Disease gene identification in Australian kelpies

Author:
Pan, Annie Ying-Hui

Publication Date:
2012

DOI:
https://doi.org/10.26190/unsworks/15445

Disease Gene Identification in Australian Kelpies

by

Annie Ying-Hui Pan

The University of New South Wales

A thesis submitted as partial fulfilment

of the requirements for the degree of

Master of Philosophy (BABS)

School of Biotechnology and Biomolecular Sciences

University of New South Wales

2012

THE UNIVERSITY OF NEW SOUTH WALES
Thesis/Dissertation Sheet

Surname or Family name: PAN

First name: ANNIE **Other name/s:** YING-HUI

Abbreviation for degree as given in the University calendar: MPhil (BABS)

School: Biotechnology and Biomolecular Sciences (BABS) **Faculty:** SCIENCE

Title: Disease Gene Identification in Australian Kelpies

Abstract 350 words maximum: (PLEASE TYPE)

The Australian Kelpie was generated from three pairs of working Scottish Collie dogs in 1870. The new breed was developed in an effort to ease the growing working demand faced by sheep farming. Like most dog breeds, inbreeding has been common within the Kelpie breed and persistent breeding of champion dogs (popular sire effect) has resulted in the spread of autosomal recessive conditions. One of these, Cerebellar Abiotrophy (CA) results in ataxia, characterized by a head tremor, poor body coordination and high stepping gait. Whole genome association and homozygosity analysis have mapped the CA locus to a 5 Mb region on chromosome 3 and identified a common SNP haplotype shared between all affecteds and some unaffected controls. This region was sequenced on a 454 platform in two affected and one control dogs. A total of 2019 differences were identified homozygous in the two affecteds compared to the control, 17 of which were synonymous substitutions in coding exons and substitutions in the untranslated regions of mRNA. On top of this, a total of 22 differences in introns and intergenic regions with high sequence conservation between other mammals and dogs were identified and investigated as possible causative mutations for CA in Kelpies. PCR and Sanger sequencing were employed to fill in 454 sequencing gaps for a total of 40 coding, non-coding exons and upstream regions of different genes within the candidate region. One intergenic (31674050) deletion conserved between other mammals and dogs was homozygous in the affecteds compared to controls and was predicted to disrupt a HSF2 transcription factor binding site for regulation of neighbouring genes. In addition, sequencing of the *DMGDH* gene in affected and unaffected control dogs identified a 223 bp insertion upstream of the last exon. This insertion may produce aberrant mRNA splicing and protein truncation, and may be a causative factor in the CA phenotype.

FOR OFFICE USE ONLY **Date of completion of requirements for Award:**

THIS SHEET IS TO BE GLUED TO THE INSIDE FRONT COVER OF THE THESIS

ORIGINALITY STATEMENT

'I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.'

Signed 

Date 20/06/12

COPYRIGHT STATEMENT

'I hereby grant the University of New South Wales or its agents the right to archive and to make available my thesis or dissertation in whole or part in the University libraries in all forms of media, now or here after known, subject to the provisions of the Copyright Act 1968. I retain all proprietary rights, such as patent rights. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.
I also authorise University Microfilms to use the 350 word abstract of my thesis in Dissertation Abstract International (this is applicable to doctoral theses only).
I have either used no substantial portions of copyright material in my thesis or I have obtained permission to use copyright material; where permission has not been granted I have applied/will apply for a partial restriction of the digital copy of my thesis or dissertation.'

Signed .....

Date 20/06/12 .....

AUTHENTICITY STATEMENT

'I certify that the Library deposit digital copy is a direct equivalent of the final officially approved version of my thesis. No emendation of content has occurred and if there are any minor variations in formatting, they are the result of the conversion to digital format.'

Signed .....

Date 20/06/12 .....

DEDICATION

This thesis is dedicated to the memory of my supervisor, Associate Professor Alan N. Wilton (1953-2011), who is now rested peacefully in the hands of angels.

A Farewell

I must learn to let it go. Alan, you are my first supervisor. I will always remember you and your teachings. Have a safe journey. And please look after yourself. Take care! I will take care of myself too. You will be greatly missed. I wish I can be fortunate enough to get to meet and know you better in the next life (if there is one). I shall see you soon!

You will always be in my heart, and you shall never be forgotten regardless of whatever may happen in the future. Good bye Alan. Please enjoy your time in the new beautiful world, a place without any pain.

ACKNOWLEDGEMENTS

I wish to thank the many breeders and owners of Australian Kelpies both within and outside Australia for providing valuable blood and brain tissue samples from their beloved Kelpie dogs for this important research. My special appreciation and gratitude goes to the owners of Kelpies that were affected with Cerebellar Abiotrophy. Without these precious samples and detailed pedigree information this research would not have been possible to continue.

I must thank my supervisor Alan Wilton for accepted me as one of his final students. He taught me a lot of genetics during the two years of my master degree and without his selfless teachings I would not have been able to complete this genetics-based project. I must also thank him for giving me so much guidance, supports, encouragements and understanding whenever I felt lost and unsure what to do next. He is always here when I need him to help me with anything with regards to my research and other personal issues. So thank you Alan. You are the best supervisor I will ever have during my time as a student. I am feeling truly lucky to have you as my supervisor and I will never forget anything that you have taught me during the past 2 years. Thank you for making my 2 years lab research experiences so much fun and memorable, I will always remember these days for the rest of my life. No words could fully describe my gratitude to you. For me, you are a friend as well as a teacher. I will always remember you.

Particular thanks to Dr. Sven Delaney, my co-supervisor for spending his precious time to go through my thesis many times (and giving me lots of comments and suggestions) while trying to organise all necessary administrative paperwork in time for the thesis submission. Without his support and understanding, this thesis would not

have been possible. So thank you Sven. I felt fortunate to have you as my co-supervisor and be always on my side while Alan has moved to the Heaven permanently.

Special thanks to Dr. Louise Lutze-Mann for giving me after-hour access to her lab so I can use her lab equipments for my Southern blot hybridisation experiments. I must also thank her for giving me counselling support and helping me getting through my emotional low tide, as a friend during the most difficult times of my master degree – when my supervisor was sick. I really appreciate her understanding and her support.

Thanks to our collaborators Professor Rosanne Taylor and Jessica Fletcher at University of Sydney for doing histopathological studies on brain tissues for the Kelpie CA research. Thanks to Dr. Martin Horan (UNSW) for helping me to analyse my results and giving me guidance for thesis while my supervisor was away on sick leave.

I also wish to thank The Working Kelpie Council of Australia and Terry Snow (Canberra Airport) who provided financial support to the project and generated awareness of the disease and the importance of submitting samples to the project. Thanks also to Marie Colyer for assistance and information.

Thanks to all the past and present members of the Wilton lab: Jeremy, Auda, Paulina, Josh, Sven, Matthew, Emilio, Joe, Thore, Zi and Pranoy, for bringing all the laughs and joy during my time working in the lab. Also thanks to people who occupied the other end of the lab (aka. Ballard lab), Professor Bill Ballard and his group: Martin, Jonci, Pann Pann, Nico, Carolina, Wen, Michael, Ida and Kylie. Also big thanks to Tonia Russell of The Ramaciotti Centre for Gene Function Analysis at UNSW, for her technical assistance in sequencing and troubleshooting.

Most importantly, I could never have managed the MPhil without support and understanding of my family and friends. To my parents, sisters, brother and the two dogs, I offer the most heartfelt thanks for their encouragement and ongoing belief in me. Without my friends: Nirmani, Erica, David, Vanessa, Marie, Ben and Jackie, I could not have the courage and motivation to move forward.

I also would like to thank my supervisor, Alan Wilton and Dr. Barbara Zangerl for their comments and corrections to the thesis.

Last but not least, I want to express my deepest gratitude and appreciation to my supervisor Associate Professor Alan Wilton again, for giving me the opportunity to take part in this ‘challenging’ project and giving me all the tools (both tangible and intangible) that I need to identify the disease gene in Kelpie.

ABSTRACT

The Australian Kelpie was generated from three pairs of working Scottish Collie dogs in 1870. The new breed was developed in an effort to ease the growing working demand faced by sheep farming. Like most dog breeds, inbreeding has been common within the Kelpie breed and persistent breeding of champion dogs (popular sire effect) has resulted in the spread of autosomal recessive conditions. One of these, Cerebellar Abiotrophy (CA) results in ataxia, characterized by a head tremor, poor body coordination and high stepping gait. Whole genome association and homozygosity analysis have mapped the CA locus to a 5 Mb region on chromosome 3 and identified a common SNP haplotype shared between all affecteds and some unaffected controls. This region was sequenced on a 454 platform in two affected and one control dogs. A total of 2019 differences were identified homozygous in the two affecteds compared to the control, 17 of which were synonymous substitutions in coding exons and substitutions in the untranslated regions of mRNA. On top of this, a total of 22 differences in introns and intergenic regions with high sequence conservation between other mammals and dogs were identified and investigated as possible causative mutations for CA in Kelpies. PCR and Sanger sequencing were employed to fill in 454 sequencing gaps for a total of 40 coding, non-coding exons and upstream regions of different genes within the candidate region. One intergenic (31674050) deletion conserved between other mammals and dogs was homozygous in the affecteds compared to controls and was predicted to disrupt a HSF2 transcription factor binding site for regulation of neighbouring genes. In addition, sequencing of the *DMGDH* gene in affected and unaffected control dogs identified a 223 bp insertion upstream of the last exon. This insertion may produce aberrant mRNA splicing and protein truncation, and may be a causative factor in the CA phenotype.

TABLE OF CONTENTS

Page No.

ABBREVIATIONS

Acc. No.	accession number
ARCAs	autosomal recessive cerebellar ataxias
BC	Border Collies
CA	Cerebellar abiotrophy
CFA	chromosome
dNTP	deoxynucleotide tri-phosphate
EDTA	ethylene diamine tetra acetic acid
emPCR	emulsion PCR
EtBr	ethidium bromide
FCG	functional candidate gene
GWAS	genome wide association study
IBD	identical by descent
IBS	identical by state
IDLs	insertion-deletion loops
LINE	long interspersed nuclear element
LOD	logarithm of the odds
MSP	methylation-specific PCR
NCBI	National Center for Biotechnology Information
ORFs	open reading frames
PCR	polymerase chain reaction
qPCR	quantitative real-time PCR
RT-PCR	reverse transcription PCR
SBA	sodium boric acid
SDS	sodium dodecyl sulphate
SINE	short interspersed nuclear element
SNP	single nucleotide polymorphism
STR	short tandem repeat
TF	transcription factor
TSS	transcription start site
UNSW	University of New South Wales
UTR	un-translated region
UV	ultra-violet
WGA	whole genome association
WK	working Kelpies

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 Introduction

1.1.1 Origin of the Australian Kelpie

The Australian Kelpie is a breed of highly intelligent working dogs generated from three pairs of imported Scottish Collie dogs in 1870 (Parsons 1992). The breed was created to work and live under extreme weather conditions in all parts of Australia and to ease the work of sheep farmers. The Australian Kelpie was originally selected for its excellent natural working ability on the farm, however some breeders in recent years have selected for physical characteristics such as coat colour. This has given rise to two Kelpie strains: Working Kelpies and Show Kelpies (Parsons 1992).

The Australian Kelpie is a medium sized dog (Figure 1-1) with an average adult height between 43 and 51 cm and adult weight between 11 and 20 kg. The most common coat colour of Australian Kelpie is black, with some coat colour variations including black and tan, red, and blue (Parsons 1992).

Figure 1-1: A typical Australian Kelpie with red coloured coat.

1.1.2 Dog disease as a model for studying human genetic diseases

There are over 400 breeds of dogs with each breed having its own unique morphological and behaviour traits (Vila *et al.* 1997). The level of genetic diversity within a breed is small with little genetic variation between dogs of the same breed (Parker and Ostrander 2005). This is largely due to the extensive use of champion dogs with desirable characteristics as founder dogs for subsequent breeding programs over many generations. The inbreeding resulting from this practice has reduced the size of the gene pool and has increased the incidence of genetic disease due to the over-representation of deleterious alleles from founder dogs (Asher *et al.* 2009; Chase *et al.* 2009). This has led to a rapid spread of recessive and dominant disease alleles with incomplete penetrance or late onset in dogs (Higgins and Nicholas 2008).

The dog has a unique population structure that makes it useful for mapping of simple genetic disorders. As a consequence of inbreeding, dogs have larger haplotype blocks compared to humans (Lindblad-Toh *et al.* 2005). Large haplotype blocks with a strong linkage disequilibrium allows mapping of disease genes in dogs to be carried out with relatively few genetic markers in genome wide association studies (GWAS) compared to human studies (Lindblad-Toh *et al.* 2005). 50,000 SNPs are sufficient for mapping most simple genetic defects in dogs using GWAS while up to ten times of that number are needed to map a simple genetic disease in humans (Lindblad-Toh *et al.* 2005). In dogs, GWAS are routinely used to map the genes involved in simple genetic disorders involving one or few genes and a minimal environmental component (Karlsson *et al.* 2007; Sutter *et al.* 2004).

Dogs make a useful animal model for the study of genetic diseases in humans. Dogs and humans share over 360 naturally-occurring genetic diseases including heart

disease, neurological disease, immune disorders, blindness, epilepsy, hip dysplasia, deafness and cancers (Ostrander et al. 2000; Patterson et al. 1982; Patterson et al. 1988; Patterson 2000). More than 250 genetic diseases in dogs and humans are caused by mutations in the same gene or by malfunction of a similar biochemical pathway (Ostrander *et al.* 2000; Parker and Ostrander 2005). Hence disease genes identified in dogs may indicate the genetic basis of similar diseases in humans (Lindblad-Toh *et al.* 2005). Conversely, disease genes identified in human may also be used to study the genetic cause of similar diseases in dogs. For example, the *CLN5* gene that is responsible for the late infantile form of NCL in humans was also the gene associated with NCL in the Border collie dogs (Melville *et al.* 2005).

1.1.3 Overview of Cerebellar abiotrophy in Australian Kelpies

Cerebellar abiotrophy (CA) is a neurodegenerative disease that has been reported in a number of species including dogs, horse, cattle, cats, rabbits and chickens (Berry and Blas-Machado 2003; Bildfell et al. 1995; Blanco et al. 2006; Brault et al. 2011; Gandini et al. 2005; Jokinen et al. 2007; Kemp et al. 1995; Mitchell et al. 1993; Negrin et al. 2006; Nibe et al. 2007; Olby et al. 2004; Sandy et al. 2002; Sato et al. 2011; Smyth et al. 1985; Tatalick et al. 1993; Thomas and Robertson 1989; Willoughby and Kelly 2002; Yasuba et al. 1988). It is a relatively common inherited disease in dogs where it has been identified in a number of different breeds including Border Collies (Sandy *et al.* 2002), miniature schnauzer (Berry and Blas-Machado 2003), Lagotto Romagnolo (Jokinen *et al.* 2007), Labrador retriever (Bildfell *et al.* 1995), American Staffordshire terriers (Olby *et al.* 2004), Brittany Spaniels (Tatalick *et al.* 1993), Beagles (Yasuba *et al.* 1988), Papillons (Nibe *et al.* 2007), English Bulldogs (Gandini *et al.* 2005) and Australian Kelpies (Thomas and Robertson 1989). The genetic

basis of CA in these species and breeds is likely to vary and is poorly understood (Brault *et al.* 2011).

In Australian Kelpies, CA is a slowly progressing autosomal recessive disease affecting less than one percent of dogs (Thomas and Robertson 1989). Since CA is inherited as a recessive disease in Kelpies, the parents of affected dogs are obligate carriers (if unaffected) and unaffected siblings have a 1 in 2 chance of being a carrier for the disease gene. Affected dogs experience ataxia as the main symptom, which includes poor body coordination, head tremors and a high-stepping gait (Shearman *et al.* 2008; Thomas and Robertson 1989). CA symptoms may become apparent at five to seven weeks in severely affected Kelpies, but may not be obvious in mildly affected dogs until twelve weeks of age (Thomas and Robertson 1989).

CA in Australian Kelpies was first reported in 1989 when three puppies were reported to show variable severity of symptoms early in life (but no progression in symptom severity over time) (Thomas and Robertson 1989). Histological reports on these three CA affected Kelpie dogs revealed a loss of Purkinje cells and granular cells in the cerebellum (Figure 1-2) (Thomas and Robertson 1989). This histopathological observation is consistent with previous reported cases of CA in Kerry Blue Terrier dogs, in which affected dogs showed a loss of Purkinje cells and progressive degeneration of other brain regions (Thomas and Robertson 1989). However in a recent histopathology study performed by Shearman *et al.* (Shearman *et al.* 2011) on four cerebellum tissue samples from affected dogs, there was a reduction in granular and molecular cell layers with only a secondary loss of Purkinje cells in severely affected individuals (Shearman *et al.* 2011). Histopathological changes in cerebellum tissues from affecteds can be used to differentiate between different possible causes of ataxia where ataxia caused by CA

displayed a unique pattern of loss of Purkinje cells, granular and molecular cells in cerebellum of affected individuals (Blanco *et al.* 2006).

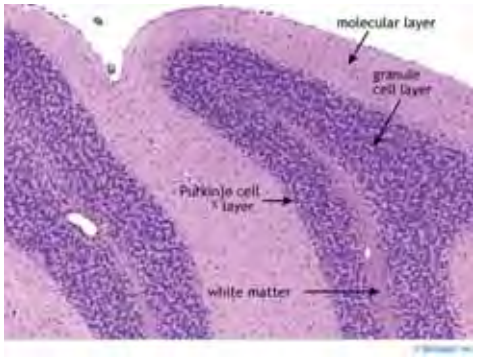


Figure 1-2: A micrograph of the cerebellar cortex showing three different neuron cell layers (Deltagen).

Purkinje cells are some of the largest neuron cells located in the cerebellar cortex of the brain between the granular and molecular layer (Dean *et al.* 2010). Purkinje cells play an important role in coordinate motor movement and body coordination in the cerebellar cortex by sending signals to cerebellar nuclei upon received a response from other neuron cells in the brain (Abrams and Zhang 2011; Loewenstein *et al.* 2006). Death of Purkinje cells shortly after birth would result in CA with ataxia as the main symptom (Berry and Blas-Machado 2003; Shearman et al. 2011; Yasuba et al. 1988).

There is no effective treatment available for Kelpies affected with CA. While CA is not always fatal and may not reduce expected lifespan the quality of life for affected dogs is often poor and many are euthanized.

1.1.4 Rationale for identifying the cause of CA in Kelpies

1.1.4.1 Costs of CA in Kelpies to breeders

The costs of an affected Kelpie dog to breeders and owners can be significant both financially and emotionally. The affected Kelpie is lost to breeding and may need to be euthanized. Since there is no way of identifying which animals are carriers of CA, relatives of the affected dog must also be excluded from breeding. This means a loss of income for breeders. Emotionally, a good and loyal companion is lost for the owner.

1.1.4.2 Benefits of identifying the genetic basis of CA in Kelpies

The identification of CA-causing mutation(s) would result in the development of a reliable genetic test that would allow breeders to identify carriers of CA and hence avoid the production of affected pups. Carriers could still be bred but only by mating with a healthy homozygous normal dog. Such a test and breeding program would help to eventually eliminate the disease allele from the Kelpie breed over a number of generations (Bell 2011). In addition, the gene responsible for CA in Kelpies could also be used as a likely candidate for studying the genetic basis of similar diseases in humans.

1.1.5 Approaches for identification of causative mutation(s)

There are three main approaches to disease gene identification: the functional candidate gene approach (Kwon and Goate 2000; Zhu and Zhao 2007), the whole genome association approach and homozygosity analysis (Jiang et al. 2009; John et al. 2004; Knapp and Becker 2003; Liu et al. 2009), and linkage analysis using microsatellites (Yue *et al.* 2005) (Table 1-1). A combination of one or more of these approaches is commonly adopted to identify the causative mutation in genetic disorders.

Table 1-1: Features, benefits, and limitations of three main approaches commonly used for disease gene identification.

Approaches	Features
Functional candidate gene (FCG)	• Candidate gene is identified based on gene function, expression level in the target tissue or because it causes a similar disease in other mammals (Zhu and Zhao 2007). • *Benefits:* A quick and easy method to identify the likely disease gene if the gene is well characterized in terms of function, expression and biochemical interactions with other genes. • *Limitations:* Can be impractical, expensive and time consuming if there are a large number of candidate genes known to cause a similar disease, since each gene must individually studied and eliminated.
Whole genome association & homozygosity analysis	• Involves the use of large numbers of genetic markers such as SNPs or microsatellites, evenly spaced across the genome, and genotyped in affecteds and controls to identify alleles associated with disease (John *et al.* 2004). • *Benefits*: Does not require prior knowledge of candidate genes. Rapidly identifies the target region which can be looked at more closely. • *Limitations*: Can be resource-intensive as the entire genome needs to be searched for any association with the disease. Often identifies a large region with many genes that can only be narrowed down and eliminated using the FCG approach.
Linkage analysis	• Based on the co-segregation of alleles and the disease within families. Can be applied to single known candidate genes or used to identify candidate regions. • *Benefits*: Can combine data across families. The LOD score can be used to confirm linkage between marker loci and the disease locus, or eliminate the disease gene from a region of the genome. • *Limitations*: Requires identification and genotyping of families segregating for the disease, and large number of individuals are needed.

1.1.6 Mapping the genetic basis of inherited disease using canine SNP arrays

A canine Affymetrix SNP array with 127,000 SNPs has been widely used to localise the candidate region associated with a disease trait in a specific dog breed, by genotyping both affecteds and unaffecteds with SNPs on the array (Lindblad-Toh *et al.* 2005). This array was recently superseded by the Illumina high-density 172K array for performing GWAS in affecteds and healthy controls. This newly developed canine SNP array contains a selection of existing and new SNPs that have been identified by re-sequencing the entire dog genome including the X chromosome (Lequarre *et al.* 2011). The Illumina high-density 172K array contains 45,000 more SNPs than the Affymetrix SNP array which allows disease gene to be mapped to a more precise region in the dog genome and also increases power of the GWAS (Lequarre *et al.* 2011).

The Illumina high-density 172K array was used by Abitbol *et al.* (Abitbol *et al.* 2010) to fine-map the genetic cause of neuronal ceroid lipofuscinosis (NCL) in American Staffordshire terriers to a 1039 kb candidate region (containing 9 genes) on chromosome 9. NCL is a lysosomal storage disorder disease that can cause oxidative stress on neuron cells and leads to apoptosis of neurons (Bellettato and Scarpa 2010). The causative mutation was mapped to exon 2 of the arylsulphatase G (*ARSG*) gene by the case-control association and gene expression studies (Abitbol *et al.* 2010). The Illumina high-density 172K array is currently being used for fine-mapping the genetic basis of a number of complex genetic diseases including dilated cardiomyopathy, myxomatous mitral valve disease and intervertebral disc disease in a variety of different dog breeds (Madsen *et al.* 2011; Mausberg *et al.* 2011; Mogensen *et al.* 2011).

1.1.7 Previous research findings

The functional candidate gene approach involves identifying candidate gene based on gene function or a similar disease in other mammals (Zhu and Zhao 2007). Using this approach, Shearman (Shearman *et al.* 2008) tested three candidate genes that were known to cause CA with mild non-progressive ataxia in humans: *SETX* on chromosome 9, *ATCAY* on chromosome 20 and *SYNE1* on chromosome 1. Since all affected dogs can be traced back to a small number of common ancestors within eight generations, it is hypothesised that all CA affected dogs would have been homozygous for the same causative mutation(s) (i.e. homozygous identical by descent) (Shearman *et al.* 2008). These genes were eliminated as the cause of CA in Australian Kelpies because affecteds were not homozygous in the region of the genes (Shearman *et al.* 2008).

Whole genome association and homozygosity analysis have been used to identify regions of homozygosity common to all individuals affected with CA (Shearman *et al.* 2011). The candidate region for the gene causing CA was mapped to a 5 Mb (28 Mb to 33 Mb) region on chromosome 3 where all CA affected dogs shared the common CA SNP haplotype (NCBI Dog build 2.1) (Shearman *et al.* 2011). There are 44 genes in the candidate region (Figure 1-3). The identification of a common haplotype containing the CA mutation using the Affymetrix canine SNP chip v2 allowed preliminary testing of suspected affected Kelpie dogs for the presence of CA and to identify phenocopies of suspected CA cases.

There was no statistically significant association between SNPs in the region and disease as would usually be expected (Shearman *et al.* 2011). The reason proposed for the common SNP haplotype (CA haplotype) observed in all affecteds and in some of the

controls was that the mutation responsible for CA has occurred on a common haplotype and that only some copies of the haplotype have the CA mutation (Shearman *et al.* 2011). This is presented as the 'common haplotype hypothesis', which assumes that the causative mutation is only present in CA affected Kelpies and is absent in unaffected control dogs with the CA haplotype. This hypothesis was tested by genotyping 4 CA affected families with 5 microsatellites of long repeat lengths and high mutation rates (Shearman *et al.* 2011). A linkage analysis was performed on the microsatellite data of these four CA affected families. The positive LOD score of 2.37 (Figure 3-8) indicates a linkage between marker loci and the disease locus (Shearman *et al.* 2011). However since the LOD score is below the significant threshold value of +3, a LOD score of 2.37 would not normally be considered significant evidence of linkage.

In an effort to identify the causative mutation, Nimblegen sequence capture array and 454 sequencing were used to obtain the sequences for all 44 genes (Figure 1-3, Table 1-2) within the candidate region in two affecteds and one unaffected control (Shearman 2011). When compared, there were 2017 differences between affecteds and the control of which 550 differences were considered to be known differences among other dog breeds and were removed from the study (Shearman 2011). There were eight non-synonymous substitutions among the remaining 1577 differences which were checked in 96 Kelpies and eliminated as causing CA in Kelpies by Sanger sequencing and association studies (Shearman 2011). This left a large number of differences to be investigated further including 17 synonymous substitutions in coding exons and substitutions in the untranslated regions (UTRs) of mRNA (Table 2-2) (Shearman 2011).

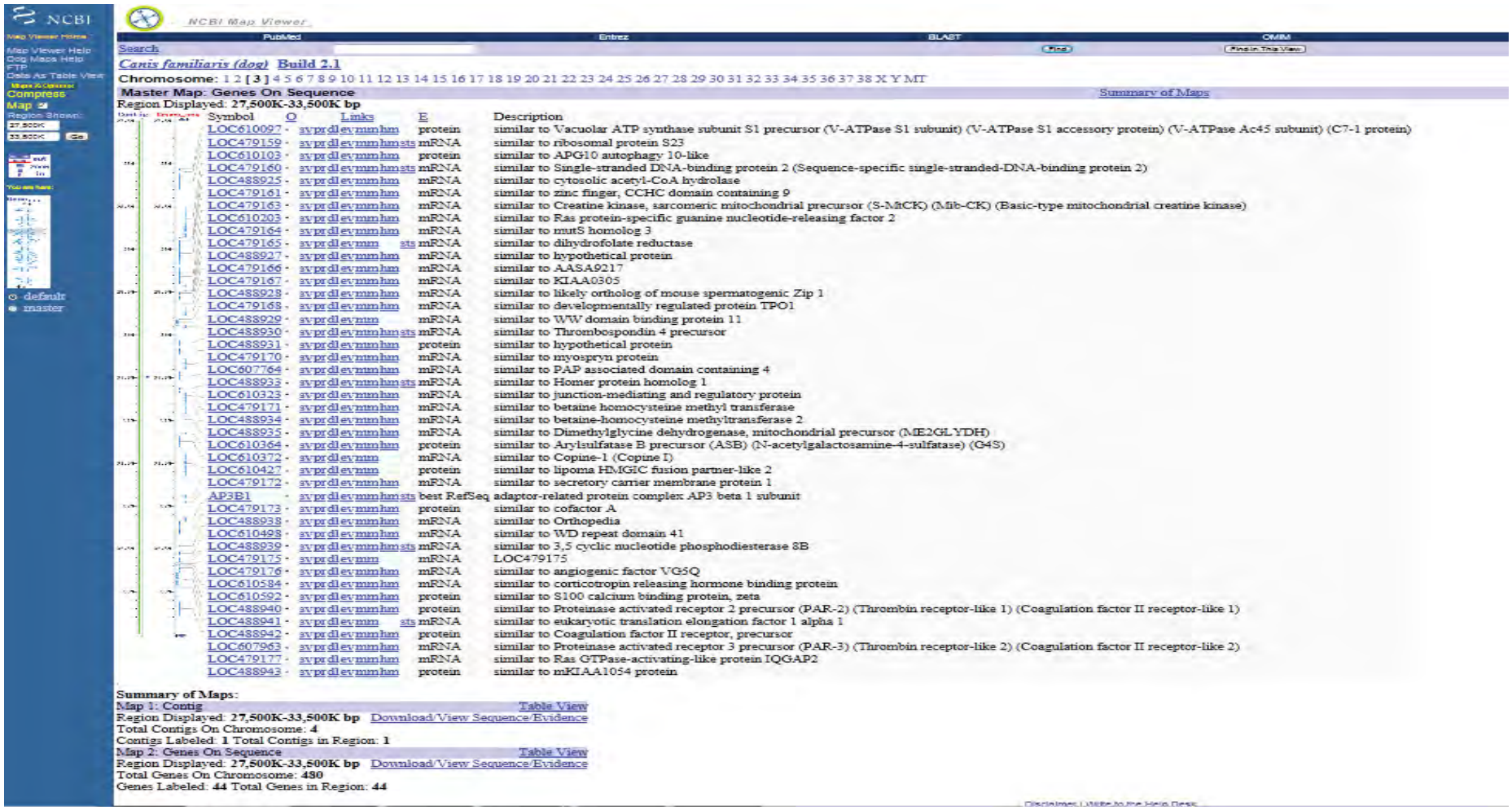


Figure 1-3: A screenshot of the NCBI gene map showing all 44 genes in the candidate region, on chromosome 3 between 28 and 33 Mb of the dog genome (NCBI Dog build 2.1).

1.1.8 Influence on gene function by DNA sequence variations in exons and non-coding DNA

The most common form of genetic variation in a DNA sequence is single base changes in protein-coding and non-coding regions of genes. Those single base changes are known as single nucleotide polymorphism (SNPs), occurring at a frequency of one in every 900 base pairs between breeds throughout the dog genome (Lindblad-Toh *et al.* 2005). Over 50% of these SNPs are located in regions that do not encode any protein, while the remainder is in protein-coding regions. The SNPs in protein-coding regions may have direct functional implications for disease susceptibility if they are non-synonymous substitutions (resulting in a change in amino acid sequence) or nonsense mutations (resulting in stop codon insertion and protein truncation). Both of these mutations can have a large effect on phenotype by changing protein structure and function (Knight 2003). In contrast, synonymous substitutions do not change the amino acid sequence of the translated protein and so are unlikely to affect protein function or phenotype (Knight 2003).

A SNP, insertion or deletion at an exon-intron boundary may also disrupt RNA splicing signals and lead to aberrant splicing of the mRNA transcript. This may result in exon skipping or intron retention or a shift in reading frame. The protein produced from the aberrantly-spliced mRNA may be truncated and/or may have a reduced function when compared to the original protein (Solis *et al.* 2008; Wang and Burge 2008). Many diseases are caused by this mechanism. For example, muscular dystrophy is an X-linked recessive disease in humans caused by mutation at a natural splice site of the dystrophin gene. This mutation causes a frame shift and internal deletion in the mRNA transcript and leads to the absence and/or under-expression of functional dystrophin protein,

which is required for maintaining the strength and movement of the skeletal muscle (Solis *et al.* 2008).

