A GENETIC AND PHYSICAL MAPPING STUDY OF THE RETINITIS PIGMENTOSA CRITICAL REGION ON THE PROXIMAL SHORT ARM OF THE X CHROMOSOME

DAWN L. THISELTON
B.Sc.
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

1997
Retinitis Pigmentosa (RP) is an inherited disease of the retina, in which progressive degeneration of the photoreceptors leads to night blindness and ultimately complete loss of vision. Genetic linkage studies have identified two major loci (RP2 and RP3) mapping to the short arm of the X chromosome which are implicated in the X-linked form of the disease: RP2 maps to Xp11.3-Xp11.22 and RP3 maps to Xp21.

The precise localisation of RP3 in Xp21 has been facilitated by the identification of cytogenetic abnormalities in affected patients and extensive physical mapping studies to a 150kb region flanked by OTC and DXS1110. At the outset of this study, several groups were actively engaged in analysing transcripts from this region as potential candidates for the disease and a gene RPGR has since been established as the site of a proportion of RP3 mutations. The focus of this thesis was to more definitively characterise the RP2 gene critical region in Xp11 by a combination of genetic and physical mapping methods, as a first step toward positional cloning of the defective gene.

The precise localisation of RP2 has been hampered since its initial assignment to Xp11 by a lack of associated cytogenetic anomalies, few informative recombination events and a paucity of polymorphic markers in proximal Xp. In order to generate novel genetic markers mapping to the region, 13 human X chromosome-specific cosmids known to contain (CA)n microsatellites were FISH mapped to X chromosome sub-regions and polymorphic microsatellites isolated from 6 that mapped to proximal Xp. This thesis describes the isolation and characterisation of one such microsatellite (DXS556) from cosmid HX20 that mapped to Xp11.4.

To determine the relationship between the new microsatellites and existing genetic markers mapping to proximal Xp, these markers were used to genetically characterise a panel of 14 XLRP families by linkage and haplotype analysis. A combination of multiply informative crossovers and physical mapping via YAC STS-content analysis enabled the successful positioning and ordering of the new microsatellites and their incorporation into the well-established framework map of proximal Xp.

In an attempt to determine the particular XLRP locus segregating in each of the XLRP families, further genotyping was performed using additional markers as they became available through the Genome Database (GDB). Key recombination events detected by haplotype analysis resulted in the assignment of one family as RP3 and four as RP2. Finer localisation of the RP2 gene was achieved by extensive genetic analysis incorporating the most recently available CEPH/Genethon markers to generate dense haplotypes spanning Xp11.3-Xp11.22. Two multiply informative crossovers in one family defined new proximal and distal boundaries for the RP2 critical region and significantly reduced its size from ~13cM to ~5cM.

During the course of this study, mutations in the TIMP-3 (tissue inhibitor of metalloproteinases-3) gene were found to cause the retinal degenerative disease Sorsby's
Fundus Dystrophy. TIMP-3 is a member of a family of proteins that play an integral role in connective tissue homeostasis. A related gene, TIMP-1, lay within the RP2 critical interval and thus immediately became a positional candidate for RP2. The genomic organisation of TIMP-1 was completed prior to mutation screening by PCR and direct sequencing of the entire coding region, exon-intron boundaries and 5' untranslated region, in affected and unaffected males from three RP2 families. Although a neutral polymorphism was discovered in exon 5, no disease-associated sequence alterations were found, suggesting that TIMP-1 is unlikely to play a causal role in the etiology of XLRP.

Further refinement of the RP2 critical region by genetic means will require the identification of more recombination events within the interval (i.e. new RP2 families or further members of existing ones) and/or mapping of additional genetic markers to this region. In parallel with such ongoing genetic studies, a YAC-based physical map of Xp11.3-11.22 has been constructed using an STS-content mapping strategy, as a tool from which to isolate novel markers and candidate genes. The map to date is centred around two contigs: a ~1.5Mb contig encompassing DXS7-MAO in Xp11.3 and a ~3Mb contig spanning the newly defined proximal RP2 boundary in Xp11.23 and extending distally within the RP2 critical interval. The high redundancy of the map has enabled the physical linkage and ordering of many STSs, genes and ESTs in Xp11.3-11.23 to a density of ~1 per 40kb. The inclusion of most known microsatellites from the region has enabled integration of the physical and genetic maps and provided precise locus assignments for several microsatellites. The use of a 'meiotic breakpoint mapping panel' composed of key recombinants from the XLRP families provided supplementary order information for key microsatellites to facilitate contig orientation in Xp11.23. YAC insert-terminal isolation has generated 14 novel region-specific STSs, two of which are identical by database similarity searches to ESTs, thus adding novel gene sequences to the physical map of Xp11.23. YAC-end STSs also proved instrumental in confirming overlaps between YACs and are being used in a 'chromosome walking' approach to bridge the remaining gaps in the contig.

This project demonstrates the effectiveness of concurrent genetic and physical mapping approaches to continually direct further efforts in the positional cloning of RP2. With such a large critical interval, in a region known to be gene-rich, mapping and ordering expressed sequences and polymorphic markers is an important adjunct to its genetic refinement. Integration of the genetic and physical data presented in this study has successfully excluded several potential candidate genes from a causal role in RP2, and provided a high resolution mapping resource with which to investigate further candidate genes for RP2 and other genetic diseases that localise to proximal Xp.
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.

Dawn L. Thiselton B.Sc.
Acknowledgements

A big thankyou to my supervisor, Professor Shomi Bhattacharya, for giving me the opportunity to undertake this research project, and for his constant encouragement and advice throughout the practical and literary aspects of this work. I also thank Dr. Susan Lindsay for excellent guidance throughout the early stages of this thesis.

My biggest thanks must go to two wonderful friends. Firstly, Dr. Alison Hardcastle, with whom I have been very lucky to work, for her enthusiasm, optimism and general brilliance, and for her invaluable help and constructive criticism of the many drafts of this thesis. I couldn’t have done it without you! Secondly, my flatmate Liz Evans, for putting up with me over the last three years and for her amazing support and understanding. Thanks to both of you for being there and for being you.

Many thanks to several people who deserve a special thankyou for extra help during the final stages (computing, formatting, editing, hugs!): Binoy, Kamal, Steph, Jimbob, Karen, Zara, Zoe, Reshma, Chris, Cheryl and Peter, and my thanks also go to everyone else in the department for making it all so interesting and for many a good beano.

I would also like to thank Nessa, Kirsty, Belinda, Andy, Rita and Neil for their much-treasured friendship throughout this thesis, and for still being there at the end of it! Also to Beavis and Mr Tooty for providing furry distractions.

Last but not least, I thank with all my heart my family, especially Mum and Geoff, Dad and Gloria, Linda and Phil, for their priceless love and support all the way along, and also my gorgeous nieces and nephew, Cassie, Charlotte and Daniel for pretending they don’t think I’m weird. Fond love and thanks also to Nanny and Grandma for helping me through the difficult bits with their cheery ways.
Dedication

This thesis is dedicated to Linda

“For there is no friend like a sister
In calm or stormy weather,
To cheer one on one’s tedious way,
To fetch one if one goes astray,
To lift one if one totters down,
To strengthen whilst one stands.”

*Goblin Market*
Christina Rossetti
Publication list


# TABLE OF CONTENTS

## CHAPTER 1

**General Introduction**

1.1 The identification of genes in human inherited disease 19  
1.1.1 Positional cloning 19  
1.2 Genetic Mapping 21  
1.2.1 Meiotic recombination 21  
1.2.2 Genetic map distance 22  
1.2.3 DNA polymorphism and the development of human genetic markers 23  
1.2.3.1 Microsatellites 24  
1.2.3.1.1 Mutation rate and mechanism of microsatellite length variation 24  
1.2.4 Marker utility: the degree of polymorphism 25  
1.2.5 Genetic maps of the human genome 26  
1.2.6 Pedigree-based genetic linkage analysis 27  
1.2.7 Refining the localisation of a disease gene 28  
1.3 Physical mapping 30  
1.3.1 Cytogenetic methods of physical mapping 30  
1.3.1.1 Somatic cell hybrids 31  
1.3.1.2 In situ hybridisation 33  
1.3.2 Molecular methods of physical mapping 34  
1.3.2.1 Pulsed field gel electrophoresis for size separation of large DNA molecules 34  
1.3.2.2 Generating long-range physical maps of cloned DNA 35  
1.3.2.3 Yeast artificial chromosomes as a genome mapping tool 35  
1.4 Identification of gene sequences from a defined genomic interval 37  
1.4.1 Northern blot analysis 37  
1.4.2 Cross-species sequence homology 37  
1.4.3 Detection of CpG islands 37  
1.4.4 cDNA selection 38  
1.4.5 Exon trapping 38  
1.4.6 Computer-based DNA sequence analysis 39  
1.5 Whole genome mapping efforts: The Human Genome Project 39  
1.5.1 A common language for physical maps: the sequenced-tagged site 40  
1.5.2 Current physical maps of the human genome 40  
1.5.3 The final phase: sequencing the human genome 41
1.6 Positional candidate approach to the isolation of human disease genes

1.7 Inherited retinal degenerations

1.7.1 Structure of the retina

1.7.1.1 Neural retina

1.7.1.1.1 Photoreceptors

1.7.1.2 Retinal pigment epithelium

1.7.2 The rod phototransduction cascade

1.7.2.1 Recovery of the dark state and the role of calcium

1.7.3 Retinal degeneration due to mutations in the photoreceptor genes

1.7.3.1 Rhodopsin

1.7.3.2 Rod cGMP phosphodiesterase α and β subunits (PDEA and PDEB)

1.7.3.3 cGMP-gated channel

1.7.3.4 Peripherin/RDS

1.7.3.5 ROM1

1.7.3.6 Retinal guanylate cyclase (RETGC)

1.7.4 The mechanism of photoreceptor cell death

1.8 The human X chromosome

1.8.1 Mapping X-linked diseases

1.8.2 Mutations in X-linked genes

1.8.3 The genetic and physical map of the human X chromosome

1.8.4 The current map of proximal Xp: Xp21.1-Xcen

1.8.4.1 Cloned genes in proximal Xp

1.8.4.2 Recent additions to the proximal Xp gene map

1.9 Genetic eye diseases mapping to proximal Xp

1.9.1 Norrie disease

1.9.2 X-linked congenital stationary night blindness

1.9.3 Aland island eye disease (AIED)

1.9.4 X-linked progressive cone dystrophy

1.9.5 X-linked retinitis pigmentosa

1.9.5.1 Genetic heterogeneity in XLRP

1.9.5.2 Phenotype/genotype correlations in XLRP

1.9.5.3 Recent progress in positional cloning of the RP3 and RP2 genes

1.9.5.3.1 RP3 - Physical mapping studies and isolation of candidate genes

1.9.5.3.2 RP2 - Further genetic mapping studies to confirm and refine locus location

1.10 Aims of this thesis
CHAPTER 2

Materials and Methods

2.1 Fluorescence in situ hybridisation (FISH)
  2.1.1 Preparation of metaphase spreads
  2.1.2 Labelling of DNA probes
  2.1.3 Column purification and precipitation of the labelled probe
  2.1.4 Biotin incorporation test
  2.1.5 In situ hybridisation
    2.1.5.1 Slide treatment
    2.1.5.2 Hybridisation mix and probe application
      -repetitive probe DXZ1
    2.1.5.3 Hybridisation mix and probe application
      -cosmid probes
  2.1.6 Post hybridisation washes
  2.1.7 Signal detection
  2.1.8 Signal visualisation

2.2 Preparation of DNA
  2.2.1 Human genomic DNA
    2.2.1.1 Standard protocol
    2.2.1.2 Rapid protocol (Nucleon II DNA extraction kit)
  2.2.2 Alkaline lysis miniprep of plasmid or cosmid DNA
  2.2.3 Alkaline lysis midiprep of plasmid or cosmid DNA
  2.2.4 Yeast DNA in agarose plugs

2.3 Restriction enzyme digestion of DNA

2.4 Size fractionation of DNA by gel electrophoresis
  2.4.1 Conventional agarose gel electrophoresis
  2.4.2 Polyacrylamide gel electrophoresis
  2.4.3 Pulsed field gel electrophoresis (PFGE)
    2.4.3.1 Casting and loading the gel
    2.4.3.2 Gel electrophoresis
    2.4.3.3 Visualisation of the size-separated yeast DNA

2.5 Detection of DNA by hybridisation
  2.5.1 Conventional Southern blotting (uncharged membrane)
  2.5.2 Alkali Southern blotting (positively charged membrane)
  2.5.3 Bidirectional blotting
  2.5.4 Hybridisation of Southern blots with radiolabelled probes
    2.5.4.1 Probe labelling
2.5.4.2 Hybridisation 89  
2.5.4.3 Post-hybridisation washes and autoradiography 89  
2.5.4.4 Membrane-stripping protocol 89  

2.6 Sequencing single and double stranded DNA 90  
2.6.1 Overview of dideoxy chain termination sequencing 90  
2.6.2 Sequencing double stranded plasmid DNA by incorporation of isotope 90  
2.6.3 Direct sequencing of PCR products using a $\gamma^{32}$P endlabelled primer 90  

2.7 Polymerase chain reaction (PCR) techniques 91  
2.7.1 General considerations 91  
2.7.2 Standard parameters for a typical PCR 91  
2.7.3 Primer design and PCR optimisation 92  
2.7.4 Microsatellite analysis 92  
2.7.4.1 End-labelled primer 93  
2.7.4.2 Incorporation of label 93  
2.7.5 YAC library screening by PCR 93  
2.7.6 Screening for YAC-positive transformant yeast clones by colony PCR 95  
2.7.6.1 PCR of intact yeast cells 95  
2.7.6.2 PCR of sphleroblasts 95  
2.7.7 Inter-Alu PCR to generate ‘fingerprints’ 95  
2.7.8 Generation of YAC insert-terminal sequences 96  
2.7.8.1 Alu-vector PCR 96  
2.7.8.2 Vectorette PCR 97  
2.7.8.2.1 Construction of Vectorette library 97  
2.7.8.2.2 PCR amplification of Vectorette library 97  
2.7.8.3 Somatic cell hybrids for chromosomal localisation of STSs 98  

2.8 Cloning techniques 98  
2.8.1 Construction of a minilibrary from cosmid HX20 by shotgun subcloning 98  
2.8.1.1 Vector and cosmid restriction digests and DNA purification 98  
2.8.1.2 Insert dephosphorylation and ligation to vector 98  
2.8.2 Isolation of a CA-repeat-containing subclone from the HX20 minilibrary 99  
2.8.2.1 Preparation of competent E.Coli XL1-Blue cells 99  
2.8.2.2 Transformation of competent cells with ligated products 99  
2.8.2.3 Preparation of replica filters 99  

2.9 Reagents and buffers used in this research 100  
2.9.1 Stock reagents in general use 100  
2.9.2 FISH 100  
2.9.3 Preparation of DNA 101  
2.9.4 Gel electrophoresis 102  
2.9.5 Southern blot reagents 102  
2.9.6 Preparation of competent E.Coli cells 102
CHAPTER 3
Subchromosomal Localisation of X-specific Cosmid Clones by FISH

3.1 Introduction 103
   3.1.1 Physical assignment of DNA sequences to chromosomal subregions 103
   3.1.2 In situ hybridisation 103
   3.1.3 FISH methodology 105
      3.1.3.1 Chromatin preparation and denaturation 105
      3.1.3.2 Probe labelling 106
      3.1.3.3 Hybridisation 107
      3.1.3.4 Post-hybridisation washes 107
      3.1.3.5 Fluorescent detection 108
      3.1.3.6 Chromosome counterstaining and banding 108
      3.1.3.7 Signal visualisation 109
   3.1.4 FISH as a mapping tool in this study 110

3.2 Materials and Methods 111
   3.2.1 In situ hybridisation of centromeric repeat probe DXZ1 111
      3.2.1.1 Probe description and evaluation 111
      3.2.1.2 Experimental details 111
   3.2.2 In situ hybridisation of cosmid clones 112

3.3 Results 112
3.4 Discussion 119

CHAPTER 4
Isolation and Characterisation of a Microsatellite from Cosmid HX20

4.1 Introduction 122
   4.1.1 The relevance of microsatellites as genetic markers in this study 122
   4.1.2 Isolation of microsatellites 123
   4.1.3 Strategy in summary 123

4.2 Materials and Methods 123
   4.2.1 Construction of a minilibrary from cosmid HX20 123
   4.2.2 Isolation of HX20 subclones containing a CA repeat 124
   4.2.3 Sequencing the CA-positive HX20 subclone 124
   4.2.4 Microsatellite PCR assay and assessment of polymorphism 125
CHAPTER 5

Genetic and Physical Mapping of Five Novel Microsatellite Markers on Xp21.1-Xp11.22

5.1 Introduction 141
   5.1.1 The locus-ordering problem 141
   5.1.2 General strategies for map construction 142
       5.1.2.1 Framework versus comprehensive genetic maps 143
   5.1.3 The ordering of novel microsatellite markers on proximal Xp 143

5.2 Materials and Methods 144
   5.2.1 Families 144
   5.2.2 Genetic markers 144
   5.2.3 Genotyping 144
   5.2.4 Linkage analysis 146
       5.2.4.1 Correction of marker allele frequencies 147
   5.2.5 Haplotype analysis 148
   5.2.6 Physical mapping studies 148

5.3 Results 148
   5.3.1 Genotype data 148
   5.3.2 ILINK two-point analysis between pairs of markers 149
   5.3.3 Ordering the loci by haplotype analysis 159
   5.3.4 Ordering DXS722 and DXS573 by physical mapping 160
   5.3.5 Ordering the loci by ILINK order of preference 162

5.4 Discussion 164
CHAPTER 6

Genetic Refinement of the RP2 Critical Region and Exclusion of Candidate Gene TIMP-1 from a Causative Role in the Disease

6.1 Introduction

6.1.1 Mapping the RP2 locus for X-linked retinitis pigmentosa
6.1.2 Investigation of a role for TIMP-1 in the pathophysiology of RP2

6.2 Materials and Methods

6.2.1 Haplotype analysis in XLRP families
6.2.1.1 Families
6.2.1.2 Genetic markers and haplotype analysis
6.2.1.3 Meiotic breakpoint panel
6.2.2 Molecular genetic analysis of RP2 candidate gene TIMP-1
6.2.2.1 Retinal expression of TIMP-1
6.2.2.2 Completion of the TIMP-1 gene genomic organisation
6.2.2.3 Screening for TIMP-1 mutations in RP2 patients

6.3 Results

6.3.1 Haplotype analysis on 14 XLRP families
6.3.2 Genetic localisation and ordering of new markers within the RP2 critical region using a meiotic breakpoint panel
6.3.3 Investigation of TIMP-1 as a candidate for RP2
6.3.3.1 Confirmation of TIMP-1 gene expression in the retina
6.3.3.2 Genomic organisation of the human TIMP-1 gene
6.3.3.3 TIMP-1 mutation screen in RP2 patients

6.4 Discussion

6.4.1 Haplotype analysis of XLRP families: refinement of the RP2 region
6.4.2 Exclusion of TIMP-1 from a causative role in RP2
6.4.3 Genetic mapping and ordering of new markers within the RP2 critical region

CHAPTER 7

Physical Mapping in Xp11.3-p11.22: Construction of a YAC Contig Spanning the RP2 Critical Interval

7.1 Introduction

7.1.1 YAC libraries and identification of specific clones
7.1.2 Assembling YAC contigs
7.1.2.1 STS / probe-content mapping
7.1.2.2 Random assembly / fingerprinting 207
7.1.2.3 Chromosome walking: Isolation of YAC ends 208
7.1.3 Construction of a YAC contig spanning the RP2 interval 209

7.2 Materials and Methods 210
7.2.1 Sources of STS sequences and STS development 210
7.2.2 Identification of YACs by PCR-based YAC library screening 214
7.2.3 YAC characterisation and contig construction 215

7.3 Results 215
7.3.1 YAC contig around the RP2 region distal boundary in Xp11.3 215
7.3.2 YAC contig construction around the RP2 proximal boundary in Xp11.23-11.22 224
7.3.2.1 YAC insert-terminal isolation to generate novel region-specific STSs 224
7.3.2.2 YAC STS-content analysis and map integration 238
7.3.3 YAC contig construction in the central RP2 interval: extension of the physical map between TIMP-1 (Xp11.23) and distal flanking marker DXS8083 (Xp11.3) 244
7.3.3.1 Identification of seed YACs to nucleate contig construction in a region of low marker density 244
7.3.3.2 A ‘best estimate’ YAC contig derived by integration of the genetic, physical and transcriptional maps of Xp11.23 249
7.3.3.3 ALU-PCR fingerprints to confirm YAC order and orientation 253
7.3.4 Summary of YAC contig construction in the RP2 gene critical interval 256

7.4 Discussion 260

CHAPTER 8 264
Overall discussion and future prospects 264

REFERENCES 270
## List of figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>A schematic of the key features of a positional cloning strategy</td>
<td>20</td>
</tr>
<tr>
<td>1.2</td>
<td>The hierarchy of approaches to chromosome mapping</td>
<td>32</td>
</tr>
<tr>
<td>1.3</td>
<td>The rise of the positional candidate approach to identify disease genes</td>
<td>43</td>
</tr>
<tr>
<td>1.4</td>
<td>Cloned and mapped genes causing retinal degeneration in humans</td>
<td>44</td>
</tr>
<tr>
<td>1.5</td>
<td>Transverse section of the vertebrate eye</td>
<td>46</td>
</tr>
<tr>
<td>1.6</td>
<td>A schematic diagram of the primate retina</td>
<td>47</td>
</tr>
<tr>
<td>1.7</td>
<td>Diagrammatic representation of mammalian rod and cone photoreceptors</td>
<td>49</td>
</tr>
<tr>
<td>1.8</td>
<td>The phototransduction cascade in vertebrate rod photoreceptor cells</td>
<td>51</td>
</tr>
<tr>
<td>1.9</td>
<td>Consensus map of the proximal short arm of the human X chromosome</td>
<td>61</td>
</tr>
<tr>
<td>1.10</td>
<td>Genetic localisations of inherited eye disorders mapping to proximal Xp</td>
<td>67</td>
</tr>
<tr>
<td>1.11</td>
<td>Fundus photograph of an XLRP retina</td>
<td>72</td>
</tr>
<tr>
<td>3.1</td>
<td>FISH results using X-specific centromeric probe DXZ1 with 2 or 3 layers of antibodies</td>
<td>113</td>
</tr>
<tr>
<td>3.2</td>
<td>FISH mapping of 9 cosmids to specific X chromosome subregions</td>
<td>114</td>
</tr>
<tr>
<td>4.1</td>
<td>Agarose gel showing restriction digestion of cosmid HX20 and vector pT7T3</td>
<td>126</td>
</tr>
<tr>
<td>4.2</td>
<td>Agarose gel showing successful ligation of HX20 subclones to pT7T3</td>
<td>126</td>
</tr>
<tr>
<td>4.3</td>
<td>Autoradiograph showing HX20 cosmid minilibrary colony hybridisation</td>
<td>128</td>
</tr>
<tr>
<td>4.4</td>
<td>Agarose gel and autoradiograph showing size of CA repeat-containing subclones</td>
<td>129</td>
</tr>
<tr>
<td>4.5</td>
<td>Autoradiograph showing sequence of HX20 subclone 3b spanning CA repeat</td>
<td>130</td>
</tr>
<tr>
<td>4.6</td>
<td>Sequence of HX20 subclones 2b and 3b</td>
<td>131</td>
</tr>
<tr>
<td>4.7</td>
<td>Autoradiograph showing sequence of HX20 subclone 2b spanning CA repeat</td>
<td>133</td>
</tr>
<tr>
<td>4.8</td>
<td>Agarose gel showing efficient PCR amplification of the HX20 (CA)$_{21}$ microsatellite from genomic DNA</td>
<td>135</td>
</tr>
<tr>
<td>4.9</td>
<td>Autoradiograph showing 8 alleles of the HX20 (CA)$_{21}$ microsatellite detected in unrelated females</td>
<td>135</td>
</tr>
<tr>
<td>4.10</td>
<td>Autoradiograph showing X-linked codominant segregation of HX20 (CA)$_{21}$ microsatellite</td>
<td>136</td>
</tr>
<tr>
<td>5.1</td>
<td>The 14 XLRP pedigrees and associated haplotypes used to genetically order cosmid-derived microsatellites on proximal Xp</td>
<td>150</td>
</tr>
<tr>
<td>5.2</td>
<td>Summary of meiotic recombination events informative for ordering the loci</td>
<td>161</td>
</tr>
<tr>
<td>5.3</td>
<td>Set of YAC clones in Xp11.23 used to physically map microsatellite DXS722</td>
<td>162</td>
</tr>
<tr>
<td>6.1</td>
<td>Schematic diagram summarising the order and approximate locations of</td>
<td></td>
</tr>
</tbody>
</table>
genetic markers used in this study on proximal Xp

**Figure 6.2** Extended genetic haplotypes for the 14 XLRP families showing key crossovers which localise the segregating XLRP gene

**Figure 6.3** Ordering markers genetically using the meiotic breakpoint panel

**Figure 6.4** PCR amplification products spanning introns 1 and 2 of the human TIMP-1 gene

**Figure 6.5** Genomic structure of the human TIMP-1 gene

**Figure 6.6** Sequence of TIMP-1 exon 5 in affected/unaffected members of the RP2 families showing a C to T polymorphism

**Figure 7.1** STS loci mapped to Xp11.3-p11.22 at the outset of this study

**Figure 7.2** Agarose gel photographs showing positives identified upon sequential screening of the ICI YAC library pools with marker DXS742

**Figure 7.3** Pulsed field gel photograph and autoradiograph demonstrating size estimation of YACs from the YAC contig in Xp11.3

**Figure 7.4** YAC contig covering the distal RP2 boundary in Xp11.3

**Figure 7.5** ALU-vector PCR products obtained from YAC 40HD1

**Figure 7.6** FASTA printout showing ALU identity of STS 9 BA10LA

**Figure 7.7** YAC contig spanning the RP2 proximal boundary in Xp11.23-p11.22

**Figure 7.8** Pulsed field showing approximate sizes of the ICRF SYN-1 YACs

**Figure 7.9** Examples of PCR products representing insert terminals from YACs in the proximal contig

**Figure 7.10** FASTA printout showing region of ALU identity in 4DF10LA YAC end-sequence

**Figure 7.11** FASTA printout showing identity of 4HG2RA YAC end-sequence to SYN-1

**Figure 7.12** FASTA printouts showing identities of 34AC5RA YAC end-sequence to 4 Genbank EST entries

**Figure 7.13** Chromosomal localisation of YAC insert-terminal STSs by PCR analysis of X-specific somatic cell hybrids

**Figure 7.14** Agarose gels showing YAC STS-content mapping results for YAC end STSs C1022RA and 4HG2RA

**Figure 7.15** Localisation of YAC end STS C01160LA to chromosome 5 or 6 by use of the Biosmap multichromosomal somatic cell hybrid panel

**Figure 7.16** FASTA printout showing identity of YAC end STS 4DF10RA to STS DXS9795

**Figure 7.17** FASTA printout showing identity of YAC end STS 27CB12LA to STS DXS8193

**Figure 7.18** Agarose gel photograph showing STS-content of YACs around OATLI for marker DXS6950

**Figure 7.19** YAC STS-content mapping results for the retinally expressed MG cDNAs

**Figure 7.20** YAC contig spanning the central portion of the RP2 interval in Xp11.23
**Figure 7.21** FASTA printout showing identity of 24BF12RA YAC end-sequence to a fibroblast EST  

**Figure 7.22** Agarose gels showing YACs positive for markers TIMP-1, DXS337, ZNF157, DXS1146 and DXS7124 by STS-content mapping  

**Figure 7.23** Agarose gel photographs showing ALU PCR fingerprints for key YACs in Xp11.23  

**Figure 7.24** Chart summarising YAC contigs assembled in Xp11.3-p11.22  

### List of tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1.1</td>
<td>The evolution of human genetic maps</td>
<td>26</td>
</tr>
<tr>
<td>Table 1.2</td>
<td>Genetic maps for the human X chromosome</td>
<td>60</td>
</tr>
<tr>
<td>Table 1.3</td>
<td>Cloned genes in proximal Xp at the outset of this study</td>
<td>62</td>
</tr>
<tr>
<td>Table 3.1</td>
<td>Subchromosomal localisation of 12 X specific cosmids by metaphase FISH</td>
<td>118</td>
</tr>
<tr>
<td>Table 4.1</td>
<td>Sizes of HX20 (CA)$_n$ allelic fragments and their frequencies in 36 unrelated Caucasian females</td>
<td>134</td>
</tr>
<tr>
<td>Table 5.1</td>
<td>Description of 5 microsatellites derived from FISH mapped cosmids</td>
<td>145</td>
</tr>
<tr>
<td>Table 5.2</td>
<td>Reference genetic markers used in genetic map construction</td>
<td>146</td>
</tr>
<tr>
<td>Table 5.3</td>
<td>Demonstration of allele recycling in family F47</td>
<td>149</td>
</tr>
<tr>
<td>Table 5.4</td>
<td>Pairwise maximum likelihood recombination fractions and corresponding lod scores</td>
<td>158</td>
</tr>
<tr>
<td>Table 5.5</td>
<td>Best orders obtained from ILINK analysis of markers distal to DXS7</td>
<td>163</td>
</tr>
<tr>
<td>Table 6.1</td>
<td>Further microsatellites used for genetic characterisation of XLRP families</td>
<td>173</td>
</tr>
<tr>
<td>Table 6.2</td>
<td>Key crossovers in Xp11.3-p11.22 comprising the meiotic breakpoint panel</td>
<td>174</td>
</tr>
<tr>
<td>Table 6.3</td>
<td>Summary table of XLRP locus assignments</td>
<td>190</td>
</tr>
<tr>
<td>Table 6.4</td>
<td>Primers and conditions for PCR amplification of the TIMP-1 gene</td>
<td>196</td>
</tr>
<tr>
<td>Table 6.5</td>
<td>Intron/exon structure and splice junctions of the TIMP-1 gene</td>
<td>198</td>
</tr>
<tr>
<td>Table 7.1</td>
<td>Additional STSs used to format the YAC contig</td>
<td>212</td>
</tr>
<tr>
<td>Table 7.2</td>
<td>Details of STSs and ESTs derived from published sequences</td>
<td>213</td>
</tr>
<tr>
<td>Table 7.3</td>
<td>YACs identified with STSs around the distal RP2 boundary in Xp11.3</td>
<td>218</td>
</tr>
<tr>
<td>Table 7.4</td>
<td>YACs identified with STSs spanning the proximal RP2 boundary in Xp11.23-p11.22</td>
<td>226</td>
</tr>
<tr>
<td>Table 7.5</td>
<td>Insert-terminal sequences derived from YACs comprising the proximal contig in Xp11.23-p11.22</td>
<td>228</td>
</tr>
<tr>
<td>Table 7.6</td>
<td>YAC end-STS chromosomal localisation for the proximal contig from STS-content analysis of somatic cell hybrids and overlapping YACs</td>
<td>233</td>
</tr>
<tr>
<td>Table 7.7</td>
<td>Confirmed positive YACs for the central RP2 interval in Xp11.3-p11.23</td>
<td>245</td>
</tr>
<tr>
<td>Table 7.8</td>
<td>Novel YAC insert-terminal STSs mapping to Xp11.3-p11.22</td>
<td>257</td>
</tr>
</tbody>
</table>
CHAPTER 1

General Introduction

1.1 The identification of genes in human inherited disease

The main problem with human disease gene identification is one of scale. The size of the human genome is approximately $3 \times 10^6$ bp whereas the size of the coding sequence of the 'average' gene is around $3 \times 10^3$ bp, one millionth of the size of the genome in which it resides. Moreover, a change in a single base in the gene may be sufficient to cause disease. Despite this, several hundred human disease related genes have been cloned and analysed since the advent of recombinant DNA technology in the mid-seventies, and thousands have been mapped to specific chromosomal locations.

There are basically two pathways towards the identification of human disease related genes. The first approach depends on prior information about the protein product and/or function of the responsible gene and is therefore termed 'functional cloning'. However, for the vast majority of human inherited diseases, the underlying biological defect is unknown. In this instance, most investigators adopt one of two more recent strategies to identify the causative genes: the 'candidate gene' approach or 'positional cloning'. Previously isolated human genes can be considered as disease locus candidates if they are known/inferred to have a role in the physiology of the diseased tissue. 'Positional cloning' assumes no functional information, and must locate the responsible gene solely on the basis of map position.

1.1.1 Positional cloning

A general outline of this approach for the identification of disease related genes is shown in Figure 1.1. The success of the positional cloning strategy (Collins 1992) lies initially in establishing a sub-chromosomal localisation for the disease gene via genetic linkage analysis in affected families (section 1.2.6). Subsequent linkage analysis can then be employed to successively reduce the candidate interval by identifying flanking markers that show very little or no recombination with the disease locus (section 1.2.7). The next step is to convert the genetic distance to the physical scale and to construct a high resolution physical map of the region. This involves the precise localisation and ordering of clones and markers, with an aim to assemble and characterise a genomic contig over the disease region (section 1.3.2). The ultimate identification of a disease locus minimally requires the demonstration of patient-specific mutations which are inconsistent with normal gene expression.

In some cases the approach to identifying the disease gene is helped enormously by the identification of disease-specific chromosomal aberrations, detected either as gross structural...
Figure 1.1 A schematic showing the key features of positional cloning (after Collins 1992)
rearrangements on a cytogenetic level, or by Southern blot analysis (section 1.8.1). A special case of DNA rearrangement, the expanded trinucleotide repeat, has also greatly facilitated those gene searches where it proved to represent the mutational basis of the disease. In these instances, a trinucleotide repeat sequence is expanded to hundreds or thousands of copies in affected individuals, resulting in a visible abnormality on Southern analysis. The length of the expanded repeat appears to correlate with severity of disease, and the discovery of around 10 such disorders in the last three years suggests that this phenomenon might be responsible for a significant percentage of human genetic disease (Sutherland and Richards 1995). Examples of genes that have been cloned with the aid of chromosomal abnormalities include Duchenne muscular dystrophy (deletions and translocations: Monaco et al 1986) choroideremia (large cytogenetic deletions: Cremers et al 1990) and Huntington’s disease (trinucleotide repeat expansion: The Huntington’s disease collaborative research group 1993).

For disorders not blessed with such genetic clues, the route to the gene involves saturation genetic and physical mapping. The key to successful isolation of such genes stemmed from three crucial developments: the discovery of genetic markers across the entire genome (Botstein et al 1980; genetic mapping), the development of efficient and accessible statistical methods for analysing the segregation of of markers and disease genes (Elson and Stewart 1971, Ott 1974; linkage analysis), and the advances in large-scale cloning and sequencing of DNA (Burke et al 1987; physical mapping).

### 1.2 Genetic mapping

Genetic mapping depends on following the segregation of alleles at two or more loci (genes or genetic markers) during meiosis. For disease gene mapping, one locus is a genetic marker and the other is the disease phenotype.

#### 1.2.1 Meiotic recombination

The genetic material of humans exists in linear arrays of genes along 23 pairs of homologous chromosomes in the nucleus of the cell. During meiotic cell division homologous chromosome pairs, each consisting of two chromatids, are aligned side by side in the centre of the cell. During this alignment DNA segments are exchanged between chromatids of homologous chromosomes. This process, which results in recombinant chromosomes, is known as crossing over. Only homologous regions of chromosomes are exchanged such that each chromosome has the same set of genetic loci. Though in a diploid cell each copy of the same chromosome comes from a different parent, they are more than 99% identical in DNA sequence. Nevertheless, each copy contains many small variations (DNA polymorphisms), the sum of which gives each individual a unique genetic makeup. Thus the formation in the germ line cells of a single chromosome derived from segments of each of the parental copies (as a result of crossover events) passes on a unique assortment of genes to the next generation.
During meiosis, each pair of homologous chromosomes undergoes at least one recombination (cross-over) between non-sister chromatids. It follows that for alleles at two different loci to segregate together during meiosis (when they are said to be genetically linked), the loci must lie in close physical proximity on a chromosome. To be linked, it is necessary but not sufficient for loci to be syntenic (located on the same chromosome) as different chromosomes assort independently. The combination of alleles at linked loci is called a haplotype.

The extent of genetic linkage between two or more loci is measured by the recombination fraction ($\theta$), which is the proportion of recombinations between two loci, out of the total number of opportunities for recombination. As each of the two sister chromatids in a chromosome has only a 50% chance of being involved in any one crossover, the maximum recombination fraction between any two loci is 0.5, corresponding to unlinked loci that segregate independently. Where two loci are in close physical proximity along a chromosome, $\theta$ will be <0.5 and for tightly linked loci, $\theta$ approaches 0, indicating that only rarely does a recombination occur between them. With increasing distance between two loci, an increasing number of crossovers is expected to occur between them, thus the recombination fraction is also a measure of the recombination frequency, and therefore serves as a measure of distance between them.

### 1.2.2 Genetic map distance

The unit of genetic distance is the Morgan, defined as the length of chromosomal segment which, on average, undergoes one exchange per individual chromatid strand. In male meiosis an average of 53 chiasmata occur over all of the autosomes during meiosis. The total male autosomal map length is thus estimated as approximately 26.5 Morgans. Based on higher recombination rates in females than in male meiosis, the female map length measures about 39 Morgans, giving a sex-averaged autosomal map length of 33 Morgans (Renwick 1969). The 'average' human chromosome is then 1.5 Morgans long; that is, it experiences an average of 1.5 crossovers. In practice, genetic distances are quoted in centimorgans (cM), whereby two loci are 1 cM apart if they recombine once in every 100 meioses. Over short chromosomal regions, $\theta$ is directly proportional to the genetic map distance, so that a $\theta$ value of 0.01 corresponds to a genetic map distance of 1 cM. However, over longer distances this linear relationship breaks down, mainly because of multiple crossovers occurring between the two loci, and also the effect of 'interference'. Positive interference describes the effect that a crossover has of reducing the probability of a second crossover in its vicinity, and operates over relatively small distances (theta < 0.10).

When many closely linked loci of known order are mapped to a chromosome segment, the simplest method of determining map distances among these loci is to estimate the recombination fractions in each interval of adjacent loci. The map distance between two more distant loci is then obtained as the sum of the map distances in the intervals between these loci. For larger distances, the recombination fraction is not an additive distance measure as values would likely exceed 0.5. Several map functions have hence been devised to transform $\theta$ into
additive map distance, by incorporating different levels of interference (reviewed in Ott 1991). The Kosambi map function is most widely used as it allows for the strongest level of interference and thus produces more realistic map distance values for θ values of 0.10-0.30.

### 1.2.3 DNA polymorphism and the development of human genetic markers

The earliest markers used for human genetic mapping were protein markers, including blood group antigens and serum protein polymorphisms (reviewed in Ott 1991). These markers were generally quite uninformative and required a wide range of biochemical and immunological techniques for their analysis. In experimental animals and other model organisms the low polymorphism of such protein markers could be surmounted by enforced matings on captive populations; obviously the same could not be done in human genetics.

The situation changed rapidly after Botstein et al (1980) observed that DNA sequence polymorphisms provided an enormous, untapped source of variation in eukaryote genomes. A subset of these normal sequence variations, such as single base changes, deletions or insertions, either create or abolish restriction enzyme cleavage sites, and thus alter the length of the DNA fragments produced upon digestion with the appropriate restriction enzyme. The size differences can be detected by Southern analysis, in which recombinant DNA probes define the locus of interest and restriction enzymes define the sequence variant. These restriction fragment length polymorphisms (RFLPs) tend to be ‘neutral’ polymorphisms in that they are presumed to have no physiological function, which makes them well suited as genetic markers as they are expected not to interfere with the phenotypic expressions at other loci.

Regions containing moderately sized arrays of tandem repeated DNA sequences (minisatellites) are another common source of detectable sequence variation (Jeffreys et al 1985). Such hypervariable minisatellites (named by analogy with the larger satellite arrays comprising the heterochromatin) are organised in over 1000 arrays of short (from 0.1 to 20kb long) tandem repeats. Such variable number tandem repeats (VNTRs) proved better genetic markers than RFLPs owing to their greater number of allelic variants. Botstein had calculated that 150 polymorphic markers would be required to generate a complete linkage map of the human genome with markers spaced at 40cM intervals. It soon became clear, however, that many more would be needed to map the 50-100,000 genes estimated to exist in humans.

Donis-Keller et al in 1987 reported the first complete linkage map of the human genome, comprising 403 RFLPs with linkage groups on all chromosomes and an average resolution of 9cM. Many of these genetic markers were found during the characterisation of cloned genes, but the majority stemmed from systematic searches of arbitrary cloned DNA fragments. By 1991, 3000 human polymorphic markers had been isolated (Human Gene Mapping 11;1991). However, 90% of these had a heterozygosity of less than 50% and they were unevenly spaced throughout the genome, with large gaps on certain chromosomes. In particular, the more promising VNTRs tended to cluster near the ends of the chromosomes (Royle et al 1988). Although maps based on these markers had contributed greatly to the primary mapping of a
number of genetic diseases, they were insufficient for refining linkage intervals to distances suitable for gene identification.

Two major advances in molecular biology made possible the rapid development of highly informative markers for genetic mapping and linkage analysis. First, the introduction of the polymerase chain reaction (PCR) as a means to amplify DNA from a template in vitro using DNA polymerase and temperature cycling (Saiki et al 1988), followed by its use to demonstrate a high level of polymorphism or allelic variation in the repeat number for simple sequence tandem repeats known as microsatellites (Weber and May 1989).

1.2.3.1 Microsatellites

Microsatellites consist of around 10-50 copies of motifs from 1 to 6 bp that can occur in perfect tandem repetition (perfect repeats), as imperfect (interrupted) repeats or together with another repeat type (compound repeats) (Weber 1990). Perfect sequences comprise about 65% of the total (CA), polymorphisms, imperfect sequences about 25% and compound sequences about 10%. Microsatellites appear frequently and abundantly in all eukaryotic DNAs examined (Stallings et al 1991) and are named by analogy with the larger minisatellite arrays or VNTRs. In humans, 76% of repeat types are A, CA, AAAN, AAN or AG, in decreasing order of abundance. Around 80% of human A, AAAN, and AAN repeats and 50% of AT microsatellites occur 3' to ALU elements (Beckman and Weber 1992) and about 17% of CA repeats are also associated with ALUs, which has led to speculation over the possibility that primate microsatellites may originate from ALU A-rich regions (Arcot et al 1995). Tri- and tetranucleotide microsatellites occur every 300-500kb on the human X chromosome (Edwards et al 1991) and appear to be interspersed at this frequency throughout the genome. On average, CA repeats occur every 30kb in human euchromatic DNA (Stallings et al 1991), distributed equally in genic 5'- and 3'-untranslated regions, introns and intergenic regions; these are the most commonly used type of microsatellite owing to their abundancy.

The broad dispersal of CA repeats throughout the genome is clear from several lines of evidence. When large-insert genomic DNA libraries are screened by hybridisation with a poly-CA probe, many different clones are labelled, 20% clones from phage libraries (Hamada et al 1984) and 40% clones from cosmid libraries (Litt and Luty 1989). These results would not be obtained if the CA repeats were clustered, and are consistent with an estimated 50,000-100,000 sequence copies of this family in the human genome by hybridisation to total human DNA (Hamada and Kakunaga 1982). Also, human chromosome-specific libraries each contain large numbers of these sequences (Luty et al 1990; Stallings et al 1991).

1.2.3.1.1 Mutation rate and mechanism of microsatellite length variation

If the mutation rate for a polymorphic DNA sequence is so high that alleles cannot be followed through families, its utility as a genetic marker for mapping disease genes and the study of evolution is compromised. Linkage disequilibrium studies (see section 1.2.7), for example, rely on linkage disequilibrium between marker alleles which in turn is highly
dependent upon mutation rate. The mutation rate for microsatellites is estimated at between \(5 \times 10^4\) and \(10^5\) (Kwiatkowski et al 1992) which is low enough to permit their use in such studies.

Strand slippage is currently the most favoured mechanism for mutation of simple-sequence tandem repeats (Schottlerer and Tautz 1992). Slippage of the newly synthesised strand along the replication fork during the process of its replication introduces or deletes small sequence units in the new DNA double strand, generating new alleles. A tandem organisation of the same sequence unit will facilitate its misalignment i.e. slippage.

1.2.4 Marker utility: the degree of polymorphism

In order to detect recombinants and non-recombinants between two marker loci in the progeny, at least one of the parents must be doubly heterozygous (i.e. possess two different alleles at each marker locus), when the mating is said to be ‘informative for linkage’. It follows that a marker’s usefulness for linkage analysis depends on the number of its alleles and their gene frequencies (i.e. its degree of polymorphism), in the sense that an increased polymorphism leads to an increased probability of heterozygosity. Informativeness of human DNA polymorphisms is one of the most important criteria of marker utility. Increasing the information content of markers can drastically reduce the number of individuals that must be typed to map genetic disease genes and can greatly improve the resolution of linkage maps produced using a fixed number of reference families.

The probability that a random individual is heterozygous is used as a measure of the degree of polymorphism and is estimated in two ways, each based on a random sample of unrelated individuals:

The first measure is the amount of heterozygosity observed (h) and is simply the proportion of heterozygous individuals observed in the sample. In human genetics, a more precise estimate is usually used, by assuming that the genotypes are in Hardy-Weinberg equilibrium (HWE), where the genotypes observed in a population will directly reflect the gene frequencies. This expected heterozygosity (or in human genetics, just heterozygosity) is defined as

\[
H = 1 - \sum p_i^2
\]

where \(p_i\) is the population frequency of the \(i\)th allele, and \(H\) is simply the probability that a random individual is heterozygous for any two alleles at a gene locus with allele frequencies, \(p_i\). Heterozygosity is thus directly proportional to the number of alleles each marker exhibits and the frequency of each allele in the general population.

In human genetics, another measure of the degree of polymorphism often used is the PIC (polymorphism information content) value (Botstein et al 1980). PIC is similar to heterozygosity, but allows for the fact that it may not be possible to deduce the parental chromosomal type for all children. The PIC value is defined as the probability that the marker
genotype of a given offspring will allow deduction of which of the two marker alleles of the affected parent it had received. PIC is calculated as

$$\text{PIC} = 1 - \sum_{i=1}^{n} p_i^2 - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 2p_i p_j$$

where $p_i$ and $p_j$ are the population frequencies of the $i$th and $j$th alleles. In the PIC value a quantity is subtracted from the heterozygosity that corresponds to the probability that offspring are uninformative, because if both parents are identically heterozygous, on average half of their children (the homozygotes) are informative and half (the heterozygotes) are uninformative. The PIC value is therefore always smaller than the heterozygosity, but for large numbers of alleles, the two tend to become very close.

Polymorphisms with only two alleles, as is the case for many RFLPs, have a maximum PIC of 0.5 (Botstein et al 1980). PIC values are highest for markers with several common alleles, such as VNTRs and microsatellites (reviewed in Goodfellow 1992).

1.2.5 Genetic maps of the human genome

Since the development of highly polymorphic microsatellite genetic markers, the progress in compiling a detailed and informative genetic map of the human genome has been astounding (see Table 1.1), greatly aided by the availability of a set of 61 families of the CEPH (Centre d'Etude du Polymorphisme Humain). The CEPH contribution allowed investigators around the world to pool data from markers developed in individual laboratories but studied on a common set of families (Dausset et al 1990).

<table>
<thead>
<tr>
<th>Group</th>
<th>Year</th>
<th>Spacing (cM)</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collaborative Research</td>
<td>1987</td>
<td>10</td>
<td>403 loci (393 RFLPs)</td>
</tr>
<tr>
<td>NIH/CEPH</td>
<td>1992</td>
<td>&lt;5</td>
<td>1416 loci (339 microsatellites) (279 genes)</td>
</tr>
<tr>
<td>Genethon</td>
<td>1992</td>
<td>4.4</td>
<td>814 CA repeats</td>
</tr>
<tr>
<td>Pittsburgh</td>
<td>1994</td>
<td>6.0</td>
<td>1663 loci (1266 microsatellites) (397 VNTRs and genes)</td>
</tr>
<tr>
<td>CHLC</td>
<td>1994</td>
<td>4.9</td>
<td>1123 loci (mixed)</td>
</tr>
<tr>
<td>Genethon</td>
<td>1994</td>
<td>2.9</td>
<td>2066 CA repeats</td>
</tr>
<tr>
<td>CHLC/Genethon/CEPH</td>
<td>1994</td>
<td>0.7</td>
<td>5840 loci (3617 microsatellites)</td>
</tr>
<tr>
<td>Genethon</td>
<td>1996</td>
<td>1.6</td>
<td>5264 CA repeats</td>
</tr>
</tbody>
</table>

In 1992, NIH/CEPH Collaborative Mapping Group published a genetic linkage map of the human genome consisting of 1416 loci, 339 of which were microsatellites. Soon after, Weissenbach et al Genethon published their ‘second generation linkage map of the human
genome’, constructed from segregation analysis of 814 novel polymorphic CA repeats in eight large CEPH families (Weissenbach et al 1992). 605 markers showed a heterozygosity above 0.7, 553 could be ordered with odds 1000:1, and the distance spanned corresponded to ~90% of the estimated length of the human genome. Both maps had an average resolution of <5cM, filling in many of the gaps in the original RFLP-based map from Collaborative Research (Donis-keller et al 1987).

With the isolation of novel markers continuing to accelerate, 1994 saw the publication of more human genetic linkage maps with even greater resolving power (see Table 1.1). Matise et al (1994) developed an automated procedure to generate a map using published genotypes at 1266 microsatellite loci and annotated the map by adding 397 VNTR and polymorphic gene markers of known physical location. The map produced by the Cooperative Human Linkage Centre (CHLC) (Buetow et al 1994) addressed lack of integration of independent marker collections by constructing a set of genome-wide maps using datasets contributed by the National Centre for Human Genome Research (NCHGR) Index Map Consortium and Genethon in 1992. This map also included many tri- and tetra-nucleotide microsatellites which, although less frequent than CA repeats, are easier to genotype as they do not have the replication slippage products of dinucleotide repeats (see section 4.4). Furthermore, these markers have a different regional bias (Sheffield et al 1995) and thus may also fill in some of the gaps in the CA-based maps as they reach their resolution limit. In September of 1994, collaborative efforts of the 3 large groups (CEPH, Genethon, CHLC) produced a genetic map combining genotype data generated over the last decade, including all recent efforts, and the resulting maps represent the culmination of work by hundreds of investigators worldwide (Murray et al 1994). It consists of 5840 loci, of which 970 are uniquely ordered, covering 4000 cM on the sex-averaged map. 3617 were PCR-formatted short tandem repeat polymorphisms, 427 were genes. The map has markers at an average marker density of 0.7cM. The final Genethon map features 5264 CA repeats of mean heterozygosity 70% (Dib et al 1996) at an average resolving power of 1.6cM. These final maps thus achieve one of the goals of the human genome project (section 1.5), a comprehensive, high-density genetic map.

1.2.6 Pedigree-based genetic linkage analysis

Linkage analysis capitalises on polymorphic genetic markers to detect crossover events in families and assess whether they segregate with less than 50% recombination, whence they are 'genetically linked'. Determination of a heterozygous parent genotype does not reveal the distribution of alleles between the two chromosomes (phase). These distributions influence the interpretation of results, because a given progeny chromosome will be scored as either recombinant or non-recombinant depending on the arrangement chosen. Because of the analytical complexities resulting from such incomplete knowledge of the genotypes of individuals in families, linkage data from humans have traditionally been evaluated by maximum likelihood estimations (Morton 1955), where an estimate of the recombination interval (θ) between two loci is obtained by determining the value of θ which gives the maximum probability for the observed data. A quantitative expression for the likelihood of
linkage is the 'odds ratio', which is the ratio of the probability of observing the observed segregation pattern of two loci if they are linked at a given value of $\theta$ ($0<0.5$) to the probability at $\theta = 0.5$ (the marker loci are genetically unlinked) i.e.

\[
\frac{\text{likelihood the two loci are linked (}\theta<0.5)}{\text{likelihood the two loci are unlinked (}\theta=0.5)}
\]

This ratio is calculated at a series of recombination values (from $\theta=0-0.5$) and is normally expressed as a logarithm, the LOD ($z$, log of the odds) score for linkage, thus defined as the log_{10} of the ratio of the probability that the data would have arisen if the loci are linked to the probability that the data would have arisen if the loci are unlinked. The $\theta$ value that gives the maximum lod score ($Z_{\text{max}}$) is the best estimate of the degree of linkage between the two loci. Lod scores of 3 or higher, representing a relative likelihood of linkage to non-linkage > or = 1000:1 are considered to be strong evidence for linkage. A lod score of -2 or less is taken as evidence against linkage, within an interval equal to the corresponding $\theta$ value from both sides of the marker locus. If there are sufficient negative data, a locus can be excluded from whole chromosomes, or even most of the genome (exclusion mapping). The search can then be focused on the remaining non-excluded locations. The apparently high likelihood ratio for proof of linkage is needed because the prior probability that any two loci are linked is about 1 in 50; the supplementary probability of 1 in 20 required to achieve a lod score of 3 therefore corresponds to only about 95% confidence that linkage exists.

Extended, multigenerational kindreds are most effective for human linkage studies, as they can sometimes provide phase information in parents from determination of the grandparental genotypes. Importantly for studies of inherited diseases, use of a single family also minimises the possibility that genes at more than one locus might be involved. Large sibships also yield a higher number of informative chromosomes, which becomes important when the allele distribution is not explicitly defined by the genotypes of the grandparents.

Linkage analysis is a powerful method for localising disease genes to a small region of the genome. To date, at least 1800 human genes have been mapped using this approach (OMIM; the on-line database of human genetic diseases; World Wide Web address http://www3.ncbi.nlm.nih.gov/Omim/).

1.2.7 Refining the localisation of a disease gene

Once a disease gene has been successfully assigned to a particular subchromosomal region by genetic linkage analysis, its location can be refined by genotyping branches of the pedigree(s) containing key recombinant individuals with further genetic markers in the vicinity of the gene, to discover the closest flanking markers by 'haplotype analysis'. This involves careful inspection of allele segregation to determine whether each new marker has recombined with the disease locus or resides within the critical interval bounded by the flanking recombination events. The approach is limited by the number of informative meioses
in family pedigrees, dependent both on marker density and heterozygosity in the region of interest, and on the number of recombinant individuals.

With the current availability of high resolution genetic maps of virtually the whole of the human genome (section 1.2.5), fine mapping can now potentially delimit the critical interval containing the disease gene to about 1cM. As the genetic length of the human genome is 33M (Renwick 1969) and the physical length of a haploid genome is \(~3 \times 10^9\) bp, 1 cM corresponds to \(~1\) Mb of DNA. This figure is variable, however, as different regions of the genome vary in recombinogenicity, such that a 1 cM critical interval may in fact equate to a physical distance of 0.5-3 Mb.

In regions where there are relatively few recombination events, or once the maximum refinement with available recombinants and markers has been achieved, linkage disequilibrium between the disease and linked DNA markers may further localise the defective gene. This is the non-random association of marker alleles at linked loci, or marker alleles and a disease mutation. In contrast to linkage mapping, linkage disequilibrium mapping attempts to glean information from all recombination events that have occurred during the history of a population rather than simply those in existing families. Among a collection of chromosomes carrying the same ancestral mutation, genetic markers nearest the disease gene will have recombined least often with the disease. To apply this method, one would determine the haplotypes of disease-bearing chromosomes for an extremely dense set of genetic markers, use these haplotypes to identify a subset of chromosomes likely to carry a common ancestral mutation, and find the genetic markers that show the strongest linkage disequilibrium with the disease on these chromosomes. This approach has the best chance of success for relatively rare, simple genetic diseases that are not very old (so that a disease-causing mutation is still likely to be associated with a specific background haplotype) and for which most of the disease chromosomes descend from just a few ancestral mutations e.g. in isolated founder populations. However, the technique is not always straightforward unless the mutation segregates with a rare marker allele, and can produce misleading results where more than one gene gives rise to the same disease phenotype (genetic heterogeneity), or where there are multiple disease-causing mutations as in Huntington’s disease (MacDonald et al 1992). Its usefulness falls as markers become very close to the disease gene (< 75kb), and the populations must also be large enough to provide an adequate sample of disease and normal haplotypes for analysis (Jorde 1995).

Once the fine genetic mapping of a disease gene has reached its limit of resolution, a positional cloning approach then reduces to a search through the critical region for all transcripts, followed by analysis of each of these for alterations in affected individuals. Other information, such as the pattern of tissue expression of the various candidate genes, or evidence of evolutionary conservation by ‘zoo-blot’ analysis, can often prioritise the search at this point (section 1.4). Before such characterisation of transcripts in the region can begin, however, the region must be recovered in cloned DNA, and high resolution physical mapping strategies must be employed to define the disease locus more accurately before attempts are undertaken to isolate the gene.
1.3 Physical mapping

A physical map represents distances between identifiable landmarks (e.g. genetic marker sequences, restriction sites) in numbers of nucleotide base pairs and is thus a more actual representation of the underlying DNA than a genetic map generated on the basis of recombination events, which are known to be nonrandom in location. At its simplest level of resolution, physical mapping involves chromosome banding, whereas the ultimate physical map would be a complete nucleotide sequence of the human genome (see Figure 1.2).

Efforts to construct a high resolution map of the human genome have intensified significantly during the past few years. Physical mapping techniques provide a complement to genetic linkage in the generation of chromosome maps; for a DNA marker to be useful in localising a disease gene by linkage, its own chromosomal localisation must be determined with a reasonable degree of accuracy. For initial chromosome assignment and precise sequence localisation, somatic cell genetics and in situ hybridisation (ISH) have been mostly used (genetic linkage mapping can then be employed to compile regional genetic maps).

1.3.1 Cytogenetic methods of physical mapping

The traditional methods of physical mapping have been the cytogenetic techniques using rodent-human hybrid cells and in situ hybridisation, both of which have been sharply improved in recent years. These techniques have in the past been considered 'low resolution' mapping approaches, but with the rapid pace of development of various new molecular technologies such distinctions are fast becoming extinct.

1.3.1.1 Somatic cell hybrids

Somatic cell hybrids derive from experimentally induced fusion of cultured cells from different species, usually human and rodent (Ruddle 1981). The preferential loss of most human chromosomes from such unstable hybrid nuclei in a random fashion can be used to generate a panel of hybrids, each of which retains multiple human chromosomes along with the complete complement of rodent chromosomes. The complement of human chromosomes in each cell line can be determined by chromosomal in situ hybridisation (section 1.3.1.2) using total human genomic DNA as a probe.

In multichromosomal cell hybrid panels, although each human chromosome may be present in several of the hybrid cell lines, a given chromosome will have a unique ‘signature’ across the entire panel. DNA markers can then be mapped to particular human chromosomes by observing the concordance and discordance of positive signals obtained after hybridisation or PCR of such a panel, with the chromosomes present in the cell lines. The success of the technique relies on the probe being human-specific, such that cross-reactivity to the rodent chromosomes does not occur.

Monochromosomal somatic cell hybrid panels contain 24 hybrid cell lines, each of which retains a different single human chromosome in addition to the rodent chromosomes. These
panels are much easier to characterise cytogenetically, and provide firmer establishment of chromosomal location where related genes or pseudogenes are involved. However, the redundancy of multichromosomal panels tends to ensure greater confidence in establishing a chromosomal location in the instance that cytogenetically undetectable rearrangements (e.g. deletions) include the locus of interest.

The use of somatic cell hybrids containing parts of human chromosomes (such as those constructed from human cells harbouring a chromosomal translocation or deletion) can permit assignment of DNA probes or markers to subchromosomal segments. These 'regional' panels are particularly useful for ordering markers in a disease gene critical region (Patel et al 1995).

The most significant recent development in somatic cell genetics has been the introduction of radiation hybrids for the physical mapping of chromosomes in the context of establishing a complete human gene map (Cox et al 1990, Walter et al 1994). This methodology provides a general method for ordering DNA markers spanning millions of bp of DNA at the 500 kb level of resolution (Figure 1.2). A high dose of X-rays is used to break the human chromosome of interest into several fragments, which are recovered in rodent cells. By estimating the frequency with which pairs of linked markers are cotransferred after irradiation it is possible to determine their linear order in a manner analogous to meiotic mapping. Simply determining that a radiation hybrid line carries two markers does not distinguish between these markers being retained on the same fragment or two independent fragments, however. This necessitates statistical techniques for analysing radiation hybrid (RH) data, similar to linkage data, whereby maximum likelihood methods calculate orders for sets of markers and the distances between them (measured in cRays, variant according to the radiation dose).

RH mapping of DNA markers can overcome several of the difficulties encountered in genetic linkage mapping. Firstly, it involves the analysis of a single copy of the human chromosome of interest, unlike meiotic mapping, in which two copies of a human chromosome must be distinguished from one another by human DNA polymorphisms. Therefore nonpolymorphic DNA markers can be used for RH mapping, which exploits human-rodent differences. Another advantage is that the range of resolution of RH mapping can be varied by altering the X-ray dose used to fragment the chromosomes. Furthermore, an unlimited number of RH cell lines can be produced from a single donor cell line, in contrast to genetic linkage mapping, where large numbers of pedigrees must be sampled in order to obtain sufficient meioses to map loci. Scoring of markers is, in principle, easier than genetic linkage mapping because it depends only the presence or absence of a band, and not on the deciphering of alleles. Relative physical distances along the chromosome are less likely to be distorted because no chromosomal radiosensitivities have been identified with current RH sets that would correspond to hotspots for genetic recombination in genetic linkage maps.
Figure 1.2 The hierarchy of approaches to human chromosome mapping. The cytogenetic map provide the lowest level of resolution; the distance between chromosomal features such as bands and breakpoints can be measured microscopically and the approximate position of markers (including genes) determined by in situ hybridization. The genetic map reflects the recombination frequency between linked markers and has a similar resolution to the radiation hybrid map. Physical maps may be generated by pulsed-field gel electrophoresis or by the ordering of overlapping sets of clones. The use of sequence-tagged sites (STSs) allows one to link physical and genetic maps and provides a starting point for a high resolution map of the DNA sequence. A broad range of FISH techniques can now complement most levels of the mapping process (section 1.3.1.2).
1.3.1.2 In situ hybridisation

In situ hybridisation (ISH) is a powerful technology for examining the chromosomal map location and the relative order of genes and DNA sequences within subchromosomal regions, in single cells, or in tissue sections. Historically, ISH has been performed using isotopically labelled probe sequences that are detected by autoradiography (Gerhard et al 1981). The comparable sensitivity of modern non-isotopic method provides a simpler and faster alternative. The most widely used procedure is fluorescence in situ hybridisation (FISH), using fluorescently-labelled probes which can be scored easily by eye under a fluorescence microscope (Trask 1991).

Hybridisation signals from 2 probes can be spatially resolved by FISH to metaphase chromosomes when the probes are only several hundred kb apart, but a minimum of 1-2Mb separation is required to establish their physical order (Trask et al 1991) because of the way genomic DNA is packaged in highly condensed metaphase chromosomes. Spatial resolution can be improved by cohybridising probes labelled with different coloured fluorescent dyes to less-condensed interphase chromatin (Trask et al 1991), where a fairly linear relationship exists between physical distance and genome order over the range 30kb - 1Mb (Trask et al 1989). The combination of metaphase and interphase nuclear mapping thus offers the opportunity to physically order genomic DNA segments with a resolution achieved by gel electrophoretic methods (section 1.3.2.1), and provides a bridge to interrelate physical and genetic linkage information (see Figure 1.2). However, the 3D arrangement of chromatin fibres in interphase nuclei prevents resolution of probes >1Mb apart (Trask et al 1989). New FISH techniques have recently been developed to extend the resolution range of visual mapping from several megabases to only a few kilobases (Heiskaren et al 1996).

For ordering of sequences across the interphase-metaphase boundary (the 1-3Mb range), the use of chromatin fibres released from interphase nuclei (free chromatin) can provide a solution. When mechanically stretched out on a microscope slide in a linear fashion, free chromatin can greatly simplify the ordering of DNA segments to a resolution of 10-20 kb (Haaf and Ward 1994). In addition, probes separated by greater distances can be mapped without the constraint of the nuclear envelope, making this technique particularly appropriate for situations requiring long-range genome mapping with high resolution, such as the orientation and ordering of YAC and cosmid contigs, including approximate sizing of gaps and overlaps (Haaf and Ward 1994). However, mechanically stretched chromosome preparations are characterised by a high variability in target stretching so this strategy cannot be used for direct measurements of probe distances (Laan et al 1996).

Depleting the chromatin of its proteins to produce further extended ‘naked DNA’ fibres (up to 200% of the original length) enables accurate distance and overlap measurements over a 1-500kb range (‘fibre-FISH’; Parra and Windle 1993). The ‘string’ of signal obtained makes this technique useful for the precise positioning of markers and genes within long-range clones (Weier et al 1995).
1.3.2 Molecular methods of physical mapping

Increased mapping resolution and the ability to determine precise physical distances at the nucleotide level have been provided by advances both in fractionation and cloning of DNA. Molecular mapping methods involve cloning a genome or fragments of a genome by recombinant DNA methods, followed by ordering of the clones to create a physical map.

1.3.2.1 Pulsed field gel electrophoresis for size separation of large DNA molecules

Conventional gel electrophoresis is limited to the separation of fragments smaller than about 50 kb, due to the pore sizes of the polyacrylamide or agarose gel matrices through which separation is carried out. Clearly this is not suitable for the large scale mapping of mammalian genomes as the regions to be mapped are often in excess of 1 Mb.

Pulsed field gel electrophoresis (PFGE) was introduced by Schwartz and Cantor (1984) as a means to separate DNA molecules as large as 10 Mb using modified electrophoretic apparatus. It has helped fuel a revolution in molecular genetics because it permits rapid examination of very large chromosomal regions through construction of long-range restriction maps, thus filling the resolution gap between cytogenetic analysis and cloning vectors that could accept 20-40kb of DNA (Figure 1.2).

The principle behind PFGE is based on the observation that very long DNA chains orient themselves in an electric field, particularly in the presence of an agarose gel matrix, and migrate end-first through pores in the matrix (termed reptation). Once oriented with a pore, all large DNA molecules (>50kb) migrate at a rate independent of size and to the same position on gels. DNAs smaller than 50kb are able to enter the agarose pores as random coils and fractionation according to size occurs. The application of electric fields that change strength or direction over time forces DNA molecules to periodically reorient in new directions, facilitating the likelihood of orienting with a pore in the gel, but only as a function of size; small molecules will reorient more readily than large ones in a given time, and thus will migrate more quickly through the gel. In this fashion, molecules greater than 50kb could be fractionated according to size. Each field is turned on for a set time (the pulse time) and then off while the second field acts. Over the course of a run this process can separate molecules in the megabase size-range.

Modifications of the original PFGE method of Schwartz and Cantor have since been developed using alternative electric field geometries. The implementation of homogeneous electric fields and obtuse reorientation angles resulted in the DNA running straight in the tracks thus improving resolution (e.g. contour-clamped homogeneous electric field or CHEF; Vollrath et al 1987).

The DNA to be run is prepared in agarose blocks to prevent shearing, and may be run intact (e.g. yeast chromosomes) or cleaved with rare-cutting restriction enzymes (e.g. mammalian genomic DNA) for long-range restriction mapping to generate detailed organisational and structural information about a chromosome region. This involves Southern blotting of the gel followed by hybridisation with appropriate DNA markers. Physical linkage
of closely spaced markers on specific restriction fragments can provide orders and physical distances between markers to generate a detailed map. Partial digests can also be used to generate maps over much larger regions and to link more widely-spaced DNA probes. The use of restriction enzymes which identify CpG islands (known to be associated with the 5' end of most genes; Bird 1987) can give clues about the density of gene coding sequences in the region, and may provide a good start for gene cloning attempts (section 1.4).

The physical size of a candidate region for the Huntington’s disease (HD) locus was defined by PFGE analysis (Bates et al. 1991). The hybridisation of different DNA markers to the same DNA fragments allowed the linking of previous partial maps and permit an accurate size of 2.5 Mb to be determined for the HD critical region. This study also demonstrated the value of physical size determination of the region as opposed to relying on correlations between genetic and physical distance. The markers D4S10 and D4S125 were genetically separated by 3.5cM but their physical separation was less than 600kb. Conversely, the markers D4S125 and D4S113 had a genetic separation of 1.2cM but were found to be physically separated by almost 2Mb.

1.3.2.2 Generating long-range physical maps of cloned DNA

Once an accurate definition of the size of the disease gene critical region has been obtained, a decision can be made on whether to attempt to clone the region and search for candidate genes, or whether more DNA markers and physical and genetic linkage information should be sought. In practice both cloning of the locus and refinement of the genetic map take place simultaneously.

The flanking and intervening DNA markers may be used to screen libraries of genomic DNA. In early studies these would have been cloned into cosmid or phage vectors. Chromosome walks are undertaken by isolating endfragments from the cloned DNA and using these to rescreen the libraries to isolate overlapping clones. In this way whole chromosomal regions could be isolated in contiguous clones (contigs), but is an extremely labour-intensive process as the size of each ‘step’ is restricted by the DNA-insert capacity of these vectors (45kb in cosmids, 23kb in phage).

1.3.2.3 Yeast artificial chromosomes as a genome mapping tool

The development of YACs as a cloning vector (Burke et al. 1987) has dramatically facilitated the cloning of large genomic regions. By providing isolated fragments of human DNA of up to 2Mb, easily exceeding the size of a typical cosmid/phage/plasmid contig, chromosome walks can be undertaken relatively quickly due to the large possible size of each ‘step’. YACs therefore affect the achievement of long-range continuity in two main ways:

- They can provide even large genes intact and in proper context, thus assisting in their isolation and permitting analysis of long-distance regulatory elements that are frequently absent from typical cosmid constructs. Where no YAC can be isolated that contains the entire gene, the high activity of yeast homologous recombination can be exploited to recombine
overlapping YACs and achieve clones with the gene intact (e.g. the CFTR gene; Green and Olsen 1990). This ability to manipulate YACs genetically and physically as yeast chromosomes can also be exploited in generating nested deletion derivatives for fine mapping of clone inserts (Pavan et al 1991).

- YACs can provide coverage of targeted regions of several Mb, thus bridging the gap that had existed between the insert capacity of previous cloning vectors and cytogenetic analysis (similar to PFGE). YAC contigs are especially useful in the targeted search for disease genes, where flanking probes obtained by cytogenetic studies or genetic linkage define a critical region of several Mb. Another advantage of such large-insert contigs is the ability to identify and physically characterise 'gene clusters', related in tissue expression, evolutionary origin, or coregulated in development (e.g. the human type-1 interferon gene cluster on chromosome 9p; Diaz et al 1994). The extensive representation of genomic sequences in YACs (Coulson et al 1988) highlighted the feasibility of complete coverage of the human genome, as a few hundred would theoretically be sufficient to span an entire chromosome. When the entire map becomes available (section 1.5.2), targeted cloning will become superfluous as YACs will have been identified for markers flanking any particular disease gene such that whole contigs spanning the region of interest can be readily obtained.

While large-insert YACs are superior to smaller-insert clones for the rapid development of long-range physical maps, they are not ideal for the fine-structure analysis that must follow. One major problem is the high percentage (40-60%) of chimaeric clones i.e. those that contain artefactually linked DNA segments from non-contiguous regions of the genome, while ~10% of YAC clones carry two independent YACs because of cotransformation. Chimaerism may be due to co-ligation of DNA inserts in vitro prior to yeast transformation, or to recombination in vivo between two DNA molecules that were introduced into the same yeast cell. Chimaerism causes problems in chromosome-walking and gene-isolation experiments but can usually be detected by methods such as metaphase FISH or YAC end isolation and subsequent mapping. A second problem with YACs is that some clones are unstable and tend to delete internal regions from their inserts. This can be deleterious when constructing physical maps of chromosome regions or in efforts to isolate genes, and is difficult to detect by FISH or YAC end characterisation. Finally, the low yield of YAC insert DNA and its structural similarity to the yeast chromosomal DNA precludes its isolation by simple methods, but is necessary for gene isolation attempts. Preparative PFGE can be used to separate the YAC from the endogenous yeast chromosomes if the YAC does not comigrate with one of the other yeast chromosomes, as can direct subcloning of the entire yeast genome followed by identification of human-specific clones derived from the YAC.

With these deficits in mind, alternative cloning systems based on bacterial host systems have been developed in the past few years, which are capable of carrying inserts of 100-300kb (Monaco and Larin 1994). These new vectors, along with the more traditional cosmid vectors, are proving to be very useful in complementing YAC contigs in positional cloning projects.
Contigs comprising these smaller, more stable clones also provide more convenient templates from which to isolate polymorphic markers for disease gene refinement, and subsequently, candidate gene sequences.

### 1.4 Identification of gene sequences from a defined genomic interval

Linkage and physical mapping studies can often localise a gene to a region 1cM in size. Although small by genetic standards, a 1cM region may contain upwards of 50 genes (Fields et al. 1994). Thus the rate limiting step in positional cloning is the identification and characterisation of genes in the region. The most common approaches utilise genomic clones spanning the disease gene critical region and fall into two categories: those based on cross-hybridisation with probes derived from RNA (dependent on the gene in question being expressed in the relevant tissue) or those detecting features of genomic DNA that are universally associated with coding sequences.

#### 1.4.1 Northern blot analysis

Screening genomic clones for hybridisation to RNA in Northern blots requires that the potential transcript be present in the tissue(s) from which the RNA was derived. This method is limited by the generation of DNA fragments small enough to function as sensitive probes (YAC inserts are too complex and must be subcloned into cosmids or phage) and the fact that each probe must be analysed individually, and is not therefore suitable for large genomic intervals. Subclones that detect transcripts can be used to screen cDNA libraries to isolate the corresponding transcript.

#### 1.4.2 Cross-species sequence homology

Based on the observation that coding sequences are much more strongly conserved during evolution than non-coding sequences, this analysis involves low-stringency Southern hybridisation of a candidate DNA fragment to DNAs from a variety of species (a ‘zoo-blot’) to determine whether the candidate cross-hybridises across species boundaries. This procedure suffers from the same limitations as Northern blot analysis (section 1.4.1).

#### 1.4.3 Detection of CpG islands

Although vertebrate DNA is highly methylated compared to that of invertebrates, a small fraction (1% of the genome) differs from bulk DNA by being non-methylated at CpG, and relatively high in GC content (Bird 1987). Sequences with these characteristics occur at discrete ‘islands’, usually 1-2kb long, that are dispersed at about 30,000 sites in the genome. The abundance of CpG in these islands is actually attributable to the absence of a CpG deficiency i.e. CpG in islands occurs at the frequency predicted by base composition, whereas elsewhere in the genome CpG is present at less than 25% of its expected frequency, due to the propensity of 5methyl-cytosine to undergo deamination to form thymidine (Cooper and
Since CpG islands are frequently associated with expressed sequences, and can be detected by cleavage with rare-cutting restriction enzymes that have CpG-containing GC-rich recognition sites, this technique provides a rapid means for assaying a large stretch of DNA (Lindsay and Bird 1987). Clones that lie near CpG islands can then be used as probes against cDNA libraries. It must be noted, however, that not all CpG islands detected in cloned DNA will be present in genomic DNA due to differences in methylation, and that not all genes are associated with CpG islands.

1.4.4 cDNA selection

A variety of methods have been devised which involve the hybridisation of PCR-amplified cDNAs to immobilised target genomic clones, and circumvent many of the drawbacks of the above procedures (reviewed in Parrish and Nelson 1993). The most widely used approach entails first quenching of repeated sequences in the immobilised target DNA (purified YAC or cosmid DNA) which is then biotinylated. PCR-amplified cDNAs are then hybridised to the genomic probe in the presence of Cot 1 DNA (enriched for human repeat sequences) to suppress repeats in the cDNA. Streptavidin-coated magnetic beads are used to capture the genomic probe and attached cDNAs, which are then eluted and PCR amplified. After several rounds of further selection substantial enrichment of those cDNAs specifically hybridising to the probe is achieved; these are then cloned, sequenced and hybridised back to the YAC or cosmid target to verify their genomic location.

This technique enables detection of rare transcripts, can detect human genes that have diverged from those of other species, can be applied to several genomic probes simultaneously, and does not require YACs to be subcloned which speeds the progression of positional cloning projects.

1.4.5 Exon trapping

Strategies based on the screening of cDNA libraries allow the isolation of those cDNAs corresponding to mRNAs expressed in the cell line being screened. However, if expression of the coding sequence is restricted to a limited number of tissue types, or a specific stage in development, or if the mRNA species has low stability or very low abundance, then such approaches may be unsuccessful. Innovative techniques have therefore been devised that enable identification of exons directly from genomic clones.

