Investigating the Genetic Basis of Cranial Cruciate Ligament Rupture in the Newfoundland Dog

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy

by

Arabella Elizabeth Gardiner Baird

University of Liverpool

September 2013

Acknowledgements

I would like to thank each of my supervisors Stuart Carter, Andrea Short and John Innes for their help, support and guidance throughout my PhD. My thanks also go to all of the staff who worked with me in the Department of Infection Biology (Institute of Infection and Global Health) at the University of Liverpool, especially to Di Isherwood and Nick Evans. Acknowledgements also go to Dr Ana Alfirevic and Dr Dan Carr in the pharmacology department, for allowing me access to the Sequenom equipment and training me to use it.

I would also like to thank all owners and vets who collected and submitted samples to the project (which would not have been able to take place without them).

Grateful thanks goes to the UK Kennel Club Charitable trust for funding this project and my stipend.

Thanks also goes to the Newfoundland Dog Club of Canada , especially the BC Region for the kind donation of money to purchase a new laptop – indispensable in my PhD writing/data analysis etc. and for their enthusiasm in the project and in sample collection.

I would also like to thank all of my family and friends for their constant support and encouragement over the years; this wouldn't have been possible without you!

Table of Contents

List of tables

List of figures

List of Appendixes

Abstract

This thesis presents work to examine the genetic basis of CCL rupture in dogs. It describes research to identify causative mutations that will help to develop a genetic screening test to identify dogs that have a high risk of developing CCL rupture.

Cranial Cruciate Ligament (CCL) rupture is the most common cause of hind limb lameness in dogs and is especially common in large and giant breeds such as Newfoundlands, Rottweilers and Staffordshire Bull Terriers.

CCL rupture cases and from two continents (Europe n = 48 and North America n = 48) were examined using genome wide association studies (GWAS). A candidate gene study was performed using Sequenom iPlex genotyping, on cases and controls from Newfoundlands (99 cases, 172 controls) and three other susceptible breeds: Labradors (124 cases, 165 controls), Rottweilers (57 cases, 81 controls) and Staffordshire Bull Terriers (13 cases, 38 controls). One hundred and eighty-six SNPs across 26 candidate genes were investigated for association with CCL rupture. To investigate downstream events, gene expression was compared between healthy and CCL rupture tissue. An *in-vitro* laboratory model of CCL rupture was investigated by examining ligamentocytes induced with and without TNFα. To investigate whether there was an auto-immune component to CCL rupture, the two main loci of the dog leucocyte antigen (DLA) system were assessed for disease association.

Principle component analysis of the GWAS data revealed population stratification within the Newfoundland breed, indicative of the continent of origin (Europe or North America). GWAS identified three main regions associated with CCL rupture (on chromosomes 1, 3 and 33). Significantly associated genes SORCS2 and SEMA5B function in neurological pathways; this may indicate that mechanotransduction, neurological and neuromuscular pathways play an important role in the pathogenesis and susceptibility to CCL rupture. Candidate gene analysis identified associations with two collagen genes (collagen type-V and collagen type-I) and three extracellular matrix proteins; Aggrecan (ACAN), Opticin (OPTC) and Latent transforming growth factor beta 2 (LTBP2). Gene expression analysis revealed significant differential expressions in COL1A1 and COL1A2. These results indicate that the strength and stability of the ligament is probably important in susceptibility to CCL rupture. Gene expression results also revealed that genes involved in degradation (TRAP and DIRC2) are upregulated, indicating that the cells are trying to repair themselves whilst a simultaneous degradative process is still on going. The TNFα model may be used as an *in-vitro* model to study CCL rupture, but may be more useful as a model for examining changes that occur after CCL rupture rather than the early stages of the disease. We showed no association with the DLA region and CCL rupture.

The identified associated regions should be further investigated and refined using next generation, targeted re-sequencing and transcriptomic approaches. This could identify the specific causative mutations involved in CCL rupture susceptibility. This study confirms that there are complex multigenetic and environmental factors involved in CCL rupture susceptibility. The work has contributed to the understanding of causative factors involved in CCL rupture susceptibility and may be an important step in the development of a screening test(s) to reduce the incidence of CCL rupture in dogs and thus improve their health and welfare.

List of abbreviations

The following lists the commonly used abbreviations in this thesis:

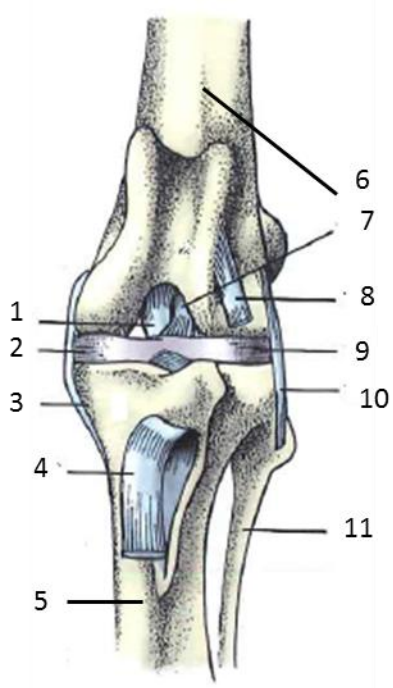
ACL	Anterior Cruciate ligament (human analogue of the CCL)
BONF	Bonferroni multiple correction
BONF *p*	Bonferroni multiple correction p-value
CCL	Cranial Cruciate Ligament
cDNA	Complementary DNA
Chr	Chromosome
CMH	Cochran-Mantel-Haenszel test for association
CT	Threshold cycle
DLA	Dog Leukocyte Antigen
DMEM	Dulbecco's modified Eagle's medium
DNA	Deoxyribonucleic acid
ECM	Extracellular matrix
EDTA	Ethylene-diamine-tetra-acetic acid
EMMAX	Efficient Mixed-Model Association eXpedited analysis
EU	European
GWAS	Genome Wide Association Study
HWE	Hardy Weinberg equilibrium
INDEL	Insertion or Deletion of a base pair in the DNA sequence
Kb	Kilobase
LD	Linkage disequilibrium
MAF	Minor allele frequency
Mb	Megabase
MHC	Major Histocompatibility Complex
NA	North American
OR	Odds ratio
p	p-value
PCA	Principle component analysis
PCR	Polymerase chain reaction
Pos	Position
QA	Quality Assurance
Q-PCR	Quantitative real time PCR
QQ Plot	Quantile-Quantile plot
RNA	Ribonucleic acid
SAP	Shrimp alkaline phosphatase
SEM	Standard error of the mean
SNP	Single Nucleotide Polymorphism
TNFα	Tumour necrosis factor alpha
VLA	Veterinary Laboratory Agency

Chapter 1.

Introduction

Cruciate ligaments – anatomy and function

Cranial cruciate ligament (CCL) rupture is the most common musculoskeletal condition that requires orthopaedic surgery in dogs *(Canis familiaris*). Complete rupture or partial tearing of the CCL is the most common cause of pelvic limb (hind leg) lameness [1-6].



1. **Caudal Cruciate Ligament**
2. Medial meniscus
3. Medial collateral ligament
4. Patellar ligament
5. Tibia
6. Femur
7. **Cranial Cruciate Ligament**
8. Stump of tendon of long digital extensor
9. Lateral meniscus
10. Lateral collateral ligament
11. Fibula

Figure 1.1. Cranial view of the left stifle joint of the dog. The patella has been removed. Adapted from Textbook of Veterinary Anatomy, 4th Edition [7]

The main ligaments that stabilise the stifle (knee) joint are the medial and lateral collateral ligaments - located on the outside of the joint [8] and the cranial and caudal cruciate ligaments which are located in an intracapsular but extrasynovial location [9]. It is the CCL that is the primary stabiliser of the stifle joint and is the

ligament most commonly ruptured in both dogs and in humans (along with the medial collateral ligament [10]). The CCL attaches to the fossa on the caudal aspect of the medial side of the lateral femoral condyle [11]. The CCL passes cranially, medially and distally across the intercondylar fossa and attaches to the cranial intercondyloid area of the tibia [11]. The ligament acts to resist hyperextension of the stifle and limits internal rotation of the tibia with respect to the femur [12, 13]. The CCL has two distinct functional bands – the craniomedial and the caudolateral bands [11]. The craniomedial band is taut during flexion and extension of the stifle joint, whereas the caudolateral band is only taut when the stifle is in extension [11]. Both bands are taut during extension of the stifle, so they act to stabilise the joint and prevent over-extension. If the CCL becomes torn, instability results and this can be associated with lameness, chronic pain and stiffness. In dogs, osteoarthritis will develop and is unpreventable [14]. The CCL is analogous to the anterior cruciate ligament (ACL) in humans. They are often inappropriately used interchangeably; in this project CCL refers to the canine cruciate ligament and the ACL refers to the human ligament.

Veterinary diagnosis and treatment of CCL rupture

One of the key diagnostic tests for CCL rupture is the “draw sign” [15]. This is where the clinician holds the femur and then manipulates the tibia. If the tibia displaces cranially relative to the femur, it indicates loss of integrity of the CCL. Radiographs may also be taken to investigate changes that might typically occur, such as joint effusion and osteophytosis. However, these are not able to detail soft

tissue (such as ligaments). Arthroscopy, ultrasound, magnetic resonance imaging (MRI) or exploratory arthrotomy may also be used to confirm the diagnosis [16, 17].

For treatment, a conservative approach, such as rest and analgesia (e.g. non-steroidal anti-inflammatory drugs) may initially be used, especially if the ligament is only stretched or partially torn, or when the patient is a small breed of dog. However, in most dogs, especially the large or giant breeds, surgical reconstruction options are the recommended approach to management and are the only way to stabilise the ruptured ligament.

There are many different surgical techniques that can be performed to stabilise the joint and manage CCL rupture [18]. The most common forms available currently, are tibial plateau levelling osteotomy (TPLO) [19], tibial tuberosity advancement (TTA) [20], lateral suture stabilisation (LSS) [21] and “over the top” (OTT) fascial grafting [22]. They all endeavour to stabilise the joint and prevent abnormal movement or further pain within the joint. Veterinary fees for such surgical procedures can place a heavy financial burden on dog owners or their pet health insurers. In the United States, it was estimated that the annual cost for treating CCL rupture (for surgery and follow up care) was $1.32 billion in 2003 [1]. These costs are increasing year on year, as the surgical techniques become more advanced requiring more specialist equipment and as dogs are becoming increasingly popular as pets. In the UK, the number of dogs in 2012 was estimated to be around 8 million [23], with the UK Kennel club currently recognising 210 different dog breeds and

registering over 200,000 pure bred dogs a year [24]. Within the UK, if a dog has had surgery to stabilise the stifle, it cannot be shown in championship shows without specific permission from the Kennel Club [24]. Following surgery, the dog may never recover 100% normal function [2, 14, 25] due to the knee failing to gain complete stability. Any underlying causes of CCL rupture or inflammation within the joint or synovium may still have an active effect whilst progressive osteoarthritis occurs simultaneously within the joint [26]. These add to the pain and inflammation in the joint.

Pathology of CCL rupture

Susceptibility

The effects of age, bodyweight, the size of dog, bone conformation, hormonal, autoimmune components and genetic factors are all key contributory factors to CCL rupture [4, 5, 14-16, 27-36], as is the underlying ligament structure/strength. There may be many other pathways involved in CCL rupture susceptibility, such as sensorimotor systems, neurological or neurovascular pathways [37-39], but the relative impacts of these have not been described in detail. These pathways may act to initiate CCL rupture and/or repair, or they may be secondary to rupture. Substance P (and other neurological pathways) have previously been associated with rheumatoid- and osteo-arthritis [40, 41] and so it is feasible that they too may be involved in CCL rupture.

Ligament structure

Ligaments are strong connective tissue structures that connect bone to bone across a joint. Ligaments have a complex hierarchical structure similar to that of tendons [42-44], with many different collagen fibres of different sizes arranged longitudinally along the length [43, 45, 46]. For illustrative purposes the structure of the tendon is shown in Figure 1.2 (adapted from Kastelic et al. [47]).

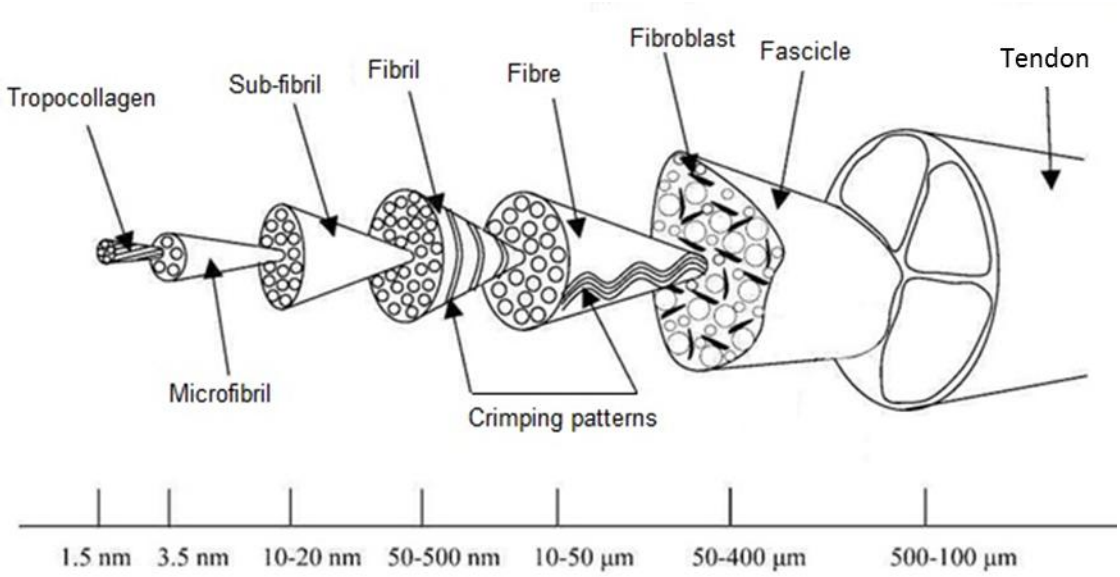


Figure 1.2. The microscopic structure of a tendon (adapted from Kastelic et al.) [47]. This is shown to illustrate the longitudinal arrangement of collagen fibrils, which is similar to that which occurs in ligaments.

Although tendons have many fibril unit classifications (as shown above), Clark et al. [42] found only two distinct classifications within the CCL – collagen fibre bundles and fascicles. The collagen fibre bundle is the basic unit of the ligament structure. Cells (from which the ECM proteins are released) are present between the collagen

fibres, arranged in a longitudinal direction [43, 46]. The cell population is mainly composed of fibroblasts, but chondrocyte-like cells may also be present [48, 49].

Ligaments are two thirds water and the main structural component in the dry weight of a ligament is collagen [37, 50]. Collagens are essential to the integrity, structure and strength of the ligament. Collagen is arranged in fibres made from many fibrils. The conversion of the fibrils to fibres involves the enzyme lysyl-oxidase (LOX) [51] which promotes cross-linking between each fibre. This gives the ligament its core structure and increased strength and stability. The fibrils, instead of being completely straight are crimped and it is thought that the crimp is important in allowing the ligament to absorb a degree of tensile load bearing without the ligament being damaged [43]. The greater the force applied to the ligament, the straighter the collagen fibres become.

Collagen comprises 75% of the dry weight of the CCL. There are several types of collagen in ligaments; the primary form (85% of total collagen present) being collagen type I (COL1A1, COL1A2) [43], the remaining collagens are type III, type V, type VI, type XI and type XIV [43]. Collagen type I has a triple helical structure [52] formed by two different alpha chains A1 and A2, whose genes reside on different chromosomes (Chromosome 9 and 14 in the dog). They are normally present in the ratio of 2:1 - COL1A1:COLA2 [53]. The remaining matrix components of ligaments are proteoglycans, actin, laminin, integrins and some uncharacterised proteins [43,

47, 54]. As these are all structural components of ligament, important in tensile strength, they can be considered as possible candidate elements for CCL rupture.

Collagen is released from cells as procollagen, which is then cleaved into tropocollagen [55-57]. This tropocollagen is converted to collagen fibrils with the aid of the enzyme lysyl-oxidase (LOX) [58]. LOX acts on lysine and hydroxylysine residues deaminating them to aldehyde groups. This allows covalent bonding to form between the different tropocollagen molecules producing collagen fibrils [51]. This post-translational modification of the procollagen chains is crucial for the formation of mature collagen molecules [59]. Collagens are renowned for their triple helix structure which is formed as a result of a repetitive motif, typically $(Gly\text{-}X\text{-}Y)_n$ where X is normally proline and Y hydroxyproline [52]. This repetitive backbone forms a very stable, tightly coiled alpha helical structure – hence the high tensile strength of collagen fibres under normal conditions.

Levels of cross-linking along with the high collagen turnover rate in the ligament may also play an important role in determining whether or not the CCL is likely to rupture [50, 60]. Furthermore, Comerford et al. [50] found that ruptured CCLs had significantly higher concentrations of immature collagen cross-links present in them compared to intact control CCLs. High levels of immature collagen cross-links in the CCL may indicate that the ligament cannot withstand day-to-day stresses making it more likely to rupture. The time at which the ligament had ruptured was unknown, so the higher levels of immature collagen may be due to the ligament trying to

repair itself (remodelling) after an initial injury or it could be an underlying cause of the injury. The various histological changes over time may reflect weakening of the ligament which then fails when loaded [50, 60, 61]. One of the key problems in investigations into CCL rupture is that it is impossible to stage the injury (start of rupture, partial rupture or complete rupture), so it is difficult to equate the changes seen with certainty, to before, during or after the rupture.

Collagen is not the only component of interest in ligaments – proteoglycans (PGs) and glycosaminoglycans (GAGs) also play an important role in integrity of ligaments [43, 62, 63]. Examples of such PGs include aggrecan, biglycan, decorin, fibromodulin and lumican [64-68]. Cells secrete the PGs directly into the extracellular environment, where they are important in the strength, structure and assembly of the extracellular matrix. Due to the electrostatic attraction between the PGs and water [63], they are important in the osmotic potential of the ligament and the ability to withstand stress and compression. An increase in water content changes the viscoelastic properties of the ligament and may lead to an increase in CCL rupture [46, 50, 69, 70]. PGs are important in collagen formation, assembly and organisation of the fibrils [66, 71, 72]. They are also to be considered ideal candidates to study for association with CCL rupture susceptibility [68, 73].

Bone conformation

A steep tibial plateau angle (TPA) has also been proposed as a possible aetiological factor for CCL rupture in dogs [74-78]. The tibial plateau angle (TPA) can be

demonstrated on the radiograph (Figure 1.3) and is the angle at the top of the tibia bone (between the tibial plateau and the long axis of the tibia). It is thought to be important in CCL rupture as the gradient of the tibia determines the strain placed on the CCL. Static loading models have indicated that the gradient of the tibial plateau correlates with the strain placed on the CCL [79, 80]. However, such models do not involve dynamic stabilisation from major muscle groups acting across the stifle joint, so this remains a controversial area. Some researchers have found a significantly greater TPA in dogs with CCL rupture compared to dogs without CCL ruptures [81], whereas other research shows no statistical evidence for tibial plateau gradients and CCL rupture [82, 83].

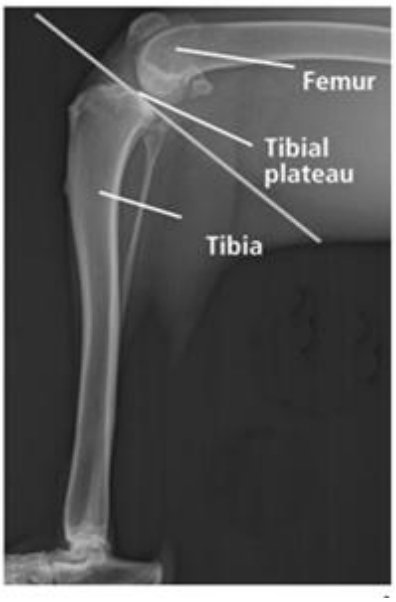
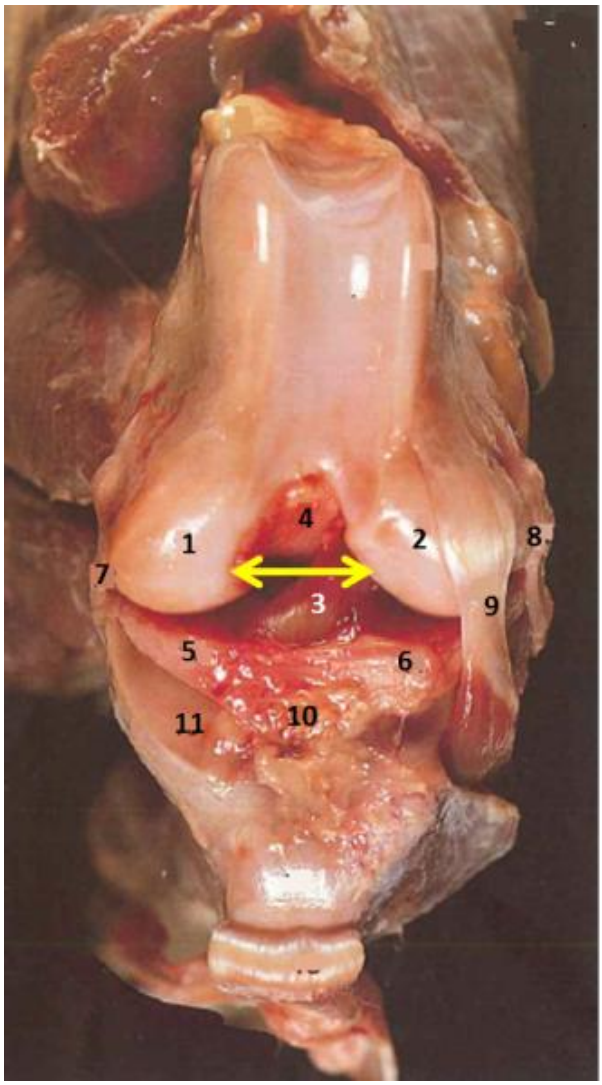


Figure 1.3. Radiograph of the stifle joint showing the tibial plateau (kindly provided by Professor John Innes, University of Liverpool).

Narrow intercondylar notch width has also been implicated in the aetiopathogenesis of CCL rupture [76, 84-86]. The intercondylar notch is the site of the origin of the CCL and the ligament passes cranially, medially and distally through the notch [11]. If the notch is stenosed, there is likely to be impingement on the CCL such that the CCL may be more likely to be mechanically damaged. An illustrative diagram depicting the anatomical location of the intercondylar notch is shown in Figure 1.4.



1. Medial condyle of femur
2. Lateral condyle of femur
3. Cranial cruciate ligament
4. Caudal cruciate ligament
5. Medial meniscus
6. Lateral meniscus
7. Medial collateral ligament
8. Lateral collateral ligament
9. Tendon of long digital extensor
10. Infrapatellar fat body
11. Medial condyle of tibia

Intercondylar notch indicated by ⟷

Figure 1.4. Cranial aspect of the left stifle joint of a dog depicting the intercondylar notch. The patella has been displaced and retracted distally. The joint is in extreme flexion and the joint capsule opened to reveal the intracapsular structures. Adapted from Colour Atlas of Clinical Anatomy of the Dog and Cat 2nd Edition [87].

Degeneration of the ligament structure

Although CCL rupture often presents as acute onset lameness, it is generally thought to occur after progressive degeneration and subsequent weakening of the ligament [4]. As such, it is typically a non-contact injury in dogs. This degeneration has been associated with changes in the extracellular matrix (ECM) [4, 50], with a reduction in the structural and functional properties of the CCL. This weakening may then enable the ligament to stretch or partially tear under what might be considered physiological loads, before a complete rupture presents itself [88]. The degeneration and weakening of the ligament may also be explained by higher than normal levels of collagenase being present in transected ligaments compared to normal, healthy ligaments. Amiel et al. [89] documented this finding when examining normal and transected CCLs of rabbits. Degeneration of the ligament may also be consistent with normal age related changes that occur over time. Over time, gradual functional and structural deterioration of the ligament structure occurs, with loss of ligamentocytes, metaplasia of the ligamentocytes to other cells (such as chondrocytes) and loss of collagen fibril organisation and strength [4, 90, 91]. Age related degenerative changes are consistent with reports that CCL rupture has an increased incidence in older dogs [27].

Size and obesity of dogs

Large breeds of dogs are more susceptible to CCL rupture than smaller breeds [92]. This can be attributed to more stresses on the joint in the larger breeds of dog. Duval et al. [27] and Whitehair et al. [93] have both documented that dogs

weighing >22kg had a higher prevalence of CCL rupture, at a younger age, than smaller dogs under 22kg. Obesity has also been documented to increase the risk of CCL rupture in Cocker Spaniels [94]. This study showed that Cocker Spaniels with CCL had a greater body weight than Cocker Spaniels without CCL rupture. Another experiment in Labrador Retrievers showed that those on a limited food intake regime had a significantly lower incidence and severity of osteoarthritis compared to those dogs with no food restrictions [95].

Hormonal influences

Another factor that also might play a part in CCL rupture is hormonal influences on tissue metabolism. In humans, ACL rupture was reported to be more common in females compared to males [96-99]. Female dogs have a higher incidence of CCL rupture than male dogs [93] and CCL rupture was found to be more common in dogs neutered early (before 8 months) compared to dogs neutered after puberty [100, 101]. Differences in oestrogen, progesterone, testosterone and relaxin between male and female athletes, as well as the continual monthly variations of hormones in females, have all been hypothesised to play a role in the increased risk of rupture of the ACL in women [30, 96, 99, 102, 103]. Research has shown a statistically increased risk of ACL damage for female athletes during the ovulatory phase of the menstrual cycle [104].

Auto-immune components

Auto-immune reactions have been hypothesised as influencing factors for CCL rupture [3, 31, 105, 106] because synovial inflammation (synovitis) is commonly found in the joint during repair of the ligament. Whether this is a consequence of the CCL rupture or a damage initiating factor is still under debate. However, synovitis has been found to increase the risk of contralateral CCL rupture [31, 107]. In rabbits, persistent synovitis has been found to cause disorganisation of the ligament framework, causing marked changes in the morphological characteristics of the CCL. This resulted in a significant decrease in tensile strength of the ligament [108].

Inflammatory cells, such as B and T lymphocytes, TRAP positive macrophages and CD1+ cells have all been found to be increased in the synovium of CCL ruptured dogs [109, 110] and also in the synovium of dogs with stifle arthritis [111]. Normally, the synovium that surrounds the stifle joint acts as a protective barrier to the cruciate ligaments which are enveloped between synovial membranes. The CCL is shielded from the immune system by a thin synovial lining. This means that any locally produced antigens will form immune complexes that will remain within the synovial space and have an unhindered local affect [105]. Due to the CCL being primarily composed of collagen type I fibres, any anti-collagen type I antibodies formed would bind to the CCL making it more susceptible to degradation and rupture [105]. If the CCL is damaged, neo-epitopes and other damage-associated molecular patterns (DAMPs) may be released from the ligament tissue and act as

antigens, causing innate and adaptive immune response to local tissues [112, 113]. This process may not act alone in causing CCL rupture, but it may well act as a promoter for further rupture. Persistence of the antigen presence and subsequent immune response, along with the prolonged inflammation, is thought to play an exacerbatory role in CCL rupture and the associated pain and inflammation within the joint.

Whether the inflammatory and auto-immune components cause CCL rupture or if they occur secondary to the rupture of the CCL is still under debate [114-116]. Exposed antigens may also play a role in initiating osteoarthritis (OA), as OA has been proposed to be, at least partially, caused by an auto-immune response with collagen as the direct initiator [112, 117]. It is unclear whether the degradation of the CCL and subsequent release of collagen antigens is a significant cause of the OA or whether the direct and prolonged inflammation within the synovial joint leads to OA [26, 117]. Whichever the case, OA occurs simultaneously with the CCL injury and is unavoidable and untreatable.

Major Histocompatibility Complex and CCL rupture

Cell mediated immunity is regulated by the major histocompatibility complex (MHC) which can have an important role in disease susceptibility and resistance [118]. MHC has been implicated in many autoimmune diseases such as diabetes [119], hypothyroid disease [120], Addison's disease [121], systemic lupus erythematosus (SLE) [122], anal furunculosis (AF) [123] and rheumatoid arthritis

(RA) [124]. In dogs the MHC region is found on chromosome 12 and the genes are called the “dog leucocyte antigen” (DLA) system.

MHC molecules are transmembrane glycoproteins, which capture and present epitopes of antigens to T-cell receptors and initiate immune response cascades. The MHC region is highly polymorphic, with many alleles existing for each gene. Expression of the genes is codominant, with most individuals being heterozygous. This allows a wide range of antigen presentation to be available and maximises the immune response to foreign antigens. Loss of heterogeneity (diversity of MHC) due to inbreeding can result in an increased susceptibility to autoimmune diseases [125].

Genetics

It should be considered that a genetic element could be a significant predisposing factor of pathological CCL changes and the subsequent degeneration. From epidemiological studies, it is apparent that CCL rupture is not a classical genetic disease, as characterised by simple Mendelian genetics [126]. It is instead thought to be a polygenic, multifactorial, complex genetic trait [36]. As it is obviously not a monogenic condition, it will not be possible to find a single, simple, genetic basis explaining the rupture susceptibility. It will however, hopefully become possible to determine which dogs are more at risk of developing a ruptured CCL. Those with predisposing alleles may be detected by a genetic screening test. This information could then be used to inform breeding strategies, whereby dogs could be screened

for the “risk” alleles before breeding. In order to achieve this, it would be necessary to identify a set of susceptibility markers.

It has been widely reported that several dog breeds are more or less susceptible to CCL rupture than others. Breeds such as the Greyhound, Dachshund, Basset Hound and Doberman Pinchers are less likely to rupture their CCL [27, 93] than dogs in the high risk category such as Newfoundlands, Labrador Retrievers, Rottweilers and Saint Bernards [2, 6, 13, 27, 60]. Dogs in the high risk category often rupture their ligament without any apparent traumatic circumstances and as a result, young dogs rupture their CCL suddenly when only performing simple day-to-day activities, such as during play, walking or running [6, 14].

Rupture of one CCL puts a dog at higher risk of rupturing (or otherwise damaging) the contralateral CCL - studies have indicated this to be around 54% risk in Labradors [107]. This value can vary between 22-54% depending on which breed is being studied [33, 107, 127, 128].

The genetic mode of inheritance of CCL rupture is postulated by some researchers to be autosomal recessive with around a 51% penetrance [6]. This means that for every dog that has the “risk” allele, just over half of the dogs will actually rupture their cruciate ligament. With a heritability of only 27% [6], this clearly implies that other factors, both intrinsic and extrinsic such as those listed above contribute to

development of CCL rupture. Nonetheless, the inherited genetic component, although relatively small is still likely to play an important role in CCL susceptibility. As one cannot tell whether or not a dog will be affected, there is a need to identify the specific genes and/or genomic regions that might be involved in the CCL rupture susceptibility, initiation or progression.

Newfoundland dogs and CCL rupture

Newfoundlands are a large breed of dog commonly used for draught and water rescue work. They are known for their gentle docile nature as well as their strength and endurance [129].

Figure 1.5. Structure and form of a Newfoundland dog[1]

[1] Reproduced with permission: ©David Cant 2013

The origins of the Newfoundland dog as a breed are vague, but the breed is thought to have arisen when a “large Mastiff-type dog” was brought back to England from Canada by European fisherman and explorers. The modern breed as known today, developed from there around the 1800s. Many, if not all American Newfoundlands have an ancestry traceable back to one imported European dog (Siki) around the end of World War 1 and whose son (Ch. Harlingen Neptune of Waseek) started the first large kennel in the United States [130].

Strict breed standards dictate which dogs can be registered within a breed in the dog clubs. In the UK the standard colours for registration are solid black, brown or grey, black with white markings, brown with white markings and Landseer (white with black markings). American standard colours are the same with the solid colour markings, but registration restrictions mean white markings on the coat are only allowed on the chin, chest, toes and tip of tail. In Canada brown and grey colours are not recognised within the breed/kennel clubs and in some European countries such as France and Germany (including the Fédération Cynologique Internationale), the Landseer is considered to be a separate breed, although in the UK and USA/Canada the breed is considered to be the same [129, 131, 132].

CCL rupture is a major clinical problem within the Newfoundland breed along with other genetic conditions including hip and elbow dysplasia [133-135], cystinuria [136], dilated cardiomyopathy (DCM) [137] and sub-aortic stenosis (SAS) [138]. Estimates from other studies in the USA, suggest that CCL rupture has a 22%

prevalence in Newfoundland dogs, a heritability of 27% [6] with a postulated recessive mode of inheritance, strongly indicating that that there is a genetic component to the CCL rupture [6]. Heritability may be under-represented however, because a dog may have the risk alleles and not rupture its CCL; hence it is likely the true value could be much higher. In the past 10 years (from 2002-2011) the mean number of Newfoundlands registered each year with the UK Kennel Club was 976 (range 838 - 1,145). Considering the estimated impact of 22% prevalence of CCL rupture in Newfoundland dogs, this would estimate to be around 215 incidences of CCL rupture a year (excluding any contralateral ligament rupture). Radiograph scoring schemes for hip and elbow dysplasia [139, 140] and DNA screening tests for DCM and SAS [141] are already in place to try and reduce the incidence of these conditions in the breed. It is hoped that a DNA screening test for CCL rupture will be added to the list at some stage in the future, to help further improve the health and welfare of the breed.

Canine genetic studies

The dog has 78 chromosomes which are arranged in pairs; in each pair, one chromosome is inherited from each parent. The canine genome was published in 2004 by the BROAD institute at MIT/Harvard using sequences from a female boxer dog and a poodle [142]. Since the publication of the canine genome, huge advances have been made in investigating conditions with genetic components in the dog. A new annotation of the canine genome (CanFam3) was released in May 2012. In this thesis, unless otherwise stated all analyses and descriptions relate to the CanFam3

genome annotation, although many SNPs and genes investigated had sequences initially chosen from CanFam2.

Pedigree dogs are seen as a suitable species to study as models for many complex human genetic diseases. Studies of many naturally occurring diseases in dogs such as CCL rupture, arthritis, epilepsy, deafness and cataracts have identified similar aetiologies and/or pathogeneses to human versions of these diseases [142-145]. Comparative genetics is much easier to translate between dogs and humans than mice and humans because dogs are genetically more similar to the human than the mouse [146]. Many genes in dogs have orthologues in humans with the same function. Therefore, not only can the research from this study be used for studies of CCL rupture in Newfoundlands and other dog breeds, but it could also have implications in human research. Causal genetic variants for diseases such as Duchenne muscular dystrophy [147], narcolepsy [148] and osteogenesis imperfecta [149] were first identified in the canine version of the disease and then were subsequently systematically applied to these Mendelian disorders in Man.

The canine genome shares substantial homology with the human genome and dogs offer great advantages over humans for studying certain diseases. Dogs have a significantly shorter lifespan than humans, so longitudinal studies are much easier to conduct in dogs than in humans and reduce the need for specialised housing for studies in mice/rats. Dogs also share the same environment with man [144] so the impact of environmental factors (housing, weather, cigarette smoke etc.) can be

taken into account. The dog has a highly conserved breed structure and comprehensive pedigree data is readily available from many generations [144]. Pure breed pedigree dogs in particular, show very high linkage disequilibrium (LD) [150]. This means that smaller numbers of a breed can be used in a genetic study to achieve the same statistical power needed for a similar study conducted on human subjects. It also means that fewer genetic variations need be studied within a breed as the SNPs are often in LD with each other.

The dog has a unique population structure with minimal variation due to the small amount of founders in each breed and "closed" breeding populations [151]. There are two distinct "bottlenecks" in the history of dog evolution [142, 151, 152]. The first bottleneck occurred when dogs became domesticated from their close ancestor the wolf. The second when humans started selectively breeding the dogs for desirable, breed-specific characteristics i.e. to hunt, to retrieve, to point, for long snouts, short legs etc. [143]. This has resulted in large phenotypic diversity between breeds. As a result of the breeding patterns that have developed to enable a breeder to produce the ideal pedigree dog, for example, the use of popular sires as studs, genetic diversity is reduced within each breed [151]. One example of such restrictive breeding practises is that for a dog to be registered as a "pure-breed" in accordance with the kennel club guidelines, both of its parents must have been registered under the same breed classification.

Genetic variations

The release of the human genome in 2001 enabled great advances in investigations into diseases and population genetics [153]. As a result the natural variations in DNA can be assessed and can provide estimation for risk of an individual to specific disease susceptibility or resistance. The most common variations in DNA are called single nucleotide polymorphisms or SNPs [154]. These are variations in the DNA sequence and involve substitution of one of the bases (A, T, C or G) in the DNA sequence with another. Other common DNA variations that have been identified in DNA include; INDELs (Insertion/deletion) [155] where a base or number of bases is/are inserted or deleted in the sequence, microsatellite markers (MSAT) [156] where 2-4 bases are repeated within the DNA sequence and the number of repeats can vary between individuals and copy number variations (CNV) [157] where a section of the DNA is replicated. MSATs have traditionally been used as the molecular markers of choice to determine genetic associations with disease [158]. More recently however, SNP chip arrays have superseded MSAT analysis due to the high frequency of SNPs within the genome [159].

SNPs within a gene may be intronic (within an intron), exonic (within an exon), synonymous (the amino acid is not changed as a result of the SNP), non-synonymous/missense (the amino acid is changed by the SNP and the protein coding sequence is altered). SNPs that cause a frameshift in the DNA sequence (by addition or deletion of a base) often lead to truncated proteins by generation of premature stop codons within the DNA reading frame [160]. The non-functional

protein generated as a result can have severe consequences in health and disease [161, 162]. SNPs can also be found in parts of the genome that are devoid of protein coding genes, i.e. intergenic regions, where the SNP may be involved in the regulation of gene expression (in promoter or enhancer regions) or non-coding regulatory RNA molecules. Intergenic SNPs that show association with disease may be in LD with SNPs that are up or downstream of an actual gene, or in LD with a SNP that is within a functional gene. The exact function of variants in non-coding regions of the genome still often remains unclear.

Intronic and synonymous SNPs are most likely to be non-functional mutations as they do not have a direct effect on the protein coding regions. This is because the mutation is spliced out before the translation of RNA to protein (an intronic SNP), or the amino acid in the sequence is not changed (a synonymous SNP). However, they may be functional if they affect splice sites or regulatory regions that affect downstream processing, protein folding or function and/or the stability of microRNA [163]. These SNPs, although sometimes significantly associated with a disease are unlikely to be the actual causative mutation but are likely to be in LD with the causative mutation which may be exonic or a non-synonymous mutation that causes a change in the amino acid. As a result of the amino acid change, the protein folding will be altered as the secondary, tertiary and quaternary protein structures may be changed (as established by Anfinsens dogma [164] – the protein folding is determined by the amino-acid sequence). The frequency of these types of mutations in a population is higher relative to mutations that have a direct effect on

protein coding, structure and formation such as non-synonymous (missense) SNPs and frameshift mutations (INDELs) [165].

In order to study genetic variation within a species, it is important to understand how the variations are linked within a species. The developments of genetic maps and advances in analysing linkage disequilibrium (LD) have made this possible. LD is the non-random association of alleles at two loci during recombination [166]. When a particular allele at one locus is found together with a specific allele at a second locus, more often than expected if the loci were segregating independently, then they are said to be in linkage disequilibrium with each other. LD is also a measure of population size and genetic structure [167]. In large, out-bred populations (such as humans), genetic diversity is huge as there have been many recombination events throughout evolution i.e. LD is low. However, large populations themselves do not necessarily correspond with low LD if there are only a small number of founders in the population (i.e isolated populations).

The dog has high LD when compared to the “out-bred” human population, due to a small number of founders and limited outbreeding [150, 168]. Genetic variants differ between dog breeds and although LD is relatively short between breeds, it is greatly extended within a breed and is much greater than that seen in human populations [150]. Of the dog breeds that have been studied to date, Labrador Retrievers show the shortest LD because the heterogeneity within the breed is high, especially when compared to breeds such as the Japanese Akita and the Pekinese

[150]. This is due to the Labradors popularity as a breed (it is the UK's most popular breed of dog [23]) in comparison to the Akita and Pekinese which are much less popular. There were severe reductions in the Akita breed numbers during World War 2 [150] which significantly reduced the number of founder dogs for the modern breed and consequently reduced LD.

Maintaining genetic diversity is of great importance when considering inherited diseases. Genetic diversity in dogs is reduced due to selection pressures (popular sires and dams), bottlenecks in breed creation and closed breeding populations [169]. This has led to a considerable number of inherited genetic diseases and defects that are overly prevalent in certain breeds. The inbreeding coefficient can be used to determine the level of genetic diversity within a breed [170]. The higher the inbreeding coefficient, the more common ancestors are present in a pedigree. The UK Kennel Club will no longer register mother/son, father/daughter or brother/sister matings, a plan that has been put into place to reduce the level of inbreeding. They have also introduced a MateSelect™ program that allows the breeder to determine the inbreeding coefficient of each mating.

The Kennel Club describes the inbreeding coefficient as:

> *"The probability that two copies of the same gene have been inherited from a common founder - that is an ancestor shared by both parents. The lower the inbreeding coefficient, the lower the probability (risk) that this will happen." [171]*

The Newfoundland breed as a whole has an inbreeding coefficient of 4.8%, which translates to approximately a 1 in 21 chance that the dog will inherit the same gene

from a common ancestor that appears in both the sire's and dam's pedigree. For a pure-bred dog this is relatively low compared to some breeds such as the King Charles Spaniel which has an inbreeding coefficient of 11.8% and suffers severe hereditary health problems. The other breeds investigated in this project have inbreeding coefficient values of 6.4% (Labrador Retrievers), 6.8% (Rottweilers) and 6.7% (Staffordshire Bull Terriers).

Quantitative trait mapping for complex polygenic disorders (such as CCL rupture) remains difficult relative to monogenic trait/disease mapping [172]. This emphasises the need for accurate and clear phenotyping so that any background "noise" – such as breed, geographical location, coat colour etc. can be reduced. Since the release of the dog genome in 2004, many advances in understanding the genetic basis of canine (and human) diseases have been made in a relatively short time frame. With the release of the newest canine genome assembly (CanFam3 in 2012), in addition to the advent of newer next-generation sequencing platforms, it should become easier to identify causative mutations that are involved in complex diseases. Large regions of association within chromosomes can be narrowed down to specific mutations, which will hopefully then be used to develop screening tests.

Aims of the Project

It is hypothesised that there is a genetic basis to CCL rupture which can be identified at a molecular level. To investigate this, the project aims were to use molecular genetic techniques (GWAS, Sequenom, Q-PCR and tissue culture) to;

1. Identify genetic mutations and putative genes which are associated with resistance or susceptibility to CCL rupture in Newfoundland dogs.
2. Determine whether there are any genetic differences between Newfoundland dogs from two different populations (Europe and North America).
3. Determine whether the genetic mutations identified within the Newfoundland breed are specific only to the Newfoundlands or can also be found in other CCL rupture susceptible dog breeds (Labrador Retrievers, Rottweilers and Staffordshire Bull Terriers).
4. To determine the expression levels of select candidate genes between ruptured and healthy CCL tissue.
5. To determine if there is an autoimmune component to CCL rupture by studying the DLA region for association.
6. To determine if TNFα stimulation of ligamentocytes in vitro can be used as a laboratory model of the pathological changes seen in CCL rupture.

Once the associated SNPs have been identified they could be used to develop genetic screening tests to identify susceptible/resistant dogs. This data can then be used to initiate breeding strategies to try and reduce the incidence of CCL rupture in dogs. This will be of great economic and emotional benefit to dogs, dog breeders, owners and all!

Chapter 2.

General Materials

& Methods

The laboratory techniques presented in this chapter are generic to all subsequent experiments described herein. Deviations from the general protocol for individual experiments are detailed in the relevant chapters of the thesis.

Clinical diagnosis and phenotypic status

The phenotypic description of the dogs was determined according to the inclusion/exclusion criteria of the study.

- Inclusion criteria (cases): Dogs with CCL rupture of any age. CCL rupture must have been confirmed during surgery.
- Exclusion criteria (controls): Healthy control dogs, with no symptoms of CCL damage, were over five years of age and showed no obvious signs of CCL rupture (hind limb lameness or pain).

Ideally, the control dogs should have been examined by a veterinary surgeon with a minimum qualification of a CertSAS to ensure no clinical features of CCL rupture (such as medial buttress formation, stifle joint effusion, positive tibial thrust or positive cranial draw) were present. This clinical assessment was not available for all control samples submitted and is a recognised limitation of the project; only 35% of control samples were subjected to this assessment.

Dogs included in the study

Samples from Newfoundland dogs that had ruptured their CCL (cases) and those that had no clinical evidence of CCL rupture (controls) had DNA extracted from either EDTA blood (excess from other veterinary tests already being carried out), or salivary swabs – see Sample collection method, page 44, for more detail.

The majority of the Newfoundland CCL rupture cases and controls were collected by attending dog shows, mailing sample collection kits to the owners and breeders or by obtaining samples sent from veterinary surgeons. The remaining CCL rupture cases and controls, as well as the cases/controls for other breeds used in the study (see below) were taken from the UK DNA archive for companion animals (http://www.liv.ac.uk/dna_archive_for_companion_animals/) and had already been collected for previous research studies [173]. Full informed, written owner consent had been given to use these samples in research studies.

Other breeds that have a high risk of developing CCL rupture were also included in this project to enable comparisons of the genetic predisposition between the breeds. These breeds were Labrador Retrievers, Rottweilers and Staffordshire Bull Terriers. The decision to use the other breeds (especially the Staffordshire Bull Terriers) was taken retrospectively after the main sample collection process had been carried out. This meant that specific collection of the samples needed (both cases and controls) could not be carried out as for the Newfoundlands. Therefore,

the majority of samples from these breeds were obtained from those already integrated into archive databases – which were limited in number.

The CCL rupture case samples for the other breeds were taken from the UK DNA archive for companion animals at Liverpool. The phenotype of the dog (breed, age, sex, CCL rupture) was known. The control samples for the Labrador Retrievers, Rottweilers, Staffordshire Bull Terriers and some Newfoundland samples were taken from the Veterinary Laboratory agency (VLA) archive (a subset of samples within the UK DNA archive for companion animals). The VLA samples represented a random population of UK dogs with no disease information and were therefore assumed to be controls with no cruciate ligament rupture – this is a recognised limitation in the project as no clear diagnostic information is available.

Specific information on breed, sex, age and numbers of cases and controls used are given in the Appendix (Appendix 1).

Each sample submitted to the study was accompanied by a consent form (shown in Appendix 2) in which the owner gave their full informed written consent. The owner supplied information about the dog, including age at sample collection, sex, whether the dog was neutered/spayed. Information about the injury (if applicable) was also provided, such as age at diagnosis/injury, how the injury was caused (slip, fall, accident or during normal activity) and surgical information was provided when

available. The owner also consented to allow the sample to be stored and integrated in the UK DNA archive for companion animals and used as appropriate in this study and other scientific research.

Samples were initially collected from dogs within the UK only, but sample submission rate was very low, despite multiple approaches and addresses to UK Newfoundland breeders and owners at their club meetings and dog shows for samples. On the basis that at the current rate of sample collection, progress would not be enough to enable the project to proceed within a suitable timeframe, the sample collection area was extended to include North America and mainland Europe.

Linblad Toh et al. (2005) [142] suggested that for a multigenic trait, at least 100 cases and 100 controls were needed to detect disease alleles. Factors including allele frequency, the interaction of the allele with other variants and the risk that having the allele confers to the dog, also determines the number of cases and controls needed to produce a significant study. They found that an allele that gives a 5-fold increase in disease risk will be detected with 97% power and an allele that increases risk 2 fold will be detected with 50% power using 100 cases/100 controls. These numbers (as well as some studies using smaller cohorts of dogs) have been successfully used to map traits and diseases such as epilepsy, diabetes and atopic and digital dermatitis [174-177]. In our study, the numbers of dogs (for the majority of breeds) are close to Linblad Toh's estimates; 99 cases and 172 controls for the

Newfoundlands, 124 cases and 165 controls for the Labrador Retrievers, 57 cases and 81 controls for the Rottweilers with the Staffordshire Bull Terriers falling short at 13 cases and 38 controls. As the dogs are pure-breeds the relationship status between each dog is such that the linkage is high enough to detect mutations in a much smaller number of dogs than would be required if the study was being conducted on a human population.

Sample Collection for DNA extraction

In order to perform the experiments in this project it was necessary to collect biological samples to obtain DNA from dogs with and without CCL rupture. Blood or saliva samples were collected from Newfoundlands and other selected dog breeds (Labrador Retrievers, Rottweilers and Staffordshire Bull Terriers) and were integrated within the UK DNA archive for companion animals at Liverpool (www.liv.ac.uk/dna_archive_for_companion_animals). The research in this project was approved by the University's Research Ethics Sub-Committee.

Collection of veterinary blood samples solely for research purposes in the UK without a home office licence is prohibited; however, residual blood remaining after a diagnostic sampling may be used for research and does not require a licence. Collecting saliva with swabs is considered a non-invasive procedure and does not require a licence.

Sample submission was either an EDTA tube of blood (residual blood remaining from routine veterinary diagnostic tests carried out on the dog), or, if blood was not available then salivary swabs were taken using the Oragene™ saliva collection kits (Oragene™, DNA Genotek, Ontario, Canada). Samples were only included in the project if the owners gave their full written consent to the blood/samples being stored and used in scientific research.

Oragene™ sampling kits were mailed to owners, who collected the sample themselves. The Oragene™ kit consisted of sponges and a collection tube containing the Oragene™ buffer solution. The owner simply had to place the sponges in the dogs' mouth (at the cheek pouch) and saturate the sponges with saliva. The sponges were then removed from the mouth, placed directly into the collection tube, shaken and then mailed back to us by standard post. The benefits of using the Oragene™ sample collection process is that the sample is stable after collection for up 6 months at room temperature and is easily mailed to and from the owners for ease of collection. At the start of the sample collection process the only kits available were the human collection kits (OG-510). Soon after beginning the project, Oragene™ released the animal collection kits (OA-400) which was subsequently used as the standard collection method. The difference between the two kits was a change in design (tubes rather than pots) and a slight (unidentifiable) change in buffer solution that required an extra processing stage in the DNA extraction method – detailed further on in the chapter in the DNA extraction section.

DNA extraction

Blood

DNA was extracted and prepared from EDTA blood using the Qiagen DNA Blood midi kit (Crawley, UK) following the manufacturer's instructions and as previously described [178]. The DNA was then quantified using a nanodrop spectrophotometer (Thermo scientific, Massachusetts, USA, www.nanodrop.com).

Saliva

DNA from Oragene™ saliva kits was extracted following the manufacturer's protocol (DNA genotek, Canada) [179], but modified such that the whole sample could be processed instead of 500µl aliquot. Briefly, samples were mixed and placed in a water bath at 50˚C for one hour. Excess liquid was squeezed from the sponges into the collection tube. Oragene™ DNA purifier was added to the tube ($1/25^{th}$ volume) and vortexed to mix. Samples were cooled on ice for 10 minutes and then centrifuged at 4000g for 17.5 minutes. The supernatant was transferred to a new labelled tube. For animal kits, 5M NaCl (60µl/ml) and an equal volume of 100% ethanol was added and then mixed by inverting 10 times before leaving at room temperature 10 minutes. The NaCl addition was omitted for the human kits. Samples were centrifuged at 4000g for 17.5 minutes. The supernatant was removed and 1ml of 70% ethanol was added and left for 1 minute. The ethanol was removed and the pellet dried for 15 minutes. The pellet was resuspended in 300µl buffer AE, vortexed for 30 seconds and placed in a 50˚C water bath one hour. The DNA was

transferred into a pre-labelled tube and was quantified using a nanodrop spectrophotometer (Thermo scientific, Massachusetts, USA, www.nanodrop.com).

Salivary DNA extraction often generates lower quality DNA than that of blood, due to potential carryover from contaminants like residual food, hair, ingested dirt etc. within the sample. Furthermore, the collection of samples from salivary collection kits may introduce additional variation as they are collected in an uncontrolled manner by the owners. Owners may not have followed the instructions adequately, for example the owner may have sampled the dog immediately after food or drink or dropped the swab on floor. Variation in the amount of saliva collected on the swabs may also have affected the amount of DNA available. Some dogs may have been sampled with a dry mouth resulting in fewer cells from which DNA can be extracted resulting in a lower yield of DNA. Pooling a large quantity of saliva into the collection kit (for example collecting “drool”) effectively overloads the extraction process and can also reduce the yield of the DNA. The quality and quantity of DNA from salivary swabs, although lower than whole blood extracted DNA, is adequate for downstream applications such as genetic studies [180, 181].

Clean-up of DNA

DNA purity was assessed using the 260:280 and 260:230 ratios obtained from the nanodrop spectrophotometer readouts. Pure DNA fluoresces at 260nm, pure uncontaminated DNA will have a 260:280 ratio between 1.8-1.9. DNA absorbs

ultraviolet (UV) light at 260nm, protein and phenol absorb UV at 280nm and carbohydrates at 230nm. Any substantial decrease in the values will indicate contamination with protein/phenol/carbohydrate. In some instances it was necessary to "clean up" the DNA to remove contaminants, as there were notably lower 260:280 and 260:230 ratios and a greatly increased peak at 230nm caused by organic carryover from the extraction process. Several methods were evaluated before deciding which method was the best to use on the samples in question.

Ethanol purification of DNA

Briefly, 1/10th volume of 3M sodium acetate and 2½ times volume of 100% ethanol were added to the "dirty" DNA. Samples were vortexed and cooled on ice for 15 minutes before centrifugation at 10000g for 15 minutes. The supernatant was removed and 500µl of 70% ethanol was added to wash the pellet. The pellet was air dried for 15 minutes and then resuspended in 300µl buffer TE. DNA concentration and purity were quantified using a nanodrop spectrophotometer.

Chloroform Clean-up of DNA

The DNA was transferred to an eppendorf tube, 1/20th volume of chloroform was added, and the sample vortexed vigorously for 10 seconds. The sample was centrifuged at maximum speed for 10 minutes. The upper aqueous layer was transferred to a new pre-labelled tube and DNA was quantified using a nanodrop spectrophotometer.

Qiagen Column clean-up of DNA

Ethanol (200µl of 100%) was added to 200µl of DNA and vortexed briefly. The sample was transferred into a Qiagen spin column (Qiagen, Crawley, UK) and centrifuged at 6000g for 1 minute. The filtrate was discarded and 50µl of buffer AW1 (Qiagen, Crawley, UK) was added and the sample centrifuged again at 6000g for 1 minute. The filtrate was discarded and 500µl of buffer AW2 (Qiagen, Crawley, UK) was added and the sample centrifuged at 20000g for 3 minutes. The filtrate was discarded and the samples centrifuged again at 20000g for 1 minute. The spin column was transferred to a new collection tube, 200µl of buffer AE (Qiagen, Crawley, UK) was added and the sample kept at room temperature for 1 minute before centrifuging at 6000g for 1 minute. The eluate was added back to the column and re-centrifuged at 6000g for 1 minute. The DNA was then transferred into a pre-labelled tube and was quantified using a nanodrop spectrophotometer.

Of the above methods the chloroform clean-up method produced the best results; on average the chloroform lowered carbohydrate contamination by 26.6% compared to 23.4% by ethanol purification (as assessed by absorbance at 230nm on the nanodrop spectrophotometer) the column clean-up diminished yield such that it was not a viable method to use – data not shown. It was not possible to remove all of the carbohydrate contamination from the sample as saliva contains large amounts of carbohydrate from the heavily-glycosylated protein mucin [182]. As a result the 260/230 absorbance ratio alone was not considered a useful method of assessing the purity of DNA extractions from saliva samples, although it is adequate

for DNA prepared from blood. The chloroform clean-up method was used to "clean-up" the contaminated DNA in this project.

Quantification, normalisation and storage of DNA

DNA was quantified using a nanodrop spectrophotometer (Thermo Scientific, Massachusetts, US) and was normalised (standardised) using molecular grade water either by hand or by using the Hamilton Microlab® Star robot (Bonaduz, Switzerland). Samples were normalised to 100ng/µl for Illumina HD genotyping and 20ng/µl for Sequenom genotyping. DNA was stored at -20˚C or-80˚C until it was required in experiments. All DNA for long term storage was integrated in the UK DNA Archive for Companion Animals at -80˚C. An XL20 robot (BioMarkerLab, California, US) was used to cherry pick samples that were already in the archive that were required for the study using the unique 2D barcodes allocated to each sample tube.

Tissue collection and culture of ligament cells

In order to examine changes in gene expression in CCL ruptured tissue it was necessary to collect both ruptured and healthy/control CCL tissues. Ruptured CCL tissue was collected at the time of reconstructive CCL surgery and was immediately placed into RNAlater (Invitrogen, California, US). The control tissue was sourced from cadavers of dogs that had been euthanased for reasons other than

degenerative joint disease and which had been donated to the University for teaching.

CCL tissue was harvested from the stifle of the cadavers by sharp dissection, using an aseptic technique. Additionally, separate scalpels were used for incision of skin and the remaining dissection and the scalpel was rinsed in 100% ethanol between the cut into joint capsule and the extraction of the CCL.

Ligamentocyte cell identification

The primary cell type within the CCL is the ligament fibrocyte, otherwise known as a ligamentocyte [46, 183, 184]. These ligamentocytes are responsible for the creation and the maintenance of the extracellular matrix (ECM). The predominant ECM protein in the CCL is collagen type I, which is essential for the integrity and support of the ligament structure. The cells show typical fibroblast-like morphology.

Monolayer cultures

Ligament tissues were harvested from the cadavers by sharp dissection of the stifle joint, using an aseptic technique as described above. The tissue was immediately washed with Hanks Balanced Salt Solution (HBSS), before being digested overnight with 0.1%w/v bacterial collagenase in Dulbecco's modified Eagle's medium (DMEM/F12) with 5% fetal bovine serum (FBS) and 1% penicillin/1% streptomycin/1% amphotericin, at 37˚C on an orbital shaker. Following overnight

digestion, the released cells were recovered by sieving through a 40µm sieve to remove debris. Cells were pelleted by centrifugation at 1000g for 10 minutes and the pellet washed by resuspension of cell pellet in 5ml media (DMEM/F12 glutamax). The cells were centrifuged at 1000g for a further 10 minutes to wash cells and remove all traces of collagenase.

The cell pellet was resuspended in 1ml of tissue culture medium (DMEM/F12 glutamax, 10% FBS, 1% penicillin, 1% streptomycin, 1% amphotericin) and the cells counted before seeding plates/flasks at appropriate density to allow cell proliferation and establishment of monolayer cultures. When cells reached 80% confluence, usually after 4-5 days at 37˚C, they were released with Trypsin/EDTA (1x). The media was replenished every 3 days. For subsequent experiments the cells were plated into six-well plates in triplicate and when confluent were either induced with TNFα (see Chapter 8) in media for 24 hours, or left in normal media as a control.

After all experiments the cells were either cryopreserved for long term storage or lysed with Trizol for the first stage of RNA extraction.

Lysis of monolayer cells using Trizol

At the end of each culture period, the media was removed from the monolayer cultures and the cells washed with HBSS. Trizol was added at a ratio of 1ml/10cm^3

(250µl for 24 well plate, 500µl for 12 well plate and 1ml for 6 well plate, 2.5ml for T25 flask). The cells were lysed by pipetting up and down several times and then left for 5 minutes at room temperature. The lysed cells were then transferred to RNAase-free eppendorf tubes and stored at -80˚C until RNA was extracted.