SNPs in coding regions can have obvious effects on protein structure and function, but it is often difficult to predict the functional effects of SNPs and other sequence variations in non-coding DNA because these variations often affect gene expression rather than protein structure (Knight 2003). Variations in non-coding regions can be found in the promoter region upstream of genes, in introns and between genes. A single base change in the promoter regions of genes may affect gene expression by affecting the binding of the RNA polymerase to the promoter and hence the initiation of transcription (Orphanides and Reinberg 2002; Tournamille *et al.* 1995; Udalova *et al.* 2000). Such mutations may cause disease by resulting in the production of a non-functional protein or a low level of functional protein (Orphanides and Reinberg 2002; Tournamille *et al.* 1995; Udalova *et al.* 2000). For example, a single nucleotide polymorphism located within intron 4 of the human growth hormone 1 (*GH1*) gene has been associated with lowered expression of the *GH1* protein and reduced secretion of human growth hormone which may increase the risk of osteoporosis while decreasing the risk of colorectal cancer (Dennison *et al.* 2004; Le Marchand *et al.* 2002; Millar *et al.* 2010).

1.1.9 Regulation of gene expression without alteration in DNA sequence (epigenetics)

Epigenetics is the study of heritable changes in gene expression that are caused by DNA methylation and post-translational changes to histone proteins, and are not caused by alternations in DNA sequence (Holliday 1987). DNA methylation generally acts to negatively-regulate gene expression (transcription repression), and involves the covalent addition of a methyl group to the 5' position of cytosines inside the CpG islands (short CpG dinucleotide regions) of promoters and first exons of genes (Brenet *et al.* 2011; Jorda and Peinado 2010). These CpG dinucleotides are often unmethylated but may become methylated through tightly-controlled processes that regulate gene transcription, including genomic imprinting and X-chromosome inactivation (Bird 1986). In contrast, CpG dinucleotides located outside CpG islands (for example, in repetitive DNA) can be methylated without affecting gene expression (Bird 2002; Brenet *et al.* 2011). An example of a disease influenced by DNA methylation is Rett syndrome, an X-linked neurodegenerative disease caused by disruption of MeCP2 protein binding to methylated DNA (Amir *et al.* 1999).

Although transcriptional silencing is linked with promoter methylation, a recent study has found that DNA methylation surrounding the transcription start site (TSS) in the first exon played a more significant role in repressing transcription (Brenet *et al.* 2011). DNA methylation of coding regions may prevent initiation of gene transcription and results in production of a non-functional protein or a low level of functional gene product (Brenet *et al.* 2011).

1.1.10 Regulation of gene expression by transposable SINE elements

Short interspersed nuclear elements (SINEs) are retrotransposons that are distributed evenly throughout the dog genome, with a frequency 10-100 fold higher than in humans (Kirkness *et al.* 2003; Wang and Kirkness 2005). A significant number of SINE elements in the dog genome are inserted in locations where they can influence gene expression and cause disease or alter non-disease phenotypes (Cordaux and Batzer 2006; Wang and Kirkness 2005). The insertion of a SINE element upstream of a gene or exon can cause disease by disrupting open reading frames (ORFs), alternative splicing of mRNA transcripts and/or the level of gene expression (Cordaux and Batzer 2006).

Conversely, the insertion of SINE elements into an intron does not affect protein function but may have an effect on the expression level of the adjacent gene by causing aberrant splicing of mRNA transcripts (Cordaux and Batzer 2006). For example, the insertion of a SINE element in the *PTPLA* gene leads to multiple splicing defects resulting in centronuclear myopathy in the Labrador Retriever (Pele *et al.* 2005) while the insertion of a SINE element into intron 3 of the hypocretin receptor gene leads to narcolepsy in the Doberman Pinscher (Lin *et al.* 1999). A recent study show the insertion of a SINE element in the *SILV* gene is responsible for the merle phenotype (coat colour) in the Shetland Sheepdog and in other dog breeds (Clark *et al.* 2006).

1.1.11 Candidate genes

In this study, the 44 candidate genes (Section 1.1.7) were prioritised using BioGPS with gene ontology and expression profiles in the human, mouse and rat brain/cerebellum (Wu *et al.* 2009). There are six different pathogenetic mechanisms that are known to cause autosomal recessive cerebellar ataxias (ARCAs) in humans when

any of these mechanism malfunction due to mutation (Vermeer *et al.* 2011). These processes are mitochondrial oxidative stress, DNA repair maintenance, protein quality control and recycling, mislocalisation of synaptic myonuclei, altered function of calcium-activated chloride channels, and altered vesicular trafficking (Vermeer *et al.* 2011). Candidate genes were classified into three groups according to molecular functions and expression level in brain tissues (as measured in microarray analyses) (Table 1-2). Genes with molecular processes that are known to cause ARCAs in humans were chosen for further examination. Of these genes, three genes (discussed in Sections 1.1.11.1 to 1.1.11.3) were identified as preferred candidates for their previous links with cerebellar ataxia in mice, dogs and horses. Seventeen more genes were also chosen as candidates for this study based on their high expression level in brain tissues and their molecular functions (Table 2-6 to Table 2-7).

1.1.11.1 Autophagy related 10 homolog (*ATG10*)

Autophagy is the protein degradation pathways that recycled intracellular proteins by engulfed cytoplasm into a double-membrane vesicles, known as the autophagosome (Komatsu *et al.* 2006; Phillips *et al.* 2008). In mice, a loss of *ATG7* in the central nervous system of affected mice can lead to neurodegeneration with neuron cell death in the cerebellum cortex of brain. Hence functional ATG7 protein is essential for autophagy and maintaining survival of neuron cells in the brain (Komatsu *et al.* 2006). In order for a mutation in *ATG10* to be causing CA in Kelpies, it would presumably need to be a down-regulated mutation that causes ATG10 protein to be deficient in brain tissues and thus make neuron cells more susceptible to accumulation of aggregates in cells and early cell apoptosis.

1.1.11.2 Arylsulfatase B precursor (*ARSB*)

ARSB and *ARSG* are both sulfatase that catalyse the hydrolysis of sulphate esters in the presence of different substrates (Sardiello *et al.* 2005). A non-synonymous substitution in exon 2 of the *ARSG* gene leads to a 75% reduction in sulfatase activity and causes neuronal ceroid lipofuscinoses (NCLs) in American Staffordshire Terrier (ASTs) (Abitbol *et al.* 2010). ARSG protein is necessary for the long term survival of neuron cells (Abitbol *et al.* 2010). ASTs affected with NCLs usually experience ataxia as the main symptom, with severe loss of Purkinje cells in the cerebellum (Abitbol *et al.* 2010). The onset age of NCLs in ASTs is between 3 and 5 years with a variable disease penetrance (Abitbol *et al.* 2010). In order for a mutation in *ARSB* to be causing CA in Kelpies, it would presumably need to be a down-regulated mutation that causes ARSB protein to be deficient in brain tissues and thus make neuron cells more susceptible to accumulation of aggregates in cells and early cell apoptosis. Mutations in *ARSB* cause mucopolysaccharidosis type VI, an autosomal recessive lysosomal storage disease characterised by an accumulation of intracellular dermatan sulphate. Clinical symptoms for mucopolysaccharidosis type VI include abnormal growth, short stature, stiff joints and skeletal malformations (Matalon *et al.* 1974; O'Brien *et al.* 1974).

1.1.11.3 mutS homolog 3 (*MSH3*)

MUTYH and *MSH3* are both DNA repair genes that can recognise and repair single base pair mismatches in response to oxidative damage in nuclear and mitochondrial DNA (Brault *et al.* 2011; Charbonneau *et al.* 2009; Gu *et al.* 2002; Marra *et al.* 1998; Sampson *et al.* 2005; Sixma 2001). A regulatory SNP located 1200 bp upstream of the MUTYH gene causes Cerebellar Abiotrophy in Arabian horses (Brault *et al.* 2011). The MSH3 gene interacts with another DNA mismatch repair gene, *MSH2*, to form a complex that repairs base mismatches and small/ large insertion-

deletion loops (IDLs) (Charbonneau *et al.* 2009; Marra *et al.* 1998; Sixma 2001). Mutations in *hMSH3* are currently associated with sporadic colon cancer and microsatellite instability in humans (Orimo *et al.* 2000). However, there is a possibility that mutations in *MSH3* could lead to apoptosis of Purkinje cells by preventing formation of the MSH2/MSH3 DNA repair complex (Charbonneau *et al.* 2009).

Table 1-2: Genes in the Kelpie cerebellar abiotrophy candidate region showing location and known gene ontology/gene expression profile in brain and cerebellum.

Canine gene	Gene position (Mb on CFA3)	Gene name	Gene symbol (Human)	Gene ontology	Gene expression level	
					brain	cerebellum
LOC610097	28.052-28.067	Vacuolar ATP synthase subunit S1 precursor	*ATP6AP1*	ATP hydrolysis, proton transport	7/4	5/0
					20/8	5/0
					5/3	1/1
LOC479159	28.092-28.093	Ribosomal protein S23	*RPS23*	Translation, elongation and termination	1/2	1/0
					2/3	0/1
					-	-
LOC610103	28.107-28.307	APG10 autophagy 10-like	*ATG10*	Autophagy, protein transport	5/3	1/1
					4/4	0/1
					3/4	0/0
LOC479160	28.499-28.783	Single-stranded DNA-binding protein 2	*SSBP2*	Transcription regulation	8/3	4/1
					21/12	2/4
					4/5	0/0
LOC488925	28.811-28.854	Cytosolic acetyl-CoA hydrolase	*ACOT12*	ATP binding	2/1	0/1
					0/5	0/1
					-	-
LOC479161	28.856-28.864	Zinc finger, CCHC domain containing 9	*ZCCHC9*	Zinc ion and metal ion binding	1/1	1/0
					1/6	0/1
					4/4	0/0
LOC479163	28.894-28.909	Creatine kinase, sarcomeric mitochondrial precursor	*CKMT2*	ATP binding. Cellular mechanism	2/4	0/4
					5/12	1/5
					-	-
LOC610203	28.925-29.170	Ras protein-specific guanine nucleotide-releasing factor 2	*RASGRF2*	Apoptosis, protein binding	7/2	2/2
					18/15	0/5
					-	-
LOC479164	29.219-29.404	mutS homolog 3	*MSH3*	DNA repair	5/4	1/2
					9/12	1/0
					2/2	0/0
LOC479165	29.404-29.430	Dihydrofolate reductase	*DHFR*	NADP binding	6/3	2/2
					7/15	1/1
					-	-
LOC488927	29.483-29.485	Ankyrin repeat domain 34B	*ANKRD34B*	Unknown	-	-
					3/2	3/0
					5/5	0/0
LOC479166	29.499-29.527	AASA9217	*AASA9217*	Unknown	-	-
					-	-
					-	-
LOC479167	29.547-29.582	Zinc finger, FYVE domain containing 16 (KIAA0305)	*ZFYVE16*	Protein binding, protein transporter	8/0	2/0
					8/11	0/4
					3/3	-

LOC488928	29.632-29.634	Likely ortholog of mouse spermatogenic Zip 1	*SPZ1*	DNA binding	2/0	1/0
					1/4	0/1
					-	-
LOC479168	29.671-29.773	Developmentally regulated protein TPO1	*SERINC5*	Lipid metabolic process	3/0	3/0
					9/5	1/1
					-	-
LOC488929	29.816-29.819	Similar to WW domain binding protein 11	*ssWBP11*	rRNA/mRNA processing, RNA splicing	5/5	1/2
					17/13	0/2
					8/6	1/0
LOC488930	29.827-29.887	Thrombospondin 4 precursor	*THBS4*	Calcium ion binding	5/3	0/3
					9/13	0/5
					2/2	0/1
LOC488931	29.887-29.916	Metaxin 3 (hypothetical protein)	*MTX3*	Protein transport	2/1	0/1
					0/2	0/0
					-	-
LOC479170	30.364-30.119	Myospryn protein	*CMYA5*	Protein binding	1/2	0/2
					5/7	3/0
					-	-
LOC607764	30.156-30.218	PAP associated domain containing 4	*PAPD4*	mRNA processing	2/1	1/1
					-	-
					3/3	0/0
LOC488933	30.283-30.406	Homer protein homolog 1	*HOMER1*	Calcium ion transport, skeletal muscle fiber development	8/4	3/1
					24/14	1/4
					-	-
LOC610323	30.434-30.529	Junction-mediating and regulatory protein	*JMY*	Transcription, DNA repair	2/2	0/0
					10/8	1/2
					-	-
LOC479171	30.616-30.635	Betaine homocysteine methyl transferase	*BHMT*	Zinc ion binding, protein methylation	2/3	1/2
					-	-
					5/6	0/1
LOC488934	30.659-30.672	Betaine homocysteine methyl transferase 2	*BHMT2*	Metal ion binding, methyltransferase activity	2/5	0/3
					4/12	0/3
					-	-
LOC488935	30.673-30.739	Dimethylglycine dehydrogenase, mitochondrial precursor	*DMGDH*	Electron carrier activity, folic acid binding, aminomethyltransferase activity	1/2	0/1
					3/8	0/1
					5/3	0/0
LOC610364	30.753-30.916	Arylsulfatase B precursor	*ARSB*	Autophagy, metal ion binding	4/6	1/2
					8/7	1/2
					7/6	0/1
LOC610372	30.986-30.988	Copine-1	*CPNE1*	Calcium dependent phospholipid binding	1/2	1/1
					9/12	3/1
					-	-
LOC610427	31.146-31.163	Lipoma HMGIC fusion partner-like 2	*LHFPL2*	Unknown	3/4	0/4
					13/11	5/0
					5/7	1/0
LOC479172	31.187-	Secretory carrier	*SCAMP1*	Protein	8/3	3/1

	31.305	membrane protein 1		binding, protein transport	22/11	3/2
					10/8	1/1
AP3B1	31.370-31.636	Adaptor-related protein complex AP3 beta 1 subunit	*AP3B1*	Protein binding, protein targeting to lysosome, protein transport	5/4	3/2
					15/16	0/5
					7/7	0/0
LOC479173	31.880-31.913	Tubulin binding cofactor A	*TBCA*	Protein metabolism	1/0	1/0
					11/13	0/2
					-	-
LOC488938	31.952-31.959	Orthopedia	*OTP*	Sequence specific DNA binding, transcription factor activity	2/2	0/1
					4/8	0/1
					-	-
LOC610498	32.129-32.174	WD repeat domain 41	*WDR41*	Unknown	6/3	0/2
					5/4	0/1
					-	-
LOC488939	32.179-32.331	3,5 cyclic nucleotide phosphodiesterase 8B	*PDE8B*	Transcription regulation	9/6	1/4
					13/9	1/1
					3/3	0/0
LOC479175	32.466-32.472	Hypothetical	*ZBED3*	DNA and metal ion binding	4/2	0/2
					6/11	0/0
					4/4	0/0
LOC479176	32.535-32.570	Angiogenic factor VG5Q	*AGGF1*	Protein binding, RNA processing	9/4	1/2
					5/3	1/0
					-	-
LOC610584	32.614-32.649	Corticotrophin releasing hormone binding protein	*CRHBP*	Signal transduction	5/1	0/1
					8/7	0/2
					12/11	0/2
LOC610592	32.661-32.688	S100 calcium binding protein, zeta	*S100Z*	Calcium ion binding	1/0	0/0
					-	-
					-	-
LOC488940	32.688-32.697	Proteinase activated receptor 2 precursor	*F2RL1*	Receptor activity	3/7	1/4
					6/13	1/3
					2/1	0/0
LOC488941	32.719-32.734	Eukaryotic translation elongation factor 1 alpha 1	*EEF1A1*	Translation elongation, protein binding, gene expression	1/2	0/2
					0/1	0/0
					-	-
LOC488942	32.760-32.777	Coagulation factor II receptor, precursor	*F2R*	Positive regulation of transcription, negative regulation of neuron apoptosis	2/7	0/6
					14/20	3/4
					5/6	0/1
LOC479177	32.782-33.063	Ras GTPase-activating-like protein IQGAP2	*IQGAP2*	Protein binding, signal transduction, actin binding	1/4	0/2
					13/16	0/4
					-	-

LOC607963	32.855-32.862	Proteinase activated receptor 3 precursor	*F2RL2*	Protein binding, blood coagulation	1/4	0/2
					4/6	0/0
					1/1	0/0
LOC488943	33.120-33.274	Synaptic vesicle glycoprotein 2C	*SV2C*	Neurotransmitter transport, trans-membrane transport	12/3	1/1
					7/6	0/2
					6/7	0/0

All gene names (except *AP3B1*) are listed as "similar to" in the dog reference genome (NCBI canine Build 2.0), but has been omitted in the gene name here. Remainder of the genes were identified based on sequence homology in the DNA sequences in human and have not been confirmed by experimental studies.

Gene expression microarray data in the human, mouse and rat (in rows) brain/cerebellum tissues (in columns) respectively were collected from the BioGPS database where available. Number before the (/) mark indicates the number of over-expressed cases, with number after the (/) mark represents the number of under-expressed cases. Missing microarray data is represented as a dash.

1.1.12 Aims

The main objective of this project is to identify the mutation(s) that cause CA in Australian Kelpies.

1.1.12.1 Specific aims

- identify likely candidate genes by the functional candidate gene approach;
- study synonymous, regulatory, and conserved differences identified by the sequencing of multiple affected and unaffected dogs, and look for any association with the CA trait;
- identify new variants within exons and upstream of candidate genes by PCR and Sanger sequencing;
- increase the power of previous linkage analysis studies (Shearman *et al.* 2011) by typing further affected and unaffected individuals

1.1.12.2 Potential outcomes

Identification of the genetic basis of CA in Kelpies will provide a foundation for the development of a reliable genetic test for this disorder. The genetic test would allow breeders to identify carriers of CA and hence avoid the production of affected pups. The gene responsible for CA in Kelpies could also be used as a likely candidate for studying the genetic basis of similar diseases in humans.

2 MATERIALS AND METHODS

2.1 Materials

2.1.1 Collection of samples

Blood samples of CA affected Australian Kelpie dogs and relatives, and unaffected Australian Kelpie dogs were collected with the kind assistance of breeders and owners, and handled in accordance with the protocols set out by the Animal and Ethics Committee of the University of New South Wales. Each blood sample received was assigned an identification number in the order in which samples were received, and added to an in-house database of Australian Kelpie samples. The samples collected in previous studies consisted of 21 ataxia affected dogs (categorised based on clinical diagnosis), 26 obligate carriers, 25 unaffected controls and 172 relatives of affecteds with unknown status (Shearman et al. 2011). The samples collected during the current study included 9 ataxia affected dogs, 9 obligate carriers and 5 relatives of affecteds with unknown status. Parents and siblings of affecteds were collected where possible. Samples collected included three affecteds from one earlier study by Thomas and Robertson (Thomas and Robertson 1989). Pedigree information was also collected from breeders and dog owners when available (Figure 2-1).

Forty samples from informative pedigrees, comprising ataxia affecteds, siblings of affecteds, and carriers were genotyped for microsatellite loci for linkage analysis. In this study, a total of 49 samples were tested for synonymous and UTR differences (Table 2-1, column 4), 10 samples were tested for non-coding evolutionarily-conserved sequence variants (Table 2-1, column 5) and 13 affected samples were typed for the 10 variants previously linked with the disease haplotype (Table 2-1, column 6) (Shearman *et al.* 2011), and 9 samples comprising two affecteds (WK6025, WK6065), four

controls (WK6059, WK6074, WK6078, WK6088), two carriers (WK6061, WK6142) and one unknown status (WK6058) were used for sequencing of exons and DNA upstream of genes. The 10 SNPs previously linked with the disease haplotype were only typed in a small number of CA affected dogs because the samples for these individuals were received after the publication of Shearman *et al.* (2011), and it was necessary to ensure that they shared the same SNP haplotype as the CA affected individuals described in that study.

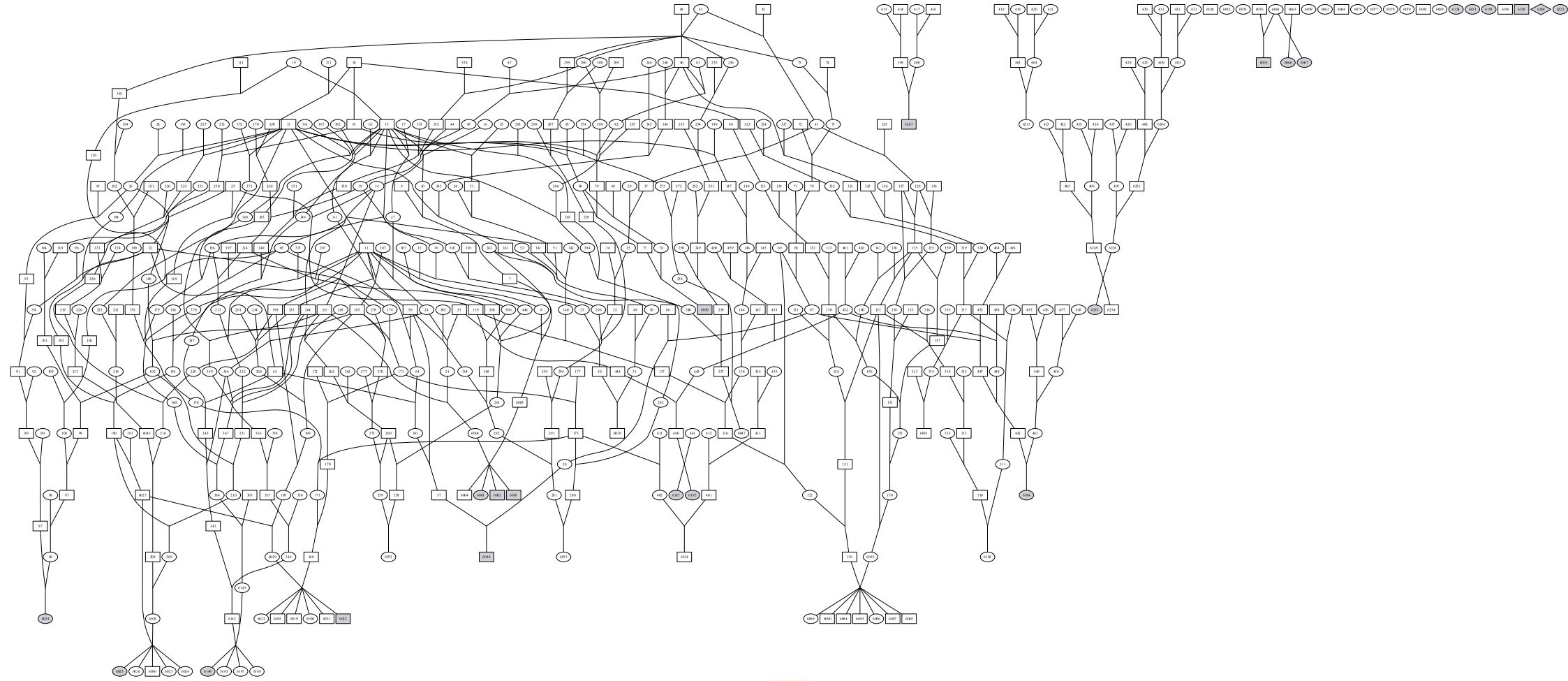
Table 2-1: Samples used for SNP typing.

Dog ID number	Affection status	Possessed the CA haplotype	Synonymous and UTR differences	Differences conserved between mammals	Selected SNPs on canine SNPs array v2.0
WK6001	Affected	**Yes**	**Tested**	Not tested	Typed (Shearman *et al.* 2011)
WK6003	Affected	**Yes**	**Tested**	Not tested	Typed (Shearman *et al.* 2011)
WK6022	Affected	**Yes**	**Tested**	Not tested	Typed (Shearman *et al.* 2011)
WK6025	Affected	**Yes**	**Tested**	**Tested**	Typed (Shearman *et al.* 2011)
WK6044	Affected	**Yes**	**Tested**	Not tested	Typed (Shearman *et al.* 2011)
WK6050	Affected	**Yes**	**Tested**	Not tested	Typed (Shearman *et al.* 2011)
WK6054	Affected	**Yes**	**Tested**	Not tested	Typed (Shearman *et al.* 2011)
WK6065	Affected	**Yes**	**Tested**	**Tested**	Typed (Shearman *et al.* 2011)
WK6066	Affected	**Yes**	**Tested**	Not tested	Typed (Shearman *et al.* 2011)
WK6067	Affected	**Yes**	**Tested**	Not tested	Typed (Shearman *et al.* 2011)
WK6124	Affected	**Yes**	**Tested**	Not tested	Typed (Shearman *et al.* 2011)
WK6149	Affected	**Yes**	**Tested**	**Tested**	Typed (Shearman *et al.* 2011)
WK6181	Affected	**Yes**	**Tested**	Not tested	**Typed in this study**
WK6182	Affected	**Yes**	**Tested**	Not tested	**Typed in this study**
WK6183	Affected	**Yes**	**Tested**	Not tested	**Typed in this study**
WK6184	Affected	**Yes**	**Tested**	Not tested	**Typed in this study**
WK6195	Affected	**Yes**	**Tested**	Not tested	**Typed in this study**
WK6203	Affected	**Yes**	**Tested**	**Tested**	**Typed in this study**
WK6204	Affected	**Yes**	**Tested**	Not tested	**Typed in this study**
WK6212	Affected	**Yes**	**Tested**	Not tested	**Typed in this study**
WK6222	Affected	**Yes**	**Tested**	Not tested	**Typed in this study**
WK6253	Affected	**Yes**	**Tested**	Not tested	**Typed in this study**
WK6254	Suspected affected	No	**Tested**	Not tested	**Typed in this study**
WK6263	Affected	**Yes**	**Tested**	Not tested	**Typed in this study**
WK6266	Affected	**Yes**	**Tested**	Not tested	**Typed in this**

					study
WK6018	Unaffected control	**Yes**	**Tested**	**Tested**	Typed (Shearman *et al.* 2011)
WK6043	Unaffected control	No	**Tested**	Not tested	Typed (Shearman *et al.* 2011)
WK6047	Unaffected control	No	**Tested**	Not tested	Typed (Shearman *et al.* 2011)
WK6059	Unaffected control	**Yes**	**Tested**	**Tested**	Typed (Shearman *et al.* 2011)
WK6062	Unaffected control	No	**Tested**	Not tested	Typed (Shearman *et al.* 2011)
WK6064	Unaffected control	**Yes**	**Tested**	**Tested**	Typed (Shearman *et al.* 2011)
WK6072	Unaffected control	No	**Tested**	Not tested	Typed (Shearman *et al.* 2011)
WK6073	Unaffected control	No	**Tested**	Not tested	Typed (Shearman *et al.* 2011)
WK6074	Unaffected control	No	**Tested**	**Tested**	Typed (Shearman *et al.* 2011)
WK6077	Unaffected control	No	**Tested**	Not tested	Typed (Shearman *et al.* 2011)
WK6078	Unaffected control	No	**Tested**	**Tested**	Typed (Shearman *et al.* 2011)
WK6079	Unaffected control	No	**Tested**	**Tested**	Typed (Shearman *et al.* 2011)
WK6082	Unaffected control	No	**Tested**	Not tested	Typed (Shearman *et al.* 2011)
WK6088	Unaffected control	No	**Tested**	Not tested	Typed (Shearman *et al.* 2011)
WK6091	Unaffected control	No	**Tested**	Not tested	Typed (Shearman *et al.* 2011)
WK6099	Unaffected control	No	**Tested**	Not tested	Typed (Shearman *et al.* 2011)
WK6102	Unaffected control	**Yes**	**Tested**	Not tested	Typed (Shearman *et al.* 2011)
WK6052	Obligate carrier	No	**Tested**	Not tested	Typed (Shearman *et al.* 2011)
WK6053	Obligate carrier	No	**Tested**	Not tested	Typed (Shearman *et al.* 2011)
WK6145	Unaffected sibling	**Yes**	**Tested**	Not tested	Typed (Shearman *et al.* 2011)
WK6147	Unaffected sibling	**Yes**	**Tested**	Not tested	Typed (Shearman *et al.* 2011)
WK6148	Unaffected sibling	**Yes**	**Tested**	Not tested	Typed (Shearman *et al.* 2011)
WK6196	Unaffected sibling	**Yes**	**Tested**	Not tested	Not typed
WK6267	Unaffected sibling	**Yes**	**Tested**	Not tested	Not typed

Figure 2-1: Pedigree of all samples used for SNPs genotyping and linkage analysis with microsatellites as genetic markers (full pedigree fold-out).

Affected dogs are shaded, dogs with the 4 digits identification number represents received samples. Pedigree dogs (no DNA samples received) were given a 2 digits identification number. Relationship is shown where known. Full pedigree information is missing for samples in the top right corner.



2.2 Methods

2.2.1 DNA extraction from blood

DNA was extracted from blood samples by the standard salting out method (Miller *et al.* 1988). Up to 10 mL of frozen blood was thawed on ice and poured into one 50 mL universal tube containing 40 mL of cold red blood cell lysis buffer (0.32 M sucrose, 10 mM Tris-HCl pH 7.5, 5 mM $MgCl_2$, 1% Triton X-100). The tube was mixed by gentle semi-inversion and centrifuged at 4300 *g* for 20 min at 4°C. The supernatant was discarded and the pellet was resuspended in 1 mL of cold TE washing buffer (20 mM Tris, 5 mM EDTA pH 8.0). The tube was then filled with 45 mL of cold TE washing buffer, and the tube was vortexed briefly before centrifugation at 1900 *g* for 10 min at 4°C. The TE wash step was repeated once or twice until a pure white pellet was obtained. The pellet was next resuspended in 5 mL of cold TE washing buffer using a Pasteur pipette, before 150 µL of Proteinase K solution (10 mg/mL) and 150 µL of 10% SDS solution were added to the tube. The tube was mixed by gentle inversion and incubated in a 55°C water bath overnight.

After the overnight incubation, a further 100 µL of Proteinase K (10 mg/mL) was added and mixed and incubation was continued for another hour at 55°C to ensure complete digestion of any proteins present in the mixture. Saturated 5 M NaCl (2 mL) and 7 mL of chloroform:isoamyl alcohol (24:1) were then added and the contents of the tube were mixed by inversion several times. The tube was centrifuged at 1900 *g* for 15 min at room temperature to separate the chloroform, protein and aqueous phases into 3 different layers. The aqueous phase (containing DNA) was carefully removed from the tube and transferred to a new 50 mL universal tube. The DNA was precipitated from the solution by adding two volumes of 96% and mixing by gentle inversion. The

precipitated DNA was spooled onto a Pasteur pipette with sealed end, and washed in 70% ethanol followed by a brief wash in 96% ethanol. The DNA pellet was then air-dried and dissolved in 500 µL of TE buffer (10 mM Tris-HCl pH 7.6, 1 mM EDTA) at room temperature for two days. The stock DNA samples were stored at -20°C until further use.

2.2.2 DNA extraction from dried blood spots

Where eluted DNA was not available for genotyping, DNA was extracted from samples dried onto Whatman FTA cards[1] using the manufacturer's protocol. A 1.2 mm sample disc was punched from the blood soaked FTA card using a coring device, and washed in 200 µL of FTA Purification Reagent (Whatman) for 5 min at room temperature. The reagent was then removed and discarded. The FTA wash step was repeated two more times. The sample punch was then incubated in 200 µL of TE buffer (10 mM Tris-HCl pH 7.6, 1 mM EDTA) for 5 min at room temperature. The buffer was then removed and discarded. The TE wash step was repeated once more. The sample disc was then air-dried overnight at room temperature, and was used directly in the PCR reaction.

2.2.3 DNA quantification

DNA concentration and purity was measured using a NanoDrop Spectrophotometer (Thermo Scientific, Wilmington, DE, USA). An aliquot of each

[1] FTA card is an index-sized card that can be used to collect and store nucleic acids contained in blood. Each card contains chemicals that can denature proteins and protect DNA from nucleases and oxidative damage.

undiluted stock DNA sample was diluted with MilliQ water to a final concentration of 50 ng/µL and stored at -20°C until further use.