The first such scheme (exon-trapping) was developed by Duyk et al (1990). The strategy involves shotgun subcloning genomic DNA into a trapping vector that contains a splice donor site. Further manipulations, including transfection into mammalian cells, then enables clones that contain a splice acceptor site to be identified. However, this method is sensitive to cryptic splice acceptor sites, necessitating rescreening of candidates to identify true exons. The exon amplification scheme of Buckler et al (1991) is more reliable as it requires the presence of both splice acceptor and donor sites in the insert DNA. An exon present in the correct orientation in the genomic insert will be spliced into the mature mRNA and will be present in
the amplified product flanked by known sequences. Exon trapping schemes are theoretically
able to detect all of the internal exons in a given region of interest, regardless of their states of
transcriptional activity. Although this method has been applied to several regions with
excellent results (e.g. The Huntington's disease collaborative research group 1993), it is not
always successful and is best used in conjunction with another technique.

1.4.6 Computer-based DNA sequence analysis

Computational advances have enabled the development of sequence-based methods for
identifying potential exons in genomic sequence data. Comprehensive databases (e.g.
Genbank) allow for sequence comparison both within and between species by computer
programs such as BLAST (Altschul et al 1990) or FASTA (Pearson and Lipman 1988);
sequence conservation may indicate the presence of an exon or gene identity may be instantly
apparent from homology to existing gene sequences. More recent analytical methods attempt
to detect novel genes in genomic DNA by recognising base patterns characteristic of coding
sequences, such as the gene recognition and analysis internet link (GRAIL) which can identify
90% of coding exons of 100bp or more (Uberbacher and Mural 1991). This strategy can thus
highlight clones containing coding sequences. One drawback to searching for genes in this
manner is that, when a large region is to be searched, huge amounts of sequence data must first
be generated. Alternative methods for identifying likely coding regions (such as those
outlined in sections 1.4.1-1.4.5) could be used to narrow the scope of the sequencing project:
computational systems could then be used to detect potential exons within these sequences.

1.5 Whole genome mapping efforts: The Human Genome Project

The Human Genome Project was instigated in 1988 as an international research initiative
to produce detailed genetic and physical maps of each of the 24 different human
chromosomes, with the ultimate aim of determining the sequence of the 3 billion nucleotides
that make up human DNA. The proposal stemmed from the realisation that detailed reference
maps of all human chromosomes would be an essential aid to the search for human genes. The
entire DNA sequence would allow determination of many important structural and functional
attributes of human individuals, as well as an unprecedented insight into genomic evolution
when compared with that of other organisms. A scientific plan for the U.S. Human Genome
Project was published in the spring of 1990, with specific goals for the first 5 years:
1. To develop a 2-5cM resolution human genetic map comprised of polymorphic STSs.
2. To assemble a whole genome physical map comprising STSs evenly-spaced at 100kb
intervals, and overlapping sets of cloned DNA with 2Mb continuity for much of the genome.
3. To improve DNA sequencing technology and simultaneously sequence 10Mb human DNA.
4. To conduct parallel mapping and sequencing on selected model organisms to assess new
technologies and provide comparative information to complement the human genome studies.
5. To develop software and databases to support large-scale mapping and sequencing projects.
6. To research the ethical, legal and social implications of the project.
   This initial set of goals has now essentially been achieved (Cox and Myers 1996). Coordination has been facilitated by staging ‘human chromosome workshops’ (managed by the Human Genome Organisation HUGO) to encourage collaboration and speed the completion of consensus chromosome maps.

1.5.1 A common language for physical maps: the sequenced-tagged site

The introduction of the sequence-tagged site (STS) as the common landmark in physical map construction by Olsen in 1989 has proved of paramount importance; by ‘translating’ all types of mapping landmarks (e.g. restriction sites, genetic markers, cloned DNA probes or gene sequences) into a short tract of single-copy DNA sequence that can be recovered at any time by PCR, map data from various mapping methods can be more effectively combined, thus maximising the world-wide resources available for formatting a high resolution physical map (Ward and Davies 1993). Using the same set of markers to generate multiple maps also allows inconsistencies in emerging maps to be identified and corrected. STS information can be readily disseminated throughout the mapping community via computer databases, such that any STS can be regenerated in any laboratory with access to primer-making facilities and appropriate libraries of cloned DNA. This alleviates the need for permanent clone archives and protects against ‘clone obsolescence’. STS-formatted YAC-based physical maps are also important intermediates in producing a ‘sequence-ready’ physical map consisting of smaller and more stable clones.

1.5.2 Current physical maps of the human genome

The CEPH/Genethon collaboration first tackled the task of creating a whole genome physical map in 1993, with their ‘first generation physical map of the human genome’ (Cohen et al 1993). Physical map construction was directly based on integration with the genetic map (Weissenbach et al 1992) by extensive analysis of the 33,000 YACs in the CEPH YAC library with >2000 genetic markers distributed over 90% of the genome. Additional mapping techniques were employed to establish homology relations among clones and permit integration of the genetic, physical and cytogenetic maps: YAC crosshybridisation, fingerprinting and metaphase FISH of genetically anchored YACs. This map has since been improved by CEPH/Genethon in collaboration with WI/MIT and other investigators worldwide (Chumakov et al 1995), exploiting the updated genetic map reported by Gyapay et al (1994) as a framework for physical map construction. New STSs were also integrated including YAC endclones and unpublished tetranucleotide microsatellites from the WI/MIT/CEPH/Genethon linkage mapping efforts, to validate the map, refine the framework order and strengthen contigs by increasing their density. This clone-based map covers about 75% of the human genome in 225 contigs having an average size of about 10Mb, and with CEPH genetic markers distributed at an average spacing of 1.2Mb.
The latest map is STS-based (Hudson et al 1995) to circumvent the problems associated with YAC chimaerism/instability and provide a more suitable scaffold for sequencing the genome. This map is the most comprehensive to date, incorporating 7000 genetic markers from Genethon and CHLC and 3000 ESTs from dbEST to give a fully integrated map of 15000 STSs spaced on average every 200kb. The distinguishing feature of this map is the use of radiation hybrid mapping (section 1.3.1.1) for 10500 STSs to supplement the genetic map in providing a global framework for long-range order and orientation of YAC contigs. This approach also gave a statistical measure of the accuracy of STS order and an estimate of the physical distance between adjacent STSs. As estimates of genome coverage based on the RH and YAC maps are 99% and 94% respectively, this map provides the necessary physical reagents for most positional cloning projects; efforts are now underway to increase the STS resolution to 100kb to bring the map to a sequence-ready format.

1.5.3 The final phase: sequencing the human genome

Now that the human genetic and physical maps fulfill the first two short-term goals of the Human Genome Project, with any gaps likely to be bridged by small groups conducting fine analysis on particular subregions, attention is turning to sequencing. Proponents of the ‘junk DNA’ hypothesis encourage first sequencing the functional genes (3% of the genome) to provide rapid access to those underlying genetic disease (Sikela and Auffray 1993). To this end, several large groups have been actively generating single-pass sequences (expressed sequence tags or ESTs) from random cDNA clones.

Adams et al (1995) produced an impressive list of 174,472 ESTs from 300 cDNA libraries constructed from 37 distinct organs and tissues. These ESTs, when combined with those in the database dbEST (Boguski et al 1993), comprised 83Mb of human DNA sequence. The inherent redundancy of random cDNA sampling (abundant mRNAs will be represented by many ESTs) was exploited to assemble ESTs into overlapping contigs, reflecting the mRNA composition of the source library, and in many cases representing full-length transcripts. This assembly process yielded 87,983 distinct sequences, 10,214 of which matched previously known genes upon database sequence similarity searches. Analysis of their tissue distribution revealed a high level of tissue-specific genes and a surprisingly low number of broadly expressed, relatively abundant transcripts. Although partially due to the sampling strategy, this nevertheless suggests much greater differences in gene expression among different cell types than was previously assumed. Hillier et al (1996), part of the Washington University-Merck Collaboration, have recently published over 300,000 EST sequences derived from ~180,000 IMAGE cDNA clones (Integrated Molecular Analysis of Genomes and Their Expression Consortium; Lennon et al 1996) which they estimate correspond to ~40,000 distinct genes.

However, only once their precise subchromosomal localisation is known will the systematic partial sequencing of thousands of random cDNA clones provide meaningful reagents for the rapid assessment of genes responsible for inherited disorders. Although cDNA sequencing in itself does not reveal the intron/exon organisation of a gene, or the nature and position of the relevant control elements, the corresponding genomic clones can be isolated.
once a map position renders them worthy of further investigation. Therefore STSs are now being developed from these ESTs to enable their placement on the physical map by somatic cell/radiation hybrid and YAC analysis (Boguski and Schuler 1995; Berry et al 1995). As most ESTs are derived from oligo(dT)-primed cDNAs, they represent the 3’ untranslated regions (UTRs) of the corresponding mRNA which facilitates STS design for two reasons: 3’UTRs rarely contain introns hence PCR product sizes are predictable and assays more reliable (Wilcox et al 1991), and these regions are more gene-specific and species-specific than the 5’UTRs or coding portions and thus allow discrimination among members of highly conserved gene families (Makalowski et al 1996). Already 13,600 ESTs have in this way been placed on the human physical map (Hillier et al 1996); when complete, the resulting transcript map should greatly accelerate the search for disease genes via the ‘positional candidate approach’.

1.6 Positional candidate approach to the isolation of human disease genes

The resources provided by the Human Genome Project have shaped a more streamlined method of identifying disease genes which combines pure positional cloning with the fortuitousness of the candidate gene approach (Collins 1995). This strategy relies on a combination of mapping to the correct chromosomal subregion (generally by linkage analysis) followed by a survey of the interval to see if attractive candidates reside there. The future expanding success of the positional candidate approach is predicted on the basis of an increasingly dense transcript map (Figure 1.3a). When a new disease gene is assigned to a specific map position it will soon be possible to ‘query’ that region of the chromosome in a computer database and have access to a list of genes assigned to the same region which thus become candidates for the disease. Strong contenders can then be screened for mutations in affected individuals. In an analogous way, when a new gene is mapped to a region one will have access to a list of all disease loci mapped to that region. Figure 1.3b shows examples of data that may be available to assist in this process of ‘partner’ identification.

1.7 Inherited retinal degenerations

Inherited diseases that cause the retina to degenerate, leading to either partial or total blindness, affect approximately 1 in 3000 people. Rapid progress is being made in identifying the genetic causes of many forms of inherited retinal disease: linkage studies of large families and candidate gene screening of known retinal genes have already identified 59 independent genetic loci that can cause retinal degeneration (Figure 1.4; Sullivan and Daiger 1996). An emerging pattern is the exceptional heterogeneity of these diseases, with growing examples of genetic heterogeneity (mutations in different genes giving rise to the same phenotype), allelic heterogeneity (different mutations in the same gene causing either the same or different
Figure 1.3(a) Trends in methods used for cloning human disease genes 1980-2010. Trends after 1995 are highly speculative (after Collins 1995).

Figure 1.3(b) Schematic representation of the positional candidate approach. When a new disease locus is assigned to a chromosomal region (shown in middle) candidate genes from the same region are analysed for features that match those of the disease. Similarly, when a new gene is assigned to a chromosomal region candidate diseases are analysed in an analogous manner. (from Ballabio 1993).
Cloned and/or mapped genes causing retinal degeneration in humans (red) and retina-specific genes (green). Disease genes are labeled to the right of the chromosome and retina-specific genes are to the left. Cases where the gene causing the disease has been identified are shown boxed in gold. Abbreviations for cloned disease-causing genes are listed below, in alphabetical order. References for all of the genes labeled can be found in the Genome Data Base, available on the World Wide Web at: http://gdbwww.gdb.org/gdb/browser/docs/topq.html. Cloned retinal genes: BCP, blue cone pigment; CHM, choroideremia protein; CNCG1, cyclic-GMP-gated channel α subunit; DMD, dystrophin; GCP, green cone pigment; MYVIIA, myosin VIIA; NDP, Nore disease protein; OAT, ornithine aminotransferase; PDE6A, cyclic cGMP phosphodiesterase, rod-specific α subunit; PDE6B, cyclic GMP phosphodiesterase, rod-specific β subunit; PLCB4, phosphoinositide-specific phospholipase Cβ4; RCP, red cone pigment; RDS, peripherin/RDS/RHO, rhodopsin; ROM1, rod outer-membrane protein; RPGR, retinitis pigmentosa GTPase regulator; SAG, S-antigen (arrestin); TIMP3, tissue inhibitor of metalloproteinase. Retinal degeneration loci: adRP, dgRP, dominant or digenic retinitis pigmentosa; adRP, dgRP, MD, dominant or digenic retinitis pigmentosa or macular degeneration; AGS, Alagille syndrome; arRP, recessive retinitis pigmentosa; BCM, blue cone monochromatism; CBD, colorblindness, deuteronopic (non-progressive); CBP, colorblindness, protanopic (non-progressive); CBT, colorblindness, tritanopic (non-progressive); CHM, choroideremia; CSNB1, congenital stationary night blindness (non-progressive); DMD, Oregon eye disease; GA, gyrate atrophy; ND, EVR, Norrie disease or familial exudative vitreoretinopathy (non-progressive); OD, recessive Oguchi disease; RP3, retinitis pigmentosa; SFD, Sorsby fundus dystrophy; USH1B, Usher syndrome 1B.
phenotypes) and clinical heterogeneity (the same mutation causing different symptoms in different individuals, even within the same family).

The online database of human genetic diseases OMIM (section 1.2.6) lists over 100 distinct diseases that include some form of retinal degeneration. These can be classified by mode of inheritance: autosomal dominant, autosomal recessive, X-linked and digenic forms are well documented (Sullivan and Daiger 1996) in addition to a small number of families where retinal degeneration is associated with other symptoms (e.g. RP and hearing loss in Usher’s syndrome). The genetic heterogeneity observed in these diseases suggests that retinal degeneration is a common endpoint for many different biochemical abnormalities, possibly due to the limited repertoire of responses that the retina shows to disease. The clinical heterogeneity in retinal disease is poorly understood, but implies that genetic background and environment must play significant roles in clinical expression. One reason for this complexity is the broad definition of ‘retinal degeneration’: this term encompasses diseases of the peripheral retina such as retinitis pigmentosa (RP) and congenital stationary night blindness (CSNB), diseases of the central retina such as macular degeneration, and many others in which the pattern of degeneration is complex.

The key to unravelling the complexities of retinal degeneration is to first understand the structure and function of the normal retina in the visual process.

1.7.1 Structure of the retina

The retina is the neural sensory layer of the eye and contains a dense array of light-sensitive photoreceptors containing visual pigment molecules that initiate the neural response to light (phototransduction) which has been focused onto the retina by the cornea and lens (Figure 1.5). Before light can reach the photoreceptors in the innermost layer of the retina, it must first penetrate the many nerve cell bodies and processes of the remaining retinal layers. Structurally and functionally the retina consists of two distinct layers: the non-neural retinal pigment epithelium (RPE), and the tightly opposed neural retina (Figure 1.6).

1.7.1.1 Neural retina

The vertebrate retina has six major neuronal cell types, organised into 5 distinct layers: 3 layers of cell bodies (the outer nuclear layer, inner nuclear layer and ganglion cell layer) and two layers of synaptic connections (the outer and inner plexiform layers). The photoreceptors (rods and cones), bipolar cells and ganglion cells form a connecting sequence that carries information from the region of light transduction in the external photoreceptor layer to the optic nerve fibres leading to the brain. The horizontal and amacrine cells form lateral synapses that relate activity across different visual fields, while interplexiform cells transmit signal back to the distal retinal layers and thus serve to modify the impulses by feedback (Figure 1.6). The visual signal flows from photoreceptors to bipolar cells and then to ganglion cells, whose axons join to form the optic nerve which relays the signal to the visual cortex in the brain for further processing and interpretation. Ganglion cells, and to a lesser extent bipolar cells, can
Figure 1.5 Transverse section of the human eye
Figure 1.6 A schematic diagram of the primate retina, showing most of the cell types and layers. The photoreceptors' terminals are greatly enlarged, to show their connections with other cell types. CC: choroidcapilaris; BM: Bruch's membrane; RPE: pigment epithelium; OS: photoreceptor outer segments; IS: photoreceptor inner segments; OLM: outer limiting membrane; ONL: outer nuclear layer; OPL: outer plexiform layer; INL: inner nuclear layer; IPL: inner plexiform layer; GCL: ganglion cell layer; OFL: optic fibre layer; ILM: inner limiting membrane.
have quite extensively spreading dendrites so that each ganglion cell may be influenced by the activity of a large number of rods and cones. The relationship varies with retinal location; in the fovea at the centre of the retina (which contains no rods) where visual acuity is highest, 200,000 densely packed cones connect with a similar number of optic nerve fibres (1:1 relationship), while in the far periphery, 10,000 rods may be connected in clusters to a single nerve fibre, such that ganglion cells here will respond to a pattern of illumination falling on a considerable area of the retina. Muller cells are glial cells whose fibres extend between the outer and inner limiting membranes and provide mechanical support for the retina.

1.7.1.1 Photoreceptors

Vertebrates possess two main types of photoreceptor, rods and cones, which can be distinguished by their morphology, absorption spectra and anatomical distributions. Rod-shaped photoreceptor cells contain the light-sensitive chromophore 11-cis-retinal, bound to the visual pigment rhodopsin (peak absorption ~500nm), are responsible for contrast sensitivity and vision in dim light (scotopic vision) and outnumber cones by about 20 to 1 in humans (100 million rods : 5 million cones). Rod-mediated vision is characterised by poor spatial resolution and high sensitivity. Cones are responsible for medium and bright light (photopic) vision, color sense and fine discrimination. In humans, three types of cone contain 11-cis-retinal bound to red, green or blue opsins, with peak absorption ($\lambda_{\text{nm}}$) at 560nm (long wavelength light), 530nm (medium) and 420nm (short) respectively, giving them different, although overlapping spectral sensitivities. The highest cone density is found in the macula, in particular the fovea, a small area of central retina on which incident light is focused. Cone density decreases sharply from the foveal centre, with a 90% reduction by 1mm in humans. Rod coverage follows an inverse pattern, increasing from 0% at the fovea to a maximum of 55-65% in far peripheral retina. Thus rods are essential for 'peripheral' vision, while cones are required for 'central' vision.

A schematic diagram of a vertebrate rod and cone is shown in Figure 1.7. The process of visual transduction, i.e. the conversion of photons of light into an electrical impulse, takes place in the outer segment. Electron microscopy of the outer segments reveals a stacked array of disc-like membranes which contain visual pigment (opsin) molecules as the major protein component. In rods the discs are constantly generated by successive basal invaginations of the outer segment plasma membrane, and as displaced by newer ones, eventually separate off from the plasma membrane and become internalised. In cones such internalisation does not take place and the discs represent a continuous, highly convoluted plasma membrane. This rod/cone difference in outer segment geometry is thought to reflect the fact that cones function in bright light and thus require rapid recycling of their pigment molecules, the continuum between the disc and plasma membranes providing rapid access of the regenerated chromophore from the cell exterior to the bleached pigment. Rod discs are continually shed from the rod apex and undergo phagocytosis by the retinal pigment epithelial (RPE) cells, thus maintaining the outer segments at a relatively uniform length. The inner segment of rods and cones is rich in mitochondria, ribosomes and golgi complexes and thus holds the
Figure 1.7 Diagrammatic representation of mammalian rod and cone photoreceptors (from Dulai 1996)
machinery necessary for highly active protein synthesis. The outer and inner segments are joined by a slender immotile cilium, which transmits components from the inner segment and cell nucleus to the discs and their plasma membrane.

1.7.1.2 Retinal pigment epithelium

The RPE comprises a single layer of cells with villous processes projecting from its apical surface interdigitating with the photoreceptor outer segments. The basal surface has numerous infoldings and is functionally linked to the choroid via Bruch’s membrane. The RPE contains light-absorbing pigment granules which control light scatter and permitting precise image formation. Tight junctions connecting the RPE cells also render it an important protective barrier between the choroidal circulation and the neural retina. Other important functions served by the RPE layer include supply of nutrients to the neural retina, production of interphotoreceptor space glycosaminoglycans, uptake and storage of retinoids and phagocytosis of rod, and to a lesser extent, cone outer segments.

1.7.2 The rod phototransduction cascade

Visual phototransduction covers the sequence of biochemical and electrophysiological events through which the absorption of a photon of light by a pigment molecule in a photoreceptor cell transmits an electrical nerve signal to the brain (Yau 1994; Koutalos and Yau 1996). This process is shown schematically in Figure 1.8 for the best-studied system, the vertebrate rod.

Absorption of a photon of light by rhodopsin generates activated rhodopsin (R*) by isomerisation of its bound 11-cis-retinal chromophore to 11-trans-retinal. The conformational change in rhodopsin exposes a binding site on the cytoplasmic surface for the G protein transducin, which is weakly bound to the disc membrane. Interaction between R* and transducin catalyses the exchange of GDP on transducin’s surface to GTP. In this process, the α subunit of transducin (Tα-GTP), charged with GTP, dissociates from the β and γ subunits and from rhodopsin, into the cytoplasm. R* is then free again to bind to a new transducin molecule. In this primary step the visual signal is amplified considerably, because one R* molecule is able to activate up to 500 transducin molecules. Tα-GTP then activates the enzyme cGMP phosphodiesterase (PDE). PDE consists of 4 subunits, α, β and two γs and is also weakly bound to the disc membrane. Activation is achieved by displacement from the enzyme of the two inhibitory γ subunits, which stimulates PDE more than 1500-fold. Activated PDE rapidly hydrolyses cGMP, which is bound to the rod photoreceptor cGMP-gated channel protein, a transmembrane cation-specific channel that in the dark, allows Na⁺ and Ca²⁺ to enter the outer segment across the plasma membrane. At least three cGMP molecules are bound to each channel protein, and the ability of one activated PDE molecule to hydrolyse thousands of cGMP per sec achieves the second signal amplification stage. The reduced levels of cGMP result in closure of the ion channels and subsequent hyperpolarisation of the rod cell plasma membrane. This generates a nerve signal by reducing the rate of transmitter release from the
Figure 1.8 Key components of the phototransduction cascade in vertebrate rod photoreceptor cells (from Polans et al 1996)

hv: incident photon
R: rhodopsin (* active)
Rec: recoverin
T: transducin
$T_\alpha^*$: activated $\alpha$ subunit of transducin
CaM: calmodulin

GC: guanylate cyclase
RK: rhodopsin kinase
PDE: cGMP phosphodiesterase (* active)
GCAP: guanylate cyclase activating protein
Arr: arrestin
Rod synapse. Channel opening is cooperative, which makes the channel highly responsive to small changes in cGMP level.

Rhodopsin, in the meantime, is deactivated as a combined result of phosphorylation by the enzyme rhodopsin kinase, and binding of a 48kd protein called arrestin blocking the further activation of transducin and therefore terminating phototransduction. All-trans retinal is released and recycled. Blocking the activated receptor is not sufficient to arrest the cascade; the Tα-GTP already formed will maintain the PDE activity by remaining bound to the γ subunit inhibitors until they lose this capacity due to the intrinsic GTP hydrolytic activity of Tα. The regenerated Tα-GDP has low affinity for the PDE inhibitors which immediately rebind to the catalytic units of PDE and block its activity. Tα-GDP then recombines with Tβγ and reassociates with the disc membrane. The system is reset to the dark-adapted state when rhodopsin is regenerated by recombining with 11-cis-retinal and when its bound phosphate is removed by a phosphatase.

1.7.2.1 Recovery of the dark state and the role of calcium

Recovery of the dark state is mediated by deactivation of PDE and activation of guanylate cyclase, in order that the rod be repeatedly responsive to light stimulation. In their open dark-adapted state the channels have a low permeability for Ca²⁺ and Ca²⁺ homeostasis is maintained by the counteracting effect of Na⁺/Ca²⁺ exchangers in the plasma membrane of the rod outer segment. After illumination, closure of the cGMP-sensitive channels blocks the influx of Ca²⁺ but its export by the exchangers continues. The resulting net drop in intracellular Ca²⁺ from about 500 to 50nM induces the activation of guanylate cyclase that in turn restores the dark level of cGMP in the cell and permits a reopening of the cGMP-gated channels. This Ca²⁺ action produces a negative feedback with two effects: firstly, a speeding of the recovery after illumination, and secondly, it provides a potential mechanism for light adaptation in steady illumination, by antagonising the light-induced increase in cGMP hydrolysis and thus down-regulating the light-sensitivity of the rod.

Ca²⁺ has several targets in the phototransduction cascade that modulate the negative feedback between cGMP and Ca²⁺ levels (reviewed in Polans et al 1996). The activation of guanylate cyclase is mediated by photoreceptor-specific guanylate cyclase activating proteins (GCAPs) in response to low Ca²⁺ levels. A second feedback pathway is mediated by another Ca²⁺ binding protein, recoverin, which inhibits the phosphorylation of activated rhodopsin at high Ca²⁺ levels with the net result when intracellular Ca²⁺ drops in the light being a decrease in the lifetime of activated cGMP PDE activity. Finally, Ca²⁺ decreases the affinity of the cGMP-gated channels for cGMP, so that when the concentration of Ca²⁺ falls in the light, some of the channels tend to reopen despite the decrease in concentration of cGMP, thus enhancing photorecovery and also leading to light adaptation.

Phototransduction in cones is thought to closely parallel that in rods, with many key rod proteins having cone counterparts (e.g. opsin, transducin, cGMP PDE, recoverin (visinin), cGMP-gated channel, Na⁺/Ca²⁺ exchanger; Yau 1994).
1.7.3  Retinal degeneration due to mutations in photoreceptor genes

When a photoreceptor cannot perform its function because of a structural or biochemical defect, it may degenerate. If the disease affects rods primarily, visual field loss progresses from mid-periphery towards the middle of the retina, leaving the patient with night blindness and 'tunnel vision' (e.g. retinitis pigmentosa). Conversely, diseases that primarily affect cones have the opposite effect, destroying the central retina (macula) where cones are most abundant. Several genes causing inherited retinal degeneration have been cloned and shown to encode photoreceptor-specific proteins, as described below. Their identification was accomplished primarily by way of the candidate gene or positional candidate approach (section 1.6; e.g. rhodopsin, Dryja et al 1990) on the basis that mutations in such indispensable photoreceptor-specific genes would likely cause retinal dysfunction. Mutations in these genes can either cause RP, CSNB or macular degeneration as outlined in several excellent reviews (Daiger et al 1994; Rosenfeld et al 1994; Dryja and Li 1995; Bird 1995; Sullivan and Daiger 1996). Such mutations can contribute to our understanding of both abnormal and normal visual processes and are summarised below.

1.7.3.1  Rhodopsin

Close to 100 mutant alleles of the rhodopsin gene have been identified as causes of RP. Almost all are dominant (comprising 25% of all ADRP cases) and most of these are missense mutations. The first mutation identified, in codon 23 (Dryja et al 1990), accounts for 12% of ADRP in the USA, but has not been found elsewhere and therefore probably represents a founder effect.

The mechanisms by which recessive or dominant rhodopsin mutations lead to photoreceptor degeneration have not yet been established. Two of three recessive alleles, a nonsense mutation and a splice-site mutation, are not likely to encode a functional rhodopsin. So it appears that photoreceptor viability cannot be maintained without functional rhodopsin. For dominant mutations, haploinsufficiency is unlikely to be the pathogenic mechanism since carriers of at least one apparent null allele are phenotypically normal, showing that the photoreceptor remains healthy with only 50% of control levels of rhodopsin. Furthermore, transgenic mice expressing dominant missense mutations in a genetic background with 2 wild-type murine rhodopsin alleles develop photoreceptor degeneration. Thus, most dominant rhodopsin mutants appear to be gain-of-function or dominant-negative alleles. The majority affect either the normal folding of the molecule, its capacity to bind 11-cis-retinal, its effective transportation from the endoplasmic reticulum following synthesis or its incorporation into rod outer segments, or some combination. Congestion in the protein transport/maturation pathways could compromise the viability of the rod photoreceptors. Alternatively, those mutant proteins that do become incorporated into the outer segment disc membranes may result in structural destabilisation of the membranes themselves, as rhodopsin comprises 90% and 50% of rod disc and plasma membranes proteins respectively. Another proposed
pathogenic mechanism for certain dominant mutations invokes the ‘constant equivalent light’ model. This was inspired by studies of mutants altering the lysine residue at codon 296. Lys296 is the attachment site for 11-cis-retinal and also participates in holding the opsin in an inactive conformation when it is not bound to the chromophore. Mutations altering this residue encode opsin that cannot incorporate 11-cis-retinal and that constitutively activate transducin in vitro. Such mutants may cause cell death by overstimulating the phototransduction pathway, analogous to exposure to constant light which is known to cause photoreceptor cell death by inducing high levels of cGMP which are toxic to the cell. However, Lys296Glu opsin in transgenic mice is inactivated by phosphorylation and binding to arrestin. Thus Lys296Glu opsin in vivo does not activate the phototransduction cascade and therefore cell death must be mediated by a mechanism other than overstimulation. Two mutations that cause dominant congenital stationary night blindness generate opsins that bind chromophore normally, but can also activate transducin in the absence of chromophore, causing inappropriate activation of transducin when all-trans retinal leaves opsin during the regeneration cycle. This minor abnormality is sufficient to diminish sensitivity to dim light but not sufficient to damage the photoreceptor.

1.7.3.2 Rod cGMP phosphodiesterase α and β subunits (PDEA and PDEB)

Progress in gene defect detection in autosomal recessive RP, accounting for up to 50% of all cases, has been slower as fewer large families are available for linkage analysis. However, mutations in the genes coding for PDEA and PDEB have recently been identified as the cause of RP in some recessive pedigrees. Nonsense mutant alleles in either gene result in the truncation of the putative catalytic domains and are thus likely to be null alleles. Such mutations in the PDEB gene and the resulting phenotype in human RP families parallel those found in the rd (retinal degeneration) mouse model of RP (Farber 1995), where loss of cGMP PDE activity leads to persistently elevated cGMP levels which are cytotoxic and cause photoreceptor death. Mutations in PDEB are also associated with autosomal dominant CSNB where the retarded hydrolysis of cGMP is presumed to be less severe, resulting only in reduced sensitivity of the photoreceptors to light.

1.7.3.3 cGMP-gated channel

Mutations in the gene encoding the α subunit of the channel have been found in a few families with autosomal recessive RP. Since the mutations are either obviously null (e.g. frameshifts early in the reading frame) or encode a channel protein that functions poorly, it appears that a paucity or absence of functional channels is deleterious to rod photoreceptors.

1.7.3.4 Peripherin/RDS

Mutations in this gene were first observed to result in a slow degeneration of photoreceptor cells in both homozygous and heterozygous mice (rds; retinal degeneration slow). The protein is localised to the outer segment disc membranes of both rods and cones,
and is thought to be essential for the assembly and physical stability of outer segment discs, and hence the outer segments themselves, since mice homozygous for the rds null allele fail to develop them, whilst heterozygous mice produce disorganised discs and shortened outer segments. This suggests that the semi-dominant phenotype in the rds null allele heterozygote (in contrast to the rhodopsin null allele heterozygote) is due to a loss of function (haploinsufficiency), a concept supported by the observation that the retinal abnormalities of rds mice are completely rescued by transgenic correction.

Mutations in the human peripherin/RDS gene have been found in families with autosomal dominant retinal degenerations. Most cases have the diagnosis of RP, while some mutations are seen to cause distinct forms of macular dystrophy (Dryja and Li 1995). Different diagnoses have been assigned even to relatives with the same mutation. Objective measures of photoreceptor function, such as ERGs, rod and cone sensitivities and rates of dark adaptation, show similar abnormalities regardless of clinical diagnosis. Hence, it is likely that the clinical heterogeneity is due not to variations in the underlying physiological defects but to epigenetic factors modulating the response of the retina to the primary genetic defect.

1.7.3.5 ROM1

Like peripherin/RDS, ROM1 is an integral membrane protein which is abundant at the rim region of rod outer segment discs. In fact, both ROM1 and peripherin/RDS interact to form a tetrameric complex, an association thought to be integral to the formation and stability of the bend in the rod disc membranes. Three RP families have been reported which segregate mutations in both the ROM1 gene and the peripherin/RDS gene (‘digenic’ inheritance). The ROM1 mutations were frameshift mutations early in the reading frame and likely to be null mutations. All 3 families segregated the same peripherin/RDS allele, the missense change Leu185Pro. Affected individuals were double heterozygotes for mutations in both genes, while heterozygote carriers of either the ROM1 or peripherin/RDS mutation were phenotypically normal.

The absence of ROM1 in cones implies that differences exist between the precise mechanisms by which peripherin/RDS stabilises outer segment membranes in the two classes of photoreceptor. In rods, the association between peripherin/RDS and ROM1 may be important, whereas in cones peripherin/RDS may bind to a different membrane protein or act alone. If the binding sites are different in rods and cones, constancy of one amino acid of the peripherin/RDS molecule may be important to rods only, and a mutation causing an abnormality at this site would cause a dystrophy falling within the category of RP, in which rods were the target cells of disease, with relative preservation of cones. Conversely, a different mutation in the peripherin/RDS gene may disrupt the metabolism or structure of either cones alone, causing macular dystrophy, or of both rods and cones.

1.7.3.6 Retinal guanylate cyclase (RETGC)

Missense and frameshift mutations in RETGC have recently been discovered to cause Leber’s congenital amaurosis, an autosomal recessive condition characterised by blindness at
birth or shortly thereafter, due to a defect in photoreceptor development or their early
degeneration (Perrault et al 1996). As cGMP production in photoreceptor cells is therefore
abolished in this disease, the excitation process is impaired due to consistent closure of the
cGMP-gated channels with hyperpolarisation of the plasma membrane. The cGMP
concentration cannot be restored to the dark level, leading to a situation equivalent to
consistent light exposure during photoreceptor development, and consequent toxicity.

1.7.4 The mechanism of photoreceptor cell death

As indicated above, for many photoreceptor mutations the cause of retinal degeneration
can be inferred, such as membrane instability, aberrant activation of phototransduction or
increased cGMP levels. Why these various conditions lead to degeneration is unclear. One
 possibility is a simple necrotic process wherein chemical/structural alterations triggered by
these mutations lead to a cessation of metabolic activity, cell lysis and phagocytic activity,
either by the RPE or infiltrating macrophages, to remove the cellular debris. However, recent
evidence implicates apoptosis, a form of programmed cell death, as the mechanism of
photoreceptor cell death in 4 different mouse models of retinal degeneration (see below).
Apoptosis occurs naturally as a normal process of embryonic development (reviewed in White
1996) and differs from necrosis in several other respects: it affects isolated cells rather than
patches of tissue, is not usually accompanied by inflammation or scarring, and apoptotic cells
are usually phagocytosed by adjacent cells, frequently without apparent involvement of
circulating macrophages. A hallmark of apoptosis is internucleosomal DNA fragmentation,
detectable by the appearance of a characteristic DNA ladder on agarose gel electrophoresis.

Neuronal death is thought to modulate the interconnections of developing neurons. In
mouse models of ADRP (rd, rds, and transgenic mice with different rhodopsin mutations) the
appearance of apoptotic cells began and/or extended beyond the stage at which developmental
neuronal cell death reaches a peak in normal animals and was restricted to the photoreceptor
layer (Chang et al 1995). Thus it appears that retinal cells use apoptosis not only during
development for fine tuning retinal cell number and interconnections, but also as a response to
aberrant stimuli such as those generated by mutations in key photoreceptor proteins. It seems
reasonable to suppose that an analogous activation of the apoptotic program occurs in the
diseased human retina.

It is not yet clear how each mutation triggers photoreceptor apoptosis, although
understanding the mechanisms that initiate photoreceptor degeneration could open the
possibility of intervention in this process with considerable impact on potential approaches to
therapy (reviewed in Gregory and Bird 1995). Internucleosomal DNA fragmentation is
thought to be mediated by a nuclear endonuclease that can be triggered by a rise in Ca\(^{2+}\) ions,
which may explain apoptosis in the rd mouse where a rise in cGMP follows lack of a
functional PDE complex. Pivotal work in the nematode C. elegans has identified 14 genes
involved in the triggering or blocking of apoptosis, two of which are essential for apoptosis
(ced-3 and ced-4) and are negatively regulated by a third gene ced-9, the human homologue
of which is bcl-2. Attempts have been made to identify gene expression during apoptosis in
the retina: for instance, clusterin is an apoptosis-inducible gene preferentially expressed in peripheral retina of the rd mouse. Clusterin expression is present during normal retinal development and increases with the onset of retinal degeneration, followed by a decrease paralleling the loss of photoreceptors in this model.

The interesting conclusion from the mouse models of retinal degeneration is that although the initial mechanisms of pathogenesis are different in the 4 models, it appears there is a final common pathway mediating cell death. Given the large number of mutations that cause RP (section 1.7) strategies of gene therapy aimed at correcting each individual mutation may be an overwhelming task. In contrast, strategies aimed at introducing a gene (e.g. bcl-2) that would interfere with the cells' ability to carry out apoptosis may be a much more practical approach because it would lead to preservation of sight despite the presence of multiple predisposing mutations. Moreover, because the degenerative changes in RP occur over decades, even modest decreases in the rate of cell death could significantly increase the number of years of useful vision.

1.8 The human X chromosome

The X chromosome is the most extensively studied of all human chromosomes due to a wide interest in X-linked diseases and the phenomenon of X chromosome inactivation (Lyon 1988). Both of these features are related to the different dosage of X in males and females; because males have only a single X, recessive diseases tend to be revealed, which accounts for the large numbers of X-linked diseases. As a result, the X chromosome was the first to have a genetic map based on RFLPs and systematic approaches to physical coverage were undertaken and expanded as part of the Human Genome Initiative. By 1993, 200 structural genes had been assigned to the X chromosome, together with 400 anonymous DNA segments (Sclessinger et al 1993). Genes for two X-linked diseases (chronic granulomatous disease and Duchenne muscular dystrophy) were the first to be isolated by positional cloning (Royer-Pokora et al 1986; Monaco et al 1986) and it was by elucidating the mutation mechanism of two X-linked diseases, fragile X syndrome and spinobulbar muscular atrophy (Sutherland and Richards 1995) that expanded trinucleotide repeats were discovered.

1.8.1 Mapping X-linked diseases

X-linked diseases have features that facilitate positional cloning. Chromosomal assignment is obvious because of the inheritance pattern, reducing mapping on the X chromosome to regional mapping. For several diseases, rare affected females have been found with balanced X-autosome translocations. In these patients, the normal X chromosome is generally inactive and the translocated X active, because of selection in early embryogenesis in order to maintain expression of genes on both autosomes. Translocations that have a breakpoint in a gene will lead to expression of the corresponding disease, as the uninterrupted copy on the normal X is inactive. Such translocations have been instrumental in the cloning of several X-linked disease genes (section 1.1.1).
Rare male patients exist with deletions (often detected cytogenetically) encompassing several megabases of DNA. In most cases the lack of function for genes in the deleted region results in a contiguous gene syndrome. In other words, several diseases are associated in a single patient, allowing very accurate mapping of the relevant genes if a series of overlapping deletions are available for analysis. This was first observed in the the case of the BB deletion encompassing part of DMD and genes for CGD (CYBB), McLeod syndrome (XK) and retinitis pigmentosa-3 (RP3) in Xp21 (Francke et al 1985). The BB deletion was instrumental for the cloning of the DMD and CYBB in 1986, and XK more recently (Ho et al 1994). It is likely that such regions have relatively low gene densities or the deletions would be lethal. In the Xp21.2 region this can be accounted for in part by the huge size of the DMD gene (~2Mb), but more generally it illustrates the great heterogeneity in gene density throughout the genome, with gene-poor regions (in general AT-rich, Giemsa dark bands) and gene-rich regions (GC-rich, Giemsa light bands) (Bickmore and Sumner 1989).

1.8.2 Mutations in X-linked genes

For X-linked diseases that severely decrease reproductive fitness in affected males, the number of mutations in each generation decreases by one-third (since males have one-third of the X chromosomes in the population) and the particular mutation becomes extinct after a few generations. The spectrum of mutations and incidence of severe X-linked diseases is due to the constant input of new mutations and directly reflects the mutational sensitivity of the gene. It is now clear that there is a striking difference in the deletion frequency for various diseases (reviewed in Mandel et al 1992); in X-linked ichthyosis, 80-90% of the mutations are large deletions encompassing the entire gene (many due to unequal recombination between flanking low-copy repetitive elements) while DMD has a high frequency of partial deletions (60-70%) and a considerable level of partial duplications (6-7%) (due in part to the large size of the dystrophin gene). For most other diseases analysed, the frequency of deletions or rearrangements detectable by Southern blot is about 5-15%. Analysis of the factor IV gene in cases of hemophilia B has revealed about 400 different point mutations and one case where the gene was disrupted by de novo insertion of an ALU repeat. The fragile X syndrome is is caused by an unstable expansion of a CGG repeat in the 5' exon of the FMR-1 gene beyond a normal threshold, which is correlated in patients with abnormal methylation of the adjacent CpG island leading to loss of expression of the gene transcript. In SMA, the mutation is a more moderate expansion of a CAG repeat in the NH2-terminal coding region of the androgen receptor gene (AR) while other heterogeneous mutations in AR result in the completely different phenotype of testicular feminisation.

The above examples illustrate the mutational diversity associated with most X-linked disorders as a result of the high proportion of families with new mutations, which can severely hamper precise localisation by linkage analysis. Even with a dense genetic map composed of highly polymorphic markers, most rare diseases will not be mapped to intervals smaller than 2-5cM (potentially containing 100 genes!), and precise localisation by linkage disequilibrium
studies is only advisable in homogeneous populations capable of showing a founder effect for the disease (section 1.2.7).

1.8.3 The genetic and physical map of the human X chromosome

A unified genetic, physical and transcriptional map of the human X chromosome is being built through a concerted, international effort. Two large groups (Nagaraja et al 1996; Crollius et al 1996) have recently published integrated maps specifically for the X chromosome, owing to its poor representation compared with the autosomes in the whole-genome physical maps (Chumakov et al 1995; Hudson et al 1995). This is principally due to the low representation of the X in the CEPH YAC library used to construct these maps, and the fact that they are based on microsatellite STSs which are lower in frequency on the X chromosome (Dietrich et al 1996). YAC contigs with an average clone depth of at least four-fold coverage now span more than 80% (125Mb) of this 160Mb chromosome (Crollius et al 1996), with an average inter-STS resolution of 85kb and a current total of around 2000 STSs in 5200 YACs (Nagaraja et al 1996).

Numerous YAC contigs have also been established in smaller regions surrounding specific disease genes in positional cloning strategies, and contigs are rapidly being linked by common STSs and clones such that the ‘consensus’ physical map is virtually at the point of completion (Nelson et al 1995). The gaps that remain suffer a variety of contigging problems: regions that are unstable or unrecovered in cloned DNA; sequence elements that are repeated along the X; clones that are chimaeric or contain large internal deletions. Closure will most likely accrue from efforts to construct higher resolution maps in bacterial cloning systems (section 1.3.2.3).

There are 6 major genetic maps for the X chromosome (Murray et al 1994, Matise et al 1994, Donnelly et al 1994, Wang et al 1994, Fain et al 1995, Dib et al 1996) consisting of RFLP’s and microsatellite markers. Basic features of these maps are summarised in Table 1.2. The maps are described as ‘framework’ (F) or ‘comprehensive’ (C) depending on the statistical support for marker order; 1000:1 (lod 3) is required for a framework map (detailed in section 5.1.2.1).

These genetic maps show a high degree of overall consistency and the genetic lengths are in good agreement with that postulated by Morton in 1991 (220cM). Differences in the resolution and genetic length among maps reflect differences in error screening, number of markers analysed, ordering criteria and map format. The higher resolutions of the 2D map (3.2cM; Fain et al 1995) and the map of Wang et al (1994) reflect the high marker density provided by map integration, and make them extremely valuable for positional cloning projects where it is more useful to know and refine the probable marker order in a dense map, both to refine disease gene intervals and to assist in locus assignment for X-linked disorders that display genetic heterogeneity (e.g. retinitis pigmentosa, Teague et al 1994; mental retardation; Lubs et al 1996). The high-confidence framework maps serve as a guide to low resolution linkage mapping.
Table 1.2 Genetic maps for the X chromosome

<table>
<thead>
<tr>
<th>Reference</th>
<th>Marker loci</th>
<th>Genetic map length</th>
<th>Resolution</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matise et al 1994</td>
<td>57 mixed markers (24 framework)</td>
<td>191 cM</td>
<td>8.9 cM (F)</td>
<td>Comprehensive map; Multimap algorithm</td>
</tr>
<tr>
<td>Donnelly et al 1994</td>
<td>62 PCR-based markers (30 framework)</td>
<td>236 cM</td>
<td>3.7 cM (C)</td>
<td>Comprehensive map; integration with cytogenetic map</td>
</tr>
<tr>
<td>Murray et al 1994</td>
<td>150 microsatellites (35 framework)</td>
<td>247 cM</td>
<td>8.2 cM (F)</td>
<td>Framework map; integration with cytogenetic map</td>
</tr>
<tr>
<td>Wang et al 1994</td>
<td>270 mixed markers (104 framework) (50 PCR-based)</td>
<td>211 cM</td>
<td>-</td>
<td>Comprehensive map; integrates genetic and physical data from XCW5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Schlessinger et al 1994)</td>
</tr>
<tr>
<td>Fain et al 1995</td>
<td>243 mixed markers</td>
<td>190 cM</td>
<td>3.2 cM (C)</td>
<td>2D comprehensive map; integrates maps of Matise et al, Gyapay et al and Murray et al (1994)</td>
</tr>
<tr>
<td>Dib et al 1996</td>
<td>216 CA repeats (81 framework)</td>
<td>198 cM</td>
<td>~ 2 cM (F)</td>
<td>Framework map; average marker HET 65%</td>
</tr>
</tbody>
</table>

Thus the human X chromosome genetic map is essentially complete in terms of marker density. Whole chromosome efforts to generate further markers would now offer diminishing returns; all that remains is consolidation of marker order and distances across the chromosome, which will most likely be established by integration with the developing physical map. With the YAC map near to completion, X chromosome mapping is now shifting towards STS-based map integration, EST assignment and long-range sequencing (Nelson et al 1995).

1.8.4 The current map of proximal Xp: Xp21.1-Xcen

Figure 1.9 shows the current published map of proximal Xp (Nelson et al 1995). The region is now well-covered with microsatellite markers which are being actively incorporated into the growing physical map. Four main YAC contigs now span this interval, generated by many groups working around specific disease genes. Gaps remain between the proximal end of the YAC contig in Xp21.1 containing the genes for XK, CYBB, RP3 and OTC and a 2-3Mb contig in Xp11.4. There is a second gap between this contig and a ~3Mb contig in Xp11.23-Xp11.22. These two contigs are discussed in detail in chapter 7.

1.8.4.1 Cloned genes in proximal Xp

At the outset of this project in 1991, 12 genes had already been cloned and localised to proximal Xp. These genes provided important anchor points for the mapping project described in this thesis, thus their localisations and functions are summarised in Table 1.3.
Figure 1.9 The proximal short arm of the human X chromosome consensus map (from Nelson et al 1995) (not to scale)

- YAC contigs are indicated by vertical bars
- an approximate genetic scale is indicated to the left (cM)
- microsatellite polymorphisms are denoted with an asterisk
- the Xp11.3-Xp11.23 contig is expanded to show detail
- DNA segments and genes shown to the left of the YAC contigs are only ordered with respect to breakpoints (BXP)
<table>
<thead>
<tr>
<th>Gene</th>
<th>Symbol</th>
<th>Location</th>
<th>Function</th>
<th>Associated disease</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>β subunit of cytochrome b</td>
<td>CYBB</td>
<td>Xp21.1</td>
<td>Reduction of oxygen to superoxide anion in the NADPH oxidase system in phagocytic cells</td>
<td>Chronic granulomatous disease (CGD); reduced resistance to infection</td>
<td>Royer-pokora et al (1986)</td>
</tr>
<tr>
<td>Ornithine transcarbamylase</td>
<td>OTC</td>
<td>Xp11.4</td>
<td>Mitochondrial enzyme that catalyses synthesis of citrulline from ornithine and carbamylphosphate in the urea cycle</td>
<td>OTC deficiency; hyperammonemia, metabolic and behavioural disturbances</td>
<td>Tuchman et al 1995</td>
</tr>
<tr>
<td>Monoamine oxidase A</td>
<td>MAOA</td>
<td>Xp11.4-11.3</td>
<td>Mitochondrial enzyme that catalyses the oxidative degradation of biogenic amines</td>
<td>MAOA deficiency; low normal intelligence, impulsive behaviour, cardiovascular problems</td>
<td>Chen et al 1992a, Brunner et al (1993)</td>
</tr>
<tr>
<td>Monoamine oxidase B</td>
<td>MAOB</td>
<td>Xp11.4-11.3</td>
<td>Mitochondrial enzyme that catalyses the oxidative degradation of biogenic amines</td>
<td>unknown</td>
<td>Chen et al (1992)</td>
</tr>
<tr>
<td>A-raf-1</td>
<td>ARAFI</td>
<td>Xp11.3-11.23</td>
<td>Proto-oncogene encoding a cytoplasmic protein kinase involved in cell growth and development</td>
<td>unknown</td>
<td>Derry and Barnard (1992), Lee et al (1994)</td>
</tr>
<tr>
<td>Tissue inhibitor of metallo-proteinase 1</td>
<td>TIMP1</td>
<td>Xp11.3-11.23</td>
<td>Glycoprotein inhibitor of matrix metallo-proteinases; regulation of connective tissue breakdown</td>
<td>unknown</td>
<td>Mattrisian 1990, Derry and Barnard (1992)</td>
</tr>
<tr>
<td>Synapsin</td>
<td>SYN1</td>
<td>Xp11.3-11.23</td>
<td>Synaptic vesicle phosphoprotein involved in regulation of neurotransmitter release</td>
<td>unknown</td>
<td>Bahler et al (1990), Derry and Barnard (1992)</td>
</tr>
</tbody>
</table>
The MAO genes play a key role in the deamination of neurotransmitters and have been localised by physical mapping within a 240kb interval 140kb centromeric to DXS7 and 50kb distal to the Norrie disease gene (NDP), in a tail-to-tail orientation (Chen et al 1992). Complete absence of MAOA activity caused by a single base substitution in exon 8 which renders the enzyme disfunctional has been linked to behavioural problems and hypertensive crises in affected males of one family (Brunner et al 1993).

Ubiquitinisation is a post-translational covalent modification of various cellular proteins that exerts a wide range of effects within all eukaryotic cells (Deshaies 1995). UBE1 catalyses the activation of ubiquitin into a high-energy intermediate, which can then be conjugated to proteins destined for degradation via ubiquitin protein ligases (Handley et al 1991). The UBE1 gene has been mapped to Xp11.3-11.23 using somatic cell hybrids and human cell lines with X chromosome rearrangements or deletions (Lafreniere et al 1991). Although it has been physically mapped to the distal end of the Xp11.23 contig as depicted in Figure 1.9, the precise localisation of UBE1 within this interval is unclear due to conflicting reports as discussed in chapter 7.

The genes for ARAF1, SYN1, TIMP-1 and PFC have been cloned and mapped to the proximal short arm of the human X chromosome by ISH and Southern blot analysis of somatic cell hybrids (Lafreniere et al 1991). ARAF-1 is a cellular oncogene first identified by virtue of its homology to the viral v-raf oncogene, and is thought to be a serine/threonine kinase (Lee et al 1994). It has been shown to function in a signal transduction pathway downstream of growth factor receptors, thus playing a critical role in cell growth and development. The SYN1 gene encodes neuronal actin-binding phosphoproteins 1a and 1b, which coat the cytoplasmic surface of small synaptic vesicles in the presynaptic terminal (Bahler et al 1990). It is proposed that SYN1 may connect synaptic vesicles to each other or to the cytoskeleton, modulated by phosphorylation of SYN1 through signal transduction pathways (Jovanovic et al 1996). Thus SYN1 may play a role in positioning synaptic vesicles in the nerve terminal, thereby regulating neurotransmitter release. TIMP-1 encodes a glycoprotein inhibitor of matrix metalloproteinases and is therefore extremely important in the regulation of connective tissue breakdown (Matrisian 1990). It also appears to act as an erythroid hematopoietic growth factor (Hayakawa et al 1992). PFC is a glycoprotein which functions as a positive regulator of the alternative pathway of complement. The importance of PFC in immune defense is illustrated by PFC-deficient individuals who display severe pyogenic bacterial infections with a high mortality rate. The recent discovery of a nucleotide change in exon 5 of PFC, generating a stop codon in affected males of one family, provided proof that PFC deficiency is caused by alterations in the PFC structural gene (Westberg et al 1995).

Genetic linkage mapping in mouse and man have demonstrated that the genes for ARAF1, SYN1, TIMP1 and PFC are clustered in Xp11.23 (Derry and Barnard 1992). These studies have been confirmed by direct physical mapping of these genes in man and mouse using YACs, PFGE and cDNA sequencing which showed that they lie within 70kb of each other in the order Xpter-5' ARAF1 3'- 3' SYN1-TIMP-1-SYN1 5'-3' PFC 5'- Xcen and that the
physical arrangement is conserved on the human and mouse X chromosomes (Derry and Barnard 1992). Detailed restriction mapping showed that TIMP-1 lies within an intron of the SYN1 gene and is transcribed in the opposite direction.

The structural gene for human ornithine aminotransferase (OAT), a mitochondrial matrix enzyme catalysing the interconversion of ornithine and D1-pyrroline-5-carboxylate, has been cloned and mapped to chromosome 10q26 (O’Donnell et al 1988). Deficiencies in the enzyme result in gyrate atrophy, an autosomal recessive disease characterised by progressive degeneration of the choroid and retina and a marked increase in plasma ornithine (Valle et al 1977). Sequences homologous to OAT, designated ‘OAT-like’ (OATL) have been localised close to the structural OAT gene at 10q26 and to proximal Xp. The localisation of the OATLX sequences has been refined using somatic cell hybrids and deletion cell lines and reveals two loci: Xp11.23 (OATLI) and Xp11.22-11.21 (OATL2) (Lafreniere et al 1991a). The hybridisation pattern of the OAT cDNA to these loci indicated that they comprise ‘clusters’ of reiterated OATL sequences, each contained in a distinct EcoRI fragment: OATLI contains 8 fragments and there are 4 in the OATL2 cluster. Sequence analysis of OATLX sequences suggests that they are processed pseudogenes as they lack introns, are flanked by direct repeats (Looney et al 1987) and do not code for OAT activity (O’Donnell et al 1988). A long-range pulsed field map spanning the OATLI locus has recently demonstrated a 275kb region containing multiple copies of several probes; several non-OATL sequences have homologous counterparts at the OATL2 cluster and are reiterated several times within each region (Chand et al 1995). One of these was discovered upon cloning translocation breakpoints t(X;18)(p11.2;q11.2) associated with synovial sarcoma and found to lie within the OATL pseudogene clusters (Crew et al 1995). The translocations produce hybrid transcripts SYT-SSX1 or SYT-SSX2 with the SYT gene derived from chromosome 18 and SSX1 or SSX2 derived from OATLI or OATL2 respectively. The SSX genes contain sequences indicative of a role in transcriptional repression; PCR primers that amplify the hybrid gene provide a diagnostic assay to detect SYT-SSX transcripts in 90% of synovial sarcomas (Crew et al 1995).

Synaptophysin (SYP) is an integral membrane glycoprotein of small synaptic vesicles in brain and endocrine cells (Wiedenmann and Francke 1985). The human gene was cloned and mapped to Xp11.23-11.22 using family linkage studies and Southern analysis of somatic cell hybrids containing different portions of the human X (Ozcelik et al 1990) and has been mapped by STS-content analysis of YACs and cosmids to the proximal end of the Xp11.23 contig as shown in Figure 1.9 (Boycott et al 1996; Schindelhauer et al 1996).

TFE3 is a basic helix-loop-helix (BHLH) and leucine zipper (LZ) motif-containing protein that binds specifically to the μE3 site of the immunoglobulin heavy-chain enhancer. TFE3 is ubiquitously expressed and functions as a positive-acting transcription factor, presumably by DNA/protein interactions mediated via its functional motifs (Beckmann et al 1990). TFE3 was initially mapped to Xp11.22 using a somatic cell hybrid panel and linkage analysis (Lafreniere et al 1991) and its physical localisation recently established by Schindelhauer et al (1996) and Boycott et al (1996). Complete loss of normal TFE3 transcripts caused by a translocation breakpoint in the promoter region of this gene that gives
rise to a novel fusion protein has recently been shown to result in papillary renal cell carcinoma (Sidhar et al 1996).

1.8.4.2 Recent additions to the proximal Xp gene map

The ZNF21, ZNF41 and ZNF81 genes encode Kruppel-type C2H2 zinc-finger proteins (ZFPs), which are thought to regulate gene expression during development by binding to DNA/RNA in a sequence-specific manner (Schuh et al 1986). Zinc fingers consist of tandemly repeated units of 28-30 amino acids which fold independently into DNA-binding domains in which the cysteine and histidine residues are specifically arranged so that they pair tetrahedrally with a single zinc ion (El-baradi and Pieler 1991). All three genes have been mapped to Xp (Huebner et al 1991; Franze et al 1991; Marino et al 1993). ZNF21 was one of 30 ZFPs mapped to specific chromosomes by hybrid analysis, when it was revealed that ZFPs tend to cluster within the genome (Huebner et al 1991) possibly reflecting local gene duplication events. STS-content mapping of YACs in Xp11.23 has since shown that ZNF21, ZNF41 and ZNF81 are similarly clustered (Knight et al 1994). ZNF157, a novel Kruppel-type zinc finger gene was isolated during the course of YAC end-rescue and has been physically mapped distal to the ZNF21, 41, 81 cluster between UBE1 and TIMP-1 by STS-content mapping of a YAC contig in which all three genes were present (Derry et al 1995, Carrel et al 1996; Figure 1.9).

In an effort to search for genes involved in the many retinal degenerations mapping to Xp11.23 (section 1.9) and to verify that the OATLI sequences do not represent functional genes, Geraghty et al (1993) used a 480kb YAC containing the OATLI locus to screen a retinal cDNA library. In addition to 2 OAT cDNAs corresponding to the structural gene on chromosome 10, four evolutionarily conserved single copy transcripts were isolated, MG21, MG44, MG61 and MG81, which were sublocalised and ordered within the OATLI YAC using YAC fragmentation vectors (Pavan et al 1991). The order proposed was Xpter–OATLI–MG61–MG81–MG21–MG44–Xcen, with the novel cDNAs spanning 350kb.

PCTAIRE-1 (PCTK1) is a member of a recently identified sub-family of cdc-2 related protein kinases of ubiquitous expression which is thought to be involved in cell cycle regulation (Meyerson et al 1992). Human clones containing PCTK1 sequences were isolated from a cosmid library with murine PCTAIRE cDNA probes and used to map the gene to Xp11 by FISH (Okuda et al 1994). PCTK1 has recently been physically linked to UBE1 in Xp11.3 by colocalisation to a 420kb YAC which may be of significance at the level of gene regulation if PCTK1 is a target for ubiquitinisation catalysed by UBE1 (Knight et al 1995).

Figure 1.9 also shows several novel ESTs mapped to proximal Xp by many groups involved in X inactivation studies or positional cloning projects. Those pertaining to this thesis will be discussed in the relevant sections.

In addition to the aforementioned tumour-related breakpoints that disrupt the TFE3 gene and OATLI pseudogene cluster, several other genes in proximal Xp have been shown through positional cloning techniques to underly specific inherited disorders. Genes for both
McLeod syndrome (Xp21.1), a late onset myopathy with neurological defects, and Dent’s disease (Xp11.22), a renal tubular disorder have been identified with the aid of deletion patients and shown to encode respectively, a membrane transport protein (XK) and voltage-gated chloride channel (CLCN5) (Ho et al 1994; Fisher et al 1994). Long-range restriction analysis of a YAC spanning a translocation breakpoint associated with Aarskog-Scott syndrome (Faciogenital dysplasia) in Xp11.21, coupled with a search for evolutionarily conserved sequences, recently led to the isolation of FGDI, a gene encoding a widely-expressed protein with guanine-nucleotide exchange activity that was shown to be nonfunctional in Aarskog families (Pasteris et al 1994). In the absence of chromosomal rearrangements, the gene for Wiskott-Aldrich syndrome (WASP), characterised by eczema, thrombocytopenia and immunodeficiency, was positionally cloned by classical genetic linkage studies followed by cDNA selection from a YAC contig spanning the critical interval (Derry et al 1994).

Inherited diseases genetically mapped to this interval but for which the corresponding genes are not yet cloned include several forms of X-linked mental retardation (XLMR; Lubs et al 1996), incontinentia pigmenti (IP; Reed et al 1994) and X-linked infantile spinal muscular atrophy (Kobayashi et al 1996).

1.9 Genetic eye diseases mapping to proximal Xp

The proximal short arm of the human X chromosome is home to a cluster of genes implicated in various inherited disorders of the eye (Figure 1.10). Several of these disorders share key symptoms with differing degrees of severity and overlapping map locations, and the growing examples of allelic heterogeneity (section 1.7; see below) suggest that different mutations in the same gene may have diverse phenotypic effects.

1.9.1 Norrie disease

Norrie disease (NDP) is an X-linked neurologic disorder characterised by bilateral retinal dysplasia leading to opacities behind the lens and consequent congenital blindness. Neuropathologic examinations have revealed poorly differentiated, malformed retinas suggesting developmental abnormalities (Warburg 1966). Approximately one-third of NDP patients develop progressive sensorineural hearing loss, while over half display mental retardation and/or psychotic behaviour.

The NDP gene was assigned to Xp11.3 by linkage studies and analysis of submicroscopic deletions in affected males (de la Chapelle et al 1985). These patients displayed complex phenotypes, presumably representing a ‘contiguous gene syndrome’ whose phenotype reflects the deletion or disruption of more than one gene. Although both deleted in these patients, the MAOA and MAOB genes were excluded as NDP candidates by the demonstration of intact MAO genes and normal levels of enzyme activity in typical NDP patients (Sims et al 1989). A single recombination event between a microsatellite marker within MAOB and NDP established MAOB as the distal boundary of the NDP region (Sims et al 1992). The delineation
Figure 1.10 Genetic localisations of inherited eye disorders on proximal Xp (not to scale)
of the proximal boundary was achieved through characterisation of a YAC clone which
contains DXS7, MAOA and MAOB. The retention of the proximal end of the YAC in one of
the deletion patients indicated that all or part of the NDP locus was contained within a 160kb
fragment of this YAC (Sims et al 1992).

The Norrie disease gene was subsequently isolated by two groups using a positional
retinal cDNA libraries using as a probe the 160kb YAC subfragment containing the disease
screening of retina and brain cDNA libraries using a cosmid probe from a contig extending
250kb proximal to the MAOB gene. In addition to mapping within the region of overlap in
several microdeletion patients, this cosmid detected aberrant genomic DNA fragments on
Southern blot analysis of nonsyndromic NDP males, thus was considered likely to span at least
part of the NDP locus.

Both groups identified an evolutionarily highly conserved cDNA whose expression was
limited to retina, brain and choroid. Furthermore, the cDNA sequence was absent or disrupted
in all Norrie-associated microdeletions. Protein sequence comparisons revealed a cysteine-rich
C-terminal domain (CT domain) in the NDP gene (termed norrin) with homology to diverse
extracellular proteins (Meindl et al 1992). An N-terminus hydrophobic region contains
features of a signal sequence, suggesting that norrin may be a secreted protein involved in the
regulation of cell differentiation and proliferation. Computer modelling of norrin predicts a
3D structure reminiscent of a family of cysteine knot growth factors, in particular
transforming growth factor β (Meitinger et al 1993). This supports the idea that norrin may
function as a growth factor involved in development and maintenance of neurons in the retina
and CNS.

A full spectrum of mutations (stop codons, missense mutations, and insertions, as well as
deletions) has been identified in Norrie patients (Meindl et al 1992; Berger et al 1992a), in
keeping with the high proportion of new mutations expected for an X-linked disorder with
reduced male reproductive fitness (section 1.8.2). The absence of simple genotype/phenotype
correlations (Berger et al 1992a) proves that ocular symptoms, mental retardation and deafness
are all pleiotropic effects of mutations in the Norrie gene. However, it has recently been
reported that missense mutations in non-conserved residues of the C-terminal domain lead to a
more benign Norrie phenotype, presumably because they produce a partially functional
protein (Meindl et al 1995). This hypothesis is supported by the description of similar norrin
mutations causing X-linked exudative retinopathy (XLFEVR), the symptoms of which overlap
that of NDP but are restricted to the eye and are less severe (Chen et al 1993). It is possible
that a temporal difference in pathogenesis may explain the clinical differences between these
allelic disorders; the events causing NDP may occur earlier in fetal life, so that by birth the
disease process is more advanced.

In order to elucidate the cellular and molecular processes involved in Norrie disease, a
mouse model has recently been generated by gene targetting (Berger et al 1996). 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. These ocular findings are consistent with observations in NDP patients, thus this mutant mouse line should prove a faithful model for unravelling the early pathogenic events in this neurological disorder.

1.9.2 X-linked congenital stationary night blindness

Night blindness is a symptom of several hereditary retinopathies. Congenital stationary night blindness (CSNB) is characterised by nonprogressive night blindness (hence the term 'stationary') without concomitant retinal dystrophy, and is associated with reduced visual acuity (Pearce et al 1990). Although ophthalmological examination may reveal a normal fundus, the electroretinogram (ERG) of affected males shows an absent or severely reduced scotopic b-wave indicative of a defect in neurotransmission from rods to bipolar cells. Autosomal dominant, autosomal recessive and X-linked forms of the disease can only be distinguished by genetic evaluation of the pedigrees, although X-linked CSNB (CSNBX) is frequently associated with myopia.

Two clinical subtypes of CSNBX have been observed; the complete form (CSNB1) lacks rod function by ERG and dark adaptometry and myopia is usually severe, whilst the incomplete type (CSNB2) shows some rod function on scotopic testing with moderate myopia to hyperopia (Miyake et al 1986). Cone function is always affected in CSNB2, but infrequently in CSNB1. Both types have been observed in single X-linked pedigrees (Aldred et al 1992; Bergen et al 1995), suggesting that CSNBX is a single entity manifesting a wide variation in clinical expression.

Genetic analyses over the last decade have established heterogeneity for CSNBX, implicating at least two distinct genes on the proximal short arm of the X chromosome. Linkage studies initially located CSNB1 close to DXS7 in Xp11.3. Analysis of informative crossovers subsequently placed CSNBX proximal to MAOA and distal to TIMP-1 (Bech-Hansen et al 1992; Aldred et al 1992), in contrast to later studies in which a key recombinant in one large family positioned CSNB1 proximal to TIMP-1 (Bech-Hansen and Pearce 1993). Tight linkage has been found between CSNB2 and markers in Xp11.3-11.22 (Musarella et al 1992), and it has been placed proximal to DXS7 by a single recombination event (Bergen et al 1994), supporting the suggestion that CSNB1 and CSNB2 may be allelic disorders.

A novel CSNBX locus has recently been genetically mapped distal to DXS7 (disease interval Xp21.1-Xp11.3) in one family, however, providing evidence that CSNBX is genetically heterogeneous (Bergen et al 1995). A recent report from Berger et al (1995) positions CSNBX in a large Dutch pedigree to the interval OTC-DXS1003 (Xp11.4-Xp11.23) which overlaps that of the two previous localisations. In light of the allelism demonstrated between autosomal forms of CSNB and the progressive degenerative retinal disease retinitis pigmentosa (RP) (section 1.9.5), which shares night blindness as one of its early symptoms, it is postulated that CSNBX may in fact be allelic to XLRP which similarly displays genetic heterogeneity with map locations in Xp11.3-p11.22 and Xp21.1 (section 1.9.5.1).
1.9.3 Aland Island Eye Disease (AIED)

This non-progressive condition is characterised by reduced visual acuity, colour vision abnormalities and myopia. Initially, AIED was thought to be a variant of ocular albinism (OAl) on the basis of the the fundal depigmentation in the original family described, and was given the locus symbol OA2, but this has since been disproved. Linkage of AIED to the pericentromeric region of the X chromosome was demonstrated by Atilato et al (1991) in the original AIED family. Subsequent refinement in a second AIED family placed the disease locus between DXS7 (Xp11.3) and DXSY1 (Xq21.3) with tight linkage to \textit{TIMP-1} and DXS255 (Schwartz and Rosenberg 1991). Linkage analysis of a third family suggested a location between \textit{DMD} and DXS255 (Glass et al 1993). Assuming homogeneity of the three pedigrees, the combined information supported a localisation of the AIED gene between DXS7 and DXS255 (Glass et al 1993). These findings contradicted a previous assignment of the AIED gene to Xp21.3-p21.2 based on a male patient with a contiguous gene deletion and complex phenotype of DMD, glycerol kinase deficiency, congenital adrenal hypoplasia and AIED. Further characterisation of the deletion in this patient showed that it spanned a region more proximal than those of several OAl patients (Xp22), supporting the argument that OAl and AIED are distinct (Pillars et al 1990). The identification of dystrophin isoforms in retinal tissue has since raised the possibility that the impaired retinal neurotransmission in this patient was a direct consequence of dystrophin deficiency as observed in many DMD/BMD patients, and the phenotype was termed Oregon eye disease to distinguish it from AIED (Fillers et al 1993).

The similarity in ERG findings suggests that AIED and CSNB2 might be a single entity (Schwartz and Rosenberg 1991; Glass et al 1993). Since the genes responsible for AIED and CSNB1 have been assigned to the DXS7-DXS255 interval, possible allelism between AIED/CSNB2 and CSNB1 is also addressed and remains open.

1.9.4 X-linked progressive cone dystrophy

X-linked progressive cone dystrophy (XLPCD; COD1) is a rare hereditary eye disorder characterised by progressive loss of visual acuity, color vision defects, myopia, prominent macular changes and photophobia (Keunen et al 1990). The cone ERG is absent or severely attenuated in affected males, and the rod ERG may also be abnormal. The severity of the fundus findings correlates with the degree of visual dysfunction and suggests an age-related progression of the condition (Hong et al 1994). The retinas of some affected males may present with a tapetal-like sheen (Jacobsen et al 1989). The clinical picture of heterozygous females ranges from asymptomatic to a widespread spectrum of cone-mediated dysfunction; colour vision testing has been reported to allow detection of 87% of obligate carriers (Keunen et al 1990).

At present, only a small number of XLPCD families have been clinically described and only four have been studied for linkage (Meire et al 1994; Hong et al 1994), localising the
gene to Xp21.1-p11. The COD1 locus has recently been refined to Xp11.4-p11.3 (Seymour et al 1996). Two obligate recombinants as defined by haplotype analysis localise the gene to a ~1Mb interval between DXS556 and DXS993, thus excluding COD1 as an allelic variant of RP3 or RP2 (section 1.9.5.1).

1.9.5 X-linked retinitis pigmentosa

Retinitis pigmentosa (RP) is a group of progressive retinal degenerations primarily affecting the (rod) photoreceptor cells and pigment epithelial layer (Pagon 1988). With a prevalence of about 1 in 4000 of the general population, it is one of the more common causes of blindness in early and middle life. The prominent symptoms are progressive development of night blindness and constriction of the peripheral visual fields ('tunnel vision'), with eventual loss of central vision in most cases (Heckenlively 1988). Fundus examination in advanced disease shows attenuated retinal vessels and bone-spicule pigmentary deposits commencing in the periphery due to invasion of the retina by cells of the retinal pigment epithelium (see Figure 1.11). Progressive deterioration of retinal function is revealed by scotopic ERG, which shows reduced amplitude and delayed response times and soon becomes unrecordable (Heckenlively 1988).

RP shows exceptional genetic heterogeneity as a consequence of the complex protein networks within the retina: families illustrating autosomal dominant (ADRP), autosomal recessive (ARRP), X-linked (XLRP), digenic inheritance and syndromal forms are well documented, with ~20 different RP loci mapped to date (Dryja and Li 1995; Berson 1996; Sullivan and Daiger 1996). Up to 50% of RP cases are isolated (simplex), with no family history of the disease. Most of these are ARRP, but some are likely to be ADRP or XLRP (preponderance of males), while others may represent new mutations or phenocopies (Jay 1982). The prevalence of the various genetic forms varies considerably from country to country (Jay 1982).

The X-linked form of the disease (XLRP) presents the most severe phenotype. XLRP males become symptomatic in the first decade of life with blindness occurring by the third or fourth decade, and female heterozygous carriers often develop phenotypic signs of the disease in middle to late life (Bird 1975). In such females random inactivation of one or other of the two chromosomes in each cell during embryogenesis (Lyon 1988) causes a wide variation in phenotype depending on the proportion of retinal cells with an active disease-bearing chromosome (Bird 1975).

1.9.5.1 Genetic heterogeneity in XLRP

Two lines of evidence firmly established the genetic heterogeneity of XLRP; genetic linkage studies and the analysis of X chromosome deletions. There are to date potentially 4 XLRP loci; RP2, RP3, RP6 and RP15. The evidence for each of these is discussed below.

Close genetic linkage between XLRP and DXS7 in Xp11.3 was first demonstrated in 5 British families by Bhattacharya et al (1984). However, ensuing reports were in conflict over
Figure 1.11 Fundus photograph of a retina in the late stages of X-linked retinitis pigmentosa: the 'bone spicule' pattern of pigmentary deposit in the periphery is a result of rod photoreceptor degeneration which leads to tunnel vision and eventual blindness.
the location of XLRP relative to DXS7. Friedrich et al (1985) identified two recombinants in a
large Danish family that positioned XLRP proximal to DXS7, while others favoured a more
distal location between DXS7 and DXS84, tightly linked to OTC in band Xp21 (Denton et al
1988; Musarella et al 1988). This was consistent with a male patient (BB; Francce et al 1985)
who suffered from DMD, McLeod syndrome, CGD and RP, and bore a cytologically visible
deletion in this band. Two further deletion patients, SB with CGD, McLeod syndrome and
XLRP, and OM, with only McLeod syndrome and CGD, further localised the XLRP gene
proximal to CYBB. The result of a large multipoint analysis of 20 XLRP families essentially
excluded the XLRP locus from the Xp21 region and favoured a location proximal to DXS7
and distal to DXS14 in Xp11 (Wright et al 1987).

The situation was clarified by heterogeneity analysis of linkage data pooled from 62
XLRP pedigrees from many countries (incorporating most of the previously published
families), whereby Ott and colleagues (1990) provided overwhelming evidence in favour of 2
loci versus a single XLRP locus with odds of 6.4 x 10^8. The majority of families (75%) were
associated with a locus 1cM distal to OTC (RP3) in Xp21.1, consistent with the earlier
localisations for XLRP from both deletion and linkage analysis of large pedigrees, while in
25% families the most likely location of the gene was 2cM proximal to DXS14 in Xp11.22.
The confidence limits for the latter locus were wide, extending 3cM proximal to DXS7 to 1cM
distal to the centromeric marker DXZ1, so that it could correspond to the Xp11 locus
proposed by earlier linkage studies (RP2). More recent heterogeneity analysis of 37 XLRP
families has placed the RP3 gene 0.4cM distal to OTC (70% of families) and the gene for RP2
6.5cM proximal to DXS7 (30% of families) (Teague et al 1994), and these have become the
‘working’ locations for the two loci.

Musarella et al (1988) proposed another XLRP locus (RP6) located between DXS28 and
the DMD locus in Xp21.3-p21.2 based on recombination data from one two-generation
family. Ott et al (1990) found some statistical evidence for a third locus located in the same
region, but patients deleted for this region have not been reported to have RP (Wright 1990),
which argues against an XLRP gene in this location.

A fourth locus for a clinical variant of XLRP was recently reported in distal Xp (McGuire
et al 1995), following linkage to DXS989 (Xp22.11) in a family originally thought to be
segregating autosomal dominant cone-rod dystrophy owing to the early cone involvement and
the presence of severely affected females. This locus may correspond to the putative RP6 locus
which is not genetically well-defined (Musarella et al 1988). The authors argued that the
progressive cone-rod degeneration seen in this family, although less common than the typical
rod-cone subtype, is a justifiable clinical variant of RP (it has also been reported in an RP2
family; Jacobsen et al 1992), and proposed the locus symbol RP15. However, the ‘dominant’
nature of the disease has been questioned on the premise that manifesting females are likely to
arise owing to the lottery of X-inactivation. In general, unless the gene in question escapes X-
inactivation, the terms ‘dominant’ and ‘recessive’ would seem inapplicable. This paper
demonstrates the inter-familial phenotypic variability evident in X-linked retinal
degenerations, strengthens the case that many may be allelic variants, and has raised important
questions concerning the wider issues of nomenclature for inherited eye diseases (Inglehearn and Hardcastle 1996).