Cryopreservation and resuscitation of cryopreserved cells

Cryopreservation was used for long term storage of cells [185]. After trypsinisation the ligamentocytes were pelleted by centrifugation at 1000g for 5 minutes. $1x10^6$ cells were placed in each cryotube. The majority of media was removed leaving the cell pellet in approximately 300µl of media to which 1ml of cryofluid (92% FBS, 8% DMSO) was added. The cell pellet was resuspended. The cells were transferred to a cryotube which was subsequently wrapped in cotton wool and placed in a padded envelope at -80˚C overnight before being transferred to liquid nitrogen for long term storage [method from Professor SD Carter: personal communication].

To re-inoculate cells in culture, frozen vials were removed from liquid nitrogen and immediately placed into a 37˚C water bath for 1 minute. The sides of the tube were wiped in ethanol before opening; 5ml of pre-warmed media was added dropwise to the cells. This was mixed gently and centrifuged at 1000g for 5 minutes. The supernatant was discarded and 1ml of media was added to the pellet. The cells were then counted and seeded at appropriate densities ($1x10^6$/ml).

RNA extraction

Samples were lysed using Trizol (Invitrogen, California, US). For tissue samples, the tissue was wrapped in foil and placed into liquid nitrogen to deep freeze the samples. The tissue was then homogenised using a dismembrator (Mikro-Dismembrator, Sartorius Stedim Biotech, France) and the powder was placed in Trizol (1ml of Trizol for 100mg tissue). For monolayer cells, the cells were lysed directly on the plate by adding 1ml of Trizol per $10cm^3$ as detailed above.

RNA was extracted using PureLink RNA mini kits (Ambion, UK) using the normal phase separation, binding, washing and elution stages as detailed in the manufacturer's protocol [186]. An added step of on-column DNase treatment (PureLink DNAase kit, Ambion, UK) [187] was carried out on the sample to remove genomic DNA contamination. RNA was eluted with 100µl RNAase free water.

Assessment of quality and quantity of RNA

Traditionally, RNA was assessed for purity using agarose gel electrophoresis and by measuring the bands of the 28S and 18S ribosomal RNA, but this was a subjective approach and required large amounts of RNA [188]. A newer method, now determined to be the gold standard for RNA quality assessment, requires only low volumes of RNA (1µl) and uses an automated comparison of the 28S and 18S RNA peaks [189]. The purified RNA was assessed for quality using the Agilent 2100 Bioanalyser and a RNA Nano 6000 Chip Kit (Agilent, California, US). The integrity of

the RNA was assessed using the RNA Integrity number (RIN) value [189]. The RIN is calculated from the electrophoretic trace and is a measure of the integrity of the RNA, it ranges from 10 (intact RNA) to 1 (completely degraded RNA). The bioanalyser also gives the quantity of RNA present in the sample, as well as the 28s:18s ratio, which for pure RNA has a 2:1 ratio.

For this project, samples that had a 28s:18s ratio of ≥1.6 and a RIN value of ≥8 were selected. If necessary, the RNA was normalised to 0.1µg/µl using RNAase free molecular grade water before conversion to cDNA.

cDNA synthesis

cDNA is more stable than single stranded RNA, therefore the mRNA samples were transcribed immediately into cDNA to avoid any degradation and deterioration of the sample. There are three commonly used methods of producing cDNA from single stranded RNA, oligo dT's, specific primers or random hexamers, as shown in Figure 2.1.

.

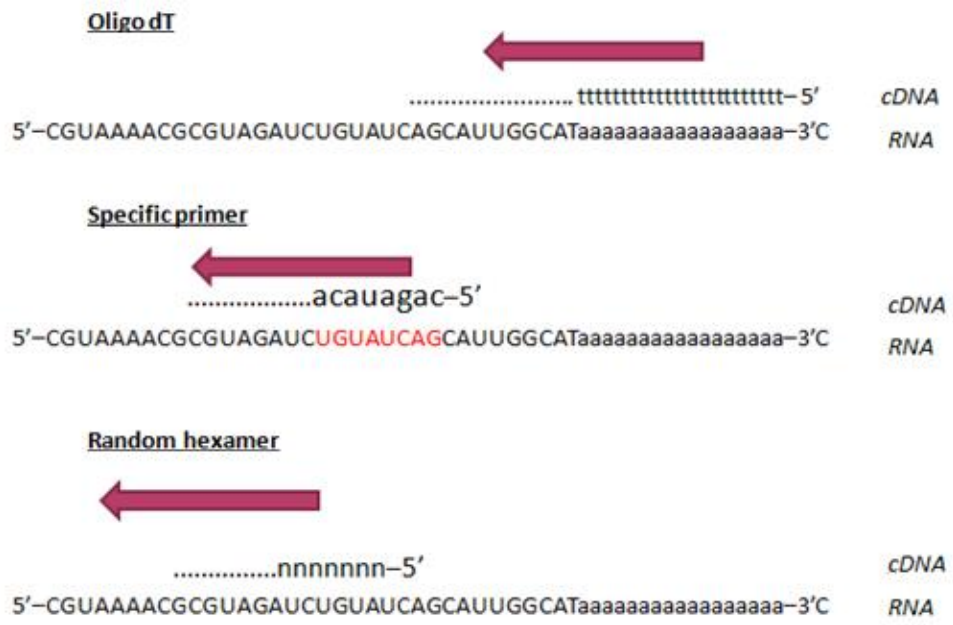


Figure 2.1. The three methods commonly used to reverse transcribe RNA into cDNA – Baird 2013.

OligoDt was chosen in this project as it only hybridises to polyadenylated mRNA at the 3' end polyA tail and therefore minimises genomic DNA contamination. This method is highly specific in only transcribing polyA mRNA, but it can be limited in the fact that transcription may only proceed as far as 1.5KB, such that very long mRNA may not be transcribed completely to the 5' end. Any primers designed for the cDNA should be designed as close to the 3' end of the RNA as possible to ensure efficient conversion.

Reverse transcription was performed using oligoDt primers and Superscript II reverse transcriptase (Invitrogen, California, US) according to the manufacturer's instructions. Briefly, 10µl RNA (0.1µg/µl) was mixed with 1µl oligo-DT (0.5µg/µl

Invitrogen, California, US) and 1µl of 10nM dNTP mix (Invitrogen, California, US). The RNA was heated at 65˚C for 5 minutes, before being quickly chilled on ice. The tube was centrifuged to sediment the contents before 4µl of 5x first strand buffer (Invitrogen, California, US), 2µl of 0.1M DTT (Invitrogen, California, US) and 1µl of RNAsin (40U/µl, Promega, Wisconsin, US) was added. The tube was gently mixed and then heated at 42˚C for 2 minutes. Superscript II reverse transcriptase (1µl, 200U/µl, Invitrogen, California, US) was added and the sample heated at 42˚C for 50 minutes, before the reaction was inactivated by heating at 70˚C for 15 minutes. The cDNA was stored at -20˚C until use in the Q-PCR gene expression studies.

General analysis and statistical methods

All genotyping data were analysed using freely available software PLINK [190, 191] implemented using the program BCgene (BC Platforms, Finland).

Quality assurance (QA) was carried out on the genotyping data to ensure only the top quality data was used for analysis. Cut off values were assigned for each of the following values and are specified in each individual chapter:

- Minor allele frequency - frequency of less common allele in a population (denoted in PLINK by --maf)
- Exact Hardy-Weinberg equilibrium in controls (denoted in PLINK by --hwe)
- SNP and individual call rates - number of SNPs that have alleles assigned (denoted in PLINK by --geno and --mind respectively).

Case control associations were performed on the genotype data using a standard case control allelic association unless otherwise detailed [192, 193]. This is a basic chi-squared test of association (using a 2x2 contingency table with 1 degree of freedom) of the allele counts between cases and controls, regardless of the genotype combination from which they arise. This assumes a multiplicative model of inheritance, so that the risk of the disease increases by factor *r* for each allele carried. For example with genotypes AA, AT and TT, the risk increases by *r* for each T allele carried – AA = 1, AT = *r*, TT = r^2 [192, 194].

A Cochran-Mantel-Haenszel (CMH) test (2x2xK) was performed on the genotype data when a stratified analysis was needed [195, 196]. This is a cluster based analysis that generates an average odds ratio that controls for confounding factors between clusters. This is referred to as a stratified analysis.

Meta-analyses use statistical methods to compare similar studies with a similar hypothesis to examine results as a whole and to make the best use of all the information from all samples available, increasing statistical power. This analysis was used where appropriate to combine results across multiple breed analyses [197]. This was run in PLINK using the command line --meta-analysis with the appropriate association data files written in.

Haploview case control analysis was performed to determine haplotype associations within the genotype datasets [198]. This was implemented using PLINK and Haploview. The standard Gabriel method was used to define haplotype blocks [199].

Significance of the data was determined by *p*-values, a raw *p*-value of ≤0.05 or a corrected *p*-value (see below) was used to infer significance in the data unless otherwise stated.

Corrections for multiple testing (where applicable) were applied to counteract multiple testing [200]. There are various methods to use to correct for multiple testing, Bonferroni, permutation testing, or the Benjamini & Hochberg false discovery rate (FDR). Bonferroni correction multiples the *p*-value by the number of tests being carried out to lower the chances of type 1 error (false positives). As the highest accepted individual raw *p*-value for the GWAS data would have to be 0.00000025 (0.00000025 x 200,000 = 0.05) this makes the test very conservative, especially given that dogs have very high LD, so the actual number of independent tests will be less than 200,000. Permutation testing involves the cases and controls being rearranged randomly and re-analysed to assess significance. It is standard to use 1000 [201] or 100,000 [202] permutations for correction in genetic studies. FDR testing consists of sorting the *p*-values in ascending order, then dividing each observed *p*-value by its percentile rank to get an estimated FDR. FDR gives the

expected percentage of false positives among all the claimed positive associations. The three methods will all be used in this thesis and the results compared.

Hardy-Weinberg equilibrium (HWE) states that in an outbred, randomly mating population the allele frequencies will remain constant from one generation to another [203, 204]. It is represented by the equation: $p^2 + 2pq + q^2 = 1$ where p and q are the frequencies of alleles for a locus in the population, p^2 and q^2 represent the proportions of each homozygote and 2pq represents the proportion of heterozygotes. In an inbred population (such as pure-breed dogs) the populations are non-randomly mated and therefore the allele frequencies will not conform to the HWE. The observed allele frequencies are compared to the expected allele frequencies using an exact test designed by Wiggington et al. 2005 [205] in PLINK, to produce a *p*-value which denotes the deviation of the SNPs from HWE.

The odds ratio (OR) is used to provide an estimation of relative risk of an allele in the population [206]. It is a measure of magnitude of association between the cases and controls and can be calculated by dividing the odds of the allele in the cases by the odds of the allele in the control population:

$$OR = \frac{s \div (1-s)}{f \div (1-f)}$$

's' is the frequency of a disease-associated variant in cases and 'f ' is its frequency in controls. The OR is used to determine the direction of association: >1 indicates that

allele 1 increases the risk of disease compared to allele 2 and <1 allele 2 increases risk of disease compared to allele 1, if the OR is equal to 1 then there is no difference in allele frequency between cases and controls (null hypothesis) [206]. The 95% confidence interval (CI) is the range at which the true value is thought to lie; 95% of the population values will fall between the two numbers given (upper and lower limit). The 95% CI of the OR is used to determine significance of the results; If the 95% CI contains 1 (null hypothesis) then the result should be considered non-significant, if the 95% CI does not contain the null hypothesis then the *p*-value should be significant, but if the window is window is wide then the certainty that the results are meaningful is reduced.

Chapter 3.

Genome Wide Association Study (GWAS)

Introduction

High density whole genome scanning provides an unbiased approach for scanning the entire genome to identify novel candidate genomic regions of disease association. Previously, candidate genes were selected based on their perceived involvement in the disease or disease process. Many genes were not screened for disease association because their function was unknown or was not thought to be relevant to disease susceptibility, onset or progression [207]. Consequently many associations may have been missed for not being in obviously "associated" genes.

Genome wide association studies (GWAS) have been successfully been conducted for many human diseases, the most prominent paper evaluating the use of GWAS in common human diseases is the Welcome Trust Case control Consortium (WTCCC) [208]. This paper evaluates genetic contribution in seven common genetic diseases using 14,000 cases and 3,000 controls. Considering the dog has a unique population structure (highly inbred, genetically isolated populations with extended LD) fewer genetic markers and smaller sample sizes are required to find disease associations in dogs than in human studies [144, 150]. In human studies, to achieve whole genome coverage you require 200,000 to 500,000 markers [144] but this has significantly increased in recent years – the newest generation of Affymetrix human arrays now contains nearly one million SNPs (906,600, Human SNP Array 6.0). In canine studies this number is much reduced with as little as 10,000 markers needed [142] and the newest canine SNP chip contains nearly 200,000 markers (Illumina CanineHD chip).

High density SNP arrays have been successfully used to determine genetic involvement in canine disorders including atopic dermatitis, epilepsy and pituitary dwarfism as well as non-disease traits such as coat colour [174, 177, 209, 210]. These arrays are available for genome wide analyses of many veterinary species including the dog and are commercially available from Illumina (www.illumina.com) and Affymetrix (www.affymetrix.com). As these arrays offer so much in terms of screening the genome for regions of disease association and generate high quality data unobtainable by other means, a genome wide approach was used to examine the canine genome for SNP associations with CCL rupture in Newfoundland dogs.

Causative mutation(s) may be ancient or quite modern in origin [168]. If they are ancient they will have survived the selective breeding pressures of breed creations to form modern dog breeds and as a result will have been fixed in distantly related dog breeds. If the mutation is relatively modern the mutation will only be present in discreet clusters of one or a few modern breeds. GWAS is a powerful and relatively inexpensive approach to screening multiple dogs from a variety of breeds to try and determine the origin of the mutation(s) and whether or not it/they is/are present in different dog breeds. Multiple breeds may share the same phenotype caused by the same causal genetic variant. This is the case for a number of conditions, such as chondrodysplasia (dwarfism) and the brachycephalic (short snout) trait [211, 212]. This phenotype sharing due to the same causal genetic variation/region of interest was identified through GWAS using multiple breeds. Although we are only investigating CCL rupture via GWAS in one breed, Newfoundlands, it would be

interesting to examine the associations in other breeds and to determine whether the same associations for CCL rupture are fixed in the different breeds – this is beyond the scope of the current project.

This is the first GWA study that has been conducted on CCL rupture using the high density CanineHD array. It is hoped that this study will allow novel mutations that have not previously been associated with CCL rupture to be determined in a cohort of Newfoundland dogs.

Methods

Sample collection and preparation

Samples collection was either via salivary swabs (Oragene™, DNA genotek, Ontario) or from residual blood excess from veterinary investigations already being carried out (for more details see the main methods Chapter 2). The DNA was extracted as described in the main methods (Chapter 2) and then normalised to 100ng/µl using a Nanodrop spectrophotometer.

DNA from 96 Newfoundlands – 48 North American (NA) and 48 European (EU) Newfoundlands (24 cases and 24 controls in each of the populations) was sent to Roslin Institute for genotyping on the Illumina CanineHD array.

Illumina genotyping

The genome wide SNP genotyping of Newfoundland dog DNA samples was performed using the Illumina CanineHD chip (San Diego, California, US), with the hybridisation being performed at the Roslin Institute in Edinburgh. The SNPs on the chip were selected from the CanFam2 assembly of the dog genome. The new SNP chip (CanineHD) contains many more SNPs than the original Canine SNP chip array from Illumina (CanineSNP20), which contained 22,362 validated SNP probes; it also has greater coverage across the whole genome. The CanineHD SNP chip contains 173,662 markers evenly spaced across the genome with an average coverage of greater than 70 markers per megabase (Mb). This allows an unbiased approach to screening the whole of the canine genome for differences associated with normal or disease phenotypes.

The genotyping was carried out using the standard manufacturer's protocol (www.illumina.com) at the Roslin Institute in Edinburgh. Briefly, the genomic DNA was denatured and neutralised before being isothermally genome wide amplified. The amplified DNA was enzymatically fragmented in a controlled process so as not to over fragment (and which does not require a gel electrophoresis). After alcohol precipitation the DNA was resuspended in hybridisation buffer. The DNA was annealed to specific probes attached to beads on the surface of a chip. Un-hybridised DNA was washed off, before a single base extension of the probe with labelled nucleotides (to confer allelic specificity). Fluorescent intensities were detected by the Illumina software and turned into an automated genotype call.

The GWAS data were analysed using a freely available whole genome association analysis toolset PLINK [190, 191] implemented using the program BCGene (BC Platforms, Finland). Graphical representation of the data was performed using the GenABEL and ggplot packages in R [213].

Quality Assurance

Raw genotypes from the Illumina final report file were directly uploaded to the BCgene program. The GWAS data were then assessed and filtered for quality assurance (QA) before association testing. Samples and SNPs were excluded if the sample call rate was <85%, the genotype call rate <90%, the control population was not in Hardy-Weinberg equilibrium (p<0.05) and/or the minor allele frequency ≤0.05. Following QA, 81 samples remained. The data were then assessed for population stratification and any outliers deviating from the main cluster(s) were removed from subsequent analyses – five samples were classed as outliers and removed, leaving 76 samples for further analysis.

Assessing population structure

As samples were sourced from two different geographical locations (North America and Europe), the data had to be assessed for population stratification. Underlying population stratification is a common confounder that could introduce bias to the analyses and reported allelic associations, resulting in higher numbers of false positive and false negative associations [214, 215]. By checking and accounting for

underlying population stratification, the number of spurious associations is greatly reduced [216]. Population stratification can be assessed in many ways [214, 217-220] including; genomic control [221], Eigenstrat principle component analysis (PCA) [216], or multidimensional scaling (MDS) [222].

In the early genetic studies, genomic control [221] was the primary method for correcting population stratification. This is where the data is adjusted by the common genomic inflation factor (λ). The data is visualised on a quantile-quantile (Q-Q) plot by plotting observed data against the expected data. Any deviation from the straight line indicates stratification within the dataset. The lambda value (λ) is the estimation of the distribution of the *p*-values and whether they conform to expected data; any stratification within the results leads to deviation of this value away from one. The genomic control method has been criticised for being over conservative in its results and also for being poor in correcting for stratification if the stratification was large, if there were discrete structures within the subjects [214, 218], or if there is a high level of relatedness within the population (which is the case with pure-bred dog breeds) [223].

An alternative approach for correcting for population substructure is the Eigenstrat principle component analysis (PCA) [216, 219]. This works by computing principle components (clusters) within the datasets and the top two principle components are indicative of the structure within the dataset. The samples are assigned clusters based on their ancestry and variation of the markers to each other. They are

assigned an axis and are plotted as if in a three dimensional space, with related samples clustering together. The clusters within the data can be visualised graphically by plotting the first two axes of variation against each other on an XY scatter plot.

A third approach for determining genetic heterogeneity within the datasets is MDS [224], which is similar to PCA. It uses a matrix of pairwise identity by state (IBS) values between individuals, to compute the principle components in the data, which are then plotted in the same way as the PCA. An increasingly popular approach to examining stratification in the analysis involves running an Efficient Mixed-Model Association eXpedited analysis (EMMAX) [217]. This takes into account the kinship structure within the subjects to correct for any stratification; it also performs a case control association analysis at the same time.

The method chosen to correct for stratification in this study was the Eigenstrat principle component analysis (PCA), which is believed to be the most robust and powerful approach for correcting for substructure within genome wide association datasets [216, 225]. If any samples are visually seen to deviate from the main cluster(s) they are classed as outliers and are removed from all subsequent analyses. If it is clear that there are two or more separate clusters of data, then any further analyses should take the grouping into account. A stratified allelic association analysis using the clusters as the groups – such as a Cochran–Mantel–Haenszel (CMH) 2x2xK, or by performing the analysis separately on each individual

cluster and combining results with a meta-analysis, or an EMMAX analysis can be run as an alternative.

Assessing Significance

To determine whether SNPs were significant or not, the significance level needed to be established. If the standard significance value was used (*p*-value ≤0.05) for the GWAS data there would be 10,000 SNPs that would reach statistical significance by chance alone (200,000 X 0.05). This is nonsensical and therefore a more appropriate level of significance is needed such that the number of SNPs reaching significance is more realistic. Human GWAS studies use a significance level that indicates a *p*-value of $1x10^{-7}$ as being very highly associated and a *p*-value of $1x10^{-5}$ being reasonably associated [208]. In this project, the significance threshold was set so that any association below $1x10^{-5}$ was considered associated with CCL rupture susceptibility in dogs. This has been considered suitable in other canine GWAS studies [177] because dogs have much higher LD than humans.

Haplotype analysis of GWAS data

Different combinations of alleles in a population that are inherited together are called a haplotype [226]. Using PLINK and Haploview, a case–control haplotype association analysis can be performed on the genome wide data. The linkage and recombination relationship between each SNP on the chromosomes is determined. A 1000 permutations test was used to correct the haplotype block raw *p*-values to counteract multiple testing issues.

Haplotype blocks were defined by the Haploview software (version 4.2), using the standard Gabriel method [199]. Recombination between the SNPs can be estimated using D'. If SNPs are in strong LD, D' = 1, if SNPs are independent of each other D' = 0. The default Gabriel settings (as implemented by the Haploview software), are that if the upper 95% confidence bound on D' is >0.98 and the lower bound is >0.7 then there is LD between the SNPs. A haplotype block is formed when less than 5% of markers show recombination (i.e. 95% of markers are in strong LD with each other).

The haplotype blocks can be visualised by using a heatmap, which is a triangular matrix of pairwise dependencies between SNPs that are within 500KB of each other. The colour depth of each block indicates the strength of the LD within each block/between each marker. Bright red indicates markers that have very high LD and are more likely to be inherited together (LOD = 2, D' = 1), white indicates no LD between markers (LOD = <2, D' = <1) and pink/pale red and blue are for markers that have intermediate LD or can be indicative of a monomorphic SNP (LOD = 2, D' = <1, or LOD = <2, D' = 1). The number shown in each diamond on the heatmap is indicative of the D' value between each SNP. Each SNP is listed above the heatmap and they are ordered by position on chromosome. The upward facing sides of the diamond indicate which SNPs are in LD with one another; an example heat map with explanation of the details is shown in Figure 3.1

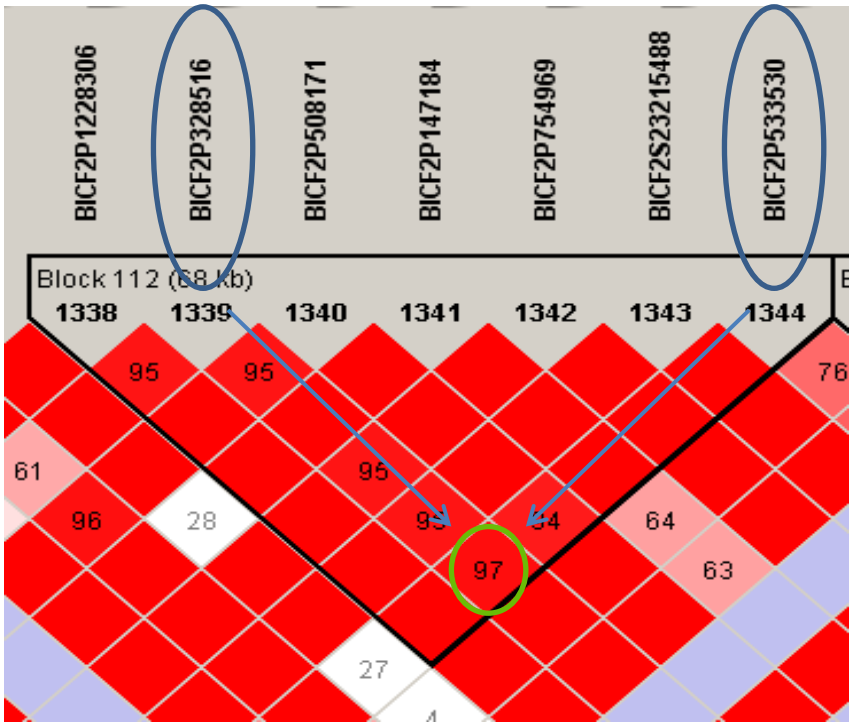


Figure 3.1. Haplotype LD plot example. Number circled in green is D' value between SNPs circled in blue. Colour is rough measure of significance. Block number (and size) is shown and detailed by black triangle. The numbers along the top block indicate the sequential order of SNPs along the chromosome. The SNP is shown vertically above each diamond.

Overview of genome wide data analyses

The overview of the workflow analyses for the genome wide association study is summarised diagrammatically in Figure 3.2

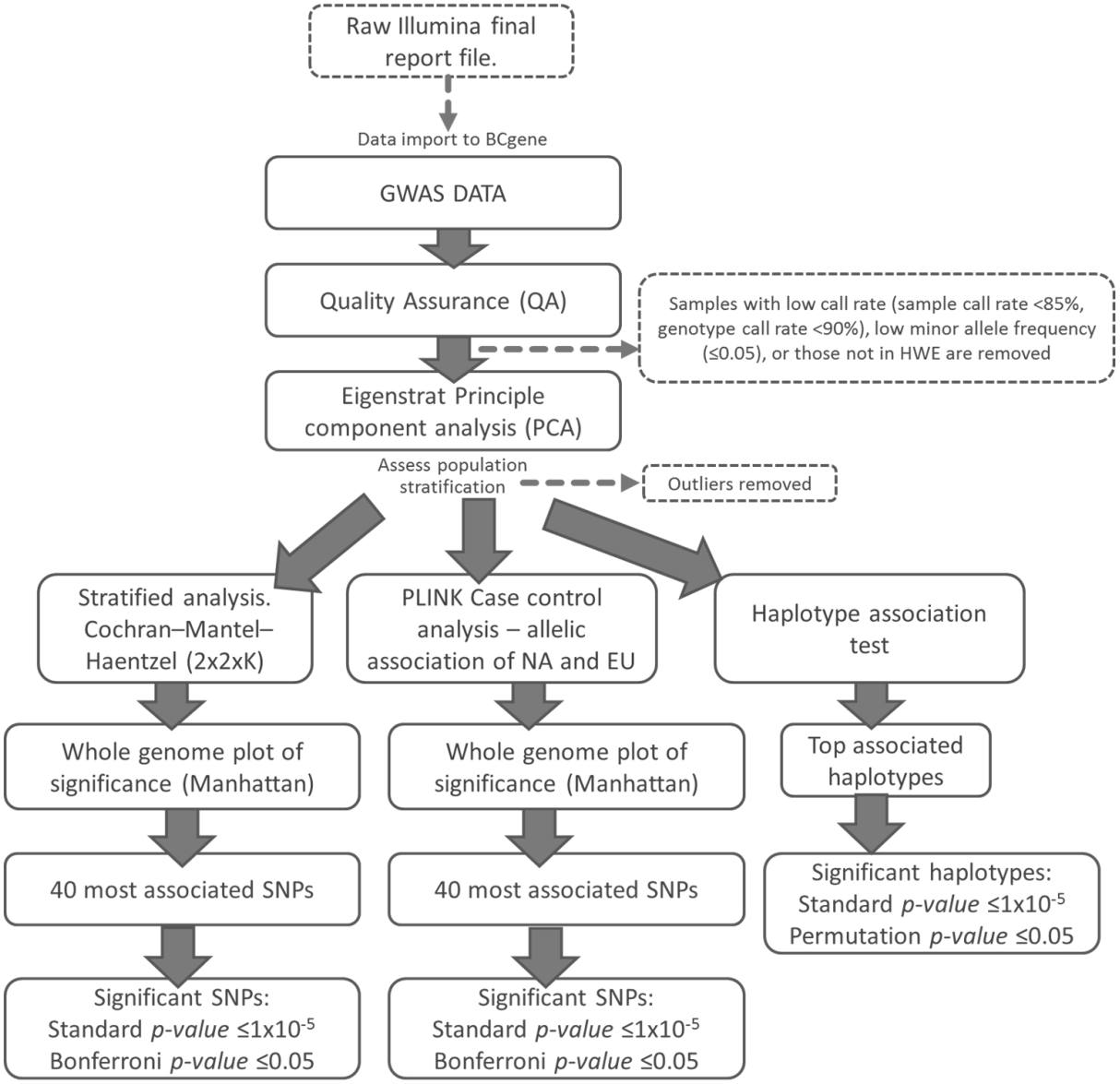


Figure 3.2. Workflow methodology for the GWAS study. GWAS data were imported into the BCgene program and subjected to QA testing. Data were excluded if the sample call rate was <85%, the genotype call rate <90%, the control population was not in Hardy-Weinberg equilibrium (*p*<0.05) and/or the minor allele frequency ≤0.05. Population stratification was then assessed and outliers removed before data were analysed either using the CMH stratified analysis or a case control chi squared test. Haplotype association tests were also conducted on the data.

Results

Population stratification within the Newfoundland dogs

PCA analysis of the GWAS data as a whole (without regard to CCL rupture status), identified two discreet clusters evident within the dataset (Figure 3.3) which were indicative of the country of origin of the Newfoundland dog. The QQ plots of the data and their respective λ values are shown in Figure 3.4. For the unadjusted dataset, λ was 1.31 and when the data were stratified by country using a CMH test, this was reduced to 1.27, after conducting an EMMAX analysis on the data, the λ was further reduced to 1.02.

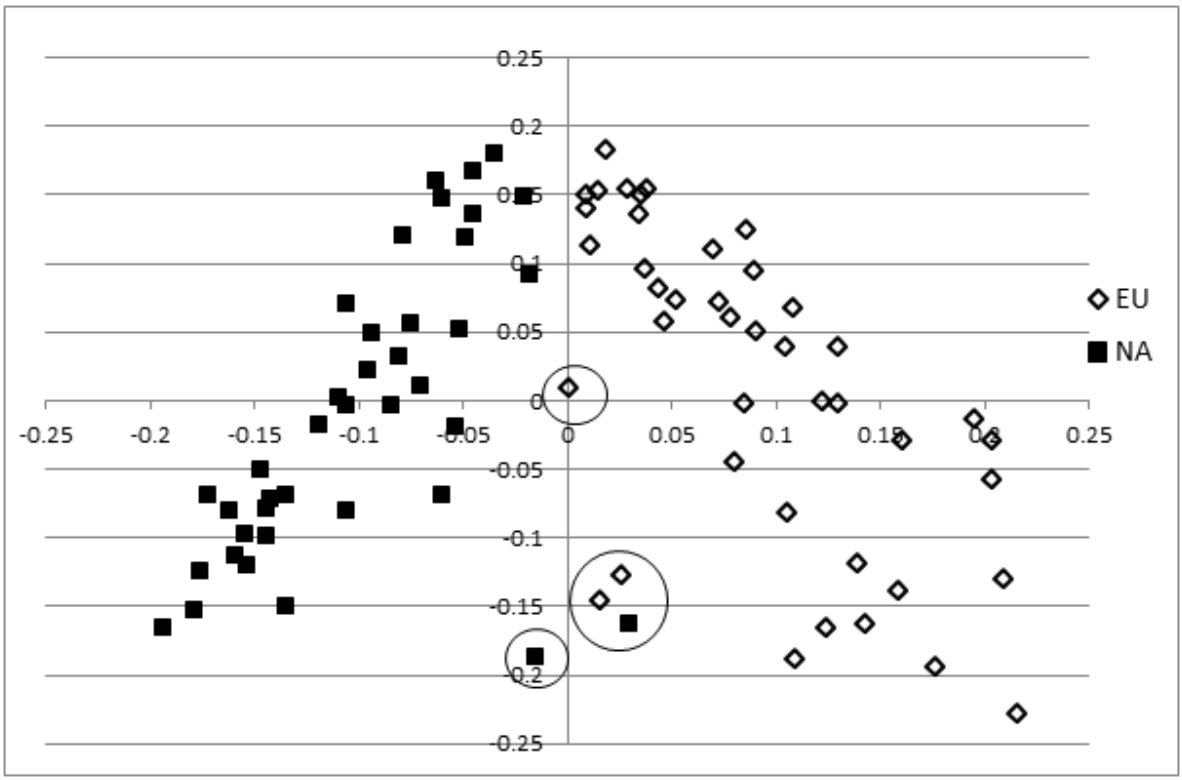


Figure 3.3. Eigenstrat PCA plot of the 81 Newfoundland dogs that satisfied the QA criteria. Outliers (n=5, circled) were excluded from any subsequent analyses of the data, as they deviated from the main clusters. Black squares represent North American dogs, white diamonds indicate European dogs.

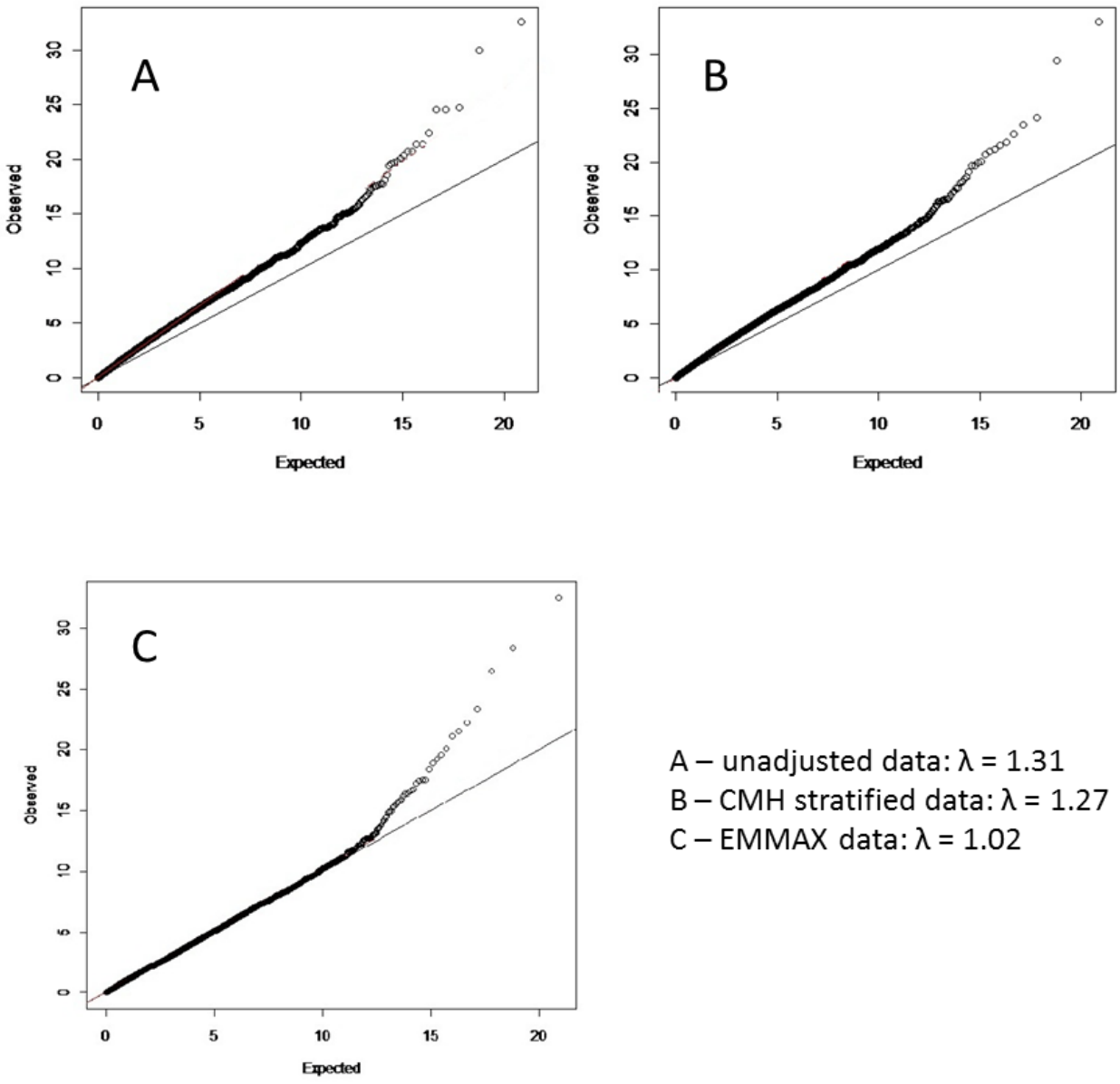


Figure 3.4. QQ plots and λ values of the GWAS data; A = unadjusted data, B = Stratified by country (CMH test), C = EMMAX analysis

High Density Illumina Array - Newfoundlands

Manhattan plots were produced for each GWAS analysis: CMH stratified (Figure 3.5), European (Figure 3.6), North American (Figure 3.7) and EMMAX (Figure 3.8). This data representation displays *p*-value*s* for each SNP based on the genotype frequencies when comparing their incidence in healthy dogs to those with CCL rupture. Each SNP is visually represented with its -log *p*-value against chromosomal location. Anything with a *p*-value of $\leq 1x10^{-5}$ (above dotted line of significance) was considered to have genome wide statistical significance indicating whether it was over or under represented in the CCL rupture group of dogs.

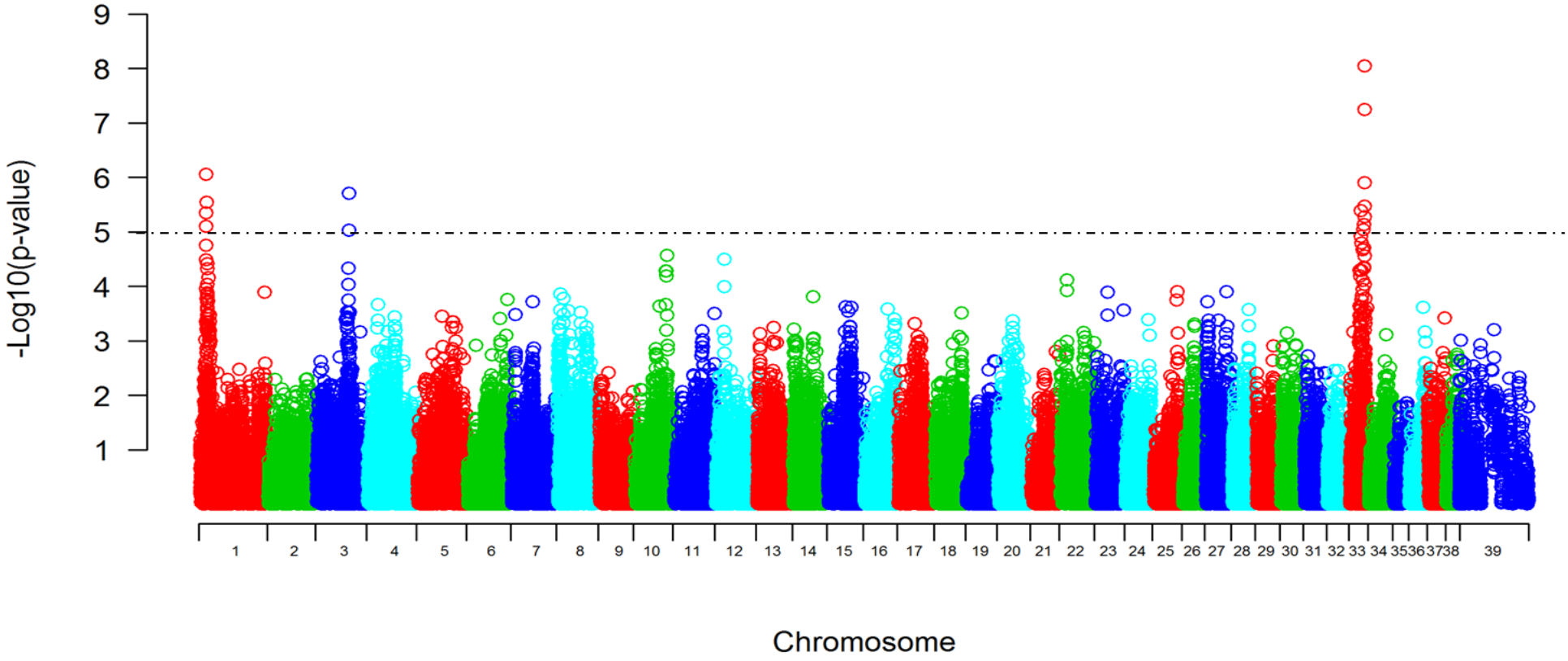


Figure 3.5. Manhattan plot for all Newfoundland dogs following the CMH stratified GWAS analysis. The dotted line represents the *p*-value threshold of $1x10^{-5}$ to indicate genome wide significance of SNPs associated with CCL rupture.

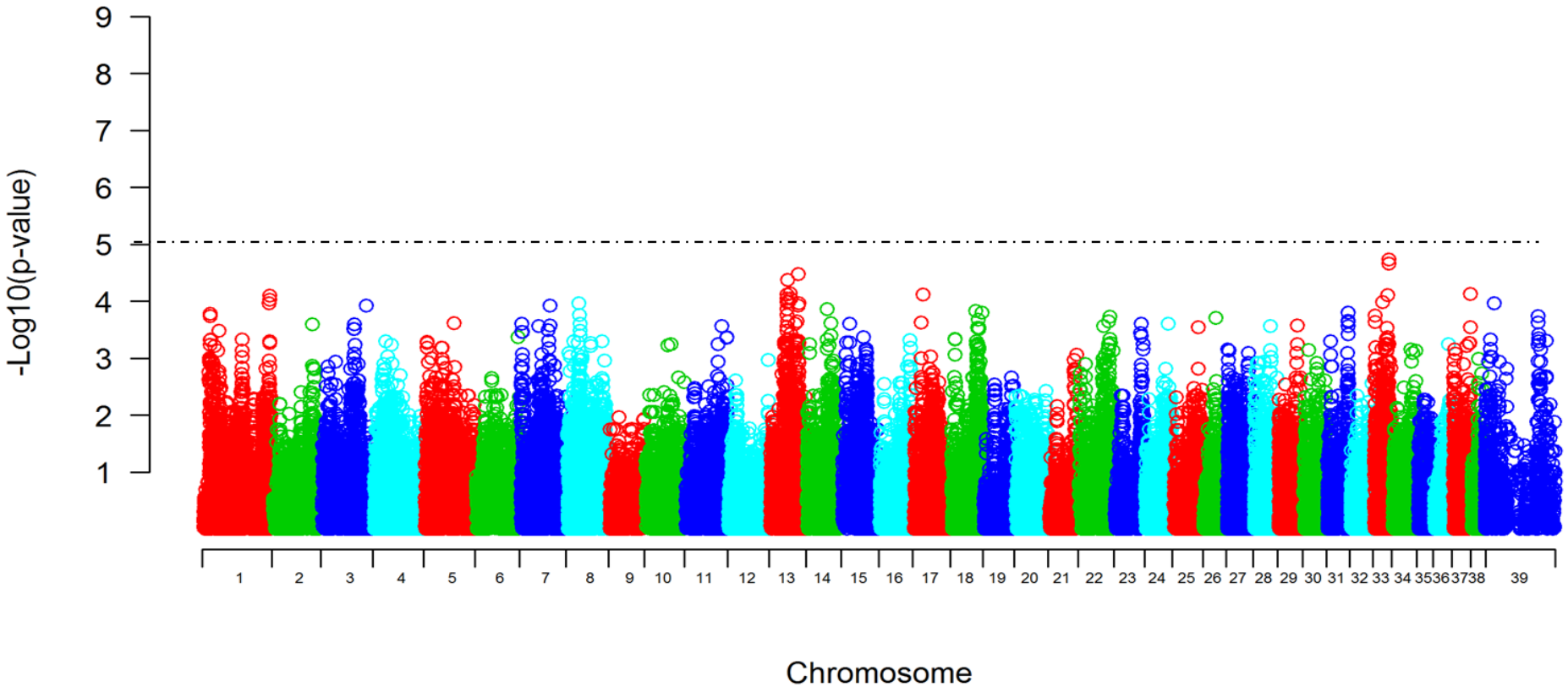


Figure 3.6. Manhattan plot for EU Newfoundland dogs following genome wide association analyses. The dotted line represents the *p*-value threshold of $1x10^{-5}$ for suggested genome wide significance of SNPs associated with CCL rupture.

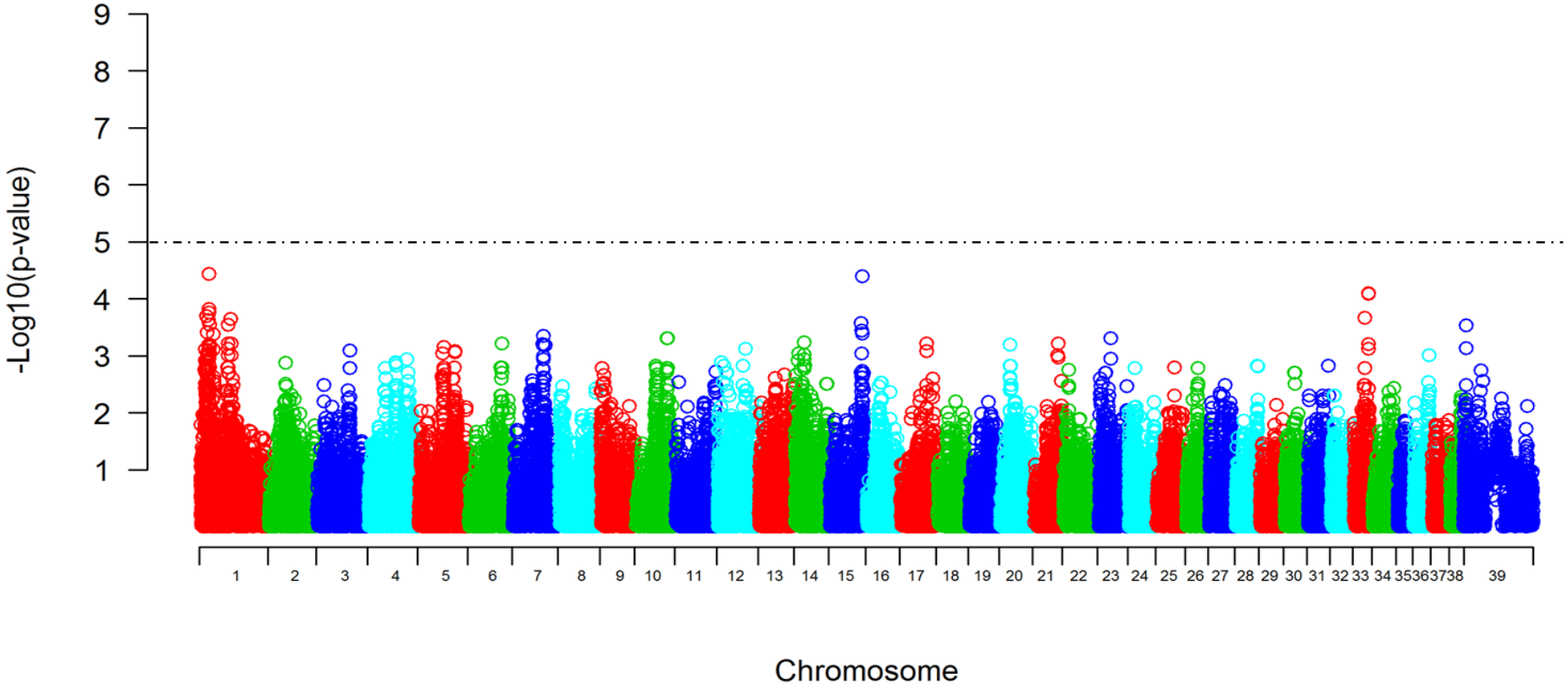


Figure 3.7. Manhattan plot for NA Newfoundland dogs following genome wide association analyses. The dotted line represents the *p*-value threshold of 1x10-5 for suggested genome wide significance of SNPs associated with CCL rupture.

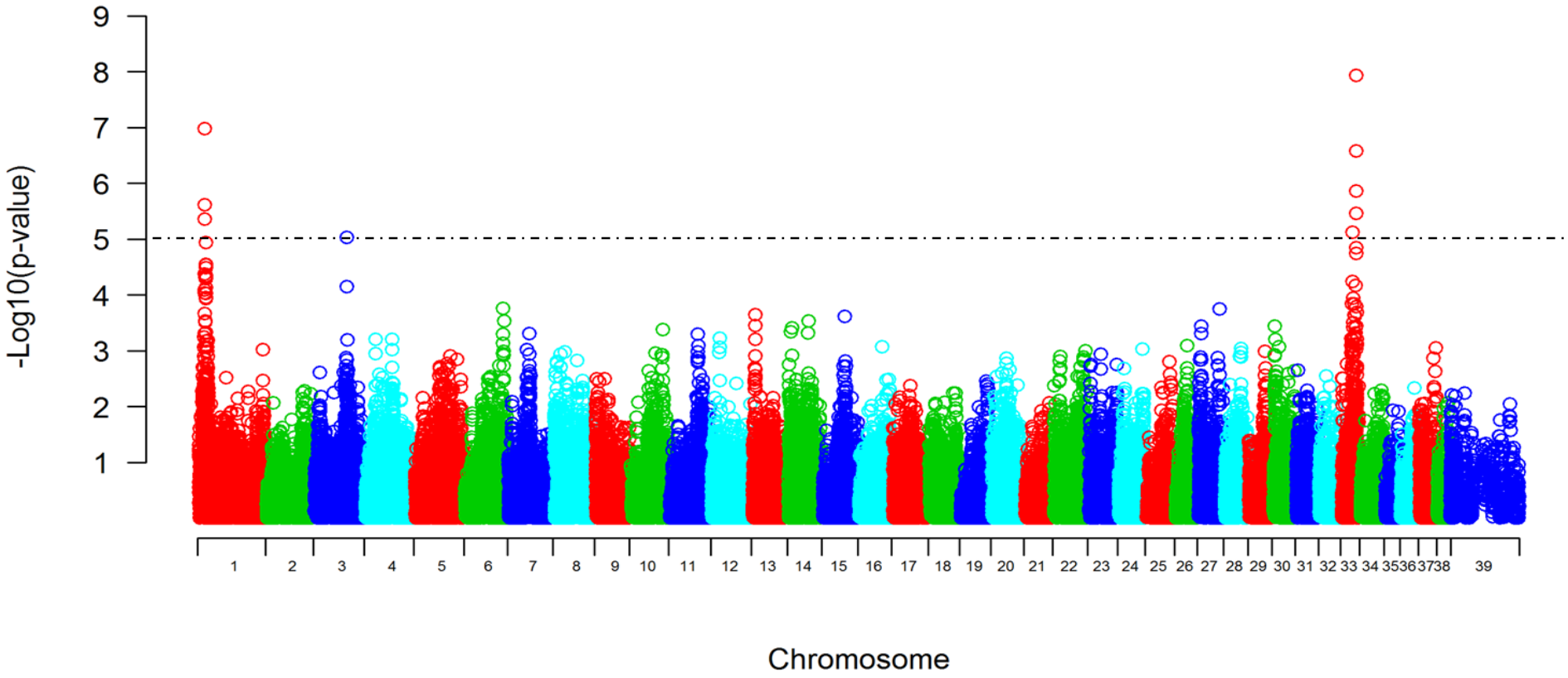


Figure 3.8. Manhattan plot for all Newfoundland dogs tested after genome wide association analyses stratified by EMMAX. The dotted line represents the *p*-value threshold of 1x10-5 for suggested genome wide significance of SNPs associated with CCL rupture.

A comparison of the different analyses (stratified, meta- and EMMAX) showed that variation was low and that similar SNPs reached statistical significance ($p \leq 1x10^{-5}$) although their rank order was different - Table 3.1 shows this point by ranking the ten most associated SNPs in each of the 3 different analyses.

Table 3.1. The ten most significant SNPs ranked by *p*-value from each different analyses of the GWAS data.

SNP	Chromosome	Rank in CMH stratified	Rank in Meta	Rank in EMMAX
BICF2S23133650	33	1	1	1
BICF2P1311062	33	2	2	3
BICF2P59100	1	3	3	2
BICF2P797863	33	4	4	4
BICF2G630343527	3	5	8	9
TIGRP2P8419_ rs8959368	1	6	9	-
BICF2P486041	33	7	-	6
BICF2G63080008	33	8	-	8
BICF2P1008542	1	9	5	5
BICF2P212467	33	10	7	-
BICF2S2318592	1	-	6	7
BICF2P1457397	33	-	10	-
BICF2P893554	1	-	-	10

The 40 SNPs showing greatest association (raw *p*-values <0.05) from the CMH stratified analysis are shown in Table 3.2

Table 3.2. The 40 SNPs showing greatest association (by raw *p*-value) from the CMH genome wide association stratified analysis.

Rank	CHR	SNP	Position	A1	MAF	A2	*p* (raw)	*p* (BONF)	*p* (PERM)	FDR Benjamini and Hochberg *p*	OR	95% CI
1	33	BICF2S23133650	25987425	G	0.32	A	9.08E-09	0.0009	0.0003	0.0009	0.10	0.04 - 0.24
2	33	BICF2P1311062	26001662	G	0.35	A	5.76E-08	0.0058	0.0014	0.0029	0.13	0.06 - 0.28
3	1	BICF2P59100	10792454	A	0.43	C	8.90E-07	0.0901	0.0213	0.0300	5.96	2.85 - 12.46
4	33	BICF2P797863	26063962	C	0.50	T	1.25E-06	0.1268	0.0296	0.0317	0.18	0.09 - 0.37
5	3	BICF2G630343527	59316812	A	0.39	T	1.99E-06	0.2019	0.0473	0.0404	0.17	0.08 - 0.36
6	1	TIGRP2P8419_ rs8959368	11497808	G	0.28	A	2.88E-06	0.2912	0.0668	0.0485	0.16	0.07 - 0.36
7	33	BICF2P486041	25983710	C	0.47	T	3.42E-06	0.3465	0.0791	0.0495	0.19	0.09 - 0.40
8	33	BICF2G63080008	19430035	G	0.33	A	4.10E-06	0.4156	0.0952	0.0507	0.18	0.09 - 0.39
9	1	BICF2P1008542	10756573	C	0.38	T	4.51E-06	0.4566	0.1032	0.0507	5.65	2.62 - 12.2
10	33	BICF2P212467	25923850	A	0.38	G	5.26E-06	0.5327	0.1179	0.0533	0.19	0.09 - 0.40
11	33	BICF2P1072408	24536212	T	0.34	A	7.49E-06	0.7588	0.1621	0.0669	0.18	0.09 - 0.40
12	1	BICF2S2318592	10822495	G	0.37	A	7.92E-06	0.8022	0.1702	0.0669	5.49	2.53 - 11.91
13	33	BICF2P325662	23291566	C	0.46	T	9.13E-06	0.9244	0.1947	0.0678	0.22	0.11 - 0.43
14	3	BICF2G630343620	59381030	C	0.39	T	9.37E-06	0.9490	0.1994	0.0678	0.20	0.09 - 0.41
15	33	BICF2G63079918	19502373	T	0.28	C	1.21E-05	1.0000	0.2486	0.0817	0.18	0.08 - 0.40
16	33	BICF2S23216989	19794189	C	0.43	T	1.62E-05	1.0000	0.3135	0.1027	0.22	0.11 - 0.45
17	1	BICF2G630711545	10238823	T	0.43	C	1.79E-05	1.0000	0.3385	0.1068	4.58	2.25 - 9.34
18	33	BICF2P1457397	24386039	T	0.36	G	2.00E-05	1.0000	0.3681	0.1126	0.22	0.11 - 0.45
19	33	BICF2P774690	22930550	T	0.45	C	2.13E-05	1.0000	0.3858	0.1138	0.22	0.11 - 0.45
20	10	BICF2S23060690	58719181	T	0.14	C	2.74E-05	1.0000	0.4579	0.1329	0.08	0.02 - 0.32
21	33	BICF2S23241903	26080255	C	0.23	T	2.75E-05	1.0000	0.4588	0.1329	0.17	0.07 - 0.40
22	12	BICF2S24413446	14748689	T	0.09	G	3.18E-05	1.0000	0.5038	0.1444	0.00	0.00 - 0.00
23	1	BICF2G630711540	10231948	G	0.43	A	3.28E-05	1.0000	0.5122	0.1444	4.43	2.16 - 9.10

Rank	CHR	SNP	Position	A1	MAF	A2	*p* (raw)	*p* (BONF)	*p* (PERM)	FDR Benjamini and Hochberg *p*	OR	95% CI
24	1	BICF2S23354689	13375903	T	0.32	C	3.89E-05	1.0000	0.5660	0.1607	0.22	0.11 - 0.47
25	33	TIGRP2P389380_rs8613571	24503510	A	0.37	G	4.44E-05	1.0000	0.6104	0.1607	0.23	0.11 - 0.47
26	3	BICF2G630343137	58878743	A	0.48	G	4.64E-05	1.0000	0.6244	0.1607	4.10	2.06 - 8.17
27	33	TIGRP2P389430_rs9122813	24703426	C	0.42	T	4.72E-05	1.0000	0.6301	0.1607	0.25	0.12 - 0.49
28	1	BICF2P893554	12724538	T	0.42	G	4.73E-05	1.0000	0.6312	0.1607	4.33	2.10 - 8.89
29	33	TIGRP2P388471_rs9063975	19813762	A	0.46	G	5.08E-05	1.0000	0.6525	0.1607	0.25	0.13 - 0.49
30	33	BICF2G63081807	16640349	C	0.38	T	5.09E-05	1.0000	0.6533	0.1607	0.23	0.11 - 0.48
31	10	BICF2S237368	57091773	G	0.13	A	5.24E-05	1.0000	0.6626	0.1607	0.08	0.02 - 0.33
32	10	BICF2S23110665	57419768	A	0.13	G	5.24E-05	1.0000	0.6626	0.1607	0.08	0.02 - 0.33
33	10	BICF2P1370043	57892601	A	0.13	G	5.24E-05	1.0000	0.6626	0.1607	0.08	0.02 - 0.33
34	33	TIGRP2P388138_rs8754330	18223573	G	0.26	A	6.43E-05	1.0000	0.7292	0.1873	0.21	0.09 - 0.46
35	10	BICF2P1384039	57622160	G	0.16	A	6.47E-05	1.0000	0.7308	0.1873	0.14	0.05 - 0.41
36	1	BICF2G630711764	14988276	C	0.38	T	6.91E-05	1.0000	0.7505	0.1944	0.25	0.12 - 0.50
37	22	BICF2G630317860	10772478	T	0.40	C	7.63E-05	1.0000	0.7807	0.2079	0.25	0.13 - 0.51
38	33	BICF2P29702	22652694	T	0.41	C	7.80E-05	1.0000	0.7871	0.2079	0.25	0.13 - 0.51
39	3	BICF2G630343157	58927121	C	0.47	T	9.23E-05	1.0000	0.8319	0.2303	3.84	1.94 - 7.61
40	1	BICF2P1077045	12732415	T	0.41	C	9.27E-05	1.0000	0.8328	0.2303	4.05	1.98 - 8.27

KEY: MAF = minor allele frequency, A1 = base for allele 1, A2 = base for allele 2, OR = odds ratio of A1, CI = confidence interval of the odds ratio. The position is the chromosomal position on the CanFam3 genome annotation. Three methods of correction for multiple testing are shown, BONF = Bonferroni, FDR and PERM = 100,000 permutations.

SNPs positively or negatively associated with CCL rupture were present on chromosomes 33, 1, 3, 10, 12 and 22. Chromosome 33, in particular showed a strong linkage with CCL rupture; 19 of the 40 SNPs showing greatest association with CCL rupture (by raw *p*-value) were located to that chromosome. Of the ten most associated SNPs (all below the standard *p*-value cut-off of $p_{raw} \leq 1x10^{-5}$), six SNPs located to chromosome 33, three SNPs to chromosome 1 and one SNP to chromosome 3.

The two highest ranking CCL rupture associated SNPs both located to chromosome 33 and were significant after correction for multiple testing using Bonferroni. When 100,000 permutations were run on the data, five SNPs showed significance (three on chromosome 33 and one SNP on chromosome 1 and 3). When FDR was applied seven SNPs were significant (four on chromosome 33, two on chromosome 1 and one on chromosome 3).

