

**MOLECULAR GENETIC STUDY OF
AUTOSOMAL DOMINANT RETINITIS
PIGMENTOSA ON CHROMOSOME 19q13.4**

Leen Abu Safieh

M.Sc.

A thesis submitted to the University College of London for the degree of
Doctor of Philosophy



Department of Molecular Genetics
Institute of Ophthalmology
University College London
University of London
Bath Street
London EC1V 9EL
UK
2003

ProQuest Number: U643756

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest U643756

Published by ProQuest LLC(2016). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code.
Microform Edition © ProQuest LLC.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

Retinitis Pigmentosa (RP) describes a clinically and genetically heterogeneous group of inherited retinal disorders characterized by progressive loss of peripheral vision and night blindness. RP can be inherited as an autosomal dominant, autosomal recessive or X-linked trait. The autosomal dominant form of RP (adRP) can be caused by mutations in 11 genes and one other locus, for which the gene remains to be identified. The work presented in this thesis relates to the adRP locus that maps to chromosome 19q13.4 (designated *RP11*). Affected members from *RP11*-linked pedigrees have type II, the regional form of RP. Uniquely all *RP11* linked families also exhibit an incomplete penetrance phenotype defined by the presence of completely asymptomatic disease gene carriers.

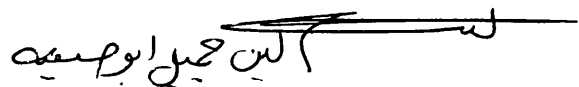
The aim of this work was to identify and characterize the *RP11* gene and to determine the molecular basis of the partial penetrance phenotype. At the time this work was initiated the *RP11* locus was confined to a ~ 500kb genomic region, DNA of which was represented by several BAC clones. As the sequence of these BAC clones was emerging due to the efforts of the Human Genome Mapping Consortium, a systematic analysis of this sequence was carried out to identify positional candidates for *RP11*. This thesis specifically describes the characterization of genes located in BAC clone 3093M3. In addition to the sequence analysis, ESTs mapping to the *RP11* interval were also characterized. The screening and exclusion of one such EST i.e. WI-17997, encoding a hypothetical gene LOC911663, is described in chapter 3. This EST was one of the few ESTs that mapped to the interval. Hence it was fully characterized and screened as a candidate for *RP11*. Screening of *insilico*-predicted genes located in BAC clone 3093M3 led to the identification of pathogenic mutations in a novel gene, *PRPF31* subsequently found to encode a pre-mRNA splicing factor. PRPF31 protein is an integral component of the human U4.U6 snRNP particle and therefore essential to the spliceosome complex, which mediates pre-mRNA splicing. Identification of the *RP11* gene is described in chapter 4.

To investigate the possibility that the partial penetrance of disease is due to the differential expression of wild type *PRPF31* allele, the level of *PRPF31* expression in symptomatic and asymptomatic individuals both at the mRNA and the protein levels was investigated (chapter 5). Real-time quantitative RT-PCR was performed on RNA from lymphoblastoid cell lines derived from the adRP family, ADRP5 that segregates an 11bp deletion in exon 11 of *PRPF31*. The mRNA levels from only the wild-type allele of *PRPF31* were assayed using a probe designed across the deletion. The *PRPF31* protein levels from symptomatic and asymptomatic individuals were also assayed by Western blot analysis using an antibody specific to the wild-type *PRPF31* protein. The use of cell lines was validated by the observation that cell transformation did not alter *PRPF31* expression in the cell lines as compared with nucleated blood cells and donor retinas. A significant difference in wild-type *PRPF31* mRNA levels was observed between symptomatic and asymptomatic individuals ($p < 0.001$). This data were supported by Western blot analysis of the *PRPF31* protein. Therefore, partial penetrance in *RP11* could be due to the co-inheritance of a *PRPF31* gene defect and a low expressed wild type allele. This study revealed a potential avenue for future therapy, as it appears that moderate over expression of wild type *PRPF31* may prevent clinical manifestation of the disease.

Declaration

I declare that work presented in this thesis for the degree of Doctor of Philosophy is composed by myself and data herein is my own original work.

Leen Abu Safieh

A handwritten signature in black ink, appearing to be in Arabic script, positioned below the printed name. The signature is fluid and cursive, with a long horizontal stroke extending to the right.

Dedication

To Ayman, my loving husband and my parents (Jojo and Um-um) for their love, support and continuous prayers. Also I dedicate this thesis to all Palestinians all over the World. May ALLAH, the greatest God bless you all.

Acknowledgment

First, my greatest thanks to ALLAH for his blessing all through my life and for giving me the power and energy to do this work.

I would like to thank my supervisor, Professor Shomi Bhattacharya for giving me the opportunity to do this PhD in his laboratory, and for his support, help and advice during the course of this degree. I will be forever grateful to the best work partner ever, Dr Eranga Vithana for her great help, in and outside the lab and for being such a great friend who was always there when needed. Words can not say enough (as you know my English fails me sometimes). I really appreciate everything you did to help me, thank you for everything, in particular thank you for being a friend.

I would like to give a big hug (in writing only) to my BIG brother Neil, who used to know my bad and good days before I said a word, and always knew how to bring back the smile on my face (the ice-cream trick), thank you Neil. Special thanks goes the bride to be, Resh, for your help during writing, Suba for reading and correcting this thesis but mainly for help in tissue culture and protein experiments. I want to pass my thanks to my office mates: Myrto, Christina, Tin, and Quincy, and to all the lovely people in the department: firstly our lovely Zara, Bart, Naheed (Bunny), Alison, Beverly, Lili, Simon, Viny and Ashwin.

I am deeply grateful to my lovely husband for his support, understanding and love, if it was not for you I would not be where I am today, I love you Ayman. I want to extend my thanks to my parents, my sister Reema, my brothers, Mohammed, Mohanad, Moayaed, Ahmad, and my lovely little brother Amjad, my in-laws especially Abu Asi, my uncle Abu Hatem, his wife and my favorite cousin Refah, and finally my grandparents, I love you all. Finally, I want to thank all my friends in Palestine, UK and all over the globe for being so supportive. Thank you Eman (Pal), Samar (Kwt), Suha (USA), Nidaa (Grm), Nawaf, Mohamad Abdullah, Prof. K. Hindi and his lovely wife Flora, Sawsan, Naser, Sabri, Sarah, Muna, Nebal, Shireen, Rania, Eyad, Aysha, Afaf, Zeen, Maryam, Fadya, thank you for being great friends. Big thanks and appreciation goes to my dearest friends in London, Khalida and Amal for being so supportive and caring. I'm really thankful to those who took part over the last few months of my suffering and helped me get over my stress. **I can't believe I did it.**

Publication list

- Chakarova CF, Hims MM, Bolz H, **Abu Safieh L**, Patel RJ, Papaioannou MG, Inglehearn CF, Keen TJ, Willis C, Moore AT, Rosenberg T, Webster AR, Bird AC, Gal A, Hunt D, Vithana EN, Bhattacharya SS. Mutations in *HPRP3*, a third member of pre-mRNA splicing factor genes, implicated in autosomal dominant retinitis pigmentosa. *Hum Mol Genet.* 2002 1; 11 (1):87-92.
- Vithana EN, **Abu Safieh L**, Allen MJ, Carey A, Papaioannou M, Chakarova C, Al-Maghtheh M, Ebenezer ND, Willis C, Moore AT, Bird AC, Hunt DM, Bhattacharya SS. A human homolog of yeast pre-mRNA splicing gene, PRP31, underlies autosomal dominant retinitis pigmentosa on chromosome 19q13.4 (*RP11*). *Mol Cell.* 2001 8 (2): 375-81.
- Payne A., Vithana E., Khaliq S., Hameed A., Deller J., **Abu Safieh L.**, Kermani S., Leroy B.P., Mehdi S.Q., Moore A.T., Bird A.C, Bhattacharya SS. RP1 protein truncating mutations predominate at the RP1 adRP locus. *Invest Ophthalmol Vis Sci.* 2000 (13): 4069-73.
- Vella M, Slater E, **Abu Safieh L**, Hussein AS, Greenwell P. (2000). The application of databases and PCR in the cloning of glycosidase genes from the protozoan *Tritrichomonas foetus*. *Mol Biotechnol*; 15 (1): 1-10
- **Leen Abu Safieh**, Eranga N. Vithana, Alan C Bird, David M. Hunt, Stephen. A. Bustin, Shomi. S. Bhattacharya. Expression of precursor RNA processing factor 31 (*PRPF31*) mRNA in patients with autosomal dominant retinitis pigmentosa: a molecular ‘clue’ for incomplete penetrance? *IVOS.*, (2003) *in press*.
- **Abu Safieh L**, Vithana EN, Papaioannou MG, Moore, A.T., Bird, A.C, Bhattacharya, S.S. Deletion mutations predominate in *PRPF31*, the autosomal dominant retinitis pigmentosa gene on 19q13.4. (2003). (manuscript in preparation).

- Vithana E.N, Chakarova C.F, Bolz H., **Abu Safieh L.**, Wilkie, S., Moore, T., Gal, A., Bird, A.C., Hunt, D.M., Bhattacharya, S.S. (2002). Identification of mutations in *HPRP3*, the gene for autosomal dominant retinitis pigmentosa on chromosome 1 and Functional analysis of mutations. *Invest. Ophthalmol. Vis. Sci.*: 43: 790.
- **Abu Safieh L**, Bhattacharya, S.S., Vithana E.N, Papaioannou M, Moore, T., Bird, A.C., Hunt D (2002). Identification of new mutations in *PRPF31*, the autosomal dominant retinitis pigmentosa gene on 19q13.4. *Invest. Ophthalmol. Vis. Sci.*: 43: 794.
- Bhattacharya S.S., Vithana E.N, **Abu Safieh L**, Hunt D.M. (2001). Transcript map of the adRP locus on chromosome 19 and investigation of candidate genes. *Invest. Ophthalmol. Vis. Sci.*: 42: 3459.
- Kaushal S., Vithana E.N., **Abu Safieh L.**, Payne, A., Bhattacharya S.S. (2000). Mutation screening of RPE G-Coupled Protein Receptor (*RGR*) in Retinitis Pigmentosa. *Invest. Ophthalmol. Vis. Sci.*: 41: 1014.

Table of Contents

	Page No.
Abstract	I
Declaration	II
Dedication	III
Acknowledgment	IV
Publication list	V
Table of contents	IX
List of figures and tables	
List of figures	XV
List of tables	XVII
Amino acid code	XIX
Abbreviations	XX
 <i>Chapter 1</i>	
<i>General Introduction</i>	
1.1 Human genome project	2
1.1.1 Sequencing strategies applied to genome sequencing	3
1.1.2 Generation of large insert clones for sequencing	3
1.1.3 Number of total genes estimated in the human genome	4
1.1.4 Data management of the human genome project	4
1.2 Positional cloning	5
1.2.1 Physical mapping	6
1.2.2 Bioinformatics, gene identification and characterization	7
1.2.2.1 Gene prediction tools	9
1.3 The eye	11

- Vithana E.N, Chakarova C.F, Bolz H., **Abu Safieh L.**, Wilkie, S., Moore, T., Gal, A., Bird, A.C., Hunt, D.M., Bhattacharya, S.S. (2002). Identification of mutations in *HPRP3*, the gene for autosomal dominant retinitis pigmentosa on chromosome 1 and Functional analysis of mutations. *Invest. Ophthalmol. Vis. Sci.*: 43: 790.
- **Abu Safieh L.**, Bhattacharya, S.S., Vithana E.N, Papaioannou M, Moore, T., Bird, A.C., Hunt D (2002). Identification of new mutations in *PRPF31*, the autosomal dominant retinitis pigmentosa gene on 19q13.4. *Invest. Ophthalmol. Vis. Sci.*: 43: 794.
- Bhattacharya S.S., Vithana E.N, **Abu Safieh L.**, Hunt D.M. (2001). Transcript map of the adRP locus on chromosome 19 and investigation of candidate genes. *Invest. Ophthalmol. Vis. Sci.*: 42: 3459.
- Kaushal S., Vithana E.N., **Abu Safieh L.**, Payne, A., Bhattacharya S.S. (2000). Mutation screening of RPE G-Coupled Protein Receptor (*RGR*) in Retinitis Pigmentosa. *Invest. Ophthalmol. Vis. Sci.*: 41: 1014.

Table of Contents

	Page No.
Abstract	I
Declaration	II
Dedication	III
Acknowledgment	IV
Publication list	V
Table of contents	IX
List of figures and tables	
List of figures	XV
List of tables	XVII
Amino acid code	XIX
Abbreviations	XX
 <i>Chapter 1</i>	
<i>General Introduction</i>	
1.1 Human genome project	2
1.1.1 Sequencing strategies applied to genome sequencing	3
1.1.2 Generation of large insert clones for sequencing	3
1.1.3 Number of total genes estimated in the human genome	4
1.1.4 Data management of the human genome project	4
1.2 Positional cloning	5
1.2.1 Physical mapping	6
1.2.2 Bioinformatics, gene identification and characterization	7
1.2.2.1 Gene prediction tools	9
1.3 The eye	11

1.3.1	Basic anatomy	11
1.3.2	Function of the eye	12
1.3.3	Structure of the retina	13
1.3.4	The photoreceptor cells	14
1.3.5	Retinal pigment epithelium	18
1.3.6	Phototransduction pathway	19
1.4	Retinal dystrophies	22
1.5	Genes causing retinal dystrophies	28
1.6	A historical overview of the <i>RP11</i> locus	30
1.6.1	<i>RP11</i> families	31
1.7	Aims of the thesis	32
	Chapter 2	33
	Materials and methods	
2.1	Preparation of PCR templates	33
2.1.1	DNA extraction	33
2.1.2	RNA extraction	34
2.1.2.1	RNA quantification	34
2.1.3	Total RNA extraction from retina and fresh blood samples	35
2.2	PCR reactions	36
2.2.1	PCR amplification	36
2.2.2	Expression studies	37
2.2.3	Real Time RT-PCR amplification	37
2.2.3.1	TaqMan primers and probe	38
2.3	Gel electrophoresis	39
2.3.1	Agarose gel electrophoresis	39
2.3.2	SDS-gel electrophoresis	40
2.3.2.1	Coomassie staining and visualising protein on SDS gels	41
2.3.2.2	Transfer of proteins from SDS gels to	41

	nitrocellulose membrane	
2.4	Mutation detection	42
2.4.1.1	Heteroduplex analysis- gel based analysis	42
2.4.1.2	Heteroduplex analysis- DHPLC WAVE ^R Nucleic Acid Fragment Analysis System	43
2.4.2	Direct sequencing	44
2.4.4	Restriction digest analysis	45
2.5	Northern blot analysis	45
2.6	Tissue culturing of transformed lymphocytes cell lines	46
2.7	Preparation of soluble cell extracts	47
	Solutions	48
	Chapter 3	49
	<i>Characterisation and mutation screening of an EST localised to the RP11 critical interval</i>	
3.1	Introduction	49
3.1.1	Mapping ESTs to the <i>RP11</i> interval: constructing a transcript map of the interval	49
3.1.2	Characterisation in of the genomic sequence using NIX analysis	50
3.2	Results and discussion	51
3.2.1	Mapping of WI-17997 EST	51
3.2.2	Expression of WI-17997 EST	55
3.2.3	Protein identities and possible function	56
3.2.4	Genomic organisation of hypothetical gene LOC91663	56
3.2.5	NIX analysis of 331H23 genomic sequence to obtain genomic organisation of the putative gene LOC91663	60
3.2.6	Mutation screening in <i>RP11</i> linked families	60

	Chapter 4	62
	<i>Identification of the RP11 gene</i>	
4.1	Introduction	62
4.2	Towards identification of the <i>RP11</i> gene	64
4.3	Methods	68
4.4	Results	68
4.4.1	<i>PRPF31</i> , the gene for <i>RP11</i>	68
4.4.1.1	Genomic organisation of <i>PRPF31</i>	68
4.4.2	Mutation screening	70
4.4.2.1	Mutation screening in <i>RP11</i> linked families	70
4.4.2.2	Screening for deletions in the <i>PRPF31</i> gene	77
4.4.2.2.1	Screening <i>PRPF31</i> for a large deletion in <i>RP11</i> linked family ADRP2	77
4.4.2.2.1.2	Possible clinical implication of the deletion in the ADRP2 family	83
4.4.2.2.2	Investigation of a putative deletion in <i>RP11</i> linked family, ADRP11	84
4.4.2.3	Screening other RP cases for mutations in the <i>RP11</i> gene	86
4.4.2.3.1	Mutation screening in small families	86
4.4.2.3.2	Mutations identified in the sporadic RP panel	87
4.4.2.3.2.1	Screening the sporadic RP panel for small deletions	91
4.4.2.4	Identification of non-pathogenic sequence variations within the <i>PRPF31</i> gene	92
4.5	Discussion	94
4.5.1	Predictive function of <i>PRPF31</i> based on protien homology	94
4.5.1.1	Pre mRNA splicing	97
4.5.2	<i>PRPF31</i> , the gene responsible for <i>RP11</i>	101

4.5.3	Possible functional consequences of <i>PRPF31</i> mutations	103
4.5.4	Why only the retina? Physiological implications of <i>PRPF31</i> , on photoreceptor cells	106
4.5.5	Splicing factors causing other forms of adRP	107
	CHAPTER 5	112
	<i>Genetic and biochemical characterisation of the partial penetrance phenotype of the RP11 locus</i>	
5.1	Introduction	110
5.1.1	Possible causes for the partial penetrance phenotype in <i>RP11</i> families	111
5.2	Aims and methods	114
5.2.1	Quantification of gene expression through realtime quantitative RT-PCR	115
5.2.2	Why the ADRP5 family	121
5.3	Results	123
5.3.1	Single nucleotide polymorphisms analysis within <i>PRPF31</i> symptomatic and asymptomatic individuals of ADRP5 and ADRP29	123
5.3.2	RNA extraction for <i>PRPF31</i> expression study	126
5.3.3	Construction of the standard curve	129
5.3.4	Analysis of quantitative RT-PCR on RNA extracted from cell lines.	131
5.3.5	Comparison of <i>PRPF31</i> copy numbers between tissue culture cells and nucleated blood cells	136
5.3.6	Quantitative RT-PCR on post-mortem retinal tissue and nucleated blood cells	137
5.3.7	Western blots analysis of PRPF31 protein extracted from cell lines	138

5.4	Discussion	140
5.4.1	Significance of the mRNA expression study	140
5.4.2	<i>PRPF31</i> expression in retina	142
5.4.3	Possible genetic cause for the partial penetrance phenotype of <i>RP11</i>	143
5.4.4	Future work	144
5.4.5	Implications for future therapy	146
	<i>CHAPTER 6</i>	147
	<i>General discussion and future prospects</i>	
6.1	Identification of the <i>RP11</i> gene	147
6.2	Significance of the <i>RP11</i> gene	148
6.3	Identification of the genetic cause of partial penetrance phenotype in <i>RP11</i>	150
6.4	Functional and biochemical analysis of <i>PRPF31</i> mutations	151
6.5	<i>RP11</i> mouse model	153
6.6	Gene Therapy	155
	<i>References</i>	157
	<i>Appendix</i>	175

List of figures and tables

List of figures

Figure No.		Page No.
1.1	Cross-section of the human eye	12
1.2	Cross section of the human retina	15
1.3	Cross-section through a human rod cell.	17
1.4	Phototransduction pathway	22
2.1	Schematic diagram showing qRT-PCR study design	39
2.2	DHPLC elution profiles for the wild type and a variant sequence of exon 7 of <i>PRPF31</i> gene.	44
3.1	Physical map of the <i>RP11</i> critical region showing ESTs WI-17997 and A006107	53
3.2	The collated sequence of the transcript represented by EST WI-17997	54
3.3	Northern blot analysis of EST WI-17997.	55
3.4	Alignment of human hypothetical protein BC013995 against the mouse Myeloid up-Regulated protein	58
3.5	Genomic sequence of Hypothetical gene <i>LOC91663</i>	59
3.6	NIX analysis of the genomic sequence of WI17997 EST.	61
4.1	Physical map of the <i>RP11</i> critical interval	63
4.2	RT-PCR of two predicted genes within the <i>RP11</i> critical interval	64
4.3	Genes in the <i>RP11</i> critical interval	67
4.4	Expression profile of <i>PRPF31</i>	69
4.5	Segregation of the 11bp deletion mutation (1115-1125del) in pedigree ADRP5	72
4.6	Pedigree of ADRP29 and electropherogram showing the mutation identified in <i>PRPF31</i>	74
4.7	Pedigree of RP1907 and electropherogram showing the	75

	mutation identified in <i>PRPF31</i>	
4.8	Segregation of the 43bp deletion mutation (IVS6-3 to -45 del) in pedigree ADRP677	75
4.9	Segregation of a one base pair deletion (1141-1142delG) in the <i>RP11</i> linked family of Japanese origin	76
4.10a	Haplotype analysis of 12 markers from 19q13.4 in ADRP2	80
4.10b	Schematic diagram of SNPs and markers analysed in family ADRP2	81
4.10c	Deletion haplotype in ADRP2 between PRKCG intragenic marker and D19S785	82
4.11	Segregation of IVS1+54T>C sequence change ADRP11 linked family	85
4.12	Small family pedigrees screened for mutations in the <i>RP11</i> gene	86
4.13	Missense mutation identified in a small Greek family	87
4.14	Electropherogram of mutations identified in sporadic RP individuals 117 and 42	88
4.15	Deletion identified in sporadic 14 and digest segregation gel image in the family	90
4.16	PCR amplification of <i>PRPF31</i> introns and exons in sporadic RP individuals	92
4.17	<i>PRPF31</i> alignment	96
4.18	Splicing pathway	100
5.1	Schematic diagram showing the polymerisation associated, 5'-3' nuclease activity of the polymerase during PCR	117
5.2	Amplification plot showing the log of the change in fluorescence plotted versus the cycle number during qRT-PCR	119

5.3	ADRP5 family pedigree showing individuals enrolled for the mRNA/ protein study	122
5.4	Analysis of polymorphisms identified in <i>PRPF31</i> gene in two sips-hips from ADRP5 family	125
5.5	The schematic diagram of the ADRP5 pedigree	128
5.6	Standard curve used to calculate mRNA copy numbers in lymphoblatoid cell lines	131
5.7	Scatter plot showing <i>PRPF31</i> mRNA copy numbers	133
5.8	ADRP5 pedigree showing mRNA copy numbers in carrier and non-carrier individuals from ADRP5 family.	135
5.9	Scatter plot showing <i>PRPF31</i> mRNA copy numbers of the ADRP5 pedigree and blood and retinal tissues from non-carrier individuals	138
5.10	Western analysis of PRPF31 protein	140

List of tables

Table no.		Page no.
1.1	Genes implicated in retinitis pigmentosa	25
1.2	Autosomal dominant RP genes and their chromosomal localization	26
2.1	Agarose concentration versus the range of resolution of linear DNA.	40
3.1	The cDNA IMAGE clones for EST WI-17997 (UniGene cluster -Hs 250700)	52
4.1	Genes identified in the <i>RP11</i> critical interval	66
4.2	<i>PRPF31</i> intronic primers used for PCR amplification and direct sequencing of <i>PRPF31</i> exons.	70
4.3	Markers and SNPs used in identifying the extent of the	79

	deletion in ADRP2 family	
4.4	Non-pathogenic sequence variations in <i>PRPF31</i> identified in <i>RP11</i> linked families and in the sporadic RP panel	93
4.5	Mutations detected in the <i>PRPF31</i> genomic sequence	102
5.1	C _t values obtained for known input copy number of the <i>PRPF31</i> amplicon from three different RT-PCR runs	130
5.2	<i>PRPF31</i> mRNA expression in ADRP5 lymphoplastoid cell lines	134
5.3	<i>PRPF31</i> expression in retina, nucleated blood cells and lymphoblastoid cells from normal/non-carrier individuals	137

Amino acids code

First base							
Second base	U		C		A	G	
U	UUU	Phe	CUU	Leu	AUU	GUU	Val
	UUC		CUC		AUC	GUC	
	UUA	Leu	CUA		AUA	GUA	
	UUG		CUG		AUG	GUG	
C	UCU	Ser	CCU	Pro	ACU	GCU	Ala
	UCC		CCC		ACC	GCC	
	UCA		CCA		ACA	GCA	
	UCG		CCG		ACG	GCG	
A	UAU	Try	CAU	His	AAU	GAU	Asp
	UAC		CAC		AAC	GAC	
	UAA	STOP	CAA	Gln	AAA	GAA	Glu
	UAG	STOP	CAG		AAG	GAG	
G	UGU	Cys	CGU	Arg	AGU	GGU	Gly
	UGC		CGC		AGC	GGC	
	UGA	STOP	CGA		AGA	GGA	
	UGG	Trp	CGA		AGG	GGG	

Alphapatic amino acids	Aromatic amino acids	Acidic amino acids	Basic amino acids	Cyclic amino acid	Amino acids with hydroxyl- or sulfur containing side chains
A alanine G glycine V valine L leucine I isoleucine	F phenylalanine Y tyrosine W tryptophan	D aspartic acid E glutamic acid N asparagine Q glutamine	R arginine K lysine H histidine	P Proline	S serine C cysteine T threonine M methionine

Abbreviations

RP: Retinitis Pigmentosa

HGP: Human Genome Project

BAC: Bacterial Artificial Chromosome

YAC: Yeast Artificial Chromosome

PAC: P₁ derived artificial chromosome

HUGO: Human Genome Organisation

STSs: Sequence Tagged Sites

ESTs: Expressed Sequence Tags

NIX: Nucleotide Identification of X

HGMP: Human Genome Mapping Project

RPE: Retinal Pigment Epithelium

ERG: Electroretinogram

adRP: autosomal dominant Retinitis Pigmentosa

arRP: autosomal recessive Retinitis Pigmentosa

xlRP: X-linked Retinitis Pigmentosa

CSNB: Congenital Stationary Night Blindness

LCA: Leber Congenital Amaurosis

CRD: Cone-rod dystrophy

DHPLC: Denaturing High Performance Liquid Chromatography

Myadm: Myeloid-associated differentiation marker

SNPs: Single Nucleotide Polymorphisms

OSCAR: Osteoclasts Associated Receptor

NDUFA3: NADH dehydrogenase (ubiquinone)1 alpha subcomplex3

TFPT: TCF3 (E2A) fusion partner

PRPF31: precursor RNA processing factor 31

snRNPs or snurps: small nuclear RiboNucleoprotein Particles

NMD: Nonsense-Mediated Decay

RT-qPCR: Real-Time quantitative PCR

C_t: threshold cycles

EPP: Erythropoietic Protoporphyria

SMA: Spinal Muscular Atrophy

SMN: Survivor of Motor Neuron

CHAPTER 1

General Introduction

Retinitis pigmentosa (RP) is a progressive form of retinal degeneration displaying high genetic heterogeneity. RP can be inherited as an autosomal dominant, autosomal recessive or X-linked trait. This thesis describes the cloning of the autosomal dominant retinitis pigmentosa gene on chromosome 19q13.42 (designated as the *RP11* locus). Based on the significant number of families linked to *RP11*, it is thought to be a major dominant RP locus. Partial penetrance is a characteristic phenomenon associated with the *RP11* locus making the identification of the gene responsible for the *RP11* phenotype an exciting prospect.

In this chapter, the relevance of the Human Genome Project (HGP) will be discussed. The genome sequence information derived from this project has facilitated the identification of genes, which in turn has led to the cloning of many genes associated with hereditary human diseases. As stated above, the thesis describes the cloning of the *RP11* gene and therefore, only a brief outline of the HGP will be discussed followed by the structure and function of the eye along with a detailed review of the genetics of inherited retinal dystrophies.

1.1 Human genome project

The Human Genome Project represents an exciting and challenging mission for biologists and geneticists. It is an international effort aimed at delivering the complete nucleotide sequence of human DNA and providing accessible information on almost 80,000 genes as initially thought to be encoded by the human genome (Fields *et al.*, 1994). It is also aimed at using the DNA sequence information to explore gene function and form a novel framework for developing new therapeutic approaches as represented by gene therapy.

Since 1995, sequencing projects progressed rapidly on two fronts. Firstly, at the level of constructing the genetic and physical map of both the human and the mouse genomes (Hudson *et al.*, 1995; Nusbaum *et al.*, 1999). At the time, these maps were key tools for many positional cloning projects and aided the identification of disease genes in humans.

Secondly, distinct progress was also made on the sequencing level. The sequencing of the human genome, paralleled by many other genome projects, seeks to obtain the full DNA sequence of a variety of organisms. Some of these projects have already been completed, such as that of the mouse and *Caenorhabditis.elegans* (*C.elegans*).

Since the start of the human genome project there have been many updates about the state and progress of this project as well as other associated projects. In 1996, a comprehensive genetic map was published, 2 years later the genomic sequence of the *C.elegans* was completed. The complete sequence of chromosomes 21 and 22 were subsequently published (Dunham *et al.*, 1999; Hattori *et al.*, 2000; Takai and Jones, 2002). The final draft sequence of the human genome is now available through both the human genome project consortium and Celera Genomics, however, gaps in the sequence still remain to be closed (Lander *et al.*, 2001; Venter *et al.*, 2001).

1.1.1 Sequencing strategies applied to genome sequencing

The international human genome sequencing project was based on a mapping and sequencing strategy that involved generating an overlapping series of clones, which covered the entire genome. It required the digestion of genomic DNA into segments of about 150kb and cloning these into bacterial artificial chromosomes (BACs) to produce a number of overlapping clones. Each clone was then digested with restriction enzymes to produce a unique pattern or fingerprint for that clone. This pattern, when compared with that of other clones, identified overlaps between clones, allowing them to be aligned according to their sites in the genome at which restriction endonuclease cleaves are charted, resulting in the establishment of the complete physical map of the human genome. These BAC clones were then subcloned into smaller clones for sequencing (Lander *et al.*, 2001).

Celera Genomics, on the other hand, adopted a different approach from the publicly funded human genome project. This involved preparing small-insert clones directly from genomic DNA and sequencing from both ends of the double stranded cloned DNA, rather than from mapped BACs. Finally, the finished sequence was assembled through a sophisticated computer program that is capable of recognising overlaps in DNA sequences between clones (Venter *et al.*, 2001).

1.1.2 Generation of large insert clones for sequencing

Yeast artificial chromosome (YAC) clones were widely used in generating the first generation physical maps of the human chromosomes. However, YAC contig maps were of limited value because the inserts were often unstable and prone to rearrangements and therefore did not provide a true representation of the original starting DNA. As a result, second generation contig maps have relied on bacterial artificial

chromosome (BACs) and P₁ derived artificial chromosome (PACs) clones. The insert sizes of these clones are much smaller than that of YACs, ranges from 70 and 250kb for PACs and BACs respectively, compared to 1Mb of YACs. However, the PAC and BAC inserts were found to be much more stable (Ioannou and de Jong, 1996).

1.1.3 Number of total genes estimated in the human genome

Both sequencing projects (HGP and Celera) predict that the number of genes encoded by the HGP is far less than that which was initially estimated. The human genome project consortium estimates that there are 31,000 genes, while Celera estimates about 26,000 genes. Both figures represent almost a third of the number initially predicted. Investigating the full length mRNA and alternative spliced products are essential when identifying novel genes (Lander *et al.*, 2001; Venter *et al.*, 2001).

1.1.4 Data management of the human genome project

Several web-based databases have been established to manage and store the large amount of data arising from these initiatives. These data are accessible through the internet to investigators world-wide for their research analysis. The Human Genome Organisation (HUGO) was established in 1988 in order to co-ordinate these international efforts. This organisation made exchanging of research resources possible. It also encouraged public debate and advice on the implications of the human genome research. Another example is the Genome Database, which is a mapping database in which human genetic data is collated. Data on such websites is changing rapidly as more sequence becomes available. This progress allows rapid identification of candidate genes *in silico*, followed by mutation screening of relevant candidates, aided by information on gene structure. Gaps within the published human genome sequence still

exist and additional library screening and sequencing is required to obtain the complete sequence across such regions.

With the completion of the human genome project, more gene sequences and functional information will be accessible using bioinformatics techniques. However, experimental approaches will always be required to prove the exact location, splicing patterns, transcriptional activity, functions and the involvement of novel genes.

1.2 Positional cloning

Positional cloning was a widely used approach for gene identification which involved mapping the chromosomal region containing the gene by linkage analysis in affected families and then scouring the region to identify the gene. Positional cloning is a powerful but very laborious and time-consuming technique. However, many disease causing genes have been identified using this approach, such as the dystrophin and CFTR genes responsible for duchenne muscular dystrophy and cystic fibrosis respectively (Monaco *et al.*, 1986; Riordan *et al.*, 1989).

Genetic markers are used to map the disease gene by linkage analysis. Chromosomal walking is then performed to construct a contig of overlapping clones which contain genomic DNA sequence that represents the region of interest. Several approaches are used to identify candidate genes within the critical region, which are then analysed for pathological mutations. Several genes implicated in retinal degeneration have been identified using positional cloning, such as the Rab Escort Protein-1 (REP), XLRP2 and XLRP3 (<http://www.sph.uth.tmc.edu/retnet/disease.htm>).

Traditional positional cloning projects involved the isolation of genes from YACs and cosmids derived from the critical region using techniques such as cDNA selection (Parimoo *et al.*, 1993), direct screening of cDNA libraries with YACs, screening for the presence of CpG islands (Craig and Bickmore, 1994) and exon trapping (Church *et al.*, 1994). However, due to the success of the Human Genome project with the sequence of the entire human genome as the ultimate target, such gene identification methods soon became redundant. Currently, most positional cloning projects are being complemented with *in silico* identification of candidate genes. Such an approach was used in the identification of the RP18 gene on chromosome 1q (Chakarova *et al.*, 2002).

1.2.1 Physical mapping

Maps that reflect the actual distance in base pairs are called physical maps. These maps chart the chromosomal localisation of genes and other sequence tagged sites (STSs). Physical maps are powerful tools for the localization and isolation of genes, studying the organisation and evolution of genomes and as a preparatory step for efficient sequencing. They are based on organising minimal overlapping clones in the correct position and orientation in the genome, they constitute a scaffold for detecting genetic rearrangements, locating genomic markers along chromosomes, and elaborating sequence-ready maps for sequencing projects. Different techniques are used in constructing a physical map, such as clone-probe hybridisation approach, radiation-hybrid mapping and fingerprinting. Cloned fragments are cut by restriction enzymes, smaller DNA fragments are obtained, which are sequenced in detail (shotgun-sequencing), and the individual clone sequence is finally assembled.

1.2.2 Bioinformatics, gene identification and characterisation

Genes represent only a small fraction of human DNA, however they comprise the major biological function of the genome. Gene identification has been the main focus of biologists and the most challenging step in functional genomics research. There are many approaches to candidate gene identification and analysis. Candidate genes can be defined on the basis of its chromosomal location and its biological and functional properties, which can lead to the identification of the disease-causing genes.

The availability of the genomic and cDNA sequence on the electronic databases has made gene identification easier and faster. Information about tissues in which cDNA are expressed has also been advantageous since such information would determine the association of a gene with a phenotype. As the sequence information was growing, the need to store and analyse the information increased (<http://www3.oup.co.uk/nar/database/summary/12>).

Bioinformatics is widely used in almost each step following the initial linkage of disease loci. Bioinformatics is a new discipline that has arisen to support genomics (the study of DNA of an organism) and proteomics (the study of the proteins of an organism).

Several internet-based programs have been developed to assist researchers in their gene identification quest. Computer programmes are currently used to identify genes by searching databases of finished sequences. These programmes include Mapviewer (<http://www.ncbi.nlm.gov/>) and/or Ensembl (<http://www.ensembl.org/>). Both tools use the same database, however, they differ in data presentation and annotation system (Lichanska and Simpson, 2002). Mapviewer enables the identification of genomic contigs within a region of interest, it also contains links to all genes identified within the genomic region, their

structure and conserved domains within the protein sequence. Despite its wide use, Mapviewer has some limitations, such as it cannot navigate through the genome or zoom in and out of the region. Ensembl is a user-friendly database; it allows easy identification of the information relevant to the region of interest by collating all information available on the sequences within that region. It also offers partial annotation of sequences by incorporating all existing genes and genes identified by computational tools for gene structure prediction, such as GeneWise, or Genscan.

Links to GenBank, SNP, EST, Unigene and SWISS-PROT databases are provided through Ensembl. It also provides evidence supporting the structure of computationally predicted genes, and are based on sequencing similarity of the genomic fragment to sequences of Expressed Sequence Tags (ESTs), clones, UniGene sequences, and SWISS-PROT protein sequences. Following the identification of an EST within a region of interest, a number of clones should be fully sequenced and compared to the genomic sequence to prevent mis-interpretation of genomic structure. Identifying the tissue-specific cDNA libraries where the desired EST is expressed provides a clue of the gene's pattern of expression.

Most genes incorporated into the genome browsers are actually computationally predicted genes, this means that some of these hypothetical genes may represent pseudogenes and therefore real genes maybe missed due to limited prediction ability, for example, intronless genes (Lichanska and Simpson, 2002).

1.2.2.1 Gene prediction tools

Tools frequently used for prediction of gene structure are divided into three groups; these are splice site prediction, exon identification and full gene prediction. Programs such as HMgene and GeneScan are used to identify the full gene within a genomic fragment. These programs rely on recognising translation signals in addition to identifying exons by recognising splice sites. It is worthwhile mentioning that the ability to identify initial and terminal exons varies within and between programs since some programs have difficulty in identifying terminal rather than initial exons (Burset and Guigo, 1996; Thanaraj, 2000).

It is essential to account for accuracy of novel sequence prediction when analysing uncharacterised fragments of genomic sequences. Identification of internal exons relies mostly on the ability of the program to predict the splice sites (GT at 5' and AG at 3' end). However, various factors can affect accuracy of prediction of genes, such as misassembly, incomplete or inaccurate sequences or the presence of unusual sequences such as repetitive motifs within exons. Using multiple gene prediction programs gives a more reliable model since the accuracy of exon prediction varies between 40-90% (Thanaraj, 2000).

Development of prediction programs relies on artificially short fragments of high coding density. Such prediction programme's power and reliability decrease when confronted with large genomic sequences. Incomplete or inaccurate sequence assembly, algorithms used in predictions, incomplete transcriptome representation in the database and presence of a large number of RNA splice variants are mainly the reasons behind discrepancies between different prediction programs.

Programs have been designed to eliminate false positive predictions, which can be used for more complex sequence analysis (Burge and Karlin, 1997; Lukashin and Borodovsky, 1998). Nucleotide Identification

of X (NIX) for example, is another internet-based tool available through the Human Genome Mapping Project (HGMP) Resource Centre, Cambridge (<http://www.hgmp.mrc.ac.uk>) used in viewing results obtained by running many DNA analysis programs on DNA sequence, allowing the comparison of predicted exon and gene structures. NIX enables simultaneous evaluation of predictions from multiple programs, these include GRAIL, Fex, Hexon, MZEF, Genemark, Genefinder, FGene, BLAST (against many databases), RepeatMasker and tRNAscan.

Despite the availability of lots of gene prediction programs, there is still a need to confirm these predictions by laboratory experiments. It is essential to confirm whether the genomic fragments have been correctly assembled, since incorrect arrangement would affect prediction. Northern blot analysis, rapid amplification of cDNA ends (RACE) analysis or *in situ* hybridisation are laboratory approaches that can be used to provide more expression information about a gene.

Gene prediction through bioinformatics is only the beginning. There is still a need to confirm gene structure and function. Gene structure can be confirmed by analysis of large IMAGE cDNA clones, which are available free for researchers all over the world. Sequences of these clones are being deposited in EST databases (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene>) and allow immediate comparison with genomic sequence to determine intron/exon boundaries without the need for library screening, cloning and sequencing. However, there are a number of drawbacks. These include poor sequence quality; the chance of having chimaeric clones, or clones containing genomic sequence. Moreover, alternative splicing and partial sequences mean that clones from the same gene may not be identified as being from the same transcript.

Once the genomic organisation and expression information, about genes of interest have been established they can then be prioritised for

mutation screening. In order to determine the biological role of these genes and similarities of their encoded proteins between human and other organisms, the BLAST database can be used for both known and novel genes. The BLAST database is widely used to identify conserved domains and functionally related proteins in other species that might have been previously characterised. Searching the genome databases of other organisms, such as the mouse genome database (<http://www.informatics.jax.org/userdocs/aboutMGD.shtml>), may provide preliminary information about protein function. Further functional information can be obtained from existing transgenic animals by searching transgenic databases. Animal models are useful for studying the pathology caused by absence or mutation of the gene and in evaluating possible treatments.

1.3 The eye

1.3.1 Basic anatomy

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

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

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

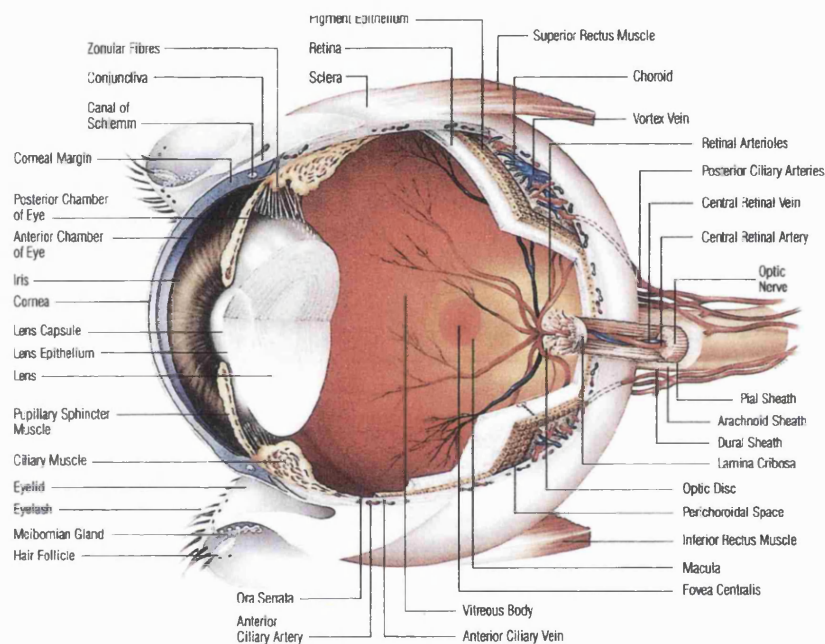


Figure 1.1 Cross-section of the human eye. (Pharmacia and Upjohn Ltd).