In support of further heterogeneity for XLRP, Aldred et al (1994) described 3 XLRP families which appear unlinked to either the RP2, RP3 or RP6 loci as shown by haplotype data and multipoint linkage analysis using markers spanning Xp22.2-Xq21.3. It was tentatively suggested that there may thus be further XLRP loci on distal Xp or Xq, although the authors recommend caution in interpretation as both families had affected males in only one generation.

In summary, although the genetic heterogeneity of XLRP is well-established, the story is far from complete. However, the overall genetic data is supportive of two predominant loci, RP3 in Xp21.1 and RP2 in Xp11.3-p11.2. As their close proximity often precludes straightforward assignment of families to one or other locus by genetic mapping, researchers have sought clinical differences between RP2 and RP3, primarily driven by the need to provide genetic counselling and prenatal diagnosis for ‘at risk’ carrier females, but also to aid gene mapping efforts.

1.9.5.2 Phenotype/genotype correlations in XLRP

Inter-familial variation in phenotype can be due to a variety of causes, including polygenic or environmental influences on a single disease locus, due to different alleles at the same genetic locus in different families, or due to different genetic loci. There are no unequivocal phenotypic differences between RP2 and RP3 that allow them to be distinguished on clinical grounds alone. However, clinical studies in families that can be distinguished genetically have led to some tentative correlations between genotype and phenotype. Features such as tapetal reflex (Denton et al 1988; Bergen et al 1995), myopia (Wright et al 1991), onset symptom (Kaplan et al 1992) and level of cone dysfunction (Wright et al 1991; Jacobsen et al 1992) have all been variably associated with either RP2 or RP3, but there is no conclusive evidence to suggest that any are pathognomonic for specific XLRP loci.

Once the precise number and location of the genes causing XLRP are established it may be possible to determine whether or not clinical manifestations of each disease provides some basis for predicting the genotype. The confounding factor of allelic differences within each genetic type make it more likely that the issue will only be clarified once all the XLRP genes are isolated.

1.9.5.3 Recent progress in positional cloning of the RP3 and RP2 genes

A comprehensive review of all published linkage data (Aldred et al 1994a), incorporating recent data of Bergen et al (1995) suggests that 36% of European families are RP2, almost twice and thrice that found in Australia (22%) and the USA (14%). Overall, 30% of families are RP2, consistent with the heterogeneity analyses of Ott et al (1990) and Teague et al (1994). Within Europe, RP2 appears particularly prevalent in the UK. Although this may reflect the
large number of British families reported, this may be a consequence of a founder effect. The genetic origin and drift of distinct RP mutations within a country or region may considerably influence the a priori probability of encountering a particular XLRP subtype in DNA diagnosis, with important implications for genetic counselling. The following sections summarise recent mapping progress on the two XLRP loci for which genetic evidence is undeniable, RP3 and RP2.

1.9.5.3.1 RP3 - Physical mapping studies and isolation of candidate genes

Fine mapping of deletion patients has provided the ultimate route toward identification of the gene causing RP3. The RP3 gene was believed to lie in the proximal portion of the BB deletion (section 1.9.5.1) because of its coincidence with the critical region delineated by linkage analysis. The two genetic markers, DXS1110 and OTC that flank the RP3 locus span a physical distance of ~520kb (Nelson et al 1995); the proximal end of the BB deletion is 40kb centromeric to DXS1110. Several groups have been searching for transcribed sequences in this region surrounding the proximal breakpoint for several years.

The BB deletion junction was initially cloned by Musarella et al (1991) using a DMD cDNA as a starting point from which to isolate genomic clones and create a long-range physical map of the proximal part of the deletion. PFGE studies restricted the RP3 locus to a 205kb SfiI fragment in which a CpG island was identified. Segments of genomic DNA adjacent to the CpG island hybridised to discrete bands in digested DNA of several species, indicating evolutionary conservation, and were subsequently used to isolate a cDNA of ubiquitous expression from retinal cDNA libraries (Roux et al 1994). Given its high similarity to the murine tctex-1 gene (thought to be involved in spermatogenesis) the function of the novel human gene (TCTEIL) was speculative, but its location in Xp21 and complete deletion in patient BB prompted investigation as a candidate for RP3. However, no disease-associated changes in the coding portion of the gene were found in 20 RP3 patients (Roux et al 1994).

At this point re-evaluation of the contiguous deletions associated with syndromes in Xp21.1 suggested that BB may have misdirected the search for RP3. A deletion in patient NF who suffered from DMD and CGD but did not have RP symptoms, extended proximal to that of BB, essentially excluding the proximal part of the BB deletion from containing RP3 (Brown et al 1996). Gene cloning efforts were thus directed to an extended interval between the TCTEIL locus and the SB proximal deletion breakpoint 50kb distal to OTC (Meindl et al 1995).

In the autumn of 1995, two groups simultaneously reported the cloning of a candidate RP3 gene, SRRP/EYX-1, mapping within this interval (Meindl et al 1995; Dry et al 1995). To identify transcribed sequences, the strategy adopted by Meindl et al (1995) was to screen retinal cDNA libraries with subfragments of a YAC containing both OTC and CYBB, paralleled by genomic sequencing of putative CpG islands and hybridisation of cosmid digests with a splice-site consensus sequence. Dry et al (1995) used YACs covering the CYBB-OTC interval to isolate cosmids upon which to perform exon amplification (section 1.4.5). The SRRP/EYX-1 transcript showed highest expression in retina and heart, encodes a putative cell surface
protein and contains sushi repeats similar to selectin cell adhesion proteins. SRPX/ETX-1 was found to lie 250kb proximal to the BB deletion junction (Meindl et al 1995) but was considered a candidate gene for RP3 on the grounds that large chromosomal rearrangements may affect the expression of nearby genes (position effect; Bedell et al 1996) or be associated with secondary deletions remote from the major site. The gene was completely deleted in XLRP patient SB, and a microdeletion was detected in XLRP patient MO encompassing the promoter region and exon 1 of SRPX (Meindl et al 1995). However, no further functionally significant mutations were detected by SSCP screening of all exons in a panel of unrelated XLRP patients nor by full-length RT-PCR sequencing in two RP3 families (Meindl et al 1995). The role of this highly conserved retinally expressed gene in the pathogenesis of RP therefore remains to be determined.

A timely collaborative effort between the two groups in searching for further transcripts mapping within the MO deletion has since resulted in the cloning of a gene which is mutated in a proportion of RP3 patients (Meindl et al 1996). The gene is evolutionarily conserved and is thought to be a 'housekeeping gene' on the basis of its ubiquitous expression and 5' CpG island (Dry et al 1996). Interestingly, this gene shows very low levels of expression in the retina and RPE, which explains why it could not be isolated through direct screening of retinal cDNA libraries (Meindl et al 1996). The gene was eventually identified by systematically subcloning and sequencing two cosmids covering the proximal part of the MO deletion, and detection of gene sequences by computational analysis (section 1.4.6). Mutations were found in highly conserved residues which segregated with disease in 7 XLRP families from a pool of 74 unrelated patients, providing evidence that this gene underlies at least some cases of RP3. The predicted protein contains a tandem repeat structure similar to RCC-1 (regulator of chromosome condensation) which regulates the GTPase Ran (Ras-related nuclear protein), known to play a role in cell cycle progression, membrane transport and RNA processing. The high rates of membrane turnover in the retina and RPE (see section 1.7.1) have led to speculation that this novel gene, termed RPGR (retinitis pigmentosa GTPase regulator) acts to regulate this process (Meindl et al 1996). It is interesting that in another X-linked eye disorder, choroideremia, the defective gene plays a role in the geranylgeranylation of different Rab proteins, another family of Ras-related GTPases (Seabra et al 1993; van Bokhoven et al 1994). In view of the clinical similarities between choroideremia and RP3 (Bird 1975), it is speculated that RPGR may be a guanine-nucleotide exchange factor for retina-specific Rab proteins.

Shortly after the publication of RPGR, Roepman et al (1996) cloned the same gene via a novel method called 'YAC representation hybridisation' and found disease-associated mutations in 5 out of 28 XLRP patients. It is surprising that so few of the XLRP patients screened (18%) revealed disease-associated mutations when RP3 accounts for ~70% of all XLRP cases (section 1.9.5.1). Perhaps mutations may lie in unidentified parts of the gene or alternative transcripts, common mutations were not identified by SSCP analysis and/or there is heterogeneity within RP3, with the major locus still to be identified. Fujita et al (1996) have recently reported a recombination event in a large RP3 family which localises the causative mutation proximal to the BB deletion. If RPGR is the major RP3 gene, questions remain as to
the presence of XLRP in patient BB. It is feasible that his large deletion may include another RP3 locus, or an RPGR regulatory element situated some distance from the gene. Without a family history, it is also difficult to exclude the possibility that he may have had an autosomal form of RP.

Studies are currently underway to determine the precise location and function of RPGR in the retina and elucidate the mechanism by which defects within it lead to retinal degeneration. This will enhance our understanding of normal retinal function and may provide clues as to the causes of other inherited retinopathies, in particular RP2. The discovery of RPGR will undoubtedly benefit women in families segregating the gene who request carrier testing, and will diagnose XLRP in a proportion of sporadic patients enabling more appropriate counselling and prognosis. Recent mutation screening of 10 CSNBX pedigrees has disclosed an RPGR mutation segregating in one family which is absent from 170 control X chromosomes (Hermann et al 1996). This provides evidence for the postulated allelism between XLRP and CSNBX (section 1.9.2) and may assist functional analysis of the gene product.

1.9.5.3.2 RP2 - further genetic mapping studies to confirm and refine locus location

The lack of associated cytogenetic abnormalities for RP2 has hampered localisation of this gene and its precise location by linkage analysis has been less well-defined. Ott et al (1990) localised RP2 to a broad region extending from DXS7 in Xp11.3 to the centromere. Multipoint linkage analysis by Wright et al (1991) on the large British kindred of Bhattacharya et al (1984) gave a maximum likelihood location of RP2 close to DXS255 in Xp11.22 and TIMP-1 in Xp11.23 in an area extending from 2cM proximal to DXS7 to 1cM distal to DXS14. This is supported by the study of Teague et al (1994), where RP2 had a maximum likelihood location 6cM proximal to DXS7. The recent heterogeneity analysis of Bergen et al (1995) produced a most likely location for RP2 at DXS255, with a confidence interval extending from DXS7 to DXS14.

In summary, all multipoint linkage studies have given differing most likely locations for RP2, albeit with varying confidence intervals which overlap to some degree. This may reflect detection of linkage by DXS7 to both RP2 and RP3 loci (only ~10-15cM separates them). More importantly, the relative rarity of the RP2-type family (30% of most XLRP populations) reduces the number of informative recombination events necessary for genetic localisation of the gene. Friedrich et al (1992) reported a large RP2 kindred in which they observed an apparent expansion of the genetic map between the centromere and proximal Xp markers. This suggested that variability in recombination due to differences in the amount of heterochromatin segregating in various pedigrees could have contributed to the uncertainty regarding the location of RP2.

At the outset of this study, overall data indicated a location for the RP2 locus between DXS7 and DXS255 (Wright et al 1991; Friedrich et al 1992), a genetic interval of 13-18cM (Mahtani et al 1991). Thus there is still scope for refining the localisation of RP2 further by genetic mapping, but given the problems of classifying families and obtaining large pedigrees
with multiply informative recombinants, other approaches are being pursued, in particular physical mapping strategies to order new markers within the RP2 region and investigation of candidate genes mapping to the critical interval (see chapter 6 and 7). Families segregating CSNBX mapping to Xp11 may also be useful for refining the localisation of RP2 if these represent allelic disorders.

Aldred et al (1994b) have described a family in which XLRP cosegregates with mental retardation and in which appears to be RP2 by linkage analysis (maximum likelihood location 0.5cM distal to TIMP-1 with Z=1.61). This may represent a new genetic syndrome due to a locus that is fortuitously located in the same region, or, more interestingly, raises the possibility of a contiguous-gene-deletion syndrome involving the RP2 gene in this family, which could significantly reduce the region of search for the RP2 gene.

1.10 Aims of this thesis

The primary aim of this thesis was to further progress the positional cloning of the gene causing one form of XLRP (RP2) on proximal Xp by using genetic and physical mapping strategies to define the disease gene critical region and provide reagents within this interval as a starting point for transcript analysis.

The first priority was to generate novel polymorphic markers for the marker-poor region in Xp21.1-p11.22 by FISH mapping of X-specific cosmids known to contain CA-dinucleotide microsatellites. The microsatellites would then be isolated from cosmids mapping to this interval and their location further refined by genetic mapping in relation to reference markers in a panel of XLRP families, thereby enhancing the genetic map of proximal Xp. Further genetic characterisation of the 14 XLRP families was undertaken with an extensive set of markers spanning Xp21.1-p11.22 in order to distinguish between those families segregating RP2 from RP3 (or other XLRP loci) and further delineate the RP2 critical region. A high resolution physical map of the RP2 gene critical interval was subsequently generated in the form of a YAC contig, to enable integration of the genetic and physical maps and assemble a transcriptional map of the target region. This contig would also provide a template from which to isolate further polymorphic markers and candidate gene sequences to facilitate identification of the gene underlying RP2.
CHAPTER 2

Materials and Methods

NB: A list of reagents and solutions used in this study is described at the end of this chapter (section 2.9)

2.1 Fluorescence in situ hybridisation (FISH)

2.1.1 Preparation of metaphase spreads

Chromosome metaphase spreads were prepared from short-term cultures (48 hours) of phytohemagglutinin (PHA) stimulated whole blood in RPMI 1640 culture medium, supplemented with 15% Fetal Calf Serum. Addition of thymidine after 48 hours blocked cell division at S phase of the cell cycle. Removal of the block after 19 hours allowed cell division to progress in a synchronised fashion. Colcemid was added 4 hours later, for 15 minutes, prior to harvesting, to inhibit spindle formation thus arresting the cells at metaphase. The cells were then treated with a prewarmed solution of 0.075M KCl and fixed with a 3:1 methanol:acetic acid solution. Cell suspensions for metaphase spreads used in this project were donated by the Cytogenetics laboratory of the Department of Human Genetics, University of Newcastle upon Tyne and derived from cells with a normal 46,XX female karyotype. To prepare chromosome spreads from these suspensions, samples were centrifuged for 10 minutes at 1500 rpm, the pellet was resuspended in 5mls freshly prepared fixative and the centrifugation repeated. The pellet was then resuspended in 0.5-1ml of fresh fixative and 2-3 drops were placed onto each microscope slide. 3 drops of fixative were added to cover each slide and these were allowed to dry. The slides were then checked under a phase-contrast microscope to ensure proper spread of the chromosomes and an adequate mitotic index.

2.1.2 Labelling of DNA probes

For each hybridisation, the probe DNA was labelled with biotin by nick translation (Rigby et al 1977) in a 20μl reaction volume as follows: to 1μg DNA were added 2μl 10x salts (0.5M Tris-HCl pH 7.8, 0.05M MgCl₂, 0.01M β-mercaptoethanol, 50μg/ml BSA), 2.5μl cold dNTPs mixture (dATP, dCTP, dGTP, 10mM each, in 1:1:1 proportion: Bethesda Research Laboratories; BRL), 2.5μl (2.5 nmol) of biotin-16-dUTP (Boehringer Mannheim) and SDW to make up the volume to 15μl. 5μl of DNAase (40pg/μl)/DNA Polymerase I (0.4 units/μl) mixture (BRL) were added and the reaction was incubated at 15°C for 90 minutes.

2.1.3 Column purification and precipitation of the labelled probe

To the nick-translated probe were added 5μl stop buffer, 1μl 5% SLS (sodium lauryl sulphate) and 25μl TNE buffer. The probe was loaded onto a Sephadex G-50 column (Pharmacia) equilibrated in TNE buffer (prewashed three times, with 400μl TNE each time).
400µl TNE buffer were passed through the column and this first eluate was discarded. A further 400µl TNE buffer were added and to this second eluate containing the purified probe, 100µg sonicated salmon sperm DNA (Sigma) and 1/10 volume 3M ammonium acetate were added and the DNA was precipitated with 2.5 volumes absolute ethanol at -20°C for 1 hour. After 5 minutes centrifugation at RT, the supernatant was removed and the pellet air dried and resuspended in 25µl TE buffer for all cosmid probes or in 50µl TE for the centromeric repeat probe. Purified probes were stored frozen at -20°C.

2.1.4 Biotin incorporation test

Biotin incorporation was assessed by immunological detection via a streptavidin-alkaline complex which, with NBT (nitro-blue tetrazolium) and BCIP (5-bromo-4 chloro-3-indolyI phosphate) as substrates, results in the production of a blue precipitate. The buffers A, B, C and D required for the method are described in section 2.9.

Four dilutions of the DNA probe and of the appropriate controls were prepared starting with a 1:1 dilution of the probe (1µl DNA and 1µl 20xSSC; 20ng DNA/µl) followed by a further three sequential 1 in 10 dilutions (using 10xSSC) until a final probe concentration of 20pg/µl was obtained. 1µl of each of the 4 probe dilutions (i.e. 20ng, 2ng, 0.2ng and 20pg respectively) was spotted onto a gridded nitrocellulose filter (Schleicher and Schnell BA 85, Cellulosenitrate; E; 0.45mm, 88x88mm) and after air-drying the DNA, the filter was placed between two 3mm Whatmann papers and baked at 80°C for two hours.

The filter was then soaked in 20ml buffer A, in a square petri dish (Sterilin, 105x105mm) for 5 minutes followed by soaking in 20ml buffer B for 30 minutes. The filter was then incubated in a fresh 20ml buffer B containing streptavidin-alkaline phosphatase complex (diluted 1:1000; Boehringer Mannheim) for 30 minutes, followed by three successive 5 minute washes in buffer A.

The filter was then equilibrated for 5 minutes in buffer C before incubation in 20ml buffer C containing 88µl of 75mg/ml NBT (in 70% dimethylformamide; Boehringer Mannheim) and 68µl of 50mg/ml BCIP (in dimethylformamide; Boehringer Mannheim) for at least 3 hours in the dark.

The reaction was stopped by washing the filter in buffer D (stop buffer) for 5 minutes. The filter was wrapped in clingfilm, photocopied to obtain a record and stored at 4°C in the dark. Blue colour was seen to develop at first and second dilution (occasionally at third dilution) if biotin had been efficiently incorporated into the probe.

2.1.5 In Situ Hybridisation

Chromosome denaturation and probe denaturation were carried out separately.

2.1.5.1 Slide treatment

Slides were dipped in 100µg/ml RNAase (Sigma) in 2xSSC at 37°C for 1 hour, then dehydrated by passage through an ethanol series (70%, 90%, 100%) allowing 5 minutes in
each. After air-drying, the chromosomes were denatured by immersing the slides in prewarmed 0.6xSSC, 70% formamide (Analar) solution for 2 minutes, then immediately passed through a second ethanol series and air dried. The slides were now ready for probe application.

2.1.5.2 Hybridisation mix and probe application - repetitive probe DXZ1

A 100μl hybridisation cocktail was prepared containing 5μl of column-purified DXZ1 probe DNA (1ng/μl final concentration), 2xSSC, 500μg/ml E.Coli t-RNA, 10% dextran sulphate and 50% formamide. The probe was denatured by heating the mixture at 75°C for 10 minutes before placing on ice for 1 minute. 40μl was applied per slide (probe DNA concentration 40ng/slide), and hybridisation was carried out at 37°C for 16-18 hours.

2.1.5.3 Hybridisation mix and probe application - cosmid probes

Approximately 600ng of labelled probe (15μl) were combined with 60μg total human DNA (1:100 fold competition) in a 100μl reaction volume with 5xSSC final concentration. The mixture was denatured at 100°C for 10 minutes, placed on ice for 1 minute and preannealed at 60°C for 10 minutes. To the preannealed probe DNA 100μl of 20% dextran sulphate in 100% formamide and 10μl of 10mg/ml carrier E. coli t-RNA were added to make a 210μl hybridisation mix containing 3ng/μl probe DNA, 2.5xSSC, 500μg/ml E. coli t-RNA, 10% dextran sulphate and 50% formamide. 40μl of this hybridisation mix were placed on a clean coverslip and a treated slide was inverted onto it ensuring that no air bubbles were included, before sealing with rubber solution. When the seal was completely dry, the slide was transferred to a high humidity chamber, saturated with 20xSSC, and incubated at 37°C overnight.

During the course of the experiments, the protocol was modified (by Smaro Kamakari) such that satisfactory probe hybridisations could be attained using lower probe concentrations and a slightly lower salt concentration. The modified reagent mix included: 80ng labelled probe (2μl), 8μg total human DNA (8μl) (1:100 fold competition), 4μl 20xSSC, 20μl 20% dextran sulphate in 100% formamide, and 2μl 10mg/ml E. Coli tRNA, making a hybridisation mix of total volume 40μl (probe concentration 2ng/μl, 2xSSC), all of which was then applied to the slide.

If two probes were cohybridised (e.g. a cosmid and DXZ1), the above hybridisation cocktail was adjusted further by using only 40ng of cosmid DNA and adding 20ng DXZ1 DNA without the need for a separate hybridisation mixture for it as described previously.

2.1.6 Post hybridisation washes

Slides hybridised with biotinylated probes were passed through a series of washes: four 3 minute washes in prewarmed Solution 1 at 45°C, followed by four 3 minute washes in Solution 2 at 45°C, with a final brief immersion in Solution 3 at RT.
2.1.7 Signal detection

Slides were initially incubated for 5 minutes at RT with 40μl per slide of blocking buffer under coverslips. 40μl of the first antibody, FITC-conjugated avidin DCS (Vector Laboratories) was then applied to each slide, at final concentration 4μg/ml in blocking buffer, and incubation allowed to proceed under coverslips for 30 minutes at 37°C. Slides were then passed through three 2 minute washes in wash Solution 3 at 45°C. 40μl of the second antibody, biotinylated goat anti-avidin (Vector Laboratories) at final concentration 5μg/ml in blocking buffer were then applied and the slides incubated at 37°C for 30 minutes. The slides were then washed as before.

This whole procedure was then repeated, finishing with a final incubation in FITC-avidin DCS. After the final wash, the slides were mounted in anti-fading glycerol (Citifluor) solution containing 0.5μg/ml propidium iodide (Sigma). Slides were dry blotted, sealed with rubber solution and observed under the fluorescence microscope.

2.1.8 Signal visualisation

Signals on metaphase chromosomes were viewed under a Leitz Diaplan microscope, using filter 13 to visualise the FITC, with a x40 oil immersion objective. Signals were recorded by photography using a Wild MPS 46/52 camera (Leitz) and 400 ASA Kodak Gold film.

2.2 Preparation of DNA

2.2.1 Human genomic DNA

Blood samples were collected in 10ml EDTA tubes then either extracted at the time of their arrival in the laboratory, or stored at -80°C until required. Two methods were used to extract the DNA from peripheral blood lymphocytes.

2.2.1.1 Standard protocol

Individual blood samples (approximately 10ml) were thawed slowly on ice, and their volumes increased to 25ml with SDW. Erythrocyte lysis was achieved by incubation with lysis buffer (25ml) for 30 minutes on ice. Intact lymphocytes were pelleted at 3500g for 15 minutes, and the supernatant reduced to 8ml by aspiration. A further 42ml of lysis buffer was added to the supernatant on ice for 10 minutes to ensure complete lysis of the erythrocytes. Intact lymphocytes were pelleted at 3500g for 15 minutes and resuspended in 3ml of suspension buffer. Overnight rotary incubation with 10% SDS and proteinase K (0.4mg/ml) at 37°C was followed by a further 2 hour rotary incubation at 37°C with 0.6M sodium perchlorate. The suspension was extracted with an equal volume of phenol, phenol:chloroform (1:1, v/v) and twice with chloroform respectively at 5000g. The DNA in the aqueous layer was precipitated with 3M sodium acetate (0.1 volume) and 100% ethanol (2.5 volumes). The precipitated DNA which formed on tube inversion was removed with a sterile needle and
washed in 70% ethanol. The DNA was dissolved in an appropriate volume of TE buffer (100-300 µl depending on the size of the pellet) at 4°C over a period of 1-2 days.

2.2.1.2 Rapid protocol (Nucleon II DNA extraction kit)

Thawed blood samples were transferred to 50ml Falcon tubes and their volume increased to 50ml with reagent A. After mixing, the tube was centrifuged at 4000rpm for 5 minutes, the supernatant discarded and 2ml reagent B were added to the pellet. The mixture was then transferred to a 5ml tube and 500µl of 5M sodium perchlorate added, and mixed slowly for 15 minutes at room temperature, followed by 25 minutes at 65°C. Tubes were cooled on ice, then 2ml of chloroform and 300µl of silica suspension were added. Samples were mixed for 5 minutes and then centrifuged for 6 minutes at 1400g. 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 then washed with 70% ethanol, allowed to dry, and resuspended in a suitable amount of TE buffer (section 2.2.1.1).

Integrity and yield of genomic DNA was estimated by electrophoresis of a 2µl 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 10ml whole blood was between 200-300 µg.

2.2.2 Alkaline lysis miniprep of plasmid or cosmid DNA (Birnboim and Doly 1979)

This method is appropriate for preparing DNA from 5ml cultures of plasmid or cosmid containing bacteria. Colonies from a bacterial stab or glycerol stock were streaked out on 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 5ml of sterile LB broth containing antibiotic. Following incubation at 37°C overnight with agitation, 1.5ml of this culture was transferred to a 1.5ml tube and spun in a microfuge for 2 minutes at 13000rpm. The supernatant was removed and the pellet resuspended in 100µl of Solution I. 400µl of Solution II (freshly prepared) was added and mixed with the resuspended pellet (without vortexing) and placed on ice for 5 minutes. 300µl of Solution III was then added and mixed thoroughly. The sample was placed on ice for a minimum of 5 minutes followed by centrifugation for 2 minutes. The supernatant was removed and the DNA precipitated with 0.6 volumes of isopropanol. The pelleted DNA was washed with 70% ethanol, air dried and resuspended in 20µl TE buffer. 2µl of the preparation were electrophoresed on a 0.8% agarose gel (section 2.4.1) to assess quality and concentration.

2.2.3 Alkaline lysis midiprep of plasmid or cosmid DNA

The method below describes the isolation and purification of larger amounts of DNA.
50ml of LB broth supplemented with the appropriate antibiotic was inoculated with a single colony obtained as described in section 2.2.2 and incubated in a sterile 200ml flask at 37°C overnight with agitation. After preparing glycerol stocks from the culture (2ml of culture mixed with 2ml of 50% glycerol under sterile conditions and stored at -20°C or -70°C), the culture was transferred into 50ml Falcon tubes and centrifuged at 2500rpm for 15 minutes. The pelleted cells were resuspended in 4ml of Solution I and left at RT for a minimum of 5 minutes. 8ml of Solution II (freshly prepared) were added and mixed with the cells which were then placed on ice for 5 minutes. 4ml of Solution III was then added and the sample thoroughly mixed and placed on ice for a minimum of 15 minutes. The mixture was then centrifuged at 2500rpm for 10 minutes. The supernatant was strained through filter paper into fresh 50ml Falcon tubes and the DNA precipitated with 0.6 volumes ice cold isopropanol. After spinning at 2500rpm for 5 minutes, the supernatant was discarded. The DNA pellet was rinsed with 70% ethanol, resuspended in 0.4ml TE buffer and transferred to a 1.5ml tube. 15μl of 3M sodium acetate and 40μl of 5mg/ml heat-treated RNAase were added to the DNA suspension which was then incubated at 37°C for 30 minutes. The DNA was extracted with one volume of phenol, followed by one volume of phenol/chloroform (1:1) and one volume of chloroform. 35μl of 3M sodium acetate was added to the final aqueous phase, and the DNA precipitated with 2.5 volumes of ice-cold absolute ethanol. After spinning at 13000rpm for 2 minutes in a microfuge, the pelleted DNA was rinsed with 70% ethanol, air-dried and resuspended in 50μl of TE buffer. 2μl of the dissolved DNA was electrophoresed on a 0.8% agarose gel (section 2.4.1) to assess yield and quality.

2.2.4 Yeast DNA in agarose plugs

During normal preparation of DNA, shear forces will break the long, thin molecules down to sizes less than 500kb. To accurately size and restriction digest defined DNA molecules which may be considerably larger than this, such as YACs, DNA must be protected from shear forces during its preparation. This is achieved by encapsulating cells prior to lysis in agarose blocks (or 'plugs').

In the following method, YAC clone DNA was prepared according to Vollrath and Davies (1987) with slight modification. Yeast clones were streaked out onto AHC plates supplemented with ampicillin and incubated at 30°C until single large colonies were obtained (usually 36-48hrs). A single, verified YAC positive colony (see section 2.7.6) was picked into 10ml AHC media supplemented with ampicillin and allowed to grow at 30°C for 24 hrs; the culture was subsequently used to seed a 100ml culture, grown with agitation at 30°C for ~40 hours. The cells were pelleted by centrifugation for 10 minutes at 4K, and resuspended in 50ml of SCE solution. The cells were repelleted and resuspended in 0.5ml SCEM. Lyticase (Sigma) was added to a final concentration of 120μg/ml and incubation performed at 30°C for 1 hour to digest the yeast cell walls. Low melting point (LMP) agarose (1%) was prepared in 1M sorbitol and kept molten at 55°C. The yeast cell suspension was then mixed with an equal volume of LMP agarose and 200μl aliquots were dispensed into precooled plastic moulds and left to set on ice. The plugs were then placed in proteinase K/sarcosyl solution (1ml per plug) and
incubated at 50°C for 2 days. Following extensive rinsing with TE, proteinase K was inactivated by treatment with 0.04mg/ml PMSF (phenyl-methyl-sulphonyl-fluoride) in TE with incubation at 50°C for 30 minutes. The plugs were then rinsed three times in TE and stored at 4°C in TE containing 0.05mM EDTA to protect the DNA from degradation. For PCR analysis of YACs (section 2.7), plugs were rinsed three times in TE buffer or SDW over a period of 1-2 hours to eliminate the EDTA, were melted at 65°C and then diluted 1:1 with SDW and stored at 4°C.

2.3 Restriction enzyme digestion of DNA

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

In principle, 1 unit of restriction enzyme digests 1μg of λ DNA in 1 hour. Crude DNA preparations often require more enzyme and/or more time for complete digestion. Therefore restriction digests were performed using 2-4 units of enzyme per μg of DNA template, typically in a 50μl reaction including 5μl of 10x restriction buffer (Pharmacia or Biolabs) and made up to volume with SDW. 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 2 hours at the recommended temperature (usually 37°C) and the products of digestion resolved by agarose gel electrophoresis (section 2.4.1). For digestions 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 digest.

2.4 Size fractionation of DNA by gel electrophoresis

2.4.1 Conventional agarose gel electrophoresis

DNA preps and DNA fragments resulting from restriction enzyme digestion of genomic DNA (section 2.3), cloned DNA or non-radioactive PCR (section 2.7.2) were size-separated through agarose gels containing 0.5μg/ml ethidium bromide DNA stain. Gels were prepared by mixing an appropriate amount of agarose and 1xTAE buffer, melting in a microwave oven, adding stain and cooling to ~55°C 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 1-20kb and 3% for fragments of 70bp-1kb (2% NuSieve:1% standard agarose). DNA samples were mixed with 1/10 volume 10x loading buffer and electrophoresed at 80V for ~1 hour in 1xTAE buffer along with an appropriate size standard (λHindIII for fragments 1-20kb and φX174/HaeIII (Promega) for smaller products). The resolved fragments were visualised under 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 sequencing and microsatellite gels was purchased from BioRad. 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 8.3M urea solution (Sequagel, National Diagnostics) was mixed with diluent (8.3M urea) and buffer (10xTBE in 8.3M urea) to give a 6% gel solution in 8.3M urea and 1xTBE (gel base; total volume 150ml for a 40x60 cm gel rig, 100ml for a 20x40 gel rig).

A plug was first prepared to seal the bottom end of the plates prior to pouring the gel proper. 150μl of 25% (w/v) ammonium persulphate and 150μl of TEMED were added to 25ml of the above gel base and quickly mixed before pouring into the sealing tray (this sets very quickly). Once the plug was set (~15 minutes), 25% (w/v) ammonium persulphate (500μl for a 40x60 cm gel rig; 300μl for a 20x40 cm gel rig) and TEMED (56 μl for the large gel rig, 34μl 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. DNA samples (sequencing reactions or radioactively labelled microsatellite PCR products) were heat denatured at 95°C for 5 minutes and ice-chilled before loading 2-3μl onto a prewarmed (50-55°C) 6% denaturing polyacrylamide gel, first flushing out excess urea from the wells. Electrophoresis was carried out at a constant power of 90W for an appropriate length of time to achieve maximum resolution in the size required; 2-5 hours for microsatellites (depending on the size range of the allele system) and 2-7 hours for sequence (depending on the size of the template and length of sequence sought) after which gels were fixed in a 10% methanol/10% acetic acid solution for 5 minutes, covered in clingfilm and vacuum dried on 3MM Whatmann paper at 70°C for 1 hour and autoradiographed overnight at 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 Pulsed Field Gel Electrophoresis (PFGE)

The use of modified electrophoretic apparatus with alternating electric fields permits the separation of DNA fragments up to 10Mb in size and was thus used to separate YAC clones from their yeast chromosomal background in order that the size of the YAC be estimated. A full description of this technique is described in section 1.3.2.1.

The PFGE apparatus used here was a CHEF-DR II (Biorad) based on the clamped homogeneous electric fields methodology (Vollrath and Davies 1987).

2.4.3.1 Casting and loading the gel

100ml of 1% agarose (Molecular Biology Certified; Biorad) in 0.5x TBE buffer was poured into the casting stand (standard casting stand; Biorad) and allowed to set at RT. Sample
plugs were loaded and sealed in place by filling the remaining gaps in each well with 1% Low Melting Point Agarose (Biorad). The agarose was left to solidify for ~15 minutes.

2.4.3.2 Gel electrophoresis

The electrophoresis chamber was rinsed briefly with 2 litres of pre-cooled (4°C) distilled H₂O. Two litres of pre-cooled 0.5x TBE were poured into the chamber and allowed to circulate briefly through the recirculating pump. The gel was placed in the centre of the tank and secured according to the manufacturer's instructions. The following electrophoresis parameters were selected to separate DNA fragments in the size range 0.1-2Mb (specific operating instructions can be found in the Biorad CHEF-DRII manual): pulsewave 760, initial A time 60 seconds, final A time 90 seconds (this pulse mode is called ramping), run time 21 hours. The temperature during electrophoresis was kept at an optimal 14-15°C by placing the whole apparatus at 4°C. The buffer flow through the pump was adjusted to a level where gentle rippling could be observed on the buffer surface in the tank, but the gel was not disturbed.

2.4.3.3 Visualisation of the size-separated yeast DNA

The pulsed field gel was placed in 0.5μg/ml freshly prepared ethidium bromide solution and stained for 30 minutes. It was then destained in distilled H₂O for 30 minutes to remove background staining, visualised on a UV transilluminator and photographed.

2.5 Detection of DNA by hybridisation

In order to identify an individual DNA fragment in a complex mixture that has been resolved by gel electrophoresis, the technique first described by Southern (1975) using capillary transfer of DNA to a membrane has been used, which can then be hybridised with a DNA probe complementary to the sequence of interest. Although based on the same principles, specific experimental details vary with the type of membrane used. (NB: For transfer of DNA fragments >20kb, gels were first depurinated in 0.25M HCl solution for 30 minutes to break it into smaller fragments).

2.5.1 Conventional Southern blotting (uncharged membrane)

The agarose gel containing electrophoresed DNA was placed in denaturing solution for 20 minutes to denature the DNA followed by neutraliser for 40 minutes to reset the pH. After equilibrating the gel in 20xSSC, DNA fragments were transferred to a nylon membrane (Hybond N; Amersham) by capillary action as follows: the gel was inverted and placed on a wick of 3MM Whatmann paper assembled over a reservoir of 20xSSC. Membranes were cut to exactly the same size as the gel and placed on top. The whole was then covered with a further two pieces of 3MM Whatmann paper soaked in 20xSSC. Care was taken not to introduce air bubbles between the various layers. A stack of paper towels was then placed on top to a height of several centimetres and a uniform weight (500-750g) added. Clingfilm was used to cover
the surround of the assembly, up to the edges of the gel, so as to eliminate buffer evaporation which would slow the transfer process. Transfer was allowed to proceed overnight, after which the membrane was marked to indicate gel lane positions, rinsed in 2xSSC to remove residual agarose, and left to dry between two pieces of 3MM Whatmann paper at RT. DNA was fixed to the filter by UV cross-linking for 10 minutes on a UV transilluminator.

2.5.2 Alkali Southern blotting (positively charged membrane)

The procedure is the same as described in section 2.5.1 with some exceptions. The blot is quicker to set up as there are no denaturing and neutralising steps as the transfer buffer in this case is 0.4N NaOH which denatures the DNA simultaneously. The membrane in this case was Hybond N+ to which the negatively-charged DNA is directly fixed and thus UV crosslinking after transfer is not necessary.

2.5.3 Bidirectional blotting

This is a modification of the basic techniques described in sections 2.5.1 and 2.5.2 whereby the DNA is transferred to two membranes simultaneously, and is particularly useful where a membrane may require several rounds of hybridisation. Treatment of the gels was as described in sections 2.5.1 or 2.5.2 but the blotting assembly was altered: paper towels were stacked on the bench to a height of several centimetres and covered with 3 sheets of 3MM Whatmann paper pre-soaked in transfer buffer. A membrane cut to the size of the gel was then placed on top of this assembly. The treated gel was then positioned on top of the membrane and another membrane placed on the gel. The ‘sandwich’ was completed by placing 3 sheets of 3MM Whatmann presoaked in transfer buffer on top of this second membrane and covering the whole with stacked paper towels and a uniform weight. The remaining steps were as described in section 2.5.1.

2.5.4 Hybridisation of Southern blots with radiolabelled probes

2.5.4.1 Probe labelling

DNA probes used in this study were labelled with $^{32}$P by one of two methods. The CA repeat probe (GT)$_n$G (section 4.2.2) was end-labelled in a 20µl reaction containing polynucleotide kinase buffer (One-Phor-All; Pharmacia), 50mCi $\gamma$-$^{32}$P-dATP and 1µl (9.3 units) polynucleotide kinase (Pharmacia) at 37°C for 30-45 minutes in a similar fashion to the sequencing primer end-labelling described in section 2.6.3.

Total human DNA was labelled for hybridisation to pulsed field blots by the method ‘random prime oligolabelling’ (Feinberg and Vogelstein 1983). The probe DNA is first denatured and then mixed with hexadeoxyribonucleotide oligomers of random sequence. These ‘random hexamers’ anneal non-specifically to the probe DNA, serving as primers for DNA synthesis by the Klenow fragment of E.Coli DNA polymerase I. By incorporation of a radiolabelled nucleotide in the synthesis reaction, probe DNA is generated with high specific
activity. 50ng of total human placental DNA (Sigma) were denatured by incubation at 95°C for 5 minutes and then quenched on ice. The denatured probe was then labelled in a 30μl reaction containing 3μl reaction mixture (random hexamer mix), 3μl of dNTPs mixture (dATP, dGTP, dTTP in 1:1:1 proportions; 0.5mM each nucleotide), 5μl of α-32P-dCTP (10μCi/μl) and 2μl (4 units/μl) of Klenow enzyme, at 37°C for 2 hours. The % incorporation of label was assessed by spotting 0.5μl of the reaction onto a 2.5cm Whatman round glass microfibre filter beneath a Geiger monitor adjusted to read 100 counts. The filter was then washed under vacuum with 15ml 5% TCA (Trichloroacetic acid) to remove unincorporated label and re-monitored. If incorporation was efficient (50-80% i.e. 50-80 counts) the probe was ready for hybridisation. Probes for which incorporation was < 25% were discarded.

2.5.4.2 Hybridisation

The membranes were pre-hybridised in a 20ml hybridisation mixture containing 0.125M NaPi solution at 65°C for 30 minutes in 12x12 round petri dishes placed in a damp container (plastic lidded box with damp paper towels in the bottom). The labelled probe was denatured by incubation at 95°C for 5 minutes and added to the hybridisation mixture. Hybridisation was allowed to proceed overnight at 65°C in a slow-speed shaking incubator.

2.5.4.3 Post-hybridisation washes and autoradiography

Washes to remove non-specific hybridisation were carried out to the required stringency with decreasing concentrations of SSC and increasing concentrations of SDS (starting with two 5 minute washes in 2xSSC/0.1%SDS at RT as the lowest stringency, then two 20 minute washes in 2xSSC/1% SDS at 65°C, progressing to two 20 minute washes in 0.1xSSC/1%SDS at 65°C as the highest stringency). The membranes were monitored with a Geiger counter at each stage to assess the need for further washes, whilst taking care to ensure that the membranes did not dry out. The membranes were then sealed damp in clear plastic and exposed to X-OMAT-AR Kodak autoradiographic or Fuji X-ray film overnight at RT. Depending on the result upon development of the film, the membranes may then have been re-exposed for a longer/shorter period at RT or at -80°C with an intensifying screen to enhance the signal.

2.5.4.4 Membrane-stripping protocol

Hybond N and N+ was stripped of hybridised probe by pouring boiling 0.5% SDS onto the membranes and allowing to cool to RT; if indicated necessary by Geiger monitoring, the stripping procedure was repeated. Satisfactory removal of the probe was confirmed by autoradiography for an appropriate time period.
2.6 Sequencing single and double stranded DNA

2.6.1 Overview of dideoxy chain termination sequencing

Sequencing reactions were carried out using Sequenase version 2.0 (USB) or T7 polymerase (Pharmacia) which are genetically engineered T7 DNA polymerases lacking 3’ to 5’ exonuclease activity. The dideoxy chain termination protocols described are modified from Sanger et al. (1977).

The ddNTPs are nucleotide analogs which lack a hydroxyl residue at the 3’ position of the deoxyribose. This prevents formation of a phosphodiester bond when incorporated into the elongating DNA molecule and leads to chain termination. By using 4 different ddNTPs in 4 separate reactions, along with all four types of dNTP nucleotide, the products of each reaction are a set of DNA chains of different lengths each terminating in a ddNTP, corresponding to the different positions occupied by every A, C, G or T in the template strand. These chains are rendered detectable by including a radiolabelled dNTP in the synthesis reaction or by end-labelling the sequencing primer which is annealed to the template strand and from which the growing DNA strand extends. The nucleotide sequence of the template strand is deduced by the pattern of autoradiographic bands produced after electrophoresing the 4 separate reaction products on high resolution denaturing polyacrylamide gels.

2.6.2 Sequencing double-stranded plasmid DNA by incorporation of isotope

5μg DNA was denatured in a 20μl volume containing 0.2M NaOH and 0.2mM EDTA at 37°C for 30 minutes. 1/10 volume of 3M sodium acetate was then added and the DNA precipitated with 2.5 volumes of absolute EtOH. The DNA was pelleted by centrifugation (13000 rpm) at 4°C for 10 minutes, washed with 70% EtOH and resuspended in 7μl SDW.

DNA sequencing was carried out using a United States Biochemicals (USB) sequencing kit as follows: the resuspended DNA was annealed to 10 pmol (1μl 10mM stock) sequencing primer (vector primer T3 or T7) in a 10μl reaction containing 2μl Sequenase buffer by placing the reaction at 70°C for 3 minutes, then cooling to 40°C over 20 minutes on the bench. To the annealed template/primer mixture 1μl 100μM DTT, 2μl diluted labelling mix (1.5mM each of dCTP, dGTP, dTTP), 2μl diluted Sequenase enzyme (1:8 in ice-cold enzyme dilution buffer) and 1μl α32P-dATP (10mCi/ml) were added, and the labelling reaction allowed to proceed for 2-3 minutes at RT. The reaction was terminated by transferring 3.5μl labelled mixture to each of 4 tubes containing 2.5μl each of the 4 chain-terminating nucleotides (ddATP, ddCTP, ddGTP, ddTTP) which had been prewarmed at 37°C. Incubation continued at 37°C for 5 minutes, before adding 4μl stop mix (EDTA) to each tube.

2.6.3 Direct sequencing of PCR products using a γ-32P endlabelled primer

Prior to sequencing, PCR products were purified by passing aliquots of < or = 25μl through a Sephacryl S-400 MicroSpin Column (Pharmacia) following the manufacturer’s instructions, to remove any excess dNTPs and DNA oligos smaller than 30bp. 10pmoles of
sequencing primer was end-labelled in a 10µl reaction containing 1µl 10x One-Phor-all buffer (Pharmacia) containing 10mM Tris-acetate, 10mM magnesium acetate and 50mM potassium acetate, 1µl γ-32P-dATP (10µCi/µl), 6µl SDW and 1µl T4 polynucleotide kinase (5-10 units) at 37°C for 30-45 minutes. 3µl of the labelled primer was added to 9µl purified PCR product and the two annealed by placing the mixture at 95°C for 5 minutes (to denature the template DNA) and then cooling on ice. After spinning briefly, the following were added, all on ice: 2µl T7 polymerase (diluted 1:4 in cold enzyme dilution buffer) (Pharmacia T7 polymerase sequencing kit), 1µl DTT (100mM) and 3µl dATP labelling mix (Pharmacia kit). 4.5µl of this mixture was then transferred to each of 4 tubes containing 2.5µl of termination mixes (ddATP, ddCTP, ddGTP, ddTTP) which had been prewarmed at 37°C. After incubation at 37°C for 5 minutes the reactions were terminated with 4µl 'stop solution' (95% formamide; Pharmacia kit) for each tube.

2.7 Polymerase chain reaction (PCR) techniques

2.7.1 General considerations

The polymerase chain reaction (PCR) 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). Two oligonucleotide primers are synthesised which derive from opposite strands of the template DNA to be amplified and which have their 3' termini facing each other. The target DNA is denatured in the presence of a large excess of the two primers and then returned to a temperature which will allow the primers to anneal to the DNA. A heat-stable DNA polymerase (most frequently that isolated from the thermophilic bacterium *T. aquaticus* and therefore called *Taq* polymerase), and all 4 nucleoside triphosphates are included in the reaction and the sample is placed at a temperature optimal for elongation by the enzyme. This produces two copies of the sequence lying between the primers. The three steps of denaturing, annealing and extension are repeated, producing another 4 copies of the sequence bracketed by the oligonucleotides primers. Two of these will have indeterminate 3' termini but two will now have termini dictated by the 5' terminus of the other primer. In subsequent cycles of reactions there will be further exponential amplification (n cycles of PCR amplify the target 2^n-fold) such that, within a few cycles, the predominant product is that defined by the 5' termini of the starting primers.

2.7.2 Standard parameters for a typical PCR

PCR was performed using *Taq* polymerase from NBL or Bioline using the manufacturer’s 10x buffer containing 10mM Tris-HCl, 50mM KCl, 15mM 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 25µl: 1x manufacturer's buffer, 0.2mM each dNTP, 0.1-0.2µM each primer, 150ng template DNA and 0.5U *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. 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 10 seconds, annealing at the appropriate temperature for 10 seconds, and extending at 72°C, allowing 20 seconds per kilobase to be extended. Cycle number varied between 30 and 40 cycles, depending upon the template and particular methodology. Once optimised, PCRs were generally performed on the same machine, as they tend to vary slightly in temperature and cycle time; either a Hybaid Omnigene or a Techne PHC-3.

2.7.3 Primer design and PCR optimisation

The optimal annealing temperature for specific primer pairs was approximated by first calculating the melting temperature ($T_m$) 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 2-5°C lower than the value obtained (e.g. a $T_m$ of 62°C would indicate an annealing temperature of 59-60°C). Primer pairs were then tested by PCR on several samples of human genomic DNA (with a 'no DNA' control) and electrophoresed on agarose gels (section 2.4.1) to assess the adequacy of the PCR conditions for subsequent experiments. Additional bands (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 bands could usually be eradicated by increasing the annealing temperature by 1-2°C (since ionic strength can affect $T_m$, an alternative solution would be to decrease the magnesium concentration).

When designing primer pairs for PCR amplification of known sequences, several rules were followed as far as possible to ensure an optimal result: random base distribution and 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 $T_m$ values of both primers and as far as possible given the above constraints, a primer length of at least 20 nucleotides (more if possible) to increase the sequence specificity. In our experience, longer primers also tended to be more robust in PCRs from difficult samples e.g. YAC library pools (section 2.7.5). PCR primers were synthesised by either the HGMP Resource Centre or commercial manufacturers such as Cruachem or GenoSys.

2.7.4 Microsatellite analysis

Both $^{32}$P and $^{35}$S may be used to label microsatellite products; the former is widely used to label one end of a primer, and either isotope may be incorporated into the product during PCR amplification (Litt 1991). Both methods worked well throughout this study, although the
easier discrimination of allele differences by end-labelling made this the preferred method for most microsatellites. It was generally found that the relative success of each method was STS specific (e.g. DXS1003 best by primer end-labelling, DXS1055 best by incorporation). This of course is subject to which primer is end-labelled, and incorporation labelling proved especially useful where one/both of a primer pair had homology to ALU (e.g. DXS337/2). The microsatellite allele size differences (multiples of 2-4bp) were resolved by polyacrylamide gel electrophoresis (section 2.4.2).

2.7.4.1 End-labelled primer

One primer was end-labelled with $\gamma^{32}$P-dATP as described in section 2.6.3, using 1 pmol primer per PCR reaction from a 10$\mu$M stock (i.e. 1$\mu$l per 10 PCR reactions) and 1$\mu$l (10$\mu$Ci) $\gamma^{32}$P-ATP per 20 PCR reactions; the final labelling volume was therefore increased to 15-20$\mu$l for large sample sets. The human template DNA samples (150ng) were dispensed into separate 0.5ml eppendorfs, overlaid with mineral oil and an appropriate volume of PCR reaction mix (section 2.7.2) dropped through the oil to give a total volume of 10$\mu$l (to minimise the exposure to radioactive isotope).

2.7.4.2 Incorporation of label

PCR reactions were performed in a total volume of 10$\mu$l with a reaction mix as described in section 2.7.2, with the final dNTP concentration modified so that cold dCTP was present at only 1/10 the concentration of the other dNTPs (i.e. 20$\mu$M rather than 200$\mu$M) and with 0.5$\mu$Ci $\alpha^{32}$P-dCTP added per reaction (i.e. 10$\mu$Ci of $\alpha^{32}$P-dCTP was added to a reaction mix for 20 reactions). The template DNA was dispensed into 0.5ml eppendorfs and overlaid with mineral oil prior to the addition of reaction mix.

After microsatellite PCR amplification using either method 2.7.4.1 or 2.7.4.2, 6$\mu$l of formamide loading buffer was added to each reaction prior to denaturing polyacrylamide gel electrophoresis (section 2.4.2).

2.7.5 YAC library screening by PCR

Three human YAC libraries were screened by PCR using a variety of STSs from the RP2 gene region on Xp11.3-Xp11.22 to identify corresponding YAC clones. Details of these YAC libraries are given below. The ICI and ICRF libraries were available from HGMP whilst the CEPH mega-YAC library was obtained directly from CEPH. YAC libraries were screened as hierarchical pools of decreasing complexity (section 7.2) until a single positive clone was identified.

ICI YAC library (Anand et al 1990): 34,500 clones with an average insert size of 350kb and a >3.5x coverage of the human genome, and was made from a human lymphoblastoid 48XXXX cell line. Data so far suggests that approximately 10% of the clones are chimaeric.
PCR pools: 40 1° pools (numbered 1-40) of ~860 YAC clones each. Each 1° pool has 9 corresponding 2° pools (A-I), which represent the 9 plates comprising each 1° pool (i.e. 95 YAC clones per 2° pool). Each of 20 tertiary pools correspond to the rows (pools 1-8) and columns (pools 9-20) of each plate (i.e. secondary pool) to establish the plate coordinates for the positive clone(s).

ICRF YAC library (Larin et al 1991): 20500 clones and combines 3 separate libraries with an overall chimaerism rate of ~30% -
4X (human female lymphoblastoid 48XXXX cell line): 15700 clones, average insert size 600kb, 3x coverage of autosomes and 6x coverage of the X chromosome.
4Y (human male lymphoblastoid cell line 49XYYYY): 2300 clones, average insert size 400-500kb.
HD (human female lymphoblastoid, Huntingdon’s disease 46XX): 2500 clones; average insert size 600kb.

PCR pools: 27 1° pools (numbered A-Z and ZZ) of ~760 YAC clones each. Each 1° pool has 8 corresponding 2° pools (A-H) which represent the 8 plates comprising each 1° pool (i.e. 95 YAC clones per 2° pool). Tertiary pools were as for the ICI library.

CEPH mega YAC library (Chumakov et al 1992): This is a subsection (plates 613-984) of the CEPH YAC library made from genomic DNA of a 46XY male, comprising 35600 larger YACs with an average insert size of 920kb and a 7-8x coverage of the human genome. The chimaerism rate has been reported to be about 40%.

PCR pools: 29 1° pools of 1228 YAC clones each.
- 2° and 3° screening performed at CEPH.

The DNA pools (with the exception of the CEPH library) were received embedded in agarose plugs stored in 5mM EDTA. For PCR, these were washed in TE for two hours, with shaking and two changes of buffer, then left in TE at 4°C overnight, to remove all inhibitory reagents from the agarose. The plugs were then melted at 65°C, diluted 1:1 with SDW and overlaid with mineral oil to prevent evaporation, which would concentrate the agarose and increase the concentration of residual inhibitory compounds. 3μl aliquots of the melted plugs or 2μl of CEPH pool DNA were used as DNA template in each 25μl PCR, which was performed according to standard conditions (section 2.6.2). The Taq polymerase was added after a ‘hot start’ of 95°C for 5 minutes to ensure the long YAC DNA molecules were fully denatured. YAC clones were received either as ‘stabs’ or streaked out on selective agar, and were restreaked onto fresh AHC + ampicillin (section 2.2.4) to initiate their growth as single colonies.
2.7.6 Screening for YAC-positive transformant yeast clones by colony PCR

2.7.6.1 PCR of intact yeast cells

A standard ‘hot start’ PCR reaction was carried out (section 2.7.5) using primers and cycling conditions specific to the STS used to identify the YAC (i.e. an STS known to reside within the YAC). 5-10 single large yeast colonies (0.5-2mm) were tested for their presence of the YAC by picking a small amount of colony with a sterile loop, mixing it with 20μl TE in a microtube, then using 2μl of this as DNA template for the PCR reaction. The PCR products were analysed by agarose gel electrophoresis (section 2.4.1) and only colonies displaying a clear band of the correct size were chosen for continued growth and DNA preparation (section 2.2.4). This method was not 100% reliable.

2.7.6.2 PCR of sphleroblasts

PCR and agarose gel electrophoresis were performed as in section 2.7.6.1. However, the yeast colony templates were treated to break down the yeast cell wall prior to PCR using a protocol reported by Ling et al (1996). A small amount of each colony was picked with a sterile loop into 10-20μl of incubation solution (1.2M sorbitol, 100mM sodium phosphate pH 7.4, and 2.5mg/ml Lyticase). The resulting enzyme/cell mixture solution was incubated at 37°C for 5 minutes. 3μl of the mixture solution was then used as a template for PCR. The remaining sphleroblast sample was stored at -20°C for repeated use if necessary. This method generally proved more reliable i.e. if 5 colonies were screened, at least one positive transformant was always detected.

2.7.7 Inter-ALU PCR to generate YAC ‘fingerprints’

ALU repeats are the most common member of the SINE (short interspersed repeat element) family of interspersed repetitive sequences (IRSs) scattered throughout the human genome, and consist of a core repeat region of about 300bp occurring on average every 4kb (Batzer and Deininger 1991). Because they are absent in the yeast genome, only human-specific DNA fragments will be amplified when a yeast clone containing a YAC is used as a template for PCR. With a single ALU primer and a YAC-clone DNA template in a PCR, amplification of one or more PCR products is expected, depending on the orientation and distance between neighbouring ALU elements in the insert. The pattern of fragments obtained when the resulting products are subject to agarose gel electrophoresis will be specific to the unique region of the genome comprising the YAC insert DNA and is thus termed a ‘fingerprint’.

Two ALU primers were used: ALE1 (5’ GCC TCC CAA AGT GCT GGG ATT ACA G 3’) and ALE3 (5’ CCA T/CTG CAC TCC AGC CTG GG 3’) (Cole et al 1991) which correspond to the most conserved regions at the extremities of the ALU consensus sequence and are oriented to amplify away from the ALU repeat, thus maximising the amount of unique
sequence in the PCR product (Nelson et al. 1989) and ensuring maximal representation of the YAC insert.

Inter-ALU PCR was carried out in a 50μl reaction using a standard PCR reaction mix (section 2.7.2) with a final concentration of 1μM ALU primer and using 3μl of YAC plug DNA as template. For each YAC, all combinations of ALU primers were used i.e. ALE1 alone, ALE3 alone and ALE1/ALE3 together (to generate a more complex set of products). After a ‘hot start’, PCR cycling conditions were 40 cycles of 95°C for 1 minute, 65°C for 1 minute, 72°C for 2 minutes (a long extension time to permit amplification between more widely-spaced ALU elements), after which 10μl of reaction products were analysed by agarose gel electrophoresis (section 2.6.1) using a 2% low-melting point (LMP) agarose gel containing 0.5μg/ml ethidium bromide. Frequent checks were made on the extent of separation by visualising the gel under UV light; it was important to run the gel far enough so that adequate resolution of bands was achieved.

2.7.8 Generation of YAC insert-terminal sequences

2.7.8.1 ALU-vector PCR

ALU-vector PCR uses ALU primers together with a primer homologous to vector sequences that lie adjacent to the cloning site. This method results in amplification of sequences between the cloning site of the vector and an ALU element within the insert, thus generating specific end-fragments from the cloned DNA. This protocol is rapid and straightforward.

The vector primers were specific for the left and right arms of the pYAC4 vector (LA: 5’CAC CCG TTC TCG GAG CAC TGT CCG ACC GC 3’, RA: 5’ ATA TAG GCG CCA GCA ACC GCA CCT GTG GC 3’). The PCR reaction mix and conditions were as in section 2.7.7 with the following primer combinations for each YAC: ALE1/LA, ALE1 alone, ALE1/RA, ALE3/LA, ALE3 alone, ALE3/RA. The inclusion of ‘ALU primer alone’ reactions acted as a control, so that any ALU-vector PCR products obtained could be deduced by comparison after agarose gel electrophoresis (section 2.4.1). Any ALU-vector bands which were absent from the ‘ALU primer alone’ reactions were verified by stabbing the band under UV illumination with a Gilson pipette tip to extract a small quantity of DNA, to act as the template in a second round of PCR using the relevant ALU-vector primer combination and identical conditions. This procedure also provided an enriched sample of the specific ALU-vector product bereft of the products generated by inter-ALU PCR. The 40μl of remaining PCR product after agarose gel electrophoresis was then purified, sequenced directly (section 2.6.3) using either a RA or LA pYAC4 vector primer internal to the PCR primers (LA: 5’GTT GGT TTA AGG CGC AAG 3’ and RA: GTC GAA CGC CCG ATC TCA AG 3’) and analysed after denaturing polyacrylamide gel electrophoresis and autoradiography (section 2.4.2).
2.7.8.2 Vectorette PCR

This method is based on that of Riley et al (1990) and is also known as Chemical Genetics since it does not involve a biological system in which to propagate cloned DNA i.e. E. Coli transformed with plasmid containing insert. Vectorette PCR permits PCR amplification of a DNA fragment when sequence from only one end is known, and is therefore directly applicable to the generation of YAC insert terminal sequences. It involves restriction enzyme digestion of DNA from the yeast clone harbouring the YAC, followed by ligation of the restriction fragments to a specially designed double-stranded 'Vectorette' linker compatible with the ends generated by digestion. Anchored PCR is then performed using YAC vector-specific left and right arm primers and a primer specific to the Vectorette unit, to amplify the DNA covering the YAC-insert/YAC arm junction. The success of this technique rests on the design of the Vectorette itself, which has a central region of non-complementarity. The universal Vectorette primer is identical to the bottom strand of this mismatched region, so that when PCR is carried out, the universal Vectorette primer cannot bind to the Vectorette. However, when a complementary copy of the bottom strand is synthesised, the universal Vectorette primer is able to bind and participate in PCR. Although many fragments of yeast will have Vectorette units ligated to them, use of the YAC vector primer to prime synthesis of the bottom strand during the first cycle of PCR ensures that the only product generated will correspond to that flanking the YAC insert terminal.

2.7.8.2.1 Construction of Vectorette library

A segment (approximately 1/4) of a 200μl YAC DNA plug was equilibrated in 1ml of cold TE overnight to remove low molecular weight impurities. The TE was replaced with 1ml of the relevant 1x restriction buffer for 1 hour on ice. The buffer was then replaced with 100μl of fresh 1x restriction buffer plus 20U of the chosen restriction enzyme and incubation was performed overnight at 37°C. The restriction enzyme/buffer solution was replaced with 1ml of 1x ligase buffer (Gibco BRL or Pharmacia) and the plug equilibrated for 1 hour on ice. The buffer was then replaced with 100μl of fresh ligase buffer and 5μl (3pmol) of the appropriate blunt-ended Vectorette units (CRB) were added. The plug was melted at 65°C for 10 minutes, followed by cooling to 37°C. ATP to final concentration of 1mM and 9U of T4 DNA ligase (Gibco BRL or Pharmacia) were added and the ligation reaction was incubated at 37°C for 2 hours. 100μl of SDW was added and the Vectorette library was stored in 20μl aliquots at -20°C.

2.7.8.2.2 PCR Amplification of Vectorette library

PCR amplification of the Vectorette library was performed using the Universal Vectorette primer (CRB) and the pYAC4 left or right arm primers (CRB). For PCR amplification, the following were combined in a sterile microtube: 1μl of Vectorette library, 10μl of 10x PCR buffer (NBL), 50μM of each dNTP, 1μM Universal Vectorette primer and
1µM of specific primer. The volume was made up to 99µl with SDW and the sample overlaid with mineral oil. Following a ‘hot start’ and the addition of 1U of Taq polymerase, PCR conditions consisted of 40 cycles of 95°C for 1 minute, 60°C for 1 minute and 72°C for 2.5 minutes, with a final extension of 72°C for 10 minutes. The Vectorette PCR products were analysed on a 1% agarose gel stained with ethidium bromide.

2.7.8.3 Somatic cell hybrids used for chromosomal localisation of STSs

STSs were localised to proximal Xp by PCR assay of DNA from human-rodent somatic cell hybrids containing various portions of the X chromosome: ThyBX (whole X), SIN176 (X deleted for Xp22.11-p11.22; Ingle et al 1985) and KAG2.3 (made from a female with translocation Xp21.2::22q13.3 and retaining the X chromosome segment Xp21.2-qter; Boyd et al 1988), with mouse and hamster DNA as controls for the parental cell lines. STSs that yielded a product of the expected size with human genomic DNA, ThyBX, KAG2.3 but not SIN176, mouse or hamster DNA were ascribed to human Xp21.1-p11.22.

The Biosmap multichromosomal somatic cell hybrid panel was used (Bios Laboratories) to establish the autosomal location of STSs that were excluded from the X chromosome by analysis of the X-specific hybrids described above. It is described in greater detail in Figure 7.15 (chapter 7).

2.8 Cloning techniques

2.8.1 Construction of a minilibrary from cosmid HX20 by ‘shotgun’ subcloning

2.8.1.1 Vector and cosmid restriction digests and DNA purification

5µg each of vector (pT7T3) and cosmid DNA were digested with BamHI and Sau3AI, respectively, as described in section 2.3. The digested DNAs were extracted with, in turn, 50µl of phenol, phenol/chloroform (1:1) and finally chloroform. 1/10 volume of 3M sodium acetate was added to the final aqueous phase, and the DNA precipitated with 2 volumes of absolute ethanol. Following centrifugation for 10 minutes at 4°C, the pelleted DNA was washed with 70% ethanol, air dried and resuspended in 10µl TE (cosmid) and 20µl TE (vector). 1µl of each purified DNA was electrophoresed on a 0.8% agarose gel (section 2.4.1) to assess completion of digestion and DNA concentration.

2.8.1.2 Insert dephosphorylation and ligation to vector

Insert DNA fragments were dephosphorylated rather than vector DNA to minimise the chance of insert coligation. Cosmid HX20/Sau3AI DNA (1µg) was dephosphorylated by combining the DNA with 1µl 50mM CaCl₂, 1µl phosphatase buffer, 1 unit HK™ phosphatase (Epicentre Technologies) and SDW to give a final volume of 10µl. The temperature of the reaction was adjusted to 30°C before the addition of phosphatase (to ensure optimal enzyme activity) and incubation continued for 1 hour. The reaction was then placed at 65°C for 30 minutes to heat inactivate the enzyme.
Vector-insert ligations were carried out in 10μl reactions using 0.1U T4 DNA ligase (Biolabs) and allowed to proceed at 12°C overnight. The efficiency of ligation was assessed by running 2μl of the reaction on a 0.8% agarose gel.

2.8.2 Isolation of a CA repeat-containing subclone from the HX20 minilibrary

2.8.2.1 Preparation of competent E.Coli XL1-Blue cells

In order to introduce recombinant DNA molecules into a host bacterial cell (transformation) for subsequent propagation and isolation, the host cells must be made receptive to the recombinant molecules. Most commonly, bacterial membranes are rendered temporarily permeable to DNA by a brief electric shock (electroporation) or exposure to divalent cations. The latter approach was adopted here using E. Coli XL1-Blue cells and gentle treatment throughout.

E. Coli cells (from a glycerol stock) were streaked onto LB agar supplemented with 12.5 μg/ml tetracycline and incubated overnight at 37°C in the dark (tetracycline is light-sensitive). A colony was picked under sterile conditions and transferred to 50ml LB broth with 12.5μg/ml tetracycline. After overnight incubation with agitation at 37°C, 0.5ml of this culture was used to inoculate a 50ml culture of medium A. The cells were grown at 37°C with aeration in a shaking incubator to mid-log phase (i.e. exponential growth phase) when A<sub>oo</sub> reached 0.38 - 0.4 as determined by spectrophotometry. The cells were kept on ice for 10 minutes, then pelleted at 2500 rpm for 10 minutes at 4°C (brake off). The cells were gently resuspended in 0.5ml precooled (4°C) medium A. 2.5ml precooled storage medium B were added and mixed thoroughly but gently with the cell suspension. The cells were divided into 250μl aliquots, flash frozen in liquid N<sub>2</sub> and stored at -70°C.

2.8.2.2 Transformation of competent cells with ligated products

2μl (40ng DNA) of the ligation mixture was added to 250μl competent cells which had been previously thawed on ice. The transformation mixtures were incubated at 4°C on ice for 30 minutes, heat-shocked at 42°C for 2 minutes and then chilled on ice for 1 minute. 800μl prewarmed LB broth was added to each mixture, which was then placed at 37°C for 1 hour. After centrifugation at low speed (4000rpm) for 1 minute, the supernatant was removed and the cells resuspended in 100μl LB broth. These cells were plated onto sterile Hybond N filters on top of agar plates supplemented with 50μg/ml ampicillin, 60μg/ml IPTG (isopropyl-β-D-galactoside) and (40μl of 2% in DMF stock) Xgal (5-bromo-4-chloro-3-indoyl-β-galactoside). The plates were incubated at 37°C overnight.

2.8.2.3 Preparation of replica filters

Master filters were peeled off each transformation plate and laid (colony side up) on sterile 3mm Whatmann paper. A fresh sterile filter was laid on each master filter. The two filters were pressed firmly together and oriented by making a series of asymmetrical needle
holes. The master filters were placed back onto their original plates and the replicas onto fresh LB agar plates supplemented with 50μg/ml ampicillin. Both master and replica plates were incubated at 37°C for 2 hours. The master plates were then stored at 4°C. Replica filters were prepared for hybridisation by laying them on a series of 3mm Whatmann papers in turn, soaked in a) 10% SDS for 3 minutes, b) denaturer for 5 minutes, c) neutraliser for 5 minutes and d) 2xSSC for 5 minutes. They were air dried and the DNA was fixed with UV irradiation for 5 minutes. For prolonged storage of master filters, the filters were placed on LB agar plates containing 50μg/ml ampicillin and 25% glycerol, incubated for 1 hour at 37°C before sealing with parafilm and placing inverted in a sealed plastic bag at -70°C.

2.9 Reagents and buffers used in this research

2.9.1 Stock reagents in general use

2M Tris-HCl (trizma hydrochloride pH 7.5, 7.8 or 8.0) : adjust pH to 7.5 or 8.0 with 5M NaOH, autoclave before use.

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

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

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

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

20xSSC : 3M NaCl, 0.3M tri-sodium-citrate.

2.9.2 FISH

Probe purification

Stop buffer : 300mM EDTA pH 8.0

TNE buffer : 0.2M NaCl, 10mM Tris-HCl pH 8.0, 1mM EDTA.

Biotin incorporation test

Buffer A : 0.1M Tris-HCl pH 7.5, 1xSSC.

Buffer B : 0.1M Tris-HCl pH 7.5, 1xSSC, 3% BSA (bovine serum albumin) fraction V.

Buffer C : 0.1M Tris-HCl pH 7.5, 1xSSC, 50mM MgCl₂.

Buffer D : 0.01M Tris-HCl pH 7.5, 1mM EDTA pH 8.0.

Hybridisation mix : see sections 2.1.5.2 and 2.1.5.3
Post hybridisation washes

Solution 1: 2xSSC, 50% formamide.
Solution 2: 2xSSC.
Solution 3: 4xSSC, 0.05% Triton-X 100.

Signal detection

Blocking buffer: 4xSSC, 0.05% Triton-X 100, 5% low-fat milk powder (0.5g low-fat milk powder were dissolved in 10ml of wash solution 3, 2.9.2.4)

1° antibody (FITC-conjugated avidin DCS) 4 μg/ml: 2μl of 2mg/ml stock in 1ml blocking buffer.

2° antibody (biotinylated goat anti-avidin) 5 μg/ml: 1μl of 0.5mg/ml stock in 100μl blocking buffer.

2.9.3 Preparation of DNA

Human genomic DNA

10x lysis buffer: 50mM Tris-HCl (pH 7.5), 25mM MgCl₂·6H₂O, 0.6M sucrose, 5% (v/v) Triton X-100

10x suspension buffer: 0.1M Tris-HCl (pH 7.5), 0.1M NaCl, 0.1M sodium EDTA (pH 8.0).

Reagent A (5x lysis buffer): 320mM sucrose, 10mM Tris HCl (pH 7.5), 5mM MgCl₂ (adjust pH to 8.0 with 5M NaOH), autoclave then add 1% Triton X-100.

Reagent B: 400mM Tris-HCl (pH 7.5), 60mM EDTA, 150mM NaCl (adjust pH to 8.0 with 5M NaOH), autoclave then add 1% SDS.

Cosmid or 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: 50mg/ml stock; plasmids-1μl per 1ml media (f.c. 50μg/ml) (NB: heat-sensitive) cosmids-1μl per 2ml media (f.c. 25μg/ml)

Tetracycline: 10mg/ml stock; plasmids- 2.5μl per ml of media (f.c. 25μg/ml) (NB: light-sensitive)

Solution I: 50mM glucose, 25mM Tris-HCl (pH 8.0), 10mM EDTA, prepared from 0.5M EDTA and 2M Tris-HCl stocks, autoclave before use.

Solution II: 0.2M NaOH, 1% (w/v) SDS (prepared from 5M NaOH and 10% SDS stocks).

Solution III: 3M potassium acetate 2M acetic acid, autoclave before use.
Yeast DNA

AHC broth: 1.7g yeast nitrogen base, 10g casein hydrolysate, 20mg 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.

SCE buffer: 1M sorbitol, 0.1M NaCl, 10mM EDTA, autoclave before use.

SCEM buffer: SCE with 30mM β-mercaptoethanol, freshly prepared.

1M sorbitol: autoclave before use.

Proteinase K/sarcosyl solution: 1% sodium lauryl sarcosine (SLS), 0.5M EDTA pH 8.0, 2mg/ml proteinase K.

10mM stock PMSF (phenylmethylsulphonylfluoride): 17.4mg PMSF in 10ml isopropanol (i.e. 1.74mg/ml). 1.14ml of stock solution were added to 50ml TE buffer to give a final concentration of 0.04mg/ml.

2.9.4 Gel electrophoresis

10xTAE buffer (pH 8.0): 0.4M tris-acetate, 10mM EDTA.

10xTBE buffer: 1M trisma base, 0.83M boric acid, 10mM EDTA.

10x loading buffers - agarose gels: 25% (w/v) ficoll, 0.25% (w/v) orange G, 10mM EDTA.

polyacrylamide gels: 95% formamide, 0.3% bromophenol blue, 0.3% xylene cyanol, 10mM EDTA.

Ethidium bromide (10mg/ml): 1g ethidium bromide/100ml distilled H2O.

2.9.5 Southern blot reagents

Denaturer: 0.5M NaOH, 1.5M NaCl.

Neutraliser: 0.5M trisma base, 3M NaCl, 0.3M Tri Na acetate.

0.4N NaOH

Depurinator: 0.25M HCl

Hybridisation mix: 0.125M NaPi (pH 7.2; 68.4ml 1M Na2HPO4 plus 31.6ml 1M NaH2PO4 per litre of distilled H2O) 1mM EDTA, 7% SDS, 10% PEG (polyethylene glycol).

2.9.6 Preparation of competent E.Coli cells

Medium A: LB broth containing sterile 10mM MgSO4.7H2O and sterile 0.2% glucose.

Medium B: LB broth containing sterile 36% glycerol, 12% polyethylene glycol (PEG), 12mM MgSO4.7H2O.
CHAPTER 3

Subchromosomal localisation of X-specific cosmid clones by FISH

3.1 Introduction

3.1.1 Physical assignment of DNA sequences to chromosomal subregions

Cytogenetic methods of physical mapping provide an important tool for the precise assignment of DNA sequences to particular chromosomal subregions (section 1.3.1). Such techniques include the use of rodent-human hybrid cell lines containing individual human chromosomes in whole or in part, and panels of well-characterised hybrids are becoming especially useful for fine mapping and ordering of nonpolymorphic DNA sequences such as ESTs and anonymous STSs in specific subregions of human chromosomes, for the generation of highly annotated physical maps to facilitate positional cloning projects (Patel et al. 1995). The recent development and availability of whole-genome irradiation hybrid mapping panels (section 1.3.1.1) now offers the possibility of mapping and ordering DNA sequences with very high resolution to defined chromosomal subregions, pending appropriate statistical analysis.

However, for precise regional localisation of DNA probes to chromosomes where high resolution hybrid panels are not available, the most direct method for identifying the chromosomal locus is by in situ hybridisation to metaphase spreads. This technology has progressed from a laborious and time-consuming approach to detect abundant nucleic acid sequences with low resolution to an approach that allows fast, highly precise, and sensitive localisation of as little as one molecule per cell (Trask 1991). Fluorescence in situ hybridisation (FISH) has the advantage of providing rapid results which can be conveniently scored by eye using a fluorescence microscope. This approach is a particularly efficient technique for identifying the relative locations and orders of DNA segments from the same chromosomal region, obviating the extensive prior characterisation and post-statistical analysis required of hybrid mapping methods. FISH is deemed to be the most efficient approach to chromosomal localisation of large-insert clones such as cosmids, as it provides a more precise regional localisation than is achieved from the degree of resolution normally permitted from regional hybrid mapping panels.

3.1.2 In situ hybridisation

In situ hybridisation (ISH) is the hybridisation-mediated detection of targeted nucleic acid sequences within structurally intact cells, tissue sections or chromosomes. As such, it
uniquely provides direct localisation of nucleic acid, either DNA or RNA, superimposed on observable cellular and subcellular structural detail.

The early phase of the development of this field relied on autoradiographic detection of radioactive probes to localise abundant sequences, such as localisation of DNA sequences in amplified polytene chromosomes and highly reiterated sequences on metaphase chromosomes. DNA probes existing as multiple, often tandem copies, made detection easier because a large signal was generated by the many labelled probe molecules concentrated in a small chromosomal region, and during the following decade, applications were generally restricted to highly represented sequences.

