A molecular genetic study of X-linked retinal diseases

Ilaria Zito
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

2001
Declaration

I declare that this 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.

Ilaria Zito B. Sc.
Abstract

The X chromosome is home to a large number of genes involved in retinal disease. Two diseases studied in detail in this thesis are X-linked retinitis pigmentosa (XLRP) and X-linked congenital stationary night blindness (CSNBX). XLRP and CSNBX share some clinical features in that patients are initially diagnosed as night blind, however XLRP is a progressive degenerative disease which eventually results in blindness, whilst CSNBX is a milder disease with no disease progression. At the outset of this study four distinct XLRP loci had been described on the short arm of the X-chromosome (RP2, RP3, RP6 and RP15) and the RPGR gene at the RP3 locus had been identified. Linkage data also suggested the presence of at least two CSNBX genes on the short arm of the X chromosome, potentially allelic with XLRP loci.

Using haplotype and linkage analysis, the clinical resource available was utilised to establish the level of genetic heterogeneity for these diseases and define the proportion of families mapping to each locus. As disease genes were identified, the mutation spectrum in the patient pool was established for RPGR, RP2 and NYX. Novel mutations were identified for RPGR and RP2 accounting for approximately 30% of disease in the XLRP families studied. In addition, sequence analysis of the RPGR gene provided evidence for a non-pathogenic founder complex allele in the population. Screening for NYX mutations revealed that mutations in this gene are the most common cause of CSNBX, and haplotype analysis around common mutations confirmed a founder mutation in three of the families studied.

Further genetic heterogeneity for XLRP has been established by the identification of a new locus in Xp22, RP23, and a positional cloning strategy has been employed towards the identification of the disease gene. Eight genes were physically mapped to Xp22 and these were screened for mutations. A search for new genes, using bioinformatics to detect ESTs, led to the characterisation of several potential transcripts.

A family clinically identified as XLRP with additional symptoms, including nonsensorineural deafness, recurrent chest infections and sinus infections, suggested an associated cilia abnormality as the cause of disease. Haplotype analysis revealed disease segregation at the RP3 locus and an RPGR mutation was subsequently identified. This is the first description of such a phenotype and associated genotype and its implications are discussed.
Acknowledgements

My first words are for my supervisor, Professor Shomi Batthacharya. Thank you for believing in me in the first place and for giving me this tremendous opportunity, and also for enlightening us with your linkage knowledge and for your kind words and warm hugs.

My biggest thanks go to my day-to-day ‘boss’ Dr. Alison Hardcastle. Thank you for making me live this PhD as a fantastic experience, and for your constant support and guidance, for your optimism, your enthusiasm and your friendship. You have been a real source of richness for me.

My next special thanks go to two very special friends. Firstly Dawn, for your sincere and unique friendship, for supporting me and helping me whenever I was in need and for seeking my help whenever you were in need. You really are a special person. Secondly Resh, for lightening up the lab and the office, for your invaluable help with the thesis, for our nice chats and for always being so cheerful and caring.

I wish to thank Naheed for sharing my passion for contemporary dance, Louise for being the sweetest girl on earth, and two key elements in the lab, Neil and Steph for being just as they are. Thanks to Zara for some nice chats in Italian and to Bart the fun Belgian guy. To Christiane and Uli for their fantastic laughs, to Simon for being so nice, to Alex for being so English and to Andrew for his medical expertise. Thanks to Eranga and Sana for telling me all about their fascinating cultures and to Vinny for her hugs. A big thank you to a big friend, Sam, who is now enjoying his life in Greece and finally a very very special thank you to Phillipa, Suba, Mel, Suzanne, Jill, Miranda, Inez, Clare, Sue, Sam and Marie for making lunch-time in our tea-room more pleasant with chats, food lovers clubs and with all the crosswords done together. And thanks again to Dawn, Jill and Neil and Quincy for their help with the thesis.

And now to my family and Italian friends. First of all an incredible thank you to my parents and my brother for your constant caring, your unconditional love, for being wonderful, for blindly thinking that your daughter/sister is a genius and because I love you very much. To my huge and incredibly special family in law, and particularly to our 9 nieces and nephews for being so wonderful and for making the return back home more pleasant. And thanks to all my friends who are still there after these 3 years abroad, to Danila for our weekly phone calls and our mutual support, to Ale, Ale and Paola, to Stefano and Beatrice and to all the others...you know who you are.

Lastly my most special thank you goes to my husband Filippo. Thanks for your constant support, your encouragement, your help, your taking me through all this day-by-day, I couldn’t have done it without you. Thank you because these 3 years in London just the two of us have been the most wonderful experience of my life, and thank you for doing even the impossible to make me happy, for your incredible and very very special love and for being the most special person in the whole world.
Dedication

This thesis is dedicated to Filippo, my source of strength and happiness, and to our child who is growing inside of me.
Publication list

Publications


Abstracts


Holder GE, Kambanarou SA, Fitzke FW, Bird AC, Zito I, Hardcastle AJ, Webster AR (2001) Congenital stationary night blindness (CSNB) and a ‘Schubert-Bornshein’ type
electrophysiology in a family with dominant inheritance. *Invest Ophthalmol Vis Sci* Program Nr. 428


**Oral presentations**


Table of contents

CHAPTER 1
General Introduction

1.1 Human retina structure, function and disease 20
   1.1.1 Structure of the retina 20
      1.1.1.1 The neural retina 22
      1.1.1.1.1 Photoreceptor cells 25
      1.1.1.2 The retinal pigment epithelium 27
   1.1.2 The phototransduction cascade 27
      1.1.2.1 The recovery phase 29
   1.1.3 Photoreceptor genes and inherited retina diseases 30
   1.1.4 Retina diseases on the X chromosome 32
      1.1.4.1 Retinitis pigmentosa 33
         1.1.4.1.1 Genetic heterogeneity in XLRP 38
         1.1.4.1.2 The disease gene at the RP2 locus 40
         1.1.4.1.3 The disease gene at the RP3 locus 42
      1.1.4.2 Congenital stationary night blindness 45
         1.1.4.2.1 The disease gene at the CSNB1 locus 46
         1.1.4.2.2 The disease gene at the CSNB2 locus 47
      1.1.4.3 Cone dystrophy 48
   1.1.4.4 Other X-linked retinopathies 48

1.2 The human genome project 52

1.3 Paths to human disease gene identification 53
   1.3.1 Positional cloning 54

1.4 Genetic mapping 56
   1.4.1 Meiotic recombination and genetic map distances 56
   1.4.2 DNA markers as a genetic mapping tool 57
   1.4.3 Linkage analysis 59
      1.4.3.1 Haplotype analysis 60

1.5 Physical mapping 61
1.5.1 Low resolution physical mapping 61
  1.5.1.1 Somatic cell hybrid 61
    1.5.1.1.1 Radiation hybrid mapping 61
  1.5.1.2 In situ hybridisation 63
1.5.2 High resolution physical mapping 63
  1.5.2.1 Long-range restriction mapping 64
  1.5.2.2 Clone contig assembly 64
    1.5.2.2.1 Yeast artificial chromosomes versus bacterial host systems 65

1.6 Methods of gene identification 66
  1.6.1 Detection of CpG islands 66
  1.6.2 Exon trapping 67
  1.6.3 cDNA selection 67
  1.6.4 Computer-based DNA sequence analysis 67
  1.6.5 Expressed sequence tags and full-length cDNAs 68

1.7 Aims of this thesis 70

CHAPTER 2
Materials and methods

2.1 DNA isolation 71
  2.1.1 Human genomic DNA preparation from blood samples 71
  2.1.2 Plasmid DNA preparation (QIAprep spin miniprep, QIAGEN) 72
  2.1.3 Yeast DNA preparation 72

2.2 Purification of DNA 73
  2.2.1 Phenol chlorophorm extraction 73
  2.2.2 Ethanol precipitation 73
  2.2.3 Sephacryl-S200 and S400 HR columns (Pharmacia) 74
  2.2.4 Centricon 100 spin columns (Centricon) 74
  2.2.5 The enzymatic method (ExoI and SAP) 74
  2.2.6 QIAquick spin columns (QIAGEN) 75
2.2.6.1 QIAquick PCR purification 75
2.2.6.2 QIAquick gel extraction 75

2.3 Restriction enzyme digestion of DNA 76

2.4 Size fractionation of DNA 76
  2.4.1 Agarose gel electrophoresis 76
  2.4.2 Polyacrylamide gel electrophoresis 77
  2.4.3 MDE gel electrophoresis 78

2.5 RNA manipulation 78
  2.5.1 RNA handling 78
  2.5.2 Isolation of total RNA from fibroblast cell lines 79
    2.5.2.1 Harvesting of cells 79
    2.5.2.2 RNeasy mini protocol (QIAGEN) 79

2.6 Polymerase chain reaction (PCR) 80
  2.6.1 Standard parameters for a typical PCR 80
  2.6.2 Primer design 81
  2.6.3 PCR optimisation 81
  2.6.4 Microsatellite and SSCP analysis 82
  2.6.5 Reverse Transcriptase (RT) PCR 83
  2.6.6 YAC library screening by PCR 83
  2.6.7 Colony PCR 84

2.7 Automated DNA sequencing 84

2.8 Computer-aided analysis 85
  2.8.1 Linkage analysis 85
  2.8.2 Sequence similarity searches 85
  2.8.3 DNA analysis 86
  2.8.4 Gene analysis 86
  2.8.5 Protein analysis 86

2.9 Reagents and buffers 87
  2.9.1 General reagents 87
  2.9.2 DNA isolation 87
  2.9.3 Gel electrophoresis 88
CHAPTER 3

Haplotype and linkage analysis in families with X-linked retina disease

3.1 Introduction

3.1.1 X-linked retinitis pigmentosa
3.1.2 X-linked congenital stationary night blindness

3.2 Materials and methods

3.2.1 Families
3.2.1.1 XLRP families
3.2.1.2 CSNB families
3.2.2 Genetic markers
3.2.3 Genotyping
3.2.4 Haplotype analysis
3.2.5 Linkage analysis

3.3 Results

3.3.1 XLRP families
3.3.1.1 Families segregating with the RP3 locus
3.3.1.2 Families segregating with RP3 and RP2
3.3.1.3 Families segregating with RP3 and RP23
3.3.1.4 Families segregating with RP2 and RP24
3.3.1.5 Families segregating with RP3, RP2 and RP24
3.3.1.6 Families segregating with RP3, RP2 and RP23
3.3.1.7 Families segregating with RP3, RP2, RP24 and RP23
3.3.1.8 Linkage analysis

3.3.2 CSNBX families
3.3.2.1 Haplotype analysis for four CSNBX families
3.3.2.2 Linkage analysis

3.4 Discussion
CHAPTER 4

Mutation screening of the RPGR, RP2 and NYX genes

4.1 Introduction

4.1.1 The RPGR gene

4.1.2 The RP2 gene

4.1.3 The NYX gene

4.1.4 Allelism in X-linked retinopathies

4.2 Materials and methods

4.2.1 Patients

4.2.1.1 Patients screened for the RPGR gene

4.2.1.2 Patients screened for the RP2 gene

4.2.1.3 CSNB patients screened for the RPGR and RP2 gene

4.2.1.4 Patients screened for the NYX gene

4.2.1.5 Controls

4.2.2 Mutation detection

4.2.2.1 PCR conditions

4.2.2.2 Single-strand conformation polymorphism analysis

4.2.2.3 Direct automated sequencing of PCR products

4.2.2.4 Restriction enzyme digestion analysis

4.2.2.5 Sequence analysis

4.3 Pathogenic mutations and polymorphic variants of the RPGR gene

4.3.1 RPGR mutations and polymorphisms identified by SSCP analysis in a subset of 27 XLRP families segregating with the RP3 locus

4.3.2 Additional RPGR mutations and polymorphisms identified by direct sequencing in a subset of 10 XLRP families segregating with the RP3 locus

4.3.3 Sequence variations within the RPGR gene: evidence for a founder complex allele

4.4 Pathogenic mutations and polymorphic variants of the RP2 gene

4.4.1 RP2 mutations identified by direct sequencing in a subset of 33 XLRP families
4.4.2 Novel frameshift mutations and polymorphisms identified by SSCP analysis in RP2 in a subset of 21 XLRP families 162

4.5 Pathogenic mutations and polymorphic variants of the NYX gene 165
   4.5.1 Gene analysis 165
   4.5.2 Mutations identified by direct sequencing in the NYX gene 165

4.6 Screening of RPGR and RP2 in CSNB families 171

4.7 Discussion 171
   4.7.1 RPGR 171
   4.7.2 RP2 175
   4.7.3 NYX 176

CHAPTER 5
Towards the identification of novel X-linked retina disease genes

5.1 Introduction 180
   5.1.1 Physical mapping and positional cloning 180
      5.1.1.1 Yeast Artificial Chromosomes (YACs) 181
      5.1.1.2 Human YAC libraries 181
      5.1.1.3 Contig assembly 182
   5.1.2 The phenomenon of X inactivation 182
   5.1.3 The genomic region of Xp11.4 183
   5.1.4 The genomic region of Xp22 184

5.2 Materials and methods 184
   5.2.1 Clinical assessment 184
   5.2.2 Computer-aided analysis 184
   5.2.3 YAC library screening by PCR 184
   5.2.4 Haplotype and linkage analysis 185
   5.2.5 Mutation screening 185
   5.2.6 EST characterization 188
   5.2.7 X inactivation studies 190

5.3 Results 192
   5.3.1 The CSNB1 locus 192
 CHAPTER 6
RP in association with other phenotypes

6.1 Introduction 227
   6.1.1 RP in syndromic forms 227
      6.1.1.1 Usher syndrome 227

6.2 Materials and methods 228
   6.2.1 Clinical assessment 228
   6.2.2 Genotype analysis 229
   6.2.3 Mutation screening 229

6.3 Results 229
   6.3.1 Family phenotype 229
   6.3.2 Family genotype 230
   6.3.3 Mutation screening of the RP2 and RPGR genes 230

6.4 Discussion 233
CHAPTER 7

Overall discussion and future prospects

7.1 Overview of work presented 236
7.2 Unique features of different genes 239
7.3 The importance of genotype-phenotype correlation 240
7.4 Photoreceptors cell death and animal models 241

REFERENCES 244
# List of Figures

<table>
<thead>
<tr>
<th>Figure No.</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Cross section of the human eye.</td>
<td>21</td>
</tr>
<tr>
<td>1.2</td>
<td>A schematic representation of a portion of the human retina showing different cell types and layers.</td>
<td>23</td>
</tr>
<tr>
<td>1.3</td>
<td>A fundus image of a human retina.</td>
<td>24</td>
</tr>
<tr>
<td>1.4</td>
<td>Schematic diagram of a rod photoreceptor cell.</td>
<td>26</td>
</tr>
<tr>
<td>1.5</td>
<td>Schematic representation of the phototransduction cascade.</td>
<td>28</td>
</tr>
<tr>
<td>1.6</td>
<td>Graph showing the increasing number of mapped and cloned retinal disease genes in the period between 1980 and 2001.</td>
<td>31</td>
</tr>
<tr>
<td>1.7</td>
<td>Ideogram of the human X chromosome showing map locations of eye diseases.</td>
<td>34</td>
</tr>
<tr>
<td>1.8</td>
<td>A scene viewed from a person with normal vision and a person with retinitis pigmentosa.</td>
<td>35</td>
</tr>
<tr>
<td>1.9</td>
<td>Fundus photographs of normal and diseased retina.</td>
<td>37</td>
</tr>
<tr>
<td>1.10</td>
<td>Coding and amino acid sequence of the \textit{RP2} gene.</td>
<td>41</td>
</tr>
<tr>
<td>1.11</td>
<td>cDNA and amino acid sequence of the \textit{RPGR} gene.</td>
<td>43</td>
</tr>
<tr>
<td>1.12</td>
<td>Schematic representation showing the key features of positional cloning.</td>
<td>55</td>
</tr>
<tr>
<td>1.13</td>
<td>Multiple levels of human chromosome mapping.</td>
<td>62</td>
</tr>
<tr>
<td>3.1</td>
<td>Ideogram of the human X chromosome showing map locations of XLRP and CSNB loci.</td>
<td>92</td>
</tr>
<tr>
<td>3.2</td>
<td>9 XLRP pedigrees segregating with the RP3 locus and associated genetic haplotypes.</td>
<td>100-104</td>
</tr>
<tr>
<td>3.3</td>
<td>3 XLRP pedigrees segregating with the RP3 and RP2 loci and associated haplotypes.</td>
<td>109-110</td>
</tr>
<tr>
<td>3.4</td>
<td>6 XLRP pedigrees segregating with the RP3 and RP23 loci and associated haplotypes.</td>
<td>112-114</td>
</tr>
<tr>
<td>3.5</td>
<td>3 XLRP pedigrees segregating with the RP2 and RP24 loci and associated haplotypes.</td>
<td>117</td>
</tr>
<tr>
<td>3.6</td>
<td>7 XLRP pedigrees segregating with the RP3, RP2 and RP24 loci and associated haplotypes.</td>
<td>119-121</td>
</tr>
</tbody>
</table>
3.7 3 XLRP pedigrees segregating with the RP3, RP2 and RP23 loci and associated haplotypes. 124
3.8 2 XLRP pedigrees segregating with the RP3, RP2, RP24 and RP23 loci and associated haplotypes. 126
3.9 4 CSNBX pedigrees and associated haplotypes. 134-135
4.1 Ideogram of the genomic structure of the RPGR, RP2 and NYX genes. 141
4.2 Ideogram of the short arm of the X chromosome showing map location of XLRP and CSNB loci. 145
4.3 Electropherograms depicting patient mutations and normal sequences in the RPGR gene. 152
4.4 Segregation analysis of two novel RPGR mutations. 153
4.5 Electropherograms depicting patient mutations and normal sequences in the RPGR gene. 155
4.6 Segregation analysis of two novel RPGR mutations. 156
4.7 Electropherograms depicting patient sequence variation in exon 6 and exon 14 of RPGR in NZ1 family and corresponding normal sequences. 159
4.8 Pedigree of the NZ1 family showing segregation of the IVS6+5G>A mutation with disease in the family and ideogram of the RPGR gene showing the complex allele and wild-type sequence. 160
4.9 Electropherograms depicting patient mutations and normal sequences in the RP2 gene. 163
4.10 Segregation analysis of two novel RP2 mutations. 164
4.11 Coding and predicted amino acid sequence of the NYX gene. 166
4.12 Electropherograms depicting patient mutations and normal sequences in the NYX gene. 168
4.13 Agarose gel showing segregation of the 355bp deletion at nucleotide Position 1122 to 1457 of the NYX gene in family RP2124. 170
4.14 Ideogram of the genomic structure and mutations identified in the RPGR, RP2 and NYX genes. 172
5.1 Ideogram of the human X-chromosome showing an expanded view of the CSNB region. 193
5.2 Physical map of the CSNBl region.  
5.3 Pedigree and haplotype analysis of family MGRP3.  
5.4 Physical map of the RP23 region.  
5.5 Agarose gels showing PCR experiments performed with different transcripts on several YAC clones.  
5.6 Agarose gel of restriction enzyme analysis of clone 609319.  
5.7 Sequence of clone 609319.  
5.8 Schematic representation of relative positions of IMAGE clones 23809 and 30615, EST W26143, PAC clone AC003666 and KIAA0455.  
5.9 Agarose gel of PCR experiments with primers 23809RH and W26143LH.  
5.10 Results of the NYX analysis carried out on part of PAC AC03666 sequence spanning exon 10 and 11 of the CLCN4 gene.  
5.11 Agarose gel of PCR experiments for different transcripts on human and mouse.  

6.1 Pedigree of family RP4462 showing haplotypes generated with Microsatellite markers on the X chromosome.  
6.2 Electropherograms depicting patient mutation and control sequence in exon 8 of the RPGR gene.
## List of Tables

<table>
<thead>
<tr>
<th>Table No.</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>RP genes cloned to date.</td>
<td>36</td>
</tr>
<tr>
<td>3.1</td>
<td>Microsatellite markers used for genetic characterisation of XLRP and CSNB families.</td>
<td>95-96</td>
</tr>
<tr>
<td>3.2</td>
<td>Other markers used for genetic characterisation of XLRP and CSNB families.</td>
<td>96</td>
</tr>
<tr>
<td>3.3</td>
<td>Integration of haplotype data for 33 XLRP families analysed.</td>
<td>99</td>
</tr>
<tr>
<td>3.4</td>
<td>Two-point lod scores for linkage between XLRP and four different markers for each of five XLRP families.</td>
<td>128-129</td>
</tr>
<tr>
<td>3.5</td>
<td>Integration of haplotype data for four X-linked CSNB families analysed.</td>
<td>132</td>
</tr>
<tr>
<td>3.6</td>
<td>Two-point lod scores for linkage between X-linked CSNB and four different markers for family RP2124.</td>
<td>136</td>
</tr>
<tr>
<td>4.1</td>
<td>Primer pairs for the two coding exons of the NYX gene.</td>
<td>148</td>
</tr>
<tr>
<td>4.2</td>
<td>Mutations identified in the RPGR gene in XLRP patients with an RP3 genotype.</td>
<td>150</td>
</tr>
<tr>
<td>4.3</td>
<td>Polymorphisms identified in the RPGR gene.</td>
<td>150</td>
</tr>
<tr>
<td>4.4</td>
<td>Sequence variations identified in RPGR in family NZ1.</td>
<td>157</td>
</tr>
<tr>
<td>4.5</td>
<td>Mutations identified in the RP2 gene in eight XLRP families.</td>
<td>161</td>
</tr>
<tr>
<td>4.6</td>
<td>Polymorphisms identified in the RP2 gene.</td>
<td>161</td>
</tr>
<tr>
<td>4.7</td>
<td>Mutations identified in the NYX gene in ten CSNBX families.</td>
<td>167</td>
</tr>
<tr>
<td>4.8</td>
<td>Haplotypes generated around the NYX gene.</td>
<td>169</td>
</tr>
<tr>
<td>5.1</td>
<td>Primer pairs for amplification and sequencing of the CLCN4 gene.</td>
<td>186</td>
</tr>
<tr>
<td>5.2</td>
<td>Primer pairs for amplification and sequencing of the TBL1 gene.</td>
<td>186</td>
</tr>
<tr>
<td>5.3</td>
<td>Primer pairs for amplification and sequencing of the CALB3 gene.</td>
<td>187</td>
</tr>
<tr>
<td>5.4</td>
<td>Primer pairs for amplification and sequencing of the RAI2 gene.</td>
<td>187</td>
</tr>
<tr>
<td>5.5</td>
<td>Two-point linkage analysis between XLRP and eight microsatellites on Xp.</td>
<td>198</td>
</tr>
<tr>
<td>5.6</td>
<td>Intron/exon boundaries of the CLCN4 gene.</td>
<td>204</td>
</tr>
</tbody>
</table>
CHAPTER 1

General Introduction

1.1 Human retina structure, function and disease

Although all parts of the eye are important for perceiving a good image, a critical layer for vision is the retina, which receives direct stimulation. Images are focused onto the retina by the cornea and lens, and the neural retina transmits this information to the brain. The retina is an outgrowth of the central nervous system deriving from the neural tube.

The human eye consists of several structures; the pupil, a black-looking aperture, which appears dark because of the absorbing pigments in the retina, which allows light to enter the eye; the iris, a coloured circular muscle, which controls the size of the pupil so that more or less light, depending on conditions, is allowed to enter the eye; the cornea, a transparent external surface covering both the pupil and the iris, which allows, together with the crystalline lens, the production of a sharp image at the retinal photoreceptor level; and the sclera, the "white of the eye", which forms part of the supporting wall of the eyeball and is continuous with the cornea.

A cross-sectional view of the eye shows three different layers (Figure 1.1), (i) the external layer, formed by the sclera and cornea, (ii) the intermediate layer, divided into two parts: anterior (iris and ciliary body) and posterior (choroid) and (iii) the internal layer, or the sensory part of the eye, the retina.

1.1.1 Structure of the retina

The human retina lies at the back of the eye (Figure 1.1), is approximately 0.5 mm thick and consists of two distinct structural and functional layers, namely the inner neural retina, and the outer pigmented layer or retinal pigment epithelium (RPE) (Figure 1.2), linked by the interphotoreceptor matrix (IPM) (Bridges 1985).
**Figure 1.1:** Cross section of the human eye (Picture taken from webvision, http://webvision.med.utah.edu/).
1.1.1.1 The neural retina

The neural retina is composed of 5 distinct layers (Figure 1.2), three layers of cell bodies (the outer nuclear layer, the inner nuclear layer and the ganglion cell layer), connected by two layers of synaptic connections (the outer and inner plexiform layers). The outer nuclear layer contains cell bodies of photoreceptor cells (rods and cones), the inner nuclear layer contains cell bodies of bipolar, horizontal and amacrine cells and the ganglion cell layer contains cell bodies of the ganglion cells. Photoreceptors, bipolar cells and ganglion cells together form a connecting sequence that carries the light signal to the optic nerve fibers leading to the brain. The remaining neural cells, the amacrine and horizontal cells, form lateral synapses that connect the three vertical neural cell populations, therefore creating further networks to influence and integrate the message being transmitted across the retina. Interplexiform cells transmit impulses back to the distal retinal layers, therefore collaborating to modify the signal by feedback (Bridges 1985). Thus the visual signal flows from photoreceptors to an intricate array of neural cells prior to leaving the eye through the axons of ganglion cells. These join to form the optic nerve, which passes on the signal to the visual cortex of the brain for further processing and interpretation. Within the retina the neurons are bound together by glial cells known as Muller cells (Chan-Ling 1994), whose fibres extend to form the outer and inner limiting membranes of the retina whilst providing a mechanical support for the retina.

The optic disc is situated in the centre of the retina (Figure 1.3) demarcating the site of entry of the major blood vessel of the retina and the site of departure of the optic nerve. The optic disc is also known as the blind spot due to lack of visual receptors and is therefore insensitive to light. To the left of the disc (Figure 1.3) a slightly oval shaped, reddish spot can be seen, this is the fovea, which is at the centre of an area known as the macula (Hendrickson and Yuodelis 1984). Central retina close to the fovea is considerably thicker than peripheral retina (Yamada 1969). This is due to the increased packing density of photoreceptors, particularly the cones, and their associated bipolar and ganglion cells in central retina compared with peripheral retina.

The centre of the fovea is known as the foveal pit (Polyak 1941) and is a highly specialised region of the retina. The foveal pit is an area where cone photoreceptors are
**Figure 1.2:** A schematic representation of a portion of the human retina showing different cell types and layers. (Picture taken from webvision)
Figure 1.3: A fundus image of a human retina. (Picture taken from webvision)
concentrated at maximum density with exclusion of the rods, and arranged at their most
efficient packing density. This is the thinnest part of the retina and the area of most acute
vision.

1.1.1.1 Photoreceptor cells

Packed in the outer layer of the retina are photoreceptor cells, the rods and cones,
which are specialised neurons that function in the primary event of vision. Rod cells are
highly sensitive to light and mediate dim light (scotopic) vision and contrast sensitivity.
Cones are responsible for bright light (photopic) vision, colour sense and fine discrimination
(reviewed in McNaughton 1990). In humans only 5% of photoreceptor cells are cones and
95% are rods (Curcio et al. 1987). Central retina is cone-dominated whereas peripheral retina
is rod-dominated, thus rods are essential for peripheral vision and cones are required for
central vision. The light sensitive chromophore 11-cis-retinal is present in all photoreceptor
cells and is bound to the visual pigment rhodopsin (peak absorption ($\lambda_{max}$) ~500 nm) in rods
and to red, green or blue opsins with $\lambda_{max}$ at 560 nm (long wavelength light), 530 nm
(medium) and 420 nm (short) respectively in different types of cones with different spectral
sensitivities (Nathans et al. 1986).

Both photoreceptors are elongated cells consisting of several morphologically and
functionally distinct regions. A schematic representation of a vertebrate rod photoreceptor
cell is shown in Figure 1.4. The outer segment is located adjacent to the RPE and is a
specialised compartment where the process of phototransduction takes place (section 1.1.2).
The outer segment is constantly renewed in a process in which newly synthesised membrane
is added at the base of the outer segment whilst packets of membrane at the tip of the outer
segment are phagocytosed by the RPE (Young and Bok 1969). Along the axial array of the
outer segment hundreds of disc-like membranes containing predominantly visual pigment
(opsin) molecules, can be observed. In rods the discs are constantly generated by
invaginations of the basal part of the plasma membrane and subsequent separation from the
plasma membrane as new invaginations take place (Steinberg et al. 1980). Therefore the disc
membrane separates an intradiscal space from the cytoplasm. In cones the discs do not
separate from the plasma membrane, which appears as a highly convoluted, continuous
entity. The photoreceptor's inner segment is a cellular compartment rich in mitochondria,
Figure 1.4: Schematic diagram of a rod photoreceptor cell. (Adapted from webvision)
endoplasmic reticulum, ribosomes and Golgi apparatus, ensuring highly active protein synthesis (Papermaster et al. 1985). A thin non-motile cilium connects the two segments and constantly transports components from the inner to the outer segment. Adjacent to the inner segment is the cell body containing the nucleus, which extends in the synaptic region where the electrical signal is transmitted to other retinal neurons.

### 1.1.1.2 The retinal pigment epithelium

The RPE is a single layer of cells adjacent to the outer segments of photoreceptors, which separates the retina from the choroidal circulation, the blood supply for the outer retina. Each RPE cell contacts approximately 45 outer segments and one of the functions of the RPE is the phagocytosis of the distal 10% of each outer segment daily (Young and Bok 1969). This phenomenon makes the RPE cell the most active phagocytic cell in the human body. Among the other functions carried out by the RPE is the synthesis of 11-cis retinal, which is then transferred to the photoreceptors for visual pigment regeneration, the active uptake of nutrients from the choriocapillaries, which are then supplied to the photoreceptors to maintain their metabolic requirements, and the active transport of ions accompanied by water from the subretinal space (Bok 1985).

### 1.1.2 The phototransduction cascade

The process of phototransduction is initiated when a photon of light is absorbed by a pigment molecule in a photoreceptor cell (Yau 1994). The process is schematically represented in Figure 1.5 for the vertebrate rod and reviewed in Molday (1998) and Yau (1994). The photoactivation of the visual pigment, which in this case is rhodopsin, involves the isomerisation of its bound 11-cis form of the chromophore retinal to 11-trans. The conformational change of rhodopsin exposes a binding site for the trimeric G-protein transducin. Interaction between the activated rhodopsin and transducin catalyses the exchange of GDP to GTP on the surface of transducin and the release of the α subunit of transducin charged with GTP (Tα-GTP). The rhodopsin molecule is now free to bind a new transducin molecule and it is estimated that each molecule of rhodopsin can bind up to 500 molecules of transducin, providing a primary amplification step of the visual signal. The free Tα-GTP proceeds to activate the enzyme cGMP phosphodiesterase (PDE), found on the
Figure 1.5: Schematic representation of the phototransduction cascade.
cytosolic surface of the disc membrane. The enzyme consists of two catalytic subunits, \( \alpha \) and \( \beta \), and two inhibitory \( \gamma \) subunits. Interaction between \( \text{T}\alpha\)-GTP and PDE promotes the activation of PDE by displacement of the two \( \gamma \) subunits. Activated PDE rapidly hydrolyses cGMP, which is bound to the cGMP-gated channel protein, to 5’GMP. This step provides the second amplification signal, as one activated PDE molecule is able to hydrolyse thousands of cGMP per second. In the dark, the transmembrane cGMP-gated channels allow \( \text{Na}^+ \) and \( \text{Ca}^{2+} \) to enter the outer segment through the plasma membrane. The rapid fall of cGMP levels in the cytoplasm on exposure to light results in the closure of the channel and hyperpolarisation of the photoreceptor plasma membrane due to the reduction of \( \text{Na}^+ \) and \( \text{Ca}^{2+} \) influx. The hyperpolarisation causes a reduction in the rate of transmitter release from the photoreceptor’s synapse generating a signal, which is passed on to the bipolar and ganglion cells of the retina.

1.1.2.1 The recovery phase

The recovery of the dark state involves active turnoff and recycling of the transduction components (reviewed in Molday 1998 and Yau 1994). To block PDE activity, \( \text{T}\alpha\)-GTP needs to be released from the \( \gamma \) subunit inhibitors of PDE, this is achieved by the intrinsic hydrolytic activity of \( \text{T}\alpha\). \( \text{T}\alpha\)-GDP has low affinity for the PDE inhibitors, which are now free to regenerate the stationary (inactive) state of PDE by rebinding to the catalytic subunits. \( \text{T}\alpha\)-GDP is also free to recombine with the \( \beta \) and \( \gamma \) subunits of transducin.

Rhodopsin needs to be regenerated by recombining with 11-cis retinal. The deactivation of rhodopsin results from a combination of phosphorylation by the enzyme rhodopsin kinase and binding of arrestin to phosphorylated rhodopsin, therefore inhibiting its association with transducin. All-trans retinal is released and recycled.

The complete reset to the dark-adapted state appears to be highly dependent on intracellular \( \text{Ca}^{2+} \) levels, which act at several different levels. In the dark state, \( \text{Ca}^{2+} \) enters the cell through the cGMP-gated channels and \( \text{Ca}^{2+} \) homeostasis is maintained through the concomitant action of a \( \text{Ca}^{2+}-\text{Na}^+ \) antiporter in the plasma membrane. Upon exposure to light, \( \text{Ca}^{2+} \) influx through the cGMP-gated channel is blocked, whilst its extrusion through the \( \text{Ca}^{2+}-\text{Na}^+ \) exporter is maintained, resulting in an overall drop in the intracellular \( \text{Ca}^{2+} \) level. One of the responses to the low intracellular \( \text{Ca}^{2+} \) level is the activation of photoreceptor-
specific guanylate cyclase activating proteins (GCAPs), which in turn activate guanylate cyclase, responsible for the synthesis of cGMP from 5'GMP, thus promoting the reopening of the cGMP-gated channel. A second pathway dependent on the Ca$$^{2+}$$ level is mediated by the Ca$$^{2+}$$ binding protein recoverin, which at low Ca$$^{2+}$$ levels, promotes the phosphorylation of activated rhodopsin. Finally, a fall in Ca$$^{2+}$$ also increases the cGMP-gated channel affinity for cGMP so that some of the channels reopen despite the decrease in concentration of cGMP.

In cones phototransduction is thought to be similar to that in rods, and many key rod proteins appear to have cone counterparts.

1.1.3 Photoreceptor genes and inherited retina diseases

Considering the complex structure and the neurodevelopmental origins of the human eye, it is not surprising that the eye is a common site of genetic disease. In fact, genetic eye diseases, both monogenic and genetically complex, are a significant cause of blindness in the industrialised world. Retinal diseases are particularly important, as it has been estimated that approximately 1 in 2000 people worldwide is affected by one of the known monogenic retinopathies (Rattner et al. 1999). In 1990 the first mutation in a phototransduction cascade gene accounting for a human retina disease was reported (Dryja et al. 1990), the gene responsible was rhodopsin (RHO) and the disease was the autosomal dominant form of retinitis pigmentosa (adRP). Since then, many other mutations in the rhodopsin gene, accounting for both the dominant and the recessive form of RP and dominant congenital stationary night blindness (CSNB), have been reported, making RHO the most common gene accounting for adRP with an estimated prevalence varying between 25 and 50% (Inglehearn 1998, Sohocki et al. 2001). In addition, many other genes and loci involved in RP, CSNB and other retinal disorders have been identified through positional cloning and candidate gene approaches (reviewed in Clarke et al. 2000, Gregory-Evans and Bhattacharya 1998, Inglehearn 1998, Molday 1998 and Rattner et al. 1999). Figure 1.6 shows an ideogram of the progress made in the identification of new loci and genes involved in inherited retinal diseases in the past 20 years. To date 126 loci accounting for different retinal disorders have been reported and 68 genes have been cloned, and as Figure 1.6 suggests, these numbers are constantly changing. Many of the genes identified code for
Figure 1.6: Graph showing the increasing number of mapped and cloned retinal disease genes in the period between 1980 and 2001. (Taken from RetNet, http://www.sph.uth.tmc.edu/Retnet/).
photoreceptor-specific proteins, which is not surprising considering the complexity and the number of genes involved in the various steps of the phototransduction process (section 1.1.2), and the extreme metabolic activity of the photoreceptor cells which makes them vulnerable to even minor changes in the genes expressed in or near them. In fact, the underlying assumption of the candidate gene approach used to clone the majority of these genes is that mutation in photoreceptor-specific genes known to be indispensable for photoreceptor function are likely to confer a phenotype of retinal degeneration or dysfunction. Among these genes, examples include both the $\alpha$ and $\beta$ subunits of PDE ($\text{PDE}_\alpha$ and $\text{PDE}_\beta$), mutations in which were shown to cause respectively recessive RP (arRP) (Huang et al. 1995) and either arRP or dominant CSNB (McLaughlin et al. 1993, Gal et al. 1994), the $\alpha$ subunit of the cGMP-gated channel ($\text{CNGA}_1$), mutations in which were shown to cause arRP (Dryja et al. 1995), and retinal guanylate cyclase ($\text{RetGC}_1$), mutations in which were found in both recessive Leber congenital amaurosis and dominant cone-rod dystrophy patients (Perrault et al. 1996, Kelsell et al. 1998). Peripherin/RDS and ROM1 both code for integral membrane proteins of the outer segment discs of rods and cones and of rods respectively. Mutations in peripherin/RDS have been shown to cause adRP and autosomal macular dystrophy (Farrar et al. 1991, Wells et al. 1993), whilst mutations in ROM1 were found in adRP patients (Bascom et al. 1992). In addition both genes were shown to be mutated in a digenic form of RP (Kajiwara et al. 1994), providing one of the first examples of molecular documentation of this type of inheritance.

It is interesting to observe that retinal diseases display remarkable genetic heterogeneity (as the majority of retinal disorders are accounted for by mutations in more than one gene) and also allelic heterogeneity, as different mutations in the same gene can cause distinct phenotypes (e.g. RHO).

1.1.4 Retina diseases on the X chromosome

Due to wide interest in X-linked disease and the phenomenon of X inactivation, related to the different dosage of X in males and in females, the X chromosome is one of the most extensively studied of all human chromosomes. Due to haploinsufficiency in affected males, X-linked diseases tend to be particularly severe, and recessive diseases are generally revealed, hence the X chromosome appears to be particularly disease-gene rich.
General Introduction

Figure 1.7 shows the number of eye disease loci and genes identified on the X chromosome, clustered particularly on the short arm of the chromosome. The level of genetic heterogeneity found on the X for at least three retinal diseases is demonstrated; RP (5 loci), CSNB (2 loci), and cone dystrophy (COD, 2 loci) (sections 1.1.4.1, 1.1.4.2 and 1.1.4.3). These disorders fall into the main categories of retinal dystrophies, those primarily affecting peripheral vision (RP and CSNB), and the central retinal dystrophies (COD). In addition there are a number of other retinal dystrophies, which will be discussed separately (section 1.1.4.4).

Some of these disorders also share overlapping map location suggesting the possibility of allelism, where different mutations in the same gene may have different phenotypic effects. This has been documented for many autosomal retinal diseases, however allelic heterogeneity with RP on the X chromosome has yet to be found.

Two retinal genes have also been identified on the long arm of the X chromosome, X-arrestin (Sakuma et al. 1998) and RetGC2 (Lowe et al. 1995), both having autosomal paralogues involved in retinal diseases, S-arrestin (SAG, Fuchs et al. 1995), involved in recessive CSNB on chromosome 2 and RetGC1 involved in recessive Leber congenital amaurosis and dominant cone-rod dystrophy patients (Perrault et al. 1996, Kelsell et al. 1998) on chromosome 17. However no association between these genes on the X chromosome and disease has been detected to date.

1.1.4.1 Retinitis pigmentosa

Retinitis pigmentosa (RP) is a group of progressive hereditary disorders of the retina caused by the gradual degeneration of photoreceptor cells (primarily rods). The symptoms include night blindness and progressive concentric reduction of the visual field resulting in characteristic tunnel vision (Figure 1.8). Disease progresses centrally leading to cone degeneration and gradual loss of central vision (Heckenlively 1988), culminating eventually in blindness and degeneration of the retina for the majority of patients. RP is one of the most common causes of blindness affecting about one in 4000 people in the western world.

Clinically RP patients display fundus changes showing attenuated retinal vessels and bone spicule pigmentary deposits more prominent in the periphery of the retina, which are thought to result from migration of retinal pigment epithelium cells in the neural layer of the
**Figure 1.7:** Ideogram of the human X chromosome showing map locations of eye diseases. OAl: Ocular Albinism 1, XLRS1: X-linked Retinoschisis 1, DDP: Deafness-Dystonia-Optic Atrophy syndrome, MLS: Microphthalmia with Linear Skin defects, NHS: Nance-Horan syndrome, representative: Retinitis Pigmentosa, CSNB: Congenital Stationary Night Blindness, COD: Cone Dystrophy, NDP: Norrie Disease, FEVR: Familial Exudative Vitreo-Retinopathy, OPA2: Optic Atrophy 2, AIED: Aland Island Eye Disease, CHM: Choroideremia, RCP: Rod Cone Pigment, GCP: Green Cone Pigment.
Figure 1.8: a) A scene as it is viewed by a person with normal vision. 
b) The same scene as it might be viewed by a person with retinitis pigmentosa.
(Picture taken from National Eye Institute, http://www.nei.nih.gov/)
General Introduction

degenerating retina (Figure 1.9). A progressive deterioration of retinal function is revealed by electoretinographic analysis (ERG), in which the electrical signal in response to flashes of light are reduced in amplitude and delayed in response times.

Most cases of RP have no additional extraocular phenotypes (non-syndromic RP), however in a number of families, RP is found in association with other disease manifestations such as hearing loss in Usher syndrome (Fishman et al. 1983), mental retardation, polydactyly, obesity, and hypogenitalism in Bardet-Biedl syndrome (Schachat and Maumenee 1982), or deafness, obesity, and diabetes mellitus in Alstrom syndrome (Alstrom et al. 1959).

Retinitis pigmentosa is an extremely genetically heterogeneous disease. Families with autosomal dominant (adRP), autosomal recessive (arRP), X-linked (XLRP), maternal (mitochondrial), and digenic inheritance are well documented, and further genetic heterogeneity is present within inheritance classes with at least 11 adRP loci, 13 arRP loci and 5 XLRP loci (reviewed in Clarke et al. 2000, Gregory-Evans and Bhattacharya 1998, Inglehearn 1998, Molday 1998, Rattner et al. 1999, and RetNet). The prevalence of the various genetic forms varies considerably from country to country (Jay 1982), and up to 50% of RP cases are sporadic cases with no family history, most of which are probably recessive or due to new mutations. Table 1.1 shows a summary of the RP genes cloned to date.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Chromosome</th>
<th>Mode of inheritance</th>
<th>Function</th>
<th>Expression</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>RHO</td>
<td>3q21-q24</td>
<td>adRP</td>
<td>phototransduction</td>
<td>retina</td>
<td>Drvia et al. 1990</td>
</tr>
<tr>
<td>CNGC</td>
<td>4p14-p13</td>
<td>arRP</td>
<td>phototransduction</td>
<td>retina</td>
<td>Dryja et al. 1995</td>
</tr>
<tr>
<td>PDEβ</td>
<td>4p16.3</td>
<td>arRP</td>
<td>phototransduction</td>
<td>retina</td>
<td>McLaughlin et al. 1993</td>
</tr>
<tr>
<td>PDEα</td>
<td>5q31.2-q34</td>
<td>arRP</td>
<td>phototransduction</td>
<td>retina</td>
<td>Huang et al. 1995</td>
</tr>
<tr>
<td>RLBP1</td>
<td>15q26</td>
<td>arRP</td>
<td>phototransduction</td>
<td>RPE</td>
<td>Maw et al. 1997</td>
</tr>
<tr>
<td>ROM1</td>
<td>11q13</td>
<td>adRP-digenic</td>
<td>Structural</td>
<td>retina</td>
<td>Bascom et al. 1992, Kajiwara et al. 1994</td>
</tr>
<tr>
<td>NRL</td>
<td>14q11.2</td>
<td>adRP</td>
<td>transcription factor</td>
<td>retina</td>
<td>Bessant et al. 1999</td>
</tr>
<tr>
<td>RP1</td>
<td>8q11-q13</td>
<td>adRP</td>
<td>Unknown</td>
<td>retina</td>
<td>Sullivan et al. 1999</td>
</tr>
<tr>
<td>TULP1</td>
<td>6p21.3</td>
<td>arRP</td>
<td>Unknown</td>
<td>retina</td>
<td>Banerjee et al. 1998</td>
</tr>
<tr>
<td>RGR</td>
<td>10q23</td>
<td>arRP</td>
<td>chromophore regeneration</td>
<td>retina</td>
<td>Morimura et al. 1999</td>
</tr>
<tr>
<td>CRBI</td>
<td>1q31-q32.1</td>
<td>arRP</td>
<td>Unknown</td>
<td>retina</td>
<td>Den Hollander et al. 1999</td>
</tr>
<tr>
<td>MERTK</td>
<td>2q14.1</td>
<td>arRP</td>
<td>Unknown</td>
<td>RPE</td>
<td>Gal et al. 2000</td>
</tr>
<tr>
<td>RPGR</td>
<td>Xp21.1</td>
<td>XLRP</td>
<td>Unknown</td>
<td>ubiquitous</td>
<td>Meindl et al. 1996</td>
</tr>
<tr>
<td>RP2</td>
<td>Xp11.3</td>
<td>XLRP</td>
<td>Unknown</td>
<td>ubiquitous</td>
<td>Schwahn et al. 1998</td>
</tr>
</tbody>
</table>
Figure 1.9 A fundus of a patient with retinitis pigmentosa (for normal fundus see figure 1.3)
In addition, there are a number of genes involved primarily in other retinal dystrophies, which were shown to also cause adRP or arRP. Examples of these are the \textit{CRX} gene, which causes dominant cone-rod dystrophy, Leber congenital amaurosis (LCA) and adRP (late onset with cone-rod involvement \cite{Freund1997,Freund1998,Sohocki1998}; \textit{RPE65}, which was shown to cause LCA and arRP \cite{Marlhens1997,Morimura1998} and \textit{ABCR}, shown to cause recessive Stargardt disease, recessive macular dystrophy (MD) and arRP \cite{Allikmets1997,Cremers1998}.

The majority of RP genes identified are involved in the phototransduction cascade, are structural proteins of the photoreceptor cells (\textit{RDS} and \textit{ROM1}, section 1.1.2), or they are expressed at high levels in the retina and RPE. The two X-linked RP (XLRP) genes identified (\textit{RPGR} and \textit{RP2}) represent what can be considered a new generation of RP genes; in addition to having unknown function, they are ubiquitously expressed and their level of expression in the retina is barely detectable. For this type of gene the candidate gene approach loses its power and their discovery depends more on purely positional methods.

The X-linked form of RP is the most severe, it usually arises within the first two decades of life and progresses to total blindness within the third or fourth decades \cite{Bird1975}. Female carriers often develop phenotypic signs of the disease in middle to late life, however they can display a wide range of phenotypes, from asymptomatic to severely affected due to the proportion of retinal cells with an active disease-bearing chromosome \cite{Bird1975}.

\subsection*{1.1.4.1.1 Genetic heterogeneity in XLRP}

Genetic heterogeneity has also been demonstrated for the X-linked form of RP. To date there appear to be 5 distinct loci: RP2, RP3, RP6, RP23 and RP24.

RP2 was the first RP locus mapped on the X-chromosome by Bhattacharya \textit{et al.} \cite{Bhattacharya1984} through linkage studies of 5 British families, demonstrating linkage between XLRP and DXS7 in Xp11.3. Ott \textit{et al.} \cite{Ott1990} subsequently localised RP2 to a region spanning DXS7 and extending in Xp11.3 to the centromere, and Coleman \textit{et al.} \cite{Coleman1990} refined the location of the RP2 interval to between markers DXS426 and DXS7. The lack of recombination events within the critical region and of associated cytogenetic abnormalities for RP2 combined with a paucity of polymorphic markers hampered localisation of the \textit{RP2}
gene and indeed the refinement of the RP2 locus, which remained broadly localised to this ~13cM interval in Xp11.3-11.22 for a long time. It was in 1996 (Thiselton et al.) that two key families enabled the refinement of the locus to a 5 cM region between markers DXS8083 and DXS6616. The RP2 gene was subsequently positionally cloned in 1998 (Schwahn et al. 1998).

Evidence towards the existence of a second XLRP locus on the short arm of the X chromosome, started to appear in 1998 (Musarella et al.) with a proposed location for an XLRP gene in Xp21, tightly linked to OTC, consistent with a male patient (BB, Francke et al. 1985) who suffered from Duchenne muscular dystrophy (DMD), McLeod syndrome, chronic granulomatous disease (CGD) and RP who had a cytologically visible deletion in this chromosome band. Heterogeneity analysis of linkage data pooled from 62 families with XLRP provided very strong evidence in favour of two distinct loci for XLRP, with 75% of families associated with a locus in Xp21.1, named RP3, and 25% of families associated with a more proximal locus in Xp11, probably corresponding to the RP2 locus (Ott et al. 1990). These data were also confirmed by Teague et al. (1994) on a survey of 37 XLRP pedigrees. Further deletion patients and analysis of recombinants placed the RP3 locus between DXS1110 and OTC, an interval of about 520 Kb (Bergen et al. 1995, Brown et al. 1996, Fujita et al. 1996). The gene at the RP3 locus (RPGR) was positionally cloned in 1996 (Meindl et al., Roepman et al.), facilitated by genomic sequencing.

A third XLRP locus (RP6) distal to RP3, in Xp21.3-21.2, was proposed by Ott et al. (1990) based exclusively on statistical evidence, however this locus has never been proved experimentally.

More recently two additional loci, RP23 and RP24, have been described on distal Xp and on Xq respectively. RP23 was proposed on the basis of haplotype and linkage studies of an American XLRP pedigree, for which disease was shown to segregate with a ~15 cM interval between markers DXS1223 and DXS7161 in Xp22 (Hardcastle et al. 2000, section 5.3.2). The RP24 locus was mapped by haplotype and linkage analysis in a single XLRP family to a 23 cM interval in Xq26-27, between markers DXS8094 and DXS8043 (Gieser et al. 1998).

One additional locus (RP15) was originally reported in a family with a presumed variant phenotype from the classic XLRP, a dominant cone-rod dystrophy (McGuire et al.)
1995). It was only more recently that the authors re-evaluated the clinical status of a female in the pedigree, identifying her as a carrier (she had an affected son), hence the phenotype was a recessive form of RP. In light of this new clinical data the locus was found to span the RP3 interval and a mutation in the \textit{RPGR} gene was indeed identified (Mears \textit{et al.} 2000), therefore RP15 no longer exists as a genetic or clinically discrete entity.

1.1.4.1.2 The disease gene at the RP2 locus

The gene responsible for the RP2 form of retinitis pigmentosa was positionally cloned by Schwahn \textit{et al.} (1998) by screening the RP2 5 cM critical interval of 26 unrelated XLRP families for chromosomal rearrangement, using the YAC representation hybridisation (YRH) technique. This enabled the identification of an L1 insertion in one of the patients, which turned out to be located in one of the introns of the \textit{RP2} gene. The gene was found to consist of 5 exons, to encode a polypeptide of 350 amino acids (Figure 1.10) and to be ubiquitously expressed. About one third (151 amino acids) of the predicted coding sequence of the gene has homology to cofactor C, providing an indication of the possible role of the RP2 protein. Cofactor C is involved in the \(\beta\)-tubulin folding pathway (Tian \textit{et al.} 1996), collaborating to produce properly folded \(\alpha\)-tubulin subunits that can form heterodimers with \(\alpha\)-tubulin to constitute the microtubules. \(\alpha\) and \(\beta\)-tubulin are expressed in photoreceptor cells of the mammalian retina (Woodford and Blancks 1989), and it can be speculated that the pathogenesis of RP2 involves the accumulation of incorrectly folded photoreceptor tubulin isoforms, followed by progressive cell death. Mutations in the \textit{RP2} gene have been identified in approximately 20\% of XLRP families (Schwahn \textit{et al.} 1998, Hardcastle \textit{et al.} 1999, Mears \textit{et al.} 1999, Thiselton \textit{et al.} 2000, section 4.4), which is in agreement with the proportion of families segregating with this locus (Ott \textit{et al.} 1990, Teague \textit{et al.} 1994). Studies conducted more recently on the RP2 protein (Chapple \textit{et al.} 2000) have demonstrated the presence of N-terminal acylation that would target the protein to the plasma membrane, suggesting that RP2 may not be functioning exclusively in tubulin folding. However the precise role of RP2 and its involvement in the aetiology of the disease still remains to be determined.
**Figure 1.10:** Coding and amino acid sequence of the RP2 gene including exon/intron boundaries (exons 1 to 5). The amino acid sequence homologous to cofactor C is highlighted in pink.
1.1.4.1.3 The disease gene at the RP3 locus

The gene accounting for retinitis pigmentosa at the RP3 locus, named \textit{RPGR} (retinitis pigmentosa GTPase regulator), was cloned in 1996 (Meindl \textit{et al.}, Roepman \textit{et al.}). The approach used by Meindl \textit{et al.} to identify the gene was subcloning and sequencing the critical interval for RP3 and detecting gene sequences within that region by computational analysis. This led the authors to identify a 19 exon gene coding for a novel 815 amino acids protein (Figure 1.11) which is ubiquitously expressed, with a low level of expression in retina and RPE, and was found to be mutated in XLRP patients. The predicted protein product contains a tandem repeat structure similar to RCC1 (regulator of chromosome condensation), known to regulate the GTPase Ran (Ras-related nuclear protein). Similarly it can be speculated that \textit{RPGR} interacts with a retina specific GTPase, acting as a guanine-nucleotide exchange factor. Independently from Meindl \textit{et al.}, Roepman \textit{et al.} (1996a) cloned the same gene via the method of YAC representation hybridisation (YRH), the method later used to clone the \textit{RP2} gene (section 1.1.4.1.2).

After these first two publications, many other groups published mutations in the \textit{RPGR} gene (Buraczynska \textit{et al.} 1997, Miano \textit{et al.} 1999, section 4.3), and it soon became evident that the proportion of XLRP patients with mutations in this gene (~20-30%) was different from the proportion of families segregating with the RP3 locus (~70%, Ott \textit{et al.} 1990, Teague \textit{et al.} 1994), suggesting that part of the gene or alternative transcripts remained to be identified, or that microheterogeneity was present at the RP3 locus (i.e. more than one RP3-causing gene in close proximity).

Kirschner \textit{et al.} (1999) reported the identification of an additional retina specific exon (exon 15a) contained within intron 15 of the originally described gene (Meindl \textit{et al.} 1996) by analysing clones derived from human retina RNA, and the description of one patient deleted for this exon. Interestingly this is the same patient that initially enabled Roepman \textit{et al.} (1996b) to identify the RP3 gene (Roepman \textit{et al.} 1996a) leaving disease in this patient unaccounted for as no exons with an open reading frame were found in the 6.4 Kb region deleted in the patient. The same patient was clinically described in 1997 (Rosenberg \textit{et al.}) as having hearing impairments in addition to RP, which is perhaps surprising considering the reported retinal specificity of the exon that appears to be deleted.
**Figure 1.11:** CDNA and amino acid sequence of the **RPGR** gene including exon/intron boundaries (exons 1 to 19). The amino acid sequence homologous to **RCC1** is highlighted in blue.

**General Introduction**
The definitive evidence towards the hypothesis of additional coding sequences accounting for the majority of mutations in XLRP patients came more recently (Vervoort et al. 2000) with the identification of a new 3' terminal exon, exon ORF15, consisting of exon 15 and extending into part of intron 15. This exon, which was found to be expressed in all tissues with a more prominent product in retina, introduces a new stop codon and encodes 567 amino acids with a repetitive domain rich in glutamic acid and glycine residues. The exon was also shown to be highly conserved among species including mouse, bovine and Fugu rubripes genes. The authors found that 80% of the RPGR mutations identified in their patient pool were localised in this new exon, which appears to be a mutational hot spot probably due to the repetitive nature of the sequence. All the mutations found lead to premature termination of translation, due to small insertions or deletions (i.e. 1-, 2-, 4-, and 5-nucleotides) in the majority of cases, and 5 different mutations were found on at least two different haplotypes, indicating recurrent mutations. Several changes thought to be benign sequence variants were also detected. These were present in control chromosomes or were silent changes and included single-nucleotide substitutions and in-frame rearrangements. The unusual predicted protein sequence of exon ORF15, with a low sequence complexity and with a high glutamic acid and glycine content, was found to have no similarities with any known protein.

The involvement of RPGR in retinitis pigmentosa and indeed its overall function in the retina still remains to be clarified. The protein has been shown to interact with the δ subunit of rod cyclic GMP phosphodiesterase (PDEδ, Linari et al. 1999). PDE is a component of the visual transduction cascade of vertebrate photoreceptor cells (section 1.1.2). PDEδ has a high level of expression in all tissues and a high level of conservation across species, which suggests a more general function of the PDEδ protein in tissues other than the retina. Because the mouse homologue of the RPGR gene (mRpgr) was found to localise in the Golgi complex (Yan et al. 1998), and because PDEδ was shown to interact with the small GTPase, Rab 13 (Marzesco et al. 1998), which is involved in the regulation of membrane trafficking, and also to control the dynamic of the association of Rab13 with the cellular membranes, RPGR was proposed to be involved in intracellular vesicular transport. More recently, the interaction of RPGR with a retina specific protein (RPGRIP, Boylan et al. 2000, Roepman et al. 2000) has been demonstrated, and the two proteins were found to co-
localise in the outer segments of rod photoreceptors (Roepman et al. 2000). The function of this protein is unknown, however its interaction with RPGR and its retina expression and co-localisation of the two proteins in rod cells suggests that RPGRIP is a strong candidate gene for human retinal degeneration and provide a clue for the retina specific pathogenesis of RPGR mutations. This is also supported by the subcellular localisation of RPGR in the mouse photoreceptor’s connecting cilium (Hong et al 2000). In this study an RPGR-deficient murine model was created by knockout. The mutant mice showed a reduced level of rhodopsin in rod cells and localisation of cone cells cone opsins in the cell body, with subsequent degeneration of both rods and cones. The mouse model should prove useful to investigate the in vivo function of RPGR and to understand the disease mechanism associated with RPGR mutations. Also of interest is a naturally occurring canine model displaying an X-linked progressive retinal atrophy (XLPRA, Zeiss et al. 2000), with initial degeneration of rods followed by cone degeneration and complete retinal atrophy in affected males. XLPRA was tightly linked to an intragenic RPGR polymorphism and the full-length canine RPGR cDNA and three additional splice variants were cloned from the dog genome. More recently the cloning of the canine exon ORF15 enabled the identification of a mutation as the cause of disease in the XLPRA dog (Acland et al. 2001, Zhang et al. 2001).

