

Molecular characterisation of canine progressive retinal atrophies

Louise Mary Downs

A thesis submitted for the degree of Doctor of Philosophy

Animal Health Trust and

UCL Institute of Ophthalmology

2013

Declaration

I, Louise Downs confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.



Louise Downs

7 October 2013

Acknowledgements

My thanks go, first and foremost to the AHT and my supervisors, without whom this thesis would not have been possible. My primary (non-UCL) supervisor Dr Cathryn Mellersh at the AHT for the opportunity, much support and freedom throughout, helping me create a thesis worth reading, and the many dog walks. I cannot thank you enough. Prof Robin Ali, my primary supervisor at UCL, for the opportunity and support from afar. Prof Alison Hardcastle, my secondary supervisor and graduate tutor at UCL, for your support and guidance throughout, and for the temporary lab and office space.

Secondly, many thanks go to everyone who contributed to the research, or my sanity, in some way. Everyone in the Canine Genetics group at the AHT for all the laughs and support, but especially Dr Mike Bournnell (for helping me learn what little I know about VBA and perl, and the many VERY useful spreadsheets), Oliver Forman (for pioneering the NGS work at the trust), my mentor Dr Debbie Guest (believe it or not, our chats helped immensely), and Lou Hayward and Dr Sally Ricketts (for putting up with Molly and me in the office; Sal, my substitute mentor, thank you especially for proofreading my thesis, advice and highly valued friendship). To external collaborators, Dr Tomas Bergström, for your valuable contributions to the Golden Retriever work and the many laughs (in the guise of “work”) at meetings/conferences. Jerold Bell, for your valuable contribution to the Gordon Setter work. Finally, everyone at the UCL IoO who helped me and made me feel welcome and at home, however briefly I was there. In particular, Alice Davidson, Panagiotis Sergouniotis and Naheed Kanuga.

Thank you to everyone who supported this work with financial contributions: the Kennel Club Charitable Trust, the European Commission (FP7-LUPA, GA-201370), the Research Fund of the Swedish Kennel Club, the European Science Foundation, the Guide Dogs for the Blind Association, Pet Plan Charitable Trust, TarTan Gordon Setter Club and dog owners and breed clubs for their donations and fundraising.

My precious family, Mom, Dad, Tish and Victor, thank you for your unwavering love, support and belief in me. Vic, you had the misfortune of sharing me with this thesis, thank you for playing second fiddle without complaint. Finally, to my extended family and friends (there are too many to mention) for the words of encouragement and lots of laughs along the way. You all helped to keep me sane.

I am blessed indeed! XoXo

Abstract

Progressive retinal atrophy (PRA) in dogs is characterised by the degeneration of the photoreceptor cells of the retina, resulting in vision loss and eventually complete blindness. The condition affects more than 100 breeds, and is known to be genetically heterogeneous between breeds. Fourteen genes have been identified that are associated with PRA in at least 45 breeds, but for the majority of breeds the mutation(s) responsible have yet to be identified. This project utilised a genome-wide association mapping approach followed by sequencing of candidate genes or associated loci to characterise the molecular basis of PRA in three breeds: Golden Retrievers (GR), Gordon Setters (GoS) and Tibetan Spaniels (TS). Homozygous recessive mutations in four genes were identified as disease causing: *SLC4A3*_{c.2601_2602insC} and *TTC8*_{c.669delA} in GRs (PRA-1 and PRA-2 respectively), *C2ORF71*_{c.3149_3150insC} in GoSs (RCD4) and *FAM161A*_{c.1758-15_1758-16ins238} in TSs (PRA-3). *SLC4A3* is a novel retinal atrophy gene. Canine isoforms of each of the genes identified were confirmed by sequencing retinal messenger RNA. DNA genotyping tests for each of the mutations were used to determine disease allele frequencies in the affected breeds. PRA-1 and PRA-2 were found to be mainly restricted to the GR breed, with allele frequencies of 0.04 and 0.01 respectively. RCD4 and PRA-3, with allele frequencies of 0.36 and 0.15 respectively, were also identified as the cause of PRA in at least one other breed. While none of the *SLC4A3* variants identified in DNA from 200 human patients with autosomal recessive retinal degeneration were considered to be disease-causing mutations, the gene remains a plausible candidate for rare forms of retinal degeneration in humans. This research has identified four new potential models for human retinal degeneration and has resulted in the development of diagnostic DNA tests for each of the canine diseases, which will facilitate the elimination of PRA from affected breeds.

Table of Contents

Declaration	2
Acknowledgements	3
Abstract	4
Table of Contents	5
List of Figures	9
List of Tables.....	11
List of Abbreviations.....	12
Publications and Presentations	15
Chapter 1 Introduction.....	16
1.1 The Eye	16
1.2 The Retina	17
1.2.1 Retinal Pigment Epithelium	17
1.2.2 The Neurosensory Retina	18
1.2.2.1 Photoreceptors	19
1.2.2.2 Connecting Cilium.....	21
1.2.3 The Phototransduction Cascade	23
1.2.4 The Visual Cycle.....	25
1.2.5 The Canine Retina.....	26
1.3 Canine Genetics.....	28
1.3.1 History of Canine Breeds	28
1.3.2 Canine Genetic Resources.....	28
1.3.3 The Canine Genome.....	29
1.4 Canine Hereditary Disease	30
1.4.1 Canine Retinopathies.....	30
1.4.2 Progressive Retinal Atrophy	36
1.4.2.1 Ophthalmoscopic and Clinical Signs of PRA	36
1.4.2.2 Diagnosis of PRA	36
1.5 Human Retinal Degenerations.....	38
1.5.1 Retinitis Pigmentosa.....	39
1.5.2 Other Non-syndromic Retinal Dystrophies.....	39
1.5.3 Syndromic Disorders: Bardet-Biedl Syndrome	40
1.6 Canine Models for Human Genetic Disease	41
1.6.1 Canine Models for Retinal Gene Therapy Studies.....	41
1.6.1.1 LCA and <i>RPE65</i>	42
1.6.1.2 X-linked Retinal Dystrophy and <i>RPGR</i>	42
1.7 Identifying Variants Associated With Canine Genetic Traits	44
1.7.1 Genetic Markers	44
1.7.2 Candidate Gene Approach	44
1.7.3 Linkage Analysis.....	45
1.7.4 Case-Control Association Analysis.....	45
1.7.5 DNA Testing	47
1.8 Breeds Under Investigation	49
1.9 Aims and Objectives	49
Chapter 2 Materials and Methods.....	50
2.1 Canine DNA Sample Collection	50
2.1.1 PRA Diagnosis and Sample Identification.....	50
2.1.1.1 Exclusion of Known PRA Mutations	50
2.1.1.2 Ophthalmoscopic Examination	51

2.1.2	Sample Collection	51
2.1.2.1	Samples for DNA extraction	51
2.1.2.2	Samples for mRNA extraction	51
2.1.3	DNA and RNA Extraction	51
2.1.3.1	DNA Extraction from Blood	51
2.1.3.2	DNA Extraction of Buccal Mouth Swabs	52
2.1.3.3	Messenger RNA Extraction from Retina	53
2.1.3.4	Messenger RNA Extraction from Blood	53
2.1.4	DNA Normalisation	54
2.2	Polymerase Chain Reaction	55
2.2.1	Primers	55
2.2.2	Thermal Cycling Reactions	56
2.2.3	Thermal Cycling Parameters	56
2.3	Genome-Wide Association Mapping	57
2.3.1	SNP Genotyping.....	57
2.3.2	SNP Genotyping Data Analysis	57
2.3.3	Microsatellite Genotyping.....	59
2.4	Sequencing of Loci and Candidate Genes.....	60
2.4.1	Sanger Sequencing	60
2.4.1.1	Canine Gene Annotation	60
2.4.1.2	Complimentary DNA First Strand Synthesis	61
2.4.1.3	Primers and Amplification	61
2.4.1.4	Agarose Gel Electrophoresis	61
2.4.1.5	Purification of PCR Products	62
2.4.1.6	Sequencing Reaction	62
2.4.1.7	Isopropanol Precipitation.....	63
2.4.1.8	Sequence Data Analysis	63
2.4.2	Targeted Next Generation Sequencing	63
2.4.2.1	Library Preparation – RNA Baits	66
2.4.2.2	Library Preparation – Target Enrichment.....	66
2.4.2.3	Cluster Generation.....	67
2.4.2.4	Next Generation Sequencing.....	68
2.4.2.5	NGS Data Analysis.....	68
2.5	Quantification of mRNA transcripts	71
2.6	Variant Screening	72
2.6.1	Variant Screening Methods	72
2.6.1.1	Amplified Fragment Length Polymorphism Analysis	72
2.6.1.2	TaqMan® Allelic Discrimination.....	72
2.6.1.3	Sanger Sequencing	73
2.6.2	Population Screening of Variants.....	73
2.6.2.1	Sample Cohorts Collected for Research.....	73
2.6.2.2	Genetic Services Testing Data.....	74
2.7	Human Variant Identification.....	75
2.7.1	Sample Identification and Collection.....	75
2.7.2	Sanger Sequencing of <i>SLC4A3</i>	75
2.7.2.1	Purification of PCR Products	75
2.7.2.2	Sequencing Product Clean-up	75
2.7.2.3	Variant Pathogenicity Analysis	76
2.8	Bioinformatics Tools.....	77
2.9	Buffers and Solutions	78
Chapter 3 Golden Retrievers.....		79
3.1	Introduction and background	79

3.2	Diagnoses and Clinical Findings.....	80
3.3	Prcd Variant Screening.....	81
3.4	SNP Genotyping.....	81
3.4.1	Genome Wide Association Mapping	81
3.4.2	Homozygosity Analysis	83
3.5	Sequencing of Candidate Gene <i>SLC4A3</i>	85
3.5.1	Characterising <i>SLC4A3</i> Retinal Transcripts in Dog Retina	85
3.5.2	Sanger Sequencing of <i>SLC4A3</i>	85
3.5.3	<i>SLC4A3</i> _{c.2601_2602insC} Variant Validation.....	88
3.6	SNP Genotyping (PRA-2).....	91
3.6.1	Genome Wide Association Mapping	91
3.6.2	Homozygosity and Haplotype Analysis.....	93
3.7	Sequencing of PRA-2 Critical Regions.....	95
3.7.1	Target Enrichment of PRA-2 Critical regions.....	95
3.7.2	NGS Data Analysis – Target Capture Efficiency.....	96
3.7.3	NGS Data Analysis – Variant Identification.....	96
3.7.4	Characterising <i>SPATA7</i> and <i>TTC8</i> Retinal Transcripts.....	97
3.7.5	<i>TTC8</i> _{c.669delA} Variant Validation.....	100
3.7.6	PRA-2 Phenotype Questionnaire	102
3.8	Population screening of <i>SLC4A3</i> _{c.2601_2602insC} and <i>TTC8</i> _{c.669delA}	103
3.8.1	Research Sample Cohorts.....	103
3.8.2	AHT Genetic Services Cohorts	105
3.9	Screening of <i>SLC4A3</i> in humans.....	108
3.9.1	Introduction and Background.....	108
3.9.2	Sample and Assay Selection	108
3.9.3	Variants identified	110
3.10	Discussion	114
3.10.1	PRA-1 and <i>SLC4A3</i>	114
3.10.2	PRA-2 and <i>TTC8</i>	117
3.10.3	Prevalence of PRA-1 and PRA-2 in GR populations.....	121
3.10.4	<i>SLC4A3</i> in Humans	122
3.11	Conclusion.....	124
Chapter 4 Gordon Setters.....		126
4.1	Introduction and background	126
4.2	Diagnoses and Clinical Findings.....	127
4.3	SNP Genotyping.....	129
4.3.1	Genome-Wide Association Mapping	129
4.3.2	Homozygosity Analysis	131
4.4	Sequencing of Candidate Gene <i>C2ORF71</i>	133
4.4.1	Characterising <i>C2ORF71</i> Retinal Transcript	133
4.4.2	Sanger Sequencing of <i>C2ORF71</i>	133
4.4.3	<i>C2ORF71</i> _{c.3149_3150insC} Variant Validation.....	135
4.5	<i>C2ORF71</i> _{c.3149_3150insC} Population Screening	136
4.5.1	Research Sample Cohorts.....	136
4.5.2	AHT Genetic Services Cohort.....	138
4.5.3	Ophthalmoscopic Follow-Up	141
4.5.4	RCD4 Haplotype Comparison	143
4.6	Discussion	144
4.7	Conclusion.....	149
Chapter 5 Tibetan Spaniels		150
5.1	Introduction and Background.....	150

5.2	Diagnoses and Clinical Signs	150
5.3	SNP Genotyping	151
5.3.1	Genome-Wide Association Mapping	151
5.3.2	Homozygosity Analysis and Microsatellite Finemapping	153
5.4	Sequencing of the TS_PRA locus on CFA10	155
5.4.1	Target Enrichment of the TS_PRA Critical Region.....	155
5.4.2	NGS Data Analysis – Target Capture Efficiency.....	155
5.4.3	NGS Data Analysis – Variant Identification.....	156
5.4.3.1	Sanger Sequencing of <i>FAM161A</i> Insertion	158
5.4.3.2	Poly(A) Tract Length	159
5.4.4	Characterising <i>FAM161A</i> Retinal Transcripts	160
5.4.5	<i>FAM161A</i> _{c.1758-15_1758-16ins238} Variant Validation	162
5.4.5.1	AFLP Genotyping Assay.....	162
5.4.5.2	<i>FAM161A</i> _{c.1758-15_1758-16ins238} Genotyping	164
5.4.5.3	Sanger Sequencing of Blood and Retinal mRNA Transcripts	166
5.4.5.4	Quantitative PCR of <i>FAM161A</i> mRNA Transcripts	167
5.4.5.5	<i>FAM161A</i> _{c.1758-15_1758-16ins238} Effect on Protein	170
5.5	<i>FAM161A</i> _{c.1758-15_1758-16ins238} (PRA-3) Population Screening	171
5.5.1	Research Sample Cohorts.....	171
5.6	Discussion	173
5.7	Conclusion.....	179
	Chapter 6 General Discussion.....	180
6.1	Aims of The Current Project	180
6.1.1	Aim 1: Underlying Molecular Basis	180
6.1.2	Aim 2: Diagnostic DNA Test.....	181
6.1.3	Aim 3: Relevance to Humans.....	182
6.2	Sample Collection	184
6.2.1	Tissue Availability	184
6.2.2	Phenotyping.....	184
6.2.3	Allele Frequency Estimations	185
6.3	GWA Population Structure	187
6.3.1	Q-Q Plots.....	187
6.3.2	Multidimensional Scaling	189
6.3.3	Correction for Population Stratification.....	190
6.4	Sequencing	191
6.4.1	Sequencing Limitations.....	191
6.4.2	NGS Challenges	192
6.5	Genotype-Phenotype Discordance	193
6.6	Future prospects	195
6.6.1	Identification of Remaining PRA Mutations	195
6.6.2	In Vitro Functional Investigations.....	196
6.7	Conclusions	197
	References.....	198
	Appendices.....	211
I.	Genome Wide Association Q-Q and MDS Plots	212
II.	Primers	214
III.	PCR and Sequencing Reaction Components	227
IV.	Thermal Cycling Parameters.....	230
V.	Genes found in PRA-Associated Critical Regions.....	232
VI.	NGS Analysis Pipelines	240
VII.	GR PRA-2 Questionnaire.....	242

List of Figures

Figure 1. Cross section of the canine eye.....	17
Figure 2. Cross section of the canine retina.....	18
Figure 3. Photoreceptor cells.....	20
Figure 4. Connecting cilium and intraflagellar transport.....	22
Figure 5. Phototransduction cascade.....	24
Figure 6. Visual cycle.....	25
Figure 7. Classification of the canine retinopathies.....	32
Figure 8. CLAD alleles in the IrS population.....	48
Figure 9. Next-generation sequencing using sequencing by synthesis technology.....	65
Figure 10. Next generation sequencing analysis pipeline.....	70
Figure 11. Fundus changes observed in typical PRA in GRs.....	80
Figure 12. Genome-wide association mapping of PRA in GRs.....	82
Figure 13. Homozygosity analysis of GRs with PRA.....	84
Figure 14. Graphical comparison of intron-exon boundaries of <i>SLC4A3</i>	87
Figure 15. Effect of <i>SLC4A3</i> _{c.A44C} and <i>SLC4A3</i> _{c.2601_2602insC} on the SLC4A3 protein.....	88
Figure 16. Segregation of PRA-1 in a GR pedigree.....	90
Figure 17. Genome-wide association mapping of PRA2 in GRs.....	92
Figure 18. Homozygosity analysis of GRs with PRA2.....	94
Figure 19. Haplotype analysis in GRs with PRA2.....	95
Figure 20. IGV display of the 1-bp deletion in <i>TTC8</i>	97
Figure 21. Graphical comparison of the intron-exon boundaries of <i>TTC8</i>	99
Figure 22. Effect of <i>TTC8</i> _{c.669delA} on the <i>TTC8</i> protein.....	99
Figure 23. Segregation of PRA-2 in a GR pedigree.....	101
Figure 24. Human <i>SLC4A3</i> gene and protein isoforms.....	110
Figure 25. Effects of potentially pathogenic SNP variants on SLC4A3 protein.....	112
Figure 26. Predicted effects of two splice-site variants on the SLC4A3 protein.....	112
Figure 27. Comparison of human and canine retinal disease mutations.....	120
Figure 28. Geographical distribution of GR PRA variants.....	125
Figure 29. Fundus changes observed in a GoS with PRA.....	127
Figure 30. Retinal function measured in a GoS with PRA.....	128
Figure 31. Genome-wide association mapping of PRA in GoSs.....	130
Figure 32. Homozygosity analysis of Gordon Setters with PRA.....	132
Figure 33. Graphical comparison of the intron-exon boundaries of <i>C2ORF71</i>	134
Figure 34. Effect of variant <i>C2ORF71</i> _{c.3149_3150insC} on the C2ORF71 protein.....	134
Figure 35. Haplotypes comparison in five RCD4-affected breeds.....	143
Figure 36. Segregation of RCD4 in a GoS pedigree.....	146
Figure 37. Genome-wide association mapping of PRA in Tibetan Spaniels.....	152
Figure 38. Homozygosity analysis of TSs with PRA.....	154
Figure 39. IGV display of the >50 bp insertion in <i>FAM161A</i>	157
Figure 40. Sequence of SINE insertion.....	158
Figure 41. AFLP assay of SINE insertion.....	159
Figure 42. Graphical comparison of the intron-exon boundaries of <i>FAM161A</i>	161
Figure 43. <i>FAM161A</i> _{c.1758-15_1758-16ins238} genotyping assay.....	163

Figure 44. Segregation of PRA-3 in a TS pedigree	165
Figure 45. Comparison of <i>FAM161A</i> mRNA isoforms	169
Figure 46. Effect of <i>FAM161A</i> _{c.1758-15_1758-16ins238} on the protein product	170
Figure 47. <i>FAM161A</i> _{c.1758-15_1758-16ins238} effect on pre-mRNA splicing.....	175
Figure 48. Q-Q plots to assess population stratification	188
Figure 49. Multidimensional scaling plot	189
Figure 50. Mutations accounting for PRA in multiple breeds	194
Figure 51. Segregation of non-RCD4 PRA in a GoS family	196

List of Tables

Table 1. Photopigments in humans	21
Table 2. Classification and known mutations of canine retinopathies.....	33
Table 3. SNP genotyping summary	59
Table 4. Ensembl gene identification numbers for alignment	61
Table 5. Quantitative PCR thermal cycling conditions.....	67
Table 6. Research samples screened for each mutation	73
Table 7. Samples screened for each mutation through the AHT Genetic Services	74
Table 8. Capture efficiency for PRA-2 targeted NGS	96
Table 9. <i>SLC4A3</i> genotypes and PRA clinical status for 709 GRs.....	103
Table 10. <i>TTC8</i> genotypes and PRA clinical status for 660 GRs	104
Table 11. GR population screening for PRA-1 and PRA-2.....	105
Table 12. PRA-1 allele frequency breakdown by country (n>19).....	106
Table 13. PRA-2 allele frequency breakdown by country (n>9).....	107
Table 14. Patients with mapped homozygous regions containing <i>SLC4A3</i>	109
Table 15. Potentially pathogenic SNP variants.....	113
Table 16. <i>C2ORF71</i> genotypes and PRA clinical status for 126 GoSs	136
Table 17. <i>C2ORF71</i> genotypes and PRA clinical status for 72 IrSs	137
Table 18. Population screening for RCD4	138
Table 19. GoS RCD4 breakdown by country (n>19)	140
Table 20. Ophthalmoscopic examination and ERG analysis of <i>C2ORF71</i> ^{-/-} dogs	142
Table 21. Capture efficiency for TS_PRA targeted NGS	156
Table 22. Standard curves of qPCR targets	168
Table 23. <i>FAM161A</i> genotypes and PRA clinical status for 247 TSs	171
Table 24. <i>FAM161A</i> genotypes and PRA clinical status for 76 TTs	172
Table 25. Population screening for PRA-3.	172
Table 26. Summary of results	183
Table 27. GWA inflation factors (λ)	190

List of Abbreviations

A	Adenine
AAV	Adeno-associated virus
ABCA4	ATP-binding cassette, subfamily 4
ACHM	Achromatopsia
ACTB	β -Actin
aD	Autosomal dominant
AD	TaqMan® Allelic Discrimination
AE3	Anion exchanger 3
AFLP	Amplified fragment length polymorphism
AHT	Animal Health Trust
aR	Autosomal recessive
ARCI	Autosomal recessive congenital ichthyosis
b	Bases (e.g. Gb)
BBS	Bardet Biedl Syndrome
bp	Base pairs
BPS	Branch point sequence
BVA	British Veterinary Association
C	Cytosine
C2ORF71	Chromosome 2 open reading frame 71
CA	Carbonic Anhydrase
CBR	Chesapeake Bay Retriever
CC	Connecting cilium
CD	Cone dystrophy
cDNA	Complementary DNA
CERF	Canine Eye Registry Foundation
CFA	Canis familiaris chromosome
cGMP	Cyclic guanosine monophosphate
CIDD	Canine Inherited Disorders Database
CLAD	Canine leukocyte adhesion deficiency
cLCA	Canine leber congenital amaurosis
CMH	Cochran-Mantel-Haenszel
CNG	Centre National de Génotypage
CRD	Cone rod dystrophy
Ct	Cycle threshold
Da	Dalton
DTT	Dithiothreitol
E	Amplification efficiency
ERG	Electroretinogram
ES	English Setter
EVS	Exome Variant Server
FAM161A	Family with sequence similarity 161, member A
FCR	Flat-Coated Retriever
FMM	Fast Mixed Model

g	Gram
G	Guanine
g (x g)	Gravitational force
GC	Guanylate cyclase
GCL	Ganglion cell layer
gDNA	Genomic DNA
GDP	Guanosine diphosphate
GMP	Guanosine monophosphate
GoS	Gordon Setter
GR	Golden Retriever
GTP	Guanosine triphosphate
GWA	Genome wide association
HRM	High-resolution melting
HTG	High Throughput Genomics (University of Oxford)
HWE	Hardy Weinberg Equilibrium
IBS	Identity-by-state
IDID	Inherited Diseases in Dogs
IFT	Intraflagellar transport
IGV	Integrated Genomics Viewer
INL	Inner nuclear layer
IRBP	Interstitial retinol-binding protein
IrS	Irish Setter
IRWS	Irish Red and White Setter
IS	Inner segment
ISDS	International Sheep Dog Society
ITGB2	Integrin beta
KC	Kennel Club
L	Litre
LA	Lhasa Apso
LCA	Leber congenital amaurosis
LD	Linkage disequilibrium
LINE	Long interspersed nuclear element
LR	Labrador Retriever
LRAT	Lecithin-retinol acyltransferase
MDS	Multidimensional scaling
MITF	Microphthalmia-associated transcription factor
MLHD	Miniature Long Hair Dachshund
MQ	Milli-Q filtered water
mRNA	Messenger RNA
NGS	Next generation sequencing
NMD	Nonsense mediated decay
OMIA	Online Mendelian Inheritance in Animals database
OMIM	Online Mendelian Inheritance in Man database
ONL	Outer nuclear layer
ORF	Open reading frame
OS	Outer segment

PCR	Polymerase chain reaction
PDE	cGMP phosphodiesterase
PRA	Progressive retinal atrophy
prcd	Progressive rod cone degeneration
qPCR	Quantitative PCR
R ²	Coefficient of correlation
RACE	Rapid amplification of cDNA ends
RBC	Red Blood Cell
RCD	Rod cone degeneration
RDH12	Retinol dehydrogenase
RDH5	11-cis-retinol dehydrogenase
RetNet	Retinal Information System database
RHO*	Rhodopsin, activated
RP	Retinitis pigmentosa
RPE	Retinal pigment epithelium
RPE65	RPE-specific isomerohydrolase
RPGR	Retinitis pigmentosa GTPase regulator
RQ	Relative expression
SARD	Sudden acquired retinal degeneration
SBS	Sequencing by synthesis
SINE	Short interspersed nuclear element
SLC4A3	Solute carrier family 4, anion exchanger, member 3
SNP	Single nucleotide polymorphism
SP	Standard Poodle
SPATA7	Spermatogenesis associated protein 7
STGD1	Startgardt disease type 1
T	Thymine
TBP	TATA box binding protein
TCEP	Tris(2-carboxyethyl)phosphine
T _m	Melting temperature
TPR	Tetratricopeptide repeat
TS	Tibetan Spaniel
TT	Tibetan Terrier
TTC8	Tetratricopeptide repeat domain 8
UFWater	Ultrapure filtered water
UTR	Untranslated region
XL	X-linked
λ	Lambda - Inflation factor
Metric prefixes	
p	Pico (10 ⁻¹²)
n	Nano (10 ⁻⁹)
μ	Micro (10 ⁻⁶)
m	Milli (10 ⁻³)
k	Kilo (10 ³)
M	Mega (10 ⁶)
G	Giga (10 ⁹)

Publications and Presentations

Publications

Downs LM, Wallin-Håkansson B, Boursnell M, Marklund S, Hedhammar Å, Truvé K, Hübinette L, Lindblad-Toh K, Bergström T and Mellersh CS. (2011) A Frameshift Mutation in Golden Retriever Dogs with Progressive Retinal Atrophy Endorses *SLC4A3* as a Candidate Gene for Human Retinal Degenerations. PLoS ONE 6(6): e21452.

Downs LM, Bell JS, Freeman J, Hartley C, Hayward LJ and Mellersh CS. (2013) Late-onset Progressive Retinal Atrophy in the Gordon and Irish Setter Breeds is Associated with a Frameshift Mutation in *C2ORF71*. Animal Genetics. 44(2): 169-177.

Downs LM, Wallin-Håkansson B, Bergström T and Mellersh CS. A Novel Mutation Associated with Progressive Retinal Atrophy in the Golden Retriever Breed Indicates the Condition is Heterogeneous within this Breed. Manuscript in preparation.

Downs LM & Mellersh CS. A SINE insertion that results in exon-skipping is associated with PRA in Tibetan Spaniels. Manuscript in preparation.

Presentations (Oral)

Downs, LM. (2010) A 700kb region is linked to PRA in Tibetan Spaniels. 32nd Conference for the International Society for Animal Genetics (ISAG), Edinburgh, UK.

Downs, LM. (2010) A Novel Mutation in *SLC4A3* is Associated with Progressive Retinal Atrophy in the Golden Retriever. 5th International Meeting of Canine and Feline Genetics, Baltimore, USA.

Downs, LM. (2012) Late-onset PRA is Associated with a Frameshift Mutation in *C2orf71*. 6th International Meeting for Advances in Canine and Feline Genomics and Inherited Diseases, Visby, Sweden.

Downs, LM. (2012) A Novel Mutation Indicates Heterogeneous PRA in the Golden Retriever. 6th International Meeting for Advances in Canine and Feline Genomics and Inherited Diseases, Visby, Sweden.

Downs, LM. (2013) A SINE Insertion that Results in Exon-skipping is Associated with PRA in Tibetan Spaniels and Tibetan Terriers. 7th International Meeting for Advances in Canine and Feline Genomics and Inherited Diseases, Cambridge, USA.

Chapter **1** Introduction

1.1 The Eye

The eye is an organ, in most animals, which responds to light and thereby provides arguably the most important of the five human senses, vision. The human eye is made up of many different parts that are all required for optimal vision. The canine eye is very similar to the human eye in size, shape and anatomical features (Figure 1), with four main differences (discussed in more detail in section 1.2.5):

- 1) Dogs and other animals have layers of additional cells directly behind the retina, termed the tapetum lucidum (tapetal fundus) or tapetum nigrum (non-tapetal fundus). The former is a semi-circular, reflective area in the upper posterior part of the eye, while the latter is dark brown and occupies the rest of the fundus¹.
- 2) Human eyes have a tiny, cone-rich area in the centre of the retina, termed the fovea, which forms the sharpest image. This is surrounded by a photoreceptor-dense area called the macula. Dogs do not have a fovea or macula, although they do have a functionally similar visual streak or area centralis.
- 3) Humans have three different types of cone photoreceptor cells enabling trichromatic vision. Dogs only have two and therefore have dichromatic vision.
- 4) Many animals have a third eyelid, known as a nictitating membrane. Humans only have what is thought to be a tiny remnant of a third eyelid in the inside corner of the eye.

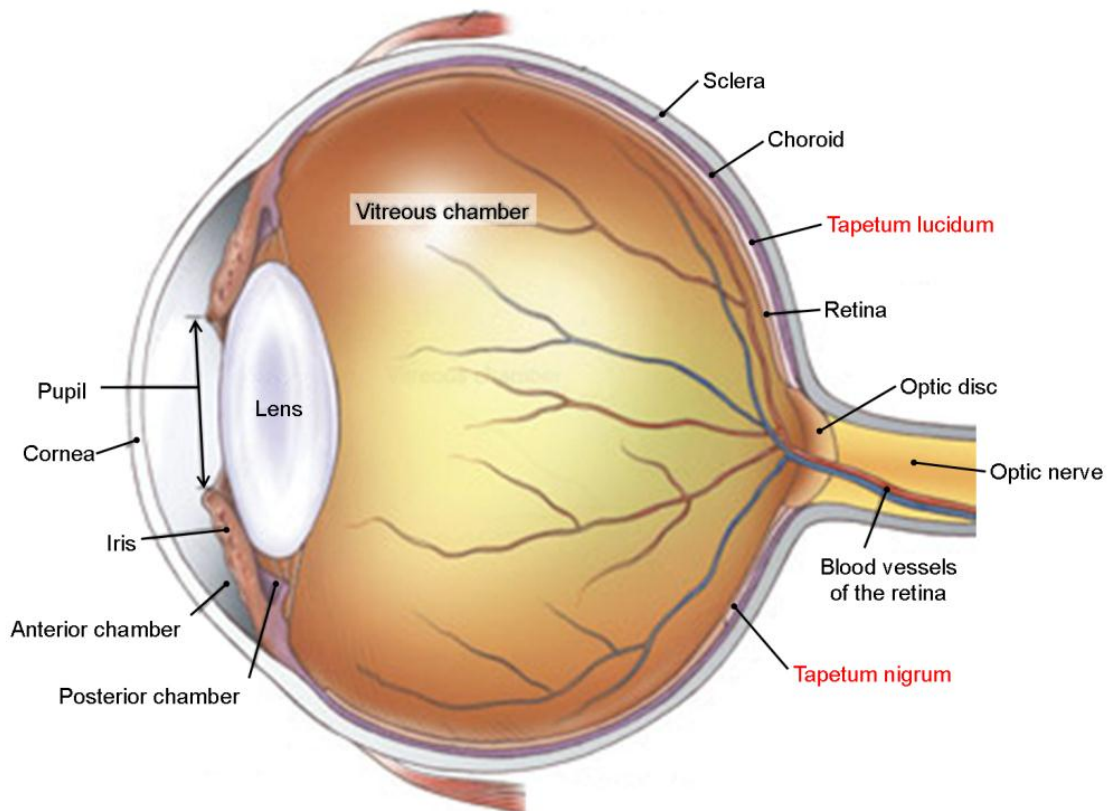


Figure 1. Cross section of the canine eye.

The canine eye and human eye are very similar in size, shape and anatomical features, but a few structures are unique to the canine eye: the 3rd eyelid, the gland of the 3rd eyelid (not shown), the tapetum lucidum and the tapetum nigrum (labelled in red). Image has been modified from www.biographixmedia.com

1.2 The Retina

The retina is the innermost layer of the eye, the main function of which is to convert captured light into neural signals to be transmitted to the brain for interpretation as vision. The retina is a complex, highly organised tissue that is composed of two main layers: the outer **retinal pigment epithelium** (RPE) and the inner **neurosensory retina** (Figure 2).

1.2.1 Retinal Pigment Epithelium

The RPE is a single layer of epithelial cells that separates the photoreceptor outer segments from the choroid. Microvilli of the RPE surround approximately one third of the photoreceptor cell outer segments and the RPE thereby maintains and supports the functions of the photoreceptor cells. Pigment granules containing melanin are found in the cytoplasm of the RPE. Their function is to prevent scattered light or light coming through the sclera from reaching the photoreceptors by absorbing it. In addition these

pigments protect the retina by scavenging free radicals. The main functions of the RPE are the formation of a barrier between the retina and blood circulation of the choroid, the transportation between the retina and choroid of ions, metabolites and water and phagocytosis of outer segments shed by photoreceptor cells. The RPE also participates in the retinoid cycle.

1.2.2 The Neurosensory Retina

The neurosensory retina is organised into two main layers: The outer nuclear layer (ONL), containing photoreceptor cells, makes up the **sensory retina** and is the site of phototransduction; the ganglion cell layer (GCL) and inner nuclear layer (INL), containing the remaining cell types, make up the **neural retina** in which the visual processing steps take place (Figure 2). The neurosensory retina is made up of six major classes of neuronal cells: rod and cone photoreceptor, bipolar, horizontal, ganglion and amacrine cells. These cells are responsible for the complex processing of signals from the photoreceptor cells to the optic nerve, which is formed by the axons of the ganglion cells. A number of non-neuronal, or glial, cells are also present: Müller cells, astrocytes and microglia. Müller cells are the predominant glia whose cell bodies occupy the INL and processes span all three layers. They are responsible for maintaining the structure and function of the retina.

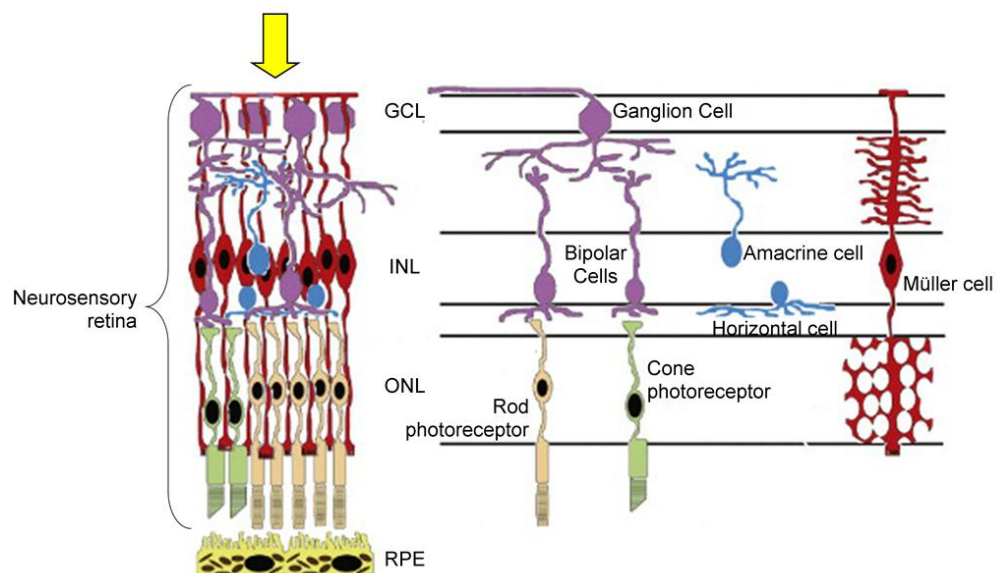


Figure 2. Cross section of the canine retina.

The direction from which light enters the retina is indicated by the yellow arrow. RPE – Retinal pigment epithelium; ONL – Outer nuclear layer; INL – Inner nuclear layer; GCL – Ganglion cell layer. Image has been modified from Smith et al.²

1.2.2.1 Photoreceptors

In humans there are approximately 120 million rod photoreceptor cells that are responsible for scotopic (low light) vision and 6 million cone photoreceptor cells that are responsible for photopic (bright light) vision³. Photoreceptor cells are most densely packed in the foveal region of humans and the visual streak or area centralis of dogs, and this is therefore where the ONL is thickest. In contrast, photoreceptor cells are absent from the optic nerve, resulting in a blind spot.

The photoreceptor cells consist of four main parts: an outer segment (OS), an inner segment (IS), a cell body and a synaptic terminal (Figure 3). The OS is made up of disks formed through folding of the photoreceptor cell membrane and that contain the photopigments, which are responsible for photon absorption. The rods have a long, cylindrical OS made up of discrete discs that are unattached to the plasma membrane. The cones have an OS made up of a short series of invaginations that are continuously connected to the membrane and that tapers distally. Photoreceptor outer segments are damaged by the light they absorb, they continuously shed their tips for phagocytosis by the RPE and new discs are formed at the base of the OS. The inner and outer segments are connected through the connecting cilium (CC).

The IS contains the basal body from which the CC originates, and other cellular organelles including mitochondria, endoplasmic reticulum, ribosomes and the Golgi apparatus. Components such as replacement proteins that are required for the renewal of outer segments are synthesized in the inner segments and transported through the cilium. In addition the energy required for phototransduction is provided by inner segment mitochondria. The cell body contains the photoreceptor nucleus and the synaptic terminal is made up of synaptic vesicles that store the chemical transmitters required for neural communication (Figure 3).

The photopigments of the photoreceptors are made up of an opsin protein that is covalently bound to a light-absorbing chromophore, 11-*cis*-retinal. The opsin varies between different cell types and makes them functionally distinguishable. Rods have only a single type of photopigment known as rhodopsin. In cones the opsin component varies and each opsin is sensitive to specific ranges of the light spectrum. In humans there are three cone types, each of which contains an opsin that absorbs light at a specific wavelength (Table 1).

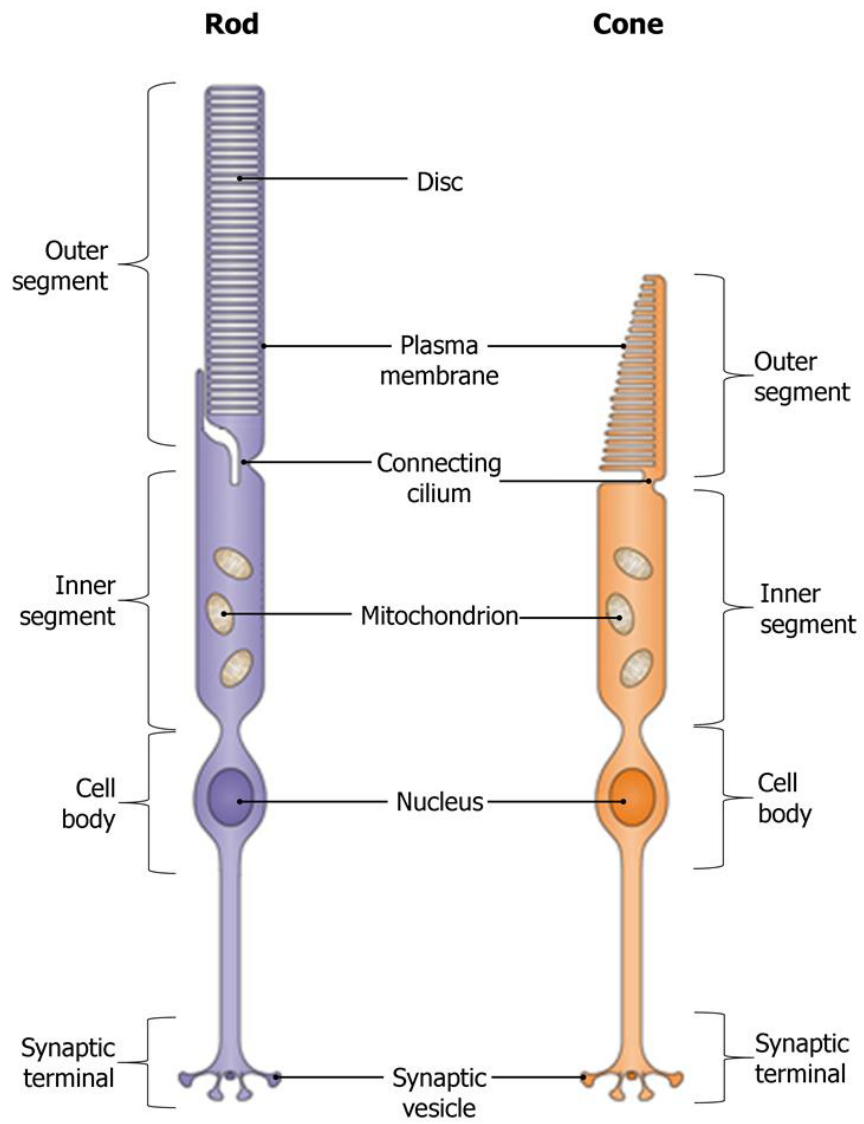


Figure 3. Photoreceptor cells

Schematic of the rod and cone photoreceptor cells. Image has been modified from Wright et al.⁴.

Table 1. Photopigments in humans

Photoreceptor type	Wavelength	λ_{\max}	Light colour
S-cones	Short	420nm	Blue
M-cones	Medium	535nm	Green
L-cones	Long	560nm	Red
Rods		498nm	

1.2.2.2 Connecting Cilium

Cilia are evolutionary conserved organelles that are essential for the development and maintenance of many cell types, tissues and organs. They are composed of a microtubule axoneme that stems from a basal body (Figure 4). The basal body is composed of nine microtubule triplets and the axoneme of nine microtubule doublets arranged in a circle. Based on their structure and ability to move cilia can be classified as two main types: motile cilia or non-motile/primary cilia. In general, motile cilia have a “9 + 2” configuration, with an additional pair of microtubules at the centre of the axoneme, while primary cilia lack the central microtubule pair and have a “9 + 0” configuration. The photoreceptor cell OS of the retina are essentially highly specialised primary cilia, other cell types and tissues with primary cilia include the kidneys, thyroid gland, liver, pancreas and endothelial cells.

In photoreceptor cells the axoneme forms the structural backbone of the OS and this together with the CC and basal body make up a highly specialised sensory cilium. The structural and functional components of the cilia, including proteins required for phototransduction, are synthesized in the IS cytoplasm, but are concentrated in the OS. The CC is effectively a bottleneck and regulatory site, as all movement of molecules between the IS and OS must proceed through this structurally complex and narrow space. The CC mediates the bidirectional transport between the IS and OS by means of intraflagellar transport (IFT). IFT is a specialised system responsible for the movement of cargo (IFT particles) into (anterograde transport) or out of (retrograde transport) the cilium using the molecular motors Kinesin-II and Dynein respectively (Figure 4). However, given the high metabolic rate of the photoreceptor cells, the number of molecules moving into the OS in light conditions would be extraordinary and probably saturate the motor-driven mechanism of IFT. It is therefore likely that movement between segments is accomplished by additional means, including simple diffusion of

soluble proteins and other facilitated mechanisms. For example, it is thought that rhodopsin trafficking involves myosin VIIa⁵.

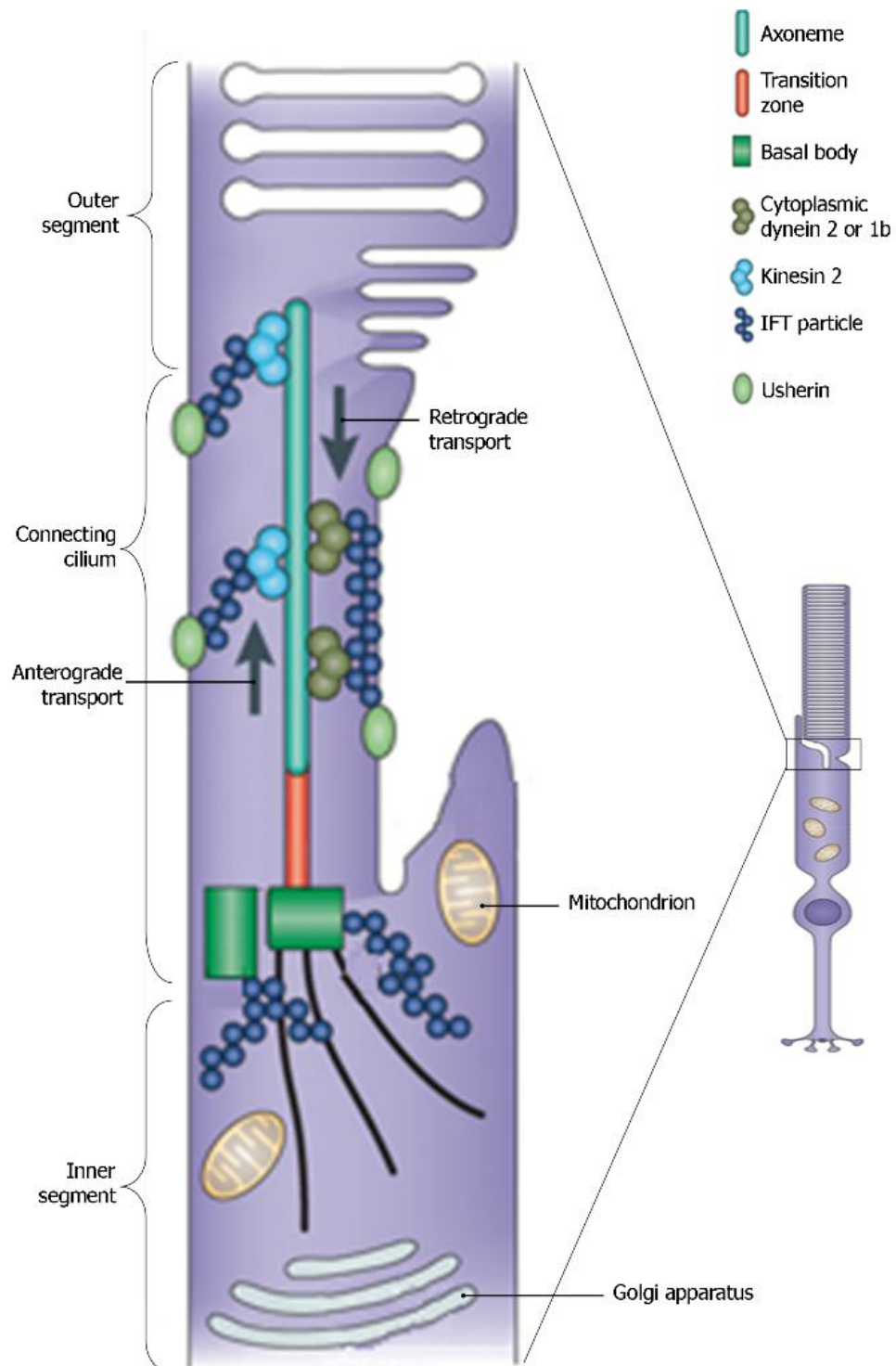


Figure 4. Connecting cilium and intraflagellar transport

The connecting cilium extends from the basal body into the OS of the photoreceptor. Transportation of particles between the IS and OS is indicated by arrows. IFT – intraflagellar transport. Image has been modified from Wright et al.⁴.

1.2.3 The Phototransduction Cascade

External information in the form of photons is captured by the photoreceptor cells, which convert it into a change in membrane resistance through the phototransduction cascade (Figure 5). Rod phototransduction is the most well characterised of all sensory transduction pathways. In the dark there is a steady current of cations, known as the “dark current”, through channels in the photoreceptor OS membrane which are kept open by the presence of cyclic guanosine monophosphate (cGMP) in the cytoplasm⁶. This results in depolarised photoreceptor cells that continuously release glutamate, a neurotransmitter, onto postsynaptic cells. Photon absorption by the photopigment (rhodopsin and 11-*cis*-retinal in rods) results in the isomerisation of 11-*cis*-retinal to all-*trans*-retinal, and this drives a conformational change in rhodopsin, activating the protein (RHO*). RHO* catalyses the activation of a G-protein, transducin, a process that includes the exchange of guanosine diphosphate (GDP) for guanosine triphosphate (GTP) with the release of the α -subunit of transducin⁶. The GTP-linked α -subunit of transducin drives the activation of cGMP phosphodiesterase (PDE), which in turn is responsible for the hydrolysis of cGMP to guanosine monophosphate (GMP). This reduction in cGMP concentration in the cytoplasm results in the closure of cGMP-gated channels in the plasma membrane and therefore the localised reduction in cation influx into the OS. This results in hyperpolarisation of the membrane and closure of synaptic glutamate-releasing calcium channels and therefore reduction or termination of glutamate release⁶.

In order to prevent further transducin activation RHO* is inactivated. Phosphorylation of RHO* allows binding of arrestin, preventing further interaction with transducin. Transducin and PDE are inactivated through the hydrolysis of GTP to GDP⁶. The falling calcium concentration in the OS results in the activation of membrane-bound retinal guanylate cyclase (GC), which in turn catalyses the regeneration of cGMP. The resulting increase in cGMP concentration results in the re-opening of cGMP-gated channels and the return of the photoreceptor to its depolarised state⁶.

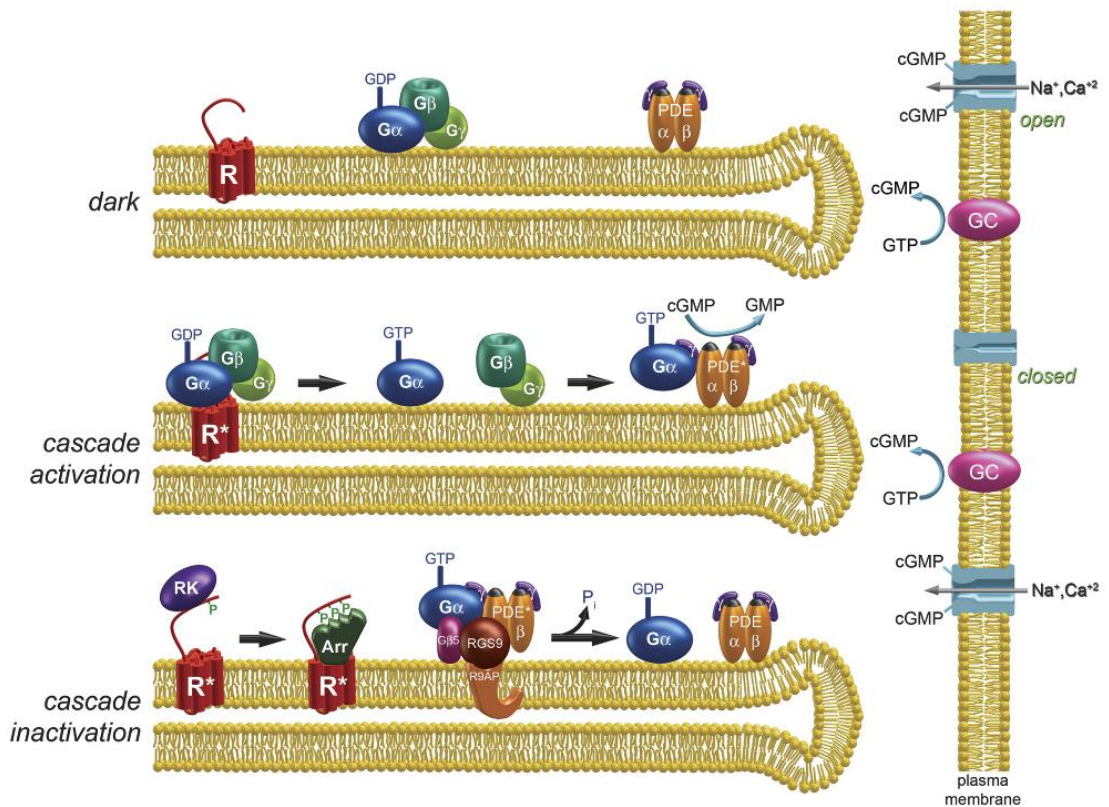


Figure 5. Phototransduction cascade

Summary of the phototransduction cascade resulting in the closure of cGMP-gated channels, followed by the inactivation of the cascade resulting in the re-opening of the channels. R – rhodopsin, R^* – activated rhodopsin, G – transducin, PDE – phosphodiesterase, cGMP – cyclic guanosine monophosphate, GC – retinal guanylate cyclase, RK – rhodopsin kinase, Arr – arrestin. Modified from Burns et al.⁶

1.2.4 The Visual Cycle

The visual cycle (Figure 6) is the process whereby the chromophore, 11-*cis*-retinal, is regenerated after photon absorption by rhodopsin. Upon activation rhodopsin releases all-*trans*-retinal, which is transported to the cytoplasm of the OS by ATP-binding cassette, subfamily A, member 4 (ABCA4). Here retinol dehydrogenase (RDH12) modifies it to all-*trans*-retinol (vitamin A). This is then transported by interstitial retinol-binding protein (IRBP) to the RPE where it undergoes a series of reactions. First, lecithin-retinol acyltransferase (LRAT) esterifies all-*trans*-retinol, followed by trans-isomerisation and further modification through the actions of RPE-specific isomerohydrolase (RPE65) and 11-*cis*-retinol dehydrogenase (RDH5) respectively to form 11-*cis*-retinal⁴. This is finally transported back to the photoreceptor cell OS.

It has been suggested that the processing of isomerised retinoid from cone cells occurs in Müller cells as well as in the RPE, although evidence is conflicting^{7,8}.

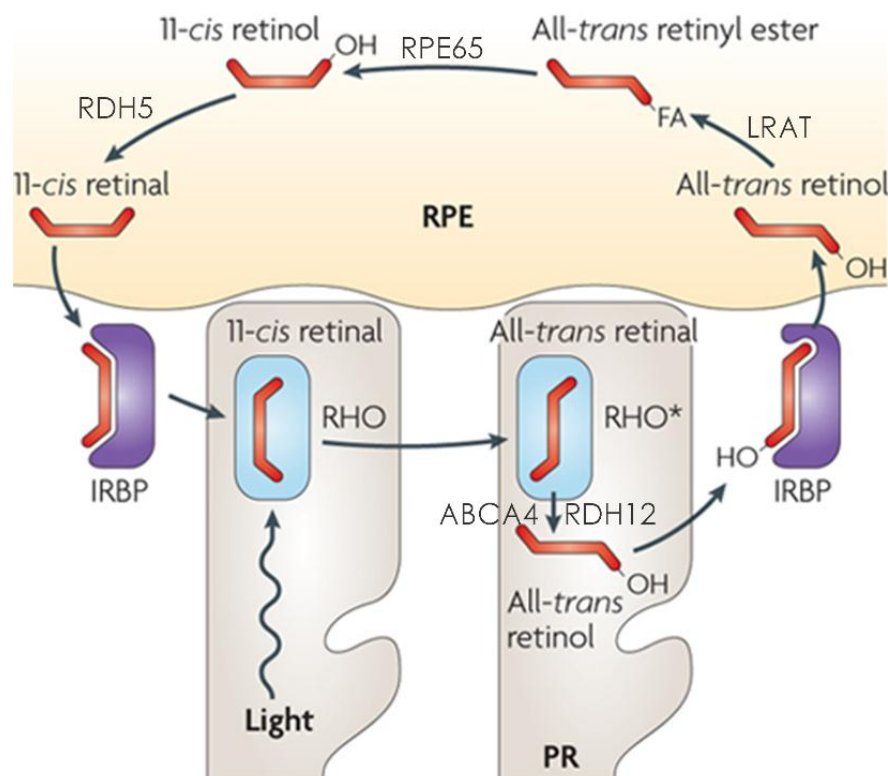


Figure 6. Visual cycle

Chromophore, 11-*cis*-retinal, is regenerated from spent all-*trans*-retinal through a number of reactions in the photoreceptor cells and RPE. PR – photoreceptor, RHO – rhodopsin, RHO* - activated rhodopsin, IRBP – interstitial retinol-binding protein, LRAT – lecithin retinol acyltransferase, RPE65 - Retinal pigment epithelium-specific 65 kDa protein, RDH5 – 11-*cis*-retinol dehydrogenase. Image has been modified from Wright et al.⁴.

1.2.5 The Canine Retina

As stated previously, the canine and human eye differs anatomically by only four major aspects, three of which involve the retina. Firstly, the canine eye has 9-20 additional layers of cells, adjacent to the RPE, called the tapetum lucidum and tapetum nigrum (Figure 1). The former is made up of reflective cells that line the top half of the retina. They receive light mostly from the darker ground and reflect light back through the retina, thereby improving scotopic vision. In lemurs (which also have these extra layers) it has been suggested that tapetal riboflavin absorbs shorter wavelengths of light (in the region of 260nm, 370nm and 450nm) and converts it to longer wavelengths (520nm) to which rhodopsin is more sensitive⁹. This may in effect brighten the appearance of a black or blue evening and also enhance the contrast between the background sky and objects in the environment. The tapetum nigrum is made up of darkly pigmented, non-reflective cells that line the bottom half of the retina. They receive light mostly from the brighter sky and may reduce the light scattering in the retina¹⁰.

Secondly, in the human eye the macula lutea is a region at the back of the eye that contains the highest concentration of pigments, which protect the retinal neurons against oxidative stress¹¹. At the centre of the macula is a single point made up entirely of densely packed cones and ganglion cells, called the fovea. It is this point that produces the sharpest image. Dog eyes do not have a macula or fovea, but similarly to other animals they do have an oval region located just above the optic nerve, termed the visual streak, enclosing a smaller circular region termed the area centralis. Similar to the fovea, the visual streak and area centralis is the region with the highest concentration of photoreceptor and ganglion cells, although only 10% are cone photoreceptor cells^{10,12}. The area centralis, found particularly in mammals with frontalised eyes, is specialised for binocular, high-resolution vision¹³. Grey wolves, from which domestic dogs are thought to have descended, have a pronounced visual streak that is thought to favour animals that live in open terrain and for which the horizon is an important feature of the field of view¹². Interestingly the distribution and density of ganglion cells varies between dog breeds and is correlated with skull shape. In general, traditional working breeds, such as those bred for hunting or herding and which would benefit from good peripheral vision, are mesocephalic. These dogs have retinas with a weak visual streak and more obvious area centralis or a pronounced visual streak and no obvious area centralis¹⁴. Conversely, many of the smaller breeds, such as Pugs and others bred almost exclusively as pets and which benefit from the ability to focus intently on a particular

object (such as a person), are brachycephalic. These dog breeds have a pronounced area centralis and virtually no visual streak¹⁴. The variation seen in dogs is most likely a result of morphological and functional trait selection and the lack of specific environmental pressures during domestication.

Finally, humans have three types of cone photoreceptors and therefore have trichromatic vision and can distinguish millions of colours. The canine retina on the other hand has only two types of cone photoreceptors, sensitive to light with wavelengths in the range of 429nm and 555nm, which appear violet and yellow-green respectively to humans. Dogs consequently have dichromatic colour vision, and light with a wavelength between 475-485nm, which appears green to humans, probably appears white or grey to dogs^{15,16}. Therefore, dogs probably see colours in a similar way to people with deuteranopia i.e. unable to discriminate between the colours red and green.

1.3 Canine Genetics

The domestic dog (*Canis lupus familiaris*) is an increasingly popular subject for the study of hereditary traits such as disease susceptibility, behavioural traits and morphological features due to its economic and social importance and unique history and population structure. The development of numerous genetic resources for the dog has resulted in considerable progress in the understanding of canine hereditary diseases and the development of DNA tests for monogenic traits. In addition the dog has recently become a prominent model species for the study of human disease.

1.3.1 History of Canine Breeds

The unique patterns of genetic variation between and within dog breeds can be attributed to their unique history and development through two major bottlenecks. The first was an ancient bottleneck that occurred when dogs were originally domesticated from the grey wolf approximately 15,000-100,000 years ago^{17,18}. The second occurred during more recent breed formation, approximately 200 years ago, when dogs were selectively bred for herding and hunting abilities, obedience and desired physical characteristics such as size, shape and coat colour. In the cases of some breeds there have been further bottlenecks associated with fluctuations in popularity, excessive use of show champion sires (the popular sire effect), geographic restrictions and catastrophic events such as world wars and economic depression. As a result there are approximately 400 distinct dog breeds worldwide. The development of breeding standards regulated by bodies such as the Kennel Club, and controlled breeding within these breeds has resulted in closed breeding populations with high levels of inbreeding and homogeneous genetic pools.

1.3.2 Canine Genetic Resources

The dog is an increasingly popular subject of scientific study for a number of reasons: 1) extensive medical surveillance and modern diagnosis and treatment options are second only to humans; 2) physical similarities with humans, such as body and organ sizes, physiology and phenotype as well as environmental exposure; 3) lifespans conducive to scientific study; 4) phenotypic homogeneity within breeds and 5) phenotypic heterogeneity between breeds. It has only been in recent years that genetic resources available for the dog have enabled extensive genetic study of the species. Linkage maps^{19,20}, a high quality radiation hybrid map²¹, and the 1.5x Poodle sequence²² were once valuable tools, but have recently been superseded by more sophisticated tools. The first high-quality (7.5x) draft sequence of the dog (a Boxer) was

made available in 2004 (CanFam1.0), and the updated assembly (CanFam2.0) in 2006²³. The assembly was updated again in 2012 (CanFam3.1), albeit only after much of this PhD project had been completed. The publication of the dog genome sequence has led to the development of a single nucleotide polymorphism (SNP) map containing >2.5 million SNPs²³, medium and high density whole genome SNP marker arrays and messenger RNA (mRNA) sequencing (RNA-seq) data from a variety of canine tissues, all accessible with the Ensembl genome browser (www.ensembl.org/Canis_familiaris).

1.3.3 The Canine Genome

The canine genome is approximately 2.4 gigabases (Gb) in size with approximately 19,000 genes²³. In a genetic study of 85 breeds it was found that dogs have similar levels of overall nucleotide diversity – 8×10^{-4} nucleotide substitutions per base/pair between random pairs of genomes – to humans. However, the variation between dog breeds is 27.5% compared with 5.4% between human populations, while the homogeneity of 94.6% within individual dog breeds is greater than that of humans (72.5%)²⁴. This contributes to dogs having unique patterns of linkage disequilibrium (LD), extending over several megabases (Mb) within breeds (10-100 fold further than in humans), but only tens of kilobases (kb) between breeds²³. LD patterns vary enormously from breed to breed, with the general trend of greater LD seen in numerically smaller breeds that have experienced significant population bottlenecks, while reduced LD is seen in numerically larger breeds such as Golden Retrievers²⁵.

1.4 Canine Hereditary Disease

The stringent breeding of dogs within closed breeding populations and population bottlenecks has resulted in elevated levels of inbreeding. This in turn has led to the emergence of hereditary diseases specific to or at higher prevalence in certain breeds, including heart disease, blindness, cancers and epilepsy²⁶. Most of these diseases occur in human populations, and clinical manifestations are often similar between the two species²⁷. Some monogenic conditions are shared by multiple breeds, but most are confined to very few breeds, while complex conditions tend to affect multiple breeds. To date more than 600 genetic traits, mostly disease traits, and 280 breeds are listed in the Inherited Diseases In Dogs (IDID) database²⁷ with a total of 1,552 entries when considering traits found in multiple breeds (<http://www.vet.cam.ac.uk/idid>, accessed on 13/11/2012). Similarly there are 583 entries for the dog in the Online Mendelian Inheritance in Animals (OMIA) database (<http://omia.angis.org.au/home/>, accessed on 13/11/2012), representing the largest number of known naturally occurring genetic diseases in any non-human species. Canine genetic disorders affect a wide variety of physiological systems including blood, cardiovascular, endocrine, ocular, gastrointestinal, immune, musculoskeletal, nervous system, respiratory, urinary and reproductive disorders as listed in the Canine Inherited Disorders Database (CIDD) (<http://www.upei.ca/~cidd>, accessed 21/06/2011). Of these inherited diseases it is thought that more than half closely resemble human diseases in terms of laboratory and clinical abnormalities²⁶.

1.4.1 Canine Retinopathies

The canine retinal dystrophies, a common cause of blindness in purebred dogs, are fairly well characterised, albeit not as well as in humans, and often based only on ophthalmoscopic findings. However, our understanding of the genetic and phenotypic characteristics of each condition, and their correlation, is continuously improving and as such their characterisation and classification is ever-changing. Causal mutations have been found for some forms of retinal diseases, but many remain undefined and novel reports of the condition being observed in breeds for the first time are commonplace.

Various methods are employed for the classification of canine retinopathies, one of which is based on three main factors (Figure 7 and Table 2):

Developmental versus degenerative – abnormal development of the retina results in developmental disease, while degeneration of a normally developed retina results in degenerative disease.

Stationary versus progressive – retinopathies in which the clinical effects remain relatively constant are described as stationary, whereas those in which the clinical effects become more severe over time are progressive.

Rod-led versus cone-led – in progressive retinal atrophy (PRA) the rod photoreceptor cells are affected first while in the cone rod dystrophy (CRD) the cone photoreceptor cells are affected first.

PRA and CRD are two groups of ocular diseases in dogs and other animals characterised by progressive retinal degeneration resulting in loss of vision. In typical PRA rod photoreceptor responses are lost first followed by cone photoreceptor responses. Therefore dogs experience nyctalopia (night blindness) followed by deterioration of daytime vision, culminating in complete blindness²⁸. In CRD cone responses are typically lost first and rod responses are affected later or to a lesser extent. Usually the ophthalmoscopic findings of PRA and CRD are similar, especially late in the progression of the disease²⁹. The only method of distinguishing between cone- or rod-led degeneration is by electroretinogram (ERG) testing in the early stages of disease. Because an ERG is not frequently performed on a canine patient as part of a routine consultation due to the expense and the necessary sedation in patients, ophthalmoscopically similar conditions such as CRD are often misdiagnosed as PRA.

Numerous forms of PRA and CRD have been documented in more than 100 dog breeds and whilst they exhibit similar clinical signs, the aetiology, age at onset and rate of progression vary between and within breeds. Typically the later onset forms of the disease progress slower than the early onset forms. Most forms have an autosomal recessive (aR) mode of inheritance^{30,31}, although an autosomal dominant (aD)³² and two X-linked³³ forms have been reported.

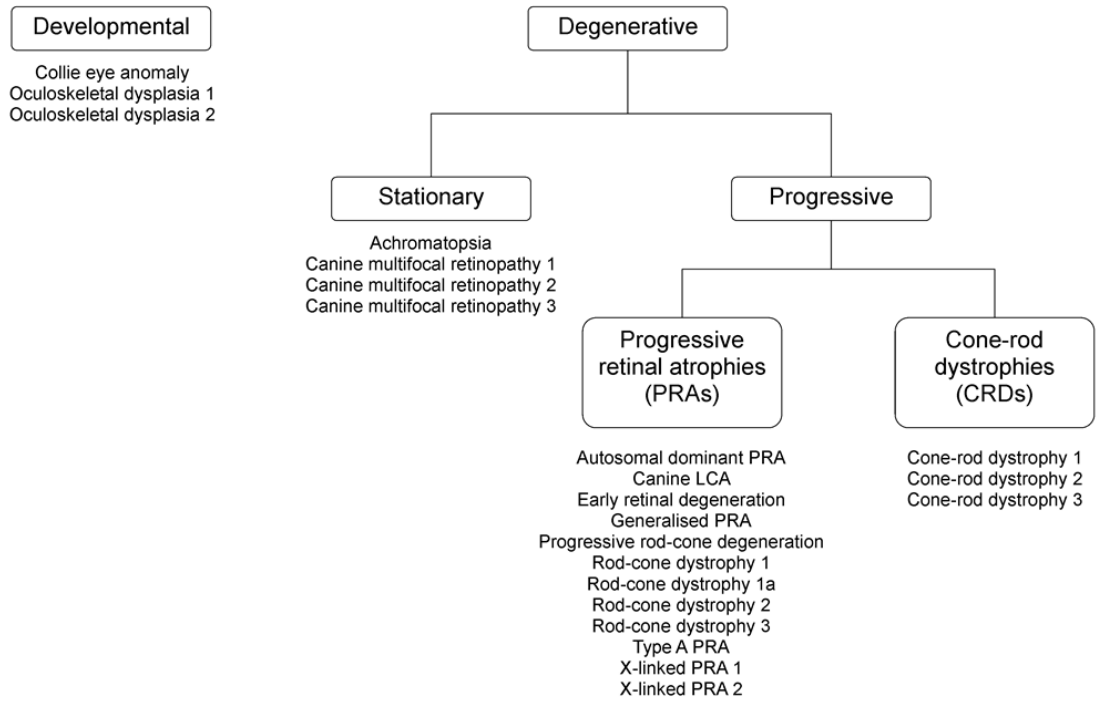


Figure 7. Classification of the canine retinopathies

Each of the canine retinopathies for which a genetic mutation has been identified is shown. Figure has been modified from Mellersh³⁴.

