

Lincoln University Digital Thesis

Copyright Statement

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

This thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- you will use the copy only for the purposes of research or private study
- you will recognise the author's right to be identified as the author of the thesis and due acknowledgement will be made to the author where appropriate
- you will obtain the author's permission before publishing any material from the thesis.

The Genetic Basis of Ovine Heritable Cataract

A thesis submitted in partial fulfilment

of the requirements of the Degree

of Doctor of Philosophy

at

Lincoln University

by

Gareth Richard Samuel Wilson

Lincoln University

2010

Abstract of a thesis submitted in partial fulfilment of the requirements for the Degree of
Doctor of Philosophy

The Genetic Basis of Ovine Heritable Cataract

by

Gareth Richard Samuel Wilson

Sheep have been bred at Lincoln University which develop cortical opacities at 1-2 months due to a genetic defect. This condition is referred to as Ovine Heritable Cataract (OHC), and has been used as an animal model for human cataracts.

Breeding records were studied to determine the manner of inheritance of the cataract trait. Offspring affected by cataracts were produced from matings between normal, unrelated animals and animals affected by cataract. Since cataracts are rare in sheep, the normal unrelated animals were probably not carriers for the OHC trait and the results suggested a dominant mode of inheritance for OHC. The cataract trait could be inherited from both sires and dams, indicating that the inheritance was autosomal dominant. Matings between affected animals, and also between affected animals and normal animals, produced significantly fewer cataract offspring than expected for an autosomal dominant trait with full penetrance. For offspring of two affected parents, the proportion of affected offspring was not significantly different from that expected from a lethal homozygote manner of inheritance, and no known homozygotes have been produced. However the changes in litter size or premature death rates expected from a lethal homozygote effect were not observed. Also, a lethal homozygote effect would not explain why fewer affected animals than expected were produced by mating affected animals with normal animals. Therefore it was concluded that OHC is inherited in an autosomal dominant manner with incomplete penetrance. From a weighted average of the results of both types of matings, a penetrance of 0.81 ± 0.04 was calculated.

In order to localise the gene responsible for OHC, a genome scan for a locus associated with OHC was carried out. DNA was extracted from a ram affected by OHC and 41 of his offspring. Ten human and mouse chromosomes with genes or regions linked to cataract were selected. The ovine homologues of these chromosomes were analysed, and no significant linkage was found. When the next ten largest chromosomes were tested, markers on ovine chromosome 6 (OAR6) showed significant linkage to the OHC locus with the most probable

position for the mutation between 112.3 and 135.5 cM from the centromere of OAR6. A higher-resolution genome scan, using more markers in the linked region and 296 animals, narrowed the most likely region to between 112.3 and 132.9 cM.

A breakpoint panel of 36 sheep with recombinations in the linked region was genotyped for eight new markers, selected for informativeness and ease of scoring from 63 bovine markers. Because of errors in the genome assembly used to select the markers, only three were within the linked region. Linkage analysis on the combination of old and new genotypes showed that the most likely position for the OHC locus was between 114.8 and 132.9 cM. A multipoint analysis performed on the same data gave a 1-LOD support interval of 123.3 to 128.8 cM.

The breakpoint panel was then genotyped for three new markers, selected from 83 bovine markers. Linkage analysis incorporating the new genotypes showed that the most likely position for the OHC locus was between 131.1 and 132.9 cM. This interval was used to select candidate genes for OHC, but errors in the marker order and pedigree were discovered that made this interval unreliable.

The candidate gene *NUDT9*, which is found in this region of chromosome 6, was sequenced to find any possible cataract mutations. A polymorphism was found with a strong association to OHC, and a large number of animals were genotyped. Some animals had a genotype inconsistent with the polymorphism being the OHC mutation, and the amino acid residue changed was not conserved in other mammals. Therefore the polymorphism was ruled out as the OHC mutation but was a useful marker.

The linkage analysis was repeated, including the *NUDT9* polymorphism as an additional marker and with corrections to the pedigree and marker order. Linkage analysis showed the most likely position for the OHC locus was between 131.1 and 131.8 cM. multipoint analysis gave a 1-LOD support interval of 131.3 to 131.7 cM from the centromere.

The gene *AFF1*, which causes autosomal dominant cataracts in mice, is also found in the linked region. Primers were designed to amplify the entire *AFF1* coding sequence, but only half of the sequence was successfully amplified. This region was sequenced in both affected

and normal animals, and no polymorphisms with an association with OHC were discovered. However, AFF1 remains the best candidate for the OHC gene.

Keywords: AFF1, cataract, cDNA, genome scan, lens, linkage, microsatellite, mRNA, NUDT9, OHC, ovine, ovine heritable cataract, RNA, sheep, SSR.

Table of Contents

Abstract	iii
Table of Contents	v
Acknowledgements	ix
List of Tables.....	x
List of Figures	xii
Abbreviations	xiv
Chapter 1 Literature Review	1
1.1 Introduction	1
1.2 Physiology of the Lens.....	1
1.3 Cataract	6
1.3.1 Introduction	6
1.3.2 Age-Related Nuclear Cataract.....	7
1.3.2 Cortical Cataract.....	9
1.3.3 Diabetes and Cataract.....	10
1.3.4 Cataracts in Non-Human Animals	10
1.4 Hereditary Cataracts.....	11
1.4.1 Human Hereditary Cataracts	11
1.4.1.1 Crystallins	11
1.4.1.2 Connexins.....	12
1.4.1.3 Transcription Factors	14
1.4.1.4 Other Genes.....	15
1.4.1.5 Summary of Human Cataract Mutations.....	18
1.4.1.6 Linked Regions	22
1.4.2 Rodents.....	22
1.4.3 Canine	23
1.4.4 Bovine	24
1.5 Ovine Heritable Cataract.....	25
1.5.1 Introduction	25
1.5.2 Inheritance.....	25
1.5.3 Pathology.....	26
1.5.4 Histology	27
1.5.5 Biochemistry	28
1.5.6 OHC as an Experimental Model	29
1.6 Linkage Mapping and Microsatellites.....	30
1.7 Genome Assemblies.....	33
1.8 Tools in Ovine Genetics.....	35
1.9 Identifying Disease-Causing Genes and Mutations	37
1.9.1 Introduction	37
1.9.2 Traditional Candidate Gene Strategy	37
1.9.3 Comparative Genomics Strategy.....	38
1.9.4 Positional Candidate Gene Strategy.....	38
1.9.5 Single Nucleotide Polymorphism Chips	39
1.9.6 Next Generation Sequencing.....	39
1.10 Hypothesis and Aims	40
Chapter 2 Mode of Inheritance of Ovine Heritable Cataract	41
2.1 Introduction	41
2.1.1 Massey University Breeding	41
2.1.2 Lincoln University Breeding.....	41

2.1.3 Incomplete Penetrance	42
2.1.4 Lethal Embryonic Homozygote	43
2.2 Massey University Breeding	44
2.3 Lincoln University Breeding	46
2.3.1 Pedigree	46
2.3.2 Breeding Records	47
2.3.3 Statistical Analysis	52
2.3.4 Litter Size	53
2.3.5 Death Rates	54
2.4 Discussion	55
2.5 Conclusion	59
Chapter 3 Genome Scanning	60
3.1 Introduction	60
3.2 Whole Population Scan	60
3.2.1 Introduction	60
3.2.2 Methods	61
3.2.2.1 Chromosome Selection	61
3.2.2.2 Pedigree Selection	63
3.2.2.3 Genotyping	66
3.2.2.4 Linkage Analysis	67
3.2.2.5 Multipoint Analysis	68
3.2.3 Results	69
3.2.3.1 Chromosome Selection	69
3.2.3.2 Linkage Analysis	69
3.2.3.3 Multipoint Analysis	72
3.2.4 Discussion	72
3.3 Breakpoint Panel Scan	78
3.3.1 Introduction	78
3.3.2 Methods	79
3.3.2.1 Breakpoint Panel	79
3.3.2.2 Fine Mapping, Part One	82
3.3.2.3 Fine Mapping, Part Two	84
3.3.3 Results	91
3.3.3.1 Fine Mapping, Part One	91
3.3.3.2 Pedigree and Marker Position Revision	93
3.3.3.3 Fine Mapping, Part Two	94
3.3.3.4 Gene Synteny	102
3.3.3.5 Breakpoint Panel Recombinations	104
3.3.4 Discussion	105
3.3.4.1 Breakpoint Panel	105
3.3.4.2 New Markers, Part One	106
3.3.4.3 New Markers, Part Two	107
3.3.4.4 Linkage Analysis	107
3.3.4.5 Gene Synteny	110
3.5 Conclusion	111
Chapter 4 Candidate Genes	112
4.1 Introduction	112
4.1.2 Criteria for Selection of Candidate Genes	112
4.2 Database Searches	113
4.2.1 Introduction	113

4.2.2 Methods.....	113
4.2.3 Results.....	114
4.2.4 Discussion.....	117
4.3 NUDT9.....	121
4.3.1 Introduction.....	121
4.3.3 Methods.....	121
4.3.3.1 Extraction of RNA from Normal Lenses.....	121
4.3.3.2 Expression of NUDT9 in Normal Lenses.....	122
4.3.3.3 Sequencing.....	124
4.3.3.4 Genotyping.....	125
4.3.3.5 Comparison of NUDT9 Protein Sequences in Different Species.....	127
4.3.3.6 Comparison of NUDT9 Exon Boundaries.....	127
4.3.4 Results.....	128
4.3.4.1 Expression of NUDT9 in Normal Lenses.....	128
4.3.4.2 Sequencing.....	130
4.3.4.3 Polymorphisms.....	132
4.3.4.4 Genotyping.....	134
4.3.4.5 Comparison of NUDT9 Protein Sequences in Different Species.....	139
4.3.4.6 Comparison of NUDT9 Exon Boundaries.....	141
4.3.5 Discussion.....	144
4.3.5.1 Gene and Protein Details.....	144
4.3.5.2 Analysis of NUDT9 as a Candidate for the OHC Gene.....	146
4.3.6 Conclusion.....	149
4.4 AFF1.....	150
4.4.1 Introduction.....	150
4.4.2 Methods.....	150
4.4.2.1 Expression of AFF1 in Lenses.....	150
4.4.2.2 Sequencing of AFF1.....	151
4.4.2.3 Mutation Analysis of AFF1.....	152
4.4.2.4 Comparison of AFF1 Exon Boundaries.....	153
4.4.3 Results.....	153
4.4.3.1 Expression of AFF1 in Lenses.....	153
4.4.3.2 Sequencing of AFF1.....	153
4.4.3.3 Mutation Analysis of AFF1.....	161
4.4.3.4 Comparison of AFF1 Exon Boundaries.....	161
4.4.4 Discussion.....	166
4.4.4.1 Gene and Protein Details.....	166
4.4.4.2 Analysis of AFF1 as a Candidate for the OHC Gene.....	168
4.4.3.4 Comparison of AFF1 Exon Boundaries.....	169
4.4.5 Conclusion.....	170
Chapter 5 General Discussion and Future Directions.....	171
5.1 General Discussion.....	171
5.1.1 Genome Scanning and Candidate Gene Analysis.....	171
5.1.2 Breeding Studies.....	174
5.1.3 Significance to Cataract Research.....	175
5.2 Future Directions.....	176
5.2.1 Genome Scanning.....	176
5.2.2 AFF1 Testing.....	177
5.2.3 Large Scale Sequencing.....	178
5.2.4 Biochemistry of OHC.....	178

5.2.5 NUDT9 Marker	179
5.2.6 Breeding an OHC Homozygote	179
Appendix A: Genome Scan Markers	181
Appendix B: Crimap Input File.....	185
References	198
Publications and Presentations Arising from This Thesis.....	222

Acknowledgements

I would like to thank my supervisors, Dr Jim Morton and Professor David Palmer.

I would also like to thank Nigel Jay, Martin Ridgeway, and the Johnston Memorial Laboratory workers for management of the sheep flock, Steve Heap for ophthalmologic testing of the sheep, and Malvern Abattoir for donation of sheeps' eyes. Karl Gately, Dr Lucinda Robertson, Dr Hannah Lee, Matthew Muir, and Joshua McDermott provided technical assistance. Large-scale genotyping was carried out by Sharon Wilson at AgResearch Invermay. Dr John McEwan and Dr Rayna Anderson from AgResearch Invermay provided advice on genetics and genotyping. Dr Ken Dodds at the same corporation carried out multipoint linkage analysis. Dr Richard Sedcole from Lincoln University provided advice on statistics.

List of Tables

Table 1: Mutations linked to human cataracts	18
Table 2: Cataract status of offspring from Massey University matings	45
Table 3: Cataract statuses of affected ram - normal ewe offspring	47
Table 4: Cataract statuses of affected ram – affected ewe offspring	50
Table 5: Litter sizes from different breeding types	53
Table 6: Pedigree of 41 lambs used for linkage analysis	63
Table 7: Pedigree of 293 lambs used for linkage analysis	64
Table 8: Human and mouse genes and chromosome regions	69
Table 9: Highest LOD scores for each chromosome	70
Table 10: Chromosome 6 LOD scores	70
Table 11: Offspring of each sire omitted	71
Table 12: LOD scores from 289 sheep	71
Table 13: Microsatellite primer sequences	82
Table 14: Target positions for new markers	85
Table 15: Microsatellite primer sequences, second batch	86
Table 16: Loglikelihoods of different OHC locus positions	92
Table 17: New marker positions	98
Table 18: Common alleles in founding sires	99
Table 19: Loglikelihoods of different OHC positions, revised positions	100
Table 20: Genes in the OHC-linked region	116
Table 21: NUDT9 primers	123
Table 22: NUDT9 sequencing primers	124
Table 23: NUDT9 polymorphism testing primers	126
Table 24: NUDT9 DNA extract primers	126

Table 25: 989G>A polymorphism genotypes	136
Table 26: 989G>A polymorphism genotypes of the OHC flock	138
Table 27: Cross-species comparison of NUDT9 polymorphisms	139
Table 28: AFF1 primers	152
Table 29: DNAMAN alignments of AFF1 sequences	158
Table 30: Markers used in partial genome scan	181
Table 31: PCR results for first batch of new markers	184
Table 32: CRIMAP input file for 11 markers on sheep chromosome 6	197

List of Figures

Figure 1: Cross-section of the human eye	1
Figure 2: Cross-section of the lens	2
Figure 3: Progression of Ovine Heritable Cataract	27
Figure 4: Early breeding at Massey University	44
Figure 5: Pedigree diagram of the cataract flock	46
Figure 6: Human-cattle Oxford grid	62
Figure 7: First stage multipoint linkage analysis	72
Figure 8: Pedigree of animals in the breakpoint panel	80
Figure 9: Breakpoint panel recombinations	81
Figure 10: Products from BMS1509593, BMS1509702, and BMS1509791	91
Figure 11: Second multipoint linkage analysis	93
Figure 12: First amplification of t6s8	95
Figure 13: Temperature gradient PCR, t6s8, bovine DNA	95
Figure 14: Temperature gradient PCR, t6s8, sheep DNA	96
Figure 15: Magnesium gradient PCR for primer set t6s8	96
Figure 16: Acrylamide gel of t6s8 products	97
Figure 17: Third multipoint linkage analysis	101
Figure 18: Sheep, cattle, and human gene positions	102
Figure 19: Cattle-human gene rearrangement	103
Figure 20: Breakpoint panel recombinations, new markers	104
Figure 21: Close-up of Figure 17	109
Figure 22: Genome scanning overview	109
Figure 23: BLAST matches to the t3s9 sequence	114
Figure 24: Recognition site of HpyCh4IV	127

Figure 25: Agarose gel of RNA extracts	128
Figure 26: NUDT9F-NUDT9R PCR product	129
Figure 27: StartF-GapR, GapF-GapR, and EndF-EndR products	130
Figure 28: StartF and EndR PCR products	131
Figure 29: NUDT9 sequences	131
Figure 30: Genedoc alignment of NUDT9 sequences	132
Figure 31: NUDT9 consensus sequence, from start to stop codon	133
Figure 32: BLDF-BLDR PCR products	134
Figure 33: Genedoc alignment of NUDT9 sequences, 20 animals	135
Figure 34: DNA2F-DNA2R PCR products	137
Figure 35: Restriction digest products	138
Figure 36: Cattle, sheep and human NUDT9 protein sequences	140
Figure 37: NUDT9 exon boundaries	143
Figure 38: AFF1C-AFF1F PCR products	154
Figure 39: AFF1FF and AFF1R products	156
Figure 40: AFF12GAP PCR products	157
Figure 41: AFF1DF sequence aligned with bovine AFF1 CDS	159
Figure 42: AFF1 sheep sequences	159
Figure 43: Sequence and primers for AFF1	161
Figure 44: AFF1 exon boundaries	164
Figure 45: Human-mouse AFF1 alignment	165
Figure 46: Sheep genome BLAST search	166

Abbreviations

AAH	Artificial aqueous humour
AFF1	Affranchised 1 gene
APS	Amphotericin, penicillin, and streptomycin
AgResearch	New Zealand Pastoral Agriculture Research Institute Limited
ASD	Anterior segment dysgenesis
BAC	Bacterial artificial chromosome
BFSP2	Beaded filament structural protein 2
BLAST	Basic local alignment search tool
bp	base pairs
Btau2.0	<i>Bos taurus</i> genome assembly 2.0
Btau3.1	<i>Bos taurus</i> genome assembly 3.1
Btau4.0	<i>Bos taurus</i> genome assembly 4.0
BUPA	(formerly) British United Provident Association
C	Cataract
°C	Degrees Celsius
cDNA	Complementary DNA
CDS	Coding sequence
CDS1	Cytosine-diphosphate-diacylglycerol synthase (phosphatidate cytidyltransferase) 1
CHX10	Ceh-10 homeodomain-containing homolog
cM	centiMorgan
CP49	Cytoskeletal protein 49
CTS	Cataract Shionogi (mouse)
CRYBB1	β -B1 crystallin gene
CRYBA1	β -A3/A1 crystallin gene
CRYGD	γ -D crystallin gene
dH ₂ O	Deionised water
DNA	Deoxyribonucleic acid
dNTP	Deoxy(adenosine/thymidine/cytosine/guanine) triphosphate
g	Acceleration of 1 gravity
GDS	Guanidine denaturing solution
GJA3	Gap junction protein, A 3

GJA8	Gap junction protein, A 8
HSF4	Heat shock transcription factor 4
hg19	Human genome assembly 19
ICR	Ihara cataract rat
JML	Johnston memorial laboratory (Lincoln University)
KCl	Potassium chloride
KH ₂ PO ₄	Potassium dihydrogen phosphate
LD	Linkage disequilibrium
LOD	Logarithm of odds
M	Molar
MAF	V-maf musculoaponeurotic fibrosarcoma oncogene homolog (avian)
mg	Milligram
MgCl ₂	Magnesium chloride
MIP	Major intrinsic protein
min	Minute
mL	Millilitre
mM	Millimolar
mRNA	Messenger ribonucleic acid
N	Normal or control
NaCl	Sodium chloride
NaHCO ₃	Sodium bicarbonate
Na ₂ HPO ₄	Disodium hydrogen phosphate
NKX6-1	NK6 transcription factor related, locus 1
nL	Nanolitre
nM	Nanomolar
NUDT9	Nudix (nucleoside diphosphate linked moiety X)-yype motif 9
OHC	Ovine heritable cataract
Oar1.0	<i>Ovis aries</i> genome assembly 1.0
OAR6	<i>Ovis aries</i> chromosome 6
PCR	Polymerase chain reaction
PBS	Phosphate-buffered saline
pH	<i>potenz</i> hydrogen (hydrogen potential)
PhD	<i>Philosophiæ doctor</i> (Doctor of Philosophy)
PLAC8	Placenta-specific 8

Poly-A	Poly-adenosine
Poly-T	Poly-thymine (primer)
Polb	DNA polymerase β
r^2	Pearson product-moment correlation coefficient, squared
RCS	Royal College of Surgeons (rat strain)
RLT	RNeasy™ lysis buffer (Qiagen)
RNA	Ribonucleic Acid
rpm	Revolutions per min
RT-PCR	Reverse transcription polymerase chain reaction
SAM-R/3	Senescence-accelerated mouse, strain R/3
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
s	Second
SHR	Spontaneously hypertensive rat
SNP	Single nucleotide polymorphism
Taq	<i>Thermus aquaticus</i> DNA polymerase
TE	Trishydroxymethylaminomethane-ethylenediaminetetraacetic acid buffer
Tris	Trishydroxymethylaminomethane
T-test	Student's <i>t</i> -test
USA	United States of America
UV	Ultraviolet
WBN/Kob	Wistar Bonn/Kobori (rat strain)
WDFY3	WD (tryptophan-aspartic acid) repeat and FYVE (Fab1, YOTB/ZK632.12, Vac1, and EEA1) domain containing 3
WGS	Whole genome shotgun
WFS1	Wolfram syndrome 1
μg	Microgram
μL	Microlitre

Chapter 1 Literature Review

1.1 Introduction

The function of the lens is to transmit light and focus it on the retina. Opacities in the lens, referred to as cataracts, scatter or absorb light and can lead to blindness in severe cases. The cause of most cataracts is unknown, and is being investigated using animal models. One animal model is Ovine Heritable Cataract (OHC), an autosomal dominant cataract used by Lincoln University to test cataract drugs and investigate cataractogenesis.

The human, mouse and bovine genomes have been fully sequenced. Along with the partial sequence of the sheep genomes, these assemblies can be used as a source of markers and cataract genes in order to find the gene responsible for OHC.

1.2 Physiology of the Lens

Eighty percent of the refraction of light entering the eye is carried out by the cornea (Hejtmancik *et al*, 2001). The lens serves to fine-tune the focussing, and can also change shape under conscious control to bring close or distant objects into focus, a process known as accommodation. When focussing on distant objects, the ciliary muscles are relaxed and the zonular fibres naturally pull the lens into an oblate shape. When focussing on near objects, the ciliary muscles contract, releasing tension on the zonular fibres and allowing the lens to assume a more spherical shape (Figure 1).

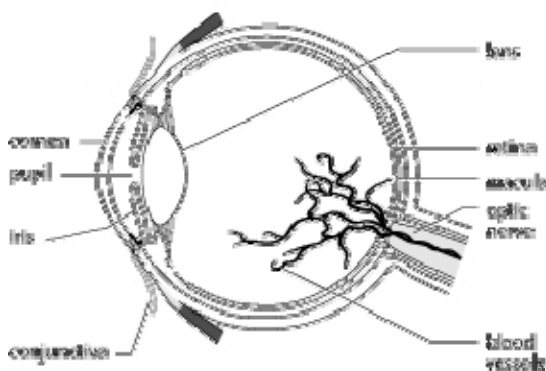


Figure 1: Cross-section of the human eye. The zonular fibres and ciliary muscles are shown at the top and bottom of the lens. From BUPA fact sheet at http://hcd2.bupa.co.uk/fact_sheets/html/lasik.html.

A transparent extracellular membrane known as the capsule surrounds the lens and contributes to the shaping of the lens during accommodation. During the normal aging process the lens becomes less elastic and less able to focus on near objects, with complete loss of accommodating power after the age of about 50 (Hejtmancik *et al*, 2001). Accommodating power declines continuously throughout life, with a 10-year-old child having twice the accommodating power of a 25-year-old.

In order to function properly the lens must be completely transparent with sufficient refractive index to focus light. To maintain the combination of transparency, refractive index, and elasticity needed for proper function, the lens has evolved as a unique tissue, consisting of cells with very high concentrations of proteins known as crystallins, and with greatly reduced or absent organelles. Most lens cells are elongated and are known as lens fibres.

Each cell in the lens follows the same general pattern of development (Figure 2). A single layer of cells, known as the epithelium, surrounds the anterior of the lens underneath the lens capsule. Just posterior of the equator, the epithelial cells divide and move laterally towards the equator (Pandey *et al*, 2002).

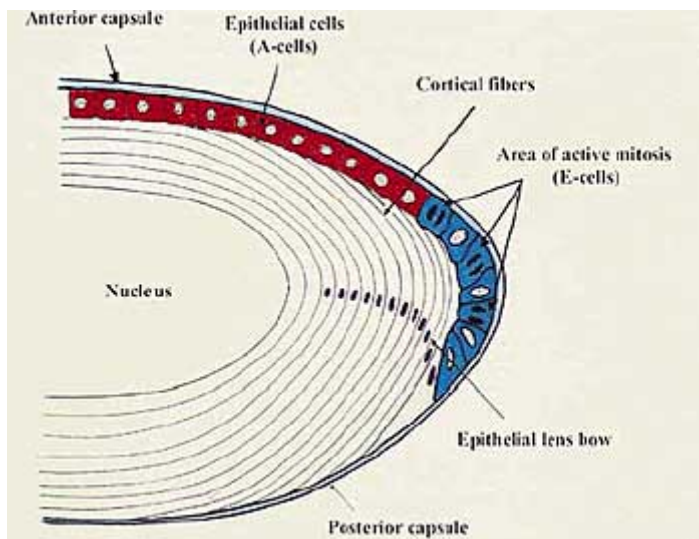


Figure 2: Cross-section of the lens. The normal epithelial cells are shown in red and the dividing cells in blue. The unshaded layers represent the differentiated cells which have lost most of their organelles. The layers continue into the nucleus of the lens but some have been omitted in this diagram. (Pandey *et al*, 2002)

The cells begin to differentiate, producing large amounts of crystallins. The cytoskeletons of the cells are modified, elongating them and making them wrap around the lens (Francis *et al*, 1999). The enzyme DNase II β is synthesised and targeted to the lysosomes (Nakahara *et al*, 2007). The lysosomes fuse with the nucleus, releasing DNase inside which breaks down the chromosomes (Nakahara *et al*, 2007). Then the nucleus itself disintegrates, releasing short fragments of DNA into the cytoplasm (Bassnett, 1997). Once the degradation of the nucleus is complete, no more transcription can occur and any remaining RNA decays rapidly in the cytoplasm (Faulkner-Jones *et al*, 2003).

Other organelles including the Golgi apparatus and the endoplasmic reticulum also break down (Bassnett, 1995). Proteins left over from the degradation are tagged with a poly-ubiquitin chain and targeted to the proteasome for destruction (Zandy and Bassnett, 2007). Organelle breakdown is thought to occur through multiple independent pathways, since degradation of the nucleus can be inhibited without affecting degradation of the other organelles (Nishimoto *et al*, 2003). The process of organelle degradation in lens cells resembles apoptosis, controlled cell death. However, the cysteine proteases known as caspases that cleave proteins in apoptosis are not required for the lens cell differentiation (Zandy *et al*, 2005).

This process produces the secondary lens fibres, which form after birth and make up the cortex of the lens (Francis *et al*, 1999). The cortex surrounds the nucleus, which consists of primary lens fibres formed before birth. These fibres originate from cells in the posterior of the lens at the first stages of development (Francis *et al*, 1999). The process of development is similar in both primary and secondary lens fibres.

Apart from the epithelium, the lens has a very low level of metabolic activity and protein turnover. The lens contains no blood vessels, but the cells making it up still consume nutrients and generate waste products. Because oxygen can only diffuse into the outermost layer of the lens and the inner layers contain no mitochondria to carry out respiration, sugar metabolism in deeper layers is anaerobic, producing lactate that diffuses out into the aqueous humour (van Heyningen, 1965).

Crystallins are water soluble structural proteins which are responsible for the transparency and elasticity of the lens. Despite the name, the lens fibres do not contain crystals of these proteins, which are instead arranged in a glass-like state with only short-range order (Delaye and Tardieu, 1983). Three types of crystallins are found in all vertebrate lenses, and these are referred to as the α -, β -, and γ -crystallins, named for the order in which they elute from a chromatography column (Lawson *et al*, 1981).

Other crystallins vary between species and are referred to as taxon-specific crystallins. These crystallins are almost always related or identical to metabolic enzymes (Piatigorsky and Wistow, 1989). For example, τ -crystallin is found in birds and reptiles and is identical to α -enolase. It is encoded by a single gene which is highly expressed in the lens for its structural role, and expressed to a lesser degree in the rest of the body for its enzyme function (Wistow *et al*, 1988).

In some cases the crystallin gene has been duplicated, allowing one form to be specialised for the lens function and one for the enzymatic function. For example, the bird and reptile δ -crystallin has two very similar forms encoded by two neighbouring genes (Piatigorsky *et al*, 1988). One form, $\delta 2$ -crystallin, has argininosuccinate lyase activity, whereas $\delta 1$ -crystallin is specialised for refraction in the lens and has no enzyme activity. The $\delta 1$ -crystallin is expressed more in the lens than $\delta 2$, but $\delta 2$ is still expressed mostly in the lens rather than in other tissues, indicating that it still functions as a crystallin (Piatigorsky *et al*, 1988).

A mammalian example of a crystallin produced by a duplicated gene is η -crystallin in elephant shrews (Macroscelidea) (Graham *et al*, 1996). The crystallin is also an aldehyde dehydrogenase and can be referred to as ALDH1 (Graham *et al*, 1996). One gene codes for the crystallin form, which is the predominant form of ALDH1 in the entire eye as well as the lens, but is expressly weakly in the liver (Graham *et al*, 1996). Another gene codes for “non-lens” ALDH1 (ALDH1-nl), which is the predominant form of ALDH1 in the liver, but is expressed weakly in the eye (Graham *et al*, 1996). Both proteins show aldehyde hydrogenase activity, and are closely related to the ALDH1 enzymes found in other vertebrates, which do not have a structural role in the lens (Graham *et al*, 1996).

An example of a taxon-specific crystallin not related to an enzyme is ι -crystallin, which was first discovered in the lenses of the diurnal gecko *Lygodactylus picturatus* (Röll *et al*, 1996). It is most closely related to cellular retinol-binding protein I (CRBP I), which stabilises and transports retinoids in the cytoplasm (Röll *et al*, 1996). In the gecko lens, ι -crystallin binds to 3-dehydroretinol and the complex absorbs UV light, protecting the retina against damage from sunlight (Röll *et al*, 1996).

The recruitment of proteins to act as crystallins, followed by duplication of genes and specialisation, may be the origin of all crystallins, with the relationship to the original enzymes obscured by evolution (Piatigorsky, 2003).

The α -crystallins are large oligomers made up of 40 copies of two subunits, α -A and α -B crystallin (Slingsby and Clout, 1999). Both subunits are 20 kDa in size and have related sequences. The C-terminal 100 residues of both subunits contain sequences related to heat-shock proteins. As well as its structural role, α -crystallin also acts as a chaperone, preventing the misfolding and aggregation of other crystallins (Horwitz, 1992). Both α -A and α -B crystallin are also expressed in other tissues (de Jong *et al*, 1993).

The β -crystallins are more complex oligomers, consisting of up to seven sequence-related polypeptides ranging in size from 20 to 30 kDa (Slingsby and Clout, 1999). The γ -crystallins are a family of monomeric proteins 21kDa in size, which share 30% sequence homology with the β -crystallins (Slingsby and Clout, 1999).

1.3 Cataract

1.3.1 Introduction

A cataract is defined as any opacity of the lens, regardless of whether it affects visual acuity or has any clinical significance (Koenigsberg, 1989). Opacities are usually caused by high molecular weight aggregations of lens proteins which scatter light. Abnormalities in the development of lens cells and disruptions to the structure of the lens fibres can also cause cataract. Specific insults such as dehydration or UV light are known to cause cataracts by disrupting the lens proteins and causing them to aggregate. Most human cataracts occur after the age of 45 with no obvious cause, and are referred to as senile or age-related cataracts (Hejtmancik *et al*, 2001). The mechanism of formation of age-related cataracts is unclear at present and is being actively investigated. Cataracts caused by a genetic defect are referred to as hereditary cataracts. Between 8 and 25 percent of congenital cataracts are hereditary and genetic factors can contribute to cataract at any age (Hejtmancik *et al*, 2001).

Cataracts can be classified as nuclear or cortical, as in the early stages of each cataract either the nucleus or the cortex develops opacities. Cortical cataracts commonly spread to the entire lens whereas nuclear cataracts remain confined to the nucleus. Nuclear cataracts are the most common form of age-related cataract and therefore the most common cataract in general (Hejtmancik *et al*, 2001).

In middle age the diffusion of small molecules between the nucleus and cortex of the lens is impeded (Sweeney and Truscott, 1998). This is referred to as the lens barrier and may be due to a layer of damaged cells in the interface between the cortex and the nucleus, or differences in cytoplasm viscosity. The lens barrier may contribute to nuclear cataract, and this would explain the distinction between cortical and nuclear cataracts.

Cataract is the leading cause of vision loss worldwide, with 20 million estimated to be blind from cataracts as of 2000 (World Health Organisation, 2000). At present the only treatment for cataract is to remove the lens and replace it with an artificial lens. This restores sight but not the ability to change the focus of the lens. However, most cataract surgery is performed on patients old enough to have already lost accommodating power. Cataract surgery is one of the most common surgical procedures, with one million extractions being performed each

year in the USA (Newell, 1992), and about 8,000 per year in New Zealand (2003/2004 figures, <http://www.beehive.govt.nz/ViewDocument.aspx?DocumentID=22876>).

1.3.2 Age-Related Nuclear Cataract

Age-related nuclear cataracts (ARN) are the most common form of cataract in humans (Truscott, 2003). Although the incidence of cataracts increases with age, cataracts are not an inevitable part of the aging process since there are many 80-year-olds with no significant opacities, and some 50-year-olds with the same cataract condition as older patients (Schmitt and Hockwin, 1990). Although the mechanism of cataract formation in ARN is not known, it is thought to involve long-term modification of the lens proteins. Since proteins in the nucleus form before birth and are not replaced (Bloemendal, 1977), a 60-year-old person has 60-year-old nucleus proteins and some modification of the lens proteins is expected over such a long time. Therefore the age-related nuclear cataract may be the product of a biochemical process that takes decades to complete but is not a result of the normal aging process.

Since UV light is known to damage tissues and the lens is exposed to it on a daily basis, it has been considered as a possible cause of ARN. However, since UV light passes through the anterior cortex of the lens before it reaches the nucleus, this would not be consistent with the opacities developing only in the nucleus. Further investigation has shown that UV light contributes to cortical but not nuclear cataract (Taylor *et al*, 1988).

Concentrations of the antioxidant glutathione (GSH) are significantly lower in the nucleus in ARN (Truscott and Augusteyn, 1977b). The concentration of the oxidised form of GSH increases in the same region, indicating that the loss of GSH is due to oxidation. Even in the most advanced cataracts the cortical concentrations of GSH can remain normal (Truscott and Augusteyn, 1977b). This means that it is unlikely that oxidants from outside the lens are responsible for the oxidation, as they would be expected to affect the cortex as well.

Oxidation is also observed in proteins in ARN. Over 90% of protein sulfhydryl (PSH) groups are lost, and half of all methionine residues are oxidised to methionine sulfoxide in the most advanced cataracts (Truscott and Augusteyn, 1977a). Oxidation of proteins can be detected in the earliest stages of the cataract, and as opacity of the nucleus increases the degree of protein oxidation also increases (Truscott and Augusteyn, 1977a). The correlation between oxidation

and opacity suggests that oxidative modification of protein is the underlying mechanism for opacification.

Normal human lenses at a variety of ages have been analysed for the same oxidative changes (Truscott, 2005). There were no significant changes in nuclear PSH content between the ages of 25 and 75, and the methionine sulphoxide content of the lens increased by less than 1% between birth and age 80. This indicates that the oxidative modifications found in ARN are a specific biochemical process, distinct from the normal changes in the lens with aging. GSH concentrations in the nucleus decrease with age, but to a lesser extent than in ARN, and there is no evidence of its oxidation products (Dickerson and Lou, 1997). GSH is generated in the cortex (Truscott, 2005) and the decline in its concentration in the nucleus with age is thought to be due to the lens barrier forming in middle age and impeding its inwards diffusion (Sweeney and Truscott, 1998).

A mechanism for ARN has been proposed (Truscott, 2005) where superoxide ion generated in mitochondria in the cortex of the lens diffuses into the nucleus and reacts with water to form hydrogen peroxide. In young lenses, GSH can penetrate the nucleus and reduce the peroxide back to water, preventing any oxidation of the nuclear proteins. In older lenses, the lens barrier impedes the diffusion of GSH into the nucleus and the peroxide is able to accumulate. This leads to oxidative modification of the nuclear proteins. Oxidised proteins are more susceptible to interchain crosslinking, which leads to aggregation and cataract. This mechanism means that cataract is more likely with increasing age, as the lens barrier forms and GSH concentration in the nucleus slowly declines. But the formation of cataracts is still dependent on several factors which are not related to age, explaining how the lens can stay free of cataracts even at an advanced age.

1.3.2 Cortical Cataract

Cortical cataracts involve changes in ion concentrations that are not observed in nuclear cataracts, including increased sodium and calcium and decreased potassium (Duncan and Bushell, 1975). Mature cortical cataracts also show a loss of dry weight due to a decrease of protein content in the lens. In human cataracts the degree of ionic disruption has been shown to be correlated with the loss of protein (van Heyningen, 1972). Experiments in cultured lenses have shown that loss of protein and opacity are independent of sodium and potassium concentrations but strongly dependent on an increase in calcium concentration (Marcantonio *et al*, 1986). Culturing human lenses in media with a relatively high calcium concentration (20mM) for 48 hours is enough to induce cortical opacities, the *in vitro* equivalent of cataracts (Hightower and Farnum, 1985).

The relationship between calcium concentration and protein loss and opacity is thought to be due to the cysteine proteases known as calpains (Sanderson *et al*, 1996). The physiological role of these enzymes is poorly understood, but they are known to be involved in cell motility (Calle *et al*, 2006) and the regulation of the cell cycle (Jánossy *et al*, 2004). Calpains are dependent on calcium for protease activity, and it has been shown that elevated calcium concentrations can overactivate calpains, leading to pathological proteolysis (Lee *et al*, 2000). Calcium-induced activation of calpain has been observed in cortical cataracts, and proteins extracted from the cataracts show cleavage characteristic of calpain proteolysis (Robertson *et al*, 2005). Specific inhibitors of calpain have been shown to prevent cortical opacities (Robertson *et al*, 2005), providing more evidence for calpain's role in cataractogenesis.

In the current model, increased calcium concentrations in the cortex of the lens activate calpains and lead to proteolysis of the crystallins. The cleaved proteins aggregate and scatter light, causing opacities. The cause of the elevated calcium concentrations is less well understood. Some genetic defects disrupt the normal transport of ions between lens cells, leading to increases in calcium (Baruch *et al*, 2001). Other external insults known to cause cataract could also trigger the same mechanism. The cortex is thought to be more vulnerable to this mechanism of cataractogenesis because it surrounds the nucleus and is exposed to external insults first.

1.3.3 Diabetes and Cataract

In humans, diabetes greatly increases the incidence of cataract (Bron *et al*, 1993). Animal models that involve feeding sugars to rodents or dogs have been developed to study diabetic cataracts, and the cataracts that are produced are referred to as sugar cataracts. These cataracts are produced by sugar alcohols accumulating in the lens and causing osmotic influx of water which damages the lens fibres (Lizak *et al*, 1996). The relevance of these cataracts to human diabetes is still unclear (Bron *et al*, 1993), but they show suture accentuation and the formation of cortical vacuoles (Lizak *et al*, 1996) that have also been reported in Ovine Heritable Cataract. (Brooks *et al*, 1983).

1.3.4 Cataracts in Non-Human Animals

Cataracts are known in a wide variety of non-human animals as well as in humans, including sheep (Brooks *et al*, 1983), cattle (Saunders *et al*, 1951), cats and dogs (Salgado *et al*, 2000) and horses (McLaughlin *et al*, 1992). Hereditary cataracts in non-human animals are described in sections 1.4.2 onwards. As well as genetic defects, cataracts in non-human animals can be caused by old age or as a complication of diabetes (Salgado *et al*, 2000).

1.4 Hereditary Cataracts

Hereditary cataracts are caused by an inherited genetic defect. The defect often causes a wide variety of other symptoms in the eye or the entire body with the cataract being a secondary development.

1.4.1 Human Hereditary Cataracts

Mutations responsible for human hereditary cataract occur in several different genes. They include crystallin genes, connexin genes, and genes for transcription factors. The inheritance pattern is usually autosomal dominant with complete penetrance. Autosomal recessive inheritance is very rarely seen (Beby *et al*, 2003).

1.4.1.1 Crystallins

Mutant crystallins often cannot form normal complexes and instead form very large aggregates of protein that scatter light and cause cataract. As described above, crystallins found in all vertebrates are divided into three groups, α , β , and γ , according to their chromatographic properties.

The α crystallin is composed of two polypeptide subunits known as α -A and α -B, coded by two genes *CRYAA* and *CRYAB*. Mutations have been observed in *CRYAA* which produce α -A proteins with an arginine replaced with a cysteine (R12C, R49C, R54C, and R116C). This leads to abnormal binding between the α -A and α -B crystallins and the formation of large aggregations of protein in the lens which scatter light. These mutations also cause abnormally small eyes and small corneas (Graw, 2009).

Another *CRYAA* mutation introduces a premature stop codon into the gene (Pras *et al*, 2000). The mutation is severe enough that the gene is not expressed at all and no abnormal protein is produced. This cataract is therefore one of the rare examples exhibiting autosomal recessive inheritance. Heterozygotes have sufficient α -A crystallin for normal lenses and no mutant protein is present, whereas homozygotes lack the protein altogether. Autosomal dominant inheritance is much more common in cataracts because the unique environment of the lens means that mutant proteins frequently cause a loss of transparency even if normal proteins are present (Andley, 2006).

Another *CRYAA* mutation produces a protein with a phenylalanine replaced with a leucine (F71L) and is associated with age-related cataract (Bhagyalaxmi, 2009). The structure of the α -crystallin complex is unchanged, but the chaperone activity of mutant α A-crystallin is decreased. This suggests that α -crystallin helps to prevent cataract by acting as a molecular chaperone.

Mutations have been reported in *CRYAB* (Graw, 2009 and Chen *et al*, 2009) which cause autosomal dominant cataracts of varied morphology. Another mutation in the same gene causes a recessive cataract (Safieh *et al*, 2009).

The β -crystallins are divided into two groups, the more acidic (β A-) and more basic (β B-) (Graw, 2009). Each subgroup has three different genes, *CRYBA1*, *CRYBA2*, *CRYBA4*, and *CRYBB1-3*. The *CRYBA1* gene codes for both the β A1 and β A3 crystallins from two in-frame AUG codons. The only difference between the two proteins is an extra 17 amino acid residues in the N-terminal region of β A3-crystallin (Andley, 2006).

Mutations in the β -crystallin genes are known to cause cataract (Graw, 2009 and Mothobi *et al*, 2009). The only β -crystallin gene not known to have cataract mutations is *CRYBA2*. The different crystallins give different cataract phenotypes (Andley, 2006).

The three γ -crystallin genes expressed at the highest levels in the human lens are *CRYGC*, *CRYGD*, and *CRYGS*. They are expressed at the beginning of embryogenesis and are therefore found in the nucleus of the lens and cause nuclear cataracts when abnormal.

Mutations associated with cataract are known in these γ -crystallins (Graw, 2009, Santana *et al*, 2009, Vanita *et al*, 2009, and Zhang *et al*, 2009). Some mutations cause the proteins to fold abnormally or to form abnormal disulphide bridges, leading to protein aggregation and cataract, whereas the mechanisms associated with other mutations are unknown (Beby *et al*, 2003).

1.4.1.2 Connexins

Connexins are proteins that form junctions between cells through which ions and metabolites can be exchanged (Beby *et al*, 2003). They are expressed in many locations throughout the body, but are particularly important in the avascular environment of the lens. After their synthesis connexins join together in groups of six in the Golgi apparatus, forming hollow

tubular structures known as connexons. As parts of the Golgi apparatus fuse with the cell's plasma membrane via transport vacuoles, the connexons become embedded in the plasma membrane, spanning the lipid bilayer. A junction is formed when two connexons on neighbouring cells bind together. Hydrophobic interactions bind the connexons together and isolate the junction channel from the intercellular space. These channels allow the exchange of ions, small proteins, and other small molecules. Several different mutations in human connexin genes are known to cause cataracts (Beby *et al*, 2003).

One example is the gene for connexin 50, known as *CX50* or *GJA8*. A mutation in this gene changes a cytosine to a thymine, which changes a proline to a serine in the connexin 50 protein. This alters the region of the protein involved in the opening and closing of the junction channel and causes a zonular cataract in humans (Shiels *et al*, 1998). Experiments with *Xenopus laevis* cells show that a cell with this mutation cannot establish a junction with a normal cell. A single mutant connexin in the connexon is enough to prevent the formation of a junction. This explains the dominant mode of inheritance of the cataract. Other mutations in the same gene cause similar effects (Arora *et al*, 2008, Hejtmancik, 2008, Lin *et al*, 2008, Schmidt *et al*, 2008, Vanita *et al*, 2008, and Yan *et al*, 2008).

Mutations in the gene for connexin 46 (*CX46* or *GJA3*) also cause cataracts, with a phenotype similar to that of mutations in *GJA8*. In one case (Baruch *et al*, 2001), it has been shown that the mutant proteins are not incorporated into the connexon. Experiments with knockout mice (Baruch *et al*, 2001) have shown that the cataracts caused by the disruption of *GJA3* are associated with an accumulation of γ -crystallin cleavage products, suggesting that proteolysis is involved in the formation of the cataracts. The cysteine protease inhibitor E-64 prevents cataract formation in the knockout lenses, and its targets are the calcium-dependent proteases m-calpain and Lp82. Increased Lp82 activity was detected in the knockout lenses, along with a decrease in the outflow and an increase in the inflow of calcium ions.

These results have been used to construct a model (Baruch *et al*, 2001) of cataract formation due to mutations in the connexin genes. The mutant connexin either prevents the formation of any connexons, or leads to the formation of abnormal connexons which cannot form gap junctions. The lack of gap junctions prevents calcium ions from being transported out of the cell, while the disruption to the cell membrane caused by the lack of gap junctions makes it more permeable and the inward flux of calcium ions increases. This leads to activation of

calcium-dependant proteases such as Lp82 and the calpains. These proteases cleave the crystallins, producing fragments that aggregate and scatter light. It is important to note that the crystallins themselves are normal and no mutations occur in the crystallin genes.

This general mechanism may also apply in other cataracts, with other inherited defects in the lens and external insults to the lens causing the initial increase in calcium instead of connexin mutations.

1.4.1.3 Transcription Factors

Transcription factors are regulatory proteins which control which genes are expressed in particular tissues. Three genes for transcription factors involved in the normal development of the eye are *HSF4*, *CHX10*, and *MAF* (Bu *et al*, 2002, Percin *et al*, 2000, Jamieson *et al*, 2002). Mutations in these genes cause cataracts, sometimes accompanied by microphthalmia and gaps in the iris (coloboma).

HSF4 is the gene for heat-shock transcription factor 4, which regulates the expression of several heat-shock genes (Tanabe *et al*, 1999). Heat-shock proteins act as molecular chaperones involved in protein synthesis, folding, repair and degeneration. They are important components of lens development. Human α -crystallins are also heat-shock proteins and carry out chaperone functions when expressed outside the lens (Piatigorsky, 2003). Mutations leading to cataract in humans have been identified in the regions of *HSF4* coding for DNA-binding regions of the protein, with a dominant pattern of inheritance (Bu *et al*, 2002).

Experiments with knockout mice (Min *et al*, 2004) show that in the absence of *HSF4*, several heat-shock proteins, proteases, and DNases are significantly less expressed in the lens. In particular, there was an 861-fold reduction in the expression of heat-shock protein HSP25. This protein is involved in regulating actin filament dynamics and cytoskeletal functions. In the lenses of knockout animals, fibre cells from the nucleus contained mitochondria, endoplasmic reticulum, and nuclei with intact membranes, showing that the normal process of organelle degradation failed to occur. The lenses never achieved transparency.

These results support a model where a mutation in *HSF4* gene produces a mutant protein which is unable to bind to DNA, reducing the expressions of many genes involved in the

normal development of the fibre cells of the lens (Min *et al*, 2004). The transparent fibre cells fail to form and some layers of the lens never become transparent. In this case the opacity comes from nearly normal epithelial tissue in the lens, not a protein aggregation. In humans the cataract occurs no earlier than 15 months and the nucleus remains transparent (Bu *et al*, 2002), whereas the knockout mice never produce any transparent lens tissue (Min *et al*, 2004).

CHX10 is a homeobox gene that codes for a transcription factor essential to the development of the eye. Mutations in this gene cause cataract in humans, along with iridal microphthalmia and coloboma, inherited in a recessive pattern (Percin *et al*, 2000).

MAF is another gene for a transcription factor where mutations cause defects in the eye as well as cataracts. Unlike *CHX10* the cataracts are cortical, more similar to OHC (Jamieson *et al*, 2002).

Other transcription factor genes associated with cataracts are *EYAI* (Azuma *et al*, 2000), *FOXE3* (Semina *et al*, 2001), *PAX6* (Hanson *et al*, 1999), and *PITX3* (Addison *et al*, 2005).

1.4.1.4 Other Genes

Another gene linked to hereditary cataracts is *AQP0*, which codes for Major Intrinsic Protein (MIP), also known as Aquaporin 0. MIP is found in the membrane of fibre cells and is responsible for the transport of water through the plasma membrane. Two different mutations have been found in this gene that lead to cataract in humans (Berry *et al*, 2000). Experiments on *Xenopus laevis* eggs have shown that cells expressing either of the mutant proteins are only one-third as permeable to water as normal cells (Francis *et al*, 2000). This suggests that mutations in *AQP0* produce mutant MIP which changes the water content of the cell and leads to the precipitation of the crystallins from solution, causing opacities. Although the opacities would be made up of intact crystallins, the crystallins themselves would be normal.

Mutations in *LIM2*, a gene for the membrane protein MP-20 which is found in junctions between lens fibre cells, are also known to cause cataracts. The function of MP-20 is poorly understood at present, so the mechanism of cataract formation is uncertain. It may be due to a disruption of enzymatic or transport function in MP-20 (Pras *et al*, 2002).

Mutations in the genes for beaded filament structural proteins 1 (*BFSP1*) (Ramachandran *et al*, 2006) and 2 (*BFSP2*) (Conley *et al*, 2000) also cause cataracts, both nuclear and cortical depending on the mutation. The two proteins combine to form beaded filaments, which are cytoskeletal structures of unknown function. The expression of *BFSP2* is restricted to the fibre cells of the lens (Beby *et al*, 2003).

Ferritin is a protein which stores iron by forming tiny iron oxide particles (Theil, 2003). The particles are stored inside the protein, which prevents the formation of an iron oxide precipitate and keeps the iron available to the metabolism. The production of ferritin is strictly regulated so that it is only produced when iron concentrations are above a threshold.

Mutations in the ferritin gene, *FTL*, disrupt this regulation and lead to increased production even when iron levels are normal (Girelli *et al*, 1997). This causes few problems in most of the body since the turnover of proteins is rapid. However protein breakdown in the lens is very slow and the increased production of ferritin causes oxidative stress and a disturbance of the equilibrium between soluble proteins. This leads to a distinctive cataract consisting of fine orange radial opacities (Girelli *et al*, 1997).

CHMP4B is gene whose product is involved in protein sorting and transport in the endosome-lysosome pathway. A mutation in *CHMP4B* causes posterior polar cataracts in humans which are inherited in an autosomal dominant manner (Shiels *et al*, 2007).

Nance-Horan syndrome is an X-linked disorder involving congenital cataracts, dental anomalies, and facial deformities. The syndrome is caused by mutations in *NHS*, a gene of unknown function on the short arm of the X chromosome (Burdon *et al*, 2003).

GCNT2 is the gene that codes for β -1, 6-*N*-acetylglucosaminyltransferase, an enzyme which is responsible for antigen development in red blood cells. Mutations in the gene have been reported which introduce a premature stop codon. This prevents the infant i antigen from being converted to the adult I antigen, and causes congenital cataracts which are inherited in an autosomal dominant manner (Pras *et al*, 2004).

EPHA2 codes for a protein tyrosine kinase receptor which is involved in tight junctions between cells. Expression of *EPHA2* is low in the anterior epithelial cells, increases with the

differentiation of cortical lens fibres, and is absent in the nucleus. Mutations in *EPHA2* are associated with age-related cortical cataracts (Jun *et al*, 2009).

The structural protein vimentin is a part of the cytoskeleton which maintains the transparency of lens fibres. A mutation in the human vimentin gene, *VIM*, is associated with a pulverulent cataract inherited in an autosomal dominant manner (Müller *et al*, 2009).

GFER is a gene on human chromosome 16 that codes for a protein that is thought to be involved in protein processing in mitochondria. A mutation in the gene causes a mitochondrial disorder in children which includes congenital cataracts (Di Fonzo *et al*, 2009).

1.4.1.5 Summary of Human Cataract Mutations

Table 1: Mutations linked to human cataracts. ASD stands for anterior segment dysgenesis.

Gene	Location	Mutation	Amino acid change	Phenotype
<i>CRYAA</i>	21q22.3	27 G→A	W9X	Age-related
		34 C→T	R12C	Cataract and microcornea
		61 C→T	R21W	Cataract, microcornea, microphthalmia
		62 G→T	R21L	Dense central
		149 C→T	R49C	Nuclear
		160 C→T	R54C	Total, microcornea
		292 G→A	G98R	Progressive total
		347 C→T	R116C	Nuclear
		213 C→A	F71L	Age-related
		<i>CRYAB</i>	11q23.1	32 G→A
58 C→T	P20S			Posterior polar
166 C→T	R56W			Varies
418 G→A	D140N			Lamellar
450 A deleted	Frameshift			Posterior polar
514 G→A	A171T			Lamellar
<i>CRYBA1</i>	17q11.2			279-81 GAG deleted
		Intron3 +1 G→A	Premature stop?	Varies
		Intron3 +1 G→C	Premature stop?	Pulverulent
<i>CRYBA4</i>	22q12.1	242 T→C	L69P	Cataract and microphthalmia
		317 T→C	F94S	Lamellar and microphthalmia
<i>CRYBB1</i>	22q12.1	667 C→T	Q223V	Nuclear
		682 T→C	S228P	Nuclear
		168 G deleted	Frameshift	Nuclear
<i>CRYBB2</i>	22q11.23	383 A→T	D128V	Ring-shaped
		607 G→A	V187M	Varies

<i>CRYBB3</i>	22q11.23	493 G→C	G165R	Nuclear
<i>CRYGC</i>	2q33.3	13 A→C	T5P	Nuclear lamellar
		212-22 GCGGC duplicated	hybrid protein	Zonular pulverulent
		327 C→A	C109X	Nuclear
		502 C→T	R168W	Lamellar
<i>CRYGD</i>	2q33.3	43 C→T	R14C	Progressive punctuate
		70 C→A	P24T	Varies
		70 C→T	P24S	Congenital non-nuclear
		109 C→A	R36S	Crystal deposition, greyish opacity
		176 G→A	R58H	Aculeiform
		181 G→T	G61C	Corraliform
		403 C→A	Y134X	No data
		418 C→T	R140X	Nuclear
		470 G→A	W157X	Central nuclear
		494 G deleted	Frameshift	Nuclear
<i>CRYGS</i>	3q27.3	105 G→T	G18V	Cortical
		116 C→G	S39C	Progressive
<i>GJA8</i>	1q21.1	68 G→C	R23T	Progressive nuclear
		131 T→A	V44E	Congenital total
		134 G→C	W45S	Jellyfish-like
		139 G→A	D47N	Congenital nuclear pulverulent
		142 G→A	E48K	Zonular pulverulent
		191 T→G	V64G	Congenital nuclear
		262 C→T	P88S	Zonular pulverulent
		263 C→A	P88Q	Lamellar pulverulent
		566 C→T	P198L	Star-shaped nuclear
		593 G→A	R198Q	Posterior subcapsular
		607 A inserted	Frameshift	Total congenital, nystagmus
		741 T→G	I247M	Zonular pulverulent
		776 G inserted	Frameshift	Triangular nuclear
		827 C→T	S276F	Congenital pulverulent

<i>GJA3</i>	13q12.11	7 G→T	D3Y	Zonular pulverulent
		32 T→C	L11S	Ant egg
		82 G→A	V28M	Cortical and capsular
		114 C→A	F32L	Nuclear pulverulent
		134 G→C	W45S	Nuclear
		176 C→T	P59L	Nuclear punctate
		188 A→G	N63S	Zonular pulverulent
		226 C→G	R76G	Total
		227 G→A	R76H	Nuclear
		260 C→T	T87M	Pearl box
		560 C→T	P187L	Zonular pulverulent
		563 A→C	N188T	Congenital nuclear pulverulent
		1138 C inserted	Frameshift	Zonular pulverulent
<i>HSF4</i>	16q22.1	341 T→C	L114P	Lamellar
		355 C→T	R120C	Marnier
		56 C→A	A20D	Lamellar
		1152 A→G	I87V	Lamellar
		Intron 12 splice	splice	Total, nystagmus
		524 G→C	R175P	Nuclear
		595-599 GGGCC deleted	deletion	No data
		218 G→A	R73H	Congenital total
<i>CHX10</i>	14q24.3	599 G→A	R200Q	Congenital
		599 G→C	R200P	Congenital
<i>MAF</i>	16q23.2	863 G→C	R288P	Lamellar pulverulent
		890 A→G	K297R	Congenital cerulean
<i>EYAI</i>	8q13.3	1320 G→A	R407Q	Congenital
		988 G→A	E330K	Congenital
		1177 G→A	G393S	Congenital, nystagmus
<i>FOXE3</i>	1p33	699 G inserted	Frameshift	Cataract, ASD
<i>PAX6</i>	11p13	553 G→T	G64V	No data
		459 G→C	A33P	No data
		489 T→C	S43P	No data

<i>PITX3</i>	10q24.32	656 17bp duplicated	Frameshift	Posterior polar congenital
		38 G→A	S13N	Congenital
		650 G deleted	Frameshift	Posterior polar congenital
<i>AQP0</i>	12q13.3	401 A→G	E134G	Varies
		413 C→G	T138R	Lamellar, sutural
		638 G deleted	Frameshift	Nuclear
<i>LIM2</i>	19q13.41	310 T→G	F104V	Presenile
<i>BFSP1</i>	20p12.1	736 966bp deletion	Frameshift	Developmental
<i>BFSP2</i>	3q22.1	697 GAA deleted	E233 deleted	Varies
		859 C→T	R287W	Lamellar
<i>FTL</i>	19q13.33	147 G→C		Varies
		A→G		Varies
		G→A		Varies
		10-38 deleted		Varies
<i>CHMP48</i>	20q11.22	386 A→T	D129V	Progressive subcapsular
		481 G→A	E161K	Posterior polar
<i>NHS</i>	Xp22.13	2387 C inserted	Frameshift	Congenital total
		3459 Frameshift	Frameshift	Congenital total
		1117 C→T	R373X	Congenital total
		718 G inserted	Frameshift	Congenital total
		400 C deleted	Frameshift	Congenital total
		3738 TG deleted	Frameshift	Congenital total
		2687 A deleted	Frameshift	Congenital total
		115 C→T	Q39X	Congenital total
		Intron A→G	splice	Congenital total
		2601 G inserted	Frameshift	Congenital total
		2635 C→T	R879X	Congenital total
		3624 C→A	C1208X	Congenital total
		1108 C→T	Q370X	Congenital total
		3908 11bp deleted	Frameshift	Prenatal
<i>GCNT2</i>	6p24.3	1043 G→A	G348E	No data

<i>GCNT2</i>		1148 G→A	R383H	No data
		Total deletion of gene		No data
		978 G→A	W326X	Congenital
<i>EPHA2</i>	1p36.13	Exon 13 G→A	R721Q	Age-related cortical
<i>VIM</i>	10p13	596 G→A	E151K	Pulverulent
<i>GFER</i>	16p13.3	581 G→A	R194H	Congenital

The mutations shown in Table 1 are taken from Andley, 2006, Arora *et al*, 2008, Bhagyalaxmi *et al*, 2009, Chen *et al*, 2009, Di Fonzo *et al*, 2009, Girelli *et al*, 1997, Graw, 2009, Hanson *et al*, 1999, Hejtmancik *et al*, 2008, Jun *et al*, 2009, Müller *et al*, 2009, Safieh *et al*, 2009, Schmidt *et al*, 2008, Vanita *et al*, 2008, and Yan *et al*, 2008.

1.4.1.6 Linked Regions

Some hereditary cataracts have been mapped to regions of the human genome without the gene responsible being identified. Posterior polar cataract has been mapped to the short arm of human chromosome 1, but there are no plausible candidate genes for the condition at present (Ionides *et al*, 1997). Age-related cortical cataract shows strong linkage to a 60 cM region of human chromosome 6 (Iyengar *et al*, 2004). There are no genes known to cause cataract in this region, but there are several genes for gap junction proteins, heat-shock proteins, and ion channels that could be possible candidates.

1.4.2 Rodents

Aside from clinical trials, experiments related to cataract research cannot be performed on humans for both ethical and practical reasons. Even extracted cataract lenses are of little use in research, as extraction methods which leave the lens intact have been mostly superseded by methods that homogenise the lens before extraction. Therefore animal models of cataract, mostly in rodents, have been the subject of extensive research.

Hereditary cataracts are known in mice, rats, and guinea pigs. The mouse models include CTS, Fraser, Lop, Philly, Nakano, Nop, cw Deer, Emory, Swiss Webster, Balb, and SAM-R/3. The rat models include ICR, Sprague-Dawley, Wistar, SHR, Dahl, WBN/Kob, and RCS.

Cataracts in mice and rats show a variety of pathologies and are caused by mutations in different genes, many analogous to human hereditary cataracts (Tripathi *et al*, 1991).

Knockout mice have been prepared for several human cataract genes, in order to determine the function of the genes and the mechanism of cataract formation.

A cataract in the guinea pig has been linked to a mutation in the taxon-specific crystallin ζ -crystallin (Huang *et al*, 1990). This crystallin is closely related to alcohol dehydrogenase and is also expressed in the liver and kidney (Huang *et al*, 1990). An AG deletion in intron 6 of the gene causes a splicing error during mRNA processing which leads to intron 7 being skipped (Rodriguez *et al*, 1992). This produces amino acid sequence 34 residues shorter than normal (Rodriguez *et al*, 1992). The cataract is congenital and inherited in an autosomal dominant manner (Huang *et al*, 1990).

The Emory mouse is particularly interesting in relation to OHC because it involves an autosomal dominant cataract that is slow to develop and involves the cortex first (Kuck, 1990). Increased calcium concentrations are also found in the cataract (Kuck, 1990). The metabolic defect involved in the Emory mouse cataract has not yet been determined.

Transgenic mice engineered to over-express the DNA polymerase- β gene develop cortical cataracts which progress to affect the entire lens (Sobol *et al*, 2003).

Rodent models have the advantage that rat and mouse genetics are very well understood and large litters can be easily bred. They have the disadvantage of having very small lenses that often need to be pooled for analysis and are difficult to dissect. Other animals develop hereditary cataracts but they have not received as much attention as the rodent models.

1.4.3 Canine

Of the 300 breeds of dog, 164 are known to exhibit hereditary cataracts (Gelatt and MacKay, 2005). In most studies of canine inherited cataract, limited numbers of dogs are examined and the mode of inheritance has not been validated (Wallace *et al*, 2005).

Studies of the Bichon Frise (Wallace *et al*, 2005) and the Miniature Schnauzer (Rubin *et al*, 1969), have discovered cataracts with autosomal recessive inheritance. This mode of inheritance is rarely observed in human cataracts. Recessive inheritance could be more commonly observed in dogs due to a higher rate of inbreeding which produces more homozygous animals. The ease with which a dominant cataract can be culled out of a

breeding population of dogs could also reduce the proportion of dominant cataracts which are observed.

Only one canine gene that causes hereditary cataract has been identified (Mellersh *et al*, 2006). This is *HSF4*, a transcription factor gene which also causes cataracts in humans (Bu *et al*, 2002) and in mice (Min *et al*, 2004). In Staffordshire Bull Terriers, Boston Terriers, and Australian Shepherds, mutations in *HSF4* cause a frameshifts that introduce a premature stop codon and produce truncated proteins. The resulting cataracts first occur a few weeks to months after birth and become total after 2-3 years. In the terriers the mutant protein has 27 aberrant amino acids whereas the mutant Australian Shepard protein has 86. The inheritance pattern is recessive in the terriers but dominant in Australian Shepherds (Mellersh *et al*, 2006).

A canine cataract that may be related to ovine heritable cataract is found in Chesapeake Bay Retrievers (Gelatt *et al*, 1979). This cataract affects the cortex of the lens and is inherited in an autosomal dominant manner with incomplete penetrance.

1.4.4 Bovine

Hereditary nuclear cataracts have been described in cattle (for example in Saunders *et al*, 1951), but the genes responsible for them have yet to be discovered.

1.5 Ovine Heritable Cataract

1.5.1 Introduction

In the late 1970s a breeder of New Zealand Romney sheep reported blindness in four out of 20 offspring of a stud ram and two related sheep. The sheep were examined by the Veterinary Science Department of Massey University and were found to have mature bilateral cataracts. Since the stud ram was unaffected, the mode of inheritance was thought to be recessive. If this were true, cataracts could represent an important genetic defect in the strain of New Zealand Romney. Therefore the development of the cataracts and their inheritance was studied. The cataract was the first reported occurrence of inherited cataracts in sheep and will be referred to as Ovine Heritable Cataract or OHC (Brooks *et al*, 1983).

1.5.2 Inheritance

The offspring of the original stud ram have been bred for several generations, producing both affected and normal animals. A flock of animals descended from the stud ram is maintained at Lincoln University. This flock will be referred to as the OHC flock, even though it contains animals that have not developed cataracts. Several years of breeding studies have shown that the pattern of inheritance of OHC is autosomal dominant (Brooks *et al*, 1983, Robertson *et al*, 2005). Since opacities are usually visible by two months after birth in affected animals (Brooks *et al*, 1983), this means OHC can easily be controlled by culling the affected animals before breeding and is not an important genetic defect from an economic perspective.

Since the original ram was unaffected, either the dams had cataracts, a mutation occurred in the germ cells of the ram, or the ram had an OHC genotype but did not develop cataracts because of incomplete penetrance. In the first three years of matings two out of 82 offspring had mature congenital cataracts and the rest were either born with normal lenses that slowly developed cataracts, or were unaffected (Brooks *et al*, 1982). Since the two lambs with congenital cataracts each had two affected parents, they were considered to be homozygous for the OHC gene. However, further breeding showed that mature congenital cataracts could be produced from breeding between normal, unrelated ewes and affected rams (Robertson *et al*, 2005). Since cataract in sheep is rare, it is very unlikely that these ewes had any alleles associated with cataract, even allowing for the possibility of incomplete penetrance.