1.1.4.2 Congenital stationary night blindness

Congenital stationary night blindness (CSNB) denotes a series of non-progressive retinal disorders (hence the term ‘stationary’) inherited as autosomal dominant (adCSNB), autosomal recessive (arCSNB), or X-linked (CSNBX) disease, which can only be distinguished by genetic evaluation of the pedigrees, although CSNBX is frequently associated with myopia. Night blindness is also a well characterised early clinical feature of the progressive degenerative disease retinitis pigmentosa (RP), and the functional relationship between RP and CSNB has been highlighted in studies describing mutations in the same gene that result in either an RP or CSNB phenotype (discussed in section 1.1.3). To date one gene has been identified as a cause of the dominant form of CSNB, the α subunit of transducin (GNAT1, Dryja et al. 1996). In addition, rare mutations in the RP genes RHO and the PDEβ were also shown to cause dominant CSNB (Dryja et al. 1993, Gal et al. 1994). Two genes which cause recessive CSNB (known as Oguchi disease) have been identified, the
S-arrestin gene (*SAG*, Fuchs *et al.* 1995) and the rhodopsin kinase gene (*RHOK*, Yamamoto *et al.* 1997). Genetic studies of X-linked CSNB families by several groups identified close linkage to markers on proximal Xp (OMIM: 300071 and 310500), and genetic heterogeneity for CSNBX was established (Bech-Hansen *et al.* 1993, Bergen *et al.* 1995). A key family of British origin defined a genetic location for CSNBX (locus named CSNB1 or CSNB4) between the retinitis pigmentosa loci RP2 and RP3, therefore excluding allelic heterogeneity with XLRP in this family (Hardcastle *et al.* 1997), and confirming the presence of at least 2 genes for CSNBX on Xp11.

The phenotype for CSNBX is described as ‘Schubert-Bornschein’ type, thought to result from decreased effectiveness of synaptic transmission between photoreceptors and second-order neurons, since electroretinograms (ERGs) suggest functional rod photoreceptor outer segments with a more proximal defect (Miyake *et al.* 1987, Hood and Birch 1990). CSNBX displays clinical heterogeneity with 2 major clinical definitions, “complete” (or CSNB1 associated with myopia), and “incomplete” (or CSNB2 associated with hyperopia and myopia), defined by the presence of residual rod function in the latter (Miyake *et al.* 1986, Miyake *et al.* 1994). Although there are evidently 2 distinct phenotypes, the clinical definition with reference to genetic location, mutation type, and association with other ocular phenotypes (e.g. hyperopia, myopia, nystagmus and strabismus) warrants further investigation (Boycott *et al.* 2000, Pearce *et al.* 1990, Bergen *et al.* 1995), since there are reported observations of both types of CSNBX within the same family (Pearce *et al.* 1990, Aldred *et al.* 1992, Bergen *et al.* 1995).


1.1.4.2.1 The disease gene at the CSNB1 locus

The gene responsible for the complete form of CSNB (CSNB1) was recently identified on Xp11.4 by positional cloning (Pusch *et al.* 2000, Bech-Hansen *et al.* 2000). Mutations in this gene, named *NYX*, appear to be a common cause of CSNBX, with mutations identified in 47 families. Transcript size and tissue distribution were not detectable by Northern analysis due to the low level of expression of this gene, however RT-PCR
results indicate expression in several tissues including retina, kidney, brain, muscle, placenta and testis (Pusch et al. 2000). The predicted amino acid sequence encodes a 481 amino acids polypeptide (nyctalopin), which is a new putative extracellular member of the small leucine-rich proteoglycan (SLRP) family. The protein contains 11 typical leucine-rich repeats (LRRs) flanked by two cysteine-rich LRRs with a C-terminal GPI anchor and N-terminal signal sequence (see section 4.1.3 and Figure 4.11). Members of the LRR superfamily are involved in regulation of cell growth, cell adhesion and axon guidance (Hocking et al. 1998). Nyctalopin has approximately 30-35% identity with other family members including Drosophila slat (Rothberg et al. 1988) and Karatocan (mutations in which have been shown to cause corneal plana, Pellegata et al. 2000). Other superfamily members that have GPI-anchors include Drosophila chaoptin, an extracellular photoreceptor neuron-specific cell adhesion molecule (Krantz et al. 1990), and Drosophila connectin, which acts as a repulsive guidance molecule during motoneurone growth cone guidance and synapse formation (Nose et al. 1994). Nyctalopin may therefore have important function as a developmental regulator of synaptic formation for the rod mediated response of the retina. Previously identified CSNB genes (e.g. RHO, PDEβ and CACNA1F) are essential for effective photoreceptor phototransduction, however NYX is predicted to have a different function and therefore presents a different disease mechanism for CSNB.

1.1.4.2.2 The disease gene at the CSNB2 locus

The gene at the CSNB2 locus responsible for the incomplete form of CSNB was cloned in 1998 (Bech-Hansen et al., Strom et al.). Mutations in this novel voltage-gated L-type calcium channel α1-subunit gene (CACNA1F) have been reported in 16 families included in the original investigations. The gene maps to Xp11.23 and is predominantly expressed in the retina. CACNA1F cDNA is 5901 bp long with a predicted protein of 1966 amino acids containing four repeat domains, each with six transmembrane domains most closely related to brain CACNA1D. Neurotransmitter release is mediated by the influx of Ca$^{2+}$ into photoreceptors through calcium channels (Schmitz et al. 1997, section 1.1.2), therefore mutations in this gene are thought to impair the influx of Ca$^{2+}$ required for tonic glutamate release from photoreceptor pre-synaptic terminals in the dark. Functional
investigation of the role of this protein in the neural retina is now required to probe this hypothesis.

1.1.4.3 Cone dystrophy

X-linked progressive cone dystrophy (COD1, OMIM: 304020 and COD2, OMIM: 300085) is a rare hereditary disorder primarily affecting the cone photoreceptors. The clinical symptoms in affected males include progressive loss of visual acuity, colour vision abnormalities, photophobia, myopia and prominent macular changes, leading to localised atrophy of the retinal pigment epithelium in older individuals (Jacobson et al. 1989). According to Pinckers (1982), the disorder begins as a peripheral cone disease and progresses to a diffuse cone disease. Affected males present with an absent or severely attenuated cone ERG, and the rod ERG may also be abnormal. Heterozygous females have diminished visual acuity and myopia, but a normal cone ERG and normal colour vision.

Due to the rarity of this form of the disease, only a small number of families have been clinically described and only a few of them have been studied for linkage. The study of four independent families mapped the COD1 locus to Xp11.3-21.1 (Meire et al. 1994, Hong et al. 1994). These studies were unable to resolve COD1 as a distinct locus from the retinitis pigmentosa RP2 and RP3 loci, and the possibility of allelism between COD1 and either of the two RP loci was suggested. The study of three additional pedigrees (Seymour et al. 1998) enabled the authors to refine the disease interval to a 1 cM interval in Xp11.4 between markers DXS993 and DXS556, excluding COD1 as an allelic variant of both RP2 and RP3.

In a family with X-linked progressive cone dystrophy Bergen and Pinckers (1997) found linkage to the Xq27 region (COD2 locus). This established a distinct locus to COD1, providing evidence for genetic heterogeneity underlying cone dystrophy on the X chromosome.

1.1.4.4 Other X-linked retinopathies

-Ocular Albinism

Ocular albinism type 1 (OA1, OMIM: 300500) is an X-linked disorder characterised by severe impairment of visual acuity, retinal hypopigmentation, nystagmus and the presence of macromelanosomes. The gene, mapped to a critical region in Xp22.3-22.2, was cloned in
1995 by Bassi et al. and was found to be expressed at high levels in RNA samples from retina, including the retinal pigment epithelium. The 424 amino acid protein encoded by the \textit{OAI} gene displays several putative transmembrane domains and shares no similarities with previously identified molecules. Schiaffino et al. (1999) provided structural evidence that the protein product of the \textit{OAI} gene, a pigment cell-specific integral membrane glycoprotein, represents a novel member of the G protein-coupled receptors (GPCR) superfamily and demonstrated that it binds heterotrimeric G proteins. Incerti et al. (2000) generated and characterised \textit{OAI}-deficient mice by gene targeting. Ophthalmologic examination showed hypopigmentation of the ocular fundus in mutant animals compared with wild type.

\textbf{-Retinoschisis}

Retinoschisis (RS, OMIM: 312700) is described as intraretinal splitting due to retinal degeneration. Affected males show cystic degeneration leading to split in the retina, detachment of the retina, and finally complete retinal atrophy with sclerosis of the choroid (George et al. 1996). Impairment of vision is slowly progressive. The typical lesions displayed in RS patients are thought to be related to a defect in retinal Muller cells. Sauer et al. (1997), suggested that as Muller cells have been shown to aid neurite outgrowth and neuronal connections, failure to establish a proper neuronal interaction may be the indirect result of a Muller-cell defect. Consequently, it has been suggested that RS may be considered a disorder of retinal development rather than a dystrophic process. The \textit{XLSRI} gene, cloned by Sauer et al. in 1997, was found to be exclusively expressed in the retina. The predicted protein sequence, named retinoschisin, was found to contain a highly conserved motif implicated in cell-cell interaction and thus may be active in cell adhesion processes during retinal development. Grayson et al. (2000) suggested that retinoschisin is released by photoreceptors and has functions within the inner retinal layers, therefore X-linked retinoschisis may be caused by abnormalities in a putative secreted photoreceptor protein.

\textbf{-Norrie disease}

Norrie disease (NDP, OMIM: 310600) is an X-linked neurological disorder characterised by congenital blindness due to bilateral retina dysplasia showing abnormal vascularisation of the peripheral retina, mental disturbances and progressive sensorineural
deafness in one-third of the patients (Warburg 1966). Using a positional cloning strategy the NDP (or norrin) gene was isolated in 1992 by two different groups (Berger et al. and Chen et al.) and was found to be exclusively expressed in retina, brain and choroid. Mutations in the same gene were also found to cause familial X-linked exudative retinopathy (FEVRX, Chen et al. 1993), a disorder characterised by retinal traction, peripheral vitreous opacities, and subretinal and intraretinal exudates (OMIM: 305390) and retinopathy of prematurity (ROP, Shastry et al. 1997) a retinal vascular disease occurring in infants with short gestational age and low birth weight, which can lead to retinal detachment. Protein sequence comparisons revealed homologies with cysteine-rich protein-binding domains of proteins implicated in the regulation of cell proliferation (Chen et al. 1993), suggesting that the NDP protein likewise may be involved in the pathway that regulates neural cell differentiation and proliferation. To elucidate the cellular and molecular processes involved in Norrie disease, Berger et al. (1996) used gene targeting technology to generate NDP mutant mice. Hemizygous mice carrying a replacement mutation in exon 2 of the NDP gene developed retrolental structures in the vitreous body and showed an overall disorganisation of the retinal ganglion cell layer with loss of photoreceptor outer segments.

-Aland Island Eye Disease

Aland island eye disease (AIED, OMIM: 300600) is a disorder described in a family from the Aland Islands in the Sea of Bothnia, characterised by albinism of the fundus, hypoplasia of the fovea, visual impairment, nystagmus, myopia, astigmatism, and colour vision abnormalities. Initially the phenotype of this family suggested that AIED was a variant of ocular albinism (OA1) and was given the locus symbol OA2. Subsequently Weleber et al. (1989) pointed out that electroretinography examinations (ERGs) in AIED patients showed many differences between this disorder and OA1, whilst similarities were found with CSNB; they thus suggested the possibility that AIED should be classified as a type of CSNB. Since the gene responsible for AIED has been assigned to the DXS7-DXS225 interval in Xp11.3 (Alitalo et al. 1991, Glass et al. 1993) allelism between AIED and one of the two CSNB loci, both located in the pericentromeric region of the short arm of the X chromosome, remains open.
General Introduction

-Choroideremia

Choroideremia (CHM, OMIM: 303100) is an X-linked condition whose clinical symptoms include progressive loss of vision, characterised by reduction of central vision, constriction of visual fields and night blindness, beginning at an early age, and the complete atrophy of choroid and retina. By linkage analysis the disease gene was located to the Xq13-q21 interval by Schwartz et al. (1986). With the aid of four different deletions found in males affected by choroideremia Cremers et al. (1990) isolated a 45 kb genomic DNA segment corresponding to this region of deletion overlap. Conserved sequences from this DNA segment were used as a probe to screen a human retinal cDNA library and a transcript expressed in choroid, retinal pigment epithelium and other cells was detected. Isolation and characterisation of the complete open reading frame of the CHM gene and its exon-intron structure was later reported by van Bokhoven et al. (1994). A gene targeting approach was used by van den Hurk et al. (1997) to disrupt the mouse gene, which was shown to be lethal in male embryos, and in females embryos only if the mutation is of maternal origin, giving evidence of an X-linked imprinted gene. In both heterozygous females and chimeras, the mutation caused photoreceptor cell degeneration. Consequently, conditional rescue of the embryonic lethal phenotype of the chm mutation may provide a faithful mouse model for choroideremia.

-Optic Atrophy

Optic atrophy (OMIM: 165500) is a disorder characterised by decreased visual acuity, colour vision deficits, and optic nerve pallor. Went et al. (1975) described a family with a very early onset of the disease, perhaps present at birth and with a very slow progression of the loss of visual acuity. Genetic analysis of this family (Assink et al. 1997) placed the gene in the Xp11.4-p11.21 interval between markers DXS993 and DXS991 (OPA2, OMIM: 311050).

-Red and Green cone pigments

Red and green cone pigments (RCP and GCP) are responsible for protanopia (red, or protan, colourblindness, PCB) and deuteranopia (green, or deutan, colour-blindness, DCP) respectively (OMIM: 303900 and 303800). Red and green opsins are bound to the
chromophore 11-cis retinal to form the visual pigment in cone photoreceptor cells and have
different absorption spectra. Colour-blindness arises from alterations in one of the genes
encoding one of the cone opsins. In western Europeans, about 8% of males are colour-blind.
Of these, about 75% have a defect in the deutan (green) series and about 25% have a defect
in the protan (red) series. The genes coding each of the two visual pigments situated on
Xq28, and the blue pigment situated on chromosome 7, were isolated by Nathans et al.
(1986).

1.2 The Human Genome Project

The Human Genome Project is the task of an international 15-year collaboration
involving over 20 laboratories and hundreds of people around the world, whose ultimate goal
is to sequence the entire human genome and to delineate the position of all genes by the year
2005. The idea of sequencing the entire human genome was first proposed at scientific
meetings during the years 1984 to 1986 (Palca 1986). However the US National Research
Council proposed to commence with a broader programme that would involve the creation of
genetic, physical and sequence maps of the human genome, as well as parallel efforts in
model organisms, and development of new advanced technologies. The programme was
launched by the US Department of Energy and the US National Institute of Health, supported
by other Institute around the world including the UK Medical Research Council and the
Wellcome Trust in Britain and other centres in France and Japan and later in Germany and
China. After the construction of genetic and physical maps of the human and mouse genomes
(Donis-Keller et al. 1987, Hudson et al. 1995, Nusbaum et al. 1999), and the partial
sequencing of the yeast and worm genomes (Oliver et al. 1992, Wilson et al. 1994), a general
strategy for genome sequencing was developed. This involves a first phase of ‘shotgun’
sequencing, in which the genome is divided into appropriately sized segments and each
segment is covered to a high degree of redundancy through the sequencing of randomly
selected subfragments, and a second ‘finishing’ phase, in which sequence gaps are closed and
remaining ambiguities are resolved through directed analysis. After launching pilot-projects
to demonstrate the feasibility of large-scale sequencing and the introduction of BACs
(Shizuya et al. 1992) in the strategy, the human genome sequencing effort moved into full
production in March 1999, with the initial goal of producing, by the year 2001, a draft
The achievement so far has already produced interesting information, not least that there appears to be about 30,000-40,000 genes in the human genome, much less than the predicted 100,000 (Antequera and Bird 1994, Liang et al. 2000), and only about twice as many as the worm or fly or a few thousand more than a plant (the C. elegans Sequencing consortium 1998, Adams et al. 2000, the Arabidopsis Genome Initiative 2000).

With the completion of sequencing chromosomes 21 and 22 and "draft" of human genome, the human genome sequencing project is entering its final stage. The "finished" sequence of the entire human genome within a few years will mark a transition from the era of physical genomics to that of functional genomics.

1.3 Paths to human disease gene identification

Different routes can be followed to identify human genes associated with disease, but all pathways eventually converge on mutation screening in a candidate gene. In general two different approaches are commonly used. The first one does not require any information on the chromosomal location of the disease locus, and depends on prior information about the protein product or function of the responsible gene, and is therefore termed 'functional cloning'. The first disease genes were identified using this method because relevant mapping information was not available. The second approach depends on map position and assumes no functional information, thus the term 'positional cloning'. With the constant generation of high resolution genetic and physical maps from the HGP initiative and the concomitant decrease of identified genes with a known function or of inherited diseases with a known underlying biological defect, this second approach has been more commonly adopted in
recent years. In reality most genes are identified through a combination of both methods integrating chromosomal location data and properties of a candidate gene in what is called the ‘positional candidate approach’. Once the chromosomal region containing the disease gene has been narrowed to a manageable size, candidate genes are selected on the basis of their specific or most abundant expression in the tissue most profoundly affected by the disease. If the gene appears to be ubiquitously expressed the candidate selection criterion might be based on homologies with known proteins in different species (orthologues) involved in similar phenotypes, or homologies with a family of proteins (paralogues) known to play crucial steps in a pathway that may lead to a similar phenotype if disrupted. However, although the power of the positional candidate approach is indisputable, the identification of some genes will depend mainly on more positional methods.

1.3.1 Positional cloning

The molecular genetics of human disease must often proceed without knowledge of the biochemical function of the responsible genes. Positional cloning methods must therefore be used to identify the gene that contains disease-causing mutations. A schematic outline showing the key features of this approach is represented in Figure 1.12. The first crucial step of positional cloning is performing genetic linkage analysis in affected pedigrees to establish a minimal chromosomal region of segregation with disease. This can sometimes localise a gene to a manageable sized interval (0.5-1 cM). More often however the minimal region is much larger (10 to 20 cM), making the successive steps much more laborious. The next phase involves construction of a high-resolution physical map represented by YAC, BAC, PAC and cosmid contigs covering the disease interval (section 1.5). This was a rate-limiting step in positional cloning for a long time. However, physical mapping is now facilitated by the availability of contigs covering most of the human genome, thus removing the labour-intensive contig building steps of disease gene identification. The next phase of positional cloning is the identification and characterisation of genes in the region. For a positional cloning effort to be effective, it is imperative to use rapid and thorough methods to identify candidate genes (section 1.6). Ultimately, the identification of a disease gene
**Figure 1.12:** Schematic representation showing the key features of positional cloning

- **Family inheritance pattern**
- **Linkage analysis and finer genetic mapping**
- **Recognised boundaries of the critical interval**
- **Physical mapping**
  - Contig of cloned fragments across the critical interval
- **Gene identification and isolation**
  - Candidate cDNAs
- **Mutation screening**
  - Normal
  - Mutation
requires the demonstration of patient-specific mutations associated with the disease phenotype.

1.4 Genetic mapping

Genetic linkage maps assign chromosomal locations to genetic landmarks, either genes or distinct short sequences of DNA, on the basis of how frequently markers are inherited together. Linkage maps exploit a phenomenon called recombination or crossing over.

1.4.1 Meiotic recombination and genetic map distances

As developing sperm and egg cells divide during meiosis, pairs of maternal and paternal chromosomes, each consisting of two chromatids, occasionally break and exchange pieces with one another, in a phenomenon called crossing over. The zones of contact between two homologous chromosomes undergoing a recombination event, known as chiasmata, reflect the occurrence of crossing over between chromatids, with at least one chiasma per pair of chromosomes.

Sets of loci on the same small chromosomal segment are unlikely to be separated by a crossover, and are termed linked, i.e. they tend to be transmitted as a block known as haplotype. In contrast, the further apart two loci are on a chromosome, the more likely it is that a crossover will separate them, the result of this recombination can be seen in the offspring and such offspring are termed recombinants. The recombination fraction (θ) is a measure of the extent of genetic linkage between two or more loci and expresses the proportion of recombination between two loci out of the total number of opportunities for recombination. On average, given any two loci segregating independently, 50% of the children will be recombinant between those two loci and 50% will be non-recombinant, the recombination fraction is therefore 0.5 or 50%. If the loci are syntenic, that is if they lie on the same chromosome, they will show deviation from the normal pattern of independent assortment. The recombination fraction will then be 0 for tightly linked loci, which always segregate together without crossovers, and it will range between 0 and <0.5 for loci that are not tightly linked, depending on how far apart they are on the chromosome. Thus the recombination fraction is a measure of the genetic distance between two loci.
The genetic map distance (in unit of Morgans) is defined as the expected number of crossovers occurring on a single chromatid between two loci. Chiasma counts in male meiosis show an average of 53 crossovers over all the autosomes. Since each crossover gives 50% recombinants, a total male genetic map length of approximately 26.5 Morgans is estimated. Because chiasmata are more frequent in female meiosis, the female genetic map length approximates 39 Morgans, thus the sex-averaged map length of the human genome equals 33 Morgans. A human chromosome is then on average 1.5 Morgans long, which means that it experiences an average of 1.5 crossovers per meiosis. Commonly genetic distances are given in centimorgans (cM), and two loci that show 1% recombination (θ=0.01) are defined as being 1 cM apart on a genetic map. This holds true for small genetic distances, but over longer distances, the occurrence of multiple crossovers and the effect of interference, whereby a crossover reduces the probability of a second crossover in its vicinity, this linear relationship is lost.

Map distance is a genetic distance, which cannot be directly correlated with physical distance (measured in number of base pairs of DNA). The only connection is that an increase in map distance translates into a larger physical distance. However it can be assumed that, with a total genetic map length of around 3000 cM and a number of $3 \times 10^9$ bp in the haploid genome, a genetic distance of 1 cM approximately corresponds to 1 Mb. This ratio however varies considerably for different chromosomal region according to different degrees of recombinogenicity. In general there is more recombination towards the telomeres of chromosomes in males, while centromeric regions have recombinants in females but not in males (Broman and Weber 1998)

1.4.2 DNA markers as a genetic mapping tool

Human genetic mapping depends largely on markers. A crucial characteristic of a genetic marker is a sufficient level of polymorphism so that any randomly chosen person stands a good chance of being heterozygous for that marker, i.e. of being 'informative for linkage'. The probability that a random person is heterozygous for a given marker is used as a measure of the degree of polymorphism for that marker. This probability is defined as:

$$H = 1 - \sum p_i^2$$
where \( p_i \) is the population frequency of the \( i^{th} \) allele. Heterozygosity is therefore directly proportional to the number of alleles presented by each marker and the frequency of each allele in the general population.

Another important characteristic of genetic markers is their density along the genome. Each sequence variant has occurred in a particular individual at some time in the past. Throughout generations the haplotype immediately surrounding an altered gene will be maintained. The size of this region, in which the genetic markers are said to be in linkage disequilibrium, will vary with the age of the variant. Assuming that a certain sequence variant is 1000 to 10,000 generations old, then the haplotype shared by two distantly related chromosomes in the vicinity of that variant, will on average be as small as 10 to 100 Kb. This is why, for efficient identification of disease associated regions, a high marker density is necessary.

The first generation of DNA markers were restriction fragment length polymorphisms (RFLPs, Botstein et al. 1980), which take advantage of the property of a normal sequence variation, such as a single base-pair change or a deletion or insertion, to create or abolish restriction fragment recognition sites, thus altering the length of the DNA fragments produced upon digestion with the appropriate restriction enzyme. These differences can be detected with Southern blot analysis. A limitation of this method is the time consuming and expensive technology, and also the limited informativeness of RFLPs, which can have only two alleles, thus their maximum heterozygosity is 0.5.

Minisatellites or VNTRs (variable number tandem repeats) consist of tandemly repeated units of 11-60 bp that can extend up to 1 Kb in length (Jeffreys et al. 1985). VNTRs represented a great improvement as they can have many alleles and a high level of heterozygosity. However the employment of Southern blots and the fact that VNTRs are not evenly spread across the genome poses a limitation to the use of these markers.

The advent of polymerase chain reaction (PCR, Saiki et al. 1988) represented a major advance in genetic mapping allowing rapid development and detection of highly polymorphic markers. The standard tools for PCR linkage analysis are microsatellites, tandem repeats of 1-6 bp, represented mostly by \((CA)_n\) repeats, and consisting of around 10-50 copies. They are found abundantly and randomly distributed in the genome. Tri- and tetranucleotide microsatellites occur approximately every 300-500 Kb on the human X
chromosome (Edwards et al. 1991), and this frequency seems to be maintained throughout the genome. CA repeats occur approximately every 30 Kb in human euchromatic DNA (Stallings et al. 1991), therefore these microsatellites are the most commonly used. Microsatellite repeats are particularly prone to strand slippage during replication, and this slippage makes them highly polymorphic in the population.

The most recent generation of genetic markers are single nucleotide polymorphisms (SNPs), which include RFLPs but also polymorphisms that do not happen to create or abolish a restriction site. The advantage of SNPs is that they can be developed in large quantities and processed on DNA chips without recourse to gel electrophoresis (Wang et al. 1998), and that they are extremely widespread along the entire genome. The latest figures provided by the Human Genome Project show the presence of 1.42 million SNPs found at a density of approximately one SNP per 1.91 Kb. Their sheer number and the possibility for automation largely offset the lower informativeness of SNPs.

1.4.3 Linkage analysis

The aim of linkage analysis is to assess whether two or more markers are genetically linked, which depends on whether or not recombination events have occurred between the markers and disease. In most cases it is not possible to identify recombinants unambiguously and count them, owing to the phase of heterozygous parent genotype. That is, the distribution of alleles between the two chromosomes, which cannot always be inferred, therefore making distinction between recombinants and non-recombinants difficult to decipher. It is possible, however, to calculate the overall likelihood of linkage, on the alternative assumption that the loci are linked ($\theta<0.5$) or not linked ($\theta=0.5$). The ratio of these two likelihoods gives the odds of linkage, which is normally expressed as a logarithm, the LOD (log of the odds, z) score for linkage, and is calculated at a series of $\theta$ ranging from 0 to 0.5. The lod score is thus defined as the $\log_{10}$ of the ratio of the probability that the data would have arisen if the loci were linked, to the probability that the data would have arisen if the loci were unlinked, hence is used as a measure of support for linkage versus absence of linkage (Ott 1985).

Positive lod scores give evidence in favour of linkage and negative lod score give evidence against linkage. The most likely recombination fraction is the one at which the lod score is highest. If there are no recombinants, the lod score will be maximum at $\theta=0$. A lod
score of 3 is the threshold for accepting linkage, with a 5% chance of error. A lod score of \(-2\) or less is taken as evidence against linkage. The values in between are considered inconclusive. A lod score of 3 corresponds to 1000:1 odds. The reason why such a stringent threshold is required lies in the improbability that two loci, chosen at random among 22 chromosomes, should be linked. For proof of linkage on the X chromosome a lod score of 2 is considered sufficient, owing to the higher probability of two randomly chosen loci on one chromosome (one order of magnitude higher in relation to 22 autosomes) to be linked.

Multipoint linkage analysis, whereby more than two loci are analysed simultaneously, can be more efficient. Multipoint analysis will never give a higher value than a two-point analysis where the marker is fully informative. The usefulness of multipoint is that the chances that at least one of the markers used gives a maximum lod score are increased.

1.4.3.1 Haplotype analysis

Linkage analysis allows one to place a disease gene within a subregion of a given chromosome. On average 1 cM of genome sequence may contain up to 50 genes in a gene-rich region, it is therefore important to exhaust all opportunities to refine the genetic interval, and this may be possible by haplotype analysis. Haplotype analysis is employed by typing further markers around the gene in key recombinant individuals of the pedigree, in order to identify the closest flanking markers to the disease gene. The segregation of a marker is followed in each of the family members to carefully determine whether each new marker has recombined with the disease locus or resides within the critical interval, thus establishing sites of meiotic recombination. If there is any uncertainty about the distribution of alleles between the two chromosomes, the most likely haplotypes are inferred by minimising the number of crossover events in each sibship.

This approach is limited by the number of informative meioses, the marker density and heterozygosity, and the number of recombinant individuals.
1.5 Physical mapping

After a gene has been assigned to a relatively small area on a chromosome through genetic linkage maps, before moving on to the characterisation of transcripts in the region, this region must be recovered in cloned DNA to create a resource from which to isolate candidate transcripts. Physical maps of DNA can have several levels of detail, from the banding patterns of the chromosomes, to clones of overlapping segments of DNA, and ultimately to the base-by-base sequence of DNA (Figure 1.13) reflecting different levels of resolution.

1.5.1 Low resolution physical mapping

The low-resolution physical maps are achieved commonly by cytogenetic methods, and the smallest map unit that can be resolved is traditionally one to several megabases of DNA. However, with the recent development of various technologies the level of resolution has increased to a maximum of 10-20 Kb.

1.5.1.1 Somatic cell hybrid

Somatic cell hybrids can be generated, under certain experimental conditions, inducing the fusion of cultured cells from different species. In human genome mapping, human and mouse or hamster hybrids are commonly used (Ruddle 1981). Due to the unstable nature of hybrid cell lines, human chromosomes tend to be randomly lost, thus generating a panel of hybrids each containing the full set of rodent chromosomes plus a few human chromosomes. Hybrids can be selected for retention of a given human chromosome if it corrects an otherwise lethal abnormality in the rodent cell.

The human chromosome in the somatic cell hybrid can be identified upon hybridisation or PCR using human specific sequences. More refined mapping can be achieved by using hybrids containing parts of human chromosomes (such as those generated from human cells that have a chromosomal translocation or deletion).

1.5.1.1.1 Radiation hybrid mapping

A major development in somatic cell genetics has been the introduction of radiation hybrids (RH) to generate physical maps (Walter et al. 1994). RH are constructed by first
**Figure 1.13:** Multiple levels of human chromosome mapping. The lowest resolution is provided by cytogenetic maps, and then increases through progressively more detailed physical maps, to achieve its maximum with the base-by-base sequence of DNA. (Picture taken from Access Excellence, http://www.accessexcellence.org/)
using X-rays to break the chromosomes of a human cell into several fragments whose sizes are a function of the dose of radiation. The irradiated cells are then fused with the rodent cells, each of which will then harbour a complete set of rodent chromosomes and a random set of human chromosome fragments. When a set of DNA markers is assayed in a panel of such radiation hybrids, the pattern of cross-reactivity can be used to construct a map, by taking advantage of a principle similar to that used in genetic linkage analysis, the nearer together the two DNA sequences are on a chromosome, the lower the probability that they will be separated by a break produced by the X-ray irradiation. Distances between markers are thus calculated and measured in centiRays (cR), which is a function of the dosage of radiation.

The range of resolution of RH mapping can be varied by altering the X-ray dose used to fragment the chromosomes.

1.5.1.2 In situ hybridisation

Chromosomal in situ hybridisation is a powerful technology to assign chromosomal location and relative order of DNA sequences to chromosomal sub-regions. Chromosomes in metaphase on a microscope slide are hybridised with a suitable probe isotopically-labelled. More recently, the sensitivity and resolution of in situ hybridisation has been largely increased by the development of fluorescent in situ hybridisation (FISH), which uses fluorescently-labelled probes that can be easily scored by eye under a fluorescence microscope (Trask 1991).

The resolution of conventional FISH on metaphase chromosomes is 1-2 Mb (Trask et al. 1991). The use of the more extended prometaphase chromosomes can permit a higher resolution (30 Kb-1 Mb). More recently, a technique has been developed, which mechanically stretches out free chromatin fibres from interphase nuclei on a microscope slide, this can achieve a resolution of 10-20 Kb (Haaf and Ward 1994).

1.5.2 High resolution physical mapping

The high-resolution physical maps, resolved through molecular methods, have a resolution that can go from hundreds of kilobases to a single nucleotide.
1.5.2.1 Long-range restriction mapping

Long-range restriction mapping enables the determination of the order and distance of genetic markers by Southern blot analysis of complete or partially digested genomic DNA. Two markers are physically linked if they co-hybridise to the same band and their maximum distance is indicated by the smallest restriction fragment shared by both probes. The resolution of restriction mapping depends on the frequency of the recognition site of the restriction enzyme used. This method can be applied to intact YAC, PAC or BAC cloned contigs, enabling identification of overlaps, rearrangements, deletions, translocation and chimerism that may be present.

The development of pulsed field gel electrophoresis (PFGE, Schwartz and Cantor 1984) has made restriction mapping possible, owing to its ability of separating large restriction fragments, which could not have been achieved by standard gel electrophoresis. Whilst standard gel electrophoresis allows separation of fragments of up to 50 Kb, PFGE allows the resolution of fragments as large as 10 Mb. This is achieved by introducing electric fields that change strength or direction over time, thus forcing the large DNA molecules to periodically reorientate in new directions. The time taken for a DNA molecule to reorient under the new electric field is strictly size-dependent, allowing very large molecules to be efficiently fractionated.

1.5.2.2 Clone contig assembly

The construction of the ultimate physical map of a region of interest requires the assembly of cloned DNA fragments, which collectively provide full representation of the given region. To ensure a complete coverage, the clones should contain overlapping inserts forming a continuous tiling path or contig. The strategy involves partial digestion of the genomic DNA and cloning of the fragments generated so that DNA sequences of each clone partially overlap with at least some other clones in the library.

One of the main approaches to determine the overlaps between clone inserts and to assess this overlap is by sequence tagged site (STS) content mapping (Kere et al. 1992). The strategy is based on PCR screening of STS markers in different clones, overlapping inserts are therefore identified when an STS is present in both. Based on the assumption that an STS
is unique to a particular genomic locus, all clones containing that STS must have originated from that same locus and must therefore overlap.

An alternative approach for the identification of overlapping clones is by fingerprint analysis. This method can be performed without first separating the clone insert from the host chromosomal background, by digestion of clones, separation by PFGE and identification of common bands by Southern analysis of clone inserts using human-specific repetitive sequence probes (e.g. Alu, Wada et al. 1990, or LINE-1, Bellane-Chantellot et al. 1992). A more rapid alternative to generate a fingerprint is to perform inter-Alu PCR (Nelson et al. 1989) using primers designed from the terminal end of the repeats, and compare the pattern of products generated by electrophoresis.

To bridge gaps in existing contigs or to extend outwards from both ends of a contig the procedure generally employed is chromosome walking. In this strategy, terminal sequences of a clone are used as a probe to screen a library for the identification and isolation of overlapping clones. Different methods have been developed for the isolation of clone ends, most of which are PCR-based (e.g. inverse PCR, Ochman et al. 1988, vectorette PCR, Riley et al. 1990, vector-Alu PCR, Nelson et al. 1991).

1.5.2.2.1 Yeast artificial chromosomes versus bacterial host systems

The development of YACs, which contain inserts of up to 2 Mb (Burke et al. 1987, Larin et al. 1991), has greatly increased the genomic cloning capacity over cosmids, which contain inserts of approximately 40 Kb. Owing to their large size, contigs of up to several megabases can be assembled over large regions of chromosomes. They can often provide cloned intact large genes, also permitting analysis of long distance regulatory elements, generally absent from cosmid constructs.

The downside of YACs is that they often contain chimeric inserts, which are DNA fragments containing two or more non-contiguous regions of the genome. This may be due to co-ligation of two different restriction fragments prior to transformation. Unstable YACs that can easily lose the insert or that have the tendency to delete internal regions from their inserts are also common, posing another disadvantage in the use of these clones. For the fine-structure analysis that must follow a long-range physical map, alternative clones based on bacterial host systems have been developed along the more traditional cosmid vectors. These
second generation clones, bacterial artificial chromosomes (BACs) and PI artificial chromosomes (PACs), have insert sizes ranging from 100 to 300 Kb (Monaco and Larin 1994). While the insert size is much smaller than that of YACs, their greater stability makes them very useful in positional cloning projects as they provide a more faithful representation of the original DNA, and they exist in many copies per cell, providing a high yield.

The easy availability of YAC libraries and the large insert capacity of the clones provide a very useful starting point to cover large disease regions and YACs have become the initial tool of choice for cloning contiguous large regions of DNA. At a later stage in contig assembly, bacterial host systems can also be incorporated to provide a more manageable and stable resource for further manipulations and for developing high-resolution physical maps.

1.6 Methods of gene identification

A critical step in positional cloning is the identification of candidate genes from a large, genetically defined region. The most commonly used methods include isolation of CpG islands, exon trapping, cDNA selection, and genomic sequencing followed by database comparisons as well as exon prediction through computational analysis of the sequence.

The development and map assignment of ESTs, which are being developed to identify expressed genes, is facilitating the construction of a gene map, thus simplifying gene identification.

1.6.1 Detection of CpG islands

CpG islands are short GC-rich sequences about 1 Kb long, which are unusually hypomethylated and are often found at 5' ends of vertebrate genes (Bird 1987). They can be easily recognised by a variety of rare-cutter restriction enzymes whose recognition site contains one or two CpG dinucleotides. This property makes them useful landmarks for identifying novel genes (Lindsay and Bird 1987). However, this method is not always valid since a substantial number of genes are not associated with CpG islands.
1.6.2 Exon trapping

The exon trapping strategy to identify genes takes advantage of the property of exons to engage in an artificial RNA splicing assay (Duyk et al. 1990, Buckler et al. 1991).

Although exon trapping has been used in the identification of many genes, including the Huntington’s disease gene (the Huntington’s Disease Collaborative Research Group, 1993) and the neurofibromatosis type 2 tumour suppressor gene (Trofatter et al. 1993), it is not always successful and is best used in combination with other strategies.

1.6.3 cDNA selection

cDNA selection involves the formation of DNA/cDNA heteroduplexes, by hybridising a PCR-amplified cDNA library to a cloned genomic DNA, such as a YAC insert, immobilised on a filter (Lovett et al. 1991, Pamiroo et al. 1991).

This technique has the advantage of enabling detection of rare or tissue specific transcripts and can be applied to several genomic clones simultaneously. A drawback is that since short fragment cDNAs constitute the selected material, to obtain a full-length coverage of a gene an entire cDNA library needs to be subsequently screened.

1.6.4 Computer-based DNA sequence analysis

Once the sequence of a DNA clone is known, computational analysis provides a powerful approach to determine whether that sequence is likely to be part of a gene. The analysis can be performed with two main types of software. The first is based on homology programs such as BLAST (Altschul et al. 1990) and FASTA (Pearson and Lipman 1988) (section 2.8.2), which searches sequence databases for coding sequences that shares similarities with the given DNA sequence both between and within species. The second is based on exon-prediction programs such as GRAIL (Uberbacher and Mural 1991), and looks for sequences in the genomic clone that share similarities with gene-specific features (i.e. ORFs, CpG islands, promoter regions, intron-exon boundaries, poly-A sites). Generally, the two types of computer programs are combined using integrated software packages such as NIX (section 2.8.4), to achieve maximum results.

Some of the limitations of in silico predictions include accurate definition of 5’ and 3’ ends of genes, and the identification of intronless transcripts and genes with small ORFs.
1.6.5 Expressed sequence tags and full-length cDNAs

The identification of a positional candidate in a disease-associated region is now often reduced to searching computer databases where lists of genes mapping to specific chromosomal regions can be obtained. The development of transcript maps, which detail the location of transcribed sequences, incorporating all available expressed sequence tags (ESTs), has made this approach more feasible. ESTs are partial sequences generated from cloned cDNA, which are being developed to allow rapid identification of expressed genes by sequence analysis. Alongside mapping of ESTs using radiation mapping, expression studies are also being performed, so that an increasing number of ESTs come with chromosomal location and expression pattern information. With the increasing number of ESTs being produced, contributing to the transcript map of the human genome, selection of ESTs as positional candidates is taking over from the “traditional” gene isolation techniques described previously.

Over several years Adams et al. (1991, 1992, 1993, 1995) contributed massively to the generation of ESTs with the production of almost 200,000 sequences. Their general approach was to generate ESTs from random primed and partial cDNA clones rather than sequencing the ends of full-length cDNA clones, which contain 5' and 3' untranslated sequences. Other genome centres who contributed to the EST sequence databases include the Genexpress Centre (Genethon, Houlgatte et al. 1995), creating 18,698 ESTs, and the Washington-University-Merck EST project in collaboration with the IMAGE consortium (Integrated Molecular Analysis of Genomes and their Expression, Lennon et al. 1996), with more than 300,000 ESTs derived from both 5' and 3' ends of cDNAs (Hillier et al. 1996). Most accumulated sequences are gathered in the dbEST division of the Genbank database (Boguski et al. 1993), which now counts 3,169,953 human ESTs. Several genome centres have reported the assembling of deposited ESTs and gene sequences into virtual transcripts representing, in some cases, full-length cDNAs. Full length cDNAs would prove more useful than the short EST sequences produced, which have often been associated with a variety of artefacts including genomic contamination, sequencing errors and improper splicing (Wolfsberg and Landsman 1997). Among the Genome centres involved in this effort is The Institute for Genomic Research (TIGR) and the UniGene division of the National Centre for Biotechnology Information (NCBI). However, because expression of a single gene may
culminate in production of several different mRNA transcripts, depending both on the gene and the source tissue, and considering also the technical challenges of converting mRNAs into cDNAs, libraries with abundant truncated products are the common result, particularly for longer mRNAs. Different groups are therefore generating and sequencing sets of unique full-length cDNA clones, which are having a major impact on the positional cloning of disease-associated genes. The European IMAGE consortium (EURO-IMAGE, [http://www.ornl.gov/meetings/wccs/euro.htm](http://www.ornl.gov/meetings/wccs/euro.htm)), founded in 1997, is among these groups. Their major goals include generating a set of non redundant cDNA clones for most human gene transcripts with a master set of unique full-length cDNA clones based upon the IMAGE Consortium resources, and obtaining high resolution and comparative functional mapping localisation in man with model organisms of genes represented in the master set. The Kazusa DNA Research Institute (Nomura et al. 1994, Nagase et al. 2000) is also conducting a large-scale collection and sequencing of full-length cDNA by isolating clones from full-length enriched human cDNA libraries made by an "Oligo-capping" method. This method involves replacement of the cap structure of mRNAs with an oligoribonucleotide. The oligo-capped mRNA is the used for first strand synthesis with dT adapter primers to construct full-length cDNA libraries. The 5' end sequences of the clones are determined and used to select putatively full-length cDNA clones, which are then entirely sequenced. Their aim is to determine the sequence of 20,000 full-length cDNA clones. The German Genome Project also founded a cDNA consortium in 1997, with the aim of characterising the complete sequences of novel human transcripts at the cDNA level. They have recently characterised 500 novel full-length cDNAs and over 1000 cDNAs corresponding to previously known genes (Wiemann et al. 2001), by using a method similar to the one employed by the Kazusa Institute. The full-length cDNA is a crucial tool both for the annotation of the human genome and for the experimental analysis of gene function, as it enables the definition of precise protein coding sequences and, in conjunction with the genomic sequence, to define gene structures and the composition of exons in alternatively spliced transcripts of the same gene. Once a full gene catalogue with full-length cDNA sequences covering every human gene has been established it will be possible to undertake large-scale and comprehensive functional analysis of human genes and proteins.
The power of comparative genomics is already being realised with the sequencing of other genomes, such as Drosophila, mouse, zebrafish and chicken (Adams et al. 2000), along with genomes of organisms already fully sequenced, i.e. Saccharomyces cerevisiae, Arabidopsis thaliana and Caenorhabditis elegans (Dujon 1996, the Arabidopsis Genome Initiative 2000, the C. elegans Sequencing consortium 1998). Parallel initiatives to characterise cDNA library ESTs from a variety of tissues for many species provides additional database resource for the characterisation of novel genes, particularly useful for understanding gene function. This allows high resolution and comparative functional mapping in human and model organisms of several genes, and is often extremely useful in the isolation of new human disease-genes.

1.7 Aims of this thesis

The primary aim of this thesis was to provide a clear genetic picture of X-linked retinal diseases and to potentially identify new disease-associated loci and genes by genetic and physical mapping and mutation screening.

The first priority was to genetically characterise all the X-linked RP and CSNB families available in the laboratory by using haplotype and linkage analysis, in order to establish the level of genetic heterogeneity for these diseases. Ultimately, for the majority of families gene screening would be the only way to follow disease in the family, and this formed the next aim of the thesis, to establish the mutation spectrum of the two known XLRP genes, RP2 and RPGR, and of one of the two CSNBX genes, NYX, by SSCP analysis and direct sequencing.

The final aim of the thesis was to employ a positional cloning strategy towards the identification of new X-linked retina disease genes, by generating a YAC contig of the RP23 region on Xp22, evaluating candidate genes by mutation screening and mapping and characterising novel transcripts contained within the region.
CHAPTER 2

Materials and methods

Note: A list of general reagents and solutions used in this study are described at the end of this chapter (section 2.9).

2.1 DNA isolation

2.1.1 Human genomic DNA preparation from blood samples

Blood samples were collected in 10 ml EDTA tubes than either extracted at the time of their arrival in the laboratory, or stored at -80°C until required.

DNA extraction was performed using the Nucleon II kit (Scotlab Bioscience). thawed blood samples were transferred to 50 ml Falcon tubes and their volume increased to 50 ml by the addition of reagent A (section 2.9.2). After mixing by inversion, the mixture was centrifuged at 4000 rpm for 5 minutes at 4°C (varifuge 3.0R, Heraeus), the supernatant was discarded and 2 ml of reagent B (section 2.9.2) was added to the pellet. The mixture was then transferred to a 5 ml tube and 500 μl of 5 M sodium perchlorate was added. After mixing briefly by hand, 2 ml of 24:1 (v/v) chloroform:isoamyl alcohol was added and the mixture was emulsified by inversion. 300 μl Nucleon® resin (provided) was than layered on top and the mixture centrifuged for 6 minutes at 1400 g. The upper DNA-containing layer was transferred to a universal tube where two volumes of ethanol were added to precipitate the DNA. After gentle inversion, the DNA pellet was picked out with a sterile needle and transferred to an eppendorf tube. It was than washed with 70% (v/v) ethanol, allowed to dry, and resuspended in a suitable amount of distilled water.

Integrity and yield of DNA was estimated by electrophoresis of a 2 μl aliquot on a 1% agarose gel (section 2.4.1). DNA samples were kept at 4°C for subsequent use, or at 20°C for long-term storage. On average the yield of DNA from 10 ml whole blood was between 200 and 300 μg.
2.1.2 Plasmid DNA preparation (QIAprep spin miniprep, QIAGEN)

The QIAprep miniprep procedure uses the modified alkaline lysis method of Birnboim and Doly (Birnboim and Doly 1979). This method is appropriate for purification of up to 20 µg of high-copy plasmid DNA from 1-5 ml overnight cultures of E. coli in LB (Luria-Bertani) medium (section 2.9.2). Colonies from a bacterial stab or glycerol stock were streaked out on LB agar plates supplemented with the appropriate antibiotic and incubated overnight at 37°C. A single colony was picked from the plate and used to inoculate 5 ml of sterile LB broth containing antibiotic. Following incubation at 37°C overnight with agitation, 3 ml of this culture was transferred in two 1.5 ml tubes and spun in a microcentrifuge for 5 minutes at 3000 rpm (centrifuge 5415D, eppendorf). The supernatant was removed and the two pellets were resuspended together in 250 µl of buffer P1 (provided). 250 µl of buffer P2 (provided) was added and the tube was gently inverted 4-6 times to mix. 350 µl of buffer N3 (provided) was then added, the tube was immediately but gently inverted 4-6 times and then centrifuged for 10 minutes at 13,000 rpm. The supernatant was applied to a QIAprep column previously placed on a 2 ml collection tube and centrifuged for 30-60 seconds at 13,000 rpm. The flow-through was discarded and the column was washed by adding 0.5 ml of buffer PB (provided) followed by centrifugation for 30-60 seconds at 13,000 rpm. The flow-through was discarded and the column was again washed by adding 0.75 ml of buffer PE (provided) and centrifuged for 30-60 seconds at 13,000 rpm. After discarding the flow-through the column was centrifuged for an additional minute to remove residual wash buffer. The column was then placed in a clean 1.5 ml tube. To elute DNA 30 µl of buffer EB (provided) was added to the centre of the column, left to stand for 1 minute, than centrifuged for 1 minute. The eluted DNA was then ready to use without additional purification steps.

2.1.3 Yeast DNA preparation

This protocol yields DNA that is fragmented but is sufficient for PCR analysis. 5 ml of AHC broth supplemented with ampicillin (35 µg/ml) was inoculated with a single YAC colony and incubated overnight at 30°C in a shaking incubator. Cells were pelleted by centrifuging at 4000 rpm for 10 minutes. After discarding the supernatant, the cells were
Materials and methods

Resuspended in 500 μl of solution I (section 2.9.2) and transferred to a 1.5 ml tube. After the addition of 50 μl of lyticase (1 mg/ml), the mixture was incubated for 60 minutes at 37°C. The cells were then pelleted by centrifugation for 1 minute at 10,000 rpm, the supernatant was discarded and the cells resuspended in 500 μl of solution II (section 2.10.1). After the addition of 50 μl of 10% SDS the solution was mixed well and incubated for 30 minutes at 65°C. 200 μl of 5 M potassium acetate was then added and the tube was placed on ice for 5 minutes. After centrifugation for 5 minutes at 10,000 rpm the supernatant was transferred to a clean 1.5 ml tube, one volume of isopropanol was added and the mixture was left for 5 minutes at room temperature. This was followed by centrifugation for 10 seconds and the supernatant was then removed. The pellet was washed once with 200 μl of 70% ethanol, air-dried for 10 minutes and resuspended in 300 μl of distilled water. Samples were stored at -20°C. 1 in 20 dilutions were used for PCR analysis.

2.2 Purification of DNA

2.2.1 Phenol chloroform extraction

For small volumes of DNA, the extraction volume was increased to 200 μl with 1x TE. Larger volumes were extracted directly. An equal volume of phenol: chloroform: isoamyl alcohol (25:24:1) was mixed gently into the solution by inversion. The samples were then centrifuged at 6000 rpm for 3 minutes to separate the aqueous and organic layers. The top aqueous layer was transferred to a clean tube and the phenol extraction was repeated if necessary, depending on the protein content of the sample. An equal volume of chloroform was added, mixed and re-spun. The aqueous samples were transferred to a fresh tube for ethanol precipitation.

2.2.2 Ethanol precipitation

For standard DNA precipitations, two volumes of absolute ethanol and 1/10th volume of 3 M sodium acetate were added to the solution and placed at -80°C for 30 minutes to allow the DNA to precipitate. The samples were then spun at 13,000 rpm for 15 minutes. The pellet of DNA was washed with 200 μl 70% ethanol and the samples were centrifuged again.
for 5 minutes. The pellet was either air-dried or was placed under vacuum for approximately 10 minutes. The samples were resuspended in distilled water.

2.2.3 Sepharly-S200 and S400 HR columns (Pharmacia)

S200 HR columns were used for desalting, buffer exchange, and the removal of labelled or unlabelled nucleotides from DNA solutions, providing the DNA fragment was >100 bp. S400 HR columns serve the same purpose, but also remove primers or excess primer dimers from PCR products prior to sequencing. The Sephacryl HR resin (sephacryl equilibrated in TE buffer, pH 7.6) was resuspended in the column by gentle vortexing. The cap was loosened and the base snapped off before placing in an open topped eppendorf. This was spun at 3000 rpm for 1 minute to compact the gel matrix. The column was transferred to a fresh tube, and 20-25 μl of the DNA sample was added to the bed of resin. The sample was collected at the bottom of the tube by centrifuging again for 1 min at 3000 rpm.

2.2.4 Centricon 100 spin columns (Centricon)

These columns were 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 DNA to remain on the membrane while all unincorporated primers and dNTPs pass through). The column was than inverted and centrifuged for 5 minutes at 3000 g allowing the DNA elute to be collected. The purified products were directly used for cycle sequencing.

2.2.5 The enzymatic method (ExoI and SAP)

This is an alternative quicker and cheaper method for purifying PCR products prior to sequencing. An aliquot of the amplification product (8 μl) was purified by the addition of 1 U shrimp alkaline phosphatase (SAP; Amersham Life Science) and 1 U Exonuclease I (United States Biochemical) in SAP buffer. Samples were incubated for 30 minutes at 37°C followed by 15 minutes at 80°C. The purified products were directly used for cycle sequencing.
Materials and methods

2.2.6 QIAquick spin columns (QIAGEN)

2.2.6.1 QIAquick PCR purification

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 (provided) were added to 1 volume of the PCR reaction mix, the sample was than applied to a QIAquick spin column previously placed in a 2 ml collection tube, and centrifuged for 30-60 seconds at 13,000 rpm. The flow-through was discarded and the column was placed back into the same tube. To wash, 0.75 ml of buffer PE (provided) was added to the column, followed by centrifugation for 30-60 seconds at 13,000 rpm. 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 than placed in a clean 1.5 tube. To elute DNA, 30 µl of buffer EB (provided) was added to the centre of the column, left to stand for 1 minute and spun for 1 minute at 13,000 rpm. The eluted DNA was directly used for cycle sequencing.

2.2.6.2 QIAquick gel extraction

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 (provided) were added to 1 volume of gel and the mixture was than 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 spun for 1 minute at maximum speed. 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 13,000 rpm. 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 13,000 rpm. 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 than
Materials and methods

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 13,000 rpm. The eluted DNA was directly used for cycle sequencing.

2.3 Restriction enzyme digestion of DNA

Restriction endonucleases are bacterial enzymes, which cleave double-stranded DNA into discrete pieces, resolvable by gel electrophoresis (section 2.4.1). 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. 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.

2.4 Size fractionation of DNA

2.4.1 Agarose gel electrophoresis

DNA preparations, DNA fragments resulting from restriction enzyme digestion of genomic DNA (section 2.3), cloned DNA and non-radioactive PCR products (section 2.6.1) were size separated through agarose gels containing 0.5 μg/ml ethidium bromide DNA stain. Gels were prepared by mixing an appropriate amount of agarose (BIO-RAD) and 1x TAE buffer, melting in a microwave oven, cooling to ~55°C and adding stain before pouring into a
sealed gel casting tray. The concentration of the gel was varied to achieve optimal resolution according to fragment size; 0.8-1% for fragments of 2-20 kb, 1-2% for fragments of 1-2 kb and 3% for fragments of 70 bp-1 kb. DNA samples were mixed with 1/5 volume 6x loading buffer (ABgene) and electrophoresed at 80V for ~1 hour (Midicell® Primo EC330 electrophoretic gel system, E-C Apparatus Corporation) in 1x TAE buffer along with an appropriate size standard marker (λHindIII for fragments 1-20 Kb and ΦX174HaeIII (Promega) for smaller products). The resolved fragments were visualised on a UV transilluminator and photographed under UV light using a Polaroid MP4 camera and Kodak plus-X Estar film.

2.4.2 Polyacrylamide gel electrophoresis

Apparatus used for running microsatellite gels was purchased from BIO-RAD. Glass plates were cleaned from dried gel and soap residues by applying ethanol to both plates and wiping dry. The back plate (holding the buffer reservoir) was silanised before assembly (Sigmacote, Sigma) to ease plate separation and reduce the risk of gel tearing on completion of electrophoresis. 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 (10x TBE in 8.3 M urea) to give a 6% gel solution in 8.3 M urea and 1x TBE (gel base; total volume 150 ml for a 40x60 cm gel rig, 100 ml for a 20x40 cm gel rig).

A plug was first prepared to seal the bottom end of the plates prior to pouring the proper gel. 150 µl of 25% (w/v) ammonium persulphate and 150 µl of TEMED were added to 25 ml of the above gel base and quickly mixed before pouring into the sealing tray (this sets very quickly). Once the plug was set (allowing at least 15 minutes), 25% (w/v) ammonium persulphate (500 µl for a 40x60 cm gel rig; 300 µl for a 20x40 cm gel rig) and TEMED (56 µl for the large gel rig, 34 µl for the small) were added to the remaining gel base and quickly mixed before pouring the gel slowly and steadily to avoid air bubbles becoming trapped. The running buffer used was 1x TBE. 6 µl of formamide loading buffer was added to 10 µl of radioactively labelled microsatellite PCR products (section 2.6.4). Samples were than heat denatured at 95°C for 5 minutes and ice-chilled before loading 2-3 µl onto a prewarmed (50-55°C) 6% denaturing polyacrylamide gel, after flushing out excess urea from
the wells. Electrophoresis was carried out at a constant power of 90-100 W (large gel) or 45-50 W (small gel) for an appropriate length of time (2-5 hours depending on the size range of the allele system) to achieve maximum resolution in the size required. Gels were then fixed in a 10% methanol/10% acetic acid (v/v) solution for 5 minutes, covered in clingfilm and vacuum dried on 3MM Whatmann paper at 70°C for 1 hour and autoradiographed overnight at room temperature (RT) using Kodak X-OMAT or Fuji X-ray film. When required, gels were autoradiographed at -80°C with an intensifying screen to enhance the signal.

2.4.3 MDE gel electrophoresis

The apparatus used for running single strand conformation polymorphism (SSCP) gels was the same as that used for microsatellite gels. Glass plates were cleaned and silanised as described above. MDE™ Gel solution (Flowgen) was mixed with 10x TBE buffer and distilled water to give a 0.5x acrylamide gel solution in 0.6x TBE (gel base; total volume 150 ml for a 40x60 cm gel rig, 100 ml for a 20x40 gel rig).

A plug was first prepared as described above adding the same amounts of 25% ammonium persulphate and TEMED. The gel was poured as described, again adding the same amounts of ammonium persulphate and TEMED. 0.6x TBE was used as running buffer. Appropriate loading buffer (section 2.9.3) was mixed 1:1 with the radioactively labelled PCR products (section 2.6.4). Denaturation and loading of samples was carried out as described above. Electrophoresis was carried out at a constant power of 8-10 W (large gel) or 4-5 W (small gel) for 14-18 hours depending on the size of the PCR product (approximately 14 hours for fragments ~200 bp, 18 hours for fragments ~300 bp). For convenience, gels were run overnight at RT, they were than fixed, vacuum dried and autoradiographed as described above.

2.5 RNA manipulation

2.5.1 RNA handling

All equipment, and work areas were kept thoroughly clean and free of commercial ribonucleases during manipulations with RNA. Diethyl pyrocarbonate (DEPC) treated water (100 µl in 1 litre of distilled water) was routinely used for all solutions. DEPC was added to
the distilled water and left to incubate overnight in a fume hood (a minimum time of 12 hours is recommended) prior to autoclaving. Where possible, the chemicals used were reserved separately from other uses. Filter tips and sterile disposable plastic-ware were utilised and gloves were used at all times.

2.5.2 Isolation of total RNA from fibroblast cell lines

2.5.2.1 Harvesting of cells

Human fibroblast cell line (293T) was grown in 75 ml flasks with the addition of 20 ml DMEM media (section 2.9.4) 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 (section 2.9.4). To dislodge cells from the bottom of the flask 1 ml of trypsin/EDTA (GibcoBRL) was added and left at RT 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 (Zeiss, lens Zeiss Ph1 F10/0.25, eye piece CPLW 10x/18). The total number of cells (n_{TOT}) was derived from the number of cells counted (n) using the following formula:

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

The cell suspension was then transferred to a 50 ml Falcon tube and cells were pelleted by centrifuging at 1200 rpm for 5 minutes at RT. The supernatant was discarded and the pellet was directly used for RNA extraction (section 2.5.2.2).