Table 2. Classification and known mutations of canine retinopathies

Gene	Form of disease	Mutation	CanFam2 Coordinates*	Mutation Type	Breeds	Ref.	Orthologous Human Disease	Gene OMIM no.†
Developmental retinal disorders								
<i>NHEJ1</i>	Collie eye anomaly (CEA)	7.8kb deletion	CFA37:28697542-28705340	Binding domain deletion	8 breeds‡	35	SCID (Severe combined immunodeficiency)	611290
<i>COL9A3</i>	Oculoskeletal dysplasia 1 (osd1/drd1)	1 bp insertion (A)	CFA24:49699847-49699850	Frame-shift	Labrador Retriever	36	Multiple Epiphyseal dysplasia	120270
<i>COL9A2</i>	Oculoskeletal dysplasia 2 (osd2/drd2)	1,267 bp deletion	CFA15:5649586-Unknown	Exon loss	Samoyed	36	Multiple Epiphyseal dysplasia Stickler syndrome, type V	120260
Stationary retinal disorders								
<i>CNGB3</i>	Achromatopsia	140kb deletion	Unknown	Gene loss	Alaskan Malamute	37	Achromatopsia-3 Juvenile macular degeneration	605080
<i>CNGB3</i>	Achromatopsia	G->A	CFA29:35840555	Missense	German Shorthaired Pointer	37		
<i>BEST1</i>	Canine multifocal retinopathy (cmr1)	C->T	CFA18:57508040	Nonsense	3 breeds§	38	Retinitis pigmentosa 50 Retinitis pigmentosa, concentric Vitelliform macular dystrophy Best macular dystrophy Bestrophinopathy Microcornea, rod-cone dystrophy, cataract and posterior staphyloma Vitreoretinchoroidopathy Maculopathy, bull's-eye	607854
<i>BEST1</i>	Canine multifocal retinopathy (cmr2)	G->A	CFA18:57505416	Missense	Coton de Tulear	38		
<i>BEST1</i>	Canine multifocal retinopathy (cmr3)	1 bp deletion (C) & G->T	CFA18:57499959 & 57499881	Frame-shift	Lapponian Herder	39		

Gene	Form of disease	Mutation	CanFam2 Coordinates *	Mutation Type	Breeds	Ref.	Orthologous Human Disease	Gene OMIM no.†
Progressive retinal disorders: Cone-rod dystrophies (CRDs)								
<i>RPGRI1</i>	Cone-rod dystrophy 1 (cord1)	44bp insertion (A ₂₉ GGAAGCA ACAGGATG)	CFA15:21338814-21338815	Frame-shift	4 breeds**	30	Cone-rod dystrophy 13 Leber congenital amaurosis 6	605446
<i>NPHP4</i>	Cone-rod dystrophy	180bp deletion	CFA5:62913591-62913770	Exon skipping	Standard Wire-haired Dachshund	40	Nephronophthisis 4 Senior-Loken syndrome 4	607215
<i>ADAM9</i>	Cone-rod dystrophy 3	24.7kb deletion	CFA16:29.400-29.423Mb	Exon skipping	Glen of Imaal Terrier	41,42	Cone-rod dystrophy 9	602713
Progressive retinal disorders: Progressive retinal atrophies (PRAs)								
<i>RPE65</i>	Canine LCA (cLCA)	4bp deletion (AAGA)	CFA6:79959789-79959792	Frame-shift	Briard	43	Leber Congenital Amaurosis 2 Retinitis pigmentosa 20	180069
<i>RHO</i>	Autosomal dominant PRA	C->G	CFA20:8641456	Missense	English Mastiff; Bull Mastiff	32,44	Night blindness, congenital stationery, autosomal dominant 1 Retinitis pigmentosa 4, autosomal dominant or recessive Retinitis punctata albescens	180380
<i>STK38L</i>	Early retinal degeneration (erd)	SINE insertion	CFA27:23459892-23459893	Exon skipping	Norwegian Elkhound	45		
<i>CCDC66</i>	Generalised PRA	1bp insertion (A)	CFA20:36742856-36742857	Frame-shift	Schapendoes	46		
<i>PRCD</i>	Progressive rod-cone degeneration (prcd)	G->A	CFA9:7186710	Missense	29 breeds††	47	Retinitis pigmentosa 36	610598
<i>PDE6A</i>	Rod-cone dystrophy 3 (rcd3)	1bp deletion (A)	CFA4:62348897	Frame-shift	Cardigan Welsh Corgi	31	Retinitis pigmentosa 43	180071

Gene	Form of disease	Mutation	CanFam2 Coordinates *	Mutation Type	Breeds	Ref.	Orthologous Human Disease	Gene OMIM no. †
<i>PDE6B</i>	Rod-cone dystrophy 1 (rcd1)	G->A	CFA3:94574275	Missense	Irish setter	48	Night blindness, congenital stationary, autosomal dominant 2 Retinitis pigmentosa 40	180072
<i>PDE6B</i>	Rod-cone dystrophy 1 (rcd1a)	8bp insertion (TGAAGTCC)	CFA3:94574247-94574248	Frame-shift	Sloughi	49		
<i>RD3</i>	Rod-cone dystrophy 2 (rcd2)	22bp insertion (gccccccccgccccgcccc)	CFA7:12832684-12834235	Frame-shift	Rough and smooth collies	50	Leber congenital amaurosis 12	180040
<i>PDC</i>	Type-A PRA	C->G	CFA7:22463847	Missense	Miniature Schnauzer	51		171490
<i>RPGR</i>	X-linked PRA 1	5bp deletion (GAGAA)	CFAX:Unknown**	Frame-shift	Samoyed and Siberian Husky	52	Cone-rod dystrophy-1 Macular degeneration, X-linked atrophic Retinitis pigmentosa 3 Retinitis pigmentosa, X-linked, and sinorespiratory infections, with or without deafness	312610
<i>RPGR</i>	X-linked PRA 2	2bp deletion (GA)	CFAX:Unknown**	Frame-shift	Mongrel (Mixed breed)	52		

* Unknown = precise coordinates or boundaries unknown due to missing from CanFam2 assembly or insufficient information in literature.

† Online Mendelian Inheritance in Man database (<http://www.ncbi.nlm.nih.gov/omim>)

‡ Australian Shepherd, Border Collie, Lancashire Heeler, Nova Scotia Duck Tolling Retriever, Rough Collie, Shetland Sheepdog, Smooth Collie, Long haired Whippet

§ Great Pyrenees, English Mastiff, Bullmastiff

** Miniature Long, Short and Wire Haired Dachshunds, English Springer Spaniel

†† American Cocker Spaniel, American Eskimo Dog, Australian Cattle Dog, Australian Shepherd, Australian Shepherd (Miniature), Australian Stumpy Tail Cattle Dog, Chesapeake Bay Retriever, Chinese Crested, Cockapoos, Dwarf Poodle, English Cocker Spaniel, Entelbacher Mountain Dog, Finnish Lapphund, Golden Retriever, Golden Doodle, Karelian Bear Dog, Kuvasz, Labradoodle, Labrador Retriever, Laponian Herder, Markiesje, Moyen Poodle, Norwegian Elkhound, Nova Scotia Duck Tolling Retriever, Poodle Miniature and Toy, Portuguese Water Dog, Spanish Water Dog, Swedish Lapphund, Yorkshire Terrier.

‡‡ Exact coordinates unknown, but somewhere between 33,026,823 and 33,027,678 bp.

1.4.2 Progressive Retinal Atrophy

1.4.2.1 Ophthalmoscopic and Clinical Signs of PRA

Disease progression is bilateral and usually symmetrical. Initially the tapetal periphery takes on a granular appearance, and with loss of thickness of the retina the tapetum becomes increasingly reflective (known as tapetal hyper-reflectivity). Some dogs do not have a tapetum, in which case the identification of retinal thinning is difficult. Further disease progression, however, results in depigmentation and pigment clumping of the non-tapetal fundus⁵³. Another typical sign of PRA is attenuation of the blood vessels. However caution must be taken as smaller dogs tend to have small vessels and can thus be misdiagnosed on the basis of their retinal vasculature. Finally, in advanced PRA, the loss of ganglion cell axons that form the optic nerve results in atrophy of the optic nerve head, giving it a flatter, grey appearance²⁸.

1.4.2.2 Diagnosis of PRA

Accurate diagnosis requires ophthalmoscopic examination at the very least and preferably also an ERG. Retinal abnormalities can be detected by ERG before ophthalmoscopic signs appear or if the view of the retina is impeded by secondary cataracts which often accompany PRA⁵⁴. An ERG is a measure of the response of the retina to a light stimulus. The retina is stimulated by a flash of light and the electrical response at the corneal surface is recorded. A detailed ERG can provide information about the retinal dysfunction including the order in which rods and cones are affected, if both are affected, and the age of onset of retinal dysfunction. ERGs tend to be most informative when carried out very early in the development of PRA, often prior to the onset of ophthalmoscopic signs²⁸.

The BVA/KC/ISDS (British Veterinary Association/Kennel Club/International Sheep Dog Society) Eye Scheme in the UK and other similar international schemes offer dog breeders the opportunity to have their dogs screened by veterinary ophthalmologists for the presence of hereditary eye disease, including PRA. The clinical data pertaining to dogs obtained through such schemes is well-documented, and as it is collected by clinical specialists, robust, albeit usually limited to ophthalmoscopic signs.

Other methods employed to assist diagnosis, usually only obtained for the purposes of research and requiring dogs bred specifically for the purpose of research, include histopathology and biochemistry. The former can identify the cell type involved in disease and also determine whether the cells involved develop normally before

degeneration, or never reach maturity at all²⁸. Biochemical assays can potentially identify any abnormalities in the cell biochemistry. For example, with certain rod-cone dystrophies elevated levels of cGMP is seen within the process of rod phototransduction⁵⁵.

The presence of phenocopies, i.e. non-hereditary retinal degenerations that mimic the effects of hereditary retinal dystrophies, complicates the accurate diagnosis of PRA. In sudden acquired retinal degeneration (SARD), as reviewed by Petersen-Jones, ophthalmoscopic signs of retinal degeneration only appear many weeks after a sudden loss of visual function, and these dogs can therefore be identified based on their history²⁸. However, it is prudent to bear in mind the possibility of misdiagnosis of PRA during any molecular or clinical investigations.

1.5 Human Retinal Degenerations

The degeneration of photoreceptor cells and/or the RPE results in retinal dystrophies, which are the major cause of familial blindness in the western world, and are incurable. Similarly to dogs, human retinal dystrophies can be characterised by which cells are affected and in which order they degenerate during disease progression e.g. cone, rod-cone etc.. Despite these classifications there is still a great deal of genetic and clinical overlap of the retinal dystrophies. In addition, retinal dystrophies present along with other symptoms in a number of syndromic conditions. The past two decades has seen an explosion in the numbers of retinal disease loci mapped and mutations found. According to the Retinal Information System (RetNet; <http://www.sph.uth.tmc.edu/Retnet/>; accessed 13/11/2012), 185 of 225 genes causing retinal disease have been identified. Incredibly, approximately one third of these genes localise to the CC or basal body of photoreceptor cells⁵⁶, and diseases caused by mutated ciliary genes are known as ciliopathies. Canine PRA has not been described as part of a syndrome in any of the breeds under investigation, and I will therefore be focusing mainly on the non-syndromic retinal dystrophies for the purposes of this project. Of 153 genes causing non-syndromic retinal disease in humans, 122 have been identified. A number of these genes have also been implicated in canine retinal disease (Table 2). Despite this success, a large number of mutations have yet to be identified in humans. It is estimated that the genes identified thus far account for approximately 65% of inherited autosomal recessive retinal dystrophies⁵⁷.

1.5.1 Retinitis Pigmentosa

PRA is considered the veterinary equivalent of **Retinitis pigmentosa (RP)** in humans and has already proved to be a valuable model for various forms of human retinal degenerations. RP is the name given to a group of common inherited retinal diseases in humans which affects 1 in 3500 to 4500 people⁵⁸. Photoreceptor cells predominantly affected are the rods and therefore clinical symptoms typically include night blindness (nyctalopia) and loss of peripheral vision. With disease progression the cones also degenerate resulting in central vision loss and eventually possibly complete blindness⁵⁹. While this heterogeneous group of diseases is highly variable with regard to age of onset, retinal appearance, progression and visual outcome, there are hallmark characteristics secondary to photoreceptor degeneration. These include “bone spicules” caused by pigment granule migration from the RPE and retinal arteriole and vein attenuation⁵⁹.

1.5.2 Other Non-syndromic Retinal Dystrophies

While RP is the most common of the hereditary retinal degenerations, there are a number of other diseases that exhibit degeneration of the retina:

Leber congenital amaurosis (LCA) is clinically similar to RP, but more severe. Patients are impaired visually from birth and present with nystagmus (involuntary eye movements). Pigmentation and arteriole attenuation can vary from mild to severe, and an electroretinogram is usually non-detectable within the first year of life. LCA is relatively rare, with a worldwide incidence of approximately 1 in 30,000 and most cases show aR inheritance⁶⁰.

Startgardt Disease Type 1 (STGD1), with a worldwide incidence of 1 in 10,000, is the most common juvenile macular dystrophy. Patients typically experience loss of visual acuity in early childhood or early adolescence. Progressive retinal and macular changes result from lipofuscin accumulation⁶¹.

Cone-rod dystrophy (CRD) is a disorder predominantly characterised by a loss of cone function, with patients experiencing photoaversion (intolerance of light), and defective colour vision in adolescence or early adult life. Mild thinning of the retinal vessels is observed, but bone spicules are usually absent⁶².

Achromatopsia (ACHM) is a stationary, congenital aR cone disorder, where diminished cone ERG responses are observed, but rod ERG responses are normal.

Patients generally experience low visual acuity, photophobia (pain caused by light exposure), nystagmus and severe defects in colour vision⁶³.

Cone dystrophy (CD) is a progressive condition in which cone function is initially normal. However patients experience progressive loss of visual acuity, colour vision disruption and increasing photophobia. Cone ERG responses begin to diminish within the first two decades of life and by the fourth decade of life visual acuity worsens to the point of legal blindness⁶³.

1.5.3 Syndromic Disorders: Bardet-Biedl Syndrome

A number of syndromes (Senior-Løken syndrome, Alström syndrome, Joubert syndrome, Bardet-Biedl syndrome and Usher syndrome), all of which are ciliopathies, have been described in which a broad range of clinical manifestations are observed, including retinal degeneration. **Bardet-Biedl Syndrome (BBS)** is probably the best characterised of the syndromes. Patients display some or all of the features common to ciliopathies including, but not limited to, RP, cystic kidneys, polydactyly, obesity, *situs inversus* (left-right reversal of organ distribution), gonadal malformations/hypogenitalism, heart disease, liver dysfunction and mental retardation. Seventeen genes to date have been implicated in the disease (RetNet; <http://www.sph.uth.tmc.edu/Retnet/>; accessed 16/11/2012 and Online Mendelian Inheritance in Man (OMIM) database; OMIM number 209900; <http://www.ncbi.nlm.nih.gov/omim>; accessed 16/11/2012). Most forms of BBS have an aR mode of inheritance, although triallelic inheritance has also been reported^{64,65}.

Clinical symptoms tend to progress in a similar way in most BBS patients: Postaxial polydactyly and hypogenitalism are present at birth. Within a few years children begin to gain weight and eventually become obese. By eight years of age visual problems begin with night-blindness, which progresses to complete blindness by the age of 15-20 years. During childhood patients also develop kidney cysts. Many patients eventually develop chronic kidney disease and end-stage renal failure is the most common cause of premature death.

1.6 Canine Models for Human Genetic Disease

The study of genetic disease in model organisms is potentially highly beneficial to humans and the identification of certain disease genes in dogs has already facilitated progress in studies of equivalent human disease. For example, autosomal recessive congenital ichthyosis (ARCI) in humans is associated with mutations in various genes that affect skin barrier function. However, in the canine equivalent (a lamellar ichthyosis in Golden Retrievers) mutations were identified in the patatin-like phospholipase domain containing 1 (*PNPLA1*) gene. This discovery led scientists to the human homologue, in which mutations were identified in six affected individuals from two families⁶⁶. Human populations with limited genetic variation, such as the Bedouins, Finns and Icelanders, have been successfully used in disease-mapping studies^{67,68}. However, most human populations have a great deal of genetic variation and as a result mapping of most disease traits that segregate in them is far more difficult. Canine breeds, with their limited genetic variation, are analogous to the former, inbred human populations. In addition, the development of canine genetic resources and the disease similarities of humans and dogs suggest that genes involved in disease may be more quickly and easily found in dogs than in humans. The dog is therefore an ideal species in which to study complex and simple genetic diseases. The prevalence and heterogeneity of canine retinal degenerations has resulted in the identification of numerous canine models for human retinal degenerations (Table 2). Due to the recent and continuing development of sophisticated resources for the dog, the rate at which novel disease- and trait-associated genes are identified is expected to increase.

1.6.1 Canine Models for Retinal Gene Therapy Studies

Animal models for human disease have also proved useful in gene therapy studies. The mouse is the most commonly used model for human retinal degenerations with spontaneous or man-made models available for many of the genes implicated in nonsyndromic retinal dystrophies. However, a disadvantage of the mouse as a model is that the size and anatomy of the mouse eye differs from that of the human, and as a result gene delivery requires different surgical techniques. In addition, due to the short lifespan of the mouse it has limited relevance to human disease.

Large animal models such as dogs are increasingly being used for gene therapy studies. Anatomically, the eyes of dogs are far more similar to humans than those of mice. While only primates have maculas, dogs do have the ganglion-enriched areas known as a visual streak or area centralis, which are functionally similar to the human macula. In

addition, the size of the eye in dogs is similar to humans, allowing the use of the same surgical approaches for gene delivery.

1.6.1.1 LCA and *RPE65*

Canine leber congenital amaurosis (cLCA) in Briards, previously known as congenital stationary night blindness, is an autosomal recessive retinopathy caused by dysfunctional rods and cones. The fundus appears ophthalmoscopically normal for up to three years before subtle clinical signs appear and photoreceptor cells only begin to change late in the progression of the disease⁶⁹. *RPE65* plays an important role in the visual cycle (Figure 6) and a mutation in the gene that is associated with cLCA has been described⁴³. Aberrant *RPE65* also accounts for approximately 6% of LCA in humans (*LCA2*)⁶⁰. The phenotype is similar to that of cLCA in that, initially, retinal architecture is normal and patients have useful visual acuity. Vision becomes severely impaired by the third decade due to progressive deterioration in photoreceptor morphology and visual function. A mouse model for *LCA2* is available, but is a sub-optimal system for development of therapies for humans due to phenotypic differences between the two species: the knockout mice retain cone function, whereas human patients exhibit rod and cone dysfunction⁷⁰.

Canine LCA is proving to be an extraordinarily valuable model for gene therapy studies involving *LCA2* in humans⁷¹. Retinal degeneration in affected dogs is slow and as a result the window of opportunity for administering treatment is relatively large. In addition, as the mutation results in a lack of gene product an appropriate treatment is simply to administer a functional product into the subretinal space, carried by a vector. *RPE65*^{-/-} dogs have been used to test the efficacy of gene therapy using an adeno-associated virus (AAV) vector carrying wildtype *RPE65*, and visual function was restored and improvements were maintained for more than three years^{72,73}. As a result, to assess the safety and efficacy of similar therapy in humans, three phase I/II clinical trials have been undertaken in *LCA2* patients in which no serious adverse effects were reported. In addition, improvements in rod-mediated visual function were also reported^{2,71,74-76}.

1.6.1.2 X-linked Retinal Dystrophy and *RPGR*

The two forms of X-linked PRA are both caused by mutations in the retinitis pigmentosa GTPase regulator (*RPGR*) gene, specifically in open reading frame 15 (ORF15), but they are clinically distinct. Type 1, found in the closely related Siberian Husky and Samoyed breeds, manifests after normal photoreceptor morphogenesis.

Affected dogs display electroretinographic abnormalities from six months of age and retinal degeneration is led by the rods. By four years of age the blood vessels are severely attenuated and the tapetum displays advanced hyper-reflectivity. Type 2, found in a mixed breed, is more severe and manifests during retinal development. Affected dogs display electroretinographic abnormalities from six weeks of age and advanced PRA by two years of age⁵². *RPGR* is a component of the connecting cilium and is thought to be involved in microtubule organisation and intracellular trafficking⁷⁷. Aberrant *RPGR* is the most common cause of retinal dystrophy in humans and more than 240 mutations have been identified in the gene, as reviewed by Shu et al.⁷⁸. Of these, 95% cause X-linked RP (XLRP), 3% X-linked cone or cone-rod dystrophy (XLCRD) and 2% cause atrophic macular dystrophy. A mutational hotspot, ORF15, contains more than half of the reported mutations^{78,79}. XLRP-affected males typically develop nyctalopia by the age of 10 years and are legally blind by the third decade of life. They have generalised rod and cone dysfunction, but due to normal visual acuity during childhood the window of opportunity for gene therapy is longer than LCA⁸⁰. An *Rpgr* knockout mouse has slow retinal degeneration, and cone and rod photoreceptor abnormalities, but the small amount of residual protein that remains suggests some residual function^{77,81}. Mutations that result in retention of partial function in humans are associated with XLCRD⁸². The knockout mouse therefore represents a better model for human XLCRD than XLRP. A second, transgenic, mouse model expressing truncated *RPGR*^{ORF15} had a more severe gain-of-function phenotype⁸³.

Recently Beltran et al. described the use of gene augmentation therapy in the canine XLGRA1 and XLGRA2 models. *RPGR*^{-/-} dogs given subretinal injections of AAV vectored with human *RPGR* proved the efficacy of the treatment with improvements seen in treated eyes only compared with untreated eyes: inner and outer segments were preserved, rod and cone photoreceptor function was greater, photoreceptor structure was normal and opsin mislocalisation was reversed⁸⁴. While this suggests that gene augmentation therapy is a viable option for the treatment of *RPGR*-associated retinopathies, the broad range of phenotypes observed and the effect of modifiers on *RPGR* expression may be a potential obstacle in selecting suitable recipients^{85,86}.

1.7 Identifying Variants Associated With Canine Genetic Traits

A number of approaches have been used to successfully map and identify canine genetic mutations, including a candidate gene approach, family-based linkage analyses and case-control association studies. The two latter approaches essentially compare the segregation of genetic markers among affected and non-affected individuals to identify loci potentially containing the mutation. The long LD within breeds makes it possible to map simple traits efficiently using few samples and a relatively sparse marker set compared with human studies²³.

1.7.1 Genetic Markers

Microsatellite markers are short repetitive DNA units, usually 2-4 base pairs, scattered throughout the genome. The number of repeat units is highly variable and each marker could have multiple alleles, making microsatellites very useful genetic markers. Minimal Screening Set 2⁸⁷ and Minimal Screening Set 3⁸⁸, containing 325 and 507 microsatellite markers respectively, were developed for use in mapping studies, but have become obsolete. Markers within these sets are separated by an average physical spacing of 9 Mb and 5 Mb respectively. As a result, both sets lack the density of markers required when studying traits in the majority of breeds as LD seldom extends >3 Mb⁸⁹.

SNP markers are defined as the change at a specific genomic position from one nucleotide to another. SNPs are usually bi-allelic and therefore less informative than microsatellite markers, but genotyping is easier to automate as a result. Despite this they have a far higher density across the canine genome with an average spacing of approximately 1 kb between SNPs, making them very useful for whole genome mapping approaches²³.

1.7.2 Candidate Gene Approach

Candidate genes are genes that have been identified based on their function, expression patterns and/or association with similar diseases or traits in other species. In the case of ocular disease the number of candidate genes could be extremely high and screening all of them for variants would be labour-intensive, time-consuming and costly. Therefore polymorphic markers within close proximity to the candidate genes are identified and co-segregation of marker alleles with the disease examined and confirmed prior to undertaking more thorough investigations. Mutations are identified by searching for

differences in the genes between affected and unaffected dogs. Prior to the development of dense SNP maps, this strategy was frequently used as an easy way to interrogate the whole genome. However, this approach has proved to be inefficient with only 3.4% of published studies reporting probable retinal disease-causing mutations in dogs⁹⁰, and as a result is seldom used nowadays.

1.7.3 Linkage Analysis

Family-based linkage studies examine the transmission of marker alleles through multiple generations of families in which a genetic disease or trait segregates. Previously microsatellite markers were most often used for linkage studies, but SNP markers can also be used. Using statistical methods, marker alleles that segregate with the trait more often than is expected at random are identified⁹⁰. To reduce and define the size of the trait-linked locus identified, additional markers are often genotyped, followed by the analysis of candidate genes within the defined locus. A number of disease loci have been mapped by linkage analysis including canine progressive rod cone degeneration, and the mutation has since been identified^{47,91,92}. A disadvantage of linkage studies is the requirement of complete data and samples from large families. Apart from the difficulty of obtaining sufficient samples and complete datasets, limited recombination within families often results in the identification of linked regions too large to follow up methodically and economically.

1.7.4 Case-Control Association Analysis

Association studies are similar to linkage studies in that the segregation of marker alleles with a genetic trait are compared in affected (case) and unaffected (control) cohorts, and statistical methods are used to ascertain whether this occurs more often than is expected if a random occurrence. The main difference from linkage studies is that unrelated dogs are used, making sample collection far easier. However, a much denser set of markers is required. Due to increased recombination within a population compared with within families, associated loci tend to be relatively small.

Genetic association studies are commonly used to identify loci associated with certain traits or disease in many species. The difficulty in human studies is that, with the exception of relatively few closed breeding populations, the average LD in human populations, while highly variable, tends to be short. As a result at least 500,000 SNP markers⁹³ and usually hundreds or even thousands of samples are required to conduct successful association studies. Currently arrays containing approximately one million SNPs are used routinely and arrays containing more than four million SNPs are

available. However, the problem of adequate sample number remains the limiting factor for the detection of alleles without a very large effect on a multifactorial trait, or very rare alleles involved in a simple mendelian trait. For example, in order to detect an allele with a twofold increased risk (odds ratio = 2) with 90% power, 1,000 samples are required if the allele frequency is 0.3 and 6,000 if the frequency is 0.03⁹⁴. In the case of heterogeneous diseases such as RP where it is difficult, if not impossible, to distinguish between diseases caused by different mutations at the ophthalmoscopic or clinical level and many of which are thought to be sporadic cases, association studies are not a feasible option for identifying mutations. In contrast, while PRA in dogs is heterogeneous between breeds, it is largely homogeneous within individual breeds. For the study of PRA in dogs therefore, genome wide association (GWA) mapping is particularly well suited.

Since the assembly of the 7.5x canine reference sequence, the development of the 2.5 million SNP map and the subsequent development of SNP arrays for the dog, case-control association studies have become the method of choice for mapping canine genetic traits. Based on the unique LD patterns seen in dogs Lindblad-Toh et al. predicted that canine GWA studies could require as few as 15,000 evenly spaced SNP markers²³. Karlsson et al. mapped the *ridgeless* allele in Rhodesian Ridgebacks to a 750 kb region on chromosome 18 using only 9 ridgeless (case) and 12 ridged (control) dogs and approximately 27,000 markers⁹⁵. The first arrays developed, which contained 22,362 SNPs (CanineSNP20 BeadChip, Illumina) and 26,578 SNPs (Version 1 Canine array, Affymetrix), have both been discontinued and replaced by denser chips containing 49,663 SNPs (50K Version 2 Canine array, Affymetrix) and 173,662 SNPs (CanineHD BeadChip, Illumina). The number of samples required depends on the inheritance patterns of the trait under investigation: 20 each of cases and controls are recommended for simple Mendelian recessive traits with full penetrance, and 50 each of cases and controls for dominant traits. For complex traits a fivefold risk allele can be detected with 100 each of cases and controls, while a twofold risk allele requires approximately 500 each of cases and controls⁹⁶. This is significantly less than the 1,000-6,000 samples required to detect a twofold risk allele in human studies⁹⁴.

Due to the relatively long LD in dogs, loci identified through GWA studies can be large, up to 10 Mb. Cross-breed comparison of disease-associated loci could significantly benefit fine-structure mapping if the trait is shared by two or more breeds, provided they share the same genetic background²³. By combining the SNP genotyping data from two

breeds, Boxers and Bull Terriers, in which the major white spotting *S* locus segregates, Karlsson et al. reduced the associated region to 102 kb, compared with 800 kb identified using boxers only, and a mutation in microphthalmia-associated transcription factor (*MITF*) was identified⁹⁵. However, in the case of rare alleles segregating in only one breed or heterogeneous conditions such as PRA this across-breed strategy is usually not possible, and it may therefore not be possible to reduce trait-associated loci to an easily managed size.

1.7.5 DNA Testing

Upon identification of a causal mutation or markers associated with a disease, a DNA test can be developed. In the case of simple recessive, dominant and X-linked conditions, the availability of DNA tests to identify carriers of disease will allow breeders to make more informed decisions regarding the use of their breeding dogs. The hope is that by carefully selecting against a disease allele, the mutant gene can eventually be eliminated from a breed without adversely affecting the genetic diversity of the breed, as has been the case with canine leukocyte adhesion deficiency (CLAD) in the Irish Setter (IrS) breed.

CLAD is a fatal immunodeficiency disease with an aR mode of inheritance. Kijas et al. identified a single missense mutation in the Integrin beta (*ITGB2*) gene⁹⁷. This led to the development of a DNA test to identify dogs that carry or are affected with CLAD i.e. *ITGB2*^{+/-} or *ITGB2*^{-/-} respectively. The Animal Health Trust Genetic Services department (www.aht.org.uk/genetics_tests.html) has been testing IrSs for this mutation since 1999 and offers advice on breeding with carrier dogs. Over the years the proportion of mutant alleles in tested IrS populations from all over the world has decreased steadily (Figure 8). Since 2005 the Kennel Club has only registered IrSs that are genetically clear of CLAD implying that the disease has, in theory at least, been eliminated from the purebred IrS breeding population in the UK, or from the vast majority of the population at the very least.

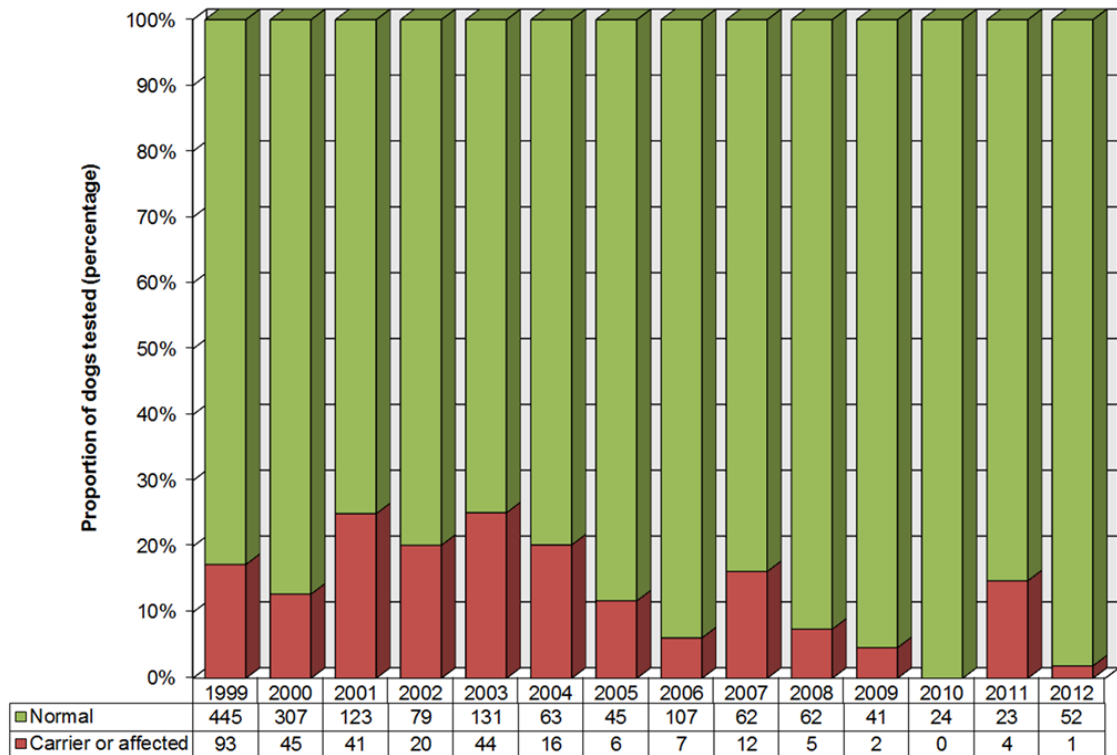


Figure 8. CLAD alleles in the IrS population.

DNA testing results for CLAD since 1999. The proportion of samples from dogs tested that have 1 or 2 copies of the *ITGB2* mutation, calculated from the figures in the table, has shown a decreasing trend. The sharp decline seen since 2005 coincides with the removal of carrier dogs from UK breeding stock by many breeders.

1.8 Breeds Under Investigation

Many dog breeds are prone to hereditary disease of some description, and more than 100 breeds are affected by PRA. For this project, three main breeds, the Golden Retriever (GR), Tibetan Spaniel (TS) and Gordon Setter (GoS) breeds, were selected for initial investigation. Follow-up studies were conducted in other breeds considered likely to share any mutations with the GR, TS or GoS.

1.9 Aims and Objectives

This study was conceived with three main aims:

- 1) Investigate the molecular basis underlying PRA.
- 2) Establish a diagnostic DNA test to be commercially and publically available.
- 3) Assess the relevance of findings to human retinal degenerations.

The primary aim of this project (comprising points 1 and 2) was to investigate the molecular basis of canine PRA in three dog breeds, GRs, TSs and GoSs. This was done initially by mapping mutation-harboring loci and then identifying probable PRA-causing mutations in these breeds with PRA of unknown aetiology. Follow up studies such as population screening and investigation of the functional effects of each mutation were conducted where possible. Population screening allowed the determination of the prevalence of the mutation, which is required to offer suitable advice for breeding strategies. This is of particular benefit to the individual breeds as PRA could, in theory, be eliminated from each breed while avoiding significant loss of genetic diversity. Functional studies, while of limited benefit to the breeders or their dogs, may allow confirmation of whether probable disease causing mutations are certainly disease-causing. Functional investigations could also reveal information about previously unknown or little understood genes or pathways.

The secondary aim was to determine if any of the PRAs investigated are suitable models for equivalent human retinopathies. While not within the scope of the current project, PRA affected dogs could form the foundation of research colonies to be used for gene therapy investigations. Alternatively the results of the canine studies could provide novel candidate genes for unknown human retinopathies such as RP or LCA.

Chapter **2** **Materials and Methods**

2.1 Canine DNA Sample Collection

2.1.1 PRA Diagnosis and Sample Identification

The diagnosis of individual dogs was determined by veterinary ophthalmologists. Many dogs were examined through the BVA/KC/ISDS Eye Scheme in the UK, the Swedish Kennel Club Eye Scheme in Sweden, or the Canine Eye Registry Foundation (CERF) in the USA, while others were independently examined. Cases were defined as dogs diagnosed as affected with PRA on the basis that they displayed ophthalmoscopic signs consistent with PRA including tapetal hyper-reflectivity, vascular attenuation and pigmentary changes in the non-tapetal fundus. Controls for the GWA studies were dogs free of inherited eye disease of any kind, and of a predetermined age or older at the time of examination (see section 2.3 and Table 3), depending on the breed under investigation. In subsequent investigations “unaffected” dogs were simply those of any age that had not been diagnosed with PRA.

2.1.1.1 Exclusion of Known PRA Mutations

PRA cases were screened for mutations known to cause PRA, if any were known, in the breed under investigation. DNA from 80 PRA-affected GRs and 46 PRA-affected LRs were screened for the *prcd* (progressive rod cone degeneration) mutation using the TaqMan® Allelic Discrimination (AD) technique (section 2.6.1.2). DNA from six IrS dogs affected with late-onset PRA were screened for the rod cone degeneration (*rcd1*) mutation using the Sanger sequencing method (section 2.6.1.3). For additional information regarding these techniques see sections 2.4.1.3-2.4.1.8 and 2.6.

2.1.1.2 Ophthalmoscopic Examination

A number of dogs had retrospective ophthalmoscopic and other examinations on the basis of their affected genotypes at PRA loci. These included fundus examination, pupil reflexes in response to light, the ability to navigate obstacles in the dark and retinal function by ERG. ERGs were conducted in a darkened room, without dark adaptation or sedation, and responses to 12 flashes were measured for each eye to mitigate for muscle artefacts.

2.1.2 Sample Collection

2.1.2.1 Samples for DNA extraction

All DNA samples were collected from privately owned pet dogs with the owners' consent and where possible with copies of eye examination certificates and 5-generation pedigrees. All information pertaining to samples was used to establish relationships between individuals and draw pedigrees using the software package Progeny Lab (Progeny Software, LLC). Samples collected by the Animal Health Trust (AHT) were received in the form of residual blood samples (drawn for the purposes of a veterinary procedure) collected into EDTA tubes or cheek cells collected with buccal swabs, and were stored at -20°C until DNA was extracted. Samples collected from collaborators were received in the form of DNA in solution and were stored at -20°C.

2.1.2.2 Samples for mRNA extraction

A retinal tissue sample from a dog of unknown breed that was free of PRA was taken post mortem, with the owner's consent. The retina was immersed in RNAlater (Ambion) to prevent RNA degradation, and stored at -20°C.

Two blood samples, one from a Tibetan Spaniel with PRA and one from a healthy dog (breed unknown), were collected into EDTA tubes, kept at room temperature while in transit, and processed immediately upon receipt to extract RNA.

2.1.3 DNA and RNA Extraction

2.1.3.1 DNA Extraction from Blood

Genomic DNA (gDNA) was extracted from whole blood using a protocol based on the Nucleon BACC2 Genomic DNA Extraction Kit (Tepnel Life Sciences). Briefly, Nucleon Reagent A (section 2.9) was added to up to 5 mL whole blood in a 50 mL conical tube (Falcon) to a total volume of 25 mL, and mixed by inversion for four

minutes to lyse the red blood cells. White blood cells were collected by centrifugation at 4,300 x g for ten minutes, discarding the supernatant. To lyse the white blood cells, 2 mL Nucleon Reagent B (section 2.9) was added to the pellet, which was then fragmented by vortexing and incubated at 37°C overnight. The solution was transferred to a 15 mL conical tube (Falcon) and protein precipitation was achieved by adding 800 µL 5M sodium perchlorate and mixing by inversion. This was followed by the addition of 2 mL chloroform and mixing by inversion for 4 minutes. Centrifugation at 4,300 x g for 5 minutes facilitated the separation of the fluid phases to form three distinct layers, chloroform at the bottom, a layer of precipitated proteins in the middle and solution containing the DNA at the top. The DNA-containing aqueous phase was transferred to a new 15 mL tube and the protein precipitation stages (addition of chloroform and centrifugation) repeated. The top aqueous phase was again transferred to the new tube, 5 mL ice cold absolute ethanol was added to precipitate the DNA, and the solution was slowly inverted until precipitated DNA became visible. Finally, a sealed glass hook was used to hook out the precipitated DNA, which was allowed to air dry and then washed off into a microcentrifuge tube with 300 µL 1X TE buffer (section 2.9).

2.1.3.2 DNA Extraction of Buccal Mouth Swabs

For samples collected as buccal mouth swabs, DNA was extracted using a QIAmp® DNA Blood Midi Kit (Qiagen) and an optimised protocol. Briefly, buccal cells were lysed by suspending up to four buccal swabs from a single donor in a solution of 1 mL phosphate buffered saline, 1 mL Lysis Buffer and 50 µL 20 mg/mL Proteinase K, in a 15 mL conical tube (Falcon). The solution was incubated at 56°C for 20 minutes, during which time it was mixed by inversion briefly every 5 minutes. DNA was then precipitated by adding 1,000 µL absolute ethanol and mixing by vortexing. Precipitated DNA was collected by passing the solution through the filter units supplied in the kit by centrifuging at 1,500 x g for 5 minutes. DNA on the filter was then purified with two wash steps using wash buffers AW1 and AW2, each followed by centrifugation at 4,300 x g for 1 minute and 15 minutes respectively. Finally the DNA on the filter was eluted by adding 150 µL Elution Buffer and centrifuged at 4,300 x g for 1 minute. The elution step was repeated twice, resulting in a final volume of approximately 450 µL. All extracted DNA samples were stored at -20°C.

2.1.3.3 Messenger RNA Extraction from Retina

RNA was extracted from retinal tissue using an RNeasy Protect Mini Kit (Qiagen) according to the manufacturer's instructions. Briefly, 15-20 mg RNAlater stabilised tissue was disrupted using a mortar and pestle, transferred to a new microcentrifuge tube and 600 μ L Buffer RLT was added. The lysate was transferred to a QIAshredder spin column (Qiagen) in a 2 mL collection tube for homogenisation. After centrifugation at 16,000 x g for 3 minutes the supernatant of the flow-through was transferred to a new microcentrifuge tube, without disturbing the pellet. 600 μ L 70% ethanol was added to the lysate, mixed and 700 μ L sample was transferred to an RNeasy spin column placed in a 2 mL collection tube. After centrifugation at 16,000 x g for 15 seconds the flow-through was discarded, and the process repeated with the remaining sample solution. RNA on the spin column membrane was washed by adding 700 μ L Buffer RW1 and centrifuging at 16,000 x g for 15 s. This was repeated a further two times using 500 μ L Buffer RPE, centrifuging first for 15 seconds and then for 2 minutes. The RNeasy spin column was then transferred to a new 1.5 mL collection tube, and 50 μ L ultrapure filtered water added to the membrane (Ultra Clear™ RO/EDI Water Purification System, hereafter termed UFWater) before centrifuging 16,000 x g for 1 minute to elute the RNA. Extracted RNA was stored at -20°C.

2.1.3.4 Messenger RNA Extraction from Blood

RNA was extracted from EDTA-stabilised whole blood using the PerfectPure RNA Blood Kit (5 Prime) according to the manufacturer's instructions. Briefly, up to 3 mL whole blood was incubated with 3 volumes RBC Lysis Solution to facilitate red blood cell (RBC) lysis, followed by centrifugation at 2,000 x g for 2 minutes and removal of the supernatant to collect the white blood cells. White blood cells were then lysed by the addition of Lysis Solution and detergent tris(2-carboxyethyl)phosphine (TCEP). Nucleic acids were extracted by passing the lysate through a Purification Column at 16,000 x g for 2 minutes, and further purified through a number of wash steps, including the removal of DNA with the addition of DNase Wash Solution. Each wash step included centrifugation at 16,000 x g for 1 or 2 minutes. Finally RNA was eluted in 50 μ L Elution Solution and collected by centrifugation at 16,000 x g for 1 minute.

2.1.4 DNA Normalisation

DNA concentration and purity was determined by a NanoDrop 1000 spectrophotometer (Thermo Scientific) and/or a Qubit® Fluorometer with the Qubit® dsDNA BR Assay Kit (Invitrogen), and normalised to the required concentrations. Samples required for GWA mapping were normalised to a final concentration of 75 ng/μL and a final volume of 50 μL. Samples for target-capture and next-generation sequencing were normalised in 1X TE buffer to a final concentration of 25 ng/μL and a final volume of 120 μL. For “in-house” investigations such as mutation screening, microsatellite genotyping and Sanger sequencing, aliquots of DNA diluted with UFWater to concentrations of approximately 20 ng/μL were used. Samples with low concentrations were precipitated using MultiScreen_{HTS} PCR filter plates or Amicon Ultra-0.5 mL Centrifugal Filters (Millipore) and resuspended in UFWater.

2.2 Polymerase Chain Reaction

The polymerase chain reaction (PCR) is the targeted amplification of a specific section of DNA, and is used in a wide variety of applications. The basic principle of PCR involves 3 steps that are repeated approximately 30 to 40 times in sequence:

- 1) **Denaturation** of the double-stranded DNA template at approximately 95°C.
- 2) **Annealing** of the primers to complementary sequence on the denatured DNA at temperatures that are dependent on the melting temperature (T_m) of the primers, usually between 55°C and 60°C.
- 3) **Extension** of the complementary strand from the 3'-end of the annealed primers, at approximately 72°C.

2.2.1 Primers

All primers for microsatellite genotyping (section 2.3.3), Sanger sequencing from gDNA or complimentary DNA (cDNA) (section 2.4.1 and 2.6.1.3), mutation screening by AD (section 2.6.1.2) or Amplified Fragment Length Polymorphism (AFLP; section 2.6.1.1) analysis and quantitative PCR (qPCR; section 2.5), were designed with Primer3 (Appendix II)⁹⁸ to meet the following criteria:

- Size: 16-24 base pairs (bp), with an optimum length of 21 bp
 GC content: 40-60%, or 20-80% in GC-rich regions
 T_m : 57°C - 63°C, with an optimum of 60°C and a difference of no more than 2°C between each primer pair.

The sequence of each primer pair was blasted against the canine reference sequence, CanFam2.0, i.e. aligned with the reference sequence using BLAST, and pairs that produced more than 100 alignments were discarded. For primers intended for microsatellite genotyping an 18 bp tail (5'- TGACCGGCAGCAAAATTG -3') was added to the 5'-end of each left/forward primer. A primer complementary to the tail was labelled with a fluorescent molecule, FAM or VIC (with blue and green fluorescence respectively) e.g. 5'-FAM-CAATTTTGCTGCCGGTCA-3'. One primer of each pair intended for AFLP was labelled with a fluorescent molecule, FAM. All primers with and without fluorescent tags, including PrimeTime Dual-Labelled Probes for AD, were purchased from Integrated DNA Technologies (IDT).

2.2.2 Thermal Cycling Reactions

PCR reaction mixes varied between applications, and the reaction mixes and reagents used are listed in Appendix III. AmpliTaq Gold® DNA Polymerase (Applied Biosystems) was used for all microsatellite genotyping reactions. HotStarTaq Plus DNA Polymerase (Qiagen) was used for most other PCR reactions. In the case of difficult to amplify templates, such as those with a particularly high GC content, an additive, Q solution, was also added. For a small number of amplicons amplification with HotStarTaq Plus was unsuccessful and either AmpliTaq Gold® 360 DNA Polymerase (Applied Biosystems) or Herculase II Fusion DNA Polymerase (Agilent Technologies) was used with 360 GC Enhancer and DMSO respectively added for GC-rich templates. All thermal cycling was performed with a MJ Research PTC-225 gradient cycler or a Bio-Rad T100™ thermal cycler.

2.2.3 Thermal Cycling Parameters

Each standard PCR reaction starts with an extended denaturation to ensure the genomic double-stranded DNA is completely denatured, followed by cycles of denaturation, annealing and extension, and finally an extended annealing period to ensure that all fragments in the mixture are completely amplified and identical. Variations of the standard reaction were used for some applications and protocols are listed in Appendix IV. The standard cycling protocol was used for AFLP and amplification for Sanger sequencing. An additional step at the end of the standard cycling protocol was added for AD to measure the fluorescent emissions in the reaction (allelic reads). Finally, PCR amplification for microsatellite genotyping made use of two sets of three-step cycles, one for amplification from the tailed-forward primer and reverse primer and another, at a lower annealing temperature, for amplification from the reverse primer and the tail-complementary fluorescent primer.

2.3 Genome-Wide Association Mapping

2.3.1 SNP Genotyping

CanineSNP20 BeadChips or CanineHD BeadChips (Illumina Inc.) were used to obtain genotype calls for 22,362 and 173,662 SNPs respectively, using DNA from PRA cases and controls from four separate PRA cohorts (Table 3). Each BeadChip is a silica slide approximately the size of a standard microscope slide and contains millions of three micron beads. Each bead is covalently linked to hundreds of thousands of copies of a single locus-specific 50-mer oligo. Single-base, allele-specific extension of the oligo, and subsequent fluorescent staining of the extension enable identification of the allele at that locus.

This genotyping was conducted externally by service laboratories (Wellcome Trust Sanger Institute in the UK; The Centre National de Génotypage (CNG) in France and; Cambridge Genomic Services in the UK), but briefly the procedure involved whole genome amplification of 750 ng genomic DNA followed by enzymatic fragmentation. Fragmented DNA was then ethanol precipitated, resuspended and hybridised to the oligos on the BeadChip. Each oligo was extended by a single base, determining the SNP allele, and fluorescently stained. An Illumina BeadArray Reader determined the fluorescence intensities.

2.3.2 SNP Genotyping Data Analysis

Processing of SNP genotyping data was almost identical for all four cohorts, regardless of service laboratory or SNP array used, with the exception of genotype calling for the GoS_PRA cohort. This data was processed through the LUPA consortium⁹⁹, and therefore raw signal intensities were not released and genotype calls were made by the service laboratory (CNG). For the three remaining cohorts genotyping fluorescence signals were clustered and SNP alleles called automatically using GenomeStudio (Illumina). The resulting genotypes were imported into Progeny and assigned to the relevant samples and pedigrees. A GWA analysis was conducted using the software package PLINK¹⁰⁰. SNPs with a minor allele frequency <0.05 or missing genotype calls >0.1 within each cohort were removed from the analysis. This significantly reduced the number of SNPs used in downstream analyses (Table 3) although it did improve the overall quality, as indicated by overall genotyping rates $>99\%$ for each cohort. All samples had $>90\%$ of SNP genotypes called, with the exception of a single case from the GR_PRA2 study for which 80-90% of SNP genotypes were called. Many of the

dogs used in each cohort were known to be related through established pedigrees and therefore the presence of population stratification was expected. The extent of population stratification, represented by the inflation factor (λ) was calculated based on median χ^2 values using the following process:

χ^2_{observed} values were sorted from largest to smallest, and assigned a “rank” from 1 to the number of SNPs analysed (n).

$$p_{\text{expected}} = (\text{rank} - 0.5)/n$$

χ^2_{expected} was calculated using the p_{expected} and 1 degree of freedom

$$\lambda = (\text{Median } \chi^2_{\text{observed}})/(\text{Median } \chi^2_{\text{expected}})$$

For additional visualisation of population stratification, χ^2_{observed} and χ^2_{expected} values were used to create quantile-quantile (Q-Q) plots and multidimensional scaling (MDS) analysis of the 2x2 matrix of identity-by-state (IBS) pairwise distances was performed (Appendix I). Population stratification was corrected by IBS clustering based on the SNP genotypes, analysed for association within individual clusters and the results combined with a Cochran-Mantel-Haenszel (CMH) meta-analysis. As a correction for multiple testing, the GWA analysis was repeated using the Max(T) permutation procedure in PLINK (100,000 permutations). p_{raw} denotes p-values that have not been corrected for multiple testing, while p_{genome} denotes p-values that have. In the event of weak or insignificant association, an additional analysis, using a mixed model algorithm to correct for population stratification, was also conducted with Fast Mixed Model (FMM)¹⁰¹. For the purposes of plotting graphs using CMH- and FMM-generated data, the Bonferroni correction was applied to the significance level of 0.05:

$$\alpha = 0.05/n$$

Where n is the number of SNPs remaining after filtering (Table 3).

Finally, visual examination of SNP genotypes for homozygosity was performed to narrow and define critical regions. For the GR_PRA2 and TS_PRA cohorts PHASE was used for haplotype reconstruction¹⁰².

Table 3. SNP genotyping summary

Cohort*	SNP Array	Cases	Controls	Age of Controls (years)	Before filtering		After filtering	
					SNP genotypes	Overall Genotyping Rate	SNP genotypes	Overall Genotyping Rate
GR_PRA1	SNP20 BeadChip	27	19	8	22,362	0.9929	14,389	0.9984
GR_PRA2	CanineHD Genotyping BeadChip	10	16	7 [†]	173,662	0.9842	103,264	0.9904
TS_PRA	SNP20 BeadChip	22	10	4	22,362	0.9903	15,674	0.9991
GoS_PRA	CanineHD Genotyping BeadChip	16	22	9	173,662	0.9895	107,982	0.9901

* GR_PRA1 and TS_PRA were processed by the Wellcome Trust Sanger Institute in Cambridge, GR_PRA2 by Cambridge Genomic Services in Cambridge and GoS_PRA by The Centre National de Génotypage (CNG) in Paris.

† 2 of the controls were younger than 7 years of age due to sample availability constraints.

2.3.3 Microsatellite Genotyping

Microsatellite markers within the targeted loci were identified by searching the reference sequence (CanFam2.0) for dinucleotide repeats with a total length of at least twenty nucleotides. Primers flanking each marker were designed as discussed in section 2.2.1, and so that the resulting products would be between 200 and 400 base pairs in size. After amplification with the third, fluorescent primer complementary to the tail (sections 2.2.2 and 2.2.3), the products were separated by size on a 3100 or 3130xl Genetic Analyzer (Applied Biosystems) and the data analysed and alleles assigned to each sample with the GeneMapper software package (Applied Biosystems).

2.4 Sequencing of Loci and Candidate Genes

2.4.1 Sanger Sequencing

If functional candidate genes could be identified in the associated loci, as was the case in the GR_PRA1 (*SLC4A3*) and GoS_PRA (*C2ORF71*) cohorts, the gene sequence was compared between PRA-affected and unaffected dogs. Sanger sequencing of candidate genes was carried out on gDNA from six dogs of the breed under investigation that were included in the GWA study and made up of two PRA cases and four controls. For the GR_PRA1 cohort two out of four controls were obligate carriers i.e. dogs that are known to have produced PRA-affected offspring. DNA from one Miniature Long-Laired Dachshund and one Border Collie were included as additional controls.

In the GR_PRA1 cohort two variants in *SLC4A3* that were considered potentially disease-causing were sequenced in 25 dogs, comprising 19 PRA cases and 6 controls. In the GR_PRA2 cohort a potentially disease-causing variant in *TTC8* was sequenced in 26 dogs, comprising 10 PRA cases and 16 controls. In the GoS_PRA cohort one variant in *C2ORF71* that was considered potentially disease-causing was sequenced in 38 dogs, comprising 16 PRA cases and 22 controls. In the TS_PRA cohort two variants in exon 1 of *FAM161A* were sequenced in 43 TSs and 76 dogs comprising 31 breeds. In addition, one potentially disease-causing variant in intron 4 was sequenced in 80 TS dogs, comprising 29 PRA cases, 10 obligate carriers and 41 unaffected dogs.

Candidate genes in retinal and blood mRNA were also sequenced using the Sanger method. Amplification and sequencing protocols were identical to those of gDNA, with the exception that 2 μ L cDNA was used as template DNA instead of gDNA.

2.4.1.1 Canine Gene Annotation

When genes annotated in CanFam2.0 are compared with orthologous genes in other species such as the human and mouse it is often found that potential exons are not identified or intron-exon boundaries are inconsistent between the dog and other species. For each gene identified an alignment was produced using the Ensembl predicted canine transcript and available known mouse and human Ensembl transcripts (Table 4). Exons to target through sequencing of gDNA and cDNA were defined based on a combination of CanFam2.0 annotation and a ClustalW¹⁰³ alignment with orthologous genes. In addition, RNA-Seq data obtained from brain tissue (sequenced by Oliver Forman) was useful for confirmation of intron-exon boundaries of candidate genes.

Table 4. Ensembl gene identification numbers for alignment

	Dog Gene	Human Gene	Mouse Gene
<i>SLC4A3</i>	ENSCAFG00000015723	ENSG00000114923	ENSMUSG00000006576
<i>C2ORF71</i>	ENSCAFG00000023626	ENSG00000179270	ENSMUSG00000044375
<i>SPATA7</i>	ENSCAFG00000017354	ENSG00000042317	ENSMUSG00000021007
<i>TTC8</i>	ENSCAFG00000017478	ENSG00000165533	ENSMUSG00000021013
<i>FAM161A</i>	ENSCAFG00000003079	ENSG00000170264	ENSMUSG00000049811

2.4.1.2 Complimentary DNA First Strand Synthesis

Reverse transcriptase PCR was used to create and amplify double-stranded DNA from single-stranded retinal and blood mRNA. SuperScript®II Reverse Transcriptase (Invitrogen) was used for first-strand synthesis of cDNA. Briefly, 2 µL RNA (1 ng – 5 µg), 1 µL Oligo(dT)15 Primer (500 µg/mL, Promega), 1 µL 10 mM dNTP mix (from 100mM stock, Thermo Scientific) and 8 µL UFWater were mixed in a 200 µL tube, incubated at 65°C for 5 minutes on a thermal cycler and cooled to 4°C on ice. Added to this was 4 µL 5X First Strand Buffer, 2 µL 0.1 M Dithiothreitol (DTT) and 1 µL RNaseOUT™, followed by incubation at 42°C for 2 minutes on a thermal cycler. Finally 1 µL SuperScript™ II Reverse Transcriptase was added and the solution was incubated at 42°C for 50 minutes and then heated to 70°C for 15 minutes on a thermal cycler, and cDNA products were stored at -20 °C.

2.4.1.3 Primers and Amplification

All primers (Appendix II) were designed according to the criteria discussed in section 2.2.1. Primers for the amplification of exons in gDNA were designed in the introns, flanking the exons. Primers for the amplification of cDNA were designed in the exons, flanking splice sites. In both cases the fragments were amplified by the standard PCR protocol. Reagents and thermal cycling conditions (sections 2.2.2 and 2.2.3) are listed in appendices II, III and IV.

2.4.1.4 Agarose Gel Electrophoresis

Prior to Sanger sequencing the success of PCR amplification was determined through agarose gel electrophoresis. Molecular grade agarose (Bioline) gels of 2% were used (e.g. 2 g agarose in 100 mL 1X TAE buffer, section 2.9). 3 µL ethidium bromide was added to the gel solution prior to casting. Once the gel had set it was completely submerged in 1X TAE buffer in the electrophoresis tank (PerfectBlue™ Gel System

Mini L, PeqLab). 2 μ L of each PCR product mixed with 5 μ L STOP (section 2.9) was loaded into each lane in the agarose gel. As a size control 5 μ L 2-Log DNA Ladder (New England Biolabs, section 2.9) was also loaded in one of the empty lanes. Electrophoresis was performed at 100-120 V for approximately 30 minutes, and gels were analysed using a FluorChem™ 5500 Imager (Alpha Innotech). Samples for which a single band of the expected size was visible were considered suitable for purification and sequencing.

2.4.1.5 Purification of PCR Products

Successful amplification was followed by purification of the PCR products to remove excess primers and PCR reagents. PCR products were purified using 96-well MultiScreen_{HTS} PCR filter plates (Millipore) according to the manufacturer's instructions. Briefly, 100 μ L UFWater was added to the PCR product and this was then loaded into a well on the filter plate. The filter plate and vacuum manifold were assembled, and a vacuum of approximately 15 inches Hg was applied for 30 minutes, or until the wells appeared empty. The plate was then removed from the manifold, 20 μ L UFWater was added to each well containing DNA, and samples were mixed on a rocking table (Luckham Ltd) for 20 minutes to allow full resuspension of the PCR products. Purified products were transferred to 96-well PCR plates for storage and further use.

In some instances, individual PCR products were excised and purified directly from the agarose gel, using the QIAquick Gel Extraction Kit (Qiagen), according to the manufacturer's instructions. Briefly, the DNA fragment was excised from the gel using a scalpel, weighed and added to Buffer QC. The agarose was solubilised by incubating at 50°C for 10 minutes, after which 100% isopropanol was added to precipitate the DNA fragments. The solution was applied to a QIAquick filter column and subjected to a series of washing steps. These comprised the addition of wash buffers, followed by centrifugation at 17,000 x g for 1 minute. Finally, the purified DNA was eluted in 30 μ L Buffer EB for storage and further use.

2.4.1.6 Sequencing Reaction

For most exons or amplicons, the primers that were used for PCR amplification were also used for Sanger sequencing. For amplicons that were larger than approximately 600 bp additional internal primers (Appendix II) were also used for Sanger sequencing to ensure complete coverage of the forward and reverse strands of the amplicon.

Amplification products were then sequenced using 2 μ L purified PCR product, 0.5 μ L BigDye Terminator v3.1 (Applied Biosystems), 1 μ L SBDD buffer (section 2.9) and 1.5 μ L UFWater on a thermal cycler.

2.4.1.7 Isopropanol Precipitation

Excess reagents were removed from the sequenced product by precipitation. Briefly 50 μ L 80% isopropanol was added to each reaction in a 96-well plate, precipitating the DNA. DNA was collected at the bottom of the well by centrifugation at 2,600 x g for 30 minutes. The supernatant was discarded by turning the plate upside down and allowing the liquid to drain out. This was followed by the addition of 100 μ L 60% isopropanol to further purify the DNA, centrifugation at 2,600 x g for 10 minutes to collect the DNA and removal of the supernatant by draining upside down. The plate was then placed upside down on tissue, centrifuged at 160 x g for one minute and then dried at 37°C to remove all traces of isopropanol, thereby ensuring the purity of the pellet. The pellet was resuspended in 10 μ L Hi-Di Formamide (Applied Biosystems) and analysed on a 3100 or 3130xl Genetic Analyzer (Applied Biosystems), which generated raw sequence data for further analysis.

2.4.1.8 Sequence Data Analysis

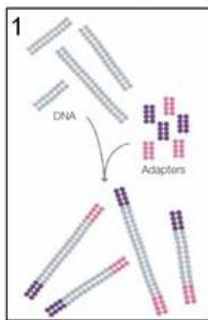
Sequence traces were assembled, analysed and compared using Pregap4 and Gap4 in the Staden Package¹⁰⁴. Raw sequencing data was processed through Pregap4 to determine where sequence quality was too low for alignment and convert the data format from ABI to ZTR, and create experiment files. Gap4 (Genome Assembly Program) is a graphical interface that was used to assemble and compare sequence reads, align and compare them to a reference sequence and view the sequence traces.

2.4.2 Targeted Next Generation Sequencing

Functional candidate genes could not be identified in a PRA-associated locus identified in the TS_PRA cohort at the time of experiment design. Multiple critical regions and candidate genes were identified in the GR_PRA2 cohort. For both of these cohorts therefore, the entire PRA-associated locus was sequenced in 10 samples, made up of PRA-affected and unaffected dogs. The TS_PRA cohort was made up of four PRA affected dogs and six unaffected dogs, including two obligate carriers. The GR_PRA2 cohort was made up of five PRA cases and five unaffected dogs, including two obligate carriers. The loci chosen for sequencing in these cohorts were large (5 Mb for TS_PRA and 3.8 Mb for GR_PRA2) and next generation sequencing (NGS) was therefore the

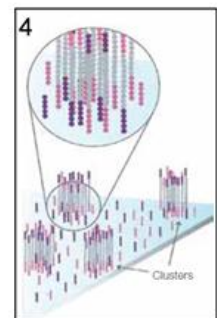
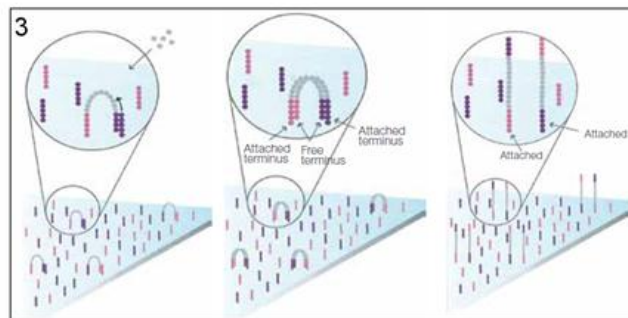
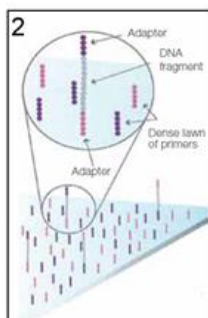
most affordable and convenient method to do this. NGS with the Genome Analyzer II or HiSeq 2000 (Illumina), which uses sequencing-by-synthesis (SBS) technology (Figure 9), enables massively paralleled sequencing and the process is made up of three parts: library preparation, cluster generation and sequencing. The library preparation was performed in-house, except where specified otherwise, but library samples were then sent to the High Throughput Genomics (HTG) group at The Wellcome Trust Centre for Human Genetics, University of Oxford, for cluster generation and sequencing on a HiSeq2000 platform (Illumina), in 51 bp paired end runs.

A. Library Preparation



1. DNA is randomly fragmented and adaptors ligated to the ends.

B. Cluster Generation

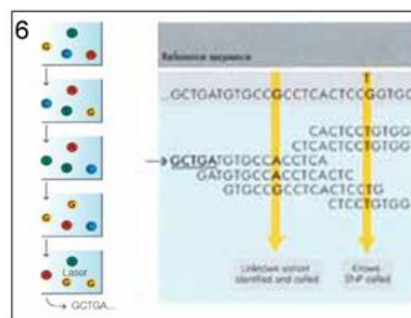
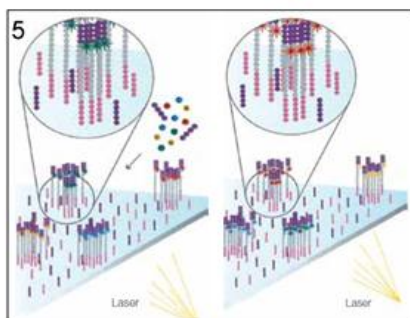


2. Single-stranded fragments randomly bound to flowcell surface.

3. Extension of bound fragments, bridge formation and amplification and denaturation.

4. Dense clusters of dsDNA generated, reverse strands removed.

C. Sequencing



5. Labeled nucleotides addition, laser excitation and first base identification. Followed by second base identification.

6. Repetition of sequencing cycles to determine base sequence, one base at a time. Data aligned and compared to reference.

Figure 9. Next-generation sequencing using sequencing by synthesis technology.

A) DNA library is prepared by fragmenting genomic DNA and adding adaptors. B) Cluster generation to amplify each DNA fragment. C) Sequencing one base at a time using fluorescently-labelled, reversibly-terminated nucleotides. Modified from http://www.illumina.com/documents/products/techspotlights/techspotlight_sequencing.pdf

2.4.2.1 Library Preparation – RNA Baits

A DNA library was prepared from each DNA sample (Figure 9A) using the SureSelectXT Custom MP4 Kits (Agilent Technologies), for the enrichment of targets between 1.5 Mb and 2.99 Mb. Each kit contained a custom Capture Library of biotinylated RNA baits, designed based on the CanFam2.0 reference sequence across the associated loci, using Agilent Technologies' eArray tool (<https://earray.chem.agilent.com/earray/>). The following Optimised Parameters were selected for the Design Options: Centred design strategy (baits are designed from the centre of a region outwards), baits 120 bp in length, a 2x bait tiling frequency and a 20 bp overlap into avoided regions. For the TS_PRA cohort baits were designed for a 5 Mb region on CFA10, from 63 to 68 Mb. For the GR_PRA2 cohort baits were designed for two regions on CFA8, one from 62.046 to 64.373 Mb (2.237 Mb) and the other from 71.147 to 72.634 Mb (1.487 Mb), covering a total of 3.813 Mb. Repeating elements identified with the RepeatMasker program (www.repeatmasker.org) were avoided, resulting in baits covering 2.7 Mb for TS_PRA and 2.3 Mb for GR_PRA2.

2.4.2.2 Library Preparation – Target Enrichment

Target enrichment was performed according to the manufacturer's instructions. Briefly, 3 µg genomic DNA was prepared by shearing it to an optimum size of 150-200 bp using a Covaris S220 (Covaris; performed by the Eastern Sequence and Informatics Hub (EASIH), Cambridge). The fragmented DNA ends were repaired to create blunt-ended fragments with 5'-phosphorylated ends, a 3'-dA overhang added, followed by the ligation of indexing-specific paired-end adaptors and amplification of the adapter-ligated library. After each step the samples were purified using Agencourt AMPure XP beads and a DynaMag™-2 Magnetic Particle Concentrator (Invitrogen). The quality and quantity of the amplified library was assessed with a NanoDrop. Half of the amplified library was then hybridised to the Capture Library (RNA baits), and hybrids selected using Dynabeads® M-280 Streptavidin (Invitrogen) magnetic beads and purified using the AMPure beads, both with the magnetic concentrator. Finally, unique Illumina index tags were added to each capture library by amplification and purified using the AMPure beads. The quality of the captured library was assessed with a 2100 Bioanalyzer (Agilent Technologies; performed by the Eastern Sequence and Informatics Hub (EASIH), Cambridge). The quantity of the captured library was assessed by qPCR using the KAPA Library Quantification Kit for the Illumina Genome Analyzer Platform (KAPA Biosystems), according to the manufacturer's instructions. Briefly, a 20 µL

reaction volume was used, made up of 10 μL 2X KAPA SYBR® FAST Universal qPCR Master Mix, 2 μL 10X Illumina Genome Analyzer Primer Premix, 4 μL UFWater and 4 μL library DNA (diluted 1:1,000 with Library Dilution Buffer, section 2.9) or 4 μL DNA standard. Reactions were performed in triplicate, on an Eco Real-Time PCR System (Illumina) (Table 5).

Table 5. Quantitative PCR thermal cycling conditions

Step Description	Temp	Duration	
Initial activation / denaturation	95 °C	5 minutes	
Denaturation	95 °C	30 sec	X35
Annealing / extension / data acquisition	60 °C	45 sec	

Average concentrations (pM) determined through qPCR were adjusted to account for differences in the average length of the DNA library and the length of the standards (452 bp) with the formula:

$$C = A \times \frac{452}{S} \times D$$

Where C is the adjusted concentration, A is the average concentration (pM) determined by qPCR, S is the average size of the fragments (bp) as determined by Bioanalyzer analysis and D is the dilution factor (1,000 in this project) used during qPCR.

Finally, all ten captured libraries were then pooled in equimolar amounts for multiplexed sequencing using the formula

$$V_i = \frac{V_f \times C_f}{n \times C}$$

Where V_i is the initial volume of sample required, V_f is the final volume required (20 μL), C_f is the final concentration required (10 nM), n is the number of samples in the pool (10) and C is the size-adjusted concentration calculated above, and converted from pM to nM. The final volume, if less than 20 μL , was adjusted with 1X TE buffer to 20 μL at 10 nM.

2.4.2.3 Cluster Generation

During cluster generation (Figure 9B) individual molecules of the prepared, pooled library are amplified in a flow-cell in preparation for sequencing. The adaptors to which the single-stranded library fragments are ligated bind to a lawn of oligos on the surface of the flow-cell. Bound fragments are extended to form double-stranded molecules and

the adapter on the unattached end also binds covalently to the flow-cell surface, forming a bridge. Through a series of expansions, bridge amplifications and denaturation each library fragment is clonally amplified, creating millions of unique clusters of DNA fragments. The reverse strands are removed, leaving only single-stranded DNA molecules attached to the flow-cell at one end. Finally the free ends are blocked and sequencing primers are hybridised to the DNA templates.

2.4.2.4 Next Generation Sequencing

The Illumina systems utilise SBS technology (Figure 9C), which means that DNA templates are sequenced in parallel on a base-by-base basis. Four fluorescently-labelled nucleotides compete to bind to the template. These bases are also reversibly-terminated so that in each round of synthesis extension by only a single base is possible. After each round clusters are excited by a laser, resulting in the emission of a colour that is unique to the newly bound base. Following laser excitation the blocking molecule and fluorescent label are cleaved from bound nucleotides, allowing the addition of the next base. The first 51 bases of each fragment, from both ends of the fragment were sequenced, resulting in 51 bp paired-end reads. The resulting data was subjected to the HTG group's standard analysis pipeline: Illumina base-calling pipeline; Illumina index tag identification and subsequent separation of generated data into ten batches, one for each sample; alignment to the CanFam2 reference genome using Stampy¹⁰⁵ and quality control.

2.4.2.5 NGS Data Analysis

Sequence data was received from the HTG group in the form of raw FASTQ files containing millions of sequence reads 51 bp in lengths and BAM files containing reads already aligned to CanFam2.0. These were run through an analysis pipeline created by PhD candidate Oliver Forman at the AHT, made up of various analysis tools (Figure 10), discussed below. I created a second analysis pipeline to assess target capture efficiency using the sequencing data. Both pipelines were written using the Perl programming language and the code, demonstration video clips and relevant files can be found on the CD in Appendix VI.