1.3.2 Function of the eye

The majority of the refraction of light (70-80%) occurs as light passes through the transparent cornea and the anterior chamber. The iris is the variable diaphragm that expands or contracts depending on the level of light and the distance of the object. Light then passes through the

posterior chamber and the lens, where a change in shape of the lens can alter its radius of curvature and focal length and where the refraction of light can accommodate the eye for near or far vision. The light then crosses a third refractive medium, the vitreous body, before striking the retina. When light hits the retina it evokes an action potential in the photoreceptors and these electrical impulses are carried to the brain by the optic nerve to be processed (Remington, 1998).

1.3.3 Structure of the retina

Studying the morphology and biological mechanisms of the retina is important for understanding retinal degeneration. The retina is the innermost layer of the eye, which is an outpouching of the central nervous system (CNS). The vertebrate neural layer has six major neuronal cell types, which forms a connecting sequence that carries light signal via the optic nerve fibres leading to the brain (figure 1.2). These are photoreceptors (rods and cones), horizontal, bipolar, amacrine and ganglion cells. They are sub-grouped into distinct layers of cell bodies (the outer nuclear layer, inner nuclear layer and inner plexiform layer) and synaptic connections.

Photoreceptor axons synapse with either the bipolar or horizontal cells in the outer plexiform layer. Cone bipolar cells synapse either directly with ganglion cells or indirectly through amacrine cells. Rod bipolar cells terminate deep in the inner plexiform layer and always synapse with intermediary amacrine cells. The horizontal and amacrine cells modify and control the impulses and form laterally directed pathways between neurones. Interplexiform cells receive input from other amacrine cells and function to transmit signals back to the distal retinal layers, forming a feedback pathway. The last neuronal cell type in the retinal layer, the ganglion cells, receive their input from amacrine and

bipolar cells and have extensively spread dendrites so each ganglion cell maybe influenced by the activity of a large number of rods and cones. The axons of the ganglion cells join to form the optic nerve, which relays the signal to the brain. Müller cells are glial cells whose fibres extend and span through all the retinal layers to provide mechanical support for the retina. The retina is 32mm (from ora to ora the horizontal meridian) in size. The average dimension of the human eye is 22mm from anterior to posterior poles. The retinal pigment epithelium (RPE), which is a derivative of the CNS, is found adjacent to the retina and separates the retina from the choroidal circulation. Each RPE cell contacts about 45 outer segments of photoreceptors. The RPE is involved in the unique exchange of retinoid processing (visual cycle) with the photoreceptor cones and rods, which plays a critical role in visual pigment regeneration.

The main function of the retina is to detect and amplify photons of light via a mechanism known as phototransduction and to send this signal to the brain for visual processing.

1.3.4 The photoreceptor cells

There are two types of photoreceptor cell in the retina, rods and cones. These are highly specialised cells with a sensory organelle that absorbs photons of light by means of rod- or cone-specific photopigment molecules (opsin). In the mammalian retina, rod photoreceptors are more abundant with a ratio to cones of about 20:1. The distribution of rod and cone cells in the retina is not uniform. The fovea centralis contains no rods but is densely packed with ~200,000 cones. The cone density in the fovea is approximately 150,000 cones per mm². In the periphery, 10,000 rods maybe connected in clusters to a single nerve fibre. The density of cones in this region is reduced to 4,500 cones per mm². In contrast, the

density of rods is highest in the mid-periphery and decreases when approaching the centre.

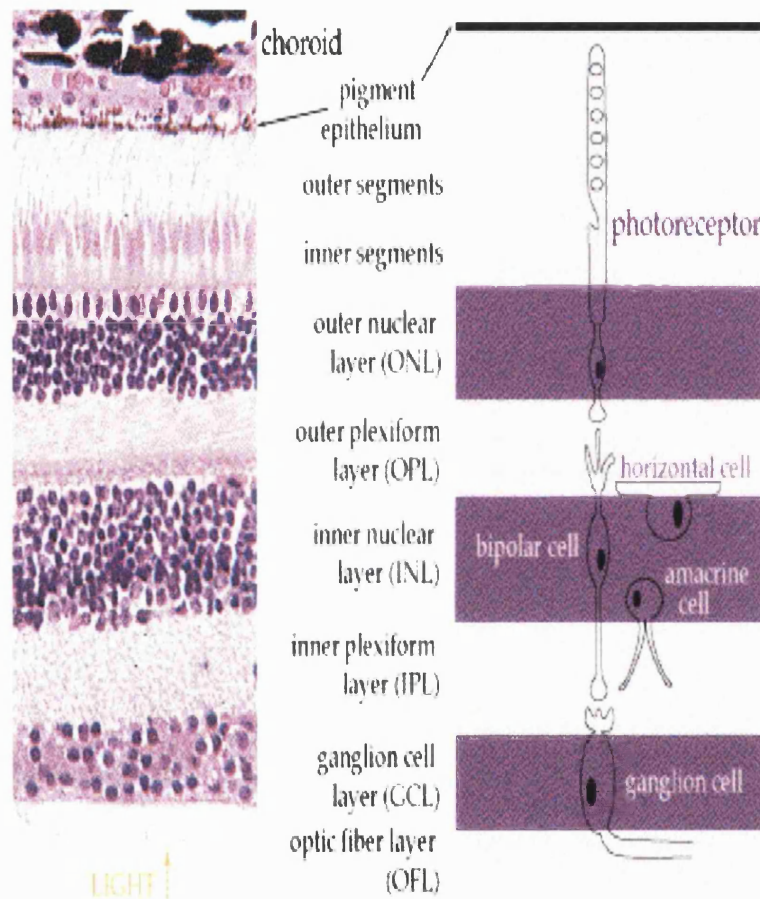


Figure 1.2 Cross section of the retina representing the different cell layers (left). On the right a schematic diagram showing the photoreceptor cell connection with other cells types within the retina (<http://thalamus.wustl.edu/course/eyeret.html>).

Morphologically, rods are thin and cylindrical whereas cones have a conical shaped outer segment. The outer segments accommodate a dense stack of disc-like membranes, which contain the visual pigment (opsin) molecules (Molday, 1998). In rods, the discs are closed, flattened sacs, which are formed by successive basal invagination of the outer segment plasma membrane (figure 1.3). Cone discs are different in that they form a continuous serrated membrane with the plasma membrane of cone outer segments (Nathans *et al.* 1986).

The visual pigment opsins are embedded in the membrane of these outer segment discs and comprise 80-90% of the total protein in the discs. These have a light sensitive chromophore, 11-*cis*-retinal, which is bound to the opsins. The rod visual pigment is rhodopsin. When the rod opsin is attached to the chromophore, it has a peak absorption (λ_{\max}) of ~500 nm and is responsible for contrast sensitivity and vision in dim light. There are three types of cone photoreceptors each containing a specific photopigment with absorption maxima in the red (560nm; long wavelength), green (530nm; medium wavelength) or blue (426nm; short wavelengths) (Nathans *et al.* 1986).

The inner segment of photoreceptors is joined to the outer segment by a slim immotile cilium. This functions to transmit cellular components from the inner segment to the discs. The inner segment is rich in mitochondria, ribosomes and golgi complexes providing the metabolic machinery necessary for active protein synthesis (Molday, 1998).

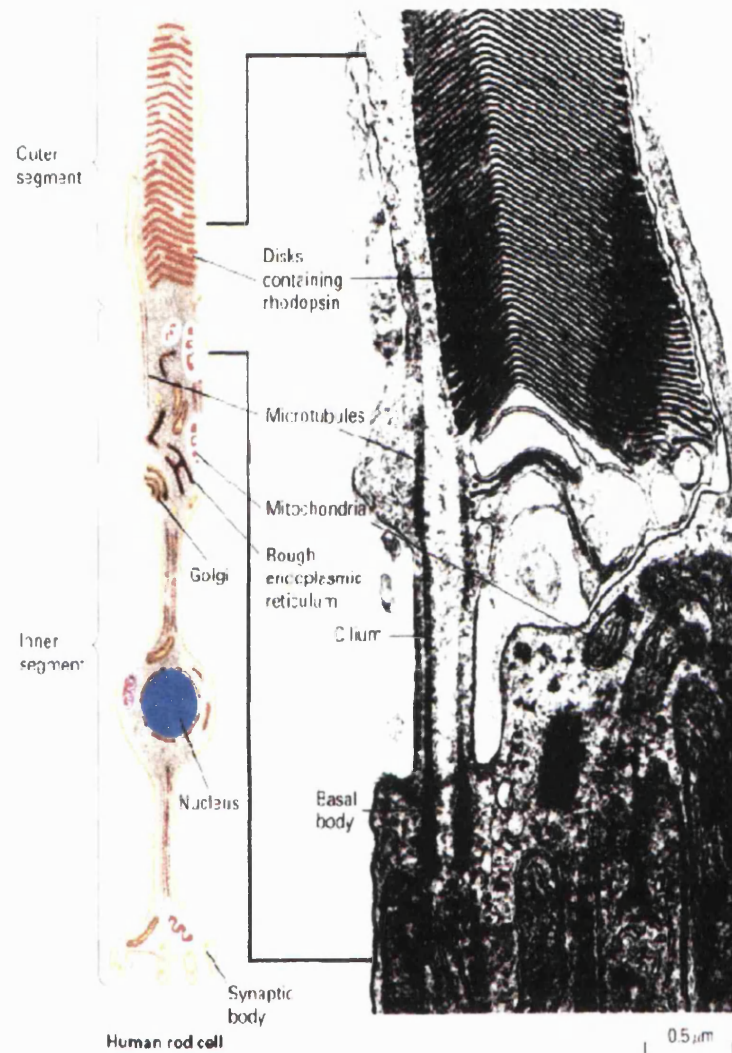


Figure 1.3 Cross-section through a human rod cell. The outer segment contains the photoreceptor molecule rhodopsin within the densely packed membrane stacks. The inner segment includes the rest of the cell body containing the nucleus and the synaptic body (<http://thalamus.wustl.edu/course/eyerod.html>).

1.3.5 Retinal pigment epithelium

The retinal pigment epithelium is a monolayer of highly differentiated cells consisting of about five million cuboidal cells. The RPE cells are important in excluding the exchange of potentially toxic substances between the choroidal circulation and the neural retina. It also has a role in the uptake, processing, transport and release of vitamin A (retinol) and some of the visual cycle intermediates (retinoids) (Shen *et al.*, 1994). The visual cycle is essential for the regeneration of *11-cis* retinaldehyde, the retinoid that serves as the chromophore for the visual pigment in the rod and cone outer segments (Shen *et al.*, 1994). The RPE is also involved in the absorption of scattered light by the melanin granules, transport of nutrients and metabolites through this extra retinal blood barrier, secretion of interphotoreceptor matrix (IPM) and maintenance of the photoreceptor microenvironment. In addition, the RPE has a role in the phagocytosis of the discarded tips of photoreceptor outer segment (Young and Bok, 1969). RPE cells have in fact been shown to engulf and digest the distal 10% of each outer segment daily, a total biomass equivalent to several average size cells. This process continues throughout life, making the RPE the most active phagocytic cell in the human body (Young and Bok, 1969; Shen *et al.*, 1994).

Different RPE specific or preferentially expressed genes have been found to be involved in different inherited retinal degenerations (Wu *et al.*, 1996) (<http://www.sph.uth.tmc.edu/retnet/disease.htm>).

1.3.6 Phototransduction pathway

The basic phototransduction pathway takes place in both rods and cones. However, there are rod-specific and cone-specific versions of the same proteins involved in the pathway. Phototransduction proteins and opsin molecules are all located in the outer segment of the photoreceptor cells (figure 1.4).

Phototransduction, also known as the visual cascade, is the process by which the energy of a photon is converted into a neuronal signal, which is eventually perceived as sight (figure 1.4). The phototransduction cascade begins with the photoactivation of the visual pigment (rhodopsin) and isomerisation of the *11-cis*-retinal of rhodopsin and dissociation into its *all-trans* isomer. The light activated rhodopsin goes through conformational changes that expose a heterotrimeric G-protein binding site, which permits the binding of photoreceptor G-protein transducin. The G-protein transducin is a peripheral membrane protein of the rod outer segment and is composed of three structural subunits α , β , and γ . The active form of transducin, the α subunit charged with GTP, dissociates from the $\beta\gamma$ subunits and is released into the cytoplasm. The α subunit of transducin activates cGMP phosphodiesterase (PDE). The PDE molecule is composed of α , β and two γ subunits. GTP molecules bind with the two inhibitory γ subunits, thus activating one PDE molecule. Activation of PDE results in hydrolysis of 3', 5'-cGMP to 5'-GMP. The cGMP-gated channel protein is composed of two subunits. The α and β subunits are located in the rod outer segment and are controlled by cGMP concentrations. Upon hydrolysis, the cytosolic concentration of cGMP decreases leading to closure of the cGMP-dependant cation channels on the plasma membrane of the rod outer segment and preventing the influx of sodium and calcium ions. This subsequent hyperpolarisation of the rod cell plasma membrane generates a nerve signal by reducing the level of neurotransmitter release from the rod synapse (Stryer, 1988).

Termination of the photoresponse occurs at several levels, including the phosphorylation of rhodopsin by rhodopsin kinase, the subsequent binding of phosphorylated rhodopsin by arrestin and by the hydrolysis of bound GTP by the intrinsic GTPase activity of transducin α subunit.

Defects in the majority of phototransduction cascade proteins have been known to cause one or more forms of retinal dystrophy (see section 1.5) (<http://www.sph.uth.tmc.edu/retnet/disease.htm>). In the following section, clinical and genetic description of retinal dystrophies and the role of proteins causing retinal diseases will be reviewed.

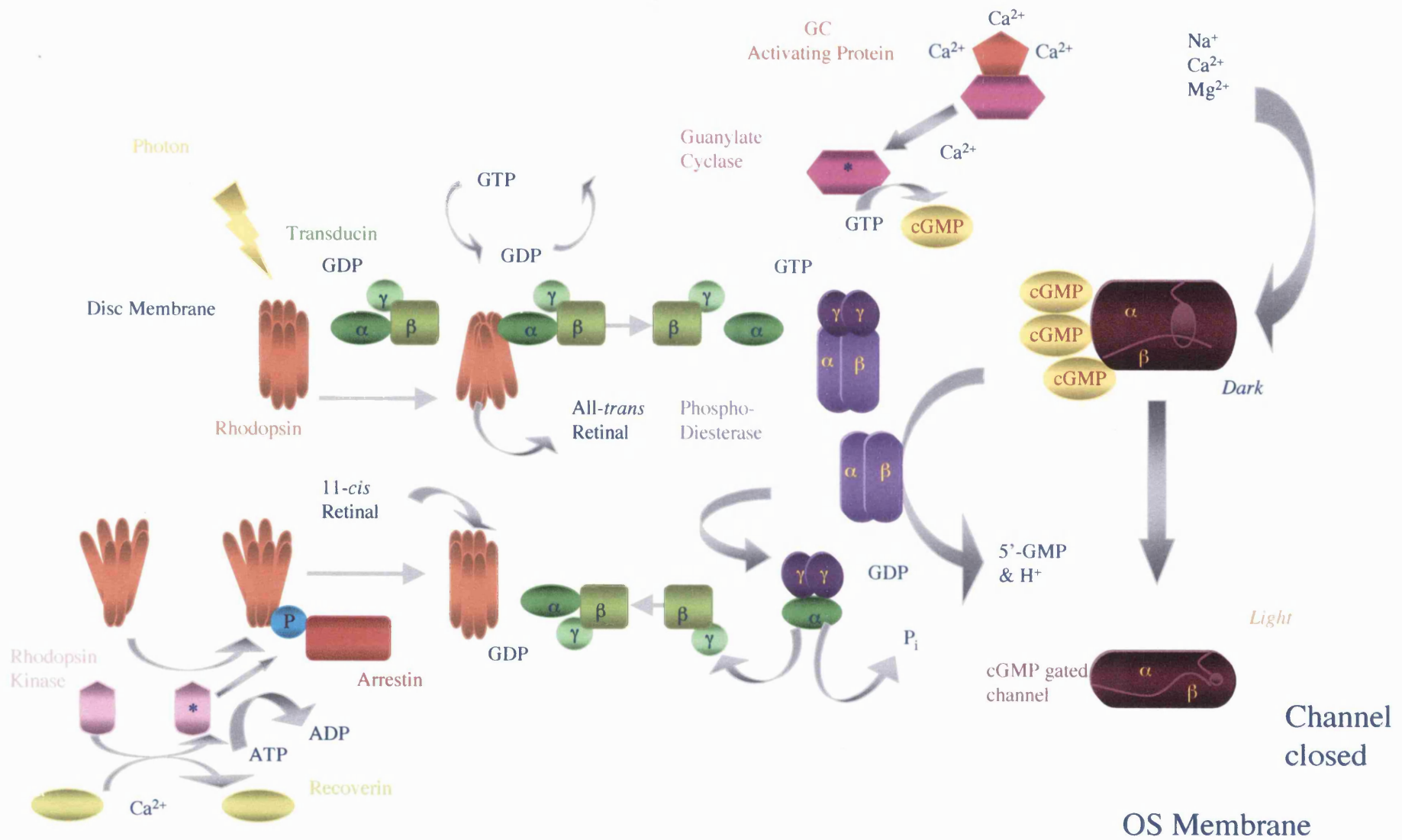


Figure 1.4 Schematic diagram of the phototransduction pathway (Dr Bart Leroy).

1.4 Retinal dystrophies

Retinal degeneration is the name given to a large number of inherited retinal dystrophies characterised by visual loss due to premature death of photoreceptor cells in the retina. This group of diseases are genetically and clinically heterogeneous affecting 1 in 3,000 individuals worldwide (Bundey and Crews, 1984; Sullivan and Daiger, 1996). To date, 131 retinal disease loci have been identified and mutations in 76 genes have already been identified to cause retinal dystrophies (<http://www.sph.uth.tmc.edu/retnet/disease.htm>) (Rivolta *et al.*, 2002). It has been suggested that the expected number of genes implicated in retinal diseases are likely to be two to three times higher than the present figure of mapped loci (Wright and Heyningen, 2001).

Clinical heterogeneity of retinal disorders is not well understood, however, the presence of genetic heterogeneity suggests that the degeneration of the retina is a common end point for many different biochemical abnormalities. The increase in understanding of genetic pathology of certain retinal dystrophies generates a rapid advance in the systematic approaches to classification of retinal diseases.

Retinal disease can be classified into two main classes; diseases of the peripheral retina, such as retinitis pigmentosa and congenital stationary night blindness; and diseases of the central retina, such as macular degeneration, cone dystrophy and cone-rod dystrophy.

Retinitis Pigmentosa (RP): this type of retinal degeneration describes a group of clinically and genetically heterogeneous inherited retinal disorders. RP is the most common form of hereditary eye diseases affecting almost 1 in every 5000 births in the Western world (Weleber and Evans, 2001).

Clinical examination of middle-aged patients with RP generally shows fundus changes that include a pale optic nerve head, attenuated retinal vessels and intra-retinal pigmentation. RP is also characterised by early night blindness, progressive loss of peripheral vision due to rod photoreceptor degeneration and changes to the optic nerve head during the course of the disease. Pigmentation observed in RP patients is the result of photoreceptor degeneration and exposing of the pigmented RPE cells (Carr and Heckenlively, 1986). These pigments accumulate in perivascular clusters (bone-spicule formations) in the neural retina. Cell degeneration results in a ring scotoma (blind or partially blind area) in the mid-periphery, which eventually encroach centrally leading to cone degeneration and often to complete blindness later during patient's lives. The age of onset of RP varies from infancy through to late middle age. Electroretinogram (ERG) abnormalities are recordable in the early stages of RP with attenuation of rod and cone responses. In advanced RP, both rod and cone ERG responses are extinguished (Carr, 1994).

Prior to molecular classifications, retinitis pigmentosa cases demonstrating an autosomal dominant pattern of inheritance (adRP) were previously classified into two main groups based on the clinical appearance of the fundus and pattern of photoreceptors, as type-1 or D-type and type-2 or R-type (Massof and Finkelstein, 1981; Lyness *et al.*, 1985). Type-1/ D-type adRP is more severe with diffused loss of rod function, which in the early stages co-exists with relatively normal cone function. Type-2 or R-type RP is more common and displays regional/patchy loss of rod photoreceptor function that is accompanied by loss of cone function. A rare class, known as sectorial RP, is characterised by retinal atrophy that occurs only in the lower nasal quadrant of the fundus. Patients with type 1/D-type disease usually exhibit night blindness before 10 years, while pigmentary changes and visual difficulties may not be detectable until 10-20 years after the onset of night blindness. Type 2-R patients have a variable age of onset. Some

type 2-R adRP pedigrees show a wide variation of severity, and the inheritance of the disease can appear to skip generations, a phenomenon known as variable or incomplete penetrance. *RP9* and *RP11* are examples of incomplete penetrance in RP.

Genetic classification of RP is mainly derived from their mode of inheritance. RP is inherited as monogenic (i.e. the disease is caused by a defect in one gene) or digenic disease (i.e. the disease is caused by two genes). Monogenic form of RP is classified into two subgroups. Non-syndromic RP where only the retina is affected by the disease, and syndromic RP in which other tissues as well as the retina are affected, such as Usher syndrome which combines RP and deafness. Non-syndromic RP can be inherited as an autosomal dominant (adRP), autosomal recessive (arRP) or an X-linked (xlRP) trait, which account for approximately 30%, 25% and 20% of all RP cases respectively. Sporadic RP or simplex cases accounts for almost 25% of all RP cases. RP in those patients could be the result of either novel dominant mutations, recessive, X-linked or digenic cases where a family history is unknown (<http://www.sph.uth.tmc.edu/retnet/disease.htm>).

Genes responsible for RP can be grouped into several functional classes. The majority of genes causing RP are expressed in the photoreceptor cells of the retina or in the RPE and/or participate in biochemical pathways specific to the retina. In addition, proteins involved in the phototransduction pathway and the vitamin A cycle in the photoreceptors and the RPE were found to cause retinal disorders (Table 1.1) (see section 1.5). At the start of this project, adRP was caused by mutations in 11 known genes, with two further loci implicated by linkage analysis, for which the gene remained to be identified (Table 1.2).

Table 1.1 Genes implicated in retinitis pigmentosa

Gene	Disease
Phototransduction cascade <i>(RHO) Rhodopsin</i> <i>PDE6α</i> <i>PDE6β</i> <i>CNGA1</i> <i>CNGB1</i>	adRP/arRP arRP arRP arRP arRP
Structural proteins <i>RDS (Peripherin)</i> <i>ROM1</i> <i>FSCN2</i>	Digenic RP/adRP Digenic RP adRP
Retinol (Vitamin A) metabolism <i>RLBP1</i> <i>RGR</i> <i>LRAT</i>	arRP arRP arRP
Enzymes <i>IMPDH1</i>	adRP
Transcription factors <i>NRL</i>	adRP
Splicing factors <i>PRPF3</i> <i>PRPF8</i>	adRP adRP
Proteins of unknown function <i>CRB1</i> <i>PROML1</i> <i>TULP1</i> <i>RP1</i> <i>RPGR (RP3)</i> <i>RP2</i> <i>PIM1K</i>	arRP arRP arRP adRP xLRP xLRP adRP

Table 1.2 Autosomal dominant RP genes and their chromosomal localisation (<http://www.sph.uth.tmc.edu/retnet/disease.htm>).

Chromosomal location	Symbol (OMIM number)	ADRP gene
Ch.1q21.1	<i>RP18</i> (601414)	<i>PRPF3</i>
Ch.3q22.1	<i>RP4</i> (180380)	<i>RHO</i>
Ch.6p21.2	<i>RP7</i> (179605)	<i>Peripherin/RDS</i>
Ch.7p15.1-p13	<i>RP9</i> (180104)	<i>PIM1K</i>
Ch.7q32.1	<i>RP10</i> (180105)	<i>IMPDH1</i>
Ch.8q12.1	<i>RP1</i> (180100)	<i>RP1</i>
Ch.11q12.3	<i>ROM1</i> (180720)	<i>ROM1</i>
Ch.14q11.2	<i>RP27</i> (162080)	<i>NRL</i>
Ch.17p13.3	<i>RP13</i> (600059)	<i>PRPF8</i>
Ch.17q22	<i>RP17</i> (600852)	?
Ch.17q25	<i>FSCN2</i> (-)	<i>FSCN2</i>
Ch. 19q13.42	<i>RP11</i> (606419)	?

Congenital Stationary Night Blindness (CSNB): CSNB is another non-progressive retinal disorder. CSNB is characterised by a decreased visual acuity and loss of night vision from early infancy. CSNB patients retain daytime vision throughout life. Genetic studies showed that CSNB is inherited as autosomal dominant, autosomal recessive and X-linked trait. To date, seven genetic loci for CSNB have been mapped; two of which are on the X chromosome (<http://www.sph.uth.tmc.edu/retnet/disease.htm>). Mutations in the genes encoding proteins of the phototransduction cascade as well as proteins of the recovery phase were found to be responsible for autosomal dominant and autosomal recessive forms of CSNB. Mutations in the calcium-channel α 1-subunit gene (*CACNA1F*) and the leucine-rich proteoglycan nyctalopin encoded

protein (*NYX*) were found to cause X-linked CSNB (Bech-Hansen *et al.*, 1998; Bech-Hansen *et al.*, 2000).

Leber Congenital Amaurosis (LCA): LCA is an early onset form of retinal degeneration inherited in a recessive manner. Infants are born blind or visually impaired, a condition confirmed with retinal electrodiagnostics. Studies estimated that LCA accounts for 10-18% of all congenital blindness (Fazzi *et al.*, 2003). For over 100 years physiological evidence has been gathered to find the causes of this heterogenic disease (Perrault *et al.*, 1999). To date, there are five genes whose mutations are known to cause LCA: *RetGC1*, *RPE65*, *CRX*, *AIPL1*, *RPGRIP1* and two other loci (LCA3, LCA5) for which the gene remains to be identified (see section 1.5) (Fazzi *et al.*, 2003).

Cone dystrophy: Patients with cone dystrophies lose all three types of cone photoreceptor cells while the rod cells remain functional. Hence, affected individuals suffer from photophobia, loss of visual acuity, colour vision and central visual field loss (Ripps *et al.*, 1987; Simunovic and Moore, 1998). Cone dystrophies are genetically heterogeneous. To date four loci have been mapped for cone dystrophies, 2 of which are inherited in an autosomal dominant manner while the other two loci were mapped to X-chromosome (<http://www.sph.uth.tmc.edu/retnet/disease.htm>). Mutations in open reading frame 15 of the *RPGR* gene has been shown to cause cone dystrophy on chromosome Xp11.4 (Yang *et al.*, 2002).

Cone-rod dystrophy (CRD): CRD is a severe form of chorioretinal dystrophy, which affects the central part of the retina, the macular region. This form of retinal degeneration is characterised by loss of colour vision followed by night blindness and loss of peripheral vision. Retinal pigmentation and chorioretinal atrophy of the central and peripheral retina are also manifested in CRD patients (Moore, 1992). As with other retinal degenerations, CRD is also genetically heterogeneous. To date 12 loci have been mapped for CRD inherited in autosomal dominant or autosomal recessive manner (<http://www.sph.uth.tmc.edu/retnet/disease.htm>).

1.5 Genes causing retinal dystrophies

The majority of the genes of the phototransduction cascade were found to be involved in at least one form of retinal degeneration. The most prominent example is the gene encoding rhodopsin. Apart from its importance in the phototransduction cascade, rhodopsin is also considered as an important structural protein of the rod outer segment localised in the disks and plasma membranes of rod outer segments (Nathans, 1992). Rhodopsin accounts for almost 50% of the protein content of the outer segments and 80% of proteins in the discs (Hargrave and McDowell, 1992). Involvement of Rhodopsin in RP was first reported by Dryja and colleagues in the early 90's and since then more than 150 mutations in rhodopsin have been reported to cause adRP, arRP and congenital stationary night blindness (CSNB) (Dryja *et al.*, 1990; Daiger *et al.*, 1995; <http://www.sph.uth.tmc.edu/retnet/disease.htm>). Most of the mutations identified in the rhodopsin gene are thought to affect either folding, stability or trafficking of the rhodopsin molecule.

Transducin, another important protein of the phototransduction cascade, was also found to cause retinal degeneration. The α subunit of

transducin (GNAT1) was found to cause autosomal dominant CSNB (Dryja *et al.*, 1996).

Mutations in α and β subunits of phosphodiesterase (PDE) were found to cause arRP. Moreover, mutations in PDE6 β were found to cause CSNB. Mutations in PDE are thought to directly affect the enzymatic activity of PDE enzyme (Farber, 1995). No disease causing mutations have yet been identified in the γ subunit of PDE.

Mutations in the membrane bound guanylate cyclase (RETGC1) were reported in Leber congenital amaurosis (LCA) and autosomal dominant cone-rod dystrophy patients (Perrault *et al.*, 2000; Duda *et al.*, 2000). RetGC encodes Retinal Guanylate Cyclase and converts GTP into cGMP which in turn enables the cGMP gated ion channel in the photoreceptor outer membrane to open again after light exposure (Perrault *et al.*, 1999). This mechanism does not occur in patients carrying this mutation, which results in a constant light-adapted state leading to improper photoreceptor cells development and/ or degeneration.

Other studies showed the involvement of the α and β subunits of cGMP phosphodiesterase in arRP (Dryja *et al.*, 1995; Bareil *et al.*, 2001). Guanylate cyclase activator 1A, a regulatory component of the phototransduction cascade was found to cause autosomal dominant cone dystrophy (Payne *et al.*, 1998).

Retinaldehyde binding protein (CRALBP) and RPE65, both proteins of the RPE were found to cause arRP (Gu *et al.*, 1997; Marlhens *et al.*, 1997). RPE65 was found to cause LCA (Morimura *et al.*, 1998). The 65kD protein thought to be involved in the conversion of *all-trans*-retinyl ester into *11-cis*-retinol in the RPE. Absence of a functional RPE65 protein leads to a constant dark-adapted state with no functional rhodopsin.

Cone-Rod Homeobox gene (CRX) is a photoreceptor specific gene encoding a transcription factor, which regulates the transcription of rhodopsin and a series of other genes. Mutations in this gene were found responsible for CRD and LCA (Freund *et al.*, 1997; Duda *et al.*, 2000). Moreover, the recently identified AIPL1 gene, which encodes an aryl-hydrocarbon interacting protein-like-1, was also found to cause LCA. This gene is involved in either protein folding and/or trafficking (van der Spuy *et al.*, 2002).

Proteins involved in the recovery phase in the phototransduction cascade were also found to cause retinal degeneration. Arrestin and rhodopsin kinase were reported to cause Oguchi disease; a rare form of autosomal recessive CSNB (Saari *et al.*, 1994).

1.6 A historical overview of the *RP11* locus

Mapping of the *RP11* locus was done by linkage analysis on a large British pedigree (ADRP5), which led to the localisation of the disease gene on chromosome 19q13.41 (*RP11*; MIM 600138) (Al-Maghtheh *et al.*, 1994). Since then, great effort was made to refine the *RP11* region. Six additional families mapped to the same locus (ADRP29, ADRP2, ADRP11, RP1907, and a Japanese family). Recombination events in the linked families have placed the *RP11* gene within a 3cM interval between the markers D19S927 (proximal) and D19S781.2 (distal) (Vithana *et al.*, 1999).

In an effort to clone the *RP11* gene, a DNA contig spanning the *RP11* interval composed of YAC, PAC, BAC and cosmid clones was established. This contig, which spanned a 3Mb genomic interval, contained 29 STSs from a wide range of sources, including 14 microsatellite markers, 3 expressed sequence tags (ESTs) and two known

genes, *PRKCG* and *RPS9*. PAC clones were isolated by screening the RPC11 PAC library (Ioannou and de Jong, 1996) with all the relevant sequence tagged sites (STS) from all high-resolution genetic and physical maps of chromosome 19q. The STS content of the region was increased by the isolation of novel microsatellite markers (e.g. D19S785, D19S781.1, D19S781.2 and D19S781.5) and non-polymorphic STSs generated by sequence sampling of PAC and cosmid clones. One of the novel microsatellite markers D19S781.2 (observed heterozygosity, 0.65) isolated from cosmid 30712 was used to refine the *RP11* interval in pedigree ADRP2 that originally showed the distal cross over with marker D19S418 (Al-Magthteh *et al.*, 1996). BAC clones of the *RP11* contig were obtained from the LLNL genome centre (http://wwwbio.llnl.gov/bbrp/genome/html/chrom_map.html), which was responsible for sequencing chromosome 19 in its entirety. Analysis of these BAC clones revealed that inserts of 6 BACs clones (3022G6: AC008753, 331H23: AC008440, 3093M3: AD012314, 402N14: AC009968, 158G19: AC022318 and 2337J16: AC010492) gave coverage over the entire *RP11* critical interval. From the insert size of these clones the *RP11* interval has been estimated to be approximately 500kb. Linkage and physical mapping and molecular analysis of candidate genes were the tools used to identify the *RP11* gene.

1.6.1 *RP11* families

Subsequent reports of British (Al-Magthteh *et al.*, 1996; Vithana *et al.*, 1998), Japanese (Xu *et al.*, 1995) and American families (McGee *et al.*, 1997) linked to this locus suggested *RP11* is perhaps as prevalent in the involvement of adRP as rhodopsin. Affected members from all these pedigrees have type II/ regional form of RP. Uniquely, they also show bimodal expressivity defined by the presence of asymptomatic individuals who have both an affected parent and affected children.

This all or none form of penetrance of *RP11* differs from the partial penetrance phenotype seen at the *RP9* locus where gene carriers exhibit a range of phenotypes from fully asymptomatic to severely affected (Kim *et al.*, 1995). The nature of the *RP11* phenotype suggests the possibility of a considerable proportion of small adRP families (that cannot be tested by linkage analysis) as well as many of the sporadic RP cases as being linked to the 19q locus. These factors further underline the fact that *RP11* could be a common and hence a very important RP locus.

1.7 Aim of the thesis

The aim of this thesis was to identify and characterise the *RP11* gene following a candidate gene approach. Initially the gene density of chromosome 19 was approximated at 1 gene per 50kb of genomic DNA (L.K. Ashworth, LLNL). However, sequence analysis from the region has indicated the *RP11* disease interval to be far more gene rich with an average density of one gene per 25kb of genomic DNA. Therefore, a systematic analysis of all the genes that map within the *RP11* genetic interval was undertaken.

Following the identification of the *RP11* gene, the genetic cause of the partial penetrance phenotype was pursued. It was previously proposed that the wild type allele of the *RP11* gene modulates the disease manifestation at this locus. Therefore, the role of the wild type allele in symptomatic and asymptomatic individuals was investigated.

CHAPTER 2

Materials Methods

2.1 Preparation of PCR templates

2.1.1 DNA extraction

DNA was extracted from blood samples using the Nucleon DNA extraction kit, Scotlab, UK, according to manufactures instructions. Briefly, blood samples were collected in 10ml EDTA tubes. Frozen blood samples were thawed at room temperature and transferred to 50ml sterile falcon tubes. Lysis of the red blood cells was performed by adding 40ml of solution A (see solutions) (nucleonTM, UK) then mixed by inverting 3-5 times, this was followed by centrifugation for 4 min at 3500rpm. The supernatant was discarded and the pellet was resuspended in 2ml of solution B. The mixture was then transferred into a 5ml screw-capped polypopylene centrifuge tube. 500µl of 5M sodium perchlorate was added and mixed by inverting the tubes seven times. 2ml of chloroform was added and mixed by inverting the tube, followed by 300µl of silica suspension (NucleonTM DNA extraction Kit, UK) and the tubes were centrifuged at ~3000g for 4 minutes. Precipitation of the DNA was performed by adding two volumes of absolute ethanol to the transferred supernatant. At this stage the DNA was visible as a thread like strand and was removed into a 1.5-ependrof tube using a sterile needle. The DNA pellet was left to air dry and prior to being dissolved in ~400µl of sterile

distilled water. Dilutions of 1/10 were kept at 4°C for routine usage and DNA stock was stored at -20°C.

2.1.2 RNA extraction

Total RNA extraction was performed on frozen cell pellets according to QIAGEN Rneasy[®] Midi protocol (Qiagen Ltd, UK). Cell lysis was performed by adding 2ml of RTL/ β -Mercaptoethanol to the cell pellet followed by a homogenizing step performed by passing the cell lysates 5-10 times through an 18-20-gauge needle. One volume of 70% ethanol was added to the homogenised lysate, which was then applied to the Rneasy midi column. The column was centrifuged at 4000 rpm for 5 minutes. The column was washed by adding 4ml of RW1 buffer and centrifuged at 4000 rpm for 5 minutes. 10 ml of buffer RPE was added to the column and centrifuged at 4000 rpm for 2 minutes. The column was washed again by adding 2.5ml of RPE buffer and centrifuged at 4000 rpm for 5 minutes. The RNA was then eluted from the column by adding 150 μ l of Rnase-free water to the silica-gel membrane. The column was then left to stand for one minute to maximise the yield of RNA eluted before being centrifuged for 3 minutes at 4000 rpm. RNA samples were then transferred into Rnase-free eppendorf tubes and stored at -70°C. Two RNA extractions were carried out from two independent cell passages from each cell line.

2.1.2.1 RNA quantification

The extracted RNA was quantified using two different methods. The first method was using a RNA chip and a bioanalyzer (Agilent Technologies, UK). The RNA6000 Nano chip was used to check for both quality and quantity of total RNA and mRNA. The Agilent bioanalyzer

software allows detection of total RNA as low as 5ng and up to 500ng. This method is useful for other applications such as northern blot analysis and total mRNA purification analysis of RT-PCR products. The RNA chip was prepared as recommended by the supplier. The gel mix was first prepared by adding 400µl of RNA gel mix to the spin filter and was centrifuged at 4000 rpm for 10 minutes. RNA dye was added to the gel mix and 9µl of the gel-dye mix was dispensed into RNA chip. The process was started by loading the chip with RNA 6000 Nano marker in all 12 wells. The next step was loading the RNA 6000 ladder in the ladder well and 1µl of the RNA samples in each of the 12 wells. The chip was then loaded on the Agilent2100 bioanalyzer. When the assay was completed the results were viewed as electropherograms, which shows the quality of RNA, and as figures of RNA.

The quantity of RNA was also measured using the eppendorf BioPhotometer (Agilent Technologies, UK). The spectrophotometer displays the concentration of the RNA sample and the absorbance at wavelength of 230, 260, 280 and 320nm and the ratios A_{260}/A_{280} as an indication of the purity of the nucleic acid samples. After calibrating the instrument with Rnase free water, 50µl of RNA sample was placed in 36mm Uvette[®] and placed in the cuvette shaft.

2.1.3 Total RNA extraction from retina and fresh blood samples

Retinal RNA was kindly provided by Dr. Dan Hornan (Moorfields Eye Hospital, UK) which was extracted using the following protocol. The donor eyes were obtained from the Eye Bank of British Columbia. Whole globes were placed in RNAlater (Ambion) after enucleation and corneal excision and stored at 4°C. The average time between death and dissection was about eight hours. Globes soaked in RNAlater were cut into segments and the retina detached from the RPE. Approximately,

10mm diameter sections of retina were removed with RNase free instruments and frozen in fresh RNAlater, or processed immediately. The Rnaqueous-4PCR kit (Ambion) was used to extract RNA according to the manufactures protocol, including DNase treatment. RNA was analysed for quantity and quality by gel electrophoresis. Nine RNA samples of equal quantity, from four donors (aged 52-64 years) were pooled to minimise inter-sample variation in gene expression.

Total RNA was extracted from 5-10 ml of fresh blood from 4 individuals of English Caucasian origin according to QIAGEN RNeasy® Midi protocol (Qiagen Ltd, UK) (see section 2.2.2).

2.2 PCR reactions

2.2.1 PCR amplification

Oligonucleotide primers for PCR reactions were synthesised by Sigma genosys and Bioline, UK. PCR reactions were carried out on ~100ng of template DNA. 50µl PCR reactions were performed which consisted of 1X PCR buffer (10X NH₄), 1.5 mM MgCl₂, 0.2 mM of each dNTP, 25 pmoles of each primer and 0.5 units of Taq polymerase. The temperature cycling profile consisted of denaturation for 1 min at 94°C, followed by 35 cycles of denaturation for 30 sec at 94°C, annealing for 30 sec and extension for 30 sec at 72°C. The reaction was completed with a final extension step for 4 min at 72°C. The annealing temperature was determined by the melting temperature(T_m) of the two primers used for amplification. The T_m was determined by the nucleotide sequence of the primer and was calculated according to the following equation: $T_m = 4(C+G) + 2(A+T)$.

2.2.2 Expression studies

Expression was assessed by PCR amplification of Quick-Clone™ human cDNAs (Clontech, UK) from retina, brain, heart, skeletal muscle, kidney, liver, colon and peripheral blood leukocytes, using the appropriate exonic primers. PCR products were then electrophoresed on a 1% (w/v) agarose gel and visualised by ethidium bromide staining. The ubiquitously expressed *PGM1* gene (Whitehouse *et al.*, 1992) was used as an amplification control.

2.2.3 Real Time RT-PCR amplification

Reverse transcription and PCR (RT-PCR) was carried out using the EZ one tube- buffer system (PE-Applied Biosystems, UK). 100ng of total RNA was used in a 25µl total volume of PCR mix. To prevent contamination of amplified DNA, the reaction was carried out in the presence of dUTP. RT-PCR thermal cycling was performed as follows. The RNA template was heated for 2 min at 50°C in the presence of 0.01U/µl AmpErase UNG (Applied Biosystems, UK). RT was performed for 30 min at 60°C and 5 min denaturing at 92. 40 cycles of PCR carried out. Cycling conditions were 20s at 92°C and 1 min at 62°C in the presence of the labelled probed. Following the target amplification, the probe annealed to the amplicon and was displaced and cleaved between the reporter and quencher dyes. The amount of product resulting in detectable fluorescence at any given cycle is within the exponential phase of PCR is proportional to the initial number of template copies. The number of PCR cycles (the threshold cycle, C_t) needed to detect the amplicon is therefore a direct measure of the template concentration. Threshold is defined as the point at which the fluorescence rises appreciably above the background fluorescence. Results were captured using SDS software from ABI and evaluated using Microsoft Excel. Copy

number calculations were performed by comparing detection values with values of accompanying standard curve.