It was not until 1981 that Gerhard and colleagues first demonstrated the feasibility of localising single-copy sequences cloned from individual genes onto metaphase chromosomes by isotopic ISH using $^{125}$I- or $^3$H-labelled probes. Two developments had facilitated this achievement: firstly, the availability of cloned genomic fragments containing structural genes and the intervening and surrounding sequences meant that the length of DNA sequence hybridised for single copy genes approached that for repetitive families, and secondly, the amplification of the signal, obtained by the addition of dextran sulphate to the hybridisation mixture. Dextran sulphate increases the rate of hybridisation by promoting the formation of extended double-stranded probe networks at the site of hybridisation (Lawrence and Singer 1985). Detection of such networks is easier because they are much larger than the individual probe molecules.

ISH using radioactively-labelled probes subsequently became popular because of its high sensitivity in nucleic acid detection. The technique is useful to show the initial existence of a DNA or RNA species, as well as its precise localisation within a tissue section or chromosome spread, and has become an invaluable tool for studying the expression of genes and for demonstrating viral infection. Despite its sensitivity, however, the use of radioactive probes has certain disadvantages: probe instability due to radiolysis, the relatively long delay necessitated by autoradiographic detection procedures (days-weeks), high background noise (identification of the site of localisation requires statistical analysis of autoradiographic grain distribution in as many as 50-100 metaphases), the fact that the radioactive signal and the underlying chromosomes cannot be visualised microscopically in the same focal plane, and the scatter of radioactive disintegrations limits the resolution to only relatively large chromosomal segments.

To overcome these limitations, in the late 1970s and early 1980s nonradioactive ISH techniques were introduced and developed. Most nonradioactive methods share the same principle of coupling a hapten to the DNA/RNA probe, followed by detection with specific antibodies or affinity reagents after the ISH has been performed.

The most widespread approach is to label probes with reporter molecules which, after hybridisation, bind fluorescent affinity reagents (Fluorescence In Situ Hybridisation or FISH). The expanding use of FISH is the result of several factors, notably: 1) advantages over isotopic methods including high specificity with increased signal:noise ratio, eliminating the need for extensive statistical analysis, improved spatial resolution, rapid results, and probe stability 2) the sensitivity of the technique approaches that of isotopic methods (1-5kb) 3) a variety of
probe-labelling schemes are available for simultaneous detection, in different colours, of two or more different sequences within the same preparation (Trask et al 1991) 4) the steadily improving quality of fluorescence microscopes over the past decade 5) the ability to specifically highlight either the entire genome of a particular species, entire chromosomes, chromosomal subregions, or single-copy sequences depending on the complexity of the probes used and 6) the facility to suppress hybridisation of species-specific repetitive sequences by prehybridisation of probes with unlabelled genomic DNA (Lichter et al 1991). This is crucial for localisation of unique sequences contained in large-insert probes, such as cosmids and yeast artificial chromosomes (YACS).

3.1.3 FISH Methodology

The technique is relatively simple: DNA or RNA probes are first labelled with reporter molecules. The probe and the target chromosomes or nuclei are denatured. Complementary sequences in the probe and target are then allowed to reanneal. After washing and incubation in fluorescently labelled affinity reagents, a discrete fluorescent signal is visible at the site of probe hybridisation.

3.1.3.1 Chromatin preparation and denaturation

Cells are hypotonically swollen and fixed on slides by standard methanol:acetic acid fixation, which provides excellent preservation of cell and chromosome morphology without altering the target nucleic acid, whilst still allowing access and hybridisation of the probe. Slides are then treated with RNAase to remove endogenous RNA which might act as a competitor for the labelled probe and result in an increased unspecific background signal. Careful washing after the RNAase treatment ensures that no enzyme remains to degrade the hybridisation probe.

Base, acid or high temperatures can be used to denature the chromosomes. The former two methods tend to be less efficient, but give good results if the DNA sequences of interest are repeated (e.g. ribosomal DNAs, satellite DNAs). Highly efficient denaturation methods are needed for unique sequence localisation; in such cases slides are incubated briefly at 70°C in a solution containing 70% formamide to dissociate the DNA into single strands, and then fixed by dehydration in cold ethanol to quench the denaturation reaction and thereby reduce strand reannealing before the addition of probe. Such heat treatment followed by ethanol dehydration also reduces the loss of morphology that occurs when part of a chromosome does not adhere tightly to the slide. Formamide alters the dielectric constant of a solution and thereby lowers the melting point of DNA approximately 0.7°C per % of formamide (McConnaughy et al 1969). This decrease in the temperature required to melt the chromosomal DNA helps to maintain good chromosome cytology. After denaturation, slides should be used for hybridisation within no more than 12 hours, because the efficiency of hybridisation decreases rapidly as a function of storage time after denaturation.
3.1.3.2 Probe labelling

A wide variety of nucleic acid probes, probe labelling reactions, and labels have been used for FISH. Total genomic DNA can be used as a species-specific label, as the repetitive sequences it contains show little evolutionary conservation and it is these sequences that reanneal more rapidly than the more highly conserved unique sequences because of their high frequency in probe and target. Chromosome-specific repeats can be used to tag particular chromosomes and are described in more detail in section 3.2.1.1. More recently, chromosome ‘painting’ probes have been developed (reviewed in Lichter et al 1991). These consist of collections of DNA sequences derived from a single human chromosome (e.g. from somatic cell hybrids carrying the desired chromosome as its only human material). Prehybridisation with unlabelled genomic DNA suppresses hybridisation of repetitive elements in the collection that are common to many chromosome types, so that only the chromosome of interest is highlighted. These probes have proved invaluable for the detection of cytogenetic abnormalities. Single-copy sequences can be mapped using FISH, although the efficiency of hybridisation site detection decreases with decreasing probe size. The efficiency of labelling large insert probes (such as cosmids or YACs) is > 90% under suppression conditions.

Although probes may be directly conjugated with fluorescent molecules (Koch et al 1992), the signals obtained are relatively weak, limiting the technique to the detection of repeated DNA sequences or abundant mRNA species. Alternative labelling techniques have been developed in an effort to produce more sensitive or convenient technology. The most widespread approach is to label probes with reporter molecules which, after hybridisation, bind fluorescent affinity reagents. Typical reporter molecules include biotin, digoxigenin, dinitrophenyl (DNP), aminoacetylfluorene (AAF), mercury and sulphonate (reviewed in Bauman et al 1990). For the first three of these reporter molecules, enzymatic incorporation of nucleotides modified with biotin, digoxigenin or DNP is usually preferred over chemical labelling techniques employing photoreactive compounds (photobiotin, photodigoxigenin, or photodNP) because of a higher labelling efficiency (Lichter et al 1991). However, chemical modification schemes using the other reporter molecules AAF, mercury or sulphonate, have been used successfully for sensitive nonradioactive detection of hybridised nucleic acids (Landegent et al 1987). Alternatively, probes can be labelled by PCR amplification between known priming sequences (primer-induced sequence-specific labelling or PRINS, Koch et al 1992) or by RNA transcription from appropriate vectors, in the presence of labelled nucleotides.

Biotin and digoxigenin labelling combined with fluorescent detection are currently the most widely used because of the high sensitivity of detection and the commercial availability of the reagents. Typically employed are dUTP or UTP analogs that contain a biotin molecule covalently attached to the C-5 position of the pyrimidine ring through an allylamine linker arm (Langer et al 1981). Such modified nucleotides can function as suitable substrates for DNA or RNA polymerases in vitro, and can thus be introduced into DNA or RNA molecules.
to form efficient labelled probes without significantly altering their hybridisation characteristics. Both biotin- and digoxygenin-labelled probes are also stable for years.

The size of the probe is critical for in situ hybridisation and should be between 100 and 500 nucleotides, a size that maximises specific hybridisation as it is ideal for the formation of the probe networks required to efficiently detect unique sequences, whilst decreasing background hybridisation (Lawrence and Singer 1985). The most convenient method for introducing label into a double-stranded probe is nick translation (Rigby et al 1977), because probes of the desired size range can be obtained readily by adjusting the DNAase concentration. DNA probes can also be prepared using short random DNA primers and the Klenow fragment of E. Coli DNA pol (Feinberg and Vogelstein 1983). However, this tends to yield probes of a shorter length, limiting the formation of probe networks (Kamakari 1994) although suitable for the detection of tandemly repeated probes. In the case of probes labelled by chemical modification, sonication can be used to achieve fragments of the required size.

3.1.3.3 Hybridisation

Labelled probes are mixed in a hybridisation buffer containing formamide, salt, dextran sulphate and unlabelled carrier DNA (e.g salmon sperm DNA) or E.Coli t-RNA. Early studies used high salt buffers (5xSSC) and elevated temperatures (60-68°C) to promote accurate base-pairing and therefore efficient hybridisation. Such conditions often adversely affected the cytology, making chromosome identification difficult after hybridisation. The presence of formamide and moderate salt concentrations in the hybridisation mixture lowers the temperature required for accurate nucleic acid annealing, permitting hybridisation at temperatures of 37-45°C (McConnaughy et al 1969) which helps to preserve chromosome morphology.

Dextran sulphate increases the rate of reassociation of DNA as described in section 3.1.1, whilst unlabelled carrier DNA or E. Coli tRNA are included to reduce non-specific binding of the probe to chromatin and glass.

For repetitive probes, the mixture is denatured and applied directly to slides after chromosome denaturation, or is directly applied to slides and denatured simultaneously with the target DNA. For unique sequence contained in large-insert probes, the probe is preannealed with excess unlabelled genomic or Cot1 DNA to reduce repetitive sequence binding to the target before applying to the slides. Hybridisation between target and probe occurs during overnight incubation (16-18 hrs) of the slides at 37°C.

3.1.3.4 Post-hybridisation washes

After hybridisation, excess unhybridised probe must be removed by extensive washing of the slides. In the case of DNA-DNA hybrids, non-specifically bound DNA probe can be removed by washes in salt solutions of varying ionic strengths, or by incubation in low salt at high temperatures. Temperatures of between 55-70°C are well below the melting point of 50% GC content DNA and will ensure fidelity of the hybrids of interest whilst providing a
stringency sufficient to remove non-specifically bound probes. However, incubations at high temperatures for prolonged periods destroys the architecture of the chromosomes. Standard post-hybridisation washes are thus performed with 50% formamide in 2xSSC followed by 2xSSC at a temperature 5-7°C higher than the hybridisation temperature. The number and duration of washes may vary according to the probe size.

3.1.3.5 Fluorescent detection

Hybridised probes are detected by incubating the slides in immunofluorescent reagents to produce a fluorescent signal at the hybridisation site. The extraordinary affinity of biotin for the glycoprotein avidin (Langer et al. 1981) is frequently exploited for the detection of biotin-labelled probes by using fluorochrome-conjugated avidin. Anti-biotin antibodies can also be used. To overcome the potential problems associated with the endogenous presence of biotin in cytological material which may cause ‘autofluorescence’ and contribute to undesirable background signals, detection of biotin-labelled probes is performed in a ‘blocking buffer’ containing high salt concentration (typically 4 x SSC) which reduces the non-specific binding of biotin affinity reagents (Lawrence et al. 1990). Probes labelled with digoxygenin, DNP, AAF or sulphonate are detected with specific antibodies followed by fluorescently labelled anti-immunoglobulins.

A variety of fluorescent labels are available for use in FISH, differing in their absorption and emission spectra, that is, the range of light wavelengths required for excitation and those re-emitted by that substance. The most commonly used fluorochromes are fluorescein isothiocyanate (FITC) which is excited by blue light of 495nm and emits in the green-yellow region at 515nm, rhodamine and Texas Red which are excited at 550nm and 595nm and both emit in the red region at 575nm and 615nm, respectively, and 7-amino-4-methyl coumarin-3-acetic acid (AMCA) which is excited by UV light (350nm) and emits in the blue region at 450nm. The sensitivity of the fluorescence detection can be increased using multiple layers of affinity reagents or antibodies e.g. for biotinylated probes, avidin may be used for primary detection, followed by alternating layers of anti-avidin antibody and avidin (sandwich amplification).

3.1.3.6 Chromosome counterstaining and banding

The most commonly used fluorescent dye for chromosome staining when banding is not required is propidium iodide which excites at 340nm and emits in the red region at 600-610nm. Chromosomes can be banded after hybridisation, and although the bands are crude they are sufficient to identify chromosomes and regionally localise probes. The assignment of a sequence to an individual band may also provide information on the general functional significance of that region since light G-bands are early replicating, rich in CpG islands, contain many housekeeping genes, and large numbers of ALU repeats, whilst dark G-bands are late replicating with fewer but tissue-specific genes and more L1 repeats (Bickmore and Sumner 1989). Banding methods that can be applied in parallel to fluorescent probe detection
and allow simultaneous visualisation of bands and fluorescent signal include the use of actinomycin/diamidinophenylindole (DAPI; emission in the blue region) staining to highlight G-bands (where band contrast may be improved by growth in 5-bromodeoxyuridine (BrdU), Hoescht staining and UV irradiation for R-bands, or ALU or LINE repetitive sequences can be added to the hybridisation mix to depict R- or G-bands respectively. Quinacrine (emits in the yellow region) can also be used to generate a Q-banding pattern, as can DAPI used alone.

3.1.3.7 Signal visualisation

Most conventional fluorescence microscopes are suitable for viewing the signals from painted whole chromosomes, stained subchromosomal regions or localised repetitive or single-copy probes, provided that they are fitted with appropriate objectives and filter sets for visualisation of the relevant fluorochromes. Excitation filters select the optimum wavelength of light to excite the fluorochrome, whilst barrier filters both suppress the excess exciting light and select out the emission wavelength of the fluorochrome to render it visible.

A filter set can be chosen specifically for a given fluorochrome (single band-pass filter sets), or for the simultaneous detection of two or more fluorochromes having different wavelength specificities, such as FITC for signal detection and propidium iodide for chromosome staining (double band-pass filter sets), or similarly for viewing three distinct fluorochromes (triple band-pass filter sets). Such multiple band-pass filter sets have greatly simplified multiple colour FISH mapping experiments by minimising the image shift that can occur when filters are changed. The consequent displaced registration of images detected on multiple exposure of photographic films can introduce errors where analysis involves fine measurements or localisation of probes that map very close to one another.

Images of hybridised cells can be collected with high resolution on high-speed colour film or into computer memory via a static CCD (charge coupled device) camera or a confocal laser scanning microscope equipped with a photomultiplier. CCD cameras and confocal laser scanning microscopes allow the generation of digitised images and thus facilitate the collection and analysis of hybridisation data by enhancing contrast and improving the signal-to-noise ratio by averaging images to allow detection of small signals. The former system is the most sensitive to date, by virtue of its high efficiency in counting emitted photons over a broad spectral range of wavelengths, enabling the recording of signals which are not visible to the observer’s eye. The second system is especially suitable for three-dimensional microscopy, where a series of optical sections through a labelled specimen is obtained. 3D reconstruction is accomplished by applying appropriate computer software to a stack of digitised images. Laser scanning confocal microscopy is designed to eliminate most of the out-of-focus fluorescence obtained in each section by conventional image devices. The generation of high quality optical sections, greatly reduces the amount of mathematical operations required to reconstruct the object.
3.1.4 FISH as a mapping tool in this study

At the outset of this project, although some linkage maps of the X chromosome were reported (NIH/CEPH collaborative mapping group 1992; Weissenbach et al 1992), their overall resolution was insufficient to sublocalise genes of interest within a small enough region to permit their isolation. In particular, large gaps in these genetic maps existed in certain regions such as proximal Xp, which had significantly impeded progress in both the initial characterisation of XLRP families as RP2 or RP3, and also in refining the genetic intervals containing the defective genes. This chapter describes how X chromosome-specific cosmids known to contain poly CA repeats were sublocalised to discrete regions of the X chromosome by FISH to human metaphase spreads, in order to identify those which mapped to the RP2 region (Xp11.3-Xp11.22; section 1.9.5.3.2). These would then be potentially useful as novel genetic markers with which to further refine the interval containing this gene.

The cosmids were derived from an X-chromosome specific somatic cell hybrid library (Lindsay and Bird 1987). The insert DNA was partially digested with the enzyme MboI and cloned into the BamHI site of cosmid vector pJBS to give an average insert size of 40kb. The cosmids used in this study had previously been shown to be both human and X-specific by dosage hybridisation to a panel consisting of 4X, XX, XY and mouse DNAs (data not shown). All cosmids included in this study were also known to contain microsatellite sequences by virtue of a positive signal upon hybridisation with a poly CA probe. Preliminary investigations using a limited panel of X chromosome somatic cell hybrids to sublocalise such cosmids proved unreliable with frequent contradictory results (data not shown). Detailed characterisation of somatic cell hybrids to determine the extent of their human chromosome content is a crucial but unfortunately timeconsuming task, and it was therefore decided to attempt FISH as an alternative, and ultimately more precise method by which to map the cosmids.

A total of 12 microsatellite-containing cosmids were mapped by FISH to specific regions of the X chromosome, 4 of which sublocalised to Xp11.3-Xp11.22, the RP2 critical region. The location of these cosmids in proximal Xp was subsequently more precisely defined by two-colour FISH with respect to cosmids containing reference markers DXS7 and DXS426 in Xp11.3 and Xp11.23 respectively (Kamakari et al 1995). Other cosmids mapping close to, but not necessarily within this critical region would also be worthy of further investigation as invaluable sources from which to generate novel genetic markers, both to facilitate the classification of families as RP2 or RP3, and to confirm the localisation of the RP2 gene in other XLRP families.
3.2 Materials and Methods

3.2.1 In situ hybridisation of centromeric repeat probe DXZ1

3.2.1.1 Probe description and evaluation

The X chromosome centromeric repeat probe DXZ1 was used initially to test the technique, and subsequently in conjunction with each cosmid undergoing FISH, to specifically highlight the X chromosome, and as a positive control. Repetitive elements that are repeated 100-5000 times on specific chromosome types, when labelled and hybridised at sufficient stringency, produce intense and compact zones of hybridisation near centromeres or in heterochromatic regions of the relevant chromosomes (Willard and Waye 1987). The highly chromosome specific organisation of such locus specific repeated sequences, notably the alphoid centromere repeats, renders them useful for tagging a particular chromosome of interest where banding is not required. Specificity of centromeric tandem arrays of alpha satellite DNA monomers for individual human chromosomes is characterised by the length of the higher order repeat unit (a highly ordered hierarchy of multimeric repeat lengths is superimposed on each array of monomers), its copy number (the size of the overall array), the primary nucleotide sequence of the higher repeat unit, and restriction enzymes used to visualise the higher-order repeat on Southern blots. DXZ1 is a 2kb BamHI X-specific centromeric probe defining the higher repeat unit of the X chromosome alpha satellite (Mahtani and Willard 1988). It was obtained from the American Type Culture Collection (ATCC). DNA was prepared using the alkaline lysis midiprep method (section 2.2.3) and resuspended in 50μl TE. 2μg DNA was digested with Bam HI (NBL) in a 20μl reaction at 37°C for 2 hours as described in section 2.3. The sample was electrophoresed on a 0.8% agarose gel to check DNA quality and insert size prior to labelling.

3.2.1.2 Experimental details

The probe was labelled with biotin-16-dUTP by nick translation (section 2.1.2). Probe purification and the biotin incorporation test have been described in sections 2.1.3 and 2.1.4 respectively.

Along with the DXZ1 hybridisation mix described in section 2.1.5.2, a “no DNA” control hybridisation cocktail was prepared, to the same specification but without probe DNA. 40μl of each mixture was applied per slide (probe concentration 40ng/slide for test slide). Probe and chromosome denaturation were initially performed simultaneously by transferring the slide to a plastic tray floating in a waterbath at 75°C for 10 minutes. Hybridisation was then carried out at 37°C for 16 hours. However, this method of denaturing both probe and chromosomes together did not give satisfactory results (data not shown) and subsequently DXZ1 was denatured separately before application to denatured slides as for cosmid probes.
Post hybridisation washes, signal detection and signal visualisation have been described in sections 2.1.6, 2.1.7 and 2.1.8 respectively.

### 3.2.2 In Situ Hybridisation of cosmid clones

DNA from all cosmids used in the study was prepared by the alkaline lysis midiprep method (section 2.2.3). Before in situ hybridisation, a 2μl aliquot was digested with Hind III (4U) for 2-4 hours at 37°C and electrophoresed along with an undigested aliquot and an appropriate size standard (λ/Hind III) on a 0.8% agarose gel to check the quality of preparation, estimate concentration and confirm that the cosmid had the expected HindIII digest pattern (section 2.4.1; data not shown). Cosmid HX74 was selected as a suitable control for FISH of cosmid clones as its location on the X chromosome was already known. PCR studies had previously shown that this cosmid contains the microsatellite marker DXS426 which maps to Xp11.23 (data not shown).

The 13 cosmids included in this study were HX2, HX3, HX4, HX8, HX20, HX39, HX43, HX64, HX72, HX80, HX89, HX91 and HX98. All were labelled with biotin-16-dUTP and detected with FITC-conjugates as outlined in section 2.1. Labelled DXZ1 was used as a positive control in the biotin incorporation test for each cosmid. Cosmids HX80 and HX89 were hybridised using the modified hybridisation mix as described in section 2.1.5.3. Two metaphase spread slides were routinely hybridised with each cosmid, one of which was cohybridised with the cosmid and DXZ1, at least 20 metaphase spreads were analysed per cosmid and a number of them were recorded using conventional photography. In each experiment, one slide was hybridised with the X-centromeric probe DXZ1 alone, as a positive control and to highlight the X chromosome, whilst a “no DNA” slide was included as a negative control for background hybridisation.

### 3.3 Results

DXZ1 was chosen as a control for the effectiveness of the FISH technique because as an X-specific alpha centromeric probe it would be easily detectable and also identify the chromosome of interest. Figure 3.1 shows the results obtained with DXZ1 using sandwich amplification with two (a) or three (b) layers of antibodies in the detection system. Although 2 layers produces a cleaner signal from the centromeric probe, with no background, this was not sufficient to generate a discernible signal from cosmid HX74 (data not shown), and all subsequent experiments employed three layers of antibodies, to produce a discrete signal from the site of cosmid hybridisation with virtually no background spots. Using this level of signal amplification with DXZ1 as a probe, weaker signals were often seen to mark the centromeres of other chromosomes (as seen in Figures 3.1b and 3.2h). Most metaphases examined showed cosmid signals on both chromatids of both homologous chromosomes with virtually no background spots, indicating that the level of repeat sequence suppression was adequate.

12 of the 13 cosmids chosen for FISH mapping were successfully sublocalised to specific X chromosome regions. Results for cosmids HX4, HX8, HX20, HX39, HX43, HX64, HX80,
Figure 3.1 FISH result with the X-specific alpha centromeric probe DXZ1 on metaphase chromosomes of a normal female (chromosomes are counterstained with propidium iodide and fluoresce red) using sandwich amplification with (a) 2 layers (b) 3 layers of antibodies in the detection system. The probe is labelled with biotin; regions of hybridisation are detected with FITC and appear yellow. With 3 layers of antibodies (b) weaker signals are seen to mark the centromeres of other chromosomes.
Figure 3.2 Sublocalisation of $9X$ specific cosmids to $X$ chromosome regions by metaphase FISH. Cosmids were labelled with biotin and detected with FITC conjugates. Discrete signals at the site of cosmid hybridisation are indicated by arrows.

(a) Cosmid HX4

Cosmid HX4 + DXZ1

(b) Cosmid HX8

Cosmid HX8 + DXZ1
(c) Cosmid HX20

Cosmid HX20 + DXZ1

(d) Cosmid HX39
(e) Cosmid HX43

(f) Cosmid HX64 + DXZ1

(g) Cosmid HX80

Cosmid HX80 + DXZ1
(h) Cosmid HX89 + DXZ1

(i) Cosmid HX98 + DXZ1
HX89, and HX98 are shown in Figure 3.2(a-i), where highly specific signals can be seen on metaphase chromosome spreads. For these cosmids, more than 80% of all metaphases showed signals on both chromatids of both homologous X chromosomes, demonstrating the efficiency of the procedure. Observation of 10 metaphases was usually sufficient to localise the cosmid. Cosmids HX2 and HX72 both appeared to map to the centromeric region, their signals overlapping that of DXZ1. In fact, it was presumed that they contained the X-alphoid repeat sequence, as the signals they generated looked identical to those of DXZ1 alone, with weaker signals observed at the centromere of other chromosomes (as in Figure 3.1b).

Repeating the experiments with DXZ1 in a fresh environment gave identical results. No signal could be obtained with cosmid HX3 despite efficient biotin labelling and clear hybridisation of DXZ1 in the same experiment. Six cosmids, HX4, HX20, HX39, HX43, HX80 and HX91 were found to map to the proximal short arm of the X chromosome. The remaining four cosmids, HX8, HX64, HX89 and HX98 were found to map to the long arm. Localisation was estimated relative to the distance of the signal from the centromere to give a broad map position. Although chromosome banding was not performed, a tentative map position for each cosmid is given in Table 3.1, which in some instances has been confirmed by other methods.

### Table 3.1 Subchromosomal localisation of 12 X-specific cosmids by metaphase FISH

<table>
<thead>
<tr>
<th>HX number</th>
<th>D number</th>
<th>FISH location</th>
<th>Confirmation</th>
</tr>
</thead>
<tbody>
<tr>
<td>HX2</td>
<td>DXS546</td>
<td>centromere</td>
<td></td>
</tr>
<tr>
<td>HX4</td>
<td>DXS549</td>
<td>Xp11.4</td>
<td></td>
</tr>
<tr>
<td>HX8</td>
<td>DXS553</td>
<td>Xq28</td>
<td></td>
</tr>
<tr>
<td>HX20</td>
<td>DXS556</td>
<td>Xp11.4-p11.3</td>
<td></td>
</tr>
<tr>
<td>HX39</td>
<td>DXS977</td>
<td>Xp11.3-p11.22</td>
<td></td>
</tr>
<tr>
<td>HX43</td>
<td>DXS567</td>
<td>Xq22</td>
<td></td>
</tr>
<tr>
<td>HX64</td>
<td>DXS568</td>
<td>centromere</td>
<td></td>
</tr>
<tr>
<td>HX80</td>
<td>DXS574</td>
<td>Xp11.4</td>
<td></td>
</tr>
<tr>
<td>HX89</td>
<td>DXS575</td>
<td>Xq22.1-q21.3</td>
<td></td>
</tr>
<tr>
<td>HX91</td>
<td>DXS574</td>
<td>Xp11.4</td>
<td></td>
</tr>
<tr>
<td>HX98</td>
<td>DXS575</td>
<td>Xq25-q26</td>
<td>Southern analysis (K.E. Davies, personal communication)</td>
</tr>
</tbody>
</table>
3.4 Discussion

Unique sequences, chromosomal subregions, or entire genomes can be specifically highlighted in metaphase or interphase cells by FISH. In this study the technique was used on metaphase spreads to directly sublocalise 12 X-specific cosmid clones to specific regions of the X chromosome. The utility of FISH for mapping closely spaced sequences depends on whether linear order is maintained in target chromosomes and how chromatin packing affects the resolution of hybridisation sites. Cosmids can be routinely mapped by FISH with a resolution of 2-3Mb on metaphase chromosomes. In this study, cosmids could be rapidly mapped to broad subchromosomal regions (equating roughly to chromosome bands) by visually estimating the distance between probe and centromere as a fraction of chromosome length.

The X-specific centromeric repeat probe DXZ1 was used as a positive control for the efficiency of the technique, and to identify the X chromosome as the chromosome of interest. The X chromosome higher order repeat unit detected by DXZ1 is 2kb long, and is comprised of 12 tandemly repeated 171bp alpha satellite monomers (Mahtani and Willard 1988). 5000 copies of the 2kb repeat constitute the centromeric heterochromatin of the human X chromosome, and with such a large target region, DXZ1 is easy to detect by FISH. Although the DXZ1 probe primarily hybridises to the X chromosome, the existence of an evolutionarily related group of monomers within the higher-order repeats of chromosomes 1,11,17 and X (Willard and Waye 1987) may explain the weaker signals sometimes generated from the centromeres of other chromosomes in this study. Inspection of metaphases where this phenomenon is evident (e.g Figure 3.2h, 3.1b) would indeed indicate the involvement of chromosomes 11 and 17 purely on the basis of size. Where present, these minor hybridisation signals on additional chromosomes did not interfere with the analysis as they were of much weaker intensity than the X signal.

The intensity of hybridisation signal detected depends both on the specific activity of the probe and on how many of the target sequences have bound the probe. It is therefore useful to use the probe at a concentration that will nearly saturate the target DNA. Higher concentrations of probe will only contribute to the background (often from the formation of excessively large, insoluble probe networks) without improving the signal. The probe concentrations used here (1ng/μl DXZ1, 2-3ng/ml cosmid) had been carefully evaluated (by S. Kamakari) prior to the work described here, and fulfilled the necessary criteria of strong and discrete signal production in the absence of significant background hybridisation (e.g. Figures 3.1a, 3.2a, 3.2d).

Biotin was chosen to label the probes, as the detection system for biotinylated probes is indirect (sandwich amplification), consisting of five steps involving alternating layers of antibodies conjugated with a fluorochrome (avidin-FITC) and antibodies conjugated with a reporter group (biotinylated anti-avidin). This system provides the opportunity to be able to amplify the signal intensity from biotinylated probes by increasing the number of layers, and is therefore advantageous for the detection of small or weakly hybridising probes. Initial experiments using cosmid HX74, whose map position was known, indicated that three layers of
antibodies would provide sufficient signal amplification to specifically detect hybridised cosmids against a relatively low level of background hybridisation, as demonstrated by results obtained with subsequent cosmids in Figure 3.2. Competition was performed with total human DNA to remove repetitive sequences from cosmid probes (CISS; Lichter et al 1991), eliminating the need to identify and isolate single-copy subclones from each large genomic DNA fragment. The presence of interspersed repetitive elements (IRS e.g. SINES and LINES) in the probe DNA would lead to hybridisation signals distributed over the whole chromosome complement, due to the ubiquitous presence of IRSs throughout the genome. Such an increase in background would have made it difficult to detect the signal arising from the single-copy portion of the cosmid.

Signals from the cosmids were visualised using conventional fluorescence microscopy and a filter set which allowed the simultaneous visualisation of FITC for signal detection and propidium iodide for chromosome staining. The map position for all but one cosmid could be roughly assigned by visual inspection, by estimating the localisation relative to the distance from the centromere to give a broad map position. Multiple chromosomes from each hybridisation reaction were analysed to define the locus more precisely. The localisation of 3 of the cosmids (HX20, HX43, HX91) has been confirmed genetically by multipoint linkage analysis or haplotype analysis using other polymorphic markers in the region (Thiselton et al 1995, Kamakari et al 1995). Cytological banding was not performed since fine mapping was desired only for cosmids mapping in the proximal short arm of the X chromosome. This was achieved using two colour FISH (Kamakari et al 1995).

The reasons for the lack of signal from cosmid HX3 are unclear. This probe appeared to label to high efficiency as evidenced by the biotin incorporation test, and a signal from DXZ1 upon cohybridisation with HX3 demonstrated that the hybridisation conditions were adequate. It has since been shown that this cosmid contains several poly CA tracts (data not shown), and it is possible that it represents a particularly repetitive portion of the X chromosome (but not the heterochromatin) such that after competing with total human DNA to the stringency used here, the single-copy regions remaining may have been so small as to be undetectable after in situ hybridisation.

Two cosmids, HX2 and HX72, appeared to map to the X centromere. Although it seems likely that both these cosmids contained portions of the DXZ1 alpha-satellite repeat sequence, as outlined in the results, this has not been proven. Furthermore, it is known that local chromatin structure plays a significant role in determining the limits of resolution of closely spaced DNA sequences, especially in heterochromatic regions. This fact, together with the high degree of DNA compaction in metaphase chromosomes, indicates that the possibility of a location distinct from the centromere, yet very close to it, cannot be entirely ruled out for these two cosmids. It would thus appear that unresolved or merged signals (from the cosmid and DXZ1) are the result of chromatin orientation or compaction. The ability to order closely linked sequences on metaphase chromosomes is limited to a few Mb because the chromosome width can encompass substantial lengths of DNA and the linear order can be disrupted.
The variation in the degree of chromosome condensation at metaphase between cells is evident from Figures 3.1a and 3.2f and demonstrates the improvement in precise map positioning when more elongated chromosomes are examined, as they give better spatial resolution. This is the basis of interphase chromatin mapping (section 1.3.1.2), whereby FISH detection of sequences in the less condensed interphase chromatin provides exceptionally high resolution analysis of closely spaced DNA sequences. The average interphase distance has a linear correlation with DNA distances up to at least 1 Mb, and signals from probes separated by as little as 50kb have been distinguished (Trask et al 1989), giving interphase mapping a 10-fold higher resolution for ordering sequences than 2-colour metaphase mapping (50-100kb vs 1Mb). Nevertheless, the latter technique is the simpler method, and interphase mapping is normally only utilised when the probes to be ordered are known to lie less than 1-2 Mb apart. In addition, interphase mapping, although useful for ordering DNA probes, cannot provide their initial subchromosomal localisation, therefore metaphase mapping was accordingly utilised in this study.

The identification of sequence-tagged sites (STSs) and polymorphic loci within a given chromosomal region is an essential resource for constructing physical and genetic maps of the human genome, a necessary tool for the identification of disease genes by positional cloning. The cosmids sublocalised here by FISH may provide potentially polymorphic markers for use in genetic linkage studies involving disease genes mapping to the relevant regions, as well as contributing to the genetic and physical maps of the human X chromosome. Furthermore, this approach has resulted in the successful localisation of 5 cosmids to the region containing the genes implicated in both RP2 and RP3 on proximal Xp (Xp11.22-Xp11.4). Four of these cosmids contain poly CA tracts (HX20, HX39, HX43, HX91) which, if polymorphic, would provide novel genetic markers to more definitively characterise XLRP families and refine the gene-containing intervals. The isolation and characterisation of the associated microsatellite is an essential prerequisite for genetic analysis using these cosmids, and the description of this procedure for one of the cosmids which mapped to the proximal short arm (HX20) is the subject of the next chapter.
CHAPTER 4

Isolation and Characterisation of a Microsatellite from Cosmid HX20

4.1 Introduction

4.1.1 The relevance of microsatellites as genetic markers in this study

In recent years, microsatellites have rapidly become the markers of choice in genetic linkage mapping of disease genes, for several reasons: their considerable degree of polymorphism between individuals (due to variation in tandem repeat length), making them very informative in most pedigrees, their small size (usually no greater than 100bp) means that they can be analysed quickly and easily by PCR (obviating the need for large quantities of template DNA as with the Southern blots necessary for studying most RFLP/VNTR variations), and their ubiquitous distribution in the genome (in humans, a microsatellite occurs on average every 30kb).

Positional cloning approaches to map disease genes on the human X chromosome suffered at the outset of this project from a lack of mapped VNTR loci (Luty et al 1990). As of 1989, the Human Gene Mapping Library listed a total of 567 anonymous DNA segments and/or cloned genes localised to the X chromosome, but only 3 of these (DXS52, DXS255, DXS278) corresponded to multiallelic VNTRs (Mandel et al 1989). Probes from the X chromosome also appeared to reveal RFLPs at a lower frequency than that seen with autosomally derived probes (Hofker et al 1986). In the region on proximal Xp encompassing the XLRP genetic loci (Xp21.1-11.22), there was a particular paucity of highly polymorphic marker loci.

Whatever factors limit the occurrence of classical VNTRs and RFLPs on the X chromosome do not appear to apply to the microsatellite repeats however, which appeared to be just as abundant and widely dispersed on this chromosome than any other (Luty et al 1990). Microsatellites were therefore isolated and characterised from the cosmids mapped to this region by FISH, with the aim of increasing the polymorphic marker density in this region of Xp. Because (CA)n microsatellites appear to be the most abundant class of dispersed simple sequence repeats (section 1.2.3.1), efforts were focused on the characterisation of polymorphisms based on variation involving this class of simple sequence.

This chapter describes the methods employed for one such cosmid, HX20, which mapped by FISH to Xp11.4. A polymorphic marker in this region would be a useful addition to the existing marker map in classifying families as either RP2 or RP3, and, depending on its precise location, in refining the genetic intervals containing either of these genes.
4.1.2 Isolation of microsatellites

Several strategies have been utilised to identify and then convert microsatellites into polymerase chain reaction (PCR) based assays for genotyping. The first set of markers were generated using information recovered from computer-based searches of DNA sequence databases (Weber 1990). Identification of new sets was first accomplished by screening large-insert genomic libraries (phage or cosmid) with the core sequence of the desired repeat (e.g. (CA)$_{20}$). A PCR-based assay was developed by identifying, subcloning, then sequencing a small restriction fragment containing the targeted repeat. This approach is tedious and limited by the difficulty in finding a suitable restriction fragment permitting the design of primers based upon a single-pass sequence.

An alternative approach has been to construct small-insert (200-500bp) genomic libraries in plasmid, M13 or phagemid vectors by ligation with genomic DNA restricted with frequent-cutting enzymes (Sau3AI or a combination of AluI, RsaI and HaeIII), then screen these libraries with the core sequence of the desired repeat in the form of a simple repetitive oligonucleotide probe. Once a clone containing the targeted repeat is identified, DNA sequences suitable for development of primers flanking the microsatellite can be readily obtained from a single sequencing run. This approach was adopted by Weissenbach et al (1992) in the first genome-wide effort to generate a genetic map based on microsatellite polymorphisms. After sequencing the CA-positive clones from a library of AluI-digested human genomic DNA, STSs were developed from those containing repeats with greater than 12 CAs. A total of 12,014 clones were sequenced, producing 814 polymorphic markers, 97% of which showed heterozygosities above 0.5 and 74.4% above 0.7. This "shotgun cloning" method was employed for the isolation of the microsatellite from cosmid HX20.

4.1.3 Strategy in summary

Briefly, a minilibrary was generated from cosmid HX20 by ligation of a sequencing vector (pT7T3) to Sau3AI-restricted HX20 DNA. The library was screened for a subclone containing the microsatellite by hybridisation with a simple oligonucleotide probe (poly dG-dT), and two positive clones were sequenced. PCR primers flanking the CA repeat in one of the clones were used to amplify genomic DNA from 36 unrelated Caucasian females, in order to establish its heterozygosity, and hence its usefulness as a genetic marker.

4.2 Materials and methods

4.2.1 Construction of a minilibrary from cosmid HX20

Restriction digestion of DNA from cosmid HX20 and the plasmid vector pT7T3 into which the resulting fragments were to be cloned, and DNA purification prior to ligation are described in section 2.8.1.1. Ligation of the cosmid fragments to pT7T3 is described in section 2.8.1.2. Three separate ligation reactions were set up with different ratios of vector:insert DNA (1:2, 1:1, 2:1) based on the number of molecules of each, with 200ng total
DNA. 1µl of the reaction was used as a control before adding 1µl (0.1 unit) of T4 DNA ligase
to the remaining 9µl reaction. Control ligations were also performed containing the following
DNA samples as the only DNA present in the reaction: (1) 100ng vector pT7T3/BamHI
(positive control) (2) 100ng dephosphorylated HX20/Sau3AI insert DNA (negative control).
(3) 100ng λHind III (positive control).

4.2.2 Isolation of HX20 subclones containing a CA repeat

Preparation and transformation of competent bacterial cells with the ligated products are
described in sections 2.8.2.1 and 2.8.2.2 respectively. Control transformations were set up
using (1) No DNA (2µl SDW, negative control) (2) 1ng self-ligated pT7T3 (positive control).
Following transformation and growth of the plated out cells, replica filters were prepared
(section 2.8.2.3) for experimental and control plates (self-ligated vector) for subsequent
hybridisation with a CA repeat probe to identify CA-positive transformants. The inclusion of
the replica of the control plate would here serve as a negative control for the hybridisation.
20pmol oligonucleotide (GT)_{11}G was end-labelled with γ^{32}P-ATP as described in section
2.5.4.1. The replica filters were hybridised with the labelled probe (section 2.5.4.2) together
with a filter containing DNA from cosmids known to harbour CA repeats, as a positive control.
After post-hybridisation washing to high stringency (section 2.5.4.3), the filters were
autoradiographed as described at -80°C with signal intensification.

Following exposure of the autoradiograph, positive signals were matched to the colonies
on the master plate with the aid of the orientation holes, thus identifying potential (CA)_{n}
positive clones. These colonies were picked and DNA prepared using the alkaline lysis DNA
miniprep method (section 2.2.2). 2µl of each preparation was electrophoresed on a 0.8%
agarose gel (section 2.4.1), the gel blotted (section 2.5.2), and the filter hybridised with the
(GT)_{11}G probe as before to identify true positive clones. These clones were then restreaked
from the remains of the culture used to make the minipreps onto LB plates supplemented with
ampicillin, IPTG and Xgal (section 2.8.2.2), and greater quantities of DNA prepared by the
alkaline lysis midiprep method (section 2.2.3).

4.2.3 Sequencing the CA-positive HX20 subclones

Before sequencing, the insert size of each subclone was determined by restriction
enzyme digestion with HindIII/EcoRI (section 2.3), HindIII to linearise the molecule, and
EcoRI to release the insert. The products of digestion were analysed by electrophoresing 2µl
on a 0.8% agarose gel (section 2.4.1). A Southern blot of this gel was hybridised with the
(GT)_{11}G probe (section 2.8.2.2) to confirm the presence of a CA repeat in the insert
fragments.

The (CA)_{n}-positive subclones were sequenced from the T3 and T7 vector primers (i.e.
both ends of the insert) according to the sequencing protocol described in section 2.6.2, and
subjected to polyacrylamide gel electrophoresis (section 2.4.2).
4.2.4 Microsatellite PCR assay and assessment of polymorphism

PCR amplification of the pHX20-2b microsatellite locus was first optimised in trial reactions using non-radioactive reagents (section 2.7.2). Products were analysed on 3% agarose gels (section 2.4.1) to ensure that efficient production of a fragment of predicted length had been obtained.

To assess the level of polymorphism of the HX20 microsatellite in the general population, PCR was performed radioactively by using γ-32P-ATP-endlabelled T3 primer on 36 unrelated Caucasian females (72 X chromosomes), and PCR products resolved on polyacrylamide gels (section 2.7.4). To estimate the size of the alleles, PCR reactions were also performed on the recombinant pHX20-2b template that had been sequenced, and the amplification product was loaded on the gel as a DNA size marker.

4.3 Results

4.3.1 Construction of a minilibrary from cosmid HX20

4.3.1.1 Vector and cosmid restriction digests and DNA purification

Figure 4.1 shows aliquots of cosmid HX20 digested with Sau3AI and an aliquot of vector pT7T3 digested with BamHl, after purification, electrophoresed on a 0.8% agarose gel alongside λ HindIII size marker. Digested vector had the expected linear size of 2.9kb, and cosmid HX20 was digested to an average fragment size of 800-900bp. Concentrations were estimated at 50ng/μl for HX20/Sau3AI and 200ng/μl for pT7T3/BamH1.

4.3.1.2 Ligation of dephosphorylated cosmid restriction fragments to plasmid vector

Figure 4.2 shows successful ligation of insert to vector, indicated by an increase in the fragment size of the sample after ligation as compared with an aliquot from before the addition of ligase. The positive controls, vector/BamHl alone and λHindIII, are also seen to have ligated, whilst as expected, the negative control HX20/SauIIIa (dephosphorylated) did not ligate to any great degree.

4.3.2 Isolation of the subclone containing the CA repeat

4.3.2.1 Transformation of competent cells with ligated products

Following transformation with the Xgal blue/white system, colonies of bacterial cells containing a non-recombinant plasmid, which had been transformed with the control sample (self-ligated plasmid DNA) would appear blue in colour, due to the conversion of the colourless chromogenic substrate Xgal into a blue product by the E.Coli enzyme β-galactosidase. Colonies of bacterial cells with recombinant plasmids appear white in colour, because the inserted DNA disrupts the plasmid β-galactosidase gene, thus preventing β-galactosidase activity.
Figure 4.1 Electrophoresis on a 0.8% agarose gel of cosmid HX20 and vector pT7T3 DNAs digested with SauIIIa and BamHI respectively.

Figure 4.2 Electrophoresis on a 0.8% agarose gel showing successful ligation of insert (HX20/SauIIIa) to vector (pT7T3/BamHI). Lanes 1, 3, 5: Unligated vector and insert DNA in ratios 1:2, 1:1, 2:1 (vector:insert) respectively. The insert DNA is not visible due to low quantity of loading. Lanes 2, 4, 6: The corresponding ligated products. Positive controls before and after ligation: Lanes 7 and 8, vector/BamHI alone; Lanes 9 and 10, λHindIII. Lanes 11 and 12: Negative control HX20/SauIIIa (dephosphorylated) before and after ligation.
The efficiency of transformation was therefore assessed by observing the ratio of white:blue bacterial colonies on the plates. The plate with the 2:1 ratio of vector:insert ligation had ~2000 bacterial colonies, almost half of which were white, and this was carried through as the experimental plate. The plate containing cells transformed with self-ligated vector (positive control for transformation) had ~3000 colonies (all blue) and as expected, no colonies were observed in the 'No DNA' control plate. Few colonies were observed on the 1:2 and 1:1 vector:insert plates, with no white colonies, the reasons for which were unclear.

4.3.2.2 Hybridisation screening with (GT)$_{11}$G probe

Positive signals were observed after autoradiography in the experimental plate and the positive control filter (Figure 4.3). No signals were observed in the negative control. Following a secondary screen of positive colony DNA preps with the (GT)$_{11}$G probe (section 4.2.2), two positive colonies were observed corresponding to DNA from subclones 2b and 3b (data not shown).

4.3.3 Characterisation of microsatellites

4.3.3.1 Sequencing subclones HX20-2b and HX20-3b

From Figure 4.4a the 2.9kb vector band is evident in the HindIII/EcoRI double digests of both subclones 2b and 3b. The insert of HX20-2b appears to contain an EcoRI site, as two fragments were released from the vector, of sizes 1.8kb and 1kb, giving a total insert size of approximately 2.8kb. The size of the 3b insert was estimated at roughly 300-400bp. The results of hybridising a Southern blot of this gel with the (GT)$_{11}$G probe (Figure 4.4b) clearly that both 2b and 3b contain a CA repeat, and also indicate partial digestion of subclone 3b.

Comparison of the sequence data revealed that sequence of the subclones were different after the BamHI cloning site (residues 217-222 in pT7T3) compared to the vector sequence, confirming presence of an insert in the subclones.

The sequence derived from subclone HX20-3b on the T3-primed strand revealed a stretch of interrupted CA-dinucleotide, $(CA)_n$ TGCACT (CA)$_7$ (Figure 4.5) from residues 66-93 (Figure 4.6a), which was also seen as the mirror (TG)$_7$AGTGCA(TG)$_4$ on the opposite strand sequenced from the T7 primer, thus confirming sequence of the entire insert (Figure 4.5). This dinucleotide repeat was unlikely to be polymorphic (see section 4.4), therefore was not characterised further.

Initial sequence data from the T3-primed strand of subclone HX20-2b (298bp from cloning site) indicated a possible stretch of (GT)$_n$ just beyond this readable sequence. Sequence from the T7 primer revealed no corresponding (CA)$_n$ sequence, therefore primers were designed within the available sequence (T3primer1, residues 254-274, Figure 4.6b; T7primer1, not shown) and used to extend the sequence of HX20-2b. This sequencing step provided T3 strand data (Figure 4.7) between residues 309-577 (Figure 4.6b) containing a (GT)$_{11}$ repeat (residues 315-356). Unfortunately no sequence was obtained using T7primer1, but sufficient data was obtained from T3primer1 to deduce the sequence of the opposite
Figure 4.3 Colony hybridisation with (Gt),G probe. (a) Experimental plate with arrows indicating positives selected (b) positive control (filter with cosmids known to contain CA repeats (c) negative control plate (colonies transformed with self-ligated vector)
Figure 4.4 (a) Agarose gel showing restriction digestion of subclones pHX20-2b and pHX20-3b with HindIII/EcoRI to establish insert size. The 2.9kb vector band is evident in both (lanes 3 and 6).

(b) Autoradiograph showing hybridisation with probe (GT)$_n$G to a Southern blot of the gel shown in (a). CA repeats are clearly present in both HX20 subclones 2b and 3b (this autoradiograph also indicates partial digestion of subclone 3b).
Figure 4.5 Autoradiograph showing the sequence of pHX20-3b subclone from both ends of the insert. The CA repeat is bounded by arrows.
Figure 4.6a: Sequence of the pHX20-3b subclone insert
(sequence range: 1 to 147)

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>20</td>
<td>30</td>
<td>40</td>
<td>50</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>GTCGCCTGCC</td>
<td>TCCCCCACCA</td>
<td>ACGCATTCAT</td>
<td>TCCCTCTCCC</td>
<td>CAGGGACACA</td>
<td>AGAGAGCTTG</td>
<td></td>
</tr>
<tr>
<td>CAGCGGACGG</td>
<td>AGGGGTGGGT</td>
<td>TGGTAAAGTA</td>
<td>AGGGAGAGGG</td>
<td>GTCCCTGTGT</td>
<td>TCTCTGAAC</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>80</td>
<td>90</td>
<td>100</td>
<td>110</td>
<td>120</td>
<td></td>
</tr>
<tr>
<td>CCGTGCAACAC</td>
<td>ACATGCACAC</td>
<td>ACAACACACAC</td>
<td>ACATCCCCCTC</td>
<td>TGCCAGGTCC</td>
<td>AGGCAAAGTC</td>
<td></td>
</tr>
<tr>
<td>CGCACGTTGTG</td>
<td>TGTACGTGA</td>
<td>TGTGTTGTTG</td>
<td>TGAGGGGGAG</td>
<td>ACGGTCAGAG</td>
<td>TCGTTCAG</td>
<td></td>
</tr>
<tr>
<td>130</td>
<td>140</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGCCTCCTCC</td>
<td>CAAGGAACCTT</td>
<td>ATGTAGA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAGGAACCTT</td>
<td>GTTCTGAAC</td>
<td>TACATCT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 4.6b: Sequence of the pHX20-2b subclone insert (from T3 primer)  
(sequence range : 1 to 577)

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>20</td>
<td>30</td>
<td>40</td>
<td>50</td>
<td>60</td>
</tr>
<tr>
<td>AGCCTACTCC</td>
<td>AGAATGAGTA</td>
<td>TTTGCCCTTC</td>
<td>TTTGAAGCCA</td>
<td>GAGAATGATG</td>
<td>TGCTGAGCCA</td>
</tr>
<tr>
<td>TGGATGAGG</td>
<td>TCTTACTCAT</td>
<td>AAACGGAAG</td>
<td>AAACCTCGGT</td>
<td>CTTTACTAC</td>
<td>ACGAGTCGT</td>
</tr>
</tbody>
</table>

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>70</td>
<td>80</td>
<td>90</td>
<td>100</td>
<td>110</td>
<td>120</td>
</tr>
<tr>
<td>CCGGAAACAT</td>
<td>CATTACTTTT</td>
<td>CCAAAATTGG</td>
<td>CGCGATTGTA</td>
<td>ACTTGAGGTG</td>
<td>GTGCCATTGT</td>
</tr>
<tr>
<td>TTACCTTGA</td>
<td>TGAATGAAA</td>
<td>GGGTTTACC</td>
<td>GCGTAAACAT</td>
<td>TGGACTACA</td>
<td>CACGGTAACA</td>
</tr>
</tbody>
</table>

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>130</td>
<td>140</td>
<td>150</td>
<td>160</td>
<td>170</td>
<td>180</td>
</tr>
<tr>
<td>TGCTGATGTG</td>
<td>AGAAGGAGCN</td>
<td>CTTGACATTT</td>
<td>TTTGGCTTG</td>
<td>ATTTCAAGTN</td>
<td>NNCTCATGCA</td>
</tr>
<tr>
<td>AGAATCACA</td>
<td>TCTTCTCGTN</td>
<td>GGAACCTGTA</td>
<td>AAAAAAGGAG</td>
<td>CAGAATAAAA</td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>190</td>
<td>200</td>
<td>210</td>
<td>220</td>
<td>230</td>
<td>240</td>
</tr>
<tr>
<td>AGAGTACTAG</td>
<td>CTCTCTTACA</td>
<td>CCAGACATTA</td>
<td>TANNTTNCA</td>
<td>GAAAAAGGATG</td>
<td>GAGAATAAAA</td>
</tr>
<tr>
<td>TCTCATGATC</td>
<td>GAGAGATGT</td>
<td>GGTCTCGAT</td>
<td>ATNGNAANGT</td>
<td>CTTTCTCTAC</td>
<td>CTCTTATTTT</td>
</tr>
</tbody>
</table>

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>260</td>
<td>280</td>
<td>290</td>
<td>300</td>
<td></td>
</tr>
<tr>
<td>T3Primer1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGAGGAGGAGC</td>
<td>CGTGATCGAG</td>
<td>ACACCCCAAC</td>
<td>TACAAAAAC</td>
<td>CAGGCCCAG</td>
<td></td>
</tr>
<tr>
<td>TCTTCTCCCG</td>
<td>GGCATACCTC</td>
<td>TGTCCGGTGA</td>
<td>ATCTTTTAGC</td>
<td>GTCCCGGCG</td>
<td></td>
</tr>
<tr>
<td>AACGTCACTC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>310</td>
<td>320</td>
<td>330</td>
<td>340</td>
<td>350</td>
<td>360</td>
</tr>
<tr>
<td>AGCAAGGCCCA</td>
<td>ATATTGTGTG</td>
<td>TGTGTGTTG</td>
<td>TGTGTGTTG</td>
<td>TGTGTGTTG</td>
<td>TGTGTGTTT</td>
</tr>
<tr>
<td>TGGTTCCGTT</td>
<td>TATATAACAC</td>
<td>ACACCAACAC</td>
<td>ACACCAACAC</td>
<td>ACACCAACAC</td>
<td>ACACACAAAA</td>
</tr>
</tbody>
</table>

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>370</td>
<td>380</td>
<td>390</td>
<td>400</td>
<td>410</td>
<td>420</td>
</tr>
<tr>
<td>ATGGCGAAAC</td>
<td>TGTATCTAGG</td>
<td>GATGCTAAAT</td>
<td>AAAAAAGATACC</td>
<td>AATTATATAA</td>
<td>TATAGTTTGT</td>
</tr>
<tr>
<td>TACCGTTTGG</td>
<td>ACACAGTTCC</td>
<td>CTACGATTGA</td>
<td>TTCTTATGG</td>
<td>TAAATATTT</td>
<td>ATATACAAAA</td>
</tr>
</tbody>
</table>

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>430</td>
<td>440</td>
<td>450</td>
<td>460</td>
<td>470</td>
<td>480</td>
</tr>
<tr>
<td>T3Primer2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAACGAGGGCC</td>
<td>CTCAAACTGT</td>
<td>TTTAAGATAG</td>
<td>TATTTCCGGTA</td>
<td>GCCATATNCT</td>
<td>TGCCATGTC</td>
</tr>
<tr>
<td>TGGTCTCCGG</td>
<td>GAGTTTGGCA</td>
<td>AATTCTTACT</td>
<td>ATAAAGCCAT</td>
<td>CGTGATANGA</td>
<td>GACGTACCAG</td>
</tr>
</tbody>
</table>

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>490</td>
<td>500</td>
<td>510</td>
<td>520</td>
<td>530</td>
<td>540</td>
</tr>
<tr>
<td>ATCATATATGT</td>
<td>ATATGTGATA</td>
<td>TGGGTGTTTG</td>
<td>TGGCAGACACA</td>
<td>CGTGCAGCTA</td>
<td>TGTCATATA</td>
</tr>
<tr>
<td>TGTATACCA</td>
<td>TATACACTAT</td>
<td>ACCCACAAC</td>
<td>ACGCGTGTTG</td>
<td>GCACGGCCTA</td>
<td>ACAGTAGTAT</td>
</tr>
</tbody>
</table>

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>550</td>
<td>560</td>
<td>570</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCCTCTCATG</td>
<td>GTTCAGAGGT</td>
<td>GAAGAGACAT</td>
<td>TCATACA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGGAGACGTA</td>
<td>CAAGTCTCCA</td>
<td>CTTTCTGTA</td>
<td>AGTATGT</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

T3Primer2
4.4.3.2. PCR assay and characterization of a microsatellite.

The PCR assay was conducted using the following primers: 5' ATTG TACG TTA GAA GGT AAG CAG TGT GTT CTG TAA AGG TGCT TAG CAA GTT GCT TAA AGG CAG TGT GGT AAG CAG TGT GTT CTG TAA AGG TGCT TAG CAA GTT GCT TAA G, and 3' TGT GTT CTG TAA AGG TGCT TAG CAA GTT GCT TAA G. The reaction was carried out in a 50 µL reaction mixture containing 1× PCR buffer, 0.5 mM dNTPs, 2 mM MgCl₂, 0.05 U Taq polymerase, and 0.2 µM of each primer. The reaction was denatured at 94°C for 5 min, followed by 30 cycles of 94°C for 1 min, 56°C for 1 min, and 72°C for 1 min, and a final extension at 72°C for 10 min.

A microsatellite within the PCR product was revealed in the autoradiograph (Figure 4.7), and the repeat unit was found to be 36 nucleotides long. The size of the repeat unit was measured as 36 nucleotides long. The microsatellite showed a perfect (GT)₃₆ repeat.

The readable sequence obtained beyond the dinucleotide stretch was used to design a ‘reverse’ primer for PCR amplification of the microsatellite.

Table 4.1: Frequencies of HX20-2b Alleles

<table>
<thead>
<tr>
<th>Allele</th>
<th>Observed</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>A2</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>A3</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>A4</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>A5</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td>A6</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td>A7</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>A8</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>A9</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>A10</td>
<td>0.05</td>
<td></td>
</tr>
</tbody>
</table>

Figure 4.7: Autoradiograph showing the sequence obtained from subclone HX20-2b using T3primer1. The microsatellite is seen to be a perfect (GT)₃₆ repeat. The readable sequence obtained beyond the dinucleotide stretch was used to design a ‘reverse’ primer for PCR amplification of the microsatellite.
strand beyond the GT repeat and thus design a second primer (T3Primer 2, residues 418-438, Figure 4.6b) with which to amplify across the repeat by PCR.

Database homology searches (FASTA) were performed on insert sequences as they were generated to assist in suitable primer design.

### 4.3.3.2 PCR assay and polymorphism of the (GT)$_{21}$ microsatellite

The flanking primers used to amplify DNA containing the microsatellite are as follows:

- pHX20/T3 Primer1 - 5' TAT GAA GAG AGC CAA CTT AGA 3' (GT strand) and pHX20/T3 primer2 - 5' AGT TTG AGG GCT TCG TTT AC 3' (CA strand) as shown in Figure 4.6b (EMBL accession number X67603). Optimal PCR conditions were found to be 30 cycles of 94°C (1 minute), 57°C (1 minute), 72°C (1 minute) followed by a final extension of 72°C for 10 minutes, as demonstrated non-radioactively (section 4.2.4) in Figure 4.8. The size of the amplification product was 184bp in cosmid HX20.

Autoradiography of radioactive PCR products from the population panel revealed 8 different allelic fragments (as shown in Figure 4.9), the sizes and frequencies of which are shown in Table 4.1. These frequencies were used to calculate a PIC of 0.72 (see section 1.2.4). The observed heterozygosity was 0.73. X-linked codominant segregation was observed in 5 informative families, an example of which is shown in Figure 4.10.

<table>
<thead>
<tr>
<th>Allele</th>
<th>Size (bp)</th>
<th>No. observed</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>192</td>
<td>3</td>
<td>0.04</td>
</tr>
<tr>
<td>A2</td>
<td>190</td>
<td>4</td>
<td>0.06</td>
</tr>
<tr>
<td>A3</td>
<td>188</td>
<td>5</td>
<td>0.07</td>
</tr>
<tr>
<td>A4</td>
<td>186</td>
<td>9</td>
<td>0.13</td>
</tr>
<tr>
<td>A5</td>
<td>184</td>
<td>27</td>
<td>0.37</td>
</tr>
<tr>
<td>A6</td>
<td>182</td>
<td>3</td>
<td>0.04</td>
</tr>
<tr>
<td>A7</td>
<td>178</td>
<td>1</td>
<td>0.01</td>
</tr>
<tr>
<td>A8</td>
<td>176</td>
<td>20</td>
<td>0.29</td>
</tr>
</tbody>
</table>

### 4.4 Discussion

The results here demonstrate the successful isolation of a highly polymorphic, perfect repeat (CA)$_{21}$ microsatellite (DXS556) from a cosmid template. It was Weber (1990) who classified poly CA repeats as either perfect, imperfect or compound, depending on their sequence, as a means of predicting their likely heterozygosity. Perfect repeats are defined as alternating, tandem CA repeats without interruption and without adjacent repeats of another sequence. On searching GenBank, Weber (1990) found that perfect repeats were the most common type of CA microsatellite (72%), with few greater than 24 repeat units in length. Of
Figure 4.8 Agarose gel demonstrating efficient PCR amplification with HX20 primers from human genomic DNA (Lanes 6 and 7). Cosmid HX20 (lane 2) and subclone 2b (lane 1) are positive controls, cosmids HX65 (lane 3), HX82 (lane 4), vector pT7T3 (lane 5) and ‘no DNA’ (lane 8) are negative controls. The amplified product is ~180 bp on comparison with size marker φX174/HaeIII.

Figure 4.9 Autoradiograph of polyacrylamide gel showing 8 alleles of marker HX20 detected after PCR amplification in 36 unrelated females e.g. Lane 1: A1/A8, Lane 41: A2/A5, Lane 35: A3/A8, Lane 37: A4/A5, Lane 38: A5/A6, Lane 13: A7/A7. Allele sizes were deduced by comparison with the amplification product from subclone 2b (184 bp: allele A5) in lane 21. Lanes 10, 19, 20 and 27 were not loaded.
Figure 4.10 Autoradiograph of polyacrylamide gel showing X-linked codominant segregation of the HX20 (CA)$_{21}$ microsatellite in XLRP family F47.
an assumed total of 35000 CA repeats in the haploid genome with 12 or more repeat units, he calculated that there are about 7000 with a PIC of 0.70 or greater. The hybridisation conditions used here are obviously sufficiently stringent to pick out these rarer, longer repeats, but also the more common shorter repeats which are unlikely to be useful as genetic markers. This is advantageous when small numbers of clones are being screened for potential polymorphic markers in a defined chromosomal region as no potentially useful marker will be missed, but for large-scale screening more stringent hybridisation procedures should be used, such as using a probe with longer repeat length e.g. \((CA)_n\) and/or increasing the temperature of hybridisation (Iizuka et al 1993).

The HX20-3b microsatellite fulfills the criteria for a single, imperfect repeat sequence (Weber 1990). However, interruptions tend to reduce the informativeness to the value expected for the longest uninterrupted part of such motifs (in this case, \((CA)_n\)) and as perfect repeats of less than 12 CAs are unlikely to be polymorphic, it was considered unworthwhile to pursue this as a genetic marker, although its conversion to an STS may facilitate physical mapping projects in Xp11.4.

Several different mutation mechanisms have been proposed for tandem simple sequence repeats, but almost all invoke the process of strand slippage during replication, repair or recombination (Schottlerer and Tautz 1992). Mutation of human CAs via strand slippage is consistent with the observed dependence of marker informativeness on numbers of repeats (apparently, the rate of strand slippage for CA/GT sequences jumps as the number of repeats exceeds about 10). The lower informativeness of repeat sequences containing imperfections may be due to the fact that strand slippage of these sequences produces structures with noncomplementary bases.

In order to design unique PCR primers for efficient genotyping of the \((CA)_n\) microsatellite, a database homology search (FASTA; Pearson and Lipman 1988) was performed on the DNA sequence flanking the repeat. Particular attention was paid to the presence of any ALU sequence, as has been observed for several tandem repetitive sequences (Zuliani and Hobbs 1990), and identity with other microsatellites. No significant homology was observed to any sequences in GenBank, thus no there were no specific constraints on primer design. For this microsatellite, there was also no evidence of any direct repeats flanking the CA-stretch which might indicate that it had arisen by insertion, as for some retrotransposons (Hamada et al 1984).

For most microsatellite markers, major alleles are clustered in size (Figure 4.9), within about 6bp on either side of the predominant allele. Primers were designed so that the amplified fragments would be small enough (within 100-400bp) so that alleles differing in size by only 2 bases would be easily resolved on polyacrylamide gels. Scoring of allelic fragments is often made easier by initially optimising the PCR conditions non-radioactively, as was done here, especially with respect to annealing temperature; the presence of spurious bands due to cross hybridisation to nontarget sequences can severely hamper interpretation.

End-labelling one of the PCR primers, followed by autoradiography, was chosen here as a means to detect the allele system after amplification and resolution. Although internal
labelling can be performed simply by adding a labelled nucleotide to the PCR, the difference in migration of the CA and GT strands can lead to confusion when trying to decipher heterozygotes from homozygotes. When endlabelling is employed, each allelic band is associated with one or more less intense, smaller bands (Figure 4.9 and 4.10). Since only one of the 2 strands is labelled, these cannot be attributed to the difference in mobility between the (GT), and (CA), strands. It is thought that they arise by slipped mispairing of Taq Polymerase in the repeat region during PCR (Weber 1990, Litt and Luty 1989). This interpretation is supported by the observation that the extra bands form a ladder with a spacing of 2nt between adjacent members of the series. In practice, these bands do not interfere with allele scoring because they are usually fainter than the major bands. In fact, they can be helpful in comparing allele sizes in different lanes.

Allele frequencies, rather than numbers of different alleles, are intrinsic to calculations of heterozygosity for a polymorphic marker. The frequencies of the most common alleles of a DNA typing system reflect the utility of that assay in practice. (A marker may have 10 alleles, but if one or two predominate up to 95% of the time, the chance of any individual being heterozygous is basically reduced to that of a 2-allele system. Thus the most informative markers for any given allele number are those for which the allele frequencies are equal). The frequency of the major DXS556 allele, A5, is only 0.37, therefore it was expected to be at least reasonably informative. This allele happened to be the one sequenced in the HX20-2b subclone. The number of repeats in the original allele sequences are generally good predictors of marker informativeness, but for any individual marker it is always possible to be misled by the chance cloning and sequencing of one of the shortest or one of the longest alleles.

For the (CA), microsatellite characterised here, the observed heterozygosity of 73% is close to the calculated PIC value 0.72 that would be expected if all 8 alleles were in HWE. Heterozygosity values for polymorphic microsatellites can reach values >80% (Litt and Luty 1989), although average heterozygosity for X chromosome microsatellites is lower (~65%), possibly reflecting a different mutation rate (each X chromosome resides in males only two-thirds as often as each autosome, and most mutations are thought to occur in the male germline) or different population genetic forces (Dib et al 1996; Dietrich et al 1996). With this in mind, the CA repeat isolated here compares favourably in terms of its polymorphic value. As informativeness has been shown to increase with the number of dimeric elements in such a repeat (Weber 1990), it was not surprising that this microsatellite should display such a level of polymorphism. Where deviations from HWE occur, these are reflected in a difference between the observed and expected heterozygosities, due to non-random mating or faulty data interpretation.

Estimation of the degree of marker polymorphism was investigated here on the basis of a panel of unrelated female individuals, because males have only one X chromosome. Depending on the number of chromosomes investigated, however, the true heterozygosity can be quite different from its estimate. A sample of 8-10 individuals may be sufficient for an initial screen of markers to discriminate between those with high and low heterozygosity, as there is a high chance of encountering most alleles of a polymorphic marker (Ott 1992). This
is a useful test when large numbers of markers are being screened with the aim of detecting markers with a high degree of heterozygosity. To characterise markers in more detail, it is more usual to look at the probability of observing all the alleles. Obviously, for rarer alleles of a polymorphic marker, more individuals will be needed to be typed to guarantee finding them. Thus although 72 unrelated chromosomes were typed in this study, we cannot exclude the possibility of unencountered alleles which would alter the allele frequencies and hence the heterozygosity or PIC. Because of the effect of sample size, and in the absence of simple methods for constructing support intervals for heterozygosity values, it is therefore general policy to state the observed heterozygosity in combination with the number of chromosomes analysed when reporting new markers (Thiselton et al 1993).

The ‘shotgun subcloning’ method employed to isolate the CA repeat from HX20 is especially suitable for isolating microsatellites from a limited number of larger clones, where the CA repeats will be sufficiently represented. For large-scale screening to generate large sets of markers for compiling linkage maps of whole genomes, however, thousands of plates need to be screened. In addition, when using sequencing vectors, best results are obtained when the microsatellite is situated 50-200bp from one end of a cloned restriction fragment. Problems can arise if the microsatellite happens to be juxtaposed with the vector cloning site, providing insufficient genomic sequence to permit a primer to be designed for developing a PCR-based assay, or if the microsatellite is situated such that two or more sequencing steps are required to identify it, as in the case of HX20-2b.

Recognition of these limitations has led to a variety of improved procedures for recovery of microsatellite polymorphisms. In order to determine sequences which are inaccessible with vector-derived primers, degenerate sequencing primers which anneal directly to (CA)n microsatellites have been used (Yuille et al 1991, Browne and Litt 1992, LeBlanc-Straceski et al 1994). Direct sequencing of plasmid, phage or cosmid clones with such primers is used to recover sequence information flanking microsatellites in large-insert clones. To ensure that this sequence has a unique start site, primers are synthesised with one or two unique bases at their 3' end. Browne and Litt used two sets of 4 primers each [(CA) or (GT)] with a twelve-fold degeneracy at the 3' end to sequence plasmid templates. Using these primer mixes, they obtained sequence from both strands by running two sets of sequencing reactions. Their method was sufficient to obtain DNA sequence from 88% of the microsatellites attempted. Yuille et al. used primers with a minimum level of degeneracy at the 3' end, individually in 6 sets of sequencing reactions to sequence one strand of plasmid templates. LeBlanc-Straceski et al used both methods in combination with cycle sequencing which allowed them to sequence directly from a cosmid template.

PCR-based procedures for the isolation of microsatellites have also been reported which tend to improve the ease of obtaining sequence as well of sequence readability. These methods exploit anchored PCR to introduce, flanking the microsatellite, a unique priming site suitable for direct DNA sequencing. Briefly, YAC, cosmid or phage DNA is digested with one of several frequently cutting restriction enzymes, followed by ligation to appropriate linkers
PCR is performed on the ligation products, using primers specific to the microsatellite and the linker to amplify the intervening sequence, which is then sequenced either directly or after cloning. The linkers used in the above reports were those described by Riley et al. (1990). Koref et al. (1993) ligated restricted DNA to dephosphorylated plasmid vectors and performed PCR on the ligated products, using a primer for the microsatellite and a primer specific to the vector. The flanking region was sequenced in a separate reaction.

These latter approaches streamline the recovery of microsatellites from large-insert clones, but still require that individual clones be processed through several steps prior to sequencing; also, two sequencing runs are necessary to obtain the flanking sequence required for the development of a PCR-based assay. More importantly, such strategies result in the loss of precise sequence identity of the microsatellite, of potentially great value in predicting the informativeness of a marker (Weber 1990). The major advantage of these methods, nevertheless, is that by avoiding the subcloning step, they allow immediate use of existing genomic libraries for the recovery of genetic markers, which could be of great assistance in large-scale projects.

Another line of research directed towards the efficiency of microsatellite recovery has focused on the problem of representation in small-insert libraries, a more serious problem with trimeric and tetrameric tandem repeats, because of their lower abundance in the genome (section 1.2.3.1). Building on earlier studies which used DNA-affinity chromatography to enrich 10-fold for CA/GT repeats, Armour et al. (1994) have used a combination of the above methods to develop a more efficient enrichment strategy based on hybridisation selection. Genomic DNA fragments were restricted with MboI and ligated to appropriate linkers, amplified with linker primers, denatured and hybridised to filters bearing mixed trimeric or tetrameric repeat arrays. The hybridised fragments were recovered and reamplified whence they were found to be 50-100 fold enriched for simple tandem repeats (approximately 30% of clones gave positive hybridisation signals).

These recent developments in strategies for the efficient isolation of highly polymorphic genetic markers therefore provide a variety of methods which can be chosen according to the particular demands of the investigation in question. In conclusion, the ‘shotgun subcloning’ method adopted here proved sufficient to isolate both CA repeats present in cosmid HX20. Although one of these turned out to be of no use as a genetic marker, the highly polymorphic (CA)$_{21}$ microsatellite DXS556 is physically mapped to Xp11.4, and should therefore be of indisputable value as a genetic marker for the linkage and fine mapping of diseases, such as XLRP, which map to this region.
CHAPTER 5

Genetic and Physical Mapping of Five Novel Microsatellite Markers on Xp21.1-Xp11.22

5.1 Introduction

This chapter describes the final part of the initial objective of increasing the polymorphic marker density in proximal Xp in order to provide additional tools for fine genetic mapping of the RP2 locus. The greater the map density, the more accurate our ability will be to determine the exact points of recombination in each informative meiosis and thus more accurately define the disease gene critical region.

In addition to the microsatellite DXS556 isolated from cosmid HX20 in Xp11.4 (chapter 4), similar methods were employed by team colleagues to identify microsatellites from other cosmids mapped by FISH to proximal Xp (HX43, HX91, chapter 3; HX81, HX82, HX97, S. Kamakari 1994). The utility of these genetic markers would be improved by establishing their precise location with respect to known markers in proximal Xp, and this chapter describes how this was achieved by a combination of genetic and physical mapping.

5.1.1 The locus-ordering problem

Locus ordering is of critical importance in the generation of the high resolution genetic maps required to refine the localisation of a disease gene. Moreover, for multipoint linkage analysis of a disease to be meaningful an accurate map is essential. For this reason, one of the crucial problems in linkage analysis is the ordering of sets of closely linked markers.

In practice, the first step in ordering a set of loci is to carry out conventional two-point linkage analysis for each locus pair (pairwise analysis), where the recombination fractions can give an approximation of genetic distance over small intervals e.g. for a set of three loci, the two showing the highest recombination fraction become the outside markers.

An alternative way to choose which order is best is to pick the one that requires the minimum number of double recombinants, because these are known to be rare events due to the phenomenon of interference (section 1.2.2). The most effective approach for determining locus order examines recombination among the offspring of individuals heterozygous for 3 or more loci, such that recombination in several intervals can be followed in the same set of chromosomes (multiply informative chromosomes). Such haplotype-based mapping procedures are generally the best for fine mapping at localised regions (10-20cM) because the drastically reduced likelihood of close multiple exchanges can provide good evidence for the elimination of most of the possible orders.
However, in typical human pedigrees all meioses are not informative and phases are not known for all pairs of markers. Multipoint analysis using a variety of computer programs can be used to solve this problem, and has much greater ordering power than two-point analysis as it can help discriminate between parental alleles of an uninformative marker and thus provide more precise localisation of recombination events. Multilocus analysis programs generally incorporate a level of interference which permits better discrimination between locus orders, the major reason being that different locus orders predict different numbers of double recombinants, which are less likely with increased interference. The ILINK program of the LINKAGE package (Lathrop et al 1984) can be used to maximise the likelihood under each possible order by estimating each intermarker recombination fraction jointly. Thus, maximum use is made of all the pedigree information, including the partially informative meioses. Typically, for a locus order to be conclusive, it must be supported by 1000:1 odds over the second best order or a log_{10}(likelihood) difference of at least 3 between them (section 5.1.2.1).

5.1.2 General strategies for map construction

For the construction of accurate genetic maps of known order and genetic distance between adjacent loci, the maximum likelihood strategy is generally applied. This amounts to computing the likelihood for each possible order; then the order with the highest associated likelihood is the best estimated order of loci.

For a large number of loci, preliminary ordering methods based on pairwise analysis (reviewed in Weeks 1991) are often applied, which can provide an approximate measure for the plausibility of the data under each order e.g. the sum of adjacent recombination fractions (SARF, the smaller the better) which is obtained by adding the recombination fraction estimates in each interval. Another measure is the sum of two-point lod scores for all pairs of adjacent loci (SAL, large values are good). Although of some value in providing broad indications of marker order, these measures also depend on other parameters such as the heterozygosity of the markers used, and are therefore not considered practically reliable for the precise ordering desired when constructing high resolution genetic maps (Terwilliger and Ott 1992).

The most popular strategy in current use is to calculate exact likelihoods for only a limited number of markers at a time under all possible orders. Initial sets of markers are selected whose order can be established unequivocally (usually the most informative markers) for building a ‘skeletal map’ which is then extended by addition of new markers in a stepwise fashion. Small groups of loci can also be provisionally ‘grouped’ on the basis of pairwise analysis, facilitating the selection of particular marker combinations for subsequent evaluation of multipoint likelihoods to infer locus order (e.g Mahtani et al 1991).