Haplotype associations

Haplotype analyses were performed on each of the three main chromosomes showing regions of association with CCL rupture (chromosome 1, 3 and 33). Significant haplotype blocks were identified in two of the three chromosomes (chromosome 33 - Figure 3.9, and chromosome 1 - Figure 3.10). Chromosome 3 did not have any associated haplotype blocks that reached statistical significance (raw $p = \leq 1x10^{-5}$). A representation of the haplotype blocks for each chromosome is shown to indicate the physical distances between each of the blocks and their

position on the chromosome (chromosome 33 - Figure 3.11 and chromosome 1 - Figure 3.12). The haplotype frequencies for each allele combination are shown in Table 3.3 for chromosome 33 and Table 3.4 for chromosome 1.

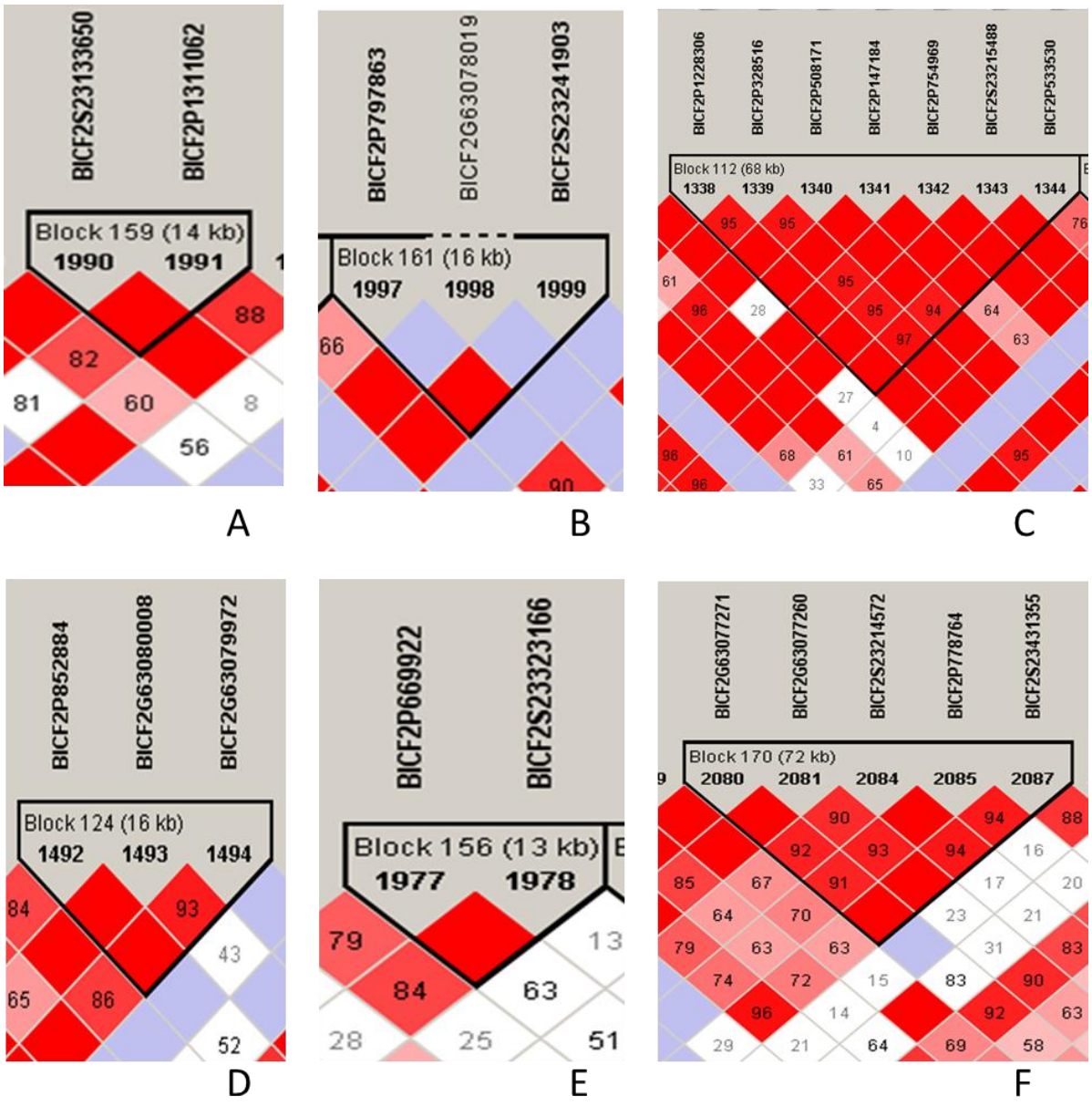


Figure 3.9. LD plots showing haplotype blocks reaching statistical significance (raw *p*= ≤1x10^{-5}) on chromosome 33

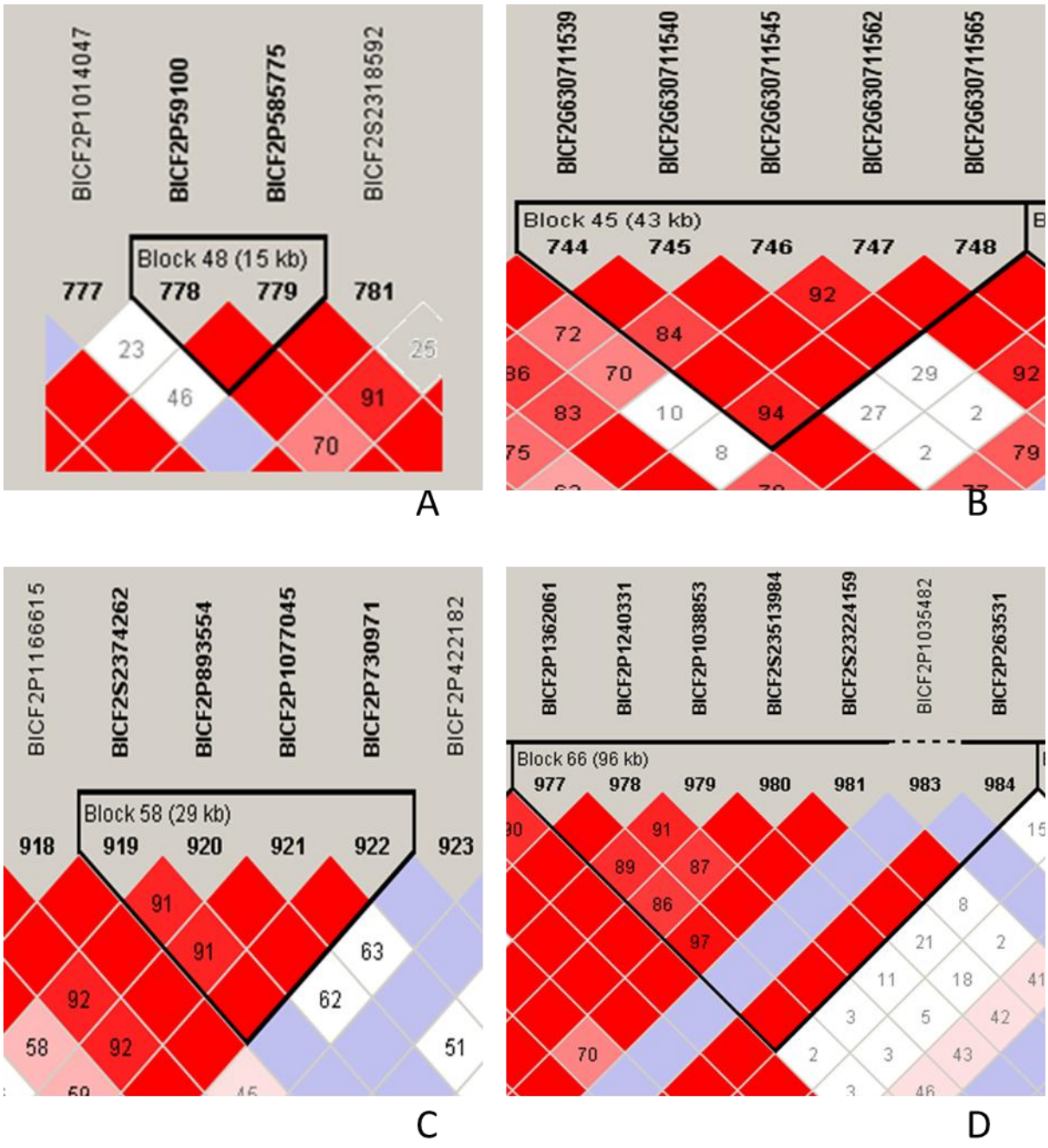


Figure 3.10. LD plots showing haplotype blocks reaching statistical significance (raw *p*= ≤1x10^{-5}) on chromosome 1.

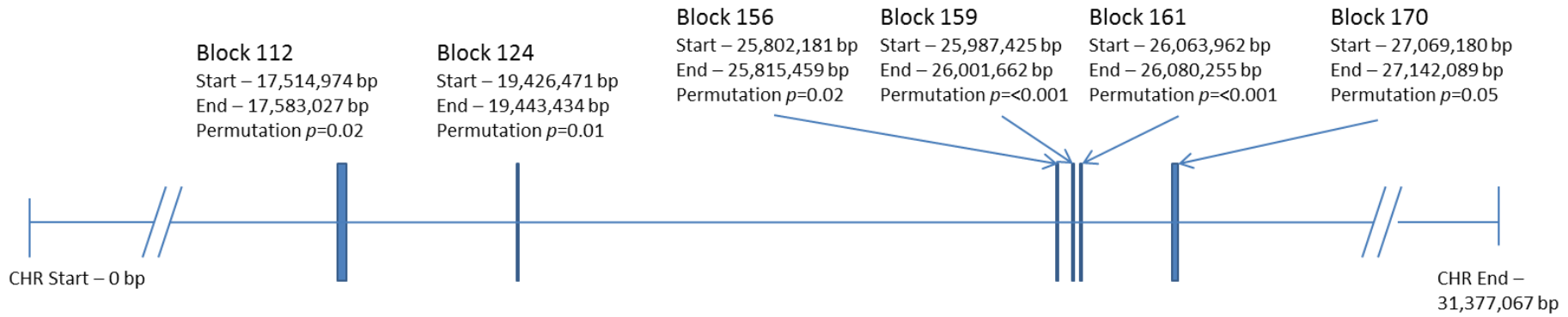


Figure 3.11. Representation of haplotype blocks on chromosome 33

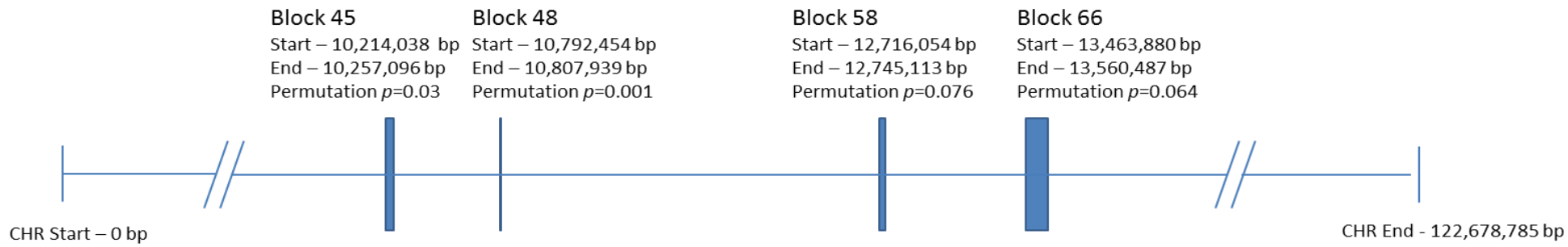


Figure 3.12. Representation of haplotype blocks on chromosome 1

Table 3.3. Haplotype blocks in chromosome 33.

Block	Haplotype	Freq.	Case, Control Ratio Counts	Case, Control Frequencies	*p*-value (raw)	*p*-value (permutation)
Block 159	GG	0.32	10.0 : 80.0, 45.0 : 37.0	0.11, 0.55	7.90E-10	<0.001
	AA	0.64	76.0 : 14.0, 34.0 : 48.0	0.84, 0.42	4.52E-09	<0.001
	AG	0.04	4.0 : 86.0, 3.0 : 79.0	0.04, 0.04	7.94E-01	1.000
Block 161	TT	0.51	62.7 : 29.3, 27.3 : 56.7	0.68, 0.33	2.42E-06	<0.001
	CC	0.23	8.9 : 83.1, 31.4 : 52.6	0.10, 0.37	1.29E-05	0.010
	CT	0.26	20.4 : 71.6, 25.3 : 58.7	0.22, 0.30	2.33E-01	1.000
Block 112	CTTCGAG	0.18	6.1 : 85.9, 25.0 : 59.0	0.07, 0.30	5.96E-05	0.020
	ATCCTGG	0.28	30.9 : 61.1, 19.0 : 65.0	0.34, 0.23	1.07E-01	1.000
	CCTCGAG	0.06	7.4 : 84.6, 3.0 : 81.0	0.08, 0.04	2.10E-01	1.000
	CCTTGAT	0.42	40.0 : 52.0, 33.0 : 51.0	0.44, 0.39	5.73E-01	1.000
	CCTCGAT	0.05	5.5 : 86.5, 4.0 : 80.0	0.06, 0.05	7.21E-01	1.000
Block 124	GAC	0.57	66.8 : 25.2, 33.7 : 50.3	0.73, 0.40	1.42E-05	0.010
	AGC	0.17	10.4 : 81.6, 20.0 : 64.0	0.11, 0.24	2.74E-02	1.000
	AGT	0.17	6.0 : 86.0, 24.0 : 60.0	0.07, 0.29	1.00E-04	0.030
	AAC	0.08	7.8 : 84.2, 6.1 : 77.9	0.08, 0.07	7.79E-01	1.000
Block 156	TC	0.43	26.0 : 66.0, 47.4 : 32.6	0.28, 0.59	4.13E-05	0.020
	AT	0.50	56.9 : 35.1, 29.0 : 51.0	0.62, 0.36	8.00E-04	0.250
	AC	0.07	9.1 : 82.9, 3.6 : 76.4	0.10, 0.05	1.74E-01	1.000
Block 170	ATCAC	0.25	12.0 : 80.0, 30.9 : 51.1	0.13, 0.38	2.00E-04	0.050
	GCCAC	0.08	11.0 : 81.0, 2.1 : 79.9	0.12, 0.03	1.97E-02	0.990
	GCTGT	0.37	40.5 : 51.5, 23.2 : 58.8	0.44, 0.28	3.10E-02	1.000
	GTCAC	0.03	3.0 : 89.0, 2.0 : 80.0	0.03, 0.02	7.52E-01	1.000
	GCTGC	0.25	22.4 : 69.6, 21.7 : 60.3	0.24, 0.26	7.53E-01	1.000
	ATTGC	0.01	1.0 : 91.0, 1.1 : 80.9	0.01, 0.01	8.73E-01	1.000

Those in red are statistically significant haplotypes (raw *p*= ≤1x10^{-5}). The one that is highlighted yellow is of greatest importance/statistical significance.

KEY – freq. = frequency of haplotype in population. Case, control ratio counts = number of cases with and without and then the number of controls with and without the haplotype. Case, control frequency = frequency of haplotype in cases and controls. Permutation *p*-value = 1000 permutations *p*-value

Table 3.4. Haplotype blocks in chromosome 1.

Block	Haplotype	Freq.	Case, Control Ratio Counts	Case, Control Frequencies	*p*-value (raw)	*p*-value (permutation)
Block 48	AA	0.45	55.7 : 34.3, 22.7 : 61.3	0.62, 0.27	3.87E-06	0.001
	CC	0.39	25.7 : 64.3, 42.7 : 41.3	0.29, 0.51	2.60E-03	0.835
	CA	0.16	8.5 : 81.5, 18.5 : 65.5	0.10, 0.22	2.21E-02	1.000
Block 45	CGTCA	0.43	52.0 : 38.0, 22.0 : 60.0	0.58, 0.27	4.25E-05	0.031
	CACTG	0.19	11.4 : 78.6, 20.8 : 61.2	0.13, 0.25	3.34E-02	1.000
	CACCA	0.12	7.0 : 83.0, 13.9 : 68.1	0.08, 0.17	6.63E-02	1.000
	AACTG	0.24	17.6 : 72.4, 24.2 : 57.8	0.20, 0.30	1.27E-01	1.000
	AATTA	0.01	1.0 : 89.0, 1.1 : 80.9	0.01, 0.01	9.09E-01	1.000
Block 58	GTTC	0.42	50.6 : 39.4, 21.9 : 60.1	0.56, 0.27	9.36E-05	0.076
	GGCC	0.33	17.4 : 72.6, 39.1 : 42.9	0.19, 0.48	7.70E-05	0.064
	TGCT	0.24	21.6 : 68.4, 19.7 : 62.3	0.24, 0.24	9.96E-01	1.000
Block 66	TGGTAT	0.40	49.9 : 42.1, 21.0 : 63.0	0.54, 0.25	7.67E-05	0.064
	AACGTT	0.15	8.3 : 83.7, 17.8 : 66.2	0.09, 0.21	2.27E-02	1.000
	AACGTC	0.37	29.7 : 62.3, 36.0 : 48.0	0.32, 0.43	1.48E-01	1.000
	AAGTTT	0.03	1.0 : 91.0, 3.3 : 80.7	0.01, 0.04	2.25E-01	1.000
	TGCGTT	0.03	2.0 : 90.0, 3.0 : 81.0	0.02, 0.04	5.81E-01	1.000

Those in red are statistically significant haplotypes (raw *p*= ≤1x10^{-5}). The one that is highlighted yellow is of greatest importance/statistical significance.

KEY – freq. = frequency of haplotype in population. Case, control ratio counts = number of cases with and without and then the number of controls with and without the haplotype. Case, control frequency = frequency of haplotype in cases and controls. Permutation *p*-value = 1000 permutations *p*-value

The most significant haplotype was a 14KB region (block 159) on chromosome 33 which contained two SNPs (BICF2S23133650 and BICF2P1311062), that were the two highest ranking hits (had the smallest raw *p*-values) in the CMH stratified dataset. The second most important haplotype block was a 15KB region on chromosome 1 (block 48), which contained the third most significant SNP in the CMH stratified dataset (BICF2P59100 raw *p*-value = 8.90E-07).

Chromosome 33 contained five haplotype blocks that reached statistical significance after 1000 permutation (Figure 3.9 A-E, blocks 159, 161, 112, 124 and 156) and one block (block 170) that neared significance after multiple correction testing (permutation *p*-value = 0.053) - Figure 3.9 F. On chromosome 33 (see Figure 3.9), the largest haplotype block (Block 170, 72KB) contained five SNPs (BICF2G63077271, BICF2G63077266, BICF2S23214572, BICF2P778764 and BICF2S23431355) all of which showed high pairwise LD in addition to extended LD across all of the SNPs in the block. Block 112 (68Kb) contained the largest number of SNPs (7) within a block (BICF2P1228306, BICF2P328516, BICF2P508171, BICF2P147184, BICF2P754969, BICF2S23215488 and BICF2P533530).

Chromosome 1 had two associated haplotype blocks (blocks 48 and 45) that were significant after permutation testing (Figure 3.10 A & B) and two blocks (blocks 58 and 66) that neared multiple correction significance Figure 3.10 C & D (permutation *p*-values for each = 0.06). On chromosome 1 (see Figure 3.10), the largest haplotype block was Block 66 (96KB); the SNPs in LD in this block were

BICF2P1362061, BICF2P1240331, BICF2P1038853, BICF2S23513984, BICF2S23224159 and BICF2P263531.

Low density Illumina Array – Labrador Retrievers and Golden Retrievers

Genotype data from a previous unrelated study (performed on an Illumina low density SNP20 canine chip containing 22,362 SNPs) was made available for analysis purposes within this project. The data included cruciate cases and controls from two breeds – Labrador Retrievers (n=49) and Golden Retrievers (n=38). The SNPs were selected from the CanFam2 genome assembly.

These data were also assessed for stratification by the Eigenstrat PCA method to check for any underlying population stratification. Two Labrador Retriever samples were removed for being outliers on the plots leaving 85 samples for analysis. PCA analysis showed that the two breeds segregated separately on the principle component plot (Figure 3.13)

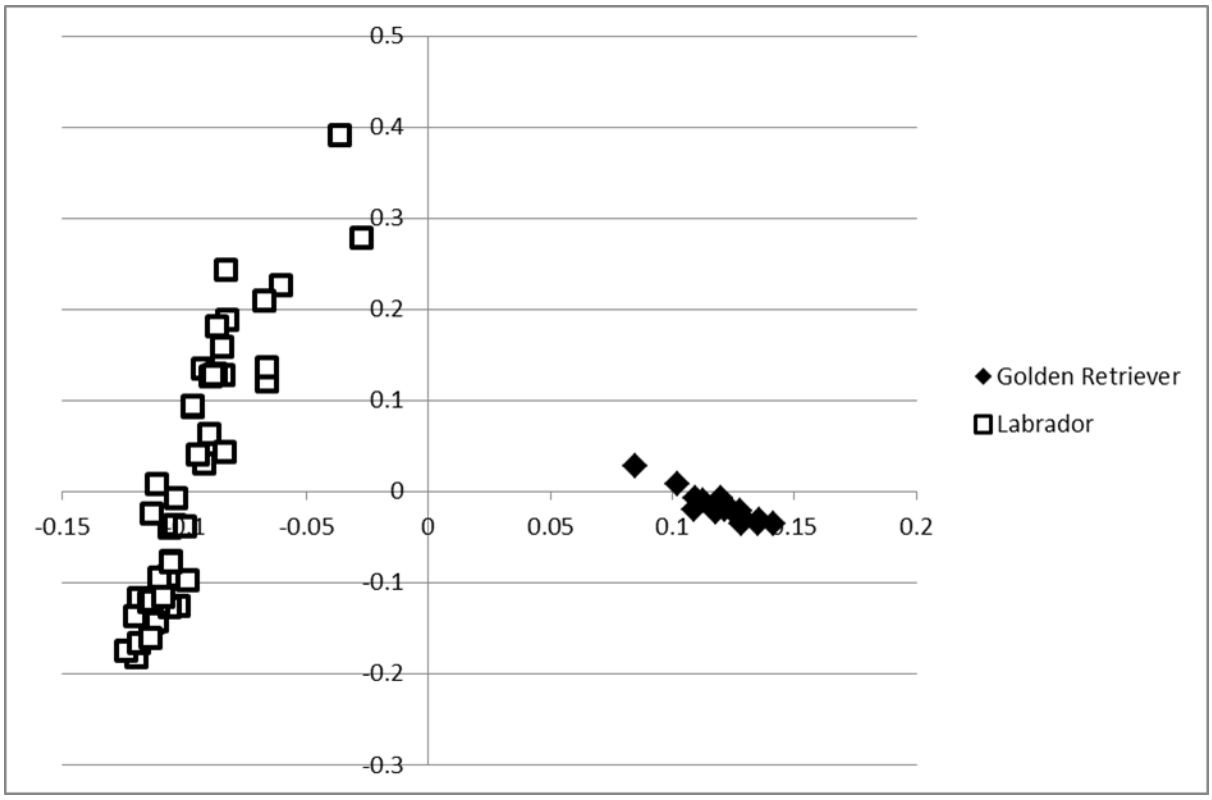


Figure 3.13. Eigenstrat population stratification plot of the low density (Illumina canine SNP20 chip) genome wide association data for Labrador Retrievers (n=47) (white squares) and Golden Retrievers (n=38) (black diamonds). Outliers have been removed from the plot.

The Manhattan plots for each individual breed are shown in Figure 3.14 and Figure 3.15 and the 40 most significant SNPs from the Golden retrievers and Labrador Retrievers are shown in Table 3.5 and Table 3.6

.

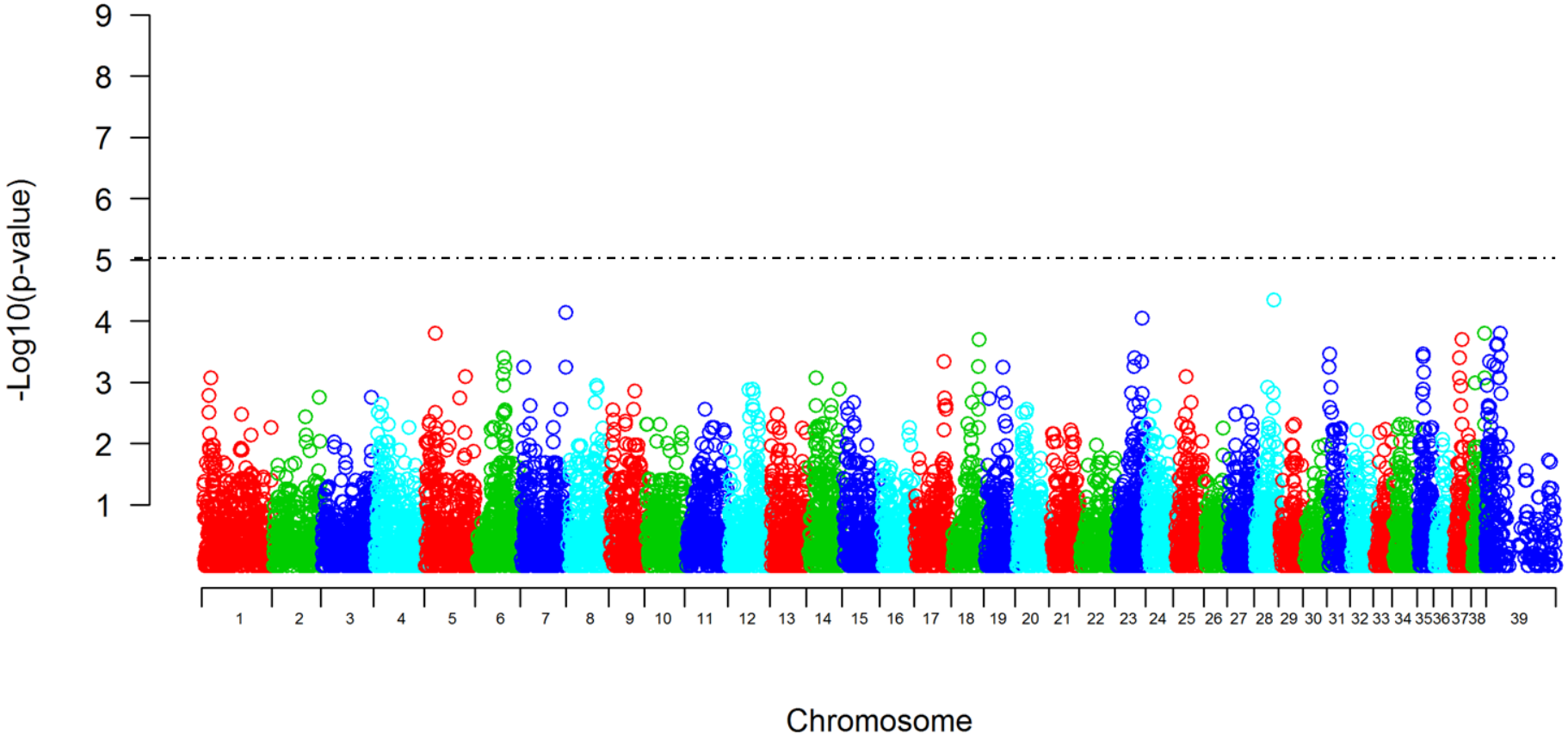


Figure 3.14. Manhattan plot showing the SNP position on each chromosome for the Golden Retrievers. The dotted line represents the *p*-value threshold of $1x10^{-5}$ to indicate genome wide significance of SNPs associated with CCL rupture.

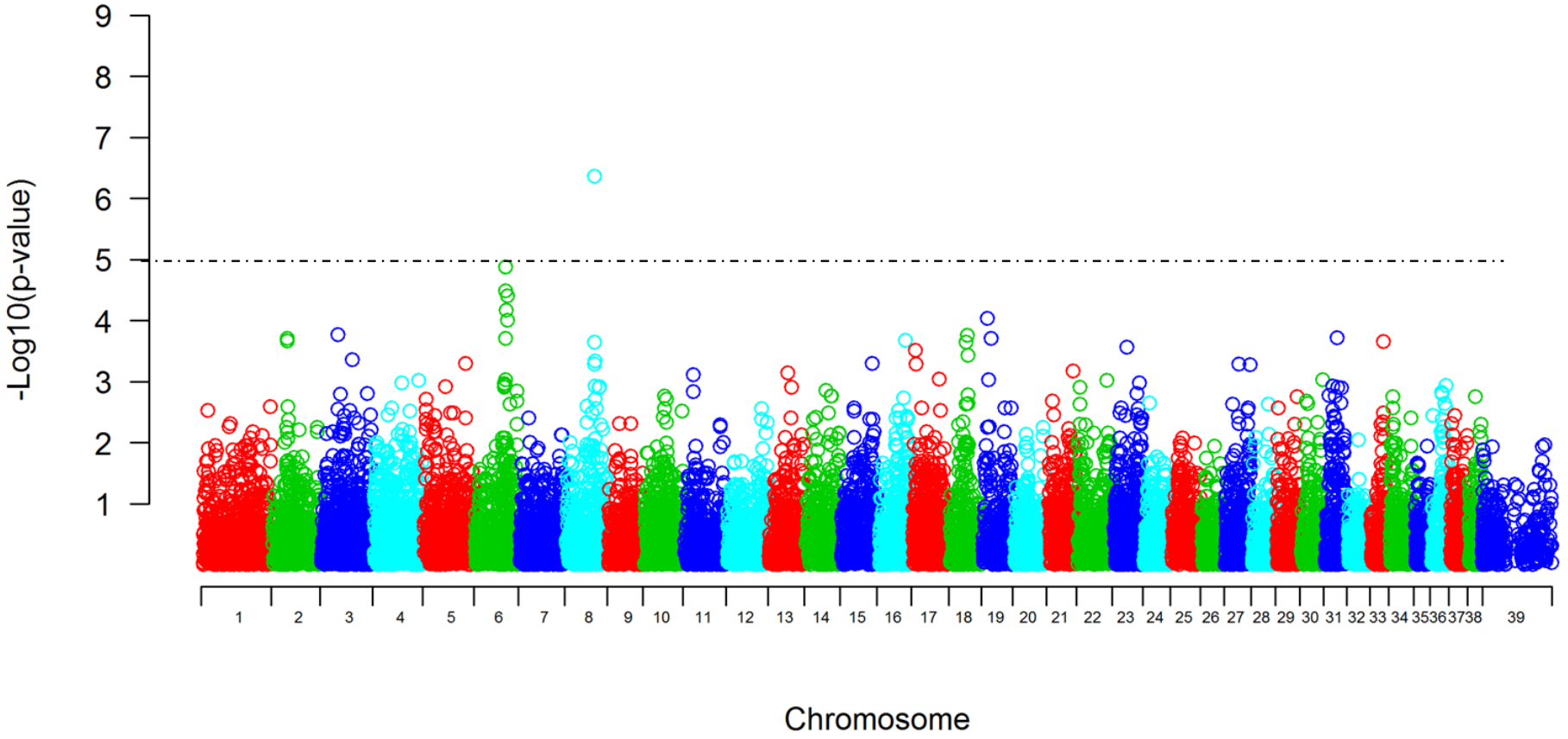


Figure 3.15. Manhattan plot showing the SNP position on each chromosome for the Labrador Retrievers. The dotted line represents the *p*-value threshold of $1x10^{-5}$ to indicate genome wide significance of SNPs associated with CCL rupture.

Table 3.5. The 40 most associated SNPs from Golden Retriever low density GWA data.

SNP	CHR	POSITION (bp)	A1	A2	p (raw)	p (BONF)	p (FDR)	p (perm)	OR
BICF2P1262009	28	36603096	A	G	0.00005	0.690	0.269	0.235	0.10
BICF2G63088517	7	83386719	G	A	0.00007	1.000	0.269	0.325	7.60
BICF2S23118727	23	50062261	A	T	0.00009	1.000	0.269	0.424	0.10
BICF2S23316643	X	26644787	T	C	0.00016	1.000	0.269	0.560	0.00
BICF2P566020	5	20811822	C	G	0.00016	1.000	0.269	0.544	13.10
BICF2G63067880	38	24803437	T	C	0.00016	1.000	0.269	0.544	13.10
BICF2P257904	18	50409186	G	A	0.00020	1.000	0.269	0.673	6.40
BICF2P678005	37	18333740	A	G	0.00020	1.000	0.269	0.673	0.20
BICF2G630535123	X	21076817	T	C	0.00024	1.000	0.269	0.719	6.10
BICF2G630535042	X	21634355	A	C	0.00024	1.000	0.269	0.727	0.20
BICF2P196724	X	18993861	T	C	0.00025	1.000	0.269	0.746	0.10
BICF2P756453	31	5469008	G	A	0.00035	1.000	0.269	0.834	11.70
BICF2G630727337	31	5487351	A	G	0.00035	1.000	0.269	0.834	11.70
BICF2P1153207	35	11394748	A	G	0.00035	1.000	0.269	0.834	11.70
BICF2G630532723	X	27757945	A	G	0.00038	1.000	0.269	0.857	0.10
BICF2G630772450	35	11171254	A	G	0.00038	1.000	0.269	0.857	19.20
BICF2S23032956	6	51890820	A	G	0.00040	1.000	0.269	0.884	0.10
BICF2P680952	23	36495440	T	C	0.00040	1.000	0.269	0.884	7.70
BICF2G630128194	37	13737113	T	C	0.00040	1.000	0.269	0.884	0.10
BICF2G630538696	X	7380657	A	T	0.00046	1.000	0.269	0.903	5.80
BICF2S23055953	17	54963129	G	A	0.00046	1.000	0.269	0.903	5.80
BICF2S23356422	23	48919581	T	C	0.00046	1.000	0.269	0.903	5.80
BICF2P1209319	X	15247146	C	T	0.00052	1.000	0.269	0.920	0.20
BICF2P778313	18	49588963	A	G	0.00055	1.000	0.269	0.930	6.60

SNP	CHR	POSITION (bp)	A1	A2	p (raw)	p (BONF)	p (FDR)	p (perm)	OR
BICF2G630373728	23	35041979	A	G	0.00055	1.000	0.269	0.930	6.60
BICF2S23219696	X	20942865	C	T	0.00056	1.000	0.269	0.939	5.40
BICF2S23060383	X	21107246	G	A	0.00056	1.000	0.269	0.939	5.40
BICF2G630575388	6	54336831	C	T	0.00056	1.000	0.269	0.939	0.20
BICF2G630549539	7	6917459	A	G	0.00057	1.000	0.269	0.948	0.10
BICF2G63052930	19	35337964	A	G	0.00057	1.000	0.269	0.948	8.50
BICF2S23227907	7	83313016	C	A	0.00058	1.000	0.269	0.954	0.20
BICF2G630726990	31	4334718	A	G	0.00058	1.000	0.269	0.954	0.20
BICF2P477037	35	12153431	T	C	0.00069	1.000	0.297	0.962	NaN
BICF2G630578492	6	50788895	A	G	0.00073	1.000	0.297	0.974	0.10
BICF2P1170982	5	75258257	A	G	0.00081	1.000	0.297	0.980	17.10
BICF2P862614	25	23258426	T	C	0.00081	1.000	0.297	0.980	0.10
BICF2S24112448	1	16271092	C	T	0.00085	1.000	0.297	0.984	6.90
BICF2P1370375	14	17971222	C	G	0.00085	1.000	0.297	0.984	6.90
BICF2G630128293	37	13794179	T	C	0.00085	1.000	0.297	0.984	0.10
BICF2G630128407	37	13923230	G	A	0.00085	1.000	0.297	0.984	0.10

KEY: OR = odds ratio. KEY: A1 = base for allele 1, A2 = base for allele 2, OR = odds ratio for A1. The data is from the CanFam2 genome annotation and the low density Illumina array

Table 3.6 The 40 most associated SNPs from Labrador Retriever low density GWA data.

SNP	CHR	POSITION (bp)	A1	A2	p (raw)	p (BONF)	p (FDR)	p (perm)	OR
BICF2S2356933	8	53525876	A	C	0.0000004	0.008*	0.008*	0.004*	10.30
BICF2P824055	6	57929723	C	T	0.0000135	0.238	0.119	0.140	7.50
BICF2G630573645	6	57338843	T	C	0.0000326	0.576	0.174	0.288	6.50
BICF2S23217009	6	60904209	A	G	0.0000393	0.694	0.174	0.364	11.30
BICF2S23213970	6	58426144	A	G	0.0000684	1.000	0.210	0.504	5.90
BICF2P115992	19	11296216	T	C	0.0000928	1.000	0.210	0.642	22.70
BICF2S2439360	19	11304497	T	G	0.0000928	1.000	0.210	0.642	22.70
BICF2P1267480	6	60820837	G	A	0.0000995	1.000	0.210	0.678	10.20
BICF2G630706178	3	33472405	A	G	0.0001700	1.000	0.210	0.820	0.10
BICF2P814871	18	34563332	A	G	0.0001740	1.000	0.210	0.833	5.30
BICF2P783963	31	24917345	A	G	0.0001930	1.000	0.210	0.842	0.20
BICF2P178000	2	30099000	T	C	0.0001950	1.000	0.210	0.852	6.50
BICF2S23651052	6	57316517	C	G	0.0001950	1.000	0.210	0.852	6.50
BICF2S23714172	19	18607492	A	G	0.0001950	1.000	0.210	0.852	6.50
BICF2G630816155	16	51480795	T	G	0.0002120	1.000	0.210	0.872	5.30
BICF2P681116	2	30117462	C	T	0.0002150	1.000	0.210	0.878	6.10
BICF2G63078520	33	24890251	A	G	0.0002240	1.000	0.210	0.892	5.40
BICF2S23737289	8	53570064	A	G	0.0002260	1.000	0.210	0.898	5.80
BICF2P778506	18	32131688	T	C	0.0002260	1.000	0.210	0.898	5.80
BICF2P721935	23	27073235	G	A	0.0002730	1.000	0.241	0.939	0.20
BICF2P1010365	17	8058457	G	A	0.0003080	1.000	0.259	0.950	0.20
BICF2S2379941	18	35534955	T	C	0.0003730	1.000	0.300	0.969	4.80
BICF2P638159	3	59502464	A	G	0.0004400	1.000	0.314	0.984	0.20
BICF2P695795	8	54831607	A	G	0.0004650	1.000	0.314	0.986	5.90

SNP	CHR	POSITION (bp)	A1	A2	p (raw)	p (BONF)	p (FDR)	p (perm)	OR
BICF2P1307704	5	77056386	T	C	0.0005100	1.000	0.314	0.991	0.20
BICF2P592926	15	58282077	C	G	0.0005120	1.000	0.314	0.992	0.20
BICF2P95470	17	8860568	T	C	0.0005130	1.000	0.314	0.992	0.10
BICF2P1240005	8	53680922	C	T	0.0005150	1.000	0.314	0.993	0.20
BICF2P759679	27	27664279	C	G	0.0005150	1.000	0.314	0.993	0.20
BICF2S23514041	27	47718008	C	G	0.0005350	1.000	0.315	0.994	5.20
BICF2P526433	21	48157647	G	C	0.0006740	1.000	0.384	0.999	0.10
BICF2P883484	13	36084406	T	C	0.0007280	1.000	0.399	0.999	0.10
BICF2P703258	11	19340622	T	C	0.0007750	1.000	0.399	0.999	6.70
BICF2S23547520	17	50927494	C	G	0.0009050	1.000	0.399	1.000	0.10
BICF2P1340532	6	58060674	T	C	0.0009400	1.000	0.399	1.000	5.90
BICF2G63037067	19	14230616	C	T	0.0009400	1.000	0.399	1.000	5.90
BICF2G630394788	30	41936591	C	A	0.0009430	1.000	0.399	1.000	0.20
BICF2S23525565	22	55319287	A	G	0.0009520	1.000	0.399	1.000	9.90
BICF2S23260589	4	84261092	G	A	0.0009560	1.000	0.399	1.000	0.20
BICF2P1266966	4	53505610	C	T	0.0010430	1.000	0.399	1.000	4.20

KEY: OR = odds ratio. KEY: A1 = base for allele 1, A2 = base for allele 2, OR = odds ratio for A1. The data is from the CanFam2 genome annotation and the low density Illumina array. Asterix (*) = significant after correction for multiple testing.

To compare the Labrador Retrievers and Golden Retrievers together as a whole cohort, a meta-analysis was carried out on the two datasets. The corresponding Manhattan plot and the 20 most associated SNPs are shown below in Figure 3.16 and Table 3.7 respectively.

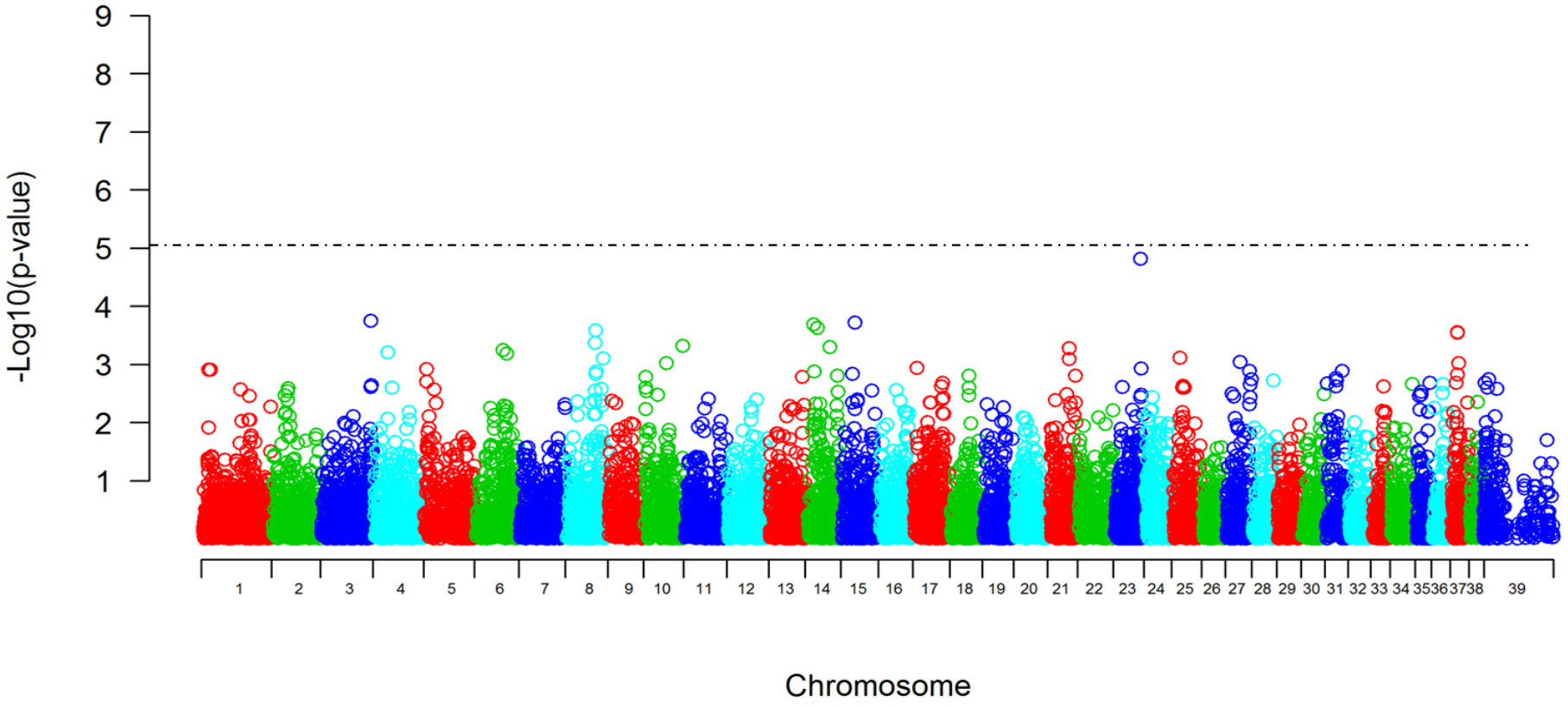


Figure 3.16. Manhattan plot of each SNP on each chromosome, for the meta-analysis of Labrador Retrievers and Golden Retrievers low density GWAS data. The dotted line represents the *p*-value threshold of $1x10^{-5}$ to indicate genome wide significance of SNPs associated with CCL rupture.

Table 3.7. The 20 most associated SNPs (raw *p*<0.05) from a meta-analysis of the Golden Retriever data and the Labrador Retriever low density GWAS data

SNP	CHR	Position (bp)	A1	A2	Raw *p*	Estimated p (BONF)	OR
BICF2S23118727	23	50062261	T	A	0.00002	0.32	4.2
BICF2P280650	3	91785665	A	G	0.00018	1.00	3.6
BICF2P990180	15	26091657	C	T	0.00019	1.00	3.4
BICF2G630522197	14	15582285	G	A	0.00021	1.00	0.2
BICF2P209484	14	22304055	G	A	0.00024	1.00	0.3
BICF2P695795	8	54831607	A	G	0.00026	1.00	3.8
BICF2G630128293	37	13794179	T	C	0.00028	1.00	0.2
BICF2G630128194	37	13737113	T	C	0.00029	1.00	0.2
BICF2P1240005	8	53680922	C	T	0.00043	1.00	0.2
BICF2G630492949	10	71446224	G	A	0.00049	1.00	0.3
BICF2G630530271	14	44484835	G	T	0.00050	1.00	3.4
BICF2P1285690	21	39448116	A	G	0.00054	1.00	0.2
BICF2S23032956	6	51890820	A	G	0.00057	1.00	0.3
BICF2P331356	4	27447828	G	A	0.00063	1.00	4.9
BICF2P1340532	6	58060674	T	C	0.00066	1.00	4.6
BICF2S23623203	25	15609495	G	A	0.00077	1.00	0.3
BICF2P562824	8	67905258	A	G	0.00079	1.00	0.3
BICF2P1210287	21	39467303	A	G	0.00081	1.00	0.2
BICF2P759679	27	27664279	C	G	0.00092	1.00	0.3
BICF2G630486552	10	41050419	G	A	0.00096	1.00	3.0

KEY: A1 = base for allele 1, A2 = base for allele 2, OR = odds ratio for A1. Based on the CanFam2 annotation of genome. p (BONF) = Correction for multiple testing. This has been estimated by correcting for 16163 tests (unique tests run in the meta-analysis).

None of the SNPs reached significance (raw *p*-value ≤ $1x10^{-5}$) in the meta-analysis of the low density Labrador Retrievers and Golden Retriever data, although the SNP showing greatest association (BICF2S23118727) on chromosome 23 generated a raw *p*-value of $1.5x10^{-5}$. These SNPs are different to those identified using the high density Newfoundland data.

In order for a comparison to be made between the Newfoundlands, Labrador Retrievers and Golden Retrievers, a meta-analysis was run include the high density data of the Newfoundlands along with the low density data of the Labrador Retrievers and Golden Retrievers. Only the SNPs common to both arrays were assessed (16163 SNPs).

The Manhattan plot of the data is shown in Figure 3.17 and the 20 most significant SNPs from the combined meta-analysis of the high density and low density data for the three breeds are shown in Table 3.8

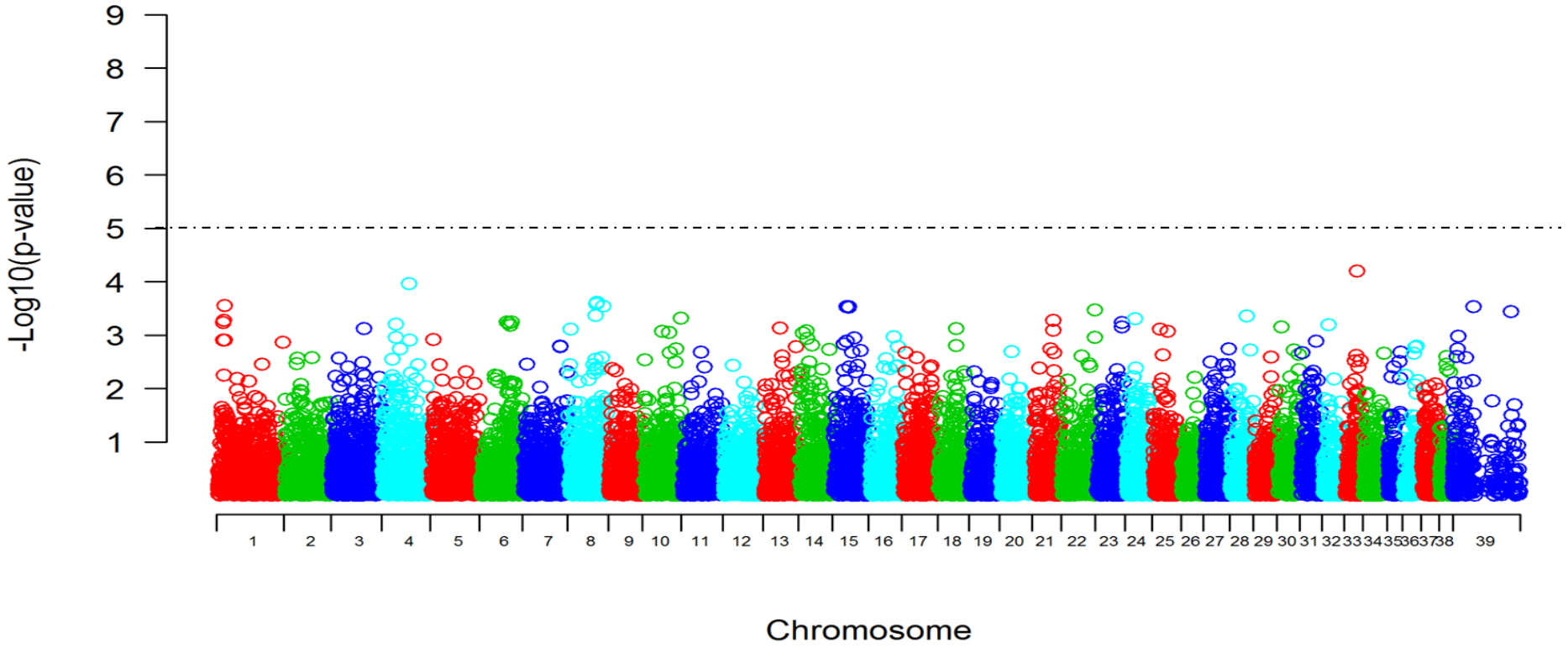


Figure 3.17. Manhattan plot of each SNP on each chromosome, for the meta-analysis of (low density array) Labrador Retrievers, Golden Retrievers and the (high density array) Newfoundlands. The dotted line represents the *p*-value threshold of $1x10^{-5}$ to indicate genome wide significance of SNPs associated with CCL rupture

Table 3.8. Twenty most significant SNPs from the meta-analysis of the low density Golden Retriever data and Labrador Retriever data and the high density Newfoundland data .

SNP	CHR	Position (bp)	A1	A2	Raw *p*	Estimated p (BONF)	OR
BICF2S22956692	33	24532210	G	T	6.33E-05	0.66	2.75
BICF2P174661	4	52672382	T	A	1.09E-04	1.00	0.29
BICF2S23756043	8	55368579	G	C	2.43E-04	1.00	2.53
BICF2P695795	8	54831607	A	G	2.64E-04	1.00	3.82
BICF2P1240331	1	16482671	G	A	2.78E-04	1.00	2.46
BICF2P562824	8	67905258	A	G	2.87E-04	1.00	0.40
BICF2P355991	39	38928542	C	T	2.96E-04	1.00	2.84
BICF2G630434646	15	30579673	T	C	2.97E-04	1.00	0.37
BICF2P990180	15	26091657	C	T	2.97E-04	1.00	2.33
BICF2G630434636	15	30593836	A	C	3.00E-04	1.00	0.37
BICF2G63093243	22	63949304	A	G	3.41E-04	1.00	0.38
BICF2P227876	39	109116340	A	C	3.68E-04	1.00	3.05
BICF2P1240005	8	53680922	C	T	4.26E-04	1.00	0.24
BICF2S23510607	28	32230015	C	G	4.36E-04	1.00	0.38
BICF2G630492949	10	71446224	G	A	4.87E-04	1.00	0.27
BICF2P1049530	24	19875480	A	C	4.90E-04	1.00	2.38
BICF2P676783	24	19944160	C	T	4.90E-04	1.00	2.38
BICF2P779428	1	14732079	T	C	5.37E-04	1.00	0.33
BICF2P1285690	21	39448116	A	G	5.37E-04	1.00	0.17
BICF2P1267480	6	60820837	G	A	5.73E-04	1.00	2.93

KEY: A1 = base for allele 1, A2 = base for allele, OR = odds ratio. Based on the CanFam2 annotation of genome P (BONF) = Correction for multiple testing. This has been estimated by correcting for 10432 tests (unique tests run in the meta-analysis).

None of the SNPs reached significance (raw *p*-value ≤ $1x10^{-5}$) in the meta-analysis of the high and low density data combined for the three individual breeds. The significant SNPs identified in the Labrador Retriever and Golden Retriever are different from the high density Newfoundland data.

Discussion

Investigations using 96 Newfoundland dogs (irrespective of disease status) identified population stratification within the dataset, due to the geographical region of origin of the dogs (Europe or North American). Population stratification is well documented as causing spurious genetic associations and is an important confounder in both human and canine genetic analyses [216, 219, 220, 227-229]. This effect has been reported previously in dogs that have originated from different geographical regions [170, 177, 230, 231] but has not previously been documented within the Newfoundland breed.

The data were assessed in the individual populations to evaluate if the results differed by country. A case control study was conducted separately on the European and the North American dogs. A meta-analysis was carried out to examine the effects of the combined European and North American case control studies and an EMMAX analysis was also conducted. This enabled a thorough comparison of the different analysis methods used for correcting population stratification. The variation in the most significant SNPs between each of the methods was minimal (Table 3.1) and the same significant associations were identified. Therefore only the stratified CMH analysis will be discussed in detail.

An analysis of the three methods used for correction for multiple testing shows that Bonferroni is the most conservative of the three methods, with FDR being the least

stringent. Permutation testing with 100,000 permutations shows intermediate stringency when used on the GWAS data.

From the stratified GWAS data in the high density array, there are three significant regions of association (seen as straight line vertical peaks, rising above the background noise on the Manhattan plot - Figure 3.5). The regions on chromosomes 1, 3 and 33 indicate a genetic association with CCL rupture and justify further investigation. The two most significant SNPs (on chromosome 33) are contained within the gene Semaphorin 5b (SEMA5B), which has functions involved in axonal regulation and neural development [232]. These two SNPs are both intronic and are present in one haplotype block – block 159 on chromosome 33. This block also reaches significance with a permutation *p*-value of ≤0.0001 for the most common allele combination. The three combinations are GG, AA or AG; the frequency of GG haplotype in the population is 32%, the AA 64% and the much rarer AG haplotype only has a 4% frequency in the population. The dogs that have the GG haplotype are at a reduced risk of CCL rupture 11% frequency in cases and 55% frequency in controls. The risk haplotype AA has a much higher frequency in the population (64%) than the protective haplotype with 84% frequency in cases and 42% frequency in the controls.

The third most associated SNP in the stratified analysis (BICF2P59100 on chromosome 1), was located in a haplotype block with one other SNP which did not reach statistical significance (ranked 839). With a frequency of 45% in the

population (62% frequency in cases and 27% frequency in controls) the AA allele combination is the only allele combination to show significance. Both SNPs in the haplotype block are intergenic and the nearest gene is CDH19 which is 0.27Mb upstream. CDH19 is a cadherin gene that is involved in cell-cell adhesion, it is not known whether the identified SNPs impact on the function of CDH19.

To date, there have been no other reported GWAS screens that have used any high density SNP chips to test the hypothesis that there is a genetic basis to CCL rupture in dogs. The only previous studies that have investigated the genetic basis of CCL rupture are microsatellite (MSAT) studies [2] and candidate gene studies [6, 233-235]. In the MSAT report the study assessed 495 MSATs in total for potential genome wide significance with CCL in the Newfoundland dog. From these markers, 86 were found to be significant with a *p*-value of ≤0.05. After correcting for multiple testing, four markers showed significance. The four markers were on different chromosomes and after validation using other markers nearby, three of the four chromosomes remained significant on chromosome 3, 5 and 13. We did not find any significant associations on either chromosome 5 or 13, but we did identify five SNPs on chromosome 3 that had a position of 68.2MB (the same position as the CPH19 MSAT) which could potentially be associated with CCL rupture although none were statistically significant (Table 3.9).

Table 3.9. SNPs on chromosome 3 at position 68.2MB in the GWAS study. Our results at this position are not significant and do not replicate the results of Wilke et al. (2009).

CHR	SNP	Position	*p*-value
3	BICF2S23344482	68205869	0.64
3	BICF2P439249	68209925	0.29
3	BICF2P7565	68219772	0.29
3	BICF2S23143373	68260652	0.16
3	BICF2S23345176	68267363	0.33

The MSATs that Wilke et al. (2009) used were spaced on average 5.5 centimorgans (cM) apart across all 38 chromosomes. The number of bases per centimorgan varies between species, but is thought to be around 1cM to 1Mb in humans [236] and in dogs is thought to be 106bp to 1cM [237]. On the Illumina CanineHD chip there are only 21 gaps that are larger than 200kB [238] and the markers on the Illumina CanineHD chip are spaced on average 13Kb apart. Therefore, as the number and size of gaps in the MSAT study is much greater than the Illumina scan, important regions and causative mutations are likely to have been missed by the MSAT study. Thus using the Illumina CanineHD arrays, the whole of the canine genome is genotyped with high coverage, depth and in more detail than any previous MSAT or low density GWA screen has been able to achieve.

Low density data

Within the meta-analysis of the low density data there were no significant associations, however one SNP - BICF2S23118727 (chromosome 23) neared

statistical significance (raw p=0.00002). This is an intergenic SNP, 0.1MB upstream of the gene MBNL2 that may have an involvement in myotonic dystrophy disease in humans [239, 240]. This disease causes muscle wasting and weakness and so functionally it is possible that this may play a role in CCL rupture susceptibility.

The lack of SNPs reaching statistical significance on the low density chip may be explained by the limited sample size (49 Labrador Retrievers and 38 Golden Retrievers) and the relatively low number of SNPs (22,362). As a consequence the gap size between markers is larger, and many SNPs will not have been tested. The causal genetic variant may have been missed as it may have fallen between the markers on the array. To determine whether there is sharing of causal mutations/regions involved in susceptibility to CCL rupture between different dog breeds it would be useful to perform a case control GWAS on multiple breeds using the high density Illumina CanineHD chip and large numbers of dogs from a number of susceptible breeds.

In isolation, the three regions of association, important SNPs and several key haplotypes found in the Newfoundland GWAS, cannot be taken as substantial evidence for genetic susceptibility to CCL rupture. Further replication and validation are needed in order for them to validated and confirmed as true associations.

Replication and validation of GWAS data

The 40 most associated SNPs found in the high density GWAS data were re-genotyped on another genotyping platform (Sequenom iPlex), with a separate cohort of Newfoundlands to determine if the genotyping results were consistent and could be replicated in a larger number of dogs.

Chapter 4.

GWAS validation study: SNP detection by Sequenom

Introduction

The GWAS SNP data generated for the Newfoundlands as described in Chapter 3 was of considerable interest. It revealed three promising chromosome association signals associated with CCL rupture on chromosomes 1, 3 and 33. These associated regions, in addition to other SNPs with the lowest *p*-values in the individual GWAS analyses (a total of 70 unique SNPs) were chosen for additional genotyping. This hoped to confirm (or not) the significance of association of the SNPs with CCL rupture. As well as validating the GWAS identified SNPs in the test population of Newfoundlands, there was also the opportunity to examine SNP associations in an additional cohort of Newfoundlands and in other breeds where CCL rupture is also considered to be a major problem. Any reproducibility would suggest a common pathogenesis and a possible shared aetiology between breeds.