2.5.2.2 RNeasy mini protocol (QIAGEN)

Depending on the total number of cells in the pellet, the appropriate amount of buffer RLT (provided) containing β-mercaptoethanol was added (350 µl for up to 5x10^6 cells and 600 µl for 5x10^6 to 10^7 cells), the mixture was then vortexed or pipetted until no more clumps were visible. The lysate was passed at least 5 times through a 20-G (Ø 0.9 mm) needle fitted to a syringe to homogenise. One volume of 70% ethanol (in DEPC treated water) was added to the homogenised lysate and mixed well by pipetting. Up to 700 µl of
sample was applied to an RNeasy mini spin column sitting on a 2 ml collection tube and centrifuged for 15 seconds at 10,000 rpm. The flow-through was then discarded and the collection tube reused. If the volume of the mixture exceeded 700 µl the remaining aliquot was successively added to the column and spun as above. 700 µl of buffer RWI (provided) was added to the column and spun for 15 seconds at 10,000 rpm to wash. The column was transferred into a new 2 ml collection tube and 500 µl of buffer RPE (provided) was added prior to spinning for 15 seconds at 10,000 rpm to wash. After discarding the flow-through and placing the column in the same collection tube another 500 µl of buffer RPE was added followed by centrifugation at maximum speed for 2 minutes to dry the RNeasy membrane and a second centrifugation for 1 min again at full speed after discarding the flow-through. The column was then transferred to a new 1.5 ml tube (supplied) prior to pipetting 40 µl of RNase-free water (provided) directly onto the membrane. This was followed by centrifugation for 1 minute at 10,000 rpm. The eluted RNA was stored at −80°C or directly used for RT-PCR (section 2.6.5).

2.6 Polymerase chain reaction (PCR)

2.6.1 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 Taq polymerase from Bioline using the manufacturer’s 10x buffer containing 10 mM Tris-HCl, 50 mM KCl, 15 mM MgCl₂ and 0.1% 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 (6µg/µl) 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. In terms of cycling conditions a routine initial denaturation step of 95°C for 3 minutes was followed by cycling parameters of denaturing at 95°C for 15 seconds, annealing at the appropriate
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. Either Hybaid Omnigene or Techne PHC-3 or Techne Genius PCR thermocyclers were used.

2.6.2 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 (i.e. self-complementarity) 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 (section 2.6.3) 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 such as Biolines or GenoSys.

2.6.3 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 (along with a ‘no DNA’ control) and electrophoresed on agarose gel (section 2.4.1) 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.
Materials and methods

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 Tm, 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 (NH₄)₂SO₄, 670 mM Tris-HCl (pH 8.8 at 25°C) and 0.1% Tween-20 was used. Appropriate amounts of 50 mM MgCl₂ 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 is to ensure that the DNA molecules were fully denatured before the polymerisation started to proceed.

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

2.6.4 Microsatellite and SSCP analysis

Radioactive PCR reactions were performed in a total volume of 10 µl with a reaction mix as described in section 2.6.1, with the final dNTPs concentration 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 0.5 µCi α-³²P-dCTP added per reaction (i.e. 10 µCi of a α-³²P-dCTP was added to a reaction mix for 20 reactions). The template DNA was dispensed into 0.5 ml eppendorfs and overlaid with mineral oil prior to the addition of the reaction mix.

The microsatellite allele size differences (multiple of 2-4 bp) and the single strand conformational polymorphism differences were resolved respectively by polyacrylamide and MDE gel electrophoresis (sections 2.4.2 and 2.4.3).
2.6.5 Reverse Transcriptase (RT) PCR

First strand cDNA synthesis was performed using the Ready-to-go You-Prime First-Strand Beads (Pharmacia). Two different aliquots (2 and 5 μl) of total RNA extracted from fibroblast cell lines (section 2.5.2.2) 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 RT for approximately 1 minute and than 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.

2.6.6 YAC library screening by PCR

Human YAC clones from the CSNB4 region in Xp11.4 and from the RP23 region in Xp22.3 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.

For the CSNB4 region a tentative physical map was constructed through resources available on the database (section 2.10, particularly using GDB and IXDB) and then confirmed and correctly ordered through STS content screening. For the RP23 region a comprehensive physical map was already published (Ferrero et al. 1995), clones from this map were ordered and confirmed again through STS content screening to make sure the in-house map was correct.

Three different human YAC libraries were used, the ICI (Anand et al. 1990) and ICRF (Larin et al. 1991) libraries available both from HGMP, and the CEPH mega-YAC library (Chumakov et al. 1992) obtained directly from CEPH.

YAC clones were received either as ‘stabs’ or streaked out on selective agar, and were restreaked onto fresh AHC+Ampicillin agar plates (section 2.9.2) to initiate their
growth as single colonies. DNA was than extracted as described in section 2.1.3 and PCR with the relative STS primers was performed as described in section 2.6.1.

### 2.6.7 Colony PCR

This method was very useful for preliminary screening prior to DNA extraction from yeast clones. The yeast 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 solution was incubated at 37°C for 5 minutes. 3 μl of the mixture was than used as a template for PCR. The method proved fast, easy and reliable.

### 2.7 Automated DNA sequencing

Automated sequencing was performed on an ABI 373A DNA sequencer (Perkin Elmer), using the ABI PRISM® BigDye™ Terminator Cycle Sequencing Ready reaction Kit which contains the sequencing enzyme *AmpliTaq*® DNA polymerase, FS. Automated cycle was adapted from the chain-termination sequencing method developed by Sanger *et al.* (1977). This is based on enzymatic synthesis of a template strand by polymerase action and termination of the extended products by the incorporation of dideoxynucleotides, which lack hydroxyl residues at the 3’ position of the deoxyribose. This prevents the formation of a phosphodiester bond with free nucleotides, thereby leading to termination of chain synthesis. The automated technique is based on using 4 fluorescent dye-labelled terminators, which are detected by lasers that analyse and convert the DNA sequence to a graphical image on a computer. Reactions are carried out in single tubes, and do not have to be separated for each terminator nucleotide. Only one lane on the denaturing gel per reaction is required, as individual labelled nucleotides can be distinguished by unique fluorescent labels. Also the results can be seen immediately in a computer analysis file.

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 using one of the methods described previously (section 2.2). The cycle
sequencing reaction for both cloned and PCR fragments consisted of 4 µl terminator ready reaction mix (includes dye terminators, buffer, magnesium chloride, dNTPs and AmpliTaq polymerase), 1.6 pmol sequence specific primer and the appropriate amount of template DNA (100-250 ng for cloned DNA, 10-50 ng for PCR product) in a total volume of 10 µl. This was subjected to 25 cycles of 96°C for 10 seconds, 50°C for 5 seconds and 60°C for 2 minutes (for PCR products) or for 4 minutes (for longer templates, i.e. clones) in a thermal cycler. Excess unincorporated fluorescent dye was removed by ethanol precipitation of the DNA; 26 µl of distilled water and 64 µl of ethanol 95% were added to each sequencing reaction vortexed briefly and left at RT for at least 15 minutes. Tubes were then spun at maximum speed for 20 minutes and the supernatant was immediately aspirated. 250 µl of 70% ethanol was added to the tubes, which were than vortexed and spun for 5 minutes at maximum speed. The supernatant was immediately aspirated and pellets were left to dry at room temperature for 30-60 minutes or in a vacuum centrifuge (Uniscience, Univap) without heat. The resultant pellet was resuspended in 3 µl of ABI loading buffer (section 2.9.3) and the sample was denatured before loading onto the denaturing acrylamide gel. The ABI 373A DNA sequencer was set up and run in accordance with the manufacturer’s instructions. The gel was generally run for 12 hours. The output data was converted to a text file and an analysis file on an Apple Macintosh computer.

2.8 Computer-aided analysis

2.8.1 Linkage analysis

Allele information was processed using the software program Cyrillic 2.1.3 (Cherwell Scientific), which also enables pedigree drawing, and two-point lod scores were calculated using the program MLINK included with the package. The pedigrees were analysed as X-linked recessive traits, with equal allele frequencies and a disease gene frequency of 0.0001.

2.8.2 Sequence similarity searches

The BLAST program (http://www.ncbi.nlm.nih.gov/BLAST/) (Altschul et al. 1990) and FASTA program (http://www.hgmp.mrc.ac.uk/Registered/Webapp/fasta/) (Pearson and Lipman 1988) are methods which utilise rapid database searching algorithms for finding
optimal sequence alignments between query and database comparisons. The BLAST program is a fast yet less sensitive mode of similarity searching as it disregards large gaps to improve the alignment, whereas FASTA does take into consideration gaps in sequences. They both determine the smallest sum probability of gaining the resultant alignment from the query and database sequences by chance. DNA sequence databases were searched with the query sequence using BLASTN, or the query sequence was translated in six reading frames to search protein databases using BLASTX. Searches were done mainly against nr (non-redundant DNA sequences), month (last month inputs in the database), dbEST (expressed sequences database) and swissprot (protein database).

2.8.3 DNA analysis

The software program DNASTAR (DNASTAR, Inc. Wisconsin USA) (including EditSeq™, GeneQuest™, MapDraw™, MegAlign™, PrimerSelect™, Protean™ and SeqMan™) was used as a tool to analyse DNA sequences. This analysis included editing of sequences, alignments, restriction maps and ORF finder.

2.8.4 Gene analysis

The NIX program (http://www.hgmp.mrc.ac.uk/Registered/Webapp/nix/) is a tool to view the results of many DNA analysis programs applied on a DNA sequence. The analysis programs include exon prediction programs (GRAIL, Fex, Hexon, MZEF), gene prediction programs (Genemark, Genefinder, Fgene), BLAST (against many databases), Polyah (polyadenylation-signal identifier), RepeatMasker, tRNAscan (transfer RNA identifier). NIX is intended as a tool to aid the identification of interesting regions in Genomic or transcribed nucleic acid sequences. Viewing the results of many such programs side by side makes it easy to see when many programs have a consensus about a feature.

2.8.5 Protein analysis

Similarly to NIX, PIX (http://www.hgmp.mrc.ac.uk/Registered/Webapp/pix/) is a tool to view the results of many peptide analysis programs applied on a peptide sequence. Analysis carried out on the peptide sequence include cell localisation and secondary structure
prediction, blast searches against sequence and domain databases, motif and domain databases searches, low complexity regions and long/short globular domains searches, coiled-coil, transmembrane, helix-turn-helix, signal peptide, antigenic regions and enzyme digest predictions.

2.9 Reagents and buffers

2.9.1 General reagents

2 M Tris-HCl (trizma hydrochloride pH 7.5, or 8): adjust pH to 7.5 or 8.0 with 5 M NaOH, autoclave before use.

0.5 M EDTA (ethylenediaminetetraacetic acid pH 8.0): adjust pH to 8.0 with 5 M NaOH, autoclave before use.

5 M NaOH (sodium hidrooxide): prepared with SDW.

10% (w/v) SDS (sodium dodecyl sulphate): prepared with SDW.

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

20x SSC: 3 M NaCl, 0.3 M tri-sodium-citrate

2.9.2 DNA isolation

Human genomic DNA

Reagent A (5x lysis buffer): 320 mM sucrose, 10 mM Tris HCl (pH 7.5), 5 mM MgCl₂ (adjust pH to 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 (adjust pH to 8.0 with 5 mM NaOH), autoclave than add 1% SDS.

Plasmid DNA

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

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

Ampicillin: 50 mg/ml stock; 1 μl per 1 ml media (f.c. 50 μg/ml)
(NB: heat-sensitive)
Materials and methods

All the reagents for the QIAgen miniprep were provided with the kit.

Yeast DNA

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, autoclave before use.

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

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

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

2.9.3 Gel electrophoresis

Ethidium bromide (10 mg/ml): 1 g ethidium bromide/100 ml distilled water.

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% bromophenol blue, 0.3% xylene cyanol, 10 mM EDTA.

MDE gels: 95% formamide, 0.05% bromophenol blue, 0.05% xylene cyanol, 20 mM EDTA. Mix with 100 mM NaOH (4:1).

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

2.9.4 Cell lines manipulation

Media: DMEM (Dulbecco’s MEM) with Glutamax-I (from GibcoBRL), with the addition of 5% (v/v) foetal calf serum and 0.1% (v/v) penicillin-streptomycin (from GibcoBRL)

PBS: 10 mM phosphate, 0.95% (w/v) NaCl pH 7.4
2.10 Electronic databases used

Human Genome Mapping Project (HGMP): http://www.hgmp.mrc.ac.uk/
The Integrated X Chromosome Database (IXDB): http://ixdb.molgen.mpg.de/
Human Genome Database (GDB): http://www.gdb.org
Genethon: http://www.genethon.fr/genethon_en.html
Whitehead: http://www-genome.wi.mit.edu/
TIGR: http://www.tigr.org/tdb/hgi/index.html
GeneCards: http://bioinfo.weizmann.ac.il/cards/
Sanger Centre: http://www.sanger.ac.uk/HGP/
RetNet: 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

Haplotype and linkage analysis in families with X-linked retina disease

3.1 Introduction

The X chromosome is home to a large number of retina disease genes as discussed in section 1.1.4. The aim of the research presented in this chapter was to use the clinical resource available through Moorfields Eye Hospital and other clinical colleagues to define disease gene locations for X-linked retinitis pigmentosa and X-linked congenital stationary night blindness.

3.1.1 X-linked retinitis pigmentosa

X-linked retinitis pigmentosa (XLRP) is the most severe clinical form of RP, with an incidence of about 1:20,000 (Jay 1982, Heckenlively 1988). Male XLRP patients generally develop night blindness and reduced visual acuity as a consequence of the degeneration of rods in the periphery of the retina. As disease progresses, cones start to degenerate producing a gradual impairment of the central vision which culminates in total blindness within the third or fourth decade of life (Bird 1975). Female carriers show variable symptoms of the disease upon ophthalmological testing.

Genetic analysis of XLRP families has shown an unexpected high degree of heterogeneity. At the time when this work was initiated there appeared to be two major loci accounting for RP on the X chromosome, RP2 and RP3 (Bhattacharya et al. 1984, Thiselton et al. 1996, Ott et al. 1990). The gene at the RP3 locus (RPGR, section 4.1.1) was cloned in 1996 (Meindl et al 1996). Other reported loci included RP6, postulated by statistical analysis and RP15 (Ott et al. 1990, McGuire et al. 1995). Figure 3.1a shows an ideogram of the X chromosome with map location of XLRP loci as it was at this stage.

During the course of this study, worldwide efforts to unravel the complete genetic picture underlying this disease led to the identification of the RP2 gene (Schwahn et al. 1998)
Haplotype and linkage analysis in families with X-linked retina disease

and of two other XLRP loci, RP23 (section 5.3.2, Hardcastle et al. 2000) and RP24 (Gieser et al. 1998). Moreover, the presumed dominant cone-rod dystrophy family previously mapped to the RP15 locus was shown to be, in fact, an RPGR family (Mears et al. 2000). Figure 3.1b shows an updated ideogram of the X chromosome with map locations of XLRP loci and genes.

In this study haplotype analysis was performed on 33 XLRP families using polymorphic markers spanning the entire X chromosome. The aim of this analysis was firstly to determine the particular XLRP gene segregating within each family, and secondly to firmly define and if possible refine each of the XLRP loci or to identify new XLRP loci (see chapter 5).

During this result chapter I will be referring to RP3 and RP2 loci rather than the RPGR and the RP2 genes, to take into account the possibility of microheterogeneity around the RP3 locus (which will be discussed in the next chapter) and the fact that at the start of this study, the RP2 gene had not been cloned.

3.1.2 X-linked congenital stationary night blindness

Congenital stationary night blindness (CSNB) is a retina disease characterised by non-progressive night blindness (hence the term stationary), and is associated with reduced visual acuity (Pearce et al. 1990). The X-linked form of this disease (CSNBX) is frequently associated with myopia.

Genetic analyses over the last decade have established genetic heterogeneity for CSNBX, implicating at least two distinct genes on the proximal short arm of the X chromosome. At the outset of this study the disease gene presumably accounting for the complete form (CSNB1) had been localised to a 10 cM region between markers DXS556 and DXS8083 in Xp11.4-p11.3 (Bergen et al. 1995, Hardcastle et al. 1997, Boycott et al. 1998), whilst the CSNB2 gene had been localised to a 1.2 cM region between markers DXS722 and DXS255 in Xp11.23 (Boycott et al. 1998). Figure 3.1c shows an ideogram of the X chromosome with map location for these two types of CSNBX.

The CSNB1 locus was then refined on Xp11.4 to between markers DXS1068 and DXS6810 (Rozzo et al. 1999) and the gene for the incomplete type of CSNB on Xp11.23 was identified as a calcium-channel alpha-1-subunit gene (CACNA1F, Bech-Hansen et al.)
Figure 3.1: Ideogram of the human X chromosome showing map locations of XLRP loci (a: at the time when this study begun, b: at present) and CSNBX loci (c)
1998, Strom et al. 1998). More recently the gene for the complete form of CSNBX, the NYX gene, has also been cloned (Bech-Hansen et al. 2000 and Pusch et al. 2000) (Figure 3.1c).

Haplotype analysis was performed on 4 X-linked CSNB families using polymorphic markers spanning the CSNB1 and CSNB2 loci on Xp11.4 and Xp11.23 respectively. The aim of this analysis was to identify the gene segregating with disease for each family, and possibly refine each of the two CSNBX regions.

During this results chapter, I will be referring to CSNB2 and CSNB1 loci rather than the CACNA1F and the NYX genes, as neither gene was cloned at the start of this analysis.
3.2 Materials and methods

3.2.1 Families

By pedigree analysis an X-linked form of RP (male family members more severely affected than female carriers and no male-to-male transmission) was established. The diagnosis of disease (either RP or CSNB) in affected members was confirmed by ophthalmological examination including fundoscopy, visual field assessment, electroretinograms (ERGs) and dark adaptation. Females at risk showing the ophthalmological changes of carrier status, which were confirmed by electrophysiological tests, were taken to be presumptive heterozygotes; otherwise the genetic status of females was not inferred unless the pedigree structure suggested they were obligate carriers. DNA was extracted from venous blood samples according to methods described in section 2.1.1.

3.2.1.1 XLRP families

A panel of 33 XLRP families was examined in this study by haplotype analysis. F53, F75, RP4, RP34, RP76, RP87, RP90, RP91, RP158, RP180, RP296, RP639, RP1571, RP1591, RP1750, RP2063, RP2689, RP2786, RP2788, RP3126, RP3877 and RP3973 are all British families and were clinically ascertained through Moorfields Eye Hospital. Two other British families, RP759 and RPO1, were obtained respectively via the Genetics Clinic at Newcastle General Hospital (Dr. Grace O’Hearn) and via the Department of Clinical Genetics at Oxford Redcliffe Hospital (Dr. Sue White). An Italian family, RP1600, was provided by Dr. Benedetto Falsini (Institute of Ophthalmology, Catholic University, Rome), DNA samples from Belgian families RP1L and RP2L were donated by Dr. Koen Devriendt (Centre for Human Genetics, Leuven), whilst blood samples from another Belgian family, RP1B, were donated by Dr. Lionel Van Maldergem (Institut de Pathologie et de Genetique, Loverval). NZ1, a New Zealand family, was provided by Dr. Ingrid Winship (Molecular Genetics Laboratory, Auckland Hospital). RP1120, an American family, was donated by Dr. Tim Stout (Department of Ophthalmology, Childrens Hospital, Los Angeles) and MGRP4, RP106 and RP107, three other American families, were provided by Dr. Michael Gorin (Department of Ophthalmology, University of Pittsburgh).
3.2.1.2 CSNBX families

Samples from 4 X-linked CSNB families (RP2124, RP3810, GC4180 and RP4914) were collected and analysed in this study. These are all British families, which were clinically ascertained through Moorfields Eye Hospital.

3.2.2 Genetic markers

54 microsatellite markers spanning the entire X chromosome were used in this study for haplotype analysis. Details of these markers, including chromosomal location, level of heterozygosity, PCR product size and annealing temperatures are shown in Table 3.1. Primer sequence information was available through GDB (http://www.gdb.org), Genethon (http://www.genethon.fr/genethon_en.html) or Thiselton et al. (1995), whilst approximate genetic distances in cM (from a marker to the next marker down in the list) were taken from the Genetic Location Database (http://cedar.genetics.soton.ac.uk/public_html/ldb.html), GDB, Genethon, Thiselton et al. (1995), or from the reports of the international workshops on X chromosome mapping.

Table 3.1 Microsatellite markers used for genetic characterisation of XLRP and CSNB families

<table>
<thead>
<tr>
<th>Marker name</th>
<th>Chromosome location</th>
<th>Heterozygosity</th>
<th>Size (in bp)</th>
<th>Annealing T (°C)</th>
<th>Primer sequence information</th>
<th>Genetic distances</th>
</tr>
</thead>
<tbody>
<tr>
<td>DXS1233</td>
<td>Xp22.33</td>
<td>0.43</td>
<td>-</td>
<td>53°C</td>
<td>GBD</td>
<td></td>
</tr>
<tr>
<td>DXS996</td>
<td>Xp22.32</td>
<td>0.82</td>
<td>153-171</td>
<td>56°C</td>
<td>Genethon</td>
<td>5 cM</td>
</tr>
<tr>
<td>DXS1223</td>
<td>Xp22.32</td>
<td>0.75</td>
<td>164</td>
<td>55.5°C</td>
<td>Genethon</td>
<td>2.9 cM</td>
</tr>
<tr>
<td>DXS7103</td>
<td>Xp22.31</td>
<td>0.62</td>
<td>-230</td>
<td>57.5°C</td>
<td>Genethon</td>
<td>1.9 cM</td>
</tr>
<tr>
<td>DXS1224</td>
<td>Xp22.31</td>
<td>0.61</td>
<td>162</td>
<td>56°C</td>
<td>Genethon</td>
<td>5.5 cM</td>
</tr>
<tr>
<td>DXS999</td>
<td>Xp22.2</td>
<td>0.75</td>
<td>260-276</td>
<td>58°C</td>
<td>Genethon</td>
<td>7 cM</td>
</tr>
<tr>
<td>DXS7593</td>
<td>Xp22.13</td>
<td>0.71</td>
<td>210-230</td>
<td>62°C</td>
<td>Genethon</td>
<td>5.3 cM</td>
</tr>
<tr>
<td>DXS7110</td>
<td>Xp22.13</td>
<td>0.71</td>
<td>135-155</td>
<td>54°C</td>
<td>Genethon</td>
<td>6.3 cM</td>
</tr>
<tr>
<td>DXS989</td>
<td>Xp22.13</td>
<td>0.82</td>
<td>170-200</td>
<td>58°C</td>
<td>Genethon</td>
<td>0.9 cM</td>
</tr>
<tr>
<td>DXS8027</td>
<td>Xp22.12</td>
<td>0.66</td>
<td>230-240</td>
<td>58°C</td>
<td>Genethon</td>
<td>0 cM</td>
</tr>
<tr>
<td>DXS8039</td>
<td>Xp21.3</td>
<td>0.80</td>
<td>145-165</td>
<td>57°C</td>
<td>Genethon</td>
<td>2.6 cM</td>
</tr>
<tr>
<td>DXS1214</td>
<td>Xp21.2</td>
<td>0.79</td>
<td>215-135</td>
<td>58°C</td>
<td>Genethon</td>
<td>3 cM</td>
</tr>
<tr>
<td>DXS8090</td>
<td>Xp21.1</td>
<td>0.81</td>
<td>150-170</td>
<td>56°C</td>
<td>Genethon</td>
<td>9.3 cM</td>
</tr>
<tr>
<td>DXS572</td>
<td>Xp21.1</td>
<td>0.29</td>
<td>155-165</td>
<td>56°C</td>
<td>Thiselton et al. 1995</td>
<td>0.5 cM</td>
</tr>
<tr>
<td>DXS1242</td>
<td>Xp21.1</td>
<td>0.77</td>
<td>210-130</td>
<td>56°C</td>
<td>GBD</td>
<td>5 cM</td>
</tr>
<tr>
<td>CYBB</td>
<td>Xp21.1</td>
<td>0.76</td>
<td>145-160</td>
<td>66°C</td>
<td>GBD</td>
<td>1 cM</td>
</tr>
<tr>
<td>DXS1110</td>
<td>Xp21.1</td>
<td>0.70</td>
<td>250-270</td>
<td>60°C</td>
<td>GBD</td>
<td>0.65 cM</td>
</tr>
<tr>
<td>DXS8349</td>
<td>Xp21.1</td>
<td>-</td>
<td>-220</td>
<td>58°C</td>
<td>GBD</td>
<td>0.43 cM</td>
</tr>
<tr>
<td>DXS8170</td>
<td>Xp21.1</td>
<td>0.45</td>
<td>170-180</td>
<td>57°C</td>
<td>GBD</td>
<td>1.2 cM</td>
</tr>
<tr>
<td>M6</td>
<td>Xp21.1</td>
<td>-</td>
<td>-200</td>
<td>57°C</td>
<td>Meindl, pers. comm. 0.2 cM</td>
<td></td>
</tr>
<tr>
<td>DXS6679</td>
<td>Xp11.4</td>
<td>-</td>
<td>120</td>
<td>60°C</td>
<td>GBD</td>
<td>1.8 cM</td>
</tr>
</tbody>
</table>
Haplotype and linkage analysis in families with X-linked retina disease

<table>
<thead>
<tr>
<th>Marker name</th>
<th>Chromosomal location</th>
<th>Heterozygosity</th>
<th>Size (in bp)</th>
<th>Annealing T (°C)</th>
<th>Primers sequence information</th>
<th>Genetic distances</th>
</tr>
</thead>
<tbody>
<tr>
<td>DXS1068</td>
<td>Xp1.4</td>
<td>0.82</td>
<td>250-260</td>
<td>57°C</td>
<td>Genethon</td>
<td>2 cM</td>
</tr>
<tr>
<td>DXS556</td>
<td>Xp1.4</td>
<td>0.76</td>
<td>175-192</td>
<td>56°C</td>
<td>GBD</td>
<td>1 cM</td>
</tr>
<tr>
<td>DXS574</td>
<td>Xp1.4</td>
<td>0.61</td>
<td>200-215</td>
<td>55°C</td>
<td>Thiselton et al. 1995</td>
<td>1.5 cM</td>
</tr>
<tr>
<td>DXS993</td>
<td>Xp1.4</td>
<td>0.79</td>
<td>290-310</td>
<td>57°C</td>
<td>GBD</td>
<td>0.5 cM</td>
</tr>
<tr>
<td>DXS8012</td>
<td>Xp1.4</td>
<td>0.79</td>
<td>178</td>
<td>62°C</td>
<td>Genethon</td>
<td>0 cM</td>
</tr>
<tr>
<td>DXS977</td>
<td>Xp1.4</td>
<td>0.60</td>
<td>148-162</td>
<td>50°C</td>
<td>GBD</td>
<td>1 cM</td>
</tr>
<tr>
<td>DXS1207</td>
<td>Xp1.4</td>
<td>0.44</td>
<td>180-190</td>
<td>56°C</td>
<td>Genethon</td>
<td>1 cM</td>
</tr>
<tr>
<td>DXS1201</td>
<td>Xp1.4</td>
<td>0.56</td>
<td>260-290</td>
<td>56°C</td>
<td>Genethon</td>
<td>0 cM</td>
</tr>
<tr>
<td>DXS8085</td>
<td>Xp1.4</td>
<td>0.62</td>
<td>171</td>
<td>60°C</td>
<td>Genethon</td>
<td>0 cM</td>
</tr>
<tr>
<td>DXS228</td>
<td>Xp1.3</td>
<td>0.53</td>
<td>190</td>
<td>57°C</td>
<td>GBD</td>
<td>1 cM</td>
</tr>
<tr>
<td>DXS8035</td>
<td>Xp1.3</td>
<td>0.78</td>
<td>120-150</td>
<td>48°C</td>
<td>Genethon</td>
<td>1 cM</td>
</tr>
<tr>
<td>DXS7</td>
<td>Xp1.3</td>
<td>0.48</td>
<td>155-170</td>
<td>60°C</td>
<td>GBD</td>
<td>0.5 cM</td>
</tr>
<tr>
<td>MAOA</td>
<td>Xp1.3</td>
<td>0.85</td>
<td>112</td>
<td>54°C</td>
<td>GBD</td>
<td>0.5 cM</td>
</tr>
<tr>
<td>DXS8080</td>
<td>Xp1.3</td>
<td>0.68</td>
<td>175-200</td>
<td>60°C</td>
<td>Genethon</td>
<td>0.5 cM</td>
</tr>
<tr>
<td>DXS8083</td>
<td>Xp1.3-23</td>
<td>0.74</td>
<td>165-185</td>
<td>58°C</td>
<td>Genethon</td>
<td>3.5 cM</td>
</tr>
<tr>
<td>DXS1003</td>
<td>Xp1.23</td>
<td>0.80</td>
<td>170-195</td>
<td>55°C</td>
<td>Genethon</td>
<td>1.9 cM</td>
</tr>
<tr>
<td>DXS426</td>
<td>Xp1.23</td>
<td>-</td>
<td>280-300</td>
<td>56°C</td>
<td>GBD</td>
<td>2.5 cM</td>
</tr>
<tr>
<td>DXS1367</td>
<td>Xp1.23</td>
<td>0.77</td>
<td>105-135</td>
<td>60°C</td>
<td>GBD</td>
<td>0.4 cM</td>
</tr>
<tr>
<td>DXS722</td>
<td>Xp1.23</td>
<td>0.47</td>
<td>185-195</td>
<td>56°C</td>
<td>GBD</td>
<td>4.8 cM</td>
</tr>
<tr>
<td>DXS1126</td>
<td>Xp1.23</td>
<td>0.68</td>
<td>230-250</td>
<td>69°C</td>
<td>GBD</td>
<td>0.7 cM</td>
</tr>
<tr>
<td>DXS573</td>
<td>Xp1.23</td>
<td>0.72</td>
<td>135-145</td>
<td>54°C</td>
<td>GBD</td>
<td>0.45 cM</td>
</tr>
<tr>
<td>DXS988</td>
<td>Xp1.22</td>
<td>0.62</td>
<td>130-145</td>
<td>60°C</td>
<td>Genethon</td>
<td>1.5 cM</td>
</tr>
<tr>
<td>DXS1204</td>
<td>Xq1.22-21</td>
<td>0.78</td>
<td>235-250</td>
<td>55°C</td>
<td>GBD</td>
<td>0.1 cM</td>
</tr>
<tr>
<td>DXS986</td>
<td>Xq2.1</td>
<td>0.76</td>
<td>150-175</td>
<td>56°C</td>
<td>Genethon</td>
<td>12 cM</td>
</tr>
<tr>
<td>DXS990</td>
<td>Xq2.33</td>
<td>0.77</td>
<td>170-180</td>
<td>60°C</td>
<td>Genethon</td>
<td>13 cM</td>
</tr>
<tr>
<td>DXS6804</td>
<td>Xq2.3</td>
<td>0.70</td>
<td>170</td>
<td>57°C</td>
<td>GBD</td>
<td>12.3 cM</td>
</tr>
<tr>
<td>DXS1220</td>
<td>Xq2.3</td>
<td>0.71</td>
<td>190-220</td>
<td>57°C</td>
<td>Genethon</td>
<td>0.5 cM</td>
</tr>
<tr>
<td>DXS424</td>
<td>Xq2.3</td>
<td>0.79</td>
<td>180-200</td>
<td>55°C</td>
<td>GBD</td>
<td>2.7 cM</td>
</tr>
<tr>
<td>DXS1192</td>
<td>Xq2.63</td>
<td>0.85</td>
<td>120-140</td>
<td>58°C</td>
<td>Genethon</td>
<td>24 cM</td>
</tr>
<tr>
<td>DXS1123</td>
<td>Xq2.7</td>
<td>0.68</td>
<td>168-180</td>
<td>55°C</td>
<td>GBD</td>
<td>29 cM</td>
</tr>
<tr>
<td>DXS1684</td>
<td>Xq2.8</td>
<td>0.81</td>
<td>128-148</td>
<td>55°C</td>
<td>GBD</td>
<td>3 cM</td>
</tr>
</tbody>
</table>

In addition to the 52 microsatellite markers, two other non-microsatellite markers (shown in Table 3.2) were used in this study for haplotype analysis.

Table 3.2 Other markers used for genetic characterisation of XLRP and CSNB families

<table>
<thead>
<tr>
<th>Marker name</th>
<th>Chromosomal location</th>
<th>Heterozygosity</th>
<th>Size (in bp)</th>
<th>Annealing T (°C)</th>
<th>Primers sequence information</th>
<th>Genetic distances</th>
</tr>
</thead>
<tbody>
<tr>
<td>RPGR/18</td>
<td>Xp2.1</td>
<td>0.39</td>
<td>239</td>
<td>58°C</td>
<td>Meindl et al. 1996</td>
<td></td>
</tr>
<tr>
<td>OTC</td>
<td>Xp1.4</td>
<td>0.44</td>
<td>116</td>
<td>50°C</td>
<td>Petty et al. 1991</td>
<td>0.5 cM</td>
</tr>
</tbody>
</table>

Note that RPGR is located between markers M6 and DXS6679 at a genetic distance of 1.5 cM and 0.3 cM respectively, while OTC is located between markers DXS6679 and DXS1068 at a genetic distance of 0.2 cM and about 2 cM respectively.
3.2.3 Genotyping

For typing of microsatellite markers, radioactive PCR was performed as described in section 2.6.4 and the allele size differences were resolved by polyacrylamide gel electrophoresis (section 2.4.2).

RPGR/18 is a single nucleotide polymorphism (SNP) found in exon 18 of the RPGR gene (T or C at position IVS18+11). The T to C change has been observed to occur with a frequency of 27% (Miano et al. 1999), therefore accounting for a level of heterozygosity of 0.44 (see section 1.4.2). Direct sequencing (described in section 2.7) of this exon was employed to detect the two different alleles.

OTC is a biallelic restriction fragment length polymorphism (RFLP), which was detected by PCR, followed by digestion with restriction enzyme Dral, and resolution on 4 % agarose gel (3 % NuSieve: 1 % standard agarose) as described (Petty et al. 1991).

Genotypes were recorded independently from the family data, and then transferred to the family pedigree where X-linked segregation was confirmed.

3.2.4 Haplotype analysis

The segregation of marker allele was followed in each of the families to establish sites of meiotic recombination. The most likely haplotypes were inferred by minimizing the number of crossover events in each sibship.

3.2.5 Linkage analysis

To generate lod scores, allele information was processed using the program MLINK to calculate two-point likelihoods of family genotype data by stepwise variation of the recombination fraction (θ) between a marker and the disease locus (either XLRP or CSNBX). For each family two-point lod scores were calculated using four different markers, two contained within the region of segregation (as ascertained by haplotype analysis) and two outside this region, at a series of θ ranging from 0 to 0.5. The pedigrees were analysed as X-linked recessive traits, with equal allele frequencies and a disease gene frequency of 0.0001 in the general population, penetrance values for carriers were set at 0.0000.
Haplotype and linkage analysis in families with X-linked retina disease

In order to achieve a maximum lod score ($Z_{\text{max}}$) of over 3, which is considered significant for linkage, one needs a minimum of eleven meioses, one to determine the phase and ten to contribute to the lod score, although for an X-linked pedigree a lod score of 2 is considered significant for linkage. Since the majority of the families analysed do not have such number of meioses, linkage analysis was carried out only on the larger families.

3.3 Results

3.3.1 XLRP families

A summary in the form of a schematic representation of the haplotype data generated with all of the 33 XLRP families studied is shown in Table 3.3. The region of segregation with disease for each family is represented as blocked boxes. Where identified, proximal and distal crossovers are also shown (represented with the symbol $\times$), whilst uninformative markers are represented as shaded boxes and markers not typed are left blank.

All pedigrees are drawn assuming the minimum number of recombination events.

3.3.1.1 Families segregating with the RP3 locus

Pedigrees of 9 XLRP families segregating exclusively with the RP3 locus are shown in Figure 3.2(a-i), together with the marker haplotypes for sampled members and key recombination events. For the majority of these families haplotypes constructed on Xp showed linkage to this part of the chromosome, however to exclude the RP24 locus from the disease interval markers on Xq would need to be tested. The results for each family will be presented in turn:

a) Family RPIL

In this Belgian family individual II:9, an obligate carrier, is a recombinant with respect to her carrier mother I:2 (whose genotype can be inferred from her affected sons; individuals II:3 and II:11, and her unaffected son individual II:4) between markers DXS989 and DXS8090. Individual IV:7 is a recombinant with respect to his grandmother (II:9) between markers DXS556 and DXS8080. Although his mother (III:11) was not typed for DXS8080 (due to PCR problems), it is clear that individual IV:7 has inherited the DXS8080
Table 3.3 Integration of haplotype data for 33 XLRP families analysed

<table>
<thead>
<tr>
<th></th>
<th>RP3</th>
<th>RP3+RP2</th>
<th>RP3+RP23</th>
<th>RP3+RP24</th>
<th>RP3+RP2+RP24</th>
<th>RP3+RP2+RP23</th>
<th>*</th>
</tr>
</thead>
<tbody>
<tr>
<td>DXS1233</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DXS996</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DXS1223</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DXS7103</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DXS1224</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DXS7593</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>DXS7110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>DXS989</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>DXS8027</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>DXS8090</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>DXS572</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DXS1242</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>CYB3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DXS1110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DXS8349</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DXS8170</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>M6</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>RPGR/18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DXS6679</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>OTC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>DXS1068</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>DXS556</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>DXS993</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DXS977</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>DXS278</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>DXS8035</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>DXS57</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>MAOA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>DXS8081</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>DXS8083</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>DXS1003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>DXS426</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>DXS1367</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>DXS722</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>DXS1126</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>DXS573</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>DXS988</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>DXS1204</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>DXS986</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>DXS980</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>DXS804</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>DXS1220</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>DXS424</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>DXS1192</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>DXS1123</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>DXS1084</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

*Families segregating with all of the four known loci
**Figure 3.2:** 9 XLRP pedigrees segregating with the RP3 locus and associated genetic haplotypes. Blocked bars represent regions of segregation with disease. Shaded bars represent uninformative markers. Males and females are represented respectively by boxes and circles. Affected males are represented by filled boxes, carrier females by circles with dots and females of unknown clinical status by circles with question marks.

**Figure 3.2a:** Family RP1L
Haplotype and linkage analysis in families with X-linked retina disease

Figure 3.2b: Family RP106

Figure 3.2c: Family RP158
Figure 3.2f: Family RP1591

Figure 3.2g: Family RP1600
Figure 3.2h: Family RP1750

Figure 3.2i: Family RP2788
allele from his unaffected grandfather (II:10). On this basis it appears that the crossovers define the distal and proximal boundaries of the XLRP interval in this family to between markers DXS989 and DXS8080 hence spanning the RP3 locus and excluding both the RP23 locus, located distal to marker DXS989 and the RP2 locus, located proximal to marker DXS8080. According to this data the two females II:12 and IV:2 appear to be carriers of the disease gene, whilst the two females II:5 and III:5 appear to be non-carriers.

b) Family RP106

In the American family RP106, a key crossover was detected in affected male III:7 between markers DXS1242 and DXS1110, with the disease-associated haplotype located proximal to the crossover therefore excluding the RP23 locus. Individual III:8 displays another recombination event between markers DXS556 and DXS993 placing the disease interval distal to this crossover and therefore defining the minimal region of segregation with disease in this family between markers DXS1242 and DXS993 also excluding the RP2 locus, hence spanning the RP3 locus only.

c) Family RP158

In this small XLRP pedigree, it is difficult to infer the haplotype of carrier female I:2, and therefore which of her children (II:1 or II:3) displays a recombination event. Assuming however that individual II:1 has inherited the affected chromosome from her mother without crossovers, her brother (individual II:3) has undergone a recombination event between markers DXS8039 and DXS8090. (Note that by assuming the opposite situation in which individual II:1 was the recombinant with respect to her mother, the result would have been the same). Individual III:3, who has a crossover between markers DXS8080 and DXS8083, provides the proximal boundary for the disease-associated region in this family, which is contained between markers DXS8039 and DXS8083. The RP23 and RP2 loci are therefore excluded from this family and disease segregates with the RP3 locus only. Individual III:2 is hence not a carrier of the disease gene.
Haplotype and linkage analysis in families with X-linked retina disease

d) Family RP296

Two key recombinant individuals define the distal and proximal boundaries of the XLRP region in this family. The unaffected individual II:1 defines the distal boundary with a recombination event between markers DXS999 and DXS7593 therefore excluding the RP23 locus. Individual II:4 is recombinant with respect to her mother, between markers M6 and DXS556, with marker DXS1068 being uninformative, placing the disease interval distal to this crossover and excluding RP2.

e) Family RP1120

In this American family the obligate carrier female II:1 is a double recombinant with respect to her obligate carrier sister II:3. The distal crossover has occurred between markers CYBB and DXS1110 therefore excluding RP23 from the disease interval, the proximal crossover has occurred between markers DXS997 and DXS228, excluding the RP2 locus from the disease interval. The minimal region of the chromosome associated with disease in this family spans the RP3 locus only. (Note that individual II:3 could be considered a double recombinant in this family, or each of the two cousins II:1 and II:3 could be considered as single recombinants, the result would have been the same). Females of unknown status (IV:1 and IV:2) appear to have inherited the disease gene.

f) Family RP1591

The distal boundary for the disease interval in this family is defined by individual III:3, an obligate carrier female (she has an affected son, not shown in the pedigree) who displays a recombination event between markers DXS1110 and DXS8170 placing disease proximal to the RP23 locus. Although DNA was not available for her mother II:3, it is clear that she must have allele 2/3 at DXS1110 because her daughter III:3 is 2/2 and her son (III:4) is 3 at the same locus. The proximal crossover in this family occurred between markers DXS556 and DXS993 in individual II:8 with respect to her brother II:1 (if individual II:1 was the recombinant, the result would be identical). This recombination event excludes the RP2 locus from the disease interval, and disease therefore only segregates with the RP3 region. Females of unknown status II:7 and IV:2 appear to have inherited the disease gene, whilst female III:9 appears not to be a carrier and status of female III:5 remains unknown.
**g) Family RP1600**

In this Italian family individual II:4, a carrier female, has a crossover between markers DXS989 and DXS1214 therefore placing disease proximal to RP23. Her unaffected brother, individual II:3, provides the proximal boundary for disease in this family with a crossover between markers DXS1068 and DXS8080 which enables us to exclude RP2 as well as RP24 from the disease interval, since markers on Xq were also tested. This also shows that female III:2 is a carrier of disease.

**h) Family RP1750**

In this XLRP family the distal crossover is observed in individual III:5. Although his mother’s DNA was not available for this analysis, he has a recombination event with respect to his two uncles (II:1 and II:4) between markers DXS8039 and DXS8090, excluding the distal part of Xp containing the RP23 locus from the disease interval. The proximal crossover is defined by individual III:1 who has a recombination event with respect to his two uncles (II:1 and II:4) between markers DXS1068 and DXS556 placing the disease interval distal to the RP2 region, hence disease in this family segregates with the RP3 locus only. It is clear from these data that female IV:2 has inherited the disease gene.

**i) Family RP2788**

In family RP2788 DNA for the carrier female II:3 was not available for the analysis. However her two sons III:2 and III:5 share all alleles tested in common, they have therefore inherited the entire affected chromosome from their mother without crossovers. Their brother III:3 is a double recombinant. The first crossover, between markers DXS999 and DXS7110 defines the distal boundary of the disease interval in this family and excludes the RP23 locus. The second crossover occurred between markers DXS1068 and DXS556, defining the minimal region of segregation with disease, which spans the RP3 locus only.

**3.3.1.2 Families segregating with RP3 and RP2**

Due to the small genetic distance between the RP3 and the RP2 loci (estimated to be about 20-22 cM), the fact that RP2 and RP3 are clinically indistinguishable and that RP3 is the major genetic locus accounting for XLRP, it has always been difficult to genetically
define families segregating with RP2 only, and the majority of families linked to the RP2 locus also include the RP3 locus in their disease interval. In this section three XLRP pedigrees segregating with both loci (Figure 3.3a-c) are described in turn:

a) Family RP2786

In family RP2786 the distal boundary for the disease interval is defined by individual II:4, a carrier female. It is clear by comparison with her two sisters haplotypes (individuals II:1 and II:3) that she has a recombination event between markers DXS989 and DXS8039, placing disease proximal to RP23. The proximal boundary is defined by individuals III:1 and IV:1 who both have a recombination between markers DXS1367 and DXS1204, placing disease proximal to RP24, therefore spanning both the RP2 and RP3 loci. It also appears from this analysis that female II:3, whose clinical status is unknown, is possibly a carrier, whilst for female III:3 the status is still unknown.

b) Family XRP90

A double recombination event in individual III:3, an obligate carrier, with respect to her mother II:1 (again an obligate carrier), defines the minimal region of segregation with disease in this family. The phases of the two sisters (III:1 and III:3) can be inferred from their father II:2, whilst the phase of individual II:1 can be inferred from her daughter III:1. The two crossovers have occurred (i) between markers DXS7593 and DXS989, both lying proximal to the RP23 locus, and (ii) between markers DXS1204 and DXS986, both lying proximal to the RP24 locus. It is clear that the disease interval in this family spans the RP2 and RP3 loci.

c) Family RP639

In this family the distal boundary of the XLRP gene is defined by individual III:2 who has undergone a recombination event with respect to his mother (II:1) between markers M6 and DXS1068. Two other markers (DXS6679 and OTC) were tested in this family to try to exclude the RP3 region, but they were both uninformative. The proximal boundary is defined by the unaffected individual III:4. He has a recombination event between markers
**Figure 3.3**: 3 XLRP pedigrees segregating with the RP3 and RP2 loci and associated genetic haplotypes.

**Family 3.3a**: Family RP2786

**Family 3.3b**: Family XRP90
Figure 3.3c: Family RP639

[Family tree with marker alleles for each individual]
MAOA and DXS986 therefore excluding RP24 that lies distal on Xq. Disease interval in this family segregates with the RP2 locus and potentially the RP3 locus.

3.3.1.3 Families segregating with RP3 and RP23

In the next set of six families (Figure 3.4a-f) distal crossovers were not identified or disease intervals included the RP23 locus, whilst proximal crossovers were always between the RP3 and RP2 loci. Markers tested in each of these families showed segregation of disease on Xp, however to exclude RP24 from the disease interval markers on Xq need to be typed. Each pedigree is described in turn:

a) Family NZ1

In this New Zealand family, a recombination is observed in the unaffected individual IV:2 between markers OTC and DXS8035 excluding RP2 from the disease interval. No distal crossover was detected in this family; therefore disease segregates with the RP23 and RP3 loci. All females of unknown status (III:2, IV:1 and IV:3) appear to be non-carriers.

b) Family F75

In this small pedigree individual III:2 is a recombinant with respect to his mother (II:2) between markers DXS8083 and DXS1003. This crossover defines the proximal boundary of disease, excluding the RP2 locus. In order to evaluate potential segregation with the RP23 locus more distal markers need to be scored in this family. At present disease is linked to the RP3 locus and possibly the RP23 locus.

c) Family XRP91

A key recombination event between markers MAOA and DXS8083 has occurred in this family in the unaffected individual III:1 who has inherited, from his carrier mother, the affected part of the chromosome lying proximal to marker DXS8083. The RP2 locus is therefore excluded in the family. A distal crossover between markers DXS1110 and DXS228 is present in individual III:4 whose clinical status is unknown and is therefore not informative. The disease interval in this family spans the RP3 locus and potentially also the RP23 locus, however additional distal markers need to be scored. Also individual III:3 whose
Figure 3.4: 6 XLRP pedigrees segregating with the RP3 and RP23 loci and associated genetic haplotypes.

Figure 3.4a: Family NZ1

Figure 3.4b: Family F75
**Figure 3.4c:** Family XRP91

**Figure 3.4d:** Family RP759
Haplotype and linkage analysis in families with X-linked retina disease

**Figure 3.4e: Family RP2689**

**Figure 3.4f: Family RP3126**
clinical status is unknown is a potential carrier, whilst the status of individual III:4 remains unknown.

d) Family RP759

In this family from Newcastle, individual III:3, an obligate carrier, is a double recombinant with respect to her obligate carrier cousins (III:1, III:5 and III:7). The distal recombination event has occurred between markers DXS1224 and DXS7593, hence the disease interval still spans part of the RP23 region. The proximal one has occurred between markers DXS1068 and DXS556, placing disease distal to the RP2 locus. Individual IV:6 is also a recombinant with respect to her mother (III:7) between markers DXS7593 and DXS1110. This crossover would enable the exclusion of the RP23 locus from the disease region if this female were a carrier, however her clinical status is unknown at present, therefore this family appears to segregate with both the RP3 and the RP23 loci. Also female IV:1 is not a carrier of the disease gene whilst status of female IV:3 still remains undetermined.

e) Family RP2689

In this family assuming that individuals II:3, II:4 and II:5 are non-recombinant, the phase of individual I:2 can be inferred. One of her affected sons (II:2) has a crossover between markers DXS556 and DXS8035 placing disease distal to RP2. A distal crossover has not been detected in this family, which appears to be segregating potentially with both the RP3 and RP23 loci. At present females II:4 and II:5, whose clinical status is unknown appear to be carrier and non-carrier of the disease gene respectively.

f) Family RP3126

Two recombination events have occurred in this family in individuals IV:2 and IV:4 with regard to their respective mothers III:2 and III:4. The first crossover between markers DXS1224 and DXS7593 defines the distal boundary of disease, which includes part of the RP23 region. The proximal boundary is defined by the second crossover, which occurred between markers DXS993 and DXS8083, excluding RP2. Individual IV:6, a carrier female, has inherited a more distal crossover between markers DXS556 and DXS993, which still
does not exclude the RP3 locus. It also appears from this analysis that individual IV:3, whose clinical status is unknown, is a carrier.

3.3.1.4 Families segregating with RP2 and RP24

In the next three families described (Figure 3.5a-c) it was possible to exclude genetically the RP3 locus, however proximal crossovers were not identified therefore, for these families, the disease interval potentially spans the RP24 locus. The results of the haplotype analysis for each family are described in turn:

a) Family RP107

In this American family the unaffected individual IV:3 is a recombinant with respect to his mother (individual III:4) between markers DXS8035 and DXS1367 placing disease proximal to the RP23 and RP3 loci. It is important to note that although the clinical status of individual III:4 is unknown it appears she has inherited, from her obligate carrier mother (individual II:1), all her affected chromosome with the exception of the most distal marker tested (DXS996). It is therefore highly likely that she is a carrier. A second recombination event has occurred in individual IV:2 between markers DXS986 and DXS1684, this crossover is on Xq and does not exclude RP24 in this family, which is therefore linked to the RP2 and RP24 loci.

b) Family RP3877

A key recombination event has occurred in this family in individual II:3 with respect to his mother (I:1) between markers DXS8035 and DXS8083. This enables disease to be placed proximal to marker DXS8035, hence proximal to RP3. Additional markers on Xq need to be tested to evaluate the possible involvement of the RP24 locus in disease for this family. At present the disease interval spans the RP2 locus and potentially RP24. Female of unknown clinical status (III:1) appears to be a non-carrier of the disease gene.
Figure 3.5: 3 XLRP pedigrees segregating with the RP2 and RP24 loci and associated genetic haplotypes.

Figure 3.5a: Family RP107

Figure 3.5b: Family RP3877

Figure 3.5c: Family RP3973
Haplotype and linkage analysis in families with X-linked retina disease

c) Family RP3973

In this small pedigree a crossover event in individual III:3 between markers DXS8080 and DXS573 places disease proximal to the RP3 locus. A distal crossover on Xq has not been identified and at present RP2 and potentially RP24 are segregating with XLRP in this family.

3.3.1.5 Families segregating with RP3, RP2 and RP24

In this section seven pedigrees are described (Figure 3.6a-g) for which distal crossovers enabled exclusion of the RP23 locus from the disease interval. Each of the seven families is presented in turn:

a) Family RPB1

In this Belgian family individual III:3 is a recombinant with respect to his mother II:2 between markers DXS7593 and DXS989. This crossover defines the distal breakpoint of the disease region in this family excluding the RP23 locus. On the other hand all of the other three loci (RP3, RP2 and RP24) are included as potential XLRP genes in this family as no distal crossovers on Xq have been identified. The clinical status of females III:1 and III:2 remains undetermined. Female III:1 is a potential carrier only in the unlikely event that the disease gene in the family is RP24, and female III:2 is a carrier only if the disease gene is either RP24 or RP2.

b) Family MGRP4

In this American family a key recombination event between markers DXS1214 and DXS1068 has occurred in individuals V:1 (with respect to his mother IV:4), and in individual IV:1 (with respect to her distant cousins in the same generation IV:2, IV:3 and IV:4). This distal crossover excludes the RP23 locus from the disease interval, however RP3 lies distal to DXS1068 (see table 3.3). Markers OTC, DXS6679 and M6 should be typed in this family in order to exclude or include the RP3 locus. This was not performed prior to mutation screening (section 4.3.2); therefore in this family the XLRP region potentially spans the RP3 locus. The proximal boundary is defined by individual IV:3. He is a recombinant with respect to his mother (III:3) between markers DXS424 and DXS1123 both located on Xq; since the RP24 region lies between these two markers, it can not be excluded from the disease interval.
Figure 3.6: 7 XLRP pedigrees segregating with the RP3, RP2 and RP24 loci and associated genetic haplotypes.

**Figure 3.6a:** Family RPBl

**Figure 3.6b:** Family MGRP4
Figure 3.6c: Family RP4

Figure 3.6d: Family RP34

Figure 3.6e: Family RP76
c) Family RP4

In this small pedigree a recombination event in individual II:4 between markers DXS572 and DXS556 places disease proximal to RP23. DXS556 is just proximal to the RP3 region whilst DXS1110, which is distal to RP3, was uninformative in the family. Disease is therefore linked to the RP2 locus and potentially the RP3 and RP24 loci, although additional markers need to be scored to evaluate disease segregation on Xq. Female II:1 whose clinical status is unknown appears to be a potential non-carrier although she too was uninformative for DXS1110.

d) Family RP34

A recombination event has occurred in this family in individual III:4 with respect to his cousin III:1 whose phase can be inferred from her son (IV:2). Note that the opposite situation can be assumed, in which individual III:1 is recombinant with respect to her cousin III:4. In any case the crossover has occurred between markers DXS989 and DXS8090 and places disease proximal to RP23, still spanning the RP3 and RP2 loci. No recombination event has been identified that would ascertain segregation or non-segregation of disease with RP24 which is therefore only potentially linked in this family. All females of unknown clinical status (III:5, IV:1 and IV:3) appear to be non-carriers of the disease gene.

e) Family RP76

In this very small family, implying the minimum recombination events, allele 1 of marker DXS1214 comes from the affected chromosome of carrier female I:2 and allele 2 at the same locus comes from the unaffected chromosome. Individual II:2 is therefore recombinant between markers DXS1214 and RPGR/18 and has inherited, from his mother, the unaffected allele (2) of marker DXS1214. This suggests exclusion of RP23 from the XLRP region and association of RP3, RP2 and RP24 with disease since no crossovers on distal Xq have been identified.

f) Family RP1571

Assuming that the phase of the affected chromosome in this family is determined by individuals III:5 and III:6, respectively affected and unaffected, a key recombination event is
observed in the affected individuals III:7, IV:1 and IV:2 between markers DXS989 and DXS8090, placing disease proximal to the RP23 locus. RP3 is linked to disease in this family whilst RP2 and RP24 are only potentially so, since the proximal two markers tested (DXS556 and DXS8080) were uninformative in the family and markers on Xq remains to be tested. It also appears from this analysis that both females of unknown status (III:2 and IV:3) are non-carriers of the disease gene.

g) Family RP2063
In family RP2063 both siblings III:4 and III:5 are recombinant with respect to their mother II:3 between markers M6 and RPGR/18 excluding RP23 from the disease interval. The markers tested, DXS573 and DXS986, were both uninformative and more distal markers on Xq need to be tested in order to exclude RP24 from the XLRP interval. At present disease in this family is linked to RP3, RP2 and potentially the RP24 locus.

3.3.1.6 Families segregating with RP3, RP2 and RP23
In this last set of XLRP families three pedigrees (Figure 3.7a-c) for which distal crossovers on Xq excluded RP24 as the potential candidate disease region are described in turn:

a) Family RPO1
In this family from Oxford the obligate carrier female III:3 is recombinant with respect to her cousin III:1 between markers DXS996 and DXS8090, placing disease proximal to DXS996. Since DXS996 is distal to the RP23 locus and DXS8090 is proximal to RP23, additional markers in between these loci need to be tested to evaluate segregation or non-segregation with this locus. At present RP23 potentially segregates with disease in this family. A second key recombination event has occurred in the unaffected male IV:2 between markers DXS8080 and DXS573 placing disease distal to DXS573 therefore spanning the RP2 and RP3 loci.
Figure 3.7: 3 XLRP pedigrees segregating with the RP3 RP2 and RP23 loci and associated genetic haplotypes.

Figure 3.7a: Family RPO1

Figure 3.7b: Family F53

Figure 3.7c: Family RP87
b) Family F53

In family F53 the distal breakpoint of the disease region is defined by individual II:6, a carrier female who is a recombinant with respect to her brothers II:1 and II:2 and to her sister II:5. The crossover is between markers DXS7103 and DXS989 still contained within the RP23 locus. The proximal boundary is defined by individual II:1. Since DNA from carrier female I:2 is not available and all the siblings in 2nd generation share the same allele (1) at the DXS986 marker, it is impossible to determine their mother’s haplotype at the same locus, which indeed seems to be homozygote, and is therefore uninformative in this family. However the crossover in individual II:1 occurred somewhere between markers DXS1204 and DXS1192 which is sufficient to exclude the RP24 locus from the XLRP region and to include RP3 and RP2.

c) Family RP87

The key recombination event in this XLRP family occurred in individual VI:2 with respect to his mother V:4. The crossover lies between markers DXS1003 and DXS426 and places disease distal to DXS426 hence spanning the RP2 and RP3 loci. Since no markers have been scored in the distal part of Xp, RP23 is also potentially considered a candidate disease locus for this family. Female VI:1 appears to be a potential carrier of disease.

3.3.1.7 Families segregating with RP3, RP2, RP24 and RP23

These last two XLRP families described (Figure 3.8a-b) have crossovers on Xp and on Xq distal to the RP23 and RP24 loci respectively or within the disease intervals, hence not excluding these loci. The two families are presented in turn:

a) Family RP2L

In this Belgian family, a recombination event has occurred in individual IV:3 with respect to his distant cousins in the 5th generation. The crossover has occurred between markers DXS1223 and DXS989 still potentially including the RP23 locus in the disease interval. This family was received recently; therefore additional markers need to be tested to identify the minimal region of segregation with disease. At present the XLRP region is linked
Figure 3.8: 2 XLRP pedigrees segregating with the RP3, RP2, RP24 and RP23 loci and associated haplotypes

Figure 3.8a: Family RP2L

Figure 3.8b: Family RP180
to RP3 and potentially RP23, RP2 and RP24. Females of unknown status (VI:1 and VI:4) appear to be carrier and non-carrier of the disease gene respectively.

b) Family RP180

In family RP180 the only recombinant is individual III:2. In fact all females in the 2nd generation are obligate carriers that could not possibly have inherited a recombination event from their affected father. The distal crossover on Xp in the affected male in the 3rd generation occurred between markers DXS996 and DXS7593 with marker DXS999 being uninformative; therefore RP23 cannot be excluded as a potential XLRP gene in this family. RP3 and RP2 both segregate with disease and potentially also RP24 since the distal crossover on Xq, observed again in the same affected male, is between markers DXS1220 and DXS1684, which is still not sufficient to exclude RP24.

