marker D5S393
Figure 3.5 Microsatellite analysis of the CHED1 family for marker D20S471
Figure 3.6 Abridged CHED1 pedigree and haplotype showing linkage to chromosome 20
Figure 3.7 Ideogram of chromosome 20
Figure 3.8 Multipoint graph of the lod scores vs recombination fraction for markers
    D20S114, D20S48 and D20S471
Figure 3.9 Novel microsatellite marker M1068E13 typed across the CHED1 pedigree
Figure 3.10 Haplotype analysis of additional markers on chromosome 20 mapping to
    the CHED1 disease interval
Figure 3.11 STS content mapping of 15 YACs using marker WI-9743
Figure 3.12 Preliminary YAC contig spanning the critical interval for CHED1 disease region to chromosome 20p11.2 using FISH
Figure 3.13 Localisation of two YACs (A) 875B10 and (B) 972H6 to the CHED1 disease
    region to chromosome 20p11.2 using FISH
Figure 3.14 Autoradiograph of radiolabelled probe WI-9743 hybridised to the
    PAC library.
Figure 3.15 Isolation of a PAC using the PCR pools for the RPCII PAC library and
    the marker D20S471
Figure 3.16 FISH of PAC196D8 to metaphase spreads
Figure 3.17 ALU- PCR of PAC196D8
Figure 3.18 Sequence of the sub-cloned PCR product obtained from Alu-PCR of 196D8 88
Figure 3.19 Minimal tiling path and sequence ready map of the CHED1 interval 90
Figure 3.20 Organisation of the family of cystatin genes on chromosome 20p 91
Figure 3.21 The expression pattern of destrin in the corneal epithelium, stroma and endothelium using RT-PCR 94
Figure 3.22 Schematic diagram of the genomic organisation of the destrin gene 95
Figure 3.23 Homology between mouse chromosome 2 and human chromosomes 1, 2, 7, 9, 10, 11,12, 15 and 20 100
Figure 4.1 An ideogram showing the CHED2 locus on chromosome 20p13 103
Figure 4.2 The opaque cornea of an affected member of the CHED2 family 105
Figure 4.3 A consanguineous Pakistani pedigree with autosomal recessive CHED 106
Figure 4.4 Schematic of the CHED2 sequence tile path 108
Figure 4.5 In situ hybridisation analysis of Cds2 in adult mouse cornea 109
Figure 4.6 Schematic diagram of the genomic organisation of the CDS2 gene on chromosome 20p13 110
Figure 5.1 Photograph of CDGG1 patient's cornea and light micrograph of a corneal section stained with alcon-blue. 115
Figure 5.2 Abridged pedigree of a family with CDGG1 used in this linkage study 117
Figure 5.3 Autoradiograph showing the alleles of microsatellite marker D5S638 in the CDGG1 pedigree 118
Figure 5.4 Haplotype data for the CDGG1 pedigree with markers on 5q 120
Figure 5.5 YAC STS content mapping for marker D5S816 122
Figure 5.6 Physical map of the CDGG1 region on 5q31-33 122
Figure 5.7 Fluorescent in situ hybridisation (FISH) of YAC 773D to metaphase spreads 123
Figure 5.8 STS content mapping of 5 YACs spanning the critical interval for CDGG1 with an STS for the LOX1 gene 125
Figure 5.9 Genomic structure of Big-H3 (TGFBI) and the reported mutations 126
Figure 5.10 Schematic representation of the domain structure of Big-H3 127
Figure 5.11 Sequence from exon 12 of Big-H3 128
Figure 5.12 Restriction digest of PCR product of exon 12 with BstX1 129
Figure 6.1 A slit lamp photograph of a patient with cornea plana. 134
Figure 6.2 Ideogram of chromosome 12 showing the position of the CNA1 and CNA2 loci. 135
Figure 6.3 Schematic diagram showing the structure of human keratocan gene. 137
Figure 6.4 A consanguineous recessive cornea plana (CNA2) Hispanic pedigree. 138
Figure 6.5 Slit lamp photograph of two individuals from the CNA2 pedigree. 139
Figure 6.6 Corneal topography of (A) an unaffected and (B) affected member of the CNA2 family 140
Figure 6.7 Electropherograms showing the A393G change in keratocan 141
Figure 6.8 A CLUSTALW alignment of human KERA with bovine, mouse, quail and chick 143
Figure 6.9 Protein prediction models of ALS and KERA 145
Tables

Table 1.1 Corneal dystrophies with the primary affected corneal layer 15
Table 3.1 Candidate loci analysed in the CHED1 pedigree 68
Table 3.2 Individuals in the pedigree used for linkage analysis 70
Table 3.3 Two-point lod scores between CHED1 and polymorphic markers 71
Table 3.4 Individuals in the pedigree used for linkage analysis. 73
Table 3.5 Two-point lod scores between CHED1 and markers on chromosome 20 75
Table 3.6 Primer sequences used to amplify novel microsatellite markers. 78
Table 3.7 Two-point lod scores between CHED1 disease phenotype and additional markers on chromosome 20. 78
Table 3.8 Individuals in the pedigree used for linkage analysis. 79
Table 3.9 Primers designed to amplify destrin cDNA 93
Table 3.10 Intronic primers designed to amplify the 4 exons of destrin 95
Table 3.11 Genes in the CHED1 region derived from genomic sequence predictions 96
Table 3.12 Homology which exists between mouse chromosome 2 and the short arm of human chromosome 20 101
Table 4.1 Primer sequences used to amplify novel microsatellite markers 106
Table 4.2 Putative genes and transcripts annotated within the CHED2 critical interval 107
Table 4.3 Primer sequences used to amplify the 13 exons of CDS2 by PCR for mutation analysis 111
Table 5.1 Microsatellite markers on chromosome 5q used in a linkage study 118
Table 5.2 Alleles and disease status of individuals scored for the marker D5S638 119
Table 5.3 Two-point lodscores for linkage between the CDGG1 phenotype and six markers spanning the locus on chromosome 5q 119
Table 5.4 Candidate genes and EST's that map within the disease interval 124
Table 5.5 Intronic primers used to identify mutations in Big-H3 127
Table 5.6 Seven different stromal dystrophies and the mutations responsible for these phenotypes 130
Table 6.1 Primer sequences and conditions used to amplify the coding region of keratocan 131
Table 7.1 Table of corneal dystrophies with chromosomal localisation, flanking markers and gene 150
**Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>adRP</td>
<td>autosomal dominant retinitis pigmentosa</td>
</tr>
<tr>
<td>APS</td>
<td>Ammonium persulphate</td>
</tr>
<tr>
<td>arRP</td>
<td>autosomal recessive retinitis pigmentosa</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine 5’-triphosphate</td>
</tr>
<tr>
<td>BLAST</td>
<td>Basic Local Alignment Search Tool</td>
</tr>
<tr>
<td>bp</td>
<td>base pair</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine Serum Albumin</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complementary deoxyribonucleic acid</td>
</tr>
<tr>
<td>DEPC</td>
<td>Diethyl pyrocarbonate</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>dNTPs</td>
<td>Deoxyribonucleic acid triphosphates</td>
</tr>
<tr>
<td>DNase</td>
<td>Deoxyribonuclease</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
</tr>
<tr>
<td>EDTA</td>
<td>1,2-diaminoethanetetraacetic acid</td>
</tr>
<tr>
<td>EGTA</td>
<td>Ethylene glycol-O,O’-bis(2-aminoethyl)-N,N,N’,N’-tetraacetic acid</td>
</tr>
<tr>
<td>FISH</td>
<td>Fluorescent <em>in situ</em> hybridisation</td>
</tr>
<tr>
<td>FITC</td>
<td>Fluoresein isothiocyanate</td>
</tr>
<tr>
<td>K+</td>
<td>Potassium ion</td>
</tr>
<tr>
<td>GITC</td>
<td>Guanidium thiocyanate</td>
</tr>
<tr>
<td>kb</td>
<td>Kilo base</td>
</tr>
<tr>
<td>Symbol</td>
<td>Term</td>
</tr>
<tr>
<td>--------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>kDa</td>
<td>kilo Daltons</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger ribonucleic acid</td>
</tr>
<tr>
<td>Na+</td>
<td>Sodium ion</td>
</tr>
<tr>
<td>ng</td>
<td>nanogram</td>
</tr>
<tr>
<td>OMIM</td>
<td>Online Mendelian Inheritance in Man</td>
</tr>
<tr>
<td>PAC</td>
<td>P1 derived artificial chromosome</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate buffered saline</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>pmole</td>
<td>picomole</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>RP</td>
<td>Retinitis pigmentosa</td>
</tr>
<tr>
<td>RNase</td>
<td>Ribonuclease</td>
</tr>
<tr>
<td>RPE</td>
<td>Retinal pigment epithelium</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>Reverse transcriptase-polymerase chain reaction</td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium dodecyl sulphate</td>
</tr>
<tr>
<td>SSC</td>
<td>Standard saline citrate</td>
</tr>
<tr>
<td>TAE</td>
<td>Tris acetate</td>
</tr>
<tr>
<td>UV</td>
<td>Ultra violet</td>
</tr>
<tr>
<td>v/v</td>
<td>Volume/volume</td>
</tr>
<tr>
<td>w/v</td>
<td>Weight/volume</td>
</tr>
<tr>
<td>YAC</td>
<td>Yeast artificial chromosome</td>
</tr>
</tbody>
</table>
CHAPTER 1

Introduction

The research described in this thesis involves genetic and molecular biology approaches including a positional cloning strategy towards the identification of the genes responsible for hereditary corneal dystrophies.

The following sections provide an overview of the structure, physiology and metabolism of the cornea for the maintenance of a transparent cornea, essential for vision. A number of inherited eye disease affect the cornea and the dystrophies affecting the different layers of the cornea are described.

The human genome project is also discussed with relation to its relevance in the identification of genes involved in ocular diseases, as well as introducing concepts in molecular biology relevant to this thesis.

1.1 The structure of the eye

1.1.1 Basic anatomy

The eye is composed of three major layers, an outer fibrous coat that consists of the sclera, the middle uveal layer (a vascular layer composed of the iris), and the retina with two major components, the neural retina and the retinal pigment epithelium (Figure 1.1).

The white sclera forms over 80% of the outer coat while the limbus forms a transition between the opaque sclera and the transparent cornea. The cornea is a highly specialised tissue that is made up of fairly rigid fibrous material which allows for the refraction and the transmission of light (Oshika et al., 1999). Ocular muscles are also attached to the sclera. The uveal layer includes the iris, the pigmented muscular structure around the pupil, the ciliary body from which the lens is suspended, and the choroid that lies between the sclera and the retina.

The retina consists of an inner neurosensory layer that contains the photoreceptor cells, rod and cones, other neuronal cells and glia. The outer layer of the retina is a monolayer of pigmented cells (RPE), which lies behind the neural retina. The main
function of the retina is to detect light and form a retinal image that can be processed by the brain (Remington 1998).

Figure 1.1 Cross-section of the human eye.
The cornea, as indicated by the red arrow lies in the anterior portion of the eye. (Pharmacia and Upjohn, Ltd)

1.1.2 Function of the eye

The majority of the refraction of light (70-80%) occurs as light passes through the transparent cornea and the anterior chamber. The iris is the variable diaphragm that expands or contracts depending on the level of light and the distance of the object. Light then passes through the posterior chamber and the lens, where a change in shape of the lens can alter its radius of curvature and focal length and where the refraction of light can accommodate the eye for near or far vision. The light then crosses a third refractive medium, the vitreous body, before striking the retina. When light hits the retina it evokes an action potential in the photoreceptors and these electrical impulses are carried to the brain by the optic nerve to be processed (Remington 1998).

1.2 The cornea

The cornea is a transparent tissue that covers the outer coat of the eye. At the limbus the cornea is continuous with the sclera (Figure 1.1). The cornea is slightly oval, with an average thickness of 650 μm at the periphery. However, at the centre it is thinner at approximately 520 μm. The adult cornea is mainly avascular, and obtains oxygen and
metabolites by diffusion from the aqueous humour across the endothelium as well as the tear film interface at the surface of the epithelium. The cornea has two main functions, to form a mechanical barrier against foreign material and to refract light. As the major refractive surface of the eye, the power of the cornea is 48 diopters, any structural change to the cornea or a loss of transparency can greatly reduce visual acuity (Bron et al., 1997; Remington 1998). The knowledge of the basic anatomy and physiology of the cornea aids the understanding of corneal diseases.

1.2.1 Anatomy

The cornea consists of five layers, the epithelium, Bowman's membrane, the stroma, Descemet's membrane and the endothelium.

![Figure 1.2 Light micrograph cross-section of an haemotoxylin-eosin stained normal cornea.](image)

The uppermost layer is the multilayered corneal epithelium (Ep) with its basement membrane. Bowman's layer separates the stroma or substantia propria (S) from the epithelium. The corneal endothelium (En) and Descemet's membrane, form the posterior layer of the cornea. Adapted from (Eagle 1999)
1.2.1.1 Epithelium

The epithelium is 50 to 90 μm thick and covers the anterior stroma (Figure 1.2). It consists of five to seven layers of stratified squamous epithelium sitting on a basement membrane. There are three types of epithelial cells (Figure 1.3). Normal epithelial growth begins in the basal layer, which has cylindrical or columnar shaped cells that have a high degree of mitotic activity.

![Light micrograph section of the anterior stroma.](image)

The epithelium showing the basal cells (C), the wing-shaped cells (W) and the flattened squamous cells (S). The anterior stroma shows keratocytes (K) and Bowman’s membrane (BM) (with permission from S.Inglis).

The dividing basal cells become the wing-shaped cells that migrate. They interdigitate with each other and adhere with large numbers of desmosomal attachments. These transform into slightly flattened squamous epithelial cells. Eventually these cells lose the attachment to the cornea and slough off into the tear film. Covering the epithelium is a layer of mucin that forms the air-tear interface which prevents the epithelium drying out and smoothes the surface of the cornea. At the junction of squamous epithelial cells there are indentations that zipper cells together (zonula occludens) and form a mechanical barrier that blocks the penetration of micro-organisms and other foreign bodies, but allows for the flow of fluids and metabolites to the stroma (Lemp et al., 1989; Mathers et al., 1992). The major components of the cytoplasm in basal, wing and squamous cells are the intermediate filament proteins, the cytokeratins.
Other major cytoskeletal proteins are actin filaments and microtubules. Actin is present in all epithelial cells but highly expressed in apical cells (Lauweryns et al., 1993; Kivela et al., 1998). Microtubules are present in mitotically active cells and are prominent in basal cells. The other cell types are neurons, melanocytes, leukocytes and modified macrophages, which are involved in the ocular hypersensitivity and immunological activity.

1.2.1.2 Bowman’s membrane

Bowman’s membrane is an acellular 12 μm thick tissue that helps to maintain corneal shape. It lies between the epithelial basement membrane and the stroma and it consists of fine collagen fibrils (Figure 1.3). There are two theories concerning the formation of this layer, either it is produced by the basal epithelial cells that also lay down the epithelial basement membrane, or it may be a homogeneus modified layer of the anterior stroma (Sevel et al., 1988). The major component of Bowman’s membrane is type VII collagen which is responsible for the binding of epithelial cells to its basement membrane and to the stroma (Gipson et al., 1987). Bowman’s membrane cannot regenerate, so in disease or trauma, lesions in the membrane are filled in with cellular scar tissue which creates permanent opacities (Cogan et al., 1964).

1.2.1.3 Stromal layer

The stroma or substantia propria is approximately 500 μm, which forms 90% of the thickness of the cornea (Figure 1.2). The main morphological characteristic of the stroma is the collagen fibrils that form stacked lamellae (Figure 1.4).

The lamellae are formed from collagen fibrils of uniform diameter, which are arranged in layers in parallel with each other and the corneal surface. There are 250 to 400 lamellae that extend from limbus to limbus and are arranged at less than 90° in the anterior stroma to almost perpendicular at the posterior stroma, where they are more compact (Hamada et al., 1974). The regular spacing of the lamellae together with the low cellular density enables the stroma to maintain the transparency of the cornea. The keratocyte is the predominant cell in the stroma and makes up 3-5% of its volume.
Keratocytes are large flat cells with a number of processes that extend out at all angles. In humans the cells lie mainly between the collagen lamellae. In transverse sections of the cornea, the keratocytes appear as long, thin, flattened cells that lie parallel to the corneal surface (McTigue 1967) (Figure 1.3). The main function of keratocytes is to synthesize and maintain the collagen fibrils and the extracellular matrix of the stroma. In response to wounding, the keratocytes are transformed into fibroblasts that proliferate and increase collagen production resulting in scar formation (Assouline et al., 1992). Other types of cells found in normal corneal stroma are lymphocytes and macrophages.

1.2.1.4 Descemet’s membrane

Descemet’s membrane lies on the posterior side of the stroma and is the basement membrane of the corneal endothelial cells (Figure 1.5). It is approximately 3 μm at birth but thickens to 12 μm by adulthood. In the adult two distinct layers can be seen, the anterior third consists of banded collagen (embryonic) while the posterior layer is composed of non-banded collagen (Schuler et al., 1996). Type VIII collagen is highly expressed in Descemet’s membrane but there is also type IV collagen, and fibronectin.
present which could form a scaffolding onto which the endothelial cells are attached (BenEzra et al., 1981).

![Figure 1.5 Light micrograph of an haemotoxylin-eosin stained section of posterior cornea.](image)

(En) Endothelial cells and its basement membrane, Descemet’s membrane (DM).
(Adapted from Eagle 1999).

### 1.2.1.5 Endothelial layer

The endothelium is the most posterior layer of the cornea and consists of a single layer of cells extending over the inner surface of Descemet's membrane (Figure 1.2 and Figure 1.5). The cells appear almost hexagonal and each cell is anchored to its neighbour with a large number of interdigitations (Foets et al., 1990; Barry et al., 1995). Cell densities vary a great deal but generally decrease with age (Figure 1.6).

![Figure 1.6 Specular microscopy showing endothelial cells at x133 magnification.](image)

(A) normal endothelium(25 year old), (B) an aged endothelium (80 year old) and (C) a dystrophic endothelium. Adapted from http://www.swmed.edu/home_pages/ophth/.
As the cells are incapable of regeneration, disease or trauma can also reduce cell density. However the endothelial cells can enlarge, reorganize and migrate in order to keep the monolayer intact (Figure 1.6 C) (Binkhorst et al., 1977).

Unlike most other cells in the cornea the endothelial cells have a large number of mitochondria as a result of the higher energy requirement due to its barrier and pump function. The endothelial cells act as a permeability barrier. If the monolayer is damaged or broken there is a resultant loss in the barrier function of the endothelium which leads to corneal oedema and a loss of transparency (Sasaki et al., 1986; Ohrloff et al., 1987; Ehlers et al., 1998).

1.2.2 Structural proteins of the cornea

There are two major structural proteins of the cornea, collagen and proteoglycan, which help to maintain the transparency of the cornea.

1.2.2.1 Collagen

Collagens are described as a number of sub-types such as fibrous collagen which includes type I, II, III, V, non-fibrous collagen including type IV and filamentous collagen such as type VI, VIII, IX and X.

The cornea unusually consists of a wide variety of these different collagen types which display a high degree of specialisation with fine and uniform fibril size and fibrillar organisation (Figure 1.7).

Collagen type I is a major constituent of the stroma, accounting for 90% of the collagens identified in this layer.

Collagen type IV is found in the basal lamina of the epithelium and is the predominant protein found in Descemet's membrane. However, collagen type V and collagen type VIII form a significant proportion of the protein in Descemet's membrane.

1.2.2.2 Proteoglycans

Proteoglycans are a class of glycosylated proteins which have covalently linked sulfated glycosaminoglycans, such as chondroitin sulfate, dermatan sulfate, heparan sulfate, or keratan sulfate which can modulate collagen fibrillogenesis.
Keratocytes of the corneal stroma secrete a number of proteoglycan molecules, mainly keratan sulfate proteoglycans (KSPGs) and dermatan sulphate proteoglycans (DSPGs) which form the extracellular matrix (Funderburgh et al., 2001). The corneal transparency is maintained by the presence of stromal collagen fibrils of small and regular diameter (22.5-35 nm), spaced at regular intervals (42-44 nm), together with the minimal hydration of the ECM.

The small leucine-rich proteoglycans (SLRPs) can be synthesised as either glycoproteins containing oligosaccharides or proteoglycans containing chondroitin/dermatan sulphate or keratan sulphate. Leucine-rich repeats (LRRs) with 20-30 amino acids, have a central common domain consisting of repeats of Asparagine (N) and Leucine (L) residues in conserved positions (LXXLXLXXNXL) which can form up to 80% of the protein moiety. The LRR is thought to function in molecular recognition processes such as cell adhesion, signal transduction, DNA repair and RNA processing (Iozzo 1999). SLRPs like lumican and keratocan are abundantly expressed in the cornea and are involved in protein-protein interactions (Figure 1.7).

Figure 1.7 Stromal collagen stained with cupromeronic blue.
The arrows indicate the interaction between proteoglycans and the collagen fibrils. (Adapted from (Bron et al., 1997)
1.2.3 Corneal metabolism

Energy is required by the cells of the cornea for continued mitosis in epithelial cells, for the regulation of hydration by the endothelial cells and to supply its metabolic needs. The peripheral cornea draws metabolites from the capillary system at the corneo-scleral limbus (Fischer et al., 1978). In order to pass through the epithelium, substances must be water soluble as well as lipid soluble. Ions and other lipid insoluble compounds are prevented from entering the cornea by the tight junctions formed by the epithelial cells. Glucose provides the energy that the epithelial cells require via three metabolic pathways, glycolysis, pentose phosphate shunt and glycogen synthesis, and is dependent on the amount of oxygen available (Friend 1979).

The aqueous humour is the liquid between the anterior and posterior chamber that maintains intraocular pressure and provides for the metabolic needs of the tissues within the anterior chamber. This includes the posterior cornea, where the endothelial cells have a high metabolic activity for the maintenance of the hydration of the stroma.

1.2.4 Corneal physiology

There are a number of characteristics of epithelial and endothelial physiology that are involved in the maintenance of corneal hydration and hence transparency.

The epithelial cells have tight zonular junctions that restrict the amount of fluid entering from the tear film. In addition there is an ionic pump mechanism (Figure 1.8) which pumps sodium and chloride ions from the tear interface towards the stroma by a sodium-potassium adenosine triphosphatase (ATPase) (Kaye et al., 1965).

The endothelial cell apical membranes have tight junctions and gap junctions between them which allows the aqueous humour access to the paracellular space, and also enables the endothelial barrier to be permeable (Hodson et al., 1976). The endothelial pump was used to explain the observation that the endothelium can transfer fluid from a swollen stroma to the aqueous humour using an energy dependent process. (Harris 1967; Kenney et al., 1984).
Figure 1.8 Schematic representation of the endothelial pump mechanism. (Adapted from Bron et al. 1997.)

The thickness of the cornea can be constantly maintained if the volume of fluid leaking into the stroma is equal to the volume of fluid actively removed from the stroma by the endothelium. (Hodson 1977; Graham et al., 1980; Hodson et al., 1981; Wigham et al., 1981; Hodson et al., 1983)

1.2.5 Corneal transparency

The cornea allows the transmission of electromagnetic radiation within the range of 365 nm to 2,500 nm which correspond to the ultraviolet and infrared regions of the spectrum respectively. The cornea absorbs the majority of ultraviolet light below 365 nm while wavelengths above 700 nm do not stimulate the photoreceptor cells (Bron et al., 1997). The transparency of the cornea is maintained as a result of a balance between its anatomy, physiology and metabolism. The endothelial and epithelial cells are arranged in a regular manner and there is an absence of vascular tissue and pigment in the cornea. The regular spacing of the collagen fibrils in a two dimensional lattice reduces light scattering as it passes through the stroma together with the scarcity of nucleated cells (Goldman et al., 1968). The endothelial pump (Figure 1.8) maintains the hydration of the stroma. If the pump fails the stroma swells and the spacing between collagen fibrils (Figure 1.9) exceeds the equal inter-fibril distance and leads to a translucent cornea.
(Maurice 1970). The corneal lamellae of the stroma contain only 75% of the water they are capable of binding in order to keep the collagen fibril distance equal. So relative dehydration of the stroma is needed for corneal transparency (Meek et al., 1993).

**Figure 1.9 Lattice arrangement of collagen fibrils in the stroma.**
Normal arrangement is shown (A), in contrast the irregular spacing of the fibrils and the distortion in the lattice framework is shown in (B) oedematous cornea. (Adapted from Smolin et al. 1994.)

**Figure 1.10 Schematic showing the development of the human cornea.**
(A) The lens induces the corneal epithelium (EP), (B) waves of migration of mesenchyme cells (MES/MS) give rise to the primordial stroma and endothelium (EN). (C) secondary waves of migration of fibroblasts (F) lays down stroma (anterior) (D) differentiated cornea with distinct epithelial, stromal and endothelial layers. (Adapted from Hay et al.; 1969)
1.2.6 Development of the cornea

The different tissues of the eye are derived from surface ectoderm, neural ectoderm, neural crest and mesoderm (Figure 1.10). The final differentiation and order of these cells are controlled by a number of complex inductive and suppressive processes which are mediated by chemical signals that work at a cellular and molecular level. The knowledge of human corneae at particular stages of development has been obtained from studies on mammals as well as those on the chicken. The avian cornea is structurally very similar to the primate cornea, as both have a Bowman’s membrane, and the developmental processes can be easily monitored in the chick. As a result avian species have been used as a model system for the development of the human cornea (Linsenmayer et al., 1998).

In the human cornea, at 33 days of gestation, the lens vesicle separates from the surface ectoderm and induces the development of the cornea (Figure 1.10 B). At this stage a thin line of basal lamina forms that gives rise to the corneal epithelium (O'Rahilly 1983).

The presence of the lens is crucial to corneal development and if the lens is removed before this stage the cornea fails to form (Coulombre et al., 1965). Neural crest derived mesenchymal cells migrate from the rim of the optic cup to form the primordial corneal endothelium. There is a second wave of migration of mesenchymal cells that later differentiate into stromal fibroblasts and keratocytes. Thereafter, detachment of the lens vesicle induces the basal layer of the epithelium to secrete collagen fibrils and glycosaminoglycans, which fills the space between the corneal epithelium and the lens forming the primary stroma. If the lens is lost prior to mesenchymal invasion this can result in microphthalmos and an opaque cornea (Hay et al., 1969). By the third month of gestation, the endothelium becomes a monolayer resting on an interrupted basal lamina, which forms part of Descemet’s membrane. This continues to grow and differentiate with the sequential secretion of linear filaments that give rise to a multilayer structure. This process occurs fairly rapidly, at 3 months there is one layer but at 6 months there are 10 layers and at birth it increases to 40 layers with a maximum thickness of 3µm. In the adult the anterior banded zone is referred to as the foetal band which remains static while the posterior non-banded zone thickens with age at the rate of 1 µm per decade.
Toward the end of the fourth month of gestation the epithelial basal lamina secretes the collagenous tissue which forms Bowman’s membrane. Between the age of 6 months to one year the majority of the growth in size of the cornea takes place. At birth the cornea is relatively large compared to the rest of the eye and the adult size of approximately 1.3 cm² is reached before the second year.

### 1.3 Inherited corneal dystrophies

Inherited corneal dystrophies are described as primary, bilateral changes of the cornea that are not associated with systemic disease or inflammation. The majority of diseases classified as corneal dystrophies tend to exhibit an autosomal dominant pattern of inheritance that presents within the first two decades of life. Corneal degenerations on the other hand, tend to be age-related alterations usually associated with other ocular or systemic diseases. The classification of corneal disease has historically been by clinical, anatomical or histopathological methods which often makes differential diagnosis a problem (Waring et al., 1978, 1978; Rodrigues et al., 1988). In this thesis an anatomical classification has been used, however as the genetic basis of these diseases are defined, a classification system based on genotype can be developed for corneal dystrophies.

The prevalence of corneal dystrophies are difficult to estimate since they are rare diseases and have generally been assessed as retrospective studies on those individuals that have undergone penetrating keratoplasty. As a result of these criteria they tend to be underdiagnosed. Fuch’s endothelial dystrophy will account for 10% of transplants while all other corneal dystrophies will account for less that 5% of transplants and of this the stromal dystrophies, granular, lattice and macular corneal dystrophies will account for 30% each. These studies have been conducted in different countries and apart from the occasional deviation the prevalences are approximately equivalent to one another (Lang et al., 1987; al Faran et al., 1991; Santo et al., 1995; Konishi et al., 1997).

Apart from Lattice type II corneal dystrophy, which was genetically mapped to chromosome 9q, none of the corneal dystrophies described had been mapped at the start of this study.
<table>
<thead>
<tr>
<th>Disease name and inheritance pattern</th>
<th>Symbol</th>
<th>OMIM number</th>
<th>Layer affected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epithelial basement membrane corneal dystrophy (AD)</td>
<td></td>
<td>121820</td>
<td>Epithelium</td>
</tr>
<tr>
<td>Meesmann’s corneal dystrophy (AD)</td>
<td></td>
<td>122100</td>
<td>Epithelium</td>
</tr>
<tr>
<td>Granular type I corneal dystrophy (AD)</td>
<td></td>
<td>121900</td>
<td>Stroma</td>
</tr>
<tr>
<td>Lattice type I corneal dystrophy (AD)</td>
<td>CDGI</td>
<td>121900</td>
<td>Stroma</td>
</tr>
<tr>
<td>Lattice type III/A corneal dystrophy</td>
<td>CDLII</td>
<td>122200</td>
<td>Stroma</td>
</tr>
<tr>
<td>Reis-Bückler corneal dystrophy (AD)</td>
<td>CDLIII/A</td>
<td>122200</td>
<td>Stroma</td>
</tr>
<tr>
<td>Avellino corneal dystrophy (AD)</td>
<td>CDRB</td>
<td></td>
<td>Stroma</td>
</tr>
<tr>
<td>Fleck Dystrophy (AD)</td>
<td></td>
<td>121850</td>
<td>Stroma</td>
</tr>
<tr>
<td>Lattice type II corneal dystrophy (AD)</td>
<td>CDLII</td>
<td>105120</td>
<td>Stroma</td>
</tr>
<tr>
<td>Thiel-Behnke corneal dystrophy (AD)</td>
<td>CDTB</td>
<td>602082</td>
<td>Stroma</td>
</tr>
<tr>
<td>Cornea plana congenita type I (AD)</td>
<td>CNA1</td>
<td>121400</td>
<td>Stroma</td>
</tr>
<tr>
<td>Cornea plana congenita type II (AR)</td>
<td>CNA2</td>
<td>217300</td>
<td>Stroma</td>
</tr>
<tr>
<td>Macular corneal dystrophy type I (AR)</td>
<td>MCDC1</td>
<td>217800</td>
<td>Stroma</td>
</tr>
<tr>
<td>Macular corneal dystrophy type II (AR)</td>
<td>MCDC2</td>
<td></td>
<td>Stroma</td>
</tr>
<tr>
<td>Congenital hereditary epithelial dystrophy (AD)</td>
<td>CHED1</td>
<td>121700</td>
<td>Endothelium</td>
</tr>
<tr>
<td>Congenital hereditary epithelial dystrophy (AR)</td>
<td>CHED2</td>
<td>217700</td>
<td>Endothelium</td>
</tr>
<tr>
<td>Posterior polymorphous dystrophy (AD)</td>
<td>PPCD</td>
<td>122000</td>
<td>Endothelium</td>
</tr>
</tbody>
</table>

Table 1.1 Corneal dystrophies with the primary affected corneal layer.
AD-autosomal dominant; AR-autosomal recessive

1.3.1 Epithelial dystrophies

The two most prevalent corneal dystrophies affecting the epithelium are epithelial basement membrane corneal dystrophy and Meesmann’s dystrophy. Both dystrophies involve episodes of recurrent cyst formation, however they are generally relatively benign and rarely result in the need for penetrating keratoplasty.

1.3.1.1 Epithelial basement membrane dystrophy

This is one of the more common anterior dystrophies of the cornea and a number of diseases can be classified as a single entity such as map-dot fingerprint, Cogan’s microcystic epithelial dystrophy OMIM No: 121820, anterior basement dystrophy and bleb (Cogan et al., 1964; Cogan et al., 1974). All of these dystrophies are inherited in an autosomal dominant fashion. These sub-classifications are considered to be part of the clinical spectrum of the disease. Symptoms occur in the third to fourth decade of life with very small cysts, occasional corneal erosion and little loss of visual acuity (Figure 1.11).
The major pathological change is in the synthesis of an abnormal basement membrane or duplication of the epithelium (Figure 1.11)(Heyworth *et al.*, 1998).

Figure 1.11 Epithelial basement membrane or Cogan's microcystic dystrophy. Slit lamp photograph and light micrograph with periodic acid-Schiff stain. Arrows show opacity caused by cellular debris trapped by duplication of the epithelium. (Adapted from Eagle 1999)

### 1.3.1.2 Meesmann's corneal dystrophy (MCDI)

Meesmann’s epithelial corneal dystrophy OMIM No: 122100, also known as juvenile epithelial corneal dystrophy, is a rare, bilateral autosomal dominant inherited disorder (Meesmann *et al.*, 1939). The patients tend to be asymptotic until the first to second decade but the epithelial cysts (Figure 1.12) can present from the first year of life. The cysts rupture causing erosion of the epithelium (Figure 1.11B). During episodic rupture of the cysts, vision can be diminished or permanently affected by subepithelial scarring (Burns 1968; Tremblay *et al.*, 1982).

Figure 1.12 Meesmann's corneal dystrophy shows cysts.  
(A) Retro-illumination of the epithelium, arrow indicates cysts.  
(B) Light micrograph stained with haemotoxylin and eosin, arrows highlight a cyst in the epithelium that lead to erosions. (Adapted from Eagle 1999).
1.3.2 Stromal dystrophies

The stroma accounts for 90% of the thickness of the cornea and the majority of dystrophies affecting the cornea also affect the stroma.

1.3.2.1 Granular corneal dystrophy type I (CDGG1)

Granular corneal dystrophy or Groenouw type I OMIM No: 121900 is an autosomal dominant, bilateral symmetrical opacity which is characterised by fine white dots or radial lines (Figure 1.13) in the central anterior stroma (Groenouw 1890).