Therefore there is no direct correlation between congenital cataracts and a homozygous genotype. The number of affected offspring is generally less than that expected for simple

dominant inheritance. All available breeding records are analysed in Chapter 2 to test the significance of this divergence.

1.5.3 Pathology

Lenses are usually normal at birth but develop focal anterior and posterior cortical opacities at 1-2 months of age months (Brooks *et al*, 1983). These coalesce to form spoke-like patterns that gradually become more diffuse, with the entire lens becoming opaque by 10-12 months (Brooks *et al*, 1983). Both cortical and nuclear fibres are eventually affected by the development of cataracts and in the mature state only a few equatorial fibres remain. In advanced cases the lens is partially absorbed, leaving a fluid-filled lens remnant. Occasionally mature cataracts are present at birth (Robertson *et al*, 2005). The general course of cataract development is similar to human cortical cataracts (Beby *et al*, 2003), but no human hereditary cataract is known which involves non-congenital opacities typically appearing in the first year of life.

When OHC is used in studies of cataract development, the eyes are examined with a slit lamp by a trained veterinarian and the lenses are classified into eight categories representing the progress of the cataract (Figure 3). When the entire lens is opaque (stages 5 and 6) the animal is totally blind to objects and motion. OHC does not significantly affect the animal's ability to forage for food, stay with the flock or reproduce (Robertson *et al*, 2005). In this investigation the same examination and classification was carried out, but for the purpose of linkage studies and mutation detection any degree of opacity was scored as "cataract".

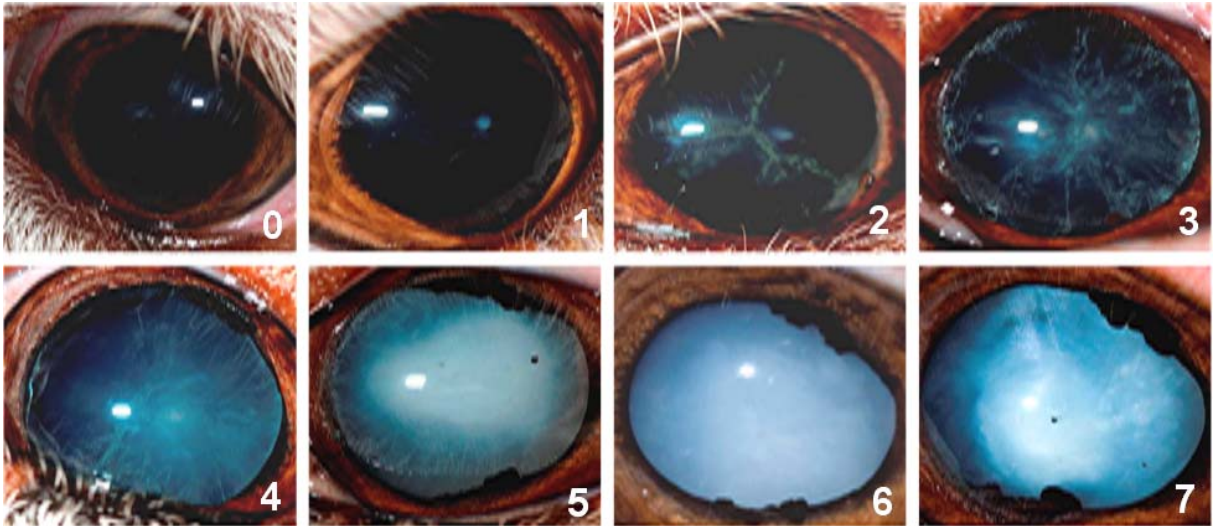


Figure 3: Progression of Ovine Heritable Cataract. The eight categories assigned to sheep lenses after inspection are shown (Robertson *et al*, 2005):

- 0) No visible opacity.
- 1) Small opacity detected at either the anterior or posterior suture line.
- 2) Small opacities detected at both suture lines.
- 3) Opacities at sutures and mild cortical involvement.
- 4) Moderate to severe cortical involvement.
- 5) Total immature cataract - reflection from retina still visible.
- 6) Total mature cataract - no reflection from retina.
- 7) Hyper-mature cataract – partial absorption of lens.

The OHC sheep seem otherwise healthy with no deformities of the eye, no abnormalities of growth, and no gross signs of any other condition. However most affected sheep are only kept for up to three years and no detailed veterinary examinations or necropsies have been performed, so it is possible that other effects of the mutation have gone undetected. The plasma glucose levels of OHC sheep are not significantly higher than normal sheep (Robertson, 2003), and therefore there is no evidence for diabetes in the OHC sheep.

1.5.4 Histology

With a light microscope, areas of fibre swelling and degeneration can be observed in the anterior cortex of lenses in the early stages of OHC. As the cataract progresses, the cortical fibres begin to contract into globules and cystic spaces develop beneath the posterior capsule. Eventually all of the lens fibres are affected, except that the formation of new fibres at the equator continues and therefore there are always a few normal fibres close to this location (Brooks *et al*, 1983).

In early cataracts, epithelial cells overlying cortical opacities develop vacuoles, whereas epithelial cells in other regions remain normal. As the cataract progresses, more epithelial cells become affected and the number and size of the vacuoles increase. Studies with an electron transmission microscope have shown that these vacuoles occur together with an increase in the amount of rough endoplasmic reticulum and may be dilatations of it. Epithelial proliferation occurs from the middle stages of cataract development and leads to clusters of epithelial cells projecting into the cortex (Brooks *et al*, 1983).

Scanning electron microscopy shows that cortical fibres in the early stages of OHC change from a hexagonal to a circular cross section and the interlocking processes between them become atrophied and irregular. In the later stages, the atrophy becomes extreme and the remaining fibres are convoluted and misshapened (Brooks *et al*, 1983).

Immunohistochemical studies of the development of the cataract (McDermott, 2007), using antigens specific for membrane proteins, confirm that the regular hexagonal pattern of the lens fibres is disrupted with progression of the cataract. In the earliest stage of cataract (stage 1 in Figure 3) normal lens morphology is observed beneath the lens capsule, down to 60 μm where there is a distinct disruption of the hexagonal structure. This transition may represent the boundary between differentiating outer-fibre cells and mature differentiated fibre cells. This boundary is the location of processing and cleavage of membrane proteins in normal lenses. The area of normal lens morphology stays intact in stage 2 of the cataract while the disruption increases below it. In stages 3 and above the morphology of the entire lens is disrupted (McDermott, 2007).

Most of these histological changes are observed in a wide range of cataracts, except for the vacuoles which develop in epithelial cells. Apart from OHC, these have only been observed in the Wistar rat strain and their origin is uncertain. They may be caused by the release of toxic substances from degenerating cortical fibres (Brooks *et al*, 1983).

1.5.5 Biochemistry

The first studies on OHC discovered increases in calcium and sodium concentration and decreases in potassium concentration with the progression of the cataract, consistent with human cortical cataracts (Brooks, 1983). Further experiments showed that the increase in calcium from normal to cataract lenses, and with the progression of the cataract, was

statistically significant ($P < 0.05$) (Robertson 2003). SDS-PAGE and immunoblotting of lens proteins showed proteolysis products characteristic of m calpain, and that m calpain activity as measured by casein zymography significantly decreased during cataract formation ($P < 0.05$) (Robertson 2003). This loss of activity is an indication of previous activation and autolysis, and has been observed in the equivalent situation in post-mortem meat. The m calpain gene was upregulated in affected lenses. A specific inhibitor of calpain known as SJA6017 significantly slowed the development of cataracts when applied to the eyes of affected sheep (Robertson 2003). All of these results indicate a mechanism in which some genetic defect produces a higher calcium concentration inside the lens, which activates calpain, breaks down crystallins and causes opacities. Mutations in connexin genes are known to have this effect in humans (Shiels *et al*, 1998) and mice (Baruch *et al*, 2001), but many other genes could also be involved. A mutation in a crystallin gene is not supported by the biochemical evidence as it would not increase the concentration of calcium.

1.5.6 OHC as an Experimental Model

OHC is a convenient model for both *in vitro* and *in vivo* studies of cataract since sheep lenses are about twice as large as human lenses and are therefore easy to dissect and manipulate. As long as sufficient land is available, as at Lincoln University, sheep are no more expensive per animal to use in experiments than the usual rodent cataract models. OHC is the only cataract model based on hereditary cataracts in a large non-human animal. The canine hereditary cataracts described in section 1.4.3 also occur in a large non-human animal but have not yet been used as an experimental model for human cataracts. Lambs with OHC have already been used in trials of a cataract prevention drug based on a calpain inhibitor (Robertson, 2003). One difficulty in using OHC as a model is that no known homozygotes are available at present. This means that all matings produce a combination of wild-type and mutant lambs, which cannot be distinguished until cataracts develop. This makes the testing of preventative treatments more difficult. A reliable genetic marker for OHC would enable the cataract status of lambs to be predicted at birth, and would allow homozygotes to be identified before breeding.

1.6 Linkage Mapping and Microsatellites

During meiosis, homologous chromosomes cross over randomly and recombine, giving recombinant chromosomes with a mixture of paternal and maternal alleles. The closer two loci are on a chromosome, the less likely it is that they will be separated during this process. If two loci are close enough that their inheritance deviates significantly from random assortment, they are said to be linked and this phenomenon is known as genetic linkage (Sturtevant, 1913). Genetic linkage is classified by the linkage phase. Two alleles on the same chromosome are linked in coupling phase, while two alleles on different homologous chromosomes are linked in repulsion phase (Liu, 1998).

Genetic linkage is measured in centiMorgans, where 1 cM is equivalent to a 1% chance of recombination in a single mating (Haldane, 1919). The physical distance equivalent to 1 cM varies with species and between males and females due to differences in recombination frequencies (Petkov *et al*, 2007). The sex-averaged recombination rate over the entire human genome is $1.30 \pm .80$ cM per megabase, equivalent to 770 kilobases per cM (Yu *et al*, 2001). Different areas on the same chromosome can also have higher or lower rates of recombination which changes the relationship between physical distance and cM (Hey, 2004).

The cM scale breaks down at large distances, since 100cM is equivalent to 100% chance of recombination and is therefore indistinguishable from the two loci being on different chromosomes, even though a single chromosome can be larger than the equivalent of 100 cM. Larger cM distances are given in Kosambi cM, which are adjusted to give a more consistent scale across the entire chromosome (Felsenstein, 1978).

Linkage mapping is a method of determining the distance between two loci on a chromosome based on their genetic linkage. Linkage can be detected by the patterns of inheritance of the phenotypes and/or genotypes associated with each locus. Linkage mapping is often used to find the position of a gene responsible for an inherited trait or disease in relation to loci with known positions. This position can be used to select candidate genes, as described in section 1.9.4.

With larger pedigrees the evidence for linkage becomes stronger. This is expressed with Logarithm of Odds (LOD) scores (Morton, 1955).

A LOD score is the logarithm to base 10 of the ratio between the probabilities of the pedigree genotypes being produced with linkage and without linkage, at a particular recombination fraction. For example, a LOD score of 2 means that a pattern of pedigree genotypes is 100 (10^2) times more probable if there is linkage than if there is no linkage, at the recombination fraction used in the calculation. A LOD score of 3, equivalent to the genotypes being 1000 times more likely with linkage, is the generally accepted standard for significant linkage (Morton, 1955).

Calculation of LOD scores become complex with large pedigrees and other factors such as incomplete penetrance and missing genotypes, so computer programs such as CRIMAP (Green *et al*, 1990), are used to calculate the LOD scores and estimate the cM distances. The distances between pairs of loci can be used to construct a linkage map of all the loci across the chromosome.

Microsatellites, also referred to as Simple Sequence Repeats (SSRs), are regions of the genome where a short sequence is repeated several times (Weber and May, 1989). The length of the sequence varies from two to five bases and the number of repeats is usually in the dozens. During replication of the chromosomes the DNA polymerase can “stutter” and add repeats of the sequence. Therefore variations in the number of repeats are much more common than other mutations (Weber and May, 1989).

PCR using primers designed to match non-variable sequences on either side of a specific repeat region can amplify the entire repeat region, showing the number of repeats by the size of the product (Weber and May, 1989).

Microsatellites are often used as markers in linkage analysis. The number of repeats is stable enough, over the time of most pedigrees, to be treated as fixed. Therefore each individual inherits an unchanged repeat region from each parent and the number of repeats can be traced down the pedigree (Litt and Luty, 1989). A problem with microsatellites is the necessity to detect changes in size of as little as two bases, meaning agarose gels are not of sufficient resolution and polyacrylamide gels or chromatographic methods are necessary. An additional problem is that the stuttering effect also occurs in the PCR cycle, producing less intense bands at different sizes that make analysis more complicated (Litt and Luty, 1989).

Genome scans are carried out on as large a pedigree as possible, starting with a few markers on as many chromosomes as possible and then using more and more markers in regions that show linkage. Ideally all individuals in the pedigree would be tested for all markers to provide the maximum amount of data. However to make the process more efficient a breakpoint panel, consisting of animals with recombinations in the linked region, can be tested in the later stages (Bartoloni *et al*, 1998).

1.7 Genome Assemblies

In this project, a large number of genetic markers were used to determine genotypes in both normal and affected animals from the OHC flock. In the early stages of this process the markers were chosen from ovine linkage maps (Maddox *et al*, 2001). In the later stages more markers were required and these had to be found in the bovine genome sequence.

Two draft sequences of the human genome were published in 2001, one by a consortium of government and university laboratories (International Human Genome Sequencing Consortium, 2001) and one by the Celera Genomics corporation (Venter *et al*, 2001). A draft mouse genome sequence was published in 2002 (Mouse Genome Sequencing Consortium, 2002).

The Celera Genomics sequence was generated by the Whole Genome Shotgun (WGS) strategy, where the entire genome is broken up into short fragments which are individually sequenced (Venter *et al*, 2001). The sequences are analysed by computer software to reconstruct the entire genome sequence. Sequencing is usually repeated several times over the genome to reduce errors. Each whole genome sequence prepared by this method is referred to as an assembly.

The final human genome sequence, covering 99% of the euchromatic genome and accurate to 1 error in 100,000 bases, was completed in 2004 (International Human Genome Sequencing Consortium, 2004). A preliminary bovine genome assembly, Btau2.0, was released in March 2005 (Womack, 2005). This had 6.2X coverage of the bovine genome and was also prepared by WGS. Comparisons between this assembly and linkage and radiation hybrid maps of the bovine genome found 133 incorrectly assigned loci and localised inversions (McKay *et al*, 2007). These errors were corrected in the preparation of the next assembly, Btau3.1.

Btau3.1 was released in August 2006 and had 7.1X coverage. It was prepared from a combination of WGS and Bacterial Artificial Chromosome (BAC) sequences (The Bovine Genome Sequencing and Analysis Consortium *et al*, 2009). Comparison of this sequence to a

radiation hybrid map of bovine chromosome 14 showed that there were still errors in the order of scaffolds on that chromosome (Marques *et al*, 2007).

The next bovine genome assembly was named Btau4.0 and was released in October 2007. It had the same degree of coverage as Btau3.1 but used different mapping methods to assign sequences to chromosomes (The Bovine Genome Sequencing and Analysis Consortium *et al*, 2009). Comparison with linkage maps showed that less than 0.8% of markers were incorrectly positioned in this assembly (The Bovine Genome Sequencing and Analysis Consortium *et al*, 2009). . Details of all assemblies are available at <http://www.hgsc.bcm.tmc.edu/project-species-m-Bovine.hgsc?pageLocation=Bovine>.

1.8 Tools in Ovine Genetics

Large numbers of microsatellites have been discovered on the sheep genome, and located on known chromosomes, with known linkage to other microsatellites and to genes. This information is collected in linkage maps. One example is the linkage map prepared by researchers from several universities (Maddox *et al* 2001), which includes 941 different microsatellites and 121 genes, with an average spacing of 3.4 cM on the autosomes. The linkage map is available at <http://rubens.its.unimelb.edu.au/~jillm/jill.htm>, along with primers and PCR recipes for amplifying each microsatellite. These resources have been used in genome scans to find the genes responsible for hereditary diseases in sheep, for example in a study which found that the gene causing Spider Lamb Syndrome was on the telomeric end of chromosome 6 and that fibroblast growth factor receptor 3 was a likely candidate (Cockett *et al*, 1999). Other examples include the study of inherited ovine arthrogryposis (Murphy *et al*, 2007), which identified ovine chromosome 5 as the likely location of the associated gene, and a study of ovine neuronal ceroid lipofuscinosis (Broom *et al*, 1998) that associated the condition with ovine chromosome 7.

Known sheep microsatellites are too widely spaced for high resolution genome scans, and high quality sheep genome sequence assemblies were not available at the time this genome scan was carried out, but the bovine genome was a useful and informative proxy.

Microsatellites found in bovine sequences are commonly present in the homologous ovine sequences. A study of 251 cattle microsatellites (Crawford *et al*, 1995) found that 184 (73%) were present in sheep, although only 133 (53%) were polymorphic and therefore useful for linkage mapping.

Sheep and cattle are closely related, and therefore whole genome sequencing of sheep has not been funded to the extent that cattle sequencing has (Dalrymple *et al*, 2007). It has been estimated that there is 90% conservation of non-coding sequence between cattle and sheep (Cockett, 2006). A “virtual” sheep genome has been constructed by locating the end sequences of sheep BACs using the cattle, canine, and human genome assemblies, then reordering the human genome sequence based on the positions of the BACs (Dalrymple *et al*, 2007). Annotations of the other genomes, such as the positions of genes and markers, can be transferred to the virtual sheep genome. The virtual sheep genome has been aligned with a

radiation hybrid map of sheep chromosome 1 (Wu *et al*, 2009). The degree of discrepancy between the two was less than 1.86%.

The virtual sheep genome has been used to construct a partial “real” sheep genome assembly, Oar1.0, which was made available in May 2008. The bovine genome was reordered based on the virtual sheep genome, sheep sequences were positioned by alignment with the bovine sequences, and then the bovine sequences were replaced with the sheep sequences. Details of this process are available on the CSIRO Livestock Genomics website at <http://www.livestockgenomics.csiro.au/sheep/oar1.0.php>. The real assembly covers only 43% of the total genome sequence and is lacking most of the annotations available for the virtual sheep genome.

1.9 Identifying Disease-Causing Genes and Mutations

1.9.1 Introduction

Identifying the genes and mutations responsible for inherited diseases can provide information on the disease mechanism and possible preventative and therapeutic measures. Three strategies for the selection of candidate genes for an inherited trait are the traditional candidate gene strategy, the comparative genomics strategy, and the positional candidate gene strategy (Zhu and Zhao, 2007).

1.9.2 Traditional Candidate Gene Strategy

The traditional approach considers all known genes of the organism of interest and selects those which are known to have functions relevant to the trait. Each candidate gene is analysed to find variations associated with the trait, and then the mechanism of the formation of the trait from the mutation is investigated.

The first and most famous application of this strategy was in the study of sickle-cell anaemia (Pauling *et al*, 1949). This condition involves changes to the shape of red blood cells when oxygen is removed from them. Since haemoglobin is the protein involved in binding oxygen in the red blood cells, it was selected as the candidate protein for investigation. The investigation pre-dated the discovery of the genetic code so it was carried out on proteins instead of DNA. A difference was found in the primary structure of haemoglobin from sickle cell patients. This was the first time a genetic disease had been linked to the mutation of a specific protein.

The candidate gene approach worked in this case because the link between the pathology of sickle-cell anaemia and haemoglobin was clear. There are many other traits with the same clear connection to known genes that have been successfully investigated using the same approach. However, this approach relies on biochemical and functional knowledge about the traits which is not always available, making the selection of candidate genes from the large number of known genes difficult (Zhu and Zhao, 2007).

1.9.3 Comparative Genomics Strategy

The comparative genomics strategy takes advantage of homologous genes in other organisms which are known to be responsible for similar traits. Information on the genetics of humans, mice, rats, and other species which have been the subjects of extensive study has frequently been used to select candidate genes for traits in livestock (Zhu and Zhao, 2007).

An example is the double-muscled phenotype in cattle. The myostatin gene was selected as a candidate for this trait because a mouse model which lacked the gene showed a similar phenotype to the cattle (Grobet *et al*, 1997). A homologue to the mouse myostatin gene was subsequently discovered in cattle. Sequencing of the cattle myostatin gene revealed an 11-base deletion leading to a premature stop codon which prevents the synthesis of functional myostatin protein and causes the double-muscled phenotype (Grobet *et al*, 1997).

A similar approach was used in this study to select chromosomes for genome scanning, as described in Chapter 3.

1.9.4 Positional Candidate Gene Strategy

The positional candidate gene strategy involves using the genetic linkage techniques described in section 1.6 to determine the chromosome region linked to the disease. Candidate genes are then chosen from the identified chromosomal region (set of genes in the region of interest), not from the entire genome (Zhu and Zhao, 2007). This approach can be combined with the comparative genomic approach to select genes in the region of interest which are associated with similar traits in other organisms.

An example is the identification of a gene for milk yield in cattle, DGAT1 (Grisart *et al*, 2002). A quantitative trait locus (QTL) for milk yield was mapped to a 3cM interval on the telomeric end of cattle chromosome 14. DGAT1 was known to be involved in triglyceride synthesis and a mouse model lacking it did not lactate. The bovine homologue of the gene was shown to lie in the region that contained the QTL, making it a positional candidate gene. When the gene was sequenced in several animals, a mutation was discovered which changed a conserved amino acid. This mutation was associated with a decrease in milk yield along with an increase in fat yield, as predicted from the QTL, so it was concluded to be the gene responsible for the QTL.

An example of this approach being used to find a gene responsible for cataract is the mutant rat strain SHR-Dca (Liška *et al*, 2008). This mutation causes cataract in heterozygotes and microphthalmia in homozygotes. The gene responsible was mapped to an 8.6 megabase region on rat chromosome 2 which contained the gene GJA8. This gene codes for the gap junction protein connexin 50, which was known to be responsible for cataracts in humans (Shiels 1998). A mutation was discovered in the gene which changed an amino acid and segregated perfectly with the mutant phenotypes (Liška *et al*, 2008). It was therefore concluded that GJA8 was responsible for the SHR-Dca cataract. As described in Section 3.2.3, chromosome 1 is the most likely location for GJA8 in sheep, and markers on this chromosome show no significant linkage to OHC. Therefore GJA8 has been ruled out as the gene associated with OHC.

1.9.5 Single Nucleotide Polymorphism Chips

Single nucleotide polymorphisms (SNPs) are natural variations in single nucleotides that occur throughout the genome. They can be used as markers in a similar way to microsatellites, although they are more difficult to detect. Recently SNP chips consisting of arrays of molecular probes have become available which allow hundreds of thousands of SNPs to be genotyped simultaneously, providing linkage data for inherited traits across the entire genome (Schymick *et al*, 2007). This method of analysis is faster and more informative than genome scans using microsatellites due to the closer spacing of SNPs. SNP chips were first created for human SNPs, but chips containing probes for sheep SNPs have recently been released (Smith *et al*, 2010). Sheep SNP chips were not available at the time the genome scan in this investigation was carried out, but they would be the most efficient method if the investigation was repeated now.

1.9.6 Next Generation Sequencing

Next generation sequencing refers to systems that produce hundreds of thousands of sequencing reactions in parallel, as opposed to capillary sequencing which generates one sequence at a time (reviewed in Turner *et al*, 2009). Sequence data can be combined to give continuous sequences of up to 2 kilobases, much longer than from capillary sequencing (Turner *et al*, 2009). Several different systems are used, mostly involving microscale solid phase components such as beads, cells, or wells. Next generation sequencing makes sequencing the entire genome of an organism much faster and cheaper. Already positional candidate approaches in the genetics of organisms such as yeast are declining in popularity

because it is easier to sequence the entire genome and search for variations. It is expected that the same approach will eventually be possible in organisms with larger genomes such as vertebrates (Turner *et al*, 2009). Even with current technology it is possible to sequence entire linked regions from linkage studies instead of individually testing candidate genes. No next generation sequencing equipment was available at the time this investigation was carried out.

1.10 Hypothesis and Aims

The hypothesis behind this investigation is that OHC is mostly caused by a mutation in a single autosomal gene. The aims of this investigation are to localise this gene on the sheep genome through a partial genome scan, use fine mapping to further refine the position of the gene, choose positional candidate genes through biochemical and clinical criteria and evaluate them as candidates for the OHC gene through sequencing.

Chapter 2 Mode of Inheritance of Ovine Heritable Cataract

2.1 Introduction

2.1.1 Massey University Breeding

OHC was originally reported in the offspring of a Romney stud ram who was unaffected himself (Brooks *et al*, 1982). This suggested a recessive mode of inheritance which could make OHC an important genetic defect in New Zealand Romney sheep depending on the frequency of the disease allele in the population. Two of the ram's affected offspring were studied by Massey University and bred with each other and with normal unrelated sheep. The results showed OHC being inherited as a simple autosomal dominant trait (Brooks *et al*, 1982). This meant that the trait could easily be removed from flocks by culling and would not be economically important. In this chapter the breeding records from Massey University (and Lincoln University) are re-analysed to confirm that they show dominant inheritance.

2.1.2 Lincoln University Breeding

Descendants of the OHC sheep were transported to Lincoln University for further study and breeding. Affected rams were bred with large numbers of normal ewes to produce offspring likely to develop cataracts. These offspring were then used to study the development of cataracts, and also as experimental animals in trials of a drug designed to prevent cataracts (Robertson *et al*, 2005). In addition, affected rams were bred with affected ewes in an attempt to produce an animal homozygous for the OHC gene. With simple dominant inheritance, a homozygote would have produced only affected offspring from normal ewes, which would have made the drug trials and cataract studies more efficient.

During the breeding at Lincoln University, fewer affected animals were produced than expected and no animal had enough affected offspring to be considered homozygous for the cataract gene, even allowing for incomplete penetrance. Two possible modes of inheritance that would explain these results are incomplete penetrance and an embryonic lethal homozygote. In this chapter, breeding records from Lincoln University will be analysed to detect any significant deviation from simple dominant inheritance.

Little information was available on the details of the Massey University breeding and phenotyping, and the relationship between animals bred by Massey University and the founding animals for the Lincoln University breeding was unclear. The Lincoln University breeding produced more than 1,000 offspring whereas the Massey University breeding produced only 87. Therefore the Lincoln University breeding was considered the most reliable source of data and no combined analysis was performed.

2.1.3 Incomplete Penetrance

Penetrance is a term for the proportion of offspring carrying a particular genotype that also express the trait (phenotype) normally associated with that genotype (Vogt, 1926). If an inherited trait shows incomplete penetrance, not all of the individuals with the genotype will show the expected phenotype and the penetrance will be less than 1. Penetrance can be calculated by dividing the observed proportion of offspring with a trait by the expected proportion for a gene with complete penetrance (Latter *et al*, 2006).

Incomplete penetrance can be caused by another gene suppressing the expected trait. In this case the penetrance reflects the frequency of the genotype that suppresses the trait in the population being studied (Blanc *et al*, 2003). Environmental factors such as temperature can also cause incomplete penetrance (Krafka, 1920).

2.1.4 Lethal Embryonic Homozygote

In lethal embryonic homozygote inheritance, a trait is inherited in a dominant manner except that homozygotes die before the trait is recorded. This is common because a homozygous genotype often leaves an individual without any functional copies of a gene which is necessary for survival. Often the homozygotes die very early in development and cannot be included in breeding records. This is one effect which can distort the proportion of offspring with each phenotype (Castle and Little, 1910).

Cataracts have been known to be inherited with a lethal homozygote effect. For example a study of γ -irradiated mice found 7 different dominant cataract mutations that were lethal in homozygotes (Kratochvilova, 1981). The genes responsible for the mutations were not identified, but none of the cataract phenotypes were the same as OHC.

2.2 Massey University Breeding

OHC was first observed in the offspring of a single ram. As described in Brooks *et al*, 1982, two of his offspring, sheep 676 and sheep 1836, were bred with normal unrelated sheep and with each other. The eyes of all offspring were inspected without equipment by the attending shepherd within a few days of birth. The eyes were also examined at least three times between 2 and 8 months after birth. During these examinations atropine drops were administered 30 minutes prior to a close examination of eye with a pen light in semi-darkness. Each lens was classified as normal or affected (Brooks *et al*, 1982). The production of affected offspring from matings between affected sheep and normal unrelated sheep suggested a dominant manner of inheritance (Figure 4). The normal sheep could also be carriers for a recessive condition, but cataracts are rare enough in sheep for this to be unlikely.

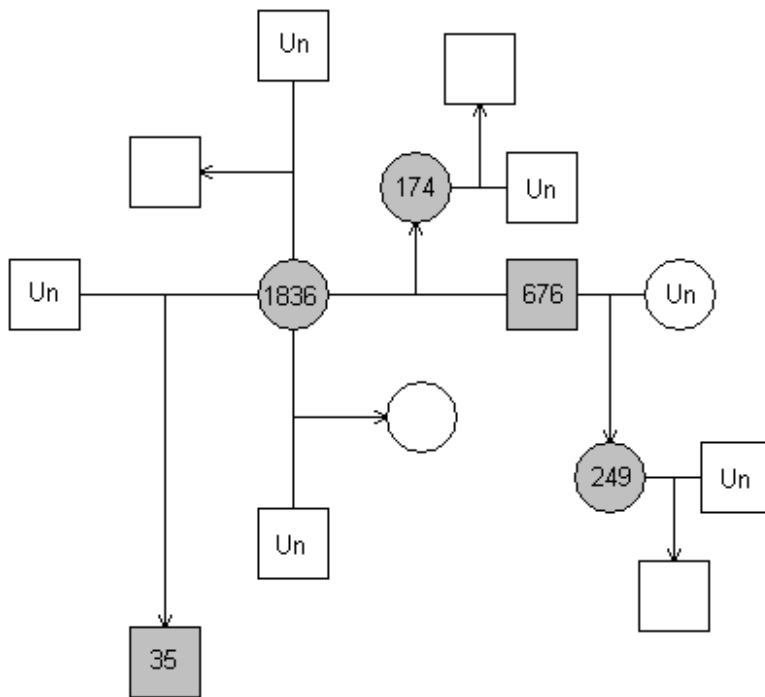


Figure 4: Early breeding at Massey University. Squares represent males and circles represent females. Blank symbols represent normal animals and shaded symbols represent affected animals. Symbols labelled "Un" represent normal unrelated sheep.

All affected animals except 174 have at least one normal parent, so they are assumed to be heterozygotes. Affected ewe 174 has a single offspring that is normal, so 174 is also assumed to be a heterozygote.

Affected ram 35 was mated with 54 normal unrelated ewes and then with 9 affected daughters from these matings, producing a total of 82 offspring (Brooks *et al*, 1982).

Table 2: Cataract status of offspring from Massey University matings. Standard errors are given by the square root of the variance of a binomial distribution. The binomial distribution is defined by the proportion of cataract offspring and the total number of offspring. Data from Brooks *et al*, 1982.

Mating Type	Normal	Affected	Total	Proportion Cataract	Standard Error
Affected-Normal	42	37	79	0.47	0.06
Affected-Affected	3	7	10	0.7	0.1
Total	45	44	89		

The affected-normal matings give 42 normal offspring and 37 (0.47 ± 0.06) affected offspring (Table 2), when the expected proportion is 0.5 affected for dominant inheritance from one heterozygote and one normal homozygote. The standard error is given by using the total number of animals ($n=79$) and the proportion of affected animals ($p=0.47$) as the parameters of a binomial distribution. The variance of the binomial distribution is given by $np(1-p)$, and the standard error is the square root of the variance, expressed as a proportion.

The affected-affected matings give 3 normal offspring and 7 (0.7 ± 0.1) affected offspring (Table 2), when the expected proportion is 0.75 affected for dominant inheritance from two heterozygotes.

Some of the affected animals in Table 2 are likely to be homozygotes, but there are no records of their offspring that would distinguish between homozygotes and heterozygotes.

For both affected-normal and affected-affected matings, the expected proportions are within the standard error of the observed proportions, and therefore simple dominant inheritance is consistent with these results. This statistical test has previously been used to detect incomplete penetrance in an inherited trait (Latter *et al*, 2006). In conclusion, the breeding carried out at Massey University shows that OHC is likely to be inherited in a dominant manner. There is no evidence from this breeding of incomplete penetrance.

2.3 Lincoln University Breeding

2.3.1 Pedigree

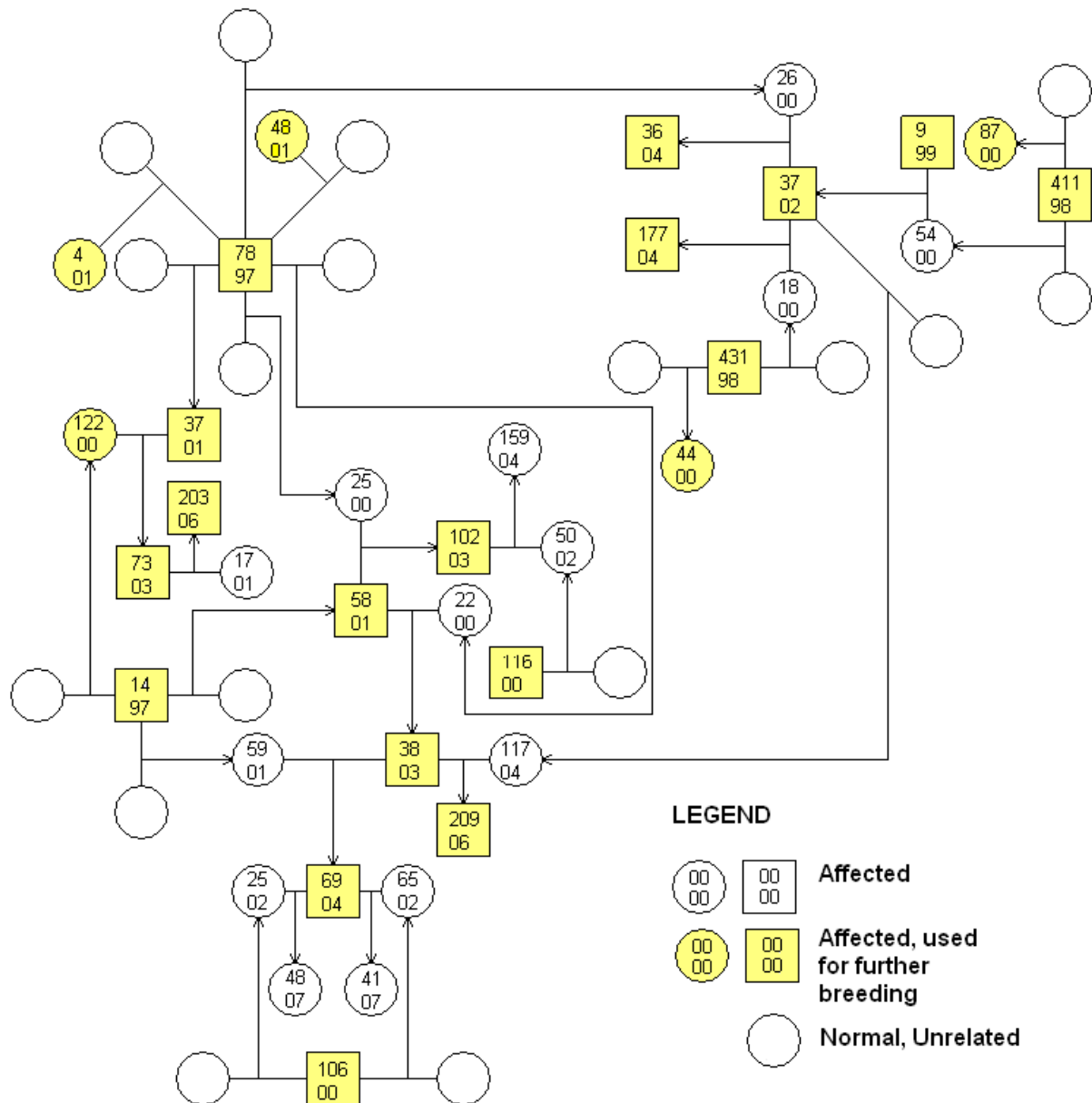


Figure 5: Pedigree diagram of the cataract flock. Squares are males and circles are females. This pedigree also suggests dominant inheritance, because affected offspring are produced by mating normal unrelated animals to affected animals. The first two digits give the birth order and the last two digits the year of birth. The sheep without known ancestors in the pedigree are 14/97, 78/97, 411/98, 431/98, 9/99, 106/00, and 116/00.

Six affected rams, 14/97, 411/98, 431/98, 9/99, 106/00, and 116/00, were bred with normal Coopworth ewes. The rams were descendants of “676” and “1836” from the Massey pedigree but the exact relationship to them was unknown. The breeding was continued from 2000 to 2008, involving both in-breeding and out-breeding to normal ewes. All matings were single-sire. This breeding did not significantly differ from normal farm practice and did not require

an ethics approval. The only parentage verification was a comparison of the marker genotypes of the sires and offspring described in section 3.2.3. The breeding records presented in this chapter have been corrected using this comparison.

The eyes of the animals were examined in darkened rooms to determine their cataract status, following established procedure (Robertson *et al*, 2005). Eye pupils were dilated with atropine (1% atropine sulfate, Sigma Pharmaceuticals, Australia) and the lenses were inspected with a slit lamp microscope (Kowa, SL/5, Japan) and an ophthalmoscope (Vista Diagnostic Instruments, Kellar, UK) at 6-8 weeks after birth by a trained veterinary ophthalmologist. Animals born in 2003 and 2004 were also examined 10-14 weeks after birth. Cataracts were assigned to the stages described in section 1.5.3, but for the linkage and mutation studies performed in this investigation, any opacity in the lens was classified as affected by OHC. OHC was likely to be inherited in a dominant manner (Figure 5).

No environmental differences are known between the Massey University and Lincoln University breeding areas. Mature congenital cataract does not significantly affect the reproduction of mature sheep or survival of lambs. No postmortems have been performed.

2.3.2 Breeding Records

Over eight years of breeding, 19 affected rams sired 1,034 lambs whose eyes were examined for cataracts, 260 from affected ewes and 774 from normal ewes. The affected-normal and affected-affected breeding records are analysed separately below. Standard errors are calculated as in Section 2.2.

Table 3: Cataract statuses of affected ram - normal ewe offspring. Each block shows the offspring of a single affected ram with normal ewes, with the bottom block showing the combined results for all of the rams shown above.

14/97	Normal	Affected	Total	Proportion Affected	Standard Error
Female	17	11	28	0.39	0.09
Male	16	11	27	0.41	0.09
Total	33	22	55	0.40	0.07
78/97	Normal	Affected	Total	Proportion Affected	Standard Error
Female	9	9	18	0.5	0.1
Male	8	4	12	0.3	0.1
Total	17	13	30	0.43	0.09
411/98	Normal	Affected	Total	Proportion Affected	Standard Error
Female	7	4	11	0.4	0.2
Male	7	6	13	0.5	0.1
Total	14	10	24	0.42	0.10
431/98	Normal	Affected	Total	Proportion Affected	Standard Error
Female	13	5	18	0.3	0.1
Male	8	6	14	0.4	0.1
Total	21	11	32	0.34	0.08
102/03	Normal	Affected	Total	Proportion Affected	Standard Error
Female	35	17	52	0.33	0.07
Male	24	29	53	0.55	0.07
Total	59	46	105	0.44	0.05
106/00	Normal	Affected	Total	Proportion Affected	Standard Error
Female	10	5	15	0.3	0.1
Male	5	8	13	0.6	0.1
Total	15	13	28	0.46	0.09
116/00	Normal	Affected	Total	Proportion Affected	Standard Error
Female	12	6	18	0.3	0.1
Male	11	4	15	0.3	0.1
Total	23	10	33	0.3	0.08
37/01	Normal	Affected	Total	Proportion Affected	Standard Error
Female	14	4	18	0.2	0.1
Male	7	3	10	0.3	0.1
Total	21	7	28	0.25	0.08
58/01	Normal	Affected	Total	Proportion Affected	Standard Error
Female	16	15	31	0.48	0.09
Male	19	12	31	0.39	0.09
Total	35	27	62	0.44	0.06
37/02	Normal	Affected	Total	Proportion Affected	Standard Error
Female	14	10	24	0.4	0.1
Male	11	7	18	0.4	0.1
Total	25	17	42	0.40	0.08

	Normal	Affected	Total	Proportion Affected	Standard Error
38/03					
Female	30	22	52	0.42	0.07
Male	36	34	70	0.49	0.06
Total	66	56	122	0.46	0.05
73/03					
Female	14	12	26	0.5	0.1
Male	13	9	22	0.4	0.1
Total	27	21	48	0.44	0.07
102/03					
Female	35	17	52	0.33	0.07
Male	24	29	53	0.55	0.07
Total	59	46	105	0.44	0.05
36/04					
Female	28	4	32	0.13	0.06
Male	32	3	35	0.09	0.05
Total	60	7	67	0.10	0.04
69/04					
Female	14	13	27	0.5	0.1
Male	12	14	26	0.5	0.1
Total	26	27	53	0.51	0.07
177/04					
Female	11	6	17	0.4	0.1
Male	7	11	18	0.6	0.1
Total	18	17	35	0.49	0.08
203/06					
Female	0	1	1	1	0
Male	2	1	3	0.3	0.3
Total	2	2	4	0.5	0.3
209/06					
Female	0	0	0	0.00	0.00
Male	2	4	6	0.7	0.2
Total	2	4	6	0.7	0.2
All					
Female	244	144	388	0.37	0.02
Male	220	166	386	0.43	0.03
Total	464	310	774	0.40	0.02

The highest proportion of affected offspring comes from ram 209/06, with 0.7 ± 0.2 (Table 3). This particular ram gave only 6 offspring which makes the proportion of affected offspring less reliable, as is shown by the standard error. The ram with the second highest proportion of affected offspring is 69/04, with 0.51 ± 0.07 from 53 offspring. Since a homozygote would give

all affected offspring at full penetrance, there is no evidence that any of the rams are homozygotes. They will all be assumed to be heterozygotes for the OHC gene in the following analysis.

The expected proportion of affected offspring from matings between affected and normal animals is 0.5. Some rams are at or above this figure but overall fewer affected offspring are produced than expected, with the proportion of affected offspring for all rams together being 0.40 ± 0.02 . If the lower proportion of affected offspring is due to incomplete penetrance, this is equivalent to a penetrance of 0.80 ± 0.04 .

Cataracts are more common in males than in females. The statistical significance of this difference is analysed below.

Table 4: Cataract statuses of affected ram – affected ewe offspring.

9/99	Normal	Affected	Total	Proportion Affected	Standard Error
Female	3	4	7	0.6	0.2
Male	1	4	5	0.8	0.2
Total	4	8	12	0.7	0.1
37/02	Normal	Affected	Total	Proportion Affected	Standard Error
Female	0	3	3	1	0
Male	1	6	7	0.9	0.1
Total	1	9	10	0.90	0.09
38/03	Normal	Affected	Total	Proportion Affected	Standard Error
Female	8	8	16	0.5	0.1
Male	6	11	17	0.7	0.1
Total	14	19	33	0.58	0.09
73/03	Normal	Affected	Total	Proportion Affected	Standard Error
Female	9	14	23	0.6	0.1
Male	5	9	14	0.6	0.1
Total	14	23	37	0.62	0.08
102/03	Normal	Affected	Total	Proportion Affected	Standard Error
Female	3	2	5	0.4	0.2
Male	4	4	8	0.5	0.2
Total	7	6	13	0.46	0.14
36/04	Normal	Affected	Total	Proportion Affected	Standard Error
Female	3	0	3	0.00	0.00
Male	1	2	3	0.7	0.3
Total	4	2	6	0.33	0.19

	Normal	Affected	Total	Proportion Affected	Standard Error
69/04					
Female	13	20	33	0.61	0.09
Male	14	22	36	0.61	0.08
Total	27	42	69	0.61	0.06
177/04					
Female	5	15	20	0.8	0.1
Male	10	12	22	0.6	0.1
Total	15	27	42	0.64	0.07
203/06					
Female	3	7	10	0.7	0.1
Male	2	4	6	0.7	0.2
Total	5	11	16	0.69	0.12
209/06					
Female	5	8	13	0.6	0.1
Male	4	5	9	0.6	0.2
Total	9	13	22	0.6	0.1
All					
Female	52	81	133	0.61	0.04
Male	48	79	127	0.62	0.04
Total	100	160	260	0.62	0.03

Ram 37/02 has the highest proportion of affected offspring, 0.90 ± 0.09 (Table 4). This may be consistent with being homozygous for the OHC gene, except that the proportion of affected offspring for this ram with normal ewes was 0.40 ± 0.08 .

The ram with offspring only from affected ewes, 9/99, gave 0.7 ± 0.1 cataract offspring. This is not consistent with the 100% affected offspring expected from homozygotes, so it is assumed that all of the rams are heterozygotes. The proportion of affected offspring from this ram is also less than the 0.75 expected from two heterozygotes.

It is possible that some of the affected ewes are homozygotes, but there have not been enough offspring from each ewe to determine their genotype. Assuming all rams are heterozygotes and all affected ewes are homozygotes, all of the offspring should have cataracts. If all the affected ewes are heterozygotes, then a cataract offspring proportion of 0.75 is expected. With full penetrance the proportion of affected offspring should be between these values, depending on the exact number of homozygous ewes. The proportion of affected offspring from all rams with affected ewes is 0.62 ± 0.03 , less than expected even if all affected ewes are

heterozygotes. If the lower proportion of affected offspring is due to incomplete penetrance, this is equivalent to a penetrance of 0.83 ± 0.04 .

2.3.3 Statistical Analysis

Statistical analysis was carried out on the combined data for all rams. This gave the largest number of offspring and the most accurate statistics. Offspring from normal ewes and from affected ewes was analysed separately. A chi-square goodness of fit test was used to test the pattern of affected and normal offspring against the pattern expected from various modes of inheritance. Another chi-square test was used to analyse the difference in cataract frequency between males and females. Although some affected animals are likely to be homozygotes, not enough breeding has been done to identify them, so all affected animals are assumed to be heterozygotes. All chi-square statistics were calculated using Microsoft Excel, and the p values were calculated from them using the online calculator at <http://www.stat.tamu.edu/~west/applets/chisqdemo.html>.

The affected-normal matings give 464 normal offspring and 310 (0.40 ± 0.02) affected offspring, when the expected numbers are 387 each (0.5) for dominant inheritance from one heterozygote and one normal homozygote. This gives a chi-square statistic of 31 for deviation from the expected distribution. With one degree of freedom this gives a p value of less than 0.0001, which is highly significant. Therefore OHC is not inherited in a simple dominant manner.

The affected-affected matings give 260 lambs, of which 160 (0.62 ± 0.03) are affected, when the numbers expected from dominant inheritance would be 195 (0.75) affected and 65 normal. These numbers give a chi-square statistic of 25 for deviation from the expected distribution. With one degree of freedom this also gives a p value of less than 0.0001. This confirms that OHC is not inherited in a simple dominant manner.

The lower percentage of affected offspring may be due to a lethal homozygote effect. If all homozygotes for the OHC allele died before eye testing at 6-8 weeks, the expected proportion of cataracts in the surviving lambs would be 0.67. This would equate to 173 affected and 87 normal offspring. When the observed data are compared to this distribution, a chi-square statistic of 2.9 is obtained for deviation from the expected distribution. With one degree of

freedom this gives a p value of 0.088. This is not significant, so a lethal homozygote effect cannot be ruled out.

For the affected-normal matings the proportion of affected animals is $0.37 \pm .02$ for females and $0.43 \pm .03$ for males. When the chi-square test is performed on the numbers of affected and normal males and females the chi-square statistic is 2.8. With one degree of freedom this gives a p value of 0.094. For the affected-affected matings the affected proportions are $0.61 \pm .04$ for females and $0.62 \pm .04$ for males. When the chi-square test is performed on the raw numbers of affected and normal males and females the chi-square statistic is 0.047. With one degree of freedom this gives a p value of 0.83.

When the numbers of affected and normal males and females are combined from both types of mating, the affected fractions are $0.43 \pm .02$ for females and $0.47 \pm .02$ for males. When the chi-square test is performed on the raw numbers of affected and normal males and females the chi-square statistic is 2.2. With one degree of freedom this gives a p value of 0.1398. None of these results are significant, so there is no evidence of significant difference in the proportion of affected animals between males and females.

2.3.4 Litter Size

The litter size in affected-affected matings will be reduced by any lethal homozygote effect that kills individuals before birth. All other factors being equal, such an effect would kill 25% of the offspring of affected-affected matings before birth while leaving all the offspring of affected-normal matings unaffected. Therefore the mean litter size in affected-affected matings would be predicted to be 75% of mean litter size in affected-normal matings.

The Lincoln University breeding produced 608 litters, 423 with affected-normal and 185 with affected-affected parents. These figures exclude lambs born in 2003, for which the birth records are unreliable. The numbers of each size of litter from 1 to 4 lambs are shown in Table 5.

Table 5: Litter sizes from different breeding types.

	Affected-normal	Affected-affected
One lamb	167	79
Two lambs	224	92
Three lambs	29	14
Four lambs	3	0
Total	423	185

The mean litter sizes are 1.69 for affected-normal and 1.65 for affected-affected matings. An unpaired T-test was carried out on the litter sizes using Microsoft Excel, giving a t statistic of 0.712. With 606 degrees of freedom this gives a p value of 0.476, which is not significant. Given the results of this t-test there is no evidence from litter size of a lethal homozygote effect.

2.3.5 Death Rates

No evidence was found of a lethal homozygote effect that killed animals before birth, so death rates were analysed to find whether the deaths were occurring after birth. As described above, with a lethal homozygote effect 25% of the offspring of affected-affected matings should die and none of the affected-normal offspring. A total of 1,019 animals were born between 2000 and 2008, 305 with two affected parents and 714 with one affected and one normal. As for the litter size analysis, lambs born in 2003 are omitted. This includes animals that were not included in the analysis of OHC inheritance above, because their cataract statuses were not recorded.

From the affected-affected matings, 31 out of 305 animals died before their cataract scores were recorded, a death rate of 0.10 ± 0.02 . For the affected-normal matings, 37 out of 714 animals died, a death rate of 0.052 ± 0.008 . A chi-square independence test comparing the two death rates from an online calculator (Preacher, 2001) gives a statistic of 8.517. With one degree of freedom this gives a p value of 0.0035, which is significant. Therefore the death rate in the affected-affected matings is significantly higher than in the affected-normal matings. The excess death rate is 0.048, which is much less than than 0.25 expected from a lethal homozygote effect.

2.4 Discussion

The breeding carried out at Massey University suggested a dominant mode of inheritance and showed no evidence for incomplete penetrance. The larger number of offspring from the Lincoln University breeding provides a more reliable test for incomplete penetrance.

For the Lincoln breeding, both the affected-normal and the affected-affected matings give significantly fewer affected offspring than expected from dominant inheritance. However, the mode of inheritance is unlikely to be recessive. This is because cataracts are rare in sheep and the unrelated normal ewes mated with affected rams are unlikely to be carriers for a recessive cataract, but these matings still produce affected offspring.

The proportion of affected animals is not significantly different between males and females. Thus there was no evidence for an X or Y-chromosome linked gene and the OHC gene is probably autosomal.

Incomplete dominance in OHC would give less opacity or slower development of opacity in heterozygotes compared to homozygotes. Since any degree of opacity was counted as affected, the former situation would not affect this analysis. In addition, affected-affected and affected-normal matings give similar deviations from the expected number of affected offspring. Homozygotes should only be produced from affected-affected matings, and in incomplete dominance this would make the results differ between the two mating types. Therefore OHC is probably not inherited in an incomplete dominant manner.

The proportion of affected offspring produced from the affected-affected matings at Lincoln is consistent with a lethal homozygote. This would also explain the absence of known homozygotes from several generations of inbreeding. However, a lethal homozygote would not explain the lower number of affected offspring from affected-normal matings.

Analysis of litter size shows no evidence of a lethal homozygote effect that kills animals before birth. However, analysis of death rates showed that significantly more offspring with two affected parents died before their eyes were examined than those with only one affected parent. A lethal homozygote is expected to kill 25% of the affected-affected offspring and

none of the affected-normal offspring, but the excess death rate is only 4.8% in this case. Therefore the increased death rate may be due to inbreeding depression. This is the situation where inbreeding results in increased homozygosity for recessive alleles throughout the genome that reduce the fitness of the animal (Sittmann *et al*, 1966). No inbreeding coefficients have been calculated for the pedigree.

Incomplete penetrance would explain both deviations from simple dominant inheritance. For the affected-normal matings a penetrance of 0.80 ± 0.04 would explain the number of affected offspring, whereas for the affected-affected matings the appropriate penetrance would be 0.83 ± 0.04 . The similarity of the two penetrance figures suggests that this explanation is correct, and that the penetrance (from a weighted average using the number of offspring from each mating type) is 0.81 ± 0.04 .

Since most animals' eyes were only examined 6-8 weeks after birth, any effect that delayed the development of opacities past that period would lead to a large number of affected animals being classified as normal. This would give the same results as incomplete penetrance. A study of the same flock of animals at the same facility examined the development of cataract over 10 months (Robertson *et al*, 2005). Of the 162 animals without opacities at 6 weeks after birth, only 9 or 6% developed cataracts later. This is not sufficient to account for the difference from the expected results from simple dominant inheritance, even if all animals were examined at 6 weeks after birth. In practice most animals were examined later than this and more would have developed opacities.

If any effect delayed the development of opacities past 10 months, affected animals would have been scored as normal by both this investigation and the previous study described above. However since OHC is a developmental cataract which sometimes occurs congenitally, it is very unlikely that the development of opacities would be delayed for so long. Any opacities that occurred after 10 months would be more likely to be the symptom of a different cataract condition.

If the gene had the same penetrance in the Massey breeding, it would produce affected offspring proportions of 0.41 for affected-normal matings and 0.62 for affected-affected matings. These figures are also within the standard errors of the observed numbers.

However since the Lincoln University breeding introduced new Coopworth genetics into the OHC flock, the penetrance is not necessarily the same as at Massey University.

A human congenital cataract has been reported to be inherited in an autosomal dominant manner with incomplete penetrance (Burdon *et al*, 2004). The gene responsible is Connexin 46/GJA3, which is found on human chromosome 13. The genome scan described in Chapter 3 showed that the cataract locus is on sheep chromosome 6, the homologous chromosome to human chromosome 4. Therefore GJA3 is unlikely to be the OHC gene.

The location of any genes which contribute to incomplete penetrance is unknown, but because of the similarity of penetrance between males and females, it is unlikely that X or Y linked genes are involved. Environmental factors could also have contributed to incomplete penetrance.

Software such as Statistical Analysis for Genetic Epidemiology (SAGE) has been used to carry out complex segregation analysis on pedigrees (Tayo *et al*, 2009). This involves constructing mathematical models of several different modes of inheritance and using the software to test each model against the pedigree. This software was not available in this investigation.

The OHC gene is expected to have two alleles, which can be represented with “C” for the dominant cataract allele and “c” for the recessive wild-type allele. In several years of breeding, no CC homozygotes have been observed. There have been 260 offspring produced at Lincoln University that have had their cataract status recorded and have two affected parents. If all the parents are assumed to be heterozygous, 65 CC homozygotes would be expected. The incomplete penetrance calculated above means that 10-15 of these would be expected not to have cataracts. The actual number of CC homozygotes would be expected to be higher because some of the parents would be CC homozygotes themselves and would be more likely to produce homozygous offspring.

CC homozygotes can only be detected through breeding if they have produced enough offspring for the proportion of affected offspring to be estimated. With full penetrance all of a CC homozygote’s offspring should have cataracts, regardless of the genotype of the other parent. With the penetrance calculated above, 77-85% of the offspring are expected to have

cataracts. Two heterozygous animals should have 58-64% cataract offspring, and a heterozygous animal mated with a cc homozygote should have 39-43% cataract offspring.

After the offspring of two affected animals were born, their eyes were inspected and they were classified as affected or normal. Normal animals were not used in further breeding and were eventually culled. As described above this would be expected to eliminate 10-15 of the 65 CC homozygotes, leaving 50-55. Of the 260 offspring, 133 were female, so 26-28 of the affected CC homozygotes would be expected to be female. Cataract ewes were used in breeding but produced much fewer offspring than rams. In the Lincoln University breeding no individual cataract ewe produced enough offspring to determine whether it was a CC homozygote. Therefore only the 24-27 affected CC homozygote rams could potentially be distinguished from heterozygotes. From the 127 affected rams with two affected parents, 19 were used in further breeding. The expected number of CC homozygotes in this group is 4, and these were expected to be the only CC homozygotes with enough offspring to be distinguished from heterozygotes.

In order to distinguish between heterozygous and homozygous affected rams, the difference in the proportion of affected offspring must be detectable. This is between 0.64 and 0.77 with affected ewes and between 0.43 and 0.77 with normal ewes. The proportion of affected offspring from each of the 19 rams is presented in Table 3 and Table 4, along with the standard error in the proportion.

For 15 of the 19 rams, all of the breeding results show a proportion of affected offspring more than 1 standard error less than that expected from a CC homozygote. These rams are probably heterozygotes. For one ram (9/99) the proportion of affected offspring is within one standard error of that expected from a CC homozygote, and also within one standard error of that expected from a heterozygote. This ram may be a CC homozygote, but only produced 12 recorded offspring before it died. It provided one of the samples for the genome scan, so its marker genotypes may be analysed further to determine its OHC genotype. For three rams (37/02, 203/06, and 209/06) the results varied between normal ewes and affected ewes – one set of offspring was not consistent with a CC homozygote, and one set was. These three rams are probably heterozygotes, but were also part of the genome scan and can have their marker genotypes studied to determine the OHC genotypes. If all four animals are CC homozygotes this will be consistent with the expected numbers.

The breeding program carried out on the OHC flock while this investigation was undertaken was designed to produce large numbers of lambs for testing a cataract prevention drug as well as providing samples for the linkage studies and candidate gene testing described in this thesis. Another goal of the breeding program was to produce a homozygous affected animal. This meant that the breeding program was not ideal for linkage analysis of OHC. For example, none of the normal ewes used in breeding were genotyped or had their eyes examined for cataracts, which makes the linkage analysis less efficient.

If a breeding program was designed for efficient linkage analysis, ewes unrelated to the OHC flock would have their eyes examined for cataracts, and the normal ewes would be genotyped for a pre-selected set of markers before being mated to affected rams, which would also be genotyped. Genomic DNA samples from all animals would be stored so that additional markers could be genotyped. All matings would be carefully controlled and all ewes kept isolated from rams in order to avoid parentage errors. The genotypes of the offspring would be used for parentage verification before linkage analysis. The offspring would have their eyes examined at different ages to detect opacities that develop later than normal. Affected rams would be mated with genotyped, unrelated normal ewes as for the previous generation, and the process would be repeated for several generations.

This breeding program maximises overall heterozygosity and therefore informative meioses, and makes linkage analysis more efficient. However it is very unlikely to produce a homozygous affected animal as the ewes are selected to be wild-type homozygotes. Since homozygous affected animals would be useful in drug trials, the breeding program could be extended to produce them. Affected ewes and rams would be selected that were homozygous for marker alleles linked to OHC. This would indicate a greater probability of being homozygous for the OHC gene. Matings between these animals would be more likely to produce homozygotes. The offspring of these matings would be genotyped for linked markers and mated with normal animals to detect whether they were homozygous. This would be easier with rams because they can produce more offspring.

2.5 Conclusion

Ovine heritable cataract is inherited in an autosomal dominant manner with incomplete penetrance. The penetrance is 0.81 ± 0.04 .

Chapter 3 Genome Scanning

3.1 Introduction

In this chapter a partial genome scan for linkage to OHC was carried out on sheep chromosomes selected by homology with human and mice chromosomes associated with cataract. New sheep microsatellite markers were generated from bovine genome assemblies and used for fine mapping of the OHC locus. A short region of the sheep genome linked to OHC was determined.

3.2 Whole Population Scan

3.2.1 Introduction

Twelve genes or chromosome regions known to be associated with cataracts in humans or mice were investigated. The mouse or human chromosomes that they were located on were determined, and the homologous sheep chromosomes were determined using Oxford grids. Sixty-seven microsatellites known to lie on the 10 homologous chromosomes (1, 2, 3, 8, 10, 11, 12, 14, 17, and 26) were genotyped in related affected and normal animals. This is similar to the approach taken in a study of malaria resistance in mice (Hernandez-Valladares *et al*, 2004). The human chromosome region 5q31-q33 is known to be linked to malaria resistance and the homologous mouse chromosomes 11 and 18 were used in a linkage study to find quantitative trait loci.

In this OHC study, as in the malaria resistance study, each entire chromosome was the target of a partial genome scan and no attempt was made to determine the homologous region of the target chromosome. This was done to allow for uncertainties in the homologies; the homologous chromosome is more likely to be correct than the homologous region. This contrasts with the approach taken in a study of familial frontal epilepsy (Lopes-Cendes *et al*, 1995), where the five human chromosome regions homologous to a mouse candidate gene were used in a linkage study. That study found no linkage to the condition.

No significant linkage to cataract was found on the first set of chromosomes, so 61 markers on the next 10 largest chromosomes were also analysed. From these chromosomes, markers on chromosome 6 showed significant linkage to OHC and no markers on the other chromosomes showed significant linkage.

More microsatellites were genotyped on this chromosome to restrict the position of the OHC locus. In this section all DNA extractions and CRIMAP linkage analyses were carried out by the author, and all PCR reactions, chromatography, and multipoint linkage analyses were carried out by AgResearch Invermay. The reason for outsourcing the PCR and chromatography is that AgResearch had equipment not available at Lincoln University which was able to generate large numbers of genotypes quickly. AgResearch also had experience in multipoint analysis using the MultExclude program which Lincoln University did not share, so this was also outsourced.

3.2.2 Methods

3.2.2.1 Chromosome Selection

Genes were used to select chromosomes for testing if they caused autosomal dominant cataracts. All the genes selected caused cataracts in either humans or mice. The characteristics of OHC were not taken into account when selecting the candidate genes, because the same gene often causes cataracts with different characteristics in different species. One example is the gamma-D crystallin gene, CRYGD. A mutation in the mouse gene causes nuclear cataracts with some opacity of the cortex (Graw *et al*, 2002), whereas a mutation in the human homologue causes a ring of small opacities around the anterior pole of the lens (Roshan *et al*, 2010)

The calcium-calpain-proteolysis model for OHC described in Section 1.5.4 is not consistent with some of the genes, but they were selected anyway as the model has not been confirmed yet. The homologous sheep chromosome for each gene was found using online Oxford grids at http://www.informatics.jax.org/searches/oxfordgrid_form.shtml. An Oxford grid is a method of displaying the location of homologous loci between different species (Edwards, J.H., 1991) (Figure 6).

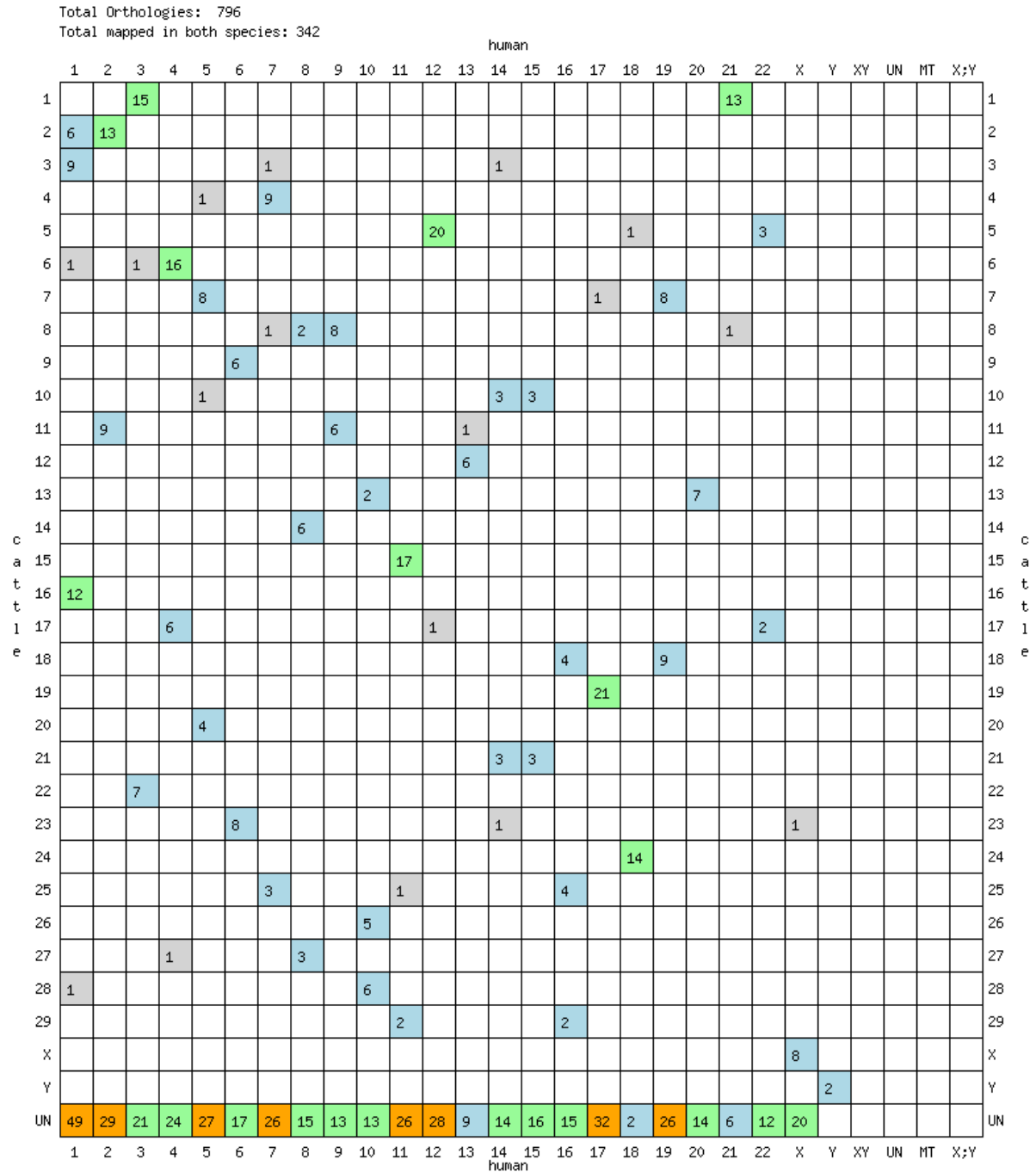


Figure 6: Human-cattle Oxford grid. Human chromosomes run from left to right and cattle chromosomes run from top to bottom. The grid squares contain numbers showing how many loci are in both chromosomes. The squares are colour-coded by the number of loci. For example, the green square at the intersection of cattle chromosome 6 and human chromosome 4 shows 16 loci which are in both chromosomes. On the original web page, http://www.informatics.jax.org/searches/oxfordgrid_form.shtml, each filled square is a link to a list of the loci. Human-cattle, cattle-sheep, mouse-sheep, and sheep-human Oxford grids were all used in this section.