2.2.4 Primer design

Primers were designed to amplify 17 synonymous or UTR substitutions identified previously (Table 2-3) (Shearman 2011), 10 selected SNPs within the candidate region from the canine SNP array v2 (Table 2-4) (Karlsson *et al.* 2007; Shearman *et al.* 2011), an additional 22 non-coding sequence variants (SNPs and indels) that are evolutionarily conserved between mammals including in dogs (Table 2-5), 37 exons from 18 genes with poor 454 sequence coverage (Table 2-6), and non-coding regulatory sequences upstream of the 12 selected candidate genes (Table 2-7). Primers were designed using Primer 3 (Rozen and Skaletsky 2000), with the dog reference sequence Build v2.1 (Genbank. NC_006585.2). RepeatMasker was used to remove all the known repeats in the target sequence (Smit 1996-2010). The optimum primer length was set to 20 bp, with GC content between 45 and 70%, and melting temperature at the optimum of 60°C. Primers were checked against the dog genome to ensure they were specific to candidate region using BLAST (NCBI) (Altschul *et al.* 1990; Mount 2007). Primers were ordered from GeneWorks, Sigma Australia, or Integrated DNA Technologies.

2.2.5 DNA amplification by the polymerase chain reaction

Distinct PCR reaction mixes and cycling conditions were used to amplify SNPs, exons, and microsatellites. A 'step down' PCR program (Multi 55) was adopted for the amplification of most SNPs and non GC-rich exons. This program consisted of 95°C for 10 min; 4 cycles of 95°C for 15 sec, 68°C for 2 min; 5 cycles of 95°C for 15 sec, 64°C for 1 min 15 sec, 68°C for 2 min; 10 cycles of 95°C for 15 sec, 58°C for 1 min 15 sec,

68°C for 2 min; 9 cycles of 95°C for 15 sec, 55°C for 1 min 15 sec, 68°C for 2 min; 10 cycles of 89°C for 15 sec, 55°C for 1 min 15 sec, 68°C for 2 min; a final extension at 68°C for 10 min; and hold at 4°C. These multiple annealing temperatures allow several primer pairs (in distinct PCR reaction mixes) with slightly different annealing temperatures to be amplified simultaneously using one PCR condition, while ensuring primer specificity at the same time. This 'step down' PCR was performed in a 20 µL reaction containing 2 µL of 10X PCR buffer (Applied Biosystems), 2 µL of $MgCl_2$ (25 mM) solution (Applied Biosystems), 3 µL of forward primer (5 µM), 3 µL of reverse primer (5 µM), 0.16 µL of dNTPs (25 mM), 0.2 µL of AmpliTaq Gold DNA Polymerase (5 U/µL) (Applied Biosystems) and 2 µL of eluted genomic DNA template (50ng/µL).

Another 'step down' PCR program (Multi 52) was used for the amplification of some GC rich exons. This program consisted of 98°C for 2 min; 95°C for 10 min; 4 cycles of 95°C for 15 sec, 64°C for 1 min 15 sec; 5 cycles of 95°C for 15 sec, 58°C for 1 min 15 sec, 68°C for 1 min 5 sec; 10 cycles of 95°C for 15 sec, 55°C for 1 min 15 sec, 68°C for 1 min; 9 cycles of 95°C for 15 sec, 52°C for 1 min 15 sec, 68°C for 1 min; 10 cycles of 89°C for 15 sec, 52°C for 1 min 15 sec, 68°C for 1 min; a final extension at 68°C for 10 min; and hold at 4°C. This 'step down' PCR was performed in a 20 µL reaction containing 2 µL of 10X PCR buffer (Applied Biosystems), 2 µL of $MgCl_2$ (25 mM) solution (Applied Biosystems), 2.7 µL of GC rich solution (Applied Biosystems), 3 µL of forward primer (5 µM), 3 µL of reverse primer (5 µM), 0.16 µL of dNTPs (25 mM), 0.2 µL of AmpliTaq Gold DNA Polymerase (5 U/µL) (Applied Biosystems) and 2 µL of eluted genomic DNA template (50ng/µL).

Some GC rich sequences were also amplified by the above 'step down' PCR program (Multi 55) except with the initial denaturation condition at 94°C for 2 min since a hot start Taq was not being used. The PCR reaction mix for amplification of GC rich sequences contained 2 µL of 10X Mg^{2+}-free DyNAzyme™ EXT Buffer (Finnzymes), 1 µL of $MgCl_2$ (50 mM) solution (Finnzymes), 1 µL DMSO (Finnzymes), 2 µL of forward primer (5 µM), 2 µL of reverse primer (5 µM), 0.16 µL of dNTPs (25 mM), 0.4 µL of DyNAzyme™ EXT DNA Polymerase (1 U/µL) (Finnzymes), and 2 µL of eluted genomic DNA template (50ng/µL), in a 20 µL reaction.

A range of different cycling conditions and distinct PCR reaction mixes (detailed in Table 2-2) were used to amplify other GC rich sequences that failed to give a PCR product with the 'step-down' PCR programs described above.

Table 2-2: PCR protocols for amplification of GC rich DNA.

PCR program for GC rich amplification		PCR components and cycling conditions
1	Touchdown at 55°C	1 µL of 10X Mg^{2+}-free DyNAzyme™ EXT Buffer (Finnzymes), 0.5 µL DMSO (Finnzymes), 1 µL of forward primer (5 µM), 1 µL of reverse primer (5 µM), 0.08 µL of dNTPs (25 mM), 0.2 µL of DyNAzyme™ EXT DNA Polymerase (1 U/µL) (Finnzymes), and 2 µL of eluted genomic DNA template (50ng/µL), in a 10 µL reaction. Initial denaturation at 94°C for 1.5 min; 20 cycles of 94°C for 30 sec, 65°C for 30 sec (-0.5°C/cycle), 72°C for 1 min; 20 cycles of 94°C for 30 sec, 55°C for 30 sec, 72°C for 1 min; a final extension at 72°C for 10 min; and hold at 4°C.
2	FastStart PCR at 60°C or 63°C, for 30/35 cycles	3 µL of GC-rich Solution (Roche Applied Science), 1.5 µL of forward primer (5 µM), 1.5 µL of reverse primer (5 µM), 0.12 µL of dNTPs (25 mM), 2 µL of eluted genomic DNA template (50ng/µL), and one of the following $MgCl_2$/Taq enzyme mix, in a 15 µL reaction. *Mix 1:* **2 mM $MgCl_2$ with 1 U/µL enzyme:** 1.5 µL of 10X PCR Buffer with $MgCl_2$ (Roche Applied Science), 0.06 µL of FastStart Taq DNA Polymerase (5 U/µL) (Roche Applied Science) *Mix 2:* **1 mM $MgCl_2$ with 2 U/µL enzyme:** 1.5 µL of 10X PCR Buffer <u>without</u> $MgCl_2$ (Roche Applied Science), 0.6 µL of $MgCl_2$ solution (25 mM) (Roche Applied Science), 0.12 µL of FastStart Taq DNA Polymerase (5 U/µL) (Roche Applied Science) *Mix 3:* **2 mM $MgCl_2$ with 2 U/µL enzyme:** 1.5 µL of 10X PCR Buffer with $MgCl_2$ (Roche Applied Science), 0.12 µL of FastStart Taq DNA Polymerase (5 U/µL) (Roche Applied Science) Initial denaturation at 95°C for 6 min; 15 cycles of 95°C for 30 sec, 60/63°C for 30 sec, 72°C for 1 min; 15/20 cycles of 95°C for 30 sec, 60/63°C for 30 sec, 72°C for 1 min (+5sec/cycle); a final extension at 72°C for 7 min; and hold at 4°C.

2.2.6 PCR optimisation

PCR reactions that failed to produce a product of the expected size under standard PCR conditions were optimised for variables including annealing temperatures, Taq DNA polymerase concentrations, initial denaturation time, extension time, number of cycles, primer concentrations and $MgCl_2$ concentrations. PCRs for amplification of GC-rich sequences were optimised with PCR additives such as the 5% DMSO or GC rich buffer (Roche Applied Science) in the PCR reaction mix. Annealing temperatures were optimised by performing temperature gradient PCR with temperature between 50°C and 62°C, while the magnesium concentrations were optimised by testing several $MgCl_2$ concentrations between 1 mM and 2.5 mM in the PCR reaction. Primer concentrations were optimised by trying several primer concentrations between 0.5 µM and 0.75 µM per PCR reaction. Other variables such as the Taq DNA polymerase concentrations, initial denaturation time, extension time, and cycle number were optimised accordingly with respect to the gel electrophoresis results.

Table 2-3: Primers used for PCR and sequencing of synonymous and UTR substitutions in the candidate region.

Gene name	Control	Affected	Location (bp) [£]	Forward primer	Reverse primer	Product size (bp)
ATP6AP1	C	T	28056759	TTGACAAGAGAGGGTGCAA	CAAATATACAGGTGCACTGAGC	381
RPS23	G	C	28092993	GGCTCGACAGCTTTGTTTCT	TGACCTTTACGACCAAATCC	492
CKMT2	T	C	28903436	AGCTGCGTCTCATCTTGTCA	GTACAGGAATCAAAGGGGACTG	562
ZFYVE16	C	T	29567856	ATTGGAGTACCTTGCAGTCTGG	AGGTACAGCTAGGAGCCCATAA	697
SPZ1	C	T	29632686	AACAGAAGCTATCAGCAGAGCA	CGTCACTACTTCCCACACTTCA	655
WBP1	A	G	29818867	GATCCCACCCTCGAAAGG	CCTCAGGATATCCGTCATCTTC	976
THBS4	C	T	29837088	CCAGAATCCATCCTTTGATGTC	CTCAGACTTCCTGTAGCAGTCC	497
PAPD4	A	G	30173203	TGGATACTTTCTGGCATACTCC	CCAGTCCTACTATGGAAACAGG	584
BHMT[¤]	G	A	30616120	TGGGAAAATCTTCGGATAGC	GGGCCCAGTTTTAAGCAG	994
BHMT[¤]	A	G	30616445	''	''	
BHMT[¤]	G	C	30616525	''	''	
BHMT[¤]	G	A	30616716	''	''	
DMGDH	A	G	30691196	CCTATCAATGCCAGTTGTGGTA	AGAGAAGACGAGACGGATGG	374
DMGDH	C	T	30704150	TTTTCGCCGCACAAACTG	GGAGTTCCACCATTCATACAGG	297
DMGDH[§]	T	C	30738644	AGTTTACTTCCTGGCGTTAGCC	CACTTAGGAAAGAGCTGAGAAAGG	525
DMGDH[§]	C	T	30738713	''	''	
ARSB	A	G	30917504	CATCTTTGCGTGGCTTCCT	CGACGTCAGCCTCTTCAGT	569

[£] Location on chromosome 3 (NCBI – dog reference genome build 2.1)

[¤] All four variants were captured with a single set of primers

[§] Both variants were captured with a single set of primers

Table 2-4: Primers used for PCR and sequencing of SNPs previously linked with the disease haplotype (Shearman *et al.* 2011).

Name [€]	Control	Forward primer	Reverse primer	Product size (bp)
Chr3.28354775	T	TCCCATTAGCTTGATATCATTGTGT	AGTGATTCCTCTCCGCTCTG	428
Chr3.29949728	C	TACCAATGGGGTCATCAACA	GATCCTAGGGAGTGCCAAAA	684
Chr3.30059175	A	AGTAACGTCCTCCACCGTGT	CAGTGCTAGGAGCCTTTTCTG	484
Chr3.30364362	T	TCAAGAAGCACTCGACCTCA	ACCCTCAAACTAGCTGACTGTT	469
Chr3.30597963	G	GCATAGCTACCTCTGCATGG	TGTCGTGTCTGGTACTGAAAGAG	485
Chr3.30684786	G	CCTGGCAGGATATAATTACACTGA	ATGATTGCGAATCTGCACAC	594
Chr3.30782110	T	CCTACAGCCCTAGGCATCTT	GCTGCTAAACGCTCCAATTT	608
Chr3.31163217	T	CTACGCATCGGCCTACAAAC	ACGGCTGCCTTAATGGTTTA	490
Chr3.32292529	A	GCTAAAGACTGCCCCCACTA	TCCTAGCTGAACCCCAGAGA	673
Chr3.33061677	C	GTTAGGCATCCATGCCAGTA	GCAAAGGTGAAAAGACATTCCT	655

[€] Primers were named based on chromosome number and genomic position in bp (e.g. C3.28354775 is on chromosome 3 at 28354775 bp).

Table 2-5: Primers used for PCR and sequencing of sequence variants evolutionally conserved in mammals and in dogs.

Name	Control	Affected	Forward primer	Reverse primer	Product size (bp)	Source
Chr3.28273654	AA	G	AGGGAAGTTAGAACATGGGTCA	GCTCTTAGCCACTTATCAGACCA	666	Intron (*ATG10*)
Chr3.29828219	C	G	CTGAAGTATCGCTGCAATGG	AAAAGGAAGCCACCGTGTAT	711	Intergenic
Chr3.29957695	-	C	GACACCGACACTTGCTTGAA	GGGTAACCGCAGAGGGAAC	487	Intron (*MTX3*)
Chr3.29984289	CTT	G	CTGTGAACCCCTCTCTCGTT	TGTTAGGTACACCCCCATCTG	483	Intergenic
Chr3.30085423	C	T	CACAAAAGCCCTCGGTGT	GGTGAAGACCCTCAATGGAT	354	Intron (*CMYA5*)
Chr3.30705732	A	G	TTGCCAAGACAATCGTTCAC	TCATATCCACCGTTGACTGC	797	Intron (*DMGDH*)
Chr3.30727425	T	C	CATCGCAATATCGTGCCTTA	ATCCCTTAAAAGCGCAACC	827	Intron (*DMGDH*)
Chr3.30829090 Chr3.30829251	A G	G A	TCATCTGTCCCACTAACATTGC	CGGGTATTTCGAGAAGTTGG	828	Intron (*ARSB*)
Chr3.30901375	G	A	ACTTAGGGCGAACCCTCAAC	AAATCCTTGCAGCACGAAAC	493	Intron (*ARSB*)
Chr3.30917959 Chr3.30917967	T G	C A	CTGGGAACTCGGCTTGAAA	AGTTTCAGGGTTAGAGGCGAAG	572	Intergenic Intergenic
Chr3.30949002	TCTT	-	ATGGCTTAGGCACCACCTAATA	GTCTCCGCGATCCTTTTACC	798	Intergenic
Chr3.31075908	T	C	TGAACGCAGAGACGACCAT	TAGGCCATCTAAAGCCAGCA	779	Intergenic
Chr3.31358376	T	A	GCTATGTCCTTAGCTTACCCACA	GCAGGCCTTACAAGAGTAGGAA	679	Intergenic
Chr3.31674050	CT	-	CGATGCGAGTTTGACTTTTC	CTCCCCGGGAAGTAGGAA	700	Intergenic
Chr3.31723991	G	A	ATTTCCAGCACAGGGAGGAT	AACCATAGGACCGGCAGTTA	936	Intergenic
Chr3.31752468	C	T	GGCAGGTCAGTCACAGATGA	TTGGTATCGATGGCACCTCT	583	Intergenic
Chr3.31758102	T	C	GGGTTGCAGGGTTCAAATAA	TGTTCACACGCACCGATT	970	Intergenic
Chr3.31960132	C	A	TTCGTCTTAAAGGAGGACAAGC	TCTTCGAAAGCGCTTCTGTT	682	Intergenic
Chr3.32500348	G	A	AAACTAGAGGCTATCTGGCAGTC	GTGACATTCATCACGTGGAA	420	Intergenic
Chr3.33001932	A	C	TGTTGTCCTCACTGTACGAACT	AGAGATGATGACCTACCTATCTGG	666	Intron (*IQGAP2*)

Table 2-6: Primers used for PCR and sequencing of exons of candidate genes.

Gene name	Exon	Forward primer	Reverse primer	Product size (bp)	PCR program
SCAMP1	1	AGAGCTGCACCCGGATCT	CAGCCCGTGCCAGTAAGG	484	Not optimised
HOMER1	8	CTCTTTCCATGAGCAAGCCATA	ACAGCACTGCTTAGAAACCGTA	955	Multi 55 + AmpliTaq gold DNA polymerase
	10	TCTGAGTCGTGAACTGACCTGT	AGAGTGGTCACCGAGTGAAA	876	Multi 55 + AmpliTaq gold DNA polymerase
RASGRF2	1	CGGGTGTGAGCGTGCAAA	GGACGGGGGTACGTGGAC	991	Not optimised
ARSB	1	CCGCCAGTTCCTGATTCTACC	GCTCCTCGGGGCCTCTCC	827	Mix 1 at 60°C for 30 cycles (see Table 2-2)
AP3B1	1	GGGTACAACCGGTTACTGCT	GGAGGTGTCCGTTCACAACT	783	Multi 55 + DyNAzyme EXT DNA polymerase
	3	GATTTCAGCTTGTTAACATCATAGC	AAAATCTAGTGATACTTACGGATCTTG	566	Multi 55 + AmpliTaq gold DNA polymerase
	6	GGGAGTAAAGCGTCCCATTT	CCAACTTGTTCTTCTTCAGACTCC	658	Not optimised
	23	CCAGATATGTGTACCTGGGAAA CCAATGTGCAACAGGTCAAT	TTGCATGCCCATAGGAAG GCTGGTATGTGACTCACTGAGAA	540 846	Multi 55 + AmpliTaq gold DNA polymerase
F2RL1	1	GAGGCTCGGGTTTCACTG	GAACGAGAGGCCCAAGAAG	635	Multi 55 + DyNAzyme EXT DNA polymerase
	2	CATTTCCCTGCGAAGTTTG	AGCCGCCTGAATCAGAGA	595	Multi 55 + AmpliTaq gold DNA polymerase
SERINC5	1	GGTGCCTGGCTTTCCAAC	CCGCACGACAGAGGAGAC	561	Touchdown at 55°C (see Table 2-2)
ANKRD34B	5'UTR	CCCCTCTCAGTATGAGTTCCAC	AAGCGTGCATCAAAGCAGTT	888	Multi 55 + AmpliTaq gold DNA polymerase
PAPD4	2	GGGGGACTTAGCGATGTG	AGTTGAGGTCTGTATTGGACTGAC	607	Multi 55 + AmpliTaq gold DNA polymerase
	9	GATCAGGAATGGTTTTGGTCAC	AAGTACCACGACTGGCATCA	847	Multi 55 + AmpliTaq gold DNA polymerase

	5'UTR	TGAGTGAGTTAGCGGCCTTC	GCGTTTCCTTTCCGTTCA	566	Multi 55 + DyNAzyme EXT DNA polymerase
BHMT	3'UTR	TCGAAAGACTAGAGAGTCCTGCT	CAACCAAGGTAGTCTCACTGCT	664	Multi 55 + DyNAzyme EXT DNA polymerase
SSBP2	1	CCAGCGTTGGGGTTTAAAAT	GCTTTGAGCGGAGGAGAG	729	Not optimised
	3	GATGTGCCTTAAGTTCCCAAAG	CTGGGCCTTCAATTAGTCTG	917	Multi 55 + AmpliTaq gold DNA polymerase
	4	CTGGCAGTTGGACGACAAAA	ACGGCATTTGGACTAACCTC	719	Multi 55 + AmpliTaq gold DNA polymerase
	7	GGGAAGGTATTATGCTTCTTGG	ATCACCACCCTCACTATCATGT	544	Multi 55 + AmpliTaq gold DNA polymerase
	10	TGGGATACTGGTGGCACTTTTA	CAGGGCTTTCAAAAGCAACA	756	Multi 55 + AmpliTaq gold DNA polymerase
	15	GATGATTGGGAGCTAGAGATGC	GGAGACACAAGCTATTGGAAGG	810	Multi 55 + AmpliTaq gold DNA polymerase
BHMT2	1	TTCTCCCGGACACACTGTAA	TGTGAAAGCGGGATAATGTG	973	Not optimised
ATG10	1	TCCATTCTACGCCCCAAA	AATGGTTTGAAGCACCACAG	535	Multi 55 + AmpliTaq gold DNA polymerase
ZFYVE16	4	CTGCCATGCTCCAGTCTACA	GGAAGCTCACAGCATTGAAA	612	Multi 55 + AmpliTaq gold DNA polymerase
	9	CTGTAATTTTCCAGCCATTATCA	ATGTGAAAACACCGCAGCAG	790	Multi 55 + AmpliTaq gold DNA polymerase
	10	TTTCTATTGTGGAGTAGTTGGAAGG	CAAGTCCATTAGAAAACAAATCTAAAA	586	Multi 55 + AmpliTaq gold DNA polymerase
	15	TTTCTTGGCGCTAACTTTGG	TGTCAAGCAGGATTACTACAGTG	706	Multi 55 + AmpliTaq gold DNA polymerase
ACOT12	1	CGGGGAGCGGGATTCGAG	CACCACACGTCCCCTTGC	583	Touchdown at 55°C (see Table 2-2)
AGGF1	1	GACAGGAAATCTCCGCGTTT	AACGTCACCAGCGAGAGTG	629	Touchdown at 55°C (see Table 2-2)
	9	TTCTATGAGAGAGAAGTTATCTGTGG	GCACCTTGGTGTATGTTAGTGA	589	Multi 55 + AmpliTaq gold DNA polymerase

	3'UTR	GAAAGTACCAGGTTGGATACAGCA	ATGCACCACTATCCCTGTTTTC	753	Multi 55 + AmpliTaq gold DNA polymerase
DHFR	5	TGCTCTAAGCAGCTAAAGGTTG	TCCTTGATGGGCAACACAGT	722	Multi 55 + AmpliTaq gold DNA polymerase
	3'UTR	ACCCAGGTGTGCTTTCTGAT	TCTTCAGATACCTAACAATCCACAT	884	Multi 55 + AmpliTaq gold DNA polymerase
DMGDH	1	GCCCAGAAACTCACCTTTTTG	ACGGCGAACGTTCTTCTCTA	834	Touchdown at 55°C (see Table 2-2)
	16	TCTCTTGGCATGAAAGTGAGC	TCCCTTGTTCAACTACTGCTTCTC	937	Multi 55 + AmpliTaq gold DNA polymerase

Table 2-7: Primers used for PCR and sequencing of regulatory sequences upstream of candidate genes.

Gene name	Genomic position (bp) [Ψ]	Forward primer	Reverse primer	Product size (bp)	PCR program
SCAMP1	31305743-31304943	AGGTGCCCTCCGGTTAAGT	AGATCCGGGTGCAGCTCTC	902	Mix 1 at 60°C for 30 cycles (see Table 2-2)
HOMER1	30282089-30282889	GGTCTCCCCGCCCACTCC	GCGACGTGCGTGCACTTTA	842	Not optimised
RASGRF2	29170307-29169507	GGGAGTTGTGAGTTGGTAGGAA	GTTGTAGCGCACGCTCTTCT	961	Not optimised
ARSB	30751953-30752753	GAATGTGCTGGTCCTAAGGCTAA	GTCCGCCAGCACGAAGAC	998	Mix 3 at 63°C for 35 cycles (see Table 2-2)
AP3B1	31368712-31369512	TGGTCTTTGGAGGAGTGGA	GCTCACTGTAAGCGAAGCTG	972	Mix 3 at 60°C for 35 cycles (see Table 2-2)
SERINC5	29670360-29671160	GGCTTGGCGCTAAGTTCT	GCGATCTCCCCTTTCTGG	906	Mix 1 at 60°C for 35 cycles (see Table 2-2)
SSBP2	28498825-28499625	GCCGTCCAGCGAGGAGGT	GGACGGCGCTGCTGTTAC	991	Not optimised
BHMT2	30673277-30672477	TCCATGAAAAGCCACAATCA	GGGGAGGCTGAATACGTC	985	Mix 2 at 60°C for 35 cycles (see Table 2-2)
AGGF1	32570516-32569716	GAAGTCTGACTCCTCCCTCGT	CAGCTCGCCTTCCAACTTCT	831	Touchdown at 55°C (see Table 2-2)
MSH3	29405211-29404411	ACAGAGCTTTGGGTGTCAATAAG G	GGGACCCTCCGTGTGACG	972	Not optimised
DHFR	29400497-29401436	GCTGTTTGACTTGCAGGAATG	GCCCAAGGACCTGTATCTGT	939	Multi 52 + AmpliTaq gold DNA polymerase
F2R	32777927-32777127	TTTCGGTGGCTGCTCAAT	GACGCTGCAGCCCTTCCT	939	Multi 52 + AmpliTaq gold DNA polymerase

[Ψ] Location on chromosome 3 (NCBI – dog reference genome build 2.1)

2.2.7 Gel electrophoresis of PCR products

PCR products were separated on a 2% (w/v) agarose gel (containing 0.03% ethidium bromide) by electrophoresis at 180V, in 1X Sodium Boric Acid Buffer (10 mM NaOH, adjusted to pH 8 with Boric Acid). For each sample, 4 µL of PCR product was mixed with 2 µL of 1X gel loading dye (New England BioLabs) and loaded into the corresponding well in the gel. 0.5 µL of the 100 bp size standard (New England BioLabs) was also mixed with 2 µL of 1X gel loading dye (New England BioLabs) and loaded into a separate well. The gel was then visualised using a GelDoc (Bio-Rad) UV transilluminator system and photographed.

2.2.8 Sequencing

Single PCR products were purified using ExoSAP-IT (Affymetrix). 5 or 10 µL of PCR product was added to 2 or 4 µL of ExoSAP-IT enzyme, in a total volume of either 7 or 14 µL respectively. The ExoSAP-IT reaction was incubated at 37°C for 30 min, followed by 80°C for 15 min, and held at 10°C when the reaction was completed.

The purified PCR product (7 µL) was combined with 1 µL of either the forward or reverse primer (5 µM), 1 µL of ABI BigDye terminator v3.1 (Applied Biosystems), 2 µL of 5X sequencing buffer (Applied Biosystems), and 9 µL of MilliQ water, making a total of 20 µL per sequencing reaction. Sequencing was performed using a standard cycling protocol consisted of 96°C for 5 sec; 25 cycles of 96°C for 10 sec, 50°C for 5 sec, 60°C for 4 min; and were then held at 11°C.

PCR product was precipitated in 60 µL of 95% ethanol and 5 µL of 125 mM EDTA. The tube was then vortexed briefly and incubated at room temperature for 15 min before centrifugation at 20,000 *g* for 20 min. The supernatant was removed and

discarded. The pellet was then washed in 70% ethanol (150 µL) to remove any remaining salt in the solution. The tube was then vortexed briefly and centrifuged at 20,000 *g* for 10 min at 4°C. The supernatant was removed and discarded as before. DNA pellet was either air-dried or dried by spinning in a DNA110 Speed Vac® (Sevant) at medium speed for 7 min.

Samples were processed on an ABI3730 Capillary Sequencer (Applied Biosystems) at The Ramaciotti Centre for Gene Function Analysis, UNSW.

2.2.9 Sequence alignments

Sequence data were aligned to the dog reference sequence Build v2.1 (Genbank. NC_006585.2) using the SeqScape software (Applied Biosystems). Sequences from affecteds, unaffected siblings of affecteds, obligate carriers, and unaffected controls were compared and any differences were noted.

2.2.10 Pedigrees

Three additional pedigree families of 17 Kelpies that included 3 affected Kelpies, 1 suspected affected Kelpie and 5 unaffected siblings to affecteds were collected for use in locating the CA disease gene in Australian Kelpies (Figure 2-2 to Figure 2-4). All affected Kelpies displayed clinical symptoms as described in Section 1.3. Many of the affected animals can be traced back to common ancestors (Figure 2-1). These three new pedigree families, together with the initial 4 pedigree families (Shearman *et al.* 2011) were included for the linkage analysis (Section 2.2.12).

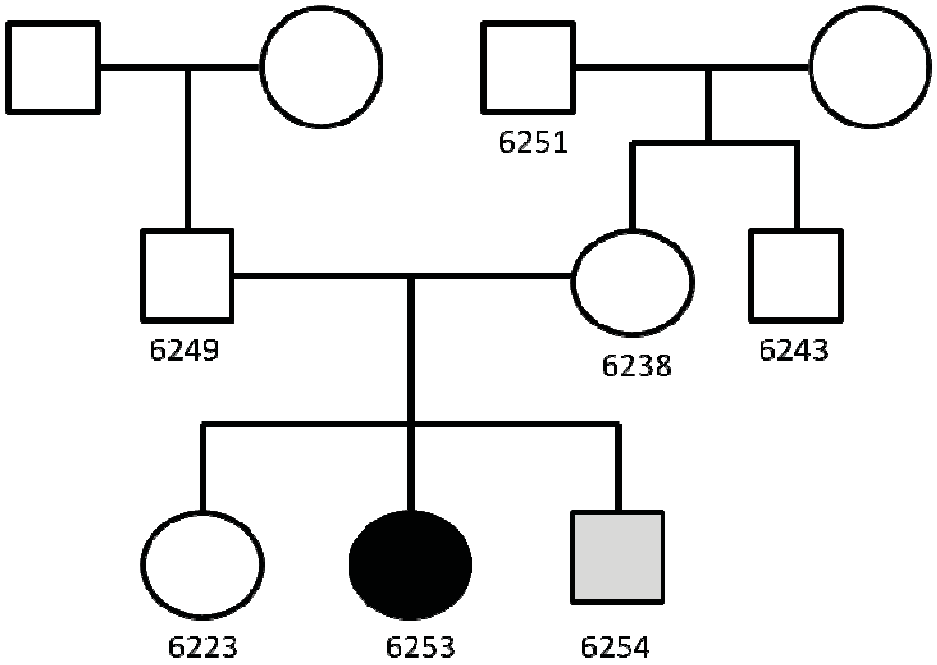


Figure 2-2: Australian Kelpie CA pedigree 1.

The following symbols are used: square (males), circles (females), filled (affected), unfilled (unaffected), grey shaded (suspected affected). The sample numbers are shown below individuals involved in the study where samples were collected. Obligate carriers in this pedigree are 6249 and 6238.

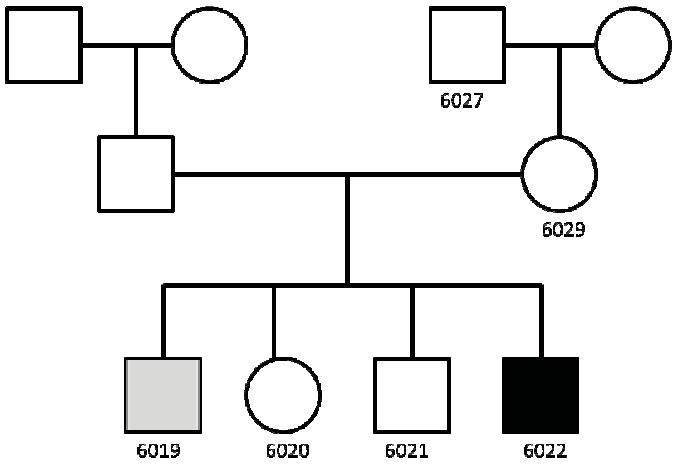


Figure 2-3: Australian Kelpie CA pedigree 2.

The following symbols are used: square (males), circles (females), filled (affected), unfilled (unaffected), grey shaded (suspected affected). The sample numbers are shown below individuals involved in the study where samples were collected. The obligate carrier in this pedigree is 6029.

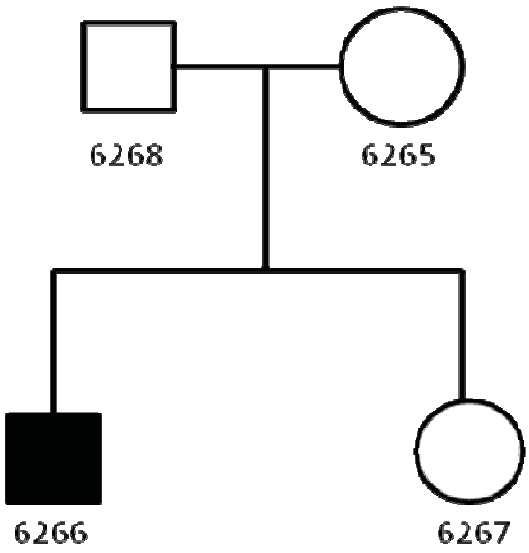


Figure 2-4: Australian Kelpie CA pedigree 3.

The following symbols are used: square (males), circles (females), filled (affected), unfilled (unaffected). The sample numbers are shown below the individual. Obligate carriers in this pedigree are 6265 and 6268.