In addition to the ILINK program, various other programs exist for the construction of multilocus linkage maps, of which perhaps CILINK, CRIMAP and MAPMAKER are the most widely used (reviewed in Terwilliger and Ott 1994). However, these programs require simple three generation families as input data and although they are therefore applicable to CEPH
family analysis (section 5.4), are rarely used for disease gene mapping where sampled pedigrees tend to have less suitable structures and often require inference of grandparental genotypes.

5.1.2.1 Framework versus comprehensive genetic maps

Framework maps are of low resolution, high support and consist of loci uniquely placed with odds of 1000:1 or greater, and are thus conservative but reliable. It is desirable that framework loci be evenly spaced, highly informative, easily assayed markers whose physical locations are known. These markers may then be used as a frame against which all other markers are localised, creating a comprehensive map with much finer resolution. The CHLC and Genethon reference maps (section 1.2.5) have been constructed using framework criteria.

Comprehensive maps are of high resolution but low local support as they consist of all syntenic markers uniquely placed in their most likely location regardless of support. These maps are most usually constructed over small chromosomal regions such as those harbouring a disease gene where focussed efforts have assembled large numbers of markers from various sources and of different types. These maps provide a unique genetic location for every locus, although they may be unreliable regarding local order and distance.

Framework maps are required if genetic maps are to be used as punctuation points in other physical mapping strategies. Comprehensive maps are important when linkage maps are used to find the chromosomal locations for disease genes. In these studies, it is more important to know approximate locations for all markers in the vicinity in efforts to home in on the gene.

5.1.3 The ordering of novel microsatellite markers on proximal Xp

In summary, the microsatellites corresponding to cosmids HX20, HX81, HX82, HX91 and HX97 were tested against a panel of XLRP families which had already been genotyped using known markers in proximal Xp, in an effort to incorporate these new markers into the existing genetic linkage map of the region. The microsatellite isolated from cosmid HX43 has been described elsewhere (Kamakari et al 1995).

Two methods of genetic mapping were employed: linkage analysis to estimate genetic distances between the loci in addition to the most likely locus order, and haplotype analysis to more definitively determine locus order.

It is now widely accepted that physical mapping methods can provide a complementary approach to genetic mapping for ordering loci and constructing accurate marker maps. Physical methods are good at ordering closely spaced loci, which is difficult by genetic mapping. Several methods for ordering loci based on radiation hybrid mapping have recently become available (section 1.3.1.1).

A combination of genetic mapping methods and physical mapping using a panel of somatic cell hybrids and YACs from the region confirmed and substantially refined the FISH localisations for these new markers on proximal Xp, and also revealed their order. By
integration with anchor markers in the same region, these new polymorphic microsatellites have been incorporated into the existing genetic map of Xp21.1-Xp11.22, and thus provide a more detailed framework from which to finely map disease genes localised to this region of proximal Xp.

5.2 Materials and methods

5.2.1 Families

The families used in this study consisted of 14 XLRP pedigrees, containing ~130 potentially informative meioses, for which ongoing genetic analysis was being undertaken as part of a positional cloning strategy to identify the RP2 gene. A more extensive genetic characterisation of these families is described in chapter 6, with specific regard to the disease phenotypes.

For the purpose of ordering genetic markers in this chapter, these families were recruited as genotype data was already available for several reference markers in proximal Xp; clinical details are of no consequence in this respect and are thus omitted for simplification. DNA was extracted from venous blood samples by the methods described in section 2.2.1.

5.2.2 Genetic markers

Table 5.1 presents a description of the novel microsatellites whose physical localisations we sought to refine by genetic mapping with respect to 9 reference markers with well-established map positions on proximal Xp, listed in Table 5.2. The order of reference markers was known to be CYBB-DXS1110-OTC-DXS228-DXS7-MAOA-(SYN1-DXS426)-DXS255 (Willard et al 1994). The groups (DXS228-DXS7-MAOA) and (SYN1-DXS426) were 'haplotyped' i.e. information from markers in each set was combined to yield haplotypes that were subsequently treated as alleles at a single locus. The purpose of this was to increase the total information obtainable from the pedigrees at that locus (one or other marker in each group may not have been informative in a given pedigree), and was considered valid as no crossovers were evident between them owing to their close proximity. From now on, these groups will be referred to as markers DXS7 and DXS426 respectively.

5.2.3 Genotyping

For typing of CA repeats, PCR reactions were performed using ~100ng DNA with one of the primers end-labelled with $^{32}$P (section 2.7.4.1). PCR products were diluted with formamide buffer, and 2-4μl aliquots were analysed on a 6% sequencing gel, followed by autoradiography for allele detection (section 2.4.2). Although the exact sizes of alleles were not established for each microsatellite, scoring was standardised across families by inclusion of already-typed members on each gel. Genotypes were recorded independently from the family data, and transferred to the family pedigree where X-linked segregation was confirmed. Genotyping inconsistencies (such as double crossovers) were subject to scrutiny and resolved.
Table 5.1: Description of the 5 novel microsatellites mapped by FISH to proximal Xp

<table>
<thead>
<tr>
<th>LOCUS</th>
<th>FISH LOCATION</th>
<th>PRIMER SEQUENCES (EMBL ACCESSION NO.)</th>
<th>Allele</th>
<th>Length (nt)</th>
<th>Frequency</th>
<th>Expected Heterozygosity</th>
</tr>
</thead>
<tbody>
<tr>
<td>DXS572 (HX81)</td>
<td>Xp21.1</td>
<td>CA: 5’GCCACCATAGAAGCAGAAAATG3’</td>
<td>A1</td>
<td>164</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>GT: 5’CTATTCCTTAGCTCATC3’</td>
<td>A2</td>
<td>162</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A3</td>
<td>160</td>
<td>0.83</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A4</td>
<td>158</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A5</td>
<td>156</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>DXS573 (HX82)</td>
<td>Xp11.2-11.22</td>
<td>CA: 5’GCCACCCAACTAAAGTGGC3’</td>
<td>A1</td>
<td>145</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>GT: 5’GGTGATGATGAGTGATAGAAGG3’</td>
<td>A2</td>
<td>143</td>
<td>0.42</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A3</td>
<td>141</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A4</td>
<td>139</td>
<td>0.30</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A5</td>
<td>137</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>DXS574 (HX91)</td>
<td>Xp11.3</td>
<td>CA: 5’ATGGAATTGTGAACACTGAG3’</td>
<td>A1</td>
<td>212</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>GT: 5’TCTACAGCCAGAACTGGA3’</td>
<td>A2</td>
<td>210</td>
<td>0.37</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A3</td>
<td>208</td>
<td>0.48</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A4</td>
<td>204</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>DXS572 (HX97)</td>
<td>Xp11.22-11.23</td>
<td>CA: 5’GCTCATACTGCTAGCTGCTG3’</td>
<td>A1</td>
<td>192</td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>GT: 5’GCCTCTGTTCCACTGAGG3’</td>
<td>A2</td>
<td>190</td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A3</td>
<td>188</td>
<td>0.70</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A4</td>
<td>186</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>DXS556 (HX20)</td>
<td>Xp11.3-11.4</td>
<td>CA: 5’AGTTTGGGCTGGTTCGTTAC3’</td>
<td>A1</td>
<td>192</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>GT: 5’TATGAGACAGCCAAGCTTGA3’</td>
<td>A2</td>
<td>190</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A3</td>
<td>188</td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A4</td>
<td>186</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A5</td>
<td>184</td>
<td>0.37</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A6</td>
<td>182</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A7</td>
<td>178</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A8</td>
<td>176</td>
<td>0.28</td>
<td></td>
</tr>
</tbody>
</table>
### Table 5.2 Reference genetic markers used in genetic map construction

<table>
<thead>
<tr>
<th>Locus</th>
<th>Chromosomal location</th>
<th>Type of polymorphism</th>
<th>Heterozygosity</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYBB</td>
<td>Xp21.1</td>
<td>microsatellite</td>
<td>0.76</td>
<td>Hardwick et al 1993</td>
</tr>
<tr>
<td>DXS1110</td>
<td>Xp21.1</td>
<td>microsatellite</td>
<td>0.68</td>
<td>Roux et al 1993</td>
</tr>
<tr>
<td>OTC</td>
<td>Xp21.1</td>
<td>a) RFLP 1*</td>
<td>0.30</td>
<td>Wright et al 1991</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b) RFLP 2</td>
<td>0.44</td>
<td>Petty et al 1991</td>
</tr>
<tr>
<td>DXS228</td>
<td>Xp11.3</td>
<td>microsatellite</td>
<td>0.53</td>
<td>Coleman et al 1991</td>
</tr>
<tr>
<td>MAOA</td>
<td>Xp11.3</td>
<td>microsatellite</td>
<td>0.72</td>
<td>Black et al 1991</td>
</tr>
<tr>
<td>SYN1</td>
<td>Xp11.23</td>
<td>microsatellite</td>
<td>0.84</td>
<td>Kirchgessner et al 1991</td>
</tr>
<tr>
<td>DXS426</td>
<td>Xp11.23</td>
<td>microsatellite</td>
<td>0.52</td>
<td>Luty et al 1990</td>
</tr>
<tr>
<td>DXS255</td>
<td>Xp11.22</td>
<td>VNTR*</td>
<td>&gt;0.90</td>
<td>Fraser et al 1987</td>
</tr>
</tbody>
</table>

* Typing for these markers was performed prior to this study

by going back to the autoradiograph and rescoring the alleles. If such inconsistencies were not due to misscoring of alleles, the reaction was repeated. Tube mixup and non-paternity were excluded from consistent genotypes obtained with additional markers.

The biallelic OTC RFLP2 polymorphism was detected by non-radioactive PCR, followed by digestion with Dral and resolution on 4% agarose gels (3% NuSieve:1% normal agarose) as described (Petty et al 1991).

### 5.2.4 Linkage analysis

The ILINK program of the LINKAGE package (version 5.1, Lathrop et al 1984) was used on an IBM PC (a) to perform two-point analysis on the marker data, to provide the best estimate of the recombination fraction between each pairwise combination and to approximate the positions of the microsatellite markers relative to the reference markers (b) to perform multipoint analysis on subsets of markers, grouped on the basis of the two-point results and haplotype analysis, to establish the most likely order for the markers.

The LINKAGE package contains three programs which calculate two-point or multipoint likelihoods in human pedigrees. MLINK calculates two-point likelihoods of family genotype data by stepwise variation of the recombination fraction ($\theta$) between two markers and associated lod scores. The LINKMAP program is used to perform multipoint analysis and calculate location scores for a gene or marker whose precise position is unknown (e.g. a disease gene). LINKMAP assumes a fixed map of markers and calculates likelihoods for a new locus at various points in each interval along the known map. ILINK is the iterative version of MLINK and finds maximum likelihood estimates of linkage between two loci; it also performs
multipoint analysis to assess the most likely order of a set of markers. ILINK was used here to obtain a maximum likelihood estimate of the pairwise recombination fraction between markers, with the corresponding highest lod score. The LINKAGE package was chosen for two-point analysis as opposed to the user-friendly LIPED system (Ott 1974) as it can compute up to 10 marker systems (LIPED can only cope with 5) and so saves time. In addition, the recombination fraction and lod score values obtained using ILINK correspond to that found by LIPED or MLINK but the estimate of θ is more precise. The ILINK program was subsequently used to generate the most likely order for the new microsatellites with respect to the reference markers.

As input, LINKAGE takes files from LINKSYS (Attwood and Bryant 1988) or PREPLINK (family data and marker data) and calculates the likelihoods of such data. To prepare the data for LINKAGE the program LINKSYS was employed: a series of ‘pedigree files’ are first generated, containing member details along with genotypes for a specified series of markers. Parameter files ‘phenobil’ and ‘locuslib’ are created to incorporate information regarding the member phenotypes and marker allele numbers and frequencies. A ‘template’ file is then created, whereby the pedigree files are selected for which linkage analysis is to be performed (i.e. combining data from more than one family). At this point a check can be run on the input thus far, such that any discrepancies between pedigree and parameter files can be corrected (e.g. incorrect family structure or locus information). The template pedigree file and parameter files are then processed by the MAKEPED program, which adds several pointers required by the LINKAGE programs prior to linkage analysis via the LINKAGE CONTROL program (LCP).

When running these programs on PCs, one practical limitation is that they can only handle 48 haplotypes at once, i.e. no. of alleles of marker 1 x marker 2 etc. must not exceed 48. When using ILINK in its multipoint capacity, therefore, to determine the most likely order of a set of markers, there would be too many possible haplotypes to permit all 10 loci to be analysed simultaneously. The task was simplified by a) ‘recycling of alleles’ to reduce the number of haplotypes in such a way that no inheritance information was lost, and b) ordering subsets of loci in turn: firstly, the 3 reference loci distal to DXS7 (CYBB, DXS1110 and OTC) and the 2 proximal to DXS7 (DXS426, DXS255), with DXS7 in common to enable an overlap of the maps. The 5 novel microsatellites were then positioned onto this ‘reference map’ sequentially, by taking relevant groups of 3 loci and finding the new best order i.e. that with the highest likelihood. The best estimates of the genetic distances between adjacent markers were then calculated using the Kosambi mapping function (section 1.2.2) with the most likely recombination fractions (θ̂_{max}) via the MAPFUN program of the LINKAGE package.

5.2.4.1 Correction of marker allele frequencies

Designation of incorrect allele frequencies for markers in linkage analysis can profoundly affect the results obtained. Assigning equal allele frequencies to markers may result in false evidence for linkage as well as failure to detect true linkage (Knowles et al 1992). These effects are most pronounced in pedigrees with key individuals unavailable for
genotyping, particularly in multigeneration families where genotypes of the top generations are to be inferred by the LINKAGE program (Freimer et al 1993). In this case, cosegregation of common alleles will greatly reduce the significance of linkage since the LINKAGE program will consider it more likely that these alleles could have come to the family from the general population pool through spouses. On the other hand, by falsely assuming lower frequency of common alleles segregating in families, false linkages will be concluded since the program will give very little significance to the fact that these alleles could have entered the family at any point via spouses.

Therefore, to produce linkage data most likely to reflect the real pedigree data, allele frequencies were calculated after marker haplotyping (section 5.2.2) and allele recycling (section 5.3.1) from unrelated individuals taken from the entire set of pedigrees (on average 42 chromosomes per marker) as a representative control sample for the family population included in this study.

5.2.5 Haplotype analysis

The segregation of marker alleles was followed in each of the families to establish sites of meiotic recombination as a means to determine locus order. The most likely haplotypes were inferred by minimising the number of crossover events in each sibship. The first step was to identify individuals who exhibited recombination events between the anchor markers. Given that the recombination breakpoints lie between the anchor markers and that the test markers map generally to the region, it follows that for each breakpoint, one or more of the test markers will lie on one side of the breakpoint and the rest on the other side. Using this information, it was possible to begin building an order among the test markers.

5.2.6 Physical mapping studies

For markers which could not be separated by genetic linkage or haplotype analysis, physical mapping was attempted using available cloned DNA resources from the relevant region of proximal Xp. Markers DXS722 and DXS573 were both known to lie in Xp11.23-Xp11.22 from FISH localisation (Table 3.1). PCR analysis was performed on a human-rodent somatic cell hybrid SIN176 (section 2.7.8.3) and a series of YACs from the region spanning DXS426 to OATL1 (section 5.3.4). These clones were well-characterised as to their human DNA component by STS-content mapping and physical studies provided additional information regarding the position and order of these 2 markers in relation to reference markers.

5.3 Results

5.3.1 Genotype data

Genotypes were generated for members of 14 XLRP families using 9 reference markers (treated as 6 distinct loci) and the 5 novel microsatellites. Pedigrees and original genotypes are
shown in Figure 5.1(a-n). Table 5.3 shows an example of allele recycling to reduce the number of haplotypes for linkage analysis by allowing a maximum of 4 alleles per marker. This procedure obviously altered the allele frequencies for those markers to which it was applied, therefore these were recalculated as described in section 5.2.4.1.

NB: Since publication of this work (Thiselton et al 1995), further information has led to modification of one pedigree (family F5) due to the discovery that one member was adopted. The results presented here are therefore adjusted accordingly. The OTC data for members II-7 and III-1 of family F45 (Figure 5.1f), and member III-1 of family F47 (Figure 5.1g) has also been removed from the linkage analysis as it represents inconsistency, the reasons for which will become clear.

Table 5.3 Demonstration of allele recycling in family F47 (Figure 5.1g)

<table>
<thead>
<tr>
<th>Person</th>
<th>Father</th>
<th>Mother</th>
<th>CYBB Original alleles</th>
<th>CYBB Recycled alleles</th>
<th>DXS1110 Original alleles</th>
<th>DXS1110 Recycled alleles</th>
<th>DXS573 Original alleles</th>
<th>DXS573 Recycled alleles</th>
</tr>
</thead>
<tbody>
<tr>
<td>I-1</td>
<td></td>
<td></td>
<td>(8 → 1)</td>
<td>(5 → 3)</td>
<td>(9 → 1)</td>
<td>(4 → 3)</td>
<td>(5 → 4)</td>
<td></td>
</tr>
<tr>
<td>I-2</td>
<td></td>
<td></td>
<td>4-allele system: 1234</td>
<td>4-allele system: 1234</td>
<td>3-allele system: 234</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>II-1</td>
<td>I-2</td>
<td>I-1</td>
<td>2-8</td>
<td>1-2</td>
<td>5-9</td>
<td>1-3</td>
<td>2-4</td>
<td>2-3</td>
</tr>
<tr>
<td>II-2</td>
<td></td>
<td></td>
<td>4-allele system: 1234</td>
<td>4-allele system: 1234</td>
<td>3-allele system: 234</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>II-3</td>
<td>I-2</td>
<td>I-1</td>
<td>2-3</td>
<td>2-3</td>
<td>2-5</td>
<td>2-3</td>
<td>2-5</td>
<td>2-4</td>
</tr>
<tr>
<td>II-4</td>
<td></td>
<td></td>
<td>4-allele system: 1234</td>
<td>4-allele system: 1234</td>
<td>3-allele system: 234</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>III-1</td>
<td>II-2</td>
<td>II-1</td>
<td>2-1</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>III-2</td>
<td>II-2</td>
<td>II-1</td>
<td>8-1</td>
<td>9</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>III-3</td>
<td>II-2</td>
<td>II-1</td>
<td>2-2</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>III-4</td>
<td>II-2</td>
<td>II-1</td>
<td>2-4</td>
<td>2-4</td>
<td>5</td>
<td>3</td>
<td>2-5</td>
<td>2-4</td>
</tr>
<tr>
<td>III-5</td>
<td>II-4</td>
<td>II-3</td>
<td>3-3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>III-6</td>
<td>II-4</td>
<td>II-3</td>
<td>2-2</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

5.3.2 ILINK two-point analysis between pairs of markers

The genotype data were analysed pairwise to determine the maximum likelihood θ value (θmax) at maximum LOD (Zmax) using the ILINK program (Lathrop et al 1984). The resulting two-point data for all pairwise combinations of markers is presented in Table 5.4, both before and after the data qualifications described above (section 5.3.1).

Let us first consider the markers distal to DXS7, where genotype input modifications had their most profound effects. It is immediately apparent from the two-point data in Tables 5.4(a) and (b) that DXS1110 and CYBB show complete cosegregation and therefore could not be separated genetically in our group of families.
Figure 5.1 14 XLRP pedigrees and associated genetic haplotypes showing recombination events (X) from which marker order information was derived.

Figure 5.1a: Family F5

Figure 5.1b: Family F11
Figure 5.1c: Family F15

Markers

- DXS572
- CD98
- DXS1110
- OTC (PCR)
- DXS556
- DXS74
- DXS7 (probe)
- DXS426
- DXS722
- DXS573
- DXS255
Figure 5.1d: Family F21

Markers
- CYBB
- DXS1110
- OTC (probe)
- DXS556
- DXS574
- DXS7 (probe)
- DXS426
- DXS255

Figure 5.1e: Family F38

Markers
- DXS572
- CYBB
- DXS1110
- OTC (probe)
- DXS556
- DXS574
- DXS7 (DXS228)
- DXS426
- DXS722
- DXS573
- DXS255
Figure 5.1f: Family F45

Figure 5.1g: Family F47
Figure 5.1h: Family F51

Figure 5.1i: Family F53
Figure 5.1j: Family F71

Markers

CYBB
DXS1110
DXS556
DXS574
DXS7 (MAOA)
DXS426
DXS722
DXS573

Figure 5.1k: Family F72

Markers

DXS572
CYBB
DXS1110
DXS574
DXS7 (DXS228)
DXS426 (SYN1)
DXS722
DXS573
Figure 5.1n: Family F83
Table 5.4: Pairwise maximum likelihood recombination fractions (below the diagonal) and the corresponding lod scores (above the diagonal)

a) genetic markers distal to DXS7 (pedigree data unmodified)

<table>
<thead>
<tr>
<th></th>
<th>DXS572</th>
<th>CYBB</th>
<th>DXS1110</th>
<th>OTC</th>
<th>DXS556</th>
<th>DXS574</th>
<th>DXS7</th>
</tr>
</thead>
<tbody>
<tr>
<td>DXS572</td>
<td>7.58</td>
<td>1.95</td>
<td>5.61</td>
<td>2.18</td>
<td>2.36</td>
<td>2.85</td>
<td></td>
</tr>
<tr>
<td>CYBB</td>
<td>0.064</td>
<td>18.57</td>
<td>11.33</td>
<td>11.71</td>
<td>9.17</td>
<td>11.55</td>
<td></td>
</tr>
<tr>
<td>DXS1110</td>
<td>0.096</td>
<td>0.000</td>
<td>2.17</td>
<td>7.17</td>
<td>5.34</td>
<td>7.43</td>
<td></td>
</tr>
<tr>
<td>OTC</td>
<td>0.033</td>
<td>0.072</td>
<td>0.154</td>
<td>6.94</td>
<td>6.81</td>
<td>7.41</td>
<td></td>
</tr>
<tr>
<td>DXS556</td>
<td>0.126</td>
<td>0.075</td>
<td>0.089</td>
<td>0.078</td>
<td>16.28</td>
<td>18.35</td>
<td></td>
</tr>
<tr>
<td>DXS574</td>
<td>0.181</td>
<td>0.118</td>
<td>0.133</td>
<td>0.100</td>
<td>0.015</td>
<td>16.59</td>
<td></td>
</tr>
<tr>
<td>DXS7</td>
<td>0.137</td>
<td>0.106</td>
<td>0.118</td>
<td>0.113</td>
<td>0.023</td>
<td>0.035</td>
<td></td>
</tr>
</tbody>
</table>

b) genetic markers distal to DXS7 (potential errors eliminated)

<table>
<thead>
<tr>
<th></th>
<th>DXS572</th>
<th>CYBB</th>
<th>DXS1110</th>
<th>OTC</th>
<th>DXS556</th>
<th>DXS574</th>
<th>DXS7</th>
</tr>
</thead>
<tbody>
<tr>
<td>DXS572</td>
<td>7.25</td>
<td>1.93</td>
<td>5.57</td>
<td>2.05</td>
<td>2.39</td>
<td>2.86</td>
<td></td>
</tr>
<tr>
<td>CYBB</td>
<td>0.071</td>
<td>17.26</td>
<td>14.65</td>
<td>11.64</td>
<td>9.13</td>
<td>12.27</td>
<td></td>
</tr>
<tr>
<td>DXS1110</td>
<td>0.102</td>
<td>0.001</td>
<td>4.10</td>
<td>6.94</td>
<td>5.30</td>
<td>8.35</td>
<td></td>
</tr>
<tr>
<td>OTC</td>
<td>0.034</td>
<td>0.016</td>
<td>0.045</td>
<td>10.01</td>
<td>7.66</td>
<td>8.26</td>
<td></td>
</tr>
<tr>
<td>DXS556</td>
<td>0.136</td>
<td>0.075</td>
<td>0.088</td>
<td>0.001</td>
<td>15.77</td>
<td>17.70</td>
<td></td>
</tr>
<tr>
<td>DXS574</td>
<td>0.179</td>
<td>0.117</td>
<td>0.133</td>
<td>0.067</td>
<td>0.015</td>
<td>16.31</td>
<td></td>
</tr>
<tr>
<td>DXS7</td>
<td>0.137</td>
<td>0.095</td>
<td>0.095</td>
<td>0.084</td>
<td>0.023</td>
<td>0.035</td>
<td></td>
</tr>
</tbody>
</table>

NB: major effects on recombination fractions and lod scores are highlighted in italics

c) genetic markers proximal to DXS7

<table>
<thead>
<tr>
<th></th>
<th>DXS7</th>
<th>DXS426</th>
<th>DXS722</th>
<th>DXS573</th>
<th>DXS255</th>
</tr>
</thead>
<tbody>
<tr>
<td>DXS7</td>
<td>10.35</td>
<td>8.95</td>
<td>9.23</td>
<td>7.97</td>
<td></td>
</tr>
<tr>
<td>DXS426</td>
<td>0.123</td>
<td>18.65</td>
<td>22.79</td>
<td>14.57</td>
<td></td>
</tr>
<tr>
<td>DXS722</td>
<td>0.107</td>
<td>0.000</td>
<td>17.21</td>
<td>9.64</td>
<td></td>
</tr>
<tr>
<td>DXS573</td>
<td>0.113</td>
<td>0.000</td>
<td>0.000</td>
<td>7.83</td>
<td></td>
</tr>
<tr>
<td>DXS255</td>
<td>0.105</td>
<td>0.016</td>
<td>0.024</td>
<td>0.047</td>
<td></td>
</tr>
</tbody>
</table>
Cosmid HX81, containing microsatellite DXS572, was already known to lie distal to OTC by virtue of its absence from a cell line containing a deletion beginning just proximal to CYBB and extending telomeric into the DMD gene (S. Lindsay, personal communication). DXS572 is not highly polymorphic (expected heterozygosity 29%), which means that there is less information regarding its position relative to the other markers (it is only informative in 4 families). Inspection of the two-point data would imply that DXS572 is closer to OTC than DXS1110 or CYBB.

The most striking effect of removing sources of potential error from the pedigree data is seen with regard to the OTC two-point data in Tables 5.4(a) and (b). This is to be expected, as genotype data for the individuals concerned involved double crossovers directly flanking OTC (section 5.3.1). Removal of these unlikely recombination events has resulted in a reduction of the pairwise recombination fractions between OTC and all other markers (except DXS572) to give much more realistic values considering published genetic distances between these markers (discussed in section 5.4). This is exemplified most dramatically between OTC and DXS1110, where $\theta_{\text{max}}$ is reduced from 0.154 ($Z_{\text{max}}=2.17$) to 0.045 ($Z_{\text{max}}=4.30$). Another consequence of adjusting the data input for ILINK two-point analysis is that OTC and DXS556 become completely cosegregating markers with a pairwise $\theta_{\text{max}}$ of 0.001 ($Z_{\text{max}}=10.01$) as opposed to 0.078 ($Z_{\text{max}}=6.94$). However, the observation that DXS556 is much more tightly linked to DXS7 ($\theta_{\text{max}}=0.023$, $Z_{\text{max}}=17.70$) than DXS1110 ($\theta_{\text{max}}=0.088$, $Z_{\text{max}}=6.94$) would indicate that DXS556 lies proximal to OTC, a pattern that is similarly observed for marker DXS574. Other slight differences in $\theta_{\text{max}}$ and $Z_{\text{max}}$ values between Tables 5.4(a) and (b) are a consequence of removing the adopted individual in family F5 from the analysis altogether. This individual displayed multiple recombination events along the entire portion of Xp in question (data not shown), which may have contributed to these variations.

For the markers proximal to DXS7, the genotype adjustments had negligible effect on the pairwise data and thus only the modified two-point data is shown in Table 5.4(c). This data reveals that DXS722 and DXS573 could not be distinguished genetically from DXS426 or from each other, as there are no crossovers internally between any of these microsatellites, preventing the resolution of their relative order.

### 5.3.3 Ordering the loci by haplotype analysis

In general, the recombination fractions derived from pairwise analysis of the loci are consistent with the order of markers as determined by haplotype analysis. However, one discrepancy is evident regarding the location of DXS572. A recombination event in Figure 5.1c; individual IV-1 confirms a location for DXS572 distal to OTC (see section 5.3.2). However, DXS572 is placed distal to CYBB by 2 recombination events shown in Figure 5.1k; individuals II-2 and II-4, which suggest that DXS572 is also distal to CYBB. This is in contrast to the information deduced from the two-point data alone (section 5.3.2) which suggests a location for DXS572 much closer to OTC. However, caution should be taken when interpreting two-point data to define order of markers: the recombination fractions between
pairs of markers provide a reasonable estimate of the genetic distance between them (for small genetic distances and pending suitable conversion), but ordering information can only accurately be determined from multipoint analysis (Ott 1991). The observation that both \textit{OTC} and DXS572 have low heterozygosity values, and were not simultaneously informative in the relevant meioses, may explain this ambiguity (Terwilliger and Ott 1992). In Figure 5.1k, the alternative position for DXS572 proximal to (\textit{CYBB-DXS1110}) generates double crossovers directly flanking DXS572 in individuals II-2 and II-4 which is highly unlikely given the small genetic interval. Additionally, the phenomenon of interference very likely precludes a second crossover so close to the first one (section 1.2.2). Therefore, from multiply informative meioses it can be concluded that the order of loci must be Xpter-DXS572-\{\textit{CYBB-DXS1110}\}-Xcen.

In Figure 5.1h, individual II-2 shows a chromosome which is recombinant between DXS1110 and DXS556/DXS574, placing these two microsatellites proximal to DXS1110. Similar crossovers seen in Figure 5.1d; individual II-2 and Figure 5.1k; II-7 support this conclusion. Individual III-7 in Figure 5.1a shows a crossover event between DXS574 and DXS556, but not between DXS556 and \textit{OTC} or between DXS574 and DXS7, supporting the order \textit{OTC-DXS556-DXS574-DXS7}. This recombination also confirms a location for DXS556 distal to DXS7, together with recombinant individual IV-9 in Figure 5.1b. Similarly, individuals III-18 and III-21 in Figure 5.1c and individual IV-3 in Figure 5.1n show crossovers between DXS7 and DXS574, but not between DXS574 and \textit{OTC} which also confirms that DXS574 lies distal to DXS7.

The remaining discussion will address the ordering of markers between DXS7 and the centromere. Multiple recombination events in this data set support a location for both DXS722 and DXS573 proximal to DXS7 e.g. Figure 5.1j; III-1, Figure 5.1f; III-7, Figure 5.1g; III-2, Figure 5.1m; II-2. In Figure 5.1b, individual IV-10 shows a recombination event between both DXS573/DXS722 and DXS255, but not between DXS573/DXS722 and DXS426, placing both DXS573 and DXS722 distal to DXS255. The absence of accurately defined crossover events between DXS426, DXS722 and DXS573 meant that haplotype analysis could not resolve their relative order. Figure 5.2 shows a summary of the 6 recombinant meiotic events informative for ordering the loci.

\subsection*{5.3.4 Ordering DXS722 and DXS573 by physical mapping}

Physical mapping studies gave additional information regarding the order of these two markers. DXS573 and DXS722 are seen to be different microsatellites by their distinctly different levels of heterozygosity (72\% and 46\% respectively), and their unique flanking sequences (data not shown). Testing each by PCR against a panel of somatic cell hybrids allowed the two microsatellites to be ordered with respect to each other (section 2.7.8.3). DXS573 is present in hybrid SIN176 (data not shown), as is DXS255 (Lafreniere et al 1991), whereas DXS722 is absent, as is DXS426 (Coleman et al 1990). Taking into account the FISH locations for DXS573 (Xp11.2) this indicates that this marker lies proximally to both DXS426 and DXS722. There was no detectable recombination in any of the families under study.
Figure 5.2 Representation of recombinant meiotic events. Open and filled circles represent alleles corresponding to one or other of the maternal chromosomes for informative loci. Horizontal bars indicate loci that were uninformative in a given individual.
between DXS426 and DXS722 by which to order these two markers. However, two-colour FISH experiments had previously placed DXS722 proximal to DXS426 (Kamakari et al. 1995). PCR analysis of a series of overlapping YACs from the region using DXS722 as an STS provided confirmation of this data which established that DXS722 is present in YAC ICRFy900F0501 (data not shown) containing OATL1 (section 1.8.4.1; a gift from Dr. A.P. Monaco) which does not cover the DXS426 region and is known to lie proximal to DXS426 (Schlessinger et al. 1993). The resulting suggested location for DXS722 is shown in Figure 5.3.

Figure 5.3 Set of YAC clones in Xp11.23 used to physically map DXS722

This schematic shows the relationship and orientation of the YACs used to physically order DXS722.

In conjunction with the haplotype analysis, therefore, this physical mapping information defines the order of markers proximal to DXS7 as Xpter-DXS7-DXS426-DXS722-DXS573-DXS255-Xcen.

5.3.5 Ordering the loci by ILINK order of preference

Finally, an ILINK order of preference was carried out on the loci. Two-point and haplotype analysis of markers proximal to DXS7 clearly indicated that they were unresolvable by genetic means, therefore the results presented here describe analysis of markers distal to DXS7. ILINK analysis of the original pedigree data produced a most likely order for the reference markers that was incorrect, thus all subsequent analysis was performed on the modified pedigree data, which generated the correct order among the reference markers, thus providing a more accurate ‘framework’ upon which to position the new markers. The most likely orders so obtained are listed in Table 5.5, along with the odds supporting these orders. From order 2 in Table 5.5 it is evident that markers CYBB and DXS1110 could not be resolved by this analysis, so CYBB was used in following calculations to represent the pair as it was most informative. An order was compiled by sequentially including the novel markers in three-point computations involving reference markers (orders 4-12 in Table 5.5). Four-point
Table 5.5: Best orders obtained from ILINK order analysis of loci distal to DXS7 with associated odds over next best order

<table>
<thead>
<tr>
<th>Order of loci</th>
<th>Odds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original pedigrees</td>
<td></td>
</tr>
<tr>
<td>1: DXS1110-DXS7-OTC</td>
<td>1.67 : 1</td>
</tr>
<tr>
<td>Modified pedigrees</td>
<td></td>
</tr>
<tr>
<td>1: DXS7-OTC-DXS1110</td>
<td>6.54 : 1</td>
</tr>
<tr>
<td>2: DXS7-OTC-CGD-DXS1110</td>
<td>10.3 : 1</td>
</tr>
<tr>
<td>DXS7-OTC-DXS1110-CYBB</td>
<td></td>
</tr>
<tr>
<td>3: DXS7-DXS574-OTC</td>
<td>23.3 : 1</td>
</tr>
<tr>
<td>4: DXS7-DXS574-CYBB</td>
<td>21.5 : 1</td>
</tr>
<tr>
<td>5: DXS7-DXS574-DXS556</td>
<td>1.67 : 1</td>
</tr>
<tr>
<td>6: DXS7-DXS574-DXS572</td>
<td>3.29 : 1</td>
</tr>
<tr>
<td>7: DXS574-DXS556-OTC</td>
<td>7.17 : 1</td>
</tr>
<tr>
<td>8: DXS574-DXS556-CYBB</td>
<td>12.9 : 1</td>
</tr>
<tr>
<td>9: DXS574-DXS556-DXS572</td>
<td>1.03 : 1</td>
</tr>
<tr>
<td>10: DXS574-OTC-CYBB</td>
<td>11.1 : 1</td>
</tr>
<tr>
<td>11: DXS574-OTC-DXS572</td>
<td>15.6 : 1</td>
</tr>
<tr>
<td>12: DXS574-CYBB-DXS572</td>
<td>145 : 1</td>
</tr>
<tr>
<td>13: DXS7-DXS574-OTC-CYBB</td>
<td>13.7 : 1</td>
</tr>
<tr>
<td>14: DXS7-DXS556-OTC-CYBB</td>
<td>16 : 1</td>
</tr>
<tr>
<td>15: DXS7-OTC-CYBB-DXS572</td>
<td>7.6 : 1</td>
</tr>
<tr>
<td>16: DXS7-DXS574-DXS556-OTC</td>
<td>33 : 1</td>
</tr>
</tbody>
</table>
analysis was finally performed via the computing facilities at HGMP to allow inclusion of all three reference loci with each of the novel markers.

From this analysis, the order of preference for the loci distal to DXS7 can be clearly deduced as:

\[
\text{Xpter-DXS572-[CYBB/DXS1110]-OTC-DXS556-DXS574-DXS7-Xcen}
\]

This confirms the two-point and haplotype data already described. The ILINK analysis could not separate the group DXS426, DXS722 and DXS573. However, these results taken in conjunction with the physical mapping studies give the final best order for these loci, with estimated genetic distances in cM, as:

\[
\text{Xpter-DXS572-7.1-CYBB-0.0-DXS1110-4.5-OTC-0.0-DXS556-1.5-DXS574-3.5-DXS7-12.5-DXS426-0.0-DXS722-0.0-DXS573-4.7-DXS255-Xcen}
\]

The data presented here indicate an overall genetic distance of 26.7cM between CYBB and DXS255. This is in close agreement with recently published data for a comparable region on the short arm of the X chromosome (27.9cM, Donnelly et al 1994; 24.5cM, Fain et al 1995).

### 5.4 Discussion

Despite the growing number of informative genetic markers placed at random in the genome, the need to saturate specific regions with highly informative genetic markers is often required to localise and clone specific disease genes. In such situations, unambiguous localisation of the component markers is of paramount importance.

The practical utility of the microsatellite markers isolated from cosmids mapped by FISH to proximal Xp in further characterising the disease gene critical regions for RP2 and RP3 necessitated their precise localisation within this interval. This was carried out by genetic mapping methods using genotype data from 14 XLRP families, complemented by a physical mapping approach using a panel of somatic cell hybrids and YACs from the region. Relative order and chromosomal orientation of markers was determined by analysis of individual recombination events and by two-point and multipoint linkage analysis between the microsatellites and markers that had been previously placed on X chromosome linkage maps (Willard et al 1994). Unambiguous recombination events between the microsatellites distal to DXS7 firmly established their relative order, while physical mapping revealed unique locations for the microsatellites proximal to DXS7. Recombination events occurring adjacent to the most terminal marker on the map were checked thoroughly (i.e. DXS572/CYBB and DXS573/DXS255). The integration of different mapping methods has enabled successful construction of a comprehensive map positioning these polymorphic markers in relation to key reference markers on proximal Xp.

Although orders and genetic distances were published for some of the reference markers, the combination of marker systems from different sources, and of different polymorphic type
meant that integration of existing genetic map data was difficult. Consequently, published distances were not used and we here generated our own genetic data using an independent subset of families. The consistency between the map constructed here and those generated by other investigators (notably Donnelly et al 1994 and Fain et al 1995), provides good confirmation of the existing data and also indicates the efficacy of the methods used to construct our map. More recently, several YAC contigs constructed in this region of proximal Xp have physically confirmed the order and positioning of three of the novel microsatellites mapped in this study (Nelson et al 1995-DXS722, DXS573, DXS556).

For meiotic maps to be sufficiently accurate for optimal utilisation, it is imperative that they be of both high resolution and high integrity. Disturbances in data integrity can result from sample mixups, incorrect interpretation of genotypes (misscoring of alleles) or data-entry errors. Data errors may have a dramatic effect on meiotic maps constructed at this level of resolution, most usually by introducing spurious crossovers which lead to an inflation of genetic distances, as reflected in the pairwise recombination fractions, and decrease support for the correct genetic order (Buetow 1991). Multipoint methods give greater weight to small distances, which increases sensitivity to double crossovers (Morton and Collins 1990). This can lead to an incorrect order when typing errors simulate double crossovers in a small interval and is the likely explanation for erroneous order generated when the discrepant OTC RFLP1 results were included in the ILINK analysis of locus order. A data-typing error for a biallelic marker informative for linkage will often be undetectable in an offspring. Such an error will likely appear as a double recombinant if flanking markers are also informative. Within a 1 cM interval, this double recombinant is approximately 10000 times less likely than no recombination events (Buetow 1991). The estimated genetic distance between DXS1110 and OTC is of the order of 1 cM (Nelson et al 1995) and their physical separation is only 270kb (Roux et al 1994) which lends support to the possibility that the double crossovers observed were due to OTC typing errors (although alternative explanations such as new mutations or gene conversion cannot be ruled out). This observation would account for the statistical support placed against the correct locus order within our subset of families. Removal of this OTC data from the linkage analysis subsequently led to a much improved genetic map. It may be advisable not to incorporate older genotyping data in linkage studies, unless it can be verified and corrected (Brzustowicz et al 1993).

It is to be noted, however, that the statistical support for the most likely orders in this study (Table 5.5) are not significant if framework criteria are applied. The discrepancies involving OTC were highlighted because its relative position on proximal Xp has been well-established by multiple physical mapping reports, work stemming primarily from to its close proximity to the RP3 gene (Nelson et al 1995). The low level of support for best orders generated in this study are mainly the result of two factors, which contribute to reduced occurrence and detection of crossovers (on which genetic mapping depends): a small dataset and inclusion of markers with relatively low heterozygosity (e.g. DXS572, OTC). The latter effect was also exacerbated by the allele recycling necessary to perform the ILINK multipoint analysis. These effects on the strength of the final map would indicate that it is prudent to
employ additional mapping methods when determining the precise location of novel markers using a small set of families.

Haplotype analysis provides a reliable means to establish locus orders when recombinant individuals are informative for the loci concerned, where the crossover minimisation strategy is applied (e.g. for marker DXS572). Physical mapping via STS-content mapping of large-insert clones containing DNA from the region can also be used to determine positional information for markers, and is particularly useful for ordering markers in small chromosomal subregions, where integration of the genetic and physical maps can simultaneously be achieved (e.g. DXS573 and DXS722 with DXS426 and DXS255). In general a combination of different mapping methods is desirable for local map construction with sets of closely spaced markers as this can provide validation of such comprehensive maps where statistical support is lacking. Power for ordering loci is very small for closely linked loci and reaches an optimum value at recombination rates of about 20% between loci (Ott 1991). An increasing density of marker loci is thus expected to lead to maps with well-supported order for not too closely-spaced loci, but ordering neighbouring loci may be difficult on purely genetic grounds.

The more accurate and reliable framework marker maps can only be generated by using much larger data sets. For this reason, most genome-wide, large-scale genetic maps have been constructed using the family panel assembled at the Centre d’Etude du Polymorphisme Humain (CEPH) in Paris. The standard set comprises 40 nuclear, three-generation families (so that grandparental genotypes can establish phase in the parents) for which many polymorphic markers have been typed (with an average resolution of 0.7cM; section 1.2.5), such that for most regions of the genome, typing any new marker will enable its position on the map to be discovered efficiently.

Despite the vast numbers of highly polymorphic markers now available for the human genome (section 1.2.5), integration of the various genetic maps is still incomplete, which means that precise locations of markers can be difficult to deduce, especially when different anchor markers have been employed. Furthermore, until recently, certain regions of the genome were still relatively scantily represented by the current genetic maps, particularly true of Xp21.1-Xp11.22 at the outset of this study. These two factors meant that utilising the CEPH panel would have not been especially advantageous for our purposes at that time, and using the XLRP pedigree genotype information already generated proved adequate for basic map construction.

The set of ordered crossovers so characterised in this study should provide valuable resources for mapping of additional markers as they are generated (e.g. CEPH/Genethon) to this region of proximal Xp in the form of a ‘mini-meiotic breakpoint panel’ in a method analogous to that in operation for the CEPH family panel, albeit on a much smaller scale (Elsner et al 1995; see chapter 6). By selecting key recombinant individuals (such as those represented in Figure 5.2) together with their parents and non-recombinant sibs (for phase determination), genotyping of such new markers should help close the ‘gap’ in proximal Xp and permit their more precise genetic localisation, such that their suitability for further genetic and physical mapping studies can be ascertained. In this regard it is fortuitous that most
breakpoints are represented in more than one individual and in different families, increasing the chances of each marker being informative and thus able to be mapped. Since this approach provides a means for examining which genotypes are affecting the location of each marker within the map, discrepancies can be observed easily.

The genetic map reported here incorporates 5 novel polymorphic microsatellites defined in physically mapped cosmids which should prove to be of great value in fine genetic mapping of the genes underlying RP2 and RP3 (see chapter 6) in addition to other disease loci assigned to this region. This map also provides landmarks that will greatly contribute to the construction of a detailed physical map for the Xp21.1-11.22 region and thus serve to integrate the developing physical map with the X chromosome linkage map. Of particular relevance to this study, DXS722 has been instrumental in the construction and orientation of a detailed YAC contig map in the RP2 critical region in Xp11.23-11.3, which will be described in chapter 7.
CHAPTER 6

Genetic Refinement of the RP2 Critical Region and
Exclusion of Candidate Gene TIMP-1 from a
Causative Role in the Disease

6.1 Introduction

6.1.1 Mapping the RP2 locus for X-linked retinitis pigmentosa

X-linked retinitis pigmentosa (XLRP) is the most severe clinical form of RP (section 1.9.5), with an incidence of about 1:20,000 (Jay 1982, Heckenlively 1988). Male XLRP patients generally develop concentric field loss before the 20th year of life leading to severe visual handicap by the age of 40 (Bird 1975). Female carriers show variable symptoms of the disease upon opthalmological testing; where present, fundus irregularities and/or ERG changes precede visual impairment usually by middle age (Arden et al 1983).

In the absence of functional clues as to the pathophysiology of XLRP, positional cloning strategies have been adopted to isolate the defective genes (section 1.1.1). Genetic linkage studies have clearly established genetic heterogeneity for XLRP, with two major loci at Xp11.3-11.23 (RP2) and Xp21.1 (RP3) (section 1.9.5.1). Although phenotypic differences have been reported between RP2 and RP3 (section 1.9.5.2), further studies have found no reliable means by which to distinguish between them clinically. Hence the identity of the particular XLRP locus segregating in each family can only be determined by the identification of recombinant meioses within the target region, via inspection of haplotypes (section 1.2.7). The close proximity of XLRP genes on proximal Xp makes such a genetic distinction of XLRP families difficult; nevertheless, from cumulative genetic data it appears that RP3 predominates in British and American families (section 1.9.5.1).

Precise localisation of the RP3 gene to Xp21.1 by genetic linkage analysis has been augmented by the molecular genetic analysis of RP3 patients with submicroscopic DNA deletions, confining RP3 to a 530kb stretch of DNA flanked by genetic markers OTC and DXS1110 (Roux et al 1994). At the outset of his study, several groups were actively engaged in analysing transcripts from this region as potential candidates for the disease (section 1.9.5.3.1), and the gene RPGR has recently been cloned and shown to underly a proportion of RP3 cases (Meindl et al 1996). We therefore focused our attention on genetic refinement of the RP2 locus, which has remained broadly localised to a ~13cM interval in Xp11.3-Xp11.22 flanked by DXS7 and DXS255 (section 1.9.5.3.2), owing to a lack of informative recombination events in the critical region, no detectable disease-associated chromosomal abnormalities, and until recently, the paucity of genetic markers. Furthermore, the results of
multipoint and heterogeneity analyses suggest different locations for the RP2 gene within this interval (6cM proximal to DXS7 in Xp11.23; Teague et al 1994, or a more proximal location in Xp11.22; Bergen et al 1995). Narrowing of the RP2 region is a necessary prerequisite toward characterisation of the gene and its mutations, and may help resolve the issue of clinical and genetic heterogeneity.

In this study haplotype analysis was performed on 14 XLRP families using 34 polymorphic markers from proximal Xp, firstly in an attempt to determine by molecular genetic means the particular XLRP gene segregating within each family, and secondly to firmly define and if possible refine the interval containing the RP2 gene. Accurate genetic localisation of flanking crossovers in families defined as RP2 was undertaken to find the minimal critical region for this gene in our data set. The use of subsets of XLRP families containing only potentially informative recombination events (a mini ‘meiotic breakpoint panel’) within the RP2 critical interval reduced the typing work-load involved in this analysis, helped order new markers that became available in the region of interest, and simultaneously led to substantial refinement of the RP2 gene critical interval.

6.1.2 Investigation of a role for TIMP-1 in the pathophysiology of RP2

The TIMP (tissue inhibitor of metalloproteinases) gene family play a significant role in the maintainance of connective tissue homeostasis. The extracellular matrix (ECM) is a complex and dynamic meshwork of proteins and proteoglycans that not only provides support to organisms, but also has a profound effect on many biological activities; cell proliferation and differentiation, cell adhesion, migration and tissue morphogenesis. Agents that modify the ECM thus have the potential to influence a wide variety of normal and pathological processes. ECM turnover is regulated by the balance between synthesis of new components and degeneration of existing structures. Matrix destruction is caused by the concerted action of an array of secreted proteinases, the most predominant of which are a family of zinc-dependant matrix metalloproteinases (MMPs) that includes collagenases, stromelysins and gelatinases (Matrisian 1990). The activity of these enzymes is controlled at several levels, of which an important one is through their interaction with specific TIMP inhibitors. Coordinated control over the expression of both MMPs and TIMPs via a small cast of nuclear transcription factors in response to agents such as serum growth factors is essential for connective tissue homeostasis and for the rapid localised remodelling that occurs during many normal physiological processes (Matrisian 1990). When such regulation is lost, an excess of MMP activity over protective TIMP activities causes tissue destruction and degenerative disease e.g. rheumatoid arthritis (McNaul et al 1990). In the developing neural retina, MMPs have been postulated to play a role in facilitating neurite extension (Sheffield and Graff 1991), and the presence of a TIMP-like protein has been reported in the interphotoreceptor matrix (Gallo et al 1994).

During the course of this study, mutations in the TIMP-3 gene on chromosome 22q13.1 were identified as the cause of Sorsby’s Fundus Dystrophy (SFD), a degenerative macula disease of the neural retina, pigment epithelium and choroid (Weber et al 1994). Diffuse
Bruchs membrane thickening reminiscent of the findings in SFD has been demonstrated in cases of XLRP (Santos-Anderson et al 1982). Retinal TIMP-3 levels are also dramatically increased in cases of simplex RP, suggesting that extensive remodelling of the retinal ECM is occurring in this ocular disease (Jones et al 1994). Another member of the TIMP gene family, TIMP-1 (or erythroid potentiating activity EPA), is located within intron 5 of the SYNI gene on chromosome Xp11.23 (section 1.8.4.1). TIMP-1 also has cell growth promoting activities, indicating that this is a multifunctional protein involved in controlling ECM breakdown and modulating cell growth (Hayakawa et al 1992). In addition, immunohistochemistry has demonstrated TIMP-1 protein expression in all nuclear layers of the retina (Johnson et al 1995), and MMPs and TIMP-1 have been identified in cultured human RPE cells (Alexander et al 1990). To date, the marker most closely linked to RP2 is DXS426 in Xp11.23 (Teague et al 1994). TIMP-1 lies approximately 200kb distal to DXS426 (Coleman et al 1994) and was therefore considered to be a good positional candidate for RP2. The complete genomic organisation of the TIMP-1 gene was elucidated in order to effectively screen all exons, exon/intron boundaries and 5' untranslated region (containing promoter and enhancer elements) for mutations in affected RP2 patients by PCR and direct sequencing.

6.2 Materials and methods

6.2.1 Haplotype analysis in XLRP families

6.2.1.1 Families

A panel of 14 XLRP families was employed for genetic analysis. Families F5, F11, F15, F21, F38, F45, F47, F51, F53, F75 and F83 are British families and were clinically ascertained through Moorfields Eye Hospital. Another British family, F74 was obtained via the genetics clinic at Newcastle General Hospital (Grace O’Hearn) and blood samples from Belgian families F71 and F72 were kindly donated by Dr. Lionel Van Maldergem (Institut de Pathologie et de Genetique, Loverval). The diagnosis of XLRP 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 psychophysical and electrophysiological tests, were taken to be XLRP heterozygotes (presumptive heterozygotes), otherwise the genetic status of females was not inferred unless the pedigree structure showed them to be obligate heterozygotes. DNA was extracted from venous blood samples by the methods described in section 2.2.1.

6.2.1.2 Genetic markers and haplotype analysis

34 polymorphic markers were used to type the families; their localisations are indicated in Figure 6.1, which also details genetic maps covering this region of proximal Xp. Fourteen of these markers are described in chapter 5 (Tables 5.1 and 5.2). Use of these earlier markers as outlined in chapter 5 had already established crossovers in these XLRP families, so that the
This order is based upon cumulative genetic and physical mapping data from many groups (Nelson et al. 1995) with some more recent additions: DXS977 has been mapped genetically by our laboratory between DXS574 and DXS228 (Kamakari et al. 1995) and DXS8080 is located on a YAC contig proximal to MAOA (chapter 7). Markers DXS8026 and DXS8083 are recent additions to the final version of the Genethon map (Dib et al. 1996). Pairs of markers for which order is unknown are bracketed. Markers with an asterisk to the left of the consensus map were of only approximate location; one of the aims of this work was to incorporate these markers into the genetic map of proximal Xp. Comparison of the various maps shows that the overall region from Xp21.1 to Xp11.22 covered by the entire marker set spans approximately 25cM. The interval between RP3 and the distal boundary of the RP2 interval is thus approximately 4-6cM, highlighting the importance of markers mapping to this interval in discriminating genetically between the two loci.
Figure 6.1: Different genetic maps spanning the RP2-RP3 regions on proximal Xp

- Wang et al. 1995
- Fain et al. 1995
- Donnelly et al. 1995
- Dib et al. 1996
- Thiselton et al. 1995

Xpter

<table>
<thead>
<tr>
<th>Marker</th>
<th>DXS572</th>
<th>CYBB</th>
<th>DXS1110</th>
<th>OTC</th>
<th>DXS1068</th>
<th>DXS556</th>
<th>DXS574</th>
<th>DXS977</th>
<th>DXS993</th>
<th>DXS228</th>
<th>DXS7</th>
<th>MAOA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Xp21.1

<table>
<thead>
<tr>
<th>Marker</th>
<th>DXS538</th>
<th>5.6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Xp11.4

<table>
<thead>
<tr>
<th>Marker</th>
<th>DXS8080</th>
<th>DXS8026</th>
<th>DXS8083</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6.4</td>
<td>8.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Xp11.3

<table>
<thead>
<tr>
<th>Marker</th>
<th>DXS7124</th>
<th>DXS337</th>
<th>DXS1146</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

RP3

<table>
<thead>
<tr>
<th>Marker</th>
<th>DXS1003</th>
<th>DXS1055</th>
<th>SYNL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.5</td>
<td>0.0</td>
<td>3.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

RP2

<table>
<thead>
<tr>
<th>Marker</th>
<th>DXS1367</th>
<th>DXS6616</th>
<th>DXS6941</th>
<th>DXS722</th>
<th>DXS1126</th>
<th>DXS573</th>
<th>DXS1039</th>
<th>DXS255</th>
<th>DXS1000</th>
<th>DXS988</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.9</td>
<td>1.2</td>
<td>5.0</td>
<td>5.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.5</td>
<td>6.4</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Xcen

<table>
<thead>
<tr>
<th>Marker</th>
<th>DXS538</th>
<th>DXS572</th>
<th>CYBB</th>
<th>DXS1110</th>
<th>OTC</th>
<th>DXS1068</th>
<th>DXS556</th>
<th>DXS574</th>
<th>DXS977</th>
<th>DXS993</th>
<th>DXS228</th>
<th>DXS7</th>
<th>MAOA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5.6</td>
<td>0.9</td>
<td>0.3</td>
<td>0.8</td>
<td>DXTS1068</td>
<td>6.1</td>
<td>5.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.8</td>
<td>0.6</td>
<td>0.1</td>
<td>0.0</td>
<td>DXTS1068</td>
<td>9.8</td>
<td>9.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Marker</th>
<th>DXS1003</th>
<th>DXS1055</th>
<th>SYNL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.5</td>
<td>0.0</td>
<td>3.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Marker</th>
<th>DXS1039</th>
<th>DXS255</th>
<th>DXS1000</th>
<th>DXS988</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.5</td>
<td>6.4</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>3.0</td>
<td>3.0</td>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>1.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>
more recently identified markers were used in various combinations in each family, to better localise the sites of these recombinations. Details of markers not previously described, all of which are microsatellites, are shown in Table 6.1. Genotyping was performed as in section 5.2.3, and haplotypes were constructed for all sampled individuals using the principle of crossover minimisation (section 5.2.5).

Table 6.1 Further microsatellites used for genetic characterisation of XLRP families

<table>
<thead>
<tr>
<th>Locus</th>
<th>Chromosomal location</th>
<th>Heterozygosity</th>
<th>Annealing T°C</th>
<th>Allele size range (bp)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>DXS1068</td>
<td>Xp21.1-11.4</td>
<td>0.82</td>
<td>55°C</td>
<td>245-259</td>
<td>Weissenbach et al 1992</td>
</tr>
<tr>
<td>DXS977</td>
<td>Xp11.4</td>
<td>0.52</td>
<td>50°C</td>
<td>148-162</td>
<td>Kamakari et al 1995</td>
</tr>
<tr>
<td>DXS993</td>
<td>Xp11.4-p11.3</td>
<td>0.79</td>
<td>55°C</td>
<td>292-312</td>
<td>Gyapay et al 1994</td>
</tr>
<tr>
<td>DXS538</td>
<td>Xp21.1-p11.21</td>
<td>0.79</td>
<td>55°C</td>
<td>154-184</td>
<td>Browne et al 1991</td>
</tr>
<tr>
<td>DXS8080</td>
<td>Xp11.3</td>
<td>0.68</td>
<td>58°C</td>
<td>177-191</td>
<td>Dib et al 1996</td>
</tr>
<tr>
<td>DXS8026</td>
<td>Xp11.3-p11.23</td>
<td>0.60</td>
<td>58°C</td>
<td>222-226</td>
<td>Dib et al 1996</td>
</tr>
<tr>
<td>DXS8083</td>
<td>Xp11.3-p11.23</td>
<td>0.74</td>
<td>58°C</td>
<td>167-173</td>
<td>Dib et al 1996</td>
</tr>
<tr>
<td>DXS7124</td>
<td>Xp11.3-p11.23</td>
<td>0.60</td>
<td>60°C</td>
<td>150-156</td>
<td>Coleman et al 1996</td>
</tr>
<tr>
<td>DXS337/2</td>
<td>Xp11.3-p11.22</td>
<td>0.67</td>
<td>54°C</td>
<td>163-193</td>
<td>Barker et al 1993</td>
</tr>
<tr>
<td></td>
<td>Xp11.3-p11.22</td>
<td>0.73</td>
<td>54°C</td>
<td>139-145</td>
<td>Barker et al 1993</td>
</tr>
<tr>
<td>DXS1055</td>
<td>Xp11.3-p11.23</td>
<td>0.72</td>
<td>58°C</td>
<td>81-93</td>
<td>Gyapay et al 1994</td>
</tr>
<tr>
<td>DXS1003</td>
<td>Xp11.3-p11.23</td>
<td>0.80</td>
<td>57°C</td>
<td>169-195</td>
<td>Gyapay et al 1994</td>
</tr>
<tr>
<td>DXS1146</td>
<td>Xp11.3-p11.23</td>
<td>0.45</td>
<td>56°C</td>
<td>130-142</td>
<td>Hong et al 1993</td>
</tr>
<tr>
<td>DXS1367</td>
<td>Xp11.3-p11.23</td>
<td>0.77</td>
<td>60°C</td>
<td>107-135</td>
<td>Schindelhauer et al 1994</td>
</tr>
<tr>
<td>DXS6616</td>
<td>Xp11.23</td>
<td></td>
<td>56°C</td>
<td>~117</td>
<td>Meindl 1996 (personal communication)</td>
</tr>
<tr>
<td>DXS6941</td>
<td>Xp11.23</td>
<td>0.59</td>
<td>65°C</td>
<td>158-166</td>
<td>Meindl 1995 (personal communication)</td>
</tr>
<tr>
<td>DXS1126</td>
<td>Xp11.3-11.23</td>
<td>0.68</td>
<td>60°C</td>
<td>230-252</td>
<td>Donnelly et al 1994</td>
</tr>
<tr>
<td>DXS1039</td>
<td>Xp11.23-p11.22</td>
<td>0.61</td>
<td>56°C</td>
<td>93-103</td>
<td>Gyapay et al 1994</td>
</tr>
<tr>
<td>DXS1000</td>
<td>Xp11.22-p11.21</td>
<td>0.35</td>
<td>55°C</td>
<td>230-236</td>
<td>Gyapay et al 1994</td>
</tr>
<tr>
<td>DXS988</td>
<td>Xp11.22-p11.21</td>
<td>0.62</td>
<td>55°C</td>
<td>134-144</td>
<td>Gyapay et al 1994</td>
</tr>
</tbody>
</table>

6.2.1.3 Meiotic breakpoint panel

The haplotype data generated using markers of known order on proximal Xp revealed several key crossovers in the Xp11.3-Xp11.23 region (see Figure 5.2 and section 6.3.1). This 'meiotic breakpoint panel' presented a potentially valuable tool for fine genetic localisation and ordering of newly reported markers mapping broadly to Xp11.3-p11.23, both to assess their utility in refining the RP2 critical interval, and also to simultaneously direct YAC
contigging efforts (chapter 7). This procedure is outlined in section 5.4 and is used to define the order of new markers on existing genetic maps by genotyping chromosomes with meiotic breakpoints known to reside between previously mapped loci. Subsets of individuals from the relevant families were therefore genotyped with some such markers in an attempt to determine their precise genetic map location with respect to markers for which haplotype data was already available. Family members were chosen in order to establish allele phase in the recombinant individual (genotypic configuration of the alleles) and are highlighted with an asterisk in Figure 6.2 (section 6.3.1; for family F11, the subset of individuals comprised more than half the family, so that for some markers the entire family was genotyped). The crossovers within the meiotic panel divided the region between DXS8080 and DXS426 into 3 distinct intervals, as shown in Table 6.2. Segregation patterns of each test marker were compared with the patterns of the anchor markers flanking the breakpoints, minimising double crossovers. Markers mapped genetically with respect to existing markers in Xp11.3-11.23 using this panel are highlighted in bold in Table 6.1 and highlighted with an asterisk in Figure 6.1.

Table 6.2  Key crossovers in Xp11.3-11.23 comprising the meiotic breakpoint mapping panel

<table>
<thead>
<tr>
<th>Family</th>
<th>Recombinant individual</th>
<th>Location of crossover</th>
</tr>
</thead>
<tbody>
<tr>
<td>F11</td>
<td>IV-3</td>
<td>DXS1055 X DXS426</td>
</tr>
<tr>
<td></td>
<td>IV-10</td>
<td>DXS1110 X DXS426</td>
</tr>
<tr>
<td>F15</td>
<td>IV-9</td>
<td>DXS993 X DXS1003</td>
</tr>
<tr>
<td>F38</td>
<td>III-3</td>
<td>DXS8080 X DXS8026</td>
</tr>
<tr>
<td></td>
<td></td>
<td>/DXS8083</td>
</tr>
<tr>
<td>F47</td>
<td>III-2</td>
<td>DXS1003 X DXS426</td>
</tr>
<tr>
<td>F51</td>
<td>II-3</td>
<td>DXS1055 X DXS426</td>
</tr>
<tr>
<td>F71</td>
<td>III-1</td>
<td>DXS8083 X DXS1055</td>
</tr>
<tr>
<td>F72</td>
<td>II-5</td>
<td>DXS8083/ X DXS1003</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DXS8026</td>
</tr>
<tr>
<td>F74</td>
<td>IV-5</td>
<td>MAOA X DXS1003</td>
</tr>
<tr>
<td></td>
<td>IV-6</td>
<td>DXS1055 X DXS722</td>
</tr>
<tr>
<td>F75</td>
<td>II-2</td>
<td>DXS8083 X DXS1003</td>
</tr>
<tr>
<td>F83</td>
<td>III-5</td>
<td>DXS8080 X DXS1003</td>
</tr>
</tbody>
</table>

Summary of genetically separable intervals

<table>
<thead>
<tr>
<th>DXS8080 X DXS8026 X DXS1003 X DXS426</th>
</tr>
</thead>
<tbody>
<tr>
<td>DXS8083 X DXS1055</td>
</tr>
</tbody>
</table>
6.2.2 Molecular genetic analysis of RP2 candidate gene TIMP-1

6.2.2.1 Retinal expression of TIMP-1

To confirm TIMP-1 expression in the retina, PCR primers were designed to the published cDNA sequence (accession X03124; F 5'-'TGGCTTCTGGCATCTCTGGTGTG-3' and R 5'-'GGCAGGAATTCAGGCTATCTGGGAC-3') to amplify a 613bp fragment covering exons 3-5 of the TIMP-1 transcript. These primers were tested against phage lysates from 2 retina cDNA libraries; an adult retina cDNA library in λZapII (Stratagene) and an adult retina cDNA library in λgt10 (gift from J. Nathans), together with a ‘no DNA’ negative control. PCR cycling conditions were 30 cycles of 94°C for 1 min, 68°C for 30 sec, 72°C for 30 sec, with a final extension of 72°C for 5 min. PCR amplification products were visualised by electrophoresis on a 2% agarose gel along with size standard φX174/HaeIII, followed by staining in EtBr and UV illumination.

6.2.2.2 Completion of the TIMP-1 gene genomic organisation

Several entries for human TIMP-1 were retrieved from Genbank; X03124, the original cDNA sequence; D11139, a partial genomic sequence encompassing exon 3 to exon 6; D26513, 5' flanking region. Positions, sequence and size of introns 1 and 2 were undetermined. After inferring the genomic structure of the gene from the afore-mentioned database entries and donor/acceptor splice site consensus, exon-specific primers were designed to amplify introns 1 and 2. Primer pair I1F 5'-CCACGAGCCAGAGAGACC-3' (in exon 1) and I1R 5'-TGCCAGAAGCCAGGGCTCA-3' (in exon 2) was used to amplify intron 1 from genomic DNA. Primer pair I2F 5'-CCACAGAGCCCTCTGCAAT-3' (exon 2) and I2R 5'-GGTCTGGTTGACTTCTGGTGTCC-3' (in exon 3) was used to amplify a region of genomic DNA containing intron 2. PCR cycling conditions used were 30 cycles of 94°C for 1 min, 62°C for 2 min and 72°C for 1 min, with a final extension of 72°C for 5 min. Amplified introns produced fragment sizes of approximately 1.0kb spanning intron 1 and 1.6kb spanning intron 2. The PCR products were purified using Microspin S-400 HR columns (Pharmacia Biotech) and partial sequences were obtained by direct sequencing (section 2.6.3) using the primers described above. PCR primers could then be designed from this intronic sequence to permit PCR amplification of exons 1, 2 and 3.

6.2.2.3 Screening for TIMP-1 mutations in RP2 patients

DNA from one affected male patient and one unaffected male (as a control) were selected from families defined as RP2 by the detection of key recombinants via haplotype analysis i.e. F15, F38 and F72 (section 6.3.1). Primer pairs were designed to amplify all 6 exons, exon/intron boundaries and the 5'UTR containing promotor and enhancer elements. PCR reaction conditions were as follows: 30 cycles of 94°C for 1 min, 62°C for 1 min and 72°C for 1 min with a final extension of 72°C for 5 min. PCR products were purified using
Microspin S-400 columns (Pharmacia Biotech). Primers designed for PCR amplification were also used to directly sequence the PCR products, using the Pharmacia T7 DNA polymerase sequencing kit (section 2.6.3). Sequencing reactions were fractionated through 6% denaturing polyacrylamide gels and visualised by autoradiography (section 2.4.2).

6.3 Results

6.3.1 Haplotype analysis on 14 XLRP families

The pedigrees for the XLRP families analysed in this study are shown in Figure 6.2(a-n), together with the marker haplotypes for sampled members and key recombination events. No deletions were detected using a total of 34 genetic markers spanning Xp21.1-Xp11.22 (i.e. the interval embracing the RP3 and RP2 genes). The results for each family will be presented in turn:

a) Family F5

In this XLRP family, individual III-7 is a recombinant with respect to her carrier mother II-4 between markers DXS556 and DXS574 (Figure 6.2a). The phase of alleles in III-7 and II-4 was established from analysis of obligate carrier III-6. Although not examined in recent years, recombinant female III-7 was diagnosed as unaffected in middle age, when her sister III-6 showed clear fundus abnormalities. On this basis it appears that the crossover defines the proximal boundary of the XLRP interval in this family, with the disease-associated haplotype located distal to DXS574, which indicates that the XLRP gene segregating in F5 is not RP2. Further genetic markers distal to CYBB need to be tested on this family in order to determine whether the disease locus corresponds to RP3, RP6, RP15, or an as yet unidentified locus.

b) Family F11

Close inspection of the haplotypes of affected males in family F11 (Figure 6.2b) reveals that in fact only 3 markers consistently segregate with the disease phenotype, OTC, DXS556 and DXS977. Haplotype reconstruction in this family necessarily required the inference of recombination events in affected male members II-1 and II-3, for whom no DNA could be obtained. A key crossover was detected in affected male IV-9 between markers DXS977 and DXS228, with the disease-associated haplotype located distal to the crossover. DXS228 therefore defines the proximal boundary of the XLRP interval and this recombination event excludes the possibility that F11 is an RP2 family. Female IV-10 appears to have undergone a double recombination event such that she has inherited a portion of her mother’s affected chromosome between markers DXS1003 and DXS255 such that she has inherited the affected alleles at DXS573, DXS722, DXS337 and DXS426. The clinical status of this female is unknown at present, therefore this data cannot help with disease gene localisation. However, the crossover in male IV-9, who is definitely affected, provides
Figure 6.2 Extended genetic haplotypes for the 14 XLRP families constructed with the markers boxed to the left of each pedigree. Black symbols denote affected individuals; white symbols denote unaffected individuals; carrier females are depicted by a black dot; a question mark denotes females of unknown status. The haplotype of the disease-bearing chromosome is boxed; hatched boxes represent markers that were uninformative. Crossovers between unaffected chromosomes are indicated by an 'X'. Asterisks mark individuals comprising the meiotic breakpoint mapping panel.