The **BWA** (Burrows-Wheeler-Alignment) tool aligns short sequences to the reference sequence (CanFam2.0) by assigning chromosomal coordinates to each read, creating a read file in the SAM (Sequence Alignment/Map) format¹⁰⁶. The “short” algorithm was used as reads were <200 bp in length. This algorithm has a low error rate (<3%),

supports paired-end reads and is one of the fastest available. **SAMtools** enables manipulation of SAM and BAM (Binary Alignment/Map) files¹⁰⁷. It was used to remove PCR duplicate reads, by retaining from multiple read pairs with identical chromosomal coordinates only the pair with the highest mapping quality. It was also used to index the reference FASTA and read BAM files. **Picard** (<http://picard.sourceforge.net>) is a java interface comprising a number of utilities that manipulate SAM files to create BAM files in which the reads were sorted by their chromosomal coordinates. The “CalculateHSMetrics” function of Picard was used in the Target Capture Efficiency pipeline to measure metrics for each sample, including total number of reads generated, average read depth and efficiency of target capture. **GATK** (Genome Analysis Toolkit) provides a Java programming framework for writing tools for the analysis of next generation sequencing¹⁰⁸. A number of built-in analysis tools were used in the analysis pipeline, including base quality score recalibration to improve the accuracy of base quality scores, identification of indels and SNPs, and depth of coverage calculations. The **Variation Effect Predictor** from Ensembl predicts the effects of variants on transcripts and regulatory regions annotated in Ensembl¹⁰⁹. **Pindel**¹¹⁰ identifies large deletions (1 bp – 10 kb) and medium insertions (1-16 bp), and their breakpoints. However, in our experience Pindel proved extremely unreliable and seldom produced any results at all. The **Integrative Genomics Viewer (IGV)** is a visualisation tool for large datasets, and was used to visually explore and examine NGS data^{111,112}. This included visual inspection for larger changes not identified by any tools in the analysis pipeline.

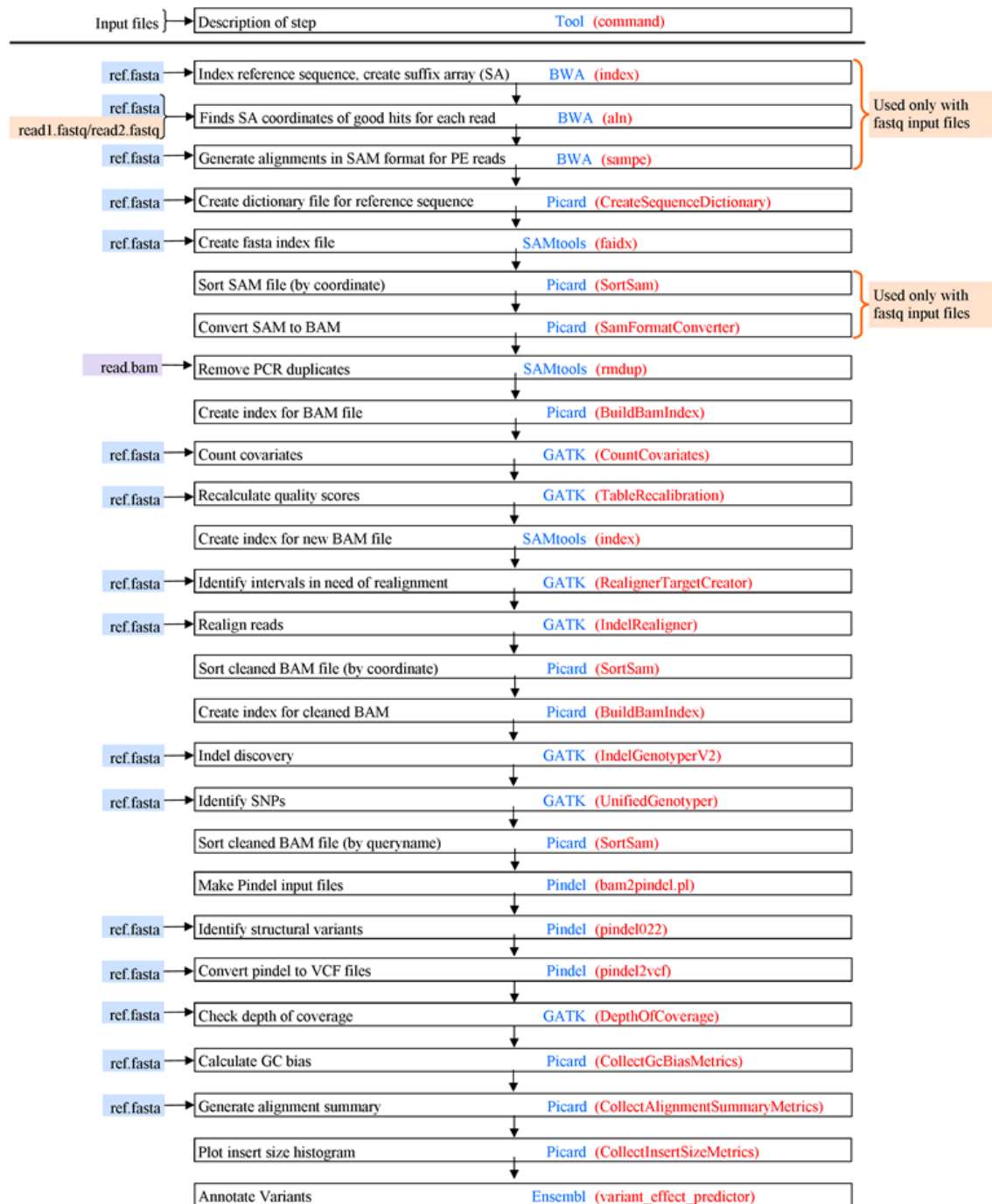


Figure 10. Next generation sequencing analysis pipeline.

The steps, tools and specific commands with tools used in the pipeline are indicated. Either two or three input files are required for the analysis of paired-end reads: a reference file in FASTA format and either two read files in FASTQ format or one read file in BAM format.

2.5 Quantification of mRNA transcripts

Quantitative real time PCR (qPCR) was performed using KAPA SYBR[®] FAST Universal 2x qPCR Master Mix (Kapa Biosystems, Appendix III) on an Eco Real-Time PCR System (Illumina, Appendix IV). Target primers (Appendix II) were designed based on sequence determined through sequencing of mRNA. For each target a standard curve was established using a 1:2 dilution series of retinal cDNA, from which the amplification efficiency (E) of each reaction was determined. Melt-curve analysis was conducted using target primers to identify the generation of primer-dimers. Expression levels of the targets were ascertained using the comparative cycle threshold (Ct) method, where two technical replicates were averaged. Due to variable E between targets, relative expression (RQ) levels were normalised to two housekeeping genes¹¹³, *ACTB* (β -Actin) and *TBP* (TATA box binding protein), using the Pfaffl method, i.e. based on E and Ct¹¹⁴, according to the formula:

$$\Delta Ct_{(\text{Target})} = \text{mean } Ct_{(\text{retinal cDNA})} - \text{mean } Ct_{(\text{blood cDNA})}$$

$$RQ_{(\text{Target})} = \frac{E^{\Delta Ct_{(\text{Target})}}}{\text{Geometric mean } [E^{\Delta Ct_{(\text{ACTB})}}, E^{\Delta Ct_{(\text{TBP})}}]}$$

2.6 Variant Screening

2.6.1 Variant Screening Methods

The method employed to detect previously known mutations or to create DNA tests for mutations identified during the course of this project were largely dependent on the type of mutation, as well as the cost and equipment available. For descriptions of primers and thermal cycling see section 2.2.

2.6.1.1 Amplified Fragment Length Polymorphism Analysis

AFLP analysis was preferred when the mutation was an insertion or deletion (indel) that resulted in a change in the size of an amplicon. A pair of primers, one of which is fluorescently-labelled, was used to amplify by PCR the region containing the mutation. In this way, products that differed by as little as a single base-pair in size could be separated and accurately identified using a 3100 or 3130xl Genetic Analyzer and GeneMapper software (Applied Biosystems).

2.6.1.2 TaqMan[®] Allelic Discrimination

AD analysis provided a convenient method to identify SNPs that do not affect the size of an amplicon. This method makes use of allele-specific fluorescently labelled probes to determine which alleles are present. Each probe is labelled with a unique fluorescent reporter on the 5'-end and a quencher on the 3'-end. As long as the reporter and quencher are in close proximity, as they are when attached to the same probe, the quencher absorbs the energy emitted by the reporter dye. However, when the primer hybridises to a complementary DNA template in a PCR reaction the quencher is cleaved by the 5'- to 3'-exonuclease activity of polymerase. The resulting spatial separation of the reporter and quencher increases the fluorescence of the reporter dye. By determining which reporter dye is fluorescing, the allele-specific primer that is hybridizing with the DNA template, and therefore the allele, can be identified. Briefly, the region containing the mutation was amplified using a pair of primers and at the same time allele-specific PrimeTime Dual-Labelled Probes, labelled with 6-FAM or HEX reporters and Black Hole Quencher[®] 1, were hybridised to the mutation. PCR amplification and AD plate read and analysis were carried out on a Techne Quantica Real Time Thermal Cycler with the Quansoft software (Bibby Scientific).

2.6.1.3 Sanger Sequencing

Sanger sequencing (sections 2.4.1.3-2.4.1.8) is preferred to screen small numbers of samples for indels or SNP mutations. The same primers are used to amplify and sequence the region containing the mutation.

2.6.2 Population Screening of Variants

An appropriate screening method (as discussed in section 2.6.1) was selected for rapid and cost-effective mutation screening. In the case of all four mutations identified in this project an AFLP assay was designed and optimised for each. Population screening of disease-causing mutations was done in two stages for each cohort.

2.6.2.1 Sample Cohorts Collected for Research

As many samples as possible from the breed under investigation and samples from related breeds, with known PRA clinical statuses and submitted to the AHT for research purposes, were screened (Table 6).

Table 6. Research samples screened for each mutation*

PRA-1	PRA-2	PRA-3	RCD4
709 GR (80 A, 39 C, 590 U)	475 GR (33 A, 5 C, 437 U)	247 TS (35 A, 16 C, 196 U)	126 GoS (22 A, 104 U)
132 LR (16 A, 116 U)	70 LR (17 A, 53 U)	76 TT (12 A, 64 U)	19 IRWS (3 A, 16 U)
49 CBR (2 A, 47 U)	48 CBR (2 A, 46 U)	23 LA (9 A, 7 C, 7 U)	72 IrS (10 A, 62 U)
92 FCR (92 U)	59 FCR (59 U)		56 ES (6 A, 5 C, 45 U)
			98 TT (17 A, 81 U)
			92 SP (1 A, 91 U)

* PRA clinical status: A = Affected, C = Obligate Carrier, U = Unaffected, LR = Labrador Retriever, CBR = Chesapeake Bay Retriever, FCR = Flat Coat Retriever, TT = Tibetan Terrier, LA = Lhasa Apso, IRWS = Irish Red and White Setter, IrS = Irish Setter, SP = Standard Poodle.

2.6.2.2 Genetic Services Testing Data

Genetic Services at the AHT offer DNA tests to the public for three of the four variants identified, PRA-1, PRA-2 and RCD4, using the same AFLP assays described above. The second stage of population screening made use of data obtained by Genetic Services over time for allele frequency calculations (Table 7). χ^2 statistics and p-values were calculated to test the hypothesis that the observed genotypes conform to Hardy Weinberg Equilibrium (HWE; $P = 0.05$, degrees of freedom = 1).

Table 7. Samples screened for each mutation through the AHT Genetic Services*

PRA-1	PRA-2	RCD4
855 GR (26 months)	425 GR (6 months)	1,788 GoS (24months)
		2,249 IrS (18 months)
		60 TT (3 months)

* The duration for which the test has been available for each breed is indicated in brackets

2.7 Human Variant Identification

2.7.1 Sample Identification and Collection

Variant screening of *SLC4A3* in samples from 200 unrelated patients with retinal degeneration was performed. These included samples from 192 sporadic cases of aRRP, as well as samples from eight individuals with familial recessive retinal degenerations (five aRRP, one rod monochromasy, one retinal dystrophy and one cone-rod dystrophy). For all eight familial samples a locus containing *SLC4A3* had previously been identified through autozygosity mapping. Samples were identified at Moorfields Eye Hospital using standard techniques, and were of mixed ethnic origin. In adherence with the Declaration of Helsinki informed consent for genetic studies was obtained. Samples were normalised to 20 ng/ μ L in ultrapure Milli-Q filtered water (Millipore), hereafter termed MQ.

2.7.2 Sanger Sequencing of *SLC4A3*

The process of Sanger sequencing of human samples, from primer design through to trace analysis with Staden, was largely similar to that of canine samples (sections 2.4.1.3-2.4.1.8). Instances where slightly different protocols were used are described here.

2.7.2.1 Purification of PCR Products

PCR products were purified using Multiscreen PCR _{μ 96} filter plates (Millipore) according to the manufacturer's instructions. Briefly, 75 μ L MQ was added to the PCR product and this was then loaded into a well on the filter plate. The filter plate and vacuum manifold were assembled, and a vacuum of approximately 20 inches Hg was applied until the wells appeared empty (approximately 10 minutes). An additional 50 μ L MQ was added to each well and the vacuum applied until wells appeared empty. The plate was then removed from the manifold, 20 μ L MQ was added to each well containing DNA, and samples were mixed on a Vortex-Genie 2 at low speed (Scientific Industries, Inc.) for 10 minutes to allow full resuspension of the PCR products. Purified products were transferred to 96-well PCR plates for storage and further use.

2.7.2.2 Sequencing Product Clean-up

Excess reagents were removed from sequenced products using the Montage SEQ96 Cleanup Kit (Millipore), according to the manufacturer's instructions. Briefly, 25 μ L Injection Solution was added to the sequencing product and this was then loaded into a well on the filter plate. The filter plate and vacuum manifold were assembled, and a

vacuum of approximately 20 inches Hg was applied until the wells appeared empty (approximately 10 minutes). An additional 25 μ L Injection Solution was added to each well and the vacuum applied until wells appeared empty. The plate was then removed from the manifold, 20 μ L Injection Solution was added to each well containing DNA, and samples were mixed on a Vortex-Genie 2 at low speed (Scientific Industries, Inc.) for 10 minutes to allow full resuspension of the sequencing products. Purified products were transferred to 96-well PCR plates and then run on a 3730 Genetic Analyzer (Applied Biosystems).

2.7.2.3 Variant Pathogenicity Analysis

The potential pathogenicity of variants identified were assessed with various bioinformatics tools, described in section 2.8. This included novelty of the variants and their predicted effect on exon splicing, amino acid sequence and protein function.

2.8 Bioinformatics Tools

The **Ensembl** genome browser (www.ensembl.org) was used for examining the canine (CanFam2.0), human (GRCh37) and mouse (GRCm38) assemblies. **BLAST** (**B**asic **L**ocal **A**lignment **S**earch **T**ool) and **BLAT** (**B**LAST-**L**ike **A**lignment **T**ool) on the Ensembl genome browser (<http://www.ensembl.org/Multi/blastview>) were used to assess primer quality and identify the exact genomic location of all variants. **ClustalW2** (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>) is a multiple sequence alignment tool, suitable for DNA and protein sequence alignment, and was used mainly to compare orthologous DNA sequences. A Splice site prediction tool, **NNSPLICE0.9**¹¹⁵ (http://www.fruitfly.org/seq_tools/splice.html), was used to identify any variants that may affect intron splicing. The Single Nucleotide Polymorphism database, **dbSNP** (<http://www.ncbi.nlm.nih.gov/projects/SNP/>), the **1000 Genomes** browser (<http://browser.1000genomes.org/index.html>) and the **Exome Variant Server** (EVS; <http://evs.gs.washington.edu/EVS/>) were used to determine whether any of the human variants discovered were novel and to determine allele frequencies of those SNPs previously identified through the 1000 Genomes Project and the NHLBI GO Exome Sequencing Project. Three tools were used to assess potentially pathogenic SNP variants. Firstly, **SIFT** (**S**orting **I**ntolerant **F**rom **T**olerant - <http://sift.jcvi.org/>) predicts whether amino acid substitutions are likely to affect protein function based on the degree of conservation of amino acid residues¹¹⁶. Secondly, **PolyPhen** (**P**olymorphism **P**henotyping - <http://genetics.bwh.harvard.edu/pph/>) predicts the possible impact of an amino acid substitution on the function and structure of a protein, using physical and comparative considerations¹¹⁷. Finally, **PMut** (<http://mmb2.pcb.ub.es:8080/PMut/>) predicts the pathological character of amino acid mutations based on the use of neural networks trained with a large database of neutral and pathological mutations¹¹⁸. The ExPaSy Bioinformatics Resource Portal (<http://expasy.org>) is a database of bioinformatics tools¹¹⁹, two of which were used: The **Translate** tool translates DNA sequence into amino acid sequence and the **Compute pI/Mw** tool calculates the theoretical isoelectric point and molecular weight of a protein.

2.9 Buffers and Solutions

Nucleon Reagent A: 6.304 g Tris Hydrochloride, 438.12 g Sucrose, 4.066 g MgCl₂, 40 mL Triton X-100, Make up to 4 L with UFWater, Adjust to pH 8.0 with 40 % NaOH.

Nucleon Reagent B: 63 g Tris Hydrochloride, 22.3 g EDTA, 8.8 g NaCl, 800 mL UFWater, Adjust to pH 8.0 with 2M NaOH, Adjust to 1 litre with UFWater and then add 10 g Sodium Dodecyl Sulphate (SDS).

1X TE Buffer (10mM Tris-HCL, 0.1mM EDTA): 10 mL 1 M Tris-hydrochloride, 2 mL 500 mM EDTA, Add UFWater to 1 L, Adjust to pH 8.0.

STOP: 10 mL Glycerol, 4 mL 500 mM EDTA, 1 mL 1M Tris, 1 mL Bromophenol blue. Diluted 1:5 with 1X TAE Buffer.

50X TAE Buffer: 242 g Tris base, 57.1 mL Glacial acetic acid, 37.2 g EDTA, Add double-distilled water to 1 L, Adjust to pH 8.0.

2-Log DNA ladder: 120 µL STOP, 240 µL 1X TAE Buffer, 40 µL 500 µg/mL 2-Log DNA ladder.

SBDD Buffer: 160 mL 1M Tris base, 3 mL 1 M MgCl₂, 50 mL Tetramethylenesulphone, 290 mL UFWater.

Library Dilution Buffer (LDB): 200 µL 1M Tris (10 mM Tris), 10 µL Tween (0.05% Tween), 19.79 mL UFWater.

Chapter **3** Golden Retrievers

3.1 Introduction and background

The GR breed is one of the most popular and recognisable breeds worldwide. As with many traditional “working” breeds, GRs can broadly be split into two groups based on whether they come from working or non-working breeding lines, with the latter consisting mostly of pets, but also show dogs. These dogs are easy to train due to their extraordinary intelligence and affable nature and are therefore very well suited to be working dogs in a wide variety of areas. While still used in their traditional role as gun dogs to retrieve game, the breed are now used in a variety of other areas including guide dogs for the blind and deaf, rescue dogs, and therapy for the physically and mentally disabled. The GR breed is one of the more genetically heterogeneous breeds and has a relatively short LD ($D' = 0.5$ at 400-700kb) in comparison with breeds such as the Akita ($D' = 0.5$ at 2.4 Mb)⁸⁹. Due to the popularity of the breed there is little importing and exporting of dogs for breeding, relative to the size of the breeding population. As such, dogs from different continents in particular tend to form distinct genetic populations¹²⁰, an important factor to consider when undertaking genetic mapping studies. GRs are prone to a large number of hereditary diseases including hip and elbow dysplasia, PRA, epilepsy, cataracts and muscular dystrophy¹²¹. The origin of the GR breed can be traced back to a mating of an ancestor of the modern day Flat-Coated Retriever (FCR) to a Tweed Water Spaniel (now extinct) in the mid-19th century. The FCR in turn, along with Labrador Retriever (LR) and Chesapeake Bay Retriever (CBR), can be traced back to the St Johns Water dogs from the St Johns region of Newfoundland, Canada, also in the 19th century¹²². All four breeds are also known to be affected by PRA and due to their common ancestors they could share PRA mutations(s). In the case of the LR breed the main form of PRA has been identified as *prcd*⁴⁷, but there is at least one additional,

unknown form of PRA in the breed. A very small number of GRs in the USA also have PRA caused by the *prcd* mutation, but it does not account for most PRA cases, and none of the European dogs tested (OptiGen®; <http://www.optigen.com>). It is therefore thought that there is at least one other form of PRA in this breed. This chapter describes the identification of the causal mutations for two additional forms of PRA in the breed, termed PRA-1 and PRA-2.

3.2 Diagnoses and Clinical Findings

PRA in the GR is ophthalmoscopically indistinct from most other forms of PRA. Typical ophthalmoscopic signs include tapetal hyper-reflectivity, attenuated blood vessels and optic nerve atrophy (Figure 11). The average age at diagnosis in our sample cohort, which is indicative of age of onset, is 6.02 years, but ranges from 1 to 11 years of age.

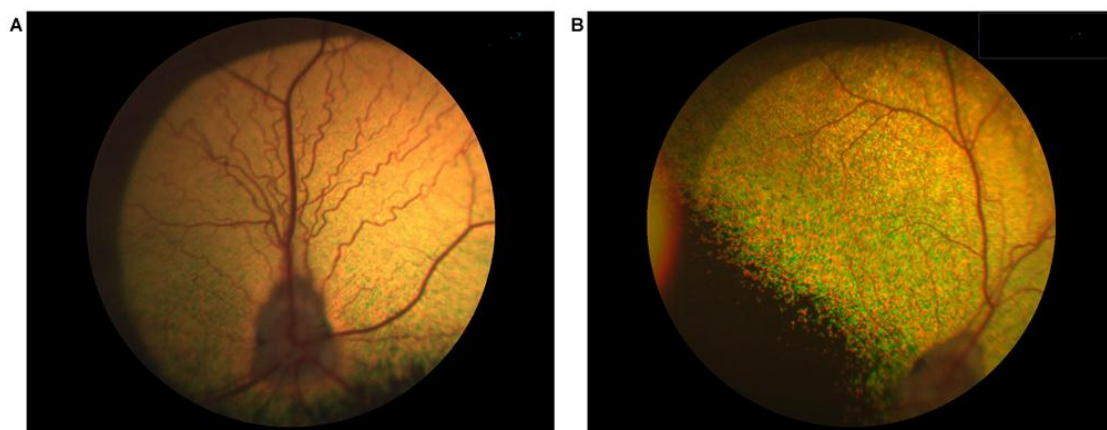


Figure 11. Fundus changes observed in typical PRA in GRs.

A) The fundus of a healthy GR. B) The fundus of a GR displaying signs typical of PRA. Due to the advanced stage of disease, the tapetum has gone dark (after a period of hyper-reflectivity), the blood vessels are attenuated and the optic disc is pale. The photos were taken on the same day from littermates that were 6 years old.

3.3 Prcd Variant Screening

DNA samples were collected from 80 GRs with PRA, of which 50 were from Sweden, and the remainder from the UK (n=9), France (n=8), Finland (n=3), Canada (n=3) and the USA (n=7). To exclude the possibility that the PRA-affected GRs were positive for the mutation already known to cause prcd, all 80 of our GR cases were investigated for the previously described, aR prcd mutation⁴⁷ (as described in sections 2.2 and 2.6.1.2). None of the affected GRs tested in this study was found to be homozygous for the prcd mutation. A single individual was heterozygous (G/A) while the remaining 79 were homozygous for the wild type (normal) allele (G/G).

3.4 SNP Genotyping

3.4.1 Genome Wide Association Mapping

GWA analysis of genotyping data from 46 GR dogs, (27 cases and 19 controls) genotyped with 14,389 SNPs (section 2.3) revealed a genome-wide significant association on chromosome 37 (CFA37; $p_{\text{raw}} = 1.14 \times 10^{-11}$, $p_{\text{genome}} = 1.0 \times 10^{-5}$). IBS clustering using genome-wide SNP marker data confirmed the presence of population stratification with a genomic inflation factor > 1 ($\lambda = 1.44$) and MDS analysis revealed a similar distribution of cases and controls (Appendix I). The signal on CFA37 remained strong ($p_{\text{raw}} = 1.94 \times 10^{-10}$) after correction for this by analysing for association with a CMH meta-analysis (Figure 12A). However, the inflation factor was increased ($\lambda = 1.46$), indicating that population stratification had not been corrected for. The signal remained significant after correction for multiple testing using permutations ($p_{\text{genome}} = 1.0 \times 10^{-5}$). Before and after CMH-correction, the strongest statistically associated SNP was BICF2G630131493 at 29.277 Mb. Alternative analysis of the data using FMM to correct for population stratification revealed similar results, with the strongest signal on CFA37 ($p_{\text{raw}} = 4.13 \times 10^{-11}$; Figure 12B) and a reduced inflation factor ($\lambda = 0.96$). After FMM-correction the strongest statistically associated SNP was BICF2S23032414 at 29.252 Mb. A number of weaker signals were seen on other chromosomes including 7, 19, 24 and 29, but none reached the level of significance ($p_{\text{genome}} = 0.05$). The statistically associated region on CFA37, defined as the region encompassing permuted SNPs statistically associated with PRA, extended from 28.331 to 29.847 Mb (data not shown) with the strongest statistically associated SNPs indicated (Figure 12C).

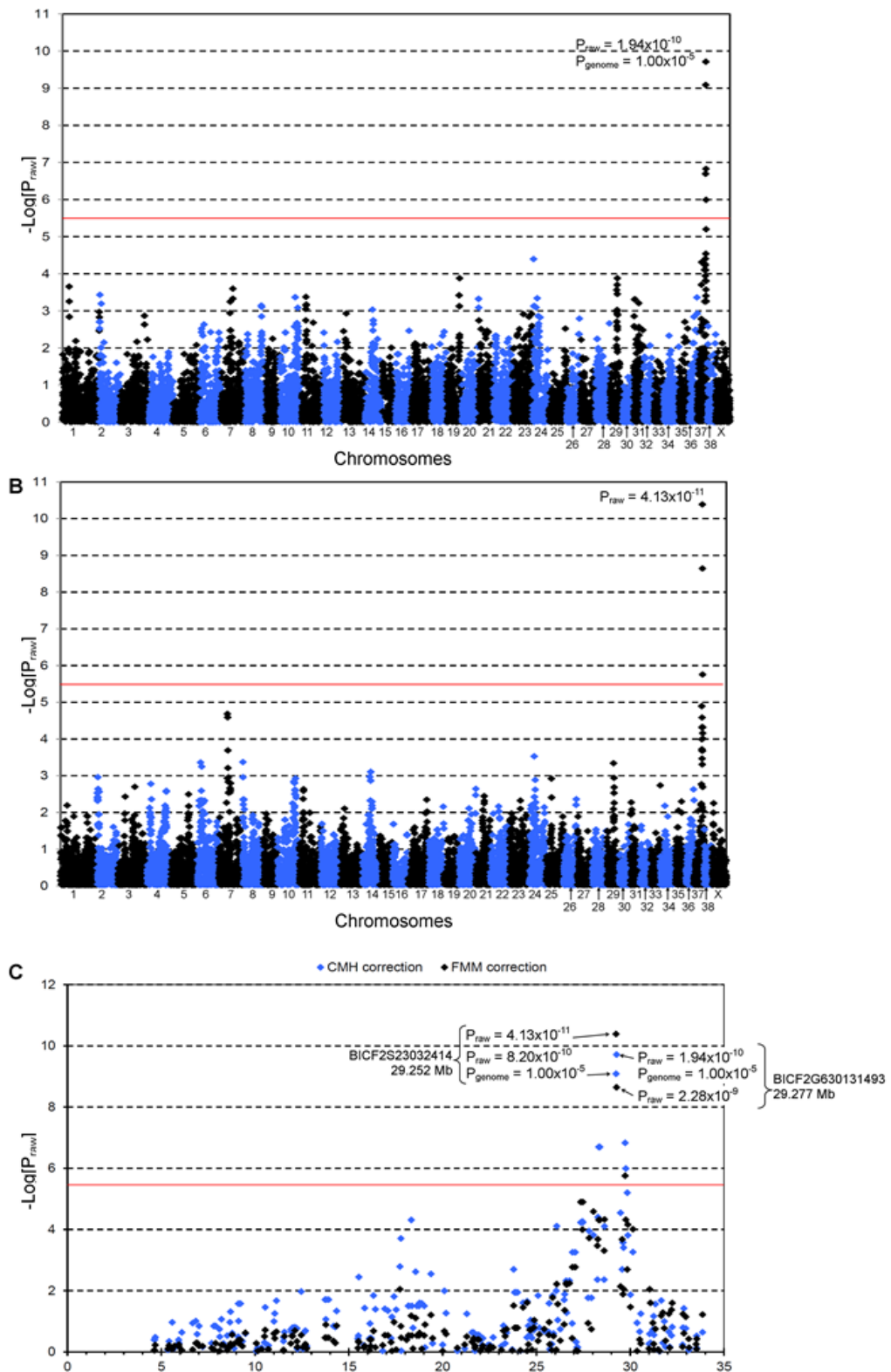


Figure 12. Genome-wide association mapping of PRA in GRs

$-\text{Log}_{10}$ of p-values after correction for population stratification. The red lines indicate the Bonferroni-corrected 5% significance level based on 14,389 SNPs. A) The CMH meta-analysis approach shows a statistically significant signal on CFA37 ($p_{raw} = 1.94 \times 10^{-10}$, $p_{genome} = 1.00 \times 10^{-5}$). B) The FMM approach also shows a statistically significant signal on CFA37 ($p_{raw} = 4.13 \times 10^{-11}$). C) Statistically associated SNPs span a 1.64 Mb region from 28.331 Mb to 29.847 Mb on CFA37.

3.4.2 Homozygosity Analysis

Visual examination of SNP genotypes revealed two overlapping homozygous regions on CFA37 from 27.813 to 29.277 Mb and 28.633 to 29.878 Mb (Figure 13A) and most of the PRA cases (23/27) analysed were homozygous for one or both of these regions. These homozygous regions overlap from 28.633 to 29.277 Mb on CFA37 (Shared Block in Figure 13A) and this 644 kb region was defined as the PRA critical region. It is likely that the four dogs that do not carry the haplotype associated with the critical region are affected with a form of PRA that is genetically distinct from that addressed here. The average age at diagnosis of the 27 cases used in the GWA study was 6.41 years. When these four cases were excluded the average age at diagnosis in the remaining 23 cases increased to 6.78 years, ranging from four to 11 years of age. These four dogs had a lower age at diagnosis (4.25, ranging from three to five years) than the rest of the dogs in the cohort. This is consistent with a different form of PRA segregating in these four dogs.

There are 27 genes within the PRA critical region, of which 25 have orthologous genes in humans (Figure 13B and Appendix V). Nonhomologous end-joining factor 1 (*NHEJ1*) has been associated with Collie Eye Anomaly, an ocular condition clinically distinct from PRA³⁵. It is therefore unlikely, although not impossible, that this gene is harbouring any PRA-causing mutations. None of the remaining 24 genes had previously been associated specifically with ocular function or retinal degeneration in humans, however a solute carrier anion exchanger gene (*SLC4A3*) located within this region was identified as a strong candidate based on function and its association with retinal degeneration in mice. *SLC4A3* mediates $\text{Cl}^-/\text{HCO}_3^-$ exchange across cellular membranes¹²³ and is expressed in various tissues including the Müller and horizontal cells of the retina¹²⁴. The effect of *SLC4A3*-deficiency has been described in a knockout mouse model showing that a selective inner retina defect is followed by photoreceptor degeneration¹²⁵, similar to the retinal degeneration seen in many forms of PRA and RP. *SLC4A3* was therefore considered an excellent candidate gene for further investigation.

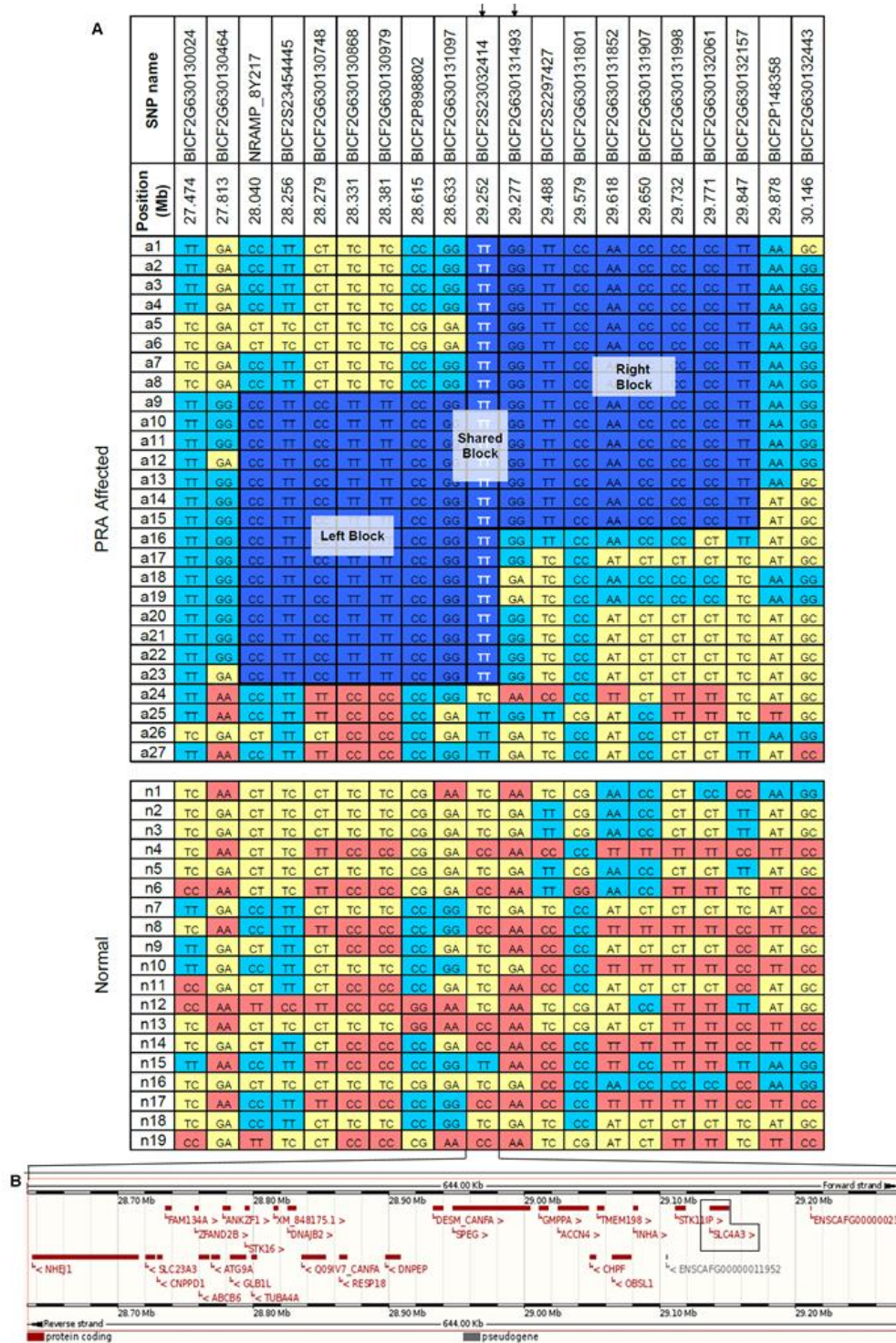


Figure 13. Homozygosity analysis of GRs with PRA

A) SNP genotypes for 27 PRA cases and 19 PRA controls, encompassing the 1.6 Mb region identified during the GWA study. Fifteen of the cases share a homozygous haplotype of eight SNPs (Left Block) and 15 cases share another homozygous haplotype of nine SNPs (Right Block). These two homozygous blocks overlap by just one SNP (Shared Block), defining a critical region of 644 kb between 28.633 Mb and 29.277 Mb on CFA37. Four of the cases do not share any haplotypes with the other 23 cases, but do share haplotypes with the controls. The strongest statistically associated SNPs are indicated by arrows (↓). B) There are 27 genes found within the critical region, as annotated on CanFam2.0.

3.5 Sequencing of Candidate Gene *SLC4A3*

3.5.1 Characterising *SLC4A3* Retinal Transcripts in Dog Retina

SLC4A3 occurs in two main isoforms, full-length (*SLC4A3_{fl}*) and cardiac (*SLC4A3_c*), formed by alternative splicing of the N-terminal exons (Figure 14)^{124,126}. Alignment of canine, human and mouse genomic and coding sequences revealed several inconsistencies and possible errors in the prediction of intron-exon boundaries of canine *SLC4A3* (Figure 14A-C). These included missing exons that are probably due to gaps in the reference sequence, perhaps due to extremely high GC content in these regions. All of the coding sequence of the *SLC4A3* retinal transcripts, from both the main isoforms (*SLC4A3_{fl}* and *SLC4A3_c*), was successfully amplified and sequenced in a healthy dog (sections 2.2 and 2.4.1) revealing that both isoforms are transcribed in the canine retina. In addition, intron-exon boundaries were confirmed to be identical to those of the human and mouse, which is in conflict with the boundaries predicted by Ensembl genebuild for the canine gene (Figure 14D). Alternative splicing of exon 5 of the full-length isoform results in two “full-length” isoforms that differ by the presence (*SLC4A3_{fl1}*) or absence (*SLC4A3_{fl2}*) of 27 nucleotides at the 5’ end. Amplification of the full 5’ and 3’ untranslated regions (UTR) for any isoform was unsuccessful. Sequencing of the retinal mRNA transcripts of all three isoforms revealed that canine *SLC4A3_{fl1}* (Genbank accession no HQ379706) contains 1237 amino acids, *SLC4A3_{fl2}* (Genbank accession no HQ379707) contains 1228 amino acids and *SLC4A3_c* (Genbank accession no HQ379708) contains 1030 amino acids, with predicted molecular weights of 136kDa, 135kDa and 114kDa respectively. *SLC4A3_{fl1}* is considered the reference transcript for the remainder of this document; the terms *SLC4A3_{fl}* and *SLC4A3_{fl1}* are therefore interchangeable.

3.5.2 Sanger Sequencing of *SLC4A3*

Primers for the sequencing of gDNA were designed to ensure sequencing coverage of the known exons as well as the probable 5’- and 3’-UTRs for both full-length and cardiac isoforms of *SLC4A3* (sections 2.2 and 2.4.1). To identify possible causal variants in *SLC4A3*, 20 of the 23 *SLC4A3* exons and splice sites were sequenced in gDNA from six GRs and two non-GR dogs (a Border Collie and a Miniature Long Hair Dachshund). The GRs comprised two PRA-affected dogs that were homozygous for the “affected haplotype” at the critical region, two obligate carriers that were heterozygous for the “affected haplotype” and two unaffected dogs that were homozygous for a

wildtype haplotype. The CanFam2.0 reference sequence and annotation for *SLC4A3* was incomplete and includes two gaps that cover exons 1, 2 and 7 of the gene (Figure 14D). All attempts at sequencing across the two gaps (including combinations of various thermal cycling parameters, primers based on canine and human reference sequence, and Taq polymerases) were unsuccessful and these three exons were therefore not sequenced. Comparison of the sequence data with the canine reference genome sequence revealed 50 sequence variants. Five of the sequence variants showed perfect segregation with PRA within the eight samples sequenced, and of these two were exonic (Figure 14D). Splice-site predictions (section 2.8), when taking the variants into consideration, revealed neither the introduction of novel splice sites nor the eradication of existing splice sites, although the prediction scores for splice sites involving two of the intronic variants were altered slightly (data not shown). Two of the five variants are exonic and non-synonymous. An A to C transversion at position 44 (c.A44C; CFA37:29,141,439) that results in a nonsynonymous substitution (p.Q15P) was found in the first exon of the cardiac specific isoform, but did not affect the full-length isoform (Figure 15B). The second exonic change was a frame shift caused by the insertion of a single cytosine in exon 16 of the full-length isoform (c.2601_2602insC; CFA37:29,147,633). It is predicted to cause a premature stop codon in exon 18 (p.E868RfsX104) possibly resulting in degradation of the mRNA by nonsense-mediated decay (NMD) or a truncated protein product (Figure 15B). Exon 16 is included in both the cardiac specific isoform and the full-length isoform of *SLC4A3*¹²⁴.

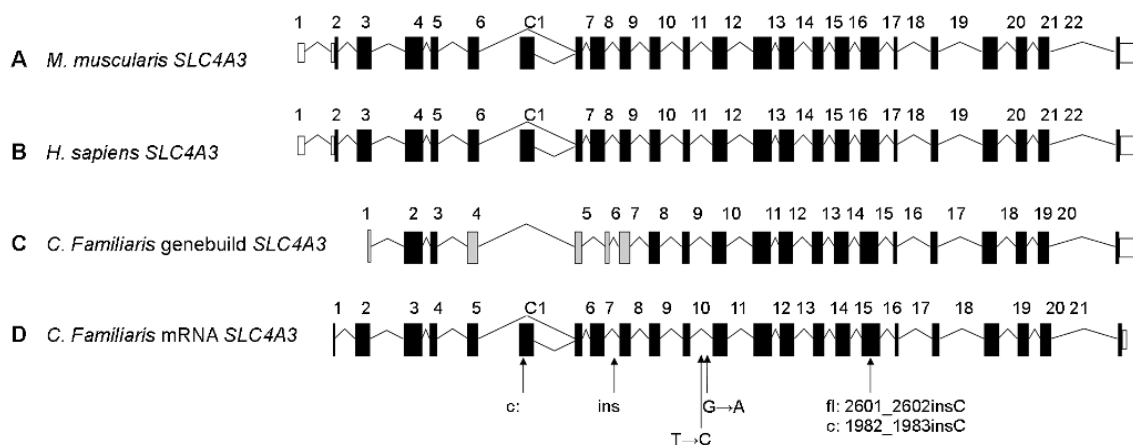


Figure 14. Graphical comparison of intron-exon boundaries of *SLC4A3*.

A) Mouse (*Mus musculus*) *SLC4A3*. B) Human (*Homo sapiens*) *SLC4A3*. C) Canine (*Canis familiaris*) *SLC4A3* as predicted by Ensembl genebuild. Sixteen of the genebuild exons predicted are identical to the mouse and human exons (black), but the intron-exon boundaries of exons 1, 4, 5 and 7 (grey) are inconsistent. Exon 6 (grey) shows no sequence similarity to its human and murine equivalents and is probably incorrect. D) Canine *SLC4A3* exons confirmed by sequencing the retinal mRNA transcript. The locations of the five fully-segregating sequence variants are indicated.

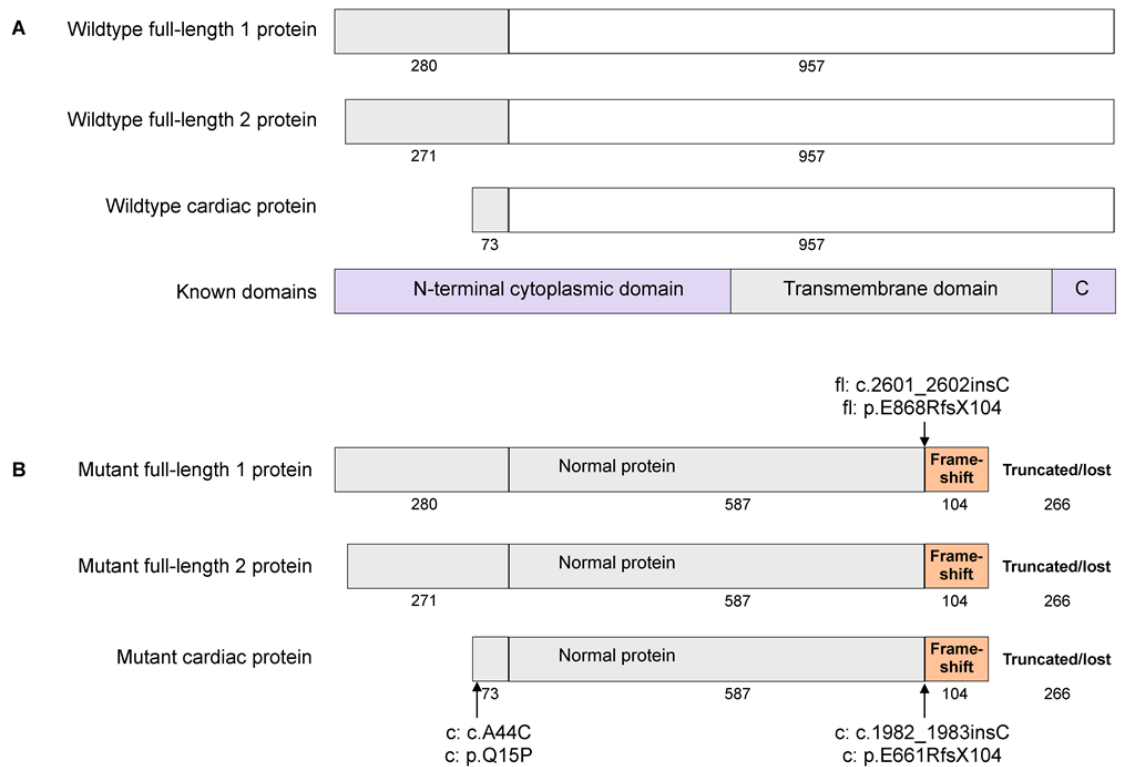


Figure 15. Effect of *SLC4A3*_{c.A44C} and *SLC4A3*_{c.2601_2602insC} on the *SLC4A3* protein.

A) The wildtype full-length 1 and 2, and cardiac isoforms are encoded by 1,237, 1,228 and 1,030 amino acids respectively. The proteins differ only at the N-terminus with 280 amino acids unique to the full-length isoform 1, and 73 unique to the cardiac isoform. The remainder of the protein is identical in all isoforms. The C-terminal cytoplasmic domain is represented by “C”. B) The cardiac isoform alone is affected by the c.A44C SNP variant (p.Q15P). All isoforms are affected by the c.2601_2602insC variant. 867 amino acids at the N-terminus of the full-length mutant protein and 660 of the cardiac isoform are unaffected. However the insertion causes a shift in the reading frame introducing 104 amino acids, leading to a premature termination codon. This results in a truncated protein product, lacking 266 residues of the C-terminus of both isoforms, if expressed.

3.5.3 *SLC4A3*_{c.2601_2602insC} Variant Validation

A total of 25 GR dogs (19 cases and 6 controls present in the GWA data set) were then screened (section 2.4.1) for the two coding variants, *SLC4A3*_{c.A44C} and *SLC4A3*_{c.2601_2602insC}, to investigate the association of both these variants with PRA and compare them with the most highly associated SNP, BICF2G630131493. The genotypes of the two variants i.e. homozygous for the reference allele, heterozygous or homozygous for the rare variant were concordant in all but two dogs (data not shown), indicating the two variants are in very close linkage disequilibrium. *SLC4A3*_{c.2601_2602insC} was more strongly associated ($p_{\text{raw}} = 4.1 \times 10^{-6}$) than *SLC4A3*_{c.A44C} or BICF2G630131493 ($p_{\text{raw}} = 1.2 \times 10^{-5}$ and $p_{\text{raw}} = 9.5 \times 10^{-5}$ respectively). Neither variant

shows complete concordance with PRA, but *SLC4A3*_{c.2601_2602insC} has a stronger likelihood of a deleterious effect on the protein i.e. it affects both transcripts shown to be present in dog retina and is a frame-shift rather than missense. *SLC4A3*_{c.A44C} was also predicted to be benign by PolyPhen and PMut. A comparison among 33 eutherian mammals (data not shown) showed that the position is not well conserved, indicating that the missense sequence variant may not be functionally important. This variant was therefore considered to be non-pathogenic and discarded from further analysis. Analysis of the segregation of *SLC4A3*_{c.2601_2602insC} with PRA in a family of Swedish ancestry (Figure 16) indicates that the form of PRA associated with this mutation, known hereafter as PRA-1, is recessive and fully penetrant. Fifty-six percent of the GR PRA cases in our sample cohort were homozygous for *SLC4A3*_{c.2601_2602insC} (*SLC4A3*^{-/-}), while eight were heterozygous (*SLC4A3*^{+/-}) and 27 were homozygous for the wildtype allele (*SLC4A3*^{+/+}, section 3.8.1). PRA-1 and prcd therefore, do not account for 35/80 PRA cases in our cohort. A second study was subsequently undertaken to identify the third form of PRA in the breed, known hereafter as PRA-2, using these 35 cases (see section 3.6). The average age at diagnosis in this sample cohort is 5.27, but ranges from 2 to 11 years of age.

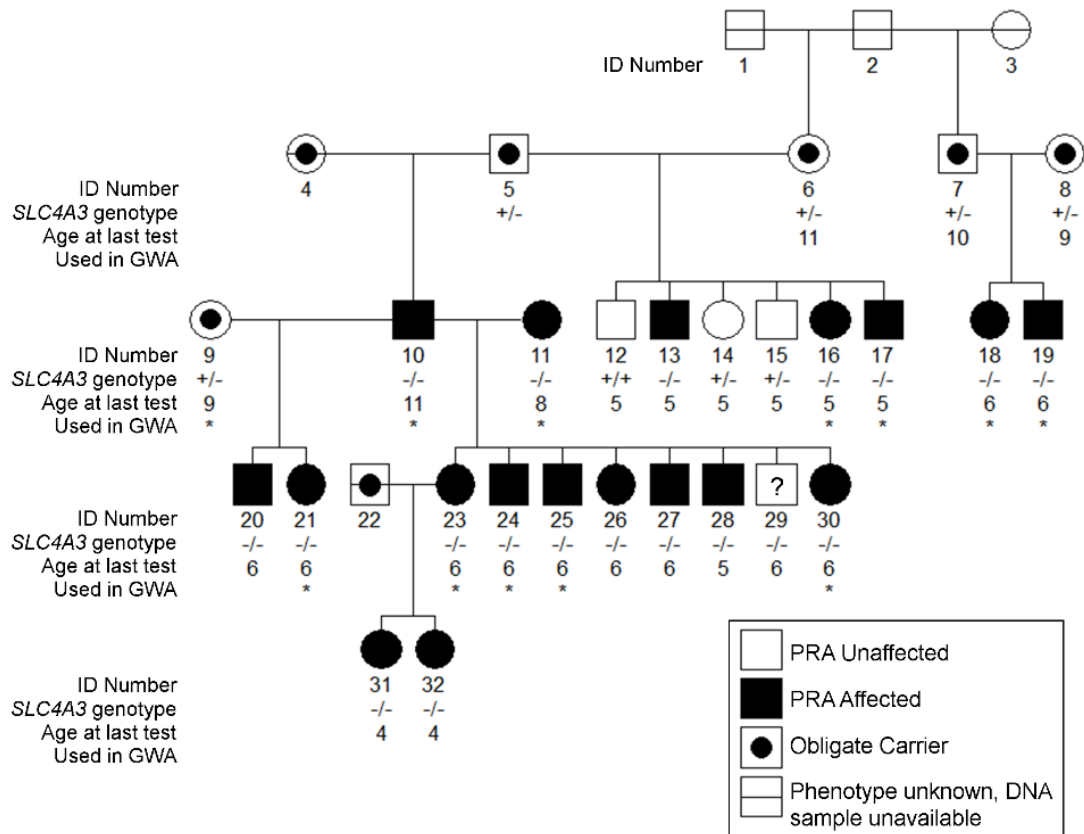


Figure 16. Segregation of PRA-1 in a GR pedigree.

The segregation of PRA-1 is consistent with an autosomal recessive mode of inheritance. The PRA-1 mutant allele (*SLC4A3*_{c.2601_2602insC}) is represented by “-” and the wildtype allele by “+”. A single dog (#29) that is homozygous for the PRA-1 mutant allele (-/-), had not been diagnosed with PRA, but did show ophthalmoscopic signs consistent with the early stages of PRA at the age of 6 years. Clinical information pertaining to and DNA samples from individuals 1-4 and 22 were unavailable. Samples used in the GWA study are indicated (*).

3.6 SNP Genotyping (PRA-2)

3.6.1 Genome Wide Association Mapping

GWA analysis of genotyping data from 26 GR dogs (10 cases that are not PRA-1 or prcd, and 16 controls) genotyped with 103,264 SNPs (sections 2.3.1 and 2.3.2) revealed a genome-wide significant association, using permutations, on chromosome 8 (CFA8; $p_{\text{raw}} = 1.30 \times 10^{-6}$, $p_{\text{genome}} = 0.036$). IBS clustering using genome-wide SNP marker data confirmed the presence of population stratification with a genomic inflation factor > 1 ($\lambda = 1.32$). MDS analysis revealed an uneven distribution of cases and controls (Appendix I), but this could be due to the low number of samples used, and could therefore not be corrected. The signal on CFA8 remained the strongest signal ($p_{\text{raw}} = 8.99 \times 10^{-5}$) after correction for this by analysing for association with a CMH meta-analysis, although it did drop below the level of genome-wide significance (Figure 17A). The inflation factor was reduced to less than optimal ($\lambda < 1$), suggesting an over-correction for population stratification ($\lambda = 0.84$). The signal remained below the level of significance after correcting for multiple testing ($p_{\text{genome}} = 0.148$). This region on CFA8, nevertheless, showed the strongest signal, 10-fold greater than that observed anywhere else on the genome. The region extended from 63.600 to 71.732 Mb with the most significantly associated SNP, BICF2G630416812, at 71.732 Mb. Further investigation of the signal on CFA8 revealed that it was formed by two apparently independent and distinct signals at 63.614 Mb (BICF2P582923; $p_{\text{raw}} = 1.16 \times 10^{-4}$; $p_{\text{genome}} = 0.197$) and 71.732 Mb (BICF2G630416812; $p_{\text{raw}} = 8.99 \times 10^{-5}$; $p_{\text{genome}} = 0.148$) respectively (Figure 17C). These loci are hereafter referred to as locus 1 and locus 2 respectively. A second analysis (section 2.3.2) using FMM to correct for population stratification identified two signals in the same regions (as described above) on CFA8 (Figure 17B and C): $p_{\text{raw}} = 1.82 \times 10^{-4}$ at 63.614 Mb (BICF2P582923) and $p_{\text{raw}} = 1.42 \times 10^{-5}$ at 71.732 Mb (BICF2G630416812). While these p-values have not been corrected for multiple testing they are consistent with the results observed during the CMH meta-analysis (Figure 17C). Using FMM to correct for population stratification resulted in a greatly reduced inflation factor ($\lambda = 0.27$), far below the optimum level of $\lambda = 1$, suggesting an over-correction for population stratification. The small sample size might also skew the corrected inflation factor.

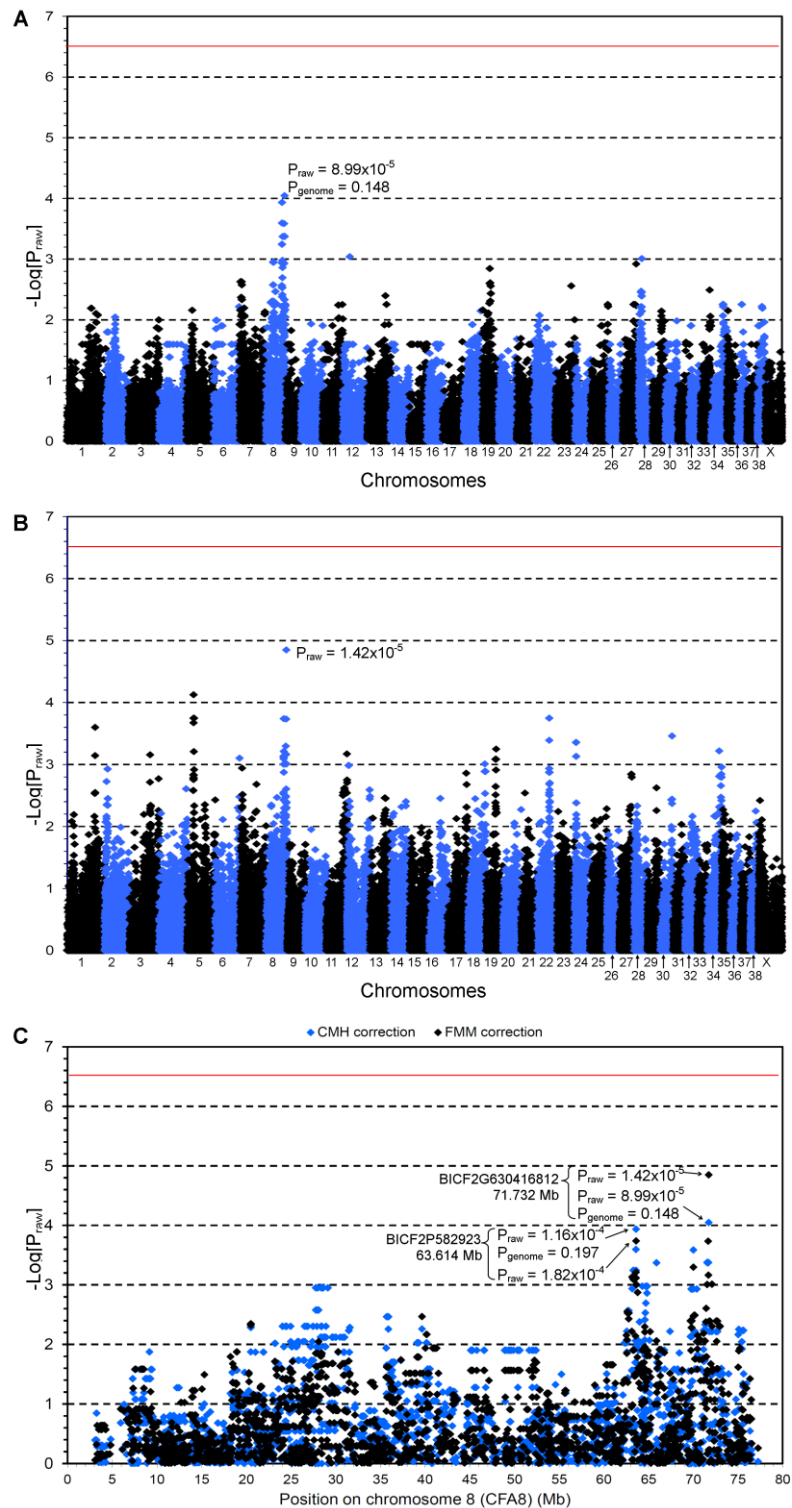


Figure 17. Genome-wide association mapping of PRA2 in GRs

$-\text{Log}_{10}$ of p-values after correction for population stratification, using two approaches. The red lines indicate the Bonferroni-corrected 5% significance level, based on 103,264 SNPs. A) The CMH meta-analysis approach shows the strongest signal on CFA8 ($p_{\text{raw}} = 8.99 \times 10^{-5}$, $p_{\text{genome}} = 0.148$). B) The FMM approach also shows the strongest signal on CFA8 ($p_{\text{raw}} = 1.42 \times 10^{-5}$). C) Two distinct signals can be seen at 63.614 Mb and 71.732 Mb.

3.6.2 Homozygosity and Haplotype Analysis

A haplotype at either locus 1 or locus 2 that was homozygous in all cases but not controls could not be identified through visual examination of SNP genotypes (Figure 18). BICF2P582923 ($p_{\text{genome}} = 0.197$) at locus 1 is homozygous (T/T) in 6/10 cases, but either heterozygous (T/C) or homozygous for the alternate allele (C/C) in all the control samples. Similarly, BICF2G630416812 ($p_{\text{genome}} = 0.148$) at locus 2 is homozygous (T/T) in 6/10 cases, but either heterozygous (T/C) or homozygous for the alternate allele (C/C) in all the control samples. The signals at loci 1 and 2 could not be clearly defined, with no obvious homozygous haplotype unique to affected individuals. However, a 437 kb region was identified upstream of locus one on CFA8 from 62.647 to 63.084 Mb, for which all of the PRA cases analysed were homozygous (“Homozygous Block”, Figure 18A). Five of the controls were also homozygous for the “affected” haplotype in this region. There was also a larger region, from 62.526 Mb to 63.263 Mb (“Affected Haplotype”, Figure 18A), over which the haplotypes of most of the cases are different from those of the controls at locus 1. Inferred phasing (section 2.3.2) of 21 SNP markers spanning this 737 kb region revealed the presence of six alleles, only one of which was homozygous in 8/10 cases but none of the controls (Haplotype number 1 in Figure 19). While it is possible this is the “PRA-2” allele, the possibility that the observed inheritance pattern of this haplotype is coincidental and the PRA-causing mutation is actually near BICF2P582923 or BICF2G630416812 could not be discounted. Therefore PRA critical regions were defined at these loci to ensure coverage of all of these elements and extend from 62.046 to 64.373 Mb and from 71.147 to 72.634 Mb, comprising regions of 2.327 and 1.487 Mb respectively, and 3.814 Mb in total. Critical region 1 contains 17 genes, 14 of which have human orthologues (Figure 18B and Appendix V). Spermatogenesis associated protein 7 (*SPATA7*) has been associated with juvenile aRRP and aRLCA¹²⁷, and tetratricopeptide repeat domain 8 (*TTC8*) with BBS and aRRP¹²⁸⁻¹³⁰. Critical region 2 contains 11 genes, all of which have human orthologues. None of these 11 genes have previously been associated with ocular function or retinal degeneration in humans. *SPATA7* and *TTC8* are both within the “Affected Haplotype” and were therefore both excellent candidate genes for further investigation. Due to the identification of two loci and two candidate genes, the critical regions surrounding both loci were selected for further investigation, specifically by targeted resequencing.

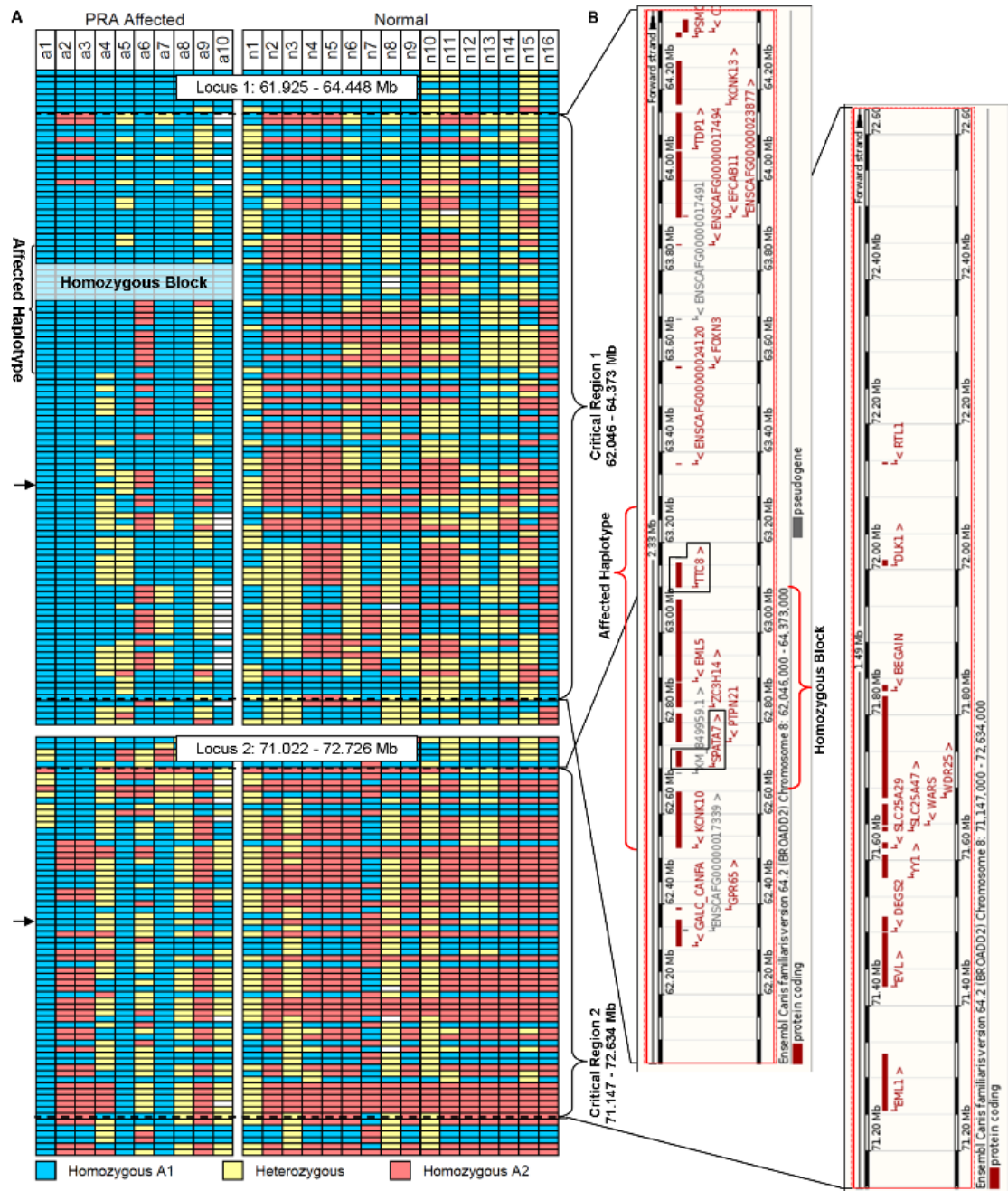


Figure 18. Homozygosity analysis of GRs with PRA2

A) SNP genotypes for 10 PRA cases and 16 PRA controls, over the two regions identified during the GWA study, encompassing loci 1 and 2. BICF2P582923 ($p_{\text{genome}} = 0.197$) is indicated with the upper arrow at locus one. BICF2G630416812 ($p_{\text{genome}} = 0.148$) is indicated with the lower arrow at locus 2. A homozygous block from 62.647 to 63.084 Mb and a larger homozygous “Affected Haplotype” are upstream of locus 1. Critical regions 1 and 2 encompass all of these elements. B) There are 17 and 11 genes found within critical regions 1 and 2 respectively, as annotated on CanFam2.0 (taken from Ensembl genome browser). *SPATA7* and *TTC8* were provocative candidate genes.

SNP name	Position	PRA Affected										Normal																
		a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	n1	n2	n3	n4	n5	n6	n7	n8	n9	n10	n11	n12	n13	n14	n15	n16	
BICF2G630416169	62,583,524	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
BICF2G630416178	62,600,578	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
BICF2G630416192	62,647,484	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
BICF2S23345277	62,659,928	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
BICF2P1456501	62,669,851	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
BICF2P745103	62,688,992	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
BICF2S23160703	62,780,858	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
BICF2S23016674	62,793,284	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
BICF2S23411200	62,897,000	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
BICF2S24410075	63,084,196	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
BICF2S2349547	63,116,670	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
BICF2S23045869	63,128,895	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
BICF2S2343538	63,139,903	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
BICF2P967520	63,155,610	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
BICF2P1468149	63,163,937	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
BICF2G630416233	63,191,730	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
BICF2G630416240	63,194,733	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
BICF2G630416251	63,211,145	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
BICF2G630416261	63,218,049	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
BICF2G630416275	63,231,285	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
BICF2G630416306	63,251,801	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
Haplotype Number		1	1	1	1	1	1	1	1	1	2	2	1	1	1	1	3	2	4	4	4	4	4	4	4	4	4	2

Figure 19. Haplotype analysis in GRs with PRA2.

SNP genotypes for 10 PRA cases and 16 PRA controls, over the 737 kb “Affected Haplotype” identified during the GWA study. Inferred phasing of the 21 SNP markers revealed six unique haplotypes, each of which was assigned a colour and number. Haplotype number 1 (yellow) is the “Affected Haplotype”.

3.7 Sequencing of PRA-2 Critical Regions

3.7.1 Target Enrichment of PRA-2 Critical regions

Custom-made RNA baits (section 2.4.2.1) were designed based on the CanFam2.0 reference sequence covering critical regions 1 (CFA8:62,045,807-64,372,549 bp) and 2 (CFA8:71,147,444-72,634,192 bp); regions 2.327 and 1.487 Mb in size respectively, and 3.814 Mb in total. As repeat masking was applied to avoid repetitive elements, baits were designed over 2.254 Mb, 59% of the total region with an average GC content of 46.9%.

Target enrichment of 10 gDNA samples (five cases, two obligate carriers and three controls) with the custom RNA baits was performed (section 2.4.2.2) and the captured libraries for each sample had concentrations ranging from 4.14 nM to 12.40 nM. Samples were pooled and the final library contained 27.7 μ L captured DNA at a concentration of 7.2 nM. While this is below the 10 nM recommended by the manufacturer, it was not possible to repeat the target enrichment. From this study we now know that libraries with concentrations <10 nM are sufficient for successful sequencing.

3.7.2 NGS Data Analysis – Target Capture Efficiency

More than 400 million reads were generated in total (section 2.4.2.5), across all 10 samples sequenced. Approximately 30% of these reads were PCR duplicates and therefore removed from further analysis. A total of >287 million reads remained, 97% of which mapped to CanFam2.0. Target capture efficiency is represented by the percentage of bases that map to the targeted regions, and was relatively similar between the 10 samples processed, with an average of 72.8% (Table 8). Approximately 74% of bases had >30x coverage and average read depth was 427x.

Table 8. Capture efficiency for PRA-2 targeted NGS

Sample no.	Clinical Status	Mean coverage of baited region	Bases mapped to targets (%)	Bases with >10X coverage (%)	Bases with >20X coverage (%)	Bases with >30X coverage (%)
1	Affected	423x	75.30	78.65	76.01	73.84
2	Affected	392x	72.73	77.27	74.48	72.21
3	Affected	441x	72.95	78.07	75.57	73.55
4	Affected	485x	74.25	77.76	75.39	73.47
5	Affected	404x	70.38	79.18	76.59	74.55
6	Carrier	388x	74.24	77.93	75.39	73.25
7	Carrier	460x	70.49	79.80	77.38	75.55
8	Unaffected	350x	74.32	77.67	74.97	72.73
9	Unaffected	427x	73.73	78.76	76.31	74.38
10	Unaffected	503x	70.04	79.12	76.84	75.06
	Average	427x	72.84	78.42	75.89	73.86

3.7.3 NGS Data Analysis – Variant Identification

More than 25,000 SNPs and 2,400 indels were identified when the NGS sequence was compared with the CanFam2.0 reference sequence (section 2.4.2.5). Of these, 658 SNPs and 168 indels segregated with the phenotype, of which only two variants were predicted to alter a protein product. One was a non-synonymous substitution in exon 12 of the *SPATA7* gene (CFA8: 62,735,867 bp; c.A1378G), resulting in a missense variant (p.Thr459Ala). The other provocative variant was a single base (adenine) deletion in an exon of *TTC8* (CFA8: 63,129,154; Figure 20).