Replication and validation of any biological data (including genotyping data) is essential to confirm that the associations seen are accurate and are not artefacts of bias in the original study [241]. It may also confirm the statistical associations of the data to validate the preliminary associations in a larger cohort of subjects [242].

Validation could be achieved by directly replicating the previous work by running a second GWAS on a further cohort of Newfoundlands and other breeds. The cost implication of whole genome genotyping (using hundreds of thousands of SNPs on an array) makes genotyping large numbers of dogs expensive; so genotyping a second cohort of dogs on the same platform (Illumina Canine HD chip) is

impractical. An alternative approach for validating associations could be to re-sequence any previously identified regions of association. If one does not know whether the areas being re-sequenced are real associations, re-sequencing a wrongly identified region would be a costly mistake. *In-vitro* and *in-vivo* functional analysis (including knock-out mice models) using the SNPs could also be run. These approaches are time consuming, expensive and are not worthwhile unless SNPs are confirmed as associated with the disease. Thus these would not be appropriate at this stage of the project.

Replication in this study was possible because of the availability of a second genotyping method – MALDI-TOF mass spectrometry (Sequenom iPlex) which enables high throughput of samples for specific SNP genotyping and is also relatively inexpensive to run. This technique targets specific SNPs, previously identified as potentially associated, by a cost-effective methodology which lends itself to studies of larger dog cohorts. This approach is suitable for genotyping small numbers of potentially important SNPs in a relatively large population of samples. This two stage approach (GWAS followed by Sequenom genotyping) has been widely accepted as a replication/validation method used by many researchers, including Wood et al. [177] who genotyped 25 atopic canine cases and 23 canine controls on the Illumina SNP20 GWAS chip. They validated their most significant associations in a total of 659 samples across eight different dog breeds using the Sequenom assay. Guo et al. [243] also used this approach to identify a novel susceptibility gene for osteoporosis in humans. They used the Affymetrix 500K

human SNP chip for the GWAS study and validated the most associated SNPs using Sequenom analysis. More recently, Tengvall et al. [244] also used this design to determine genetic associations with canine atopic dermatitis in German Shepherd dogs using the high density CanineHD GWAS array and then validating their 54 most significant associations in 185 German Shepherd dogs.

This chapter describes the use of a Sequenom genotyping platform to confirm (or refute) GWAS-identified SNPs from the previously genotyped Newfoundlands. It assesses the SNP associations in additional Newfoundlands and also in other breeds of dogs with high susceptibility for CCL rupture (Labrador Retriever, Rottweiler and Staffordshire Bull terrier).

Methods

Sample selection

Additional samples and breeds were included to determine whether the GWAS associations could be replicated in (a) a larger cohort of Newfoundlands and (b) different breeds of dog. Along with the 96 Newfoundlands previously analysed by GWAS, another 175 Newfoundlands were genotyped. Three other dog breeds at high risk for cruciate rupture were also included; Labrador Retrievers (n = 289), Rottweilers (n = 138) and Staffordshire Bull Terriers (n = 51). The samples were obtained from the UK companion animal DNA archive and the VLA database as described previously in the “Sample Collection for DNA extraction” section of the

Methods. A breakdown of the cases and controls is shown below in Table 4.1. A full table detailing the age and sex of the dogs is shown in the Appendix (Appendix 1)

Table 4.1. Cases and controls used for the replication/validation of GWAS study

Breed	Case	Control	Total
Newfoundlands (Total)	99	172	271
Newfoundlands (EU)	56	72	128
Newfoundlands (NA)	43	100	143
Labrador Retrievers	124	165	289
Rottweilers	57	81	138
Staffordshire Bull Terriers	13	38	51

Sequenom genotyping

Sequenom iPlex is a multiplex, high quality customised genotyping platform, capable of running up to 40 SNPs in one reaction (plex). Samples are processed on 384 well plates. Each SNP is chosen by the user and primers are designed and automatically assigned into plexes using the Sequenom assay design software (www.mysequenom.com). The DNA undergoes a multiplex, locus specific PCR reaction followed by a single base, allele specific primer extension using modified oligos. The PCR product is nanospotted onto a silica chip and placed into a Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometer; each spot is shot with a laser which ionizes the samples. The ions travel through a vacuum tube to an ion detector. Time of flight measures the difference in time different molecules hit the detector and the software calculates the mass of the fragment. Smaller molecules travel faster than larger ones; this happens in real time and the software analyses this and produces the allele calls.

Assay design

A total of 70 SNPs were chosen for replication from the 40 most associated SNPs in each of the GWAS analyses carried out (meta, EMMAX and CMH stratified), as detailed in Chapter 3.

The SNPs of interest (including 100bp flanking sequence) were retrieved from Ensembl, using the mining tool Biomart [245] and exported in FASTA format. Files were converted to a suitable format for use in the Sequenom assay design software (version 1.0) using a customised PERL script. ProxSNP and Prextend quality control procedures were run on the sequences as part of the design process. Any unsuitable primers, i.e. those that may form primer-dimers were rejected, as were sequences that could not have suitable primers designed. SNP primers that could not be plexed with others were also rejected. Five SNPs were excluded in this process leaving 65 SNPs for running and analyses.

Primers and probes were ordered from Sigma Aldrich (Dorset, UK) in the lyophilised form on 96 deep well plates. The primers were diluted to 50µM and pooled together into their respective plexes; each primer in the pool was at a working concentration of 0.5µM. The 65 SNPs were pooled into three plexes, 2x22plex and 1x21plex. The probes were diluted to 300µM.

The Sequenom laboratory work was carried out at the Wolfson Centre for Personalised Medicine in the Pharmacology Department at the University of Liverpool.

The primer and probe sequences for the 65 SNPs used for replication are shown in Appendix III.

PCR

PCR reactions were carried out in 384 well plates using a Veriti® 384-Well Thermal Cycler (Applied Biosystems, California, US). The PCR mix for each well consisted of 0.10µl Hot start Taq (5U/µl, Sequenom, Hamburg, Germany), 1µl of the primer pool mix (0.5µM), 0.10µl dNTP mix (25mM, Sequenom, Hamburg, Germany), 0.33µl $MgCl_2$ (25mM, Sequenom, Hamburg, Germany), 0.63µl PCR buffer (10x, Sequenom, Hamburg, Germany) and 2.85µl of nanopure water. The reaction mixture was added to wells containing DNA that had been dried down previously (2µl of 20ng/µl DNA). The amplification reaction was carried out using the following amplification conditions:

94˚C – 15 minutes
94˚C – 20 seconds
56˚C – 30 seconds
72˚C – 1 minute
} X45
72˚C – 3 minutes
4˚C - ∞

Unincorporated dNTPs were neutralised by the addition of shrimp alkaline phosphatase (SAP) enzyme (1.7 U/µl, Sequenom, Hamburg, Germany). This was prepared by adding 1.53µl of nanopure water, 0.17µl of hME buffer (10x, Sequenom, Hamburg, Germany), 0.30µl of SAP enzyme was added to each well making a total reaction volume of 7µl. This was then cycled at 37˚C for 40 minutes and 85˚C for 5 minutes.

iPLEX extension reaction

Due to the inverse relationship between peak intensity and mass of product [246] it was important to amend the concentration of the extension primers with a four stage adjustment to level out the analyte peaks in the final spectra. To do this, the lowest mass primers were mixed to a final concentration of 7µM, second lowest at 9.3µM, the next at 11.66µM and the highest mass group at 14µM. The iPLEX cocktail mixture was made by adding 0.76µl of nanopure water, 0.2µl iPLEX buffer (10x, Sequenom, Hamburg, Germany), 0.2µl iPLEX-termination mix (Sequenom, Hamburg, Germany), 0.04µl of iPLEX enzyme (Sequenom, Hamburg, Germany) with 0.80µl of the primer mix. This reaction mixture was added to each well of the 384 well plate to make a final reaction volume of 9µl, which was thermocycled using the following conditions:

94˚C – 30 seconds
94˚C – 5 seconds
52˚C – 5 seconds
80˚C – 5 seconds
52˚C – 5 seconds
80˚C – 5 seconds
X50
72˚C – 3 minutes
4˚C - ∞

The iPLEX reaction products were cleaned by adding 6mg of resin (Sequenom, Hamburg, Germany) to each well. Nanopure water (16µl) was added to each well and the 384 well plate was then rotated 360˚ along its long axis for 10 minutes before centrifugation for 5 minutes at 1500g.

Mass Spectrometry

The Sequenom chip was spotted with the PCR product using the nanodispenser (Sequenom, Hamburg, Germany) and placed in a MALDI-TOF mass spectrometer to acquire the spectra. A biallelic SNP can yield three different genotypes. For example, an A/G marker could produce the genotype of A homozygote, G homozygote and A/G heterozygote. Representative images of the spectra is shown in Figure 4.1

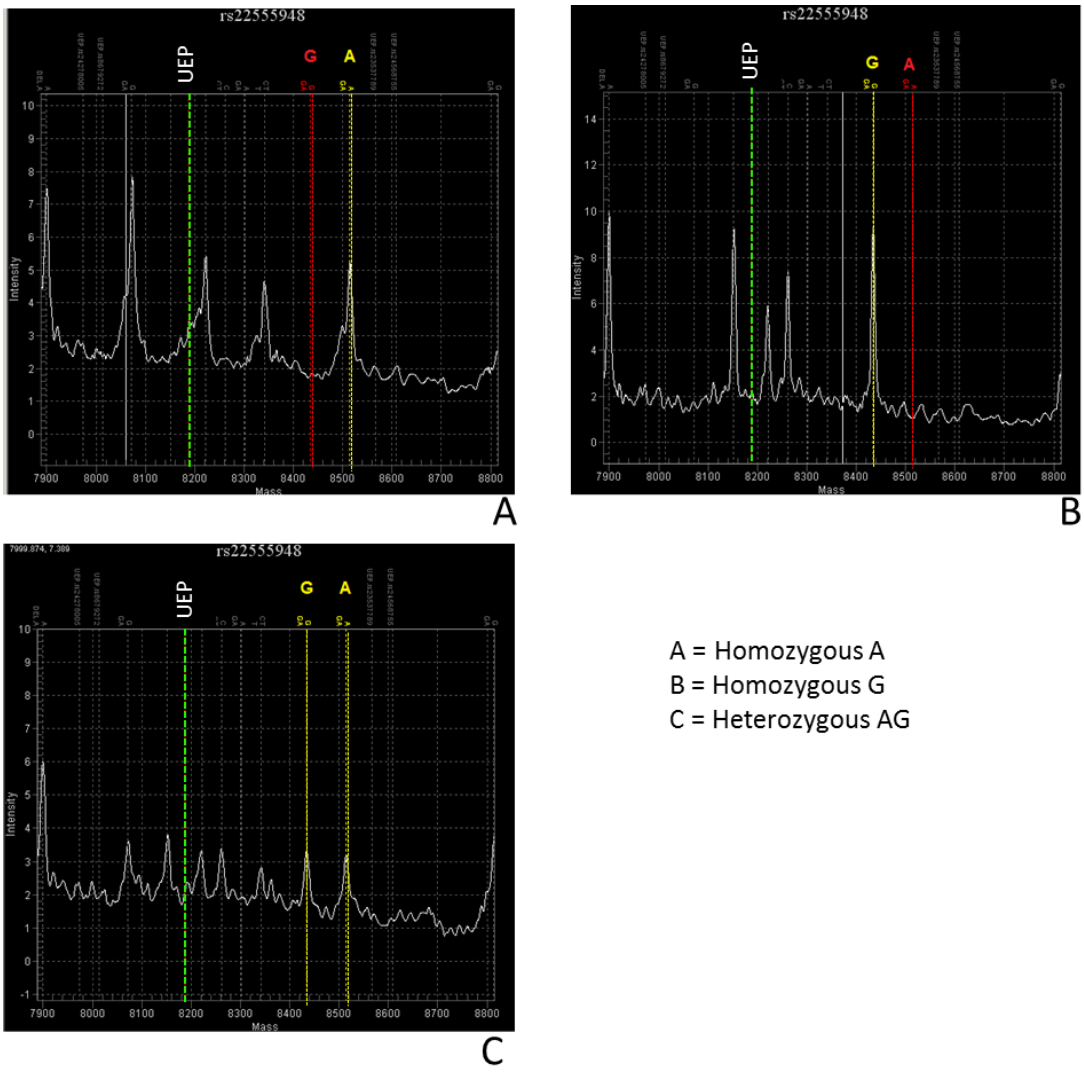


Figure 4.1. An example spectra from MALDI-TOF showing genotype calls for a bialleleic SNP A/G. The unextended primer (UEP) is shown in green, the alleles are shown in red and yellow - yellow peaks indicate the allele call for the sample. If the sample is homozygous there is one large peak as indicated by figures A & B, hetrozygote samples shows two peaks at both allele sites. The remaining peaks on the spectra are either from other SNPs in the plex or undetermined products. If the iplex extension failed then there would be a large peak at the UEP position.

The data were checked in the massARRAY TYPER program, which uses a "traffic light" system to show the quality of each reaction. The clusters and call rates for each SNP were visually assessed for correct genotype calls. The data were exported in a format suitable for input into the data analysis program PLINK [190, 191] implemented by BCgene (BC Platforms, Finland).

Quality Assurance

Before statistical analyses were performed, data were assessed for QA and excluded if the sample call rate was <75%, the genotype call rate <75%, the control population was not in Hardy-Weinberg equilibrium (p<0.001) and/or the minor allele frequency ≤0.01. Following QA, the samples and SNPs that remained for analyses differed for each breed and are shown in Table 4.2

Table 4.2. Numbers of SNPs and CCL rupture cases and controls remaining for analyses after QA

BREED	SNPs	SAMPLES		
		Case	Control	Total
Newfoundlands (Total)	58	78	136	214
Newfoundlands (EU)	59	46	56	102
Newfoundlands (NA)	59	32	80	112
Labrador Retriever	53	91	126	217
Rottweilers	53	50	64	114
Staffordshire Bull Terriers	56	10	21	31

The call rate used to filter samples during QA for the Sequenom genotyping was lower than that used for the GWAS investigation (Chapter 3) because some of the samples that were used for Sequenom were of marginally lower quality due to differences in the sample collection methods. The majority of the DNA samples that were used for the Illumina genotyping were extracted from EDTA blood samples, whereas the DNA that was used in the Sequenom genotyping was mainly extracted from Oragene™ swabs. Despite methods to clean up the DNA and remove contaminants from the salivary samples, the DNA from the Oragene™ kits was of lower quality than the DNA extracted from the EDTA blood samples. It was, however, adequate to be used for Sequenom genotyping, as documented previously [247] [181].

Results

Concordance between Illumina and Sequenom genotyping

Sixty-five SNPs in 96 samples were genotyped using both the Illumina and Sequenom platforms to check the consistency of the genotypes between the two platforms. Concordance between the two platforms was found to be 93%.

Sequenom genotyping data

The Sequenom data were examined on a breed by breed, case control basis. The Newfoundland group was analysed using the CMH test (stratified by country) in addition to separate European and North American case control studies. The

significantly associated SNPs (p<0.05) that replicated in the Newfoundland breed analyses (stratified, European, North American) are shown in Table 4.3 - Table 4.5 and the significant SNPs in the other breed analyses (Labradors Retrievers, Rottweilers and Staffordshire Bull Terriers) are shown in Table 4.6.

Table 4.3. Significantly-associated CCL rupture SNPs (p<0.05) in Newfoundlands (Stratified analysis).

Chr	SNP	Gene or intergenic	Position (bp)	A1	MAF	A2	p (raw)	p (BONF)	p (perm)	p (FDR)	OR	95% CI	Gene function
3	rs23569097	SORCS2	59381030	C	0.47	T	0.0006	0.03*	0.04*	0.02*	0.48	0.32 - 0.74	
3	rs23566023	Intergenic	58878743	A	0.43	G	0.0010	0.06	0.06	0.02*	2.00	1.32 - 3.02	-
33	rs23829198	Intergenic	19430035	G	0.39	A	0.0011	0.06	0.06	0.02*	0.48	0.31 - 0.75	-
33	rs8754330	ZDHHC23	18223573	G	0.29	A	0.0023	0.13	0.11	0.03*	0.47	0.29 - 0.77	
33	rs23835082	SEMA5B	25987425	G	0.28	A	0.0024	0.14	0.12	0.03*	0.47	0.29 - 0.77	
1	rs21977372	Intergenic	10792454	A	0.47	C	0.0039	0.23	0.18	0.04*	1.84	1.22 - 2.79	-
15	rs22346338	STAB2	42269107	T	0.18	C	0.0044	0.25	0.19	0.04*	0.42	0.23 - 0.77	
15	rs22418471	Intergenic	37221186	G	0.33	T	0.0073	0.42	0.28	0.05*	0.55	0.35 - 0.85	-
33	rs23806044	Intergenic	19502373	T	0.29	C	0.0073	0.42	0.28	0.05*	0.52	0.32 - 0.84	-
1	rs22014075	Intergenic	10756573	C	0.42	T	0.0086	0.50	0.32	0.05*	1.75	1.15 - 2.66	-
33	rs23801537	DIRC2	25923850	A	0.31	G	0.0097	0.56	0.34	0.05*	0.54	0.34 - 0.86	
33	rs23811087	Intergenic	26105553	T	0.29	C	0.0107	0.62	0.37	0.05*	0.53	0.32 - 0.86	-
33	rs23827731	SEMA5B	26001662	G	0.30	A	0.0128	0.74	0.42	0.05*	0.55	0.34 - 0.88	
1	rs9094558	RNF152	14988276	G	0.37	A	0.0132	0.77	0.43	0.05*	0.58	0.38 - 0.89	
33	rs23835584	Intergenic	26080255	C	0.21	T	0.0137	0.79	0.44	0.05*	0.51	0.30 - 0.87	-
33	rs8613571	STXBP5L	24503510	T	0.30	C	0.0139	0.81	0.44	0.05*	0.55	0.34 - 0.89	
4	rs24114482	LRRTM3	18350538	T	0.27	C	0.0152	0.88	0.47	0.05*	0.56	0.35 - 0.90	
1	rs21956708	Intergenic	10822495	G	0.41	A	0.0186	1.00	0.53	0.06	1.64	1.09 - 2.47	-
1	rs21960910	Intergenic	10408315	G	0.40	A	0.0191	1.00	0.54	0.06	0.60	0.39 - 0.92	-
33	rs23801997	Intergenic	26063962	C	0.42	T	0.0207	1.00	0.56	0.06	0.61	0.40- 0.93	-
15	rs22383474	ANO4	39884807	G	0.25	A	0.0234	1.00	0.60	0.06	0.56	0.34 - 0.93	
33	rs8748659	STXBP5L	24536212	A	0.30	T	0.0247	1.00	0.61	0.06	0.59	0.37 - 0.93	

Chr	SNP	Gene or intergenic	Position (bp)	A1	MAF	A2	p (raw)	p (BONF)	p (perm)	p (FDR)	OR	95% CI	Gene function
1	rs21987554	Intergenic	10403384	G	0.40	A	0.0251	1.00	0.62	0.06	0.61	0.39 - 0.94	-
33	rs23798518	STXBP5L	24386039	T	0.30	G	0.0269	1.00	0.64	0.06	0.59	0.37 - 0.94	
8	rs24496570	Intergenic	11069517	T	0.30	A	0.0274	1.00	0.65	0.06	0.58	0.36 - 0.95	-
27	rs23333120	CLSTN3	37900799	G	0.19	A	0.0287	1.00	0.66	0.06	0.55	0.32 - 0.95	
36	rs23935034	Intergenic	24968515	A	0.28	G	0.0289	1.00	0.67	0.06	1.66	1.05 - 2.63	-
33	rs9122813	Intergenic	24703426	G	0.33	A	0.0350	1.00	0.72	0.07	0.62	0.40 - 0.97	-
33	rs23795335	POPDC2	23291566	C	0.40	T	0.0454	1.00	0.80	0.09	0.65	0.43 - 0.99	
1	rs21998646	Intergenic	13375903	T	0.28	C	0.0457	1.00	0.80	0.09	0.61	0.38 - 0.99	-
33	rs23800347	KALRN	27503643	A	0.20	G	0.0466	1.00	0.81	0.09	0.58	0.34 - 0.99	
1	rs21882298	Intergenic	10231948	G	0.49	A	0.0499	1.00	0.83	0.09	1.50	1.00 - 2.25	-
1	rs8959368	Intergenic	11497808	G	0.04	A	0.0520	1.00	0.84	0.09	0.23	0.04 - 1.08	-
33	rs23828846	SEMA5B	25983710	C	0.39	T	0.0532	1.00	0.85	0.09	0.66	0.43 - 1.01	

KEY: A1 = base for allele 1, A2 = base for allele 2, MAF = minor allele frequency, p(raw) = raw *p*-value. Three different tests for correction for multiple testing were performed: p (BONF) = Bonferroni *p-value*, p (FDR) = *p*-value for FDR, p(perm) = *p*-value for 100,000 permutations. OR = Odds Ratio (≥1 implies A1 increases risk compared to A2, if ≤1 A2 increases risk compared to A1). An asterix (*) means significance after correction for multiple testing. The row highlighted in yellow is of most significance.

Function of the genes are colour coded:

Purple - gene function involved in neuronal systems and development and maintenance of nervous system (KALRN, LRRTM3, SEMA5B, SORCS2, STXBP5L).

Blue - synaptic junctions and synaptic signalling (ANO4, CLSTN3, ZDHHC23).

Green - polyubiquination and apoptosis or were apoptosis enhancers (RNF152, STAB2).

Orange – miscellaneous function (cancer and cellular transport; DIRC2) and (heart development; POPDC2)

Thirty-four of the 58 SNPs that replicated in the Newfoundland group as a whole (CMH stratified) had a raw *p*-value of ≤0.05. From the 34 SNPs, 18 were intergenic and the rest fall within 12 genes (KALRN, LRRTM3, SEMA5B, SORCS2, STXBP5L, ANO4, CLSTN3, ZDHHC23, RNF152, STAB2, DIRC2 and POPDC2). The twelve genes included five genes involved in neuronal systems and the development and maintenance of the nervous system. They are KALRN (Kalirin), LRRTM3 (leucine rich repeat transmembrane neuronal 3), SEMA5B (Semaphorin 5b), SORCS2 (sortillin-related vacuolar protein sorting 10-domain containing protein) and STXBP5L (Syntaxin binding protein 5 like). Also identified as CCL rupture-associated were three genes that are involved in synaptic junctions and synaptic signalling: ANO4 (Anoctamin 4), CLSTN3 (Calsytenin 3) and ZDHHC23 (Zinc finger DHHC-type containing 23). Two genes were involved in polyubiquination and apoptosis or were apoptosis enhancers RNF152 (Ring finger protein 152) and STAB2 (Stabilin 2). Two genes with miscellaneous functions were involved in cancer and cellular transport; DIRC2 (Disrupted in Renal Carcinoma 2) and heart development; POPDC2 (Popeye domain containing 2).

The most associated SNP that replicated in this breed was rs23569097 on chromosome 3, in gene SORCS2. This is highlighted in yellow in Table 4.3. It showed significance after correction for multiple testing in each of the three different tests (Bonferroni *p*-value = 0.03, permutation *p*-value = 0.04 and FDR *p*-value = 0.02).

Table 4.4. Significantly associated CCL rupture SNPs (*p*≤0.05*)* in European Newfoundlands .

Chr	SNP	Gene or intergenic	A1	F_A	F_U	A2	p (raw)	p (BONF)	p (perm)	p (FDR)	OR	95% CI	Gene function
3	rs23569097	SORCS2	T	0.57	0.32	C	0.0005	0.03*	0.03*	0.03*	2.74	1.55 - 4.86	
3	rs23566023	Intergenic	A	0.51	0.28	G	0.0012	0.07	0.07	0.04*	2.63	1.45 - 4.76	-
33	rs23829198	Intergenic	G	0.31	0.52	A	0.0036	0.21	0.18	0.07	0.42	0.23 - 0.76	-
15	rs22418471	Intergenic	G	0.27	0.46	T	0.0054	0.32	0.23	0.08	0.43	0.24 - 0.78	-
1	rs22014075	Intergenic	C	0.52	0.33	T	0.0086	0.51	0.33	0.08	2.20	1.22 - 3.96	-
8	rs8987872	Intergenic	T	0.26	0.44	C	0.0088	0.52	0.33	0.08	0.45	0.25 - 0.82	-
33	rs8754330	ZDHHC23	G	0.20	0.38	A	0.0099	0.58	0.36	0.08	0.42	0.21 - 0.82	
8	rs24496570	Intergenic	T	0.22	0.39	A	0.0150	0.89	0.47	0.09	0.43	0.22 - 0.86	-
15	rs22346338	STAB2	T	0.17	0.32	C	0.0167	0.99	0.51	0.09	0.43	0.21 - 0.87	
8	rs24526391	Intergenic	G	0.27	0.43	A	0.0169	1.00	0.51	0.09	0.48	0.27 - 0.88	-
33	rs23806044	Intergenic	T	0.20	0.35	C	0.0184	1.00	0.53	0.09	0.45	0.23 - 0.88	-
1	rs21956708	Intergenic	G	0.52	0.36	A	0.0185	1.00	0.54	0.09	1.97	1.12 - 3.47	-
4	rs24114482	LRRTM3	T	0.25	0.40	C	0.0221	1.00	0.59	0.09	0.50	0.27 - 0.91	
15	rs22380841	Intergenic	A	0.26	0.42	G	0.0231	1.00	0.60	0.09	0.50	0.27 - 0.91	-
1	rs21977372	Intergenic	A	0.51	0.35	C	0.0233	1.00	0.60	0.09	1.95	1.09 - 3.50	-
27	rs23333120	CLSTN3	G	0.13	0.25	A	0.0275	1.00	0.65	0.10	0.44	0.21 - 0.92	
1	rs21882298	Intergenic	G	0.55	0.41	A	0.0410	1.00	0.78	0.14	1.79	1.02 - 3.12	-
15	rs22383474	ANO4	G	0.20	0.33	A	0.0463	1.00	0.81	0.15	0.51	0.26 - 0.99	

For key please see next page.

KEY: A1 = base for allele 1, A2 = base for allele 2, F_A/U = frequency of allele in affected/unaffected animals, p(raw) = raw *p*-value. Three different tests for correction for multiple testing were performed: p (BONF) = Bonferroni *p-value*, p (FDR) = *p*-value for FDR, p(perm) = *p*-value for 100,000 permutations. OR = Odds Ratio (≥1 implies A1 increases risk compared to A2, if ≤1 A2 increases risk compared to A1). An asterix (*) means significance after correction for multiple testing. The row highlighted in yellow is of most significance.

Function of the genes are colour coded:

Purple - gene function involved in neuronal systems and development and maintenance of nervous system (LRRTM3, SORCS2).

Blue - synaptic junctions and synaptic signalling (ANO4, CLSTN3, ZDHHC23).

Green - polyubiquination and apoptosis or were apoptosis enhancers (STAB2).

Table 4.5. Significantly associated CCL rupture SNPs ($p \leq 0.05$) in North American Newfoundlands.

Chr	SNP	Gene or intergenic	A1	F_A	F_U	A2	p (raw)	p (BONF)	p (perm)	p (FDR)	OR	95% CI	Gene function
33	rs23827731	SEMA5B	G	0.15	0.32	A	0.0107	0.63	0.37	0.22	0.37	0.17 - 0.81	
33	rs23811087	Intergenic	T	0.17	0.34	C	0.0141	0.83	0.44	0.22	0.40	0.19 - 0.84	-
33	rs23835082	SEMA5B	G	0.16	0.33	A	0.0147	0.87	0.44	0.22	0.40	0.19 - 0.85	
33	rs23801537	DIRC2	A	0.21	0.38	G	0.0181	1.00	0.50	0.22	0.44	0.22 - 0.88	
33	rs23795335	POPDC2	C	0.26	0.42	T	0.0223	1.00	0.57	0.22	0.47	0.25 - 0.91	
33	rs23835584	Intergenic	C	0.12	0.25	T	0.0316	1.00	0.69	0.22	0.40	0.17 - 0.94	-
33	rs23798518	STXBP5L	T	0.20	0.35	G	0.0320	1.00	0.70	0.22	0.46	0.23 - 0.95	
33	rs8748659	STXBP5L	A	0.20	0.35	T	0.0320	1.00	0.70	0.22	0.46	0.23 - 0.95	
33	rs8613571	STXBP5L	T	0.20	0.35	C	0.0343	1.00	0.71	0.22	0.47	0.23 - 0.95	
33	rs23828846	SEMA5B	C	0.27	0.43	T	0.0381	1.00	0.74	0.22	0.51	0.27 - 0.97	
1	rs8959368	Intergenic	G	0.00	0.07	A	0.0418	1.00	0.78	0.22	0.00	0.00 – 0.00	-
33	rs23800347	KALRN	A	0.11	0.23	G	0.0472	1.00	0.82	0.23	0.42	0.18 - 1.01	
1	rs9094558	RNF152	G	0.25	0.39	A	0.0510	1.00	0.84	0.23	0.53	0.28 - 1.01	
1	rs21998646	Intergenic	T	0.17	0.29	C	0.0541	1.00	0.85	0.23	0.48	0.22 - 1.02	-

KEY: A1 = base for allele 1, A2 = base for allele 2, F_A/U = frequency of allele in affected/unaffected animals, p(raw) = raw *p*-value. Three different tests for correction for multiple testing were performed: p (BONF) = Bonferroni *p-value*, p (FDR) = *p*-value for FDR, p(perm) = *p*-value for 100,000 permutations. OR = Odds Ratio (≥1 implies A1 increases risk compared to A2, if ≤1 A2 increases risk compared to A1). An asterix (*) means significance after correction for multiple testing.

The row highlighted in yellow is of most significance.

Function of the genes are colour coded:

Purple - gene function involved in neuronal systems and development and maintenance of nervous system (KALRN, SEMA5B, STXBP5L).

Green - polyubiquination and apoptosis or were apoptosis enhancers (RNF152).

Orange – miscellaneous function (cancer and cellular transport; DIRC2) and (heart development; POPDC2)

In the European Newfoundlands (Table 4.4), 18 of the 59 SNPs replicated with a raw *p*-value of ≤0.05 and were associated with CCL rupture. The most associated SNP, rs23569097 on chromosome 3 and in gene SORCS2, showed significance after correcting for multiple testing in all three tests ($p = 0.03$ in each).

In the North American Newfoundlands Table 4.5), 14 of the 59 SNPs replicated with a raw *p*-value of ≤0.05 and were significantly associated with CCL rupture. After correcting for multiple testing, no SNPs were significant.

Table 4.6. Significant CCL rupture associated SNPs for the Labrador Retrievers, Rottweilers and Staffordshire Bull Terriers

Breed	Chr	SNP	Gene or intergenic	A1	F_A	F_U	A2	p (raw)	p (BONF)	p (perm)	p (FDR)	OR	95% CI	Gene function
Labrador Retrievers	8	rs24526391 ▪	Intergenic	G	0.29	0.15	A	0.0003	0.02*	0.02*	0.02*	2.37	1.47 - 3.80	-
	1	rs21895425	SERPINB13	C	0.38	0.29	G	0.0347	1.00	0.85	0.51	1.55	1.03 - 2.33	
	8	rs8987872 •	Intergenic	T	0.37	0.28	C	0.0393	1.00	0.88	0.51	1.53	1.02 - 2.31	-
	3	rs23569097	SORCS2	T	0.24	0.16	C	0.0406	1.00	0.89	0.51	1.64	1.02 - 2.64	
Rottweilers	8	rs24526391 ▪	Intergenic	A	0.33	0.51	G	0.0074	0.39	0.34	0.25	0.48	0.28 - 0.82	-
	22	rs23048917	Intergenic	T	0.53	0.36	C	0.0095	0.50	0.39	0.25	2.04	1.19 - 3.51	-
	4	rs24114482	LRRTM3	T	0.51	0.35	C	0.0167	0.88	0.57	0.29	1.92	1.12 - 3.29	
	8	rs8987872 •	Intergenic	C	0.29	0.44	T	0.0250	1.00	0.76	0.32	0.53	0.30 - 0.93	-
	33	rs23828846	SEMA5B	C	0.28	0.42	T	0.0305	1.00	0.82	0.32	0.54	0.31 - 0.95	
Staffordshire Bull Terriers	12	rs22184669	CYP39A1	T	0.25	0.00	G	0.0007	0.04*	0.05*	0.04*	0.00	0.00 - 0.00	
	15	rs22380841	Intergenic	A	0.40	0.13	G	0.0148	0.83	0.55	0.41	4.67	1.28 - 17.05	-
	1	rs21998646	Intergenic	T	0.20	0.05	C	0.0276	1.00	0.78	0.49	0.25	0.07 - 0.90	-
	10	rs22030434	CCDC85A	G	0.10	0.00	A	0.0419	1.00	0.93	0.49	0.00	0.00 - 0.00	
	1	rs9094558	RNF152	A	0.17	0.44	G	0.0439	1.00	0.94	0.49	0.25	0.06 - 1.02	

KEY: A1 = base for allele 1, A2 = base for allele 2, F_A/U = frequency of allele in affected/unaffected animals, p(raw) = raw *p*-value. Three different tests for correction for multiple testing were performed: p (BONF) = Bonferroni *p-value*, p (FDR) = *p*-value for FDR, p(perm) = *p*-value for 100,000 permutations. OR = Odds Ratio (≥1 implies A1 increases risk compared to A2, if ≤1 A2 increases risk compared to A1). An asterix (*) means significance after correction for multiple testing, "•"/"▪"= significant more than one breed. The row highlighted in yellow is of most significance.

CONTINUED ON NEXT PAGE…

Function of the genes are colour coded:

Purple - gene function involved in neuronal systems and development and maintenance of nervous system (LRRTM3, SEMA5B, SORCS2).

Green - polyubiquination and apoptosis or were apoptosis enhancers (RNF152).

Orange – miscellaneous functions - differentiation of keratinocytes; SERPINB13, bile-acid metabolism; CYP39A1 and no known function; CCDC85a

In the Labrador Retriever, four of the 53 SNPs that passed QA, replicated with a raw *p*-value ≤0.05 and were significantly associated with CCL rupture. The most associated SNP in this breed, rs24526391 on chromosome 8, showed significance after correcting for multiple testing in each of the three tests (*p*=0.02 in each of the three tests).

In the Rottweiler breed, five of the 53 SNPs that passed QA reached significance with a raw *p*-value of ≤0.05. After correcting for multiple testing, no SNPs were significant.

In the Staffordshire Bull Terrier breed, five of the 56 SNPs that passed QA were significantly associated with CCL rupture (raw *p*-value of ≤0.05), the most associated SNP (rs22184669) in this breed was on chromosome 12, in gene CYP39A1. This SNP was significant after correcting for multiple testing using all three methods (Bonferroni *p*-value = 0.04, permutation *p*-value = 0.05 and FDR *p*-value = 0.04).

SNPs that are common in more than one breed are shown in Table 4.7 for intragenic SNPs and Table 4.8 for intergenic SNPs.

Table 4.7. Intragenic SNPs shared by more than one breed

Chr	SNP	Gene	Newfoundland (stratified)	EU	NA	Labrador Retriever	Rottweilers	Staffordshire Bull Terriers
3	rs23569097	SORCS2	+	+	-	+	-	-
33	rs23801537	DIRC2	+	-	+	-	-	-
33	rs23835082	SEMA5B	+	+	+	-	+	-
15	rs22346338	STAB2	+	+	-	+	+	-
33	rs8754330	ZDHHC23	+	+	+	+	-	-
15	rs22383474	ANO4	+	+	-	-	-	-
27	rs23333120	CLSTN3	+	+	-	-	+	-
33	rs23800347	KALRN	+	-	+	-	+	-
4	rs24114482	LRRTM3	+	+	-	-	+	+
33	rs23795335	POPDC2	+	-	+	-	-	+
1	rs9094558	RNF152	+	-	+	-	-	+
33	rs23827731	SEMA5B	+	-	+	-	+	-
33	rs8613571	STXBP5L	+	-	+	+	-	-
33	rs8748659	STXBP5L	+	-	+	+	-	-
33	rs23798518	STXBP5L	+	-	+	+	-	-
10	rs22030434	CCDC85A	-	-	-	-	-	+
12	rs22184669	CYP39A1	-	-	+	-	-	+
33	rs23799015	IGSF11	-	-	+	+	-	+
33	rs23828846	SEMA5B	-	-	+	+	+	-
1	rs21895425	SERPINB13	-	-	-	+	+	-
33	rs23804256	TAGLN3	-	-	-	+	-	+

KEY: raw *p*-value is indicated by colour:

Yellow =p≤0.001, blue = p≤0.01, green = p≤0.05, red = non-significant. A minus (-) means that the SNP is not present in the analysis after QA.

EU = European Newfoundlands, NA = North American Newfoundlands.

Table 4.8. Intergenic SNPs shared by more than one breed

Chr	SNP	SNP position	Newfoundland (stratified)	EU	NA	Labrador Retriever	Rottweilers	Staffordshire Bull Terriers
3	rs23566023	Intergenic	+	+	-	+	+	-
15	rs22418471	Intergenic	+	+	-	+	-	+
1	rs21977372	Intergenic	+	+	+	-	+	-
33	rs23806044	Intergenic	+	+	-	-	-	+
33	rs23829198	Intergenic	+	+	-	-	-	-
1	rs22014075	Intergenic	+	+	-	-	-	-
36	rs23935034	Intergenic	+	-	+	+	-	+
33	rs9122813	Intergenic	+	-	-	+	-	+
1	rs21956708	Intergenic	+	+	-	-	+	-
1	rs21882298	Intergenic	+	+	-	-	+	-
1	rs21998646	Intergenic	+	-	+	-	-	+
33	rs23811087	Intergenic	+	-	+	-	-	-
33	rs23835584	Intergenic	+	-	+	-	-	-
1	rs21960910	Intergenic	+	-	-	-	-	-
33	rs23801997	Intergenic	+	-	-	-	-	-
1	rs21987554	Intergenic	+	-	-	-	-	-
8	rs24496570	Intergenic	+	+	-	-	-	-
8	rs24526391	Intergenic	-	+	-	+	+	+
13	rs8712319	Intergenic	-	-	-	+	+	+
8	rs8987872	Intergenic	-	+	-	+	+	-
22	rs23048918	Intergenic	-	-	-	+	+	-
33	rs23813763	Intergenic	-	-	+	+	-	+
15	rs22380841	Intergenic	-	+	-	-	-	+

KEY: raw *p*-value is indicated by colour

Yellow =p≤0.001, blue = p≤0.01, green = p≤0.05, red = non-significant. A minus (-) means that the SNP is not present in the analysis after QA.

EU = European Newfoundlands, NA = North American Newfoundlands.

SNPs shaded in grey are in each of the breed analyses (Labrador Retrievers, Rottweilers and Staffordshire Bull Terriers) apart from the Newfoundlands

A meta-analysis was run on the data to determine the SNPs that were significant in the whole cohort. The 20 most significant SNPs from the meta-analysis are shown in Table 4.9

Table 4.9. Meta-analysis of the 20 most CCL rupture-associated SNPs. The meta-analysis was run using association data from the Stratified Newfoundlands, EU Newfoundlands, NA Newfoundlands, Labrador Retrievers, Rottweilers and Staffordshire Bull Terriers.

Chr	SNP	Position	Gene or intergenic	A1	A2	p (raw)	Estimated p (BONF)	OR	Sig.	Gene function
3	rs23569097	59381030	SORCS2	T	C	0.0003	0.02*	1.73	***	
33	rs8748659	24536212	STXBP5L	A	T	0.0026	0.15	0.64	**	
33	rs23828846	25983710	SEMA5B	T	C	0.0037	0.21	1.46	**	
33	rs23827731	26001662	SEMA5B	G	A	0.0043	0.24	0.59	**	
33	rs23798518	24386039	STXBP5L	T	G	0.0043	0.24	0.65	**	
33	rs8754330	18223573	ZDHHC23	G	A	0.0071	0.40	0.69	**	
33	rs8613571	24503510	STXBP5L	T	C	0.0119	0.67	0.68	*	
33	rs23835082	25987425	SEMA5B	G	A	0.0152	0.85	0.69	*	
33	rs23811087	26105553	Intergenic	T	C	0.0171	0.96	0.58	*	-
33	rs23806044	19502373	Intergenic	T	C	0.0337	1.00	0.74	*	-
33	rs23800347	27503643	KALRN	G	A	0.0366	1.00	1.39	*	
1	rs21956708	10822495	Intergenic	G	A	0.0376	1.00	1.53	*	-
1	rs21998646	13375903	Intergenic	C	T	0.0382	1.00	1.34	*	-
1	rs22014075	10756573	Intergenic	C	T	0.0422	1.00	1.40	*	-
33	rs23829198	19430035	Intergenic	G	A	0.0525	1.00	0.76	*	-
1	rs21977372	10792454	Intergenic	A	C	0.0526	1.00	1.33	*	-
8	rs24526391	11289491	Intergenic	G	A	0.0545	1.00	1.30	*	-
33	rs23795335	23291566	POPDC2	C	T	0.0586	1.00	0.76	ns	
33	rs23801537	25923850	DIRC2	A	G	0.0588	1.00	0.70	ns	

KEY: A1 = base for allele 1, A2 = base for allele 2, OR = odds ratio (≥1 implies A1 increases risk compared to A2, if ≤1 A2 increases risk compared to A1). An estimated correction for multiple testing was performed, correcting for 56 tests (average number of SNPs in each of the breed analyses). * = Significant after correction for multiple testing. Significance of raw *p*-value indicated by asterix (*) scale= p≤0.05, **=p≤0.01, ***=p≤0.001, ns = not significant. Function of the genes are colour coded:

Purple - gene function involved in neuronal systems and development and maintenance of nervous system (KALRN, SEMA5B, SORCS2, STXBP5L).

Blue - synaptic junctions and synaptic signalling (ZDHHC23).

Orange – miscellaneous function (cancer and cellular transport; DIRC2) and (heart development; POPDC2)

Genes SORCS2, STXBP5L, SEMA5B, ZDHHC23 and KALRN showed significant CCL rupture-associations (*p*-value ≤0.05) with nine SNPs being found within these genes. The remaining eight of the 17 significant CCL rupture-associated SNPs in the meta-analysis were all intergenic SNPs. SNP rs23795335 within the gene POPDC2 and SNP rs23801537 within DIRC2, neared statistical significance with a raw *p*-value of 0.058 for each of the two SNPs. The SNP with the lowest *p*-value rs23569097 were significant after the applying the estimated correction for multiple testing.

Discussion

Many of the associated SNPs in the Newfoundland CMH analysis (18 out of 34) were intergenic. The functions of intergenic SNPs are largely unknown. They may alter the splicing pattern, transcription factor binding sites or affect non-coding RNA in some other way which has yet to be identified [163, 248]. The remaining SNPs were all within genes (except for rs9094558) and are all intronic. They probably do not have a direct effect on protein structure, synthesis or function, as they are spliced out when mature mRNA is formed [249]. They may be in LD with associated SNPs that are in exonic/protein coding regions that have not been screened in the study. SNP rs9094558 is downstream of the gene RNF152 and so its effect on protein function is also unknown; it may affect mRNA stability.

Having confirmed the association of some SNPs in the Newfoundland breed, it was important to determine if the same associations were also present in other breeds.

Three other breeds with a high incidence of CCL rupture were available for comparison – Labrador Retrievers, Rottweilers and Staffordshire Bull Terriers. Only 14 SNPs from the Newfoundland analysis were validated in the other breeds after QA (Table 4.6), seven are intergenic and seven were within genes. All SNPs were intronic except for rs9094558 (as mentioned above), rs21895425 which is in the 3'UTR of SERPINB13 and rs21946171 which is upstream of SERPINB13. Table 4.7 and Table 4.8 show the SNPs that are common to more than one breed analysis. None of the SNPs were common to all breeds studied. This could be indicative of different genetic components being involved in CCL rupture susceptibility in different dog breeds, however, it could be that the SNPs have different levels of penetrance in the different breeds [177, 250]

Certain genes identified in the analyses are of particular interest because of their potential relevance to CCL rupture susceptibility. SORCS2 had the highest associated SNP in the meta-analysis (after correction for multiple testing) and was identified in three of the six individual analyses (Stratified Newfoundlands, European Newfoundlands and Labrador Retrievers). It is strongly expressed in the nervous system [251-254]. SEMA5B (contains the two most significant GWAS SNPs in the Newfoundlands) is involved in axonal regulation and sends positive guidance cues to the axons during development [232]. This, coupled with the functions of several other genes of interest, involving synaptic membranes (ZDHHC23) [255], calcium-mediated postsynaptic signalling (CLSTN3 and ANO4) [256-258] and neuronal plasticity (KALRN) [259] indicate a possible role of neurological and/or

neuromuscular pathways in the susceptibility and incidence of CCL rupture. These could also impact on the severity of the disease.

The neuromuscular system can control joint stability [38, 39]. The wiring of the nervous system is dependent on correct positioning of axons and dendrites, which is carried out by axonal guidance molecules - of which the Semaphorins (including SEMA5B) are a key group [232]. Axonal guidance involves activation of signal transduction cascades that can lead to re-organisation of the cytoskeleton [260]. The cytoskeleton plays an important role in distributing mechanotransduction signals to the extracellular matrix. Shultz et al. [261] were the first to document that cruciate ligaments contain mechanoreceptors which impact on knee proprioception.

Rupture of the cruciate ligament causes mechanoreceptor impairment, which in turn leads to loss of proprioception (which can persist even after surgery to stabilise the knee). Joint instability, which is a defining feature of CCL rupture, may not only be caused by loss of restraint in the traditional sense (CCL being the primary stabiliser of the knee joint), but also due to the loss of proprioception and kinaesthesia [262, 263]. Loss of proprioception in human knees after ACL injury has been well documented [264-266] and has been linked with joint damage and knee function [263, 264, 267, 268]. Interestingly, in human research, it has been demonstrated that it is possible to significantly lower the incidence of ACL rupture during a sports season [269-271]. This was achieved by developing a training

programme for the neuromuscular system (proprioceptive training), focusing on factors such as proper landing techniques using wobble boards and balance mats. Clearly such training programs are not suitable for dogs, but it does add corroboration to the theory that neuromuscular responses, neurological system and proprioception are important factors when considering and preventing rupture of the CCL in dogs. Muscle imbalance, atrophy and changes in gait characteristics have also been documented as important predisposing factors for CCL rupture [272-275]. Mostafa et al. [275] have suggested that a preventative rehabilitation program may be implemented to help reduce the incidence of CCL rupture by strengthening the key muscles in the rear limbs of the dog, this too substantiates the neuromuscular argument.

The SORCS2 gene has been shown to have a direct link to the neuropeptide activity pathway, as well as activating Substance P receptors (Neurokinin-1 receptors) [251, 276]– see Figure 4.2. It is also associated within the neutrophin signalling pathway. This is important as neutrophins promote the survival of neurons and axonal growth – see Figure 4.3. Neuropeptides have been associated with joint conditions such as osteo- and rheumatoid arthritis [40, 277, 278]. High levels of neuropeptides (Substance P and Calcitonin Gene related peptide (CGRP)) have been found to be increased in inflamed joints and also in joints that have been injected with inflammatory mediators to mimic inflammatory disorders [279-283]. Substance P can induce synoviocytes (either directly or indirectly) to increase expression of noxious substances such as collagenase and free oxygen radicals that can damage

the synovium and cartilage [284]. Substance P also causes the release and activation of immune system inflammatory mediators such as histamine and cytokines [277, 281] and as such, the neural system appears to integrate with the immune system. Neuronal circuitry has also been implicated in immunological homeostasis [285] and may exacerbate inflammatory responses to CCL rupture.

Bilviciute et al. [282] showed a bilateral increase in neuropeptides in the synovial fluid of rats that had one knee injected with pro-inflammatory substances and the other knee injected with saline for control purposes. Higher concentrations of neuropeptides in both stifle joints would mean that inflammatory changes and associated joint pain [279] in the contralateral knee would be instigated sooner, last longer and the rate of destruction sped up compared to normal. Brydges et al. [286] detailed an increase in pain and other sensory sensitivity in dogs with CCL rupture compared to dogs with no CCL rupture. They, and others [287] also detected alterations in the gait and weight distribution towards the contralateral side of the body and a shift in distribution of weight towards the fore limbs. This, one would presume, is a compensatory mechanism to help the dog cope with the pain and joint instability in the affected joint. It may also explain why there is a high incidence of dogs damaging the contralateral ligament at some stage after the initial injury [33, 107]. This is consistent with the reports that the neuromuscular system is important in CCL rupture susceptibility [275].

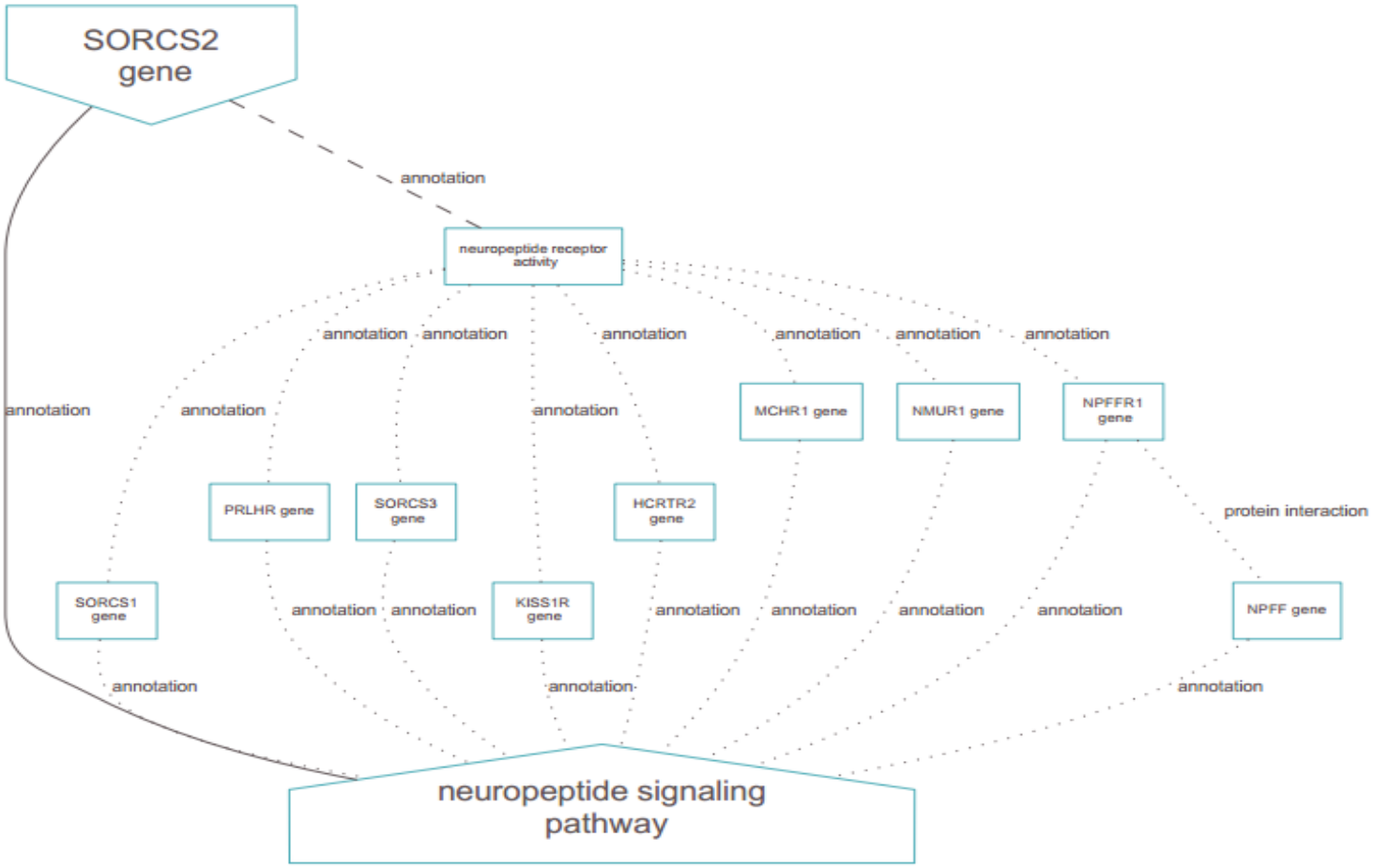


Figure 4.2. Putative pathway linking SORCS2, Substance P and other neuropeptide receptors. Reproduced with permission from http://biograph.be/concept/graph/C1539804/C1155443 [288].

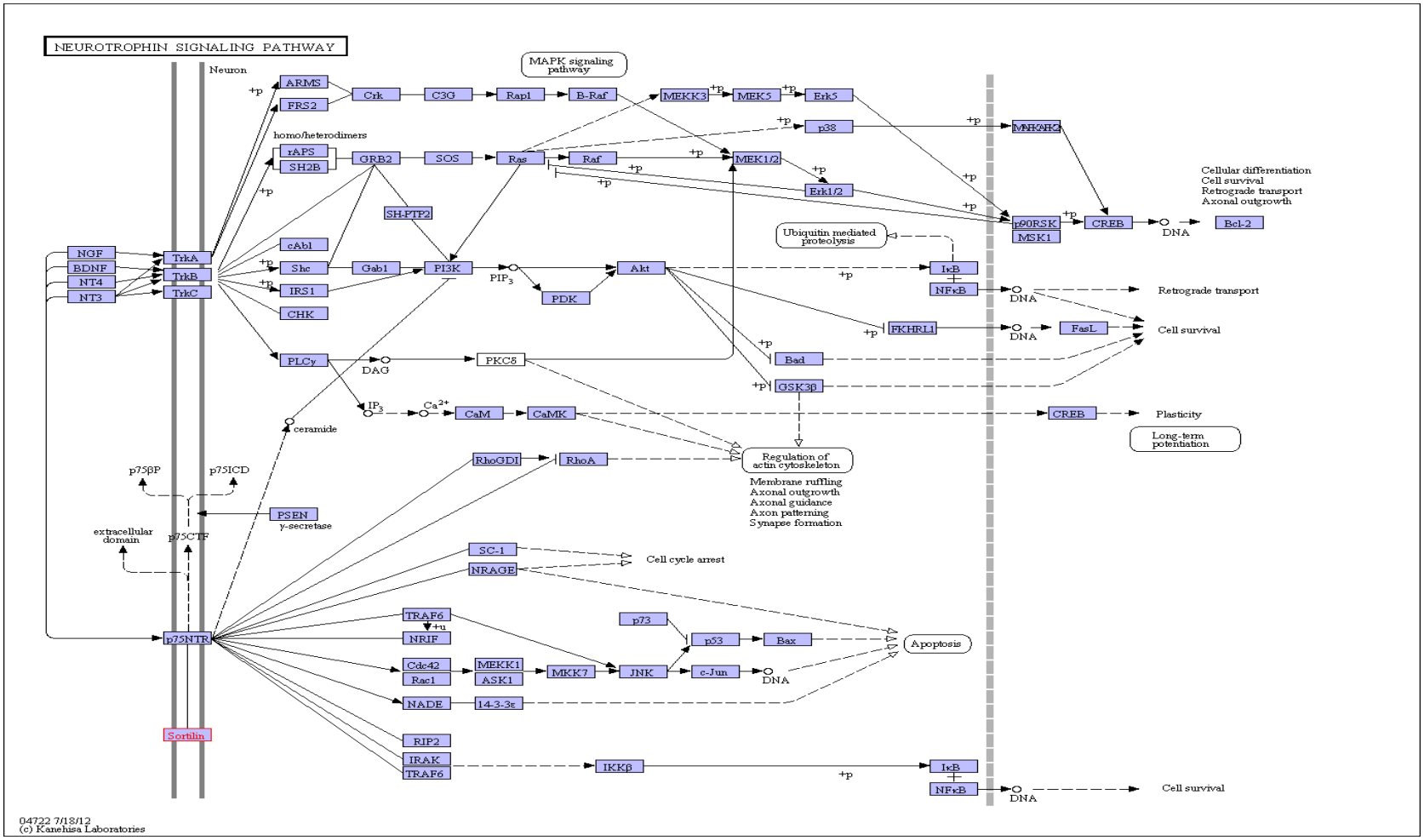


Figure 4.3. Sortilin and the neutrophin pathway. Pathway reproduced with permission from KEGG pathways (http://www.genome.jp/kegg-bin/show_pathway?map=hsa04722&show_description=show) PATHWAY: cfa04722. [289, 290]

Taken together all of the above evidence suggests that mechanotransduction, neurological and neuromuscular pathways could significantly alter the risk of CCL rupture and prolong the injury effect and associated pain. This has not previously been reported for canine CCL rupture investigations, although several human studies substantiate our findings.

The low number of statistically significant SNPs found in the Labrador Retrievers, Rottweilers and Staffordshire Bull Terriers are most likely due to the relatively small sample size which resulted in low power. Associations that may be present in these breed could have been missed as a result (false negatives). We performed the stratified CMH Newfoundland analysis as it combines the numbers of Newfoundlands from both the European and North American analyses into one study increasing sample size and power. By further combining the results of each individual breed analysis into one meta-analysis, the low power issue is reduced again. The number of cases and controls are considered as one large group and the results should identify SNPs common to CCL rupture in all of the breeds studied. Given the fact that this study has yielded significant associations it shows that the study worked with the numbers of dogs used. Taking into account that the SORCS2 SNP (rs23569097) was the most associated SNP in the stratified, European and the meta-analyses, as well as being significant in the Labrador Retriever breed analysis, it is highly plausible that this gene may have a role in CCL rupture susceptibility or resistance.

Additional validation of the associations found in this study should be conducted using larger cohorts (to increase statistical power) and also in more breeds (to determine if multiple breeds share associated SNPs). A more refined, next generation, targeted re-sequencing approach could also be carried out on the significant regions to identify additional SNPs and determine precisely which SNPs/regions of interest are involved in CCL rupture susceptibility. To explore the genetic association to canine CCL rupture in more detail, a whole genome re-sequence could also be carried out. A cheaper method that just sequences the coding regions of the genome (exons) - exome sequencing is also a valid method; transcriptome sequencing or RNA-sequencing (RNAseq) may identify new SNPs that alter the RNA or cause gene expression changes. SNPs in the coding regions of the genome are more likely to have a profound effect on a disease due to a direct effect on protein function/folding/structure.

In conclusion, analysis of Sequenom genotyping confirmed genetic associations with CCL rupture on chromosomes 1, 3 and 33 that the GWAS study reported in Chapter 3. It indicated a potential role for neurological and neuromuscular pathways in the pathogenesis and susceptibility to CCL rupture, with genes such as SORCS2 and SEMA5B indicated as being highly associated with CCL rupture in multiple breeds. The data also indicated that there are likely to be breed-specific variations in the genetic factors involved in CCL rupture, as demonstrated by the association of different SNPs in different breeds of dog. However, there are also some similarities in the genes and SNPs among the different breeds indicating that

there is some degree of shared pathogenicity of CCL rupture between different dog breeds which may become more evident in subsequent validation studies.

Chapter 5. Candidate gene association with CCL rupture

Introduction

In the previous chapters, GWAS (and Sequenom validation) identified novel chromosomal regions and SNPs of interest that may be associated with CCL rupture, using no prior knowledge of the disease or disease process to find them. In this chapter, genes were selected from the literature that may have a potential functional association with the disease initiation or progression. SNPs within these genes were tested for association with CCL rupture, for this reason this method is called the "Candidate Gene" approach. This is a purely hypothesis driven approach and relies on a prior knowledge and understanding of the disease, disease components and pathways to select genes that may be of potential interest in CCL rupture. This is a limiting factor and many disease associations within genes may be missed for not being present in obviously associated or functional genes. Before the advent of genome wide arrays, candidate gene analyses were the standard method for determining association of genes and SNPs for a variety of different diseases in humans and animals [235, 291-293].