2.2.11 Autozygosity analysis and linkage analysis by microsatellites

DNA of CA affecteds and their families were typed for both the short (20-30 di-nucleotide repeats, low variation rate) and long (>30 tetra-nucleotide repeats, high mutation rate) microsatellites using a standard PCR reaction. The 10 µL reaction consisted of 1 µL of 10X PCR buffer (Applied Biosystems), 1 µL (or 0.8 µL) of 25 mM $MgCl_2$ solution (Applied Biosystems), 0.4 µL of forward primer (2 µM), 0.4 µL of reverse primer (20 µM), 0.4 µL of fluorescent labelled forward M13 tag (20 µM), 0.08 µL of dNTPs (100 mM), 0.1 µL of AmpliTaq Gold DNA Polymerase (5 U/µL) (Applied Biosystems) and 2 µL of genomic DNA template (50 ng/µL). PCR products were fluorescently labelled for fragment analysis using the universal priming method (Neilan *et al.* 1997). PCR was performed using primers previously designed by Shearman (Shearman 2011). The PCR conditions and the type of fluorescent labelled M13 tag used for the amplification of each of the 6 short microsatellites and 5 long microsatellites are shown in Table 2-8 and Table 2-9 below.

Table 2-8: Annealing temperature, $MgCl_2$ concentration, and M13 labelled tags used for the amplification of short microsatellites.

Microsatellite name ¥	M13 labelled tag	Group number	Annealing temperature (°C)	Magnesium concentration (mM)
C3.2805	6-FAM	1	61	2.5
C3.2856	NED	2	52	2.5
C3.2869	6-FAM	2	55	2.5
C3.2984	VIC	2	55	2.5
C3.3228	VIC	1	50	2.5
C3.3265	NED	1	50.3	2.5

Table 2-9: Annealing temperature, $MgCl_2$ concentration, and M13 labelled tags used for the amplification of long microsatellites.

Microsatellite name[¥]	M13 labelled tag	Group number	Annealing temperature (°C)	Magnesium concentration (mM)
C3.2871	6-FAM	3	55	2
C3.2997	6-FAM	3	55	2
C3.3193	VIC	3	60	2.5
C3.3274	NED	3	61	2
C3.3301	VIC	3	55	2

[¥] Microsatellites were named based on chromosome number and genomic position in the 10s of kb (i.e. C3.2805 is on chromosome 3 at 28.05Mb).

The PCR program used for the amplification of short and long microsatellites consisted of 95°C for 10 min; 31 cycles of 95°C for 15 sec, the indicated annealing temperature for 20 sec, 72°C for 60 sec; a final extension at 72°C for 5 min; and held at 11°C.

PCR products were amplified in separate PCR reactions and then pooled by adding 1 µL of each PCR product into the same tube as per the group number. PCR product of similar sizes were labelled with one of the three M13 fluorescent tags (6-FAM, VIC, NED) and assigned with a group number. The grouping system allows PCR product of similar sizes to be combined and analysed in single fragment analysis reaction (Neilan *et al.* 1997). Each sample from 3 pools was run on the ABI3730 Capillary Sequencer at The Ramaciotti Centre for Gene Function Analysis, UNSW.

Fragment analysis results were analysed using the GeneMapper software (Applied Biosystems) to determine the size of PCR products.

2.2.12 Linkage analysis

Linkage analysis was performed using GENEHUNTER-IMPRINTING (Strauch *et al.* 2000), with 7 pedigrees of CA affected dogs. Due to the lack of accurate data on the incidence of CA in Australian Kelpies, linkage analysis was performed assuming the frequency of the disease allele was 2%, with complete penetrance. Frequencies of microsatellite alleles in the population were estimated from 131 Australian Kelpie dogs consisting of CA affecteds, relatives of CA affecteds, and unaffected controls.

2.2.13 Prediction of alternative splice site

The Alternative Splice Site Predictor (ASSP) software, described by Wang and Marin (2006) and available at http://es.embnet.org/~mwang/assp.html, was used to locate the potential splice sites which flank synonymous and UTR substitutions, conserved differences between mammals and dogs and insertion or deletion at the exon-intron boundary of genes (Wang and Marin 2006).

2.2.14 Prediction of transcription factor binding sites

The TFSEARCH v1.3 software, described by Heinemeyer *et al.* (1998) and available at http://molsun1.cbrc.aist.go.jp/research/db/TFSEARCH.html, was used to predict possible transcription factor binding sites that flank synonymous and UTR substitutions and conserved differences between mammals and dogs (Heinemeyer *et al.* 1998).

2.2.15 Identification of CpG islands

The EMBOSS CpGPlot software, described by Rice *et al.* (2000) and available at http://www.ebi.ac.uk/Tools/emboss/cpgplot/, was used to identify and plot CpG rich areas within regulatory region of genes that are prone to DNA methylation (Rice *et al.* 2000).

2.2.16 Prediction of protein structure and identification of conserved amino acids

The Protein Homology/analog Y Recognition Engine v2.0 (Phyre2) software, described by Kelley and Sternberg (2009) and available at http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index, was used to predict the secondary and tertiary structures of wild form and mutated form of the DMGDH protein based on protein structure homology (Kelley and Sternberg 2009). The PyMOL software, developed by Schrodinger (2010) and available for download at http://www.pymol.org/, was used to generate a 3D tertiary structure model of wild type and mutated type of the DMGDH protein using PDB files generated by the Phyre2 software (Schrodinger 2010).

The PRALINE multiple sequence alignment software (PRALINE), described by Simossis *et al.* (2005) and available at http://www.ibi.vu.nl/programs/pralinewww/, was used to align and identify evolutionally conserved amino acid residues between mammals and worm (*C. elegans*), for the DMGDH protein (Simossis *et al.* 2005). Protein sequences (dog, human, chimpanzee, cattle, mouse, rat and worm) were downloaded from the NCBI database.

2.2.17 Prediction of secondary structure of RNA

The RNAfold software, described by Gruber *et al.* (2008) and available at http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi, was used to predict the minimum free energy (MFE) secondary structure of single-stranded RNA sequences using the algorithm proposed by Zuker and Stiegler (Zuker and Stiegler 1981).

3 RESULTS

3.1 Results

3.1.1 Sequencing of synonymous and UTR substitutions

A total of 17 synonymous and UTR substitutions were investigated as a possible cause of CA in Kelpies. Eleven of these were synonymous substitutions in the coding region of eight different genes and six were single nucleotide substitutions in the 5' and 3' UTRs of genes (Table 3-1). 16 differences were assessed by sequencing 23 CA affecteds, 17 unaffected controls (4 of which possessed the CA haplotype[2] found in affecteds) and 4 unaffected siblings of CA affecteds. The last synonymous substitution in the *CKMT2* gene, was sequenced in 10 CA affecteds, 11 unaffected controls (1 of which possessed a CA haplotype) and 4 unaffected siblings of CA affecteds. 14 of the 17 differences were eliminated by comparison with sequences obtained from 14 unaffected controls not carrying the CA haplotype, 3 obligate carriers and 3 Border Collie samples (Table 3-1). In detail, these polymorphisms were mostly segregated with the CA trait and were present in unaffected dogs with CA haplotype and in unaffected siblings (Figure 3-1). This suggests that these differences do not cause CA in Kelpies, if the disease penetrance is complete.

The missing SNP calls indicated in Figure 3-1 were the result of unsuccessful PCR reactions and could be due to two different causes: disruption of primer binding sites by recombination events and/or poor DNA quality.

[2] CA haplotype referred to the 5 Mb common SNP haplotype (identified by homozygosity analysis using SNPs on the Affymetrix SNP array v2 (Table 1-1)) shared between all CA affected Kelpies.

The remaining 3 differences were present in all CA affected dogs and in all 4 unaffected controls carrying the CA haplotype. All three of them (30616120 bp, 30616525 bp, 30616716 bp) are all located in the 3' UTR of *BHMT* (Table 3-1). These differences may influence *BHMT* transcription by disrupting transcription factor or miRNA binding sites. However the sequences upstream of these differences does not match any known transcription factor binding sites (Heinemeyer *et al.* 1998) and may not influence the expression of genes directly downstream. These three differences also do not disrupt recognition sites for any of the reported conserved miRNAs since there are no known miRNA target sequences located in the CA region of dog chromosome 3 or in the human homologous region (Lewis *et al.* 2005).

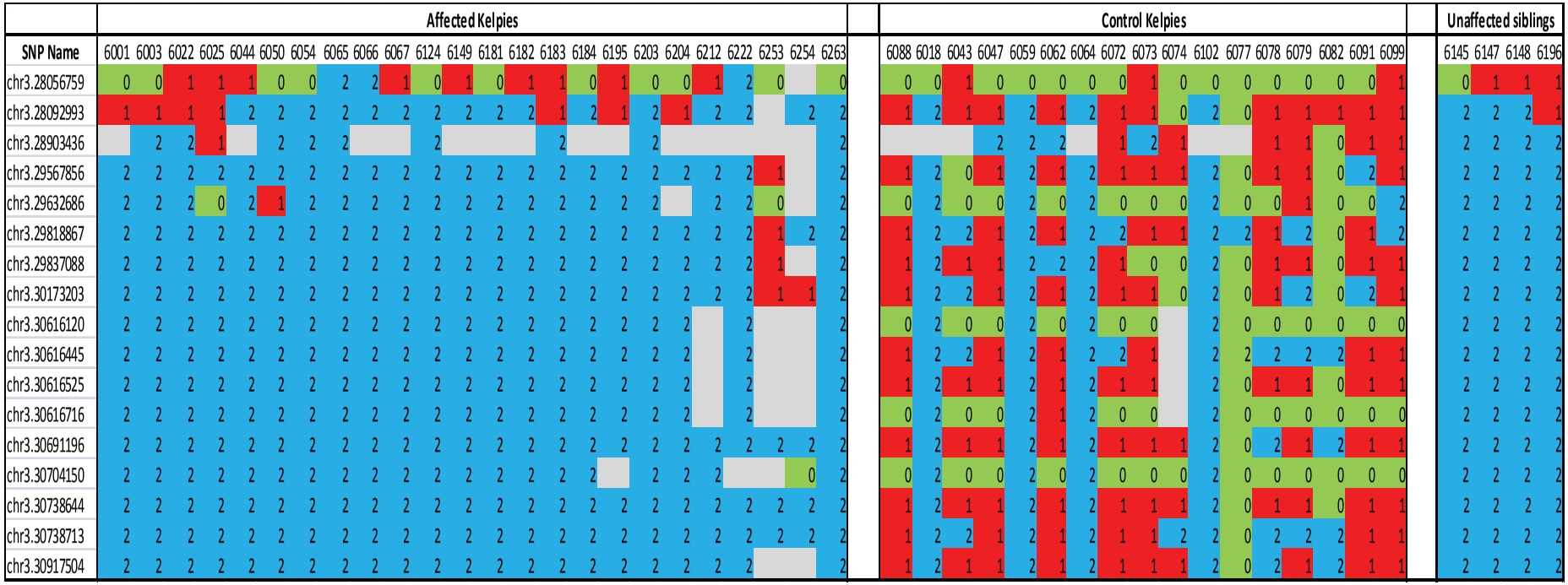


Figure 3-1: SNP genotypes in candidate region for 24 affecteds, 17 unaffected controls and 4 unaffected siblings to affecteds 6149 and 6195, for synonymous and UTRs SNPs between 28 and 33 Mb on chromosome 3 (NCBI Dog build 2.1).

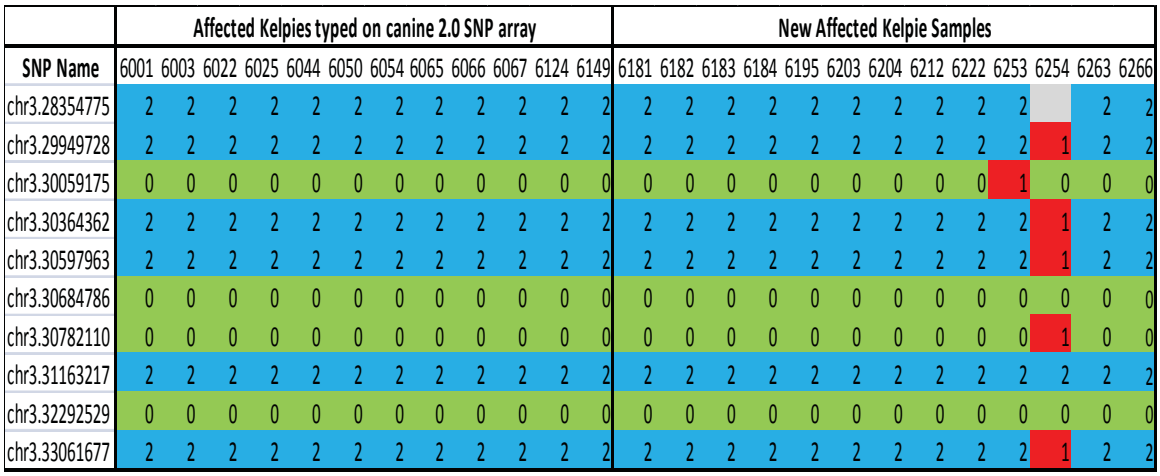
Alleles are colour coded with green (0) and blue (2) represent homozygotes at alternate alleles and red (1) represents heterozygotes. Missing SNP calls are shaded in grey. Samples across the figure are divided into affecteds, unaffected controls and unaffected siblings, with SNP names (genomic position) going down the figure.

Table 3-1: Synonymous and UTR substitutions identified by gene sequencing and used for CA case-control association study.

Gene name	Location (bp)	Type	Sequence preceding the SNP	Present outside the CA haplotype
ATP6AP1	28056759	Synonymous	5′ TGTCAGCAGC 3′ [C/T]	**Yes**
RPS23	28092993	Synonymous	5′ GTGCGTCCGG 3′ [G/C]	**Yes**
CKMT2	28903436	Synonymous	5′ GATGGCCCGT 3′ [T/C]	**Yes**
ZFYVE16	29567856	Synonymous	5′ CATTTTGCTC 3′ [C/T]	**Yes**
SPZ1	29632686	3' UTR	5′ AAACTGTAAC 3′ [C/T]	**Yes**
WBP1	29818867	5' UTR	5′ GTAGAAGAGA 3′ [A/G]	**Yes**
THBS4	29837088	Synonymous	5′ AAAAACATTC 3′ [C/T]	**Yes**
PAPD4	30173203	Synonymous	5′ TCAAGCTCCA 3′ [A/G]	**Yes**
BHMT	30616120	3' UTR	5′ ATAAGTATCG 3′ [G/A]	No
BHMT	30616445	3' UTR	5′ GATGAGACAA 3′ [A/G]	**Yes**
BHMT	30616525	3' UTR	5′ GTGGCCAGGG 3′ [G/C]	No
BHMT	30616716	3' UTR	5′ TTTCAAAGTG 3′ [G/A]	No
DMGDH	30691196	Synonymous	5′ GCAGGACTCA 3′ [A/G]	**Yes**
DMGDH	30704150	Synonymous	5′ AGTAGGGGTC 3′ [C/T]	**Yes**
DMGDH	30738644	Synonymous	5′ TCTGGCGTTT 3′ [T/C]	**Yes**
DMGDH	30738713	Synonymous	5′ CAGAAATTAC 3′ [C/T]	**Yes**
ARSB	30917504	Synonymous	5′ CTGGGGCCCA 3′ [A/G]	**Yes**

3.1.2 Sequencing of additional affected dogs with selected SNPs on the canine 2.0 SNP array

Ten SNPs located across the candidate region were analysed by sequencing in 13 CA affected Kelpie dogs that had not been previously genotyped. These SNPs were considered to be informative markers for checking the presence of SNP haplotype in CA affected dogs since only some unaffected control dogs were homozygous for these 10 SNPs based on previous genotyping results (Shearman *et al.* 2011). The purpose of this was to see if these CA affected individuals shared a common SNP haplotype with 12 other CA affected dogs previously processed on the Affymetrix canine SNP array v2 (Shearman *et al.* 2011). Affected individuals with a different SNP haplotype could be phenocopies (i.e. dogs with a disease other than CA but with CA-like symptoms). Of the 13 affected dogs genotyped for these 10 selected SNPs, 12 dogs shared the same SNP haplotype as previously identified CA haplotype (Figure 3-2). One affected dog (WK6254) was heterozygous at several markers across the region segregating a different haplotype.



	Affected Kelpies typed on canine 2.0 SNP array												New Affected Kelpie Samples											
SNP Name	6001	6003	6022	6025	6044	6050	6054	6065	6066	6067	6124	6149	6181	6182	6183	6184	6195	6203	6204	6212	6222	6253	6254	6263	6266
chr3.28354775	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2		2	2
chr3.29949728	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2
chr3.30059175	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
chr3.30364362	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2
chr3.30597963	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2
chr3.30684786	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
chr3.30782110	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
chr3.31163217	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
chr3.32292529	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
chr3.33061677	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2

Figure 3-2: SNP genotypes of 25 affected dogs for 10 selected SNPs on the canine 2.0 SNP array.

Alleles are colour coded with green (0) and blue (2) represent homozygotes at alternate alleles and red (1) represents heterozygotes. Missing SNP calls are shaded in grey. Samples across the figure are divided into old affecteds and new affecteds, with SNP names (genomic position) going down the figure.

3.1.3 Sequencing of differences conserved between mammals and dogs

A total of 22 differences between affecteds and unaffecteds in evolutionally conserved introns and intergenic regions were investigated as a possible cause of CA in Kelpies (Table 3-2). The conservation of these regions between the dog genome and other mammalian genomes suggests that mutations in these regions may have functional implications for gene regulation (Boffelli *et al.* 2004; Dermitzakis *et al.* 2005). Previous studies have demonstrated that mutations in non-coding conserved sequences may influence gene transcription by disrupting transcription factor binding sites (Bejerano *et al.* 2004). These differences were genotyped by sequencing in 4 CA affecteds and up to 6 unaffected controls (3 of which possessed the CA haplotype found in affecteds). These polymorphisms were homozygous in all 4 CA affecteds and in 3 unaffected controls with the CA haplotype (Figure 3-3). This suggests that these differences do not cause CA in Kelpies, if the disease penetrance is complete. 12 of the 22 differences were eliminated by comparison with sequences obtained from 3

unaffected controls not possessed the CA haplotype (Table 3-2). This left 10 sequence variants that were present in all CA affected dogs and in all 3 unaffected controls that possessed the CA haplotype.

These 10 differences (Table 3-3) were analysed for their potential to disrupt or create new transcription factor binding sites (Heinemeyer *et al.* 1998), while variants located within introns of genes were analysed for the presence of alternative splice sites (Wang and Marin 2006). No alternative splice sites were predicted for all four mutations located in introns of genes. A 2 nucleotide deletion (CT → -) located at 31674050 bp in the intergenic region was predicted to disrupt an HSF2 transcription factor binding site (Heinemeyer *et al.* 1998). The function of HSF2 transcription factor is to regulate and facilitate protein folding, protein degradation, protein-protein interactions and protein localisation in cells in the presence of heat stress (Richter *et al.* 2010). Conversely an indel (AA → G) located in an intron of *ATG10* and a SNP (G → A) located at 32500348 bp in the intergenic region between *ZBED3* and *AGGF1* genes were not predicted to disrupt transcription factor binding sites (Table 3-3) (Heinemeyer *et al.* 1998).

Table 3-2: Differences highly conserved between dogs and other mammals, identified by gene sequencing and used for CA case-control association study.

Gene name	Location (bp)	Source	Sequence preceding the SNP	Present outside the CA haplotype
ATG10	28273654	Intron	5′ TATAAAAAAA 3′ [AA/G].	No
-	29828219	Intergenic	5′ TTTCATTTAC 3′ [C/G]	No
MTX3	29957695	Intron	5′ GAGGGGACCT 3′ [–/C]	Yes
-	29984289	Intergenic	5′ AGCTATTCTT 3′ [CTT/G]	No
CMYA5	30085423	Intron	5′ CCTGGCTGAC 3′ [C/T]	No
DMGDH	30705732	Intron	5′ AAGTGTCTGA 3′ [A/G]	No
DMGDH	30727425	Intron	5′ TGAGGACCTT 3′ [T/C]	No
ARSB	30829090	Intron	5′ ACTTTTATCA 3′ [A/G]	Yes
ARSB	30829251	Intron	5′ TTTTATTTTG 3′ [G/A]	Yes
ARSB	30901375	Intron	5′ CACAGGCACG 3′ [G/A]	Yes
-	30917959	Intergenic	5′ TAATGGTTCT 3′ [T/C]	Yes
-	30917967	Intergenic	5′ CTGCTGGCCG 3′ [G/A]	Yes
-	30949002	Intergenic	5′ ACTCTTTCTT 3′ [TCTT/–]	Yes
-	31075908	Intergenic	5′ AATGGAGATT 3′ [T/C]	No
-	31358376	Intergenic	5′ ATTTAAAATT 3′ [T/A].	Yes
-	31674050	Intergenic	5′ AGAACACTCT 3′ [CT/–].	No
-	31723991	Intergenic	5′ AGTGTCAATG 3′ [G/A]	Yes
-	31752468	Intergenic	5′ ATTTTGTCAC 3′ [C/T]	Yes
-	31758102	Intergenic	5′ TTGCTTGCTT 3′ [T/C]	Yes
-	31960132	Intergenic	5′ AGGCTGTGTC 3′ [C/A]	No
-	32500348	Intergenic	5′ CTATTAAACG 3′ [G/A].	No
IQGAP2	33001932	Intron	5′ GGCATAAATA 3′ [A/C].	Yes

	Affected Kelpies					Control Kelpies					
SNP Name	6025	6065	6149	6203		6018	6059	6064	6074	6078	6079
chr3.28273654	1	2	2	2		2	2	2	0	1	
chr3.29828219	2	2	2	2		2	2	2		1	1
chr3.29957695	2	2	2	2		2	2	2	0	2	
chr3.29984289	2	2	2	2		2	2	2	0	1	1
chr3.30085423	2	2	2	2		2	2	2	0	1	1
chr3.30705732	2	2	2	2		2	2	2	1	1	0
chr3.30727425	2	2	2	2		2	2	2	1	1	1
chr3.30829090	2	2	2	2		2	2	2	1	2	
chr3.30829251	2	2	2	2		2	2	2	1	2	
chr3.30901375	2	2	2	2		2	2	2	1	2	
chr3.30917959	2	2	2	2		2	2	2	1	2	
chr3.30917967	2	2	2	2		2	2	2	1	2	
chr3.30949002	-	-	-	-		-	-	-	0	-	
chr3.31075908	2	2	2	2		2	2	2	1	1	1
chr3.31358376	2	2	2	2		2	2	2	1	2	
chr3.31674050	-	-	-	-		-	-	-	0	0	
chr3.31723991	2	2	2	2		2	2	2	0	2	
chr3.31752468	2	2	2	2		2	2	2	1	2	
chr3.31758102	2	2	2	2		2	2	2		2	
chr3.31960132	2	2	2	2		2		2		0	
chr3.32500348	2	2	2	2		2	2	2	1	1	1
chr3.33001932	2	2	2	2		2	2	2	1	2	

Figure 3-3: SNP genotypes in candidate region for 4 affecteds and 6 unaffected controls, for differences in gene regulatory regions highly conserved between mammals and dogs (NCBI Dog build 2.1).

Alleles are colour coded with green (0) and blue (2) represent homozygotes at alternate alleles and red (1) represents heterozygotes. Purple (-) represents deletion of allele. Missing SNP calls are shaded in grey. Samples across the figure are divided into affecteds, unaffected controls and unaffected siblings, with SNP names (genomic position) going down the figure.

Table 3-3: Differences highly conserved between mammals and their predicted effects on transcription factor binding sites and alternative splicing.

Gene name	Location (bp)	Source	SNP	Transcription factor binding		Alternative splicing	
				Wild type	Mutated type	Wild type	Mutated type
ATG10	28273654	Intron	AA/G	92.9% match for CdxA TF binding site.	92.9% match for CdxA TF binding site.	No effect	No effect
-	29828219	Intergenic	C/G	No TF binding.	No TF binding.	N/A	N/A
-	29984289	Intergenic	CTT/ G	No TF binding.	No TF binding.	N/A	N/A
CMYA5	30085423	Intron	C/T	No TF binding.	No TF binding.	No effect	No effect
DMGDH	30705732	Intron	A/G	No TF binding.	No TF binding.	No effect	No effect
DMGDH	30727425	Intron	T/C	No TF binding.	No TF binding.	No effect	No effect
-	31075908	Intergenic	T/C	No TF binding.	No TF binding.	N/A	N/A
-	31674050	Intergenic	CT/-	87.2% match for HSF2 TF binding site.	No TF binding.	N/A	N/A
-	31960132	Intergenic	C/A	No TF binding.	No TF binding.	N/A	N/A
-	32500348	Intergenic	G/A	92.9% match for CdxA TF binding site.	92.9% match for CdxA TF binding site.	N/A	N/A

Table 3-4: A summary of the potential functional implications created by the polymorphisms located in UTRs of genes and evolutionarily conserved regions of the genome.

Gene name	Location (bp)	Type	Differences	Potential functional implications
BHMT	30616120	3' UTR	G/A	No
BHMT	30616525	3' UTR	G/C	No
BHMT	30616716	3' UTR	G/A	No
ATG10	28273654	Intron	AA/G	92.9% match for CdxA TF binding site.
-	29828219	Intergenic	C/G	No
-	29984289	Intergenic	CTT/G	No
CMYA5	30085423	Intron	C/T	No
DMGDH	30705732	Intron	A/G	No
DMGDH	30727425	Intron	T/C	No
-	31075908	Intergenic	T/C	No
-	31674050	Intergenic	CT/-	87.2% match for HSF2 TF binding site.
-	31960132	Intergenic	C/A	No
-	32500348	Intergenic	G/A	92.9% match for CdxA TF binding site.

3.1.4 Microsatellites genotyping

Three additional CA families were genotyped for long microsatellites that had a high degree of variation between individuals within the pedigree (Figures 3-4 to 3-6). One of the families was also genotyped for short microsatellites that showed no genetic variation between affecteds and controls and hence these markers were not informative for the linkage analysis (Figure 3-7).

Long Microsatellites

Name	**Repeat**	
C3.2871	AAAG × 38	
C3.2997	AAAG × 36	
C3.3193	AAAG × 22	AAAAG × 18
C3.3274	TTTC × 20	CCTTT × 26
C3.3301	AAAGG × 14	AAAAG × 17

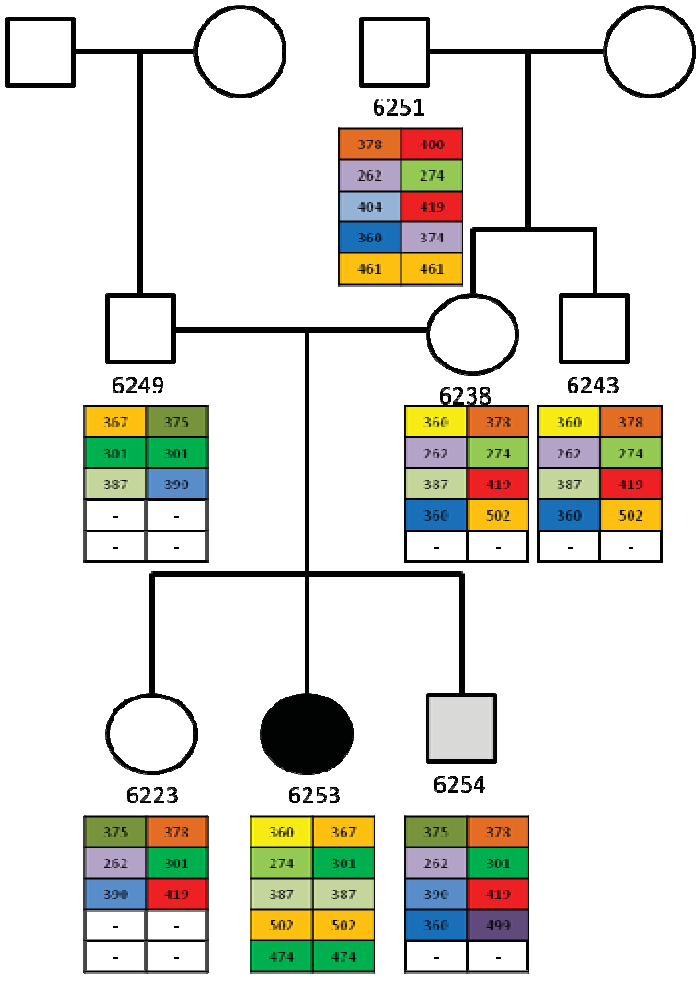


Figure 3-4: Inheritance patterns in affected CA Pedigree 1 for long microsatellites.

Alleles have been colour coded per microsatellite for ease of viewing. Affected individual is shaded in black. Suspected affected individual is shaded in grey. Missing data is represented as a dash.

Long Microsatellites

Name	Repeat	
C3.2871	AAAG × 38	
C3.2997	AAAG × 36	
C3.3193	AAAG × 22	AAAAG × 18
C3.3274	TTTC × 20	CCTTT × 26
C3.3301	AAAGG × 14	AAAAG × 17

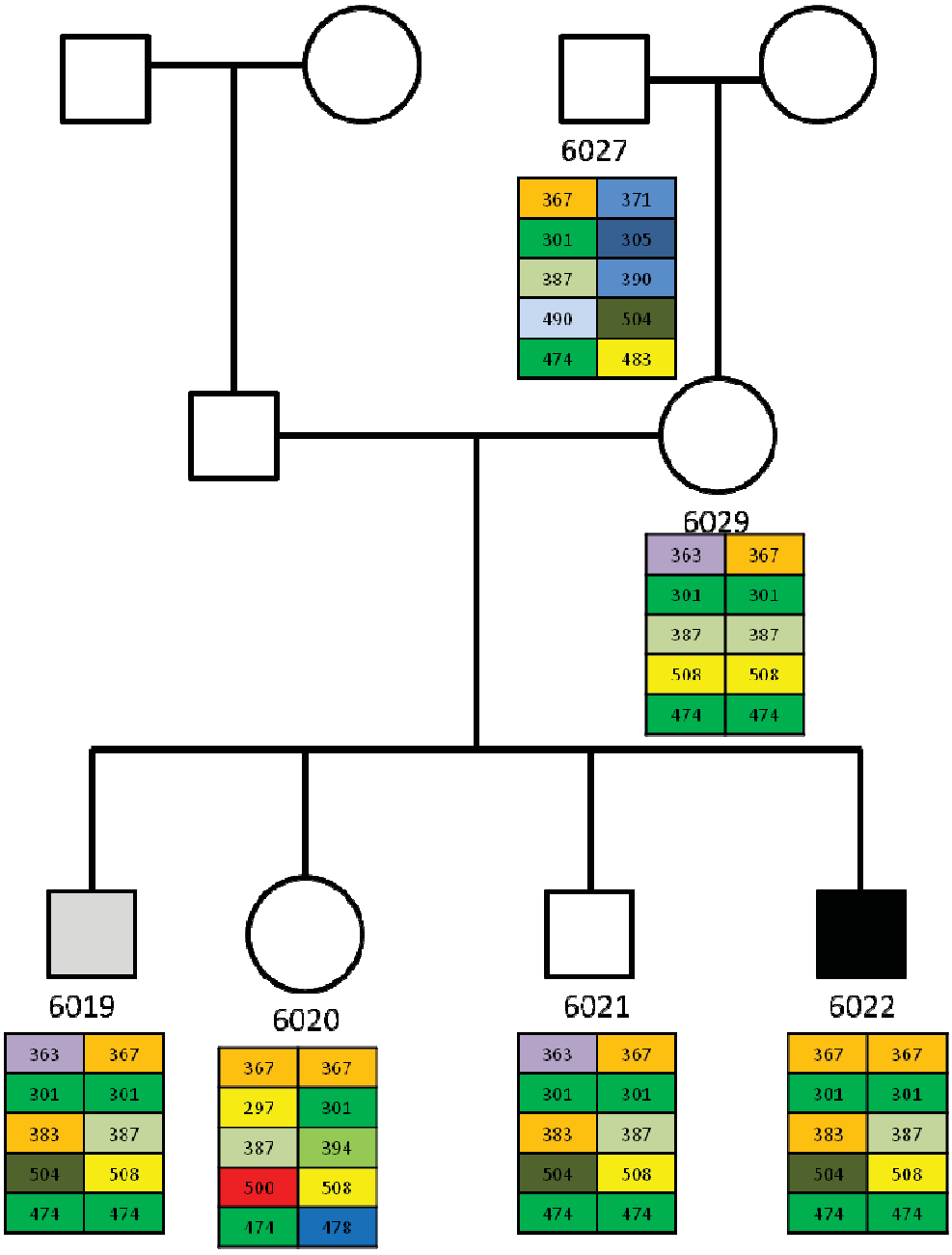


Figure 3-5: Inheritance patterns in affected CA Pedigree 2 for long microsatellites.