2.2.3.1 TaqMan primers and probe

PCR primers and TaqMan probes specific for the target molecules were designed using Primer Express Software (Applied Biosystems, UK) and synthesised by MWG Biotech (Germany). PCR reactions were performed in the presence of three sequence specific oligonucleotides, these were forward and reverse primers (5'-AAGATGAAGGAGCGGCTGG-3', and 5'-CCTCCTGGTAGGCGTCCTC-3') respectively and the hydrolysis probe (5'CCGGAAGCAGGCCAACCGTATG-3'). The TaqMan hydrolysing probe is an oligonucleotide labelled with a reporter dye (carboxy-florescien, FAM) at the 5' end and a quencher dye (6-carboxy-tetramethylrhodamine, TAMRA) at the 3' end. PCR primers binding to exon 11 and 12 of *PRPF31* generate a 91bp PCR fragment while the probe hybridises to sequence in exon 11 encompassing the deletion in the mutant allele in the AD5 family used in this study. This will enable the amplification and detection of only the wild type allele. This assay would therefore enable the detection of the level of the two different wild type alleles *PRPF31* RNA (figure 2.1).

Taq DNA polymerase cleaves the TaqMan probe into fragments by its 5' to 3' endonucleolytic activity, and thus the reporter dye was separated from the quenching dye, resulting in an increase in fluorescence emission. The amount of the florescence detected by the ABI7700 reflects the amount of amplification of the mRNA molecule.

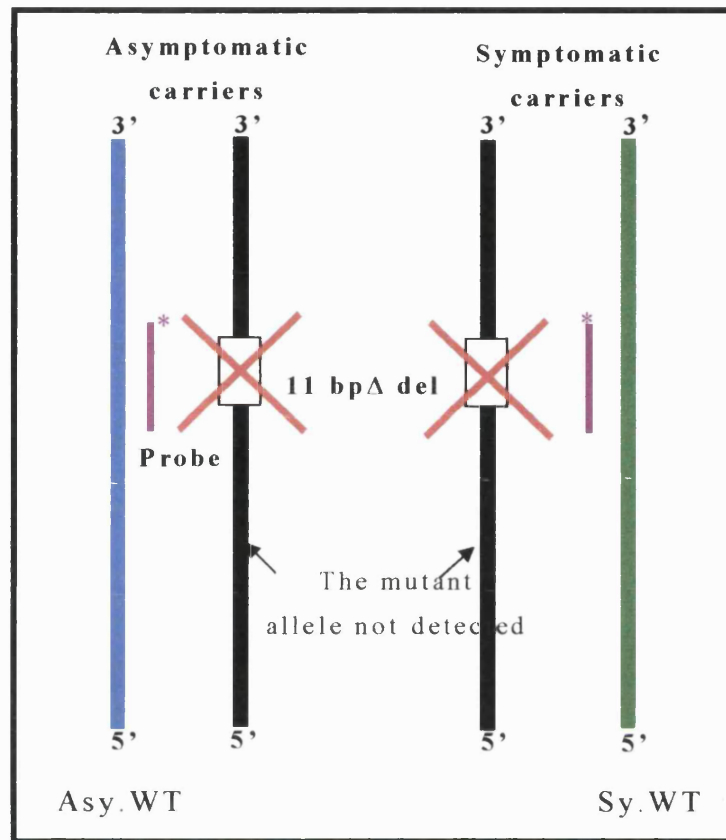


Figure 2.1 Schematic diagram showing the study design for quantitative RT-PCR.

2.3 Gel electrophoresis

2.3.1 Agarose gel electrophoresis

This method is used for separation of DNA fragments between 0.5-25kb. PCR fragments, enzyme digests as well as neat genomic DNA were run on agarose gels and visualised using ethidium bromide to a final concentration of 0.5µg/ml. Agarose gels were prepared to the appropriate concentration (table 2.1) and volume in 1X TAE. The mixture

was then heated until the agarose completely dissolved then was left to cool before being poured into a sealed loading tray which included a comb at one end of the tray. DNA samples were then prepared by adding 10X of loading dye (1µl to 9µl of sample) and loaded into the polymerised wells. Size standard (e.g. ϕ X) was always included in each run. The gel was then placed in an electrophoresis tank containing sufficient volume of 1X TAE buffer, with the wells positioned near the cathode end. Gels were visualised on UV trans-illuminator.

Table 2.1 Agarose concentration versus the range of resolution of linear DNA.

Agarose (%w/v)	Range of resolution of DNA (kb)
0.3	5.0- 60
0.6	1.0- 20
1.0	0.5- 10
1.5	0.2- 6.0
2.0	0.1- 2.0
3.0	0.05- <0.1

2.3.2 SDS-gel electrophoresis

Whole protein extract of cell lines was run on 10% (v/v) SDS-polyacrylamide gels. The glass plates (10x10cm) of the Bio-Rad protein electrophoresis system were assembled according to the manufacturer's instructions (Bio-Rad laboratories Ltd, UK). The resolving acrylamide gel solution was poured into the gap between the two glass plates leaving 1cm empty space for the stacking gel and left in a vertical position to polymerise for 30 minutes at room temperature. Once the resolving gel had polymerised, the stacking gel was poured on top and combs were

inserted immediately with care so as not to create air bubbles. While the stacking gel is polymerising, the samples were prepared by working out the volume necessary to get equal loading in μg of total protein on the gel (as a control for the experiment) and adding 2x loading buffer then denaturing the protein samples to 100°C for 3 minutes. Before loading the samples, the comb was removed and the wells were washed with 1x Tris-glycine running buffer. $15\mu\text{l}$ as a total volume of each sample was loaded into the bottom of the wells along with $5\mu\text{l}$ of rainbow marker (Amersham, UK). The gel was run first at 150 volts until the dye front moved into the resolving gel, then the voltage increased to 170 volts. The run finished when the bromophenol blue reached the bottom of the resolving gel. The glass plates were removed from the electrophoresis apparatus and separated using a plastic spatula. The gel was then either stained overnight with coomassie stain or was prepared for transfer on a nitrocellulose membrane and subsequent immunoblotting. Two identical gels were run at the same time; one used for coomassie staining to check for equal loading and the other for immunoblotting.

2.3.2.1 Coomassie staining and visualising protein on SDS gels

When the electrophoresis was completed, the gel was taken out from the glass cassettes and was stained overnight in the coomassie stain. De-staining with 10% acetic-acid helped to visualise the protein bands on the SDS-gel. Several washes with 10% acetic-acid were needed before the bands started to appear. Stained gels were dried using the Bio-Rad drier and scanned.

2.3.2.2 Transfer of proteins from SDS gels to nitrocellulose membranes

After separation of the protein on the SDS-gel, the gel was removed from the glass cassettes and incubated for 20 minutes at room temperature in the transfer buffer. The two sheets of pre-cut filter paper and the nitrocellulose membrane were cut to exactly the same size as the gel and were incubated in the transfer buffer as well. The gel membrane and the filter paper were placed on the blotter to allow transfer of the protein from the gel to the membrane. Transfer took place at 15V for 60 minutes. After transfer, the membrane was blocked at room temperature for 2 hours or over night at 4°C in a solution consisting of 5% (w/v) milk in PBS/0.05% (v/v) Tween. The membrane was then washed 3 times for 10 minutes each with PBS/Tween before it was incubated with the primary antibody for one hour at room temperature. *PRPF31* monoclonal antibody was raised in rabbit against a C-terminus peptide; amino acid residues 484-497 (Makarova *et al.*, 2002). The anti-61k antibody was used at 1:500 dilutions for western blotting. The membrane was washed again as above before it was incubated for an hour with the secondary anti-rabbit monoclonal antibody (Bio-Rad), which was used at 1:2000 dilutions. The membrane was washed as previously and the immunoreactive protein was detected using enhanced chemiluminescence (National Diagnostics, UK).

2.4 Mutation detection

2.4.1.1 Heteroduplex analysis- gel based analysis

Mutation detection by this method is based on the different migration rates between the heteroduplex and the homoduplex DNA strands since each adopts a different structural conformation.

Heteroduplexes occur when a normal DNA strand anneals with a mutant complementary DNA strand. These heteroduplexes migrate on acrylamide gels at a slower rate than homoduplexes. The electrophoresis apparatus (40cm X20cm glass plates with a 1mm spacer) was vertically assembled according to the manufacture's instructions (J.T Baker, USA) and clamped within a casting tray. The gel mix was poured between the glass plates avoiding the formation of air bubbles. Then the comb was inserted and clamped into place and the gel was allowed to polymerise for 1 hour. Prior to loading the PCR product (100-400bp) the comb was removed and the wells were rinsed with 1X TBE buffer. 10µl of the DNA sample containing 100-200ng PCR products in 2µl of sucrose loading buffer was loaded into each well. The gel cassette was mounted on the electrophoresis apparatus and sufficient 1X TBE was added to the upper and lower buffer container. Then electrophoresis was formed at 150volts for ~14 hours. The gels were stained with ethidium bromide (0.5µg/ml) and UV photographed.

2.4.1.2 Heteroduplex analysis- DHPLC WAVE^R Nucleic Acid Fragment Analysis System

Denaturing High Performance Liquid Chromatography (DHPLC) and direct sequencing was used for mutation detection. For DHPLC, PCR products were analysed using the WAVE^R Nucleic Acid Fragment Analysis System (Transgenomic, Inc., San Jose, CA). DNA fragment elution profiles were captured online and visually displayed using the Transgenomic WAVEMAKERTM software. Chromatograms were compared with those of normal controls to detect samples with altered elution profiles (figure 2.2).

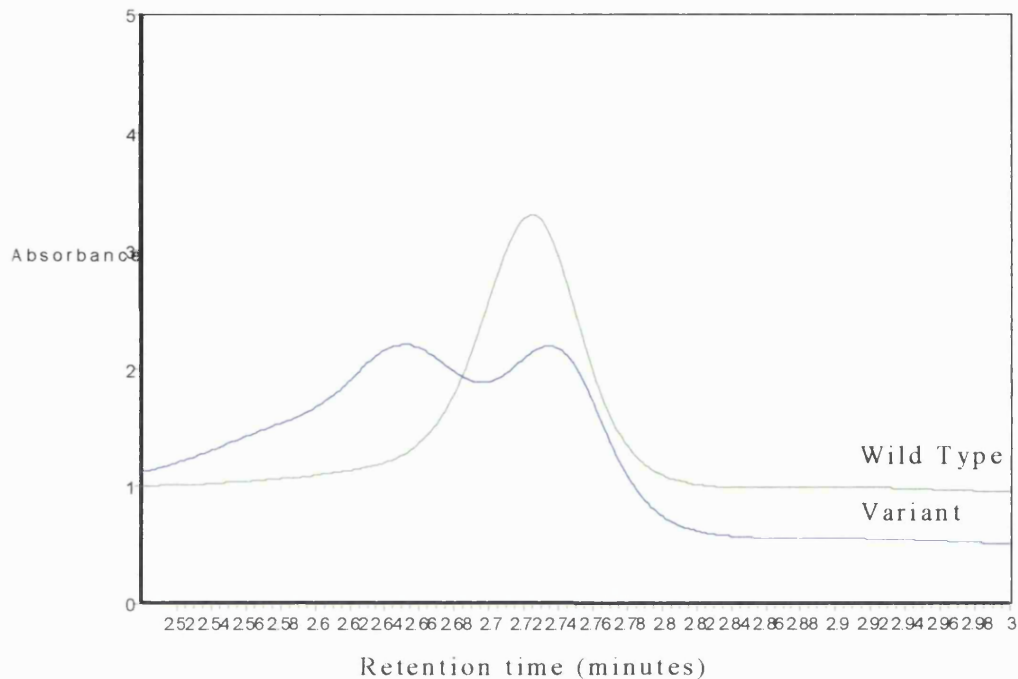


Figure 2.2 DHPLC elution profiles for the wild type and a variant sequence of exon 7 of *PRPF31* gene.

2.4.2 Direct sequencing

Prior to sequencing, PCR products were purified using QIAGEN PCR purification kit to remove unincorporated primers and dNTPs. Cycle sequencing was carried out using $\sim 1\mu\text{g}$ of the purified PCR product was used with $3.2\mu\text{M}$ of either the forward or reverse prime and $4\mu\text{l}$ of Big Dye mix (containing the fluorescent dye) mix in a $10\mu\text{l}$ reaction mix. The reaction was done using Perkin Elmer GeneAmp 9600 machine in 25 cycles of maximum temp of 96°C followed by 50°C for 5 sec and 60°C for 240 sec. DNA was then precipitated by 100% and 70% ethanol to remove excess fluorescent dye. Prior to loading the samples on the ABI sequencing gel, $3\mu\text{l}$ of the formamide/EDTA loading dye was added to

the pellet, mixed well, then denatured at 94°C for 3 min and snap frozen on ice. Sequencing was performed on ABI 373A with the XL adaptation or ABI3100 DNA sequencer (Applied Biosystems).

The 373a on 3100 DNA sequencers were run according to the manufacture's instructions. The gel was generally run for 13 hours or the appropriate time for the capillaries used on the 3100. A computer linked to the sequencer was programmed to analyse the data.

2.4.4 Restriction digest analysis

This method was applied to confirm changes found in the DNA sequences. Gene Work program was used to identify restriction sites in both the wild type sequence and the mutated sequence.

All restriction enzymes used in this project were obtained from Promega (UK). Restriction digests was carried out in a 20µl total volume on 1.5µg DNA, 0.5µl of 10u/µl enzyme, 2µl of 10X RE buffer and 0.2 µl of 10µg/µl of Bovine Serum Albumin to enhance activity. The reaction was incubated for 4 hours at 37°C, prior to being loaded on 2.5% agarose gel and visualised on a UV trasilluminator.

2.5 Northern blot analysis

Multiple tissue northern blots (Clontech, UK) were used to examine gene expression. The 155338 IMAGE clone was used as a probe for hybridisation with the multiple tissue membrane. The membrane contains purified poly A⁺ RNA from brain, heart, muscles, colon, thymus, spleen, kidney, liver, small intestine, placenta, lung and leukocyte. RNA was run on a denaturing 1% formaldehyde agarose gel and blotted onto a

positively charged nylon membrane. The membrane was hybridised with radioactively labelled probe.

Fresh [^{32}P] with an activity of 3000Ci/mmol for [α - ^{32}P] dCTP and a 6000Ci/mmol for [γ - ^{32}P] dATP was used to label the probe. Unincorporated [^{32}P]-labelled nucleotides were removed using the S400 columns (Promega, UK).

Radioactively labelled probe was denatured at 94°C for 4 min then snapped on ice quickly. 5ml of the radioactively labelled probe were added to fresh ExpressHyb. The container was closed and the solution was distributed evenly making sure that there were no air bubbles. The blot was then incubated at 68°C overnight. After the incubation period the membrane was washed with solution 1 several times at room temperature (RT). This step was followed by another wash with solution 2 for 3 hours at 50°C. the membrane was then removed using forceps and covered immediately with a plastic wrap. The membrane was exposed to X-ray film for 3 hours at -80°C and overnight at RT. Sometimes the membrane needed further stripping. The membrane was removed from the plastic wrap and placed into heated H₂O/0.5% SDS solution. The membrane was then incubated on a shaking tray for 3 hours at 50°C before the membrane was exposed autoradiograph.

2.6 Tissue culturing of transformed lymphocytes cell lines

A lymphocyte cell line from ADRP5 family deposited into the European Collection of Cell Culture (ECACC, UK). These were transformed into growing cell lines by immortalisation with Epstein-Barr virus (EBV).

Growing cells arrived in 25cm² tissue culture flasks filed with RPMI 1640+2Mm Glutamine +10% Fetal Bovine Serum (FBS). Cultures were kept in the 5% CO₂ 37°C incubator for 24 hours before sub-culturing with RPMI 1640 containing 10% FCS and 1% penicillin-streptomycin medium. Cell lines were sub-cultured every 2-3 days in order to maintain a healthy looking media of orange-yellow colour and cell number of 3x10⁵-2x10⁶. Cell counting was usually performed after 2 weeks of sub-culturing using the hemocytometer. Cells were harvested at the exponential phase of cell growth and pelleted by centrifuging the appropriate volume of cell culture; cell numbers of 5x10⁵-2x10⁶cells/ml, at 2500rpm for 5 minutes. Cells pelletes were washed with phosphate buffered saline (PBS) before they were stored at -20°C. Three cell pelletes were made of each sample containing approximately the same cell number. The first cell pellet was used for RNA extraction while the second one was used for protein extraction and the last one was stocked in liquid nitrogen.

2.7 Preparation of soluble cell extracts

Cell lysis was performed on frozen cell pellets prepared as mentioned in section 2.2.2. Cells were washed with PBS to remove any remains of the media that might interfere with the lysis process. The cell pellet was resuspended in 100µl of lysis buffer and incubated on ice for 10 min to lyse all membranes and maximise the amount of protein extracted. Homogenisation by repeated pippetting was also performed. The cell lysate was then centrifuged at 13000rpm for 15 minutes at 4°C, the resultant supernatant containing the soluble protein was preserved at -20°C until needed. Protein samples were quantified using Bradford protein assay using BSA to form the standard curve (Bradford, 1976).

Solutions

Sequencing gel (ABI373)

40ml Sequagel
10ml Sequagel complete buffer
reagent
0.04g Ammonium Persulfate
(APS)

Heteroduplex gel

50ml MDE
40ml dH₂O
6ml 10X APS
45µl TEMED
15g Urea

Sucrose loading buffer

40% Sucrose
0.25% Orange G dye
0.25% Xylene cyanol dye
0.25% Bromophenol blue dye

DNA extraction solutions

1- Solution A:
10mM Tris-HCl, pH8.0
320mM Sucrose
5mM MgCl₂
1% Triton (X-100)

2- Solution B:
400mM Tris-HCl, pH8.0
60mM EDTA
150mM NaCl
1% SDS

TBE

89mM Tris base
89mM Boric acid
2mM EDTA

Northern blot washing solutions

Wash Solution 1:
2X SSC
0.05% SDS

Wash Solution 2:

0.1X SSC
0.1% SDS

DHPLC buffers

Buffer A: 0.1M triethylammonium acetate (TEAA)
Buffer B: 0.1M TEAA with 25% acetonitrile.

Cell lysis buffer

20mM HEPES, pH 7.8
0.4mM EDTA
450mM NaCl
0.5mM DTT
0.5 mM PMSF

Coomassie Blue Stain for SDS-gel

200mg Coomassie Brilliant blue
160ml Ethanol
40ml Acetic acid
200ml water

SDS loading buffer

1M Tris (pH 6.8)
10% SDS
20% Glycerol
0.2% Bromophenol Blue
5% β-Mechaptoethanol

Transfer buffer, 1000ml

4g Tris base
14.4g Glycine
200ml Methanol

CHAPTER 3

Characterisation and mutation screening of an EST localised to the *RP11* critical interval

3.1 Introduction

Expressed Sequence Tags (ESTs) are short CND sequences usually derived from the 3' end of genes. EST mapping is a powerful and important tool used in the identification of genes that map within a given genetic interval. ESTs assist the construction of primary information about novel genes and are an immense resource for transcript analysis at the sequence level. They act as useful tags for isolating the full-length minas, from which they were derived and therefore help construct the complete sequence as well as identify alternate transcripts of novel genes. Moreover, ESTs give expression information as well as Exxon position within genes.

3.1.1 Mapping ESTs to the *RP11* interval: constructing a transcript map of the interval

In an effort to find the disease gene, the *RP11* physical connoting was used to fine map numerous ESTs localised to the 19q13.4 region by the gene-mapping consortium (Scholar *et al.*, 1996; Delouses *et al.*,

1998). In addition to the ESTs placed on the genemap, ESTs localised to 19q13.4 by NCBI genome centre (<http://www.ncbi.nlm.nih.gov/genemap/map.cgi?BIN=569&MAP=GB4>) and other major genome centres were also ascertained for analysis.

More than 100 ESTs were mapped to the 19q13.4 interval by radiation hybrid mapping. Therefore, the exact placement of these ESTs within the *RP11* critical interval needed to be determined prior to them being considered as positional candidates. Furthermore, the expression pattern of these ESTs also needed to be established.

ESTs chosen for mapping were transcripts expressed predominantly in the brain and retina. In total, 45 such ESTs were tested which led to the mapping of 6 ESTs within the *RP11* contig (WI-17997, A006I07, WI-15661, WI-22592, WI-15638 and SHGC13495). Only 3 of these ESTs (WI-17997, A006I07 and WI-22592) were subjected to further analysis as the rest were mapped outside the *RP11* critical interval. These 3 ESTs were re-tested for retinal expression by PCR amplification on retinal cDNA. Once retinal expression was confirmed, the cDNA clones (available through HGMP resource centre) of these ESTs were sequenced to obtain their full-length cDNA sequence.

The genomic organisations of the encoding genes were characterised utilising genomic clones that contained the given EST. Only WI-17997 was analysed to a considerable degree as the *RP11* gene was identified during the characterisation of these ESTs.

3.1.2 Characterisation in of the genomic sequence using NIX analysis

Nucleotide Identification X (NIX) is an electronic programme available through HGMP (<http://www.hgmp.mrc.ac.uk>) which incorporates and utilises programs such as BLAST (against nucleotide

and protein databases) Polyah, RepeatMasker, tRNAscan) and several exon prediction programs (e.g. GRAIL, Fex, Hexon, MZEF, Genemark, Genefinder, Fgene) for a comprehensive analysis of the genomic DNA, thus giving the opportunity to compare predicted exons and gene structure from several gene prediction programs. Despite all information given by NIX analysis, the information is only preliminary and needs to be confirmed by laboratory work. NIX is designed to make use of all the information available on the database. However, some of this information may be mis-described, poorly sequenced or annotated, which may lead to false positive or negative misleading predictions. For example, small exons could be missed in this analysis. Therefore, it is recommended that the only exons taken for further examination are the ones strongly predicted by one or more programs since having an exon predicted by different detection approaches makes it a more reliable result.

3.2 Results and discussion

3.2.1 Mapping of WI-17997 EST

Mapping of WI-17997 EST within the *RP11* contig indicated that it is present in five genomic clones (figure 3.1), all of which also contained the *PRKCG* gene, previously excluded from involvement in adRP (Al-Maghtheh *et al.*, 1998). As WI-17997 was one of the few ESTs, which mapped to the interval, placed between D19S572 and D19S785, it was fully characterised and subsequently screened as a candidate for *RP11*.

UniGene cluster (Hs 250700) for EST WI-17997 included more than 300 cDNA clones from a variety of libraries, indicating the ubiquitous expression of the gene (Table 3.1). UniGene is a system for automatically partitioning GenBank sequences and hundreds of thousands of novel EST sequences into a non-redundant set of gene-oriented clusters. Each

cluster contains sequence, which represents a unique gene and provides expression and mapping information of that gene. In order to obtain the full sequence for this gene, 9 IMAGE clones of different sizes and derived from a variety of tissues including brain were sequenced. Vector primers (M13 forward and reverse) as well as several internal primers were used to obtain the entire sequence of these clones. The collated sequence is 2.2kb in size (figure 3.2).

Table 3.1 The cDNA IMAGE clones for EST WI-17997 (UniGene cluster -Hs 250700)

IMAGE clone ID	Insert size	Accession number	Derivative tissue
531448	2.1kb	AA075868	Brain
155338	1.7kb	R69484	Breast
1523203	1.1kb	AA909044	-
1086010	1.0kb	AA582143	Germ Cell
1552077	1.0kb	AA582143	Germ Cell
1703512	0.8kb	AI602211	Brain
1572322	0.8kb	AA938008	Kidney
1561787	0.6kb	AA953018	
1694786	-	AI093195	Heart
1253927	-	AA938468	Prostate

All cDNA sequences were compared for evidence of alternative splicing, but no splice variants were detected for this gene. The largest IMAGE clone sequenced (insert size 2.2kb) (531448:AC AA075868) contained the longest open reading frame (ORF) corresponding to a putative protein of 322 amino acids. This clone was used for further characterisation of the gene represented by WI-17997 EST.

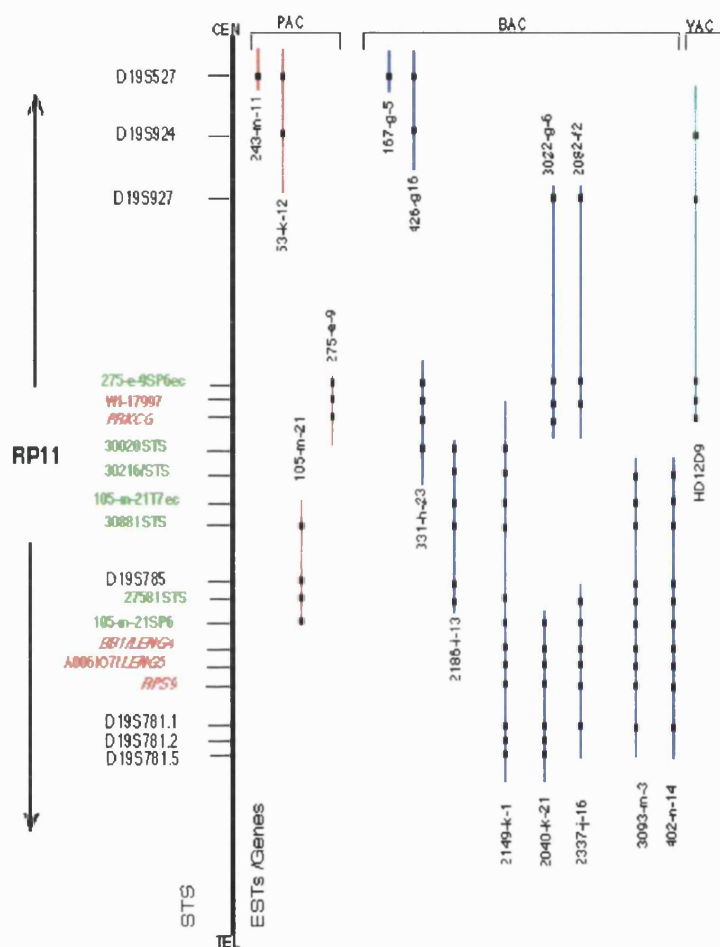


Figure 3.1 Physical map of the *RP11* critical region showing ESTs *W1-17997* and *A006107*. Genetic markers STSs from cosmids and known genes mapping within the refined interval *D19S927* and *D19S781.2* were used to construct the physical map based on PAC, YAC and BAC clones. Only few of the clones of the contig are presented. Clones being sequenced by LLNL are drawn in red.



Figure 3.2 The collated sequence of the transcript represented by EST WI-17997. The 5' UTR, the entire coding sequence and the 3' UTR region are indicated. The start and the stop codons are in red. Arrows represents primers used for sequencing the cloned insert to get the entire cDNA sequence.

3.2.2 Expression of WI-17997 EST

A pre-made Northern blot containing 1µg of polyA⁺ RNA lane from different human tissues was used for Northern blot analysis of W-17997 EST (Clontech, USA). The 1.7kb insert of IMAGE clone 155338 (AC R69484) was used as a probe. Northern blot analysis confirmed the presence of one transcript of approximately 2.2kb in all of the tissues analysed (figure 3.3), in addition that the full-length transcript is similar in size to the largest cDNA clone sequenced.

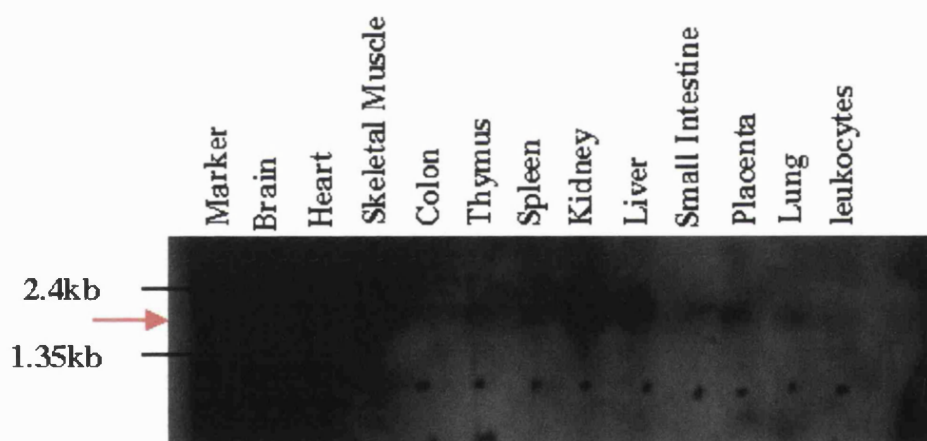


Figure 3.3 Northern blot analysis of WI-17997. One transcript of approximately 2.2kb is shown in all of the tissues analysed (as indicated by the **red** arrow). The Northern blot was probed with the (P^{32}) radiolabelled 1.7kb insert of IMAGE clone 155338 (AC R69484).

3.2.3 Protein identities and possible function

Hypothetical protein BC013995 (322 amino acids) represented by EST WI-17997 shows homology of 81%, over aligned region of 244 amino acids to a mouse protein encoded by a gene called *Myadm* (Myeloid-associated differentiation marker) (Pettersson *et al.*, 2000). This predicted 32-kDa protein contains multiple putative hydrophobic transmembrane segments and has several potential consensus sites for phosphorylation (figure 3.4).

Unlike the human gene, *Myadm* exhibits a restricted expression pattern within the hematopoietic system. In view of this expression pattern, *Myadm* has been suggested to serve as a new marker for hematopoietic differentiation. The ubiquitous expression pattern of the human gene is however suggestive of it having additional functions besides a possible involvement in myeloid differentiation.

Moreover, protein similarities were also identified with *H.sapiens* mucin protein DKFZp434B0635.1 (pir: T345949) (26%/ 341aa), *R.norvegicus* plasmolipin (pir: A55046) (25%/ 159aa), *A.thaliana* glycine-rich protein (pir: T49100) (29%/ 188aa) and *C.elegans* hypothetical protein T05C12.10 (pir: T23754) (24%/ 277aa) for each of the homologies percent identities and length of the aligned region are shown (figure 3.4).

3.2.4 Genomic organisation of hypothetical gene LOC91663

The genomic organisation of this novel gene, now identified as a hypothetical gene LOC91663, was established by aligning the collated cDNA sequence (see section 3.2.1) with the genomic sequence of the BAC clone 331H23 using the BLAST program to identify exon/intron boundaries (Pearson and Limpman, 1988; Altschul *et al.*, 1990).

LOC91663 was found to consist of 2 exons, 86bp and of 924bp in size respectively for exon 1 and 2. Translation was initiated and terminated within the second exon. The two exons follow the GT/AG rule for exon/intron boundaries. The gene spans 8.6kb of genomic DNA sequence (figure 3.5). Based on the genomic sequence, the stop codon of LOC91663 was found to be approximately 9kb far from the start codon of *PRKCG* (Al-Magthteh *et al.*, 1998). The genomic organisation of the hypothetical LOC91663 gene was further aided by the analysis of the genomic sequence of 331H23 by NIX program.

Human BC013995	1	MPVTVTRTTITTTTTSSSGLGSPMIVGSPRALTQPLGLLRLLQLVSTCVAFSLVASVGAW	60
		MP T TT ++ S VGS RALTQPLGLLRLLQL+STCVAFSLVASVGAW	
Mouse mye.upprtn		MP-----VTVTRTTITTTTTSSSTTVGSARALTQPLGLLRLLQLISTCVAFSLVASVGAW	54
Human BC013995	61	TGSMGNWSMFTWCFCFSVTLIIILIVELCGLQARFPLSWRNFPI TFACYAALFCLSASIIY	120
		TG MGNW+MFTWCFCF+VTLIIILIVEL GLQA FPLSWRNFPI TFACYAALFCLS+SIY	
Mouse mye.upprtn	55	TGPMGN <u>WAMFTWCFCFAVTLIIILIVE</u> LGGLQAHFPLSWRN <u>FPITFACYAALFCLSSSIY</u>	114
Human BC013995	121	PTTYVQFLSHGRSRDHAI AATFFSCIA CVAYATEVAWTRARPGEITGYMATVPGLLKVLE	180
		PTTYVQFL+HGR+RDHAI AAT FSC+AC+AYATEVAWTRARPGEITGYMATVPGLLKV E	
Mouse mye.upprtn	115	<u>PTTYVQFLAHGRTRDHAI AATTFSCVACLAYATEVA</u> WTRARPGEITGYMAT <u>VPGLLKVFE</u>	174
Human BC013995	181	TFVACII FAFISDPNLYQHQP ALEWC VAVYAICFILAAIAILLNLGECTNVLPIPFPSFL	240
		TFVACII FAFI +P LY +PALEWC VAVYAICFILA + ILLNLG+CTNVLPIPF+FL	
Mouse mye.upprtn	175	<u>TFVACII FAFIGE</u> PLLYNQKPAL <u>EWCVAVYAICFILAGVTILL</u> NLGDCTN <u>VLPIPFPTFL</u>	234
Human BC013995	241	SGLA	244
		SGLA	
Mouse mye.upprtn	235	<u>SGLA</u>	238

Figure 3.4 Alignment of human hypothetical protein BC013995 against the mouse *Myeloid* up-Regulated protein. Amino acids highlighted in red represents the mouse transmembrane regions.

ACTAGTTCTC TCTCTCTCTC TCCGTCTTTG CTTAGCCGC AGTCGCCACT GGCTGCCTGA
 GNGTGAGGTT GTGGGGACCC CAGAGGGCAT AAGCAACAGT CTAGGGTCTT TATGGCCAGA
 GAGAGGAATG GGCTGGGTCT CAGAATGTCT GGGTCCCTAG AAAGACTGGG GCTGGGCCAG
 GATTCTTGGG GTTCTGCAGA GGAGAGGGCT GAGGACCTGG ACTGTGGAGG TCATACGTCT
 GGAAGGGGGA TCCTCTAGAG TCGACCAACA GCATGCAGCT GCCTGCTG //
 ATGAGGTCAC TCTAGAGGTA TCCCCAGCCA AACAGTTAAC TGAACGGGGA GGGGAGGGGG
 CGGGGCTTAC CCGGGGTCAC CGCCAGGCCC TGCTGGGTTC CACGCCACGT TCCTGTCTCC
 GCAGGGCGTT CGATCCCTAG AGGGAGGAGC CTGTCCAAAC GGACGCTAAC GGCCTACCTC
 CCCCTCAGGT GCTCTTACAG CCTGTTCCAA GTGTGGCTTA ATCCGTCTCC ACCACCAGAT Exon 1
 CTTTCTCCGT GGATTCTCTT GCTAAGACCG CTGAGTGAGT AGGCAGCGGG CCTGGACTCC
 GGGGTCCGAG GGAGGAGGGG CTGGGGGCTT GGACTCCGGG GTCCGAGGGA GGAGGGGCTG
 GGGGCTTGGG CTCCTCGGTC CGAGGGAGGA GGGGCTGGGG GCCTGGGGGC CTGGGCTCCT
 GGGTCCGAGG GAGGAGGGGC TGGGGGCTTG GGCTCCTGGG TCCGAGGGAG GAGGGGCTGG
 GGGTCTGGAC TCCGGGGTCC GAGGGAGGAG GGGCTGGGGT CTGGACTCCT GGGTCCGAGG
 AAGGAGGGGC TGGGTCTTGG ACTCCTGGGT CCGAGGAAGG AGGGGCTGGG GGCCTGGGCT
 GCTGGGTCCG AGGGAGGAGG GGCTGG // GGGT GACTAGAGGT TCCAAACGAA
 AGCCCCATC TGGGAACCTC CTAATTAGAG GTCATATTAA TTTGTCCCTG GGGTAGGCAA
 TGGCTAAGGG ACCTGGATTG TTATGTTTGT GACTGTATAA GGACACTGTC TTTCCCTTTT
 TTGCACTCA TGGCAGTGAC GGTAACCCGC ACCACCATCA CAACCACCAC GACGTCTCTT
 TCGGGCCTGG GGTCCCCAT GATCGTGGGG TCCCTCGGG CCCTGACACA GGGGCTGGGT
 CTCCTTCGCC TGCTGCAGCT GGTGTCTACC TGGTGGGCTT TCTCGCTGGT GGCTAGCGTG
 GGCCTCTGGA CGGGGTCCAT GGGCAACTGG TCCATGTTCA CCTGGTGCTT CTGCTTCTCC
 GTGACCTGA TCCTCTCAT CGTGGAGCTG TGGGGGCTCC AGGCCCGCTT CCCCTGTCT
 TGGCGCAACT TCCCCATCAC CTTGCGCTGC TATGCGGCC TCTTCTGCCT CTCGGCCTCC
 ATCATCTACC CCACCACCTA TGTCCAGTTC CTGTCCACG GCCGTTGCGG Exon 2
 ATCGCGCCA CCTTCTTCTC CTGCATCGCG TGTGTGGCTT ACGCCACCGA AGTGGCCTGG
 ACCCGGGGCC GGCCCGGCGA GATCACTGGC TATATGGCCA CCGTACCCGG GCTGCTGAAG
 GTGCTGGAGA CCTTCGTTGC CTGCATCATC TTCGCGTTCA TCAGCGACCC CAACCTGTAC
 CAGCACCAGC CGGCCCTGGA GTGGTGCCTG GCGGTGTACG CCATCTGCTT CATCCTAGCG
 GCCATCGCCA TCCTGCTGAA CCTGGGGGAG TGCACCAACG TGCTACCCAT CCGCTTCCCC
 AGCTTCCTGT CGGGGCTGGC CTTGCTGTCT GTCTCTCTCT ATGCCACCGC CTTGTCTCTC
 TGGCCCTCT ACCAGTTCGA TGAGAAGTAT GCGGCCAGC CTCGGCGCTC GAGAGATGTA
 AGCTGCAGCC CCAGCCATGC CTACTACGTG TGTGCTGGG ACCGCCGACT GAGCTGTGGC
 ATCCTGACGG CCATCAACCT ACTGGCGTAT GTGGCTGACC TGGTGCATCT TGCCACCTG
 GTTTTGTGCA AGGTCTAAGA CTCTCCCAAG AGGCTCCCCT TCCCTCTCCA ACCTCTTTGT
 TCTTCTTGCC CGAGTTTCTT TTATGGAGTA CTTCTTTCTT CCGCCTTTCC TCTGTTTTCC
 TCTTCTGTC TCCCCTCCCT CCCACCTTTT TCTTCTCTTC CCAATTCTT GCACCTAAC
 CAGTTCTTGG ATGCATCTT TCCTTCCCT TTCTCTTGC TGTTCCTTC CTGTGTTGTT
 TTGTTGCCCA CATCTGTTT TCACCCCTGA GCTGTTTCTC TTTTCTTTT CTTTCTTTT
 TTTTTTTTTT TTAAGACGGA TTCTACTCTT GTGGCCAGG CTGGAGCGCA GTGGTGCAT
 CTCGACTCAC TGCAACCCCC GCCTCCTGGG TTCAAGCGAT TCTCCTGCCC CAGCCTCCCA
 AGTAGCTGGG AGGACAGGTG TGAGTGCCGC ACCAGCCTG TTTCTCTTT TCCACTCTC
 TTTTCTCTCA TCTCTTTCT GGGNTTCTGT CGGCTTTCTT ATCTGCCTGT TTTGCAAGCA
 CCTTCTCCTG TGTCTTGGG AGCCCTGAGA CTTCTTTCTC TCCTTGCCCT CACCCACCTC
 CAAAGGTGCT GAGCTCACAT CCACACCCCT TGCAGCCGTC CATGCCATCA GCCCCAAGGG
 GCCCCATTGC CAAAGCATGC CTGCCACCC TCGCTGTGCC TTAGTCAGTG TGTACGTGTG
 TGTGTGTGTG TGTGTTGGGG GTGGGGGGTG GGTAGCTGGG GATTGGGCCC TCTTCTCTCC
 AGTGGAGGAA GGTGTGCAGT GTACTTCCCC TTAAATTAA AAAACATATA TATATATATA
 TTTGGAGGTC AGTAATTTCC AATGGGCGGG AGGCATTAAG CACCGACCCG TGGTCTCTA
 GGCCCGCCTT GGCATCAGC CTTGCCAGAG ATTGGCTCCA GAATTTTTGC CAGGCTTACA
 GAACACCCAC TCCTAGAGG CATCTTAAAG GAAGCAGGG CTGGATGCCT TTCATCCCA
 CTATTCTCTG TGGTATCAAA AAGAAAAAAG AAAAAAAG AAGGAGTCGG GGCCGGGCGT
 GGTGGCTCAC GCCTGTAATC CCAGCACTTT GGGAGACCAA GTCAGGCAAT CATCTGAAGT
 CAGGAGTTCA AGA

Figure 3.5 Genomic sequence of Hypothetical gene LOC91663. Both the start and the stop codon in the last exon are highlighted in red. Exonic sequence of the gene are highlighted in blue. Coloured arrows represent primer pairs used for amplifying and sequencing overlapping PCR fragments from exon 2, the coding sequence of the gene .

3.2.5 NIX analysis of 331H23 genomic sequence to obtain genomic organisation of the putative gene LOC91663

NIX analysis was performed on the partial genomic sequence of the BAC clone 331H23, which was known to contain EST-17997 to identify the genomic structure of this putative gene. NIX predicted the presence of strong promoter motifs upstream of the first exon of the predicted gene (TSSW/ Promoter), shown to consists of two exons (GRAIL, Genefinder, Fex,Hexon) (figure 3.6). A strong poly adenylation signal/motif was also predicted in the sequence analysed (GENESCAN/poly A, figure 3.6). There were several exons predicted 5' to LOC91663, however, these exonic sequences were not part of the known transcript of LOC91663 (figure 3.2) indicating these exons to be part of another gene or that the mRNA sequence of LOC91663 is incomplete. The latter possibility is unlikely given that the Northern blot analysis indicated the LOC91663 transcript to be less than 2.4kb (figure 3.3). Furthermore, the location of the promoter motif indicates that the first exon (as indicated in figure 3.6) is the initial exon of this gene.