3.3.1.8 Linkage analysis

Linkage analysis was carried out only on the largest families (with a minimum of 10 meioses): RP1L, RP106, RP1120, RP1591 and RP3126. Table 3.4(a-e) shows the results of this analysis for the five families, which are presented here in turn:

a) Family RP1L

Family RP1L (Table 3.4a and Figure 3.2a) is the largest XLRP family collected in the laboratory containing 23 meioses. The maximum lod score (Zmax=4.9; θmax=0) was generated with marker DXS556, which, by haplotype analysis, appeared to be linked to disease and to be informative in all the females, tested. Marker DXS8090 also gave a significant two-point lod score (Zmax=4.6; θmax=0) although lower than the one generated with DXS556 due to the fact that female III:11 was uninformative for this marker therefore it is not possible to determine whether her son’s allele derives from her affected or from her unaffected chromosome. The genetic distance between DXS556 and RPGR is estimated to be 8 cM; therefore disease in this family is linked to the RP3 locus.
Haplotype and linkage analysis in families with X-linked retina disease

Table 3.4 Two-point lod scores for linkage between XLRP and four different markers for each of five XLRP families: (a) RPIL, (b) RP106, (c) RP1120, (d) RP1591, (e) RP3126

<table>
<thead>
<tr>
<th>θ values</th>
<th>DXS989</th>
<th>DXS8090</th>
<th>DXS556</th>
<th>DXS8080</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-∞</td>
<td>4.6</td>
<td>4.9</td>
<td>-∞</td>
</tr>
<tr>
<td>0.05</td>
<td>-0.58</td>
<td>4.2</td>
<td>4.5</td>
<td>1.2</td>
</tr>
<tr>
<td>0.1</td>
<td>0.12</td>
<td>3.8</td>
<td>4.1</td>
<td>1.3</td>
</tr>
<tr>
<td>0.15</td>
<td>0.43</td>
<td>3.4</td>
<td>3.7</td>
<td>1.25</td>
</tr>
<tr>
<td>0.2</td>
<td>0.58</td>
<td>3.2</td>
<td>3.2</td>
<td>1.1</td>
</tr>
<tr>
<td>0.25</td>
<td>0.62</td>
<td>2.5</td>
<td>2.7</td>
<td>1</td>
</tr>
<tr>
<td>0.3</td>
<td>0.6</td>
<td>2.06</td>
<td>2.2</td>
<td>0.84</td>
</tr>
<tr>
<td>0.35</td>
<td>0.52</td>
<td>1.5</td>
<td>1.7</td>
<td>0.66</td>
</tr>
<tr>
<td>0.4</td>
<td>0.39</td>
<td>1.01</td>
<td>1.1</td>
<td>0.45</td>
</tr>
<tr>
<td>0.45</td>
<td>0.22</td>
<td>0.5</td>
<td>0.5</td>
<td>0.23</td>
</tr>
<tr>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

b)

<table>
<thead>
<tr>
<th>θ values</th>
<th>DXS1242</th>
<th>DXS1110</th>
<th>DXS556</th>
<th>DXS993</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-∞</td>
<td>2.29</td>
<td>2.33</td>
<td>-∞</td>
</tr>
<tr>
<td>0.05</td>
<td>0.67</td>
<td>2.08</td>
<td>2.12</td>
<td>0.84</td>
</tr>
<tr>
<td>0.1</td>
<td>0.79</td>
<td>1.87</td>
<td>1.9</td>
<td>0.93</td>
</tr>
<tr>
<td>0.15</td>
<td>0.78</td>
<td>1.64</td>
<td>1.67</td>
<td>0.88</td>
</tr>
<tr>
<td>0.2</td>
<td>0.72</td>
<td>1.4</td>
<td>1.42</td>
<td>0.78</td>
</tr>
<tr>
<td>0.25</td>
<td>0.61</td>
<td>1.5</td>
<td>1.17</td>
<td>0.64</td>
</tr>
<tr>
<td>0.3</td>
<td>0.48</td>
<td>0.88</td>
<td>0.9</td>
<td>0.48</td>
</tr>
<tr>
<td>0.35</td>
<td>0.33</td>
<td>0.62</td>
<td>0.62</td>
<td>0.32</td>
</tr>
<tr>
<td>0.4</td>
<td>0.18</td>
<td>0.34</td>
<td>0.34</td>
<td>0.17</td>
</tr>
<tr>
<td>0.45</td>
<td>0.05</td>
<td>0.1</td>
<td>0.11</td>
<td>0.05</td>
</tr>
<tr>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
### RP1120

<table>
<thead>
<tr>
<th>θ values</th>
<th>CYBB</th>
<th>DXS1110</th>
<th>DXS556</th>
<th>DXS228</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-∞</td>
<td>1.47</td>
<td>1.51</td>
<td>-∞</td>
</tr>
<tr>
<td>0.05</td>
<td>-0.05</td>
<td>1.32</td>
<td>1.36</td>
<td>0.9</td>
</tr>
<tr>
<td>0.1</td>
<td>0.39</td>
<td>1.16</td>
<td>1.2</td>
<td>0.44</td>
</tr>
<tr>
<td>0.15</td>
<td>0.57</td>
<td>1</td>
<td>1.03</td>
<td>0.61</td>
</tr>
<tr>
<td>0.2</td>
<td>0.63</td>
<td>0.83</td>
<td>0.86</td>
<td>0.66</td>
</tr>
<tr>
<td>0.25</td>
<td>0.63</td>
<td>0.66</td>
<td>0.68</td>
<td>0.65</td>
</tr>
<tr>
<td>0.3</td>
<td>0.58</td>
<td>0.49</td>
<td>0.51</td>
<td>0.59</td>
</tr>
<tr>
<td>0.35</td>
<td>0.49</td>
<td>0.32</td>
<td>0.33</td>
<td>0.49</td>
</tr>
<tr>
<td>0.4</td>
<td>0.36</td>
<td>0.17</td>
<td>0.17</td>
<td>0.36</td>
</tr>
<tr>
<td>0.45</td>
<td>0.19</td>
<td>0.05</td>
<td>0.05</td>
<td>0.19</td>
</tr>
<tr>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

### RP1591

<table>
<thead>
<tr>
<th>θ values</th>
<th>DXS8090</th>
<th>DXS1110</th>
<th>DXS556</th>
<th>DXS993</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.04</td>
<td>0.78</td>
<td>0.7</td>
<td>-∞</td>
</tr>
<tr>
<td>0.05</td>
<td>0.95</td>
<td>0.7</td>
<td>0.63</td>
<td>-1.16</td>
</tr>
<tr>
<td>0.1</td>
<td>0.85</td>
<td>0.63</td>
<td>0.56</td>
<td>-0.63</td>
</tr>
<tr>
<td>0.15</td>
<td>0.74</td>
<td>0.55</td>
<td>0.49</td>
<td>-0.35</td>
</tr>
<tr>
<td>0.2</td>
<td>0.64</td>
<td>0.47</td>
<td>0.41</td>
<td>-0.18</td>
</tr>
<tr>
<td>0.25</td>
<td>0.54</td>
<td>0.39</td>
<td>0.34</td>
<td>-0.07</td>
</tr>
<tr>
<td>0.3</td>
<td>0.43</td>
<td>0.32</td>
<td>0.27</td>
<td>-0.005</td>
</tr>
<tr>
<td>0.35</td>
<td>0.32</td>
<td>0.24</td>
<td>0.2</td>
<td>0.03</td>
</tr>
<tr>
<td>0.4</td>
<td>0.22</td>
<td>0.16</td>
<td>0.13</td>
<td>0.04</td>
</tr>
<tr>
<td>0.45</td>
<td>0.11</td>
<td>0.08</td>
<td>0.06</td>
<td>0.03</td>
</tr>
<tr>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

### RP3126

<table>
<thead>
<tr>
<th>θ values</th>
<th>DXS1224</th>
<th>DXS7593</th>
<th>DXS993</th>
<th>DXS8083</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-∞</td>
<td>1.4</td>
<td>1.44</td>
<td>-∞</td>
</tr>
<tr>
<td>0.05</td>
<td>-1.29</td>
<td>1.28</td>
<td>1.31</td>
<td>-1.25</td>
</tr>
<tr>
<td>0.1</td>
<td>-0.77</td>
<td>1.16</td>
<td>1.18</td>
<td>-0.73</td>
</tr>
<tr>
<td>0.15</td>
<td>-0.49</td>
<td>1.03</td>
<td>1.05</td>
<td>-0.46</td>
</tr>
<tr>
<td>0.2</td>
<td>-0.31</td>
<td>0.9</td>
<td>0.91</td>
<td>-0.29</td>
</tr>
<tr>
<td>0.25</td>
<td>-0.2</td>
<td>0.76</td>
<td>0.77</td>
<td>-0.18</td>
</tr>
<tr>
<td>0.3</td>
<td>-0.12</td>
<td>0.62</td>
<td>0.63</td>
<td>-0.11</td>
</tr>
<tr>
<td>0.35</td>
<td>-0.06</td>
<td>0.48</td>
<td>0.48</td>
<td>-0.05</td>
</tr>
<tr>
<td>0.4</td>
<td>-0.02</td>
<td>0.33</td>
<td>0.33</td>
<td>-0.02</td>
</tr>
<tr>
<td>0.45</td>
<td>-0.006</td>
<td>0.17</td>
<td>0.17</td>
<td>-0.005</td>
</tr>
<tr>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Haplotype and linkage analysis in families with X-linked retina disease

b) Family RP106

The maximum two-point lod scores generated for the second family presented in this section, RP106 (Table 3.4b and Figure 3.2b), demonstrate how the lod score depends largely on the number of meioses present in the family under study. In family RP106 there are 13 meioses, although this number is sufficient to produce a significant lod score and although a significant lod score was indeed generated with both markers associated with disease in this family, DXS1110 (Zmax=2.29; θmax=0) and DXS556 (Zmax=2.33; θmax=0), these values are significantly lower than the ones generated for family RP1L, where 23 meioses are present. Also both markers were uninformative in female III:3, reducing the lod score. However both Zmax values are over 2, which is considered significant for linkage in an X-linked pedigree. Given the fact that DXS556 is approximately 8 cM from RPGR and DXS1110 is approximately 1 cM from RPGR, with the two markers lying on either side of the gene, it is clear that disease is linked to RP3 in this family.

c) Family RP1120

Although there are 13 meioses in family RP1120 (Table 3.4c and Figure 3.2e), the two-point lod scores generated with the two markers that from haplotype analysis appear to be linked with disease (DXS1110 and DXS556), are below the threshold value for significance of 2. This is due to the fact that both markers were uninformative in individuals III:7 and III:8 and because the haplotype of female III:1 cannot be deduced from her children (who apparently have inherited all the same alleles from those two markers). Therefore part of the information equal to 5 meioses (three meioses from individual III:1 to her children IV:1, IV:2 and IV:3 and two meioses from individual III:8 to her two daughters IV:5 and IV:6) is lost. More over, by looking at the maximum lod scores obtained with the two markers there is a difference (Zmax=1.47; θmax=0 for DXS1110 vs. Zmax=1.51; θmax=0 for DXS556) even though the informativeness of the two markers is the same in all individuals. This is due to the fact that DXS556, which gives a higher Zmax, has a higher level of heterogeneity (3 alleles vs. 2 alleles for marker DXS1110). The lod scores obtained in this pedigree are not significant and therefore linkage to the RP3 locus is not statistically proved.
Haplotype and linkage analysis in families with X-linked retina disease

d) Family RP1591

In this pedigree (Table 3.4d and Figure 3.2f), again both markers tested for linkage that appeared to be linked with disease by haplotype analysis (DXS1110 and DXS556) did not give significant maximum lod scores even with 15 meioses in the family ($Z_{\text{max}}=0.78$; $\theta_{\text{max}}=0$ and $Z_{\text{max}}=0.7$; $\theta_{\text{max}}=0$ respectively). This is due partly to individual III:1 for which marker DXS1110 was not scored and marker DXS556 was uninformative, therefore information relative to two meioses (from this individual to her two children IV:1 and IV:2) is lost. Moreover, individuals IV:3 and IV:4 are carriers because they have been clinically evaluated and they show signs of disease (i.e. altered ERG), however they are not obligate carriers and the linkage program is not able to incorporate this information. Linkage to the RP3 locus is therefore not proved in this family.

e) Family RP3126

In family RP3126 (Table 3.4e and Figure 3.4f) there are 12 meioses but both females II:1 and II:3 are uninformative for the markers associated with disease according to the haplotype data (DXS7593 and DXS993). For 5 meioses (two from individual I:2 to her daughters II:1 and II:3, two from individual II:1 to her daughters III:1 and III:2 and one from individual II:3 to her daughter III:4), segregation information is lost. Therefore, although the number of meioses in this family is potentially sufficient to generate a significant lod score, both markers give a maximum value that is below the minimum level required for a marker to be significantly linked to disease ($Z_{\text{max}}=1.4$; $\theta_{\text{max}}=0$ for DXS7593 and $Z_{\text{max}}=1.44$; $\theta_{\text{max}}=0$ for DXS993). No statistical proof of linkage to the RP3 locus could therefore be obtained in this family.

3.3.2 CSNBX families

Table 3.5 shows the region of segregation with disease (represented as blocked boxes) for each of the four X-linked CNSB (CSNBX) families studied. Where identified, proximal or distal crossovers are also shown (represented with the symbol $X$), whilst uninformative markers are represented as shaded boxes.

During this section I will be referring to CSNB1 and CSNB2 loci rather than the NYX and the CACNA1F genes, as at the outset of this study the genes had not been cloned.

131
Table 3.5 Integration of haplotype data for four X-linked CSNB families analysed

<table>
<thead>
<tr>
<th></th>
<th>RP2124</th>
<th>RP3810</th>
<th>GC4180</th>
<th>RP4914</th>
</tr>
</thead>
<tbody>
<tr>
<td>DXS556</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DXS574</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DXS993</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DXS8012</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DXS1207</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DXS1201</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DXS8085</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DXS228</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DXS7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAOA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DXS8080</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DXS426</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DXS573</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NYX → CSNB1

CSNB2 ← CACNA1F
Each of the four CSNBX pedigrees (Figure 3.9a-d), which were presented assuming the minimum number of recombination events, will be discussed in turn:

3.3.2.1 Haplotype analysis for four CSNBX families

a) Family RP2124

In this large pedigree a recombination event was detected in individual V:1 with respect to his affected brothers (V:2, V:3 and V:5) between markers DXS574 and DXS993 placing disease proximal to DXS574. Individual VI:3 is also a recombinant with respect to his mother (V:9) between markers DXS574, which lies distal to CSNBl, and MAOA, which is proximal to CSNBl. However both markers DXS993 and DXS7 were uninformative in the carrier mother therefore it is not possible to exclude or include CSNBl from the disease interval in this family, which appears to segregate with both CSNBX loci. From this analysis it is also clear that female V:4 has inherited the disease gene.

b) Family RP3810

No recombination events were detected in family RP3810 using markers spanning the disease critical intervals. The two affected individuals (IV:1 and IV:4) have inherited the same haplotype which is different from that shared by the two unaffected individuals (IV:2 and IV:3). In this family therefore the disease interval spans both the CSNBl and CSNB2 loci.

c) Family GC4180

Although this is a very small pedigree, it was possible to detect a recombination event in individual II:2 with respect to his mother I:1. The crossover has occurred somewhere between markers DXS8085 and DXS426 with DXS228 and DXS8080 being uninformative, placing disease distal to the CSNB2 locus. Note that it is not possible to determine the phase of individual I:1 who could be either 1/2 or 2/1 at both markers DXS426 and DXS573. Depending on the phase of these markers either one or the other of her two sons might have inherited the crossover that places disease distal to DXS426.
Figure 3.9: 4 CSNBX pedigrees and associated genetic haplotypes.

Figure 3.9a: Family RP2124
Figure 3.9b: Family RP3810

Figure 3.9c: Family GC4180

Figure 3.9d: Family RP4194

Haplotype and linkage analysis in families with X-linked retinopathy.
d) Family RP4914

This is another small pedigree for which no crossovers were observed. It is clear from haplotype data that both affected males (II:3 and III:1) have inherited the same haplotype from their respective mothers. It is therefore not possible to discriminate whether disease in this family is linked to CSNB1 or CSNB2.

3.3.2.2 Linkage analysis

From this group of X-linked CSNB families linkage analysis was performed on family RP2124 where 12 meioses are present. This analysis was carried out with 4 different markers shown in Table 3.6.

From the haplotype data (Figure 3.9a) it appeared that markers MAOA and DXS426 were segregating with disease in this family, whilst DXS7 was uninformative and DXS574 was not segregating. MAOA did indeed give a significant two-point lod score (Table 3.6, Zmax=2.81; θmax=0), whilst the value for DXS426 is below the threshold of significance (Zmax=1.81; θmax=0) due to the fact that individual IV:5 is homozygote at this locus therefore the information relevant to two meioses (from individual IV:5 to her children V:5 and V:6) is lost. In addition, individual VI:1 was not scored for this marker therefore information from this meiosis (from his mother V:9 to him) is also lost. MAOA lies between the CSNB1 and the CSNB2 loci at a genetic distance from both loci of less than 10 cM. It is therefore not possible in this family to define linkage to one or the other of the two loci.

Table 3.6 Two-point lod scores for linkage between X-linked CSNB and four different markers for family RP2124

<table>
<thead>
<tr>
<th>θ values</th>
<th>DXS574</th>
<th>DXS7</th>
<th>MAOA</th>
<th>DXS426</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-∞</td>
<td>0.61</td>
<td>2.81</td>
<td>1.81</td>
</tr>
<tr>
<td>0.05</td>
<td>-0.49</td>
<td>0.57</td>
<td>2.56</td>
<td>1.62</td>
</tr>
<tr>
<td>0.1</td>
<td>-0.06</td>
<td>0.53</td>
<td>2.29</td>
<td>1.43</td>
</tr>
<tr>
<td>0.15</td>
<td>0.11</td>
<td>0.48</td>
<td>2.02</td>
<td>1.24</td>
</tr>
<tr>
<td>0.2</td>
<td>0.19</td>
<td>0.42</td>
<td>1.73</td>
<td>1.04</td>
</tr>
<tr>
<td>0.25</td>
<td>0.21</td>
<td>0.35</td>
<td>1.43</td>
<td>0.83</td>
</tr>
<tr>
<td>0.3</td>
<td>0.19</td>
<td>0.29</td>
<td>1.12</td>
<td>0.63</td>
</tr>
<tr>
<td>0.35</td>
<td>0.15</td>
<td>0.22</td>
<td>0.81</td>
<td>0.44</td>
</tr>
<tr>
<td>0.4</td>
<td>0.11</td>
<td>0.15</td>
<td>0.51</td>
<td>0.27</td>
</tr>
<tr>
<td>0.45</td>
<td>0.05</td>
<td>0.07</td>
<td>0.23</td>
<td>0.12</td>
</tr>
<tr>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Family</td>
<td>Diagnosis</td>
<td>Disease Locus</td>
<td>Mutation Screen*</td>
<td>Disease Gene</td>
</tr>
<tr>
<td>------------</td>
<td>-----------------------------------------------------</td>
<td>---------------------</td>
<td>------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>RP4462</td>
<td>X-linked RP with associated ciliary abnormalities</td>
<td>Spans RP2 and RP3</td>
<td>RP2</td>
<td>RPGR</td>
</tr>
<tr>
<td>MGRP3</td>
<td>XLRP</td>
<td>RP23</td>
<td>XLR3S1, PPEF, GLRA2, CLCN4, TBL1, APXL, CALB3, RAI2</td>
<td>CACNA1F</td>
</tr>
<tr>
<td>GC1006</td>
<td>CSNBX</td>
<td>RP2, RPGR, CACNA1F</td>
<td>CACNA1F</td>
<td></td>
</tr>
<tr>
<td>RP3532</td>
<td>CSNBX</td>
<td>CSNB1 and CSNB2</td>
<td>CACNA1F</td>
<td>CACNA1F</td>
</tr>
<tr>
<td>AD101</td>
<td>CSNBX</td>
<td>CACNA1F</td>
<td>CACNA1F</td>
<td></td>
</tr>
<tr>
<td>RP3810</td>
<td>CSNBX</td>
<td>CSNB1 and CSNB2</td>
<td>RP2, RPGR, CACNA1F, NYX</td>
<td></td>
</tr>
<tr>
<td>RP4611</td>
<td>CSNBX</td>
<td>RP2, RPGR, CACNA1F</td>
<td>NYX</td>
<td></td>
</tr>
<tr>
<td>RP1002</td>
<td>CSNBX</td>
<td>RP2, RPGR, CACNA1F</td>
<td>NYX</td>
<td></td>
</tr>
<tr>
<td>CDT</td>
<td>CSNBX</td>
<td>CACNA1F, NYX</td>
<td>NYX</td>
<td></td>
</tr>
<tr>
<td>RP3665</td>
<td>CSNBX</td>
<td>RP2, RPGR, CACNA1F</td>
<td>NYX</td>
<td></td>
</tr>
<tr>
<td>CCD</td>
<td>CSNBX</td>
<td>CACNA1F, NYX</td>
<td>NYX</td>
<td></td>
</tr>
<tr>
<td>CRP</td>
<td>CSNBX</td>
<td>CACNA1F, NYX</td>
<td>NYX</td>
<td></td>
</tr>
<tr>
<td>RP2427</td>
<td>CSNBX</td>
<td>CSNB2</td>
<td>CACNA1F</td>
<td>NYX</td>
</tr>
<tr>
<td>RP4126</td>
<td>CSNBX</td>
<td>RP2, RPGR, CACNA1F</td>
<td>NYX</td>
<td></td>
</tr>
<tr>
<td>RP2124</td>
<td>CSNBX</td>
<td>CSNB2</td>
<td>NYX</td>
<td>NYX</td>
</tr>
<tr>
<td>RP4914</td>
<td>CSNBX</td>
<td>CSNB1 and CSNB2</td>
<td>CACNA1F</td>
<td>NYX</td>
</tr>
<tr>
<td>RP3623</td>
<td>CSNBX</td>
<td>RP2, RPGR</td>
<td>NYX</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CACNA1F</td>
<td>NYX</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
<td>---------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>RP4121</td>
<td>CSNBX</td>
<td>RP2</td>
<td>NYX</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RPGR</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CACNA1F</td>
<td>NYX</td>
<td></td>
</tr>
<tr>
<td>GC4180</td>
<td>CSNBX</td>
<td>CSNB1 and CSNB2</td>
<td>RP2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RP3</td>
<td>RPGR</td>
<td></td>
</tr>
<tr>
<td>RP1L</td>
<td>XLRP</td>
<td>RP2 and RP3</td>
<td>RP2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RP3</td>
<td>RPGR</td>
<td></td>
</tr>
<tr>
<td>RRP</td>
<td>XLRP</td>
<td>RP3</td>
<td>RPGR</td>
<td></td>
</tr>
<tr>
<td>BRP</td>
<td>XLRP</td>
<td>RP2 and RP3</td>
<td>RP2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RP3</td>
<td>RPGR</td>
<td></td>
</tr>
<tr>
<td>RP51</td>
<td>XLRP</td>
<td>RP3</td>
<td>RPGR</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RP2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RP66</td>
<td>XLRP</td>
<td>RP3</td>
<td>RPGR</td>
<td></td>
</tr>
<tr>
<td>RP147</td>
<td>XLRP</td>
<td>RP3</td>
<td>RPGR</td>
<td></td>
</tr>
<tr>
<td>RP380</td>
<td>XLRP</td>
<td>RP3</td>
<td>RPGR</td>
<td></td>
</tr>
<tr>
<td>RP509</td>
<td>XLRP</td>
<td>RP2 and RP3</td>
<td>RP2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RP3</td>
<td>RPGR</td>
<td></td>
</tr>
<tr>
<td>RP1206</td>
<td>XLRP</td>
<td>RP2 and RP3</td>
<td>RP2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RP3</td>
<td>RPGR</td>
<td></td>
</tr>
<tr>
<td>RP1484</td>
<td>XLRP</td>
<td>RP3 and RP23</td>
<td>RP2</td>
<td></td>
</tr>
<tr>
<td>F51</td>
<td>XLRP</td>
<td>RP3</td>
<td>RPGR</td>
<td></td>
</tr>
<tr>
<td>RP106</td>
<td>XLRP</td>
<td>RP3</td>
<td>RPGR</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RP3</td>
<td>RPGR</td>
<td></td>
</tr>
<tr>
<td>RP158</td>
<td>XLRP</td>
<td>RP3</td>
<td>RPGR</td>
<td></td>
</tr>
<tr>
<td>RP296</td>
<td>XLRP</td>
<td>RP3</td>
<td>RPGR</td>
<td></td>
</tr>
<tr>
<td>RP1120</td>
<td>XLRP</td>
<td>RP3</td>
<td>RPGR</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RP2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RP1591</td>
<td>XLRP</td>
<td>RP3</td>
<td>RPGR</td>
<td></td>
</tr>
<tr>
<td>RP1600</td>
<td>XLRP</td>
<td>RP3</td>
<td>RPGR</td>
<td></td>
</tr>
<tr>
<td>RP1750</td>
<td>XLRP</td>
<td>RP3</td>
<td>RPGR</td>
<td></td>
</tr>
<tr>
<td>RP2788</td>
<td>XLRP</td>
<td>RP3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RP2786</td>
<td>XLRP</td>
<td>RP2 and RP3</td>
<td>RP2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RP3</td>
<td>RPGR</td>
<td></td>
</tr>
<tr>
<td>XRP90</td>
<td>XLRP</td>
<td>RP2 and RP3</td>
<td>RP2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RP3</td>
<td>RPGR</td>
<td></td>
</tr>
<tr>
<td>RP639</td>
<td>XLRP</td>
<td>RP2 and RP3</td>
<td>RP2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RP3</td>
<td>RPGR</td>
<td></td>
</tr>
<tr>
<td>NZ1</td>
<td>XLRP</td>
<td>RP3 and RP23</td>
<td>RP2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RP3</td>
<td>RPGR</td>
<td></td>
</tr>
<tr>
<td>F75</td>
<td>XLRP</td>
<td>RP3 and RP23</td>
<td>RP2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RP3</td>
<td>RPGR</td>
<td></td>
</tr>
<tr>
<td>XRP91</td>
<td>XLRP</td>
<td>RP3 and RP23</td>
<td>RP2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RP3</td>
<td>RPGR</td>
<td></td>
</tr>
<tr>
<td>RP759</td>
<td>XLRP</td>
<td>RP3 and RP23</td>
<td>RP2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RP3</td>
<td>RPGR</td>
<td></td>
</tr>
<tr>
<td>RP2689</td>
<td>XLRP</td>
<td>RP3 and RP23</td>
<td>RP2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RP3</td>
<td>RPGR</td>
<td></td>
</tr>
<tr>
<td>RP3126</td>
<td>XLRP</td>
<td>RP3 and RP23</td>
<td>RP2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RP3</td>
<td>RPGR</td>
<td></td>
</tr>
<tr>
<td>RP107</td>
<td>XLRP</td>
<td>RP2 and RP24</td>
<td>RP2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RP3</td>
<td>RPGR</td>
<td></td>
</tr>
<tr>
<td>RP3877</td>
<td>XLRP</td>
<td>RP2 and RP24</td>
<td>RP2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RP3</td>
<td>RPGR</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>F15</td>
<td>XLRP</td>
<td>RP2 and RP24</td>
<td>RP2</td>
<td>RP2</td>
</tr>
<tr>
<td>RP3973</td>
<td>XLRP</td>
<td>RP2 and RP24</td>
<td>RP2</td>
<td></td>
</tr>
<tr>
<td>RPBI</td>
<td>XLRP</td>
<td>RP2, RP3 and RP24</td>
<td>RPGR</td>
<td>RP2</td>
</tr>
<tr>
<td>MGRP4</td>
<td>XLRP</td>
<td>RP2, RP3 and RP24</td>
<td>RP2</td>
<td>RPGR</td>
</tr>
<tr>
<td>RP4</td>
<td>XLRP</td>
<td>RP2, RP3 and RP24</td>
<td>RP2</td>
<td>RPGR</td>
</tr>
<tr>
<td>RP34</td>
<td>XLRP</td>
<td>RP2, RP3 and RP24</td>
<td>RP2</td>
<td>RPGR</td>
</tr>
<tr>
<td>RP76</td>
<td>XLRP</td>
<td>RP2, RP3 and RP24</td>
<td>RP2</td>
<td>RPGR</td>
</tr>
<tr>
<td>RP1571</td>
<td>XLRP</td>
<td>RP2, RP3 and RP24</td>
<td>RP2</td>
<td>RPGR</td>
</tr>
<tr>
<td>RP2063</td>
<td>XLRP</td>
<td>RP2, RP3 and RP24</td>
<td>RP2</td>
<td>RPGR</td>
</tr>
<tr>
<td>RP277</td>
<td>XLRP</td>
<td>RP2, RP3 and RP24</td>
<td>RP2</td>
<td>RP2</td>
</tr>
<tr>
<td>RP01</td>
<td>XLRP</td>
<td>RP2, RP3 and RP23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F53</td>
<td>XLRP</td>
<td>RP2, RP3 and RP23</td>
<td>RP2</td>
<td>RPGR</td>
</tr>
<tr>
<td>RP87</td>
<td>XLRP</td>
<td>RP2, RP3 and RP23</td>
<td>RPGR</td>
<td>RPGR</td>
</tr>
<tr>
<td>RP2L</td>
<td>XLRP</td>
<td>RP2, RP3, RP23 and RP24</td>
<td></td>
<td>RPGR</td>
</tr>
<tr>
<td>F21</td>
<td>XLRP</td>
<td>RP2, RP3, RP23 and RP24</td>
<td>RP2</td>
<td>RP2</td>
</tr>
<tr>
<td>RP180</td>
<td>XLRP</td>
<td>RP2, RP3, RP23 and RP24</td>
<td>RP2</td>
<td>RPGR</td>
</tr>
<tr>
<td>F71</td>
<td>XLRP</td>
<td>RP2, RP3, RP23 and RP24</td>
<td>RP2</td>
<td>RP2</td>
</tr>
<tr>
<td>F72</td>
<td>XLRP</td>
<td>RP2</td>
<td>RP2</td>
<td>RP2</td>
</tr>
<tr>
<td>NRP</td>
<td>XLRP</td>
<td>RP2</td>
<td>RP2</td>
<td>RP2</td>
</tr>
<tr>
<td>GT1</td>
<td>XLRP</td>
<td>RP2</td>
<td>RP2</td>
<td>RP2</td>
</tr>
</tbody>
</table>

* RPGR screening of exons 1-19 including exon 15a but excluding the recently identified ORF15
# CACNA1F screening was performed in Cambridge by Dr. L. Allen
3.4 Discussion

Different XLRP subtypes cannot be identified clinically. Although some reports suggest there may indeed be phenotypic differences (Kaplan et al. 1992), other reports state that there is not such difference (Flaxel et al. 1999) and the data cannot be used as a diagnostic tool. Genetic analysis therefore provides the only means of distinguishing between them. The clustering of XLRP loci on Xp also makes it difficult to genetically discriminate within XLRP families, especially between the two prevalent subtypes RP3 and RP2. It is often necessary to obtain large numbers of samples from an extended family in search of a key recombination. Nevertheless, when the genetic distinction is possible, it appears that RP3 is the predominant form (60-90\%) of XLRP in different Caucasian populations (Ott et al. 1990, Teague et al. 1994, Meindl et al. 1996, Fujita et al. 1997), whilst 10-30\% of XLRP appears to segregate with RP2 (Thiselton et al. 1996), only a minor proportion segregating with the remaining loci (RP23 and RP24).

Abridged pedigrees are shown to represent samples corrected and analyzed. It should be noted that although potential disease haplotypes are drawn on the X chromosome, resulting from the analysis described, some of the pedigrees are too small to say with certainty that they indeed represent X-linked inheritance (e.g. RP158, RP4). However, some small pedigrees were subsequently shown to have mutations in an XLRP gene (e.g. F75, RP2689, RP76, see table 4.2).

In this study the haplotype analysis conducted on 33 XLRP families has confirmed the data so far reported. As seen in Table 3.3 the majority of these families segregate with the RP3 locus. In some cases (see first nine families in Table 3.3) proximal and distal crossovers identified in close proximity to the RP3 region defined a minimal region of disease segregation at this locus. In other families the distal and proximal crossovers were a greater genetic distance apart including the RP23 or the RP2 locus and occasionally, where tested, even the recently described RP24 locus on Xq. In some families a distal crossover (families NZ1, F75, XRP91 RP2689, and RP87) or proximal crossover (families RP3877, RP3973, RP4, RP34, RP1571 and RP2L) were not identified using the markers described, these families are therefore considered as also potentially spanning the RP23 or the RP24 locus.
Haplotype and linkage analysis in families with X-linked retina disease

There are no families segregating exclusively with the RP2 locus, this is due to the small genetic interval of about 20-22 cM separating RP2 and RP3 which, as already mentioned, makes it difficult to discriminate between the two loci within XLRP pedigrees. This is also one of the reasons why the cloning of the RP2 gene in the absence of deletion patients was delayed until recently after the first mapping of the gene in 1984 (Bhattacharya et al.).

The conclusive genetic picture of XLRP may still be ill defined, since the genetics has changed even over the past last few years. Four loci were initially proposed on the short arm of the X chromosome, RP2, RP3, RP6 and RP15 (Figure 3.1a). The genes at the two main loci RP2 and RP3 have now been identified. RP6 still remains as a possible XLRP locus, although it has only been demonstrated by statistical evidence, and RP15 in fact does not exist (section 3.1.1). In addition two new loci have been identified, RP23 and RP24. It appears, therefore, that there are at least 4 XLRP genes. The frequencies for disease segregation at the RP2 and RP3 loci have been estimated and differ slightly according to the population studied, whilst for RP23 and RP24 we can only conclude that they are both very rare. Further genetic studies world wide may identify an XLRP family segregating with the proposed RP6 locus, or may reveal new XLRP loci.

Genetic heterogeneity has also been demonstrated for CSNBX with the localisation of at least two distinct loci (CSNB1 and CSNB2) on Xp. Although it has been reported that CSNBX is also clinically heterogeneous, with a so-called complete and incomplete forms with respect to the level of rod function in the disease, a clear correlation was not always observed between phenotypic descriptions and different map locations (Pearce et al 1990, Aldred et al. 1992, Bergen et al. 1995). Again genetic analysis is the most accurate way to distinguish between the different loci.

Haplotype analysis was carried out on 4 CSNBX families. It was possible to discriminate between the two known loci for one of these families (Table 3.5), whilst for two other families no crossovers were identified with the markers tested, and for another family (RP2124) the crossover detected did not enable the exclusion of either CSNBX loci. Ultimately gene screening would be the only way to identify the disease locus in these
families, unless new individuals from the families could be tested, who may reveal recombinations.

Due to the complexity of the genetics of XLRP and CSNBX, and lack of conclusive phenotypic definition of different loci, a systematic haplotype analysis of all families is still important. Firstly it helps to discriminate between different loci hence providing a clearer picture of the relevance of each locus and secondly it may provide new evidence for the presence of different loci associated with the disease (see chapter 5). Such a study is also very useful with respect to subsequent gene screening (see chapter 4), as families not segregating with a particular gene can be excluded from the analysis, and microheterogeneity can be evaluated (see chapter 4).
CHAPTER 4

Mutation screening of the RPGR, RP2 and NYX genes

4.1 Introduction

4.1.1 The RPGR gene

In 1996 a gene named RPGR (retinitis pigmentosa GTPase regulator) was cloned from the RP3 region and pathological mutations were found to be associated with disease (Meindl et al. 1996, Roepman et al. 1996a). The RPGR gene is ubiquitously expressed and consists of 19 exons (Figure 4.1a shows an ideogram of the gene). The putative RPGR gene product, of 815 amino acids, shows a high degree of homology to the RCC1 (regulator of chromosome condensation) protein. This region of homology is localised within the N-terminal half of the predicted protein, including exons 3 to 9 plus part of exons 2 and 10 (Figure 4.1a), and consists of six complete tandem repeats of 52-54 amino acids (preceded by one incomplete repeat) with a high degree of conservation across vertebrates. RCC1 regulates a GTPase known as Ras-related nuclear protein by its tandemly arranged repeats that act as the catalytic site for the exchange of guanine nucleotides (Bischoff and Ponsingl 1991, Klebe et al. 1995). Similarly the RPGR gene product may interact with a retina specific Ras-related nuclear protein. It has subsequently been shown that the RCC1-like domain of RPGR interacts with the δ subunit of rod cyclic GMP phosphodiesterase (PDEδ), and that disruption of this interaction may be the cause of, or contribute to, retinal disease (Linari et al. 1999). However the role of RPGR in the retina and its involvement in the aetiology of retinitis pigmentosa have not yet been resolved. More recently interaction of RPGR with a retina specific protein of unknown function (RPGRIP) has been demonstrated (Boylan et al. 2000, Roepman et al. 2000). Disease associated mutations of RPGR have been shown to disrupt this interaction, and the two proteins co-localise in the outer segment of the photoreceptors. These results provide a clue for the retina specific pathogenesis of RPGR.
Figure 4.1: Ideogram of the genomic structure of the (a) RPGR (b) RP2 and (c) NYX genes. Exons are numbered and are shown as boxes (drawn to scale) and introns are shown as lines between boxes (not drawn to scale). In a) the blue-shaded boxes represent region of homology with RCC1. In b) the pink-shaded box represents region of homology with cofactor C. In c) the green-shaded boxes represent the coding region of the gene and the patterned box represents region containig 11 LRRs plus 2 cysteine-rich LRRs.
Mutation screening of the RPGR, RP2 and NYX genes

mutations. In 1999 a new retina specific exon (exon 15a) was identified (Kirschner et al.). This exon is contained within intron 15, it introduces a new stop codon producing a predicted gene product of 646 instead of 815 amino acids and it was found to be deleted in one patient.

Mutations in RPGR seem to account for approximately 20-25% of XLRP families (Meindl et al. 1996, Roepman et al. 1996a, Buraczynska et al. 1997, Zito et al. 1999, Miano et al. 1999, Zito et al. 2000), which does not correlate with the large proportion of families segregating with the RP3 locus (70-75%). The general consensus at the stage when this study was undertaken was that additional coding sequences, the regulatory region of the gene or even a different gene close to RPGR may be involved in the proportion of familial disease still unaccounted for.

The hypothesis of microheterogeneity (i.e. a different gene lying proximal to RPGR accounting for a proportion of disease) was particularly attractive, especially in view of the fact that the reports describing RPGR mutations generally lacked information on the genetic mapping of disease in the families used in their studies, the exception being Fujita et al. (1997).

For the purpose of this study, to investigate the proportion of disease in genetically defined RP3 families caused by mutations in RPGR, only families previously characterised in which disease segregates with this genetic interval were screened.

In a recently published paper (Vervoort et al. 2000), the issue of the large proportions of XLRP families not accounted for by mutations in the known genes (RPGR and RP2) has apparently been solved with the identification of 2 new overlapping RPGR exons (ORF14 and ORF15). Exon ORF14 is generated by the retention of intron 14, whilst exon ORF15 contains exon 15 and extends into part of intron 15. The transcript containing ORF15 appears to be widely expressed with more prominent RT-PCR products in retina, and contains an unusual region of low complexity with high glutamic acid and glycine content. A proportion of 80% of the RPGR mutations they have identified in their XLRP patients seem to be localised in exon ORF15, which appears to be a mutation hot spot.

The RPGR screen presented in this chapter does not include this newly identified exon.
4.1.2 The *RP2* gene

*RP2* was the first genetically mapped retinitis pigmentosa locus (Bhattacharya *et al.* 1984) (Xp11.3-p11.23). Genetic mapping studies suggested that *RP2* is rare (15-20% of disease, Teague *et al.* 1994) in relation to the more common RP3, and the close proximity of the two loci has hampered the search for the *RP2* gene for many years. It was fourteen years after this initial genetic mapping that the gene was finally cloned (Schwahn *et al.* 1998).

*RP2* consists of 5 exons, encodes a polypeptide of 350 amino acids and, like *RPGR*, is ubiquitously expressed (Figure 4.1b shows an ideogram of the gene). The only functional clue for the *RP2* gene is the homology over 151 codons (30.4%) of the *RP2* predicted amino acid sequence with cofactor C, which is involved with β-tubulin folding (Schwahn *et al.* 1998) suggesting that *RP2* may also have a role in tubulin biogenesis. The identification of pathogenic point mutations at residues that are conserved with cofactor C (Schwahn *et al.* 1998, Hardcastle *et al.* 1999) supports this hypothesis. Recently *RP2* has been principally localised to the plasma membrane (Chappie *et al.* 2000) and potential sites for N-terminal acylation that would target the protein to the plasma membrane have been identified suggesting that *RP2* may not be functioning exclusively in tubulin folding. The overall function and specific role of the protein in the retina are yet to be unravelled.

To date mutations identified in the *RP2* gene seem to account for approximately 10-20% of XLRP families (Schwahn *et al.* 1998, Harcastle *et al.* 1999, Mears *et al.* 1999, Thiselton *et al.* 2000), which is in agreement with genetic mapping studies that suggest that about 15-25% of families segregate with the *RP2* locus (Teague *et al.* 1994).

4.1.3 The *NYX* gene

The gene responsible for the complete form of CSNB on the X chromosome at the CSNB1 locus (in Xp11.4) has been recently cloned by two different groups (Pusch *et al.* 2000 and Bech-Hansen *et al.* 2000) and has been named *NYX* (for nyctalopin on the X chromosome). The gene consists of 3 exons spanning 28 Kb of genomic sequence and has a translation start site (ATG) lying near the 3’ end of exon 2 (Figure 4.1c). The full-length cDNA has an ORF of 1443 bp encoding a protein of 481 amino acids that contains an N-terminal signal peptide and a C-terminal glycosylphosphatidylinositol (GPI)-anchor presumably anchoring the protein to the extracellular membrane. The gene has been shown to
be expressed in retina and in other tissues including testis, placenta, skeletal muscle, brain, kidney and retinoblastoma (Pusch et al. 2000).

The ORF contains 11 typical leucine-rich domains (LRRs) flanked on each side by a cysteine-rich LRR. This motif is important for protein-protein interaction and members of the LRR superfamily are involved in cell adhesion and axon guidance (Hocking et al. 1998). It is suggested that nyctalopin may be required for establishing or maintaining functional contacts between rod photoreceptor cells and post-synaptic neurons including bipolar and amacrine cells.

A number of mutations have been identified in patients with the complete form of CSNB by both groups (Pusch et al. 2000 and Bech-Hansen et al. 2000).

4.1.4 Allelism in X-linked retinopathies

The X chromosome hosts several different genes involved in retinal disease (i.e. RP, CSNB, COD, RS, section 1.1.4). Considering the high density of loci identified especially on the short arm of the X chromosome and particularly in the region spanning Xp21.1 to Xp11.23 where two RP loci (RP2 and RP3) and two CSNB loci (CSNB1 and CSNB2) have been localised (Figure 4.2), it was tempting to speculate on the possibility of these retinopathies being allelic.

RP and CSNB share clinical characteristics such as night blindness and visual field constriction, the main difference being the stationary condition of the patients suffering CSNB related to the progressive nature of RP. Hence, it is possible that different mutations in the same gene may account for these different phenotypes and indeed there often appears to be a correlation between the severity of a disease and the localisation or the extent of the mutation that causes that disease.

To give strength to this hypothesis, mutations are reported in rhodopsin that can cause an autosomal dominant form of RP as well as an autosomal dominant form of CSNB, and mutations in rod cGMP phosphodiesterase β-subunit that can cause a recessive form of RP as well as an autosomal dominant form of CSNB (Gal et al. 1994, Rao et al. 1994).

Many of the XLRP or CSNBX families cannot be genetically defined precisely to determine which is the only locus segregating with disease. A considerable number of families with XLRP were still unaccounted for by mutation screen of the RPGR and the RP2
Figure 4.2: Ideogram of the short arm of the X chromosome showing map location of XLRP and CSNB loci.
gene (both CSNB genes, \textit{CACNAIF} and \textit{NYX}, were cloned fairly recently and there has not yet been such an extensive worldwide screen of these genes in CSNB families), therefore any of these genes were considered as potential candidate genes for any of these diseases.

4.2 Materials and methods

4.2.1 Patients

4.2.1.1 Patients screened for the \textit{RPGR} gene

A total of 38 XLRP patients were screened for mutations in the \textit{RPGR} gene. In the first round of mutation detection carried out by SSCP analysis, 27 patients were screened. The majority of these patients are from families already presented in section 3.2.1.1: RP106, RP158, RP296, RP759, RP1120, RP1591, RP1750, RP3126, RP87, XRP90, RP2063, RP2786, F75, RPB1, RP4, RP34, RP180 and F53 (Table 3.3). Another eight patients were also included in this initial study from the following families: RRP, BRP, RP51, RP66, RP147, RP380, RP509, RP1206 and RP1484, these eight families have all been previously haplotyped in the laboratory and they were all shown to be linked to the RP3 locus.

In a second round of mutation screening, carried out this time by direct sequencing, another ten patients were included from the following families: RPIL, RP1600, MGRP4, RP639, XRP91, RP2689, RP2L, RP76, RP1571, all of these previously presented (section 3.2.1.1, Table 3.3) plus an additional patient from a British family, F51, clinically ascertained through Moorfields Eye Hospital, and again previously shown to be linked to RP3 by haplotype analysis in the laboratory.

Family NZ1, which was described in section 3.2.1.1, was also screened for mutations in \textit{RPGR} by direct sequencing but will be presented in a separate section (4.3.3).

Families RP2788 and RPO1, which were discussed in section 3.2.1.1 and disease was shown to be linked to the RP3 locus (Table 3.3), were not included in this study because they became available after the gene screening was pursued.

4.2.1.2 Patients screened for the \textit{RP2} gene

A total of 54 patients from XLRP families were screened for mutations in the \textit{RP2} gene. In the first round of screening, from a total of 59 XLRP families available at the time this study started, 26 were excluded on the basis of linkage data or mutation data (e.g.
mutations already identified in the *RPGR* gene), hence 33 families presented in Hardcastle *et al.* 1999 were assessed. The following families already described in section 3.2.1.1 and for which disease was shown to segregate with the RP2 locus (Table 3.3), were included in this screening: MGRP4, XRP90, RP639, RP2786, RP107, RP3877, RP3973, RPB1, RP4, RP34, RP76, RP180, RP1571, RP2063, F53, RP87. Families RPB1, RP2L and RPO1, which were also shown to be linked to the RP2 locus (Table 3.3) were not included because they were not available at the time of screening. The second round of screening comprised 21 families presented in Thiselton *et al.* 2000.

### 4.2.1.3 CSNB patients screened for the *RPGR* and *RP2* gene

The following 10 CSNBX British families clinically ascertained through Moorfields Eye Hospital, and Addenbrooks Hospital (Cambridge), were screened for mutations in both the *RPGR* and *RP2* genes: RP3665, RP4126, RP3810, AD1001, RP1002, RP3623, GC1006, RP4121, RP4611, GC4180.

### 4.2.1.4 Patients screened for the *NYX* gene

A total of 21 CSNB families were screened for mutations in the *NYX* gene including eight of the patients already screened for *RPGR* and *RP2* (section 4.2.1.3), two of the families (AD1001 and GC1006) being excluded on the basis of a parallel study done in Cambridge (Dr. L. Allen, personal communication) where they were shown to have mutations in the *CACNA1F* gene. Another 5 CSNBX families were included in this screening; RP4914, RP2124, CRP, CDT and CCG, plus an additional 8 sporadic patients with no family history.

### 4.2.1.5 Controls

For each genetic alteration identified a minimum number of 50 control females (=100 chromosomes) was screened, with the exception of exon 1 of the *RPGR* gene for which 40 control females were analysed.
4.2.2 Mutation detection

4.2.2.1 PCR conditions

Exon fragments for each of the three genes were amplified with the respective intronic primer pairs. The RPGR gene primers used were as described in Meindl et al. (1996) for exons 1 to 19 and in Kirschner et al. (1999) for exon 15a. Additional primers for exon 8 were designed also spanning the exon and intron/exon boundaries (8b-F: atcctgtttgatattgccc and 8b-R: ccagtctataaatataacag). The RP2 gene primers used were as described in Schwahn et al. (1998). For the NYX gene, primers amplifying the two coding exons (exons 2 and 3) were as described by Pusch et al. (2000), subsequently primers for exon 3 were redesigned. Primers amplifying the non-coding exon 1 were designed from the genomic sequence contained in PAC Z93015 (section 4.5.1). Table 4.1 shows newly designed primer pairs for exons 1, primer pairs for exon 2 (as in Pusch et al. 2000) and primer pairs for exon 3 (split into 4 fragments) redesigned prior to this study. PCR product size, buffer used with the optimum magnesium concentration and annealing temperature are also shown in the table.

### Table 4.1: Primer pairs for the 2 coding exons of the NYX gene

<table>
<thead>
<tr>
<th>Exon 1</th>
<th>Forward primer (5'-3')</th>
<th>Reverse primer (5'-3')</th>
<th>Size (bp)</th>
<th>Buffer*</th>
<th>Ann. T</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acctttactctctctcaacca</td>
<td>ggcatacgtacaaccacccgc</td>
<td>260</td>
<td>KCl (incl. 15mM MgCl₂)</td>
<td>60°C</td>
<td></td>
</tr>
<tr>
<td>Exon 2</td>
<td>ggcatccaggaagaaggctg</td>
<td>aagactctttggcaggaagc</td>
<td>238</td>
<td>KCl (incl. 15mM MgCl₂)</td>
<td>60°C</td>
</tr>
<tr>
<td>Exon 3a</td>
<td>ggctttggctgacggtgc</td>
<td>gtgcaggacggtgctgc</td>
<td>418</td>
<td>NH₄ (+ 15 mM MgCl₂)</td>
<td>60°C</td>
</tr>
<tr>
<td>Exon 3b</td>
<td>acaaccttgctctctcaacccgc</td>
<td>cagttgacgcggagcgag</td>
<td>497</td>
<td>NH₄ (+ 10 mM MgCl₂)</td>
<td>60°C</td>
</tr>
<tr>
<td>Exon 3c</td>
<td>gactgtggcgtcttgagcgc</td>
<td>ggtcaccgtgctgactgc</td>
<td>450</td>
<td>NH₄ (+ 10 mM MgCl₂)</td>
<td>60°C</td>
</tr>
<tr>
<td>Exon 3d</td>
<td>atggaggctcgagcag</td>
<td>ttaccacacaaccactcaagcc</td>
<td>450</td>
<td>NH₄ (+ 15 mM MgCl₂)</td>
<td>60°C</td>
</tr>
</tbody>
</table>

* From Bioline

All PCRs were carried out as described in materials and methods (section 2.6.1) with special conditions being used for exon 1 of the RPGR gene and exon 3 of the NYX gene both of which have a high GC content. For amplification of RPGR exon 1, 5% formamide and 10% glycerol were added to PCR reactions and NH₄ buffer with 10 mM MgCl₂ was used. For amplification of the NYX exon 3, 1 μl of DMSO per 25 μl reaction was added and again NH₄ buffer with different magnesium concentrations was used as described in table 4.1.
4.2.2.2 Single-strand conformation polymorphism analysis

SSCP analysis was carried out as described in sections 2.6.4 and 2.4.3 for exons 1 to 19 of the RPGR gene in 27 of the patients screened, and for the 5 coding exons of the RP2 gene in 21 of the patients screened.

4.2.2.3 Direct automated sequencing of PCR products

Direct sequencing was carried out as described in section 2.7 for PCR products of patients screened for the NYX gene, on the remaining patients screened for the RPGR and the RP2 genes and on all PCR products of samples showing a band shift on SSCP.

4.2.2.4 Restriction enzyme digestion analysis

PCR fragments of RPGR exons 5 and 8 were digested with HaeIII restriction enzyme (Promega) and BssSI restriction enzyme (New England BioLabs) respectively, using the appropriate buffers provided by the manufacturers. PCR fragments for exon 2 of NYX were digested with SphI restriction enzyme (Promega), using the appropriate buffer provided by the manufacturer. 20 μl of the amplification reaction was brought to a final volume of 30 μl by the addition of 10X Buffer, 6 U enzyme and distilled water. Digestion was performed at 37°C for 16 hours.

4.2.2.5 Sequence analysis

The Blast program (http://www.ncbi.nlm.nih.gov/BLAST/) (section 2.8.2) and the DNASTAR software program (including EditSeq™ and MapDraw™) (section 2.8.3) were used as a tool to compile the cDNA sequence of the NYX gene.

4.3 Pathogenic mutations and polymorphic variants of the RPGR gene

A total of 10 mutations (Table 4.2) in 11 different families and 5 polymorphisms (Table 4.3) were identified in the 38 families screened for the RPGR gene in this study.

The results of this screening will be presented in the three following separate sections (4.3.1, 4.3.2 and 4.3.3).
Mutation screening of the RPGR, RP2 and NYX genes

Table 4.2: Mutations identified in the RPGR gene in XLRP patients with an RP3 genotype

<table>
<thead>
<tr>
<th>Family</th>
<th>Intron/Exon</th>
<th>DNA Position*</th>
<th>Change</th>
<th>Amino acid no.</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>RP106a, MGRP4b</td>
<td>Intron 1</td>
<td>86+1</td>
<td>G→A</td>
<td>-</td>
<td>Putative splicing</td>
</tr>
<tr>
<td>RP1120a</td>
<td>Intron 1</td>
<td>86+5</td>
<td>G→A</td>
<td>-</td>
<td>Putative splicing</td>
</tr>
<tr>
<td>RP1600b</td>
<td>Exon 5</td>
<td>421</td>
<td>T→C</td>
<td>121</td>
<td>Leu→Pro</td>
</tr>
<tr>
<td>NZ1c</td>
<td>Intron 6</td>
<td>678+5</td>
<td>G→A</td>
<td>-</td>
<td>Putative splicing</td>
</tr>
<tr>
<td>F75a</td>
<td>Exon 7</td>
<td>807-808</td>
<td>2-bp deletion</td>
<td>250</td>
<td>Frameshift</td>
</tr>
<tr>
<td>RP76b</td>
<td>Exon 8</td>
<td>950-951</td>
<td>2-bp deletion</td>
<td>297</td>
<td>Frameshift</td>
</tr>
<tr>
<td>RP1571b</td>
<td>Exon 8</td>
<td>963</td>
<td>T→C</td>
<td>302</td>
<td>Cys→Arg</td>
</tr>
<tr>
<td>RP2689b</td>
<td>Exon 8</td>
<td>Uncharacterised</td>
<td>Deletion</td>
<td>-</td>
<td>Deletion</td>
</tr>
<tr>
<td>XRP90a</td>
<td>Exon 10</td>
<td>1297-1298</td>
<td>2-bp insertion</td>
<td>413</td>
<td>Frameshift</td>
</tr>
<tr>
<td>RP87a</td>
<td>Exon 11</td>
<td>1435-1436</td>
<td>2-bp deletion</td>
<td>459</td>
<td>Frameshift</td>
</tr>
</tbody>
</table>

*Presented in section 4.3.1
bPresented in section 4.3.2
cPresented in section 4.3.3

Table 4.3: Polymorphisms identified in the RPGR gene

<table>
<thead>
<tr>
<th>Intron/Exon</th>
<th>DNA Position*</th>
<th>Change</th>
<th>Function</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exon 3</td>
<td>286</td>
<td>G→T</td>
<td>76</td>
<td>Ser→Ile 10%</td>
</tr>
<tr>
<td>Exon 10</td>
<td>1223</td>
<td>G→A</td>
<td>388</td>
<td>Silent 18%</td>
</tr>
<tr>
<td>Exon 11</td>
<td>1333</td>
<td>G→A</td>
<td>425</td>
<td>Arg→Lys **</td>
</tr>
<tr>
<td>Exon 14</td>
<td>1756</td>
<td>G→A</td>
<td>566</td>
<td>Gly→Glu **</td>
</tr>
<tr>
<td>Intron 18</td>
<td>2300+11</td>
<td>T→C</td>
<td>-</td>
<td>- 37%</td>
</tr>
</tbody>
</table>

*According to Figure 1.11
**Observed only once but previously reported as a polymorphism (Burackzynska et al 1997)

4.3.1 RPGR mutations and polymorphisms identified by SSCP analysis in a subset of 27 XLRP families segregating with the RP3 locus

A total of 10 genetic alterations were identified by SSCP analysis in seven different exons of the RPGR gene (from exons 1 to 19 plus exon 15a) in 27 XLRP families segregating with the RP3 locus (Zito et al. 1999). Five of the ten changes were classified as polymorphisms (Table 4.3) either because they were previously described as polymorphisms
Mutation screening of the RPGR, RP2 and NYX genes

or because they were also present in unaffected control chromosomes. The remaining five changes were classified as potential novel mutations found in four different exons. Each mutation is described as a correspondent change at the nucleotide and/or the amino acid level (Table 4.2a, see also Figure 1.11).

In families RP106 and RP1120 a nucleotide change was found in the donor splice site respectively at the first and at the fifth nucleotide downstream of the exon/intron boundary of exon 1 (Figure 4.3a and 4.3b). Exon 1 was sequenced in 80 control chromosomes and these changes were not detected. In addition for both families the segregation of the nucleotide change with disease was demonstrated by sequencing 3 affected and 3 unaffected males (data not shown) and analysis of the entire gene in both families did not reveal any other sequence alterations. Biological proof that these nucleotide changes are causing the disease in these families requires further experimentation initially by reverse transcription PCR. Unfortunately fresh patient blood was unavailable. Nevertheless all intron splice acceptor and donor sequences of the human RPGR gene follow the GT-AG rule and the first nucleotide of a donor splice sequence is 100% conserved (Horowitz and Krainer 1994), hence a G→A transition at this site is very likely to affect RNA splicing. The fifth nucleotide is 80% conserved, therefore a G→A transition at this position is also likely to affect RNA splicing.