![Figure 1.13 Granular corneal dystrophy showing granular deposits in the stroma](image)

These changes are often noted in the first decade of life but rarely affect vision at this stage. As time progresses the opacities coalesce and become larger and as they multiply, visual acuity is compromised (Sornson 1965). Occasionally corneal erosion is seen as well as mild photophobia. In the fifth to sixth decade of life visual acuity can be severely decreased but rarely warrants penetrating keratoplasty. The white deposits are composed of non-collagenous protein called hyaline and are thought to be a result of abnormal synthesis of protein or phospholipids (Waring et al., 1978; Rodrigues and Krachmer 1988).
1.3.2.2 Reis-Bucklers' dystrophy (CDRB1)

In 1917 Reis first noted the condition and later Bucklers more fully described the bilaterally symmetrical central cloudy dystrophy with an autosomal dominant pattern of inheritance (Bucklers 1949; Grayson et al., 1966). There is opacification at the level of Bowman's membrane (Figure 1.14). CDRB1 OMIM No: 121900 presents in the first to second decade with recurrent attacks of photophobia and irritation. CDRB1 is phenotypically similar to Thiel-Behnke dystrophy (CDTB), however CDTB has a honeycomb appearance while CDRB has 'curly fibres' (Thiel et al., 1967).

Figure 1.14 Slit lamp photograph and light micrograph of CDRB1. (A) the arrow shows the sub-epithelial location of the gray deposits. (B) A light micrograph shows the haemotoxylin-eosin staining of the sub-epithelial deposits of abnormal material instead of normal Bowman's layer. (Adapted from Eagle 1999).

1.3.2.3 Lattice corneal dystrophy (CDL)

Lattice corneal dystrophy OMIM No: 122200/105120 CDLI, CDLII and CDLIII are inherited as an autosomal dominant trait while CDLIII is inherited as autosomal recessive dystrophy (Figure 1.15 and Table 1.1). The common pathological feature is the deposition of amyloid material in the stroma which are often seen as refractile lines. Lattice corneal dystrophy type I (CDLI) is inherited as an autosomal dominant disease with variable expressivity and is usually bilateral and symmetrical, although unilateral and asymmetrical cases have been reported (Seitz et al., 1993). The disease presents in the first or second decade of life with corneal erosions or decrease in visual acuity. Initially fine refractile lines with a lattice network are seen in the central stroma, together with a
number of white dots or flecks. The deposits consist of amyloid, which are of glycoprotein origin and made up of fibrils, giving the lattice appearance.

No systemic deposition of amyloid has been detected in other tissues of the body. Later as the opacities develop, the lines spread to the periphery of the stroma and the area between the lattice lines develop a haze that reduces visual acuity. In some patients, as a result of corneal erosions and the increasing loss of vision, penetrating keratoplasty is necessary (Malbran et al., 1988; Rodrigues and Krachmer 1988).

![Figure 1.15](image)

**Figure 1.15 Lattice corneal dystrophy with refractile lines and amyloid deposits.** Slit lamp photographs (A) CDL type I, (B) CDL type III and (C) CDL type IIIA. Refractile lines are shown by arrows. (D) Light micrograph using a congo red stain to highlight the amyloid deposits (arrow) seen in the stroma in lattice dystrophies. (Adapted from Eagle 1999).

Lattice corneal dystrophy type II (CDLII) is also called Meretoja syndrome or familial amyloid polyneuropathy type IV (Meretoja 1972). The disease is associated with systemic amyloidosis with deposits reported in the brain, kidney, heart and the cornea. The disease is inherited in an autosomal dominant fashion, although one case of homozygosity has been reported (Meretoja 1973; Haltia et al., 1992; Asaoka et al., 1993; Kivela et al., 1994). The changes in the cornea are often noted in the third decade and the lattice lines in the cornea tend to be more radially oriented and less dense than CDLI. Visually acuity is less severely impaired as there are fewer flecks between the lines and less of a haze. The major constituent of the amyloid laid down in this disease is a gelsolin protein (Meretoja et al., 1978; Loeffler et al., 1992). The gene for gelsolin maps to chromosome 9q34 and it has been shown to be important in the severing of actin
filaments. A G187A mutation in the gelsolin gene has been identified in three large families with CDLII and has been shown to be responsible for causing the disease (Haltia et al., 1992). Subsequently, mutations were also identified in Japanese families with this disease (Akiya et al., 1996).

Lattice corneal dystrophy type III (CDLIII) (Figure 1.15) is inherited as an autosomal recessive trait and has a late adult onset (70-90 years). It was recently reported in two Japanese families with no evidence of systemic amyloidosis (Hida et al., 1987). The disorder occurs unilaterally or bilaterally in the form of thick, ropy lattice lines extending from limbus to limbus. Visual acuity is not greatly affected until the sixth decade and corneal erosions are rare.

Lattice Corneal Dystrophy type IIIA (CDLIIIA) differs from CDL type III as the mode of inheritance is autosomal dominant rather than recessive and is mainly been reported in Caucasian patients (Stock et al., 1991). The lattice lines are thick and ropy and reduce visual acuity. The corneal erosions represent a major component of the disease.

1.3.2.4 Avellino corneal dystrophy

Avellino corneal dystrophy OMIM No: 121900/122200 was the name used to describe an Italian family from Avellino that showed the clinical and histological characteristics of both granular and lattice corneal dystrophy type I (Folberg et al., 1988).

Figure 1.16 Slit lamp photograph of Avellino corneal dystrophy. Arrows indicate stromal deposits with granular and lattice characteristics. (Adapted from Eagle 1999).

The disease is inherited as an autosomal dominant trait with central anterior stromal opacities and a deeper lattice network of deposits (Figure 1.6). The granular stromal
opacities are seen first with the lattice lines developing later together with a corneal haze (Figure 1.16). Recurrent corneal erosions can appear in some patients. Histology shows that hyaline deposits as well as amyloid deposits are present and suggest that CDGGI, CDLI and ACD may be part of a spectrum of the same disease (Folberg et al., 1988).

### 1.3.2.5 Fleck dystrophy

Fleck dystrophy OMIM No: 121850 was described by Fancois who described a family with 31 affected patients in 3 generations that had small dots or flecks of varying size and opacity throughout the stroma (Figure 1.17), which was inherited in an autosomal dominant fashion (Francois et al., 1957).

![Figure 1.17 Slit lamp photograph of Fleck dystrophy with mid-stromal deposits. Arrow shows speckled of fleck appearance within the stromal layer. (Adapted from Eagle 1999).](image)

The flecks are seen in the first few years of life and may be congenital, however it is not a progressive dystrophy and the same changes are seen in patients in their seventies (Toselli et al., 1966). Vision is unaffected and no corneal sensitivity is reported. It has been noted that central cloudy dystrophy occurred in a family with Fleck dystrophy, demonstrating that both dystrophies may be part of a spectrum of the same disease. Central cloudy dystrophy has more numerous and dense opacities, and like Fleck dystrophy, vision is unimpaired (Goldberg et al., 1977). There are also reports of Turkish and British pedigrees with Fleck dystrophy (Akova et al., 1994; Assi et al., 1999).
1.3.2.6 Cornea plana

Cornea plana is a rare congenital anomaly in which the cornea is flattened with a corneal curvature of less than 43 diopters, the radius of curvature may reach levels as low as 20 or 30 diopters. It is sometimes difficult to differentiate clinically, from peripheral sclerocornea (Forsius et al., 1998).

Cornea plana has two modes of inheritance: a severe autosomal recessive form (CNA2; OMIM, 217300) and a milder autosomal dominant form (CNA1; OMIM, 121400). Both types share many clinical features including reduction of the corneal power, broadening of the limbus, frequent hyperopia and arcus lipoides at an early age (Tahvanainen et al., 1996). For the autosomal dominant type, visual acuity may be normal, the corneal parenchyma is clear and the diopteric power of the cornea is nearly 3-7 diopters less than the normal value (normal diopteric power of the cornea is 43 diopters). On the other hand, in the recessive form shows a markedly reduced refractive power to about 20-30 diopters, widened limbal zone, smaller corneal diameter, strong hyperopia, shallow anterior chamber and rounded opaque central thickening nearly 5 mm in diameter are observed (Forsius et al., 1998). Additional anomalies are often present in the recessive form, such as malformation of the iris, a slit-like pupil and adhesion between the iris and the cornea. However, the most distinctive feature of the recessive form is the presence of a central corneal opacity (Tahvanainen et al., 1995)

1.3.3 Endothelial dystrophies

The pump and barrier function of the endothelium plays a crucial role in the maintenance of a clear cornea. Any alterations in function through aging, trauma or other insults to the endothelium can cause the loss of clarity and loss of visual acuity. There are only three recognised endothelial dystrophies, Fuch’s endothelial dystrophy, congenital hereditary endothelial dystrophy and posterior polymorphous dystrophy.

1.3.3.1 Fuch’s endothelial dystrophy (FECD)

Fuch’s described a condition in 1910 in which patients had corneal epithelial oedema, stromal clouding and decreased sensitivity, later shown to result from endothelial
dysfunction. Fuch’s dystrophy OMIM No: 136800 usually begins in the fifth to sixth decade and is often referred to as late onset hereditary endothelial dystrophy, with females reportedly more severely affected and two and a half times more frequently affected than men.

![Image](image_url)

**Figure 1.18 Slit lamp and light micrograph of a patient with Fuchs’ dystrophy.** Arrows indicate an opaque cornea (A) and dystrophic endothelium and thickened Descemet’s membrane (B) (Adapted from Eagle 1999).

It is also the most common endothelial dystrophy, inherited in an autosomal dominant fashion (Magovern *et al.*, 1979). The condition is normally bilateral and is characterised by guttatae, which are excrescences on Descemet’s membrane that give the endothelium a beaten metal appearance. These guttatae are produced by abnormal endothelial cells and are also seen in aging and inflammation (Figure 1.18). Confluent and bilateral guttatae are a characteristic of the disease especially if they occur centrally and become pigmented. Histology has shown that the guttatae are composed of clumps of collagenous tissue on Descemet’s membrane (Brooks *et al.*, 1991). The foetal anterior band is normal, the posterior non-banded zone is thinned and a posterior collagenous band may be seen as well as an attenuated endothelium. The barrier and pump functions of the endothelium are compromised and result in oedema that causes the changes first seen by Fuch’s in the stroma and epithelium (Burns *et al.*, 1981).
1.3.3.2 Congenital hereditary endothelial dystrophy (CHED)

Congenital hereditary corneal dystrophy can be inherited as an autosomal dominant disease and as an autosomal recessive disease and was first described by Maumenee in 1960 (Maumenee 1960). The autosomal dominant disease is designated CHED1 OMIM No: 121700 while the more severe, recessive condition is designated CHED2 OMIM No: 217700. Clinically it is difficult to differentiate between them. Disease is characterised by diffuse oedema and varies in severity from a mild haze to an opaque cornea (Figure 1.19).

![Slit lamp photograph of a CHED patient with corneal haziness.](Adapted from Eagle 1999).

Generally for CHED 2 severe corneal clouding is present at birth, or shortly after, with little progression over time (Kirkness et al., 1987). CHED 1 is characterised by corneal clouding which can be present at birth or evident within the first two years and is slowly progressive. The changes seen in the cornea are very similar to Fuch’s dystrophy, the major difference being age of onset. CHED is a rare condition with more recessive families reported than dominant.

1.3.3.3 Posterior polymorphous corneal dystrophy (PPCD)

Posterior polymorphous corneal dystrophy PPCD OMIM No: 122000 is an autosomal dominant disease with variable expressivity that was first described by Koepppe in 1916. It is a bilateral disorder of the corneal endothelium characterised by thickening of Descemet’s membrane, the presence of vesicles, gray plaques or band shaped lesions and the multi-layering of the endothelium (Figure 1.20). Occasionally corneal oedema is present as well as changes in the iris and anterior chamber angle. PPCD can be associated
with ocular disorders such as glaucoma, keratoconus and Alport syndrome (Bechara et al., 1991; Colville et al., 1997). Four different cell types are present on the posterior cornea, normal endothelial cells, degenerated or attenuated cells, fibroblast-like cells and the epithelial-like cells usually associated with PPCD (de Felice et al., 1985).

Figure 1.20 Cornea of patients with PPCD.
(A) Slit lamp photograph with arrow showing the corneal opacity. (B) Light micrograph- arrow highlights multilayering of the endothelium. (C) Immunohistochemistry with pan-cytokeratin. Arrows show staining of the epithelium and the epithelialisation of the endothelium. (Adapted from Eagle 1999).

The alterations seen in Descemet's membrane, the abnormal collagen deposits and guttatae are all secondary to the changes that occur to the endothelium. Occasionally penetrating keratoplasty is needed if the oedema and endothelial changes result in an opaque cornea.

1.4 Strategies to identify genes implicated in human disease

A number of different methods can be employed to identify genes implicated in human disease, ultimately leading to the screening of a candidate gene for mutations. Two major strategies are functional cloning and positional cloning. Functional cloning requires prior knowledge of the protein product, function of the gene, or biochemical understanding of the disease. However for the majority of inherited human disease the underlying biochemical defect is unknown. A positional cloning strategy assumes no functional knowledge and is dependent on the chromosomal location of the disease locus.
determined by genetic linkage analysis. A combination of the two approaches is often used and is referred to as a positional candidate gene approach. This involves the initial localisation of the disease to a chromosomal region and the subsequent selection of candidate genes based on criteria such as gene expression pattern or homology to genes or proteins previously implicated in similar disease processes or pathways.

The availability of the sequence of the human genome has greatly facilitated the elucidation of the genetic basis of human disease.

1.5 Genetic mapping

Genetic mapping follows the segregation of alleles at a locus for a genetic marker during meiosis.

1.5.1 Meiotic recombination

Human genomic DNA is distributed between 22 pairs of autosomes and two sex chromosomes. During meiotic cell division homologous chromosome pairs, each consisting of two chromatids, are aligned together and DNA segments are exchanged between chromatids of homologous chromosomes. This process is referred to as crossing over and results in recombinant chromosomes. As a result of these events, a unique chromosome is formed and inherited by the next generation. Loci which lie physically further apart are more likely to recombine while loci which lie closer together on the chromosome are less likely to recombine and can therefore be considered linked to one another.

The frequency of recombination can be calculated by comparing the inherited alleles of offspring to those of their parents. Linkage analysis relies on the frequency of crossovers to infer the distance between genetic markers in a family pedigree. Using linkage analysis, genetic markers can be placed onto genetic maps of chromosomes and the segregation of inherited diseases in a family can be followed. Linkage analysis in family pedigrees expressing a disease phenotype can be used to define a chromosomal locus, leading to the isolation of the causative gene.
1.5.2 Recombination frequency

The number of recombinants expressed as a fraction of the total number of gametes is called the recombination fraction ($\theta$) and it is a measure of the genetic distance between any two loci. The further apart two loci are, the greater the possibility of crossovers being observed and the recombination fraction will then approach 0.5 which indicates independent segregation. Conversely, if two loci lie close together on a chromosome and no crossovers are observed then recombination fraction will be zero. If the recombination fraction between a disease and a genetic marker at a specific locus is less than 0.5 it suggests a deviation from Mendel's law of independent segregation, which is suggestive of linkage to that locus.

1.5.3 Genetic map distance

The genetic map distance (in units of Morgans) is defined as the length of chromosomal segment, which on average undergoes one crossover per chromatid strand. The male autosomal map length is estimated to be 26.5 Morgans based on an average occurrence of 53 chiasmata. The recombination rate in females is higher than in males and is estimated to be 39 Morgans. Thus the sex-averaged autosomal map is 33 Morgans which implies that the average length of a human chromosome is 1.5 Morgans or that it undergoes 1.5 crossover per meiosis (Renwick 1969). However, genetic distance is usually quoted in centimorgans (cM), where two loci are 1 cM apart if they recombine once in 100 meiosis or show a recombination fraction ($\theta$) of 0.01. There is a linear correlation between $\theta$ and genetic map distance over short distances however, over longer distances multiple crossovers can occur between two loci and the values of $\theta$ are not additive. Several mapping functions have been used to provide a mathematical relationship between $\theta$ and genetic map distance over long distances as $\theta$ can never exceed 0.5. Haldane's function assumes that crossovers randomly occur across a chromosome pair and that they have no influence on each other. However, Kosambi's function is widely used as it allows for the effects of interference (Ott 1997). Positive interference describes the effect that a crossover has on reducing the probability of a second crossover occurring in its vicinity.
1.5.4 Polymorphic markers used in linkage analysis

Genetic mapping requires markers, that are polymorphic, display mendelian segregation and are distributed at regular interval throughout the genome in order to provide a useful tool for the chromosomal localisation of a disease.

1.5.5 RFLPs and mini-satellite DNA

In order to map genes using linkage analysis it is necessary to have genetic markers to which these genes can be linked. Restriction fragment length polymorphisms (RFLPs) exploit the fact the human DNA can be variable and that these variations or polymorphisms may result in the alteration of a restriction enzyme site. RFLPs were typed by the hybridisation of Southern blots of restriction digested genomic DNA with a probe tagged with radioisotope and subsequent exposure to autoradiograph film. For the first time, DNA polymorphisms were used to provide a marker set spanning the entire genome (Botstein et al., 1980). Most RFLP systems have only two alleles with a maximum heterozygosity of 50%. A major drawback of the use of RFLPs is that it is time consuming and the informativeness of markers was low. The polymorphism information content (PIC) value was used to calculate the informativeness of a marker in a given population and is dependent on the number of alleles in that population.

Minisatellite or variable number tandem repeats (VNTRs) are tandemly repeated units of 11-60 bp that can extend up to 1 kb. Polymorphism occurs as a result of the difference in the number of repeats. As such, VNTRs are more informative than RFLPs and have higher PIC values hence they become more useful for linkage studies. The major limitation of minisatellites is the tendency for clustering around the telomeres and the time consuming method of analysis (Jeffreys et al., 1985).

1.5.6 Microsatellite markers (Short Tandem Repeat Polymorphisms)

Microsatellites are simple tandem repeat sequences that can be found at the 5’ and 3’ untranslated region of genes, as well as within introns and non-coding DNA (Weber et al., 1989)

The use of microsatellites overcame many of the limitations of minisatellite DNA as they are distributed more evenly throughout the genome at approximately 30kb intervals.
In addition, the ease of analysis with amplification by PCR, is more efficient (Litt et al., 1989). Microsatellites consist of tetra, tri and di-nucleotide repeats with the most common being the dinucleotide CA repeat. A measure of the informativeness of a microsatellite marker can be obtained from its polymorphic information content (PIC) value (Weber and May 1989).

A number of genetic maps composed of microsatellite markers have been published, and the final Genethon map in 1996 brought to an end this phase of the human genome project (Dib et al., 1996)

1.5.7 Single nucleotide polymorphisms (SNPs)

Single nucleotide polymorphism (SNP) is a biallelic system based on a single nucleotide change at the sequence level.

SNPs are the most frequently occurring polymorphic markers available estimated at 1 SNP per 100 to 300 bp (http://www.ncbi.nlm.nih.gov/SNP). A genetic map of SNPs was created using a number of microsatellite markers as anchor points and the automation of the analysis was achieved using a genotyping chip that would allow the simultaneous analysis of 500 SNPs (Wang et al., 1998). The SNP database at The National Center for Biotechnology Information serves as a central repository for both single base nucleotide substitutions and short deletion and insertion polymorphisms. It is estimated that 60,000 SNPs fall within exons and that 85% of exons are within 5 kb of the nearest SNP (Sachidanandam et al., 2001). As genetic markers, SNPs can be used for conventional linkage analysis. However, due to the availability of high density maps and the improvements in technology that allow for the automation of screening SNPs they are a useful resource for linkage disequilibrium (LD) and association studies.

1.5.8 Degree of marker polymorphism

For linkage analysis it is necessary to have informative meiosis, however to achieve this it is necessary for a marker to be polymorphic. The greater the number of alleles the greater the probability of heterozygosity. Increasing the information content of markers can reduce the number of individuals that need to be typed to genetically map disease genes.
The probability that a random individual is heterozygous is used as a measure of the degree of polymorphism and can be estimated based on a random sample of unrelated individuals.

The degree of observed heterozygosity is the proportion of heterozygous individuals observed in the sample. This can be calculated from the Hardy-Weinberg equation (HWE), where the observed allele frequencies for a given marker must add up to 1. The observed genotypes in a population reflect the allele frequencies such that the chance that an individual is $A_1A_1$ is $p^2$, the chance that they are $A_1A_2$ is $2pq$ and the chance that they are $A_2A_2$ is $q^2$.

$$p^2 + 2pq + q^2 = 1$$

For example, for a dimorphic marker with allele frequencies of 0.7($p$) and 0.3($q$) we expect to observe the following number of individuals with the given genotypes: homozygous $p$ and $q$ and heterozygous $pq$.

$$0.7^2 + 2(0.7x0.3) + 0.3^2 = 1$$

or

$$0.49 + 2(0.21) + 0.09 = 1$$

or

$$49\% + 42\% + 9\% = 100\%$$

### 1.6 Linkage analysis

Linkage analysis involves the use of polymorphic genetic markers to determine the segregation of alleles within families with a mendelian trait and to determine whether the markers are linked to the disease.

### 1.6.1 Maximum likelihood estimate (MLE)

The principle of maximum likelihood states that the hypothesis with the greatest likelihood is that for which the probability of the observations is maximised. This maximum is obtained by finding that value of the recombination frequency between the disease and the marker that maximises the probability of the data. This value is called the maximum likelihood estimate.

$L(\theta)$ is the likelihood of an observation of recombination at a given recombination fraction ($\theta$). $L(\theta)$ can be calculated for a range of recombination fractions between 0 and
0.5 and the value of $\theta$ that produces the highest $L(\theta)$ is the MLE. This is often denoted as odds ratios, $L(\theta)/L(0.5)$ which expresses the probability of linkage versus non-linkage.

### 1.6.2 Lod score (Z)

The lod score method is based on the likelihood ($L$) of an observation of recombination at a given recombination fraction ($\theta$).

Lod score ($Z$) is the decimal logarithm of the odds ratio which, for a specific recombination fraction can be calculated as:

$$Z(\theta) = \log_{10} \left\{ \frac{L(\theta)}{L(0.5)} \right\}$$

where:

- $L(\theta) = \text{the likelihood of obtaining the data if the two loci are linked with a recombination fraction } \theta$
- $L(0.5) = \text{the likelihood of obtaining the data if the two loci are not linked}$

Lod scores are normally calculated over a number of values of $\theta$. The recombination fraction that gives the maximum lod score ($Z$) is the best estimate of the degree of linkage between the two loci. Positive lod scores are suggestive of linkage with a lod score of 3 or greater considered definitive evidence for linkage. This roughly equates to 1000 to 1 odds in favour of linkage. Conversely a negative lod score of −2 at a given value of $\theta$ is considered as evidence against linkage within an interval equal to $\theta$ from either side of the locus.

### 1.6.3 Multipoint linkage analysis

Multipoint analysis allows for the simultaneous analysis of three or more genetic loci. Detailed mapping of a disease locus involves the ordering of new loci with respect to other loci in the region and its placement on a map containing markers of known order and genetic distance. For a given locus order, lod scores are calculated at different recombination fractions. This procedure is repeated sequentially for different marker orders and the largest likelihood observed should correspond to the most likely order of marker loci to the disease. Multipoint analysis can be used as an important tool when
markers within the disease interval are partially informative, as it provides a cumulative result based on the assessment of all markers at the same time.

1.7 Physical mapping

One of the goals of the human genome project was to create a series of human gene maps of each chromosome sequence at increasingly finer resolutions, with the ultimate aim of obtaining the entire sequence of each chromosome. Towards this end it was necessary to fragment the chromosomes into yeast or bacterial clones that can be propagated and manipulated. Once a series of overlapping ordered fragments of DNA (a contig) is obtained, the nucleotide sequence of the clones can be determined. The physical map charts the chromosomal localisation of genes and other sequence tagged sites (STSs).

1.7.1 Low resolution mapping

The lowest resolution physical map is the cytogenic map based on banding patterns of chromosomes stained with particular dyes. These maps can place DNA markers within 2-5 Mb. However, if fluorescent in situ hybridisation (FISH) is performed on interphase spreads the resolution can be increased to 0.1 Mb (Trask 1991; Trask et al., 1993). FISH can be used to order genomic clones or genes to particular loci on chromosomes.

Somatic cell hybrids are a relatively quick and efficient way of assigning a segment of DNA, a gene, EST or STS, to one particular chromosome or chromosomal region. They are created by the fusion of rodent and human cells, with the former acting as a host cell line. The human donor cell often confers a selective advantage, such as drug resistance, to the hybrid cell in order to be retained during chromosomal segregation. A number of polychromosomal somatic cell hybrid panels exist as well as a monochromosomal panel (Rao et al., 1992; Kelsell et al., 1997). The segment of DNA to be amplified by PCR can be a marker, STS or gene that is then assigned to a particular chromosome or chromosomal region.
1.7.2 High resolution mapping

To obtain a high resolution physical map of the genome one of the methods used to represent the order of the DNA fragments is to create a YAC, BAC or PAC contig.

YACs are Yeast Artificial Chromosomes that have three types of DNA sequence elements, telomers, centromeres and autonomous replicating sequences. These sequences are essential for yeast chromosome function and can be combined together with large fragments of human genomic sequence to create stable vectors. The YACs can vary in size from 150 kb to over 1 Mb in length with their mitotic segregation behaving like those of a natural yeast chromosome (Burke 1991). One of the major drawbacks of using YACs to construct a physical map of a chromosomal region is the instability of the clones as well as the high percentage of chimaerism that exists (Monaco et al., 1994). This can make identification of overlapping clones by chromosome walking very difficult. To overcome this alternative bacterial host cloning vectors like BACs and PACs were created.

P1 derived artificial chromosomes (PACs) are based on the P1 bacteriophage cloning system, which carries a gene for kanamycin resistance. PACs have a P1 plasmid replicon that maintains the vector at one copy per cell and an average insert size of 130-150 kb (Ioannou et al., 1994). PACs have a 5% level of chimaerism compared to the 40-60% estimated for YACs. As a result of the easy amplification of DNA by growing Escherichia coli cultures these vectors are ideal for constructing contigs and as sequencing templates (Ioannou et al., 1994)

Bacterial artificial chromosomes (BACs) are a low copy number vector, based on E. coli and its single-copy plasmid F factor, which can contain DNA inserts of up to 300 kb. BACs are relatively stable with a low level of chimaeric clones. A number of genome centres involved in large scale sequencing projects prefer these clones to PACs because of the higher insert to clone ratio of sequence obtained (Shizuya et al., 1992)

1.7.3 Radiation hybrid (RH) mapping

Radiation hybrid mapping panels were developed by fusing irradiated human donor cells with recipient rodent cells. The resolution of these panels depends on the fragment size, which is dependent on the radiation used to fragment the human DNA and the
retention frequency. The Genebridge 4 (GB4) panel was created by irradiating a human fibroblast cell line with 3000 rads that resulted in 168 cell lines. 93 of these hybrids showing the highest retention of human chromosomal segments and are available as a mapping tool (Gyapay et al., 1996).

Radiation hybrid mapping has provided an independent method of assigning STSs, DNA markers and ESTs to particular chromosomes. In addition to positioning markers, it has had major impact on the sequencing of the human genome as the PACs and BACs are now ordered according to their RH mapping data (Hudson et al., 1995; Schuler et al., 1996; McPherson et al., 1997; Deloukas et al., 1998; Weissenbach 1998).

1.8 Mutation detection

Once a disease gene has been linked to a particular locus on a chromosome the next step is the identification of the gene mutations responsible. A number of different mutation detection methods can be used to detect gene mutations.

1.8.1 Heteroduplex analysis

Heteroduplex analysis is a technique that can be routinely used to detect variants in a given sequence. It is based on the fact that two complementary DNA strands, which differ in sequence by even one base pair, will have mismatched positions when base paired. These double stranded heteroduplex molecules tend to show altered mobility in non-denaturing polyacrylamide gels when compared to the homoduplexes of either allele (Keen et al., 1991). Recently denaturing high-pressure liquid chromatography (DHPLC) technology has been applied to identify mutations such as the Transgenomics Wave machine. DHPLC uses triethylammonium acetate (TEAA), which acts as an ion-pairing reagent that binds to the hydrophobic DNA giving it hydrophilic characteristics. For the Wave machine (Transgenomics), the TEAA (0.1 M) coats the DNA, which allows it to bind to the divinylbenzine beads on a column. There is a proportional relationship between the association of TEAA and the binding to the DNA separation column. Dissociation of the TEAA-bound DNA from the column is accomplished by an elution gradient comprising 0.1 M TEAA and 25% acetonitrile. The proportions of reagents are changed over time and allow the elution of DNA. Smaller DNA fragments, having less TEAA bound, are eluted before the larger fragments and the technique allows for a sizing
application as well as its use for mutation detection. The DNA fragments eluted are detected using a UV detector and visualised as a graph.

1.8.2 Single strand conformation polymorphism (SSCP)

The sequence of interest is amplified by PCR and separated as single-stranded molecules by resolution on native polyacrylamide gels. Sequence variants tend to show a shift in mobility due to changes in the tertiary structure of the single-stranded DNA (Orita et al., 1989)

No method of mutation detection is 100% efficient and if changes are seen in either method they require confirmation by direct sequencing of the DNA product. Direct sequencing of the gene sequence is the ultimate and most accurate mutation technique available.

1.9 Human genome project

The Human Genome Project was instigated as an international research initiative to produce detailed genetic and physical maps of each of the 24 human chromosomes. The ultimate goal of the initiative, following the construction of physical maps in large insert clones, is to produce a single continuous sequence for each chromosome and define the positions of all genes. Similar analyses of several other genomes and model organisms were also initiated, and the methods and resources used to determine the significantly smaller genome sequence of the nematode *Caenorhabditis elegans* (97 million base pairs) greatly aided the planning and execution of the Human Genome Project (approx. 3 billion base pairs). Throughout the progression of the Human Genome Project (HGP) new methods and resources were developed and complimentary approaches used to create, store and analyse the data (Jordan 1994; Haldi et al., 1996; Lander 1996; Schuler et al., 1996).

Major achievements of the HGP include;
1995 – first generation physical map
1996 – comprehensive genetic map
1998 – *C. elegans* genome completed
1999 – sequence of chromosome 22 completed
2001 – complete draft sequence of the human genome
The draft human genome sequence was published in 2001 by the HGP consortium and Celera Genomics (International human genome sequencing consortium 2001; Venter et al., 2001). The HGP project estimates that there are approximately 31,000 genes and Celera estimates an even smaller number of genes, about 26,000. This is in stark contrast to initial estimates of human gene number which ranged from 50,000 to 100,000, and indicates that the density of genes in the human genome is much lower than for any other genome sequenced so far. Both groups used computational algorithms to model and predict gene sequences, however these methods are known to be inaccurate through over and under prediction. To arrive at a more accurate description of the number of genes in the human genome we will continually rely upon individual gene and protein characterisation. Initiatives to characterise full length mRNAs have already highlighted the fact that the current human genome annotation has not detected a significant number of gene sequences (Wiemann et al., 2001). It is also evident that alternative splicing of genes in the human genome is common (over 60% of genes have alternative transcripts) which would result in a larger number of protein products.

Much work remains to be done to complete the sequence and be assured of its accuracy, but the vast amount of information that has become available through this initiative provides us with basic information on genome evolution and structure, and revolutionises the field of medical genetics.

1.10 Bioinformatics and human genome resources

Over the past few decades, major advances in the field of molecular biology, coupled with advances in genomic technologies, have led to an explosive growth in the biological information generated by the scientific community. This deluge of genomic information has, in turn, led to an absolute requirement for computerised databases to store, organize and index the data, and for specialised tools to view and analyse the data. Bioinformatics is the term used to describe the discipline which brings together biology and computer science. The ultimate goal of the field is to enable discovery of new biological insights. Currently the focus on the human genome project has enabled the analysis and interpretation of various types of data including nucleotide and amino acid sequences, the
development of tools which enable efficient access to different types of information, and the development of new algorithms to assess and analyse relationships between data sets.

Alongside the development of new methods for establishing maps of the genome and new sequence strategies a great deal of effort is concentrated on designing databases and programs to accurately represent and analyse the information generated. One major challenge was to collect, store, distribute, analyse and retrieve data created as whole genomes were physically mapped and sequenced. A *C. elegans* database (ACeDB) was developed by Durbin and Thierry-Mieg in order to manage the data from the nematode sequencing project and integrate the data with other databases using both text and graphics. Sequence annotation is accomplished in ACeDB incorporating search and analysis tools, which look at intrinsic sequence properties to predict, for example, exons and protein structure. ACeDB incorporates GRAIL, Fex, Hexon, GeneFinder and Fgenesh for gene prediction (Eeckman *et al.*, 1995). These original databases and programs are constantly evolving and many suites of programs and databases now exist on web sites world wide providing a remarkable resource for the field of molecular genetics.

### 1.10.1 Map integration

A major goal for the human genome project, facilitated by the Whitehead Institute/MIT, was the construction of an STS content map of the human genome consisting of 10,000 STSs screened in the CEPH YAC library (Hudson *et al.*, 1995). When this goal was achieved in 1995, efforts were then directed towards the construction, integration and validation of maps. A radiation hybrid map of 14,665 markers, in addition to the 5,264 marker Genethon map, was constructed covering an estimated 94% of the genome. Genome maps have subsequently integrated PAC and BAC clone data and the ultimate map, the genome sequence, is being deposited in these databases in accessible formats for analysis. More recently emphasis has shifted towards the establishment of a SNP map covering the genome. Regional and chromosome maps can be viewed at many centres, each with their own graphic interface and options to view and analyse.
1.10.2 Expressed sequences

The human mapping project also involved mapping expressed sequence tag (EST) sequences with the Genbridge 4 radiation hybrid panel and improving map quality. The EST maps and cDNA sequence resources are invaluable for disease gene identification projects. The main centres include the Institute for Genomic Research (TIGR) and the expressed gene anatomy database (EGAD), the National Center for Biotechnology Information (NCBI) with access to Unigene, and others, such that a non-redundant repository of genes is created and their expression profiles and genomic map position are instantly obtainable.