3.2.2.2 Pedigree Selection

Cataract statuses of the sheep were determined as described in section 2.3.1. A ram (ID 38/03) was selected that was affected and had sired 41 lambs with normal, unrelated ewes (Table 6).

Table 6: Pedigree of 41 lambs used for linkage analysis. Ram 38/03 was the sire of all the lambs. Where a dam is not shown it is the same as the animal above it. Affected animals are highlighted in yellow. All dams are unrelated Coopworth animals.

Sire	Dam	Offspring	Sire	Dam	Offspring
38/03	Y484	1/04	38/03	Y460	98/04
	Y550	2/04			99/04
	R278	6/04		R330	101/04
		7/04		Y532	105/04
	Y515	11/04			106/04
	Y173	15/04		W212	114/04
	Y223	23/04			115/04
		24/04		W131	121/04
	R335	51/04			122/04
	R112	53/04		Y521	123/04
	Y551	61/04			124/04
	W43	62/04		R13	129/04
		63/04			131/04
	W102	64/04		W12	132/04
		65/04			133/04
	R347	78/04		R187	134/04
	R105	82/04		Y513	161/04
		83/04		Y225	170/04
	W175	88/04			171/04
		89/04			
	W211	96/04			
		97/04			

Nineteen of the lambs were affected and 22 remained unaffected. When the linked chromosome was identified, samples from three more rams and their 341 offspring with normal ewes were also genotyped. In addition, samples were extracted and genotyped from six affected ewes and nine of their offspring with the four rams. The total number of sheep genotyped was 356. As described below in section 3.2.3, Linkage Analysis, 67 animals had genotypes incompatible with any of the sires and were excluded from the analysis, leaving 289 animals (Table 7). The parentage of the remaining animals in Table 7 has been corrected.

Table 7: Pedigree of 293 lambs used for linkage analysis. This includes 289 animals that contributed DNA samples and four ancestors that did not. The latter are shown with black boxes. Animals are arranged by sire and then by dam, with each sire's offspring shown below him. The order of animals runs down each column on a page, then across to the next column, and then on to the next page. Where the sire or dam field is blank it is identical to the animal above it. Affected animals are highlighted in yellow.

Sire	Dam	ID	Sire	Dam	ID	Sire	Dam	ID
Unknown	Unknown	186/96	38/03	r335	51/04	38/03	y152	65/05
186/96	CPW1	14/97		r347	78/04			66/05
14/97	CPW2	58/01		w1015	63/05		y173	15/04
58/01	22/00	38/03			64/05		y223	23/04
38/03	9999	136/05		w102	64/04			24/04
	b1128	127/05			65/04		y225	170/04
		128/05		w12	132/04			171/04
		129/05			133/04		26/00	154/05
	b1132	9/05		w131	121/04			155/05
		10/05			122/04		y444	51/05
	b1136	55/05		w132	134/05			52/05
		56/05		w17	103/05		y460	98/04
	b1147	269/05			104/05			99/04
	b1157	18/05		w175	88/04		y469	49/05
	b27	89/05			89/04			50/05
	b777	214/05		w20	125/05		y484	1/04
	g1159	5/05			126/05		y494	193/04
		6/05		w211	91/05			194/04
	g1164	230/05			92/05		y510	11/05
		231/05			96/04		y513	161/04
		232/05			97/04		y515	11/04
	g1172	3/05		w212	114/04		y521	123/04
		4/05			115/04			124/04
	g1187	22/05		w43	62/04			233/05
	g642	144/05			63/04			234/05
		151/05		w61	77/05		y526	107/05
	g650	289/05			78/05			108/05
		290/05			79/05		y532	105/04
		292/05		y1002	45/05			106/04
	r10	186/05		y1006	237/05		y541	351/05
		187/05		y1011	206/05			352/05
	r105	2/05			207/05		y547	67/05
		82/04		y1017	223/05			68/05
		83/04			224/05		y550	2/04
	r112	53/04		y1020	145/05		y551	61/04
	r13	129/04			146/05		y87	13/05
		131/04		y1024	220/05	58/01	25/00	102/03
	r1728	23/05		y1029	212/05	102/03	b1112	255/05
		24/05		y1033	203/05			259/05
		26/05		y1036	242/05		b1133	95/05
	r187	134/04		y1044	19/05			96/05
	r278	6/04			20/05		b1137	273/05
		7/04		y1062	27/05			275/05
	r330	101/04		y1064	53/05		b1159	93/05
		48/05			54/05			94/05

Sire	Dam	ID	Sire	Dam	ID	Sire	Dam	ID
102/03	b1168	256/05	102/03	y1025	29/05	78/97	y744	26/00
	b1172	270/05		y1037	69/05	Unknown	Unknown	9/99
	b353	88/05			70/05	9/99	54/00	37/02
	g1153	179/05		y1038	138/05	37/02	26/00	36/04
	g1173	33/05		y105	274/05	36/04	b1108	201/05
		34/05		y1053	296/05		b1145	169/05
	g1188	139/05		y1063	297/05			170/05
		140/05		y1065	354/05		b1155	84/05
	g621	192/05		y113	100/04		b1155	85/05
	g652	46/05		y122	3/04		g1157	160/05
		47/05		y152	181/04		g1186	229/05
	g664	247/05			182/04		r124	16/05
	NT	135/04		y225	238/05		r278	263/05
		136/04			240/05		r3	86/05
	r15	42/05		y470	194/05		r3	87/05
	r177	154/04			195/05		w1013	121/05
	r191	164/04		y482	58/04		w1065	246/05
		165/04			59/04		w175	35/05
	r199	76/04		y507	82/05			36/05
	r206	279/05			83/05		y1023	123/05
	r208	80/04		y510	185/04		y484	61/05
	r221	162/04			186/04			62/05
		163/04		y512	155/04	37/02	b27	25/04
		41/05			156/04		p51	74/04
	r268	109/04		y524	94/04			75/04
	r332	142/04			95/04		r10	157/04
		143/04		y526	150/04		r124	43/04
	r335	30/05			151/04		r204	126/04
	w1029	281/05		y528	355/05			127/04
	w11	105/05		y532	80/05			128/04
	w136	7/05		y532	81/05		r206	176/04
	w144	45/04		y543	243/05		r3	144/04
	w185	110/04			244/05		r331	112/04
		111/04		y548	120/04			113/04
	w212	75/05			184/05		w11	137/04
		76/05			185/05		w132	166/04
	w27	22/04		y549	91/04			167/04
	w36	109/05		y551	360/05		w136	20/04
	w36	110/05		y552	252/05		w36	187/04
	W48	284/05			261/05		w37	38/04
	w73	135/05	14/97	y744	W17			39/04
	y1000	235/05	Unknown	Unknown	188/96		w73	158/04
		236/05	188/96	CPW3	78/97		y205	140/04
	y1013	137/05	78/97	CPW6	22/00			141/04
	y1018	31/05		O487	W4		y444	118/04
		32/05	78/97	w497	W48		y444	119/04
	y1022	14/05		y744	25/00		y462	90/04

Sire	Dam	ID
37/02	y471	116/04
		117/04
	y477	40/04
	y495	84/04
		85/04
	y504	13/04
	y530	92/04
		93/04
	y534	8/04
	y537	86/04
	y539	145/04
		146/04
Unknown	Unknown	411/98
411/98	CPW7	54/00
	O51	Y87

3.2.2.3 Genotyping

Eighty-eight markers on the selected chromosomes (1, 2, 3, 8, 10, 11, 12, 14, 17, and 26) were tested in the cataract ram's sample. The 67 markers that were heterozygous in the ram were tested on his offspring's samples. Sixty-one markers on the ten largest remaining chromosomes (4, 5, 6, 7, 9, 13, 15, 16, 18, and 19) were tested on the ram and the offspring, as above. All 61 were heterozygous in the ram. A total of 149 markers were tested, 128 of which were heterozygous in the ram and used to genotype his offspring. The average separation of the 128 markers on the 20 chromosomes was 20 cM.

Once chromosome 6 was identified as the location of the OHC locus, the markers on chromosome 6 were tested on the expanded pedigree described above. Four new markers on chromosome 6 were also tested to increase the density of marker coverage.

Venopuncture blood samples of 5-8mL were taken from the animals and stored at 4°C. The blood sampling was approved by the Lincoln University Animal Ethics Committee on December 4, 2007, Application Number 221.

Vacurette brand venopuncture tubes containing K₃EDTA (Greiner Bio-one, Bad Haller Straße 32, A-4550 Kremsmünster, Austria) were used to collect and store the blood. One mL of each sample was mixed with 45 mL lysis buffer (0.32M sucrose, 10mM

tris(hydroxymethyl)aminomethane (Tris), 5mM MgCl₂, 10.7 g/L Triton X-100 (4-octylphenolpolyethoxylate, Dow Chemical, Midland, Michigan, USA)) incubated at room temperature for 3 min and centrifuged at 1500g for 5 min. The pellet was mixed with phosphate-buffered saline (0.137M NaCl, 2.68mM KCl₂, 10mM Na₂HPO₄, 1.76mM KHPO₄) and centrifuged again, then mixed with guanidine denaturing solution (4M guanidine thiocyanate, 0.25M sodium acetate, 0.119M β-mercaptoethanol) and shaken overnight at room temperature. DNA was precipitated with 5mL isopropanol, washed twice with ethanol, and dissolved in pH 8 TE. This method has been previously described (Ciulla *et al*, 1988). Seven ancestors of the affected sheep were genotyped from archived DNA samples extracted by the same method.

All genotyping PCR reactions were carried out as described in Schuelke, 2000, in order to produce PCR products with an attached fluorescent label.

Each PCR reaction mixture consisted of a 1:10 dilution of venopuncture DNA extract, 200μM dNTP mix (Invitrogen), 0.4μM reverse PT primer (Integrated DNA Technologies, 1710 Commercial Park, Coralville, Iowa, USA), 0.1μM forward M-13 primer (Integrated DNA Technologies), 1.5mM MgCl₂ (Invitrogen), 0.4μM M-13 colour, and 0.25 units per μL of Platinum *Taq* DNA polymerase (Invitrogen). The final volume of each mixture was 10 μL. The magnesium concentration was adjusted to 2mM or 3mM if necessary for each marker.

The PCR cycle involved an initial denaturation, 94°C 5 min, followed by 5 cycles of 94°C 30s, 60°C 30 s, and 72°C 40 s. The 5 cycles were repeated with an annealing temperature of 58°C, 56°C, 54°C (twice), 50°C, and 52°C for 7 cycles, for a total of 37 cycles. There was a final elongation step, 72°C 45 min.

PCR products were analysed by capillary chromatography using an ABI3730 DNA analyser with fluorescence detection, and the results were used to assign genotypes using Applied Biosystems GENEMAPPER software, version 3.7. The genotypes of offspring were compared to those of the sires to correct any mistakes in parentage.

3.2.2.4 Linkage Analysis

The genotypes of the first ram and his offspring were combined into a pedigree file for the CRIMAP genetic linkage program (Green *et al*, 1990). The data was arranged as one family

with one sire, 26 dams, and 41 offspring. The dams for all the lambs were included, scored as wild-type (homozygous normal) for the OHC locus and unknown for all the markers. The ram was treated as heterozygous for the OHC locus, consistent with the frequency of cataracts in his offspring.

The pedigree file was used with the PREPARE option of CRIMAP to prepare the auxiliary data files (.dat, .par, and .loc). All CRIMAP input files are shown in appendix B. These files were analysed using the TWOPOINT option of CRIMAP, to find significant linkage between markers and the OHC locus. The ALL option of CRIMAP was used to find the most probable position of the OHC locus on the chromosome with significant linkage to OHC.

For the expanded pedigree, the data were structured as a single family with four related rams and six related affected ewes, plus unrelated normal ewes and 279 offspring. Ancestors of the affected sheep were included in the pedigree. When DNA samples were not available for the ancestors they were scored as unknown for all markers. All affected animals were scored as heterozygous for OHC, since there is no evidence from breeding of a homozygote. The CRIMAP input file is shown in Appendix B.

3.2.2.5 Multipoint Analysis

A multipoint linkage analysis was also carried out on the data from all the genotypes, using MultExclude, a program developed at AgResearch Invermay by Dr Ken Dodds for SAS software based on algorithms developed for the construction of multilocus genetic linkage maps in humans (Lander and Green, 1987). The 1-LOD support interval and the most likely position of the OHC locus were determined. The multipoint linkage analysis was carried out by Dr Ken Dodds.

3.2.3 Results

3.2.3.1 Chromosome Selection

Twelve genes or chromosome regions associated with cataract in humans or rodents were found by studying the literature (Table 8).

Table 8: Human and mouse cataract genes and chromosome regions. “Gene” shows either the gene associated with cataract or the type of cataract that has a chromosome region associated with it. “Human” shows the human chromosome region. “Mouse” shows the mouse chromosome. “Sheep” shows the predicted sheep chromosome.

Gene	Human	Mouse	Sheep	Comment	Reference
BFSP2	3q21-22		1	Membrane protein	Conley <i>et al</i> , 2000
GJA8	1q21.1		1	Connexin	Shiels <i>et al</i> , 1998
CRYGD	2q33-35		2	Crystallin	Rogaev <i>et al</i> ,
MIP	12q14		3	Membrane protein	Berry <i>et al</i> , 2000
Age-related cortical	6p12-q12		8	Unknown gene	Iyengar <i>et al</i> ,
GJA3	13q11-12		10	Connexin	Bennett <i>et al</i> ,
CRYBA1	17q11.2-12		11	Crystallin	Padma <i>et al</i> , 1995
Posterior polar	1p terminus		12	Unknown gene	Ionides <i>et al</i> ,
MAF	16q23.2	8	14	Crystallin	Jamieson 2002
HSF4	16q21		14	Heat Shock Protein	Bu <i>et al</i> , 2002
CRYBB2	22q11.23	5	17	Crystallin	Litt <i>et al</i> , 1997
Polb		8	26	Transgenic over-expression	Sobol <i>et al</i> , 2003

The sheep chromosomes shown in Table 8 were determined by correspondences between sheep, cattle, and human or mice chromosomes, as described in Section 3.2.2.

3.2.3.2 Linkage Analysis

Genotypes of 42 affected and normal animals were generated for 67 markers on the 10 chromosomes listed in Table 8. Analysis of the genotypes did not find any parentage errors. The CRIMAP program was used to find linkage between the markers and the OHC locus.

Analysis of these ten chromosomes did not find any significant linkages between the OHC locus and any markers, so the next ten chromosomes were selected from the remaining 16 on the basis of size, largest to smallest. They were chromosomes 4, 5, 6, 7, 9, 13, 15, 16, 18, and 19. Sixty-one markers were tested on them. Of the seven markers tested on chromosome 6, three showed significant linkage to OHC, whereas the markers on the other chromosomes showed no significant linkage (Table 9 and Table 10). Marker LS55 on chromosome 7 shows

a LOD score of 1.12, which means that the set of genotypes is 13 times more likely with linkage than without linkage, which is not significant.

Table 9: Highest LOD scores for each chromosome from a CRIMAP twopoint analysis with the OHC locus. The chromosome and marker with the highest LOD score is highlighted. RF shows the recombination fraction for each non-zero LOD score.

	Chromosome	Marker with highest LOD	LOD	RF
Initial 10	1	CSSM04	0.04	0.41
	2	All	0.00	
	3	All	0.00	
	8	All	0.00	
	10	MCM470	0.02	0.42
	11	CSSM08	0.07	0.40
	12	All	0.00	
	14	All	0.00	
	17	All	0.00	
	26	All	0.00	
Second 10	4	HH64	0.21	0.36
	5	All	0.00	
	6	JMP8	4.24	0.1
	7	LS55	1.12	0.26
	9	MAF33	0.04	0.41
	13	All	0.00	
	15	MCMA16	0.14	0.36
	16	MCM506A	0.25	0.35
	18	OY5	0.15	0.38
	19	HMH1R	0.28	0.36

Table 10: Chromosome 6 LOD scores from a CRIMAP twopoint analysis with the OHC locus. Scores from all markers tested on chromosome 6 are shown, based on genotypes from one ram and 41 of his offspring. The most probable position of the OHC locus, from the output of the ALL option of CRIMAP, is shown in bold. RF shows the recombination fraction for each LOD score. All markers positions are from independent linkage mapping supplied by AgResearch Invermay.

Marker	Position cM	Lod Score	RF
MCM53	29.7	0.00	0.50
JMP36	62.1	0.28	0.37
BM4621	88	1.8	0.26
JMP4	112.3	1.7	0.24
OHC			
JMP8	135.5	4.24	0.1
JMP12	152.2	3.77	0.15
BL1038	152.2	3.05	0.18

The ALL option of CRIMAP was used with the genotypes of the markers on chromosome 6 to find the most likely position for the OHC locus. This was determined to be between JMP4

and JMP8, equivalent to between 112.3 and 135.5 cM from the centromere of chromosome 6. This is a region spanning about 23.2 cM.

In order to fine map the region linked to OHC, four additional markers on chromosome 6 were selected and more animals were genotyped. All 11 markers were tested on 356 animals from the cataract flock. These animals included the 42 animals used in the previous analysis.

Analysis of the genotypes of sires and offspring found 89 animals with genotypes inconsistent with their sires. Of these, 22 could be assigned to one of the other three sires. The remaining 67 had genotypes incompatible with any of the four samples and were omitted from the linkage analysis, leaving 289 animals (Table 11).

Table 11. Offspring of each sire omitted from the analysis or reassigned to another sire.

Sire	37/02	38/03	102/03	36/04	Total
Omitted	2	5	7	53	67
Reassigned	1	7	6	8	22

Strong linkage was detected between six of the markers and the OHC locus (Table 12). The most likely position for the OHC locus was between markers JMP4 and DK1183A. This is equivalent to between 112.3 and 132.9 cM from the centromere of sheep chromosome 6, a distance of 20.6 cM.

Table 12: LOD scores from 289 sheep, from a CRIMAP twopoint analysis with the OHC locus. Scores from all 11 markers on chromosome 6 are shown, and the most likely position of the OHC locus from the ALL option of CRIMAP is shown in bold. All marker positions are from the sex averaged linkage map version 4.7 at <http://rubens.its.unimelb.edu.au/~jillm/jill.htm>. JMP12 is omitted because it is at the same position as BL1038.

Marker	Position cM	Lod Score
CP125	2.6	0.00
MCM53	29.7	0.43
AE101	49.9	0.00
JMP36	62.2	0.13
JMP1	76.1	1.15
BM4621	88.0	10.21
EL13	100.1	11.56
JMP4	112.3	14.13
OHC		
DK1183A	132.9	38.71
JMP8	134.7	24.71
BL1038	151.2	7.35

3.2.3.3 Multipoint Analysis

A multipoint linkage analysis carried out on the chromosome 6 marker genotypes by Dr Ken Dodds, using the MultiExclude program, showed that the most likely position for the OHC locus is 126.4 cM from the centromere of sheep chromosome 6 (Figure 7). This is consistent with the CRIMAP results. The 1-LOD support interval is between 123.3 and 128.8 cM, a distance of 5.5cM. Over this interval the LOD score for the OHC locus is within 1 of the maximum value. The interval is approximately equivalent to a 90% confidence interval.

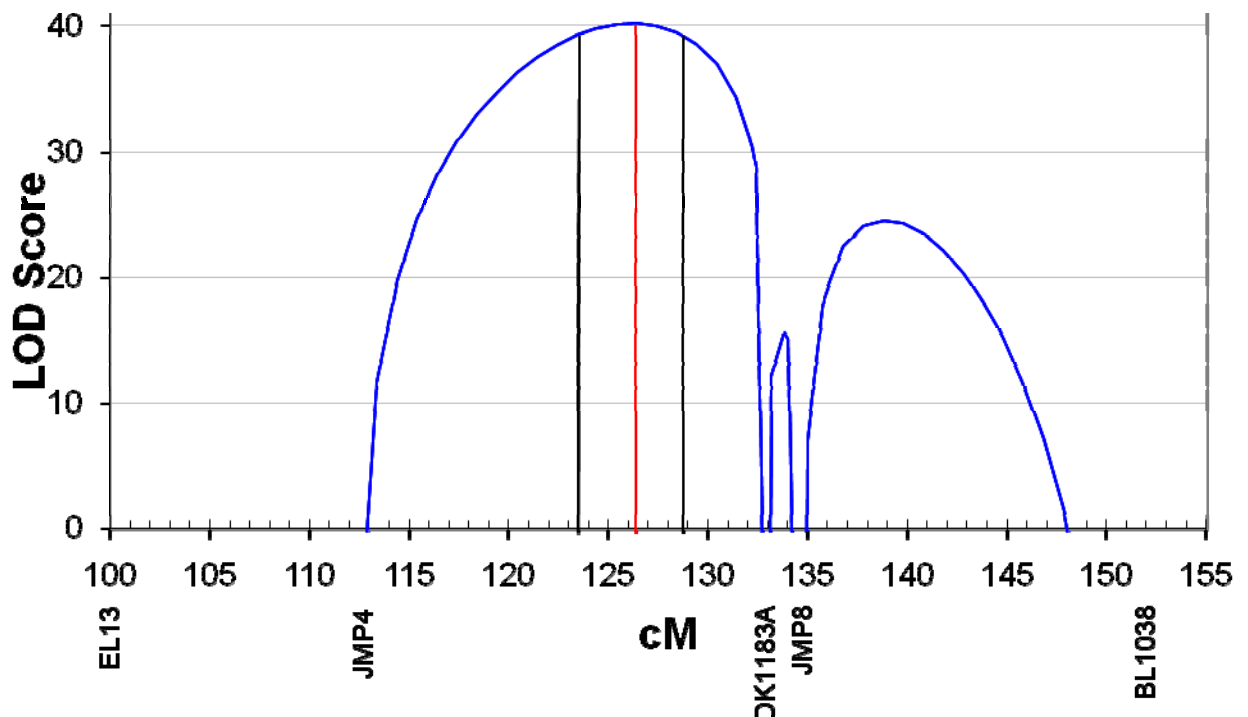


Figure 7: First multipoint linkage analysis. The analysis was performed on genotypes of 11 markers in 289 sheep. Markers are shown by the vertical text, and six markers are outside the borders of the graph. The red vertical line shows the most likely position for the OHC locus in this analysis, 126.4 cM, and the two vertical black lines show the edges of the 1-LOD support interval, 123.3 and 128.8 cM. Graph drawn by the author from data provided by Dr Ken Dodds.

3.2.4 Discussion

When selecting human or mouse genes to help select the target chromosomes, only autosomal dominant conditions were considered. The mode of inheritance of a cataract condition can vary in a single gene depending on the specific mutation, so some recessive conditions may involve the homologue of the OHC gene. An example of this is the variation between recessive and dominant inheritance of cataracts arising from mutations in the HSF4 gene in different breeds of dog, as described in section 1.4.3. However, any recessive condition may

be in a gene that is unlikely to cause a dominant condition. An example of this is an enzyme where a mutation eliminates enzyme activity. In heterozygotes the transcription of the enzyme gene can usually be easily up-regulated to compensate for the missing allele, and the mutant enzyme causes no disruption. One example of this is recessive genetic condition phenylketonuria, caused by a mutation in the phenylalanine hydroxylase gene (Lidsky *et al*, 1984). A mutation in the regulatory regions of the enzyme gene could cause a gain-of-function condition and therefore a dominant trait, but this is not guaranteed simply by the observation of a recessive condition. Therefore a recessive cataract associated with a gene is not sufficient evidence for its selection as a gene for choosing chromosomes.

Analysis of the breeding records (Chapter 2) indicates that OHC is inherited in an autosomal dominant manner with a penetrance of 0.81 ± 0.04 . No allowance was made for incomplete penetrance in selecting the pedigree to be genotyped. This means that animals with the cataract genotype could fail to develop cataracts and be scored as having the normal genotype. In linkage analysis this would increase the calculated recombination rate between the OHC locus and any linked markers above the actual rate. This problem occurs with any pedigree with a combination of normal and affected offspring, as is required for any linkage analysis. Multiple linkage studies (for example Köhn *et al*, 2009, Sundin *et al*, 2006, and Hedera *et al*, 2006) have been carried out on autosomal dominant conditions with incomplete penetrance without any selection of the pedigree taking the penetrance into account.

The distortion due to incomplete penetrance could have been removed by scoring each normal animal with an affected parent as unknown for the OHC locus. This would be consistent with all the genotypes that could produce a normal animal, but would remove information from the pedigree file and weaken the analysis more than assuming each normal animal had the wild-type genotype. Therefore this was not done and the pedigree file was constructed as if the OHC locus had complete penetrance.

Some linkage analysis software can be programmed with the estimated penetrance of the locus under analysis and adjust its calculations of linkage to take the penetrance into account. An example is the FASTLINK program which has been used to analyse autosomal dominant conditions with incomplete penetrance (Köhn *et al*, 2009, Hedera *et al*, 2006). The MENDEL program also has this feature and was available during this investigation. However the collaborators at AgResearch were more familiar with the CRIMAP program, which does not

allow the input of a penetrance parameter. This meant they were able to instruct the author in its use and would not be able to do so for MENDEL. Therefore the CRIMAP program was chosen. This meant that all linkage analysis was done with an incorrect model that assumed the OHC locus was fully penetrant.

A mathematical study (Clerget-Darpoux *et al*, 1986) of lod score analysis with penetrance parameters different from the actual penetrance showed that this error tends to lower lod scores. However this effect is slight, and in any case it would not produce a false positive result. Any marker showing significant linkage in this analysis would also show at least as much linkage in an analysis that took into account the incomplete penetrance. However, the analysis described in this section may be underestimating the number of markers with significant linkage and the size of linked regions. This also applies to analysis in the rest of this chapter.

A study of Fuchs corneal dystrophy (FCD) (Sundin *et al*, 2006) used the MLINK program with the assumption that the locus had full penetrance. Further investigation showed that the locus had incomplete penetrance, but MLINK still detected markers with significant linkage.

Multipoint linkage analysis is much more sensitive to penetrance in models of inheritance than simple lod score analysis (Risch and Giuffra, 1992). Therefore the multipoint analysis carried out in this section may not be reliable. The error would be of the same type as in lod score analysis: an underestimation of linkage and size of linked regions. No false positive results would be produced. The FCD study (Sundin *et al*, 2006) also used a multipoint analysis and did not detect significant linkage until the parameters were adjusted to reflect incomplete penetrance.

Oxford grids were used to compare homologous genes between sheep and humans, sheep and cattle, and cattle and humans to find sheep chromosomes likely to contain cataract genes. At the time the study was carried out this was the simplest way to investigate homology between different mammals. Chromosome painting results were also available but were harder to interpret. The website listed in this section no longer displays Oxford grids, these have been replaced with other online homology comparison tools.

Of the 356 animals that were genotyped, 89 or a proportion of 0.25 had marker genotypes incompatible with their sires in the original pedigree. Although all arranged matings were single-sire, such a large proportion of parentage errors indicates that other matings occurred and offspring were produced from different sires. A total of 22 of the 89 offspring had a genotype that was compatible with one of the other three affected sires. The pedigree was changed to reassign these offspring to the new sires, and the linkage analysis proceeded using the new pedigree. The remaining 67, 0.19 of the original 356, had genotypes incompatible with any of the affected sires. This indicated that rams from outside the OHC flock at the same sheep facility had mated with the ewes, producing offspring with unknown sires. These animals were omitted from the pedigree.

Although the large proportion of animals with genotypes incompatible with the expected sires indicates poor control over the mating, the remaining animals have genotypes of several highly polymorphic markers compatible with their expected sires. This is unlikely to occur by chance with another sire and their parentage is probably reliable.

Two software packages, CRIMAP and MULTEXCLUDE, were used in this section. CRIMAP is a linkage analysis program that accepts a pedigree file and can output a variety of data, including twopoint LOD scores, loglikelihoods of orders of loci, and estimated haplotypes. CRIMAP cannot accept penetrance or allele frequency parameters – all parameters are either calculated from a Mendelian inheritance model or from the pedigree file itself. When finding the position of a locus, CRIMAP can only give the most likely interval between two existing markers. With a set of markers at known positions, CRIMAP can only accept the relative order of the markers and the actual position of the interval in cM must be taken from the positions of the markers. In order to estimate the locus' position inside this interval, a multipoint linkage analysis must be performed with a software package such as MULTEXCLUDE.

MULTEXCLUDE uses both a pedigree file and a list of marker positions in cM. A hypothetical position for the locus under investigation is generated and a LOD score is calculated for the position from the linkage to the other markers. Then the hypothetical position is moved a short distance, usually 0.1cM in this investigation, and the process is repeated. With enough iterations a curve of LOD scores can be plotted through the region of

interest. The peak of the curve gives the most likely position for the locus, and the region within a LOD score of 1 of the peak gives the equivalent of a 90% confidence interval.

The LOD score results differ between the two software packages because they represent different results; CRIMAP gives the twopoint LOD score between the OHC locus and each marker, whereas MULTEXCLUDE gives the LOD scores for the presence of the OHC locus at each position.

Significant linkage has been detected between the OHC locus and six markers on sheep chromosome 6, indicating that the gene causing OHC is located on that chromosome. Chromosome 6 contains none of the candidate genes used to select the first 10 chromosomes; it was part of the second partial genome scan. Therefore the OHC gene is probably not homologous to any of the cataract genes used to select the chromosomes.

Sheep chromosome 6 is homologous to human chromosome 4, and at the time of this investigation the only gene on this chromosome associated with human cataracts is WFS1. Mutations in this gene in humans cause Wolfram Syndrome, a recessive genetic condition which involves diabetes and optic atrophy (Khanim *et al*, 2001). Cataracts are sometimes found in this condition and are considered a secondary effect of the diabetes (Al-Till *et al*, 2002). WFS1 was not included in the original list of cataract genes because it does not always cause cataracts and the condition is inherited in a recessive manner. A search of the latest sheep genome assembly, Oar1.0, indicates that WFS1 is within this linked region. The position of WFS1 in relation to other genes is shown in Table 20. Since there is no evidence for diabetes or optic atrophy in OHC sheep, WFS1 is probably not the OHC gene.

The gene AFF1 is found on cattle chromosome 6, which is homologous to sheep chromosome 6. Mutations in AFF1 are associated with cataracts in mice which are inherited in an autosomal dominant manner (Isaacs *et al*, 2003). Therefore AFF1 is a candidate gene. However when the chromosomes were being selected the references to AFF1 causing cataracts were not found, so its chromosome was not selected. After the chromosomes had been selected another literature search using the alias “AF4” for AFF1 found references to its association with cataracts. AFF1 is evaluated as a candidate gene for OHC in Chapter 4.

The linked region given by the CRIMAP analysis covers about 20.6 cM, and the 1-LOD support interval given by the multipoint analysis covers 5.5 cM. Before candidate genes were selected for testing, these regions had to be narrowed to reduce the number of genes covered. No more characterised sheep microsatellites were available, therefore bovine microsatellites were selected from the homologous region of the bovine genome and tested to determine whether they were present and informative in sheep. Informative markers were tested on a subset of the pedigree to narrow the linked region.

3.3 Breakpoint Panel Scan

3.3.1 Introduction

In the following section, the region of sheep chromosome 6 linked to OHC was narrowed by finding new markers and testing them in a breakpoint panel selected from the cataract flock. New markers were selected from the bovine genome assembly and tested in the four rams used in the genome scan. Markers that were informative when tested in the rams were sent to Invermay to genotype the breakpoint panel.

In order to narrow the region of sheep chromosome 6 linked to OHC, more markers were selected from the region of interest from the latest bovine genome assembly available at the time, Btau3.1. They were tested in sheep by a similar procedure to that described in section 3.3. After the initial testing, Btau3.1 was reported to also have contigs in the wrong order, so the latest assembly, BosTau4, was searched for the new markers. The majority of the markers were still within the linked region so they were used to genotype the breakpoint panel.

The two oldest affected ancestors of the animals with available DNA were studied to determine which markers had a common allele. The absence of a common allele between two affected animals means that there must have been a recombination between the marker and the OHC locus. Without a recombination, two related affected animals would inherit the same allele of the marker on the same chromosome as the OHC allele. The microsatellite markers used in this analysis are unlikely to mutate over the length of time covered by the pedigree, so the alleles can be treated as constant.

Together with the order of the markers, the presence or absence of common alleles can be used to restrict the position of the OHC locus. Several adjacent markers that share a common allele in affected sheep would give the linkage disequilibrium (LD) region with OHC. The name refers to the fact that the different markers have a non-random association of alleles with the OHC locus due to linkage. The OHC locus is likely to lie inside the LD region.

3.3.2 Methods

3.3.2.1 Breakpoint Panel

The linked region is between markers JMP4 and DK1183A. The CHROMPIC option of CRIMAP was used to find the 19 animals that had a recombination between these two markers (Figure 9).

The CHROMPIC option also found 15 animals with a recombination between markers EL13 and DK1183A (Figure 9). The three markers are in the order EL13-JMP4-DK1183A, so a recombination between EL13 and DK1183A must either be between EL13 and JMP4 or between JMP4 and DK1183A. A lack of phase information for JMP4 for the 15 animals means that the two recombinations cannot be distinguished and both were shown as between the two markers with phase information. Since some of these recombinations could be between JMP4 and DK1183A, all 15 animals were included in the breakpoint panel. Three animals had recombinations between markers JMP4 and JMP8 (Figure 9). These could also include recombinations between JMP4 and DK1183A, so all three animals were included in the breakpoint panel.

Ten ancestors of these animals with DNA samples available were selected for genotyping. A total of 47 animals were included in the breakpoint panel, along with one full sibling of a breakpoint panel animal (Figure 8). Animal 43/05 was not genotyped for the first set of new markers due to an error by AgResearch at Invermay and was excluded from the analysis.

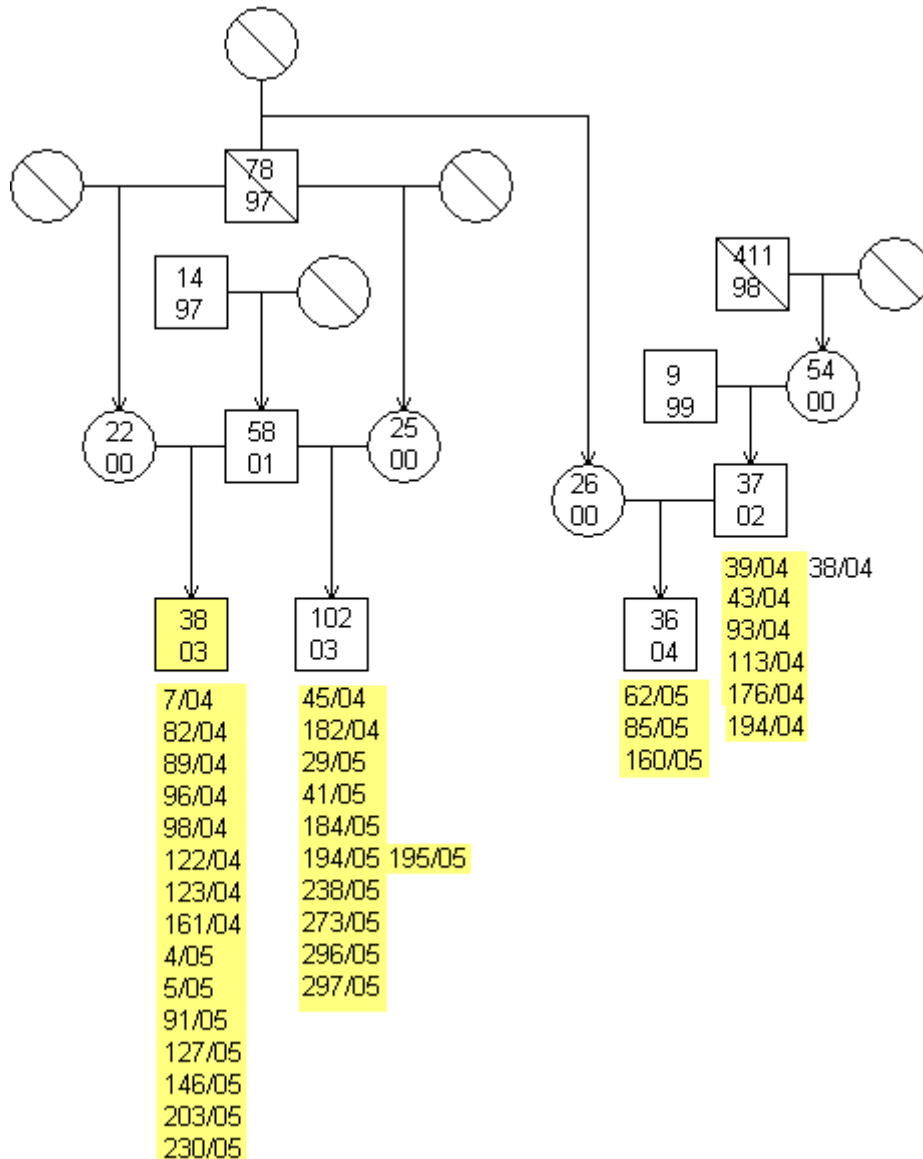


Figure 8: Pedigree of animals in the breakpoint panel. The 36 recombinant animals are highlighted in yellow. The animals listed below the four sires are their offspring with normal unrelated ewes, with full siblings shown on the same line. Sheep 38/03 is both a sire and part of the breakpoint panel. Sheep 38/04 has no known recombinations, but is the full sibling of 39/04 and was also genotyped. Animals with a diagonal slash have no DNA available. All other animals have a DNA sample included in the analysis. All animals with numbers are affected, all animals without numbers are normal unrelated ewes.

	EL13	JMP4	DK1183A	JMP8	Cataract Status	Recombinant Chromosome
14/97						
58/01						
38/03	_____				Cataract	Paternal
7/04		_____			Cataract	Paternal
82/04		_____	_____		Cataract	Paternal
89/04	_____				Cataract	Maternal
96/04		_____			Normal	Paternal
98/04		_____			Cataract	Paternal
122/04		_____			Cataract	Paternal
123/04		_____			Normal	Paternal
161/04		_____			Normal	Paternal
4/05		_____			Cataract	Paternal
5/05		_____			Normal	Paternal
91/05		_____			Normal	Maternal
127/05		_____	_____		Normal	Paternal
146/05		_____	_____		Normal	Paternal
203/05	_____				Cataract	Paternal
230/05		_____			Normal	Paternal
102/03						
45/04	_____				Normal	Paternal
182/04	_____				Normal	Paternal
29/05	_____				Normal	Paternal
41/05	_____				Cataract	Paternal
184/05	_____				Normal	Maternal
194/05	_____				Cataract	Paternal
195/05	_____				Normal	Paternal
238/05	_____				Cataract	Both
273/05	_____				Cataract	Paternal
296/05	_____				Cataract	Paternal
297/05	_____				Normal	Paternal
9/99						
37/02						
38/04					Cataract	Paternal
39/04		_____			Normal	Paternal
43/04		_____			Normal	Paternal
93/04		_____			Normal	Paternal
113/04		_____			Normal	Paternal
176/04		_____			Normal	Paternal
194/04	_____				Cataract	Paternal
36/04						
43/05		_____			Cataract	Paternal
62/05		_____			Normal	Paternal
85/05		_____			Normal	Paternal
160/05		_____			Normal	Paternal

Figure 9: Breakpoint panel recombinations. Markers are shown at the top. The horizontal lines show which pair of markers the recombination is between for each animal. The 37 animals in the breakpoint panel are listed along with their sires and grandsires. The colours indicate which generation each animal belongs to. The first generation is orange and the second, third, and fourth generations are yellow, green, and blue respectively. Offspring of each sire are listed below him in birth order. The black boxes indicate pairs of full siblings. Sheep 38/04 has no known recombinations, but is the full sibling of a sheep with a recombination and was also genotyped.

3.3.2.2 Fine Mapping, Part One

The homologous region of the bovine genome was searched for microsatellites. A total of 63 microsatellites were selected from the Btau2.0 assembly of chromosome 6, between the 50 cM and 65 cM positions. The 300 bases surrounding the repeat sequence were used to design primers using the Primer3 program available at <http://frodo.wi.mit.edu/primer3/> (Rozen and Skaletsky, 2000) (Table 13). Each primer set was used in a touchdown PCR reaction using bovine DNA and DNA from the four rams as templates.

Table 13: Microsatellite primer sequences from Btau2.0 assembly.

Marker	Forward Primer 5'-3'	Reverse Primer 5'-3'
BMS1508899	ACGTCGATGTTGACATTTGG	TATCGACCCAAGGACTGGAC
BMS1508976	GGAGGGGACCTGGGTATATCT	AGCAGCAGCAATGCACATAA
BMS1509059	AGGATGAAATCCCAATCCA	TTCAGAATGCCCATTTGTCA
BMS1509106	CCACTCTGATATGAAGAGGGAGA	AATCGCCCAATATGTCTCGT
BMS1509192	GAAGCAACCTAAGTGTCATCA	ATCCATGTTTTTGCAAGTGG
BMS1509206	GTC AAGCTTCCCTCATCCTG	AGCTCCGGAGTAACTGAGCA
BMS1509297	GATGCCCAGAAACAAACAGA	GCAATTAACCACTGCCATA
BMS1509387	CCTTGGCCTCTGTAGGAAAG	TGAAGTGGGTTTCTTGTAGACTCA
BMS1509482	GGCAAATTGTAAGTTCACTGGA	CCTCAACCAAGGTGAGGTTT
BMS1509593	TCGATGGACGTGAGTTTGAG	AAGTGGGATGCTGCCATAAT
BMS1509702	GTTACTTACGTATTTTAGCCACCT	CGCCACAATTTGTATCTGAGAG
BMS1509791	CATGTAAGTAGGACTTTCCATTGAC	GCCTGAGTTCAGTCCTTGGT
BMS1509869	GGAGACACTCGCACACACAT	CCGATCAAAAATCCACCATC
BMS1509996	TGTGAAGAGGAAAAGGCATCA	CTCCTCACGGGCAAGTAAGT
BMS1510131	GATGCACTAACAGTGAGCTTATCA	CCATAGGGTCTCAAAGAGTTGG
BMS1510292	CGCAGCGAGGAAAATGTTTA	GGTCAGAATGACCTGGTCTCA
BMS1510434	TGGGGGCTCAACTCTAAATG	GTGGGATGTGGAGAAGGTGT
BMS1510511	GGTTTGAGCCACTCTCCTGA	TCCAAGGTTGCAAAGAGTCA
BMS1510639	TGTTCTCAACTCTGCCTGTCTT	CAGTCCACAGGGTCAAAAG
BMS1510724	AAATCCGATGGACAGTGGAG	CTCTGCTTCCTTATGCACCA
BMS1510829	CCTCCAAATCCATCCATGTC	GCAATCTAAGTGTCCATCAACAG
BMS1510869	CATCTTCTGCATGTGTTTGA	CGACTGAGCGATTGAACTGA
BMS1511009	CCCAAAGTACCTGAGCCAGA	GCTTTCTTTTACCTGGATAACTCC
BMS1511091	TCCTAGAAAAGCGCATCTTACA	CCATCTGAGCCTGAAGGTTT
BMS1511155	CCACAGAAGACAACAGTCTACCA	GCTGTGTAATGGTCCTTTGG
BMS1511218	AGGAATCTGTCCTGCAAACC	TGTCCCTGCCAATTCTTCT
BMS1511300	TGGCTGTCTACAGTCCATGC	ACGTGTGCAAAGCAATTCAT
BMS1511389	GGAAAACATAACGCCAGACG	AGCCATCTTCAAATCGGTGT
BMS1511509	TAAATGCCCAGAGTGGGAAA	TTCTCGAGGTTAATCCATGTTG
BMS1511603	AGGCTGAGGAGAATTTGCAG	GGTGCACCACAAAGTGTGAG
BMS1511687	GTAGGGGCAGGACATCTTGA	CCAAGGGGAGGAAAGAATGT
BMS1511787	GGGGATGTGTCATTTTTGCT	TTGACATCCTAGTGTCCCAAT
BMS1511896	GAGGTTTGGTTGGTTTGTG	AGATAAGGAGGGCCGACTGT
BMS1511966	AACAACCTTAGGGCCATTTTG	AAAGTTCCAACCCACAACCTCA
BMS1512041	AGGGCTTATAGGCGTGATGA	CCAGGCACTTTGTTTCCCTA
BMS1512131	AGCCTCCTGGTGTATGCTTG	GAAGGCTTCAGGTCAGCCTA

BMS1512218	GGGGGAGTGTGTCTGAGTGT	GCTCCACTCATGGGACTGAT
BMS1512327	TTTGCCCAAGAGATGGATTG	GCCTGGAAGACTACGGTCTG
BMS1512413	CAATAGCCAAGACATGGAAGC	TCTAGGTCCATCCATGTTGC
BMS1512533	TCACTTTGCCAACAAGGTCT	GTCCAGAGGGTTGCAGAGTC
BMS1512589	TGCTCTGCTTCAGGGTCTTT	GAAGCTAGCCAGGCAGAAAA
BMS1512718	GGGGGAAGAATCTGAAAAGG	CACCACACAGATTTGGAGCA
BMS1512786	GATGGGAGAGCTGGAGACTG	CCTGGACTCTGCTCATCAAA
BMS1512864	CTTACTGGTGGGCTGAGGAG	ACCCTTGTTTTCACCTCTCG
BMS1512958	TGTATTTTGGGGGATGGAAA	ACATGCCTTGAGGGCCAAA
BMS1513060	GGAGAAAATTGAGGCAGCTT	CTGCAGCTCTGCAAAACCTA
BMS1513088	ATATCACGGAGGCCAACTGC	CCTTCCAGTGACCATTCAGG
BMS1513151	GTCCCACTTCTGCCTCTTTG	TGTCAGTTGCCTACAATGTGC
BMS1513195	TGGAGAATCCCACAGACAGA	CTGGCACACAGTGGATTTGT
BMS1513254	AAGCGACTTAGCACAGCAGA	GTCTGGTTCAGTTTTTGTTC
BMS1513364	ACGGTACTTTAGGGGGTCTCT	CACACCTCCCTATTGAATTTCC
BMS1513442	GCCAGCTTACAGTGAAGAAAGA	CCAACCTCAAACCAGATAAGCAA
BMS1513522	TTTCCTGGTGCCTGAAGAAG	AAGATCGAACCCAGGTCTCC
BMS1513600	CGAGCGCCACACTAAAATG	GCGCTCTTCCAAATTCCTTA
BMS1513647	AAAGCGCAAGGACTAAGGTG	TGCTCAGTGGGGGACATT
BMS1513679	TGAGGGGTGGCCACTTATAG	CCATGTTGTGTTAGTTTCAGGTG
BMS1513738	ACCTAGAGCTCCTGGTGTCC	ATTTGCAAACACATGCACCA
BMS1513755	GCTCTGTTGTGTGCATGGAT	TTTCTCAAACGTCCCTGACC
BMS1513823	CAATCACGGCTTCTCTGGAT	TGGAGAATTCCATGGACAGA
BMS1513888	CTGGAAGCCATCAAAATTGC	GCCACTGTGAAAGGATGACA
BMS1513944	GCACTGCGCTATTGGTGTTA	TTGTGGGGACACAGTTCAA
BMS1514069	CAGGGAAAGGTGGTTAGTGG	GCAGAAATGCCATCAGGACT
BMS1514207	ACGGCTGTCTGCTCCAGTAT	AGCTCCCTGTGGGGATAGAG

Every forward primer had the sequence TGTAACGACGGCCAGT attached to its 5' end. This sequence allows the product to be inserted into an M13 plasmid easily. Every reverse primer had the sequence GTTTCTT attached to its 5' end. This is a "pigtail" sequence used to ensure that an extra adenine is consistently added to the PCR product by the DNA polymerase. Without the pigtail sequence the addition of the extra A is unpredictable and can make microsatellites harder to score (Brownstein *et al*, 1996).

.Each PCR reaction mixture consisted of a 1:10 dilution of venopuncture DNA extract, 200µM dNTP mix (Invitrogen), 1µM of each primer, 5mM MgCl₂, and 0.5 units per µL of Platinum *Taq* DNA polymerase (Invitrogen). The magnesium concentration is relatively high, to make the PCR conditions less stringent and allow the primers designed from bovine sequence to amplify sheep DNA. The final volume of each mixture was 10 µL.

The PCR cycle involved an initial denaturation, 94°C 5 min, followed by 5 cycles of 94°C 30s, 60°C 30 s, and 72°C 40 s. The 5 cycles were repeated with an annealing temperature of

58°C, 56°C, 54°C (twice), and 50°C, for a total of 30 cycles. There was a final elongation step, 72°C 45 min. The PCR products were visualised under UV light on a 2% agarose gel, following electrophoresis at 90V for 1 h with ethidium bromide. Primer pairs that gave single bands with a size from 100-200 bp, in at least two rams, were selected for further testing at Invermay.

A fluorescent labelling PCR reaction was performed with the selected primer pairs and the same ram DNA, to determine the ram genotypes. The PCR products were resolved by a capillary and fluorescence detection method as described in Section 3.2.2. Markers can only be used to detect recombinants if a parent is heterozygous for them, so markers were only selected if at least one ram was heterozygous. These are referred to as informative markers. The 37 sheep in the breakpoint panel and their 10 ancestors were genotyped for the polymorphic markers.

The genotypes derived with the new markers were used along with the genotypes derived with the other markers for linkage analysis using the CRIMAP program to further refine the position of the OHC locus.

Animals that were not part of the breakpoint panel were included, scored as unknown for all of the new markers. A multipoint graph was prepared from the genotypes by Dr Ken Dodds at AgResearch Invermay using the MultExclude program. This involved placing the OHC locus at a series of positions less than 1cM apart throughout the linked region and evaluating the LOD score at each position, using all the available genotypes.

3.3.2.3 Fine Mapping, Part Two

Eight target positions on bovine chromosome 6 were selected to be searched for new markers for the next stage of genome scanning (Table 14). The eight target positions are labelled T1-T8. Positions T4-T7 were selected to coincide with genes inside the 1-LOD support interval given by the multipoint analysis shown in Figure 11. These four genes (PLAC8, NKX6-1, CDS1, and WDFY3) have no known association with cataract. Position T8 is the position of NUDT9. This gene is just outside the 1-LOD support interval, but was known to be expressed in the lens and to be a good candidate for the OHC gene, as described in Section 4.3.1.1. Positions T1-T3 are roughly evenly spaced between marker BMS1509996 and the 1-LOD support interval.

Table 14: Target positions for new markers. The eight target positions are labelled T1 to T8. The start and end of the 1-LOD support interval and the most likely position for the OHC locus are highlighted in yellow. Existing markers are highlighted in blue. Where a target position is the same as the position of a gene it is highlighted in green. The positions are given in kilobases from the centromere of bovine chromosome 6, taken from bovine genome assembly Btau3.1.

Targets	Gene/Marker	Bovine kilobases
	BMS1509996	87,453
T1		89,259
T2		91,065
T3		92,870
	Start	93,492
T4	PLAC8	93,630
T5	NKX6-1	94,244
T6	CDS1	94,346
T7	WDFY3	94,446
	OHC	94,482
	End	95,180
T8	NUDT9	95,225
	BMS1512327	96,023
	DK1183A	96,481

A total of 146 microsatellites were selected from the vicinity of the eight positions, using the SPUTNIK program (Abajian, 1994) and the bovine genome assembly BTau3.1. Each marker was labelled with a number indicating which position it came from on the bovine chromosome, and its number in the set of markers at that position. For example t8s16 is the 16th marker tested from the 8th position. T stands for “target” and s stands for “SSR”. The web program REPEATMASKER, available at <http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker> (Tarailo-Graovac *et al*, 2009) was used to find repetitive elements in the sequence surrounding each microsatellite. The web program Primer3 (Rozen and Skaletsky, 2000), was used to design primers for each microsatellite from the sequence surrounding the dinucleotide repeat, avoiding repetitive elements.

The microsatellite primers were synthesised with the modifications described in section 3.3.2.2 and used in touchdown PCR reactions with DNA from rams 37/02, 38/03, 102/03, and 36/04, and also from an unrelated ewe, to test whether they amplified a product of the appropriate size in sheep. The PCR conditions are described above. Primer pairs were tested in batches covering all eight positions. More primers were tested from positions where the markers in the first batch were not informative. In total, 83 primer pairs were tested (Table

15). The remaining 63 markers either had too much repetitive sequence to design effective primers, or were in positions close to previously tested markers.

Table 15: Microsatellite primer sequences, second batch.

Marker	Forward Primer	Reverse Primer
t1s1	TGTATCCCAATGCTCACTGC	CAGCCATGTCTCTGCAAATG
t1s4	CGACCCTGGTTTTCACAACT	TCTGGTCCTTTTCAATTGGTG
t1s4a	GCTCCCTCAGCAACAGCTAC	TCCCACTGACACTCATATTGCT
t1s8	CGACCCTGGTTTTCACAACT	TCTGGTCCTTTTCAATTGGTG
t1s10	CCTGCTGTGCACTGTCCTAA	CTAAACCGCAGCAAAAGAGG
t2s1	TGAGTTGGCACCTAGGATCTC	CTCGATGGACGTGAGTCTGA
t2s2	GCAAGTATTTGGGCTCTTG	CATTGGCAGGTGGATTCTTT
t2s3	GCCCCAGGTTAAGAGCTGTA	AGTCACCCTATTTTGGGAAGGA
t2s4	TTGGAGAAGGAAATGGCAAC	ACAGCGAATGCTGCTCAATA
t2s5	CCCACATGGTTAGTTCAGCA	ACGCCTCATCCTAAGAGACC
t2s6	GGAAGATCCCCTGGAGTAGG	GGCACAAAGAAGGAGAGAAGAA
t2s7	AGAAGGGGACGACAGAGGAT	TCGTTTCATGGAATTTCTCCAG
t2s8	TGAAGACCTGACTGGAAGGAA	AATCCTGGATGGTCTTTGCT
t3s1	TGAAATAAAAGTGCACAAATGGTT	CCTGGGATTTGATGGAGTCT
t3s4	GGAAATGGCAACCCACTCTA	GAACAACAATGCCCATTTTC
t3s5	TAGGAAATGGCAACCCACTC	TCAACCTGCTGTCAAAGGTG
t3s7	GAGTCGGACACAGCTTAGCA	GCCATGTCCGTTGATAGCTT
t3s8	TCCCTTCAGAGCCTTTTAACC	TGGAAGAAGGCTGTTGCTTT
t3s9	TTTCCAGGCACTTTGTTTCC	AGGGCTTATAGGCGTGATGA
t3s11	AAGCCACAGAGGGTCAAAG	GTCACAGGCCCTCAGTGGTA
t3s13	AGGCGGTTTTACATCCATGA	CTAAAACCTTGCATGGCACA
t3s14	ACCACACTGTGCCATGAGAA	TCTGAAGAATGCCCAAATC
t3s15	TGCAGAGAACAGCAATTTG	CACAGGCCCTAACCTTTCAA
t3s16	GACAAGGTAAGTCAGATTTATGCAG	CCCTTTAAACCACAGCCTTC
t3s17	AAAGACCTGCAGGAACCAGA	GGGGCAGCATATACACTCCT
t3s19	ATGCTTGCTCCAAACACTCC	TGAATATTTGCCTCATATTTTCATCC
t3s20	TGTATATTTCCAATCCAGGTGTTG	TCCCACTTTCTTCTTGGGAAT
t4s1	CCAGAGCTTGTGGAACCTGT	CACTGGACAACCAGGGAAGT
t4s2	GAGTCGGACACAACCTGAGCA	CCCCTGGTCTAAGTGTGCAT
t4s3	ATTCGCATTGCACATCTCTG	CGCAAGGAGTTGGACATGAT
t4s4	GTGGGATCCAGGAGTCTGAA	GAAATCCAAAAGGGAGGAGA
t4s5	TGGGGTGAATGTGCATCTAA	CCAGGGAAGCTCATGCTAAA
t4s6	CCCCTTCTGATGCCTATGTT	GGTGTGGGAAAAGGAATGA
t4s7	CCCTTGGAGGAGAAAATGGT	CTGGAAACAGTTGGCATTCA
t4s9	TTCAGGGATGATAATGGTATTGC	TGACTCTCTTCTCTCTCACATCC
t4s10	TCCCATAGGAATCTGGACTCA	CCCAGTCTAAAGGGTGATGG
t4s11	CACTTCTGTGCTGCTGTGGT	TCAATCATCCTCCAAACACAA
t5s1	TCCTCATCCTCCGAAGTCTC	ACTCGGACGACGAGAAAATC
t5s2	TCACTCCTGAAATGCATTGAT	CCAGGGAATCAAATTGGAGA
t5s3	TCACTCCTGAAATGCATTGAT	CCAGGGAATCAAATTGGAGA
t5s4	CAACCTCCTTCGCTGATCTC	GAGTGGACCTTGCAAACCTC
t5s9	AGTGAAGCATCTGCCAAGT	GGAGGAGCTCCATATTTCCA
t6s1	TGTATGATCACAATGCACACAAA	GGGGATGACAGAGGATGAGA
t6s2	AGAAGGGGACGACAGAGGAT	CTGCAACAACATGGATGACC
t6s3	AGATAAGGAGGGCCGACTGT	TTTTCTGGCCAAAACCAAAG
t6s4	ATGCTGTGTGGCAGAGTCAG	ATGGACAGAGGAGCCTGATG

t6s5	GCCAAAACTTTGTTGAAACATC	TCCCTCACACCGTGTACAAA
t6s6	TACAGCAGTTCATGGGGTCA	CTGCACCCATAGCCTTGACT
t6s7	GCATTAGGGAGATGTTGGAAA	GCCAGGCAAGAGTACTGGAG
t6s8	TGTGAAAGATGTCACAAATCACTG	GGAAGATTCCCTGGAGAAGG
t7s1	GAGAAGGACATGGCAAGGAG	TGTGAAAGATGTCACAAATCACTG
t7s2	CATGGGTTGGAGCCTAACAG	GGACCAAGGATAGGACCAA
t7s3	TTGTCTTAAGAAATCCAGGGTCA	TGGGTCAATAAGAGAAACAGACAA
t7s4	GAAAAGGGTAGGCAAGCACA	TTCCCATGGACCACTGCTAC
t7s5	AGCCTTCTCCCACTCCTTTC	GGGGATGTGTCATTTTTGCT
t7s6	AAACCAAATGGTCCTTTCTCC	GCTTTACGGTTCAGGGAGTG
t7s7	TTGAAAAGACCATGCCCTCTA	TGCTGGATTGTGTGAAATCTG
t7s8	CCACTGGTCAAGAGTGACAGG	TTTGAGTTTGGATGTGCAACTT
t7s9	TTGTTTTCTGCCCTCATTCC	TGATGAACCATAAGGGAAAAGAA
t7s10	TCCGACAGGACTGAGTGAGTT	GAGTCTGGGGGAGAATGGAT
t7s11	GACCATGGAGTATGCTATCTGC	GCCACTGGAACAAACCATTT
t7s12	TCCGACAGGACTGAGTGAGTT	GAGTCTGGGGGAGAATGGAT
t7s14	TTTGGTCTTTACCTCGTCCA	GCTTTTGTAAAGTGGGCAGCA
t7s16	CTTTTCACATGAGGGGAAGC	TCAGCAAAGAGCCCTTTAAGTC
t7s17	AATCATCTTCAGGGCCACAG	CCAAAGATCAGGGTGACTGC
t7s18	GCCTCCTCTGGGGAATGTAT	CTTGCTGTGTGCTAGGTGGA
t7s20	TGTCCCTCCAAAGAGCATT	TTTGTGGTCTTTGGCTACTGAA
t7s21	TCCATCAACCTCTCCTCCAC	TGAGCTCAACAAGGCATTG
t8s1	GAGTCGGACACGACTGAACA	CCTTGCACATTTCTGGGTAAGT
t8s2	TGAAAGTAAGGAGCCCTGGA	CCAGGTTCAATCCCTGGTC
t8s3	TCGATGGACGTGAGTCTGAG	AGACCCAGATTTGCAGCCTA
t8s4	GCTGCAAAGAGTCAGACACG	TACAGTCCATGGGGTCACAA
t8s5	AGCCGATTCTCAATTCCTT	GCCATCTGTGGTCCAGTCTT
t8s6	GCTGCAAAGAGTCAGACACG	TACAGTCCATGGGGTCACAA
t8s7	TGGCTGGTGTCTTGAATTG	AGGGTTGCACAGAGTTGGAC
t8s8	TTGCCTTTCCTTACCACTTTG	CTGTGGGGGAAAATTCAAGA
t8s9	GTGGATATGGCACGCTCTCT	CACTGACAGATGTGGGCATT
t8s10	AGCCTCCTGGTGTATGCTTG	GAAGGCTTCAGGTCAGCCTA
t8s11	AGGACATTGGGCCATGTTAG	GACCTAGACCTGGTGGACA
t8s13	ATTTCTGGGGGAAAAGAA	CTTCTGTCCCTGTTCCAAA
t8s14	AAAGTTGGGGAGGGAGAGAG	GGGAACAGTGGGGACAATTA
t8s15	TTCCCAGCACAGATTTTACA	GCTCGCATCCTCTCCTACAG
t8s16	TCCCTGGTGACAGAGGACTT	TACGAACCGCAAAACATTGA

The PCR products were visualised on agarose gels with ethidium bromide and UV light. Often multiple bands were present on the gel as well as the desired band. PCR products that amplified a band of the appropriate size were used in temperature gradient and magnesium concentration range PCRs to find the optimal conditions for amplification of the appropriate band. The temperature range was 55°C to 65°C and the magnesium concentration was between 1 and 4 mM. Bovine DNA was used in all of the optimisation experiments to confirm it that it could still be used as a positive control whatever the conditions. Once the PCR conditions had been optimised, PCRs were run using template DNA from the four sires of the genome scan offspring. Template DNA from four of the offspring and a normal,

unrelated ewe was also used. The products were run on a 6% acrylamide, 7M urea gel, 1500 V, 3h. Markers were considered informative if at least one of the sires had a different genotype to the majority of the sheep.

For each informative marker, a PCR was performed using template DNA from a sire. The PCR product was run on an agarose gel with ethidium bromide at 90 V, 1 h, and the products were visualised with UV light. The band of the appropriate size was cut out and extracted with a MinElute kit (Qiagen). One μL of the gel extract was run on an agarose gel alongside genomic DNA samples of known concentration. This allowed the quality of the gel extract to be checked and the concentration roughly estimated. Additional gel extract was used to prepare two sequencing mixtures, one mixture containing the forward primer and one containing the reverse primer. Each sequencing mixture contained 3.2 μL of a 1 μM solution of the primer, the equivalent of 20ng of PCR product DNA, and sterile deionised water to 15 μL . The amount of DNA needed was calculated by allowing 2ng per 100 bases of the sequence.

The mixtures were sequenced with a BigDyeTM Terminator Version 3.1 Ready Reaction Cycle Sequencing Kit (Applied Biosystems) and a ABI3730 capillary instrument by the Allan Wilson Centre Genome Service at Massey University. When the sequencing results returned, ambiguous bases were corrected where possible by studying the electropherogram and choosing the highest peak. The sequences were aligned with the homologous sequences from the latest bovine genome assembly at the time, Btau3.1, and also used in BLAST searches against the entire bovine genome. The results of the alignments and searches were used to confirm that each pair of primers was amplifying the expected sequence of each marker.

A new bovine genome assembly, Btau4.0, became available after the markers had been selected and tested. The positions of the markers were found on the new assembly and only one, t2s8, was outside the linked region given by the previous CRIMAP analysis. This marker was abandoned and the rest of the markers were used in further analysis in their new positions.

Rams 14/97 and 9/99 are the oldest ancestors that were genotyped. They are related but the details of the pedigree connecting them are unknown. Both were affected, and their genotypes for all of the markers were studied to find alleles common to both animals and

generate a Linkage Disequilibrium region where both affected rams shared alleles in adjacent markers. Ram 9/99 had no genotype for four markers, so alleles were reconstructed for it from the genotypes of his son, ram 37/02. This was possible because 37/02 was homozygous for those markers, and must have inherited one of the alleles from his sire.

Genotypes of the new markers were used with the CRIMAP program to narrow down the position of the OHC locus, as described above. The linked region given by this analysis was used to find candidate genes on bovine chromosome 6, as described in Chapter 4.

After the candidate genes had been investigated, a multipoint linkage analysis was performed to further restrict the position of the OHC locus and determine which genes were still candidates. In order for a multipoint graph to be generated, all markers had to have a known position in cM. For the older markers, these positions were taken from version 4.7 of the sex-averaged linkage map prepared by Dr Jill Maddox, available at <http://cmap.sg.angis.org.au>.

For the new markers, the position in bases on the physical map of the sheep genome was determined by a BLAST search of Ovine genome assembly Oar1.0, using the bovine sequence surrounding each marker as the query. The ovine genome database is available at <https://isgdata.agresearch.co.nz>.

Positions in bases were also found for the older markers by using their names in text searches on the sheep genome browser at <http://www.livestockgenomics.csiro.au/perl/gbrowse.cgi/oar1.0>. This viewer uses the same database as used for the BLAST searches, Oar1.0.

A Microsoft Excel chart was created using the linkage map positions of the older markers as the X-axis values and the physical map positions as the Y-axis values. A linear trendline was constructed by an automated least-squares method which gave the best estimate of the relationship between the positions. This was used to estimate the linkage map positions in cM of the new markers from their physical map positions. This procedure has been previously used to generate linkage map distances for human and mouse microsatellite markers (Bahlo *et al*, 2004).

Three markers (BMS1509192, t6s8, and BMS3944) could not be found on the sheep genome assembly. BLAST searches were carried out to find their positions on Btau4.0, the latest cattle genome assembly. The chart and trendline interpolation described above was repeated using Btau4.0 positions instead of sheep assembly positions, to estimate the linkage map positions of the three new markers in cM.

When the gene *NUDT9* was tested as a candidate, a G/A polymorphism was discovered at position 989 of the CDS, as described in Section 4.3.4.3. The polymorphism correlated with cataract status so it was tested in 238 animals to determine whether it could be the OHC mutation. As discussed in Section 4.3.5, the polymorphism is probably not the OHC mutation, but since the gene is within the linked region its genotypes were added to the pedigree and it was treated as a new marker.

The quality of the bovine genome assembly was checked by comparing the order of genes between the sheep, bovine, and human genomes. The new genotypes of the breakpoint panel were analysed using the CHROMPIC option of CRIMAP to find the positions of their recombinations.