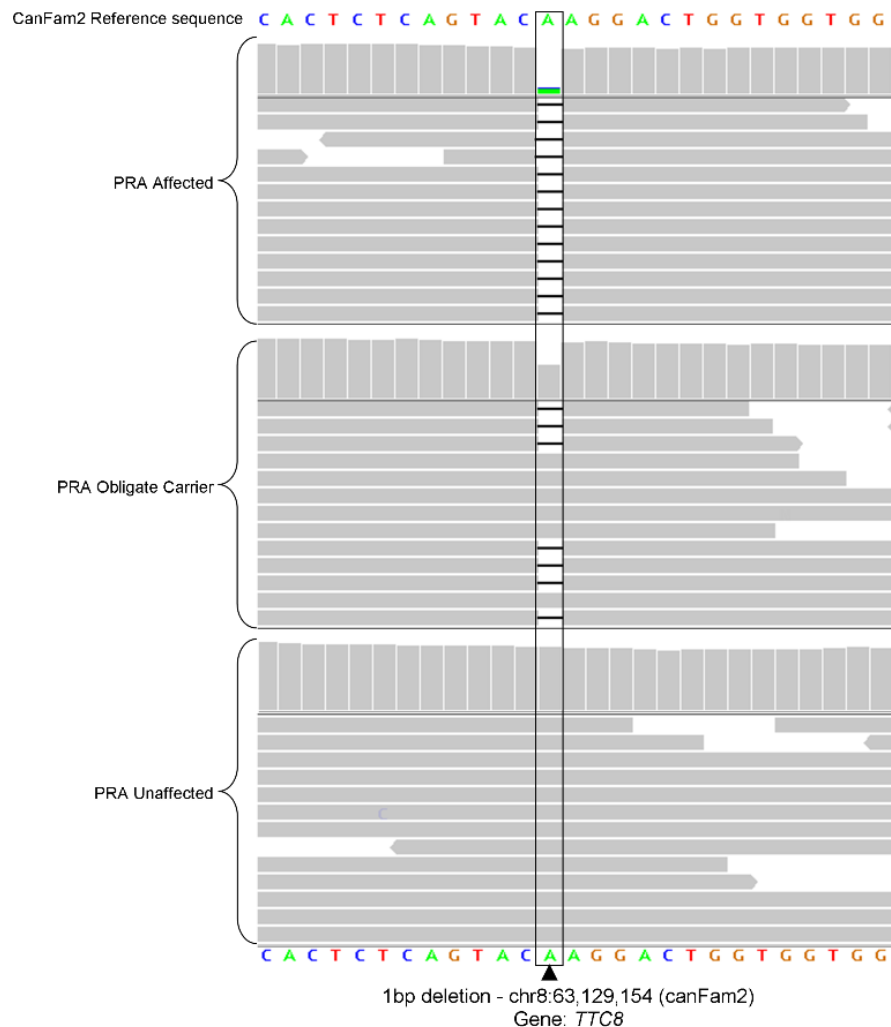


Figure 20. IGV display of the 1-bp deletion in *TTC8*

Each of the three samples (PRA-affected, obligate carrier and control) viewed in IGV are represented by two panels. The upper panel is a histogram where the height of each column is representative of the read depth at that location. The lower panel is a graphical view of a subset of the reads that align to that location. The “A” nucleotide indicated is absent in almost all reads in the PRA-affected sample, approximately half the reads in the obligate carrier sample and none of the reads in the PRA-unaffected (control) sample.

3.7.4 Characterising *SPATA7* and *TTC8* Retinal Transcripts

TTC8 occurs in three isoforms in humans. Two isoforms are created by alternative splicing of exon 5¹²⁸, the functional implications of which are unknown. Both isoforms are ubiquitously expressed and only the larger isoform (containing exon 5, *TTC8*) was therefore taken as the reference sequence. A third retinal isoform containing exon 2A (*TTC8*_{2A}) has also been identified¹³⁰. Only a single isoform of *SPATA7* has been reported¹³¹. As was the case with *SLC4A3* in section 3.5.1, several inconsistencies and possible errors were identified in the canine *SPATA7* and *TTC8* gene predictions (*SPATA7*: data not shown; *TTC8*: Figure 21).

All of the coding sequence of the *SPATA7* and *TTC8* retinal transcripts were successfully amplified and sequenced in a healthy dog (sections 2.2 and 2.4.1) revealing that both genes are transcribed in the canine retina. The presence of the *SPATA7* variant, c.A1378G, in the retinal mRNA transcript from a healthy dog as well as limited conservation in 35 eutherian mammals (data not shown) suggested the variant is unlikely to be pathogenic and it was therefore eliminated from further investigation. A *TTC8* isoform lacking exon 5 was not detected, however two isoforms of *TTC8* were found to be transcribed in the canine retina and intron-exon boundaries are identical to those of the human and mouse (Figure 21), which is in conflict with the boundaries predicted by Ensembl genebuild for the canine gene. One canine *TTC8* retinal isoform was found to contain exon 2A. Amplification of the full 5' and 3' UTRs was unsuccessful. Sequencing of the retinal mRNA transcripts of both isoforms revealed that canine *TTC8* (Genbank accession no JQ941743) encodes 505 amino acids and *TTC8*_{2A} (Genbank accession no JQ941742) encodes 515 amino acids, with predicted molecular weights of 57.571 kDa and 58.247 kDa respectively.

The exonic 1 bp deletion detected by NGS was a frame-shift variant, caused by the deletion of a single adenine in exon 8 (c.669delA, CFA8:63,129,154). It is predicted to cause a premature stop codon (p.Lys223ArgfsX15), possibly resulting in the degradation of mRNA by NMD or a truncated protein product (Figure 22). *TTC8*_{c.669delA} affects both isoforms of the protein.

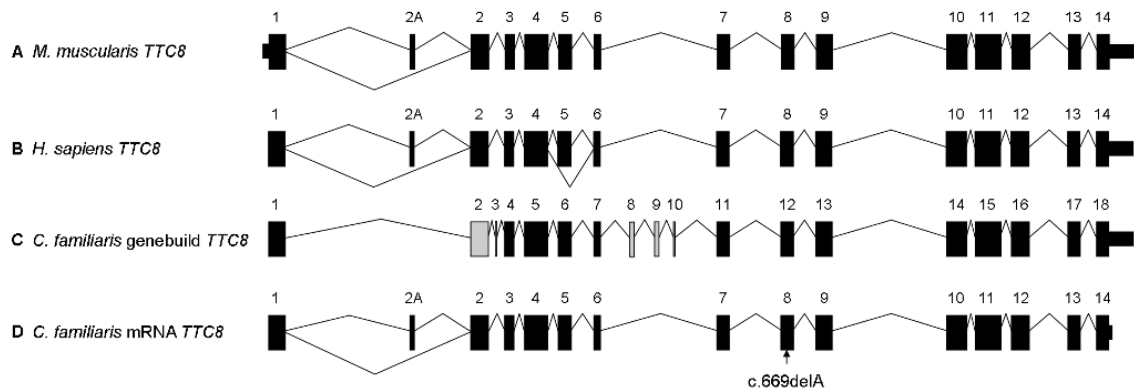


Figure 21. Graphical comparison of the intron-exon boundaries of TTC8.

A) Mouse (*Mus musculus*) *TTC8*. B) Human (*Homo sapiens*) *TTC8*. C) Canine (*Canis familiaris*) *TTC8* as predicted by Ensembl genebuild. Thirteen of the genebuild exons predicted are identical to the human and mouse exons (black). Exon 2 (grey) differs at the 3' intron-exon boundary. Exons 3, 8, 9 and 10 (grey) show no sequence or size similarity to their human or murine equivalents and are probably incorrect. D) Canine *TTC8* exons confirmed by sequencing of the retinal mRNA transcript. Exon 2A was not predicted by Ensembl genebuild. The location of the sequence variant is indicated.

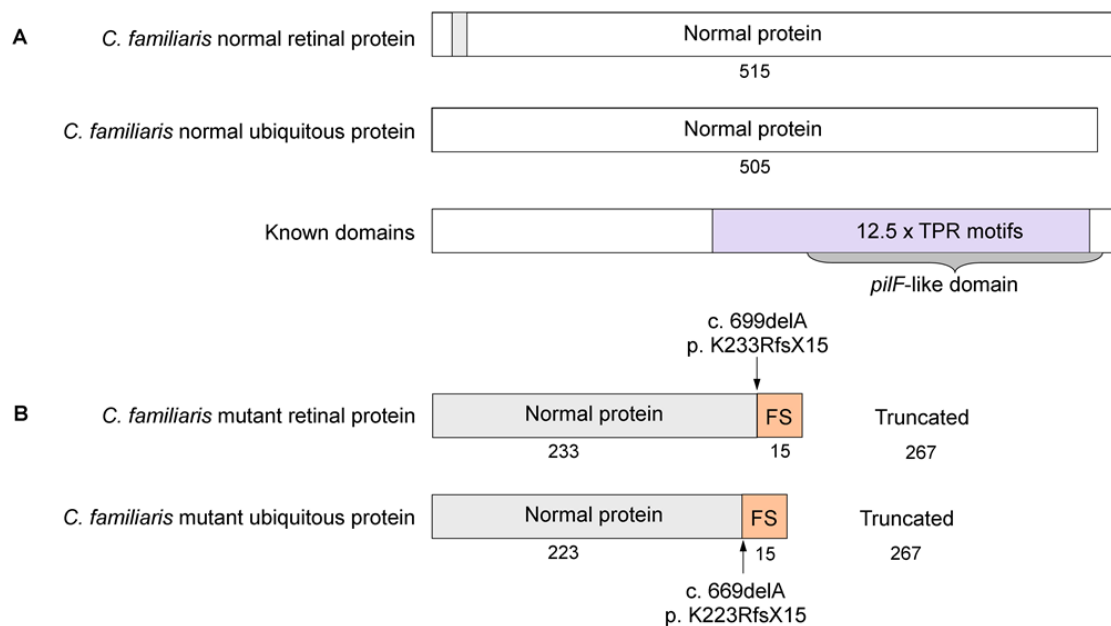


Figure 22. Effect of *TTC8*_{c.669delA} on the *TTC8* protein

A) The normal retina-specific and ubiquitous isoforms encode 515 and 505 amino acids respectively. The proteins differ only with the presence or absence of 10 amino acids unique to the retina-specific isoform (grey). The remainder of the protein is identical in both isoforms. B) Both isoforms are affected by the c.669delA mutation. 233 amino acids at the N-terminus of the retina-specific protein and 223 of the ubiquitous isoform are normal. However the deletion causes a shift in the reading frame of 15 amino acids, leading to a premature termination codon. This results in a truncated protein product, lacking 267 residues of the C-terminus of both isoforms. More than half of the protein is therefore absent.

3.7.5 *TTC8*_{c.669delA} Variant Validation

All 26 GR dogs (10 cases and 16 controls) that participated in the GWA study for the coding variant, *TTC8*_{c.669delA}, were then screened (section 2.4.1) to confirm the association of this variant with PRA and compare it with the most associated SNP markers, BICF2P5829231 and BICF2G630416812, without correcting for population stratification. *TTC8*_{c.669delA} showed significant allelic association with PRA ($P_{\text{raw}} = 6.31 \times 10^{-7}$, $P_{\text{genome}} = 0.019$) and was more strongly associated than BICF2P582923 ($P_{\text{raw}} = 5.79 \times 10^{-6}$, $P_{\text{genome}} = 0.109$) or BICF2G630416812 ($P_{\text{raw}} = 1.30 \times 10^{-6}$, $P_{\text{genome}} = 0.036$). Eight out of ten PRA cases and none of the controls were homozygous for *TTC8*_{c.669delA}. While the variant shows incomplete association with PRA, it has a strong likelihood of a deleterious effect on the protein. In addition, the nucleotide and the amino acid affected by *TTC8*_{c.669delA} is conserved in 32 eutherian mammals (data not shown). Analysis of the segregation of *TTC8*_{c.669delA} with PRA in a family (Figure 23) indicates that the form of PRA (known as PRA-2) associated with this variant is recessive and fully penetrant.

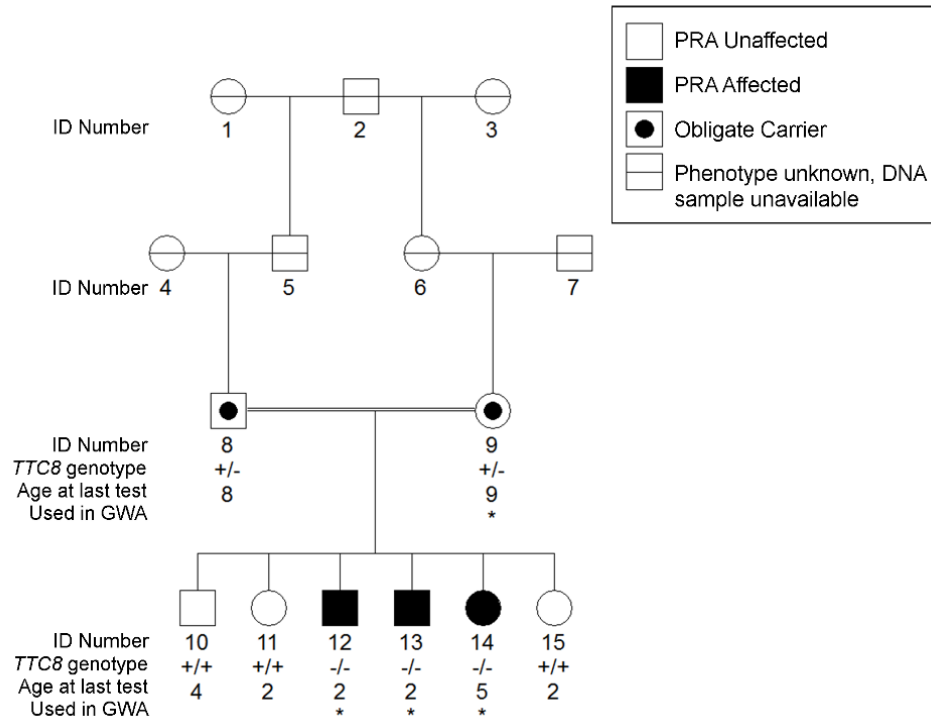


Figure 23. Segregation of PRA-2 in a GR pedigree

The segregation of PRA-2 is consistent with an autosomal recessive mode of inheritance with full penetrance. The PRA-2 variant allele is represented by “-” and the wildtype allele by “+”. Clinical information pertaining to and DNA samples from individuals 1-7 was unavailable. Samples used for the GWA study are indicated (*).

3.7.6 PRA-2 Phenotype Questionnaire

Mutations in *TTC8* have been associated with RP and BBS in humans¹²⁸⁻¹³⁰, but only one of the GRs in our cohort that was homozygous for *TTC8*_{c.669delA} was reported to have clinical signs other than PRA. These were “hormonal changes and thyroid problems”, the precise nature of which are unknown.

A questionnaire (Appendix VII) was sent to the owners of dogs with PRA-2 to assess the presence of systemic clinical signs in addition to PRA. Only two of these questionnaires were returned:

- 1) A bitch, diagnosed with PRA at the age of 5.9 years, was reported to be in good health. She was small for the breed, at 30kg was overweight for her size and gained weight easily. She had undergone metabolism tests, and while not treated, the results were considered borderline. The owner also reported a behavioural anomaly i.e. she had never been able to catch a ball in the air, which is unusual for a GR.
- 2) A dog, diagnosed with PRA at 2.9 years, was reported to be in good health until his death at 9 years, from cancer. At 52 cm tall, he was considered small for the breed (males typically range from 58 cm to 61 cm), but was not overweight. The owner also reported a behavioural anomaly i.e. he was aggressive towards certain people when in the home, but not outside. Interestingly, he also appeared to have a poor sense of smell, which is very unusual for a GR.

3.8 Population screening of *SLC4A3*_{c.2601_2602insC} and *TTC8*_{c.669delA}

3.8.1 Research Sample Cohorts

An additional 661 GR dogs were screened to confirm that the *SLC4A3*_{c.2601_2602insC} mutation is not a commonly occurring polymorphism in this breed, making a total of 709 GRs tested (Table 9). Of the confirmed PRA cases, 56.3% were homozygous for *SLC4A3*_{c.2601_2602insC} (*SLC4A3*^{-/-}) and 87.2% of obligate carriers were heterozygous (*SLC4A3*^{+/-}). The majority of dogs (99.6%) known to be clinically free of PRA were either carriers of the mutant allele (9.2%; *SLC4A3*^{+/-}) or homozygous for the wild type allele (90.4%; *SLC4A3*^{+/+}). These GRs were from a number of different countries including known PRA cases from the UK, Sweden, France and the USA (data not shown). The mutation described here accounts for 55.6% and 76.0% of PRA cases from the UK and Sweden respectively. However it accounts for only a minority of PRA cases from France and Finland, 12.5% and 33.3% respectively, and none of the cases from the USA.

Table 9. *SLC4A3* genotypes and PRA clinical status for 709 GRs

	PRA Affected	PRA Carrier	PRA Unaffected	Unknown	Total
<i>SLC4A3</i> ^{-/-}	45	1	2	0	48
	56.3%	2.6%	0.4%	0%	6.8%
<i>SLC4A3</i> ^{+/-}	8	34	47	11	100
	10.0%	87.2%	9.2%	13.9%	14.1%
<i>SLC4A3</i> ^{+/+}	27	4	462	68	561
	33.8%	10.3%	90.4%	86.1%	79.1%
Total	80	39	511	79	709

An additional 634 GR dogs were tested for *TTC8*_{c.669delA} to confirm that the mutation is also not a commonly occurring polymorphism in this breed, making a total of 660 GRs (Table 10). Of the 35 confirmed PRA cases used in the study 22 (62.9%) were homozygous for *TTC8*_{c.669delA} (*TTC8*^{-/-}) and all 513 dogs known to be clinically free of PRA at their last eye examination, including five obligate carriers of PRA, were either carriers of the mutant allele (3.9%; *TTC8*^{+/-}) or homozygous for the wild type allele

(96.1%; $TTC8^{+/+}$). PRA-2 accounts for the majority of cases of PRA that are not PRA-1 in all countries, excluding Canada (data not shown).

Table 10. $TTC8$ genotypes and PRA clinical status for 660 GRs

	PRA Affected	PRA Carrier	PRA Unaffected	Unknown	Total
$TTC8^{-/-}$	22	0	0	0	22
	62.9%	0.0%	0.0%	0.0%	3.3%
$TTC8^{+/-}$	1	2	18	4	25
	2.9%	40.0%	3.5%	3.6%	3.8%
$TTC8^{+/+}$	12	3	490	108	613
	34.3%	60.0%	96.5%	96.4%	92.9%
Total	35	5	508	112	660

The allele frequency of PRA-1 is 0.14 based on all 709 GRs submitted to the AHT for the research purposes (Table 11). This is not unusually high, even for a recessive disease allele. The allele associated with primary lens luxation in terrier breeds has similar frequencies in many terrier breeds¹³². The allele frequency of PRA-2 is 0.05, based on all 660 GRs submitted to the AHT for the research purposes (Table 11). This is also not unusually high for an autosomal recessive disease allele in dogs. The GR cohorts screened contained samples collected for the purposes of research, including all available samples from dogs clinically affected with PRA, and samples from PRA-unaffected dogs closely related to affected dogs and to one another. This cohort was therefore subject to considerable sampling bias. This, combined with the relatively small number of samples screened, has resulted in a cohort that does not conform to HWE (Table 11). A subset of the GR cohort that was unrelated at the parent level and from the UK, was selected for allele frequency estimations (n=88, Table 11, collected and genotyped by Bryan McLaughlin and Lou Hayward, AHT). These samples were collected for the purpose of estimating disease allele frequencies in the breed, and as such clinical history pertaining to these dogs is unknown. Analysis of PRA-1 and PRA-2 using only this cohort resulted in reduced allele frequencies, and this cohort does conform to HWE (Table 11). This suggests that the high degree of relatedness between samples and sampling bias are a major cause of deviation from HWE.

Table 11. GR population screening for PRA-1 and PRA-2

Breed	Observed Genotypes*				HWE		1-in-__ Expected		
	-/-	-/+	+/+	TOTAL	Allele Freq	χ^2	P	Affected	Carrier
Research Cohorts									
PRA-1	48	100	561	709	0.138	118.00	1.73×10^{-27}	52	4
PRA-1 UK (Unrelated)	0	7	81	88	0.040	0.15	0.70	632	13
PRA-2	22	25	613	660	0.052	251.82	1.04×10^{-56}	366	10
PRA-2 UK (Unrelated)	0	2	86	88	0.011	0.01	0.91	7744	45
AHT Genetic Services Cohorts									
PRA-1	2	83	770	855	0.051	0.02	0.88	386	10
PRA-2	0	27	388	415	0.033	0.47	0.49	945	16

* The mutant allele is represented by “-” and the wildtype allele by “+”.

To determine whether *SLC4A3*_{c.2601_2602insC} is present in healthy dogs of other breeds or is causal for PRA in other breeds, a cohort/sample set of 273 dogs from three closely related breeds made up of PRA cases of unknown aetiology and unaffected dogs was screened: 132 LR (16 cases), 92 FCR, 49 CBR (two cases). All 273 dogs were homozygous wild-type (*SLC4A3*^{+/+}).

Similarly, to determine whether *TTC8*_{c.669delA} is present in healthy dogs of other breeds or is causal for PRA in other breeds, a cohort/sample set of 174 dogs from three closely related breeds made up of PRA cases of unknown aetiology and unaffected dogs was screened: 70 LR (17 cases), 59 FCR, 48 CBR (two cases). One LR with PRA was homozygous for the variant (*TTC8*^{-/-}). The remaining 173 dogs were homozygous wild type (*TTC8*^{+/+}).

3.8.2 AHT Genetic Services Cohorts

The DNA test for PRA-1 was made available commercially on 15 November 2010. Over a period of 26 months, 855 GRs from 24 countries were genotyped by the AHT Genetic Services (Table 11). Overall, two were *SLC4A3*^{-/-} (0.2%) and 83 were *SLC4A3*^{+/-} (10%), indicating an overall allele frequency of 0.05. Similarly, the DNA test for PRA-2 was made available commercially on 13 August 2012. Over a period of six months, 415 GRs from 19 countries were genotyped by the AHT Genetic Services (Table 11). Overall, none were *TTC8*^{-/-} and 27 were *TTC8*^{+/-} (6.5%), indicating an

overall allele frequency of 0.03. These larger and more diverse AHT Genetic Services cohorts did conform to HWE and are therefore likely to be more representative of the general population than the research cohorts.

The allele frequency of *SLC4A3*_{c.2601_2602insC} in each country was calculated for those represented by at least 20 samples (Table 12) and range from 0.01 to 0.15. Each cohort represented by >20 samples conforms to HWE. The allele frequency of 0.115 in dogs from the USA is surprising as none of the PRA cases from the USA were found to be homozygous for *SLC4A3*_{c.2601_2602insC}. This may be due to the relatively small number of samples tested (n = 26), and only time will tell if this estimate is inflated. The allele frequency of 0.149 in Denmark is also high, although could be explained by the geographical proximity to Sweden, where the allele frequency of PRA-1 is approximately 9.3% (personal communication, Tomas Bergström).

Table 12. PRA-1 allele frequency breakdown by country (n>19)

Country	Observed Genotypes*					HWE		1-in-__ expected		Region [†]
	-/-	-/+	+/+	TOTAL	Allele Freq	χ^2	P	Affected	Carrier	
France	0	8	56	64	0.063	0.28	0.59	256	9	1
Germany	0	5	80	85	0.029	0.08	0.78	1156	18	1
Spain	0	1	47	48	0.010	0.01	0.94	N/A	N/A	2
Denmark	0	11	26	37	0.149	1.13	0.29	45	4	3
Finland	0	4	33	37	0.054	0.12	0.73	342	10	3
Norway	0	5	37	42	0.060	0.17	0.68	282	9	3
UK	2	32	367	401	0.045	1.93	0.17	496	12	3
Canada	0	2	20	22	0.045	0.05	0.82	484	12	4
USA	0	6	20	26	0.115	0.44	0.51	75	5	4
Total	2	74	686	762	0.051	8.63x10⁻⁶	1.00	382	10	

* The mutant allele is represented by “-“ and the wildtype allele by “+”.

† 1 = Western Europe, 2 = Southern Europe, 3 = Northern Europe, 4 = North America

Due to the relatively short period of time for which the PRA-2 DNA test has been available, the numbers of dogs tested from individual countries is still small. Therefore, the allele frequency of $TTC8_{c.669delA}$ for each country was calculated for those represented by at least 10 samples (Table 13) and range from 0 to 0.15. Each cohort represented by >10 samples conforms to HWE. The allele frequency of 0.15 in dogs from Austria is surprising as none of the PRA cases are from Austria. This may be due to the relatively small number of samples tested ($n = 10$), and only time will tell if this estimate is inflated.

Table 13. PRA-2 allele frequency breakdown by country (n>9)

Country	Observed Genotypes*					HWE		1-in-__ expected		Region†
	-/-	-/+	+/+	TOTAL	Allele Freq	χ^2	P	Affected	Carrier	
USA	0	1	23	24	0.021	0.01	0.92	2304	25	1
Finland	0	1	10	11	0.045	0.02	0.87	484	12	2
UK	0	13	275	288	0.023	0.15	0.70	1963	23	2
Italy	0	0	17	17	0.000	N/A	N/A	N/A	N/A	3
Spain	0	2	10	12	0.083	0.10	0.75	144	7	3
Austria	0	3	7	10	0.150	0.31	0.58	44	4	4
France	0	2	13	15	0.067	0.08	0.78	225	8	4
Total	0	22	355	377	0.029	0.34	0.56	1175	18	

* The mutant allele is represented by “-“ and the wildtype allele by “+”.

† 1 = North America, 2 = Northern Europe, 3 = Southern Europe, 4 = Western Europe

3.9 Screening of *SLC4A3* in humans

3.9.1 Introduction and Background

As discussed in Chapter 1, while 185 of the genes that contain mutations that cause retinal degeneration in humans have been identified, it is estimated that the genes involved in approximately 35% of aR retinal dystrophies remain unknown⁵⁷. While PRA is widely considered to be the veterinary equivalent of RP, the limited characterisation of many forms of PRA at the cellular and molecular level suggests it may, in some cases, be the equivalent of other retinal dystrophies. Alvarez et al. identified *SLC4A3* as a candidate gene for human vitreoretinal degenerations based on their findings of blindness and retinal degeneration in knockout mice¹²⁵. However, prior to the identification of *SLC4A3*_{c.2601_2602insC} in GRs with PRA in this study, *SLC4A3* had not been associated with naturally occurring, spontaneous ocular disease in any species. The identification of the variant therefore endorses the status of *SLC4A3* as a candidate gene for human retinal degenerations. The *SLC4A3* gene was therefore screened in DNA from human patients with retinal degeneration of unknown aetiology in order to determine whether mutations in the gene cause retinal degeneration.

3.9.2 Sample and Assay Selection

Little is known about the phenotype associated with PRA-1 in GRs beyond ophthalmoscopic observations. Clinically the fundus appears identical to other forms of PRA and histological and detailed ERG analysis has not been reported. In addition the age of onset is difficult to define, although the age at which dogs with PRA-1 are diagnosed is approximately 6-7 years. This is suggestive of a relatively late-onset condition, a hypothesis consistent with the findings in the *SLC4A3* knockout mouse in which a selective inner retina defect is followed by photoreceptor degeneration at eight months¹²⁵. As six of the 10 genes already implicated in both canine and human progressive retinal disease are involved in RP (see Chapter 1 Table 2), *SLC4A3* is a strong candidate for humans with RP. A cohort of 192 DNA samples from individuals with aRRP of unknown aetiology was therefore selected for screening.

Eight additional human patients with various forms of retinal degeneration caused by unknown variants were also selected for screening. In each patient, autozygosity mapping previously conducted at the Institute of Ophthalmology had identified large regions that potentially harboured disease-causing variants (personal communication,

Alice Davidson). These regions contained *SLC4A3*, amongst other genes (Table 14), and these patients were therefore prioritised and screened first.

A number of DNA screening methods were considered as a cost-effective alternative to Sanger sequencing, particularly high-resolution melting (HRM) analysis. This commonly used technique identifies melting changes of a DNA duplex caused by variation in sequence. As with any screening methods, however, additional sequencing is required to identify and confirm specific variants. More than 5,000 variations have been identified in human *SLC4A3*, which appear to be evenly spread across the gene (data not shown). As a result it is likely that HRM analysis would identify variations in most, if not all exons that would require subsequent sequencing. In addition, due to the number of exons involved ($n = 25$), a great deal of HRM optimisation would be required. Therefore screening of all exons of the gene by Sanger sequencing was considered the least costly and time-consuming approach.

Table 14. Patients with mapped homozygous regions containing *SLC4A3*

Patient ID	Disease	Region containing <i>SLC4A3</i>			Other regions identified		
		Size (Mb)	No of genes	Retinal genes*	Size (Mb)	No of genes	Retinal genes*
1	aRRP	28.78	193	3	19.74	74	0
2	aRRP	19.24	158	4	88.69	503	5
3	Rod monochromasy	9.42	39	1	280.70	2218	21
4	Recessive retinal dystrophy	22.59	182	3	222.93	2010	22
5	aRRP	36.36	271	3	94.87	1271	20
6	aRRP	13.91	99	1	172.98	1334	17
7	aRRP (pseudo dominant)	14.16	93	1	369.70	1774	20
8	cone-rod dystrophy	16.07	102	3	34.38	281	1

* Genes previously associated with retinal disease in humans

As is the case in the dog, two main isoforms of *SLC4A3* have been described in humans: a full-length (*SLC4A3_{fl}*) isoform made up of 23 exons and a cardiac (*SLC4A3_c*) isoform made up of 18 exons¹³³. These differ at the N-terminus but have identical C-terminal regions (Figure 24). Alternative splicing of exon six of *SLC4A3_{fl}* results in two “full-length” isoforms that differ by 81 bp, or 27 amino acids. *SLC4A3_{fl1}* (Transcript ID: ENST00000358055; Accession no: NM_005070) is the shorter version and results in a protein made up of 1,232 amino acids. *SLC4A3_{fl2}* (Transcript ID: ENST00000273063; Accession no: NM_201574) results in a protein made up of 1,259 amino acids (Figure 24), and this is the isoform referred to throughout the remainder of this document. In *SLC4A3_c* exon C1 replaces exons one to six of the full-length transcript, resulting in a shorter protein product made up of 1,034 amino acids.

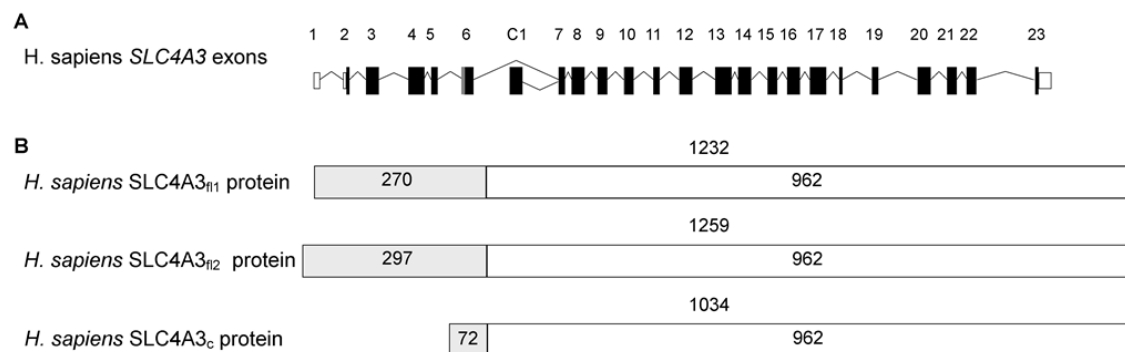


Figure 24. Human *SLC4A3* gene and protein isoforms

A) The three isoforms of *SLC3A3* are created by alternative splicing of exons 1-6 and C1, and by alternative splicing of exon 6 (grey). B) *SLC4A3_{fl1}* and *SLC4A3_{fl2}* proteins are created from alternative splicing of exon 6 and differ only with regards to the absence and presence respectively of 27 amino acids. *SLC4A3_c* is created by alternative splicing of exon C1 and the first 72 amino acids are completely different than the full-length isoforms (grey). All three isoforms are identical over the 962 amino acids at the C-terminal end.

3.9.3 Variants identified

Sequencing of all known exons and intron-exon boundaries of human *SLC4A3* (section 2.2 and 2.7.2) in 200 patients revealed 50 SNP variants affecting at least one individual when compared with the human reference sequence (hg19, data not shown). Of these, 21 are located in exons, 23 in introns, four in the 5'-UTR and two in the 3'-UTR. The majority (n=39) have previously been identified and have entries in the dbSNP database. Variants that could be pathogenic were retained for further investigation, i.e. those that may affect exon splicing or result in amino acid changes that are predicted to be pathogenic by at least one of the SNP pathogenicity prediction tools utilised.

After elimination of variants unlikely to be pathogenic from the investigation five SNP variants remained that were predicted to affect exon splicing and/or change the amino acid sequence of the protein (Figure 25 and Table 15). Fourteen patients in total, all of which come from the panel of 192 aRRP cases, carried one of these five variants.

SNPs_1-4 are non-synonymous SNPs that result in amino acid changes that are predicted to be pathogenic by at least one of the in silico prediction tools (PolyPhen, SIFT and PMut). SNP_4 and SNP_5 are predicted to affect exon splicing and due to their location in exons 22 and 18 respectively would affect all three isoforms of the protein. If the predictions are accurate, SNP_4 (c.G3674A, p.R1225Q) could introduce a new acceptor site. This would result in a 147 bp, in-frame deletion from exon 22, decreasing the size of the exon from 174 to 27 bp. This would, in turn, produce a deletion of 49 amino acids near the C-terminus of the protein, resulting in a 1,210 amino acid protein (Figure 26B) with the first 1,176 and last 34 amino acids identical to the wildtype protein. SNP_5 (c.G2865A, p.G955G) could remove an existing acceptor site, which would result in a 180 bp increase in the size of exon 18, from 90 to 270 bp. The first 949 amino acids of the resulting protein would be normal, followed by 41 different amino acids and a premature termination codon, resulting in a loss of 269 amino acids (Figure 26C).

These five variants all have potentially pathogenic, loss-of-function effects on the SLC4A3 protein. Only the A allele of SNP_1 was present in the homozygous state in a single sample as well as in the heterozygous state in an additional four samples. However, the presence of this allele in the homozygous state in the 1000 genomes and EVS populations suggests it is highly unlikely to be disease causing. SNPs_2-5 were all only identified in the heterozygous state. In addition no patients carried more than one variant (i.e. not compound heterozygotes).

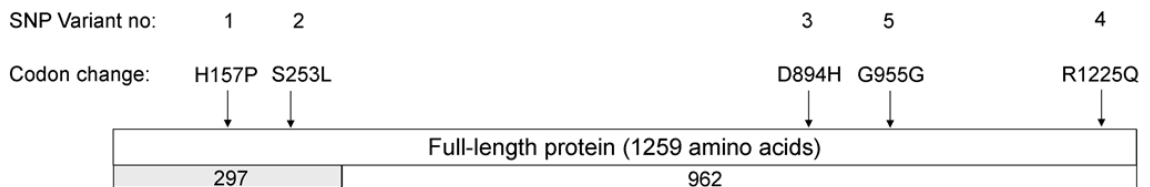
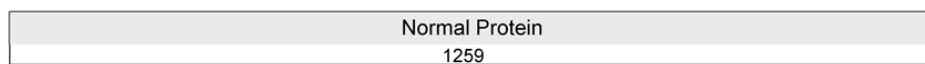


Figure 25. Effects of potentially pathogenic SNP variants on SLC4A3 protein

The number of amino acids that make up each protein domain are indicated. All five SNP variants affect the full-length isoform of SLC4A3, with two occurring in the first part of the N-terminal cytoplasmic domain unique to the full-length isoform (grey). Only three variants affect the cardiac and full-length isoforms.

A Wildtype protein



B SNP_4



C SNP_5

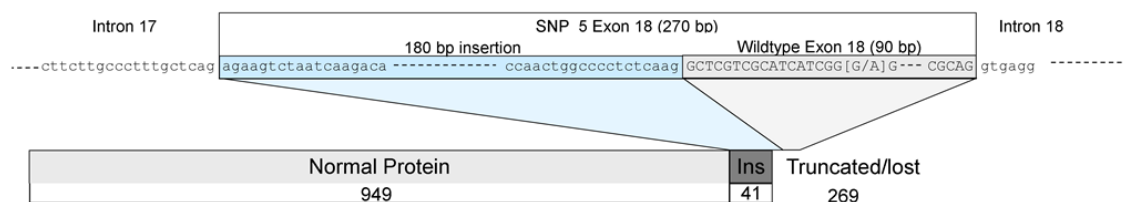


Figure 26. Predicted effects of two splice-site variants on the SLC4A3 protein.

Graphical representation of the wildtype human SLC4A3 protein and the predicted effects of splicing variants on the protein. The numbers of amino acids are indicated, and the nucleotides affected by the variants are bracketed. A) Normal full-length (1,259 amino acids) protein. B) SNP_4 could introduce an acceptor site resulting in the reduction in the size of exon 22 from 174 to 27 bp. This equates to an in-frame deletion of 49 amino acids near the C-terminus of the protein, and a final protein 1,210 amino acids in size. C) SNP_5 could remove an acceptor site resulting in an increase in the size of exon 18 from 90 to 270 bp, which equates to an insertion of 41 amino acids (Ins), a premature termination codon, and the loss of 269 amino acids at the C-terminus.

Table 15. Potentially pathogenic SNP variants

Variants								200 Patient Panel				1,000 Genomes Data (n=1,092)				EVS Data (n≈6,500)			
SNP _#	Genomic Location	A1*	A2†	Possible effect on splice sites	Codon change	rs Number	Pathogenic by‡	A1/A1	A2/A1	A2/A2	A2 Freq	A1/A1	A2/A1	A2/A2	A2 Freq	A1/A1	A2/A1	A2/A2	A2 Freq
1§	2:220494118	C	A		H157P	rs597306	PP	195	4	1	0.015	1013	74	5	0.038	5891	552	54	0.051
2	2:220494940	C	T		S226L	rs36068948	PP	199	1	0	0.003	1084	8	0	0.004	6453	48	0	0.004
3	2:220502366	G	C		D894H	rs61729101	PP, S	194	6	0	0.015	1079	13	0	0.006	6357	146	0	0.011
4	2:220505656	G	A	Introduction of acceptor site (0.00-0.81)	R1225Q	rs150952379	PP, S, PM	199	1	0	0.003	1091	1	0	0.000	6496	7	0	0.001
5	2:220502903	G	A	Loss of acceptor site (0.45-0.00)	G955G			199	1	0	0.003								

* Allele 1 = Reference allele

† Allele 2 = Minor/variant allele

‡ PP = PolyPhen; S = SIFT; PM = PMut

§ Allele 2 (A) is actually the reference allele, while allele 1 (C) is the ancestral allele

3.10 Discussion

3.10.1 PRA-1 and *SLC4A3*

Using a GWA mapping approach, a novel locus on chromosome 37 that is statistically associated with PRA in the GR breed ($p_{\text{genome}} = 1.0 \times 10^{-5}$) has been identified. A critical region of 644 kb, formed by two overlapping homozygous blocks, was defined through homozygosity analysis using the SNP genotypes. Most of the cases (23/27) and none of the controls were homozygous for one of the homozygous blocks.

SLC4A3 was identified as a strong candidate causal locus within the PRA-critical region as it encodes the anion exchanger 3 (AE3) protein, which mediates $\text{Cl}^-/\text{HCO}_3^-$ exchange across cellular membranes¹³⁴ and is expressed in various tissues including the Müller and horizontal cells of the retina¹²⁴. Interestingly, the effect of AE3-deficiency has been described in a knockout mouse model showing that a selective inner retina defect is followed by photoreceptor degeneration¹²⁵, similar to the retinal degeneration seen in many forms of PRA and RP.

SLC4A3 occurs in two main isoforms, *SLC4A3_{fl}* and *SLC4A3_c*, which have identical transmembrane and C-terminal regions while the N-terminal amino acids differ (Figure 15)^{124,126}. *SLC4A3_{fl}* is expressed predominantly in the brain, but it is also found in the gut, kidney, heart and Müller cells of the retina^{124,133-136}. The cardiac isoform is expressed predominantly in the heart and also in the horizontal neurons of the retina^{124,134,136}.

Sequencing of *SLC4A3* transcripts from healthy retinal mRNA served three purposes. Firstly it confirmed the presence of “cardiac” and “full-length” mRNA transcripts in the normal canine retina. Secondly it revealed that the intron-exon boundaries predicted by genebuild for the dog were incorrect for five of the first eight exons. They are instead identical to the human and mouse boundaries. Thirdly it revealed exons orthologous to part or all of human exons 1, 2, 3 and 8, that were absent from the Ensembl canine predictions (Figure 14). Two large gaps in the reference sequence were found upstream of canine exons 2 and 8, and it is very likely that the missing exons lie within these gaps. As is the case in humans and mice, canine *SLC4A3_{fl}* is alternatively spliced to produce two “full length” isoforms (*SLC4A3_{fl1}* and *SLC4A3_{fl2}*). The functional implication of the difference between these isoforms is currently unknown.

The two exonic and nonsynonymous variants identified in this study (*SLC4A3_{c.A44C}* and *SLC4A3_{c.2601_2602insC}*) are in tight linkage disequilibrium for the 25 dogs screened. In two

dogs (one case and one control) however, variant alleles are not inherited together. *SLC4A3*_{c.2601_2602insC} is more strongly associated with PRA than *SLC4A3*_{c.A44C} or BICF2G630131493, and it appears to be fully penetrant.

In order to further test the validity of the insertion mutation, 709 GRs were screened for the insertion (Table 9). The majority, 56.3%, of the PRA cases, 87.2% of the obligate PRA carriers and 99.6% of clinically unaffected dogs (which could be clear of the mutation or carry a single copy) had *SLC4A3* genotypes that are concordant with their clinical status. There were two groups of dogs with genotypes discordant with their phenotypes. The first comprises three PRA-unaffected dogs that were *SLC4A3*^{-/-} (one of which is an obligate carrier). All three dogs were examined by a veterinary ophthalmologist at 4-5 years of age and as average age of onset is approximately seven years they would not necessarily have been displaying clinical signs at the time of their examination. Thus 93.8% of dogs (45/48) homozygous for the mutation (*SLC4A3*^{-/-}) have developed PRA, suggesting that the mutation is fully penetrant, or nearly so. The inheritance observed in a pedigree of 32 dogs (18 cases) is supportive of a recessive mode (Figure 16). The second group of discordant dogs comprised 35 PRA-affected dogs that were not homozygous for *SLC4A3*_{c.2601_2602insC} and four obligate carriers that did not carry *SLC4A3*_{c.2601_2602insC}. It is formally possible that the mutation has a dominant mode of inheritance with incomplete penetrance, or complex trait or compound heterozygote effects. However, the PRA cases discordant with *SLC4A3*_{c.2601_2602insC} segregation (*SLC4A3*^{+/+} and *SLC4A3*^{+/-}) tended to develop PRA at an earlier age, with an average age at diagnosis of 5.27 years (data not shown), which is consistent with the segregation of a third form of PRA in the GR breed (PRA-2). It is becoming increasingly apparent that even in breeds with small effective population sizes, mendelian conditions may be genetically heterogeneous. For example day blindness in the Alaskan Malamute breed is caused by two or more mutations, only one of which is currently known¹³⁷. These discordant cases were therefore used to undertake a subsequent study (sections 3.6 and 3.7) in which a third locus and mutation were identified. Due to the limited knowledge regarding the specific clinical pathology of the PRA form associated with *SLC4A3*_{c.2601_2602insC} it was termed PRA-1. None of the affected GR dogs are known to have any cardiac or neurological abnormalities. This is consistent with previous findings that loss of AE3 alone is insufficient to cause cardiac or neurological disease in mice^{138,139}.

The AE3 protein is part of a group of Na⁺-independent anion exchangers and it exchanges chloride for bicarbonate^{134,140}. Little is known about the structure of the protein, but it is thought to be similar to family member SLC4A1 (AE1). SLC4 proteins are made up of three structural domains (Figure 15). At the N-terminus there is a hydrophilic, cytoplasmic domain of between 400 and 700 amino acids, followed by a hydrophobic, polytopic transmembrane domain of approximately 500 amino acids, comprising up to 14 transmembrane spans, and lastly a cytoplasmic domain of between 30 and 100 amino acids at the C-terminal end^{133,135}.

The mutation, c.2601_2602insC (p.E868RfsX104) in *SLC4A3fl* and c.1981_1982insC (p.E661RfsX104) in *SLC4A3c*, occurs in the transmembrane region and as a result is likely to affect both full length and cardiac isoforms of the protein (Figure 15). The shift in the reading frame results in an incorrect sequence of 104 amino acids and the loss of 266 residues at the C-terminal end of the protein. These residues constitute a large portion of the transmembrane region and the complete C-terminal cytoplasmic region. Retinal tissue from a dog affected with PRA was not available and *SLC4A3* mRNA could not be isolated from available tissues (blood or buccal cells); therefore confirmation of whether the protein is expressed in a truncated form or not expressed at all due to NMD of the mRNA transcript has not been possible. In the event of truncated protein expression, functionally important residues are lost. Glycosylphosphatidylinositol-linked carbonic anhydrase IV (CAIV)¹⁴¹ and transmembrane carbonic anhydrase IX (CAIX) interact with exofacial portions of the transmembrane domains of AE3¹⁴²; the C-terminal tail of the SLC4 polypeptides contains acidic motifs that may serve as cytoplasmic carbonic anhydrase II (CAII) binding sites^{143,144}. Coexpression of each of the CA proteins with AE3 results in increased HCO₃⁻ transport activity. A number of possible functions for *SLC4A3fl* in Müller cells have been proposed:

- 1) Intracellular CO₂ released by photoreceptors is converted into HCO₃⁻ and H⁺ by carbonic anhydrases. It is thought that removal of these products into the blood or vitreous is facilitated by transporters such as *SLC4A3* in the Müller cell end feet¹⁴⁵.
- 2) *SLC4A3fl* may facilitate the exchange of intracellular HCO₃⁻ for extracellular Cl⁻, thereby contributing to pH maintenance¹²⁴ and a lack of bicarbonate in the retina results in loss of retinal function¹⁴⁶.

- 3) Finally it may play a role in Cl^- level maintenance which potentially affects Cl^- -dependent transporter activity e.g. transporters of γ -Aminobutyric acid (GABA) and taurine¹⁴⁷.

Horizontal cells also express various carbonic anhydrases, which could in turn create significant levels of intracellular HCO_3^- . By exchanging this HCO_3^- with extracellular Cl^- , *SLC4A3_c* may contribute to pH_i homeostasis¹²⁴.

PRA caused by the mutation described here has an average age at diagnosis of 6.67 years and this is indicative of a late age of onset (data not shown). Two affected individuals that were diagnosed at the relatively young age of four years (individuals 31 and 32 in Figure 16) were examined at the same time as a number of relatives including their mother. In these two dogs the disease was at an early stage. The late age at diagnosis corresponds to the late-onset findings of Alvarez et al. where *SLC4A3* knockout mice appeared completely normal at four months of age¹²⁵. However on closer investigation the mice were observed to be affected by an inner retinal defect from birth, which led to a failure of phototransduction at four months, followed by pathological signs of photoreceptor degeneration at eight months and eventually complete blindness. Pathological signs included defects in retinal blood vessels, the optic nerve head and rod bipolar cells, as well as reduced a-wave ERG amplitudes and b-wave depression¹²⁵, consistent with most forms of PRA. *SLC4A3* deficiency also resulted in increased expression of CAII and CAXIV enzymes as well as the $\text{Na}^+/\text{HCO}_3^-$ co-transporter, NBC1, in the horizontal and Müller cells. The authors hypothesised that while these changes possibly compensated for the loss of *SLC4A3* to some extent, the ability of the cells to maintain acid-base balance is still compromised, leading to late-onset photoreceptor cell death¹²⁵.

3.10.2 PRA-2 and *TTC8*

Using a GWA mapping approach, a novel locus on chromosome 8 that is associated, albeit not with genome wide significance, with PRA in the GR breed ($p_{\text{genome}} = 0.148$) was identified. An “Affected Haplotype” of 737 kb was defined through homozygosity and haplotype analysis using the SNP genotypes. Most of the cases (8/10) and none of the controls were homozygous for the “Affected Haplotype”.

SPATA7 and *TTC8* were identified as candidate causal genes due to their position within the PRA critical region, their expression in the retina and their association with human retinal disease. *SPATA7* was first identified in human spermatocytes and is thought to be involved in chromatin preparation for the initiation of meiotic recombination¹³¹. In

addition the protein is expressed in the retina and shows a uniform distribution in the inner segment cytoplasm. Nonsense and loss-of-function variants have been identified in patients with juvenile recessive RP and LCA from distinct ethnic groups¹²⁷. *TTC8* (a.k.a. *BBS8*) is one of a number of BBS proteins that likely function in membrane trafficking to the primary cilium¹⁴⁸. Homozygous and heterozygous variants have been identified in a small number of BBS families^{128,129} and a homozygous variant in the splice acceptor site of a retina-specific exon has been found to cause aRRP¹³⁰.

Sequencing of *SPATA7* and *TTC8* transcripts from healthy retinal mRNA served four purposes. Firstly, it confirmed the presence of both mRNA transcripts in the normal canine retina. Secondly, the presence of the *SPATA7*_{c.A1378G} variant in mRNA from a healthy dog allowed the elimination of this variant from further investigation. Thirdly, it revealed that the intron-exon boundaries predicted by genebuild for *TTC8* in the dog were incorrect for five exons. They are instead identical to the human and mouse boundaries. Finally, it revealed an exon orthologous to human exon 2A, that is absent from the Ensembl canine prediction (Figure 21). As is the case in humans and mice, canine *TTC8* is alternatively spliced to produce two isoforms (*TTC8* and *TTC8*_{2A}). The precise functional difference between the two isoforms is unknown, but it is thought *TTC8*_{2A} plays an important role in the function of the protein in the photoreceptor cell-containing outer nuclear layer of the retina¹³⁰.

In order to further test the validity of the deletion mutation, *TTC8*_{c.669delA}, 660 GRs were screened for the mutation (Table 10). This analysis revealed that 62.9% of the PRA cases, 40% of the obligate PRA carriers and 100% of clinically unaffected dogs (which could be clear of the variant or carry a single copy) have *TTC8* genotypes that are concordant with their clinical status. All 22 dogs homozygous for the mutation i.e. *TTC8*^{-/-}, have developed PRA, suggesting that the mutation is fully penetrant. The inheritance observed in a family of eight dogs (three cases) is supportive of a recessive mode (Figure 23). There was a group of dogs with genotypes discordant with their phenotypes, comprising 13 PRA-affected dogs that are not homozygous for *TTC8*_{c.669delA} and three obligate carriers do not carry *TTC8*_{c.669delA}. It is formally possible that the mutation has a dominant mode of inheritance with incomplete penetrance, or complex trait or compound heterozygote effects. However, given that three distinct loci have now been implicated in PRA in the breed, it is more likely that still more loci are responsible for the discordant cases. PRA caused by the mutation described here has an average age at diagnosis of 4.51 years and this is indicative of a late age of onset (data

not shown). The discordant GR PRA cases i.e. *TTC8*^{+/+} and *TTC8*^{+/-} tended to develop PRA at a later age, with an average age at diagnosis of 6.38 years (data not shown), which is consistent with the segregation of a fourth form of PRA in the GR breed.

TTC8 encodes the protein tetratricopeptide repeat (TPR) domain 8 and was recognised as a candidate gene due to its previous implication in BBS and aRRP in humans¹²⁸⁻¹³⁰. Human *TTC8* is made up of 14 exons and alternative splicing of the fifth exon results in two isoforms that are widely expressed in ciliated tissue¹²⁸. In addition a retina-specific isoform, expressed exclusively in the retina and primarily in photoreceptors, is created by the alternative splicing of exon 2A¹³⁰. The gene is comprised of 12.5 TPR motifs, which represent a protein-protein interaction module found in functionally diverse proteins that facilitates specific interactions between partner proteins^{128,149,150}. Part of the gene is also similar to the prokaryotic domain *pilF*, which is involved in twitching mobility and type IV pilus assembly^{128,151,152}. *TTC8* is one of seven highly conserved BBS proteins that form a stable complex known as the BBSome, which functions primarily at the ciliary membrane and is thought to play a role in ciliogenesis^{148,149}. The structural elements of the proteins that make up the BBSome resemble those of coat protein (COP) complexes involved in intracellular transport (COPI, COPII). BBS proteins, while not required for ciliogenesis *per se*, are thought to transport a specific set of transmembrane proteins to the cilia^{148,149}.

BBS is a pleiotropic disease and characteristic symptoms include obesity, retinal degeneration, kidney malformation, hypogonadism, polydactyly and learning disabilities. While the syndrome is usually inherited in an autosomal recessive manner, triallelic inheritance has also been observed⁶⁴. While most mutations that significantly alter the structure of the *TTC8* protein cause BBS, those that result in skipping of the 10 amino acids encoded by exon 2A cause non-syndromic RP (Figure 27). The *TTC8*_{c.669delA} mutation in the GR is predicted to have a significant effect on the structure of the protein, including the loss of tetratricopeptide repeats and the *pilF*-like domain near the carboxyl terminus¹²⁸.

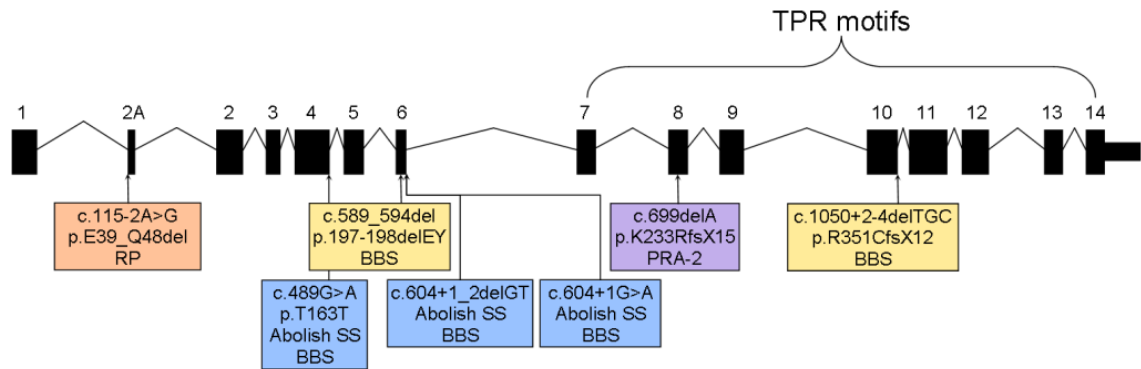


Figure 27. Comparison of human and canine retinal disease mutations

In humans, only mutations in exon 2A cause RP (orange¹³⁰), while mutations elsewhere have been associated with BBS (yellow¹²⁸ and blue¹²⁹). *TTC8*_{2Ac.699delA} (purple) is associated with PRA in GRs.

Most of the affected dogs, however, presented with typical PRA and no additional clinical signs. It is possible that other clinical signs in the dog have simply not been reported to, or by, the examining ophthalmologists. One dog with PRA-2 was reported to have “hormonal changes and thyroid problems”, the precise nature of which are unknown. Attempts to gain further information pertaining to this dog have thus far been unsuccessful. Signs seen in two other dogs (small stature, obesity, aberrant metabolism and olfaction, and unusual behaviour), as reported by their owners, suggest there may be more to the phenotype of PRA-2 than solely PRA. As the signs reported are subjective, an objective, clinical investigation is warranted. Attempts to conduct a clinical examination of one of these dogs have been unsuccessful. Nevertheless, it is clear there are phenotypic differences between human BBS and canine PRA-2. It is unclear whether these phenotypic differences are purely due to species differences or some other as yet unknown mechanism, such as the effect of modifier genes. The *TTC8* protein is clearly critical to ensure efficient function of ciliated tissue in humans, as demonstrated by the BBS phenotype. In contrast, it may be that the *TTC8* protein is critical for photoreceptor function in dogs, but less so in other tissues, which would explain a retina-specific or possibly less severe systemic phenotype in dogs. Alternatively, there may be other proteins that compensate for the loss of *TTC8* in dogs more than in humans. While PRA is widely considered to be the veterinary equivalent of RP, the limited characterisation of PRA at a cellular level is insufficient to fully justify this comparison. Further investigations are required to understand the cellular processes involved in this form of PRA, including whether the rod or cone

photoreceptor cells are affected first and whether other signs indicative of a syndrome are present in affected dogs.

3.10.3 Prevalence of PRA-1 and PRA-2 in GR populations

The absence of the *SLC4A3*_{c.2601_2602insC} allele from non-GR breeds tested, including some dogs affected with PRA, indicates that the mutation is rare and probably confined to GRs. In contrast, identification of a LR (with clinically apparent PRA) homozygous for the *TTC8*_{c.669delA} allele suggests it may be present in the LR breed. However, as only 1/17 LR PRA cases in our sample set, is caused by this mutation, it may be only a minor cause of PRA in the breed. The absence of the mutation from the other 69 LRs could be attributed to the small number of samples tested. The absence of *TTC8*_{c.669delA} from FCR and CBR dogs tested, including some dogs affected with PRA, indicates that the mutation is rare and is probably confined to the GR and LR breeds.

The mutant allele frequencies of *SLC4A3*_{c.2601_2602insC} and *TTC8*_{c.669delA} (4% and 1.1% respectively) within the samples collected for research purposes (n = 88, Table 11) would indicate that 1 in 632 and 1 in 7,744 GRs in the UK is affected with PRA-1 and PRA-2 respectively. These samples were unrelated at the parent level and were collected specifically to estimate the allele frequency in the breed. However this could still be an over-estimate as any sample cohort collected for research is unlikely to be representative of the wider GR population as not all dog owners will be aware of the research to begin with. In addition, the cohort size is also relatively small, thereby reducing the accuracy of population statistics and predictions.

In contrast, sample cohorts collected through AHT Genetic Services are larger and more diverse, as they were not recruited for any particular study, and therefore more representative of the wider population. However, the *SLC4A3*_{c.2601_2602insC} allele frequency of 4.5% in the UK fraction of this cohort was actually slightly higher than the unrelated research cohort (4%). Similarly, the *TTC8*_{c.669delA} allele frequency of 2.3% in the UK fraction of the AHT Genetic Services cohort was also slightly higher than the unrelated research cohort (1.1%) This suggests the strategy of collecting unrelated samples specifically for the purposes of estimating allele frequencies may be sufficient, assuming sufficient numbers are available, and the resulting data presents reliable evidence on which to base advice offered to breeders.

The presence of both mutations in GR dogs from all over the world (Table 12 and Table 13) suggests the mutations may have arisen prior to the geographic dispersion of the

breed. The *SLC4A3*_{c.2601_2602insC} allele frequency of 5.1% indicates that 1 in 386 GRs worldwide will develop PRA-1, and 1 in 10 is expected to carry it. Similarly, the *TTC8*_{c.669delA} allele frequency of 3.3% indicates that only one in 945 GRs worldwide will develop PRA-2, although 1 in 16 is expected to carry it.

3.10.4 *SLC4A3* in Humans

While *SLC4A3* has not been directly linked with disease in humans, the p.Ala867Asp variant confers susceptibility to idiopathic generalised epilepsy (IGE)¹⁵³. Hentschke et al. described a *SLC4A3* knockout mouse that appeared healthy, but had a reduced seizure threshold and increased seizure-induced mortality¹³⁸. In addition, AE3_c has a key role in myocardial pH_i recovery from alkaline loads¹⁵⁴. While the loss of AE3 alone has no known adverse effects on the heart the combined loss of AE3 and sodium/potassium/chloride transporter 1 (NKCC1, aka SLC12A2) impairs cardiac function¹³⁹. Similarly, loss of AE3 in the TM180 transgenic mouse (with a Glu180Gly substitution in the α -tropomyosin gene) led to more rapid decompensation and heart failure than the TM180 mouse alone¹⁵⁵.

The identification of a frameshift insertion in *SLC4A3* in GR dogs with PRA endorses the status of this gene as a candidate for human retinal degenerations. Screening of the exons of *SLC4A3* in DNA samples from 200 patients with retinal degeneration resulted in the identification of five SNP variants that may be pathogenic. Fourteen individuals with aRRP carry one (n = 13) or two (n = 1) copies of the minor allele at one of these loci. None of the heterozygous individuals carry more than one pathogenic variant and do not therefore appear to be compound heterozygotes. SNPs_1-3 are unlikely to cause RP due to their frequency in the 1000 Genomes and EVS populations (Table 15). The two remaining variants (SNP_4 and SNP_5) are each found in one individual and could have large effects on exon splicing (Figure 26). This could be confirmed through analysis of mRNA transcripts if they could be detected in easily accessible tissue such as blood or buccal cells. However, as canine *SLC4A3* mRNA is undetectable in blood or buccal cells, human *SLC4A3* mRNA is also likely to be undetectable and this was not pursued. In addition, the individuals affected by these variants carry only a single copy and the variants therefore remain insufficient to cause aRRP on their own.

While it appears that the variants in *SLC4A3* identified in this study are unlikely to cause aRRP in the cohort screened, at least on their own, this is insufficient evidence to exclude the gene as a candidate for retinal degeneration. There are a number of scenarios in which *SLC4A3* could be involved in retinal degeneration:

- 1) Intronic mutations or upstream elements controlling transcription that cannot be identified by sequencing exons could be pathogenic by affecting exon splicing or regulation of gene expression.
- 2) It is possible disease-causing mutations in this gene are rare and affected individuals have yet to be screened. For example, mutations in the *C2ORF71* gene have been associated with RP^{156,157}. However, in a study screening the gene in DNA from 286 individuals affected with LCA (n=95) or aRRP (n=191), a biallelic variant responsible for retinal degeneration could not be identified¹⁵⁸.
- 3) It is also possible that mutant *SLC4A3* actually causes retinal degeneration other than RP. Given the late onset of clinical signs in the dog and mouse models, early onset degenerations such as LCA are unlikely to be caused by variants in *SLC4A3*. Alvarez et al. concluded that their results in the knockout mouse linked aberrant *SLC4A3* to vitreoretinal degeneration¹²⁵. While there are similarities to vitreoretinal disorders, the data presented for the mouse phenotype is not pathognomonic of a particular human phenotype (personal communication, Michel Michaelides). In addition, vitreoretinal disorders are rare and DNA samples may be difficult to obtain. The other abnormalities reported by Alvarez et al., altered ERG and abnormal retinal vessels, are shared by many vitreoretinal and retinal degenerations. Currently there is no evidence to suggest a human phenotype more likely to be caused by variants in *SLC4A3* than aRRP and screening of patients with other phenotypes was therefore not undertaken.

NGS is now commonly performed in patients with retinal dystrophies by many groups world-wide, so it is possible that human *SLC4A3* mutations may be identified through these efforts.

3.11 Conclusion

Several forms of PRA segregate in the GR, including *prcd*, but this form of PRA does not account for the majority of cases, particularly in Europe. Using a GWA analysis approach, two novel candidate mutations were identified that are likely to represent major susceptibility loci for PRA in the GR. *SLC4A3*_{c.2601_2602insC} (p.E868RfsX104) accounts for the majority of cases in our sample cohort (56%) and *TTC8*_{c.669delA} (p.Lys233ArgfsX15) accounts for a further 28% (Figure 28). While these variants do not explain all cases of PRA in this study, suggesting that there are additional loci causing PRA in this breed, they do appear to be fully penetrant and together cause the majority of PRA cases in the breed. These findings therefore indicate that PRA is caused by at least four variants in this breed.

These findings also comprise the first report of an association between aberrant *SLC4A3* and retinal degeneration in a spontaneous animal model. As a result, *SLC4A3* has become a provocative candidate gene for human retinal degenerations. All of the variants identified through the screening of *SLC4A3* in human patients with retinal degeneration were eliminated from further investigation due to insufficient evidence of pathogenicity. However, the absence of possible causal variants in ~200 human aRRP patients is insufficient evidence to exclude the gene as a candidate entirely. It may be associated with disease in a very small number of patients, or even patients with a different disease type altogether.

In conclusion, the findings presented here established PRA-1 as a model for human retinal disease and PRA-2 as a model for human RP, and potentially BBS. Both forms of PRA in the GR may prove to be a valuable models for further studies to enhance our understanding of visual pathways and gene therapy investigations. It is unknown whether the remaining cases are caused by variants in one or more genes, and this, along with the relatively few samples available, will make identification of further variants in the breed challenging (discussed further in Chapter 6).

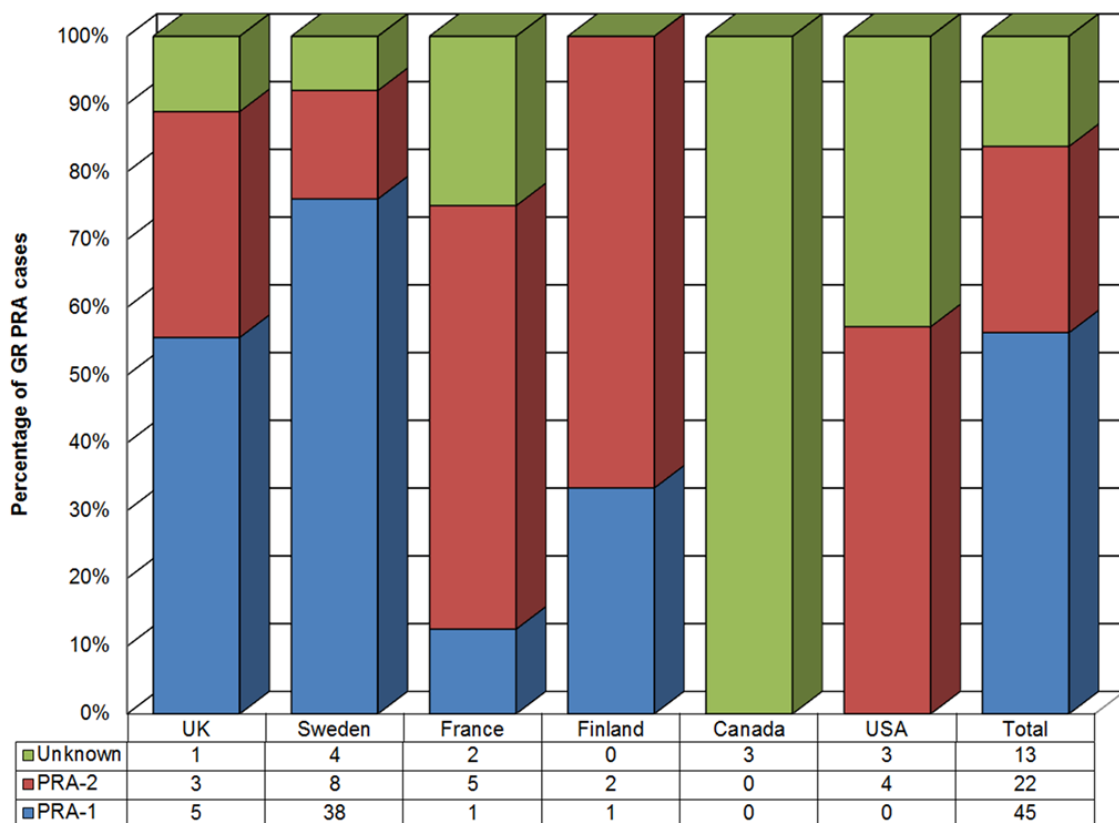


Figure 28. Geographical distribution of GR PRA variants

More than half (56%; 45/80; blue) of the PRA cases tested are the PRA-1 form of the disease, while 28% (22/80; red) are PRA-2 and the remaining 16% (13/80; green) have PRA of unknown aetiology.

Chapter **4** Gordon Setters

4.1 Introduction and background

GoSs are a healthy breed and are reported to be predisposed to only a handful of hereditary conditions. These include lethal astrocytosis, canine juvenile cellulitis, hip and elbow dysplasia and cerebellar degeneration¹²¹. GoSs were originally known as Black and Tan Setters and were popular in the Midland counties of the United Kingdom in the 17th century¹²². The first ever documented description of PRA was in a GoS over 100 years ago, described by Magnusson as stated by Barnett^{159,160}. However, since that time there have been few reports of the condition until the 21st century, when a number of dogs were diagnosed with a late-onset PRA¹⁶¹. Irish Setters (IrS) and Irish Red and White Setters (IRWS) are two distinct breeds that are thought to have originated from common founders, and both are known to be predisposed to PRA. It is unclear how closely related GoSs are to IrSs, but anecdotal evidence suggests they have been interbred in the past. In addition there are currently breeding kennels that breed with both; it is therefore possible that interbreeding is still occurring. IrSs are known to be affected by an early-onset form of PRA, rod-cone degeneration 1 (rcd1)¹⁶². However, in recent years in particular, a number of older dogs that are homozygous for the wildtype allele at the rcd1 locus have been diagnosed with PRA. Evidence therefore suggests there is at least one other form of PRA, with a later age of onset, segregating in this breed in addition to rcd1¹⁶³. Anecdotal evidence suggests that PRA in the IRWS is also a late-onset form of the disease. In light of this it is possible that the GoS, IrS and IRWS breeds share the late onset form of PRA, caused by an identical mutation.

4.2 Diagnoses and Clinical Findings

PRA in the GoS is clinically indistinguishable from PRA in other breeds. Dogs typically present with tapetal hyper-reflectivity, optic nerve atrophy and attenuated vessels (Figure 29). Dogs in which clinical signs are present also display an absence of retinal function as detected by ERG (Figure 30). The mode of inheritance is consistent with an autosomal recessive trait. The age at diagnosis (which is indicative of age of onset) was reported for 16 dogs in our sample cohort. The average was 8.91, but ranged from 3 to 12 years of age with only four younger than eight years of age.