Alleles have been colour coded per microsatellite for ease of viewing. Affected individual is shaded in black. Suspected affected individual is shaded in grey.

Long Microsatellites

Name	Repeat	
C3.2871	AAAG × 38	
C3.2997	AAAG × 36	
C3.3193	AAAG × 22	AAAAG × 18
C3.3274	TTTC × 20	CCTTT × 26
C3.3301	AAAGG × 14	AAAAG × 17

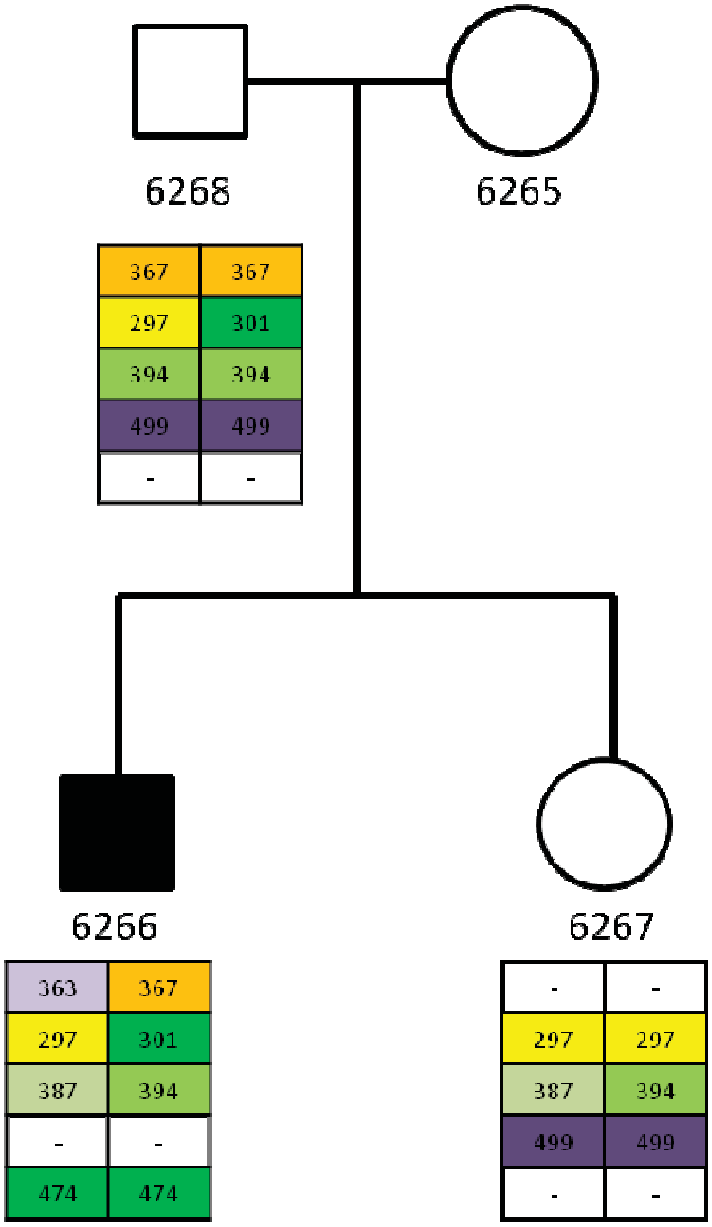


Figure 3-6: Inheritance patterns in affected CA Pedigree 3 for long microsatellites.

Alleles have been colour coded per microsatellite for ease of viewing. Affected individual is shaded in black. Missing data is represented as a dash.

Short Microsatellites

Name	Repeat
C3.2805	AC × 21
C3.2856	TG × 25
C3.2869	TG × 21
C3.2984	AG × 28
C3.3228	TG × 22
C3.3265	AC × 25

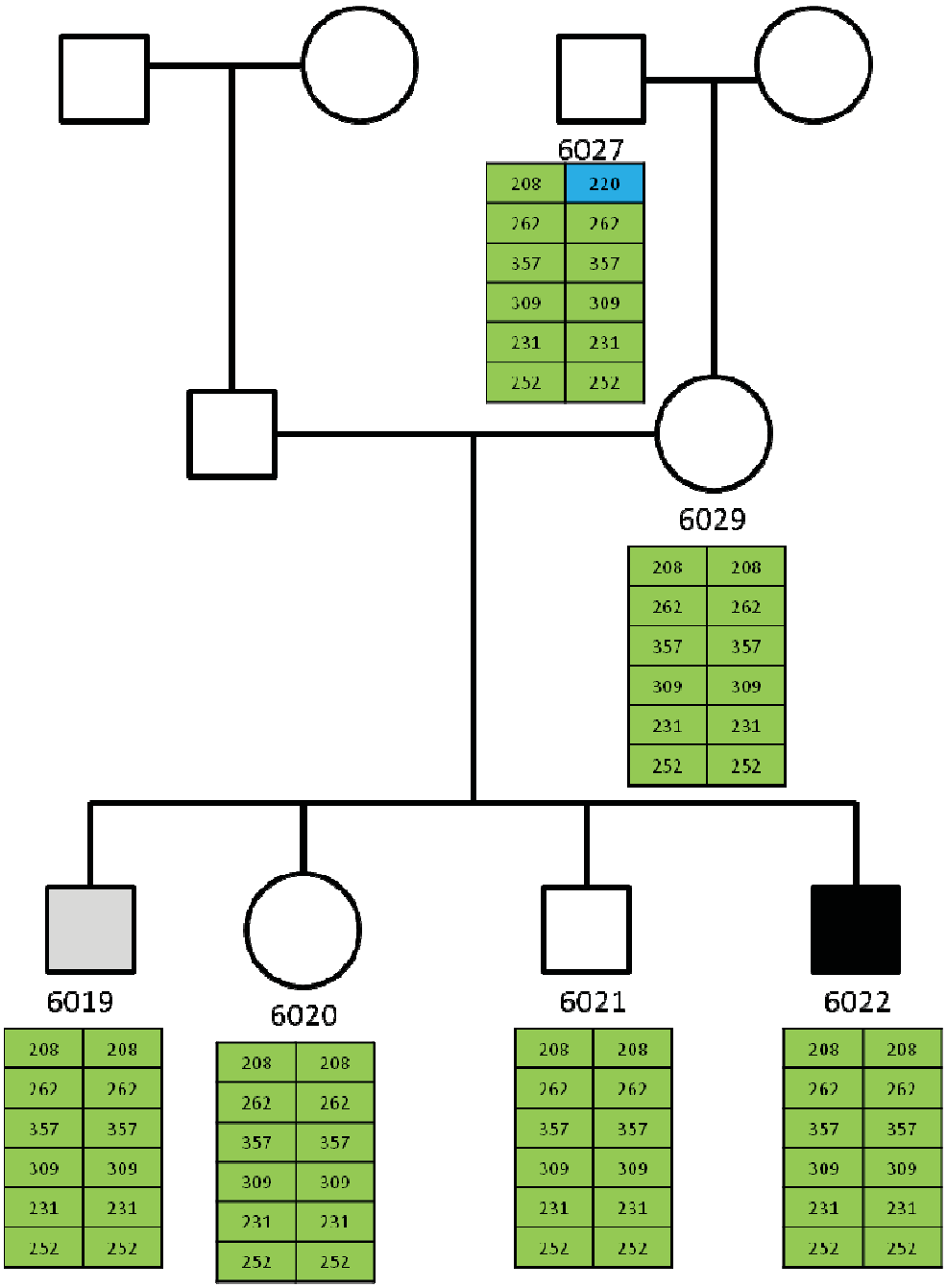


Figure 3-7: Inheritance patterns in affected CA Pedigree 2 for short microsatellites.

Alleles have been colour coded per microsatellite for ease of viewing. Affected individual is shaded in black. Unknown individual is shaded in grey.

3.1.5 Linkage analysis

Multipoint linkage analysis was performed between five long microsatellite markers located across the candidate region and CA disease gene locus, with 7 CA affected families (including the 3 new CA affected families genotyped in this study). A peak LOD score of 2.98 at 29.97 Mb on chromosome 3 is right on the threshold of significance (+3) and hence has reached the commonly accepted level of significance for linkage between these five marker loci and disease locus (Figure 3-8). This is supportive of the weak association identified in previous studies (Shearman *et al.* 2011). A multipoint linkage analysis was not performed between six short microsatellite markers positioned across the candidate region and CA disease gene locus since these short microsatellite markers were not considered as informative markers for the linkage analysis.

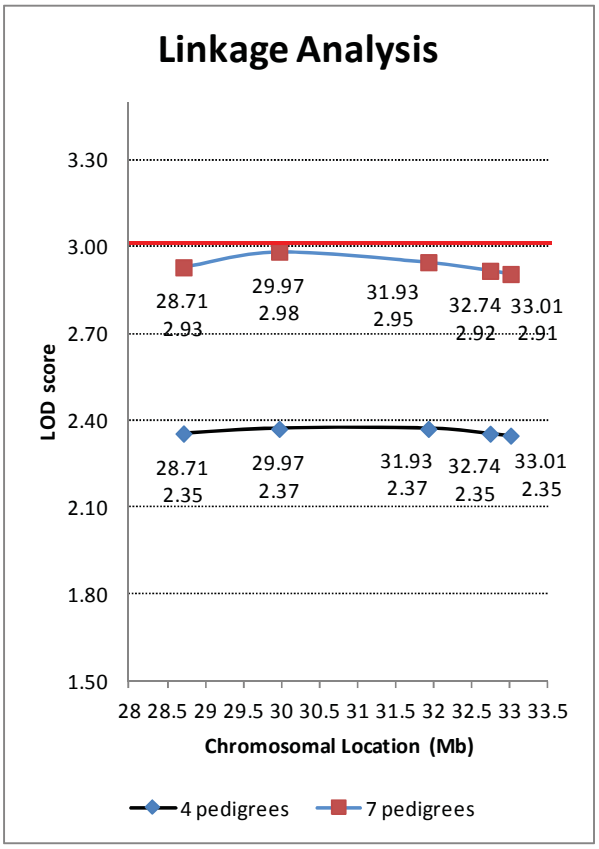


Figure 3-8: Multipoint linkage analysis of long microsatellite group, with four CA affected families and seven CA affected families respectively.

Significant LOD score threshold is indicated by the red line.

3.1.6 Candidate gene sequencing

3.1.6.1 Sequencing of exons

PCR and Sanger sequencing were used to fill in gaps in 454 sequences (Shearman 2011) for a total of 37 coding exons and 5'UTR and 3'UTR from 18 different genes within the candidate region (Tables 2-6 to 2-7). However PCR was only successful in producing products of the expected size for 32 exons (Table 3-5) with little success in amplifying the remaining 5 exons (Table 3-6). All 32 PCR products yielded good quality sequences and were aligned to the Boxer dog reference sequence Build v2.1 as described earlier (Sections 2.2.8 and 2.2.9). Sequences for all 32 exons were obtained from 2 CA affecteds, 2 carriers, 1 unknown and up to 4 controls (1 of which possessed the CA haplotype) for any sequence differences that may be the cause of CA in Kelpies. No sequence differences were observed between CA (affected) and other (unaffected) samples for 30 of the 32 exons sequenced (Table 3-4).

An affected (WK6025), two carriers (WK6061 and WK6142) and one unaffected individual carrying the CA haplotype (WK6059) were homozygous for the 'G' allele at position 30615886 in the *BHMT* 3'-UTR sequence, while one unaffected (WK6074) and one individual of unknown phenotype (WK6058) were found to be homozygous for 'A' at this SNP. This base substitution may be due to sequence variation between individuals since one control WK6059 is also homozygous for the 'G' allele (Figure 3-9). The sequence before this SNP (5′ATTGAGCCCCACA 3′) is not a known transcription factor binding site (Heinemeyer *et al.* 1998), and hence this single base change is not expected to impact on the transcriptional regulation of *BHMT*.

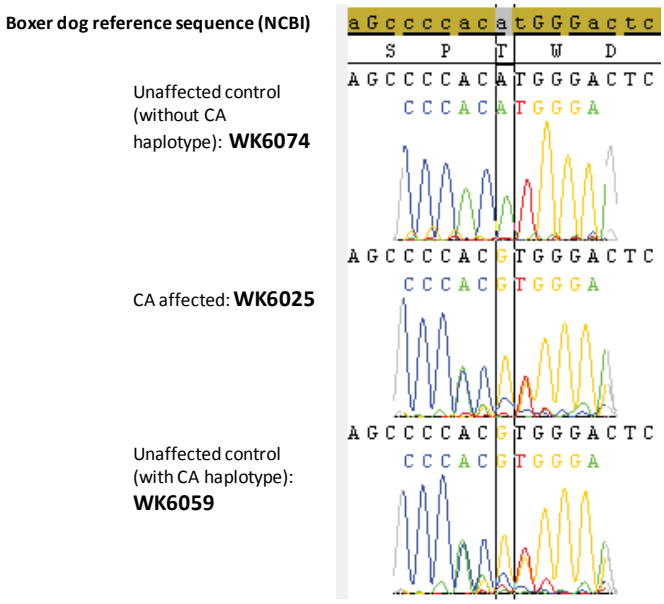


Figure 3-9: Electropherogram showing aligned sequences of 3' UTR of *BHMT* with a single base substitution (A/G) in affected and control with CA haplotype.

Agarose gel electrophoresis of PCR products amplified with the primer set for *DMGDH* Exon 16 revealed different PCR band patterns for individuals of different affection status (Figure 3-10). All CA affected dogs and unaffected dogs with CA haplotype were homozygous for a larger allele (~ 1200 bp) while unaffected dogs without CA haplotype were homozygous for a smaller allele (937 bp) predicted from the dog reference genome sequence. Some unaffected control dogs and dogs of unknown affection status were heterozygous for both alleles (examples shown in Figure 3-10).

Sequencing results showed a 223 bp insertion near the 5' end of amplified DNA sequence in all affected dogs and unaffected dogs with CA haplotype (Figure 3-11). An insertion might disrupt the gene by creating an alternative splice site which spliceosome can binds to initiate splicing of pre-mRNA sequences. This could results in exon skipping or intron retention and affects the level of gene expression. Using the Alternative Splice Site Predictor software described by Wang and Marin (2006), two potential constitutive acceptor splice sites were located within the 223 bp inserted

sequence (Figure 3-11). These constitutive (acceptor) splice sites are located at positions 367 bp [tttttttATTTTATTTA] and 248 bp [gggtctccagGATCGCGCCC] respectively upstream of the original acceptor splice site [tgtttttcagGTGGTTGGCA]. Hence, the constitutive (acceptor) splice site at 367 bp upstream of the original acceptor splice site (since it has a score similar to the original splice site) is predicted to splice the final exon at this position and produce a truncated protein that is 72 amino acids shorter in length. There was no sequence difference observed between Boxer reference sequence and unaffected dogs without CA haplotype.

Amino acid residues of the DMGDH protein that are evolutionally conserved between dogs, other mammals and worm (*C. elegans*) were identified and colour coded using PRALINE multiple sequence alignment software (Section 2.2.16). Of the 72 deleted amino acid residues (position 822 to 892), 57 amino acids (position 822 to 878) were highly conserved between dog, other mammals (human, chimpanzee, cattle, mouse, rat) and worm (Figure 3-12).

Using the PyMOL software developed by Schrodinger (2010), 3D tertiary structures of wild form and mutated form of the DMGDH protein (Figure 3-13) were constructed with PDB files generated by the Phyre2 software (Section 2.2.16). Positions of the deleted amino acid residues (Figure 3-12) were visually represented as pink dots on the 3D protein model (Figure 3-14). These deleted amino acid residues were mainly located on the surface of protein where protein-protein and protein-DNA interactions can occur.

RNA secondary structures spanning all three constitutive acceptor splice sites in the RNA sequence were predicted for the CA affected dog (with the 223 bp insert) and

unaffected control dog (wild type RNA sequence) (Figures 3-15 to 3-18). Stem-loop structures were observed at the original constitutive acceptor splice site for the unaffected control Kelpie (Figure 3-15) and at the first constitutive acceptor splice site for the CA affected Kelpie (Figure 3-16). The stem-loop structure observed in unaffected control (that spanned the original constitutive acceptor splice site) was not seen in the CA affected Kelpie (positions 446 to 455) (Figure 3-18). Hence splicing of the final exon is more likely to be initiated at the first constitutive acceptor splice site in the CA affected dog, rather than at the second or the third (original) constitutive acceptor splice sites.

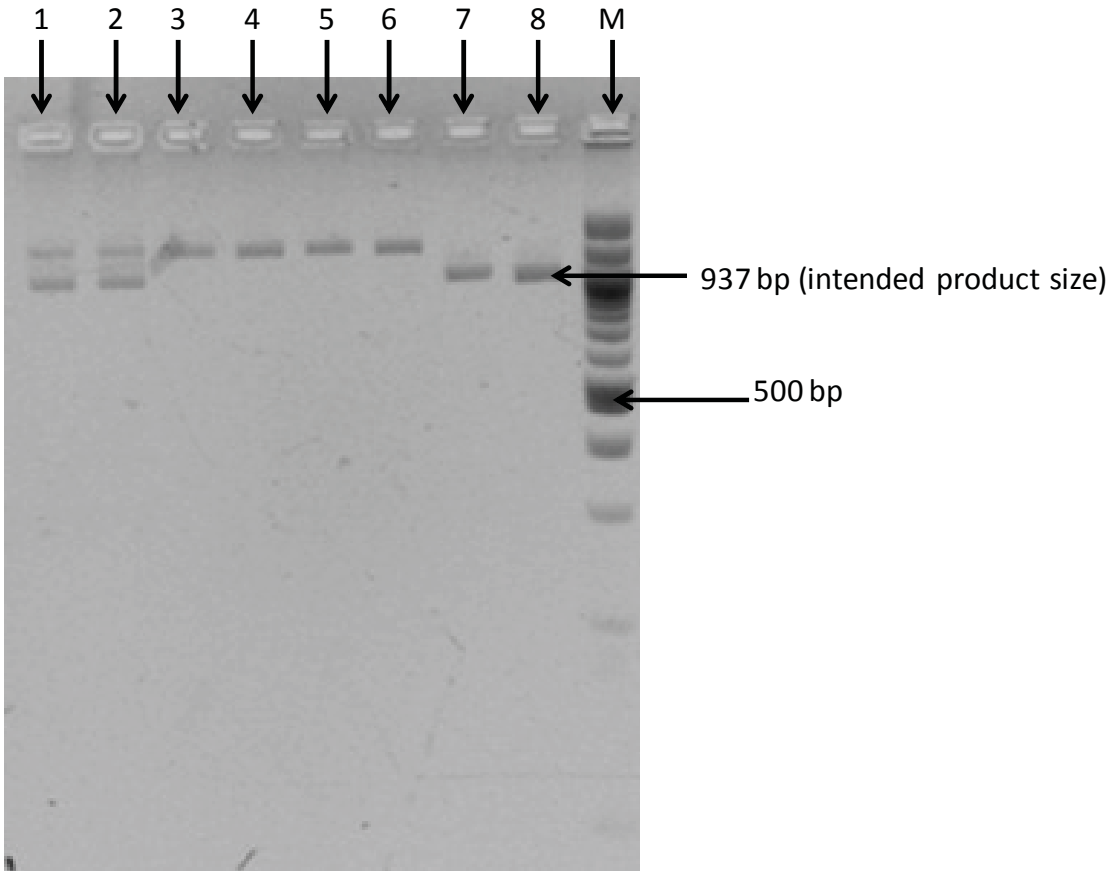


Figure 3-10: PCR product amplified with primer set for *DMGDH* Exon 16 and visualised after electrophoresis using a 2% (w/v) agarose gel stained with 0.03% ethidium bromide.

M: 100 bp size standard; Lane 1: WK6088 unaffected control without CA haplotype; Lane 2: WK6058 unknown affection status; Lane 3: WK6025 CA affected; Lane 4: WK6059 unaffected control with CA haplotype; Lane 5: WK6065 CA affected; Lane 6: WK6145 unaffected sibling of CA affected with CA haplotype; Lane 7: WK6205 unknown affection status without CA haplotype; Lane 8: BC2923 (Border Collie CA unaffected).

```
CLUSTAL 2.1 multiple sequence alignment

Control         GAAGTTTGAGGGGGTTGATATGTATTTTGAAAATAATTTTTCATAACTTTT---------- 51
Affected        GAAGTTTGAGGGGGTTGATATGTATTTTGAAAATAATTTTTCATAACTTTTTTTTTTTT 60
                **************************************************

Control         ------------------------------------------------------------
Affected        TTTTTTTTTTTTTTTTAAAGATTTTATTTATTTATTCATGATAGTCACAGAGAGAGAGAGA 120

Control         ------------------------------------------------------------
Affected        GAGGCAGAGACACAGGCAGAGGGAGAAGCAGGCTCCATGCACCGGAGCCCGATGTGGGAT 180

Control         ------------------------------------------------------------
Affected        TCGATCCCGGGTCTCCAGGATCGCGCCCTGGGCCAAAGGCAGGCGCCAAACCGCTGCGCC 240

Control         ----------------------------------AAACCTAATATTCACAAAAGTAAAGT 77
Affected        ACCCAGGGATCCCAAATAATTTTTCATAACTTTTAAACCTAATATTCACAAAAGTAAAGT 300
                                                  **************************

Control         TTACTTCCTGGCGTTAGCCCCATGGGCAATAATCCCTAAAACTTGCAGATTGTAAATCAA 137
Affected        TTACTTCCTGGCGTTAGCCCCATGGGCAATAATCCCTAAAACTTGCAGATTGTAAATCAC 360
                ************************************************************

Control         ACCTATGTTTTAAGTGTTTAATCTGGAGAGGAGGACCATAAAGCTGATGGCCAAGATGTT 197
Affected        ACCTATGTTTTAAGTGTTTAATCTGGAGAGGAGGACCATAAAGCTGATGGCCAAGATGTT 420
                ************************************************************

Control         GAGGGTTACTGGCTCTTGTTTTTCAGGTGGTTGGCAACACGACATCTGGAACCTACAGTT 257
Affected        GAGGGTTACTGGCTCTTGTTTTTCAGGTGGTTGGCAACACGACATCTGGAACCTACAGTT 480
                ************************************************************

Control         ACAGCATACAGAAGAGTCTGGCGTTTGCCTATGTCCCTATAGAACTAAGTAAAGTAGGAC 317
Affected        ACAGCATACAGAAGAGTCTGGCGTTCGCCTATGTCCCTATAGAACTAAGTAAAGTAGGAC 540
                ************************* ********************************

Control         AACAGGTGGAGGTTGAACTGTTAGGCAGAAATTACCCAGCAAGCATCATACAGGAGCCCT 377
Affected        AACAGGTGGAGGTTGAACTGTTAGGCAGAAATTATCCAGCAAGCATCATACAGGAGCCCT 600
                ********************************** ***********************

Control         TGGTATTGACTGAACCAACCAGAAACCGACTTCGGAAGAAAAGCAGAAAGGACAAAATTT 437
Affected        TGGTATTGACTGAACCAACCAGAAACCGACTTCGGAAGAAAAGCAGAAAGGACAAAATTT 660
                ************************************************************

Control         GAAAAAAGACCTTCAGCAGGCAACTGAATTATGTTAGCTCTATGACTGTACTAAAAATCA 497
Affected        GAAAAAAGACCTTCAGCAGGCAACTGAATTATGTTAGCTCTATGACTGTACTAAAAATCA 720
                ************************************************************

Control         TAACTGCTTTTTGGGAGAATATAAGAAATCAAGAATTCATTCAACCCCTCAAAATCTAGA 557
Affected        TAACTGCTTTTTGGGAGAATATAAGAAATCAAGAATTCATTCAACCCCTCAAAATCTAGA 780
                ************************************************************

Control         TTTATAATCTTGATGAAACCTTTCTCAGCTCTTTCCTAAGTGAGAAACAATAATGGATTA 617
Affected        TTTATAATCTTGATGAAACCTTTCTCAGCTCTTTCCTAAGTGAGAAACAATAATGGATTA 840
                ************************************************************

Control         GTGATTAACATTGTTCTTTAAAATGTAGACGTTGGAAATGAATGTTTTTATTAACCTATA 677
Affected        GTGATTAACATTGTTCTTTAAAATGTAGACGTTGGAAATGAATGTTTTTATTAACCTATA 900
                ************************************************************

Control         TGGTTAATGAAACTCAGCAAAACATTTAGGTGTTTGCTAGTATATAATCACTATTATAAC 737
Affected        TTGTTAATGAAACTCAGCAAAACATTTAGGTGTTTGCTAGTATATAATCACTATTATAAC 960
                * ************************************************************

Control         TCAATTAGAGGACATTGTGTAATTTTAGGAATATATGGTTCTTGTGACCTGAAAACAAAT 797
Affected        TCAATTAGAGGACATTGTGTAATTTTAGGAATATATGGTTCTTGTGACCAGAAAACAAA- 1019
                ********************************************** *********

Control         TGTTTTACATAAAAATAAATGCAGTACTATTGATGGATTTGTTTCATTGAG 848
Affected        ---------------------------------------------------
```

Figure 3-11: (Caption is on the next page).

Figure 3-11: Multiple sequence alignment of sequences spanning the *DMGDH* Exon 16 region in CA affected and unaffected control dogs. Constitutive acceptor splice sites located within the 223 bp insert in CA affected dog were shown.

The 223 bp insert sequence (in affected) is highlighted in yellow with constitutive acceptor splice sites indicated in red surrounded with a border. Constitutive original acceptor splice site was located at position 437 bp and 214 bp for CA affected and unaffected control respectively. Stop codon/s are indicated in bold and underlined. Multiple sequence alignment was performed using ClustalW2 described by Larkin *et al.* (http://www.ebi.ac.uk/Tools/msa/clustalw2/) (Larkin *et al.* 2007).

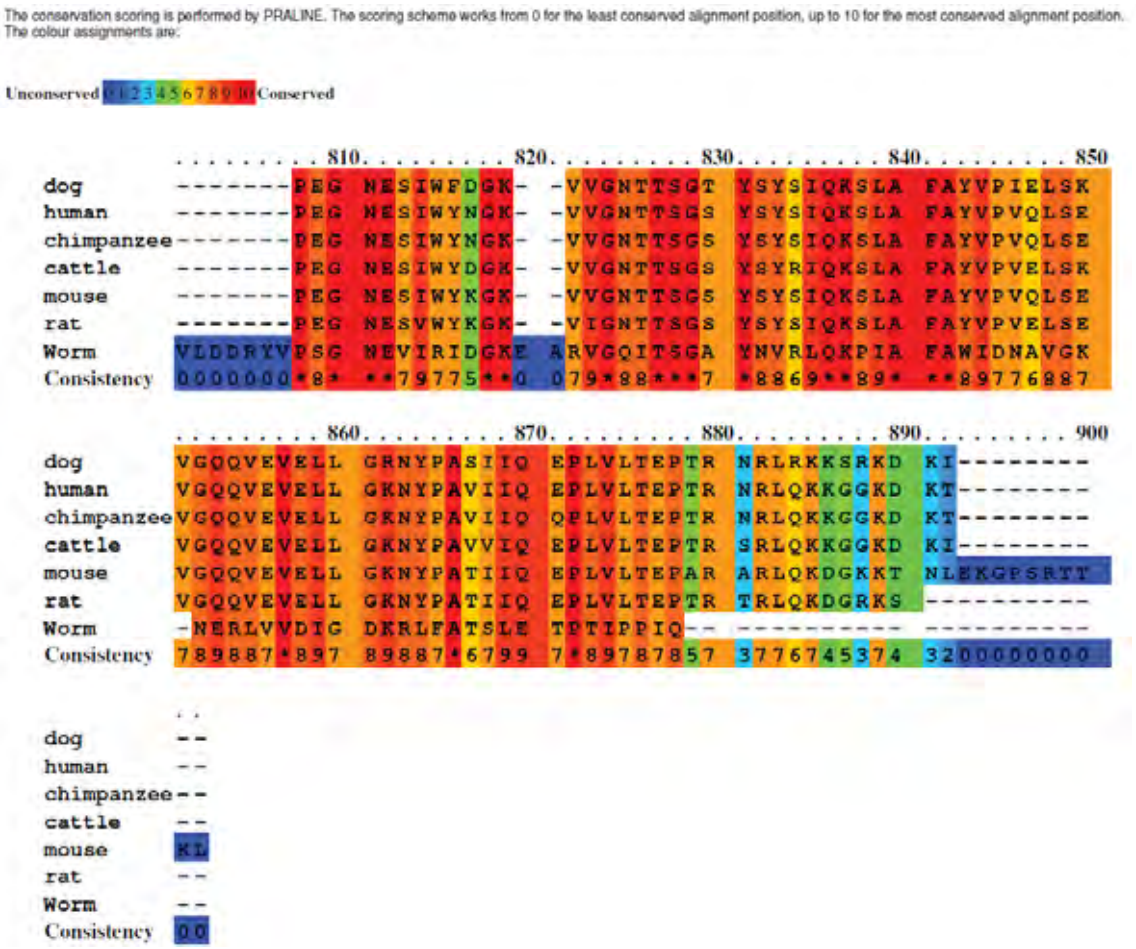


Figure 3-12: Multiple sequence alignment for the DMGDH protein sequence showing amino acid residues that were evolutionally conserved between dog, other mammals and worm encoded by the last exon (Exon 16) of mRNA transcript.

a)

b)

Figure 3-13: PyMOL generated 3D tertiary structures of DMGDH protein in a) unaffected control – wild type, and in b) CA affected – mutated type.

72 amino acids encoded by the last exon of wild type DMGDH protein were deleted in the mutated DMGDH protein. Protein models were predicted and constructed based on protein homolog template c1pj6A (crystal structure of dimethylglycine oxidase of arthrobacter2 globiformis in complex with folic acid), with 100% confidence and 28% sequence identity between protein sequence and the homolog template.

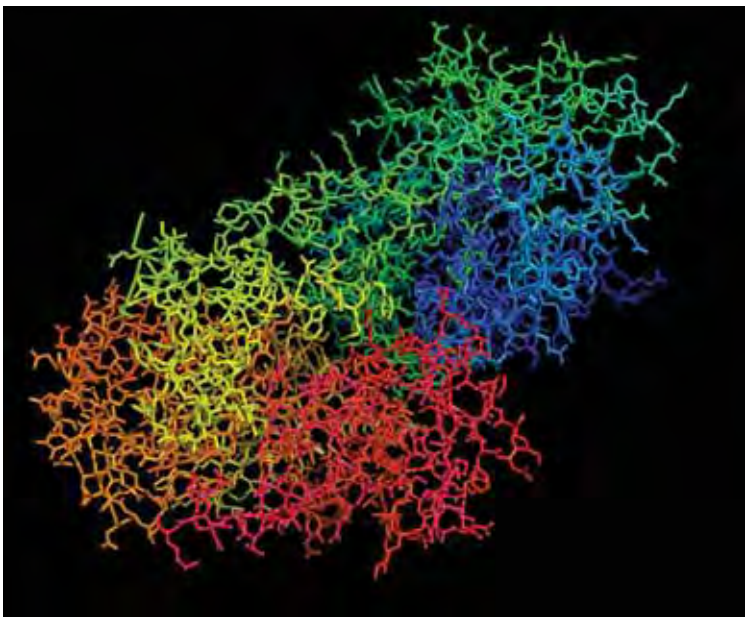


Figure 3-14: A PyMOL generated 3D structure of DMGDH protein with amino acid residues deleted in the CA affected represented as pink dots.