3.2.6 Mutation screening in *RP11* linked families

LOC91663 was subjected to mutation screening in *RP11* linked families by heteroduplex analysis and sequencing. The disease status of all individuals enrolled in this study had been established in clinical examinations at Moorfields Eye hospital, London, UK. Overlapping primers were designed to enable mutation screening of the entire coding sequence, intronic primers were also designed to screen the un-translated first exon (figure 3.5).

Screening of the coding second exon of this gene failed to identify any disease causative mutations. Moreover, pathogenic mutations were

CHAPTER 4

Identification of the *RP11* gene

4.1 Introduction

In the construction of the *RP11* physical map it became apparent that the inserts of 5 BAC clones, all sequenced by LLNL (http://wwwbio.llnl.gov/bbrp/genome/html/chrom_map.html), cover the entire *RP11* candidate region. These overlapping BAC clones, along with YAC and PAC clones were used to position genes and ESTs on chromosome 19q, which included the disease interval for *RP11* (figure 4.1). The emerging and often unannotated sequence of these BAC clones were then analysed using gene identification programs to identify putative candidate genes for *RP11* (see section 1.2.2).

The main program used for such gene identification was NIX, available through the Human Genome Mapping Project (HGMP) Resource Centre web site (<http://www.hgmp.mrc.ac.uk>) (see section 3.2.5). In addition to NIX, Ensemble, a joint project between EMBL-EBI and the Sanger Centre that produces and maintains automatic annotation on eukaryotic genomes was also utilised to obtain information on putative human genes on chromosome 19 (<http://www.ensemble.org/>).

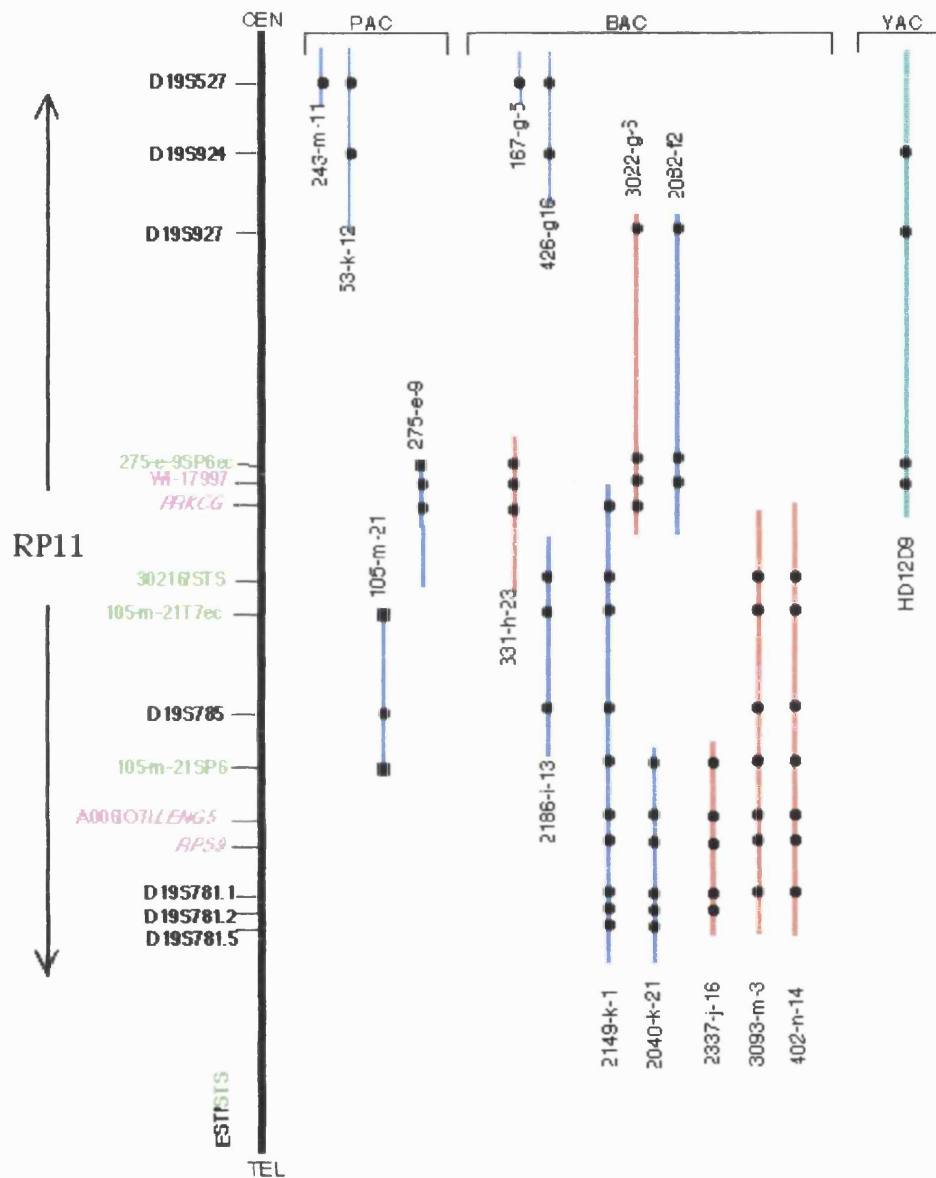


Figure 4.1 Physical map of the *RP11* critical interval. Fully sequenced clones are drawn in red. ESTs and genes identified in the region are in pink. Polymorphic markers and STSs are in black and green respectively.

4.2 Towards identification of the *RP11* gene

The analysis of BAC sequences led to the identification of several potential candidate genes including two known genes, *PRKCG* and *RPS9*, both of which were previously excluded from involvement in *RP11* (Al-Maghtheh *et al.*, 1998; Vithana *et al.*, 1999). The predicted genes were initially tested for expression in retina, RPE, brain and blood by RT-PCR, an example of expression profile on two genes identified within the *RP11* critical interval is shown in figure 4.2. The negative result in the brain is probably due to RNA degradation of that sample and not to a lack of expression in this tissue. Later analysis showed that *PRPF31* is expressed in brain tissue (see figure 4.4). Approximately 21 genes were localised to the *RP11* critical interval (figure 4.3). However none showed an exclusive expression in the retina. During this work, the strategy to systematically characterise the *RP11* locus, define the genomic organisation of all the genes identified in the *RP11* critical region and screen those for pathogenic mutations in the linked families was adopted.

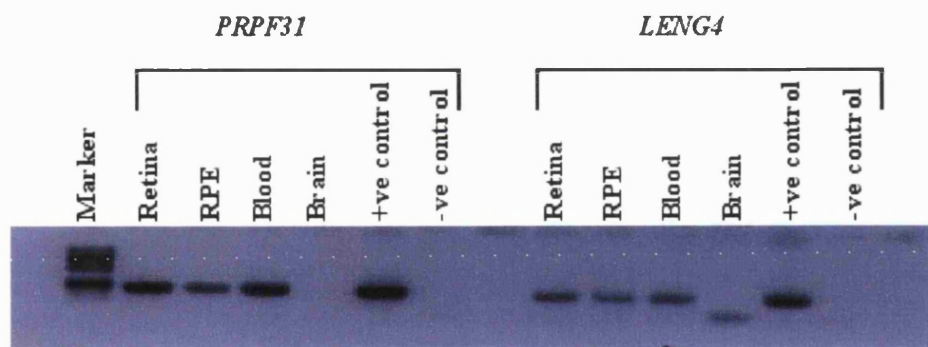


Figure 4.2 RT-PCR of two predicted genes within the *RP11* critical interval of 500kb. The expression profile of *PRPF31* and *LENG4* in retina, RPE, blood and brain is shown.

Following the exclusion of *PRKCG* and the putative *LOC91663* gene (Al-Maghtheh *et al.*, 1998; see chapter3), both of which were part of BAC clone 331H23, attention was focused on the analysis of BAC clone CTD-3093M3 (accession number AC012314), the biggest clone in the region (size 180kb). Sequence from an overlapping BAC clone 402N14 (size 162kb) was also analysed to confirm gene predictions from 3093M3. Despite the fact that there was sequence gaps within 3093M3 clone, a number of genes were predicted from the sequence of the clone. A previously excluded gene, *RPS9* (Al-Maghtheh *et al.*, 1998) and a number of other genes, such as *LENG4*, *LENG5*, *TFPT*, *CNOT3* and *PRPF31* were identified and fully characterised within that clone (table 4.1, figure 4.3).

Table 4.1 Genes identified from the analysis of the genomic sequence of BAC clones within the *RP11* critical interval.

HUGO ID	Lab ID	No. of exons	Tissue expression	Protein size	Putative function/ Homology
<i>WI-17997</i>	WI-17997	2 exons, start and termination codons are in the second exon	Ubiquitous including retina and RPE	322 amino acids	Hypothetical protein BC013995, homologous to mouse myeloid up-regulated protein.
<i>NDUFA3</i>	-	4 exons	Ubiquitous including retina	83 amino acids, 9kD	NADH-ubiquinone oxidoreductase b9 subunit function in electrons transfer from NADH to the respiratory chain.
<i>TFPT</i>	-	7 exons	Hemopoietic cell lines	270 amino acids, 28.3 kD protein	No homology to known genes. Possible role in development of leukemogenesis.
<i>PRPF31</i>	PRPF31	14 exons, start codon in the second exon	Ubiquitous including retina	499 amino acids	34% similar to yeast splicing protein PRP31.
<i>CNOT3</i>	KIAMINO ACIDS0691	17 exons	Brain. Retina, RPE, blood	753 amino acids	Homology to yeast <i>NOT3</i> gene (nearer the NH2 end and the COOH end). <i>NOT3</i> negatively regulates the basal and activated transcription of many genes.
<i>LENG4</i>	-	2 exons	cDNA isolated from bladder tumour cell cDNA library.	199 amino acids	Malignant cell expression enhanced gene. No homology to known genes. Possible role in tumour progression.
<i>LENG5</i>	A006107	5 exons, 1 st exon untranslated.	Ubiquitous including retina and RPE	310 amino acids	A putative tRNA splicing endonuclease γ subunit.

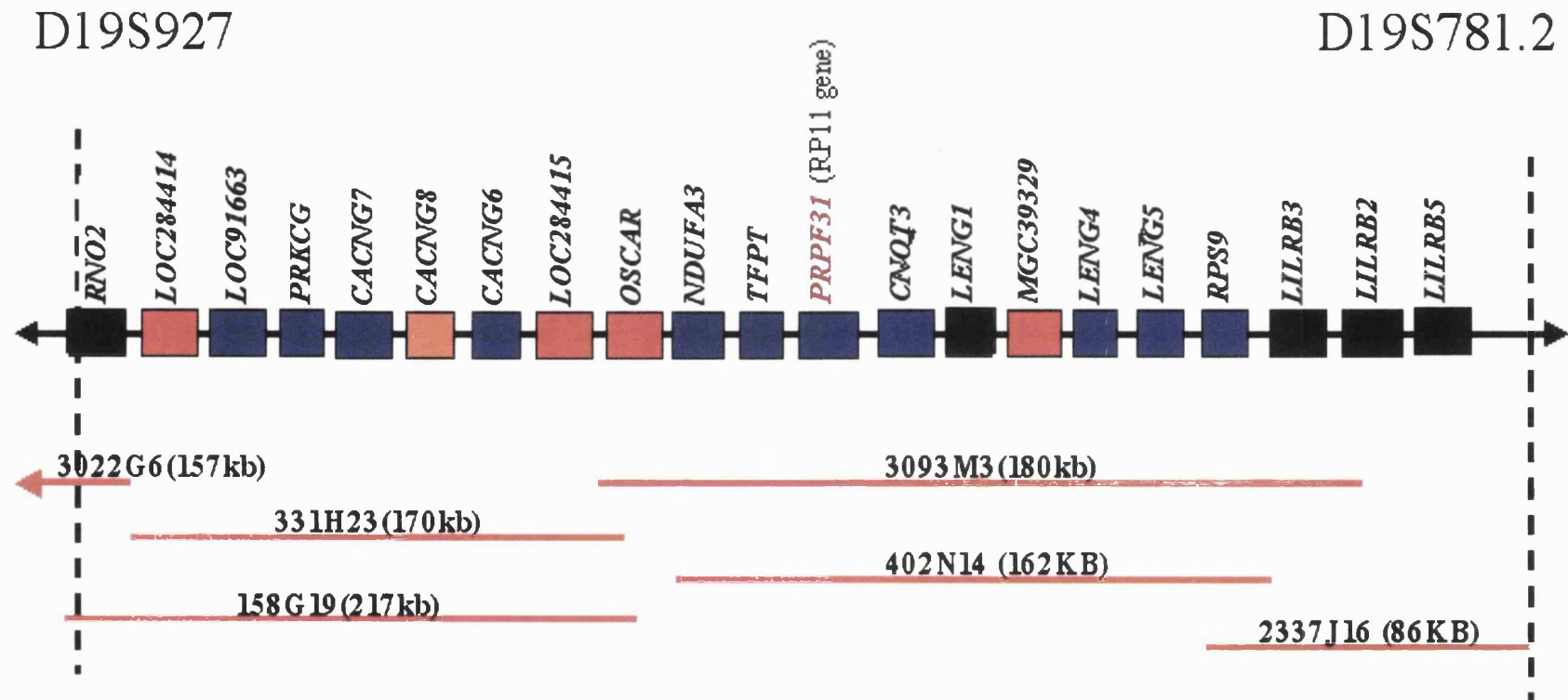


Figure 4.3 RP11 critical interval between markers D19S927 and D19S781.2. An estimate of approximately 500kb for the *RP11* interval was derived from the sizes of the 5 individual BAC clones. Twenty one genes have been identified in this region and are represented as boxes. Genes shown in **blue** represent genes that have been fully characterised and excluded. Genes shown in **black** represent those for which only the genomic organisation was established. Pseudogenes identified in the region are shown in **orange**. Genes in **red** represent the recently identified genes as sequence gaps were filled and completed. The above figure is not an exact representation of the RP11 genomic region in terms of gene sizes and distances between genes.

4.3 Methods

NIX analysis was the main tool used in identifying genes within the BAC clones spanning the *RP11* genetic interval. The genomic organisation of predicted genes was further confirmed by comparing the cDNA sequences of these genes to that of BAC clones. PCR primers were then designed and PCR reactions were carried out on DNA from affected individuals from all *RP11* linked families. PCR fragments were screened for mutations using the DHPLC (Wave[®]) machine and direct sequencing (see section 2.5). Segregation of mutations in the families was shown on agarose, and MDE denaturing gels.

4.4 Results

4.4.1 *PRPF31*, the gene for *RP11*

Screening of NIX predicted genes in the affected members of *RP11* linked families led to the identification of mutations in the *PRPF31* gene, which encodes a protein found to share homology to a yeast pre-mRNA splicing factor PRP31p (Weidenhammer *et al.*, 1996; Weidenhammer *et al.*, 1997).

4.4.1.1 Genomic organisation of *PRPF31*

The genomic organisation of *PRPF31* (precursor RNA processing factor 31) was determined by the comparison of the cDNA (accession number AL050369) and genomic sequence obtained from the BAC clones CTD-3093M3 and *RP11*-402N14 (accession number AC009968). The *PRPF31* gene comprises of 14 exons, spans approximately 18kb of DNA, and encodes a protein of 499 amino acids. PCR analysis of cDNAs from a variety of normal tissues including retina with a primer pair that

amplifies exons 13 and 14 of *PRPF31* as a single amplicon, produced the expected 350bp product in all tissues (figure 4.4). This result is in agreement with the EST database (www.ncbi.nlm.nih.gov/UNIGENE/), which shows *PRPF31* gene to be expressed in a wide range of tissues including brain and retina.

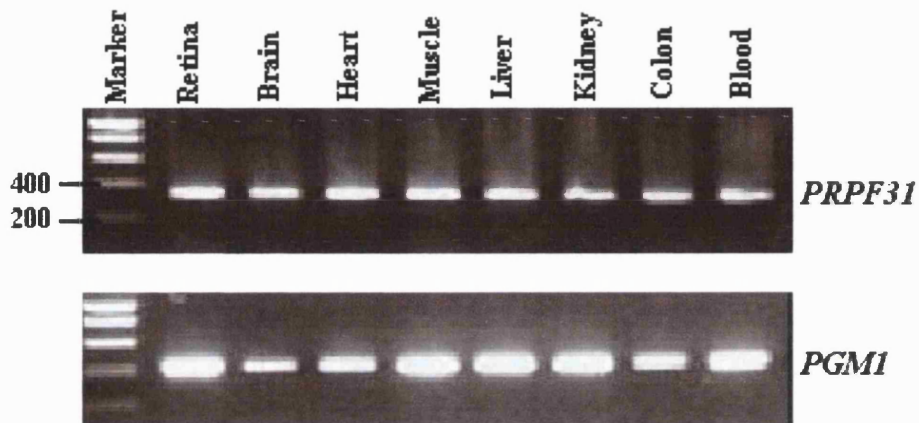


Figure 4.4 Expression profile of *PRPF31*. cDNA expression profile for exons 13-14 of *PRPF31* showing amplification of the 350bp product from all tissues including retina. An ubiquitously expressed gene *PGMI* (Whitehouse *et al.*, 1992) was used as a positive control; its 400bp fragment was also observed in all tissues.

4.4.2 Mutation screening

4.4.2.1 Mutation screening in *RP11* linked families

Mutation analysis of the *PRPF31* gene was carried out by amplifying the exons from genomic DNA using primers located in flanking intron and UTR sequences (Table 4.2).

Table 4.2 *PRPF31* intronic primers used for PCR amplification and direct sequencing of *PRPF31* exons.

Exon	Primer sequence 5' - 3'	Mg ⁺² (mM)/Tm
Ex1	F CTGCTCCAATTCCATCTCCG R GTTGCAATGAGCCGAGATCG	1.5/59°C
Ex2	F CATCGCTCAGTAATAAGGAGG R TGGGAGATTCTGTCTAGAGC	1.5/55°C
Ex3	F AGTGCTGGATTCTGACTGTC R GAAGCATGAGCTTTACAATGC	1.5/55°C
Ex4	F TAGATACTCACACCCATGCC R AGGTCAACCTCGATCTGAGC	1.5/55°C
Ex5	F GAGTCTCCTTCCATCTCACC R CAGAGCAGACCACTGAGCC	1.5/59°C
Ex6	F GAGCAAGAGAGGTTCTCG R CAGCCTAATCCCCAATCC	1.5/58°C
Ex7	F GGAGATCCAGGAGGCTGG R GGTCACAGTGTCAGCAGACC	1.0/61°C
Ex8	F CTCTCTGCTTTCTTCTGACC R TGAGTGCTACCGTCAGCT	1.5/60°C
Ex9	F GGTTGCTTTGCTGTTACCTC R CCTTGGTAGGACAGTGCTC	1.5/55°C
Ex10	F TAAGGCACGTGGATACTCG R CATCTTGCGGTACCTGGG	1.5/61°C
Ex11	F CCCAGGTACCGCAAGATG R CCTAGACATGAGCCAGGG	1.5/59°C
Ex12-13	F CTGGTCGCTGAACTGCAG R CCTTGCCAAGGTCTCAGG	1.0/64°C
Ex14	F CTCACCTAACCCATCATCC R GCTCTGATTCTTCTAGTTGCC	1.5/55°C
Promoter	AF GTGGAAAGGCAGAGATTGC CR CAGATTGAAGAGGTCGGAAGC	1.5/58°C

Mutation analysis was undertaken in seven adRP families (ADRP5, ADRP29, ADRP11, ADRP2, RP677, RP1907, and a Japanese family). Six of these families were previously reported as *RP11* linked families (Vithana, 1998) and one pedigree were subsequently identified during the mutation screening process. The mutations identified in these families co-segregated with the disease in all cases. None of the mutations were present in 100 control subjects, thereby making it improbable that the identified mutations are rare polymorphisms.

The ADRP5 family initially mapped to the *RP11* locus (Al-Maghtheh *et al.*, 1996) showed a 11bp deletion (1115-1125del) in exon 11, which was found to segregate in all affected members of the family (figure 4.5). From the sequence it maybe inferred that an aberrant truncated protein of 469 residues maybe produced.

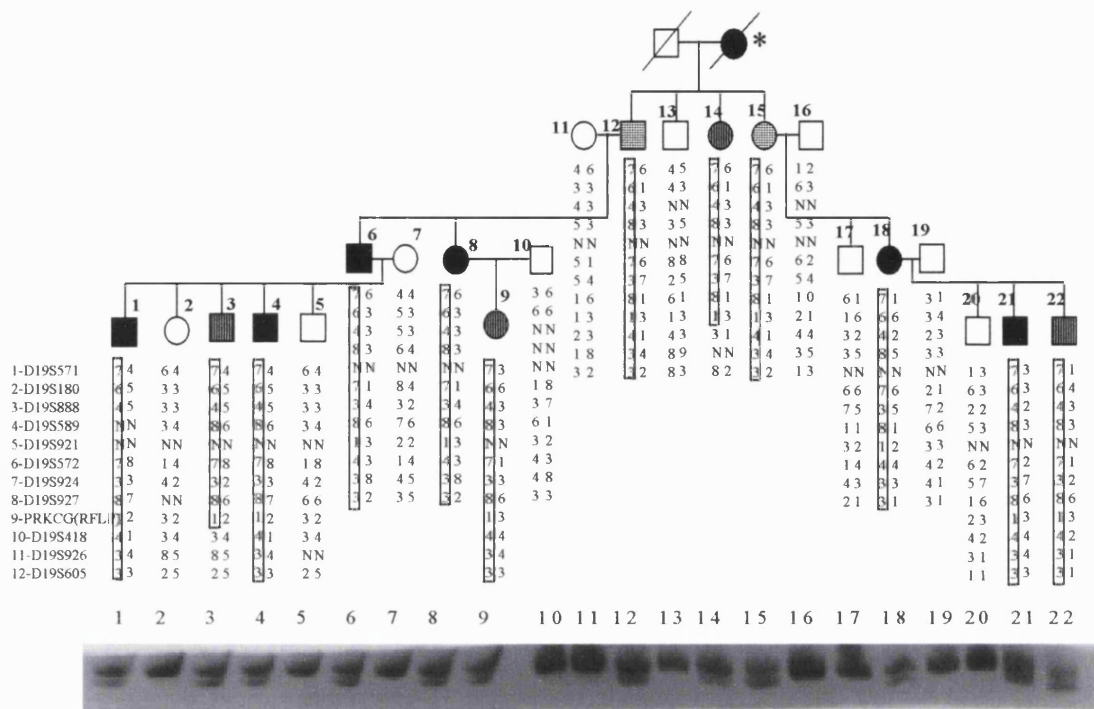


Figure 4.5 Segregation of the 11bp deletion mutation (1115-1125del) in pedigree ADRP5 on MDE heteroduplex gel. Only a branch of ADRP5 pedigree is shown along with the haplotypes for the *RP11* markers, the asterisk denotes individual II-2 in the original pedigree (Al-Maghtheh *et al.*, 1994). Lane assignment (1-22) corresponds to samples of the respective individuals in the pedigree. Individuals 12 and 15 are asymptomatic obligate carriers, whilst individuals 3, 9, 14 and 22 denote asymptomatic disease haplotype carriers. All affected haplotype carriers exhibit the deleted allele.

In family ADRP29 (Al-Maghtheh *et al.*, 1996) a missense mutation was identified in exon 7 in all affected individuals (figure 4.6). This C>G transversion missense mutation at position 646 results in an Ala216Pro substitution, a residue which is widely conserved across homologous genes in *S. cerevisiae*, *S. pombe*, *D. melanogaster* and *A. thaliana*.

In family RP1907 (Al-Maghtheh *et al.*, 1996) a sequence change, that segregates in the family, was identified in the third base of the donor site of intron 6 (IVS6+3A>G) (figure 4.7). A neural-network computer program for recognising splice sites (available at <http://www.fruitfly.org/seq-tools/splice.html>) predicted a reduction in the likelihood of splicing at this site (probability reduced from 0.99 to 0.68). Retention of intron 6 would result in a truncated protein of 186 amino acids. A similar result would be expected from the mutation in the previously unreported *RP11* linked family RP677 where a deletion in intron 6 (IVS6-3 to -45 del) removes the essential polypyrimidine tract located between the intron branch point and the splice acceptor site (figure 4.8). RT-PCR or an *in vitro* splicing assay would give more precise information for the effect of both of the mutations (the IVS6+3A>G and IVS6-3 to -45 del) on the splicing of the mutant transcript. However, it is unlikely that these transcripts would produce a functional protein. If intron 6 is retained, the mutant transcript would produce an aberrant protein lacking the highly conserved NOP domain.

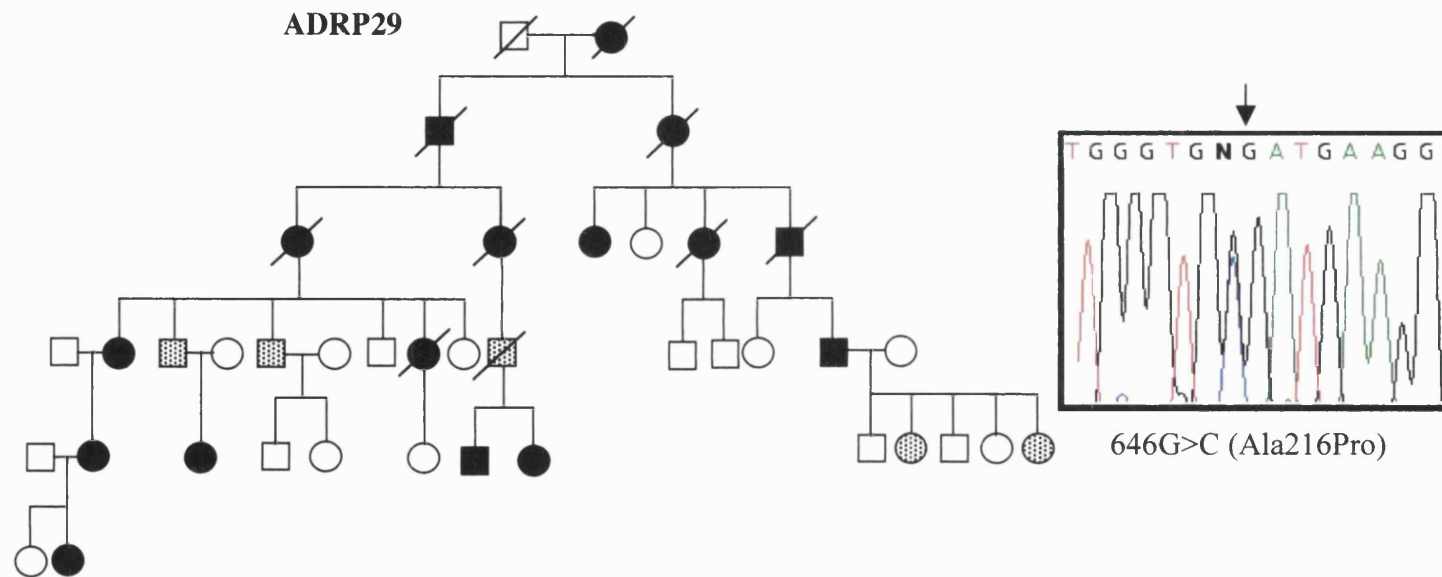


Figure 4.6 Pedigree of ADRP29. Right is part of exon 7 sequence, the pointed arrow is at the C>G transversion missense mutation at position 646. This mutation was found by sequencing to segregate with all affected individuals in this family.

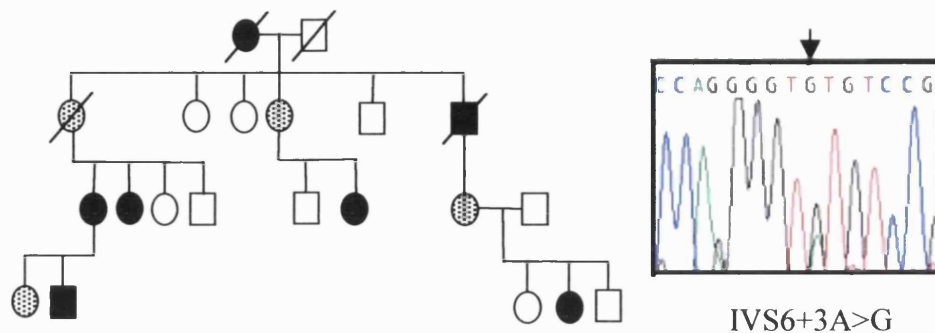


Figure 4.7 Pedigree of RP1907. Right is part of intron 6 sequence, the arrow pointed at the sequence change (A>G) at the third base in the donor site of intron 6 (IVS6+3A>G). This mutation was found by sequencing to segregate with all affected individuals in this family.

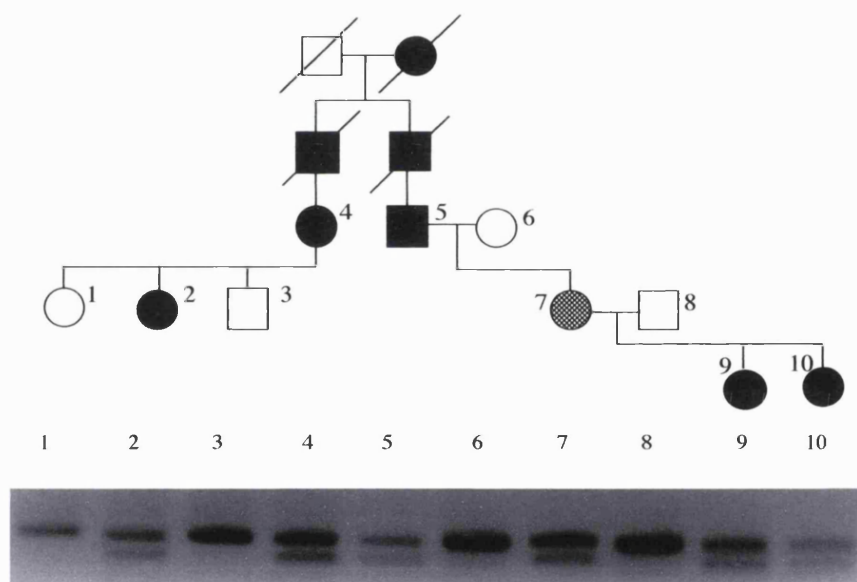


Figure 4.8 Segregation of the 43bp deletion mutation (IVS6-3 to -45 del) in pedigree ADRP677 as shown by agarose gel electrophoresis. Lane assignment (1-10) corresponds to samples of the respective individuals in the pedigree. Individual 7 in the pedigree is an asymptomatic obligate carrier and demonstrates the incomplete penetrance in this pedigree. All affected individuals exhibit the deleted allele.

A previously reported Japanese family linked to the *RP11* locus (Xu *et al.*, 1995) was also included in the *PRPF31* mutation analysis. Sequence analysis in this family revealed a single base pair deletion (1141-1142delG) in exon 11, which would produce a truncated protein of 411 amino acids, if the aberrant mRNA was translated. Segregation of this deletion in this family is shown in figure 4.9 on MDE gel.

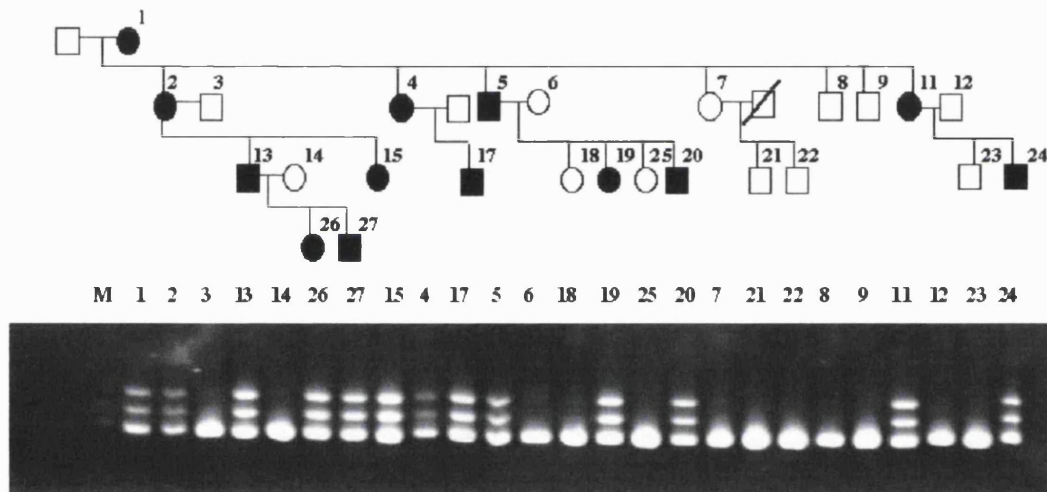


Figure 4.9 Segregation of a one base pair deletion (1141-1142delG) in the *RP11* linked family of Japanese origin. PCR fragments of exon 11 were separated using MDE gel electrophoresis. Lane assignment (1-27) corresponds to samples of the respective individuals in the pedigree. The third band in lanes of the affected individuals is due to conformational changes as a result of the deletion in the mutant allele.

4.4.2.2 Screening for deletions in the *PRPF31* gene

4.4.2.2.1 Screening *PRPF31* for a large deletion in *RP11* linked family ADRP2

The routinely used mutation screening method in this project was DHPLC analysis (wave analysis) and sequencing. When the *RP11* linked family ADRP2 was first screened, no obvious mutation was identified in the *PRPF31* gene. However, close scrutiny of the data revealed that all affected individuals were hemizygous for all Single Nucleotide Polymorphisms (SNPs) within the *PRPF31* gene indicating a large intragenic deletion. However, the extent of the deletion was unknown, and whether the deletion includes only the *PRPF31* gene or other neighbouring genes as well remains to be studied.

Since *PRPF31* gene lies between the *PRKCG* intragenic marker and the D19S785, it was important to re-analyse the segregation of these two markers in the ADRP2 family to identify whether these markers were also deleted. The RFLP marker (an *MspI* polymorphism) associated with the *PRKCG* gene and the microsatellite marker D19S785 were both found to segregate in the family (figure 4.10a), indicating that the deletion does not extend as far as these two markers.

A number of polymorphic markers and Single Nucleotide Polymorphisms (SNPs) have been identified within the genomic region between *PRKCG* and D19S785 and they were typed in all members of the ADRP2 family to identify the size of the deletion (table 4.3, figure 4.10b). Analysis of SNPs JST060265, JST060264 and JST114222, located far upstream of *PRPF31*, did not show a loss of heterozygosity. However, marker 77965(GATA)_n was shown to be hemizygous, indicating that the deletion start point is between marker 77965(GATA)_n and JST114222. The 77965(GATA)_n marker is positioned approximately 13kb upstream of *PRPF31*.

Analysis of SNPs markers within *PRPF31* showed that all were hemizygous, therefore the deletion includes most of the gene. However, analysis of the 109617(GTTT)_n marker, which is at the 3' end of *PRPF31* was found to be heterozygous. In an attempt to get closer to the deletion termination point another SNP in exon 14 of *PRPF13* (1461C>T) was typed in the entire family. Despite the fact that this SNP was uninformative in most of the family members, it was found to be heterozygous in the symptomatic individual III-5. This suggested that a genomic region of approximately 30kb in length, encompassing upstream genes osteoclasts associated receptor (*OSCAR*), NADH dehydrogenase (ubiquinone)1 alpha subcomplex3 *NDUFA3*, TCF3 (E2A) fusion partner (*TFPT*) and *PRPF31* gene up to its fourteenth exon are deleted in this family (figure 4.10b and 4.10c).

Table 4.3 Markers and SNPs used in identifying the extent of the deletion in ADRP2 family

STS	Position	Primers sequence Forward and reverse/5'-3'	Polymorphism
JST060265 JST060264	~ 75kb 5' <i>PRPF31</i>	F -TGGGGTAGAAAGAGTATG R -TCTATTTTCTCCAAAGACGG	T/C
JST114222	~ 13kb 5' <i>PRPF31</i>	F -GAAAGGGGTGACTCACAGAG R -TTCCACAGTGCCATAGGGCT	T/C
77965	5' <i>NDUFA3</i>	F -TTTGCACCTATTATGTGCCA R -TCACAGCAAGGGCCTGGCAC	(GATA) _n
SNP's within <i>PRPF31</i>	*IVS5+81 IVS5+82 735 IVS8+39 IVS9-70 IVS11-9 1461	**Exon 5 primers Exon 5 primers Exon 8 primers Exon 8 primers Exon 10 primers Exon 12 primers Exon 14 primers	T/C C/G C/T(Pro245Pro) G/A C/T C/T(Val489Val) C>T
109617	3' <i>PRPF31</i>	F -GGATGATCCTGGCTCATG R -GAGAATCACTTCAACCTGG	(GTTT) _n
D19S785	~12kb 3' <i>PRPF31</i>	F -AGGTCATCCCTCAGACAAGT R -TCAGCTCTTCACAGCTTTTCG	(CA) _n

* Nucleotide designation commencing 1 at position 36 of GenBank entry AL050369.

** See table 4.2 for primers sequences.

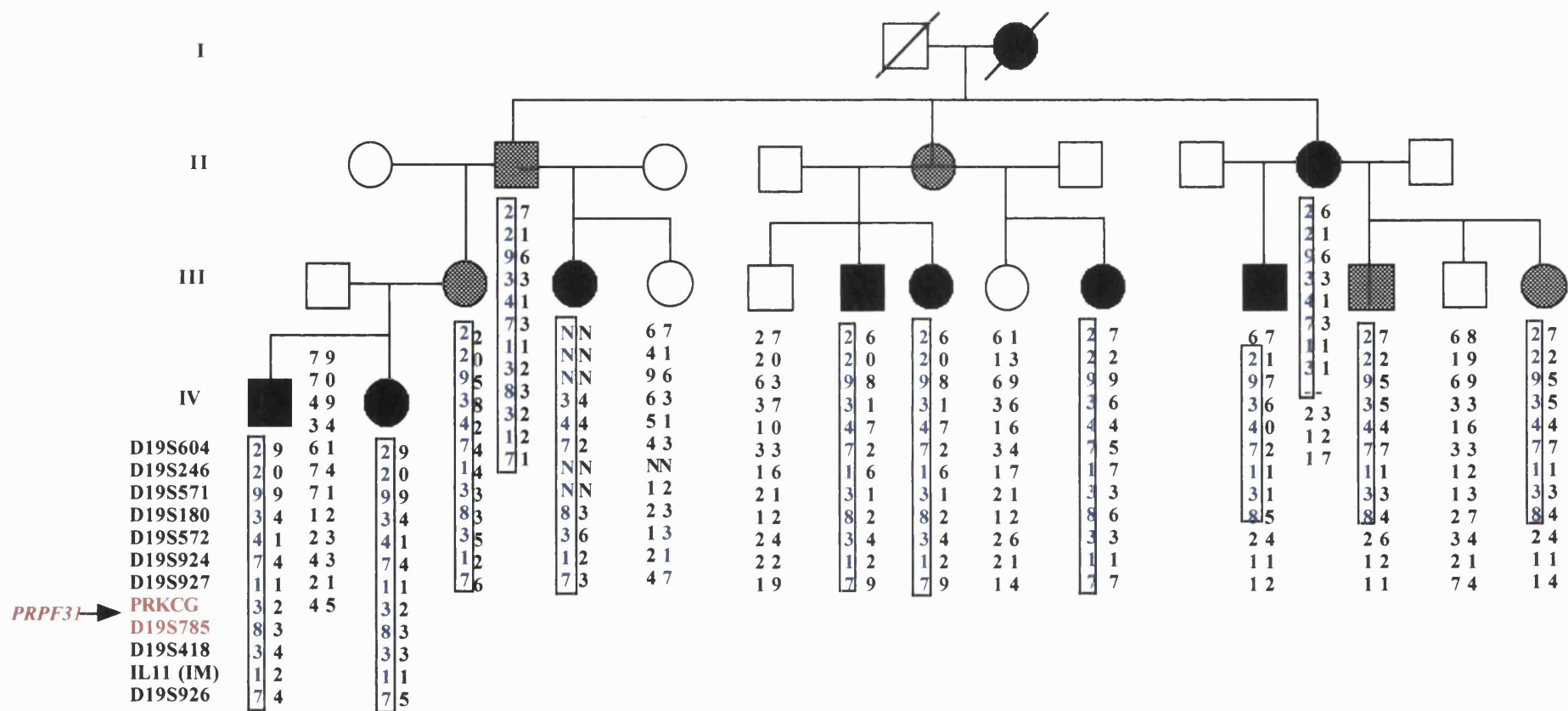


Figure 4.10a Haplotype analysis of 12 markers from 19q13.4 in ADRP2 (Vithana, 1998). The genetic interval between intragenic PRKCG marker and D19S785 was analysed to investigate the deletion size found in this family.

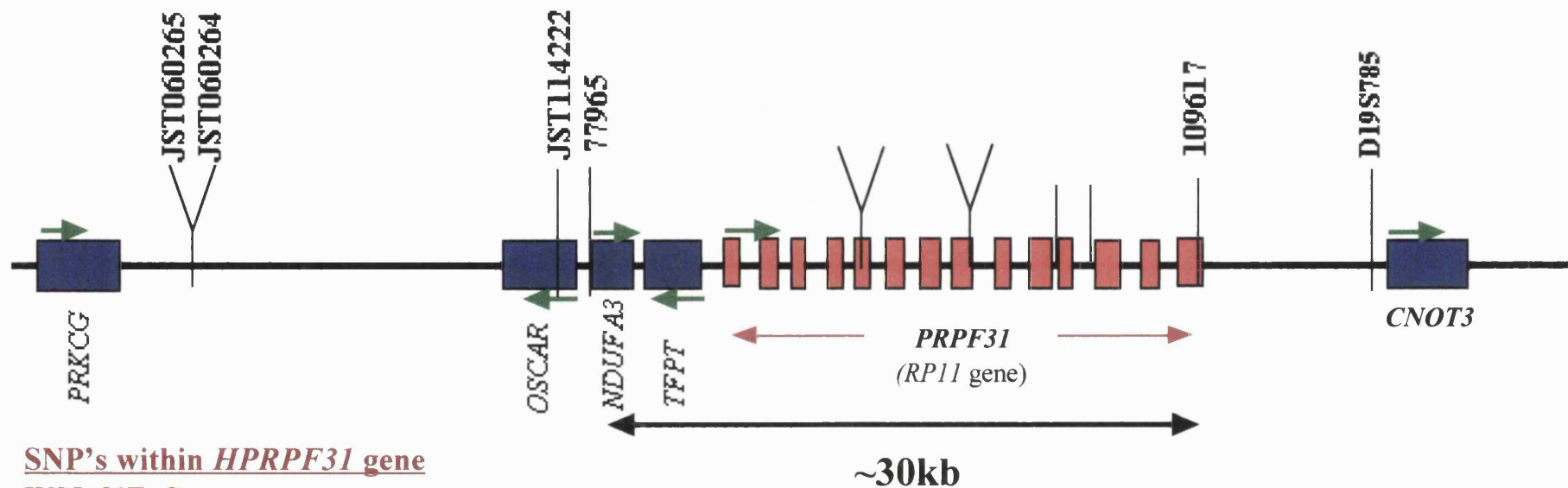


Figure 4.10b Schematic diagram of SNPs and markers analysed in family ADRP2 between *PRKCG* and *CNOT3* genes. Genes in this genetic interval are presented as boxes, the genomic structure of the *RP11* gene is shown. Green arrows show translation direction. The SNPs are pointed by upright lines and **bolded font**. Sizes of genes and distance between markers/SNPs are approximate.