Although the GWAS whole SNP arrays have recently superseded traditional candidate gene studies, candidate gene analyses still remain a valid approach to investigate the genetic basis of diseases. Whole genome SNP chip arrays have gaps in the SNP distribution and some regions/genes do not necessarily have high coverage. By manually selecting SNPs within certain "key" genes of interest which may or may not be covered by the whole genome arrays, you improve the

opportunity of detecting an association in the specific gene. It is possible to then to confirm or eliminate it as a candidate in CCL rupture susceptibility or resistance.

Candidate gene studies are based on selecting genes and SNPs that may be implicated in the disease process (or SNPs that are in linkage with the causative markers). As CCL rupture is a complex disorder, there are likely to be many genes implicated in initiation and progression of the disease. It is unlikely that all of the key genes have been characterised and selected for analyses. As each SNP may only have a small contribution to the disease risk, it is likely to be the additive effect of multiple variations from multiple genes that will impact on the overall risk for the condition. Thus, identifying individual variants that contribute more or less to the risk of CCL rupture will be difficult, especially with small sample sizes. Against that argument, the high LD present in individual dog breeds means smaller sample sizes are needed to identify significant SNP associations than if the study was being conducted in humans. Hence, any differences in allele frequency between canine cases and controls may indicate that the SNP is involved in the disease (or in linkage with another SNP that is functional). In this way, SNP associations may narrow down the gene/genomic area for further detailed investigation. This could then be further investigated by fine mapping, resequencing, as well as by using *in-vivo* and *in-vitro* functional association tests.

In previous canine candidate gene studies, using a cohort of Newfoundlands, the researchers tested FBN1, COMP and COL9A1 for CCL rupture associations; they

found no significant associations [233]. Other work carried out within the Boxer breed also discounted Collagen type-IX (COL9A1, COL9A2, COL9A3) [235] and COL2A1 as potential candidates for association with CCL rupture [294]. Clements et al. [292] performed a candidate gene study for 20 genes that had been all found to be previously associated with osteo-arthritis (OA) in dogs. They found no significant associations between their 20 genes and CCL rupture and also no common genetic factors between the diseases either within the same breed or between the two breeds studied.

Published candidate gene studies in human research have proven to be more productive. They have suggested that polymorphisms within COL5A1 are relevant to the risk of ACL rupture and in tendinopathy of the Achilles [234, 295]. COL1A1 has also been linked with susceptibility to ACL rupture [296, 297]. In the latter work Khoschnau et al. performed a study evaluating association of the SP1 transcription factor binding site with ACL rupture and shoulder dislocations. The SP1 binding site is located on intron 1 of the COL1A1 gene and causes a G-T mutation [297]. The mutation causes an increase in binding of SP1 to the gene and an increase in COL1A1 transcription which leads to a greater tensile strength of the ligament. However, the imbalance between the COL1A1 and COL1A2 chains (normally in ratio of two units COL1A1: one unit COL1A2) can cause osteoporosis and other bone density problems.

This chapter will assess whether select candidate genes are associated with CCL rupture in a population of Newfoundland dogs and in three other breeds that are at increased risk for CCL rupture (Labrador Retrievers, Rottweilers and Staffordshire Bull Terriers).

Methods

Sample selection

In addition to the Newfoundland dogs studied previously in this project, a further three breeds (Labrador Retrievers, Rottweilers and Staffordshire Bull Terriers) were studied; these were chosen because of the increased risk status of their breed for cruciate rupture [93]. They were included to determine whether the same candidate genes were associated with CCL rupture in the different breeds which would indicate a common pathogenesis to CCL rupture across susceptible breeds.

The same 749 samples that were used in Chapter 4 (see Table 4.1, page 116) were also used in this study.

Candidate gene and SNP selection

Genes were selected for investigation into CCL rupture based on a potential involvement in disease initiation or progression, or functional association with ligament structure, function and tensile strength. Pubmed (http://www.ncbi.nlm.nih.gov), Scopus (http://www.scopus.com) and other

scientific library databases were used to compile the list of candidate genes. Search terms included cruciate ligaments, ligament structure, cruciate ligament injury and connective tissues. SNPs were selected from Ensembl and were chosen to be spread as evenly as possible throughout the genes of interest. Mutations that were non-synonymous or frameshift were given priority inclusion over intronic SNPs (the frequency of these types of mutations was lower than exonic synonymous mutations). An even coverage was selected so as to maximise likely associations within the genes. Given the high LD present in individual dog breeds (spanning MB regions in some cases) the likelihood of chosen SNPs being causative or being in linkage with a causative SNP was high. The genes chosen, reason for selection and the number of SNPs typed in each gene are shown in Table 5.1

Table 5.1. Selected candidate genes and reason for inclusion in the study. The number of SNPs chosen in each gene is also shown.

Gene name	Gene Symbol	Number SNPs chosen	Reason for choice	Literature references
Fibromodulin	FMOD	1	Part of ECM. Affects rate of fibril formation and fibrillogenesis	[298, 299]
Decorin	DCN	9	Affects rate of fibril formation	[300, 301]
Elastin	ELN	10	Involved in elastic fibre formation which is key to CCL structure and stability	[302-304]
Opticin	OPTC	13	Part of ECM. Important in collagen fibril morphology, spacing and organisation	[305, 306]
Latent-transforming growth factor binding protein B2	LTBP2	9	Important in elastic-fibre architectural organisation and assembly within the ECM. Structural component of microfibrils	[307, 308]
Biglycan	BGN	1	Important in collagen fibre assembly and organisation of ECM	[67, 72]
Fibrillin 1	FBN-1	8	Structural component of ECM. Important for calcium-binding in microfibrils which provide the force bearing support to ligaments	[309-311]
Cartilage oligomeric matrix protein	COMP	1	Part of ECM. Role in structural integrity of cartilage via interaction with collagens and fibronectin. May have a role in pathogenesis of OA	[233, 312, 313]
Serpin peptidase inhibitor, clade H member 1	Serpin-H1	7	Involved in biosynthetic pathway of collagen formation and helps bind collagens	[149, 314]
Procollagen-lysine, 2-oxoglutarate 5-dioxygenase 1	PLOD-1	7	Attachment site for carbohydrates in collagens essential for cross link stability in collagen fibres	[315, 316]
Lysyl oxidase	LOX	6	Important part of collagen formation and strength. Promotes crosslinking of fibres to fibrils	[317, 318]
Collagen type I	COL1A1, COL1 A2	3 9	Main component of ligament	[319, 320]
Collagen type III	COL3A1	7	Component of ligament	[320]

Gene name	Gene Symbol	Number SNPs chosen	Reason for choice	Literature references
Collagen type V	COL5A1 COL5A2 COL5A3	12 7 13	Component of ligament. Important in fibril assembly and formation	[234, 321-324]
Collagen type VI	COL6A1 COL6A3	1 9	Component of ligament. Part of ECM	[325]
Collagen type XI	COL11A1 COL11A2	11 7	Important in fibrillogenesis by controlling lateral growth of collagen fibrils. Component of ligament	[326, 327]
Collagen type XII	COL12Al	11	Interacts with COL type I. Component of ligament	[328, 329]
Collagen type XIV (also named Undulin)	COL14A1	7	Regulates fibrillogenesis by integrating collagen bundles Component of ligament	[330]
Collagen type XXIV	COL24A1	9	Regulation of fibrillogenesis. Mineralisation of bone	[331]
Matrix metalloproteinase - 1	MMP-1	6	Cleaves collagen	[332, 333]
Cathepsin K	CTSK	5	Important in collagen fragmentation. Involved in osteoclastic bone resorption and bone remodelling, by causing ECM degradation	[334, 335]
Aggrecan	ACAN	6	Major component in ECM. Has a role in inflammatory joint diseases and OA	[66, 336, 337]
SIX homeobox 1	SIX-1	1	Involved in ligament, tendon and limb development	[338, 339]

The candidate gene SNPs for this study were initially chosen from the CanFam2 alignment of the canine genome. However, as a new build of the genome was released in May 2012 (CanFam3) the data were subsequently realigned to the new assembly. The majority of the SNPs retained their location within the same gene as the CanFam2 annotation, but one gene Elastin, was realigned from chromosome 9 to chromosome 6. As a result of this realignment, the seven SNPs initially chosen because they located within the Elastin gene were no longer within that gene. Four of the SNPs (rs24535705, rs24568755, rs24535784 and rs24535768) are now intergenic and three of the SNPs (rs24553222, rs24566251 and rs24566255) located to other genes. These are a novel protein coding gene (rs24553222) and U6 spliceosomal RNA (rs24566251 and rs24566255). One other SNP, rs22372241 had also been re-aligned to a novel protein coding gene instead of the COL1A1 gene.

Once genes had been selected SNPs present within the gene or in regulatory regions (5' or 3' UTRs) were identified using the Ensembl database (http://www.ensembl.org). The SNPs of interest (and 100bp flanking sequence) were mined from Ensembl using Biomart and exported in FASTA format. They were converted to a format suitable to run through the Sequenom assay design software using a customised PERL script.

In total, 196 SNPs from 28 genes were chosen for investigation in the candidate gene study.

The genotyping of the SNPs was performed using Sequenom iPlex MassARRAY® technology at the Wolfson Centre for Personalised Medicine in the Pharmacology Department at the University of Liverpool as in Chapter 4.

Assay design

The 196 SNPs chosen for the candidate gene study were pooled into seven plexes (reactions) by the assay design software – 5x29 plex, 1x28 plex and 1x13 plex; nine SNPs were excluded as suitable primers could not be designed (rs8817977, rs8907731, rs22913607, rs22913610, rs9113092, rs24584869, rs23655986, rs24044145 and rs8619084) and one SNP (rs8742180) formed a single plex that was not economically viable to be run and was excluded. This left 186 SNPs to be investigated. The primer and probe sequences for each SNP are shown in Appendix V.

Quality assurance

Following genotyping, the data were assessed for QA and data were excluded if the sample call rate was <75%, the genotype call rate <75%, the control population was not in Hardy-Weinberg equilibrium ($p<0.001$) and/or the minor allele frequency ≤0.01. Following QA, the numbers of samples and SNPs that remained for analyses differed for each breed and are shown in Figure 5.1

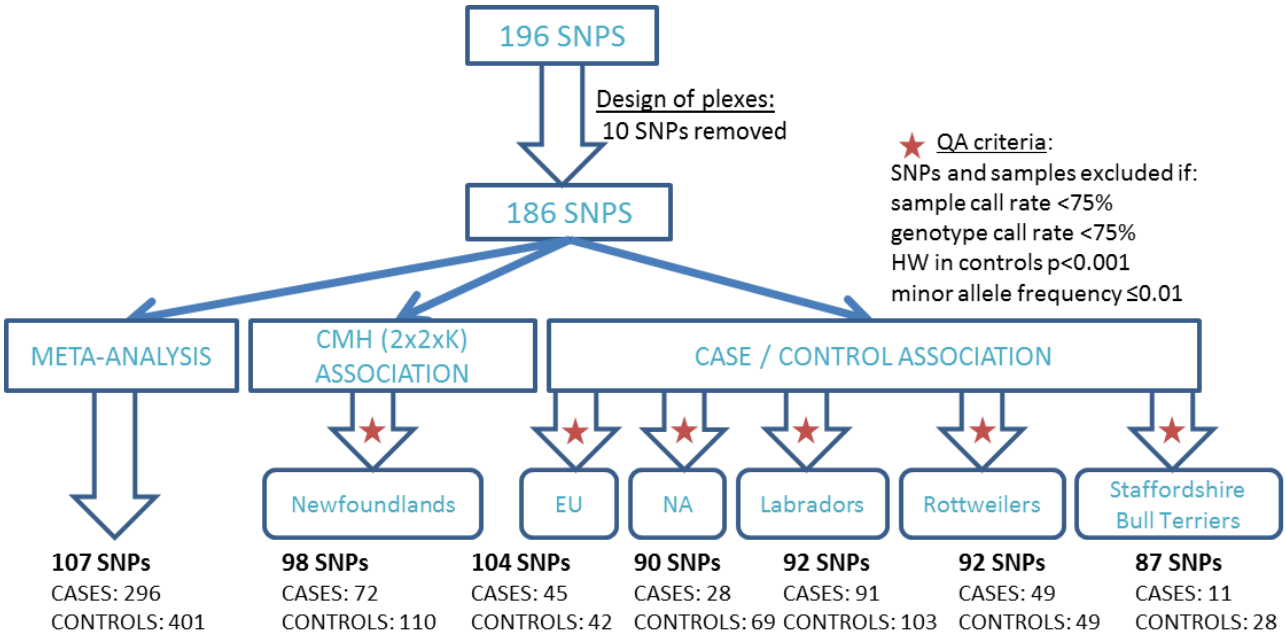


Figure 5.1. The number of SNPs and cases/controls remaining for each breed after QA in the candidate gene study.

The data were examined on a breed by breed case control basis and the Newfoundland group was analysed using a Cochran–Mantel–Haenszel (CMH) 2x2xK (stratified by country) analysis, as well separate analysis for the North American and European groups. A meta-analysis was also performed to determine associations across every breed.

Correction for multiple testing was performed using three methods: 100,000 permutations, Benjamini & Hochberg false discovery rate (FDR) and a modified "Bonferroni-style" correction where the *p*-value is divided by the number of genes in the analyses. When meta-analyses were run, multiple testing was not automatically carried out, therefore an estimated correction was applied after meta-analyses were run, using the modified "Bonferroni-style" correction.

The modified "Bonferroni-style" correction was used instead of the standard Bonferroni correction (dividing the *p*-value by number of tests i.e SNPs), as this was considered too severe for multiple correction given the LD structure of dogs. Correction for the number of genes was considered to be a suitable alternative because of the high LD with each other within each gene.

Results

The 20 most associated SNPs for the stratified CMH Newfoundland analysis are shown in Table 5.2. The SNPs that were associated with CCL rupture in the European and North American Newfoundlands, as well as in the other breeds (Labrador Retrievers, Rottweilers and Staffordshire Bull Terriers) analyses are shown in Table 5.3

Table 5.2. Twenty most CCL rupture associated SNPs in the Newfoundland CMH stratified candidate gene analysis

CHR	SNP	Gene	Consequenence	A1	MAF	A2	POSITION	p (raw)	p (BONF STYLE)	p (perm)	p (FDR)	OR	95% CI
36	rs8587007	COL 5A2	Intronic	T	0.18	A	33604230	0.006	0.15	0.45	0.54	0.40	0.21 - 0.76
9	rs24560587	COL 1A1	Intronic	C	0.31	A	29515328	0.024	0.60	0.91	0.70	0.52	0.30 - 0.92
17	rs22597612	Cathepsin K	Downstream	T	0.13	C	63008191	0.030	0.75	0.95	0.70	0.43	0.20 - 0.94
9	rs24546447	COL 5A1	Intron splice variant	T	0.13	G	54147364	0.035	0.88	0.97	0.70	2.31	1.09 - 4.91
36	rs23942103	COL 3A1	Intronic	G	0.06	C	33548721	0.043	1.00	0.98	0.70	2.87	1.02 - 8.11
6	rs8652327	COL 11A1	Synonymous	C	0.05	T	50628874	0.048	1.00	0.99	0.70	0.26	0.06 - 1.02
11	rs22190862	LOX	Intronic	A	0.32	G	15041235	0.058	1.00	1.00	0.70	0.62	0.38 - 1.02
38	rs24048794	OPTC	3' UTR	C	0.16	A	3222563	0.070	1.00	1.00	0.70	1.76	0.94 - 3.29
9	rs24553222	Novel	Intronic	T	0.03	C	24898151	0.086	1.00	1.00	0.70	0.27	0.05 - 1.32
11	rs8679272	LOX	Intronic	T	0.38	C	15038128	0.087	1.00	1.00	0.70	0.64	0.39 - 1.07
3	rs23546710	ACAN	Synonymous	A	0.26	G	54860309	0.100	1.00	1.00	0.70	0.64	0.38 - 1.10
9	rs24546410	COL 5A1	synonymous	A	0.20	G	54141470	0.101	1.00	1.00	0.70	1.60	0.92 - 2.78
9	rs8680884	COL 5A1	Intronic	G	0.35	A	54093629	0.110	1.00	1.00	0.70	1.46	0.92 - 2.31
9	rs24535705	N/A (Was Elastin)	Intergenic	A	0.03	G	24951335	0.111	1.00	1.00	0.70	0.34	0.09 - 1.34
30	rs23648407	FBN1	Intronic	T	0.20	C	17707531	0.112	1.00	1.00	0.70	1.61	0.90 - 2.89
9	rs24566251	U6	Downstream	A	0.03	C	24922431	0.114	1.00	1.00	0.70	0.29	0.06 - 1.47
9	rs24578426	COL 5A1	Missense	T	0.02	C	54169612	0.126	1.00	1.00	0.73	0.20	0.02 - 1.82
9	rs8680883	COL 5A1	Intronic	C	0.34	T	54093637	0.142	1.00	1.00	0.76	1.42	0.89 - 2.26
17	rs22595730	Cathepsin K	Intronic	G	0.02	C	63012099	0.148	1.00	1.00	0.76	0.27	0.04 - 2.02
9	rs24546719	COL 1A1	Synonymous	A	0.01	G	29522007	0.175	1.00	1.00	0.80	0.00	0.00 - 0.00

KEY: A1 = base for allele 1, A2 = base for allele 2, MAF = minor allele frequency, p(raw) = raw *p*-value. Three different tests for correction for multiple testing were performed: p (BONF STYLE) = modified "Bonferroni-style" corrected *p*-value (corrected for 25 tests, the number of genes used in this study), p (FDR) = *p*-value for FDR, p(perm) = *p*-value for 100,000 permutations. OR = odds ratio, 95% CI = 95% confidence interval of the OR.

Six of the 98 SNPs that passed QA in the Newfoundland group as a whole (CMH analysis) had a raw *p*-value of ≤0.05. Five of these SNPs were located within collagen genes (COL5A2, COL1A1, COL5A1, COL3A1 and COL11A1) and one SNP (the third most associated) was found in Cathepsin K. After correction for multiple testing using all three methods, no SNPs were significant.

Table 5.3. SNPs significantly associated with CCL rupture in Newfoundlands from Europe North America, Labrador Retrievers, Rottweilers & Staffordshire Bull Terriers in the candidate gene analyses

	CHR	SNP	Gene	Consequence of SNP	A1	F_A	F_U	A2	p (raw)	p (BONF STYLE)	p (perm)	p (FDR)	OR	95% CI
EU	9	rs24546447	COL 5A1	Intron splice site	T	0.3	0.09	G	0.0023	0.06	0.20	0.24	4.23	1.59 - 11.23
	36	rs23942103	COL 3A1	Intronic	G	0.2	0.07	C	0.0263	0.66	0.90	0.64	3.29	1.10 - 9.80
	9	rs24546410	COL 5A1	Synonymous	A	0.31	0.17	G	0.0320	0.8	0.95	0.64	2.27	1.06 - 4.86
	3	rs23546710▪	ACAN	Synonymous	A	0.17	0.3	G	0.0417	1.00	0.98	0.64	0.46	0.22 - 0.98
NA	36	rs8587007	COL5A2	Intronic	T	0.06	0.25	A	0.00273	0.06	0.30	0.24	0.18	0.05 - 0.60
	38	rs24048794	OPTC	3' UTR	C	0.23	0.09	A	0.01696	0.39	0.80	0.75	3.08	1.20 - 7.86
	9	rs24560587	COL1A1	Intronic	C	0.25	0.46	A	0.02576	0.59	0.90	0.76	0.39	0.18 - 0.85
Labrador Retriever	6	rs24306291	COL 24A1	Intronic	T	0.51	0.29	C	0.00002	<0.01*	<0.01*	<0.01*	2.56	1.66- 3.97
	6	rs24306964	COL 24A1	Synonymous	A	0.57	0.41	G	0.00458	0.11	0.41	0.21	1.86	1.21- 2.85
	8	rs8721919	LTBP2	Intronic	G	0.53	0.4	A	0.01443	0.36	0.79	0.37	1.74	1.12- 2.72
	12	rs22217603	COL 11A2	Synonymous	T	0.05	0.12	G	0.01828	0.46	0.85	0.37	0.37	0.16- 0.87
	38	rs24048817	OPTC	Splice region synonymous	A	0.04	0.01	G	0.02313	0.58	0.90	0.37	7.91	0.96- 64.96
	3	rs23546710▪	ACAN	Synonymous	A	0.17	0.3	G	0.02435	0.61	0.91	0.37	0.46	0.22 - 0.98
	38	rs24038028	OPTC	Downstream	C	0.03	0.01	T	0.04115	1.00	0.98	0.48	6.82	0.81- 57.24
	38	rs24029015	OPTC	3'UTR	T	0.03	0.01	C	0.04148	1.00	0.98	0.48	6.80	0.81- 57.06
Rottweilers	9	rs24535768	N/A (Was Elastin)	Intergenic	C	0.01	0.14	A	0.00098	0.02*	0.08	0.09	0.07	0.01 - 0.53
	9	rs8680883	COL 5A1	Intronic	C	0.53	0.34	T	0.00802	0.20	0.47	0.37	2.20	1.22 - 3.94
	30	rs23637297	FBN1	Intronic	A	0.32	0.47	C	0.02980	0.75	0.92	0.5	0.52	0.29 - 0.94
	30	rs23651397	FBN1	Intronic	T	0.33	0.47	A	0.04391	1.00	0.97	0.5	0.55	0.30 - 0.99
	15	rs22378757	Decorin	Upstream	T	0.11	0.03	C	0.04500	1.00	0.98	0.5	3.66	0.96 - 14.01
SBT	6	rs24278005	COL 11A1	Intronic	A	0.14	0.02	G	0.03693	0.96	0.95	1.00	8.37	0.82 - 85.42
	12	rs22255739	COL 11A2	Synonymous	A	0.25	0.08	G	0.04670	1.00	0.97	1.00	4.00	0.95 - 16.83
	12	rs22184922	COL 12A1	Synonymous	G	0.27	0.52	A	0.05014	1.00	0.98	1.00	0.35	0.12 - 1.02

KEY: A1 = base for allele 1, A2 = base for allele 2, EU = European Newfoundlands, NA = North American Newfoundlands, SBT = Staffordshire Bull Terriers, A1 = allele 1, A2 = allele 2 FA/U = frequency of allele in affected/unaffected dogs. p(raw) = raw *p*-value. Three different tests for correction for multiple testing were performed: p (BONF STYLE) = modified "Bonferroni-style" corrected *p*-value (EU, Labrador retrievers, Rottweilers corrected for 25 tests, NA corrected for 23 tests, Staffordshire Bull Terriers corrected for 26 tests), p (FDR) = *p*-value for FDR, p(perm) = *p*-value for 100,000 permutations. OR = odds ratio, 95% CI = 95% confidence interval of the OR. Asterix (*) = significant after multiple correction, "▪" = significant in more than one breed.

Four of the 104 SNPs that passed QA in the European subgroup of Newfoundlands had a raw *p*-value of ≤0.05. The SNPs were found within collagen genes (COL5A1 and COL3A1) and the fourth SNP was found in the ACAN gene. After correcting for multiple testing with all three methods, none were significant.

Three of the 90 SNPs that passed QA in the North American subgroup of Newfoundlands had a raw *p*-value of ≤0.05. The SNPs were in collagen genes (COL5A2 and COL1A1) and in OPTC gene. After correcting for multiple testing using all three methods none were significant.

Eight of the 92 SNPs in the Labrador Retrievers that that passed QA had a raw *p*-value of ≤0.05. The most significant SNP in this breed (rs24306291) was significant after correcting for multiple correction in all three tests ("Bonferroni Style" *p*=0.001, permutation *p*=0.003, FDR *p*=0.002). This was an intronic SNP in COL24A1. The second most associated SNP was also present in COL24A1 and the SNP ranked fourth was in COL11A2, the remaining significant SNPs were within genes LTBP2, OPTC and ACAN.

Five of the 92 SNPs that passed QA in the Rottweiler group had a raw *p*-value of ≤0.05. The SNP that ranked number two was in a collagen gene - COL5A1. The other associated SNPs were in FBN1 and Decorin. The most associated SNP rs24535768 (originally chosen for being in the Elastin gene in the CanFam2 annotation) is

intergenic in the new CanFam3 annotation. After correcting for multiple testing none were significant.

Three of the 87 that passed QA in the Staffordshire Bull Terrier group had a raw *p*-value of ≤0.05. All of the SNPs were within collagen genes (COL11A1, COL11A2 and COL12A1). After correcting for multiple testing using all three methods, none were significant.

A meta-analysis was carried out on the data to compare the breeds in a case control study as a whole cohort. Table 5.4 shows SNPs that were associated with CCL rupture from this analysis.

Table 5.4. SNPs associated with CCL rupture from the meta-analysis of the candidate gene data of the stratified CMH Newfoundland data, the European Newfoundlands, North American Newfoundlands datasets and the Labrador Retriever, Rottweiler and Staffordshire Bull Terriers datasets.

Gene	CHR	Position	SNP	Consequence of SNP	A1	A2	P	Estimated p (BONF STYLE)	OR
COL 5A1	9	50752492	rs8680883	Intronic	C	T	0.0011	0.03*	1.46
COL 1A1	9	26190947	rs24560587	Intronic	A	C	0.0018	0.04*	1.52
COL 5A1	9	50806219	rs24546447	Intron splice variant	T	G	0.0022	0.05*	2.45
COL 3A1	36	30519543	rs23942103	Intronic	G	C	0.0036	0.08	3.06
COL 5A1	9	50752484	rs8680884	Intronic	G	A	0.008	0.18	1.38
COL 5A2	36	30575052	rs8587007	Intronic	T	A	0.0114	0.26	0.65
N/A (Was Elastin)	9	21631629	rs24535768	Intergenic	C	A	0.0177	0.41	0.15
COL 5A1	9	50752527	rs8680882	Intronic	T	C	0.0195	0.45	1.38
COL 11A1	6	47575177	rs8652327	Synonymous	C	T	0.0218	0.50	0.31
FBN1	30	14717529	rs23648407	Intronic	C	T	0.0249	0.57	0.72
LOX	11	12036988	rs22190862	Intronic	A	G	0.026	0.60	0.76
COL 24A1	6	62282300	rs24306291	Intronic	T	C	0.0288	0.66	1.31
Novel (Was Elastin)	9	21569516	rs24553222	Intronic	T	C	0.0291	0.67	0.54
COL 5A1	9	50800325	rs24546410	Synonymous	A	G	0.0357	0.82	1.33
N/A (Was Elastin)	9	21622700	rs24535705	Intergenic	A	G	0.0358	0.82	0.57
COL 5A2	36	30548758	rs8867454	Intronic	T	C	0.0498	1.15	0.74
LTBP2	8	47686777	rs24469329	Missense	T	C	0.0499	1.15	1.56

KEY: A1 = base for allele 1, A2 = base for allele 2, Novel = novel protein coding gene, U6 = U6 spliceosomal RNA. OR = odds ratio. Estimated correction for multiple testing was run using the modified Bonferroni correction, correcting for 23 tests (genes common to all breed analyses).

Seventeen of the 107 SNPs which were common in two or more breed analyses were significant with a raw *p*-value of ≤0.05. Of the 17 SNPs that were associated with CCL rupture, 11 were within collagen genes – seven of which were collagen type-V (five in COL5A1 and two in COL5A2). The remaining collagens that contained

associated SNPs were COL1A1, COL3A1, COL11A1 and COL24A1. The other genes containing associated SNPs were FBN1, LOX, LTBP2, a novel gene and two SNPs were intergenic. The three SNPs with the lowest raw *p*-values were significant after estimated correction for multiple testing. They were in COL5A1 and COL1A1, indicating potential involvement of both the Collagen type I and type V genes with CCL rupture susceptibility.

One mutation, rs24469329 (in gene LTBP2) is a missense mutation and causes a change in the codon from GAA-AAA. The altered codon causes an amino acid substitution E^{744}-K, an acidic amino acid (Glutamic acid) being substituted with a basic one (Lysine). From the amino acid sequence the secondary structure was predicted for LTBP2 using the prediction software algorithm implemented by Chou & Fasman [340, 341]. The mutation is at amino acid residue 744 and is present in helix 721. There have been no NMR or X-ray crystallography structural models prepared for this protein that encompass the residue in question and so it is not possible to map on a 3D structure where the mutation lies (binding site, internal or external etc). Further functional analysis is required to determine the role of this mutation in CCL rupture susceptibility

Discussion

The candidate gene data were analysed on a breed by breed case control basis to determine whether any SNPs associated with CCL rupture were in multiple susceptible breeds. It was hoped that this approach might identify risk (or protection) associated SNPs specific for each breed and common to all (or several) breeds. A meta-analysis was conducted to enable the data to be examined as a single cohort, this increased statistical power relative to the individual breed analyses. It considered genetic influences on CCL rupture susceptibility across all breeds studied.

The most associated CCL rupture SNP in the Rottweilers (rs24535768) was an intergenic SNP (see Table 5.3) that was significant after correction for multiple testing with a "Bonferroni-style" *p*-value = 0.02. The SNP is ~50Kb upstream of the non-coding RNA for the U6 spliceosomal gene and is also ~50kb downstream of several Keratin-like associated genes. Keratin and keratin associated proteins are involved in the strength and structure of hair and skin and have functional roles in wound healing, epithelial repair and cytoskeletal protein synthesis [342, 343]. The homeostatic response to tissue damage in epithelial cells involves a complex response, including the upregulation of keratin filament genes. It has been suggested that a similar process to that which occurs during the epithelial regeneration, may also occur in other tissues (such as ligaments) in response to their injury [342]. Association of mutations in keratin genes with CCL rupture is therefore functionally feasible.

Two SNPs that were significant after correction for multiple testing in the meta-analysis were in Collagen type-V gene. This suggests that this could be a particularly important gene in relation to CCL rupture susceptibility. Collagen type-V is an important candidate gene because it is involved in the formation and assembly of the collagen fibrils [234, 321, 323], which are key to the composition and strength of the ligaments. A third SNP which also significant after correction for multiple testing was in collagen type-I, the most abundant collagen type found in ligaments [43]. As collagen is a key component of ligament, a number of collagen genes were chosen for investigation. Mutations that directly (or are in linkage disequilibrium with one that do) alter the collagen protein sequence, formation, structure and/or organisation of collagen fibrils, such that they are unable to retain their normal organised structures and strength may cause weakness and loss of key stability in the CCL. This would mean that a ligament formed with the mutated collagen fibrils may be more susceptible to rupture under normal conditions – i.e. non-contact rupture.

Other genes that had significant raw *p*-values (although significance was not seen after correction for multiple testing) include: Cathepsin K, Aggrecan, OPTC, LTBP2, FBN1, Decorin and LOX. Cathepsin K is a matrix degradation protease that is involved in bone resorption and the degradation of cartilage and the ECM [344]. Levels are often found to be increased after injury [345] and defects in Cathepsin K have been linked to skeletal disorders such as osteoporosis [346] as well inflammatory conditions such as rheumatoid arthritis [347, 348]. There is an

emerging role for Cathepsins in the innate immune response [349], where they are believed to initiate apoptosis of inflammatory cells such as neutrophils [350]. Cathepsin K has been shown to be differentially expressed between ruptured and intact CCLs [351] and as such was seen as an important candidate gene to consider in this study. The associated SNPs in this gene may be in LD with SNPs that are in regulatory regions of the gene. It is possible that SNPs may cause an increase or decrease of proteolytic/collagenolytic activity of Cathepsin K within ligaments. Alternatively, a SNP may cause a change in localisation of the enzyme which may contribute to disease susceptibility.

SNPs in several extracellular matrix (ECM) proteins were found to be associated (raw *p*-values) with CCL rupture in this study, although after correction for multiple testing they did not show significance. They were Fibrillin-1 (FBN1), Decorin, Lysyl oxidase (LOX), Aggrecan (ACAN), Opticin (OPTC) and Latent transforming growth factor beta 2 (LTBP2). FBN1, Decorin, ACAN and OPTC are heavily glycosylated glycoproteins (proteoglycans), important in providing the structural support and integrity of the ECM and the ligaments [352, 353]. The integrity of the ECM is important in growth and repair [354], cell matrix communication [355] and is also important in withstanding load compression [356]. Altered ECM composition may lead to an increase in CCL rupture risk, as the ligaments are structurally altered such that they are unable to withstand normal load and compression.

Aggrecan (ACAN) is involved in cartilage formation which helps resist compression of cartilage and ligaments and is therefore critical in the structure and strength of a joint [66, 336, 357]. If ACAN is mutated such that the ability of the ligament and/or cartilage to resist compression and stretch becomes compromised, it may lead to premature rupturing of the ligament. This theory is supported by Pratta et al. [358] and many degenerative joint diseases such as osteoarthritis are characterised by proteolytic destruction of the matrix and loss of ACAN [336].

Latent transforming growth factor beta 2 (LTBP2), which has a high degree of homology with fibrillins [307], helps in the formation of elastin microfibrils [359]. This gene may have a key function in the ligament's role as a connective tissue, providing recoil and resilience properties. It also has a role in cell adhesion [360, 361], a key process in tissue repair [362].

Fibrillin 1 (FBN1) is essential for the formation of the elastic fibres and is involved in structural support of connective tissues such as elastin [363]. Like LTBP2, it is essential in maintaining the elasticity of the ligaments.

Lysyl Oxidase (LOX) is an extracellular enzyme that aids in the crosslinking of elastin and collagen fibres [51] by causing the catalysis of specific lysine and hydroxylysine residues in the collagen amino acid chain. The crosslinking determines the strength of the collagen fibres and the integrity of the ECM and the ligament. Mutations

within this gene may impact on the strength of the collagen and lead to a ligament that is unable to withstand daily stresses and is more susceptible to rupture than a ligament formed with normal crosslinks.

Decorin is involved in matrix assembly and fibril formation and organisation [364] and can help bind growth factors such as epidermal growth factor (EGF) as well as collagens [365]. Disruptions in the decorin gene may lead to altered packing and/or assembly of the collagen fibrils, disrupting the normal strength of the fibres and may lead to a ligament that is more likely to rupture.

Finally, Opticin (OPTC) helps to regulate collagen fibril assembly [305] and organisation. Mutations in this gene may have some impact on the ligament's basic structure by compromising the ligament's strength, leading to increased CCL rupture susceptibility. Although opticin has predominately been found localised in the ocular system [366], it has also been detected in several other canine tissues including shoulder, elbow, rib chondrocyte, ligament, skin, muscle and testes [305]; therefore association with CCL rupture can be tentatively inferred.

All of the above genes, although they were not significant after correction for multiple testing, were chosen for analysis because they are key components in the ECM and are structural constituents of ligaments. A defect in any one of these genes may significantly affect the structural integrity of the ligament, resulting in its

failure during normal activity. Failure to retain significance after correction for multiple testing may be due to the small samples numbers. Further validation of these data in a larger cohort of dogs would verify the true significance of the results.

Comparing the alleles that confer risk to the disease is important for development of genetic screening tests and for determining risk of each dog. With the exception of SNP rs23546710 between the European Newfoundlands and Labrador Retrievers, the "risk" allele is the same in each breed. SNP rs23546710 has the same minor allele (A) in each breed, but the odds ratio shows the risk allele to be 'G' for European Newfoundlands and 'A' in the Labrador Retrievers. This ambiguity may be attributed to the SNP being in LD with the associated mutation rather than being associative itself. These mutations may be different and present with variable frequencies in the breeds studied, or there could have been a genetic cross-over during breed creation. Alternatively, the SNPs could be subject to genotyping error and may have been miss-called by the analysis software - this is unlikely however, as all SNPs were visually checked for correct clustering. Once the associated mutations have been identified and the risk alleles have been determined for each breed, a screening test can be developed. The genotype of each dog would be examined and those that have the risk alleles at each loci would be significantly more at risk compared to those dogs that have fewer or no risk alleles [367].

To date, no candidate gene studies have shown significant results for canine CCL rupture. This study is the first of its kind to show potential associations of certain SNPs/genes with CCL rupture in dogs. The results need further validation in a larger number of dogs and breeds to confirm the associations and also to confirm (or refute) that different SNPs may be important in different breeds for susceptibility to CCL rupture. Fine mapping and next generation re-sequencing of the associated SNP regions should also be carried out to further examine regions of association. Expression work such as transcriptomics using Q-PCR should also be conducted for the significantly associated genes found in this study.

In summary, this study confirms conclusions of several other studies investigating genetic components of cruciate rupture; collagens are important genes in CCL rupture susceptibility. Collagen type-V and type-I are the predominant genes that have statistically significant SNPs associated with CCL rupture in this study.

Chapter 6.

DLA typing of Newfoundlands with cruciate rupture

Introduction

Previous chapters have considered genetic involvement in CCL rupture susceptibility, but have not focused their approach on any specific detailed genomic regions. As there may be an autoimmune component to CCL rupture [3, 31, 105] the main genes that control the immune system - the Major Histocompatibility complex (MHC) will be studied. MHC genes control the strength and nature of immunity and in dogs this is called the dog leucocyte antigen (DLA) system. Variations in the DLA region can control susceptibility of dogs to diseases. DLA associations have been found for several canine autoimmune diseases such as hypothyroidism, polyarthritis and diabetes [119, 120, 124].

There are two classes of MHC molecules, class I and class II. Class I have the role of presenting antigens from intracellular origin to CD8 cytotoxic T-cells [368]. Class II molecules can only present antigens from an extracellular origin, which have been taken up from vesicles of phagocytic cells such as macrophages, dendritic cells and B cells [369]. The DLA molecules then present the antigens to CD4 T-cells [368]. These are helper T-cells which activate other T cells and B lymphocytes to generate effective immune responses [369]. Destructive autoimmune responses may occur if self-antigens are recognised and targeted.

There are three genes that code for the MHC class II molecules in dogs and these are called DLA-DRB1, DLA-DQA1 and DLA-DQB1. It is the hypervariable region in the second exon of each of these genes that determines the efficiency of binding to different antigens [370]. There are many different alleles for each gene >250 for DRB1 (206 with official names), >45 for DQA1 (37 with official names), >140 for DQB1 (100 with official names) and >280 haplotype combinations found in three or more dogs [personal communication from Dr LJ Kennedy]. Many of these are breed specific and each breed has only a limited diversity of DLA alleles [371-374]. The Centre for Integrated Genomic Medical Research (CIGMR) at Manchester University has considerable expertise in studying DLA alleles and their disease associations; because of the possible autoimmune nature of CCL rupture, DLA-typing our Newfoundland dog populations was considered to be a viable strategy. There is high linkage disequilibrium within DLA, especially between DLA-DQA1 and DLA-DQB1 [375], because of this it is possible to determine haplotypes of the DLA group by typing just two of the three known genes (DLA-DRB1 and DLA-DQB1) to assess if there is an DLA association with susceptibility to CCL rupture.

Methods

Sample selection

Ninety-five Newfoundland dogs (the same 95 that have previously been genotyped by Illumina GWAS and Sequenom) were characterised for the two main loci of DLA class II (DLA-DRB1 and DLA-DQB1) by sequence based typing of DNA using locus

specific, intronic primers. A negative control containing no DNA was also included to identify any contamination or crossover of reagents.

Primer design

The primers were produced by Metabion (Martinried, Germany) as previously described [372, 376], and are detailed in Table 6.1

Table 6.1. DLA primers used for genotyping the DLA

Primer		SEQUENCE
DRB1 forward	DRBIn1	CCGTCCCCACAGCACATTTC
DRB1 reverse	DRBIn2-T7	TAATACGACTCACTATAGGGTGTGTCACACACCTCAGCACCA
DQB1 forward	DQB1BT7	TAATACGACTCACTATAGGGCTCACTGGCCCGGCTGTCTC
DQB1 reverse	DQBR2	CACCTCGCCGCTGCAACGTG

The product sizes are 303bp for DRB1 and 300bp for DQB1.

PCR

DNA was normalised to 5ng/µl. The PCR reaction was carried out in 96 well plate format, with 25ng DNA in a 25µl reaction volume, containing 20µl of master mix and 5µl of DNA (5ng/µl). For each well the master mix included: 2.5µl 10x PCR buffer (20mM, Roche, Hertfordshire, UK), 0.5µl dNTP mix (10mM, Roche, Hertfordshire, UK), 0.13µl of fast start Taq (5U/µl, Roche, Hertfordshire, UK), 0.5µl of forward primer (50µM), 0.5µl of reverse primer (50µM) and 15.87µl of dH_20. The PCR was run on a tetrad gradient cycler in a touch-down PCR cycle with the following conditions:

DQB

95˚C – 5 minutes
95˚C – 30 seconds
73˚C – 1 minute **-1°C each cycle** } X14
72˚C – 1 minute
95˚C – 30 seconds
66˚C – 1 minute } X20
72˚C – 1 minute
72˚C – 10 minutes

DRB

95˚C – 5 minutes
95˚C – 30 seconds
62˚C – 1 minute **-1°C each cycle** } X14
72˚C – 1 minute
95˚C – 30 seconds
55˚C – 1 minute } X20
72˚C – 1 minute
72˚C – 10 minute

To confirm successful amplification, 5µl of each resulting PCR product was run on a 2% agarose gel. The PCR products were purified using shrimp alkaline phosphatase (SAP) to remove any unincorporated dNTPs. A 1:8 dilution of SAP (USB® ExoSAP-IT®, Affymetrix, UK) was made and 4µl was mixed with 1µl of PCR product. This was then left for one hour at 37˚C and then heated for 15 minutes at 80˚C.

DLA sequencing

The samples were sequenced at SourceBioscience (Nottingham, UK). The sequence analysis and allele/haplotype designation was carried in conjunction with Dr LJ Kennedy, CIGMR[2].

Allele designation of sequence peaks was conducted using the SBTengine software (http://www.gendx.com) – see Figure 6.1.

[2] Dr Lorna Kennedy, CIGMR (Centre for Integrated genomic research), Faculty of Medical and Human Sciences Manchester University, Manchester, M13 9PT, UK

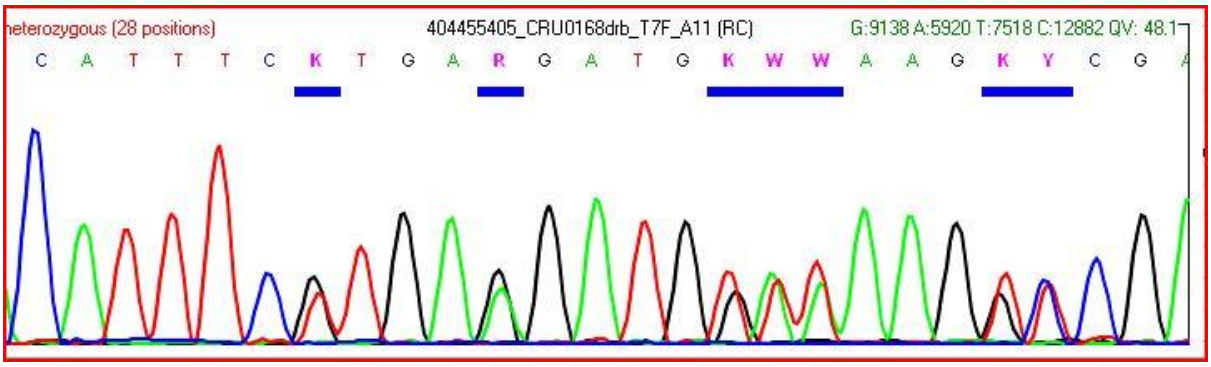


Figure 6.1. An example chromatogram showing sequencing of DLA haplotypes. Allele designation was carried out in standard way: colour of peaks – allele: Green = A, Red = T, Blue = C, Black = G. Double peak in one position = heterozygous (C+G = S, C+T = Y, C+A = M, G+T = K, G+A = R, T+A = W).

The software aligned the generated sequences to a consensus sequence and each polymorphic locus was analysed by the software and checked manually. The annotated sequence was matched to a library reference sequence (taken from http://www.ebi.ac.uk/ipd/mhc/dla/index.html).

Analysis of haplotypes

Haplotype frequencies were determined for each analysis (Case control, European case control, North American case control and European versus North America) and the statistical analyses was performed using STATCALC [377]. Contingency tables (2×2) were used to determine a Yates corrected chi-squared value and corresponding *p*-value which are tabulated in each results table. Correction for multiple testing is not currently carried out when analysing DLA data [121, 173].

Results

Ninety-five Newfoundlands were characterised for their DLA loci and nine different DLA haplotypes were found in the cohort as a whole; DRB*00601/DQB*00701, DRB*00201/DQB*00101, DRB*00601/DQB*02001, DRB*00901/DQB*008011, DRB*01201/DQB*013017, DRB*01301/DQB*00201, DRB*01501/DQB*02301, DRB*02001/DQB*01303 and DRB*03001/DQB*00301.

Analysis was split into standard case control analyses (Table 6.2) and case control analyses were carried out on the European (Table 6.3) and North American (Table 6.4) samples separately. A comparison between the European and North American dogs was also carried out irrespective of cruciate rupture status (Table 6.5). Significant haplotypes ($p \leq 0.05$) are highlighted in yellow in following the tables.

Table 6.2. DLA haplotype frequencies, full cohort case control.

DLA Haplotype	Number haplotypes: Cases	Number haplotypes: Controls	*p*-value	OR	95% CI
DRB*00201 / DQB*00101	16	16	0.953	0.95	0.41 - 2.17
DRB*00601 / DQB*00701	7	5	0.825	1.37	0.37 - 5.20
DRB*00601 / DQB*02001	4	5	0.947	0.76	0.16 - 3.38
DRB*00901 / DQB*008011	0	5	0.062	0.00	0.00 - 1.09
DRB*01201 / DQB*013017	16	22	0.288	0.63	0.29 - 1.38
DRB*01301 / DQB*00201	25	10	0.013	2.90	1.22 - 6.99
DRB*01501 / DQB*02301	8	6	0.847	1.30	0.39 - 4.45
DRB*02001 / DQB*01303	13	12	0.907	1.04	0.42 - 2.62
DRB*03001 / DQB*00301	5	9	0.358	0.51	0.14 - 1.74

KEY: Yellow = significant (*p*-value ≤0.05), OR = Odds ratio

Table 6.3. DLA haplotype frequencies, European case control study.

DLA Haplotype	Number Haplotypes: EU case	Number Haplotypes: EU control	*p*-value	OR	95% CI
DRB*00201 / DQB*00101	12	13	1.000	0.90	0.33-2.46
DRB*00601 / DQB*00701	5	5	0.738	1.00	0.23-4.37
DRB*00601 / DQB*02001	3	5	0.712	0.57	0.10-3.00
DRB*00901 / DQB*008011	0	0	-	-	-
DRB*01201 / DQB*013017	11	14	0.642	0.72	0.26-1.98
DRB*01301 / DQB*00201	11	5	0.171	2.56	0.73-9.41
DRB*01501 / DQB*02301	1	2	1.000	0.49	0.02-7.24
DRB*02001 / DQB*01303	3	0	0.241	-	-
DRB*03001 / DQB*00301	2	4	0.673	0.48	0.06-3.27

KEY: EU = European, OR = Odds ratio

Table 6.4. DLA haplotype frequencies, North American case control study.

DLA Haplotype	Number Haplotypes : NA case	Number Haplotypes: NA control	*p*-value	OR	95% CI
DRB*00201 / DQB*00101	4	3	0.900	1.24	0.21-7.55
DRB*00601 / DQB*00701	2	0	0.515	-	-
DRB*00601 / DQB*02001	1	0	0.963	-	-
DRB*00901 / DQB*008011	0	5	0.051	0.00	0.00-1.01
DRB*01201 / DQB*013017	5	8	0.436	0.52	0.13-1.97
DRB*01301 / DQB*00201	14	5	0.064	3.24	0.94-11.69
DRB*01501 / DQB*02301	7	4	0.628	1.71	0.40-7.65
DRB*02001 / DQB*01303	10	12	0.622	0.69	0.24-2.03
DRB*03001 / DQB*00301	3	5	0.613	0.52	0.09-2.72

KEY: NA = North American, OR = Odds ratio

Table 6.5. DLA haplotype frequencies, European versus North America (irrespective of cruciate status).

DLA Haplotype	Number haplotypes: EU	Number haplotypes: NA	*p*-value	OR	95% CI
DRB*00201 / DQB*00101	25	7	0.00238	4.07	1.56 - 11.08
DRB*00601 / DQB*00701	10	2	0.05290	5.00	0.99 - 34.10
DRB*00601 / DQB*02001	8	1	0.05500	7.91	0.97 - 172.13
DRB*00901 / DQB*008011	0	5	0.05560	0.00	0.00 - 1.04
DRB*01201 / DQB*013017	25	13	0.08840	2.03	0.91 - 4.58
DRB*01301 / DQB*00201	16	19	0.50800	0.73	0.33 - 1.62
DRB*01501 / DQB*02301	3	11	0.03420	0.23	0.05 - 0.92
DRB*02001 / DQB*01303	3	22	0.00004	0.10	0.02 - 0.36
DRB*03001 / DQB*00301	6	8	0.65400	0.67	0.20 - 2.23

KEY: EU = European, NA = North American, Yellow = significant (*p*-value ≤0.05), OR = Odds ratio

A visual representation of the haplotype data for each analysis (using pie-charts) is shown in Figure 6.2.

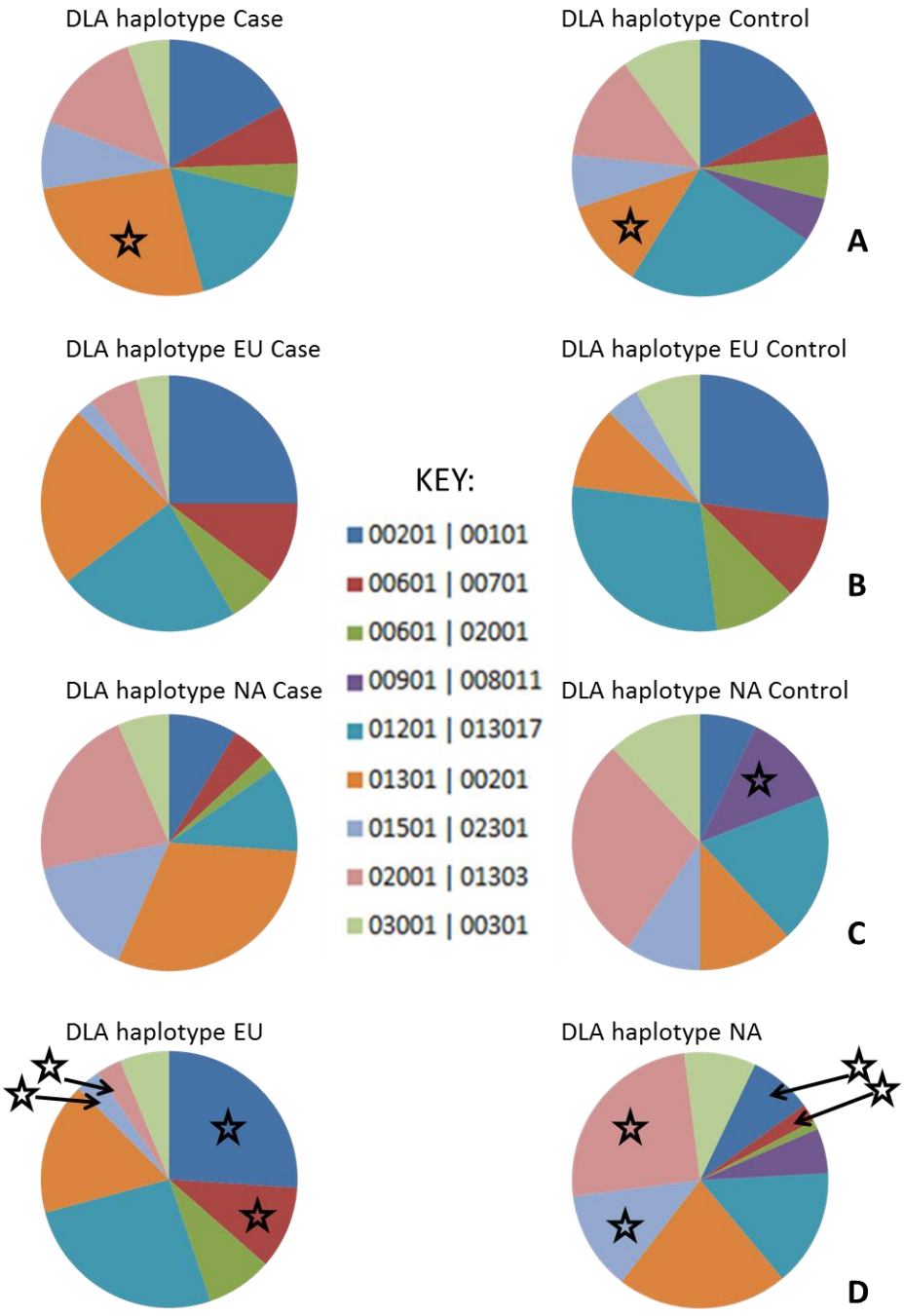


Figure 6.2. Pie charts showing the DLA haplotype frequencies in the different analyses. A star (☆) indicates significantly different haplotype (p≤0.05) EU = European, NA = North American. A – Case control analysis, B – EU case control analysis, C – NA case control analysis, D – EU/NA analysis

Discussion

DLA associations have been found for many canine auto-immune diseases including SLE and autoimmune haemolytic anaemia [122, 378]. However, it has previously been reported, in a cohort of Labrador Retrievers and Golden Retrievers, that there is no association between DLA haplotypes and CCL rupture [379]. At the time of this study, there had been no investigation into DLA haplotype association with CCL rupture in Newfoundlands, nor had there been a study investigating DLA haplotypes across geographical groups within the same breed of dog.

Comparing DLA haplotypes in a case control study using Newfoundlands with cruciate rupture compared to healthy controls (using the whole, unstratified, cohort of 95 Newfoundlands), there was only one haplotype that was significant, haplotype DRB*01301/DQB*00201 (p=0.013). However, having demonstrated in previous chapters that it is inappropriate to analyse North American and European Newfoundlands in the same analysis without first accounting for population sub-structure, this would need to be validated in the separate European and North American cohorts before the haplotype can be classed as being truly associated with cruciate rupture susceptibility in Newfoundland dogs.

When the data was split into separate European and North American cohorts and the DLA haplotypes then assessed for association with cruciate rupture on a case control basis, there were no significant haplotypes associated in the European cohort, but there was one haplotype significant in the North American cohort. This

haplotype, DRB*00901/DQB*008011, was only found in the North American Newfoundlands.

Haplotype DRB*01301/DQB*00201 had the lowest p-value in the European and the second lowest *p*-value in the North American haplotypes (*p*-value 0.17 and 0.06 respectively). It failed to reach significance in either analysis, which may in part be due to the small individual sample group sizes i.e. 47 dogs from the European group and 48 from the North American group. To confirm whether this haplotype is definitely significantly associated with CCL rupture in the Newfoundland breed the analysis should be repeated with a much larger cohort of Newfoundlands from each country.

When comparing the European Newfoundlands with those from North America (irrespective of the cruciate status), there were eight DLA haplotypes common to both subgroups of Newfoundlands, four of which exhibited significant differences between the two populations. These were DRB*02001/DQB*013030 (*p*=0.00004), DRB*00201/DQB*00101 (*p*=0.002), DRB*01501/DQB*02301 (*p*=0.03) and DRB*00601/DQB*00701 (*p*=0.053). There were also two haplotypes that were nearing statistical significance in the dataset DRB*00601/DQB*02001 (*p*=0.055) and DRB*00901/DQB*008011 (*p*=0.056). The four significant haplotypes and two that were nearing significance showed that there are clear DLA differences between European and North American Newfoundlands. It also confirmed the genetic population stratification within the Newfoundland breed from different

geographical origins that was found in the other datasets in this project. The stratification is probably due to limited intercontinental breeding. This effectively acts as if the populations are isolated from each other and will lead to different allele frequencies between the two groups. The stratification within the Newfoundland breed will have an impact on any future genetic investigations into canine immune-related conditions using the breed. It would suggest that any genetic studies (small or large) have a PCA (or similar population stratification test) conducted on the data (even from dogs within the same breed) to determine if stratification exists before performing any genetic and statistical phenotype association analyses. If there are differences within the populations, then the stratification must be accounted for or the results analysed separately.

Previous research has shown that there are differences in the DLA haplotype frequencies found between different breeds of dog [371, 373, 380]. This is due to breed isolation, breed creations and other bottlenecks. There has been limited research has been conducted into different haplotypes within the same breed of dog but from different populations – such as geographical location. Isolated populations of feral dogs, such as those found on islands like Bali, Australia and New Guinea have large genetic diversity of DLA [381]. There are no selective breeding pressures or physical barriers to prevent cross breeding, populations are large, outbred and heterozygosity is high. Pedigree dogs however, are isolated by breed rules and strict breeding standards, hence in these circumstances

heterozygosity and genetic diversity is low and resistance to certain diseases, including auto-immune diseases is reduced.

In conclusion, this chapter supports the work of Clements et al. (2011) that there is no DLA association involved in CCL rupture susceptibility whether this be in the Labrador Retriever, Golden Retriever or Newfoundland breeds of dog. There may still be autoimmune components to cruciate ligament rupture, but the SNPs and genes involved are not part of the DLA region. This chapter confirms that there is genetic stratification within the Newfoundland breed, which can be attributed to the geographical origin of the dog.

Chapter 7. Quantitative gene expression analysis between ruptured & healthy CCL tissue

Introduction

In previous chapters, candidate genes and genomic regions associated with CCL rupture were identified either by whole genome approaches (GWAS) or by selection as potential candidates using the existing knowledge of the disease process. These genetic approaches considered the inheritance of CCL rupture, but did not take into account potential downstream effects of these genetic changes. The first important step after genomic changes is the actual expression of a gene under different circumstances such as growth, repair or disease. It is possible that many genomic differences could be detected between the different phenotypes but only a few would be relevant to the development of a phenotype. It is for this reason, so much attention has been given to gene expression in the last 20 years, particularly in relation to health and disease [382-387].

By investigating gene expression patterns, one may identify particular processes which are functioning to either cause a condition or attempting to prevent the repair of disease processes. Another reason that the science of transcriptomics has become so well used in comparative biology is because it is now relatively easy to do and is cheap to generate quantitative data with fairly inexpensive laboratory kit. Indeed, it is far easier, because of the genome annotations now available, to assess gene expression for almost any molecule in almost any species and this would be virtually impossible to achieve at the level of proteins, even with the recent advances in proteomics.

Gene expression changes have been described for many diseases [386, 388-391]. Genetic variations that influence gene expression, such as changes in the DNA sequence in the promoter or enhancer regions of a gene, can impact on disease phenotype. Previously, gene expression would have been assessed by RNA microarrays [392, 393], but these have largely been superseded by Q-PCR. [385, 394] and more recently, global transcriptomics (RNAseq) [395]. Quantitative-PCR (Q-PCR) can be used to examine changes in gene expression over time in a disease, or during treatments regimens and as such is a useful diagnostic aid, as well as a tool for research.

Q-PCR is an innovative version of a standard PCR reaction that evaluates gene expression changes in real time. It measures the kinetics of the reaction in the early stages of the cDNA amplification (exponential phase). Assuming 100% reaction efficiency, the amount of product doubles with each cycle of PCR that is carried out. There is a quantitative relationship between amount of starting product and amount of target product at any given cycle.