3.3.3 Results

3.3.3.1 Fine Mapping, Part One

In order to find new markers to narrow the cataract linked region, 63 microsatellites were tested on four affected rams and a cattle DNA positive control. Of these 47 showed some amplification in the sheep samples and 29 gave single bands between 100 and 200 bp in at least two rams.

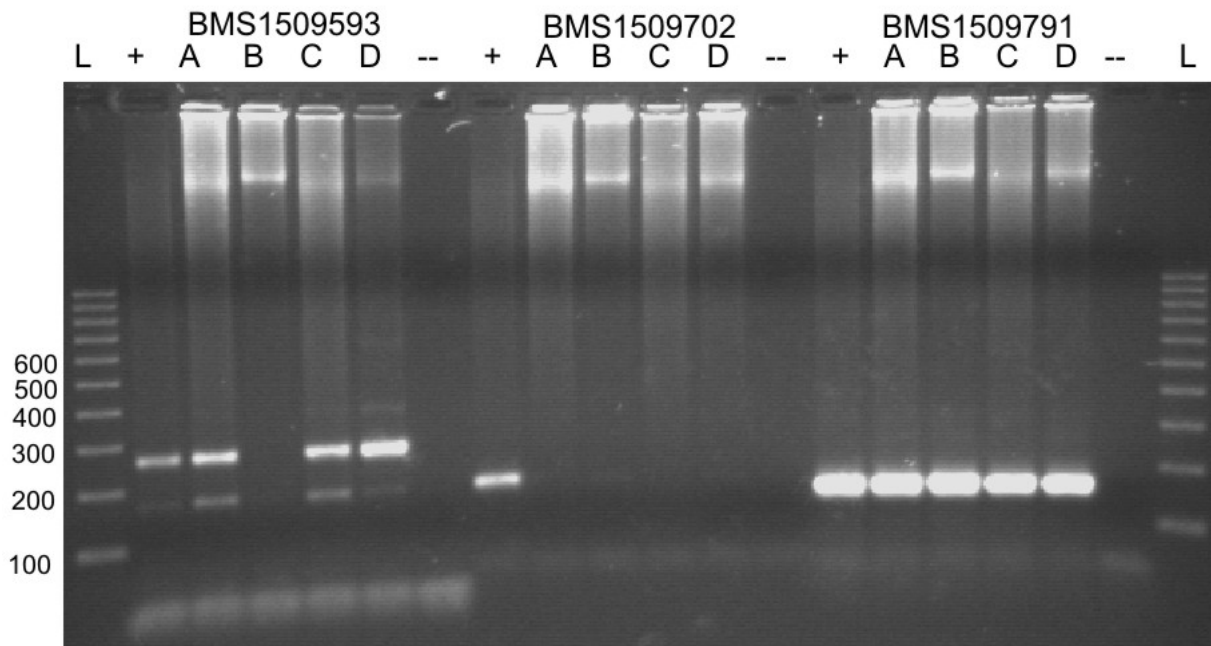


Figure 10: Products from BMS1509593, BMS1509702, and BMS1509791. The lanes labelled L contain 100 base-pair ladder (Invitrogen). Bands of the ladder are labelled in bp. The + sign represents the cattle DNA positive control and the – symbol represents the blank. Sheep samples A, B, C, and D are from rams 38/03, 102/03, 37/02, and 36/04, respectively. The relationship between these rams is shown in Figure 8.

In Figure 10, marker BMS1509593 shows bright bands of the appropriate size in both the cattle DNA and three of the four rams. However they also show dimmer bands at an unexpected size. Marker BMS1509702 show a band of the appropriate size in the cattle but none in any of the rams. This indicates that the microsatellite is not present in sheep, or that the conditions need further optimisation to display it in sheep. Marker BMS1509791 all have bright bands at the appropriate size, with no extra bands. From this group only marker BMS1509791 was submitted to Invermay for testing.

Of the 29 markers submitted to Invermay for testing, 8 were informative and easy to score; BMS1509192, BMS1509996, BMS1510292, BMS1512327, BMS1512786, BMS1512958, BMS1513060, and BMS1513944. The bovine genome assembly used to find the new markers, Btau2.0, was found to have contigs arranged in an order inconsistent with homologous regions of the human genome. When the new markers were found on an updated bovine assembly, Btau3.1, only three markers (BMS1510292, BMS1509996, and BMS1512327) were actually located in the linked region.

When the pedigree file for CRIMAP was being assembled, two animals in the breakpoint panel gave noninheritance errors. The genotypes of these animals and their parents were investigated and no obvious inconsistencies were discovered, but the noninheritance error persisted so these animals were omitted from the analysis. No noninheritance errors of this kind were observed in the previous analysis, and these errors were found after the parentage errors found in the previous analysis were corrected. A total of 45 animals had genotypes for the new markers included in the analysis.

The genotypes of all eight markers were used in the CRIMAP analysis, and this only slightly altered the calculated position. The cataract locus is almost equally as likely to lie between BMS1509996 and BMS1512327 as between BMS1512327 and DK1183A (Table 16). Therefore the entire interval between BMS1509996 and DK1183A is the region linked to cataract in this analysis.

Table 16: Loglikelihoods of different OHC locus positions. Each loglikelihood is the log to base 10 of the likelihood of producing the set of genotypes given each position of the OHC locus. The very small likelihoods are due to the large size of the pedigree; they decrease exponentially with the number of animals. The top two positions are almost equally likely, and they are both about 1800 times more likely than the position third from the top. The OHC locus is shown in red.

Position 1	Position 2	Position 3	Position 4	Position 5	Loglikelihood
BMS1509996	OHC	BMS1512327	DK1183A	BMS1513060	-380.288
BMS1509996	BMS1512327	OHC	DK1183A	BMS1513060	-380.301
BMS1509996	BMS1512327	DK1183A	OHC	BMS1513060	-384.552

At the time this analysis was carried out, the position of BMS1509996 on the linkage map was generated by interpolation from the physical map Btau3.1. It was determined to be 114.8 cM from the centromere of sheep chromosome 6. Therefore the linked region is between 114.8 cM and 132.9 cM from the centromere, compared to the old position between 112.3 and 132.9 cM. The distance has reduced from 20.6 cM to 18.1 cM.

The Btau3.1 physical map was found to be unreliable, so the position of marker BMS1509996 has been revised using the Btau4.0 map (Table 17). The position of marker DK1183A at 132.9 cM has been determined through independent linkage mapping and is not derived from any physical map. The interval between 114.8 cM and 132.9 cM is the best estimate of the linked region with the information available at the time the analysis was performed. The linked region derived from the latest physical map and analysis of more markers is described in the next section and Table 19.

A multipoint graph generated from the genotypes gave a 1-LOD support interval of 123.7-129 cM from the centromere, with the highest LOD score at 126.8 cM (Figure 11). These results are consistent with the conventional linkage analysis.

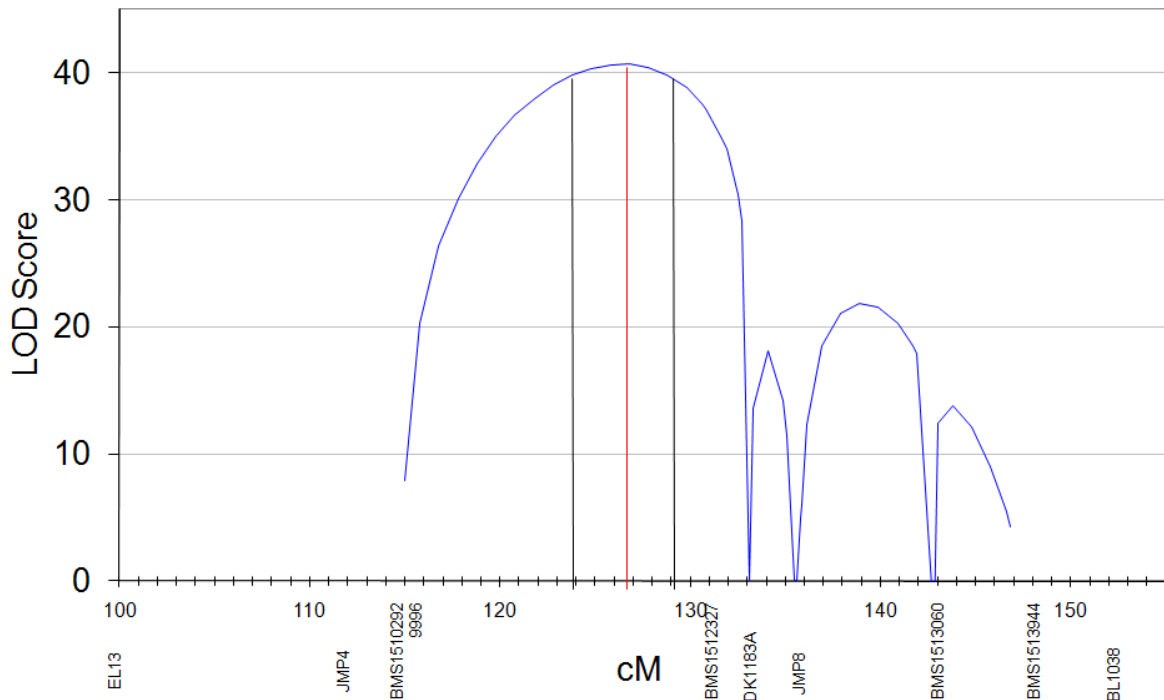


Figure 11: Second multipoint linkage analysis. This analysis is based on the genotypes of the original 11 markers in 296 animals, plus the genotypes of 8 markers in 48 of the 296 animals. Black lines enclose the 1-LOD support interval and the red line shows the point of highest LOD score. Markers are indicated by vertical text along the bottom. Graph drawn by the author from data provided by Dr Ken Dodds.

3.3.3.2 Pedigree and Marker Position Revision

To find errors in OHC status, the CHROMPIC option of CRIMAP was used to find animals where the OHC locus was out of phase with the loci on either side of it. This is equivalent to having two recombinations, one on each side of the OHC locus, which is unlikely. This

pattern was found in three animals entered as normal (136/04, 164/04, and 34/05), and two animals entered as cataract (80/04 and 58/04). The last two digits of the animal code number give the year of birth.

The cataract status records for these five animals were checked to determine whether their OHC statuses in the pedigree were correct. All of the animals born in 2004 had been examined for cataracts twice, one month apart, and the two observations agreed for all four animals. Animal 34/05 had been examined for cataracts only once, nine weeks after birth.

Animal 164/04 had been entered as normal, but records showed that it actually had cataracts. Animal 58/04 had cataracts, but had the same code number as a normal animal. No errors were found in the OHC statuses of the other four animals. The pedigree was revised to change the OHC statuses of 58/04 and 164/04.

When cM positions for all of the markers, including NUDT9, had either been found or generated, the order of the markers was different to that used in the previous CRIMAP analysis. The analysis was repeated with the markers in the new order, corrected OHC statuses, and with the NUDT9 genotypes included. Only results from the revised pedigree and marker order are presented. After the CRIMAP analysis, the marker genotypes and positions were used in a multipoint analysis.

3.3.3.3 Fine Mapping, Part Two

In order to further restrict the position of the OHC locus, the section of the bovine genome homologous to the linked region was searched for more microsatellites. A total of 146 markers were identified. Pairs of primers were designed for only 126 markers, as 20 markers had too much repetitive sequence around the repeat to allow useful primer design. Of these 126 markers, 83 were tested while the other 43 either had repetitive sequences discovered after primer design or were in similar positions to markers already known to be informative.

Of the 83 markers tested, six gave bands of the appropriate size and were informative when tested on acrylamide gels. These markers were t1s10, t2s8, t3s9, t3s19, t4s11, and t6s8. No informative markers were found for positions t5, t7, or t8. Figures 12-16 show the results from the testing and optimisation of marker t6s8 as an example.

The marker t8s16 has a 40-base insertion which occurs in some rams but not others.

Therefore this marker is technically informative, but it lies just outside the linked region and therefore is not useful for restricting the position of the OHC locus. There are other untested microsatellites in position t5 but not in t7 or t8.

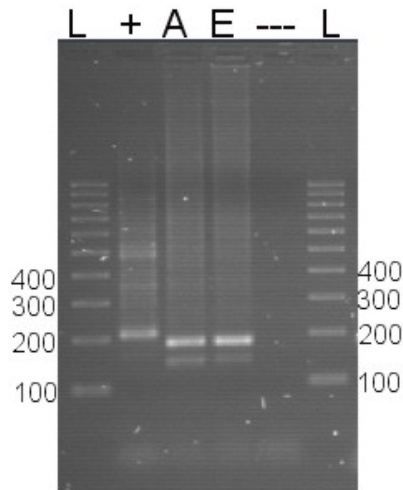


Figure 12: First amplification of t6s8. L indicates the 100 base pair ladder, + indicates the bovine DNA used as a positive control, A is a sample from affected ram 38/03, E is a sample from a normal unrelated ewe, and --- is the blank. The gel shows a clear band at an appropriate size. The predicted size from the cattle genome is 213 bases. There are secondary bands which were minimised through PCR optimisation.

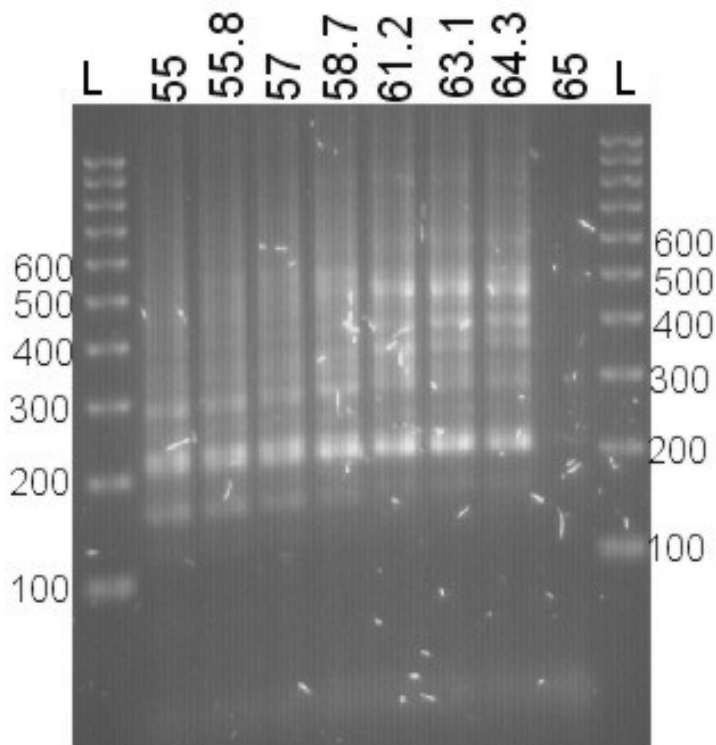


Figure 13: Temperature gradient PCR, t6s8, bovine DNA. Annealing temperatures are from 55 to 65 degrees. In this gel the template is the bovine DNA.

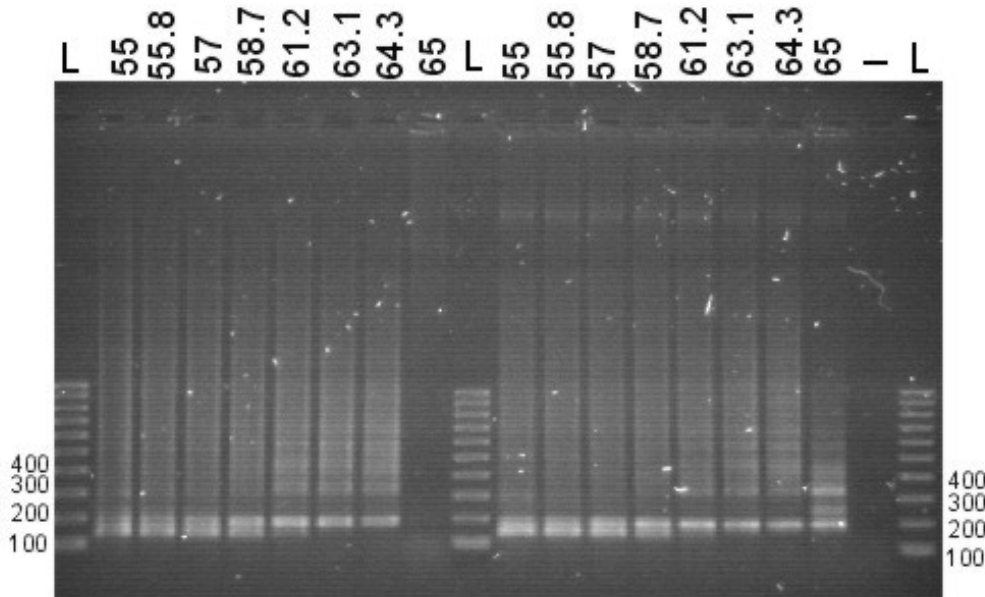


Figure 14: Temperature gradient PCR, t6s8, sheep DNA. Products from sheep A (ram 38/03) are shown on the left and products from sheep B (ram 102/03) on the right. Over both samples, the annealing temperature with the fewest extraneous bands is 57 degrees.

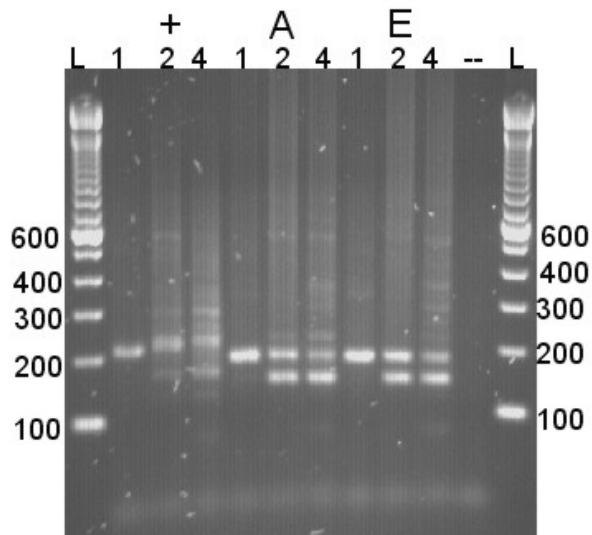


Figure 15: Magnesium gradient PCR for primer set t6s8. The concentration of magnesium varies from 1 to 4 mM, and the annealing temperature was 57 degrees. A concentration of 1mM gives a single clear band with all samples. The templates are bovine DNA (+), ram 38/03 DNA (A), and unrelated ewe DNA (E). The – symbol represent the blank.

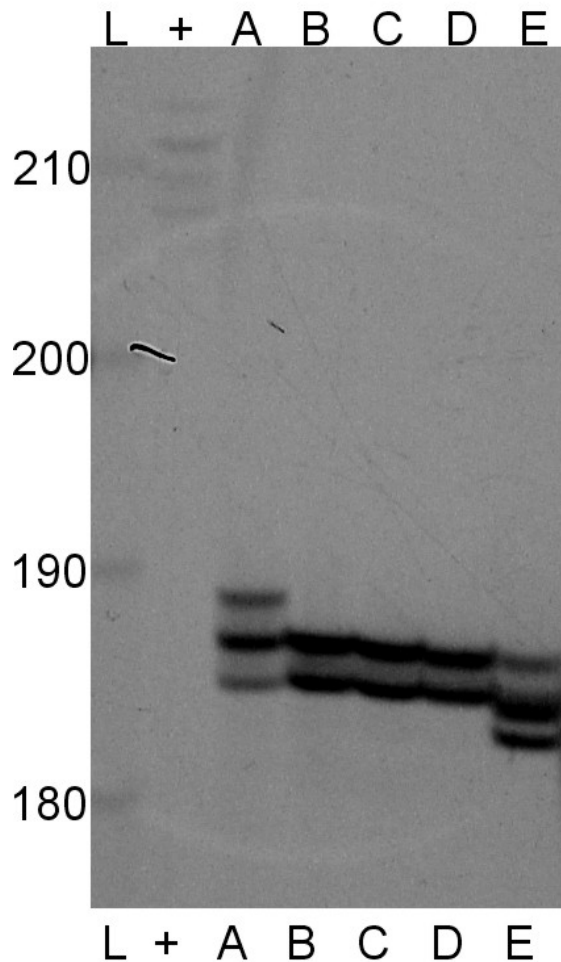


Figure 16: Acrylamide gel of t6s8 products. Lanes A, B, C, and D are cataract rams 38/03, 102/03, 37/02, and 36/04. The relationship between these rams is shown in Figure 8. Lanes + and E are bovine DNA and unrelated ewe DNA. This gel shows ram 38/03 and ewe E as heterozygotes with different genotypes, while the other rams are identical homozygotes. Single alleles are represented by pairs of bands on the gel. Because this marker shows variation among the sires it is likely to be informative. Therefore the marker was tested further by sequencing.

Markers t1s10, t2s8, t4s11, and t6s8 were successfully sequenced. When aligned against the equivalent sequences from the bovine genome, all showed at least 90% identity. When BLAST searches were carried out against the entire bovine genome, the sequences of t2s8, t4s11, and t6s8 had their best match with the appropriate region of cattle chromosome 6. The sequence of t1s10 had its best match with cattle chromosome 2. This showed that the t1s10 primers were amplifying a marker on chromosome 2 rather than the t1s10 marker on chromosome 6, so that set of primers was abandoned. The attempts at sequencing t3s9 and t3s19 were not successful, but they were used in the analysis anyway.

A new bovine genome assembly, BosTau4, became available online in October 2007 and the markers were located on it. Markers t3s9, t3s19, t4s11 and t6s8 were found within the linked region given by the previous analysis, so they were submitted to Invermay to test on the breakpoint panel. Marker t2s8 was found outside the linked region so it was not submitted. There was too much non-specific amplification in t3s19 and genotypes could not be generated from it, but genotypes were generated from the other three markers. The predicted positions of these three markers in relation to the existing markers is shown in Table 17.

Table 17: New marker positions. Positions are given in megabases from the centromere of bovine chromosome 6 on bovine genome assembly Btau4.0. New markers are shown in bold. Positions in red are calculated through interpolation using cM positions on sheep chromosome 6. The linked region given by the previous CRIMAP analysis is between BMS1510292 and BMS1512327.

Marker	Megabases
CP125	1.59
MCM53	18.17
AE101	37.27
JMP36	46.34
JMP1	59.48
BM4621	70.65
EL13	80.36
JMP4	90.16
BMS1509192	96.36
BMS1510292	98.74
BMS1509996	99.30
t4s11	102.95
t6s8	103.47
t3s9	105.77
BMS1512327	107.23
DK1183A	107.89
JMP8	108.58
BMS1512786	108.62
BMS1512958	108.83
BMS1513060	109.48
BMS1513944	118.88
BL1038	122.16

The genotypes of affected rams 14/97 and 9/99 were analysed to find any markers without common alleles between the two (Table 18). Ram 9/99 did not have a genotype for three of the markers, so one allele for each marker was reconstructed from the genotypes of his son, ram 37/02.

Table 18: Common alleles in founding sires, all three are affected. The two founding sires, 14/97 and 9/99, are shown along with 37/02, which is the son of 9/99. NUDT9 indicates 989G>A of the NUDT9 gene, as described in Chapter 4. All other markers are microsatellites, and the numbers show the PCR product size for each allele. Asterisks show where no genotype was available for 9/99 and one allele has been reconstructed from the genotype of the son. Alleles common to all three sheep are shown in blue. The alleles have been arranged so that common alleles are on the left, but the haplotype is unknown. The shaded markers are excluded from the LD region with OHC. Marker positions are given in cM from the centromere of sheep chromosome 6. Positions from independent mapping as described in section 3.2.3.2 are shown in black. Positions interpolated from Oar1.0 are shown in green. Positions interpolated from Btau4.0 are shown in red.

Marker	Position	14/97	9/99	37/02
CP125	2.60	127 138	127 133	127 133
MCM53	29.70	104 107	121*	121 121
AE101	49.90	117 117	135*	135 135
JMP36	62.20	167 167	159 165	159 165
JMP1	76.10	112 112	112 118	112 118
BM4621	88.10	168 181	158 179	179 181
EL13	100.10	160 165	160 172	160 172
JMP4	112.40	143 155	143 161	143 161
BMS1509192	119.69	180 171	180 180	180 180
BMS1510292	122.76	189 193	189 241	189 241
BMS1509996	123.46	199 199	199 199	199 199
t4s11	127.75	262 258	262 254	262 254
t6s8	128.06	211 211	211 213	211 213
t3s9	131.11	203 190	203 214	203 214
NUDT9	131.81	A G	A*	A A
BMS1512327	132.82	153 159	153 155	153 155
DK1183A	132.90	245 248	241 241	241 248
BMS1512958	134.42	243 241	243 245	243 241
BMS1512786	134.68	179 179	191 191	179 191
JMP8	134.70	151 147	151*	151 151
BMS1513944	146.28	177 177	177 189	177 189
BMS1513060	139.97	174 178	184 184	178 184
BL1038	151.20	134 134	116 130	130 134

There is a continuous region of markers with common alleles in all three sheep, running from EL13 to BMS1512327. Markers BM4621 and DK1183A do not have common alleles between the founding sires and their locations are the boundaries of the LD region with OHC. The LD region is between 101.10 and 132.82 cM from the centromere of sheep chromosome 6, a distance of 31.7 cM. This is more than twice the size of the region given by the previous CRIMAP analysis, 18.1 cM. A larger region is expected because only three animals were used in the analysis, rather than the entire pedigree.

The CRIMAP program gave noninheritance errors for 9 animals in the pedigree. For one animal the dam had been recorded incorrectly, and this was corrected. Five animals had

genotypes incompatible with their sires, so three were assigned to new sires and two were deleted as they did not match any sire. The remaining three animals did not seem to have any genotypes incompatible with their sires or dams. The loci that triggered the noninheritance errors were all set to “unknown”. When the CRIMAP analysis was repeated with the edited pedigree no noninheritance errors were given.

These errors occurred in a CRIMAP input file that was prepared from a spreadsheet of genotypes with corrected parentages, as described in section 3.2.3.2. The reassignments and deletions are in addition to those described in that section. The CRIMAP file was prepared from the same source as the analysis described in section 3.3.3.1, so the noninheritance errors described there are included in the errors described here.

The TWOPOINT option of CRIMAP was used to find the LOD score for linkage between the cataract locus and each marker, with the revised pedigree and marker order. Marker t3s9 had a LOD score of 3.69 and the NUDT9 polymorphism had a LOD score of 5.55, both indicating significant linkage to OHC. There were 40 informative meioses involving t3s9 and 52 involving NUDT9. Marker t4s11 had a LOD score of 0.08, which is not significant. There were 14 informative meioses involving t4s11. Marker t6s8 had a LOD score of 0.00, indicating that there was no evidence of linkage to the OHC locus even though the marker is within the linked region. There were only 4 informative meioses involving t6s8.

When the ALL option of CRIMAP was used to insert the OHC locus in an ordered list of all markers, using the revised marker positions and pedigree, the interval between t3s9 and NUDT9 was 27 times as likely to contain the locus as the next most likely interval (Table 19).

Table 19: Loglikelihoods of OHC positions, with revised pedigree and marker positions. The positions from 5' to 3' are labelled from 1 to 10 across the top of the table. “Loglike” stands for loglikelihood. The top position is about 27 times more likely than the positions second and third from the top.

1	2	3	4	5	6	7	8	9	10	Loglike
t6s8	t3s9	OHC	NUDT9	BMS1512327	DK1183A	BMS1512958	BMS1512786	JMP8	BMS1513060	-389.578
t6s8	t3s9	NUDT9	BMS1512327	OHC	DK1183A	BMS1512958	BMS1512786	JMP8	BMS1513060	-391.008
t6s8	t3s9	NUDT9	OHC	BMS1512327	DK1183A	BMS1512958	BMS1512786	JMP8	BMS1513060	-391.008
t6s8	t3s9	NUDT9	BMS1512327	DK1183A	OHC	BMS1512958	BMS1512786	JMP8	BMS1513060	-394.308
t6s8	t3s9	NUDT9	BMS1512327	DK1183A	BMS1512958	OHC	BMS1512786	JMP8	BMS1513060	-396.499
t6s8	t3s9	NUDT9	BMS1512327	DK1183A	BMS1512958	BMS1512786	JMP8	OHC	BMS1513060	-396.613
t6s8	OHC	t3s9	NUDT9	BMS1512327	DK1183A	BMS1512958	BMS1512786	JMP8	BMS1513060	-396.818

This interval is equivalent to between 111.4 and 112.0 megabases from the centromere of sheep chromosome 6, a distance of 0.6 megabases. In terms of linkage it is between 131.1 and 131.8 cM from the centromere, a distance of 0.7 cM. This interval includes AFF1 (Figure 21), which has been selected as a candidate gene for OHC as described in Chapter 4. Taking into account the loglikelihoods of the other positions, the OHC locus being 3' of marker t3s9 is about 37,000,000 times more likely than the locus being 5' of t3s9.

A multipoint analysis was performed using the revised pedigree and marker positions, giving a 1-LOD support interval of 131.3-131.7 cM from the centromere, with the highest LOD score at 131.5 cM (Figure 17). The width of the interval is 0.4 cM. This is the equivalent of 111.6 to 111.9 megabases from the centromere, with the highest LOD score at 111.7 megabases, a total distance of 0.3 megabases. The 1-LOD support interval generated in this analysis is consistent with the most likely linked region from the latest CRIMAP analysis, and also contains the candidate gene AFF1 (Figure 21).

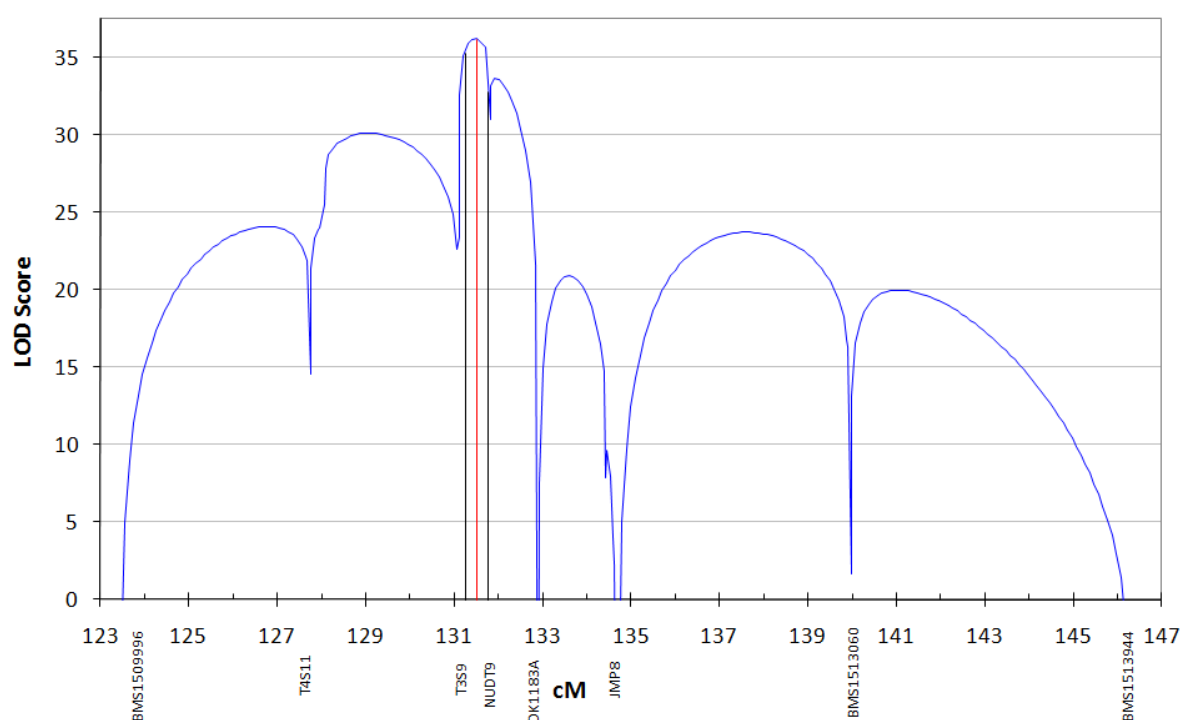


Figure 17: Third multipoint linkage analysis. This analysis is based on the genotypes of 11 markers in 296 animals, plus the genotypes of 12 more markers in 48 of the 296 animals. The revised pedigree and marker positions were used to prepare this graph. The two black lines indicate the 1-LOD support interval and the red line indicates the highest LOD score. Markers are indicated by vertical text along the bottom, and some have been omitted for clarity. Graph drawn by the author from analysis carried out by Dr Ken Dodds.

3.3.3.4 Gene Synteny

In order to determine the usefulness of the latest bovine genome assembly, the order of genes was compared between the real sheep genome assembly described in Section 1.6, the BTau.4.0 genome assembly, and the latest human genome assembly, hg19 (Figure 18).

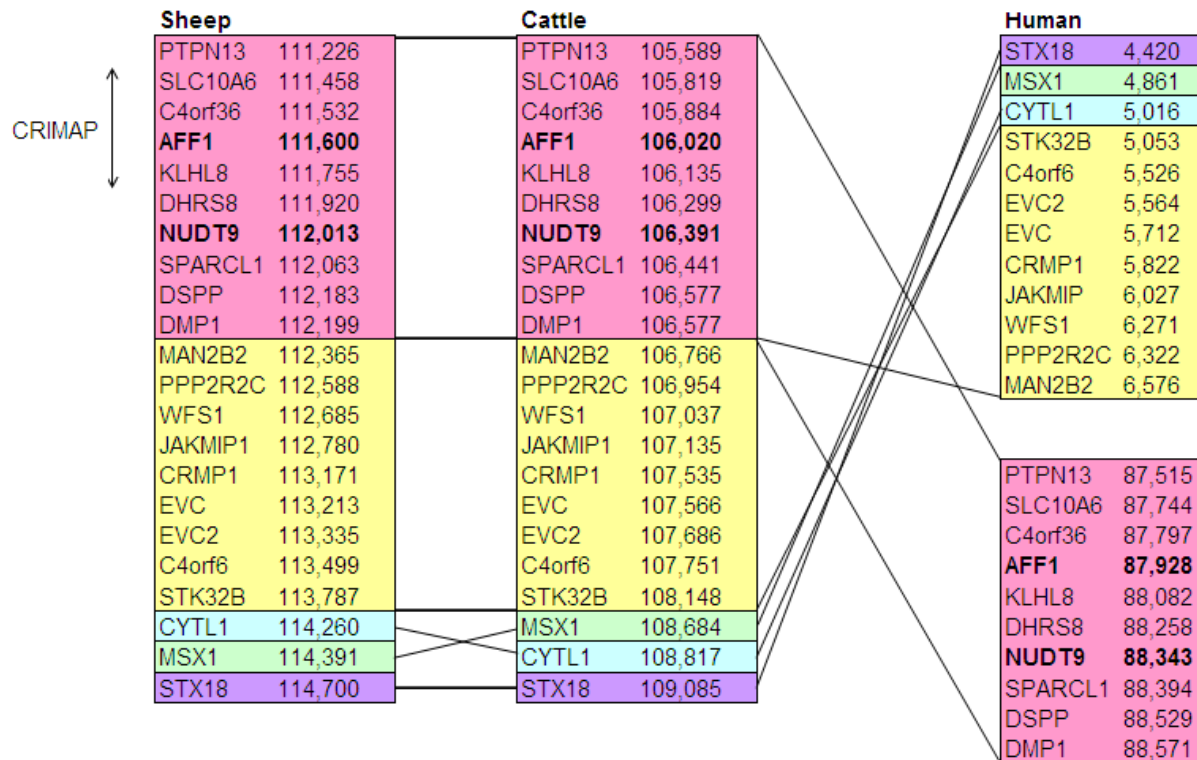


Figure 18: Sheep, cattle, and human gene positions. All positions are given in kilobases from the centromere of sheep chromosome 6, bovine chromosome 6, or human chromosome 4, which are all homologous. The assemblies that positions come from are Oar1.0 for the sheep, Btau4.0 for the cattle, and hg19 for the human. The two blocks of genes which are in the same order in cattle and sheep are highlighted in pink and yellow. Individual genes are also shown in the same colour in all three assemblies. The linked region from the CRIMAP analysis shown in Table 19 is indicated by the double-headed arrow labelled CRIMAP. The candidate genes AFF1 and NUDT9 are shown in bold.

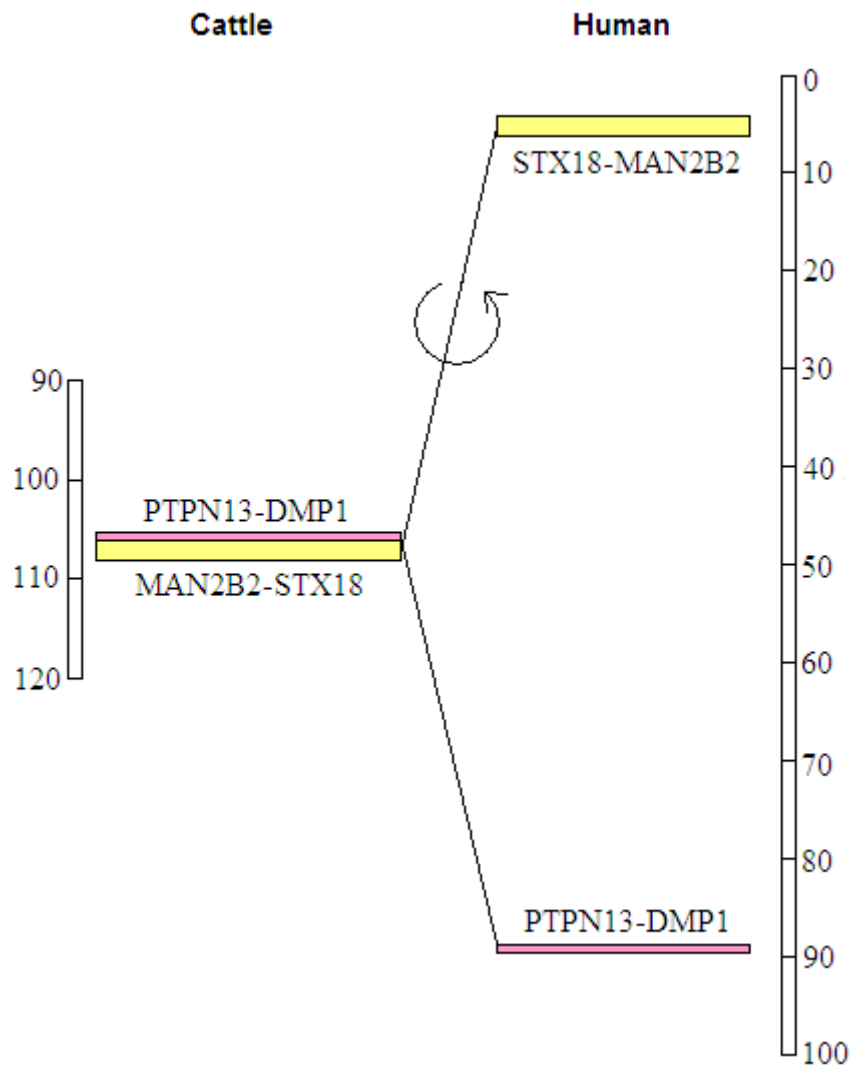


Figure 19: Cattle-human gene rearrangement. This scale diagram shows how the positions of two blocks of genes change between the cattle and human genomes. Changes in the position of individual genes are not shown. The yellow block is inverted between the cattle and human genome. The scale on the left shows megabases from the centromere of bovine chromosome 6, from assembly Btau4.0. The scale on the right is the equivalent for human chromosome 4, from assembly hg19. Both scales have the same relationship between distance and number of bases.

When the order of genes is compared between the sheep and bovine genome assemblies, two differences are found (Figure 18). Genes DSPP and DMP1 are separated by 16 kilobases in the sheep assembly, but in the bovine assembly their positions overlap. These genes are separated by about 42 kilobases in the human assembly, so the overlap is only present in the bovine assembly and is probably an error in the position of one of the genes. Also, the order of genes CYTL1 and MSX1 is reversed between the sheep and cattle assemblies. Taking into account the inversion described below, the order of the genes is the same in the human and sheep assemblies, with the bovine assembly as the outlier.

The continuous block of genes from MAN2B2 to STX18 in the bovine assembly is inverted and found closer to the centromere in the human genome assembly (Figure 19). The sheep assembly has this block in the same order as the bovine assembly with the exceptions noted above, so this is the only change where the human assembly is the outlier.

The region linked to cataract given by the latest CRIMAP analysis is not altered by any of these rearrangements. Therefore even if the rearrangements are due to errors in the bovine or ovine genome assemblies, they do not affect the outcome of the linkage analysis.

3.3.3.5 Breakpoint Panel Recombinations

The CHROMPIC option of CRIMAP was used to find recombinations in the 37 animals in the breakpoint panel (Figure 20).



Figure 20: Breakpoint panel recombinations, new markers. Markers are shown across the top. The four markers shown in red were used to select the breakpoint panel (Figure 9). The horizontal lines show which pair of markers the recombination is between for each animal. The red horizontal line shows a second recombination. The 37 animals in the breakpoint panel are listed along with their sires and grandsires, and dams and granddams are omitted. The colours indicate which generation each animal belongs to. The first generation is orange and the second, third, and fourth generations are yellow, green, and blue respectively. Offspring of each sire are listed below him in birth order. The cataract status of each animal is shown in Figure 9. The most likely position for the OHC locus is between the two markers shown in blue, t3s9 and NUDT9. All of the animals shown in this diagram are affected by OHC.

Of the 37 animals that had recombinations detected between EL13 and JMP8 when the breakpoint panel was selected, 33 show recombinations in the same region with the new

genotypes (Figure 20). The four animals with missing recombinations are discussed in the following section. No recombinations are between pairs of primers outside the JMP4-DK1183A interval, which is the CRIMAP linked region used to select the breakpoint panel.

3.3.4 Discussion

3.3.4.1 Breakpoint Panel

The 37 animals in the breakpoint panel were all selected because they had a recombination between markers EL13 and JMP8 (Figure 9). Of the 37 animals, 19 were selected because they originally had recombinations between JMP4 and DK1183A, the two markers that defined the most likely interval for the OHC locus. The other 18 animals were selected because they had recombinations between EL13 and DK1183A, or between JMP4 and JMP8, both of which could actually be between JMP4 and DK1183A. With the new genotypes no recombination is definitely outside the JMP4-DK1183A interval (Figure 20), so no member of the breakpoint panel can be eliminated for having a recombination in the wrong position and the choice to expand the selection interval was correct.

Four animals, 38/03, 123/04, 194/04, and 43/05, do not show recombinations in Figure 20 despite being selected for the breakpoint panel. Ram 38/03 does not show a recombination because the estimation of linkage phase has changed with the new genotypes. On 38/03's paternal chromosome, the EL13 allele was estimated by CRIMAP to come from 38/03's grandsire. On the same chromosome the DK1183A allele was estimated to come from the granddam. This suggested there was a recombination between EL13 and DK1183A, so 38/03 was included in the breakpoint panel.

With the new genotypes both the EL13 and DK1183A alleles are estimated to come from the granddam, so there is now no evidence of a recombination. There are nine other alleles between EL13 and DK1183A, and six are estimated to come from the granddam while the origin of three is unknown. This is further evidence against a recombination in this region of the chromosome.

Animals 123/04 and 194/04 were part of the breakpoint panel, but showed noninheritance errors when the pedigree file with the new genotypes was prepared, so they were left out of the analysis. To test the information provided by these animals, the genotypes of the markers causing the noninheritance errors were set to "unknown" and the CRIMAP analysis was

repeated with the ALL option. The loglikelihoods of the OHC locus positions (Table 19) all decreased by 5.3-5.5 due to the extra animals, but the positions themselves and their relative likelihoods were unchanged. This indicates that leaving the animals out did not significantly affect the results of the linkage analysis. As described in Section 3.3.2, animal 43/05 was not genotyped for the first group of new markers due to an error by AgResearch, so it was omitted from the analysis.

3.3.4.2 New Markers, Part One

Of the 63 cattle primer pairs tested in the first batch, 47 or 75% showed amplification in sheep. This indicates that about 75% of the cattle microsatellites were also present in sheep, which is consistent with other testing of cattle microsatellites in sheep (Crawford *et al*, 1995). However, only 41 or 65% of the primers showed amplification in cattle. Of the primer pairs that showed amplification, 18 had multiple bands and were abandoned.

For the first batch of markers, no attempt was made to optimise the PCR conditions to eliminate the multiple bands. Some of the markers that were rejected for having multiple bands in the first batch could be used for genome scanning after optimisation.

The PCR products from the next batch of markers were sequenced to confirm that the primers had amplified the correct sequences from sheep DNA (Section 3.3.3.3). This step was not taken for the first batch of markers, so the PCR products used to genotype animals in the linkage studies may not come from the markers that were selected, if the primer pairs were amplifying other regions of the sheep genome instead. This is possible because the primers are designed from bovine sequence and the homologous sheep sequence may be different enough that the primers may not bind to it effectively and also bind somewhere else. This would usually be visible on a gel as the two amplified regions would usually be of different sizes, giving two bands.

Some of the cases where multiple bands appear on the agarose gel (see the BMS1509593 bands in Figure 10) may be due to this situation. However, when a single intense band is present at the expected size (see the BMS1509791 bands in Figure 10) it is likely that the correct region is being amplified. This is because the correct region is likely to show some amplification, which would appear on a gel as a band at a different size to the incorrect region. A single band indicates all the amplification occurs at one region, which is probably

the correct one. Only markers with intense single bands at the appropriate size were used for genotyping in this section, so those primers are likely to be amplifying the correct regions of the genome.

The CRIMAP and multipoint analyses were performed on genotypes of all eight markers. Because only three of the markers were inside the linked region the most likely position of the OHC locus was only slightly changed, from between 112.3 and 132.9 cM to between 114.8 and 132.9 cM. The 1-LOD support given by the multipoint analysis has shifted from 123.3-128.8 cM to 123.7-129.0 cM, which is also a small change. These changes may not be meaningful, considering that both intervals are only the most likely positions given by the analysis and there is a significant probability of the locus being outside them. In order to significantly restrict the position of the OHC locus, more markers were selected using the Btau3.1 physical map and tested on the breakpoint panel, as described in the following section.

3.3.4.3 New Markers, Part Two

Four new markers were tested in the next batch. The NUDT9 polymorphism was tested on 238 animals and showed significant linkage to OHC. Marker t3s8 was only tested on the 37 breakpoint panel animals and 11 related animals, but still showed significant linkage to OHC.

T6s8 showed no linkage to cataract, but was only involved in 4 informative meioses. An informative meiosis occurs when genotypes of an animal's parents enable any recombinations to be detected. With a large enough number of informative meioses, the recombination fraction can be estimated and the degree of linkage calculated. With only 4 informative meioses, any linkage between T6S8 and the OHC locus cannot be detected.

T4s11 was involved in 14 informative meioses, but still showed no significant linkage to cataract, with a LOD score of only 0.08. The number of informative meioses may still be too low to detect significant linkage. As a comparison, the nearby marker BMS1509192 was involved in 10 informative meioses and had a LOD score of 0.00, despite being close to the linked region.

3.3.4.4 Linkage Analysis

Two CRIMAP analyses with the ALL option were carried out with genotypes from all markers. The first analysis used a pedigree with two errors in OHC status and an incorrect

marker order. The results of this analysis are not presented in this section, but it gave a linked region of 131.1-132.9 cM from the centromere of sheep chromosome 6. This region was used to select candidate genes for OHC as described in Chapter 4.

The two genes selected as candidates for OHC were AFF1 and NUDT9. A polymorphism was detected in NUDT9 but was ruled out as the OHC mutation. AFF1 was not completely sequenced and remains a strong candidate for the OHC gene (See Chapter 4).

After the pedigree and marker order was corrected, the CRIMAP analysis was repeated, including the NUDT9 polymorphism as a new marker. The results of this analysis are presented in the previous section. The new linked region was 131.1-131.8 cM, 1.1 cM narrower (Figure 21). The NUDT9 polymorphism is the new 3' boundary of the linked region, while the 5' boundary remains the same. The AFF1 gene stretches between 131.3 and 131.5 cM, so it still lies within the linked region despite being selected from the wider region (Figure 21).

The 1-LOD support interval given by the multipoint graph is 131.3-131.7 cM from the centromere. This lies within the most likely region given by the latest CRIMAP analysis, and is 0.4 cM wide compared to the 0.7 cM width of the CRIMAP region (Figure 21). AFF1 is also within this narrower interval, consistent with experimental evidence linking the mouse homologue to cataracts. The incomplete penetrance of OHC makes results from multipoint analyses less reliable, as described in section 3.2.4. Therefore this narrower region may be less reliable than the wider region from the CRIMAP analysis. Figure 22 summarises the results of all the linkage analysis performed in this investigation, including the preliminary results used to select the candidate genes.

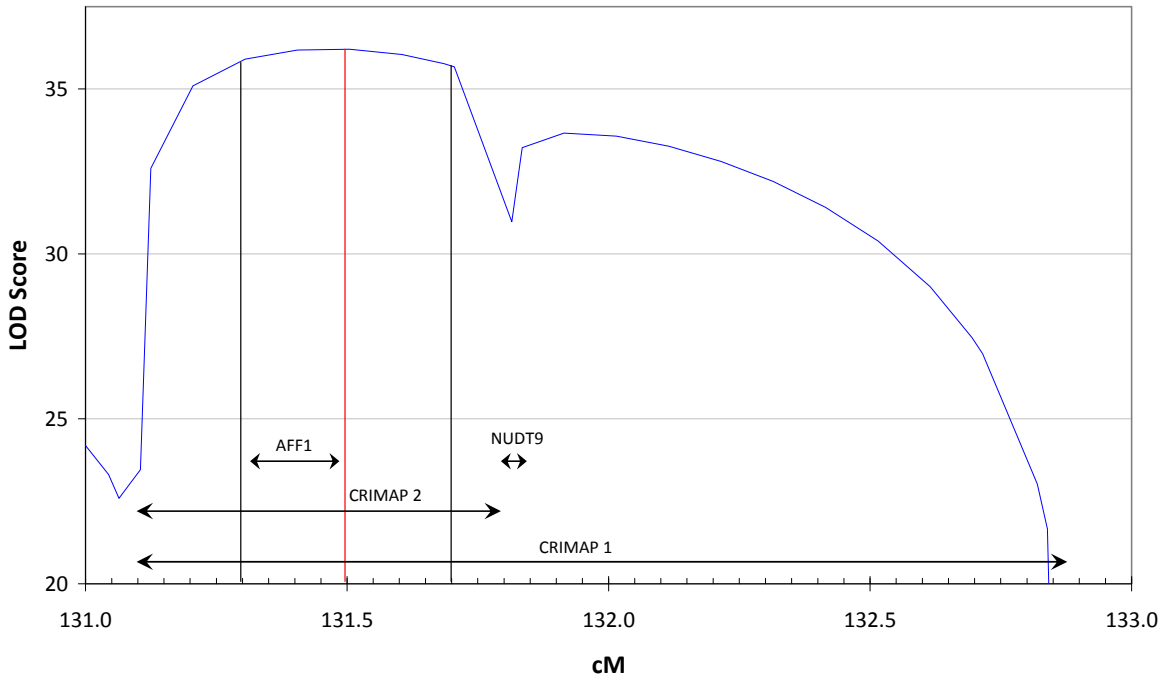


Figure 21: Close-up of Figure 17. The blue curve shows the LOD score. The two black vertical lines show the boundaries of the 1-LOD support interval, and the red vertical line shows the point with the highest LOD score. The double-headed arrow labelled CRIMAP 1 shows the linked region from the CRIMAP analysis that was used to select OHC candidate genes. The double-headed arrow labelled CRIMAP 2 shows the linked region from the latest CRIMAP analysis. The double-headed arrows labelled AFF1 and NUDT9 show the boundaries of those genes, taken from the Oar1.0 physical map and interpolated to the linkage map.

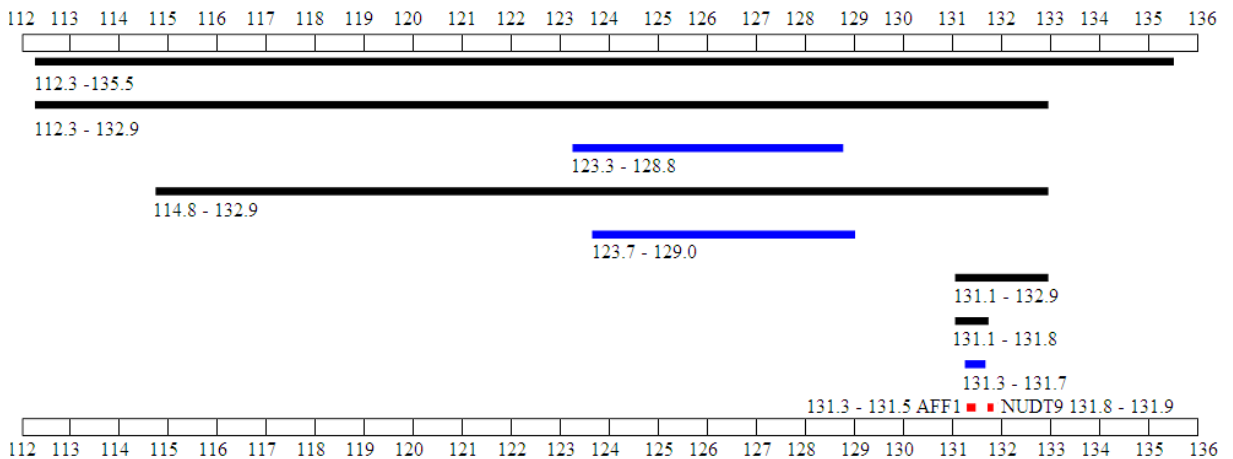


Figure 22: Genome scanning overview. The most likely intervals from the five CRIMAP analyses (black) and the 1-LOD support intervals from the three multipoint analyses (blue) are shown, in the order of generation from top to bottom. The positions of the candidate genes AFF1 and NUDT9 are shown in red below the last 1-LOD support interval. All positions are given in cM from the centromere of sheep chromosome 6, rounded to the nearest 0.1 cM.

As described in section 3.2.4, the incomplete penetrance of OHC was not taken into account when performing the linkage analysis in this chapter. This means that LOD scores for linkage between the cataract locus and each marker, the number of markers with significant linkage, and the size of the linked region will be underestimated (Clerget-Darpoux *et al*, 1986). This effect will be worse with the multipoint analysis (Risch and Giuffra, 1992).

3.3.4.5 Gene Synteny

The order of genes was compared between sheep, cattle, and human assemblies to determine whether there were any errors that would affect the results of the candidate gene searches or linkage analysis (Figure 18). In the linked region given by the latest CRIMAP analysis, the order of genes is consistent between sheep, cattle and human. The 1-LOD support interval is also inside this region.

In the wider linked region used to select the candidate genes (105.7 to 107.8 megabases from the centromere of bovine chromosome 6), the only inconsistency between the bovine and sheep assemblies is in the positions of DSPP and DMP1. These two genes are separate in the sheep and human assemblies but overlap in the bovine assembly. This error could not be caused by incorrect assembly of the contigs, because DSPP and DMP1 would be on the same contig in the bovine assembly if the overlap was correct. It is likely that the inconsistency is caused by an error in annotation which incorrectly assigned either DSPP or DMP1 to the position shown in Figure 18. There is an inversion between the human gene order and the sheep and cattle orders, but it is on a large enough scale to be the result of natural evolution. The order of genes CYTL1 and MSX1, which is consistent between human and sheep, is reversed in the bovine assembly. This change is on a small enough scale to indicate an error in the assembly of contigs, but occurs outside the region used to select candidate genes (Figure 18). Since there is no evidence of rearrangement in the region used to select candidate genes, the gene assemblies are reliable enough to make comparisons and predict the positions of genes between different genomes.

3.5 Conclusion

The most likely position for the OHC locus is between 111.4 and 112 megabases from the centromere of sheep chromosome 6. In terms of linkage this is between 131.1 and 131.8 cM from the centromere. This interval includes AFF1, which is a candidate gene for OHC. The interval is calculated using a model which assumes full penetrance. Because OHC shows incomplete penetrance, there is a significant probability that the OHC locus lies outside this interval.

Chapter 4 Candidate Genes

4.1 Introduction

This study follows a positional candidate strategy with comparative genomics to generate candidate genes for OHC. The genome scan and breakpoint panel analysis described in Chapter 3 was used to restrict the OHC locus to a narrow region of sheep chromosome 6. A database search was carried out on the latest bovine genome assembly (Btau4.0) to find bovine genes in the equivalent region of bovine chromosome 6. Positional candidate genes were selected on the basis of their function and any possible link to cataracts.

4.1.2 Criteria for Selection of Candidate Genes

As described in Chapter 3, significant linkage was found between the OHC locus and markers on sheep chromosome 6. This chromosome is equivalent to bovine chromosome 6 and human chromosome 4. Genome scanning revealed that the most probable location for the OHC gene was between markers t3s9 and DK1183A. On the latest assembly of the sheep genome this is equivalent to between 111.4 and 112.9 megabases from the centromere of chromosome 6, a distance of 1.5 megabases. Genes homologous to those causing cataracts in other organisms or where a plausible mechanism could be invoked for cataract formation are good candidates. As described in Chapter 3, errors in the marker order and OHC statuses used in the analysis were found, and the most likely region for the cataract locus was revised to between 111.4 to 112.0 megabases from the centromere. The search for candidate genes was carried out before the errors were discovered, so the older region was used.

4.2 Database Searches

4.2.1 Introduction

The bovine genome has been the subject of an intensive sequencing project, and as a result many more bovine genes are known than ovine genes. Because the bovine genome is so similar to the sheep genome, the latest bovine genome assembly was used to select candidate genes rather than any sheep assembly. First the bovine equivalents of t3s9 and DK1183A were found on the bovine genome using a BLAST search and a text search on a genome viewer. One of the marker sequences originally came from the bovine genome but refinements in the assembly had made the old position unreliable and it was rechecked. The positions of these markers on the bovine genome were recorded, and a list of bovine genes between these positions was generated using a genome assembly viewer. When the linked region was narrowed after a repeat of the CRIMAP analysis, the list of bovine genes was revised.

4.2.2 Methods

A 658-base bovine sequence that included the microsatellite t3s9 was used in a BLAST search of the latest bovine genome assembly, Btau4.0. The search was carried out online at <http://www.ncbi.nlm.nih.gov/genome/seq/BLASTGen/BLASTGen.cgi?taxid=9913>. The database was genome (reference only), the program was megaBLAST and the expectation filter was 0.01. The BLAST hit was located on the bovine genome using the NCBI map viewer.

Marker DK1183A is located inside the gene EVC2 in sheep. A text search was carried out on the bovine genome viewer at <http://genome.ucsc.edu/cgi-bin/hgGateway?hgsid=107188599&clade=vertebrate&org=Cow&db=0>, using the string “EVC2”, in order to find the equivalent position of the DK1183A marker in the bovine genome. The genome viewer was set to display results only from Btau4.0.

The sequences of eight other markers with known positions on the sheep genome surrounding DK1183A were used in BLAST searches on the bovine genome. The positions of the markers on the bovine genome were recorded and used to construct a graph relating sheep and

bovine genome positions with Microsoft Excel. A linear trendline was used to interpolate a position on the bovine genome for DK1183A, to confirm that it was located within EVC2.

The online bovine genome viewer was used to display all known genes between the two positions found in the searches. This included all matches between bovine sequence and gene sequences from humans and other mammals, including hypothetical genes. Therefore the set of genes generated by the search represents the best available information about genes in that interval and it is unlikely that there are any other genes inside it. All searches were performed on the 15th of March, 2009. The positions of the genes were rechecked on the first of October of the same year. The source of the information was the “Cow Oct. 2007 Assembly”, which is the latest assembly, Btau 4.0. The bovine genes displayed in the genome viewer had previously been identified through chained BLAST searches against the human genome sequence. Genes with higher-scoring matches on chromosomes other than 6 were eliminated from the list.

When the CRIMAP analysis was repeated with corrected OHC statuses and marker order, the end of the linked region was the polymorphism in NUDT9, instead of marker DK1183A. The start of the linked region was still t3s9. Therefore the genes inside the new linked region are a subset of the genes inside the old linked region, including every gene before NUDT9.

4.2.3 Results

BLAST searches were carried out using a 658-base bovine sequence surrounding the t3s9 marker. The NCBI map viewer showed that the BLAST matches were located between 105,767,511 and 105,768,152 bases from the centromere of bovine chromosome 6 (Figure 23).

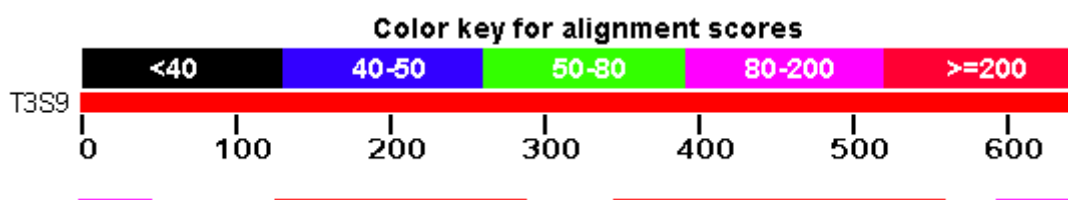


Figure 23: BLAST matches to the t3s9 sequence. All matches were found in the *Bos taurus* chromosome 6 genomic contig NW 001495231.2. This contig is based on the then latest bovine genome assembly for the 15th of March, 2009, BTau_4.0.

A text search of the bovine genome viewer showed that the EVC2 gene is between 107,686,099 and 107,864,164 bases from the centromere of bovine chromosome 6.

To confirm that DK1183A is within the EVC2 gene the equivalent position of DK1183A was found on bovine chromosome 6 by interpolation. This showed that the position of DK1183A is about 107.6 megabases from the centromere of bovine chromosome 6, probably within the EVC2 gene. Thus, EVC2 can be used as the 3' boundary of the linked region.

From the positions found above, the region of the bovine genome equivalent to the linked region of the sheep genome is from 105.7 to 107.9 megabases from the centromere of bovine chromosome 6. The bovine genome viewer displayed 16 genes in this interval (Table 20).

Table 20: Genes in the OHC-linked region of the bovine genome. Positions are given in megabases from the centromere on bovine chromosome 6, from Btau4.0. The entire table shows the sixteen genes in the earlier linked region. The genes highlighted in yellow are present in the linked region given by the latest CRIMAP analysis and shown in Table 19. Protein functions and references are given where available. The full names of each gene are:

PTPN13 protein tyrosine phosphatase, non-receptor type 13

SLC10A6 solute carrier family 10 (sodium/bile acid cotransporter family), member 6

MGC26744 mammalian gene collection 26744

AFF1 affranchised 1

KLHL8 kelch-like 8

DHRS8 hydroxysteroid (17-beta) dehydrogenase

NUDT9 nudix (nucleoside diphosphate linked moiety X)-type motif 9

SPARCL1 SPARC-like 1 (hevin)

DSPP dentin sialophosphoprotein

DMP1 dentin matrix acidic phosphoprotein 1

MAN2B2 mannosidase, alpha, class 2B, member 2

PPP2R2C protein phosphatase 2, regulatory subunit B, gamma

WFS1 Wolfram syndrome 1 (wolframin)

JAKMIP2 janus kinase and microtubule interacting protein 2

FLJ4648 full-length long Japan 4648

CRMP-1 collapsin response mediator protein 1

EVC Ellis van Creveld syndrome

EVC2 Ellis van Creveld syndrome 2

Gene	Protein Function	Reference	Megabases
PTPN13	Protein tyrosine phosphatase	Sato 1995	105.59-105.81
SLC10A6	Sodium-dependent organic ion transporter	Geyer 2004	105.82-105.85
MGC26744	Hypothetical gene, unknown function	Gerhard 2004	105.88-105.89
AFF1	Transcription factor, cataract	Isaacs 2003	106.02-106.12
KLHL8	Ubiquitination	Nam 2009	106.14-106.15
DHRS8	Androsterone synthesis	Chai 2003	106.30-106.35
NUDT9	Hydrolyses ADP-ribose	Perraud 2002	106.39-106.43
SPARCL1	Cell adhesion, migration, proliferation	Sullivan 2004	106.44-106.49
DSPP	Dentine extracellular matrix	Gu 2000	106.58-106.59
DMP1	Transcription factor, bone and teeth matrix	Narayanan 2003	106.58-106.59
MAN2B2	Mannosidase	Park 2005	106.73-106.78
PPP2R2C	Serine/threonine protein phosphatase	Eichhorn 2007	106.95-107.01
WFS1	Transmembrane protein, Wolfram syndrome	Khanim 2001	107.04-107.05
JAKMIP2	Unknown, binds microtubules	Steindler 2004	107.14-107.26
FLJ4648	Hypothetical gene, unknown function		107.39-107.41
CRMP-1	Axonal guidance, neuronal differentiation	Shih 2003	107.50-107.56
EVC	Unknown, Ellis-van Creveld syndrome	Ruiz-Perez 2003	107.57-107.67
EVC2	Unknown, Ellis-van Creveld syndrome	Ruiz-Perez 2003	107.69-107.86

After NUDT9 was evaluated as a candidate for OHC, the CRIMAP analysis was repeated using NUDT9 as a new marker, a new order of the other markers, and two corrected OHC statuses. The new linked region was between t3s9 and NUDT9, and the genes within it are highlighted in yellow in Table 20.

4.2.4 Discussion

Of the genes found in the previous section, only two are obvious candidates for the OHC gene. AFF1 is a transcription factor that is known to cause leukemia when chromosome translocations fuse it to another gene (Bizarro *et al*, 2007). A mutation in AFF1 causes cataracts and ataxia in mice (Isaacs *et al*, 2003). The cataract is inherited in an autosomal dominant manner, just as in OHC. Therefore AFF1 is a viable candidate gene for OHC. However, sheep affected by OHC show no signs of ataxia. This means that OHC is significantly different from the mouse cataract condition and is evidence against AFF1 being the OHC gene.

NUDT9 is an enzyme that hydrolyses ADP-ribose into AMP and ribose-6-phosphate (Perraud *et al*, 2002). A calcium channel protein with significant similarities to NUDT9 is activated by ADP-ribose. It is possible that a mutation in NUDT9 could cause a build up of ADP-ribose, leading to inappropriate activation of the calcium channel and an influx of calcium ions. The increased calcium concentration could overactivate calpain and lead to degradation of lens proteins and cataract. Therefore NUDT9 is also a viable candidate gene for OHC. However there is no known association between mutations in NUDT9 homologues and cataract in any animal. Also, OHC is inherited in a dominant manner and is less likely to be caused by a mutation in the gene for an enzyme (see Section 4.3.1.1). This means that NUDT9 is not as strong a candidate for the OHC gene as AFF1 is.