4.4.2.2.1.2 Possible clinical implication of the deletion in the ADRP2 family

A large deletion, which includes a number of functional genes may result in a more complex and non-specific phenotype. As mentioned earlier, three genes in addition to the *PRPF31* are deleted in the ADRP2 family, however the only phenotype observed in this family is RP. The initial clinical reports on this family identified some affected members who suffered from macular degeneration (Moorfields Eye Hospital, UK). However, the cone rod dystrophy locus proximal to the *RP11* locus was excluded by typing markers D19S246 and D19S604 (Vithana, 1998). Further clinical testing of this family was not possible. Therefore, the phenotype description remains to be fully described.

The first gene included in this deletion, *OSCAR* is a novel member of the leukocyte receptor complex, which is highly expressed in preosteoclasts or mature osteoclasts (Suda *et al.*, 1999). It was reported that the *OSCAR* protein might have a role in regulating osteoclast function or differentiation. Studies showed that *OSCAR* is also required for the formation of mature osteoclasts from bone marrow precursor cells (Nacksung *et al.*, 2002). There is no evidence for *OSCAR* involvement in any disorder, however defects in this gene may have an effect on osteoblast maturation and as a result on the integrity of bone structure.

NDUFA3 is the second gene deleted in the ADRP2 family. This enzyme is part of a large subgroup of the NADH ubiquinone oxidoreductase (complex I), which is a complicated multi-protein enzyme complex found in the inner mitochondrial membrane (Loeffen *et al.*, 1998). This complex functions in the transfer of electrons from NADH to the respiratory chain. Moreover, deficiency in this protein complex I was found to be a major cause of abnormalities in the energy generating system within the mitochondria. However, it is unlikely that the heterozygous defect in this gene would show a phenotype since most

enzymatic defects are the result of homozygous changes and results in a recessive disease phenotype.

The third gene deleted in this family is *TFPT*. This gene is a ubiquitously expressed gene with higher levels of expression in the brain compared to other tissues. Expression was also observed in hematopoietic cell lines (Brambillasca *et al.*, 2001). This gene shares protein similarity with Amida encoded by the rat ortholog of *TFPT*. Amida was found to induce apoptosis when expressed in cultured cells. Therefore, based on protein homology, *TFPT* may have a role in apoptosis. It was also suggested that the level of expression of *TFPT* might play an important role in leukemogenesis (Brambillasca *et al.*, 2001).

There have not been any reports of direct involvement of these genes in any human disorders. Therefore, it is difficult to speculate on a particular phenotype within the *ADRP2* family although a thorough clinical assessment may reveal a more subtle phenotype. However, it could very well be that one copy of any of these genes does not lead to loss of function.

4.4.2.2 Investigation of a putative deletion in *RP11* linked family, *ADRP11*

As presented above, mutations have been identified in all but one of *RP11* linked families. Pathological mutations have not been identified in the coding or flanking intronic regions of *PRPF31* in the linked *ADRP11* family (Vithana *et al.*, 1998). Screening of the *ADRP11* family by heteroduplex analysis showed a doublet band corresponding to a large PCR fragment (750pb) spanning the untranslated exon 1 region of *PRPF31* (figure 4.11). The pattern that appeared on the denaturing hetroduplex gel suggested a deletion, however cloning and sequencing of

that fragment failed to show any aberrant sequence. Only one sequence change was identified within that fragment (IVS1+54T>C). Interestingly this sequence change was shown to retard the migration of the fragment in non-denaturing PAGE gels as well. Heteroduplex and sequence analysis confirmed the IVS1+54T>C change in all affected members of the family. Therefore, the band pattern was confirmed as being due to this sequence change and not a deletion within that fragment. Since this change segregates with the disease, although it appears to be a polymorphism, its rarity (0.5% in sporadic RP panel) makes it important to investigate its possible involvement in disease. The localisation of IVS1+54 T>C in intron 1 of *PRPF31* suggests that it maybe part of an enhancer element that modulates the activity of the promoter. It is also possible, however, that this sequence change is in linkage disequilibrium with the real mutation in *PRPF31* gene in this family. The mutation in ADRP11 family may have been missed due to the insensitivity of the detecting methods used.

ADRP11

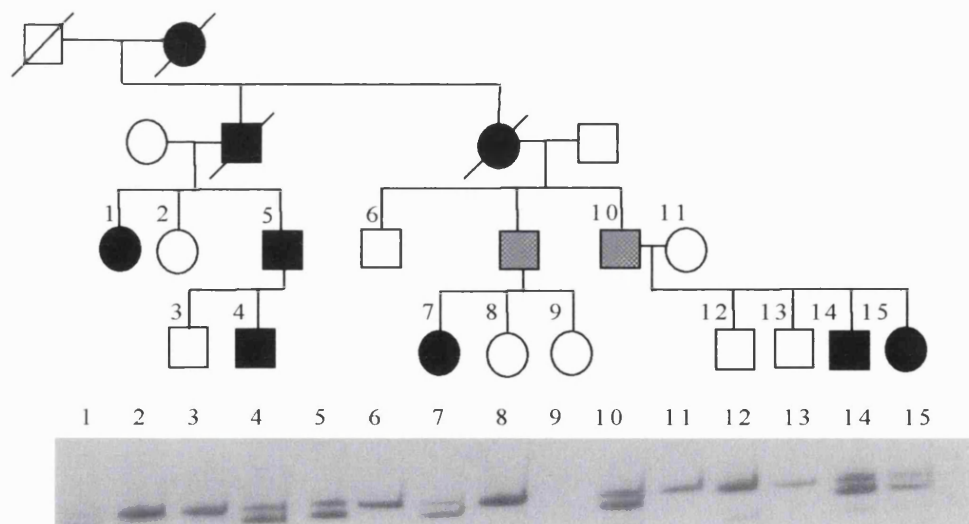


Figure 4.11 Segregation of IVS1+54T>C sequence change in *RP11* linked family on denaturing heteroduplex gel. Line assignment (1-15) corresponds to samples of respective individuals in the pedigree.

4.4.2.3 Screening other RP cases for mutations in the *RP11* gene

4.4.2.3.1 Mutation screening in small families

Additional small nuclear families were also included in the mutation screening of *PRPF31*. The number of DNA samples available from these families was too small to prove their linkage the *RP11* locus (figure 4.12). No mutations were identified in the majority of these families. However, mutations screening led to the identification of a missense mutation (R211Q) in a small family of Greek origin (figure 4.13). This mutation was not found in 60 normal Greek individuals, making it unlikely that this change is a polymorphism within the Greek population. This change was also not identified in 100 normal controls of British origin.

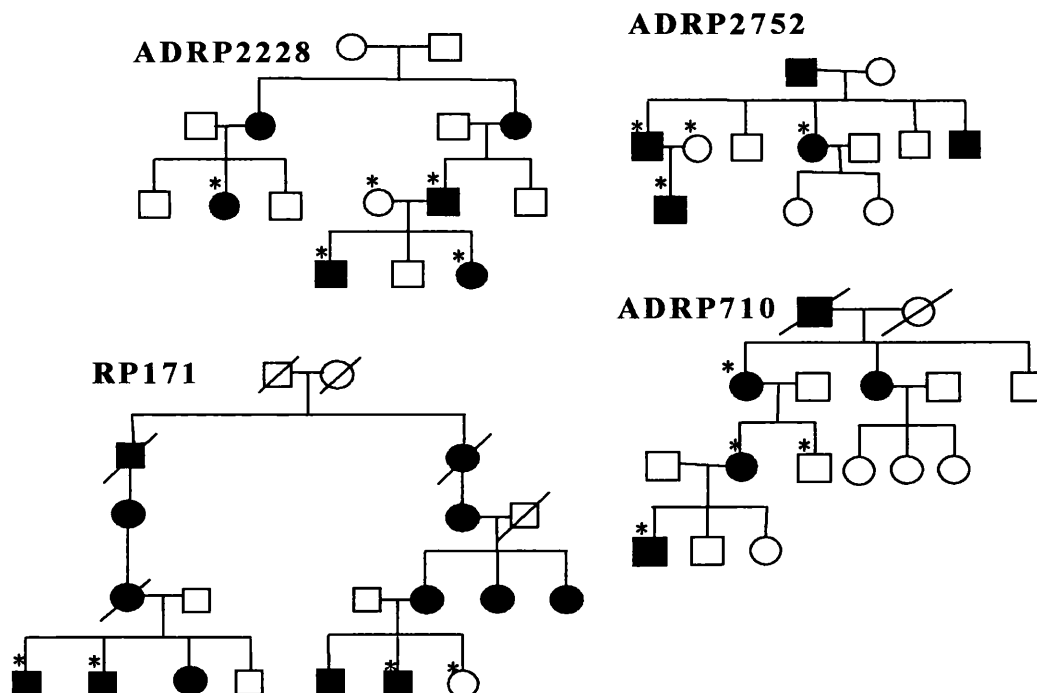


Figure 4.12 Small family pedigrees screened for mutations in the *RP11* gene. The star (*) marks the individuals where DNA was available.

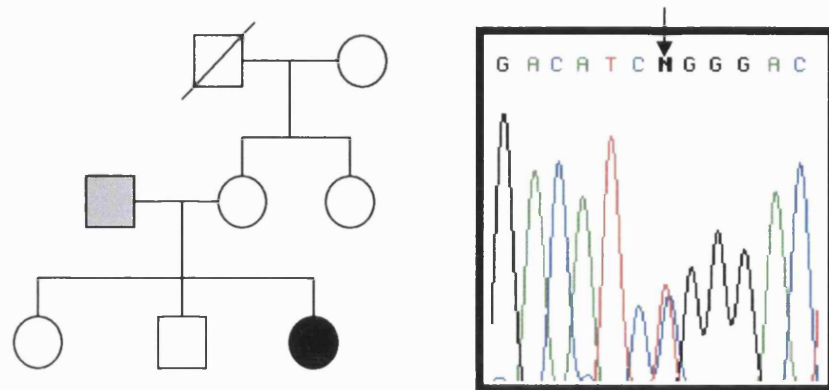


Figure 4.13 Missense mutation in a small Greek family (G>A; R211Q), the complementary strand of exon 7 is shown.

4.4.2.3.2 Mutations identified in the sporadic RP panel

RP cases for which no family history of disease was evident were grouped in a panel of sporadic or simplex cases. This panel, consisting of 300 sporadic RP cases, was also subjected to mutation screening by DHPLC WAVE[®] and sequence analysis. Additional mutations and polymorphisms were identified in the coding as well as in the non-coding sequence of *PRPF31*. The three mutations identified in sporadic RP individuals include a missense and two insertion mutations. In individual SP117, a single base pair insertion (769-770insA) causes a frameshift mutation. This mutation could produce a truncated protein of 277 residues if the transcript was to be translated (figure 4.14a). In individual SP42, a 581C>A transversion causing the non-conservative amino acid change A194E was identified (figure 4.14b).

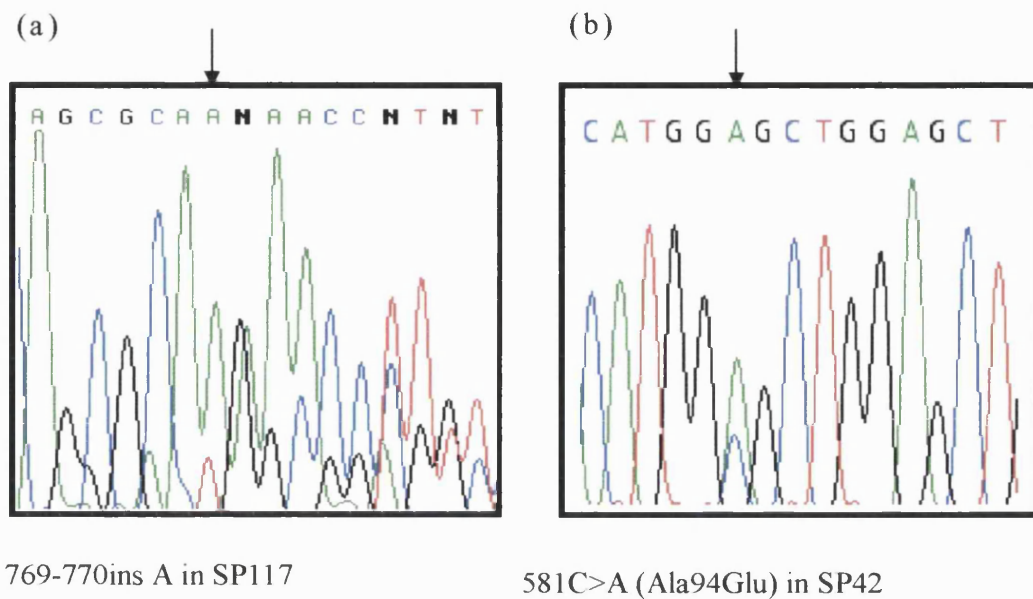


Fig 4.14 Mutations identified in sporadic RP individuals. Figures a and b show insertion and point mutations identified in sporadic patients 117 and 42, respectively.

In individual SP14, a 33bp insertion (580-581dup33bp) involving codon 194, the same codon mutated in SP42, duplicates residues ELERLEEACDM producing a slightly longer protein of 510 residues (figure 4.15a). Familial clinical records were examined when any of these sporadic cases showed sequence variations in a candidate gene to confirm previous family history of RP. Clinical records of individual SP14 showed a family history of RP, however, the mother of SP14 had a rhodopsin mutation (Gly106Arg), which abolishes *ApaI* site (family RP1038) (Moore, 1992). SP14 was screened and found to be negative for the rhodopsin mutation (figure 4.15b). Therefore, it appears that RP in SP14 was due to the *PRPF31* mutation inherited from his father, who was clinically normal indicating incomplete penetrance; typical of the *RP11*

phenotype. Segregation in this small family was carried out and confirmed that both SP14 and the father carry the *PRPF31* mutation (figure 4.15c).

All of the above mutations showed distinct elution profiles on DHPLC that were not found in 100 control subjects, confirming that these mutations are not rare polymorphisms but are very likely to be pathogenic mutations.

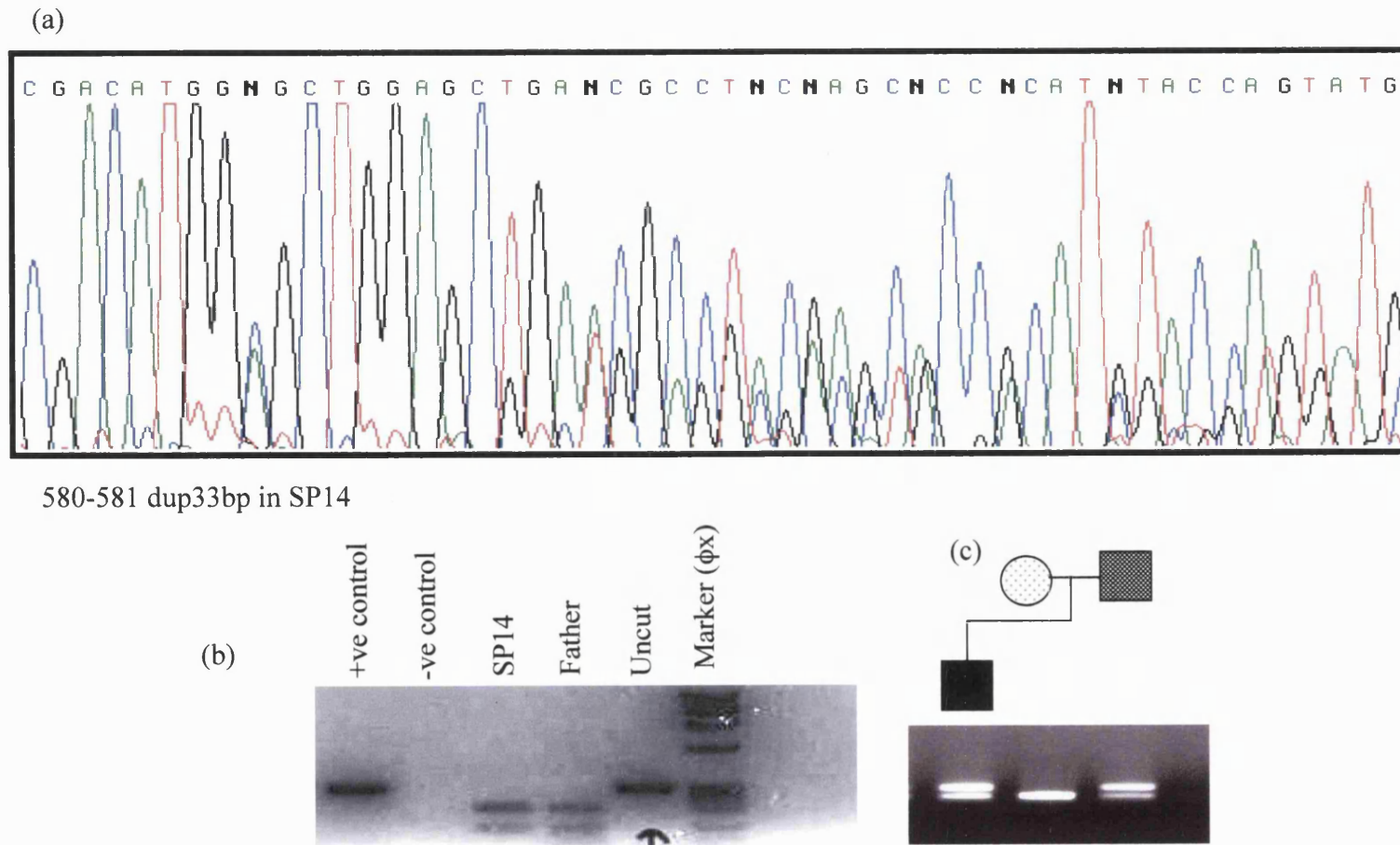


Figure 4.15 Electropherogram of insertion (duplication) in exon 7 found in sporadic patient 14. Figure b shows the *Apal* enzyme digest. Segregation of the *PRPF31* mutation is shown in figure c.

4.4.2.3.2.1 Screening the sporadic RP panel for small deletions

The nature of the mutations identified in the *RP11* families and sporadic cases strongly suggest that deletions, insertions and sequence rearrangements are more prevalent in the *PRPF31* gene than point mutations. Such mutations are more difficult to identify using routine PCR based screening method. In an effort to identify further mutations in the sporadic RP panel (325 individuals) long PCRs spanning one or more intronic/exonic sequences were amplified. Introns 2, 4, 5 6, 7 and 8 were analysed by running these PCR fragments on low percent agarose gels (figure 4.16). Unfortunately, no deletions/insertions were identified in the panel. Insertions/deletion as small as 30bp can be visualised on low percentage agarose gels (see figure 4.15c). Introns 10 and 12 were analysed by DHPLC and by direct sequencing since they were small in size (exons 10/11 and 12/13 were contained in two PCR fragments). Introns 1, 3, and 9 were too large to be analysed using this approach.

Based on the number of mutations identified in the panel composed of sporadic British RP patients, the frequency of *PRPF31* mutations is ~ 3.3%. This value is likely to be a gross underestimate as it is likely that a significant proportion of the sporadic RP individuals in the panel have recessive mutations. Also the PCR based mutation detection method used for large-scale screening may have failed to detect large rearrangement mutations that seem to be more common in *PRPF31*. The reason a sporadic RP panel was screened was due to the fact that the nature of the *RP11* phenotype (i.e. partial penetrance) suggested the possibility of many sporadic RP cases having mutations in *PRPF31*.

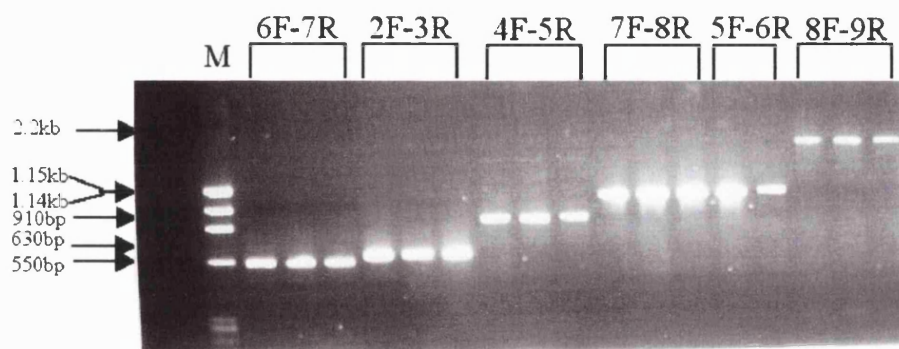


Figure 4.16 PCR amplification *PRPF31* introns and exons in sporadic RP individuals. Fragments amplified and the size of these fragments are shown (i.e. 6F-7R the fragment includes exons 6, 7 and intron 6).

4.4.2.4 Identification of non-pathogenic sequence variations within the *PRPF31* gene

Sequencing a number of sporadic RP individuals with distinct DHPLC wave patterns also led to the identification of exonic and intronic sequence changes (Table 4.4). None of the changes identified disease-causing mutations, and it is highly unlikely that any of these polymorphisms would be pathogenic. However, functional assessment of these polymorphisms is required to determine their role in *RP11* pathogenicity. Sequence variations identified were either intronic changes or silent mutations. Despite the fact that some of polymorphisms identified do not alter *PRPF31* protein sequence, namely (Glu188Glu) and (Pro245Pro), it is likely that they may have an effect on mRNA splicing or stability. Silent changes have been shown to affect splicing,

such sequences form part of exon's splicing enhancer (ESE) sequences (Cartegni *et al.*, 2002).

Table 4.4 Non-pathogenic sequence variations in *PRPF31* identified in *RP11* linked families and in the sporadic RP panel. Location and frequency of these sequence variations in the British population are also shown.

Location	Sequence variation	Frequency in panel
Intron 1	IVS1+14A>G	?
Intron 1	IVS1+54T>C	0.5%
Intron 3	IVS3+115C>T	26%
Intron 5	IVS5+81T>C	4%
Intron 5	IVS5+82C>G	38.5%
Intron 8	IVS7-31T>C	2%
Exon 7	564G>A (Glu188Glu)	3.4%
Exon 8	735C>T (Pro245Pro)	9%
Intron 8	IVS8+39G>A	9%
Intron 10	IVS9-70T>C	?
Intron 11	IVS11-9C>T	15%
Exon14	1461C>T (Val489Val)	15%
Intron 14	1652G>A	15%

4.5 Discussion

4.5.1 Predictive function of *PRPF31* based on protein homology

At the time mutations were identified in *PRPF31*, very little was known about the function of PRPF31 protein. Therefore, a bioinformatics approach was used to assess the likely function of the PRPF31 (accession number NP-056444). Homology screening of the protein databases (<http://www.ncbi.nlm.nih.gov:80/BLAST/>) revealed homologous proteins in *D. melanogaster* (accession number AAF49655) and *A. thaliana* (accession number T02269) as well as two PRP31p proteins from *S. cerevisiae* (accession number NP-011605) and *S. pombe* (accession number CAA17928) (Weidenhammer *et al.*, 1996; Weidenhammer *et al.*, 1997; Bishop *et al.*, 2000) (figure 4.17). However, bioinformatics screening did not reveal any homologues of PRPF31 in human, implying that PRPF31 is very likely to be present as a single copy gene and is therefore the functional equivalent of yeast PRP31p in the human. A number of splicing factors in yeast have been identified in screens of temperature-sensitive strains for those exhibiting a splicing defect at the non-permissive temperature. These are referred to as *PRP* genes, for precursor RNA processing. The *PRP31* gene in yeast was also identified in such a screen (see section 4.5.3).

PRPF31 protein alignment was done using Clustal W software, a general purpose multiple alignment program for DNA and/or proteins (EMBL/EBI, Hinxton, UK). This program identified a modest overall protein sequence identities between PRPF31 and the two yeast proteins (24.74% with *S. cerevisiae* PRP31p, 40.68% with *S. pombe* PRP31+, 61.11% with *D. melanogaster* CG6876-PA protein, and 44.89% with *A. thaliana* T13D8.6 hypothetical protein) (figure 4.17). Additionally, certain domains within the protein were found to share a much higher level of identity (up to 43%) suggesting a common function (figure 4.17). In particular, all these proteins contain a region highly homologous to

Nop, a putative snoRNA-binding domain (Pfam accession number PF01798, Bateman *et al.*, 2000). This domain is present in various pre-RNA processing ribonucleoproteins such as Nop5p (Wu *et al.*, 1998) and Nop56p (Gautier *et al.*, 1997). In PRP31p, the Nop domain is believed to mediate protein-RNA interactions required for spliceosome assembly. Further evidence for a nuclear role for PRPF31 is the presence of putative nuclear localisation signal in the medial region of the protein. All this was suggestive of a role for *PRPF31* in pre-mRNA splicing. Subsequently with the characterisation of the human pre-mRNA splicing pathway the role of PRPF31 as an essential splicing factor in human was established.

	1	1	
H._Sap	-----MSLADELLADLEEAEE-----	16	
S._cerv	-----MSSEEDYFDELEYDLAD-----	17	
D._mel	-----MSLADELLADLEEDNDN-----	17	
S._pom	-----MSLADELLADLDDIEETTESTITDELGPDAKKR	33	
A._tha	MVSVSITLFFSSLFQLLAIVFSVYCNRIATLEDSFLADLDELSDNEA-----	46	
NOP	-----		
H._Sap	----EEEGGSYGEEEE-----PAIEDVQEETQLDLS---GDSVKTIAKLWDSKMF AEIM	64	
S._cerv	--EVNEEKEDIQTKKLT-----TVNCQTEKFNPF EILP---ESI E L F R T L A L I S P D R L S L	66	
D._mel	--ELEEE D S E M A S A E D E S L L A E K L A K P A P N L M D V D V T ---VQSVRELCKLRDSERLKNTL	72	
S._pom	RLELQLEEGNGISAELENDLDITKISDSAQKLPSEVANKFNDNNENIYQLLNSTRLRDII	93	
A._tha	--ELDENDGDVGKEEED-----VMDMDADLETLN---YDDL D N V S K L Q S Q R Y A D I M	93	
NOP	-----		
H._Sap	MKIEEYISKQAKASEVMGPV E A A P E Y R V I V D A N N L T V E I E N E L N I I H K F I R D K Y S K R F P E	124	
S._cerv	SETAQILPKIVDLKRILQQQ E I D - F I K L L P F F N E I I P L I K S N I K L M H N F L I S L Y S R R F P E	125	
D._mel	QQIEHYASRQRTAAEMLG S V E S D P E Y C L I V D A N A I A V D I D N E I S I V H K F T K E K Y Q K R F P E	132	
S._pom	EGTEKYKG--TEKQAITGNI E D D L E Y H L I V D S N S I A M E I D D E I L R L H R L V K E W Y H D R F P E	151	
A._tha	HKVEEALGKDSGAEKGT V L E D D P E Y K L I V D C N Q L S V D I E N E I V I V H N F I K D K Y K L K F Q E	153	
NOP	-----		
H._Sap	LES L V P N A L D Y I R T V K E L G N -----S I D K C K N N E N L Q Q I L T N A T I M V V S V T A S T T Q G --	176	
S._cerv	L S S L I P S P L Q Y S K V I S I L E N E N Y S K N E S D E L F F H L E N K A K L T R E Q I L V L T M S M K T S F K N K	185	
D._mel	L D S L I V G E I E Y L L A V K E L G N -----D I D Q V K N N E K L Q A I L T Q A T I M I V S V T A S T T Q G --	184	
S._pom	L S S L V L N A F D Y C K T V S S L L N -----D L D N S K T K ---L S F L P S A T V M V I A T T A T T T V G --	200	
A._tha	LES L V H H P I D Y A C V V K K I G N -----E T D L A L V D --L A D L L P S A I I M V V S V T A L T T K G --	203	
NOP	-----		
H._Sap	QQLSEEEELERLDEACDMALEFNASKHRIEYVE SRMSFTAPNLSIIIGASTAPKIMGVAG	236	
S._cerv	EPLDIKTRTQILEANSILENLWKLQEDIGQYIASKISIIAPNVCFLVQPEIAPQLIAHAG	245	
D._mel	TMLTPAEKAKIDEACEMAIEINNFKSKIYEVVSRMTFIAPNLSMIVGASTAPKLLGIAG	244	
S._pom	KPLPDEMINKNVKNCCEAIQQGEEKQKII EYVQSRISVVAPNLSAVVGSTTAPNLIGIAG	260	
A._tha	SALPEDVLQKVLEACDRALILDSARKKVLEFVFSKMGSIAPNLSAIVGSAVAPKLMGTAG	263	
NOP	-----NILEFCERILSLAEYRKQLEEYIESKMSKIAPNLTALVQELCGRLISHAG	51	

H._Sap	GLTNLSKNPACNIMLLGAQRKTLSGFSS-TSVLPHTCYIYHSDIVCSLPPDLRRKAARLV	295
S._cerv	GVLEFSRI [●] PSCNIASIGKNKHLSHELHTLESGVRQEGYLFASDMICKFPVSVHKQMIKML	305
D._mel	GLSKLSKNPACNVQVLGAQKKTLSGFSSQ-TQMLPHTCYVYYSQIVQDTAPDLRRKAARLV	303
S._pom	GLTRLGKEPACNLPALGKRRLLTTIGINN-PAVSGDYCYFLYMSEIVCKTTPDVRKQAIKMT	319
A._tha	GLSALAKNPACNVQVLGHKRNLAGFSS-ATSQSRVGYLEQTEIYOSTPPGLQARAGSLV	322
NOP	SLTNLAKLPASTIQVLGAEKALFRALKT-KAKTPKMLIYHSPLCKAPPKVRGKIARML	110
..		
H._Sap	PAHCTLAARVLSFH--ESTEG-KVGYELKDEIERKFTFWQEPFPPVKQVKPLPAPLDGQRK	353
S._cerv	CAEVSLAARVLAGQK-NGDRNTVLAHKWKAELSCKAFRLSEAPSISETKALPIPEDQPKK	365
D._mel	PAESVLAARVLAACH--ESVHG-EIGLRFKEDVEKKLNLQEPFPPVKFIKPLPKPIEGSKK	361
S._pom	PAHVALAARVLSIH--EYPDG-SFGISARKEVERKIEFLLEPPSQKPTVALPVPDDRPRKR	377
A._tha	PAHSTLAARVLAATR--GDPLG-ISGKAFFREEIRKKIEFWQEPFPPARQPKPLVPDSEPKK	380
NOP	PAHVALAARVLAFAFS--EEDGVSFVGIELREELEKRLERLEEG-----	150
..		
H._Sap	KRGGRIRYRKMKERLGLTEIRKQANRMSFGEIEEDAYQ-EDLGFSLGHLGKSGSGRVRQTQ	413
S._cerv	KRAGRKFRKYKEKFRLSHVRLQNRMFEGKQEQTVLDSYGEEVGLGMSNTSLQQAVGATS	423
D._mel	KRGGRKVRVKMKERYALTEFRKQANRMNFGDIEEDAYQ-GDLGYSRGTIGKTGTGRIRLPQ	421
S._pom	RRGGRIRYRKMKERYAVTELRLRLQNRVAFGKEEAEVFN-FDETEGLGMLGQEGEGKIRAVS	437
A._tha	RRGGRRLRKMKERYQVTDMRKLANRMAFGTPEESSLG-DGLGEGYGMGLGQAGSNRLRVSS	440
NOP	-----	
H._Sap	VNEATKARISKTLQRTLQKQSVVYGGKSTIRDRSSGTASSVAFTPLQGLEIVNPQAAEKK	473
S._cerv	GSRRSAGNQAKLTQVMKHRIS---EANQQADEFLISLGHNTE-QPNLSPQMVQMHKKQHT	482
D._mel	VDEKTKVRISKTLHKNLQKQ-QVYGGNTTVKRQISGTASSVAFTPLQGLEIVNPQAAERS	480
S._pom	IDSRTKLRLPKARKAQLQSMQAQ---KNPLAA----SGLQSSLSFTPIQGIELVNPQLQRQQ	491
A._tha	VPSKLKIN-AKVAKKLKERQ---YAGGATT----SGLTSSLAFTPVQGIELCNPPQALGL	492
NOP	-----	
H._Sap	VAEANQKYFSSMAEFLKVKGEKSGLMST--	499
S._cerv	NPEEETNWFSGHG-----	493
D._mel	QTEANAKYFSNTSGFMSVGQRTT-----	501
S._pom	KVEEANKWFRD-GVFTQIKKDSNEPKNKFS	518
A._tha	GSGTQSTYFSESGTFSKLKKI-----	511
NOP	-----	

Figure 4.17 Protein alignment of human full-length PRPF31 (DKFZP566J153 protein, residues 1 to 499) with PRP31-like proteins from the CG6876 gene product of *D. melanogaster*, *A. thaliana* hypothetical protein T13D8.1, prp31_ from *S. pombe*, and Prp31p from *S. cerevisiae*, and Nop, a putative snoRNA binding domain. Amino acids which exhibit complete identities across species are in shown red, whereas amino acids identical to those of the Nop snoRNA domain are boxed. The region between residues 351 and 364 of PRPF31 (highlighted in blue) with the consensus sequence 351(R/K)(R/K)(R/K) RXX(R/K)(R/K)XRKXKE364 harbors three putative nuclear localization signals (NLS) in agreement with the proposed consensus sequence, K(K/R)X(K/R), for monopartite NLSs (Chelsky *et al.*, 1989). Red circles (●) indicate the most likely signal sequence of the three according to PSORT, a program that identifies subcellular localization signals.

4.5.1.1 Pre-mRNA splicing

Pre-mRNA splicing is an essential process in eukaryotic gene expression. Splicing of nuclear pre-mRNA is catalysed by a large ribonucleoprotein complex known as the spliceosome. This macromolecular machine carries out the excision of introns; the non-coding intervening sequences in the genes, from eukaryotic pre-mRNAs and ligates together exons to create a continuous coding sequence. There are two types of nuclear pre-mRNA introns. The major type, the most common one, has the sequence GT-AG found as a consensus in the introns 5' and 3' ends, respectively. On the other hand, the minor type has the sequence AT-AC at the introns 5' and 3' ends. The major and the minor spliceosomes mediate splicing of the major and the minor introns respectively.

snRNP particles: The components of the major spliceosome are four RNA-protein assemblies called the small nuclear ribonucleoprotein particles (snRNPs) or 'snurps' for short (Burge *et al.*, 1999). They are named after their uridine rich snRNA components: U1, U2 and U5 snRNPs contain U1, U2 and U5 snRNAs respectively, whereas the U4/U6 snRNP contain U4 and U6 snRNAs that are extensively base-paired. The base paired U4 and U6 form a phylogenetically highly conserved Y-shaped U4/U6 interaction domain, consisting of two intermolecular helices (stem I and II) separated by the 5' stem loop of U4. Splicing of the minor intron type is mediated by the minor spliceosome composed of snRNP particles U11, U12 and U4/U6 and U5 snRNP particles also common to the major spliceosome. The protein component of snRNPs is classified into two main groups. Group 1 consists of specific proteins that are found only in a given snRNP and group 2 consists of the core proteins that are common to all snRNPs. In HeLa cells seven such core proteins have been identified: B/B', D1, D2, D4, E, F, and G. These proteins are also called Sm proteins because of their reactivity with Sm serotype from

patients with systemic lupus erythematosus. The core Sm proteins are assembled at the Sm site, a uridine rich sequence present in U1, U2, U4 and U5 snRNAs (Will and Lührmann, 2001). As for the particle specific proteins there are many, each snRNP particle contains approximately 6-10 particle specific proteins. For example the human 13S U4/U6 snRNP contains five proteins specific to this particle. They are three proteins of 20kDa (encoded by *PPIH*, cyclophilin H), 60kDa (encoded by *PRPF3*) and 90kDa (encoded by *PRPF4*) that form a stable heteromeric complex (20/60/90K complex), another 15.5kDa protein that possesses a RNA binding domain binds directly to U4 snRNA and finally the 61kDa protein encoded by *PRPF31*, the *RP11* gene.

Other non-snRNP proteins are also involved in splicing. These proteins have functions in RNA annealing activities, RNA recognition, and pre-mRNA processing activity and helicase activity. This class contains many of the Prp (pre-mRNA processing) proteins and DNA helicases, as well as the initiation factor eIF-4A known for its RNA-unwinding activity (Tsonis, 2003).

Pre-mRNA splicing process: The pre-mRNA splicing consists of two highly conserved sequential trans-esterification reaction steps. (Kramer, 1996; Burge *et al.*, 1999). Splicing begins with the recognition interaction of the complementary sequence within the U1 and U2 snRNP with the conserved sequence at the splice site and branch point, respectively, forming the pre-spliceosome (complex A). The mature spliceosome (complex B) is formed by the subsequent association of U4/U6•U5. The catalytic active state of the mature spliceosome requires conformational changes in the RNAs of the U4/U6•U5 tri-snRNP. U4/U6 snRNA binding proteins play an important role in stabilisation/destabilisation of the U4/U6 RNA duplex. The intermolecular helices of U4/U6 snRNP (stems I and II), separated by the 5' stem-loop of U4 are disrupted during spliceosome activation. Upon

spliceosomal assembly the extensive base pairing between U4 and U6 snRNAs is unwound, and U6 snRNA subsequently base-pairs with U2 snRNA and the 5' splice site forming the catalytic centre of the spliceosome, thereby displacing U1 snRNA. Conformational rearrangements in the first trans-esterification reaction result in the cleavage of 5' exon, and the formation of a lariat intron intermediate. A highly conserved loop in U5 snRNA aligns the 5' and 3' splice site for the second trans-esterification reaction. Following the excision of the intron, ligation of the two exons takes place to form the mature mRNA species (Will and Luhrmann, 2001, Nottrott *et al.*, 2002) (figure 4.18).

Function of PRPF31 within the U4/U6 snRNP: Studies have showed PRPF31 as the human functional equivalent of yeast PRP31p and being integral to the U4/U6•U5 tri- snRNP complex (Teigelkamp *et al.*, 1998). U4/U6•U5 trimer is an important particle in the formation of the spliceosome complex during mRNA splicing as described earlier. PRPF31 was reported as an U4/U6-specific protein and biochemical analysis of the spliceosome complex showed that PRPF31 interacts with an U5-associated 102K protein (Makarova *et al.*, 2002). Thus, the primary role of PRPF31 is thought to be to recruit and physically tether U5 to U4/U6 to yield the tri-snRNP particle.

According to studies in *S. cerevisiae*, PRP31p is also involved in recruiting the U4/U6•U5 tri-snRNP to the prespliceosome complex and/or in stabilising these interactions to form the mature spliceosomal complex (Weidenhammer *et al.*, 1997). However in yeast the role of PRP31p is thought to be essentially one of tethering the tri-snRNP to the spliceosome complex rather than being integral to an snRNP particle.

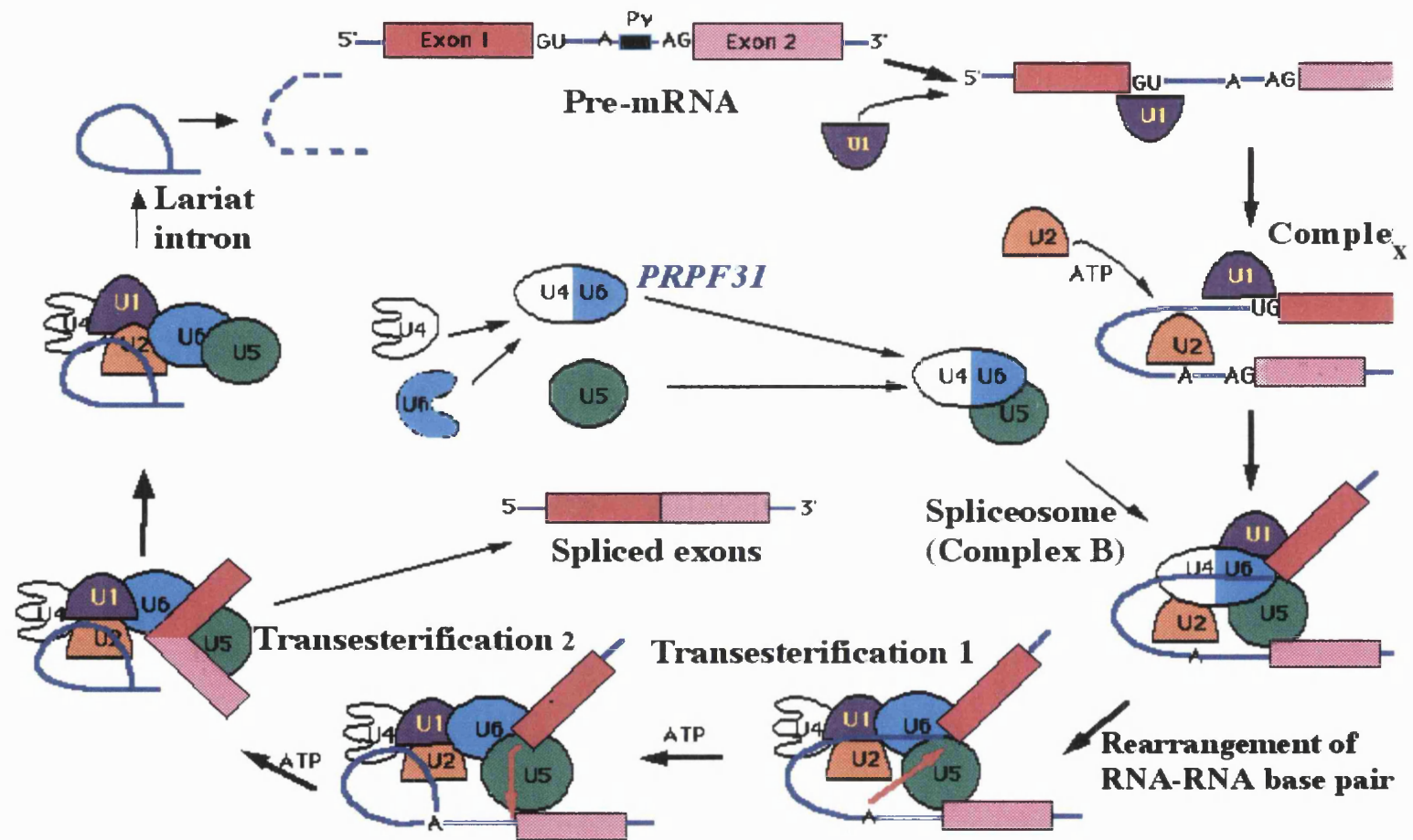


Figure 4.18 Schematic diagram shows the splicing cycle

4.5.2 *PRPF31*, the gene responsible for *RP11*

At the start of this study the genomic organisation was established for a number of genes within the 3093M3 BAC clone, such as *LENG5*, *LENG4*, *LOC91663* and *PRPF31*. The choice of genes to be screened for mutations in the *RP11* families was based on gene expression and function. Priority in screening genes was given to those expressed in the retina or genes encoding proteins most likely to be involved functionally in the eye or the retina.