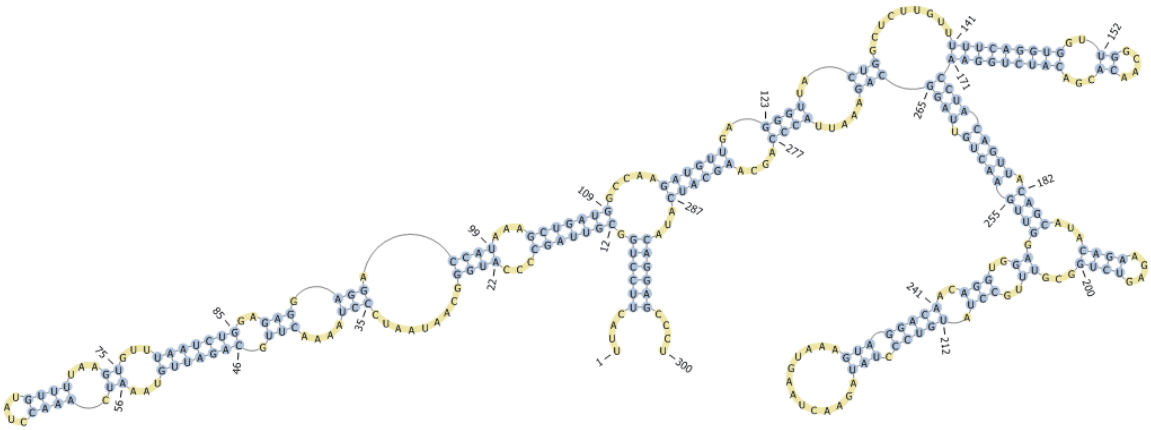


Figure 3-15: RNA secondary structure of unaffected control, spanning the original constitutive acceptor splice site of *DMGDH* pre-mRNA (positions 137 to 146 bp).

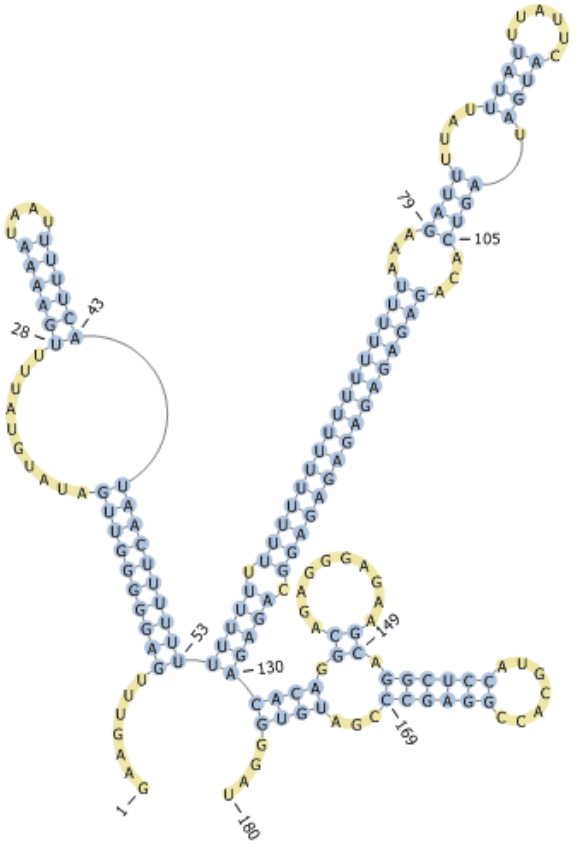


Figure 3-16: RNA secondary structure of CA affected dog, spanning the first constitutive acceptor splice site (in the inserted sequence) of *DMGDH* pre-mRNA (positions 70 to 79).

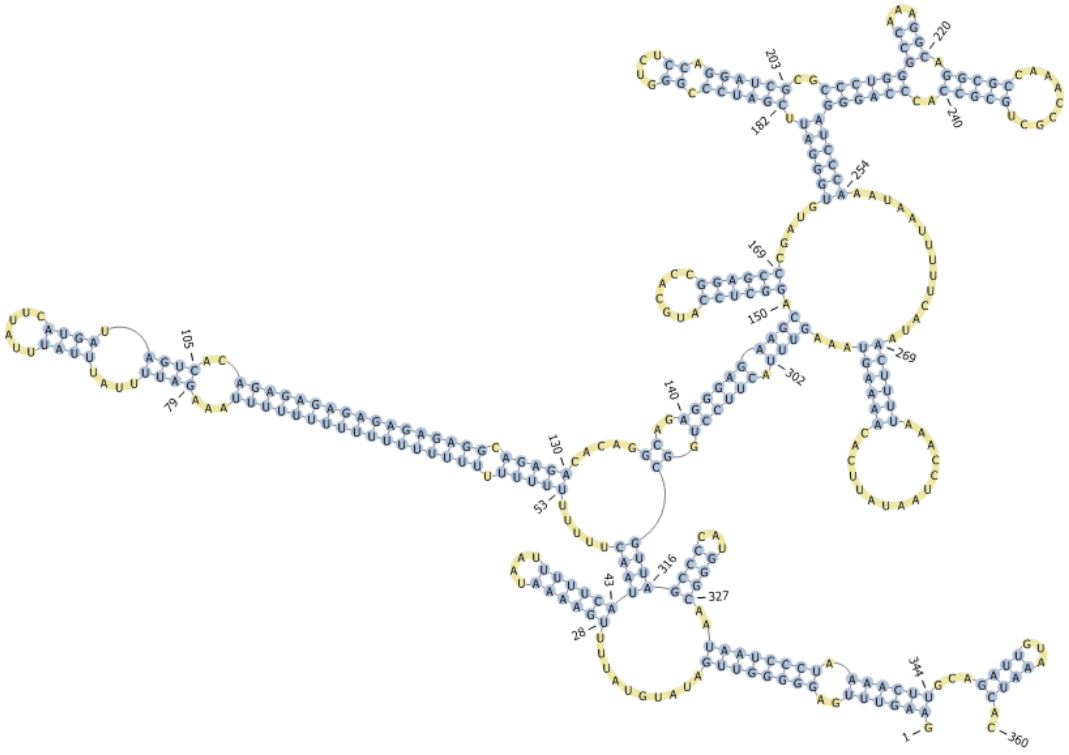


Figure 3-17: RNA secondary structure of CA affected dog, spanning the first two constitutive acceptor splice sites (in the inserted sequence) of *DMGDH* pre-mRNA (positions 70 to 79 and positions 193 to 202).

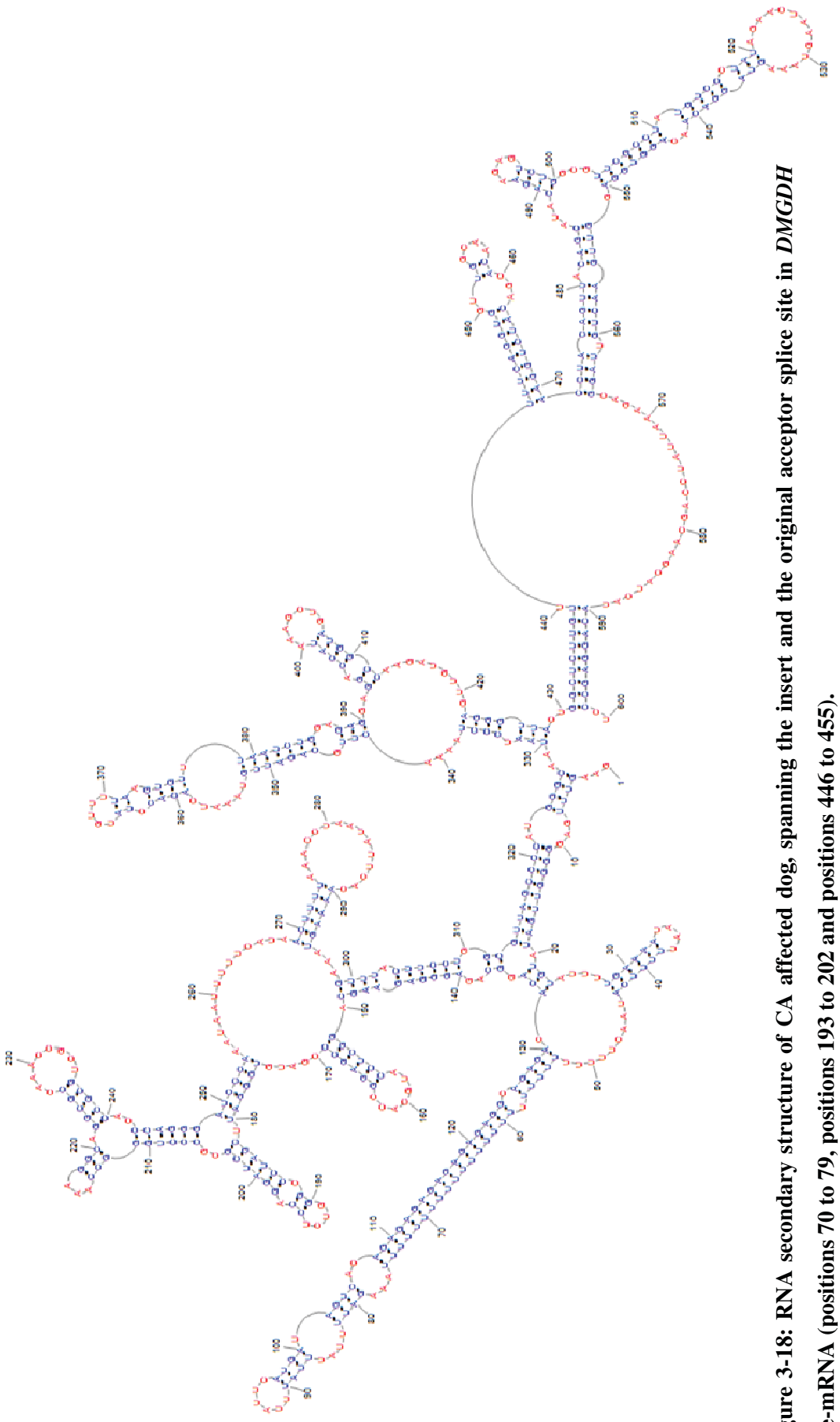


Figure 3-18: RNA secondary structure of CA affected dog, spanning the insert and the original acceptor splice site in *DMGDH* pre-mRNA (positions 70 to 79, positions 193 to 202 and positions 446 to 455).

Table 3-5: Sequenced exons of candidate genes and genomic positions of potential CpG islands.

Gene name	Exon	Matched with Boxer reference sequence	GC Content (%)	CpG island region genomic positions (bp) (Section 2.2.15)
HOMER1	8	Yes	34.6	No
HOMER1	10	Yes	34.1	No
ARSB	1	Yes	78.60	30752800-30753016
AP3B1	1	Yes	72.7	31369559-31369698
AP3B1	3	Yes	37.3	No
AP3B1	23	Yes	37.9	No
F2RL1	1	Yes	76.8	32697471-32697445
F2RL1	2	Yes	50.4	No
SERINC5	1	Yes	77.4	29671161-29671238
ANKRD34B	5'UTR	Yes	27.7	No
PAPD4	2	Yes	40	No
PAPD4	9	Yes	38.8	No
PAPD4	5'UTR	Yes	31.5	No
SSBP2	3	Yes	41.9	No
SSBP2	4	Yes	41.2	No
SSBP2	7	Yes	52.2	No
SSBP2	10	Yes	51	No
SSBP2	15	Yes	44.8	No
ATG10	1	Yes	37	No
ZFYVE16	4	Yes	42	No
ZFYVE16	9	Yes	36.2	No
ZFYVE16	10	Yes	35.2	No
ZFYVE16	15	Yes	38.7	No
ACOT12	1	Yes	78	28811165-28811236
AGGF1	1	Yes	70.6	32569669-32569571
AGGF1	9	Yes	31.4	No
AGGF1	3'UTR	Yes	29.8	No
DHFR	5	Yes	38	No
DHFR	3'UTR	Yes	32.7	No
DMGDH	1	Yes	72.7	30672475-30672537

Table 3-6: Unamplified exons of candidate genes and genomic positions of potential CpG islands.

Gene name	Exon	GC Content (%)	CpG island region genomic positions (bp) (Section 2.2.15)
SCAMP1	1	65.2	31304844-31304789
RASGRF2	1	70.8	29169459-29169274
AP3B1	6	31.3	None
SSBP2	1	72.2	28499468-28499631
BHMT2	1	69.7	None

3.1.6.2 Sequencing of regulatory regions of genes

The upstream regions of 12 different genes within the candidate region had previously been sequenced by 454 sequencing (Shearman 2011). PCR and Sanger sequencing were used to fill in gaps in this initial sequencing. However PCR was only successful in producing products of the expected size for 8 genes (Table 3-7) and the remaining 4 upstream regions could not be sequenced (Table 3-8). All 8 PCR amplified products yielded good quality sequences and were aligned to the Boxer dog reference sequence Build v2.1 as described earlier (Sections 2.2.8 and 2.2.9). Sequences for the upstream regions of all 8 genes were obtained for 2 CA affecteds, 2 carriers, 1 unknown and up to 4 controls (1 of which possessed the CA haplotype) for any sequence differences that could be the cause of CA in Kelpies.

No significant sequence differences in upstream regulatory regions were observed between CA affecteds and unaffecteds for 5 of the 8 genes sequenced (Table 3-6).

Complete sequencing of 800 bp upstream of the *SCAMP1* gene revealed new sequence that was previously unknown in the boxer reference sequence (Figure 3-19). This GC-rich region is between 31305460 and 31305055 bp on chromosome 3 of the Boxer dog genome (NCBI Dog build 2.1) (Shearman *et al.* 2011). No sequence variants were seen between CA affected and unaffected dogs in this region.

CCCGGGTCGGTACCGCGAGGCCGGCCTCCGTGATTTGTCAGCAGCCGCGCGTCCCGGGACCG
GAAGACGAGGGGCTGCGGGGGCAGGTGGGGCGGGTGGGGCGCGTGCGCCGCGCTCGGCGGCG
AGGTCGGGGCCCGGAGCGCGGTTGAGTCGCCGCCCCCTGCCCAGGTGCGCAGCCTGCGCGTC
GGGCCGCCTGCCGGGCCTGGGAGGCGGGGTCCGCGCGCGTCCTCCCCGCTCACAGGTGCCCC
CCCGTCACGTTGGACCTCAGGGAGCCAGCGTGCAGCGGCTGGCGGTGGGATGGAGTCCAGCC
GCACGCGGGGGCTCAGCCCGGCCCGCCGCTCCCGGCTCCCTCCGCCCCTGGTCCCTGCGGTC
GAAGCCAAAGCCCGGAACGTGGATCCCGGCCCGG

Figure 3-19: Masked repeat sequence in the upstream region of *SCAMP1* gene.

A 5 bp insertion of CCGAG (genomic positions 30752305 bp to 30752309 bp) was observed in the upstream region of *ARSB* in two affecteds, two carriers and one control with the CA haplotype (WK6059). The insertion was not present in controls without the CA haplotype (WK6078 and WK6088). This insertion may be a polymorphism between individuals rather than a cause of CA because one control WK6059 is also homozygous for the 5 bp insertion (Figure 3-20). There is no transcription factor binding site spanning the 5 bp insertion sequences according to the TFSEARCH database (Heinemeyer *et al.* 1998).

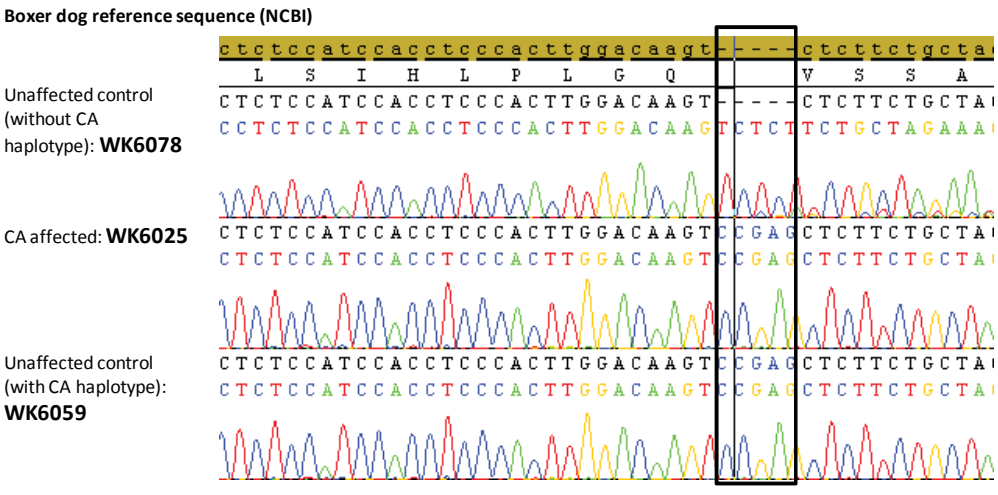


Figure 3-20: Electropherogram showing aligned sequences of upstream regions of *ARSB* with the 5 bp insertion (CCGAG) in affected and control with CA haplotype.

Complete sequencing of 710 bp upstream of the *SERINC5* gene revealed another region of new sequence that was previously unknown in the boxer reference sequence (Figure 3-21). The GC-rich region is between 29670706 and 29671041 bp on chromosome 3 of the Boxer dog genome (NCBI Dog build 2.1) (Shearman *et al.* 2011). No new sequence variants were observed between CA affected and unaffected dogs in this region.

```
GTGCCCGGGTCCCAGGGAGCCGAGGAGGGAGGTGAGGCGCCCCGTGCAGCTCTGGGC
TCCCTGCAGCAGGCATCCAGCCGGTGCATCGGGCCGCGGGCGCCGGGGGCCGCGAGC
GGGCGCGGTAGCACGGGAACGCGGGCCCCGCCCTCCGGTAAGTGAGCGCAGCCGCCG
CGCCGCCTGACGCCAGCGGAAGGGCGGGCCCCGCGCGAGCCTGCTCGGCCGCCGGCG
CCCGAGAGGGGGCGGGGCCTGCGGCGCCGCGGCCGGGCGGGGCGGGGCTGCGGCGCG
AGGGGCGGGGCCTGCGGCTCGAGGGGGCGGGGCTCGAGGGGCGGGGCGGGG
```

Figure 3-21: Masked repeat sequence in the upstream region of *SERINC5* gene.

Table 3-7: Sequenced upstream regulatory regions of candidate genes and genomic positions of potential CpG islands.

Gene name	Genomic position	Matched with Boxer reference sequence	GC Content (%)	CpG island region genomic positions (bp) (Section 2.2.15)
SCAMP1	31305743-31304943	Yes.	58.5	31305557-31305452
ARSB	30751973-30752753	5 bp insertion from 30752305 to 30752309 bp.	60.2	30752442-30752698
AP3B1	31368712-31369512	Yes. No causative variation observed.	69.4	31368762-31369457
SERINC5	29670340-29671050	Yes.	64	29670398-29670533 & 29670599-29670711
BHMT2	30673327-30672610	Yes. No causative variation observed.	68.3	30673108-30672830
AGGF1	32570436-32569716	Yes. No causative variation observed.	67.3	32570307-32569770
DHFR	29400690-29401436	Yes. No causative variation observed.	57.4	None
F2R	32777827-32777127	Yes. No causative variation observed.	42.6	None

Table 3-8: Unamplified upstream regulatory regions of candidate genes and genomic positions of potential CpG islands.

Gene name	Genomic position	GC Content (%)	CpG island region genomic positions (bp) (Section 2.2.15)
HOMER1	30282089-30282889	83.2	30282138-30282832
RASGRF2	29170307-29169507	65.3	29170257-29170096 29170072-29169916 29169754-29169562
SSBP2	28498825-28499625	77.3	28498914-28499282 28499291-28499567
MSH3	29405211-29404411	45.2	None

4 DISCUSSION

4.1 Discussion

4.1.1 Homozygosity analysis and linkage analysis

The identification of a common haplotype containing the CA mutation using the Affymetrix canine SNP chip v2 allowed preliminary genetic testing of suspected affected Kelpie dogs for the presence of CA (Shearman *et al.* 2011). This is to ensure all suspected affected Kelpie dogs shared the common CA haplotype before being used in the study since ataxia can be caused by many reasons other than CA (Vermeer *et al.* 2011). The haplotype was used to identify phenocopies of suspected CA cases. Most newly affected dogs display symptoms of CA while also sharing the same CA haplotype as other affected dogs that were typed on the canine SNP chip (Figure 3-2). However, to date one suspected affected dog (WK6254) did not show any evidence of sharing the same SNP haplotype as other affected dogs and therefore may be a phenocopy of CA (Figure 3-2). Some carriers and unaffected siblings of affecteds shared the CA haplotype but have not show any symptoms of CA based on physical observations.

The CA candidate region in the Australian Kelpie dog was previously mapped to a 5 Mb region (CFA3: 28 and 33 Mb) through a genome wide association study with the Affymetrix canine SNP array v2 (Shearman *et al.* 2011). In that study a single affected was heterozygous up to 29.8 Mb, suggesting the candidate region was between 29.8 and 33.8 Mb. Genotyping additional affected Kelpies supports this reduced candidate region; a single observed recombination event (WK6253) further reduces it to between 30.2 Mb and 33.8 Mb on chromosome 3 (Figure 3.1) which contains 28 genes (Table 7-3).

The presence of CA haplotype in phenotypically unaffected Kelpies leads to the possibility that CA may not have complete penetrance in affected Kelpies. This means the age of onset and severity of symptoms is variable among Kelpies affected with CA. Hence it is possible for mildly affected dogs with CA haplotype to show no symptoms of the disease until several months of age and be classified as unaffected by their owners and their vets at the time of clinical diagnosis.

However it is also possible that the mutation causing CA occurs in a common haplotype where only some copies of the haplotype have the CA mutation. This means the haplotype cannot be used to identify the disease-causing mutation (Shearman *et al.* 2011). The inheritance patterns of the short microsatellite haplotypes in CA pedigree 2 (Figure 2-3) also showed no genetic variation between affected and unaffected members of the family (Figure 3-7). This result suggests these short microsatellites are not informative markers for the linkage analysis.

A multipoint linkage analysis was performed on long microsatellites data from seven Kelpie CA families (Figures 3-4 to 3-6). These results suggest a positive linkage between 5 marker loci and the CA disease locus, with a maximum logarithm of odds (LOD) score of 2.98 (significance threshold of 3) (Figure 3-8). This multipoint linkage analysis was performed as an extension to the linkage analysis performed previously by Shearman *et al.* with four Kelpie CA affected families and gave a LOD score of 2.37 (Figure 3-8) (Shearman *et al.* 2011). The positive LOD score (of 2.98) is supporting evidence that the region on chromosome 3 from 28 to 33 Mb does contain the mutation causing CA in Kelpies. The significant LOD score threshold would probably be reached by typing further CA affected and unaffected individuals.

4.1.2 Common haplotype hypothesis

The 'common haplotype hypothesis' assumes that the causative mutation is only present in CA affected Kelpies and is absent in unaffected controls with the CA haplotype. If this hypothesis is true then none of the differences studied (Tables 3-1 to 3-2) and none of the new variants identified in exons and upstream region of genes (Sections 3.1.6.1 and 3.1.6.2) can be causing CA in Kelpies since all unaffected controls with the CA haplotype were homozygous for the same sequences as CA affected Kelpies.

The most effective way of identifying consistent differences between CA affecteds and unaffected controls (which possessed the CA haplotype found in affecteds) would be to re-sequence the entire 5 Mb candidate region in affected and unaffected individuals. Therefore it may be necessary to sequence the entire candidate region in unaffected control dogs with the CA haplotype using next generation sequencing. This would help identify variants that are only present in some copies of the CA haplotype and thus reduce the number of candidate differences significantly.

4.1.3 Identification of functional differences

The presence of a common CA haplotype in phenotypically affected and unaffected Kelpies suggests that the mutation(s) that cause CA may show variable penetrance. Under this hypothesis, any of the 13 differences that were not present outside the CA haplotype can be involved in CA in Kelpies (Table 3-4). These sequence differences are located in the 3' UTR of *BHMT*, in introns of genes and in regions between genes, and hence may play a regulatory role in gene transcription. Of these 13 differences chosen as potentially affecting expression of a downstream gene, 3 obvious

differences located in transcription factor binding sites could be investigated further as a possible cause of CA in Kelpies (Table 3-4). The remaining 10 differences could be more conclusively eliminated as possible causes of CA by PCR and sequencing experiments on additional unaffected control samples without the CA haplotype (Table 2-1).

Sequences upstream of the three obvious differences *ATG10* (28273654) indel, the intergenic (31674050) deletion and the intergenic (32500348) were searched against known transcription factor binding sites and have revealed that these differences are immediately downstream of putative binding sites for the transcription factors CdxA and HSF2 (respectively) (Table 3-4). In chicken, CdxA is a homeobox transcription factor required for marking the boundary between the stomach and intestine and is mostly expressed in epithelial-mesenchymal tissues, and is not expressed in pancreas, liver and allantois (Ishii *et al.* 1997). Each member of the homeobox genes family is required for maintaining correct segment identity. Ten homeobox genes that are homologous to the *Drosophila* gene *caudal* (homologue of *CdxA* in *Drosophila*) have been isolated in vertebrates (Blumberg *et al.* 1991; Duprey *et al.* 1988; Frumkin *et al.* 1991; Gamer and Wright 1993; German *et al.* 1992; James and Kazenwadel 1991; Joly *et al.* 1992; Northrop and Kimelman 1994; Serrano *et al.* 1993).

HSF2 is a member of the heat shock transcription factor family needed to regulate and facilitate protein folding, protein degradation, protein-protein interactions and protein localisations in cells in the presence of heat stress (Richter *et al.* 2010). Dysregulation of the HSF2 protein encourages the accumulation of misfolded proteins in cells, and can contribute to a number of neurodegenerative diseases including

Alzheimer's disease, Parkinson's disease and polyglutamine diseases (Powers *et al.* 2009).

HSF2 is not the main regulator for heat shock response in mammals but plays an important role in the development of the brain and reproductive organs (Chang *et al.* 2006; Kallio *et al.* 2002; Wang *et al.* 2003) and interacts directly with HSF1, the major regulator of the heat shock response in vertebrates (Sandqvist *et al.* 2009). There are four members in the vertebrate HSF family that bind to heat shock elements in response to heat stress (Shinkawa *et al.* 2011). HSF2 knock-out mice have a shorter lifespan than wild-type mice due to an accumulation of misfolded polyQ aggregates in the brain and symptoms similar to Huntington's disease (Shinkawa *et al.* 2011). The disruption of the HSF2 transcription factor binding site at 31674050 bp on chromosome 3 in CA affected Kelpies and unaffected control Kelpies with the CA haplotype may make them more susceptible to an accumulation of protein aggregates in the brain since they do not have HSF2 transcription factor to regulate gene expression. Gene expression assay such as the luciferase reporter assay could be used to see whether the sequence difference (31674050) that was predicted to disrupt the HSF2 transcription factor binding site influences the expression of gene/s.

Luciferase reporter assays could be used to determine whether these sequence variants influence the expression of adjacent genes. This assay would require a DNA fragment containing the wild type allele and the variant allele to be PCR amplified and subcloned into an appropriate position upstream of a luciferase reporter vector such as pGL3 (Promega, WI) and then transfected into brain or other cell cultures. The effect of the variant allele on gene transcription could then be compared with wild-type allele using a dual-luciferase assay. To date, this approach exceeded the scope of this project.

4.1.4 Analysis of mutation in DMGDH exon 16

Dimethylglycine dehydrogenase (*DMGDH*) is a mitochondrial enzyme that is able to convert dimethylglycine to sarcosine through choline metabolism by removing one methyl group from dimethylglycine (Binzak *et al.* 2001). It uses folate and flavin adenine dinucleotide (FAD) as cofactors for the catabolism of choline in mitochondria (Cook *et al.* 1984; Wagner *et al.* 1984). A non-synonymous substitution of histidine by arginine at codon 109 (A→G) located in the *DMGDH* gene has been associated with deficient expression of the *DMGDH* protein and elevated levels of dimethylglycine in serum and urine of patients (which often display a fish-like body odour and chronic muscle fatigue) (Binzak *et al.* 2001). However none of these clinical signs are observed in CA affected Kelpies.

A recent study by Pierson *et al.* (2011) in human subjects affected with an early onset spastic-ataxia-neuropathy syndrome using whole-exome sequencing and homozygosity scan identified two candidate genes; *DMGDH* and *AFG3L2*. This syndrome is characterised by spasticity of the lower extremities, peripheral neuropathy, ptosis, oculomotor apraxia, dystonia, cerebellar atrophy, and progressive myoclonic epilepsy. A mutation in the *AFG3L2* gene has been previously associated with SCA28, an autosomal dominant form of cerebellar ataxia, and may be a more likely candidate gene for the spastic-ataxia-neuropathy syndrome in humans. A non-synonymous substitution at codon 1847 (A→G) located in the *AFG3L2* gene was functionally confirmed as the causative mutation using the yeast complementation assays (Pierson *et al.* 2011). However in order for this result to be conclusive, it is necessary to perform gene expression assay such as the luciferase reporter assay to confirm this non-synonymous substitution to be the causative mutation in a human brain cell line. Alternatively, a gene-knock out mice carrying this mutation can also be used to study

the phenotype associated with a down-regulated expression of the AFG3L2 protein in an animal model.

In rat tissues, *DMGDH* has the highest expression in liver and kidney, as shown by Western and Northern blots (Lang *et al.* 1994). Detectable expression levels were found at the RNA and protein level in brain, heart, lung and spleen tissues (Lang *et al.* 1994). In addition, *DMGDH* was found to be expressed in 14-days old rat embryos which further suggest the importance of this protein in choline catabolism (Lang *et al.* 1994).

An analysis of the 223 bp inserted sequence near the 3' end of *DMGDH* gene in CA affecteds revealed this sequence is located in the last intron of the gene and 396 bp upstream of the original splice site for the last exon (Exon 16). The insertion sequence is repeated farther upstream in the same intron. Analysis of the sequence using BlastN shows that many copies of the sequence are also present elsewhere in the dog genome. Some of the features of the sequence are superficially similar to those found in retrotransposon-derived elements such as SINEs (for example, the poly-T region at the 5' end of the sequence), but the BLAST results fail to show substantial homology with SINEs or other mobile genetic elements. Hence, the origin of the sequence remains uncertain.

The inserted sequence contains two predicted constitutive acceptor splice sites. As shown in Figure 3-11, the first of these acceptor splice sites is located at 367 bp, while the second splice acceptor site is 248 bp upstream of the original acceptor splice site (Figure 3-11). A putative stop codon (TGA; underlined in Figure 3-11) located immediately after the first acceptor splice site is predicted to produce a truncated protein

that is 72 amino acids shorter than the wild type DMGDH protein. The RNA stem-loop structure spanning the first constitutive acceptor splice site in CA affecteds could attract the spliceosome and so cause aberrant splicing of the last exon of *DMGDH* (Figure 3-16).

A multiple sequence alignment of amino acids for the DMGDH protein in dogs, other mammals (human, chimpanzee, cattle, mouse and rat) and worm (*C. elegans*) revealed amino acids that were deleted in CA affected Kelpies were highly conserved throughout evolutionary time (Figure 3-12). This high degree of amino acid conservation between different species suggests that these amino acids are important in maintaining the tertiary structure of functional DMGDH protein (Figure 3-13). In addition, many of the amino acid residues that are deleted in the truncated protein are normally located on the surface of DMGDH protein (Figure 3-14). These amino acid residues are likely to form active sites for other enzymes (protein-protein interactions) and binding sites for DNA-binding proteins.