No homologues of any of the other genes are known to be associated with cataracts in any species. The functions of each of the other genes, where they are known, are briefly summarised in Table 20. There is no obvious mechanism for cataract formation associated with most of these genes, but several have functions which could lead to cataracts if disrupted.

WFS1 codes for a transmembrane protein which regulates cellular calcium concentrations (Takei *et al*, 2006). Mutations in this gene cause Wolfram Syndrome in humans, a recessive condition which is characterised by deafness, diabetes, and optic atrophy (Khanim *et al*, 2001). Cataracts have been reported in Wolfram Syndrome (Al-Till *et al*, 2002), but only as complications of diabetes.

OHC is thought to be caused by increased calcium concentrations leading to overactivation of calpain and proteolysis of lens proteins. Therefore WFS1 could be the OHC gene, with a

mutation leading to increased intracellular calcium. However, no homologue of WFS1 is known to be associated with a cataract disease in any species, since Wolfram Syndrome usually does not include cataracts. Most Wolfram Syndrome patients have normal lenses despite not having any functional copies of WFS1 atrophy (Khanim *et al*, 2001), which suggests it is not required for lens transparency.

SPARCL1 codes for a protein involved in cell adhesion, migration, and proliferation (Sullivan *et al*, 2004). If SPARCL1 is necessary for adhesion between lens fibres, it is possible that OHC is caused by a mutation in SPARCL1 disrupting the lens fibres.

KLHL8 codes for a protein involved in ubiquitination, a process where proteins are tagged for transport to the proteasome where they are degraded (Nam *et al*, 2009). Ubiquitination is part of the lens differentiation process, as proteins are degraded to produce lens fibres. Mutations in KLHL8 interfering with ubiquitination and leading to opacities from leftover proteins in the lens is a possible mechanism for cataract formation. However this does not seem to be the mechanism in OHC, as the opacities seem to be due to proteolysis of lens proteins rather than improper differentiation of the lens fibres. Therefore KLHL8 is unlikely to be the gene for OHC unless ubiquitination is involved in cataract formation in OHC, which there is no evidence of at present.

PPP2R2C codes for a regulatory subunit of protein phosphatase 2 (Eichhorn *et al*, 2007). This enzyme is a serine/threonine phosphatase that removes phosphate groups from proteins. It has diverse roles in signal transduction, apoptosis and cell division (Torres *et al*, 2010). The subunit coded for by PPP2R2C itself may modulate substrate selectivity or catalytic activity. Since protein phosphatase 2 is involved with many different pathways, it is possible that it is also involved in a pathway that is necessary for lens transparency and that PPP2R2C is the OHC gene. However there is no evidence of this and homologues of PPP2R2C are not known to be associated with cataracts in any species.

EVC and EVC2 are genes of unknown function that are associated with Ellis-van Creveld syndrome (Ruiz-Perez *et al*, 2003). Both of their sequences contain domains which indicate that they code for transmembrane proteins. If EVC or EVC2 are involved in maintaining osmotic balance, cell adhesion or calcium concentrations in lens fibres, a mutation in one of these genes could disrupt the fibres and cause cataract.

PTPN13 codes for a protein tyrosine phosphatase that removes phosphate groups from the Fas receptor, inhibiting apoptosis (Sato *et al*, 1995). Lens differentiation involves the degradation of organelles in a process similar to apoptosis, but it is not identical to apoptosis and is not known to involve the Fas receptor. Therefore the inhibition of apoptosis by PTPN13 is unlikely to be relevant to OHC, and PTPN13 is unlikely to be the gene for OHC unless it is involved in other metabolic pathways.

DMP1 codes for a transcription factor which regulates gene expression in osteoblasts (Narayanan *et al*, 2003). It is also found in the extracellular matrix of bone and teeth where it regulates mineralisation. DMP1 is unlikely to be the OHC gene unless it acts as a transcription factor in the lens in a metabolic pathway that is necessary for lens transparency.

CRMP-1 codes for a signalling protein expressed mainly in the nervous system and involved in axonal guidance and neuronal differentiation (Shih *et al*, 2003). The expression of CRMP-1 in cancer cells inhibits their invasive ability (Shih *et al*, 2003), and the CRMP-1 protein may also be involved in regulating the cell cycle. None of these functions have an obvious relevance to OHC. Disruption of the cell cycle might interfere with the differentiation of lens cells and lead to cataract, but this does not appear to be the mechanism of cataract formation in OHC.

Because CRMP-1 codes for a signalling protein involved in multiple metabolic pathways, it is possible that this gene is involved in another pathway which is relevant to OHC. In that case CRMP-1 would be a strong candidate for the OHC gene, but there is no evidence of that at present.

JAKMIP2 codes for a protein which binds to both microtubules and Janus kinases, which are enzymes involved in cytokine signalling (Steindler *et al*, 2004). It is part of a family of three genes, JAKMIP1-3, whose functions are poorly understood. They may be involved in microtubule transport and cytoskeleton rearrangements. It is possible that JAKMIP2 is involved in a pathway that is relevant to OHC, but there is no evidence of that at present.

DSPP, DHRS8, MAN2B2, and SLC10A6 code for proteins with functions that are unlikely to contribute to OHC. DSPP codes for a dentine protein (Gu *et al*, 2000), DHRS8 codes for an enzyme involved in androsterone synthesis (Chai *et al*, 2003), MAN2B2 codes for a

mannosidase, an enzyme which breaks down oligosaccharides (Park *et al*, 2005), and SLC10A86 codes for a protein that transports large organic ions, including steroid precursors (Geyer *et al*, 2004). Neither DSPP, DHRS8, MAN2B2, or SLC10A6 are likely to be the OHC gene.

Both MGC26744 (Gerhard *et al*, 2004) and FLJ4648 are hypothetical genes of unknown function. While it is possible that one of these genes is the OHC gene, they cannot be evaluated as candidates until information on their function becomes available.

Further testing ruled out NUDT9 as the OHC gene (Section 4.3), but this gene contained a polymorphism that was used as a marker in a CRIMAP analysis. The analysis gave a narrower linked region that contained only the genes PTPN13, SLC10A6, MGC26744, AFF1, KLHL8, and DHRS8. Since OHC shows incomplete penetrance which is not taken account of in the linkage analysis, the size of the linked region may be underestimated. Therefore the previous, larger interval may actually be more useful for selecting candidate genes.

As described above, the only gene in the narrower region with a known link to cataracts is AFF1. Since AFF1 is associated with dominant cataracts in mice and is located inside both the narrower linked region and the 1-LOD support interval from the most recent multipoint analysis (Figure 21), it is the strongest candidate gene for OHC.

4.3 NUDT9

4.3.1 Introduction

The purpose of this section is to determine whether or not there is a mutation in NUDT9 which is strongly linked to OHC status. If there is, this would suggest that NUDT9 is likely to be the gene responsible for OHC. Mutations were detected by sequencing the CDS in both affected and normal animals, and then finding polymorphisms that correlate with OHC status.

The bovine NUDT9 CDS was used to design primers that would amplify a section of the sequence in sheep. PCRs were carried out using these primers and cDNA was generated from normal cultured sheep lenses. The PCR products were sequenced and compared to bovine NUDT9 sequences to confirm that NUDT9 is expressed in normal cultured lenses, and that the primers are capable of amplifying it.

The sheep sequences generated with the first pair of primers were used to design additional primers that amplified the entire NUDT9 CDS. The primers were used in PCRs with cDNA from normal and affected animals from the cataract flock. PCR products were sequenced and compared between affected and normal samples to identify polymorphisms that were candidates for the OHC mutation.

A combination of PCR, sequencing and restriction digest analysis was used to genotype many animals for the polymorphisms. The genotypes were studied to determine whether the polymorphism was consistently correlated with OHC status.

4.3.3 Methods

4.3.3.1 Extraction of RNA from Normal Lenses

RNA was extracted from cultured sheep lenses and used as the template for PCRs with primers designed to amplify NUDT9. Cultured lenses were used because a study had been carried out investigating gene expression in cultured lenses with induced opacities. RNA had been extracted from control lenses with no opacities and was available for further analysis.

Eyes were extracted by the author from normal lambs unrelated to the cataract lambs, within 1 h of slaughtering. No more than 4 h later, the eyes were dissected to remove the lenses. The lenses were placed in 10mL of Eagle's minimum essential medium containing 0.02mg/mL

getamycin, 2.5µg/mL amphotericin, and 26mM NaHCO₃. After all the dissections were finished, the lenses were transferred to warm sterile medium containing APS (amphotericin (0.25µg/mL), penicillin (100units/mL), and streptomycin (100µg/mL)). The lenses were incubated at room temperature for 30 min, and then transferred to culture medium without APS, one lens to 10mL. After culture for 1-3 days at 35°C, lenses were inspected and any with opacities discarded. The lenses were then transferred to artificial aqueous humour (AAH), each lens in 10mL of AAH, and cultured 35°C, 7 h. For the two lenses (from two different animals) used in this section, no opacity-inducing substances were added to the culture medium.

Lenses were removed from the culture medium and rolled on clean filter paper to remove fragments of the iris. The epithelia of the lenses were dissected out with a scalpel and tweezers. The epithelia were placed in a microcentrifuge tube, frozen with liquid nitrogen and stored at -80°C.

The frozen epithelium was pulverised with a micropestle and 350µL RLT lysis buffer from a Qiagen RNeasy Minikit (Qiagen GmbH, D-40724 Hilden, Germany) was added. The contents were homogenised through a 20 gauge needle until no large clumps remained, and RNA extraction was continued according to the manufacturers' instructions, eluting the RNA with 50µL of deionised water. A 2µL sample of each RNA extract was visualised on an agarose gel with ethidium bromide and UV light to check quality, and UV absorbances of the extracts at 260 and 280 nm were taken to determine concentration and purity. The RNA extracts were frozen at -80°C. This RNA was used as a representative of normal gene expression in the ovine lens. The RNA extract was run on a 2% agarose gel containing ethidium bromide to check quality and concentration before use.

4.3.3.2 Expression of NUDT9 in Normal Lenses

A 10.5 µL sample of the RNA extract from the animal designated "L" (see Figure 25) was used to prepare cDNA with Superscript III reverse transcriptase (Invitrogen). The oligo dT primer provided with the enzyme was used in order to produce cDNA from all mRNA sequences present in the RNA sample.

The cattle CDS of the NUDT9 gene (from the NCBI database at <http://www.ncbi.nlm.nih.gov/nuccore> , accession number BC151586) was used to design a

pair of primers to amplify most of the sheep CDS, using the online Repeatmasker (Tarailo-Graovac *et al*, 2009) and Primer3 (Rozen and Skaletsky, 2000) programs (Table 21).

Table 21: NUDT9 primers. The product size predicted from the bovine sequence is 996bp.

Name	Sequence	T _m (°C)
NUDT9F	TCCTGGGGAAGACTTTAGCC	60.6
NUDT9R	CATCTCGTTCTCTGCCACA	60.0

PCR was carried out using the cDNA as a template. Each PCR reaction mixture consisted of a 1:20 dilution of cDNA, 125nM dNTP mix (Invitrogen), 250nM of each primer, 1X Q-solution (Qiagen), 1mM MgCl₂, and 0.25 units per µL of *Taq* DNA polymerase (Qiagen). The final volume of each mixture was 20 µL.

The PCR cycle involved an initial denaturation, 94°C 10 min, followed by 35 cycles of 94°C 30 s, 54°C 30 s, 72°C 30 s. There was a final elongation step, 72°C 10 min.

The entire PCR product from the reactions described above were mixed with loading dye and loaded onto 2% agarose gels. The gels were run at 90V, 1 h. The bands, one per PCR product, were cut out and extracted with an Axyprep DNA gel extraction kit (Axygen Biosciences, 33210 Central Avenue, Union City, California 94587, USA). The concentration of the gel extracts were determined by running samples on agarose gels, along with genomic DNA samples of known concentration.

These gel extracts were used as templates in re-amplification PCRs. The recipe was identical to the previous PCRs except that 1 µL of gel extract was used instead of cDNA. The PCR cycle had an annealing temperature of 57°C rather than 54°C but was otherwise identical to the previous PCR. The PCR products were purified on agarose gels as described above. The gel extracts from the re-amplification PCR were used to prepare two sequencing mixtures, one with the forward primer and one with the reverse primer, as described in Section 3.4.2. The mixtures were sequenced by Massey University as described in the same section, and the sequences were aligned with the cattle NUDT9 CDS using the DNAMAN (Lynnon Corporation, 383 av St-Louis, Pointe-Claire, Quebec, Canada) program. The two sequences did not overlap, due to the length of the CDS.

4.3.3.3 Sequencing

The two new sequences generated in the previous section were used to design a pair of primers to amplify the gap between them, named GapF and GapR. The sequence generated from these primers was used in a BLAST search to find sheep Expressed Sequence Tags (ESTs) that covered areas of the sheep NUDT9 CDS not amplified by the existing primers. Not enough sheep sequence was available for a BLAST search before this sequence was amplified.

The ESTs EE803518 and FE029171 were found on the NCBI database at <http://www.ncbi.nlm.nih.gov/nucest>. The ESTs covered the 5' and 3' ends, respectively, of the NUDT9 CDS not covered by existing sequences, and were used to design primers to amplify those regions. The 5' primers were named StartF and StartR and the 3' primers were named EndF and EndR. The three pairs of primers (Table 22) were used to amplify and sequence the NUDT9 cDNA from the RNA from two cultured lenses.

Table 22: NUDT9 sequencing primers. The PCR product sizes predicted from bovine sequence are also shown. The positions of their primers on the CDS are shown in Figure.

Name	Sequence	T _m (°C)	Size (bp)
StartF	GCGGCTCAAGCTAAGAGCTA	60.0	283
StartR	TGCGCTGAACTTTTGAACCT	60.9	
GapF	TTTCTGTCTTGGCTGGACCT	59.8	378
GapR	TCTTTTCAGTGTGGCGCTAA	59.6	
EndF	ATGCTGGAAAGGTGAAATGG	59.9	168
EndR	GGCTTTTGGCTTACATGGAG	59.7	

All PCRs were carried out with appropriate blank samples for each pair of primers. The blanks were run alongside the other samples on agarose gels to check for contamination. For some PCR reactions the gels were repeated without blanks in order to show more PCR products on a single gel.

Three affected (20/07, 47/07, and 56/07) and three normal sheep (37/07, 45/07, and 50/07) were selected from the OHC flock, from cataract statuses previously determined as described in section 2.3.1. All six animals were half-siblings, the offspring of one affected ram and six different affected ewes. RNA was extracted from venopuncture blood samples using a PureLink Total Blood RNA Purification Kit (Invitrogen). Samples of the RNA were run on

agarose gels to determine the approximate concentration before the cDNA was prepared and the RNA extracts were used to prepare cDNA as described above.

PCRs were carried out using the cDNA as a template, using the three pairs of primers described above. The PCR recipes and cycles were as described in Section 4.3.3.2 except that the annealing temperature was 65°C. The reaction mixtures were not inserted into the PCR cyclers until the temperature had reached 80°C. This hot-start technique ensured that the Taq enzyme did not start elongation until the template was completely denatured, reducing non-specific amplification. The concentration of the PCR products was determined as described above.

Sequencing mixtures were prepared as above, except that the PCR products were purified using an AxyPrep PCR Cleanup Kit (Axygen), and the purified products sequenced as described above. Each sequence was aligned with an assembly consisting of sheep ESTs and a short sheep sequence generated from cultured lens RNA. The degree of homology between the sequence and the assembly showed the quality of the sequence.

Consensus sequences were assembled for each animal from the six sequences generated from their cDNA. Any ambiguous bases were resolved by examining the electropherograms for each sequence, and also previously generated sequences from the same animal. Sequences which appeared to be genuine heterozygotes were indicated with a two-base ambiguity code. The six consensus sequences, one from each animal, were aligned with the bovine CDS using Genedoc (Nicholas *et al*, 1997). The sequences were arranged in normal and affected groups and were examined to detect any differences which correlated with OHC status.

4.3.3.4 Genotyping

The normal sheep NUDT9 sequence generated above was used to design primers to amplify both position 855 and 989 from cDNA (Table 23), which both showed polymorphisms that segregated with OHC status. The online Primer3 program (Rozen and Skaletsky, 2000) was used, and the product was limited to shorter than 300 bases to enable the entire product to be sequenced in one step.

Table 23: NUDT9 polymorphism testing primers. The PCR product is predicted to be 277 bases from the bovine sequence.

Name	Sequence	T _m (°C)
NUDT9BLDF	GCTCTTCAGCCAGGAACATC	60.0
NUDT9BLDR	GCAGCCAGTCTCAGAGTCCT	59.7

Nine more sheep, including 5 affected (5/07, 21/07, 46/07, 53/07, and 61/07) and 4 normal animals (7/07, 54/07, 59/07 and 60/07), were selected from the OHC flock. The 4 normal animals could may have the OHC genotype and have normal lenses due to incomplete penetrance. Blood RNA was extracted and used to prepare cDNA as described above. PCRs were carried out using cDNA from 21 animals, including the 6 original animals, the 9 new animals, and cultured lens cDNA from 6 animals from a commercial meat flock. This included the “L” lens cDNA used previously. The PCR recipes and cycles were as described in Section 4.3.3.2, except that the annealing temperature was 60°C. For each PCR product, a 2 µL sample was run on an agarose gel to check that the amplification was sufficient and the band size was as expected. The PCR products were sequenced as above.

Primers were designed to amplify the position 989 allele of NUDT9 from DNA extracts. The optimal reverse primer was identical to NUDT9BLDR but was still named NUDT9DNA2R. PCRs were carried out using DNA extracts from affected and normal sheep (Table 24).

Table 24: NUDT9 DNA extract primers. The product size predicted from bovine sequences is 151bp. NUDT9DNA2R is identical to NUDT9BLDR.

Name	Sequence	T _m (°C)
NUDT9DNA2F	CCTTGCCTTAGAAGCTGGAG	59.2
NUDT9DNA2R	GCAGCCAGTCTCAGAGTCCT	59.7

The PCR recipes and cycles were as described above, except that the annealing temperature was 59°C. A 2µL sample of each PCR product was run on an agarose gel to check that the amplification was sufficient and the band size was as expected. For 9 samples the PCR products were sequenced as described above, using the primers in Table 24.

It was determined that the restriction enzyme HpyCh4IV (New England Biolabs, 240 County Road, Ipswich MA 01938-2723, USA) would cut the PCR product once, only when the position 989 base was G. The recognition site is shown in Figure 24.

**A!CGT
TGC!A**

Figure 24: Recognition site of HpyCh4IV. Exclamation points show the cleavage sites.

The 151-base PCR product from the DNA extract primers would be cut into a 97-base fragment and a 54-base fragment when the base is G, and remain intact when the base is A.

Restriction digests were performed on 5 μ L of each PCR product. Each digest mixture contained 2.5 μ L of 10X buffer, 0.25 μ L (2.5U) of HpyCh4IV, and 17.25 μ L of dH₂O to bring the total volume to 25 μ L. The mixtures were incubated, 37°C 1 h. Ten μ L of each digest were run on 2.5% agarose gels, 90 volts 30 min.

To determine the accuracy of the polymorphism as a genetic test for OHC, 6-11 weeks after birth blood samples were taken from 81 related sheep from the OHC flock, and each sheep had its eyes examined to check for cataracts at the same time. The eyes were examined using the procedure described in section 2.3.1. DNA was extracted from each sample and was genotyped for the NUDT9 allele as described previously. The genotypes were used to predict the phenotypes, using AA and AG as affected and GG as normal. The predictions were checked against the actual phenotypes.

4.3.3.5 Comparison of NUDT9 Protein Sequences in Different Species

The NUDT9 sequences of several different vertebrates were investigated to see if the polymorphisms involved amino acid residues that were conserved across different species. A sheep NUDT9 protein sequence was generated from the NUDT9 cDNA sequence determined in section 4.3.3.3, and it was aligned with the human and cattle equivalents using DNAMAN (Lynnon).

4.3.3.6 Comparison of NUDT9 Exon Boundaries

The human and mouse cDNA sequences were aligned with the NUDT9 cDNA sequence generated in section 4.3.3.3 using DNAMAN (Lynnon). Exon boundaries for the human and mouse sequences were determined from genomic DNA sequences, and the alignment was used to determine whether there were any changes in exon boundaries between the sheep, human and mouse sequences.

4.3.4 Results

4.3.4.1 Expression of NUDT9 in Normal Lenses

All RNA extracts from normal cultured lenses were run on 2% agarose gels containing ethidium bromide to check their quality. All extracts still showed two clear bands corresponding to the 18S and 28S ribosomal RNA subunits, indicating that they were of sufficient quality for preparing cDNA (Figure 25).

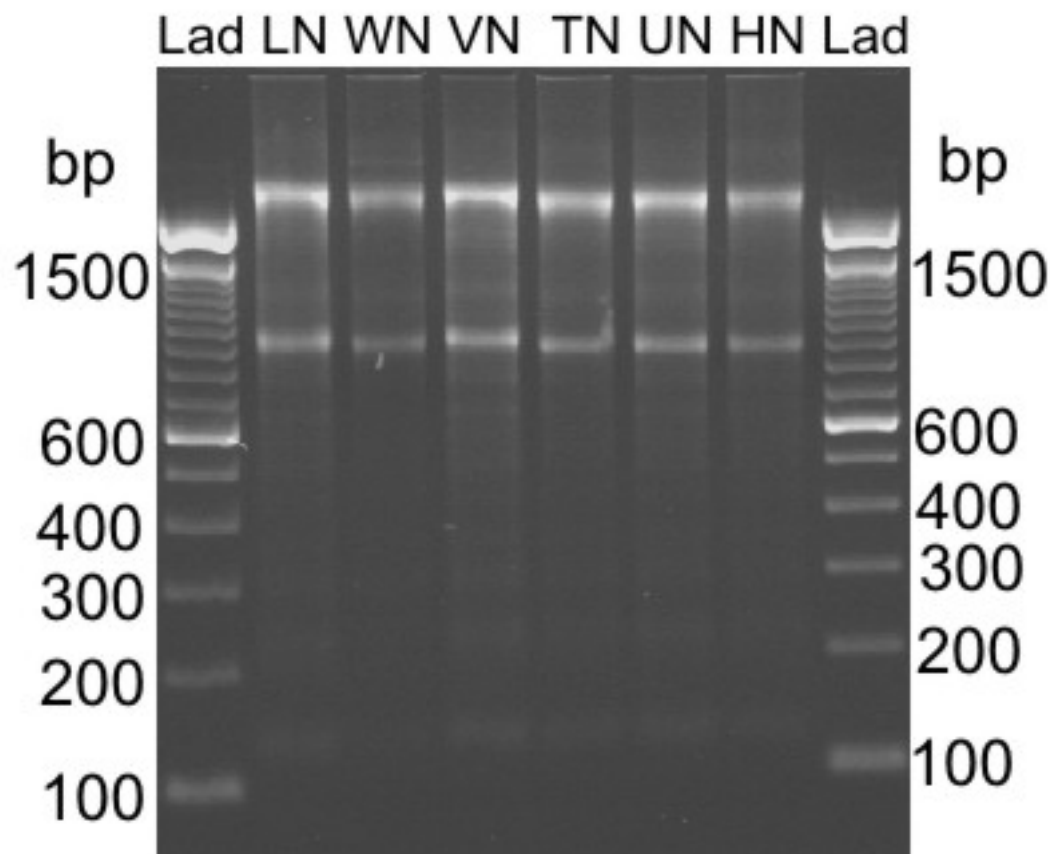


Figure 25: Agarose gel of RNA extracts. The RNA samples originally came from cultured lenses from animals designated with the arbitrary letters L, W, V, T, U and H. The N in the name of each sample indicates that it has not been exposed to opacity-inducing substances during lens culture. The lanes labelled LAD are 100 base pair ladders. The RNA extracts used in the NUDT9 experiments are LN and UN.

The NUDT9 primers were used in PCRs with cDNA prepared from the LN RNA. This RNA was extracted from a cultured lens from animal L, and was not exposed to any opacity-inducing substances, hence the letter N for normal. A reamplification PCR was performed using the PCR product as a template. The agarose gel of the PCR products from both the first PCR and the re-amplification PCR showed a single band at about 1,000 bases, consistent with

the 996bp PCR product size predicted from the cattle CDS (accession number BC151586) (Figure 25).

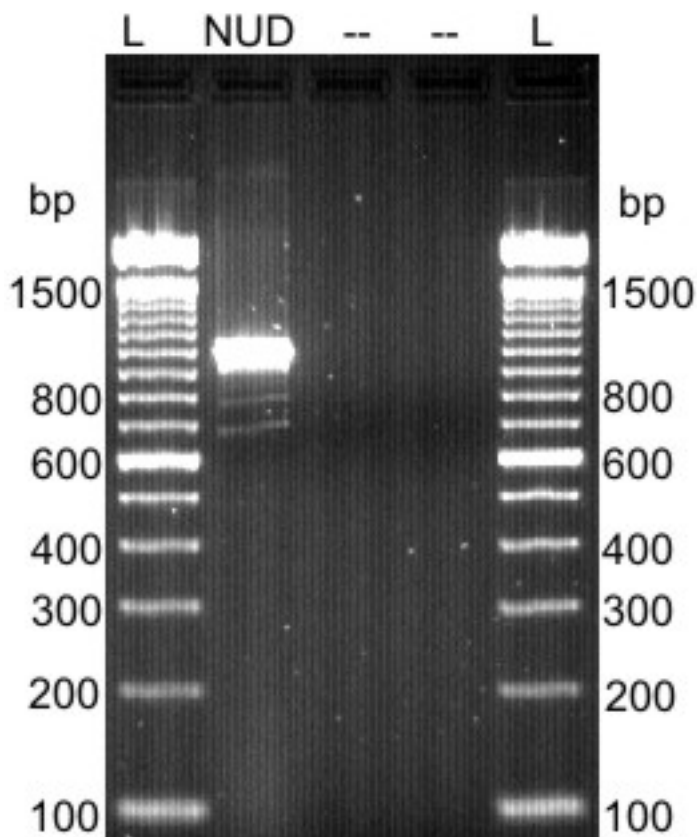


Figure 26: NUDT9F-NUDT9R PCR product. The template for this reaction was normal cDNA. The expected size is 996bp. The lanes labelled L are 100 base pair ladders.

Bands from the gels were extracted for reamplification and sequencing. The first gel extract had a concentration of approximately 3.5ng per μL . The gel extract of the re-amplification PCR product had a concentration of about 10 ng per μL , and was used to prepare sequencing mixtures.

The sequence generated from the gel extract with the forward primer gave a 302-base sequence that showed 96% homology with the cattle CDS. The sequence generated with the reverse primer gave a 477-base sequence that showed 95% homology with the cattle CDS.

4.3.4.2 Sequencing

When the primers were used in PCRs with cDNA generated from both affected and normal RNA, the Start pair of primers gave no amplification, so the PCR was repeated using the StartF primer with the GapR primer (Figure 27). As well as the Gap and End pairs, the Start F and End R primers were used together to amplify the entire CDS (Figure 28). All of the four primer combinations gave products of the expected sizes.

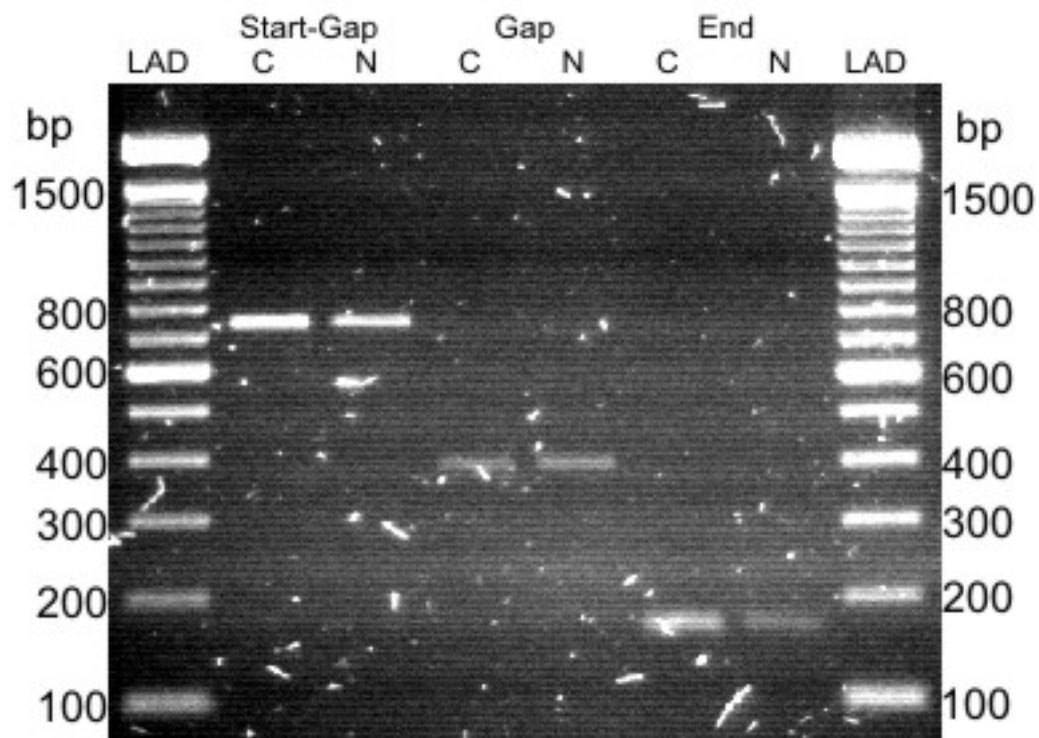


Figure 27: StartF-GapR, GapF-GapR, and EndF-EndR products. The expected sizes are 738bp for StartF-GapR, 378bp for GapF-GapR, and 168bp for EndF-EndR. The lanes labelled LAD are 100 base pair ladders. All of the C lanes were prepared using cDNA from affected sheep 47/07, and all of the N lanes were prepared using cDNA from normal sheep 50/07.

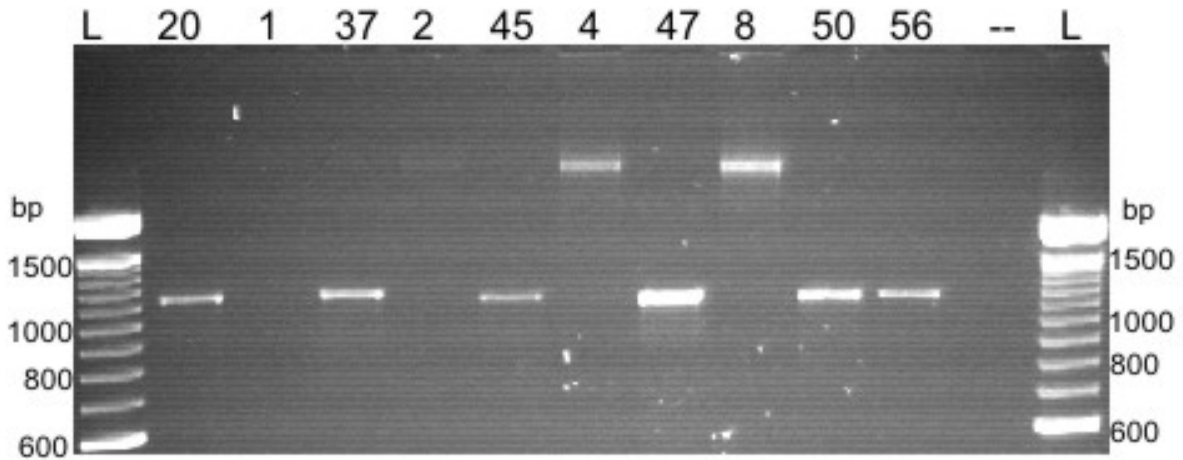


Figure 28: StartF and EndR PCR products. Lanes 20, 47, and 56 are from cDNA from affected sheep, and lanes 37, 45, and 50 are from cDNA from normal sheep. All sheep were born in 2007, and the number of the lane is the birth order. The predicted size from bovine sequence for all the PCR products is 1,132 bp. Lanes 1, 2, 4, and 8 are genomic DNA samples used as concentration standards. The concentrations of the samples are 1, 2, 4, and 8 ng per μL . L indicates a 100 base ladder, and - - stands for the blank. All six products were sequenced.

When the PCR products were sequenced, the sequences generated were long enough to cover the entire cattle CDS. Consensus sequences were assembled for all six animals from the PCR products produced from the StartF and EndR, StartF and GapR, and EndF and EndR primer combinations (Figure 29).

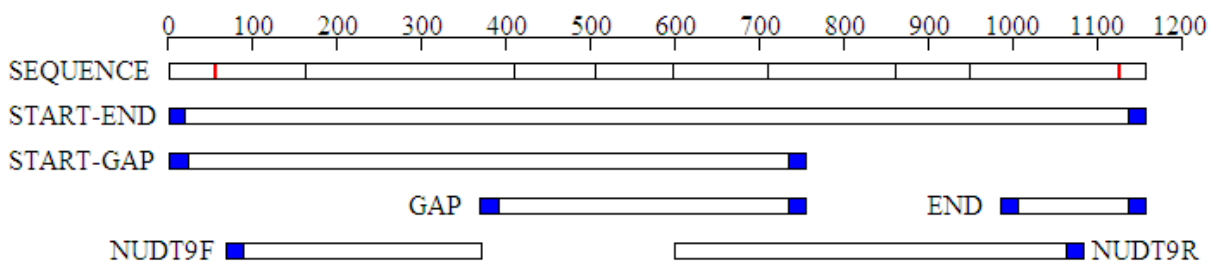


Figure 29: NUDT9 sequences. The consensus sequence is shown at the top. The scale gives distances in bases. Red lines indicate the start and stop codons and the primers are shown in blue. Black vertical lines in the total sequence show the exon boundaries, as discussed in 6.3.5 below.

Inside the CDS, all six of the consensus sequences were identical except for the polymorphisms described in section 4.3.4.3. The consensus of all six sequences is identical to the sequence from the normal sheep 45/07 and 50/07. This sequence was aligned with the bovine NUDT9 CDS and showed 96% homology (Figure 31).

4.3.4.3 Polymorphisms

Three polymorphisms were found in the six consensus sequences (Figure 30).

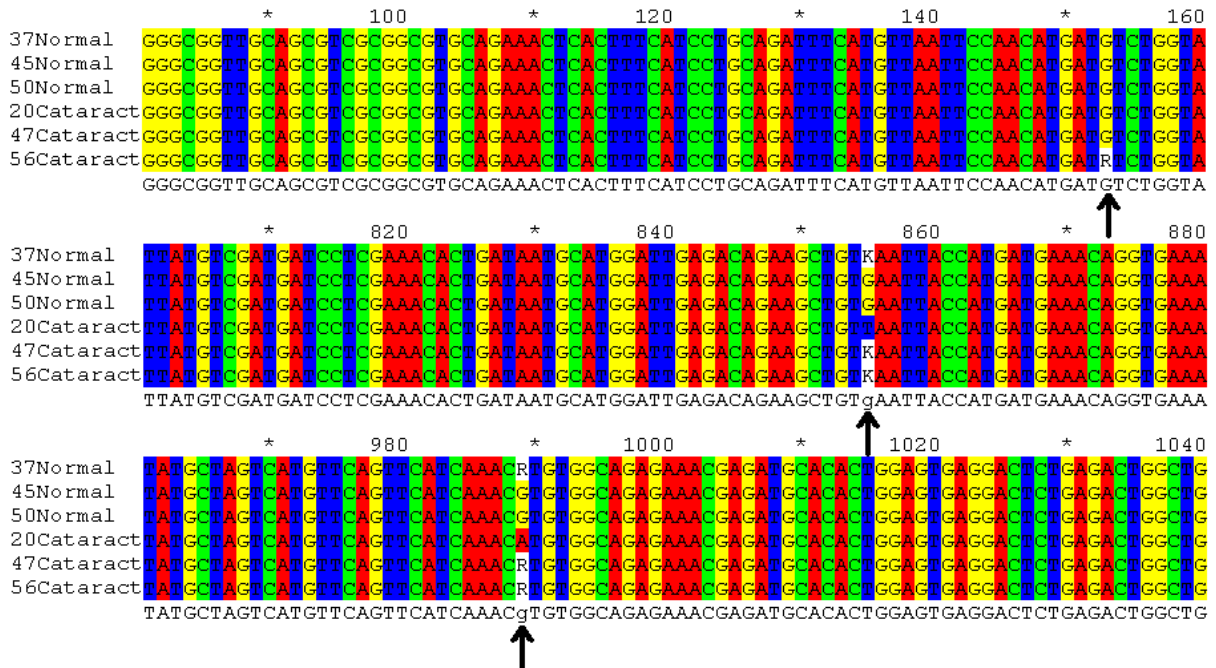


Figure 30: Genedoc alignment of NUDT9 sequences. The alignment shows three regions of the CDS, bases 80-160, 800-880, and 960-1040. The polymorphisms are indicated by arrows. R stands for a G/A heterozygote and K stands for a G/T heterozygote. “Cataract” in the name of an animal means it is affected by OHC.

The three polymorphisms are 153G>A, 855G>T, and 989G>A. The latter two polymorphisms are correlated with cataract status, allowing for incomplete penetrance.

Polymorphism 153G>A causes a change from methionine to isoleucine in the amino acid sequence of the protein. Polymorphism 855G>T is a silent mutation, with both forms coding for valine. Polymorphism 989G>A would cause a change from arginine to histidine. The positions of 855G>T and 989G>A in relation to the start and stop codons are shown in Figure 31.

Upper line: CowCDS BC151586, from 1 to 1050

Lower line: Sheep CDS, from 1 to 1050

CDS_Cow.txt:50Normal.txt identity= 96%

```

1   ATGGCCGGAGGCTTCCTGGGGAAGACTTTAGCCGCCGTGTCCCTCCCTGTCGCCCTGGCC
   |||
1   ATG|GCCGGAGGCTTCCTGGGGAAGACTTTAGCCGCCGTGTCCCTCTCTGTCGCCCTGGCC
61  TCTGTGACTGTTAAGTCCTTGGGCAGTTGCAGCGTCGCGGCGTGCAGAAACTCACTTTCA
   |||
61  TCTGTGACTGTCAAGTCCTTGGGCGGTTGCAGCGTCGCGGCGTGCAGAAACTCACTTTCA
121 TCCTGCAGATTTTCATGTTAATTCCAACATAATGTCTGGTTTTAATGGTACCAAAGAGAAT
   |||
121 TCCTGCAGATTTTCATGTTAATTCCAACATGATGTCTGGTATTAATGGTGCCATAGAGAAT
181 TCCCACAACAAGGCTCGGACATCTCCTTATCCAGGTTCAAAAGTTCAGCGCAGCCAGGTT
   |||
181 TCCCACAACAAGGCTCGGACATCGCCTTACCCAGGTTCAAAAGTTCAGCGCAGCCAGGTT
241 CCTAATGATAAAGTGGACTGGCTTGTGAGTGGCATGACTATAATCCTGTGGAATACACT
   |||
241 CCCAATGATAAAGTGGGCTGGCTTGTGAGTGGCAGACTATAATCCTGTGGAATACACT
301 GCAGTTTCTGTCTTGGCTGGACCTCAGTGGGCAGATCCTCAGATCACTGAAAGCAACTTC
   |||
301 GCAGTTTCTGTCTTGGCTGGACCTCAGTGGGCAGATCCTCAGATCAATGAAAGCAACTTC
361 TCTCCCAAGTTTAAATGAAAAGGATGGGCAGGTTGAGAGAAAAGAGCCAGAATGGCCTGTAT
   |||
361 TCTCCCAAGTTTAAATGAAAAGGATGGGCAGGTTGAGAGAAAAGAGCCAGAATGGCCTGTAT
421 GAGGTTAAAAATGGAAGACCTAGAAATCCTGCAGGACGGACTGGACTGGTTGGCCGGGGG
   |||
421 GAGGTTGAAAATGGAAGACCTAGAAATCCTGCAGGGCGGACTGGACTGGTTGGCCGGGGG
481 CTTTTGGGGCGGTGGGGCCCCAATCATGCTGCAGATCCCATCATAACCAGGTGGAAAAGG
   |||
481 CTTTTGGGGCGATGGGGCCCCAATCATGCTGCAGATCCCATCATAACCAGGTGGAAAAGG
541 GATAGAAGTGAAAATAAAATCACCCATCCCATTTCTGGGAAAAACATCTTGCAGTTTGT
   |||
541 GATAGAAGTGAAAATAAAATCACCCATCCCATTTCTGGGAAAAACATCTTGCAGTTTGT
601 GCAATAAAAAGGAAAAGACTGTGGAGAATGGGCAATCCAGGGGGGATGGTGGATCCAGGA
   |||
601 GCAATAAAAAGGAAAAGACTGTGGAGAATGGGCAATCCAGGGGGGATGGTGGATCCAGGA
661 GAGAAGATTAGTGCCACACTGAAAAGAGAATTTGGTGAGGAAGCTCTGAACTCCTTACAG
   |||
661 GAGAAGATTAGCGCCACACTGAAAAGAGAGTTTGGTGAGGAAGCTCTCAACTCCTTACAG
721 AAGTCCAGCGCTGAAAAGAGAGAATTACAGGAAAAGTTGCAGAAGCTCTTCAACCAGGAA
   |||
721 AAGTCCAGTGCTGAAAAGAGAGAATTACAGGAAAAGCTGCACAAGCTCTTCAACCAGGAA
781 CATCTAGTGATATATAAAGTTATGTTGATGATCCTCGAAACACTGATAATGCATGGATG
   |||
781 CATCTAGTGATATATAAAGTTATGTCGATGATCCTCGAAACACTGATAATGCATGGATT
841 GAGACAGAAGCTGTGAATTACCATGATGAAACAGGTGAAATAATGGATAACCTTACTTTA
   |||
841 GAGACAGAAGCTGTGAATTACCATGATGAAACAGGTGAAATAATGGATAACCTTGCCTTA
901 GAAGCTGGAGATGATGCTGGAAAAGTGAAATGGATGGATATCAGCGATAAACTCAAGCTT
   |||
901 GAAGCTGGAGATGATGCTGGAAAAGTGAAATGGATGGATATCAGCGATAAACTCAAGCTT
961 TATGCTAGTCACGCTCAGTTCATCAAAGTGTGGCAGAGAAACGAGATGCACACTGGAGT
   |||
961 TATGCTAGTCATGTTCAAGTTCATCAAAGTGTGGCAGAGAAACGAGATGCACACTGGAGT

```

```

1021  AAGGACTCTGAGACTGGCTGCCCTGGGTAA
      |||||
1021  GAGGACTCTGAGACTGGCTGCCCTGGGTAA

```

Figure 31: NUDT9 consensus sequence, from start codon to stop codon. The consensus sequence of all six animals, which is identical to the sequence of normal sheep 45/07 or 50/07, is aligned with the bovine equivalent (accession number BC151586). Bovine sequence is shown in black and sheep sequence in blue. The start and stop codons and polymorphisms are highlighted in red.

4.3.4.4 Genotyping

When the new primers were used in PCRs with cDNA from normal and affected sheep, all products gave bands of the appropriate size on an agarose gel (Figure 32).

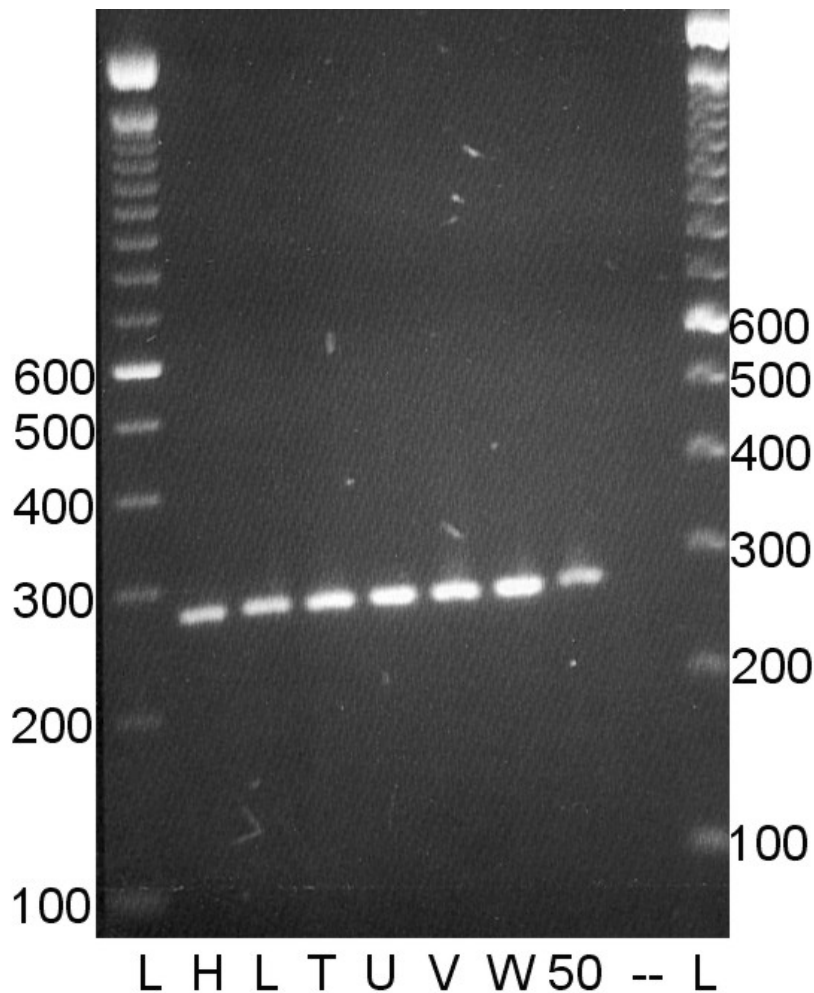


Figure 32: BLDF-BLDR PCR products. The sources of the cDNA are normal cultured lenses HN, LN, TN, UN, WN, and VN, and a blood sample from normal sheep 50/07. The expected size is 277bp.

The PCR products were extracted and used to prepare sequencing reactions. Twenty of the 21 PCR products gave good sequences with both primers, while the PCR product from cataract sheep 61/07 did not. Consensus sequences were prepared for each of the 20 animals (Figure 33). Two more polymorphisms were discovered in the new genotypes but they were not correlated with cataract status.

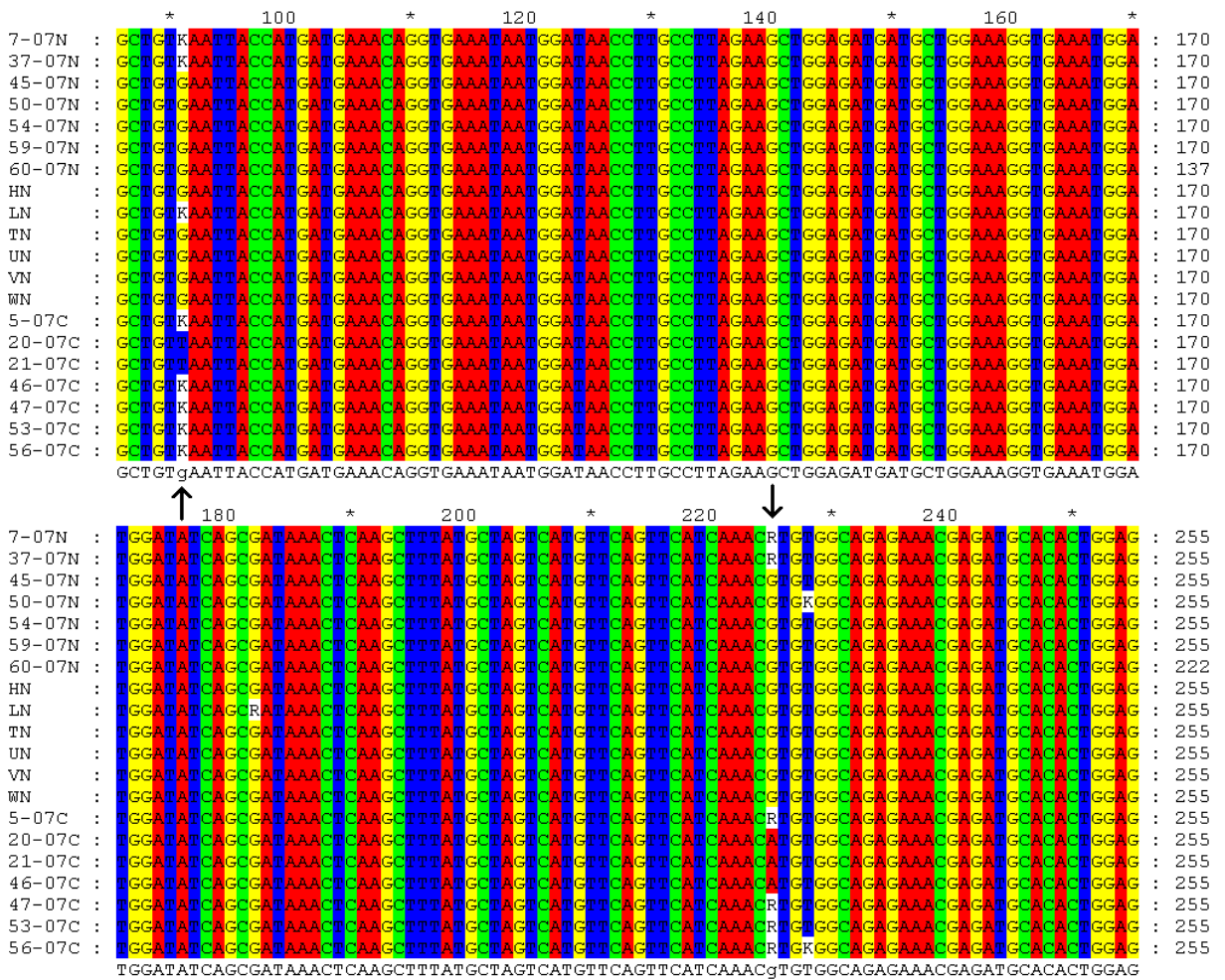


Figure 33: Genedoc alignment of NUDT9 sequences, 20 animals. Names of the animals are shown on the left of the sequence. All animals with names ending in N are normal, whereas all animals with names ending in C are affected. HN, LN, TN, UN, VN, and WN indicate cultured lenses taken from normal animals unrelated to the cataract flock, as described in Section 3.2.1. All other animals are from the OHC flock. The first, silent polymorphism occurs at base 91 in this diagram, whereas the second polymorphism, which causes an amino acid change, occurs at base 225. Both are indicated by arrows. On the CDS these polymorphisms are at 855 and 989 bases from the start codon respectively.

Table 25: 989G>A polymorphism genotypes.

Genotype	Affected	Normal
GG	0	11
AG	4	2
AA	3	0

Table 25 shows the genotypes of 989G>A taken from the new sequencing. All affected animals have either the AG or AA genotype. Most normal animals have the GG genotype. This is consistent with the polymorphism being the OHC mutation, assuming that the mode of inheritance is dominant with incomplete penetrance as shown in Chapter 2. If the polymorphism is the mutation, the OHC allele would be A and the normal allele would be G. The two normal animals with the heterozygote genotype would have an OHC genotype and a normal phenotype due to incomplete penetrance. The genotypes of 855G>T are also consistent with it being the OHC mutation, but since it is a silent polymorphism it is unlikely to be the cause of OHC.

Samples of genomic DNA from both affected and normal animals were used in PCRs with the new primers. Most of the PCRs gave bands of the appropriate size on agarose gels (Figure 34). In total, 149 samples gave acceptable PCR products.

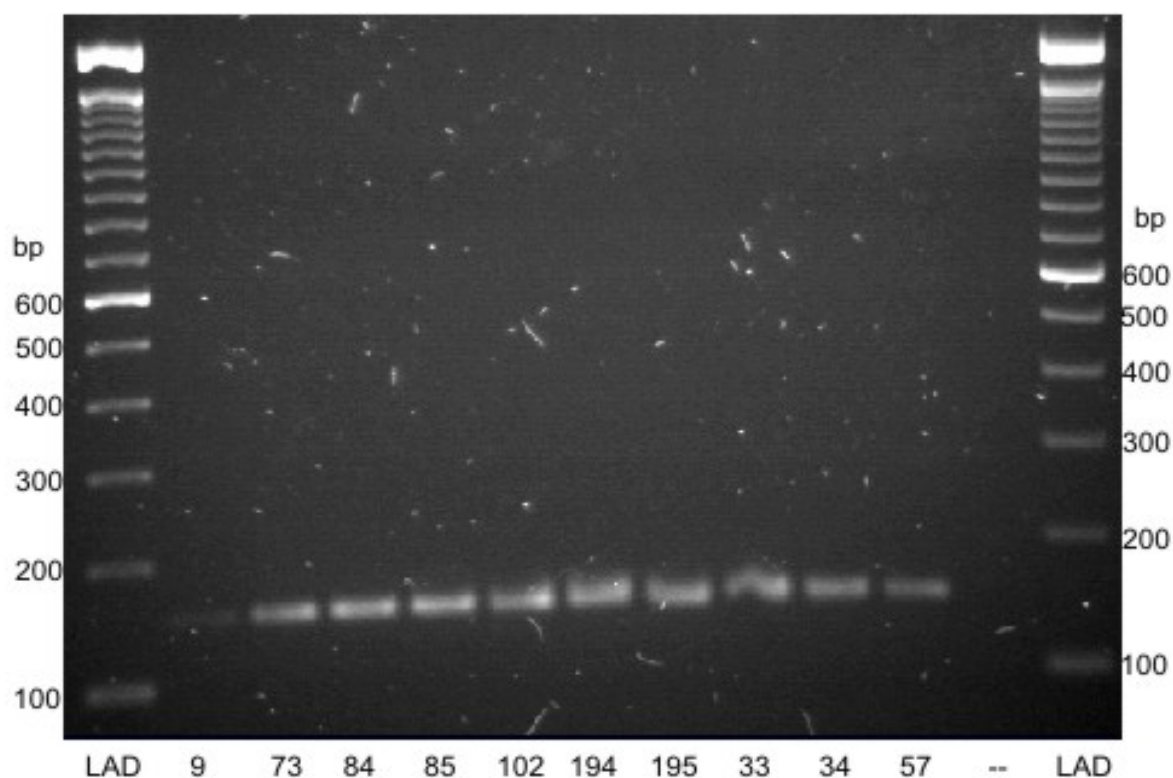


Figure 34: DNA2F-DNA2R PCR products. The templates for this reaction are DNA extracts from both normal and affected animals. The expected size is 151 bp. In this gel all samples except number 9 give bands of sufficient intensity for sequencing.

Nine of the PCR products were sequenced to determine the genotype of 989G>A. Once the genotypes were known, all 148 PCR products were digested with the restriction enzyme HpyCh4IV and run on agarose gels, to generate fragments that depended on the genotype of the same polymorphism.

Of the nine samples with a known genotype from sequencing, all of the restriction digests showed a pattern of fragments consistent with the genotypes (Figure 35). Samples with the homozygote genotype are cut once by the enzyme, giving two fragments of 54 and 97 bp. In samples with the heterozygote genotype, one allele is cut and the other allele remains uncut, giving two fragments and an undigested band of 151 bp at the same intensity. Samples with the AA genotype are not cut by the enzyme, giving a single 151 bp band.

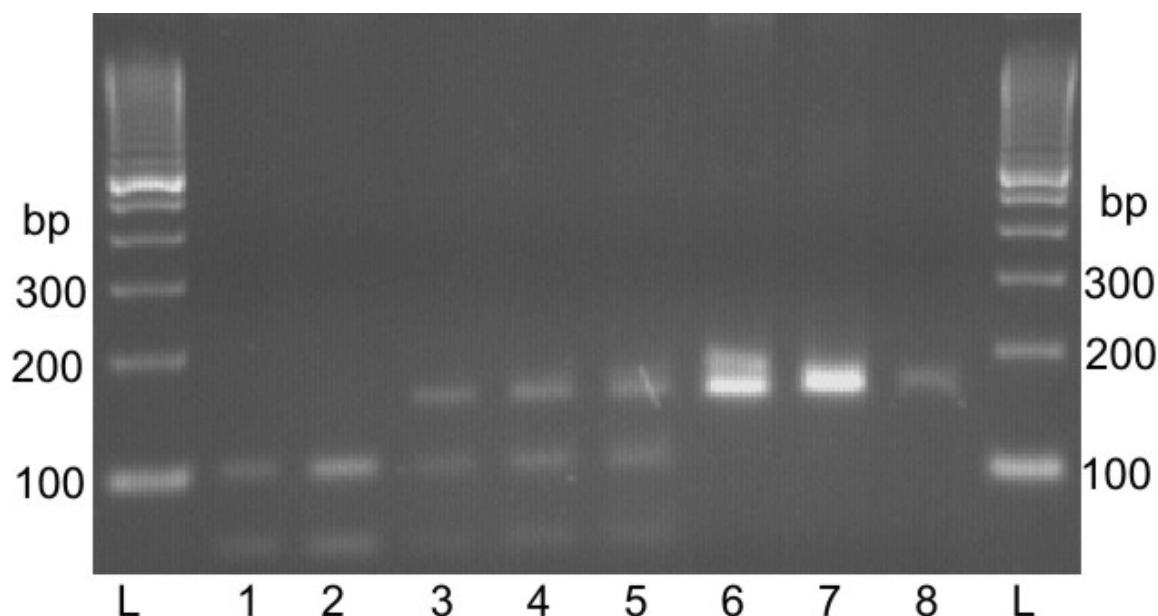


Figure 35: Restriction digest products. Animals with all three genotypes are represented in the samples. L indicates the 100 bp ladder and the other lanes' animal IDs, 989G>A genotypes, and cataract statuses are as follows:

- 1 85/05, GG, normal**
- 2 33/07, GG, normal**
- 3 73/03, AG, affected**
- 4 84/05, AG, affected**
- 5 194/05, AG, affected**
- 6 20/07, AA, affected**
- 7 21/07, AA, affected**
- 8 57/07, AA, affected**

The remaining 141 digested samples were assigned genotypes based on their banding patterns on agarose gels. All of the genotypes of 989G>A are summarised in Table 26.

Table 26: 989G>A genotypes of the OHC flock, with cataract statuses. These include the genotypes from the original six animals used for sequencing.

	Affected	Normal
GG	2	44
AG	125	28
AA	38	1
Total	165	73

Two animals have the genotype GG despite being affected, which is not consistent with the 989G>A allele being the OHC mutation (Table 26). The PCR products for both of these animals were sequenced as described above, using the same primers. Both gave good quality sequences from both primers. The sequences and electropherograms confirmed that both animals were GG homozygotes.

Of the 81 animals tested to determine the accuracy of the polymorphism as a genetic test for OHC, 19 had the GG genotype, 45 had the GA genotype, and 17 had the AA genotype. If this test was accurate, the predicted number of affected sheep would be 62 and the predicted number of normal lambs would be 19. The actual OHC statuses were 49 affected and 32 normal. The differences were all normal animals that had the genotype GA rather than GG. Of the 81 predictions made, 84% were correct and 16% were incorrect. Of the 62 animals that had either the AA or AG genotype, 49 or 79% were affected.

4.3.4.5 Comparison of NUDT9 Protein Sequences in Different Species

Protein sequences from other species were analysed to determine whether the residues equivalent to the polymorphisms are conserved across species (Table 27).

Table 27: Cross-species comparison of NUDT9 polymorphisms. All residues are at the equivalent position to the polymorphisms.

Species	153G>A	855G>T	989G>A
<i>Danio rerio</i>	Serine	Valine	Threonine
<i>Xenopus laevis</i>	Glycine	Valine	Isoleucine
<i>Gallus gallus</i>	Methionine	Valine	Leucine
<i>Ovis aries</i> (allele1)	Methionine	Valine	Arginine
<i>Ovis aries</i> (allele2)	Isoleucine	Valine	Histidine
<i>Bos taurus</i>	Methionine	Valine	Serine
<i>Rattus norvegicus</i>	Methionine	Valine	Leucine
<i>Mus musculus</i>	Methionine	Valine	Leucine
<i>Homo sapiens</i>	Methionine	Valine	Leucine
<i>Pan troglodytes</i>	Methionine	Valine	Leucine

The residue equivalent to 153G>A is conserved across all the species investigated, except *Danio rerio* and *Xenopus laevis*. One affected animal has a variation in the cDNA sequence which would change the methionine residue to an isoleucine, but the other two affected animals that were sequenced do not have this variation.

855G>T would not change the amino acid residue, and the residue it is equivalent to is conserved across all species investigated.

For 989G>A the arginine residue found in normal sheep is not found in any other species that was investigated, but neither was the histidine residue found in affected sheep. When the adjacent residues are examined in each protein sequence, no arginines are found.

The sheep NUDT9 protein sequence was aligned with the cattle and human homologues using DNAMAN (Lynnon). The sheep-cattle and sheep-human alignments gave homologies of 95% and 86% respectively. The affected and normal sheep sequences gave the same degree of homology to both cattle and human sequences. The cattle-human alignment gave a homology of 85%. The area of poorest homology between sheep and human sequences is found close to the N-terminal, in the region where the crystal structure has not been determined (Shen *et al*, 2003). This area of poor homology is also found in the cattle-human alignment, showing that it is probably a genuine species difference. The difficulty in determining the crystal structure of the N-terminal region may indicate that the conformation of this region can vary without affecting the protein's function. In that case the amino acid sequence could also vary between species, giving poor homology between the different sequences.

ORIGIN		
Cattle	MAGGFLGKTLAAVSLPVALASVTVKSLGSCSVAACRNSLS	40
Sheep	MAGGFLGKTLAAVSLsVALASVTVKSLGgCSVAACRNSLS	40
Human	MAGrLLGKaLAAVSLslALASVTirSsrcrgiqAfrNSfS	40
Consensus	mag lgk laavsl alasvt s a rns s	
Cattle	SCRFHVNSNIMSGFNGTKENSHNKARTSPYPGSKVQRSQV	80
Sheep	SCRFHVNSNmSGiNGaiENSHNKARTSPYPGSKVQRSQV	80
Human	SswFHlNtNvMSGsNGsKENSHNKARTSPYPGSKVeRSQV	80
Consensus	s fh n n msg ng enshnkartspypgskv rsqv	
Cattle	PNDKVDWLVEWHDYNPVEYTAVSVLAGPQWADPQITESNF	120
Sheep	PNDKVgWLVEWHDYNPVEYTAVSVLAGPQWADPQInESNF	120
Human	PNeKVgWLVEWqDYkPVEYTAVSVLAGPrWADPQIsESNF	120
Consensus	pn kv wlvev dy pveytavsvlagp wadpqi esnf	
Cattle	SPKFNEKDGQVERKSQNGLYEVKNGRPRNPAGRTGLVGRG	160
Sheep	SPKFNEKDGQVERKSQNGLYEveNGRPRNPAGRTGLVGRG	160
Human	SPKFNEKDGhVERKSkNGLYEieNGRPRNPAGRTGLVGRG	160
Consensus	spkfnekdg verks nglye ngrprnpagrtglvgrg	
Cattle	LLGRWGPnHAADPIITRWKRDRSGNKIITHPIsGKNILQFV	200
Sheep	LLGRWGPnHAADPIITRWKRDRSGNKIITHPIsGKNILQFV	200
Human	LLGRWGPnHAADPIITRWKRdSGNKImHPvSGKhILQFV	200
Consensus	llgrwgpnhaadpiitrwkrd sgnki hp sgk ilqfv	
Cattle	AIKRKDCGEWAIPGGMVDpGEKISATLkREFGEEALNSLQ	240
Sheep	AIKRKDCGEWAIPGGMVDpGEKISATLkREFGEEALNSLQ	240
Human	AIKRKDCGEWAIPGGMVDpGEKISATLkREFGEEALNSLQ	240
Consensus	aikrkdcgewaipggmvdpgekisatlkrefgeaalnslq	
Cattle	KSSAEKRELQEKlQKLfNQEHLVIYkGYVDDPRNTDnAWM	280
Sheep	KSSAEKRELQEKlHkLkLfSQEHLVIYkGYVDDPRNTDnAWi	280
Human	KtSAEKREieEKlHkLkLfSqdHLVIYkGYVDDPRNTDnAWM	280
Consensus	k saekre ekl klf q hlviykygyvddprntdnaw	

Cattle	ETEAVNYHDETGEIMDNLTLEAGDDAGKVKWMDISDKLKL	320
Sheep	ETEAVNYHDETGEIMDNLaLEAGDDAGKVKWMDISDKLKL	320
Human	ETEAVNYHDETGEIMDNLMLEAGDDAGKVKWvDInDKLKL	320
Consensus	eteavnnyhdetgeimdnl leagddagkvkw di dklkl	
Cattle	YASHAQFIKSVAEKRDHWSKDSETGCPG	349
Sheep	YASHvQFIKvVAEKRDHWSedSETGCPG	349
Human	YASHsQFIKlVAEKRDHWSedSEadChal	350
Consensus	yash qfik vaekrdahws dse c	

Figure 36: Cattle, sheep and human NUDT9 protein sequences. The cattle sequence's accession number is AAI51587 and the human sequence's accession number is AAH00542, both from <http://www.ncbi.nlm.nih.gov/protein>. The sheep sequence has been generated by translating the normal sheep NUDT9 cDNA sequence shown in Figure 31. Polymorphisms are highlighted in red. The cattle sequence is shown at the top. The sheep sequence is shown below it, with residues in lower case where they do not match the cattle sequence. The human sequence is shown below the cattle and sheep sequences, with residues in lower case where they do not match both of the other sequences. The consensus sequence, with residues that are present in all three sequences, is shown at the bottom. Each sequence is shown from the N-terminal to the C-terminal. Pairs of sequences were also aligned to find degrees of homology (not shown).

4.3.4.6 Comparison of NUDT9 Exon Boundaries

The normal sheep NUDT9 cDNA sequence generated in this section was aligned with the human and mouse homologues using DNAMAN (Lynnon), in order to compare the exon boundaries (Figure 37). The human cDNA sequence and exon boundaries were taken from NCBI Reference Sequence NM_024047.3, and the mouse equivalents from NCBI Reference Sequence NM_028794.3.