Direct automated sequencing of DNA from another 3 patients that presented a shift on an SSCP gel led to the identification of 3 different frameshift mutations. Family F75 (Figure 4.3c) has a 2 base pair deletion in exon 7 corresponding to nucleotides 807 and 808 in the normal sequence. In family XRP90 (Figure 4.3d) 2 base pairs have been inserted in exon 10 between nucleotides 1297 and 1298 of the normal sequence. Family RP87 (Figure 4.3e) has a 2 base pair deletion in exon 11 corresponding to nucleotides 1435 and 1436 in the normal sequence. In all three cases, the resulting frameshift gives rise to a considerably truncated protein predicted to lead to lack of function. Segregation of the mutations with disease was demonstrated in each family by SSCP analysis, examples of which are seen in Figure 4.4a-b (families F75 and XRP90 respectively).
Figure 4.3: Electropherograms depicting patient mutations and normal sequences for the *RPGR* gene. Arrows indicate the sites of mutations.

a) A G>A change at the first nt downstream of the exon1/intron1 boundary in an affected male from family RP106 and normal sequence of the corresponding region.

b) A G>A change at the fifth nt downstream of the exon1/intron1 boundary in an affected male from family RP1120 and normal sequence of the corresponding region.

c) A 2-bp deletion (GT) in exon 7 in an affected male from family F75 and normal sequence of the corresponding region.

d) A 2-bp insertion (AG) in exon 10 in an affected male from family XRP90 and normal sequence from the corresponding region.

e) A 2-bp deletion (TC) in exon 11 in an affected male from family RP87 and normal sequence of the corresponding region.
Figure 4.4: Segregation analysis of two novel RPGR mutations. Arrows on the right depict the informative bands. Pedigrees are aligned to correspond to the respective gel lane.

a) Segregation of the exon 7 frameshift mutation in family F75.

b) Segregation of the exon 10 frameshift mutation in family XRP90.
4.3.2 Additional *RPGR* mutations and polymorphisms identified by direct sequencing in a subset of 10 XLRP families segregating with the RP3 locus

Direct sequencing of PCR products of exons 1 to 19 plus exon 15a of *RPGR* in ten families segregating with the RP3 locus, led to the identification of five mutations in five families (summarised in Table 4.2b, see also figure 1.11) (Zito *et al.* 2000a). Polymorphisms identified in this subset of families are as described in section 4.3.1 (Table 4.3) in exon 10 and intron 18.

In family MGRP4, the mutation identified was a G>A transition 1 nucleotide downstream of the exon 1/ intron 1 boundary. Segregation of the mutation with disease was demonstrated by sequencing the available members of the family (data not shown). This mutation has also been identified in another family RP106 (section 4.3.1, Figure 4.3a and Zito *et al.* 1999), therefore strengthening its role as a pathogenic mutation.

Family RP1600 carries a T>C change at nucleotide position 421 in exon 5 (Figure 4.5a) which results in a Leu>Pro change at codon 121. This codon is conserved across species including mouse, dog, cow and fugu. The T>C change creates a recognition site for *Hae*III restriction enzyme and this was used as a tool to demonstrate segregation of the mutation with disease in the family (data not shown) and to exclude the presence of the same change in 110 control chromosomes. No other changes were observed in this family in the other 18 exons plus exon 15a.

In family RP76 a 2 bp deletion was observed at nucleotide position 950-951 (Figure 4.5b). This deletion causes a frameshift that is predicted to create 47 novel amino acids and a new stop codon that is predicted to produce a protein product of 344 amino acids with 471 amino acids missing. By sequence analysis the mutation was shown to segregate with the disease in the family (data not shown).

In family RP1571 a T>C transition was detected at position 963 (Figure 4.5c), producing a Cys>Arg change at codon 302, which is conserved in human, mouse and dog. This change creates a restriction site for *Bss*SI enzyme, which was used to demonstrate segregation of the mutation with disease in the family (Figure 4.6a) and lack of the same mutation in 100 control chromosomes. Sequencing of the remaining 18 exons plus exon 15a in this family showed no other variations.
Figure 4.5: Electropherograms depicting patient mutations and normal sequences for the *RPGR* gene. Arrows indicate the sites of mutations.

a) A G>A change at in exon5 in an affected male from family RP1600 and normal sequence of the corresponding region.

b) A 2-bp deletion (AA) in exon8 in an affected male from family RP76 and normal sequence of the corresponding region.

c) A T>C change in exon8 in an affected male from family RP1571 and normal sequence of the corresponding region.
Figure 4.6: Segregation analysis of two novel RPGR mutations. Pedigrees are aligned to correspond to the respective gel lane.

a) Segregation of the exon 8 missense mutation in family RP1571 by restriction enzyme digestion analysis.

b) Segregation of the exon 8 deletion mutation in family RP2689 by gel electrophoresis analysis.
Mutation screening of the RPGR, RP2 and NYX genes

In family RP2689 exon 8 failed to amplify in the affected members of the family. To exclude the possibility of sequence alterations within the primer sequences primers 8b-F and 8b-R were used (section 4.2.2.1). The new primers again failed to amplify in the affected members of the family. This deletion needs further characterisation, unfortunately the full genomic sequence for RPGR is not available and PCR across exons 7-9 has been carried out with no resulting amplification product (probably due to the size of the introns). Southern blot analysis is currently unachievable due to a shortage of DNA. However, by PCR and agarose gel electrophoresis the mutation was shown to segregate with disease in the family (Figure 4.6b) and the remaining 18 exons plus exon 15a present no alterations.

4.3.3 Sequence variations within the RPGR gene: evidence for a founder complex allele

Sequencing analysis of exons 1 to 19 of RPGR gene in an affected member of family NZ1 showed 8 different sequence variations (Table 4.4) (Zito et al 2000b).

<table>
<thead>
<tr>
<th>Intron/exon</th>
<th>Position*</th>
<th>Change</th>
<th>Amino acid no</th>
<th>Consequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intron 6</td>
<td>678+5</td>
<td>G→A</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Intron 8</td>
<td>Primer sequence</td>
<td>Not characterised</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Exon 11</td>
<td>1333</td>
<td>G→A</td>
<td>425</td>
<td>Arg→Lys</td>
</tr>
<tr>
<td>Intron 13</td>
<td>1631+11</td>
<td>A→G</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Exon 14</td>
<td>1635-1637</td>
<td>3-bp deletion</td>
<td>526</td>
<td>Gln deleted</td>
</tr>
<tr>
<td>Exon 14</td>
<td>1657</td>
<td>C→T</td>
<td>533</td>
<td>Thr→Met</td>
</tr>
<tr>
<td>Exon 14</td>
<td>1756</td>
<td>G→A</td>
<td>566</td>
<td>Gly→Glu</td>
</tr>
<tr>
<td>Intron 18</td>
<td>2300+11</td>
<td>C→T</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

* According to Figure 1.11

Of these variations, three changes identified (1333G>A, IVS13+11A>G and IVS18+11C>T) have all been previously described as polymorphic variants (Roepman et al. 1996a, Buraczynska et al. 1997, Fujita et al. 1997, Zito et al. 1999). The fourth variant, IVS6+5G>A (Figure 4.7a), is likely to affect splicing at this site. The fifth change is contained in the reverse primer sequence of exon 8. This exon failed to amplify with published primers 8-F and 8-R (Meindl et al. 1996), however a product of the expected size
was obtained using primer 8-F and primer 8b-R, and sequencing analysis showed that exon 8 and the intron/exon boundaries are not altered in this family.

The remaining three changes are all contained within exon 14 (180 bp) (Figure 4.7b); 1635-1637delCAA resulting in the loss of a Glu at codon 526, 1657C>T resulting in a Thr>Met change at codon 533, and 1756G>A resulting in a Gly>Glu change at codon 566, i.e. complete loss of one amino acid and two non-conservative amino acid substitutions.

All the exons for which sequence alterations were identified in the patient were analysed in the other members of this family. The sequence variation identified in intron 6 (IVS6+5G>A) is the only one that appeared to segregate with disease in the pedigree (Figure 4.8a) and is therefore the only potential disease-causing mutation. This nucleotide is part of the 5' consensus donor splice site and is 80% conserved in mammals (Horowitz and Krainer 1994). In addition exon 6 was sequenced in 275 control chromosomes and this change was not detected.

The other 7 sequence variations identified in the family, comprising 6 coding single nucleotide polymorphisms (cSNPs) and 1 coding multiple nucleotide polymorphism (i.e. 3 bp deletion), have no apparent effect on the resulting phenotype.

In order to provide evidence of polymorphism for the three protein altering variants observed in exon 14, the exon was sequenced in 344 control chromosomes from individuals chosen for their healthy status, predominantly of European origin. Surprisingly, 15 out of 344 chromosomes appeared to have all of the same three protein-altering changes in exon 14. None of the control samples had the 526delGlu or the Thr533Met change in isolation, while the Gly566Glu change, previously described as a polymorphism (Buraczynska et al. 1997) was found in one chromosome in isolation.

Sequencing of the remaining exons of the RPGR gene in these 15 chromosomes revealed a common configuration of the 7 sequence variations in the RPGR gene identified in family NZ1. The calculated prevalence of this complex haplotype is 4.3% (15/344) in the European population, demonstrating the presence of a founder effect. Figure 4.8b shows an ideogram of the complex haplotype compared to the wild type sequence.
**Figure 4.7:** Electropherograms depicting patient sequence variation in (a) exon 6 and (b) exon 14 of *RPGR* in NZ1 family and corresponding normal sequences. Arrows depict the site of sequence variations.

a) A G-to-A change at position 678+5 in intron 6.

b) I. A 3-bp deletion (CAA) at position 1635-1637 in exon 14.
   II. A C-to-T change at position 1657 in exon 14.
   III. A G-to-A change at position 1756 in exon 14.
Figure 4.8: a) Pedigree of the NZ1 family showing segregation of the IVS6+5G>A mutation with disease in the family and non-segregation of the complex allele. The boxed G represents the wild type sequence in intron 6 at position 678+5. The boxed shaded A represents the sequence variation at the same position.  
I represents the wild type sequence at positions: 1333, 1631+11, 1635-1637, 1657, 1765, 2300+11.  
I represents sequence variations at the same positions.  
The sequence variation in intron 8 is omitted.  
b) Ideogram of the RPGR gene showing the complex allele (I) and the wild-type sequence (II). Arrows pinpoint the sites where changes have occurred. The star depicts a non characterized change. Letters indicate the relative nucleotides and bars indicate deleted nucleotides. Numbers underneath refer to the exons. Exons are not drawn to scale and exons 2 to 7 are omitted.
4.4 Pathogenic mutations and polymorphic variants of the RP2 gene

A total of six mutations (Table 4.5) in 8 different families and four polymorphisms (Table 4.6) were identified from the RP2 screening carried out on 54 XLRP families. The results of this analysis will be presented in the next two sections (4.4.1 and 4.4.2).

Table 4.5: Mutations identified in the RP2 gene in eight XLRP families

<table>
<thead>
<tr>
<th>Family</th>
<th>Exon</th>
<th>DNA Position*</th>
<th>Change</th>
<th>Amino acid no.</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>F72a</td>
<td>2</td>
<td>353</td>
<td>G→A</td>
<td>118</td>
<td>Arg→His</td>
</tr>
<tr>
<td>F15a, F71a, RP277a</td>
<td>2</td>
<td>358</td>
<td>C→T</td>
<td>120</td>
<td>Tyr→Stop</td>
</tr>
<tr>
<td>NRPa</td>
<td>2</td>
<td>688-692</td>
<td>5-bp deletion</td>
<td>230</td>
<td>Frameshift</td>
</tr>
<tr>
<td>F21b</td>
<td>2</td>
<td>723</td>
<td>1-bp deletion</td>
<td>242</td>
<td>Frameshift</td>
</tr>
<tr>
<td>GT1b</td>
<td>3</td>
<td>798-801</td>
<td>4-bp deletion</td>
<td>266</td>
<td>Frameshift</td>
</tr>
<tr>
<td>RP3877a</td>
<td>4</td>
<td>929</td>
<td>insT</td>
<td>311</td>
<td>Frameshift</td>
</tr>
</tbody>
</table>

*aPresented in section 4.4.1

Table 4.6: Polymorphisms identified in the RP2 gene

<table>
<thead>
<tr>
<th>Exon</th>
<th>DNA Position*</th>
<th>Change</th>
<th>Amino acid no.</th>
<th>Change</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>105</td>
<td>A→T</td>
<td>35</td>
<td>Silent</td>
<td>3%</td>
</tr>
<tr>
<td>2</td>
<td>597</td>
<td>T→C</td>
<td>199</td>
<td>Silent</td>
<td>42%</td>
</tr>
<tr>
<td>3</td>
<td>844</td>
<td>C→T</td>
<td>282</td>
<td>Arg→Trp</td>
<td>2%</td>
</tr>
<tr>
<td>5</td>
<td>1012</td>
<td>G→T</td>
<td>338</td>
<td>Asp→Tyr</td>
<td>31%</td>
</tr>
</tbody>
</table>

*According to Figure 1.10

4.4.1 RP2 mutations identified by direct sequencing in a subset of 33 XLRP families

From the first round of screening of the RP2 gene, comprising 33 patients from XLRP families, four mutations were identified in six different families (Table 4.5a, see also Figure 1.10). Much of this work, performed by direct sequencing of the five coding exons of the gene, was done in collaboration with Dr. D. Thiselton and Dr A. Hardcastle.
A G>A change at nucleotide position 353 in exon 2 was observed in family F72, this change results in a Arg>His change at codon 118 (Figure 4.9a). To rule out the possibility of this mutation being a non-pathogenic sequence variant, it was analysed and excluded by SSCP analysis in 50 independent female controls. This mutation was also previously identified in another XLRP family (Schwahn et al 1999), strengthening its role as a pathogenic mutation.

The most common mutation found in the sample set was a C>T change at nucleotide position 358 in exon 2 (Figure 4.9b) resulting in a nonsense mutation (Tyr>Stop) at codon 120. This mutation was found in three apparently unrelated pedigrees, namely, F15, F71 and RP227.

In family NRP a 5 bp deletion in exon 2 comprising nucleotides 688 to 692 was detected (Figure 4.9c). This deletion causes a frameshift that creates 2 novel amino acids and a new stop codon producing a predicted truncated protein product with 117 amino acids missing.

In family RP3877 a 1 bp insertion (T) was observed at nucleotide position 929 in exon 4 causing a frameshift, which is predicted to create 18 novel amino acids and a new stop codon, with 23 amino acids missing from the resulting protein product.

Segregation of mutation with disease was demonstrated in each family by SSCP analysis, examples of which are shown in Figure 4.10a-b (segregation of the C358T mutation in family F71 and segregation of the 929insT in family RP3877).

4.4.2 Novel frameshift mutations and polymorphisms identified by SSCP analysis in RP2 in a subset of 21 XLRP families

For this second batch of families assessed for mutations in the RP2 gene the screening was carried out by SSCP analysis and this led to the identification of two novel mutations in two different families and 4 polymorphic variants (Tables 4.5b and 4.6, see also Figure 1.10).

In family F21, a deletion of 1 bp (T) at nucleotide position 723 in exon 2 was detected, causing a frameshift which creates 12 novel amino acids and a premature termination signal with 98 amino acids missing from the resulting protein product.
**Figure 4.9:** Electropherograms depicting patient mutations and normal sequences of the RP2 gene. Arrows indicate the sites of mutations.

a) A G>A change at position 353 in exon 2 in an affected male from family F72.
b) A C>T change at position 358 in exon 2 in an affected male from family RP227.
c) A 5-bp deletion (AAGAG) at position 688-692 in exon 2 in an affected male from family NRP.
d) A 4-bp deletion (GACA) at position 798-801 in exon 3 in an affected male from family GT1.
Figure 4.10: Segregation analysis of two novel RP2 mutations. Arrows on the right depict the informative bands. Pedigrees are aligned to correspond to the respective gel lane.

a) Segregation of the exon 2 C358T mutation in family F71
b) Segregation of the exon 4 929 insT mutation in family RP3877
Family GT1 has a 4 bp deletion mutation at nucleotide position 798 to 801 in exon 3 (Figure 4.9d) resulting in a frameshift mutation, which creates 3 novel amino acids and a new stop codon that produces a truncated protein product with 80 amino acids missing.

Segregation of both mutations with disease in each family was demonstrated by SSCP analysis (data not shown). SSCP analysis of a further 100 chromosomes did not detect either deletion, suggesting that these mutations are not variant alleles in the general population and that they underlie disease in these XLRP families.

During this mutation screen, four single nucleotide changes were also identified in the coding portion of the gene. These changes appear to represent polymorphic variants of RP2 (Table 4.6). Two of these can be detected with restriction enzymes (BsmF1 and ThlII restriction sites created by the T597C polymorphism and Sth132I restriction site abolished by the C844T polymorphism), providing a simple method for distinguishing these alleles in genetic mapping or population studies.

4.5 Pathogenic mutations and polymorphic variants of the NYX gene

4.5.1 Gene analysis

The genomic sequence of the NYX gene was initially found by Blast analysis using available primer sequences. A PAC (accession No Z93015) containing the genomic sequence of both the coding exons (2 and 3) of the gene was identified. By comparing the genomic sequence with the published protein sequence (Pusch et al. 2000) the complete coding sequence of the gene was compiled (Figure 4.11).

4.5.2 Mutations identified by direct sequencing of the NYX gene

A total of 7 different mutations segregating with disease were identified in the NYX gene in 10 different families (Table 4.7).
Figure 4.11: Coding and amino acid sequence of the NYX gene including exon/intron boundaries (exons 2 and 3). The amino acid sequence homologous to the leucine-rich repeats is highlighted in green. The signal sequence is underlined in blue and the GPI anchor is underlined in red.
Table 4.7: Mutations identified in the NYX gene in ten CSNBX families

<table>
<thead>
<tr>
<th>Family</th>
<th>Exon</th>
<th>DNA Position*</th>
<th>DNA Change</th>
<th>Amino acid no.</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>RP3810</td>
<td>2</td>
<td>37+1</td>
<td>G→C</td>
<td></td>
<td>Splicing</td>
</tr>
<tr>
<td>CDT</td>
<td>3a</td>
<td>48-64</td>
<td>17-bp deletion</td>
<td>17</td>
<td>Frameshift</td>
</tr>
<tr>
<td>RRP3665, RP3623**</td>
<td>3a</td>
<td>339-353</td>
<td>15-bp deletion</td>
<td>114-118</td>
<td>ELRLAadel</td>
</tr>
<tr>
<td>CCD, CRP**</td>
<td>3b</td>
<td>647</td>
<td>A→G</td>
<td>216</td>
<td>Asn→Ser</td>
</tr>
<tr>
<td>RP4126</td>
<td>3c</td>
<td>895</td>
<td>C→T</td>
<td>299</td>
<td>Gln→Stop</td>
</tr>
<tr>
<td>RP2124, RP4914</td>
<td>3c-d</td>
<td>1122-1457</td>
<td>335-bp deletion</td>
<td>374-481</td>
<td>Truncated</td>
</tr>
<tr>
<td>RP4121</td>
<td>3d</td>
<td>1309</td>
<td>delC</td>
<td>437</td>
<td>Frameshift</td>
</tr>
</tbody>
</table>

*According to Figure 4.11

**Mutations already described (Pusch et al. 2000, Bech-Hansen et al. 2000)

A splice site mutation, G>C, was detected in family RP3810 at position +1 in intron 2 (Figure 4.12a). The first nucleotide of a donor splice sequence is generally 100% conserved (Horowitz and Krainer 1994), hence the transition identified at this site is very likely to affect RNA splicing. The change introduces a restriction site for SphI restriction enzyme, which was used to show segregation of the mutation with disease in the family and absence of the same mutation in 100 unrelated control chromosomes (data not shown).

A 17 bp deletion was found in patient CDT (Figure 4.12b), spanning nucleotides 48 to 64. This deletion produces a frameshift, predicted to create 91 novel amino acids and a premature stop codon.

An in-frame 15 bp deletion corresponding to nucleotides 339 to 353 was identified in families RP3665 and RP3623 (Figure 4.12c), resulting in the loss of 5 amino acids (ELRLA) and a smaller protein (477 amino acids). This mutation is predicted to alter the structure of the protein and has also been identified in one other family (Pusch et al. 2000).

One missense mutation A>G at nucleotide position 647 was identified in 2 patients, CCD and CRP (Figure 4.12d), resulting in an Asn to Ser change at amino acid position 216. This mutation was also identified in one family from Canada (Bech-Hansen et al. 2000) and appears to involve a conserved residue within the LRR repeats of nyctalopin.

In family RP4126 a nonsense mutation C>T at nucleotide position 895 (Figure 4.12e) was detected, resulting in protein truncation at amino acid 299 with loss of the 11th LRR, the C-terminal cysteine rich LRR and GPI-anchor.
Figure 4.12: Electropherograms depicting patient mutations and normal sequences of the NYX gene. Arrows or diagonal bars indicate the sites of mutations.

a) A G>C change at the first nucleotide downstream of the exon2/intron2 boundary in an affected male from family RP3810.
b) A 17 bp deletion in exon 3a in patient CDT
c) A 15 bp deletion in exon 3a in an affected male from family RP3665.
d) An A>G change in exon 3b in patient CCD
e) A C>T change in exon 3c in an affected male from family RP4126.
f) A 1-bp deletion (C) in exon 3d in an affected male from family RP4121 (Non-coding strand sequence is shown).
A 335 bp deletion corresponding to nucleotides 1122 to 1457 (11 nucleotides inside the 3' untranslated region), was detected in families RP2124 and RP4914 and results in a truncated protein with 107 amino acids missing, causing disruption of the C-terminal cysteine rich LRR and loss of the GPI-anchor. This mutation has also been identified by Pusch et al. (2000) in a family previously described in the laboratory (Hardcastle et al. 1997). Segregation of this deletion with disease in family RP2124 is shown in figure 4.13.

Loss of the GPI-anchor is also the predicted result of a 1 bp deletion in family RP4121 at nucleotide position 1309 (Figure 4.12f) and therefore emphasises the crucial role of the GPI-anchor for nyctalopin function.

Interestingly, no polymorphisms were detected for NYX, perhaps due to the structural constraints of the protein.

To evaluate the possibility of a founder effect among the families sharing the same mutations (families RP2124, RP4914 and the original family described in Hardcastle et al. (1997, family RP2427), families RP3665 and RP3623 and families CCD and CRP) microsatellite markers were amplified to create a haplotype around the NYX gene. Markers DXS574, DXS993, DXS8012 and DXS1201 were amplified in affected males from the families, to reveal a common haplotype in families RP2124, RP4914 and RP2427 suggesting that the 335bp deletion of NYX occurred as a single event in a common ancestor. The other families analysed appear to be unrelated as no common haplotype was observed in families with the same mutation. Table 4.8 is a representation of the haplotypes generated with the four markers and the respective NYX mutation for each of the families analysed. The shaded blocks represent the common haplotype found in the three families with the 355 bp deletion. Previous NYX gene screens also revealed recurring mutations, most notably a 24 bp deletion identified in 7 families originating from the USA with a common founder (Bech-Hansen et al. 2000).

<table>
<thead>
<tr>
<th></th>
<th>RP2124</th>
<th>RP4914</th>
<th>RP2427</th>
<th>RP3665</th>
<th>RP3623</th>
<th>CD</th>
<th>RP</th>
</tr>
</thead>
<tbody>
<tr>
<td>DXS574</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>DXS993</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>DXS8012</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>NYX</td>
<td>Δ335</td>
<td>Δ335</td>
<td>Δ335</td>
<td>Δ15</td>
<td>Δ15</td>
<td>A216G</td>
<td>A216G</td>
</tr>
<tr>
<td>DXS1201</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 4.8: Haplotypes generated around the NYX gene
Figure 4.13: Agarose gel showing segregation of the 355bp deletion at nucleotide position 1122 to 1457 of the NYX gene in family RP2124. The pedigree of the family is aligned to correspond to the respective gel lanes.
+ represents positive control in a control DNA.
- represents negative control in no DNA.
4.6 Screening of RPGR and RP2 in CSNB families

A set of 10 CSNBX families were screened for mutations in both the RPGR and RP2 genes to evaluate the possibility of allelism between RP and CSNB on the X chromosome. The screening was carried out for both genes by direct sequencing of exons 1 to 19 of RPGR and of the 5 coding exons of RP2. No mutations were identified in either of the two genes.

4.7 Discussion

In Figure 4.14a-b-c an ideogram of each of the 3 genes screened for mutations in this study (RPGR, RP2, and NYX) is represented, together with all the mutations (represented with different symbols according to different mutations) reported to date for each gene by all groups. Differences include type of mutations; for example RPGR has all types of mutation, whilst RP2 has predominantly truncation mutations (either splicing or frameshift or nonsense mutations, with fewer missense mutations) and in NYX the majority of mutations are missense, with fewer frameshift and nonsense mutations. Moreover for the NYX gene several mutations have been identified in more than one family, while this situation is relatively rare for the other two genes. It is also striking that for all of the three genes the majority of mutations lie within the respective regions of homology with other proteins (RCC1 for RPGR, cofactor C for RP2 and LRR domains for NYX). The possible implications of the number and type of mutations identified in the three genes will be discussed in turn.

4.7.1 RPGR

Linkage and haplotype analysis of XLRP families has shown segregation of the disease with the RP3 region in 60-90% of cases (Teague et al. 1994, Meindl et al. 1996), while the reported percentage of mutations in the RPGR gene prior to the identification of the newly identified exon ORF15 (Vervoort et al. 2000) ranges between 10% and 30% (Meindl et al. 1996, Roepman et al. 1996a, Buraczynska et al. 1997). In this study, mutation screening conducted on 38 RP3 genotype pedigrees has confirmed this data (section 4.3). A total of 10 different putative mutations in 11 families were identified in this subset of families, defining a mutation rate in RPGR of 29% (11/38 assuming that these families are unrelated).
Figure 4.14: Ideogram of the genomic structure of the (a) RPGR (b) RP2 and (c) NYX genes as described in figure 4.1, showing above each diagram mutations identified so far in each of the three genes. ★ missense mutations, × nonsense mutations, ▼ frameshifts, ▲ splice site mutations, – in frame deletions, ▽ in frame insertions.

a) The RPGR gene

b) The RP2 gene

c) The NYX gene
Mutation screening of the RPGR, RP2 and NYX genes

Ten of the eleven mutations reported here (from a pool of predominantly British families) are all different and novel suggesting a high new mutation rate for the RPGR gene and no founder effect, and supporting the assertion of Fujita et al (1996) that most RPGR mutations are of independent origin.

It is interesting to observe that in previous reports describing RPGR mutations, exon 1 was often excluded from the analysis, due to amplification difficulties. However in this study, amplification of this exon was successful and three potential mutations were identified out of a total of 59 mutations reported for the gene to date (Meindl et al. 1996, Roepman et al. 1996a, Buraczynska et al. 1997, Fujita et al. 1997, Miano et al. 1999, Zito et al. 1999, Zito et al. 2000a, Sharon et al. 2000, this study), excluding exon ORF15. The three exon 1 mutations are all within the 5' consensus donor splice site and two are identical. In addition one of the exon 1 mutation presented here has been recently described in another family (Sharon et al. 2000). Considering that exon 1 comprises less than 2% of the entire coding region of the gene, this donor splice site may be considered a potential mutational hot spot of the gene.

Of interest is the fact that with the three mutations identified in this study in exon 8, the total number of different mutations described for this exon is 11 (Figure 4.14a), out of a total of 59 different mutations identified to date in exons 1 to 19 of the gene (Meindl et al 1996, Roepman et al. 1996a Buraczynska et al. 1997, Miano et al. 1999, Sharon et al 2000, Zito et al. 2000a). This suggests that exon 8 has an important role in the aetiology of retinitis pigmentosa especially considering that three of these mutations are missense mutations pinpointing at least three crucial amino acids contained within this exon.

Most of the mutations reported so far lie within the region homologous to RCC1 (particularly exons 7 and 8, see Figure 4.14a), which is obviously a major functional part of the protein, however several mutations have been localised outside the RCC1-homologous region, and biochemical characterisation of these mutations should result in understanding other important domains of the protein related to its function, its involvement in protein-protein interaction or DNA binding sites.

When this study was carried out the new open reading frame of the gene (ORF15) which seems to account for the majority of mutations (Vervoort et al 2000) had not been discovered and the low mutation rate generally detected in the RPGR gene could have had
Mutation screening of the RPGR, RP2 and NYX genes

other explanations such as mutations in the promoter region, microheterogeneity at the RP3 locus, allelism with other retinopathies, or inefficiency of the detection method.

These issues appear to have been resolved as the majority of mutations appear to be localised within the new exon. Indeed it is now necessary to analyse exon ORF15 in the genetically well defined families for which mutations have not been found, to confirm the hypothesis that the majority of families for which cause of disease is still unknown are accounted for by mutations in this exon.

Inefficiency of the detection method should always be taken into account especially when comparing the data obtained with the first set of RPGR screening (section 4.3.1) which was carried out by SSCP analysis and yielded a mutation rate of 18% (5/27), with the second set (section 4.3.2) carried out by direct sequencing for which the mutation rate was 50% (5/10).

A complex haplotype comprising seven different sequence alterations has been identified (section 4.3.2, Zito et al. 2000b). This haplotype has no apparent effect on the resulting phenotype, as it is present in the general population, but generally highlights the specific characteristics of the RPGR gene. The cSNPs forming the complex haplotype occur outside the RCC1-like domain; hence it is not possible to predict their significance, but confirms the highly polymorphic nature of RPGR.

Complete or nearly complete association between polymorphisms has been frequently observed (Cambien et al. 1999, Behague et al. 1996). Such associations suggest the existence of ancestral haplotypes with potentially beneficial functional consequences. It is possible that each sequence alteration arose as an independent event, and the complex haplotype became fixed in the population either as a result of neutral drift or, if functional, by positive selection. This haplotype may have been introduced in the population under study as a founder effect.

This study emphasises the importance of obtaining a more complete picture of the allelic variation within a gene, suggests a cautious interpretation of phenotypic association with a single SNP, and highlights the potential problems associated with interpreting genetic studies. This is emphasised by the description of a missense mutation in the XLRS1 gene in 8 different families (Sauer et al. 1997, the Retinoschisis consortium, 1998), which was not found in a total of 190 control chromosomes screened by the two groups. This putative
mutation was subsequently re-evaluated as non disease-causing, since it was found in randomly selected control individuals of different ethnic backgrounds (Hiroaka et al. 1999).

4.7.2 RP2

Through linkage analysis, the RP2 locus has been shown to segregate with disease in approximately 20-25% of XLRP families (Teague et al. 1994). The results of a mutation screen in the RP2 gene carried out on a total of 54 XLRP families (section 4.4) has led to the identification of 6 different mutations in 8 families which gives a total mutation rate of 15% (8/54), not dissimilar from the linkage data.

The results of the mutation rate are quite different depending on the mutation detection method used. In section 4.4.1 where direct sequencing was used to screen the gene the analysis gave a final mutation rate for RP2 of 18% (6/33) in contrast to the 9.5% (2/21) obtained using SSCP analysis (section 4.4.2).

To date, no mutations have been identified in exon 5 (Figure 4.14b) suggesting that the C-terminal region encoded by this exon may not significantly affect protein function, and only one mutation has been described for exon 3 (described in this chapter, Table 4.5). Two polymorphisms were detected, which alter amino acid residues in both exons 3 and 5, Arg282Trp and Asp338Tyr respectively (Table 4.6) suggesting that non-conservative changes at less critical or conserved sites can be tolerated by the RP2 protein without pathogenic effect.

In contrast a high density of mutations are localised within exon 2 with 11 different mutations identified (Figure 4.14b), out of a total of 20 different mutations described for the gene to date (Schwahn et al. 1998, Hardcastle et al. 1999, Mears et al. 1999, Thiselton et al. 2000, Wada et al. 2000, Sharon et al. 2000). Most of the remaining mutations are protein truncations or frameshifts in exon 1, which also lead to a disrupted translation of exon 2. Considering that the homology with cofactor C has been identified within exon 2, these results highlight the probable functional importance of this homology domain.

To date the only RP2 mutation for which a functional significance has been demonstrated is a Ser6 deletion in exon 1 that has been shown to disrupt the correct cellular targeting of the protein (Chapple et al 2000).
Mutation screening of the RPGR, RP2 and NYX genes

Until a full understanding of the function of the protein is unravelled, the significance of the distribution of the mutations found in the RP2 gene will remain solely speculative. Interestingly, the spectrum of mutations has a trend toward severe protein truncation, with only 3 different missense mutations identified out of a total of 20 different mutations described (Figure 4.14b). This suggests a critical role for these residues, namely Cys86, Arg118 (mutated in three apparently unrelated families) and Leu253 (Schwahn et al. 1998, Hardcastle et al. 1999, Sharon et al. 2000 and Wada et al. 2000), on protein function, protein folding or protein-protein interaction. Indeed, although the RP2 gene is widely expressed, the exclusive retinal defect resulting from mutations in this gene may be due to one interacting partner specific to normal retina function and the mutation of any particular residue of RP2 directly involved in this presumed interaction would be crucial for the correct functioning of the retina.

The data presented show that out of a total of 59 XLRP families were available in the laboratory, the majority of which were screened for the 2 known XLRP genes, mutations were found in only 20. Twelve of these mutations were detected in the RPGR gene and 8 in the RP2 gene (sections 4.3 and 4.4). For the remaining 39 families disease is still unaccounted for. Since most of these families are linked to the RP3 locus, it is likely that they will display a mutation in the new exon ORF15 of RPGR.

The cloning of both the RPGR and RP2 genes has been a significant step towards the understanding of the genetics of XLRP, describing the number and variety of gene mutations is also important for genotype-phenotype correlation studies. In turn, these studies may help provide insight into the function of the proteins and physiological consequences of different mutations and may provide a possibility of phenotypically distinguishing these two types of XLRP. This has positive implications for the molecular diagnosis of patients and hopefully this research will aid genetic counselling in the clinic.

4.7.3 NYX

The NYX gene was very recently identified by two different groups (Pusch et al. 2000, Bech-Hansen et al. 2000) as the gene responsible for the complete form of CSNBX on Xp11.4. Both groups have also presented several different mutations identified in the gene in
Mutation screening of the RPGR, RP2 and NYX genes

their patient pool (represented in Figure 4.14c together with the mutations described in this study). Among the mutations described in this study (Table 4.7) three have been previously reported, namely the 15 bp deletion at position 339 to 353 and the 335 bp deletion at position 1122 to 1457 were previously described by Pusch et al., and the A>G change at position 647 was described by Bech-Hansen et al. Out of a total of 33 different mutations described for the gene to date 13 have been identified in more than one family, accounting for 39% of the mutations. This suggests the presence of different ancient mutations that may have been introduced in the population as a founder effect. By comparing these data with those presented for the RPGR and RP2 genes (Figure 4.14a-b-c) it is noticeable that the number of same mutations reported in different families for this gene is considerably higher than that for the other two genes, respectively 6 (10%) and 2 (10%). This may be explained by the fact that CSNB patients have a milder phenotype compared to the XLRP patients and therefore does not present such a debilitating disorder. Demonstration of the existence of a founder effect has been shown for one of the mutations identified, a 355bp deletion (Table 4.8).

Another feature of mutations identified in NYX is that almost 50% of the mutations described (16/33) are missense mutations. This type of mutation may have such a deleterious effect on nyctalopin because the LRR domains of the protein form a very particular tertiary structure (Pusch et al. 2000) that may be disrupted by amino acid substitution, which consequently would affect the correct function of the protein.

No polymorphisms in the NYX gene have been observed. This is also probably due to the complex tertiary structure of the protein that may not tolerate amino acid change.

Because the number of protein truncation mutation in the NYX gene is relatively low, accounting for 18% of mutations (6/33), it can also be speculated that more serious mutations may lead in some cases to a more serious phenotype. Therefore, as it was assumed that milder mutations in RPGR and RP2 could lead to CSNB, although no mutations were found in the CSNB patients as described, it may be equally interesting to screen XLRP patients for mutations in the NYX gene.

From a pool of 21 families screened for mutations in the NYX gene 7 mutations were identified in ten different patients. No mutations were identified in any of the 8 sporadic patients, suggesting that they either have a mutation in the CACNA1F gene or the disease maps to the autosomes. This work was done in collaboration with Dr. L. Allen in Cambridge,
Mutation screening of the RPGR, RP2 and NYX genes

who screened the same X-linked CSNB families for mutations in the CACNAIF gene (excluding families GC4180 and RP2124). This collaborative study led Dr. Allen to identify a correlation between genotype and phenotype in the CSNBX families, by electrophysiological and psychophysical testing of at least one affected member of each pedigree. The scotopic (rod derived) responses tended to be of lower amplitude in the subjects with NYX gene mutations than those with CACNAIF mutations, while the photopic (cone derived) responses tended to be more attenuated in the CACNAIF mutation group. Oscillatory potentials (derived from various cell types in the inner retina) were absent in all the subjects with NYX gene mutations but recordable in all subjects with CACNAIF mutations, which may prove to be a distinguishing feature of the two phenotypes.

These results indicate that mutations in the NYX and CACNAIF genes account for almost all X-linked CSNB studied, with NYX mutations being more prevalent. In three families no mutations were detected in either gene using methods described. Family GC4180 was not included in Dr. Allen screen of CACNAIF, suggesting that this may be the causative gene for disease in this family. Although haplotype data shows segregation of disease with both CSNBX loci in the family (section 3.3.2), the pedigree was too small to obtain a significant lod score, and the possibility of an autosomal form of the disease segregating in this family cannot be excluded. Two families (RP4611 and RP1002) presented large X-linked pedigrees and they were screened for mutations in both genes. The phenotype of family RP4611 fits exactly with that described for patients with a NYX mutation, therefore it is likely that this family suffers from a NYX mutation either in the undefined promoter region of this gene or deep within the introns, since mutations screening for both genes was negative. Family RP1002 has not been clinically evaluated and therefore in the absence of mutation detection no such prediction can be made.

This study provides the first evidence towards a correlation of visual function with mutations identified in both genes, confirming presence of a different phenotype associated with genotype. Both protein products evidently have an undefined essential function in the retina, with a similar clinical outcome in the event of mutation, however predicted functions reveal a potentially different role for each protein. Future functional studies of these proteins may reveal their specific roles in the neural retina and also highlight the mechanisms by
which they cause a stationary disease primarily affecting the rod-mediated response to low light levels.
CHAPTER 5

Towards the identification of novel X-linked retina disease genes

5.1 Introduction

Although not yet fully demonstrated, genetic heterogeneity for retina eye diseases is much greater than initially expected and more focused research of new families and their corresponding disease intervals should help to define and refine new loci associated with disease. Examination of these genomic intervals physically will provide new candidate genes for disease, posing the basis for a positional cloning approach. Once the disease interval has been refined as much as possible using available families and genetic marker resources, the next step toward positionally cloning the gene is to isolate DNA and characterise genes from the critical region. Physical maps arrayed with yeast artificial chromosomes (YACs) provide long range-continuous coverage, generating a resource for the fine mapping of regional markers (sequence tagged sites (STTs) and microsatellites) and expressed sequence tags (ESTs). This resource can then be used to isolate further polymorphic markers for disease locus refinement, and novel candidate genes.

5.1.1 Physical mapping and positional cloning

The construction of a physical map involves the assembly of overlapping clones (contig) that would faithfully represent the genomic region of interest in a linear order. The contig can then be used to map pre-characterised genes to the disease interval, or to isolate and characterise novel transcripts through one of several techniques (section 1.6; e.g. detection of CpG islands, Lindsay and Bird 1987, exon trapping, Duyk et al. 1990, cDNA selection, Pamiro et al. 1991, YAC representation hybridisation, Roepman et al. 1996b). This strategy is called positional cloning as it allows gene identification, without prior functional information, based solely on the basis of its map position.
5.1.1.1 Yeast Artificial Chromosomes (YACs)

Yeast Artificial Chromosomes (YACs) are generally created by ligating large fragments of genomic DNA that have been partially digested with restriction enzymes (usually EcoRI), between two vector arms (Schlessinger 1990). The cloning vector contains indispensable chromosomal elements, namely two telomeric sequences positioned at both arms termini for the complete replication of linear molecules, a centromeric sequence required for mitotic segregation and a replication origin for autonomous replication. Selectable markers to stabilise the YAC in the yeast host are also present. The vector sequence totals approximately 10 Kb.

The pYAC4 vector used in the construction of the main YAC libraries used in the UK (CEPH, ICRF and ICI) contains a multicloning site of the bacterial plasmid pBR322. This site resides within an intron of the SUP4 gene, a suppressor tRNA gene that overcomes the effect of the ade-2 ochre mutation present in the yeast host cell used in these libraries. The ade-2 mutation is involved in adenine metabolism and results in the accumulation of a red pigment when propagated in the presence of adenine. The SUP4 gene product, overcoming the effect of the ade-2 mutation, restores wild-type activity, resulting in colourless colonies. Cloning of a foreign DNA fragment into the SUP4 gene causes disruption of the gene function, resulting in the mutant (red colour) phenotype.

5.1.1.2 Human YAC libraries

Different groups have carried out several modifications to the original YAC vector establishing a number of human YAC libraries in which the human genome is represented. The libraries vary in the representation of the human genome and in the lengths of the clone inserts.

The ICI library (Anand et al. 1990) consists of 34,500 clones with an average insert size of 350 Kb, it has over 3.5 x coverage of the human genome and was made from a 48XXXXX cell line. The library is available as 40 primary pools each containing 864 YAC clones. It is suggested that approximately 10% of the clones are chimeric.

The ICRF library (Larin et al. 1991) consists of 20,500 clones with average insert size of 400-600 Kb and combines three separate libraries made from cell lines 48XXXXX,
Towards the identification of novel X-linked retina disease genes

49XYYYY and 46XX. The library is available as 27 primary pools of YAC clones and has an overall chimerism of ~30%.

The CEPH mega YAC library (Albertsen et al. 1990) retains the largest inserts ranging from 100 Kb to 2 Mb, with an average size of 600 Kb. It comprises approximately 95,000 clones and a 7-8 x coverage of the human genome. The chimerism rate has been reported to be about 40%.

5.1.1.3 Contig assembly

To identify clones with overlapping inserts a variety of methods have been developed, such as STS content mapping, fingerprinting and chromosome walking (section 1.5.2.2).

5.1.2 The phenomenon of X inactivation

X chromosome inactivation is the process by which dosage compensation of X-linked gene products between XY males and XX females is achieved. In mammals this phenomenon results from the transcriptional silencing of one X chromosome in each female somatic cell (Lyon 1961). The mechanism of X inactivation, which occurs early in embryonic development, involves different steps, including counting the number of X chromosomes in a cell, randomly choosing all but one X chromosome to be inactivated per diploid cell, and initiating and spreading the inactivation signal in cis throughout the chromosome (Lyon 1961). Because the phenomenon of choosing which X chromosome will be inactivated in each cell is generally random, in a female embryo some cells will inactivate the paternal X and some the maternal X. Once this initial choice is made, it is remembered, meaning that the daughter cells inactivate the same X as the mother cell (Monk 1986). An adult female will therefore result in a mosaic of clones derived from different embryonic cells with all cells in the clone having the same inactive X. In a female carrier of an X-linked disease the pattern of clones with different inactive Xs can have major implications. The process of X-inactivation is under the control of a unique region called the X-inactivation centre (Xic). The Xic is a complex locus that controls the initiation and propagation of the inactivation process in cis. It also ensures that the correct number of X chromosomes undergo inactivation and influences which X chromosome becomes inactivated (reviewed in Willard 1996). Two genes at Xic
have been shown to be involved in the early steps of X inactivation, *Xist* (Borsani et al. 1991) and *Tsix* (Lee et al. 1999).

It has been demonstrated however that a number of X-linked genes escape X inactivation and are expressed in both the active (Xa) and inactive (Xi) X chromosomes. Genes contained in the pseudoautosomal regions, which have counterparts on the Y chromosome, are among the ones escaping X inactivation, as are the majority of other genes that have Y chromosome homologues and that appear to cluster together in regions of the short arm of the X chromosome. However genes without Y chromosome homologues and genes on Xq have also been shown to escape X inactivation (reviewed in Disteche 1995). It is now clear that genes that escape X inactivation are not necessarily clustered together as originally hypothesised, but are also interspersed with genes subject to inactivation along the chromosome (Carrel et al. 1999).

It is of general interest to evaluate the X inactivation status of a gene for the clinical assessment of female carriers of an X-linked disorder, and also for understanding the chromosomal mechanism underlying the phenomenon of X inactivation.

### 5.1.3 The genomic region of Xp11.4

The CSNB1 locus for the complete form of CSNB was assigned to a 5 cM interval in the Xp11.4 region between markers DXS556 and DXS228 (Bergen et al. 1995, Boycott et al. 1998, Hardcastle et al. 1997 and Rozzo et al. 1999). A gene responsible for cone-rod dystrophy (Seymour et al. 1998) is also contained within the same region.

The majority of the Xp11.4 region was part of the sequencing efforts conducted by the Sanger Centre therefore numerous data were available on the database particularly through GDB and IXDB.
5.1.4 The genomic region of Xp22

Because many diseases have been mapped to Xp22, this region has been extensively studied by many different groups (Biancalana et al. 1994, Ferrero et al. 1995, Wapenaar et al. 1993, Wapenaar et al. 1994, Walpole et al. 1997, Warneke-Wittstock et al. 1998) and a number of maps have been developed. Among the diseases described for this interval are at least 4 eye diseases including Ocular Albinism, Retinoschisis, Nance-Horan Syndrome (nuclear cataract, microcornea, hutchinsonian teeth) and Microphthalmia with linear skin defects Syndrome (Meindl et al. 1993, van de Vosse et al. 1996, Bergen et al. 1994, Wapenaar et al. 1993). The genes for two of these diseases, Ocular Albinism and Retinoschisis, have been identified (Bassi et al. 1995, Sauer et al. 1997).

5.2 Materials and methods

5.2.1 Clinical assessment

Family MGRP3 is an American family that was kindly donated by Dr. M. B. Gorin. By pedigree analysis an X-linked form of RP (male family members more severely affected than female carriers and no male-to-male transmission) was established (RP23). Ophthalmological examinations in the family included fundoscopy, visual field assessment, electroretinograms (ERGs) and dark adaptation. DNA was extracted from venous blood samples according to methods described in section 2.1.1.

5.2.2 Computer-aided analysis

Computer-aided analyses including Blast analysis and Nix analysis were carried out as described in sections 2.8.2 and 2.8.4. Several of the electronic databases presented in section 2.10 were used to collect data towards the construction of the CSNB1 and the RP23 physical maps.

5.2.3 YAC library screening by PCR

PCR screening of YAC libraries generated encompassing the CSNB1 and the RP23 regions was carried out as described in section 2.6.6.

Yeast DNA preparations were carried out as described in section 2.1.3.
Towards the identification of novel X-linked retina disease genes

5.2.4 Haplotype and linkage analysis

Microsatellite analysis was performed as described in section 2.6.4 and the allele size differences were resolved by polyacrylamide gel electrophoresis (section 2.4.2) in order to construct haplotypes for family MGRP3. Linkage analysis was carried out using the MLINK program as described in section 2.8.1.

5.2.5 Mutation screening

Eight different genes (XLRS1, PPEF, GLRA2, CLCN4, TBL1, APXL, CALB3, RAI2) were screened for mutations in family MGRP3 by PCR and direct sequencing as described in sections 2.6.1 and 2.7 respectively.

Primers used to amplify and sequence the XLRS1 gene were as described in Sauer et al. (1997).

Primers and conditions used to amplify PPEF were as described (van de Vosse et al. 1997) for the majority of exons. Some exons failed to amplify, due predominantly to errors in the published primers sequence, hence they were redesigned as follows (Ta: 60°C for all):

3F: 5'-GCACCTACTTCTCCTAACAGG-3'
3R: 5'-CTTCATTCTTTCAATAACAAAGG-3'
4F: 5'-TCGGGTGAACCAACCAACGC-3'
4R: 5'-CCATCTTCAGAATTTACCGCTC-3'
11F: 5'-TATGTGCAGATAACAGAGG-3'
11R: 5'-TAGTGTCATTGTTGCCACTGCA-3'
12F: 5'-GAAAGATATATCTCCTCAGTGAAG-3'
12R: 5'-TTAGGGAAAGGGATGTTTATGGG-3'

To amplify GLRA2, primers and conditions used were as described in Cummings et al. 1998 with the exception of exon 4 for which the following new primers were designed (Ta: 60°C):

4F: 5'-CCAAAGCTGTATCAAATCAGGG-3'
4R: 5'-ATCCTACCCTGCAGAATCAGG-3'

Primers spanning the intron/exon boundaries of the 11 exons of the CLCN4 gene were designed from genomic sequence (Acc No: AC002364 and AC003666) and used for PCR and sequencing reactions. Primer sequences, PCR product sizes for each fragment and annealing temperatures (Ta) are shown in Table 5.1.
Towards the identification of novel X-linked retina disease genes

Table 5.1: Primer pairs for amplification and sequencing of the CLCN4 gene

<table>
<thead>
<tr>
<th>Exon no.</th>
<th>Product size (bp)</th>
<th>Forward primer (5'→3')</th>
<th>Reverse primer (5'→3')</th>
<th>Ta (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>215</td>
<td>TATTCCTTCTGATGGCGCATGG</td>
<td>TTGAATGAACACATGGGCTTAGC</td>
<td>60</td>
</tr>
<tr>
<td>2</td>
<td>237</td>
<td>CCGCCGCTGTGCTCATTTC</td>
<td>TCCCGAGAGGAGATGGGAAAGC</td>
<td>60</td>
</tr>
<tr>
<td>3</td>
<td>317</td>
<td>CCCAGGGAACGACATGGGAGC</td>
<td>TTTTGAAGGCTTCGGCCTTCG</td>
<td>60</td>
</tr>
<tr>
<td>4</td>
<td>259</td>
<td>GGAAGATGGTGATGGGCTTG</td>
<td>GACAGTCTACAGAAAGAATTG</td>
<td>60</td>
</tr>
<tr>
<td>5</td>
<td>351</td>
<td>CAGAGGAAATATAGCTCTTTC</td>
<td>CAAGTCAGACAGCAAGGAGG</td>
<td>60</td>
</tr>
<tr>
<td>6</td>
<td>231</td>
<td>CTCCTCTGCTGCTTCTTCG</td>
<td>CTGGATTTCAACCTACAGGG</td>
<td>60</td>
</tr>
<tr>
<td>7a</td>
<td>381</td>
<td>GGGCTTGGGCTCTGAGCAGC</td>
<td>GATTGGGATGGCAAATGAGT</td>
<td>62</td>
</tr>
<tr>
<td>7b</td>
<td>369</td>
<td>GCAAGACGACAGCGCTGGG</td>
<td>GGTTGCTCCGCCAGAAGG</td>
<td>62</td>
</tr>
<tr>
<td>8</td>
<td>312</td>
<td>TCTTGGGAGGCTGGTAAAGC</td>
<td>TCCCTGAAGGACCATGTTCC</td>
<td>60</td>
</tr>
<tr>
<td>9a</td>
<td>329</td>
<td>CACCTCAGTCATCTGAGC</td>
<td>GGGTGACGCAGCACAGTCG</td>
<td>62</td>
</tr>
<tr>
<td>9b</td>
<td>323</td>
<td>GGCACGACATCCATATGGA</td>
<td>ATGGTAGGGATCAAGAGGAT</td>
<td>62</td>
</tr>
<tr>
<td>10</td>
<td>344</td>
<td>AGTTGGGAGGCTGGAAATGG</td>
<td>ATGTCTACATCATCAATCTTC</td>
<td>62</td>
</tr>
<tr>
<td>11</td>
<td>274</td>
<td>TTAGGGAGGAGGAGTGTCG</td>
<td>AAGCTTCTTCTGCTGTCAG</td>
<td>60</td>
</tr>
</tbody>
</table>

Table 5.2: Primer pairs for amplification and sequencing of the TBLI gene

<table>
<thead>
<tr>
<th>Exon no.</th>
<th>Product size (bp)</th>
<th>Forward primer (5'→3')</th>
<th>Reverse primer (5'→3')</th>
<th>Ta (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>262</td>
<td>ACAACAAATAGGGATATAGTGCA</td>
<td>CTATAGGCTATGGAGAAGG</td>
<td>60</td>
</tr>
<tr>
<td>3</td>
<td>250</td>
<td>CTAAGGAACGGAAGGATGGTG</td>
<td>GCATTTCTGCTTCCAAACAG</td>
<td>60</td>
</tr>
<tr>
<td>4</td>
<td>269</td>
<td>GGTGAAACGAAATGACAGAG</td>
<td>ACCCGGCGACATGCAAC</td>
<td>58</td>
</tr>
<tr>
<td>5</td>
<td>270</td>
<td>AATGATTTAATTATCTTCTCA</td>
<td>ATCTCTTCCTGTTGAAGATGC</td>
<td>60</td>
</tr>
<tr>
<td>6</td>
<td>305</td>
<td>ACCCTCAATCTCCTAATGCTG</td>
<td>CATGTAGTCCTCCAAATGTC</td>
<td>58</td>
</tr>
<tr>
<td>7</td>
<td>378</td>
<td>TCCCTGCGCTCTCTCTGCCAG</td>
<td>GCGAAGAAAGAACAAATGAA</td>
<td>60</td>
</tr>
<tr>
<td>8</td>
<td>240</td>
<td>GTGTCTCTGCTGTATGGTAA</td>
<td>CAGCTTCTGCTCCCATCTC</td>
<td>58</td>
</tr>
<tr>
<td>9</td>
<td>291</td>
<td>GTGGAGAGGCGTCTTATCCCA</td>
<td>CACCTTGCTCTATACG</td>
<td>60</td>
</tr>
<tr>
<td>10</td>
<td>178</td>
<td>GGTGCGCTGGCTCTCTAAGTG</td>
<td>CACTCAAGGATTAAATGTCG</td>
<td>62</td>
</tr>
<tr>
<td>11</td>
<td>245</td>
<td>GATGAGGAGGACCATGCAGAT</td>
<td>CGACTTCCAGAAACACAGA</td>
<td>56</td>
</tr>
<tr>
<td>12</td>
<td>209</td>
<td>TTATGATATAATGGTTCCAGA</td>
<td>CACTGGGCTCGACTGAAT</td>
<td>60</td>
</tr>
<tr>
<td>13</td>
<td>273</td>
<td>TCTCTGATGCGCTGGATCTTG</td>
<td>CACACAGAAGACGCTCAG</td>
<td>60</td>
</tr>
<tr>
<td>14</td>
<td>214</td>
<td>GAAGTTGTGACCTACCTACTC</td>
<td>GTGACGCTGAGACTGGCAT</td>
<td>56</td>
</tr>
<tr>
<td>15</td>
<td>252</td>
<td>GAATGATTTAATCGCTTCCC</td>
<td>TTTCTTGGTGGCTTTTGTAAC</td>
<td>56</td>
</tr>
<tr>
<td>16</td>
<td>276</td>
<td>TGAAACCCAGCTGTGTATG</td>
<td>TGGAGTCTGGCATTAGTTG</td>
<td>58</td>
</tr>
<tr>
<td>17</td>
<td>233</td>
<td>GAGCGCAGCATCTCAGACCT</td>
<td>CGTAGAGCTCATTAGG</td>
<td>56</td>
</tr>
<tr>
<td>18</td>
<td>175</td>
<td>GCCGCTAGCTTGCGAATCTCTC</td>
<td>GAGCTGCAACCCCTACATTC</td>
<td>64</td>
</tr>
</tbody>
</table>

Primers spanning the intron/exon boundaries of the 17 coding exons of the TBLI gene were designed from genomic sequence (Acc No: AC003036) and used for PCR and sequencing reactions. Primer sequences, PCR product sizes for each fragment and annealing temperatures (Ta) are shown in Table 5.2. Primers amplifying exon 6 of the TBLI gene were used for YAC library screening.

Table 5.2: Primer pairs for amplification and sequencing of the TBLI gene

<table>
<thead>
<tr>
<th>Exon no.</th>
<th>Product size (bp)</th>
<th>Forward primer (5'→3')</th>
<th>Reverse primer (5'→3')</th>
<th>Ta (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>262</td>
<td>ACAACAAATAGGGATATAGTGCA</td>
<td>CTATAGGCTATGGAGAAGG</td>
<td>60</td>
</tr>
<tr>
<td>3</td>
<td>250</td>
<td>CTAAGGAACGGAAGGATGGTG</td>
<td>GCATTTCTGCTTCCAAACAG</td>
<td>60</td>
</tr>
<tr>
<td>4</td>
<td>269</td>
<td>GGTGAAACGAAATGACAGAG</td>
<td>ACCCGGCGACATGCAAC</td>
<td>58</td>
</tr>
<tr>
<td>5</td>
<td>270</td>
<td>AATGATTTAATTATCTTCTCA</td>
<td>ATCTCTTCCTGTTGAAGATGC</td>
<td>60</td>
</tr>
<tr>
<td>6</td>
<td>305</td>
<td>ACCCTCAATCTCCTAATGCTG</td>
<td>CATGTAGTCCTCCAAATGTC</td>
<td>58</td>
</tr>
<tr>
<td>7</td>
<td>378</td>
<td>TCCCTGCGCTCTCTCTGCCAG</td>
<td>GCGAAGAAAGAACAAATGAA</td>
<td>60</td>
</tr>
<tr>
<td>8</td>
<td>240</td>
<td>GTGTCTCTGCTGTATGGTAA</td>
<td>CAGCTTCTGCTCCCATCTC</td>
<td>58</td>
</tr>
<tr>
<td>9</td>
<td>291</td>
<td>GTGGAGAGGCGTCTTATCCCA</td>
<td>CACCTTGCTCTATACG</td>
<td>60</td>
</tr>
<tr>
<td>10</td>
<td>178</td>
<td>GGTGCGCTGGCTCTCTAAGTG</td>
<td>CACTCAAGGATTAAATGTCG</td>
<td>62</td>
</tr>
<tr>
<td>11</td>
<td>245</td>
<td>GATGAGGAGGACCATGCAGAT</td>
<td>CGACTTCCAGAAACACAGA</td>
<td>56</td>
</tr>
<tr>
<td>12</td>
<td>209</td>
<td>TTATGATATAATGGTTCCAGA</td>
<td>CACTGGGCTCGACTGAAT</td>
<td>60</td>
</tr>
<tr>
<td>13</td>
<td>273</td>
<td>TCTCTGATGCGCTGGATCTTG</td>
<td>CACACAGAAGACGCTCAG</td>
<td>60</td>
</tr>
<tr>
<td>14</td>
<td>214</td>
<td>GAAGTTGTGACCTACCTACTC</td>
<td>GTGACGCTGAGACTGGCAT</td>
<td>56</td>
</tr>
<tr>
<td>15</td>
<td>252</td>
<td>GAATGATTTAATCGCTTCCC</td>
<td>TTTCTTGGTGGCTTTTGTAAC</td>
<td>56</td>
</tr>
<tr>
<td>16</td>
<td>276</td>
<td>TGAAACCCAGCTGTGTATG</td>
<td>TGGAGTCTGGCATTAGTTG</td>
<td>58</td>
</tr>
<tr>
<td>17</td>
<td>233</td>
<td>GAGCGCAGCATCTCAGACCT</td>
<td>CGTAGAGCTCATTAGG</td>
<td>56</td>
</tr>
<tr>
<td>18</td>
<td>175</td>
<td>GCCGCTAGCTTGCGAATCTCTC</td>
<td>GAGCTGCAACCCCTACATTC</td>
<td>64</td>
</tr>
</tbody>
</table>

Primers used to screen the APXL genes were as described by Schiaffino et al. (1995).