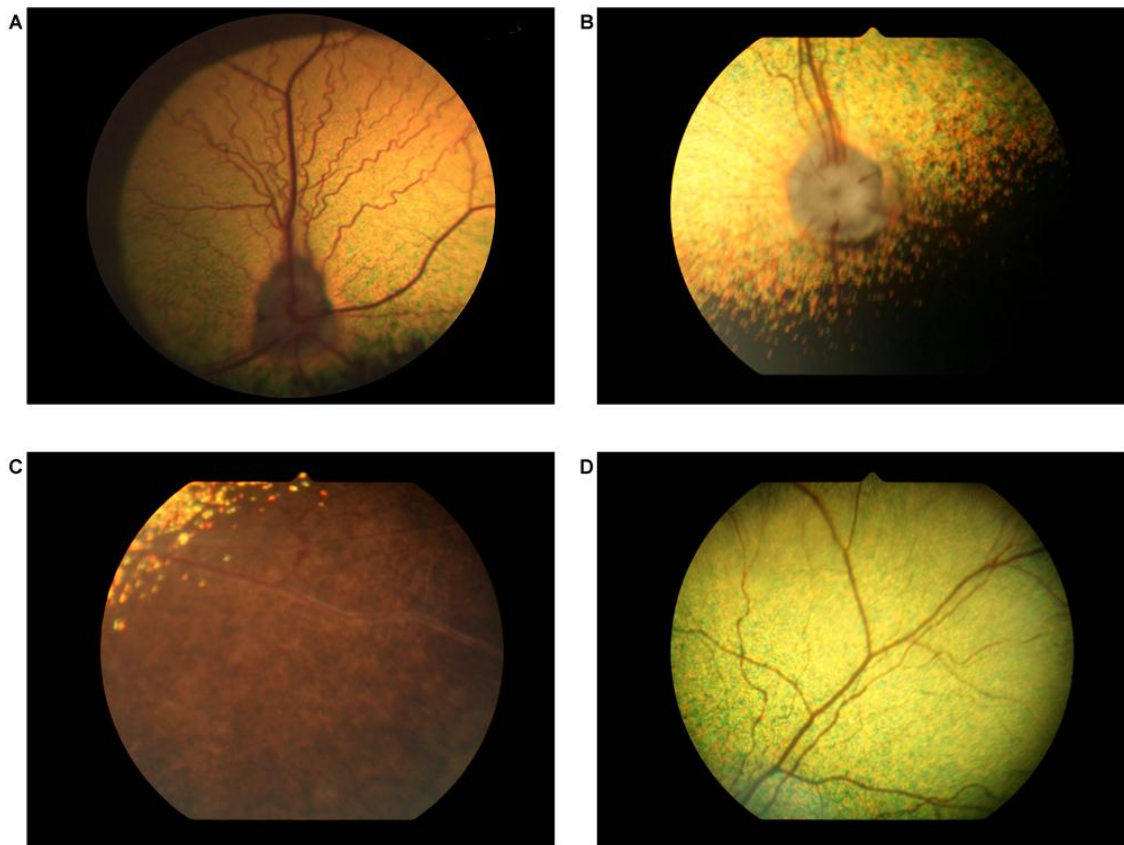


Figure 29. Fundus changes observed in a GoS with PRA.

A) The fundus of a healthy dog. B-D) The fundus of a Gordon Setter, age 11 years, displaying signs typical of PRA in most breeds: Hyper-reflective tapetum (B), pigmentary changes in the non-tapetal fundus (C) and vascular attenuation (D).

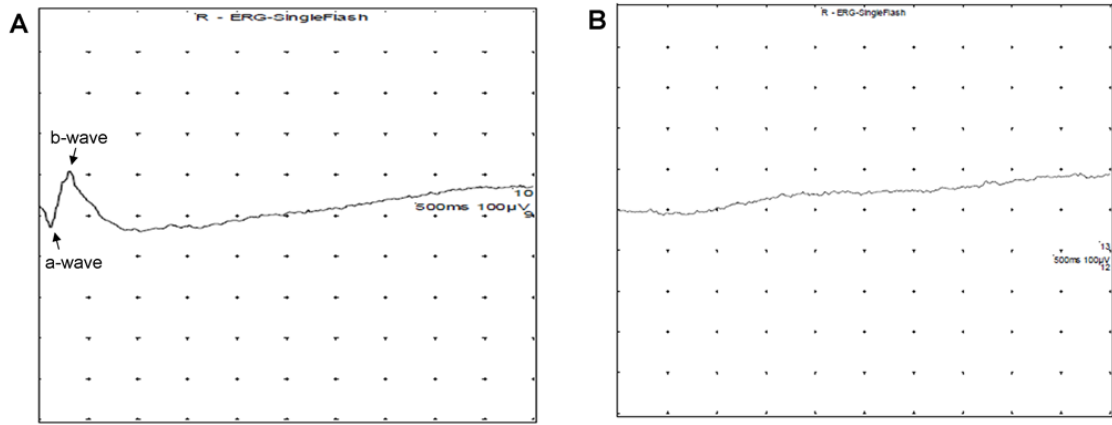


Figure 30. Retinal function measured in a GoS with PRA.

Scotopic ERGs conducted without sedation, showing the average of up to 12 tests. A) Control ERG of a dog with normal vision. The characteristic a-wave and b-wave are evident. B) ERG of a Gordon Setter affected with PRA. The flat line is indicative of no retinal function.

4.3 SNP Genotyping

4.3.1 Genome-Wide Association Mapping

GWA analysis of 38 GoS dogs, 16 PRA cases and 22 controls over the age of 8 when last examined, which were genotyped with 107,982 SNPs (section 2.3) revealed a genome-wide significant association with PRA on chromosome 17 (CFA17; $p_{\text{raw}} = 2.76 \times 10^{-10}$, $p_{\text{genome}} = 1 \times 10^{-5}$). Additional genome-wide significant associations were seen on chromosomes 10 ($p_{\text{raw}} = 7.47 \times 10^{-7}$, $p_{\text{genome}} = 0.02$), 21 ($p_{\text{raw}} = 6.1 \times 10^{-7}$, $p_{\text{genome}} = 0.02$) and 28 ($p_{\text{raw}} = 3.2 \times 10^{-7}$, $p_{\text{genome}} = 0.01$) (data not shown). IBS clustering confirmed the presence of substantial population stratification, with an inflation factor > 1 ($\lambda = 2.44$, based on median χ^2). MDS analysis revealed an uneven distribution of cases and controls (Appendix I), suggesting that many of the cases were more closely related to one another than to the controls, and vice versa. The signal on CFA17 remained strong ($p_{\text{raw}} = 2.22 \times 10^{-8}$) and the inflation factor was reduced ($\lambda = 1.56$) when corrected for population stratification with a CMH meta-analysis (Figure 31A). This signal remained significant after correction for multiple testing ($p_{\text{genome}} = 2 \times 10^{-5}$) while all other signals fell below the level of genome-wide significance (data not shown). The statistically associated region on CFA17, defined as the region encompassing statistically associated SNPs, spanned 1.64 Mb from 25.825 to 27.466 Mb with the most statistically associated SNP markers (BICF2S23023516, BICF2P960008, BICF2S23914151 and BICF2S23419454) at 25.83, 25.84, 26.06, and 27.19 Mb (Figure 31B). Alternative analysis of the data using FMM to correct for population stratification revealed similar results with the only statistically significant association on CFA17 ($p_{\text{raw}} = 3.96 \times 10^{-9}$), although the most statistically associated SNP (BICF2P91126) was at 27.47 Mb (data not shown). However, the greatly reduced inflation factor ($\lambda = 0.34$) suggests an over-correction for population stratification.

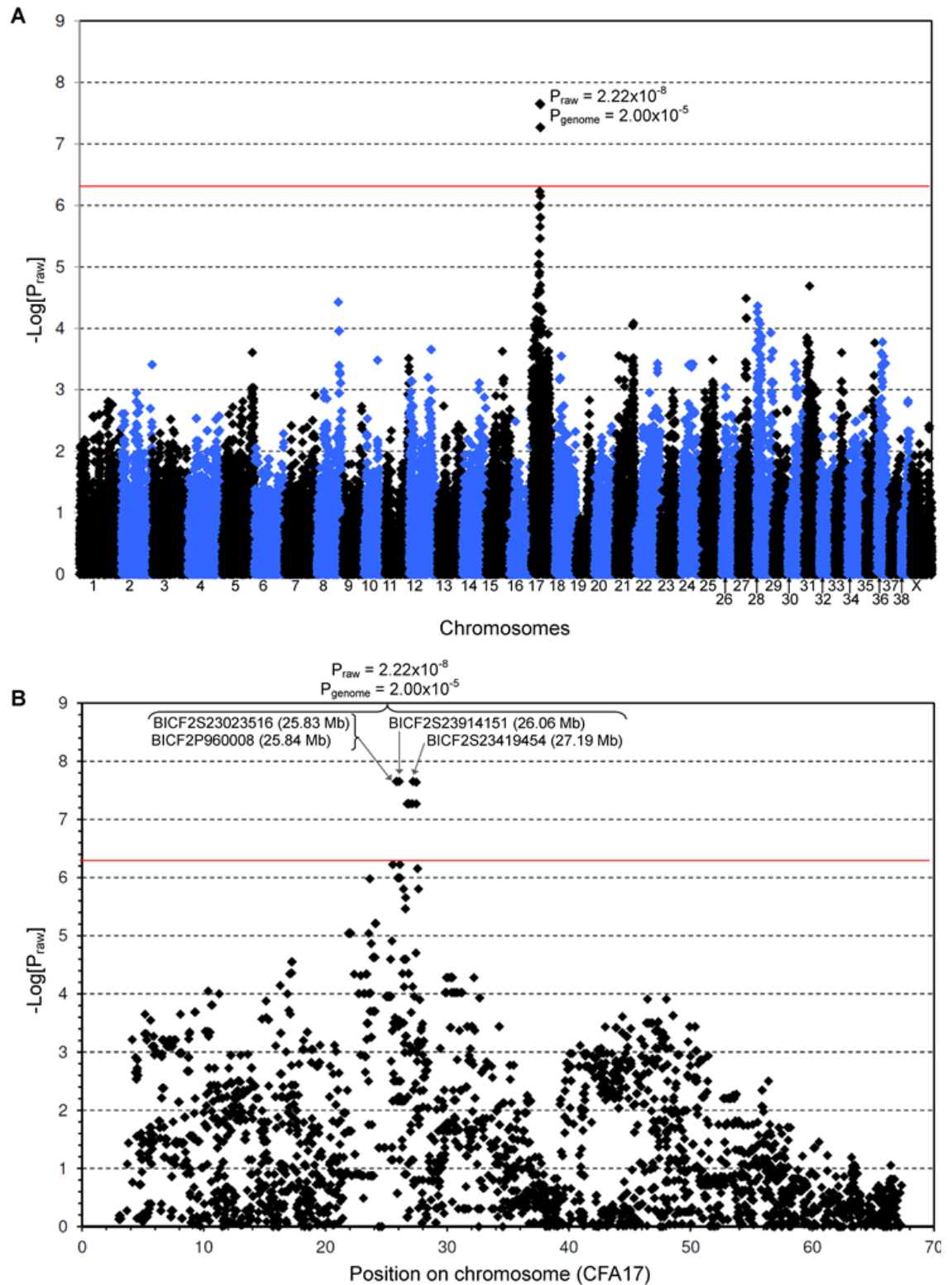


Figure 31. Genome-wide association mapping of PRA in GoSs

$-\text{Log}_{10}$ of p-values after correction for population stratification. The red lines indicate the Bonferroni-corrected 5% significance level based on 107,982 SNPs. A) Results show a strong statistical signal on CFA17 ($p_{\text{raw}} = 2.22 \times 10^{-8}$, $p_{\text{genome}} = 2.00 \times 10^{-5}$). B) Statistically associated SNPs span a 1.64 Mb region from 25.825 Mb to 27.466 Mb on CFA17.

4.3.2 Homozygosity Analysis

A haplotype of 3.221 Mb, extending from 24.629 to 27.851 Mb, was identified through homozygosity analysis for which 15 of 16 cases and only 1 of 22 controls were homozygous (Figure 32). Although PRA in the GoS is likely to be inherited as an aR trait, and all affected dogs would therefore be expected to be homozygous for the disease-associated haplotype as well as the causal mutation, it is also possible that more than one form of PRA is segregating in the breed. The case that is not homozygous for the region was diagnosed at 4.5 years of age and this is younger than expected for the GoS breed. For this reason this disease-associated region was not excluded from further investigation on the basis of the single discordant case that was not homozygous as it may have been suffering from a genetically distinct form of PRA. There are 27 genes within the PRA critical region, 22 of which have human orthologues (Figure 32B and Appendix IV). One of the genes, chromosome 2 open reading frame 71 (*C2ORF71*), has been associated with RP in humans, is expressed in the retina and is thought to play a role in retinal function, although its exact role remains unknown^{156-158,164}. None of the other 26 genes are known to have an ocular or retinal function, or have been associated with vision loss. *C2ORF71* alone was therefore identified as the most provocative candidate gene for further investigation.

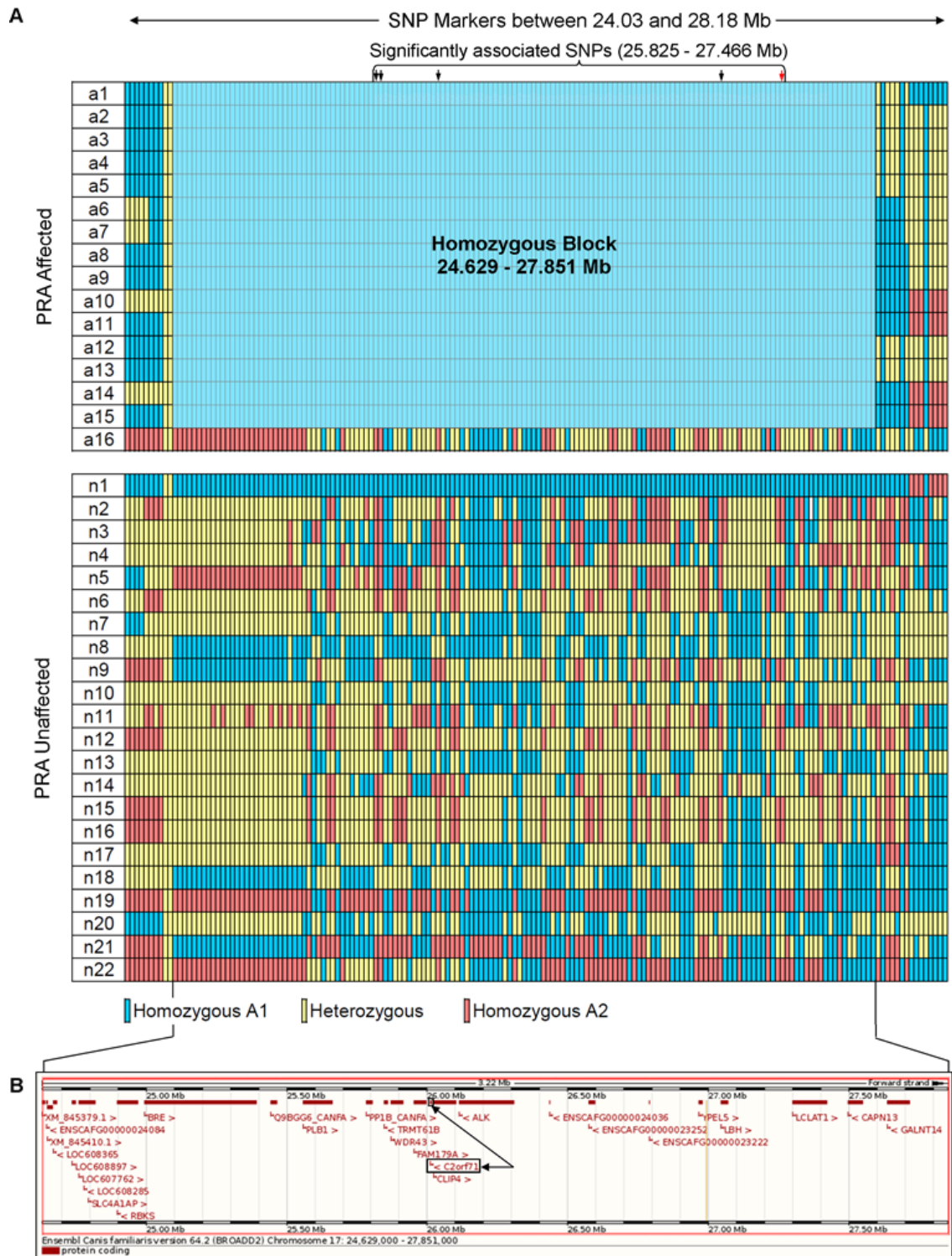


Figure 32. Homozygosity analysis of Gordon Setters with PRA

A) SNP genotypes for 16 PRA cases and 22 PRA controls, over the 1.6 Mb region identified during the GWA study. Fifteen of the cases share a homozygous haplotype of 147 SNPs (Homozygous Block), defining the critical region of 3.221 Mb between 24.629 Mb and 27.851 Mb on CFA17. One of the cases does not share this haplotype with the other 15 cases, while one of the controls does. The five most significantly associated SNPs are indicated by arrows (↓; CMH-corrected in black, FMM-corrected in red). B) The critical region contains 27 genes, as annotated on CanFam2.0 (taken from the Ensembl genome browser), of which *C2ORF71* is the most provocative candidate gene.

4.4 Sequencing of Candidate Gene *C2ORF71*

4.4.1 Characterising *C2ORF71* Retinal Transcript

Alignment of canine, human and mouse genomic and coding sequences revealed several inconsistencies and possible errors in the prediction of intron-exon boundaries of canine *C2ORF71* (Figure 33). All of the coding sequence of the *C2ORF71* retinal mRNA transcript was successfully amplified and sequenced (sections 2.2 and 2.4.1) in a healthy dog, confirming that the gene is transcribed in the canine retina. In addition, intron-exon boundaries were identical to those of the human and mouse, which is in conflict with the boundaries predicted by Ensembl genebuild for the canine gene (Figure 33). All attempts at sequencing the full 5' and 3' UTRs were unsuccessful. Sequencing of the retinal mRNA transcript revealed that the canine *C2ORF71* protein (Genbank accession no JN390834) contains 1,308 amino acids with a predicted molecular weight of 139 kDa.

4.4.2 Sanger Sequencing of *C2ORF71*

Primers for the sequencing of gDNA were designed to ensure coverage of the known exons as well as probable 5'- and 3'-UTRs. The *C2ORF71* gene contains only two exons, both of which were sequenced (sections 2.2 and 2.4.1) by Sanger sequencing in two PRA affected and four PRA unaffected GoSs, as well as one Miniature Long Hair Dachshund and one Flat Coat Retriever. Comparison of the sequence data with the canine reference genome sequence (CanFam2.0) revealed 31 sequence variants. Only one of the sequence variants segregated consistently with PRA within the eight samples sequenced, was exonic and potentially pathogenic (Figure 33D). This variant was an inserted cytosine in exon 1 (c.3149_3150insC; CFA17:26,010,485). It is predicted to shift the reading frame and cause a premature stop codon (p.C1051VfsX90) potentially resulting in degradation of the mRNA by NMD or a truncated protein product (Figure 34). While the position of the variant is not particularly well conserved at the nucleotide or amino acid level between 31 eutherian mammals (data not shown), the predicted effect of the variant is severe and it therefore remains significant.

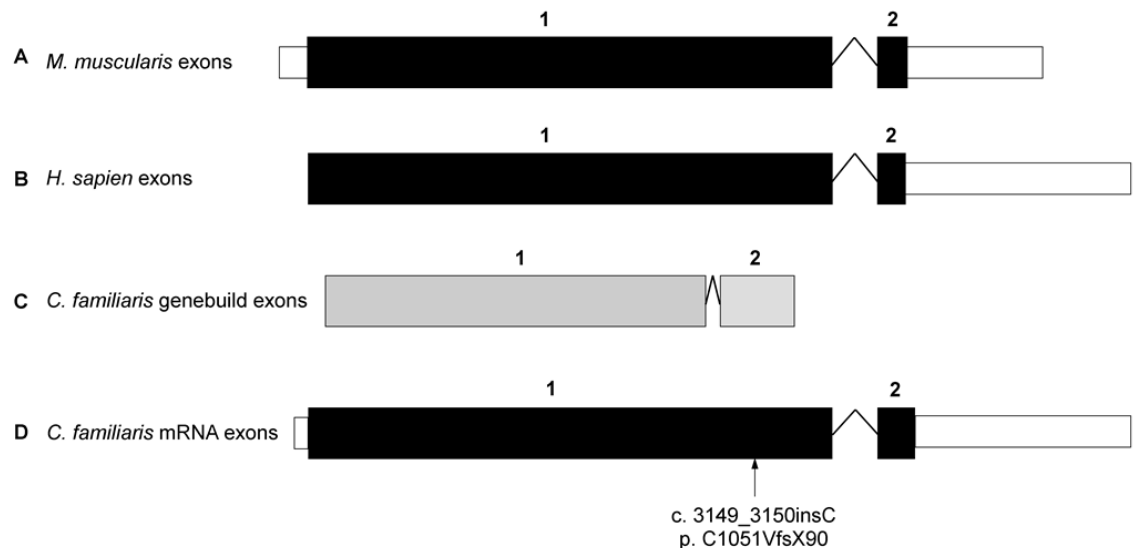


Figure 33. Graphical comparison of the intron-exon boundaries of *C2ORF71*.

A) Mouse (*Mus musculus*) *C2ORF71*. B) Human (*Homo sapiens*) *C2ORF71*. C) Canine (*Canis familiaris*) *C2ORF71* as predicted by Ensembl genebuild. Both exons predicted (grey) show sequence similarity with human and mouse exon 1, although the 5'- and 3'-boundaries are inconsistent. D) Canine *C2ORF71* exons confirmed by sequencing the retinal transcript. The location of the fully-segregating sequence variant is indicated, and is exonic and nonsynonymous.

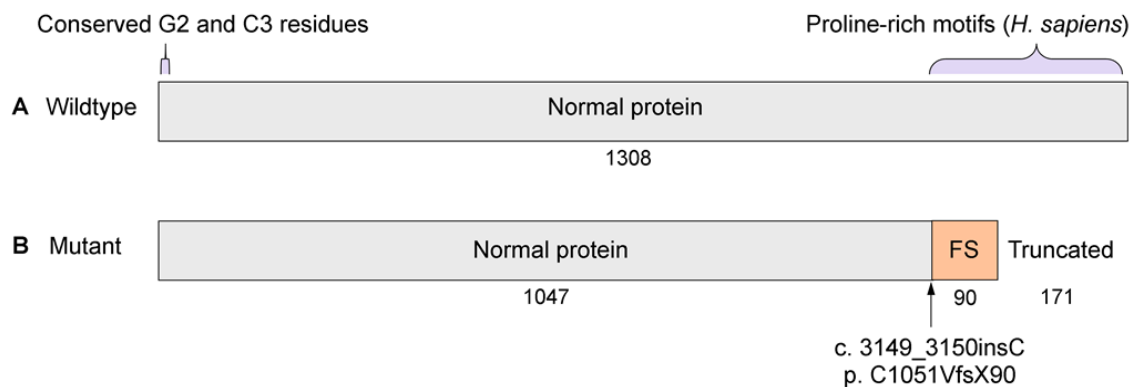


Figure 34. Effect of variant *C2ORF71*_{c.3149_3150insC} on the *C2ORF71* protein.

A) The wildtype canine *C2ORF71* protein contains 1308 amino acids. B) The effect of the insertion on the protein is significant. 1047 amino acids at the N-terminus of the protein are normal. However the insertion causes a shift in the reading frame that results in 90 aberrant amino acids and leads to a premature termination codon. This is predicted to result in a truncated protein product, lacking 171 residues of the C-terminus. Therefore a large part of the protein is absent.

4.4.3 *C2ORF71*_{c.3149_3150insC} Variant Validation

All 38 GoSs (16 cases and 22 controls) that participated in the GWA study were screened (section 2.4.1) by Sanger sequencing for the coding variant, *C2ORF71*_{c.3149_3150insC}, to confirm the association of this mutation with PRA and compare it with the most highly associated SNPs, BICF2S23023516, BICF2P960008, BICF2S23914151 and BICF2S23419454. The genotypes of the mutation and all four SNPs were perfectly concordant in all dogs i.e. homozygous for the reference allele (*C2ORF71*^{+/+}), heterozygous (*C2ORF71*^{+/-}) or homozygous for the mutant allele (*C2ORF71*^{-/-}). *C2ORF71*_{c.3149_3150insC}, along with all four markers showed significant and equal allelic association with PRA ($p_{\text{raw}} = 2.758 \times 10^{-10}$). It also showed almost complete segregation with PRA (15/16 cases) and is highly likely to have a deleterious effect on the protein. Seven GoSs were carriers of the mutation but all were clinically free of PRA at their last eye examination (average age 12yo, range 10-14yo) which substantiates the hypothesis that PRA in the GoS has an aR mode of inheritance.

4.5 *C2ORF71*_{c.3149_3150insC} Population Screening

4.5.1 Research Sample Cohorts

Eighty-eight additional GoSs, (seven clinically affected with PRA and 81 unaffected dogs of any age) were screened (section 2.6.1.1), making a total of 126 GoSs tested for *C2ORF71*_{c.3149_3150insC} (Table 16) to confirm that the mutation was not a commonly occurring polymorphism in this breed. Of the 22 PRA cases used in the study 19 (86.4%) were homozygous for the mutation (*C2ORF71*^{-/-}) and 84/104 (80.8%) dogs known to be clinically free of PRA were either carriers of the mutant allele (44.2%; *C2ORF71*^{+/-}) or homozygous for the wildtype allele (36.5%; *C2ORF71*^{+/+}). For 12/20 clinically unaffected dogs that were homozygous for the mutation, their age when they were last examined was reported (average = 9 years). Interestingly, seven were at least 10 years old and three subsequently underwent follow-up ophthalmoscopic examination at the AHT (sections 2.1.1.2 and 4.5.3). Anecdotal evidence from owners suggests that affected GoSs often present with nyctalopia (night-blindness) initially, prior to more extensive visual impairment, indicating that this form of PRA may be a rod-cone degeneration (RCD), known hereafter as RCD4 to distinguish it from forms of RCD described previously.

Table 16. *C2ORF71* genotypes and PRA clinical status for 126 GoSs

Genotype	PRA Affected	PRA Unaffected	Total
<i>C2ORF71</i> ^{-/-}	19	20	39
	86.4%	19.2%	31.0%
<i>C2ORF71</i> ^{+/-}	1	46	47
	4.5%	44.2%	37.3%
<i>C2ORF71</i> ^{+/+}	2	38	40
	9.1%	36.5%	31.7%
Total	22	104	126

To determine whether RCD4 accounts for PRA in other breeds, a further 91 dogs were screened from two closely related breeds. These were 19 IRWSs and 72 IrSs, including three and 10 respectively clinically affected with PRA. All 19 IRWS dogs, including three PRA cases, were homozygous for the wild-type allele (*C2ORF71*^{+/+}). Of the 10 IrSs with PRA, seven were homozygous for the mutation (*C2ORF71*^{-/-}), while the remaining three PRA cases were either carriers (*C2ORF71*^{+/-}; n=2) or homozygous wildtype (n=1). In addition a further five IrS dogs that had not been diagnosed with

PRA were also homozygous for the mutation (Table 17). Only one of these dogs had previously had an eye examination, at the age of six years, and was free of PRA at the time. Another subsequently underwent follow-up ophthalmoscopic examination at the AHT (sections 2.1.1.2 and 4.5.3).

Table 17. *C2ORF71* genotypes and PRA clinical status for 72 IrSs

Genotype	PRA Affected	PRA Unaffected	Total
<i>C2ORF71</i> ^{-/-}	7	5	12
	70.0%	8.1%	16.7%
<i>C2ORF71</i> ^{+/-}	2	25	27
	20.0%	40.3%	37.5%
<i>C2ORF71</i> ^{+/+}	1	32	33
	10.0%	51.6%	45.8%
Total	10	62	72

In addition to these setter breeds, *C2ORF71*_{c.3149_3150insC} has been identified in Standard Poodles (SP) Tibetan Terriers (TT) and English Setters (ES) (Table 18). Six out of nine *C2ORF71*^{-/-} ESs had been clinically diagnosed with PRA. Interestingly in this breed, the *C2ORF71*_{c.3149_3150insC} allele was found exclusively in dogs from Norway. Out of three SPs and six TTs that were *C2ORF71*^{-/-}, one and three respectively had been diagnosed with PRA. The five dogs without a PRA diagnosis were sampled for different studies and therefore were too young to show signs of RCD4 and/or were not ophthalmoscopically examined.

The allele frequency of *C2ORF71*_{c.3149_3150insC} in the GoS, IrS, ES and TT breeds is extremely high for an autosomal recessive disease allele. The cohorts screened contained samples collected for the purposes of research, including all available samples from dogs clinically affected with PRA, and samples from PRA-unaffected dogs closely related to affected dogs and to one another. These cohorts were therefore subject to considerable sampling bias. This combined with the relatively small number of samples screened has resulted in cohorts that do not conform to HWE, with the exception of IrSs (Table 18). Analyses using only 36 dogs with unique sires and dams resulted in a reduced disease allele frequency, and this cohort does conform to HWE (Table 18). This suggests that the high degree of relatedness between samples is a major cause of deviation from HWE, and must be considered when making disease frequency predictions and subsequent breeding recommendations.

Table 18. Population screening for RCD4

Breed	Observed Genotypes*					HWE		1-in-__ Expected	
	-/-	-/+	+/+	TOTAL	Allele Freq	χ^2	P	Affected	Carrier
Research Cohort									
Gordon Setters	39	47	40	126	0.50	8.12	4.00x10 ⁻³	4	2
Gordon Setters Unrelated	6	14	16	36	0.36	0.89	0.35	8	2
Irish Setters	12	27	33	72	0.35	2.34	0.13	8	2
English Setters	9	7	40	56	0.22	22.90	1.70x10 ⁻⁶	20	3
Standard Poodle	3	3	86	92	0.05	38.81	4.66x10 ⁻¹⁰	418	11
Tibetan Terrier	6	11	81	98	0.12	20.57	5.75x10 ⁻⁶	73	5
AHT Genetic Services Cohort									
Gordon Setters	208	747	833	1788	0.33	4.14	0.04	9	2
Irish Setters	158	824	1267	2249	0.25	2.28	0.13	16	3
Tibetan Terrier	1	24	35	60	0.22	1.91	0.17	21	3

* The mutant allele is represented by “-“ and the wildtype allele by “+”.

4.5.2 AHT Genetic Services Cohort

The DNA test for RCD4 was made available commercially on 21 March 2011, initially only for GoSs and later for IrSs and TTs also. Over a period of 24 months 1,788 GoSs from 28 countries were genotyped by AHT Genetics Services (Table 18). Overall, 208 (12%) were *C2ORF71*^{-/-} and 747 (42%) were *C2ORF71*^{+/-}, indicating an overall allele frequency of 0.325. Similarly over a period of 18 months 2,249 IrSs from 31 countries were genotyped for *C2ORF71*_{c.3149_3150insC} (Table 18). Overall, 158 (7%) were *C2ORF71*^{-/-} and 824 (37%) were *C2ORF71*^{+/-}, equal to an overall allele frequency of 0.253. TTs have recently been added to the breeds tested, and over a period of three months 60 TTs from two countries were genotyped for *C2ORF71*_{c.3149_3150insC} (Table 18). Overall, only one (2%) was *C2ORF71*^{-/-} and 24 (40%) were *C2ORF71*^{+/-}, equal to an overall allele frequency of 0.217. The larger number of samples resulted in a

distribution of genotypes closer to HWE than research samples; only the GoS cohort did not conform to HWE. The Genetic Services cohorts are therefore likely to be more representative of the general population than the research cohorts.

The allele frequency for each country was calculated for those represented by at least 20 samples (Table 19) and range from 0.16 to 0.5 in GoSs and 0.003 to 0.33 in IrSs. The allele frequency in TTs in the UK is 0.22. Interestingly, each cohort represented by >20 samples, with the exception of the Finnish GoS cohort, conforms to HWE. Additionally, the frequency of *C2ORF71*_{c.3149_3150insC} is higher than expected for the recessive disease trait in most cohorts.

Table 19. GoS RCD4 breakdown by country (n>19)

Country	Observed C2ORF71 Genotypes*					HWE		1-in-__ expected		Region [†]
	-/-	-/+	+/+	TOTAL	Allele Freq	χ^2	P	Affected	Carrier	
Gordon Setters										
Australia	10	37	42	89	0.320	0.18	0.67	10	2	1
Canada	0	11	22	33	0.167	1.32	0.25	36	4	2
USA	11	102	189	302	0.205	0.37	0.54	24	3	2
Poland	7	30	19	56	0.393	0.85	0.36	6	2	3
Russia	5	19	14	38	0.382	0.13	0.71	7	2	3
Norway	2	17	45	64	0.164	0.06	0.80	37	4	4
Sweden	5	18	24	47	0.298	0.33	0.56	11	2	4
Finland	1	28	22	51	0.294	5.30	0.02	12	2	4
Italy	5	16	9	30	0.433	0.22	0.64	5	2	5
UK	70	203	182	455	0.377	1.14	0.28	7	2	6
Austria	0	12	11	23	0.261	2.87	0.09	15	3	6
Switzerland	2	20	19	41	0.293	1.30	0.25	12	2	6
Germany	42	158	153	353	0.343	0.02	0.90	9	2	6
Netherlands	14	29	34	77	0.370	2.85	0.09	7	2	6
Belgium	11	20	11	42	0.500	0.10	0.76	4	2	6
Total	185	720	796	1701	0.320	1.34	0.25	10	2	
Irish Setters										
Australia	3	14	22	39	0.256	0.13	0.71	15	3	1
USA	0	1	195	196	0.003	0.00	0.97	153,664	197	2
Poland	1	10	22	33	0.182	0.01	0.92	30	3	3
Hungary	1	9	14	24	0.229	0.09	0.76	19	3	3
Norway	0	7	30	37	0.095	0.40	0.53	112	6	4
Sweden	4	39	72	115	0.204	0.21	0.65	24	3	4
Finland	11	55	89	155	0.248	0.38	0.54	16	3	4
UK	81	410	482	973	0.294	0.22	0.64	12	2	6
Belgium	10	61	77	148	0.274	0.20	0.65	13	3	6
Germany	25	111	133	269	0.299	0.07	0.79	11	2	6
Netherlands	10	59	52	121	0.326	1.43	0.23	9	2	6
Total	146	776	1188	2110	0.253	1.56	0.21	16	3	
Tibetan Terriers										
UK	1	24	34	59	0.220	2.00	0.16	21	3	6
Total	1	24	34	59	0.220	2.00	0.16	21	3	

* The mutant allele is represented by “-“ and the wildtype allele by “+”.

† 1 = Australasia; 2 = North America; 3 = Eastern Europe; 4 = Northern Europe; 5 = Southern Europe; 6 = Western Europe

4.5.3 Ophthalmoscopic Follow-Up

Three of the GoSs and one of the IrSs not previously diagnosed with PRA but identified as homozygous for the *C2ORF71*_{c.3149_3150insC} allele during the screening of research samples described above were subsequently examined by an ophthalmologist (section 2.1.1.2; Table 20). The three older dogs (GoS#1, GoS#2 and IrS#2) displayed clinical signs consistent with PRA and this was confirmed by flat ERG traces in two of them (data not shown). GoS#3 did not display any clinical signs associated with PRA, but did have a reduced ERG response (data not shown).

Table 20. Ophthalmoscopic examination and ERG analysis of *C2ORF71*^{-/-} dogs

ID	Age (years)	Ophthalmoscopic examination	ERG Method	ERG Result	Information from owner
GoS#1	11	Poor pupillary light reflexes	No sedation	Flat traces	Nyctalopia i.e. tentative and nervous in dark
		Bilateral cataracts	Darkened room, without dark adaptation		
		Tapetal hyper-reflectivity	Average of 12 flashes		
		Mild/moderate blood vessel attenuation			
GoS#2	11	Subtle changes in tapetum	Dark adapted	Flat traces	Retains enough vision to run an agility course
		Mild vessel attenuation	Under anesthesia		
			Average of 8 flashes		
Gos#3	6.5	Normal retina	Dark adapted	Diminished responses	
		Incipient cataracts	Under anesthesia		
			Average of 8 flashes		
IrS#1	13	Nuclear sclerosis	None	None	Completely blind for 2 years at time of examination
		Senile cataract formation			
		Areas of tapetal hyper-reflectivity			
		Mild vascular attenuation.			

4.6 Discussion

Using a GWA mapping approach, a novel locus on chromosome 17 that is significantly associated with PRA in the GoS breed ($p_{\text{genome}} = 2.00 \times 10^{-5}$) has been identified. A critical region of 3.221 Mb was defined through homozygosity mapping using the SNP genotypes. Most of the cases (15/16) and one control (1/22) were homozygous for the same haplotype across the critical region. *C2ORF71* was identified as a strong candidate causal locus within the PRA-critical region as it has been implicated in human retinitis pigmentosa^{156,157,164}. *C2ORF71* is highly expressed in the eye in humans, and is localised to the photoreceptors, specifically the outer segment or connecting cilium¹⁵⁷. The phenotypes of individuals affected with mutations in *C2ORF71* vary in that while RP is usually diagnosed as an adult, individuals with a more severe and early-onset retinal dystrophy have also been described¹⁵⁷. The human protein contains no known functional domains. However, a proline-rich region near the C-terminal end of the protein is annotated on the Ensembl GRCh37 assembly (Figure 34). While the functional significance of this domain in *C2ORF71* is unknown, proline-rich domains in other proteins have been implicated in protein interactions¹⁶⁵. In addition, evidence indicates that residues G2 and C3 are highly conserved and undergo post-translational lipid modification¹⁵⁷. Finally, Collin et al. also show that *C2ORF71* plays an important role in the mature retina, and may also be required in the developing retina¹⁵⁶. A novel candidate mutation in the *C2ORF71* gene was identified that is likely to represent a major causal mutation for PRA in the GoS. This mutation was a frame-shifting insertion that is predicted to result in a truncated protein product (c.3149_3150insC, p.C1051VfsX90).

Sequencing of *C2ORF71* from healthy retinal mRNA served three purposes. Firstly it confirmed the presence of the *C2ORF71* mRNA transcript in the normal canine retina. Secondly it revealed that the intron-exon boundaries predicted by Genebuild for the dog are incorrect. They are instead identical to the human and mouse boundaries. Thirdly it revealed an exon orthologous to human exon 2, that is absent from the Ensembl canine predictions (Figure 33). No alternative isoforms of *C2ORF71* were observed, and none have been reported for the mouse or human gene.

In order to further test the validity of the insertion mutation, 126 GoSs were screened for the mutation (Table 16). Most 19/22 (86.4%) of the PRA cases and 80.8% of clinically unaffected dogs (which could be clear of the mutation or carry a single copy) have *C2ORF71* genotypes that are concordant with their clinical status. There are two

groups of dogs with genotypes discordant with their phenotypes. The first comprises 20 PRA-clear dogs that are *C2ORF71*^{-/-}. Three of these dogs had cataracts and as a result a comprehensive fundus exam was impossible, and an ERG was not conducted. It is possible the cataracts in these dogs may be secondary to undiagnosed PRA. A further four dogs, while not officially diagnosed with PRA, were described to have abnormal blood vessels (narrow or atretic). These signs are consistent with the very early stages of PRA. We were unable to obtain clinical information pertaining to seven of the dogs, and the remaining six dogs were examined between the ages of 2 and 11 years. The age of diagnosis of RCD4 is highly variable, ranging from 5-12 years, with an average of 9.8 years in our cohort, and as such these eight discordant dogs could have been examined prior to the onset of their clinical signs. Of the 20 discordant dogs, three were subsequently re-examined (Table 20). GoS#1 and GoS#2, both 11 years old, displayed clinical signs and a flat ERG trace (indicative of limited or no retinal function) consistent with PRA. GoS#3, at only 6.5 years old had a normal retina, but a reduced ERG trace (indicative of reduced retinal function), suggestive of very early sub-clinical PRA. This dog also had incipient cataracts which could be secondary to PRA. This suggests the onset of retinal degeneration in middle-age and slow progression, a hypothesis that may be confirmed by following the progress of this dog in coming years. Thus eight of the 20 discordant dogs showed signs of early PRA and if these are considered to be affected with PRA, 69.2% of dogs (27/39) homozygous for the mutation (*C2ORF71*^{-/-}), have developed PRA, suggesting that the mutation is highly, if not fully penetrant, assuming the dogs live long enough to develop the disease. The observation that only one of the 47 RCD4 heterozygotes has been diagnosed with PRA, including 21 dogs ranging from 9-14 years of age when examined, and the analysis of pedigrees (Figure 36) is supportive of a recessive mode of inheritance. The RCD4 heterozygote is included in the second group of discordant dogs which comprises three PRA-affected dogs that are not *C2ORF71*^{-/-}. Interestingly these dogs have a younger age at diagnosis (less than 4.5 years) than the average of 9.8 for all other cases. This suggests a locus other than *C2ORF71* may be associated with PRA in these dogs.

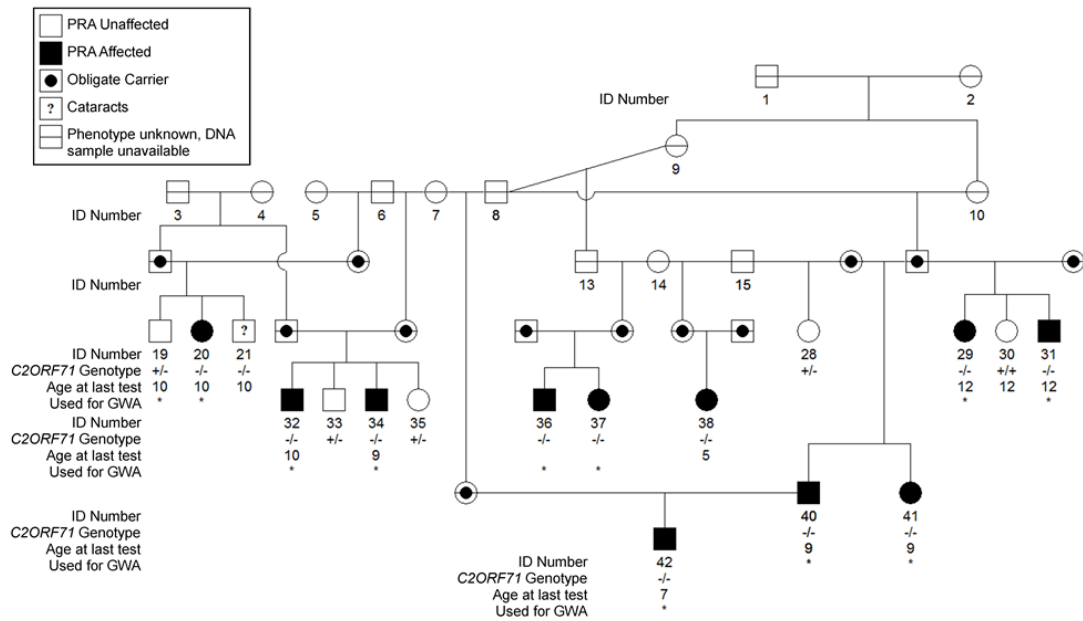


Figure 36. Segregation of RCD4 in a GoS pedigree

The segregation of RCD4 is consistent with an autosomal recessive mode of inheritance. A single dog (#21) is homozygous for RCD4 but has not been diagnosed with PRA. This dog does have cataracts which prevented fundus examination and could be secondary to undiagnosed PRA. The RCD4 mutant allele is represented by “-” and the wildtype allele by “+”.

IrSs and IRWSs are now distinct breeds, although they share ancestors and were once considered coat colour variants of a single breed. They are therefore closely related to one another, and to GoSs, and as a result are the breeds most likely to share the *C2ORF71*_{c.1349_1350insC} mutation. Screening of 72 IrS and 19 IRWS dogs, including ten and three late-onset, non-rcd1, PRA cases respectively, revealed that the mutation is present in IrSs, but is absent from IRWSs. *C2ORF71* genotypes of 70.0% of the IrS PRA cases and 91.9% of clinically unaffected dogs (which could be clear of the mutation or carry a single copy) were concordant with their clinical status. The discordant dogs could be explained with similar arguments to the GoS above. Four of the five PRA-unaffected dogs genetically affected with RCD4 (*C2ORF71*^{-/-}), were not examined beyond their sixth birthdays, and would therefore probably not yet have developed PRA. Subsequent to these findings the fifth dog was examined ophthalmoscopically at the age of 13 (IrS#1, Table 20), by which time she had been clinically blind for 2 years. None of the changes observed were sufficient to cause complete blindness, and were also insufficient to be diagnosed as PRA. However, due to the restricted view of the fundus it is possible that further PRA-associated changes were missed. In addition a confirmatory ERG was not available.

As the ES breed is part of the Setter family, along with the GoS and IrS breeds, the detection of RCD4 in this breed was anticipated. RCD4 was only detected in ESs from Norway. However, the 36 ESs that make up the remainder of the cohort are all from related UK breeding lines with only seven unique sires and ten unique dams between them, and none were clinically affected with PRA. These samples therefore provide insufficient evidence to infer the absence of RCD4 from the UK ES population.

The detection of RCD4 in TTs and SPs was surprising considering these breeds are relatively distantly related to one another and to the Setters²⁴. However, this is not an unprecedented scenario in dogs as *prcd* has been detected in many diverse breeds⁴⁷. One hypothesis to explain the detection of RCD4 in diverse breeds is the presence of a mutation hotspot, in which the same effective mutation, the insertion of a C somewhere in a poly(C) tract, arose more than once as independent events. The detection of the same alleles, 315 and 295 of markers CFA17_25.99 and CFA17_26 respectively, inherited with *C2ORF71*_{c.1349_1350insC} mutation in all five breeds (Figure 35) is in conflict with this hypothesis. Rather the evidence indicates a single mutation event occurred on a chromosome containing alleles 315 and 295, prior to individual breed formation. One hypothesis is that the “315 295” haplotype was common in the population when the *C2ORF71*_{c.1349_1350insC} variant arose, resulting in copies of this haplotype that do not carry the mutation. This is demonstrated in the IrS and ES dogs genotyped, and suggests that the mutation may have originated in ancestors of the IrS and ES breeds, prior to dispersal to other breeds.

The mutation allele frequency in four of the five breeds screened is unusually high for a recessive disease allele (Table 18). These frequencies are likely over-estimates, as research sample cohorts are subject to sampling bias and therefore are unlikely to be representative of the wider population. This hypothesis is supported by the observation that only the IrS and unrelated GoS cohorts conform to HWE. Nevertheless these data suggest the mutation is common in all four breeds.

The data obtained from the AHT Genetics Services cohorts (Table 18), while also unlikely to be completely representative of breed populations, is arguably more so than research cohorts, particularly as a larger dataset is available. The mutation allele frequency in GoSs of 0.33, much lower than that of the research cohort, indicated that 1-in-9 GoSs worldwide are genetically affected with RCD4 and almost half (0.44) carry the mutation. The mutant allele frequency in IrSs of 0.25 indicated that 1-in-16 IrSs worldwide are genetically affected with RCD4 and more than one third (0.38) carry the

mutation. In addition the GoS allele frequencies observed in 15 different countries are similar, ranging from 0.16-0.5. IrS allele frequencies observed in 10 different countries (excluding the USA) range from 0.01-0.33. Interestingly for both breeds the highest allele frequencies are observed in European countries, followed by Australasia and North America. The presence of the mutation in TTs (with an allele frequency of 0.22 in the UK), SPs and setter breeds, as well as the results of microsatellite marker genotyping, suggests it may have arisen early in, if not prior to the development of the breeds, before geographically distinct populations were formed. This could also explain the low frequency of *C2ORF71*_{c.3149_3150insC} in IrSs from the USA: the US population may have been founded with a small cohort of dogs in which the mutant allele was absent or very rare. However the apparent absence of *C2ORF71*_{c.1349_1350insC} from the IRWS breed is surprising. While the mutation has not been found in the population, and does not account for the three late-onset PRA cases tested (with an average age of diagnosis of 10 years), the possibility that RCD4 may be present in the IRWS population cannot be ruled out as the number of cases tested was very small. Alternatively the RCD4 allele may have been lost when the breed underwent a severe bottleneck towards the end of the 19th century, reaching the brink of extinction. The extraordinarily high mutant allele frequency in most breeds could be a result of a lack of selection against the disease as dogs would only begin to show clinical signs well after the normal breeding age. The late age of onset may also account for the diverse range of breeds affected. It is possible, if not highly likely, that many older dogs go blind due to late-onset PRA, potentially RCD4, but a diagnosis is not sought as the loss of sight is attributed to old age. It is also possible that RCD4 segregates in additional breeds but has thus far gone undetected.

Interestingly, this mutation still does not account for all cases of PRA in the GoS or non-rcd1 PRA in the IrS breeds, indicating the presence of at least one additional mutation for PRA in each breed that remains to be identified. This additional variant may be extremely rare in GoS and IrS dogs and as such it will be difficult to collect a sufficient number of cases to conduct a GWA study and alternative approaches need to be considered (discussed in Chapter 6).

4.7 Conclusion

PRA in the GoS has not previously been associated with any genetic mutations. Using a GWA analysis approach, a novel candidate mutation, c.3149_3150insC, was identified in the *C2ORF71* gene that is likely to represent a causal mutation for PRA in the GoS and IrS. While this mutation does not account for all cases of PRA in this study, suggesting that there are additional loci causing PRA in these breeds, it does appear to be highly penetrant and a major cause of PRA in these breeds. Given that RCD4 is also present in at least three other breeds it is possible, if not probable, that more breeds are affected. Screening of dogs from as many breeds as possible for RCD4 is therefore warranted. The development and availability of a DNA test will enable breeders to eventually eliminate RCD4 from the breeds affected entirely.

Chapter **5** Tibetan Spaniels

5.1 Introduction and Background

Tibetan Spaniels are fairly healthy as a breed, and in general are currently predisposed to only a few hereditary conditions including PRA, cataracts and entropion¹²¹. The origin of the breed, along with the Lhasa Apso (LA) and Tibetan Terrier (TT) breeds (which are also predisposed to PRA) can be traced back to ancient Tibetan monasteries¹²². Apart from the small number of clinically affected TTs that are homozygous for the RCD4 mutation that have been identified during this study as (as described in Chapter 4), the mutations responsible for PRA in any of these Tibetan breeds have yet to be identified. While initial investigations will focus on TSs, due to their similar origins it is possible all three breeds share an ancestral mutation causing PRA.

5.2 Diagnoses and Clinical Signs

Typical ophthalmoscopic signs observed in TSs with PRA include tapetal hyper-reflectivity, attenuated blood vessels and optic nerve atrophy. The average age at diagnosis in our sample cohort, which is indicative of age of onset, is 4.69, but ranges from 1 to 11 years of age. The wide range of ages-at-diagnosis could be due to a cohort comprising genetically different forms of PRA or a single form of PRA with a variable phenotype.

5.3 SNP Genotyping

5.3.1 Genome-Wide Association Mapping

GWA analysis (section 2.3.2) of genotyping data from 32 TS dogs (22 cases and 10 controls over the age of four when last examined) genotyped with 15,674 SNPs (section 2.3.1) revealed a genome-wide significant association on chromosome 10 (CFA10; $p_{\text{raw}} = 1.77 \times 10^{-7}$, $p_{\text{genome}} = 0.004$). Two SNP markers 1.86 Mb apart (BICF2P729624 at 62.0 Mb and BICF2S23250878 at 63.86 Mb) were equally the most associated with PRA. IBS clustering using genome-wide SNP marker data confirmed the presence of population stratification with a genomic inflation factor > 1 ($\lambda = 1.69$ based on the median χ^2) and MDS revealed an even distribution of cases and controls (Appendix I). The inflation factor was reduced to an acceptable level ($\lambda = 1.06$) after correcting for population stratification with a CMH meta-analysis. While the signal on CFA10 ($p_{\text{raw}} = 2.01 \times 10^{-5}$, Figure 37A) dropped below the level of Bonferroni-corrected significance, the permutation-corrected signal remained statistically associated ($p_{\text{genome}} = 0.014$). Alternative analysis of the data using FMM to correct for population stratification revealed similar results, with the strongest signal on CFA10 ($p_{\text{raw}} = 5.67 \times 10^{-5}$, Figure 37B) and a reduced inflation factor ($\lambda = 1.27$). Using both CMH and FMM corrections the most associated SNP was BICF2S23422025 at 66.74 Mb. The signal on CFA10 was at least 10-fold stronger than a number of signals seen on other chromosomes including 1, 8 and 38. While only a single SNP (BICF2S23422025 at 66.74 Mb) was statistically significantly associated, the signal on CFA10 extended from 62 to 67.37 Mb (Figure 37C). The use of only ten controls would have reduced the power of the GWA analysis. Additional controls would therefore probably have resulted in a stronger statistical signal.

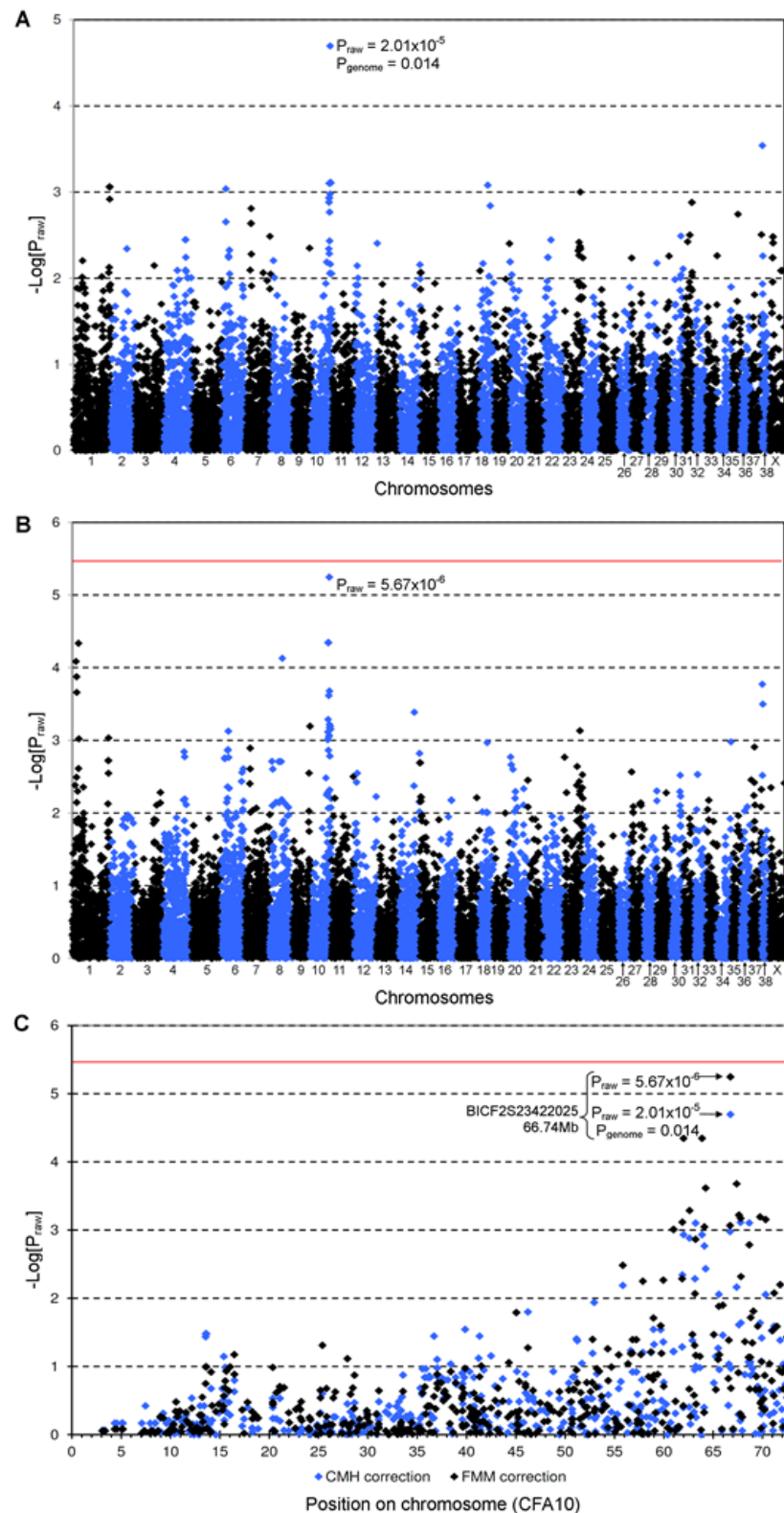


Figure 37. Genome-wide association mapping of PRA in Tibetan Spaniels

$-\text{Log}_{10}$ of p-values after correction for population stratification. The red lines indicate the Bonferroni-corrected 5% significance level based on 15,674 SNPs. A) The CMH meta-analysis approach shows the strongest signal on CFA10 ($p_{\text{raw}} = 2.01 \times 10^{-5}$, $p_{\text{genome}} = 0.014$). B) The FMM approach also shows a prominent signal on CFA10 ($p_{\text{raw}} = 5.67 \times 10^{-6}$). C) The signal spans a region of 5.37 Mb from 62 to 67.37 Mb on CFA10.

5.3.2 Homozygosity Analysis and Microsatellite Finemapping

A haplotype homozygous in cases, but not in controls, could not be easily identified through homozygosity analysis (data not shown). The most highly-associated SNP in CMH and FMM analyses, BICF2S23422025 ($p_{\text{genome}} = 0.014$) is homozygous (A/A) in most of the cases (19/22), but also in 3/10 controls. The SNP20 BeadChip contains a relatively small number of SNPs resulting in a low genotyping resolution – one SNP approximately every 114 kb if all SNPs are informative. In the case of the TS cohort, only 15,674 SNPs were informative, resulting in 1 SNP approximately every 159 kb on average. This made it difficult to identify homozygous haplotypes from the SNP data alone so additional microsatellite markers from the region were genotyped to provide additional haplotype information (Figure 38). Due to sample availability constraints (low DNA quantity) only 18 of the original 22 cases and five of the original ten controls (one of which is an obligate carrier) used in the GWA study were used in the microsatellite marker genotyping study. Even with the increased resolution it was not possible to define a haplotype that was homozygous in all cases. A number of factors may contribute to this: there may be more than one form of PRA in the breed and the cases are therefore heterogeneous; alternatively the causal mutation may be within a haplotype too small to be easily identified at the current marker resolution. Due to the uncertainty of the precise boundaries, a broad critical region of 3.794 Mb, from 63.935 Mb to 67.729 Mb on CFA10 was identified. This region is almost completely homozygous in most of the cases (12/18) and none of the controls, but heterozygous in the obligate carrier. This region contains 31 genes, 29 of which have human orthologues (Figure 38B and Appendix IV).

At the time this work was undertaken, none of the genes in the region could be identified as strong functional candidates. However RP28 had been mapped to a locus in humans (2p14-15)^{166,167}, and part of the genomic region in dogs syntenic to the RP28 locus overlapped with the PRA critical region identified here. Due to the lack of identifiable candidate genes, sequencing of the whole critical region with next generation sequencing was chosen for further investigation of the TS_PRA locus. Subsequently, mutations in *FAM161A* have been associated with RP28 in humans and this gene was therefore identified as an excellent candidate gene for further investigation^{168,169}. None of the other genes in the region are known to be involved in ocular or retinal function or are associated with vision loss in any species.

5.4 Sequencing of the TS_PRA locus on CFA10

5.4.1 Target Enrichment of the TS_PRA Critical Region

Custom-made RNA baits were designed based on the CanFam2.0 reference sequence (section 2.4.2.1), to cover the 3.794 Mb critical region (CFA10:63-68 Mb). As repeat masking was applied to avoid repeat-rich regions, baits were designed over 2.723 Mb, 54% of the total region with an average GC content of 41.0%.

Target enrichment of 10 gDNA samples (four cases, two obligate carriers and four controls) with the custom RNA baits was performed (section 2.4.2.2) and the captured libraries for each sample had concentrations ranging from 1.71 nM to 13.22 nM. Samples were pooled and the final library contained 54.5 µL captured DNA at a concentration of 3.7 nM. While this is less than the 10 nM recommended by the manufacturer, it proved to be sufficient for successful sequencing.

5.4.2 NGS Data Analysis – Target Capture Efficiency

More than 280 million reads were generated in total (section 2.4.2.5), for all 10 sample sequences combined. Approximately 30% of these reads were PCR duplicates and therefore removed from further analysis. A total of >190 million reads remained, 96% of which mapped to CanFam2.0. Approximately 72% of bases mapped to the targeted regions, and almost 65% of all reads had >30x coverage (Table 21). Target capture efficiency is represented by the percentage of bases that map to the targeted regions, and was relatively similar between the 10 samples processed, with an average of 71.8%. Baited regions also had an average read depth of 236 reads.

Table 21. Capture efficiency for TS_PRA targeted NGS

Sample no.	Clinical Status	Mean coverage of baited region (X)	Bases mapped to targets (%)	Bases with >10X coverage (%)	Bases with >20X coverage (%)	Bases with >30X coverage (%)
1	Affected	289	74.68	72.54	69.06	66.33
2	Affected	273	73.68	72.16	68.59	65.82
3	Affected	248	74.42	71.82	68.24	65.39
4	Affected	253	74.76	71.98	68.38	65.61
5	Carrier	201	72.31	71.49	67.39	64.10
6	Carrier	248	73.90	70.43	66.68	63.83
7	Normal	169	69.98	69.85	65.31	61.75
8	Normal	238	71.93	73.36	69.37	66.25
9	Normal	275	72.33	72.15	68.63	65.97
10	Normal	168	59.66	70.10	65.65	62.08
	Average	236	71.77	71.59	67.73	64.71

5.4.3 NGS Data Analysis – Variant Identification

More than 19,000 SNPs and 3,700 indels were identified when the sequence was compared with the CanFam2.0 reference sequence (section 2.4.2.5). Of these, 194 SNPs and 81 indels segregated with the phenotype, but none of these variants were predicted to alter the protein product. Visual analysis of sequence data in IGV revealed 16 additional variants (larger insertion and deletions not identified using the NGS analysis pipeline), of which seven segregated with the phenotype. One of these variants is an insertion flanked by a 14 bp repeat motif (Figure 39). This repeat is visualised in IGV as an increase in the read depth. The length of the inserted sequence is longer than the length of the NGS reads (>50 bp) and the precise sequence of the insertion could therefore be only partly determined (Figure 39B). Only this variant, which was predicted to be located near a splice acceptor site of the *FAM161A* gene (CFA8: 64,974,130; Figure 39), could potentially alter the protein product, by interfering with exon splicing.

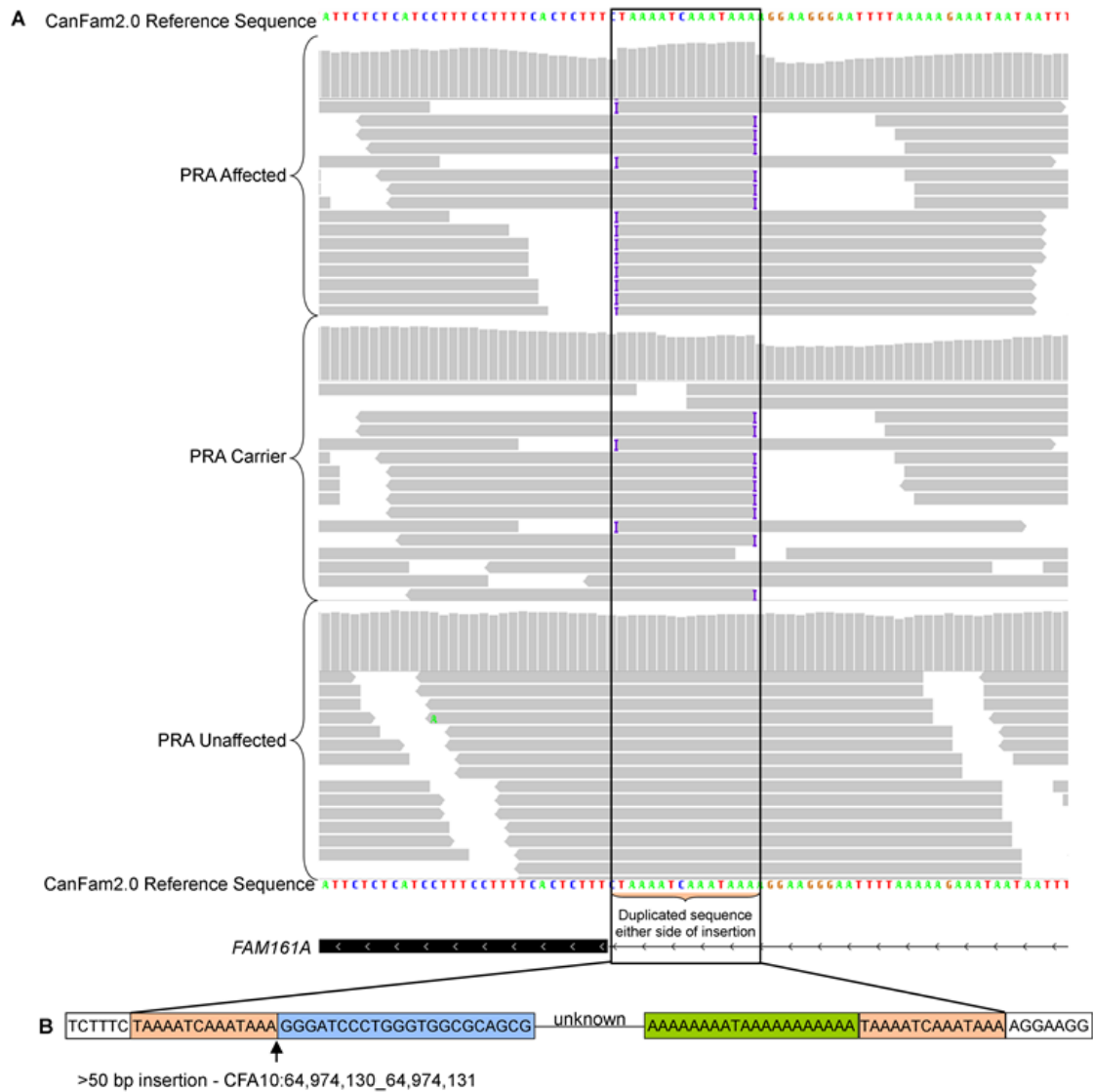


Figure 39. IGV display of the >50 bp insertion in *FAM161A*

A) Each of the three samples (PRA-affected, obligate carrier and control) viewed in IGV are represented by two panels. The upper panel is a histogram where the height of each column is representative of the read depth at that location. The increased read depth in the affected panel is characteristic of a duplication, caused by the repeat motif flanking the insertion. The lower panel is a graphical view of a subset of the reads that align to that location. The sudden termination of reads and the insertion symbol (I) either side of the duplicated sequence is also characteristic of an insertion flanked by the duplicated sequence. The inserted sequence is present in all reads in the PRA-affected sample, approximately half the reads in the obligate carrier and none of the reads in the PRA-unaffected (control) sample. B) Inserted sequence (blue and green) as determined from NGS data, flanked by 14 bp repeats (orange).

5.4.3.1 Sanger Sequencing of *FAM161A* Insertion

The full sequence of the insertion was determined by Sanger sequencing (sections 2.4.1.3-2.4.1.8) using primers flanking exon 5 of *FAM161A*, including the insertion site in gDNA from 80 TS dogs (29 affected with PRA, 10 obligate carriers and 41 unaffected). Using agarose gel electrophoresis, a single band of the expected size (720 bp) was visible for 40 unaffected samples, while a band approximately 230 bp larger (~950 bp) was visible for 17 of the PRA affected samples and none of the carrier or unaffected samples (data not shown). Carriers of the insertion were more difficult to identify from gel electrophoresis alone (the larger band was not as bright as the smaller band, and sometimes not visible at all), presumably due to preferential amplification of the shorter DNA fragment lacking the insertion. However, carriers were identifiable by Sanger sequencing. Sequencing of the ~230 bp insertion revealed that it contains a 132 bp short interspersed nuclear element (SINE), a retroposon that is distributed widely throughout the canine genome¹⁷⁰. As is characteristic, the SINE is followed at the 3' end by a dinucleotide repeat, (CT)₈, and a poly(A) tract (interrupted by the occasional T) at least 45 bp in length. The nucleotides at the 3'-end of the poly(A) tract are duplicated at the 5'-end of the SINE (Figure 40). The number of adenine nucleotides that comprise a portion of the poly(A) tract (underlined in Figure 40A) could not be determined accurately due to difficulties amplifying homopolymers with synthetic taq polymerases, specifically polymerase slippage along the poly(A) tract. However, based on the sequence traces, there appear to be 35-50 adenine nucleotides.

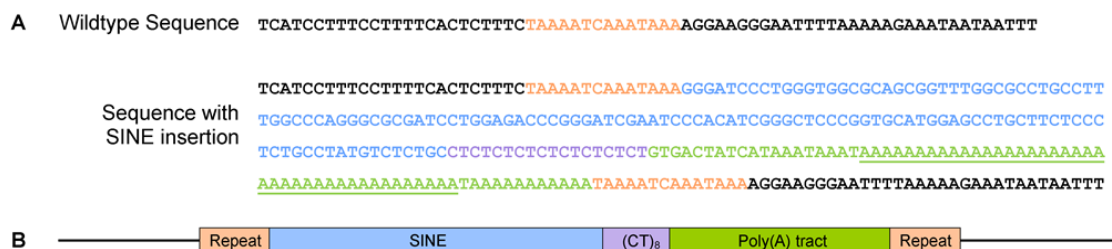


Figure 40. Sequence of SINE insertion

A) The sequence and B) graphical representation of the SINE insertion. The precise number of nucleotides that comprise the underlined portion of the poly(A) tract remains unclear, but is approximately 35-50.

5.4.3.2 Poly(A) Tract Length

An AFLP assay (section 2.2) using primers flanking the SINE insertion was used to further define the length of the insertion and therefore the poly(A) tract. As a result of Taq polymerase slippage on the poly(A) tract, amplification of the SINE insertion resulted in multiple products varying in size, creating a “hedgehog” effect (Figure 41). The SINE insertion was amplified from 21 dogs that were either homozygous ($n = 14$) or heterozygous ($n = 7$) for the SINE insertion, as determined by sequencing (Section 5.4.3.1). For most dogs (15/21), the highest peak was at 391 bp (Figure 41A). The size of the wildtype amplicon is 153 bp, and this suggests the SINE insertion is 238 bp in size. Therefore, the poly(A) tract (represented by the underlined text in Figure 40A) is 44 nucleotides in length. In 5/21 samples assayed the highest peak was at 396 bp (Figure 41B), suggesting a poly(A) tract five nucleotides longer i.e. 49 nucleotides. One dog carried a single copy each of the 391 and 396 bp alleles (Figure 41C). The age at diagnosis was known for 12 of the dogs assayed that were clinically affected with PRA (Figure 41). The age at diagnosis of PRA in dogs with only the 391 bp allele ranges from 2.9 to 10.2, but there is insufficient data for the 396 bp allele. These data are consistent with the length of the poly(A) tract having no effect on the age at diagnosis.