1.10.3 Sequence analysis tools

The Basic Alignment Search Tool (BLAST) is the main program used to compare sequence similarity, with algorithms for nucleotide and amino acid sequence analysis (Altschul et al. 1990, Altschul et al. 1997). By comparing sequence identity and similarity of a gene or protein with all known sequences (GenBank), function can often be inferred. Gene and protein sequences can also be analysed using suites of programs such as NIX and PIX available at HGMP, which combine many analysis tools in one package. In addition, NIX is a useful tool for analysing genomic sequence which runs gene prediction programmes similar to ACeDB as well as data mining and comparison programs. Many analysis programs have been developed for specific applications, however it is important to note that all bioinformatic tools merely provide a guide for future experimental work.

Problems can arise when utilising bioinformatic tools for data analysis. One major draw back is that they all rely on the accuracy of the archived data, as well as our current understanding of gene families, protein function, exon structure etc. As we learn more about the genome and specific gene/protein functions the bioinformatic applications will improve.
1.11 Aims of this thesis

The aim of this thesis was to study inherited corneal dystrophies using genetic and molecular biology in order to identify the underlying genetic defects. The goal of the study was to make a contribution to the understanding of the physiology and pathogenesis of diseases affecting the cornea.

Congenital hereditary endothelial dystrophy (CHED 1) is inherited as an autosomal dominant endothelial disease. The largest reported CHED 1 pedigree was mapped to chromosome 20 using linkage analysis and the critical interval was subsequently refined. A physical map of this region was constructed initially using a YAC and PAC contig in order to map sequence tagged sites (STSs) and expressed sequence tagged sites (ESTs) to the region. Candidate genes for CHED were excluded by sequence analysis (Chapter 3).

The locus for CHED 2 spans approximately 3 Mb of DNA on chromosome 20p13. In collaboration with the Chromosome 20 group at the Sanger Institute, a contig was created across the disease interval and subsequently sequenced. A positional candidate gene approach was undertaken to determine the gene responsible for CHED2. To facilitate gene identification a small Pakistani family, which had previously been linked to the locus on 20p13, was used to analyse candidate genes for mutations (Chapter 4).

Granular Corneal Dystrophy (GCD 1) is an autosomal dominant stromal disease. A GCD pedigree was linked to chromosome 5q. A physical map of the interval between markers D5S393 and D5S396 was constructed using YACs from the CEPH YAC library and the Whitehead database and ESTs were mapped into the disease interval. The keratoepithelin gene was screened and a mutation identified in this family (Chapter 5).

Cornea Plana can be inherited as a mild autosomal dominant disease (CNA1) or a more severe, autosomal recessive form (CNA2). A small CNA2 pedigree was ascertained and screened for mutations within the keratocan gene, in which mutations causing CNA2 had previously been identified (Chapter 6).
CHAPTER 2

Materials & Methods

2.1 Materials

2.1.1 General reagents

All reagents were purchased from Sigma, Promega, GibcoBRL, Invitrogen, Boehringer Mannheim or Amersham Pharmacia Biotech unless otherwise stated; Oxoid supplied the broth stocks.

Tris-EDTA (TE) buffer: 10 mM Tris-HCl (pH 7.5), 1 mM EDTA (pH 7.5-8.0), autoclave before use.

Phosphate buffered saline (PBS): 10 mM phosphate, 0.9% (w/v) NaCl pH 7.4.

2.1.2 Reagents for DNA isolation

Human genomic DNA using the Nucleon II kit

Reagent A (5x lysis buffer): 320 mM sucrose, 10 mM Tris HCl (pH 7.5), 5 mM MgCl$_2$ (adjusted to pH 8.0 with 5 M NaOH), autoclave the add 1% (v/v) Triton X-100.

Reagent B: 400 mM Tris-HCl (pH 7.5), 60 mM EDTA, 150 mM NaCl (adjusted to pH 8.0 with 5 mM NaOH), autoclaved then SDS added to a final concentration of 1% (v/v).

YAC DNA

Solution I: 1 M sorbitol, 0.1 M EDTA pH 7.5, autoclaved before use.

Solution II: 50 mM Tris-HCl pH 7.4, 20 mM EDTA, autoclaved before use.

YRB: 1.2 M Sorbitol,10 mM Tris, 20 mM EDTA

Plasmid or PAC DNA

P1 solution: 15 mM Tris-HCl pH 8, 10 mM EDTA pH 8, 100 μg/ml RNase A

P2 solution: 2 M NaOH, 1% (w/v) SDS

P3 solution: 3 M KOAc pH 5.5
2.1.3 Buffers, Solutions and Media

2.1.3.1 Media

**DMEM** (Dulbecco's MEM) with Glutamax-I, with the addition of 5% (v/v) foetal calf serum and 0.1% (v/v) penicillin-streptomycin (GIBCOBRL)

**LB broth**: 1% (w/v) bactotryptone, 0.5% (w/v) yeast extract, 1% (w/v) NaCl, autoclaved before use.

**LB agar**: LB broth containing 1.5% (w/v) agar, autoclaved before use.

**Ampicillin**: 50 mg/ml stock at final concentration 50 μg/ml

**Kanamycin**: 50 mg/ml stock at final concentration 50 μg/ml

**Tetracycline**: 15 mg/ml stock at final concentration 25 μg/ml

**AHC broth**: 6.7 g yeast nitrogen base, 10 g casein hydrolysate, 20 g D-glucose, 20 mg adenine hemisulphate, made up to 1 litre with water, autoclaved before use.

**AHC agar**: AHC broth containing 1.5% (w/v) agar, autoclaved before use.

2.1.3.2 Hybridisation solutions and buffers

**20x SSC**: 3M NaCl, 0.3M Sodium Citrate.

**Denaturing Solution**: 0.5M NaOH, 1.5M NaCl

**Neutralising Solution**: 1.5M NaCl, 1M Tris Base, 1mM EDTA (pH 7.2)

**100x Denhardt’s**: 2% Ficoll (w/v), 2% Polyvinyl pyrrolidine (PVP) (w/v), 2% Bovine Serum Albumin (w/v)

**Church’s Buffer**: First a phosphate buffer was made (1 M NaPi) that consisted of 34 ml of 1 M Na$_2$HPO$_4$ and 16 ml of 1 M NaH$_2$PO$_4$. To this 35 ml of 20% SDS was added together with 200 μl of 0.5 M EDTA. The solution was made up to 100 ml with distilled water.

2.1.4 Gel electrophoresis

**Ethidium bromide** (10 mg/ml)

**10x TAE buffer (pH 8.0)**: 0.4 M Tris-acetate, 10 mM EDTA.
10x TBE buffer: 1 M Trizma base, 0.83 M boric acid, 10 mM EDTA.

10x loading buffers- agarose gels: 6x Type II loading buffer from ABgene: 15%w/v Ficoll 400, 0.06% (w/v) Bromophenol blue, 0.06% (w/v) Xylene cyanol FF, 30 mM EDTA.

Polyacrylamide gels: 95% (v/v) formamide, 0.3% (w/v) bromo-phenol blue, 0.3% (w/v) xylene cyanol, 10 mM EDTA.

Formamide loading buffer: 5:1 (v/v) formamide:50 mM EDTA with 50 mg/ml dextran blue

Acrylamide gel: 0.04 g Ammonium persulphate (APS), 40 ml Sequagel-6 (National Diagnostics), 10 ml Sequagel complete.

2.2 Cell lines and Libraries

An SV40 transformed lymphoblast cell line was established from one affected and one unaffected individual from the CHED1 family.

The endothelial cell line (HCN) was obtained from S.Wilson (Seattle, U.S.A). The stromal cell line, (EK1BR) was kindly donated by R. Farragher (Brighton, U.K.) M.Griffiths (Ottawa, Canada) donated the RNA from epithelial, stromal and endothelial cell lines.

All gridded libraries were obtained from the Human Genome Mapping Project Resource Centre (HGMP-RC), Hinxton, UK (http://www.hgmp.mrc.ac.uk).

2.3 Kits

Kits for the preparation of DNA, gel purification and PCR product purification were supplied by Qiagen. The oligolabelling kit and the cDNA synthesis kits were supplied by Amersham Pharmacia Biotech. All other kits used are cited in appropriate text.

2.4 Oligonucleotide primers

All primers used for polymerase chain reaction (PCR) and for sequencing were supplied by HGMP, Cruachem, Bioline or Sigma-Genosys Ltd.
2.5 Methods

2.5.1 Preparation of Nucleic Acids

2.5.1.1 Extraction of DNA from Blood Samples

10 ml tubes containing sodium EDTA, were used to collect peripheral blood samples that were stored at -20 °C or -40 °C. The DNA was extracted using the Nucleon II kit from Scotlab Biosciences. The blood samples were allowed to thaw at room temperature before being transferred to 50 ml Falcon tubes. 40 ml of Reagent A was added (section 2.1.2) and mixed by inverting several times before being centrifuged at 4000 g for 10 minutes. The supernatant was gently poured into another 50 ml Falcon tube, making sure that the pellet of white blood cells was not discarded. The pellet was resuspended in 2 ml of Reagent B (section 2.1.2). The suspension was transferred to a 5 ml tube to which 500 mM sodium perchlorate was added. The samples were then incubated at room temperature for 15 minutes in a rotary mixer and then at 65 °C for 25 minutes. The samples were cooled on ice for five minutes before 2 ml of chloroform and 300 µl of Nucleon silica suspension were added and rotary mixed for 5 minutes at room temperature. Following centrifugation at 1400 g for 6 minutes, the supernatant was aspirated off into 30 ml sterile universal tubes to which two volumes of absolute ethanol were added. The samples were gently mixed and the DNA was visible as strands that clumped together. The DNA was transferred into eppendorf tubes containing 70% ethanol and was centrifuged at 10,000 rpm in a micro centrifuge for 2 minutes and the supernatant discarded. The pellet was allowed to air dry for 10 minutes before 400 µl of sterile water was used to resuspend the pellet. A 1/10 dilution was made of the stock DNA and an aliquot was stored at -20 °C, 1 µl of the dilution was used in subsequent PCR reactions.

2.5.1.2 Extraction of DNA from PAC/BAC and plasmid clones

This is an alkaline lysis mini-preparation method that is used for isolating DNA from large BAC/PAC clones as well as smaller plasmids and cosmids.

A single colony was used to inoculate 5 ml of LB media (section 2.1.3) supplemented with 25 mg of kanamycin. The culture was grown overnight at 37 °C with shaking at 300
rpm. Glycerol stocks were made by adding 0.5 ml culture and 0.5 ml 40% (v/v) sterile glycerol. 2 x 1.5 ml of the overnight culture were pelleted in a centrifuge at 8000 g for 5 minutes. The supernatant was discarded and the pellet was resuspended in 300 µl of P1 solution (section 2.1.2). After 5 minutes 300 µl of P2 (section 2.1.2) was added, gently mixed and left for 5 minutes at room temperature. As the cell lysis occurs the appearance changes from turbid to translucent. In order to precipitate protein and E.coli DNA 300 µl of P3 solution (section 2.1.2) was added to the tubes and left on ice for 5 minutes. The samples were centrifuged at 8000 g for 10 minutes. The supernatant was carefully transferred to a new eppendorf without disturbing the white precipitate. 0.8 volume of isopropanol was added to the supernatant, gently mixed and left on ice for 5 minutes. The samples were centrifuged for 15 minutes at 8000 g, the supernatant was discarded and the pellet washed in 70% ethanol. The tubes were then centrifuged for 5 minutes at 8000 g, the supernatant carefully aspirated off and the pellet air dried for 10 minutes before resuspending the pellets in 50 µl of TE.

2.5.1.3 Extraction of DNA from yeast artificial chromosome (YAC) clones

A single YAC colony was used to inoculate 5 ml of AHC media (section 2.1.3) and grown overnight at 30 °C. The next day glycerol stocks were made, by adding 0.5 ml culture and 0.5 ml 40% sterile glycerol. The cells were pelleted by centrifugation at 4000g for 10 minutes. The supernatant was discarded and the cells resuspended in 500 µl of 1 M sorbitol and 0.1 M EDTA pH 7.5 and transferred to a 1.5 ml eppendorf. 50 µl of lyticase (1 mg/ml) was added and incubated at 37 °C for 60 minutes. The cells were pelleted in a benchtop centrifuge at 8000 g for 1 minute. The supernatant was discarded and the cells resuspended in 500 µl of 50 mM Tris pH 7.4 and 20 mM EDTA, 50 µl of 10% SDS was added and mixed well before incubating at 65 °C for 30 minutes. 200 µl of 5 M KOAC was added and the eppendorf tubes placed on ice for 5 minutes before centrifugation for 5 minutes at 8000 g. The supernatant was transferred to a fresh eppendorf tube and 1 volume of isopropanol was added and left at room temperature for 5 minutes before centrifugation for 10 seconds at 8000 g. The supernatant was aspirated off and the pellet dried for 10 minutes before resuspension in 300 µl of TE (section
2.1.1). An aliquot of this sample was sufficient for FISH mapping and when diluted 1/20 was suitable for PCR and STS content mapping.

2.5.1.4. RNA extraction

All equipment and work areas were kept thoroughly clean and free of commercial ribonucleases during manipulations with RNA. 0.1% (v/v) Diethyl pyrocarbonate (DEPC) treated water was routinely used for all solutions. DEPC was added to the distilled water and left to incubate overnight in a fume prior to autoclaving. Filter tips and sterile disposable plastic-ware were utilised and gloves were used at all times.

2.5.1.4.1 Extraction of total RNA from cultured cells

The cultured cells used were either freshly harvested or those which had previously been stored at -80 °C. Approximately 2x10^6 cells were used per RNA isolation using the Purescript kit (Genta Systems) which uses a modified salt precipitation procedure together with effective inhibitors of RNase activity. The cells were placed in a 1.5 ml eppendorf tube with culture medium, which was removed by centrifugation at 8000 g for 5 seconds and aspiration of all but 20 μl of the liquid. The tube was vigorously vortexed to resuspend the pellet. 300 μl of the cell lysis solution was added and the solution was pipetted 3 times to lyse the cells. 100 μl of protein DNA precipitation solution was added to the cell lysate and the eppendorf inverted 10 times and then placed on ice for 5 minutes. The sample was centrifuged at 8000 g for 3 minutes, the white pellet of proteins and DNA was left behind while the supernatant was transferred to a fresh eppendorf. 300 μl of isopropanol was added to the supernatant and the tube mixed by gently inverting 5 times. The sample was then centrifuged at 8000 g for 3 minutes, the supernatant was carefully aspirated off and the translucent pellet washed with 300 μl of 70% ethanol. The sample was then centrifuged at 8000 g for 1 minute, 70% ethanol removed and the pellet air-dried for 10 minutes. 50 μl of the RNA hydration solution was added and the pellet was allowed to rehydrate for 30 minutes on ice. RNA samples were stored at -80 °C.
2.5.2. Precipitation of nucleic acids

2.5.2.1 Phenol/Chloroform extraction

To the DNA or RNA an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) was added. The samples were mixed by inversion until an emulsion formed and were then centrifuged at 8000 g for 5 minutes. The aqueous phase was aspirated into a clean eppendorf tube and the process repeated until the sample was free from protein contamination.

2.5.2.2 Ethanol precipitation

The aqueous phase from the phenol chloroform extraction was precipitated by the addition of two volumes of absolute ethanol together with 1/10 volume of 3 M sodium acetate. The samples were placed at -20 °C for 30 minutes to precipitate the DNA. The samples were centrifuged at 8000 g for 15 minutes. The DNA pellet was washed in 70% ethanol air dried and resuspended in sterile distilled water.

2.5.3 Restriction enzyme digestion of DNA

Restriction endonucleases are bacterial enzymes, which cleave double-stranded DNA into discrete pieces, resolvable by gel electrophoresis (section 2.5.4.1). They cleave at, or very close to, specific recognition sequences within the DNA, the length of which (usually 4-8 bp) determines the frequency of cleavage.

In principle, 1 unit of restriction enzyme digests 1 µg of λ DNA in 1 hour at 37 °C. Genomic DNA preparations often require more enzyme and/or more time for a complete digestion. Therefore restriction digests were performed using 2-4 units of enzyme per µg of DNA template, typically in a 50 µl reaction including 5 µl of 10x restriction buffer (Promega, Pharmacia or Biolabs), and made up to volume with distilled water. The volume of restriction enzyme added should be less than 1/10 the final reaction volume, as glycerol in the storage buffer may inhibit digestion. Reactions were incubated for a minimum of 2 hours at the recommended temperature (usually 37°C) and the products of digestion resolved by agarose gel electrophoresis. For digestion involving two enzymes simultaneously, digestion was first performed using the enzyme with the lower restriction
buffer salt concentration. Appropriate quantities of the second enzyme and its specific buffer were then added to complete the double digestion.

2.5.4 Resolving DNA using electrophoresis

2.5.4.1 Agarose gel electrophoresis

Commercially available agarose (Biorad, U.K.) was used to prepare gels used in routine PCR analysis, restriction digests and in estimation of DNA concentration. Agarose was melted in 1 X TAE buffer until dissolved, poured into the desired moulds with the appropriate combs, and allowed to polymerise at room temperature. Once polymerised, the comb was removed and the DNA samples were mixed with loading buffer and loaded in wells. Commercially available markers (Promega) were also used to estimate the size and concentration of DNA. A low percentage (0.8 %) agarose was used to separate relatively large molecules of DNA (PCR products of ~900 bp) and a high percentage gel (2 – 3%) was used to separate small fragments (PCR products of ~150 bp).

Gels were subjected to an electric field and the negatively charged DNA migrated from the cathode to the positive anode in 1 X TAE buffer.

At low voltages, rate of migration of linear DNA is proportional to the voltage applied. Gels were stained with ethidium bromide (0.5 μg/ml) which is a fluorescent intercalating dye and reduces the mobility of DNA by 15 %. The dye intercalates between stacked base pairs of DNA, extending the length of linear and nicked circular DNA and makes them more rigid. Gels were viewed under UV light and photographed.

2.5.4.2 Denaturing polyacrylamide gel electrophoresis.

To resolve small differences in size between DNA fragments, a high resolution denaturing gel was prepared to fractionate single stranded DNA fragments. Samples were run in 6%, 0.4 mm thick denaturing polyacrylamide gels, on 50 cm length Bio-Rad electrophoresis apparatus. Prior to assembly, the back plate (which was attached to the buffer reservoir) was silanised with Sigmacote (Sigma). Acrylamide concentrate (19:1 acrylamide : bisacrylamide) in 8.3 M urea solution (Sequagel, National Diagnostics), was
mixed with diluent (8.3 M urea) and buffer (10 x TBE in 8.3 M urea) to give a 6% gel solution in 8.3 M urea and 1 x TBE (section 2.1.4). A 150 ml solution was made of which a 50 ml acrylamide plug was set at the bottom of the plate, by the addition of 100 µl TEMED (BDH) and 300 µl 25% (w/v) ammonium persulphate. For polymerisation of the remaining 100 ml gel solution, 600 µl of 25% (w/v) ammonium persulphate, and 60 µl TEMED was added, mixed and quickly poured into the apparatus, taking care to prevent the introduction of any air bubbles.

The samples were denatured by heating at 95°C for 3 min, and immediately placed on ice. They were then loaded onto a pre-warmed (50-55°C) 6% gel and electrophorised in 1 x TBE buffer at constant power (50 W for 25 cm wide gels, and 90 W for 40 cm wide gels). After electrophoresis for the required time, the gels were fixed in a 10% methanol-10% acetic acid solution for 5 min, transferred to 3MM Whatmann paper and wrapped in cling film before vacuum drying. The gel was then exposed to autoradiograph film overnight before being developed.

2.5.5 Polymerase Chain Reaction (PCR)

2.5.5.1 Standard parameters for a typical PCR

The polymerase chain reaction has become one of the most valuable techniques in molecular biology by allowing the synthesis of microgram amounts of specific nucleic acid sequences from any part of the genome (Saiki et al. 1988). PCR was performed with Thermophilus aquaticus (Taq) polymerase from Bioline using the manufacturer’s 10x buffer containing 10 mM Tris-HCl, 50 mM KCl, 15 mM MgCl₂ and 0.1% (w/v) non-ionic detergent. Unless otherwise stated, the following constitute a standard PCR reaction which was carried out in a total volume of 25 µl: 1x manufacturer’s buffer, 0.2 mM each dNTP, 0.1-0.2 µM each primer, 150 ng template DNA and 0.5 U Taq polymerase. The use of a master mix of all the reaction components except the DNA ensured consistency in the amplification reactions; thus differences between samples were due to the DNA added. The reactions were overlaid with mineral oil to prevent evaporation, unless performed in equipment with heated lids. Routine cycling conditions were: an initial denaturation step of 95°C for 3 minutes, followed by cycling parameters of denaturing at
95°C for 15 seconds, annealing at the appropriate temperature for 15 seconds, and extending at 72°C for 30 seconds (extension time was increased for products larger than 1 kb). Cycle number varied between 30 and 40, depending upon the template and particular methodology. A final extension step at 72°C for 5 minutes followed. Once optimised, PCRs were generally performed on the same machine, as they tend to vary slightly in temperature and cycle time. Hybaid Omnigene, Techne PHC-3 or Techne Genius PCR thermocyclers were used.

2.5.5.2 Primer design

When designing primer pairs for PCR amplification of known sequences, several rules were followed as far as possible to ensure an optimal result: a similar GC content for both primers, an anchoring C or G at the 3’ end of each primer, minimal secondary structure and low complementarity to each other particularly in the 3’ region to reduce the incidence of ‘primer dimer’ formation, no greater than 4°C difference between the melting temperature (Tm) values of the two primers and a primer length of at least 20 nucleotides (more if possible) to increase the sequence specificity. PCR primers were synthesised by commercial manufacturers.

2.5.5.3 PCR optimisation

The optimal annealing temperature for specific primer pairs was approximated by first calculating the Tm of each primer which is dependant on the nucleotide sequence and was derived using the following formula:

\[ 4 \, (G+C) + 2 \, (A+T) = T_m \]

then assigning an annealing temperature (Ta) 3-5°C lower than the value obtained (e.g. a Tm of 62°C would indicate a Ta of ~58°C). Primer pairs were then tested by PCR on human genomic DNA or equivalent positive control (along with a ‘no DNA’ control) and electrophoresed on agarose gels (section 2.5.4.1) to assess the adequacy of the PCR conditions for subsequent experiments. Additional products (ghost bands) to the authentic PCR product suggested cross-hybridisation of the primers to sequences within the
genomic DNA that bear some degree of homology to the intended target sequence. These extra products could usually be eradicated by increasing the annealing temperature by 1-2°C whereas when no products were detectable on the gel the annealing temperature was decreased by 1-2°C.

Since ionic strength can affect the $T_m$, an alternative solution to obtain a good amplification product is to alter the magnesium concentration. In this instance an alternative 10x reaction buffer (provided) containing 160 mM (NH$_4$)$_2$SO$_4$, 670 mM Tris-HCl (pH 8.8 at 25°C) and 0.1% (w/v) Tween-20 was used. Appropriate amounts of 50 mM MgCl$_2$ were added to the buffer in order to obtain different final concentrations of magnesium (10, 15, 20 and 25 mM). A master mix was then prepared with the primer pair to be tested adding all the reaction components except the buffer, thus differences between samples were due to the different concentrations of magnesium added.

Depending on the sequence specificity of the primer pair used, sometimes “ghost bands” proved particularly difficult to eliminate, a hot start was then performed: PCR reactions including all the components except the Taq polymerase were prepared and put in the thermocycler to denature for 5 minutes prior to adding the enzyme. This was to ensure that the DNA molecules were fully denatured before the polymerisation started to proceed.

With templates particularly rich in GC content, either dimethylsulfoxide (DMSO) (1 μl per reaction) or 5% formamide and 10% (v/v) glycerol were added to PCR reactions to release DNA from possible secondary structures.

2.5.5.4 Microsatellite analysis

2.5.5.4.1 Incorporation labelling of PCR product

Radioactive PCR reactions were performed in a total volume of 10 μl with a reaction mix in which the final dNTP concentration was modified so that dCTP was present at only 1/10 the concentration of the other dNTPs (i.e. 20 μM rather than 200 μM) and with 1 μCi $\alpha^{32}$P-dCTP added per reaction (i.e. 10 μCi of a $\alpha^{32}$P-dCTP was added to a reaction mix for 40 reactions). The template DNA was dispensed into 0.5 ml eppendorfs and overlaid with mineral oil prior to the addition of the reaction mix.
2.5.5.4.2 End labelling of primer

Rather than incorporating the radionucleotide into the PCR mix, which can give rise to an autoradiograph with a high level of background, the following method was also used to label one of the PCR primers. 20 pmol of the oligonucleotide was end-labelled in 10 μl volumes containing 1 x reaction buffer (One-Phor-All buffer), 3.6 Mbq [γ²P]dATP and 5 units of T4 polynucleotide kinase. The reactions were incubated at 37°C for 45 minutes and 2 μl of this was used directly in a PCR reaction with one labelled and one unlabelled primer. 6 μl of formamide loading buffer (section 2.1.4) was added to the reaction. Samples were denatured at 95°C immediately before loading onto acrylamide gels.

The microsatellite allele size differences (multiple of 2-4 bp) were resolved by denaturing polyacrylamide gel electrophoresis (2.5.4.2).

2.5.5.4.3 Fluorescent automated genotyping

Microsatellite markers were synthesised using the fluorescent labels FAM of HEX. PCR conditions were optimised (2.5.5.3) and the PCR products were checked on an agarose gel. The PCR products were diluted in formamide which contained a fluorescently tagged DNA ladder denatured and loaded on an ABI 3100 (Applied Biosystems). The Genotype software was used to determine the alleles that were then used to calculate a lod score.

2.5.5.5 Reverse Transcriptase (RT) PCR

First strand cDNA synthesis was performed using the Ready-to-go You-Prime First-Strand Beads (Amersham Biotech). Two different aliquots (2 and 5 μl) of total RNA extracted from fibroblast cell lines (section 2.5.1.4) were brought to a volume of 25 μl in an RNase-free microcentrifuge tube using DEPC treated water. The mixture was heated for 10 minutes at 65°C, chilled on ice for 2 minutes and transferred to the tube of First-Strand reaction mix beads, after checking that the beads were visible at the bottom of the tube, taking care not to mix the solution at this stage. 0.5 μg of Oligo(dT)₁₂₋₁₈ and DEPC-treated water to a final volume of 33 μl were subsequently added to the mixture. The
mixture was left at room temperature for approximately 1 minute and then gently vortexed and briefly spun to collect the contents at the bottom of the tube prior to incubating at 37°C for 60 minutes. The completed first-strand reaction was stored at 4°C or used immediately for PCR amplification.

2.5.5.6 Inter-ALU PCR

Throughout the genome, there are interspersed repetitive sequences (IRS) of which the short interspersed repeat element (SINE) is an example. ALU repeats are the most commonly occurring SINE and consist of a repeat element of approximately 300 bp every 4 kb (Batzer et al., 1991). As Alu sequences are not represented in the yeast genome, only the human DNA in a YAC clone will be able to act as a template in PCR amplification. Two ALU primers were used:

Ale1 5' GCC TCC CAA AGT GCT GGG ATT AC AG 3'
Ale3 5' CCA T/CTG CAC TCC AGC CTG GG 3'

Inter-ALU PCR was carried out in a 50 μl reaction using standard PCR reaction mix (2.6.5.5.3) with a final concentration of 1 μM ALU primer and 150ng YAC DNA as a template. A combination of primers were used, Ale1 only, Ale3 only and Ale1/Ale3. The PCR conditions used over 40 cycles were:

95 °C for 1 minute
65 °C for 1 minute
72 °C for 2 minutes

10 μl of the PCR products were analysed on an agarose gel (2.5.4.1).

2.5.6 YAC/PAC library screening by PCR

Human YAC clones were screened by PCR using a variety of STSs, ESTs and microsatellite markers from the respective regions to identify corresponding YAC clones, to determine overlaps between adjacent YACs and to detect any chimerism present. YAC clones were received either as ‘stabs’ or streaked out on selective agar, and were restreaked onto fresh AHC+Ampicillin agar plates (section 2.1.3.) to initiate their growth.
as single colonies. DNA was then extracted as described in section 2.5.1.3 and PCR with the relative STS primers was performed as described in section 2.5.6.1.

2.5.6.1 Colony PCR

This method was very useful for preliminary screening prior to DNA extraction from YAC clones. The YAC colony templates were treated to break down the yeast cell wall prior to PCR using a protocol reported by Ling et al. (1995). A small amount of each colony was picked with a sterile tip into 10 µl of incubation solution (1.2 M sorbitol, 100 mM sodium phosphate pH 7.4 and 2.5 mg/ml lyticase). The resulting enzyme/cell mixture was incubated at 37°C for 5 minutes. 3 µl of the mixture was then used as a template for PCR. The method proved fast, easy and reliable.

2.5.6.2 PAC library screening by PCR

There are 21 primary pools (A-U) for the PAC RPCI1 library which is screened by PCR. Positive signals obtained for any pool, involves the further screening of fifteen plates. The positive plate is then tested to obtain two positive bands that correspond to a row and a column. This gives a PAC clone of plate (number), row (letter), column (number), for example clone 705D8.

2.5.7 Purification of PCR products

It is often necessary to carry out a post PCR purification in order to remove primers, dNTPs or to exchange buffers. A number of different methods have been used.

2.5.7.1 Sephacryl® HR columns (S200/400)

These columns (Pharmacia) provide a convenient method for desalting, buffer exchange, removal of primers and labelled and unlabelled nucleotides from DNA solutions. They can be used for purification of plasmid minipreps prior to sequencing and for purification of PCR products. The Sephacryl (HR resin sephacryl equilibrated in TE buffer, pH 7.6) was initially resuspended in the column by gentle vortexing and then
placed in an open-topped Eppendorf tube and pelleted at 1000 g for 1 min to compact the column. The column was removed into a clean Eppendorf tube and 20-25 μl of DNA sample was applied to the top centre of the compacted bed being careful not to disturb the matrix. Centrifugation for 1 minute allowed DNA to elute into the eppendorf tube below. S200 columns were routinely used for purification of labelled probes prior to hybridisation. These columns also facilitated buffer exchange and desalting for purification of probes before hybridisation. S400 columns were used for removal of excess primers and ‘primer dimers’ from PCR products prior to sequencing. Both gel matrices removed unincorporated dNTPs provided the DNA fragment was greater than 100 bp in length.

2.5.7.2. Centricon 100 spin columns (Centricon, USA)

These were routinely used for the purification of PCR products prior to ABI sequencing. Columns were assembled according to manufacturer’s guidelines and 5-10 μl of PCR product in 2 ml of sterile distilled water was added to the upper reservoir of each column. These were centrifuged at 1000 g for 15 minutes (allowing the product DNA to remain on the membrane while all unincorporated primers and dNTPs pass through). The column was then inverted and centrifuged for 5 minutes at 3000 g allowing the DNA eluate to be collected. The purified products were directly used for cycle sequencing.

2.5.7.3. QIAquick spin columns (QIAGEN)

This method is designed to purify PCR products ranging from 100 bp to 10 kb eliminating primers, nucleotides, polymerases and salts.

Five volumes of buffer PB were added to 1 volume of the PCR reaction mix, the sample was then applied to a QIAquick spin column previously placed in a 2 ml collection tube, and centrifuged for 30-60 seconds at 8000 g. The flow-through was discarded and the column was placed back into the same tube. To wash, 0.75 ml of buffer PE was added to the column, followed by centrifugation for 30-60 seconds at 8000 g. After discarding the flow-through the column was centrifuged for an additional minute at maximum speed to eliminate any residual ethanol from buffer PE. The column was then
placed in a clean 1.5 ml tube. To elute DNA, 30 µl of elution buffer was added to the centre of the column, left to stand for 1 minute and spun for 1 minute at 8000 g. The eluted DNA was directly used for cycle sequencing.

2.5.7.4. QIAquick gel extraction (Qiagen)

This method is designed to extract and purify DNA of 70 bp to 10 kb from agarose gel in TAE buffer.

The DNA fragment of interest was excised from the agarose gel with a clean, sharp scalpel, placed in a 1.5 tube and weighed. Three volumes of buffer QG was added to 1 volume of gel and the mixture was then incubated at 50 °C for 10 minutes or until the gel slice was completely dissolved. Occasionally, during the incubation time, the tube was vortexed to help dissolve the gel slice. After it was completely dissolved, 1 gel volume of isopropanol was added to the sample and mixed. To bind DNA, the sample was applied to a QIAquick column previously placed in a 2 ml collection tube and centrifuged for 1 minute at 8000 g. After discarding the flow-through and placing the column back in the same collection tube, 0.5 ml of buffer QG was added to the column followed by 1 minute of centrifugation at 8000 g. The flow-through was again discarded. To wash, 0.75 ml of buffer PE was added to the column, followed by centrifugation for 30-60 seconds at 8000 g. After discarding the flow-through, the column was centrifuged for an additional minute at maximum speed to eliminate any residual ethanol from buffer PE. The column was then placed in a clean 1.5 ml tube. To elute DNA, 30 µl of buffer EB was added to the centre of the column, left to stand for 1 minute and spun for 1 minute at 8000 g. The eluted DNA was directly used for cycle sequencing.

2.5.8 Automated fluorescent DNA sequencing.

All sequencing methods used here were adapted from the chain-termination sequencing method (Sanger et al., 1977).

Automated sequencing was performed on an ABI 373A with the XL adaptation or ABI 3100 DNA sequencer (Applied Biosystems), using the ABI PRISM Amplitaq DNA polymerase FS, Dye Terminator Cycle Sequencing Ready reaction Kit or the Big dye
terminator kit which gave signals of more even intensity.

Both cloned material and PCR products could be used for cycle sequencing. PCR products were directly sequenced after purifying the template to remove unincorporated primers and dNTPs by passing through Centricon-100 spin columns (section 2.7.7.2). A fraction of the purified product was quantified on an agarose gel (section 2.5.4.1). The cycle sequencing reaction for both cloned and PCR fragments consisted of 4 μl terminator ready reaction mix (includes labelled dye terminators, buffer, and dNTPs), 1 μl sequence specific primer (3.2 pmole) and 0.5 mg template DNA (cloned DNA or PCR product) in a total volume of 10 μl. This was subjected to 25 cycles at 96 °C for 10 seconds, 50 °C for 5 seconds and 60 °C for 4 minutes in a Perkin Elmer GeneAmp 9600. Excess unincorporated fluorescent dye was removed by ethanol precipitation of the DNA with 2.5 x volume of absolute ethanol. The samples were then washed with 70% ethanol and vacuum dried. The resultant pellet was resuspended in 5 μl of ABI loading buffer, and the sample was denatured at 95 °C for 3 minutes before loading onto the denaturing acrylamide gel. The ABI 373A/ABI 3100 DNA sequencer was set up and run in accordance to the manufacturer’s instructions. The gel was generally run for 13 hours or the appropriate time for the capillaries used on the 3100.