ORIGIN		
humanNUDT9cds	ATGGCGGGACGCCTCCTGGGAAAGGCTTTAGCCGCGGTGT	40
mouseNUDT9cds	ATGGCcGGcCGctcCCTGGGgcAGGCggTAGCCaCcGTGT	40
50Normal	ATGGCcGGAgGctTCCTGGGgAAGaCTTTAGCCGcCGTGT	40
Consensus	atggc gg gc cctggg ag c tagcc c gtgt	
humanNUDT9cds	CTCTCTCTCTGGCCTTGGCCTCTGTGACTATCAGGTCCCTC	80
mouseNUDT9cds	CgCTCTCggTGGCCcTcGCgTCaGTGACTgTCCGGTCCCTC	80
50Normal	CcCTCTCTgTcGCCcTGGCCTCTGTGACTgTCAaGTCCCTt	80
Consensus	c ctctc t gcc t gc tc gtgact tc gtcct	
humanNUDT9cds	GCGCTGCCGCGGCATCCAGGCGTTCAgAACTCGTTTTCA	120
mouseNUDT9cds	agcCTGCCGCGcCgTCCccGCaccCAGAAACaCaTTTcCA	120
50Normal	GgGCgGttGCaGCgTCgcGGCGTgCAGAAACTCacTTTCA	120
Consensus	c g gc c tc gc cagaaac c tt ca	
humanNUDT9cds	TCTTCTTGGTTTTCATCTTAATACCAACGTCATGTCTGGTT	160
mouseNUDT9cds	aCcTgTgGGTTcCAcCTTAAcgcCCAACaTCATGTCTGGTT	160
50Normal	TCcTgcaGaTTTTcATgTTAAAtCCAACaTgATGTCTGGTa	160
Consensus	c t g tt ca ttaa ccaac t atgtctggt	
humanNUDT9cds	CTAATGGTTCCAAAGAAAATTCTCACAATAAGGCTCGGAC	200
mouseNUDT9cds	CTAATGGTgCCAAAGAgAATTCCcACAAcAAGGCcCGGAC	200
50Normal	tTAATGGTgCCAtAGAgAATTCCcACAAcAAGGCTCGGAC	200
Consensus	taatggt cca aga aattc cacaa aaggc cggac	

humanNUDT9cds	GTCTCCTTACCCAGGTTCAAAAAGTTGAACGAAGCCAGGTT	240
mouseNUDT9cds	tTCTCCTTACCCAGGTTCAAAAAGTcGAgCGAAGtCAGGTT	240
50Normal	aTCgCCTTACCCAGGTTCAAAAAGTTcAgCGcAGCCAGGTT	240
Consensus	tc ccttaccaggttcaaaagt a cg ag caggtt	
humanNUDT9cds	CCTAATGAGAAAAGTGGGCTGGCTTGTGAGTGGCAAGACT	280
mouseNUDT9cds	CCgAATGAGAAgGTGGGCTGGCTcGTTGAGTGGCAgGACT	280
50Normal	CCcAATGAtAAAAGTGGGCTGGCTTGTGAGTGGCAcGACT	280
Consensus	cc aatga aa gtgggctggct gttgagtggca gact	
humanNUDT9cds	ATAAGCCTGTGGAATACACTGCAGTCTCTGTCTTGGCTGG	320
mouseNUDT9cds	AcAAcCCaGTGGAgTACAcCGcGTCTCTGTCCtGGCTGG	320
50Normal	ATAAtCCTGTGGAATACACTGCAGTtTCTGTCTTGGCTGG	320
Consensus	a aa cc gtgga tacac gc gt tctgtc tggctgg	
humanNUDT9cds	ACCCAGGTGGGCAGATCCTCAGATCAgTGAAAGTAATTTT	360
mouseNUDT9cds	ACCCcaGTGGGcGgAcCCTCAGATCAgTGAAAGcAAcTTc	360
50Normal	ACcTcaGTGGGCAGATCCTCAGATCAaTGAAAGcAAcTTc	360
Consensus	acc gtgggc ga cctcagatca tgaaag aa tt	
humanNUDT9cds	TCTCCCAAGTTTAAcGAAAAGGATGGGCATGTTGAGAGAA	400
mouseNUDT9cds	TCTCCCAAGTTTAAAtGAAAAGGAcGGGCATGTTGAGAGAA	400
50Normal	TCTCCCAAGTTTAAAtGAAAAGGATGGGCAGgTTGAGAGAA	400
Consensus	tctcccaagtttaa gaaaagga gggca gttgagagaa	
humanNUDT9cds	AGAGCAAGAATGGCCTGTATGAGATTGAAAATGGAAGACC	440
mouseNUDT9cds	AGAGCcAGAATGGCCTGTATGAGATTGAAAAcGGAAGACC	440
50Normal	AGAGCcAGAATGGCCTGTATGAGgTTGAAAATGGAAGACC	440
Consensus	agagc agaatggcctgtatgag ttgaaaa ggaagacc	
humanNUDT9cds	GAGAAATCCTGCAGGACGGACTGGACTGGTGGGCCGGGGG	480
mouseNUDT9cds	cAGAAAcCCTGCAGGgaGGACaGGACTGGTtGGCCGGGGt	480
50Normal	tAGAAATCCTGCAGGgCGGACTGGACTGGTtGGCCGGGGa	480
Consensus	agaaa cctgcagg ggac ggactggt ggccgggg	
humanNUDT9cds	CTTTTGGGGCGATGGGGCCAAATCACGCTGCAGATCCCA	520
mouseNUDT9cds	CTTcTGGGGaGATGGGGCCAAATCAAtGCcGCAGATCCtA	520
50Normal	CTTTTGGGGCGATGGGGCCcAAATCAAtGCTGCAGATCCCA	520
Consensus	ctt tgggg gatggggccc aatca gc gcagatcc a	
humanNUDT9cds	TTATAACCAgATGGAAAAGGGATAGCAGTGGAAATAAAAT	560
mouseNUDT9cds	TTATAACCAgTGGAAAAGGGAcgagAGcGGgAATAAgAT	560
50Normal	TcATAACCAgTGGAAAAGGGATAGaAGTGGAAATAAAAT	560
Consensus	t ataaccag tggaaaaggga ag gg aataa at	
humanNUDT9cds	CATGCATCCTGTTTCTGGGAAGCATATCTTACAATTTGTT	600
mouseNUDT9cds	CACaCAcCCTGTcTCTGGGAAAtgcATCTTgCAgTTcGTT	600
50Normal	CACcCATCCcaTTTCTGGGAAaaAcATCTTgCAgTTTgTT	600
Consensus	ca ca cc t tctgggaa atctt ca tt gtt	
humanNUDT9cds	GCAATAAAAAAGGAAAGACTGTGGAGAATGGGCAATCCCAG	640
mouseNUDT9cds	GCcATcAAgAGGAAAGACTGTGGAGAgTGGGCAATCCCgG	640
50Normal	GCAATAAAAAAGGAAAGACTGTGGAGAATGGGCAATCCCAG	640
Consensus	gc at aa aggaaagactgtggaga tgggcaatccc g	
humanNUDT9cds	GGGGGATGGTGGATCCAGGAGAGAAGATTAGTGCCACACT	680
mouseNUDT9cds	GGGGGATGGTGGAcCCcGGAGAGAAGATcAGTGCCACcCT	680
50Normal	GGGGGATGGTGGATCCAGGAGAGAAGATTAGcGCCACACT	680
Consensus	gggggatggtgga cc ggagagaagat ag gccac ct	
humanNUDT9cds	GAAAAGAGAATTTGGTGAGGAAGCTCTCAACTCCTTACAG	720

mouseNUDT9cds	GAAAAGgGAgTTTGGTGAAgGAAGCcCTgAACTCgTTACAG	720
50Normal	GAAAAGAGAgTTTGGTGAGGAAGCTCTCAACTCCTTACAG	720
Consensus	gaaaag ga tttggtga gaagc ct aactc ttacag	
humanNUDT9cds	AAAACCAGTGCTGAGAAGAGAGAAAATAGAGGAAAAGTTGC	760
mouseNUDT9cds	AAgtCCAGTGCCcGAGAAGAGAGAgATcGAGGAgAAacTGC	760
50Normal	AAgtCCAGTGCTGAAaAGAGAGAAATAcAGGAAAAGcTGC	760
Consensus	aa ccagtgc ga aagagaga t agga aa tgc	
humanNUDT9cds	ACAAACTCTTCAGCCAAGACCACCTAGTGA TATATAAGGG	800
mouseNUDT9cds	ACgcgCTCTTCAGCCAAGAgCAcCTcGTGA TATATAAGGG	800
50Normal	ACAAgCTCTTCAGCCAgGAaCAcCTAGTGA TATATAAaGG	800
Consensus	ac ctcttcagcca ga ca ct gtgatataaa gg	
humanNUDT9cds	ATATGTTGATGATCCTCGAAACACTGATAATGCATGGATG	840
mouseNUDT9cds	gTATGTTGAcGAcCCTCGgAACACTGAcAATGCgTGGATG	840
50Normal	tTATGTcGATGATCCTCGAAACACTGATAATGCATGGATt	840
Consensus	tatgt ga ga cctcg aacactga aatgc tggat	
humanNUDT9cds	GAGACAGAAGCTGTGAACTACCATGACGAAACAGGTGAGA	880
mouseNUDT9cds	GAGACgGAAGCcGTGAACTACCATGAtGAAACAGGgGAGA	880
50Normal	GAGACAGAAGCTGTGAAATACCATGAtGAAACAGGTGAaA	880
Consensus	gagac gaagc gtgaa taccatga gaaaca gg ga a	
humanNUDT9cds	TAATGGATAATCTTATGCTAGAAGCTGGAGATGATGCTGG	920
mouseNUDT9cds	ccATGGAcAAcCTcAccCTgGAgGCcGGAGATGAcGCaGG	920
50Normal	TAATGGATAAcCTTgcctTAGAAGCTGGAGATGATGCTGG	920
Consensus	atgga aa ct t ga gc ggagatga gc gg	
humanNUDT9cds	AAAAGTGAAATGGGTGGACATCAATGATAAACTGAAGCTT	960
mouseNUDT9cds	gAAgGTGAAgTGGGTGGACATCAgcGAccAgCTcAAGtTg	960
50Normal	AAAgGTGAAATGGaTGGAtATCAgcGATAAACTcAAGCTT	960
Consensus	aa gtgaa tgg tgg a atca ga a ct aag t	
humanNUDT9cds	TATGCCAGTCACTCTCAATTCATCAAACCTGTGGCTGAGA	1000
mouseNUDT9cds	TAcGCCAGTCACTCTCAgTTCATCAAACCTcGTGGCAGAGA	1000
50Normal	TATGCTAGTCAtgtTCAGTTCATCAAACgTGTGGCAGAGA	1000
Consensus	ta gc agtca tca ttcatcaaac gtggc gaga	
humanNUDT9cds	AACGAGATGCACACTGGAGCGAGGACTCTGAAGCTGACTG	1040
mouseNUDT9cds	AACGAGATGCcCACTGGAGCGAGGACcacGccGCTGACaG	1040
50Normal	AACGAGATGCACACTGGAGtGAGGACTCTGAgACTGgCTG	1040
Consensus	aacgagatgc cactggag gaggac g ctg c g	
humanNUDT9cds	CCATGCGTTGTAG	1053
mouseNUDT9cds	CCgTGgGcTaTAG	1053
50Normal	CCcTGgGTaa	1050
Consensus	cc tg g	

Figure 37: NUDT9 exon boundaries. The human and mouse cDNA sequences are aligned with the consensus for normal sheep 50/07. Lower case is used as described in Figure 36. In the human sequence the adjacent bases that represent the end of one exon and the start of another are coloured blue. The mouse and sheep homologues to these bases are coloured blue if they are identical and red if they are different.

There are seven exon boundaries in the human and mouse sequences. In six of the boundaries the adjacent bases bracketing the boundaries are identical between human and mouse sequences, while in one boundary a mouse exon starts with a G while the homologous human

exon starts with an A (Figure 37). A single base substitution between human and mouse sequences is consistent with normal evolution and does not change the reading frame or the position of the boundary. Therefore the seven positions can be used for cross-species comparisons.

The exon boundaries of sheep NUDT9 cannot be determined from the sequencing performed in this section, because only the cDNA sequence and short fragments of genomic DNA inside an exon were sequenced. When the sheep cDNA sequence was aligned with the human and mouse sequences, for five of the seven exon boundaries both of the surrounding bases were identical in all three animals (Figure 37). For one boundary one sheep base matched the mouse base but not the human base. In all six of these cases the sheep sequence is identical to either the human or mouse sequence in the bases surrounding a known exon boundary. Therefore the exon boundaries are likely to be at the same position in the sheep cDNA sequence.

For one boundary, both the mouse and human sequences have a G where the sheep sequence has an A (Figure 37). As described above, this substitution does not change the reading frame or the positions of the other bases. Therefore the positions of all seven exon boundaries are the same in the sheep sequence as in the human and mouse sequences. No sequence has been inserted or deleted in any exon. The length of the introns has not been determined and may vary between species.

4.3.5 Discussion

4.3.5.1 Gene and Protein Details

NUDT9 codes for an enzyme that acts specifically on ADP-ribose, hydrolysing it into AMP and ribose-5-phosphate (Perraud *et al*, 2001). ADP-ribose has been shown to damage proteins by glycooxidation (Cervantes-Laurean *et al*, 1996). This is a process where reducing sugars react with proteins in an uncontrolled manner, followed by oxidation. The aldehyde form of the sugar reacts with amine groups on the protein to form Schiff base adducts, and these adducts undergo the Amadori rearrangement to form ketoamine adducts. Oxidation of the ketoamine adducts can lead to a series of complex modifications that disrupt the functions of proteins and promote protein-protein cross-links (Cervantes-Laurean *et al*, 1996).

Increased concentrations of glycoxidation products have been detected in both senile and diabetic cataracts (Zarina *et al* 2000). The products are mostly found in high molecular weight aggregates, supporting a model of cataract formation where glycoxidation leads to cross-linking of crystallin proteins and then the formation of aggregates large enough to scatter light and cause opacities, as described in Section 1.3.2. Senile cataracts may result from the exposure of lens proteins to sugars over decades, combined with the development of oxidising conditions in the lens. The cause of diabetic cataracts is still poorly understood, but high sugar concentrations in the blood may increase sugar concentration in the lens and trigger glycoxidation.

Mutations in NUDT9 may cause OHC by preventing the breakdown of ADP-ribose, which may lead to glycoxidation. This is not the current hypothesis for the formation of OHC cataracts in sheep, but since there has been no investigation of glycoxidation in the sheep cataracts it cannot be ruled out.

The calcium channel protein TRPM2 has a region of 50% homology with NUDT9 and has been shown to hydrolyse ADP-ribose with the same specificity, but a lower level of activity (Perraud *et al*, 2001). Binding of ADP-ribose activates the calcium channel, allowing calcium to enter the cell. The actual hydrolysis of ADP-ribose is not necessary for activation of the channel (Kühn and Lückhoff, 2004).

The TRPM2 gene is found on human chromosome 21, so it is not a candidate gene for OHC. However, a build-up of ADP-ribose caused by mutations in NUDT9 could over-activate the TRPM2 calcium channel, leading to increased calcium in the lens cells. This could in turn over-activate calpain and lead to degradation of the lens proteins and cataract formation. This mechanism is consistent with the current model for cataract formation in OHC, as described in Section 1.5.4.

The close homology between NUDT9 on human chromosome 4 and TRPM2 on human chromosome 21 is unexplained at present. TRPM2 contains a ~600 amino acid N-terminal domain and a ~300 amino acid channel-forming domain that are both homologous to other ion channel proteins, and unrelated to NUDT9 (Perraud *et al*, 2005).

Both NUDT9 and TRPM2 contain a Nudix box. Nudix boxes are sequences found in several different enzymes in a variety of species, all of which catalyse the hydrolysis of nucleoside diphosphate derivatives (Bessman *et al*, 1996). The name indicates that the enzymes act on NUCleoside Diphosphates bound to some other group, X, hence “NUDIX”. The name NUDT9 is derived from the name of the nudix box and stands for NUDix-Type motif 9.

The human NUDT9 protein consists of 350 amino acid residues arranged in 17 β -strands and seven α -helices (Shen *et al*, 2003). The human protein has a site of enhanced susceptibility to proteolysis, which tends to cleave the protein into a 163-residue N-terminal domain and a 187-residue C-terminal domain.

The C-terminal domain includes a four-stranded β -sheet, a three-stranded β -sheet, and a triangle of α -helices which form the core of the catalytic domain. The 25-residue Nudix box is found in an α -helix in the C-terminal domain and is necessary for catalysis. The substrate binding site of NUDT9 is located in a cleft between the N-terminal and C-terminal domains, and two other sites in the C-terminal domain bind magnesium ions.

A study of gene expression in the eye using DNA microarrays found that NUDT9 was expressed four times as much in the lens as in the rest of the eye (Diehn *et al*, 2005). NUDT9 was the only gene in the linked region with evidence of expression in the eye.

One difficulty with any cataract mechanism for NUDT9 is that OHC has a dominant mode of inheritance, so heterozygotes with one normal copy of the cataract gene still develop cataracts. If one normal copy of NUDT9 was available, it is possible that its expression could be upregulated to break down any excess of ADP-ribose and prevent cataract formation. For similar reasons few dominant conditions are caused by mutations in enzymes (Phadnis and Fry, 2005). An example of such a condition is sialuria, which is caused by a mutation that prevents the inhibition of the enzyme but leaves the enzyme activity intact (Leroy, J. G. *et al*, 2001).

4.3.5.2 Analysis of NUDT9 as a Candidate for the OHC Gene

Three polymorphisms were found in the first analysis of the NUDT9 cDNA sequence. Since 153G>A is different in only one animal out of six, it is probably not associated with OHC. 855G>T is a silent mutation, so it is unlikely to be the cause of OHC. 989G>A changes

arginine to histidine. Further sequencing found two additional polymorphisms, neither of which were correlated with cataract.

The structure of the human NUDT9 enzyme is described in Section 4.3.1.1. 989G>A in the DNA sequence is equivalent to residue 330 in the sheep protein sequence. In the human enzyme, this position is part of an α -helix in the C-terminal catalytic domain (Shen *et al*, 2003). This is a different α -helix from the helix that contains the Nudix box. Residue 324, six residues towards the N-terminal from it, is a histidine residue which lies within a loop at the end of the α -helix and binds to the reaction product in the active site (Shen *et al*, 2003). If the change from arginine to histidine at residue 330 disrupted the α -helix and changed the position of residue 324, this could interfere with the function of the protein.

153G>A is equivalent to residue 51 of sheep NUDT9 protein. This is close to the N-terminus of the sheep protein, in a region where the crystal structure has not been determined. This polymorphism was only observed to vary in one animal.

855G>T is equivalent to residue 285 of the sheep protein. This is 50 bases away from the Nudix box in the direction of the C-terminus and part of one strand of a four-strand β -sheet. Since the polymorphism is silent it does not change the structure of the protein.

For 989G>A, the genotypes of 20/07, 45/07, 47/07, 50/07, and 56/07 are consistent with the A form being a dominant allele for OHC. However 37/07 is a G/A heterozygote despite being normal. This either indicates that the position is not involved in OHC or that there is incomplete penetrance which prevents the development of OHC phenotype in an OHC genotype. There is evidence from breeding records of such incomplete penetrance, as described in Chapter 2.

A polymorphism that is physically close to the OHC locus will be linked to it and each allele will segregate with a particular OHC status even if there is no other connection with OHC. Therefore 989G>A is not necessarily responsible for OHC.

Genotyping a wider population revealed that two animals have the GG genotype and also have cataracts. Since these cataracts occur at the same age as OHC and look identical to OHC

cataracts, they are almost certainly caused by OHC. Therefore 989G>A is probably not the OHC mutation.

If 989G>A were the OHC mutation, the genotype AA would represent an animal that was homozygous for the OHC gene. Thirty-eight affected animals have this genotype. Of these, 36 have no offspring. One has four affected and 1 normal offspring with cataract ewes. This is consistent with a homozygote at the 82% penetrance of OHC determined through breeding studies (Chapter 2). The remaining ram has 27 normal offspring and 9 affected offspring with normal ewes. For a homozygote this would require OHC to have a penetrance of only 25%, which is not consistent with the breeding studies described in Chapter 2.

The comparison of NUDT9 protein sequences in different species is further evidence against 989G>A being the OHC mutation, because the arginine residue it replaces is not found in equivalent or adjacent positions in any other normal animal, so it is unlikely to be necessary to prevent cataracts.

The human genome browser at <http://genome.ucsc.edu/cgi-bin/hgTracks> was used to search for SNPs in positions equivalent to the polymorphisms. There were 128 SNPs in the human NUDT9 gene, including introns and untranslated regions. The nearest SNP to the equivalent to 989G>A was at 943 bases from the human start codon. This SNP also alternated between an A and a G. There are no SNPs in the equivalent exons for 153G>A or 855G>T. Therefore there is no evidence that any of the polymorphisms seen in the sheep cDNA sequence are equivalent to any human SNPs. This is consistent with other comparisons of SNPs between species (Sechi *et al*, 2009), which show that even closely related animals such as cattle and sheep have too much sequence divergence to have a significant proportion of SNPs in common. Since sheep and humans are more distantly related, it is even less likely that they have SNPs in common.

Because the entire CDS of NUDT9 has been sequenced and compared between affected and normal animals without finding any other candidates for the OHC mutation, NUDT9 is probably not the OHC gene. Mutations in the non-coding sequence of genes have been known to cause disease (Liu *et al*, 1999 and Wen *et al*, 1990), but ruling out mutations in the coding sequence still makes it much less likely that NUDT9 is the OHC gene. However, the

genotype of 989G>A is strongly correlated with OHC status and could be the basis for an indirect genetic test.

Of the 81 sheep which were tested for OHC using 989G>A, the only sheep with phenotypes that were not predictable from their genotypes were normal sheep with the AG genotype. These errors could be due to recombination between the OHC locus and the NUDT9 allele, but it also could be due to incomplete penetrance. In the latter case the sheep would actually have inherited the OHC allele, but its effect would have suppressed by another gene. None of the 81 animals had the combination of an affected phenotype and a GG genotype. In the previous analysis of 238 animals only two had this combination, so it is not likely to occur often when testing animals and will not be a significant factor in the accuracy of the test. The expected error rate from recombination alone is 0.008.

As described in Chapter 2, there is evidence from breeding records of OHC having incomplete penetrance, and the results from the 81 sheep are consistent with this. The error rate of 16% is apparently high, but incomplete penetrance sets a limit on how accurate a genetic test for OHC could be. Even if the mutation itself was tested, about 20% of those animals with an OHC genotype would have a normal phenotype. Therefore 989G>A has the expected accuracy for a marker for OHC.

The accuracy of the test could be improved by breeding from animals that showed the phenotype predicted from the 989G>A genotype, and whose ancestors showed the same pattern. If the incomplete penetrance is due to the influence of other genes this would select against those alleles that suppress the OHC phenotype. It would also select against animals with recombinations between 989G>A and the OHC locus.

4.3.6 Conclusion

NUDT9 is probably not the OHC gene, but it contains a polymorphism, 989G>A, which is correlated with OHC status and can be used as a genetic marker.

4.4 *AFF1*

4.4.1 Introduction

The purpose of this section is to determine whether a mutation in *AFF1* is strongly linked to OHC status. The strategy for testing *AFF1* was the same as for *NUDT9*, except that primers were designed from both the ovine *AFF1* sequence and the bovine equivalent. Multiple sets of primers were designed to amplify the entire CDS, but only half of the sequence was successfully amplified. For the region where sequences could be amplified from normal cultured lens cDNA, the PCRs were repeated with cDNA from affected sheep. Affected and normal sequences were compared.

4.4.2 Methods

4.4.2.1 Expression of *AFF1* in Lenses

A sheep CDS sequence for *AFF1* was assembled using BLAST matches between the cattle CDS (accession number NM_001191525, from <http://www.ncbi.nlm.nih.gov/nucleotide>) and the sheep sequences available at

<http://www.ncbi.nlm.nih.gov/genome/seq/BlastGen/BlastGen.cgi?taxid=9940>. The sheep CDS was used to design primers using the online Repeatmasker (Tarailo-Graovac *et al*, 2009) and Primer3 (Rozen and Skaletsky, 2000) programs. Primer sequences were designed only to match areas of known sheep sequence. Seven pairs of primers were designed, and each primer was named with the name of the gene, the letters A-F, and an F for “forward” or an R for “reverse”. The primers designed to amplify the entire gene were named *AFF1F* and *AFF1R* (Table 28).

The lens RNA used in the following procedures came from normal cultured lenses UN, VN, and TN. As described in Section 4.3.4.1, the letter N means that the lens were not exposed to opacity-inducing substances and the other letters refer to the animal the lens came from.

These lenses were cultured and RNA was extracted as described in Section 4.3.3.1. , Leukocyte RNA from the affected sheep 47/07 which was used in *NUDT9* sequencing was also used.

10.5 µL of the RNA extract and oligo dT primers (Invitrogen) were used to prepare cDNA with Superscript III reverse transcriptase (Invitrogen). PCR was carried out using the cDNA

as a template, and the primer pair with the smallest product, AFF1FF and AFF1FR, was used (Table 28). The smallest product was chosen to ensure that the product would be sequenced at high fidelity. The PCR recipe and cycle was as described in section 4.3.3.2, except that an annealing temperature of 59°C was used. A 2µL sample of each PCR product was run on 2% agarose gel, 90V 1 h.

The PCR product was used to prepare two sequencing mixtures, one with the forward primer and one with the reverse primer. Each sequencing mixture contained 3.2 µL of a 1µM solution of the primer, the equivalent of 6.1ng of PCR product DNA, and sterile deionised water to 15 µL. The amount of DNA needed was calculated by allowing 2ng per 100 bases of the sequence. The sequencing mixtures were submitted to the Allan Wilson Centre Genome Service at Massey University, and sequenced as described in section 3.3.2.3. A BLAST search was performed to confirm that the sequence was from AFF1 cDNA and therefore that AFF1 is expressed in the cultured lens.

4.4.2.2 Sequencing of AFF1

The other primer pairs (Table 28) were used in PCRs to amplify regions of the AFF1 CDS. PCRs and agarose gels were carried out as in section 4.3.3.2. If there was no amplification or if the band was too faint, the PCR was repeated at an annealing temperature of 55°C. PCR products were purified with an AxyPrep PCR Cleanup Kit (Axygen).

The PCR products were sequenced as described in section 3.3.2.3. Each sequence was used in a BLAST search of the bovine genome to confirm that it matched only the AFF1 sequence. The DNAMAN (Lynnon) program was used to align each sequence with the bovine AFF1 CDS, and also its counterpart sequence generated from the other primer in the pair. The bovine sequence was used because the sheep sequence was too incomplete to align with most of the sequences. Sequences were cropped to the region of highest homology with the bovine sequence and each alignment was repeated. Sequences were assembled together using the Genedoc program (Nicholas *et al*, 1997), placing normal and affected sequences adjacent.

Only half of the AFF1 CDS was successfully sequenced using the original set of primers, so additional primers were designed and used in more PCRs. These were AFF1DNAF and AFF1DNAR, designed from bovine DNA sequence to amplify a 921-base region from genomic DNA, AFF1TOPF and AFF1TOPR, designed from bovine cDNA sequence to

amplify the start of the CDS from cDNA, and AFF1B2F, designed from bovine cDNA sequence to be used with AFF1DR to amplify a 1500-base section of the CDS from cDNA (Table 28). Two additional primers were designed from new sequence generated from the AFF1B primers and normal cDNA. These were named AFF12GAPF and AFF12GAPR (Table 28). The generation of this sequence is described in section 4.4.3.2.

Table 28: AFF1 Primers. The sizes predicted from bovine sequence (accession number NM_001191525) are given in bp. There are two forward primers designed for AFF1BR, and sizes for both sets are given. The positions of these primers are shown in Figure.

Name	Sequence	Size (bp)	T _m (°C)
AFF1F	AGCAGGTAGTCCCGTCACC	3986	60.1
AFF1R	AAACCATCCTGTCCAACCAA		60.2
AFF1AF	AGTCTCCACGCCAAAAGCTA	691	60.0
AFF1AR	GCCATGAGTGGGTCATTTTC		60.3
AFF1BF	GGCAGCAGACCTTTGAGAAA	661	60.5
AFF1B2F	GCTCCTAGCGAGTCTCCTGA	615	59.9
AFF1BR	TGACTTTGGTCAGCCAGTTG		59.9
AFF1CF	CCTCAGCACATTCCAGCAG	561	60.6
AFF1CR	CTTTGGAGGGCTGGTCCTT		61.5
AFF1DF	CTGCCTGCAGGTGGAGAG	546	60.7
AFF1DR	CTGGCGTTTTTCTGAGCTTCT		59.8
AFF1EF	CAGAAGACAAACAGCCACCA	800	59.9
AFF1ER	ACGCATGCATAAAACAGCAA		60.3
AFF1FF	GGAGACCGTGGACCTCATTA	303	59.9
AFF1FR	CTTGACTGGGACCCTATGGA		59.9
AFF1DNAF	ACGGCATTATGTGACGTGGT	1057	61.3
AFF1DNAR	ATCTTCCTTCCCAGCAAAT		60.0
AFF1TOPF	ATGGACGGCAGACCTTGA	301	60.2
AFF1TOPR	TTCCCCAGCCTATTTTCAGA		59.6
AFF12GAPF	AGTTCTGCAGTCCCAAACCA	153	60.7
AFF12GAPR	GAAAGGTGGCTTCTCTGAGG		59.0

4.4.2.3 Mutation Analysis of AFF1

Affected and normal sequences were compared to find candidate OHC mutations. The 47/07 affected sheep is assumed to be heterozygous for the cataract locus. Therefore if the AFF1 sequence contains the OHC mutation, the 47/07 cDNA will be a mixture of the normal and affected alleles while the normal cDNA will only have the normal allele. The sequencing results will show a two-base code for the cataract sequence and a one-base code for the same position in the normal sequence. The electropherogram should indicate that the two-base code is a true heterozygote with roughly equal peaks. It is also possible that 47/07 is a homozygote, in which case the cataract sequence will have one base and the same position in the normal

sequence will have another. In either case, the forward and reverse sequencing results and results from overlapping primer sets should be identical.

4.4.2.4 Comparison of AFF1 Exon Boundaries

The human AFF1 gene has 20 exons with 19 exon boundaries other than the start and stop codons. In order to compare the exon boundaries between human and sheep sequences, the human AFF1 CDS was aligned with the sheep cDNA sequence generated in this chapter.

4.4.3 Results

4.4.3.1 Expression of AFF1 in Lenses

The AFF1FF and AFF1FR primers were used in PCRs with normal lens and cataract blood cDNA. The agarose gel of the PCR products showed a single band in both the normal and cataract samples at just above 303 bases, consistent with the PCR product size predicted from the cattle CDS.

The sequences generated from both primers and both cDNA sources showed 96-97% homology when aligned with both the known bovine and sheep sequences, showing that the primers were amplifying the AFF1 gene from both normal lens and affected leukocyte cDNA.

4.4.3.2 Sequencing of AFF1

PCRs were performed using the AFF1F, AFF1R, AFF1AF, AFF1AR, AFF1BF, AFF1BR, AFF1CF, AFF1CR, AFF1DF, AFF1DR, AFF1EF, AFF1ER, AFF1FF, and AFF1FR primers. Bands of the expected size were generated for the primer pairs AFF1DF and AFF1DR, AFF1EF and AFF1ER, and AFF1FF and AFF1FR from both normal and affected cDNA (Figure 38). Primers AFF1CF and AFF1CR gave a band of the expected size with affected cDNA. With the best normal cDNA there was an additional larger band. Primers AFF1BF and AFF1BR gave a band of the expected size with normal cDNA, but also a band at about half the expected size. With the affected cDNA there was no amplification.

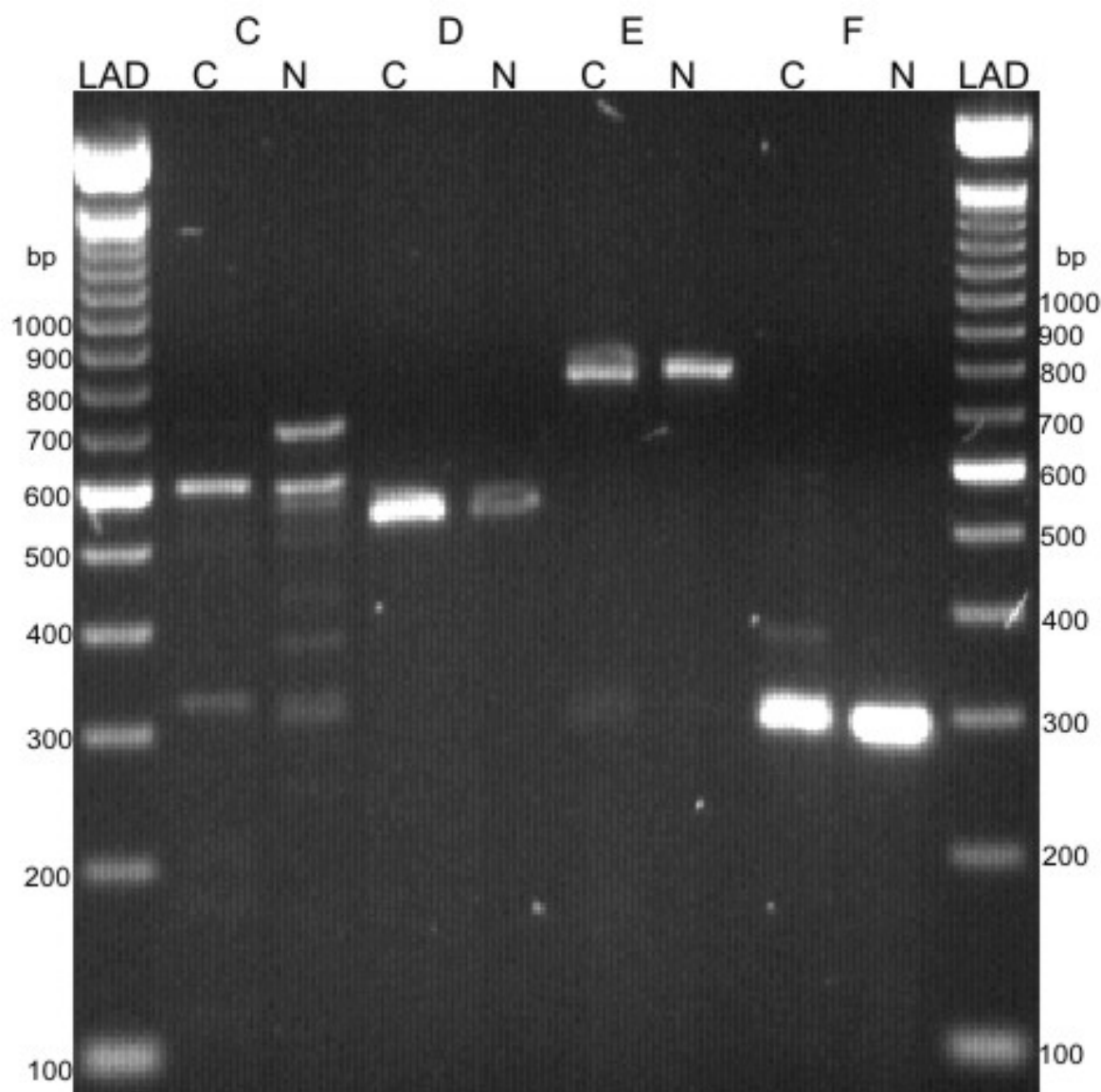


Figure 38: AFF1C-AFF1F PCR products. All PCR reactions also included blanks, which were all free of bands on agarose gels and are not shown here. The lanes labelled LAD contain a 100-base ladder. The top line shows the primer pair: AFF1C, AFF1D, AFF1E, and AFF1F. The second line shows the cDNA source, C for affected and N for normal.

In Figure 38, CC is the purified PCR product from primers AFF1CF and AFF1CR, using cDNA from affected sheep 47/07. CN is the PCR product from the same primers, using cDNA from normal cultured lens TN. The expected size for both samples is 561 bp, but the brightest bands in CC and CN are larger than 600 bp. This indicates that they are not the expected product, and this is confirmed by the sequencing and BLAST search performed below. A larger band is visible in CN and smaller, dimmer bands are visible in both.

DC is the purified PCR product from primers AFF1DF and AFF1DR, using cDNA from sheep 47/07. DN is the purified PCR product from the same primers, using cDNA from

normal cultured lens TN. The expected size for both samples is 554 bp, and the observed size is between 500 and 600 bp.

EC is the purified PCR product from primers AFF1EF and AFF1ER, using cDNA from sheep 47/07. EN is the purified PCR product from the same primers, using cDNA from normal cultured lens VN. The expected size for both samples is 800 bp, and this is the approximate size of the observed bands.

FC is the PCR product from primers AFF1FF and AFF1FR, using cDNA from affected sheep 56/07. FN is the PCR product from the same primers, using cDNA from normal cultured lens TN. The expected size for both samples is 303 bp, and the observed bands are at about 300 bp.

After checking combinations of primers for complementarity with each other using the DNAMAN (Lynnon) program, different combinations of primers were used with the same cDNA. The combinations tried were AFF1F- AFF1AR, AFF1AF- AFF1BR, AFF1BF- AFF1CR, AFF1FF- AFF1R, AFF1F- AFF1DR, AFF1BF- AFF1DR, and AFF1CF- AFF1DR. Only the AFF1FF- AFF1R primer pair gave an intense band at between 600 and 700 bp (Figure 39). The expected size is 757 bp.

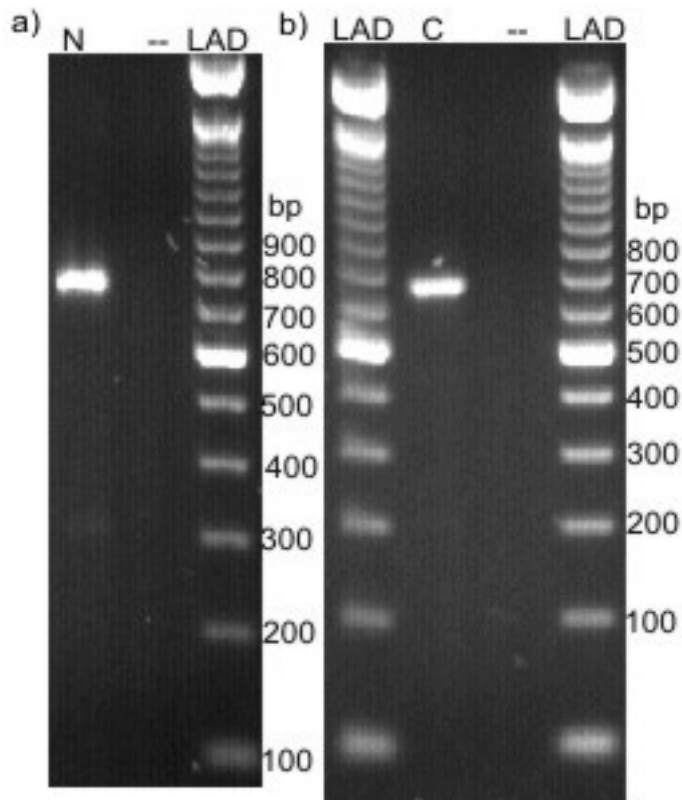


Figure 39: AFF1FF and AFF1R products. Lane N in gel a shows the product from normal cDNA TN. The -- symbol indicates the blank in both gels and L indicates a 100-base ladder. Lane C in gel b shows the product from cataract cDNA from sheep 56/07. The expected size of the band is 757 bp.

The PCRs were repeated with the primer combinations AFF1DNAF- AFF1DNAR, AFF1TOPF- AFF1TOPR, and AFF1B2F-AFF1DR. The DNA primers were used with genomic DNA and the others with cDNA. No amplification was observed. A product between 100 and 200 bp in size was generated with the AFF12GAP primers (Figure 40). The expected size is 153 bp.

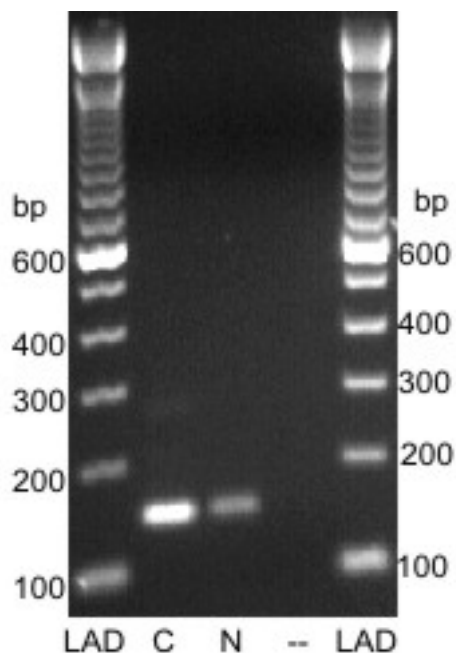


Figure 40: AFF12GAP PCR products. Lane C shows the product from cDNA from affected sheep 47/07. Lane N shows the product from cDNA from normal cultured lens HN. The expected size of the product from bovine sequence is 153 bp. The - - symbol indicates the blank and LAD indicates a 100-base ladder.

Sequences were generated from the AFF1B, AFF1C, AFF1D, AFF1E, AFF1F, AFF1FFR, and AFF12GAP PCR products. For the B primers the band at the expected size was cut out and purified with an AxyPrep Gel Extraction Kit (Axygen) before sequencing. Only the sequence from the reverse primer gave a sequence of adequate quality, so it was the only sequence used in the BLAST searches and DNAMAN (Lynnon) alignments. The searches and alignments showed that all of sequences matched the bovine AFF1 with a high degree of homology (Table 29), except for the AFF1C sequences. These matched a region of bovine chromosome 8 rather than the expected region on chromosome 6. It was concluded that the AFF1C primers were amplifying the wrong sheep cDNA so they were abandoned.

Mismatches between forward and reverse sequences, and between PCR product sequences and the bovine sequence, were mostly found at the 5' ends of each sequence where the sequence was of low quality with several ambiguous bases. Sequences were cropped to remove these areas and the alignments were repeated (Table). Figure 41 shows the alignment between the bovine AFF1 CDS and the AFF1DF sequence. There is a nine-base deletion between the bovine sequence and the PCR product sequence. This is identical in the affected and normal sequences.

Table 29: DNAMAN alignments of AFF1 sequences. The alignments are between cropped sheep sequences and bovine AFF1 cDNA, and between sequences from the same pair of primers.

Sequence	Cropped length (bp)	Bovine alignment	Pair
AFF12GAPF Normal	103	99%	100%
AFF12GAPR Normal	91	98%	
AFF12GAPF Affected	110	97%	100%
AFF12GAPR Affected	90	98%	
AFF1BR Normal	461	96%	
AFF1DF Normal	340	97%	98%
AFF1DR Normal	504	97%	
AFF1DF Affected	473	97%	99%
AFF1DR Affected	495	97%	
AFF1EF Normal	738	96%	99%
AFF1ER Normal	725	96%	
AFF1EF Affected	469	98%	99%
AFF1ER Affected	726	96%	
AFF1FF Normal	241	97%	99%
AFF1FR Normal	265	97%	
AFF1FF Affected	244	97%	98%
AFF1FR Affected	188	96%	
AFF1FF (with AFF1R) Normal	383	96%	98%
AFF1R Normal	696	96%	
AFF1FF (with AFF1R) Affected	618	96%	99%
AFF1R Affected	681	96%	

Fast alignment of DNA sequences BovineAFF1.txt and TN.AFF1DFcrop

Ktuple=2 Gap_penalty=7

Upper line: BovineAFF1.txt, from 2186 to 2534

Lower line: Ovine AFF1DF cropped, from 1 to 340

BovineAFF1.txt:TN.AFF1DFcrop identity= 97.06%(330/340)

gap=2.58%(9/349)

```

2186  CGGGAGCCCAAGCCGCGGTGCCACCCCCAGCGACCCGGAAGAAGCACCGCAGCGGCCCC
      |||
1     CGGGAGCCCAAGCCAGCGGTGCCACCCCCAGCSACCGGAAGAAGCACCGCAGTGGCCCC

2246  CCGAAGGCGCCCCGAAGGACGCGGCGGCAGAGGACCCGAGCCCCGAACACTTTGCGCTC
      |||
61    CCGAAGGCGCCCCGAAGGACGCGGCGGCAGAGGACCCGAGCCCCGAACACTTTGCGCTC

2306  GTTCCCCTGACCCAGAGCCAGGGCCCTGTCCGCGGTGGCAGCGGCGGCGCCGCCAGG
      |||
121   GTTCCCCTGACCCAGAGCCAGGGCCCTGTCCGCGGT.....GGCGCCGCGCCAGG

2366  ACTAGTGGCTGCAGGCTGGCTGTGGTGGTCCAGGAGGACCGCCGCAAAGACAAACCCCGG
      |||
172   ACTAGCGGCTGCAGGCTGGCCGTGGTGGTCCAGGAGGACCGCCGCAAAGACAAACCCCGG

```

```

2426 GTGCCTTCGAGAGACACCAAGCTGCTCTCCCCGCTCAGGGACACTCCGCCGCCACAAAGC
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
232  GTGCCTTCGAGAGACACCAAGCTGCTCTCCCCGCTCAGGGACGCTCCGCCACCACAAAGC

2486 TTGGTGGTGAAGATCACACTCGACCTCCTGTCTCGGATACCCCAGCCCC
      ||||||||||||||||||| |||||||||||||||||||||||||||||||||||
292  TTGGTGGTGAAGATCACGCTCGACCTCCTGTCTCGGATACCCCAGCCCC

```

Figure 41: AFF1DF sequence aligned with bovine AFF1 CDS, accession number NM_001191525 from <http://www.ncbi.nlm.nih.gov/nucore>. The AFF1DF sequence was generated from normal sheep cDNA from lens TN, and it was originally 521 bases long and was 88% homologous with the bovine AFF1 CDS. The sequence shown here was cropped to remove areas of poor homology at the 5' and 3' ends. It is 340 bases long and shows 97% homology with the bovine AFF1 CDS.

The sequence from normal cDNA and the AFF1B primers contained two 36-base regions not covered by the sheep sequence the other primers were designed from. One primer was designed from the sequence of each region (Table 28, above), and the primers were used in PCR reactions with normal and affected cDNA. The primers were named AFF12GAPF and AFF12GAPR and as described above they gave bands of the expected size and good quality sequence with both affected and normal cDNA.

The sequences generated from these primers were equivalent to about half of the CDS of AFF1 (Figure 42). There is no obvious relationship between the exon boundaries of the bovine CDS and the sections which could not be amplified.

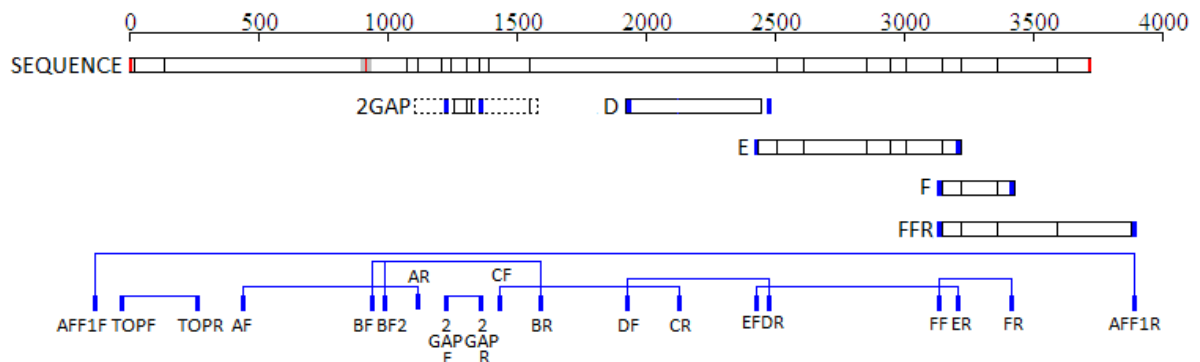


Figure 42: AFF1 sheep sequences. The bovine CDS is shown at the top. The scale gives distances in bp. Sequence that has been found in both affected and normal animals is shown with solid lines, whereas dotted lines show sequence only found in one animal. The maximum sequence given by combining affected and normal sequences is shown in Figure 43. Red lines indicate the start and stop codons and the homologous location of the mutation in robotic mouse AFF1. The gray region surrounding the mouse mutation is conserved in all known AFF1 homologues. Black vertical lines show the exon boundaries of the bovine sequence, and these are shown on the sheep sequences if they have been found in BLAST searches. Primers are shown in blue. Some primer combinations did not give high-quality sequence in the entire region between the primers. In that case gaps are visible between the primers and the ends of the sequences. The locations of all the primers are shown to scale, below the sequences, with horizontal lines connecting F and R primers in each pair.

GAGGCTGCGC TGAGGACGGC CGGGCTCCAG GAGCAGGTAG TCCCGTCACC TCGGGGAGTC
 GCGCTGGGGA ATCGTCCCCG CCCGGCCCCG GCAAAATGGTG AGCGCGGCGC CGGCAGGAGG
 GCCGCGGCTT GGAGGCGCTG ATGGACGGCA GACCTTGA GC TCTACAGACT GAATTATGGC
 AGCCCAGTCA AGCTTGTACA ATGAAGACAG AAACCTGCTT CGAATTAGAG AGAAGGAAAG
 ACGCAACCAG GAAGCTCACC AAGATAAAGA GGCATTTTCT GAAAAGATCC CCCTTTTTGG
 AGAGCCATAC AAGACAGAGA AAGGTGATGA GCTATCCAGT CGGATACAGA ACATGCTGGG
 AAACCTATGAA GAGATGAAGG AGTTCCTTAG TAGCAAGTCC CACCCCCAGC GCTTGGATGC
 TTCTGAAAAT AGGCTGGGGA AGCCGAGATA CCCTTTATTT CCTGAGAAGG GGAGCAGCCT
 TCCACCCAC TCCTTCCACA CCAGTGTCCA CCACCAGCCC GTTAACACTC CTGCCTGTGG
 GCCACATCCT GCTGGTAACA TCGGCCACCA CCCAAAGATG GCGCAGCCAA GAATGGAACC
 AATGCCAAGT CTCCACGCCA AAAGCTATGG CCCACCGGAC AGCCAGCACC TGACCCAAGA
 GCGCCTCGGT CAGGAGGGGT ACGGCTCTAG TCATCACAAA AAGGGTGACC GCAGGGCTGA
 TGGAGACCAC TGTGCTTTGA CGACAGACTC TGCTCCGGAG AGGGAAC TTT CACCTCTGTT
 CTCTTTGCC TCCCAGTCC CCCCTTTGTC ACCTGTACAT TCCAACCAGC AGACTGCTCC
 CAGGACGCAA GGAAGCGGCA AAGTTCCAG CAGTAGCAGT AGCAATAAAG GCTACTGCCT
 GGCCAAGTCT CCAAGGACC TAGCGGTAAG GGTGCACGAT AAAGAGACCC CTCAAAGCAG
 CACAGGGGCA GTGGCCAGCC TTGGGGTAGC CCCTCCTCAG CCACCTTCTC AGACCTTCCC
 CCCACCTCC CTCCCCTCAA AAAGCCTTGC AATGCAGCAG AAGCCCACGG CTTATGTCCG
 GCCCATGGAC GGTCAAGACC AGGCTCCTAG CGAGTCTCCT GA ACTGAAAC CCCTGCCGGA
 GGAGTACC GG CAGCAGACCT TTGAGAAATC AGACTTAAAA GTGCCTGCCA AAGCCAAGCT
 CACGAAGCTG AAAATGCCTT CTCAGTCAGT CGAGCAGACC TACTCCAATG AAGTCCATTG
 TGTGAAAGAG ATTCTGAAAG AAATGACCCA CTCATGGCCT CCTCCTTTGA CAGCAATACA
 CAGCCTAGT ACAGCTGAGC CCTCCAAATT TCCTTTTCCC ACAAAGGATT CTCAGCACAT
 TAGTTCTGCA GTCCCAAACC AAAAAACAATA TGATACATCT TCCAAAACCTC ACAGTAATTC
 TCAGCAGGGA ACATCCATGC TCAAAGATGA CTTTCAGCTC AGTGACAGTG AGGAGGACAG
 TGACAGCGAT CAAACCTCAG AGAAGCCACC TTTCGCACCT GCACCTCCAA GCGCTCCGCC
 GGCGCTGC CAGCAGCCGA GGCTCAGCA CATTCCAGCA GCGCCGAATC GGAGAGCACCAG
 CGACTCCG ACAGCTCCTC GGACTCGGAG AGCGAGAGCA GCTCCAGTGA CAGCGAGGAGAA
 CGAGCCCC GAGAAGCCCC GGCTCCCGAG CCCGAGCCTC CGACAACCAA CAAATGGCAGCT
 GGA CAACT GGCTGACCAA AGTCAGCCAG CCCGCGGTGC CCACCAGGAC CCCGGGTAGCAC
 GGAGCCCC CGGCCCGACC CCCGGACACT AAGGGCAAGG GCGGCGACAG CGCCACCGCGGG
 TCACGAGC GCCCAGAGTC CAAAGAGCCT CCCCCAAAG GCTCCAGCAA AGCCCCCGGGT
 CCCATCCG AAGCCCCCA CGCGGGCAAG AGGAGCTGTC AGAAGTCCC TACCCAGCAGGA
 GCCCCC AGCGGCAAAC CGTGGGAAGC AAACAGCCGA AAAAGCCCAT CAAGCTTCCGC
 CCCCGGGC ACGCGGATGC GCCTGCAGAT GCGCGTGC CT GCCTGCAGGT GGAGAGCGAGCC
 CGGACTTC CTCCCCGGC CTCC AAGGAC CAGCCCTCCA AAGACAAGCC CAAGGTGAAGAC
 GAAAGGGC GGCCCCGCGC CGCAGACAGC CGGGAGCCCA AGCCAGCGGT GCCACCCCCAG
 CGACCGGA AGAAGCACCG CAGTGGCCCC CCGAAGGCGC CCCCAGAAAG CGCAGCGGCAGA
 GGACCGGA GCCCGAACA CTTTGCCTC GTTCCCCTGA CCCAGAGCCA GGGCCCTGTCCG
 CGGTGGCG CCGGCGCCAG GACTAGCGGC TGCAGGCTGG CCGTGGTGGT CCAGGAGGACCG
 CCGCAAAG ACAAACCCC GGTGCCTTCG AGAGACACCA AGCTGCTCTC CCCGCTCAGGGA
 CGCTCCGC CACCACAAAG CTTGGTGGTG AAGATCACGC TCGACCTCCT GTCTCGGATACC
 CCAGCCCC CAGGGAAGGG CAGCCGCCCC AAGAAAC CAG AAGACAAACA GCCACCAGCAGG
 GAAGAAGC CAGATCCTGA GAAGAGAAGC TCAGAAAACG CCAGCAAGTT GGCCAAAAGAG
 AAAGGGTG AAGCAGAAAA AGACCACGAT AGCAAGAAAG TCAGGCTGGA GAAGGAAGTCAA
 GTCACAGC CGTCATCTTC CTCTTCCCAC AAGGAACCTT CCAAAATCAA AATGTCCAGGCC
 CTCCTCTG AGCCCTCAA GAAGGAAATG CTTCCCCCTT CACTCGTGTC GTCTCGTCTC
 CCAGAAGC CAGCCAAACC TGCACAGAAA AGGCCACGAC AGGAAGACCC CGGTGGCCAGGA
 CCCCCCA AAAGTGCCAG TAGTACCAAG GGCAGCCACA GAGACCCTTC TGCTCCAAACA
 CAGAAAAG CAGAGGGGAA GGTCTTTGGA AGCTCCACAG AGCACAAAGG ATCTTCTGGAGA
 TACTGCAA ACCGTTTTCC AGTGCCTTCT TTGCCAAATG GTAACCTCAA ACCAGGGAAGCC
 TCACGTGA AGTTTGACAA ACAACAAGCC GATTTTCACC TGAAGGAAGC CAAAAGTTGAA
 GGACAAAG CAGAATTAGT GACGGACAAG GTAGGGAAGG CCTTTAAATA CCTGGAGGCTGC

CTTGTCCT TTATCGAGTG CGGTGTCGCC CTGGAGTCGG AGAGCCCCGC GTCCAAGTCGGC
 GTACTCCA TCTACTCGGA GACCGTGGAC CTCATTAGT TCATATTGTC ATTAAAATCCTT
 CTCGGATG CCACAACGCC AACACAGGAG AAGATATTG CTGTTTTATG CATGCGTTGCCA
 GTCCATTT TGAACATGGC AATGTTTCGT TGTAATAAAG ATATGGCAAT AAAGTATTCCCG
 CACGCTTA ACGAACACTT CTTCAAGACT TCTTCCAAGG TTGCTCAGGC ACCTTCTCCATG
 CATCGCAA GAAGCACAGG CACACCGTCC CCCCTCTCTC CGATCCCTTC TCCGGCCAGC TC
 CATAGGGT CCCAGTCAAG CGCTGGCAGC GTGGGGAGCG GGGCGCTGCC CGCCACCATCAG
 CACCCCCG TCACCATCCA GAACATGACA TCCTCCTACG TCAGCATTAC TTCCCATGTCCT
 TACCGCCT TTGACCTGTG GGAACAGGCT GAGGTCTTCA CCAGGAAGAA CAAAGAATTCTT
 TGCTCAGC TCAGCACAAG TGTGTGCGCG CTGGCCCTCA ACAGCAGTTT GATGGATCTGGT
 GCACTATA CAAGACAGGG TTTCCAGCGG CTAATAACAAG TAACCAAAAC ACCTTAACGGAG
 GCCCAGGT GGATTCGATG CTTTGGGAAC TATTTTGCAC ATTGGAAGCC TCAAAACAGTCC
 AGACATTT GTCTCTTTAG AATACCAAAC TCAAGAAGAG AAGCACCATG AGATGGCCAGGA
 CCTTGTC A TACTGAAAC TCGAGGACTG TGGAATCGTT GGTGGACAG GATGGTTTGC
 GAAGCAGA GAGGAAGGAG GCTGGCCTGA GAGATGCTCT TGCCTTTCCT AACCAAAGGATG
 GAAGTGCA GTGTAGCCTG AGTGGGCGTA TTCTTGGTCT AGAAACATTT CCATATTTTT
 AFF1F/R
 AFF12GAPF/R
 AFF1DF/R
 AFF1EF/R
 AFF1FF/R
 AFF1TOPF/R
 AFF1AF/R
 AFF1BF/F2/R
 AFF1CF/R
 START/STOP
 MOUSE MUTATION

Figure 43: Sequence and primers for AFF1. Sheep sequence generated by the primers is indicated by underlining. The equivalent position to the mouse mutation is shown for reference. Sequences are included if they were the consensus between at least two different sequencing runs. Therefore slightly more sequence is shown than in Figure 42, which only shows the extent of sequences from both affected and normal animals. The rest of the sequence is the equivalent bovine AFF1 cDNA, accession number NM_001191525 from <http://www.ncbi.nlm.nih.gov/nucleotide>.

4.4.3.3 Mutation Analysis of AFF1

No candidate OHC mutations were found in the AFF1D, AFF1E, and AFF1F sequences, or in the sequences from the AFF12GAP primers.

4.4.3.4 Comparison of AFF1 Exon Boundaries

The human AFF1 CDS was aligned with the sheep CDS generated in this chapter to determine if there were any changes in the exon boundaries. The human sequence and exon boundaries came from NCBI Reference Sequence NM_005935.2 and the sheep sequence was the longer consensus sequence shown in Figure 43.

1939 ACGAAAGGACGGCCCCGGGCCGAGCAAGCAACGAACCCAAGCCAGCAGTGCCCCCTCC
 ||||| ||||| ||||| ||| || ||||| ||||| ||||| |||||
 898 ACGAAAGGGCGGCCCGCGCCGAGACAGCCGGGAGCCCAAGCCAGCGGTGCCACCCCC

 1999 AGTGAGAAGAAGAAGCACAAAGAGCTCCCTCCCTGCCCCCTCTAAGGCTCTCTCAGGCCCA
 || || ||||| ||||| || ||||| ||||| ||||| ||||| |||||
 958 AGCGACCGGAAGAAGCACCGCAG.....TGGCCCCCGAAGGCGCCC.....CCG

 2059 GAACCCGCGAAGGACAATGTGGAGGACAGGACCCCTGAGCACTTTGCTCTTGTTCCTTG
 || ||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1003 AAAGACGCAGCGG.....CAGAGGACCGGAGCCCCGAACACTTTGCGCTCGTTCCCTTG

 2119 ACTGAGAGCCAGGGCCACCCACAGTGGCAGCGGCAGCAGGACTAGTGGCTGCCGCCAA
 || ||||| ||||| || ||||| ||||| ||||| ||||| ||||| |||||
 1057 ACCCAGAGCCAGGGCCCTGTCCGCGGTGGCGCCGGCGCCAGGACTAGCGGCTGCAGGCTG

 2179 GCCGTGGTGGTCCAGGAGGACAGCCGCAAAGACAGACTCCCATTGCCTTTGAGAGACACC
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1117 GCCGTGGTGGTCCAGGAGACCGCCGCAAAGACAAACCCCGGTGCC TTCGAGAGACACC

 2239 AAGCTGCTCTCACCGCTCAGGGACACTCCTCCCCACAAAGCTTGATGGTGAAGATCACC
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1177 AAGCTGCTCTCCCCGCTCAGGGACGCTCCGCCACCACAAAGCTTGGTGGTGAAGATCAGG

 2299 CTAGACCTGCTCTCTCGGATACCCAGCCTCCCGGAAGGGGAGCCGCCAGAGGAAAGCA
 || ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1237 CTCGACCTCCTGTCTCGGATACCCAGCCCCAGGGAAGGGCAGCCGCCCAAGAAACCA

 2359 GAAGATAAACAGCCCGCCGAGGGAAGAAGCACAGCTCTGAGAAGAGGAGCTCAGACAGC
 ||||| ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1297 GAAGACAAACAGCCACCAGCAGGGAAGAAGCCAGATCCTGAGAAGAGAAGCTCAGAAAAC

 2419 TCAAGCAAGTTGGCCAAAAAGAGAAAAGGTGAAGCAGAAAGAGACTGTGATAACAAGAAA
 | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1357 GCCAGCAAGTTGGCCAAAAAGAGAAAAGGTGAAGCAGAAAGACCACGATAGCAAGAAA

 EXON 11 EXON 12
 2479 ATCAGACTGGAGAAGGAAATCAAATCACAGTCATCTTCATCTTCATCCTCCACAAAGAA
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1417 GTCAGGCTGGAGAAGGAAGTCAAGTCACAGCCG...TCATCTTCCTCTTCCACAAGGAA

 2539 TCTTCTAAAACAAAGCCCTCCAGGCCCTCCTCACAGTCTCAAAGAAGGAAATGCTCCCC
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1474 CCTTCAAATCAAATATGTCCAGGCCCTCCTCTGAGCCCTCAAAGAAGGAAATGCTTCCC

 EXON 12 EXON 13
 2599 CCGCCACCCGTGTC...CTCGTCTCCAGAAGCCAGCCAAGCCTGCACTTAAGAGGTCA
 || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1534 CCTTCACTCGTGTCTCGTCTCCAGAACCCAGCCAAACCTGCACAGAAAAGGCCA

 2656 AGGCGGGAAGCAGACACCTGTGGCCAGGACCCCTCCAAAAGTGCCAGCAGTACCAAGAGC
 ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1594 CGGCAGGAA...GACCCCGGTGGCCAGGACCCCCCAAAGTGCCAGTAGTACCAAGGGC

 2716 AACCACAAAGACTCTTCCATTCCCAAGCAGAGAAGAGTAGAGGGGAAGGGCTCCAGAAGC
 | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1651 AGCCACAGAGACCCCTTCTGCCTCCAAACACAGAAAAGCAGAGGGGAAGGTCTTTGGAAGC

 2776 TCCTCGGAGCACAAAGGTTCTTCCGGAGATACTGCAAATCCTTTTCCAGTGCCTTCTTTG
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1711 TCCACAGAGCACAAAGGATCTTCTGGAGATACTGCAAACCGTTTCCAGTGCCTTCTTTG

 EXON 13 EXON 14

```

2836 CCAAATGGTAACTCTAAACCAGGGAAGCCTCAAGTGAAGTTTGACAACAACAAGCAGAC
|||
1771 CCAAATGGTAACTCCAAACCAGGGAAGCCTCACGTGAAGTTTGACAACAACAAGCCGAT
EXON 14 EXON 15
2896 CTTCACATGAGGGAGGCAAAAAAGATGAAGCAGAAAGCAGAGTTAATGACGGACAGGGTT
|||
1831 TTTCACCTGAAGGAAGCCAAAAGGTTGAAGGACAAAGCAGAATTAGTGACGGACAAGGTA
EXON 15 EXON 16
2956 GGAAAGGCTTTTAAGTACCTGGAAGCCGTCCTTGTCCCTTCATTGAGTGCAGGAATTGCCACA
|||
1891 GGAAGGCTTTAAATACCTGGAGGCTGCCTTGTCCCTTATCGAGTGCAGGTGTCGCCCTG

3016 GAGTCTGAAAAGCCAGTCATCCAAGTCAGCTTACTCTGTCTACTCAGAACTGTAGATCTC
|||
1951 GAGTCGGAGAGCCCCGCTCCAAGTCGGCGTACTCCATCTACTCGGAGACCGTGGACCTC

3076 ATTAATTCATAATGTCATTAAAAATCCTTCTCAGATGCCACAGCGCCAACACAAGAGAAA
|||
2011 ATTAAGTTCATATTGTCATTAAAAATCCTTCTCGGATGCCACAACGCCAACACAGGAGAAG
EXON 16 EXON 17
3136 ATATTTGCTGTTTATGCATGCGTTGCCAGTCCATTTTGAACATGGCGATGTTTCGTTGT
|||
2071 ATATTTGCTGTTTATGCATGCGTTGCCAGTCCATTTTGAACATGGCAATGTTTCGTTGT
EXON 17 EXON 18
3196 AAAAAAGACATAGCAATAAAGTATTCTCGTACTCTTAATAAAACTTC...GAGAGTTCT
|||
2131 AAAAAAGATATGGCAATAAAGTATTCCCGCACGCTTAACGAACACTTCTTCAAGACTTCT

3253 TCCAAAGTCGCCAGGCACCTTCTCCATGCATTGCAAG...CACAGGCACACCATCCCCCT
|||
2191 TCCAAGTTGCTCAGGCACCTTCTCCATGCATCGCAAGGACACAGGCACACCGTCCCCC
EXON 18 EXON 19
3310 CTTTCCCCAATGCCTTCTCCTGCCAGCTCCGTAGGGTCCCAGTCAAGTGCTGGCAGTGTG
|||
2251 CTCTCTCCGATCCCTTCTCCGGCCAGCTCCATAGGGTCCCAGTCAAGCGTGGCAGCGTG

3370 GGGAGCAGTGGGGTGGCTGCCACTATCAGCACCCAGTACCATCCAGAATATGACATCT
|||
2311 GGGAGCGGGGCGCTGCCCGCCACCATCAGCACCCCGTACCATCCAGAATATGACATCC

3430 TCCTATGTCACCATCACATCCCATGTTCTTACCGCCTTTGACCTTTGGGAACAGGCCGAG
|||
2371 TCCTACGTCAGCATTACTTCCCATGTCCTTACCGCCTTTGACCTGTGGGAACAGGCTGAG

3490 GCCCTCACGAGGAAGAATAAAGAATTCTTTGCTCGGCTCAGCACAAATGTGTGCACCTTG
|||
2431 GTCTCACCAGGAAGAACAAGAATTCTTTGCTCAGCTCAGCACAAAGTGTGTGCGCGCTG
EXON 19 EXON 20
3550 GCCCTCAACAGCAGTTTGGTGGACCTGGTGCCTATAACAGCAGGGTTTTCAGCAGCTA
|||
2491 GCCCTCAACAGCAGTTTGGTGGACCTGGTGCCTATAACAAGACAGGGTTTTCAGCAGCTA

3610 CAAGAATTAACCAAAACACCTTAA
|||
2551 AAACAAGTAACCAAAACACCTTAA

```

Figure 44: AFF1 exon boundaries. The alignment between the AFF1 cDNA sequence generated in this chapter (Figure 43) and the human AFF1 CDS is shown. The human sequence is shown at the top and the sheep sequence at the bottom. Vertical lines connect identical bases. Sixteen exon boundaries are covered by the AFF1 sequence, the other three are 5' of the sequence shown here. The human exon boundaries are indicated by blue bases as in Figure

37. The number of each exon is labelled below the human and sheep sequences, on either side of each boundary. Where the sheep bases either side of an exon boundary are the same they are coloured blue, and substitutions and insertions are coloured red. The repeated N's show the gap in the sheep sequence, with the likely number of N's determined through alignment with the bovine sequence.