To date, nine Kelpies were found to carry the CA haplotype although not showing the clinical symptoms of CA based on owners' and vets' observations. The presence of CA haplotype in phenotypically unaffected Kelpies leads to the possibility that CA may not have complete penetrance in affected Kelpies. This means the age of onset and severity of symptoms is variable among Kelpies affected with CA. Some severely affected Kelpies displayed frequent head tremors and easily recognisable symptoms of ataxia while mildly affected Kelpies only show occasional high stepping gait and a very mild intention head tremor. Hence it is possible for mildly affected dogs with CA haplotype to show no symptoms of the disease early in life, resulting in mis-classification as unaffected by their owners and vets. If CA in Kelpies is indeed caused

by a lack of *DMGDH* or a reduction in its expression, then the variability of the CA phenotype may be attributable to varying amounts of dimethylglycine present in the brains of CA affected dogs. A simple, quick and inexpensive HPLC-UV method described by Laryea *et al.* (1998) using commercially available derivatization reagents could be used to measure the level of *N,N*-dimethylglycine present in the blood and urine of CA affected and unaffected control Kelpie dogs (Laryea *et al.* 1998), before proceeding with a genetic analysis of the *DMGDH* mutation using the molecular biology-based methods described below.

However, the inheritance pattern is consistent with CA in Kelpies being a recessive inherited disease with a high penetrance (Shearman *et al.* 2011; Thomas and Robertson 1989). The mutation may have been inherited from a number of champion dogs that were frequently used for breeding (Thomas and Robertson 1989). Therefore it may be concluded that the DMGDH mutation in Kelpie and Border collies may have come from a common ancestor, the collies (Neff *et al.* 2004).

Reverse-transcription PCR (RT-PCR) analyses using RNA from brain tissues of CA affected and unaffected control Kelpie dogs (and/or RNA from transfected cells), could be used to identify the exact *DMGDH* mRNA sequence that is being produced in the presence and absence of the *DMGDH* insertion.

To test whether the two constitutive acceptor splice sites and their corresponding stop codon (located within the 223 bp inserted sequence) affect DMGDH protein activity, a *Myc-FLAG-DMGDH* expression vector (containing the wild-type sequence and the 223 bp insert sequence) could be expressed in a mammalian cell line such as HEK293T (a human kidney embryo cell line). Human cell line such as HEK293T

(derived from adenovirus transformation of human embryonic kidney (HEK) cells) have been shown to express neurofilament subunits, α-internexin and several other proteins found in neurons by immunostaining and microarray analysis (Shaw *et al.* 2002). The HEK293T cell line were used by Abitbol *et al.* (2010) to assess the effect of a SNP on ARSG activity in healthy and NCL (neuronal ceroid lipofuscinosis) affected American Staffordshire Terrier dogs (Abitbol *et al.* 2010). Western blotting with an anti-Myc antibody could be used to detect the DMGDH protein in transfected HEK293T cells. The activity and functionality of the DMGDH protein could be tested using the anaerobic ETF (electron transfer flavoprotein) reduction assay as described by Frerman and Goodman (1985) (Frerman and Goodman 1985), with 0.5M dimethyglycine as substrate (Binzak *et al.* 2001). This assay could be used to determine activities of DMGDH protein in cultured brain cells between samples using dimethyglycine as substrate for the fluorimetric reaction (Frerman and Goodman 1985).

To confirm the loss of *DMGDH* activity caused by the presence of a 223 bp insert under physiological conditions, a qRT-PCR may be performed on cDNA generated from cerebellar tissues of CA affected and unaffected control Kelpies to measure gene transcription in these dogs. *Synaptophysin* (*SYP*) could be used as a reference gene for the qRT-PCR reaction since this gene has a constant and specific expression in neuron cells of brain tissues (Pavai *et al.* 2010). The *DMGDH* and *SYP* genes could be amplified together in the same reaction using specific primers for each of the gene. The use of *SYP* gene as an internal control gene would allows quantification of the mRNA for *DMGDH* gene between CA affected and unaffected control dog samples to be normalised for any differences in the amount of total RNA that may be present in each of the reaction (Pavai *et al.* 2010). Lower levels of *DMGDH* could result in higher levels of dimethylglycine in brain tissues. The level of dimethylglycine present in CA

affected Kelpie dogs could be variable enough to account for variation in age of onset of CA and severity of disease between individuals.

Additional unaffected control, CA affected (of different severity), obligate carrier and unknown status samples would need to be typed for the alleles present to ensure the larger allele (containing the 223 bp insert) is only present in CA affecteds and is not observed in obligate unaffected control dogs. Conversely obligate carrier dogs should be heterozygote for both of the alleles.

4.1.5 Analysis of DNA methylation of candidate genes

DNA methylation is another mechanism that could influence gene transcription and may help explain the variation in clinical symptoms observed in CA affected dogs. The candidate genes investigated in this study show few sequence differences between CA affecteds and unaffected controls, but their methylation status has not been assessed (Tables 3-5 and 3-7). Differential methylation of the candidate gene promoters and/or first exons could explain the genetic basis of CA, and would have to be investigated thoroughly before these genes could be excluded as candidate CA genes.

DNA methylation influences cell development and differentiation by regulating chromatin structure and gene expression (Urdinguio *et al.* 2009). Disruption in epigenetic mechanisms can result in gene silencing or activation leading to cancers and a number of neurodegenerative diseases including Rett syndrome, Alzheimer's disease, Parkinson's disease and Huntington's disease in humans (Urdinguio *et al.* 2009).

Mutations in genes contributed by environmental factors can cause disruption in epigenetic mechanisms that lead to neurodegenerative diseases (Holliday 1987;

Urdinguio *et al.* 2009). The degree of DNA methylation in multiple genes expressed in the neuron cells from brain tissues of affecteds can be measured by qPCR using bisulfite converted cDNA (Dietrich *et al.* 2009). The purpose of bisulfite conversion step is to convert all unmethylated cytosines to thymines while methylated cytosines remain unchanged, so heavily-methylated regions can be identified by PCR and Sanger sequencing (Frommer *et al.* 1992). Bisulfite-treated DNA could be prepared with genomic DNA extracted from whole blood or with mRNA extracted from cerebellum tissues as the template. Once heavily-methylated regions were identified by Sanger sequencing, qPCR could be used to study the relationship between methylation and gene transcription.

Methylation-specific PCR (MSP) is a simple, sensitive and fast method that is commonly used to detect methylation of CpG islands in bisulfite-treated DNA (Herman *et al.* 1996). It needs small amount of cDNA generated from cerebellum tissue samples and requires the use of methylation-specific primers for amplification of unmethylated or methylated DNA sequences. A melting-curve generated from the quantitative MSP can be used to differentiate methylated from unmethylated sequences since methylated sequences have lower melting temperature due to the conversion of cytosines to thymines (Jorda and Peinado 2010).

4.1.6 Re-assessment of CA candidate region using Illumina high-density 172K array

A canine Affymetrix SNP array v2 with 127,000 SNPs combined with SNPs homozygosity analysis, was used to localise the candidate region associated with CA in Kelpies (Shearman *et al.* 2011). The candidate region for the gene causing CA was mapped to a 5 Mb (28 Mb to 33 Mb) region on chromosome 3 (NCBI Dog build 2.1)

(Shearman *et al.* 2011). There are 44 genes in the 5 Mb candidate region (Figure 1-3). Illumina high-density 172K array is one of the new tools that could be used in this study to fine-map the CA disease gene to a more precise region in the Kelpie genome by genotyping CA affecteds and unaffected controls for the extra SNPs present on the Illumina high-density 172K array. A refined candidate region could then be defined by combining the new SNP haplotype data with existing linkage data from microsatellites genotyping.

A smaller candidate region would limit the number of genes that needed to be further investigated by sequencing. Next generation sequencing on a platform that produced long reads (for example, the 454) could be employed to sequence the revised candidate region to look for sequence differences between CA affecteds and unaffected Kelpie with CA haplotype. PCR and Sanger sequencing may be required to fill in any remaining gaps in 454 sequences.

4.1.7 Sequence GC-rich regions using 454 Next Generation Sequencing technology

The dog genome was sequenced and assembled at the Broad Institute in 2005 using capillary Sanger sequencing technology with a 7.5X coverage and an averaged reads length of 800 bp (Lindblad-Toh *et al.* 2005). This was sufficient to obtain a reference dog genome sequence but it is estimated that there is still 3% of the dog genome that could not be assembled due to the presence of highly repetitive sequences such as microsatellites and minisatellites (Lindblad-Toh *et al.* 2005).

454 sequencing technology has been used in previous studies for sequencing the CA candidate region in Kelpies (Shearman 2011) since it has a longer read length (400-700 bp) when compared to other next generation sequencing platforms, and enables

much higher depth of coverage than Sanger sequencing for the same cost and effort (Schatz *et al.* 2010). Generally, longer reads allow better assembly of repetitive sequences since they can cover more repeats per read (Schatz *et al.* 2010). However one of the limitation of standard 454 sequencing is the inability of emulsion PCR (emPCR) to amplify GC rich sequences (Grossmann *et al.* 2011). This may explain the presence of several exonic and upstream regulatory region gaps in the 454 sequence that needed to be filled in by Sanger sequencing, since the majority of these were GC rich (Tables 3-5 to 3-8).

Most housekeeping genes, tumour-suppressor genes and approximately 40% of tissue-specific genes have high GC content (usually greater than 60%) in the upstream regulatory region and first exon (Hube *et al.* 2005). These GC rich regions contain important gene regulatory elements such as promoters, transcription factor binding sites and enhancers and so are closely related to gene regulation, transcriptional silencing and disease progression (Wilson *et al.* 1997). Therefore it is important to be able to detect mutations present in these GC rich regions to see how these mutations may affect function and regulation of candidate genes (Wei *et al.* 2010). These GC rich regions are also prone to DNA methylation and gene silencing (Section 4.1.5).

A new emulsion PCR condition was developed by Grossmann *et al.* (2011) to assist with amplification and sequencing of amplicons with GC content of up to 77%, by the 454 Next Generation Sequencing platform (Grossmann *et al.* 2011). This new emPCR condition replaced water with emPCR additive in the emPCR master mix, enabling a more robust detection of sequence differences in GC rich regions of genes (Grossmann *et al.* 2011). A combination of Sanger sequencing and modified 454 chemistry procedures could be used to obtain sequences of GC rich regions and to

correct and fill in sequence gaps. This may help to improve the quality of alignment and the depth of sequence coverage for exonic and upstream regulatory regions that have high GC content. Finally, it is likely improved sequencing technologies will be introduced within the next few years that could help resolve the issue associated with sequencing GC rich regions.

4.1.8 Limitations of the study

The control samples used in this project were taken from unaffected dogs that were more than five generations removed from a previously confirmed case of CA. The classification of dogs into different disease categories was solely based on physical clinical symptoms without being confirmed by histopathological analysis of the changes in brain tissues. Therefore mildly-affected dogs with CA genotype may have been mis-classified as phenotypically unaffected by their owners and vets.

It is also not possible to perform gene expression studies with mRNA extracted from brain tissues at this stage due to the lack of access to fresh/frozen cerebellar tissue samples from CA affected and unaffected control Kelpie dogs. Confirmation of the effect of a mutation on the level of gene transcription *in vivo* (using qPCR) would require mRNA from both CA affected and unaffected control dogs.

5 CONCLUSION

The 5 Mb candidate region on chromosome 3 of the dog genome was captured and sequenced previously on the 454 Next Generation Sequencing platform with two CA affecteds and one unaffected control Kelpie dogs. Assuming CA has complete penetrance in affected Kelpies, seventeen synonymous substitutions in coding exons and substitutions in the untranslated regions of mRNA were sequenced (with Sanger sequencing) but none were consistent with causing CA since they were also present in unaffected control dogs that possessed the same haplotype as affecteds. Twenty-two non-coding differences that were evolutionarily conserved between mammals and dogs were identified and sequenced using Sanger technology but most of them were eliminated as a possible cause of CA in Kelpies. However the possibility that CA does not have complete penetrance in affected Kelpie dogs could not be definitely excluded. One intergenic deletion (31674050) that is conserved between mammals and dogs was predicted to disrupt the HSF2 transcription factor binding site and could be causing CA in Kelpies.

Sanger sequencing was employed to fill in the 454 sequencing gaps in 40 exons and upstream regulatory region of genes. Several exonic and regulatory differences were identified and eliminated as possible causes of CA in Kelpies. Sequencing of the *DMGDH* gene in CA affected and unaffected control dogs has identified a 223 bp insertion upstream of last exon of the gene. This insert sequence contains two constitutive acceptor splice sites where the spliceosome may bind to produce a 72 amino acid truncated protein. However there is also a possibility that the two constitutive acceptor splice sites do not have any effect on the splicing of pre-mRNA and a normal

functional protein can still be produced. Gene expression studies are required to fully investigate this possibility.

Apart from looking for DNA sequence differences between affected and unaffected control dogs, other possibilities like epigenetic modification also need to be considered as a possible cause of CA in Kelpies. Variable onset age and severity of clinical symptoms suggest that CA in Kelpies could be caused by a regulatory mutation (epigenetics) rather than a change in the DNA sequence. Identifying the causative mutation caused by epigenetics may require detailed assessment of the methylation status of GC-rich gene regions and gene expression level using qPCR.

It is also noted that other methods such as Southern blotting may be needed in order to identify large insertion and deletion events caused by transposable elements or gene duplications that may be present in the candidate region. Additional sequencing of unaffected Kelpies with the same haplotype as affecteds (on the 454 platform) could also help to reduce the number of differences to investigate. The candidate region could be reduced by genotyping some CA affected and healthy dogs (previously processed on the Affymetrix canine SNP array v2) with new additional SNPs on the Illumina high-density 172K array.

6 REFERENCES

Abitbol M, Thibaud JL, Olby NJ, Hitte C, Puech JP, Maurer M, Pilot-Storck F, Hedan B, Dreano S, Brahimi S et al., 2010. A canine Arylsulfatase G (ARSG) mutation leading to a sulfatase deficiency is associated with neuronal ceroid lipofuscinosis. Proceedings of the National Academy of Sciences USA. 107:14775-80.

Abrams ZR, Zhang X, 2011. Signals and circuits in the purkinje neuron. Front Neural Circuits. 5:11.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ, 1990. Basic local alignment search tool. Journal of Molecular Biology. 215:403-10.

Amir RE, Van den Veyver IB, Wan M, Tran CQ, Francke U, Zoghbi HY, 1999. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nature Genetics. 23:185-188.

Asher L, Diesel G, Summers JF, McGreevy PD, Collins LM, 2009. Inherited defects in pedigree dogs. Part 1: disorders related to breed standards. The Veterinary Journal. 182:402-11.

Bejerano G, Pheasant M, Makunin I, Stephen S, Kent WJ, Mattick JS, Haussler D, 2004. Ultraconserved elements in the human genome. Science. 304:1321-5.

Bell JS, 2011. Researcher responsibilities and genetic counseling for pure-bred dog populations. The Veterinary Journal. 189:234-235.

Bellettato CM, Scarpa M, 2010. Pathophysiology of neuropathic lysosomal storage disorders. Journal of Inherited Metabolic Disease 33:347-62.

Berry ML, Blas-Machado U, 2003. Cerebellar abiotrophy in a miniature schnauzer. The Canadian Veterinary Journal. 44:657-9.

Bildfell RJ, Mitchell SK, de Lahunta A, 1995. Cerebellar cortical degeneration in a Labrador retriever. The Canadian Veterinary Journal. 36:570-2.

Binzak BA, Wevers RA, Moolenaar SH, Lee YM, Hwu WL, Poggi-Bach J, Engelke UF, Hoard HM, Vockley JG, Vockley J, 2001. Cloning of dimethylglycine dehydrogenase and a new human inborn error of metabolism, dimethylglycine dehydrogenase deficiency. American Journal of Human Genetics. 68:839-47.

Bird A, 2002. DNA methylation patterns and epigenetic memory. Genes and Development. 16:6-21.

Bird AP, 1986. CpG-rich islands and the function of DNA methylation. Nature. 321:209-13.

Blanco A, Moyano R, Vivo J, Flores-Acuna R, Molina A, Blanco C, Monterde JG, 2006. Purkinje cell apoptosis in arabian horses with cerebellar abiotrophy. Journal of Veterinary Medicine. A-Physiology, Pathology, Clinical Medicine. . 53:286-7.

Blumberg B, Wright CV, De Robertis EM, Cho KW, 1991. Organizer-specific homeobox genes in Xenopus laevis embryos. Science. 253:194-6.

Boffelli D, Nobrega MA, Rubin EM, 2004. Comparative genomics at the vertebrate extremes. Nature Reviews Genetics 5:456-465.

Brault LS, Cooper CA, Famula TR, Murray JD, Penedo MC, 2011. Mapping of equine cerebellar abiotrophy to ECA2 and identification of a potential causative mutation affecting expression of MUTYH. Genomics. 97:121-9.

Brenet F, Moh M, Funk P, Feierstein E, Viale AJ, Socci ND, Scandura JM, 2011. DNA methylation of the first exon is tightly linked to transcriptional silencing. PLoS One. 6:e14524.

Chang Y, Ostling P, Akerfelt M, Trouillet D, Rallu M, Gitton Y, El Fatimy R, Fardeau V, Le Crom S, Morange M et al., 2006. Role of heat-shock factor 2 in cerebral cortex formation and as a regulator of p35 expression. Genes and Development. 20:836-47.

Charbonneau N, Amunugama R, Schmutte C, Yoder K, Fishel R, 2009. Evidence that hMLH3 functions primarily in meiosis and in hMSH2-hMSH3 mismatch repair. Cancer Biology and Therapy. 8:1411-20.

Chase K, Jones P, Martin A, Ostrander EA, Lark KG, 2009. Genetic mapping of fixed phenotypes: disease frequency as a breed characteristic. Journal of Heredity. 100 Suppl 1:S37-41.

Clark LA, Wahl JM, Rees CA, Murphy KE, 2006. Retrotransposon insertion in SILV is responsible for merle patterning of the domestic dog. Proceedings of the National Academy of Sciences USA. 103:1376-81.

Cook RJ, Misono KS, Wagner C, 1984. Identification of the covalently bound flavin of dimethylglycine dehydrogenase and sarcosine dehydrogenase from rat liver mitochondria. The Journal of Biological Chemistry. 259:12475-80.

Cordaux R, Batzer MA, 2006. Teaching an old dog new tricks: SINEs of canine genomic diversity. Proceedings of the National Academy of Sciences USA. 103:1157-8.

Dean P, Porrill J, Ekerot C-F, Jorntell H, 2010. The cerebellar microcircuit as an adaptive filter: experimental and computational evidence. Nature Reviews Neuroscience. 11:30-43.

Deltagen. Cerebellum (Cortex) 10x [cited]. Available from: URL|.

Dennison EM, Syddall HE, Rodriguez S, Voropanov A, Day IN, Cooper C, 2004. Polymorphism in the growth hormone gene, weight in infancy, and adult bone mass. The Journal of Clinical Endocrinology & Metabolism 89:4898-903.

Dermitzakis ET, Reymond A, Antonarakis SE, 2005. Conserved non-genic sequences - an unexpected feature of mammalian genomes. Nature Reviews Genetics 6:151-7.

Dietrich D, Lesche R, Tetzner R, Krispin M, Dietrich J, Haedicke W, Schuster M, Kristiansen G, 2009. Analysis of DNA methylation of multiple genes in microdissected cells from formalin-fixed and paraffin-embedded tissues. Journal of Histochemistry & Cytochemistry. 57:477-89.

Duprey P, Chowdhury K, Dressler GR, Balling R, Simon D, Guenet JL, Gruss P, 1988. A mouse gene homologous to the Drosophila gene caudal is expressed in epithelial cells from the embryonic intestine. Genes and Development. 2:1647-54.

Frerman FE, Goodman SI, 1985. Fluorometric assay of acyl-CoA dehydrogenases in normal and mutant human fibroblasts. Biochemical Medicine. 33:38-44.

Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL, Paul CL, 1992. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proceedings of the National Academy of Sciences. 89:1827-1831.

Frumkin A, Rangini Z, Ben-Yehuda A, Gruenbaum Y, Fainsod A, 1991. A chicken caudal homologue, CHox-cad, is expressed in the epiblast with posterior localization and in the early endodermal lineage. Development. 112:207-19.

Gamer LW, Wright CV, 1993. Murine Cdx-4 bears striking similarities to the Drosophila caudal gene in its homeodomain sequence and early expression pattern. Mechanisms of Development. 43:71-81.

Gandini G, Botteron C, Brini E, Fatzer R, Diana A, Jaggy A, 2005. Cerebellar cortical degeneration in three English bulldogs: clinical and neuropathological findings. Journal of Small Animal Practice. 46:291-4.

German MS, Wang J, Chadwick RB, Rutter WJ, 1992. Synergistic activation of the insulin gene by a LIM-homeo domain protein and a basic helix-loop-helix protein: building a functional insulin minienhancer complex. Genes and Development. 6:2165-76.

Grossmann V, Schnittger S, Schindela S, Klein H-U, Eder C, Dugas M, Kern W, Haferlach T, Haferlach C, Kohlmann A, 2011. Strategy for robust detection of insertions, deletions, and point mutations in CEBPA, a GC-rich content gene, using 454 next-generation deep-sequencing technology. The Journal of Molecular Diagnostics. 13:129-136.

Gu Y, Parker A, Wilson TM, Bai H, Chang DY, Lu AL, 2002. Human MutY homolog, a DNA glycosylase involved in base excision repair, physically and functionally interacts with mismatch repair proteins human MutS homolog 2/human MutS homolog 6. The Journal of Biological Chemistry. 277:11135-42.

Heinemeyer T, Wingender E, Reuter I, Hermjakob H, Kel AE, Kel OV, Ignatieva EV, Ananko EA, Podkolodnaya OA, Kolpakov FA et al., 1998. Databases on transcriptional regulation: TRANSFAC, TRRD and COMPEL. Nucleic Acids Research. 26:362-7.

Herman JG, Graff JR, Myohanen S, Nelkin BD, Baylin SB, 1996. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. Proceedings of the National Academy of Sciences USA. 93:9821-6.

Higgins A, Nicholas FW, 2008. The breeding of pedigree dogs: time for strong leadership. The Veterinary Journal. 178:157-8.

Holliday R, 1987. DNA methylation and epigenetic defects in carcinogenesis. Mutation Research. 181:215-7.

Hube F, Reverdiau P, Iochmann S, Gruel Y, 2005. Improved PCR method for amplification of GC-rich DNA sequences. Molecular Biotechnology. 31:81-4.

Ishii Y, Fukuda K, Saiga H, Matsushita S, Yasugi S, 1997. Early specification of intestinal epithelium in the chicken embryo: a study on the localization and regulation of CdxA expression. Development, Growth and Differentiation. 39:643-53.

James R, Kazenwadel J, 1991. Homeobox gene expression in the intestinal epithelium of adult mice. The Journal of Biological Chemistry. 266:3246-51.

Jiang H, Orr A, Guernsey DL, Robitaille J, Asselin G, Samuels ME, Dube MP, 2009. Application of homozygosity haplotype analysis to genetic mapping with high-density SNP genotype data. PLoS One. 4:e5280.

John S, Shephard N, Liu G, Zeggini E, Cao M, Chen W, Vasavda N, Mills T, Barton A, Hinks A et al., 2004. Whole-genome scan, in a complex disease, using 11,245 single-nucleotide polymorphisms: comparison with microsatellites. American Journal of Human Genetics. 75:54-64.

Jokinen TS, Rusbridge C, Steffen F, Viitmaa R, Syrja P, De Lahunta A, Snellman M, Cizinauskas S, 2007. Cerebellar cortical abiotrophy in Lagotto Romagnolo dogs. Journal of Small Animal Practice. 48:470-3.

Joly JS, Maury M, Joly C, Duprey P, Boulekbache H, Condamine H, 1992. Expression of a zebrafish caudal homeobox gene correlates with the establishment of posterior cell lineages at gastrulation. Differentiation. 50:75-87.

Jorda M, Peinado MA, 2010. Methods for DNA methylation analysis and applications in colon cancer. Mutation Research. 693:84-93.

Kallio M, Chang Y, Manuel M, Alastalo TP, Rallu M, Gitton Y, Pirkkala L, Loones MT, Paslaru L, Larney S et al., 2002. Brain abnormalities, defective meiotic chromosome synapsis and female subfertility in HSF2 null mice. The EMBO Journal. 21:2591-601.

Karlsson EK, Baranowska I, Wade CM, Salmon Hillbertz NH, Zody MC, Anderson N, Biagi TM, Patterson N, Pielberg GR, Kulbokas EJ, 3rd et al., 2007. Efficient mapping of mendelian traits in dogs through genome-wide association. Nature Genetics. 39:1321-8.

Kelley LA, Sternberg MJ, 2009. Protein structure prediction on the Web: a case study using the Phyre server. Nature Protocols. 4:363-71.

Kemp J, McOrist S, Jeffrey M, 1995. Cerebellar abiotrophy in Holstein Friesian calves. Veterinary Record. 136:198.

Kirkness EF, Bafna V, Halpern AL, Levy S, Remington K, Rusch DB, Delcher AL, Pop M, Wang W, Fraser CM et al., 2003. The dog genome: survey sequencing and comparative analysis. Science. 301:1898-903.

Knapp M, Becker T, 2003. Family-based association analysis with tightly linked markers. Human Heredity. 56:2-9.

Knight JC, 2003. Functional implications of genetic variation in non-coding DNA for disease susceptibility and gene regulation. Clinical Science (London). 104:493-501.

Komatsu M, Waguri S, Chiba T, Murata S, Iwata J, Tanida I, Ueno T, Koike M, Uchiyama Y, Kominami E et al., 2006. Loss of autophagy in the central nervous system causes neurodegeneration in mice. Nature. 441:880-4.

Kwon JM, Goate AM, 2000. The candidate gene approach. Alcohol Research & Health 24:164-8.

Lang H, Minaian K, Freudenberg N, Hoffmann R, Brandsch R, 1994. Tissue specificity of rat mitochondrial dimethylglycine dehydrogenase expression. Biochemical Journal. 299:393-8.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R et al., 2007. Clustal W and Clustal X version 2.0. Bioinformatics. 23:2947-2948.

Laryea MD, Steinhagen F, Pawliczek S, Wendel U, 1998. Simple method for the routine determination of betaine and N,N-dimethylglycine in blood and urine. Clinical Chemistry. 44:1937-1941.

Le Marchand L, Donlon T, Seifried A, Kaaks R, Rinaldi S, Wilkens LR, 2002. Association of a common polymorphism in the human GH1 gene with colorectal neoplasia. Journal of the National Cancer Institute. 94:454-60.

Lequarre AS, Andersson L, Andre C, Fredholm M, Hitte C, Leeb T, Lohi H, Lindblad-Toh K, Georges M, 2011. LUPA: A European initiative taking advantage of the canine genome architecture for unravelling complex disorders in both human and dogs. The Veterinary Journal. 189:155-9.

Lewis BP, Burge CB, Bartel DP, 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell. 120:15-20.

Lin L, Faraco J, Li R, Kadotani H, Rogers W, Lin X, Qiu X, de Jong PJ, Nishino S, Mignot E, 1999. The sleep disorder canine narcolepsy is caused by a mutation in the hypocretin (orexin) receptor 2 gene. Cell. 98:365-76.

Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, Clamp M, Chang JL, Kulbokas EJ, 3rd, Zody MC et al., 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature. 438:803-19.

Liu W, Ding J, Gibbs JR, Wang SJ, Hardy J, Singleton A, 2009. A simple and efficient algorithm for genome-wide homozygosity analysis in disease. Molecular Systems Biology. 5:304.

Loewenstein Y, Mahon S, Chadderton P, Kitamura K, Sompolinsky H, Yarom Y, Hausser M, 2006. Purkinje cells in awake behaving animals operate in stable upstate membrane potential. Nature Neuroscience 9:461-461.

Madsen MB, Olsen LH, Haggstrom J, Hoglund K, Ljungvall I, Falk T, Wess G, Stephenson H, Dukes-McEwan J, Chetboul V et al., 2011. Identification of 2 Loci associated with development of myxomatous mitral valve disease in cavalier king charles spaniels. Journal of Heredity. 102 Suppl 1:S62-7.

Marra G, Iaccarino I, Lettieri T, Roscilli G, Delmastro P, Jiricny J, 1998. Mismatch repair deficiency associated with overexpression of the MSH3 gene. Proceedings of the National Academy of Sciences USA. 95:8568-73.

Matalon R, Arbogast B, Dorfman A, 1974. Deficiency of chondroitin sulfate N-acetylgalactosamine 4-sulfate sulfatase in Maroteaux-Lamy syndrome. Biochemical and Biophysical Research Communications. 61:1450-1457.

Mausberg TB, Wess G, Simak J, Keller L, Drogemuller M, Drogemuller C, Webster MT, Stephenson H, Dukes-McEwan J, Leeb T, 2011. A locus on chromosome 5 is associated with dilated cardiomyopathy in Doberman Pinschers. PLoS One. 6:e20042.

Melville SA, Wilson CL, Chiang CS, Studdert VP, Lingaas F, Wilton AN, 2005. A mutation in canine CLN5 causes neuronal ceroid lipofuscinosis in Border collie dogs. Genomics. 86:287-294.

Millar DS, Horan M, Chuzhanova NA, Cooper DN, 2010. Characterisation of a functional intronic polymorphism in the human growth hormone (GH1) gene. Human Genomics. 4:289-301.

Miller SA, Dykes DD, Polesky HF, 1988. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Research. 16:1215.

Mitchell PJ, Reilly W, Harper PA, McCaughan CJ, 1993. Cerebellar abiotrophy in Angus cattle. The Australian Veterinary Journal 70:67-8.

Mogensen MS, Karlskov-Mortensen P, Proschowsky HF, Lingaas F, Lappalainen A, Lohi H, Jensen VF, Fredholm M, 2011. Genome-wide association study in dachshund: identification of a major locus affecting intervertebral disc calcification. Journal of Heredity. 102 Suppl 1:S81-6.

Mount DW, 2007. Using the Basic Local Alignment Search Tool (BLAST). Cold Spring Harbor Protocols. 2007:pdb top17.

Neff MW, Robertson KR, Wong AK, Safra N, Broman KW, Slatkin M, Mealey KL, Pedersen NC, 2004. Breed distribution and history of canine mdr1-1Δ, a pharmacogenetic mutation that marks the emergence of breeds from the collie lineage. Proceedings of the National Academy of Sciences USA. 101:11725-11730.

Negrin A, Bernardini M, Baumgartner W, Castagnaro M, 2006. Late onset cerebellar degeneration in a middle-aged cat. The Journal of Feline Medicine and Surgery 8:424-9.

Neilan BA, Wilton AN, Jacobs D, 1997. A universal procedure for primer labelling of amplicons. Nucleic Acids Research. 25:2938-9.

Nibe K, Kita C, Morozumi M, Awamura Y, Tamura S, Okuno S, Kobayashi T, Uchida K, 2007. Clinicopathological features of canine neuroaxonal dystrophy and cerebellar cortical abiotrophy in Papillon and Papillon-related dogs. The Journal of Veterinary Medical Science. 69:1047-52.

Northrop JL, Kimelman D, 1994. Dorsal-ventral differences in Xcad-3 expression in response to FGF-mediated induction in Xenopus. Developmental Biology. 161:490-503.

O'Brien JF, Cantz M, Spranger J, 1974. Maroteaux-Lamy disease (mucopolysaccharidosis VI), subtype A: deficiency of a N-acetylgalactosamine-4-sulfatase. Biochem Biophys Res Commun. 60:1170-7.

Olby N, Blot S, Thibaud JL, Phillips J, O'Brien DP, Burr J, Berg J, Brown T, Breen M, 2004. Cerebellar cortical degeneration in adult American Staffordshire Terriers. Journal of Veterinary Internal Medicine. 18:201-8.