Figure 6.2a: Family F5

Figure 6.2b: Family F11
Figure 6.2c: Family F15

<table>
<thead>
<tr>
<th>Markers</th>
<th>Channel 1</th>
<th>Channel 2</th>
<th>Channel 3</th>
<th>Channel 4</th>
<th>Channel 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>DXS572</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>DXS538</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>CYBB</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>OTC(PCR)</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>DXS556</td>
<td>10</td>
<td>10</td>
<td>6</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>DXS574</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>DXS538</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>DXS977</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>DXS529</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>DXS7(probe)</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>DXS993</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>DXS228</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>DXS722</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>DXS1125</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>DXS73</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>DXS255</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

Markers not informative

DXS1110 DXS7124
DXS8080 DXS1055
DXS8083 DXS1146
DXS337/6 DXS988
DXS3072 (89-13)

* *
conclusive evidence that F11 is not an RP2 family, and would therefore suggest that female IV-10 is not an XLRP carrier. Further information regarding the location of the XLRP gene segregating in this family can be obtained by comparing the haplotypes of all affected males and obligate carrier females, whereupon it can be seen that a crossover has occurred between markers CYBB/DXS1110 and the disease (inferred in individual II-3). This would place the XLRP gene proximal to DXS1110, firmly establishing that F11 is an RP3 family.

e) Family F15

Linkage analysis of this large family by Bhattacharya et al in 1984 led to the first locus assignment for XLRP to probe L1.28 (DXS7) in Xp11.3, and was thus the first RP2 pedigree discovered. The family has been well-characterised clinically (Bhattacharya et al 1985) because it is one of only a few large RP2 families so far reported. DNA samples were obtained for 11 affected males, 8 unaffected males, 3 obligate carrier females, 3 presumptive carrier females and 7 females of unknown clinical status. Five individuals show recombination events which localise the XLRP gene in this family (Figure 6.2c). Three female individuals show crossovers defining the distal boundary, where allelic phase has been established by comparison with the haplotypes of affected and unaffected brothers. Obligate carrier III-3 has a crossover between markers DXS977 and DXS993, placing the XLRP gene proximal to DXS977 and thereby excluding RP3 or XLRP loci located more distally on Xp; presumptive carrier III-13 is recombinant between markers DXS7 and DXS8080 with the disease-associated haplotype proximal to DXS7, a result which has been reported previously (Coleman et al 1990) and which ascribes DXS7 as the distal flanking marker for RP2 in F15; a crossover in presumptive carrier III-17 between DXS993 and DXS228 positions the disease locus proximal to DXS993. Crossovers in 2 status-known males provide firmer evidence for RP2 segregation; between DXS993 and DXS228 in affected male III-19, placing XLRP proximal to DXS993 (i.e. RP2) and between DXS7 and DXS8080 in unaffected male IV-10, such that he has fortuitously inherited the unaffected haplotype proximal to DXS7.

d) Family F21

In this small XLRP pedigree, individual II-2 is recombinant with respect to his carrier mother I-1 between markers DXS1110 and OTC (Figure 6.2d). The phase of the maternal alleles was inferred from the haplotypes of her unaffected sons. The disease-associated haplotype in affected male II-2 is located proximal to DXS1110, therefore the XLRP locus in this family could be either RP2 or RP3 (although a different location more proximal on Xp or on Xq cannot be ruled out at this stage).

e) Family F38

In family F38 individual III-6 is a recombinant with respect to his carrier mother II-3 between markers DXS255 and DXS988 (Figure 6.2e). The phase of the maternal alleles (II-3) was deduced from the haplotypes of her affected son II-5 and obligate carrier sister II-1 (with
Figure 6.2d: Family F21

Markers

<table>
<thead>
<tr>
<th>Markers</th>
<th>CYBB</th>
<th>DXS1110</th>
<th>OTC(probe)</th>
<th>DXS556</th>
<th>DXS574</th>
<th>DXS228</th>
<th>DXS7(probe)</th>
<th>DXS1003</th>
<th>DXS1146</th>
<th>DXS5426</th>
<th>DXS1126</th>
<th>DXS255</th>
<th>DXS988</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>6</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

Figure 6.2e: Family F38

Markers

| Markers     | CYBB | DXS572 | DXS1110 | OTC(probe) | DXS556 | DXS574 | DXS977 | DXS228 | DXS7(probe) | DXS8080 | DXS8026 | DXS8083 | DXS337/6 | DXS1003 | DXS1146 | DXS5426 | DXS722 | DXS1126 | DXS573 | DXS255 | DXS988 |
|-------------|------|--------|---------|------------|--------|--------|--------|--------|-------------|---------|---------|--------|-----------|---------|---------|---------|--------|---------|--------|--------|--------|--------|--------|
|             | 1    | 4      | 2       | 1          | 4      | 2      | 1      | 4      | 2           | 4       | 2       | 1      | 4         | 2       | 4      | 2       | 1      | 4       | 2     | 1      |
|             | 2    | 2       | 2       | 5          | 6      | 5      | 6      | 5      | 6           | 5       | 6       | 5      | 6         | 5       | 6      | 5       | 6      | 5       | 6     | 5      |
|             | 1    | 4       | 2       | 2          | 2      | 2      | 2      | 2      | 2           | 2       | 2       | 2      | 2         | 2       | 2      | 2       | 2      | 2       | 2     | 2      |
|             | 1    | 4       | 2       | 2          | 2      | 2      | 2      | 2      | 2           | 2       | 2       | 2      | 2         | 2       | 2      | 2       | 2      | 2       | 2     | 2      |

Markers not informative

<table>
<thead>
<tr>
<th>Markers</th>
<th>DXS337/6</th>
<th>DXS722</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DXS573</td>
<td>DXS1003</td>
</tr>
</tbody>
</table>

Markers not informative

<table>
<thead>
<tr>
<th>Markers</th>
<th>DXS7124 (II-3)</th>
<th>DXS1039</th>
<th>DXS1000</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DXS573</td>
<td>DXS1039</td>
<td>DXS1000</td>
</tr>
</tbody>
</table>
whom she has to share her father’s haplotype). The disease-associated haplotype in affected male III-6 is located distal to DXS988, which therefore defines the proximal boundary of the XLRP interval in this family. A key crossover was also detected in female III-3 between DXS8080 and DXS1003, with the disease-associated haplotype located distal to DXS1003. Now depending on the clinical status of III-3, this recombination event would place the XLRP gene segregating in this family either distal to DXS1003, and therefore either RP2, RP3, RP6, RP15 (if she is a carrier), or proximal to DXS8080, in which case the segregating gene would be RP2 (if she is unaffected). The importance of female III-3 to the assignment of the XLRP locus in family F38 lead to her recent clinical re-examination at Moorfields Eye Hospital, whereby it was confirmed that she has no fundus abnormalities at age >40 years. This information thus enables us to define with reasonable confidence F38 as an RP2 family and places the defective gene between markers DXS8080 and DXS988. As DXS8080 has been physically mapped proximal to MAOA (chapter 7), this simultaneously refines the distal boundary of the RP2 interval.

f) Family F45

Two key recombinant individuals define the proximal boundary of the XLRP gene in this family (Figure 6.2f). Individual III-10 harbours a crossover between markers DXS228 and DXS8080, placing the disease gene distal to DXS8080. A crossover in obligate carrier III-7 similarly positions the disease distal to DXS8080. The results of haplotype analysis in family F38 established DXS8080 as the new distal flanking marker for RP2, which indicates that F45 cannot be an RP2 family. Genotyping of markers distal to CYBB may determine whether the XLRP locus segregating in F45 corresponds to RP3, RP6, RP15 or a different locus on Xp.

g) Family F47

Genetic mapping studies in family F47 have been described by Coleman et al (1990). In individual III-1 (Figure 6.2g), a recombination event between OTC and DXS7 suggested that the XLRP locus in this family lies proximal to OTC, hence the disease was deduced to be RP2. The crossover occurring in individual II-2 additionally placed the disease distal to DXS426. Coleman et al (1990) discussed the possibility that if the disease loci in F15 and F47 were identical, then the RP2 gene would be localised between DXS7 and DXS426. Our studies show that markers just distal to OTC (CYBB and DXS1110) are non-recombinant with the disease gene and in fact a point mutation has recently been found in exon 4 of RPGR that segregates with the disease in this family, allowing definitive classification as RP3 (A.F. Wright, personal communication).

h) Family F51

In family F51 there are 2 recombination events defining both the distal and proximal boundaries of the XLRP interval (Figure 6.2h). The phase of the maternal alleles was established by the haplotypes of affected and unaffected male members of generation II.
Individual II-2 is recombinant between markers DXS1110 and DXS1068 on comparison with his obligate carrier mother, with the unaffected haplotype proximal to DXS1110. As II-2 is unaffected, this crossover positions XLRP proximal to DXS1110, thereby excluding the RP6 and RP15 loci from a causal role in RP in this family. Female II-3 has a crossover between markers DXS1055 and DXS426, such that she has inherited the affected haplotype from her mother proximal to DXS1055. Upon clinical re-examination in November 1995, this female was found to have no fundal, electrophysiological or psychophysical signs of disease and is therefore considered unlikely to be a carrier. The overall haplotype data therefore localise XLRP in this family between DXS1110 and DXS426, so that the XLRP gene segregating could be either RP3 or RP2.

i) Family F53

The generation of haplotypes for all sampled individuals of this small pedigree revealed no recombination events from Xp21.1 to Xpll.22, therefore the XLRP locus segregating in this family could not be localised by this method (Figure 6.2i). However, F53 provides a good example of how genetic mapping studies may complement clinical data for counselling purposes with ‘at risk’ females in XLRP families. Individual II-4 suffers from night blindness and has mild RPE changes upon fundus ophthalmoscopy, features which demonstrate her status as an XLRP carrier. This is supported by the genetic analysis described here, as she can be seen to have inherited her mother’s affected chromosome. Although electrophysiologically and psychophysically normal, female II-3 was initially considered to be a presumptive XLRP carrier owing to RPE atrophy noted upon fundus examination. The haplotype data clearly shows that she has inherited her mother’s normal X chromosome, however, from Xp21.1 to Xpll.22. It may be that the XLRP locus segregating in F53 lies elsewhere on the X chromosome, but a more recent clinical examination diagnosed her RPE changes as ‘tigroid fundus’, a rare but normal variant, consistent with the genetic data. Analysis of further markers distal to CYBB may identify recombination events implicating the RP6 or RP15 loci, but F53 is of no further use at present for refined localisation of the RP2 gene. Efforts are underway to sample further members of the family who may harbour multiply informative recombination events in proximal Xp such that an XLRP locus assignment can be made.

j) Family F71

In this Belgian family, individual III-1 is a recombinant with respect to her obligate carrier mother between markers DXS8083 and DXS1003, with the disease-associated haplotype proximal to DXS8083 (Figure 6.2j). The phase of the maternal alleles was established from the haplotypes of her affected and unaffected sons, who are non-recombinant in this region. Here we have a situation similar to that seen for family F38, where the only key crossover of use in classifying the family has occurred in a female who is not an obligate carrier and whose clinical status is thus of key importance. If individual III-1 is not a carrier, this places the XLRP gene segregating in F71 distal to DXS1003, in which case no XLRP locus
assignment can be made. If, however, she is a carrier of the disease, the defective gene must lie proximal to DXS8083, categorising F71 as an RP2 family, and substantially refining the distal boundary of the RP2 critical interval. Re-examination of the case notes of individual III-1 (Dr. Van Maldergem) revealed that she had mild pigment irregularities in the peripheral fundus similar to that observed in her mother. Considering the family history, it was concluded that III-1 might well be a carrier of XLRP. On this basis the haplotype data would suggest that this is indeed an RP2 family, with a new distal flanking marker DXS8083 in Xp11.23-11.3, although this conclusion must be treated as tentative. We are in regular contact with our Belgian collaborators for further information regarding the clinical status of III-1 or that of any future offspring in order to consolidate these results.

k) Family F72

In family F72 the phase of the alleles in grandmother I-1 was based upon the haplotypes of her 5 children using the principle of crossover minimisation, to give 6 recombination events informative for localising the segregating XLRP gene (Figure 6.2k). The crossovers in individuals II-2 and II-4 have already been discussed in section 5.3.3 with regard to the genetic localisation of marker DXS572; both localise the disease gene proximal to DXS572, thus excluding RP6 or RP15 segregation in this family (thorough clinical examination of II-4 has confirmed that she is unlikely to be a carrier). A recombination event has occurred in affected male II-7 that places the disease proximal to DXSlll0, but does not enable us to distinguish between RP3 and RP2. However, affected male II-5 is a recombinant with respect to his mother between markers DXS8083 and DXS1003, such that he has inherited the affected haplotype proximal to DXS8083. This excludes the RP3 locus from a causative role here in XLRP, defines F72 as an RP2 family, and provides a new distal flanking marker for the RP2 critical region. In addition, the crossover that has occurred in female III-1, whom clinical data suggests is not a carrier, would also localise the disease proximal to MAOA, consistent with the unaffected portion of her haplotype. A recombination event in individual III-2, an unaffected male, defines the proximal boundary of the RP2 interval in this family. A crossover has occurred between markers DXS6616 and DXS6941, positioning the disease-containing interval distal to DXS6941. In summary, the haplotype data clearly establishes RP2 as the XLRP gene segregating in F72, and also, within one family substantially refines both the distal and proximal boundaries of the RP2 critical region.

l) Family F74

The haplotype data for this family indicate that it is not segregating RP2 (Figure 6.2l). Allele phase for the females in generation III has been inferred from other members of the family using the crossover minimisation principle. This procedure assumes a recombination event in either II-1 or II-3 between markers DXS8080 and DXS1003 (Figure 6.2l shows the situation if II-1 is the recombinant) such that the disease-associated haplotype lies distal to DXS1003. This allele configuration reveals a second recombination event in affected male IV-
5 localising the disease gene distal to DXS1005. Phase in female IV-6 is known as she has to have inherited her father’s haplotype, and obligate carriers III-3 and III-4 must have an identical affected haplotype inherited from their father. A crossover in female IV-6 between markers DXS1055 and DXS722 is therefore evident, with the affected haplotype distal to DXS722. Although her clinical status is unknown at present, all known XLRP loci map distal to DXS722 which suggests that she is probably an XLRP carrier. However, it is only by comparing the haplotype of affected individual IV-2 with that of his affected and obligate carrier relatives that information is gained regarding the localisation of the disease gene in this family. This comparison reveals that only markers CYBB, DXS1110 and DXS1068 are linked to the disease; with all markers proximal to DXS1068 recombination has occurred, most likely in either or both of individuals I-2 and II-2. Therefore although it is not possible to pinpoint precisely the individuals in which the crossover(s) has arisen, the data clearly excludes RP2 as the locus involved in disease in this family. Markers distal to CYBB may establish whether mutations in genes at either the RP3, RP6 or RP15 loci underly the phenotype.

m) Family F75

At first glance, the mode of inheritance in this RP pedigree is not obvious due to its small size (Figure 6.2m). A diagnosis of XLRP was made, however, on the early onset and severity of symptoms in affected male II-2, and the relatively mild phenotype displayed by his mother I-2. Although allele phase in I-2 could not be determined for markers proximal to DXS8083, it is evident from the haplotypes of individuals II-1 and II-2 that a recombination event has occurred between DXS8083 and DXS1003 localising the gene distal to DXS1003. An XLRP locus assignment cannot therefore be made at this stage. Testing markers more distally located on Xp and obtaining further family members may detect other crossovers that provide useful information.

n) Family F83

This five-generation family has only 2 affected males, both in generation IV, and is composed of two branches, with the disease gene transferred through carrier females in the top 3 generations (Figure 6.2n). Haplotypes were constructed with minimal crossovers by assuming that obligate carrier females III-1 and III-3 had the same non-recombinant affected haplotype. Two recombination events were then detected which localise the segregating XLRP gene. Affected male IV-4 is recombinant between markers DXS574 and DXS228, with the disease-associated haplotype distal to DXS228, which excludes RP2. A crossover in individual III-5, an unaffected male, is apparent upon comparison with the haplotypes of his sister III-3 and brother III-6, placing the disease gene distal to DXS1003. The results indicate that F83 is not an RP2 family, although incorporation of genotype data using more distal markers may resolve whether RP3, RP6 or RP15 underlies the disease.
Figure 6.21: Family F74

![Family F74 diagram with markers and genotypes.]

Markers not informative: DXS572, OTC(PCR), DXS714(III-3 and III-4), DXS1000.

Figure 6.2m: Family F75

![Family F75 diagram with markers and genotypes.]

Markers not informative: DXS572, DXS574, DXS1110, DXS8026, OTC(PCR), DXS1146, DXS556, DXS337/6, DXS1126.
Figure 6.2n: Family F83

Markers

<table>
<thead>
<tr>
<th>Marker</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
<th>VI</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYGB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OTC(PCR)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DXS556</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>3.3</td>
<td>3.6</td>
<td>3.8</td>
</tr>
<tr>
<td>DXS74</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>DXS228</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>DXS71(A)</td>
<td>1</td>
<td>5</td>
<td>3.8</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>MAOA</td>
<td>1.7</td>
<td></td>
<td>6</td>
<td>11</td>
<td>15</td>
<td>3</td>
</tr>
<tr>
<td>DXS8000</td>
<td>3.6</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>DXS1003</td>
<td>5.6</td>
<td>3.8</td>
<td>6</td>
<td>11</td>
<td>15</td>
<td>3</td>
</tr>
<tr>
<td>DXS1055</td>
<td>5.8</td>
<td></td>
<td>6</td>
<td>11</td>
<td>15</td>
<td>3</td>
</tr>
<tr>
<td>SYN1</td>
<td>3.8</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>DXS7124</td>
<td>6</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>DXS3778</td>
<td>6</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>DXS3772</td>
<td>6</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>DXS722</td>
<td>3.3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>DXS1126</td>
<td>13</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>DXS73</td>
<td>5.4</td>
<td></td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>DXS1039</td>
<td>5.5</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>DXS988</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

Markers not informative

<table>
<thead>
<tr>
<th>Marker</th>
<th>II</th>
<th>IV</th>
<th>VI</th>
</tr>
</thead>
<tbody>
<tr>
<td>DXS572</td>
<td></td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>DXS1146</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>DXS426</td>
<td></td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>DXS1000</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>
In summary, genetic haplotype characterisation of these 14 XLRP families using polymorphic markers spanning the critical regions for RP3 and RP2 has enabled us to designate one as RP3 (F11) and 4 as RP2 (F15, F38, F71 and F72). The overall results for all families studied are summarised in Table 6.3. Analysis of recently published markers in recombinant individuals from families F71 and F72 has also led to the definition of new proximal and distal boundaries for the RP2 gene critical interval and significantly reduce its size from ~13cM to ~5cM.

Table 6.3 Summary table of XLRP locus assignments

<table>
<thead>
<tr>
<th>FAMILY</th>
<th>RP2</th>
<th>RP3</th>
<th>RP6</th>
<th>RP15</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>F5</td>
<td>-</td>
<td>-</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>F11</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>F15</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>F21</td>
<td>?</td>
<td>?</td>
<td>-</td>
<td>-</td>
<td>?</td>
</tr>
<tr>
<td>F38</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>F45</td>
<td>-</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>F47</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>F51</td>
<td>?</td>
<td>?</td>
<td>-</td>
<td>-</td>
<td>?</td>
</tr>
<tr>
<td>F71</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>F72</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>F74</td>
<td>-</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>F75</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>F83</td>
<td>-</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

6.3.2 Genetic localisation and ordering of new markers within the RP2 critical region using a meiotic breakpoint panel

The set of ordered crossovers comprising the meiotic breakpoint panel proved very useful in localising new markers to specific intervals within the RP2 critical region. The results are summarised diagrammatically for markers DXS538, DXS1146, DXS7124 and DXS337 in Figure 6.3, which shows the distribution of recombination events. In the following discussion, recombinant numbers refer to the number at the top of each meiotic recombination event shown in Figure 6.3.

Upon testing marker DXS538 in family F15 it immediately became clear that it mapped outside the RP2 critical interval (e.g. F15; III-3 and III-19 where it cosegregated with
Figure 6.3: Ordering markers genetically using the meiotic breakpoint mapping panel

<table>
<thead>
<tr>
<th>Markers</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>DXS572</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DXS538</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYBB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OTC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DXS1068</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DXS228</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAOA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DXS8080</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DXS8083</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DXS1146</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DXS1003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DXS1055</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DXS337/2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DXS337/6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DXS7124</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DXS426</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DXS722</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DXS1126</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

KEY
- ● disease-associated haplotype
- ○ unaffected haplotype
  - marker not informative; not tested
- * key recombinants for ordering markers in bold by haplotype analysis

SUMMARY
- DXS572
- CYBB
- OTC
- DXS538
- DXS1068
- DXS228
- MAOA
- DXS8080
- DXS8083
- DXS1146
- DXS1003
- DXS1055
- DXS337/2
- DXS337/6
- DXS7124
- DXS426
- DXS722
- DXS1126
markers distal to DXS993 and DXS228 respectively). Although mapping to the Xp11.3 region in the 6th XCW report (see Figure 1.9), DXS538 was localised distal to both DXS1068 and OTC by the recombination events seen respectively in individuals IV-1 in family F15 (Figure 6.2c) and recombinant 5 in Figure 6.3, and proximal to DXS572 by recombinant 8. This is consistent with the current MIT/Genethon map (Hudson et al 1995; http://www-genome.wi.mit.edu/) which places DXS538 between markers DXS1219 and DXS8102, which both lie in Xp21.1 and are separated by 6.5cM. Although the physical location of DXS8102 has not been reported, DXS1219 maps within the DMD gene. This data firmly places DXS538 outside the region of RP2, and enabled us to exclude DXS538 from our physical mapping study of the RP2 critical region (see chapter 7). It does, however, represent a useful marker for the genetic discrimination of XLRP families segregating RP3 or RP6, as it maps between the two loci.

The other 3 markers were all mapped to the Xp11.3-11.23 interval using the meiotic breakpoint panel. The positioning of these markers with respect to markers of known order is summarised in Figure 6.3. The reported localisation of DXS1146 was achieved by linkage alone (Hong et al 1993; 4.6cM proximal to DXS7 and 2.3cM distal to DXS255) using the CEPH reference pedigrees (section 5.4). DXS1146 is unfortunately not informative in most of the key recombinants in our dataset (heterozygosity 0.45), precluding its precise placement by haplotype analysis. Nevertheless, 2 individuals (recombinants 3 and 2) were sufficient to position this marker between DXS8080 and DXS426, simultaneously (in recombinant 2) placing DXS1146 distal to DXS337. This firmly placed DXS1146 in Xp11.3-p11.23 in the proximity of the RP2 gene distal flanking marker DXS8083, confirming its potential value for use both in refining the RP2 gene in future families, and as an STS in constructing a physical map of the RP2 gene critical region (chapter 7).

DXS337 consists of two microsatellites isolated from separate restriction fragments of a single phage clone RX9 (which limits their maximum separation to ~20kb). Barker et al (1993) physically localised RX9 to Xp11 with a genetic map position between DXS7 (Xp11.3) and DXS255 (Xp11.22), an interval of approximately 13cM. The 2D X chromosome genetic map of Fain et al (1995) gives a partial refinement, locating DXS337 between MAOA and DXS255, with one recombinant placing it distal to DXS1126 (Xp11.23). Testing both DXS337 microsatellites against the crossovers present in our meiotic breakpoint panel (for maximum information) the precise genetic position of this marker was established. One recombination event localised DXS337-6 distal to DXS426 (Xp11.23) in family F51 (recombinant 6 in Figure 6.3) and another in family F74 (recombinant 11), where both microsatellites were informative, positioned DXS337 proximal to DXS1055 (Xp11.3-p11.23). This distal refinement is supported by 3 recombinants placing DXS337 proximal to DXS8083 (7, 9 and 12) and 3 crossovers placing DXS337 proximal to DXS1003 (recombinants 1, 2 and 4). The DXS1055-DXS426 interval spans only 2.5cM (Fain et al 1995), thus the genetic location of DXS337 has been substantially refined by this procedure. With regards to the localisation of RP2, in the 2 families presented with distal flanking marker DXS8083 earlier
(section 6.3.1; F71 and F72), the crossovers are located between DXS8083 and DXS1003 (i.e. distal to DXS337) such that DXS337 lies within the RP2 critical interval but does not refine the distal boundary. It does, however, become a very useful marker for testing new XLRP families with RP2 designation and was instrumental in constructing a YAC contig in the RP2 region, where its newfound genetic relation to other markers in Xp11.3-p11.23 proved of primary importance (chapter 7).

DXS7124 is a microsatellite isolated from a cosmid containing the UBE1 gene (Coleman et al 1996). Although UBE1 is known to map to Xp11.3-p11.23, there has been some discrepancy in the literature regarding its position in relation to other markers in the region based on YAC STS-content mapping. As discussed in (section 7.3.3.2), this is most likely due to chimaerism or internal deletion/rearrangement of some or all of the YACs involved. The close proximity of DXS7124 and UBE1 (maximum separation ~35kb) afforded an alternative route to ordering UBE1 in relation to other markers in Xp11 via genetic mapping. To assist in our own physical mapping efforts (chapter 7) and in attempt to determine whether UBE1 lay within the RP2 critical region, DXS7124 was therefore tested against the meiotic breakpoint panel. Figure 6.3 shows 3 key recombination events informative for positioning DXS7124; between DXS8083 and DXS1055 in recombinant 7, where DXS7124 mapped proximal to DXS8083 and therefore within the RP2 interval, between DXS1055 and DXS426 in recombinant 6, with DXS7124 mapping distal to DXS426, and between DXS1003 and DXS426 in recombinant 4, where DXS7124 showed cosegregation with DXS426 and is thus positioned proximal to DXS1003. This places DXS7124 and therefore UBE1 well within the RP2 critical region defined by family F72 (at least 1.9cM, the genetic distance between DXS8083 and DXS1003), supporting the inclusion and characterisation of UBE1-containing YACs in a physical map of the RP2 region (see chapter 7). Furthermore, until further refinement of the RP2 critical is achieved, UBE1 becomes a potential candidate for the disorder (see chapter 8).

6.3.3 Investigation of TIMP-1 as a candidate for RP2

6.3.3.1 Confirmation of TIMP-1 gene expression in the retina

PCR amplification of phage lysates from 2 retina cDNA libraries was performed using primers designed to the TIMP-1 cDNA sequence (section 6.2.2.1). The successful amplification of a product of the expected size (613bp) from both libraries (data not shown) confirms the presence of the TIMP-1 transcript, verifying that the TIMP-1 gene is expressed in retinal tissue.

6.3.3.2 Genomic organisation of the human TIMP-1 gene

When point mutations or small deletions/insertions in protein-coding genes lead to disease, it is generally found that the underlying mutation lies within the coding portion of the gene, or affects a splice junction or a promoter element. Small deletions/insertions may
interrupt the open reading frame of the transcript, causing a frameshift, which results in a jumbled message downstream of the mutation and an aberrant protein product. Point mutations causing genetic disease introduce an alternative base at a conserved residue such that the resultant protein is malfunctional due to altered tertiary structure or charge. Both frameshifts and point mutations can also introduce a stop codon, resulting in a truncated protein or its absence altogether, or affect the gene promoter and hence its expression. Mutations occurring in the exon/intron boundaries will affect the normal splicing of introns that gives rise to the gene transcript, usually by deleting an exon altogether. Promotors and enhancers are upstream regulatory elements composed of short sequence motifs each with a specific function in conferring inducibility, tissue specificity or general enhancement of transcription and are usually binding sites for nuclear proteins. When investigating a candidate gene for mutations in affected individuals it is therefore important to screen the entire coding region, all intron/exon splice junctions and promoter elements. In order to screen these regions of the TIMP-1 gene for mutations in RP2 patients (section 6.3.3), sequence from the ends of introns 1 and 2 needed to be obtained so that appropriate PCR primers could be designed to amplify exons 1, 2 and 3. Primers were designed from the cDNA sequence to amplify introns 1 and 2 from genomic DNA (section 6.2.2.2). Figure 6.4 shows PCR amplification products spanning introns 1 and 2 which were then used as templates for direct double-stranded sequencing. PCR primers were designed from the sequence obtained (corresponding to the ends of introns 1 and 2) to amplify exons 1, 2 and 3. These primers are listed in Table 6.4 along with primers designed to amplify the remaining exons and promoter region. The completed genomic structure of the human TIMP-1 gene is shown schematically in Figure 6.5, and the intron/exon boundary sequences are given in Table 6.5.

6.3.3 TIMP-1 mutation screen in RP2 patients

The small size of the TIMP-1 gene facilitated a comprehensive approach to mutation screening via direct sequencing. Primers designed for PCR amplification were also used to directly sequence the PCR products from one affected and one unaffected male from each family classified unambiguously as RP2 upon haplotype analysis (section 6.3.1). Screening of all exons, exon/intron boundaries and ~750bp of 5'UTR containing promoter and enhancer elements revealed no disease-specific sequence anomalies. Only one apparently neutral polymorphism was identified due to a silent base substitution in exon 5 at nucleotide position 434 (accession X03124) and is shown in Figure 6.6. This base change converts TTC phenylalanine to TTT phenylanaline which is presumably a result of the common C to T conversion at CpG dinucleotides and would be expected to have no deleterious effect on the structure/function of the resultant protein.
Figure 6.4 Agarose gel photograph showing PCR amplification products spanning introns 1 and 2 of the human TIMP1 gene, using human genomic DNA as template. By comparison with size marker φX174/HaeIII, sizes of introns 1 and 2 were estimated at ~1.0kb and ~1.6kb respectively.
<table>
<thead>
<tr>
<th>Region amplified</th>
<th>Primer pairs (sense/antisense, 5' to 3')</th>
<th>Fragment size</th>
<th>Annealing temp.</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'UTR 1</td>
<td>AACCGGTACCCTCTCAGAGAAGTGGAGGAGATAAGGCAG</td>
<td>268 bp</td>
<td>62 °C</td>
</tr>
<tr>
<td>5'UTR 2</td>
<td>GACTCTGTGATCCTGGGTAAGAGTCTCCAGAATCTATCCAAATCA</td>
<td>279 bp</td>
<td>62 °C</td>
</tr>
<tr>
<td>5'UTR 3</td>
<td>CACAATAAAATGTTGAAGGGCTGAAGTGACCGCTAGAGGATTTATGT</td>
<td>383 bp</td>
<td>62 °C</td>
</tr>
<tr>
<td>Exon 1</td>
<td>GATGAGTAATGCATCCAGGAGCGAGCGCAAGGCTGAGTAGA</td>
<td>255 bp</td>
<td>62 °C</td>
</tr>
<tr>
<td>Exon 2</td>
<td>GGTGGATGACCTGCCAGGAGCTGGCACTGCTGAGTAGA</td>
<td>299 bp</td>
<td>62 °C</td>
</tr>
<tr>
<td>Exon 3</td>
<td>GGCTCATGCAGCTCCATTTGACGAGAACATGCTGGGAAAGAGGC</td>
<td>163 bp</td>
<td>62 °C</td>
</tr>
<tr>
<td>Exon 4</td>
<td>TGGTGCGAGAAAGTTGGCTGGAGCTGTTGGTGTCAGAGCAG</td>
<td>210 bp</td>
<td>62 °C</td>
</tr>
<tr>
<td>Exon 5</td>
<td>ATCACAAAGCTGCTTGCTGGTCCGAGGATCTCCGGTGGAG</td>
<td>209 bp</td>
<td>62 °C</td>
</tr>
<tr>
<td>Exon 6</td>
<td>GCCCGCGCTCAGTGTGGAAAGAACCAACATGGCATTCTCATC</td>
<td>459 bp</td>
<td>62 °C</td>
</tr>
</tbody>
</table>
Figure 6.5  Genomic structure of the human TMP1 gene (not to scale)

Sequence not determined
<table>
<thead>
<tr>
<th>Exon</th>
<th>3' Splice Junction*</th>
<th>Exon Size</th>
<th>5' Splice Junction*</th>
<th>Intron Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>43bp</td>
<td>GAGgtaagc</td>
<td>~1000bp</td>
</tr>
<tr>
<td>2</td>
<td>cattggcccaacagAAC</td>
<td>129bp</td>
<td>TCGgtgagt</td>
<td>~1600bp</td>
</tr>
<tr>
<td>3</td>
<td>tcctctctgcagTCA</td>
<td>80bp</td>
<td>AAGgtatcc</td>
<td>188bp</td>
</tr>
<tr>
<td>4</td>
<td>tcatccatcaacagATG</td>
<td>127bp</td>
<td>CTGgtgagg</td>
<td>205bp</td>
</tr>
<tr>
<td>5</td>
<td>gcctttctcttagGAA</td>
<td>125bp</td>
<td>ACAgtgagt</td>
<td>804bp</td>
</tr>
<tr>
<td>6</td>
<td>ctctctcttcagGTG</td>
<td>267bp</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Consensus) yyyyyyyyyynagyNNN NWGgtwrgt

*Capital letters represent coding sequence. Splice junction consensus is represented by w: A or T, r: A or G, y: C or T, s: G or C (Mount 1982)
Figure 6.6 Autoradiograph showing forward strand sequence of TIMP1 exon 5 from 3 affected RP2 patients of families F72, F38 and F15 and one unaffected control individual from family F15. Each of the dideoxy nucleotide (A, C, G, T) reactions were grouped for all individuals on the sequencing gel as shown, for ease of interlane comparison to facilitate detection of any changes. The C/T base change is indicated by an arrow; its presence in both affected and unaffected individuals suggests that this mutation is a neutral polymorphism.
6.4 Discussion

6.4.1 Haplotype analysis of XLRP families: refinement of the RP2 region

Evaluation of key recombination events is of critical importance in the fine mapping of disease genes. Identification of close flanking markers is virtually a prerequisite for the isolation and characterisation of mutated genes, particularly in the absence of candidate genes or focussing cytogenetic abnormalities. In an attempt to genetically define the position of the RP2 gene, the segregation of 34 polymorphic markers spanning the critical region for the two most prevalent forms, RP3 and RP2 (Xp21.1-Xp11.22) was analysed in 14 XLRP families by inspection of haplotypes. Key recombinants enabled XLRP locus assignment for 5 families, one as RP3 (F11) and 4 as RP2 (F15, F38, F71 and F72). The basis of RP in family F47 has since been found to be due to mutations in \textit{RPGR}, thus this dataset contains at least two RP3 families. The bias in detection of RP2 families in this dataset, despite it being rarer than RP3 in most populations studied (section 1.9.5.1), is a consequence of our interest in detecting RP2 families as a first stage toward defining the location of the gene. The high density of markers spanning the interval between the proximal flanking marker for RP3 (\textit{OTC}) and the RP2 interval also allowed us to exclude a further 4 families from segregating RP2 (F5, F45, F74 and F83). Haplotype analysis using genetic markers located more distally on Xp may resolve which of the RP3, RP6 or RP15 genes underly the affected phenotype in these families. The same applies to families F53 and F75, although such investigation in the absence of crossovers in the RP3 region will at best only exclude RP6 or RP15 and leave the question of RP2/RP3 unresolved, as is the situation after the analysis undertaken here for families F21 and F51. If further members of families of this type are unobtainable, it may be feasible to embark upon mutation screening for \textit{RPGR} which underlies RP3 in a growing number of cases (Meindl et al 1996). However, as microheterogeneity within RP3 is still a possibility, detection of no disease-associated mutations would not definitively implicate RP2.

The heterogeneity analysis of Teague et al (1994) was performed on 40 XLRP families, 9 of which were also examined in this study (F5,F11, F15, F21, F38, F45, F47, F51, F53). Interestingly, the XLRP locus designation differs for two families, F5 and F38. Our haplotype data suggests that F5 is not an RP2 family, whilst F38 does segregate RP2 (section 6.3.1). Teague et al (1994) give a 0.95 probability that F5 is an RP2 family and 0.67 probability that F38 has RP3. This discordance most likely rests on the fact that the critical crossovers defining the location of the XLRP gene in both families exists in females who are not obligate carriers. Recent re-examination of the case notes of III-7 in F5 (Figure 6.2a) and clinical re-examination of III-3 in F38 (Figure 6.2e) at Moorfields Eye Hospital confirmed the most likely clinical status of these women as indicated in Figure 6.2, thus we are confident that our locus designations are correct.
The designation of RP2 in family F38 also assumes that there are no as yet unreported XLRP located more proximally on Xp or even on Xq, as has been suggested by Aldred et al (1994). In this instance, it is fortuitous that in F72, both distal and proximal boundaries of the RP2 critical interval were evident, providing a much-needed focus upon which to direct positional cloning efforts to isolate the defective gene. Haplotype analysis of multiply informative crossovers established DXS8083 (Xp11.3-p11.23) and DXS6941 (Xp11.23) as the new flanking markers for RP2, spanning a genetic distance of approximately 5-6cM (the genetic distance between DXS8083 is not known directly but has been estimated from genetic distances to nearby markers; see Figure 6.1). This work has thus reduced the RP2 critical interval by at least half, and provides the first instance of both distal and proximal flanking markers being defined within the same family (proving that this locus definitely exists!). The degree of refinement now renders the construction of a physical contig spanning the RP2 critical region, and mapping/isolation of candidate transcripts therein, a more manageable task. Integration of the genetic data with a physical map of of Xp11.23 subsequently allowed the positioning of several genes and ESTs proximal to DXS6941, thereby excluding them from a causal role in the disease (see chapter 7). Concurrent genetic analyses undertaken by a colleague (R. Mark Hampson) on other XLRP pedigrees have revealed further refinement of the proximal boundary in a family with MAOA as the distal marker flanking the disease gene region and which could therefore be designated RP2. Assuming the defective gene is the same in both families, the combined data suggest a refined RP2 interval flanked by DXS8083 and DXS6616 (Xp11.23), spanning ~4-5cM. In the event that these two families are segregating different loci, microheterogeneity within this interval may only become evident once the disease gene for one of these families has been cloned (Thiselton et al 1996).

Several points regarding pedigree interpretation and haplotype construction are worth noting. In those cases where the structure of the family reveals only few meioses with known phase, a direct comparison of the haplotypes of each affected/obligate carrier individual can localise the minimal number of recombination events between the marker loci and the XLRP locus segregating and thus provide information as to the position of the disease gene (as in F11, F74, F83). Indeterminate allele phase can also inhibit precise crossover localisation in key females where the father’s haplotype is not available (e.g. in families F11 and F74). Furthermore, where key recombination events for localising the segregating XLRP locus occur in females who are not obligate carriers, clinical status is important but not 100% reliable. For defining the location of a progressive X-linked disease, as in XLRP, a positive result can therefore only be firmly established when the crossovers occur in males of age ~10 years or over, whose clinical status is indisputable. Affirmation of the disease diagnosis was also critical to the haplotype analysis described here; in particular, misdiagnosis of choroideremia families (which maps to Xq21 and with which XLRP shares many phenotypic characteristics; Bird 1975) can lead to fruitless investigations on Xp.

The female XLRP heterozygotes in the families comprising this study showed a wide range of clinical severity, which has been previously documented and is attributed to the
lottery of X-inactivation. Most carriers will in fact develop symptoms in later life owing to the deterioration of cells in which the normal X chromosome is inactivated. Carrier identification is therefore especially problematic in young daughters of female carriers as they are at risk of transmitting the disease to their sons and may not show retinal abnormalities. Genetic diagnosis can be used for counselling purposes on the basis of haplotype data where the segregating XLRP locus is known e.g. in RP2 family F15 (Figure 6.2c), we may predict that individuals IV-4 and IV-8 are likely to carry the defective gene, whilst IV-2, IV-6 and IV-7 are normal.

Some females are very severely affected (e.g. female III-7 in family F45 is registered blind in middle age) which is attributed to disproportionate X-inactivation of the normal allele of the XLRP gene in foetal cells destined to form the retina (Bird 1975). The usual randomness of X-inactivation makes every woman a mosaic in terms of X-linked gene expression. This is generally beneficial, as females heterozygous for detrimental X-linked mutations may show no effect of the mutant gene they express. The population of cells that express the normal allele often provides enough of an essential gene product to correct the defect in cells that express the mutant allele, or at least circumvent the lethality of a mutation. However, if by chance the X-inactivation pattern is non-random such that most cells express the mutant gene, females may manifest disease usually found only in males. The occurrence of partially affected females, resulting from mosaic inactivation of X-linked genes at the cellular level, has led some geneticists to question the applicability of the terms 'X-linked recessive' and 'X-linked dominant' used to describe the phenotypic inheritance of X-linked disorders. Rare families manifesting female-female transmission of X-linked traits, such as haemophilia B (Taylor et al 1991) may indicate a co-inherited defect in the X-inactivation centre (XIC, possibly the gene XIST; Penny et al 1996), resulting in the exclusive inactivation of the normal X. Alternatively, allelic variants at the XIC could induce skewed X-inactivation, as in the murine Xce (X-controlling element) alleles which appear to vary in strength (Cattanach et al 1975). The X-inactivation status of the X chromosome is related to the methylation of cytosine residues; analysis of differential DNA methylation at polymorphic X-linked loci provides a basis for distinguishing the active and inactive X chromosomes and is used for the study of biased inactivation in female carriers of some X-linked diseases (Hendriks et al 1992), which may be used in predictive counselling.

As demonstrated in section 1.7, genetic heterogeneity and allelism are evident in many inherited retinal degenerations. Hence the refinement of the RP2 critical interval described here may have significant implications for the localisations of other inherited X-linked disorders which have overlapping map locations in Xp11.3-Xp11.22 and which may be allelic to RP2 such as CSNBX (section 1.9.2). On the other hand, genetic refinement of the disease gene-containing intervals in families with these disorders may have a similar bearing on studies centred on RP2.
6.3.4.2 Exclusion of TIMP-1 from a causative role in RP2

The refined RP2 interval is known to be extremely gene-rich, consistent with a cytogenetic Giemsa-light band. Genes known to lie in this interval include ZNF21, ZNF41, ELKI and the (PFC-SYN1-TIMP1-ARAF1) gene cluster in Xp11.23 (section 1.8.4.1 and 1.8.4.2) and several CpG islands have been identified that correspond to as yet, unknown genes (Coleman et al 1994). None of these genes would appear to be strong candidates for RP2 on considering the etiology of the disease, although the recent discovery that mutations in the TIMP-3 gene cause Sorsby’s fundus dystrophy (Weber et al 1994), a macular degeneration, led us to investigate TIMP-1 for a causal role in RP2 (manuscript submitted). The expression of TIMP-1 in the retina has been established (Alexander et al 1990; Johnson et al 1995) and confirmed in this study by PCR amplification from retina cDNA libraries. Potentially a defect in this gene may lead to disruption of photoreceptor-matrix interactions and induce apoptosis (Frisch and Francis 1994), the mechanism responsible for photoreceptor cell death in several rodent models of RP (section 1.7.4). Therefore mutation screening in affected individuals from 3 RP2 families was performed by PCR and direct sequencing. Although labour intensive, sequence analysis is the most effective method for mutation detection as it both detects and defines all mutations within a DNA fragment simultaneously, so enabling its effect on gene expression/coding potential to be unravelled. As the TIMP-1 gene contains only 6 relatively small exons (see Figure 6.5), these advantages offset any extra manipulations involved.

Other methods exist for screening a gene for mutations in affected individuals, the most popular of which are ‘single strand conformational polymorphism’ or SSCP and ‘heteroduplex analysis’ which have proven useful for the detection of multiple mutations and polymorphisms (e.g. TIMP3 mutations in SFD, Weber et al 1994; rhodopsin mutations in RP, Keen et al 1991). These simple strategies visualise mutations in a DNA segment by revealing altered gel migration. In SSCP analysis, mutations cause the denatured DNA strands to adopt different secondary structures, thereby influencing their mobility during nondenaturing polyacrylamide gel electrophoresis (Orita et al 1989). In heteroduplex analysis, DNA from a potentially heterozygous individual is amplified by PCR and the products are denatured and allowed to renature to form heteroduplexes. Heteroduplex DNA molecules containing internal mismatches can then be separated from correctly matched molecules by electrophoresis in Hydrolink gels (Keen et al 1991). However, subtle differences between some mutant DNA fragments and their wild-type counterparts mean that these methods reliably detect only 80% of mutations in DNA fragments of optimal size; they also do not localise the mutations within the DNA fragment (Cotton 1993). The sensitivity of SSCP can vary greatly with gel matrix, temperature, DNA fragment size and sequence context (Liu and Sommer 1995). Despite the disadvantages, once optimal conditions are established, the simplicity of these techniques will ensure their future use, especially when genes containing large coding regions are to be analysed as they can easily be applied to a large number of samples.
Although it is possible that a pathogenic mutation in this gene might lie outside the regions included in our screen, the results of the present study effectively exclude mutations in the 5'UTR promoter region, coding sequence and adjacent splice sites of \textit{TIMP-1} from a causative role in the pathogenesis of XLRP. The precise role of TIMP-1 in maintaining retina structure and function has yet to be elucidated, and it remains to be seen what disease, if any, is caused by defects in the \textit{TIMP-1} gene.

\textbf{6.3.4.3 Genetic mapping and ordering of new markers within the RP2 critical region}

Using a meiotic breakpoint panel comprising defined recombinants from the XLRP families outlined in this study assembled for the Xp11.3-Xp11.23 region, 3 new previously unplaced genetic markers have been positioned relative to anchor markers assigned to this interval; markers DXS7124, DXS337 and DXS1146 have all been incorporated into the genetic map of Xp11.3-Xp11.23. Analysis with the new markers failed to identify a closer proximal boundary for the RP2 region, thus additional markers and families will be necessary to refine the position of the defective gene. The \textit{UBE1}-associated microsatellite DXS7124 was positioned proximal to DXS1003 in two multiply informative recombinants and therefore falls within the RP2 critical region. Several other genes have recently been mapped close to \textit{UBE1} in Xp11.3; \textit{PCTKI}, DXS8237E and \textit{ZNFI57} (section 1.8.4.2). These genes therefore become possible candidates for the disease and have been more finely mapped using a YAC contig assembled in this region (chapter 7).

The marker order so derived by this genetic analysis has provided a framework for ordering and orienting YAC clones containing these markers in a developing physical map spanning the RP2 region (chapter 7). The meiotic breakpoint panel has therefore allowed efficient localisation of new polymorphic markers which should be useful for precise mapping of other retinal diseases mapping to the region (section 1.9). Furthermore, this set of ordered crossovers will be valuable resources for mapping of additional markers relative to those flanking RP2, which may be useful in excluding potential candidate genes from involvement in the disease.
CHAPTER 7

Physical Mapping in Xp11.3-p11.22: Construction of a YAC Contig Spanning the RP2 Critical Region

7.1 Introduction

Once a disease gene has been localised to a subchromosomal region by genetic analysis and its position refined as much as possible using available family and genetic marker resources (section 1.2.7), the next step toward positionally cloning the gene is conventionally to isolate DNA from the critical region (section 1.3.2.2). YAC clones (section 1.3.2.3) are generally used to provide maximum coverage in compiling a physical map as a resource for the fine mapping of regional markers and ESTs (thus providing/excluding candidate genes in the latter case), and from which to isolate further polymorphic markers for disease locus refinement, and novel candidate genes.

7.1.1 YAC libraries and identification of specific clones

The majority of YAC libraries have been created by ligating genomic DNA that has been partially digested with the cloning site restriction enzyme (EcoRI for pYAC4) between the two arms of the YAC vector (Schlessinger 1990). Techniques such as size-fractionation of the digested DNA by PFGE to remove small fragments and addition of polyamines during the cloning procedure to prevent insert DNA degradation have been used in order to achieve a larger average YAC size (Anand et al 1990; Larin et al 1991). Most primary libraries have been constructed as gridded arrays of recombinants such that all clones are equally represented. Filters are prepared from these gridded arrays either manually or by the use of a robotic device, for subsequent screening by hybridisation (Anand et al 1990). Although the difficult and labour-intensive nature of the process has resulted in a limited number of human YAC libraries in widespread use, the use of common resources by different investigators facilitates the creation of consensus maps.

In order to provide a useful resource it is important that specific clones be identified quickly and efficiently. Multiple screening rounds of numerous filters by colony hybridisation can be avoided by subdividing the library and screening DNA from pools of colonies for a target sequence by PCR. In the original scheme of Green and Olsen (1990a) DNA was prepared from 36 pools each containing 1920 clones. Subsets (each of 384 clones) of all pools scoring positive in the first round were then tested in a second round of PCR screening, to identify which pool(s) of 384 clones contained the clone(s) of interest. The clone
was then identified by hybridisation of a single filter containing an array of clones stamped from the appropriate 4 microtitre plates. Further modification of this approach entailed the use of PCR pools representing rows and columns of either individual or grouped microtitre plates to identify the exact location of the positive clone by PCR, thus eliminating the final hybridisation step (Bentley et al 1992).

The choice of screening method depends upon the application. Screening by PCR is convenient, sensitive, non-radioactive, more efficient when searching for a single clone, and is particularly suited to the adoption of a strategy for genome analysis based on sequence-tagged sites by the Human Genome Project (section 1.5). Hybridisation screening is, however, preferable when using a probe which detects a multigene family, or when using complex probes with large numbers of corresponding YAC clones requiring an excessive amount of PCRs for their identification.

7.1.2 Assembling YAC contigs

A major application of YACs is to determine the order and physical distance between landmarks that have been previously assigned to specific chromosomal regions. It is usually necessary to assemble a contig of overlapping YACs to represent the region completely. The general principle of all contig-building approaches is the assumption that two clones overlap if they share one or more common reference points. The strategy used depends on the density of probes available, the lengths of the YAC inserts in the clones, and the complexities of the libraries being screened.

7.1.2.1 STS / probe-content mapping

The most powerful approach for building contigs identifies unique sequence-tagged-site (STS) markers (section 1.5.1) which are present in two or more YACs by PCR analysis. The logic behind the approach is as follows: if an STS is unique to a particular genomic locus, then all clones that contain that STS must have originated from the same genomic locus and must therefore overlap. One major advantage of STS-content mapping is that, as an integral part of the mapping process, an ordered set of STSs is generated for the genomic region under study. This can be used to order other clones or to isolate clones from other libraries, and constitutes a valuable asset which can be readily disseminated to the scientific community.

Contig-building by STS-content mapping is particularly applicable to large-insert clones such as YACs because the success of the procedure depends directly on the size of the DNA fragments cloned. However, the depth of coverage (redundancy) and average size of YACs required to link STSs will depend on the density of STSs and the desired map resolution (i.e. average distance between ordered STSs); smaller YACs require more STS landmarks for their assembly into a contig, but the final resolution will be much higher. If YACs with larger inserts are being used, then greater redundancy must be attained to achieve the same result. The order and orientation of contigs can only be determined using ‘anchors’ (i.e. STSs for
which positional information is known), which also provide a means to integrate the physical map with genetic and/or cytogenetic maps of the region.

The initial YAC contigs to be assembled were in regions where the available probes or sequence data were of sufficient density both to isolate all the necessary YACs and detect the majority of overlaps (Green and Olsen 1990; Coffey et al 1992). In chromosomal regions with insufficient probes/STSs to assemble complete contigs directly, novel region-specific STSs must usually be generated; often the rate-limiting step in YAC-based projects. Chromosome walking (section 7.1.3.3) is often used to bridge gaps (to identify an overlap between YACs flanking a gap or create a new STS for the isolation of bridging clones). DNA sources for novel chromosome/region-specific STSs include libraries of flow-sorted chromosomes (e.g. X-specific cosmid library; Nizetic et al 1991), microdissected chromatin (Ludecke et al 1989) or hybrid cells containing the region of interest as their only human component (Gusella et al 1980). A more efficient method is to use human repeat-specific primers to amplify DNA between repeats that are sufficiently close together (Nelson et al 1989). ALU represents the major family of short interspersed repeats (SINEs) in mammalian genomes, with approximately $10^6$ copies of a 300bp sequence spaced ~ every 4kb in the human genome (Batzer and Deininger 1991). Cole et al (1991) were able to obtain ALU PCR products at an estimated density of 1 per 75-100kb from a hybrid containing a segment of Xq26, and generate sufficient STSs to identify YACs spanning the entire region (Cole et al 1992). STSs may also be generated from the clones already comprising the contig, for example by YAC endcloning (section 7.1.3.3) or ALU PCR (Nelson et al 1989). This approach will, however, carry any biases inherent in the clones through to the production of STSs.

7.1.2.2 Random assembly / fingerprinting

An alternative strategy for YAC contig construction when handling large numbers of clones, or for regions where STSs are sparse, is to compare YACs by generating a characteristic fingerprint of the genomic insert within each clone. This can be achieved without first separating the YAC from the yeast chromosomal background by Southern analysis of individual YACs using human-specific repetitive sequence probes (e.g. ALU; Wada et al 1990; L1, Bellane-Chantellot et al 1992). A simpler strategy is to perform PCR using primers designed to human repeat sequences and compare the pattern of bands after electrophoresing the PCR products. ALU PCR is most frequently used, because of the relative abundance of this type of repeat within the human genome. The number of bands produced depends on the prevalence, orientation and spacing of ALU elements within the YAC insert; on average 5-10 bands of 0.1 - 2kb will be generated per YAC (Nelson et al 1991).

Since many regions of the genome share identically sized restriction fragments or inter-ALU PCR products, the fingerprinting approaches are usually accompanied by other methods to confirm YAC overlaps and are best applied in determining the relative orientation and extent of overlap of a limited set of YACs already known to be in close proximity by virtue of their STS content (Coffey et al 1992). ALU-PCR might also help determine which
YACs are likely to extend a contig furthest and can provide evidence of noncontiguous ligation or other artefacts.

Inter-ALU PCR is limited, however, by the apparent asymmetric distribution of ALU elements in the genome (Moyzis et al 1989), a problem which is confounded if the ALU repeats are not in the correct orientation for inter-ALU sequence amplification. In these situations a more complex YAC fingerprint may be attained by using a mixture of ALU primers in opposite orientations or an ALU primer in conjunction with a primer for another, less abundant repeat element such as L1 or (CA), (Zietkiewitz et al 1994). Another method which may generate a more complex pattern of PCR products is ALU-Vectorette PCR (Munroe et al 1994).

7.1.2.3 Chromosome walking: Isolation of YAC ends

Chromosome walking relies on isolation of a DNA fragment at or near an end of a cloned insert for use as a probe to screen a YAC library and identify overlapping clones. Markers from the ends of YACs are most efficient for walking strategies to maximise the size of each step. The procedure is generally used to bridge gaps in existing contigs, provide additional STSs for detection/confirmation of YAC overlap and orientation during contig formation, to extend outwards from both ends of a contig into regions devoid of STSs, and to identify chimaeric clones (Kere et al 1992). Although the most obvious way to isolate YAC endsequences is to subclone the yeast DNA and screen for YAC vector sequences, this is a labour-intensive approach and a variety of PCR-based strategies have been developed. These methods permit amplification of a DNA fragment when sequence from only one end (i.e. the vector arm) is known. They can be broadly divided into two categories: those that require manipulation of the template to create a second primer site prior to amplification, or those that initiate amplification from repeat sequences or random sites within the insert.

(1) In inverse PCR (Ochman et al 1988) DNA from a YAC clone is digested with a restriction enzyme, ligated under conditions favouring circularisation of individual restriction fragments, and used as a template for PCR with a pair of diverging oligonucleotide primers that recognise part of the YAC vector. The only PCR product generated is derived from the circularised restriction fragment containing the cloning junction between vector and insert. Inverse PCR can also be used to unravel the genomic structure of a gene by identifying intron/exon boundaries.

(2) ALU-vector PCR (Nelson et al 1991) results in amplification of sequences between the cloning site in the vector and an ALU element within the insert, thus generating specific endfragments from the cloned DNA (section 2.7.8.1). ALU-vector PCR is the simplest method for obtaining insert ends although its success depends on the chance proximity of an ALU repeat close to the cloning site.

(3) Primer-ligation methods (e.g. Riley et al 1990) involve restriction enzyme digestion of the YAC followed by ligation of the restriction fragments to a specially designed vectorette linker such that PCR amplification using YAC vector primers and a vectorette primer results in
products encompassing the insert ends (section 2.7.8.2). Vectorette PCR can be performed using any known sequence (e.g. genomic DNA) to amplify adjacent unknown sequence.

(4) 'Junction trapping' (Patel et al 1993) involves partially digesting the YAC-containing yeast DNA with a frequent-cutting enzyme such as SauIIIa, ligation to a plasmid vector and selection of the YAC insert termini using primers specific to each YAC vector arm and plasmid vector. The partial digestion strategy potentially allows the generation of more than one endfragment in the same reaction (such that a product of sufficient size for STS generation is obtained), therefore avoiding the need to use a number of enzymes to achieve the same effect (as in primer-ligation methods).

(5) More recently developed methods obviate the need for preparative digestions or ligations prior to PCR by using an arbitrary primer to initiate amplification from the unknown DNA segment bordering the insert-vector cloning junction. These procedures involve random annealing of the non-specific or degenerate primer to the template at low stringency, followed by addition of YAC vector-specific primers at higher stringency for specific PCR amplification of insert-terminal sequences (Wesley et al 1994; Dominguez and Lopez-Larrea 1994). Nontarget products are eliminated by nested PCR or by alternating reduced and high stringency PCR cycles to selectively amplify the target molecules (Thermal asymmetric interlaced PCR or (TAIL)-PCR; Liu and Whittier 1995).

This list is by no means exhaustive; the plethora of methods and modifications available for YAC-end isolation indicates that they each have advantages and limitations. In practise a combination of these is required to complete a study involving a large number of YACs.

7.1.3 Construction of a YAC contig spanning the RP2 interval

Until very recently, genetic mapping studies were unable to further delimit the RP2 gene critical region beyond a ~13cM region between markers DXS7 and DXS255 in proximal Xp (section 1.9.5.3.2). Although many markers have now been mapped to this interval, their finer ordering has been hindered by a relatively poor genetic and physical map in the region (Weissenbach et al 1992; Cohen et al 1993). The generation of an ordered marker map is essential for the localisation and isolation of disease genes.

Construction of a YAC contig was therefore initiated within this region, using an STS-content strategy, to establish physical linkage, order and approximate distances between loci spanning the RP2 critical interval. The large size of the RP2 interval made imperative continued efforts to refine the RP2 region by haplotype analysis, therefore genetic mapping studies were continued in parallel (see chapters 5 and 6), as new genetic markers/family members became available, to constantly redirect YAC contigging efforts.

Contig construction was initiated around marker DXS426 in Xp11.23 which was located in the middle of the region of interest and from heterogeneity analysis (Teague et al 1994) appeared to be the most closely linked marker to RP2. Seed YACs were also identified
for flanking markers DXS7 and DXS255 and other interim markers of known location as anchor points on the developing physical map (see Figure 7.1). STS-content mapping can be used to generate complete contigs without the need for YAC walking only for regions of the genome where large numbers of probes are available. As the interval between GATA1 and DXS255 was devoid of published STSs or microsatellites, novel STSs were created from available sequence data, in order to identify corresponding YAC clones directly and use YAC walking as a means to bridge the gaps.

Although still in progress, the work to date has resulted in one major contig of ~2.5Mb in Xp11.23 centred around DXS426 and extending proximally to DXS1240 and distally to UHXI, with potential physical linkage to a cluster of YACs around DXS1003/DXS1055 in Xp11.3-Xp11.23. Satellite contigs have been assembled around DXS1039 (Xp11.22), MAO (Xp11.3) and DXS8083 (Xp11.3-11.23). During the course of contig assembly, new genetic markers became available which led to substantial refinement of the RP2 region (as presented in chapter 6) to ~5cM flanked by DXS8083 (Xp11.3) and DXS6941 (Xp11.23). This enabled us to exclude Xp11.22 further physical analysis and focus our efforts on bridging the DXS8083-DXS426 interval (Xp11.3-11.23) to form a complete contig spanning the RP2 critical region. This STS-formatted contig provides a first step towards estimating the physical distance between the RP2 flanking markers. It will also provide a valuable resource for finer analysis of this region: localising and ordering new genetic markers, STSs and ESTs emanating from the Human Genome Project (section 1.5), and the isolation of further polymorphic markers and expressed sequences to facilitate identification of the gene underlying RP2.

7.2 Materials and Methods

7.2.1 Sources of STS sequences and STS development

Genetic markers used in contig construction were as follows, with current order from Xp11.3 to Xp11.22 (Nelson et al 1995; WI map via Netscape version 3.01): DXS7 - MAOA - MAOB - DXS8080 - (DXS8026, DXS8083) - (DXS1003/DXS1055) - DXS426 - DXS1367 - DXS6616 - DXS6941 - DXS722 - DXS1126 - DXS1240 - DXS1039 - DXS255. Markers DXS337, DXS7124 and DXS1146 (Xp11.23) were later included as discussed in sections 7.3.3.1. Loci mapped to the Xp11.3-p11.23 interval at the outset of this work are shown in black in Figure 7.1. The loci in pink represent STSs that were either of broad location or became available during the course of the study and were subsequently incorporated into the physical map. PCR conditions and relevant citations for all these markers except MAOB and DXS1240 have been previously presented in chapters 5 and 6. Details for MAOB, DXS1240 and other STSs used in contig construction are given in Table 7.1.

STSs were also designed to existing sequence data to nucleate YAC screens. Published sequences for human genes GATA1, SYP, TFE3, the DNA segment containing VNTR locus DXS255, and also 4 retinaally expressed cDNAs mapped close to OATL1 (MG21, MG44, MG61, MG81; section 1.8.4.2) were retrieved from Genbank. For all genes but SYP, genomic
Figure 7.1: Loci mapped to Xp11.3-p11.22 at X chromosome workshop 1993
(Approximate locations of later published loci are shown in pink)

RP2 intervals
(see section 7.3)

Distal

Central

Proximal

Proximal SIN176 deletion breakpoint

Xp11.3

DXS7

MAOA

MAOB

NDP

DXS742

UBE1

DXS1055

DXS1003

SYN1

TIMP1

DXS426

PFC

ELK1

ZNF21

ZNF81

OATL1

DXS226

GATA1

SYP

TFE3

DXS255

DXS8080

DXS1239

UHX1

DXS8083

DXS8026

PCTK1

DXS8237E

DXS7124

DXS1146

DXS337

DXS1004E

DXS1367

DXS1011E

DXS6616

DXS722

MG21 MG44 MG61 MG81

DXS6941

DXS1126

DXS1240

DXS573

DXS1007E

DXS1039

MG61 MG81

DXS6941
### Table 7.1 Additional STSs used to format the contig
(NB: unpublished primer pairs are shown)

<table>
<thead>
<tr>
<th>STS</th>
<th>Marker type</th>
<th>T&lt;sub&gt;r&lt;/sub&gt; (°C)</th>
<th>Product size (bp)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAOB</td>
<td>microsatellite</td>
<td>59</td>
<td>~170</td>
<td>Konradi et al 1992</td>
</tr>
<tr>
<td>NDP</td>
<td>EST</td>
<td>57</td>
<td>249</td>
<td>Ishii et al 1995</td>
</tr>
<tr>
<td>DXS6670</td>
<td>YAC end F0214</td>
<td>56</td>
<td>153</td>
<td>Black et al 1995</td>
</tr>
<tr>
<td>DXS1239</td>
<td>microsatellite</td>
<td>58</td>
<td>~210</td>
<td>Glass et al 1993</td>
</tr>
<tr>
<td>B013LA</td>
<td>YAC end ICRFy900B013</td>
<td>58</td>
<td>176</td>
<td>Black et al 1995</td>
</tr>
<tr>
<td>DXS6671</td>
<td>YAC end ICRFy900B0721</td>
<td>53</td>
<td>112</td>
<td>Black et al 1995</td>
</tr>
<tr>
<td>B0126LA</td>
<td>YAC end ICRFy900B0126</td>
<td>55</td>
<td>137</td>
<td>Black et al 1995</td>
</tr>
<tr>
<td>DXS742</td>
<td>ALU PCR product</td>
<td>56</td>
<td>~200</td>
<td>Bergen et al 1993</td>
</tr>
<tr>
<td>UBE1</td>
<td>EST</td>
<td>57</td>
<td>~700</td>
<td>Handley et al 1991</td>
</tr>
<tr>
<td>TIMP1</td>
<td>EST</td>
<td>65</td>
<td>3100</td>
<td>Aldred and Wright 1991</td>
</tr>
<tr>
<td>TIMP1</td>
<td>EST (exon 5)</td>
<td>62</td>
<td>209</td>
<td>This study; chapter 6</td>
</tr>
<tr>
<td>PFC</td>
<td>microsatellite</td>
<td>60</td>
<td>224</td>
<td>Coleman et al 1991</td>
</tr>
<tr>
<td>F0701LA</td>
<td>YAC end ICRFy900F0701</td>
<td>64</td>
<td>149</td>
<td>Fisher et al 1995</td>
</tr>
<tr>
<td>33CA11LA</td>
<td>YAC End 33CA11</td>
<td>45</td>
<td>111</td>
<td>Thiselton et al 1996</td>
</tr>
<tr>
<td>33CA11RA</td>
<td>YAC End 33CA11</td>
<td>56</td>
<td>134</td>
<td>Thiselton et al 1996</td>
</tr>
<tr>
<td>ELK1</td>
<td>EST F: 5′caacatagcagagccccgctc 3′</td>
<td>60</td>
<td>~125</td>
<td>A. Meindl (personal comm. 1994)</td>
</tr>
<tr>
<td></td>
<td>R: 5′tettgctttgcctagtcgcte 3′</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DXS1004E</td>
<td>EST</td>
<td>54</td>
<td>110</td>
<td>Parrish and Nelson 1993</td>
</tr>
<tr>
<td>DXS6849</td>
<td>microsatellite</td>
<td>54</td>
<td>~180</td>
<td>Kwan et al 1995</td>
</tr>
<tr>
<td>ZNF81</td>
<td>EST</td>
<td>53</td>
<td>?</td>
<td>Marino et al 1993</td>
</tr>
<tr>
<td>DXS6950</td>
<td>microsatellite</td>
<td>59</td>
<td>~225</td>
<td>Kwan et al 1995</td>
</tr>
<tr>
<td>DXS1011E</td>
<td>EST</td>
<td>55</td>
<td>~130</td>
<td>Parrish and Nelson 1993</td>
</tr>
<tr>
<td>DXS1240</td>
<td>microsatellite</td>
<td>57</td>
<td>~170</td>
<td>GDB (A. Nemeth 1993)</td>
</tr>
<tr>
<td>DXS1266</td>
<td>ALU PCR product from YAC ICRFy900E1076</td>
<td>60</td>
<td>95</td>
<td>Coleman et al 1994</td>
</tr>
<tr>
<td>DXS1264</td>
<td>LA YAC End CEPH 76E10</td>
<td>55</td>
<td>67</td>
<td>Coleman et al 1994</td>
</tr>
<tr>
<td>ZNF157</td>
<td>EST F:5′ctttttctgtctcaccatt 3′</td>
<td>55</td>
<td>296</td>
<td>J. Derry (personal comm. 1995)</td>
</tr>
<tr>
<td></td>
<td>R:5′aagctgactacatgtgtgcttc 3′</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DXS8237E</td>
<td>EST</td>
<td>63</td>
<td>~1200</td>
<td>Coleman et al 1996</td>
</tr>
<tr>
<td>PCTK1</td>
<td>EST F:5′gacatcagacacagccctac 3′</td>
<td>60</td>
<td>~125</td>
<td>Meyerson et al 1992</td>
</tr>
<tr>
<td></td>
<td>R:5′tagctgagcagcggagg 3′</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UHX1</td>
<td>EST F: 5′aggaatatccctccctcaggct 3′</td>
<td>60</td>
<td>118</td>
<td>Swanson et al 1996</td>
</tr>
<tr>
<td></td>
<td>R: 5′agcttggacctggctgcag 3′</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DXS1007E</td>
<td>EST</td>
<td>58</td>
<td>124</td>
<td>Parrish and Nelson 1993</td>
</tr>
</tbody>
</table>
sequence data was unavailable, thus the coding regions were established from the cDNA sequence by computer analysis using the MAP program (GCG package) via the computing facilities at The HGMP Resource Centre. EST PCR primers were designed to amplify 100-200bp of DNA from the 3'-untranslated regions (3'UTRs; section 1.5.3). SYP primers were designed within exon 6 of the genomic sequence (Ozcelik et al 1990). For DXS255, two sets of primers were designed: one flanking the VNTR, for potential use as a genetic marker, the other pair to one side of this repeat was used for YAC library screening.

Sequences of the PCR primers, sizes of the STSs and assay conditions for the STSs developed to known sequences during the course of this study are provided in Table 7.2. Complete sequence data has been deposited in Genbank and STSs submitted to GDB.

<table>
<thead>
<tr>
<th>STS Name</th>
<th>Source</th>
<th>PCR Primers</th>
<th>Tc (°C)</th>
<th>Product size</th>
<th>Accession no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>GATA1</td>
<td>Gene 3'UTR</td>
<td>5' ACAGAGCATGCGCTCCAGAG 3' 3' GTCCTTGAAGGTTCAAGCCAGG 3'</td>
<td>62</td>
<td>122 bp</td>
<td>DXS9855E</td>
</tr>
<tr>
<td>SYP</td>
<td>Gene exon 6</td>
<td>5' AGCCTGACTATGGTCAACCAG 3' 5' GACTACACTGTTGAGGAAGG 3'</td>
<td>60</td>
<td>122 bp</td>
<td>DXS9881E</td>
</tr>
<tr>
<td>TFE3</td>
<td>Gene 3'UTR</td>
<td>5' TACCAGAAACCAGTACCAG 3' 5' AGATTCCTCAGTCCGGAGCG 3'</td>
<td>58</td>
<td>203 bp</td>
<td>DXS9891E</td>
</tr>
<tr>
<td>DXS255</td>
<td>VNTR locus</td>
<td>5' GATCATGGGTAAGTTCAGGA 3' 5' AGCAATAGTAGTGAGGTAGC 3'</td>
<td>58</td>
<td>328 bp</td>
<td>DXS9882</td>
</tr>
<tr>
<td>Primers 1/3</td>
<td>(flanking)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DXS255</td>
<td>VNTR locus</td>
<td>5' GATCATGGGTAAGTTCAGGA 3' 5' TAGGTATCTCTAGTGGTACC 3'</td>
<td>58</td>
<td>190 bp</td>
<td>DXS9882</td>
</tr>
<tr>
<td>Primers 1/4</td>
<td>(to one side)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MG21</td>
<td>Gene 3'UTR</td>
<td>5' ATTAATTGCTGGACTCCAGCTCA 3' 5' CTGCGCTCCGAAGACTGTGAACGGT 3'</td>
<td>62</td>
<td>323 bp</td>
<td>DXS9867E</td>
</tr>
<tr>
<td>MG44</td>
<td>Gene 3'UTR</td>
<td>5' TGGAGGGCTGTTCTGCTGACAGTGA 3' 5' TAGCTCTTGCGCTCGACTG 3'</td>
<td>62</td>
<td>494 bp</td>
<td>DXS9868E</td>
</tr>
<tr>
<td>MG61</td>
<td>Gene 3'UTR</td>
<td>5' ACTCTGGGACCCCTTACATAACCTCTCTT 3' 5' GGATCTCCTCTTCTCTGGTGG 3'</td>
<td>62</td>
<td>252 bp</td>
<td>DXS9869E</td>
</tr>
<tr>
<td>MG81</td>
<td>Gene 3'UTR</td>
<td>5' ATCTCCATTGTCAGCTGAGG 3' 5' AAGCACAGGCCTCCACACTTAGG 3'</td>
<td>62</td>
<td>123 bp</td>
<td>DXS9870E</td>
</tr>
</tbody>
</table>

Either ALU-vector PCR or vectorette PCR was used to recover YAC insert terminal sequences (sections 2.7.8.1 and 2.7.8.2). For ALU-vector PCR, all combinations of an end with an internal ALU primer were tested along with the ALU primers alone. Only the bands that were not present in the ALU-ALU amplification were retained for subsequent analysis. For vectorette PCR, YAC clone DNA in the presence of host yeast DNA was digested with HaeIII, Rsal, AluI, DraI, EcoRV and PvuII and then ligated to blunt-ended Vectorette units. These 'Vectorette library' DNAs were then used as substrates in PCR with pYAC4 L and R end-specific and Vectorette-specific oligonucleotides to amplify insert-terminal DNA. After a
few insert-terminal STSs had been isolated using the 4bp cutters (HaeIII, AluI and Rsal) as recommended in the Chemical Genetics manual (CRB), it became apparent that many of the products obtained were too small for successful STS generation; thereafter 6bp cutters EcoRV and PvuII were routinely used to increase the chance of getting a reasonable length of insert DNA. Fragments > 100bp were thus obtained from insert ends for more than 90% of those tested using this approach. Double-stranded PCR products were sequenced directly after purification from either end using 5'32P-labelled primers complementary to the pYAC4 vector or the Vectorette (section 2.7.8.2). The resulting sequence was analysed by FASTA database identity search (Pearson and Lipman 1988) to highlight repetitive elements and any identity to published sequences. Primer pairs were designed manually to single-copy sequence following the constraints in section 2.7.3 and the conditions of PCR were optimised for each STS (section 2.7.3).

The X chromosomal localisation of each novel STS was ascertained by PCR assay of DNA from human-rodent somatic cell hybrids containing various portions of the X chromosome (section 2.7.8.3). STSs that yielded a product of the expected size with total human DNA, ThyBX, KAG2.3 but not SIN176, mouse or hamster DNA were ascribed to human Xp21.1-p11.22 and the parent YAC deemed not to be chimaeric if both ends followed this pattern. Since the a priori probability that a coligated YAC from a 4X YAC library has both ends from this region of the X is ~3% this served as a fairly reliable screen for artefacts of this kind. Further localisation was established by analysis of overlapping YACs.

7.2.2 Identification of YACs by PCR-based YAC library screening

Screening the YAC libraries (section 2.7.5) was a continuous process as new STSs were mapped to the region by the human genetics mapping community, to bypass as far as possible the need for YAC endcloning, which is time-consuming and can generate STSs of uncertain location. Additional YACs were obtained from the ICRF Reference Library Database (RLDB; Zehetner and Lehrach 1994) and CEPH database (Cohen et al 1993) to increase coverage of particular regions as described in the relevant results sections. For the sake of clarity, the names of YACs from the ICRF YAC library have been abbreviated in the text by omitting the library name ICRFy900.

For screening a YAC library, a PCR assay must be sufficiently robust to allow detection of STSs present in relatively low concentrations within yeast-derived DNA preparations. Therefore, for STSs likely to map to the region of interest, each corresponding PCR assay was routinely subjected to preliminary testing on human genomic DNA to establish acceptable conditions for detecting clones within a YAC library. Generally, a PCR assay that was not robust at the initial testing stage proved to be problematic on YAC library screening, but may have proved sufficient for detecting overlaps in neighbouring YAC clones (lower DNA complexity).

All library screens included human genomic DNA and 'no DNA' as positive and negative controls.
7.2.3 YAC characterisation and contig construction

YAC DNA was prepared in agarose plugs from verified positive single colonies as described in section 2.7.6. YAC identity was confirmed by PCR of plug preps with the STS used to identify the YAC from library pools. Size and purity of YACs in each isolate were determined by PFGE of yeast chromosomes (section 2.4.3), followed by alkali Southern blotting and hybridisation with radiolabelled sheared human genomic DNA (sections 2.5.2 and 2.5.4). YACs were often visible after ethidium bromide staining of the pulsed-field gel and exposure to UV light but hybridisation of Southern blots with a radiolabelled sheared human DNA probe highlighted the YAC where it comigrated with a yeast chromosome and would also reveal any yeast cells harbouring more than one YAC. STS content of each YAC was determined by scoring the presence or absence of a correctly sized PCR amplification product using the appropriate primers with positive (human genomic DNA) and negative controls (yeast DNA, pYAC4 DNA and ‘no DNA’), and contigs were assembled by comparing STS content of individual YACs (section 7.1.2.1). This process worked both ways: STSs of known relative location helped order the YACs, and comparison of YAC STS content generated orders for most STSs.

A subset of YACs in Xp11.3-Xp11.23 were selected for additional assessment of overlap by ALU-PCR fingerprinting (sections 7.1.2.2 and 2.7.7) using primers ALE1 and ALE3. Detailed manual comparison of fingerprints was performed to confirm potential overlaps and to provide a suggested order of clones based on subsets of shared and nonshared bands.

7.3 Results

Owing to the large size of the RP2 region when contig assembly was initiated, the physical mapping results are split into three sections, corresponding to the distal boundary for RP2 (Xp11.3), the proximal boundary (Xp11.23-11.22) and the central RP2 interval (Xp11.23) as indicated in Figure 7.1.

7.3.1 YAC contig around the RP2 region distal boundary in Xp11.3

To span the distal flanking marker for RP2 at the outset of this study (MAOA; unpublished data), 6 YACs (see Table 7.3) were obtained from the ICRF Reference Library Database (RLDB) that had been identified by a probe F9/4 derived from a cosmid isolated using the proximal end of an MAOA YAC (Gunther Zehetner, personal communication). A further 6 YACs were isolated from the ICI library using STS DXS742 (Figure 7.1) which was reported to lie within a few Mb of the Norrie disease locus (NDP) in Xp11.3 (Bergen et al 1993). Figure 7.2 shows examples of successive rounds of YAC library screening for DXS742 to identify cognate clones.

Three new Genethon microsatellite markers were subsequently mapped to the region and used to screen the ICI and ICRF libraries on the basis of their genetic location: DXS8080
Figure 7.2 The following series of agarose gel photographs show the positive pools identified upon sequential screening of the ICI YAC library 1\(^0\) (a), 2\(^{nd}\) (b) and 3\(^{rd}\) (c) pools with marker DXS742. The tertiary screen is only shown for positive secondary pools 22C and 21C, which identified YAC clones 22CA9 and 21CB12 and/or 21CF12. 21CF12 was found to be a false positive upon further analysis. H: human genomic DNA (positive control), P: pYAC4 DNA, Y: yeast DNA, -: no DNA (negative controls)

(a) 1\(^{st}\) pools 1-40

(b) 2\(^{nd}\) pools 3, 9, 21, 22, 31, 34
c) 3rd pool 22C and 21C
mapped 3cM distal to DXS1055, DXS8026 mapped 1cM distal to DXS1055 and DXS8083 was inseparable from DXS1055, 1cM distal to DXS1003 (to give the order DXS8080-DXS8026-DXS1055/DXS8083-DXS1003; J. Weissenbach, personal communication 1994). YACs isolated with the latter two markers would be particularly useful in seeding a physical map in the relatively unchartered region of Xp11.3-Xp11.23 between MAO and DXS1003 (see Figure 7.1). Details of all YACs identified and verified positive with these screening markers are given in Table 7.3. Figure 7.3 shows the DXS742 YACs sized by PFGE and hybridisation with total human DNA.

All YACs were tested for the presence/absence of STSs mapping to the region and thereby ordered into an overlapping array (Figure 7.4). The colocalisation of DXS8026 and DXS8083 in YAC 40HD1 indicated that DXS8083 lay distal to DXS1055, which was absent from 40HD1 by STS-content analysis (data not shown), and also that DXS8026 and DXS8083 lay within 365kb of each other (assuming 40HD1 is not chimaeric/deleted). The density of YACs in the distal contig allowed unique ordering of almost all STSs. YACs 9BA10 and 40HD1 were selected for insert-terminal STS generation to provide walking probes to bridge the DXS742-DXS8083/DXS8026 gap, for which no other markers were available. 40HD1 was chosen because it contains both markers DXS8083 and DXS8026, and as the relative order of these markers is unknown, presented the most effective YAC from which to walk. 9BA10 was chosen in preference to 3BD10 or 21CB12, which extend further in a proximal direction, because STSs generated from 9BA10 that localised to proximal Xp by hybrid analysis could be more faithfully anchored to the region by presence in overlapping YACs in the contig, and would hence serve as more reliable probes from which to walk.

Left-arm insert-terminal STSs were isolated from both 9BA10 and 40HD1 using the ALU-vector PCR method (section 2.7.8.1). Figure 7.5 shows PCR amplification products obtained from YAC 40HD1 before and after purifying the appropriate band by 'stab' PCR. These products were sequenced directly from the pYAC4 LA sequencing primer (section 2.6.3) to yield 191bp and 210bp of insert sequence respectively (data not shown). FASTA database identity search revealed that the last 80bp of 9BA10LA showed ~80% identity to a variety of human genes and ~70% to primate ALU elements (as demonstrated in Figure 7.6). Further investigation of the region of identity to the human gene entries showed it to be intronic in all instances and in the majority of cases noted to be ALU. Therefore PCR primers were designed from the first 110bp to avoid non-specific amplification. Similar analysis of 40HD1LA sequence revealed no overall homology to Genbank sequences thus there were no such constraints on primer design. Details of the novel STSs generated from these YAC ends are given in Table 7.8.