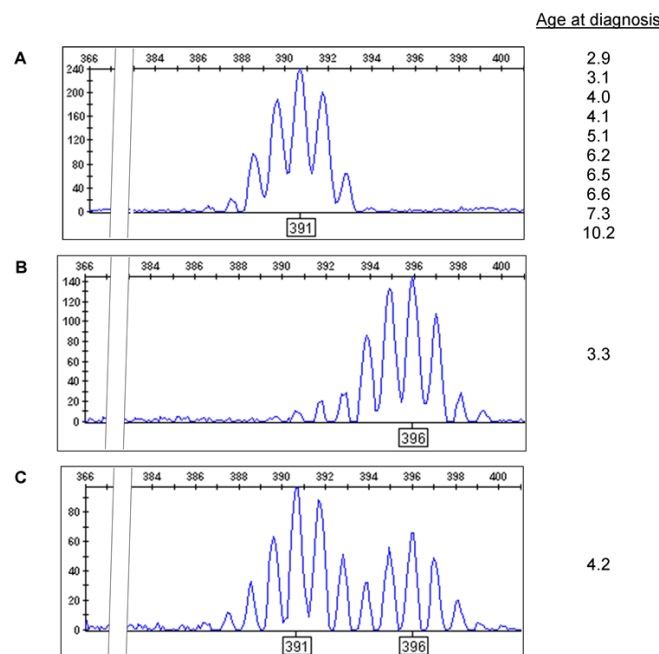


Figure 41. AFLP assay of SINE insertion

Amplification of the SINE insertion resulted in multiple products. A) The highest peak was at 391 bp in most dogs, and the age at diagnosis ranged from 2.9 to 10.2 years. B) In some dogs the highest peak was at 396 bp. One of these dogs was diagnosed at 3.3 years of age. C) A single dog carried both allele 391 and allele 396, and was diagnosed at 4.2 years of age.

5.4.4 Characterising *FAM161A* Retinal Transcripts

In humans *FAM161A* occurs in two main isoforms, full-length (*FAM161A_{fl}*) and short (*FAM161A_{sh}*), formed by alternative splicing of exon 4 (Figure 42)^{168,169}. Four incomplete canine *FAM161A* isoforms were predicted by Ensembl, but alignment with human and mouse genomic and coding sequences (section 2.4.1.1) revealed several inconsistencies and possible errors in the prediction of canine intron-exon boundaries (Figure 42A-C). Most of the coding sequence of the *FAM161A* retinal transcripts (section 2.1.2.2), from both the main isoforms (*FAM161A_{fl}* and *FAM161A_{sh}*), was successfully sequenced (sections 2.4.1.2-2.4.1.8) in a healthy dog, excluding the first 46 nucleotides of the coding sequence (Figure 42D and E). Sequencing revealed that both isoforms are transcribed in the canine retina. In addition, intron-exon boundaries are identical to those of the human and mouse, which is in conflict with the boundaries predicted by Ensembl genebuild for the canine gene. Canine retinal mRNA sequence differed from CanFam2.0 reference sequence in exon 1, and two variants were identified: the first was a SNP (C→A) that changes an alanine residue to an aspartic acid residue; the second was a 6 bp insertion resulting in the in-frame insertion of two alanine residues (Figure 42E). To assess the prevalence of these variants, gDNA from 43 TSs and 76 dogs comprising 31 breeds were sequenced. The SNP variant was present in the homozygous and heterozygous state in both cohorts (i.e. TSs and multiple breeds). Eight dogs from five breeds, excluding TSs, were heterozygous for the insertion variant and none were homozygous. It is unclear what effect, if any, these variants have on the protein, but these data suggest they are unlikely to be pathogenic. Sequencing of the full 5'- and 3'-UTRs and the beginning of exon 1 was unsuccessful, probably due to high GC content. Sequencing revealed that canine *FAM161A_{fl}* contains 716 amino acids and *FAM161A_{sh}* contains 660 amino acids, with predicted molecular weights of 83kDa and 76kDa respectively. The SINE insertion occurs near the acceptor splice site of intron 4 i.e. near the boundary of intron 4 and exon 5 (*FAM161A_{fl}*: c.1758-15_1758-16ins238; *FAM161A_{sh}*: c.1590-15_1590-16ins238) (Figure 42D).

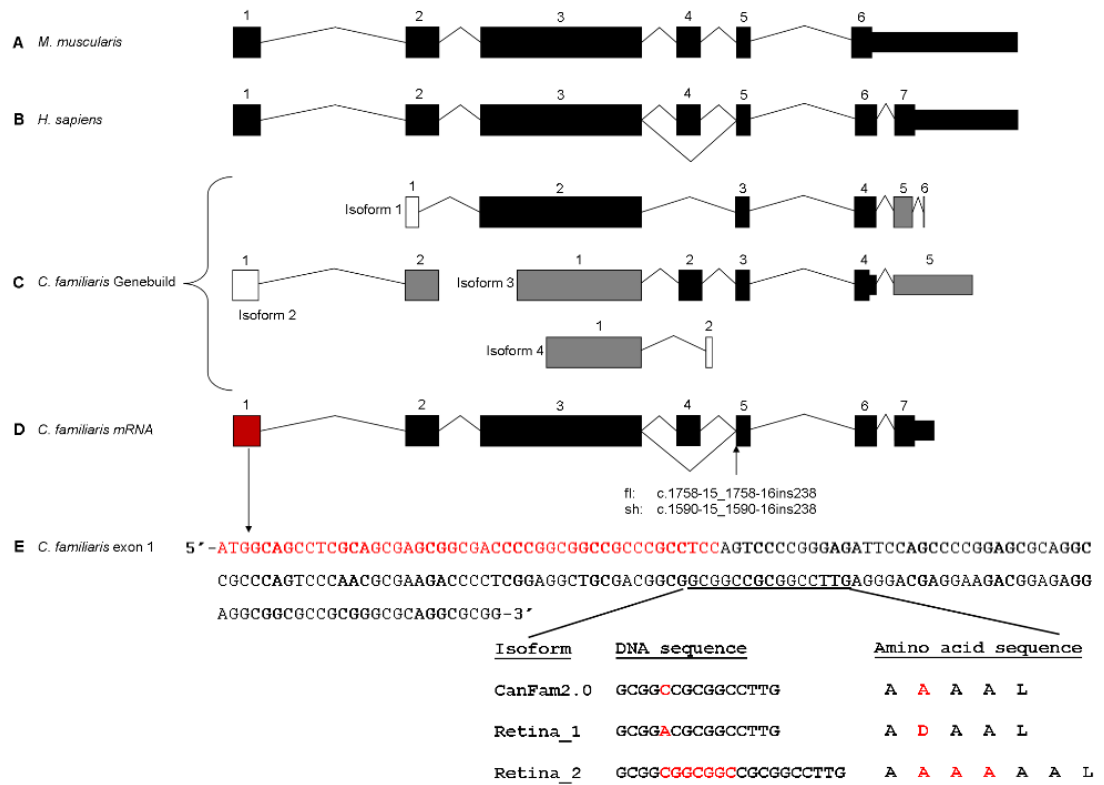


Figure 42. Graphical comparison of the intron-exon boundaries of *FAM161A*.

A) Mouse (*Mus musculus*) *FAM161A*. B) Human (*Homo sapiens*) *FAM161A*. C) Canine (*Canis familiaris*) *FAM161A* as predicted by Ensembl genebuild. Six genebuild exons (black) are identical to the mouse and human exons, while the intron-exon boundaries for five exons (grey) are inconsistent and four exons (white) bear little or no resemblance to human and mouse exons. D) Canine *FAM161A* exons confirmed by sequencing the retinal mRNA transcript. Exon 1 (red) was only partially sequenced. The location of the splice site insertion is indicated. E) Only the 3'-end of exon 1 (black) was successfully sequenced. Two alternative alleles were identified in retinal mRNA that differed from the CanFam2.0 reference sequence.

5.4.5 *FAM161A*_{c.1758-15_1758-16ins238} Variant Validation

5.4.5.1 AFLP Genotyping Assay

The straightforward AFLP assay used in section 5.4.3.2 (i.e. amplification of the region surrounding the SINE insertion with a pair of flanking primers, and comparison of amplicon lengths to determine the presence or absence of the SINE; Figure 43A) proved inefficient as a genotyping assay. Specifically, carriers of *FAM161A*_{c.1758-15_1758-16ins238} in particular, but also homozygous dogs, could not always be identified with confidence, for two reasons (Figure 43B):

- 1) The length of the wildtype and mutant amplicons differed by the length of the SINE insertion (~238 bp). As a result, preferential amplification of the smaller amplicon resulted in a far stronger fluorescence signal for the wildtype allele than for the mutant allele.
- 2) As discussed in section 5.4.3.2, these primers generate multiple products due to Taq polymerase slippage, resulting in a “hedgehog” effect. This further diluted the fluorescence emitted by the mutant allele, which often could not be distinguished with confidence from background fluorescence levels.

To create the most efficient and accurate assay possible three primers were used instead of two (Figure 43C; sections 2.2 and 2.6): A single, fluorescently-labelled forward primer, a wildtype-specific reverse primer and a mutant-specific reverse primer. These primers do not flank the poly(A) tract and result in amplicons of similar size (135 bp and 137 bp), thereby generating wildtype and mutant amplicons that emit roughly equal fluorescence (Figure 43D). These primers were used for all subsequent *FAM161A* genotyping.

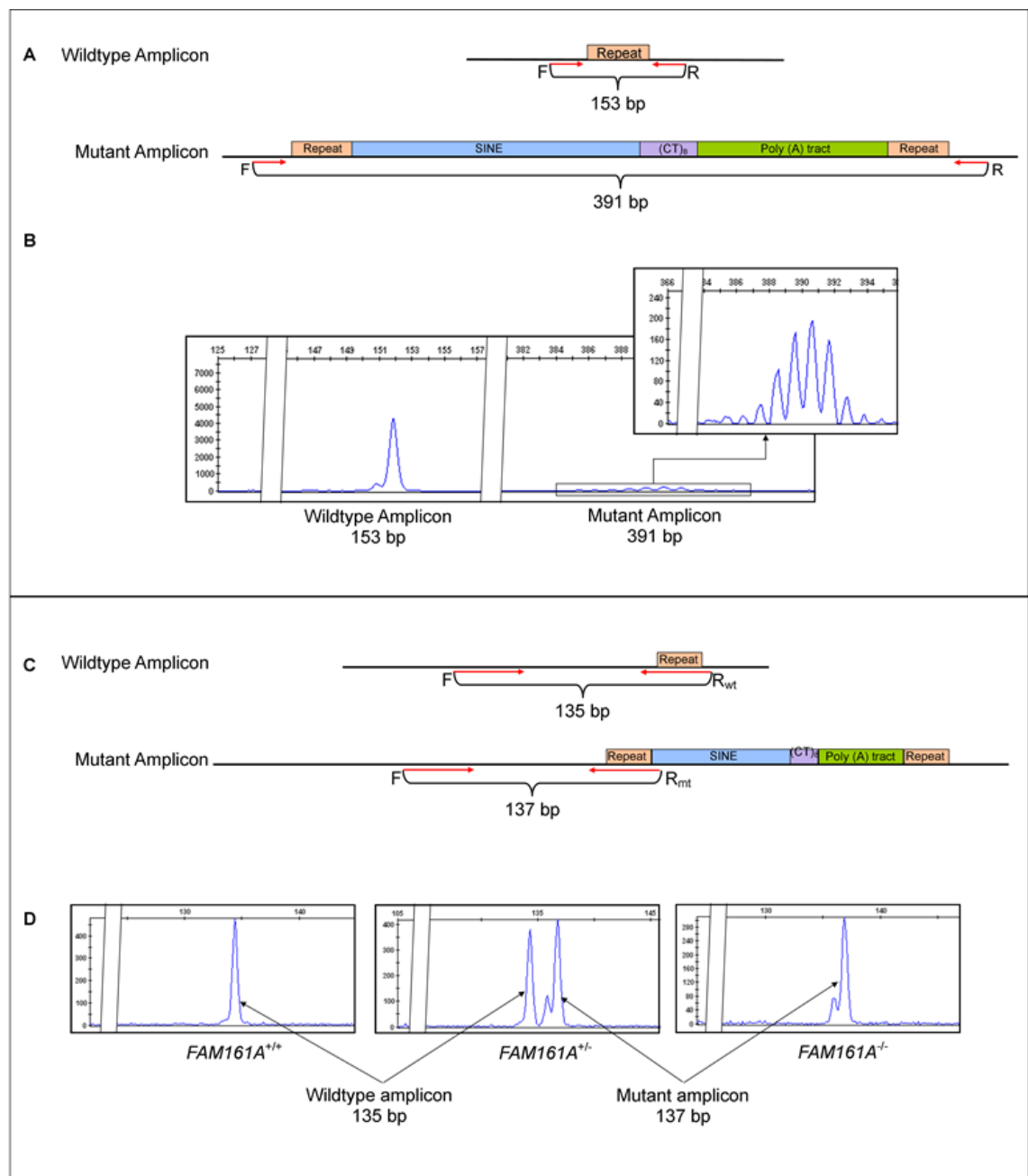


Figure 43. *FAM161A*_{c.1758-15_1758-16ins238} genotyping assay

A) A straightforward AFLP assay using flanking primers (red arrows) proved inefficient. B) The fluorescence signal emitted by the mutant amplicon (approximately 391 bp) was always far lower than that emitted by the wildtype amplicon (153 bp). C) The use of allele-specific reverse primers (wildtype = R_{wt}, mutant = R_{mt}) resulted in amplicons of similar length (135 bp and 137 bp respectively). D) The wildtype and mutant amplicons also emitted roughly equal fluorescence. Therefore dogs homozygous for the wildtype allele (135 bp, *FAM161A*^{+/+}), heterozygous (*FAM161A*^{+/-}) and homozygous for *FAM161A*_{c.1758-15_1758-16ins238} (137 bp, *FAM161A*^{-/-}) could be reliably identified.

5.4.5.2 *FAM161A*_{c.1758-15_1758-16ins238} Genotyping

All 32 TS dogs (22 cases and 10 controls) that participated in the GWA study were screened for the SINE insertion, *FAM161A*_{c.1758-15_1758-16ins238}, to confirm the association of this variant with PRA and compare it with two of the most highly associated SNP markers, BICF2S23422025 and BICF2S23250878. *FAM161A*_{c.1758-15_1758-16ins238} showed significant allelic association with PRA ($p_{\text{raw}} = 5.03 \times 10^{-7}$). The SNP markers also showed significant allelic association, although BICF2S23422025 was less associated ($p_{\text{raw}} = 6.28 \times 10^{-7}$). However, BICF2S23250878 was more highly associated ($p_{\text{raw}} = 1.77 \times 10^{-7}$) than *FAM161A*_{c.1758-15_1758-16ins238}, but this can be attributed to two PRA cases that are heterozygous for the SNP (i.e. carry the minor allele) but homozygous for the wildtype *FAM161A* allele. Fifteen out of 22 PRA cases and none of the controls were homozygous for *FAM161A*_{c.1758-15_1758-16ins238}. Analysis of the segregation of *FAM161A*_{c.1758-15_1758-16ins238} with PRA (Figure 44) indicates that the form of PRA associated with this variant is recessive and fully penetrant. The form of PRA that is associated with *FAM161A*_{c.1758-15_1758-16ins238} is known hereafter as PRA-3.

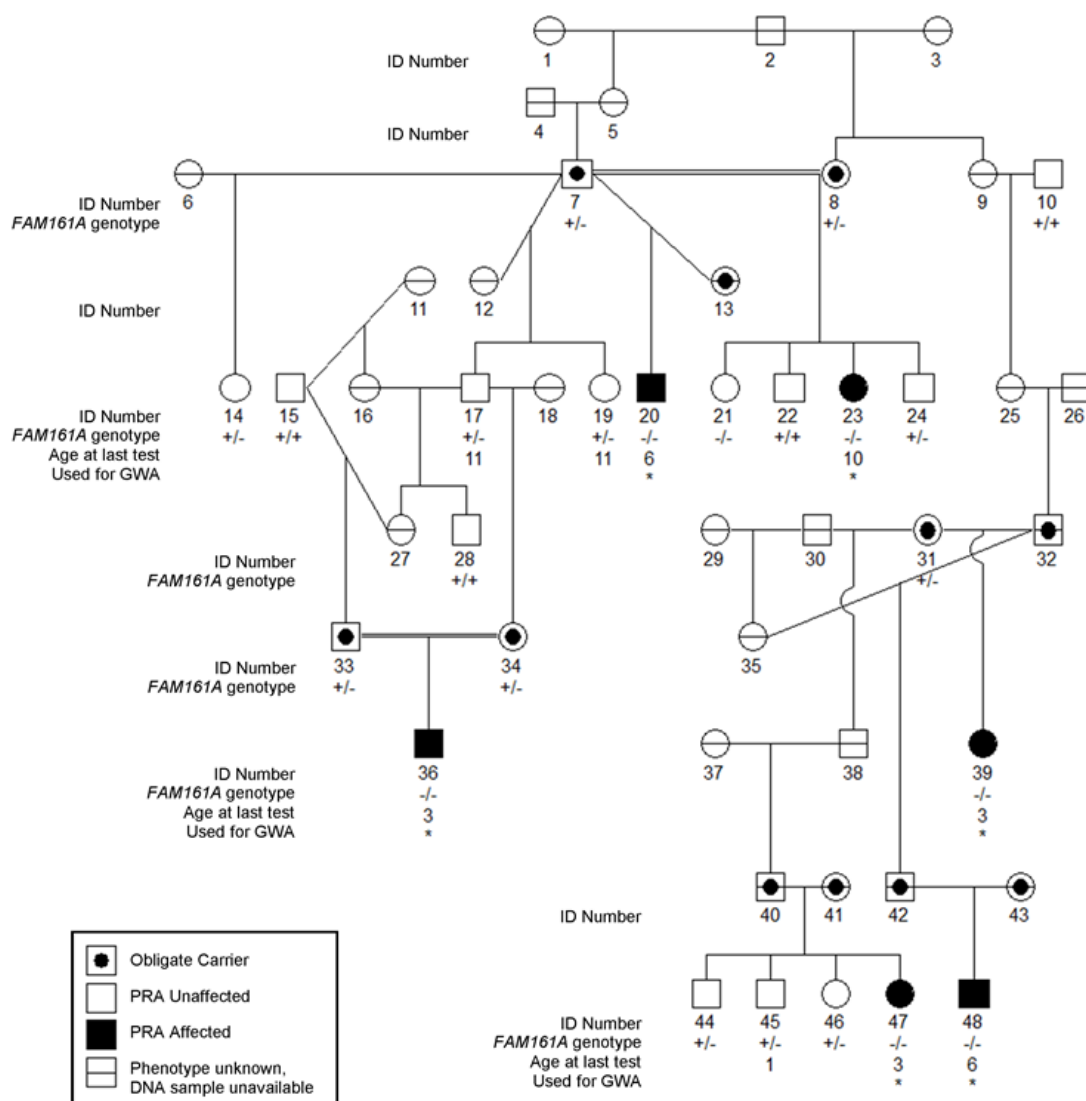


Figure 44. Segregation of PRA-3 in a TS pedigree

The segregation of PRA-3 is consistent with an autosomal recessive mode of inheritance. The PRA-3 mutation ($FAM161A_{c.1758-15_1758-16ins238}$) allele is represented by “-“ and the wildtype allele by “+”. Clinical information pertaining to and DNA samples from 26 individuals was not available. Clinical information about a single dog (#21) that was homozygous for $FAM161A_{c.1758-15_1758-16ins238}$ was unavailable.

5.4.5.3 Sanger Sequencing of Blood and Retinal mRNA Transcripts

The location of the SINE insertion near the splice acceptor site of exon 5 suggests exon splicing may be affected, possibly resulting in the skipping of exon 5. To assess this hypothesis, mRNA transcripts were compared between a TS dog homozygous for the SINE insertion and two dogs of unknown breed. In the absence of suitable retinal tissue, RNA was purified from the blood (section 2.1.3.4) of the affected and one of the unaffected dogs, while retinal tissue was available from the other unaffected dog. Primers in exons 3 and 6 were used to amplify across exons 4 and 5 (Figure 45A; sections 2.2 and 2.4.1.2-2.4.1.8). A number of products were produced for all three samples, each of which was individually sequenced (Figure 45B):

- 1) Bands 6 and 9 comprise the 421 bp amplicon (*FAM161A_{fl}*), and bands 8 and 12 comprise the 253 bp amplicon (*FAM161A_{sh}*). Both of these isoforms are detectable in normal blood and retina, but not affected blood.
- 2) Bands 3 and 11 comprise the 321 bp amplicon (*FAM161A_{fl-5}*) and bands 4 and 13 comprise the 153 bp amplicon (*FAM161A_{sh-5}*). Both of these amplicons are detectable in affected blood. Interestingly these bands are also detectable in normal retina, albeit at lower levels than *FAM161A_{fl}* and *FAM161A_{sh}*, but not unaffected blood.
- 3) Bands 7 and 10 comprise *FAM161A_{fl}-FAM161A_{sh}* hybrid amplicons i.e. sequence consistent with both isoforms was detectable in these bands. These could be attributed to depletion of PCR reagents in the latter stages of thermal cycling, resulting in annealing of the two isoforms with one another instead of with primers. These bands were absent from PCR experiments utilising a reduction in the number of cycles or a 30% increase in the reaction volume, and therefore in the reagents present (data not shown). A high number of cycles ($n = 40$) were necessary in order to visualise PCR amplicons generated from the affected blood.
- 4) Band 1 comprises an amplicon containing *FAM161A_{fl-5}* and intron 3, which could be a result of gDNA contamination. Band 2 comprises an amplicon containing *FAM161A_{sh-5}* and part of intron 5, while band 5 comprises an amplicon containing *FAM161A_{fl}* and two separate parts of intron 5. These two amplicons could be a result of incomplete or inaccurate exon splicing.

These results suggest that the difference between *FAM161A* in PRA-3 affected and normal blood is the absence and presence respectively of exon 5, supporting the exon-skipping hypothesis.

5.4.5.4 Quantitative PCR of *FAM161A* mRNA Transcripts

To further compare the levels of the four possible transcripts present in blood from an affected dog and from a normal dog, relative qPCR was undertaken using SYBR® green chemistry and isoform-specific primers (Figure 45A; section 2.5). Expression levels of targets were normalised against two housekeeping genes, *TBP* and *ACTP*, and compared by relative quantification. Exons 2 and 3 are not known to be affected by any alternative splicing, nor are they expected to be affected by the SINE insertion, and it is therefore reasonable to assume that all *FAM161A* isoforms will contain these exons. Expression levels of these exons therefore constitute 100% of *FAM161A* transcripts present ($FAM161A_{\text{total}}$).

The coefficient of correlation (R^2) and amplification efficiency (E) of PCR using each primer pair was ascertained from standard curves generated for each isoform target (Table 22), which in turn was used to determine expression levels in the blood. R^2 is an indication of how well the data points lie on a single line, or the linearity of the standard curve, and should be > 0.985 . E is the percentage of template that is amplified during each cycle, and should be between 90% and 110%. Many of the PCR reactions have suboptimal R^2 and/or E values, suggesting these assays and subsequent results are potentially unreliable. $R^2 < 0.95$ is indicative of pipetting error or the presence of inhibitory elements. As the only difference between each of the reactions was the primers used, and not all of the assays have poor R^2 values, the presence of inhibitory elements is unlikely. In this instance it is likely the suboptimal R^2 values can be attributed to the low reaction volumes used (6 μl) in each assay. Due to sample availability constraints, and the need to use the same thermal cycling parameters for each assay, the assays could not be individually optimised. These results are therefore preliminary and future investigations using fluorescent probes instead of SYBR® green should be undertaken. These probes, whilst more costly, are more reliable than SYBR® green.

Table 22. Standard curves of qPCR targets

Target	Slope	R ²	Efficiency (%)
TBP	-3.512	0.985	92.630
ACTP	-3.375	0.637	97.837
<i>FAM161A</i> _{total}	-3.563	0.985	90.823
<i>FAM161A</i> _{fl}	-3.551	0.982	91.256
<i>FAM161A</i> _{fl-5}	-5.569	0.791	51.209
<i>FAM161A</i> _{sh}	-3.405	0.828	96.665
<i>FAM161A</i> _{sh-5}	-3.121	0.528	109.132

While bearing in mind the shortcomings of the qPCR results, they do support the sequencing results discussed above (section 5.4.5.3) and aberrant splicing hypothesis (Figure 45). The similar levels of *FAM161A*_{total} between affected and normal blood indicate that mutated *FAM161A* is not subjected to NMD. *FAM161A*_{fl} and *FAM161A*_{sh} are more highly expressed in the normal blood than *FAM161A*_{fl-5} and *FAM161A*_{sh-5}, while in the affected blood the reverse is true (Figure 45C). Taken together, *FAM161A* in affected blood comprises predominantly mutant transcripts lacking exon 5 ($FAM161A_{mt} = FAM161A_{fl-5} + FAM161A_{sh-5} = 84\%$), while *FAM161A* in normal blood comprises predominantly wildtype transcripts ($FAM161A_{wt} = FAM161A_{fl} + FAM161A_{sh} = 71\%$). This supports the hypothesis that the SINE insertion results in skipping of exon 5 during pre-mRNA splicing.

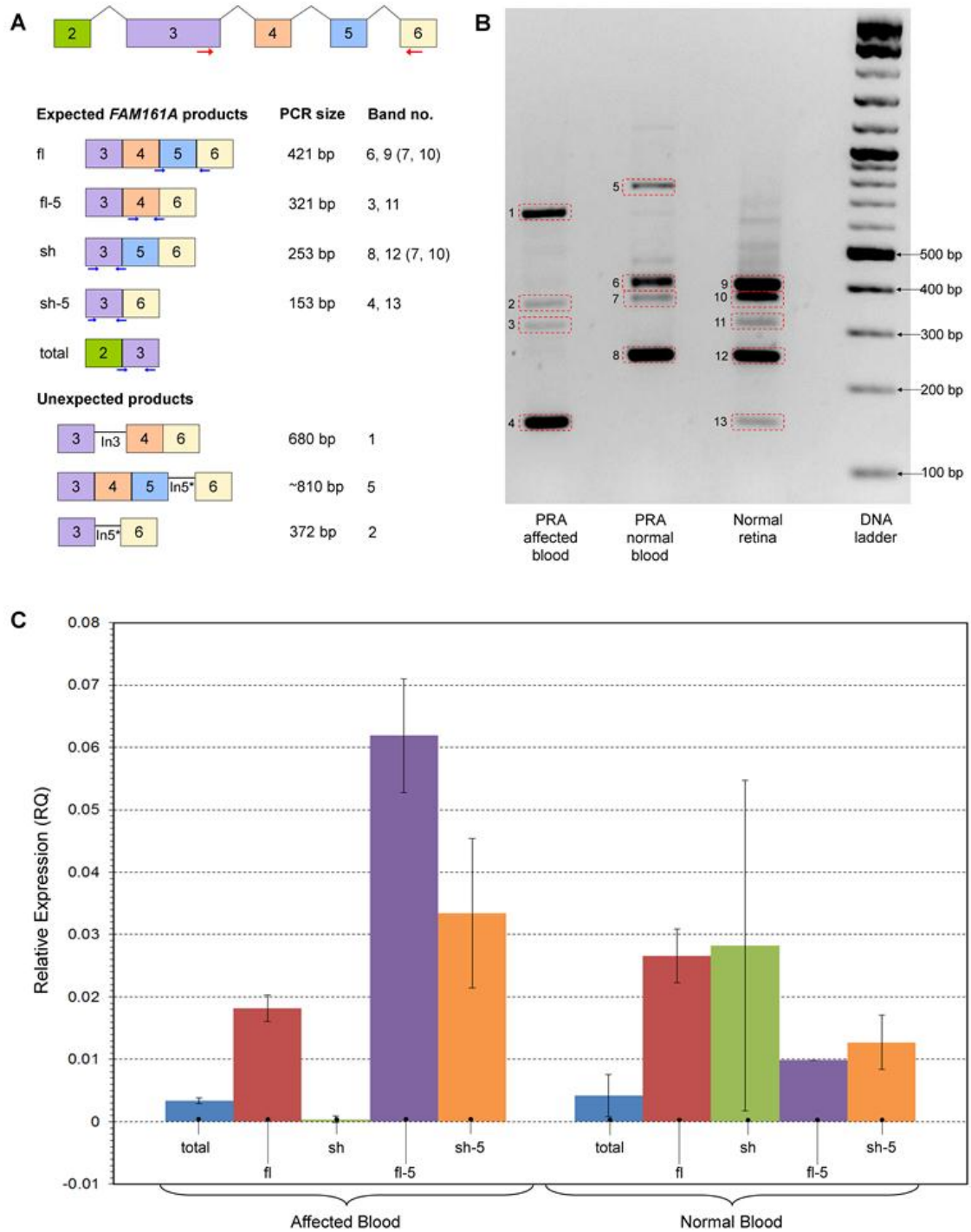


Figure 45. Comparison of *FAM161A* mRNA isoforms

PCR and qPCR to compare expression levels of *FAM161A* isoforms in blood from an affected dog and an unaffected dog. A) Primers in exons 3 and 6 (red arrows) were used to amplify *FAM161A* isoforms created by alternative splicing of exons 4 and 5, resulting in four possible amplicons expected (fl, fl-5, sh, sh-5; sizes indicated). Primers pairs for qPCR (blue arrows) were designed to be specific for each isoform. B) Agarose gel electrophoresis of PCR amplicons. C) Relative expression of each target, determined by qPCR. Error bars represent standard deviation.

5.4.5.5 *FAM161A*_{c.1758-15_1758-16ins238} Effect on Protein

The data presented in sections 5.4.5.3 and 5.4.5.4 support the hypothesis that the SINE insertion near the splice acceptor site of exon 5 of *FAM161A* results in aberrant splicing i.e. the skipping of exon 5 (Figure 46A). It is predicted to shift the reading frame and cause a premature stop codon (p.S588MfsX13) possibly resulting in a truncated protein product (Figure 46C) or degradation of the mRNA by NMD, although experimental evidence for either hypothesis is required.

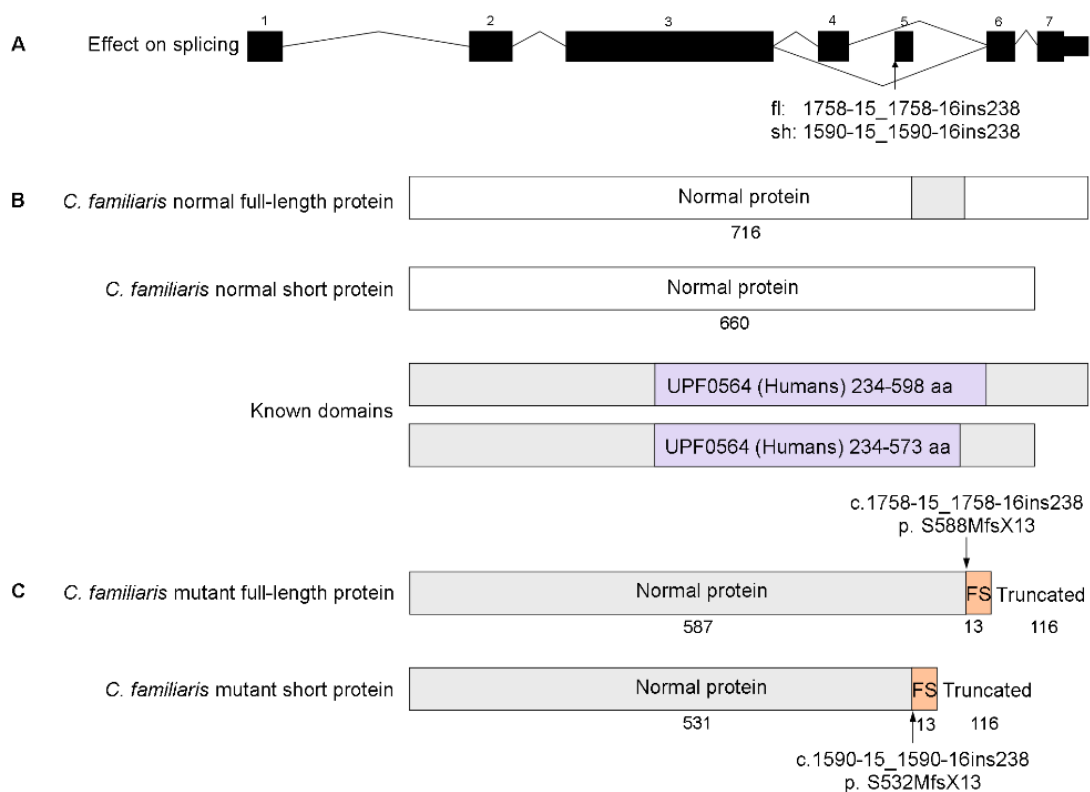


Figure 46. Effect of *FAM161A*_{c.1758-15_1758-16ins238} on the protein product

A) *FAM161A*_{c.1758-15_1758-16ins238} results in aberrant splicing i.e. the skipping of exon 5. B) The normal full-length and short *FAM161A* proteins differ only by the presence and absence respectively of exon 4-generated residues (grey). Domain UPF0564 in humans is critical to the correct functioning of the protein. C) Skipping of exon 5 results in a shift in the reading frame (FS) and a premature termination codon. The number of amino acids comprising each section are indicated.

5.5 *FAM161A*_{c.1758-15_1758-16ins238} (PRA-3) Population Screening

5.5.1 Research Sample Cohorts

To confirm that the variant is not a commonly occurring polymorphism in this breed, we screened 215 additional TS dogs (section 2.6), resulting in a total of 247 TSs tested for *FAM161A*_{c.1758-15_1758-16ins238} (Table 23). Of the 35 PRA cases used in the study 22 (62.9%) were homozygous for *FAM161A*_{c.1758-15_1758-16ins238} (*FAM161A*^{-/-}) and all 116 dogs known to be clinically free of PRA at their last eye examination, including 16 obligate carriers of PRA, were either carriers of the mutant allele (14.7%; *FAM161A*^{+/-}) or homozygous for the wild type allele (85.3%; *FAM161A*^{+/+}). PRA-3 therefore accounts for the majority of cases of PRA in our TS cohort.

Table 23. *FAM161A* genotypes and PRA clinical status for 247 TSs

Genotype	PRA Affected	PRA Carrier	PRA Unaffected	Unknown	Total
<i>FAM161A</i> ^{-/-}	22	0	0	2	24
	62.9%	0.0%	0.0%	2.1%	9.7%
<i>FAM161A</i> ^{+/-}	2	9	8	22	41
	5.7%	56.3%	8.0%	22.9%	16.6%
<i>FAM161A</i> ^{+/+}	11	7	92	72	182
	31.4%	43.8%	92.0%	75.0%	73.7%
Total	35	16	100	96	247

To determine whether *FAM161A*_{c.1758-15_1758-16ins238} is associated with PRA in related breeds we screened a further 99 dogs from two closely related breeds most likely to share polymorphisms with the TS breed. These were 23 LAs and 76 TTs (Table 24), including nine LAs and 12 TTs affected with PRA. All 23 LA dogs, including nine PRA cases, were homozygous for the wild-type allele (*FAM161A*^{+/+}). PRA-3 is therefore absent from this LA cohort, but as the number of dogs tested was small, it cannot be eliminated entirely as a form of PRA in the breed. Of the 12 TTs with PRA, four were homozygous for *FAM161A*_{c.1758-15_1758-16ins238} (*FAM161A*^{-/-}), while the remaining eight PRA cases were either carriers (*FAM161A*^{+/-}; n=1) or homozygous wildtype (n=7). In addition, all TTs known to be free of PRA (n = 10) were homozygous for the wildtype allele (79.7%). PRA-3 is therefore present in the TT breed.

Table 24. *FAM161A* genotypes and PRA clinical status for 76 TTs

Genotype	PRA Affected	PRA Unaffected	Unknown	Total
<i>FAM161A</i> ^{-/-}	4	0	0	4
	33.3%	0.0%	0.0%	5.3%
<i>FAM161A</i> ^{+/-}	1	0	13	14
	8.3%	0.0%	24.1%	18.4%
<i>FAM161A</i> ^{+/+}	7	10	41	58
	58.3%	100.0%	75.9%	76.3%
Total	12	10	54	76

The frequency of the PRA-3 allele is 0.18 in TSs, based on all 247 research samples tested (Table 25). This is not unusually high for an aR disease allele in dogs. However, the cohort screened contained samples collected for the purposes of research (as in Chapters 3 and 4) and was therefore subject to considerable sampling bias. This combined with the relatively small number of samples screened has resulted in populations that do not conform to HWE (Table 25). Analysis of 120 dogs that had unique sires and dams resulted in a reduced allele frequency, but this cohort also does not conform to HWE (Table 25). This suggests that the high degree of relatedness between samples is only one cause of deviation from HWE, and sampling bias is still significant.

Table 25. Population screening for PRA-3.

Breed	Observed Genotypes*					HWE		1-in-__ Expected	
	-/-	-/+	+/+	TOTAL	Allele Freq	χ^2	P	Affected	Carrier
Research Cohort									
Tibetan Spaniels	24	41	182	247	0.180	47.41	5.77×10^{-12}	31	3
Tibetan Spaniels Unrelated	10	16	94	120	0.150	27.32	1.73×10^{-7}	44	4
Tibetan Terriers	4	14	58	76	0.145	4.98	0.03	48	4

* The mutant allele is represented by “-” and the wildtype allele by “+”.

5.6 Discussion

Using a GWA mapping and homozygosity analysis approach, a novel 3.794 Mb locus on chromosome 10 that is associated with PRA in the TS was identified. This region overlapped with part of the region syntenic to the human RP28 locus¹⁶⁶. The entire critical region was sequenced and a single provocative variant was identified in the *FAM161A* gene. *FAM161A* was subsequently identified as a strong candidate causal locus due to the identification of *FAM161A* mutations in humans with RP and the localisation of *FAM161A* to the photoreceptors of the retina^{168,169}.

Sequencing of *FAM161A* from healthy retinal mRNA served three purposes: Firstly, it confirmed the presence of *FAM161A* mRNA transcripts in the normal canine retina. Secondly, it revealed that the intron-exon boundaries predicted by genebuild for *FAM161A* in the dog are incorrect for three exons. They are instead identical to the human and mouse boundaries. Thirdly, as is the case in humans, canine *FAM161A* is alternatively spliced to produce two isoforms, one containing and one lacking exon 4 (*FAM161A_{fl}* and *FAM161A_{sh}* respectively).

Sanger sequencing and preliminary qPCR results indicate that *FAM161A* mRNA transcripts in retinal tissue and blood from dogs not affected with PRA comprise predominantly the wildtype *FAM161A* isoforms (*FAM161A_{fl}* and *FAM161A_{sh}*). Conversely, *FAM161A* mRNA transcripts in blood from a dog affected with PRA-3 (i.e. homozygous for *FAM161A_{c.1758-15_1758-16ins238}*) comprise predominantly the aberrant *FAM161A* isoforms lacking exon 5 (*FAM161A_{fl-5}* and *FAM161A_{sh-5}*). This supports the hypothesis that the SINE insertion results in skipping of exon 5 during pre-mRNA splicing in blood. While it is likely that *FAM161A_{c.1758-15_1758-16ins238}* has the same effect of aberrant splicing in other tissues, the possibly that tissue-specific splicing negates this effect in the retina cannot be excluded. Further investigation using retinal tissue from a *FAM161A^{-/-}* dog is necessary to substantiate the hypothesis of alternative splicing. Interestingly, aberrant *FAM161A* isoforms (*FAM161A_{fl-5}* and *FAM161A_{sh-5}*) were also present in retinal tissue from a dog not affected with PRA, albeit at much lower levels than the wildtype isoforms. These are most likely a result of naturally-occurring alternative splicing, which is a common occurrence. At least 74% of human multi-exon genes are alternatively spliced¹⁷¹ and up to 30% of alternative transcripts contain premature termination codons¹⁷². These are usually targets of NMD, although Lewis et al. observed that 4.3% of RefSeq mRNAs (i.e. experimentally identified mRNAs that have not been degraded) are truncated by >50 amino acids¹⁷². While these aberrant

proteins may well be expressed in healthy retinal tissue, it is clear from the PCR/electrophoresis and qPCR data presented here that they are a minor product compared with the normal, functional protein.

Similar instances of aberrant splicing due to SINE insertions that are associated with canine traits have been reported: An insertion 35 bp upstream of an acceptor site of the *HCRTR2* (hypocretin (orexin) receptor 2) gene has been associated with canine narcolepsy in the Doberman breed¹⁷³, and an insertion nine bp upstream of an acceptor site of the *SILV* (a.k.a *PMEL*; premelanosome protein) gene has been associated with the merle pigmentation pattern in multiple dog breeds¹⁷⁴. The authors of both studies hypothesised that the intronic SINE insertion near a splice acceptor site displaced a putative lariat branch point sequence (BPS), resulting in skipping of the adjacent exon during mRNA splicing. Lin et al. presented data in support of this hypothesis: *HCRTR2* cDNA from dogs homozygous for the SINE insertion associated with canine narcolepsy lacked the sequence corresponding to the entire fourth exon¹⁷³. The pre-mRNA splicing mechanism requires at least three consensus intronic sequences for optimal function. One of these is the 3' consensus sequence 6PyNCAG (where Py is a pyrimidine base and N is any base) of the acceptor site and another, the BPS, is the site of lariat formation¹⁷⁵. In eukaryotes the latter is typically, but not always, 20 to 50 nucleotides upstream of the splice junction. The consensus sequence of the BPS, to which the U2 component of the spliceosome binds, is also variable (PyNPyTPuAPy, where Pu is a purine base), although the adenine base is of primary importance for lariat formation^{176,177}. There is no sequence within 50 nucleotides of the *FAM161A* intron 4-exon 5 splice site that corresponds to the BPS consensus sequence. However, a putative BPS located 76 nucleotides from the splice site, does correspond to the consensus sequence (Figure 47). While the SINE insertion does not affect the AG sequence of the acceptor site it will push the BPS beyond its optimal range from the acceptor site, which is likely to be the cause of aberrant splicing of exon 5 in affected dogs.

While a causal mutation was not identified, the authors suggested that this second locus on CFA15 could be a modifier of *RPGRIP1*^{-/-}, or it could be the second allele of a digenic condition¹⁸¹. A third hypothesis was that variations in the length of the poly(A) tract could account for phenotypic differences¹⁸⁰. Longer poly(A) tracts in human L1 LINEs (long interspersed nuclear elements) have been associated with disease in humans¹⁸² and the longer SINE in the *SILV* gene results in the merle phenotype in dogs, while the shorter SINE does not¹⁷⁴. However, Miyadera et al. reported that variable poly(A) tract length in the *RPGRIP1* SINE insertion did not correlate with variable phenotype¹⁸⁰. While the precise length of the poly(A) tract in the *FAM161A* SINE insertion is unknown, data presented suggests the length of the poly(A) tract does vary, with two alleles identified (A₄₄ and A₄₉). However, as the predominant allele (A₄₄) segregates in dogs that were diagnosed with PRA at between two and 10 years of age, it is unlikely that the poly(A) tract length variation has much effect on the PRA phenotype or severity.

In order to further test the validity of the insertion mutation, *FAM161A*_{c.1758-15_1758-16ins238}, we screened 247 TSs for the mutation (Table 23). We found that 62.9% of the PRA cases, 56.3% of the obligate PRA carriers and 100% of clinically unaffected dogs (which could be clear of the mutation or carry a single copy) have *FAM161A* genotypes that are concordant with their clinical status. There are two groups of dogs with genotypes discordant with their phenotypes. The first comprises two dogs that are homozygous for the mutation and have not been diagnosed with PRA. Clinical information pertaining to one of these dogs was unavailable, although it is known to have had at least one PRA-affected sibling. The other dog had not been seen by an ophthalmologist, but its owner reported no significant loss of sight by the time it died. However, it had lost one eye in an accident and developed a cataract in the other eye around nine years of age, which could have been secondary to PRA. The observation that 91.7% (22/24) of dogs homozygous for *FAM161A*_{c.1758-15_1758-16ins238} (*FAM161A*^{-/-}) have developed PRA suggests the mutation is fully penetrant, or nearly so. The inheritance observed in a family of 49 dogs (seven cases) is supportive of a recessive mode (Figure 44). The second group of discordant dogs comprises 13 PRA-affected dogs that are not homozygous for *FAM161A*_{c.1758-15_1758-16ins238} and seven obligate carriers do not carry *FAM161A*_{c.1758-15_1758-16ins238}. It is formally possible that the mutation has a dominant mode of inheritance with incomplete penetrance, or complex trait or compound heterozygote effects. However, as is the case with PRA in GRs and

GoSs (Chapter 3 and Chapter 4 respectively) it is more likely that additional loci are responsible for the discordant cases.

Anecdotal evidence of their shared origins in Tibetan monasteries suggests that TTs and LAs are the most closely related breeds to the TS, and as a result these are the breeds most likely to share the PRA-3 variant. Screening of 76 TT and 23 LA dogs, including 12 and nine PRA cases respectively, revealed that the mutation is present in TTs, but is absent from the LAs screened. *FAM161A* genotypes of 33.3% of the TT PRA cases and 100% of clinically unaffected dogs (which could be clear of the mutation or carry a single copy) were concordant with their clinical status. Interestingly, two of the eight PRA-affected TTs that were not homozygous for *FAM161A*_{c.1758-15_1758-16ins238} (i.e. *FAM161A*^{+/-} or *FAM161A*^{+/+}), were in fact homozygous for the mutation associated with RCD4 (Chapter 4). It is therefore likely that PRA in the remaining six cases in the breed is caused by a third unknown mutation.

The allele frequency of *FAM161A*_{c.1758-15_1758-16ins238} in both breeds is not unusually high, even for a recessive disease allele, as discussed in section 3.8.1 (Table 25). Estimates for both breeds indicate that up to one in 31 TS dogs and one in 48 TT dogs are affected with this form of PRA. However, these frequencies are likely to be inflated, as samples collected for research are unlikely to be representative of the wider population. This hypothesis is supported by the high Hardy Weinberg χ^2 values for the TS breed. Nevertheless these data suggest *FAM161A*_{c.1758-15_1758-16ins238} is a major cause of PRA in both breeds. A DNA test for PRA-3 is expected to be made commercially available from the AHT Genetic Services department in 2013, and screening of the wider population for the mutation or a sample cohort collected specifically for allele frequency estimations will give a more representative indication of the mutant allele frequency.

FAM161A encodes the family with sequence similarity 161, member A protein and was recognised as an appealing candidate gene due to its involvement in RP in humans^{168,169}. The gene occurs in two main isoforms, that differ by the presence or absence of exon 4 (*FAM161A*_{fl} and *FAM161A*_{sh} respectively)¹⁶⁸. Both isoforms are expressed in multiple tissues including the retina and testes, and at lower levels in the heart, liver, kidney, brain, muscle, lung and thyroid gland¹⁶⁹. Specifically, the protein has been localised to the connecting cilium and basal body in the IS of rod and cone photoreceptor cells, and to the basal body and centrosome of ciliated cells of different origins^{183,184}. *FAM161A* has been shown to interact with the CRX (Cone-rod homeobox-containing) transcription factor¹⁶⁹ and Lebercilin¹⁸³, both of which have also

been implicated in retinal degeneration in humans^{185,186}. Only a single evolutionary conserved domain (UPF0564) has been identified, which is vital for binding to and stabilising microtubules^{168,184}. This region is also required for homotypic FAM161A interactions, as well as heterotypic interactions with paralog FAM161B (family with sequence similarity 161, member B)¹⁸⁴. FAM161B interacts with TACC3 (transforming, acidic coiled-coil containing protein 3), which in turn is involved in centrosome-dependent microtubule assembly, kinetochore attachment, chromosome alignment and mitotic exit¹⁸⁷. FAM161A could therefore be involved in maintenance of the microtubule axoneme along the connecting cilium or protein transport between the IS and OS^{183,184}.

*FAM161A*_{c.1758-15_1758-16ins238} affects splicing of exon 5 of the gene, resulting in a truncated protein, including the loss of approximately 44 amino acids of the UPF0564 conserved domain. However, Bandah-Rozenfeld et al. reported that the N-terminus of the UPF0564 domain is sufficient for homotypic and heterotypic interaction with FAM161B¹⁶⁸. The truncated protein product is therefore expected to be functional in this regard. As the discovery of *FAM161A* involvement in retinal disease was relatively recent, very little is known about the structure and function of the protein in visual and other pathways. Further investigations are required to elucidate the precise pathways in which FAM161A is involved, which may lead to the identification of novel functional domains in the C-terminus of the protein. To this end, the canine model described here could be particularly useful, as no other animal models have been reported.

The presence of *FAM161A* mutant mRNA transcripts in the blood of an affected dog, and the similar expression levels of *FAM161A*_{total} in affected and normal blood, imply that the truncated transcript is not subjected to NMD. A truncated protein may therefore be expressed, although this would need to be confirmed by comparing FAM161A protein levels in *FAM161A*^{-/-} dogs with protein levels in *FAM161A*^{+/+} dogs. If this is the case, the truncated protein product must be sufficient to cause retinal degeneration. It is possible the truncated protein is toxic and therefore actively causes retinal decay. While none of the PRA-3 carriers (*FAM161A*^{+/-}) have developed PRA, they may have signs of degeneration that are not typical of PRA and therefore easily missed (or even ignored) by ophthalmologists. Alternatively, the truncated protein may cause subclinical degeneration in carrier dogs. More likely, the truncated protein is insufficient to maintain normal retinal function, resulting in retinal degeneration.

PRA caused by the variant described here has an average age at diagnosis of 4.89 years and this is indicative of a late age of onset (data not shown) and consistent with observations in human patients in which the age of onset was in the 2nd or 3rd decade¹⁶⁹. Given that FAM161A is expressed in multiple tissues, it would be interesting to determine whether a more severe change to the protein, such as a knock-out, would result in a more severe retinal or even systemic phenotype.

The discordant TS PRA cases i.e. *FAM161A*^{+/+} and *FAM161A*^{+/-} tended to develop PRA at a later age, with an average age at diagnosis of 7.01 years (data not shown), which is consistent with the segregation of a second form of PRA in the TS breed.

5.7 Conclusion

PRA in the TS has not previously been associated with any genetic mutations. Using a GWA mapping approach, a novel candidate mutation, *FAM161A*_{c.1758-15_1758-16ins238}, was identified that is likely to represent a major causal mutation for PRA in the TS. While this mutation does not account for all cases of PRA in this study, suggesting that there are additional loci causing PRA in this breed, it does appear to be highly penetrant and a major cause of PRA in this breed. While the PRA-3 mutation is also present in TTs, as they are closely related and it has not been found in any other breeds, the mutation appears to be confined to these two breeds. The development and availability of a DNA test will enable breeders to eventually eliminate PRA-3 from the breeds affected entirely.

Chapter 6 General Discussion

6.1 Aims of The Current Project

PRA has been identified in more than 100 dog breeds, but in most cases the underlying molecular basis has yet to be elucidated. This study was conceived with three main aims for each of three breeds (GR, GoS and TS):

- 1) Investigate the molecular basis underlying PRA.
- 2) Establish a diagnostic DNA test to be commercially and available to the public.
- 3) Assess the relevance of findings to human retinal degenerations.

All three aims have been achieved, to some extent, as summarised in Table 26.

6.1.1 Aim 1: Underlying Molecular Basis

Four mutations were identified that were associated with PRA in three principal breeds: *SLC4A3*_{c.2601_2602insC} and *TTC8*_{c.669delA} in GRs, *C2ORF71*_{c.3149_3150insC} in GoSs and *FAM161A*_{c.1758-15_1758-16ins238} in TSs. These mutations were hypothesised to cause aR PRA in homozygous dogs. The segregation of these mutations with PRA in their respective breeds as well as the predicted effect of each on their respective proteins, as presented in previous chapters, supports this hypothesis. The function of the genes, particularly *FAM161A* and *C2ORF71*, is not yet well understood and the effects aberrant proteins i.e. gain or loss of function, cannot be precisely predicted. However, as inheritance of all four PRAs is consistent with an aR mode and dogs that carry the variants (heterozygotes) do not usually develop PRA, the variants are likely to result in loss of function.

All four mutations identified are predicted to result in truncated protein products, which could function inefficiently or be degraded through **nonsense mediated decay (NMD)**. At least 74% of multi-exon genes in humans undergo alternative splicing¹⁷¹. It is estimated that 35% of these alternative gene isoforms contain premature termination codons¹⁷² and expression of these transcripts could result in harmful proteins. NMD, therefore, is a cellular safeguard against the expression of harmful proteins from incorrectly spliced mRNAs. However, in the case of a homozygous mutation that affects the only or major functional isoform, NMD effectively removes all mRNA transcripts and no functional protein will be expressed. One way to test for NMD is to compare the levels of mRNA between dogs with the variant and those without. Retinal tissue was preferable for this study (see section 6.2 for further discussion), but the same principle applies to any tissue in which the gene is expressed, including blood. Retinal tissue from dogs homozygous for any of the four mutations identified was unavailable, and mRNA of three of the genes (*SLC4A3*, *TTC8* and *C2ORF71*) was undetectable in blood. It therefore remains unclear whether or not these genes are subjected to NMD. Evidence against NMD of aberrant *FAM161A* mRNA is presented in Chapter 5, but this would need to be confirmed by comparing protein expression levels in retinal tissue.

The four mutations identified are all likely to result in loss of protein function. However, experimental evidence demonstrating the detrimental effect of the variants on their respective proteins, and subsequently on the biological pathways in which they are involved, was not obtained. The primary reason for this was the difficulty of obtaining the appropriate tissue from an affected dog (as discussed in section 6.2). The possibility that the variants are not causal themselves cannot be excluded. They could instead be in LD with unidentified mutations that are causal. DNA in the PRA-associated regions and retinal mRNA remains unsequenced, including the 5'- and 3'-UTRs of all four genes. These regions are often difficult to sequence, and numerous attempts using varying primer combinations, altered thermal cycling protocols, multiple Taq polymerases and RACE (Rapid Amplification of cDNA Ends) kits failed to yield positive results. Nevertheless, all four variants remain provocative candidate causal mutations.

6.1.2 Aim 2: Diagnostic DNA Test

A DNA test for each mutation has been developed and is (or will be) available to the public from AHT Genetic Services. The availability of diagnostic DNA tests will enable breeders to make sensible breeding decisions with the long term goal of eliminating PRA from the breed entirely whilst not adversely affecting the genetic health of the

breed. Breeders may be tempted to avoid breeding with dogs that are heterozygous or homozygous for any of the mutations. However, this reduces the **effective population size** (N_e ; the number of individuals that contribute offspring to the next generation) and genetic diversity of the breeds, and may bring about the emergence of other recessive diseases. Whilst risky, breeding with a carrier dog (for example) can be made safe in a two ways. Firstly, the dog should only mated to a bitch that is homozygous for the wildtype allele. Secondly, the puppies produced should be endorsed i.e. they cannot be bred from themselves until they have been tested genetically. Ideally, breeding advice would be tailored to each individual breed and will take into consideration factors such as **effective population size** and **coefficient of inbreeding** (F ; the probability that two alleles in an individual are identical by descent). However, as this usually not possible, the AHT generally advises breeders to continue breeding with dogs that carry a mutation until the frequency of the mutation is <0.01 . This is particularly critical in the case of RCD4 as, due to the extremely high prevalence of the mutant allele in affected breeds, elimination of individuals heterozygous or homozygous for the mutation would drastically reduce the effective population size of the breed, which could in turn be highly detrimental to the overall health of the breed.

6.1.3 Aim 3: Relevance to Humans

The molecular characterisation of four new forms of PRA have resulted in the identification of a large animal model for at least three human retinal degenerations. Dogs homozygous or heterozygous for a variant could be used to establish colonies, which in turn could be used to undertake further molecular investigations. This could be particularly beneficial as a mouse model (knock-out) has only been reported for one of the four genes (*SLC4A3*). Interestingly, this is the only gene (of the four identified) that has not yet been implicated in retinal degeneration in humans, but remains a provocative candidate gene. The development of mouse models for the other three genes (*TTC8*, *C2ORF71* and *FAM161A*) would be useful to understand the associated retinal degenerations better. Further molecular investigations could reveal more about the function of each protein, and contribute to further understanding of the networks and pathways involved in retinal function. In addition, several canine models have been used to test the proof of principal and the safety of novel therapeutic strategies, such as corrective gene therapies, neuroprotective agents and retinal prostheses, as reviewed by Beltran¹⁸⁸. Further investigations could therefore also elucidate the potential use of these dog models in the development of therapeutic strategies.

Table 26. Summary of results

	Cohort	Golden Retriever PRA1	Golden Retriever PRA2	Gordon Setter PRA	Tibetan Spaniel PRA
Experimental work	GWA critical region	CFA37, 644 kb	CFA8, 3.814 Mb	CFA17, 3.221 Mb	CFA10, 2.352 Mb
	Variant identification (Sequencing)	Sanger, Candidate Gene	Targeted NGS	Sanger, Candidate Gene	Targeted NGS
	Disease-associated variant	<i>SLC4A3</i> _{c.2601_2602insC}	<i>TTC8</i> _{c.669delA}	<i>C2ORF71</i> _{c.3149_3150insC}	<i>FAM161A</i> _{c.1758-15_1758-16ins238}
	Effect on protein	p.E868RfsX104	p.K233RfsX15	p.C1051VfsX90	p.S588MfsX13, splicing
	AHT Genetics Services (Commercial DNA Test)*	855 GRs; AF = 0.05	415 GRs; AF = 0.03	1788 GoSs; AF = 0.33 2249 IrSs; AF = 0.25 60 TTs; AF = 0.22	Test developed, launch TCB
	Canine PRA	Disease name/identifier	PRA-1	PRA-2	RCD4
Phenotype		Typical PRA	Typical PRA, potentially other systemic clinical signs	Typical PRA	Typical PRA
Age-at-diagnosis (years)		6.7	4.5	9.8	4.9
Human Disease	Disease association	Idiopathic generalised epilepsy. ¹⁵³ Not associated with retinal disease.	BBS ^{128,129} , arRP ¹³⁰	arRP ^{156,157,164}	arRP ^{168,169}
Gene	Gene name	Solute carrier family 4, anion exchanger 3	Tetratricopeptide repeat domain 8	Chromosome 2 open reading frame 71	Family with sequence similarity 161, member A
	Gene function	Cl ⁻ /HCO ₃ ⁻ exchanger; pH regulation, Cl ⁻ level maintenance ¹²⁴	Part of BBSome, role in ciliogenesis ¹⁴⁸	Physiological function unknown, important for retinal development and photoreceptor function ¹⁵⁶	Involved in developing & adult retina. Possibly maintenance of microtubule axoneme or protein transport. ^{183,184}
	Isoforms	Full length (fl) and Cardiac (c) ¹²⁴	Full length (fl1 & fl2) ¹²⁸ , & retina-specific (2A) ¹³⁰	Full length	Full length (fl) & short (sh) ¹⁶⁸
	Tissue expression	Brain, heart, retina, gut, kidney ¹²⁴	Ciliated tissue (fl) ¹²⁸ , 2A retina specific ¹³⁰	Eye ¹⁵⁷	Retina, testes, heart, liver, kidney, brain, muscle, lung & thyroid gland ¹⁶⁹
	Retinal localisation	Müller (fl) and Horizontal (c) cells ¹²⁴	Photoreceptors, ciliary membrane	OS or CC of Photoreceptors ¹⁵⁷	CC & BB of photoreceptors ^{183,184}
	Animal Models	Knockout mouse ¹²⁵	None reported	Zebrafish ¹⁵⁷	None reported

* AF = PRA-associated mutant allele frequency

6.2 Sample Collection

The manner in which genetic research samples are collected by the AHT is simultaneously highly beneficial and highly restrictive. DNA samples are collected from pet dogs, by their owners, using buccal/cheek swabs. As a result, an impressive bank of samples has been collected over the years (> 23,000 samples from > 180 breeds). We are therefore able to obtain, with relative ease for many studies, a sufficient number of samples to undertake GWA mapping studies. In addition we do not have to consider the costs, labour and ethical considerations associated with breeding colonies of dogs for the purposes of research.

6.2.1 Tissue Availability

However, a drawback of sourcing samples from pet dogs is that obtaining any other tissue is difficult. It is even difficult to obtain blood samples for the purposes of research, as a Home Office licence is required. As such, it is often not possible to conduct functional studies subsequent to the identification of mutations. This difficulty is compounded for PRA studies because the disease is not lethal and dogs can adapt and function remarkably well without sight. In contrast, dogs with severe or lethal diseases, such as canine Neonatal cerebellar cortical degeneration (NCCD), are usually euthanized and tissue can therefore be obtained for functional investigations¹⁸⁹. Were suitable tissues available, a number of functional investigations could be undertaken. Protein and mRNA experiments could be undertaken to assess the effect of mutations on the protein, including whether or not the proteins are expressed in a truncated form or affect the expression levels or locations of other retinal proteins. Histological and immunohistochemistry experiments could be used to assess the effect of the mutations on retinal cells and structure. For this project it was not possible to obtain the retinal tissue that would have enabled further study of the mutations or diseases identified.

6.2.2 Phenotyping

Another drawback of sourcing samples from pet dogs is the inconsistency of disease phenotyping. All dogs defined as PRA cases were examined by veterinary ophthalmologists and therefore have robust diagnoses. However, there is variation in the qualifications (certificates versus diplomas) and experience of ophthalmologists as well as the inevitable variation introduced when multiple clinicians contribute to sample collection. In addition, the phenotypic information available for each case varies from virtually nothing to quite detailed (age at diagnosis, fundus photographs,

ophthalmological observations and ERG results). This makes avoiding a heterogeneous pool of cases all but impossible. For most dogs with PRA the only additional information available is the age of the dog when it was first diagnosed. This information was used in this project as an indication of the age of onset of each form of PRA. However, apart from the possible variability in the age of onset of clinical disease, the age at diagnosis will vary due to factors such as how observant the owners are with regards to changes in their dogs behaviour and how frequently the dogs eyes are examined. Finally, the disease progression and the presence of additional clinical signs cannot be easily monitored, as demonstrated in GRs with PRA-2 (Chapter 3), due to a number of factors. Firstly, as samples are collected over many years, many of PRA cases used in these studies are no longer alive, or the contact details of the owners are no longer accurate. Secondly, most of the dogs sampled are not in close proximity to the AHT. Therefore, either the owners have to travel large distances to be examined by AHT clinicians, or additional, willing, external clinicians need to be identified and involved in the study.

6.2.3 Allele Frequency Estimations

An estimate of the frequency of any mutant allele is required in order to properly advise breeders on a suitable breeding strategy that will eventually eliminate PRA from the breed entirely. For this estimate to be accurate, however, the sample cohort must be representative of the general breed population. Sampling bias will always be a factor in sample cohorts collected for the purposes of investigating a particular trait or condition (PRA in this discussion) for a number of reasons: Firstly, such a cohort will comprise a disproportionate number of dogs with confirmed PRA. Secondly, it will also contain a disproportionate number of dogs closely related to dogs with PRA, and therefore a disproportionate number of dogs that carry PRA mutations. Thirdly, samples will only be submitted by owners who are aware of the research, often only because their dog is related to dogs with PRA or they are members of Breed Clubs that advertise and promote research projects. It is likely that many pet owners will not be aware of the disease, much less the research project, unless they visit a veterinary ophthalmologist. However, many older dogs that develop PRA will not be examined by an ophthalmologist as the vision loss will simply be attributed to old age. Sampling bias can, however, be minimised, or at least reduced by changing sample collection criteria. For example, a cohort of GR samples used in Chapter 3 was collected specifically for allele frequency estimations. Samples were taken from dogs that had been selected by a

representative of the Golden Retriever Breed Council (Margaret Woods, Breed Health Co-ordinator) to meet a number of criteria: Firstly, they were of unique parentage within the cohort i.e. none of the dogs shared a sire or dam. Secondly, the dogs came from a variety of different kennels and breeding lines. Thirdly, with the aim of selecting dogs that represented the GR breeding population, the dogs were of breeding age (defined as the ages the KC allows for breeding, between one and seven years old). An alternative method, used by the AHT to investigate the frequency of the mutation associated with Neonatal cerebellar cortical degeneration (NCCD) in Beagles, involved the use of the Kennel Club (KC) registration database. Representatives of the KC sent buccal swab (sampling) kits to the owners of 500 Beagles randomly selected from the KC transfer of ownership database (which contains information pertaining to every single pedigree dog in the UK that is KC-registered, whose registration has been transferred from the breeder to the new owner), with the request to take a sample from their dog and return it to the AHT. Of the 500 kits sent out, 152 were returned to the AHT, representing a return rate of 30% (personal communication, Oliver Forman). This strategy is preferred over that used in the GR studies, as sample selection is completely random and objective; and dogs from pet, show and working populations are represented. It also has the added benefit of reaching more of the public with information about the disease under investigation. However, sample collection is still biased towards dogs that are KC registered, and it is estimated that less than 1/3 dogs in the UK is KC registered¹⁹⁰.

6.3 GWA Population Structure

As has been demonstrated in the literature in recent years GWA is a popular method for the identification of trait-associated loci in humans and other species. It is a particularly appropriate strategy in dogs, especially for monogenic traits, due to the inherent population structure present as a result of relatively recent domestication and centuries of selective breeding. However, unknown population structure (a.k.a stratification) has a confounding effect on GWA and can result in false negatives (type II errors) or especially false positives (type I errors). The extent of population stratification is represented by the inflation factor (λ), which is simply a comparison (ratio) of the expected and observed median χ^2 values (see section 2.3.2). Population stratification will result in inflated χ^2 values (i.e. $\lambda > 1$; $\chi^2_{\text{observed}} > \chi^2_{\text{expected}}$) and can be visualised with a Q-Q (quantile-quantile) plot.

6.3.1 Q-Q Plots

A Q-Q plot (Figure 48) compares the distribution of observed test statistics (χ^2_{observed}) with those expected under the null hypothesis (χ^2_{expected}). Deviation from the null across the entire distribution indicates the presence of population stratification (Figure 48B). Deviation from the null only at the highly significant end of the range (i.e. the highest χ^2 values) indicates a true association with the trait under investigation (Figure 48D).

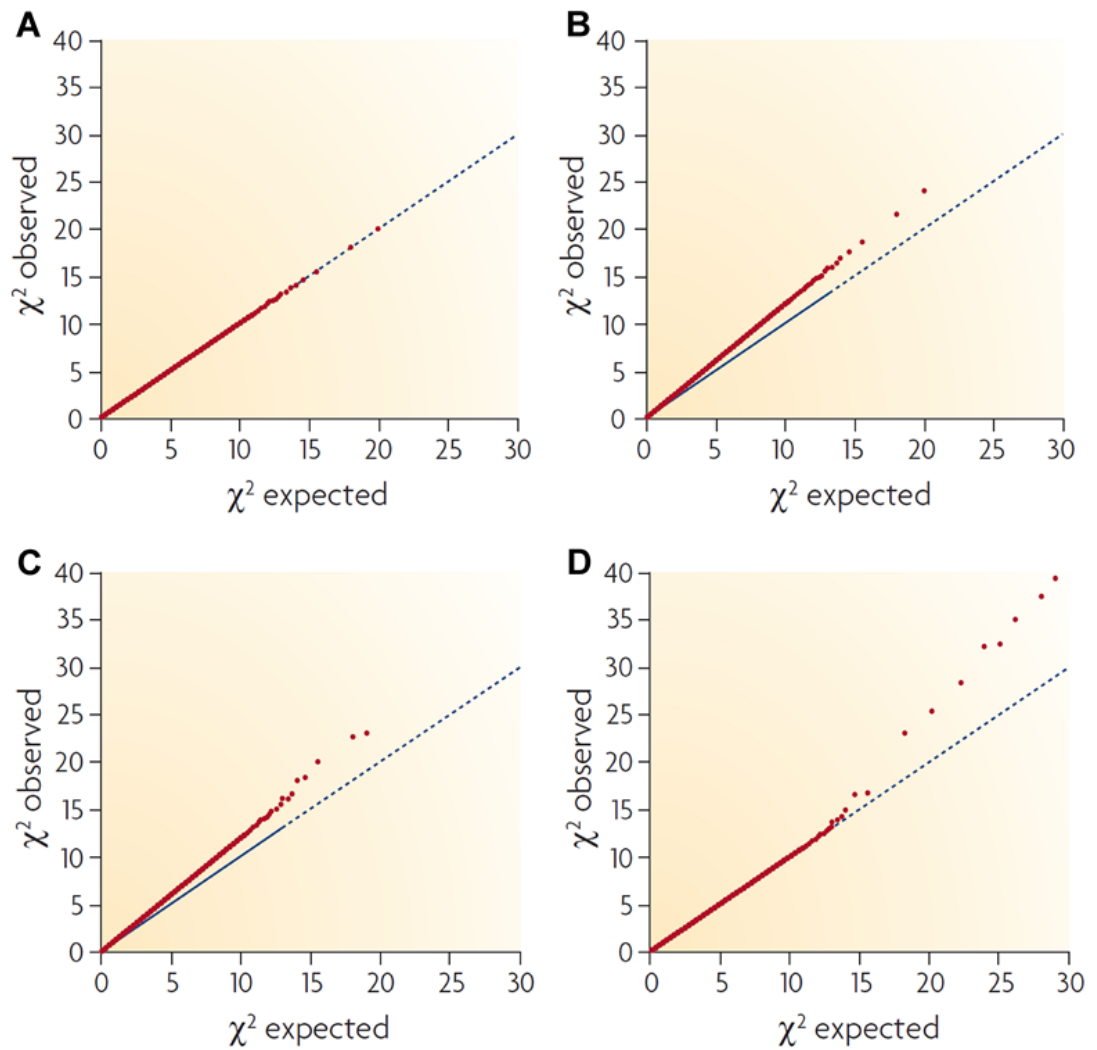


Figure 48. Q-Q plots to assess population stratification

Q-Q plots are generated by plotting observed against expected χ^2 test statistics, and comparing the distribution of χ^2_{observed} (red circles) with the distribution expected under the null hypothesis (blue line). A) No evidence of stratification or association. B) Evidence of stratification only. C) Evidence of stratification and association. D) Evidence of association only. Image modified from McCarthy et al.¹⁹¹.

6.3.2 Multidimensional Scaling

Another useful tool for visualising population structure is multidimensional scaling (MDS) of IBS pairwise distances (Figure 49). Ideally, the distribution of cases and controls should be similar, to minimise population stratification and resulting type I and type II errors. This allows identification and removal from further analysis of outlier samples that may generate spurious association results. It also allows the identification of possible sampling errors, such as the inclusion of samples from a different breed, that may also generate spurious association results. However, as with all statistical analyses, the larger the sample cohort is, the more accurate and useful the results will be. If the sample cohort is small to begin with (as were the cohorts used in this investigation), MDS plots are for information only, for the most part, as the removal of samples from further analysis will greatly reduce the power of the study.