2.5.9 DNA Hybridisation

2.5.9.1 Random priming labelling of probes.

The random primed labelling technique is a primer extension method in which hexanucleotides of random sequence are annealed to denatured DNA probe. These act as primers for incorporation of [α-32P]dCTP into newly synthesised strands by the action of Klenow polymerase. A commercial kit (Ready to Go dCTP labelling Kit, Pharmacia) was used following the manufacturer’s instructions. PCR probes were spun through S-200 columns prior to labelling (section 2.5.7.1). 25-50 ng of probe DNA was denatured in 27 μl dH2O, and immediately quenched on ice. 20 μl of dH2O was used to resuspend a lyophilised mix containing dATP, dGTP, dTTP and random hexanucleotides in 1 x reaction buffer. 3 μl (1.1 MBq) of [α-32P]dCTP and the denatured DNA was added to the
mix and the reaction was incubated at 37°C for 15-30 min. 0.5 μl was removed and the degree of incorporation was tested by the precipitation of the labelled probe with 5% trichloroacetic acid onto a glass fibre disc (Whatmann GF/B). Unincorporated dCTP was removed by passing the labelled mix through Sephacryl S-200 spin columns (section 2.5.7.1). The labelled probe was denatured at 95°C for 3 min immediately before adding it to pre-heated hybridisation solution.

2.5.9.2 Hybridisation

DNA was hybridised in (i) 25 ml 5 X SSC, 5 X Denhardt’s solution (section 2.1.3.2) and 0.5 % (w/v) SDS with 0.5 ml (1mg/ml) of denatured sonicated non-homologous DNA added to the prehybrisation solution for greater sensitivity or (ii) DNA was hybridised in 20 ml of pH 7.2 Church’s buffer (section 2.1.3.2). Filters were prehybridised for 2-3 hours and hybridised overnight at 58 °C in an oven.

2.5.9.3 PAC Library Screening

The PAC library used was constructed by Pieter de Jong and his group at the Roswell Park Cancer Institute, Buffalo, using the vector pCYPAC2. The DNA source is a normal male blood donor, with an average insert size of 110 kb. The library consists of approximately 120,000 clones in 314 384-well microtitre plates. 25% of the clones lacked insert. The RPCI11 library available from HGMP contains plates 1 – 321 and is supplied either as pools for screening by PCR (2.5.6.2), or as high density gridded filters for screening by hybridisation. Probes for PAC filter hybridisation were prepared as described in section 2.5.9.1 and hybridised as described in 2.5.9.2.

2.5.9.4 Post hybridisation washes

Filters were washed to remove non-specific hybridisation with decreasing concentration of SSC and increasing temperature. Initially, a low stringency wash for 10 minutes at room temperature in 2 X SSC and 0.1 % (w/v) SDS was carried out. If background counts were still high on monitoring, the stringency of wash was increased
by gradually decreasing SSC concentration to a maximum of 0.1% SSC with 0.1% (w/v) SDS and raising the temperature to 65 °C. Washed filters were exposed to Kodak or Fuji X-ray film between two intensifying screens.

**2.5.9.5 Fluorescent *in situ* hybridisation**

PAC or YAC DNA was isolated as described in sections 2.5.1.2 and 2.5.1.3. DNA was labelled with biotin-14-dATP using nick translation as described in the Bionick kit, (GIBCO-BRL Life Technologies). The labelled probe was precipitated with Cot-1 DNA and herring sperm DNA (75μg/ml) and resuspended in a hybridisation mixture containing 50% (w/v) formamide and 10% (w/v) dextran sulphate pH 7.0.

Human metaphase chromosome were obtained from lymphocyte cultures synchronised by the addition of thymidine to block DNA synthesis. This was later removed before the incorporation of 5-bromodeoxyuridine (BrdU) and harvesting the cells. Standard cytogenetic techniques were used for colcemid arrest, hypotonic treatment, fixing of cells, the preparation of slides and the denaturing of the DNA.

The probe was denatured and was allowed to preanneal at 37 °C for 30 minutes before being hybridised on the slide at 37 °C overnight.

The next day the slides went through a series of washes at 42 °C in 50% (w/v) formamide in 2xSSC and the stringency was increased by reducing the salt to 0.1xSSC. The following steps were carried out at room temperature, the slides were blocked with non-fat milk/4xSSC for 20 minutes and washed in 0.05% (w/v) Tween 20/4XSSC. In order to detect the fluorescent signal the slides were incubated in fluorescein isothiocyanate (FITC)-conjugated avidin 5ug/ml (Vector laboratories) in skimmed milk/4xSSC. This was amplified with biotinylated anti-avidin avidin 5ug/ml (Vector laboratories) in skimmed milk/4xSSC and then followed by a second round of amplification.

The chromosome banding patterns were visualised using fluorescent counterstains diadainophenylindole (DAPI) and propidium iodide which were added to the anti-fade mounting medium (Vectashield, Vector Laboratories).
2.5.10 Micosatellite markers

The microsatellite markers used were generally derived from published Genethon/CHLC marker maps (http://www.gdb.org). Unless otherwise stated the markers were analysed as described in section 2.5.5.4.

2.5.11 Mutation Detection

Genes were analysed for mutations using a combination of different methods.

2.5.11.1 Heteroduplex Analysis on DHPLC

Heteroduplex analysis can be used to establish differences between any two similar DNA sequences. This method is based on the principle that two complementary strands of DNA that correspond to two differing alleles will have a region of mismatch when paired together. This mismatch in the double stranded heteroduplex leads to the retarded migration of the DNA when compared to the homoduplexes of each allele which do not have any mismatches.

PCR was used to amplify a segment of DNA, generally 100-400 bp. The DNA was then denatured at 95°C for 5 minutes and then cooled at 1°C per minute to room temperature. This allowed the formation of double stranded DNA of both homoduplexes and the heteroduplex. The samples were then analysed on a denaturing high pressure liquid chromatograph (DHPLC) (Transgenomic) following the manufacturers instructions.

2.5.11.2 Direct sequencing

Direct sequencing of PCR products (2.5.8) and the subsequent alignment of affected and unaffected sequences using software such as DNASTAR (2.5.13.1) to identify any sequence change was the preferred method. If a sequence change was detected, DNASTAR was used to identify restriction sites (section 2.5.3) that a mutation might create or abolish within a PCR fragment which would then be used as a test for segregation within a family or be used to rapidly analyse control individuals.
2.5.12 Tissue culture techniques

2.5.12.1. Harvesting of cells

The cell lines used were SV40 transformed leucocyte cell line from two affected and 1 unaffected individual from the CHED1 pedigree, EK1BR a corneal stromal cell line and HCEC, a human corneal endothelial cell line. The cell lines were grown in 75 ml flasks with the addition of 20 ml DMEM media (section 2.1.3.1) at 37°C in a humidity oven for 2-3 days until 80-90% confluent.

Media was removed using a 10 ml plastic pipette and the cells were washed twice with 10 ml PBS. To dislodge cells from the bottom of the flask, 1 ml of trypsin/EDTA (GIBCOBRL) was added and left at room temperature for 3-4 minutes, flasks were then tapped vigorously to facilitate cells dislodging. 10 ml of fresh media was then added and cells were resuspended by pipetting a few times. 10 µl of cell solution was placed on a clean haemocytometer (Neubauer improved, Assistent®) to perform the cell count and cells in the 5x5 reticulate were counted under a light microscope. The total number of cells \( n^{TOT} \) was derived from the number of cells counted \( n \) using the following formula:

\[
 n^{TOT} = n \times 10^4 \; \text{cells/ml (dilution factor)} \times 11 \; \text{ml (total volume)} 
\]

The cell suspension was then transferred to a 50 ml Falcon tube and cells were pelleted by centrifuging at 800 g for 5 minutes at RT. The supernatant was discarded and the pellet was directly used for RNA extraction (section 2.5.1.4).

2.5.12.2 Freezing of cultured cells

The same cell culture freezing medium with DMSO (GIBCOBRL) can be used for both suspension and adherent cell lines.

Suspension cells were gently pelleted by centrifugation at 800 g for 5 minutes. The medium was removed down to the smallest volume above the cell pellet without disturbing the cells. The cells were gently resuspended in the cell culture freezing medium at a concentration of between \( 1 \times 10^7 \) cells/ml to \( 1 \times 10^8 \) cells/ml and transferred to a pre-cooled cryovial. Prior to this, the number of cells was counted with a haemocytometer.
Adherent cells needed to be detached from the culture flasks on which they were grown. This was done by trypsinizing the cells for 2-3 minutes with 1 ml of trypsin for each 75 cm² flask. The detached cells were resuspended in 10 ml of the medium they were grown in to neutralise the effect of the trypsin. The cells were then pelleted by centrifugation at 1000 rpm for 5 minutes before being counted and gently resuspended, in cell culture freezing medium, at a concentration of 5 X 10⁶ cells/ml to 1 X 10⁷ cell/ml.

For both suspension and adherent cell lines, once the cells were resuspended in the freezing medium, they were left on ice for 5 minutes. The cryovials were placed in a precooled freezing block containing isopropanol which gently cooled the cells to -80°C, the following day the cyovials were placed in a liquid nitrogen container for long term storage.

### 2.5.12.3 Reviving frozen cells

The cells were quickly thawed in a 37°C water bath and 1 ml of frozen cells was slowly mixed with 10 ml of complete growth medium. The cells were gently mixed and then pelleted by centrifugation at 800 g for 5 minutes. The supernatant was aspirated off and the cell pellet was gently resuspended in 10 ml of complete medium in an appropriate growth vessel.

### 2.5.13 Computational tools and databases

#### 2.5.13.1 DNA analysis

The software programs, Geneworks (Oxford Molecular Group) and DNASTAR (DNASTAR, Inc. Wisconsin USA including EditSeq™, GeneQuest™, MapDraw™, MegAlign™, PrimerSelect™, Protean™ and SeqMan™) were used as tools to analyse DNA sequences. This analysis included editing of sequences, alignments, restriction maps and open reading frame finder.

#### 2.5.13.2 Computer packages for Linkage

Linkage analysis was performed using the LINKSYS (Attwood et al., 1988), Cyrillic (FamilyGenetix) and LINKAGE packages (Lathrop et al., 1984). The LINKSYS and
Cyrillic package facilitate the management of genetic data to be used in conjunction with the analytical packages of LINKAGE and LIPED. The LS4 data management package of LINKSYS, organises and processes genotype data, allele frequencies of markers and pedigree information. The LINKAGE package includes programs MLINK, ILINK and LINKMAP. Two-point analysis was carried out using MLINK that tabulates two-point lod scores, while LINKMAP was used for multipoint analysis.

Multipoint analysis was performed through the computing facilities at HGMP Resource Centre.

2.5.13.3 Bioinformatics

The web sites for HGMP, NCBI and the Sanger Institute provided access to the majority of databases and bioinformatic tools used in this thesis. The BLAST algorithms were used extensively for gene and protein sequence analysis, including BLASTN, BLASTX, BLASTP, TBLASTX, TBLASTN. (http://www.ncbi.nlm.nih.gov)

Suites of programs at HGMP were used for gene analysis; (NIX) genomic sequence analysis (NIX) and protein sequence analysis (PIX). (http://www.hgmp.mrc.ac.uk)

For genomic map viewing Mapview (NCBI), WebACE (Sanger) and Ensembl (EBI) were used.

EST and expression databases used included Body Map (http://bodymap.ims.u-tokyo.ac.jp/), Unigene (NCBI) and TIGR (http://www.tigr.org/tdb/hgi/index.html).

2.5.13.4 Web sites and databases used

Human Genome Database (GDB): http://www.gdb.org
Genethon: http://www.genethon.fr/genethon_en.html
Whitehead: http://www-genome.wi.mit.edu/
GeneCards: http://bioinfo.weizmann.ac.il/cards/
Sanger Centre: http://www.sanger.ac.uk/HGP/
CorneaNet: http://www.sph.uth.tmc.edu/Retnet/home.htm
Genetic Location Database (LDB):
http://cedar.genetics.soton.ac.uk/public_html/ldb.html
CHAPTER 3

Congenital Hereditary Endothelial Dystrophy

CHED1

3.1 Introduction

Congenital Hereditary Endothelial Dystrophy (CHED) belongs to a group of disorders known as the corneal endothelial dystrophies, which also includes Posterior Polymorphous Dystrophy (PPCD), Fuch’s endothelial dystrophy and Iridocorneal Endothelial Syndrome (ICE), in which the major pathology is attributed to the abnormal development of the corneal endothelium (section 1.2.6). CHED was initially reported in the English literature in 1936 (Fischer et al., 1936), subsequently this phenotype was more fully described by Maumenee (1960). CHED can be inherited either as an autosomal dominant (CHED1) or autosomal recessive disease (CHED 2) (section 1.3.4 and Chapter 4). It has been suggested that the term congenital be used to describe only the recessive condition, as the opacities are seen at birth, or within the neonatal period (Judisch et al., 1978). In the dominant condition, opacities are rarely seen at birth often appear in the first few years of life and are slowly progressive (Maumenee, 1960). Judisch and colleagues (1978) suggested that the term infantile or juvenile might therefore be more appropriate.

CHED1 is characterised by diffuse, bilateral corneal opacities during the first few years of life (Figure 3.1). As a result, it can often be misdiagnosed as infantile glaucoma and treated accordingly, however, the intraocular pressure and the diameter of the cornea are generally normal in CHED1 (Pearce et al., 1969; Kenyon et al., 1971; Kirkness 1989). Electron and light microscopy have shown that the changes seen in the corneal endothelium are the major pathological features in this disease. The main changes seen in CHED1 corneal tissue is an abnormal Descemet’s membrane and endothelium (Figure 3.2). The endothelium is either absent or markedly dystrophic with an absence of the
hexagonal morphology of normal cells. The dystrophic endothelial cells are irregular and multi-nucleated. The endothelium also shows an increase in permeability to fluorescein which would suggest a failure in the barrier function of the endothelium (Burns et al., 1981; Geroski et al., 1985; McCartney et al., 1987; Ehlers et al., 1998).

Figure 3.1 Slit lamp photograph of a cornea from a CHED1 patient.
Arrow indicates the opacity that gives the cornea a bluish-white haze (with permission from D.F.Larkin).

Figure 3.2 Light micrograph sections of corneal tissue.
Arrows indicate the atrophy of the endothelial cells (blue) and thickening of the Descement’s membrane (black) when comparing diseased (B) and unaffected (A) sections stained with haemaotoxylin and eosin. (Adapted from Eagle, 1999).
Changes in the epithelium and stromal layers have also been noted but are thought to be secondary to the changes seen in the endothelium and Descemet’s membrane. Areas of Bowman’s membrane are sometimes replaced with disorganised collagen fibrils, while the stromal layer loses its regular arrangement of collagen fibrils within the lamellae and the keratocytes often have swollen organelles. These changes are thought to arise from the loss of endothelial cells and hence a loss of the osmotic pump function of the endothelium and which results in oedema (McCartney et al., 1988; Ehlers et al., 1998).

3.2 CHED1 family pedigree

A seven generation British CHED1 family attending Moorfields Eye Hospital has previously been reported (Pearce et al., 1969). An autosomal dominant mode of inheritance with complete penetrance was observed. Of the 123 individuals in this pedigree (Figure 3.3), 40 (19 affected, 13 unaffected and 8 spouses) were enrolled in this study to locate and characterise the gene segregating in the family (Figure 3.3).

All affected subjects had the distinctive changes in the cornea normally associated with CHED1. The characteristic feature of the disease in this family was the milky opacification of the cornea with some degree of photophobia (Figure 3.1). Although no corneal opacities were seen at birth, photophobia was reported shortly afterward with corneal opacities becoming apparent in the first to second year. The clinical picture was similar to that seen in Fuch’s dystrophy patients, however, no corneal guttata were seen (section 1.3.3.1).

Variation of the phenotype within affected individuals of the family was noted, in some individuals one eye was less severely affected. In all cases, penetrating keratoplasty was necessary to restore vision. It was also noted that a number of these patients have had multiple grafts due to the recurrence of the disease. The major finding, using electron microscopy, was the increase in Descemet’s membrane caused by an increase in thickness of the non-banded zone (section 1.2.6) which was between 4 and 15 times thicker in affected patients and was thickest in older patients (Pearce et al., 1969). No additional medical condition was found to segregate with the disease.
Figure 3.3 CHED1 pedigree including new generation
Filled symbols denote affected individuals and ? Indicates individuals of unknown status.
Circles represent females and squares males. Lines through symbols represent deceased individuals.
3.3 Linkage analysis of the CHED1 pedigree

The initial strategy of this study was to use linkage analysis across a number of candidate loci to determine the locus for CHED1. These candidate loci were chosen as a result of either the function of the protein or because they were implicated in other ocular dystrophies. Collagen type VIII (section 1.2.2.1) is a major component of Descemet’s membrane therefore any changes in the expression of this gene could result in a thickened membrane (Muragaki et al., 1991; Tamura et al., 1991). The collagen type IV genes Col4A3 and Col4A4 were mapped to chromosome 2q36 and are an important component of basement membranes (section 1.2.2.1).

Stickler syndrome is a dominantly inherited arthro-ophthalmolopathy that was linked to chromosome 6. The ophthalmic symptoms include vitreo-retinal degeneration, myopia and glaucoma and mutations were identified within the Col2A1gene (Knowlton et al., 1989; Bonaventure et al., 1992). The aniridia locus on chromosome 11 was selected because mutations in Pax 6 show a wide range of phenotypic expression, including Peter’s anomaly with changes to the corneal endothelium (Mannens et al., 1989; Jordan et al., 1992). The Reiger’s syndrome locus on chromosome 4 with anterior segment abnormalities (Murray et al., 1992) and the stromal dystrophies mapping to chromosome 5 (Stone et al., 1994) were all considered as candidate loci.

<table>
<thead>
<tr>
<th>Candidate Disease/Gene region</th>
<th>Chromosomal localisation</th>
<th>Markers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collagen IV</td>
<td>2q36</td>
<td>D2S126</td>
</tr>
<tr>
<td>Collagen VIII</td>
<td>3q12</td>
<td>D3S1302</td>
</tr>
<tr>
<td>Reiger’s Syndrome</td>
<td>4q25</td>
<td>D4S622</td>
</tr>
<tr>
<td>Stromal Dystrophies</td>
<td>5q31</td>
<td>D5S393</td>
</tr>
<tr>
<td>Stickler Syndrome</td>
<td>6p21</td>
<td>D6S260</td>
</tr>
<tr>
<td>Aniridia and Peter’s anomaly</td>
<td>11p13</td>
<td>D11S907</td>
</tr>
<tr>
<td>Cone-Rod Dystrophy</td>
<td>19q13</td>
<td>D19S211</td>
</tr>
<tr>
<td>Posterior Polymorphous Dystrophy</td>
<td>20p12-20q</td>
<td>D20S112</td>
</tr>
</tbody>
</table>

Table 3.1 Candidate loci and the associated markers analysed in the CHED1 pedigree.
3.3.1 Microsatellite analysis of candidate loci

The microsatellite markers used were either dinucleotide markers from the CEPH/Genethon maps or tetranucleotide markers from the CHLC maps. For the analysis of candidate genes or loci, markers closely linked to the locus were tested (Table 3.1).

The microsatellite markers were amplified using PCR and labelled with radioactive isotopes, either by incorporation labelling or by end-labelling one of the primers, in the PCR reaction (section 2.5.5.4). The radiolabelled PCR products were resolved on a 6% denaturing polyacrylamide gel by electrophoresis (section 2.5.4.2) and visualised by autoradiography.

3.3.1.1 Calculation of lod scores for microsatellite markers

Lod scores for each marker analysed were calculated. The analytical software package LINKAGE 5.1 was used to calculate the two point lod scores with the MLINK program (Lathrop and Lalouel, 1984; Attwood and Bryant, 1988). In order to use this program, a compatible output file from either LINKSYS or Cyrillic was used. These computer programs were used to process the familial, allele frequency, genotype and phenotype data (section 2.5.10).

3.3.1.2 Exclusion of candidate loci for CHED1

The markers linked to candidate loci listed in Table 3.1 were analysed across the CHED1 pedigree as shown in Figure 3.4 and the haplotypes that were used to calculate the lod score are shown in Table 3.2.

Figure 3.4 Microsatellite analysis of the CHED1 family for marker D5S393.
Individuals are identified by a pedigree number (Figure 3.6) which correspond to lane numbers 1-40 in the autoradiograph.
<table>
<thead>
<tr>
<th>Ped Nos</th>
<th>Lane Nos</th>
<th>Status</th>
<th>Alleles</th>
<th>Ped Nos</th>
<th>Lane Nos</th>
<th>Status</th>
<th>Alleles</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-8</td>
<td>35</td>
<td>N</td>
<td>34</td>
<td>6-4</td>
<td>20</td>
<td>N</td>
<td>23</td>
</tr>
<tr>
<td>4-9</td>
<td>31</td>
<td>A</td>
<td>35</td>
<td>6-5</td>
<td>22</td>
<td>A</td>
<td>23</td>
</tr>
<tr>
<td>5-1</td>
<td>24</td>
<td>N</td>
<td>33</td>
<td>6-6</td>
<td>29</td>
<td>N</td>
<td>23</td>
</tr>
<tr>
<td>5-2</td>
<td>23</td>
<td>A</td>
<td>33</td>
<td>6-7</td>
<td>6</td>
<td>N</td>
<td>33</td>
</tr>
<tr>
<td>5-3</td>
<td>19</td>
<td>A</td>
<td>13</td>
<td>6-8</td>
<td>3</td>
<td>N</td>
<td>33</td>
</tr>
<tr>
<td>5-4</td>
<td>18</td>
<td>N</td>
<td>23</td>
<td>6-9</td>
<td>4</td>
<td>A</td>
<td>33</td>
</tr>
<tr>
<td>5-5</td>
<td>21</td>
<td>N</td>
<td>13</td>
<td>6-10</td>
<td>5</td>
<td>A</td>
<td>13</td>
</tr>
<tr>
<td>5-7</td>
<td>34</td>
<td>A</td>
<td>33</td>
<td>6-11</td>
<td>8</td>
<td>N</td>
<td>33</td>
</tr>
<tr>
<td>5-8</td>
<td>1</td>
<td>N</td>
<td>33</td>
<td>6-12</td>
<td>7</td>
<td>A</td>
<td>33</td>
</tr>
<tr>
<td>5-9</td>
<td>2</td>
<td>A</td>
<td>13</td>
<td>6-13</td>
<td>16</td>
<td>N</td>
<td>33</td>
</tr>
<tr>
<td>5-10</td>
<td>14</td>
<td>N</td>
<td>33</td>
<td>6-14</td>
<td>17</td>
<td>A</td>
<td>33</td>
</tr>
<tr>
<td>5-11</td>
<td>15</td>
<td>A</td>
<td>34</td>
<td>6-15</td>
<td>26</td>
<td>A</td>
<td>13</td>
</tr>
<tr>
<td>5-12</td>
<td>32</td>
<td>A</td>
<td>34</td>
<td>6-16</td>
<td>39</td>
<td>N</td>
<td>34</td>
</tr>
<tr>
<td>5-13</td>
<td>33</td>
<td>N</td>
<td>--</td>
<td>6-17</td>
<td>27</td>
<td>N</td>
<td>23</td>
</tr>
<tr>
<td>5-15</td>
<td>25</td>
<td>A</td>
<td>33</td>
<td>7-1</td>
<td>13</td>
<td>N</td>
<td>33</td>
</tr>
<tr>
<td>5-16</td>
<td>28</td>
<td>A</td>
<td>33</td>
<td>7-2</td>
<td>12</td>
<td>N</td>
<td>33</td>
</tr>
<tr>
<td>5-17</td>
<td>36</td>
<td>N</td>
<td>33</td>
<td>7-3</td>
<td>11</td>
<td>N</td>
<td>33</td>
</tr>
<tr>
<td>6-1</td>
<td>38</td>
<td>A</td>
<td>33</td>
<td>7-4</td>
<td>10</td>
<td>A</td>
<td>33</td>
</tr>
<tr>
<td>6-2</td>
<td>37</td>
<td>N</td>
<td>33</td>
<td>75</td>
<td>9</td>
<td>A</td>
<td>33</td>
</tr>
<tr>
<td>6-3</td>
<td>30</td>
<td>A</td>
<td>12</td>
<td>7-6</td>
<td>40</td>
<td>A</td>
<td>13</td>
</tr>
</tbody>
</table>

Table 3.2 Individuals in the CHED1 pedigree used for linkage analysis. Alleles are shown for marker D5S393. Disease status, lane number and pedigree identification is also shown.

The calculated lod scores are shown in Table 3.3. A lod score of -2 or less was taken as evidence against linkage, within an interval equal to the corresponding θ value from both sides of the marker locus. Exclusion mapping of the loci shown in Table 3.3 was undertaken in collaboration with N.M.G. Toma.
Table 3.3 Two-point lod scores between CHED1 disease phenotype and polymorphic markers.

The markers D2S126, D3S1302, D4S622, D5S393, D6S260, D11S907 and D19S211 were analysed. A lodscore of -2.00 or less was taken as evidence against linkage at that locus and is shown in bold.

The candidate loci tabulated in Table 3.3 were excluded, due to the observation of a number of recombination events between the markers from these loci and the disease in this family resulting in a lod score of -2 or less.

While markers from these candidate loci were being analysed for linkage in the CHED1 family, Heon and colleagues (1995) reported the mapping of posterior polymorphous dystrophy to the long arm of chromosome 20. As this dystrophy affects the corneal endothelium and shares some features of CHED it became an excellent candidate locus, therefore microsatellite markers from this region were analysed.
3.3.2 Linkage of the CHED1 disease phenotype to chromosome 20

Linkage analysis localised PPCD to a 30 cM interval on chromosome 20 flanked by microsatellite markers D20S98 (distal) and D20S108 (proximal) (Heon et al., 1995). Using markers for the candidate loci from Table 3.1, the disease in the CHED1 family was shown to segregate with markers on chromosome 20 spanning the PPCD locus. The CEPH/Genethon and CHLC genetic maps were integrated by using framework markers, such as D20S54 (Figure 3.6), to further refine the disease interval. Detection of recombinants within the CHED1 family confirmed the marker order shown in Table 3.5.

3.3.2.1 Chromosome 20 microsatellite and haplotype analysis

An example of microsatellite marker analysis across the CHED1 pedigree is shown in Figure 3.5. The alleles were scored independently of the family pedigree data (Table 3.4) before combining these data to determine the segregation of alleles (Figure 3.6).

![Figure 3.5 Microsatellite analysis of the CHED1 family for marker D20S471. Individuals are identified by a pedigree number (Figure 3.5) which correspond to lane numbers 1-40 in the autoradiograph.](image)
<table>
<thead>
<tr>
<th>Ped Nos</th>
<th>Lane Nos</th>
<th>Status</th>
<th>Alleles</th>
<th>Ped Nos</th>
<th>Lane Nos</th>
<th>Status</th>
<th>Alleles</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-8</td>
<td>35</td>
<td>N</td>
<td>35</td>
<td>6-4</td>
<td>20</td>
<td>N</td>
<td>37</td>
</tr>
<tr>
<td>4-9</td>
<td>31</td>
<td>A</td>
<td>34</td>
<td>6-5</td>
<td>22</td>
<td>A</td>
<td>39</td>
</tr>
<tr>
<td>5-1</td>
<td>24</td>
<td>N</td>
<td>55</td>
<td>6-6</td>
<td>29</td>
<td>N</td>
<td>68</td>
</tr>
<tr>
<td>5-2</td>
<td>23</td>
<td>A</td>
<td>34</td>
<td>6-7</td>
<td>6</td>
<td>N</td>
<td>23</td>
</tr>
<tr>
<td>5-3</td>
<td>19</td>
<td>A</td>
<td>37</td>
<td>6-8</td>
<td>3</td>
<td>N</td>
<td>15</td>
</tr>
<tr>
<td>5-4</td>
<td>18</td>
<td>N</td>
<td>35</td>
<td>6-9</td>
<td>4</td>
<td>A</td>
<td>33</td>
</tr>
<tr>
<td>5-5</td>
<td>21</td>
<td>N</td>
<td>47</td>
<td>6-10</td>
<td>5</td>
<td>A</td>
<td>33</td>
</tr>
<tr>
<td>5-7</td>
<td>34</td>
<td>A</td>
<td>38</td>
<td>6-11</td>
<td>8</td>
<td>N</td>
<td>59</td>
</tr>
<tr>
<td>5-8</td>
<td>1</td>
<td>N</td>
<td>33</td>
<td>6-12</td>
<td>7</td>
<td>A</td>
<td>33</td>
</tr>
<tr>
<td>5-9</td>
<td>2</td>
<td>A</td>
<td>23</td>
<td>6-13</td>
<td>16</td>
<td>N</td>
<td>18</td>
</tr>
<tr>
<td>5-10</td>
<td>14</td>
<td>N</td>
<td>48</td>
<td>6-14</td>
<td>17</td>
<td>A</td>
<td>38</td>
</tr>
<tr>
<td>5-11</td>
<td>15</td>
<td>A</td>
<td>13</td>
<td>6-15</td>
<td>26</td>
<td>A</td>
<td>35</td>
</tr>
<tr>
<td>5-12</td>
<td>32</td>
<td>A</td>
<td>33</td>
<td>6-16</td>
<td>39</td>
<td>N</td>
<td>22</td>
</tr>
<tr>
<td>5-13</td>
<td>33</td>
<td>N</td>
<td>34</td>
<td>6-17</td>
<td>27</td>
<td>N</td>
<td>33</td>
</tr>
<tr>
<td>5-15</td>
<td>25</td>
<td>A</td>
<td>35</td>
<td>7-1</td>
<td>13</td>
<td>N</td>
<td>13</td>
</tr>
<tr>
<td>5-16</td>
<td>28</td>
<td>A</td>
<td>35</td>
<td>7-2</td>
<td>12</td>
<td>N</td>
<td>13</td>
</tr>
<tr>
<td>5-17</td>
<td>36</td>
<td>N</td>
<td>34</td>
<td>7-3</td>
<td>11</td>
<td>N</td>
<td>13</td>
</tr>
<tr>
<td>6-1</td>
<td>38</td>
<td>A</td>
<td>35</td>
<td>7-4</td>
<td>10</td>
<td>A</td>
<td>--</td>
</tr>
<tr>
<td>6-2</td>
<td>37</td>
<td>N</td>
<td>45</td>
<td>7-5</td>
<td>9</td>
<td>A</td>
<td>35</td>
</tr>
<tr>
<td>6-3</td>
<td>30</td>
<td>A</td>
<td>33</td>
<td>7-6</td>
<td>40</td>
<td>A</td>
<td>25</td>
</tr>
</tbody>
</table>

Table 3.4 Individuals in the pedigree used for linkage analysis.
Alleles are shown for marker D20S471. Disease status, lane number and pedigree identification are also shown.
Figure 3.6 Abridged CHED1 pedigree and haplotype showing linkage to chromosome 20
Table 3.5 Two-point lod scores between CHED1 disease phenotype and polymorphic markers on chromosome 20.

<table>
<thead>
<tr>
<th>Markers</th>
<th>θ</th>
<th>0.0</th>
<th>0.05</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>Z_{MAX}</th>
</tr>
</thead>
<tbody>
<tr>
<td>D20S112</td>
<td>-∞</td>
<td>3.28</td>
<td>3.05</td>
<td>2.30</td>
<td>1.45</td>
<td>0.64</td>
<td>3.28</td>
<td></td>
</tr>
<tr>
<td>D20S114</td>
<td>-∞</td>
<td>7.07</td>
<td>6.48</td>
<td>4.88</td>
<td>3.02</td>
<td>1.18</td>
<td>7.20</td>
<td></td>
</tr>
<tr>
<td>D20S48</td>
<td>-∞</td>
<td>2.00</td>
<td>1.38</td>
<td>0.68</td>
<td>2.35</td>
<td>0.88</td>
<td>6.05</td>
<td></td>
</tr>
<tr>
<td>D20S471</td>
<td>-∞</td>
<td>4.94</td>
<td>4.57</td>
<td>3.48</td>
<td>2.23</td>
<td>0.98</td>
<td>4.99</td>
<td></td>
</tr>
<tr>
<td>D20S54</td>
<td>-∞</td>
<td>2.26</td>
<td>2.61</td>
<td>2.38</td>
<td>1.53</td>
<td>0.67</td>
<td>3.04</td>
<td></td>
</tr>
<tr>
<td>D20S200</td>
<td>-∞</td>
<td>2.98</td>
<td>3.00</td>
<td>2.39</td>
<td>1.54</td>
<td>0.68</td>
<td>3.04</td>
<td></td>
</tr>
<tr>
<td>D20S106</td>
<td>-∞</td>
<td>5.58</td>
<td>5.06</td>
<td>3.89</td>
<td>2.58</td>
<td>1.27</td>
<td>5.58</td>
<td></td>
</tr>
</tbody>
</table>

Two-point linkage analysis (section 1.6) was used to obtain significant lod scores with a Zmax of >3 for D20S106 and D20S112 (Table 3.5). The recombination events observed with these markers in the CHED1 family (Figure 3.6) localised the disease to an approximately 14 cM interval (Gyapay et al., 1994). Further polymorphic markers within this genetic interval were then tested in the family in order to refine the critical interval (Table 3.5). The highest lod score (Zmax) of 7.20 was obtained with marker D20S114 at θ= 0.026.

The distal flanking microsatellite markers for the CHED1 disease interval were D20S112, D20S114, D20S118 and D20S48 and the proximal tetranucleotide marker was D20S471 (Figure 3.6, Figure 3.7 and Table 3.5). The distance between the distal marker D20S112 and proximal marker D20S471 was estimated to be approximately 4 cM on the short arm of chromosome 20p11.2 (Gyapay et al., 1994; Murray et al., 1994).
Figure 3.7 Ideogram of chromosome 20 showing the disease interval for CHED1 and PPCD.
Flanking markers D20S112, D20S114, D20S118 and D20S471 (pink) indicate the CHED1 locus lies within the critical interval for PPCD (blue). The pink bar represents the refined localisation of the disease interval for CHED (adapted from http://www.ncbi.nlm.nih.gov).