The aim of this study was to evaluate gene expression changes between ruptured CCL tissues compared to normal healthy CCL tissue samples in a standard case control experiment.

Methods

Sample collection and storage

Five clinical CCL rupture cases (ruptured ligament tissue) were collected during reconstructive CCL surgery. The breeds were varied and included a Border Terrier, a Jack Russell Terrier, an English Springer Spaniel, a Bulldog and Labrador Retrievers. These were compared with healthy control CCL tissue collected from five cadavers that had been euthanased for reasons other than degenerative joint diseases (as described previously in Chapter 2). The control breeds comprised of Beagles and Staffordshire Bull Terrier type breeds.

Once collected, the tissue was immediately placed into RNAlater solution (Invitrogen California, US) to quickly and efficiently stabilise the RNA in the sample. This alleviates the necessity to process it immediately or snap freeze the samples in liquid nitrogen. When required for analysis, the sample was removed from RNAlater and processed as if it were normal fresh tissue just collected.

RNA Extraction

Tissue samples were removed from the RNAlater solution and homogenised with 1ml of trizol reagent (Invitrogen, California, US) before RNA extraction. The RNA was extracted from samples using the PureLink RNA mini kits (Ambion, UK) according to the standard manufacturer's instructions, for more detail see materials and methods (Chapter 2).

The RNA quality and quantity was assessed as detailed in Chapter 2 using the Agilent 2100 Bioanalyser (Agilent, California, US).

cDNA synthesis

Reverse transcription was performed using oligo dT primers and Superscript II reverse transcriptase (Invitrogen, California, US) according to the manufacturer's instructions; for a more detailed method see materials and methods (Chapter 2).

Design of primers and probes

Target sequences were chosen for Q-PCR by predicted functional association with CCL rupture: MMP-2 (matrix metalloproteinase-2), BGN (Biglycan), TRAP (tartrate resistant acid phosphatase), CTSK (Cathepsin-K), MMP-1 (matrix metalloproteinase-1), COL1A1 (collagen type I alpha 1), COL1A2 (collagen type I alpha 2), IL-6 (interleukin 6), TNFα (tumour necrosis factor alpha). In addition, the nine most associated genes from the GWAS were also selected; SEMA5b (semaphorin type 5), CDH19 (cadherin 19), GAP43 (Growth Associated Protein 43), DIRC2 (disrupted in renal carcinoma 2), CDH7 (cadherin 7), PARP-14 (poly ADP ribose polymerase 14), PARP-9 (poly ADP ribose polymerase 9), PARP-15 (poly ADP ribose polymerase 15) and PDIA5 (protein disulfide isomerase family A member 5).

Transcript sequences of the cDNA were obtained from Ensembl (http://www.ensembl.org, CanFam2) and primer and probe reactions were

designed for each gene of interest using the Universal Probe Library Assay Design Centre [396] (Roche, Hertfordshire, UK). GAP43 was excluded from the study because assay design failed, leaving a total of 17 genes to design and validate primers and probes for.

Each primer and probe sequence was compared against the canine genome using BLAST (http://blast.ncbi.nlm.nih.gov/) to ensure specificity for chosen target only. Primers were designed with the following constraints; a G-C content within the range of 30-80%, identical runs of bases (especially guanine) were avoided and the T_m was in the range of 58-60˚C. The primers were designed to cross intron/exon boundaries so that any contamination with genomic DNA could not be detected – only cDNA from mRNA gene transcripts would be amplified, not genomic copies of the gene. Chosen genes and the associated primer/probe sequences and details are shown in Table 7.1. The probe for the PARP-14 assay was not found in our library and was excluded from the experiments, this left 16 genes for experimental detection.

Lyophilised primers were resuspended to a stock concentration of 100µM. The primer mix was created for each assay; 10µl of forward primer plus 10µl reverse primer and 180µl of molecular grade water – each primer in the mix was at a concentration of 5µM.

Table 7.1. Genes (including 3 reference genes) used for expression analysis in CCL tissue

Gene symbol	CHR	Transcript ID	Roche Probe #	Probe sequence	Amplicon length (nt)	Primers	TM	% GC
CDH19	**1**	ENSCAFT00000000075	5	cagccaca	75	Left - tgatcgagagattagtgcttgg	59	45
						Right - gcagaaatttgttgcacattg	59	38
CDH7	**1**	ENSCAFT00000000076	89	ggatgctg	114	Left - aattcttggatggtccgtacac	60	45
						Right - gggcactgttgccatatgta	59	50
MMP2	**2**	ENSCAFT00000014986	14	ctgggaga	95	Left - gggggagactctcactttga	59	55
						Right - cagtactctccgtcggcatt	60	55
MMP-1	**5**	ENSCAFT00000023900	48	actgggaa	94	Left - tcgatgctgctgtttctgac	60	50
						Right - tccatggattgcttatattcatca	60	33
COL1A1	**9**	ENSCAFT00000026953	65	ctggagga	94	Left - ctttgctccccaaatgtctta	59	43
						Right - gagaccacgaggaccagaag	59	60
TNF	**12**	ENSCAFT00000000804	29	cttctgcc	119	Left – cctcagcctcttctccttcc	60	60
						Right - tgattagttggaggccatttg	59	43
COL1A2	**14**	ENSCAFT00000031580	81	ggccctgg	66	Left - gcccagtatgatggaaaagg	59	50
						Right - acctctaggtcccatcaaacc	59	52
IL6	**14**	ENSCAFT00000004340	65	tcctccag	114	Left - ccaatctgggttcaatcagg	60	50
						Right - acattttccttatcaccctcatagtt	59	35
CTSK	**17**	ENSCAFT00000019202	38	ggaagcag	78	Left – caatgggacctttggaagaa	59	45
						Right - ccaaattaaacgccgagaga	60	45
TRAP	**20**	ENSCAFT00000027439	20	ctggctgg	111	Left - ccaacgacaagaggtttcg	59	53
						Right - agacgttccccaggtggt	60	61
DIRC2	**33**	ENSCAFT00000019042	5	tgtggctg	113	Left – tctccagctgggacatcg	60	61
						Right - cagaagcacagttatccgaaga	59	45
PARP-9	**33**	ENSCAFT00000018954	39	aggtggag	60	Left - tgaaaactgtggtttgcaggt	60	43
						Right - gccatgagaacggaattgtc	60	50
PDIA5	**33**	ENSCAFT00000019138	20	ctggctgg	85	Left - aatgccacatttccagaagg	59	45
						Right - ttcaaattccgaggggtaga	59	45
SEMA5B	**33**	ENSCAFT00000019107	14	tctcccag	109	Left - gggtcttcaccaccaacgta	60	55
						Right - gggttctcctggtatcgaaat	59	48
PARP-14	**33**	ENSCAFT00000019011	35	agaagagga	91	Left - aaagaacatgtcaaagagccaga	60	39
						Right - cacattcttctgaggtatcttcca	59	42
PARP-15	**33**	ENSCAFT00000018979	19	ggctggag	64	Left – tcccaatgaggacaagaatctc	60	45
						Right - gctggagaccatcgtgaac	59	58
BGN	**X**	ENSCAFT00000030457	22	ctccacca	84	Left – tgcagaacaacgacatctcag	60	48
						Right - tgttgttcaccaggacgaga	60	50
C7orf28B Ref Gene	**6**	ENSCAFT00000036998	72	gccaggaa	89	Left – gcaggaagggattctccagt	60	55
						Right -tcattaaggttcaagggtcca	59	43
ATIC Ref Gene	**37**	ENSCAFT00000022733	62	cagcaggt	79	Left - cgctgcctctttcaaacat	59	47
						Right - gcagactttggcctcatctt	59	50
HIRP5 Ref Gene	**10**	ENSCAFT00000005296	39	aggtggag	77	Left – aattcagaacatgctgcaatttta	60	29
						Right - tttcatctgattcatcatccataac	59	32

Reference gene selection

To account for variations in RNA extraction, the amount of initial sample or RT-variability, it was necessary to include reference genes, to which each unknown sample was normalised [397]. Reference genes (often referred to in the literature as housekeeping genes) need to be stably expressed in the tissue of interest and the expression levels of such genes should not alter with disease state or environmental factors [398].

Historically, the same reference genes were used universally for all experiments across all tissue types and experiments. However, more recently, it has been found that there is huge variability in the expression of commonly used reference genes such as GADPH and beta-actin in different tissues and under different experimental conditions [399, 400]. It has been necessary therefore, to use statistical algorithms such as geNORM [401], BestKeeper [397] and NormFinder [402] to determine which genes are stably expressed in the specific tissue being studied. This is done by testing a panel of SNPs in the chosen tissue on either microarray or Q-PCR platforms. Reference genes that have already been proven as stably expressed in CCL tissue via these methods were used in this study; ATIC, HIRP5, C7orf28B, MRPS7 and ORMLD2 [13]. These five genes were further reduced in number by selecting the three most stably expressed genes (ATIC, HIRP5 and C7orf28B) which were then taken forward as reference genes for the normalisation of the Q-PCR data within this project. Sequences and probe information for the selected reference genes are included on the end of Table 7.1.

Q-PCR

Each Q-PCR reaction was carried out using the Applied Biosystems 7900HT PCR machine in 384 well formats. There were three negative (no template) controls for each assay. Each reaction was in 10µl volume - containing 5µl of master mix (Taqman 2X PCR master mix, Applied Biosystems, California, US), 0.2µl of primer mix (each forward and reverse primer in mix at 5µM, Roche, Hertfordshire, UK), 0.1µl of 10µM probe (Roche, Hertfordshire, UK), 0.1µl distilled water and 4.6µl of sample cDNA (or water for the negative controls). The PCR reaction was carried out using the following thermal cycling conditions:

50˚C – 2 minutes
95˚C – 10 minutes
95˚C – 15 seconds
60˚C – 1 minute
} X40

The Q-PCR was carried out using TaqMan style probes, from the Universal Probe Library (Roche, Hertfordshire, UK). The probes were dual-labelled with a reporter fluorophore (FAM) on the 5' end and a quencher dye on the 3' end. When Taq polymerase extends through the probe it is hydrolysed and the fluorophore is released from the inhibitory effect of the quencher. A fluorescent signal is produced that is then detected by the camera in real time. This process is illustrated in Figure 7.1.

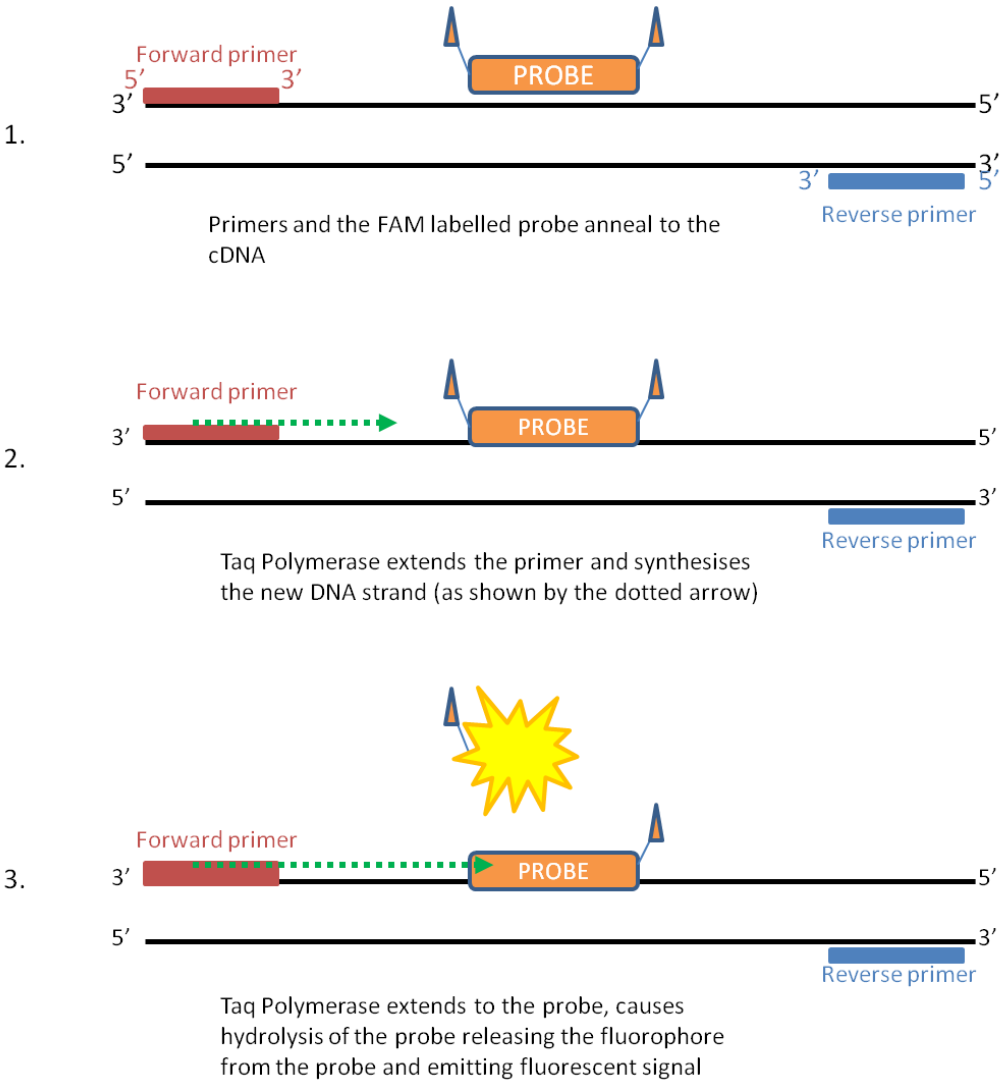


Figure 7.1. Diagrammatic representation of TaqMan probe chemistry (adapted from www.Invitrogen.com).

1. The probe anneals to the cDNA sequence downstream of the primer.

2. Taq DNA polymerase extends the primer and synthesises a new DNA strand.

3. Taq DNA polymerase extends into the probe causing hydrolysis of the probe and release of the fluorophore. This emits a fluorescence signal that is proportional to the amount of amplicon product.

Primer probe validation

The primers and probes needed to be validated to check they worked effectively and to check the efficiency of the reactions. A serial dilution of the template was made with six dilutions from 1:10 to 1:1,000,000. A negative control was included to form a standard curve. Each dilution was run in triplicate with the standard Q-PCR mix, for the standard cycling conditions (see above).

The efficiency of a Q-PCR reaction is the rate at which the amplicon is generated. In the geometric phase of amplification, the product rate should double with each PCR cycle, so a reaction should have an efficiency of 100%. The slope of the reaction can be used to determine the efficiency (E), using the following equation $E = 10^{(-1/slope)-1}$. The ideal slope should have a value of -3.32 which correlates with an efficiency of 100% i.e. 100% of the target is amplified. A slope of between -3.1 and -3.7 is considered to be the acceptable with upper and lower values for an efficiency of between 110% and 86%. A reaction that has an efficiency of greater than or less than this range indicates a a non-linear reaction. This may occur when too much target is added (causes inhibition and saturation of the PCR reaction), too little target is added (too low a concentration of sample to detect a CT), pipetting errors when creating the standard curve, detection of a PCR inhibitor, primer-dimer formation or bad design of primers/probes. Any primer/probe sets that were out of the efficiency range in this project were excluded from the study.

A 1:10 dilution of sample was chosen for use in experiments in this project. This concentration of sample universally produced the best amplification curves during primer testing and showed no inhibition of reaction due to too much template or too little amplification due to small amounts of target.

Data analysis

Determination of CT values

A threshold value is decided in the exponential phase of the amplification reaction and the cycle threshold (CT) is defined as the number of cycles required for the fluorescent signal to cross the threshold. CT levels are inversely proportional to the amount of target in the sample – the larger the amount of target sample the lower the number of PCR cycles required for the fluorescence to exceed the threshold, hence the smaller the CT value. CT values are used to calculate fold change.

Fold Change

The fold change between experimental condition and controls was calculated using relative quantification using the delta-delta CT method in the following manner:

First, the mean and standard deviations of the CT data were determined. Any values falling outside ±2 SD of the mean were excluded as outliers. Technical replicates were averaged before analysis and biological replicates averaged after delta-delta CT calculation.

The delta CT (ΔCT) was calculated using the geometric mean of the reference genes (ATIC and HIRIP5) using the formula:

$$\Delta CT = CT\ target\ gene\ of\ interest\ (GOI) - CT\ geometric\ mean\ reference\ genes\ (REF)$$

The SD of the ΔCT was calculated by using the SD of GOI and SD of REF in the following formula:

$$SD\ \Delta CT = \sqrt{SD\ GOI^2 + SD\ REF^2}$$

The delta-delta CT (ΔΔCT) was calculated using the ΔCT of experimental condition and the control condition in the following formula:

$$\Delta\Delta CT = \Delta CT\ experiment - \Delta CT\ control$$

The SD of the ΔΔCT is the same value as the SD for the ΔCT of the experimental condition

The fold change is calculated using the following formula

$$Fold\ Change = 2^{-\Delta\Delta CT}$$

Statistics

An unpaired t-test was used to compare the fold change values between the ruptured/healthy samples to determine whether or not the difference in expression was significant. The data was non parametric in its distribution (Shapiro-Wilk test for normality was applied to the data – see Appendix VI.) and subsequently non parametric versions of a t-test were run; the Mann Whitney test was performed as the samples are independent of each other.

The fold change was deemed to be significantly different in the experimental samples compared to control/normal if the resulting raw *p*-value was ≤0.05.

Results

Efficiency of primer/probe sets

The efficiency of the primers was evaluated after running standard curves. The efficiency of the reaction for the primers/probes used was on average 105%, with an R^2 value of ≥0.9. Primers and probes with low efficiency and/or poor amplification were removed from subsequent experiments – these were CDH7, SEMA5b, CDH19, PARP-15 and TNFα. For a full table of all primer efficiencies, R^2 values and the threshold values see Appendix VII.

Fold change

Any samples that had failed to amplify, or had a CT value that deviated ±2SD from the mean were excluded from analyses. The spread of variation of the remaining samples is plotted in Figure 7.2

There was sometimes large variation between each replicate sample, as indicated by the standard error of the mean (SEM) - values ranged from 4.89 - 8.02. This was due to no biological replicates because of the small amount of ligament available from each dog, particularly in the ruptured tissue group. The amount of RNA obtained from the extraction was also low in many cases.

Fold change and consequent *p*-values are shown in Table 7.2. They are graphically represented in Figure 7.3. Significant results ($p \leq 0.05$) are indicated by an asterix (*).

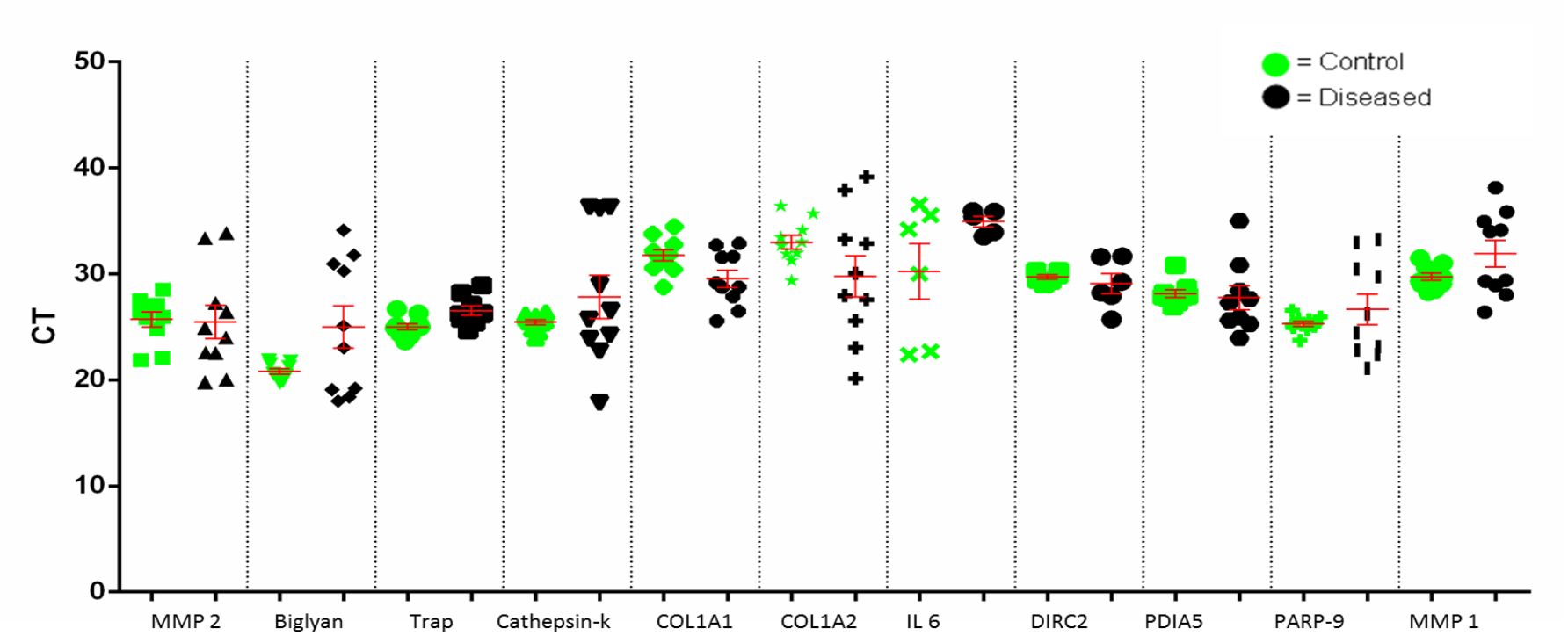


Figure 7.2 Gene expression changes in CCL rupture tissues. Clustering of the individual ruptured (n=10, five samples with two technical replicates) and healthy samples (n=10, five samples with two technical replicates) for each gene (minus samples ±2SD from mean). Error bars (standard error of the mean) are shown in red. CT = Cycle threshold

Table 7.2. Fold change values for ruptured vs. healthy experiment.

	Fold Change Ruptured	± SEM	*p* value (Mann Whitney)
MMP 2	9.19	6.85	0.1153
Biglycan	0.28	7.89	0.9677
Trap	3.67	5.11	0.0172*
Cathepsin K	1.47	8.02	0.4427
COL 1A1	36.58	5.40	0.0001*
COL 1A2	72.87	7.77	0.0172*
IL6	0.31	4.89	0.6547
DIRC2	12.14	5.29	0.0028*
PDIA5	2.86	6.74	0.1153
PARP 9	3.01	6.60	0.4429
MMP 1	1.71	6.19	0.4429

KEY: * - significant result (*p* ≤0.05). SEM = standard error of the mean.

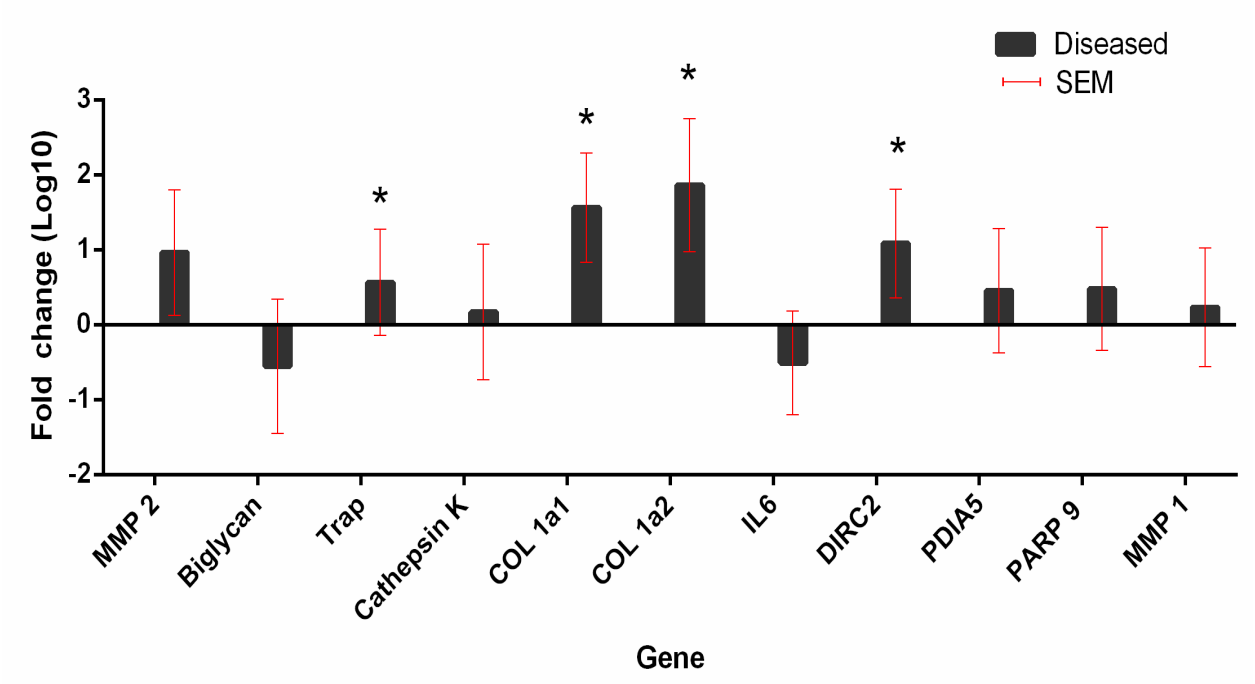


Figure 7.3. Gene expression in ruptured CCL. Fold change differences between healthy and ruptured CCL tissue. Asterix (*) = significant result (*p* ≤0.05). Data is shown on a log10 scale.

There were four genes that have significant differential expression levels between the ruptured and healthy tissues, TRAP (p=0.0172), COL1A1 (p=≤0.0001), COL1A2 (p=0.0172) and DIRC2 (p=0.0028). They all showed increased expression in the ruptured ligament tissues compared to the control tissue.

Discussion

Analysis of gene expression data is a very powerful way to study the pathogenesis of disease. It can help elucidate the underlying mechanisms of disease initiation and progression by investigating differential gene expression between disease or experimental conditions and controls [391]. This chapter aimed to determine the expression levels of select candidate genes identified as potentially having a role in the susceptibility or resistance to CCL rupture.

Q-PCR evaluates changes in expression at the time of tissue collection. Although the cases have ruptured their CCL, it may have been days/weeks/months or years after the initiation, diagnosis and subsequent surgery to collect the sample. Gene changes important for CCL rupture may have already occurred and may not have been present at the time of sampling and analysis. Indeed, changes in expression of genes that promote repair may have been prominent instead. Despite this significant reservation, it is still valid to investigate gene dysregulation in CCL rupture, as long as the above provisos are considered when data interpretation is concerned.

The data showed two potentially important genes which have significantly upregulated expression in the ruptured tissues. These are COL1A1 and COL1A2 with *p*-values of 0.0001 and 0.0172 respectively. Type-I collagen is the predominant collagen protein type found within ligaments [43] and its production is crucial in the ligament repair processes and during healing/scar formation [403, 404]. Upregulated expression of these genes in the ruptured ligaments compared to healthy ligaments denotes that a reparative process may be being attempted by the cells. The process of ligament healing after injury is thought to involve four phases; the inflammatory phase, an epiligamentous repair phase, a proliferative phase and then a remodelling phase [405]. Collagen is involved primarily in the proliferative and remodelling phases of this process [406, 407]. These results and conclusions are confirmed in other independent studies [408, 409] [50]; the authors proposed that the differential gene regulation was thought to be due to a reparative process within the ligament after the injury/rupture.

A study by Clements et al. (2008) investigated the expression profile of CCL in dogs, using normal and ruptured ligaments from Labrador Retrievers. They also compared normal Labrador Retriever CCLs to normal Greyhound CCLs (a breed that is less susceptible to CCL rupture) [13]. There were no differences between the high and low risk breeds, but 16 genes were differentially expressed between the Labrador Retrievers with CCL rupture and healthy Labrador Retrievers. COL1A2 was one of several matrix associated proteins (including COL5A1, COL3A1, lumican, aggrecan, tenascin-c) that were upregulated in the ruptured tissues. Our study not

only confirms the work of Clements et al., that COL1A1 and COL1A2 are upregulated in CCL ruptured tissue compared to normal CCL tissue samples, but also confirms their hypothesis that this is due to a remodelling and repair of the matrix of the after ligament injury. We did not wish to repeat every gene that Clements et al. studied (this thesis is an original piece of work), so were not able to confirm similar trends with any other genes. This section of the project was performed concurrently to the Sequenom candidate gene study which had not been analysed at the time of choosing genes. This is also why certain key candidate genes such as Collagen type V were not studied in this project, but they should be included in any subsequent expression studies.

We have showed that collagen mRNA levels are increased in ruptured compared to healthy CCLs (indicative of an attempted reparative process being carried out), but Young et al. 2011 [410] found that the actual collagen levels in CCL tissue were reduced. There are several flaws to directly comparing the two studies. Firstly, the collagen produced within the cells may be rapidly degraded before secretion to the ECM and therefore not able to be detected by the hydroxyproline assays used. Secondly, levels of mRNA do not necessarily correlate to protein synthesis and expression because of complex stages between transcription and translation [411]. It is feasible that both hypotheses are correct. Indeed lower levels of ECM proteins are consistent with a structurally altered ligament that is unable to withstand normal loading and therefore more susceptible to rupture.

Two other genes showed significant differences in gene expression between healthy and ruptured CCLs. These were tartrate resistant acid phosphatase (TRAP) and disrupted in renal carcinoma-2 (DIRC2), with *p*-values of 0.0172 and 0.0028 respectively. TRAP is of interest, as it has been found, along with Cathepsin-k and some other matrix-metalloproteinases, to be expressed during chronic inflammation and it induces leucocyte mediated collagenolysis [412-414]. TRAP is secreted by osteoclasts and macrophages [415, 416] and can initiate collagenolysis by forming free-radicals and reactive oxygen species (ROS). These are highly toxic to cells and can cause degradation and loss of structure/strength of the collagens (and therefore ligaments) [414]. Over expression of TRAP has been shown to increase amounts of ROS, which may subsequently increase the amount of collagen destruction [413, 417].

DIRC2 does not have such an obvious role in the pathology or susceptibility to CCL rupture. DIRC2 is a member of the major facilitator superfamily (MFS) transporters [418]. They are a group of single polypeptides which act as secondary carriers and are capable of transporting small solutes across membranes in response to chemiosmotic ion gradients [419]. DIRC2 has been found localised in many organelles including lysosomal membranes. Lysosomal enzymes can degrade collagen fibres in periodontal ligaments [420], so it is feasible that if DIRC2 is upregulated then the enzymes within the lysosome are released more abundantly and cause degradation and loss of structure/strength of the collagens and subsequently the ligaments.

The upregulation of TRAP and DIRC2 seen in ruptured CCLs compared to healthy CCLs indicates that degradation is occurring within the ligament simultaneously with the attempted reparative process. This is conclusive with the limited healing potential seen in a CCL after rupture [421, 422]. Although the limited healing is in part attributable to reduced vascular response [423], it may also be due to this protease degradation.

The weakness of this Q-PCR experiment was that the breed, age, sex, neutering status or weight of the dogs was not collected. Also, more importantly, the history of trauma pre-initiating the rupture was unknown. This part of the project was focussed on investigating the disease process and was not concerned with any inherited genetic components of the condition; hence, the breed of dog was less important. CCL rupture-associated changes in expression are likely to be the same regardless of breed and so this should not impact the biological significance of the results.

Only very small amounts of RNA are extractable from CCL tissue and given the high sample consumption of Q-PCR experiments this limited the number of genes that could be studied by this approach. For this reason, the number of reference genes used, was reduced from three to two (C7orf28B was dropped from the list because it was reported as being the least stably expressed [13]) and only the most efficient Q-PCR reactions were selected for inclusion in the experiments. The small amount of RNA also meant that only a small number of replicate samples could be used in

the experiment (two per sample). The standard errors were large due to only having duplicate samples (rather than triplicates, which is the standard for Q-PCR experiments) the error bars represent this on the fold change chart (Figure 7.3).

Alternative methods of studying gene expression involve the use of microarray chips to examine a wide selection of mRNA sequences that have been hybridised to a chip [393]. Differential labelling of the case and control RNA samples can identify genes that haven been up or down regulated by measuring the intensity of the colour on the microarray [424]. This method uses an unbiased approach to identify novel genes that may be up or down regulated in a disease process. Data analysis is lengthy and requires validation by Q-PCR to avoid false positive data. Due to budgetary and time constraints, this approach was impractical and is why gene expression was evaluated by Q-PCR instead.

Innovative, next generation sequencing approaches have recently been developed. These sequence the entire RNA library (transcriptome) of individual tissue samples to generate information pertaining to the differential gene expression between cases and controls [425, 426]. This technology was extremely new when conducting this project and as such the cost implication for running this would have been prohibitively expensive. This would be a good way to follow up the results of this experiment, to confirm and replicate the gene expression changes seen. In addition this approach may identify novel genes that may not have been identified in the current literature as candidate genes.

In conclusion, this study revealed that collagen type I genes (COL1A1 and COL1A2), as well as TRAP and DIRC2 were significantly upregulated in ruptured versus healthy CCL tissue samples. This indicates that repair and proliferation is being attempted by the ligaments after the initial injury, but that a degradative process is also still occurring. The data shown here most likely represents gene changes that have occurred after the injury and not initiating factors in ligament rupture. It is impossible to predict when a dog will rupture its ligament (or the stage of rupture), so investigating pathological changes in expression before rupture occurs is not feasible. Our approach is best viewed as a way to help elucidate changes that occur after injury. Treatments can be targeted to the effects, thereby hopefully limiting the damage caused to the contra-lateral ligament.

Chapter 8. Quantitative gene expression analysis in an *in-vitro* model of CCL rupture

Introduction

In previous chapters, the CCL rupture disease process and genetic susceptibility factors were investigated using case control association studies. Such studies require large numbers of both case and control samples to provide sufficient statistical power. For gene expression work, there is a finite amount of ligament tissue available, which hampers that approach to disease investigations. A laboratory model of CCL rupture which could be simulated in a controlled laboratory situation may facilitate investigations into the causes and treatments of diseases and would reduce the need for laborious collection of samples. Various methods of disease modelling are available from the more traditional *in-vitro* primary cell cultures [427, 428], isogenic/immortalised cell line models [429, 430] and *in-vivo* animal and transgenic models [431, 432], to the more recently developed 3D-cell culture technologies [433].

Cell culture models (primary or secondary/isogenic cell lines) have been used as *in-vitro* models to map disease and tumour progression in many cancers such as cervical cancer (HeLa cells [430]), breast cancer (MCF-7 cells [429]), colorectal cancer (HCT-116 cells [430]) and in other diseases such as cystic fibrosis (ΔF508 cells [434]) and Alzheimer's disease (IMR-32 neuroblastoma cells [435]). The advantages of using cell lines as models for disease are that they are cost effective, relatively easy to maintain and are reproducible under the defined conditions. Furthermore they reduce the number of animals that are needed in research. Disadvantages are that the cells may not behave like they would *in-situ* because the morphological

and spatial development may be different *in-vitro* relative to the disease process being modelled [436]. Such variation can be minimised by using primary cultures that have been taken directly from the tissue of interest in preference to secondary (immortal) cell lines. Cultures should be monitored for senescence and/or transformation due to genetic drift and should be discarded before this happens [437]. Alternatively, animals models can be developed (*in-vivo* research), but this can be costly [438].

In-vivo transection of the CCL in animals has been used to as an experimental laboratory model to study the developmental process of osteoarthritis associated with CCL rupture [432, 439, 440]. This model is considered to be suitable for studying downstream effects after the initial CCL rupture, but not the early disease changes which occur before the rupture. There are few verifiable models for early stage ligament degeneration. Lopez et al. [441] developed a model that simulates partial rupture/damage of the CCL, which will then progress to complete CCL rupture.

Surrogate platforms may only have limited clinical relevance, as cell lines and surgically induced animal models do not always fully represent the normal disease process that is seen in the human or dog [442]. A surgically induced transection (or otherwise tampered with) ligament may not behave in the same way as a naturally stretched ligament and subsequent degeneration and rupturing may be altered as a result. Despite their drawbacks, these laboratory models of disease can enable

proof or denial of physiological processes that may increase disease susceptibility, initiation and progression or can be useful for testing therapeutic agents to treat or prevent a disease.

Models that effectively model a disease process and enable a reduction or replacement of animals used for research are very useful [443]. They will allow more general laboratory work to be carried out within a more ethical framework, without the restrictions of home office licences and animal husbandry etc. Currently, there are no *in-vitro* models for CCL rupture and it is for this reason we aimed to try and develop an *in-vitro* model for the condition.

Tissue culture modelling can be accomplished by three main methods: cell culture, explant culture or organ/biopsy culture [444-446]. Primary cell culturing is conducted via enzymatic or mechanical disruption of a tissue, to disperse the cells into a suspension and cells are then grown as a monolayer culture [447]. Explant culture involves a fragment of tissue being placed directly onto a culture plate. During culture with appropriate media, the cells slowly migrate and outgrow from the tissue to form a monolayer culture [444, 448]. This method is time consuming and carries a high risk of bacterial contamination from the original tissue sample. This method was attempted but was not taken forward in the project as no cell migration was seen within a week of culturing and after two weeks only a small number of cells had migrated out. It was not seen as a viable option for monolayer culturing within the project's time frame. An organ biopsy is similar to an explant

culture but the experiment is carried out on the piece of tissue itself [446]. The histological and cell architecture is retained, but the reproducibility is low and intersample variation is high, making this method redundant in most studies. Microscopic monitoring of the cells daily to check for signs of contamination is important for all of the methods.

TNFα is a pro-inflammatory cytokine which is commonly used in a laboratory environment to model chronic inflammatory and autoimmune disorders such as osteo- and rheumatoid arthritis, Crohn's disease and inflammatory bowel diseases [428, 449-452]. Levels of the TNFα have been shown to be elevated in ruptured ACLs in man [453, 454]. The aim of this chapter was to compare and evaluate gene expression in cultured ligamentocytes in the presence or absence of TNFα, to determine whether TNFα induced ligamentocyte fibroblasts were valid as an *in-vitro* model for CCL rupture.

Methods

Sample collection

Samples were collected by sharp dissection of CCL ligament from five cadaveric dogs that had been recently euthanased for reasons other than degenerative joint diseases and which had been donated to the University for teaching purposes. The tissue was placed directly into sterile culture medium (as described in the Chapter

2), to obtain primary ligamentocyte cell lines. The breed could not be pre-selected, but the majority of cadavers were of the Staffordshire Bull Terrier type.

Sample preparation

Once collected, the tissue was immediately processed for tissue culturing via enzymatic disruption as described previously, to generate ligamentocyte monolayer cultures (see tissue culture of ligament cells in materials and methods – Chapter 2). Once confluent, the cells were cultured for 24 hours with or without recombinant canine TNFα (10ng/ml), as described below.

Ligamentocyte stimulation with TNFα

Recombinant canine TNFα (R&D systems, Minneapolis, US) was used. This was reconstituted to a stock volume of 100ug/ml in sterile PBS with 0.1% bovine serum albumin. When used in stimulation experiments it was used at a working concentration of 10ng/ml. The cells were exposed to TNFα for 24 hours (as described in the monolayer section detailed in materials and methods - Chapter 2) prior to assessment of changes in gene expression.

Experiments were carried out in triplicate on three independent days to allow for technical and biological replicates – see Figure 8.1. Cells in this project were never taken beyond passage 8 to avoid cellular transformation or senescence.

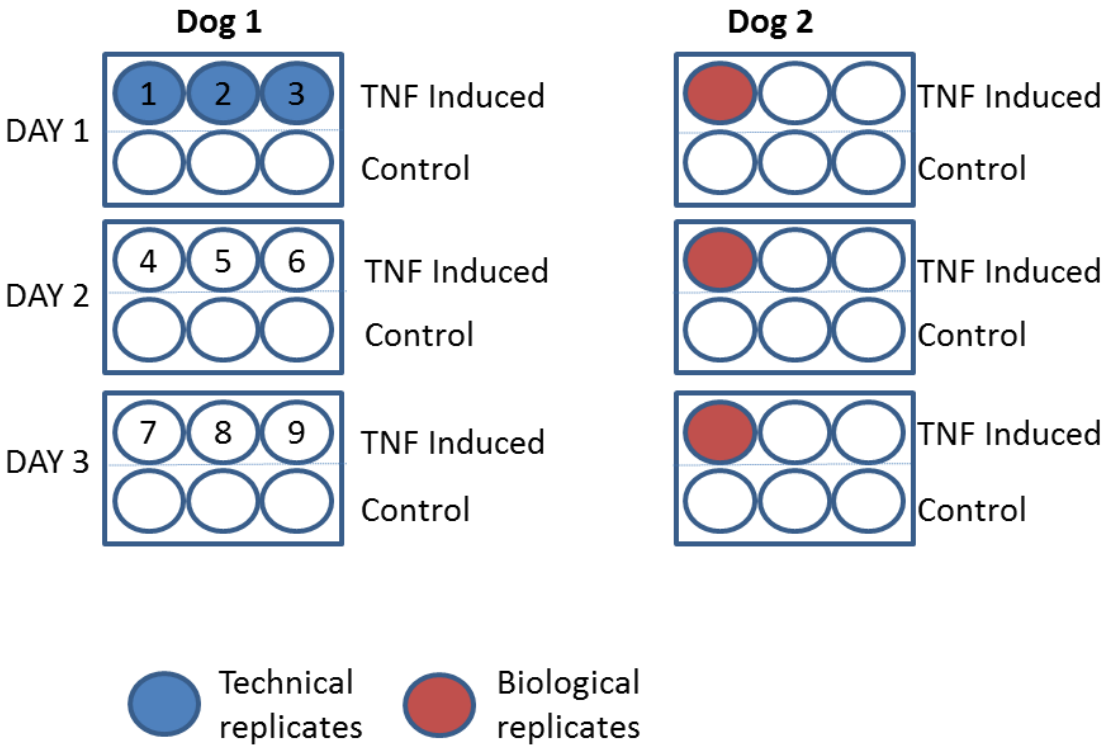


Figure 8.1. Plate setup for the tissue culture induced/control expression experiment, shown for two dogs. Colours indicate the biological (red circles) and technical replicates (blue circles) which were used in each experiment.

CCL ligamentocytes were cultured in 6-well plates, either with or without TNFα (10ng/ml) as shown above. Each experiment was executed in triplicate on three separate days, for five dogs.

For each dog there are nine induced samples (numbered) and nine control samples. For the five dogs used, there were a total of 45 TNFα induced samples and 45 control samples to be analysed by Q-PCR. Technical replicates from each day (1-3, 4-6, 7-9) were averaged before delta-delta CT calculations were performed (15 induced values - three from each of the five dogs, and 15 control values - three from each of the five dogs) to determine gene expression changes between normal and TNFα induced ligamentocytes.

RNA Extraction

Post culture, the adherent cells were lysed directly on the plate with Trizol before continuing with RNA extraction. The RNA extraction protocol was as described in Chapter 7.

cDNA synthesis

Reverse transcription of RNA was performed using oligo dT primers and Superscript II reverse transcriptase (Invitrogen, California, US) according to the manufacturer's instructions. Further details are in Chapter 7.

Design of primes and probes

Target sequence and reference gene selection was performed as in Chapter 7, using the same genes and primer sets.

Primer probe validation

The primers had been found to be efficient from the experiments reported in Chapter 7.

Q-PCR

Each Q-PCR reaction was carried out in triplicate using the Applied Biosystems 7900HT PCR machine in 384 well formats, as detailed in Chapter 7. A 1:10 dilution

of each cDNA was used for experiments in this project, as this concentration of sample produced the best amplification curves during primer testing and showed no inhibition of reaction.

Data analysis

Fold Change

The fold change between experimental condition and controls was calculated using the delta-delta CT method as detailed previously in Chapter 7.

Statistics

A paired t-test was used to compare the fold change values between induced/control samples to determine whether or not the difference in expression was significant. The data was non parametric in its distribution (Shapiro-Wilk test for normality was applied to the data – see Appendix VI.) so non-parametric versions of the t-test were used; The Wilcoxon test was performed as the samples were paired. The fold change was deemed to be significantly different between the experimental cultures compared to control cultures if the resulting raw *p*-value was ≤0.05.

Results

Fold Change

The variation between each individual replicate of induced/control samples was minimal; once samples which varied ±2 SD from the mean were removed, the clustering of the 45 individual replicate samples for each gene was evaluated and the results presented in Figure 8.2

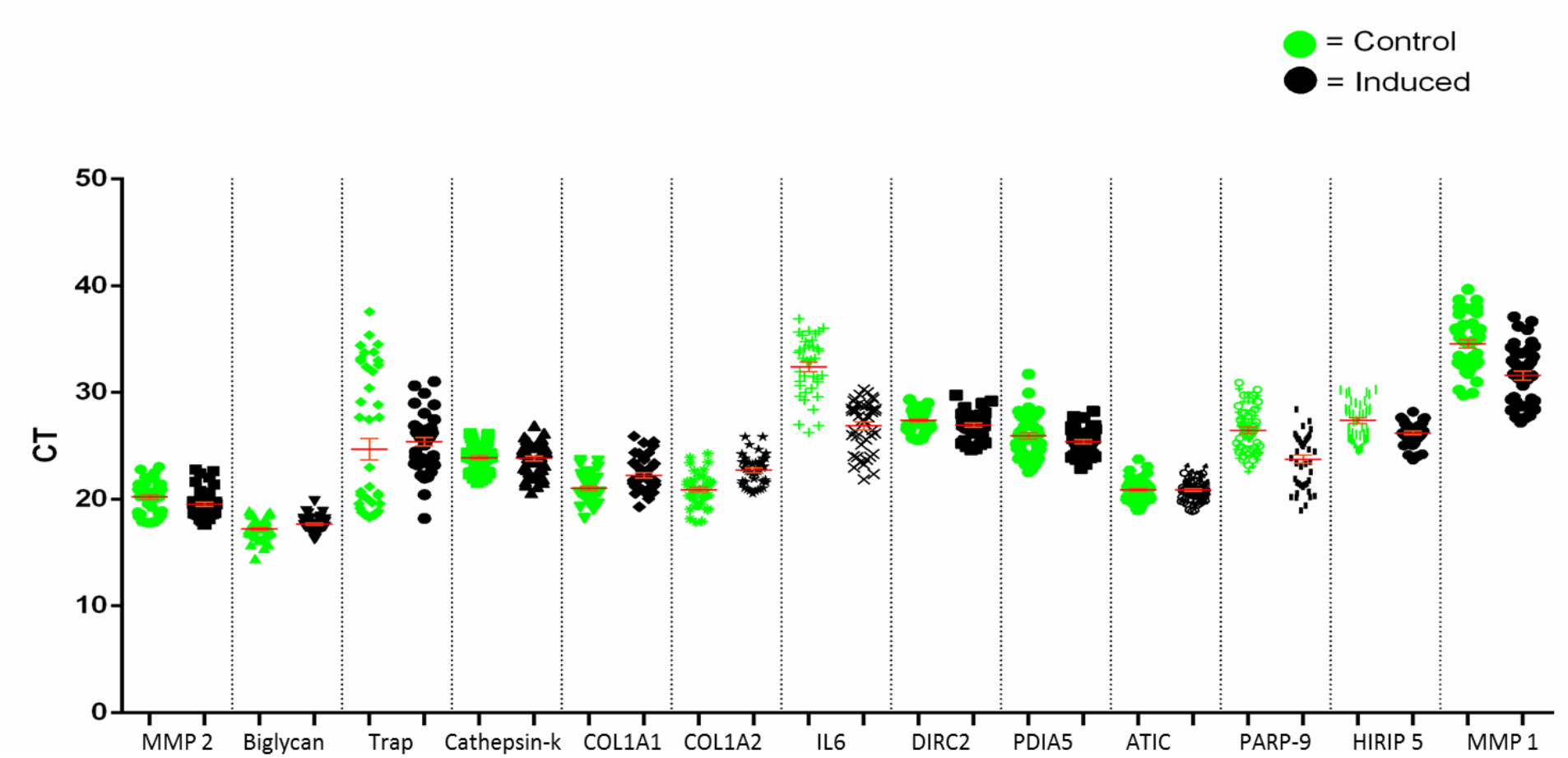


Figure 8.2. Gene expression in canine ligamentocytes induced by TNFα. Clustering of the induced (n=45) and control (n=45) samples for each gene (minus samples ±2SD from mean). Error bars (Standard error of the mean) are shown in red.

The fold changes and the *p*-values for each gene are shown in Table 8.1 and are graphically represented in a grouped bar chart (Figure 8.3). Significant results (*p*-value ≤0.05) are indicated by an asterix (*).

Table 8.1. Fold change values for CCL cells that had been cultured with 10ng/ml TNFα for 24hrs versus normal, control cells.

Gene	Fold change Induced	± SEM	*p*-value (Wilcoxon)
MMP 2	1.33	1.69	0.0640
Biglycan	0.56	1.40	0.0180*
Trap	0.70	2.61	0.3780
Cathepsin K	0.85	2.16	0.7610
COL 1A1	0.33	1.80	0.0130*
COL 1A2	0.28	1.81	0.0160*
IL6	45.78	2.59	0.0001*
DIRC2	1.12	1.70	0.2010
PDIA5	1.09	1.64	0.2080
PARP 9	5.81	2.99	0.0001*
MMP 1	6.57	2.90	0.0009*

KEY: * - significant result (*p*-value ≤0.05), SEM = standard error of the mean

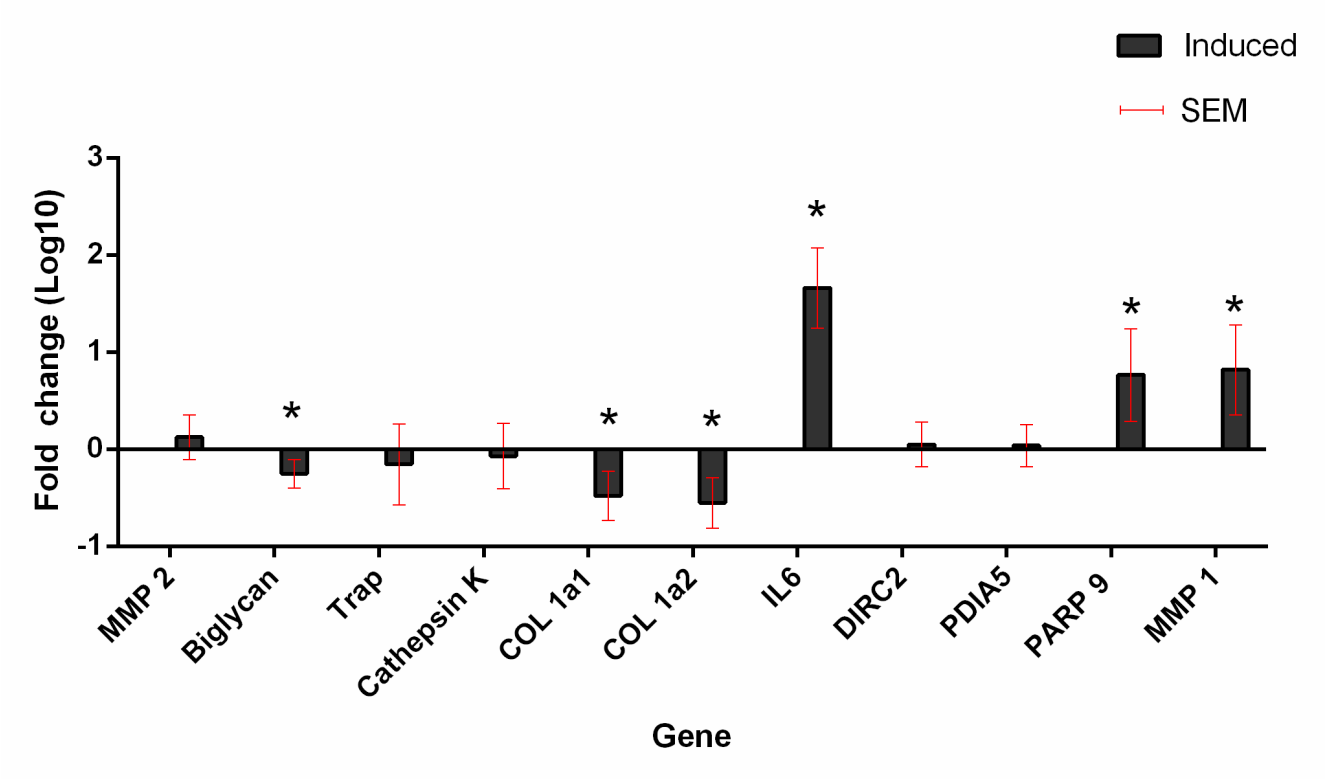


Figure 8.3. Gene expression in cultured CCL cells: Fold Change between normal cultured CCL cells and CCL cells that had been cultured with 10ng/ml TNFα. * = significant result (*p*-value ≤0.05)

Six genes showed significant differential expression between the induced ligamentocytes and the normal cells, these were Biglycan (p=0.0180), COL1A1 (p=0.0130), COL1A2 (p=0.0160) which had reduced expression with TNFα stimulation and IL6 (p=0.0001), PARP 9 (p=0.0001) and MMP-1 (p=0.0009) which were upregulated by TNFα stimulation. The expression of IL-6 after TNFα induction was 45 times that of control cultures.

Discussion

We evaluated gene expression changes in cultured ligamentocytes in the presence or absence of TNFα. This method was used not only to determine the gene up/downregulation in response to TNFα, but also to determine if this system was a valid *in-vitro* model for CCL rupture. Readily identifiable downstream inflammatory or degradative molecules are induced when cell cultures (in a variety of tissues) are stimulated with TNFα [449, 455]. This is seen here with the ligamentocyte cultures that are induced to over express PARP 9, MMP-1 and most strongly, IL-6. Similarly, the ligamentocytes cultured with TNFα had reduced expression of extracellular matrix molecules Biglycan, COL1A1 and COL1A2.

Upregulation of IL-6 is highly relevant as it confirms that the TNFα used in this project was viable and that the experiment induced a pro-inflammatory response. When cells are challenged with TNFα it causes the release of further inflammatory cytokines such as IL-6 [456] and as such the significant upregulation seen here, is a direct result of the TNFα induction.

MMP-1 was also upregulated significantly in the TNFα-treated ligamentous fibroblasts. Matrix metalloproteinases are commonly involved in many diseases where inflammation is present [457, 458] and have a key role in tissue remodelling and degradation [332]. Given that the cells have been challenged with TNFα (an inflammatory mediator), it indicates that the cells are responding to TNFα stimulation in both an inflammatory and degradative manner.

PARP-9 (poly ADP ribose-9) catalyses the post-translational modification of proteins by the addition of multiple ADP-ribose moieties (ADP ribosylation) and may be induced by cellular processes involving inflammatory mediators, such as TNFα, macrophages, IL-1, IL-6 and nitrite molecules (NO_2) [459]. PARP can also interact with other pro-inflammatory mediators such as NF-κB. As such it will have an inflammatory effect on cells [460] via these other pathways, which are also likely to be activated/induced under inflammatory conditions. Upregulation of this gene is therefore also consistent with the inflammation caused by the TNFα stimulation.

Biglycan, a small leucine-rich proteoglycan, is functional in the formation of connective tissue and mineralised tissue (bone) matrices, due to its mineralisation deposition properties [461]. It also has a role in elastic fibre formation (during the elastogenic phase) in fibres such as tropoelastin and fibrillin-containing microfibrils [462] as well as in tissue remodelling [463]. As the TNFα-treated cells are undergoing an inflammatory response, the ability of the cells to form matrices is hindered – hence the downregulation of this gene. This downregulation is consistent with other research (in man) that has found that stimulation of fibroblasts with TNFα, decreases expression of Biglycan [464-466].

The downregulation of the COL1A1 and COL1A2 gene expression is also consistent with the degradative response from the TNFα stimulation, which hinders repair and proliferation of the matrix.

We aimed to determine whether TNFα was a suitable model for studying CCL rupture in the laboratory. Expression data from the induced ligamentocytes compared to normal ligamentocytes varied considerably from the expression of ruptured CCL tissue versus healthy tissue (as detailed in Chapter 7). No genes were simultaneously up or downregulated to the same extent in both studies. Ruptured CCL tissue is considered to be abnormal, and changes relating to CCL rupture may have already occurred, therefore comparative investigations into the disease process using this method may not be suitable. The TNFα model (using healthy CCL tissue) may instead be used as a successful model to investigate the CCL rupture disease process.

TNFα can be used legitimately to study inflammation in ligamentocytes (and other cells), and the early stages of CCL rupture may be modelled using this cytokine via this model. However, as factors other than inflammatory process are considered to be involved in the initiation and progression of CCL rupture it may also be deemed unsuitable. CCL rupture is considered a complex disease with many environmental and genetic factors playing a role in the initiation and progression of the disease [36], as a result It is very difficult to study early stage CCL rupture in animals. Furthermore, it is important to consider the nature of the cells in the test system. It is difficult to know in any meaningful way how much they can represent the function of ligamentocytes and whole ligament *in-vivo,* as the morphological and spatial characteristics will be different *in-vitro* compared to *in-vivo* situations.

As the factors involved in early stages of CCL rupture are still largely unknown, studying this process *in-vitro* will be difficult and *in-vivo* models (using rabbits etc) are still primarily the best model to investigate changes that occur after CCL transection in a laboratory environment.

In conclusion, we showed that TNFα may be suitable as a laboratory model to study CCL failure as an *in-vitro* model, and will be useful to identify individual pathways that may be important after CCL failure in dogs.

Chapter 9.

General Discussion & Conclusions

Introduction

This thesis aimed to investigate the genetic basis of cranial cruciate ligament (CCL) rupture in Newfoundland dogs, but also considered other susceptible breeds and investigated downstream events from genomic variations. By using a whole genome genotyping approach (GWAS) to look for SNP associations, a large amount of data was generated that was useful for identifying a list of key genes, considered likely to have some involvement in CCL rupture susceptibility. Subsequent studies were able to concentrate on specific candidate genes and their potential role and expression in both diseased tissues and in a potential laboratory model for CCL rupture.

Stratification

A primary confounder in our association studies was that the genetic structure of the two main Newfoundland groups, Europe and North America were significantly different. Stratification of genetic data between different ethnicities in humans [467] as well as in different breeds of dog [170, 231] is well recognised and is expected given the wide variation in shape, size and colour between different breeds of dog but is not necessarily expected between dogs of the same breed. Stratification within the Newfoundland breed has not been documented before. To avoid spurious associations within datasets it is essential to use cases and controls from the same breeds and location. However, to increase the power to detect disease associations in studies, researchers have often included dogs from the same breed but other locations with little regard to how this potential sub-structure

would affect underlying stratification in the datasets i.e. assuming same breed, same genetics approach [231]. Stratification within the same breed of dog from different geographical locations has been reported in recent years [177, 230] and as such, it is now commonplace to assess stratification before carrying out analyses to ensure no underlying bias will be present in the data.

GWAS results

Analysis of the GWAS data accounting for the stratification using a CMH stratified by country analysis, identified three main chromosomal regions associated with CCL rupture on chromosomes 1, 3 and 33. The two most associated SNPs from the analysis (BICF2S23133650 and BICF2P1311062) reached genome wide significance and were significant after testing for multiple correction. These two SNPs were both on chromosome 33 in the gene SEMA5B and were found in a statistically significant haplotype block with each other.

The key SNPs identified by the GWAS in the Newfoundlands were subsequently considered in a validation study using Sequenom iPlex technology in a larger cohort of Newfoundlands and in three other highly susceptible breeds (Labrador Retrievers, Rottweilers and Staffordshire Bull Terriers). The regions of association on chromosomes 1, 3 and 33 were confirmed in the Newfoundlands but did not validate significantly in the other breeds. Interesting associations with CCL rupture were found in genes involved in neurological pathways such as SORCS2 (Chr 3) and SEMA5B (Chr 33) indicating that mechanotransduction, neurological and

neuromuscular pathways may play an important role in the pathogenesis and susceptibility to CCL rupture; a novel process that hasn't previously been documented with regards to CCL rupture susceptibility in dogs.