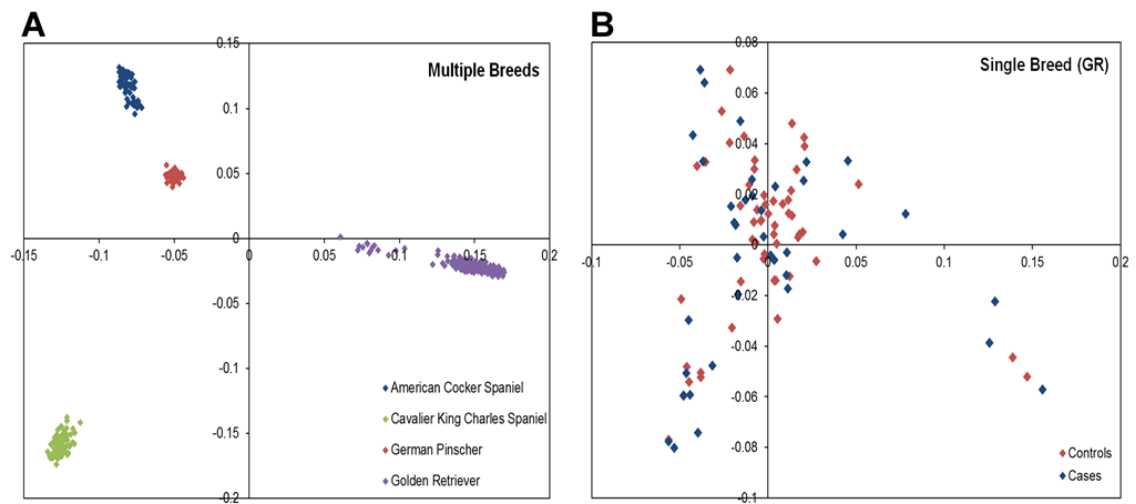


Figure 49. Multidimensional scaling plot

MDS plots using GWA data. A) Data from four breeds separates into four distinct clusters, one for each breed. B) Cases and controls in a GWA study from a single breed should have a similar pattern of distribution.

6.3.3 Correction for Population Stratification

A number of algorithms have been devised to correct for population stratification in a study population, and consequently reduce the number of type I and type II errors. The first method used during this project, meta-analysis of clusters of individuals using the Cochran–Mantel–Haenszel (CMH) test, is routinely used for the correction of population stratification in human populations. It corrects for **unrecognized population stratification arising from remote common ancestry between individuals**, and **cryptic relatedness arising from recent common ancestry in ostensibly unrelated individuals**, based on the proportion of alleles shared that are IBS^{100,192}. Animals on the other hand, particularly domestic animals such as dogs, often have a third level of relatedness that most outbred human populations do not, **family structure arising from familial relatedness**¹⁰¹. This simply means that dogs of any given population/breed are expected to share large portions of the genome. As this is not the case in humans a CMH correction might not correct adequately for stratification, a theory supported by suboptimal inflation factors ($\lambda > 1$) generated after correction in 3/4 cohorts studied (Table 27). Mixed-model algorithms, such as fast mixed model (FMM), are routinely used in animal breeding applications and are efficient models for analysing complex kinship patterns¹⁰¹. However, 2/4 cohorts have inflation factors substantially < 1 , suggesting this method tends to result in deflated association results (Table 27). From these data, it is clear that there is no single correct method for adjusting for population structure and it is therefore vital to use and consider the results from more than one. In each of the GWA studies listed in Table 27, the same signals were seen regardless of the analyses, and this conferred a great deal of confidence in the GWA results.

Table 27. GWA inflation factors (λ)

Cohort	Before correction	CMH correction	FMM correction
GR_PRA1	1.44	1.46	0.96
GR_PRA2	1.32	0.84	0.27
GoS_PRA	2.44	1.56	0.34
TS_PRA	1.69	1.10	1.27

6.4 Sequencing

Until recently the most widely used method to determine the sequence of nucleotides in a DNA molecule was Dye-terminator (Sanger) sequencing¹⁹³. The method was employed in many ground-breaking studies including the production of human and canine draft sequences^{23,194} and the identification of mutations associated with disease, including RP and PRA^{30,156,195}. While arguably still considered the gold standard for sequencing small fragments of DNA, the Sanger method does have limitations. It is labour-intensive, time-consuming and costly. The sequencing of the human genome of a single individual (Craig Venter) using this method required more than 340,000 runs and cost approximately \$70 million^{196,197}. Next-generation sequencing (NGS) enables cost-effective massively paralleled sequencing of multiple samples. Around the same time as Venter's genome was sequenced the newly developed Roche/454 NGS platform was used to sequence the genome of another individual (James Watson) and required only 234 runs and cost \$1 million^{197,198}. The labour and costs have since been considerably reduced and it now costs less than \$10,000 to sequence a human-sized genome (<http://www.genome.gov/sequencingcosts/>). A number of NGS platforms and technologies have been developed. The Illumina technology was selected for this project due to its popularity and subsequent availability at numerous service laboratories. The Illumina systems enable simultaneous sequencing of hundreds of millions of clusters of DNA fragments. The chemistry, which reads only a single base in each synthesis cycle, ensures highly accurate sequencing through difficult regions such as homopolymers and repeated sequence.

6.4.1 Sequencing Limitations

Sequencing, whether by the Sanger method or more recent next generation methods, is critical for the identification of the precise DNA mutations that cause disease. The Sanger method is still convenient and cost effective for determining small quantities of sequence. In this project it was successfully used to identify the mutations associated with PRA-1 in GRs and RCD4 in GoSs (*SLC4A3*_{c.2601_2602insC} and *C2ORF71*_{c.3149_3150insC} respectively). However, had the mutations associated with PRA in these breeds been anywhere other than in the exons of these candidate genes, they would not have been as easily identified. NGS, on the other hand, is becoming the method of choice for many studies because it allows the generation of massive amounts of sequence data. In this project NGS targeting specific regions was employed to identify the mutations associated with PRA-2 in GRs and PRA-3 in TSs (*TTC8*_{c.669delA} and *FAM161A*_{c.1758-}

15_1758-16ins238 respectively). However, had the variants associated with PRA in these breeds been within highly repetitive or repeat masked regions, they would have been more challenging to identify, as repeat masked regions were not targeted during library preparation. Current NGS technologies are not capable of creating reads long enough to be accurately aligned to highly repetitive regions of the genome. These regions i.e. repeat masked regions, therefore, cannot be sequenced accurately. Structural variation (such as large insertions and deletions) remains difficult to identify by NGS methods. While changes to read depth can be indicative of such variation, the target enrichment approach also introduces variable read depth, which is difficult to discern from structural variation induced read depth changes. While Pindel identifies such variations and was included in the NGS analysis pipeline (Chapter 2, Section 2.4.2.4), it was unreliable and inconsistent, due to a combination of software and hardware (i.e. server/processing) deficiencies. Indeed, Pindel failed to generate results for either of the NGS data sets analysed during this project, and as a result structural variations had to be identified visually where possible (section 5.4.3).

6.4.2 NGS Challenges

One of the biggest challenges with regards to NGS is the analysis of the extraordinary amount of data generated. New and improved analysis tools are made available to the public continuously, and it is challenging to keep abreast of the latest developments. In addition, there is no single package with which to undertake all of the analysis from start to finish, and few of the tools are designed with user-friendly interfaces. It is therefore becoming crucial to be proficient in the use of UNIX (the operating system in which most of the analysis tools run) and also programming languages such as perl, R and VBA (Visual Basic for Applications).

6.5 Genotype-Phenotype Discordance

While the four mutations identified account for almost 80% (108/137) of PRA cases in the three principal breeds investigated (GR, GoS and TS; Figure 50), a further 29 dogs with PRA are not homozygous for any of the variants. A number of factors could account for this genotype-phenotype discordance:

Phenocopies are non-genetic retinopathies that resemble PRA, such as SARD and Vitamin A deficiency-induced nightblindness. It is possible that at least some of the “PRA cases” have been misdiagnosed as such, although this is unlikely to account for many of the 29 discordant cases.

Genetic heterogeneity is a well-known confounding factor in the study of human retinal degenerations. While PRA within dog breeds may be less heterogeneous than RP or LCA in human populations, it is becoming increasingly apparent that PRA is not necessarily homogeneous within individual breeds, even numerically small breeds. Considering the number of genes that have been implicated in human retinal degenerations (as discussed in Chapter 1), it is hardly surprising to find that some breeds may have more than one form of PRA segregating with the breed.

In addition, 20 dogs that were homozygous for the RCD4 variant (Chapter 4) and three that were homozygous for the PRA-1 variant (Chapter 1), had not been diagnosed with PRA at the time of the study. A number of factors could account for this genotype-phenotype discordance also:

These variants could have **reduced** or **incomplete penetrance** due to environmental factors or stochastic fluctuations in gene expression¹⁹⁹. Age-related incomplete penetrance could be a significant factor in the penetrance of the RCD4 variant in particular. The average age at which dogs are diagnosed with RCD4 is 9 years, and as such, taking into consideration phenotypic heterogeneity, clinical signs of disease may not manifest during the dog’s lifetime. This hypothesis is supported by the observation that 8/18 GoSs that were homozygous for the RCD4 mutation and over the age of 9 years when they were examined (ranging from 9.1 to 11.8 years) were not diagnosed with PRA.

While these forms of PRA appear to be monogenic, we cannot discount the possibility that they may in fact be **complex** or **multifactorial traits**, with multiple genetic contributions. The interaction of multiple genes and the environment are characteristic of **multigenic** traits, whereas the interaction of few genes is

characteristic of **oligogenic** traits. The effects of the genes are said to be **modifying** or **synergistic**. Heterozygous, synergistic mutations in the *ROM1* and *RDS* genes result in digenic RP only when both mutations are present²⁰⁰. At least two modifiers manage the expression of *PRPF31* mRNA to affect the penetrance of RP caused by this gene²⁰¹.

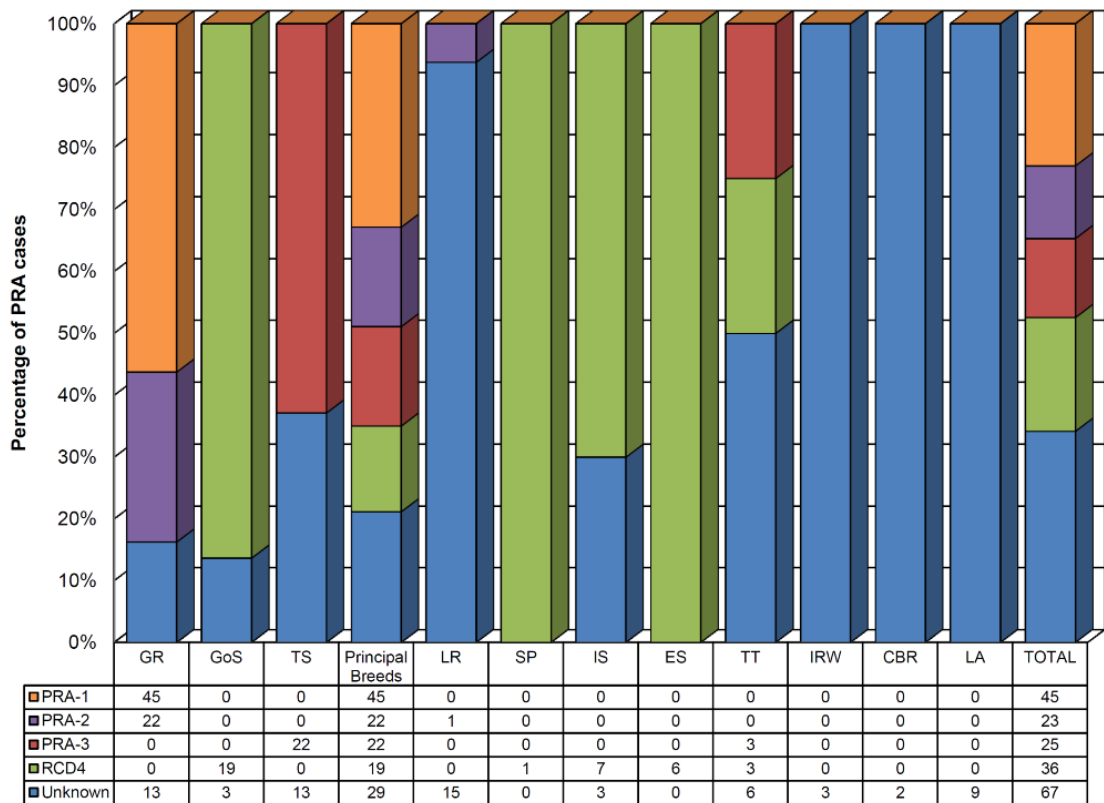


Figure 50. Mutations accounting for PRA in multiple breeds

The mutations identified account for the majority of PRA cases in the three principal breeds investigated (GR, GoS and TS). A form of PRA of unknown aetiology (blue) still segregates in most of the breeds investigated.

6.6 Future prospects

6.6.1 Identification of Remaining PRA Mutations

The molecular characterisation of the major forms of PRA in three principal breeds also revealed that PRA is heterogeneous in these breeds. Indeed, PRA of unknown aetiology segregates in 9/11 breeds investigated (Figure 50). As only a single SP with PRA was tested, the possibility that other forms of PRA, in addition to RCD4, segregate in the breed cannot be eliminated. Due to the low numbers of samples available from dogs with unknown PRA, a GWA approach will not be an appropriate strategy for the identification of additional genetic variants. However, with the continuing development of NGS technologies, **whole genome** or **whole exome sequencing** could be undertaken. The cost of enriching canine whole exome for sequencing is still prohibitively expensive (approximately £250 per sample), while similar kits designed for human use are a fraction of the cost (<£100 per sample). In addition, as demonstrated throughout this project, the gene annotation of the canine genome is incomplete. While this aspect has been greatly improved with the recent upgrade from CanFam2.0 to CanFam3.1, it is still not as well annotated at the human genome, and whole exons are often missing. In a recent trial conducted by Oliver Forman (AHT), sequencing of canine gDNA using the Nextera Human Exome kit resulted in capture and sequencing of 79% of canine exons (personal communication, Oliver Forman). In addition, 18/19 known PRA mutations (including the four mutations identified in this project) are exonic, and the locations of 15 were captured and sequenced during this trial (personal communication, Sally Ricketts). Therefore, almost 80% (15/19) of known PRA mutations could, in theory at least, have been identified through whole exome sequencing using the human exome enrichment kit. Of course, the reason that most of the mutations identified to date have been exonic is most likely that they are the easiest mutations to find, while intronic or intergenic mutations are not only more difficult to identify in the first place, but it is also difficult to prove their pathogenicity. Exome sequencing will therefore not always yield positive results.

Alternatively, an **autozygosity mapping** approach could be undertaken to identify candidate regions, genotyping PRA cases and closely related controls. Two of the GoS PRA cases that are not attributed to RCD4 (as discussed in Chapter 4) are closely related (Figure 51). Obtaining DNA samples from the close relatives that haven't yet been sampled, and obtaining SNP genotypes for all of these dogs using a canine SNP microarray, could result in the identification of loci homozygous in both affected dogs,

but not in all of the controls. A recent development is the combination of exome sequencing and autozygosity mapping. Carr et al. demonstrated the use of whole exome sequence data to simultaneously identify regions of autozygosity and identify all possible deleterious variants in those regions²⁰².

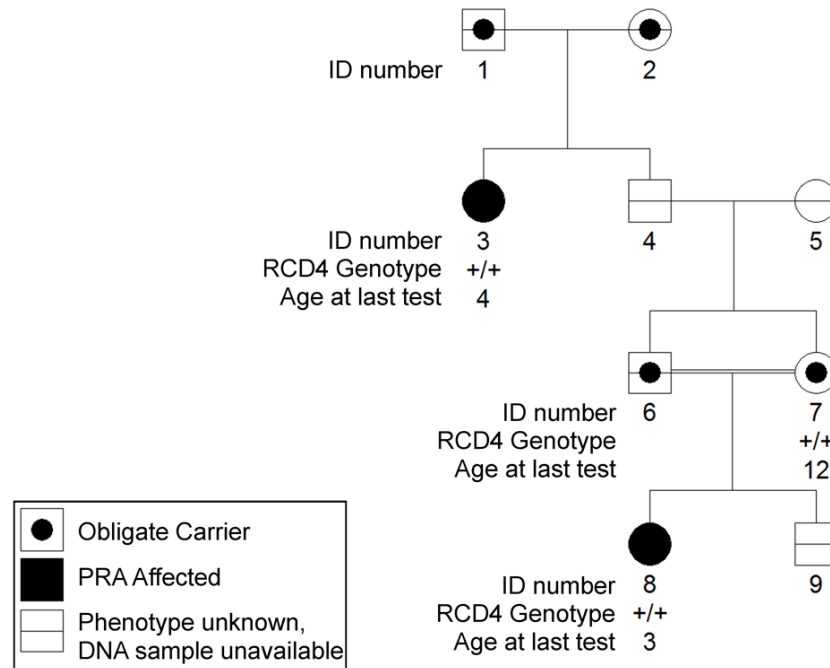


Figure 51. Segregation of non-RCD4 PRA in a GoS family

Two closely related individuals (3 and 8) are affected with a form PRA distinct from RCD4. DNA samples have been obtained from only three dogs (3, 7 and 8), and all three are homozygous for the wildtype allele at the RCD4-associated locus (*C2ORF71*^{+/+}).

6.6.2 In Vitro Functional Investigations

In the absence of suitable (retinal) tissues from dogs in which to investigate the functional consequences of the mutations described, it may be possible to conduct investigations in vitro. Site-directed mutagenesis could be undertaken to introduce the desired mutations into the cDNA of the gene of interest, which could then be transfected into an appropriate cell line. Sun et al. investigated the expression of wildtype *ABCA4* (mutations in which have been implicated in Stargardt disease, macular degeneration, aRCD and aRRP in humans, as reviewed by Koeneke²⁰³) in human embryonic kidney (293) cells, and compared it with 37 *ABCA4* mutants²⁰⁴. Their experiments revealed a wide variety of biochemical defects including reduced protein levels and defective ATP-binding.

6.7 Conclusions

This study demonstrates the value of dogs from the pet population in the search for disease or trait mutations, as the potential sample pool is very large. However, there are disadvantages to be considered. Unlike dogs in research colonies it is extremely difficult, if not virtually impossible, to obtain consistent and detailed clinical information due to the variety of ophthalmologists involved, the desires of the owners and ethical considerations. In addition downstream investigations, such as RNA or functional assays, are limited by the availability of appropriate tissues.

Nevertheless, in this study four new mutations were identified that are associated with PRA. To date, eight breeds have been identified that are affected by at least one of these mutations, and breeders can now benefit from this information by testing their dogs for the relevant variants and thereby improve the overall health of their breeds.

References

1. Barnett, K.C. Genetic anomalies of the posterior segment of the canine eye. *J Small Anim Pract* **10**, 451-5 (1969).
2. Smith, A.J., Bainbridge, J.W. & Ali, R.R. Prospects for retinal gene replacement therapy. *Trends Genet* **25**, 156-65 (2009).
3. Koenekoop, R.K. Why do cone photoreceptors die in rod-specific forms of retinal degenerations? *Ophthalmic Genet* **30**, 152-4 (2009).
4. Wright, A.F., Chakarova, C.F., Abd El-Aziz, M.M. & Bhattacharya, S.S. Photoreceptor degeneration: genetic and mechanistic dissection of a complex trait. *Nat Rev Genet* **11**, 273-84 (2010).
5. Liu, X., Udovichenko, I.P., Brown, S.D., Steel, K.P. & Williams, D.S. Myosin VIIa participates in opsin transport through the photoreceptor cilium. *J Neurosci* **19**, 6267-74 (1999).
6. Burns, M.E. & Arshavsky, V.Y. Beyond counting photons: trials and trends in vertebrate visual transduction. *Neuron* **48**, 387-401 (2005).
7. Mata, N.L., Radu, R.A., Clemmons, R.C. & Travis, G.H. Isomerization and oxidation of vitamin a in cone-dominant retinas: a novel pathway for visual-pigment regeneration in daylight. *Neuron* **36**, 69-80 (2002).
8. Kanan, Y., Kasus-Jacobi, A., Moiseyev, G., Sawyer, K., Ma, J.X. et al. Retinoid processing in cone and Muller cell lines. *Exp Eye Res* **86**, 344-54 (2008).
9. Pirie, A. Crystals of riboflavin making up the tapetum lucidum in the eye of a lemur. *Nature* **183**, 985-6 (1959).
10. Miller, P.E. & Murphy, C.J. Vision in dogs. *J Am Vet Med Assoc* **207**, 1623-34 (1995).
11. Khachik, F., Bernstein, P.S. & Garland, D.L. Identification of lutein and zeaxanthin oxidation products in human and monkey retinas. *Invest Ophthalmol Vis Sci* **38**, 1802-11 (1997).
12. Peichl, L. Topography of ganglion cells in the dog and wolf retina. *J Comp Neurol* **324**, 603-20 (1992).
13. Rapaport, D.H. & Stone, J. The area centralis of the retina in the cat and other mammals: focal point for function and development of the visual system. *Neuroscience* **11**, 289-301 (1984).
14. McGreevy, P., Grassi, T.D. & Harman, A.M. A strong correlation exists between the distribution of retinal ganglion cells and nose length in the dog. *Brain Behav Evol* **63**, 13-22 (2004).
15. Neitz, J., Geist, T. & Jacobs, G.H. Color vision in the dog. *Vis Neurosci* **3**, 119-25 (1989).
16. Jacobs, G.H., Deegan, J.F., 2nd, Crognale, M.A. & Fenwick, J.A. Photopigments of dogs and foxes and their implications for canid vision. *Vis Neurosci* **10**, 173-80 (1993).
17. Vila, C., Savolainen, P., Maldonado, J.E., Amorim, I.R., Rice, J.E. et al. Multiple and ancient origins of the domestic dog. *Science* **276**, 1687-9 (1997).

18. Savolainen, P., Zhang, Y.P., Luo, J., Lundeberg, J. & Leitner, T. Genetic evidence for an East Asian origin of domestic dogs. *Science* **298**, 1610-3 (2002).
19. Mellersh, C.S., Langston, A.A., Acland, G.M., Fleming, M.A., Ray, K. et al. A linkage map of the canine genome. *Genomics* **46**, 326-36 (1997).
20. Neff, M.W., Broman, K.W., Mellersh, C.S., Ray, K., Acland, G.M. et al. A second-generation genetic linkage map of the domestic dog, *Canis familiaris*. *Genetics* **151**, 803-20 (1999).
21. Breen, M., Hitte, C., Lorentzen, T.D., Thomas, R., Cadieu, E. et al. An integrated 4249 marker FISH/RH map of the canine genome. *BMC Genomics* **5**, 65 (2004).
22. Kirkness, E.F., Bafna, V., Halpern, A.L., Levy, S., Remington, K. et al. The dog genome: survey sequencing and comparative analysis. *Science* **301**, 1898-903 (2003).
23. Lindblad-Toh, K., Wade, C.M., Mikkelsen, T.S., Karlsson, E.K., Jaffe, D.B. et al. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438**, 803-19 (2005).
24. Parker, H.G., Kim, L.V., Sutter, N.B., Carlson, S., Lorentzen, T.D. et al. Genetic structure of the purebred domestic dog. *Science* **304**, 1160-4 (2004).
25. Ostrander, E.A. & Wayne, R.K. The canine genome. *Genome Res* **15**, 1706-16 (2005).
26. Ostrander, E.A., Galibert, F. & Patterson, D.F. Canine genetics comes of age. *Trends Genet* **16**, 117-24 (2000).
27. Sargan, D.R. IDID: inherited diseases in dogs: web-based information for canine inherited disease genetics. *Mamm Genome* **15**, 503-6 (2004).
28. Petersen-Jones, S. Advances in the molecular understanding of canine retinal diseases. *J Small Anim Pract* **46**, 371-80 (2005).
29. Barnett, K. Inherited eye disease in the dog and cat. *J Small Anim Pract* **29**, 462-475 (1988).
30. Mellersh, C.S., Boursnell, M.E., Pettitt, L., Ryder, E.J., Holmes, N.G. et al. Canine RPGRIP1 mutation establishes cone-rod dystrophy in miniature longhaired dachshunds as a homologue of human Leber congenital amaurosis. *Genomics* **88**, 293-301 (2006).
31. Petersen-Jones, S.M., Entz, D.D. & Sargan, D.R. cGMP phosphodiesterase-alpha mutation causes progressive retinal atrophy in the Cardigan Welsh corgi dog. *Invest Ophthalmol Vis Sci* **40**, 1637-44 (1999).
32. Kijas, J.W., Cideciyan, A.V., Aleman, T.S., Pianta, M.J., Pearce-Kelling, S.E. et al. Naturally occurring rhodopsin mutation in the dog causes retinal dysfunction and degeneration mimicking human dominant retinitis pigmentosa. *Proc Natl Acad Sci U S A* **99**, 6328-33 (2002).
33. Acland, G.M., Blanton, S.H., Hershfield, B. & Aguirre, G.D. XLPRA: a canine retinal degeneration inherited as an X-linked trait. *Am J Med Genet* **52**, 27-33 (1994).
34. Mellersh, C.S. Genetics of Eye Disorders in the Dog, in *The Genetics of the Dog* (eds. Ostrander, E.A. & Ruvinsky, A.) (CAB International, 2012).

35. Parker, H.G., Kukekova, A.V., Akey, D.T., Goldstein, O., Kirkness, E.F. et al. Breed relationships facilitate fine-mapping studies: a 7.8-kb deletion cosegregates with Collie eye anomaly across multiple dog breeds. *Genome Res* **17**, 1562-71 (2007).
36. Goldstein, O., Guyon, R., Kukekova, A., Kuznetsova, T.N., Pearce-Kelling, S.E. et al. COL9A2 and COL9A3 mutations in canine autosomal recessive ocular skeletal dysplasia. *Mamm Genome* **21**, 398-408 (2010).
37. Sidjanin, D.J., Lowe, J.K., McElwee, J.L., Milne, B.S., Phippen, T.M. et al. Canine CNGB3 mutations establish cone degeneration as orthologous to the human achromatopsia locus ACHM3. *Hum Mol Genet* **11**, 1823-33 (2002).
38. Guziewicz, K.E., Zangerl, B., Lindauer, S.J., Mullins, R.F., Sandmeyer, L.S. et al. Bestrophin gene mutations cause canine multifocal retinopathy: a novel animal model for best disease. *Invest Ophthalmol Vis Sci* **48**, 1959-67 (2007).
39. Zangerl, B., Wickstrom, K., Slavik, J., Lindauer, S.J., Ahonen, S. et al. Assessment of canine BEST1 variations identifies new mutations and establishes an independent bestrophinopathy model (cmr3). *Mol Vis* **16**, 2791-804 (2010).
40. Wiik, A.C., Wade, C., Biagi, T., Ropstad, E.O., Bjerkas, E. et al. A deletion in nephronophthisis 4 (NPHP4) is associated with recessive cone-rod dystrophy in standard wire-haired dachshund. *Genome Res* **18**, 1415-21 (2008).
41. Kropatsch, R., Petrasch-Parwez, E., Seelow, D., Schlichting, A., Gerding, W.M. et al. Generalized progressive retinal atrophy in the Irish Glen of Imaal Terrier is associated with a deletion in the ADAM9 gene. *Mol Cell Probes* **24**, 357-63 (2010).
42. Goldstein, O., Mezey, J.G., Boyko, A.R., Gao, C., Wang, W. et al. An ADAM9 mutation in canine cone-rod dystrophy 3 establishes homology with human cone-rod dystrophy 9. *Mol Vis* **16**, 1549-69 (2010).
43. Veske, A., Nilsson, S.E., Narfstrom, K. & Gal, A. Retinal dystrophy of Swedish briard/briard-beagle dogs is due to a 4-bp deletion in RPE65. *Genomics* **57**, 57-61 (1999).
44. Kijas, J.W., Miller, B.J., Pearce-Kelling, S.E., Aguirre, G.D. & Acland, G.M. Canine models of ocular disease: outcross breedings define a dominant disorder present in the English mastiff and bull mastiff dog breeds. *J Hered* **94**, 27-30 (2003).
45. Goldstein, O., Kukekova, A.V., Aguirre, G.D. & Acland, G.M. Exonic SINE insertion in STK38L causes canine early retinal degeneration (erd). *Genomics* **96**, 362-8 (2010).
46. Dekomien, G., Vollrath, C., Petrasch-Parwez, E., Boeve, M.H., Akkad, D.A. et al. Progressive retinal atrophy in Schapendoes dogs: mutation of the newly identified CCDC66 gene. *Neurogenetics* **11**, 163-74 (2010).
47. Zangerl, B., Goldstein, O., Philp, A.R., Lindauer, S.J., Pearce-Kelling, S.E. et al. Identical mutation in a novel retinal gene causes progressive rod-cone degeneration in dogs and retinitis pigmentosa in humans. *Genomics* **88**, 551-63 (2006).
48. Suber, M.L., Pittler, S.J., Qin, N., Wright, G.C., Holcombe, V. et al. Irish setter dogs affected with rod/cone dysplasia contain a nonsense mutation in the rod cGMP phosphodiesterase beta-subunit gene. *Proc Natl Acad Sci U S A* **90**, 3968-72 (1993).
49. Dekomien, G., Runte, M., Godde, R. & Epplen, J.T. Generalized progressive retinal atrophy of Sloughi dogs is due to an 8-bp insertion in exon 21 of the PDE6B gene. *Cytogenet Cell Genet* **90**, 261-7 (2000).

50. Kukekova, A.V., Goldstein, O., Johnson, J.L., Richardson, M.A., Pearce-Kelling, S.E. et al. Canine RD3 mutation establishes rod-cone dysplasia type 2 (rcd2) as ortholog of human and murine rd3. *Mamm Genome* **20**, 109-23 (2009).
51. Zhang, Q., Acland, G.M., Parshall, C.J., Haskell, J., Ray, K. et al. Characterization of canine photoreceptor phosphodiesterase cDNA and identification of a sequence variant in dogs with photoreceptor dysplasia. *Gene* **215**, 231-9 (1998).
52. Zhang, Q., Acland, G.M., Wu, W.X., Johnson, J.L., Pearce-Kelling, S. et al. Different RPGR exon ORF15 mutations in Canids provide insights into photoreceptor cell degeneration. *Hum Mol Genet* **11**, 993-1003 (2002).
53. Petersen-Jones, S.M. Progressive Retinal Atrophy: An Overview. in *28th World Congress of the World Small Animal Veterinary Association* (Bangkok, Thailand, 2003).
54. Clements, P.J., Sargan, D.R., Gould, D.J. & Petersen-Jones, S.M. Recent advances in understanding the spectrum of canine generalised progressive retinal atrophy. *J Small Anim Pract* **37**, 155-62 (1996).
55. Farber, D.B. & Lolley, R.N. Cyclic guanosine monophosphate: elevation in degenerating photoreceptor cells of the C3H mouse retina. *Science* **186**, 449-51 (1974).
56. Insinna, C. & Besharse, J.C. Intraflagellar transport and the sensory outer segment of vertebrate photoreceptors. *Dev Dyn* **237**, 1982-92 (2008).
57. Berger, W., Kloeckener-Gruissem, B. & Neidhardt, J. The molecular basis of human retinal and vitreoretinal diseases. *Prog Retin Eye Res* **29**, 335-75 (2010).
58. Bunker, C.H., Berson, E.L., Bromley, W.C., Hayes, R.P. & Roderick, T.H. Prevalence of retinitis pigmentosa in Maine. *Am J Ophthalmol* **97**, 357-65 (1984).
59. Phelan, J.K. & Bok, D. A brief review of retinitis pigmentosa and the identified retinitis pigmentosa genes. *Mol Vis* **6**, 116-24 (2000).
60. den Hollander, A.I., Roepman, R., Koenekeop, R.K. & Cremers, F.P. Leber congenital amaurosis: genes, proteins and disease mechanisms. *Prog Retin Eye Res* **27**, 391-419 (2008).
61. Walia, S. & Fishman, G.A. Natural history of phenotypic changes in Stargardt macular dystrophy. *Ophthalmic Genet* **30**, 63-8 (2009).
62. den Hollander, A.I., Black, A., Bennett, J. & Cremers, F.P. Lighting a candle in the dark: advances in genetics and gene therapy of recessive retinal dystrophies. *J Clin Invest* **120**, 3042-53 (2010).
63. Michaelides, M., Hunt, D.M. & Moore, A.T. The cone dysfunction syndromes. *Br J Ophthalmol* **88**, 291-7 (2004).
64. Katsanis, N., Ansley, S.J., Badano, J.L., Eichers, E.R., Lewis, R.A. et al. Triallelic inheritance in Bardet-Biedl syndrome, a Mendelian recessive disorder. *Science* **293**, 2256-9 (2001).
65. Badano, J.L., Kim, J.C., Hoskins, B.E., Lewis, R.A., Ansley, S.J. et al. Heterozygous mutations in BBS1, BBS2 and BBS6 have a potential epistatic effect on Bardet-Biedl patients with two mutations at a second BBS locus. *Hum Mol Genet* **12**, 1651-9 (2003).

66. Grall, A., Guaguere, E., Planchais, S., Grond, S., Bourrat, E. et al. PNPLA1 mutations cause autosomal recessive congenital ichthyosis in golden retriever dogs and humans. *Nat Genet* **44**, 140-7 (2012).
67. Chiang, A.P., Beck, J.S., Yen, H.J., Tayeh, M.K., Scheetz, T.E. et al. Homozygosity mapping with SNP arrays identifies TRIM32, an E3 ubiquitin ligase, as a Bardet-Biedl syndrome gene (BBS11). *Proc Natl Acad Sci U S A* **103**, 6287-92 (2006).
68. van Duijn, C.M., Dekker, M.C., Bonifati, V., Galjaard, R.J., Houwing-Duistermaat, J.J. et al. Park7, a novel locus for autosomal recessive early-onset parkinsonism, on chromosome 1p36. *Am J Hum Genet* **69**, 629-34 (2001).
69. Wrigstad, A., Narfstrom, K. & Nilsson, S.E. Slowly progressive changes of the retina and retinal pigment epithelium in Briard dogs with hereditary retinal dystrophy. A morphological study. *Doc Ophthalmol* **87**, 337-54 (1994).
70. Redmond, T.M., Yu, S., Lee, E., Bok, D., Hamasaki, D. et al. Rpe65 is necessary for production of 11-cis-vitamin A in the retinal visual cycle. *Nat Genet* **20**, 344-51 (1998).
71. Cideciyan, A.V., Hauswirth, W.W., Aleman, T.S., Kaushal, S., Schwartz, S.B. et al. Human RPE65 gene therapy for Leber congenital amaurosis: persistence of early visual improvements and safety at 1 year. *Hum Gene Ther* **20**, 999-1004 (2009).
72. Acland, G.M., Aguirre, G.D., Ray, J., Zhang, Q., Aleman, T.S. et al. Gene therapy restores vision in a canine model of childhood blindness. *Nat Genet* **28**, 92-5 (2001).
73. Acland, G.M., Aguirre, G.D., Bennett, J., Aleman, T.S., Cideciyan, A.V. et al. Long-term restoration of rod and cone vision by single dose rAAV-mediated gene transfer to the retina in a canine model of childhood blindness. *Mol Ther* **12**, 1072-82 (2005).
74. Bainbridge, J.W., Smith, A.J., Barker, S.S., Robbie, S., Henderson, R. et al. Effect of gene therapy on visual function in Leber's congenital amaurosis. *N Engl J Med* **358**, 2231-9 (2008).
75. Hauswirth, W.W., Aleman, T.S., Kaushal, S., Cideciyan, A.V., Schwartz, S.B. et al. Treatment of leber congenital amaurosis due to RPE65 mutations by ocular subretinal injection of adeno-associated virus gene vector: short-term results of a phase I trial. *Hum Gene Ther* **19**, 979-90 (2008).
76. Maguire, A.M., High, K.A., Auricchio, A., Wright, J.F., Pierce, E.A. et al. Age-dependent effects of RPE65 gene therapy for Leber's congenital amaurosis: a phase 1 dose-escalation trial. *Lancet* **374**, 1597-605 (2009).
77. Khanna, H., Hurd, T.W., Lillo, C., Shu, X., Parapuram, S.K. et al. RPGR-ORF15, which is mutated in retinitis pigmentosa, associates with SMC1, SMC3, and microtubule transport proteins. *J Biol Chem* **280**, 33580-7 (2005).
78. Shu, X., Black, G.C., Rice, J.M., Hart-Holden, N., Jones, A. et al. RPGR mutation analysis and disease: an update. *Hum Mutat* **28**, 322-8 (2007).
79. Vervoort, R., Lennon, A., Bird, A.C., Tulloch, B., Axton, R. et al. Mutational hot spot within a new RPGR exon in X-linked retinitis pigmentosa. *Nat Genet* **25**, 462-6 (2000).
80. Sundaram, V., Moore, A.T., Ali, R.R. & Bainbridge, J.W. Retinal dystrophies and gene therapy. *Eur J Pediatr* **171**, 757-65 (2012).

81. Hong, D.H., Pawlyk, B.S., Shang, J., Sandberg, M.A., Berson, E.L. et al. A retinitis pigmentosa GTPase regulator (RPGR)-deficient mouse model for X-linked retinitis pigmentosa (RP3). *Proc Natl Acad Sci U S A* **97**, 3649-54 (2000).
82. Sharon, D., Sandberg, M.A., Rabe, V.W., Stillberger, M., Dryja, T.P. et al. RP2 and RPGR mutations and clinical correlations in patients with X-linked retinitis pigmentosa. *Am J Hum Genet* **73**, 1131-46 (2003).
83. Hong, D.H., Pawlyk, B.S., Adamian, M. & Li, T. Dominant, gain-of-function mutant produced by truncation of RPGR. *Invest Ophthalmol Vis Sci* **45**, 36-41 (2004).
84. Beltran, W.A., Cideciyan, A.V., Lewin, A.S., Iwabe, S., Khanna, H. et al. Gene therapy rescues photoreceptor blindness in dogs and paves the way for treating human X-linked retinitis pigmentosa. *Proc Natl Acad Sci U S A* **109**, 2132-7 (2012).
85. Walia, S., Fishman, G.A., Swaroop, A., Branham, K.E., Lindeman, M. et al. Discordant phenotypes in fraternal twins having an identical mutation in exon ORF15 of the RPGR gene. *Arch Ophthalmol* **126**, 379-84 (2008).
86. Fahim, A.T., Bowne, S.J., Sullivan, L.S., Webb, K.D., Williams, J.T. et al. Allelic heterogeneity and genetic modifier loci contribute to clinical variation in males with X-linked retinitis pigmentosa due to RPGR mutations. *PLoS ONE* **6**, e23021 (2011).
87. Guyon, R., Lorentzen, T.D., Hitte, C., Kim, L., Cadieu, E. et al. A 1-Mb resolution radiation hybrid map of the canine genome. *Proc Natl Acad Sci U S A* **100**, 5296-301 (2003).
88. Sargan, D.R., Aguirre-Hernandez, J., Galibert, F. & Ostrander, E.A. An extended microsatellite set for linkage mapping in the domestic dog. *J Hered* **98**, 221-31 (2007).
89. Sutter, N.B., Eberle, M.A., Parker, H.G., Pullar, B.J., Kirkness, E.F. et al. Extensive and breed-specific linkage disequilibrium in *Canis familiaris*. *Genome Res* **14**, 2388-96 (2004).
90. Aguirre-Hernandez, J. & Sargan, D.R. Evaluation of candidate genes in the absence of positional information: a poor bet on a blind dog! *J Hered* **96**, 475-84 (2005).
91. Acland, G.M., Ray, K., Mellersh, C.S., Gu, W., Langston, A.A. et al. Linkage analysis and comparative mapping of canine progressive rod-cone degeneration (prcd) establishes potential locus homology with retinitis pigmentosa (RP17) in humans. *Proc Natl Acad Sci U S A* **95**, 3048-53 (1998).
92. Goldstein, O., Zangerl, B., Pearce-Kelling, S., Sidjanin, D.J., Kijas, J.W. et al. Linkage disequilibrium mapping in domestic dog breeds narrows the progressive rod-cone degeneration interval and identifies ancestral disease-transmitting chromosome. *Genomics* **88**, 541-50 (2006).
93. Kruglyak, L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat Genet* **22**, 139-44 (1999).
94. Altshuler, D., Daly, M.J. & Lander, E.S. Genetic mapping in human disease. *Science* **322**, 881-8 (2008).
95. Karlsson, E.K., Baranowska, I., Wade, C.M., Salmon Hillbertz, N.H., Zody, M.C. et al. Efficient mapping of mendelian traits in dogs through genome-wide association. *Nat Genet* **39**, 1321-8 (2007).

96. Karlsson, E.K. & Lindblad-Toh, K. Leader of the pack: gene mapping in dogs and other model organisms. *Nat Rev Genet* **9**, 713-25 (2008).
97. Kijas, J.M., Bauer, T.R., Jr., Gafvert, S., Marklund, S., Trowald-Wigh, G. et al. A missense mutation in the beta-2 integrin gene (ITGB2) causes canine leukocyte adhesion deficiency. *Genomics* **61**, 101-7 (1999).
98. Rozen, S. & Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* **132**, 365-86 (2000).
99. Copeland, H., Dukes-McEwan, J., Sargan, D., Kennedy, L., Starkey, M. et al. LUPA - studying human diseases using dog genetics. *Vet Rec* **163**, 550 (2008).
100. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).
101. Astle, W. & Balding, D.J. Population Structure and Cryptic Relatedness in Genetic Association Studies. *Statist. Sci.* **24**, 451-471 (2009).
102. Stephens, M., Smith, N.J. & Donnelly, P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* **68**, 978-89 (2001).
103. Thompson, J.D., Higgins, D.G. & Gibson, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673-80 (1994).
104. Bonfield, J.K., Smith, K. & Staden, R. A new DNA sequence assembly program. *Nucleic Acids Res* **23**, 4992-9 (1995).
105. Lunter, G. & Goodson, M. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res* **21**, 936-9 (2011).
106. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-60 (2009).
107. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-9 (2009).
108. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-303 (2010).
109. McLaren, W., Pritchard, B., Rios, D., Chen, Y., Flicek, P. et al. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* **26**, 2069-70 (2010).
110. Ye, K., Schulz, M.H., Long, Q., Apweiler, R. & Ning, Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **25**, 2865-71 (2009).
111. Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S. et al. Integrative genomics viewer. *Nat Biotechnol* **29**, 24-6 (2011).
112. Thorvaldsdottir, H., Robinson, J.T. & Mesirov, J.P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* (2012).

113. Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N. et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* **3**, RESEARCH0034 (2002).
114. Pfaffl, M.W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* **29**, e45 (2001).
115. Reese, M.G., Eeckman, F.H., Kulp, D. & Haussler, D. Improved splice site detection in Genie. *J Comput Biol* **4**, 311-23 (1997).
116. Kumar, P., Henikoff, S. & Ng, P.C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* **4**, 1073-81 (2009).
117. Sunyaev, S., Ramensky, V., Koch, I., Lathe, W., 3rd, Kondrashov, A.S. et al. Prediction of deleterious human alleles. *Hum Mol Genet* **10**, 591-7 (2001).
118. Ferrer-Costa, C., Gelpi, J.L., Zamakola, L., Parraga, I., de la Cruz, X. et al. PMUT: a web-based tool for the annotation of pathological mutations on proteins. *Bioinformatics* **21**, 3176-8 (2005).
119. Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D. et al. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* **31**, 3784-8 (2003).
120. Quignon, P., Herbin, L., Cadieu, E., Kirkness, E.F., Hedan, B. et al. Canine population structure: assessment and impact of intra-breed stratification on SNP-based association studies. *PLoS ONE* **2**, e1324 (2007).
121. Gough, A. & Thomas, A. *Breed Predispositions to Disease in Dogs and Cats*, (Blackwell Publishing, 2004).
122. Fogle, B. *The Encyclopedia of the Dog*, (DK Publishing, 1995).
123. Kopito, R.R. Molecular biology of the anion exchanger gene family. *Int Rev Cytol* **123**, 177-99 (1990).
124. Kobayashi, S., Morgans, C.W., Casey, J.R. & Kopito, R.R. AE3 anion exchanger isoforms in the vertebrate retina: developmental regulation and differential expression in neurons and glia. *J Neurosci* **14**, 6266-79 (1994).
125. Alvarez, B.V., Gilmour, G.S., Mema, S.C., Martin, B.T., Shull, G.E. et al. Blindness caused by deficiency in AE3 chloride/bicarbonate exchanger. *PLoS ONE* **2**, e839 (2007).
126. Linn, S.C., Kudrycki, K.E. & Shull, G.E. The predicted translation product of a cardiac AE3 mRNA contains an N terminus distinct from that of the brain AE3 Cl⁻/HCO₃⁻-exchanger. Cloning of a cardiac AE3 cDNA, organization of the AE3 gene, and identification of an alternative transcription initiation site. *J Biol Chem* **267**, 7927-35 (1992).
127. Wang, H., den Hollander, A.I., Moayed, Y., Abulimiti, A., Li, Y. et al. Mutations in SPATA7 cause Leber congenital amaurosis and juvenile retinitis pigmentosa. *Am J Hum Genet* **84**, 380-7 (2009).
128. Ansley, S.J., Badano, J.L., Blacque, O.E., Hill, J., Hoskins, B.E. et al. Basal body dysfunction is a likely cause of pleiotropic Bardet-Biedl syndrome. *Nature* **425**, 628-33 (2003).

129. Stoetzel, C., Laurier, V., Faivre, L., Megarbane, A., Perrin-Schmitt, F. et al. BBS8 is rarely mutated in a cohort of 128 Bardet-Biedl syndrome families. *J Hum Genet* **51**, 81-4 (2006).
130. Riazuddin, S.A., Iqbal, M., Wang, Y., Masuda, T., Chen, Y. et al. A splice-site mutation in a retina-specific exon of BBS8 causes nonsyndromic retinitis pigmentosa. *Am J Hum Genet* **86**, 805-12 (2010).
131. Zhang, X., Liu, H., Zhang, Y., Qiao, Y., Miao, S. et al. A novel gene, RSD-3/HSD-3.1, encodes a meiotic-related protein expressed in rat and human testis. *J Mol Med* **81**, 380-7 (2003).
132. Gould, D., Pettitt, L., McLaughlin, B., Holmes, N., Forman, O. et al. ADAMTS17 mutation associated with primary lens luxation is widespread among breeds. *Vet Ophthalmol* **14**, 378-84 (2011).
133. Yannoukakos, D., Stuart-Tilley, A., Fernandez, H.A., Fey, P., Duyk, G. et al. Molecular cloning, expression, and chromosomal localization of two isoforms of the AE3 anion exchanger from human heart. *Circ Res* **75**, 603-14 (1994).
134. Kopito, R.R., Lee, B.S., Simmons, D.M., Lindsey, A.E., Morgans, C.W. et al. Regulation of intracellular pH by a neuronal homolog of the erythrocyte anion exchanger. *Cell* **59**, 927-37 (1989).
135. Alper, S.L. Molecular physiology and genetics of Na⁺-independent SLC4 anion exchangers. *J Exp Biol* **212**, 1672-83 (2009).
136. Sterling, D. & Casey, J.R. Transport activity of AE3 chloride/bicarbonate anion-exchange proteins and their regulation by intracellular pH. *Biochem J* **344 Pt 1**, 221-9 (1999).
137. Seddon, J.M., Hampson, E.C., Smith, R.I. & Hughes, I.P. Genetic heterogeneity of day blindness in Alaskan Malamutes. *Anim Genet* **37**, 407-10 (2006).
138. Hentschke, M., Wiemann, M., Hentschke, S., Kurth, I., Hermans-Borgmeyer, I. et al. Mice with a targeted disruption of the Cl⁻/HCO₃⁻ exchanger AE3 display a reduced seizure threshold. *Mol Cell Biol* **26**, 182-91 (2006).
139. Prasad, V., Bodi, I., Meyer, J.W., Wang, Y., Ashraf, M. et al. Impaired cardiac contractility in mice lacking both the AE3 Cl⁻/HCO₃⁻ exchanger and the NKCC1 Na⁺-K⁺-2Cl⁻ cotransporter: effects on Ca²⁺ handling and protein phosphatases. *J Biol Chem* **283**, 31303-14 (2008).
140. Alper, S.L., Darman, R.B., Chernova, M.N. & Dahl, N.K. The AE gene family of Cl⁻/HCO₃⁻ exchangers. *J Nephrol* **15 Suppl 5**, S41-53 (2002).
141. Sterling, D., Alvarez, B.V. & Casey, J.R. The extracellular component of a transport metabolon. Extracellular loop 4 of the human AE1 Cl⁻/HCO₃⁻ exchanger binds carbonic anhydrase IV. *J Biol Chem* **277**, 25239-46 (2002).
142. Morgan, P.E., Pastorekova, S., Stuart-Tilley, A.K., Alper, S.L. & Casey, J.R. Interactions of transmembrane carbonic anhydrase, CAIX, with bicarbonate transporters. *Am J Physiol Cell Physiol* **293**, C738-48 (2007).
143. Vince, J.W. & Reithmeier, R.A. Identification of the carbonic anhydrase II binding site in the Cl⁻/HCO₃⁻ anion exchanger AE1. *Biochemistry* **39**, 5527-33 (2000).
144. Sterling, D., Reithmeier, R.A. & Casey, J.R. Carbonic anhydrase: in the driver's seat for bicarbonate transport. *JOP* **2**, 165-70 (2001).

145. Newman, E.A. & Astion, M.L. Localization and stoichiometry of electrogenic sodium bicarbonate cotransport in retinal glial cells. *Glia* **4**, 424-8 (1991).
146. Winkler, B.S., Simson, V. & Benner, J. Importance of bicarbonate in retinal function. *Invest Ophthalmol Vis Sci* **16**, 766-8 (1977).
147. Kanner, B.I. Structure and function of sodium-coupled neurotransmitter transporters. *Soc Gen Physiol Ser* **48**, 243-50 (1993).
148. Nachury, M.V., Loktev, A.V., Zhang, Q., Westlake, C.J., Peranen, J. et al. A core complex of BBS proteins cooperates with the GTPase Rab8 to promote ciliary membrane biogenesis. *Cell* **129**, 1201-13 (2007).
149. Jin, H., White, S.R., Shida, T., Schulz, S., Aguiar, M. et al. The conserved Bardet-Biedl syndrome proteins assemble a coat that traffics membrane proteins to cilia. *Cell* **141**, 1208-19 (2010).
150. Blatch, G.L. & Lassar, M. The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *Bioessays* **21**, 932-9 (1999).
151. Maurer, L. & Orndorff, P.E. Identification and characterization of genes determining receptor binding and pilus length of Escherichia coli type 1 pili. *J Bacteriol* **169**, 640-5 (1987).
152. Watson, A.A., Alm, R.A. & Mattick, J.S. Identification of a gene, pilF, required for type 4 fimbrial biogenesis and twitching motility in Pseudomonas aeruginosa. *Gene* **180**, 49-56 (1996).
153. Sander, T., Toliat, M.R., Heils, A., Leschik, G., Becker, C. et al. Association of the 867Asp variant of the human anion exchanger 3 gene with common subtypes of idiopathic generalized epilepsy. *Epilepsy Res* **51**, 249-55 (2002).
154. Chiappe de Cingolani, G.E., Ennis, I.L., Morgan, P.E., Alvarez, B.V., Casey, J.R. et al. Involvement of AE3 isoform of Na(+)-independent Cl(-)/HCO(3)(-) exchanger in myocardial pH(i) recovery from intracellular alkalization. *Life Sci* **78**, 3018-26 (2006).
155. Al Moamen, N.J., Prasad, V., Bodi, I., Miller, M.L., Neiman, M.L. et al. Loss of the AE3 anion exchanger in a hypertrophic cardiomyopathy model causes rapid decompensation and heart failure. *J Mol Cell Cardiol* **50**, 137-46 (2011).
156. Collin, R.W., Safieh, C., Littink, K.W., Shalev, S.A., Garzozzi, H.J. et al. Mutations in C2ORF71 cause autosomal-recessive retinitis pigmentosa. *Am J Hum Genet* **86**, 783-8 (2010).
157. Nishimura, D.Y., Baye, L.M., Perveen, R., Searby, C.C., Avila-Fernandez, A. et al. Discovery and functional analysis of a retinitis pigmentosa gene, C2ORF71. *Am J Hum Genet* **86**, 686-95 (2010).
158. Sergouniotis, P.I., Li, Z., Mackay, D.S., Wright, G.A., Borman, A.D. et al. A survey of DNA variation of C2ORF71 in probands with progressive autosomal recessive retinal degeneration and controls. *Invest Ophthalmol Vis Sci* **52**, 1880-6 (2010).
159. Barnett, K.C. Canine Retinopathies - I. History and Review of the Literature. *J Small Anim Pract* **6**, 41-55 (1965).
160. Magnusson, H. Über Retinitis Pigmentosa und Konsanguinität beim Hunde. *Arch. Vergh. Ophthal.* **2**, 147-163 (1911).

161. Scougall, B. & Horrocks, Y. Progressive Retinal Atrophy (PRA). in *Dog World* (Breed Notes, 2009).
162. Ray, K., Baldwin, V.J., Acland, G.M., Blanton, S.H. & Aguirre, G.D. Cosegregation of codon 807 mutation of the canine rod cGMP phosphodiesterase beta gene and rcd1. *Invest Ophthalmol Vis Sci* **35**, 4291-9 (1994).
163. Djajadiningrat-Laanen, S.C., Boeve, M.H., Stades, F.C. & van Oost, B.A. Familial non-rcd1 generalised retinal degeneration in Irish setters. *J Small Anim Pract* **44**, 113-6 (2003).
164. Audo, I., Lancelot, M.E., Mohand-Said, S., Antonio, A., Germain, A. et al. Novel C2orf71 mutations account for approximately 1% of cases in a large French arRP cohort. *Hum Mutat* **32**, E2091-103 (2011).
165. Kay, B.K., Williamson, M.P. & Sudol, M. The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. *FASEB J* **14**, 231-41 (2000).
166. Gu, S., Kumaramanickavel, G., Srikumari, C.R., Denton, M.J. & Gal, A. Autosomal recessive retinitis pigmentosa locus RP28 maps between D2S1337 and D2S286 on chromosome 2p11-p15 in an Indian family. *J Med Genet* **36**, 705-7 (1999).
167. Kumar, A., Shetty, J., Kumar, B. & Blanton, S.H. Confirmation of linkage and refinement of the RP28 locus for autosomal recessive retinitis pigmentosa on chromosome 2p14-p15 in an Indian family. *Mol Vis* **10**, 399-402 (2004).
168. Bandah-Rozenfeld, D., Mizrahi-Meissonnier, L., Farhy, C., Obolensky, A., Chowers, I. et al. Homozygosity mapping reveals null mutations in FAM161A as a cause of autosomal-recessive retinitis pigmentosa. *Am J Hum Genet* **87**, 382-91 (2010).
169. Langmann, T., Di Gioia, S.A., Rau, I., Stohr, H., Maksimovic, N.S. et al. Nonsense mutations in FAM161A cause RP28-associated recessive retinitis pigmentosa. *Am J Hum Genet* **87**, 376-81 (2010).
170. Minnick, M.F., Stillwell, L.C., Heineman, J.M. & Stiegler, G.L. A highly repetitive DNA sequence possibly unique to canids. *Gene* **110**, 235-8 (1992).
171. Johnson, J.M., Castle, J., Garrett-Engele, P., Kan, Z., Loerch, P.M. et al. Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* **302**, 2141-4 (2003).
172. Lewis, B.P., Green, R.E. & Brenner, S.E. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc Natl Acad Sci U S A* **100**, 189-92 (2003).
173. Lin, L., Faraco, J., Li, R., Kadotani, H., Rogers, W. et al. The sleep disorder canine narcolepsy is caused by a mutation in the hypocretin (orexin) receptor 2 gene. *Cell* **98**, 365-76 (1999).
174. Clark, L.A., Wahl, J.M., Rees, C.A. & Murphy, K.E. Retrotransposon insertion in SILV is responsible for merle patterning of the domestic dog. *Proc Natl Acad Sci U S A* **103**, 1376-81 (2006).
175. Fairbanks, D.J. & Anderson, W.R. (eds.). *Genetics: The Continuity of Life*, (Brooks/Cole Publishing Company, 1999).

176. Reed, R. & Maniatis, T. The role of the mammalian branchpoint sequence in pre-mRNA splicing. *Genes Dev* **2**, 1268-76 (1988).
177. Reed, R. & Maniatis, T. Intron sequences involved in lariat formation during pre-mRNA splicing. *Cell* **41**, 95-105 (1985).
178. Kuznetsova, T., Iwabe, S., Boesze-Battaglia, K., Pearce-Kelling, S., Chang-Min, Y. et al. Exclusion of RPGRIP1 ins44 from primary causal association with early-onset cone-rod dystrophy in dogs. *Invest Ophthalmol Vis Sci* **53**, 5486-501 (2012).
179. Miyadera, K., Kato, K., Aguirre-Hernandez, J., Tokuriki, T., Morimoto, K. et al. Phenotypic variation and genotype-phenotype discordance in canine cone-rod dystrophy with an RPGRIP1 mutation. *Mol Vis* **15**, 2287-305 (2009).
180. Miyadera, K., Brierley, I., Aguirre-Hernandez, J., Mellersh, C.S. & Sargan, D.R. Multiple mechanisms contribute to leakiness of a frameshift mutation in canine cone-rod dystrophy. *PLoS ONE* **7**, e51598 (2012).
181. Miyadera, K., Kato, K., Bournsnell, M., Mellersh, C.S. & Sargan, D.R. Genome-wide association study in RPGRIP1(-/-) dogs identifies a modifier locus that determines the onset of retinal degeneration. *Mamm Genome* **23**, 212-23 (2011).
182. Roy-Engel, A.M., Salem, A.H., Oyeniran, O.O., Deininger, L., Hedges, D.J. et al. Active Alu element "A-tails": size does matter. *Genome Res* **12**, 1333-44 (2002).
183. Di Gioia, S.A., Letteboer, S.J., Kostic, C., Bandah-Rozenfeld, D., Hetterschijt, L. et al. FAM161A, associated with retinitis pigmentosa, is a component of the cilia-basal body complex and interacts with proteins involved in ciliopathies. *Hum Mol Genet* (2012).
184. Zach, F., Grassmann, F., Langmann, T., Sorusch, N., Wolfrum, U. et al. The retinitis pigmentosa 28 protein FAM161A is a novel ciliary protein involved in intermolecular protein interaction and microtubule association. *Hum Mol Genet* (2012).
185. Freund, C.L., Gregory-Evans, C.Y., Furukawa, T., Papaioannou, M., Looser, J. et al. Cone-rod dystrophy due to mutations in a novel photoreceptor-specific homeobox gene (CRX) essential for maintenance of the photoreceptor. *Cell* **91**, 543-53 (1997).
186. den Hollander, A.I., Koenekoop, R.K., Mohamed, M.D., Arts, H.H., Boldt, K. et al. Mutations in LCA5, encoding the ciliary protein lebercilin, cause Leber congenital amaurosis. *Nat Genet* **39**, 889-95 (2007).
187. Gomez-Baldo, L., Schmidt, S., Maxwell, C.A., Bonifaci, N., Gabaldon, T. et al. TACC3-TSC2 maintains nuclear envelope structure and controls cell division. *Cell Cycle* **9**, 1143-55 (2010).
188. Beltran, W.A. The use of canine models of inherited retinal degeneration to test novel therapeutic approaches. *Vet Ophthalmol* **12**, 192-204 (2009).
189. Forman, O.P., De Risio, L., Stewart, J., Mellersh, C.S. & Beltran, E. Genome-wide mRNA sequencing of a single canine cerebellar cortical degeneration case leads to the identification of a disease associated SPTBN2 mutation. *BMC Genet* **13**, 55 (2012).
190. Dean, S. Dog breeding. *Vet Rec* **169**, 448 (2011).
191. McCarthy, M.I., Abecasis, G.R., Cardon, L.R., Goldstein, D.B., Little, J. et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* **9**, 356-69 (2008).

192. Sillanpaa, M.J. Overview of techniques to account for confounding due to population stratification and cryptic relatedness in genomic data association analyses. *Heredity* **106**, 511-9 (2010).
193. Sanger, F., Nicklen, S. & Coulson, A.R. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**, 5463-7 (1977).
194. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C. et al. Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).
195. Zangerl, B., Johnson, J.L., Acland, G.M. & Aguirre, G.D. Independent origin and restricted distribution of RPGR deletions causing XLPRA. *J Hered* **98**, 526-30 (2007).
196. Levy, S., Sutton, G., Ng, P.C., Feuk, L., Halpern, A.L. et al. The diploid genome sequence of an individual human. *PLoS Biol* **5**, e254 (2007).
197. Metzker, M.L. Sequencing technologies - the next generation. *Nat Rev Genet* **11**, 31-46 (2010).
198. Wheeler, D.A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L. et al. The complete genome of an individual by massively parallel DNA sequencing. *Nature* **452**, 872-6 (2008).
199. Raj, A., Rifkin, S.A., Andersen, E. & van Oudenaarden, A. Variability in gene expression underlies incomplete penetrance. *Nature* **463**, 913-8 (2010).
200. Kajiwara, K., Berson, E.L. & Dryja, T.P. Digenic retinitis pigmentosa due to mutations at the unlinked peripherin/RDS and ROM1 loci. *Science* **264**, 1604-8 (1994).
201. Rio Frio, T., Civic, N., Ransijn, A., Beckmann, J.S. & Rivolta, C. Two trans-acting eQTLs modulate the penetrance of PRPF31 mutations. *Hum Mol Genet* **17**, 3154-65 (2008).
202. Carr, I.M., Bhaskar, S., O'Sullivan, J., Aldahmesh, M.A., Shamseldin, H.E. et al. Autozygosity mapping with exome sequence data. *Hum Mutat* **34**, 50-6 (2013).
203. Koenekoop, R.K. The gene for Stargardt disease, ABCA4, is a major retinal gene: a mini-review. *Ophthalmic Genet* **24**, 75-80 (2003).
204. Sun, H., Smallwood, P.M. & Nathans, J. Biochemical defects in ABCR protein variants associated with human retinopathies. *Nat Genet* **26**, 242-6 (2000).