3.3.2.3 Multipoint Linkage Analysis

In order to determine the most likely location of the gene responsible for CHED1 relative to the 3 most closely linked markers within the critical interval, a multipoint linkage analysis was performed. The relative order of the markers tel-D20S114 -1.3 cM-D20S48 -2.7 cM- D20S471-cen, and the interlocus distance was estimated from published data (Gyapay et al., 1994; Murray et al., 1994). The analysis was performed...
between the markers and the disease phenotype using the LINKMAP program that is part of the LINKAGE package (section 2.5.13.2).

The markers D20S48 and D20S471 are the distal and proximal flanking markers respectively, although D20S114 (distal to D20S48) gave a maximum two-point lod score of 7.20. As the LINKMAP program limits the number of alleles it can analyse, it was necessary to reduce the number of haplotypes by recycling the alleles so that the informativeness of the alleles was unaffected. Two-point lod scores were then recalculated for each marker to demonstrate that the recycling of alleles had not significantly altered the lod scores before the multipoint analysis was undertaken.

The data generated from the multipoint analysis is represented in graphical form (Figure 3.8). The multipoint analysis with markers D20S114, D20S48 and D20S471 gave a maximum lod score of 9.34 in a 2.7 cM interval between the markers D20S48 and D20S471. This data suggests that the gene for CHED1 may be localised within this interval, however, the possibility remains that the disease gene may actually lie in a larger critical interval.

Figure 3.8 Multipoint graph of the lod scores vs recombination fraction for markers D20S114, D20S48 and D20S471. The red arrow indicates the maximum lod score of 9.34.
3.3.2.4 Further Analysis of the CHED1 pedigree with new polymorphic markers

New markers were identified within the CHED1 interval on chromosome 20 and the segregation of alleles for each marker analysed was tracked in the CHED1 pedigree to determine sites of meiotic recombination and the order of the marker loci. Three new markers were reported by Genethon (D20S182, D20S605, D20S139; (Dib et al., 1996)) An intragenic marker was utilised (PCSK2; GDB) and two novel markers generated from the physical map and sequence data (Table 3.6, see section 3.4 and 3.5).

<table>
<thead>
<tr>
<th>Name</th>
<th>Primer sequence (Forward)</th>
<th>Primer sequence (Reverse)</th>
<th>Ta °C</th>
<th>Amplimer (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M189K21</td>
<td>gaattcattgctageaatctacc</td>
<td>ccttctttgtgttgtattggtc</td>
<td>55</td>
<td>316</td>
</tr>
<tr>
<td>M1068E13</td>
<td>ttccagtctcattctttggagg</td>
<td>caacatagctaaagccaccc</td>
<td>57</td>
<td>385</td>
</tr>
</tbody>
</table>

Table 3.6 Primer sequences used to amplify novel microsatellite markers within the CHED1 interval.

<table>
<thead>
<tr>
<th>Markers</th>
<th>θ</th>
<th>0.0</th>
<th>0.05</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>Z(MAX)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCSK2</td>
<td>-∞</td>
<td>2.533</td>
<td>2.881</td>
<td>2.572</td>
<td>1.786</td>
<td>0.830</td>
<td>2.881</td>
<td></td>
</tr>
<tr>
<td>D20S182</td>
<td>-∞</td>
<td>2.834</td>
<td>2.704</td>
<td>2.076</td>
<td>1.301</td>
<td>0.570</td>
<td>2.834</td>
<td></td>
</tr>
<tr>
<td>D20S605</td>
<td>-∞</td>
<td>3.465</td>
<td>3.186</td>
<td>2.328</td>
<td>1.418</td>
<td>0.616</td>
<td>3.465</td>
<td></td>
</tr>
<tr>
<td>M189K21</td>
<td>-∞</td>
<td>5.450</td>
<td>5.094</td>
<td>3.950</td>
<td>2.609</td>
<td>1.253</td>
<td>5.450</td>
<td></td>
</tr>
<tr>
<td>D20S139</td>
<td>-∞</td>
<td>3.146</td>
<td>3.153</td>
<td>2.541</td>
<td>1.676</td>
<td>0.764</td>
<td>3.146</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.7 Two-point lod scores between CHED1 disease phenotype and additional markers on chromosome 20.
Figure 3.9 Novel microsatellite marker M1068E13 typed across the CHED1 pedigree.

<table>
<thead>
<tr>
<th>Ped Nos</th>
<th>Lane Nos</th>
<th>Status</th>
<th>Alleles</th>
<th>Ped Nos</th>
<th>Lane Nos</th>
<th>Status</th>
<th>Alleles</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-8</td>
<td>35</td>
<td>N</td>
<td>77</td>
<td>6-4</td>
<td>20</td>
<td>N</td>
<td>67</td>
</tr>
<tr>
<td>4-9</td>
<td>31</td>
<td>A</td>
<td>46</td>
<td>6-5</td>
<td>22</td>
<td>A</td>
<td>46</td>
</tr>
<tr>
<td>5-1</td>
<td>24</td>
<td>N</td>
<td>44</td>
<td>6-6</td>
<td>29</td>
<td>N</td>
<td>27</td>
</tr>
<tr>
<td>5-2</td>
<td>23</td>
<td>A</td>
<td>34</td>
<td>6-7</td>
<td>6</td>
<td>N</td>
<td>37</td>
</tr>
<tr>
<td>5-3</td>
<td>19</td>
<td>A</td>
<td>47</td>
<td>6-8</td>
<td>3</td>
<td>N</td>
<td>24</td>
</tr>
<tr>
<td>5-4</td>
<td>18</td>
<td>N</td>
<td>16</td>
<td>6-9</td>
<td>4</td>
<td>A</td>
<td>47</td>
</tr>
<tr>
<td>5-5</td>
<td>21</td>
<td>N</td>
<td>56</td>
<td>6-10</td>
<td>5</td>
<td>A</td>
<td>34</td>
</tr>
<tr>
<td>5-7</td>
<td>34</td>
<td>A</td>
<td>47</td>
<td>6-11</td>
<td>8</td>
<td>N</td>
<td>12</td>
</tr>
<tr>
<td>5-8</td>
<td>1</td>
<td>N</td>
<td>37</td>
<td>6-12</td>
<td>7</td>
<td>A</td>
<td>47</td>
</tr>
<tr>
<td>5-9</td>
<td>2</td>
<td>A</td>
<td>47</td>
<td>6-13</td>
<td>16</td>
<td>N</td>
<td>77</td>
</tr>
<tr>
<td>5-10</td>
<td>14</td>
<td>N</td>
<td>47</td>
<td>6-14</td>
<td>17</td>
<td>A</td>
<td>47</td>
</tr>
<tr>
<td>5-11</td>
<td>15</td>
<td>A</td>
<td>47</td>
<td>6-15</td>
<td>26</td>
<td>A</td>
<td>24</td>
</tr>
<tr>
<td>5-12</td>
<td>32</td>
<td>A</td>
<td>47</td>
<td>6-16</td>
<td>39</td>
<td>N</td>
<td>26</td>
</tr>
<tr>
<td>5-13</td>
<td>33</td>
<td>N</td>
<td>67</td>
<td>6-17</td>
<td>27</td>
<td>N</td>
<td>26</td>
</tr>
<tr>
<td>5-15</td>
<td>25</td>
<td>A</td>
<td>46</td>
<td>7-1</td>
<td>13</td>
<td>N</td>
<td>47</td>
</tr>
<tr>
<td>5-16</td>
<td>28</td>
<td>A</td>
<td>46</td>
<td>7-2</td>
<td>12</td>
<td>N</td>
<td>47</td>
</tr>
<tr>
<td>5-17</td>
<td>36</td>
<td>N</td>
<td>67</td>
<td>7-3</td>
<td>11</td>
<td>N</td>
<td>47</td>
</tr>
<tr>
<td>6-1</td>
<td>38</td>
<td>A</td>
<td>4-4</td>
<td>7-4</td>
<td>10</td>
<td>A</td>
<td>14</td>
</tr>
<tr>
<td>6-2</td>
<td>37</td>
<td>N</td>
<td>34</td>
<td>75</td>
<td>9</td>
<td>A</td>
<td>27</td>
</tr>
<tr>
<td>6-3</td>
<td>30</td>
<td>A</td>
<td>46</td>
<td>7-6</td>
<td>40</td>
<td>A</td>
<td>46</td>
</tr>
</tbody>
</table>

Table 3.8 Individuals in the pedigree used for linkage analysis.
Alleles scored for novel marker M1068E13. Disease status, lane number and pedigree identification are also shown.
The first microsatellite marker to give a positive lod score above 3 was D20S106 (section 3.3.2.2). Further microsatellite markers spanning the CHED1 locus were genotyped in this family and a haplotype created to refine the disease interval. From the initial linkage to a 14 cM interval on chromosome 20 flanked by the markers D20S112 and D20S106, the disease interval was refined to a 2.7 cM interval between the markers D20S48 and D20S471 (Figure 3.6).

All new markers were typed in the CHED1 family and lod scores calculated (Table 3.7). Familial segregation and alleles for novel marker M1068E13 (derived from analysis of sequence data for PAC 1068E13) are shown in Figure 3.9 and Table 3.8.

Further refinement of the CHED1 locus proved to be unsuccessful using the microsatellite markers shown in Table 3.7. The disease was flanked distally by recombination events seen in individuals V:3 and V:6 for the markers D20S118, D20S114, D20S112 and D20S48 (Figure 3.6). The proximal flanking marker was D20S471 and the recombination event was seen in individuals V:14 and V:15 (Figure 3.6). However, markers PCSK2, D20S605, M189K21 and M1068E13 were linked without any observed recombination (Figure 3.10). Lod scores ranging from 2.8 to 5.4 were achieved with these new markers (Table 3.7).

Microsatellite marker D20S139 showed the same recombination events as marker D20S471 when analysed in this family. However, with the availability of sequence data from chromosome 20, the order of the markers was resolved and D20S139 refined the interval by approximately 100kb (Figure 3.19). The flanking markers for disease interval therefore are D20S48 (distal) and D20S139 (proximal).

The disease interval for CHED1 was further characterised using a physical mapping and positional cloning strategy in order to identify candidate genes and new markers.
Figure 3.10 Haplotype analysis of additional markers on chromosome 20 mapping to the CHED1 disease interval
3.4 Physical mapping of the CHED1 disease interval

Positional cloning strategies are most effective when the disease locus has been refined to a small genetic or physical interval (section 1.9). Using all the previously described molecular markers mapping to human chromosome 20p11.2, the CHED1 locus was localised to an approximate 2.7 cM region between the markers D20S48 and D20S471 (Figure 3.6). Many of these markers were either uninformative (D20S182) or gave the same recombinants as markers previously tested (such as D20S114, D20S112 and D20S118) (Figure 3.6). To further refine the disease locus, it was necessary to identify additional microsatellite markers from within the disease interval.

3.4.1 Isolation of YAC clones

Initial YAC seeding was based on a partial contig, WC20.1, from the Whitehead database (http://www.mit.edu). More than fifteen YACs spanning the critical interval for CHED1 were obtained from HGMP (Hinxton, Cambs). Single colonies were propagated and solution DNA isolated (section 2.5.1). STS content mapping of the YAC DNA was carried out by PCR for fourteen markers (Figure 3.11 and Figure 3.12) using primer sequences and conditions described in GDB (http://www.gdb.org; section 2.5.13.3). A preliminary YAC contig was assembled (Figure 3.12) representing the disease interval.

![Figure 3.11 STS content mapping of 15 YACs using marker WI-9743.](image)

YACs 769E6 (lane 5), 875B10 (lane 9) and 972H6 (lane 12) with a positive signal are shown. Lane 18 is a positive control of human genomic DNA, lane 16 is a negative control.
Figure 3.12 Preliminary YAC contig spanning the critical interval for CHED1. The flanking markers are shown in green, other microsatellite markers in dark blue and STSs in light blue. The yellow circles indicate a positive signal while the blue circles denote a negative signal. The sequences of all STSs used were obtained from http://www.gdb.org.
3.4.2 FISH mapping of critical YACs

The initial linkage report of PPCD defined a 30 cM interval on the long arm of chromosome 20 based on the location of markers on the maps available at that time (Heon et al., 1995). Linkage analysis in the CHED1 pedigree mapped the locus for the disease to an overlapping, but smaller, region on chromosome 20p. In order to demonstrate that the CHED1 locus was on the short arm of chromosome 20, YAC 875B10 and YAC 972H6, that were positive for the proximal flanking marker were used for FISH analysis. DNA from the clones was extracted, labelled and hybridised to chromosome spreads (section 2.5.1.3 and 2.5.9.5). The distal markers were firmly anchored on 20p11.2 therefore clones positive for these microsatellite markers were not tested. FISH experiments were performed by M.Fox, UCL.

The FISH result confirmed that 972H6 is a chimaeric clone with a positive signal on chromosome 20p11.2 and 4q (Figure 3.13). Similarly YAC 875B10 gave a positive signal on 20p11.2 (Figure 3.13).

Therefore, the CHED1 locus was localised entirely to the short arm of chromosome 20 at 20p11.2.

Figure 3.13 Localisation of two YACs (A) 875B10 and (B) 972H6 to the CHED1 disease region to chromosome 20p11.2 using FISH.
3.4.3 Isolation and characterisation of PAC clones within the CHED1 interval

Although a partial YAC contig of the disease interval was created (Figure 3.12), to overcome the problems of chimaerism and to create a redundant and accurate copy of 20p11.2, the RPCI1PAC library was screened. Initially, radiolabelling an STS probe followed by hybridisation to the gridded PAC library (section 2.5.9) was carried out, subsequently the PAC library was screened by PCR (section 2.5.6.2).

An STS, WI-9743, with reported homology to members of the collagen family was considered a potential candidate gene (Figure 3.11, http://www-genome.wi.mit.edu). Therefore, the PAC library was screened with a hybridisation probe, using a representative 201bp PCR fragment for WI-9743. Five positive PAC clones 183B12, 181D15, 251C16, 191I10 and 124H3 were identified, clone 251C16 is shown in Figure 3.14. STS content mapping of these positive clones, identified through hybridisation, was then verified by PCR. Subsequently, WI-9743 was found to be spuriously annotated and does not represent an expressed sequence. However, this STS provided an important anchor for physical mapping of the CHED1 interval.

![Figure 3.14 Autoradiograph of radiolabelled probe WI-9743 hybridised to the PAC library. Positive signal shown is plate 5 on membrane PAC6 of the library that corresponds to plate 251, row C, column 16 (PAC 251C16).](image-url)
As YAC 972H6 was shown to be chimaeric (Figure 3.13), PAC libraries were screened to identify non-chimaeric clones with the proximal flanking marker D20S471. The PCR product for D20S471 is a tetranucleotide repeat therefore, a probe generated by PCR could not be used for hybridisation because of the repetitive DNA content. Instead a single primer was synthesized of 36 bp, end-labelled (section 2.5.5.4.2) and used to probe the PAC filters (section 2.5.9.3) No positive clones were identified therefore the PAC library was screened using PCR pools (A-U) (section 2.5.6.2). Initial screening identified four positive PAC pools B, E, F and M (Figure 3.15). Fifteen secondary pools were screened for each positive pool, plates 16-30, 61-75, 76-90 and 184 –198 respectively. Only plate 196 from pool M was positive (Figure 3.15b) and analysis of this plate showed that row D and column 8 were positive identifying PAC clone 196 D8.

![Figure 3.15 Isolation of a PAC using the PCR pools for the RPCI1 PAC library and the marker D20S471. The PCR products shown are for (A) pools A-U with pools B, E, F and M giving a positive signal and Co, the positive human genomic DNA control and W the negative control (B) plates 184-198 from pool M with plate 196 giving a positive signal.](image-url)
In order to determine whether PAC clone 196D8 was chimaeric, DNA was extracted, labelled and hybridised to metaphase spreads (sections 2.5.1.2 and 2.5.9.5). PAC 196D8 mapped to chromosome 20p11.2 with no detectable chimaerism (Figure 3.16).

![Figure 3.16 FISH of PAC196D8 to metaphase spreads](with permission from M.Fox). The arrow indicates localisation to chromosome 20p11.2 only.

PAC clone 196D8 was therefore the most reliable proximal clone from the disease interval. Inter-Alu PCR of the clone was performed in order to identify a novel STS that could subsequently be used to identify overlapping clones (chromosome walking) by repeated screening of the PAC library.

![Figure 3.17 ALU- PCR of PAC196D8](A unique PCR product (shown by the arrow) was subcloned and sequenced.)
Inter-Alu PCR was performed using Ale1 and Ale3 primers (section 2.5.7) for chromosome walking. A unique PCR product (Figure 3.17) was obtained which was subsequently extracted from the gel and sequenced (Figure 3.18, section 2.5.8).

As part of a collaboration with the Sanger Institute, DNA from clone 196D8 was labelled and hybridised to a number of PAC clones which had previously been mapped to chromosome 20. PAC clones 999L4 and 122P22 were identified and were ultimately incorporated into the minimum sequence tile path (Figure 3.19).

![Sequence of the sub-cloned PCR product obtained from Alu-PCR of 196D8](image)

The novel sequence obtained from the Alu-PCR product of 196D8 (Figure 3.18) also aligned to the finished sequence of PAC 999L4 (Accession number: AL121761 from 28633bp-28753bp) demonstrating that the two clones are contiguous (Figure 3.19).
3.5 Integration of physical and genetic maps for the CHED1 interval

Using a combination of methods previously described, the first priority was to order the microsatellite markers within the region. The distal markers D20S118, D20S112 and D20S48 could not be assigned a genetic distance with respect to each other in the recombinant individuals IV:5, V:3 and V:6 (Figure 3.6). The YAC physical map did not resolve the order of these markers (Figure 3.12) however, the PAC contig did resolve the marker order (Figure 3.19). The PACs were shown to be contiguous through fingerprinting (data from Sanger Institute) and STS content mapping confirmed the final order (Figure 3.19). D20S48 is therefore the distal flanking marker for the CHED1 locus.

3.6 Genetic refinement

In order to refine the CHED1 locus, novel microsatellites need to be identified within the 2.7 cM interval between the flanking markers D20S48 and D20S471. Unfortunately, genetic maps were unable to resolve distances between markers in this interval. Data retrieved from Genethon suggested that the markers D20S605, D20S161 and D20S868 were localised to 20p11.2. Physical mapping of these microsatellite markers resulted in the accurate positioning of these markers relative to each other and the flanking markers (Figure 3.19 and section 3.3.2.4).

Since the preliminary localisation of CHED1 to the centromeric region of chromosome 20, additional microsatellite markers from the region were subsequently analysed. Although this analysis successfully identified linked markers supporting the linkage to 20p11.2, the critical region was not significantly refined (section 3.3.2.4).

3.7 Identification of candidate genes

The region on chromosome 20p11.2 that spans the CHED1 disease interval is relatively gene poor (Deloukas et al., 2001). An initial literature search resulted in the identification of several genes; Thrombomodulin (THBD), Paired Box gene1 (PAX1), Somatostatin Receptor (SSTR), PCSK2 (Proprotein Convertase, Subtilisin/Kexin-Type 2) (PCSK2) and brain glycogen phosphorylase (PYGB;) which mapped to chromosome 20p11.2 (Schnittger et al., 1992; Yasuda et al., 1993).
Figure 3.19 Minimal tiling path and sequence ready map of the CHED1 interval adapted from the Sanger Institute data. The dark green arrow is placed between the flanking markers D20S114 and D20S471 and represents the disease interval for CHED1.
The following order was established: 20pter–PCSK2–SSTR4–THBD–PYGB—cen, with PCSK2 the only gene that localised to the disease interval (Figure 3.19). An intragenic marker for PCSK2 (GDB) was analysed in the family but no recombinant individuals were observed.

### 3.7.1 Cystatin Genes

The cystatins are a family of cysteine proteinase inhibitors that work by forming tight reversible complexes that block the proteinase active site. They have been found mainly in body fluids like cerebrospinal fluid, saliva, milk and tears. This protein family consists of 3 types, with type II genes located on chromosome 20p11.2 (Dickinson et al., 1994; Thiesse et al., 1994). There are nine cystatin genes, two are pseudogenes and one denoted S-type is highly expressed in tears, saliva and seminal fluid. A putative function of these genes is in the remodeling of the extracellular matrix. Therefore, the cystatins were considered candidates for corneal dystrophies.

As part of a collaboration with Dr D. Dickinson (Houston, Texas), who had characterised the cystatin genes, 4 YAC clones, 3 cosmid clones and 14 bacteriophage \( \lambda \) clones spanning the gene cluster were obtained (Figure 20).

---

**Figure 3.20 Organisation of the family of cystatin genes on chromosome 20p.**

STS content mapping, using PCR, of the flanking polymorphic markers CHED1as well as STSs from the disease interval was performed. None of the markers were positive for the cystatin clones screened, although the genomic DNA controls were positive (data...
not shown). This implied that the cystatin genes mapped outside the disease interval, although the possibility remained that the markers were not represented on the clones tested. Subsequent sequence data from the Sanger Institute placed the cluster of cystatin genes centromeric to the disease interval on PAC clones dJ322G13, dJ333B15, dJ568C11 and dJ850N15.

3.7.2 VSX1

VSX1 or RINX is a novel human paired-like homeodomain transcription factor that was localised to 20p11.2 (Hayashi et al., 2000; Semina et al., 2000). Although initially isolated from a human embryonic craniofacial cDNA library, VSX1 transcripts were also identified in human adult retinal and corneal cDNA.

The gene spans 6.6kb of genomic DNA and consists of five exons. VSX1 is closely related to Chx10, which has been implicated in ocular disease (Ferda Percin et al., 2000), and they are both highly expressed in the retinal inner nuclear layer (Chow et al., 2001).

A primer pair, which amplified a fragment of VSX1 was kindly donated by J.Murray (Iowa). The amplifier was used as an EST on the YAC/PAC contig spanning the disease interval, however no positive signal was obtained from any clone although the human genomic control was positive (data not shown). This placed the gene outside the critical interval for CHED1. Subsequently, Hayashi and colleagues showed that VSX1 mapped to the region on chromosome 20p11.2 in the proximity of thrombomodulin and the cluster of cystatin genes. More recently, Heon and colleagues showed that mutations in VSX1 were responsible for causing posterior polymorphous dystrophy, a related corneal endothelial dystrophy and keratoconus (Heon et al., 2002).

3.7.3 Destrin

Actin plays a significant role in the development of the corneal endothelium and its basement membrane, together with the extracellular matrix glycoproteins laminin and fibronectin. Actin has been implicated in the migration of corneal epithelial and endothelial cells as it is associated with cell motility, morphology and adhesion (Moriyama et al., 1990). Gelsolin is an actin modulating protein that has been mapped to chromosome 9q34 and mutations in this gene are responsible for Lattice Corneal
Dystrophy Type II (LCD II) (section 1.3.2.3). Mutations in this gene (Asp187Asn or G654A) result in the deposition of amyloid in the cornea and plaques in the brain. Although there is no nucleotide sequence homology between gelsolin and destrin, it has been shown that there is protein homology between segment 1 of gelsolin and destrin using NMR to deduce tertiary structure (Hatanaka et al., 1996). Given the important role of actin in the corneal endothelium, the mapping of destrin to 20p11.2 (bA462D18 in Figure 1.18) and the fact that a member of the same family of proteins causes a corneal dystrophy, destrin was considered a good candidate gene to investigate.

### 3.7.3.1 Destrin expression in the cornea

Destrin is widely expressed, including the corneal endothelium in rabbits (Fujimaki et al., 1999). To confirm corneal expression, the primer pairs Des F and Des R were used to perform RT-PCR (section 2.5.5.5) using RNA from human retina as well as RNA from human primary cells and cell lines derived from corneal epithelium, stroma and endothelium (section 2.5.12 and Table 3.9).

Figure 3.20 shows that destrin expression was detected at similar levels in the corneal epithelium, stroma and endothelium.

<table>
<thead>
<tr>
<th>Forward Primers</th>
<th>Reverse Primer</th>
<th>Size bp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Des F ATGGCCCTCAGGAGTGCAAGTA</td>
<td>Des R CTACACAGGGCAGTCTCTCAAG</td>
<td>570</td>
</tr>
<tr>
<td>166F GTTGGGATGTTGGTGAAC</td>
<td>166R GTTACACCAACATCTCCAA</td>
<td></td>
</tr>
</tbody>
</table>

### Table 3.9 Primers designed to amplify destrin cDNA.
The expected product size was 570bp and primers 166F and 166R were used as internal sequencing primers.

### 3.7.3.2 Amplification of destrin from patient RNA and mutation screening

The human cDNA sequence of destrin was obtained from Genbank (S65738). As the coding region is relatively small (497 bp), primers were designed to amplify the entire coding region from cDNA (DesF–DesR) (Table 3.9). A lymphoblast cell line established from one affected individual and an unaffected individual from the CHED1 pedigree through ECACC was used as patient source material (section 2.2). The cell lines were grown to confluency (section 2.5.12) and total RNA extracted from the cells (section
2.5.1.4.1). Reverse transcriptase was used to make first strand cDNA (section 2.5.5.5) which was subsequently amplified using destrin primers (Table 3.9). The PCR products were sequenced (section 2.5.8) and no differences were detected between the affected and unaffected individuals.

Figure 3.21 The expression pattern of destrin in the corneal epithelium, stroma and endothelium using RT-PCR.
Lanes 1-3 RNA from primary cells of the cornea, lanes 4-6 indicates RNA from corneal cell lines, Ep epithelium, St stroma and En endothelium. PGM1 and β Actin were controls.

3.7.2.3 Genomic structure of destrin

The genomic sequence of destrin spans 38kb on BAC 462D18 (Figure 3.22) with the coding region of the gene predicted to consist of four exons. Exon 1 is 129bp with 126 bp of 5’ untranslated sequence and the ATG start codon.
Exon 2 is 308 bp, while exon 3 is 77 bp. The fourth exon spans 1205 bp however, only 110bp are coding, while the remaining sequence is the 3’ untranslated sequence.

Figure 3.22 Schematic diagram of the genomic organisation of the destrin gene.

<table>
<thead>
<tr>
<th>Exon(bp)</th>
<th>Forward primer</th>
<th>Reverse primer</th>
<th>Size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exon 1</td>
<td>aaggtctgcatactcgctg</td>
<td>gatggatttcacagagcgg</td>
<td>408</td>
</tr>
<tr>
<td>Exon 2A</td>
<td>gttgaatgtttccctaccctg</td>
<td>gcatcataaagcataagcgac</td>
<td>490</td>
</tr>
<tr>
<td>Exon 2B</td>
<td>aagttcgtaaatgctccacacc</td>
<td>ctcccgagtgcttaacaaaaatg</td>
<td>314</td>
</tr>
<tr>
<td>Exon 3</td>
<td>ctggtcaacctgagtgattctg</td>
<td>aagggcttctcagaaaaaggtg</td>
<td>367</td>
</tr>
<tr>
<td>Exon 4</td>
<td>gagaaagaggtaactgtgcctg</td>
<td>aaaaagatgacaagacagctccc</td>
<td>286</td>
</tr>
</tbody>
</table>

Table 3.10 Intronic primers designed to amplify the 4 exons of destrin.

Primers in the introns of the destrin gene were designed, PCR products amplified and sequenced. No differences in the cDNA sequence were detected between the two affected and one unaffected individuals from the CHED1 family (Table 3.10).
<table>
<thead>
<tr>
<th>Clone</th>
<th>Gene Description</th>
<th>Exons</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>dJ298M8.1</td>
<td>Zinc finger protein pseudogenes</td>
<td>1</td>
<td>Pseudogene</td>
</tr>
<tr>
<td>dJ298M8.2</td>
<td>PCSK2 Proprotein convertase subtilisin/kexin type 2</td>
<td>12</td>
<td>Known</td>
</tr>
<tr>
<td>dJ531H16.2</td>
<td>Protein 91/23 (mouse dynein light chain, TCTEX-1 like)</td>
<td>1</td>
<td>Pseudogene</td>
</tr>
<tr>
<td>dJ531H16.3</td>
<td>BFSP1 (Filensin) Beaded filament structural protein</td>
<td>8</td>
<td>Known</td>
</tr>
<tr>
<td>dJ531H16.4</td>
<td>Ubiquitin-40S ribosomal protein</td>
<td>1</td>
<td>Pseudogene</td>
</tr>
<tr>
<td>bA462D18.1</td>
<td>RAC1 (small GTP binding protein Rac1)</td>
<td>1</td>
<td>Pseudogene</td>
</tr>
<tr>
<td>bA462D18.2</td>
<td>Destrin (actin depolymerising factor ADF) (ACTDP)</td>
<td>4</td>
<td>Known</td>
</tr>
<tr>
<td>bA462D18.3</td>
<td>RBP1 ribosome binding protein 1 (dog homolog)</td>
<td>25</td>
<td>Known</td>
</tr>
<tr>
<td>dJ803K15.1</td>
<td>protein similar to barrier-to-autointegration factor</td>
<td>3</td>
<td>Novel CDS</td>
</tr>
<tr>
<td>bA504H3.1</td>
<td>SNX5 (sorting nexin 5)</td>
<td>14</td>
<td>Known</td>
</tr>
<tr>
<td>bA504H3.2</td>
<td>PTMA (prothymosin alpha) pseudogene</td>
<td>1</td>
<td>Pseudogene</td>
</tr>
<tr>
<td>bA504H3.3</td>
<td>Novel zinc finger protein similar to Drosophila Ovo</td>
<td>4</td>
<td>Novel CDS</td>
</tr>
<tr>
<td>bA504H3.4</td>
<td>Novel protein</td>
<td>5</td>
<td>Novel CDS</td>
</tr>
<tr>
<td>dJ726N1.2</td>
<td>Putative novel transcript</td>
<td>2</td>
<td>Novel CDS</td>
</tr>
<tr>
<td>dJ726N1.3</td>
<td>60S ribosomal protein L15 (RPL15) pseudogene</td>
<td>1</td>
<td>Pseudogene</td>
</tr>
<tr>
<td>dJ717M23.1</td>
<td>CRP2 binding protein (CSRP2BP)</td>
<td>11</td>
<td>Known</td>
</tr>
<tr>
<td>dJ717M23.2</td>
<td>Novel pseudogenes</td>
<td>1</td>
<td>Pseudogene</td>
</tr>
<tr>
<td>dJ568F9.1</td>
<td>ZNF133 (Zinc finger protein 133 [clone pHZ-13])</td>
<td>5</td>
<td>Known</td>
</tr>
<tr>
<td>dJ568F9.2</td>
<td>Novel protein (FLJ10600)</td>
<td>16</td>
<td>Known</td>
</tr>
<tr>
<td>bA189K21.2</td>
<td>Gluocamyl (N-acetyl) transferase 1 (GCNT1)</td>
<td>1</td>
<td>Pseudogene</td>
</tr>
<tr>
<td>bA189K21.3</td>
<td>RNA polymerase III subunit RPC39 (POLR3F)</td>
<td>9</td>
<td>Known</td>
</tr>
<tr>
<td>bA189K21.4</td>
<td>60S ribosomal protein L21 (RPL21) pseudogene</td>
<td>2</td>
<td>Pseudogene</td>
</tr>
<tr>
<td>bA189K21.5</td>
<td>retinoblastoma binding protein (RBBP9)</td>
<td>5</td>
<td>Novel CDS</td>
</tr>
<tr>
<td>bA189K21.6</td>
<td>Sec23 (S. cerevisiae) homolog B</td>
<td>20</td>
<td>Known</td>
</tr>
<tr>
<td>bA189K21.7</td>
<td>40S ribosomal protein S19 (RPS19) pseudogene</td>
<td>1</td>
<td>Pseudogene</td>
</tr>
<tr>
<td>bA189K21.8</td>
<td>Novel protein</td>
<td>3</td>
<td>Novel CDS</td>
</tr>
<tr>
<td>bA379J5.2</td>
<td>Putative novel transcript</td>
<td>2</td>
<td>Putative</td>
</tr>
<tr>
<td>bA379J5.4</td>
<td>Homeobox protein pseudogenes</td>
<td>2</td>
<td>Pseudogene</td>
</tr>
<tr>
<td>bA555E18.1</td>
<td>Novel protein - bacterial histidyl-tRNA synthetase, 6</td>
<td>Novel CDS</td>
<td></td>
</tr>
<tr>
<td>dJ1068E13.1</td>
<td>Putative novel protein</td>
<td>4</td>
<td>Novel CDS</td>
</tr>
<tr>
<td>dJ1068E13.2</td>
<td>Novel protein -bovine SCP2 (Sterol carrier protein 2)</td>
<td>1</td>
<td>Novel CDS</td>
</tr>
<tr>
<td>dJ1068E13.3</td>
<td>Novel transcript</td>
<td>5</td>
<td>Novel</td>
</tr>
<tr>
<td>dJ1068E13.4</td>
<td>EEF1A1 (eukaryotic translation elongation factor1 alpha1)</td>
<td>1</td>
<td>Pseudogene</td>
</tr>
<tr>
<td>dJ1027G4.1</td>
<td>Slc24A3</td>
<td>17</td>
<td>Known</td>
</tr>
<tr>
<td>dJ999L4.1</td>
<td>Ribosomal protein L12 (RPL12)</td>
<td>1</td>
<td>Novel CDS</td>
</tr>
</tbody>
</table>

Table 3.11 Genes in the CHED1 region derived from genomic sequence predictions.
3.8 Discussion

Autosomal dominant congenital hereditary endothelial dystrophy (CHED1) is a rare dystrophy in which there is atrophy of the endothelium which results in corneal oedema. The locus for CHED1 was initially localised to a 14 cM centromeric interval on chromosome 20 which was subsequently refined to a 2.7 cM region on 20p11.2 corresponding to a physical distance of 2.73 Mb which is represented by 28 PAC and BAC clones that have been sequenced. At the last annotation (15/2/2002) there were predicted to be 11 genes of known function, 1 putative transcript, 11 novel genes and 12 pseudogenes (Table 3.11 and Figure 3.19).

In total, there are 196 exons predicted within the CHED1 disease interval and one strategy is to amplify all the exons in two affected and one unaffected individual from the family and test for mutations. However, based on published studies mapping novel cDNAs to finished and annotated chromosomal sequences many genes will be either underrepresented or not predicted (Wiemann et al., 2001). Therefore, it is likely that the 196 exons do not accurately represent all the coding sequence within the CHED1 critical interval.