Mutation screening of *PRPF31* gene resulted in the identification of mutations in *RP11* linked families. Further screening of the *RP11* families and a panel of RP patients led to the identification of mutations in six out of seven *RP11* linked families and additional four mutations in sporadic and small RP families (table 4.5). Despite the fact that neither the expression nor function of this gene is exclusive to the retina, identification of mutations in the majority of *RP11* linked families provides convincing evidence that *PRPF31* is indeed the disease-causing gene in this form of dominant RP.

Table 4.5 Mutations detected in the *PRPF31* genomic sequence.

Family/ Individual	Exon/ Intron	Nucleotide Alteration*	Protein Alteration/Predicted change
RP677	Intron 6	IVS6-3 to -45 del	Inactivation of splice acceptor site
RP1907	Intron 6	IVS6+3A>G	Inactivation of donor splice site
ADRP29	Exon 7	646G>C	Ala216Pro
ADRP5	Exon 11	1115-1125del	Frameshift, 98 novel aa then STOP.
SP14	Exon 7	580-581dup33bp	Frameshift and insertion of 11 aa inframe.
SP42	Exon 7	581C>A	Ala194Glu
SP117	Exon 8	769-770insA	Frameshift, 20 novel aa then STOP
Japanese family	Exon 11	1141-1142delG	Framshift, 31 novel aa then stop
Greek family	Exon 7	632G>A	Arg211Gln
ADRP2		<i>OSCAR</i> , and <i>NDUFA3</i> , <i>TCF3</i> <i>PRPF31</i>	Deletion size~ 30kb
* Nucleotide designation commencing 1 at position 36 of GenBank entry AL050369			

4.5.3 Possible functional consequences of *PRPF31* mutations

The majority of the mutations identified in the *RP11* gene are deletions, insertions and splice site mutations (table 4.5). Such changes normally result in a frameshift of the transcript. As a result, the mutant allele could produce an aberrant protein product that has lost its function, or it could give rise to an unstable mRNA product that may get degraded by nonsense-mediated decay (NMD). These results suggest that the pathophysiological basis of adRP at this locus is functional loss of one allele resulting in haploinsufficiency.

This argument has been strengthened further by the functional studies carried out in our laboratory on the A194E and A216P mutations identified in ADRP29 family and in SP42, respectively. This study showed that these two mutations did not exert a dominant negative effect on the wild type protein (Deery *et al.*, 2002). The authors focused on studying the effects of these two missense mutations on the localisation of the PRPF31 protein, growth of transfected yeast cells and on the splicing machinery. Expression of both the PRPF31 mutants in COS7 cells followed by immunofluorescence microscope analysis showed accumulation of the mutant PRPF31 protein in the cytoplasm with a weak signal in the nucleus when compared with the wild type protein. This result was also confirmed by western blot analysis of both the mutant and the wild type PRPF31. SDS-PAGE analysis of whole cell extracts (soluble fraction cytosolic) and nuclear proteins from cells transfected with pTriEx-1 expression construct of wild type and mutant PRPF31 showed that less protein was present in the nucleus compared to that of the wild type. Moreover, western blot analysis of the cytosolic extract from mutant and wild type transfected cells showed that high proportion of the mutant protein was insoluble. These results indicate that both missense mutations affected the solubility and localisation of the protein, hence, affecting the level of functional protein in the nucleus.

Screening a panel of EMS-generated temperature-sensitive yeast mutants, defective in pre-mRNA led to the identification of the JWY2857 yeast strain with a defect in the *PRP31* gene (Maddock *et al.*, 1996). This strain was able to grow at the permissive temperature of 23°C but failed to grow at the restrictive temperature of 37°C. Transformation of this strain with plasmid expressing the wild type yeast PRP31p fully complemented the defective gene. Deery and colleagues also tested the functional equivalence of the yeast and the human protein by transforming the *PRP31* deficient strain with a plasmid expressing the human PRPF31 protein. However the human protein was unable to complement the function of yeast PRP31p. This failure maybe explained by the low degree of sequence similarity between yeast PRP31p and human PRPF31 (24.74% identity; see section 4.5.1) leading to potential problems in protein-protein interactions within the spliceosome.

The A216P mutation, identified in ADRP29 family, occurs at a highly conserved residue across species, including *S.cerevisiae* in which the A226 residue is the equivalent site (figure 4.17). Deery and colleagues (2002) then investigated the effect of the A226P substitution in yeast PRP31p on its ability to complement the defect in the PRP31p-deficient strain. They found that the strain containing the A226P mutation was able to grow at both 23 and 37°C although, it grew more slowly at the higher temperature than when complemented with the wild type yeast PRP31p. This finding suggested that the A226P mutant does not fully complement the temperature-sensitive mutant when grown at the restrictive temperature of 37°C. The A194E mutation was also investigated in a similar manner. When the leucine residue at position 204 in yeast PRP31p, the equivalent of the A194 residue in human PRPF31, was replaced with glutamate the resulting expression construct was found to fully complement the temperature-sensitive strain. Thus the yeast L204E variant would appear to have sufficient activity to support

normal growth at 37°C. This can be explained on the basis that the position is not highly conserved (see figure 4.17).

These yeast complementation studies along with the results from the localisation experiments suggest that the mutant alleles might have reduced functional activity. Although these findings apply to yeast, future animal models should provide more information of the effects of these missense mutations *in vivo*. Additionally toxicity to the yeast cells was not shown when mutant PRP31p was over expressed, which suggested that having lower levels of the protein is pathogenic to the cells and not the opposite. This finding opens the possibility of gene therapy for the *RP11* phenotype by increasing gene product of *PRPF31*.

Finally, the effect of over expressed mutant protein (in the presence of the endogenous wild type protein) on splicing of bovine rod opsin was also tested in human HEK293T cells. There was no sign of the presence of any aberrant products or a failure of the splicing process by the mutant PRPF31 (Deery *et al.*, 2002). This indicates that over expression of the mutant allele had no effect on the normal function of the wild type protein, hence eliminating the possibility of a dominant negative effect of the mutant allele over the wild type allele. It is worth mentioning that having a dominant negative effect by mutations in such essential proteins or homozygous mutations would be lethal to the cell.

The third missense mutation identified in the *PRPF31* gene was found in a family of Greek origin (R211Q). Functional analysis of this mutation failed to show a significant difference between the mutant and the wild type protein (personal communications: S. Wilkie). However, it is still possible that this mutation may have an effect on folding or functional properties of the protein. Further investigation is required to test such effects.

4.5.4 Why only the retina? Physiological implications of *PRPF31*, on photoreceptor cells

Although the *PRPF31* gene is widely expressed and would be expected to play a central role in mRNA processing in most tissues, it is only within the rod photoreceptors that haploinsufficiency appears to result in a disease phenotype. The answer may lie in the biology of the rod photoreceptor cells themselves. The disk membranes in the outer segments of rods are continuously renewed and it has been shown that in vertebrate retinæ the level of mRNA for opsin, the most highly abundant transcript (Hamm and Bownds, 1986) in rod cells, fluctuates with a daily rhythm. The highest level of mRNA for rod opsins has been observed before light onset prior to disc shedding and during disk assembly (Korenbrod and Fernald, 1989; von Schantz *et al.*, 1999). There is therefore a very high demand for rod opsin synthesis at these times. Consequently, any defect in the splicing machinery may severely affect the photoreceptor cell, resulting in reduced levels of rod opsin production required for the maintenance of rod outer segment ultrastructure and phototransduction pathway. Alternatively, insufficiency may lead to incomplete intron removal and the production of aberrant proteins suggesting that the high rate of rod opsin production may make the rod photoreceptors particularly susceptible to such a mutant effect. Also, since photoreceptors are terminally differentiated cells, the accumulation of such aberrant protein overtime may also cause cell death. The apparent lack of any effect in all other cell types suggests that the cellular level of *PRPF31* from one wild type allele is sufficient for normal cell function.

Since the identification of *RP11* gene, more splicing factors have been reported to cause RP (see section 4.5.5). This confirms that photoreceptor cells in particular are sensitive to splicing deficiency.

4.5.5 Splicing factors causing other forms of adRP

Following the identification of the *RP11* gene, two other splicing factors were found to cause adRP. Mutations in *PRPF8* and *PRPF3* were found to be responsible for pathogenesis in *RP13* and *RP18* respectively (McKie *et al.*, 2001; Chakarova *et al.*, 2002). The *RP13* locus was first mapped by linkage analysis in a South African family of British ancestry (Greenberg *et al.*, 1994). Subsequently, the number of linked families increased to five from different ethnic origins. Physical mapping approach was used in gene identification within the 3cM genetic interval between markers D17S1529 and D17S831. Mutation screening of candidate genes led to the identification of mutations in the ubiquitously expressed *PRPC8* gene, renamed as *PRPF8* (McKie *et al.*, 2001). This gene was found to consist of 42 exons spanning approximately 36kb of genomic sequence and encoding a protein of 220kD. To date, all mutations identified in this gene are in the last exon, which suggests an important functional role for this exon. *PRPF8* was found to be the human orthologue of yeast splicing factor Prp8p (Brown and Beggs, 1992). In yeast, this protein is thought to be the core component of the U5snRNP required for the formation of the U4/U6•U5 tri-snRNP, which is an important element in the assembly of the spliceosome complex (Brown and Beggs, 1992). Additionally, a region near the C-terminus approximately 400 amino acids away from the mutation site was shown to crosslink with the 5' splice site consensus sequence. Prp8 was also thought to have a role in unwinding of the U4 and U6 snRNP during activation of the spliceosome to the catalytic complex C (Kuhn and Brow, 1999). Therefore, in yeast Prp8 has essential role in catalysing and regulating mRNA processing.

Initially, linkage analysis on a large Danish family localised the *RP18* locus to a 2cM interval on chromosome 1q between markers D1S442 and D1S2858. Further haplotype analysis in a British family

refined the *RP18* critical interval to 1cM between markers D1S442 and D1S498 (Inglehearn *et al.*, 1998). Unlike *RP11* and *RP13*, the *RP18* gene was identified through candidate gene screening following the reports of two splicing factors involved in RP. Chromosomal localisation of *PRPF3*, previously known as *HPRP3* on chromosome 1q21.1 made it a good candidate gene for RP. The gene consists of 16 exons and encodes a protein of 683 amino acids with a calculated molecular weight of 77kDa. Mutation screening of *PRPF3* in the British family identified a missense mutation in exon 11 (1482 C>T) which results in amino acid substitution of threonine to methionine at codon 494 (Thr494Met). The same mutation was also found to segregate in the affected individuals of the Danish family, in which the original linkage to chromosome 1 was identified (Xu *et al.*, 1996). Additionally three individuals were also found to have the same mutation. Another missense mutation in exon 11 (1478C>T) causing the non-conservative amino acid change Pro493Ser was identified during a screen of a panel consisting of simplex RP cases.

In order to investigate the possibility that the Danish family and the other three individuals carrying the Thr494Met mutation share a common ancestral haplotype or whether they represent independent mutation events, six markers located within the *HPRP3* gene were typed in all affected individuals. This revealed two different haplotypes segregating with disease among the English and the Danish family. This data provides convincing evidence that the T494M (1482C>T) mutation arose independently in at least the Danish and the large English family indicating the possibility of a mutation hot spot at position 1482.

The specific function of *PRPF3* protein is not yet known, however, it is known to be a protein specific for the U4/U6 snRNP particle and that it interacts with *PRPF4* (Gonzalez-Santos *et al.*, 2002). The mutations are localised in a highly conserved central domain of *PRPF3*, which

according to recent studies is important in maintaining association with the U4/U6 snRNA (Gonzalez-Santos *et al.*, 2002).

The fact that three pre-mRNA splicing factors (*PRPF31*, *PRPF3* and *PRPF8*) are involved in adRP suggests the existence of a novel mechanism of photoreceptor degeneration. It also indicates that these splicing factors, due to their key role in splicing, play a critical role in the maintenance of normal rod function and structure. Until the exact disease mechanism is established, the reasons behind photoreceptor cell death remain a mystery and open to speculation. However, functional evidence so far indicates that the effect on down stream proteins is most likely to be the cause of disease in these forms of RP rather than defects in the protein itself.

In conclusion, mutations in the splicing factor *PRPF31* were found to be responsible for this form of RP localised to chromosome 19q13.4. *PRPF31* mutations trigger a novel mechanism of photoreceptor degeneration distinct from those operating in other forms of adRP, and suggest that *PRPF31* play a critical role in the maintenance of normal rod function. Currently, work is ongoing to fully understand the pathology of RP at this locus with special reference to resolving the cause of partial penetrance in this disease. Preliminary experiments were carried out during this project in an attempt to understand this phenomenon (see chapter 5).

CHAPTER 5

Genetic and biochemical characterisation of the partial penetrance phenotype of the *RP11* locus

5.1 Introduction

Phenotypic variability is an established fact for autosomal dominant retinitis pigmentosa (adRP), as different clinical presentations have been observed along with genetic heterogeneity. In accordance with this fact, the clinical features observed in adRP families linked to chromosome 19q are also quite distinct and unique (Moore *et al.*, 1993; Evans *et al.*, 1995). The adRP phenotype of *RP11* has been classified as R type (regional/ typeII) due to the presence of matching areas of both cone and rod functional loss (Lyness *et al.*, 1985) but it is the presence of asymptomatic obligate carriers, indicating partial penetrance of the disease trait in *RP11*, which sets this locus apart from all other adRP loci (Moore *et al.*, 1993; Evans *et al.*, 1995). Autosomal dominant inheritance in these families is evident with symptomatic individuals in each generation and male-to-male transmission. All symptomatic individuals have severe and early onset disease. Typical fundus features of extensive peripheral degeneration were observed in even the youngest examined symptomatic patient who was 24 years of age. Older symptomatic individuals also showed macular atrophy, oedema and secondary cataract formation. The asymptomatic (obligate) disease haplotype carrier

individuals had normal fundi appearance and minimal or no psychophysical or electroretinographic abnormalities.

The difference in clinical presentation between members of the same family, as shown in families linked to the *RP11* locus, is known as variable expressivity. Variable expressivity and 'R type' adRP are also features of the dominant retinitis pigmentosa locus linked to chromosome 7p (*RP9*) (Kim *et al.*, 1995). However, the variable expressivity phenotype described for *RP9* is a graded disease among affected members of the same family. The clinical presentation of the disease ranges from mild, moderate and severely affected and is independent of age. This contrasts with the unusual polarity of phenotype seen in *RP11* where individuals with the disease haplotype seem either to be severely afflicted or are asymptomatic. Both *RP9* and *RP11* therefore depict two contrasting phenotypes of incomplete penetrance, the 'all or nothing' form of incomplete penetrance shown in that *RP11* led to the description of the *RP11* phenotype as 'bimodal expressivity'.

Further analysis of *RP11* families also suggested that although the offspring of parents with the disease haplotype were at 50% risk of having the genetic defect, the risk of being symptomatic was only 35%. The disease penetrance defined as the proportion of affected individuals among the susceptible disease gene carriers, for *RP11* was estimated at 0.7. This value was derived from the haplotype data of the complete pedigree of ADRP5; the largest *RP11* linked family known to date.

5.1.1 Possible causes for the partial penetrance phenotype in *RP11* families

The penetrance of disease can be influenced by genetic as well as environment factors. Environmental influence as the cause of reduced penetrance in *RP11* has been ruled out since it is more likely to produce a

graded rather than the all or nothing phenotype, typical of *RP11* (Evans *et al.*, 1995). Moreover, environmental factors have not been proved to have that great an influence on disease severity in RP (Heckenlively, 1988).

Genomic imprinting and anticipation are two genetic mechanisms associated with variable expression of disease phenotype. Analysis of *RP11* families failed to show any association between the sex of the disease-carrying parent or the sex of the carrier individuals with the disease phenotype, therefore excluding genomic imprinting and also the possibility that the partial penetrance phenotype is sex related. Anticipation is a phenomenon in which the severity of the disease is increased in successive generations. However, this was also not observed in *RP11* families, thus excluding anticipation to be the cause of the variation in the *RP11* phenotype.

The possibility that the *RP11* phenotype is due to digenic inheritance was also proposed. Digenic inheritance describes the need for a second mutation or polymorphism in a second locus for the expression of the disease phenotype. In the cases of digenic diseases symptomatic individuals are more likely to have symptomatic children rather than disease haplotype carriers, however, this has not been observed in the *RP11* linked pedigrees. Close analysis of the ADRP5 and the ADRP29 pedigrees showed that higher percentage of symptomatic children were born to asymptomatic rather than symptomatic individuals (76% as opposed to 50%). However, the difference in percentage was not significant (Evans *et al.*, 1995).

The results of a study carried out by McGee *et al.*, (1997) does support the hypothesis that the wild type alleles at the *RP11* locus or a closely linked locus (i.e. modifier gene), inherited from the non-carrier parent (i.e. the unaffected-normal parent) is a major factor influencing the penetrance of disease at this locus. They studied the correlation

between the inheritance of wild type alleles from the non-carrier parents and the presence of disease in carrier offspring by sib pair analysis with data derived from ADRP5, ADRP29 and the Japanese family. They found that pairs of carrier siblings with the same phenotype always inherited the same haplotype from the non-carrier parent. Among pairs of carrier siblings with contrasting phenotype (i.e. one affected and one unaffected) all inherited a different haplotype from the unaffected parent.

In hereditary elliptocytosis (or haemolytic anemia) the disease penetrance is specified by wild type alleles in *trans* at the disease locus and stands as a good example for allelic effect. Hereditary elliptocytosis is caused by dominant mutations in the gene coding for the α subunit of spectrin (Gratzer, 1994), which is a cytoskeletal protein essential for the normal morphology of red blood cells. Carriers of the dominant spectrin mutation could either be severely affected or be unaffected, which is similar to the *RP11* phenotype. The penetrance of mutations is specified by a high frequency, otherwise silent polymorphism at this locus which determines the relative level of expression of α -spectrin alleles (Wilmotte *et al.*, 1993). An individual with a high expressing wild type allele in *trans* of a pathogenic mutation will be unaffected or very mildly affected whereas an individual with low expressing wild type allele in *trans* would have severe haemolytic anaemia unless the pathogenic mutation by chance is also on a low expressing allele (Randon *et al.*, 1994).

5.2 Aims and methods

Following the identification of mutations in *PRPF31*, the elucidation of the genetic mechanism behind the partial penetrance phenotype typical of the *RP11* locus was of the utmost importance.

The wild type allele of *PRPF31* appears to be a major influencing factor of disease penetrance at this locus. In order to identify the secondary genetic lesion(s), present in the normal copy of *PRPF31* that set the two phenotypic categories apart from each other, symptomatic and asymptomatic individuals of the ADRP5 and the ADRP29 families were analysed as separate entities and their *PRPF31* genomic sequence was compared. These two families were chosen as they contained sib-ships with several asymptomatic and symptomatic disease carrier individuals.

The *PRPF31* expression both at the mRNA and protein level was also investigated in symptomatic and asymptomatic disease gene carrier individuals from ADRP5, the largest *RP11* family, along with normal individuals. The rationale behind this work was to investigate if there was a detectable difference in mRNA or/and protein levels between symptomatic and asymptomatic individuals and to see if the partial penetrance was due to the differential expression of the *PRPF31* wild type allele.

Lymphocyte of both patients groups (symptomatics and asymptomatics) from ADRP5 family were transformed into lymphoblastoid cell lines (European Collection of Cell Cultures, ECACC, Wiltshire, UK). These cell lines provided the major source of RNA in this study, for the comparison of expression levels of *PRPF31* between the two different phenotypic categories. Cell lines from normal individuals within the ADRP5 pedigree were also part of the study and the *PRPF31* expression in these cell lines was compared with that of retina tissues and fresh blood samples from normal individuals to

investigate the validity of using RNA from transformed lymphoblastoid cell lines.

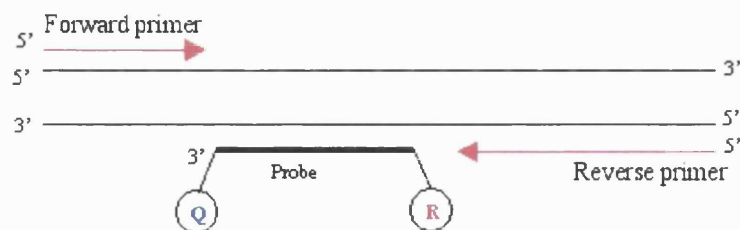
Western-blot analysis was used to compare the levels of PRPF31 protein from soluble cell extracts of the symptomatic and asymptomatic cell lines.

5.2.1 Quantification of gene expression through real time quantitative RT-PCR

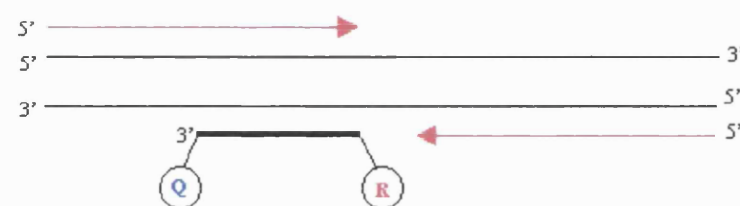
There are several methods available to quantify gene expression; these include Northern blotting, *in situ* hybridisation (ISH), RNase protection assays, cDNA arrays, and reverse transcription polymerase chain reaction (RT-PCR). RT-PCR is a more sensitive method compared to all other traditional RNA assays. This simple and powerful method is most suitable for detection of low amounts of mRNA molecules and therefore the expression of low abundant transcripts. However, RT-PCR is not suitable for very accurate quantification. The real-time quantitative PCR technique has been developed to overcome this limitation. Real-time quantitative PCR (RT-qPCR) is considered to be highly sensitive and precise. The sensitivity of real-time PCR probe allows measurement of PCR product during the exponential phase of the reaction, combining amplification and detection in one single step. Several techniques are available for detection of PCR products; these include agarose electrophoresis with ethidium bromide staining, radioactive labelling and Southern blotting. However, these techniques involve laborious post-PCR manipulation, which increase the risk of contamination. Real-time PCR is very useful in that it detects PCR products with the help of fluorogenic hybridisation probes during the amplification step without the need for purification or separation by gel electrophoresis.

The TaqMan assay takes advantage of a probe technology that exploits the 5'→3' exonuclease activity of Taq polymerase (Holland *et al.*, 1991). The fluorescent reporter dye attached covalently at the 5' end of the probe is FAM (6-carboxyfluorescein), while the reporter quencher dye attached to the 3' end by a linker arm could be either TAMRA (6-carboxytetramethylrhodamine) or DABCYL [4-(4'-dimethylaminophenyl-azo) benzoic acid] (Kreuzer *et al.*, 2001). When the probe is intact the quencher dye absorbs the fluorescence of the reporter dye due to their close proximity. During PCR amplification, 5'-exonuclease activity of the Taq polymerase hydrolyses the probe and the reporter dye is separated from the quencher, resulting in an increase in fluorescence emission. This increase is measured cycle by cycle and is a direct consequence of the amplification process (figure 5.1) (Heid *et al.*, 1996). This method is highly specific due to the complementarity between the set of primers, the internal probe and the target mRNA sequence. The fluorescence signal is generated only when the probe anneals to the target sequence during PCR amplification. High ramping rates, limited annealing and elongation time, the rapid cycle PCR in the detection systems all are very stringent PCR reaction conditions which leads to a primer sensitive and template specific PCR. In this study, quantification of the levels of *PRPF31* mRNA was achieved by using real-time RT-PCR with a fluorescent TaqMan probe on the ABI Prism 7700 sequence detection system or the TaqMan (Applied Biosystems, UK).

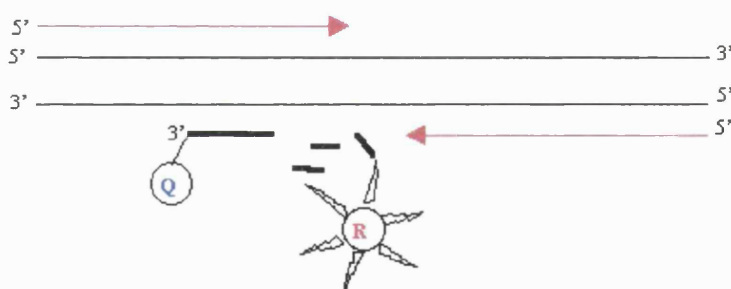
Polymerisation



Strand displacement



Cleavage



Polymerisation completed

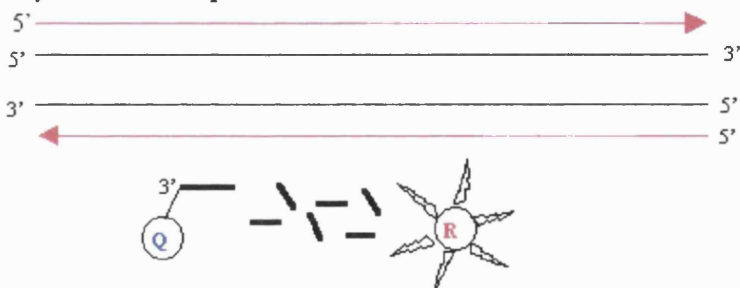


Figure 5.1 Schematic diagram showing the polymerisation associated, 5'-3' nuclease activity of the Taq polymerase during PCR.

The concept of the threshold cycle (C_t) is at the heart of accurate and reproducible quantification using fluorescence-based RT-PCR (Higuchi *et al.*, 1993). As mentioned earlier, fluorescence values are recorded during every cycle. These values represent the amount of product amplified to that point in the amplification reaction. The more the template present at the beginning of the reaction, the fewer the number of cycles it takes to reach a point where the fluorescent signal is first recorded as statistically significant above background (Gibson and Williams, 1996). This point, known as the C_t is always in the exponential phase of amplification (figure 5.2). Therefore, quantification is not affected by any reaction component becoming limited in the plateau phase, which results in a systematic bias against the more abundant templates and makes any quantification based on measurements of overall product yield intrinsically unreliable. The reporter signal is normalised to the fluorescence of an internal reference dye in the Taqman to allow corrections in the fluorescent fluctuations caused by changes in concentration or volume, and C_t value is reported for each sample. This value can be translated into a quantitative result by constructing a standard curve (see section 5.3.3).

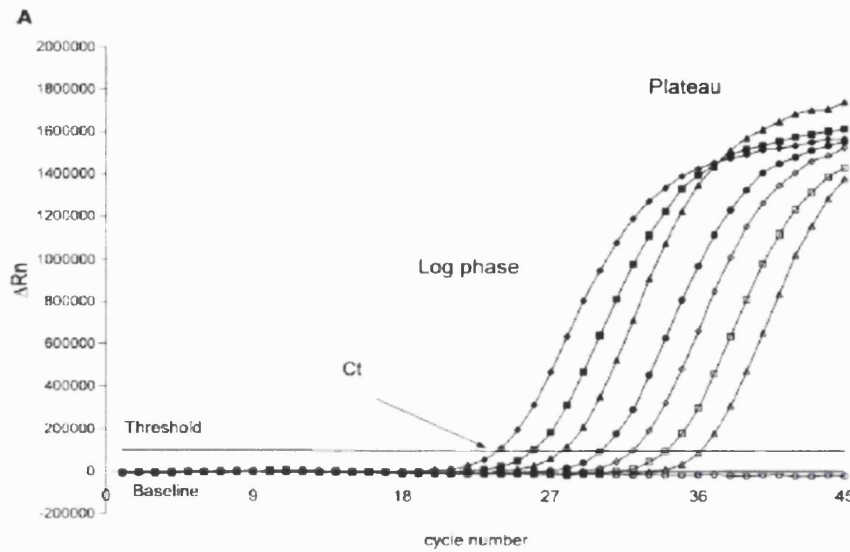


Figure 5.2 Amplification plot showing the log of the change in fluorescence plotted versus the cycle number as documented by the ABI Prism 7700 sequence detection system's analysis software (Applied Biosystems, Warrington, UK).

RT-PCR can be used for relative or absolute quantification. Relative quantification compares transcript abundance of multiple samples, using a co-amplified internal control for sample normalisation (Bustin, 2000). Normalisation to a housekeeping gene is usually used where a correction for minor variations is needed due to differences in input RNA amount or inefficiencies of reverse transcription. The most suitable housekeeping gene used for normalisation should be expressed at a constant level in different tissues of an organism at all stages of development and should not be affected by the experimental treatment itself. The most commonly used genes are β -actin, glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), tubulins, actins, cyclophilin and rRNA. Results are expressed as ratios of the gene specific signal to the internal control signal. This yields a corrected relative value for the gene

specific product in each sample. Choosing the ideal housekeeping gene is important to ensure credibility. However, using housekeeping genes as an internal control for the study can be problematic. Numerous studies have shown that expression of these genes fluctuate in given situations (Lemay *et al.*, 1996, Chang *et al.*, 1998). This might be explained by the fact that housekeeping proteins are not only implicated in the basal cell metabolism but also participate in other functions. As a consequence the use of housekeeping genes as internal controls should be examined carefully in relation to the cell type and the cell metabolism, otherwise they may result in erroneous quantification results.

Absolute quantification, on the other hand, allows the precise determination of copy number per cell, total RNA concentration, or unit mass of tissue. It requires a dilution series of a standard with known starting copy number. This standard is usually a single stranded sense-strand oligodeoxynucleotide that specifies the gene template to be quantified, by plotting the log of the starting copy number against the measured C_t values a standard curve can be generated. On this standard curve, the starting copy number of the gene in an unknown sample can be interpolated. A more detailed description of the way in which a standard curve is constructed is given in section 5.3.3.

PRPF31 expression in the ADRP5 family was quantified by relating the PCR threshold cycles (C_t) to a standard curve, generated by using a synthetic oligonucleotide. In this study, *PRPF31* mRNA copy numbers were related to total RNA rather than a single housekeeping gene.

5.2.2 Why the ADRP5 family

All previously reported *RP11* families show incomplete penetrance. Yet the ADRP5 family was chosen for the expression study as it had a number of features making it the most suitable family of all. ADRP5 is the largest family amongst all of the known *RP11* families; it has 20 symptomatic (affected carrier individuals), 20 asymptomatic (non-affected carriers individuals) and 61 normal individuals (non-carrier individuals) including married in normal (figure 5.3). This allowed the recruitment of a significant number of symptomatic, asymptomatic and normal individuals for this study. Furthermore, blood samples of this family were deposited into the European Collection of Cell Cultures (ECACC, Wiltshire, UK) to be made into cell lines in 1994, which were available upon request. These transformed cell lines were the main source of RNA and protein used in this study.

More importantly, all the disease gene carriers in the ADRP5 family have a deletion in exon 11 of *PRPF31* gene (see section 4.4.2.1). This mutation made it possible to design the RT-PCR-Taqman experiment in such a manner so that only the expression of the wild type allele would be measured. The labelled probe spanned the 11bp-deleted region. Therefore, only expression of the wild type allele in both symptomatic and asymptomatic individuals was measured. Moreover, the PRPF31 antibody used in this study was raised against the C-terminus peptide (amino acid 484-497) (Makarova *et al.*, 2002), which made it specifically recognise the wild type and not the putative mutant protein that may or may not be produced due to the 11bp deletion in ADRP5. This enabled the detection and quantification of only the wild type PRPF31 protein in symptomatic and asymptomatic carrier individuals of the ADRP5 family.

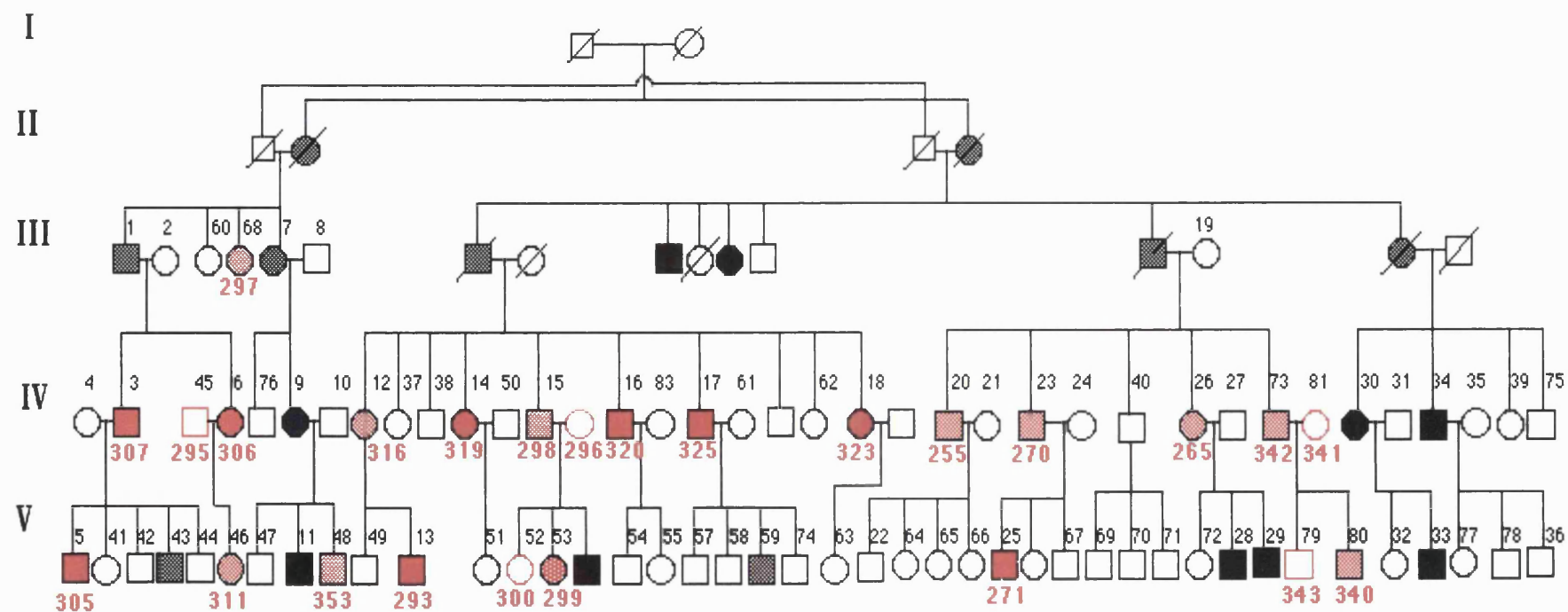


Figure 5.3 ADRP5 family, individuals enrolled for the mRNA/ protein study are in red. numbers in black represents the original family numbering, while numbers in red are cell lines numbers.

5.3 Results

5.3.1 Single nucleotide polymorphisms analysis across *PRPF31* in symptomatic and asymptomatic individuals of ADRP5 and ADRP29

Allelic effect was suggested as the possible genetic mechanism for the reduced penetrance of mutations in the *RP11* locus. In the sib pair analysis carried out by McGee *et al.* (1997), they analysed the correlation between the inheritance of the wild type haplotype from the non-carrier parent and the presence of disease in carrier offspring. Since the haplotype was made up of polymorphic markers across the *RP11* critical interval, the penetrance could be influenced by either the wild type allele of the *RP11* gene or that of a closely linked locus (i.e. a modifier gene). Therefore, in order to investigate whether symptomatic and asymptomatic carrier individuals from a given sib-ship inherit different wild type alleles of *PRPF31* from their non-carrier parent, single nucleotide polymorphisms (SNPs) within *PRPF31* were analysed in these individuals. Typing these polymorphisms in sib-ships of ADRP5 and ADRP29 families showed that symptomatic individuals within a given sib-ship share one wild type allele, whilst asymptomatic individuals show the other wild type alleles from their non-carrier parent (figure 5.4).

Despite the fact that a number of single nucleotide polymorphisms (SNPs) have been identified across the *PRPF31* gene (table 4.4), the genetic change(s) that differentiate the two phenotypic categories apart have not been identified as yet. None of the SNPs identified to date segregate with a given disease status across more than one family or a sib-ship of a given family. For example, an interesting sequence change was identified in the first intron of *PRPF31*, which maybe part of the putative promoter sequence (IVS1+14A>G) with the potential to affect the expression of *PRPF31*. The A nucleotide was found to segregate concordantly with the asymptomatic individuals, whilst the G co-segregated with the symptomatic individuals within ADRP5 family

(figure 5.4). Yet this was only found to be the case in the ADRP5 family but not in other families. This was also the case with another intronic change identified in intron 5 (IVS5+82C>G), which was found to segregate in only two sib-ships of the ADRP5. Moreover, a synonymous (silent) mutation was also identified in the coding sequence of *PRPF31* gene (735C>T/ Pro245Pro), however it was found to co-segregate with the symptomatic and asymptomatic status in only one of the sib-ships of the ADRP5 family (figure 5.4).

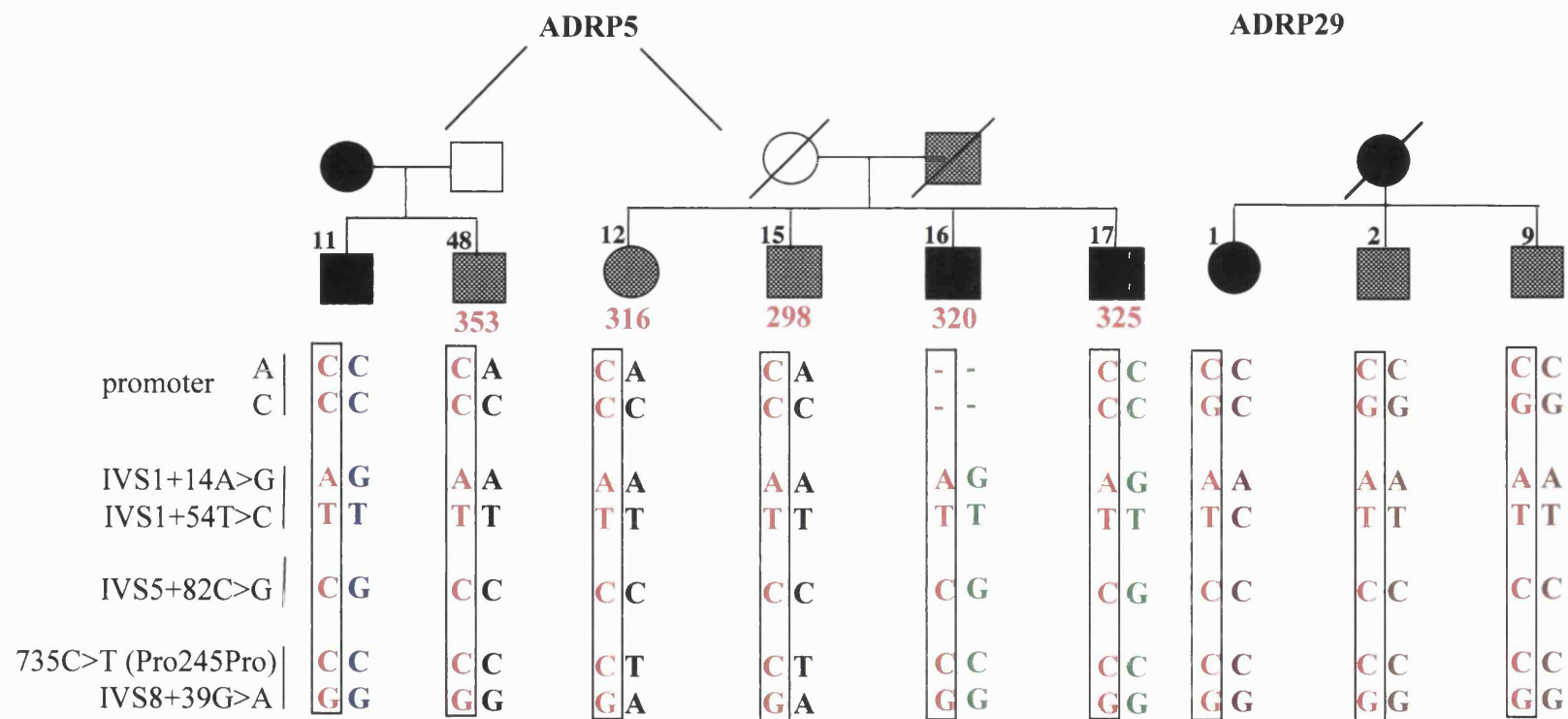


Figure 5.4 Analysis of polymorphisms identified in *PRPF31* gene in two sibships from ADRP5 family. Shaded blocks represent asymptomatic individuals. Numbers in black are the original numbering in the whole family, while numbers in red are for cell lines order numbers (see figure 5.2). The affected allele is in **Red** and boxed, the wild type allele in symptomatic individuals is in **Blue**, **Green**, **Purple** and **Brown** and the wild type allele in asymptomatic individuals is shown in **Black**.

5.3.2 RNA extraction for *PRPF31* expression study

The cohort size for the expression study was initially set at 24, and lymphocyte samples of these individuals (8 symptomatic, 9 asymptomatic and 7 non-carriers) (figure 5.5) were to be transformed into cell lines. The disease status of the individuals enrolled in the study had been established in clinical examinations and has been previously reported (Moore *et al.*, 1993; Evans *et al.*, 1995).