Candidate gene results

The "traditional" candidate gene studies investigate genes that may have a potential functional involvement in CCL rupture initiation or progression. Across all of the CCL rupture-susceptible breeds studied, the data suggests that mutations in the collagen genes have the most significant role in CCL rupture susceptibility, with 11 of the 17 most significant associations from the meta-analysis being within this gene family. This was expected as collagen is the main component of ligament structure [37]. The collagen and collagen-associated genes have a prime function in assembly and formation of collagen fibrils that are important in the integrity of the ligament extracellular matrix and the ligaments strength and stability [72, 301, 311, 318, 329, 359]. Other CCL rupture associated mutations within the different breeds were also found in Aggrecan (ACAN), Opticin (OPTC) and Latent transforming growth factor beta 2 (LTBP2); key proteins involved in the ECM formation and maintenance [66, 305, 359]. Mutations within these genes may compromise the ligament's strength and composition leading to an increased risk of CCL rupture for dogs with the associated mutations. It is clear therefore, that mutations affecting the strength and structure of the ligament and its composition may be important in CCL rupture susceptibility.

Changes in gene expression

Having investigated gene expression changes between diseased and healthy tissue samples, we found that there were significant differential expressions seen in certain genes. Collagen type I genes (COL1A1 and COL1A2) as well as TRAP and DIRC2 were found to be significantly up-regulated in CCL rupture tissues. This suggests that the ligament is undergoing proliferation as part of a repair process following the initial rupture, but that a degradative process is still in progress which hinders the actual repair of the ligament. These changes, although unlikely to be indicative of pre-rupture changes, are important for understanding changes that occur after injury and may be useful for developing treatment regimens to minimise further damage to joint and/or contralateral ligament.

We also showed that a TNFα induced cell line model may be a suitable system to study CCL rupture *in-vitro*, although it is more useful to mimic the inflammatory changes seen after the injury and hence be useful in studying the mechanisms of new treatment regimens rather than the early changes that precede CCL rupture.

DLA association

Having assessed the main loci of the DLA region, for association with CCL rupture, we showed no association of the DLA region with the condition. We did identify differences in haplotype frequencies between the Newfoundlands from North America and those from Europe. This confirms the genetic stratification within the Newfoundland breed that was seen within the GWAS data. Without further

replication in a larger cohort of Newfoundlands from both North America and Europe this DLA result cannot be confirmed, but it does support other studies indicating that there is no association of DLA regions with CCL rupture [379].

Conclusions

These data have shown that there are complex multigenic interactions involved in CCL rupture in Newfoundland dogs. There are likely to be different variations involved in susceptibility between different breeds of dog. Furthermore, the involvement of environmental factors is highly likely, making the determination of risk for CCL rupture even more complicated.

This research has made a small but significant understanding in CCL research. It is gratifying that newly developed technology has potentially enabled the identification of key genes that may be important in CCL rupture susceptibility. These associations obviously require further investigation and analysis in Newfoundlands and in other dog breeds with CCL rupture, as well as in humans where this is also a major clinical problem. These findings are only a small part of a much bigger picture, but they may be used to help develop advances in treatment and/or prevention of CCL rupture and limit damage to contralateral ligaments.

In summary, to take this work forward, the associated regions and genes associated with CCL rupture need to be further investigated in larger numbers of dogs

(samples and breeds) using more up-to-date technologies such as targeted next generation sequencing or transcriptomic approaches such as RNAseq. These technologies would allow the regions/genes to be examined in greater detail and could potentially identify the exact causative mutations that are involved in CCL rupture. Further work involving functional analysis, for example knock-out or transgenic models, could also be considered to determine the effect of the mutations in a mouse model. Once the causative mutations have been determined in each breed, a screening test can be developed and used, to allow an informed breeding program to be instigated that should reduce the incidence of the disease in dogs. This would be of great benefit to dog owners, breeders and of course, to the dogs themselves!

Chapter 10.

References

1. Wilke, V, D Robinson, R Evans, M Rothschild and M Conzemius, *Estimate of the annual economic impact of treatment of cranial cruciate ligament injury in dogs in the United States.*, J Am Vet Med Assoc, 2005, 227(10): 1604
2. Wilke, V, S Zhang, R Evans, M Conzemius and M Rothschild, *Identification of chromosomal regions associated with cranial cruciate ligament rupture in a population of Newfoundlands.*, Am J Vet Res, 2009, 70(8): 1013
3. Harasen, G, *The latest in orthopaedics*, Can. Vet. J., 2009,
4. Hayashi, K, J Frank, C Dubinsky, H Zhengling, M Markel, et al., *Histologic changes in ruptured canine cranial cruciate ligament.*, Vet Surg, 2003, 32(3): 269
5. Hayashi, K, P A Manley and P Muir, *Cranial Cruciate Ligament Pathophysiology in Dogs With Cruciate Disease: A Review*, J Am Anim Hosp Assoc, 2004, 40(5): 385
6. Wilke, V L, M G Conzemius, B P Kinghorn, P E Macrossan, W Cai, et al., *Inheritance of rupture of the cranial cruciate ligament in Newfoundlands*, J Am Vet Med Assoc, 2006, 228(1): 61
7. Dyce, K M, W O Sack and C J G Wensing, *Textbook of Veterinary Anatomy, 4th Edition*2010.
8. Eleswarapu, S V, D J Responte and K A Athanasiou, *Tensile Properties, Collagen Content, and Crosslinks in Connective Tissues of the Immature Knee Joint*, PLoS ONE, 2011, 6(10): e26178
9. O'Donoghue, D H, G R Frank, G L Jeter, W Johnson, J W Zeiders, et al., *Repair and Reconstruction of the Anterior Cruciate Ligament in Dogs factors influencing long-term results*, The Journal of Bone & Joint Surgery, 1971, 53(4): 710
10. Phisitkul, P, S L James, B R Wolf and A Amendola, *MCL injuries of the knee: current concepts review*, The Iowa orthopaedic journal, 2006, 26(77
11. Arnoczky, S P and J L Marshall, *The cruciate ligaments of the canine stifle: an anatomical and functional analysis*, Am J Vet Res, 1977, 38(11): 1807
12. Miyaki, S, K-i Tezuka, T Ushida, T Akimoto, H Akaogi, et al., *Identification of membrane and secreted proteins in anterior cruciate ligament derived cells using "signal-sequence-trap", a retrovirus-mediated expression screening method*, Materials Science and Engineering: C, 2004, 24(3): 397
13. Clements, D N, S D Carter, J F Innes, W E Ollier and P J Day, *Gene expression profiling of normal and ruptured canine anterior cruciate ligaments*, Osteoarthritis Cartilage, 2008, 16(2): 195
14. Bennett, D, B Tennant, D G Lewis, J Baughan, C May, et al., *A reappraisal of anterior cruciate ligament disease in the dog*, J Small Anim Pract, 1988, 29(5): 275
15. Harasen, G, *Diagnosing rupture of the cranial cruciate ligament.*, Can. Vet. J., 2002, 43(6): 475
16. Harasen, G, *Canine cranial cruciate ligament rupture in profile.*, Can. Vet. J., 2003, 44(10): 845
17. Han, S, H Cheon, H Cho, J Kim, J H Kang, et al., *Evaluation of partial cranial cruciate ligament rupture with positive contrast computed tomographic arthrography in dogs*, Journal of veterinary science, 2008, 9(4): 395

18. Innes, J F, *Management of cruciate ligament rupture: what is 'best practice'?*, Veterinary Ireland Journal 2012, 2(1):
19. Slocum, B and T D Slocum, *Tibial plateau leveling osteotomy for repair of cranial cruciate ligament rupture in the canine*, Vet. Clin. North Am. Small Anim. Pract., 1993, 23(4): 777
20. Lafaver, S, N A Miller, W P Stubbs, R A Taylor and R J Boudrieau, *Tibial tuberosity advancement for stabilization of the canine cranial cruciate ligament-deficient stifle joint: surgical technique, early results, and complications in 101 dogs*, Vet Surg, 2007, 36(6): 573
21. DeAngelis, M and R E Lau, *A lateral retinacular imbrication technique for the surgical correction of anterior cruciate ligament rupture in the dog*, J Am Vet Med Assoc, 1970, 157(1): 79
22. Arnoczky, S P, G B Tarvin, J L Marshall and B Saltzman, *The over-the-top procedure: a technique for anterior cruciate ligament substitution in the dog [Surgery]*, Journal-American Animal Hospital Association, 1979, 15(
23. *Pet Food Manufacturers' Association (PFMA)*, http://www.pfma.org.uk/statistics/, July 2013
24. The UK Kennel Club, *The Kennel Club*, http://www.thekennelclub.org.uk/, July 2013
25. Innes, J, D Bacon, C Lynch and A Pollard, *Long-term outcome of surgery for dogs with cranial cruciate ligament deficiency.*, Vet. Rec., 2000, 147(12): 325
26. Muir, P, S L Schaefer, P A Manley, J P Svaren, W E Oldenhoff, et al., *Expression of immune response genes in the stifle joint of dogs with oligoarthritis and degenerative cranial cruciate ligament rupture*, Vet Immunol Immunopathol, 2007, 119(3-4): 214
27. Duval, J M, S C Budsberg, G L Flo and J L Sammarco, *Breed, sex, and body weight as risk factors for rupture of the cranial cruciate ligament in young dogs.*, J Am Vet Med Assoc, 1999, 215(6): 811
28. Witsberger, T, J Villamil, L Schultz, A Hahn and J Cook, *Prevalence of and risk factors for hip dysplasia and cranial cruciate ligament deficiency in dogs.*, J Am Vet Med Assoc, 2008, 232(12): 1818
29. Harasen, G, *Canine cranial cruciate ligament rupture in profile: 2002-2007.*, Can. Vet. J., 2008, 49(2): 193
30. Slauterbeck, J, C Clevenger, W Lundberg and D M Burchfield, *Estrogen level alters the failure load of the rabbit anterior cruciate ligament*, J Orthop Res, 1999, 17(3): 405
31. Erne, J B, R L Goring, F A Kennedy and W C Schoenborn, *Prevalence of lymphoplasmacytic synovitis in dogs with naturally occurring cranial cruciate ligament rupture*, J Am Vet Med Assoc, 2009, 235(4): 386
32. Cook, J L, *Cranial cruciate ligament disease in dogs: biology versus biomechanics.*, Vet Surg, 2010, 39(3): 270
33. Doverspike, M, P B Vasseur, M F Harb and C M Walls, *Contralateral Cranial Cruciate Ligament rupture – Incidence in 114 dogs* J Am Anim Hosp Assoc, 1993, 29(2): 167
34. Flynn, R K, C L Pedersen, T B Birmingham, A Kirkley, D Jackowski, et al., *The familial predisposition toward tearing the anterior cruciate ligament: a case control study*, Am J Sports Med, 2005, 33(1): 23

35. Harasen, G, *A retrospective study of 165 cases of rupture of the canine cranial cruciate ligament.*, Can. Vet. J., 1995, 36(4): 250
36. Jerram, R M and A M Walker, *Cranial cruciate ligament injury in the dog: pathophysiology, diagnosis and treatment.*, N Z Vet J, 2003, 51(4): 149
37. de Rooster, H, T de Bruin and H van Bree, *Morphologic and functional features of the canine cruciate ligaments*, Vet Surg, 2006, 35(8): 769
38. Riemann, B L and S M Lephart, *The Sensorimotor System, Part II: The Role of Proprioception in Motor Control and Functional Joint Stability*, J Athl Train, 2002, 37(1): 80
39. Riemann, B L and S M Lephart, *The sensorimotor system, part I: the physiologic basis of functional joint stability*, J Athl Train, 2002, 37(1): 71
40. Menkes, C J, M Renoux, S Laoussadi, A Mauborgne, J Bruxelle, et al., *Substance P levels in the synovium and synovial fluid from patients with rheumatoid arthritis and osteoarthritis*, J Rheumatol, 1993, 20(4): 714
41. Grimsholm, O, S Rantapää-Dahlqvist, T Dalén and S Forsgren, *Observations favouring the occurrence of local production and marked effects of bombesin/gastrin-releasing peptide in the synovial tissue of the human knee joint–Comparisons with substance P and the NK-1 receptor*, Neuropeptides, 2008, 42(2): 133
42. Clark, J M and J A Sidles, *The interrelation of fiber bundles in the anterior cruciate ligament*, J Orthop Res, 1990, 8(2): 180
43. Frank, C B, *Ligament structure, physiology and function.*, Journal of Musculoskeletal and Neuronal Interactions, 2004, 4(2): 199
44. Amiel, D, C Frank, F Harwood, J Fronek and W Akeson, *Tendons and ligaments: a morphological and biochemical comparison*, J Orthop Res, 1983, 1(3): 257
45. Frank, C, D Bray, A Rademaker, C Chrusch, P Sabiston, et al., *Electron microscopic quantification of collagen fibril diameters in the rabbit medial collateral ligament: a baseline for comparison*, Connect. Tissue Res., 1989, 19(1): 11
46. Frank, C B, D A Hart and N G Shrive, *Molecular biology and biomechanics of normal and healing ligaments—a review*, Osteoarthritis Cartilage, 1999, 7(1): 130
47. Kastelic, J, A Galeski and E Baer, *The Multicomposite Structure of Tendon*, Connect. Tissue Res., 1978, 6(1): 11
48. Lekic, P and C A G McCulloch, *Periodontal ligament cell populations: The central role of fibroblasts in creating a unique tissue*, The Anatomical Record, 1996, 245(2): 327
49. Kumagai, K, K Sakai, Y Kusayama, Y Akamatsu, K Sakamaki, et al., *The extent of degeneration of cruciate ligament is associated with chondrogenic differentiation in patients with osteoarthritis of the knee*, Osteoarthritis Cartilage, 2012, 20(11): 1258
50. Comerford, E J, J F Innes, J F Tarlton and A J Bailey, *Investigation of the composition, turnover, and thermal properties of ruptured cranial cruciate ligaments of dogs.*, Am J Vet Res, 2004, 65(8): 1136
51. Maki, J M, R Sormunen, S Lippo, R Kaarteenaho-Wiik, R Soininen, et al., *Lysyl oxidase is essential for normal development and function of the respiratory*

system and for the integrity of elastic and collagen fibers in various tissues, Am J Pathol, 2005, 167(4): 927

52. Lodish, H, A Berk, S Zipursky and e al., *Section 22.3. Collagen: The Fibrous Proteins of the Matrix.*, in *Molecular Cell Biology. 4th edition*, W. Freeman, Editor 2000.
53. Gauba, V and J D Hartgerink, *Self-assembled heterotrimeric collagen triple helices directed through electrostatic interactions*, J Am Chem Soc, 2007, 129(9): 2683
54. Hoffmann, A and G Gross, *Tendon and ligament engineering in the adult organism: mesenchymal stem cells and gene-therapeutic approaches*, Int. Orthop., 2007, 31(6): 791
55. Bornstein, P and H Sage, *Structurally Distinct Collagen Types*, Annu Rev Biochem, 1980, 49(1): 957
56. Church, R L, S E Pfeiffer and M L Tanzer, *Collagen biosynthesis: synthesis and secretion of a high molecular weight collagen precursor (procollagen)*, Proceedings of the National Academy of Sciences, 1971, 68(11): 2638
57. Amiel, D and M E Nimni, *The collagen in normal ligaments*, Iowa Orthop. J., 1993, 13(49
58. Eyre, D R, M A Paz and P M Gallop, *Cross-linking in collagen and elastin*, Annu Rev Biochem, 1984, 53(1): 717
59. Barnard, K, N D Light, T J Sims and A J Bailey, *Chemistry of the collagen cross-links. Origin and partial characterization of a putative mature cross-link of collagen*, Biochem J, 1987, 244(2): 303
60. Comerford, E J, J F Tarlton, J F Innes, K A Johnson, A A Amis, et al., *Metabolism and composition of the canine anterior cruciate ligament relate to differences in knee joint mechanics and predisposition to ligament rupture.*, J Orthop Res, 2005, 23(1): 61
61. Comerford, E J, J F Tarlton, A Wales, A J Bailey and J F Innes, *Ultrastructural differences in cranial cruciate ligaments from dogs of two breeds with a differing predisposition to ligament degeneration and rupture.*, J Comp Pathol, 2006, 134(1): 8
62. Scott, J E, *Proteoglycan-fibrillar collagen interactions*, Biochem J, 1988, 252(2): 313
63. Yanagishita, M, *Function of proteoglycans in the extracellular matrix*, Acta Pathol. Jpn., 1993, 43(6): 283
64. Danielson, K G, H Baribault, D F Holmes, H Graham, K E Kadler, et al., *Targeted Disruption of Decorin Leads to Abnormal Collagen Fibril Morphology and Skin Fragility*, J Cell Biol, 1997, 136(3): 729
65. Chakravarti, S, T Magnuson, J H Lass, K J Jepsen, C LaMantia, et al., *Lumican Regulates Collagen Fibril Assembly: Skin Fragility and Corneal Opacity in the Absence of Lumican*, J Cell Biol, 1998, 141(5): 1277
66. Finkelstein, J E, K Doege, Y Yamada, R E Pyeritz, J M Graham, Jr., et al., *Analysis of the chondroitin sulfate proteoglycan core protein (CSPGCP) gene in achondroplasia and pseudoachondroplasia*, Am J Hum Genet, 1991, 48(1): 97

67. Wiberg, C, D Heinegard, C Wenglen, R Timpl and M Morgelin, *Biglycan organizes collagen VI into hexagonal-like networks resembling tissue structures*, J Biol Chem, 2002, 277(51): 49120
68. Iozzo, R V and A D Murdoch, *Proteoglycans of the extracellular environment: clues from the gene and protein side offer novel perspectives in molecular diversity and function*, FASEB J, 1996, 10(5): 598
69. Miles, C A and M Ghelashvili, *Polymer-in-a-box mechanism for the thermal stabilization of collagen molecules in fibers*, Biophys J, 1999, 76(6): 3243
70. Thornton, G M, N G Shrive and C B Frank, *Altering ligament water content affects ligament pre-stress and creep behavior*, J Orthop Res, 2001, 19(5): 845
71. Iwasaki, S, Y Hosaka, T Iwasaki, K Yamamoto, A Nagayasu, et al., *The modulation of collagen fibril assembly and its structure by decorin: an electron microscopic study*, Arch Histol Cytol, 2008, 71(1): 37
72. Schönherr, E, P Witsch-Prehm, B Harrach, H Robenek, J Rauterberg, et al., *Interaction of Biglycan with Type I Collagen*, J Biol Chem, 1995, 270(6): 2776
73. Plaas, A H K, S Wong-Palms, T Koob, D Hernandez, L Marchuk, et al., *Proteoglycan Metabolism During Repair of the Ruptured Medial Collateral Ligament in Skeletally Mature Rabbits*, Arch Biochem Biophys, 2000, 374(1): 35
74. Wingfield, C, A Amis, A Stead and H Law, *Comparison of the biomechanical properties of rottweiler and racing greyhound cranial cruciate ligaments.*, J Small Anim Pract, 2000, 41(7): 303
75. Wingfield, C, A Amis, A Stead and H Law, *Cranial cruciate stability in the rottweiler and racing greyhound: an in vitro study.*, J Small Anim Pract, 2000, 41(5): 193
76. Griffon, D J, *A Review of the Pathogenesis of Canine Cranial Cruciate Ligament Disease as a Basis for Future Preventive Strategies*, Vet Surg, 2010, 39(4): 399
77. Griffin, L Y, J Agel, M J Albohm, E A Arendt, R W Dick, et al., *Noncontact Anterior Cruciate Ligament Injuries: Risk Factors and Prevention Strategies*, J. Am. Acad. Orthop. Surg., 2000, 8(3): 141
78. Slocum, B and T Devine, *Cranial tibial thrust: a primary force in the canine stifle*, J Am Vet Med Assoc, 1983, 183(4): 456
79. Reif, U, D A Hulse and J G Hauptman, *Effect of tibial plateau leveling on stability of the canine cranial cruciate-deficient stifle joint: an in vitro study*, Vet Surg, 2002, 31(2): 147
80. Warzee, C C, L M Dejardin, S P Arnoczky and R L Perry, *Effect of Tibial Plateau Leveling on Cranial and Caudal Tibial Thrusts in Canine Cranial Cruciate–Deficient Stifles: An In Vitro Experimental Study*, Vet Surg, 2001, 30(3): 278
81. Morris, E and A J Lipowitz, *Comparison of tibial plateau angles in dogs with and without cranial cruciate ligament injuries*, J Am Vet Med Assoc, 2001, 218(3): 363
82. Reif, U and C W Probst, *Comparison of tibial plateau angles in normal and cranial cruciate deficient stifles of Labrador retrievers*, Vet Surg, 2003, 32(4): 385

83. Wilke, V, M Conzemius, M Besancon, R Evans and M Ritter, *Comparison of tibial plateau angle between clinically normal Greyhounds and Labrador Retrievers with and without rupture of the cranial cruciate ligament.*, J Am Vet Med Assoc, 2002, 221(10): 1426
84. Comerford, E, J Tarlton, N Avery, A Bailey and J Innes, *Distal femoral intercondylar notch dimensions and their relationship to composition and metabolism of the canine anterior cruciate ligament.*, Osteoarthritis Cartilage, 2006, 14(3): 273
85. Lewis, B A, D A Allen, T D Henrikson and T W Lehenbauer, *Computed tomographic evaluation of the canine intercondylar notch in normal and cruciate deficient stifles*, Veterinary & Comparative Orthopaedics & Traumatology, 2008, 21(2): 119
86. Shelbourne, K D, T J Davis and T E Klootwyk, *The Relationship Between Intercondylar Notch Width of the Femur and the Incidence of Anterior Cruciate Ligament Tears*, Am J Sports Med, 1998, 26(3): 402
87. Boyd, J S and C P Mosby, *Color Atlas of Clinical Anatomy of the Dog and Cat. 2nd Edition* 2001.
88. Muir, P, P A Manley and Z Hao, *Collagen fragmentation in ruptured canine cranial cruciate ligament explants*, The Veterinary Journal, 2006, 172(1): 121
89. Amiel, D, K K Ishizue, F L Harwood, L Kitabayashi and W H Akeson, *Injury of the anterior cruciate ligament: The role of collagenase in ligament degeneration*, J Orthop Res, 1989, 7(4): 486
90. Vasseur, P B, R R Pool, S P Arnoczky and R E Lau, *Correlative biomechanical and histologic study of the cranial cruciate ligament in dogs*, Am J Vet Res, 1985, 46(9): 1842
91. Smith, K D, A Vaughan-Thomas, D G Spiller, P D Clegg, J F Innes, et al., *Variations in cell morphology in the canine cruciate ligament complex*, The Veterinary Journal, 2012, 193(2): 561
92. Nečas, A, J Zatloukal, H Kecova and M Dvořák, *Predisposition of dog breeds to rupture of the cranial cruciate ligament*, Acta Vet Brno, 2000, 69(4): 305
93. Whitehair, J G, P B Vasseur and N H Willits, *Epidemiology of cranial cruciate ligament rupture in dogs.*, J Am Vet Med Assoc, 1993, 203(7): 1016
94. Brown, D C, M G Conzemius and F S Shofer, *Body weight as a predisposing factor for humeral condylar fractures, cranial cruciate rupture and intervertebral disc disease in Cocker Spaniels*, Veterinary and Comparative Orthopaedics and Traumatology, 1996, 9(2): 75
95. Kealy, R D, D F Lawler, J M Ballam, G Lust, D N Biery, et al., *Evaluation of the effect of limited food consumption on radiographic evidence of osteoarthritis in dogs*, J Am Vet Med Assoc, 2000, 217(11): 1678
96. Adachi, N, K Nawata, M Maeta and Y Kurozawa, *Relationship of the menstrual cycle phase to anterior cruciate ligament injuries in teenaged female athletes*, Basic Res Cardiol, 2008, 128(5): 473
97. Dragoo, J L, R S Lee, P Benhaim, G A M Finerman and S L Hame, *Relaxin receptors in the human female anterior cruciate ligament*, Am J Sports Med, 2003, 31(4): 577

98. Faryniarz, D A, M Bhargava, C Lajam, E T Attia and J A Hannafin, *Quantitation of estrogen receptors and relaxin binding in human anterior cruciate ligament fibroblasts*, J Immunother, 2006, Animal. 42(7): 176
99. Wojtys, E M, L J Huston, M D Boynton, K P Spindler and T N Lindenfeld, *The effect of the menstrual cycle on anterior cruciate ligament injuries in women as determined by hormone levels*, Am J Sports Med, 2002, 30(2): 182
100. Slauterbeck, J R, K Pankratz, K T Xu, S C Bozeman and D M Hardy, *Canine ovariohysterectomy and orchiectomy increases the prevalence of ACL injury*, Doc Ophthalmol Proc Ser, 2004, 429): 301
101. Torres de la Riva, G, B L Hart, T B Farver, A M Oberbauer, L L M Messam, et al., *Neutering Dogs: Effects on Joint Disorders and Cancers in Golden Retrievers*, PLoS ONE, 2013, 8(2): e55937
102. Beynnon, B D, R J Johnson, S Braun, M Sargent, I M Bernstein, et al., *The Relationship Between Menstrual Cycle Phase and Anterior Cruciate Ligament Injury*, Am J Sports Med, 2006, 34(5): 757
103. Zazulak, B T, M Paterno, G D Myer, W A Romani and T E Hewett, *The effects of the menstrual cycle on anterior knee laxity: a systematic review*, Sports Med., 2006, 36(10): 847
104. Myklebust, G, S Maehlum, I Holm and R Bahr, *A prospective cohort study of anterior cruciate ligament injuries in elite Norwegian team handball*, Scandinavian Journal of Medicine & Science in Sports, 1998, 8(3): 149
105. Doom, M, T de Bruin, H de Rooster, H van Bree and E Cox, *Immunopathological mechanisms in dogs with rupture of the cranial cruciate ligament*, Vet Immunol Immunopathol, 2008, 125(1-2): 143
106. El-Hadi, M, C Charavaryamath, A Aebischer, C W Smith, C Shmon, et al., *Expression of interleukin-8 and intercellular cell adhesion molecule-1 in the synovial membrane and cranial cruciate ligament of dogs after rupture of the ligament*, Can J Vet Res, 2012, 76(1): 8
107. Muir, P, Z Schwartz, S Malek, A Kreines, S Y Cabrera, et al., *Contralateral Cruciate Survival in Dogs with Unilateral Non-Contact Cranial Cruciate Ligament Rupture*, PLoS ONE, 2011, 6(10): e25331
108. Goldberg, V M, A Burstein and M Dawson, *The influence of an experimental immune synovitis on the failure mode and strength of the rabbit anterior cruciate ligament*, Macromolecules, 1982, 64(6): 900
109. Lemburg, A K, A Meyer-Lindenberg and M Hewicker-Trautwein, *Immunohistochemical characterization of inflammatory cell populations and adhesion molecule expression in synovial membranes from dogs with spontaneous cranial cruciate ligament rupture*, Vet Immunol Immunopathol, 2004, 97(3-4): 231
110. Klocke, N W, P W Snyder, W R Widmer, W Zhong, G P McCabe, et al., *Detection of synovial macrophages in the joint capsule of dogs with naturally occurring rupture of the cranial cruciate ligament*, Am J Vet Res, 2005, 66(3): 493
111. Tsuboi, H, Y Matsui, K Hayashida, S Yamane, M Maeda-Tanimura, et al., *Tartrate resistant acid phosphatase (TRAP) positive cells in rheumatoid synovium may induce the destruction of articular cartilage*, Ann Rheum Dis, 2003, 62(3): 196

112. Bari, A S, S D Carter, S C Bell, K Morgan and D Bennett, *Anti-type II collagen antibody in naturally occurring canine joint diseases*, Br. J. Rheumatol., 1989, 28(6): 480
113. Fosang, A J, H Stanton, C B Little and L M Atley, *Neoepitopes as biomarkers of cartilage catabolism*, Inflamm Res, 2003, 52(7): 277
114. Lawrence, D, S Bao, P J. Canfield, M Allanson and A J Husband, *Elevation of immunoglobulin deposition in the synovial membrane of dogs with cranial cruciate ligament rupture*, Vet Immunol Immunopathol, 1998, 65(1): 89
115. Bleedorn, J A, E N Greuel, P A Manley, S L Schaefer, M D Markel, et al., *Synovitis in dogs with stable stifle joints and incipient cranial cruciate ligament rupture: a cross-sectional study*, Vet Surg, 2011, 40(5): 531
116. de Rooster, H, E Cox and H van Bree, *Prevalence and relevance of antibodies to type-I and -II collagen in synovial fluid of dogs with cranial cruciate ligament damage*, Am J Vet Res, 2000, 61(11): 1456
117. Carter, S D, S C Bell, A S Bari and D Bennett, *Immune complexes and rheumatoid factors in canine arthritides*, Ann Rheum Dis, 1989, 48(12): 986
118. Fernando, M M, C R Stevens, E C Walsh, P L De Jager, P Goyette, et al., *Defining the role of the MHC in autoimmunity: a review and pooled analysis*, PLoS Genet., 2008, 4(4): e1000024
119. Catchpole, B, L J Kennedy, L J Davison and W E Ollier, *Canine diabetes mellitus: from phenotype to genotype*, J Small Anim Pract, 2008, 49(1): 4
120. Kennedy, L J, S Quarmby, G M Happ, A Barnes, I K Ramsey, et al., *Association of canine hypothyroidism with a common major histocompatibility complex DLA class II allele*, Tissue Antigens, 2006, 68(1): 82
121. Massey, J, A Boag, A D Short, R A Scholey, P S Henthorn, et al., *MHC class II association study in eight breeds of dog with hypoadrenocorticism*, Immunogenetics, 2013,
122. Wilbe, M, P Jokinen, C Hermanrud, L J Kennedy, E Strandberg, et al., *MHC class II polymorphism is associated with a canine SLE-related disease complex*, Immunogenetics, 2009, 61(8): 557
123. Kennedy, L J, T O'Neill, A House, A Barnes, K Kyöstilä, et al., *Risk of anal furunculosis in German Shepherd dogs is associated with the major histocompatibility complex*, Tissue Antigens, 2008, 71(1): 51
124. Ollier, W E, L J Kennedy, W Thomson, A N Barnes, S C Bell, et al., *Dog MHC alleles containing the human RA shared epitope confer susceptibility to canine rheumatoid arthritis*, Immunogenetics, 2001, 53(8): 669
125. Sommer, S, *The importance of immune gene variability (MHC) in evolutionary ecology and conservation*, Frontiers in Zoology, 2005, 2(1): 16
126. September, A V, M P Schwellnus and M Collins, *Tendon and ligament injuries: the genetic component*, Br. J. Sports Med., 2007, 41(4): 241
127. Buote, N, J Fusco and R Radasch, *Age, tibial plateau angle, sex, and weight as risk factors for contralateral rupture of the cranial cruciate ligament in Labradors*, Vet Surg, 2009, 38(4): 481
128. Moore, K W and R A Read, *Cranial cruciate ligament rupture in the dog--a retrospective study comparing surgical techniques*, Aust Vet J, 1995, 72(8): 281

129. The UK Kennel Club, *Newfoundland Breed Standard* ,http://www.thekennelclub.org.uk/item/57, July 2013
130. American Kennel Club (AKC), *Guide to the Newfoundland Club Of America Collection* ,http://www.akc.org/about/depts/archive/NCA_collection.cfm, July 2013
131. American Kennel Club (AKC), *Newfoundland Breed Standard* ,http://www.akc.org/breeds/newfoundland/, July 2013
132. The Canadian Kennel Club, *Newfoundland Breed Standard*, http://www.ckc.ca/en/Default.aspx?tabid=99&BreedCode=NWO, July 2013
133. Coopman, F, G Verhoeven, J Saunders, L Duchateau and H van Bree, *Prevalence of hip dysplasia, elbow dysplasia and humeral head osteochondrosis in dog breeds in Belgium*, Vet. Rec., 2008, 163(22): 654
134. Kirberger, R M and S L Fourie, *Elbow dysplasia in the dog: pathophysiology, diagnosis and control*, J S Afr Vet Assoc, 1998, 69(2): 43
135. Ginja, M M, A M Silvestre, J M Gonzalo-Orden and A J Ferreira, *Diagnosis, genetic control and preventive management of canine hip dysplasia: a review*, The Veterinary Journal, 2010, 184(3): 269
136. Henthorn, P S, J Liu, T Gidalevich, J Fang, M L Casal, et al., *Canine cystinuria: polymorphism in the canine SLC3A1 gene and identification of a nonsense mutation in cystinuric Newfoundland dogs*, Hum Genet, 2000, 107(4): 295
137. Lee, B H, J Dukes-McEwan, A T French and B M Corcoran, *Evaluation of a novel doppler index of combined systolic and diastolic myocardial performance in Newfoundland dogs with familial prevalence of dilated cardiomyopathy*, Veterinary Radiology & Ultrasound, 2002, 43(2): 154
138. Pyle, R L, D F Patterson and S Chacko, *The genetics and pathology of discrete subaortic stenosis in the Newfoundland dog*, Am Heart J, 1976, 92(3): 324
139. British Veterinary Association, 2013, *British Veterinary Association. Hip Scheme*,http://www.bva.co.uk/hip_scheme.aspx, August 2013
140. Association, B V, 2013, *British Veterinary Association. Elbow dysplasia*,www.bva.co.uk/public/documents/elbow_dysplasia.pdf, August 2013
141. Animal DNA Diagnostics, 2013,https://www.animaldnadiagnostics.co.uk/, August 2013
142. Lindblad-Toh, K, C M Wade, T S Mikkelsen, E K Karlsson, D B Jaffe, et al., *Genome sequence, comparative analysis and haplotype structure of the domestic dog*, Nature, 2005, 438(7069): 803
143. Karlsson, E K and K Lindblad-Toh, *Leader of the pack: gene mapping in dogs and other model organisms*, Nat Rev Genet, 2008, 9(9): 713
144. Starkey, M P, T J Scase, C S Mellersh and S Murphy, *Dogs really are man's best friend--canine genomics has applications in veterinary and human medicine!*, Briefings in Functional Genomics and Proteomics, 2005, 4(2): 112
145. Parker, H G, A L Shearin and E A Ostrander, *Man's Best Friend Becomes Biology's Best in Show: Genome Analyses in the Domestic Dog*, in *Annual Review of Genetics, Vol 44*, A. Campbell, M. Lichten and G. Schupbach, Editors. 2010. p. 309.

146. Kirkness, E F, V Bafna, A L Halpern, S Levy, K Remington, et al., *The Dog Genome: Survey Sequencing and Comparative Analysis*, Science, 2003, 301(5641): 1898
147. Sharp, N J, J N Kornegay, S D Van Camp, M H Herbstreith, S L Secore, et al., *An error in dystrophin mRNA processing in golden retriever muscular dystrophy, an animal homologue of Duchenne muscular dystrophy*, Genomics, 1992, 13(1): 115
148. Lin, L, J Faraco, R Li, H Kadotani, W Rogers, et al., *The sleep disorder canine narcolepsy is caused by a mutation in the hypocretin (orexin) receptor 2 gene*, Cell, 1999, 98(3): 365
149. Drogemuller, C, D Becker, A Brunner, B Haase, P Kircher, et al., *A missense mutation in the SERPINH1 gene in Dachshunds with osteogenesis imperfecta*, PLoS Genet., 2009, 5(7): e1000579
150. Sutter, N B, M A Eberle, H G Parker, B J Pullar, E F Kirkness, et al., *Extensive and breed-specific linkage disequilibrium in Canis familiaris.*, Genome Res, 2004, 14(12): 2388
151. Parker, H G and E A Ostrander, *Canine Genomics and Genetics: Running with the Pack*, PLoS Genet., 2005, 1(5):
152. Lou, X, R Todhunter, M Lin, Q Lu, T Liu, et al., *The extent and distribution of linkage disequilibrium in a multi-hierarchic outbred canine pedigree.*, Mamm Genome, 2003, 14(8): 555
153. Lander, E S, L M Linton, B Birren, C Nusbaum, M C Zody, et al., *Initial sequencing and analysis of the human genome*, Nature, 2001, 409(6822): 860
154. Hinds, D A, L L Stuve, G B Nilsen, E Halperin, E Eskin, et al., *Whole-Genome Patterns of Common DNA Variation in Three Human Populations*, Science, 2005, 307(5712): 1072
155. de la Chaux, N, P Messer and P Arndt, *DNA indels in coding regions reveal selective constraints on protein evolution in the human lineage*, BMC Evol Biol, 2007, 7(1): 191
156. Ellegren, H, *Microsatellite mutations in the germline:: implications for evolutionary inference*, Trends Genet, 2000, 16(12): 551
157. Stranger, B E, M S Forrest, M Dunning, C E Ingle, C Beazley, et al., *Relative impact of nucleotide and copy number variation on gene expression phenotypes*, Science, 2007, 315(5813): 848
158. Tautz, D, *Hypervariability of simple sequences as a general source for polymorphic DNA markers*, Nucleic Acids Res, 1989, 17(16): 6463
159. Wang, D G, J-B Fan, C-J Siao, A Berno, P Young, et al., *Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome*, Science, 1998, 280(5366): 1077
160. Isbrandt, D, J J Hopwood, K von Figura and C Peters, *Two novel frameshift mutations causing premature stop codons in a patient with the severe form of Maroteaux-Lamy syndrome*, Hum Mutat, 1996, 7(4): 361
161. Ogura, Y, D K Bonen, N Inohara, D L Nicolae, F F Chen, et al., *A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease*, Nature, 2001, 411(6837): 603

162. Guo, Q-S, B Xia, Y Jiang, Y Qu and J Li, *NOD2 3020insC frameshift mutation is not associated with inflammatory bowel disease in Chinese patients of Han nationality*, WORLD JOURNAL OF GASTROENTEROLOGY., 2004, 10(7): 1069
163. Hunt, R, Z E Sauna, S V Ambudkar, M M Gottesman and C Kimchi-Sarfaty, *Silent (synonymous) SNPs: should we care about them?*, Methods Mol Biol, 2009, 578(23
164. Anfinsen, C B, *Principles that govern the folding of protein chains*, Science, 1973, 181(4096): 223
165. Wong, G K, Z Yang, D A Passey, M Kibukawa, M Paddock, et al., *A population threshold for functional polymorphisms*, Genome Res, 2003, 13(8): 1873
166. Morton, N E, *Linkage disequilibrium maps and association mapping*, J Clin Investig, 2005, 115(6): 1425
167. Waples, R S and P R England, *Estimating contemporary effective population size on the basis of linkage disequilibrium in the face of migration*, Genetics, 2011, 189(2): 633
168. Parker, H G, L V Kim, N B Sutter, S Carlson, T D Lorentzen, et al., *Genetic Structure of the Purebred Domestic Dog*, Science, 2004, 304(5674): 1160
169. Wade, C M, *Inbreeding and genetic diversity in dogs: results from DNA analysis*, The Veterinary Journal, 2011, 189(2): 183
170. Calboli, F C F, J Sampson, N Fretwell and D J Balding, *Population Structure and Inbreeding From Pedigree Analysis of Purebred Dogs*, Genetics, 2008, 179(1): 593
171. The UK Kennel Club, *Mate select program*,http://www.the-kennel-club.org.uk/services/public/mateselect/Default.aspx, July 2013
172. Iakoubova, O A, H Dushkin and D R Beier, *Genetic analysis of a quantitative trait in a mouse model of polycystic kidney disease*, Am J Respir Crit Care Med, 1997, 156(4 Pt 2): S72
173. Massey, J, S Rothwell, C Rusbridge, A Tauro, D Addicott, et al., *Association of an MHC Class II Haplotype with Increased Risk of Polymyositis in Hungarian Vizsla Dogs*, PLoS ONE, 2013, 8(2): e56490
174. Seppala, E H, L L Koskinen, C H Gullov, P Jokinen, P Karlskov-Mortensen, et al., *Identification of a novel idiopathic epilepsy locus in Belgian Shepherd dogs*, PLoS ONE, 2012, 7(3): e33549
175. Short, A D, N M Saleh, B Catchpole, L J Kennedy, A Barnes, et al., *CTLA4 promoter polymorphisms are associated with canine diabetes mellitus*, Tissue Antigens, 2010, 75(3): 242
176. Scholey, R A, R W Blowey, R D Murray, R F Smith, J Cameron, et al., *Investigating host genetic factors in bovine digital dermatitis*, Vet Rec, 2012, 171(24): 624
177. Wood, S H, X Ke, T Nuttall, N McEwan, W E Ollier, et al., *Genome-wide association analysis of canine atopic dermatitis and identification of disease related SNPs.*, Immunogenetics, 2009, 61(11-12): 765
178. Clements, D N, S Wood, S D Carter and W E Ollier, *Assessment of the quality and quantity of genomic DNA recovered from canine blood samples by three different extraction methods.*, Res Vet Sci, 2008, 85(1): 74
179. Genotek, D, 2013, *Collect superior samples for animal genetics: Laboratory protocol for manual purification of DNA*

http://www.dnagenotek.com/ROW/products/animalkits.html
http://www.dnagenotek.com/ROW/pdf/PD-PR-006.pdf, July 2013

180. Lee, Y-H and D T Wong, *Saliva: an emerging biofluid for early detection of diseases*, Am. J. Dent., 2009, 22(4): 241
181. Rogers, N L, S A Cole, H C Lan, A Crossa and E W Demerath, *New saliva DNA collection method compared to buccal cell collection techniques for epidemiological studies*, Am J Hum Biol, 2007, 19(3): 319
182. Zalewska, A, K Zwierz, K Zolkowski and A Gindzienski, *Structure and biosynthesis of human salivary mucins*, Acta Biochim Pol, 2000, 47(4): 1067
183. Breshears, L A, J L Cook, A M Stoker, D B Fox and J K Luther, *The effect of uniaxial cyclic tensile load on gene expression in canine cranial cruciate ligamentocytes*, Vet Surg, 2010, 39(4): 433
184. Hofer, D, S Forterre, A Schweighauser, M Krayer, M Doherr, et al., *Selective iNOS-inhibition does not influence apoptosis in ruptured canine cranial cruciate ligaments*, Veterinary and Comparative Orthopaedics and Traumatology, 2009, 22(3): 198
185. Sigma Life Science, *Cell Culture Manual 2011-2014*, ed. Sigma-Aldrich.
186. Invitrogen, 2013, *TRIzol ® Plus PureLink® RNA Mini Kit*,http://products.invitrogen.com/ivgn/product/12183555

http://tools.invitrogen.com/content/sfs/manuals/Trizol_Plus_man.pdf, July 2013
187. Invitrogen, 2013, *PureLink® DNase Set* ,http://products.invitrogen.com/ivgn/product/12185010,
188. Imbeaud, S, E Graudens, V Boulanger, X Barlet, P Zaborski, et al., *Towards standardization of RNA quality assessment using user-independent classifiers of microcapillary electrophoresis traces*, Nucleic Acids Res, 2005, 33(6): e56
189. Schroeder, A, O Mueller, S Stocker, R Salowsky, M Leiber, et al., *The RIN: an RNA integrity number for assigning integrity values to RNA measurements*, BMC Mol Biol, 2006, 7(1): 3
190. Purcell, S, *PLINK Software V1.07* http://pngu.mgh.harvard.edu/purcell/plink/,
191. Purcell, S, B Neale, K Todd-Brown, L Thomas, M Ferreira, et al., *PLINK: a toolset for whole-genome association and population-based linkage analysis.*, Am J Hum Genet, 2007, 81(
192. Lewis, C M, *Genetic association studies: Design, analysis and interpretation*, Briefings in Bioinformatics, 2002, 3(2): 146
193. Lewis, C M and J Knight, *Introduction to Genetic Association Studies*, Cold Spring Harbor Protocols, 2012, 2012(3): pdb.top068163
194. Sasieni, P D, *From genotypes to genes: doubling the sample size*, Biometrics, 1997, 53(4): 1253
195. Png, E, B Alisjahbana, E Sahiratmadja, S Marzuki, R Nelwan, et al., *A genome wide association study of pulmonary tuberculosis susceptibility in Indonesians*, BMC Medical Genetics, 2012, 13(1): 5
196. Kasperaviciute, D, C B Catarino, E L Heinzen, C Depondt, G L Cavalleri, et al., *Common genetic variation and susceptibility to partial epilepsies: a genome-wide association study*, Brain, 2010, 133(Pt 7): 2136

197. Cantor, R M, K Lange and J S Sinsheimer, *Prioritizing GWAS Results: A Review of Statistical Methods and Recommendations for Their Application*, Am J Hum Genet, 2010, 86(1): 6
198. Barrett, J C, B Fry, J Maller and M J Daly, *Haploview: analysis and visualization of LD and haplotype maps.*, Bioinformatics, 2005,
199. Gabriel, S B, S F Schaffner, H Nguyen, J M Moore, J Roy, et al., *The Structure of Haplotype Blocks in the Human Genome*, Science, 2002, 296(5576): 2225
200. Duggal, P, E Gillanders, T Holmes and J Bailey-Wilson, *Establishing an adjusted p-value threshold to control the family-wide type 1 error in genome wide association studies*, BMC Genomics, 2008, 9(1): 516
201. Moskvina, V, N Craddock, P Holmans, I Nikolov, J S Pahwa, et al., *Gene-wide analyses of genome-wide association data sets: evidence for multiple common risk alleles for schizophrenia and bipolar disorder and for overlap in genetic risk*, Mol Psychiatry, 2009, 14(3): 252
202. Tsai, K, R Noorai, A Starr-Moss, P Quignon, C Rinz, et al., *Genome-wide association studies for multiple diseases of the German Shepherd Dog*, Mamm Genome, 2012, 23(1-2): 203
203. Short, A D, L J Kennedy, A Barnes, N Fretwell, C Jones, et al., *Hardy–Weinberg Expectations in Canine Breeds: Implications for Genetic Studies*, J Hered, 2007, 98(5): 445
204. Schaap, T, *The applicability of the Hardy-Weinberg principle in the study of populations*, Ann Hum Genet, 1980, 44(211
205. Wigginton, J E, D J Cutler and G R Abecasis, *A Note on Exact Tests of Hardy-Weinberg Equilibrium*, Am J Hum Genet, 2005, 76(5): 887
206. Szumilas, M, *Explaining odds ratios*, Journal of the Canadian Academy of Child and Adolescent Psychiatry, 2010, 19(3): 227
207. Witte, J S, *Genome-Wide Association Studies and Beyond*, Annu Rev Public Health, 2010, 31(1): 9
208. The Wellcome Trust Case Control Consortium, *Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls.*, Nature, 2007, 447(7145): 661
209. Tsai, K L, R E Noorai, A N Starr-Moss, P Quignon, C J Rinz, et al., *Genome-wide association studies for multiple diseases of the German Shepherd Dog*, Mamm Genome, 2012, 23(1-2): 203
210. Cadieu, E, M W Neff, P Quignon, K Walsh, K Chase, et al., *Coat variation in the domestic dog is governed by variants in three genes*, Science, 2009, 326(5949): 150
211. Parker, H G, B M VonHoldt, P Quignon, E H Margulies, S Shao, et al., *An expressed fgf4 retrogene is associated with breed-defining chondrodysplasia in domestic dogs*, Science, 2009, 325(5943): 995
212. Bannasch, D, A Young, J Myers, K Truvé, P Dickinson, et al., *Localization of canine brachycephaly using an across breed mapping approach*, PLoS ONE, 2010, 5(3): e9632
213. R Development Core Team, *R: A Language and Environment for Statistical Computing*, 2011, R Foundation for Statistical Computing: Vienna, Austria.

214. Bouaziz, M, C Ambroise and M Guedj, *Accounting for Population Stratification in Practice: A Comparison of the Main Strategies Dedicated to Genome-Wide Association Studies*, PLoS ONE, 2011, 6(12): e28845
215. Yu, K, Z Wang, Q Li, S Wacholder, D J Hunter, et al., *Population Substructure and Control Selection in Genome-Wide Association Studies*, PLoS ONE, 2008, 3(7): e2551
216. Price, A L, N J Patterson, R M Plenge, M E Weinblatt, N A Shadick, et al., *Principal components analysis corrects for stratification in genome-wide association studies*, Nat Genet, 2006, 38(8): 904
217. Kang, H M, J H Sul, S K Service, N A Zaitlen, S Y Kong, et al., *Variance component model to account for sample structure in genome-wide association studies*, Nat Genet, 2010, 42(4): 348
218. Zhang, F, Y Wang and H W Deng, *Comparison of Population-Based Association Study Methods Correcting for Population Stratification*, PLoS ONE, 2008, 3(10):
219. Patterson, N, A L Price and D Reich, *Population Structure and Eigenanalysis*, PLoS Genet., 2006, 2(12): e190
220. Price, A L, N A Zaitlen, D Reich and N Patterson, *New approaches to population stratification in genome-wide association studies*, Nat Rev Genet, 2010, 11(7): 459
221. Devlin, B and K Roeder, *Genomic control for association studies*, Biometrics, 1999, 55(997
222. Wang, D, Y Sun, P Stang, J Berlin, M Wilcox, et al., *Comparison of methods for correcting population stratification in a genome-wide association study of rheumatoid arthritis: principal-component analysis versus multidimensional scaling*, BMC Proceedings, 2009, 3(Suppl 7): S109
223. Forman, O P, J Penderis, C Hartley, L J Hayward, S L Ricketts, et al., *Parallel Mapping and Simultaneous Sequencing Reveals Deletions in BCAN and FAM83H Associated with Discrete Inherited Disorders in a Domestic Dog Breed.*, PLoS Genet., 2012, 8(1): e1002462
224. Li, M, M P Reilly, D J Rader and L S Wang, *Correcting population stratification in genetic association studies using a phylogenetic approach*, Bioinformatics, 2010, 26(6): 798
225. Liu, L, D Zhang, H Liu and C Arendt, *Robust methods for population stratification in genome wide association studies*, BMC Bioinformatics, 2013, 14(1): 132
226. Strachan, T and A Read, *Chapter 11, Genetic mapping of mendelian characters.*, in *Human Molecular Genetics. 2nd edition*1999, Wiley-Liss: New York.
227. Zondervan, K T and L R Cardon, *Designing candidate gene and genome-wide case-control association studies*, Nature Protocols, 2007, 2(10): 2492
228. Amos, C I, *Successful design and conduct of genome-wide association studies*, Hum Mol Genet, 2007, 16(R2): R220
229. Biffi, A, C D Anderson, M A Nalls, R Rahman, A Sonni, et al., *Principal-Component Analysis for Assessment of Population Stratification in Mitochondrial Medical Genetics*, Am J Hum Genet, 2010, 86(6): 904

230. Chang, M L, J S Yokoyama, N Branson, D J Dyer, C Hitte, et al., *Intrabreed Stratification Related to Divergent Selection Regimes in Purebred Dogs May Affect the Interpretation of Genetic Association Studies*, J Hered, 2009, 100(suppl 1): S28
231. Quignon, P, L Herbin, E Cadieu, E F Kirkness, B Hédan, et al., *Canine Population Structure: Assessment and Impact of Intra-Breed Stratification on SNP-Based Association Studies*, PLoS ONE, 2007, 2(12): e1324
232. Yazdani, U and J R Terman, *The semaphorins.*, Genome biology, 2006, 7(3): 211
233. Wilke, V, M Conzemius and M Rothschild, *SNP detection and association analyses of candidate genes for rupture of the cranial cruciate ligament in the dog.*, Anim Genet, 2005, 36(6): 519
234. Posthumus, M, A V September, D O'Cuinneagain, W van der Merwe, M P Schwellnus, et al., *The COL5A1 Gene Is Associated With Increased Risk of Anterior Cruciate Ligament Ruptures in Female Participants*, Am J Sports Med, 2009, 37(11): 2234
235. Temwichitr, J, H A Hazewinkel, M A van Hagen and P A Leegwater, *Polymorphic microsatellite markers for genetic analysis of collagen genes in suspected collagenopathies in dogs*, Journal of veterinary medicine. A, physiology, pathology, clinical medicine, 2007, 54(9): 522
236. Li, W and J Freudenberg, *Two-parameter characterization of chromosome-scale recombination rate*, Genome Res, 2009, 19(12): 2300
237. Foundation., A K C C H, *Centimorgan glossary* http://www.akcchf.org/canine-health/glossary/centimorgan.html, August 2013
238. Vaysse, A, A Ratnakumar, T Derrien, E Axelsson, G Rosengren Pielberg, et al., *Identification of Genomic Regions Associated with Phenotypic Variation between Dog Breeds using Selection Mapping*, PLoS Genet., 2011, 7(10): e1002316
239. Hao, M, K Akrami, K Wei, C De Diego, N Che, et al., *Muscleblind-like 2 (Mbnl2) -deficient mice as a model for myotonic dystrophy*, Dev Dyn, 2008, 237(2): 403
240. Holt, I, V Jacquemin, M Fardaei, C A Sewry, G S Butler-Browne, et al., *Muscleblind-Like Proteins: Similarities and Differences in Normal and Myotonic Dystrophy Muscle*, Am J Pathol, 2009, 174(1): 216
241. Chanock, S J, T Manolio, M Boehnke, E Boerwinkle, D J Hunter, et al., *Replicating genotype–phenotype associations*, Nature, 2007, 447(7145): 655
242. Zeggini, E, M N Weedon, C M Lindgren, T M Frayling, K S Elliott, et al., *Replication of Genome-Wide Association Signals in UK Samples Reveals Risk Loci for Type 2 Diabetes*, Science, 2007, 316(5829): 1336
243. Guo, Y, L-J Tan, S-F Lei, T-L Yang, X-D Chen, et al., *Genome-Wide Association Study Identifies ALDH7A1 as a Novel Susceptibility Gene for Osteoporosis*, PLoS Genet., 2010, 6(1): e1000806
244. Tengvall, K, M Kierczak, K Bergvall, M Olsson, M Frankowiack, et al., *Genome-Wide Analysis in German Shepherd Dogs Reveals Association of a Locus on CFA 27 with Atopic Dermatitis*, PLoS Genet., 2013, 9(5): e1003475
245. Ensembl, 2013, *Ensembl BioMart*

http://www.ensembl.org/biomart/martview/,
246. Gabriel, S, L Ziaugra and D Tabbaa, *SNP genotyping using the Sequenom MassARRAY iPLEX platform.*, Current Protocols in Human Genetics, 2009, Chapter 2(Unit 2.12
247. Mitsouras, K and E Faulhaber, *Saliva as an alternative source of high yield canine genomic DNA for genotyping studies*, BMC Research Notes, 2009, 2(1): 219
248. Kumar, V, H-J Westra, J Karjalainen, D V Zhernakova, T Esko, et al., *Human Disease-Associated Genetic Variation Impacts Large Intergenic Non-Coding RNA Expression*, PLoS Genet., 2013, 9(1): e1003201
249. Shapiro, M B and P Senapathy, *RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression*, Nucleic Acids Res, 1987, 15(17): 7155
250. McCarthy, M I and J N Hirschhorn, *Genome-wide association studies: potential next steps on a genetic journey.*, Hum Mol Genet, 2008, 17(R2): R156
251. Hampe, W, M Rezgaoui, I Hermans-Borgmeyer and H C Schaller, *The genes for the human VPS10 domain-containing receptors are large and contain many small exons*, Hum Genet, 2001, 108(6): 529
252. Hermey, G, *The Vps10p-domain receptor family*, Cell. Mol. Life Sci., 2009, 66(16): 2677
253. Hermey, G, N Plath, C A Hübner, D Kuhl, H C Schaller, et al., *The three sorCS genes are differentially expressed and regulated by synaptic activity*, J Neurochem, 2004, 88(6): 1470
254. Sjøgaard, S L S, *Structural and functional characterization of the Vps10pdomain receptor SorCS2*, in *Faculty of Health Sciences of the University of Aarhus*2006.
255. Saitoh, F, Q B Tian, A Okano, H Sakagami, H Kondo, et al., *NIDD, a novel DHHC-containing protein, targets neuronal nitric-oxide synthase (nNOS) to the synaptic membrane through a PDZ-dependent interaction and regulates nNOS activity*, J Biol Chem, 2004, 279(28): 29461
256. Hintsch, G, A Zurlinden, V Meskenaite, M Steuble, K Fink-Widmer, et al., *The calsyntenins--a family of postsynaptic membrane proteins with distinct neuronal expression patterns*, Mol Cell Neurosci, 2002, 21(3): 393
257. Tian, Y, R Schreiber and K Kunzelmann, *Anoctamins are a family of Ca2+-activated Cl- channels*, J Cell Sci, 2012, 125(Pt 21): 4991
258. Kunzelmann, K, Y Tian, J R Martins, D Faria, P Kongsuphol, et al., *Anoctamins*, Pflugers Arch., 2011, 462(2): 195
259. Murray, P S, C M Kirkwood, M C Gray, M D Ikonomovic, W R Paljug, et al., *β-Amyloid 42/40 ratio and kalirin expression in Alzheimer disease with psychosis*, Neurobiol Aging, 2012, 33(12): 2807
260. Jassen, A K, H Yang, G M Miller, E Calder and B K Madras, *Receptor regulation of gene expression of axon guidance molecules: implications for adaptation*, Mol Pharmacol, 2006, 70(1): 71
261. Schultz, R A, D C Miller, C S Kerr and L Micheli, *Mechanoreceptors in human cruciate ligaments. A histological study*, Macromolecules, 1984, 66(7): 1072

262. Adachi, N, M Ochi, Y Uchio, J Iwasa, K Ryoke, et al., *Mechanoreceptors in the anterior cruciate ligament contribute to the joint position sense*, Acta Orthopaedica, 2002, 73(3): 330
263. Valeriani, M, D Restuccia, V D Lazzaro, F Franceschi, C Fabbriciani, et al., *Central nervous system modifications in patients with lesion of the anterior cruciate ligament of the knee*, Brain, 1996, 119(5): 1751
264. Bali, K, M Dhillon and S Prabhakar, *Proprioception in anterior cruciate ligament deficient knees and its relevance in anterior cruciate ligament reconstruction*, Indian Journal of Orthopaedics, 2011, 45(4): 294
265. Roberts, D, E Ageberg, G Andersson and T Friden, *Clinical measurements of proprioception, muscle strength and laxity in relation to function in the ACL-injured knee*, Knee Surg. Sports Traumatol. Arthrosc., 2007, 15(1): 9
266. Rozzi, S L, S M Lephart, W S Gear and F H Fu, *Knee Joint Laxity and Neuromuscular Characteristics of Male and Female Soccer and Basketball Players*, Am J Sports Med, 1999, 27(3): 312
267. O'Connor, B L, D M Visco, K D Brandt, M Albrecht and A B O'Connor, *Sensory Nerves Only Temporarily Protect the Unstable Canine Knee Joint from Osteoarthritis - Evidence That Sensory Nerves Reprogram the Central Nervous System After Cruciate Ligament Transection*, Arthritis Rheum, 1993, 36(8): 1154
268. Roberts, D, G Andersson and T Friden, *Knee joint proprioception in ACL-deficient knees is related to cartilage injury, laxity and age - A retrospective study of 54 patients*, Acta Orthop Scand, 2004, 75(1): 78
269. Mandelbaum, B R, H J Silvers, D S Watanabe, J F Knarr, S D Thomas, et al., *Effectiveness of a Neuromuscular and Proprioceptive Training Program in Preventing Anterior Cruciate Ligament Injuries in Female Athletes*, Am J Sports Med, 2005, 33(7): 1003
270. Myklebust, G, L Engebretsen, I H Braekken, A Skjølberg, O E Olsen, et al., *Prevention of noncontact anterior cruciate ligament injuries in elite and adolescent female team handball athletes*, Instr. Course Lect., 2007, 56(407
271. Hewett, T E, T N Lindenfeld, J V Riccobene and F R Noyes, *The effect of neuromuscular training on the incidence of knee injury in female athletes. A prospective study*, Am J Sports Med, 1999, 27(6): 699
272. Ragetly, C A, D J Griffon, M K I Hsu, L M Klump and E T Hsiao-Wecksler, *Kinetic and kinematic analysis of the right hind limb during trotting on a treadmill in Labrador Retrievers presumed predisposed or not predisposed to cranial cruciate ligament disease*, Am J Vet Res, 2012, 73(8): 1171
273. Ragetly, C A, D J Griffon, L M Klump and E T Hsiao-Wecksler, *Pelvic limb kinetic and kinematic analysis in Labrador Retrievers predisposed or at a low risk for cranial cruciate ligament disease*, Vet Surg, 2012, 41(8): 973
274. Ragetly, C A, R Evans, A A Mostafa and D J Griffon, *Multivariate analysis of morphometric characteristics to evaluate risk factors for cranial cruciate ligament deficiency in Labrador retrievers*, Vet Surg, 2011, 40(3): 327
275. Mostafa, A A, D J Griffon, M W Thomas and P D Constable, *Morphometric characteristics of the pelvic limb musculature of Labrador Retrievers with and without cranial cruciate ligament deficiency*, Vet Surg, 2010, 39(3): 380