Appendices

I. Genome Wide Association Q-Q and MDS Plots	212
Figure A1. Golden Retrievers (PRA-1)	212
Figure A2. Golden Retrievers (PRA-2)	212
Figure A3. Gordon Setters (RCD4)	213
Figure A4. Tibetan Spaniels (PRA-3).....	213
II. Primers.....	214
Table A1. Canine Mutation Screening Primers	214
Table A2. Microsatellite Marker Primers	215
Table A3. Canine Candidate Gene Sequencing Primers.....	218
Table A4. Canine Candidate Gene mRNA Sequencing Primers	221
Table A5. Primers for qPCR of FAM161A	224
Table A6. Human SLC4A3 Sequencing Primers.....	225
III. PCR and Sequencing Reaction Components	227
Table A7. TaqMan [®] Allelic Discrimination	227
Table A8. Amplification for AFLP Analysis	227
Table A9. Amplification for Sequencing (HotStarTaq Plus).....	228
Table A10. Amplification for Sequencing (Amplitaq Gold 360)	228
Table A11. Amplification for Sequencing (Herculase II).....	228
Table A12. Amplification for Microsatellite Genotyping	229
Table A13. Sanger Sequencing (Canine DNA)	229
Table A14. Sanger Sequencing (Human DNA).....	229
Table A15. Quantitative PCR of <i>FAM161A</i> mRNA.....	229
IV. Thermal Cycling Parameters.....	230
Table A16. TaqMan [®] Allelic Discrimination	230
Table A17. Amplification for Sequencing and AFLP	230
Table A18. Touchdown PCR	230
Table A19. Microsatellite Genotyping.....	231
Table A20. Sanger Sequencing (Canine DNA)	231
Table A21. Sanger Sequencing (Human DNA).....	231
Table A22. Quantitative PCR	231
V. Genes found in PRA-Associated Critical Regions	232
Table A23. Genes Within GR_PRA1 Critical Region.....	232
Table A24. Genes Within GR_PRA2 Critical Regions One and Two	234
Table A25. Genes Within TS_PRA Critical Region.....	236
Table A26. Genes Within GoS_PRA Critical Region	238
VI. NGS Analysis Pipelines	240
VII. GR PRA-2 Questionnaire	242

I. Genome Wide Association Q-Q and MDS Plots

Figure A1. Golden Retrievers (PRA-1)

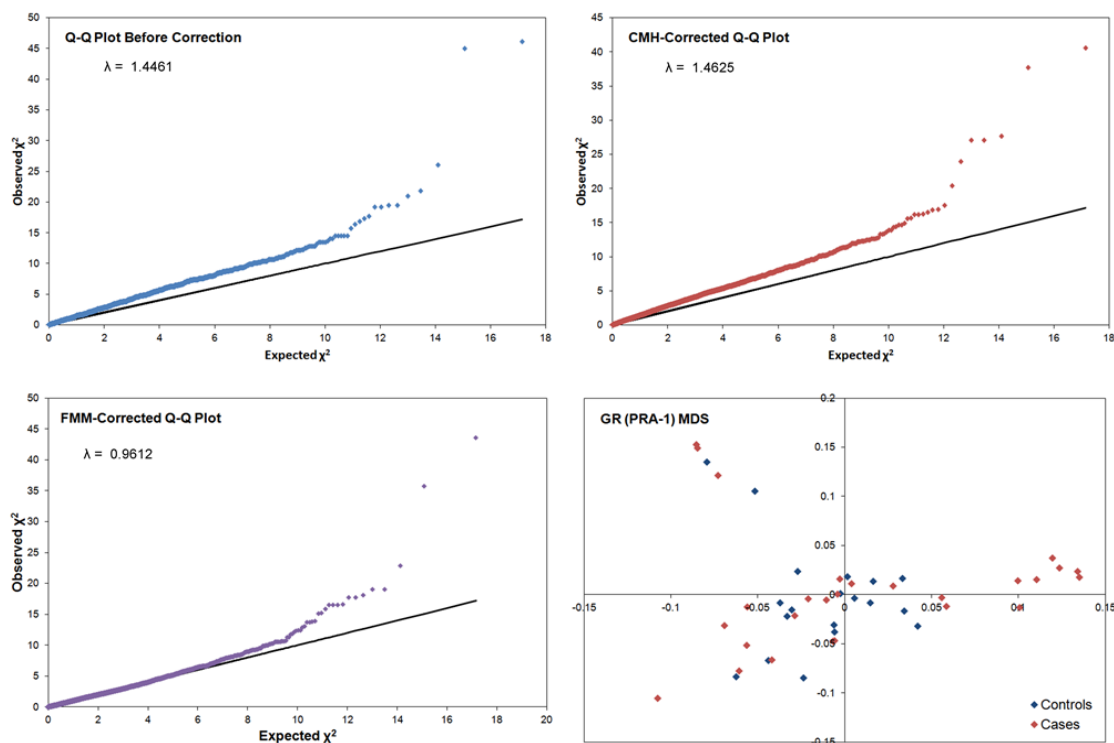


Figure A2. Golden Retrievers (PRA-2)

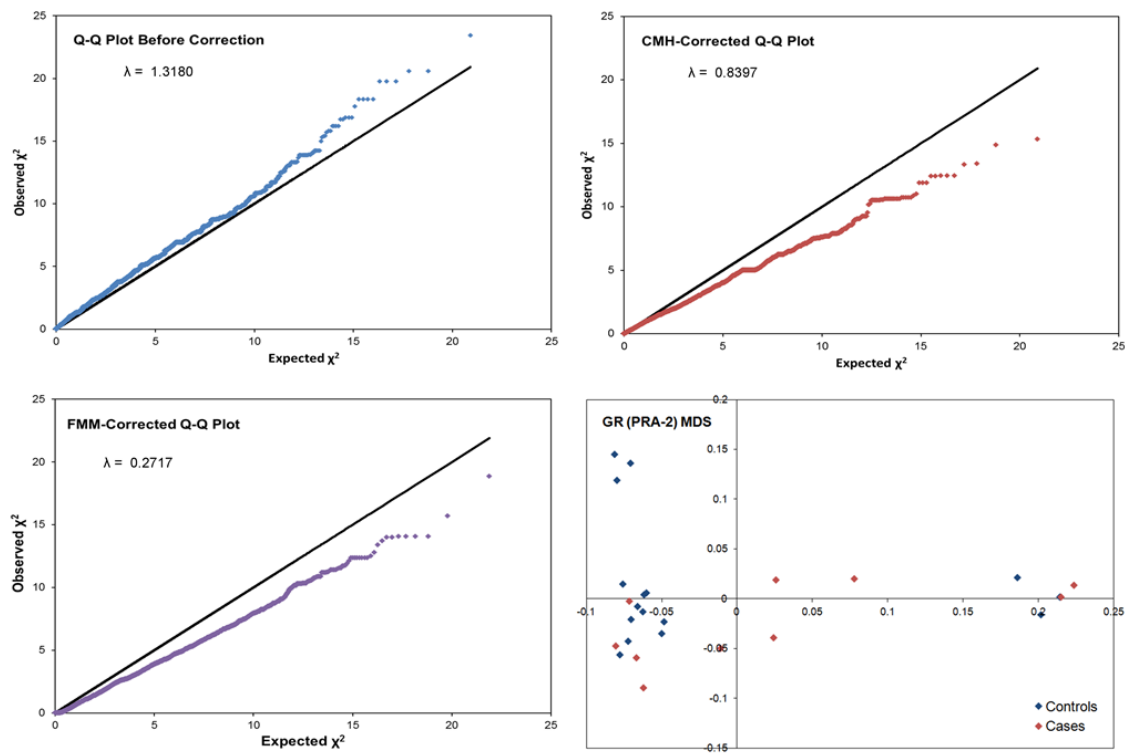


Figure A3. Gordon Setters (RCD4)

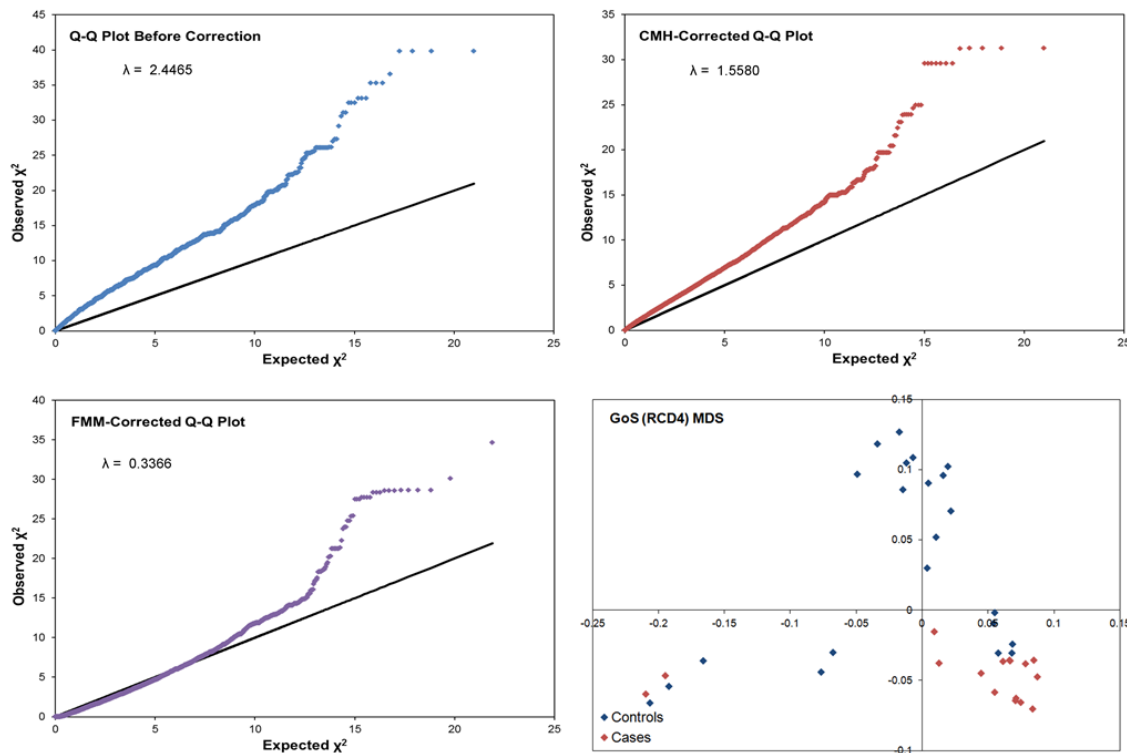
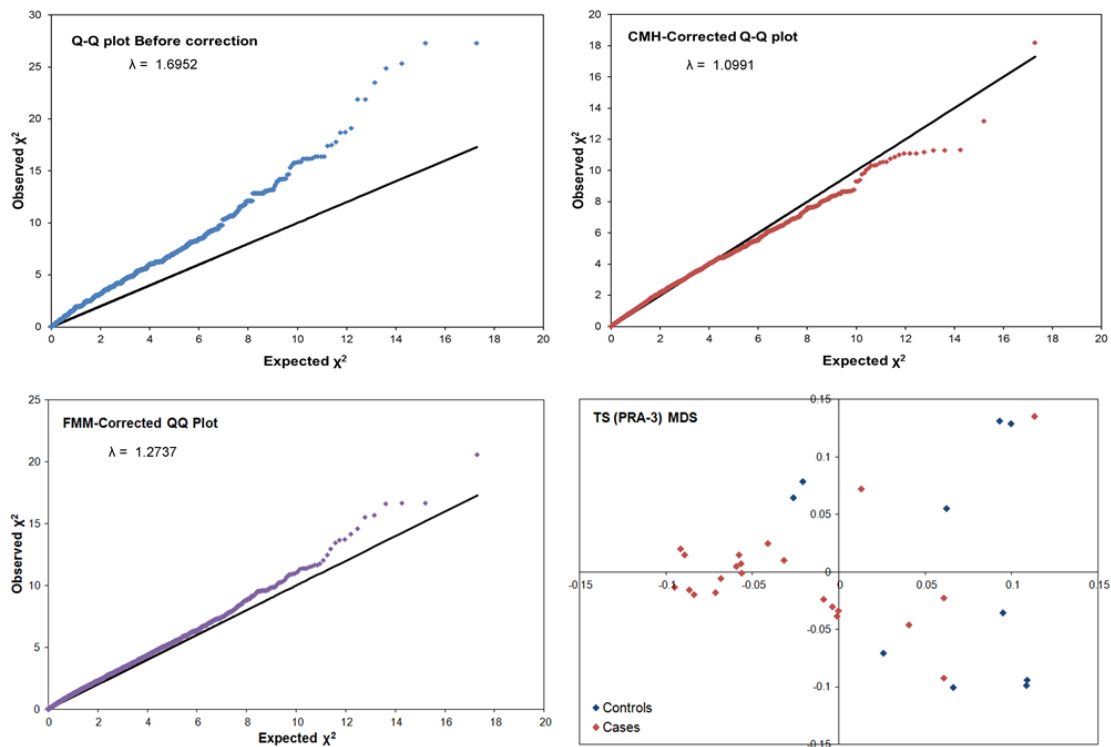


Figure A4. Tibetan Spaniels (PRA-3)



II. Primers

Table A1. Canine Mutation Screening Primers

Gene/Locus	Mutation	Method*	Primer Name	Primer Sequence	Amplicon Size (bp) Wildtype/Affected	T _m (°C)	Chr	Amplicon Position (bp) [†]
PRCD	G/A SNP	AD	PRCD_F	GGCCTTTCCTGCAGACT	134	57	9	7,186,629- 7,186,763
			PRCD_R	GGCCTTTCCTGCAGACT				
			PRCD_NORMAL	FAM-AGCCATGTGCACCACCCTCT-BHQ				
			PRCD_AFFECTED	HEX-TGAGCCATGTACACCACCCTCT-BHQ				
RCD1	G/A SNP	Sanger Sequencing	RCD1_F	CTTTCTTGGCTGTCGTCCTGTCCCT	157	60	3	94,574,186- 94,574,342
			RCD1_R	GAGTTTTCCCGTTTCCACGAAG				
PRA-1	C ins	AFLP	GR_PRA1_F	FAM-AGAGCAACCTTGTAACCCGTA	205/206	60	37	29,147,576- 29,147,781
			GR_PRA1_R	GGAAGAAGGCAATGAGAAAGG				
PRA-2	A del	AFLP	GR_PRA2_F	FAM-TGCCCTTCCACAGAGCAC	110/109	60	8	63,129,125- 63,129,234
			GR_PRA2_R	GCCATGTCTAAGCCCTTCACAA				
PRA-3v1	SINE ins	AFLP	TS_PRA_F	FAM-TGTGGTAAATTGTAAGCCATGC	153/391	57	10	64,973,910- 64,974,062
			TS_PRA_R	TGTTGGTATTCTCTCATCCTTTCC				
PRA-3v2	SINE ins	AFLP	TS_PRA_AFF_F	GGATCCCTTTATTTGATTTTAGAAAG	135/137	60.5	10	64,974,003- 64,974,139
			TS_PRA_NOR_F	TCCCTTCCTTTTATTTGATTTTAGAAAG				
			TS_PRA_R	FAM-CAACAAACACAACCTGAGCAA				
RCD4	C ins	AFLP; Q	RCD4_F	FAM-CCGAGTGCTCCCTCTGTG	146/147	57	17	26,010,390- 26,010,534
			RCD4_R	GGCTGCAGGCCTCGTC				

* AD = TaqMan[®] Allelic Discrimination; AFLP = Amplified Fragment Length Polymorphism analysis; Q=Additive Q solution used

[†] CanFam2.0

Table A2. Microsatellite Marker Primers

Microsatellite *	Position (bp) [†]	Tail-F primer Seq (5'-3') [‡]	R primer Seq (5'-3')	Size (bp)	T _m (°C)
Primers for Finemapping of TS_PRA locus on CFA10					
(CA)14	61,659,255	TAGACCCAGGCAATTCAAATC	TTGAGGATATCAAACCCTTCAC	193	56
(CA)21	61,752,065	ATGGAAGTCCATATGATCCA	GCATGTCCCTTGTATGGCTA	277	56
(GT)20	61,868,609	AGCTGGGTCCCTTTCAGTTA	CAGCCAGTACGTCTGATGCTT	236	56
(GT)15	61,963,812	CCTTTGTGCATTCTCTGTTCC	TCCCTAGGCCTACAGGATTT	234	56
(GT)18	62,337,915	TCTGCTCTTTAAGGCCTGATG	GAATGCACTTCAACCACCATT	212	56
(GT)15	62,434,740	TCTGTTCAAGAGCTGTGTGGAT	TTCTGTGGCACTGCTTCCTT	274	56
(CA)16	62,487,468	ACTGAGGGATGCAGAAGTGAA	ATATCTGAGCGCAGCATGAAG	174	56
(CA)20	62,552,700	TTTAAGAGCCAACCACTTCCA	TGCTTCTTCACCCAGATTGTT	229	56
(GT)15	62,786,640	GCCAAATTATGCAAGTAGACCA	GCATGATACCCTTTAGATTACCC	176	56
(CA)17	62,978,760	CCTGCATGTTCTCACAAAGGT	GTGGTGGTAGTTGGACAATGC	219	56
(CT)21	63,008,220	ACCTCACTGCAGGCTCACTT	TGAATGGAAGCTGTGTTTGTTT	247	56
(CA)15	63,199,200	ATCACCTCAAATCCACCACAA	TATCCCTTGGGCAGGTTTAC	241	56
(CA)17	63,331,680	ATGAAAGCAAAGCAGGGAGAT	ATTGAAGGCTGAGGGAGCTAA	209	56
(CA)21	63,402,960	TAGCACCAGAACCTTCCACTC	GCATCACAACACTCACTCTGC	205	56
(CA)18	63,509,760	GGAGTTTACCATAAGTAAGCTTCTCC	TGCTTCTGACCATTCCACAAT	228	56
(GT)16	63,705,840	TCCTCTCCTCCAAACACTTACC	ATGCATACAGCACTTGCACAC	184	56
(GT)17	63,820,800	CGCAGCCTTTAAATCATCCTT	GCAGCTCACAATGAACCATTT	249	56
(GT)22	63,859,200	TTCCATTCTGAAGGCCCTAAT	AAGGTGTGCTGAAAGGAAAGC	244	56
(CT)17	63,913,440	CTCCACCACTTGCTCTCTCAT	GAACCCTATGAAGCAACAGCA	205	56
(GA)16	63,934,627	CCCTACCAAGTTCTACCCTTT	TTCAGCATGACAAGGGACAC	242	56
(GT)19	63,998,363	GCAGCACTATCAATAGCCAAA	TCTACTTCATCCATGTCCTTGC	177	58
(CA)19	64,075,600	CCCTGTATTGATTAGAGCACGA	GATCACCTTGGTTTCTGGTCA	235	56

Microsatellite*	Position (bp) [†]	Tail-F primer Seq (5'-3') [‡]	R primer Seq (5'-3')	Size (bp)	T _m (°C)
(AT)17	64,112,403	GCATCTGCAATCAATTTGAGG	CTCCCTCTCCATTTGTCCTCT	265	58
(GA)14	64,113,386	CATCAGTTCATCCCGTAGCAT	GGCACAGTAGGAGATTAACAGCA	194	56
(GT)14	64,186,104	ACTGCACCCATGTCTTGTTCT	GGTGTTCAGTATGTGGCTATGTTC	237	56
(GA)16	64,201,234	AAGTTCATGCCAATATCCTAGAGTC	TTGTTGTTGCCTATTGGGTTC	366	56
(CA)16	64,217,353	GCAGTTGAATAACCTGCTTGC	TATCCTTTCTTCCTGCCACCT	246	56
(TG)14	64,508,603	AGACCAGCTTTTTGCTTCAGA	CTGCGATTAATGGTCAAAGG	267	57
(AC)11	64,572,573	CATGTTGTCATGGAATGCAAG	AGTGAACCCCGACAGGTTAGT	309	57
(TG)12	64,713,419	CCCCTAAGCCTGATCCAATAA	TGAAAGGAAATGAGTAGTTTTGCT	275	57
(AC)18	65,182,153	CAAGGGTCTAATTTCCCTTTCC	TGGGTACCAAGTGAATTTTG	127	57
(AG)16	65,447,202	TGAACTCCTCAGACCTGCAAT	CTCTGCTCCAGTCAGATTTGG	304	57
(TG)18	65,529,249	AGCAGTTTTGAAGCATCATGG	AAAAGTAGCCCAAGATCAACCA	120	57
(TA)20	65,625,939	CAGATCTTGAGAGGGAAAGAGAA	CATAGTTTCTTGCCAAGTACTGTCT	220	56
(TC)14	65,784,034	G TTCACCTCCACTCCCATTTA	GGCACAGAACAAAACATCCAT	314	57
(TG)11(TA)10	65,931,460	CCTGTTGTGTTTCAGCAAGGAT	AGTTATCGCTGCCTTCTCCTC	358	57
(GT)12	66,013,662	TGCAAAAGTGAGAGACAGATCC	TGCTGTGTTATCTGCAGCTGTT	322	57
(AT)20	66,058,427	CCGTATTCCTCCCTGTAGTCC	TCCAGATTTACAGCATGATTC	317	57
(AC)15	66,162,019	CTCTGTGTTTATTGGCGCATT	GGCGAGACCTCATTCTTTTCT	183	56
(GT)14	66,303,329	GCAGACAGCCTTTTGTTTGAC	AGTTCTACTTTGCCCGGCTAC	202	60
(CA)11	66,379,163	GGGAAGAAAGAACTCTTGAA	TCCCTTCTTTCCATTACCTCTG	269	57
(TA)14	66,395,178	CCCCACACATCAACATAGGTC	GCTATGCAGGCATTTATTTTGG	296	57
(TA)18	66,448,543	GTGAAATGTATTCCGCATGGT	ATGTCCTATGGCCTTTTGTT	242	57
(TC)14	66,509,954	GGAGAAGAGGAAGTTTGCAG	GCCAAAGTCTCAAAGGGATCT	314	56
(AG)12	66,557,299	CACTGGCACCAGGTTAAATGT	CCTATGGGGCTCTGGATAAAT	348	57
(AC)20	66,611,818	CTCCTACTGGGAGCCAAAAC	CCACATCCTTCCTACCACTGA	212	57

Microsatellite*	Position (bp) [†]	Tail-F primer Seq (5'-3') [‡]	R primer Seq (5'-3')	Size (bp)	T _m (°C)
(TC)21(TG)20	66,684,676	GCTGCATTGCTGCTTATTTTC	AGGTTTAAGGCAGCACCAATC	296	57
(TG)16	66,710,025	CCACCCTCATGTTTCATCTGTT	CACCATCACTACCAATAGGG	301	57
(CT)15	66,985,774	TTAGGATTGTGGGGTTCAGTG	TCAGGTGAAAGGAAGAAGCAA	304	57
(AC)15	67,056,476	GAAAGTGGAGGGGAAAGTGAC	AACCGGGTTGAACCATTTTT	236	57
(CA)21	67,164,626	CTGCACTGTTTCCTTCCTCTG	ATCTGGGGTCTTTTTGCTGTT	270	57
(GT)13	67,385,860	CCAGGTTTGGGGTGTATTTCT	CCAGTGGTCACCTCAACAAGT	244	57
(TC)19	67,482,874	TGAACCTCCGAAGTCACTTGT	GGAAAAGTGTAAGGGGACCA	258	57
(GA)13	67,729,134	TGCTTCTGAGGCACGAGTATT	TGCTTGCGTTGCTATTAAGTTC	302	57
(GT)10	67,940,561	GGTTCATGCCCTCTAATAGGC	ATGCCCTCACACTGAAATCAC	237	57
(AC)20	68,043,943	CTCCGGGGACTCTCAAATTAC	TCTGACTTGGGAAAGTGGTTG	263	57
(GT)32	68,110,195	TCTAGGCTGGGCTTTGAGAAT	TGTCAATTGATCCCACTCCAT	277	57
(GT)17	68,422,023	TTTCCTTGCTCCATTGATGTC	CAGAATGTGGGTGCTTTTCAT	314	57
(TC)24	68,544,337	GTGCACCATCTGGACTACCAT	ACAGTGCATTACAGGGACCAG	199	57
(CA)15	68,607,662	AAAACCTTTCTCCTCCCCTTC	ATGTAGGCCCACTTTTTGCTT	184	57
Primers for Analysis of GoS_PRA locus on CFA17					
(AG)12	25,989,874	GGGGGTGCTAGGATAATTTTT	GGAAAGGGGGATTTTTAACCT	297	57
(GT)17	26,009,511	GGGATCAGGAAGTTTGAGCAT	GCCTTCATCTGTCCCTCATT	276	57

* The number is the number of times the dinucleotide in brackets is repeated

[†] Location of the F primer according to CanFam2.0

[‡] Tail= 5'-TGACCGGCAGCAAAATTG-3'

Table A3. Canine Candidate Gene Sequencing Primers

F primer Name	F primer Seq (5'-3')	R primer Name	R primer Seq (5'-3')	Size (bp)	Tm (°C)	Additive/alteration*	Position (bp) [†]
Primers for PCR amplification and Sanger sequencing of <i>SLC4A3</i> (CFA37)							
SLC4A3ex2_F	GTCCGGGTGCCCTTGGAG	SLC4A3ex2_R	ACTCAAAGTCACGCTCGCT	166	56	98, Q	29,137,454
SLC4A3ex3.1_F	GGAAGGACTAGGGTGACCTTG	SLC4A3ex3.1_R	AGATTCCCCTTCTCTTCCTC	347	58	ATG360, 98	29,137,955
SLC4A3ex3.2_F	TGAGGCCCTGAGCTAGAATAA	SLC4A3ex3.2_R	TTCAAGGGCACACCCCTAC	394	58	ATG360, 98	29,137,978
SLC4A3ex4_F	GAAAGCAAAGGTAGGGGTGTG	SLC4A3ex4_R	CTCTCCTCCCCTTGACAGTCT	327	54	TD, Q, 98	29,138,343
SLC4A3ex5.1_F	GTGGTTCCCATCAAGTCC	SLC4A3ex5.1_R	GCTCCCGCAGGTCGTAG	132	54	98	29,140,268
SLC4A3ex5.2_F	CGAGTCGCTGGAGAGAAGAG	SLC4A3ex5.2_R	GAGAGCACAGAGCTTCTCAA	257	58	98, Q	29,140,339
SLC4A3ex1-C_F	GAGGGCGTGTTTACAGACTCA	SLC4A3ex1-C_R	CACCACCTCCATCGAGTCAC	566	56	H7	29,141,220
SLC4A3ex6_F	GTTGTGTAGACGTGGCCTCTG	SLC4A3ex6_R	AGCCTCTGGGGAGGACAG	312	58	TD, Q, 98	29,141,911
SLC4A3ex8.1_F	CTCCTCCGGCCTGTTTG	SLC4A3ex8.1_R	AGACACGATCATGGTCTCCAC	168	56	98	29,142,901
SLC4A3ex8.2_F	TCCTGGACCTGGAGCAGAC	SLC4A3ex8.2_R	ATTCAGGCTGAGGCTAGAAGG	209	58	98, Q	29,143,001
SLC4A3ex9-11_F	CTGCCTTACTTGACAGCTGACT	SLC4A3ex9-11_R	CCGTCGAGTCTCATGTCTCAT	1831	58		29,143,418
SLC4A3ex12-13_F	GGAAGCTCCTGGAGGAATAGA	SLC4A3ex12-13_R	CTAAGGTCCCATGACCACAGA	897	58		29,145,430
SLC4A3ex14-15_F	GTCTGTGGTCATGGGACCTTA	SLC4A3ex14-15_R	GGTTGCTCTAGCACTCACCTG	1280	58		29,146,305
SLC4A3ex16-17_F	GCAGAGGGTACAAGGAAGAGG	SLC4A3ex16-17_R	TGAAGGTGAACAGGTTTGGAG	1270	58		29,147,228
SLC4A3ex18_F	GTTTCTGGAACGCTGGTACT	SLC4A3ex18_R	GTGAGCATCACTGGGTGAGAG	487	58		29,148,499
SLC4A3ex19-21_F	TGAGAACACCACTTTGGGAAC	SLC4A3ex19-21_R	ACTTTTCAAACCCCTCCTCCA	1519	58		29,149,094
SLC4A3ex22_F	GGCGTTGCTGTCACTGTTACT	SLC4A3ex22_R	TAATGAGCTCAAAGCCTGTGC	466	58		29,150,683
Internal primers for Sanger sequencing of <i>SLC4A3</i> (CFA37)							
SLC4A3ex9_IntF	CACCTCTCTCCCTAACCTGC	SLC4A3ex9_IntR	ATCATCCCATCCACACCTTC				
SLC4A3ex10_IntF	GAAGGTGTGGATGGGATGAT	SLC4A3ex10_IntR	GGAGGTGAAAAGTTAACAGTCTCC				
SLC4A3ex11_IntF	GGAGACTGTAACTTTTACCTCC	SLC4A3ex12_IntR	GGCTCAGTCCCTCCCATG				

F primer Name	F primer Seq (5'-3')	R primer Name	R primer Seq (5'-3')	Size (bp)	Tm (°C)	Additive/alteration*	Position (bp) [†]
SLC4A3ex13_IntF	CATGGGAGGGACTGAGCC	SLC4A3ex14_IntR	ACAGGGGTGAGCTCCACAG				
SLC4A3ex14_IntF	GAGTTGGGTCCCCAGGTC	SLC4A3ex15_IntR	CTTGTACCCTCTGCTCCCAAG				
SLC4A3ex15_IntF	CTGTGGAGCTCACCCCTGT	SLC4A3ex16_IntR	TGAGCTGGTCCCTCGAAATC				
SLC4A3ex16_IntF	GTTGCCTGGTCAGGTACGTG	SLC4A3ex19_IntR	CAGACTCAGAAGGGACCTGAG				
SLC4A3ex17_IntF	GATTCGAGGGACCAGCTCA	SLC4A3ex20_IntR	GGGTCACTCTGTCCGTCTC				
SLC4A3ex20_IntF	CTCAGGTCCTTCTGAGTCTG						
SLC4A3ex21_IntF	GAGACGGACAGAGTGACCC						
Primers for PCR amplification and Sanger sequencing of <i>TTC8</i> (CFA8)							
TTC8_ex8_F	GAGGTGCTTTGAATCATTGGA	TTC8_ex8_R	TGAGGAAGTTTCACCCAGAAA	532	57		63,128,964
Primers for PCR amplification and Sanger sequencing of <i>FAM161A</i> (CFA10)							
FAM161A_ex5_F	TTCATCTCTTCAGGCCAGTGT	FAM161_ex5_R	ACTGCAGTAGCAGCGTGTCT	720	57		64,973,585
FAM161A_ex1_F	CCTGTAGCGTCCCTGGTTAC	FAM161A_ex1_R	AAATGGAGGCTCTCCAAGAAA	570	57	98, Q	64,991,088
Primers for PCR amplification and Sanger sequencing of <i>C2ORF71</i> (CFA17)							
C2OR71_exProm_F	AGGGACCAGCCTTGAGTTTTA	C2OR71_exProm_R	CACCCACACTCTTTGCAATTT	854	58		26,013,661
C2OR71_ex1.1_F	TCCACAAAAGCTCTTCCAGA	C2OR71_ex1.1_R	CCACACCCTCTTTGTACTCA	1483	56	35	26,012,688
C2OR71_ex1.2_F	AAAGAAAGCTGAGATGCCACA	C2OR71_ex1.2_R	CCTCCACGTGACTCTTCAGAC	1374	58		26,011,830
C2OR71_ex1.3_F	AGCACAGGCCAGAAAATACT	C2OR71_ex1.3_R	TCCAAGGTGAGCTTTCTGGTA	1475	58		26,010,857
C2OR71_ex1.4_F	CTCTGGAAATCCTGATGGACA	C2OR71_ex1.4_R	CCTGAAATGAGGGACAGATGA	1363	56	Q	26,009,761
C2OR71_ex2.1_F	GGGTGAGGACCTGAGAGAATC	C2OR71_ex2.1_R	GGAATAAAGTTGGCCGAGAAG	1329	58		26,004,158
C2OR71_ex2.2_F	ATTTGAACAGGTGTGCCTGAC	C2OR71_ex2.2_R	TTTCTCCAGACCTTGGGAAGT	1477	58		26,003,225
C2OR71_ex2.3_F	CACCCCTCTATTGGGAAATGT	C2OR71_ex2.3_R	GTTGATGATTGCTGGACACCT	1979	58		26,001,698
Internal primers for Sanger sequencing of <i>C2ORF71</i> (CFA17)							
C2OR71_exProm_IntF	GCAATTATGGCTCATTTCTC	C2OR71_exProm_IntR	GAGAAATGAGCCATAATTGC				
C2OR71_ex1.1_IntF1	CTGGACCAACTAGTAAGGCTG	C2OR71_ex1.1_IntR1	CAGCCTTACTAGTTGGTCCAG				

F primer Name	F primer Seq (5'-3')	R primer Name	R primer Seq (5'-3')	Size (bp)	Tm (°C)	Additive/alteration*	Position (bp)[†]
C2OR71_ex1.1_IntF2	GAAAGCTGAGATGCCACACAT	C2OR71_ex1.1_IntR2	ATGTGTGGCATCTCAGCTTTC				
C2OR71_ex1.2_IntF1	CTACCGCCTGCCACTTGGGAA	C2OR71_ex1.2_IntR1	TTCCCAAGTGGCAGGCGGTAG				
C2OR71_ex1.2_IntF2	ATACTACTGCCAGGCCTGGAG	C2OR71_ex1.2_IntR2	CTCCAGGCCTGGCAGTAGTAT				
C2OR71_ex1.3_IntF1	CTGTGAGGCAGAGAAGGTCCC	C2OR71_ex1.3_IntR1	GGGACCTTCTCTGCCTCACAG				
C2OR71_ex1.3_IntF2	CAAGCCCCTGCCTCAGGAAGT	C2OR71_ex1.3_IntR2	ACTTCCTGAGGCAGGGGCTTG				
C2OR71_ex1.4_IntF1	TGCAGACAGGGCCACCAGCCT	C2OR71_ex1.4_IntR1	AGGCTGGTGGCCCTGTCTGCA				
C2OR71_ex1.4_IntF2	GTTTGAAGCCGCCACCCGG	C2OR71_ex1.4_IntR2	CCGGGTGGGGCGGCTTCAAAC				
C2OR71_ex2.1_IntF1	CAGCCCCAGATTCTGGTTTGC	C2OR71_ex2.1_IntR1	GCAAACCAGAATCTGGGGCTG				
C2OR71_ex2.1_IntF2	GTTGTACAAATGGTTGGAAAG	C2OR71_ex2.1_IntR2	CTTTCCAACCATTTGTACAAC				
C2OR71_ex2.2_IntF1	CGGGAAGATGCACGCACTCAC	C2OR71_ex2.2_IntR1	GTGAGTGCCTGCATCTTCCCG				
C2OR71_ex2.2_IntF2	CCACCAACAGGGGCCACAAGC	C2OR71_ex2.2_IntR2	GCTTGTGGCCCCTGTTGGTGG				
C2OR71_ex2.3_IntF1	AGGGACGCCATTCATGAAATG	C2OR71_ex2.3_IntR1	CATTTTCATGAATGGCGTCCCT				
C2OR71_ex2.3_IntF2	GCCAGACTCCTGCTATATACC	C2OR71_ex2.3_IntR2	GGTATATAGCAGGAGTCTGGC				
C2OR71_ex2.3_IntF3	TCCACTTCTTGAGCTCAAAGC	C2OR71_ex2.3_IntR3	GCTTTGAGCTCAAGAAGTGGA				

* 98 = Initial denaturation at 98°C; Q = Q solution; ATG360 = AmplitaqGold 360 DNA Polymerase, 2mM Mg, 2x GC enhancer; TD = Touchdown PCR, annealing temperature decreasing from 68°C to 54°C; H7 = Herculase Taq polymerase, with 7% DMSO; 35 = 35 amplification cycles.

† According to CanFam2.0

Table A4. Canine Candidate Gene mRNA Sequencing Primers

F primer Name	F primer Seq (5'-3')	R primer Name	R primer Seq (5'-3')	Tm (°C)	Size (bp)	Additive/alteration*
Primers for PCR amplification and Sanger sequencing of canine <i>SLC4A3</i> mRNA						
SLC4A3exn1-2_RT-F	CATGGCCAACGGAGTGAT	SLC4A3exn1-2_RT-R	GGATGTGTGCCGGTGAA	58	235	Q
SLC4A3exn3_RT-F	AGCTACAGCGAGCGTGACTTT	SLC4A3exn3_RT-R	ATCGCTGCCGATGGAGAA	58	324	Q
SLC4A3exn4-5_RT-F	GAGGAAGAGGAAGGGGAATCT	SLC4A3exn4-5_RT-R	CGAGACGGCTTCTTCACTAGG	58	554	
SLC4A3exn6-8_RT-F	CGAGTCGCTGGAGAGAAGAG	SLC4A3exn6-8_RT-R	GTCCTTGTTCATCATTGGGATG	58	654	
SLC4A3exn9-11_RT-F	CTGCGTACCCTCCTACTGAAA	SLC4A3exn9-11_RT-R	GATGGAGGAATCACGATGCT	58	590	
SLC4A3exn12-13_RT-F	GTTCGCTTCCTCTTTGTGATG	SLC4A3exn12-13_RT-R	ATCAGCCCTTCCGTCTTCTC	58	560	
SLC4A3exn14-15_RT-F	CCGTGCTTTCATCTACTTCG	SLC4A3exn14-15_RT-R	CAGGAAGAAGGCAATGAGAAA	58	587	
SLC4A3exn16_RT-F	GCCTTCCTCATCTCGCTTATT	SLC4A3exn16_RT-R	GGAGTAATCCACGAGGACCAT	58	343	
SLC4A3exn17-18_RT-F	CCTTTCTCATTGCCTTCTTCC	SLC4A3exn17-18_RT-R	GCATAACAGTCAGTGC GTTCA	58	483	
SLC4A3exn19-20_RT-F	TCCTCATCCTCATCTTCATGG	SLC4A3exn19-20_RT-R	ACTTGACCACCCAGAGCAGT	58	534	
SLC4A3exn21-22_RT-F	CAGCGTCTGTTGCTTATCCTC	SLC4A3exn21-22_RT-R	TAAGAATGACCAGACCGACCA	58	581	
SLC4A3exn1_C_RT-F	GCTCTGGTTGAACAAAGACCA	SLC4A3exn1_C_RT-R	CGAGACGGCTTCTTCACTAGG	58	376	Q
Primers for PCR amplification and Sanger sequencing of canine <i>SPATA7</i> mRNA						
SPATA7_exon1-3_F	CGGGAAGGTCGGAAGG	SPATA7_exon1-3_R	ATTTGATGCTGGCACTTATGC		298	
SPATA7_exon4-5_F	CCACCTGGTCAAGAATCACAT	SPATA7_exon4-5_R	GGATGGAAATCCATTCACTTCT		298	
SPATA7_exon6_F	ACGATGTGAAAGGGAGTTGAA	SPATA7_exon6_R	TTCTTGGTGTCAACTGTCTCAA		600	
SPATA7_exon7-8_F	ATGGAAGCTGAAACCCAGACT	SPATA7_exon7-8_R	AAGCTGCAGGGAATACTGACA		243	
SPATA7_exon9-11_F	ACCATGGGATGAGATCAAAGA	SPATA7_exon9-11_R	CTTCAGGATATGCAGCAGGTG		271	
SPATA7_exon12_F	GAACGACTGTTTGAGCGACAT	SPATA7_exon12_R	TGGCCATTTAGGATATTTATTGATG		665	
Primers for PCR amplification and Sanger sequencing of canine <i>TTC8</i> mRNA						
TTC8_mRNA_F	ATGGGCTCGGAGATGGA	TTC8_mRNA_R	TCTGTGGGACAATCAGAGCA	57	1560	
		TTC8_mRNA_R2	GTGGAACTTGAGCAATAGCATT	57	265	

F primer Name	F primer Seq (5'-3')	R primer Name	R primer Seq (5'-3')	Tm (°C)	Size (bp)	Additive/ alteration*
Internal Primers for Sanger sequencing of canine <i>TTC8</i> mRNA						
TTC8_mRNA_IF1	GAACTGCCTACACAGCTCG	TTC8_mRNA_IR1	CGAGCTGTGTAGGCAGTTC			
TTC8_mRNA_IF2	GTTTCCAGGAGAAGTAACCC	TTC8_mRNA_IR2	GGGTTACTTCTCCTGGAAAC			
TTC8_mRNA_IF3	CAACTTGGGCCACATAGCTG	TTC8_mRNA_IR3	CAGCTATGTGGCCCAAGTTG			
TTC8_mRNA_IF4	CTTACGACCAGGAACCAGATC	TTC8_mRNA_IR4	GATCTGGTTCCTGGTCGTAAG			
TTC8_mRNA_IF5	GTCACCTCTGGAACCAGATC	TTC8_mRNA_IR5	GATCTGGTTCAGAGGTGAC			
TTC8_mRNA_IF6	GTCTCAGGCAGCTTGGATT	TTC8_mRNA_IR6	AATCCAAGCTGCCTGAGAC			
TTC8_mRNA_IF7	GTCCCCTTACGACCAGGTATTG	TTC8_mRNA_IR7	CAATACCTGGTCGTAAGGGGAC			
TTC8_mRNA_IF8	CACCTCTGGCAGCTTGG	TTC8_mRNA_IR8	CCAAGCTGCCAGAGGTG			
TTC8_mRNA_IF9	CCAGGCAGCTTGGATTTTA	TTC8_mRNA_IR9	TAAAATCCAAGCTGCCTGG			
Primers for PCR amplification and Sanger sequencing of canine <i>FAM161A</i> mRNA						
FAM161A_cDNA1_F	CATGGCAGCCTCGCAG	FAM161A_cDNA1_R	AAGTTGGGCAACTCATCTTCA	57	540	Q, 98
FAM161A_cDNA2_F	CCACTTGAAACTATGGCAA	FAM161A_cDNA2_R	TTTGCACCTGCACTTCTCCT	57	1203	
FAM161A_cDNA3_F	TTTTCCCCCTTATCATGACC	FAM161A_cDNA3_R	CCAGGCTACAGATGACTTTGC	57	1423	
FAM161A_cDNA4_F	ACACGGTCTTTTCTGTCTCCA	FAM161A_cDNA4_R	GCTTTTCTGCTGCCATTCTT	57	421	X40
Primers for PCR amplification and Sanger sequencing of canine <i>C2ORF71</i> mRNA						
C2OR71_amp1_F	CTGGACCAACTAGTAAGGCTG	C2OR71_amp1_R	CCACACCCCTCTTTGTACTCA	58	1048	
C2OR71_amp2_F	AAAGAAAGCTGAGATGCCACA	C2OR71_amp2_R	CCTCCACGTGACTCTTCAGAC	58	1374	
C2OR71_amp3_F	AGCACAGGCCAGAAAATACT	C2OR71_amp3_R	TCCAAGGTGAGCTTTCTGGTA	58	1475	Q
C2OR71_amp4_F	GACCTGCTACCCAGCAAGAG	C2OR71_amp4_R	CTTGACAGAGGGAGCACT	56	445	Q, 98
C2OR71_amp5_F	CACAAGAGGCTCCTGGACAC	C2OR71_amp5_R	GAGGGGGAAGTGGCCTCT	56	243	Q, 98
C2OR71_amp6_F	GAGTGCTCCCTCTGTGCAA	C2OR71_amp6_R	AAGCGGTCCTTCTGATGAAA	56	479	98
C2OR71_amp7_F	CTGTGCGGCTCAACC	C2OR71_amp7_R	AGTCTGATGCTTGGCTGTCTC	56	381	Q, 98
C2OR71_amp8_F	AGGCTAATCAACAGGGTGACA	C2OR71_amp8_R	CTGTGAGACCTCAGGCAAGAC	57	462	

F primer Name	F primer Seq (5'-3')	R primer Name	R primer Seq (5'-3')	Tm (°C)	Size (bp)	Additive/alteration*
C2OR71_amp9_F	CATGTTTCCTGGTTTTGCAGAT	C2OR71_amp9_R	GAGGGCTTTGGGTCTGTACTT	57	379	
C2OR71_amp10_F	CAGCATTCAAGTCACCATTGA	C2OR71_amp10_R	GGAATAAAGTTGGCCGAGAAG	57	393	
C2OR71_amp11_F	TGCAAACAAGCAGCCTATTCT	C2OR71_amp11_R	CTTGTTCTCGAGCCATGATCT	57	490	
C2OR71_amp12_F	CTGCTGCAAAGCTGATTTTCT	C2OR71_amp12_R	TTCCCACACACTTACCTCAGC	57	454	
C2OR71_amp13_F	CACCCCTCTATTGGGAAATGT	C2OR71_amp13_R	TTTCTCCAGACCTTGGGAAGT	57	454	
C2OR71_amp14_F	CAAGCAGGAGAAGAAATGCTG	C2OR71_amp14_R	GGTGTCCACACCTAGCTTTCA	57	364	
C2OR71_amp15_F	TCTGGGAACCAGGATAGGAAT	C2OR71_amp15_R	ACGACATCATGTATGGGGGTA	57	469	
C2OR71_amp16_F	CTGCTCCCTCCACATCTACTG	C2OR71_amp16_R	GTTCTCAAGGAGGGGTGAAAC	57	402	
C2OR71_amp17_F	TTTTTGGCTACCATCACAACC	C2OR71_amp17_R	GCGTTGGGCCTATTGGTATAA	57	371	
Internal primers for Sanger sequencing of canine <i>C2ORF71</i> mRNA						
C2OR71_amp1_IntF	GAAAGCTGAGATGCCACACAT	C2OR71_amp1_IntR	ATGTGTGGCATCTCAGCTTTC			
C2OR71_amp2_IntF	CTACCGCCTGCCACTTGGGAA	C2OR71_amp2_IntR	TTCCAAGTGGCAGGCGGTAG			
C2OR71_amp3_IntF	CAAGCCCCTGCCTCAGGAAGT	C2OR71_amp3_IntR1	GGGACCTTCTCTGCCTCACAG			
		C2OR71_amp3_IntR2	ACTTCCTGAGGCAGGGGCTTG			

* Q = Q solution; 98 = Initial denaturation at 98°C; x40 = 40 PCR cycles

Table A5. Primers for qPCR of FAM161A

F primer Name	F primer Seq (5'-3')	R primer Name	R primer Seq (5'-3')	T_m (°C)	Size (bp)
FAM161A _{total} _F	GTGTCTCTCCAGGTCTGAATCA	FAM161A _{total} _R	AAGTTGGGCAACTCATCTTCA	57	130
FAM161A _{sh-5} _F	ACACGGTCTTTTCTGTCTCCA	FAM161A _{sh-5} _R	TTGCATTTTTCTTGTGGCTTG	57	135
FAM161A _{sh} _F	ACACGGTCTTTTCTGTCTCCA	FAM161A _{sh} _R	TTCACTCTTTCTTGTGGCTTG	57	135
FAM161A _{fl-5} _F	CGGATCCTAACTAAGCAGAAGC	FAM161A _{fl-5} _R	CTTGCATTTTTCTGAGAGATTTG	57	136
FAM161A _{fl} _F	CAGAGTCAAATCTCTCAGAAAGAGTG	FAM161A _{fl} _R	CTGCTGCCATTCTTGCATT	57	140
TBP_F	CATGACTCCTGGAATCCCTATC	TBP_R	CTCCAGAAGAGAAAGGCTGTTG	57	100
ACTP_F	CCAACCGTGAGAAGATGACC	ACTP_R	AGAGGCGTACAGGGACAGC	57	95

Table A6. Human SLC4A3 Sequencing Primers

F primer Name	F primer Seq (5'-3')	R primer Name	R primer Seq (5'-3')	Size (bp)	Tm (°C)	Additive/ alteration*	Position (bp) [†]
Primers for PCR amplification and Sanger sequencing of human <i>SLC4A3</i> (Chromosome 2)							
SLC4A3_ex_1.1F	AGGCAAGGCTGTGGTAGAAAT	SLC4A3_ex_1.1R	TCTCGCAGCGCAGAGC	444	58	Q, 98	220,491,773
SLC4A3_ex_1.2F	CCCGTCCCATTGTGC	SLC4A3_ex_1.2R	CCGAGACCCAGGTGAGAAG	443	58	Q, 98	220,491,773
SLC4A3_ex_2F	GGGATTGTGGGAATATCG	SLC4A3_ex_2R	GGGCTACGGATCTGGAAAATA	849	58	Q, 98	220,492,089
SLC4A3_ex_3F	CAGATCCGTAGCCCTCTCTCT	SLC4A3_ex_3R	CACAGAGTCCAGCACAGTGTC	498	58		220,492,924
SLC4A3_ex_4F	GACTAGGGTGCCTTGTTTG	SLC4A3_ex_4R	CAACACGCACTCCTCTCCTC	464	58		220,493,757
SLC4A3_ex_5F	AGGGGTACAGAGAGGAGGAGA	SLC4A3_ex_5R	GTCACTCTCTCCTGGGGAAC	484	58		220,494,187
SLC4A3_ex_6F	GAGAGGGTGAGGAGAAAGGTG	SLC4A3_ex_6R	AGAGCCTCTCCAAGGTCCTC	475	58		220,494,677
SLC4A3_ex_6Fb	ACCCTGAGAAAAGAGGAGCAG	SLC4A3_ex_6Rb	AGAGGACTCAGCCCTTAGCAG	660	58		220,494,523
SLC4A3_ex_7F	ATGGCCTGTTGGTACTGTAG	SLC4A3_ex_7R	G TTCAGCTCCACGAACACCT	430	58		220,496,572
SLC4A3_ex_8F	ATCCTTCGCAGGAAGAAGAAG	SLC4A3_ex_8R	GAGAGTGTTCCTGGGAACTGA	457	58		220,496,788
SLC4A3_ex_9F	GTCAGTCAACAAATGGCTTCC	SLC4A3_ex_9R	CTCTGTCACTTGTGGCATTCA	489	58		220,497,353
SLC4A3_ex_10F	TGAATGCCACAAGTGACAGAG	SLC4A3_ex_10R	CAACCCACTCAGTGAAGTCGT	447	58		220,497,821
SLC4A3_ex_11F	TGCAGTACCTGCCTACTCTGG	SLC4A3_ex_11R	CAGGGGACTACGAAAGAAAGG	402	58		220,498,859
SLC4A3_ex_12F	TTCTTTGCGCAGTTTACAACC	SLC4A3_ex_12R	ACTGTGGACAATGAGGGACAG	373	58		220,499,378
SLC4A3_ex_13F	CCTGTAGCTCAGTGACCCAAC	SLC4A3_ex_13R	GATCAGGGATGCAAAGGACTC	430	58		220,499,899
SLC4A3_ex_14F	GAGAGCGTGAACAGACCAAAG	SLC4A3_ex_14R	ATTCCCTGATCCAGCATCTCT	620	58		220,500,755
SLC4A3_ex_15F	GTCACTCTAAGGGGCTGCTCT	SLC4A3_ex_15R	CAGAACTAGGTGGGGGTAGGA	499	58		220,500,908
SLC4A3_ex_16F	ATTCTCCCTCTTCCCTCGGAGT	SLC4A3_ex_16R	CTAGAACCCTCACTCCAAC	430	58		220,501,301
SLC4A3_ex_17F	CACAGGTGTA CTGATGACGATG	SLC4A3_ex_17R	AAGAAGACACCCAGGGAAGAG	466	58		220,502,227
SLC4A3_ex_18F	CTCTTCCCTGGGTGTCTTCTT	SLC4A3_ex_18R	TCCACTGGGAAAGTTGAGTTC	423	58		220,502,672
SLC4A3_ex_19F	TTCATTTCCCTGCCCTCTATT	SLC4A3_ex_19R	CCCAGCACCTGATACTGATA	470	58		220,503,232
SLC4A3_ex_20F	TGAAAGAAGAAGCTGGCAGTC	SLC4A3_ex_20R	ACACAGCACATTGGGAGATTC	496	58		220,504,538

F primer Name	F primer Seq (5'-3')	R primer Name	R primer Seq (5'-3')	Size (bp)	Tm (°C)	Additive/ alteration*	Position (bp)[†]
SLC4A3_ex_21F	TCCTTCTCTTGTTTCAGGGTGA	SLC4A3_ex_21R	CCTTCATATTCATCCCATCCA	452	58		220,504,964
SLC4A3_ex_22F	TGAAGGAGAAGCTGTGACCTG	SLC4A3_ex_22R	GGCCTTTACCTGGACTAGCAC	499	58		220,505,410
SLC4A3_ex_23F	AGGAGGAGCTGGAGAATGG	SLC4A3_ex_23R	CTCAAACCTCTCCCTGGTTTT	500	58		220,506,302
SLC4A3_ex_c1F	GGGAGGGCGTGTTTACAGA	SLC4A3_ex_c1R	TTCGCTCCCTTCCTCCTGT	489	58		220,495,763
Internal primers for Sanger sequencing of human <i>SLC4A3</i>							
SLC4A3_ex_2IntF	CTGTCCCGGGCGTGGCTGGG						

* Q = Q solution; 98 = Initial denaturation and denaturation steps at 98°C.

† GRCh37

III. PCR and Sequencing Reaction Components

Table A7. TaqMan[®] Allelic Discrimination

Component	Volume/reaction (µL)	Final Concentration
Genomic DNA (~20ng/µL)	2.00	
dNTPs (1.5mM)	2.93	0.2 mM
Forward Primer (20µM)	0.44	0.4 µM
Reverse Primer (20µM)	0.44	0.4 µM
Fluorescent Probe 1 (20µM)	0.22	0.2 µM
Fluorescent Probe 2 (20µM)	0.22	0.2 µM
HST+ 10x Buffer	2.20	1X
HotStarTaq Plus Polymerase (5 U/µL)	0.22	0.05 U/µL
Water (MilliQ)	13.33	
Total (µL)	22.00	

Table A8. Amplification for AFLP Analysis

Component	Volume/reaction (Final concentration)				
	PRA-1	PRA-2	PRA-3v1	PRA-3v2	RCD4
Genomic DNA (~20ng/µL)	2 µL	2 µL	2 µL	2 µL	2 µL
dNTPs (1.5mM)	1.6 µL (0.2 mM)	1.6 µL (0.2 mM)	1.6 µL (0.2 mM)	1.6 µL (0.2 mM)	1.6 µL (0.2 mM)
Forward Primer (10µM)	0.05 µL (0.04 mM)	0.05 µL (0.04 mM)	0.05 µL (0.04 µM)	0.1 µL (0.08 µM)	0.3 µL (0.25 µM)
Forward Primer 2 (10µM)				0.1 µL (0.08 µM)	
Reverse Primer (10µM)	0.05 µL (0.04 mM)	0.05 µL (0.04 mM)	0.05 µL (0.04 µM)	0.1 µL (0.08 µM)	0.3 µL (0.25 µM)
Q Solution	0	0	0	0	2.4 µL
10x Buffer	1.2 µL (1X)	1.2 µL (1X)	1.2 µL (1X)	1.2 µL (1X)	1.2 µL (1X)
HotStarTaq Plus Polymerase (5 U/µL)	0.07 µL (0.03 U/µL)	0.07 µL (0.03 U/µL)	0.12 µL (0.05 U/µL)	0.12 µL (0.05 U/µL)	0.12 µL (0.05 U/µL)
Water (MilliQ)	7.03 µL	7.03 µL	6.98 µL	6.78 µL	4.08 µL
Total (µL)	12 µL	12 µL	12 µL	12 µL	12 µL

Table A9. Amplification for Sequencing (HotStarTaq Plus)

Component	Volume/reaction (μL)		Final Concentration
	Without Q solution	With Q solution	
Genomic DNA ($\sim 20\text{ng}/\mu\text{L}$)	2	2	
dNTPs (1.5mM)	1.6	1.6	0.2 mM
Forward Primer (20 μM)	0.3	0.3	0.5 μM
Reverse Primer (20 μM)	0.3	0.3	0.5 μM
Q Solution	0	2.4	1X
HotStarTaq Plus 10x Buffer	1.2	1.2	1X
HotStarTaq Plus Polymerase (5 U/ μL)	0.12	0.12	0.05 U/ μL
Water (MilliQ)	6.48	4.08	
Total (μL)	12	12	

Table A10. Amplification for Sequencing (Amplitaq Gold 360)

Component	Volume/reaction (μL)	Final Concentration
Genomic DNA ($\sim 20\text{ng}/\mu\text{L}$)	2	
dNTPs (1.5mM)	1.6	0.2 mM
Mg solution (25mM)	0.96	2.0 mM
Forward Primer (20 μM)	0.6	0.2 μM
Reverse Primer (20 μM)	0.6	0.2 μM
GC Enhancer	2	
ATG360 10x Buffer	1.2	1X
Amplitaq Gold 360 Polymerase (5 U/ μL)	0.12	0.05 U/ μL
Water (MilliQ)	2.92	
Total (μL)	12	

Table A11. Amplification for Sequencing (Herculase II)

Component	Volume/reaction (μL)	Final Concentration
Genomic DNA ($\sim 20\text{ng}/\mu\text{L}$)	2	
dNTPs (1.5mM)	0.4	0.05 mM
Forward Primer (20 μM)	0.15	0.25 μM
Reverse Primer (20 μM)	0.15	0.25 μM
DMSO	0.84	7%
Herculase 5x Buffer	2.4	1X
Herculase Polymerase (5 U/ μL)	0.12	0.05 U/ μL
Water (MilliQ)	5.94	
Total (μL)	12	

Table A12. Amplification for Microsatellite Genotyping

Component	Volume/reaction (μL)	Final Concentration
Genomic DNA ($\sim 20\text{ng}/\mu\text{L}$)	2	
dNTPs (1.5mM)	1.6	0.2 mM
Forward Primer (20 μM)	0.1	0.17 μM
Reverse Primer (20 μM)	0.25	0.42 μM
Fluorescent Primer (10 μM)	0.6	0.5 μM
Amplitaq Gold 10x Buffer	1.2	1X
Amplitaq Gold DNA Polymerase (5 U/ μL)	0.12	0.05 U/ μL
Water (MilliQ)	6.48	
Total (μL)	12	

Table A13. Sanger Sequencing (Canine DNA)

Component	Volume/reaction (μL)	Final Concentration
Template (PCR product)	2.00	
2.5X BigDye Ready Reaction Mix	0.50	0.2X
10X SBDD Buffer	1.00	1.7X
Primer (1.6 μM)	1.00	0.27 μM
Water (MilliQ)	1.50	
Total (μL)	6.00	

Table A14. Sanger Sequencing (Human DNA)

Component	Volume/reaction (μL)	Final Concentration
Template (PCR product)	2.00	
2.5X BigDye Ready Reaction Mix	0.50	0.125X
5X Sequencing Buffer	2.50	1.25X
Primer (20 μM)	0.50	0.27 μM
Water (MilliQ)	4.50	
Total (μL)	10.00	

Table A15. Quantitative PCR of *FAM161A* mRNA

Component	Volume/reaction (μL)	Final Concentration
Template (cDNA)	2.00	
2X KAPA SYBR FAST	3	1X
Forward Primer (5 μM)	0.12	0.2 μM
Reverse Primer (5 μM)	0.12	0.2 μM
Water (MilliQ)	0.76	
Total (μL)	6.00	

IV. Thermal Cycling Parameters

Table A16. TaqMan[®] Allelic Discrimination

	Temperature	Duration	Cycles
Initial Denaturation	95°C	5 min	
Denaturation	95°C	30 s	40
Annealing	T _m *	30 s	
Extension	72°C	30 s	
Final Extension	72°C	10 min	
Allelic Reads	25°C	10 s	

* Refer to Appendix II, Table A1.

Table A17. Amplification for Sequencing and AFLP

	Temperature	Duration	Cycles
Initial Denaturation	94°C	10 min	
Denaturation	94°C	1 min	35
Annealing	T _m *	1 min	
Extension	72°C	2 min	
Final Extension	72°C	10 min	

* Refer to Appendix II, Table A1, Table A3, Table A4 and Table A6.

Table A18. Touchdown PCR

	Temperature	Duration	Cycles
Initial Denaturation	98°C	10 min	
Denaturation	94°C	1 min	9
Annealing	68-55°C*	1 min	
Extension	72°C	1 min	
Denaturation	94°C	1 min	26
Annealing	54°C	1 min	
Extension	72°C	1 min	
Final Extension	72°C	10 min	

* Starting at 68°C for the first cycle, decreasing for each subsequent cycle to 66°C, 64°C, 62°C, 60°C, 58°C, 57°C, 56°C and 55°C.

Table A19. Microsatellite Genotyping

	Temperature	Duration	Cycles
Initial Denaturation	94°C	4 min	
Denaturation	94°C	1 min	30
Annealing	T _m *	1 min	
Extension	72°C	1 min	
Denaturation	94°C	1 min	8
Annealing	50°C	1 min	
Extension	72°C	1 min	
Final Extension	72°C	30 min	

* Refer to Appendix II, Table A2.

Table A20. Sanger Sequencing (Canine DNA)

	Temperature	Duration	Cycles
Initial Denaturation	96°C	30 s	
Denaturation	92°C	4 s	44
Annealing	55°C	4 s	
Extension	60°C	110 s	

Table A21. Sanger Sequencing (Human DNA)

	Temperature	Duration	Cycles
Initial Denaturation	96°C	2 min	
Denaturation	96°C	10 s	25
Annealing	50°C	5 s	
Extension	60°C	4 min	

Table A22. Quantitative PCR

	Temperature	Duration	Cycles
Initial Denaturation	95°C	10 min	
Denaturation	95°C	15 s	40
Annealing	57°C	30 s	
Extension	72°C	30 s	
Denaturation	95°C	15 s	1
Melt Curve Analysis	55°C	15 s	
	95°C	15 s	

V. Genes found in PRA-Associated Critical Regions

Table A23. Genes Within GR_PRA1 Critical Region

Canine Gene	Position (CanFam2.0)	Human homologue	Full Name
NHEJ1	37:28636730-28714886	NHEJ1	Nonhomologous end-joining factor 1
SLC23A3	37:28720831-28727019	SLC23A3	Solute carrier family 23 (nucleobase transporters), member 3
C2orf24	37:28729402-28732994	C2orf24	Chromosome 2 open reading frame 24
FAM134A	37:28735172-28739313	FAM134A	Family with sequence similarity 134, member A
ZFAND2B	37:28757195-28759520	ZFAND2B	Zinc finger, AN1-type domain 2B
ABCB6	37:28759982-28767548	ABCB6	ATP-binding cassette, sub-family B (MDR/TAP), member 6
ATG9A	37:28769198-28775393	ATG9A	ATG9 autophagy related 9 homolog A (<i>S. cerevisiae</i>)
ANKZF1	37:28777601-28783188	ANKZF1	Ankyrin repeat and zinc finger domain containing 1
GLB1L	37:28783030-28794586	GLB1L	Galactosidase, beta 1-like
STK16	37:28793976-28796802	STK16	Serine/threonine kinase 16
TUBA4A	37:28798653-28802356	TUBA4A	Tubulin, alpha 4a
XM_848175.1	37:28815425-28818299	No homologues	
DNAJB2	37:28825498-28831692	DNAJB2	DnaJ (Hsp40) homolog, subfamily B, member 2
Q09IV7_CANFA	37:28835626-28853430	PTPRN	Protein tyrosine phosphatase, receptor type, N
RESP18	37:28863932-28869434	RESP18	Egulated endocrine-specific protein 18 homolog (rat)
DNPEP	37:28897527-28908427	DNPEP	Aspartyl aminopeptidase
DESM_CANFA	37:28933165-28939896	DES	Desmin
SPEG	37:28947555-29004449	SPEG	SPEG complex locus
GMPPA	37:29011301-29017577	GMPPA	GDP-mannose pyrophosphorylase A
ACCN4	37:29024803-29047411	ACCN4	Amiloride-sensitive cation channel 4, pituitary
CHPF	37:29048752-29052916	CHPF	Chondroitin polymerizing factor

Canine Gene	Position (CanFam2.0)	Human homologue	Full Name
TMEM198	37:29054074-29058919	TMEM198	Transmembrane protein 198
OBSL1	37:29064807-29078905	OBSL1	Obscurin-like 1
INHA	37:29080814-29083645	INHA	Inhibin, alpha
ENSCAFG00000015687	37:29111682-29119204	STK11IP	Serine/threonine kinase 11 interacting protein
SLC4A3	37:29137454-29150949	SLC4A3	Solute carrier family 4, anion exchanger, member 3
ENSCAFG00000023993	37:29211561-29211881	No homologues	

Table A24. Genes Within GR_PRA2 Critical Regions One and Two

Canine Gene	Position (CanFam2.0)	Human homologue	Full Name
Critical Region One			
GALC_CANFA	8:62306855-62363910	GALC	Galactocerebrosidase Precursor
GPR65	8:62387856-62391052	GPR65	Psychosine receptor
KCNK10	8:62524243-62648292	KCNK10	Potassium channel subfamily K member 10
SPATA7	8:62703376-62736328	SPATA7	Spermatogenesis-associated protein 7
PTPN21	8:62758986-62821262	PTPN21	Tyrosine-protein phosphatase non-receptor type 21
ZC3H14	8:62835738-62888899	ZC3H14	Zinc finger CCCH domain-containing protein 14
EML5	8:62892470-63072182	EML5	Echinoderm microtubule-associated protein-like 5
TTC8	8:63100700-63153499	TTC8	Tetratricopeptide repeat protein 8
ENSCAFG00000024120	8:63372283-63372582	No homologues	
FOXN3	8:63585920-63586534	FOXN3	Forkhead box protein N3
ENSCAFG00000017494	8:63856375-63857294	No homologues	
EFCAB11	8:63917901-64062130	EFCAB11	EF-hand domain-containing protein C14orf143
ENSCAFG00000023877	8:63920525-63920698	No homologues	
TDP1	8:64069226-64148150	TDP1	Tyrosyl-DNA phosphodiesterase 1
KCNK13	8:64167521-64262903	KCNK13	Potassium channel subfamily K member 13
PSMC1	8:64316698-64326521	PSMC1	26S protease regulatory subunit 4
C14orf102	8:64327945-64354840	C14orf102	UPF0614 protein C14orf102
Critical Region Two			
EML1	8:71254961-71332255	EML1	Echinoderm microtubule associated protein like 1
EVL	8:71426908-71499520	EVL	Enah/Vasp-like
DEGS2	8:71501806-71521839	DEGS2	Degenerative spermatocyte homolog 2, lipid desaturase (Drosophila)
YY1	8:71576514-71607017	YY1	YY1 transcription factor

Canine Gene	Position (CanFam2.0)	Human homologue	Full Name
SLC25A29	8:71617342-71623790	SLC25A29	Solute carrier family 25, member 29
SLC25A47	8:71640534-71645117	SLC25A47	Hepatocellular carcinoma-downregulatedmitochondrial carrier protein
WARS	8:71648349-71677009	WARS	Tryptophanyl-tRNA synthetase
WDR25	8:71686511-71825757	WDR25	WD repeat domain 25
BEGAIN	8:71833229-71840503	BEGAIN	Brain-enriched guanylate kinase-associated homolog (rat)
DLK1	8:72005936-72012495	DLK1	Delta-like 1 homolog (Drosophila)
RTL1	8:72145530-72147890	RTL1	Retrotransposon-like 1

Table A25. Genes Within TS_PRA Critical Region

Canine Gene	Position (CanFam2.0)	Human homologue	Full Name
PAPOLG	10:64032104-64056830	PAPOLG	Poly(A) polymerase gamma
REL	10:64149903-64173385	REL	V-rel reticuloendotheliosis viral oncogene homolog (avian)
PUS10	10:64185188-64250210	PUS10	Pseudouridylate synthase 10
PEX13	10:64255013-64279554	PEX13	Peroxisomal biogenesis factor 13
KIAA1841	10:64295804-64345069	KIAA1841	KIAA1841
C2orf74	10:64380504-64383662	C2orf74	Chromosome 2 open reading frame 74
USP34	10:64418779-64623882	USP34	Ubiquitin specific peptidase 34
XPO1	10:64693472-64740735	XPO1	Exportin 1
FAM161A	10:64965191-64991535	FAM161A	Family with sequence similarity 161, member A
CCT4	10:65011057-65025975	CCT4	Chaperonin containing TCP1, subunit 4 (delta)
COMD1_CANFA	10:65037469-65210770	COMMD1	COMM domain-containing protein 1 (Protein Murr1)
B3GNT2	10:65296758-65298488	B3GNT2	Beta-1,3-N-acetylglucosaminyltransferase 2
TMEM17	10:65507545-65512665	TMEM17	Transmembrane protein 17
EHBP1	10:65779934-65976775	EHBP1	EH domain binding protein 1
OTX1	10:65984099-65987213	OTX1	Orthodenticle homeobox 1
ENSCAFG00000024199	10:66114901-66115059	WDPCP	WD repeat containing planar cell polarity effector
ENSCAFG00000023673	10:66187681-66187962	WDPCP	WD repeat containing planar cell polarity effector
LOC474613	10:66290089-66428549	WDPCP	WD repeat containing planar cell polarity effector
ENSCAFG00000003121	10:66426233-66428132	No homologues	
MDH1	10:66467514-66485173	MDH1	Malate dehydrogenase 1, NAD (soluble)
UGP2	10:66682947-66748401	UGP2	UDP-glucose pyrophosphorylase 2
VPS54	10:66749716-66804065	VPS54	Vacuolar protein sorting 54 homolog (S. cerevisiae)
PELI1	10:66879734-66893812	PELI1	Pellino homolog 1 (Drosophila)

Canine Gene	Position (CanFam2.0)	Human homologue	Full Name
ENSCAFG00000023240	10:66885275-66886402	No homologues	
ENSCAFG00000003168	10:67173703-67174591	DPPA4	Developmental pluripotency associated 4
XM_538509.2	10:67196363-67199847	AC008074.1	Galectin-related protein
AFTPH	10:67281829-67321772	AFTPH	Aftiphilin
SERTAD2	10:67356859-67357892	SERTAD2	SERTA domain containing 2
SLC1A4	10:67633294-67656064	SLC1A4	Solute carrier family 1
CEP68	10:67698865-67709087	CEP68	Centrosomal protein 68kDa
RAB1A_CANFA	10:67714196-67726087	RAB1A	RAB1A, member RAS oncogene family

Table A26. Genes Within GoS_PRA Critical Region

Canine Gene	Position (CanFam2.0)	Human homologue	Full Name
XM_845379.1	17:24609166-24640801	ZNF512	Zinc finger protein 512
ENSCAFG00000024084	17:24647124-24647912	CCDC121	Coiled-coil domain containing 121
XM_845410.1	17:24649301-24667835	No homologues	GPN-loop GTPase 1
LOC608365	17:24670387-24681479	SUPT7L	Suppressor of Ty 7 (<i>S. cerevisiae</i>)-like
LOC608897	17:24737529-24749996	ZNF512	Zinc finger protein 512
LOC607762	17:24762266-24777008	No homologues	
LOC608285	17:24779556-24790704	SUPT7L	Suppressor of Ty 7 (<i>S. cerevisiae</i>)-like
SLC4A1AP	17:24793005-24819519	SLC4A1AP	Solute carrier family 4 (anion exchanger), member 1, adaptor protein
RBKS	17:24898729-24971223	RBKS	Ribokinase
BRE	17:24993320-25391957	BRE	Brain and reproductive organ-expressed (TNFRSF1A modulator)
Q9BGG6_CANFA	17:25444774-25464193	FOSL2	FOS-like antigen 2
PLB1	17:25558985-25660756	PLB1	Phospholipase B1
PP1B_CANFA	17:25783355-25803428	PPP1CB	Protein phosphatase 1, catalytic subunit, beta isozyme
TRMT61B	17:25845651-25860445	TRMT61B	tRNA methyltransferase 61 homolog B (<i>S. cerevisiae</i>)
WDR43	17:25872355-25915232	WDR43	WD repeat domain 43
FAM179A	17:25954776-25995537	FAM179A	Family with sequence similarity 179, member A
C2orf71	17:26010127-26013342	C2orf71	Chromosome 2 open reading frame 71
CLIP4	17:26024584-26102807	CLIP4	CAP-GLY domain containing linker protein family, member 4
ALK	17:26113404-26307961	ALK	Anaplastic lymphoma receptor tyrosine kinase
ENSCAFG00000024036	17:26435738-26435941	No homologues	
ENSCAFG00000023252	17:26576350-26597037	No homologues	
ENSCAFG00000023222	17:26790612-26791325	No homologues	
YPEL5	17:26966691-26979394	YPEL5	Yippee-like 5 (<i>Drosophila</i>)

Canine Gene	Position (CanFam2.0)	Human homologue	Full Name
LBH	17:27046659-27069749	LBH	limb bud and heart development homolog (mouse)
LCLAT1	17:27300558-27421174	LCLAT1	Lysocardiolipin acyltransferase 1
CAPN13	17:27496324-27548426	CAPN13	Calpain 13
GALNT14	17:27638511-27715023	GALNT14	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 14 (GalNAc-T14)

VI. NGS Analysis Pipelines

The accompanying CD contains all files and folders required for and generated by the NGS analysis pipeline (script written by Oliver Forman) and the target capture efficiency analysis pipeline (script written by Louise Downs). Only the two perl scripts, **ngs15_3.pl** and **ngs_capture_v4.3.pl**, were used for the analysis of data in this project. The input and output files described below are abridged files used for rapid demonstration purposes only.

Folders and files on the CD:

1. Input Files

The following data files were used for the demonstration of both analysis pipelines.

- test1.fastq → Raw sequence file (first of a pair) used in **ngs_fastq.avi**.
- test2.fastq → Raw sequence file (second of a pair) used in **ngs_fastq.avi**.
- test.fasta → Reference genome used in **ngs_fastq.avi** and **ngs_bam.avi**.
- test1.bam → Binary, aligned sequence file, created from test1.fastq and test2.fastq, used in **ngs_bam.avi** and **capture.avi**.
- test2.bam → Binary, aligned sequence file used in **capture.avi**.
- test_baits.bed → Co-ordinates of biotinylated RNA baits, generated by the Agilent Technologies eArray tool and used in **capture.avi**.

2. NGS Analysis

- ngs.pdf → A PDF of the NGS analysis pipeline perl script (**ngs15_3.pl**).
- ngs_fastq.avi → A video clip capturing the use of the NGS analysis pipeline using the test fastq files described above. The files created during this analysis are in the folder **results_test1**.
- ngs_bam.avi → A video clip capturing the use of the NGS analysis pipeline using the test bam file described above. The files created during this analysis are in the folder **results_test_bam**.

2.1. results_test1

All files generated are summarised in **README.rtf**. However, the main files used for subsequent analysis are:

- best_align_test.bam → Binary, aligned sequence file, with PCR duplicates removed, used for loading in IGV (Integrated Genomics Viewer), along with best_align_test.bai, to view sequence reads graphically. A similar file, raw_align_test.bam, has not had PCR duplicates removed.
- INDELS_test.bed → a list of the insertions and deletions identified.
- SNPS_test.vcf → A list of the SNPs identified.

2.2. results_test1_bam

The files are similar to those described in folder **results_test1**.

3. Target Capture Efficiency

- capture.pdf → A PDF of the target capture efficiency pipeline perl script (**ngs_capture_v4.3.pl**).
- capture.avi → A video clip capturing the use of the target capture efficiency analysis pipeline using the test bam files described above. The files created during this analysis are in the folder **capture_results_test**.

3.1. capture_results_test

Three file types are generated:

- Log_test.rtf → A record of the analysis and user-specified parameters.
- Text files → There are two files for each sample analysed, one with PCR duplicates removed and one without PCR duplicates removed.
- Excel files → There are two files, one with PCR duplicates removed and one without PCR duplicates removed. Each is a summary of the results generated and is compiled from the individual text files.

VII. GR PRA-2 Questionnaire



Please return to:
 Louise Downs
 Genetics Department
 Animal Health Trust
 Lanwades Park
 Kentford
 Newmarket
 CB8 7UU
 United Kingdom
 E-mail: louise.downs@aht.org.uk

Golden Retriever Information Update

Please fill in information and check the boxes (☐) as required.

Owner's Name	
Owner's Address	
Telephone no.	
Email	
Name of Dog	
Date of birth	

How would you describe the general and overall health of your dog?

- Good Usually only visit the vet for routine vaccinations.
- Average Veterinary care (for illness) is required at 1-2 times per year.
- Poor Veterinary care (for illness) is required at 3 or more times per year.

Does your dog have any chronic conditions i.e. that require constant or regular treatment?

- No
- Yes

If Yes, please describe in more detail:

Based on the enclosed "Body Condition System" chart, what number (1-9) would you give your dog?

How much does your dog weigh (in kg)?

Is your dog on a weight-loss or weight-control diet?

- No
- Yes

If Yes, please describe in more detail:

Has your dog ever had or showed signs of any of the following?


- Kidney abnormalities
- Digit abnormalities (e.g. more toes than normal, or toes fused together)
- Obesity
- Reproductive abnormalities (e.g. problems mating)
- Poor sense of smell
- Diabetes mellitus
- Social or behavioural problems
- Other
- None of the above

If you have checked any of the above boxes, please explain:

Is there anything else you would like us to know?







To my knowledge the above information is correct:

Please signDate.....



Nestlé PURINA

BODY CONDITION SYSTEM

TOO THIN	1	Ribs, lumbar vertebrae, pelvic bones and all bony prominences evident from a distance. No discernible body fat. Obvious loss of muscle mass.	  1
	2	Ribs, lumbar vertebrae and pelvic bones easily visible. No palpable fat. Some evidence of other bony prominence. Minimal loss of muscle mass.	
	3	Ribs easily palpated and may be visible with no palpable fat. Tops of lumbar vertebrae visible. Pelvic bones becoming prominent. Obvious waist and abdominal tuck.	
IDEAL	4	Ribs easily palpable, with minimal fat covering. Waist easily noted, viewed from above. Abdominal tuck evident.	  3
	5	Ribs palpable without excess fat covering. Waist observed behind ribs when viewed from above. Abdomen tucked up when viewed from side.	
TOO HEAVY	6	Ribs palpable with slight excess fat covering. Waist is discernible viewed from above but is not prominent. Abdominal tuck apparent.	  5
	7	Ribs palpable with difficulty; heavy fat cover. Noticeable fat deposits over lumbar area and base of tail. Waist absent or barely visible. Abdominal tuck may be present.	
	8	Ribs not palpable under very heavy fat cover, or palpable only with significant pressure. Heavy fat deposits over lumbar area and base of tail. Waist absent. No abdominal tuck. Obvious abdominal distention may be present.	
	9	Massive fat deposits over thorax, spine and base of tail. Waist and abdominal tuck absent. Fat deposits on neck and limbs. Obvious abdominal distention.	
	9		

The **BODY CONDITION SYSTEM** was developed at the Nestlé Purina Pet Care Center and has been validated as documented in the following publications:

Mawby D, Bartges JW, Moyers T, et al. *Comparison of body fat estimates by dual-energy x-ray absorptiometry and deuterium oxide dilution in client owned dogs.* *Compendium* 2001; 23 (9A): 70

Lafamme DP. *Development and Validation of a Body Condition Score System for Dogs.* *Canine Practice* July/August 1997; 22:10-15

Kealy, et al. *Effects of Diet Restriction on Life Span and Age-Related Changes In Dogs.* *JAVMA* 2002; 220:1315-1320

Call 1-800-222-VETS (8387), weekdays, 8:00 a.m. to 4:30 p.m. CT