It was essential to get as many symptomatic and asymptomatic individuals for the study to obtain a statistically significant result. Difficulties were encountered at the cell transformation stage; some lymphocyte samples failed to transform and some were contaminated with fungal spores. Therefore, only 20 cell lines were included in this study (8 symptomatic, 7 asymptomatic and 5 normal).

Total RNA was to be used from transformed lymphoblastoid cell lines and fresh blood samples of symptomatic, asymptomatic and normal individuals from the ADRP5 family, to compare *PRPF31* expression in both the lymphoblastoid cell lines and nucleated blood cells of a given individual. The cell lines are a useful resource, and in some respects more appropriate for a comparative study such as this, provided that the cell transformation did not influence the expression of the chosen gene. Cell lines facilitate multiple RNA extractions, which would not be possible with blood samples. Moreover, all cell lines were grown and maintained under the same conditions, allowing for similar RNA yield per cell line. It has been shown that RNA yield from the blood of the same individual differ at different times especially if the immune system has been challenged. Unfortunately, RNA extracted from fresh blood samples of individuals of the ADRP5 family was poor in both quantity and quality and hence excluded from the study. RNA obtained from transformed cell lines was used to compare the levels of *PRPF31* mRNA in symptomatic, asymptomatic and normal individuals with the real-time

PCR technique. However, the validity of using RNA from transformed lymphoblastoid cell line was tested by analysing the *PRPF31* expression in nucleated blood cells of control individuals and comparing the expression with that of cell lines of non-carrier individuals. RNA extracted from retina of control individuals was also used to determine whether *PRPF31* mRNA levels in retina is comparable to nucleated blood cells.

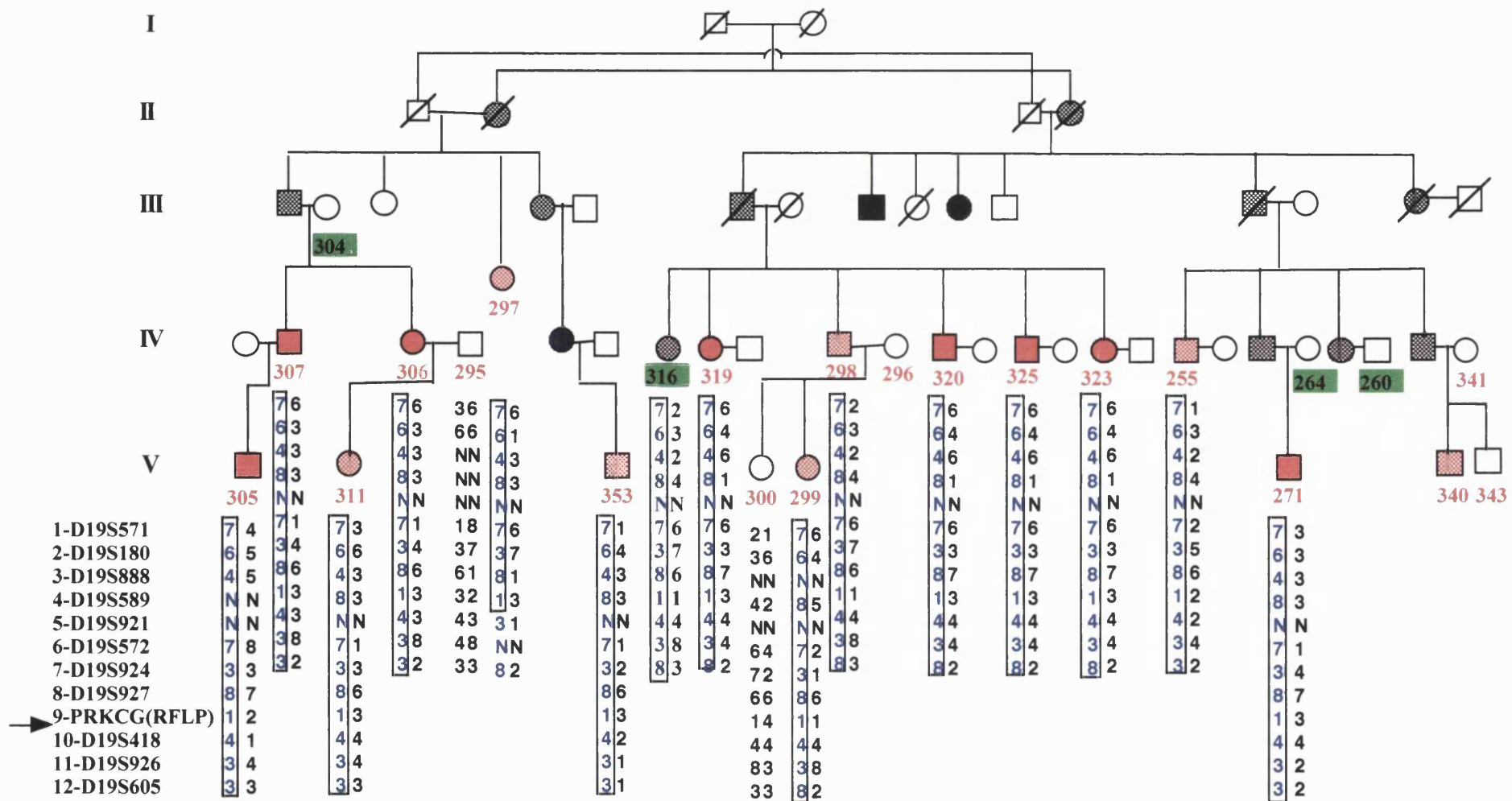


Figure 5.5 The schematic diagram of the ADRP5 pedigree. Individuals enrolled for the *PRPF31* mRNA quantitation study are depicted in red. individuals whose cells failed to transform are highlighted in green boxes. Haplotypes for the *RP11* markers are shown underneath the individuals, with the disease haplotype boxed. The arrow indicates the position of *PRPF31* gene within the *RP11* markers.

5.3.3 Construction of the standard curve

Quantitation of *PRPF31* gene expression was carried out by relating the PCR threshold cycle obtained from samples to a *PRPF31* standard curve. A 110 bp single stranded sense oligonucleotide specifying the *PRPF31* amplicon was synthesised (MWG Biotech, Ebersberg, Germany) and serially diluted from 1×10^9 molecules to 10 molecules and used in TaqMan RT PCR reactions. One microgram of a 110 bp ssDNA contains 1.7×10^{13} molecules.

To maximise accuracy, the dilutions were made over the range of copy number that includes the amount of target mRNA expected in the experimental RNA samples. Two independent serial dilutions were made of the oligonucleotide standard and used for three independent RT-PCR reactions, with each dilution assayed in duplicate (resulting in twelve reactions in total for each dilution), together with negative control (no template) for reactions. Of the C_t values obtained for a given dilution, the highest and the lowest C_t values were discarded to correct for any pipetting errors and the remaining values were averaged to give the final C_t value for that dilution. The C_t values obtained thus are shown in table 5.1 and the standard curve used to calculate the mRNA copy number of *PRPF31* in cell lines was constructed by plotting the mean (see appendix) C_t values against the logarithm of known copy numbers (figure 5.6). The copy numbers (x) of unknown samples were calculated from the linear regression line according to the formula: $\log x = (C_T - b)/m$, where C_T is the threshold cycle, b is the y-intercept, and m is the slope of the standard line. C_t values for the same amplicon concentration from different runs were comparable and RT-PCR reactions consistently failed for dilutions 10^3 - 10^5 . Odd C_t values were encountered only when the fluorogenic probe was near expiry. A different standard curve was used to obtain *PRPF31* expression in blood and the retina due to a time gap between RT PCR assays (formula used $y = -4.096x + 51.135$). However,

samples from cell lines were also included in retina/blood RT-PCR run to investigate the reproducibility of results and to see whether *PRPF31* expression in retina and nucleated blood cells are comparable with that of the cell lines.

For normalisation, the *PRPF31* copy numbers were related to total RNA concentration (in the RT-PCR) rather than to a single housekeeping gene for reasons outlined in section 5.2.1. *PRPF31* expression level is therefore presented as the mRNA copy number per microgram of total RNA.

Table 5.1 C_t values obtained for known input copy number of the *PRPF31* amplicon from three different RT-PCR runs. Each dilution was done in duplicate and the highest and the lowest C_t values were excluded.

Copy number	RT-PCR run	Mean Value
10^9	15.33, 15.22, 15.22, 15.23, 14.79, 15.07, 14.38, 14.49, 14.32, 14.45	14.84
10^8	18.31, 18.23, 18.1, 18.1, 18.07, 18.03, 18.87, 18.2, 18.1	18.22
10^7	22.18, 21.67, 22.11, 22.1, 22.63, 22.29, 21.92, 22.21	22.13
10^6	25.73, 25.5, 25.01, 24.83	25.26

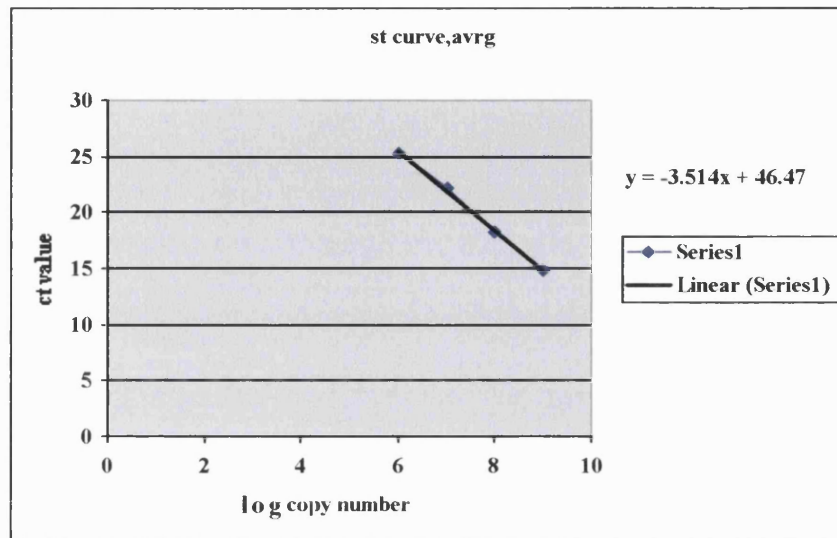


Figure 5.6 Standard curve constructed by plotting the mean C_t values against the known input copy number of a synthesised oligonucleotide (see table 5.1). The regression line formula on the left was used to calculate the mRNA copy number of *PRPF31* in ADRP5 cell lines. Different standard curve and formula were used to calculate *PRPF31* mRNA copy number in retina and blood of normal controls ($y = -4.096x + 51.135$).

5.3.4 Analysis of quantitative RT-PCR on RNA extracted from cell lines.

Figure 5.5 shows the abridged pedigree ADRP5 depicting all of the individuals whose cell lines were included in this study, along with their marker haplotype data (Vithana, 1998).

For each cell line, total RNA was extracted from two independent cell growth passages (A+B). RT-PCR reactions were carried out twice and each RNA extraction was analysed in duplicate (2x(2A+2B)) resulting in eight analyses per cell line. In order to confirm the reproducibility and sensitivity of the RT-PCR assays, dilutions of the

oligonucleotide standard were included in each RT-PCR run. The mean C_t value for each cell line was calculated, following the exclusion of the highest and the lowest C_t values to minimise preparation and run errors, this also prevents bias against the results. *PRPF31* mRNA copy numbers were calculated using the linear regression of the standard curve (figure 5.6).

The non-parametric Mann-Whitney U test (see appendix) statistical method was used to compare the median (see appendix) copy number in pairs of the three groups, namely the symptomatic and asymptomatic carriers of the mutant *PRPF31* allele and non-carrier individuals. The median *PRPF31* mRNA copy number was significantly higher in the non-carrier individuals compared to both the symptomatic and asymptomatic groups ($p < 0.01$) (table 5.2). This was expected, as the non-carrier individuals have two wild type alleles of *PRPF31* gene whilst both symptomatic and asymptomatic individuals have a single *PRPF31* wild type allele and the Taqman probe was designed to detect only wild type alleles of *PRPF31*.

Interestingly, the median mRNA copy number of the wild type *PRPF31* allele was also higher in asymptomatic compared with symptomatic individuals ($p < 0.001$) (figure 5.7). Indeed, a marked difference in mRNA copy number was observed between asymptomatic individual 298 and symptomatic individuals 319, 320, 323 and 325 who were part of the same sib-ship (figure 5.8). Unfortunately, cell transformation failed for asymptomatic individual 316 whose mRNA copy number would have enabled the comparison of all symptomatic and asymptomatic individuals in the entire sib-ship. Symptomatic and asymptomatic individuals in this sib-ship inherit two different wild type alleles from their non-carrier parent (figure 5.3 and 5.5). The mean mRNA copy numbers of the 4 siblings (figure 5.8) (319, 320, 323 and 325) with identical wild type allele show some variability (mean

$2.47 \times 10^7 \pm 3.93 \times 10^6$) with individual 323 showing the lowest copy number. *PRPF31* mRNA levels do not appear to be related to age of individuals although the mean age of the symptomatic group was higher than that of asymptomatic and the non-carrier individuals.

RNA degradation in cell line samples was evident in all subsequent RT-PCR reactions carried out after the second RT-PCR experiment. Moreover, the probe lifetime affected the sensitivity of the assay. This was also noticeable in the RT-PCR results obtained from the oligonucleotide standard in the latter stages of the study. This precluded further assays, however, the results obtained from the two early RT-PCR assays of the cell lines was sufficient to obtain statistically significant results.

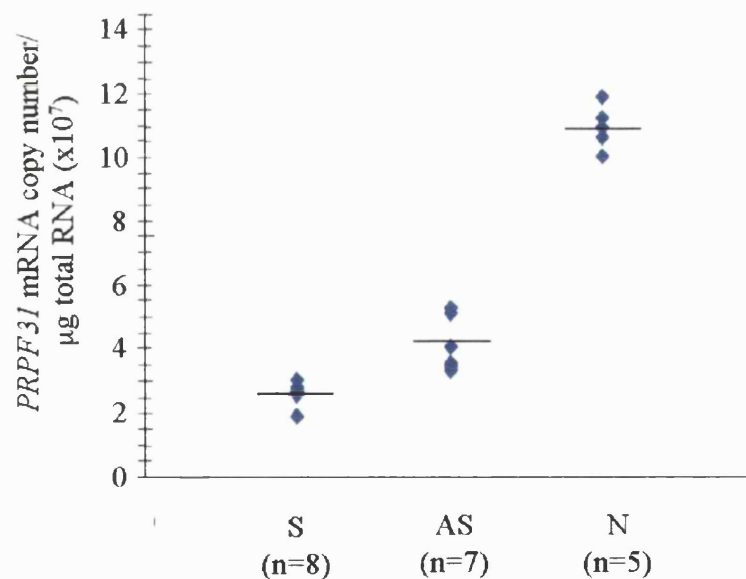


Figure 5.7 Scatter plot showing *PRPF31* mRNA copy numbers in symptomatic (S), asymptomatic (AS) and non-carrier (N) individuals of the ADRP5 pedigree. The horizontal bar indicates the mean copy number.

Table 5.2 *PRPF31* mRNA expression in symptomatic (S), asymptomatic (AS) individuals carrying the (1115-1125 del) *PRPF31* mutation, and in non-carrier (N) individuals of the ADRP5 family. Raw data (i.e. C_t values) are included in appendix I.

Cell line	Clinical status/Age	RNA copy number/ μ g RNA
307	S 58yrs	2.56×10^7
306	S 47yrs	2.97×10^7
305	S 23yrs	2.63×10^7
319	S 64yrs	2.66×10^7
320	S 56yrs	2.75×10^7
323	S 49yrs	1.89×10^7
325	S 55yrs	2.58×10^7
271	S 22yrs mean 47yrs	2.62×10^7 mean $2.58 \times 10^7 \pm 3.09 \times 10^6$
297	AS 76yrs	5.25×10^7
298	AS 61yrs	5.06×10^7
255	AS 56yrs	3.47×10^7
311	AS 23yrs	3.26×10^7
353	AS 23yrs	3.48×10^7
299	AS 26yrs	4.04×10^7
340	AS 21yrs mean 41yrs	5.25×10^7 mean $4.26 \times 10^7 \pm 9.02 \times 10^6$
300	N 32yrs	1.09×10^8
343	N 24yrs	1.06×10^8
295	N 50yrs	1.12×10^8
296	N 50yrs	1.18×10^8
341	N 50yrs mean 41yrs	1.0×10^8 mean $1.09 \times 10^8 \pm 6.71 \times 10^7$

5.3.5 Comparison of *PRPF31* copy numbers between tissue culture cells and nucleated blood cells

PRPF31 copy numbers determined from nucleated blood cells of four British Caucasians showed that *PRPF31* copy numbers between non-carrier lymphoblastoid cell lines and nucleated blood cells were similar and comparable, validating the use of RNA from lymphoblastoid cell lines (table 5.3). A degree of individual variability is apparent in the *PRPF31* copy numbers from blood lymphocytes from different individuals. A larger number of individuals would have to be assayed to investigate the depth of variation found in *PRPF31* mRNA levels in nucleated blood cells in a given population. The *PRPF31* copy numbers/ μ g of total RNA obtained from nucleated blood cells was comparable to copy numbers obtained for *GAPDH* in other studies (Bustin, 2000). This may reflect relative abundance of *PRPF31*.

Table 5.3. *PRPF31* expression in retina, nucleated blood cells and lymphoblastoid cells from non-carrier individuals.

Tissue	<i>PRPF31</i> copy number/μg of total RNA (mean/ SD)
Peripheral retina pool of 4 retinae	1.32×10^8
Blood Lymphocyte BL 1 BL2 BL3 BL4	1.35×10^8 1.27×10^8 1.43×10^8 1.05×10^8
Non-carrier cell lines N1 N2 N3 N4 N5	1.09×10^8 1.06×10^8 1.12×10^8 1.18×10^8 1.0×10^8
Mean $1.21 \times 10^8 \pm 1.39 \times 10^7$	

5.3.6 Quantitative RT-PCR on post-mortem retinal tissue and nucleated blood cells

PRPF31 mRNA levels in retina and nucleated blood cells were measured to see if the two tissues are comparable (table 5.3 and figure 5.9). *PRPF31* copy numbers in the retina compared with blood cells from 4 control individuals were similar, with little variability between the two tissues. Some variability is apparent in the *PRPF31* copy numbers for blood lymphocytes from different control individuals.

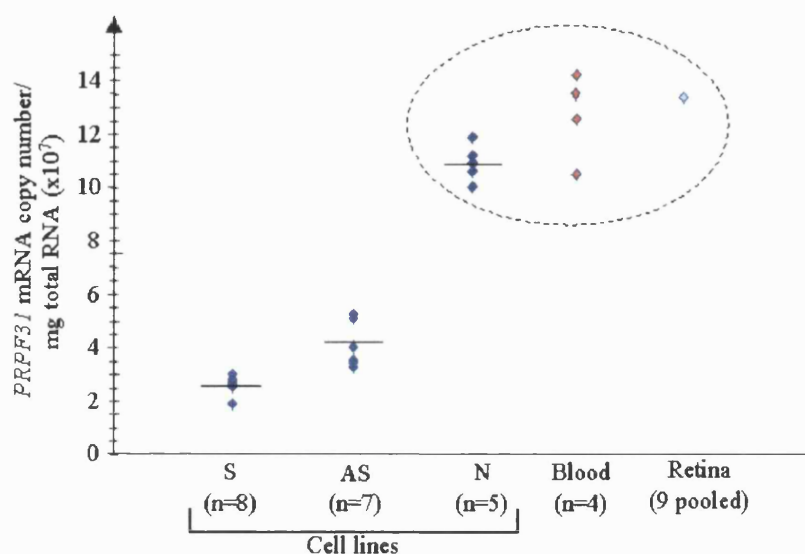


Figure 5.9 Scatter plot showing *PRPF31* mRNA copy numbers in symptomatic (S), asymptomatic (AS) and non-carrier (N) individuals of the *ADRP5* pedigree and blood and retinal tissues from normal individuals. The horizontal bar indicates the mean copy number.

5.3.7 Western blots analysis of PRPF31 protein extracted from cell lines

In order to investigate PRPF31 protein levels between symptomatic and asymptomatic individuals, the soluble fraction of whole cell lysates was examined by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) with Coomassie blue staining and Western blot analysis. The PRPF31 antibody (Anti-61K) used as a probe had been raised in rabbits against a C-terminal peptide (amino acid residues 484–497) and therefore only recognised the wild type protein (499 amino acids) seen as a band of approximately 61kDa on an immuno blot (Makarova *et al.*, 2002). Lack of additional bands suggests the specificity of the probe for the wild type protein, as the smaller (469 amino acids) putative mutant protein is likely

to run at a different molecular weight. The PRPF31 protein band (61kDa) was less intense in soluble cell extracts from symptomatic individuals compared with asymptomatic and non-carrier individuals. This demonstrates lower expression of the PRPF31 protein in symptomatic individuals (figure 5.10 B1 and B2). The band intensities were analysed using the Kodak digital science electrophoresis documentation analysis system 120 and the sum intensities of bands from the symptomatic group was ~50-60% that of the asymptomatic group (where sum intensities are defined as the sum of all pixel intensities in the band rectangle). There were also differences in the sum band intensities between the non-carrier group and the asymptomatic group with the asymptomatic band strength being ~ 80% that of the non-carrier group, although by eye the band intensities appear similar. The corresponding Coomassie blue stained gel showed banding patterns of approximately equal intensity for all samples, confirming uniform loading of samples and thus suggesting that differing PRPF31 band intensities were due to differences in the level of protein expression (figure 5.10 (A1) and (A2)). Similar data were reproduced in an independent experiment in which newly extracted cell lysates of the cell lines from another growth passage were analysed (data not shown).

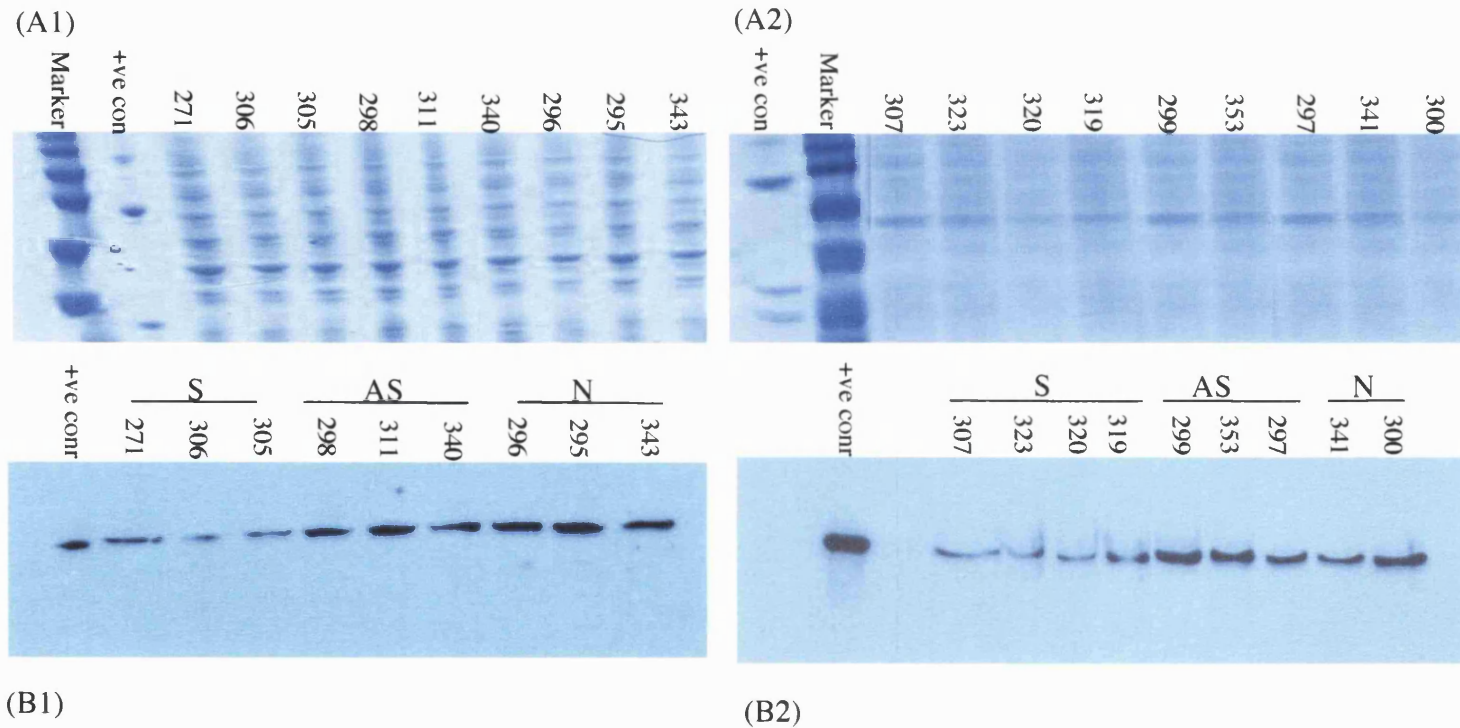


Figure 5.10 Western analysis of PRPF31 protein. Soluble cell lysate ($\sim 5 \mu\text{g}/\text{lane}$) was separated by 10% SDS/PAGE and transferred to nitro cellulose and was analysed with Anti 61K antibody. Coomassie Blue staining of the gel separated by SDS-PAGE showing equal loading (A1 and A2). Immunoblotting of the PRPF31 protein. A single band of $\sim 61 \text{ kDa}$ was detected in symptomatic (S), asymptomatic (AS), and non-carrier/ normal (N) lanes (B1 and B2).

5.4 Discussion

The *RP11* phenotype is of particular interest to clinicians and researchers as it is characterised with the presence of symptomatic and asymptomatic carrier individuals. In this study, the molecular basis of the partial penetrance phenomenon was investigated by studying the expression of *PRPF31* in symptomatic and asymptomatic individuals at both the mRNA and the protein level.

5.4.1 Significance of the mRNA expression study

Expression analysis showed a statistically significant difference in wild type *PRPF31* mRNA copy numbers between symptomatic and asymptomatic individuals carrying a *PRPF31* gene mutation in the largest adRP pedigree linked to the *RP11* locus. The mRNA copy number data were based on mRNA from lymphoblastoid cell lines. The use of cell lines was validated by the observation that cell transformation did not alter *PRPF31* expression in the cell lines as compared with nucleated blood cells. It is important to note that differences in *PRPF31* copy numbers between the two phenotypes were considered and not absolute *PRPF31* copy numbers. Difference in mRNA levels can be directly attributed to an underlying genetic variation that exists in the two phenotypic groups under investigation.

The mRNA data from cell lines indicate that symptomatic patients inherit a relatively poorly expressed *PRPF31* wild-type allele from their non-carrier parent compared with asymptomatic patients. Western blot analysis provided supporting evidence for the mRNA data, which showed correspondingly lower levels of PRPF31 protein in symptomatic cell lines compared to asymptomatic cell lines. Therefore, it appears that the clinical manifestation of RP in ADRP5 is modulated by the low expression of wild type *PRPF31* allele in *trans* with the mutant allele.

On Western analysis, non-carrier and asymptomatic cell line PRPF31 protein levels appeared similar, although the copy numbers of *PRPF31* in non-carrier individuals were approximately twice that of asymptomatic individuals. Based on the sum intensity values of the bands it appears that asymptomatic bands are 80% as intense as the normal/non-carrier PRPF31 protein bands. The correlation between the number of mRNA and protein molecules is generally not strong enough to predict one value from the measurement of the other (Anderson and Seilhamer, 1997; Gygi *et al.*, 1999). Therefore, a high mRNA copy number as seen in the non-carrier individuals may not necessarily translate to a correspondingly high level of protein. It is possible that there are mechanisms in place to maintain a certain steady state or basal level of PRPF31 protein within cells. Therefore, the asymptomatic PRPF31 protein levels maybe close to the non-carrier PRPF31 protein levels, as was observed by Western analysis.

5.4.2 *PRPF31* expression in retina

Another important question to consider is whether the mRNA and protein expression patterns seen in the lymphocytes of symptomatic and asymptomatic individuals can be extended to the rod photoreceptors of these individuals. Comparison of *PRPF31* expression in the peripheral retina and blood lymphocytes showed similar results for these very different tissues. It is reasonable to infer that relatively low PRPF31 protein levels are likely to be present within the rod photoreceptors of symptomatic individuals. The clinical manifestation of RP may subsequently arise due to the sensitivity of rod photoreceptors to the level of the PRPF31 protein at times of increased mRNA synthesis, for example during disc shedding and turnover (see section 4.5.3).

5.4.3 Possible genetic cause for the partial penetrance phenotype of *RP11*

There are several possible mechanisms for the difference in mRNA and protein levels between symptomatic and asymptomatic individuals. Firstly, different transcriptional activity of the promoter maybe a factor. Also, post-transcriptional regulatory events such mRNA translation and decay may also vary for different wild type alleles of *PRPF31*. Other factors include protein isoforms (resulting from different wild type alleles) with different biological half-lives.

Scanning of the *PRPF31* genomic sequence of symptomatic and asymptomatic individuals from ADRP5 and ADRP29 has identified several polymorphisms, both in coding and non-coding sequences. Analysis of these polymorphisms clearly shows that in a given sib-ship the two contrasting phenotypes inherit a different haplotype from their non-carrier parent. However, to date only one change in *PRPF31*, located in intron 1 (IVS1+14A>G), has been shown to segregate concordantly amongst all of the symptomatic and asymptomatic individuals within the ADRP5 pedigree. Since this change has only been found in ADRP5, it is unlikely to be directly involved in the low expression of *PRPF31* but rather may exist in linkage disequilibrium with an as yet unidentified sequence variation(s).

Many studies have reported the importance of intronic sequence variations and their effect on skipping of flanking exons and alternate splicing. Such events may produce a transcript with a premature stop codon that then subjects the transcript to non-sense mediated decay (NMD) (Cartegni *et al.*, 2002). Such sequence variants within the normal copy of a given gene could determine the disease status by affecting the wild type mRNA levels and hence leading to different levels of the wild type protein production.

A similar phenomenon is encountered in erythropoietic protoporphyria (EPP). EPP is a rare autosomal dominant disorder of heme biosynthesis, characterised by partial decrease in ferrochelatase (FECH) activity. FECH is the terminal enzyme of heme biosynthetic pathway and catalyses the insertion of ferrous iron into protoporphyrin IX to form heme. EPP, like *RP11*, exhibits incomplete penetrance and it has been demonstrated that clinical expression of EPP requires the co-inheritance of a wild type *FECH* allele with low expression and a mutant *FECH* allele (Gouya *et al.*, 1999). Furthermore, the underlying cause for low expression has been identified as an intronic SNP, IVS3-48T/C, which modulates the use of a constitutive aberrant acceptor splice site. The aberrantly spliced mRNA is degraded by NMD mechanism, producing a decreased steady state level of mRNA resulting in decreased FECH enzyme activity necessary for EPP phenotypic expression (Gouya *et al.*, 2002).

In the largest adRP pedigree with a deletion mutation in *PRPF31*, the clinical manifestation of RP could well be due to the co-inheritance of a *PRPF31* gene defect and a wild type low-expressed allele. In order to show that this phenomenon is generally involved in adRP caused by mutations in *PRPF31* and not restricted to a single family (ADRP5), more adRP families need to be studied.

5.4.4 Future work

Results of this study pertain to the ADRP5 family alone but similar studies need to be corroborated in other *RP11* linked families to show that this phenomenon applies to all families with mutations in *PRPF31*. Quantification of expression of the wild type alleles in symptomatic and asymptomatic individuals would be difficult in those families that segregate missense mutations (e.g. ADRP29), as the probe would have to distinguish between the mutant and the wild type mRNA based on a

single nucleotide change. Amongst all of the *RP11* linked families, ADRP2 is the most suitable family to confirm the results obtained with ADRP5. This family has a large deletion (~30kb), which encompasses four genes including the *PRPF31* gene up to exon 14 (see section 4.4.2.2.1). Due to this deletion, the same reagents (i.e. primers, amplicon and the antibody) used in this study can be utilised again to measure the expression of *PRPF31* wild type allele in symptomatic and asymptomatic carrier individuals of the ADRP2 family. The mutation in ADRP2 family provides further evidence that it is the wild type allele that is responsible for disease manifestation in *RP11*, as the disease allele is effectively a null allele.

Despite the significance of the results presented in this study they are only preliminary and require confirmation. They also represent the first steps towards resolving the mechanism behind the partial penetrance phenotype in *RP11*. The next immediate task is the identification of the actual genetic cause for the differential expression of wild type *PRPF31*. In order to achieve this, a thorough examination of the entire genomic sequence of *PRPF31* in symptomatic and asymptomatic individuals is required. A systematic analysis would entail the comparison of the genomic sequence of symptomatic and asymptomatic individuals within a given family to first identify the sequence of the mutant allele followed by all the variants of wild type alleles. Then all the wild type allele sequences will need to be grouped into two (symptomatic and asymptomatic) and compared to identify the concordant sequence motif(s) within the non-mutant/wild type alleles of *PRPF31* in a given group. This would be followed by the necessary functional assays to prove that the identified sequence change(s) is associated with differential expression of *PRPF31* mRNA.

5.4.5 Implications for future therapy

In this study, it was shown that the low expression of *PRPF31* mRNA in the symptomatic individuals resulted in correspondingly low protein levels when compared with the asymptomatic and normal individuals. Therefore, clinical manifestation of the disease at this locus could be due to the differential expression of the wild type *PRPF31* allele. This has revealed a potential avenue for future therapy, as it appears that increased expression of wild type *PRPF31* may prevent manifestation of the disease.

Moreover, identification of the actual genetic cause of differential expression of *PRPF31* alleles would enable the design of a PCR based diagnostic tool/protocol that can be used in offspring of carrier individuals to test for the presence of either a high or low expressing wild type allele. This will enable a more accurate prognosis of disease progression in pre-symptomatic children and allow any therapy to be targeted to those patients where retinal disease would be expected. This is highly significant as successful gene therapy in RP patients may depend upon early diagnosis and intervention in the earliest stages of disease.

CHAPTER 6

General discussion and future prospects

6.1 Identification of the *RP11* gene

The *RP11* locus was first mapped in our laboratory in a large British family, ADRP5 (Al-Maghtheh *et al.*, 1994). Following the refinement of the *RP11* genetic interval, work focused on identifying candidate genes that lie within the interval and screening these genes for pathological mutations in the linked families (Vithana, 1998). At the start of this project, the size of the *RP11* critical interval was estimated to be approximately 500kb, confined between the markers D19S927 and D19S781.2. Twenty-one genes have been identified within this interval. This thesis describes the screening of the hypothetical gene *LOC91663*, a gene characterized from an EST that maps within the *RP11* critical interval (chapter 3) and also genes located in the BAC clone 3093M3 (chapter 4). During the course of this work, the *PRPF31* splicing factor gene was identified as the *RP11* disease gene.

Mutations in *PRPF31* have been identified in all but one *RP11*-linked family, and in several sporadic RP cases. These mutations include large and small deletions (sizes range between 1bp–30kb), insertions/duplications (between 1-33bps) and missense mutations, some of which are splice site mutations (e.g. in the RP1907 family, see section 4.4.2.1). ADRP11 is the only linked family where mutations in *PRPF31* have yet to be identified. Small intronic deletion or insertion in this family, and in

other sporadic RP cases may have been missed due to the insensitivity of the PCR based methods used for large-scale mutation detection.

Based on the number of mutations identified in this study, the proportion of adRP attributed to *PRPF31* mutations is estimated to be 3.3% within the British population. This value is likely to be a gross underestimate, due to the insensitivity of the detection method for large rearrangement mutations, which appear to be the more prevalent type of mutations in *PRPF31*. However, since the identification of the *RP11* gene, more adRP families with *PRPF31* mutations have been reported across the world (personal communications). Therefore, the overall data is still compatible with *RP11* being a major locus for adRP.

6.2 Significance of the *RP11* gene

Prior to the identification of the *RP11* gene, the majority of the known RP genes encoded proteins that function within the phototransduction cascade and those involved in vitamin A metabolism (<http://www.sph.uth.tmc.edu/retnet/disease.htm>, table 1.1). Identification of the *RP11* (*PRPF31*), the *RP13* (*PRPF8*) (McKie *et al.*, 2001) and the *RP18* (*PRPF3*) (Chakarova *et al.*, 2002) genes revealed an entirely novel class of genes that can cause RP. These three genes all encode splicing factors that function within the pre-RNA splicing pathway. Therefore, a defective splicing pathway appears to be a novel entry point leading to photoreceptor cell death and retinal degeneration. The reason why only the photoreceptor cells are sensitive to splicing defects is discussed in detail in section 4.5.4.

The fact that many ubiquitously expressed genes are now known to cause RP, such as *RPGR* (Dryja *et al.*, 2001), *RP2* (Chapple *et al.*, 2000), *IMPDH1* (Bowne *et al.*, 2002) and *PIM1K* (Keen *et al.*, 2002), indicates

that photoreceptor cells are very susceptible to mutations in such house-keeping genes. This finding perhaps reflects the finely tuned nature of this remarkable cell and their intolerance towards any biochemical or biological defect. Furthermore, future investigators need to be aware about selection of candidate genes based on their expression profile; rather more thought should be given to their function and the biology of the retina/ photoreceptor cell.

Another example in which defects in a splicing factor cause disease is in the autosomal recessive spinal muscular atrophy (arSMA), a common childhood mortality disorder, where the survival of motor neuron gene (*SMN*) is implicated (Wirth 2000). SMA is characterized by progressive loss of spinal cord motor neurons, which results in denervation of skeletal muscles resulting in weakness, atrophy, and paralysis of voluntary muscles. The *SMN* gene is an ubiquitously expressed gene required for cell viability in vertebrates, and is found in a complex which interacts and masters the assembly of U1, U2, U4, and U5 small ribonucleoproteins (snRNP) particles (Terns and Terns, 2001). In support of the proposed hypothesis that photoreceptor cells, which are neuronal cells, are sensitive to decreased levels of splicing factors, the association between the *SMN* protein levels and the severity of the SMA disease was investigated (Lefebvre *et al.*, 1997; Jablonka *et al.*, 2000). Correlation between *SMN* protein levels, loss of motor neurons and disease severity was shown in both affected individuals and SMA mouse models. This indicates that the terminally differentiated neuronal cells, unlike other metabolically active cells, are highly sensitive to reduced levels of splicing factors and require a certain level of these proteins throughout their lifetime.

6.3 Identification of the genetic cause of partial penetrance phenotype in *RP11*

Based on the mRNA and protein study in the ADRP5 family, mRNA levels of the wild type allele appears to determine the disease status in carrier individuals. Further evidence that the wild type allele is indeed responsible for the manifestation of the disease is provided by the mutation in the ADRP2 family. This family has a large deletion, which spans the *PRPF31* gene and three other genes (see section 4.4.2.2.1). Therefore, no PRPF31 protein is produced from the mutant allele. Hence, disease manifestation in carrier individuals of the ADRP2 is based on which wild type PRPF31 allele they inherit from their non-carrier parent. Despite the genetic evidence, families such as ADRP2 family still need to be investigated at the RNA and protein level to confirm the role of the wild type allele in the manifestation of the disease phenotype and also to show that this phenomenon is generally involved in adRP caused by mutations in *PRPF31* and not restricted to a single family i.e. ADRP5.

In order to identify the actual genetic cause responsible for the differential expression of the wild type *PRPF31* alleles, the entire genomic sequence of the *PRPF31* gene in symptomatic and asymptomatics needs to be compared. Initial comparison of coding sequences has not yet revealed any concordant sequence variants; therefore in future analyses particular attention should be given to the promoter region, intronic and the 3' UTR sequence of *PRPF31*. This may reveal sequence variations that maybe part of a specific element/motif that confers different promoter strength and therefore different transcription rates, or changes in introns and 3'UTR that may affect the mRNA stability. Any one of these could result in different wild type mRNA levels in symptomatic and asymptomatic individuals. For example, it is likely that sequence variations deep within introns may result in alternate splicing (Gouya *et al.*, 1999). A similar case was identified in Erythropoietic protoporphyria

(EPP) (5.4.4). Such intronic changes could result in the production of transcripts with premature termination codons, which will be subjected to nonsense mediated decay (NMD) and therefore result in the reduction of wild type mRNA levels from a given allele. Identification of sequence variation(s) that functionally distinguish the two kinds of wild type alleles would also help in estimating the frequency of these alleles in the normal population.

6.4 Functional and biochemical analysis of *PRPF31* mutations

Understanding the effect of a given mutation on the functionality of the protein and the following consequences on cell biology is very important to understand a given disease mechanism. The functional and biochemical studies of the mutant and the wild type *PRPF31* are necessary to understand the role of the mutant protein in photoreceptor cell death and the disease mechanism.

Two of the three missense mutations identified in the *PRPF31* gene (A216P and A194E) lie within the Nop homology domain of *PRPF31*, a conserved central domain spanning residues 97-328 (figure 4.5). Functional analysis of A216P and A194E mutations showed that the mutant *PRPF31* does not translocate efficiently into the nucleus compared to the wild type protein and therefore resulting in a reduction of the level of functional protein in the nucleus (Deery *et al.*, 2002) (see section 4.5.3). These functional assays have so far supported the hypothesis that the pathophysiological basis of disease in *RP11* is still haplo-insufficiency due to the functional absence at the site of activity (i.e the nucleus) of the product of one allele. Since there is so far no evidence of a dominant negative effect of the mutant product on the wild type it appears that the disease manifestation is reliant on the level of the wild type protein.

However since some transport of mutant protein to the nucleus still occurs (Deery *et al.*, 2002) it is important to investigate whether these mutant proteins are functionally active within the nucleus. This can be investigated by using the known PRPF31 protein-RNA and protein-protein interactions crucial for the dynamics of the tri snRNP and spliceosome assembly. For example previous studies have shown that in the presence of the 15.5K protein, PRPF31 protein binds to the U4 snRNA and the U4/U6 snRNA duplex (Nottrott *et al.*, 2002). Also by two-hybrid analysis, the 61K PRPF31 protein has been shown to interact specifically with U5 snRNP protein 102K (encoded by *C20orf14* gene on chromosome 20q13.33) (Makarova *et al.*, 2002). The two PRPF31 mutants have already been created by site directed mutagenesis (Deery *et al.*, 2002). These recombinant mutant PRPF31 proteins could be used for co-immunoprecipitation experiments with radiolabeled U4 snRNA, 15.5K protein and the anti-61K antibody. Interaction between the mutant PRPF31 and the 102K protein could also be analyzed using recombinant 102K and mutant/ wild type PRPF31.