186
Towards the identification of novel X-linked retina disease genes

Primers amplifying the 3 coding exons of the \textit{CALB3} gene were designed from genomic sequence (Acc No: L13042) prior to gene screening in the family. Table 5.3 shows primer sequences, PCR product sizes and annealing temperatures used to amplify and sequence each exon.

Table 5.3: Primer pairs for amplification and sequencing of the \textit{CALB3} gene

<table>
<thead>
<tr>
<th>Exon no.</th>
<th>Product size (bp)</th>
<th>Forward primer (5'-3')</th>
<th>Reverse primer (5'-3')</th>
<th>Ta (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>197</td>
<td>GGAAATTTTCATATCAAGGTTGG</td>
<td>TATCTTCACACCATCTAAAGG</td>
<td>60</td>
</tr>
<tr>
<td>2</td>
<td>392</td>
<td>CGTGCCTGGGCTAAGATGCT</td>
<td>CCCTCCACCATGCTCAGC</td>
<td>60</td>
</tr>
<tr>
<td>3</td>
<td>300</td>
<td>AGTAACTGATTCATTTTTTGTTGG</td>
<td>AATAATAAGTCAACCTCACAGCC</td>
<td>60</td>
</tr>
</tbody>
</table>

Primers used to amplify and sequence the \textit{RAI2} gene were as described in Walpole \textit{et al.} (1999a-b), and paired up as described in table 5.4.

Table 5.4: Primer pairs for amplification and sequencing of \textit{RAI2}

<table>
<thead>
<tr>
<th>Fragment</th>
<th>Forward primer</th>
<th>Reverse primer</th>
<th>Product size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>389A20-P21</td>
<td>389A20-P14</td>
<td>412 bp</td>
</tr>
<tr>
<td>2</td>
<td>389A20-P17</td>
<td>389A20-P27</td>
<td>443 bp</td>
</tr>
<tr>
<td>3</td>
<td>389A20-P22</td>
<td>389A20-P23</td>
<td>364 bp</td>
</tr>
<tr>
<td>4</td>
<td>389A20-P20</td>
<td>389A20-P16</td>
<td>179 bp</td>
</tr>
<tr>
<td>5</td>
<td>389A20-P11</td>
<td>389A20-P25</td>
<td>414 bp</td>
</tr>
<tr>
<td>6</td>
<td>389A20-P19</td>
<td>389A20-P15</td>
<td>282 bp</td>
</tr>
<tr>
<td>7</td>
<td>389A20-P10</td>
<td>389A20-P28</td>
<td>340 bp</td>
</tr>
<tr>
<td>8</td>
<td>389A20-P18</td>
<td>389A20-P9</td>
<td>127 bp</td>
</tr>
</tbody>
</table>

Primers amplifying the last few nucleotides and the 3' untranslated region of the gene were designed from the published sequence (Walpole \textit{et al.} 1999a) (Ta: 60°C, product size: 388 bp):

9F: \textit{5'-TAAAGTTAAGAAAGTGGAAGTCCC-3'}
9R: \textit{5'-CACTTTTGGAATGAAAATAGC-3'}

and a new Forward primer was designed to amplify fragment 2 due to an error in the published sequence in the middle of the primer sequence (Ta: 60°C):

2F: \textit{5'-TCCCTGGTTAAGAAACGTGG-3'}
Towards the identification of novel X-linked retina disease genes

5.2.6 EST characterisation

cDNA and IMAGE clones were propagated and DNA isolated as described in section 2.1.2. Vector primers used to sequence or to amplify the inserts are as follows:

M13 F 5'-GTAAAACGACGGCCAGT-3'
M13 R 5'-GGAAACAGCTATGACCATG-3'
λgt10F 5'-AGCAAGTTCCGCTGTTAAG-3'
λgt10R 5'-CTTATGAGTATTTCCTTCCAGGTA-3'

Primers used to amplify EST stsg30779 are as follows (Ta: 54°C, product size: 133 bp):

stsG30779.a 5'-TGTGTTTTCTGCGGGAGGAGAAG-3'
stsG30779.b 5'-TCAGGGCTGCTTTCTTCATC-3'

Primers used to sequence clone 609319 are as follows (positions of each primer are shown in Figure 5.7):

F1 5'-TTGCTTTCTGGACGGAGAAG-3'
F2 5'-TGAAAAGAATTCTAAACCAACATG-3'
F3 5'-GCAGAACAACACTTGGGAGG-3'
F4 5'-AATGGCCTTGTGCTTGGTAGG-3'
F5 5'-AACCTTTTCATCTCAGCTAC-3'
F6 5'-AGGACAGTTTCCAGAATCTAGC-3'
F7 5'-AGGAAACATCTTGAGCCACC-3'
R1 5'-TGATATAAGGGAAGGAGGACGC-3'
R2 5'-AACACTGGAGACTGAATAAGACC-3'
R3 5'-GGAGATTAGTTGGCTTGGACC-3'
R4 5'-GAAATGGAATCTTCTTACAGTGCAC-3'
R5 5'-GATGTCTATTTACACAAACC-3'
R6 5'-ATTGCCGCTTTCAGCAGACG-3'
R7 5'-TAAAGAGAGATCTGTTTTAATCCC-3'

Primers used to sequence clone 30615 are as follows:

F1 5'-TTAATTAGGTACCAGTGGAAT-3'
R1 5'-GAAACAAAGTGGCTCAAAAGGG-3'

Primers used to sequence clone 23809 are as follows:

F1 5'-GGCTCAGACAAACCCCATGG-3'
R1 5'-TTTCTCATCTTTCCAGCTTACG-3'

Primers used to amplify clone 23809 and EST W26143 are as follows (positions of each primer are shown in Figure 5.8 and annealing temperatures varied according to how the
Towards the identification of novel X-linked retina disease genes

primers were coupled; 23809iF+23809iR product size: 324 bp, W26143iF+W26143iR product size: 224 bp):

23809iF 5'-TAGAGCAGTTGTTCCTACTAGG-3'
23809iR 5'-GTAGGCTCATCTCCACAGCC-3'
23809iLH 5'-GTGGTTTCTACTGACCTTTGG-3'
23809iRH 5'-TCAGCCAGATAGAAAGAAATGC-3'
W26143iF 5'-AGAAGACACACACTCTCTCTCC-3'
W26143iR 5'-GATTTAAGCCAGAGGAATCC-3'
W26143iLH 5'-TTTGTTGTTGTTGTTCTGCG-3'
W26143iRH 5'-CGCAGAAAACCACAACCAAC-3'

Primers used to amplify from the 3' end of the CLCN4 gene are as follows (Annealing temperatures varied according to how the primers were coupled):

CL-UTR I 5'-GAGATTTGCTGTATTGCACTGAG-3'
CL-UTR II 5'-TAGGCCACACTAGTGTGGGA-3'
CL-UTR III 5'-TGATATCCACTAAGACAGATTG-3'
CL 11i 5'-AGGATGTCTGAGACATATGGC-3'
CL 10i 5'-GAGATCTCTCCGAAACTGGG-3'
CL 9i 5'-ACCACACTAACGGCTTTCC-3'

Primers used to amplify EST AA663455 are as follows (Ta: 60°C, product size: 197 bp):

663455F 5'-ATTGAAATAAGTGTAGTGTGC-3'
663455R 5'-TGTGTTAACAGAGACTGCAG-3'

Primers used to amplify EST H86729 are as follows (Ta: 60°C, product size: 258 bp):

86729F 5'-AGCAGGGTGCACAGGGGTTC-3'
86729R 5'-AAAGAGAGGCCTTCTCTCAT-3'

Primers used to amplify EST AA019388 are as follows (Ta: 60°C, product size: 199 bp):

019388F 5'-ATGCCACATGCAGGTCC-3'
019388R 5'-TTGGAGGGCTTTCTCCTC-3'

Primers used to amplify EST AA179886 are as follows (Ta: 60°C, product size: 276 bp):

179886F 5'-AACACTCTATAAGATGTTGCTTGC-3'
179886R 5'-CCATCAGGGTCTGTCAGATG-3'

Primers used to amplify EST AA481766 are as follows (Ta: 60°C, product size: 271 bp):
Towards the identification of novel X-linked retina disease genes

481766F  5'-AAAGAAGCCAGACACAAAAGACC-3'
481766R  5'-AAATTCACGTAAACGATGTAATACC-3'

Primers used to amplify EST H83780 are as follows (Ta: 60°C, product size: 289 bp):
83780F  5'-GTTTCGATTGAGCCATGTTTGG-3'
83780R  5'-GTAAGGGGAAGCCACTGTGC-3'

5.2.7 X inactivation studies

X inactivation studies were initiated in house, and in collaboration with Professor Huntington Willard’s Lab (Case Western Reserve University, Cleveland) several cell lines were examined. Preliminary studies involved RT-PCR experiments on human and rodent fibroblast cell lines.

Total RNA from human fibroblast cell lines (293T) was isolated as described in sections 2.5.2.1 and 2.5.2.2, and RT-PCR was performed as described in section 2.6.5.

Rodent fibroblast cDNA, R12: Swiss 3T3 Fibroblast (Log phase) and R13: Swiss 3T3 Fibroblast (Stationary phase), was obtained through HGMP (full length cDNA libraries: http://www.hgmp.mrc.ac.uk/Biology/descriptions/specialcdna_list.html); 1μl of each was directly used for PCR experiments, R12 (Log phase) cDNA provided more consistent results than R13 cDNA. PCR experiments were carried out as described in section 2.6.1.

Control primers for the rodent fibroblast cDNA were designed from the mouse γ-actin 3’end cDNA sequence (EST J0056G08) and were as follows (Ta: 60°C, product size: 261 bp):
M-ACTG-F  5'-AGGTGTGGATGCAAAGCTCC-3'
M-ACTG-R  5'-CACGCTGATGATGAAAGCC-3'

Primers for the PGM (Phosphoglucomutase-1) gene were used as control primers for the human fibroblast cDNA. Conditions used to amplify the gene were as follows: denaturation at 94°C for 3 minutes, 35 cycles of denaturation at 94°C for 30 seconds, annealing at 57°C for 20 seconds and extension at 72°C for 20 seconds, followed by a final extension at 72°C for 5 minutes. A product of 417 bp is expected from cDNA and a larger product is expected from genomic DNA, since primers were designed to span an intron.

Primer sequences were as follows:
PGM-F  5'-GAAAAATCAAGCCATTGTTGGA-3'
PGM-R  5'-GGCACCGAGTTTCACAGAAG-3'
Primers used to amplify the $GLRA2$ gene were as follows (Ta: 59°C, product size: 280 bp, intron spanning):

\begin{align*}
cGLRA2-F & \quad 5'\text{-CCATGGACTTGAAGAACTTTCC-3'} \\
cGLRA2-R & \quad 5'\text{-AGGCTTGGGATGTACATCTGG-3'} \\
\end{align*}

Primers used to amplify the $PPEF$ gene were as follows (Ta: 59°C, product size: 280 bp, intron spanning):

\begin{align*}
cPPEF-F & \quad 5'\text{-CAGACTTGAATTTACTCCACCG-3'} \\
cPPEF-R & \quad 5'\text{-TCTGGTCCAAAATAGCAGCCC-3'} \\
\end{align*}

Primers used to amplify the $APXL$ gene were as follows (Ta: 60°C, 309 bp, intron spanning):

\begin{align*}
cAPXL-F & \quad 5'\text{-CCTGCTGTCGTGCCTCAGGC-3'} \\
cAPXL-R & \quad 5'\text{-TTCAGCTGCTTTCACCAAGG-3'} \\
\end{align*}

Primers used to amplify the $CALB3$ gene were as follows (Ta: 59°C, product size: 229 bp):

\begin{align*}
cCALB-F & \quad 5'\text{-AGATGATCTCTTCAAGAACTGG-3'} \\
cCALB3-R & \quad 5'\text{-AATAATAACTGAACCTCACAGCC-3'} \\
\end{align*}

An IMAGE clone containing the $GLRA2$ gene (79705) was obtained through HGMP. For the $PPEF$ gene no IMAGE clones were available, Professor Jeremy Nathans (whose group identified the $PPEF$ gene, Sherman et al. 1997) was contacted and he kindly provided a pBluescript plasmid containing their longest hPPEF cDNA. The plasmid was shipped on a square piece of Whatman 3 MM paper as a drop of 1 μl of 10 mM Tris, pH 7.6, containing 500 ng of plasmid. For plasmid recovery the circle containing the plasmid solution was cut out and rehydrated for 5 minutes in an eppendorf tube with the addition of 50 μl of 10 mM Tris, pH 7.6 as described by Rosman and Miller (1990). After brief centrifugation, 1, 2, 5 and 8 μl of the supernatant liquid was used to transform competent DH5α cells by adding each amount of plasmid solution to 100 μl of DH5α bacterial cells. The mixture was incubated on ice for 30 minutes, heated at 42°C for 2 minutes and quenched on ice for 1 minute. 200 μl of LB-broth (section 2.9.2) was then added and the mixture was shaken at 37°C for 1 hour. 150
μl of the transformation solution was plated on LB plates containing Ampicillin (section 2.9.2) followed by incubation at 37°C overnight.

5.3 Results

5.3.1 The CSNB1 locus

5.3.1.1 Construction of a physical map and STS content mapping

The first step towards the identification of a gene is the construction of a physical map of the region linked with the disease. For this purpose current databases were scanned (particularly IXDB and GDB) for all possible YACs, PACs and BACs contained within the 5 cM CSNB1 disease interval between markers DXS556 and DXS228 (Figure 5.1). By comparing mapping data from several different databases (including NCBI, Whitehead, Genethon, and Sanger Centre; see section 2.10 for URL addresses), a provisional map was constructed comprising 25 YAC clones.

The next step was to construct, from this preliminary data, an in-house map by obtaining the relevant YACs, and through PCR analysis of all the STSs known to map in this interval confirming clone coverage of the region. In this way the order of markers in this region was established (Figure 5.2). The in-house map consisted of 13 YAC clones covering the region between markers DXS574 and DXS228, the remaining 12 clones were excluded because they were consistently shown to be negative with all the markers tested (probably due to internal deletions or re-arrangements). No data were available on the database for the interval between markers DXS556 and DXS574 therefore this region remains to be covered. To bridge the gap chromosome walking, including isolation of DNA near the end of clone 897e5 (Figure 5.2) for use as a probe for identifying overlapping clones, would need to be performed. This work was not undertaken because at that time other groups working on the same region were close to cloning the gene responsible for CSNB1.

5.3.1.2 Genes and ESTs located in the CSNB1 region

A database search of all ESTs (expressed sequence tags) mapping to the region was also accomplished. The next priority was to map these ESTs within the contig and to evaluate which could represent a good candidate gene. This data was shared with colleagues who were close to cloning the disease gene (Dr A. Meindl and colleagues). The CSNB1 gene has
Figure 5.1: Ideogram of the human X-chromosome showing an exploded view of the CSNB1 region. Map locations of the RPGR and RP2 genes are also shown.
Towards the identification of novel X-linked retina disease genes

Figure 5.2: Physical map of the CSNB1 region. YACs are shown horizontally in purple. Markers are shown vertically in green (microsatellites) or in pink (STSs). Dots represent positive scores. Map location of the NYX gene is shown on the right of the map (between markers DXS8012 and DXS7512, Push et al. 2000).
Towards the identification of novel X-linked retina disease genes

recently been successfully isolated (Bech-Hansen et al. 2000, Pusch et al. 2000, section 4.1.3), and the physical map position of this gene is shown in Figure 5.2.

5.3.2 The RP23 locus

Haplotype and linkage analysis constantly performed on all the families received in the laboratory enabled to identify in an American family (MGRP3) a new locus for XLRP on the short arm of the X chromosome in Xp22.13-22.31, distinct from previously reported gene localisations for XLRP. This new locus was named RP23 (Hardcastle et al. 2000).

5.3.2.1 Phenotype

The phenotype of family MGRP3 presented some unique features compared to classical XLRP. As a main distinction the onset of poor vision in male members of this family appeared to be unusually early, manifesting before the age of two years, and at least one affected male had no recordable ERG response at age 21. In contrast both obligate carriers examined had normal fundi and waveforms. Carrier females normally display different degrees of the disease, ranging from totally asymptomatic to moderately affected, and this is due to the mechanism of random X inactivation. However carrier females generally show at least some electroretinogram (ERG) alterations (Fishman et al 1988). The totally asymptomatic status of the carrier females in family MGRP3 suggests the possibility that the causative gene for this disease may escape X inactivation, as this would ensure that each cell has one functional copy of the gene, regardless of which X chromosome is inactive.

5.3.2.2 Haplotype and linkage analysis

Figure 5.3a shows the pedigree of family MGRP3. Haplotype analysis was performed on the members of the family for which DNA was available using 34 microsatellite markers spanning the entire X chromosome (Figure 5.3b). Initial efforts were concentrated on microsatellites surrounding the RP2 and RP3 region, but it soon became evident that disease in this family was not segregating with these two loci, or with the other XLRP loci reported at the time (RP6 and RP15). Other relevant loci (e.g. CSNB1 and 2 and COD1) and retina specific genes (e.g. X-arrestin and RetGC2) were also excluded (Figure 5.3b).
Towards the identification of novel X-linked retina disease genes

Figure 5.3: a) Pedigree of family MGRP3. DNA was available from individuals shown in pink.

b) Haplotype analysis of family MGRP3. Regions already known to harbor disease genes or retina specific genes are highlighted in the key to the right.
Towards the identification of novel X-linked retina disease genes
Towards the identification of novel X-linked retina disease genes

The distal crossover in this family is defined by individual III:2 (Figure 5.3b), an obligate carrier female, who displays a recombination event with respect to her affected uncle (II:1) between markers DXS1223 and DXS7103. Her affected son, individual IV:4, who has recombined between markers DXS999 and DXS7161, defines the proximal crossover. The minimal region of segregation with disease in this family is localised between markers DXS1223 and DXS7161 in Xp22. From this analysis it also appears that individual IV:6 is a possible carrier, whilst individual IV:8 may not be at risk.

Two-point linkage analysis was also carried out in the family with 8 informative markers on Xp (Table 5.5). Positive lod scores were obtained with markers DXS7103 (Zmax=1.89, θ=0), DXS1224 (Zmax=1.96, θ=0) and DXS999 (Zmax=1.89, θ=0) and not with any of the other markers that, based on haplotype data, appeared not to be linked.

Table 5.5: Two-Point linkage analysis between XLRP and eight microsatellites on Xp

<table>
<thead>
<tr>
<th>θ values</th>
<th>DXS996</th>
<th>DXS1223</th>
<th>DXS7103</th>
<th>DXS1224</th>
<th>DXS999</th>
<th>DXS989</th>
<th>DXS1110</th>
<th>DXS426</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-∞</td>
<td>-∞</td>
<td>1.89</td>
<td>1.96</td>
<td>1.89</td>
<td>-∞</td>
<td>-∞</td>
<td>-∞</td>
</tr>
<tr>
<td>0.05</td>
<td>-0.42</td>
<td>0.87</td>
<td>1.73</td>
<td>1.79</td>
<td>1.73</td>
<td>0.52</td>
<td>-1.82</td>
<td>-2.04</td>
</tr>
<tr>
<td>0.1</td>
<td>0.04</td>
<td>1.01</td>
<td>1.57</td>
<td>1.63</td>
<td>1.57</td>
<td>0.62</td>
<td>-1.24</td>
<td>-1.23</td>
</tr>
<tr>
<td>0.15</td>
<td>0.24</td>
<td>1.01</td>
<td>1.40</td>
<td>1.45</td>
<td>1.40</td>
<td>0.70</td>
<td>-0.91</td>
<td>-0.80</td>
</tr>
<tr>
<td>0.2</td>
<td>0.33</td>
<td>0.95</td>
<td>1.23</td>
<td>1.27</td>
<td>1.23</td>
<td>0.67</td>
<td>-0.68</td>
<td>-0.52</td>
</tr>
<tr>
<td>0.3</td>
<td>0.34</td>
<td>0.72</td>
<td>0.85</td>
<td>0.88</td>
<td>0.85</td>
<td>0.52</td>
<td>-0.37</td>
<td>-0.20</td>
</tr>
<tr>
<td>0.4</td>
<td>0.21</td>
<td>0.40</td>
<td>0.44</td>
<td>0.46</td>
<td>0.44</td>
<td>0.29</td>
<td>-0.16</td>
<td>-0.05</td>
</tr>
</tbody>
</table>

The family under study originates from United States; therefore lod scores were recalculated with European allele frequencies for linked markers DXS7103 (allele 1 at 0.25 and allele 2 at 0.75), DXS1224 (allele 1 at 0.5, allele 2 at 0.2 and allele 3 at 0.3), and DXS999 (allele 1 at 0.6 an allele 2 at 0.4). Resultant lod scores at θ=0 were Zmax=1.97 for DXS7103, Zmax 2.01 for DXS1224 and Zmax=1.85 for DXS999, slightly increasing the chances for linkage at these loci. The possibility that the disease in this family is linked to an autosomal dominant gene with partial penetrance can not be completely ruled out. To test this hypothesis marker D19S785, which lies within the RP11 disease interval on chromosome 19q13.4 (Al-Maghtheh et al. 1994), and has a penetrance of 65% (Al-Maghtheh et al. 1996), was typed in the family. No linkage was observed between the marker and the disease gene in the family, excluding the RP11 locus as the disease interval for RP23. Further work should involve testing linkage with other incomplete penetrance RP loci (RP1 and RP9).
Towards the identification of novel X-linked retina disease genes

5.3.2.3 Physical map

As mentioned in section 5.1.4 the region in Xp22 has been extensively studied by several different groups and a number of different maps have been generated. Data were retrieved from the most comprehensive map (Ferrero et al. 1995) and a minimum tiling path of YAC clones was obtained (from HGMP and CEPH) to build an in-house map. This would serve as a resource to map any candidate genes or new interesting ESTs inside or outside of the region of interest. The in-house physical map is shown in Figure 5.4. The disease interval, contained between markers DXS1223 and DXS7161, is entirely covered with the exception of two gaps. The first gap is between markers DXS1006E and DXS6703 and is due to the fact that clone 966G1 scored negative for marker DXS6703, probably due to a partial deletion of the clone, and because clone 781F11 (not shown in the picture), which was expected to contain the region spanning these two markers, repeatedly failed to amplify even when reordered. The second gap is between markers DXS1317 and DXS999. In this case a large clone, 939H7 (2.2 Kb), expected to span the region including markers DXS418 and DXS443 (beyond the proximal marker in the map) repeatedly failed to amplify. The authors (Ferrero et al.) also stated that the clone undergoes frequent rearrangements. Other clones contained within this microregion gave similar problems or were not available through accessible resources on the databases indicating that this region of the chromosome is unstable. Chromosome walking will need to be carried out in order to bridge these gaps, probably using more stable clones such as PACs, BACs or cosmids. Genes and ESTs mapped in this study are shown in blue on the right of Figure 5.4 (discussed in sections 5.3.2.4 and 5.3.2.5).

5.3.2.4 Evaluation of candidate genes in the RP23 region

Eight genes mapping to the critical interval for RP23 were chosen as candidate genes for disease in this family either on the basis of their expression patterns or based on their proposed functional homologies. All the exons and splice junctions of all genes were PCR-amplified and directly sequenced in an affected member of the family. The eight genes analysed will be presented in turn:
Figure 5.4: Physical map of the RP23 region. YACs are shown horizontally in purple. Markers are shown vertically on the left in green (microsatellites), pink (STSs), or blue (genes or ESTs). Genes or ESTs mapped in this study are shown on the right. Dots represent positive scores.
Towards the identification of novel X-linked retina disease genes

a) XLRSl

The X-linked retinoschisis gene (XLRSl) was isolated in 1997 (Sauer et al.) in a region of the X chromosome contained between markers DXS418 and DXS999, hence within the RP23 region. The predicted protein sequence of retinoschisin has a highly conserved discoidin domain, thought to be implicated in cell-cell adhesion and phospholipid binding, a function that is in agreement with the observed splitting of the retina in retinoschisis patients, indicating that the RS gene is important for retinal development.

The XLRSl gene was shown to be retina specific and in situ hybridisation and immunohistochemistry studies have shown that the gene is expressed only in the photoreceptor layer, but the protein product, retinoschisin, is present both in the photoreceptors and within the inner portions of the retina (Grayson et al. 2000). The authors suggested that retinoschisin is released by photoreceptors, has functions within the inner retinal layers, and that X-linked retinoschisis may be caused by abnormalities in a putative secreted photoreceptor protein.

Confirmation of the gene location was evaluated by PCR screening the physical map, but due to the gap in the map present between the two markers DXS418 and DXS999, a product could not be obtained. However by blast analysis a PAC clone (Z94056) containing the first exon of the XLRSl gene and the PPEF gene was identified, positioning the XLRSl gene approximately 20 Kb distal to the PPEF gene. Although RS is clinically different from RP and the lesions in the RS patients are thought to be related to a defect in retinal Muller cells rather than photoreceptor cells, the XLRSl gene is exclusively expressed in the retina, has been shown to cause a retina disease and is located in the RP23 interval region. The gene was therefore considered a candidate for RP23. The 6 exons and splice junctions of the gene were PCR amplified and directly sequenced in a patient from family MGRP3. No mutations were identified, suggesting that XLRSl is not involved in RP23.

b) PPEF

The PPEF (protein phosphatase with EF calcium-binding domain) gene (Montini et al. 1997) is the human homologue of the Drosophila retina degeneration gene C (rdgC). In Drosophila, mutations in this gene prevent light-induced retinal degeneration (Steele and O'Tousa 1990). The gene has been previously screened and excluded as a candidate for
retinoschisis (van de Vosse et al. 1997). It consists of 17 exons and encodes a serine/threonine protein phosphatase, which is 61.7% homologous to rdgC. The gene is mainly expressed in the brain and basal ganglia and expression has not been detected in the eye, suggesting that PPEF may have a different function from the Drosophila rdgC gene. This does not necessarily exclude PPEF as a candidate gene for RP23, since both the XLRP genes identified to date (RPGR and RP2) have a very low level of expression in the eye (Meindl et al 1996, Schwann et al 1998). PPEF has been shown to be contained in four different YACs including 939H7, which consistently failed to amplify in this study (section 5.3.2.3). YAC clone 718B8 was positive for PPEF, confirming the physical location of this gene within the RP23 critical interval (Figure 5.5a). Direct sequencing of the gene in the RP23 family failed to show any mutation, suggesting that this is not the causative gene for RP23.

c) GLRA2

The GLRA2 gene encodes the α2 subunit of the glycine receptor, which is a multisubunit chloride channel that mediates inhibitory neurotransmission (Betz 1992). Because of its function and its expression in embryonic and neonatal spinal cord and brain it is suggested that mutations in the gene may lead to some neurological disorder (Malosio 1991). Although there was no evidence of retina expression for this gene, during the search for retina genes and ESTs within the RP23 region, two ESTs were identified, which were derived from foetal retina RNA and were identical to GLRA2. As shown in Figure 5.4, the gene was previously mapped in the critical interval for RP23 (Ferrero et al. 1995). Due to its location and its expression pattern, the gene was considered a candidate for disease in this family. No mutations were identified by directly sequencing the 9 exons and intron/exon boundaries of GLRA2, excluding the gene from the list of possible candidates for RP23.
Figure 5.5: Agarose gels showing PCR experiments performed with different transcripts on several YAC clones.

a) PPEF tested on: 1.939H7, 2.718B8, 3.938H6, 4.genomic DNA, and 5.no DNA
b) TBLI tested on: 1.145E5, 2.186H3, 3.210B5, 4.914B9, 5.131D2, 6.A187H5, 7.A118A8, 8.genomic DNA and 9.no DNA
c) stsG30779 tested on: 1.A118A8, 2.184F6, 3.921C1, 4.95B4, 5.genomic DNA and 6.no DNA
d) W26143 tested on: 1.210B5, 2.914B9, 3.131D2, 4.A187H5, 5.A118A8, 6.genomic DNA and 7.no DNA
e) AA663455 tested on: 1.966G11, 2.926D3, 3.810E1, 4.genomic DNA and 5.no DNA
f) AA481766 tested on: 1.764D7, 2.161D11, 3.896F9, 4.939H7, 5.genomic DNA and 6.no DNA
g) H83780 tested on: 1.810E1, 2.764D7, 3.161D11, 4.genomic DNA and 5.no DNA
Towards the identification of novel X-linked retina disease genes

d) CLCN4

The CLCN4 (chloride channel 4) gene is part of a family of voltage-gated chloride channel genes, which share significant sequence and structural similarities. The gene, located in the Xp22.3 region, was cloned in 1994 (van Slegtenhorst et al.) and the cDNA sequence was derived from an adult retina cDNA library clone. The gene was shown to contain at least 10 exons spanning 60 to 80 Kb of genomic sequence, however no genomic structure was provided (van Slegtenhorst et al. 1994). Because of its high level of expression in adult retina, its location within the critical RP23 region (Figure 5.4, Ferrero et al. 1995) and its function (voltage-gated channel have been shown to cause retinal diseases, i.e. the CACNAI1F gene), CLCN4 was an interesting candidate gene for RP23. In order to characterise the intron/exon structure of the gene the cDNA sequence was used to identify two human PAC clone sequences spanning the entire cDNA sequence, AC002364 and AC003666, by blast analysis. By comparing the cDNA sequence with the genomic sequence 11 exons, with splice-sites conforming to the AG-TG rule (underlined in Table 5.6), were detected in clones AC002364 and AC003666, and the intron/exon boundaries were derived (Table 5.6). Intronic primers were designed (Table 5.1) to enable the gene screen in family MGRP3. This was performed by direct sequencing and no mutations were detected, suggesting that CLCN4 is not the disease-causing gene for RP23.

<table>
<thead>
<tr>
<th>Exon number</th>
<th>3' splice site</th>
<th>5' splice site</th>
<th>Size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ATGGTC</td>
<td>AGGAA</td>
<td>144</td>
</tr>
<tr>
<td>2</td>
<td>tctgctttgtctag</td>
<td>TGGCGG</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>tgtactctttccag</td>
<td>TCAGAG</td>
<td>188</td>
</tr>
<tr>
<td>4</td>
<td>gttctctgtctctag</td>
<td>CAGAGG</td>
<td>123</td>
</tr>
<tr>
<td>5</td>
<td>catctctgtttctag</td>
<td>CAGAG</td>
<td>207</td>
</tr>
<tr>
<td>6</td>
<td>ccttttttttttag</td>
<td>GAGAGA</td>
<td>81</td>
</tr>
<tr>
<td>7</td>
<td>acttcttttttcag</td>
<td>ATGAAG</td>
<td>636</td>
</tr>
<tr>
<td>8</td>
<td>ctgcctctctcaag</td>
<td>TCCCGG</td>
<td>187</td>
</tr>
<tr>
<td>9</td>
<td>tctctctctctctag</td>
<td>CAATAA</td>
<td>399</td>
</tr>
<tr>
<td>10</td>
<td>ctacccctctctag</td>
<td>AGAACG</td>
<td>217</td>
</tr>
<tr>
<td>11</td>
<td>tgtctctctctag</td>
<td>GAGACTA</td>
<td></td>
</tr>
</tbody>
</table>
Towards the identification of novel X-linked retina disease genes

e) TBL1

*TBL1* (transducin (β)-like 1) is a novel gene mapping to the Xp22.3 region, which shares significant homology with members of the β-transducin WD40 repeat-containing protein family (Bassi *et al.* 1999). The gene was isolated through cDNA selection from an adult retina cDNA library and it was found to contain 18 exons spanning approximately 150 Kb adjacent to the ocular albinism (*OAl*) gene. Northern blot analysis indicated that the *TBL1* gene is ubiquitously expressed as 2 transcripts of approximately 2.1 and 6.0 Kb.

By genomic analysis, Bassi *et al.* (1999) found that the *TBL1* gene was either partly or entirely deleted in patients carrying Xp22.3 terminal deletions. The complexity of the contiguous gene syndrome though, makes it difficult to assign a specific function to all the genes present in the deletion region. One patient carrying a microinterstitial deletion involving the 3′ portion of both the *TBL1* and *OAI* genes showed the OAI phenotype associated with X-linked late-onset sensorineural deafness (OASD) and this led the authors to postulate an involvement of *TBL1* in the pathogenesis of the late-onset sensorineural deafness phenotype.

However no mutations involving *TBL1* only have been identified to date and it may still be possible that the sensorineural deafness phenotype in this patient is due to pleiotropic effects of *OAI* mutations.

Because of its position in the critical interval for RP23, which was also confirmed in this study through PCR screening in the YAC contig (Figure 5.5b), its expression in the retina and its homology with members of the β-transducin family, the *TBL1* gene was considered a candidate for disease in family MGRP3. The 17 coding exons and splice junctions of the gene (exon 1 is a non-coding exon) were screened, but no mutations were detected. *TBL1* was therefore excluded from the possible candidate genes for RP23.

f) APXL

During the search for the gene involved in *OAl*, a novel human gene sharing homology to the Xenopus laevis APX gene, which is implicated in amiloride-sensitive sodium channel activity, was identified and named *APXL* (APX-like, Schiaffino *et al.*, 1995). The gene contains 10 exons and spans approximately 160 Kb of Xp22.3. The full-length mRNA is approximately 7.5 Kb, and Schiaffino *et al.* (1995) isolated several clones from a
Towards the identification of novel X-linked retina disease genes

retinal cDNA library that corresponded to this mRNA. The authors found that, along with retina, the gene is expressed in melanoma cells, brain, placenta, lung, kidney, and pancreas.

Because of its expression in the retina and its position within the RP23 critical interval (Figure 5.4, Ferrero et al. 1995) the gene was chosen as a candidate for disease. However no mutations were identified by directly sequencing the entire gene, suggesting that this is not the gene causing RP in this family.

g) CALB3

Calbindin D9K, the vitamin D-dependent calcium-binding protein (CaBP9K), was first described in rat intestine. It is a cytosolic protein of molecular weight 9000 and belongs to a family of calcium-binding proteins. This calbindin is found in the mammalian intestine, placenta, uterus, and kidney. Its exact function is unknown. The intestinal protein is vitamin D-dependent and its expression correlates with calcium transport activity. The human homologue of the calbindin D9K, also called CALB3, was cloned in 1992 (Jeung et al., Howard et al.) and was mapped to the short arm of the X chromosome. The deduced protein sequence is homologous to other mammalian calbindins. A single abundant mRNA transcript was detectable in proximal small intestine but not in kidney, uterus, or placenta. Jeung et al. (1994) determined the structure of the CALB3 gene as a small (5.5 Kb) gene comprising 3 exons.

Because of its location in the critical interval (Figure 5.4, Ferrero et al. 1995) and because of its deduced calcium-binding function and the fact that calcium regulation plays a crucial role in the phototransduction pathway, the gene was screened as a possible candidate in the family. However, no mutations were detected, excluding CALB3 as a possible candidate gene for RP23.

h) RAI2

Walpole et al. (1999a) identified a novel human gene in the course of studies of a 1.3 Mb region of Xp22 encompassing the RS region, which they named RAI2 (retinoic acid-induced-2). RAI2 is a single-exon gene, which was found to be expressed in several foetal and adult tissues, including retina. The 530 amino acids protein (57 kD predicted mass) displayed 94% homology with a mouse retinoic acid-induced gene product and contained a
Towards the identification of novel X-linked retina disease genes

novel proline-rich (39%) domain of 68 amino acids. Retinoic acid is involved in vertebrate anteroposterior axis formation and cellular differentiation and has been shown to modulate gene expression controlling early embryonic development, suggesting a developmental role for \textit{RAI2}. The gene was also screened as a candidate for Nance-Horan syndrome (Walpole \textit{et al.} 1999b).

The gene is thought to be contained in the region on the X chromosome between markers DXS418 and DXS999 (Walpole \textit{et al.} 1999a). PCR screening in the physical map gave negative results with all the YACs tested, probably due to the gap present in the map between these two markers.

The \textit{RAI2} gene was screened as a possible candidate gene in the family for its presumed location in the critical interval for RP23 and for its expression in the retina. Failing to find any mutation in the coding region of the gene suggested that \textit{RAI2} is not involved in causing disease in this family.

\textbf{5.3.2.5 Characterisation of retina ESTs in the RP23 region}

Positional candidates in the Xp22 region were identified by searching the UniGene and TIGR databases (see section 1.6.5), where a list of expressed sequence tags (ESTs) that map to specific regions can be obtained. The ESTs are generally organised in clusters, which represent part of the same transcript and may have originated from different tissues. Preliminary expression patterns are therefore often available as well as physical location of the ESTs.

\textbf{a) stsG30779}

The first EST analysed in this study is called \textit{stsG30779}, which, according to UniGene data (Hs.187608), was located on the X chromosome in the interval between the telomere and DXS1061 and whose cDNA source was the eye. Blast analysis of this EST identified a PAC clone (AC002349) containing a portion of Xp22. Just distal to the EST of interest, the PAC sequence contained a gene called HSXPRF, also called \textit{MIDI}, which is a gene involved in Opitz G/BBB syndrome (Quaderi \textit{et al.} 1997). This gene (\textit{MIDI}) was also shown to be located on YAC 184F6, contained in the physical map (Figure 5.4). This result was confirmed through PCR screening, where the EST was shown to be present both on
Towards the identification of novel X-linked retina disease genes

YAC 184F6 and YAC 921C1 (Figure 5.5c). The EST was thus expressed in the eye and was contained in the region of interest. A foetal retina IMAGE clone (609319) was obtained through HGMP, and DNA was amplified with the EST specific primers (section 5.2.6) for confirmation. The clone was then digested with three different enzymes (EcoRI, BamHI and NotI) and electrophoresed on an agarose gel to estimate the length of the cDNA insert (Figure 5.6). The DNA was sized at 8 Kb, which was a linear form of vector (pBluescript, 2958 bp) plus insert; therefore the cDNA insert was an estimated 5 Kb.

The clone was entirely sequenced, initially using vector primers M13F and M13R. By blast analysis, the sequences generated were shown to be identical to sequences on the PAC clone AC002349. Since the length of the insert appeared to be identical to the genomic sequence on the PAC, this entire genomic sequence of 4759 bp was analysed for an open reading frame (ORF), however no obvious ORF was detected.

The entire insert was then sequenced by designing new primers to each sequence generated (section 5.2.6, Figure 5.7) to ensure there were no errors in compiling the published genomic sequence of PAC AC002349 that would conceal an ORF (Figure 5.7). However the entire sequence of the insert appeared to be exactly the same as the genomic sequence suggesting that the IMAGE clone, which indeed did not contain a poly-A tail, was a genomic contaminant not derived from a cDNA source.

b) W26143

The second EST analysed in this study is W26143. According to UniGene data (Hs.293838) this EST was located on Xp22 in PAC clone AC003666, which also contains part of the CLCN4 gene (section 5.3.2.4-d). PCR screening in the physical map was positive for W26143 on YAC A118-A8 (Figure 5.5d). The EST was part of a cluster of ESTs derived from different libraries, and this particular EST (W26143) was from a human retina randomly primed sublibrary. The EST sequence produced several other hits of clones derived from the human retina randomly primed sublibrary on blast analysis. The only IMAGE clone available for this cluster was an infant brain clone (30615), which was obtained through HGMP. The clone was sequenced using M13F and M13R primers plus newly designed internal primers. Analysis of the sequence generated showed chimerism. From the M13R direction the clone (30615) showed overlaps with EST W26143, PAC AC003666 and with
Towards the identification of novel X-linked retina disease genes

**Figure 5.6:** Agarose gel of restriction enzyme analysis of clone 609319 with the following restriction enzymes: 2) EcoRI, 3) BamHI, 4) NotI, with two different size markers: 1) ΦX174-HaeIII, 5) 1Kb ladder. Sizes of products for both markers (in bp) are shown on each side.
Figure 5.7: Sequence of clone 609319. Green boxes highlight primers used to generate the sequence, with arrows indicating the direction of each primer. Reverse primers (−→) are designed to the complementary strand. Primer names are shown on the right.
Towards the identification of novel X-linked retina disease genes
Towards the identification of novel X-linked retina disease genes

an IMAGE clone from a brain library (23809), the sequence of which began where EST W26143 sequence ended (Figure 5.8). From the M13F direction identities were detected with the Homo sapiens mRNA for KIAA0455 protein, which is located on chromosome 1, and again with PAC clone AC003666 and IMAGE clone 23809. IMAGE clone 23809 was obtained through HGMP. The sequence of the entire clone was available on the database (Acc. No. AF052117) and was 1622 bp long containing a poly-A signal and a 30 bp long poly-A tail. M13F and M13R primers plus newly designed internal primers were used to sequence the cDNA clone (23809) and to check the published sequence. The cDNA clone (23809) sequence is identical to the genomic sequence on PAC clone AC003666 and no obvious ORFs were detected, suggesting that the clone is part of the 3' untranslated region of a gene.

PCR experiments were performed on a retina cDNA library with primers W26143iF and iR, and 23809iF and iR (Figure 5.8) to demonstrate retina expression of W26143 and 23809, which proved positive in both cases. PCR, both on retina cDNA and brain cDNA with primers 23809RH and W26143LH was also carried out to demonstrate that clone 23809 and EST W26143 form a contiguous sequence in both tissues, therefore they are part of the same transcript (Figure 5.9a). To evaluate whether an additional untranslated region was present downstream of the poly-A tail of clone 23809, PCR was performed on retina cDNA library with primers 23809LH and λgt10F and λgt10R respectively, however no products were detected. On the other hand PCR with retina and brain cDNA was carried out with primers W26143RH and CL-UTR I, II and III respectively, to demonstrate that the transcript comprising 23809 and W26143 is also connected to the UTR of the CLCN4 gene. Whilst products were obtained for primer W26143RH with primers CL-UTR I and II (Figure 5.9b), which are contained in the UTR described by Kawasaki et al. (1999) (Acc. No AB019432), no product was obtained with primer CL-UTR III, which is contained in the UTR described by van Slegtenhorst et al. (1994) (Acc. No NM_001830). In the latter case genomic DNA also failed to produce an amplification product, suggesting that PCR conditions may need to be modified for this product to be amplified. PCR was also performed on retina and brain cDNA libraries using primer W26143RH with primers CL-11i, CL10i and CL9i contained respectively within exon 11, exon 10 and exon 9 of the CLCN4 gene. This experiment was done to evaluate whether the UTR represented by cDNA clone 23809 and EST W26143 is
Figure 5.8: Schematic representation of positions of IMAGE clone 23809 and EST W26143 relative to PAC clone AC003666, and of IMAGE clone 30615 relative to 23809, W26143, AC003666 and to KIAA0455 on chromosome 1.
* at position 107293 pinpoints the end of the 3'-UTR of the CLCN4 gene according to Kawasaki et al. (1999).
** at position 108652 pinpoints the end of the 3'-UTR of the CLCN4 gene according to van Slegtenhorst et al. (1994).
*** at position 109190 pinpoints the end of the last CLCN4 exon (exon 11).
Primers used for intra- and inter- transcript amplification are represented by arrows with respective primer names.
Direction of M13 primers used to sequence clone 30615 are also represented by arrows.
Figure 5.9: a) Agarose gel of PCR experiments with primers 23809RH and W26143LH. Lane 1=genomic DNA, 2=retina cDNA, 3=brain cDNA, 4=no DNA. Estimation of size of the PCR product is shown on the right of the picture and is based on φX174HaeIII product sizes which are mimicked and represented as horizontal lines on the left of the picture. b) Agarose gel of PCR experiments for primers W26143RH with CL-UTR I (lanes A), CL-UTR II (lanes B) and CL-UTR III (lanes C). 1=retina cDNA, 2=brain cDNA, 3=genomic DNA, with size marker φX174HaeIII.
Towards the identification of novel X-linked retina disease genes

attached to the \( CLCN4 \) gene taking into consideration possible alternative splicing that might occur splicing out one or more of the last exons of \( CLCN4 \) with the respective UTR. However no products were detected with any of the primers tested.

The sequence of PAC clone AC03666, spanning exon 10 and 11 and the 3'-UTR of \( CLCN4 \), the EST W26143 and cDNA clone 23809, was analysed with the NIX program (Figure 5.10) to see whether any of the gene programs detected possible alternative transcripts. Different gene prediction programs found additional potential exons, the most interesting of which are marked with letters (A, B, B', C, D, E and 10B). Using the translation tool of EditSeq in the DNASTAR program the potential exons, all of which have correct 3' and 5' splice sites apart from exon C (GA instead of AG in the 3' splice site) and exon E (5' splice site missing), were connected in all possible combinations and then translated. All but three combinations (exons 10+A+D, exons 10+A+B'+10B+11+C and exons 10+A+B'+10B+11+D) gave a continuous ORF. Because original expression studies of the \( CLCN4 \) gene (van Slegtenhorst \textit{et al.} 1994) revealed the presence of a 7.5 Kb transcript on Northern analysis, whilst the authors only characterised a 3.2 Kb cDNA transcript, it is likely that clone 23809 represents additional 3' untranslated region of the gene, which may be connected to the coding region through potential alternative splicing, possibly including one or more of the predicted exons.

c) AA663455

AA663455 is another retinal EST located on Xp22, which is part of a UniGene cluster of ESTs (Hs.23202) derived from different tissues. The physical location of AA663455 was tested by PCR on the YAC map, where the EST was shown to be located on YAC 926D3 (Figure 5.5e). A stratagene foetal retina clone (853594) containing this EST was obtained through HGMP and sequenced with M13F and M13R primers. By blasting the sequence generated the Homo sapiens m6b1 mRNA was detected. The \( M6B \) gene was isolated by Olinsky \textit{et al.} (1996) and is a member of the proteolipid protein gene family. It is expressed early during development of the central nervous system in both neurons and glial cells and the protein is a neuronal membrane glycoprotein that may have an important role in neural development. It is considered a good candidate for one of the neurological disorders that have been mapped to Xp22.2 (i.e. Rett syndrome, several X-linked mental retardations,
Figure 5.10: Results of the NIX analysis carried out on part of PAC AC03666 sequence spanning exon 10 and 11 of the CLCN4 gene, EST W26143 and clone 23809. The EST, the clone, the CLCN4 exons plus the 3'UTR are indicated by white arrows and writings.
Towards the identification of novel X-linked retina disease genes

Charcot-Marie-Tooth neuropathy), and this led Narayanan et al. (1998) to screen the gene in patients with Rett syndrome.

The seven exons of the gene remain to be screened for mutations in a patient from family MGRP3.

d) H86729

H86729 appeared to be an interesting EST because it was part of a UniGene cluster (Hs.227583) where more than 50% of the clones were from eye. IMAGE clone 223703 was obtained through HGMP and sequenced with vector primers. Blast analysis of the clone sequence identified a human clone within the Xp11.23 region, containing the CACN1A1 gene, hence EST H86729 is not located in the region of interest.

e) AA019388

AA019388 is an eye EST whose IMAGE clone (362780, from adult retina) was obtained through HGMP and sequenced with vector primers. The clone sequence was entirely available on the database (Acc. No AL079294). Blast analysis of the clone sequence resulted in identities with several different chromosomes. By re-blasting against month database (sequence generated in the last month, this was carried out in May 2000), a working draft with unordered pieces of a Homo Sapiens X clone (AC006209) was identified, with no mapping information. This analysis was repeated more recently (February 2001), but no X clones were identified. By using different pieces of the working draft clone AC006209 in blast analysis, one of the fragments was identical to part of the X chromosome located on Xp11.21-11.23, therefore proximal to the RP23 region. Using the electronic PCR (e-PCR, Schuler 1998) feature available at NCBI to test the PAC sequence for the presence of sequence tagged sites, ten markers were detected and seven of these were mapped to Xp22.32, distal to the RP23 interval. PCR screening of this EST in the physical map also gave negative results, although this may be due to the two gaps in the map.

Preliminary data suggest that EST AA019388 is not contained in the critical interval for RP23.
Towards the identification of novel X-linked retina disease genes

f) AA179886, AA481766 and H83780

Another three ESTs were more recently evaluated, AA179886, AA481766 and H83780. Each of the three ESTs is part of a UniGene cluster of ESTs (Hs.28896, Hs.6716 and Hs.65406 respectively) derived from different tissues including eye. Whilst EST AA179886 did not amplify from YACs in the map, both AA481766 and H83780 scored positive (Figure 5.5f-g). Two foetal retina clones, 609395 and 838844 and an adult retina clone, 219536, one for each of the three ESTs respectively, were obtained through HGMP.

Clone 609395 is a 2310 bp cDNA that has been completely sequenced (Acc. No. AL079296). Blast analysis of this entire sequence did not produce any significant identities with genomic regions on the X chromosome. By NIX analysis two potential poly-A signals were detected, and an ORF encoding a minimum of 240 amino acids. This amino acid sequence is part of a potential new protein with no significant homologies with other known proteins, therefore its function is unknown. Although UniGene data place clone 609395 on the X chromosome between the telomere and DXS1061, PCR YAC screening and blast analysis were inconclusive, therefore it still remains to be ascertained whether this clone is contained in the critical interval for RP23.

These three new genes present further candidates for disease and will require characterisation prior to mutation screening, as detailed above.

5.3.2.6 X inactivation studies

As mentioned in section 5.3.2.1, the totally asymptomatic status of the carrier females segregating with the RP23 locus suggests the possibility that the causative gene for this disease may escape X inactivation. Genes and ESTs that were evaluated as possible candidates in this study were therefore analysed for their X inactivation status.

Among the 8 genes screened in this study for mutations in the family (section 5.3.2.4), the TBL1 gene was previously shown to escape X inactivation (Willard Lab Data: http://genetics.gene.cwru.edu/willard/data.htm and Bassi et al. 1999), whilst the RAI2 gene was shown to be subject to X inactivation (Willard Lab Data: http://genetics.gene.cwru.edu/willard/data.htm). In addition, the XLRSl gene was already under investigation in Professor Willard’s lab.
Towards the identification of novel X-linked retina disease genes

Among the ESTs evaluated as possible candidate genes (section 5.3.2.5), AA179886, AA481766 and H83780 were not included in these X inactivation experiments because their identification was subsequent to these investigations.

X inactivation studies were carried out in collaboration with Professor Huntington Willard. The experiments, which involve RT-PCR analysis of each gene in eight to twelve rodent/human somatic fibroblast cell hybrids, each containing a different cytogenetically normal human inactive X (Xi), or active X (Xa) (Carrel et al. 1999), were carried out in Professor Willard’s lab. Preliminary experiments were required, these involved checking positive expression of each gene on human fibroblast cell lines and negative expression in rodent fibroblast cell lines. The results of this expression analysis, through RT-PCR experiments, are shown in Figure 5.11. W92467 (CLCN4), AA663455, stsg30779, AA019388 and H86729 (lanes A, B, C, D and E) all gave positive results when amplified in genomic DNA, human fibroblast cDNA and their respective clones (positive control) and negative results when amplified in rodent fibroblast cDNA. GLRA2 and PPEF (lanes F and G) also gave positive results when amplified in human fibroblast cDNA and their respective clones and negative results in rodent fibroblast cDNA, whilst APXL (lanes H) gave positive results with human fibroblast cDNA and negative results with genomic DNA (primers used were intron-spanning and a longer extension time was needed to visualise the larger product expected) and rodent fibroblast cDNA. CALB3 (lanes I) gave a positive result with genomic DNA, a negative result with rodent fibroblast cDNA and inconsistent results with human fibroblast cDNA, however primers for this gene were included in further analyses by Professor Willard. Lanes J show the results obtained with EST W26143, which is not expressed in human fibroblast cDNA, hence this EST was not tested for inactivation.

Each transcript was examined in 3 Xa and 9 Xi hybrids where inactivation status is preserved. Genes that show expression in no more than 2 Xi hybrids are considered subject to X inactivation, genes that are expressed in at least 7 Xi hybrids are considered to escape inactivation, and those in the middle are thought to show variable or heterogeneous patterns of inactivation. Three of the transcripts were shown to escape inactivation, CALB3, AA019388 and AA663455. GLRA2, stsg30779 and H86729, were shown to be subject to X inactivation. XLRSl, already underway in Willard’s lab, is also likely subject to inactivation (it is expressed in 2/9 Xi hybrids). The remaining three transcripts showed variable results;
Figure 5.11: Agarose gel of PCR experiments for different transcripts on human and mouse fibroblast cDNA. Lanes A) W92467, B) AA663455, C) stsG30779, D) AA019388, E) H86729, F) GLRA2, G) PPEF, H) APXL, I) CALB3, J) W26143. For lanes A, B, C, D, E and J: 1=genomic DNA, 2=human fibroblast cDNA, 3=rodent fibroblast cDNA, 4=corresponding cDNA clone. For lanes F and G: 1=human fibroblast cDNA, 2=rodent fibroblast cDNA, 3=corresponding cDNA clone. For lanes H and I: 1=genomic DNA, 2=human fibroblast cDNA, 3=rodent fibroblast cDNA. Lanes K) controls: 1=rodent fibroblast cDNA amplified with γ-actin mouse primers, 2=genomic DNA amplified with human PGM primers (intron-spanning) 3=human fibroblast cDNA amplified with PGM primers showing no genomic DNA contamination in RNA prep.
Towards the identification of novel X-linked retina disease genes

PPEF was not tested due to very poor expression in hybrids, APXL gave inconclusive results that could not be interpreted and W92467 appeared to be heterogeneous (expressed in 5/9 Xi hybrids). These data will be published on the Willard web page.

5.4 Discussion

In this chapter the relevance of thoroughly investigating genetically each new family received in the laboratory is revealed. Haplotype and linkage analysis carried out in one of the numerous families, MGRP3, has enabled the identification of a new XLRP locus (RP23) on the distal part of the short arm of the X chromosome.

The power of biocomputing is also highlighted and presented, and its integration with laboratory work has huge potential. Genetic and physical maps now available on the databases have enabled construction of physical maps of regions of interest by simply searching the database, thus eliminating many steps such as library screening and end cloning. Also, the increasing number of ESTs present in the database provided a number of potential genes to characterise in the disease region of RP23, eliminating the need of employing techniques such as exon trapping or cDNA selection. This demonstrate how progress made by the Human Genome Mapping Project and computational advances have provided a means to overcome the time-consuming manual construction of a contig and the identification of potential genes. With the future total human genome sequence available it will become easier to analyse DNA sequences, evaluate the clones location on the chromosome, identify new ESTs or IMAGE clones for a region of a chromosome, and compare cDNA and genomic DNA sequences, thus easily identifying exons of a particular gene.

5.4.1 Current state of the Xp11.4 and the Xp22 maps

In a positional cloning approach, the identification of a gene involved in a specific disease is based on the construction of a physical map of the chromosomal region where the disease has been mapped. This strategy has been applied to the two regions on the X chromosome presented in this chapter, the Xp11.4 region where a gene for CSNB (CSNB1) has been mapped (Bergen et al. 1995, Hardcastle et al. 1997), and the Xp22 region where a gene for RP (RP23) has been mapped (Hardcastle et al. 2000).
Towards the identification of novel X-linked retina disease genes

The main purpose of establishing these in-house maps was to provide a resource that would enable the mapping of genes inside or outside each region of interest.

-Xp11.4

The physical map of the region in Xp11.4 (Figure 5.2) consists of 13 overlapping YAC clones covering the region between markers DXS574 and DXS228. The more distal part of the disease region (between markers DXS556 and DXS574), was not covered by clones since data available on the database were not sufficient to identify an overlapping clone. STS-content mapping was also pursued to establish a correct order of the relative YACs and markers in the map. The map was left at these early stages due to collaborative efforts with other groups. Indeed a high-resolution 1.4 Mb BAC/PAC physical map within the critical region for CSNBl in Xp11.4 was published in 2000 (Sparkes et al.) and shortly afterwards two different groups independently cloned the CSNBl gene, which was called NYX (Bech-Hansen et al. 2000 and Pusch et al. 2000, sections 1.1.4.2 and 4.1.3). The order of the markers in the contig generated in this study is in agreement with the order of markers presented in the high-resolution map recently published (Sparkes et al. 2000).

-Xp22

For the Xp22 region a comprehensive physical map had already been published (Ferrero et al. 1995). 23 of the YACs contained in this map and spanning the region between markers DXS1223 and DXS7161, the two flanking markers delimitating the RP23 region in MGRP3 family, were obtained (Figure 5.4). As discussed, the map presents two gaps, the first between markers DXS143 and DXS7103, and the second between markers DXS1317 and DXS999. Chromosome walking techniques and database searches to look for additional YACs or even smaller clones (BACs, PACs or cosmids) are necessary to bridge the gaps. This work is underway. Genes and ESTs evaluated in this study were either previously mapped to the region or were mapped in this study (i.e. TBL1, W26143, stsG30779, AA663455 (M6B1), H83780, AA481766 and PPEF). Some coding sequences could not be mapped onto the YAC contig, highlighting the need to bridge the gaps in order to evaluate if the ESTs that tested negative are indeed not present in Xp22. Ultimately the main purpose of creating an in-house map of this region of the chromosome was to have a resource that
Towards the identification of novel X-linked retina disease genes

enabled validation of the location of each transcript prior to its evaluation as a candidate gene for disease in the RP23 region.

This region of the X chromosome located between markers DXS1223 and DXS7161 is very large, approximately 15 cM and spanning 3 chromosomal bands. Although the Xp22.2 band, which is mostly represented in this region, is a giesma positive dark staining band, associated with being gene poor, parts of the flanking bands (Xp22.31 and Xp22.13) are also represented, and these are giesma negative light staining regions, which tend to be gene rich. Considering that 1 cM equals approximately 0.5-1 Mb of DNA and that it has been proposed that 1 Mb of DNA may contain between 30 and 50 genes, a large number of genes can be expected in the RP23 region. The possibility of identifying the disease-causing gene among such a large number is an extremely challenging task. Currently however this has been the only possible approach, and the most useful candidates were examined. In turn this may prove valuable for any other disease that map to the region. Eight genes were screened for mutations in the family (section 5.3.2.4), which were chosen either for their expression patterns (i.e. retina) or for their functional homologies. Sequence analysis showed no mutations in any of these genes, however chromosomal rearrangements or mutations in intronic sequences (i.e. large insertions or deletions) and in the UTRs can not be excluded and Southern blots need to be prepared to test these possibilities.

It is imperative to continue haplotype analysis of all the new X-linked families that become available to try to identify at least one other family that carries the same form of RP as the MGRP3 family, and that may hopefully help to refine the region. A collection of more members of the same family is also underway, which may provide additional crossovers to refine the region. In this respect the in-house physical map is a useful resource, as new polymorphic markers may need to be generated to refine the disease interval.