For 11 of the exon boundaries, the bases surrounding each boundary were identical in the human and sheep sequences (Figure 44). In 4 of the boundaries, one of the surrounding bases is different between the two species. As described in Section 4.3.5, these changes are not likely to affect the position of the exon boundaries. Three other boundaries are in the 5' section of the AFF1 CDS that has not been sequenced yet, and are not shown in Figure 44.

For the boundary between exons 18 and 19, the sequence AAG appears in the sheep sequence, between the bases that surround the exon boundary. This shows that one of the sheep exons is three bases longer than the corresponding human exon. The AAG sequence is either on the 3' end of the exon 18 or the 5' end of the exon 19. The same insertion is found when the mouse AFF1 CDS, from NCBI Reference Sequence NM_001080798.1, is aligned with the human CDS (Figure 45).

```

3253  TCCAAAGTCGCCCAGGCACCTTCTCCATGCATTGCAAG. . .CACAGGCACACCATCCCCCT
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3295  TCCAAAGTGGCCCAGGCACCTTCTCCATGCACGGCAAGAAGCCACAGGCGTTCCGTCCCCC
                        EXON18          EXON 19

```

Figure 45: Human-mouse AFF1 alignment. The human sequence is at the top and the mouse sequence at the bottom. The bases are coloured as in Figure 44. Only a 60-base section of the alignment including the AAG insertion is shown.

An insertion of three bases adds one amino acid residue to the protein, but does not change the reading frame and therefore leaves the rest of the protein unchanged. There are 11 other insertions or deletions between the human and sheep sequences (Figure 44), but they are all in multiples of three bases so they also do not change the reading frame. Going from the human sequence to the sheep sequence, 12 amino acid residues have been inserted and 10 have been deleted. This includes 8 amino acid residues in a row for which the corresponding bases have not been sequenced in sheep. Alignment with the bovine sequence suggests that there are 24 extra bases in this region compared to the human sequence.

A 240 base sequence including the AAG insertion in the sheep cDNA was used in a BLAST search of the sheep genome, assembly Oar1.0, to determine which exon the insertion belonged to (Figure 46). The sheep BLAST page at <https://isgdata.agresearch.co.nz/> was used.

```

cDNA   1          acgcttaacgaacacttcttcaagacttcttcca 33
          |||
Genome 111732493 acgcttaacgaacacttcttcaagacttcttcca 111732526

cDNA   34          aggttgctcaggcaccttctccatgcatcgcaag 68
          |||
Genome 111732527 aggttgctcaggcaccttctccatgcatcgcaag 111732560
                      EXON 18

cDNA   67          agaagcacaggcacaccgtccccctcttccga 100
          |||
Genome 111734988 agaagcacaggcacaccgtccccctcttccga 111735019
                      EXON 19

cDNA   101         tcccttctccggccagctccataggg 126
          |||
Genome 111735020 tcccttctccggccagctccataggg 111735047

```

Figure 46: Sheep genome BLAST search. In each block the top sequence is from the cDNA sequence generated in this chapter, and the bottom sequence is from the sheep genome assembly Oar1.0. The number of the exon is given below the genomic sequence. The numbers give the position of the nearest base from either the start of the 240-base query or the centromere of sheep chromosome 6. The insertion is shown in red and the two bases that surround the exon boundary in the human sequence are shown in blue. Only 126 of the original 240 bases are shown.

The cDNA sequence gave only two matches to the genome assembly, 2428 bases apart (Figure 46). In both matches all bases were identical between the query and the genomic DNA. The first match ends with the same sequence as the end of the human exon, and the insertion is not present. The insertion is present in the second match, along with the last two bases of the first match. This two-base overlap is probably due to two bases at the end of the intron coincidentally matching the cDNA. Therefore the insertion is present in exon 19 and not exon 18.

4.4.4 Discussion

4.4.4.1 Gene and Protein Details

The AFF1 protein is a transcription factor, part of a complex which stimulates RNA polymerase II to transcribe RNA from DNA and also maintains the chromatin in an open conformation (Bitoun *et al*, 2007). The details of which genes are targeted are still unclear, but mutations in AFF1 are known to have effects on lymphocyte development (Isnard *et al*, 2000), the Purkinje cells of the cerebellum (Isaacs *et al*, 2003), and the cell cycle of yeast (Sipiczki, 1988). AFF1 is an abbreviation of “affranchised”, from its affect on the cell cycle of yeast when mutated, letting the cell escape the part of the cycle associated with meiosis and return to normal metabolism (Sipiczki, 1988).

A chromosome translocation is present in 50% to 70% of infant leukemia cases which fuses human chromosomes 4 and 11 together (Bizarro *et al*, 2007). This joins AFF1 with the mixed-lineage leukaemia (MLL) gene to create a fusion gene. AFF1 is known to be critical for normal lymphocyte development (Isnard *et al*, 2000) and MLL is also known to be a transcription factor, and in the fusion protein both the DNA-binding site of MLL and the interaction domains of AFF1 are retained. The fusion of the two proteins is thought to cause leukaemia by disruption of the normal transcriptional regulatory pathways, as the complex promotes the transcription of MLL's usual targets rather than its own. (Bitoun *et al*, 2007). This is the origin of an alternative name for AFF1, AF4. This stands for "Acute lymphoblastic leukaemia Fusion gene from chromosome 4".

A mutation in AFF1 has been discovered in mice, which causes a jerky, ataxic gait due to adult-onset Purkinje cell loss, along with cataracts (Isaacs *et al*, 2003). The mutation is referred to as "robotic mouse" after the appearance of the gait. This mutation is dominant and the cataracts are bilateral and develop gradually, in a similar way to how cataracts develop in OHC. The dominant mode of inheritance also matches the genetics of OHC. There are no signs of ataxia in the sheep affected by OHC, and no evidence of Purkinje cell loss, but together with AFF1's position in the linked region, the mouse cataracts suggest that AFF1 is a strong candidate OHC gene.

The mutation in robotic mouse changes a T to a C in the second base of codon 280 (Isaacs *et al*, 2003). This changes a valine to an alanine in the amino acid sequence. The mutation occurs in a region which is highly conserved, in a sequence of 13 amino acids which show 100% homology with all known AFF1 homologues. The equivalent region in the bovine AFF1 protein sequence is between residues 295-307.

AFF1 is degraded in the body by the Ub-proteasome pathway, and the first stage of this pathway is an interaction between AFF1 and the ligases Siah1 and Siah2. The robotic mouse mutation has been shown to disrupt binding to Siah1 (Oliver *et al*, 2004), which suggests that the robotic mouse phenotype is due to excess mutant AFF1 building up in cells due to reduced degradation. This mechanism would work in both heterozygotes and homozygotes for the robotic mouse mutant, explaining the dominant mode of inheritance. OHC may be caused by a similar build up of mutant AFF1.

The latest CRIMAP analysis includes AFF1 in the region linked to OHC, and the latest multipoint analysis includes AFF1 in the 1-LOD support interval (Chapter 3).

4.4.4.2 Analysis of AFF1 as a Candidate for the OHC Gene

Only half of the CDS was amplified in this investigation. This is despite trials of several different sets of primers designed from both sheep and bovine sequences. No amplification was observed from primers designed from bovine sequence. The region of the AFF1 CDS that was amplified was derived from cDNA from normal cultured lenses and from affected sheep leucocytes and compared. This comparison could have produced false positives if any polymorphisms varied between the cataract flock as a whole and the commercial meat flock that the cultured lenses came from, without correlating with cataract status inside the cataract flock. However, any real cataract mutation would show up in this comparison.

No polymorphisms were discovered which correlated with OHC status. There were some variations in the sequence but none were consistent between different sequencing runs on the same source of DNA, so there was no evidence that they were polymorphisms rather than sequencing artifacts. However, since mutations in AFF1 cause autosomal dominant cataracts in mice, and the gene is found in the region of chromosome 6 linked to cataract, it remains the strongest candidate for the OHC gene. As described in Section 3.4.4, both the latest CRIMAP and latest multipoint analyses give a linked region that includes AFF1. No other gene in the linked region has a homologue that is known to be consistently associated with cataracts, so based on the results of this thesis AFF1 is still the strongest candidate.

As described in Section 4.4.1.1, a mutation is known at the second base of codon 280 of mouse AFF1 which causes cataracts. This mutation changes an amino acid residue inside a 13-residue region of the AFF1 protein which is conserved across all known AFF1 homologues. The conserved region is probably important for the normal function of the AFF1 protein. The OHC mutation may be in the homologous position to the mouse mutation, or it may be in another position inside the conserved region. The sequencing performed in this chapter did not amplify the homologue of the mouse mutation or any of the sequence coding for the conserved region. Therefore even though no polymorphisms correlating with OHC status have been found in the sequencing carried out so far, the equivalent of the mouse mutation cannot be ruled out as the OHC mutation.

Further work at Lincoln University, not by the author, has sequenced the entire CDS of AFF1 in a normal sheep not related to those affected by OHC, and also in a sheep affected by OHC. RNA was extracted from muscle tissue and used to prepare cDNA which was the template for PCR reactions using the primers designed in this investigation. Further primers were designed from newly generated sequence in order to amplify the entire CDS. The success of this work suggests that the initial problems with sequencing were due to problems with RNA or cDNA quality, possibly in combination with problems with primer binding. A single silent mutation was discovered between the normal and the affected sheep. This occurs outside the conserved region and is unlikely to be the OHC mutation. These results appear to rule AFF1 out as the OHC gene, but are based on analysis of only a single normal and a single affected sheep. More animals must be tested to conclusively exclude a mutation correlated with cataract status in the coding sequence of AFF1, and a mutation in the non-coding sequence may also be responsible for OHC.

4.4.3.4 Comparison of AFF1 Exon Boundaries

Of the 19 exon boundaries in human AFF1, 15 have identical flanking bases in the sheep cDNA sequence. Six exons have changed length due to insertions or deletions. One exon boundary has shifted three bases due to an insertion at the 5' end of one exon, and the positions of three boundaries are unknown because the sequencing in this chapter did not cover them. The insertion is also present in the homologous mouse sequence. This indicates that the insertion is a species difference separating mice and sheep from humans, not an error in the sheep sequence.

With 11 insertions or deletions and a number of point mutations, there are significant differences between the human and sheep AFF1 sequence. What effect this has on protein function would depend on the role of the amino acid residues. AFF1 is known to code for a functional transcription factor in humans (Bitoun *et al*, 2007). There has been no investigation of the role of AFF1 in sheep, so its function may have significantly changed. It is even possible that the changes have rendered AFF1 non-functional in sheep. However, since it is still expressed in sheep leucocytes and lenses, it is likely to still be functional.

4.4.5 Conclusion

2,039 bases of the AFF1 CDS have been amplified in both affected and normal sheep and no polymorphism correlating with OHC status has been discovered. If the AFF1 CDS is the same size as the bovine CDS, another 2,099 bases remain to be sequenced, so AFF1 cannot be ruled out as the location of the OHC mutation. Since AFF1 is inside the region of chromosome 6 linked to OHC, and is known to be associated with cataracts in mice, it remains a strong candidate for the cataract gene.

Chapter 5 General Discussion and Future Directions

5.1 General Discussion

5.1.1 Genome Scanning and Candidate Gene Analysis

As described in Chapter 3, a genome scan was performed on animals from the OHC flock, in order to determine the OHC locus. Significant linkage was detected between markers on sheep chromosome 6 and the OHC locus.

In order to restrict the position of the cataract locus further, a breakpoint panel was genotyped for new markers. With these new genotypes, the most likely interval for the OHC locus was determined to be between 131.1 and 132.9 cM from the centromere.

As described in Chapter 4, the interval 131.1 – 132.9 cM from the centromere of sheep chromosome 6 was used to select candidate genes for OHC by investigating the homologous region of the bovine genome. Sixteen genes were found in this region but only NUDT9 and AFF1 had plausible mechanisms for cataract formation.

NUDT9 lies within the region linked to OHC and is known to be expressed in the human lens. It codes for an enzyme with a substrate that controls a calcium channel and could therefore be involved in a calcium-calpain-proteolysis mechanism of cataract formation.

Sequencing of NUDT9 mRNA in both normal and affected lenses revealed a polymorphism, 989G>A, that correlated strongly with OHC. 989G>A changed the amino acid sequence of the NUDT9 protein, so it was initially considered as a candidate OHC mutation. Further investigation showed that the amino acid that was changed was not conserved in other mammals. Also, two affected animals were found with a genotype associated with normal animals, but were affected by OHC. This is not consistent with the OHC genotype, so 989G>A was ruled out as the OHC mutation and NUDT9 is probably not the OHC gene.

To test the accuracy of 989G>A as a marker, lambs were tested for OHC status and genotyped for the polymorphism. The genotypes correctly predicted the OHC status of 84%

of the sheep. Due to incomplete penetrance even the OHC mutation itself would show a similar number of mismatches. Therefore 989G>A has the expected accuracy for a tightly linked marker for OHC.

AFF1 is a gene located on sheep chromosome 6 which codes for a transcription factor. Both the latest CRIMAP analysis and the latest multipoint analysis include AFF1 in the region most likely to contain the OHC gene. Mutations in mouse AFF1 cause cataracts which are inherited in an autosomal dominant manner, just as OHC is. As the only gene known to cause cataracts in the linked region, AFF1 is the best candidate gene for OHC. In mice, the cataracts are caused by a mutation in the AFF1 gene which interferes with the degradation of the AFF1 protein. The protein builds up in both the brain and the lens, causing ataxia and cataracts.

There is no evidence of ataxia in the cataract sheep. If AFF1 is the gene responsible for OHC, either the protein builds up in the lens but not the brain, or the sheep brain is sufficiently different that the protein does not cause ataxia. It is also possible that a different mutation in AFF1 causes OHC in sheep via a different mechanism.

Half of the CDS of AFF1 has been sequenced in both normal and affected sheep in order to determine whether there are any polymorphisms which correlate with OHC status. Such polymorphisms would be candidates for the OHC mutation, particularly if they changed the amino acid sequence of the AFF1 protein. No polymorphisms which correlate with OHC status were discovered. Several different sets of primers have been designed from both bovine and sheep sequence to amplify the missing sections of the AFF1 CDS, but none have been successful in this investigation in either affected or normal animals. The missing sections do not correspond to particular exons.

Further work at Lincoln University, not by the author, has sequenced the entire AFF1 CDS in one normal and one affected sheep (Fredriksson, 2010, unpublished results). The RNA source was muscle tissue and a combination of the primers used in this investigation and new primers were used. Only a single silent polymorphism was detected. This was outside of the conserved region and is not likely to be the mutation responsible for OHC. No polymorphism that changes the amino acid sequence was detected between affected and normal sheep, which suggests that AFF1 may not be the gene responsible for OHC. However, samples from more

animals need to be sequenced and the non-coding regions of AFF1 need to be investigated before it is ruled out as the OHC gene.

After the candidate genes had been selected, the linkage analysis was repeated using 989G>A as a new marker. The most likely interval for the OHC locus in this analysis was 131.1 to 131.8 cM from the centromere. A multipoint analysis was carried out using the same data, and gave a 1-LOD support interval of 131.3 to 131.7 cM from the centromere. The candidate gene AFF1 was inside both intervals. Therefore it was still a plausible candidate for the OHC gene. Only five other genes are known in 131.1-131.8 cM interval, and none has a plausible mechanism for cataract formation.

Both the latest interval from the CRIMAP analysis and the latest 1-LOD support interval for OHC were outside the first two 1-LOD support intervals. This means that the original 1-LOD support intervals excluded what was subsequently determined to be the most likely position for the OHC locus. Sheep markers with positions determined by independent mapping were used to generate the first 1-LOD support interval, so any inaccuracy cannot be due to errors in the bovine assemblies. The second 1-LOD support interval was generated using markers positions from bovine assembly Btau3.1, but it is very similar to the first, 123.7-129.0 cM compared to 123.3-128.8 cM. These results show the limits of multipoint analysis, as it is always possible that the locus lies outside the 1-LOD support interval, which is only equivalent to a 90% confidence interval.

The three most challenging aspects of this part of the project were finding appropriate markers for the sheep linkage analysis from the bovine genome, finding accurate marker positions from the bovine genome as it was updated, and sequencing the AFF1 CDS. A total of 146 bovine markers were tested and only 6 were suitable for genotyping and inside the interval of interest. However, these new markers did significantly restrict the position of the OHC locus to 0.4 cM, so the fine mapping studies were a success.

The bovine genome assembly was revised twice during the course of this project. Assemblies Btau2.0, Btau3.1, and Btau4.0 were all used to select markers or to determine marker positions. Significant errors were reported in the contig orders of both Btau2.0 and Btau3.1, but at the time that they were used they were the most accurate assemblies available. No

errors in contig order have been reported for Btau4.0, and the latest results use marker positions from this assembly.

As discussed in section 3.2.4, OHC is likely to be inherited with incomplete penetrance and this was not taken into account in the linkage analysis. This means that the size of the linked region is likely to be larger than the results of the linkage analysis suggest. Therefore genes outside the linked region may be worth considering as candidates for the OHC gene if AFF1 is ruled out.

5.1.2 Breeding Studies

Breeding records and OHC statuses were studied in order to determine the manner of inheritance of the OHC trait. All cataracts in sheep are rare, so if OHC was recessive the number of carriers for it in the general sheep population would be very small. The proportion of affected offspring produced from mating between affected rams and unrelated normal sheep, 0.4, is not consistent with a small number of carriers for a recessive trait. Therefore the mode of inheritance is likely to be dominant. There is no significant difference between the rate of OHC in males and females, so the inheritance is likely to be autosomal dominant.

With both normal and affected dams, the proportion of offspring that are affected is significantly lower than expected for simple dominant inheritance, indicating another factor in the inheritance that reduces the proportion of affected offspring. A lethal homozygote effect would explain the proportion of affected offspring from affected dams. For the offspring of two heterozygotes, a lethal homozygote effect predicts that two-thirds of the surviving offspring will show the dominant phenotype, and the proportion of affected offspring is not significantly different.

However, a lethal homozygote would not explain the lower proportion of affected offspring of normal ewes. There is also no evidence of fewer lambs in the affected dams, which would be expected if a lethal homozygote killed lambs before birth. The animals with two affected parents do show a significantly higher postnatal death rate than those with one, but the difference is much smaller than would be expected from a lethal homozygote.

Incomplete penetrance would explain the lower proportion of affected offspring from both affected and normal dams. The penetrance of OHC has been calculated at 0.80 ± 0.04 for

normal dams and 0.83 ± 0.04 for affected dams. If the trait is showing incomplete penetrance the two figures should be identical, and taking uncertainties into account they are. Taking the weighted mean of the two penetrances gives a penetrance of 0.81 ± 0.04 for OHC overall.

The cause of the incomplete penetrance in OHC is unknown at present. The two main possibilities are other genes influencing the development of cataracts, and environmental conditions. All animals are raised at the same facility so the only variation in environment is from year to year. The variation in penetrance with birth years could be investigated along with environmental records to determine if a significant environmental effect on penetrance exists. There is no evidence for significant linkage to OHC in other regions of the ovine genome, but it is still possible that other genes contribute to the penetrance of OHC. Six of the 26 sheep chromosomes remained untested for linkage to OHC, and a partial genome scan of these chromosomes could identify any genes contributing to incomplete penetrance. Analysis of subsets of the pedigree for inherited variations in penetrance could also help to identify these genes.

5.1.3 Significance to Cataract Research

OHC is the only large animal model currently being used in cataract research. Although it has been used in trials of cataract prevention drugs (Robertson, 2003), the cause of the inherited cataract remains unknown. This investigation has discovered that OHC is linked to a region of sheep chromosome 6. The human homologue of this region does not contain any genes associated with cataract. Therefore OHC is probably not the sheep equivalent of any known human cataract. The results of this investigation suggest that AFF1 is the gene responsible for OHC. If this is confirmed, studies of the cataract mutation in AFF1 and its effect on the lens will help to determine the mechanism of cataract formation in OHC and its relevance to human cataracts. This investigation has also discovered a genetic marker for OHC which may improve the efficiency of cataract drug trials.

5.2 Future Directions

5.2.1 Genome Scanning

Genome scanning and linkage studies have restricted the position of the OHC locus to between 131.3 and 131.7 cM from the centromere of sheep chromosome 6, a distance of 0.4 cM. This is equivalent to between 111.6 and 111.9 megabases from the centromere, a distance of 300 kilobases. If the interval were to be restricted further, both animals with recombinations in the interval and new markers inside the interval would be needed.

The interval is between marker t3s9 and NUDT9, and only 10 animals have a recombination which could be inside this interval (Figure 20). None of these recombinations is definitely within the interval and most are probably outside. In order to restrict the interval further, more breeding would have to be carried out to produce animals with recombinations in the interval. Such animals would make up only a small fraction of the total offspring.

When the SPUTNIK program (Abajian, 1994) was used to find microsatellites on the bovine chromosome, they were an average of 19 kilobases apart in this area of the chromosome, and testing 83 markers yielded only 4 suitable for genotyping sheep. Therefore if the process is repeated with the latest interval the expected number of markers that will be found inside it is 15, and it is unlikely that any of these markers will be suitable for genotyping animals. The combination of a small number of animals with recombinations and a low probability of finding suitable markers makes further restriction of the interval using microsatellite markers impractical.

Single Nucleotide Polymorphisms (SNPs) are much more common than microsatellites. For example there are 128 known SNPs in the human NUDT9 gene alone. Microarrays containing probes for large numbers of SNPs in sheep have recently become available (Smith *et al*, 2010). Therefore it would be easier to use SNPs than microsatellites as markers to further restrict the interval, but the low number of animals with recombinations would still make this approach inefficient.

If this investigation was started at the time of writing (late 2010) the easiest approach to isolating the OHC locus would be to use a sheep SNP microarray to genotype large numbers of sheep from the OHC flock for linkage analysis. Using microsatellites instead makes genotyping slower and is less efficient due to the wider spacing of the markers, but this was the only method available for linkage analysis at the time the investigation started (2004).

5.2.2 AFF1 Testing

The entire CDS of AFF1 has been sequenced at Lincoln University, as described in section 5.1.1. No polymorphisms have been discovered which are likely to be the OHC mutation, but since only one affected and one normal sheep have been sequenced it is possible that errors in the sequencing are preventing the detection of polymorphisms. The sequencing should be repeated with enough affected and normal animals to conclusively determine whether there are any polymorphisms which are likely to be the OHC mutation.

If polymorphisms are found in AFF1 which correlate with OHC status, they should be tested on a large number of affected animals to determine whether any animals have unexpected genotypes that would rule out the polymorphisms as OHC mutations. Normal animals are less useful for these tests because incomplete penetrance means that they can have the OHC mutation without having cataracts. However, normal animals with an affected parent should also be tested to check that the proportion of OHC genotypes is consistent with the penetrance calculated from breeding studies. If all affected animals and enough normal animals have the expected genotypes, AFF1 is probably the OHC gene. Any change in the amino acid sequence of the AFF1 protein caused by the OHC mutation should be investigated to determine what change it would make to the function of the protein and how this might contribute to cataract formation.

If polymorphisms correlating with cataract are not found in the coding sequence of AFF1, the gene is a strong enough candidate that the introns and other non-coding regions should also be investigated for polymorphisms. If none of these regions show any suitable polymorphisms, the OHC gene may be one of the other five other genes known in the region linked to OHC, or an unknown gene in the same region. Because incomplete penetrance may be leading to an underestimate of the size of the linked region, genes outside the region may be candidates for the OHC gene if there is clinical or biochemical evidence linking them to cataract.

5.2.3 Large Scale Sequencing

If AFF1 has been ruled out as the OHC gene, an alternative to sequencing each of the other genes in turn is to sequence the latest 1-LOD support interval entirely, in one affected and one normal related sheep. Any polymorphisms between the sequences are candidates for the OHC mutation, and the sequence surrounding them can be separately amplified in several affected and normal animals to determine whether they correlate with OHC status. The smallest interval generated by the linkage studies, the 1-LOD support interval, is 300 kilobases in length, so generating a complete sequence will require large-scale sequencing.

A 650kb region of human chromosome 13 linked to leukemia has been sequenced by selecting BACs which covered the region, shearing them into fragments, cloning the fragments using transformed bacteria, and sequencing each cloned fragment (Kitamura *et al*, 2000). The fragments are assembled by computer software into a complete sequence of the region, with only seven BACs were required to cover the 650kb sequence. Libraries of sheep BACs are available (Dalrymple *et al*, 2007), so the same process can be used to sequence the latest 1-LOD support interval.

Another option for sequencing the linked region is next generation sequencing, which as described in section 1.9.6 uses a massively parallel approach to sequence up to 2 kilobases at once (Turner *et al*, 2009). This technology is rapidly advancing and it may soon be practical to sequence the entire 300 kilobase linked region in several animals and look for polymorphisms that correlate with OHC status. Next generation sequencing also allows cheaper sequencing of short sequences in large numbers of samples. This may make it practical to use direct sequencing as a genetic test for OHC once the OHC gene is identified.

5.2.4 Biochemistry of OHC

If AFF1 is the OHC gene, the mechanism of cataract formation in sheep may be similar to the mouse cataract caused by the same gene. The biochemistry of OHC lenses should be investigated to confirm that mutant AFF1 protein is present, and to investigate whether it is being degraded normally. Western blots and other antibody techniques can be used to detect the presence of AFF1 protein in the lens, and to estimate its concentration. Any results should be compared to normal lenses from related sheep. The AFF1 protein should be extracted from both affected and normal lenses and sequenced to confirm that predicted changes in the amino acid sequence occur. OHC does not appear to include the ataxia that is a feature of the

robotic mouse trait. This may be due to a different mechanism of cataract formation in OHC, or a slower development of ataxia in sheep that is not detected before the animals are culled. Neurological studies could be carried out on the affected sheep to resolve this question.

5.2.5 NUDT9 Marker

989G>A has been shown to be an accurate marker for OHC status. This marker can be used to identify affected animals from a venopuncture sample, before the cataracts develop. In future drug trials, animals could be tested soon after birth and only those which the test predicted would be affected could be included in the trials, which could start before the development of cataracts. When affected rams are bred with normal ewes and the NUDT9 marker is used to select animals with OHC genotypes, 79% of the animals develop cataracts. Since only 40% of all the offspring with normal ewes develop cataracts, this is a significant improvement in terms of avoiding testing drugs on animals which will not be affected.

The accuracy of the test can be further improved by breeding from subsets of the pedigree where the 989G>A genotype consistently predicts the OHC phenotype. This selects against any alleles responsible for incomplete penetrance and also against recombinations between 989G>A and the cataract locus.

Affected rams are often bred with normal, unrelated ewes that may carry the OHC-linked allele of 989G>A, and these ewes are usually not genotyped for any markers. Assigning a haplotype for 989G>A and the cataract locus by analysing the genotype and phenotype of both parents would also improve the accuracy of the test. The OHC-linked allele of 989G>A must be on the same chromosome as the OHC locus to predict the formation of cataracts.

5.2.6 Breeding an OHC Homozygote

Despite several generations of inbreeding, there is no evidence of any animal homozygous for the OHC gene. Homozygotes are currently only detectable through the proportion of affected offspring, so any affected animal with two affected parents may be a homozygote, if they have produced too few offspring to be ruled out. 989G>A can be used to find homozygotes, because an animal homozygous for the marker is more likely to be homozygous for the gene.

When the OHC gene has been identified, it can be tested in any probable homozygotes to confirm that they carry two copies of the OHC allele. Both tests can be carried out soon after

birth, without having to breed from the animal or even test the eyes for cataracts. Homozygotes are much more likely to be discovered by this method. Once a homozygote has been identified, it can be used to produce affected offspring more efficiently than a heterozygote. Allowing for incomplete penetrance, a homozygote animal mated to normal animals should produce 78%-86% affected offspring, compared to 40% from a heterozygote. This higher rate of affected offspring will make drug trials more efficient in a similar way to the 989G>A genetic test for OHC. The two methods can be combined for maximum efficiency in the production and identification of affected animals. The OHC mutation itself may be used as a genetic test for cataract, but it may need sequencing to genotype. 989G>A requires only PCR and a restriction digest and may be a cheaper and faster method of predicting cataract status even if the OHC mutation is known.

Appendix A: Genome Scan Markers

Table 30: Markers used in partial genome scan. Chr indicates the sheep chromosome and cM shows the position along it in centiMorgans. The primer sequences are shown where available. Informative markers were heterozygous in the ram and were used to genotype his offspring.

Marker	Chr	Position	Forward Primer	Reverse Primer	Inform?
BMS835	1	46.1	TCATGTGCATGGGGTTTG	ATCTGCCTACCTGGGCATC	YES
ILSTS44	1	68.9	AGTCACCCAAAAGTAACTGG	ACATGTTGTATTCCAAGTGC GAGGTAAGTCTA	YES
MCM120	1	87.7	TAGCTGTCAGGCCGTGCAGT	TAAGGCTCCCCTG	NO
ILSTS29	1	97.7	TGTTTTGATGGAACACAGCC CTGGGTCTGTAT	TGGATTTAGACCAGGGTTGG CAGAACAATAAAC	YES
MCM58	1	114.0	AAGCACGTCTCC	GCTAAACCAGAGC	YES
BMS963	1	130.9	GGAGGATGAAGGAGTCTTTGG	AATTTACCACAGTCCACCGC	YES
BM6438	1	150.0	TTGAGCACAGACACAGACTGG	ACTGAATGCCTCCTTTGTGC	YES
MCMA8	1	181.3	TATTGACATTTTTTATTCATTTGC	AGGAATTCCTTTTTGTATGG	NO
BMS574	1	157.5	ATGTTCTTTGACCACATGGATT	GAACAAGCATTCTGACCATAGC	YES
CSSM04	1	200.9	ATGCGTCTAGAACTTGAGATTG GACATGACTAAAAG	GAAATCATCTGGTCATTATCAGTG	YES
DB6	1	218.6	CAATTTAGCATGC	TGGACTACAGTCCATAGCGTCTC	YES
MCM137	1	233.4	AGGGGAGCCCCAGTAAAGTATCA	AAACAAGTGGGGATGTTAGCTCTT	YES
MCM130	1	256.5	AAACTTTGTGCTGTTGGGTGTATC	CTCACCTCTGCCTTTCTATCTCTCT	YES
BMS1789	1	284.0	CTGGAAACTGGAACTAGTGGG	GTGAGGCATTATCAAGAAGCTG	YES
BM1824	1	294.1	GAGCAAGGTGTTTTCCAATC	CATTCTCCAACGTCTTCCTTG	YES
EPCDV01	1	325.5	GTTGCCTCTCAAAGTGCCTC	GGCCTTATGGAAGATTTATGCT	NO
EPCDV13	1	348.4	CCATCTGTGCTGTCAAGTCA TAGAGAAGGTGG	GAGCCTGGTGGGTTATAGT AAACAGCAGTTG	NO
CSR65	2	14.4	CTGATGTCAGAG TCCGATGTTAGA	CTTGGGTGTACG AGCTGGTATCTG	NO
MCM147	2	39.8	TGACTTTTTGTGC	TGTCTGTCAATCC	YES
BM3412	2	52.5	GTCCCAGCAGACATCACATG	ACAAAGAATATGTGCAGCAAGC	YES
BMS1341	2	74.8	CCTACCTACTGCACAGTTTTGC	CTCCCATATAAGTTACCCACCC	YES
BL1080	2	96.0	TTCTGAATGCACCCTTGTTAG	CTGGGCACTAACTAATCCTGG GATAGAAATATA	YES
BMS1591	2	121.5	GACAAGATAGGCTTTGCATG GCTGTAAGAATC	CCAGGAGCTCACA CCTGATACATGC	YES
BM81124	2	143.8	TTCATTAAGCACT	TAAGGTTAAAAAC	YES
INRA40	2	151.9	TCAGTCTGGAGGAGAGAAAAC	CTCTGCCCTGGGGATGATTG ACTCTGGGTATG	YES
TEXAN2	2	160.2	ACATTGTCATGTGGTTGCTAAC	TATATGTGCAAG	YES
ILSTS30	2	180.5	CTGCAGTTCTGCATATGTGG AAATGTGTTTAAAG	CTTAGACAACAGGGGTTTGG GGAAAACCCCCA	YES
FCB20	2	194.2	ATTCCATACAGTG	TATATACCTATAC	YES
BMS1126	2	215.3	AGCCAGCAGCAATCAAGG TGGGAAGACCTC	GCCAGCATCAAGTCAGCTC TFACTTAAACAT	YES
CSSM53	2	236.1	AAAGAGAAATGT CTGAAGTGAAG	TAAAACAGAGGG GGAATTAGAATATC	NO
MCM512	2	250.2	GAAAGGGGACAC	ATTCCTTCATCGTG	YES
INRA135	2	272.8	TCACTGGAAAGGGCTATTGC	GGGGTTGCAGAGGGTAAGAC	YES
BMS356	2	280.0	ACCTCAGAGATGACGCAAGG	TTGAAGTTTTTGTGCTGTTTGG	NO
MCMA4	2	295.4	AGATGCGTGCCTGCTGAC	GTCACAGACCAGGAGCAGG	YES
ILSTS45	3	32.0	TTCTGGCAAACCTATTCCACC	CATGAAAGACACAGATGACC	NO
BM746	3	45.4	CTATCTGTTTCCCCAGCTTCC GCGACTTAGCA	GCCCAGGTTAATGACAGAGC CATCAAGAGATGAAT	YES
FCB129	3	73.1	GCAGCAGCATCC	GAGTAAAGAAGATG	YES
TGLA67	3	89.9	GTGTCAAGTAAGACTGTTCAAA	TATGGAATTGCAAAGAATTGGAT	YES
INRA131	3	111.5	GGTAAAATCCTGCAAAACACAG	TGACTGTATAGACTGAAGCAAC	YES
BM304	3	133.1	CTGGTGTTCCTTTTATATCAACC	GGCACGTAACCTGTAAAACC	YES
BMS2569	3	153.0	AGAGAGGCCAAAGCTGGG	TTTCCTTGGGCTTCAGGAG	YES
ILSTS42	3	159.2	AGGATTCTTTACCTCCAAGG GACCTGACCTT	TGGCTCTTCGTATCAGGTGG	NO
FCB5	3	186.3	ACTCTTCTCACTC	AAGTTAATTTTCTGGCTGGAAAACCC	YES
LYZ	3	204.1	AGTGAGCAAGGCATTACAGCATG	ATGGATGTTACCAGTTATTGTGGT	YES

CP43	3	215.7	TATGGCTGAGGA ATATTTTCTGTGTTG	GAATGGATAAAG AAGATGTGAGACTGAG	YES
CSR111	3	234.7	CCATCACTCCCAAAGGATATAGTA	ACAGTAAGAGGTAGAAAAGGAGGCA	YES
BMS1248	3	251.2	GTAATGTAGCCTTTTGTGCCG	TCACCAACATGAGATAGTGTGC	YES
CSAP39E	3	268.7	CACCAAAGAAGACAAGGGGAAAC	CAGCAACTCTGCTAAAAGCAAACA	NO
CSSME76	3	290.0	CGACTTTGGAATTGCTAAGTAG ACAAGCCATGCTG AATGGTCTTATGTG	GTTTTTCTCCCAGCTCCC	NO
AGLA22	3	313.4	ACGTCCAGATTCAGATTTCTTG	GAAATGCAGTGTGTGTCAGCCAAGC	YES
BMS1788	4	2.3	GATCCTAGCATCAGTCTCCAGATG	GGAGAGGAATCTTGCAAAGG	YES
MCM218	4	26.5	GCAGGACTCTACGGGGCCTTTGC	CACTAAAAGCTTATGAAAGTTCCAGC	YES
MAF70	4	61.4	GTAGACTACTCATGA AAATCAGGTCTTAGG GGCCTAACAGAATT	CACGGAGTCACAAAGAGTCAGACC	YES
MAF50	4	69.2	CAGATGATGTTGC AATTGCATTAGTAT CTTTAACATCTGGC	GGGACATGCAGCTATACACTTGAG	YES
CP26	4	98.7	AGTCTCATGGATTTGCGTCC CGTTCCCTCACTATG GAAAGTTATATATGC GCTTTAAGAACTG	GTCACCATACTGACGGCTGGTCC ATGAAAATATAAGA GAATGAACCACACGG	YES
HH35	4	114.9	AGTCTCATGGATTTGCGTCC CGTTCCCTCACTATG GAAAGTTATATATGC GCTTTAAGAACTG	ATTCAGCCAGGATGGAAGG CACTCTATTGTAAGA ATTTGAATGAGAGC	YES
BM3212	4	137.8	CTCTCAAGGCAC GCCCCTGAAGGAATGGTG	CTGGGTGAGAAATGGGTAGTAGG CCAAAAGGTCCTATCTCCAAA	YES
HH64	4	143.6	CCAGCAGAAGAGAAAGATACTGA AGCATGTATCCAATCTACCCTG	AGTGGTAGAACTCCATCTCACA GCAAGCCCTAGGCTGAAA	YES
TGLA176	5	17.8	GTTGACTTGTTAATCACTGACAGCC AATCCAGTGTGTG AAAGACTAATCCAG	CCTTAGACACACGTGAAGTCCAC GTAGATCAAGATATA GAATATTTTTCAACACC	YES
BM741	5	54.5	GTCCATTGCCTCAAATCAATTC TGAGCAGCACACAACCCAAATATT CATGGAGTTGTA GAGTCAGACATGA	AAACCACTTGACTACTCCCCAA TCTTCTTACATAAAATGCGTGCGC	YES
BMS2258	5	88.0	CCCCTTTCTGGAAGGCAGAAATG CAAATTGACTTATCCTTGCGTG GTAACGAGGCAT GGTGCCTACAAC CGGGATGATCTTC	GAGCAAAGGTCATGTCAGGTGT CTTATTGTGTTTTCTGCCAGGGAG TGTAACATCTGGGCTGCATC CCAACATCTTGC TATGTGCATGTATG CATTTGCTTTGGCT TCAGAACCAGAG	YES
BMS792	5	92.7	TGTCCAAATATGC CAAAGAATTGGCC ACAAGTGAAGTGC	GCCTGCTTTCTTTGCGTACAACCTG GCAAAGTCTAGGTGAAATGCC	YES
TGLA137	5	113.0	GAGCAGACTGGATGTGGGGA CCAAACCCTAGTTCACAGCAAT GGTATTTTGAGAATGTGGGC	CCTTGTCTTTCCCTTCAAGAAC AAGGGATCTGCAAGTAATGGTGGT TCTTTGACCACTACCTATCC	YES
AE129	5	116.3	CACTAGCACTACCTGGGAAGCC TTCTTGCTGGAAGCAGAGG CTACAGCTCTGATGAGAACC	CAGGTTCTTTACCACTCCACCTG CAAAGAAGCCAGGAAAGTATCGT CGTTTTCTCAAACCTTCATTGCC	YES
MCM527	5	125.4	GCTTCAAGTGTATAGCAAAGTATG TTAAGAATTCTCTCCCTCCC CATGCTGTTTTATTC TTGTTTGTATGTGC	GATCTGAAAAGAAATGAATAGATGT TGACCTCATACAGATTGATGTTG	YES
CSR134	5	140.7	CACTTTTCTGGAAGGCAGAAATG CAAATTGACTTATCCTTGCGTG GTAACGAGGCAT GGTGCCTACAAC CGGGATGATCTTC	YES	
MCM53	6	29.7	TGTCCAAATATGC CAAAGAATTGGCC ACAAGTGAAGTGC	GCCTGCTTTCTTTGCGTACAACCTG GCAAAGTCTAGGTGAAATGCC	YES
JMP36	6	62.1	GAGCAGACTGGATGTGGGGA CCAAACCCTAGTTCACAGCAAT GGTATTTTGAGAATGTGGGC	CCTTGTCTTTCCCTTCAAGAAC AAGGGATCTGCAAGTAATGGTGGT TCTTTGACCACTACCTATCC	YES
BM4621	6	88.0	CACTAGCACTACCTGGGAAGCC TTCTTGCTGGAAGCAGAGG CTACAGCTCTGATGAGAACC	CAGGTTCTTTACCACTCCACCTG CAAAGAAGCCAGGAAAGTATCGT CGTTTTCTCAAACCTTCATTGCC	YES
JMP4	6	112.3	GCTTCAAGTGTATAGCAAAGTATG TTAAGAATTCTCTCCCTCCC CATGCTGTTTTATTC TTGTTTGTATGTGC	GATCTGAAAAGAAATGAATAGATGT TGACCTCATACAGATTGATGTTG	NO
JMP8	6	135.5	CACTTTTCTGGAAGGCAGAAATG CAAATTGACTTATCCTTGCGTG GTAACGAGGCAT GGTGCCTACAAC CGGGATGATCTTC	CTGCTACTGTTAACCTCCCTCC ATAAGGAGAATCTGAAGAGCCAAG CACTGCATGCTTTTCCAAAC AACTGAGCTGTATGGTGGACG	YES
JMP12	6	152.2	GGCAAGCTAGAGTCAGACACG GAGCAGACTGGATGTGGGGA CCAAACCCTAGTTCACAGCAAT	AACTGAGCTGTATGGTGGACG ACATGACAGCCAGCTGCTACT AATTTAATGCACTGAGGAGCTTGG	YES
BL1038	6	152.2	GGTATTTTGAGAATGTGGGC CACTAGCACTACCTGGGAAGCC TTCTTGCTGGAAGCAGAGG	CTAAAATGCAGAGCCCTACC GATCATCTGAGTG TGAGTATATACAG GTTCTCCATTGAACCAACTTCA	YES
LS55	7	47.9	CTACAGCTCTGATGAGAACC GCTTCAAGTGTATAGCAAAGTATG TTAAGAATTCTCTCCCTCCC CATGCTGTTTTATTC TTGTTTGTATGTGC	CTGCTACTGTTAACCTCCCTCC ATAAGGAGAATCTGAAGAGCCAAG CACTGCATGCTTTTCCAAAC AACTGAGCTGTATGGTGGACG	YES
CYP19B	7	82.0	GATCACCTTGCCACTATTTCT ACACAAATCCTTTCTGCCAGCTGA GCTTGCTACATGGAAAGTGC GATCTTTGTTTCA ATCTATTCCAATTTTC	CTAAAATGCAGAGCCCTACC GATCATCTGAGTG TGAGTATATACAG GTTCTCCATTGAACCAACTTCA	YES
ILSTS70	7	107.0	GAATCCCATCACTCTCTCAGC CATCTTTCAAAG AACTCCGAAAGTG	CTGCTACTGTTAACCTCCCTCC ATAAGGAGAATCTGAAGAGCCAAG CACTGCATGCTTTTCCAAAC AACTGAGCTGTATGGTGGACG	YES
MCM149	7	118.6	GCGCAAGTTTCTCATGC TGATTTAGATGCT TTGCTAATGCCA	CTGCTACTGTTAACCTCCCTCC ATAAGGAGAATCTGAAGAGCCAAG CACTGCATGCTTTTCCAAAC AACTGAGCTGTATGGTGGACG	YES
MCM156	7	120.5	CGTATCAACTAACACGATGCTG TGTATAGCACAGGGAAGTCTTC	GGATTCTTTACCACTAGCCCCACCT ATTAACCTTTGTGGCATCTGAGC TGTGTGAATTACATTGGATCTGTC	YES
INRA127	8	17.2	CGTATCAACTAACACGATGCTG TGTATAGCACAGGGAAGTCTTC	YES	
RM216A	8	31.6	CGTATCAACTAACACGATGCTG TGTATAGCACAGGGAAGTCTTC	YES	
MCMA5	8	48.8	CGTATCAACTAACACGATGCTG TGTATAGCACAGGGAAGTCTTC	YES	
CP21	8	64.9	CGTATCAACTAACACGATGCTG TGTATAGCACAGGGAAGTCTTC	YES	
CSR129	8	86.0	CGTATCAACTAACACGATGCTG TGTATAGCACAGGGAAGTCTTC	YES	
BM4208	8	102.8	CGTATCAACTAACACGATGCTG TGTATAGCACAGGGAAGTCTTC	YES	
BMS1967	8	132.8	CGTATCAACTAACACGATGCTG TGTATAGCACAGGGAAGTCTTC	YES	
ETH225	9	9.4	CGTATCAACTAACACGATGCTG TGTATAGCACAGGGAAGTCTTC	YES	
CSSM66	9	24.2	CGTATCAACTAACACGATGCTG TGTATAGCACAGGGAAGTCTTC	YES	
ILSTS11	9	40.1	CGTATCAACTAACACGATGCTG TGTATAGCACAGGGAAGTCTTC	YES	
MAF33	9	60.0	CGTATCAACTAACACGATGCTG TGTATAGCACAGGGAAGTCTTC	YES	
BM302	9	70.3	CGTATCAACTAACACGATGCTG TGTATAGCACAGGGAAGTCTTC	YES	
MCM42	9	78.4	CGTATCAACTAACACGATGCTG TGTATAGCACAGGGAAGTCTTC	YES	
BM4513	9	100.2	CGTATCAACTAACACGATGCTG TGTATAGCACAGGGAAGTCTTC	YES	
RJH1	9	126.2	CGTATCAACTAACACGATGCTG TGTATAGCACAGGGAAGTCTTC	YES	
MNS64	10	22.7	CGTATCAACTAACACGATGCTG TGTATAGCACAGGGAAGTCTTC	NO	
CSSM48	10	32.3	CGTATCAACTAACACGATGCTG TGTATAGCACAGGGAAGTCTTC	NO	

ILSTS56	10	61.4	GCTACTGAGTGATGGTAGGG CAGTATTCTTGC	AATATAGCCCTGGAGGATGG	YES
MCM470	10	70.0	CTGGAACCTCCA CTATACAGTTTCTG	TCAGAGCAACTTTCTAAGACATCAA	YES
DB3	10	93.2	CTGTATAGNAAAGTG CATCAAAACAGTGA	CAGGTGCCAAGTATTACATGC CGAATCTCTTCTAG	YES
TGLA94	11	27.7	AGGATGATTGCCAG	GGATTGAGACTGTG	YES
SRCRSP6	11	59.9	CATAGTTCATTACAAATATGGCA	CATGGAGTCACAAAGAGTTGAA	NO
BM17132	11	65.4	ATCTGCCAGTATCACATCAACA	GTTACTTTTCCAGGCATGAAGC	YES
MCMA24	11	73.7	TGACCTCAGGGCCTATTCC	GATCCCAGTTACTCCTCACAGG	NO
CSSM08	11	112.4	CTTGGTGTACTAGCCCTGGG	GATATATTTGCCAGAGATTCTGCA	NO
EPCDV23	11	79.9	CATCTCCTCCCTCCTAACA	CAAGAATGGTTTTCTAGGGGC	YES
HUJ614	12	7.6	CGCCAGGCATGGTGAAGTCG	CCCAGCACAGTACAGGCTGC	YES
BM4025	12	24.0	TCGAATGAACTTTTTGGCC CAGCAGACAGCT	CACTGACTATGTGACTTTGGGC CTTTCAGAAATAGT	YES
TGLA53	12	39.3	GCAAGAGTTAGC	TTGCATTTCATGCAG	YES
MCM507	12	57.8	CTGAATGACCCATGACAGGTG	GGATAGGGGCTTGCTTGTAGAA	YES
BM719	12	86.0	TTCTGCAAATGGGCTAGAGG TGTTAAATCAAC	CACACCCTAGTTTGTAAAGCAGC	YES
BM8115A	13	31.8	ATCCAGATTTCTT	CCACCCCAAAGACCTTTCT	YES
SCYAMS	13	37.4	CTGAGCAGGAGACCATAGCACT	CAAAAGCAGTAATCCCAAAGCACA	YES
MCMA2	13	58.3	TCACCCAACAATCATGAAAC	TTAAATCGAGTGTGAATGGG	YES
HUJ616	13	65.0	TTCAAACACTACACATTGACAGGG	GGACCTTTGGCAATGGAAGG	YES
MMP9	13	115.0	CTTGCCCTTCTCATGCTGGGACT CTTTTTAATGGCTC	GTGAGGATTAGCACTTGGTCTGGCT CATCAGAGGAATG	YES
AE16	13	128.3	GGTAATATTCCTC	GGTGAAGACGTGG	YES
CSR247	14	25.5	GGACTTGCCAGACTCTGCAAT	CACTGTGGTTTGCATTAGTCAGG	YES
BMS2213	14	33.8	ATGGGCAGCTTAGGGATTG	CTTCAAGAGCCTTCAGTGGG	YES
INRA63	14	65.1	ATTTGCACAAGCTAAATCTAACC	AAACCACAGAAATGCTTGGAAAG	YES
ILSTS02	14	73.1	TCTATACACATGTGCTGTGC	CTTAGGGGTGAAGTGACACG	YES
LS30	14	94.4	GTGGATTTGTCTGTTCAAGC	CAAAGAGTTGGACACAACCTG	YES
MCMA16	15	0.0	CTGCAAGGTATTCTGTTGTG	AGGAAAATGGCCAAATAGAG	YES
IL18	15	21.3	GCCACTGGACCATCAGGGAAAT	CATGTATTCCACATCCACTTCCACCT	YES
BMS812	15	93.0	TGGACAGGACTGAGTATGCA	AGGTATCCAATAACACAGCCA	YES
BMS2076	15	105.4	AGCACCTGTACCATCTGTTC	TCCATAGGCTCACAAGAGTTG	YES
RM106	16	4.3	TGCAGTTCGACCATGGTGA GTTTACCACTAAG	CAAATGCTATCATTCTTGGACAA CCTTACACTCTGAT	YES
CP99	16	25.0	CCACCAGGGAAGC TAATGAGTCAAAC	AGAGGATATTAC AATATATAAAGAGA	YES
HH62	16	48.8	ACTACTGAGAGAC ATGGTAACCCAC	AAAGCTGGGGTGCC CAGTACTCTTGC	YES
MCM506A	16	57.5	TCCAGTATTTATC CAGACACGGCTTA	CTGGAGAATTC GTGGGGATGAAT	YES
CP49	17	28.5	GCAACTAAACGC TCATGCACTTAAG	ATTCCTTCATAAAG GATCACAAAAGTT	YES
MAF209	17	48.0	TATGTAGGATGCTG	GGATACAACCGTGG	NO
BM8125	17	86.8	CTCTATCTGTGAAAAGGTGGG	GGGGTTAGACTTCAACATACG	YES
ILSTS58	17	91.4	GCCTTACTACCATTTCCAGC	CATCCTGACTTTGGCTGTGG	NO
EPCDV009	17	106.6	GAGCTTTAAGGGTCCACGA	CTTAACCTTCTGGAGGTGAC	NO
TGLA322	17	125.0	CATGCCACCTCTTGTCTGAAA	CTTTAACATGGTTTAAATGACTATT	YES
ILSTS52	18	36.4	CTGTCCTTAAGAACAAACC	TGCAACTTAGGCTATTGACG	YES
BMS2815	18	58.0	TGATATTCAAACCTCAATGAACCC TTTATTGACAAACTC	CTTGCAATATGCTCATCATTATCA GTAGTTATTTAAAAAAA	YES
HH47	18	77.9	TCTTCCTAACTCCACC	TATCATACCTCTTAAGG	YES
TGLA122	18	88.7	CCCTCCTCCAGGTAATCAGC	AATCACATGGCAAATAAGTACATAC	YES
OB2	18	95.8	CTGCCCGATCCTTCTGCTT	AAAGGGCAGATTTCAGTATCCA	YES
OY5	18	118.0	CTACAGTCCACGGGCCAC AGGAAAGAGTAG	GGATAATGATAGCTCCCTGGC	YES
INRA26	19	0.0	GAAGAACTAGTC	AACTAAAGAGCCCTGTGCTGTAAC	YES
BM1558	19	19.0	TGAGGAAAGCCTTGGCAG	ACTGGGCCTAGCTCCTTCTC	YES
BM1303	19	25.7	CTTGGGAAAAATTGCCAGC CCCTAGGAGCTTT	CTCTGCCGCTTGCCTCTC	YES
FCB304	19	66.1	CAATAAAGAATCGG	CGCTGCTGTCAACTGGGTCAGGG	YES
HMH1R	19	73.3	GGCTTCAACTCACTGTAACACATT	TTCTTCAAGTATCACCTCTGTGGCC	YES
BM6526	26	16.4	CATGCCAAACAATATCCAGC	TGAAGGTAGAGAGCAAGCAGC	YES
CSSM43	26	40.2	AAAACCTCTGGAACTTGAAAACCTA	GTTACAAATTTAAGAGACAGAGTT	YES
JMP23	26	53.5	GTATCTGGGAG	GTCCAGATGGGAATTGTCTCCAC	YES

BM203 26 71.1 CCTGTGGTTTATC GGGTGTGACATTTTGTCCC CTGCTCGCCACTAGTCCTTC YES

Table 31: PCR results for first batch of new markers. Rams Amplified shows the number of rams in which at least one band was amplified. Markers were submitted to Invermay if a single band was amplified in at least two rams, but not if multiple bands were amplified.

Marker	Rams Amplified	Multiple bands?	Submitted?
BMS1508899	4	No	Yes
BMS1508976	0	No	No
BMS1509059	0	No	No
BMS1509106	3	No	Yes
BMS1509192	4	No	Yes
BMS1509297	0	No	No
BMS1509387	4	No	Yes
BMS1509482	0	No	No
BMS1509593	3	Yes	No
BMS1509702	0	No	No
BMS1509791	4	No	Yes
BMS1509869	3	No	Yes
BMS1509996	4	No	Yes
BMS1510131	3	No	Yes
BMS1510292	4	No	Yes
BMS1510434	4	No	Yes
BMS1510511	4	Yes	No
BMS1510639	4	No	Yes
BMS1510724	4	No	Yes
BMS1510829	4	Yes	No
BMS1510869	4	Yes	No
BMS1511009	4	No	No
BMS1511091	2	No	Yes
BMS1511155	0	No	No
BMS1511218	1	No	No
BMS1511300	2	No	No
BMS1511389	3	No	No
BMS1511509	4	No	Yes
BMS1511603	2	Yes	No
BMS1511687	3	Yes	No
BMS1511787	0	No	No
BMS1511896	4	Yes	No
BMS1511966	4	No	Yes
BMS1512041	4	No	Yes
BMS1512131	0	No	No
BMS1512218	4	Yes	No
BMS1512327	4	No	Yes
BMS1512413	4	No	Yes
BMS1512533	4	Yes	No


```

5  4  2  0 127 127 0 0  0 0 165 165 112 118 179 181 160 163 143 153 241
248 0 0 116 134 1 2
6  5  3  1 127 133 0 0  0 0 159 165 112 118 179 181 160 172 143 161 241
248 0 0 130 134 1 2
7  0  0  0  0 0 0 0  0 0  0 0 0 0  0 0  0 0 0 0  0 0  1 1
8  7  1  0 127 138 0 0  0 0  0 0 112 118 158 158 158 160 143 161 241
258 0 0 116 134 1 2
9  8  6  1 127 138 0 0  0 0  0 0 112 118 158 181 158 160 143 161 248
258 0 0 116 134 1 2
10 0 0 1 127 138 104 107 117 117 167 167 112 112 168 181 160 165 143
155 245 248 147 151 134 134 1 2
11 0 0 0  0 0 0 0  0 0  0 0 0 0  0 0  0 0 0 0  0 0  1 1
12 11 10 1 127 138 107 107 117 117 165 167 112 116 162 181 160 165
143 155 245 248 149 151 130 134 1 2
13 0 0 0  0 0 0 0  0 0  0 0 0 0  0 0  0 0 0 0  0 0  1 1
14 13 1 0  0 0 111 123 0 0 159 165 0 0 158 158 160 172 143 159 241
241 149 151 130 134 1 2
15 14 12 1  0 0 107 111 0 0 159 167 0 0 158 181 160 165 143 155 241
245 149 151 130 134 1 2
16 0 0 0  0 0 0 0  0 0  0 0 0 0  0 0  0 0 0 0  0 0  1 1
17 16 1 0 127 133 0 0 135 135 165 165 118 118 168 187 158 160 0 0
241 258 151 151 134 134 1 2
18 17 12 1 127 138 0 0 117 135 165 167 112 118 168 181 158 160 0 0
241 245 149 151 130 134 1 2
19 0 0 0  0 0 0 0  0 0  0 0 0 0  0 0  0 0 0 0  0 0  1 1
20 19 9 0 127 131 0 0  0 0  0 0 112 120 177 181 160 172 155 161 241
258 0 0 116 130 1 1
21 0 0 0  0 0 0 0  0 0  0 0 0 0  0 0  0 0 0 0  0 0  1 1
22 21 9 1 127 127 0 0  0 0  0 0 114 118 158 168 158 160 153 161 245
258 0 0 116 134 1 1
23 21 9 1 127 127 0 0  0 0  0 0 116 118 158 179 158 172 153 161 241
258 0 0 116 116 1 1
24 0 0 0  0 0 0 0  0 0  0 0 0 0  0 0  0 0 0 0  0 0  1 1
25 24 9 1 127 138 0 0  0 0  0 0 112 118 158 158  0 0 143 161 245 258
0 0 116 130 1 1
26 24 9 0 127 127 0 0  0 0  0 0 112 116 158 181 160 165 143 143 245
258 0 0 116 130 1 1
27 0 0 0  0 0 0 0  0 0  0 0 0 0  0 0  0 0 0 0  0 0  1 1
28 27 9 1 127 131 0 0  0 0  0 0 112 116 179 181 160 172 143 153 245
248 0 0 130 134 1 2
29 27 9 1 131 138 0 0  0 0  0 0 112 116 179 181  0 0 143 153 254 258
0 0 116 134 1 1
30 0 0 0  0 0 0 0  0 0  0 0 0 0  0 0  0 0 0 0  0 0  1 1
31 30 9 0 127 138 0 0  0 0  0 0 112 118 158 168 158 165 143 161 245
258 0 0 116 134 1 1
32 30 9 0 133 138 0 0  0 0  0 0 112 118 158 177 158 165 143 161 245
258 0 0 116 116 1 1
33 0 0 0  0 0 0 0  0 0  0 0 0 0  0 0  0 0 0 0  0 0  1 1
34 33 9 1 127 138 0 0  0 0  0 0 112 112 177 181 158 172 155 161 254
258 0 0 116 130 1 1
35 0 0 0  0 0 0 0  0 0  0 0 0 0  0 0  0 0 0 0  0 0  1 1
36 35 9 0 127 127 0 0  0 0  0 0 112 118 158 179 158 160 153 161 241
258 0 0 116 116 1 1
37 0 0 0  0 0 0 0  0 0  0 0 0 0  0 0  0 0 0 0  0 0  1 1
38 37 9 1 127 138 0 0  0 0  0 0 112 112 168 181  0 0 143 155 245 258
0 0 116 116 1 1
39 0 0 0  0 0 0 0  0 0  0 0 0 0  0 0  0 0 0 0  0 0  1 1
40 39 9 1 131 138 0 0  0 0  0 0 116 118 158 189 158 165 153 161 245
258 0 0 116 134 1 1
41 39 9 0 131 138 0 0  0 0  0 0 112 116 181 189 160 165 143 153 245
248 0 0 116 134 1 2

```



```

79 50 15 1 0 0 107 107 0 0 165 167 0 0 175 181 165 176 153 155 245
245 149 151 130 130 1 1
80 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
81 80 15 1 0 0 111 121 0 0 159 167 0 0 158 179 158 160 143 153 229
241 151 151 130 134 1 2
82 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
83 82 15 1 0 0 107 111 0 0 159 165 0 0 158 181 163 165 153 155 245
245 147 149 134 134 1 1
84 82 15 0 0 0 107 111 0 0 159 165 0 0 158 158 160 163 143 153 229
241 147 151 134 134 1 2
85 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
86 85 15 1 0 0 111 121 0 0 159 165 0 0 158 158 160 163 143 146 241
245 147 151 116 130 1 2
87 85 15 1 0 0 107 121 0 0 165 167 0 0 158 181 160 165 143 153 241
254 151 151 134 134 1 2
88 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
89 88 15 0 0 0 107 121 0 0 165 167 0 0 177 181 160 165 153 155 229
245 147 149 130 134 1 1
90 88 15 1 0 0 107 121 0 0 159 165 0 0 158 168 165 165 155 159 245
249 147 149 116 130 1 1
91 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 192 91
15 1
0 0 107 109 0 0 165 167 0 0 177 181 160 165 143 155 245 245 147 149 134
134 1 193 91 15 0
0 0 109 111 0 0 165 167 0 0 177 181 160 165 143 155 245 245 147 149 134
134 1 1
94 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
95 94 15 1 0 0 111 121 0 0 167 167 0 0 177 181 165 172 153 155 245
254 145 149 116 130 1 1
96 94 15 1 0 0 107 121 0 0 167 167 0 0 158 177 160 172 143 159 241
245 147 151 130 134 1 2
97 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
98 97 15 1 0 0 107 109 0 0 165 167 0 0 179 181 160 165 153 155 229
245 149 151 127 130 1 1
99 97 15 0 0 0 109 111 0 0 159 165 0 0 179 181 160 165 153 155 229
245 149 151 130 134 1 1
100 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
101 100 15 1 0 0 107 111 0 0 0 0 0 0 158 168 158 160 143 153 241
241 151 151 130 130 1 2
102 100 15 0 0 0 107 107 0 0 165 167 0 0 168 181 165 172 155 155 245
245 149 154 130 134 1 1
103 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
104 103 15 1 0 0 98 107 0 0 159 165 0 0 158 175 160 165 153 155 245
258 147 149 130 132 1 1
105 103 15 0 0 0 98 107 0 0 159 165 0 0 158 175 160 160 143 153 241
258 151 154 132 134 1 2
106 103 15 0 0 0 107 113 0 0 165 167 0 0 158 175 160 160 143 153 241
248 147 151 130 132 1 2
107 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
108 107 15 0 0 0 111 111 0 0 159 167 0 0 179 181 165 165 150 155 245
254 149 151 118 130 1 1
109 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
110 109 15 1 0 0 107 111 0 0 159 165 0 0 158 158 160 160 155 155 245
264 149 151 116 130 1 1
111 109 15 1 0 0 107 111 0 0 165 167 0 0 168 181 158 165 143 143 241
264 151 151 116 130 1 2
112 0 0 0 0 0 104 121 0 0 165 167 0 0 158 168 158 165 155 161 248
248 147 151 134 134 1 2
113 112 15 0 0 0 107 121 0 0 159 165 0 0 158 168 165 165 155 155 245
248 149 151 130 134 1 2