For a positional candidate strategy of disease gene identification, it is essential to define a minimum genetic interval. Novel microsatellite markers were designed in an attempt to refine the disease interval from 2.7 cM. Unfortunately, the new markers did not refine the interval, however further microsatellites and SNPs should be typed in order to reduce the number of genes, which needs to be tested for mutations.

3.8.1 Candidate genes

Specific criteria for determining candidate genes can be based on protein function (e.g. a member of a protein family involved in fluid transport) or gene expression in the cornea. For CHED1, gene expression in the corneal endothelium would indicate a potential role at the site of pathology.

A number of genes, which map to the critical interval for CHED1 (Table 3.11), could be considered good candidates, and some of these are discussed.
3.8.1.1 Filensin or beaded filament structural protein 1, (BFSP1)

The intermediate filaments (IF) are a diverse group of proteins that are abundantly expressed in most cells of the body. In the eye, the IF proteins can form two distinct polymers, 10 nm IF as well as a beaded filament. In the lens, two major beaded intermediate filaments proteins, filensin and phakinin have been identified. Filensin (BFSP1) was localised to chromosome 20p12-11.2 (Hess et al., 1995).

Rendtorff and colleagues (1998) used 8 overlapping YAC clones and showed that filensin mapped between markers D20S112 and D20S605 (Rendtorff et al., 1998). PAC clone 531H16 from the Sanger Institute was later sequenced and includes the genomic sequence for filensin (Figure 3.19). On the basis that filensin maps within the critical interval and is expressed in the eye, this gene represents a candidate for CHED1.

3.8.1.2 Solute carrier 24A3 (SLC 24A3)

The maintenance of the hydration of cornea is crucial as the stroma can absorb fluid that would result in corneal oedema and a loss of transparency (section 1.2.4). The transport of liquid is secondary to the transport of ions across the endothelium and depends on Na⁺K⁺ATPase activity (Rae et al., 1996; Hara et al., 1999; Rae et al., 2000). Sodium/calcium (-potassium) exchangers (NCX and NCKX), now designated as solute carriers (SLC), are critical for the rapid extrusion of calcium. Mammalian NCKX3 (SLC24A3) has five membrane-spanning domains in the NH2 terminus separated from the sixth membrane domain at the COOH-terminal end by a large intracellular loop.

The SLC24A3 gene that encodes NCKX3 maps to 20p11.2, spans 510254 bp and six clones 542H21, 718D20, 97N19, 110K14, 1027G4 and 122P22 (Figure 3.18). Kraev et al (2001) predicted a start site before two potential glycosylation sites as being the true start site since the protein is slightly glycosylated. They determined tissue expression using a Multiple Tissue Expression Array from Clontech, which showed that SLC24A3 is ubiquitously expressed with higher levels of expression seen in the brain. As SLC24A3 maps within the critical interval and has an important function of maintaining homeostasis it is considered a candidate gene for CHED1.
3.8.1.3 CRP2BP

A novel gene from PAC dJ717M23 was identified and the genomic organisation was characterised using a bioinformatics approach, which resulted in the prediction of a gene, CRP2BP, with 10 exons but no functional homology to other proteins (Figure 3.19). It was recently shown that CRP2BP, is the binding partner of CRP2 (cysteine rich protein family2) which has a double LIM domain (Weiskirchen et al., 2000).

Another LIM domain encoding gene, LMX1B, on 9q32-q34 has recently been shown to cause nail-patella syndrome and possibly open angle glaucoma (Dreyer et al., 1998; Vollrath et al., 1998). It has also been shown that LMX1B is expressed in the anterior segment of the eye and that mutations in this gene can affect the expression of keratocan, which causes cornea plana (Pressman et al., 2000). Interestingly, there was a case report of an individual having the characteristic changes of the cornea usually associated with CHED also having nail hypoplasia (Stirling et al., 1994). Although ubiquitously expressed, CRP2BP is an interesting candidate gene for CHED1.

As a priority for future candidate gene analysis, human corneal RNA extracted from corneal endothelium derived cells should be tested for expression of all genes within the region (Table 3.8).

3.8.2 Synteny between human chromosome 20 and mouse chromosome 2.

The telomeric end of the long arm of mouse chromosome 2 shares a remarkable degree of synteny with the whole of human chromosome 20 as shown in Figure 3.23.

The region of homology encompasses all of human chromosome 20 to the telomeric region on the long arm of mouse chromosome 2 (Table 3.12). This region on mouse chromosome 2 was mapped using markers, which determined the order of genes on the chromosome.
Figure 3.23 Homology between mouse chromosome 2 and human chromosomes 1, 2, 7, 9, 10, 11, 12, 15 and 20.
The length of the bar indicates the number of human/mouse orthologies identified. The figure was adapted from:

Subsequently, the human orthologs of these genes were shown to reside on human chromosome 20, in much the same order as that found on the mouse chromosome 2 (Schnittger et al., 1992; Siracusa et al., 1994; Zuberi et al., 1996). Genes and disease mapping to this syntenic region on mouse chromosome 2 could therefore be considered candidates for CHED1 (Table 3.12).

A mouse coloboma mutant was described with small, sunken eyes as well as having behavioral defects, this was deletion mapped to the syntenic region of human 20p11.2 (Hess et al., 1994). However, further studies showed that the deletion did not extend to Pskk2 and was therefore unlikely to lie within the disease interval of CHED1 (Hess et al., 1995; Hess et al., 1996). Another mouse mutant on chromosome 2 with an eye phenotype
is the Corn1 mouse. This mouse, is a model for corneal neo-vascularisation which was mapped to a syntenic region on human chromosome 20q and although this locus overlaps the region of PPCD, it does not overlap the CHED1 locus (Smith et al., 1996).

<table>
<thead>
<tr>
<th>Mouse Gene symbol</th>
<th>Mouse Locus(cM)</th>
<th>Human Gene symbol</th>
<th>Human Locus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avp</td>
<td>73.2</td>
<td>AVP</td>
<td>20p13</td>
</tr>
<tr>
<td>Oxt</td>
<td>73.2</td>
<td>OXT</td>
<td>20p13</td>
</tr>
<tr>
<td>Snrbp</td>
<td>73.5</td>
<td>SNRBP</td>
<td>20p12</td>
</tr>
<tr>
<td>Pdyn</td>
<td>73.5</td>
<td>PDYN</td>
<td>20pter-p12.2</td>
</tr>
<tr>
<td>Ptpra</td>
<td>74.0</td>
<td>PTPRA</td>
<td>20p13</td>
</tr>
<tr>
<td>Itp</td>
<td>74.1</td>
<td>ITPA</td>
<td>20p</td>
</tr>
<tr>
<td>Cenpb</td>
<td>74.2</td>
<td>CENPB</td>
<td>20p13</td>
</tr>
<tr>
<td>Sn</td>
<td>75.0</td>
<td>SN</td>
<td>20p13</td>
</tr>
<tr>
<td>Prnp</td>
<td>75.2</td>
<td>PRNP</td>
<td>20pter-p12</td>
</tr>
<tr>
<td>Pcna</td>
<td>75.2</td>
<td>PCNA</td>
<td>20pter-p12</td>
</tr>
<tr>
<td>Chgb</td>
<td>75.6</td>
<td>CHGB</td>
<td>20pter-p12</td>
</tr>
<tr>
<td>Bmp2</td>
<td>76.1</td>
<td>BMP2</td>
<td>20p12</td>
</tr>
<tr>
<td>Plcb1</td>
<td>76.7</td>
<td>PLCB1</td>
<td>20</td>
</tr>
<tr>
<td>Plcb4</td>
<td>77.0</td>
<td>PLCB4</td>
<td>20</td>
</tr>
<tr>
<td>Pcsk2</td>
<td>81.4</td>
<td>PCSK2</td>
<td>20p11.2</td>
</tr>
<tr>
<td>Pax1</td>
<td>82.0</td>
<td>PAX1</td>
<td>20p11.2</td>
</tr>
<tr>
<td>Thbd</td>
<td>84.0</td>
<td>THBD</td>
<td>20p11.2</td>
</tr>
<tr>
<td>Pycgb</td>
<td>84.0</td>
<td>PYGB</td>
<td>20p11.2</td>
</tr>
<tr>
<td>Cst3</td>
<td>84.0</td>
<td>CST3</td>
<td>20p11.2</td>
</tr>
</tbody>
</table>

Table 3.12 Homology which exists between mouse chromosome 2 and the short arm of human chromosome 20.

Currently there are no obvious mouse mutant phenotypes that are good models for the CHED1 phenotype.

The sequencing and characterisation of the genes on mouse chromosome 2 may aid in the identification of the gene for CHED1.
CHAPTER 4

Congenital Hereditary Endothelial Dystrophy CHED2

4.1 Introduction

Congenital Hereditary Endothelial Dystrophy (CHED) can be inherited either as an autosomal dominant disease (CHED1) or as a more severe autosomal recessive form (CHED2) in which the corneal endothelium is the major site of pathology (section 1.3.3.2.). CHED is a very rare condition with more recessive CHED families reported than dominant (Sajjadi et al., 1995; Ehlers et al., 1998).

Callaghan and colleagues (1999) reported a large 5 generation consanguineous Irish family with CHED2. Homozygosity mapping and linkage analysis showed that CHED2 was genetically distinct from CHED1 and PPCD. Similarly, Kanis and colleagues (1999) using a large consanguineous autosomal recessive Saudi Arabian pedigree reported exclusion from the CHED1/PPCD locus on 20p11.2. Subsequently, Hand and colleagues (1999) showed that the locus for CHED2 lies on the short arm of chromosome 20 within an 8 cM interval flanked by the markers D20S113 and D20S882. More recently, a small Pakistani pedigree with CHED2 was linked to an overlapping region on 20p13 between the flanking markers D20S906 and D20S916, spanning 9.6 cM (Mohamed et al., 2001).

The collaboration with the Sanger Institute to identify the gene for CHED1 was extended to include the identification of the gene for CHED2. The CHED2 locus lies in an 8-9.6 cM interval on 20p13 (Figure 4.1). Although this is a very large genetic interval, the physical distance is relatively small, estimated to be 3.7 Mb. The CHED2 disease interval is spanned by 41 PAC and BAC clones, which have been sequenced by the chromosome 20 group at the Sanger Institute. The disease interval for CHED2 lies in a relatively gene-rich region (Deloukas et al., 2001). At the last annotation (15/2/2002) there were predicted to be 37 genes of known function, 25 putative or novel gene
transcripts and 10 pseudogenes. A positional candidate approach was adopted to identify the gene for CHED2.

Figure 4.1. An ideogram showing the CHED2 locus on chromosome 20p13. The flanking markers D20S906 and D20S916 define an interval of 9.6 cM (Mohamed et al., 2001) and D20S113 and D20S882 an interval of 8 cM (Hand et al., 1999).
4.2. Pedigree and clinical data

The family used in this study was a consanguineous pedigree originating from Pakistan that was obtained through Dr Mohamed (St. James Hospital, Leeds). The parents are first cousins and neither they nor their other relatives are affected, however, three of their 6 children were affected (Figure 4.3).

![Figure 4.2. The opaque cornea of an affected member of the CHED2 family.](image)

The three affected girls all had severe non-progressive corneal opacities from birth with a typical “ground glass” appearance and nystagmus (Figure 4.2). One of the affected children underwent penetrating keratoplasty (corneal graft) and histology of the diseased cornea was consistent with a diagnosis of CHED as there was an absence of endothelial cells, a thickening of Descemet’s membrane and the stroma was oedematous.

In addition to this family, a small panel of patients with CHED (8), Fuch’s (10) and PPCD (7) was used for screening candidate genes.
4.3 Linkage analysis of the CHED2 pedigree

Microsatellite markers spanning the locus for CHED1 were analysed and linkage was excluded to this locus. However, linkage was confirmed in this family using the markers D20S906 and D20S916 that span the CHED2 locus on chromosome 20p13 (Figure 4.1). Initial linkage analysis on this family, was performed by Mohamed and colleagues and a maximum lod score of 2.64 was achieved using multipoint linkage analysis.

In order to refine the disease interval, two novel microsatellite markers were identified from the clones dJ686C3 and dJ189G13 which lie within the contig spanning the CHED2 locus (Figure 4.4). The PAC and BAC contig was analysed using bioinformatics (section 2.5.13) and CA repeats were identified in clones dJ686C3 which had 27 CA repeats and dJ189G13 which has 22 CA repeats. The primers and conditions used to amplify these microsatellite markers in the family are shown in Table 4.1. These markers were amplified (section 2.5.5), radioactively labelled (2.5.5.4) and resolved on an acrylamide gel (2.5.4.2).

In this family, these markers proved to be uninformative and further refinement of this locus was not achieved.
Table 4.1 Primer sequences used to amplify novel microsatellite markers.

<table>
<thead>
<tr>
<th>Clone</th>
<th>Primer sequence (Forward)</th>
<th>Primer sequence (Reverse)</th>
<th>Ta °C</th>
<th>Amplimer (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>dJ686C3</td>
<td>gacagtatgcagaactttgg</td>
<td>aagctcatgactgacaactc</td>
<td>58</td>
<td>460</td>
</tr>
<tr>
<td>dJ189G13</td>
<td>atctgcaatgttccttgacac</td>
<td>tgcataatgtgacatcacc</td>
<td>60</td>
<td>390</td>
</tr>
</tbody>
</table>

4.4 Positional candidate gene analysis within the CHED2 disease interval

The PAC/BAC clones spanning the interval have been sequenced and bioinformatic techniques were used to predict the presence of transcripts and genes (sections 1.13 and 2.5.13). The putative genes and transcripts identified using these methods are shown in Table 4.2 and Figure 4.4 and were considered as potential candidate genes for CHED2. Of the 62 predicted genes or transcripts annotated (Figure 4.4), 25 of these are novel and have no known homology and a further 10 are annotated as potential pseudogenes. Detailed examination of the potential candidate genes was based on predicted function and expression profiles in order to prioritise disease gene screening and characterisation.
Table 4.2 Putative genes and transcripts annotated within the CHED2 critical interval.
Pseudogenes and novel transcripts with no known homology are not included in this table.
Figure 4.4. Schematic of the CHED 2 sequence tile path.
PAC/BAC clones form the tiling path (yellow). Microsatellite markers (green), predicted genes and transcripts (dark blue), novel CDS (light blue) and novel transcripts (orange) are shown. Green arrows indicate pseudogenes and candidate genes are underlined in red. The red arrow demarcates the critical interval for CHED2. The clones shown in purple and green represent the 686C3 and 189G13. Adapted from Sanger Institute data.
4.4.1 CDP-diacylglycerol synthase 2 (CDS2)

One of the research projects within the department focused on identifying the human orthologues of *Drosophila* genes expressed in the eye. The *Drosophila eye-cds* gene has two human orthologues, one of which, is CDP-diacylglycerol synthase 2 (CDS2) which maps to chromosome 20p13. CDS2 encodes an enzyme that catalyses the formation of diacylglycerol from phosphatidic acid and is thought to be a key regulator of the regeneration of the signalling molecule phosphatidylinositol 4,5-bisphosphate (PIP$_2$) (Halford *et al.*, 1998). *In situ* hybridisation using the murine orthologue, Cds2, was carried out on cross sections of adult mouse eye to determine the expression pattern of the gene (Figure 4.5, section 2.5.9.5).

![Figure 4.5. In situ hybridisation analysis of Cds2 in adult mouse cornea (with kind permission from S.Ingles).](image)

(A) haematoxylin and eosin (H&E) stained section (red arrows), (B) probed with the Cds2 sense probe, no expression was detected and (C) probed with the Cds2 anti-sense probe. High levels of expression were detected in the endothelium (En) and the epithelium (Ep) with no significant expression in the stroma (S) as indicated by the blue arrows.

Cds2 is expressed in the adult mouse retina (S.Ingles personal communication) however there is also a high level of expression within the cornea (Figure 4.5). The corneal section shows heavy staining of all layers of the epithelium as well as staining of
the endothelial cells. There is very little staining of the stroma, with low levels of expression in the keratocytes. No signal was detected with the control sense probe.

As the gene lies within the disease interval and the *in situ* hybridisation in mouse demonstrated Cds2 expression in the corneal epithelium as well as the endothelium, mutation detection was undertaken.

### 4.4.1.1 Genomic organisation of CDS2

CDS2 has 13 exons spanning 65 kb of genomic DNA (Figure 4.6). The open reading frame starts in exon 1 and the stop codon is in exon 13. Intron 1 is the largest intron at an estimated 47 kb (S. Halford personal communication).

![Figure 4.6 Schematic diagram of the genomic organisation of the CDS2 gene on chromosome 20p13.](image)

The 13 exons are depicted as boxes and slashed lines represent large regions of genomic DNA. The start and stop codons are also indicated. H, *Hind*III; B, *Bam*HI; R, *Eco*RI.

### 4.5 Mutation screening of CDS2

All 13 exons of CDS2 were amplified in two affected members and one unaffected member of the CHED2 Pakistani family as well as the sporadic panel, using the primer pairs listed in Table 4.3, at a Ta of 56°C. The PCR products were directly sequenced and compared to the known exon sequences (section 2.55 and 2.58). No changes were detected in the coding region of the gene, making it unlikely that CDS2 is responsible for the corneal dystrophy seen in this family.
### Table 4.3. Primer sequences used to amplify the 13 exons of CDS2 by PCR for mutation analysis.

<table>
<thead>
<tr>
<th>EXON</th>
<th>PRIMER SEQUENCE</th>
<th>AMPLIFIER</th>
</tr>
</thead>
</table>
| 1    | F 5' -GGCTGCTAAGGAATGTGAG-3'  
     | R 5' -AAGGGTCAGAGAACAAGACGAC-3' | 305bp |
| 2    | F 5' -ATGACAGGCATAGACTTCTC-3'  
     | R 5' -CCACTGCTAAGCACTAGATT-3' | 383bp |
| 3    | F 5' -GTAAGCTGGAAGGCTGCA-3'  
     | R 5' -AGTCAAGATGTTATCTAGGC-3' | 369bp |
| 4    | F 5' -ATTTTCTCACCTACGCTCA-3'  
     | R 5' -CTGCTAGGTCACCAGTTAT-3' | 286bp |
| 5    | F 5' -CTGCTGGGAATAATTAGCC-3'  
     | R 5' -CCTTCATGCTCTAGATGCAT-3' | 446bp |
| 6    | F 5' -ACTGCCCTATCTACCTGCA-3'  
     | R 5' -CTAGTTTCTCTGCTGCACT-3' | 355bp |
| 7    | F 5' -GACATGTTAGAAGATGGA-3'  
     | R 5' -AGAGGCTACCTGCTAGTG-3' | 366bp |
| 8    | F 5' -CATGAACCTCTGAGCATCA-3'  
     | R 5' -GAGCAATTCTCATCCTCACT-3' | 309bp |
| 9    | F 5' -GGCCACAATGAGGATGAAAT-3'  
     | R 5' -CTAAATTCAGTGCATGGGT-3' | 433bp |
| 10   | F 5' -TGTCCTAATATAGCCTACCC-3'  
     | R 5' -ATTAACAGAGGAGCAGGAC-3' | 429bp |
| 11   | F 5' -GTCTCACTCCTACCTGTTAAT-3'  
     | R 5' -GTCATTGTCTCAGTCACC-3' | 343bp |
| 12   | F 5' -GGCTGCTAAGGAATGTGAG-3'  
     | R 5' -AAGGGTCAGAGAACAAGACGAC-3' | 305bp |
| 13   | F 5' -ATGACAGGCATAGACTTCTC-3'  
     | R 5' -CCACTGCTAAGCACTAGATT-3' | 383bp |
| 14   | F 5' -GTAAGCTGGAAGGCTGCA-3'  
     | R 5' -AGTCAAGATGTTATCTAGGC-3' | 369bp |
| 15   | F 5' -ATTTTCTCACCTACGCTCA-3'  
     | R 5' -CTGCTAGGTCACCAGTTAT-3' | 286bp |
| 16   | F 5' -CTGCTGGGAATAATTAGCC-3'  
     | R 5' -CCTTCATGCTCTAGATGCAT-3' | 446bp |
| 17   | F 5' -ACTGCCCTATCTACCTGCA-3'  
     | R 5' -CTAGTTTCTCTGCTGCACT-3' | 355bp |
| 18   | F 5' -GACATGTTAGAAGATGGA-3'  
     | R 5' -AGAGGCTACCTGCTAGTG-3' | 366bp |
| 19   | F 5' -CATGAACCTCTGAGCATCA-3'  
     | R 5' -GAGCAATTCTCATCCTCACT-3' | 309bp |
| 20   | F 5' -GGCCACAATGAGGATGAAAT-3'  
     | R 5' -CTAAATTCAGTGCATGGGT-3' | 433bp |
| 21   | F 5' -TGTCCTAATATAGCCTACCC-3'  
     | R 5' -ATTAACAGAGGAGCAGGAC-3' | 429bp |
| 22   | F 5' -GTCTCACTCCTACCTGTTAAT-3'  
     | R 5' -GTCATTGTCTCAGTCACC-3' | 343bp |

4.6 Discussion

Autosomal recessive congenital hereditary endothelial dystrophy (CHED2) is a rare corneal disease in which there is a reduction in the number of endothelial cells, remaining cells become dystrophic and there is a thickening of Descemet’s membrane. The loss of endothelial cells results in a loss of function of the endothelial pump, which normally maintains homeostasis, and leads to an oedematous cornea. These pathological changes are seen at birth, together with the oedematous changes in the epithelium and stroma that lead to the opacification of the cornea.

A distinct locus for CHED2 was reported on chromosome 20p13 which spans 8 cM and corresponds to a physical distance of approximately 3.7 Mb (Callaghan et al., 1999;
Hand et al., 1999; Kanis et al., 1999; Mohamed et al., 2001). This telomeric region of chromosome 20 is in a Giemsa negative band that is gene rich. Sequence analysis of chromosome 20 predicts at least 62 genes or transcripts within the CHED2 interval on 20p13. This makes it a challenging task to identify the disease causing gene. Novel microsatellite markers were typed in the CHED2 family, the disease locus was not refined. It is possible that the localisation of the recombination events with respect to the disease gene has been fully achieved in this family. This is likely, due to the fact that this is a consanguineous family from a geographically isolated region, such that shared haplotypes are common.

4.6.1. Candidate genes
4.6.1.1. CDS2

CDS2 was assessed as a positional candidate gene, which is expressed in the corneal endothelium. The 13 exons of CDS2 were screened in the family and no differences were observed in the coding region of the gene between affected and unaffected individuals. Therefore it is unlikely that the CDS2 gene is responsible for the corneal dystrophy seen in this family.

The role of CDS2 in the cornea, given its high level of expression, remains to be determined. CDS2 is an enzyme involved in the regeneration of the signaling molecule PIP2. Phosphoinositide-mediated signaling pathways are ubiquitous and it is possible that one or more of these pathways may play an important role within the cornea. It has previously been demonstrated that several proteins with high levels of expression within the cornea, such as ALDH1 and ALDH3, play a dual role within the cornea. ALDH1 and ALDH3 function both as enzymes as well as structural proteins (corneal “crystallin”) (Jester et al., 1999). Since CDS2 is an enzyme that is also abundantly expressed in the cornea, it is possible that it also has additional functions in this tissue.

4.6.1.2 Transglutaminase

The transglutaminases are a family of proteins that help to stabilise protein assemblies by γ-glutamyl-ε-lysine cross-links. These enzymes catalyse a Ca2+ dependent reaction between γ-carboxyamide group of glutamine residues and the ε-amino group of lysine
residues (Candi et al., 1998). Seven members of this family have been identified and each has a specialised protein function or tissue expression. For instance, mutations in transglutaminase 1 can result in congenital recessive ichthyoses (Candi et al., 1998). This is a disorder of cornification that mainly affects the epidermis and hair and can vary from fine white scaling of the skin to the formation of large plate-like brown scales.

In addition to an enzymatic role in the cytosol there is evidence that tissue transglutaminase may play a role in the organisation of the extracellular matrix. It was shown that reduced expression of tissue transglutaminase in endothelial cells resulted in reduced cell spreading, cell adhesion and a decrease in the cross-linking of fibronectin. (Jones et al., 1997).

Two transglutaminase genes TGM3 and TGM3-L, map to the CHED2 interval on PAC clones dJ816K17 and dJ734P14 (Table 4.2 and Figure 4.4). If TGM3 or TGM-L is shown to be expressed in the cornea they represent good candidates for CHED2.

4.6.1.3 Solute carriers

There are two genes in the CHED2 disease interval, which have homology to members of the solute carrier family. SLC23A1 is a member of the sodium dependent vitamin C transporter family (also designated SVCT2) which is widely expressed in neuronal, neuroendocrine and endothelial cells. Expression of SLC23A1 was also detected in the cornea and retina and is believed to play an important role in the protection of the eye from radiation-induced damage. (Tsukaguchi et al., 1999)

The endothelial pump maintains homeostasis of the cornea with the use of sodium potassium exchangers and a bicarbonate pump. A predicted novel gene in the CHED2 region has homology to SLC4, which is an important part of the bicarbonate transport mechanism and as such is an excellent candidate for CHED2.

Before screening genes for mutations, expression in the cornea should be confirmed, as it is one method of prioritising the candidate genes. There are many candidates in this gene rich interval that may have implications in eye development, as structural proteins or be involved in ion and hence fluid transport.
Future efforts to identify the CHED2 disease gene would be greatly aided by the identification of another family. This may allow the disease interval to be refined and once a mutation in a gene is identified, would also enable the segregation of the mutation in the families to be tested.
5.1. Introduction

Granular corneal dystrophy or Groenouw Type I (CDGG1, OMIM No: 121900) is an inherited autosomal dominant disease that is fully penetrant. The corneal opacity is characterised by fine white dots that can be punctate or nodular, and mainly affects the central anterior stroma (Figure 5.1). These changes are often noted in the first decade of life but rarely affect vision at this stage. However, over time the opacities coalesce and become larger and as they multiply, visual acuity is compromised. The opacities consist of a non-collagenous protein called hyaline and are thought to result from the abnormal synthesis of proteins or phospholipids (Groenouw 1890; Rodrigues et al., 1984).

Figure 5.1 Photograph of CDGG1 patient's cornea and light micrograph of a corneal section stained with alcon-blue (with permission from Ian A. Cree). (A) arrow indicates the gray/white granular material deposited in the anterior corneal stroma. (B) layers of the epithelium stain pink while the stroma appears blue with the granular deposits staining pink.

Less that 5% of corneal grafts are performed on patients with corneal dystrophies, and of these, the stromal dystrophies are the most prevalent. A number of retrospective
studies have shown that the incidence of granular corneal dystrophy accounts for approximately 1.5% of all corneal transplants (Lang et al., 1987; al Faran et al., 1991; Santo et al., 1995; Konishi et al., 1997).

5.2. Granular corneal dystrophy type I pedigree and phenotype

A large British family with granular corneal dystrophy type I (CDGG1) participated in this linkage study to identify the gene causing the disease (Figure 5.2). Twenty members of this family were ascertained and 2x10 ml of whole blood was obtained from which DNA was isolated (section 2.5.1.1).

13 affected individuals and 7 unaffected individuals of the family were assessed by slit lamp examination and all the subjects classified as affected had the white granular deposits in the corneal stroma usually associated with CDGG1 (Figure 5.1a). The diagnosis was confirmed by histology in those individuals who had undergone keratoplasty (Figure 5.1b).

A general medical history of all the subjects was taken and did not show any other medical condition segregating with the disease.

5.3. Linkage study for CDGG1

Prior to the ascertainment of this pedigree very few genetic studies had been performed on the stromal dystrophies. Initial linkage analysis using protein and blood group markers found no evidence for linkage in a large granular family (Kompf et al., 1989), while a study by Moller and colleagues (1989) obtained a lod score of over one on chromosome 12p (Moller et al., 1989).

Using microsatellite markers from the long arm of chromosome 5 it was shown that CDGG1 was linked to a region near the interleukin 9 (IL9) gene on 5q31 (Eiberg et al., 1994; Stone et al., 1994). It was also shown that Avellino Corneal Dystrophy, Reis-Bucklers dystrophy and Lattice Corneal Dystrophy Type1 (see sections 1.3.3.2, 1.3.3.3, 1.3.3.4) mapped to this locus (Stone et al., 1994). These findings raised the possibility that, either these dystrophies were caused by mutations in different genes that mapped to 5q, or changes within a single gene were responsible for the differing phenotypes. The CDGG1 disease interval had been localised to a 10 cM gene rich region on chromosome.
Figure 5.2 Abridged pedigree of a family with CDGG1 used in this linkage study. Solid symbols indicate affected members and open symbols indicate unaffected individuals. All individuals who were ascertained for this study are shown by the symbol *.
5q flanked by the markers IL9 and D5S436 when the linkage study, described in this thesis, was initiated (Stone et al., 1994).

Microsatellite markers spanning the locus on chromosome 5q (Table 5.1) were genotyped in the family (Eiberg et al., 1994; Stone et al., 1994). The markers were PCR amplified (section 2.5.5 and Table 5.1), radioactively labelled (section 2.5.5.4) and resolved on an acrylamide gel (section 2.5.4.2).

<table>
<thead>
<tr>
<th>Marker</th>
<th>T\text{Anneal}</th>
<th>Amplimer</th>
<th>Marker</th>
<th>T\text{Anneal}</th>
<th>Amplimer</th>
</tr>
</thead>
<tbody>
<tr>
<td>D5S421</td>
<td>55 °C</td>
<td>165bp</td>
<td>D5S399</td>
<td>60 °C</td>
<td>116bp</td>
</tr>
<tr>
<td>D5S393</td>
<td>53 °C</td>
<td>162bp</td>
<td>D5S396</td>
<td>62 °C</td>
<td>122bp</td>
</tr>
<tr>
<td>D5S816</td>
<td>56 °C</td>
<td>235bp</td>
<td>D5S638</td>
<td>60 °C</td>
<td>118bp</td>
</tr>
</tbody>
</table>

Table 5.1 Microsatellite markers on chromosome 5q used in a linkage study. PCR conditions and marker size are shown. The primer sequences were obtained from the Genome Database (http://www.gdb.org).

Figure 5.3 Autoradiograph showing the alleles of microsatellite marker D5S638 in the CDGG1 pedigree.
Using the data generated from six microsatellite markers analysed (Figure 5.3 and Tables 5.2 and 5.3), lod scores were calculated using Cyrillic (section 2.5.13.2).

<table>
<thead>
<tr>
<th>Lane</th>
<th>Ped Nos</th>
<th>Status</th>
<th>Alleles</th>
<th>Lane</th>
<th>Ped Nos</th>
<th>Status</th>
<th>Alleles</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>4-1</td>
<td>A</td>
<td>12</td>
<td>12</td>
<td>5-9</td>
<td>U</td>
<td>24</td>
</tr>
<tr>
<td>1</td>
<td>4-2</td>
<td>U</td>
<td>22</td>
<td>15</td>
<td>5-10</td>
<td>A</td>
<td>22</td>
</tr>
<tr>
<td>20</td>
<td>4-3</td>
<td>A</td>
<td>22</td>
<td>16</td>
<td>5-11</td>
<td>U</td>
<td>23</td>
</tr>
<tr>
<td>19</td>
<td>4-4</td>
<td>A</td>
<td>22</td>
<td>8</td>
<td>6-1</td>
<td>A</td>
<td>22</td>
</tr>
<tr>
<td>5</td>
<td>4-6</td>
<td>A</td>
<td>23</td>
<td>9</td>
<td>6-2</td>
<td>A</td>
<td>34</td>
</tr>
<tr>
<td>4</td>
<td>5-1</td>
<td>A</td>
<td>12</td>
<td>10</td>
<td>6-3</td>
<td>A</td>
<td>34</td>
</tr>
<tr>
<td>3</td>
<td>5-2</td>
<td>U</td>
<td>-</td>
<td>13</td>
<td>6-4</td>
<td>A</td>
<td>24</td>
</tr>
<tr>
<td>6</td>
<td>5-6</td>
<td>A</td>
<td>23</td>
<td>14</td>
<td>6-5</td>
<td>U</td>
<td>24</td>
</tr>
<tr>
<td>7</td>
<td>5-7</td>
<td>U</td>
<td>24</td>
<td>17</td>
<td>6-7</td>
<td>U</td>
<td>23</td>
</tr>
<tr>
<td>11</td>
<td>5-8</td>
<td>A</td>
<td>24</td>
<td>18</td>
<td>6-8</td>
<td>A</td>
<td>23</td>
</tr>
</tbody>
</table>

Table 5.2 Alleles and disease status of individuals scored for the marker D5S638
A= affected and U= unaffected.

<table>
<thead>
<tr>
<th>Markers</th>
<th>0</th>
<th>0.05</th>
<th>0.01</th>
<th>0.02</th>
<th>0.03</th>
<th>0.04</th>
<th>Z_{max}</th>
</tr>
</thead>
<tbody>
<tr>
<td>D5S421</td>
<td>–α</td>
<td>-1.30</td>
<td>-0.52</td>
<td>0.02</td>
<td>0.15</td>
<td>0.12</td>
<td>0.12</td>
</tr>
<tr>
<td>D5S393</td>
<td>2.93</td>
<td>2.58</td>
<td>2.23</td>
<td>1.54</td>
<td>0.90</td>
<td>0.38</td>
<td>2.93</td>
</tr>
<tr>
<td>D5S816</td>
<td>3.23</td>
<td>2.87</td>
<td>2.49</td>
<td>1.742</td>
<td>1.03</td>
<td>0.43</td>
<td>3.23</td>
</tr>
<tr>
<td>D5S399</td>
<td>–α</td>
<td>1.99</td>
<td>1.91</td>
<td>1.47</td>
<td>0.92</td>
<td>0.42</td>
<td>1.99</td>
</tr>
<tr>
<td>D5S396</td>
<td>3.10</td>
<td>2.76</td>
<td>2.41</td>
<td>1.71</td>
<td>1.04</td>
<td>0.46</td>
<td>3.10</td>
</tr>
<tr>
<td>D5S638</td>
<td>–α</td>
<td>-1.30</td>
<td>-0.61</td>
<td>-0.14</td>
<td>-0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Table 5.3 Two-point lodscores for linkage between the CDGG1 phenotype and six markers spanning the locus on chromosome 5q.
θ is the recombination fraction and Z_{max} is the maximum lod score for each microsatellite marker.
Figure 5.4 Haplotype data for the CDGG1 pedigree with markers on 5q.
The circles represent females, the squares males, filled symbols indicate affected status and box indicate the affected haplotype.
5.3.1 Refinement of the critical interval

The CDGG1 pedigree was mapped to a 20 cM interval on chromosome 5q31-33 with the flanking markers D5S421 and D5S399 (Figure 5.4 and Figure 5.6). The individuals IV:1, IV:3, IV:4 and V:1 define a proximal crossover with marker D5S421.