5.4.2 EST characterisation

Five ESTs from the Xp22 region, plus another three whose study is at an initial stage, where characterised (section 5.3.2.5). Some of these ESTs appear to be outside the disease interval (e.g. H86729, AA019388 and AA179886). Another EST, AA663455, was part of a gene already characterised (the M6B gene) but had not been previously mapped to Xp22, screening of this gene is now required in the MGRP3 family. The clone containing EST
stsG30779 was probably a genomic contaminant due to lack of introns, poly-A tail and ORFs. The remaining ESTs, AA481766, H83780 and W26143, are potentially worth pursuing. The study of AA481766 and H83780 is underway. Whilst W26143 appeared to be particularly interesting, it is likely that it represents an extension of the 3' UTR of the CLCN4 gene with an additional polyadenylation signal. The CLCN4 transcript described by Kawasaki et al. (Acc. No AB019432) is 4454 bp long and does not contain a polyadenylation signal. In contrast another transcript of the same gene reported by van Slegtenhorst et al. (1994) (Acc. No NM_001830) is 3214 bp long and has a non-canonical polyadenylation signal (ATTAAA) located 21 bp from the poly-A tail. The presumed additional 3'UTR region presented here extends for another 2177 bp downstream of the UTR described by Kawasaki et al. Although the connection between the additional sequence identified in this study with the 3'UTR reported by Kawasaki et al. was made (Figure 5.9b), conclusive evidence require Northern blot analysis. In their paper, van Slegtenhorst et al. reported that their expression studies revealed the presence of a 7.5 Kb transcript by Northern analysis, and they then went on to characterise a 3.2 Kb cDNA clone. An additional 4 Kb of transcript therefore was not characterised in this article. This 4 Kb may be accounted for by the sequence presented by Kawasaki et al. plus the additional sequence including EST W26143 and clone 23809 reported here, which in total amounts to approximately 3.5 Kb. If this is the case, the CLCN4 gene contains two polyadenylation signals in the 3'UTR, and possible alternative transcripts may be identified in future, perhaps with differing functions or expression patterns. Numerous examples of transcription units with multiple poly-A sites (up to seven in the dehydrofolate reductase gene, Frayne et al. 1984) within a single 3'-terminal exon have been described (reviewed in Edwalds-Gilbert et al. 1997). It is not clear whether the use of these multiple sites may be regulated or if they are randomly used. For a large proportion of the genes described with multiple poly-A sites the distribution of the 3'ends generated varies depending on the time in development or tissue in which they are expressed. It has also been demonstrated, for a number of genes, that the different mRNA products have differential stability or translatability and this may explain how differential poly-A sites could influence protein expression.

Interestingly, apart from tandem poly-A sites, potential alternative poly-A site selection can occur in genes with composite internal or terminal exons or with exon skipping.
Towards the identification of novel X-linked retina disease genes

(Edwalds-Gilbert et al. 1997). In the first case the gene has an exon which is a composite of 3' and 5' splice sites followed by a poly-A site, these exons can serve either as internal or as terminal exons based on circumstances. This is not true for the *CLCN4* gene, where the last exon described does not have a 5' splice site. In the second case a gene can have two or more alternative 3' terminal exons encoding different C-termini that can be differentially chosen through exon skipping. This is an interesting possibility that is worth considering for the *CLCN4* gene, highlighted by the NIX analysis results (Figure 5.10), as there are at least 7 additional potential exons which, when connected in different combinations with a computer program, gave rise to a continuous ORF in the majority of cases. RT-PCR experiments are now needed to test this hypothesis and to evaluate which of these, if any, are the real exons or alternative transcripts. Due to the described function and expression pattern of *CLCN4*, the newly identified 3' terminal region of this gene described in this thesis should be screened for mutations in RP23 patients.

5.4.3 X inactivation

The X inactivation status of the majority of genes and ESTs presented in this study was examined. The reason this work was pursued is based on the assumption that if the females of this family are totally asymptomatic, in that they do not show any sign of the disease with ERG examinations, it is possible that this is due to the fact that each cell in their body contains a functional copy of the gene. This could be achieved if the gene escapes X inactivation, hence is always expressed in each cell regardless of which X chromosome is inactive, therefore mutations would not present disease in the traditional recessive description.

The X inactivation studies, which were done in collaboration with Professor H. Willard, showed that among the genes screened *CALB3* and *TBL1* escape X inactivation, *GLRA2* and *XLRSl* are subject to X inactivation and *PPEF*, *APXL* and *CLCN4* gave variable results. Among the ESTs evaluated AA019388 and AA663455 where shown to escape X inactivation whilst stsg30779 and H86729 are subject to X inactivation; W26143 was not tested as no expression was detected in fibroblast cDNA, and AA179886, AA481766 and H83780 were not included as their evaluation started after the X inactivation studies. The results of these studies show a lack of clustering of genes escaping X inactivation.
Towards the identification of novel X-linked retina disease genes

In families affected by retinoschisis, the female carriers are also unaffected (George et al. 1995) and X inactivation studies showed that the gene is subject to X inactivation. However retinoschisin, the retinoschisis gene product, is a putative secreted photoreceptor protein (Grayson et al. 2000). Considering the situation in which a protein is secreted by the cell prior to explicating its function outside the cell, then not all cells need to carry a functional gene in order for its product to function in the tissue. This is another possible explanation of how carrier females of an X-linked disorder can be asymptomatic, that can be applied to the causative gene for RP23. However, knowing the inactivation status of any gene on the X chromosome is of general interest as it provides useful information for clinical genetics and insights into the genomic organisation of the X chromosome. Inactivation profiles of genes on the X chromosome are also useful for understanding the chromosomal mechanisms of inactivation and the role that X inactivation plays in individuals with X aneuploidy and carrier females of X-linked diseases.
CHAPTER 6

RP in association with other phenotypes

6.1 Introduction

6.1.1 RP in syndromic forms

Retinitis pigmentosa is a retinal disorder caused by degeneration of rod photoreceptor cells leading to loss of peripheral vision and night blindness (section 1.1.4.1). Involvement of cone photoreceptor cells at a later stage, leads to total blindness for the majority of patients. The disease can be inherited as an autosomal dominant trait, autosomal recessive or X-linked, and there is also a reported case of digenic RP due to mutations at the peripherin/RDS and ROM1 loci (Kajiwara et al. 1994). In addition retinitis pigmentosa is observed in a number of other conditions in association with several different phenotypes. Among the syndromes reported is Alstrom syndrome, a combination of RP, deafness, obesity, and diabetes mellitus (Alstrom et al. 1959), Refsum syndrome including retinitis pigmentosa, chronic polyneuropathy, and cerebellar signs (Leys et al. 1989), Bardet-Biedl syndrome characterised by mental retardation, pigmented retinopathy, polydactyly, obesity, and hypogenitalism (Schachat et al. 1982), Laurence-Moon syndrome which includes mental retardation, pigmented retinopathy, hypogenitalism, and spastic paraplegia (Laurence and Moon 1866), Cockayne syndrome a condition displaying dwarfism, precociously senile appearance, pigmented retinal degeneration, optic atrophy, deafness, marble epiphyses in some digits, sensitivity to sunlight, and mental retardation (Paddison et al. 1963), and pallidal degeneration where RP is associated with dysarthria and progressive extrapyramidal rigidity (Winkelman 1932).

6.1.1.1 Usher syndrome

The most common syndrome reported in association with RP is Usher syndrome, which is a genetically heterogeneous group of disorders characterised by the combination of retinitis pigmentosa and congenital sensorineural hearing loss (Fishman et al. 1983). Approximately 1 in 25,000 persons are affected by the disease, which can be classified in
three different forms according to the severity of the auditory dysfunction: Type 1 Usher syndrome (OMIM: 276900) is characterised by profound hearing loss, Type 2 (OMIM: 605242) has mild hearing loss and Type 3 (OMIM: 276902) has progressive hearing loss. Along with phenotypic heterogeneity, a wide genetic heterogeneity has also been demonstrated for Usher syndrome, with ten different recessive loci described (RetNet). X-linked inheritance was suggested by Davenport et al. (1978), however autosomal recessive inheritance was considered most likely. 75% of type 1 Usher syndrome patients (Usher 1B type) are accounted for by mutations in the Myosin VIIA gene (Weil et al. 1995). Interestingly it has also been demonstrated that a distinct group of mutations in this gene are responsible for recessive isolated deafness (DFNB2, Weil et al. 1997), demonstrating that Usher 1B and DFNB2 are allelic defects of the same gene. Similarly a missense mutation in the USH2A gene has been found in 4.5% of 224 patients with recessive RP (Rivolta et al. 2000), demonstrating allelism between Usher syndrome type 2A and recessive RP.

Photoreceptors, auditory hair cells, and vestibular hair cells develop from ciliated progenitors. Several lines of evidence suggest that a generalised abnormality of axoneme structure is present in patients with Usher syndrome. Hunter et al. (1986) found a high proportion of abnormal axonemes in retinal photoreceptor cells of a patient with Usher syndrome and Shinkawa and Nadol (1986) found a decrease in outer ciliary cells in the lower part of the cochlea in this syndrome. Structural and functional evidence for abnormal nasal cilia has been found in Usher syndrome and in other patients with reportedly isolated retinitis pigmentosa (Arden and Fox, 1979). Finally, abnormal sperm motility, velocity, and structure have been described in Usher syndrome, a feature probably related to the markedly decreased fertility of these patients (Hunter et al. 1986). Bonneau et al. (1993) reported the association of type 1 Usher syndrome with bronchiectasis, chronic sinusitis and reduced nasal mucociliary clearance in 2 brothers and suggested that USH1 could be a primary ciliary disorder.

6.2 Materials and methods

6.2.1 Clinical assessment

Four subjects from one three-generation pedigree (family RP4462) were recruited to the study and information regarding 1 member of the family living overseas was obtained
RP in association with other phenotypes

from his physician. Clinical and genetic criteria were used to determine X-linked inheritance of retinitis pigmentosa in the family as previously described (section 5.2.1).

Phenotype characterisation was carried out by Dr. Susan Downes (Oxford Eye Hospital) and included an ophthalmic and systemic history, visual field testing and fundus examination. In addition fundus photography and electrophysiological evaluation, including electro-oculographic responses (EOG), full field electroretinography (ERG), and pattern electroretinograms (PERG), were performed. Further particulars regarding general health were obtained from patient's physicians.

6.2.2 Genotype analysis

Haplotype analysis was carried out on family RP4462 with markers along the long and the short arms of the X chromosome as previously described (section 2.6.4 for microsatellite analysis and section 2.4.2 for polyacrylamide gel electrophoresis).

6.2.3 Mutation screening

Direct sequencing of PCR products of the RPGR and RP2 genes was employed for mutation screening in this family. Methods for PCR and direct sequencing are described in sections 2.6.1 and 2.7 respectively.

6.3 Results

6.3.1 Family phenotype

Ophthalmic findings and electrophysiological testing in affected males were within the range of findings described in XLRP, and the 2 carrier females although asymptomatic, had sparse intraretinal bone spicule pigmentation and subnormal amplitudes in flicker ERGs. Both affected males and females had severe recurrent sinus infections and recurrent ear infections from early childhood, continuing into adulthood and leading to deafness requiring hearing aids in males. The three affected males also had recurrent chest infections, again occurring in early childhood and continuing into adulthood, and subject III/1 also had nasal polyps and renal failure.
6.3.2 Family Genotype

Figure 6.1 shows the pedigree of family RP4622 and the relative haplotypes generated for the individuals tested. A double crossover was observed in the affected male III:2 between markers DXS1214 and DXS8090 and between markers DXS556 and DXS988 with respect to his carrier mother (II:2), defining the distal and proximal boundaries of disease segregation in the family. The region of segregation with disease in the family is therefore included between markers DXS1214 and DXS998 spanning both the RPGR and the RP2 genes.

6.3.3 Mutation screening of the RP2 and RPGR genes

The RPGR and the RP2 genes were both screened for mutations in an affected male from family RP4462.

The RP2 gene was entirely sequenced and the only sequence variation identified was a C>T at position 844 in exon 3 (Thiselton et al 2000). The change did not segregate with disease in the family therefore refining the disease interval distal to the RP2 locus.

Sequence analysis of the RPGR gene showed a 2 bp deletion in exon 8 (845-846delTG) (Figure 6.2). This frameshift mutation appeared to segregate with disease in the family. The mutation introduces 19 new amino acids and a premature stop codon, which is predicted to result in a truncated protein of 280 amino acids. All the exons upstream of exon 8 were also entirely sequenced to exclude the possibility of a second sequence alteration that might have a combinatorial effect with the exon 8 mutation.

It is evident therefore that the mutation in RPGR is the cause of retinitis pigmentosa in this family, however the family suffers from additional systemic phenotypes and therefore the question remains, does the mutation in RPGR contribute to the broader phenotype observed or are other genetic and/or environmental factors involved?
**Figure 6.1:** Pedigree of family RP4462 showing haplotypes generated with microsatellite markers on the X chromosome.
**Figure 6.2:** Electropherograms depicting patient mutation and control sequence in exon 8 of the *RPGR* gene.

a) A TG deletion at nucleotide position 845-846 in an affected male from family RP4462

b) Control sequence of the corresponding region
6.4 Discussion

The retinitis pigmentosa described in this family is typical of XLRP with no particularly distinguishing features from other XLRP phenotypes. The female hemizygotes did not have tapetal reflexes, but had abnormalities consistent with X-linked carrier status. In the family, the RP phenotype is also invariably associated with hearing defects and chest and sinus impairment as a result of constant infection, suggesting that disease in this family could be a primary ciliary disorder. A databases search to try to identify a selection of genes mapping to the X-chromosome with a proposed compatible function was therefore employed.

One such gene is filamin on Xq28. This is an actin binding protein and has been localised to the circumferential/microfilament bundle in the retinal pigment epithelium and plays a role in maintaining cell contact and shape (Kodama et al. 1991). Another well-characterised example is caltractin, located between DXS1193 and qter. This is a calcium binding, phosphoprotein recognised as a structural component of the centrosome and is present in mammalian retina particularly in the connecting cilium (Wolfrum 1995). Any of the deafness loci and the RP loci on the X chromosome or a combination of them in a contiguous deletion syndrome could also be considered candidates for the phenotype of the family.

Haplotype data located the disease in the family on the short arm of the X chromosome between markers DXS1214 and DXS1068 therefore immediately excluding genes like caltractin and filamin and the majority of deafness and RP loci.

However the region of segregation with disease was still spanning both the RP2 and the RPGR genes, both genes were therefore screened in the family.

The identification of a polymorphism in the LP2 gene (C844T in exon 3), which did not segregate with disease in the family, enabled further refinement of the disease interval. Sequence analysis of the RPGR gene gave a more surprising result, a 2 bp deletion in exon 8 causing a frameshift and a severe protein truncation, segregating with disease in the family.

The RPGR gene is the major gene accounting for X-linked retinitis pigmentosa. Different mutations in RPGR have been described in many different RP families (section 4.7.1 and Figure 4.14a). Missense, nonsense, frameshift and splicing mutations have all been reported and they all appear to be associated with a classic form of XLRP, where symptoms arise at a very early stage in affected males and progression of disease invariably leads to
blindness, with no apparent association with other phenotypes. The *RPGR* mutation type identified in this family truncates the protein, however more 5' protein truncation mutations have been described as a cause of XLRP without additional phenotypes (Buraczynska et al. 1997, Meindl et al. 1996, Miano et al. 1998, Weleber et al. 1997, Zito et al. 1999, section 4.7.1). Exons 1 to 8 were directly sequenced to search for a second hit, however no additional alterations were identified in family RP4622. This is the first report of a mutation in *RPGR* in a family in which recurrent severe upper respiratory tract infections occur in all affected members, but the level at which the mutation is exerting its effect is not clear.

A thorough examination of the literature provided information that enables discussion of the findings presented in this family.

As already mentioned (section 6.1) retinitis pigmentosa, although mainly affecting the eye has also been frequently found in association with other phenotypes in syndromic forms like Usher syndrome, where sensorineural deafness is present with RP. Independent studies have also reported varying percentages of RP patients with hearing impairments without other syndromic features (Arden and Fox 1979, McDonald et al. 1988, Rosenberg et al. 1997). Among these at least three families (Rosenberg et al. 1997) have been described as X-linked RP (XLRP), including one that was previously reported as having a mutation in the *RPGR* gene (Roepman et al. 1996b). This mutation is described as a 6.4 Kb deletion, which disrupts the 3' end of *RPGR* removing the last 6 exons (downstream of exon 15). Both affected males and carrier females from this family displayed hearing difficulties (Rosenberg et al. 1997), however these features were not discussed in the mutation paper.

In a family with retinitis pigmentosa presumed to be RP3 because of linkage analysis, van Dorp et al. (1992) found that some affected males had recurrent respiratory infections as a result of a condition indistinguishable from the immotile cilia syndrome (OMIM: 242650) where a ciliary dysfunction leads to chronic sinusitis, serous otitis and recurrent episodes of bronchitis. The phenotype of this family resembles the one described in the family presented here, although they give no description of the audiological status of their patients. They raised the possibility that previously observed ciliary abnormalities in XLRP patients may be associated specifically with the RP3 locus mutation. Indeed after the *RPGR* gene was cloned in 1996 (Meindl et al.) a mutation (IVS5+1 G>T) was identified in the gene in this family (Dry et al. 1999), although the authors give no discussion of the previously described
additional phenotype associated with RP in the family. These data strengthen and support the data presented in this chapter, as there is at least one family with an *RPGR* mutation that has RP in association with hearing impairments, another one with a mutation in *RPGR* with RP in association with respiratory infections and there is also the family here described that has an *RPGR* mutation and presents RP in association with both hearing impairments and respiratory infections.

Both eye and ear contain a cilium-like structure respectively in the photoreceptor cells and in the cochlea, hence an association between visual and hearing impairments is not unexpected. Abnormalities of cilia have in fact been reported in X-linked and autosomal types of RP, based on axonemal structure examination in sperm cells or nasal cilia (Arden and Fox, 1979; Hunter *et al.* 1988). The recent subcellular localisation of RPGR in the photoreceptor connecting cilia (Hong *et al.* 2000) also provides strong evidence of a general ciliary role for RPGR and, as the authors pointed out, the possibility that certain alleles of *RPGR* might present with RP in conjunction with an immotile cilia defect. On the other hand a more thorough examination of the clinical phenotype of patients with RP may give more evidence towards additional phenotypes for RP3.

This data together with previous findings, and the recent localisation of RPGR in the connecting cilia of photoreceptor cells, suggest a role for *RPGR* in a more complex phenotype than the classic XLRP possibly connected with ciliary function and involved in hearing impairment.
CHAPTER 7

Overall discussion and future prospects

7.1 Overview of work presented

The genetic and phenotypic heterogeneity of human retinal diseases is a direct consequence of the complexity of vision. The more the development and maintenance of vision is understood, the more we can learn about the genes and regulatory elements that govern it, and vice versa. By studying families with human diseases we can pinpoint the region on a chromosome responsible for the disease phenotype, and eventually isolate the causative gene for disease. Once the gene is identified, tissue expression profiles and functional studies can be pursued to provide a better understanding of the disease mechanism.

The classical strategy for a positional cloning approach relies initially on linkage analysis, which enables low resolution mapping of a disease to a cytogenetic region using family information. Additional refinement and high resolution physical mapping is then required and this is achieved through collection of additional family members or additional pedigrees and with the assembly of the region of interest in a manner that enables further manipulation. The next phase involves isolation and characterisation of novel transcripts and evaluation of candidate genes through mutation screening.

The work presented in this thesis embraces the above-mentioned strategies, ranging from linkage and haplotype analysis of families with retinal disorders and mutation screening of the known genes mapping to the relevant region, to physical mapping and screening of candidate genes in the attempt to unravel the molecular defects underlying two retinal diseases on the X chromosome, XLRP and CSNBX.

Due to the complexity of the genetics of XLRP and CSNBX and lack of conclusive phenotypic definition of different loci, a systematic haplotype analysis of all families is very important as it helps to discriminate between different loci thus providing a clearer picture of the relevance of each locus, whilst also providing a means to potentially identify new loci associated with disease. The study presented provides confirmation of the clinical impact of
Overall discussion and future prospects

the two XLRP loci, RP3 and RP2 (Teague et al. 1994, Thiselton et al. 1996), accounting for 60-90% and 10-20% of disease respectively. In addition the classification of each family with respect to the locus segregating with disease is also useful in view of subsequent gene screening, as families not segregating with a particular gene can be excluded from the analysis, and microheterogeneity can be evaluated (section 4.7.1). Indeed the haplotype study presented was preliminary work prior to mutation screening of the relative XLRP and CSNBX genes in the genetically defined families, once the disease genes were identified. The screening of RPGR and RP2 in the patient pool enabled the identification of several novel mutations, but left disease in the majority of the families still unaccounted for (section 4.7.2). These mutations are likely to be in the newly identified additional exon of the RPGR gene that will now need to be screened in the patients. In contrast the screening of the NYX gene in the CSNBX families, concomitant to the screening of the CACNA1F gene done in Cambridge (Dr. L. Allen) on the same families, enabled the identification of mutations in almost all available patients (section 4.7.3).

The other advantage of carrying out extensive haplotype and linkage analysis is the potential for identifying new loci associated with a particular disease. This is particularly relevant for XLRP, due to the high level of genetic heterogeneity of this disease and to the high proportion of XLRP families which still appear to be unaccounted for by mutations in the two known genes (sections 4.7.1 and 4.7.2, Zito et al. 1999). This study has indeed enabled the identification of a new XLRP locus, RP23, located on the distal part of the short arm of the X chromosome (section 5.3.2, Hardcastle et al. 2000), establishing further genetic heterogeneity for XLRP. The research then progressed from linkage mapping to physical mapping, which resulted in the construction of a framework map that should provide the resource for a detailed search of the gene responsible for disease. Taking advantage of the large and increasing amount of data available on the database a physical map of the RP23 region in Xp22 as well as of the CSNB1 region in Xp11.4 was generated (sections 5.3.2.3 and 5.3.1.1). In addition, several positional candidates for RP23 were chosen from the Xp22 interval and screened in the affected family. Again with the aid of bioinformatics several ESTs were also mapped to the region and characterised to provide candidate transcripts for future screening. It is fascinating to witness the increasing potential of biocomputing approaches in combination with laboratory based work and how progress made by the
Human Genome Mapping Project and computational advances have provided a means to overcome the time-consuming manual construction of a contig and the identification of potential genes.

Because of the unusual completely asymptomatic status of females with the RP23 form of XLRP, which may be due to the expression of the disease gene from both active and inactive X chromosomes, the X inactivation status of several genes and ESTs in the RP23 region has been evaluated (section 5.3.2.6). Knowing the inactivation status of any gene on the X chromosome is also of general interest for understanding the distribution and mechanism underlying this phenomenon, why some genes appear to escape inactivation and what consequences this may pose, especially in individuals with X aneuploidy and in carrier females of X-linked diseases.

The identification of a new gene does not necessarily mean that all the questions relative to that gene and to the disease associated with the gene are answered. Also, it is not necessarily true that a phenotype is caused by a single gene and that a gene causes a single phenotype (i.e. genetic heterogeneity and allelic heterogeneity). There are well reported examples of allelism in the literature, with several examples of allelism in retinal diseases such as RP and CSNB (i.e. rhodopsin). Can this always be explained by different types of mutations whereby for example severe protein truncations cause a more severe phenotype than more subtle mutations? And how big is the influence played by genetic background in the phenotypic expression? A family identified in the clinic as XLRP with additional associated clinical symptoms including non-sensorineural deafness, recurrent chest infections and sinus infections suggested an associated cilia abnormality as the cause of disease (section 6.3). Haplotype analysis revealed disease segregation at the RP3 locus and an \( \text{RPGR} \) mutation was subsequently identified. This mutation although novel, is not functionally different from many others identified in the gene, which result in isolated RP. However literature searching enabled the identification of at least one other \( \text{RPGR} \) family with associated deafness (Roepman \textit{et al.} 1996b, Rosenberg \textit{et al.} 1997) and one with additional clinical symptoms similar to a condition called the immotile cilia syndrome (Dry \textit{et al.} 1999, van Dorp \textit{et al.} 1992). In addition, abnormalities of axonemal structure of nasal cilia and sperm cells have been detected in patients with XLRP (Arden and Fox 1979) and the mouse \( \text{RPGR} \) gene has been localised in the connecting cilium of photoreceptor cells (Hong \textit{et al.}
Factors which may be involved when considering the additional phenotype in this family are (i) genetic background, (ii) linkage disequilibrium with a mutation in another gene close to \textit{RPGR}, or (iii) that recurrent infections may exist in many XLRP patients clinically evaluated for their ophthalmic phenotype, which have been overlooked. To unravel this interesting issue, mouse models using different genetic backgrounds, as well as evaluation of candidate genes for the cilia abnormalities phenotype near the RP3 locus and examination of axonomal structure of these patients together with clinical re-examination of \textit{RPGR} patients will need to be investigated.

### 7.2 Unique features of different genes

Each gene has its own specific characteristics. In this study the analysis of three different genes, \textit{RPGR}, \textit{RP2} and \textit{NYX}, has highlighted distinctive features. Whilst both \textit{RPGR} and \textit{RP2} present various degrees of polymorphism, no such changes have been detected in the \textit{NYX} gene. Also the \textit{RP2} gene has a trend towards protein truncation mutations, more so than \textit{RPGR}, for which all types of mutation are equally represented, and even more than \textit{NYX}, where the majority of mutations appear to be missense. In addition, the \textit{NYX} gene appears to have a higher level of ancient mutations when compared with both the other genes, where the majority of mutations appear to be different and new.

The specific characteristics of a gene depend on the role that each gene plays in the aetiology of a certain disease, on the tertiary structure of the protein and on the number and complexity of interactions with other proteins. Until we learn more about the function of each protein we can only speculate on the reasons why these three genes present these different features. In the meantime however the number, type and site of mutations and polymorphisms identified in the genes can provide essential information that will help to unravel the function of their respective protein products. For example, the clustering of mutations in the region of homology with other proteins for each of the three genes, respectively RCC1, cofactor C and LRR domains for \textit{RPGR}, \textit{RP2} and \textit{NYX}, highlights the relevance of these regions, particularly in the case of missense mutations. The absence of polymorphic changes in the \textit{NYX} gene suggests that the protein has a complex tertiary structure unable to tolerate any amino acid change as each residue takes part in the maintenance of the three-dimensional structure. Indeed the three-dimensional structure of the
protein has been modelled and structural implications of different mutations have been investigated (Pusch et al. 2000); 8 missense mutations have been shown to affect highly conserved amino acids of the LRRs responsible for the maintenance of the 3D structure. In addition the large proportion of ancient mutations in the NYX gene provides evidence for founder effects, which may be more common in comparison with the other two genes due to the milder phenotype of CSNB in relation to RP, thus mutations in this gene may be more easily conserved throughout the generations.

In some cases mutations can be experimentally reproduced to biochemically test their effect. This has been achieved for two of the mutations identified in the RP2 gene, R118H and ΔS6 (Chappie et al. 2000), by testing the effects of the mutations on the subcellular localisation of the protein. A correct localisation for the R118H mutation and a mis-localisation of the ΔS6 mutation was shown. This result was particularly interesting because patients with the R118H mutation appeared to have a more aggressive form of the disease in contrast with patients with the ΔS6 mutation, associated with a milder phenotype, and this may be due to the fact that the ΔS6 mutation affect the proper targeting of RP2 to the correct cellular location, while the R118H mutation affects the function of correctly localised protein.

In this study the identification of a complex haplotype comprising seven different sequence alterations in the RPGR gene, with no apparent effect on the resulting phenotype, highlights the specific characteristics of the RPGR gene. Because these changes are in linkage disequilibrium and most of them have never been found in isolation, it is interesting to propose a potential beneficial effect of the complex haplotype that promoted its maintenance throughout generations. Of course this is only speculation, hopefully with the increasing studies of disease causing mutations and polymorphisms, other similar situations will be detected in RPGR or in any other genes, which may help to elucidate whether this is an example of a ‘positive genetic background’ that protects the individuals from other diseases, or if it was simply introduced in the general population as a result of neutral drift.

It is now important to ask the questions, why does a gene expressed in multiple tissues only cause an ocular phenotype? And what are the interacting partners of these novel proteins? Research emphasis should now shift towards functional studies of the proteins and understanding the cell biology behind pathogenic mechanisms.
7.3 The importance of genotype-phenotype correlation

The cloning of two XLRP genes and the two CSNBX genes has had significant impact towards the understanding of the genetics of these diseases, and describing the number and variety of gene mutations is important for genotype-phenotype correlation studies. Only by identifying disease-causing mutations can we relate the molecular defect to the clinical outcome and understand more about the mechanism of pathogenesis. The ability to phenotypically distinguish different forms of the diseases and relate each form to the causative gene has positive implications for the molecular diagnosis of patients and will aid genetic counselling in the clinic.

Whilst the two predominant forms of XLRP cannot be clinically distinguished and the integration of the genetic data with the clinical examination needs to be expanded in order to predict the gene causing disease in different XLRP patients, this study provides the first evidence towards a correlation of visual function with mutations identified in both CSNBX genes, confirming presence of a different phenotype associated with genotype. It will be important to further exploit this type of study so that by clinical examination we may be able to predict, not only which gene is involved, but also what type of mutation. This implies a more comprehensive knowledge of the function of the proteins, their specific roles in the neural retina and the mechanism by which they cause disease. In turn the ability to predict the type of mutation involved in each disease could have important consequences for genetic counselling and prenatal diagnosis, and for possible future treatments of the disease. It is only by understanding the precise effects of a mutation in the aetiology of disease, that complete and effective treatments can be developed.

7.4 Photoreceptors cell death and animal models

Despite the genetic heterogeneity of RP, each form of this disorder involves a common pattern of retinal degeneration, progressive degeneration of rod photoreceptor cells (night blindness and tunnel vision) followed by progressive cone degeneration over decades. Ultimately, loss of cones results in patients being registered blind. Also, it is now generally accepted that in animals, the common pathway leading to photoreceptor cell death is through
Overall discussion and future prospects

a mechanism called apoptosis or programmed cell death (Chang et al. 1993), and a similar mechanism may occur in human retinal degeneration. Unlike simple necrotic processes, where chemical or structural alterations lead to the arrest of metabolic activity, cell lysis and phagocytic activity, apoptosis is a highly genetically and chemically regulated cellular mechanism and acts naturally as a normal process in embryonic development and normal cell turnover (Gerscherson and Rotello 1992). Apoptosis affects single cells, which are usually phagocytosed by adjacent cells with minimal effects to the microenvironment, with the aim of removing damaged cells or cells no longer required as an organism’s defence mechanism. The pathway of an apoptotic event involves chromatin condensation, DNA fragmentation, inability to repair the multiple damages, breaking down of the cell into membrane bound apoptotic bodies and phagocytosis by neighbouring cells (Schwartzman and Cidlowski 1993). Evidence for apoptosis in the retina came from the identification of the typical hallmarks of apoptosis in a number of animal models with retinal dystrophies, such as the retinal degeneration slow (rds) mouse caused by a peripherin mutation, the retinal degeneration (rd) mouse, caused by a defect in cGMP phosphodiesterase, and transgenic mice with different rhodopsin mutations (Chang et al. 1993, Portera-Cailliau et al. 1994). Due to the complexity of apoptosis and the number of genes and regulatory factors involved, the entire mechanism is not yet completely understood. However pivotal studies in the nematode Caenorhabditis elegans have identified 14 different genes involved in the process, and a number of studies have also implicated Ca²⁺ levels as a key event in triggering the signal leading to apoptosis (reviewed in Gregory and Bird 1995). In the majority of mouse models the onset of retinal degeneration is earlier in comparison with humans and this may reflect the different genetic backgrounds in the different mammals. Also, the animal models are usually derived from an inbred colony with a unique genetic background. Nonetheless these models provide a valuable source for studying the pathophysiology of the disorder, whilst providing a unique opportunity to develop therapeutic strategies to delay the onset of disease. The importance of the discovery of apoptosis as a common pathway to photoreceptor cell death is that therapeutic approaches can be developed to target this specific phenomenon, and if the same mechanism applies to human retinal degeneration it will hopefully be possible to create a single form of therapy valid for all forms of RP, rather than trying to correct the specific genetic defect in each of them.
Overall discussion and future prospects

Much of our knowledge of inherited retinal degeneration comes from studies on several animal models. Lack of natural occurring models for XLRP and CSNBX has hampered the genetic understanding of these diseases. It is only more recently that a naturally occurring dog model for RP3 (Zeiss et al. 2000) and a naturally occurring mouse model for CSNBX (nob, Pardue et al. 1998) have been identified, and a knock-out mouse for the RPGR gene has been created (Hong et al. 2000). The RPGR dog gene has been isolated and completely sequenced in the affected dog, however no mutations were identified in the original sequence but recent studies report the presence of a mutation in the new exon ORF15 of RPGR (A clandest et al. 2001, Zhang et al. 2001). The nob mouse shows several key features in common with patients with the complete form of CSNB (CSNB1) and it will now be interesting to screen the yet unidentified mouse orthologue of the NYX gene in this model. The RPGR-deficient mouse model exhibits a phenotype typical of XLRP, however more recently a different knock-out mouse created with the replacement of exon 4 with a neomycin cassette, which leads to a 21 amino acids in-frame deletion, shows absence of retinal degeneration (Ruether et al. 2001), demonstrating the different effect of the mouse genetic background on the phenotypic outcome when compared to human.

These animal models should allow disease progression to be followed from development to advanced stages in more detail than has been possible in humans, providing a better understanding of the pathophysiology of these diseases. When studying a particular mutation, the use of gene targeting to create a transgenic animal, which harbours the specific mutation, should help in understanding its effects. The effect on an animal model or organism of the absence of a gene can also be evaluated by gene replacement or gene knockout (Capecchi 1994).

The impact of these animal models coupled with the molecular knowledge of the role of genes in the pathogenesis of disease, lies in the potential to move beyond the level of genetic counselling and non-specific therapies to the direct transfer of therapeutic genes or pharmacological agents into the retina (Bennet et al. 1996, Pepose and Leib 1994). A gene therapy approach is particularly relevant for X-linked retinal diseases where male patients suffer from haploinsufficiency, therefore the insertion of a normal copy of the gene may be efficient in recovering the normal phenotype or at least to slow disease progression.
References


References


Birnboim HC and Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nuc Acids Res* 7:1513-1523


References
References


References


References


References


References


252
References


Grayson C, Reid SNM, Ellis JA, Rutherford A, Sowden JC, Yates JRW, Farber DB, Trump D (2000) Retinoschisin, the X-linked retinoschisis protein, is a secreted photoreceptor protein, and is expressed and released by Weri-Rb1 cells. Hum Molec Genet 9: 1873-1879


Hiraoka M, Trese MT, Shastry BS (1999) Intragenic polymorphic missense mutations in the XLRS1 gene in families with juvenile X-linked retinoschisis. Hum Genet 104:526-527


knock-out: new insights on the pathogenesis of ocular albinism type 1. *Hum Molec Genet* 9:2781-2788


Kajiwara K, Berson EL, Dryja TP (1994) Digenic retinitis pigmentosa due to mutations at the unlinked peripherin/RDS and ROM1 loci. *Science* 264:1604-1608


References


Larin Z, Monaco AP, Lehrach H (1991) Yeast artificial chromosome libraries containing large inserts from mouse and human DNA. *Proc Natl Acad Sci USA* 88:4123-4127

Laurence JZ, Moon RC (1995) Four cases of retinitis pigmentosa occurring in the same family and accompanied by general imperfection of development. *Obes Res* 2:32-41


References


References


Monaco AP, Larin Z (1994) YACs, BACs, PACs and MACs: artificial chromosomes as research tools. Trends Biotech 12:280-286


References


Pardue MT, McCall MA, LaVail MM, Gregg RG, Peachey NS (1998) A naturally occurring mouse model of X-linked congenital stationary night blindness. *Invest Ophthal Vis Sci* 39(12):2443-3449


References


References


262


References


Sparkes RL, Summer CG, Boycott KM, Zahorchak RJ, Bech-Hansen T (2000) Development of a 1.4-Mb BAC/PAC contig and physical map within the critical region for complete X-linked congenital stationary night blindness in Xp11.4. *Genomics* 68:97-100


References


van den Hurk JA JM, Hendriks W, van de Pol DJR, Oerlemans F, Jaisisse G, Ruther K, Kohler
Mouse choroideremia gene mutation causes photoreceptor cell degeneration and is not
transmitted through the female germline. Hum Molec Genet 6: 851-858

van de Vosse E, Bergen AAB, Meershoek EJ, Oosterwijk JC, Gregory S, Bakker B,
YAC contig encompassing the disease loci for RS, KFSD, CLS, HYP and RP15: refined
localization of RS. Eur J Hum Genet 4:101-104

van de Vosse E, Franco B, van der Bert P, Montini E, Hanauer UOA, Tijmes N, van Ommen
GJB, Ballabio A, den Dunnen JT, Bergen AAB (1997) Exclusion of PPEF as the gene
causing X-linked juvenile retinoschisis. Hum Genet 101:235-237

Van Dorp DB, Wright AF, Carothers AD, Bleeker-wagemakers EM (1992) A family with
RP3 type of X-linked retinitis pigmentosa: an association with ciliary abnormalities. Hum
Genet 88:331-334

van Slegtenhorst MA, Bassi MT, Borsani G, Wapenaar MC, Ferrero GB, de Conciliis L,
shares homology with voltage-gated chloride channels. Hum Mol Genet 3(4):547-5 52

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M et

Vervoort R, Lennon A, Bird AC, Tulloch B, Axton R, Miano MG, Meindl A, Meitinger T,
Ciccodicola A, Wright AF (2000) Mutational hot spot within a new RPGR exon in X-linked

Wada M, Little RD, Abidi F, Porta G, Labela T, Cooper T, Valle GD, D'Urso M,
Schlessinger D (1990) Human Xq24-Xq28: Approaches to mapping with yeast artificial

in a Japanese family with X-linked retinitis pigmentosa. Invest Ophthalmo l Vis Sci
41(1):290-293

Walpole SM, Nicolau A, Howell GR, Whittaker A, Bentley DR, Ross MT, Yates JRW,
Trump D (1997) High-resolution physical map of the X-linked retinischisis interval in Xp22.
Genomics 44:300-308

Walpole SM, Hiriyanak KT, Nicolaou A, Bingham EL, Durham J, Vaudin M, Ross MT, Yates
JRW, Sieving PA, Trump D (1999a) Identification and characterization of the human
homologue (RAI2) of a mouse retinoic acid-induced gene in Xp22. Genomics 55:275-283


References


References


References


References


245


Birnboim HC and Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nuc Acids Res* 7:1513-1523


References


References


References


References


Grayson C, Reid SNM, Ellis JA, Rutherford A, Sowden JC, Yates JRW, Farber DB, Trump D (2000) Retinoschisin, the X-linked retinoschisis protein, is a secreted photoreceptor protein, and is expressed and released by Weri-Rb1 cells. *Hum Molec Genet* 9: 1873-1879


Hiraoka M, Trese MT, Shastry BS (1999) Intragenic polymorphic missense mutations in the XLRS1 gene in families with juvenile X-linked retinoschisis. Hum Genet 104:526-527


References

knock-out: new insights on the pathogenesis of ocular albinism type 1. *Hum Molec Genet* 9:2781-2788


Kajiwara K, Berson EL, Dryja TP (1994) Digenic retinitis pigmentosa due to mutations at the unlinked peripherin/RDS and ROM1 loci. *Science* 264:1604-1608


Larin Z, Monaco AP, Lehrach H (1991) Yeast artificial chromosome libraries containing large inserts from mouse and human DNA. *Proc Natl Acad Sci USA* 88:4123-4127

Laurence JZ, Moon RC (1995) Four cases of retinitis pigmentosa occurring in the same family and accompanied by general imperfection of development. *Obes Res* 2:32-41


References


References


Monaco AP, Larin Z (1994) YACs, BACs, PACs and MACs: artificial chromosomes as research tools. Trends Biotech 12:280-286


References


References


Pardue MT, McCall MA, LaVail MM, Gregg RG, Peachey NS (1998) A naturally occurring mouse model of X-linked congenital stationary night blindness. *Invest Ophthal Vis Sci* 39(12):2443-3449


References


Sparkes RL, Summer CG, Boycott KM, Zahorchak RJ, Bech-Hansen T (2000) Development of a 1.4-Mb BAC/PAC contig and physical map within the critical region for complete X-linked congenital stationary night blindness in Xp11.4. *Genomics* 68:97-100


References


References


References


**MUTATION IN BRIEF**

Sequence Variation within the *RPGR* Gene: Evidence for a Founder Complex Allele

Ilaria Zito 1, Alex Morris 1, Phil Tyson 2, Ingrid Winship 3, Dianne Sharp 3, Dale Gilbert 3, Dawn L. Thiselton 1, Shomi S. Bhattacharya 1, and Alison J. Hardcastle 1

1 Department of Molecular Genetics, Institute of Ophthalmology, UCL, London, UK; 2 Department of Psychiatry, School of Medicine, University of Hull, Hull, UK; 3 Molecular Genetics Laboratory, Auckland Hospital, Auckland, New Zealand

Correspondence to Dr. Alison J. Hardcastle, Department of Molecular Genetics, Institute of Ophthalmology, University College London, 11-43 Bath Street, London, EC1V 9EL; Tel: 44 171 608 6945; Fax: 44 171 608 6863; E-mail: ahardcas@hgmp.mrc.ac.uk

Contract grant sponsor: Wellcome Trust; Contract grant number: 051733/z/97

Communicated by Mark H. Paalman

In our study of sequence variation within the *RPGR* gene associated with X-linked retinitis pigmentosa, we and others have observed a high rate of new mutation within this gene, as all reported mutations are unique or uncommon. In this article we report the identification in a single family of a complex allele of 7 sequence variants in linkage disequilibrium, of which four result in amino-acid alterations (Arg425Lys, ΔGlu, Thr533Met and Gly566Glu). This complex allele was initially found in a family with XLRP. However, further study revealed an estimated prevalence of 4.3% (15/344 chromosomes) with this complex allele in the European population indicating the non-pathogenic nature of this allele and, along with previously reported polymorphisms, further supporting a high level of human protein diversity for *RPGR*. This common complex allele may have been established in the population as a founder effect. Complete gene sequencing identified a potential pathogenic sequence variant in the family described (IVS6+5G>A). This study emphasises the need to create a more complete picture of the allelic variation within a gene, suggests cautious interpretation of a phenotypic association with variant sequences, and highlights the potential problems associated with interpreting genetic studies for diagnostic purposes.

© 2000 Wiley-Liss, Inc.

**KEY WORDS:** SNPs; single nucleotide polymorphisms; complex haplotype; linkage disequilibrium; *RPGR*; retinitis pigmentosa GTPase regulator; founder effect

**INTRODUCTION**

X-linked retinitis pigmentosa (XLRP) is a progressive degenerative disease of the retina, which is phenotypically and genetically heterogeneous. Two major genetic loci have been identified (*RP2* and *RP3*) on the short arm of the X chromosome (Thiselton et al. 1996, Ott et al. 1990) and the respective causative genes cloned (Meindl et al. 1996, Roepman et al. 1996, Schwann et al. 1998). Two minor genetic loci have more recently been identified (*RP23* and *RP24*), however the causative genes at these loci have not been cloned (Geiser et al., 1998;

Received 16 May 2000; Revised manuscript accepted 13 June 2000.

© 2000 WILEY-LISS, INC.
While genetic studies have shown that disease in the majority of XLRP families (over 70%) co-segregates with the RP3 locus (Ott et al., 1990; Teague et al. 1994), it appears from genetic studies that disease in these families co-segregates with the RP3 locus (Ott et al., 1990; Teague et al. 1994). The disease gene at this locus (RPGR or retinitis pigmentosa GTPase regulator; MIM#312610) has been extensively studied and different mutations and polymorphisms have been reported (Meindl et al., 1996; Roepman et al., 1996; Buraczynska et al., 1997; Miano et al., 1999; Zito et al., 1999; Zito et al. 2000).

In addition to the thirty-nine different mutations identified in the gene to date, seventeen different polymorphisms have been reported (Meindl et al., 1996; Roepman et al., 1996; Buraczynska et al., 1997; Miano et al., 1999; Zito et al., 1999).

The importance of developing a catalog of human sequence variations is generally acknowledged. Single nucleotide polymorphisms (SNPs) occur at a frequency of approximately one per Kilobase of DNA (Chakravarti 1999), with a decreased prevalence for those polymorphisms occurring within regulatory and coding regions (cSNPs, Wang et al., 1998), consistent with greater constraints due to protein sequence conservation within such gene sequences. The presence of a number of cSNPs however raises the possibility that several of these polymorphisms might be functional (Cambien et al., 1999).

It is also important to consider that the high level of polymorphism of the human genome may easily mislead the interpretation of molecular findings. This highlights the necessity for supporting each putative mutation identified with functional analysis to ensure what is described as a mutation is indeed disease-causing.

During the course of screening the RPGR gene in a sample of XLRP families and control chromosomes, we identified a haplotypic combination comprising seven different sequence variants contained both in coding and in non-coding regions of the gene thus creating a complex allele, with no apparent effect on the resulting phenotype.

MATERIALS AND METHODS

Patients and controls

Appropriate informed consent was obtained from the XLRP family NZ1 and European control volunteers under investigation. An X-linked form of retinitis pigmentosa was established by pedigree analysis, clinical examination, and ophthalmological tests, which include electoretinogram (ERG) recordings, fundus examination, visual field assessment and fluorescein angiography.

Blood samples were collected from each available member of the family and from control individuals and DNA was extracted using the Nucleon II Kit (Scotlab Limited) according to manufacturer instructions.

Haplotype analysis

Fifteen microsatellite markers spanning the short arm of the X chromosome were used to generate haplotypes for all sampled individuals in the NZ1 family. Haplotypes were constructed assuming the minimal number of recombination events. PCRs were carried out in 10 µl reactions in the presence of 1 µCi α32P-dCTP, 0.5 U Taq polymerase, 200 µM each of dATP, dGTP, dTTP and 20 µM of dCTP, 50 pmol of each primer, 30 mM Tris-HCl, pH 8.5, 50 mM KCl, 1.5mM MgCl2, 0.01% gelatine. Amplification conditions were 95°C for 5 min, followed by 35 cycles of denaturation at 95°C for 15 sec, annealing at the primer specific temperature for 15 sec, and extension at 72°C for 30 sec. A final extension followed for 5 min at 72°C. Amplified products were mixed with 6 µl of formamide sample buffer and 3 µl aliquots were electrophoresed in 6% denaturing polyacrylamide gels. The gels were dried onto Whatman paper and analysed by autoradiography.

Detection of sequence variants

SSCP analysis

Exon fragments were amplified (see description of PCR above) with intronic primers (Meindl et al., 1996). Amplified DNA was mixed 1:1 with formamide sample buffer containing 20 mM NaOH. The products were analysed for single strand conformational polymorphism using non-denaturing 0.5X mutation-detection-electrophoresis gels (MDE; FMC Bioproducts) in 0.6x TBE buffer and run at 8-10 Watts constant power for 14-18 h. The gels were dried onto Whatman paper and analysed by autoradiography.
Genetic Variation within the \textit{RPGR} Gene

\textbf{Sequence analysis}

PCR products were examined by agarose gel electrophoresis prior to sequencing. An aliquot of the amplification product (8 \(\mu\)l) was then purified by the addition of 1 U shrimp alkaline phosphatase (SAP; Amersham Life Science) and 1 U Exonuclease I (United States Biochemical) in SAP buffer, and incubation at 37\(^\circ\)C for 30 min followed by 80\(^\circ\)C for 15 min. Five \(\mu\)l of the purified DNA sample was then used for cycle sequencing using Big Dye Terminator cycle sequencing kit following manufacturers instructions. Reactions were then electrophoresed on an ABI 373A automated sequencer.

\textbf{Linkage disequilibrium analysis}

Linkage disequilibrium was estimated using the Estimation of Haplotypes (EH) program test for association (Terwilliger & Ott, 1994). The EH program requires genotype data so that both males and females were treated as biallelic.

\textbf{RESULTS}

Haplotype analysis was conducted on an XLRP family (NZ1) with fifteen microsatellite markers localized on the short arm of the X chromosome. A cross-over was identified between markers OTC and DXS556 (data not shown) placing the disease distal to DXS556 hence spanning the \textit{RP3} locus.

SSCP and sequencing analysis of the \textit{RPGR} gene in an affected member of the family showed 8 different sequence variations (Table 1).

\begin{table}[h]
\begin{tabular}{|c|c|c|c|}
\hline
Intron/exon & Position* & Change & Amino acid no & Consequence \\
\hline
Intron 6 & 678+5 & G->A & - & - \\
Intron 8 & Primer sequence & Not characterized & - & - \\
Exon 11 & 1333 & G->A & 425 & Arg->Lys \\
Intron 13 & 1631+11 & A->G & - & - \\
Exon 14 & 1635-1637 & 3-bp deletion & 526 & Gln deleted \\
Exon 14 & 1657 & C->T & 533 & Thr->Met \\
Exon 14 & 1765 & G->A & 566 & Gly->Glu \\
Intron 18 & 2300+11 & C->T & - & - \\
\hline
\end{tabular}
\caption{Sequence variations identified in \textit{RPGR} in family NZ1}
\end{table}

Of these variations, three changes identified (1333 G>A, IVS13+11A>G and IVS18+11 C>T) have all been previously described as polymorphic variants (Roepman et al., 1996; Buraczynska et al., 1997; Fujita et al., 1997; Zito et al., 1999). The fourth variant, IVS6+5G>A, is likely to affect splicing at this site. The fifth change is contained in the reverse primer sequence of exon 8. This exon failed to amplify with published primers 8-F and 8-R (Meindl et al. 1996), however a product of the expected size was obtained using primer 8-F and a newly designed primer (8b-R: cagttctataaataatataacag) still spanning the intron/exon boundary. Exon 8 and the intron/exon boundaries are therefore not altered in this family.

The remaining three changes are all contained within exon 14 (180 bp); 1635-1637delCAA resulting in the loss of a Glu at codon 526, 1657C>T resulting in a Thr>Met change at codon 533, and 1765G>A resulting in a Gly>Glu change at codon 566, i.e. complete loss of one amino acid and two non-conservative amino acid substitutions.

All the exons for which sequence alterations were identified in the patient were analyzed in the other members of this family. The sequence variation identified in intron 6 (IVS6+5G>A) is the only variation which appeared to segregate with disease in the pedigree (Figure 1a) and is therefore the only potential disease-causing mutation. This nucleotide is part of the 5' consensus donor splice site and is 80% conserved in mammals (Horowitz and Krainer, 1994). In addition exon 6 was sequenced in 275 control chromosomes and this change was not detected.

The other 7 sequence variations identified in the family have no apparent effect on the resulting phenotype.

In order to provide evidence of polymorphism for the three protein altering variants observed in exon 14, the exon was sequenced in 344 control chromosomes from individuals chosen for their healthy status, predominantly
of European origin. Surprisingly 15 out of 344 chromosomes appeared to have all of the same three protein altering changes in exon 14. None of the control samples have the 526delGlu or the Thr533Met change in isolation; while the Gly566Glu change, previously described as a polymorphism (Buraczynska et al., 1997) was found in one chromosome in isolation.

Sequencing of the remaining exons of the \textit{RPGR} gene in these 15 chromosomes revealed a common configuration of the 7 sequence variations in the \textit{RPGR} gene identified in family NZ1. The calculated prevalence of this complex haplotype is 4.3\% (15/344) in the European population, demonstrating the presence of a founder effect. Figure 1b shows an ideogram of the complex haplotype compared to the wild type sequence.

Linkage disequilibrium between the three variant sites in exon 14 was estimated. Out of 344 chromosomes screened, fifteen were found with the DelCAA T A haplotype of which five were heterozygous females. The other 324 chromosomes contained the CAA C G haplotype except for one individual with a CAA C A haplotype. As expected for three sites within a 130 bp interval, a very highly significant association was obtained ($\chi^2=205$, df=4 $p<<10^{-10}$). If these sites had recombined independently, the frequency of the DelCAA T A haplotype would be 0.0083\% compared to the 4.3\% found in our population which is consistent with a founder effect.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{A: Pedigree of the NZ1 family showing segregation of the IVS6+5G>A mutation with disease in the family and non-segregation of the complex allele. The boxed G represents the wild type sequence in intron 6 at position 678+5. The boxed shaded A represents the sequence variation at the same position. I represents the wild type sequence at positions: 1333, 1631+11, 1635-1637, 1657, 1765, 2300+11. represents sequence variations at the same positions. The sequence variation in intron 6 is omitted. B: Ideogram of the \textit{RPGR} gene showing the complex allele (I) and the wild-type sequence (II). Arrows pinpoint the sites where changes have occurred. The star depicts a non characterized change. Letters indicate the relative nucleotides and bars indicate deleted nucleotides. Numbers underneath refer to the exons. Exons are not drawn to scale and exons 2 to 7 are omitted.}
\end{figure}

\section*{DISCUSSION}

Many evolutionary changes or population specific alterations have no effect whatsoever on phenotype, however others can alter the function of a particular protein and ultimately lead to a disease state. The relative proportion
Genetic Variation within the \textit{RPGR} Gene

and type of mutation or polymorphism are found to vary from different gene loci and the specific characteristics of individual genes are extremely important and generally overlooked.

Here we report some interesting features regarding the \textit{RPGR} gene. A complex haplotype comprising seven different sequence alterations has been identified. This haplotype has no apparent effect on the resulting phenotype as it is present in the general population but generally highlights the specific characteristics of the \textit{RPGR} gene. The cSNPs forming the complex haplotype occur outside the RCC1-like domain, hence it is not possible to predict their significance.

The level of sequence variation has been investigated for a number of genes (Rieder et al., 1999; Nickerson et al., 1999). All emphasize the importance of obtaining a more complete picture of the allelic variation within a gene, suggest a cautious interpretation of phenotypic association with a single SNP, and highlight the potential problems associated with interpreting genetic studies (Hiroaka et al., 1999).

Complete or nearly complete association between polymorphisms has been frequently observed (Cambien et al., 1999; Behague et al., 1996). Such associations suggest the existence of ancestral haplotypes with potentially beneficial functional consequences. It is probable that each sequence alteration arose as an independent event, and the complex haplotype became fixed in the population either as a result of neutral drift or, if functional, by positive selection. Other combinations of SNPs may be observed in different ethnic groups, potentially with maintained complex alleles which vary from the European common complex allele described here.

We therefore suggest that the haplotype here described results from selection or random genetic drift or from some combination of drift-selection occurring successively at individual nucleotide sites (Grant 1991) and resulting in a combination of forces that collaborate to determine its final shape. This haplotype may have then been introduced to the population under study as a founder effect. We conclude that there appears to be a low level of constraint of human genetic variation within the \textit{RPGR} gene and that the study of naturally occurring sequence variants (mutations and polymorphisms) is vital for understanding human genome pathology, in particular the relationships between i) variants of protein primary structure and tertiary structure and function, and ii) relationships between genotype and phenotype.

REFERENCES


Dry KL, Manson FDC, Lennon A, Bergen AA, VanDrop DB, Wright AF. 1999. Identification of a 5' splice site mutation in the \textit{RPGR} gene in a family with X-linked retinitis pigmentosa (RP3). Hum Mutat 13 (2):141-145


Hiroaka M, Trese MT, Shastry BS. 1999. Intragenic polymorphic missense mutations in the XLRS1 gene in families with juvenile X-linked retinoschisis. Hum Genet 104:526-527


Evidence for a New Locus for X-Linked Retinitis Pigmentosa (RP23)

Alison J. Hardcastle, Dawn L. Thibelton, Ilaria Zito, Neil Ebenezer, Tammy S. Mah, Michael B. Gorin, and Sishi S. Bhattacharya

Purpose. X-linked retinitis pigmentosa (XLRP) is a degenerative disease of the retina characterized in the early stages of disease by night blindness as a result of rod photoreceptor loss, progressing to severe disease with loss of central vision by the third decade in affected males. XLRP displays exceptional genetic heterogeneity, with five reported loci on the human X-chromosome. To investigate the level of heterogeneity for XLRP in the patient pool in the current study, extensive haplotype analysis, linkage analysis, and mutation screening were performed.

Methods. Haplotype analysis of a family with diagnosed XLRP was scored with more than 54 polymorphic markers spanning the entire X-chromosome, including regions already identified as harboring XLRP genes and retina-specific genes. Two-point and multipoint lod scores were calculated. Affected male DNA was amplified with primers specific for the retinoschisis gene (XLRSl), and the products were screened for nucleic acid alterations by direct automated sequencing.

Results. In this article haplotype and linkage data are presented identifying a new locus for XLRP on the short arm of the X-chromosome, distinct from previously reported gene localizations for XLRP. The phenotype is atypical, in that the onset of vision loss in the male members of this family is unusually early, and female obligate carriers have normal fundi and waveforms. Informative recombination events in this family define a locus for XLRP (RP23) on Xp22 between the markers DXS1223 and DXS7161, spanning approximately 15 cM. A maximum lod score of 2.1 was calculated for the locus order DXS7103-8 cm-(RP23/DXS1224)-4 cM-DXS999. This new locus (RP23) encompasses the retinoschisis disease gene; therefore, XLRSl was screened for a mutation. No sequence alteration was identified indicating that mutations in the coding region of the gene responsible for retinoschisis do not cause RP23.

Conclusions. The results describe evidence for a new locus for XLRP (RP23), adding to the established genetic heterogeneity for this disease and the number of genes expressed in ocular tissue residing on the X-chromosome. (Invest Ophthalmol Vis Sci. 2000;41:2080-2086)

The human X-chromosome is home to a large number of genes involved in inherited diseases of the eye, of which retinal diseases comprise a majority (Fig. 1). These genetic disorders are a significant cause of visual impairment and blindness and include retinitis pigmentosa (RP), congenital stationary night blindness (CSNB), and progressive cone dystrophy (COD). X-linked retinitis pigmentosa (XLRP) is the most severe form of RP, with male patients showing concentric visual field loss before the 20th year of life, leading to severe visual handicap. Female carriers show variable clinical symp-

Inspired by the Department of Molecular Genetics, Institute of Ophthalmology, University College London, United Kingdom, and the Department of Ophthalmology, School of Medicine, University of Pittsburgh, Pennsylvania.

Supported by The Wellcome Trust Grant 051733/Z/97/AJH; and The Guide Dogs for the Blind Association Grant 95-52A (DLT), London, United Kingdom; The Eye & Ear Foundation of Pittsburgh, Pittsburgh, Pennsylvania (MEG); and unrestricted funds from Research to Prevent Blindness.

Submitted for publication May 7, 1999; revised August 20, 1999, and January 24, 2000; accepted February 10, 2000.

Commercial relationships: None.

Corresponding author: Alison J. Hardcastle, Department of Molecular Genetics, Institute of Ophthalmology, University College London, 11-43 Bath Street, London, EC1V 9EL, UK. ahardcas@hgmp.mrc.ac.uk
A New Locus for XLRP

**Figure 1.** Ideogram of the human X-chromosome showing map locations of various eye diseases incorporating a genetic map of Xp22.11-Xp22.32. Genetic distances (Genethon 1996) taken from The Integrated X-Chromosome Database (http://ixdb.mpim-berlin-dahlem.mpg.de/). Superscript numbers refer to studies identifying the genes.
genetic mapping studies (see Fig. 1), however, mutation detection in the causative genes creates a more complex picture that leaves more than 60% of disease currently unaccounted for. In fact, the isolation of the RPRG and RP2 genes has made haplotype analysis leading to crossover detection an even more essential part of XLRP research because of the potential existence of genetically undefined loci on the X-chromosome.