```

114 112 15 0 0 0 107 121 0 0 165 167 0 0 158 181 158 165 155 161 245
248 147 149 130 134 1 1
115 0 1 1
116 115 15 1 0 0 107 111 0 0 159 167 0 0 179 181 158 165 143 153 241
241 151 151 118 130 1 2
117 115 15 1 0 0 107 121 0 0 167 167 0 0 179 181 158 165 153 155 241
245 149 154 130 134 1 2
118 0 1 1
119 118 15 1 0 0 107 113 0 0 167 167 0 0 179 181 160 165 143 153 241
248 147 151 132 134 1 2
120 118 15 0 0 0 107 113 0 0 167 167 0 0 179 181 160 165 143 155 245
264 149 151 127 134 1 1
121 0 1 1
122 121 15 1 0 0 111 111 0 0 165 167 0 0 175 181 165 172 153 155 241
245 151 151 127 134 1 1
123 121 15 1 0 0 107 109 0 0 159 165 0 0 158 158 160 172 143 143 241
241 151 154 116 134 1 2
124 121 15 0 0 0 107 109 0 0 165 167 0 0 175 181 160 165 143 143 241
241 151 154 116 134 1 2
125 0 1 1
126 125 15 1 0 0 107 109 0 0 159 175 0 0 158 179 160 167 143 159 241
245 143 151 116 134 1 2
127 0 1 1
128 127 15 0 0 0 98 107 0 0 159 165 0 0 158 177 160 160 143 153 241
241 151 151 116 134 1 1
129 0 1 1
130 129 15 1 0 0 107 111 0 0 165 167 0 0 158 177 160 165 153 155 241
245 149 151 130 130 1 1
131 0 1 1
132 131 15 1 0 0 107 121 0 0 159 165 0 0 158 179 160 172 143 155 241
245 151 154 116 134 1 2
133 131 15 0 0 0 111 121 0 0 159 167 0 0 179 181 158 165 155 155 241
245 151 154 116 134 1 1
134 129 15 0 0 0 107 107 0 0 165 167 0 0 168 181 160 165 143 153 241
241 149 151 134 134 1 2
135 0 0 0 0 0 121 121 0 0 159 165 0 0 158 158 158 160 143 161 241
258 151 151 116 134 1 2
136 135 15 0 0 0 111 121 0 0 165 167 0 0 158 181 158 165 155 155 245
248 149 151 130 134 1 2
137 135 15 0 0 0 111 121 0 0 159 165 0 0 158 181 158 165 155 161 245
258 149 151 116 130 1 1
138 0 1 1
139 138 15 1 0 0 107 121 0 0 165 167 0 0 168 181 160 165 143 155 241
245 151 151 130 134 1 2
140 138 15 1 0 0 111 121 0 0 159 165 0 0 158 168 160 165 155 155 245
245 147 149 130 134 1 1
141 0 1 1
142 141 15 0 0 0 111 111 0 0 165 167 0 0 158 181 165 172 155 155 245
245 147 149 130 134 1 1
143 141 15 0 0 0 107 107 0 0 159 165 0 0 170 181 165 172 155 155 245
245 147 149 134 134 1 1
144 0 1 1
145 144 15 0 0 0 107 121 0 0 165 167 0 0 168 181 160 165 143 153 241
245 149 151 128 130 1 1
146 0 1 1
147 146 15 0 0 0 107 111 0 0 165 167 0 0 158 181 163 165 153 155 245
245 147 149 130 130 1 1
148 0 1 1
149 148 15 1 0 0 0 0 165 167 0 0 175 181 160 165 153 155 241 245
149 151 134 134 1 1
150 1 1


```

188 186 15 0 127 131 98 107 0 0 159 165 0 0 158 158 160 165 153 155
245 258 149 154 132 134 1 1
189 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
190 189 15 1 116 127 107 107 0 0 165 167 0 0 177 181 163 165 155 159
245 245 149 154 130 130 1 1
191 189 15 0 116 127 107 107 0 0 159 167 0 0 158 168 163 165 155 159
245 245 149 154 130 130 1 1
192 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
193 192 15 0 127 131 111 111 0 0 165 167 0 0 158 179 158 160 143 153
241 245 151 151 134 134 1 2
194 52 15 1 127 129 98 107 0 0 159 165 0 0 158 158 165 165 155 161
241 245 147 151 118 134 1 2
195 52 15 1 127 129 111 120 0 0 159 165 0 0 158 158 160 165 143 153
241 245 147 151 116 134 1 2
196 21 15 1 127 127 107 111 0 0 165 167 0 0 179 181 165 172 153 155
241 245 149 151 130 130 1 1
197 21 15 0 127 127 107 107 0 0 165 167 0 0 158 179 160 172 143 153
241 245 147 151 116 130 1 2
198 109 15 0 127 127 111 120 0 0 159 165 0 0 158 158 160 160 143 155
245 245 147 149 116 130 1 1
199 109 15 1 127 127 111 120 0 0 159 165 0 0 158 158 160 160 143 155
241 245 147 151 134 134 1 2
200 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
201 200 15 0 127 131 107 111 0 0 159 165 0 0 158 168 163 165 155 155
241 241 151 151 134 134 1 2
202 200 15 1 127 127 107 109 0 0 159 165 0 0 158 168 160 163 143 155
241 245 147 151 116 134 1 2
203 80 15 1 127 133 109 111 0 0 165 167 0 0 179 181 158 165 153 155
229 245 149 151 130 134 1 1
204 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
205 204 15 1 127 133 98 107 0 0 165 167 0 0 168 181 165 165 155 159
241 245 149 151 118 130 1 2
206 204 15 0 127 133 111 111 0 0 159 165 0 0 158 168 160 165 143 159
241 245 151 154 134 136 1 2
207 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
208 207 15 0 127 127 107 111 0 0 165 167 0 0 181 181 0 0 155 161 245
245 149 151 130 134 1 1
209 207 15 0 127 135 107 111 0 0 167 167 0 0 181 181 165 165 155 161
245 245 149 151 130 134 1 1
210 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
211 210 15 1 127 133 111 120 0 0 159 165 0 0 177 181 160 165 153 155
245 248 149 149 116 130 1 1
212 210 15 0 127 135 107 111 0 0 165 167 0 0 177 181 160 165 155 161
235 241 151 151 116 134 1 2
213 156 15 1 127 140 107 120 0 0 159 165 116 116 158 168 158 160 143
153 245 248 149 151 116 130 1 1
214 156 15 1 127 129 107 120 0 0 165 167 0 0 179 181 165 172 153 155
245 248 149 151 130 134 1 1
215 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
216 215 15 1 127 131 107 111 0 0 165 167 0 0 168 181 165 176 155 155
241 245 149 154 130 134 1 1
217 215 15 1 127 140 107 111 0 0 159 165 0 0 168 181 0 0 155 155 241
245 147 149 130 130 1 1
218 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
219 218 15 0 127 131 98 107 0 0 165 167 0 0 177 181 160 165 153 155
245 245 147 149 130 130 1 1
220 218 15 0 127 131 98 111 0 0 159 165 0 0 158 177 160 160 143 155
241 258 151 154 127 134 1 1
221 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
222 221 15 1 127 133 111 120 0 0 159 165 0 0 158 158 160 176 143 153
241 254 149 151 134 134 1 2

```



```

335 334 18 1 127 133 0 0 135 135 165 165 112 118 168 175 158 165 0 0
241 254 151 151 134 134 1 2
336 334 18 1 133 138 0 0 135 135 165 167 112 118 168 181 160 163 0 0
245 245 149 154 130 134 1 1
337 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
338 337 18 1 131 138 0 0 135 135 165 165 118 118 168 168 158 165 0 0
241 245 147 151 116 134 1 1
339 337 18 1 127 138 0 0 135 135 165 165 116 118 158 168 158 172 0 0
241 254 151 151 116 130 1 2
340 97 18 1 133 138 0 0 117 135 167 167 112 118 179 181 160 172 0 0
245 245 149 154 130 134 1 1
341 97 18 0 133 138 0 0 135 135 165 165 118 118 168 179 158 172 0 0
245 245 149 151 127 130 1 1
342 63 18 1 0 0 0 0 135 135 165 165 118 118 168 177 158 160 0 0
241 245 147 151 134 134 1 2
343 63 18 0 127 133 0 0 135 135 165 167 112 118 168 181 160 160 0 0
241 245 147 151 127 134 1 1
344 244 18 1 120 127 0 0 117 135 167 167 112 116 162 181 160 160 0 0
245 245 143 149 116 130 1 1
345 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
346 345 18 0 127 133 0 0 117 135 165 167 118 118 158 168 158 165 0 0
241 245 151 151 116 134 1 2
347 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
348 347 18 0 131 138 0 0 135 135 165 165 112 118 168 179 158 165 0 0
245 245 147 149 130 134 1 1
349 180 18 1 127 140 0 0 117 129 167 167 112 118 168 168 158 158 0 0
241 254 151 151 134 134 1 2
350 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
351 350 18 1 127 133 0 0 117 135 167 167 112 118 158 181 158 160 0 0
245 245 143 149 130 134 1 1
352 350 18 0 133 138 0 0 117 129 165 167 112 112 158 181 0 0 0 0
241 245 147 151 130 134 1 1
353 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
354 353 18 0 127 138 0 0 129 135 165 165 114 118 168 189 158 172 0 0
241 245 151 151 130 134 1 2
355 353 18 0 127 138 0 0 133 135 155 165 114 118 168 181 158 158 0 0
241 245 151 151 130 134 1 1
356 334 18 1 133 138 0 0 117 135 167 167 112 112 168 181 160 163 0 0
241 245 151 154 134 134 1 2
357 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
358 357 18 0 127 133 0 0 129 135 165 167 116 118 168 170 158 165 0 0
241 245 147 151 116 130 1 1
359 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
360 359 18 0 127 138 0 0 135 139 165 165 118 118 168 168 158 165 0 0
241 245 149 151 130 134 1 1
361 359 18 0 133 138 0 0 117 129 165 165 118 118 168 168 158 165 0 0
241 241 151 151 134 134 1 2
362 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
363 362 18 0 131 138 0 0 117 135 165 167 112 116 175 181 160 172 0 0
241 245 147 149 130 134 1 1
364 362 18 0 127 138 0 0 117 139 165 167 116 118 168 175 158 172 0 0
241 241 151 151 130 134 1 2
365 207 18 0 127 138 0 0 117 135 165 167 112 116 168 179 158 165 0 0
241 245 151 151 134 134 1 2
366 207 18 0 135 138 0 0 117 129 167 167 112 116 168 181 158 158 0 0
241 245 151 154 130 134 1 2
367 204 18 1 127 133 0 0 117 129 165 167 112 112 168 181 160 165 0 0
245 245 149 154 130 136 1 1
368 204 18 0 133 138 0 0 117 135 165 165 112 118 168 168 158 165 0 0
241 245 151 154 134 136 1 2
369 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1

```


405 404 18 1 129 138 0 0 133 135 165 167 116 118 168 175 158 158 0 0
 235 241 145 151 116 134 1 2
 406 404 18 1 138 140 0 0 129 135 165 165 116 118 168 179 158 165 0 0
 235 241 145 151 116 134 1 2
 407 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
 408 407 18 1 127 131 0 0 117 135 165 167 112 112 158 177 160 160 0 0
 241 245 151 151 134 134 1 2
 409 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
 410 409 18 1 120 138 0 0 117 129 165 165 118 118 168 177 158 158 0 0
 241 245 0 0 134 134 1 2
 411 234 18 0 127 133 0 0 117 129 165 167 112 112 0 0 160 165 0 0
 229 245 147 149 116 130 1 1
 412 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
 413 412 18 1 127 135 0 0 117 135 165 167 112 118 158 181 160 176 0 0
 241 241 147 151 127 134 1 2
 414 412 18 1 120 127 0 0 117 135 159 165 112 112 179 181 165 165 0 0
 0 0 147 151 130 134 1 2
 415 0 0 0 0 0 0 135 135 165 167 118 118 158 158 158 160 0 0 241
 248 147 151 116 134 1 2
 416 415 18 0 138 138 0 0 135 135 167 167 112 118 158 181 160 160 0 0
 245 248 147 149 116 130 1 1
 417 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
 418 417 18 0 123 127 0 0 117 133 165 167 112 116 177 181 160 163 0 0
 241 241 147 151 130 134 1 2
 419 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
 420 419 18 0 0 0 0 0 117 135 165 165 112 118 168 181 158 176 0 0
 245 254 149 151 130 130 1 1
 421 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
 422 421 18 0 127 138 0 0 117 129 165 167 112 116 158 181 160 160 0 0
 245 254 149 154 130 134 1 1
 423 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
 424 423 18 1 127 138 0 0 117 129 165 167 112 116 179 181 160 160 0 0
 241 245 149 151 130 130 1 1
 425 184 18 0 127 127 0 0 129 135 165 167 112 112 162 181 160 165 0 0
 229 245 147 149 130 130 1 1
 426 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
 427 426 18 0 127 135 0 0 0 0 0 0 112 112 177 181 158 160 143 153
 245 254 0 0 132 134 1 1
 428 159 18 0 0 0 107 107 0 0 165 165 0 0 175 181 160 165 143 155 241
 245 149 154 130 134 1 1
 429 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
 430 429 18 1 0 0 107 107 0 0 159 165 0 0 168 179 158 172 143 153 241
 245 147 151 116 134 1 2
 431 429 18 1 0 0 107 111 0 0 165 165 0 0 158 168 158 172 143 153 241
 245 147 151 116 130 1 2
 432 161 18 1 0 0 107 107 0 0 165 165 0 0 168 168 158 165 0 0 241
 241 151 151 130 130 1 2
 433 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
 434 433 18 1 0 0 107 107 0 0 165 167 0 0 168 170 158 165 143 153 241
 245 151 151 134 134 1 2
 435 292 18 1 0 0 107 121 0 0 165 165 0 0 168 168 158 165 143 155 241
 245 151 151 134 134 1 2
 436 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
 437 436 18 1 127 138 0 0 0 0 165 167 112 112 177 181 160 160 143
 143 245 245 0 0 130 130 1 1
 438 436 18 0 127 138 0 0 0 0 165 165 112 118 168 168 158 165 143
 153 241 245 0 0 134 134 1 1

References

- Abajian, C. (1994). Sputnik DNA microsatellite repeat search utility. Available at <http://espressoftware.com/sputnik/index.html>
- Addison, P.K.F., Berry, V., Ionides, A.C.W., Francis, P.J., Bhattacharya, S.S., and Moore, A.T. (2005). Posterior polar cataract is the predominant consequence of a recurrent mutation in the PITX3 gene. *British Journal of Ophthalmology* **89**: 138-141.
- Al-Till, M., Jarrah, N.S., Ajlouni, K.M. (2002). Ophthalmologic findings in fifteen patients with Wolfram syndrome. *European Journal of Ophthalmology* **12**: 84-8.
- Andley, U. P. (2006). Crystallins and hereditary cataracts: molecular mechanisms and potential for therapy. *Expert Reviews in Molecular Medicine* **8** Digital Object Identifier: 10.1017/S1462399406000111 at <http://www-ermm.cbcu.cam.ac.uk/>
- Arora, A., Minogue, P.J., Liu, X., Addison, P.K., Russel-Eggitt, I., Webster, A.R., Hunt, D.M., Ebihara, L., Beyer, E.C., Berthoud, V.M., and Moore, A.T. (2008). A novel connexin50 mutation associated with congenital nuclear pulverulent cataracts. *Journal of Medical Genetics* **45**: 155-160.
- Azuma, N., Hirakiyama, A., Inoue, T., Asaka, A., and Yamada, M. (2000). Mutations of a human homologue of the *Drosophila eyes absent* gene (*EYAI*) detected in patients with congenital cataracts and ocular anterior segment anomalies. *Human Molecular Genetics* **9**: 363-366.
- Bahlo, M., Xing, L., and Wilkinson, C. R. (2004). HumanMSD and MouseMSD: generating genetic maps for human and murine microsatellite markers. *Bioinformatics* **20**: 3280-3282.
- Baruch, A., Greenbaum, D., Levy, E. T., Nielsen, P.A., Gilula, N.B., Kumar, N.M., and Bogyo, M. (2001). Defining a link between gap junction communication, proteolysis, and cataract formation. *Journal of Biological Chemistry* **276**: 28999-29006.

- Bartoloni, L., Horrigan, S.K., Viles, K.D., Gilchrist, J.M., Stajich, J.M., Vance, J.M., Yamaoka, L.H., Pericak-Vance, M.A., Westbrook, C.A., and Speer, M.C. (1998). Use of a CEPH meiotic breakpoint panel to refine the locus of limb-girdle muscular dystrophy type 1A (LGMD1A) to a 2-Mb interval on 5q31. *Genomics* **54**: 250-255.
- Bassnett, S. (1995). The fate of the Golgi apparatus and the endoplasmic reticulum during lens fiber cell differentiation. *Investigative Ophthalmology and Visual Science* **36**: 1793-1803.
- Bassnett, S. (1997). Fiber cell denucleation in the primate lens. *Investigative Ophthalmology and Visual Science* **38**: 1678-87.
- Beby, F., Morle, L., Michon, L., Ederly, P., Burillon, C., Denise, P., (2003). The genetics of hereditary cataract. (2003). *Journal Français d'Ophthalmologie* **26**: 400-408.
- Berry, V., Francis, P., Kaushal, S., Moore, A., and Bhattacharya, S. (2000). Missense mutations in *MIP* underlie autosomal dominant "polymorphic" and lamellar cataracts linked to 12q. *Nature Genetics* **25**: 15-17.
- Bessman, M.J., Frick, D.N., and O'Handley, S.F. (1996). The MutT proteins or "Nudix" hydrolases, a family of versatile, widely distributed, "housecleaning" enzymes. *The Journal of Biological Chemistry* **271**: 25059-25062.
- Bitoun, E., Oliver, P.L., and Davies, K.E. (2007). The mixed-lineage leukemia fusion partner AF4 stimulates RNA polymerase II transcriptional elongation and mediates coordinated chromatin remodeling. *Human Molecular Genetics* **16**: 92-106.
- Bhagyalaxmi, S.G., Srinivas, P.N.B.S., Barton, K.A., Kumar, K.R., Vidyavathi, M., Petrash, J.M., Reddy, G.B., and Padma, T. (2009). A novel mutation (F71L) in α A-crystallin with defective chaperone-like function associated with age-related cataract. *Biochimica et Biophysica Acta* (Article in press, doi:10.1016/j.bbabdis.2009.06.011)
- Bizarro, S., Cerveira, N., Correia, C., Lisboa, S., Peixoto, A., Norton, L., and Teixeira, M.R. (2007). Molecular characterisation of a rare *MLL-AF4* (*MLL-AFF1*) fusion rearrangement in infant leukemia. *Cancer Genetics and Cytogenetics* **178**: 61-64.

Blanc, I., Bach, A., Lallemand, Y., Perrin-Schmitt, F., Guénet, J.-L., and Robert, B. (2003). A new mouse limb mutation identifies a *Twist* allele that requires interacting loci on Chromosome 4 for its phenotypic expression.

Bloemendal, H. (1977). The vertebrate eye lens. *Science* **8**: 127-38.

Bovine Genome Sequencing and Analysis Consortium (2009). The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science* **324**: 522-8.

Brooks, H.V., Jolly, R.D., West, D.M., and Bruere, A.N. (1982). An inherited cataract in New Zealand Romney sheep. *New Zealand Veterinary Journal* **30**: 113-114.

Brooks, H.V., Jolly, R.D., and Paterson, C.A. (1983). The pathology of an inherited cataract of sheep. *Current Eye Research* **2**: 625-632.

Broom, M. F., Zhou, C., Broom, J. E., Barwell, K. J., Jolly, R. D., and Hill, D. F. (1998). Ovine neuronal ceroid lipofuscinosis: a large animal model syntenic with the human neuronal ceroid lipofuscinosis variant CLN6. *Journal of Medical Genetics* **35**: 717-721.

Brownstein, M. J., Carpten, J. D., and Smith, J. R. (1996). Modulation of non-templated nucleotide addition by Taq DNA polymerase: Primer modifications that facilitate genotyping. *BioTechniques* **20**: 1004-1010.

Bron, A.J., Sparrow, J., Brown, N.A.P., Harding, J.J., and Blakytyn, R. (1993). The lens in diabetes. *Eye* **7**: 260-275.

Bu, L., Jin, Y., Shi, Y., Chu, R., Ban, A., Eiberg, H., Andres, L., Jiang, H., Zheng, G., Qian, M., Cui, B., Xia, Y., Liu, J., Hu, L., Zhao, G., Hayden, M.R., and Kong, X. (2002). Mutant DNA-binding domain of HSF4 is associated with autosomal dominant lamellar and Marner cataract. *Nature Genetics* **31**: 276-278.

Burdon, K.P., McKay, J.D., Sale, M.M., Russell-Eggitt, I.M., Mackey, D.A., Wirth, M.G., Elder, J.E., Nicoll, A., Clarke, M.P., FitzGerald, L.M., Stankovich, J.M., Shaw, M.A., Sharma, S., Gajovic, S., Gruss, P., Ross, S., Thomas, P., Voss, A.K., Thomas, T., Géczy, J., and Craig, J.E. (2003). Mutations in a novel gene, *NHS*, cause the pleiotropic effects of

Nance-Horan Syndrome, including severe congenital cataract, dental anomalies, and mental retardation. *American Journal of Human Genetics* **73**: 1120-1130.

Burdon, K.P., Wirth, M.G., Mackey, D.A., Russell-Eggit, I.M., Craig, J.E., Elder, J.E., Dickinson, J.L., and Sale, M.M. (2004). A novel mutation in the Connexin 46 gene causes autosomal dominant congenital cataract with incomplete penetrance. *Journal of Medical Genetics* **41**: e106. Erratum in **42**: 288 and **45**: 256.

Cai, R.-L., Meng, W., Lu, H.-Y., Lin, W.-Y., Jiang, F., and Shen, F.-M. (2003). Segregation analysis of hepatocellular carcinoma in a moderately high-incidence area of East China. *World Journal of Gastroenterology* **9**: 2428-2432.

Calle, Y., Carragher, N. O., Thrasher, A. J., and Jones, G. E. (2006). Inhibition of calpain stabilises podosomes and impairs dendritic cell motility. *Journal of Cell Science* **119**: 2375-85.

Castle, W. E., and Little, C. C. (1910). On a modified Mendelian ratio among yellow mice. *Science* **32**: 868-870.

Cervantes-Laurean, D., Jacobson, E. L., and Jacobson, M.K. (1996). Glycation and glycooxidation of histones by ADP-ribose. *The Journal of Biological Chemistry* **271**: 10461-10469.

Chai, Z., Brereton, P., Suzuki, T., Sasano, H., Obeyesekere, V., Escher, G., Saffery, R., Fuller, P., Enriquez, C., and Krozowski, K. (2003). 17 β -hydroxysteroid dehydrogenase type XI localises to human steroidogenic cells. *Endocrinology* **144**: 2084-2091.

Chakraborty, R., Kimmel, M., Stivers, D. N., Davison, L. J., and Deka, R. (1997). Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. *Proceeding of the National Academy of Sciences of the United States of America* **94**: 1041-6.

Chambers, C., and Russell, P. (1991). Deletion mutation in an eye lens β -crystallin. *Journal of Biological Chemistry* **266**: 6742-6746.

Chen, Q., Ma, J., Yan, M., Mothobi, M.E., Liu, Y., and Zheng, F. (2009). A novel mutation in *CRYAB* associated with autosomal dominant congenital nuclear cataract in a Chinese family. *Molecular Vision* **15**: 1359-1365.

Ciulla T.A., Sklar R.M., and Hauser S.L. (1988). A simple method for DNA purification from peripheral blood. *Analytical Biochemistry* **174**: 485-488.

Clerget-Darpoux, F., Bonaïti-Pellié, and Hochez, J. (1986). Effects of misspecifying genetic parameters in lod score analysis. *Biometrics* **42**: 393-399.

Cockett, N.E., Shay, T.L., Beever, J.E., Nielsen, D., Albretsen, J., Georges, M., Peterson, K., Stephens, A., Veronon, W., Timofeevskaia, O., South, S., Mork, J., Maciulis, A., and Bunch, T.D. (1999). Localization of the locus causing Spider Lamb Syndrome to the distal end of ovine Chromosome 6. *Mammalian Genome* **10**: 35-38.

Cockett, N.E. (2006). The sheep genome. In *Vertebrate Genomes Chapter 79*. Volff, J.-N. (ed) Genome Dynamics, Basel, Switzerland 2: 79-85.

Conley, Y.P., Ertuk, D., Keverline, A., Mah, T.S., Keravala, A., Barnes, L.R., Bruchis, A., Hess, J.F., FitzGerald, P.G., Weeks, D.E., Ferrell, R.E., and Gorin, M.B. (2000). A juvenile-onset, progressive cataract locus on chromosome 3q21-q22 is associated with a missense mutation in the beaded filament structural protein-2. *American Journal of Human Genetics* **66**: 1426-1431.

Crawford, A.M., Dodds, K.G., Ede, A.J., Pierson, C.A., Montgomery, G.W., Garmonsway, H.G., Beattie, A.E., Davies, K., Maddox, J.F., Kappes, S.W., Stone, R.T., Nguyen, T.C., Penty, J.M., Lord, E.A., Broom, J.E., Buitkamp, J., Schwaiger, W., Epplen, J.T., Matthew, P., Matthews, M.E., Hulme, D.J., Beh, K.J., McGraw, R.A., and Beattie, C.W. (1995). An autosomal genetic linkage map of the sheep genome. *Genetics* **140**: 703-724.

Dalrymple, B. P., Kirkness, E.F., Nefedov, M., McWilliam, S., Ratnakumar, A., Barris, W., Zhao, S., Shetty, J., Maddox, J.F., O'Grady, M., Nicholas, F., Crawford, A., Smith, T., de Jong, P.J., McEwan, J., Oddy, V.H., and Cockett, N. (2007). Using comparative genomics to reorder the human genome sequence into a virtual sheep genome. *Genome Biology* **8**: R152 (doi: 10.1186/gb-2007-8-7-r152).

de Jong, W. W., Leunissen, J. A. M., and Voorter, C. E. M. (1993). Evolution of the α -crystallin/small heat-shock protein family. *Molecular Biology and Evolution* **10**: 103-126.

Di Fonzo, A., Ronchi, D., Lodi, T., Fassone, E., Tigano, M., Lamperti, C., Corti, S., Bordoni, A., Fortunato, F., Nizzardo, M., Napoli, L., Donadoni, C., Salani, S., Saladino, F., Moggio, M., Bresolin, N., Ferrero, I., and Comi, G.P. (2009). The mitochondrial disulfide relay system protein GFER is mutated in autosomal-recessive myopathy with cataract and combined respiratory-chain deficiency. *American Journal of Human Genetics* **84**: 594-604.

Dickerson, J. E., and Lou, M.F. (1997). Free cysteine levels in normal human lenses. *Experimental Eye Research* **65**: 451-4.

Diehn, J.J., Diehn, M., Marmor, M.F., and Brown, P.O. (2005). Differential gene expression in anatomical compartments of the human eye. *Genome Biology* **6**: R74, supplementary information at <http://microarray-pubs.stanford.edu/eyecompartments/figures.shtml>.

Duncan, G., and Bushell, A.R. (1975). Ion analyses of human cataractous lenses. *Experimental Eye Research* **20**: 223-230.

Edwards, J.H. (1991). The Oxford Grid. *Annals of Human Genetics* **55**: 17-31.

Eichhorn, P.J.A., Creighton, M.P., Wilhelmsen, K., van Dam, H., and Bernards, R. (2007). A RNA interference screen identifies the protein phosphatase 2A subunit PR55 γ as a stress-sensitive inhibitor of c-SRC. *PLoS Genetics* **3** e218.

Faulkner-Jones, B., Zandy, A. J., and Bassnett, S. (2003). RNA stability in terminally differentiating fibre cells of the ocular lens. *Experimental Eye Research* **77**: 463-76.

Felsenstein, J. (1978). A mathematically tractable family of genetic mapping functions with different amounts of interference. *Genetics* **91**: 769-75.

Francis, P. J., Berry, V., Moore, A. T., and Bhattacharya, S. (1999). Lens biology: development and cataractogenesis. *Trends in Genetics* **15**: 191-196.

Francis, P., Chung, J.-J., Yasui, M., Berry, V., Moore, A., Wyatt, M. K., Wistow, G., Bhattacharya, S. S., and Agre, P. (2000). Functional impairment of lens aquaporin in two families with dominantly inherited cataracts. *Human Molecular Genetics* **9**: 2329-34.

Gelatt, K.N., and MacKay, E.O. (2005). Prevalence of primary breed-related cataracts in the dog in North America. *Veterinary Ophthalmology* **8**: 101-111.

Gelatt, K.N., Whitley, R. D., Lavach, J.D., Barrie, K.P., and Williams, L.W. (1979). Cataracts in Chesapeake Bay Retrievers. *Journal of the American Veterinary Medicine Association* **175**: 1176-1178.

Gerhard, D.S., Wagner, L., Feingold, E.A., Shenmen, C.M., Grouse, L.H., Schuler, G., Klein, S.L., Old, S., Rasooly, R., Good, P., Guyer, M., Peck, A.M., Derge, J.G., Lipman, D., and Collins, F.S. (2004). The status, quality and expansion of the NIH full-length cDNA project: The mammalian gene collection (MGC). *Genome Research* **14**: 2121-2127.

Geyer, J., Godoy, J.R., Petzinger, E. (2004). Identification of a sodium-dependent organic anion transporter from rat adrenal gland. *Biochemical and Biophysical Research Communications* **316**: 300-306.

Girelli, D., Corrocher, R., Bisceglia, L., Olivieri, O., Zelante, L., Panozzo, G., and Gasparini, P. (1997). Hereditary hyperferritinemia-cataract syndrome caused by a 29-base pair deletion in the iron responsive element of ferritin L-subunit gene. *Blood* **90**: 2084-8.

Graham, C., Hodin, J., and Wistow, G. (1996). A retinaldehyde dehydrogenase as a structural protein in a mammalian eye lens. *Journal of Biological Chemistry* **271**: 15623-15628.

Graw, J., Löster, J., Soewarto, D., Fuchs, H., Reis, A., Wolf, E., Balling, R., Hrabé de Angelis, M. (2002). V76D mutation in a conserved γ D-crystallin region leads to dominant cataracts in mice. *Mammalian Genome* **13**: 452-455.

Graw, J. (2009). Genetics of crystallins: Cataract and beyond. *Experimental Eye Research* **88**: 173-189.

Green, P., Falls, K., and Crooks, S. (1990). Documentation for CRIMAP, version 2.4. Washington Univ., School Med. St. Louis, MO. Available online at <http://linkage.rockefeller.edu/soft/crimap/>.

Grisart, B., Coppieters, W., Farnir, F., Karim, L., Ford, C., Berzi, P., Cambisano, N., Mni, M., Reid, S., Simon, P., Spelman, R., Georges, M., and Snell, R. (2002) Positional candidate cloning of a QTL in dairy cattle: Identification of a missense mutation in the bovine *DGATI* gene with major effect on milk yield and composition. *Genome Research* **12**: 222-231.

Grobet, L., Martin, L.J.R., Poncelet, D., Pirottin, D., Brouwers, B., Riquet, J., Schoeberlein, A., Dunner, S., Ménéssier, F., Massabanda, J., Fries, R., Hanset, R., and Georges, M. (1997). A deletion in the bovine myostatin gene causes the double-muscling phenotype in cattle. *Nature Genetics* **17**: 71-74.

Gu, K., Chang, S., Ritchie, H. H., Clarkson, B. H., and Rutherford, R. B. (2000). Molecular cloning of a human dentin sialophosphoprotein gene. *European Journal of Oral Science* **108**: 35-42.

Haldane, J. B. S. (1919). The combination of linkage values, and the calculation of distances between the loci of linked factors. *Journal of Genetics* **8**: 299-309.

Hanson, I., Churchill, A., Love, J., Axton, R., Moore, T., Clarke, M., Meire, F., and van Heyningen, V. (1999) Missense mutations in the most ancient residues of the PAX6 paired domain underlie a spectrum of human congenital eye malformations. *Human Molecular Genetics* **8**: 165-172.

Hedera, P., Ma, S., Blair, M. A., Taylor, K. A., Hamati, A., Bradford, Y., Abou-Khalil, B., and Haines, J. L. (2006). Identification of a novel locus for febrile seizures and epilepsy on chromosome 21q22. *Epilepsia* **47**: 1622-1628.

Hejtmancik, J.F., (2008). Congenital cataracts and their molecular genetics. *Seminars in Cell & Developmental Biology* **19**: 134-149.

Hejtmancik, J.F., Kaiser, M.I., and Piatigorsky, J. (2001). Molecular biology and inherited disorders of the eye lens. In *The metabolic and molecular bases of inherited disease* **Chapter 241**. Scriver, C., Childs, B., Kinzler, K. W., Vogelstein, B. (eds) McGraw-Hill, New York, N.Y. 6033-6061.

Hernandez-Valladares, M., Rihet, P., ole-MoiYoi, O. K., and Iraqi, F. A. (2004). Mapping of a new quantitative trait locus for resistance to malaria in mice by a comparative mapping approach with human chromosome 5q31-q33. *Immunogenetics* **56**: 115-117.

Hey, J. (2004). What's so hot about recombination hotspots? *Public Library of Science Biology* **2**: 0730-3.

Hightower, K. R., and Farnum, R. (1985). Calcium induced opacities in cultured human lenses. *Experimental Eye Research* **41**: 565-8.

Horwitz, J. (1992). α -Crystallin can function as a molecular chaperone. *Proceedings of the National Academy of Sciences of the United States of America* **89**: 10449-53.

Hu, P., Yu, L., Zhang, M., Zheng, L., Zhao, Y., Fu, Q., and Zhao, S. (2000). Molecular cloning and mapping of the brain-abundant B1 γ subunit of protein phosphatase 2A, *PPP2R2C* to human chromosome 4p16. *Genomics* **67**: 83-86.

Huang, Q.-L., Du, X.-Y., Stone, S. H., Amsbaugh, D. F., Datiles, M., Hu, T.-S., and Zigler, J. S. Jr (1990). Association of hereditary cataracts in strain 13/N guinea-pigs with mutation of the gene for ζ -crystallin. *Experimental Eye Research* **50**: 317-325.

Ionides, A.C.W., Berry, V., Mackay, D.S., Moore, A.T., Bhattacharya, S.S., and Shiels, A. (1997). A locus for autosomal dominant posterior polar cataract on chromosome 1p. *Human Molecular Genetics* **6**: 47-51.

International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. *Nature* **409**: 860-921.

International Human Genome Sequencing Consortium (2004). Finishing the euchromatic sequence of the human genome. *Nature* **431**: 931-45.

Isaacs, A.M., Oliver, P.L., Jones, E.L., Jeans, A., Potter, A., Hovik, B.H., Nolan, P.M., Vizor, L., Glenister, P., Simon, A. K., Gray, I.C., Spurr, N.K., Brown, S.D.M., Hunter, A.J., and Davies, K.E. (2003). A mutation in *Af4* is predicted to cause cerebellar ataxia and cataracts in the robotic mouse. *The Journal of Neuroscience* **23**: 1631-1637.

Isnard, P., Coré, N., Naquet, P., and Djabali, M. (2000). Altered lymphoid development in mice deficient for the *MAF4* proto-oncogene. *Blood* **96**: 705-710.

Iyengar, S.K., Klein, B.E.K., Klein, R., Jun, G., Schick, J.H., Millard, C., Liptak, R., Russo, K., Lee, K.E., and Elston, R.C. (2004). Identification of a major locus for age-related cortical cataract on chromosome 6p12-q12 in the Beaver Dam Eye Study. *Proceedings of the National Academy of Sciences of the USA* **101**: 14485-14490.

Jamieson, R.V., Perveen, R., Kerr, B., Carette, M., Yardley, J., Heon, E., Wirth, M.G., van Heyningen, V., Donnai, D., Munier, F., and Black, G.C.M. (2002). Domain disruption and mutation of the bZIP transcription factor, *MAF*, associated with cataract, ocular anterior segment dysgenesis and coloboma. *Human Molecular Genetics* **11**: 33-42.

Jánossy, J., Ubezio, P., Apáti, A., Magócsi, M., Tompa, P., and Friedrich, P. (2004). Calpain as a multi-site regulator of cell cycle. *Biochemical Pharmacology* **67**: 1513-21.

Jun, G., Guo, H., Klein, B.E.K., Klein, R., Wang, J.J., Mitchell, P., Miao, H., Lee, K.E., Joshi, T., Buck, M., Chugha, P., Bardenstein, D., Klein, A.P., Bailey-Wilson, J.E., Gong, X., Spector, T.D., Andrew, T., Hammond, C.J., Elston, R.C., Iyengar, S.K., and Wang, B. (2009). *EPHA2* is associated with age-related cortical cataract in mice and humans. *Public Library of Science Genetics* **5**: e1000584. doi:10.1371/journal.pgen.1000584

Kaul, H., Riazuddin, S. A., Yasmeen, A., Mohsin, S., Khan, M., Nasir, I. A., Khan, S. N., Husnain, T., Akram, J., Hejtmancik, J. F., and Riazuddin, S. (2010). A new locus for autosomal recessive congenital cataract identified in a Pakistani family. *Molecular Vision* **16**: 240-245.

Khanim, F., Kirk, J., Latif, F., and Barrett, T.G. (2001). WFS1/Wolframin mutations, Wolfram syndrome, and associated diseases. *Human Mutation* **17**: 357-367.

Kilpinen, H., Ylisaukko-oja, T., Rehnström, K., Gaál, E., Turunen, J. A., Kempas, E., von Wendt, L., Varilo, T., and Peltonen, L. (2009). Linkage and linkage disequilibrium scan for autism loci in an extended pedigree from Finland. *Human Molecular Genetics* **18**: 2912-2921.

Kitamura, E., Su, Guanfang, Sossey-Alaoui, K., Malaj, E., Lewis, J., Pan, H., Hawthorn, L., Roe, B., and Cowell, J. K. (2000). A transcription map of the minimally deleted region from 13q14 in B-cell chronic lymphocytic leukemia as defined by large scale sequencing of the 650 kb critical region. *Oncogene* **19**: 5772-80.

Klug W.S., and Cummings, M.R. (1994). *Concepts of Genetics* Prentice-Hall, Inc.

Koenigsberg, R. (managing editor) (1989). *Churchill's Medical Dictionary* Churchill Livingstone Inc. Page 308.

Köhn, L., Bowne, S. J., Sullivan, L. S., Daiger, S. P., Burstedt, M. S. I., Kadzhaev, K., Sandgren, O., and Golovleva, I. (2009). Breakpoint characterisation of a novel ~59 kb genomic deletion on 19q13.42 in autosomal-dominant retinitis pigmentosa with incomplete penetrance. *European Journal of Human Genetics* **17**: 652-655.

Kühn, F. J. P., and Lückhoff, A. (2004). Sites of the NUDT9-H domain critical for ADP-ribose activation of the cation channel TRPM2.

Kuck, J.F.R. (1990). Late onset hereditary cataract of the Emory mouse. A model for human senile cataract. *Experimental Eye Research* **50**: 659-664.

Krafka, J. (1920). The effect of temperature upon facet number in the bar-eyed mutant of *Drosophila*: Part I. *Journal of General Physiology* **2**: 409-32.

Kratochvilova, J. (1981). Dominant cataract mutations detected in offspring of γ -irradiated male mice. *Journal of Heredity* **72**: 302-7.

- Lander, E. S., and Green, P. (1987). Construction of multilocus genetic linkage maps in humans. *Proceedings of the National Academy of Sciences of the United States of America*.
- Latter, M.R., Latter, B.D.H., Wilkins, J.F., and Windsor, P.A. (2006). Inheritance of proportionate dwarfism in Angus cattle. *Australian Veterinary Journal* **84**: 122-128.
- Lawson, E.Q., Schubert, C.F., Lewis, R.V., and Middaugh, C.R. (1981). The solubility of bovine lens crystallins. *The Journal of Biological Chemistry* **256**: 6523-6525.
- Lee, M., Kwon, Y. T., Mingwei, L., Peng, J., Friedlander, R. M., and Tsai, L.-H. (2000). Neurotoxicity induces cleavage of p35 to p25 by calpain. *Nature* **405**: 360-4.
- Leroy, J. G., Seppala, R., Huizing, M., Dacremont, G., De Simpel, H., Van Coster, R. N., Orvisky, E., Krasnewich, D. M., and Gahl, W. A. (2001). Dominant inheritance of Sialuria, an inborn error of feedback inhibition. *American Journal of Human Genetics* **68**: 1419-1427.
- Lidsky, A. S., Robson, K. J. H., Thirumalachary, C., Barker, P. E., Ruddle, F. H., and Woo, S. L. C. (1984). The PKU locus in man is on chromosome 12. *American Journal of Human Genetics* **36**: 527-33.
- Liška, F., Chylíková, B., Martínek, J., and Křen, V. (2008). Microphthalmia and cataract in rats with a novel point mutation. *Molecular Vision* **14**: 823-828.
- Lin, Y., Liu, N.N., Lei, C.T., Fan, Y.C., Liu, X.Q., Yang, Y., Wang, J.F., Liu, B., and Yang, Z.L. (2008). A novel GJA8 mutation in a Chinese family with autosomal dominant cataract. *Chinese Journal of Medical Genetics* **25**: 59-62. Article in Chinese, English abstract at <http://www.ncbi.nlm.nih.gov/pubmed>, ID 18247306.
- Litt, M., and Luty, J. A. (1989). A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *American Journal of Human Genetics* **44**: 397-401.
- Liu, B. H. (1998). Statistical Genomics: Linkage, mapping, and QTL analysis. CRC Press LLC, Boca Raton, Florida.

Liu, L., Dilworth, D., Gao, L., Monzon, J., Summers, A., Lassam, N., and Hogg, D. (1999). Mutation of the *CDKN2A* 5' UTR creates an aberrant initiation codon and predisposes to melanoma. *Nature Genetics* **21**: 128-132.

Lizak, M.J., Mori, K., Ceckler, T.L., Balaban, R.S., and Kador, P.F. (1996). Quantitation of galactosemic cataracts in dogs using magnetization transfer contrast-enhanced magnetic resonance imaging. *Investigative Ophthalmology and Visual Science* **37**: 2219-2227.

Lopes-Cendes, I., Phillips, H. A., Scheffer, I. E., Mulley, J. C., Desbiens, R., Andermann, E., Cendes, F., Verret, S., Andermann, F., Berkovic, S. F., and Rouleau, G. A. (1995). Genetic linkage studies in familial frontal epilepsy: Exclusion of the human chromosome regions homologous to the *El-1* mouse locus. *Epilepsy Research* **22**: 227-233.

McKay, S.D., Schnabel, R.D., Murdoch, B.M., Aerts, J., Gill, C.A., Gao, C., Li, C., Matukumalli, L.K., Stothard, P., Wang, Z., Van Tassell, C.P., Williams, J.L., Taylor, J.F., and Moore, S.S., (2007). Construction of bovine whole-genome radiation hybrid and linkage maps using high-throughput genotyping. *Animal Genetics* **38**: 120-125.

Maddox, J.F., Davies, K.P., Crawford, A.M., Hulme, D.J., Vaiman, D., Cribs, E.P., Freking, B.A., Beh, K.J., Cockett, N.E., Kang, N., Riffkin, C.D., Drinkwater, R., Moore, S.S., Dodds, K.G., Lumsden, J.M., van Stijn, T.C., Phua, S.H., Adelson, D.L., Burkin, H.R., Broom, J.E., Buitkamp, J., Cambridge, L., Cushwa, W.T., Gerard, E., Galloway, S.M., Harrison, B., Hawken, R.J., Hiendleder, S., Henry, H.M., Medrano, J.F., Paterson, K.A., Schibler, L., Stone, R.T., and van Hest, B. (2001). An enhanced linkage map of the sheep genome comprising more than 1000 loci. *Genome Research* **11**: 1275-1289.

Marcantonio, J. M., Duncan, G., and Rink, H. (1986). Calcium-induced opacification and loss of protein in the organ-cultured bovine lens. *Experimental Eye Research* **42**: 617-30.

Marques, E., de Givry, S., Stothard, P., Murdoch, B., Wang, Z., Womack, J., and Moore, S.S., (2007). A high resolution radiation hybrid map of bovine chromosome 14 identifies scaffold rearrangement in the latest bovine assembly. *BMC Genomics* **8**: 254.

McLaughlin, S. A., Whitley, R. D., and Gilger, B. C. (1992). Diagnosis and treatment of lens diseases. *Veterinary Clinics of North America: Equine Practice* **8**: 575-85.

Mellersh, C.S., Pettitt, L., Forman, O.P., Vaudin, M., and Barnett, K.C. (2006). Identification of mutations in *HSF4* in dogs of three different breeds with hereditary cataracts. *Veterinary Ophthalmology* **9**: 369-378.

Min, J.-N., Zhang, Y., Moskophidis, D., and Mivechi, N.F. (2004). Unique contribution of heat shock transcription factor 4 in ocular lens development and fiber cell differentiation. *Genesis* **40**: 205-217.

Morton, N.E. (1955). Sequential tests for the detection of linkage. *American Journal of Human Genetics* **8** 80-96.

Mothobi, M.E., Guo, S., Liu, Y., Chen, Q., Yussuf, A.S., Zhu, X., and Fang, Z. (2009). Mutation analysis of congenital cataract in a Basotho family identified a new missense allele in *CRYBB2*. *Molecular Vision* **15**: 1470-1475. Müller, M., Bhattacharya, S.S., Moore, T., Prescott, Q., Wedig, T., Herrmann, H., and Magin, T.M. (2009). Dominant cataract formation in association with a vimentin assembly disrupting mutation. *Human Molecular Genetics* **18**: 1052-1057.

Mouse Genome Sequencing Consortium (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**: 520-62.

Murphy, A. M., MacHugh, D. E., Park, S. D. E., Scraggs, E., Haley, C. S., Lynn, D. J., Boland, M. P., and Doherty, M. L. (2007). Linkage mapping of the locus for inherited ovine arthrogyrosis (IOA) to sheep chromosome 5. *Mammalian Genome* **18**: 43-52.

Nakahara, M., Nagasaka, A., Koike, M., Uchida, K., Kawane, K., Uchiyama, Y., and Nagata, S. (2007). Degradation of nuclear DNA by DNase II-like acid DNase in cortical fiber cells of mouse eye lens. *FEBS Journal* **274**: 3055-64.

Nam, S., Min, K., Hwang, H., Lee, H., Lee, J.H., Yoon, J., Lee, H., Park, S., and Lee, J. (2009). Control of rapsyn stability by the CUL-3-containing E3 ligase complex. *Journal of Biological Chemistry* **284**: 8195-8206.

Narayanan, K., Ramachandran, A., Hao, J., He, G., Park, K.W., Cho, M., and George, A. (2003). Dual functional roles of dentin matrix protein 1. *Journal of Biological Chemistry* **278**: 17500-17508.

Newell, F.W., (1992). Ophthalmology principles and concepts. Seventh edition. Mosby Year Book, St Louis.

Nicholas, K.B., Nicholas H.B. Jr., and Deerfield, D.W. II. (1997) GeneDoc: Analysis and Visualization of Genetic Variation, *EMBNEW.NEWS* 4:14 available at <http://www.psc.edu/biomed/genedoc>

Nishimoto, S., Kawame, K., Watanabe-Fukunaga, R., Fukuyama, H., Ohsawa, Y., Uchiyama, Y., Hashida, N., Ohguro, N., Tano, Y., Morimoto, T., Fukuda, Y., and Nagata, S. (2003). Nuclear cataract caused by a lack of DNA degradation in the mouse eye lens. *Nature* **424**: 1071-4.

Oliver, P.L., Bitoun, E., Clark, J., Jones, E.L., and Davies, K.E. (2004). Mediation of Af4 protein function in the cerebellum by Siah proteins. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 14901-14906.

Pandey, S. K., Thakur, J., Werner, L., Wilson, E., Werner, L. P., Izak, A. M., and Apple, D. J. (2002). The human crystalline lens, ciliary body, and zonules: their relevance to presbyopia. In *Presbyopia: a surgical textbook Chapter 2*. Agarwal, A. (ed) Slack Incorporated, Thorofare, New Jersey.

Park, C., Meng, L., Stanton, L.H., Collins, R.E., Mast, S.W., Yi, X., Strachan, H., and Moremen, K.W. (2005). Characterization of a human core-specific lysosomal α 1,6-mannosidase involved in N-glycan catabolism. *Journal of Biological Chemistry* **280**: 37204-37216.

Pauling, L., Itano, H.A., Singer, S.J., and Wells, I.C. (1949). Sickle cell anemia, a molecular disease. *Science* **110**: 543-548.

Percin, E. F., Ploder, L. A., Yu, J. J., Arici, K., Horsford, D. J., Rutherford, A., Bapat, B., Cox, D. W., Duncan, A. M. V., Kalnins, V. I., Kocak-Altintas, A., Sowden, J. C., Traboulsi,

- E., Sarfarazi, M., and McInnes. (2000). Human microphthalmia associated with mutations in the retinal homeobox gene *CHX10*. *Nature Genetics* **25**: 397-401.
- Perraud, A.-L., Fleig, A., Dunn, C.A., Bagley, L.A., Launay, P., Schmitz, C., Stokes, A.J., Zhu, Q., Bessman, M.J., Penner, R., Kinet, J.-P., and Scharenberg, A.M. (2001). ADP-ribose gating of the calcium-permeable LTRPC2 channel revealed by Nudix motif homology. *Nature* **411**: 595-598.
- Perraud, A.-L., Shen, B., Dunn, C.A., Rippe, K., Smith, M.K., Bessman, M.J., Stoddard, B.L., and Scharenberg, A.M. (2002). NUDT9, a member of the nudix hydrolase family, is an evolutionarily conserved mitochondrial ADP-ribose pyrophosphatase. *The Journal of Biological Chemistry* **278**: 1794-1801.
- Perraud, A.-L., Takanish, C.L., Shen, B., Kang, S., Smith, M.K., Schmitz, C., Knowles, H.M., Ferraris, D., Li, W., Zhang, J., Stoddard, B.L., and Scharenberg, A.M. (2005). Accumulation of free ADP-ribose from mitochondria mediates oxidative stress-induced gating of TRPM2 cation channels. *The Journal of Biological Chemistry* **280**: 6138-6148.
- Petkov, P. M., Broman, K. W., Szatkiewicz, J. P., and Paigen, K. (2007). Crossover interference underlies sex differences in recombination rates. *Trends in Genetics* **23**: 539-42.
- Phadnis, N., and Fry, J. D. (2005). Widespread correlations between dominance and homozygous effects of mutations: implications for theories of dominance. *Genetics* **171**: 385-92.
- Piatigorsky, J. (2003). Crystallin genes: specialization by changes in gene regulation may precede gene duplication. *Journal of Structural and Functional Genomics* **3**: 131-137.
- Piatigorsky, J., O'Brien, W. E., Norman, B. L., Kalumuck, K., Wistow, G. J., Borrás, T., Nickerson, J. M., and Wawrousek, E. F. (1988). Gene sharing by δ -crystallin and argininosuccinate lyase. *Proceedings of the National Academy of Sciences of the United States of America* **85**: 3479-83.
- Piatigorsky, J., and Wistow, G. J. (1989). Enzyme/Crystallins: Gene sharing as an evolutionary strategy. *Cell* **57**: 197-9.

Pras, E., Frydman, M., Levy-Nissenbaum, E., Bakhan, T., Raz, J., Assia, E. I., Goldman, B., and Pras, E. (2000). A nonsense mutation (W9X) in *CRYAA* causes autosomal recessive cataract in an inbred Jewish Persian family. *Investigative Ophthalmology and Visual Science* **41**: 3511-5.

Pras, E., Levy-Nissenbaum, E., Bakhan, T., Lahat, H., Assia, E., Geffen-Carmi, N., Frydman, M., Goldman, B., and Pras, E. (2002). A missense mutation in the *LIM2* gene is associated with autosomal recessive presenile cataract in an inbred Iraqi Jewish family. *American Journal of Human Genetics* **70**: 1363-1367.

Pras, E., Raz, J., Yahalom, V., Frydman, M., Garzosi, H.J., Pras, E., and Hejtmancik, J.F. (2004) A nonsense mutation in the glucosaminyl (*N*-acetyl) transferase 2 gene (*GCNT2*): Association with recessive congenital cataracts. *Investigative Ophthalmology and Visual Science* **45**: 1940-1945.

Preacher, K. J. (2001). Caculation for the chi-square test: An interactive calculation tool for chi-square tests of goodness of fit and independence [computer software]. Available from <http://quantpsy.org>.

Ramachandran, R.D., Perumalsamy, V., and Hejtmancik, J.F. (2006). Autosomal recessive juvenile onset cataract associated with mutation in *BFSP1*. *Human Genetics* **121**: 475-482.

Risch, N., and Giuffra, L. (1992). Model misspecification and multipoint linkage analysis. *Human Heredity* **42**: 242-53.

Robertson, L. J. G. (2003) The involvement of Calpain II in the formation of the ovine heritable cataract. PhD. Thesis, Lincoln University, New Zealand.

Robertson, L. J. G., Morton, J. D., Yamaguchi, M., Bickerstaffe, R., Shearer, T.R., and Azuma, M. (2005). Calpain may contribute to hereditary cataract formation in sheep. *Investigative Ophthalmology and Visual Science* **46**: 4634-4640.

- Rodriguez I. R., Gonzalez, P., Zigler, J. S. Jr, and Borrás, T. (1992). A guinea-pig hereditary cataract contains a splice-site mutation in a crystallin gene. *Biochimica et Biophysica Acta* **1180**: 44-52
- Röll, B., Amons, R., and de Jong, W. W. (1996). Vitamin A₂ bound to cellular retinol-binding protein as ultraviolet filter in the eye lens of the gecko *Lygodactylus picturatus*. *Journal of Biological Chemistry* **271**: 10437-10440.
- Roshan, M., Vijaya, P. H., Lavanya, G. R., Shama, P. K., Santhiya, S. T., Graw, J., Gopinath, P. M., and Satyamoorthy, K. (2010). A novel human *CRYGD* mutation in a juvenile autosomal dominant cataract. *Molecular Vision* **16**: 887-896.
- Rothwell, N.V. (1993) Understanding genetics: A molecular approach. Wiley-Liss, New York. Pages 68-69.
- Rozen S, and Skaletsky H. (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods in Molecular Biology* **132**:365-86.
- Rubin, L.F., Koch, S.A., and Huber, R.J. (1969). Hereditary cataracts in Minature Schnauzers. *Journal of the American Veterinary Medical Association* **154**: 1456-1458.
- Ruiz-Perez, V.L., Tompson, S.W.J., Blair, H.J., Espinoza-Valdez, C., Lapunzina, P., Silva, E.O., Hamel, B., Gibbs, J.L., Young, I.D., Wright, M.J., and Goodship, J.A. (2003). Mutations in two nonhomologous genes in a head-to-head configuration cause Ellis-van Creveld syndrome. *American Journal of Human Genetics* **72**: 728-732.
- Safieh, L.A., Khan, A.O., and Alkuraya, F.S. (2009). Identification of a novel *CRYAB* mutation associated with autosomal recessive juvenile cataract in a Saudi family. *Molecular Vision* **15**: 980-984.
- Sanderson, J., Marcantonio, J.M., and Duncan, G. (1996). Calcium ionophore induced proteolysis and cataract: inhibition by cell permeable calpain antagonists. *Biochemical and biophysical research communications* **218**: 893-901.

- Santana, A., Waiswol, M., Arcieri, E.S., de Vasconcellos, J.P.C., and de Melo, M.B. (2009). Mutation analysis of *CRYAA*, *CRYGC*, and *CRYGD* associated with autosomal dominant congenital cataract in Brazilian families. *Molecular Vision* **15**: 793-800.
- Sato, T., Irie, S., Kitada, S., and Reed, J. C. (1995). FAP-1: A protein tyrosine phosphatase that associates with Fas. *Science* **268**: 411-415.
- Saunders, L.Z., and Fincher, M.G. (1951). Hereditary multiple eye defects in Grade Jersey calves. *Cornell Veterinarian* **41**: 351-366.
- Schmidt, W., Klopp, N., Illig, T., and Graw, J. (2008). A novel *GJA8* mutation causing a recessive triangular cataract. *Molecular Vision* **14**: 851-856.
- Schmitt, C., and Hockwin, O. (1990). The mechanisms of cataract formation. *Journal of Inherited and Metabolic Disease* **13**: 501-16.
- Schuelke, M. (2000). An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology* **18**: 233-234.
- Schymick, J. C., Scholz, S. W., Fung, H.-C., Britton, A., Arepalli, S., Gibbs, J. R., Lombardo, F., Matarin, M., Kasperaviciute, D., Hernandez, D. G., Crews, C., Bruijin, L., Rothstein, J., Mora, G., Restagno, G., Chió, Singleton, A., Hardy, J., and Traynor, B. J. (2007). Genome-wide genotyping in amyotrophic lateral sclerosis and neurologically normal controls: first stage analysis and public release of data. *Lancet Neurology* **6**: 322-28.
- Sechi, T., Coltman, D. W., and Kijas, J. W. (2009). Evaluation of 16 loci to examine the cross-species utility of single nucleotide polymorphism arrays. *Animal Genetics* **41**: 199-202.
- Semina, E.V., Brownell, I., Mintz-Hittner, H.A., Murray, J.C., and Jamrich, M. (2001). Mutations in the human forkhead transcription factor *FOXE3* associated with anterior segment ocular dysgenesis and cataracts. *Human Molecular Genetics* **10**: 231-236.
- Shen, B. W., Perraud, A.-L., Scharenberg, A., and Stoddard, B.L. (2003). The crystal structure and mutational analysis of human NUDT9. *Journal of Molecular Biology* **332**: 385-398.

- Shiels, A., Mackay, D., Ionides, A., Berry, V., Moore, A., and Bhattacharya. (1998). A missense mutation in the human connexin50 gene (*GJA8*) underlies autosomal dominant “zonular pulverulent” cataract, on chromosome 1q. *American Journal of Human Genetics* **62**: 526-532.
- Shiels, A., Bennett, T.M., Knopf, H.L.S., Yamada, K., Yoshiura, K., Niikawa, N., Shim, S., and Hanson, P.I. (2007). *CHMP4B*, a novel gene for autosomal dominant cataracts linked to chromosome 20q.
- Shih, J.-Y., Lee, Y.-C.G., Yang, S.-C., Hong, T.-M., Huang, C.-Y.F., and Yang, P.-C. (2003). Collapsin response mediator protein-1: A novel invasion-suppressor gene. *Clinical and Experimental Metastasis* **20**: 69-76.
- Sipiczki, M. (1988). The role of sterility genes (*ste* and *aff*) in the initiation of sexual development in *Schizosaccharomyces pombe*. *Molecular and General Genetics* **213**: 529-534.
- Sittmann, K., Abplanalp, H., and Fraser, R. A. (1966). Inbreeding depression in Japanese quail. *Genetics* **54**: 371-9.
- Slingsby, C., and Clout, N. J. (1999). Structure of the crystallins. *Eye* **13**: 395-402.
- Smith, W. J. M., Li, Y., Ingham, A., Collis, E., McWilliam, S. M., Dixon, T. J., Norris, B. J., Mortimer, S. I., Moore, R. J., and Reverter, A. (2010). A genomics-informed, SNP association study reveals *FBLN1* and *FABP4* as contributing to resistance to fleece rot in Australian Merino sheep. *BMC Veterinary Research* **6**: 27.
- Sobol, R.W., Foley, J.F., Nyska, A., Davidson, M.G., and Wilson, S.H. (2003). Regulated over-expression of DNA polymerase β mediates early onset cataract in mice. *DNA Repair* **2**: 609-622.
- Steindler, C., Li, Z., Algarté, M., Alcover, A., Libri, V., Ragimbeau, J., and Pellegrini, S. (2004). Jamip1 (*Marlin1*) defines a family of proteins interacting with janus kinases and Sturtevant microtubules. *Journal of Biological Chemistry* **279**: 43168-43177.

Sturtevant, A. H. (1913). The linear arrangement of six sex-linked factors in *Drosophila*, as shown by their mode of association. *Journal of Experimental Zoology* **14**: 43-59.

Sundin, O. H., Broman, K. W., Chang, H. H., Vito, E. C. L., Stark, W. J., and Gottsch, J. D. (2006). A common locus for late-onset Fuchs corneal dystrophy maps to 18q21.2-q21.32. *Investigative Ophthalmology and Visual Science* **47**: 3919-3926.

Sullivan, M.M., and Sage, E.H. (2004). Hevin/SC1, a matricellular glycoprotein and potential tumor-suppressor of the SPARC/BM-40/Osteonectin family. *International Journal of Biochemistry and Cell Biology* **36**: 991-996.

Sweeney, M.H.J., and Truscott, R.J.W. (1998). An impediment to glutathione diffusion in older normal human lenses: a possible precondition for nuclear cataract. *Experimental Eye Research* **67**: 587-595.

Takei, D., Ishihara, H., Yamaguchi, S., Yamada, T., Tamura, A., Katagiri, H., Maruyama, Y., and Oka, Y. (2006). WFS1 protein modulates the free Ca^{2+} concentration in the endoplasmic reticulum. *FEBS Letters* **580**: 5635-5640.

Tanabe, M., Sasai, N., Nagata, K., Liu, X.-D., Liu, P. C. C., Thiele, D. J., and Nakai, A. (1999). The mammalian *HSF4* gene generates both an activator and a repressor of heat shock genes by alternative splicing. *The Journal of Biological Chemistry* **274**: 27845-6.

Tarailo-Graovac, M., and Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics* **4**:1-14.

Taylor, H. R., West, S. K., Rosenthal, F. S., Muñoz, B., Newland, H. S., Abbey, H., and Emmett, E. A. (1988). Effect of ultraviolet radiation on cataract formation. *The New England Journal of Medicine* **319**: 1429-33.

Tayo, B. O., DiCioccio, R. A., Liang, Y., Trevisan, M., Cooper, R. S., Lele, S., Sucheston, L., Piver, S. M., and Odunsi, K. (2009). Complex segregation analysis of pedigrees from the Gilda Radner Familial Ovarian Cancer Registry reveals evidence for Mendelian dominant inheritance. *PloS ONE* **4**: e5939.

Theil, E. C. (2003). Ferritin: At the crossroads of iron and oxygen metabolism. *The Journal of Nutrition* **133**: 1549S-1553S.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**: 4673-4680.

Torres, J. Z., Ban, K. H., and Jackson, P. K. (2010). A specific form of phosphor protein phosphatase 2 regulates anaphase-promoting complex/cyclosome association with spindle poles. *Molecular Biology of the Cell* **21**: 897-904.

Tripathi, B.J., Tripathi, R.C., Borisuth, N.S., Dhaliwal R., and Dhaliwal, D., (1991). Rodent models of congenital and hereditary cataract in man. *Lens and Eye Toxicity Research* **8**: 373-413.

Truscott, R. J. W. (2003). Human cataract: the mechanisms responsible; light and butterfly eyes. *The International Journal of Biochemistry and Cell Biology* **35**: 1500-4.

Truscott, R. J. W. (2005). Age-related nuclear cataract – oxidation is the key. *Experimental Eye Research* **80**: 709-725.

Truscott, R. J., and Augusteyn, R. C. (1977a). Oxidative changes in human lens proteins during senile nuclear cataract formation. *Biochimica et Biophysica Acta* **492**: 43-52.

Truscott, R. J., and Augusteyn, R. C. (1977b). The state of sulphhydryl groups in normal and cataractous human lenses. *Experimental Eye Research* **25**: 139-48.

Turner, D. J., Keane, T. M., Sudbery, I., Adams, D. J. (2009). Next-generation sequencing of vertebrate experimental organisms. *Mammalian Genome* **20**: 327-338.

van Heyningen, R. (1965). The metabolism of glucose by the rabbit lens in the presence and absence of oxygen. *Biochemical Journal* **96**: 419-431.

van Heyningen, R. (1972). The human lens, 1: a comparison of cataracts extracted in Oxford (England) and Shikarrpur (W. Pakistan). *Experimental Eye Research* **13**: 136-47.

Vanita, V., Singh, J.R., Singh, D., Varon, R., and Sperling, K. (2008). A novel mutation in *GJA8* associated with cataract in a family of Indian origin. *Molecular Vision* **14**: 323-326.

Vanita, V., Singh, J.R., Singh, D., Varon, R., and Sperling, K. (2009). Novel mutation in the γ -S crystallin gene causing autosomal dominant cataract. *Molecular Vision* **15**: 476-481.

Venter, *et al* (273 authors) (2001). The sequence of the human genome. *Science* **291**: 1304-1351.

Vogt, O. (1926) Psychiatrically relevant information from zoological and botanical systematics. *Zeitschrift für die gesamte Neurologie und Psychiatrie* **101**: 805-832. (German)

Weber, J.L., and May, P.E. (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *American Journal of Human Genetics* **44**: 388-396.

Wen, J.-K., Osumi, T., and Hashimoto, T. (1990). Molecular analysis of human acatalasemia: Identification of a splicing mutation. *Journal of Molecular Biology* **211**: 383-393.

Wallace, M.R., MacKay, E.O., Gelatt, K.N., and Andrew, S.E. (2005). Inheritance of cataract in the Bichon Frise. *Veterinary Ophthalmology* **8**: 203-205.

Wistow, G. J., Lietman, T., Williams, L. A., Stapel, S. O., de Jong, W. W., Horwitz, J., and Piatigorsky, J. (1988). τ -Crystallin/ α -enolase: One gene encodes both an enzyme and a lens structural protein. *The Journal of Cell Biology* **107**: 2729-36.

Womack, J. E. (2005). Advances in livestock genomics: Opening the barn door. *Genome Research* **15**: 1699-1705

World Health Organisation (2000). Blindness: Vision 2020 - control of major blinding diseases and disorders. Online fact sheet at <http://www.who.int/mediacentre/factsheets/fs214/en/>

Wu, C.H., Jin, W., Nomura, K., Goldammer, T., Hadfield, T., Dalrymple, B.P., McWilliam, S., Maddox, J.F., and Cockett, N.E. (2009). A radiation hybrid comparative map of ovine chromosome 1 aligned to the virtual sheep genome. *Animal Genetics* **40**: 435-455.

Yan, M., Xiong, C., Ye, S.Q., Chen, Y., Ke, M., Zheng, F., and Zhou, X. (2008). A novel connexin 50 (*GJA8*) mutation in a Chinese family with a dominant congenital pulverulent cataract.

Yang, Z., Xiao, X., Li, S., and Zhang, Q. (2009). Clinical and linkage study on a consanguineous Chinese family with autosomal recessive high myopia. *Molecular Vision* **15**: 312-318.

Yu, A., Zhao, C., Fan, Y., Jang, W., Mungall, A. J., Deloukas, P., Olsen, A., Doggett, N. A., Ghebranious, N., Broman, K. W., and Weber, J. L. (2001). Comparison of human genetic and sequence-based maps. *Nature* **409**: 951-953.

Zandy, A. J., and Bassnett, S. (2007). Proteolytic mechanisms underlying mitochondrial degradation in the ocular lens. *Investigative Ophthalmology and Visual Science* **48**: 293-302

Zandy, A. J., Lakhan, S., Zheng, T., Flavell, R. A., and Bassnett, S. (2005). Role of the executioner caspases during lens development. *The Journal of Biological Chemistry* **280**: 30263-72.

Zarina, S., Zhao, H.-R., and Abraham, E.C. (2000). Advanced glycation end products in human senile and diabetic cataractous lenses. *Molecular and Cellular Biochemistry* **210**: 29-34.

Zhang, L., Fu, S., Ou, Y., Zhao, T., Su, Y., and Liu, P. (2009). A novel nonsense mutation in *CRYGC* is associated with autosomal dominant congenital nuclear cataracts and microcornea. *Molecular Vision* **15**: 276-282.

Zhang, X.-J., Liu, J.-B., Gui, J.-P., Li, M., Xiong, Q.-G., Wu, H.-B., Li, J.-X., Yang, S., Wang, H.-Y., Gao, M., Yang, J., and Yang, Q. (2004). Characteristics of genetic epidemiology and genetic models for vitiligo. *Journal of the American Academy of Dermatology* **51**: 383-390.

Zhu, M., and Zhao, S. (2007). Candidate gene identification approach: progress and challenges. *International Journal of Biological Sciences* **3**: 420-427.

Publications and Presentations Arising from This Thesis

Morton, J. D., Wilson, G. R., Anderson, R. M., Wilson S. J., Gately, K., McEwan, J. C., and Robertson, L. J. G. (2006). The gene causing inherited cataracts in sheep maps to chromosome 6. *Investigative Ophthalmology and Visual Science* **47**: E-2520.

Wilson, G. R., Morton, J. D., Anderson, R. M., Wilson S. J., Gately, K., McEwan, J. C., and Robertson, L. J. G. (2006). The gene causing inherited cataracts in sheep is located in a narrow region of chromosome 6. *Poster presented at ComBio 2006 conference in Brisbane, Australia.*

Wilson, G. R. (2006). Mapping a gene for inherited sheep cataract. Presentation to the 2nd New Zealand Molecular Mapping Workshop in Palmerston North, New Zealand, 9th-10th October 2006.

Wilson, G. R., Gately, K., Morton, J. D. (2009). *NUDT9* contains a genetic marker for heritable cataract in *Ovis aries*. *Poster presented at ComBio 2009 conference in Christchurch, New Zealand.*