While the unaffected individual VI:5 defines the distal crossover with marker D5S399 (Figure 5.4). The microsatellite markers D5S393 and D5S816 did not show any recombinant individuals while the marker D5S396 was not informative in individual VI:5 for the distal crossover and thus the marker shows no recombination. The published marker order is (cen-D5S421-D5S393-D5S816-D5S399-D5S396-tel) and this data is consistent with observation of crossovers in this family.

Assuming a single gene is causative for the multiple corneal dystrophies mapping to chromosome 5q, data can be combined to define the disease locus. The disease region in lattice corneal dystrophy (CDL1) was refined to a 2 cM interval between the markers D5S393 and D5S396 (Gregory et al., 1995). The disease locus was then further reduced using CDL1 and CDGG1 pedigrees, to a 1 cM interval between the markers D5S393 and D5S399 (Korvatska et al., 1996).

As the region had been refined by others, and was approximately one megabase of DNA, a physical map of the region was created.

5.4 Physical Mapping of 5q31

The construction of a physical map spanning the disease locus involves the assembly of a series of overlapping clones to create a contiguous region that represents that genomic interval of interest (section 1.8). To this end, 13 YACs were obtained from a physical map of the region represented on the Whitehead database (www-genome.wi.mit.edu). The disease locus lies within the contig WC5.10 which consists of 47 YACs.

The 13 YACs identified from the Whitehead map were obtained from HGMP, propagated and the DNA extracted (section 2.5.1.3). Microsatellite markers used to map the family to chromosome 5q31 (Table 5.1) were used for STS content mapping of these YACs. Five YACs, 880G9, 721B6, 745F8, 773D3 and 796D11, were shown to span the
disease interval. An example of STS content mapping of the YACs is shown (Figure 5.5) for marker D5S816. The resulting physical map is shown in Figure 5.6.

![Figure 5.5 YAC STS content mapping for marker D5S816.](image)

An agarose gel showing amplification of D5S816 on a panel of 13 YACs spanning the CDGG1 locus. YAC 1-5 are positive, C is human genomic DNA control, W is negative control without DNA. 1=880G9, 2=721B6, 3=745F8, 4=773D3, 5=796D1, 6=638E12, 7=709A3, 8=737B9, 9=750C11, 10=781E10, 11=800B2, 12=919D4 and 13=919D6.

![Figure 5.6 Physical map of the CDGG1 region on 5q31-33.](image)

Key: Microsatellite markers (black) and genes (pink) are shown. YAC clones are represented as red bars.
5.4.1 FISH analysis of YAC clones on 5q31

The five YAC clones spanning the disease region (Figure 5.6) were used for FISH analysis. DNA was labelled and hybridised against a chromosomal metaphase spread (section 2.5.9.5). For the five YACs analysed, a hybridisation signal was only observed on chromosome 5q indicating that these YACs were not chimaeric (Figure 5.7). An example of a FISH experiment is shown in Figure 5.7. (FISH experiments were performed by M.Fox, UCL.)

![FISH experiment](image)

**Figure 5.7 Fluorescent in situ hybridisation (FISH) of YAC 773D to metaphase spreads.**
The arrows indicate the positive signal, which is only detected on chromosome 5q31.

5.5 Candidate Genes mapping to the critical interval on 5q31

The five YAC contig was used as a mapping tool to place genes or ESTs within the critical interval for CDGG1. The 1 cM CDGG1 locus on 5q31 is in a relatively gene dense region (Dixon *et al.*, 1994). 10 ESTs were chosen from the contig WC5.10 spanning the critical interval and are shown in table 5.4. In addition, the gene LOX1 was assessed as a candidate gene.
<table>
<thead>
<tr>
<th>Identifier</th>
<th>Primer Pair</th>
<th>Amplimer bp</th>
<th>Known Homology/Expression</th>
</tr>
</thead>
</table>
| IB3085     | 5' - TGCACTGTCAGAAATCCACGT - 3'  
             | 5' - TCACACTGTCAGAAATCCACGT - 3' | 240 | Infant brain |
| βig-H3     | 5' - CTGCATGGGAATTCAGGCACACG - 3'  
             | 5' - TCAACCATGTCAGAAATCCACGT - 3' | 210 | Adenocarcinoma TGF-β induced gene |
| SGC32895   | 5' - TTTCACACAGATTGCTACTTCAG - 3'  
             | 5' - TACATGGGAATTCAGGCACACG - 3' | 138 | Fetal/ placental |
| WI-7012    | 5' - ACTACACATCCTATGTCATGCCC - 3'  
             | 5' - GCATGGGAATTCAGGCACACG - 3' | 337 | Glucocorticoid receptor |
| WI-18406   | 5' - CGCTAAATATTGCTCTGGACC - 3'  
             | 5' - ATTTTTAGAGATTTTCTTCATCCATG - 3' | 130 | fetal brain |
| SGC35217   | 5' - ATGTTATGATTAGATGATGCCC - 3'  
             | 5' - ATTTTTAGAGATTTTCTTCATCCATG - 3' | 150 | fetal liver |
| WI-16140   | 5' - AGTATCAGCCTTTATTTTTCA - 3'  
             | 5' - CCAGAGTTATATTGAGAATGTT - 3' | 165 | Brain/spleen |
| TIGR-A002114 | 5' - CTTTTCTGAAAGCAAGCTAAGA - 3'  
             | 5' - GACATGTACATGCAATGACAC - 3' | 128 | cytokine |
| WI-13437   | 5' - AAGAAGTTTTGTTTTTATGTCG - 3'  
             | 5' - CTCTGTGACCGGCTGGATC - 3' | 150 | solute carrier family 25 |
| WI-9259-Testican  | 5' - GCCAATGGATTTATCTCTATGACG - 3'  
             | 5' - ATGTTGGGTTTTTTTGTGGACA - 3' | 125 | SPOCK-like proteoglycan |

Table 5.4 Candidate genes and EST’s mapping to the disease interval.

5.5.1 LOX1

Lysyl oxidase (LOX1) is an extracellular copper enzyme that initiates the crosslinking of collagens and elastin by catalyzing oxidative deamination of the ε-amino group in certain lysine and hydroxylysine residues of collagens and lysine residues of elastin (Kagan et al., 1991). This enzyme is expressed in a wide range of tissues such as skin fibroblasts, placenta, kidney and eye including the cornea (Kao et al., 1983; Svinarich et al., 1992; Fujimaki et al., 1999). The LOX1 gene spans 5 kb and consists of 7 exons. The gene encodes a protein of 417 amino acids and there is evidence of
alternative splice variants (Hamalainen et al., 1991). It was localised to the long arm of chromosome 5 (5q23.3-31.1) by in situ hybridization as well as by somatic cell panel mapping (Mariani et al., 1992). As the LOX1 gene was mapped to the disease locus, is expressed in the cornea, and is possibly involved in cross-linking of extracellular matrix proteins. This made the LOX1 gene a good candidate for CDGG1.

However, STS content mapping demonstrated that none of the 5 YACs spanning the CDGG1 locus were positive for the LOX1 STS implying that the gene lies outside the critical interval (Figure 5.8). This finding was corroborated by Korvatska and colleagues (1996) who also demonstrated that both LOX1 and SPARC were excluded as candidate genes for this disease based on recombinant individuals within the pedigrees that they studied (Figure 5.6).

Figure 5.8 STS content mapping of 5 YACs spanning the critical interval for CDGG1 with an STS for the LOX 1 gene.
Key: 1=880G9, 2=721B6, 3=745F8, 4=773D3 and 5=796D11

5.5.2. βig-H3

Munier and colleagues (1997) identified mutations in the βig-H3 gene in patients with CDGG1, CDLCD and Avellino dystrophy.

Transforming growth factor β (TGF-β) belongs to a family of proteins that are involved in the regulation of differentiation and growth of many cells. It has been shown that TGF–β is involved in the synthesis and secretion of fibronectin and collagen (Ignatzi et al., 1986; Ignatzi et al., 1987). Skonier and colleagues (1992) used a cDNA library from a lung adenocarcinoma cell line (A549) and obtained a clone of 68-kD called TGF–β induced gene -H3 (βig-H3), which was upregulated when induced by TGF–β (Skonier et al., 1992). The gene encodes a protein of 683 amino acids, with a secretory signal peptide, a carboxy-terminal RGD binding domain and four internal homologous
domains. The gene was mapped to human chromosome 5q31 and mouse chromosome 13 using FISH (Skonier et al., 1994). Subsequently βig-H3 was shown to be expressed in the cornea as well as the ocular ciliary epithelium and as such was a good candidate gene for ocular disease (Escribano et al., 1994).

### 5.5.3 Genomic structure of βig-H3

The cDNA sequence is 2052 nucleotides long divided into 17 exons (Figure 5.9). There are 140 amino acid residues that are repeated 4 times between positions 139-275, 276-410, 411-537, 538-683 (Figure 5.10). There are internal homologous domains that share identities between 16% and 31%, with domain 3 being the most divergent (Skonier et al., 1992). The internal homologous domains share limited homology to fascilin-I, a Drosophila neuronal cell adhesion molecule.

![Figure 5.9 Genomic structure of βig-H3/(TGFBI) and the reported mutations. Blue boxes indicate exons (not to scale).](image)

On the basis of this homology it was postulated that βig-H3 may act as a surface recognition protein. The protein also has a secretory leader sequence and an Arg-Gly-Asp (RGD) motif that has been known to act as an integrin recognition signal.
These sequences may also allow βig-H3 to function as a cellular adhesion molecule or to be involved in signaling between cells.

<table>
<thead>
<tr>
<th>Exon</th>
<th>Primer sequence</th>
<th>Exon</th>
<th>Primer sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exon 1F</td>
<td>5' gcgcctgccgctcggtc3'</td>
<td>Exon 1R</td>
<td>5' ggcgcgccagccggctc 3'</td>
</tr>
<tr>
<td>Exon 2F</td>
<td>5' tgaatatgtgtgaccc3'</td>
<td>Exon 2R</td>
<td>5' tggcggcatccgctc 3'</td>
</tr>
<tr>
<td>Exon 3F</td>
<td>5' ggcagtcagccgactc3'</td>
<td>Exon 3R</td>
<td>5' ggctggtccagctcgtc 3'</td>
</tr>
<tr>
<td>Exon 4F</td>
<td>5' ttctctctctctctcctc3</td>
<td>Exon 4R</td>
<td>5' agactcccattcatctggcc 3'</td>
</tr>
<tr>
<td>Exon 5F</td>
<td>5' tggcctctctctctcctc3</td>
<td>Exon 5R</td>
<td>5' ccacatggaagctgaatgg 3'</td>
</tr>
<tr>
<td>Exon 6F</td>
<td>5' gggctggggctgactcgc 3'</td>
<td>Exon 6R</td>
<td>5' gccgcagagtttgaggtgc 3'</td>
</tr>
<tr>
<td>Exon 7F</td>
<td>5' accagtgaagctgtgtgc3'</td>
<td>Exon 7R</td>
<td>5' tggcaggctgagttgtc 3'</td>
</tr>
<tr>
<td>Exon 8F</td>
<td>5' gcgcctgggctgactcgc3'</td>
<td>Exon 8R</td>
<td>5' ccacatggaagctgaatgg 3'</td>
</tr>
<tr>
<td>Exon 9F</td>
<td>5' cctctctctctctctcctc3</td>
<td>Exon 9R</td>
<td>5' ggcgcagagtttgaggtgc 3'</td>
</tr>
<tr>
<td>Exon 10F</td>
<td>5' ttctctctctctctcctc3</td>
<td>Exon10R</td>
<td>5' ttacctaatataagtacaagcc 3'</td>
</tr>
<tr>
<td>Exon 11F</td>
<td>5' ggcgcctgggctgactcgc3'</td>
<td>Exon 11R</td>
<td>5' tccacatggaagctgaatgg 3'</td>
</tr>
<tr>
<td>Exon 12F</td>
<td>5' cctctctctctctcctc3</td>
<td>Exon 12R</td>
<td>5' gccgcagagtttgaggtgc 3'</td>
</tr>
<tr>
<td>Exon 13F</td>
<td>5' cctctctctctctcctc3</td>
<td>Exon 13R</td>
<td>5' cctctctctctctcctc3'</td>
</tr>
<tr>
<td>Exon 14F</td>
<td>5' ggcgcctgggctgactcgc3'</td>
<td>Exon 14R</td>
<td>5' tccacatggaagctgaatgg 3'</td>
</tr>
<tr>
<td>Exon 15F</td>
<td>5' ggcgcctgggctgactcgc3'</td>
<td>Exon 15R</td>
<td>5' ggcgcctgggctgactcgc3'</td>
</tr>
<tr>
<td>Exon 16F</td>
<td>5' cctctctctctctcctc3</td>
<td>Exon 16R</td>
<td>5' cctctctctctctcctc3'</td>
</tr>
<tr>
<td>Exon 17F</td>
<td>5' cctctctctctctcctc3</td>
<td>Exon 17R</td>
<td>5' ggcgcctgggctgactcgc3'</td>
</tr>
</tbody>
</table>

Table 5.5 Intronic primers used to identify mutations in βig-H3
5.5.4 βig-H3 mutation detection

There were two mutation hotspots previously described for exon 4 and exon 12 of βig-H3 (Figure 5.9). Mutations in codon 555 were identified in patients with CDGG1 and CDRB1 therefore exon twelve, which codes for this domain of the protein, was initially screened in this family (Table 5.5). The exon was PCR amplified in patients and controls (section 2.5.5.1), purified and sequenced (sections 2.5.7.2 and 2.5.8). Sequencing of two affected individuals and one unaffected family member identified a R555W (CGG/TGG) substitution for codon 555 (Figure 5.11). Although the remaining exons were sequenced in affected and control individuals no further changes were noted.

![Figure 5.11 Sequence from exon 12 of βig-H3 of two affected (A and B) and one unaffected individual (C).](image)
The arrows (red) indicate the position of the substitution from CGG to TGG in affected individuals A and B. The wild type CGG is shown by the arrow (purple) in (C) the unaffected family member.

To show the segregation of this change with affected status in the family, exon twelve was amplified in all members of the family and the restriction enzyme Bst X1 was used to digest the PCR product (section 2.5.3). The products were then resolved on an agarose gel (2.5.4.1).

PCR products from all unaffected individuals remained uncut and appeared as a single product of 300 bp (Figure 5.12). However, the PCR products from the affected individuals that shared the same change, a C to T at cDNA position 1710, created a BstX1
restriction digestion site, which resulted in three products. The unaffected allele, which
does not digest (300 bp) while the mutated allele, which did digest, resulted in 2 products
of 200bp and 100bp respectively (Figure 5.12). The restriction enzyme digest showed
that the disease in this family segregated with the mutation identified.

![Figure 5.12 Restriction digest of PCR product of exon 12 with BstX1.](image)

PCR product from the unaffected individuals remain undigested while the arrows
highlights the digestion pattern seen in affected patients.

5.6 Discussion

5.6.1. Stromal Dystrophies mapping to chromosome 5q

Seven different clinical phenotypes of corneal stromal dystrophies are caused by
mutations within βig-H3 (Table 5.6). There appears to be a genotype-phenotype
correlation where each phenotype is associated with a specific mutation. However, there
is also some overlap in phenotypic expression. The mutation R124L has been described
as a variant of granular dystrophy, yet as the deposits are seen between the epithelium
and Bowman’s membrane this phenotype resembles Reis-Buckler’s dystrophy (Okada et
al., 1998; Mashima et al., 1999). Generally, the phenotypes can be separated by the
location and nature of the deposits within the cornea. In CDRB1 the deposits are non-
amyloid and are localised to the epithelium and Bowman’s membrane, while in CDGG1
the deposits are also non-amyloid but are found in the sub-epithelial layer and the stroma
(section 1.3.2.1 and 1.3.2.2). Lattice corneal dystrophy can be differentiated into CDL1,
CDLIIIA, intermediate CDL1/CDLIIIA and CDLIV and the deposits are amyloid but
lattice type depends on whether it is located in the anterior, mid or posterior stroma and
also on mode of inheritance and age of onset (section 1.3.2.3). CDA is a combination of phenotypes between CDGG1 and CDL1 and has both amyloid and non-amyloid deposits (section 1.3.2.4).

One of the most striking findings concerning the mutations in keratoepithelin, is codon R124 and codon R555, which are mutation hotspots, and account for the majority of patients with CDGG1, CDRB1, CDL1 and CDA worldwide (Konishi et al., 1998; Korvatska et al., 1998; Konishi et al., 1999; Mashima et al., 1999; Mashima et al., 2000). Subsequently, mutations within other parts of the protein have been identified but account for a relatively small number of patients (Table 5.6 and Figure 5.9).

<table>
<thead>
<tr>
<th>Corneal dystrophy phenotypes</th>
<th>Mutations</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDGG1</td>
<td>R555W</td>
<td>(Munier et al., 1997)</td>
</tr>
<tr>
<td></td>
<td>R124S</td>
<td>(Stewart et al., 1999)</td>
</tr>
<tr>
<td></td>
<td>Δ125–126</td>
<td>(Dighiero et al., 2000)</td>
</tr>
<tr>
<td></td>
<td>R124L</td>
<td>(Dighiero et al., 2000)</td>
</tr>
<tr>
<td>CDRB1</td>
<td>R555Q</td>
<td>(Munier et al., 1997)</td>
</tr>
<tr>
<td></td>
<td>R124L</td>
<td>(Okada et al., 1998)</td>
</tr>
<tr>
<td></td>
<td>ΔF540</td>
<td>(Rozzo et al., 1998)</td>
</tr>
<tr>
<td></td>
<td>G623D</td>
<td>(Afshari et al., 2001)</td>
</tr>
<tr>
<td>CDL1</td>
<td>R124C</td>
<td>(Munier et al., 1997)</td>
</tr>
<tr>
<td></td>
<td>L518P</td>
<td>(Endo et al., 1999)</td>
</tr>
<tr>
<td>CDLIIIA</td>
<td>P501T</td>
<td>(Yamamoto et al., 1998)</td>
</tr>
<tr>
<td></td>
<td>A622H</td>
<td>(Stewart et al., 1999)</td>
</tr>
<tr>
<td></td>
<td>H626A</td>
<td>(Stewart et al., 1999)</td>
</tr>
<tr>
<td>CDL1/ CDLIIIA</td>
<td>A546T</td>
<td>(Dighiero et al., 2000)</td>
</tr>
<tr>
<td></td>
<td>N544S</td>
<td>(Schmitt-Bernard et al., 2000)</td>
</tr>
<tr>
<td></td>
<td>H626R</td>
<td>(Stewart et al., 1999)</td>
</tr>
<tr>
<td>CDLIV</td>
<td>L527R</td>
<td>(Fujiki et al., 1998)</td>
</tr>
<tr>
<td></td>
<td>N544S</td>
<td></td>
</tr>
<tr>
<td>CDA</td>
<td>R124H</td>
<td>(Munier et al., 1997)</td>
</tr>
</tbody>
</table>

Table 5.6 Seven different stromal dystrophies and the mutations responsible for these phenotypes.

In an a recent study, 36 families from the Moorfields register that were screened for mutations in βig-H3 showed four missense mutations which altered protein sequence for each of the following phenotypes: R124C for LCD1, R124H for ACD, R555Q for CDRB1 and R555W for CDGG1 (El Ashry personal communication). All of these
mutations had previously been described by Munier and co-workers (1997). The large CDGG1 pedigree described in this chapter also had the common change, R555W.

After the initial identification of a mutational hotspot there appeared to be a correlation between the codon R124 mutations responsible for CDL1 and CDA, which had amyloid deposits and the R555 mutations for CDGG1 and CDRB1 that had non-amyloid deposits (Korvatska et al., 1998). However, it was later shown that the R124 mutations could result in the deposition of non-amyloid material and that R555 mutations could result in the deposition of amyloid material. In the Japanese population, for granular dystrophy the R124H mutation is the most common change identified in keratoepithelin (Konishi et al., 1999; Mashima et al., 2000).

5.6.2. TGFBI (BigH3) gene and keratoepithelin protein

BigH3 was initially identified when a lung adenocarcinoma cell line was induced with TGF-β (Skonier et al., 1992). TGF-β has been shown to upregulate the mRNA of TGFBI in a number of cell types including cornea (Escribano et al., 1994; El-Shabrawi et al., 1998). In addition, proteins such as collagens and proteoglycans have also been shown to be upregulated in the presence of TGF-β (El-Shabrawi et al., 1998).

BigH3 encodes a 683 amino acid protein containing an amino-terminal secretory signal sequence, a carboxy-terminal integrin binding RGD domain and four homologous internal domains. The four domains are 140 residues each and could possibly fold to form a potential bivalent structure which would act as a bridge between cells expressing the appropriate ligand (Skonier et al., 1992). The four internal domains of keratoepithelin share limited similarity with the fascilin 1, which has a putative role as a surface recognition molecule involved in growth cone guidance. Keratoepithelin also shares similarity of 43% to the protein OSF-2 that has a putative function as a cell adhesion molecule in bone formation (Kawamoto et al., 1998). Fascilin I has shared similarity to OSF-2 and it is possible that these three proteins have similar structural and functional roles. Interestingly, both OSF-2 and keratoepithein are both upregulated by TGF-β.

5.6.3 The role of keratoepithelin (KE) in the cornea
Mutations in the BigH gene have now been extensively reported. It has been shown that there is some correlation between genotype and the phenotype, such that certain mutations within this gene will involve the deposition of amyloid material while different mutations will deposit non-amyloid material, even though the same codon may be involved. However, relatively little is known about the role of keratoepithelin (KE) in the cornea. A protein identified that co-purified with collagen type VI and it was shown that this was keratoepithelin (Rawe et al., 1997). This filamentous type of collagen is involved in cell adhesion and the spreading of corneal fibroblasts. There are 11 Arg-Gly-Asp (RGD) motifs in the helical domain of the collagen that can interact with proteins such as KE and proteoglycans like decorin. One possible function of KE is to facilitate corneal cell-ECM interaction as it has amino acid homology to the microfibrillar component of elastic fibres and also binds Type II collagen and fibronectin.

Korvatska and colleagues (1999) generated two antibodies, KE-15 and KE-2 against peptides that were not overlapping and corresponded to the amino acid regions 69-364 and 426-682 respectively. These antibodies were used in immunohistology on CDL1, CDA and CDGG1 patient’s corneas after keratoplasty. The amyloid deposits in CDL1 patient’s cornea stained positive with the KE-2 but not for KE-15, while the CDGG1 and CDA patient’s corneas gave a positive signal for both KE-15 and KE-2 (Konishi et al., 2000). Thus KE is present in both the amyloid and non-amyloid deposits. The different staining patterns of amyloid and non-amyloid deposits in cornea shown by the two antibodies against the amino and carboxy terminal ends of KE suggest that the misfolding of KE may be pathogenic (Korvatska et al., 1999).

Each mutation was associated with characteristic changes of protein turnover in corneal tissue. Amyloidogenesis in R124C (CDL1) corneas was accompanied by the accumulation of N-terminal keratoepithelin fragments (KE-2), whereby species of 44 kDa were the major constituents of amyloid fibrils. R124H (CDA) corneas with non-amyloid inclusions showed accumulation of a new 66- kDa species altogether with the full size 68- kDa form. Finally, in R124L (CDGG1) corneas with non-amyloid deposits, only the accumulation of the 68- kDa form was found. Two dimensional gels revealed mutation-specific changes in the processing of the full size protein in all affected corneas. It appears that substitutions at the same residue (Arg-124) can result in cornea-specific
deposition of keratoepithelin via distinct aggregation pathways each involving altered turnover of the protein in corneal tissue (Korvatska et al., 2000)

The characterisation of the mouse Big-H3 gene and its embryonic expression will aid the development of animal models of these diseases and improve our understanding of the role of KE in the cornea (Schorderet et al., 2000).

5.6.4. Future Work and Conclusion

The identification of βigH3 as the gene responsible for causing a number of stromal dystrophies has made a significant impact on the diagnosis of these dystrophies.

Although a correlation can be drawn between the disease genotype and the phenotypic expression in light of the mutation hotspots in exon 4 and exon 12, a more accurate description will be based on the genotype and the phenotype it represents eg. amyloid or non-amyloid deposits.

The function of KE remains to be elucidated, considering the fact that although KE is fairly ubiquitously expressed, the only pathogenic changes reported to date have been those concerning the cornea.

The creation of animal models for these diseases will allow the study of the dystrophies and any possible systemic changes that may exist.
CHAPTER 6

Cornea Plana

6.1 Introduction

6.1.1. Cornea plana

Cornea Plana can be inherited as either a milder, dominant disease (CNA1; MIM 121400) or as a more severe, recessively inherited form (CNA2; MIM 217300 section 1.3.3.9). CNA2 is a rare disorder in which the curvature of the cornea is flattened which results in a decrease in the refractive power of the cornea and hence a corresponding reduction in visual acuity (Eriksson et al., 1973). Cornea plana is characterised by a hazy corneal limbus, peripheral scleralisation of the cornea, often with deep central corneal opacities (Figure 6.1). Although, a rare disease, it is relatively highly prevalent in Finland where 78 cases have been reported compared to 35 cases worldwide (Forsius et al., 1998).

![Figure 6.1 A slit lamp photograph of a patient with cornea plana.](http://www.bausch.com)

The red arrow indicates the flattened corneal surface while the blue arrow highlights the indistinct limbus. Adapted from http://www.bausch.com.
6.1.2. Linkage analysis and mutation identification in CNA2 families

Tahvanainen and colleagues (1995) reported the initial linkage of CNA2 to a 10 cM interval on chromosome 12q22 in a Finnish pedigree. This was subsequently refined to a 3 cM region between the markers D12S82 and D12S351. Using linkage disequilibrium analysis in 32 Finnish families, an allelic association between CNA2 and the marker D12S351 was shown (Figure 6.2), allowing the region to be refined to a 1 cM region (Tahvanainen et al., 1995).

A Cuban family with an autosomal dominant pattern of inheritance for cornea plana (CNA1) was linked to an overlapping 3 cM region on chromosome 12q22 between the markers D12S82 and D12S351 (Tahvanainen et al., 1996).

![Figure 6.2 Ideogram of chromosome 12 showing the position of the CNA1 and CNA2 loci. The family of small leucine rich proteoglycans (SLRPs) mapping to this locus is also shown.](image)

135
Two Finnish CNA1 families were excluded from the locus on chromosome 12, indicating that there is genetic heterogeneity as well as phenotypic heterogeneity for this dystrophy (Tahvanainen et al., 1996).

More recently, Pellegata and colleagues (2000) reported mutations within the keratocan gene that were responsible for causing CNA2. In an American patient of Chinese origin, they identified a homozygous CAG to TAG transversion in exon 2 causing an amino acid change from glutamine to a stop codon (Q174X). This was predicted to result in a truncated protein of only 173 amino acids compared to a full length protein of 352 residues. Using a panel of 46 Finnish families, they identified a homozygous AAC to AGC transition in exon 2 at codon 247 in all affected individuals, indicating a founder effect in the Finnish population. The codon 247 mutation causes an amino acid substitution from Asn (N) to Ser (S). The Asn residue lies within the leucine rich repeat (LRR) motif and is highly conserved (section 6.4).

A novel mutation in a Bangladeshi family with CNA2, and a mild form of microphthalmia, was recently identified in exon 2 altering codon 215 of keratocan. The ACA to AAA change results in an amino acid substitution from Thr (T) to Lys (K) and occurs in the highly conserved LRR region (Lehmann et al., 2001).

6.1.3 Keratocan

Keratocan encodes the core protein of a major corneal keratan sulphate proteoglycan, but is also present in non-corneal tissue, primarily as a non-sulfated glycoprotein. Although keratocan is abundantly expressed in the cornea and sclera, it is detected in much lesser quantities in skin, ligament, cartilage, artery and striated muscles. Only in the cornea was KERA found to contain large, sulfated keratan sulfate chains due to post-translational modification (Corpuz et al., 1996; Iozzo et al., 1999). The keratocan gene spans 7.65 kb of genomic DNA, consists of three exons with a cDNA of 2160 bp and encodes a protein of 352 amino acids, designated KERA. Exon 1 is untranslated while exon 2 contains the start codon and encodes the central leucine-rich domains of KERA (Figure 6.3). KERA contains an N-terminal signal peptide followed by a highly conserved region containing ten LRR motifs (1.2.2.2).
Figure 6.3 Schematic diagram showing the structure of human keratocan gene. The three exons are represented by boxes and indicated by numbers. The translated regions are represented by red filled boxes and the untranslated regions by orange filled boxes. (Adapted from Pellegata et al., 2000).

Northern blot analysis and in situ hybridization demonstrated that keratocan mRNA is selectively expressed in the corneal tissue of the adult mouse (Liu et al., 1998). During embryonic development (section 1.2.5), keratocan mRNA is first detected in periocular mesenchymal cells migrating toward developing corneas on embryonic day 13.5 (E13.5). Its expression is gradually restricted to corneal stromal cells at E14.5 and timepoints thereafter. Keratocan mRNA can be detected in the scleral cells of E15.5 embryos, however by E18.5 there are no detectable levels. In adults, keratocan mRNA can be detected in corneal keratocytes but not in scleral cells. In addition the expression pattern of KERA during development suggests that this molecule might be important in developing and maintaining corneal transparency (Tasheva et al., 1998; Dunlevy et al., 2000; Pressman et al., 2000).

The amino acid sequence of KERA displays a high degree of similarity between mammalian species and there is evolutionary conservation between the KERA proteins.
6.2. CNA2 pedigree and phenotype

A two-generation Hispanic pedigree diagnosed with autosomal recessive cornea plana (CNA2) was ascertained (Figure 6.4). 2x 10 ml of whole blood in EDTA tubes were collected and DNA was extracted (section 2.5.1.1). The parents are first cousins and neither they, nor their other relatives, are affected. However, they have five children of which 3 are affected (Figure 6.4).

![Figure 6.4 A consanguineous recessive cornea plana (CNA2) Hispanic pedigree. A first-cousin marriage with three affected children and two unaffected children. Males are represent by squares and females by circles. Filled symbols denote an affected individual.](image)

Ophthalmic examination included slit-lamp measurement of intraocular pressure, biomicroscopy, gonioscopy, optic disc assessment and A-scan ultrasonography to determine axial length.

The clinical characteristics of the 3 affected individuals are as follows: all exhibited bilateral cornea plana with an indistinct limbal region, varying degrees of central corneal opacity (Figure 6.5) with reduced visual acuity and hypermetropia while intraocular pressure and axial lengths were within normal limits.
Figure 6.5 Slit lamp photograph of two individuals from the CNA2 pedigree. (A) Unaffected father. The slit lamp photographs highlight the curvature of the cornea (blue arrow). (B) Affected child. The slit lamp photograph shows a flattened, uneven cornea with scleralisation of the cornea, an indistinct limbus and corneal opacity (red arrows).
6.2.1 Corneal Topography

Corneal topography (CT) maps graphically represent a picture of corneal curvature. The CT map is a two dimensional representation of a 3-D shape. Thus, colours are used to represent curvature values with the blue shades indicating flatter curvatures than the red colours, which indicate steeper curvatures.

The affected individuals have an irregular, flattened corneal surface with the power of the cornea reduced to less than 38 diopters, while the unaffected individuals have a regular surface with the power of the cornea of 43 diopters or greater (Figure 6.6).

![Figure 6.6 Corneal topography of (A) an unaffected and (B) affected member of the CNA2 family.](image)
The coloured bar on the left indicates the degree of curvature of the cornea and the power in diopters.

6.3 Keratocan gene screening

PCR amplification (section 2.5.5.1) of the keratocan gene was performed using intronic oligonucleotide primers (Table 6.1) to amplify the coding region. As exon 2 is 893 bp long, it was amplified in three overlapping segments. The samples were sequenced bi-directionally on an ABI 3100 automated sequencer using standard conditions (section 2.5.8).
Table 6.1 Primer sequences and conditions used to amplify the coding region of keratocan.

<table>
<thead>
<tr>
<th>Exon</th>
<th>Sequence (Forward)</th>
<th>Sequence (Reverse)</th>
<th>T_a(°C)</th>
<th>Amplimer (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>tggtgactgggacgagtagg</td>
<td>ctttttcagaatagggttttg</td>
<td>61</td>
<td>273</td>
</tr>
<tr>
<td>2a</td>
<td>tggtgcataatctcctctcc</td>
<td>agggctcttttttaattcc</td>
<td>59</td>
<td>477</td>
</tr>
<tr>
<td>2b</td>
<td>ctgtagggtctataatggtgcgg</td>
<td>gggttcaggtctccagattg</td>
<td>58</td>
<td>522</td>
</tr>
<tr>
<td>2c</td>
<td>ctcatgcagctaaacatggc</td>
<td>gatcaaggtgaaggtgtgc</td>
<td>59</td>
<td>292</td>
</tr>
<tr>
<td>3</td>
<td>ttgggggaaacagataggg</td>
<td>gaaaaatggtggtccgagagc</td>
<td>59</td>
<td>463</td>
</tr>
</tbody>
</table>

Figure 6.7 Electropherograms showing the A393G change in keratocan.

The solid symbols indicate affected individuals that are homozygous G, while the open symbols represent unaffected heterozygous individuals A/G at position 393. This causes an amino acid substitution of Asn(N)131 Asp(D).
Sequencing of keratocan revealed a single-nucleotide substitution in exon 2 that segregated with the disease phenotype (Figure 6.5). The sequence changes from AAT at 391bp to GAT and results in an amino acid substitution of Asn (N) to Asp (D) in codon 131.

Both parents and unaffected children were heterozygous (A/G), however all three affected siblings were homozygous (G/G) for this change (Figure 6.7).

In excess of 130 control individuals (260 chromosomes) were tested for the sequence change using a combination of DHPLC (section 2.5.11.1) and sequencing (section 2.5.8). No alterations were detected confirming that this sequence change is unlikely to be a common polymorphism. However, it should be noted that the control chromosomes tested were of British Caucasian origin rather than Hispanic and it remains a possibility that this change may be a polymorphism in the Hispanic population.