Orimo H, Nakajima E, Yamamoto M, Ikejima M, Emi M, Shimada T, 2000. Association between single nucleotide polymorphisms in the hMSH3 gene and sporadic colon cancer with microsatellite instability. Journal of Human Genetics. 45:228-230.

Orphanides G, Reinberg D, 2002. A unified theory of gene expression. Cell. 108:439-51.

Ostrander EA, Galibert F, Patterson DF, 2000. Canine genetics comes of age. Trends Genetics. 16:117-24.

Parker HG, Ostrander EA, 2005. Canine genomics and genetics: running with the pack. PLoS Genetics. 1:e58.

Parsons T 1992. *The Australian Kelpie: The essential guide to the Australian working dog*, Penguin Books Australia Ltd.

Patterson DF, 2000. Companion animal medicine in the age of medical genetics. Journal of Veterinary Internal Medicine. 14:1-9.

Patterson DF, Haskins ME, Jezyk PF, 1982. Models of human genetic disease in domestic animals. Advances in Human Genetics. 12:263-339.

Patterson DF, Haskins ME, Jezyk PF, Giger U, Meyers-Wallen VN, Aguirre G, Fyfe JC, Wolfe JH, 1988. Research on genetic diseases: reciprocal benefits to animals and man. Journal of the American Veterinary Medical Association. 193:1131-44.

Pavai Z, Pap Z, Orban-Kis K, Szilagyi T, 2010. Quantitative characterization of regional differences in the GABAA-receptor alpha1-subunit mRNA expression in the rat brain. Romanian journal of morphology and embryology 51:43-7.

Pele M, Tiret L, Kessler JL, Blot S, Panthier JJ, 2005. SINE exonic insertion in the PTPLA gene leads to multiple splicing defects and segregates with the autosomal recessive centronuclear myopathy in dogs. Human Molecular Genetics. 14:1417-27.

Phillips AR, Suttangkakul A, Vierstra RD, 2008. The ATG12-conjugating enzyme ATG10 Is essential for autophagic vesicle formation in Arabidopsis thaliana. Genetics. 178:1339-53.

Pierson TM, Adams D, Bonn F, Martinelli P, Cherukuri PF, Teer JK, Hansen NF, Cruz P, Mullikin For The Nisc Comparative Sequencing Program JC, Blakesley RW et al., 2011. Whole-exome sequencing identifies homozygous AFG3L2 mutations in a spastic ataxia-neuropathy syndrome linked to mitochondrial m-AAA proteases. PLoS Genet. 7:e1002325.

Powers ET, Morimoto RI, Dillin A, Kelly JW, Balch WE, 2009. Biological and chemical approaches to diseases of proteostasis deficiency. Annual Review of Biochemistry. 78:959-91.

Rice P, Longden I, Bleasby A, 2000. EMBOSS: the European molecular biology open software suite. Trends in Genetics. 16:276-277.

Richter K, Haslbeck M, Buchner J, 2010. The heat shock response: life on the verge of death. Molecular Cell. 40:253-66.

Rozen S, Skaletsky H, 2000. Primer3 on the WWW for general users and for biologist programmers. Methods in Molecular Biology. 132:365-86.

Sampson JR, Jones S, Dolwani S, Cheadle JP, 2005. MutYH (MYH) and colorectal cancer. Biochemical Society Transactions. 33:679-83.

Sandqvist A, Bjork JK, Akerfelt M, Chitikova Z, Grichine A, Vourc'h C, Jolly C, Salminen TA, Nymalm Y, Sistonen L, 2009. Heterotrimerization of heat-shock factors 1 and 2 provides a transcriptional switch in response to distinct stimuli. Molecular Biology of the Cell. 20:1340-7.

Sandy JR, Slocombe RE, Mitten RW, Jedwab D, 2002. Cerebellar abiotrophy in a family of Border Collie dogs. Veterinary Pathology. 39:736-8.

Sardiello M, Annunziata I, Roma G, Ballabio A, 2005. Sulfatases and sulfatase modifying factors: an exclusive and promiscuous relationship. Human Molecular Genetics. 14:3203-17.

Sato J, Sasaki S, Yamada N, Tsuchitani M, 2011. Hereditary cerebellar degenerative disease (Cerebellar cortical abiotrophy) in rabbits. Veterinary Pathology. 10.1177/0300985811402840:

Schatz MC, Delcher AL, Salzberg SL, 2010. Assembly of large genomes using second-generation sequencing. Genome Research. 20:1165-73.

Schrodinger, LLC 2010. The PyMOL Molecular Graphics System, Version 1.3r1.

Serrano J, Scavo L, Roth J, de la Rosa EJ, de Pablo F, 1993. A novel chicken homeobox-containing gene expressed in neurulating embryos. Biochemical and Biophysical Research Communications. 190:270-6.

Shaw G, Morse S, Ararat M, Graham FL, 2002. Preferential transformation of human neuronal cells by human adenoviruses and the origin of HEK 293 cells. The FASEB Journal. 16:869-71.

Shearman JR. 2011. *Canine disease gene identification.* Doctor of Philosophy, University of New South Wales.

Shearman JR, Cook RW, McCowan C, Fletcher JL, Taylor RM, Wilton AN, 2011. Mapping cerebellar abiotrophy in Australian Kelpies. Animal Genetics. 42:675-678.

Shearman JR, Lau VM, Wilton AN, 2008. Elimination of SETX, SYNE1 and ATCAY as the cause of cerebellar abiotrophy in Australian Kelpies. Animal Genetics. 39:573.

Shinkawa T, Tan K, Fujimoto M, Hayashida N, Yamamoto K, Takaki E, Takii R, Prakasam R, Inouye S, Mezger V et al., 2011. Heat shock factor 2 is required for maintaining proteostasis against febrile-range thermal stress and polyglutamine aggregation. Molecular Biology of the Cell. 22:3571-3583.

Simossis VA, Kleinjung J, Heringa J, 2005. Homology-extended sequence alignment. Nucleic Acids Research. 33:816-824.

Sixma TK, 2001. DNA mismatch repair: MutS structures bound to mismatches. Current Opinion in Structural Biology. 11:47-52.

Smit A, Hubley, R & Green, P. 1996-2010. RepeatMasker Open-3.0 [cited]. Available from: URL|.

Smyth JR, Jr., Jerszyk MM, Montgomery N, 1985. Congenital quiver, an inherited neurological defect in the chicken. Journal of Heredity. 76:263-6.

Solis AS, Shariat N, Patton JG, 2008. Splicing fidelity, enhancers, and disease. Front Biosci. 13:1926-42.

Strauch K, Fimmers R, Kurz T, Deichmann KA, Wienker TF, Baur MP, 2000. Parametric and nonparametric multipoint linkage analysis with imprinting and two-locus-trait models: application to mite sensitization. American Journal of Human Genetics 66:1945-57.

Sutter NB, Eberle MA, Parker HG, Pullar BJ, Kirkness EF, Kruglyak L, Ostrander EA, 2004. Extensive and breed-specific linkage disequilibrium in Canis familiaris. Genome Research. 14:2388-96.

Tatalick LM, Marks SL, Baszler TV, 1993. Cerebellar abiotrophy characterized by granular cell loss in a Brittany. Veterinary Pathology 30:385-8.

Thomas JB, Robertson D, 1989. Hereditary cerebellar abiotrophy in Australian kelpie dogs. Australian Veterinary Journal. 66:301-2.

Tournamille C, Colin Y, Cartron JP, Le Van Kim C, 1995. Disruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy-negative individuals. Nature Genetics. 10:224-8.

Udalova IA, Richardson A, Denys A, Smith C, Ackerman H, Foxwell B, Kwiatkowski D, 2000. Functional consequences of a polymorphism affecting NF-kappaB p50-p50 binding to the TNF promoter region. Molecular and Cellular Biology. 20:9113-9.

Urdinguio RG, Sanchez-Mut JV, Esteller M, 2009. Epigenetic mechanisms in neurological diseases: genes, syndromes, and therapies. The Lancet Neurology. 8:1056-72.

Vermeer S, van de Warrenburg BP, Willemsen MA, Cluitmans M, Scheffer H, Kremer BP, Knoers NV, 2011. Autosomal recessive cerebellar ataxias: the current state of affairs. Journal of Medical Genetics. 48:651-9.

Vila C, Savolainen P, Maldonado JE, Amorim IR, Rice JE, Honeycutt RL, Crandall KA, Lundeberg J, Wayne RK, 1997. Multiple and ancient origins of the domestic dog. Science. 276:1687-9.

Wagner C, Briggs WT, Cook RJ, 1984. Covalent binding of folic acid to dimethylglycine dehydrogenase. Archives of Biochemistry and Biophysics 233:457-61.

Wang G, Zhang J, Moskophidis D, Mivechi NF, 2003. Targeted disruption of the heat shock transcription factor (hsf)-2 gene results in increased embryonic lethality, neuronal defects, and reduced spermatogenesis. Genesis. 36:48-61.

Wang M, Marin A, 2006. Characterization and prediction of alternative splice sites. Gene. 366:219-27.

Wang W, Kirkness EF, 2005. Short interspersed elements (SINEs) are a major source of canine genomic diversity. Genome Research. 15:1798-808.

Wang Z, Burge CB, 2008. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. RNA. 14:802-13.

Wei M, Deng J, Feng K, Yu B, Chen Y, 2010. Universal method facilitating the amplification of extremely GC-rich DNA fragments from genomic DNA. Analytical Chemistry. 82:6303-6307.

Willoughby K, Kelly DF, 2002. Hereditary cerebellar degeneration in three full sibling kittens. Veterinary Record. 151:295-8.

Wilson AG, Symons JA, McDowell TL, McDevitt HO, Duff GW, 1997. Effects of a polymorphism in the human tumor necrosis factor α promoter on transcriptional activation. Proceedings of the National Academy of Sciences. 94:3195-3199.

Wu C, Orozco C, Boyer J, Leglise M, Goodale J, Batalov S, Hodge CL, Haase J, Janes J, Huss JW, 3rd et al., 2009. BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. Genome Biology. 10:R130.

Yasuba M, Okimoto K, Iida M, Itakura C, 1988. Cerebellar cortical degeneration in beagle dogs. Veterinary Pathology. 25:315-7.

Yue Q, Apprey V, Bonney GE, 2005. Which strategy is better for linkage analysis: single-nucleotide polymorphisms or microsatellites? Evaluation by identity-by-state-identity-by-descent transformation affected sib-pair method on GAW14 data. BMC Genetics. 6 Suppl 1:S16.

Zhu M, Zhao S, 2007. Candidate gene identification approach: progress and challenges. International Journal of Biological Sciences. 3:420-7.

Zuker M, Stiegler P, 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Research. 9:133-148.

7 APPENDICES

7.1 Genotyping of Australian Kelpies with long and short microsatellite markers

Table 7-1: Genotyping of long microsatellite markers in CA candidate region.

Sample ID	C3.2871		C3.2997		C3.3193		C3.3274		C3.3301	
WK6001	367	371	301	301	387	390	500	508	474	478
WK6002	367	371	301	301	387	387	504	508	474	474
WK6003	367	371	297	301	387	390	500	508	474	478
WK6004	367	371	301	301	387	387	504	508	474	474
WK6005	371	371	301	301	387	390	500	504	474	478
WK6006	363	367	301	301	383	387	508	508	474	474
WK6007	371	371	301	305	390	390	500	500	478	478
WK6008	371	394	251	301	356	387	417	504	474	474
WK6009	371	371	301	305	387	390	500	504	474	478
WK6010	367	375	301	305	387	394	490	508	474	483
WK6011	371	375	301	305	390	390	494	500	478	488
WK6012	371	375	301	305	390	390	500	508	474	478
WK6013	371	375	301	305	390	390	494	500	483	483
WK6014	363	371	301	301	383	390	500	508	474	483
WK6015	371	375	301	305	390	390	494	500	478	483
WK6016	363	371	301	301	383	390	500	508	474	483
WK6017	363	371	301	301	383	390	500	508	474	478
WK6018	375	375	-	-	-	-	490	494	-	-
WK6019	363	367	301	301	383	387	504	508	474	474
WK6020	367	367	297	301	387	394	500	508	474	478
WK6021	363	367	301	301	383	387	504	508	474	474
WK6022	367	367	301	301	383	387	504	508	474	474
WK6023	356	371	305	305	390	390	-	-	483	483
WK6024	356	367	301	305	387	390	486	504	474	483
WK6025	360	371	305	305	390	390	490	490	483	483
WK6026	367	367	301	301	387	387	504	504	474	474
WK6027	367	371	301	305	387	390	490	504	474	483
WK6028	356	367	301	305	387	390	490	504	474	483
WK6029	363	367	301	301	387	387	508	508	474	474
WK6030	367	367	301	301	387	387	504	504	474	474
WK6032	363	367	297	301	387	394	500	508	474	478
WK6033	371	375	262	309	387	394	430	494	469	478
WK6034	371	371	262	301	387	390	430	500	478	478
WK6035	367	367	301	301	383	387	504	508	474	474
WK6036	375	378	305	309	394	412	494	520	469	483
WK6037	378	378	301	305	390	412	500	520	478	483
WK6038	371	378	262	305	387	412	430	520	478	483
WK6040	356	367	282	301	387	399	417	504	474	478
WK6041	363	371	301	305	387	390	490	508	474	483
WK6042	356	371	301	305	390	390	490	500	483	483
WK6043	356	375	258	305	387	390	430	490	478	483
WK6044	371	371	297	309	387	394	494	504	474	483
WK6045	371	375	258	301	387	390	430	500	478	483

Sample ID	C3.2871		C3.2997		C3.3193		C3.3274		C3.3301	
WK6046	340	363	265	301	387	390	508	508	469	474
WK6047	340	371	265	301	390	390	500	508	469	478
WK6048	340	375	258	265	390	390	430	508	469	478
WK6049	371	375	258	301	387	390	430	500	478	478
WK6050	367	371	301	309	387	399	500	504	474	478
WK6054	371	378	305	309	390	394	490	500	483	483
WK6055	363	371	301	301	387	387	504	508	474	474
WK6056	371	371	301	301	390	394	500	500	478	478
WK6058	344	375	270	305	390	394	490	508	469	483
WK6059	371	375	305	305	390	390	486	490	483	483
WK6060	367	375	301	305	390	416	363	500	478	509
WK6061	363	367	301	301	390	390	500	504	474	478
WK6062	344	371	262	301	394	412	500	537	474	478
WK6063	371	375	301	305	390	390	494	500	478	483
WK6064	375	375	305	305	390	390	490	494	-	-
WK6065	363	371	301	301	390	390	500	504	474	478
WK6066	363	371	301	301	390	390	500	504	474	478
WK6067	-	-	301	305	-	-	-	-	-	-
WK6070	371	375	301	305	390	390	490	500	478	483
WK6071	371	375	301	305	390	390	494	500	478	483
WK6072	375	404	254	301	394	394	387	504	474	478
WK6073	340	371	265	301	390	390	500	504	469	478
WK6075	367	400	254	279	399	422	340	340	464	478
WK6076	367	404	254	279	399	433	340	340	464	464
WK6077	400	404	251	254	422	426	336	340	-	-
WK6078	367	408	270	309	390	404	417	441	478	478
WK6079	367	404	262	305	387	390	435	486	474	483
WK6082	363	367	274	282	394	412	376	376	464	524
WK6083	340	340	265	265	390	394	508	517	469	469
WK6084	340	371	265	309	394	394	494	512	469	483
WK6085	340	371	265	309	394	394	-	-	469	483
WK6086	-	-	265	309	387	390	502	508	-	-
WK6088	340	371	265	309	387	394	504	520	469	474
WK6091	371	371	279	305	390	399	423	490	474	483
WK6099	363	367	258	301	356	390	423	508	469	469
WK6102	371	375	305	305	390	394	486	490	483	488
WK6110	394	404	254	261	390	390	423	499	483	483
WK6111	404	404	261	261	390	390	499	499	483	483
WK6112	394	404	251	261	390	390	423	499	-	-
WK6124	371	375	305	305	390	390	-	-	483	483
WK6126	367	371	301	305	383	387	-	-	-	-
WK6142	363	371	301	301	387	390	499	508	474	478
WK6143	371	375	301	305	390	390	490	499	478	483
WK6144	371	375	301	301	390	390	499	499	478	478
WK6145	371	375	301	301	390	390	494	499	478	478
WK6146	371	375	301	301	-	-	499	499	478	478
WK6147	363	371	301	305	387	390	490	508	474	483
WK6148	363	375	301	301	-	-	-	-	-	-
WK6149	363	371	301	305	387	390	490	508	474	483
WK6181	375	375	301	305	-	-	490	490	-	-
WK6182	371	371	301	305	-	-	490	502	478	483
WK6183	367	371	301	301	-	-	-	-	-	-
WK6184	371	371	301	305	-	-	490	499	478	483

Sample ID	C3.2871		C3.2997		C3.3193		C3.3274		C3.3301	
WK6185	-	-	301	301	-	-	-	-	-	-
WK6190	360	363	270	274	402	404	374	408	-	-
WK6191	360	404	261	274	404	426	338	374	-	-
WK6192	360	367	274	279	404	419	374	450	474	474
WK6193	363	367	270	279	402	419	408	450	474	474
WK6194	363	404	261	270	402	426	338	408	474	474
WK6195	363	375	301	301	387	387	502	512	474	474
WK6196	363	371	301	301	387	387	502	512	474	474
WK6203	371	375	301	313	390	390	490	499	478	483
WK6204	363	371	301	301	383	390	499	508	474	478
WK6205	388	400	259	262	390	412	499	533	464	474
WK6206	375	378	305	305	390	409	490	517	483	483
WK6212	363	371	301	305	383	390	490	508	474	483
WK6213	363	371	301	305	387	390	490	508	474	483
WK6214	347	378	259	262	408	419	364	502	461	474
WK6222	371	371	301	301	387	387	502	502	469	474
WK6223	375	378	262	301	390	419	-	-	-	-
WK6238	360	378	262	274	387	419	360	502	-	-
WK6243	360	378	262	274	387	419	360	502	-	-
WK6249	367	375	301	301	387	390	-	-	-	-
WK6251	378	400	262	274	404	419	360	374	461	461
WK6253	360	367	274	301	387	387	502	502	474	474
WK6254	375	378	262	301	390	419	360	499	-	-
WK6256	371	408	265	301	390	404	419	499	478	478
WK6257	371	375	305	309	394	394	490	494	483	483
WK6258	375	400	254	305	387	394	490	502	474	483
WK6259	371	389	254	301	387	430	343	502	474	478
WK6260	371	400	254	309	394	430	343	494	478	483
WK6261	371	389	254	309	394	430	343	494	478	483
WK6263	371	378	305	305	390	390	486	490	483	483
WK6266	363	367	297	301	387	394	-	-	474	474
WK6267	-	-	297	297	387	394	499	499	-	-
WK6268	367	367	297	301	394	394	499	499	-	-
WK6270	371	371	301	301	387	387	508	508	474	474
WK6271	-	-	301	301	387	387	508	508	474	474
WK6272	371	375	301	305	387	387	508	508	474	483
WK6273	-	-	301	305	387	387	486	508	474	483
WK6274	363	371	301	301	383	387	-	-	-	-

Table 7-2: Genotyping of short microsatellite markers in CA candidate region.

Sample ID	C3.2805		C3.2856		C3.2869		C3.2984		C3.3228		C3.3265	
WK6001	208	220	262	262	357	357	309	309	231	231	252	252
WK6002	208	208	262	262	357	357	309	309	231	231	252	252
WK6003	208	220	262	262	357	357	309	309	231	231	252	252
WK6004	208	208	262	262	357	357	309	309	231	231	252	252
WK6005	208	220	262	262	357	357	309	309	231	231	252	252
WK6006	208	208	262	262	357	357	309	309	231	231	252	252
WK6007	208	220	262	262	357	359	309	309	231	231	252	252
WK6008	208	216	252	262	349	357	292	309	-	-	-	-
WK6009	208	208	262	262	357	359	309	309	231	231	252	252
WK6010	208	220	262	262	357	357	309	309	231	231	252	252
WK6011	208	220	262	262	357	359	309	309	231	231	252	252
WK6012	208	220	262	262	357	359	309	309	231	231	252	252
WK6013	220	220	262	262	357	357	309	309	231	231	252	252
WK6014	208	220	262	262	357	357	309	309	231	231	252	252
WK6015	208	220	262	262	357	359	309	309	231	231	252	252
WK6016	208	220	262	262	357	357	309	309	231	231	252	252
WK6017	208	208	262	262	357	359	309	309	231	231	252	252
WK6018	220	220	262	262	357	357	309	309	231	231	-	-
WK6019	208	208	262	262	357	357	309	309	231	231	252	252
WK6020	208	208	262	262	357	357	309	309	231	231	252	252
WK6021	208	208	262	262	357	357	309	309	231	231	252	252
WK6022	208	208	262	262	357	357	309	309	231	231	252	252
WK6024	208	210	248	262	357	365	309	309	231	231	252	252
WK6025	210	220	248	262	357	365	309	309	231	231	252	252
WK6026	208	210	262	262	357	357	309	309	231	231	252	252
WK6027	208	220	262	262	357	357	309	309	231	231	252	252
WK6028	210	210	248	262	357	365	309	309	231	231	252	252
WK6029	208	208	262	262	357	357	309	309	231	231	252	252
WK6030	208	210	262	262	357	357	309	309	231	231	252	252
WK6031	-	-	-	-	361	363	-	-	-	-	-	-
WK6032	208	208	262	262	357	357	309	309	231	231	252	252
WK6033	220	220	262	266	357	357	292	309	230	231	252	252
WK6034	208	220	262	266	357	363	292	292	230	231	-	-
WK6035	208	208	262	262	357	357	309	309	231	231	252	252
WK6036	208	220	248	262	357	365	292	309	231	231	252	252
WK6037	208	208	248	262	363	365	292	292	231	231	252	252
WK6038	208	220	248	266	357	365	292	292	230	231	252	252
WK6039	-	-	-	-	-	-	-	-	-	-	252	252
WK6040	208	210	248	262	357	365	292	309	217	231	252	252
WK6041	208	220	262	262	357	357	309	309	231	231	252	252
WK6042	210	220	248	262	357	365	309	309	231	231	252	252
WK6043	210	220	248	266	357	365	292	309	230	231	252	252
WK6044	208	220	262	262	357	357	309	309	231	231	252	252
WK6045	220	220	262	266	357	357	292	309	230	231	252	252
WK6046	208	218	256	262	357	359	292	309	231	231	252	252
WK6047	218	220	256	262	357	359	292	309	231	231	252	252
WK6048	218	220	256	266	357	359	292	292	230	231	252	252
WK6049	220	220	262	266	357	357	292	309	230	231	252	252
WK6050	220	220	258	262	357	357	309	309	231	231	-	-
WK6054	220	220	262	262	357	357	309	309	231	231	252	252

Sample ID	C3.2805		C3.2856		C3.2869		C3.2984		C3.3228		C3.3265	
WK6055	208	220	262	262	357	357	309	309	231	231	252	252
WK6056	208	216	262	262	357	357	309	309	231	231	252	252
WK6058	218	220	256	262	357	359	292	309	231	231	252	252
WK6059	220	220	262	262	357	357	309	309	231	231	252	252
WK6060	208	208	262	262	357	357	309	309	231	234	252	252
WK6061	208	208	262	262	357	357	303	309	231	231	252	252
WK6062	218	220	256	262	357	359	292	309	231	231	252	252
WK6063	208	220	262	262	357	357	309	309	231	231	252	252
WK6064	220	220	262	262	357	357	309	309	231	231	252	252
WK6065	208	208	262	262	357	357	303	309	231	231	252	252
WK6066	208	208	262	262	357	357	303	309	231	231	252	252
WK6067	208	220	262	262	357	357	-	-	231	231	-	-
WK6070	208	220	262	262	357	357	309	309	231	231	252	252
WK6071	208	220	262	262	357	357	309	309	231	231	252	252
WK6072	216	220	260	262	357	357	292	309	231	238	252	252
WK6073	208	218	256	262	357	359	292	309	231	231	252	252
WK6074	216	223	248	258	361	365	292	292	231	242	252	252
WK6075	216	216	248	260	357	363	292	292	217	238	252	262
WK6076	216	216	248	260	357	363	292	292	217	238	252	262
WK6077	216	216	260	260	357	357	292	292	238	238	252	252
WK6078	214	220	262	262	357	357	292	313	217	231	252	252
WK6079	214	220	262	264	355	357	292	309	230	231	252	252
WK6082	216	220	248	248	363	363	292	292	217	240	252	262
WK6083	218	218	256	256	359	359	292	292	231	231	252	252
WK6084	218	220	256	262	357	359	292	309	231	231	252	252
WK6085	218	220	256	262	357	359	292	309	231	231	252	252
WK6086	218	220	256	262	357	359	292	309	231	231	252	252
WK6088	218	220	256	262	357	359	292	309	231	231	252	252
WK6091	208	220	262	262	357	363	292	309	217	231	252	252
WK6099	208	220	262	262	357	361	292	309	231	238	252	252
WK6102	220	220	262	262	357	357	309	309	231	231	252	252
WK6110	216	216	252	260	349	357	292	292	231	238	252	252
WK6111	216	216	260	260	357	357	292	292	231	231	252	252
WK6112	216	216	252	260	349	357	292	292	231	238	252	252
WK6124	220	220	262	262	357	357	309	309	231	231	252	252
WK6142	208	220	262	262	357	357	309	309	231	231	252	252
WK6143	220	220	262	262	357	357	309	309	231	231	252	252
WK6144	220	220	262	262	357	357	309	309	231	231	252	252
WK6145	220	220	262	262	357	357	309	309	231	231	252	252
WK6146	220	220	262	262	357	357	309	309	231	231	252	252
WK6147	208	220	262	262	357	357	309	309	231	231	252	252
WK6148	208	220	262	262	357	357	309	309	231	231	252	252
WK6149	208	220	262	262	357	357	309	309	231	231	252	252
WK6181	220	220	262	262	357	357	309	309	231	231	252	252
WK6182	208	220	262	262	357	357	309	309	231	231	252	252
WK6183	208	208	262	262	357	357	309	309	231	231	252	252
WK6184	220	220	262	262	357	357	309	309	231	231	252	252
WK6185	208	216	248	248	357	357	306	309	230	230	-	-
WK6190	208	220	248	248	365	369	292	292	230	240	242	252
WK6191	208	216	248	260	357	369	292	292	238	240	252	252
WK6192	208	216	248	248	363	369	292	292	230	240	252	252
WK6193	216	220	248	248	363	365	292	292	230	230	242	252
WK6194	216	220	248	260	357	365	292	292	230	238	242	252

Sample ID	C3.2805		C3.2856		C3.2869		C3.2984		C3.3228		C3.3265	
WK6195	208	208	262	262	357	357	309	309	231	231	252	252
WK6196	208	208	262	262	357	357	309	309	231	231	252	252
WK6203	220	220	262	262	357	357	309	309	231	231	252	252
WK6204	208	220	262	264	355	355	309	309	231	231	252	252
WK6205	214	222	262	264	357	361	292	292	230	231	242	254
WK6206	208	220	248	262	357	365	292	292	231	231	252	254
WK6212	208	220	262	262	357	357	309	309	231	231	252	252
WK6213	208	220	262	262	357	357	309	309	231	231	252	252
WK6214	216	218	256	258	361	361	292	292	231	242	252	252
WK6222	208	208	262	262	357	357	309	309	231	231	252	252
WK6253	208	208	-	-	357	369	292	309	231	231	252	252
WK6254	216	220	258	262	357	361	292	309	231	242	252	252
WK6256	208	214	262	262	357	357	292	309	217	231	252	252
WK6257	220	220	-	-	357	357	309	309	231	231	252	252
WK6258	216	220	-	-	357	357	292	309	231	231	252	252
WK6259	208	216	260	262	357	357	292	309	231	238	252	252
WK6260	216	220	260	262	357	357	292	309	231	238	252	262
WK6261	216	220	260	262	357	357	292	309	231	238	252	252
WK6263	220	220	262	262	357	357	309	309	231	231	252	252

7.2 SNP haplotype

Table 7-3: Genotype results for the 5 Mb CFA3 interval for 16 new genetic markers and 10 SNPs from Shearman *et al.* (2011) in 23 CA affected Australian Kelpie dogs (including the 11 previously published Kelpies in Shearman *et al.* (2011)).

Kelpie ID numbers are listed across the top with each sample occupying a single column and the markers and genomic location on CFA3 are listed per row. Markers that are homozygous are indicated by a single entry and markers that are heterozygous are showed as allele1/allele2 and are shaded.

		Kelpie ID number																					
Marker	CFA3 location	6001	6003	6025	6044	6050	6054	6065	6066	6067	6124	6149	6022	6181	6182	6183	6184	6195	6203	6204	6212	6222	6253	6263
chr3.28056759	28056759	C	C	C/T	C/T	C	C	T	T	C/T	C	C/T	C/T	C	C/T	C/T	C	C/T	C	C	C/T	T	C	C
chr3.28354775	28354775	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
chr3.29567856	29567856	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	C/T	T
chr3.29632686	29632686	T	T	C	T	C/T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	C	T
chr3.29818867	29818867	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	A/G	G
chr3.29837088	29837088	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	C/T	T
chr3.29949728	29949728	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
chr3.30059175	30059175	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A/G	A
chr3.30173203	30173203	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	A/G	G
chr3.30364362	30364362	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
chr3.30597963	30597963	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
chr3.30616120	30616120	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
chr3.30616445	30616445	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
chr3.30616525	30616525	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
chr3.30616716	30616716	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
chr3.30684786	30684786	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
chr3.30691196	30691196	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
chr3.30704150	30704150	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
chr3.30738644	30738644	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
chr3.30738713	30738713	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
chr3.30782110	30782110	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
chr3.30917504	30917504	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
chr3.31163217	31163217	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
chr3.32292529	32292529	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
chr3.33061677	33061677	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T

7.3 Clinical symptoms checklist for CA (ataxia) in Australian Kelpies

Clinical Signs Checklist

- In general, affected dogs are normal at birth with progressive clinical signs of cerebellar dysfunction develop at an age of 3 to 7 weeks
- Most dogs suspected with Cerebellar Abiotrophy were diagnosis as soon as physical symptoms become apparent
- Dogs with Cerebellar Abiotrophy have ataxia as the main symptom, with slow progression of disease
- The severity of clinical signs between affected dogs may vary from mild to severe

Dog Owner Questionnaire	**Y**	**N**
Is your dog less than 7 weeks old when you first noticed his/her unusual physical behaviour and body coordination?		
Does your dog easily jump up onto or over objects or does it hit it legs when attempted to do so?		
Has your dog had a head tremor? If so, how often and how severe is the shaking? *Comment:*		
Does your dog walk normally or does it have an unusual or unsteady walk? Does it fall regularly when chasing for objects?		
Does your dog lift its front legs more so than normal dogs when walking?		
Does your dog have more, less or a normal amount of muscle in the leg?		
Have you noticed changes in your dog's standing position? Does it have the wide legged stance in the rear legs?		
Has your dog had difficulty while drinking water from a bowl?		
Does your dog overheat after doing a small amount of exercise (much less than what an average healthy dog can easily handle) to the point where it has to lay down and rest?		
Does your dog overheat while sleeping in the summer and shivering while sleeping indoor in the winter?		
Has your dog ever had any small seizure? If so how often, how severe and when did it start? *Comment:*		

Physical Examination Checklist			
Clinical Sign	**(Y/N)**	**Clinical Sign**	**(Y/N)**
Intention head tremor		Absent perineal reflex	
Head tilt and/or wide base stance in the rear legs		High stepping gait	
Abnormal body weight/ body size		Stumble and fall when jumping up a step	
Poor postural reactions when hopping		Ataxia in all limbs at time of euthanasia	

Figure 7-1: Clinical signs checklist for Cerebellar abiotrophy in Australian Kelpies.

The checklist was compiled with clinical symptoms observed and provided (with the samples) by the owners/vets of CA affected dogs.