Prior to YAC library screening, primer pairs for 9BA10LA and 40HD1LA were tested for robustness and Xp localisation. PCR products of the correct size amplified from five different human genomic DNA templates, and their human specificity and localisation to proximal Xp was verified by hybrid analysis (section 7.2.1; data not shown). STS content analysis of Xp11.3 YACs also showed that 9BA10LA was present in YACs 21CB12 and 3BD10, anchoring this STS firmly in Xp11.3 and oriented at the proximal end of 9BA10.
Table 7.3 YACs identified with STSs around the distal RP2 boundary in Xp11.3

<table>
<thead>
<tr>
<th>YAC name</th>
<th>Identifying marker</th>
<th>YAC library</th>
<th>YAC size (kb)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>y900F0214</td>
<td>F9/4 (RLDB)</td>
<td>ICRF</td>
<td>600</td>
<td>intact (Black et al 1995)</td>
</tr>
<tr>
<td>y900E01642</td>
<td>F9/4 (RLDB)</td>
<td>ICRF</td>
<td>490</td>
<td></td>
</tr>
<tr>
<td>y900G0630</td>
<td>F9/4 (RLDB)</td>
<td>ICRF</td>
<td>850</td>
<td></td>
</tr>
<tr>
<td>y900B0721</td>
<td>F9/4 (RLDB)</td>
<td>ICRF</td>
<td>365</td>
<td></td>
</tr>
<tr>
<td>y900F01125</td>
<td>F9/4 (RLDB)</td>
<td>ICRF</td>
<td>540</td>
<td>chimaeric</td>
</tr>
<tr>
<td>y900B013</td>
<td>F9/4 (RLDB)</td>
<td>ICRF</td>
<td>400</td>
<td></td>
</tr>
<tr>
<td>31GA5</td>
<td>DXS8080</td>
<td>ICI</td>
<td>390</td>
<td></td>
</tr>
<tr>
<td>38EA5</td>
<td>DXS8080</td>
<td>ICI</td>
<td>500</td>
<td></td>
</tr>
<tr>
<td>13HF11</td>
<td>DXS8080</td>
<td>ICI</td>
<td>485</td>
<td></td>
</tr>
<tr>
<td>21CB12</td>
<td>DXS742</td>
<td>ICI</td>
<td>650</td>
<td></td>
</tr>
<tr>
<td>3BD10</td>
<td>DXS742</td>
<td>ICI</td>
<td>260</td>
<td></td>
</tr>
<tr>
<td>9BA10</td>
<td>DXS742</td>
<td>ICI</td>
<td>460</td>
<td></td>
</tr>
<tr>
<td>22CA9</td>
<td>DXS742</td>
<td>ICI</td>
<td>150</td>
<td></td>
</tr>
<tr>
<td>22GA9</td>
<td>DXS742</td>
<td>ICI</td>
<td>150</td>
<td></td>
</tr>
<tr>
<td>40IB7</td>
<td>9BA10LA</td>
<td>ICI</td>
<td>nd</td>
<td></td>
</tr>
<tr>
<td>y900C0562</td>
<td>9BA10LA</td>
<td>ICRF</td>
<td>nd</td>
<td></td>
</tr>
<tr>
<td>y900G077</td>
<td>9BA10LA</td>
<td>ICRF</td>
<td>nd</td>
<td></td>
</tr>
<tr>
<td>40HD1</td>
<td>DXS8026</td>
<td>ICI</td>
<td>365</td>
<td></td>
</tr>
<tr>
<td>12AF3</td>
<td>DXS8083</td>
<td>ICI</td>
<td>nd</td>
<td></td>
</tr>
<tr>
<td>y900F05131</td>
<td>DXS8083</td>
<td>ICRF</td>
<td>nd</td>
<td></td>
</tr>
<tr>
<td>y900H0765</td>
<td>DXS8083</td>
<td>ICRF</td>
<td>260</td>
<td></td>
</tr>
<tr>
<td>y900A0947</td>
<td>40HD1LA</td>
<td>ICRF</td>
<td>nd</td>
<td></td>
</tr>
</tbody>
</table>
Figure 7.3 Size estimation of the DXS742 YACs and YAC 31GA5 by pulsed field gel electrophoresis (a) followed by Southern blot hybridisation with total human DNA (b). Only YACs 22CA9 and 22GA9 were clearly visible on ethidium bromide gel staining (~150 kb). Although the autoradiograph is in the opposite orientation, the remaining YACs are highlighted and confirmed to comigrate with yeast chromosomes. YAC 34HA11 was identified from the ICI YAC library with marker DXS742, but was unconfirmed positive by plug PCR. The autoradiograph shows two separate plugs of this YAC; one appears to have lost its human insert, indicating that YAC 34HA11 may be unstable.
Figure 7.4  YAC contig encompassing the RP2 distal boundary in Xp11.3 (not to scale). The horizontal line with bidirectional arrows represents the chromosome, with STSs above in the order determined by YAC content. The deduced location of probe F9/4 (see text) is shown. The bold horizontal lines represent YACs (see Table 7.3 for sizes and source library). A filled rectangle at the end of a YAC indicates a terminal sequence STS that maps to overlapping YACs in the contig. A dot-filled rectangle denotes a terminal sequence STS derived by Black et al (1995). Vertical broken lines show the positions of STSs in YAC clones, with a solid circle indicating a verified positive. The shaded bar depicts the region of the contig contained within the newly defined RP2 critical interval, with the new distal flanking marker, DXS8083, highlighted in bold. STSs DXS6670 and DXS6671 correspond to YAC end STSs F0214RA and B0721LA respectively (Black et al 1995).
Figure 7.5  ALU-vector PCR for insert-terminal isolation from YAC 40HD1.
(a) An ALE1/pYAC4 left arm (LA) primer combination generates a ~1.2kb product which is clearly not a result of inter-ALU amplification.
(b) This product ready for sequencing after purification by ‘stab’ PCR.
Figure 7.6 Printout from FASTA database similarity search for 9BA10LA showing examples of identities to primate ALU repeat sequence.

<table>
<thead>
<tr>
<th>HSAT3 X68793 H.sapiens gene for antithrombin III</th>
<th>240 240</th>
</tr>
</thead>
<tbody>
<tr>
<td>86.7% identity in 83 nt overlap</td>
<td></td>
</tr>
<tr>
<td>80 90 100 110 120 130</td>
<td></td>
</tr>
</tbody>
</table>
| 9ALA, GCAAATTCCTTTTAAAATACACAGATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
Both STSs proved sufficiently robust to identify further YACs to extend the contigs upon library screening, as detailed in Table 7.3 and shown in Figure 7.4.

STS-content mapping of 22 YACs isolated with 5 markers has enabled the assembly of two contigs of overlapping YACs spanning a portion of Xp11.3-Xp11.23 containing the Norrie disease locus (NDP) and the distal flanking boundary of the RP2 region (Figure 7.1). A total of 12 STSs have been used to format the distal contig, all of which except the pair MAOB and NDP are ordered in relation to each other by YAC STS content analysis. NDP is known to lie proximal to MAOB from long-range physical mapping and deletion mapping of Norrie patients (Chen et al 1992a). Microsatellite DXS1239 was isolated from a cosmid identified using an endclone of a 640kb YAC containing DXS7, MAOA, MAOB and NDP and has been mapped approximately 250kb proximal to MAOA (Glass et al 1993). The physical map generated here further positions DXS1239 proximal to NDP and distal to DXS8080 and DXS742 in Xp11.3. The overall physical order generated by this map is: Xpter - DXS7 - MAOA - (MAOB/NDP) - DXS1239 - DXS6670 - B013LA - DXS8080 - DXS6671 - B0126LA - DXS742 - 9BA10LA - Xcen. This order is consistent with that recently published by Black et al (1995) and also adds a new STS (9BA10LA/DXS9883) and two microsatellite markers (DXS1239 and DXS8080) to the physical map of Xp11.3. The estimated lengths of the two contigs, based on minimum tiling path YACs, are ~1Mb for the NDP contig and ~500kb for the DXS8083/DXS8026 contig. These are minimal estimates since the contigs extend both proximally and distally with YACs whose entire content is not included in the estimate.

The extent of YAC chimaerism in these contigs is unpredictable without further analysis. YAC F0124 has been extensively characterised (Chen et al 1992b) and appears to be intact. Comparison of the YAC sizes with their STS-content (Figure 7.4) would however suggest that F01125 (540kb) is chimaeric as it contains only DXS6670, without either neighbouring STS DXS1239 or B013LA, which are deduced to be a maximum of 365kb apart from the size of B013 which contains them both. Furthermore, YAC E01642 would appear to be internally deleted for DXS1239 based on the most likely marker order generated upon STS content analysis of overlapping YACs, taking into account that DXS6670 is reportedly the proximal end of YAC F0214.

A gap remains between 9BA10LA and the DXS8083/DXS8026 YAC cluster of approximately 1-2Mb (the genetic distance between DXS8080 and DXS8026/8083 is now estimated to be 3cM and DXS8026 and DXS8083 are no longer genetically resolved; Dib et al 1996). The size of this gap could be better estimated by procedures such as radiation hybrid mapping (section 1.3.1.1) or high resolution FISH (section 1.3.1.2). However, there are no intentions to bridge this gap since the refinement of the RP2 gene critical region described in chapter 6 effectively excludes this region and localises RP2 proximal to DXS8083. Work is now in progress to ascertain the orientation of the DXS8083 contig so that walks can proceed in a proximal direction.
7.3.2 YAC contig construction around the RP2 proximal boundary in Xp11.23-p11.22

Construction of the contig depicted in Figure 7.7 is described below.

In the proximal part of the RP2 interval (Figure 7.1), YACs were identified by YAC library screening for markers TIMP-1, PFC, DXS426, ZNF81, DXS722, MGB1, GATA1, DXS1126, DXS1240, SYP, DXS1039 and DXS255 and obtained from the ICRF RLDB for markers SYN1 and 2bC6 (DXS226) in Xp11.23. 2bC6 is a genomic clone isolated from a from an X chromosome-enriched phage library with a foetal muscle mRNA probe (Paulsen et al 1986). Markers DXS573 and TFE3 failed to identify YACs from either the ICI or ICRF YAC libraries (data not shown). All verified positive YACs and sizes where determined by PFGE are listed in Table 7.4.

DXS255 primer pair 1-3 (Table 7.1) flanking the VNTR repeat produced a product of the expected size (328bp) from YAC 14BD2 but proved to amplify genomic DNA inconsistently (data not shown). DXS255 contains a VNTR motif characterised by a 26bp repeat unit that shows extensive variation in copy number (Fraser et al 1989). Fragments that amplified successfully were often > 1.5kb, therefore to avoid the possibility that absence of product could be due to particularly large VNTR alleles beyond the limits of the PCR assay, DXS255 primer pair 1-4 (Table 7.1) were used for YAC library screening. YAC A0120 (420kb; Figure 7.8) was not analysed further as different sizes for this YAC were noted in the literature (440 or 500 kb; Coleman et al 1994, Hagemann et al 1994) suggesting instability of this clone.

7.3.2.1 YAC insert-terminal isolation to generate novel region-specific STSs

Novel STSs were generated from YAC insert-terminals of YACs 4HG2, A1220, C1022, 27GF2, 30DH10, 34AC5, C01160, 4DF10, 14BD2 and 27CB12 (the only DXS1039 clone containing a single YAC). A total of 18 YAC insert terminals were successfully isolated: details specific to each YAC end are summarised in Table 7.5 and examples of PCR products from which insert terminal sequence was successfully obtained are shown in Figure 7.9. The combination of two different methods proved beneficial as neither method alone was successful in every case.

Following FASTA database identity searches, the majority of YAC end sequences showed low level identities (~50%) to a random array of species-diverse Genbank sequences (i.e. no identity); for these YAC ends primers were designed according to the usual constraints (section 2.7.3) to amplify as large a DNA fragment as possible. For several YAC end-STSs showing identity to repetitive sequences such as ALU or L1, one primer was designed within the region of identity owing to insufficient sequence of a suitable nature elsewhere (C1022RA, 4DF10RA, 4DF10LA, 27CB12RA) as shown for 4DF10LA in figure 7.10. It was assumed in such cases that viable STSs would still be generated owing to the locus specificity conferred by the unique sequence of the second primer.

Matches were detected for 2 YAC insert terminals with gene sequences; 4HG2RA lies within exon 2 of the SYN1 gene (Figure 7.11), anchoring this end of the YAC firmly in
Figure 7.7 YAC contig covering the proximal RP2 region in Xp11.23-p11.22 (not to scale). The horizontal line with bidirectional arrows represents the chromosome, with STSs above in the order determined by YAC content. The approximate position of the OATL1 pseudogene cluster is indicated by the grey box (Nelson et al 1995). The bold horizontal lines represent YACs (see Table 7.4 for sizes and source library). A filled rectangle at the end of a YAC depicts a terminal sequence STS that maps to overlapping YACs in the contig. A hatched box at the end of a YAC depicts a terminal sequence STS that does not derive from this region of Xp11.23. The asterisk denotes the map position of a new EST in Xp11.23, corresponding to 34AC5RA (see text). Vertical broken lines show the positions of STSs in YAC clones, with a solid circle indicating a verified positive. The shaded bar denotes the region of the contig contained within the newly defined RP2 critical interval. The new proximal flanking marker for RP2, DXS6616, is boxed, and the retinally expressed cDNAs thereby excluded from a role in the disease are highlighted in bold.
Table 7.4  YACs identified with STSs spanning the proximal RP2 region in Xp11.23-11.22

<table>
<thead>
<tr>
<th>YAC name</th>
<th>Identifying marker</th>
<th>YAC library</th>
<th>Size (kb)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>y900C1022</td>
<td>SYN1.2</td>
<td>ICRF (RLDB)</td>
<td>410</td>
<td></td>
</tr>
<tr>
<td>y900C1228</td>
<td>SYN1.2</td>
<td>ICRF (RLDB)</td>
<td>320</td>
<td></td>
</tr>
<tr>
<td>y900A1220</td>
<td>SYN1.2</td>
<td>ICRF (RLDB)</td>
<td>365</td>
<td>potentially unstable</td>
</tr>
<tr>
<td>y900A0120</td>
<td>SYN1.2</td>
<td>ICRF (RLDB)</td>
<td>420</td>
<td></td>
</tr>
<tr>
<td>31CC3</td>
<td>PFC</td>
<td>ICI</td>
<td>320</td>
<td></td>
</tr>
<tr>
<td>2GD6</td>
<td>TIMP1</td>
<td>ICI</td>
<td>225</td>
<td></td>
</tr>
<tr>
<td>38HB4</td>
<td>4HG2RA</td>
<td>ICI</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td>4HG2</td>
<td>DXS426</td>
<td>ICI</td>
<td>245</td>
<td>intact</td>
</tr>
<tr>
<td>33CA11</td>
<td>DXS426</td>
<td>ICI</td>
<td>290</td>
<td>intact</td>
</tr>
<tr>
<td>3EB3</td>
<td>ZNF81</td>
<td>ICI</td>
<td>420</td>
<td>screen by R.M.Hampson</td>
</tr>
<tr>
<td>24GH5</td>
<td>ZNF81</td>
<td>ICI</td>
<td>250</td>
<td>screen by R.M.Hampson</td>
</tr>
<tr>
<td>31AG10</td>
<td>ZNF81</td>
<td>ICI</td>
<td>260</td>
<td>screen by R.M.Hampson</td>
</tr>
<tr>
<td>241A11</td>
<td>MG61</td>
<td>ICI</td>
<td>250</td>
<td>screen by R.M.Hampson</td>
</tr>
<tr>
<td>34AC5</td>
<td>DXS722</td>
<td>ICI</td>
<td>610</td>
<td>chimaeric</td>
</tr>
<tr>
<td>27GF2</td>
<td>DXS722</td>
<td>ICI</td>
<td>530</td>
<td>unstable / rearranged</td>
</tr>
<tr>
<td>31GA2</td>
<td>DXS722</td>
<td>ICI</td>
<td>550</td>
<td></td>
</tr>
<tr>
<td>191H2</td>
<td>DXS722</td>
<td>ICI</td>
<td>340</td>
<td></td>
</tr>
<tr>
<td>y900C01160</td>
<td>2bc6</td>
<td>ICRF (RLDB)</td>
<td>100</td>
<td>chimaeric</td>
</tr>
<tr>
<td>23BC7</td>
<td>MG81</td>
<td>ICI</td>
<td>320</td>
<td></td>
</tr>
<tr>
<td>30DH10</td>
<td>GATA1</td>
<td>ICI</td>
<td>360</td>
<td></td>
</tr>
<tr>
<td>32EH3</td>
<td>DXS1240</td>
<td>ICI</td>
<td>150</td>
<td></td>
</tr>
<tr>
<td>4DF10</td>
<td>SYP</td>
<td>ICI</td>
<td>200</td>
<td></td>
</tr>
<tr>
<td>27CB12</td>
<td>DXS1039</td>
<td>ICI</td>
<td>225</td>
<td>intact</td>
</tr>
<tr>
<td>16CD8</td>
<td>DXS1039</td>
<td>ICI</td>
<td>295, 450</td>
<td>multiple YACs</td>
</tr>
<tr>
<td>34AB12</td>
<td>DXS1039</td>
<td>ICI</td>
<td>140, 365, 410</td>
<td>multiple YACs</td>
</tr>
<tr>
<td>22BB8</td>
<td>27CB12RA</td>
<td>ICI</td>
<td>nd</td>
<td></td>
</tr>
<tr>
<td>4BG7</td>
<td>27CB12RA</td>
<td>ICI</td>
<td>nd</td>
<td></td>
</tr>
<tr>
<td>17AF5</td>
<td>27CB12RA</td>
<td>ICI</td>
<td>nd</td>
<td></td>
</tr>
<tr>
<td>36HB8</td>
<td>27CB12LA</td>
<td>ICI</td>
<td>nd</td>
<td></td>
</tr>
<tr>
<td>22AB3</td>
<td>27CB12LA</td>
<td>ICI</td>
<td>nd</td>
<td></td>
</tr>
<tr>
<td>y900G07133</td>
<td>27CB12LA</td>
<td>ICRF</td>
<td>nd</td>
<td></td>
</tr>
<tr>
<td>14BD2</td>
<td>DXS255</td>
<td>ICI</td>
<td>260</td>
<td>intact</td>
</tr>
<tr>
<td>unknown</td>
<td>DXS255</td>
<td>ICRF</td>
<td>1100</td>
<td>chimaeric (Kwan 1994 personal communication)</td>
</tr>
</tbody>
</table>
Figure 7.8  Pulsed field gel electrophoresis of ICRF SYN1 YACs with YAC names above each lane. The YACs are clearly visible with ethidium bromide staining alone. By comparison with the yeast chromosomal size standard approximate YAC sizes are: C1228 ~320kb, A1220 ~365kb (comigrating with yeast chromosome III), A0120 ~420kb and C1022 ~410kb.
Table 7.5  Insert-teminal Isolation for YACs in the Proximal Portion of the RP2 Interval

<table>
<thead>
<tr>
<th>YAC END</th>
<th>Method</th>
<th>PCR product</th>
<th>Insert sequence obtained</th>
<th>FASTA database homology search</th>
</tr>
</thead>
<tbody>
<tr>
<td>4HG2 LA</td>
<td>Vectorette PCR</td>
<td>~ 750bp from EcoRV library</td>
<td>238 bp</td>
<td>No identity to Genbank sequences</td>
</tr>
<tr>
<td>4HG2 RA</td>
<td>Vectorette PCR</td>
<td>~ 550bp from Dral library</td>
<td>188 bp</td>
<td>94% identity to SYN1 exon 2 (83 nt)</td>
</tr>
<tr>
<td>C1022 RA</td>
<td>ALU-vector PCR (nested)</td>
<td>~ 1150bp with ALE1 primer</td>
<td>326 bp</td>
<td>~ 76% identity to L1 (183 nt)</td>
</tr>
<tr>
<td>A1220 LA</td>
<td>Vectorette PCR</td>
<td>~ 450bp from Rsal library</td>
<td>169 bp</td>
<td>Human specific sequence overall</td>
</tr>
<tr>
<td>A1220 RA</td>
<td>Vectorette PCR (nested)</td>
<td>~ 450bp from EcoRV library</td>
<td>180 bp</td>
<td>No identity to Genbank sequences</td>
</tr>
<tr>
<td>34AC5 LA</td>
<td>Vectorette PCR (nested)</td>
<td>~ 300bp from Rsal library</td>
<td>212 bp</td>
<td>No identity to Genbank sequences</td>
</tr>
<tr>
<td>34AC5 RA</td>
<td>Vectorette PCR (nested)</td>
<td>~ 300bp from Rsal library</td>
<td>180 bp</td>
<td>Fwd strand: 99.5% identity (210nt) to a human infant brain cDNA (EST accession H09726) Rev strand: 99% identity (209nt) to a human placental cDNA (EST accession N41839)</td>
</tr>
<tr>
<td>27GF2 LA</td>
<td>Vectorette PCR</td>
<td>~ 900bp from EcoRV library</td>
<td>131 bp</td>
<td>No identity to Genbank sequences</td>
</tr>
<tr>
<td>YAC End</td>
<td>Method</td>
<td>PCR product</td>
<td>Insert sequence obtained</td>
<td>FASTA Database Homology</td>
</tr>
<tr>
<td>----------</td>
<td>-------------------------</td>
<td>--------------------------------------</td>
<td>--------------------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>27GF2 RA</td>
<td>Vectorette PCR</td>
<td>~ 700bp from Rsal library</td>
<td>182 bp</td>
<td>No identity to Genbank sequences</td>
</tr>
<tr>
<td>30DH10 RA</td>
<td>ALU-vector PCR (nested stab)</td>
<td>~ 500bp with ALE3 primer</td>
<td>242 bp</td>
<td>No identity to Genbank sequences</td>
</tr>
<tr>
<td>C01160 LA</td>
<td>Vectorette PCR</td>
<td>~ 450bp from PvuII library</td>
<td>168 bp</td>
<td>No identity to Genbank sequences</td>
</tr>
<tr>
<td>C01160 RA</td>
<td>Vectorette PCR</td>
<td>~ 500bp from PvuII library</td>
<td>182 bp</td>
<td>No identity to Genbank sequences</td>
</tr>
<tr>
<td>4DF10 LA</td>
<td>ALU-vector PCR</td>
<td>~ 550bp with ALE1 primer</td>
<td>127 bp</td>
<td>90.5% identity to ALU (42nt)</td>
</tr>
<tr>
<td>4DF10 RA</td>
<td>ALU-vector PCR</td>
<td>~ 700bp with ALE3 primer</td>
<td>174 bp; very A-rich and repetitive</td>
<td>98.1% identity to DXS9795 (156nt) 70% identity with ALU (99nt)</td>
</tr>
<tr>
<td>27CB12 LA</td>
<td>ALU-vector PCR</td>
<td>~ 900bp with ALE1 primer</td>
<td>193 bp</td>
<td>93.6% identity to DXS8139 (188nt)</td>
</tr>
<tr>
<td>27CB12 RA</td>
<td>ALU-vector PCR</td>
<td>~ 800bp with ALE3 primer</td>
<td>217 bp</td>
<td>~ 80% identity with ALU (118nt)</td>
</tr>
<tr>
<td>14BD2 LA</td>
<td>Vectorette PCR</td>
<td>~ 480bp from Rsal library</td>
<td>240 bp</td>
<td>No identity to Genbank sequences</td>
</tr>
<tr>
<td>14BD2 RA</td>
<td>ALU-vector PCR</td>
<td>~ 900bp with ALE1 primer</td>
<td>258 bp</td>
<td>No identity to Genbank sequences</td>
</tr>
</tbody>
</table>
Figure 7.9 Insert-terminal isolation from YACs in the proximal contig: (a) The Vectorette system applied to YAC 4HG2. The Vectorette library from which each product derives is indicated above each lane, along with the YAC vector primer (LA or RA). The relatively clean large EcoRV/LA and Dral/RA amplification products were chosen for sequencing. Lanes 13 and 14 are negative controls. (b) ALU-vector PCR applied to YAC 27CB12 fortuitously produced amplification products spanning both the left and right insert-ends as this figure shows.
Figure 7.10 FASTA printout showing region of ALU identity in 4DF10LA. Sequences of PCR primers designed to generate an STS are boxed.
Figure 7.11 FASTA printout showing identity of 4HG2 RA YAC terminal sequence to exon 2 of the human SYN1 gene.
Xp11.23, whilst 34AC5RA showed several identities to different Genbank human EST entries as shown in Figure 7.12. Comparison of the FASTA results suggests these ESTs are derived from the same region of a single cDNA, with the variation in % identity a result of EST sequencing errors; ESTs are unedited single-pass sequences and are thus by their very nature prone to error, mostly from base miscalling and lane tracking errors inherent in the automated fluorescent sequencing apparatus used (Hillier et al 1996). It thus appears that 34AC5RA is identical to a cDNA expressed in at least 3 tissues, infant brain, placenta and ovary tumour.

Physical map location of the novel YAC end-STSs as derived by somatic cell hybrid and YAC STS content analysis is summarised in Table 7.6. Primer pairs for those mapping to proximal Xp are listed in Table 7.8.

### Table 7.6 YAC End STS chromosomal localisation for proximal contig by STS-content analysis of hybrids and overlapping YACs

<table>
<thead>
<tr>
<th>YAC End</th>
<th>Human DNA</th>
<th>ThyBX</th>
<th>Sin 176</th>
<th>Keg 2.3</th>
<th>Mouse DNA</th>
<th>Hamster DNA</th>
<th>Positive YACs extra to parent YAC</th>
<th>Map location</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1022RA</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>A1220, C1228</td>
<td>Xp11.23</td>
</tr>
<tr>
<td>A1220RA</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>C1022, C1228, A1220, 38HB4</td>
<td>Xp11.23</td>
</tr>
<tr>
<td>4HG2RA</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>C1022, 4HG2, 33CA11</td>
<td>Xp11.23</td>
</tr>
<tr>
<td>A1220LA</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>C1022, 4HG2, 33CA11</td>
<td>Xp11.23</td>
</tr>
<tr>
<td>4HG2LA</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>C1022, 33CA11</td>
<td>Xp11.23</td>
</tr>
<tr>
<td>30DH10RA</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>34AC5, 31GA2</td>
<td>Xp11.23</td>
</tr>
<tr>
<td>34AC5RA</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>31GA2</td>
<td>Xp11.23</td>
</tr>
<tr>
<td>34AC5LA</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Xpter-p22</td>
</tr>
<tr>
<td>C01160RA</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>19IH2, 27GF2, 30DH10</td>
<td>Xp11.23</td>
</tr>
<tr>
<td>C01160LA</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>5/6 (Biosmap)</td>
<td></td>
</tr>
<tr>
<td>27GF2LA</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Xp21.1-p11.22</td>
</tr>
<tr>
<td>27GF2RA</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Xp21.1-p11.22</td>
</tr>
<tr>
<td>27CB12RA</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Xp11.22</td>
</tr>
<tr>
<td>27CB12LA</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>34AB12</td>
<td>Xp11.22</td>
</tr>
<tr>
<td>14BD2RA</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>CEPH 255 YAC</td>
<td>Xp11.22</td>
</tr>
<tr>
<td>14BD2LA</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>CEPH 255 YAC</td>
<td>Xp11.22</td>
</tr>
</tbody>
</table>

Of the 18 insert terminals isolated, 13 appeared to map to proximal Xp (4HG2LA, 4HG2RA, C01160RA, 34AC5RA, 30DH10RA, C1022RA, A1220RA, 14BD2LA, 14BD2RA, 27CB12RA, 27CB12LA, 27GF2LA, 27GF2RA) upon somatic cell hybrid analysis (examples are given in figure 7.13). However, the hybrid panel used here was only capable of localising STSs to the region Xp21.1-Xp11.22 (section 2.7.8.3). Verification of localisation to Xp11.3-11.22 was attained for the first 10 STSs above by their presence in overlapping YACs (see Figure 7.7), which also indicated that YACs 4HG2 and 14BD2 are intact (non-chimaeric). Figure 7.14 shows examples of YAC STS-content mapping results for STSs 4HG2RA and C1022RA. Insert terminal STSs C01160LA, 34AC5LA and A1220RA map to regions other than Xp11.3-11.22 upon hybrid analysis, indicating that the parent YACs are chimaeric;
Figure 7.12 FASTA database similarity search reveals identity of 34AC5RA to an EST expressed in brain, placenta and ovary tumour tissues.

```
>>GB;H09726 H09726 yl90el0.sl Homo sapiens cDNA clone 45 (611 nt)
initn: 555 initi: 555 opt: 1029 z-score: 934.6 E(): 0
99.5% identity in 210 nt overlap

34AC5R
TACACTGTGATTTGTTGCCAGCATTTATAACTT
X:;;;;;;;;;;;;;;;;;;;;;;;;;

GB:H09 TGGTGGAAAATCCCGGTCAGTTTTTACTCTGATTTTGGATTCTTCACGATTTAAT
160 170 180 190 200 210
40 50 60 70 80 90

34AC5R
GGGAAGCGGCTAAAGGGTTACTGCTTTTACTACTGTGATTTGTTGCCAGCATTTATAACTT

GB:H09 GGGAGTAAGGCTAAATTTCAGTTTCATTGCTGAAAATAAAGATGTAACATTTTCTTCCAT
220 230 240 250 260 270
100 110 120 130 140 150

34AC5R
CAAGTTCATGGTTACCCCTGGCTTCTATCCAGGTTAAGAATCCCTGCCTTTAGGGAAAA

GB:H09 CAAGTTCATGGTTACCCCTGGCTTCTATCCAGGTTAAGAATCCCTGCCTTTAGGGAAAA
280 290 300 310 320 330
160 170 180 190 200 210

34AC5R
TTCTGGACATAATCAGGACACTCCTGAAGAGGTTTAAAGAAGAGGTAAGACCTCACTCAA

GB:R46366 R46366 yg50a06.sl Homo sapiens cDNA clone 35 (470 nt)
initn: 546 initi: 546 opt: 727 z-score: 662.3 E(): 5.5e-29
93.0% identity in 185 nt overlap

34AC5R
TACACTGTGATTTGTTGCCAGCATTTATAACTT
X:;;;;;;;;;;;;;;;;;;;;;;;;;

GB:R46 TGCGAAAATCCCGGTCAGTTTTTACTCTGATTTTGGATTCTTCACGATTTAAT
200 210 220 230 240 250
40 50 60 70 80 90

34AC5R
GGGAAGCGGCTAAAGGGTTACTGCTTTTACTACTGTGATTTGTTGCCAGCATTTATAACTT

GB:R46 GGGAGTAAGGCTAAATTTCAGTTTCATTGCTGAAAATAAAGATGTAACATTTTCTTCCAT
280 290 300 310 320 330
100 110 120 130 140 150

34AC5R
CAAGTTCATGGTTACCCCTGGCTTCTATCCAGGTTAAGAATCCCTGCCTTTAGGGAAAA

GB:R46 CAAGTTCATGGTTACCCCTGGCTTCTATCCAGGTTAAGAATCCCTGCCTTTAGGGAAAA
320 330 340 350 360 370
160 170 180 190 200 210

34AC5R
TTCTGGACATAATCAGGACACTCCTGAAGAGGTTTAAAGAAGAGGTAAGACCTCACTCAA

GB:R46 TTCTGGACATAATCAGGACACTCCTGAAGAGGTTTAAAGAAGAGGTAAGACCTCACTCAA
340 350 360 370 380 390

LOCUS H09726
DEFINITION yl90el0.sl Homo sapiens cDNA clone 45 3'.
ACCESSION H09726
NID g874548
KEYWORDS EST.
SOURCE human clone=45355 library=Soares infant brain INIB

>>GB:R46366 R46366 yg50a06.sl Homo sapiens cDNA clone 35 (470 nt)
initn: 546 initi: 546 opt: 727 z-score: 662.3 E(): 5.5e-29
93.0% identity in 185 nt overlap

34AC5R
TACACTGTGATTTGTTGCCAGCATTTATAACTT
X:;;;;;;;;;;;;;;;;;;;;;;;;;

GB:R46 TGCGAAAATCCCGGTCAGTTTTTACTCTGATTTTGGATTCTTCACGATTTAAT
200 210 220 230 240 250
40 50 60 70 80 90

34AC5R
GGGAAGCGGCTAAAGGGTTACTGCTTTTACTACTGTGATTTGTTGCCAGCATTTATAACTT

GB:R46 GGGAGTAAGGCTAAATTTCAGTTTCATTGCTGAAAATAAAGATGTAACATTTTCTTCCAT
280 290 300 310 320 330
100 110 120 130 140 150

34AC5R
CAAGTTCATGGTTACCCCTGGCTTCTATCCAGGTTAAGAATCCCTGCCTTTAGGGAAAA

GB:R46 CAAGTTCATGGTTACCCCTGGCTTCTATCCAGGTTAAGAATCCCTGCCTTTAGGGAAAA
320 330 340 350 360 370
160 170 180 190 200 210

34AC5R
TTCTGGACATAATCAGGACACTCCTGAAGAGGTTTAAAGAAGAGGTAAGACCTCACTCAA

GB:R46 TTCTGGACATAATCAGGACACTCCTGAAGAGGTTTAAAGAAGAGGTAAGACCTCACTCAA
340 350 360 370 380 390

LOCUS R46366
DEFINITION yg50a06.sl Homo sapiens cDNA clone 35951 3'.
ACCESSION R46366
NID g822365
KEYWORDS EST.
SOURCE human clone=35951 library=Soares infant brain INIB
```
LOCUS H61174 352 bp mRNA EST
DEFINITION yu40a06.sl Homo sapiens cDNA clone 236242 3'.
ACCESSION H61174
NID gi014006
KEYWORDS EST.
SOURCE clone=236242 primer=Promega -21ml3 library=Soares ovary tumor

LOCUS N41839 492 bp mRNA EST
DEFINITION yw72a05.rl Homo sapiens cDNA clone 257744 5'.
ACCESSION N41839
NID gi1014006
KEYWORDS EST.
SOURCE human clone=257744 primer=T7 library=Soares placenta 8to9weeks
Figure 7.13 X-chromosomal localisation of novel YAC-insert terminal STSs by PCR analysis of X-specific somatic cell hybrids. (a) 14BD2RA (b) A1220LA (c) 34AC5RA (d) 27CB12LA. 27CB12LA was simultaneously tested against other YACs in the vicinity. The size standard in all examples is φX174/HaeIII. H: human genomic DNA (positive control), M: mouse DNA, Ha: hamster DNA, Y: yeast DNA, - - No DNA (negative controls).
Figure 7.14 PCR analysis of YACs in the proximal contig using the STSs from (a) 4HG2RA and (b) C1022RA with human genomic DNA (H) and ‘No DNA’ as positive and negative controls respectively.
34AC5LA and A1220RA map to distal Xp (the latter has recently been confirmed by Schindelhauer et al 1996) and C01160LA was mapped to chromosome 5 or 6 by PCR analysis of the Biosmap multi-chromosomal somatic cell hybrid panel (Figure 7.15). The RA and LA insert terminals of YAC 27GF2 mapped to the region of interest on hybrid analysis (data not shown), but did not map back to overlapping YACs in the contig suggesting that this YAC may be rearranged. STSs derived from 4DF10RA and LA failed to successfully amplify products of the expected size from human/parent YAC DNA; perhaps owing to the biased base composition and repetitive nature of the DNA preventing specific primer binding (see Figures 7.10 and 7.16).

Repeat FASTA database identity searches for all YAC End STSs recently identified two further matches to Genbank sequences. 4DF10RA was found to be 98% identical over 156nt to X chromosome STS sWXD1471 (Genbank ID L76151; Figure 7.16). Interestingly, sWXD1471 corresponds to an insert terminal STS (DXS9795) generated from SYP YAC I31F10 (Boycott et al 1996), and is positioned proximal to SYP. It can be seen from Figure 7.16 that DXS9795 PCR primers are virtually identical to those designed in this study for 4DF10RA, therefore the reason for failure of the latter cannot be attributed to unusual DNA sequence. Boycott et al (1996) note, however, that further YACs were unobtainable for DXS9795 or DXS573 (whose location they refined to between SYP and DXS1039) in Xp11.22, highlighting the possibility that DNA from the SYP region is unclonable in YACs. This may explain the difficulties encountered in this study in obtaining a stable isolate of the one SYP YAC identified, 4DF10, and our inability to identify DXS573 YACs. This data does, however, confirm the localisation of 4DF10RA to Xp11.22 by virtue of its presence in cosmid clones containing DXS573 (Boycott et al 1996). The second match was found between 27CB12LA and unpublished X chromosome STS sWXD2505 (Genbank ID L47500; DXS8193; Schlessinger 1995), of 93.6% over 188nt (Figure 7.17). As 27CB12LA is anchored to the region by virtue of its presence in another DXS1039 YAC 34AB12 (Figure 7.13), this therefore maps DXS8139 to within 225kb of DXS1039 in Xp11.22.

7.3.2.2 YAC STS-content analysis and map integration

STS-content mapping generated a fully integrated YAC contig in the proximal portion of the RP2 critical region comprising 34 YACs and encompassing 38 markers including 11 microsatellites, 13 gene/ESTs and 13 STSs derived from YAC insert terminals, 11 of which are novel to this study. The major contig in Xp11.23 contains 22 YACs and 31 markers; with an estimated contig length of ~1Mb based on the additive size of minimum tiling path YACs 31CC3, 24GH5, 23BC7 and 32EH3 this corresponds to an STS density of approximately 1 every 33kb. The depth of the contig permitted physical ordering of most STSs/ESTs assigned to the contig, as shown in Figure 7.7. The order so derived is as follows: SYN1(CA) - TIMP-1 - C1022RA - 4HG2RA - PFC(CA) - 33CA11LA - (DXS426, F0701LA) - A1220LA - 4HG2LA - (DXS1367, ELK1) - 33CA11RA - (ZNF8I, DXS6849, DXS1004E) - DXS6616 - DXS6950 - 34AC5RA - 30DH10RA - (MG6I, DXS6941) - MG8I - (DXS722, MG2I) - (DXS1011E, MG44) - C01160RA - GATA1 - DXS1126 - DXS1240 - Xcen, and is in general agreement
Figure 7.15 The BIOSMAP multichromosomal somatic cell hybrid panel consisted of 25 human-hamster hybrid cell lines which are listed in the Table below together with their human chromosome component. PCR analysis with COI 160LA shows clear concordance with hybrids 860, 904 and 909, which indicates that this STS derives from chromosome 5 or 6.

| Human Chromosome | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | X | Y |
|------------------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
|                  |   |   |   |   |   |   |   |   |   | 15%| 15%| 15%| 15%| 15%| 15%| 15%| 15%| 15%| 15%| 15%| 15%| 15%| 15%| 15%| 15%|
|                  |   |   |   |   |   |   |   |   |   | 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%|
|                  |   |   |   |   |   |   |   |   |   | 55%| 55%| 55%| 55%| 55%| 55%| 55%| 55%| 55%| 55%| 55%| 55%| 55%| 55%| 55%|
|                  |   |   |   |   |   |   |   |   |   | 45%| 45%| 45%| 45%| 45%| 45%| 45%| 45%| 45%| 45%| 45%| 45%| 45%| 45%| 45%|
|                  |   |   |   |   |   |   |   |   |   | 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%|
|                  |   |   |   |   |   |   |   |   |   | 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%|
|                  |   |   |   |   |   |   |   |   |   | 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%|
|                  |   |   |   |   |   |   |   |   |   | 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%|
|                  |   |   |   |   |   |   |   |   |   | 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%|
|                  |   |   |   |   |   |   |   |   |   | 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%|
|                  |   |   |   |   |   |   |   |   |   | 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%|
|                  |   |   |   |   |   |   |   |   |   | 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%|
|                  |   |   |   |   |   |   |   |   |   | 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%| 65%|

Human Chromosomes (D: Deleted at 9p12.1 - 9p13.3, Dq: multiple deletions at 9p)

Percentage values are the percent of the cell population retaining the noted chromosome. "+" indicates > 75%
Figure 7.16 FASTA printout showing identity of 4DF10 RA YAC terminal sequence to STS sWXD1471 (DXS9795). Boxed sequence represents PCR primers designed in this study to generate an STS (the dashed box indicating that this primer had opposite strand sequence). The comparable primer sequences of Boycott et al (1996) are indicated above the alignment.
Amplimer (PCR Primer Pair) sWXD2505

Name: sWXD2505
Accession ID: GDB:637651
Owner: Schlessinger, David
Status: Active

<table>
<thead>
<tr>
<th>Alias Name</th>
<th>Authority</th>
</tr>
</thead>
<tbody>
<tr>
<td>DXS8193</td>
<td>HUGO DNA</td>
</tr>
<tr>
<td>sWXD2505</td>
<td>Schlessinger, David</td>
</tr>
<tr>
<td>sWXD2505a</td>
<td>Schlessinger, David</td>
</tr>
<tr>
<td>sWXD2505b</td>
<td>Schlessinger, David</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Left Marker</th>
<th>Right Marker</th>
<th>Map Element ID</th>
<th>Map</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>Xpl1.21</td>
<td>Xpl1.21</td>
<td>GDB:1101564</td>
<td>X Chromosome Map</td>
</tr>
<tr>
<td>X</td>
<td>Xpl1.21</td>
<td>Xpl1.21</td>
<td>GDB:122152578</td>
<td>X Chromosome Map</td>
</tr>
</tbody>
</table>

Related Sequences: L47500

<table>
<thead>
<tr>
<th>Relationship</th>
<th>Marker</th>
<th>Observation</th>
<th>Order ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overlaps</td>
<td>Amplimer (PCR Primer Pair): 27CB12 LA Shares sequence identity with</td>
<td>GDB:5386088</td>
<td></td>
</tr>
</tbody>
</table>

Nucleic Acid Sequences: L47500

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Primer Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>sWXD2505a</td>
<td>GCTCCTGTCTGATGCTAAG</td>
</tr>
<tr>
<td>sWXD2505b</td>
<td>TTCTCCTAGAATACGGCATC</td>
</tr>
</tbody>
</table>

Restrictions: Yes

Figure 7.17 FASTA printout showing identity of 27CB12 LA terminal sequence to STS sWXD2505 (DXS8193) and corresponding annotated GDB entry.
with other comprehensive physical maps covering the region (Boycott et al 1996, Schindelhauer et al 1996).

The transcriptional map of the region has been enriched significantly by the high resolution mapping of several ESTs to the contig. Thus the physical localisations of DXS1011E, DXS1004E and DXS1007E (Parrish and Nelson 1993) have been refined, DXS1007E to *SYP* YAC 4DF10 in Xp11.22, DXS1011E colocalises with retinally expressed gene *MG44* in Xp11.23 and DXS1004E colocalises with ZNF8I in Xp11.23 (Figure 7.7). FASTA database searches for these three ESTs showed that DXS1007E is identical to *SYP* and DXS1004E corresponds to *ZNF41*, supporting the notion, based on sequence similarity, that *ZNF41* and ZNF8I arose by duplication (Knight et al 1994). DXS1011E does not show identity to *MG44* or *MG21* and thus represents an as yet uncharacterised gene. Failure of DXS1007E to identify YACs upon library screening provided further evidence of instability around *SYP* (section 7.3.2.1). Interestingly, our localisation of DXS1004E (*ZNF41*) differs from that in published reports (Knight et al 1994, Mazzarella et al 1994, Carrel et al 1996). In contrast to the report of Knight et al (1994), repeated testing in our laboratory confirmed YAC C1022 to be positive for DXS1004E. The additional presence of DXS1004E in YACs 3EB3, 24GH5 and 31AG10 which overlap C1022, positioned ZNF4I proximal to ELKI and distal to DXS6616 (Figure 7.7). The presence of STS 34AC5RA in YAC 31GA5 but not 30DH10, 19IH2 or 27GF2 positions this STS at the distal end of 34AC5 (Figure 7.7), thereby mapping a new EST to the physical map of Xp11.23 between genes ZNF8I and MG61.

The incorporation of five recently published microsatellites (DXS1367, DXS6849, DXS6616, DXS6950 and DXS6941) and STS F0701LA into the Xp11.23 contig consolidated YAC overlaps and in the case of DXS6950 confirmed the potential overlap between YACs 23BC7, 31GA2 and 31AG10, simultaneously refining the localisation of 34AC5RA to between DXS6950 and MG61. This physical link across the OATLI region is depicted in Figure 7.7, while Figure 7.18 shows products obtained from key YACs using DXS6950. Furthermore, inclusion of markers F0701LA, DXS6849 and DXS6950, together with the presence of ICRF YACs A1220, C1022, C1228 and C01160 has enabled integration with other partial contigs already existing for Xp11.23 (Fisher et al 1995, Kwan et al 1995, Coleman et al 1994, Knight et al 1994, Hagemann et al 1994), which reflect the interest in this gene-rich region of Xp, to which many genetic diseases have been mapped (Nelson et al 1995; section 1.8.4).

Three gaps remain in the overall contig between DXS1240 and *SYP*, *SYP* and DXS1039 and DXS1039 and DXS255, since the refinement of the RP2 proximal boundary focussed efforts more distally in Xp11.23. The relative order of *SYP* and DXS1039 has recently been reported in the physical maps of other investigators (Fisher et al 1995, Boycott et al 1996). Using strategies akin to those of this study, the former group generated two YAC contigs in the Xp11.23-p11.22 interval which established physical linkage between GATAI and SYP and between DXS1039 and DXS255, thus placing SYP distal to DXS1039. Several YACs and markers generated during their work are identical to those of this thesis, namely YACs C01160 and 27GF2 (in Xp11.23), 27CB12, 16CD8, 36HB8 and 22AB3 (Xp11.22) and marker 27CB12LA (Xp11.22). This confirms the identity of these YACs and also orients 27CB12LA
Figure 7.18 PCR amplification of key YACs in the proximal YAC contig with marker DXS6950 establishing physical linkage across the OATL1 pseudogene cluster.

Figure 7.19 STS-content mapping results of YACs encompassing OATL1, with ESTs derived from the MG cDNAs is shown below, with the YAC name above each lane. YAC 34AC5 is absent from this MG61 screen, but has proved consistently positive for MG61 upon separate testing.
(and hence DXS8193; section 7.3.2.1) to the proximal end of YAC 27CB12 (Fisher et al 1995), as shown in Figure 7.7. The authors also found that YACs containing GATA1 in Xp11.23 (including 27GF2) and SYP in Xp11.22 are susceptible to deletion and propose region-specific instability around these loci, which is supported by our study (section 7.3.2.1). Boycott et al (1996) report a 2Mb physical contig composed of YACs and cosmids spanning ZNF21 (Xp11.23) to DXS255 (Xp11.22), confirming the order Xpter-GATA1-TFE3-SYP-DXS1039-DXS255-Xcen.

The physical order and orientation of the retinally expressed cDNAs Xpter-MG61-MG81-MG44-MG21-Xcen by YAC STS-content mapping is in agreement with that of Geraghty et al (1993) and as depicted on the recently published physical map of Boycott et al (1996). Figure 7.19 shows the results of STS-content mapping all YACs in the OATL1 region with the MG cDNA PCR primers. Based on the most parsimonious order, YAC 34AC5 appears to be internally deleted for MG81 (Figure 7.7). The physical mapping data presented here clearly demonstrates that DXS6616, the new proximal flanking marker for RP2, lies distal to retinally expressed genes MG21, MG44, MG61 and MG81, thereby excluding them from involvement in the disease.

YACs in the distal portion of this contig were also tested for more distally located markers i.e. microsatellites DXS1003, DXS1055 and gene UBEI (Figure 7.1). Although DXS1003 and DXS1055 did not map to any of these YACs, UBEI was found to be present in YAC C1228 (data not shown). This apparent physical linkage of UBEI to the ARAFI/SYNI/TIMP1/PFC cluster has been found by other investigators based on YAC C1228 (Hagemann et al 1994) and is discussed more fully in the next section.

7.3.3 YAC contig construction in the central RP2 interval: extension of the physical map between TIMP1 (Xp11.23) and distal flanking marker DXS8083 (Xp11.3)

7.3.3.1 Identification of seed YACs to nucleate contig construction in a region of low marker density

This section describes the assembly of the YAC contig depicted in Figure 7.20. Contig construction in the Xp11.3-Xp11.23 region was initiated by PCR-based screening of the 4X ICI YAC library with microsatellite markers DXS1003 and DXS1055 and primers for gene UBEI (Figure 7.1). The resulting YACs and sizes are listed in Table 7.7. No other published sequence data was available for this unchartered region of proximal Xp from which to create further STSs to facilitate contig construction. In order to provide novel anchored STSs from which to walk bidirectionally from the UBEI and DXS1003/DXS1055 YAC clusters, YAC insert-terminal STSs were therefore sought by either ALU-vector PCR or Vectorette PCR (section 2.7.8) for 37FG2 and 37GA5 (the largest UBEI YACs) and DXS1003 YAC 24BF12 (parallel studies were performed by R.M.Hampson for DXS1055 YACs 34AH4 and 35CC10). Three novel STSs were successfully generated corresponding to 37FG2LA,
**Table 7.7: Confirmed positive YACs within the RP2 interval in Xp11.3-p11.23**

<table>
<thead>
<tr>
<th>YAC name</th>
<th>Identifying marker</th>
<th>YAC library</th>
<th>YAC size (kh)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>8DF4</td>
<td>UBE1</td>
<td>ICI</td>
<td>270</td>
<td></td>
</tr>
<tr>
<td>11AD10</td>
<td>UBE1</td>
<td>ICI</td>
<td>380</td>
<td></td>
</tr>
<tr>
<td>37FG2</td>
<td>UBE1</td>
<td>ICI</td>
<td>680</td>
<td></td>
</tr>
<tr>
<td>37GA5</td>
<td>UBE1</td>
<td>ICI</td>
<td>430</td>
<td>Chimaeric</td>
</tr>
<tr>
<td>36GB12</td>
<td>DXS1003</td>
<td>ICI</td>
<td>1000</td>
<td>Screen and sizing by S. Kamakari</td>
</tr>
<tr>
<td>19DH7</td>
<td>DXS1003</td>
<td>ICI</td>
<td>?</td>
<td>Screen by S. Kamakari</td>
</tr>
<tr>
<td>24BF12</td>
<td>DXS1003</td>
<td>ICI</td>
<td>nt</td>
<td></td>
</tr>
<tr>
<td>y900D1151</td>
<td>DXS1003</td>
<td>ICRF</td>
<td>530</td>
<td>RLDB (Coleman et al. 1994)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Potentially deleted; this study</td>
</tr>
<tr>
<td>714D3</td>
<td>DXS1003</td>
<td>CEPH</td>
<td>490</td>
<td>Chimaeric (Xp/17;infoclone)</td>
</tr>
<tr>
<td>964C7</td>
<td>DXS1003</td>
<td>CEPH</td>
<td>1750</td>
<td></td>
</tr>
<tr>
<td>925E9</td>
<td>DXS1003</td>
<td>CEPH</td>
<td>880,1140,1520</td>
<td></td>
</tr>
<tr>
<td>995E4</td>
<td>DXS1003 n</td>
<td>CEPH</td>
<td>1240</td>
<td></td>
</tr>
<tr>
<td>18CG12</td>
<td>DXS1003</td>
<td>ICI</td>
<td>500</td>
<td>Coleman et al 1994</td>
</tr>
<tr>
<td>34AH4</td>
<td>DXS1055</td>
<td>ICI</td>
<td>400</td>
<td>Screen and sizing by RM Hampson</td>
</tr>
<tr>
<td>35CC10</td>
<td>DXS1055</td>
<td>ICI</td>
<td>360</td>
<td>Screen and sizing by RM Hampson</td>
</tr>
<tr>
<td>779D11</td>
<td>DXS1055</td>
<td>ICI</td>
<td>1000</td>
<td></td>
</tr>
<tr>
<td>20DD2</td>
<td>DXS1146</td>
<td>ICI</td>
<td>280</td>
<td></td>
</tr>
<tr>
<td>y900F11105</td>
<td>DXS1146</td>
<td>ICRF</td>
<td>nt</td>
<td></td>
</tr>
<tr>
<td>y900H11108</td>
<td>DXS1146</td>
<td>ICRF</td>
<td>nt</td>
<td></td>
</tr>
<tr>
<td>21IA10</td>
<td>DXS337</td>
<td>ICI</td>
<td>150</td>
<td></td>
</tr>
<tr>
<td>21IA12</td>
<td>DXS337</td>
<td>ICI</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>21IB10</td>
<td>DXS337</td>
<td>ICI</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>y900D076</td>
<td>DXS337</td>
<td>ICRF</td>
<td>nt</td>
<td></td>
</tr>
<tr>
<td>17AE5</td>
<td>DXS337</td>
<td>ICI</td>
<td>nt</td>
<td></td>
</tr>
<tr>
<td>y900C1023</td>
<td>PCTK1</td>
<td>ICRF</td>
<td>420</td>
<td>Chimaeric; Knight et al 1995</td>
</tr>
<tr>
<td>6AG10</td>
<td>ZNF157</td>
<td>ICI</td>
<td>nt</td>
<td></td>
</tr>
<tr>
<td>y900C0528</td>
<td>ZNF157/DXS1266</td>
<td>ICRF</td>
<td>nt</td>
<td></td>
</tr>
<tr>
<td>y900C0628</td>
<td>ZNF157/DXS1266</td>
<td>ICRF</td>
<td>nt</td>
<td></td>
</tr>
<tr>
<td>y900C0728</td>
<td>ZNF157/DXS1266</td>
<td>ICRF</td>
<td>nt</td>
<td></td>
</tr>
<tr>
<td>y900B10123</td>
<td>DXS1266</td>
<td>ICRF</td>
<td>nt</td>
<td></td>
</tr>
</tbody>
</table>
Figure 7.20 YAC contig spanning the central portion of the RP2 critical interval in Xp11.23 (not to scale). The horizontal line with bidirectional arrows represents the chromosome, with STSs above in the most likely order determined by YAC content. The bold horizontal lines represent YACs (see Table 7.7 for sizes and source library). A filled rectangle at the end of a YAC indicates a terminal sequence STS that maps to overlapping YACs in the contig. A hatched box at the end of a YAC depicts a terminal sequence STS that does not derive from this region of Xp11.23. Vertical broken lines show the positions of STSs in YAC clones, with a solid circle indicating a verified positive. Based on the most parsimonious order, several YACs appear to contain internal deletions ( ). YACs with right-hand arrows link up with the proximal contig (Figure 7.7), while those with dashed lines have not been tested for more proximal markers. SYN1 has not been tested on all YACs.
37GA5LA, 37GA5RA and 24BF12RA. The 71bp insert sequence obtained from ~480bp PvuII Vectorette library PCR products revealed that 37FG2LA and 37GA5LA were identical (data not shown). Sequences representing 37GA5RA and 24BF12RA were obtained by ALU-vector PCR; 168bp from ~570bp ALE3-pYAC4RA PCR product for 37GA5RA and 182bp from a ~400bp ALE1-pYAC4RA PCR product for 24BF12RA (data not shown).

FASTA database identity searches revealed 85-90% identity over the entire 37FG2/37GA5LA sequence to >25 different human Genbank DNA sequence entries. Retrieval of a few such entries revealed that they shared a stretch of THE-LTR repeat corresponding to the region of identity with 37FG2/37GA5LA (data not shown). These interspersed elements are estimated to occur every 100-300kb in human DNA (Fields et al 1992) and the occurrence of one such repeat at the LA insert-terminal of both 37FG2/37GA5 precluded STS development from this sequence. Similar analysis for 37GA5RA demonstrated no identity to known sequences and therefore primers were designed within the usual constraints (section 2.7.3) to amplify as large a DNA fragment as possible. However, this STS showed no amplification product from all the X-chromosomal hybrids (data not shown), suggesting that it is of autosomal origin and that YAC 37GA5 is chimaeric. Similar analysis for 24BF12RA, however, suggested a location within the region of interest on proximal Xp and this novel STS was subsequently used to confirm overlaps within several YACs (see Figure 7.20). FASTA analysis of this YAC end-sequence showed 99.5% reverse strand identity over 183nt to a human fibroblast cDNA (IMAGE Consortium; Genbank accession W47657) and 100% forward strand identity over 54nt to a different Genbank entry from the same cDNA library, EST accession W47658 (Figure 7.21). Comparison of these two EST sequences showed that they represent opposite strands of the same cDNA (data not shown). The first ~50 bp of 24BF12RA forward strand sequence is ALU-similar, as confirmed by this region showing further identities of ~90% to several other human ESTs, some of which were noted for ALU identity (data not shown). Although the forward PCR primer for 24BF12RA is therefore ALU-similar, the reverse primer designed to unique sequence enabled successful STS generation and the mapping of this EST to Xp11.3-p11.23 by hybrid analysis.

In the absence of further markers in the region, an alternative strategy to increase YAC coverage was to search the CEPH database via the QUICKMAP Infoclone program for megabase-sized YACs corresponding to existing STSs. Many larger YACs were available for markers DXS1003 and DXS1055 including putative neighbouring YACs based on ALU fingerprint overlaps (Cohen et al 1993). A total of 16 YACs were selected of various sizes in an attempt to span as wide an interval as possible. Only 5 of these YACs were confirmed positive for DXS1003 or DXS1055 in this study, and are listed in Table 7.7. Recent data from the CEPH database suggests that the DXS1003/DXS1055 ‘neighbouring YACs’ selected for this study on the basis of ALU-PCR relation are of questionable X location or likely to derive from autosomes.

Additional markers and YACs were incorporated during the course of this study to extend and consolidate the developing contig and are included in Tables 7.2 and 7.7 respectively:
Figure 7.21 FASTA printouts showing (a) forward and (b) reverse strand identity of 24BF12 RA YAC end sequence to an EST derived from a human fibroblast cDNA. The forward strand region of homology is ALU-similar.
YACs D1151, 18CG12 (Coleman et al 1994) and C1023 (Knight et al 1995), microsatellites DXS337, DXS1146 and DXS7124 broadly localised to Xp11.3-p11.23, STSs DXS1264 (LA insert end of DXS1003/DXS1055 CEPH YAC 76E10; Coleman et al 1994) and DXS1266 (ALU PCR product physically mapped between DXS1146 and TIMP-1; Coleman et al 1994), and gene markers PCTK1, ZNF157, UHX1 and DXS8237E which had been physically mapped close to UBE1 and TIMP-1 (see Figure 7.1). PCTK1 and ZNF157 are briefly described in section 1.8.4.2 in connection with the 1995 XCW map. DXS8237E is a widely expressed transcript that derives from a cosmid containing UBE1 (Coleman et al 1996); detailed physical mapping studies by Carrel et al (1996) have placed PCTK1, UBE1 and DXS8237E within a 50kb interval, in the order PCTK1-UBE1-DXS8237E, with PCTK1 ~5kb from UBE1. UHX1 is a novel ubiquitin C-terminal hydrolase gene localised to Xpl 1.4-11.23 (Swanson et al 1996). These newer STSs were tested on existing YACs and YAC library screening performed with PCTK1, DXS337, DXS1146, DXS1266 and ZNF157 to identify further YACs (Table 7.7).

7.3.3.2 A ‘best estimate’ YAC contig derived by integration of the genetic, physical and transcriptional maps of Xp11.23

STS-content mapping of all identified YACs with 14 markers DXS1264, DXS1055, DXS1003, 24BF12RA, DXS1146, DXS1266, UHX1, PCTK1, UBE1, DXS8237E, DXS7124, ZNF157, DXS337 and TIMP-1 established three clusters of YACs corresponding to markers (DXS1264, DXS1003, 24BF12RA, DXS1055, DXS1146), (UHX1, PCTK1, UBE1, DXS8237E, DXS7124) and (ZNF157, DXS337, TIMP-1) as depicted in Figure 7.20. Figure 7.22 shows STS-content results for all YACs with markers TIMP-1, DXS337, ZNF157, DXS1146 and DXS7124. This map localises two new microsatellite markers, DXS1146 and DXS337 to the RP2 critical region in Xp11.23 by physical linkage to markers DXS1003 and TIMP-1 respectively. The order Xpter-DXS1055-DXS1003-DXS1146-Xcen is shown in Figure 7.20 in accordance with the genetic order Xpter-DXS1055-DXS1003-Xcen derived by Weissenbach et al (1992). The presence of both DXS1003 and DXS1146 in YAC 714D3 places DXS1146 within 490kb of DXS1003, substantially refining its genetic localisation between DXS8080 and DXS337 obtained via meiotic breakpoint mapping (section 6.3.2). From the YACs amplified by 24BF12RA (see Figure 7.20), the EST it represents is mapped between DXS1055 and DXS1003, and may therefore correspond to one of the CpG islands mapped to this interval by Coleman et al (1994). DXS7124 is seen to colocalise with UBE1 in all cases (Figure 7.20) and thus is not physically separable from UBE1 with this set of YACs. Physical linkage between microsatellite DXS337 and TIMP-1 in 6 separate YACs (see Figure 7.20) strongly indicates a location for DXS337 within 100kb distal to TIMP-1 (the size of YAC 21IA12). This is supported by meiotic breakpoint mapping data placing DXS337 between DXS426 and DXS1003/DXS1055 in Xp11.23 (section 6.3.2), and significantly refines this genetic localisation. STS-content data positions ZNF157 distal to DXS337 and places ZNF157, DXS337 and TIMP-1 within a 100kb interval as defined by the size of YAC 21IA12 which.
Figure 7.22 Agarose gels showing STS-content results for all YACs in the Xp11.3- p11.23 'central' RP2 region contig with the markers indicated. The cluster of YACs around DXS8083/DXS8026 (Figure 7.4) were also included. The size standard on all gels is φX174/HaeIII; human genomic DNA (H) and 'no DNA' (-) are positive and negative controls.
ZNF157: all YACs not shown were negative for this EST.
contains them all (see Figure 7.20). This localisation of ZNF157 confirms the physical map location of Derry et al (1995) and Carrel et al (1996) between UBE1 and TIMP-1.

These studies also show that ZNF157 is present in YAC 8DF4 (Figure 7.22), thus physically links the UBE1 gene cluster to the TIMP1/ARAF1/SYN1/PFC gene cluster in Xp11.23, a result which has been previously reported based on coretention of UBE1 and TIMP-1 in YAC C1228 (Hagemann et al 1994). However, other investigators have excluded UBE1 from a YAC contig extending from DXS1055 to OATLI and which included YAC C1228 (Coleman et al 1994, Knight et al 1995). However, the C1228 YAC in these studies sized at only 240kb, whereas the version in the map presented here and that of Hagemann et al (1994) was ~320kb (Figure 7.8), thus it is possible that the former isolate of this clone is deleted for sequences around UBE1. In this study, meiotic breakpoint mapping and the analysis of critical crossovers (chapter 6) have served to position DXS7124, a microsatellite situated very close to UBE1 (section 6.3.2) proximal to DXS1003/DXS1055, thus supporting the order shown in Figure 7.20. The report of a different YAC shown to contain both UBE1 and ZNF157 in Carrel et al (1996) provides further evidence that UBE1 lies proximal to DXS1003/DXS1055. On this basis our isolate of YAC C1228 appears to be internally deleted for ZNF157 (Figure 7.22). It would be interesting to find out if the smaller version of this YAC reported in Coleman et al (1994) contained any of the STSs DXS8237E, DXS7124, ZNF157 or DXS337.

The gene UHX1 was successfully incorporated into the physical map of the region by its presence in YACs 11AD10 and 37FG2 (data not shown), thus locking this gene into the PCTK1/UBE1/DXS8237E gene cluster and indicating that YACs 11AD10 and 37FG2 extend distal to C1023 and 37GA5 (Figure 7.20). However, this EST did not detect overlaps between the UBE1 YACs and DXS1003 YACs, and was insufficiently robust to detect further YACs by screening the library pools, therefore a potential gap here still remains.

7.3.3.3 ALU-PCR fingerprints to confirm YAC order and orientation

ALU-PCR fingerprinting on selected YACs was performed in an attempt to confirm the YAC overlaps established by shared internal STSs and the physical localisation of UBE1 proximal to DXS1003/DXS1055. Figure 7.23 shows the fingerprints obtained using primers ALE1 and ALE3 as described in section 2.7.7. The complex pattern of bands indicated that this region is ALU-rich, consistent with its high gene density (Bickmore and Sumner 1989).

A cursory glance at these fingerprints clearly demonstrates shared bands and therefore likely overlaps within the DXS1003/DXS1055/DXS1146 YAC cluster and the TIMP-1/DXS337/ZNF157 YAC cluster but is not so obvious for the UBE1 YACs. However, several key ALU products provide support for the relative order of YAC clusters given in Figure 7.20. A strong ~400bp ALE1 fragment is common to YACs C1228, A1220, C0628, 8DF4, 6AG10 and faintly in YAC 37FG2, but is not amplified from YACs 925E9, D1151, 20DD2, 714D3, 24BF12, 964C7 or 779D11, providing strong support for the overlap between the UBE1 and TIMP-1 YACs deduced from STS-content mapping. Similar conclusions can be drawn by comparing the YACs which share an ~250bp ALE1 fragment (Figure 7.23a). Several ALE1 products of YAC 20DD2 appear identical in size to those of YAC 11AD10 (Figure 7.23a)
Figure 7.23 Agarose gels showing ALU-PCR fingerprints of key YACs in the central RP2 region contig in Xp11.3-11.23. Shared bands indicative of YAC overlaps are discussed in the text (section 7.3.3.3) and are highlighted above by arrows.
which potentially indicate physical linkage between DXS1146 and the UBEI gene cluster, undetected by STS-content analysis alone. This is supported by a ~350bp ALE1 fragment common to YACs D1151, 714D3, 24BF12, 964C7, 925E9 and 37FG2 suggesting a potential overlap between the latter YAC (the longest UBEI YAC) and the DXS1003 YACs. As this fragment is not present in the ZNF157/DXS337 YACs the order Xpter-DXS1003-UBEI-ZNF157/DXS337-Xcen is strongly favoured by this result, which also suggests that DXS1055 lies distal to DXS1003 by absence of this fragment from YAC 779D11 (Figure 7.23a) and thus confirms the orientation of the distal portion of the contig (Figure 7.20).

Examination of the ALE3 fingerprint reveals a ~450bp PCR product common to YACs C1228, 8DF4, 11AD10 and 6AG10 alone, further supporting the notion of physical linkage between the UBEI and TIMP-1 YACs via ZNF157 and suggesting that DNA close to ZNF157 must be present in YAC C1228 (Figure 7.23b;7.20). A ~300bp fragment shared by YACs C1228, C0628, 8DF4, 11AD10 is also present in YACs 925E9 and D1151 and potentially in 714D3 and 24BF12 (Figure 7.23b). This result would be consistent with the PCR product deriving from DNA close to marker DXS1266, which by STS content analysis of all available YACs (data not shown) appears to map to two alternative locations within the region, one between DXS1146 and PCTKI and the second between DXS7124 and ZNF157 (Figure 7.20). Either this sequence is duplicated within Xp11.23, or certain YACs must contain internal deletions for the UBEI gene cluster. If YACs D1151 and 925E9 are intact, then DXS1266 must lie proximal to DXS1146 and distal to the UBEI gene cluster, requiring that YAC C1228 be doubly deleted at separate sites for ZNF157 and PCTKI and that YACs C0528, C0628, C0728 and B10123 be deleted for PCTKI, UBE1, DXS8237E and DXS7124. The alternative location just distal to ZNF157 is more favourable as it requires that only 3 YACs are internally deleted; 8DF4 for DXS1266, and 925E9 and D1151 for the UBEI gene cluster; this is the order shown in Figure 7.20. The localisation of DXS1266 between DXS1146 and SYN1 by Coleman et al (1994) is based on its presence in YACs D1151 and C1228, detecting an overlap which the authors claim links DXS1003 to the TIMP-1/ARAF1/SYN1/PFC gene cluster. As discussed above, this data is discordant with that of Carrel et al (1996) linking UBE1 to TIMP-1, and is hard to reconcile with our genetic data that places UBE1-associated microsatellite DXS7124 proximal to DXS1003 (section 6.3.2).

The combination of information on marker order from meiotic breakpoint mapping (chapter 6) and STS-content/ALU fingerprint data on the 34 YACs presented here would in summary appear to suggest the following provisional order for markers within Xp11.3-p11.23:


If this order is correct, and the physical linkage between UBE1 and TIMP-1 can be confirmed, then the distance between UBE1 and ZNF157 is ~270kb (the size of YAC 8DF4)
and that between \textit{UBE1} and \textit{TIMP-1} is likely to be no greater than 320kb (size of YAC C1228). This is a smaller estimate than that of Carrel et al (1996) of ~550kb between \textit{UBE1} and \textit{TIMP-1}, based on the smaller YACs used in our analysis. Although distance estimates between markers are tentative in this region at present where so many YACs appear to be internally deleted/chimaeric, coretention in YACs D1151 and 925E9 suggest that DXS1055, 24BF12RA, DXS1003 and DXS1146 are clustered in a 530kb interval, and from STS-content analysis of YAC 34HA4, that DXS1264 is no more than 360kb from DXS1055. Thus the DXS1264-\textit{TIMP-1} interval may encompass no more than ~900kb.

Attempts are presently underway to further characterise this set of YACs in order to verify the weak link in the contig between DXS7124 and \textit{ZNF157}, which although strongly supported by YAC ALU fingerprinting is only firmly established in YAC 8DF4 by STS-content (Figure 7.20). Strategies being employed include YAC insert-terminal STS generation from key YACs and STS-content mapping of cosmids reported to contain \textit{ZNF157} and \textit{UBE1} (Derry et al 1995; Coleman et al 1996). Further characterisation of YAC ICRFy900E1076 from which DXS1266 is derived may also provide further information regarding relative placement of this marker in Xp11.23, together with identification of DXS1266 ICI YACs which may then be tested for other STSs in the region. The putative overlaps between the DXS1003/DXS1055 YAC cluster and \textit{UBE1} YACs indicated by the ALU fingerprint data may be confirmed by end-cloning key YACs in the distal portion of the contig to create additional STSs between DXS1146 and \textit{UHX1}. Indeed, as cloned DNA is now isolated for all available STSs in this region, generating novel STSs may be the simplest way to establish a definitive contig with this set of YACs. However, as the region around \textit{UBE1} appears to be inherently unstable in YACs, as demonstrated by multiple internal deletions and YAC chimaerism (A1220, section 7.3.2.1; 37GA5, section 7.3.3.1; C1023, Knight et al 1995) a physical contig encompassing the Xp11.3-p11.23 interval which faithfully represents the underlying DNA may only be achieved in the form of smaller, more stable DNA clones such as cosmids, which may be small enough to avoid containing a number of sequences that jointly cause instability.

7.3.4 Summary of YAC contig construction in the RP2 gene critical interval

A total of 30 STSs mapped to the Xp11.3-Xp11.22 region have been used to identify YAC clones from the ICI, ICRF and CEPH megaYAC libraries by a PCR-based screening strategy. 35 additional STSs were used in PCR analysis of these and supplementary YACs from published reports to determine the STS content of each YAC and hence to establish overlaps and assemble contigs. The order of most markers could be established by the presence of two or more markers within a single YAC and comparison with overlapping clones. The results show that STS content of the YACs could provide a finer level of resolution than the genetic map (e.g. for DXS337, DXS8080).

A total of 23 YAC insert-terminal STSs were generated to increase regional marker density and thus the efficacy of the YAC map. All except 3 fulfilled STS criteria by
### Table 7.8  Novel YAC insert-terminal STSs mapping to Xp11.3-p11.22

<table>
<thead>
<tr>
<th>STS name</th>
<th>PCR primer sequences</th>
<th>$T_a$ (°C)</th>
<th>Product size</th>
<th>Map location</th>
<th>Accession number</th>
</tr>
</thead>
<tbody>
<tr>
<td>9BA10 LA</td>
<td>F: 5' AAGGAGGAACAGATAAAAGTGTG 3' R: 5' ATTAAGCAGAATTTCACAGATG 3'</td>
<td>59</td>
<td>96 bp</td>
<td>Xp11.3</td>
<td>DXS9883</td>
</tr>
<tr>
<td>40HD1 LA</td>
<td>F: 5' CCAAATATGCTAATCTACTATGC 3' R: 5' CTTTTAAGCAGATGTGGTTAG 3'</td>
<td>59</td>
<td>172 bp</td>
<td>Xp21.1-11.22</td>
<td>DXS9884</td>
</tr>
<tr>
<td>24BF12 RA</td>
<td>F: 5' AGCCCTGTTGCTCTGAAGTTTC 3' R: 5' CCTACATAGCGTACAGCAGC 3'</td>
<td>67</td>
<td>140 bp</td>
<td>Xp11.23</td>
<td>DXS9886E</td>
</tr>
<tr>
<td>C1022 RA</td>
<td>F: 5' ATTAATACCGACACGGCATGTTG 3' R: 5' AGAGGAGCTTCTGGTGTGGC 3'</td>
<td>69</td>
<td>125 bp</td>
<td>Xp11.23</td>
<td>DXS9875</td>
</tr>
<tr>
<td>4HG2 LA</td>
<td>F: 5' CTTGAAAAGGACACGACAAGC 3' R: 5' TGTCCTTGAGGTGATAGGC 3'</td>
<td>57</td>
<td>162 bp</td>
<td>Xp11.23</td>
<td>DXS9872</td>
</tr>
<tr>
<td>4HG2 RA</td>
<td>F: 5' TGCCACTCAGTTGGCAGTATG 3' R: 5' CCTCTCCAGGAAATACCTTC 3'</td>
<td>63</td>
<td>125 bp</td>
<td>Xp11.23</td>
<td>DXS9873E</td>
</tr>
<tr>
<td>A1220 LA</td>
<td>F: 5' TGCAGGCACAGATGTAATCCAGT 3' R: 5' AGGTGTTGGCATAAAAAGTCAGCGC 3'</td>
<td>62</td>
<td>144 bp</td>
<td>Xp11.23</td>
<td>DXS9892</td>
</tr>
<tr>
<td>34AC5 RA</td>
<td>F: 5' TACACTGTTGTGGTGGCAGC 3' R: 5' AGGAGGTCCTGATATGTCGC 3'</td>
<td>64</td>
<td>177 bp</td>
<td>Xp11.23</td>
<td>DXS9853E</td>
</tr>
<tr>
<td>30DH10 RA</td>
<td>F: 5' AGGTATACCTGCAAGCAAGAC 3' R: 5' TACAGACTGGGATGGCTCGAG 3'</td>
<td>61</td>
<td>129 bp</td>
<td>Xp11.23</td>
<td>DXS9854</td>
</tr>
<tr>
<td>C01160 RA</td>
<td>F: 5' ACACCTGAGAGGCAATACCTGG 3' R: 5' TGCCATTTGCGTCTCGGTGG 3'</td>
<td>58</td>
<td>172 bp</td>
<td>Xp11.23</td>
<td>DXS9874</td>
</tr>
<tr>
<td>27CB12 LA</td>
<td>F: 5' CAGATTGTAATATGTACCTGG 3' R: 5' CTTTCCAGCTTCCTCAATG 3'</td>
<td>60</td>
<td>154 bp</td>
<td>DXS9887</td>
<td></td>
</tr>
<tr>
<td>27CB12 RA</td>
<td>F: 5' AACATATCTTCTTGGAATAGACTG 3' R: 5' AGGTGACGTGGGACATG 3'</td>
<td>60</td>
<td>146 bp</td>
<td>Xp11.22</td>
<td>DXS9888</td>
</tr>
<tr>
<td>14BD2 LA</td>
<td>F: 5' TGAATGTGGCGCAAGATGCTGG 3' R: 5' TACAAAACATGGGCTGACTAG 3'</td>
<td>60</td>
<td>141 bp</td>
<td>Xp11.22</td>
<td>DXS9890</td>
</tr>
<tr>
<td>14BD2 RA</td>
<td>F: 5' AACAGGAACGACTACGACTCCAGC 3' R: 5' AGAGAATCCCTGTCCTGTCG 3'</td>
<td>62</td>
<td>141 bp</td>
<td>Xp11.22</td>
<td>DXS9889</td>
</tr>
</tbody>
</table>
amplifying a unique site from human genomic DNA as evidenced by a single PCR product of the expected size. FASTA database identity searches thus proved generally sufficient to highlight repetitive sequence stretches and aid suitable primer design. PCR analysis of somatic cell hybrids and overlapping YACs, together with published map locations for 2 STSs indicated that 16 of these novel markers map back to Xp11.3-p11.22, 2 map elsewhere on the X chromosome and 2 are autosomal in origin. Five novel STSs mapping to the region showed identity to published sequences; 3 to expressed sequences (4HG2RA, 34AC5RA, 24BF12RA), 2 to STSs (4DF10RA, 27CB12LA).

Figure 7.24 shows a chart summarising the YAC contigs assembled in Xp11.3-p11.23 comprising 43 YACs, the minimum number required to establish orders for most STSs. The remaining YACs were omitted for simplicity of presentation but the overall number of YACs positive for each marker is reported. A variable redundancy is apparent, but it is not possible to determine whether this is the result of an unequal representation in the original libraries or is due to incomplete detection during the screening procedure. As many YAC addresses were incomplete (missing either the row or column dimension) it is likely that the latter was partially responsible, although attempts were made to resolve such ambiguities by screening with adjacent STSs or screening all possible clones in the degenerate set of addresses.

There are 5 clear gaps in the contig as indicated in Figure 7.24. The DXS1240-SYP, SYP-27CB12LA and 27CB12RA-DXS255 gaps have recently been bridged by other researchers (e.g. Boycott et al 1996). Within the contigs, each set of markers is ‘doubly-linked’ to adjacent markers (i.e. both present in 2 separate YACs) with 3 exceptions: ‘singly linked’ intervals exist between markers DXS6670 and B013LA in Xp11.3, DXS6616 and DXS6950 in Xp11.23 and DXS7124 and DXS1266 in Xp11.3-p11.23. The first two linkages are supported by the physical maps of other groups (Black et al 1995; Kwan et al 1995) while the last is being verified in this laboratory as discussed in section 7.3.3.3.

The order of markers presented in Figure 7.24 generated the minimum number of internal deletions and is thus considered to be the most reliable. The relative positions of several markers could not be resolved because the YAC hits were identical (e.g. DXS722/MG21, UBE1/DXS7124) and there were no YACs with only partial overlap to split the STSs. As the depth of coverage in these regions is substantial (Figure 7.24) this suggests these STSs lie in very close physical proximity, as confirmed by the presence of both sets of STSs within single cosmid clones (this study, data not shown; Coleman et al 1996). The close association of genes in the Xp11.23 region of the contig (18 ESTs in the Xp11.23 contig which spans ~ 2Mb) correlates well with the apparent rule that light-staining Giemsa bands are densely populated with expressed sequences (Bickmore and Sumner 1989).