6.4 Alignment of conserved residues

Further evidence that this sequence change may be responsible for causing cornea plana was obtained from a multiple sequence alignment using CLUSTALW (section 2.5.13.1), which was performed to determine if this residue was conserved.

The N131D substitution occurs within the third leucine-rich repeat (LRR) motif of KERA, which is a highly conserved residue within the LRR (1.2.1.2).

The CLUSTALW alignment (Figure 6.8) indicated that this residue was highly conserved in the KERA proteins from human, bovine, mouse, quail and chick.

Similarly, the mutations previously identified also affect highly conserved residues (Pellegata et al., 2000; Lehmann et al., 2001). The T215K change seen in a Bangladeshi family occurs at a highly conserved Thr residue in the seventh LRR. The N247S change, seen the in the Finnish population, involves the substitution of the conserved Asn (LXXLXLXXNXL) within the eighth LRR.
**Figure 6.8 A CLUSTALW alignment of human KERA with bovine, mouse, quail and chick.**

The mutation detected in this family is highlighted with a green arrow while the red arrows indicate mutations previously identified within the keratocan gene. The red line indicates the first 20 amino acids that form the signal peptide (SP) while the blue lines show the leucine-rich repeat regions from 1 to 10.
6.5 KERA protein modeling

KERA was found to have a relatively high degree of amino acid sequence identity (30.5%) to the acid-labile sub unit of serum insulin-like growth factor (ALS), which also contains a series of LRR motifs. The availability of a theoretical structural model for ALS facilitated the generation of a partial 3D model for KERA (Janišová et al., 1999). Although the model is incomplete at the N- and the C-termini, it was possible to examine the structural implication of the mutation at site 131.

The model was created using the Swiss Model server program (Guex et al., 1997) with Swiss Pdb Viewer v3.7b2 (Guex and Peitsch 1997) and Rasmol v2.6 (Roger Stayle, Glaxo Wellcome, Stevenage, UK) viewing programs.

The structural model for ALS, which has 20 LRRs (Figure 6.9A), forms a doughnut structure with an inner core. The predicted structure of the LRRs of KERA follows this model with its 10 LRRs also facing inwards forming a horseshoe structure (Figure 6.9B). The LRRs fold the protein into a series of parallel β-strands, which stack into an arched β-sheet array, creating an arch-shaped three-dimensional structure with keratan sulphate (KS) attachment sites positioned so that the KS chains extend from the convex side of the arch (Iozzo 1997). The N131D mutation is not predicted to cause a change in the structure of the protein (Figure 6.9C). Similarly neither does the N247S mutation which is responsible for the founder effect seen in the Finnish population. However, the T215K mutation seen in a Bangladeshi family does predict a conformational change, such that the mutation reduces the effective length of LRR7 and causes the loop connecting LRR6 to LRR7 to be laterally displaced (Figure 6.9D, Lehmann et al., 2001).

Although the N131D and N247S mutations are not predicted to cause a conformational change based on the protein model used, it is probable that the amino acid substitution at a highly conserved residue will have a deleterious effect.
Figure 6.9 Protein prediction models of ALS and KERA.

(A) Protein model of the acid-labile subunit (ALS) of insulin-like growth factor depicted as a ribbon with the LRRs lining the inner surface of the protein. (B) Predicted partial structure of wild-type KERA based on ALS. (C) Predicted structure of KERA with the mutations N131D (blue arrow) and N247S (red arrow) that show no change when compared to wt KERA. (D) The T215K mutation predicted to cause a conformational change in the predicted structure of KERA (red arrow) is also shown.
6.4 Discussion

Cornea plana is a rare corneal dystrophy, which displays both phenotypic and genetic heterogeneity. The dominant form of the disease CNA1, is a relatively mild disorder with little loss of visual acuity. The refractive power of the cornea is reduced from 45 to 38-42 diopters. The recessive form of this dystrophy, CNA2, results in a significant loss of visual acuity as the refractive power of the cornea is reduced to 25-35 diopters. Thus, these two dystrophies can be distinguished based on phenotype, as CNA2 presents as a more severe disease. Although, the two forms of the disease have been linked to an overlapping locus on chromosome 12q22, mutations within the keratocan gene have as yet only been identified in CNA2 families. However, there are at least two Finnish pedigrees with CNA1 that do not map to the locus on chromosome 12, indicating genetic heterogeneity (Tahvanainen et al., 1996).

Pellegata and colleagues (2000) identified mutations in keratocan that have been shown to cause CNA2 in both a large cohort of Finnish patients and an American patient of Chinese origin. They identified N247S (Asn 247 Ser) and Q174X (Glu 174 stop) homozygous changes respectively. More recently, a Bangladeshi CNA2 family was identified with a novel T215K (Thr 215 Lys) mutation in KERA. This family had a broader phenotype in which a mild form of microphthalmia co-segregated with the cornea plana dystrophy. These findings implicated KERA in a wider role in the formation of the structure of the eye (Liu et al., 1998; Pressman et al., 2000).

The homozygous mutation, N131D (Asn 131 Asp), described in this chapter is the fourth novel pathogenic change described in the keratocan gene. The family studied is of Hispanic origin and the affected children are from a consanguinous marriage. The mutation affects the asparagine residue in the third LRR which is highly conserved and this may lead to the destabilisation of the protein structure or alter the ability of the mutant KERA to interact with other proteins (Figure 6.8).

KERA is one of the small leucine-rich repeat proteoglycans (SLRPs) that can bind to collagen and other membrane proteins and may be involved in ocular development. Pressman and colleagues (2000) showed that mice that are homozygous for a targeted mutation of the LIM-homeodomain transcription factor Lmx1b, develop iris and ciliary
body hypoplasia as well as corneal stromal defects. Two cDNAs, mfl and mfh, normally downregulated in the developing cornea were expressed in the mutant and in these mice there were no detectable levels of keratocan mRNA. Additionally the ultrastructural analysis of mutant corneae indicated an abnormal pattern of collagen fibrillogenesis.
Chapter 7

Discussion

7.1 Overview of the work presented

At the start of this research little was known about the genetic basis of corneal dystrophies apart from the identification of mutations in the gelsolin gene in patients with LCD type II. However, over the years, seven genes have been identified which are implicated in causing corneal dystrophies. In addition, four other loci involving corneal dystrophies have been genetically mapped (Table 7.1).

The locus for congenital hereditary endothelial dystrophy (CHED1) was mapped to chromosome 20p11.2 using the largest reported family. The CHED1 locus was refined to a 2.7 cM interval on chromosome 20 that overlapped with the locus for PPCD. A physical map of the interval was created using a contig of YACs, PACs and BACs. This contig was used to exclude the cystatin genes and VSX1 as candidates for CHED1. Destrin, an actin depolymerising factor, was also excluded by sequencing the cDNA from two affected patients. The contig spanning the disease region of CHED1 is now complete and 23 genes have been predicted within this interval.

The locus for CHED2 was previously shown to be distinct from that of CHED1 and was localised to 20p12-13. A small consanguineous pedigree was used to try to refine the disease interval. Affected patients from this family were screened for mutations in CDS2, considered as a candidate gene for CHED2. Although highly expressed in the cornea, no mutations were identified in the coding sequence of CDS2.

Granular corneal dystrophy (CDGG1) is one of a number of stromal dystrophies that have been localised to chromosome 5q31-33. Linkage analysis was carried out on a British family and the disease was mapped in this family to 5q31-33. A physical map was created using YACs spanning the disease interval and this was used to exclude the LOX1 gene as a candidate for CDGG1. Following the identification of the causative gene βig-
H3 for the stromal dystrophies mapping to 5q, mutation detection in this family was performed and an R555W mutation in keratoepithelin was detected.

Mutation analysis was performed on a small family with autosomal recessive cornea plana (CNA2). The three exons of keratocan were sequenced and a change was identified in exon 2 that caused a highly conserved asparagine residue at position 131 to be altered to an aspartic acid residue which is likely to be disease causing.

The research presented highlights the genetic and phenotypic heterogeneity of corneal dystrophies.

7.2 Genetic heterogeneity

The corneal dystrophies are a heterogeneous group of inherited corneal diseases and the recent mapping and characterisation of a number of causative genes has enhanced our understanding of disease pathogenesis (Figure 7.1). These dystrophies display a remarkable degree of genetic heterogeneity.

Meesmann's dystrophy, which affects the corneal epithelium, shows genetic heterogeneity with loci on chromosome 12 and 17, however the phenotype is indistinguishable. Mutations have been identified in Keratin3 (K3 on chromosome 12) and Keratin12 (K12 on chromosome 17) as a cause of Meesmann's dystrophy. The keratins are divided into type I (acidic) and type II (neutral or basic) which dimerise with each other. Keratins are often tissue specific and K3 and K12 are expressed in corneal epithelial cells. A number of mutations have now been identified in K3 and K12 as a cause of Meesmann's dystrophy (Irvine et al., 1997; Nishida et al., 1997; Corden et al., 2000; Corden et al., 2000).

For posterior polymorphous dystrophy (PPCD) the main morphological change is the growth of epithelial-like cells with the capacity to proliferate and replace the corneal endothelium which is often accompanied by iridocorneal peripheral adhesions sometimes associated with other ocular disease such as glaucoma or keratoconus. Recently, genetic heterogeneity was demonstrated for PPCD with the mapping of a locus to the centromeric region of chromosome 20 and the identification of mutations in some families in the transcription factor VSX1. A locus on chromosome 1 was also implicated in PPCD when
<table>
<thead>
<tr>
<th>CHROMOSOME</th>
<th>DISEASE NAME</th>
<th>SYMBOL</th>
<th>LOCUS</th>
<th>FLANKING MARKER</th>
<th>GENE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SCHNYDER CRYSTALLINE DYSTROPHY</td>
<td>SCCD</td>
<td>1p36-p34.1</td>
<td>D1S2633-D1S228</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td>GELATINOUS DROP-LIKE DYSTROPHY</td>
<td>GDLD</td>
<td>1p31-32</td>
<td>D1S2890-D1S2801</td>
<td>M1S1</td>
</tr>
<tr>
<td>5</td>
<td>GRANULAR TYPE I</td>
<td>CDGI</td>
<td>5q31</td>
<td>D5S396-D5S399</td>
<td>βig-H3</td>
</tr>
<tr>
<td>LATTICE TYPE I</td>
<td>LATTICE TYPE III A</td>
<td>CDLII</td>
<td>9q34</td>
<td>D9S258-D9S1821</td>
<td>GELSOLIN</td>
</tr>
<tr>
<td>9</td>
<td>FAMILIAL AMYLOIDOSIS V</td>
<td>CDLII</td>
<td>9q34</td>
<td>D9S258-D9S1821</td>
<td>GELSOLIN</td>
</tr>
<tr>
<td>10</td>
<td>THIEL-BEHNKE DYSTROPHY</td>
<td>CDTB</td>
<td>10q23-q24</td>
<td>D10S677-D10S1671</td>
<td>-</td>
</tr>
<tr>
<td>12</td>
<td>CORNEA PLANA CONGENITA TYPE I</td>
<td>CNA1</td>
<td>12q21</td>
<td>D12S82-D12S327</td>
<td>KERATOCAN</td>
</tr>
<tr>
<td>CORNEA PLANA CONGENITA TYPE II</td>
<td>CNA2</td>
<td>12q21</td>
<td>D12S82-D12S327</td>
<td>KERATOCAN</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>MEESMANN'S DYSTROPHY</td>
<td>KRT3</td>
<td>12p 12-q13</td>
<td>D12S390-D12S325</td>
<td>KERATIN 3</td>
</tr>
<tr>
<td>16</td>
<td>MACULAR TYPE I</td>
<td>MCDC1</td>
<td>16q22</td>
<td>D16S18-D16S12</td>
<td>CHST6</td>
</tr>
<tr>
<td>MACULAR TYPE II</td>
<td>MCDC2</td>
<td>16q22</td>
<td>D16S18-D16S12</td>
<td>CHST6</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>MEESMANN'S DYSTROPHY</td>
<td>KRT12</td>
<td>17q12-q21</td>
<td>D17S800-D17S930</td>
<td>KERATIN 12</td>
</tr>
<tr>
<td>20</td>
<td>CONGENITAL HEREDITARY EPITHELIAL DYSTROPHY (DOMINANT)</td>
<td>CHED1</td>
<td>20p11.2-q11.2</td>
<td>D20S48-D20S471</td>
<td>-</td>
</tr>
<tr>
<td>20</td>
<td>CONGENITAL HEREDITARY EPITHELIAL DYSTROPHY (RECESSIVE)</td>
<td>CHED2</td>
<td>20p13</td>
<td>D20S113-D20S882</td>
<td>-</td>
</tr>
<tr>
<td>20</td>
<td>POSTERIOR POLYMORPHOUS DYSTROPHY</td>
<td>PPCD</td>
<td>20p</td>
<td>D20S98 - D20S108</td>
<td>VSX 1</td>
</tr>
</tbody>
</table>

Table 7.1 Table of corneal dystrophies with chromosomal localisation, flanking markers and genes
mutations in the gene COL8A2 were shown to cause both PPCD and Fuch’s endothelial dystrophy (Table 7.1).

Thiel-Behnke dystrophy has been described as a “honeycomb dystrophy” that causes corneal stromal opacities. In some patients, mutations in the \( \beta \text{ig-H3} \) gene have been identified (Munier \textit{et al.}, 2002). However, there is genetic heterogeneity such that a locus on chromosome 10 has been described in a family with this phenotype (Yee \textit{et al.}, 1997).

Although corneal dystrophies are rare these examples highlight the fact that clinical and genetic characterisation of new families will reveal new genetic loci and disease genes. In addition to the extensive genetic heterogeneity, wide phenotypic variation is also observed.

7.3 Genotype / Phenotype correlation

With the identification of the causative gene for a number of corneal dystrophies, it is important to try to relate a specific molecular defect to a corresponding clinical phenotype.

Corneal stromal dystrophies with distinct phenotypes were mapped to the same locus on chromosome 5q (Stone \textit{et al.}, 1994; Small \textit{et al.}, 1996). Mutations in the \( \beta \text{ig-H3} \) gene are responsible for causing these dystrophies (Table 7.1, Chapter 5). Mutations in the \( \beta \text{ig-H3} \) gene, therefore, cause a spectrum of different corneal stromal phenotypes. There are two mutation hot spots in keratoepithelin, one at position R124 and the other at R555, with approximately 50% of mutations targeting these two amino acids. However, there are four corresponding phenotypes, R124C in CDLI, R555W, R124L and R124S in CDGGI, R124H in CDA, R124L, and R555Q in CRRB. In addition, mutations in other domains cause CDGG1 (A125-126), CDRB1 (AF540, G623D), CDL1 (P551Q, L518P), CDLI1IA (P501T, A622H, H626A), CDL1/CDLI1IA (A546T, H626R, NVP629-630 ins) and CDLIV (L527, N544S). There is also a report of a homozygous R124H mutation in which the patient had an extremely severe granular dystrophy phenotype (Mashima \textit{et al.}, 1998). Korvatska and colleagues (1999) used immunohistochemistry in an attempt to determine the pathogenesis of keratoepithelin. It was shown that the corneae from
patients with the different phenotypes of 5q31 linked stromal dystrophies had accumulation of keratoepithelin. Antibodies were raised against the amino and carboxyl termini of keratoepithelin, and the different staining patterns of amyloid and nonamyloid deposits seen, suggested that two mechanisms of keratoepithelin misfolding are implicated in the pathogenesis of these stromal dystrophies.

Macular corneal dystrophy type I (MCDC1) and macular corneal dystrophy type II (MCDC2) are both autosomal recessive stromal dystrophies in which there is a respective absence and presence of sulphated keratan sulphate in the patient’s corneae and serum. Onset of macular corneal dystrophy is usually between ages 5 and 9, slowly progressive and there is the development of small, grey, punctate opacities that leads to a reduction in visual acuity. Mutations have been identified in a carbohydrate sulphotransferase gene (CHST6) mapping to chromosome 16q22 (Table 7.1). The MCDC1 mutations reported are either homozygous or compound heterozygous changes, which results in undetectable levels of sulphated keratan sulphate in the cornea and serum. However, MCDC2 mutations also involve large deletions and/or replacements caused by homologous recombination in the upstream region of CHST6, which results in the loss of cornea-specific expression of CHST6. Apart from the detection of sulphated keratan sulphate in the corneae and serum of patients with MCDC2, clinically MCDC1 and MCDC2 have indistinguishable phenotypes.

The identification of mutations in the genes implicated in corneal dystrophies affords the opportunity to conduct further research to understand the mechanism of pathogenesis of the disease. Specific functional assays need to be developed for each gene and each mutation identified to unravel the different pathogenic processes and subsequent clinical outcomes of disease.

7.4 Advances in the human genome project

The human genome project has made a significant contribution to the mapping and characterisation of genes implicated in human disease through microsatellite analysis, physical mapping and ultimately genome sequencing. Novel microsatellite markers are now easier to identify with the sequence available. In addition, these traditional markers can now be supplemented by single nucleotide polymorphism (SNP) maps with increased
density, enabling the fine mapping of disease loci. This increased density of polymorphisms has greatly enhanced the prospects of studying multifactorial or complex traits (Chanock 2001)

Once a disease locus has been genetically refined to the smallest region possible, the identification of genes within this interval has been facilitated by the sequence of the human genome and related strategies such as cDNA sequencing. Traditional positional cloning of disease causing genes has now been superseded by positional candidate gene strategies. To ensure success, however, one must first identify and prioritise all genes that map to the critical disease interval.

The human genome was estimated to be comprised of 60-80 000 genes (Fields et al., 1994) however, the publication of the sequence of the human genome (Lander et al., 2001; Venter et al., 2001) led to the revised estimate of 30-40 000 genes. Recently, Wiemann and colleagues (2001) sequenced 500 novel cDNAs and when these sequences were aligned to the sequences of finished chromosomes 21 and 22 a large proportion of the novel cDNAs had either been completely missed by the bioinformatic analysis of the genomic sequencing or had been incorrectly predicted. If this finding is extrapolated to the remaining chromosomes in the genome, then perhaps the estimate of 30-40 000 genes was too conservative. In addition, the complexity of the human genome, with many alternatively spliced products, translates to a greater number of proteins with modified locations and/or functions.

As individual research groups take on the challenge of in depth functional analysis of gene and protein families ('post genomics') the bioinformatic tools will become more accurate and descriptive and will evolve with the research. This in turn will aid annotation of cDNA and protein, as well as genomic, sequence.

In the post genomic era, the techniques collectively known as proteomics are useful for characterising the protein phenotype of a particular tissue or cell, as well as quantitatively identifying differences in the levels of individual proteins following modulation of a tissue or cell. These tools will enable us to monitor changes in the expression of a given corneal protein(s) and its post-translational modification, identify novel therapeutic targets and evaluate pharmacological effects on a given metabolic pathway. These technologies are now set to provide us with an immense amount of data.
for future analysis, which will hopefully pave the way to understanding disease pathogenesis and producing targeted therapies.

7.5 Animal models

The development of animal models for human diseases are essential for the study of the progression, cause, and development of therapeutic drugs to combat human disease. Unfortunately, naturally occurring animal models are rare for corneal dystrophies (as are the human equivalent dystrophies). However, in dogs a number of different corneal phenotypes have been described that resemble the human condition, such as Fuch’s endothelial dystrophy, crystalline corneal dystrophy and corneal epithelial erosions (Dice 1980; Crispin 1988; Cooley et al., 1990). Corneal dystrophies, due to the obvious nature of the phenotype, should be relatively easy to identify, but are perhaps overlooked during screening of mutants. To overcome the lack of available models, one commonly used and powerful approach is to create a null mutant.

The lumican knockout mouse displays a phenotype with corneal opacities and fragile corneal epithelia and skin. Affected mice show abnormal collagen formation in the dermis and the cornea similar to that seen in decorin null mice. Subsequently, it was shown that lumican plays a key role in the posterior stroma, maintaining normal fibril architecture (Chakravarti et al., 2000). However no corresponding human phenotype has been described.

7.6 Future perspectives for therapy

The ultimate goal of molecular ophthalmology is the early detection and therapeutic treatment of eye disease. The corneal dystrophies lend themselves to the development of therapeutic intervention, as the cornea is located in the anterior portion of the eye it is relatively easily accessed.

The autosomal recessive dystrophies discussed such as cornea plana or macular dystrophy are caused by haploinsufficiency, therefore the prospect of introducing the normal protein product to recover the phenotype is a possibility (Veske et al., 1999; Van
Hooser et al., 2000). For autosomal dominant diseases, such as the 5q31 linked stromal dystrophies in which there is an accumulation of an aberrant protein, the use of antisense technology would seem appropriate to block the increase in the accumulation of the aberrant protein, and thus slow the progression of the disease. The classical antisense approach is based on the selective blocking of a specific gene in vivo, which is responsible for a particular disease. Antisense oligonucleotides are short, generally 15 to 25 bases long, single stranded DNA fragments, which are targeted against a specific mRNA, which can be modified in order to increase their stability in vivo (Tamm et al., 2001). Oligonucleotides are designed in a complementary (antisense) orientation to their target (sense) mRNA to which they hybridise and thus block translation. They can also be designed to bind to genomic DNA in the nucleus and thus block transcription or may also be utilised to bind to a target protein. The antisense strategy is also a useful research tool for the identification of specific gene-protein functions. The first in vivo animal studies and clinical experiences have been carried out in the fields of cardiovascular medicine, oncology and virology yielding promising results (Culman 2000).

As we learn more about the development of corneal disease pharmaceutical intervention also becomes a more realistic prospect in the near future, particularly when patients who are pre-symptomatic can be identified through genetic testing.
REFERENCES


Linkage of congenital hereditary endothelial dystrophy to chromosome 20


Department of Ophthalmology, Royal London Hospital, London E1 1BB, UK

Received September 5, 1995; Revised and Accepted October 9, 1995

Congenital hereditary endothelial dystrophy (CHED) is a rare autosomal dominant disorder of the cornea. We have performed genetic linkage analysis with microsatellite markers on a seven generation British pedigree. Two-point linkage analysis revealed significant linkage of CHED (lod score >3) with seven marker loci mapping to chromosome 20. The highest observed lod score was 7.20 (θ = 0.026) with marker D20S114. Multipoint analysis gave a maximum lod score of 9.34 between D20S48 and D20S471. This 2.7 cM region lies within the 30 cM region recently assigned to posterior polymorphous dystrophy (PPD). PPD and CHED may therefore be allelic, or alternatively it is possible that more than one gene in this region is responsible for these two corneal dystrophies.

INTRODUCTION

Congenital Hereditary Endothelial Dystrophy (CHED) belongs to a group of disorders collectively known as the corneal endothelial dystrophies, in which the primary abnormality is attributed to abnormal development of the corneal endothelium. Other disorders in this group include Fuch’s endothelial dystrophy, Posterior polymorphous dystrophy (PPD) and the Iridocorneal Endothelial Syndrome (ICE) (1). CHED was first documented in the English literature by Maumenee in 1960 (2), although the condition had been described previously by several authors in the European literature (3–9). It is inherited in an autosomal dominant or autosomal recessive fashion, and is characterised clinically by diffuse corneal oedema and thickening of Descemet’s membrane affecting both eyes, usually symmetrically, presenting at or soon after birth. The clinical features remain stationary or progress slowly. The visual acuity is usually severely affected and the majority of patients require penetrating keratoplasties to obtain visual improvement (10).

A British family with CHED demonstrating autosomal dominant inheritance formed the basis of this genetic linkage study to identify the chromosomal location of the gene responsible for this disease. It is the largest family with autosomal dominant inheritance of CHED reported in the literature and was initially reported on by Pearce et al. (11) and further commented on by Kirkness et al. (10). Expressivity is constant in the family and the disease appears fully penetrant. Using a candidate locus strategy we have mapped the disease in this family to a 2.7 cM interval on chromosome 20, as follows.

RESULTS

A 129 member, seven generation family tree was constructed based on the information obtained from family members and previous pedigree studies. Forty family members were enrolled into a clinical and genetic linkage study after obtaining ethics committee approval. This included 19 affected and 13 unaffected subjects as well as eight spouses, as shown in Figure 1. A detailed description of the results of clinical examination in this family is in preparation. All the subjects assigned as affected in this study demonstrated the distinctive corneal changes recognised in CHED. General medical histories and a basic general physical examination failed to identify other medical conditions segregating with the disease.

Candidate loci were tested in the family, using markers tightly linked to each locus. These loci were implicated because they either showed linkage to or were genes known to be involved in other disorders affecting the anterior segment of the eye. They included the locus for Reiger’s syndrome on chromosome 4 (12), the chromosome 5 locus for lattice type I, granular and Avellino dystrophies (13), the locus for aniridia on chromosome 11 (14), the locus for type VIII collagen on chromosome 3 (15) and finally the locus on chromosome 20 to which PPD has recently been linked (16).

Linkage was identified on chromosome 20 with the same markers as have been linked to PPD. Significant positive two point lod scores (>3) were obtained with seven markers in this region. The two point maximum likelihood data for 10 markers from chromosome 20 are summarised in Table 1. The highest two point lod score was 7.20 and was obtained with D20S114 at θ = 0.026. Multipoint analysis with markers D20S114, D20S48 and D20S471 gave a maximum lod score of 9.34, in the interval D20S48 to D20S471 (Fig. 2), suggesting that the gene for CHED lies within the 2.7 cM interval between these two markers. This was confirmed by haplotype analysis in critical crossovers, as shown in Figure 1. The upper limit of the interval is defined by a crossover in individual IV-4 with markers D20S114 and D20S48. The lower limit is marked by a crossover involving the branches of the family to the right of Figure 1, with markers D20S471 and D20S54.

*To whom correspondence should be addressed
Figure 1. Abridged pedigree of the CHED family used in the study illustrating the segregation of four chromosome 20 loci. These are, from top to bottom, D20S114, D20S48, D20S471, D20S54. Squares and circles indicate males and females respectively. Solid symbols indicate affected and open symbols indicate non-affected members of the family. Slashed symbols show deceased individuals.

Table 1. Two-point lod scores for linkage between CHED and 10 markers on chromosome 20

<table>
<thead>
<tr>
<th>Marker</th>
<th>0.00</th>
<th>0.01</th>
<th>0.05</th>
<th>0.10</th>
<th>0.20</th>
<th>0.30</th>
<th>0.40</th>
<th>Z max</th>
<th>θ max</th>
</tr>
</thead>
<tbody>
<tr>
<td>D20S172</td>
<td>-∞</td>
<td>0.55</td>
<td>2.79</td>
<td>3.05</td>
<td>2.30</td>
<td>1.40</td>
<td>0.84</td>
<td>3.29</td>
<td>0.118</td>
</tr>
<tr>
<td>D20S112</td>
<td>-∞</td>
<td>2.99</td>
<td>3.28</td>
<td>3.05</td>
<td>2.30</td>
<td>1.45</td>
<td>0.84</td>
<td>3.28</td>
<td>0.042</td>
</tr>
<tr>
<td>D20S114</td>
<td>-∞</td>
<td>7.06</td>
<td>7.07</td>
<td>6.48</td>
<td>4.88</td>
<td>3.02</td>
<td>1.18</td>
<td>7.20</td>
<td>0.028</td>
</tr>
<tr>
<td>D20S48</td>
<td>-∞</td>
<td>5.89</td>
<td>5.95</td>
<td>5.42</td>
<td>3.98</td>
<td>2.36</td>
<td>0.88</td>
<td>6.05</td>
<td>0.028</td>
</tr>
<tr>
<td>D20S471</td>
<td>-∞</td>
<td>4.78</td>
<td>4.94</td>
<td>4.57</td>
<td>3.48</td>
<td>2.23</td>
<td>0.98</td>
<td>4.99</td>
<td>0.033</td>
</tr>
<tr>
<td>D20S54</td>
<td>-∞</td>
<td>0.60</td>
<td>2.26</td>
<td>2.61</td>
<td>2.33</td>
<td>1.82</td>
<td>0.73</td>
<td>2.82</td>
<td>0.115</td>
</tr>
<tr>
<td>D20S200</td>
<td>-∞</td>
<td>2.05</td>
<td>2.99</td>
<td>2.99</td>
<td>2.38</td>
<td>1.53</td>
<td>0.67</td>
<td>3.04</td>
<td>0.073</td>
</tr>
<tr>
<td>D20S106</td>
<td>-∞</td>
<td>5.34</td>
<td>5.58</td>
<td>5.06</td>
<td>3.89</td>
<td>2.58</td>
<td>1.27</td>
<td>5.53</td>
<td>0.031</td>
</tr>
<tr>
<td>SRC11B</td>
<td>-∞</td>
<td>1.08</td>
<td>2.18</td>
<td>2.35</td>
<td>2.00</td>
<td>1.38</td>
<td>0.68</td>
<td>2.35</td>
<td>0.097</td>
</tr>
<tr>
<td>D20S107</td>
<td>-∞</td>
<td>-1.31</td>
<td>1.11</td>
<td>1.82</td>
<td>1.96</td>
<td>1.50</td>
<td>0.77</td>
<td>2.01</td>
<td>0.161</td>
</tr>
</tbody>
</table>

DISCUSSION

This study demonstrates the localisation of the genetic mutation responsible for CHED to a 2.7 cM interval on chromosome 20. This lies within the 30 cM region on the same chromosome recently reported by Heon et al. to be linked to PPD (16). Although both CHED and PPD share clinical, histological and embryological similarities, significant differences between the phenotypic expression of these two disorders exist (1,17–20). CHED is characterised clinically by marked bilateral corneal oedema associated with thickening of Descemet's membrane presenting within the first few years of life. In its dominant form it is fully penetrant and demonstrates little variation in expressivity. All affected patients are symptomatic and the majority require corneal grafting to improve the visual outcome. Histologically the corneal endothelium demonstrates atrophy with vacuolation, focal absence of cells, multilayering and melanin deposition (10,11,21,22). In contrast PPD is a milder disease characterised clinically by the presence of vesicles and/or bands on the endothelial surface of the cornea. This is
quite often asymptomatic, although some cases do demonstrate marked corneal oedema requiring corneal grafting. PPD has been reported with other anterior segment anomalies including irido-corneal adhesions which may be associated with pupillary ectropion and corectopia. Histologically, the characteristic change in the corneal endothelium in PPD is its epithelialisation. This is associated with the development of multi-layered cells linked by desmosomes, with surface microvilli, and intracytoplasmic 10 nm filaments (19,20,22). The pathogenesis of both CHED and PPD is considered to be due to a primary dysfunction of corneal endothelium. This is precipitated by an abnormal terminal differentiation of secondary mesenchyme in the later stages of foetal development. This postulate is based on the abnormal appearance of Descemet's membrane on electron microscopy which demonstrates a normal anterior banded zone and abnormal posterior non-banded zone thus providing an archaeological record of the insults suffered by this membrane during its development (23). The severity and nature of the resulting endothelial dysfunction, which may be genetically controlled, ultimately determine the range of clinical signs that manifest in the form of CHED or PPD (1,21,24).

The establishment by linkage analysis of a common location for these phenotypically distinct disorders may be explained in several ways. It is possible that more than one gene in this region is responsible for these two dystrophies, perhaps implying a cluster of genes with related function. Alternatively, it is possible that the different phenotypes are allelic.

The identification of candidate genes for CHED and PPD is complicated by the apparently large physical region of chromosome 20 to which this region corresponds. Heon et al. placed PPD in the 30 cM interval between markers D20S98 and D20S108, and stated that this assigns the PPD locus to 20q12 (16). However the marker D20S18, not used in their analysis but clearly within this interval, has been assigned to 20p11.2 (25). The first marker in the literature clearly distal to PPD on the p arm is D20S27, at 20p12, while the first marker distal on the q arm is PLC1, which has been assigned to 20q12–13.1 (26). Thus based on published cytogenetic localisations PPD should more correctly be placed at 20p12-q13.1, which in fact includes most of the chromosome. Our own data, though providing a much more refined genetic localisation for CHED, does not provide any further insight into the physical assignment. For now, therefore it seems appropriate to merely refer to these loci as mapping to the pericentromeric region of chromosome 20.

The formation of the anterior segment of the eye requires the co-ordinated development of a number of tissues with vital roles to play in the establishment and maintenance of normal structure and function. Genetic mapping of several anterior segment disorders has been achieved including aniridia, for which the gene has been characterised (14), Rieger's syndrome (12), lattice dystrophy (13), Peter's anomaly (27), PPD (16) and now CHED. Cloning of the genes responsible for inherited anterior segment disorders will make a contribution to understanding the genetic control of the development of the eye, and aid in the diagnosis and treatment of human ocular disease.

**MATERIALS AND METHODS**

Blood samples were obtained from 19 affected members, 13 non-affected members and eight spouses for linkage analysis. Genomic DNA was extracted from these samples using a Nucleon II DNA extraction kit (Scotlab Bioscience). PCR amplification (28) from 100 ng DNA samples was performed utilising primers labelled with radioactive 32P. PCR conditions included 30 cycles under the following conditions: 94°C for 1 min, 55°C for 1 min, 72°C for 1 min. The products were separated by denaturing polyacrylamide gel electrophoresis and visualised by autoradiography.
Linkage analysis
Data was estimated using the LINKSYS (version 3.1) data management package (29). Two-point linkage analysis was performed using the MLINK (version 5.10) submodule of the LINKAGE package (30). Allele frequencies were obtained from the eight spouses in the family. Maximum lod scores were obtained using LINK. Multipoint analysis was performed with the LINKMAP program using the computing resources of the Human Genome Mapping Project. The CHED phenotype in this family was analysed as an autosomal dominant trait with complete penetrance, infantile onset and a gene frequency of 0.0001 for the affected allele.

ACKNOWLEDGEMENTS
We would like to thank Marcelle Jay for her assistance with pedigree information and Noel Rice for his advice and support for this project. We would also like to thank the HGMP Resource Centre for on line bio-informatics facilities. This work was generously supported by a Locally Organised Research Scheme (LORS) grant, Moorfields Eye Hospital. CFl is a Wellcome Senior Research Fellow (Grant No. 035535/Z/92/Z).

ABBREVIATIONS
CHED: Congenital hereditary endothelial dystrophy; PPD: Posterior polymorphous dystrophy.

REFERENCES