We have extensively haplotyped many XLRP families, and in this article we report a new locus for atypical XLRP in a single family that is not associated with any previously described loci.

**MATERIALS AND METHODS**

**Clinical Assessment**

This study, which involved human subjects, conformed to the tenants of the Declaration of Helsinki. Individual IV-2 was familiar with the clinical histories of three other males in the family, all of whom had the onset of poor vision before the age of 2 years. Thus she sought an ophthalmic evaluation for her son (individual V-2) at age 2 years, when he was noted to have limited central vision and poor night vision and frequently ran into objects.

On examination, the patient fixed and followed with both eyes and had no evidence of tropias, phorias, or nystagmus. His retinoscopic refraction was +3.50 sphere and +1.25 sphere for the right and left eyes, respectively. The anterior segment examination findings were completely normal, and the fundus examination demonstrated normal optic nerves (cup/disc ratio [C/D] = 0.3), abnormal greyish macular reflexes, and extensive whitish gray spots (discrete and not fleck like) distributed throughout the midperiphery of the posterior pole. These spots appeared at the level of the pigment epithelium and were not associated with any overlying intraretinal pigment migration. In many areas these white spots were coalescent. When he was examined 4 years later, the patient had no nystagmus, but he clearly demonstrated eccentric fixation with preferential use of the left eye and more symptomatic photophobia. His visual acuities were estimated to be in the 20/900 range in both eyes.

Findings in a fundus examination were virtually unchanged, except that retinal arteriole attenuation was now evident. Karyotype evaluation in 1997 demonstrated normal i6XY chromosomes in all cells analyzed.

Individual V-2 was most recently clinically examined in 1998 at age 11. The patient reported no photophobia, and his visual acuity was limited to counting fingers at 2 to 3 feet. His peripheral vision was markedly constricted, and the changes in the pigment epithelium were nearly confluent in the midperiphery, giving rise to an overall grayish appearance (see Fig. 2). Small patches of retinal pigment epithelium (RPE) atrophy were now present in the midperiphery, and a few small patches of intraretinal pigment were noted. There was a circular area of approximately 1.5 disc diameters, centered over the fovea that appeared to have relatively normal retina and RPE. The retinal arterioles were clearly attenuated in all quadrants, but there was no optic nerve pallor. Electroretinograms (ERGs) and formal visual field testing of the child were not performed because of his hyperactive nature, because of the family’s reluctance to have him sedated, and because such testing would not alter his diagnosis, prognosis, or treatment.

**Figure 2.** Photomontage of the fundus of affected male V-2 showing attenuated retinal arterioles and near-confluent changes in the pigment epithelium in the midperiphery.

Subject IV-4 was initially evaluated by a genetic retina specialist in Boston at age 21. At that time, extensive electrophysiologic and psychophysical testing demonstrated central visual acuities of 2/400 with mild myopia (~1.00–2.00 × 180 OU), nondetectable ERG recordings bilaterally, abnormal color vision, and only residual temporal and inferior fields of vision. The clinical examination was remarkable for the absence of foveal reflexes in both eyes, diffuse granularity of the RPE in the central macula, and moderate intraretinal pigment in the retinal midperipheries. A follow-up examination was conducted when the patient was 35 years of age by one of the authors (MBG). At that time, his visual acuities had deteriorated to hand motion vision in both eyes, and his macular and retinal midperipheries were relatively unchanged except for more notable areas of RPE atrophy and increased amounts of intraretinal pigment in the retinal peripheries. Posterior subcapsular cataracts, which were not present at age 21, were clearly evident by age 35. Both retina specialists independently concluded that the diagnosis was consistent with RP, with a severe phenotype and macular involvement.

Both the mother (individual IV-2) and the maternal grandmother (individual III-2), who were obligate carriers for the condition based on the pedigree, were examined and underwent ERG testing. The mother’s uncorrected Snellen visual acuities were 20/25-2 and 20/15-1 in the right and left eyes, respectively. The visual acuities in the grandmother were 20/25-1 and 20/25-2 in the right and left eyes, respectively. The clinical fundus examinations were completely normal, with no evidence of retinal or RPE changes for either individual. The ERGs were conducted using Jet electrodes with an LKC EPIC-XL instrument (LKC Technologies, Inc., Gaithersburg, MD) with a Ganzfeld chamber, in compliance with the international protocol standards and compared with age-matched control standards. Both carrier females demonstrated normal wave-
forms, with amplitudes and implicit times that were within the normal range (data not shown). For this study, the participation of family members was approved by the University of Pittsburgh Biomedical Institutional Review Board, in accordance with the guidelines of the Office for the protection from Research Risks, and informed consent was obtained from the patients before their participation.

Haplotype Analysis

The forward primer for each microsatellite was end labeled with [32P]-yATP at 37°C for 45 minutes using T4 polynucleotide kinase (New England Biolabs, Hertfordshire, UK). Polymerase chain reaction (PCR) was performed as previously described. Alleles were detected by electrophoresing the PCR products on 6% denaturing polyacrylamide gels (Promega, Southampton, UK). Details of primer sequences and PCR conditions for all microsatellites used in this study are available from The Genome Database (http://www.genome.ucsc.edu/)

Linkage Analysis

Two-point linkage analysis for XLRP and informative markers on chromosome Xp were scored by computer (Linkage, version 5.1 using Mlink; Columbia University, New York, NY). The frequency of the XLRP gene in the general population was taken to be 0.0001. Penetrance values for carriers were set at 0.0000. Alleles at marker loci were assumed to have equal frequency. Alleles were detected by electrophoresing the PCR products on 6% denaturing polyacrylamide gels (Promega, Southampton, UK).

Sequence Analysis

All six exon fragments of the XLRP gene were amplified with intronic primers described. PCR were performed in 25-μl reactions in the presence of 0.5 U Taq polymerase (Biotaq from Bioline, London, UK); 200 μM each of dATP, dCTP, dGTP, and dTTP; 200 picomoles of each primer and 1X KCl reaction buffer (from Bioline) including 1.5 mM MgCl2. PCR conditions were as described. Amplification products were purified with centricon concentrators (Amicon, Gloucestershire, UK) according to the manufacturer's instructions. The purified DNA sample (5 μl) was cycle sequenced in both directions (with primers used for amplification) using a termination cycle sequencing kit (ABI prism Ready Reaction Dye FS kit, Perkin-Elmer, Applied Biosystems, Worthington, UK) according to the manufacturer's instructions. Reactions were then electrophoresed on a sequencer (model 373A, Perkin-Elmer Applied Biosystems).

RESULTS

Clinical Characteristics

XLRP has been reported to be particularly severe compared with many of the autosomal forms, and macular lesions have also been noted in a higher percentage of X-linked cases; however, the onset of vision loss in the male members of this family was unusually early. There was clear evidence of progressive retinal and RPE changes with both rod and cone involvement (see Fig. 2), and affected male IV-4 had no recordable ERG response at age 21. Both obligate female carriers examined had normal fundi and waveforms.

Haplotype Analysis

The family was analyzed with more than 34 polymorphic marker loci spanning the entire X-chromosome. Figure 3 details haplotype results for the family members. Initial efforts were concentrated on microsatellites surrounding the RP2 and RP3 loci on the short arm of the X-chromosome, because these are the reportedly common loci for XLRP. As the haplotype of this portion of the X-chromosome was constructed, it became evident that the disease in this particular pedigree did not segregate with these intervals (see Fig. 3). Other loci (e.g., CSNB4/COD1157, and RP1515) and markers linked to known retina-specific genes (e.g., X-arrestin1156 and RhlGC2158) were excluded.

The distal boundary of XLRP in this family was defined by a recombination event observed in individual III-2, an obligate carrier female who had inherited the disease-associated haplotype at DXS7103, placing the disease proximal to DXS1223. This event was also observed in her obligate carrier daughter (IV-2), affected son (IV-4), daughter of unknown status (IV-6), and affected grandson (V-2). The proximal boundary was defined in individual IV-4, an affected male who was recombinant, relative to his carrier mother, between the markers DXS999 and DXS161. In summary, haplotype data clearly defined a locus for XLRP (RP23) between the loci DXS1223 and DXS161 in Xp22 with markers DXS7103, DXS1224, and DXS999 cosegregating with disease. This analysis would predict that individual IV-6 was a carrier and IV-8 was not at risk.

Linkage Analysis

Table 1 describes two-point linkage analysis results, demonstrating that disease in this family (RP23) was not linked to informative markers mapping to previously identified XLRP loci. Significant lod scores were obtained with markers DXS7103 (Zmax = 1.49; θ = 0), DXS1224 (Zmax = 1.96; θ = 0), and DXS999 (Zmax = 1.89; θ = 0). For markers telomeric to the aforementioned loci significant lod scores were not obtained—i.e., DXS996 (Zmax = 0.36; θ = 0.25) and DXS1223 (Zmax = 1.01; θ = 0.1). Insignificant lod scores were also obtained with markers DXS989 (linked to the RP15 locus; Zmax = 0.70; θ = 0.15), DXS1110 (linked to the RP3 locus; Zmax = 0; θ = 0.5) and DXS426 (linked to the RP2 locus; Zmax = 0; θ = 0.5). Multipoint analysis was performed to determine the most likely location of XLRP in relation to DXS7103, DXS1224, and DXS999 (see Figs. 1 and 3). A maximum lod score of 2.06 was scored for the locus order DXS7103—DXS1224—DXS999. The family under study originates from the United States; therefore, lod scores were recalculated with European allele frequencies for linked markers DXS7103 (allele 1 at 0.25 and allele 2 at 0.75), DXS1224 (allele 1 at 0.5, allele 2 at 0.2, and allele 3 at 0.3), and DXS999 (allele 1 at 0.6 and allele 2 at 0.4). Resultant lod scores were Z = 1.97 at θ = 0 for DXS7103, Z = 2.01 at θ = 0 for DXS1224, and Z = 1.85 at θ = 0 for DXS999, slightly increasing the statistical significance for linkage at these loci. To confirm the observation of X-linkage in this family, an autosomal dominant model with partial penetrance was compared with a fully penetrant X-
Figure 3. Haplotype analysis of family MGRP-3. Obligate carrier females are represented by half filled circles, affected males by filled boxes, and females of unknown clinical status by circles with question marks. Open circles and squares denote noncarrier females and unaffected males. The affected haplotype is shown as a shaded chromosome. The order of microsatellites on the X-chromosome is shown in the key to the right of the figure. Regions already known to harbor disease genes or retina-specific genes are highlighted in the key (also see Fig. 1). Unmarked alleles were not scored.
linked model using MLINK with a dummy marker. With a range of assessed penetrances the X-linked model was favored by at least 2 orders of magnitude. The relative likelihood (odds) of observing this segregation pattern if the disease is X-linked versus autosomal dominant is \((1/2)^{10}(1/2)^{10} \approx 1 \times 10^3\). These data add significant statistical evidence for X-linked inheritance in this family.

**XLRSl Mutation Screen**

All six exons and exon-intron boundaries were sequenced in affected males IV-4 and V-2 and no sequence alterations were detected when compared with the normal gene sequence. Mutations in the coding region of **XLRSl** are therefore excluded as causing disease in this family.

**DISCUSSION**

An emerging pattern for inherited retinal diseases is the exceptional heterogeneity of these disorders, with growing examples of genetic heterogeneity, allelic heterogeneity, and phenotypic heterogeneity. The complexity of the situation is amply demonstrated on the X-chromosome (Fig. 1). High-resolution mapping by haplotype analysis for X-linked retinal disease has enabled us to define the boundaries for **RP2** and establish a new genetic locus for **CSNBX (CSNB^p, RP22)**, which has potential to be allelic with the gene for **COD1**. The data presented here were obtained in our efforts to establish the number of distinct genes on the X-chromosome involved in eye disease, define their locations, and address the likelihood of allelic heterogeneity.

The family described had an atypical XLRP phenotype, in that the age of onset of disease in affected males is unusually early. However, both obligate carriers tested demonstrated normal waveforms with amplitudes and implicit times that were within the normal range.

Haplotype analysis demonstrated disease segregation with markers on Xp22, excluding all other known locations for XLRP (**RP2**, **RP3**, **RP6**, **RP15**, and **RP24**). The location of disease in this family also excluded potential allelism with X-linked progressive cone dystrophy loci (**COD1** and **COD2**) and **CSNBX** locus (Fig. 1) as well as several candidate genes known to be retinal specific that reside on the X-chromosome (**RCC1**, **X-arrestin**, **RCP**, and **GCP**).

Disease in this family is clinically distinct from the phenotype described by McGuire et al. for **RP15** and segregates with markers distal to this locus (Fig. 1). This novel locus for XLRP has been designated **RP23**, adding to the level of heterogeneity for XLRP loci and RP loci in general.

The critical interval maps to Xp22.32-Xp22.13 spanning approximately 15 cm, cosegregating with the markers DXS7103, DXS1224, and DXS999. Multipoint linkage analysis scores a maximum lod of 2.06 for the locus order DXS7103-XDXS1224-DXS999.

The **RP23** disease interval encompasses the retinoschisis locus. **XLRSl** is a neural retina-specific gene potentially involved in cell–cell interactions on membrane surfaces. **XLRSl** was the primary candidate for disease in the family in this study, and the gene was therefore screened for a mutation. All exons, and exon-intron boundaries were analyzed by sequencing an affected male patient. Although this result was negative, indicating that the gene responsible for retinoschisis does not cause **RP23**, mutations outside the coding region cannot be excluded.

The critical interval for **RP23** is large, and international efforts will ultimately result in the entire sequence of this genomic region being publicly available; therefore, we envisage that many genes will be implicated as candidates for this disease. Ongoing analysis of XLRP pedigrees may lead to further refinement of the **RP23** disease interval and facilitate identification of the causative gene.

**Acknowledgments**

The authors thank the participating family.

**References**


Novel Frameshift Mutations in the \textit{RP2} Gene and Polymorphic Variants

Dawn L. Thiselton\textsuperscript{1}, Ilaria Zito\textsuperscript{1}, Catherine Plant\textsuperscript{1}, Marcelle Jay\textsuperscript{1}, Shirley V. Hodgson\textsuperscript{2}, Alan C. Bird\textsuperscript{1}, Shomi S. Bhattacharya\textsuperscript{1}, and Alison J. Hardcastle\textsuperscript{1}

\textsuperscript{1}Departments of Molecular Genetics and Clinical Ophthalmology, Institute of Ophthalmology, University College London, London, UK; \textsuperscript{2}South Thames Regional Genetics Centre, Guy's Hospital, Division of Medical and Molecular Genetics, London, UK

Correspondence to Alison Hardcastle, Department of Molecular Genetics, Institute of Ophthalmology, University College London, 11-43 Bath Street, London EC1V 9EL, UK; Tel: 44 171 608 6945; Fax: 44 171 608 6863; E-mail: ahardcas@hgmp.mrc.ac.uk

Contract grant sponsor: Guide Dogs for the Blind and The Wellcome Trust; Contract grant number: 95-52A and 051733/Z/97

Communicated by Daniel F. Schorderet

Mutations in the \textit{RP2} gene located on Xp11.23 are associated with X-linked retinitis pigmentosa (XLRP), a severe form of progressive retinal degeneration which leads to complete loss of vision in affected males. To date, 14 different mutations in the \textit{RP2} gene have been reported to cause XLRP, the majority of which lead to a coding frameshift within the gene and predicted truncation of the protein product. We here report two novel frameshift mutations in \textit{RP2} identified in XLRP families by PCR-SSCP and direct sequencing, namely 723delT and 796-799del. Four single nucleotide polymorphisms (SNPs) within the coding region of \textit{RP2} are also described (105A>T, 597T>C, 844C>T, 1012G>T), the first polymorphisms to be reported within this gene of unknown function, two of which alter the amino acid sequence. The current study extends the XLRP mutation profile of \textit{RP2} and highlights non-pathogenic coding sequence variations which may facilitate both functional studies of the gene and analysis of intragenic allelic contribution to the phenotype.

\textcopyright{} 2000 Wiley-Liss, Inc.

KEY WORDS: X-linked retinitis pigmentosa; RP2; cSNP

\textbf{INTRODUCTION}

X-linked retinitis pigmentosa (XLRP) is a genetically heterogeneous retinal degeneration characterised by night blindness, pigmentary retinopathy and progressive loss of vision, owing to a primary degeneration of rod photoreceptor cells in the retina (Bird 1975). The two predominant XLRP loci, RP2 and RP3, map to Xp11.23 (Thiselton et al 1996) and Xp21.1 (Musarella et al 1990), and account for 20-25\% and 70\% XLRP accordingly from genetic mapping studies (Teague et al 1994; Ott et al 1990). Although XLRP genes have been cloned in both genomic intervals (respectively \textit{RP2}; Schwahn et al 1998 and \textit{RPGR}; Meindl et al 1996, Roepman et al 1996) mutations in these genes do not account for disease in all families linked to these intervals, and the genetic complexity of XLRP is expanding with additional loci reported in single families at Xp22.3-p22.2 (RP23);

Received 21 January 2000; Revised manuscript accepted 29 March 2000.
The recent cloning of RP2 (MIM# 312600; Schwahn et al 1998) has facilitated the screening of XLRP family collections for disease-associated mutations in this gene. RP2 consists of 5 exons and encodes a protein of 350 amino acids with no known function but with predicted homology to cofactor C, involved in the folding of β-tubulin (Schwahn et al 1998). To date, 14 different mutations have been described, causative of 10-20% XLRP in European and North American populations studied, and localised to exons 1, 2 and 4 of the RP2 gene (Schwahn et al 1998; Mears et al 1999; Hardcastle et al 1999). All families mapping exclusively to the RP2 locus by haplotype analysis have so far been found to have an underlying mutation in the RP2 gene (Hardcastle et al 1999).

We here report two novel RP2 mutations that were found in patients from British XLRP families. In the course of screening for disease-associated sequence alterations in RP2, we also identified four single nucleotide polymorphisms in the coding portion of the gene (cSNPs) that appear to have no clinical consequence.

MATERIALS AND METHODS

For this study, 21 unrelated XLRP patients were screened for mutations in the RP2 gene. X-linked inheritance was ascertained from pedigree and clinical evaluation and appropriate informed consent was obtained from patients and relatives. Genomic DNA was extracted from peripheral blood leukocytes using the Nucleon II kit (Scotlab Limited, Strathclyde, Scotland) according to manufacturer’s instructions. SSCP analysis was performed using intronic PCR primers (Schwahn et al 1998) flanking each of the 5 RP2 exons and intron/exon splice sites. PCR-SSCP protocols have previously been described (Hardcastle et al 1999). Each sample showing a conformational variant was sequenced directly from the purified PCR product (as in Hardcastle et al 1999), and further SSCP analysis performed for this product on other available family members and 50 additional European control samples (comprising 100 X chromosomes).

RESULTS

PCR-SSCP analysis followed by direct sequencing of the exonic fragments, including splice sites, for RP2 revealed 2 novel mutations in XLRP patients and 4 polymorphic variants in this study (Table la and lb). All were confirmed by repeat PCR to ensure against artefacts.

In family F21, a deletion mutation (723delT) of one base pair in exon 2 at position 723 (Hardcastle et al 1999) was detected, causing a frameshift and premature termination signal at codon 254. Family GT1 have a 4 bp deletion mutation (796-799delCAGA) in exon 3 resulting in a frameshift and premature termination signal at codon 271. Segregation of the mutation with disease was confirmed in both families by SSCP analysis (data not shown). SSCP analysis of a further 100 X chromosomes did not detect either deletion, suggesting that these mutations are not variant alleles in the general population and that they underly disease in these XLRP families. Both mutations are predicted to truncate the protein product, consistent with the majority of RP2 mutations to date, although this is the first report of a mutation in exon 3 causing XLRP. The detection of new independent mutations further emphasises a high new mutation rate for RP2 with little or no founder effect.

During the mutation screen, we identified four single nucleotide polymorphisms in the coding portion of the gene (cSNPs) that appear to represent polymorphic variants of RP2 (Table lb). Two of these can be detected with restriction enzymes, providing a simple method for distinguishing these alleles in genetic mapping or population studies (Table lb). The 844C>T transition in exon 3 (CGG— >TGG) at a hypermutable CpG site changes codon 282 from amino acid arginine (basic polar) to a tryptophan residue (nonpolar aromatic) and does not appear to be a common allele in the population (occurring at a frequency of 1.7%). Despite the non-conservative nature of this substitution, comparison with putative murine RP2 sequence compiled from orthologous ESTs would indicate that this amino acid residue is not evolutionarily conserved (data not shown). The fourth SNP lies in exon 5 (G 1012G>T) substituting an aspartic acid residue (acidic polar) for tyrosine (uncharged polar aromatic) at codon 338, but with allele frequencies of 69% (G) and 31% (T) in the control population sample, is deduced to be of no phenotypic consequence.
Novel Mutations and Polymorphisms in \textit{RP2}

Table 1a: Novel mutations identified in the RP2 gene

<table>
<thead>
<tr>
<th>Family</th>
<th>Intron/Exon</th>
<th>Nucleotide Change*</th>
<th>Amino Acid</th>
<th>Consequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>F21</td>
<td>Exon 2</td>
<td>723delT</td>
<td>242</td>
<td>Frameshift, 12 novel aa and 98 aa missing</td>
</tr>
<tr>
<td>GT1</td>
<td>Exon 3</td>
<td>796-799del</td>
<td>266</td>
<td>Frameshift, 3 novel aa and 80 aa missing</td>
</tr>
</tbody>
</table>

Table 1b: Summary of cSNPs identified in the RP2 gene

<table>
<thead>
<tr>
<th>EXON</th>
<th>Nucleotide Change*</th>
<th>Consequence</th>
<th>Frequency</th>
<th>Restriction Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>105A&gt;T</td>
<td>Val35Val</td>
<td>3%</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>597T&gt;C</td>
<td>Val199Val</td>
<td>42%</td>
<td>BsmFI, created</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Tth111I, created</td>
</tr>
<tr>
<td>3</td>
<td>844C&gt;T</td>
<td>Arg282Trp</td>
<td>2%</td>
<td>Sth132I, abolished</td>
</tr>
<tr>
<td>5</td>
<td>1012G&gt;T</td>
<td>Asp338Tyr</td>
<td>31%</td>
<td></td>
</tr>
</tbody>
</table>

* nucleotide position taken from Hardcastle et al 1999

DISCUSSION

In summary, we report 2 novel frameshift XLRP mutations and 4 cSNPs in the \textit{RP2} gene. This brings the total number of reported \textit{RP2} gene mutations so far to 16, highlights further domains or residues which may have important functional implications and will aid DNA-based diagnosis for XLRP. The majority of \textit{RP2} mutations create or lead to premature stop codons and result in absence of RP2 protein via mRNA instability or a severely truncated protein product which may be rapidly degraded. Although the \textit{RP2} gene is widely expressed, the exclusively retinal defect resulting from such mutations may be due to one particular interacting partner specific to normal retinal function. No mutations have to date been identified in exon 5 of \textit{RP2}, indicating that the C terminal region encoded by this exon may not significantly affect protein function. Indeed, the detection of 2 cSNPs in this study which alter amino acid residues in exon 3 (Arg282Trp) and exon 5 (Asp338Tyr) would suggest that non-conservative changes at less critical or conserved sites can be tolerated by RP2 without pathogenic effect. It has been shown experimentally that genes containing one or more cSNPs can give rise to two or more allelic mRNAs that differ in secondary or higher order structure (Shen et al 1999). It is feasible therefore that the presence or absence of specific cSNP alleles may influence the phenotypic expression of particular \textit{RP2} mutations. All 4 cSNPs described in this study lie outside the cofactor C-homologous domain and their discovery may focus experimental studies to unravel the biological function of the RP2 protein, in addition to facilitating genotype-phenotype correlations for XLRP.

REFERENCES


Novel Mutations of the RPGR Gene in RP3 Families

Ilaria Zito¹, Michael B. Gorin², Catherine Plant¹, Alan C. Bird¹, Shomi S. Bhattacharya¹, and Alison J. Hardcastle¹

¹Departments of Molecular Genetics and Clinical Ophthalmology, Institute of Ophthalmology, University College London, London, UK; ²Department of Ophthalmology, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

*Correspondence to Ilaria Zito, Department of Molecular Genetics, Institute of Ophthalmology, University College London, 11-43 Bath Street, London, EClV 9EL; Tel: 44 171 608 6987; Fax; 44 171 608 6863; E-mail; ilaria.zito@ucl.ac.uk

Contract grant sponsor: Wellcome Trust; Contract grant number: 051733/z/97

Communication by Haig H. Kazazian

X-linked retinitis pigmentosa is a severe form of retinal degeneration characterized by night blindness and visual field constriction, leading to complete blindness within the third decade of life. Mutations in the RPGR gene (retinitis pigmentosa GTPase regulator), located on Xp21.1 in the RP3 region, have been associated with an RP phenotype. Further to our previous mutation screening of RPGR in families segregating with the RP3 locus, we have expanded this study to include other 8 RP3 pedigrees. Here we report the results of this expanded study and the identification of five mutations in RPGR, four of which are novel (IVS6+5 G>A, 950-951 del AA, 963 T>C, EX5del) and one of which occurs in the donor splice site of intron I (IVS1+1 G>A). These findings bring the proportion of “RP3 genotypes” with a mutation in this gene to 27% (10/37). © 2000 Wiley-Liss, Inc.

KEY WORDS: retinitis pigmentosa GTPase regulator, RPGR, RP3, X-linked retinitis pigmentosa, XLRP, mutation analysis

INTRODUCTION

X-linked retinitis pigmentosa (XLRP) is a severe form of retinal degeneration manifesting in the first years of life of affected males due to haploinsufficiency. The first clinical symptoms of the disease include night blindness and reduced visual acuity as a consequence of the degeneration of rods in the periphery of the retina. As disease progresses cones start to degenerate producing a gradual impairment of the central vision which culminates in total blindness within the third or fourth decade of life.

Two genes causing retinitis pigmentosa have been cloned on the short arm of the X chromosome: the RPGR gene (retinitis pigmentosa GTPase regulator; MIM# 312610) on Xp21.1 in the RP3 region (Meindl et al. 1996) and the RP2 gene on Xp11.3 in the RP2 region (Shwann et al. 1998). By a combination of haplotype and linkage analysis these appear to be the two major XLRP loci accounting respectively for over 70% and for 20-25% of familial disease (Ott et al. 1980, Teague et al. 1994).

Worldwide effort in trying to unravel the complete genetic picture underlying this devastating disease led to the identification of another two XLRP loci in two single families. These two loci (RP23 on Xp22.3, Harcastle et al. submitted and RP24 on Xq24, Gieser et al. 1998) appear to account for only a small proportion of disease in XLRP.
The RP3 gene (RPGR) was cloned in 1996 (Meindl et al.) and is composed of 19 exons encoding an ubiquitously expressed protein product of 815 amino-acids. More recently a new retina-specific exon (exon 15a) has been identified (Kirschner et al. 1999). This exon is contained within intron 15 and it introduces a new stop codon producing a predicted gene product of 646 instead of 815 amino acids. The N-terminal portion of the gene shares sequence homology with RCC1, a regulator of chromosome condensation that interacts with small nuclear GTPases (Meindl et al. 1996, Roepman et al. 1996). It has recently been demonstrated that the RCC1-like domain is able to interact with the δ subunit of phosphodiesterase (δPDE, Linari et al. 1999). However the role of RPGR in the retina and its involvement in the etiology of retinitis pigmentosa have not yet been resolved.

Mutations in RPGR seem to account for approximately 20-25% of XLRP families (Meindl et al. 1996, Roepman et al. 1996, Buraczynska et al. 1997, Zito et al. 1999, Miano et al. 1999) which does not correlate with the large proportion of families segregating with the RP3 locus (70-75%). Additional coding sequences, the regulatory region of the gene or even a different gene close to RPGR may be involved in the proportion of familial disease still unaccounted for.

Further to our previous study (Zito et al. 1999) we here report the results of a screening carried out on eight new families segregating with the RP3 locus and the identification of five RPGR mutations, bringing the proportion of familial “RP3 genotypes” with a mutation in this gene to 27%.

**MATERIALS AND METHODS**

**XLRP families and controls**

By pedigree and clinical analysis an X linked form of RP was established. The diagnosis was based on ophthalmological tests including fundus examination, visual field assessment, fluorescein angiography and electroretinogram (ERG) recordings. Blood samples were collected for each family and DNA was extracted using the Nucleon II kit (Scotlab Limited, Strathclyde, Scotland) according to manufacturer instructions. The ethnic background of the families used in this study is as follows: family A is from the USA, family B is from New Zealand, families C, D and E are from the UK, the other three families that were screened in this study and in which mutations were not identified are from Belgium (two families) and from the UK (one family).

**Mutation detection by direct automated sequencing of PCR products**

All 19 exon fragments plus exon 15a of the RPGR gene were amplified with intronic primers described in Meindl et al. (1996) and in Kirschner et al. (1999). PCRs were carried out as previously described and special conditions were used for the amplification of exon 1 (Zito et al. 1999). PCR products were examined by agarose gel electrophoresis prior to sequencing. An aliquot of the amplification product (8 μl) was then purified by the addition of 1 U shrimp alkaline phosphatase (SAP; Amersham Life Science, Buckinghamshire, UK) and 1 U Exonuclease I (United States Biochemical) in SAP buffer, and incubation at 37°C for 30 min followed by 80°C for 15 min. Five μl of the purified DNA sample was then used for cycle sequencing using Big Dye Terminator cycle sequencing kit following manufacturers instructions. Reactions were then electrophoresed on an ABI 373A automated sequencer.

**Restriction enzyme digestion analysis**

Exon 8 PCR fragments were digested with BssSI restriction enzyme (New England BioLabs, Hertfordshire, UK). PCR reactions were carried out as described and 5μl of the PCR product were checked on an agarose gel prior to digesting. The remaining 20μl were brought to a final volume of 30μl by the addition of 10X NE Buffer 3 (New England BioLabs, Hertfordshire, UK), 6U BssSI enzyme and H2O. Digestion was performed at 37°C for 16-18 h.
RESULTS AND DISCUSSION

Direct sequencing of PCR products of all the 19 exons plus exon 15a of RPGR in eight families segregating with the RP3 locus, led to the identification of five mutations in five families (summarized in table 1).

Table 1: Mutations identified in the RPGR gene in four XLRP families with an RP3 genotype

<table>
<thead>
<tr>
<th>Family</th>
<th>Intron/Exon</th>
<th>Nucleotide alteration*</th>
<th>Amino acid no</th>
<th>Consequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Intron 1</td>
<td>IVS1+1 G&gt;A</td>
<td>-</td>
<td>Splicing</td>
</tr>
<tr>
<td>B</td>
<td>Intron 6</td>
<td>IVS6+5 G&gt;A</td>
<td>-</td>
<td>Splicing</td>
</tr>
<tr>
<td>C</td>
<td>Exon 8</td>
<td>950-951delAA</td>
<td>297</td>
<td>Frameshift</td>
</tr>
<tr>
<td>D</td>
<td>Exon 8</td>
<td>963 T&gt;C</td>
<td>302</td>
<td>C302R</td>
</tr>
<tr>
<td>E</td>
<td>Exon 8</td>
<td>Deletion (uncharacterized)</td>
<td>-</td>
<td>Deletion</td>
</tr>
</tbody>
</table>

*Position according to GenBank No. U57629

In family A, the mutation identified was a G>A transition 1 nucleotide downstream of the exon1/intron1 boundary (IVS1+1 G>A). Fresh blood of patients from this family is not available for RNA extraction hence we cannot demonstrate that this mutation gives rise to an aberrant splicing product. However exon 1 was sequenced in 80 control chromosomes and this change was not detected. In addition sequence analysis of exon 1 in the family demonstrated segregation of the mutation with disease and sequence analysis of the entire gene in the family did not reveal any other sequence alterations, therefore it is likely that the G>A transition observed affects normal splicing and is the causative mutation in this family. This mutation has also been already identified in another family (Zito et al 1999), therefore strengthening its role as a pathogenic mutation.

It is interesting to observe that in previous reports describing RPGR mutations, exon 1 was often excluded from the analysis, due to amplification difficulties. However this is the third potential mutation described for this exon out of a total of 44 mutations reported for the gene to date (Meindl et al. 1996, Roepman et al. 1996, Buraczynska et al. 1997, Zito et al. 1999, Miano et al. 1999, this study). The three exon 1 mutations are all within the 5' consensus donor splice site and two of them are identical. Considering also that exon 1 comprises less than 2% of the entire coding region of the gene, this donor splice site may be considered a potential mutational hot spot of the gene.

In family B carries a novel G>A change five nucleotide downstream of the exon6/intron6 boundary (IVS6+5 G>A). This nucleotide is part of the 5' consensus donor splice site and is 80% conserved in mammals (Horowitz and Krainer 1994), it is therefore likely that a sequence variation at this position affects normal RNA splicing. This change cosegregates with disease in the family and over 200 control chromosomes were sequenced to exclude the possibility of a polymorphism.

In family C a novel 2-bp deletion was observed at nucleotide position 950-951 (950-951delAA). This deletion causes a frameshift that is predicted to create 47 novel amino acids and a new stop codon that produces a protein product of 471 amino acids with 471 amino acids missing. The mutation segregates with the disease in the family.

In family D we observed a novel 963T>C transition, producing a Cys>Arg change at codon 302. This is a non-conservative amino acid change at a conserved codon. The T>C change creates a restriction site for BssS I enzyme which was used to demonstrate segregation of the mutation with disease in the family and lack of the same mutation in 100 control chromosomes. Sequencing of the entire coding region of the gene in this family showed no other variations.

In family E exon 8 failed to amplify in the affected members of the family. To exclude the possibility of sequence alterations within the primer sequences new primers were designed still spanning the exon and intron/exon boundaries (8b-F: atctgtgcgtgactgcc and 8b-R: ccagttctataataaataacag). The new primers again failed to amplify in the affected members of the family. The deletion needs further characterization, as unfortunately the full genomic sequence for RPGR is not available and PCR across exons 7-9 has been carried out with no resulting amplification product (probably due to the size of the introns). Southern blot analysis is currently
unachievable due to a shortage of DNA. However, the mutation segregates with the disease in the family and the remaining coding sequence of the gene presents no alterations.

With these three mutations in exon 8 the number of mutations described for this exon to date is 9 (Meindl et al. 1996, Roepman et al. 1996 Buraczynska et al. 1997, Miano et al. 1999) out of a total of 44 mutations (20.5%). This suggests that exon 8 has an important role in the etiology of RP, especially considering that three of these mutations are missense mutations pinpointing at least three crucial amino acids contained within this exon.

REFERENCES


Identification of novel \textit{RPGR} (retinitis pigmentosa GTPase regulator) mutations in a subset of X-linked retinitis pigmentosa families segregating with the \textit{RP3} locus

Received: 13 April 1999 / Accepted: 15 June 1999 / Published online: 21 July 1999

Abstract The X-linked form of retinitis pigmentosa (XLRP) is a severe disease of the retina, characterised by night blindness and visual field constriction in a degenerative process, culminating with complete loss of sight within the third decade of life. Genetic mapping studies have identified two major loci for XLRP: \textit{RP3} (70\%–75\% of XLRP) and \textit{RP2} (20\%–25\% of XLRP). The \textit{RPGR} (retinitis pigmentosa GTPase regulator) gene has been cloned within the \textit{RP3} genomic interval and its position is 10\%–20\% of XLRP families have mutations in this gene. Here, we describe a single-strand conformational polymorphism-based mutation screening of \textit{RPGR} in a pool of 29 XLRP families for which the disease segregates with the \textit{RP3} locus, in order to investigate the proportion of \textit{RP3} families with \textit{RPGR} mutations and to relate the results to previous reports. Five different new mutations have been identified: two splice site mutations for exon 1 and three frameshift mutations in exons 7, 10 and 11. The percentage of \textit{RPGR} mutations identified is 17\% (5/29) in our genetically well-defined population. This figure is comparable to the percentage of \textit{RP2} gene mutations that we have detected in our entire XLRP patient pool (10\%–15\%). A correlation of \textit{RPGR} mutations with phenotype in the families described in this study and the biochemical characterisation of reported mutations may provide insights into the function of the protein.

Introduction

Retinitis pigmentosa (RP) is a progressive hereditary disorder of the retina; it affects one in 4000 people and is caused by the gradual degeneration of photoreceptor cells. Manifestations of the disease include night blindness and visual field constriction with gradual loss of the central vision (Heckenlively 1988). Progression of the disease culminates for most of the patients in complete loss of sight and degeneration of the retina. The X-linked form of RP (XLRP) is the most severe, usually arising within the first two decades of life and progressing to total blindness within the third decade (Bird 1975).

Genetic analysis of XLRP families has shown an unexpectedly high degree of heterogeneity. There are two major loci (\textit{RP2} and \textit{RP3}), which have been mapped to the short arm of the X chromosome. Other loci include \textit{RP6}, \textit{RP15}, \textit{RP23} and \textit{RP24} (Ott et al. 1990; McGuire et al. 1995; Gieser et al. 1998; Hardcastle et al., in preparation).

Retinitis pigmentosa type 3 (\textit{RP3}) has been shown to be the predominant form of XLRP from genetic linkage studies and segregates with the disease in 60\%–90\% of affected families (Ott et al. 1990; Teague et al. 1994; Fujita et al. 1997).

In 1996, a gene named \textit{RPGR} (retinitis pigmentosa GTPase regulator) was cloned from the \textit{RP3} region (Meindl et al. 1996; Roepman et al. 1996). The \textit{RPGR} gene is ubiquitously expressed and consists of 19 exons. The putative \textit{RPGR} gene product of 815 amino acids shows a high degree of homology to the \textit{RCC1} (regulator of chromosome condensation) protein. This region of homology is localised within the N-terminal half of the predicted protein, includes exons 3–9 plus part of exons 2 and 10 (Fig. 1) and consists of six complete tandem repeats of 52–54 amino acids (preceded by one incomplete repeat) with a high degree of conservation across vertebrates. \textit{RCC1} regulates a GTPase known as Ras-related nuclear protein by its tandemly arranged repeats that act as the catalytic site for the exchange of guanine nucleotides (Bischoff and Ponstingl 1991; Klebe et al. 1995). Similarly, the \textit{RPGR}
gene product may interact with a retina-specific Ras-related nuclear protein. Recently, it has been shown that the RCC1-like domain of RPGR interacts with the β subunit of rod cyclic GMP phosphodiesterase (PDEβ) and that disruption of this interaction may be the cause of, or may contribute to, retinal disease (Linari et al. 1999).

Since the identification of RPGR, various groups have independently reported disease-causing mutations within this gene in XLRP patients. Surprisingly, the percentage of mutations found is 15%–20% of XLRP (Buraczynska et al. 1997; Meindl et al. 1996; Roepman et al. 1996), which is a remarkably low frequency compared with the high percentage of familial disease linked to the RP3 locus (60%–90%; Ott et al. 1990; Teague et al. 1994).

The reports describing RPGR mutations however generally lack information on the genetic mapping of disease in the families used in their studies, the exception being Fujita et al. (1997). Hence, the consensus 15%–20% of mutations reported is described as a proportion of total XLRP families, some of which may have little evidence of disease segregation at the RP3 locus.

In this study, 29 previously characterised families in which the disease segregates with this genetic interval have been screened in order to investigate the proportion of disease caused by mutations in RPGR in genetically defined RP3 families.

Materials and methods

XLRP families and controls

An X-linked form of RP (with male family members being more severely affected than female carriers and no male-to-male transmission) was established by pedigree and clinical analysis. The diagnosis was based on ophthalmological tests, including fundus examination, visual field assessment, fluorescein angiography and electroretinogram recordings. Blood samples were collected from each family and DNA was extracted by using the Nucleon II kit (Scotlab, Strathclyde, Scotland) according to the manufacturer’s instructions.

Extensive haplotype analysis identified 29 XLRP pedigrees with disease segregation at the RP3 locus (data not shown). These families have genomic disease intervals that encompass the RPGR gene and will be referred to as “RP3 genotype” families throughout this text. These 29 RP3 genotype families were analysed for RPGR mutations in this study. For each genetic alteration identified, a number of control chromosome ranging from 30 to 80 were also screened.

Mutation detection

by single-strand conformation polymorphism analysis

All 19 exon fragments of the RPGR gene were amplified with intronic primers described in Meindl et al. (1996). Polymerase chain reactions (PCRs) were carried out in 10 µl reactions in the presence of 0.2 µCi α-32P-dCTP (Amersham Life Science, Buckinghamshire, UK), 0.2 U Taq polymerase (Biotaq; Bioline, London, UK), 200 µM each of dATP, dGTP and dTTP and 20 µM dCTP, 0.2 µM each primer, 1 × KCl reaction buffer (Bioline) including 1.5 mM MgCl2. Amplification conditions were: 95°C for 5 min, followed by 35 cycles of denaturation at 95°C for 15 s, annealing at the primer-specific temperature for 15 s (Meindl et al. 1996), extension at 72°C for 30 s and a final extension for 5 min at 72°C. Special PCR conditions were used to amplify exon 1: 1 × NH reaction buffer (Bioline), with the addition of 1.0 mM MgCl2, was used instead of KCl buffer. To help release DNA strands, 5% formamide and 10% glycerol were also added to the PCR mix, as exon 1 has a high GC content. Moreover, denaturation, annealing and extension times were all increased to 45 s and the number of cycles was reduced from 35 to 30.

Amplified DNA was mixed 1:1 with formamide sample buffer (95% deionised formamide, 0.05% bromophenol blue, 0.05% xylene cyanole, 20 mM EDTA) containing 20 mM NaOH. The products were analysed for single-strand conformational polymorphism (SSCP) by using non-denaturing 0.5% MDE gels (mutation-detection-electrophoresis gels; Flowgen Instruments, Staffordshire, UK) in 0.6 × TBE buffer (1 × TBE buffer = 0.09 M TRIS-borate, 0.002 M EDTA, pH 8.3) run at 8–10 W constant power for 14–18 h. The gels were dried onto Whatman paper and analysed by autoradiography.

Direct automated sequencing of PCR products

PCR products of samples showing a bandshift on SSCP were checked by agarose gel electrophoresis prior to sequencing. The remaining amplification product was purified with centricron concentrators (Amicon Limited, Gloucestershire, UK) following the manufacturer’s instructions. A 5 µl sample of the purified DNA was cycle-sequenced (with primers used for amplification) with an ABI prism Ready Reaction Dye Terminator cycle sequencing kit (FS kit, Perkin-Elmer Applied Biosystems, Warrington, UK) following the manufacturer’s instructions. Reactions were then electrophoresed on an ABI 373A automated sequencer.

Results

A total of 10 genetic alterations was identified in seven different exons of the RPGR gene in 29 XLRP families segregating with the RP3 locus. Five of the ten changes were classified as polymorphisms (Table 1), either because they had previously been described as polymorphisms or because they were also present in unaffected control chromosomes. The remaining five changes were classified as potential novel mutations found in four different exons. Each mutation is given with its corresponding change at the amino acid and/or the nucleotide level in Table 2. These results indicate that, from a subset of 29 XLRP families with an RP3 genotype, 17% (5/29) display an RPGR mutation.
Table 1 Polymorphisms identified in the RPGR gene

<table>
<thead>
<tr>
<th>Intron/exon</th>
<th>DNA Position</th>
<th>Change</th>
<th>Function Amino acid no. Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exon 3</td>
<td>286</td>
<td>G→T</td>
<td>76 Ser→Ile</td>
</tr>
<tr>
<td>Exon 10</td>
<td>1223</td>
<td>G→A</td>
<td>388 Silent</td>
</tr>
<tr>
<td>Exon 11</td>
<td>1333</td>
<td>G→A</td>
<td>425 Arg→Lys</td>
</tr>
<tr>
<td>Exon 14</td>
<td>1765</td>
<td>G→A</td>
<td>566 Gly→Glu</td>
</tr>
<tr>
<td>Intron 18</td>
<td>2300 plus 11</td>
<td>T→C</td>
<td>Intronic</td>
</tr>
</tbody>
</table>

Table 2 Mutations identified in the RPGR gene in XLRP patients with an RP3 genotype

<table>
<thead>
<tr>
<th>Family</th>
<th>Intron/exon</th>
<th>DNA Position</th>
<th>Change</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>RP106</td>
<td>Intron 1</td>
<td>86 plus 1</td>
<td>G→A</td>
<td>Putative splicing</td>
</tr>
<tr>
<td>RP1120</td>
<td>Intron 1</td>
<td>86 plus 5</td>
<td>G→A</td>
<td>Putative splicing</td>
</tr>
<tr>
<td>F75</td>
<td>Exon 7</td>
<td>807–808</td>
<td>2-bp deletion</td>
<td>250 Frameshift</td>
</tr>
<tr>
<td>XLRP90</td>
<td>Exon 10</td>
<td>1297</td>
<td>2-bp insertion</td>
<td>413 Frameshift</td>
</tr>
<tr>
<td>RP87</td>
<td>Exon 11</td>
<td>1435–1436</td>
<td>2-bp deletion</td>
<td>459 Frameshift</td>
</tr>
</tbody>
</table>

*According to GenBank no. U57629

Direct automated sequencing of the DNA from patients that presented a shift on an SSCP gel led to the identification of three frameshift mutations. Family F75 (Fig. 2A) has a 2-bp deletion in exon 7 corresponding to nucleotides 807 and 808 in the normal sequence. In family XLRP90 (Fig. 2B), two base pairs have been inserted at nucleotide 1297 of the normal sequence. Family RP87 (Fig. 2C) has a 2-bp deletion in exon 11 corresponding to nucleotides 1435 and 1436 in the normal sequence. In all three cases, the resulting frameshift gives rise to a considerably truncated protein leading to lack of function. Segregation of the mutations with disease was demonstrated in each family by SSCP analysis (Fig. 3 shows the results obtained with family XLRP90).

Fig. 2A–E. Electropherograms depicting patient mutations and normal sequences in the RPGR gene. Arrows depict the sites of mutations. A A 2-bp base pair deletion (GT) in exon 7 in an affected male from family F75 and the normal sequence of the corresponding region. B A 2-bp insertion (AG) in exon 10 in an affected male from family XLRP90 and the normal sequence of the corresponding region. C A 2-bp deletion (TC) in exon 11 in an affected male from family RP87 and the normal sequence of the corresponding region. D A G-to-A change in the donor splice site at the first nucleotide downstream of the exon 1/intron 1 boundary in an affected male from family RP106 and the normal sequence of the corresponding region. E A G-to-A change in the donor splice site at the fifth nucleotide downstream of the exon 1/intron 1 boundary in an affected male from family RP1120 and the normal sequence of the corresponding region.
In families RP106 and RP1120, a nucleotide change was found in the donor splice site respectively at the first and at the fifth nucleotide downstream of the exon/intron boundary of exon 1 (Fig. 2D, E). In both families, the segregation of the nucleotide change with disease was demonstrated by sequencing three affected and three unaffected males (data not shown). Neither of these sequence alterations was identified in 80 sequenced control chromosomes (data not shown).

Discussion

Linkage and haplotype analysis of XLRP families has shown segregation of the disease with the RP3 region in 60%-90% of cases (Teague et al. 1994; Meindl et al. 1996), whereas the reported percentage of mutations in the RPGR gene ranges between 10% and 20% (Buraczynska et al. 1997; Meindl et al. 1996; Roepman et al. 1996). In this study, mutation screening conducted for 29 RP3 genotype pedigrees has largely confirmed this data. A total of five different potential mutations has been identified in this subset of families, defining a mutation rate in RPGR of 17% (5/29).

The five mutations reported here (from a pool of predominantly British families) are all different and new. These data therefore suggest a high new mutation rate for the RPGR gene and no founder effect, supporting the assertion of Fujita et al. (1996) that most RPGR mutations are of independent origin.

Three of the mutations found cause a frameshift, which leads to a severely truncated polypeptide and which clearly correlates with a loss of function of the protein. Of these, one occurs in exon 7 with a predicted 531 amino acids missing and is contained in the RCC1 domain (see Fig. 1). Another occurs in exon 10 with 391 amino acids missing, the third being found in exon 11 with 353 amino acids missing (Table 2).

As exon 1 has proved to be difficult to amplify in previously published reports, mutation analysis of the RPGR gene has been conducted predominantly on the remaining 18 exons and mutations in this exon may have been largely undetected. In this study, analysis of exon 1 has been successful and two patients (from a subset of 29 patients) have been shown to have sequence alterations in the exon 1 PCR product. The sequence changes identified are single nucleotide substitutions in the splice donor region, one at the first and one at the fifth nucleotide downstream of the exon/intron boundary. SSCP analysis of the entire gene from these patients has not revealed any other sequence alterations. Definitive biological evidence that these nucleotide changes cause disease in these families requires further experimentation, initially by reverse transcription/PCR. However, fresh patient blood is unfortunately unavailable. Nevertheless, all intron splice acceptor and donor sequences of the human RPGR gene follow the GT-AG rule and the first nucleotide of a donor splice sequence is 100% conserved (Horowitz and Krainer 1994); hence, a G→A transition at this site is highly likely to affect RNA splicing. The fifth nucleotide is 80% conserved; a G→A transition at this position also probably affects RNA splicing. These predicted alterations to the protein product have not been demonstrated experimentally and may not be pathogenic. To exclude the possibility of polymorphic alternative splicing within the gene, an occurrence that could give rise to a variant, less common, but still functional protein, 80 control chromosomes were sequenced. No sequence alterations were detected in the control samples but the number of control chromosomes tested must be increased before non-pathogenic sequence variation can be excluded.

The level of pathogenic and non-pathogenic sequence variation in the RPGR gene warrants further investigation. Evidently, a search for sequence variations occurring in, or in proximity to, exon 1 could result in the increased detection of RPGR mutations associated with XLRP.

All mutations reported to date lie within exons 2-15, with the exclusion of exon 12 for which no mutations have been reported, together with exons 16, 17, 18 and 19 (Fig. 1, Table 3). The potential splice mutations described here in intron 1 are the first reported and the most N-terminal of the sequence alterations found to date.

The correlation between the nature of the RPGR gene defect and the resulting phenotype has been investigated (Andreasson et al. 1997); a more severe visual impairment has been shown for those individuals carrying a mutation in the RCC1-homologous domain. Most of the mutations reported so far lie within this region of homology (particularly exons 7 and 8, see Fig. 1), which is obviously a major functional part of the protein, as demonstrated by its interaction with PDE8 (Linari et al. 1999). The role of RPGR in phototransduction or maintenance of retina function is still unclear and interaction of RPGR with another molecule, perhaps a retina-specific GTPase, should...
be investigated. Several mutations have been localised outside the RCC1-homologous region; the biochemical characterisation of these mutations should result in an understanding of other important domains of the protein related to its function, its involvement in protein-protein interaction or DNA-binding sites.

A description of the number and variety of RPGR mutations is important for genotype-phenotype correlation studies. In turn, these studies may help provide insights into the function(s) of the protein and the physiological consequences of various mutations. The recent identification of the RP2 gene (Schwahn et al. 1998) is also important and complementary to these studies. Positive implications for the molecular diagnosis of patients come from RPGR and RP2 mutation screening and the possibility of phenotypically distinguishing these two types of XLRP. As a result, this research should aid genetic counselling in the clinic. The work described in this paper confirms previous data relating to the percentage of RPGR mutations detected in XLRP patients. Various explanations may account for the low mutation rate detected. Mutations in the promoter region or in an as yet unidentified part of the gene or an alternatively spliced transcript may increase the proportion of XLRP attributable to RPGR mutations; however, an equally likely explanation would be microheterogeneity at this locus. The possibility that an as yet unidentified gene or genes tightly linked to the RP3 region may be implicated in XLRP cannot be excluded.

By working on a subset of genetically characterised families in which the disease segregates with the RP3 genomic interval, this study gives supportive evidence of microheterogeneity at this locus, since only 17% of families with an RP3 genotype have RPGR mutations. Moreover, mutation screening of the RP2 gene (Hardcastle et

---

### Table 3 Spectrum of reported RPGR mutations

<table>
<thead>
<tr>
<th>Intron/exon</th>
<th>Position</th>
<th>Change</th>
<th>Function</th>
<th>No. of cases</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intron 1</td>
<td>86 plus 1</td>
<td>G→A</td>
<td>Putative splicing</td>
<td>1</td>
<td>This paper</td>
</tr>
<tr>
<td>Intron 1</td>
<td>86 plus 5</td>
<td>G→A</td>
<td>Putative splicing</td>
<td>1</td>
<td>This paper</td>
</tr>
<tr>
<td>Exon 2</td>
<td>213</td>
<td>G→T</td>
<td>Gly→Stop</td>
<td>3</td>
<td>Buraczyńska et al. (1997), Jacobson et al. (1997), Meindl et al. (1996)</td>
</tr>
<tr>
<td>Exon 3</td>
<td>238</td>
<td>G→T</td>
<td>Gly→Val</td>
<td>2</td>
<td>Buraczyńska et al. (1997)</td>
</tr>
<tr>
<td>Exon 4</td>
<td>353</td>
<td>C→A</td>
<td>Frameshift</td>
<td>1</td>
<td>Miano et al. (1998)</td>
</tr>
<tr>
<td>Exon 4</td>
<td>354</td>
<td>2-bp insertion</td>
<td>99</td>
<td>Frameshift</td>
<td>1</td>
</tr>
<tr>
<td>Intron 4</td>
<td>369 plus 3</td>
<td>A→G</td>
<td>Splicing</td>
<td>1</td>
<td>Buraczyńska et al. (1997)</td>
</tr>
<tr>
<td>Exon 5</td>
<td>430</td>
<td>1-bp deletion</td>
<td>124</td>
<td>Frameshift</td>
<td>1</td>
</tr>
<tr>
<td>Exon 5</td>
<td>449</td>
<td>T→G</td>
<td>Phe→Cys</td>
<td>1</td>
<td>Roepman et al. (1996)</td>
</tr>
<tr>
<td>Intron 5</td>
<td>528 plus 1</td>
<td>G→T</td>
<td>Splicing</td>
<td>1</td>
<td>Dry et al. (1999)</td>
</tr>
<tr>
<td>Exon 6</td>
<td>545</td>
<td>1-bp deletion</td>
<td>162</td>
<td>Frameshift</td>
<td>1</td>
</tr>
<tr>
<td>Exon 6</td>
<td>551</td>
<td>G→A</td>
<td>Trp→Stop</td>
<td>1</td>
<td>Buraczyńska et al. (1997)</td>
</tr>
<tr>
<td>Exon 6</td>
<td>640</td>
<td>G→A</td>
<td>Trp→Stop</td>
<td>1</td>
<td>Meindl et al. (1996)</td>
</tr>
<tr>
<td>Exon 7</td>
<td>703</td>
<td>G→T</td>
<td>Gly→Val</td>
<td>1</td>
<td>Meindl et al. (1996)</td>
</tr>
<tr>
<td>Exon 7</td>
<td>762</td>
<td>C→T</td>
<td>Pro→Ser</td>
<td>1</td>
<td>Roepman et al. (1996)</td>
</tr>
<tr>
<td>Exon 7</td>
<td>765</td>
<td>C→T</td>
<td>Gln→Stop</td>
<td>1</td>
<td>Buraczyńska et al. (1997)</td>
</tr>
<tr>
<td>Exon 7</td>
<td>807</td>
<td>T→C</td>
<td>Cys→Arg</td>
<td>1</td>
<td>Meindl et al. (1996)</td>
</tr>
<tr>
<td>Exon 7</td>
<td>807–808</td>
<td>2-bp deletion</td>
<td>250</td>
<td>Frameshift</td>
<td>1</td>
</tr>
<tr>
<td>Intron 7</td>
<td>837 plus 5</td>
<td>G→A</td>
<td>Splicing</td>
<td>1</td>
<td>Buraczyńska et al. (1997)</td>
</tr>
<tr>
<td>Exon 8</td>
<td>882</td>
<td>G→A</td>
<td>Gly→Ser</td>
<td>2</td>
<td>Roepman et al. (1996)</td>
</tr>
<tr>
<td>Exon 8</td>
<td>928</td>
<td>1-bp deletion</td>
<td>290</td>
<td>Frameshift</td>
<td>1</td>
</tr>
<tr>
<td>Exon 8</td>
<td>945–959</td>
<td>15-bp deletion</td>
<td>296–300</td>
<td>5-aa missing</td>
<td>1</td>
</tr>
<tr>
<td>Exons 8–10</td>
<td>–</td>
<td>Deletion</td>
<td>–</td>
<td>Splicing and frameshift</td>
<td>1</td>
</tr>
<tr>
<td>Exon 10</td>
<td>1179</td>
<td>G→T</td>
<td>374</td>
<td>Frameshift</td>
<td>1</td>
</tr>
<tr>
<td>Exon 10</td>
<td>1297</td>
<td>2-bp insertion</td>
<td>413</td>
<td>Frameshift</td>
<td>1</td>
</tr>
<tr>
<td>Intron 10</td>
<td>1304 plus 3</td>
<td>A→G</td>
<td>Splicing</td>
<td>1</td>
<td>Fujita et al. (1997)</td>
</tr>
<tr>
<td>Exon 11</td>
<td>1320–1338</td>
<td>19-bp deletion</td>
<td>421</td>
<td>Frameshift</td>
<td>1</td>
</tr>
<tr>
<td>Exon 11</td>
<td>1345–1436</td>
<td>2-bp deletion</td>
<td>459</td>
<td>Frameshift</td>
<td>1</td>
</tr>
<tr>
<td>Exon 11</td>
<td>1461–1464</td>
<td>4-bp deletion</td>
<td>468</td>
<td>Frameshift</td>
<td>1</td>
</tr>
<tr>
<td>Exon 13</td>
<td>1571–1572</td>
<td>2-bp deletion</td>
<td>504</td>
<td>Frameshift</td>
<td>1</td>
</tr>
<tr>
<td>Intron 13</td>
<td>1632 minus 8</td>
<td>A→G</td>
<td>Splicing</td>
<td>1</td>
<td>Fujita et al. (1997)</td>
</tr>
<tr>
<td>Exon 14</td>
<td>1641–1644</td>
<td>4-bp deletion</td>
<td>528</td>
<td>Frameshift</td>
<td>1</td>
</tr>
<tr>
<td>Exons 14–15</td>
<td>–</td>
<td>Deletion</td>
<td>–</td>
<td>In frame deletion</td>
<td>1</td>
</tr>
</tbody>
</table>
al. 1999) has identified only one of the 29 families analysed in this study with a mutation in RP2. Therefore, for the majority of families described here (24/29), the cause of the disease is unknown. Although not yet fully characterized, heterogeneity for XLRP is much greater than originally expected and further research in new and established XLRP families and their corresponding disease intervals should help define new XLRP loci (e.g. RP24, Gieser et al. 1998; RP23, Hardcastle et al., in preparation) and identify new candidate genes for this disease.

Acknowledgements The authors wish to thank all the participating families who have contributed to this study. This research was supported by the Wellcome Trust (grant 051733/z/97 to A.J.H), the British RP Society, the Eye and Ear Foundation of Pittsburgh, and Research to Prevent Blindness, N.Y.

References