By summing the estimated lengths of each contig it can be inferred that approximately 4.5-5.0 Mb of the Xp11.3-p11.23 interval have been recovered in cloned DNA by this work. As the region between DXS7 and DXS255 has a genetic distance of ~13cM, this equates to roughly half the region. Of particular relevance to positional cloning strategies to identify the RP2 gene, the newly defined critical interval of ~5cM flanked by markers DXS8083 and DXS6616 (chapter 6) is now extensively represented in overlapping YACs to an extent of
Figure 7.24: Summary chart of YAC contigs generated in Xp11.3-p11.22

<table>
<thead>
<tr>
<th>YACs</th>
<th>Orientation</th>
<th>Overlap</th>
<th>STSs</th>
<th>No. YACs</th>
<th>No. YACs containing STSs</th>
</tr>
</thead>
<tbody>
<tr>
<td>F02 14</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>E01640</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>G0630</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>B013</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>F01125</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>3BEA</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>B0721</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>13BF11</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>9BEA10</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>22CA9</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>40BE7</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>A094 7</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>40HD1</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>12AF3</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>19DNF</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>D1151</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>24BF12</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>71D3</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>20D32</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>11AD10</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>37GA5</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>8DF4</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>C122B</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>C0628</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>A120</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>17A55</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>20DE6</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>C1022</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>44G2</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>32CA11</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>3BEA</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>31GA10</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>23BC7</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>31GA2</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>34A5S</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>30DH10</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>24A11</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>19HE</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>52EH3</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>40F10</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>22B99</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>27CB12</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>22AB9</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>14BD2</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>
~2.5Mb. This serves as an excellent resource from which to initiate targetted walking strategies and bridge the remaining gap between DXS8083 and DXS1264.

The extent of chimaerism or instability within this set of YACs cannot be ascertained at this stage as many of the YACs have not been sufficiently characterised. In some instances, however, where a cluster of YACs were available for end-cloning (such as for *UBE1* and DXS722), this study shows that although tempting to choose the largest YACs to extend the walk furthest into neighbouring regions, such YACs are more likely to be chimaeric or rearranged (e.g. 37GA5, 27GF2), possibly because of increased chances for homologous recombination between repetitive human DNA elements (e.g. ALU, L1) which are present more frequently in larger YACs. It is thus more advisable to walk from smaller YACs whose ends are internal to the ends of the contig (e.g. 9BA10) so that insert-terminal STS location can be verified by presence in overlapping YACs. Two regions of instability were discovered; a region close to *GATA1* and *SYP* in Xp11.23-p11.22, as noted by other investigators (section 7.3.2.1) and around the *UBE1* gene cluster in Xp11.3-p11.23 (section 7.3.3.3). Instability could be due to repetitive DNA with a tendency to delete out of YACs (as has been postulated for the *OATL1* region; Chand et al 1995), or the high GC content characteristic of gene-rich regions may be incompatible with the high AT environment of yeast DNA (Palmieri et al 1994). Despite the fact that 55 (64%) of the YACs comprising the contigs originate from the ICI library with the lowest reported level of chimaerism (Nagaraja et al 1994; Anand et al 1990), and the overall level of redundancy is high, it may therefore be that YAC chimaerism/instability in the *UBE1* region will necessitate supplementation in a different cloning vehicle (section 1.3.2.3) to form a continuous contig spanning the RP2 critical region.

### 7.4 Discussion

When this work was initiated, the RP2 critical region extended between *MAOA* in Xp11.3 and DXS255 in Xp11.22. For this region, estimated to contain ~13Mb of DNA, the following markers were available in STS form: DXS7, *MAOA, MAOB*, DXS742, DXS1055, DXS1003, *UBE1, SYN1, TIMP-1, PFC*, DXS426, *ELK1, ZNF81*, DXS722, DXS1126 and DXS573 (Figure 7.1). Together with STSs generated in this study to published sequences (*SYP, TFE3, GATA1, DXS255, MG21, MG44, MG61, MG81*) these were used as primary tools to isolate YACs for the region and assemble them into preliminary contigs by STS-content mapping. Multiple YAC libraries were used to ensure reasonable redundancy and provide extra opportunities for ordering STSs by obtaining overlapping YACs of different extents. Additional STSs were recruited to enrich the physical map from either YAC endsequences or following their publication by other investigators.

A total of 65 STSs have been used to date, to identify and overlap 84 YAC clones via an STS-based approach to form a series of contigs within the RP2 critical region on proximal Xp. Altogether the YACs are estimated to span approximately 5Mb of the Xp11.3-p11.22 region and ~2.5Mb of the newly defined RP2 critical region flanked by markers DXS8083 and DXS6616. By using a variety of STSs comprising polymorphic genetic markers, random STSs and ESTs and known genes, integration of the genetic, physical and transcriptional maps
of the region has been accomplished. 21 of the STSs were developed during the course of this study from either published sequences (7 STSs) or YAC insert-terminal sequences (14 STSs) and their analysis integrated with previously described markers.

The relative order of STSs on this physical map is in general agreement with other partial contigs generated across the region (as discussed in section 7.3) and further extends these findings; the map has incorporated 5 new microsatellites into the Xp11.3-Xp11.23 physical map (DXS1239, DXS8080, DXS1146, DXS7124, DXS337) and refined the localisations of 3 others (DXS1055, DXS1003, DXS1126). Their physical localisation will render these microsatellites excellent markers for linkage studies of the various diseases involved in this region (sections 1.8.4.2 and 1.9). The order for DXS1003 and DXS1055 generated by our YAC contig suggests that the original order Xpter-DXS1055-DXS1003-Xcen of Weissenbach et al (1992) may have been correct. The reverse order obtained more recently by the MIT Genome Centre (Hudson et al 1995) may be a result of complex rearrangements that are known to exist in the radiation hybrids used in the mapping process. The current WI YAC map suggests no physical links with neighbouring anchored markers from which this order may have been derived.

The discovery by FASTA database searches of 4 identities between YAC end-sequences and recently developed ESTs/STSs demonstrates the importance of regular reanalysis of sequences obtained during contig assembly. As found in this study, this may provide positional information for the queried sequence, or alternatively may localise unmapped ESTs/STSs to the region of interest. An existing STS of previously unspecified X-location has been mapped to the Xp11.23-11.22 interval by identity to a YAC insert-terminal sequence (DXS8193) and by a similar process 2 ESTs have been newly assigned to loci between DXS6950 and DXS6941 in Xp11.23 (DXS9853E) and between DXS1055 and DXS1003 (DXS9886E). Furthermore, the positioning of the newly defined proximal flanking marker for RP2, DXS6616, distal to DXS9853E and 4 retinally expressed genes MG21, MG44, MG61 and MG81 excludes these genes from a causative role in the disease.

The use of STSs or other internal landmarks to determine the colinearity between clones is a very rapid approach and is not hampered over small regions by cloning artefacts that give rise to chimeric clones, but is sensitive to internal deletions in regions of low clone density. Furthermore, STS-content mapping alone may not detect short overlaps, and in such cases it is necessary to resort to other methods to determine continuity. This phenomenon was highlighted in this study in the central portion of the region (section 7.3.3.3) where an unbroken tiling path at present remains elusive despite extensive testing of YACs verses STSs in that region. This may be due in part to false positive and negative YAC/STS results; although results have been confirmed as far as possible, internal deletions within YAC clones still remain. All postulated orders of STSs (within the known framework) in this region results in a number of gaps and deleted YACs which require further investigation. Genetic order for key microsatellite markers derived from meiotic breakpoint mapping (chapter 6) has gone part way to resolving the situation, but further end-cloning to assess which YACs are deleted/chimaeric and create further STSs in the region for confirmation of overlaps is needed.
and is underway. The estimation of the size of this contig will also become more accurate when the position of end-STSs within YACs are known.

ALU-PCR fingerprinting of selected YACs in Xp11.3-p11.23 was used to reinforce links between adjacent loci DXS337/ZNF157/DXS7124 and in addition indicated undetected but suspected overlaps between the DXS1003 and UBE1 YAC clusters. In contrast to analysis using an individual STS, this technique assays multiple specific sites in each clone and can be applied to overlapping clones in the absence of any prior information about the region under study. It was therefore particularly applicable to the Xp11.3-p11.23 region surrounding UBE1 for which discrepant marker orders had been observed (section 7.3.3.3). The potential overlaps detected will require confirmation by other methods e.g. by PCR analysis using STSs derived from YAC insert-terminals or cross-hybridisation of YAC ALU-PCR products.

It is also acknowledged that additional methodologies such as FISH mapping and restriction mapping of key YACs are needed to confirm the integrity of the cloned inserts between STS landmarks and obtain a finer scaffold map of STSs. Restriction mapping also provides a more accurate assessment of the extent of YAC overlaps and hence the physical distance between STSs. By choosing rare-cutting restriction enzymes whose recognition sites lie in CpG islands this procedure may also be a useful indicator of the locations of gene sequences (section 1.4.3). To gain additional confidence that the entire region has been cloned, it will be helpful to align the YAC map with a long-range genomic map of the region. Although preliminary long-range mapping data has been generated by a number of groups for the well-characterised interval in Xp11.23-p11.22 between TIMP-1 and DXS255 (e.g. Schindelhauer et al 1996, Boycott et al 1996, Derry et al 1994) a map of this kind would be potentially more useful in the Xp11.3-p11.23 interval between DXS8083 and TIMP-1, for which the tendency toward instability in YACs will be prohibitive to the determination of precise inter-marker physical distances.

The YAC contigs presented here significantly improve the physical map of the region, especially for Xp11.3-p11.23, which is still under-represented in both the Chumakov et al (1995) and Hudson et al (1995) whole genome physical maps, possibly because of under-representation of the X as a whole in the CEPH mega-YAC library used to construct the maps. The latter STS-based map, is however, updated daily and accessible via the Internet (http://www-genome.wi.mit.edu/); indeed, newly-mapped STSs (many corresponding to ESTs) are already being used to consolidate and extend our map. The increased map resolution may resolve local order for many STSs (such as DXS7124 and DXS1266 in Xp11.23) by linking and thus orienting adjacent contigs.

Several gaps exist within the YAC contig, the most sizeable of which span the 9BA10LA - DXS8083/DXS8026 interval (Xp11.3) and the DXS8083/DXS8026-DXS1264 interval (Xp11.3-p11.23). The latter gap requires closure in order to complete a YAC map of the current RP2 critical region; the genomic distance between DXS8083 and DXS1264 (genetic distance 2cM; Dib et al 1996) is yet to be established but its size can be estimated by comparison with other maps of the region: based on the physical distance between DXS1264 and DXS255 of 2.7Mb (Boycott et al 1996) which corresponds to a genetic distance of ~ 6cM
(Dib et al 1996), and the recently available X chromosome YAC map of Crollius et al (1996) which estimates the physical distance between DXS742 and PFC (see Figure 7.1) to be only 2Mb, the DXS8083-DXS1264 gap is likely to be no more than 1Mb and thus the majority of the region is now in cloned form. Alternative procedures such as high resolution FISH mapping (section 1.3.1.2) or radiation hybrid mapping (section 1.3.1.1) may also be applied to confirm contig alignment and orientation (e.g. using markers DXS1055, DXS1146, DXS7124, ZNF157) by generating a ‘framework’ order for markers >500kb apart.

The major YAC contig spanning the Xp11.3-p11.23 region (Figure 7.24) forms the framework for a detailed physical map of the RP2 critical region and will be useful in its present form for the identification and characterisation of genes that reside in the ~2.5Mb array of overlapping clones so far allocated within it. The ordered set of transcripts generated in this study will be of great benefit to positional cloning projects to identify disease genes in the region by way of the positional candidate approach (section 1.6). YAC STS-content analysis with the growing number of ESTs being mapped to proximal Xp (Hudson et al 1995; Hillier et al 1996) and/or their localisation to the contig by identity to YAC-derived sequences will also undoubtedly provide a rapid means to exclude candidate genes, as well as anchoring and confirming the YAC map. In addition, the isolation of new polymorphic markers from key YACs of defined location may provide new tools for narrowing the RP2 distal boundary in families F71 and F72 (chapter 6), by either haplotype analysis or the generation of linkage disequilibrium data, where a dense set of genetic markers is essential (section 1.2.7). With the application of region-specific gene identification techniques (section 1.4) this set of ordered DNA fragments will provide molecular access to new, previously unidentified genes for the Xp11.3-p11.23 region. As several other inherited eye disorders map to this interval (section 1.9), the YAC contig described here potentially has many applications in ophthalmic genetics; by facilitating the analysis of expressed sequences mapping therein, it should either reinforce the hypothesis that the RP2 gene is allelic to one form of CSNBX/AIED, or lead to the isolation of other genes within this region that have a role in retinal physiology.
CHAPTER 8

Overall Discussion and Future Prospects

The plethora of human genetic eye disorders is a direct consequence of the biochemical and physiological complexity of the visual process and the fact that mutations in the governing genes will generally not hinder the viability of individuals who carry them. A major component are the retinal degenerations, for which an ever-increasing number of genetic loci are becoming evident (section 1.7). Identifying the repertoire of genes responsible for these diseases and elucidating their role in both normal and defective vision is a current challenge for both biochemists and geneticists alike.

Several of the genes underlying RP have been identified in recent years primarily by way of the candidate gene approach (section 1.1). The candidate gene approach has now been extended however, since many genes specific to the phototransduction pathway have already been identified, mapped and studied in diseased families. It is interesting to observe that mutations in genes with such a specific function have only resolved a small proportion of disease phenotypes. Many other proteins that are critical to normal photoreceptor development and function but not necessarily exclusively or abundantly expressed in the retina have become strong contenders for the disease genes yet to be identified. Furthermore, although RP is thought to be primarily a disorder of the rod (and to a lesser extent cone) photoreceptors, any of a large number of biochemical abnormalities could potentially cause a degenerative change associated with rod-cone dysfunction and pigmentary retinopathy, and the primary defect could lie in any of the interacting cellular elements of the neural retina and RPE or even in a tissue remote from the eye.

In the absence of suitable candidate genes and disease-associated cytogenetic abnormalities, a positional cloning strategy has been adopted to identify the gene underlying RP2. At the outset of this project, RP2 remained broadly localised to a ~13cM region of proximal Xp flanked by markers DXS7 (Xp11.3) and DXS255 (Xp11.22) (section 1.9.5.3.2), owing to a lack of informative recombination events and a paucity of suitably placed polymorphic markers. To generate additional polymorphic markers for the RP2 critical region, 12 X chromosome-specific cosmids known to contain (CA)₇ microsatellites were physically mapped to X chromosome subregions by metaphase FISH. The corresponding microsatellites were subsequently isolated and characterised from key cosmids that mapped to Xp11.4-Xp11.22 and investigated for polymorphism. This thesis describes this procedure for cosmid HX20 that mapped to Xp11.4, from which a highly polymorphic microsatellite (DXS556) was isolated (Thiselton et al 1993).

A combination of genetic linkage analysis, haplotype analysis and YAC STS-content mapping was then used to refine the localisations of 5 polymorphic microsatellites mapping to
the XLRP critical interval and incorporate them into the genetic map of proximal Xp (Thiselton et al 1995). In addition to increasing the regional marker density for XLRP gene localisation in this study, these microsatellites have become useful genetic landmarks for other researchers mapping disease genes on proximal Xp; DXS556 in particular has been instrumental in refining the localisation of COD1 (section 1.9.4). They have also provided important STSs to integrate the genetic map with the developing physical map of the region, and all but one (DXS574) are now anchored in YAC contigs (Nelson et al 1995).

These and other microsatellites mapped to the region from whole-genome mapping efforts were subsequently used in this study to generate extensive haplotypes for members of 14 XLRP families in an effort to determine the locus segregating within each family and define the localisation of the RP2 gene. Key recombination events enabled classification of one family as RP3 and four as RP2. The detection of 2 critical recombination events in males of known clinical status defined both the distal and proximal boundaries of the RP2 critical interval in a single family. This interval was thus substantially refined from ~13cM to ~5cM between DXS8083 (Xp11.3-p11.23) and DXS6941 (Xp11.23). Although one family (F47) has been defined as RP3 independently by discovery of a mutation in RPGR (section 6.3.1), further genetic studies are necessary to establish the segregating XLRP locus in the remaining 8 families. This may add to the number of families classified as RP6/RP15, or substantiate reports of still further loci for this disorder located on the X chromosome (section 1.9.5.1). Screening for RPGR mutations in the RP3 family F11 or those designated RP2/RP3 may either implicate this gene as causative of the disease, or support the suggestion of microheterogeneity within the RP3 locus (section 1.9.5.3.1).

Speculation already exists for phenotype/genotype correlation with XLRP, specifically associating a tapetal reflex with the RP3 locus (section 1.9.5.2). Eventually it may be possible to predict the genotype of a patient simply by electrodiagnostics and fundoscopy in the clinics. However, the exact number and position of XLRP loci is still under investigation (section 6.4).

In parallel with the genetic studies, a high resolution YAC contig has been constructed in Xp11.3-p11.22 by an STS-content mapping approach and comprises 84 YACs formatted with 65 STSs, 21 of which are novel to this thesis, including 14 YAC-end STSs and 7 STSs derived from published sequences. The map to date is centered around two contigs: a ~1.5Mb contig encompassing DXS7-MAO in Xp11.3 and a ~3Mb contig spanning the newly defined proximal RP2 boundary and extending distally within the RP2 critical interval. The high redundancy of the map has enabled the physical linkage and ordering of many STS, genes and ESTs in Xp11.3-p11.23 to a density of 1 per 33kb, which constitutes an important contribution to the transcriptional map of proximal Xp. The use of a 'meiotic breakpoint mapping panel' composed of key recombinants from the XLRP families, provided supplementary order information for key microsatellites to facilitate contig orientation in Xp11.23. Integration of the genetic and physical data has provided precise locus assignments for 6 microsatellites of prior broad localisation to the region, which will be a major aid to genetic mapping of XLRP, CSNBX and other disorders mapping to this interval (section 6.4).
Furthermore, as most of the YACs used to construct the contigs described here are novel to the region, they provide a useful template for testing additional STSs and so contribute to the fidelity of the overall regional map. One of the critical functions of this physical map will be its use as a tool for positioning potential RP2 candidate ESTs and genes which now saturate the databases.

STSs derived from YAC insert-terminals in the work described here proved highly informative for two reasons: chimaeric YACs could usually be detected by analysis of only three X-specific somatic cell hybrids and overlapping YACs, and sequence analysis demonstrated identities of three STSs to gene sequences in Xp11.23, SYN1 and 2 ESTs. Although the relative location of SYN1 was known, the depth of the contig immediately gave a precise physical map location to the 2 ESTs. Integration of the physical and genetic maps at this point was able to exclude one of these from the RP2 critical region. The power of sequence analysis cannot be underestimated, with the ever-increasing number of ESTs being deposited in Genbank and dbEST, database identity searches with YAC-end STSs can be a simple and valuable way of physically mapping novel genes with high resolution to a disease gene region. Furthermore, owing to the redundancy of most ESTs in the databases, comparison of the ‘highest hits’ can often reveal an expression profile of the corresponding gene which may prioritise further work for those that map within the region of interest (e.g. DXS9853E in section 7.3.2.1). Developments in molecular genetics have recently led to an influx of vast amounts of data from centres worldwide as part of the Human Genome Mapping Project (section 1.5). As a result, groups interested in particular disease regions must constantly evaluate these databases, of largely non-overlapping sets of data, and attempt to integrate them with their own physical and genetic maps. In this way the fortuitous fine mapping of an EST as described above, may provide you with your disease gene, forgoing the necessity for implementing complex techniques such as cDNA selection.

Consideration of the possible functions of the RP2 gene may highlight potential candidates from the vast array now mapped within the RP2 critical interval and in this respect the functions of genes involved in other retinal degenerations can provide important clues. In addition to the aforementioned photoreceptor-specific components, an increasing body of evidence is accumulating to implicate disturbances in protein processing in retinal degeneration. For example, both the putative RP3 gene (RPGR), and the choroideremia gene (a Rab escort protein) encode proteins involved in the regulation of GTPase activity and hence cellular signalling (section 1.9.5.3.1). TIMP-3, the gene responsible for Sorsby’s fundus dystrophy, encodes a tissue inhibitor of metalloproteinases which is important in the processing of extracellular matrix proteins (section 6.1.2). Although these genes are expressed in many tissues, mutations appear to result solely in an ocular phenotype, which suggests that photoreceptors, with their high levels of protein synthesis and trafficking, are particularly sensitive to disturbances in protein turnover and processing. Several genes located within the RP2 critical region are strong ‘positional candidates’ in this regard. The TIMP-1 gene has therefore been screened in this study for mutations in affected males of three RP2 families by
PCR and direct sequencing of all coding sequence, splice junctions and 5'UTR promotor region. Although a neutral polymorphism was discovered in exon 5, no disease-associated sequence alterations were found, suggesting that TIMP-1 is unlikely to play a causal role in the etiology of XLRP. Given the importance of correct protein processing to retinal health and the role of the ubiquitin system in protein modification and targetting (sections 1.8.4.1 and 1.8.4.2), the genes UBE1, UHX1 and PCTK1 in Xp11.23 are also worthy of investigation. Such investigations are now underway for UHX1, a ubiquitin C-terminal hydrolase, as its high level of retinal expression and similarity to the Drosophila gene fat facets, the gene product of which influences the fate of cells destined to become photoreceptors in the developing compound eye (Fischer-Vize et al 1992), and map location to Xp11.23 established in this thesis, make it a strong positional candidate for RP2.

The extensive knowledge of gene functions in other model organisms may also suggest likely candidates for retinal disorders. Banfi et al (1996) have taken advantage of the extensive database facilities now available covering the human and fruit fly genomes. By searching a subset of the human EST database with Drosophila genes, they have identified a number of human genes significantly related to a Drosophila counterpart which they call Drosophila related expressed sequences (DRES). Sub-chromosomal localisation of these DRES genes by FISH and RH mapping has provided positional candidates for many genetic diseases. One of these, DRES14, is a fat facets human homologue and is another member of the family of UBE1 C-terminal hydrolases (Jones et al 1996). The map location for DRES14 in Xp11.4 excludes it from a role in RP2 but makes it a prime positional candidate for COD1 (section 1.9.4).

The relative order of markers in the distal portion of the newly-defined RP2 interval between DXS8083 and DXS1003 requires further investigation as YAC STS-content mapping, despite the overall clone density, has not resolved this issue owing to regional instability. Increasing the local marker density which is currently poor, may contribute to establishing the most likely order firstly by incorporating recently published STSs and ESTs (Hudson et al 1995; http://www-genome.mit.wi.edu/) followed by the isolation of novel YAC-insert STSs. Those mapping to the distal ends of the Xp11.23 contig and the extremities of the DXS8083 contig can also be used to screen the libraries for overlapping YACs in an effort to bridge the DXS8083-DXS1264 gap. Any resulting non-chimaeric YACs can be used as templates from which to isolate novel polymorphic markers to refine the distal boundary of the RP2 interval in family F72. Although microsatellites can be isolated from YAC clones via methods based on analysis of YAC-derived PCR products (e.g. ALU-PCR; Sinnett et al 1990, linker-based methods; Brown et al 1995, Chen et al 1995), bi-allelic single nucleotide polymorphisms will also be sought as the microsatellite density in Xp11.23 reaches its limit. These have been found to occur about every 1kb and are forming the basis of a ‘third generation genetic map of the human genome (Wang et al 1996). A complementary approach for orientating and ordering markers within this unchartered interval would be to build up the publicly available radiation hybrid maps which currently do not represent the region of interest. Also a large spectrum of FISH technologies with different levels of sensitivity, resolution and accuracy,
combined with speed and easy interpretation, now offer an alternative system that can be implemented in positional cloning projects (section 1.3.1.2). The feasibility of this approach has been exemplified by Laan et al (1996) who on the basis of FISH techniques in conjunction with genetic mapping data refined the initial 4cM critical region for neuronal ceroid lipofuscinosisis (CLN5) to a high resolution map of only 400kb in length.

In the event that further YACs turn out to contain confounding deletions or chimaerism, a long-range restriction map of the region will be concurrently established to verify marker order and also establish physical distances between them. Use of rare-cutter restriction enzymes will also pinpoint the locations of CpG islands in the region as potential coding sequences; comparison with the YAC map will then highlight clones from which to isolate the corresponding gene(s). The YACs will also provide a resource for additional methods to locate transcripts within them such as BLOCK PCR (d’Esposito et al 1994) which utilises primers designed to amplify ‘ancient conserved regions’ or motifs of several gene families (e.g. zinc-finger transcription factors) or screening with a consensus oligonucleotide probe for splice site sequences. cDNA selection (section 1.4.4) is already underway in our laboratory using YACs reported in this thesis as it appears to be the most successful method for isolating novel genes (Harshman et al 1995). However, it will only identify cDNAs that are present in the source library and therefore other strategies will be employed in parallel such as those already described to enhance the overall efficiency. It is envisaged that with the YAC resources available so far it will be possible to build up a high resolution transcriptional map of the region for constant integration with the genetic map to exclude candidate genes, as has already been demonstrated in this thesis. Thus the genetic analysis of RP2 families and physical mapping of the interval should proceed hand in hand; in such a gene-rich region the exclusion of candidate genes is essential due to the amount of effort required to establish the genomic structure of a gene and screen for mutations.

In addition to mapping new genetic markers to the DXS8083-DXS6616 interval in Xp11.23, further refinement of the RP2 critical region by genetic means will require the identification of more recombination events within the critical interval. In view of the relative rarity of RP2 pedigrees and the likely allelism between XLRP and other retinal degenerations with overlapping map locations such as CSNBX (sections 1.9.2), further genetic analysis will be extended to families with these disorders. Any refinement of these genetic intervals may then guide the search for RP2 in this gene rich region. The complexity of the situation is compounded by the fact that the RP2 gene may not be expressed in the retina or only at very low levels (akin to the RP3 gene RPGR; section 1.9.5.3.1) consequently, a wider spectrum of transcripts must now be evaluated. The recent genetic localisation of a CSNBX gene between OTC (Xp11.4) and DXS1003 (Xp11.23) could potentially be of assistance in this regard. The distal boundary for this locus clearly excludes allelism with RP3 (which maps distal to OTC) and highlights the possibility that the gene segregating in this family is potentially allelic with RP2. The functional relationship between CSNBX and RP has been demonstrated by genetic allelism of both X-linked and autosomal forms of these diseases (sections 1.7; 1.9.5.3.1). If this is the case, then the search becomes directed to the interval between DXS8083 and
DXS1003, combining the locations of both genes. Once a definitive order of markers in this interval is achieved (as outlined in section 7.4) this may affect the RP2 candidacy of the UBE1 gene cluster.

Other strategies may also establish the most likely location of the RP2 gene. Microdeletions are not evident in the families studied here on the basis of microsatellite analysis, but may be detected upon Southern analysis of affected and unaffected individuals with genes or other sequences that map to the region, or by more extensive STS PCR analysis of patient DNA. The possibility of a founder effect for RP2 in the European population (section 1.9.5.3) may be investigated by linkage disequilibrium studies (section 1.2.7), although the success of this technique depends on a larger set of affected and unaffected chromosomes than are at present available in our laboratory. Furthermore, preliminary analysis of marker loci closely linked to RP2, based on the haplotypes generated in this study, has shown different alleles segregating with the disease phenotype, thus making a close relationship between the various RP2 families unlikely (data not shown).

The mouse is often utilised as a model organism for a diverse range of human genetic diseases. The human and mouse genomes contain ~150 conserved segments with nearly identical gene content (Meisler 1996). Conserved linkage relationships make it possible to use gene identification and map position in one species to predict location in the other and to recognise true homologies between mouse mutations and human disease. The human-mouse comparative map has in this way facilitated the positional cloning of several human disease genes, e.g. the the gene responsible for Usher syndrome 1B, the most frequent cause of deaf-blindness in humans, was identified on the basis of homology to the mouse shaker-1 deafness gene (myosin VIIA), that mapped to a conserved linkage group in the mouse genome (Weil et al 1995). Since no known mouse model exists for XLRP a suitable application in the future may be to use gene targeting for gene replacement or knock out studies (Capecchi 1994) since gene function is not always inferred from primary sequence database identities. This technology allows the evaluation of the effect on the organism of absent gene function or presence of dysfunctional product. Although the physiological abnormalities in the mouse are frequently not identical to those in human patients, these mouse models can be extremely valuable for testing disease interventions (Searle et al 1994). If the mutation responsible for RP2 turns out to be of retina-specific expression, gene therapeutic strategies may eventually compensate at least partially for the defect (Ali et al 1996). The wide phenotypic variability seen both between and within families suffering from most inherited retinal degenerations indicates that the individual's own genetic background may well influence the activity or expression of the disease-causing gene. Alternatively, environmental factors could have a significant effect on the expression of mutated genes. Dietary intake of large doses of vitamin A has indeed been shown to improve night vision in patients with Sorsby's Fundus Dystrophy (Jacobsen et al 1995) and identification of epigenetic components which might be involved in the suppression of an RP phenotype is a major and exciting challenge.
REFERENCES


285


LETTER

Mapping the RP2 Locus for X-linked Retinitis Pigmentosa on Proximal Xp: A Genetically Defined 5-cM Critical Region and Exclusion of Candidate Genes by Physical Mapping

Dawn L. Thiselton, R. Mark Hampson, Manimekelei Nayudu, Lionel Van Maldergem, Mitchel L. Wolf, Bratin K. Saha, Shomi S. Bhattacharya, and Alison J. Hardcastle

1Department of Molecular Genetics, Institute of Ophthalmology, London, UK; 2Rayne Institute, St. Thomas' Hospital, London, UK; 3Institut de Pathologie et de Génétique A.S.B.L., B-6280, Gerpinnes (Loverval), Belgium; 4Jewish Hospital, St. Louis, Missouri 63108; 5Winship Cancer Centre, Emory University School of Medicine, Atlanta, Georgia 30322

Genetic linkage studies have implicated at least two loci for X-linked retinitis pigmentosa (XLRP) on proximal Xp. We now report a defined genetic localization for the RP2 locus to a 5-cM interval in Xpl1.3–ll.23. Haplotype analysis of polymorphic markers in recombinant individuals from two XLRP families has enabled us to identify DXS8083 and DXS6616 as the new distal and proximal flanking markers for RP2. Using STS-content and YAC end-clone mapping, an ~1.2 Mb YAC contig has been established encompassing the proximal RP2 boundary and extending from TIMP1 to DXS1240 in Xpl2.23. Several ESTs have been positioned and ordered on this contig, one of which is novel to the region, identified by sequence database match to a physically mapped YAC insert terminal STS. Integration of the genetic and physical data has placed four retinally expressed genes proximal to DXS6616, and thereby excluded them from a causative role in RP2. This work now provides a much needed focus for positional cloning approaches to isolation of the defective gene.

Retinitis pigmentosa (RP) is a group of hereditary progressive retinal degenerations characterized by night blindness, visual field impairment, and degenerative pigmentary changes in the retina. RP exists as autosomal dominant, autosomal recessive, and X-linked forms and displays considerable genetic heterogeneity with at least 15 distinct loci so far assigned to human chromosomes (for review, see Dryja et al. 1995). X-linked retinitis pigmentosa (XLRP) is the most severe clinical form, accounting for 7–30% of all cases, depending on the population studied, with an incidence of ~1:20,000 (Jay 1982; Heckenlively 1983). Male XLRP patients generally develop concentric visual field loss before the 20th year of life leading to severe visual handicap by the age of 40 (Bird 1975). Female carriers show variable symptoms of the disease on ophthalmological testing, with visual impairment usually beginning in middle age, although absence of ocular abnormalities does not exclude the carrier state (Arden et al. 1983).

In the absence of functional clues as to the pathophysiology of XLRP, positional cloning strategies have been adopted to isolate the defective genes. Following the first genetic linkage of an RP gene (designated RP2) to Xp11.3 in a panel of British families (Bhattacharya et al. 1984), subsequent genetic analyses have indicated the existence of at least three other XLRP loci (RP3, RP6, and RP15) located more distally on Xp (Musarella et al. 1990; Ott et al. 1990; Teague et al. 1994; McGuire et al. 1995) and the fact that the disease in some families maps to none of these locations suggests the possibility of even more XLRP loci (Aldred et al. 1994). As evidence for RP6 is to date only statistical (Ott et al. 1990), and RP15 has been demonstrated in only one family (which is

*Corresponding author.
E-MAIL ahardcas@hgmp.mrc.ac.uk; FAX 44-171-608-6863.
reported as a cone-rod degeneration; McGuire et al. 1995), the majority of XLRP families fall into the categories of RP2 or RP3.

The ability to distinguish between RP2 and RP3 is dependent on the detection of crossovers dissecting the target region by genetic linkage/haplotype analysis, attributable to the lack of reliable clinical differences between the two disease entities (Wright et al. 1991). The clustering of XLRP genes on proximal Xp makes such genetic distinction of XLRP families difficult; however, from cumulative genetic data it appears that RP3 predominates in British and American families (Musarella et al. 1990; Ott et al. 1990; Teague et al. 1994).

Precise localization of the RP3 gene to Xp21.1 by genetic linkage analysis has been augmented by the molecular genetic analysis of RP3 patients with submicroscopic DNA deletions, confining RP3 to a 530-kb stretch of DNA flanked by genetic markers OTCl and DXS1110 (Roux et al. 1994). The focusing of efforts afforded by the identification of M.O., a male RP patient harboring a 75-kb deletion within the RP3 critical region, has recently led to the isolation of a gene (RPGR) in which mutations account for XLRP in a proportion of RP3 patients (Meindl et al. 1996).

The RP2 gene has remained broadly localized to an ~13-cm interval in Xp11.22-11.3 flanked by DXS7 and DXS255, owing to a lack of recombination events in the critical region (Friedrich et al. 1992; Wright et al. 1991) and no detectable disease-associated deletions. Various multipoint and heterogeneity analyses further suggest different locations for the RP2 gene within this interval (Xp11.23; Teague et al. 1994; Xp11.22; Bergen et al. 1995). Narrowing of the RP2 region will improve reliability of carrier detection and may facilitate characterization of the gene and its mutations and help resolve the issue of clinical and genetic heterogeneity.

We present here two XLRP families in which the gene responsible is consistent with an RP2 location, and where key recombinants have been identified that define both the proximal and distal boundaries of the RP2 critical interval within each family. Integration of the genetic and physical maps of the region has led to the exclusion of several retinally expressed candidate genes located in Xp11.23, and the degree of refinement now renders the construction of a physical contig spanning the RP2 critical region, and isolation of candidate transcripts therein, a more manageable task.

RESULTS

Haplotype Analysis

Key recombination events within families NRP and F72 are shown in Figure 1. No deletions were detected using the 19 microsatellites listed.

Family NRP

In family NRP (Fig. 1a) individual V-1, an affected male, is a recombinant with respect to his carrier mother IV-1 between markers MAOB and DXS1055. The phase of the maternal alleles was established firmly from analysis of two obligate carrier sibs (IV-2 and IV-3) who share the mother's affected haplotype. This crossover defines the proximal boundary of the XLRP interval, with the disease-associated haplotype located proximal to MAOB, providing firm evidence for RP2 segregating in family NRP. A similar crossover is seen in individual III-4, a carrier female who has inherited her mother's affected chromosome proximal to MAOB and passed this recombinant chromosome onto her affected son IV-7. The phase of the alleles in II-2 could be clearly deduced from the haplotypes of her affected sons III-3 and III-6, assuming crossover minimization.

The proximal boundary of the RP2 interval is defined by an inferred crossover in individual IV-4 between markers DXS1055 and DXS6616, which has been transmitted to her affected son V-3. It is assumed that this recombination event occurred in individual III-2, as individual IV-5 has the same affected haplotype as her maternal uncle III-1 (as seen in his daughters IV-1, IV-2, and IV-3). The disease-associated haplotype in individuals IV-4 and V-3 is therefore located distal to DXS6616. A crossover event in individual III-5 provides additional support for the location of the XLRP gene in this family. This affected male is a recombinant with respect to his carrier mother II-2 between markers DXS6616 and DXS6941, with the disease-associated haplotype located distal to DXS6941.

In summary, key recombination events in family NRP are consistent with RP2 segregation, and indicate an RP2 critical region flanked by MAOB (Xp11.3) and DXS6616 (Xp11.23).

Family F72

In family F72 (Fig. 1b) individual II-5 is a recombinant with respect to his carrier mother I-1 between markers DXS8083 and DXS1003. The phase of the maternal alleles was deduced from
Figure 1  Pedigrees of the XLRP families used in this study, showing haplotypes constructed with the markers listed. Solid bars indicate those alleles that are linked to the XLRP mutation (i.e., disease-associated haplotype). In the case of recombinant individuals (marked by an X) the solid bar is used to depict only those alleles that can be unambiguously linked to the parental "affected" haplotype. Hatched lines represent markers that were uninformative. (A) Family NRP, with recombinant individuals IV-7, V-1, and V-3 localizing the \( \text{RP2} \) gene to the region between markers MAOB and DXS6616. (B) Family F72, with recombinant individuals II-5 and III-1 positioning the \( \text{RP2} \) gene between markers DXS8083 and DXS6941.

The haplotypes of her phase-known unaffected and obligate carrier daughters. The disease-associated haplotype in affected male II-5 is located proximal to DXS8083, confirming the segregation of \( \text{RP2} \) in this family, and providing a new distal flanking marker for the \( \text{RP2} \) critical region.

A recombination event in individual III-1, an unaffected male, defines the proximal boundary of the \( \text{RP2} \) interval in family F72. A crossover has occurred between markers DXS6616 and DXS6941, positioning the disease-containing interval distal to DXS6941. The phase of the maternal alleles was firmly established from analysis of the grandparental genotypes.

The crossover data clearly defines F72 as an \( \text{RP2} \) family, with the disease gene critical interval flanked distally by DXS8083 (Xp11.3-11.23) and proximally by DXS6941 (Xp11.23).

YAC Contig Construction; Generation and Physical Ordering of Sequence-tagged Sites in Xp11.23

Sequence-tagged site (STS)-content mapping of yeast artificial chromosomes (YACs) identified
using markers from Xp11.23 established a contig spanning ~1.2 Mb (based on the additive size of minimal tiling path YACs) from TIMPI to DXS1240 (Fig. 2). YACs were initially isolated from the library by PCR screening using markers DXS426, ZNF81, MG61, DXS722, and GATA1. STSs were derived from the left and/or right ends of the human DNA inserts in several YACs (detailed in Table 1) and proved instrumental in confirming YAC overlaps and integrity. Of the 10 total YAC insert terminals isolated, two proved to derive from regions other than Xp11.23 on somatic cell hybrid analysis; 34AC5LA maps to distal Xp and CO1160LA maps to chromosome 6, indicating that these YACs are chimeric. FASTA data-base identity searches detected matches for two YAC insert terminals with gene sequences: 4HG2LA lies within exon 2 of the SYN1 gene, anchoring this end of the YAC firmly in Xp11.23, whereas 34AC5RA showed 99.5% identity over 210 bp to a human infant brain cDNA [Integrated Molecular Analysis of Genomes and their Expression (IMAGE) Consortium; expressed sequence tag (EST) accession no. H09726] or 99% identity over 209 bp to a human placental cDNA (IMAGE Consortium; EST accession no. N41839) for forward and reverse strands, respectively. This result suggests that these cDNAs are identical, and maps a new EST (expressed in brain and placenta) to the physical map of Xp11.23 (Fig. 2).

The contig comprises 18 YACs and encompasses 30 markers including 11 microsatellites, 10 gene/ESTs and eight new STSs derived from YAC insert terminals, to give a STS density of ~1 every 40 kb. Key markers (e.g., F0701LA, DXS6849, and DXS6950) and two YACs, ICFy900C1022 (SYN1) and ICRFy900C01160 (2bc6), have been included to allow integration with other partial YAC contigs already existing for Xp11.23 (Coleman et al. 1994; Hagemann et al. 1994; Knight et al. 1994; Fisher et al. 1995; Kwan et al. 1995), which reflect the interest in this gene-rich region of Xp, to which many genetic diseases have been mapped (Nelson et al. 1995). The depth of the contig permitted physical ordering of most STSs/ESTs assigned to the contig, as shown in Figure 2. The order so derived is as follows: Xpter-SYN1(CA)n-TIMPI-C1022RA-4HG2RA-PFC-(CA)n-33CA11LA-(DXS426,F0701LA)-4HG2LA-(DXS1367,E1K1)-33CA11RA-(ZNF81,DXS6849,DXS1004E)-DXS6616-DXS6950-34AC5RA-30DH10RA-(MG81,DXS5941)-MG81-(DXS722,MG21)-(DXS1011E,MG44)-C01160RA-GATA1-DXS1126-DXS1240-Xcen. Interestingly, our lo-

**Figure 2** YAC contig encompassing the RP2 proximal flanking marker, DXS6616, and adjacent markers in Xp11.23 (not to scale). The horizontal line with bidirectional arrows represents the chromosome, with STSs above in the order determined by YAC content. Where two STSs proved inseparable, they are assigned to a single position. The approximate position of the OATL1 pseudogene cluster is indicated by the broken horizontal line (Nelson et al. 1995). The bold horizontal lines represent YACs with sizes given where known. All YACs are derived from the ICI YAC library (Anand et al. 1990) except ICRFy900C1022 and ICRFy900C01160 (Larin et al. 1991). A filled rectangle at the end of a YAC indicates a terminal sequence STS that maps to overlapping YACs in the contig. A hatched box at the end of a YAC depicts a terminal sequence STS that does not derive from this region of Xp11.23. The asterisk denotes the map position of a new EST in Xp11.23, corresponding to 34AC5RA (see text). Vertical broken lines indicate a terminal sequence STS that maps to overlapping YACs in the contig. The shaded bar denotes the region of the contig contained within the newly defined RP2 critical interval. The new proximal flanking marker for RP2, DXS6616, is underscored, and the retinally expressed cDNAs thereby excluded from a role in the disease are in boldface.
## Table 1. Details of Gene-based and YAC Insert STSs Derived in This Study

<table>
<thead>
<tr>
<th>STS name</th>
<th>Source</th>
<th>PCR primers (5’-3’)</th>
<th>(T_a) (°C)</th>
<th>Product size (bp)</th>
<th>Accession no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>GTA1</td>
<td>Gene 3’UTR</td>
<td>ACAGAGCATGGCCTCCAGAGAGCTTGGAAGGTTCAGGCCAGG</td>
<td>62</td>
<td>112</td>
<td>DXS9855E</td>
</tr>
<tr>
<td>MG21</td>
<td>Gene 3’UTR</td>
<td>ATTATGTCGGCGGGAACGCTCTGCTGCTT</td>
<td>62</td>
<td>323</td>
<td>DXS9867E</td>
</tr>
<tr>
<td>MG44</td>
<td>Gene 3’UTR</td>
<td>TGAGGCTGTGTGGGCTGCTGCTTA</td>
<td>62</td>
<td>494</td>
<td>DXS9868E</td>
</tr>
<tr>
<td>MG61</td>
<td>Gene 3’UTR</td>
<td>ATCTGTGACACCTCATAACCTCCTTGGATCCTGCTGCTT</td>
<td>62</td>
<td>252</td>
<td>DXS9869E</td>
</tr>
<tr>
<td>MG81</td>
<td>Gene 3’UTR</td>
<td>ATCTCCATTTTGCCATGAGGAAGCAGACGTCCACATACTT AAC</td>
<td>62</td>
<td>123</td>
<td>DXS9870E</td>
</tr>
<tr>
<td>C1022RA</td>
<td>YAC end</td>
<td>ATTAATCCGACGCTGCTGCTT</td>
<td>64</td>
<td>125</td>
<td>DXS9875</td>
</tr>
<tr>
<td>4H2G2LA</td>
<td>YAC end</td>
<td>CTGGAAAGGAACCAAGCAACGCTTGGGCTGCTGCTT</td>
<td>57</td>
<td>162</td>
<td>DXS9872</td>
</tr>
<tr>
<td>4H2G2RA</td>
<td>(SYN1 exon2)</td>
<td>TGCCACGTGATTTGCGAGATG</td>
<td>63</td>
<td>125</td>
<td>DXS9873E</td>
</tr>
<tr>
<td>33CA11LA</td>
<td>YAC end</td>
<td>ACTACGGAATTTCCACCTTACTTACTTGGTCTTGGTCTT</td>
<td>45</td>
<td>111</td>
<td>DXS9876</td>
</tr>
<tr>
<td>33CA11RA</td>
<td>YAC end</td>
<td>GATTTCCAGCTGAGAAATATGCAAGGCTGGTTGCTAAAAATACCT</td>
<td>56</td>
<td>134</td>
<td>DXS9877</td>
</tr>
<tr>
<td>34AC5RA</td>
<td>YAC end</td>
<td>TACGTGCTAGTTGTTGACGGTCTT</td>
<td>64</td>
<td>177</td>
<td>DXS9853E</td>
</tr>
<tr>
<td>30DH10RA</td>
<td>YAC end</td>
<td>AGGTATACAGCGACACACCAGGCTGGTTGCTGCGATGAC</td>
<td>64</td>
<td>129</td>
<td>DXS9854</td>
</tr>
<tr>
<td>C01160RA</td>
<td>YAC end</td>
<td>ACAGTAGGAGGAATCTTGGTGCCATGAGGCTGCTGCTT</td>
<td>58</td>
<td>172</td>
<td>DXS9874</td>
</tr>
</tbody>
</table>

Calibration of DXS1004E (ZNF41) differs from that in published reports (Knight et al. 1994; Carrel et al. 1996). In contrast to the report of Knight et al. (1994), repeated testing in our laboratory has confirmed YAC ICRFy900C1022 to be positive for DXS1004E. The additional presence of DXS1004E in YACs 3EB3, 24GH5 and 31AG10 which overlap C1022, positions ZNF41 proximal to ELK1, and distal to DXS6616.

The physical ordering and orientation of the retinal cDNAs MG21, MG44, MG61, and MG81 on our contig is in agreement with that depicted on the recently published physical map of Boycott et al. (1996). Although orders could not be established between all pairs of markers, our physical mapping data demonstrates clearly that DXS6616, the new proximal flanking marker for RP2, lies distal to retinally expressed genes MG21, MG44, MG61, and MG81.

### DISCUSSION

XLRP is a progressive degenerative disease of the retina that is clinically and genetically heterogeneous. In this study we performed haplotype analysis in two XLRP families with 20 polymorphic microsatellite markers spanning the critical region for the most prevalent forms RP3 and RP2 (Xp21.1–Xp11.22). Analysis of newly positioned markers in recombinant individuals from these families has enabled us to define new proximal and distal boundaries for the RP2 gene critical...
interval and reduce significantly its size. The new flanking markers in family NRP are MAOB (Xp11.3) and DXS6616 (Xp11.23), whereas those in family F72 are DXS8083 (Xp11.3–11.23) and DXS6941 (Xp11.23). Taken together, assuming the defective gene is the same in both families, this data suggests a refined RP2 interval flanked by DXS8083 and DXS6616 on proximal Xp, spanning ~4–5 cm. Positional cloning efforts can now be directed at this greatly reduced interval, facilitating the search for an RP2 candidate gene. In the event that these two families are segregating different loci, microheterogeneity within this interval may only become evident once the disease gene for one of these families has been cloned.

Genetic heterogeneity is evident in many inherited retinal degenerations, reflecting the eye’s limited repertoire of responses to a variety of genetic lesions. Another emerging pattern in ophthalmic genetics is that of “gene sharing” (allelism) in which different mutations within the same gene can cause clinically distinct ocular diseases. Two forms of autosomal congenital stationary night blindness (CSNB) have been shown to be allelic to RP: Both CSNB and RP can result from mutations in (1) the rhodopsin gene (Dryja et al. 1993; Rao et al. 1994); and (2) the gene encoding the β-subunit of the rod cGMP phosphodiesterase (Gal et al. 1994). The refinement of the RP2 critical interval described here may have significant implications for the localization of other inherited X-linked retinal disorders that have overlapping map locations in Xp11.3–Xp11.22 and which may be allelic to RP2 that is, CSNBX (Aldred et al. 1992; Bech-Hansen et al. 1992; Berger et al. 1995) and X-linked progressive cone dystrophy (Hong et al. 1994; Meire et al. 1994).

The identification of DXS6616 as the new proximal flanking marker for RP2 has also enabled us to exclude several genes as potential candidates for this disorder. Four retinaally expressed genes have been mapped to the OATL1 region in Xp11.23 by direct selection using an OATL1 YAC to screen a retinal cDNA library (Geraghty et al. 1993). We have positioned and ordered these genes with respect to DXS6616 on a 1.2-Mb YAC contig spanning the RP2 proximal boundary and show that DXS6616 lies distal to this cluster of retinaally expressed genes, thereby excluding them from involvement in RP2. With an STS density of ~1 every 40 kb, including seven novel STSs and a newly mapped EST, this contig significantly adds to the available maps of Xp11.23. Furthermore, the contig presented here is comprised primarily of YACs from the ICI 4X library (Anand et al. 1990), 15 of which have not previously been reported, and therefore provides a useful, alternative resource for more detailed analysis of this region of Xp11.23, well noted for its region-specific instability in YACs (Chand et al. 1995; Fisher et al. 1995; Boycott et al. 1996).

The refined RP2 interval is known to be extremely gene-rich, consistent with a cytogenetic Giemsa light band. Genes known to lie in this interval include ZNF21, ZNF41, ZNF81, ELK1, and the (PFC–SYN1–TIMP1–ARA1) gene cluster in Xp11.23 (Nelson et al. 1995) and several CpG islands have been identified that correspond to as yet unknown genes (Coleman et al. 1994). None of the known genes would appear to be strong candidates for RP2 on considering the etiology of the disease, although the recent discovery that mutations in the TIMP3 gene cause Sorsby’s fundus dystrophy (Weber et al. 1994), a macular degeneration, has led us to investigate TIMP1 for a causal role in RP2 (A.J. Hardcastle, D.L. Thiselton, M. Nayudu, R.M. Hampson, and S.S. Bhatcharya, in prep.). Recently, other genes have recently been mapped close to UBE1 in Xp11.3: PCTK1, DXS8237E, and ZNF157 (Nelson et al. 1995; Carrel et al. 1996). A UBE1-associated microsatellite, DXS7124 (Coleman et al. 1996), has been positioned proximal to DXS8083 by genetic analysis of a mini “meiotic breakpoint panel” (Gerken et al. 1995) comprising defined recombinants from XLRP families in our laboratory, and therefore falls within the RP2 critical region (data not shown). Unfortunately, DXS7124 is not informative in family F72, therefore possible involvement of these genes in RP2 remains open and is being addressed.

In addition to the meiotic breakpoint panel, we are currently employing a variety of resources to physically map and order genes and markers in the RP2 critical interval flanked by DXS6616 and DXS8083. This combination of genetic and physical mapping methods will enable us to further define the location of the defective gene.

METHODS

Subjects and Samples

Two families are presented: family NRP from the USA, comprising four generations with DNA for nine affected males and eight obligate carrier females, and family F72 from Belgium, comprising three generations with DNA for
four affected males and two obligate carrier females. A diagnosis of XL RP was based on detailed family history and comprehensive ophthalmological tests including fundus examination, visual field assessment, fluorescein angiography, and electroretinogram (ERG) measurements. DNA examination, visual field assessment, fluorescein angiogram comprehensive ophthalmological tests including fundus extraction from peripheral whole blood was performed using the Nucleon II kit (Scotlab).

Detection of Microsatellite Polymorphisms

The forward primer for each microsatellite was end-labeled with [γ-32P]ATP by incubating the primer at 37°C for 45 min with T4 polynucleotide kinase (New England Biolabs). The dinucleotide repeats were then amplified from 100 ng of genomic DNA as described previously (Thiselton et al. 1995). Alleles were detected by electrophoresing the PCR products on 6% denaturing polyacrylamide gels (Pro-mega), followed by exposure to X-ray film. Specific primer details and PCR conditions for each microsatellite can be obtained from GDB (1995).

Haplotype Analysis

Nineteen microsatellite markers spanning ~25 cm from Xp21.1 to Xp11.22 (Pain et al. 1995; Thiselton et al. 1995) were used to generate haplotypes for all sampled individuals. From Xp21.1 to Xp11.22 the order of markers is known: CYBB-DSX1110-DSX556-DSX574-DSX977-DSX228-DSX7-MAOB-DSX8080 (afmc012zc1)-DSX8083 (afmc024xc5)-DSX1003-DSX1055-SYN1-DSX426-DSX1367-DSX6616-DSX6941-DSX1126-DSX7 (Nelson et al. 1995; Dib et al. 1996). Haplotypes were constructed assuming the minimal number of recombination events.

Construction of a YAC Contig Spanning the RP2 Proximal Boundary in Xpll.23

Sources of Initial STSs and ESTs

Details of STSs corresponding to genetic markers (SYN1, PFC, DXS426, DXS1367, DXS6616, DXS6941, DXS722, DXS1126, and DXS1240) and expressed sequences TIMP-1, DXS1004E (ZNF41), and DXS1011E may be obtained from GDB (1995). Primer pairs for microsatellite DXS6941 were kindly provided by A. Meindl (Kinderpoliklinik der Universität München, Germany). Published sequence information for other genes assigned to the region was used to design EST markers from the 3' untranslated regions (GATA1, MG21, MG44, MG61, and MG81) and these are described in Table 1. STSs for genes ELK1 and ZNF81 were isolated from somatic cell hybrids containing portions of the X chromosome as their only human component; Kag 2.3 (Xp21.1-Xpter) and Sin 176 (deleted for Xp22.1-Xp11.22; Lafreniere et al. 1991) to provide an initial indication that they derived from proximal Xp.

Identification and Initial Characterization of YACs

YACs were identified for a STS or EST marker by PCR-based screening of hierarchical pools of clones from the ICI 4X YAC library (Anand et al. 1990). Intact DNA from each positive clone was prepared in agarose plugs and analysed by PCR to verify STS/EST content, and by pulsed-field gel electrophoresis (PFGE) (CHEF DRRI; Bio-Rad) to assess the size and purity of the YACs present. PFGE conditions were run time 17-22 hr, 5 V/cm, with a pulse time ramped from 60-90 sec in 0.5 x TBE at 14°C. YAC sizes were estimated by comparison with yeast chromosome size standards (Bio-Rad) on 1% agarose gels by visual inspection after staining the gel with ethidium bromide. If no distinct YAC was visible, Hybond N+ (Amersham) blots were probed from acid-nicked gels by Southern transfer (Sambrook et al. 1989) and hybridized at 65°C overnight with [a-32P]CTP-labeled total human DNA. The contig was supplemented with YACs for markers SYN1 and 2bC6 (DSX226) from the ICRF YAC library (Larin et al. 1991) through the ICRF Reference Library Database.

Creation of YAC Insert-end STSs

YAC insert ends were isolated by two methods: ALU-vector PCR using human-specific ALU primers AE1 (5'-GCTTCCTCAAAGTGCTGGAGATTACAG-3') or AE3 (5'-CCAT/CTGCACCTACGCTTGGG-3') and primers specific for the left and right arms of the pYAC vector (LA 5'-TACCCCGTTCTCGAGACATCCTGGCGG-3'; RA 5'-ATATAGGCAGCCAAGCCAGCTGGTGCGG-3'), or Vectorette PCR (Riley et al. 1990), using restriction enzymes PvuII, DraI, and EcoRI with 6-bp recognition sequences, which were found generally to yield large insert terminal PCR products. PCR products were purified and sequenced directly using an internal Vectorette unit primer or pYAC4 primer (LA 5'-GTGGGTTAAGGGCAAG-3'; RA 5'-GTCGAAAAGCCCGATCTCAA-3'). End-fragment sequences were subjected to FASTA data-base identity searches before PCR primer design. The new STSs developed from YAC insert ends were tested against a pair of somatic cell hybrids containing portions of the X chromosome as their only human component; Kag 2.3 (Xp21.1-Xpter) and Sin 176 (deleted for Xp22.1-Xp11.22; Lafreniere et al. 1991) to provide an initial indication that they derived from proximal Xp.

YAC Contig Construction by STS-content Mapping

The YAC contig was constructed by determining the STS content of each YAC and therefore establishing overlaps between clones. The orientation of the contig was deduced from the known order of genetic markers mapping to the contig, that is, Xpter-PFC-DSX426-DSX722-XS1126-Xcen (Nelson et al. 1995).

ACKNOWLEDGMENTS

We gratefully acknowledge the Human Genome Mapping Project Resource Centre for invaluable YAC and primer provision, and thank the Ulverscroft Foundation (D.L.T.) and Wellcome Trust (A.J.H.; grant no. 042019/2/94/Z/ WRE/MB/JAT) for their support. Sample collection and diagnosis for family NRP was supported by the National Society to Prevent Blindness awarded in memory of Silas Adelheim (B.K.S.).

The publication costs of this article were defrayed in part by payment of page charges. This article must there-
fore be hereby marked “advertisement” in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES


GENETIC AND PHYSICAL MAPPING OF THE RP2 LOCUS


Received July 2, 1996; accepted in revised form September 6, 1996.
SHORT COMMUNICATION

Genetic and Physical Mapping of Five Novel Microsatellite Markers on Human Xp21.1–p11.22

D. L. Thiselton, † S. Lindsay, † S. Kamakari, * A. J. Hardcastle, * P. Roustan, † and S. S. Bhattacharya *

*Department of Molecular Genetics, Institute of Ophthalmology, Bath Street, London EC1V 9EL, United Kingdom; †Molecular Genetics Unit, Division of Human Genetics, University of Newcastle upon Tyne, United Kingdom; and †Laboratoire de Génétique Moléculaire de la Neurotransmission et des Processus Neurodégénératifs, CNRS, Bat 32 rde. Avenue de la Terasse, 91 198 Gif-sur-Yvette, Paris, France

Received April 18, 1994; revised October 12, 1994

Five polymorphic CA-dinucleotide repeats, identified in cosmids from the short arm of the human X chromosome, have been characterized and localized to Xp21.1 (DXS572), Xp11.4 (DXS556, DXS574), and Xp11.22–p11.23 (DXS722, DXS573). Genetic mapping with respect to five reference markers that include the gene for CGD (CYBB in Xp21.1), complemented by physical mapping information, has indicated the order tel – DXS572 – CYBB – DXS1110 – DXS556 – DXS574 – DXS7–DXS426–DXS722–DXS573–DXS255–cen. © 1995 Academic Press, Inc.

The proximal short arm of the human X chromosome is the site of several genetic eye disease loci, e.g., Aland Island eye disease, retinitis pigmentosa form 2, X-linked congenital stationary night blindness (12), and X-linked progressive cone dystrophy (10). As yet, there are still relatively few reported microsatellites identified in this region as compared to other areas of the X chromosome (3). The microsatellites described here should, therefore, be helpful in genetic and physical mapping studies involving the proximal short arm (16).

We have isolated 5 cosmids (HX20, HX81, HX82, HX91, and HX97) from a library that has been previously described (8), and these have been assigned Human Gene Mapping (HGM) symbols DXS556, DXS572, DXS573, DXS574, and DXS722, respectively. Sublocalization of the cosmids was carried out by in situ hybridization, showing that DXS572 maps to Xp21.1, DXS556 and DXS574 map to Xp11.4–p11.3, and DXS722 and DXS573 map to Xp11.23–p11.22 (Table 1). These cosmids were shown to contain a microsatellite by hybridization with a poly(dCA·dGT) probe (Pharmacia). Cosmids were digested with Sau3A and subcloned into the plasmid vector pT7T3 (Pharmacia). Microsatellite-containing subclones were identified and sequenced, and flanking PCR primers were designed as described previously (9).

Table 1 shows the primer sequences for DXS572, DXS574, and DXS722, along with the expected heterozygosities, calculated using the allele frequencies shown. The allele details for DXS556 and DXS573 have been described previously (13, 15).

The 5 loci were mapped genetically with respect to five reference markers whose map positions are well established (12): CYBB (Xp21.1) (4), DXS1110 (Xp21.1) (14), DXS7 (Xp11.3–p11.4), DXS426 (Xp11.23), and DXS255 (Xp11.22). Their order on Xp is also known to be tel–CYBB–DXS1110–DXS7–DXS426–DXS722–DXS573–DXS255–cen (12). The most significant lod scores (Zmax) and corresponding maximum likelihood recombination fractions (θmax) for the new markers are given in Table 2.

Figure 1 shows six recombinant meiotic events informative for ordering the loci. These give the order tel–DXS572–(CYBB, DXS1110)–DXS556–DXS574–DXS7–(DXS426, DXS722, DXS573)–DXS255–cen.

No crossovers were observed among DXS426, DXS722, and DXS573, as evidenced by the two-point linkage data in Table 2. However, physical mapping studies provided information regarding the order of these two markers. DXS573 and DXS722 are seen to be different microsatellites by their distinctly different levels of heterozygosity (72 and 46%, respectively) and their unique flanking sequences. Testing each against a panel of somatic cell hybrids allowed the two microsatellites to be ordered with respect to each other. Sin 176 is a human–mouse hybrid cell line containing the X chromosome as its only human component, which is deleted for the region Xp21.1–p11.22 (6). By PCR analysis (data not shown), DXS573 is present in this hybrid, as is DXS255 (6), whereas DXS722 is absent, as is DXS426 (1). This clearly indicates that DXS573 lies proximally to both DXS426 and DXS722. DXS722 was shown to lie proximal to DXS426 by its presence in YAC ICRFY900F0501, containing the distal ornithine aminotransferase-like sequence cluster (OATL1), which does not cover the DXS426 region and is known to lie proximal to DXS426 (12).

Finally, an ILINK order of preference was carried out on the marker loci (7). As there were too many

Table 1  shows the primer sequences for DXS572, DXS574, and DXS722, along with the expected heterozygosities, calculated using the allele frequencies shown. The allele details for DXS556 and DXS573 have been described previously (13, 15).

The 5 loci were mapped genetically with respect to five reference markers whose map positions are well established (12): CYBB (Xp21.1) (4), DXS1110 (Xp21.1) (14), DXS7 (Xp11.3–p11.4), DXS426 (Xp11.23), and DXS255 (Xp11.22). Their order on Xp is also known to be tel–CYBB–DXS1110–DXS7–DXS426–DXS722–DXS573–DXS255–cen (12). The most significant lod scores (Zmax) and corresponding maximum likelihood recombination fractions (θmax) for the new markers are given in Table 2.

Figure 1 shows six recombinant meiotic events informative for ordering the loci. These give the order tel–DXS572–(CYBB, DXS1110)–DXS556–DXS574–DXS7–(DXS426, DXS722, DXS573)–DXS255–cen.

No crossovers were observed among DXS426, DXS722, and DXS573, as evidenced by the two-point linkage data in Table 2. However, physical mapping studies provided information regarding the order of these two markers. DXS573 and DXS722 are seen to be different microsatellites by their distinctly different levels of heterozygosity (72 and 46%, respectively) and their unique flanking sequences. Testing each against a panel of somatic cell hybrids allowed the two microsatellites to be ordered with respect to each other. Sin 176 is a human–mouse hybrid cell line containing the X chromosome as its only human component, which is deleted for the region Xp21.1–p11.22 (6). By PCR analysis (data not shown), DXS573 is present in this hybrid, as is DXS255 (6), whereas DXS722 is absent, as is DXS426 (1). This clearly indicates that DXS573 lies proximally to both DXS426 and DXS722. DXS722 was shown to lie proximal to DXS426 by its presence in YAC ICRFY900F0501, containing the distal ornithine aminotransferase-like sequence cluster (OATL1), which does not cover the DXS426 region and is known to lie proximal to DXS426 (12).

Finally, an ILINK order of preference was carried out on the marker loci (7). As there were too many

Table 1  shows the primer sequences for DXS572, DXS574, and DXS722, along with the expected heterozygosities, calculated using the allele frequencies shown. The allele details for DXS556 and DXS573 have been described previously (13, 15).

The 5 loci were mapped genetically with respect to five reference markers whose map positions are well established (12): CYBB (Xp21.1) (4), DXS1110 (Xp21.1) (14), DXS7 (Xp11.3–p11.4), DXS426 (Xp11.23), and DXS255 (Xp11.22). Their order on Xp is also known to be tel–CYBB–DXS1110–DXS7–DXS426–DXS722–DXS573–DXS255–cen (12). The most significant lod scores (Zmax) and corresponding maximum likelihood recombination fractions (θmax) for the new markers are given in Table 2.

Figure 1 shows six recombinant meiotic events informative for ordering the loci. These give the order tel–DXS572–(CYBB, DXS1110)–DXS556–DXS574–DXS7–(DXS426, DXS722, DXS573)–DXS255–cen.

No crossovers were observed among DXS426, DXS722, and DXS573, as evidenced by the two-point linkage data in Table 2. However, physical mapping studies provided information regarding the order of these two markers. DXS573 and DXS722 are seen to be different microsatellites by their distinctly different levels of heterozygosity (72 and 46%, respectively) and their unique flanking sequences. Testing each against a panel of somatic cell hybrids allowed the two microsatellites to be ordered with respect to each other. Sin 176 is a human–mouse hybrid cell line containing the X chromosome as its only human component, which is deleted for the region Xp21.1–p11.22 (6). By PCR analysis (data not shown), DXS573 is present in this hybrid, as is DXS255 (6), whereas DXS722 is absent, as is DXS426 (1). This clearly indicates that DXS573 lies proximally to both DXS426 and DXS722. DXS722 was shown to lie proximal to DXS426 by its presence in YAC ICRFY900F0501, containing the distal ornithine aminotransferase-like sequence cluster (OATL1), which does not cover the DXS426 region and is known to lie proximal to DXS426 (12).

Finally, an ILINK order of preference was carried out on the marker loci (7). As there were too many
TABLE 1
PCR Conditions and Allele Details for DXS572, DXS574, and DXS573

<table>
<thead>
<tr>
<th>Locus</th>
<th>Location</th>
<th>Primer sequences (EMBL Accession no.)</th>
<th>Allele</th>
<th>Length (nt)</th>
<th>Frequency</th>
<th>Expected heterozygosity</th>
</tr>
</thead>
<tbody>
<tr>
<td>DXS572 (pHX81)</td>
<td>Xp21.1</td>
<td>CA: 5'-GCACCATAGAAGCAGAAAAGTG-3'</td>
<td>A1</td>
<td>164</td>
<td>0.01</td>
<td>29%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GT: 5'-CCTATTTCCCTAGTCATC-3'</td>
<td>A2</td>
<td>162</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A3</td>
<td>160</td>
<td>0.93</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A4</td>
<td>158</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A5</td>
<td>156</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>DXS574 (pHX91)</td>
<td>Xp11.3-p11.4</td>
<td>CA: 5'-ATGGATATTTGAAGCAGCTAG-3'</td>
<td>A1</td>
<td>212</td>
<td>0.04</td>
<td>61%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GT: 5'-TCATCAAGCCGAATATGGA-3'</td>
<td>A2</td>
<td>210</td>
<td>0.38</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A3</td>
<td>208</td>
<td>0.48</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A4</td>
<td>204</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>DXS722 (pHX97)</td>
<td>Xp11.23</td>
<td>CA: 5'-GCTCATAGCTCTAGTGC-3'</td>
<td>A1</td>
<td>192</td>
<td>0.14</td>
<td>47%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GT: 5'-GATCCTGTCCTACAGGG-3'</td>
<td>A2</td>
<td>190</td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A3</td>
<td>188</td>
<td>0.70</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A4</td>
<td>186</td>
<td>0.02</td>
<td></td>
</tr>
</tbody>
</table>

Note. PCR amplifications were carried out for 30 cycles using one labeled primer as follows: denaturation at 94°C for 1 min, annealing at 56°C (DXS572, DXS574) or 57°C (DXS722) for 1 min, and extension at 72°C for 1 min, with a final elongation step at 72°C for 10 min. Each reaction was carried out in a 10-µl volume using 0.5 U of Taq polymerase and manufacturer’s buffer to give a final concentration of 1.5 mM MgCl₂ (Bioline). Allele frequencies were estimated from 59 (118 chromosomes) unrelated Caucasian females for DXS572, 25 (50 chromosomes) for DXS574, and 28 (56 chromosomes) for DXS722.

possible haplotypes to permit all 10 loci to be analyzed simultaneously, the task was simplified by initially ordering the 2 reference loci distal to DXS7 (CYBB, DXS1110) and the 2 proximal to DXS7 (DXS426, DXS255), with DXS7 in common to enable an overlap of the maps. The 5 novel microsatellites were then positioned onto this “reference map” in turn, by taking relevant groups of 3 loci and finding the new best order. This analysis confirmed the order derived from the haplotype data already described. The ILINK analysis could not separate the group DXS426, DXS722, and DXS573. However, these results taken in conjunction with the physical mapping studies give the final order for these loci, with estimated genetic distances in centimorgans, as tel-DXS572-6.4-CYBB-0.0-DXS1110-
9.0-DXS556-1.5-DXS74-3.5-DXS7-12.5-DXS426-0.0-DXS722-0.0-DXS573-4.7-DXS255-cen. The genetic distances between adjacent markers have been calculated from the recombination fractions (θmax) in Table 2 and the Kosambi mapping function (5) using the MAPFUN program (11).
The data presented here indicate an overall genetic distance of 31 cM between CYBB and DSX255. This is in close agreement with the recently published data of

TABLE 2
Two-Point Analyses of Locus Relationships

<table>
<thead>
<tr>
<th>Locus 1</th>
<th>Locus 2</th>
<th>θmax (1 vs 2)</th>
<th>Zmax</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYBB</td>
<td>DXS572</td>
<td>0.064</td>
<td>7.58</td>
</tr>
<tr>
<td>CYBB</td>
<td>DXS556</td>
<td>0.075</td>
<td>11.71</td>
</tr>
<tr>
<td>CYBB</td>
<td>DXS574</td>
<td>0.118</td>
<td>9.17</td>
</tr>
<tr>
<td>CYBB</td>
<td>DXS1110</td>
<td>0.000</td>
<td>15.57</td>
</tr>
<tr>
<td>DXS1110</td>
<td>DXS556</td>
<td>0.089</td>
<td>7.17</td>
</tr>
<tr>
<td>DXS556</td>
<td>DXS574</td>
<td>0.015</td>
<td>16.28</td>
</tr>
<tr>
<td>DXS556</td>
<td>DXS7</td>
<td>0.023</td>
<td>18.35</td>
</tr>
<tr>
<td>DXS574</td>
<td>DXS7</td>
<td>0.035</td>
<td>16.59</td>
</tr>
<tr>
<td>DXS7</td>
<td>DXS426</td>
<td>0.123</td>
<td>10.53</td>
</tr>
<tr>
<td>DXS7</td>
<td>DXS722</td>
<td>0.107</td>
<td>8.95</td>
</tr>
<tr>
<td>DXS7</td>
<td>DXS573</td>
<td>0.113</td>
<td>7.97</td>
</tr>
<tr>
<td>DXS426</td>
<td>DXS722</td>
<td>0.000</td>
<td>18.65</td>
</tr>
<tr>
<td>DXS426</td>
<td>DXS573</td>
<td>0.000</td>
<td>22.79</td>
</tr>
<tr>
<td>DXS722</td>
<td>DXS573</td>
<td>0.000</td>
<td>17.21</td>
</tr>
<tr>
<td>DXS722</td>
<td>DXS255</td>
<td>0.024</td>
<td>9.64</td>
</tr>
<tr>
<td>DXS573</td>
<td>DXS255</td>
<td>0.047</td>
<td>7.83</td>
</tr>
</tbody>
</table>

FIG. 1. Representation of recombinant meiotic events. Open and filled circles represent alleles corresponding to one or the other of the maternal chromosomes for informative loci. Horizontal bars indicate loci that were uninformative in a given individual.
Donnelly and colleagues (2) for a comparable region on the short arm of the X chromosome.

ACKNOWLEDGMENT

We gratefully acknowledge The Ulverscroft Foundation in funding D. L. Thiselton for this research.

REFERENCES


Dinucleotide repeat polymorphism at the DXS556 locus

D.L. Thiselton+, S. Kamakari+, A.R.J. Curtis, M.P. Coleman1, K.E. Davies1, S. Lindsay and S. Bhattacharya*+
Molecular Genetics Unit, University of Newcastle upon Tyne, NE1 7RU and *Institute of Molecular Medicine, John Radcliffe Hospital, Oxford OX3 9DU, UK

Source and Description of Clone: Cosmid HX20 was isolated from a library constructed using DNA from a mouse–human hybrid cell line, Thy-B1-33-12, which has the X chromosome as the only detectable human chromosome (1). It was then selected by hybridisation to a poly (dCA,dGT) probe (Pharmacia). pHX20 is a Sau3AI subclone of this cosmid which contains a (CA)10 microsatellite. pHX20 was sequenced and the sequences flanking the repeat (EMBL accession number X67603) were used to design PCR primers.

PCR Primers:

pHX20A = 5'-AGTTTGAGGGCTTCGTTTAC-3' (CA strand)
pHX20B = 5'-TATGAAGACAGCCAACTTAGA-3' (GT strand)

Polymorphism: Allelic fragments were detected on DNA sequencing gels. Lengths (nt) are:

A1 = 192
A2 = 190
A3 = 188
A4 = 186
A5 = 184
A6 = 182
A7 = 178
A8 = 176

Subclone pHX20 contains A5.

Frequencies: Estimated from 36 (72 chromosomes) unrelated Caucasian families:

<table>
<thead>
<tr>
<th>Allele</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>0.04</td>
</tr>
<tr>
<td>A2</td>
<td>0.06</td>
</tr>
<tr>
<td>A3</td>
<td>0.07</td>
</tr>
<tr>
<td>A4</td>
<td>0.13</td>
</tr>
<tr>
<td>A5</td>
<td>0.37</td>
</tr>
<tr>
<td>A6</td>
<td>0.04</td>
</tr>
<tr>
<td>A7</td>
<td>0.01</td>
</tr>
<tr>
<td>A8</td>
<td>0.29</td>
</tr>
</tbody>
</table>

Observed heterozygosity = 0.73.

Chromosomal Localization: DXS556 has been assigned to Xp11.23–p11.13 by hybridization of the cosmid HX20 to a panel of somatic cell hybrids DNAs and confirmed by fluorescence in situ hybridisation.

Mendelian Inheritance: X-linked co-dominant segregation was observed in 5 informative families.

PCR Conditions: PCR amplifications were carried out for 30 cycles using 32P end labelled primer pHX20A as follows: denaturation at 94°C for 1 minute, annealing at 57°C for 1 minute and extension at 72°C for 2 minutes, with a first denaturation step at 94°C for 3 min and a final elongation step at 72°C for 10 min.

Acknowledgements: This work was supported by the British Retinitis Pigmentosa Society, the Sir Jules Thorn Trust and the Wellcome Trust (SL grant no. 033229/Z/91/Z/1.27/DG).


* To whom correspondence should be addressed at *present address:
Department of Molecular Genetics, Institute of Ophthalmology, London EC1V 9EL, UK

---

Dinucleotide repeat polymorphism at D21S49 (21q22.3)

I.N. Bespalova1,2, S. Kiousis3, J.S. Chamberlain3 and M. Burmeister1,2,3,*
1Mental Health Research Institute, 2Department of Psychiatry and 3Department of Human Genetics, University of Michigan, Ann Arbor, MI 48109-0720, USA

Source/Description: A (GT)9 repeat was identified in pha SF58, which defines the locus D21S49 (1). A 1 kb Eco fragment containing the repeat was subcloned and partial sequenced. (Genbank accession no. L07094).

Primer Sequences:

Primer 1 = 5'-TTCTCGTAAGTCACCTGACTAC-3'
Primer 2 = 5'-CGACTGTCTTTCTAGCTTCTGAC-3'

Frequency: 80 chromosomes from the parents of the CEP pedigrees were analyzed. Heterozygosity = 0.70

<table>
<thead>
<tr>
<th>Allele size (bp)</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>166</td>
</tr>
<tr>
<td>A2</td>
<td>160</td>
</tr>
<tr>
<td>A3</td>
<td>158</td>
</tr>
<tr>
<td>A4</td>
<td>156</td>
</tr>
<tr>
<td>A5</td>
<td>154</td>
</tr>
<tr>
<td>A6</td>
<td>152</td>
</tr>
</tbody>
</table>

Reference CEPH genotypes:

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>13291-1</td>
<td>156/160</td>
</tr>
<tr>
<td>13291-2</td>
<td>158/160</td>
</tr>
</tbody>
</table>

Chromosomal Location: D21S49 is known to be on chromosome 21q22.3, between Mx and PKL, near D21S42 and D21S19 (2).

Mendelian Inheritance: Mendelian inheritance was observed in four CEPH pedigrees.

PCR Conditions: Amplification is performed in a volume of 2 µl. Each reaction contains 20 ng genomic DNA, 250 µM each of dATP, dGTP, dTTP, and dCTP, 0.5 U Taq polymerase, 30 mM Tris-HCl, pH 8.5, 50 mM KCl 2 mM MgCl2, 5 mM 2-mercaptoethanol, 0.01% gelatin, 0.1% Thesit. Amplifications conditions are 40 cycles of 92°C for 30 sec, 56°C for 30 sec, and 72°C for 1 min. Amplification product can be resolved on 15% polyacrylamide gels.

Acknowledgement: Supported by a grant from the Epilepsy Foundation of America (M.B.).