Functional analysis of the deletions/insertion mutations identified in *PRPF31* is also worth investigating, particularly in cases where the deletion/insertion in the gene cause a frame shift in the coding sequence i.e. the deletion mutation in ADRP5. If a stable mutant protein is produced as a result, such products should also be tested for any functional activity to see whether they are also excluded from the nucleus as are the A216P or A194E mutants.

During the mutation screening process a number of non-pathogenic mutations were identified in some of the exons and introns of the *PRPF31* gene. Some of these exonic sequence variations may have an effect on mRNA stability and some of the intronic polymorphisms may affect the splicing probability of the flanking exon. Therefore, it may be

worthwhile to investigate the effect of these changes on the integrity of mRNA transcript and the protein.

6.5 *RP11* mouse model

Animal models are very important in medical research since they allow detailed examination of the physiological and biochemical basis of a given disease. They are also used in testing novel treatments before conducting clinical trials on humans. Several models, mainly mice, have been created for retinal degenerations. Over the years animal models resulted in much progress in the understanding of disease mechanisms and they have been very useful in gene therapy attempts (Hafezi *et al.*, 2000). Therefore, it is important to develop an appropriate animal model that mimics the human disease in *RP11*.

As mentioned earlier (see section 4.5.1.1), *PRPF31* is an ubiquitously expressed protein and an essential component of the pre-mRNA splicing pathway. Therefore, the *Prpf31* knockout mouse (*Prpf31*^{-/-}) is expected to be lethal. However, heterozygous (*Prpf31*^{+/-}) mouse would model the effect of *PRPF31* mutations and the haploinsufficiency of the disease, and a conditional knockout in which *Prpf31* is targeted for heterozygous deletion only in retina should be generated.

The phenotype of the heterozygous mouse thus generated should be examined for degenerative or other changes in histological sections of the retina. Electroretinography can be used to monitor the progress of visual loss by comparing the electroretinograms (ERG) recorded with brief light flashes and flickering light in wild type and *Prpf31*^{+/-} mice as described by Olsson *et al.* (1992). The wild type and the heterozygote mice from the F1 generation, obtained by crossing the chimera with a C57BL/6

female can be used. These two groups will be completely isogenic except for the mutated locus, which will circumvent concerns regarding comparability of genetic background. The effect of reduced PRPF31 splicing factor protein on splicing and the turnover of retinally expressed genes, especially rhodopsin, could be investigated by Northern blot analysis and RT-PCR of retinal RNA. The role of apoptosis could also be followed in *Prpf31*^{+/-} mouse by standard TUNEL methods.

To explore the possible influence of different wild type *Prpf31* alleles on photoreceptor viability and function, the null *Prpf31* allele can be backcrossed into unrelated inbred genetic backgrounds, e.g. 129Sv, C57BL/6 and CBA (Beck *et al.*, 2000). The choice of inbred strain would be determined according to the relative expression of *Prpf31* wild type alleles between these strains as determined by quantitative RT-PCR. The resulting congenic heterozygote mice can be subjected to the same comparative structural and functional analysis as detailed above. It will be interesting to see if the non-penetrance phenotype can also be observed in mice. Such experiments could provide further insight into the genetic mechanism behind the partial penetrance in *RP11* and will enable further examination into the effect of the wild type and the mutant allele of the transgenic animal on manifestation of the disease.

Knock-in mice with the A216P and A194E mutations could also be created to explore the possibility of a dominant negative effect. Finally, these models would be useful for testing novel therapeutic drugs and they would serve as experimental models for gene therapy.

6.6 Gene Therapy

Gene therapy in principle aims to cure genetic defects by either correcting the genetic material or by introducing genes to counter or correct the consequences of the mutated genes while the actual genetic cause underlying the genetic defect is overlooked. Gene replacement and the introduction of antisense RNA or ribozymes in recessive and dominant disorders, where mutations cause a dominant negative effect or the gain/loss of function, is most likely to ameliorate the phenotype. Retinitis pigmentosa is caused by genetic defects that result in photoreceptor cell death, therefore any proposed therapeutic approach must be based on decreasing the rate of photoreceptor cell loss rather than restoring lost cells (Bessant *et al.*, 2001). Such approach if successful, would be suitable for the majority of retinal dystrophies.

Based on mRNA and protein results presented here (see chapter 5), the clinical manifestation appears to be due to decreased levels of the protein. Moreover, functional studies carried out to date supports the hypothesis that *PRPF31* mutations are not likely to have a dominant negative effect and that pathogenicity is due to haploinsufficiency of the *PRPF31* gene in the photoreceptor cells (Deery *et al.*, 2002). Therefore, increasing the expression level of *PRPF31* by introducing a normal copy of the gene maybe sufficient to ameliorate the disease phenotype. The generation of *Prpf31*^{+/-} mouse would facilitate the examination of gene replacement therapy using either Adeno-associated virus (AAV) or lentiviral mediated gene transfer on rescuing the RP phenotype in this animal. AAV have been widely used in gene therapy trials, they are effective with long lasting results and are known for being safe (Bennett *et al.*, 1999). Gene replacement therapy in the retinal degeneration slow mice (*rds*), homozygous for a null mutation in *Prph2* gene, using an AAV vector, has resulted in the development of the outer segments and has shown an improvement in photoreceptor function (Ali *et al.*, 2000).

This thesis described the identification of the *RP11* gene, probably the second most common adRP gene after rhodopsin. Identification of the *RP11* gene is only the starting point for a large body of work that still remain to find a diagnostic tool for the clinical manifestation of the disease and also a cure for this form of RP. More promisingly, according to data presented in this thesis it appears that gene therapy by introducing an additional copy of the gene may rescue the phenotype in this form of RP.

References

- Al-Maghtheh M., Inglehearn C.F., Keen T.J., Evans K., Moore A.T., Jay M., Bird A.C., Bhattacharya S.S. (1994). Identification of a sixth locus for autosomal dominant retinitis pigmentosa on chromosome 19. *Hum Mol Genet.*, 3: 351- 354.
- Al-Maghtheh M., Vithana E., Tartelin E., Jay M., Evans K., Moore T., Bhattacharya S., Inglehearn C.F. (1996). Evidence for a major retinitis pigmentosa locus on 19q13.4 (*RP11*) and association with a unique bimodal expressivity phenotype. *Am. J Hum Genet.*, 59: 864- 871.
- Al-Maghtheh, M. Vithana E.N., Inglehearn C.F., Moore T., Bird A.C, Bhattacharya S.S. (1998). Segregation of a *PRKCG* mutation in two *RP11* families. *Am J Hum Gene*, 62: 1248-1252.
- Ali, R.R., Sarra, G.M., Stephens, C., Alwis, M.D., Bainbridge, J.W., Munro, P.M., Fauser, S., Reichel, M.B., Kinnon, C., Hunt, D.M., Bhattacharya, S.S. and Thrasher, A.J. (2000) Restoration of photoreceptor ultra structure and function in retinal degeneration slow mice by gene therapy. *Nat. Genet.*, **25**, 306-310.
- Altschul SF., Gish W., Miller W., Myers EW. and Lipman DJ. (1990). Basic local alignment search tool. *J Mol Biol.* 215 (3): 403- 410.
- Anderson KM., Cheung PH., Kelly MD.. (1997). Rapid generation of homologous internal standards and evaluation of data for quantitation of messenger RNA by competitive polymerase chain reaction. *J Pharmacol Toxicol Methods* 38 (3): 133-140.
- Bareil C, Hamel CP, Delague V, Arnaud B, Demaille J, Claustres M. (2001). Segregation of a mutation in *CNGB1* encoding the beta-subunit of the rod cGMP-

gated channel in a family with autosomal recessive retinitis pigmentosa. *Hum Genet.*, 108 (4): 328-334.

- Bateman A., Birney E., Durbin R., Eddy S.R., Howe K.L. and Sonnhammer E.L. (2000). The Pfam protein families database. *Nucleic Acids Res*, 28: 263-266.
- Beck, J.A, Lloyd, S., Hafezparast, M., Lennon-Pierce, M., Eppig, J.T., Festing, M.F., Fisher, E.M. (2000) Genealogies of mouse inbred strains. *Nat Genet.*, 24, 23-25.
- Bennett J., Maguire AM., Cideciyan AV., Schnell M., Glover E., Anand V., Aleman TS., Chirmule N., Gupta AR., Huang YJ., *et al.* (1999). Stable transgene expression in rod photoreceptors after recombinant adeno-associated virus-mediated gene transfer to monkey retina. *Proc. Natl. Acad. Sci. USA*, 96: 9920-9925.
- Bessant D.A. Ali R.R. and Bhattacharya S.S. (2001). Molecular genetics and prospects for therapy of the inherited retinal dystrophies. *Current Opinion in Genetics and Development*, 11: 307-316.
- Bishop D.T., McDonald W.H., Gould K.L., Forsburg S.L. (2000). Isolation of an essential *Schizosaccharomyces pombe* gene, *prp31(+)*, that links splicing and meiosis. *Nucleic Acids Res.*, 28: 2214-2220.
- Bowne S. J., Sullivan L. S., Blanton S. H., Cepko C. L., Blackshaw S., Birch D. G., Hughbanks-Wheaton D., Heckenlively J. R., Daiger S. P. (2002). Mutations in the inosine monophosphate dehydrogenase 1 gene (*IMPDH1*) cause the RP10 form of autosomal dominant retinitis pigmentosa. *Hum. Molec. Genet.* 11: 559-568.
- Bradford M.M. (1976). A rapid and sensitive method for the quantification of microgram quantities of protein using the method of protein dye bonding. *Analytical Biochemistry*, 72: 248-254.

- Brambillasca F., Giuliana M., Ballabio E., Biondi A. (2001). Promoter analysis of TFPT(FB1), a molecular partner of TCF3 (E2A) in childhood acute lymphoblastic leukemia. *Biochemical and Biophysical Research Communications*, 288: 1250-1257.
- Brown JD., Beggs JD. (1992). Roles of PRP8 protein in the assembly of splicing complexes. *EMBO J.*, 11: 3721-3729.
- Bunday S., and Crews SJ. (1984). A study of retinitis pigmentosa in the city of Birmingham. J. Prevalence. *J. Med. Genet.*, 21: 417-420.
- Burge C.B., Tuschl T. and Sharp P.A. (1999). Splicing of precursors to mRNAs by the spliceosomes. In *The RNA World II*, Gestland, R.F., Chech, T.R. and Atkins, J.F, ed. (New York: Cold Spring Harbor laboratory Press) p: 525-560.
- Burge C. and Karlin S. (1997). Prediction of Complete Gene Structures in Human Genomic DNA. *J. Mol. Biol.*, 268: 78-94
- Burset M. and Guigo R. (1996). Evaluation of gene structure prediction programs. *Genomics*, 34 (3): 353-367.
- Bustin S A. (2000). Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *Jor. Mol. Endocrinology*, 25: 169-193.
- Carr RE., Heckenlively JR. (1986). Hereditary pigmentary degenerations of the retina. Tasman W, Jaeger, EA editors. *Clinical Ophthalmology*. Vol 3. Philadelphia: Lippencott, Williams, and Wilkins; p1-28.
- Carr RE. (1994). Abnormalities of cone and rod function. Ryan SJ, editor. *Retina* 2nd ed. St. Louiso: Mosby, p 502-514.
- Cartegni L, Chew SL, Krainer AR. (2002) Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet.* , 3 (4): 285-298.

- Chakarova C.F., Hims M.M., Bolz H., Abu-Safieh L., Patel R.J., Papaioannou M.G., Inglehearn C.F., Keen T.J., Willis C., Moore A.T., Rosenberg T., Webster A.R., Bird A.C., Gal A., Hunt D., Vithana E.N., and Bhattacharya S.S. (2002). Mutations in *HPRP3*, a third member of pre-mRNA splicing factor genes, implicated in autosomal dominant retinitis pigmentosa. *Hum. Mol. Genet.*, 11: 87-92.
- Chang T.J., Juan C.C., Yin P.H., Chi C.W., Tsay H.J., (1998). Up-regulation of beta-actin, cyclophilin and GAPDH in N1S1 rat hepatoma. *Oncol. Rep.*, 5: 469-471.
- Chapple J. P., Hardcastle A. J., Grayson C., Spackman L. A., Willison K. R., Cheetham M. E. (2000). Mutations in the N-terminus of the X-linked retinitis pigmentosa protein RP2 interfere with the normal targeting of the protein to the plasma membrane. *Hum. Molec. Genet.* 9: 1919-1926.
- Church DM, Stotler CJ, Rutter JL, Murrell JR, Trofatter JA, Buckler AJ. (1994). Isolation of genes from complex sources of mammalian genomic DNA using exon amplification. *Nat Genet.*, 6 (1): 98-105.
- Craig JM, and Bickmore WA. (1994). The distribution of CpG islands in mammalian chromosomes. *Nat Genetics.*, 7 (3): 376-382.
- Daiger S. P., Sullivan L. S. and Rodriguez J. A. (1995). Correlation of phenotype with genotype in inherited retinal degeneration. *Behav. Brain Sci.*, 18: 452-467.
- Deery E. C., Vithana E. N., Newbold R. J., Gallon V. A., Bhattacharya SS., Warren M. J., Hunt D. M. and Wilkie S. E. (2002). Disease mechanism for retinitis pigmentosa (*RP11*) caused by mutations in the splicing factor gene *PRPF31*. *Human Molecular Genetics*, 11 (25): 3209-3219.
- Deloukas P., Schuler G.D., Gyapay g., Beasley E.M., *et al* (1998). A physical map of 30,000 human genes. *Science*, 282: 744-746.

- Dryja T. P., Finn J. T., Peng Y. W., McGee T. L., Berson E. L., Yau K. W. (1995). Mutations in the gene encoding the alpha subunit of the rod cGMP-gated channel in autosomal recessive retinitis pigmentosa. *Proc Natl Acad Sci USA* 92 (22): 10177-10181.
- Dryja T. P., McGee T. L., Reichel E., Hahn L. B., Cowley G. S., Yandell D. W., Sandberg M. A. and Berson E. T. (1990). A point mutation of the rhodopsin gene in one form of retinitis pigmentosa. *Nature*, 343: 364-366.
- Dryja T. P., Hahn L. B., Reboul T., Arnaud B. (1996). Missense mutation in the gene encoding the alpha subunit of rod transducin in the Nougaret form of congenital stationary night blindness. *Nat Genet.*, 13 (3): 358-360.
- Dryja T. P., Adams S. M., Grimsby J. L., McGee T. L., Hong D.-H., Li T., Andreasson S., Berson E. L. (2001.). Null *RPGRIP1* alleles in patients with Leber congenital amaurosis. *Am. J. Hum. Genet.* 68: 1295-1298.
- Duda T., Venkataraman V., Jankowska A., Lange C., Koch KW. and Sharma RK. (2000). Impairment of the rod outer segment membrane guanylate cyclase dimerization in a cone-rod dystrophy results in defective calcium signaling. *Biochemistry*, 39:12522-12533.
- Dunham I., Shimizu N., Roe BA., Chisoe S., Hunt AR., Collins JE., Bruskiewich R., Beare DM., Clamp M., Smink LJ., Ainscough R., Almeida JP., Babbage A., Bagguley C., Bailey J., Barlow K., Bates KN., Beasley O., Bird CP., Blakey S., Bridgeman AM., Buck D., Burgess J., Burrill WD., O'Brien KP., *et al.*, (1999). The DNA sequence of human chromosome 22. *Nature*, 2; 402(6761): 489-495.
- Evans K., Al-Magthteh M., Fitzke F.W., Moore A.T., Jay M., Inglehearn C.F., Arden G.B., Bird A.C. (1995). Bimodal expressivity in dominant retinitis pigmentosa genetically linked to chromosome 19q. *Br J Ophthalmol.*, 79: 841-846.

- Farber D.B. (1995). From mice to men: the cyclic GMP phosphodiesterase gene in vision and disease. *Invest. Ophthalmol. Vis. Sci.*, 36: 263-275.
- Fields C, Adams MD, White O, Venter JC. (1994). How many genes in the human genome? *Nat. Genet.*, 7 (3): 345-346.
- Freund CL, Gregory-Evans CY, Furukawa T, Papaioannou M, Looser J, Ploder L, Bellingham J, Ng D, Herbrick JA, Duncan A, Scherer SW, Tsui LC, Loutradis-Anagnostou A, Jacobson SG, Cepko CL, Bhattacharya SS, McInnes RR. (1997). Cone-rod dystrophy due to mutations in a novel photoreceptor-specific homeobox gene (CRX) essential for maintenance of the photoreceptor. *Cell*, 14; 91 (4): 543-553.
- Gautier T., Berges T., Tollervey D. and Hurt E. (1997). Nucleolar KKE/D repeat proteins Nop56p and Nop58p interact with Nop1p and are required for ribosome biogenesis. *Mol Cell Biol.*, 17: 7088-7098.
- Gibson UE., and Williams PM. (1996). A novel method for real time quantitative RT-PCR. *Genome Research*, 6: 995-1001.
- Gonzalez-Santos JM, Wang A, Jones J, Ushida C, Liu J, Hu J. (2002). Central region of the human splicing factor hprp3p interacts with hprp4p. *J Biol Chem* 277(26): 23764-23772.
- Gouya L., Puy H., Lamoril J., Da Silva V., Grandchamp B., Nordmann Y., and Deybach JC. (1999). Inheritance in erythropoietic protoporphyria: a common wild-type ferrochelatase allelic variant with low expression accounts for clinical manifestation. *Blood.*, 93: 2105-2110.
- Gouya, L. Puy, H. Robreau A.M., Bourgeois M., Lamoril J., Da Silva V., Grandchamp B., and Deybach JC. (2002). The penetrance of dominant erythropoietic protoporphyria is modulated by expression of wildtype FECH. *Nat. Genet.*, 30: 27-28.

- Gratzner W. (1994). Silence speaks in spectrin. *Nature*, 372: 620-621.
- Greenberg J., Goliath R., Beighton R., and Ramesar R. (1994). A new locus for autosomal dominant retinitis pigmentosa on the short arm of chromosome 17. *Hum. Mol. Genet.*, 3: 915-918.
- Gu SM., Thompson DA., Srikumari CR., Lorenz B., Finckh U., Nicoletti A., Murthy KR., Rathmann M., Kumaramanickavel G., Denton MJ., and Gal A. (1997). Mutations in RPE65 cause autosomal recessive childhood severe onset retinal dystrophy. *Nat Genet.*, 17: 194-197.
- Gygi S.P., Rochon Y., Franza R. B. and Aebersold R. (1999). Correlation between protein and mRNA abundance in yeast. *Mol. Cell Biol.*, 19: 1720-1730.
- Hafezi F, Grimm C, Simmen BC, Wenzel A, Reme CE. (2000) Molecular ophthalmology: an update on animal models for retinal degenerations and dystrophies. *Br J Ophthalmol.*;84 (8): 922-927.
- Hamm H.E. and Bownds M.D. (1986). Protein complement of rod outer segments of frog retina. *Biochemistry*, 25: 4512-4523.
- Hargrave PA., and McDowell JH. (1992). Rhodopsin and phototransduction: a model system for G-protein-linked receptors. *FASEB J.* 6: 2323-2331.
- Hattori M., Fujiyama A., Taylor TD., Watanabe H., Yada T., Park HS., Riesselmann L., Dagand E., Haaf T., Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F., Lehrach H., Reinhardt R., Yaspo ML. *et al.*, (2000). The DNA sequence of human chromosome 21. *Nature*, 18; 405 (6784): 311-319.
- Heckenlively J.R. (1988). Retinitis Pigmentosa. (Lippincott, Philadelphia, PA).
- Heid CA., Stevens J., Livak KJ., Williams PM. (1996). Real time quantitative PCR. *Genome Res.*, 6(10): 986-994.

- Higuchi R., Fockler C., Dollinger G., and Watson R. (1993). Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Biotechnology*, 11: 1026-1030.
- Hudson TJ., Stein LD., Gerety SS., Ma J., Castle AB., Silva J., Slonim DK., Baptista R., Kruglyak L., Xu SH., et al. (1995). An STS-based map of the human genome. *Science*, 22; 270 (5244): 1945-1954.
- Inglehearn C. F., Tarttelin E. E., Keen T. J., Bhattacharya S. S., Moore A., Taylor R., Bird A. C. (1998) A new dominant retinitis pigmentosa family mapping to the *RP18* locus on chromosome 1q11-21. *J. Med. Genet.* 35: 788-792.
- Ioannou P.A. and de Jong P. J. (1996). Construction of Bacterial Artificial chromosome libraries using the modified P1 (PAC) system. In Current Protocols in Human Genetics. Dracopoli *et al*, ed. (New York: John Wiley and sons).
- Jablonka S., Schrank B., Kralewski M., Rossoll W., and Sendtner M. (2000). Reduced survival motor neuron (*SMN*) gene does in mice leads to motor neuron degeneration: An animal model for spinal muscular atrophy III. *Hum. Mol. Genet.*, 9: 341-346.
- Keen T. J., Hims M. M., McKi A. B., Moore A. T., Doran R. M., Mackey D. A., Mansfield D. C., Mueller R. F., Bhattacharya S. S., Bird A. C., Markham A. F., Inglehearn C. F. (2002) Mutations in a protein target of the Pim-1 kinase associated with the RP9 form of autosomal dominant retinitis pigmentosa. *Europ. J. Hum. Genet.* 10: 245-249.
- Kim R.Y., Fitzke F.W., Moore A.T., Jay M., Inglehearn C., Arden G.B., Bhattacharya S.S., Bird A.C. (1995). Autosomal dominant retinitis pigmentosa mapping to chromosome 7p exhibits variable expression. *Br J Ophthalmol.*, 79: 23-27.
- Korenbrot J.I and Fernald R.D. (1989). Circadian rhythm and light regulate opsin mRNA in rod photoreceptors. *Nature*, 337: 454-457.

- Kramer A. (1996). The structure and function of proteins involved in mammalian pre-mRNA splicing. *Annu. Rev. Biochem.*, 65: 367-409.
- Kreuzer KA., Bohn A., Lupberger J., Solassol J., le Coutre P., Schmidt CA. (2001). Simultaneous absolute quantification of target and control templates by real-time fluorescence reverse transcription-PCR using 4-(4'-dimethylaminophenylazo) benzoic acid as a dark quencher dye. *Clin Chem.*, 47 (3): 486- 490.
- Kuhn A.N., and Brow Li. Z. (1999). Splicing factor Prp8 governs U4/U6 RNA unwinding during activation of the spliceosome. *Mol. Cell*, 3: 65-75.
- Lander ES., Linton LM., Birren B., Nusbaum C., Zody MC., Baldwin J., Devon K., Dewar K., Doyle M., FitzHugh W., Funke R., Gage D., Harris K., Heaford A., Howland J., Kann L., Lehoczky J., Rogers J., Sulston J., Ainscough R., Beck S., Bentley D., Burton J., Clee C., Carter N., Coulson A., Cook LL., Fulton RS., Johnson DL., Minx PJ., Clifton SW., Hawkins T., Branscomb E., Predki P., Richardson P., *et al.*, (2001). Initial sequencing and analysis of the human genome. *Nature*, 15; 409(6822): 860-921.
- Lefebvre S., Burlet P., Liu Q., Bertrand S. Clermont O., Munnich A., Dreyfuss G., and Melki J. (1997). Correlation between severity and SMN protein levels in spinal muscular atrophy. *Nat. Genet.*, 16: 265-269.
- Lemay S., Changchun M., Singh A.K., (1996). Cytokine gene expression in the MRL/lpr model of lupus nephritis. *Kidney Int.*, 50: 85-93.
- Lichanska A.M. and Simpson D.A..(2002). Application of in silico positional cloning and bioinformatic mutation analysis to the study of eye disease. *Briefings in Bioinformatics*, 3 (1): 59-72.
- Loeffen J.L.C.M., Triepels R.H., van den Heuvel L.P., Schuelke M., Buskens C.A.F., Smeets R.J.P., Trijbels J.M.F., and Smeitink A.M. (1998). cDNA of eight nuclear encoded subunits of NADH: ubiquinone oxidoreductase: human complex

I cDNA characterization completed. *Biochemical and Biophysical research communications*, 253: 415-422.

- Lukashin A. V. and Borodovsky M. (1998). GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Research*, 26 (4): 1107-1115.
- Lyness A. L., Ernst W., Quinlan M. P., Clover G. M., Arden G B., Carter R. M., Bird A. C. and Parker J. A. (1985). A clinical, psychophysical and electroretinographic survey of patients with autosomal dominant retinitis pigmentosa. *Br. J. Ophthalmol.* 69 (5): 326-339.
- Maddock JR., Roy J., and Woolford JL. (1996). Six novel genes necessary for pre-RNA splicing in *Saccharomyces cerevisiae*. *Nucleic Acidss Research*, 24: 1037-1044.
- Makarova O.V., Makarov E.M., Liu, S., Vornlocher H-P. and Luhrmann R. (2002). Protein 61 K, encoded by a gene (*PRPF31*) linked to autosomal dominant retinitis pigmentosa, is required for U4/U6.U5 tri-snRNP formation and pre-mRNA splicing. *EMBO J.*, 21, 1148–1157.
- Marlhens F., Bareil C., Griffoin JM., Zrenner E., Amalric P., Eliaou C., Liu SY., Harris E., Redmond TM., Arnaud B., Claustres M., Hamel CP. (1997). Mutations in RPE65 cause Leber's congenital amaurosis. *Nat Genet*, 17 (2): 139-141.
- Massof R. W. and Finkelstein D. (1981). Two forms of autosomal dominant primary retinitis pigmentosa. *Doc. Ophthalmol.*, 51 (4): 289-346.
- McGee T.L., Devoto M., Ott J., Berson E.L., Dryja T.P. (1997). Evidence that the penetrance of mutations at the *RP11* locus causing dominant retinitis pigmentosa is influenced by a gene linked to the homologous *RP11* allele. *Am J Hum Genet.* 61: 1059-1066.
- McKie A.B., McHale J.C., Keen T.J., Tarttelin E.E., Goliath R., van Lith-Verhoeven J.J.C., Greenberg J., Ramesar R.S., Hoyng C.B., Cremers F.P.M.,

- Mackey D.A., Bhattacharya S.S., Bird A.C., Markham A.F., and Inglehearn C.F. (2001). Mutations in the pre-mRNA splicing factor gene *PRPC8* in autosomal dominant retinitis pigmentosa (RP13). *Hum. Mol. Genet.*, 10: 1555-1562.
- Molday R.S. (1998). Photoreceptor membrane proteins, phototransduction, and retinal degenerative diseases. *Inv. Ophthalm. And Visual Science*, 39 (13): 2493-2513.
 - Monaco AP, Neve RL, Colletti-Feener C, Bertelson CJ, Kunit DM, Kunkel LM. (1986) Isolation of candidate cDNAs for portions of the Duchenne muscular dystrophy gene. *Nature*. 16-22; 323 (6089): 646-650.
 - Moore A.T., Fitzke F.W., Kemp C.M., Arden G.B., Keen T.J., Inglehearn C.F., Bhattacharya S.S., and Bird, A.C. (1992). Abnormal dark adaptation kinetics in autosomal dominant sector retinitis pigmentosa due to rod opsin mutation. *British Journal of Ophthalmology*, 76: 465-469.
 - Moore A.T., Fitzke F., Jay M., Arden G.B., Inglehearn C.F., Keen T.J., Bhattacharya S.S., Bird, A.C. (1993). Autosomal dominant retinitis pigmentosa with apparent incomplete penetrance: a clinical, electrophysiological, psychophysical, and molecular genetic study. *Br J Ophthalmol.*, 77: 473-479.
 - Morimura H., Fishman GA., Grover SA., Fulton AB., Berson EL., Dryja T. P. (1998). Mutations in the RPE65 gene in patients with autosomal recessive retinitis pigmentosa or leber congenital amaurosis. *Proc Natl Acad Sci USA*. Mar 17; 95 (6): 3088-3093.
 - Nacksung K., Masamichi T., Jaerang R., Regis J., and Yongwon C. (2002). A novel membrane of the leukocyte receptor complex regulates osteoclast differentiation. *J.Exp.Med.*, 195 (2): 201-209.
 - Nathans J., Thomas D., Hogness DS. (1986). Molecular genetics of human color vision: the genes encoding blue, green, and red pigments. *Science*, 11; 232 (4747): 193-202.

- Nathans J. (1992). Rhodopsin: structure, function and genetics. *Biochemistry*, 31: 4923-4931.
- Nottrott S., Urlaub H., and Luhrmann R. (2002). Hierarchical, clustered protein interactions with U4/U6 snRNP: a biochemical role for U4/U6 proteins. *EMBO*, 21 (20): 5527-5538.
- Nusbaum C., Slonim DK., Harris KL., Birren BW., Steen RG., Stein LD., Miller J., Dietrich WF., Nahf R., Wang V., Merport O., Castle AB., Husain Z., Farino G., Gray D., Anderson MO., Devine R., Horton LT. Jr, Ye W., Wu X., Kouyoumjian V., Zemsteva IS., Wu Y., Collymore AJ., Courtney DF., et al. (1999). A YAC-based physical map of the mouse genome. *Nat Genet.*, 22(4): 388-393.
- Olsson, J.E., Gordon, J.W., Pawlyk, B.S., Roof, D., Hayes, A., Molday, R.S., Mukai, S., Cowley, G.S., Berson, E.L. and Dryja, T.P. (1992) Transgenic mice with a rhodopsin mutation (Pro23His): a mouse model of autosomal dominant retinitis pigmentosa. *Neuron* 9,815-817.
- Oshika T, Klyce SD, Applegate RA, Howland HC.(1999) Changes in corneal wavefront aberrations with aging. *Invest Ophthalmol Vis Sci.*, 40 (7): 1351-1355.
- Parimoo S., Kolluri and R. Weissman SM. (1993). cDNA selection from total yeast DNA containing YACs. *Nucleic Acids Res.*, 11;21(18): 4422-4423.
- Payne A.M., Downes S.M., Bessant D.A.R., Taylor R., Holder, G.E., Warren M.J., Bird A.C. and Bhattacharya S.S. (1998). A mutation in guanylate cyclase activator 1A (GUCA1A) in autosomal dominant cone dystrophy pedigree mapping to a new locus on chromosome 6p21.1. *Hum. Mol. Genet.*, 7: 273-277.
- Pearson WR. and Lipman DJ. (1988). Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA* 85 (8): 2444-2448.

- Perrault I., Rozet JM., Gerber S., Ghazi I., Leowski C., Ducroq D., Souied E., Dufier JL., Munnich A., Kaplan J. (1999). Leber congenital amaurosis. *Mol Genet Metab.*, 68 (2): 200-208.
- Perrault I., Rozet JM., Gerber S., Ghazi I., Ducroq D., Souied E., Leowski C., Bonnemasion M., Munnich A., and Kaplan J. (2000). Spectrum of retGC mutations in Lebers congenital amaurosis. *Eur. J. Hum. Genet.*, 8: 578-582.
- Pettersson M., Dannaëus K., Nilsson K., Jonsson JL, (2000). Isolation of MYADM, a novel hematopoietic-associated marker gene expressed in multipotent progenitor cells and up-regulated during myeloid differentiation. *J Leukoc Biol.*, 67(3): 423-431.
- Randon J, Boulanger L, Marechal J, Garbarz M, Vallier A, Ribeiro L, Tamagnini G, Dhermy D, Delaunay J. (1994) A variant of spectrin low-expression allele alpha LELY carrying a hereditary elliptocytosis mutation in codon 28. *Br J Haematol.* 88 (3): 534-540.
- Remington L.A. (1998). Clinical anatomy of the visual system. Boston, Butterworth-Heinemann.
- Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, Grzelczak Z, Zielenski J, Lok S, Plavsic N, Chou JL, et al. (1989). Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science*, 8; 245 (4922): 1066-1073.
- Rivolta C., Sharon D., DeAngelis M.M. and Dryja T.P. (2002). Retinitis pigmentosa and allied disease: numerous diseases, genes, and inheritance patterns. *Human Molecular Genetics*, 11 (10): 1219-1227.
- Saari JC., Bredberg DL., Noy N. (1994). Control of substrate flow at a branch in the visual cycle. *Biochemistry*, 33 (10): 3106-3112.

- Schuler G.D., Boguski M.S., Stewart E.A., Stein L.D., Gyapay G., Rice K., White R.E., *et al* (1996). A gene map of the human genome. *Science* 274: 540-546.
- Shen D., Jiang M., Hao W., Tao L., Salazar M., Fong HK. (1994). A human opsin-related gene that encodes a retinaldehyde-binding protein. *Biochemistry*, 8; 33 (44): 13117-13125.
- Stryer L. (1988). Molecular basis of visual excitation. *Cold Spring Harb. Symp. Quant. Biol.*, 53: 283-294.
- Suda T., Takahashi N., Udagawa N., Jimi E., Gillespie, M.T., and Martin T.J. (1999). Modulation of osteoclast differentiation and function by the new members of the tumor necrosis factor receptor and ligand families. *Endocr. Rev.*, 20: 345-357.
- Sullivan L. S. and Daiger S. P. (1996). Inherited retinal degeneration: exceptional genetic and clinical heterogeneity. *Molecular Medicine Today*, 380-386.
- Takai D., Jones PA. (2002). Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc Natl Acad Sci USA*, 99 (6): 3740-3745.
- Terns M.P. and Terns R.M. (2001). Macromolecular complexes: SMN- the master assembler. *Curr. Biol.*, 11: R862-R864.
- Thanaraj T.A. (2000). Positional characterisation of false positives from computational prediction of human splice sites. *Nucleic Acids Research*, 28 (3), 744-754.
- Teigelkamp S., Achsel T., Mundt C., Gotherl S.F., Cronshagen U., Lane W.S., Marahiel M. and Luhrmann R. (1998). The 20 kD protein of human [U4/U6.U5] tri snRNPs is a novel cyclophilin that forms a complex with the U4/U6-specific 60 kD and 90 kD proteins. *RNA*, 4: 127-141.
- Tsonis P. A. (2003). Anatomy of gene regulation, a three dimensional structural analysis. Cambridge University Press, chapter seven, 153-171.

- Van der Spuy J, Chapple JP, Clark BJ, Luthert PJ, Sethi CS, Cheetham ME. (2002). The Leber congenital amaurosis gene product AIPL1 is localized exclusively in rod photoreceptors of the adult human retina. *Hum Mol Genet.*, 1;11 (7): 823-831.
- Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, et al., (2001). The sequence of the human genome. *Science* 16;291 (5507): 1304-1351.
- Vithana E.N. (1998). Towards positional cloning of the gene responsible for autosomal dominant retinitis pigmentosa mapped to chromosome 19q13.4. Thesis submitted to University College London. p: 136-141.
- Vithana E.N., Al-Maghtheh M., Bhattacharya S.S., Inglehearn C.F. (1998). *RP11* is the second most common locus for dominant retinitis pigmentosa. *J Med Genet.*, 35: 174-175.
- Vithana E.N., Al-Maghtheh M., Kenmochi N., Higa S. and Bhattacharya S.S. (1999). Physical map of the *RP11* locus for autosomal dominant retinitis pigmentosa and invetsigation of a candidate gene. *Am. J Hum Genet.*, 65: 2390.
- Von Schantz M., Lucas RJ., and Foster, R.G. (1999). Circadian oscillation of photopigment transcript levels in the mouse retina. *Brain Res Mol Brain Res.*, 72, 108-114.
- Weleber R.G. and Gregory-Evans K. (2001). Retinitis pigmentosa and allied disorders. In Ryan, S.J. (ed.), *Retina*. Mosby, St. Louis. 362-470
- Weidenhammer E.M., Singh M., Ruiz-Noriega M., Woolford J.L. (1996). The PRP31 gene encodes a novel protein required for pre-mRNA splicing in *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, 24: 1164-1170.

- Weidenhammer E.M., Ruiz-Noriega M., Woolford J.L. (1997). Prp31p promotes the association of the U4/U6 x U5 tri-snRNP with prespliceosomes to form spliceosomes in *Saccharomyces cerevisiae*. *Mol Cell Biol.*, 17: 3580-3588.
- Whitehouse DB, Putt W, Lovegrove JU, Morrison K, Hollyoake M, Fox MF, Hopkinson DA, Edwards YH. (1992). Phosphoglucomutase 1: complete human and rabbit mRNA sequences and direct mapping of this highly polymorphic marker on human chromosome 1. *Proc Natl Acad Sci U S A.*, 1; 89 (1): 411-415.
- Will C.L. and Luhrmann R. (2001). Spliceosomal UsnRNP biogenesis, structure and function. *Current Opinion in Cell Biology*, 13: 290-301.
- Wilmotte R, Marechal J, Morle L, Baklouti F, Philippe N, Kastally R, Kotula L, Delaunay J, Alloisio N. (1993). Low expression allele alpha LELY of red cell spectrin is associated with mutations in exon 40 (alpha V/41 polymorphism) and intron 45 and with partial skipping of exon 46. *J Clin Invest.*, 91(5): 2091-2096.
- Wirth B. (2000). An update of the mutation spectrum of the survival motor neuron gene (*SMN1*) in autosomal recessive spinal muscular atrophy (SMA). *Hum. Mutat.*, 15: 228-237.
- Wright A.F. and van Heyningen V. (2001). Short cut to disease genes. *Nature*, 414: 705- 706.
- Wu GS., Swiderek KM., Rao NA. (1996). A novel retinal pigment epithelial protein suppresses neutrophil superoxide generation. II. Purification and microsequencing analysis. *Exp Eye Res.*, 63 (6): 727- 737.
- Wu P., Brockenbrough J.S., Metcalfe AC., Chen S. and Aris JP. (1998). Nop5p is a small nucleolar ribonucleoprotein component required for pre-18S rRNA processing in yeast. *J Biol Chem.*, 273, 16453- 16463.

- Xu S., Nakazawa M., Tamai M. and Gal A. (1995). Autosomal dominant retinitis pigmentosa locus on chromosome 19q in a Japanese family. *J. Med. Genet.* 32, 915- 916.
- Xu S. J., Schhwartz M., Rosenberg T., Gal A. (1996). A ninth locus (RP18) for autosomal dominant retinitis pigmentosa maps in the pericentromeric region of chromosome 1. *Hum. Mol. Genet.* 5, 1193-1197.
- Yang Z, Peachey NS, Moshfeghi DM, Thirumalaichary S, Chorch L, Shugart YY, Fan K, Zhang K. (2002). Mutations in the RPGR gene cause X-linked cone dystrophy. *Hum Mol Genet.* 1; 11 (5): 605- 611.
- Young RW and Bok D (1969). Participation of the retinal pigment epithelium in the rod outer segment renewal process. *J Cell Biol*, 42: 392– 403.

Appendix

C_t values obtained for cell lines of ADRP5 family from two RT-PCR assays that were subsequently used to calculate the mRNA copy number using the standard curve formula $y = -3.514x + 46.47$

Cell line	Clinical status	Ct value
307	Symptomatic	22.53, 22.17, 21.88, 21.46
305	Symptomatic	21.31, 21.16, 21.44, 21.33, 21.26, 21.09, 21.57
306	Symptomatic	21.01, 21.19, 21.19, 21.42, 21.68, 20.94
319	Symptomatic	21.40, 21.14, 21.13, 21.41, 21.41, 21.55, 21.73
320	Symptomatic	20.86, 20.71, 21.08, 21.92, 21.31, 21.43, 21.87, 20.58
323	Symptomatic	21.84, 21.60, 21.30, 21.38
325	Symptomatic	21.20, 21.24, 21.42, 21.30, 21.84, 21.57
271	Symptomatic	21.55, 21.20, 21.57, 21.11, 21.16, 21.37
340	Asymptomatic	21.05, 20.96, 20.29, 20.36
299	Asymptomatic	21.01, 20.62, 20.50, 20.47, 20.71
298	Asymptomatic	20.52, 20.42, 20.52, 20.64
353	Asymptomatic	21.10, 20.66, 21.00, 20.75
297	Asymptomatic	21.12, 20.56, 20.94, 21.11, 20.53
255	Asymptomatic	21.26, 20.87, 21.39
311	Asymptomatic	21.02, 21.02, 21.33, 20.94, 21.31
300	Normal	19.10, 19.68, 19.35, 19.35, 19.13
343	Normal	19.69, 19.33
295	M. Normal	19.19, 19.36, 19.63, 19.29
296	M. Normal	19.43, 19.27, 19.77, 19.89
341	M. Normal	19.75, 19.72, 19.90,

Mann-Whitney-U Test (independent group comparison test): The Mann-Whitney-U test is a non-parametric statistical procedure that can be used to compare two independent groups of sampled data.

This, like many non-parametric tests, uses the ranks of the data rather than their raw values to calculate the statistic. The test statistic for the Mann-Whitney test is U. This value is compared to a table of critical values for U based on the sample size of each group. If U exceeds the critical value for U at some significance level (usually 0.05) it means that there is evidence to reject the null hypothesis in favor of the alternative hypothesis. P is the probability that the variates are not related. The low P value of <0.05 indicates that the two variates probably are related (see chapter 5).

Arithmetic mean

The arithmetic mean, commonly called the mean or average, is calculated by summing all values in a particular samples group and dividing this sum by the number of items in that group. In chapter 5 in this thesis the arithmetic mean was used to calculate the mean of C_t values obtained from RT-qPCR experiments (see table 5.1, 5.2, 5.3).

The median

The median is defined as the value of variable (in an ordered array) that has an equal number of items on either side of it. Thus, the median divides a frequency distribution into two halves. When the number in the sample is even, the median is conventionally calculated as the midpoint between two variates (see figure 5.7).

The standard deviation

A desirable measure of dispersion that would take all items of a distribution into consideration, weighting each item by its distance from the center of distribution. Three steps are necessary for computing the standard deviation: (1) finding the $\sum y^2$, the sum squares, (2) dividing by n-1 to give the variance and (3) taking the square root of the variance to obtain the standard deviation (see table 5.2, 5.3).