276. Gene Ontology Annotation (GOA) Database, *Relations between 'SORCS2 gene' and 'neuropeptide signaling pathway*, http://biograph.be/concept/graph/C1539804/C1155443, July 2013
277. Arnalich, F, E de Miguel, C Perez-Ayala, M Martinez, J J Vazquez, et al., *Neuropeptides and interleukin-6 in human joint inflammation. Relationship between intraarticular substance P and interleukin-6 concentrations*, Neurosci Lett, 1994, 170(2): 251
278. Grimsholm O Fau - Guo, Y, T Guo Y Fau - Ny, S Ny T Fau - Rantapaa-Dahlqvist, S Rantapaa-Dahlqvist S Fau - Forsgren and S Forsgren, *Are neuropeptides important in arthritis? Studies on the importance of bombesin/GRP and substance P in a murine arthritis model*, 0077-8923 (Print)):
279. Neugebauer, V, F Weiretter and H G Schaible, *Involvement of substance P and neurokinin-1 receptors in the hyperexcitability of dorsal horn neurons during development of acute arthritis in rat's knee joint*, J. Neurophysiol., 1995, 73(4): 1574
280. Marshall, K W, E Theriault and D A Homonko, *Distribution of substance P and calcitonin gene related peptide immunoreactivity in the normal feline knee*, J Rheumatol, 1994, 21(5): 883
281. Larsson, J, A Ekblom, K Henriksson, T Lundeberg and E Theodorsson, *Concentration of Substance P, Neurokinin A, Calcitonin Gene-related Peptide, Neuropeptide Y and Vasoactive Intestinal Polypeptide in Synovial Fluid from Knee Joints in Patients Suffering from Rheumatoid Arthritis*, Scand J Rheumatol, 1991, 20(5): 326
282. Bileviciute, I, T Lundeberg, A Ekblom and E Theodorsson, *Bilateral changes of substance P-, neurokinin A-, calcitonin gene-related peptide- and neuropeptide Y-like immunoreactivity in rat knee joint synovial fluid during acute monoarthritis*, Neurosci Lett, 1993, 153(1): 37
283. Hernanz, A, E De Miguel, N Romera, C Perez-Ayala, J Gijon, et al., *Calcitonin gene-related peptide II, Substance P and Vasoactive Intestinal Peptide in plasma and synovial fluid from patients with inflammatory joint disease*, Rheumatology, 1993, 32(1): 31
284. Partsch, G, M Matucci-Cerinic, S Marabini, S Jantsch, A Pignone, et al., *Collagenase Synthesis of Rheumatoid Arthritis Synoviocytes: Dose-dependent Stimulation by Substance P and Capsaicin*, Scand J Rheumatol, 1991, 20(2): 98
285. Andersson, U and K J Tracey, *Neural reflexes in inflammation and immunity*, J Exp Med, 2012, 209(6): 1057
286. Brydges, N M, D J Argyle, J R Mosley, J C Duncan, S Fleetwood-Walker, et al., *Clinical assessments of increased sensory sensitivity in dogs with cranial cruciate ligament rupture*, The Veterinary Journal, 2012, 193(2): 545
287. Korvick, D L, G J Pijanowski and D J Schaeffer, *Three-dimensional kinematics of the intact and cranial cruciate ligament-deficient stifle of dogs*, J Biomech, 1994, 27(1): 77
288. Liekens, A M, J De Knijf, W Daelemans, B Goethals, P De Rijk, et al., *BioGraph: unsupervised biomedical knowledge discovery via automated hypothesis generation*, Genome biology, 2011, 12(6): R57

289. Kanehisa, M and S Goto, *KEGG: kyoto encyclopedia of genes and genomes*, Nucleic Acids Res, 2000, 28(1): 27
290. Kanehisa, M, S Goto, Y Sato, M Furumichi and M Tanabe, *KEGG for integration and interpretation of large-scale molecular data sets*, Nucleic Acids Res, 2012, 40(Database issue): D109
291. Lester, K J, J L Hudson, M Tropeano, C Creswell, D A Collier, et al., *Neurotrophic gene polymorphisms and response to psychological therapy*, Translational Psychiatry, 2012, 2(e108
292. Clements, D N, A D Short, A Barnes, L J Kennedy, J F Ferguson, et al., *A Candidate Gene Study of Canine Joint Diseases*, J Hered, 2010, 101(1): 54
293. Tabor, H K, N J Risch and R M Myers, *Candidate-gene approaches for studying complex genetic traits: practical considerations*, Nat Rev Genet, 2002, 3(5): 391
294. van Hagen, M A E and P A J Leegwater, *Genetic evaluation of collagen COL2a1 in canine cranial cruciate ligament rupture*, in *Department of Clinical Science of Companion Animals*2004, Utrecht University.
295. September, A V, M Posthumus and M Collins, *Application of genomics in the prevention, treatment and management of Achilles tendinopathy and anterior cruciate ligament ruptures*, Recent Patents on DNA & Gene Sequences, 2012,
296. Posthumus, M, A September, M Keegan, D O'Cuinneagain, W Van der Merwe, et al., *Genetic risk factors for anterior cruciate ligament ruptures: COL1A1 gene variant.*, Br. J. Sports Med., 2009, 43(5): 352
297. Khoschnau, S, H Melhus, A Jacobson, H Rahme, H Bengtsson, et al., *Type I collagen alpha1 Sp1 polymorphism and the risk of cruciate ligament ruptures or shoulder dislocations*, Am J Sports Med, 2008, 36(12): 2432
298. Matheson, S, H Larjava and L Hakkinen, *Distinctive localization and function for lumican, fibromodulin and decorin to regulate collagen fibril organization in periodontal tissues*, J Periodontal Res, 2005, 40(4): 312
299. Geng, Y, D McQuillan and P J Roughley, *SLRP interaction can protect collagen fibrils from cleavage by collagenases*, Matrix Biol, 2006, 25(8): 484
300. Raspanti, M, M Viola, M Sonaggere, M E Tira and R Tenni, *Collagen fibril structure is affected by collagen concentration and decorin*, Biomacromolecules, 2007, 8(7): 2087
301. Fleischmajer, R, L W Fisher, E D MacDonald, L Jacobs, Jr., J S Perlish, et al., *Decorin interacts with fibrillar collagen of embryonic and adult human skin*, J Struct Biol, 1991, 106(1): 82
302. Yu, J and J P Urban, *The elastic network of articular cartilage: an immunohistochemical study of elastin fibres and microfibrils*, J Anat, 2010, 216(4): 533
303. Rodeo, S A, K Suzuki, M Yamauchi, M Bhargava and R F Warren, *Analysis of collagen and elastic fibers in shoulder capsule in patients with shoulder instability*, Am J Sports Med, 1998, 26(5): 634
304. Reiser, K, R J McCormick and R B Rucker, *Enzymatic and nonenzymatic cross-linking of collagen and elastin*, FASEB J, 1992, 6(7): 2439

305. Pellegrini, B, G M Acland and J Ray, *Cloning and characterization of opticin cDNA: evaluation as a candidate for canine oculo-skeletal dysplasia*, Gene, 2002, 282(1-2): 121
306. Monfort, J, G Tardif, P Roughley, P Reboul, C Boileau, et al., *Identification of opticin, a member of the small leucine-rich repeat proteoglycan family, in human articular tissues: a novel target for MMP-13 in osteoarthritis*, Osteoarthritis Cartilage, 2008, 16(7): 749
307. Moren, A, A Olofsson, G Stenman, P Sahlin, T Kanzaki, et al., *Identification and characterization of LTBP-2, a novel latent transforming growth factor-beta-binding protein*, J Biol Chem, 1994, 269(51): 32469
308. Li, B, J P Urban and J Yu, *The distribution of fibrillin-2 and LTBP-2, and their co-localisation with fibrillin-1 in adult bovine tail disc*, J Anat, 2012, 220(2): 164
309. Sakai, L Y, D R Keene and E Engvall, *Fibrillin, a new 350-kD glycoprotein, is a component of extracellular microfibrils*, J Cell Biol, 1986, 103(6 Pt 1): 2499
310. Lonnqvist, L, A Child, K Kainulainen, R Davidson, L Puhakka, et al., *A novel mutation of the fibrillin gene causing ectopia lentis*, Genomics, 1994, 19(3): 573
311. Dietz, H C and R E Pyeritz, *Mutations in the human gene for fibrillin-1 (FBN1) in the Marfan syndrome and related disorders*, Hum Mol Genet, 1995, 4 Spec No(1799
312. Wislowska, M and B Jablonska, *Serum cartilage oligomeric matrix protein (COMP) in rheumatoid arthritis and knee osteoarthritis*, Clin. Rheumatol., 2005, 24(3): 278
313. Thur, J, K Rosenberg, D P Nitsche, T Pihlajamaa, L Ala-Kokko, et al., *Mutations in cartilage oligomeric matrix protein causing pseudoachondroplasia and multiple epiphyseal dysplasia affect binding of calcium and collagen I, II, and IX*, J Biol Chem, 2001, 276(9): 6083
314. van Gent, D, P Sharp, K Morgan and N Kalsheker, *Serpins: structure, function and molecular evolution*, Int J Biochem Cell Biol, 2003, 35(11): 1536
315. Tasker, P N, H Macdonald, W D Fraser, D M Reid, S H Ralston, et al., *Association of PLOD1 polymorphisms with bone mineral density in a population-based study of women from the UK*, Osteoporos. Int., 2006, 17(7): 1078
316. Takaluoma, K, M Hyry, J Lantto, R Sormunen, R A Bank, et al., *Tissue-specific changes in the hydroxylysine content and cross-links of collagens and alterations in fibril morphology in lysyl hydroxylase 1 knock-out mice*, J Biol Chem, 2007, 282(9): 6588
317. Viguet-Carrin, S, P Garnero and P D Delmas, *The role of collagen in bone strength*, Osteoporos. Int., 2006, 17(3): 319
318. Nagan, N and H M Kagan, *Modulation of lysyl oxidase activity toward peptidyl lysine by vicinal dicarboxylic amino acid residues. Implications for collagen cross-linking*, J Biol Chem, 1994, 269(35): 22366
319. Collins, M, M Posthumus and M P Schwellnus, *The COL1A1 gene and acute soft tissue ruptures*, Br. J. Sports Med., 2010, 44(14): 1063
320. Romani, W A, P Langenberg and S M Belkoff, *Sex, collagen expression, and anterior cruciate ligament strength in rats*, J Athl Train, 2010, 45(1): 22

321. Wenstrup, R J, J B Florer, E W Brunskill, S M Bell, I Chervoneva, et al., *Type V collagen controls the initiation of collagen fibril assembly*, J Biol Chem, 2004, 279(51): 53331
322. Collins, M and M Posthumus, *Type V collagen genotype and exercise-related phenotype relationships: a novel hypothesis*, Exerc. Sport Sci. Rev., 2011, 39(4): 191
323. Birk, D E, *Type V collagen: heterotypic type I/V collagen interactions in the regulation of fibril assembly*, Micron, 2001, 32(3): 223
324. Birk, D E, J M Fitch, J P Babiarz, K J Doane and T F Linsenmayer, *Collagen fibrillogenesis in vitro: interaction of types I and V collagen regulates fibril diameter*, J Cell Sci, 1990, 95(4): 649
325. Neurath, M F and E Stofft, *Structure and function of matrix components in the cruciate ligaments. An immunohistochemical, electron-microscopic, and immunoelectron-microscopic study*, Acta Anat. (Basel). 1992, 145(4): 387
326. Warner, L R, R J Brown, S M Yingst and J T Oxford, *Isoform-specific heparan sulfate binding within the amino-terminal noncollagenous domain of collagen alpha1(XI)*, J Biol Chem, 2006, 281(51): 39507
327. Smith, G N, E A Mickler and K D Brandt, *Synthesis of a phenotypically abnormal type XI: type II collagen ratio by chondrocytes from canine osteoarthritic cartilage*, Osteoarthritis Cartilage, 1994, 2(3): 165
328. Posthumus, M, A V September, D O'Cuinneagain, W van der Merwe, M P Schwellnus, et al., *The association between the COL12A1 gene and anterior cruciate ligament ruptures*, Br. J. Sports Med., 2010, 44(16): 1160
329. Agarwal, P, D Zwolanek, D R Keene, J N Schulz, K Blumbach, et al., *Collagen XII and XIV, new partners of cartilage oligomeric matrix protein in the skin extracellular matrix suprastructure*, J Biol Chem, 2012, 287(27): 22549
330. Ansorge, H L, X Meng, G Zhang, G Veit, M Sun, et al., *Type XIV Collagen Regulates Fibrillogenesis: PREMATURE COLLAGEN FIBRIL GROWTH AND TISSUE DYSFUNCTION IN NULL MICE*, J Biol Chem, 2009, 284(13): 8427
331. Koch, M, F Laub, P Zhou, R A Hahn, S Tanaka, et al., *Collagen XXIV, a vertebrate fibrillar collagen with structural features of invertebrate collagens: selective expression in developing cornea and bone*, J Biol Chem, 2003, 278(44): 43236
332. Vincenti, M P and C E Brinckerhoff, *Transcriptional regulation of collagenase (MMP-1, MMP-13) genes in arthritis: integration of complex signaling pathways for the recruitment of gene-specific transcription factors*, Arthritis research, 2002, 4(3): 157
333. Cawston, T, *Matrix metalloproteinases and TIMPs: properties and implications for the rheumatic diseases*, Mol. Med. Today, 1998, 4(3): 130
334. Garnero, P, O Borel, I Byrjalsen, M Ferreras, F H Drake, et al., *The collagenolytic activity of cathepsin K is unique among mammalian proteinases*, J Biol Chem, 1998, 273(48): 32347
335. Hayami, T, Y Zhuo, G A Wesolowski, M Pickarski and T Duong le, *Inhibition of cathepsin K reduces cartilage degeneration in the anterior cruciate ligament transection rabbit and murine models of osteoarthritis*, Bone, 2012, 50(6): 1250

336. Neame, P J and J D Sandy, *Cartilage aggrecan. Biosynthesis, degradation and osteoarthritis*, J. Fla. Med. Assoc., 1994, 81(3): 191
337. Dodge, G R, A Diaz, C Sanz-Rodriguez, A M Reginato and S A Jimenez, *Effects of interferon-gamma and tumor necrosis factor alpha on the expression of the genes encoding aggrecan, biglycan, and decorin core proteins in cultured human chondrocytes*, Arthritis Rheum, 1998, 41(2): 274
338. Le Grand, F, R Grifone, P Mourikis, C Houbron, C Gigaud, et al., *Six1 regulates stem cell repair potential and self-renewal during skeletal muscle regeneration*, J Cell Biol, 2012, 198(5): 815
339. Gordon, B S, D C Delgado Diaz, J P White, J A Carson and M C Kostek, *Six1 and Six1 cofactor expression is altered during early skeletal muscle overload in mice*, Journal of Physiological Sciences, 2012, 62(5): 393
340. Chou, P Y and G D Fasman, *Prediction of protein conformation*, Biochemistry, 1974, 13(2): 222
341. Chou, P Y and G D Fasman, *Conformational parameters for amino acids in helical, β-sheet, and random coil regions calculated from proteins*, Biochemistry, 1974, 13(2): 211
342. Kim, S, P Wong and P A Coulombe, *A keratin cytoskeletal protein regulates protein synthesis and epithelial cell growth*, Nature, 2006, 441(7091): 362
343. Moll, R, M Divo and L Langbein, *The human keratins: biology and pathology*, Histochem Cell Biol, 2008, 129(6): 705
344. Li, Z, W-S Hou, C R Escalante-Torres, B D Gelb and D Brömme, *Collagenase Activity of Cathepsin K Depends on Complex Formation with Chondroitin Sulfate*, J Biol Chem, 2002, 277(32): 28669
345. Cheng, X W, M Kuzuya, T Sasaki, K Arakawa, S Kanda, et al., *Increased expression of elastolytic cysteine proteases, cathepsins S and K, in the neointima of balloon-injured rat carotid arteries*, Am J Pathol, 2004, 164(1): 243
346. Zaidi, M, B Troen, B S Moonga and E Abe, *Cathepsin K, Osteoclastic Resorption, and Osteoporosis Therapy*, J Bone Miner Res, 2001, 16(10): 1747
347. Yasuda, Y, J Kaleta and D Brömme, *The role of cathepsins in osteoporosis and arthritis: Rationale for the design of new therapeutics*, Adv Drug Delivery Rev, 2005, 57(7): 973
348. Hou, W S, Z Li, R E Gordon, K Chan, M J Klein, et al., *Cathepsin k is a critical protease in synovial fibroblast-mediated collagen degradation*, Am J Pathol, 2001, 159(6): 2167
349. Bell, E, *Innate immunity: New role for cathepsin K*, Nature Reviews Immunology, 2008, 8(3):
350. Conus, S and H U Simon, *Cathepsins and their involvement in immune responses*, Swiss Med Wkly, 2010, 140(w13042
351. Barrett, J G, Z Hao, B K Graf, L D Kaplan, J P Heiner, et al., *Inflammatory changes in ruptured canine cranial and human anterior cruciate ligaments*, Am J Vet Res, 2005, 66(12): 2073
352. Hasegawa, A, H Nakahara, M Kinoshita, H Asahara, J Koziol, et al., *Cellular and extracellular matrix changes in anterior cruciate ligaments during human knee aging and osteoarthritis*, Arthrit. Res. Ther., 2013, 15(1): R29

353. Petersen, W and B Tillmann, *Structure and vascularization of the cruciate ligaments of the human knee joint*, Anat Embryol, 1999, 200(3): 325
354. Schultz, G S and A Wysocki, *Interactions between extracellular matrix and growth factors in wound healing*, Wound Repair Regen, 2009, 17(2): 153
355. Juliano, R L and S Haskill, *Signal transduction from the extracellular matrix*, J Cell Biol, 1993, 120(3): 577
356. Geiger, B, A Bershadsky, R Pankov and K M Yamada, *Transmembrane crosstalk between the extracellular matrix--cytoskeleton crosstalk*, Nat Rev Mol Cell Biol, 2001, 2(11): 793
357. Benjamin, M and J R Ralphs, *Fibrocartilage in tendons and ligaments--an adaptation to compressive load*, J Anat, 1998, 193 (Pt 4)(481
358. Pratta, M A, W Yao, C Decicco, M D Tortorella, R-Q Liu, et al., *Aggrecan Protects Cartilage Collagen from Proteolytic Cleavage*, J Biol Chem, 2003, 278(46): 45539
359. Gibson, M A, G Hatzinikolas, E C Davis, E Baker, G R Sutherland, et al., *Bovine latent transforming growth factor beta 1-binding protein 2: molecular cloning, identification of tissue isoforms, and immunolocalization to elastin-associated microfibrils*, Mol Cell Biol, 1995, 15(12): 6932
360. Vehvilainen, P, M Hyytiainen and J Keski-Oja, *Latent transforming growth factor-beta-binding protein 2 is an adhesion protein for melanoma cells*, J Biol Chem, 2003, 278(27): 24705
361. Mohanty, K, M Tanwar, R Dada and T Dada, *Screening of the LTBP2 gene in a north Indian population with primary congenital glaucoma*, Mol Vis, 2013, 19(78
362. Sinha, S, A M Heagerty, C A Shuttleworth and C M Kielty, *Expression of latent TGF-beta binding proteins and association with TGF-beta1 and fibrillin-1 following arterial injury*, Cardiovasc Res, 2002, 53(4): 971
363. Kielty, C M, C Baldock, D Lee, M J Rock, J L Ashworth, et al., *Fibrillin: from microfibril assembly to biomechanical function*, Philos Trans R Soc Lond B Biol Sci, 2002, 357(1418): 207
364. Kavanagh, E and D E Ashhurst, *Distribution of Biglycan and Decorin in Collateral and Cruciate Ligaments and Menisci of the Rabbit Knee Joint*, Mo. Dent. J., 2001, 49(7): 877
365. Danielson, K G, H Baribault, D F Holmes, H Graham, K E Kadler, et al., *Targeted disruption of decorin leads to abnormal collagen fibril morphology and skin fragility*, J Cell Biol, 1997, 136(3): 729
366. Friedman, J S, M Faucher, P Hiscott, V L Biron, M Malenfant, et al., *Protein localization in the human eye and genetic screen of opticin*, Hum Mol Genet, 2002, 11(11): 1333
367. Mellersh, C, *DNA testing and domestic dogs*, Mamm Genome, 2012, 23(1-2): 109
368. Bennett, S R M, F R Carbone, F Karamalis, J F A P Miller and W R Heath, *Induction of a CD8+ Cytotoxic T Lymphocyte Response by Cross-priming Requires Cognate CD4+ T Cell Help*, The Journal of Experimental Medicine, 1997, 186(1): 65
369. Janeway, C, P Travers, M Walport and E al., *Immunobiology: The Immune System in Health and Disease. 5th edition.*2001, New York: Garland Science.

370. Kennedy, L J, S D Carter, A Barnes, S Bell, D Bennett, et al., *DLA-DRB1 polymorphisms in dogs defined by sequence-specific oligonucleotide probes (SSOP)*, Tissue Antigens, 1999, 53(2): 184
371. Kennedy, L J, A Barnes, G M Happ, R J Quinnell, O Courtenay, et al., *Evidence for extensive DLA polymorphism in different dog populations*, Tissue Antigens, 2002, 60(1): 43
372. Kennedy, L J, A Barnes, G M Happ, R J Quinnell, D Bennett, et al., *Extensive interbreed, but minimal intrabreed, variation of DLA class II alleles and haplotypes in dogs*, Tissue Antigens, 2002, 59(3): 194
373. Kennedy, L J, A Barnes, A Short, J J Brown, J Seddon, et al., *Canine DLA diversity: 3. Disease studies*, Tissue Antigens, 2007, 69(292
374. Kennedy, L J, *14th International HLA and Immunogenetics Workshop: report on joint study on canine DLA diversity*, Tissue Antigens, 2007, 69 Suppl 1(269
375. Bexfield, N H, P J Watson, J Aguirre-Hernandez, D R Sargan, L Tiley, et al., *DLA class II alleles and haplotypes are associated with risk for and protection from chronic hepatitis in the English Springer spaniel*, PLoS ONE, 2012, 7(8): e42584
376. Wagner, J L, R C Burnett, S A DeRose and R Storb, *Molecular analysis and polymorphism of the DLA-DQA gene*, Tissue Antigens, 1996, 48(3): 199
377. Dean, A G, K M Sullivan and M M Soe, *OpenEpi: Open Source Epidemiologic Statistics for Public Health*,www.OpenEpi.com,
378. Kennedy, L J, A Barnes, W E Ollier and M J Day, *Association of a common dog leucocyte antigen class II haplotype with canine primary immune-mediated haemolytic anaemia*, Tissue Antigens, 2006, 68(6): 502
379. Clements, D N, L J Kennedy, A D Short, A Barnes, J Ferguson, et al., *Risk of canine cranial cruciate ligament rupture is not associated with the major histocompatibility complex*, Veterinary and Comparative Orthopaedics and Traumatology, 2011, 24(4): 262
380. Angles, J M, L J Kennedy and N C Pedersen, *Frequency and distribution of alleles of canine MHC-II DLA-DQB1, DLA-DQA1 and DLA-DRB1 in 25 representative American Kennel Club breeds*, Tissue Antigens, 2005, 66(3): 173
381. Runstadler, J A, J M Angles and N C Pedersen, *Dog leucocyte antigen class II diversity and relationships among indigenous dogs of the island nations of Indonesia (Bali), Australia and New Guinea*, Tissue Antigens, 2006, 68(5): 418
382. Silva, A S, S H Wood, S van Dam, S Berres, A McArdle, et al., *Gathering insights on disease etiology from gene expression profiles of healthy tissues*, Bioinformatics, 2011, 27(23): 3300
383. Morley, M, *Genetic analysis of genome-wide variation in human gene expression*, Nature, 2004, 430(743
384. Merryman-Simpson, A E, S H Wood, N Fretwell, P G Jones, W M McLaren, et al., *Gene (mRNA) expression in canine atopic dermatitis: microarray analysis*, Veterinary Dermatology, 2008, 19(2): 59
385. Livak, K J and T D Schmittgen, *Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method*, Methods, 2001, 25(4): 402

386. Lewis, P A and M R Cookson, *Gene expression in the Parkinson's disease brain*, Brain Res Bull, 2012, 88(4): 302
387. Kort, E J, P Norton, P Haak, B Berghuis, S Ramirez, et al., *Review Paper: Gene Expression Profiling in Veterinary and Human Medicine: Overview of Applications and Proposed Quality Control Practices*, Vet. Pathol., 2009, 46(4): 598
388. Dubois, P C, G Trynka, L Franke, K A Hunt, J Romanos, et al., *Multiple common variants for celiac disease influencing immune gene expression*, Nat Genet, 2010, 42(4): 295
389. Alter, M D, R Kharkar, K E Ramsey, D W Craig, R D Melmed, et al., *Autism and Increased Paternal Age Related Changes in Global Levels of Gene Expression Regulation*, PLoS ONE, 2011, 6(2):
390. Alcina, A, M Abad-Grau Mdel, M Fedetz, G Izquierdo, M Lucas, et al., *Multiple Sclerosis Risk Variant HLA-DRB1*1501 Associates with High Expression of DRB1 Gene in Different Human Populations*, PLoS ONE, 2012, 7(1):
391. Dermitzakis, E T, *From gene expression to disease risk*, Nat Genet, 2008, 40(5): 492
392. Whitchurch, A K, *Gene expression microarrays*, Potentials, IEEE, 2002, 21(1): 30
393. Schena, M, D Shalon, R W Davis and P O Brown, *Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray*, Science, 1995, 270(5235): 467
394. Maccoux, L J, F Salway, P J R Day and D N Clements, *Expression profiling of select cytokines in canine osteoarthritis tissues*, Vet Immunol Immunopathol, 2007, 118(1-2): 59
395. Wang, Z, M Gerstein and M Snyder, *RNA-Seq: a revolutionary tool for transcriptomics*, Nat Rev Genet, 2009, 10(1): 57
396. Roche, *Universal ProbeLibrary Assay Design Center*,http://www.roche-applied-science.com/shop/CategoryDisplay?catalogId=10001&tab=&identifier=Universal+Probe+Library, July 2013
397. Pfaffl, M W, A Tichopad, C Prgomet and T P Neuvians, *Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper--Excel-based tool using pair-wise correlations*, Biotechnol Lett, 2004, 26(6): 509
398. Silver, N, S Best, J Jiang and S Thein, *Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR*, BMC Mol Biol, 2006, 7(1): 33
399. Huggett, J, K Dheda, S Bustin and A Zumla, *Real-time RT-PCR normalisation; strategies and considerations*, Genes & Immunity, 2005, 6(4): 279
400. Ayers, D, D Clements, F Salway and P Day, *Expression stability of commonly used reference genes in canine articular connective tissues*, BMC veterinary research, 2007, 3(1): 7
401. Vandesompele, J, K De Preter, F Pattyn, B Poppe, N Van Roy, et al., *Accurate normalization of real-time quantitative RT-PCR data by geometric averaging*

of multiple internal control genes, Genome biology, 2002, 3(7): RESEARCH0034
402. Andersen, C L, J L Jensen and T F Orntoft, *Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets*, Cancer Res, 2004, 64(15): 5245
403. Guo, S and L DiPietro, *Factors Affecting Wound Healing*, J Dent Res, 2010, 89(3): 219
404. Liu, S H, R-S Yang, R Al-Shaikh and J M Lane, *Collagen in Tendon, Ligament, and Bone Healing: A Current Review*, Clin. Orthop., 1995, 318(265
405. Murray, M M, S D Martin, T L Martin and M Spector, *Histological changes in the human anterior cruciate ligament after rupture*, Macromolecules, 2000, 82-A(10): 1387
406. Singer, A J and R A Clark, *Cutaneous wound healing*, N Engl J Med, 1999, 341(10): 738
407. Raghow, R, *The role of extracellular matrix in postinflammatory wound healing and fibrosis*, FASEB J, 1994, 8(11): 823
408. Boykiw, R, P Sciore, C Reno, L Marchuk, C B Frank, et al., *Altered levels of extracellular matrix molecule mRNA in healing rabbit ligaments*, Matrix Biol, 1998, 17(5): 371
409. Lo, I K Y, L L Marchuk, D A Hart and C B Frank, *Comparison of mRNA levels for matrix molecules in normal and disrupted human anterior cruciate ligaments using reverse transcription-polymerase chain reaction*, J Orthop Res, 1998, 16(4): 421
410. Young, K, T Samiric, J Feller and J Cook, *Extracellular matrix content of ruptured anterior cruciate ligament tissue*, The Knee, 2011, 18(4): 242
411. Greenbaum, D, C Colangelo, K Williams and M Gerstein, *Comparing protein abundance and mRNA expression levels on a genomic scale*, Genome biology, 2003, 4(9): 117
412. Muir, P, G M Schamberger, P A Manley and Z Hao, *Localization of Cathepsin K and Tartrate-Resistant Acid Phosphatase in Synovium and Cranial Cruciate Ligament in Dogs with Cruciate Disease*, Vet Surg, 2005, 34(3): 239
413. Halleen, J M, S Räisänen, J J Salo, S V Reddy, G D Roodman, et al., *Intracellular Fragmentation of Bone Resorption Products by Reactive Oxygen Species Generated by Osteoclastic Tartrate-resistant Acid Phosphatase*, J Biol Chem, 1999, 274(33): 22907
414. Janckila, A J, S P Slone, S C Lear, A Martin and L T Yam, *Tartrate-resistant acid phosphatase as an immunohistochemical marker for inflammatory macrophages*, Am J Clin Pathol, 2007, 127(4): 556
415. Hayman, A R, A J Bune, J R Bradley, J Rashbass and T M Cox, *Osteoclastic Tartrate-resistant Acid Phosphatase (Acp 5): Its Localization to Dendritic Cells and Diverse Murine Tissues*, Mo. Dent. J., 2000, 48(2): 219
416. Kirstein, B, T J Chambers and K Fuller, *Secretion of tartrate-resistant acid phosphatase by osteoclasts correlates with resorptive behavior*, J Cell Biochem, 2006, 98(5): 1085

417. Forterre, S, A Zurbriggen and D Spreng, *Nitric oxide induces cell death in canine cruciate ligament cells by activation of tyrosine kinase and reactive oxygen species*, BMC veterinary research, 2012, 8(40
418. Savalas, L R, B Gasnier, M Damme, T Lubke, C Wrocklage, et al., *Disrupted in renal carcinoma 2 (DIRC2), a novel transporter of the lysosomal membrane, is proteolytically processed by cathepsin L*, Biochem J, 2011, 439(1): 113
419. Pao, S S, I T Paulsen and M H Saier, Jr., *Major facilitator superfamily*, Microbiol Mol Biol Rev, 1998, 62(1): 1
420. Wang, H M, *Detection of lysosomal enzymes derived from pig periodontal ligament fibroblasts and their ability to digest collagen fibrils and proteoglycan*, Arch Oral Biol, 1982, 27(9): 715
421. Hefti, F L, A Kress, J Fasel and E W Morscher, *Healing of the transected anterior cruciate ligament in the rabbit*, J. Bone Joint Surg. Am., 1991, 73(3): 373
422. Murray, M M and K P Spindler, *Anterior cruciate ligament healing and repair*, Sports Med. Arth. Rev., 2005, 13(3): 151
423. Bray, R C, C A Leonard and P T Salo, *Correlation of healing capacity with vascular response in the anterior cruciate and medial collateral ligaments of the rabbit*, J Orthop Res, 2003, 21(6): 1118
424. Cui, X and G Churchill, *Statistical tests for differential expression in cDNA microarray experiments*, Genome biology, 2003, 4(4): 210
425. Shendure, J and H Ji, *Next-generation DNA sequencing*, Nat Biotechnol, 2008, 26(10): 1135
426. Chepelev, I, G Wei, Q Tang and K Zhao, *Detection of single nucleotide variations in expressed exons of the human genome using RNA-Seq*, Nucleic Acids Res, 2009, 37(16): e106
427. Allen, D D, R Caviedes, A M Cardenas, T Shimahara, J Segura-Aguilar, et al., *Cell lines as in vitro models for drug screening and toxicity studies*, Drug Dev. Ind. Pharm., 2005, 31(8): 757
428. Yang, X, S Zhang, X Pang and M Fan, *Pro-inflammatory cytokines induce odontogenic differentiation of dental pulp-derived stem cells*, J Cell Biochem, 2012, 113(2): 669
429. Petanidis, S, M Hadzopoulou-Cladaras and A Salifoglou, *Cadmium modulates H-ras expression and caspase-3 apoptotic cell death in breast cancer epithelial MCF-7 cells*, J Inorg Biochem, 2013, 121C(100
430. Jingushi, K, T Nakamura, F Takahashi-Yanaga, E Matsuzaki, Y Watanabe, et al., *Differentiation-Inducing Factor-1 Suppresses the Expression of c-Myc in the Human Cancer Cell Lines*, Journal of pharmacological sciences, 2013,
431. Bluteau, G, J Gouttenoire, T Conrozier, P Mathieu, E Vignon, et al., *Differential gene expression analysis in a rabbit model of osteoarthritis induced by anterior cruciate ligament (ACL) section*, Biorheology, 2002, 39(1-2): 247
432. Yoshioka, M, R D Coutts, D Amiel and S A Hacker, *Characterization of a model of osteoarthritis in the rabbit knee*, Osteoarthritis Cartilage, 1996, 4(2): 87

433. Carletti, E, A Motta and C Migliaresi, *Scaffolds for Tissue Engineering and 3D Cell Culture*, in *3D Cell Culture*, J.W. Haycock, Editor 2011, Humana Press. p. 17.
434. Zeiher, B G, E Eichwald, J Zabner, J J Smith, A P Puga, et al., *A mouse model for the delta F508 allele of cystic fibrosis*, J Clin Investig, 1995, 96(4): 2051
435. Neill, D, D Hughes, J A Edwardson, B K Rima and D Allsop, *Human IMR-32 neuroblastoma cells as a model cell line in Alzheimer's disease research*, J Neurosci Res, 1994, 39(4): 482
436. Brandon, E F A, C D Raap, I Meijerman, J H Beijnen and J H M Schellens, *An update on in vitro test methods in human hepatic drug biotransformation research: pros and cons*, Toxicol Appl Pharmacol, 2003, 189(3): 233
437. ATTC, *Passage Number effects in cell lines. Technical bulletin No.7*,http://www.atcc.org/~/media/PDFs/Technical%20Bulletins/tb07.ashx, July 2013
438. Kilkenny, C, W Browne, I C Cuthill, M Emerson and D G Altman, *Animal research: reporting in vivo experiments: the ARRIVE guidelines*, Br J Pharmacol, 2010, 160(7): 1577
439. Marshall, K W and A D Chan, *Arthroscopic anterior cruciate ligament transection induces canine osteoarthritis*, J Rheumatol, 1996, 23(2): 338
440. Pond, M J and G Nuki, *Experimentally-induced osteoarthritis in the dog*, Ann Rheum Dis, 1973, 32(4): 387
441. Lopez, M J, D Kunz, R Vanderby, Jr., D Heisey, J Bogdanske, et al., *A comparison of joint stability between anterior cruciate intact and deficient knees: a new canine model of anterior cruciate ligament disruption*, J Orthop Res, 2003, 21(2): 224
442. Gillet, J P, S Varma and M M Gottesman, *The clinical relevance of cancer cell lines*, J. Natl. Cancer Inst., 2013, 105(7): 452
443. W.M.S., R and B R.L., *The Principles of Humane Experimental Technique*1959.
444. Adams, A M, J V Soames and R F Searle, *Cultural and morphological characteristics of human periodontal ligament cells in vitro*, Arch Oral Biol, 1993, 38(8): 657
445. Jaffe, E A, R L Nachman, C G Becker and C R Minick, *Culture of human endothelial cells derived from umbilical veins. Identification by morphologic and immunologic criteria*, J Clin Invest, 1973, 52(11): 2745
446. Browning, T H and J S Trier, *Organ culture of mucosal biopsies of human small intestine*, J Clin Invest, 1969, 48(8): 1423
447. Invitrogen, *Dissociation of Cells from Primary Tissue*,http://www.invitrogen.com/etc/medialib/en/filelibrary/pdf.Par.18492.File.dat/Dissociation_Cells_Y14477_Dissociation.pdf, July 2013
448. Muir, P, P A Manley and Z Hao, *COL-3 inhibition of collagen fragmentation in ruptured cranial cruciate ligament explants from dogs with stifle arthritis*, The Veterinary Journal, 2007, 174(2): 403
449. Treede, I, A Braun, P Jeliaskova, T Giese, J Fullekrug, et al., *TNF-alpha-induced up-regulation of pro-inflammatory cytokines is reduced by phosphatidylcholine in intestinal epithelial cells*, BMC Gastroenterology, 2009, 9(1): 53

450. Li, P and E M Schwarz, *The TNF-alpha transgenic mouse model of inflammatory arthritis*, Springer Semin Immunopathol, 2003, 25(1): 19
451. Kollias, G, *TNF pathophysiology in murine models of chronic inflammation and autoimmunity*, Semin Arthritis Rheum, 2005, 34(5, Supplement 1): 3
452. Keffer, J, L Probert, H Cazlaris, S Georgopoulos, E Kaslaris, et al., *Transgenic mice expressing human tumour necrosis factor: a predictive genetic model of arthritis*, EMBO J., 1991, 10(13): 4025
453. Akesen, B, B Demirag and F Budak, *Evaluation of intra-articular collagenase, TIMP-1, and TNF-alpha levels before and after anterior cruciate ligament reconstruction*, Acta Orthop Traumatol Turc., 2009, 43(3): 214
454. Apostolaki, M, M Armaka, P Victoratos and G Kollias, *Cellular mechanisms of TNF function in models of inflammation and autoimmunity*, Curr Dir Autoimmun, 2010, 11(1
455. Treede, I, A Braun, R Sparla, M Kuhnel, T Giese, et al., *Anti-inflammatory effects of phosphatidylcholine*, J Biol Chem, 2007, 282(37): 27155
456. Tanabe, K, R Matsushima-Nishiwaki, S Yamaguchi, H Iida, S Dohi, et al., *Mechanisms of tumor necrosis factor-α-induced interleukin-6 synthesis in glioma cells*, J Neuroinflammation, 2010, 7(1): 1
457. Andonovska, B, C Dimova and S Panov, *Matrix metalloproteinases (MMP-1, -8, -13) in chronic periapical lesions*, Vojnosanit. Pregl., 2008, 65(12): 882
458. Parks, W C, C L Wilson and Y S Lopez-Boado, *Matrix metalloproteinases as modulators of inflammation and innate immunity*, Nat. Rev. Immunol., 2004, 4(8): 617
459. Hauschildt, S, P Scheipers, W Bessler, K Schwarz, A Ullmer, et al., *Role of ADP-ribosylation in activated monocytes/macrophages*, Adv Exp Med Biol, 1997, 419(249
460. Hassa, P and M Hottiger, *The functional role of poly (ADP-ribose) polymerase 1 as novel coactivator of NF-κB in inflammatory disorders*, Cell. Mol. Life Sci., 2002, 59(9): 1534
461. Waddington, R J, H C Roberts, R V Sugars and E Schonherr, *Differential roles for small leucine-rich proteoglycans in bone formation*, Journal of European Cells and Materials, 2003, 6(12
462. Reinboth, B, E Hanssen, E G Cleary and M A Gibson, *Molecular interactions of biglycan and decorin with elastic fiber components: biglycan forms a ternary complex with tropoelastin and microfibril-associated glycoprotein 1*, J Biol Chem, 2002, 277(6): 3950
463. Tufvesson, E and G Westergren-Thorsson, *Biglycan and decorin induce morphological and cytoskeletal changes involving signalling by the small GTPases RhoA and Rac1 resulting in lung fibroblast migration*, J Cell Sci, 2003, 116(23): 4857
464. Zandvoort, A, D Postma, M Jonker, J Noordhoek, J Vos, et al., *Smad gene expression in pulmonary fibroblasts: indications for defective ECM repair in COPD*, Respiratory Research, 2008, 9(1): 83
465. Ungefroren, H and N B Krull, *Transcriptional Regulation of the Human Biglycan Gene*, J Biol Chem, 1996, 271(26): 15787

466. Ameye, L, D Aria, K Jepsen, A Oldberg, T Xu, et al., *Abnormal collagen fibrils in tendons of biglycan/fibromodulin-deficient mice lead to gait impairment, ectopic ossification, and osteoarthritis*, FASEB J, 2002, 16(7): 673
467. Tang, H, T Quertermous, B Rodriguez, S L Kardia, X Zhu, et al., *Genetic structure, self-identified race/ethnicity, and confounding in case-control association studies*, Am J Hum Genet, 2005, 76(2): 268

Chapter 11.

Appendices

I. Sample dog data

Appendix 1. Summary of age and sex data for each dog breed: including age (minimum and maximum), sex of dogs (number of males, females, unknown sex and the number of dogs spayed/neutered) and the total number of dogs in each category.

	Age of dogs			Sex of dogs							
Breed	**Minimum age**	**Maximum age**	**Mean age**	**Total females**	**Female (entire)**	**Female (spayed)**	**Total males**	**Male (entire)**	**Male (neutered)**	**Unknown sex**	**Total number of dogs**
Newfoundland Case	4 m	12 yr	3 yr 3 m	53	27	26	45	24	21	1	99
Newfoundland Control	5 yr	13 yr	7 yr 6 m	93	40	53	45	30	15	34	172
Labrador Case	1 yr	12 yr 1 m	5 yr 3 m	68	23	45	51	25	26	5	124
Labrador Control	5 yr	13 yr	7 yr 3 m	76	43	33	87	62	25	2	165
Rottweilers Case	1 yr	10 yr	3 yr 4 m	30	17	13	22	19	3	5	57
Rottweilers Control	5 yr	14 yr	7 yr 2 m	26	20	6	24	21	3	31	81
Staffordshire Bull Terriers Case	1 yr	10 yr	4 yr 2 m	6	1	5	5	3	2	2	13
Staffordshire Bull Terriers Control	5 yr	8 yr 3m	6 yr 4 m	19	11	8	19	15	4	0	38
											749

Key: yr = years, m = months

II. Sample submission form which accompanied each sample submitted

Appendix 2. Sample submission form which accompanied each sample submitted

Canine cruciate disease in Newfoundlands

Client Name.....

Age.....y.....m Gender: M F

Neutered/Spayed: YES NO

Referring Clinician.....

Address.....

.....

Tel.....

Kennel club number .....

Age of onset .....

Family history .....

Bodyweight .....

History of trauma .....

Owner Name .....

Dog Name .....

Cruciate information

Age of onset (index joint) .....days/weeks/months/years

Age of onset (contra lateral joint) .....days/weeks/months/years Joints affected: LEFT RIGHT BOTH

Radiographic findings (if available)

Index joint Effusion ☐ Osteophyte 0 1 2 3

Contra lateral joint Effusion ☐ Osteophyte 0 1 2 3

Synovial fluid analysis (if available)

Index joint WBC ..... $x10^9$/L PMN's ..... %

Contra lateral joint WBC ..... $x10^9$/L PMN's ..... %

Surgical findings

Partial tear ☐ Complete tear ☐

Med. Meniscal injury ☐ Lat. Meniscal injury ☐

Information sheet for owner

Great advances in veterinary medicine have been made recently and many of these have centred on new developments in body imaging, new treatments and surgical procedures, and the identification of genes which cause disease. Major developments in molecular biology have taken place in the last few years, making it possible to quickly analyse the DNA of both humans and animals. This is helping scientists work out what the underlying causes are for diseases and why some individuals become ill, whereas others remain well. Many of the diseases seen in companion animals including dogs, cats and horses are caused by a combination of genes from their parents (this is often referred to as "nature") and the external or environmental factors they have experienced during their lives (this is called "nurture").

Most scientists now accept that for the majority of features about ourselves and our animals are the result of a mixture of nature and nurture. For example, body weight, and height are in part caused by which genes are inherited and in part caused by our nutritional intake. In the same way, diseases such as diabetes in dogs, sarcoid in horses, and renal failure in cats are likely to be caused by a combination of both nature and nurture. The analogy often given to explain why such diseases develop is that of requiring both seed (nature) and the soil (nurture) before a plant can grow.

If researchers can determine which environmental factors (for example vaccination, diet, or drugs) and which genes interact together to cause disease, we may be able to make huge strides in improving animal welfare by, for example, advising owners which food their pets should have or avoid in order to reduce risk of developing a particular disease.

Researchers from all six UK Veterinary schools are now beginning to investigate the genetic and environmental factors underlying a wide range of diseases in companion animals. To do this it is important to collect large numbers of DNA samples from animals where the clinical features of the disease are clearly defined. The Vet Schools in the UK have agreed to work together in assembling one national UK DNA archive to avoid many small or duplicated sample collections.

The information collected will be kept strictly confidential. The samples and clinical data may be made available, through application to a review committee, to *bona fide* research groups working on projects that are ethically sound. It is also possible that samples will be made available to research groups working in partnership with non-academic industrial partners.

The DNA samples being submitted to the archive will be usually derived from blood leftover from routine pathology tests being performed. Samples will only be included if the owners give their written consent. The sample will be anonymous once it is entered in the archive. The owner retains the right to remove the sample from the archive in the future if so wished. No information regarding tests performed on the DNA sample will be given back to the owner. This is because it will only be possible to find out which genes and environmental factors are important by identifying patterns in large numbers of affected and unaffected animals.

Should you require further clarification of any issues raised please contact Wendy Hallows Archive coordinator Tel: 01517947208 whallows@liv.ac.uk

Owner Informed consent

1. I have read and understood the accompanying leaflet explaining the UK DNA archive for Companion Animals

2. I appreciate that in order to advance our understanding of veterinary diseases there is a need to determine how a particular condition relates to the genetic profile of the animal

3. I understand that any genetic tests relating to my animal will not provide specific information about its condition but will contribute to the general body of knowledge about the disease in the species. I realise that no specific information regarding genetic tests on my animal will be reported back to me.

4. I agree to DNA being extracted from blood that may be taken from my animal for the agreed clinical investigations and that is surplus to requirements for clinical tests. I agree that this will be used entirely for research purposes. I give consent for the material to be stored and made available to *bona fide* scientists in the field of animal disease and genetics.

5. I understand that all information I give will be held in strict confidence and the source of the archived DNA will not be divulged

6. I understand that this research will not benefit my animal directly, but in the future may be of benefit to other animals.

7. I understand that the custodianship of the DNA resides with the University of Liverpool but I retain the right to remove my animal's sample from the archive in the future if so wished.

Signed..... **Date**.....

III. Table detailing full gene name and abbreviation

Full gene name	Abbreviation
Aggrecan	ACAN
Anoctamin 4	ANO4
Biglycan	BGN
Calsytenin 3	CLSTN3
Cartilage oligomeric matrix protein	COMP
Cathepsin K	CTSK
Coiled-coil domain containing 85a	CCDC85a
Collagen type I	COL1A1, COL1 A2
Collagen type III	COL3A1
Collagen type V	COL5A1, COL5A2, COL5A3
Collagen type VI	COL6A1 COL6A3
Collagen type XI	COL11A1, COL11A2
Collagen type XII	COL12Al
Collagen type XIV (also named Undulin)	COL14A1
Collagen type XXIV	COL24A1
Cytochrome P450, family 39, subfamily A, polypeptide 1	CYP39a1
Decorin	DCN
Disrupted in Renal Carcinoma 2	DIRC2
Elastin	ELN
Fibrillin 1	FBN-1
Fibromodulin	FMOD
KALRN	Kalirin
Latent-transforming growth factor binding protein B2	LTBP2
Leucine rich repeat transmembrane neuronal 3	LRRTM3
Lysyl oxidase	LOX
Matrix metalloproteinase - 1	MMP-1
Opticin	OPTC
Popeye domain containing 2	POPDC2
Procollagen-lysine, 2-oxoglutarate 5-dioxygenase 1	PLOD-1
Ring finger protein 152	RNF152
Semaphorin 5b	SEMA5B
Serpin peptidase inhibitor clade B (ovalbumin) member 13	SERPINB13
Serpin peptidase inhibitor, clade H member 1	Serpin-H1
SIX homeobox 1	SIX-1
Sortilin-related VPS10 domain containing receptor 2	SORCS2
Stabilin 2	STAB2
Syntaxin binding protein 5 like	STXBP5L
Zinc finger DHHC-type containing 23	ZDHHC23

IV. Primer and probe sequences for the replication study detailed in Chapter 4

Appendix 3. Primer and probe sequences for the replication study

Plex #	SNP_ID	Forward Primer (5' - 3')	Reverse Primer (5' - 3')	Probe (5' - 3')
1	rs22051140	ACGTTGGATGGACAG GTATAGGTTCTAAAC	ACGTTGGATGCTTCTC AGATCATCCTCTCC	CATCCTCTCCTCACCAA
1	rs24496570	ACGTTGGATGAGTAC AGCCAGAGAACACAC	ACGTTGGATGGGCCC CAGAACACCAATATC	TTAGCTGTGCCACAGAC
1	rs23827731	ACGTTGGATGCTTAG GGAATTATGCTTGGG	ACGTTGGATGTGATCC CATCCTAACTCAGC	CCTAACTCAGCATTTCCA
1	rs9063975	ACGTTGGATGCCAGT GGCCTATTATGTTCC	ACGTTGGATGGTGGG ACTGTGATTAGGAAG	CAAGTGCTGTAGGTGATT
1	rs21971362	ACGTTGGATGTTCCCT TCCCACTTACGTAG	ACGTTGGATGTCTTCG TGTGTTGTTCTCCC	TGTTCTCCCCAAGGGTACA
1	rs23801997	ACGTTGGATGCCCTGT GTAAATCAAAGCTG	ACGTTGGATGCGGTT GGGATCGCTTTTAAC	TATACAGGGACGGAAGAAG
1	rs23048918	ACGTTGGATGCTTGA GGCTGTAGGATTTGG	ACGTTGGATGGAAAA TGTGGTAAACTTCCAG	GTAAACTTCCAGAAATGGTC
1	rs23935034	ACGTTGGATGAAGGA GTCTGCTCAGATGTG	ACGTTGGATGAACCAC CTTCTAATGCCAGC	CTTTCTAATGCCAGCCTCATA
1	rs9094558	ACGTTGGATGATGTCC TCGAGTTCTGTGTG	ACGTTGGATGTTGCAC TAAGAAACACAGGG	CCAAGAAACACAGGGCTCTAC
1	rs8568925	ACGTTGGATGGGTGT TCTTTGAATGCCGTG	ACGTTGGATGGGAAC ACCATGCACATGAAG	CAGGGGATAAGGATCCCGTGA
1	rs23813763	ACGTTGGATGAATGCT TGGTTGGACCACAG	ACGTTGGATGAAATCC TGCAATTTCGTGGG	CTCGCGGGGTTTAACGGCCAAC
1	rs23795335	ACGTTGGATGTTGACA GACAGTGGGAACAG	ACGTTGGATGCCCAG GGATTTCCTTCCAAA	GGTAAAAGTGCAAGGGACTAAG
1	rs23566023	ACGTTGGATGACCAAT GATGTGTGTGTCTG	ACGTTGGATGTTTTGA GGATGTGGCTCTGG	GCAGTGTGCACGCCTTCCATTTG
1	rs22074260	ACGTTGGATGGGAAC AAAAATAATGAGCTG	ACGTTGGATGAGCTG GTAGATAACTGTGTC	AACAGAAAAGTTGCATTTGATTT
1	rs23812432	ACGTTGGATGATTAAT GGATCACAACCTGC	ACGTTGGATGGACATC TCTTTTTCTTCAG	CATCTCTTTTTCTTCAGATATTTT
1	rs23835082	ACGTTGGATGTGGGT AGACTGAAAGCAAAC	ACGTTGGATGAATGTA GGGAGCTGAGAGCG	ATAAAAGCTGTTTGGCCAGAATTG
1	rs23333120	ACGTTGGATGGGGAT ATGGGTATACAAGGG	ACGTTGGATGACACTC CCTTCCTCATAGAC	CTACCATGACGGTGAGTCTATCTC C
1	rs23829198	ACGTTGGATGTATCAC CAAGTGGTTGCCTC	ACGTTGGATGGAGGA ACAGAAAGGAGGAAC	CTGAACAGAAAGGAGGAACTAAT GT
1	rs9253138	ACGTTGGATGTACTTC TTGACATTCCGGAG	ACGTTGGATGGGGCC ACTGGTTTATAACAC	CGCCACTGGTTTATAACACATACTT C
1	rs22418471	ACGTTGGATGACATCC TCACCAAACTCAGC	ACGTTGGATGCGCAC ACATCATAGAGAAAC	TGAACAAAATTATGGGCAATCCTG TC
1	rs22017302	ACGTTGGATGAGTAG	ACGTTGGATGCCCACC	GAAAAGTTAGAAGATGCCAGGGA

Plex #	SNP_ID	Forward Primer (5' - 3')	Reverse Primer (5' - 3')	Probe (5' - 3')
		GTTCTTTGTAGAGGC	GTATGATTACTCTG	CAC
1	rs21960910	ACGTTGGATGGAAAT AAAAAGGTCTACGGG	ACGTTGGATGTAGAAT CCTGCTTCCAGCAC	CTCTCCCAGCACTCTGTAAAAGCA AAA
2	rs8748659	ACGTTGGATGGGCCT CAGTATTACAAAGAC	ACGTTGGATGGATTTG ACATCAGAGATCAC	CACTTTCATGCTTCGTC
2	rs23804256	ACGTTGGATGCCAGTA GCTACCATTTGTGC	ACGTTGGATGAGCCA ATAATGGTTGCAGGG	TGCAGGGATGGCTTTTA
2	rs23835584	ACGTTGGATGTATTCA GGACTGCGTCCTTG	ACGTTGGATGACGCTA GACTCATACACCTC	CTCATACACCTCAAGAGA
2	rs21956708	ACGTTGGATGCAGCC ATGATACCTGAACTC	ACGTTGGATGTCCTCA AATGACATCACCCC	TCATTCTCCCCACAGCAGG
2	rs23569097	ACGTTGGATGGTATG CAAGGCTGGGAAAAG	ACGTTGGATGAACCA AACACAGAAGTCCCC	TGCCCCACTGTAAAGGTGA
2	rs21976781	ACGTTGGATGTCCCAC TCCCTCTGCTCCT	ACGTTGGATGCTGTG GACTGCTTTATGAATC	TTTTTGAGAGAGTGTGAAC
2	rs22035867	ACGTTGGATGAAAGA ATCTCGTGAGTGAGG	ACGTTGGATGTCAAAA CAGGCTCCCCATAC	CCCCATACTCTACATGGAAA
2	rs8987872	ACGTTGGATGTGCGT GGTTAACAAAACTGG	ACGTTGGATGGATGC TTCTGCAAGAAAGGG	CCTTGGCCTCAACGTATCCCA
2	rs8498499	ACGTTGGATGCCTGAA GCTGAGTGAGGTTG	ACGTTGGATGAAAAT AGCCGGTCTGGCTTC	AGCCGGTCTGGCTTCTATTCT
2	rs24526391	ACGTTGGATGAGTGA GATACTAACCTGGAG	ACGTTGGATGCACTCA AGTAGTGACATCATC	GAGTTTGTGACAATCAAGTTA
2	rs22030434	ACGTTGGATGTGGAA AGCAAAGTGAGGCTG	ACGTTGGATGAGTAA CTGTGAAAACTTGG	CTTGGATCTAATTGAAGCTCCT
2	rs21882305	ACGTTGGATGATTTTG CATTTCCTCTTCC	ACGTTGGATGCATCTC TCCATGGACCAAAG	TTCCAGAGATAGAGAAATAGAT
2	rs21880947	ACGTTGGATGGCTGA CTTTCTCCTTGAGAC	ACGTTGGATGGGCTT GGAACATTCCTTCAG	TTTCCTTGTTTACTAGAGTACTG
2	rs24114482	ACGTTGGATGGTATCA ATCCAGGCAGCCAC	ACGTTGGATGCATACT TCCAAATCTTCGGTG	CATCTTCGGTGATAATGGAGACA
2	rs9122813	ACGTTGGATGGGCTTT CACCTGCACAGTTA	ACGTTGGATGCAGCAT TCAGAAGACTGAGC	GGAATGTTGCAGCCTAAAGTACCC
2	rs23828846	ACGTTGGATGACTGCA GACTTCCTATCTGG	ACGTTGGATGAGGTC AATCAGACAGTGGTG	GGCAATCAGACAGTGGTGAGCAG C
2	rs21946491	ACGTTGGATGGCCCTT ATAAAAGGTCATTGC	ACGTTGGATGTTCCTT CTAGAGGAATATCG	GGAAGAGATAAAGAGAAAGAAAC A
2	rs21895425	ACGTTGGATGCACCAA TTAGTGCACAAAGC	ACGTTGGATGCGAGG CTCATTAGTAATCTC	TTGATTTCTTCTGAAAGTAAGTCAA
2	rs23800347	ACGTTGGATGTGGAG TTTGAAACCCAAGCC	ACGTTGGATGAAACT GGGAAATTCTGAGAC	GGAACTTCTGAGACTAGAATTCAT CG
2	rs23798518	ACGTTGGATGTCATAG ACGCAGTACATGCC	ACGTTGGATGTGAGA TGTGGAGTGTGGTTC	CATTCCTGGGGAAGATGCTTGATG GT
2	rs21882298	ACGTTGGATGCTCGAT AATAAGATAGCCAC	ACGTTGGATGCACTAA CACTCAAGGACTTTG	CCCCATAAATCCATTCTGAGCTTAG TT
3	rs23799015	ACGTTGGATGTGAGA AAAGCGATGAAGCTG	ACGTTGGATGTCACAA TGTGGACCTGAACC	TGAACCTTCCAACCTCC
3	rs23048917	ACGTTGGATGGCACA	ACGTTGGATGAAACCC	GGTGTGAGCACTGTCCC

Plex #	SNP_ID	Forward Primer (5' - 3')	Reverse Primer (5' - 3')	Probe (5' - 3')
		GGCGATAAAGCTTTC	ATGATGCCTCACTG	
3	rs23811087	ACGTTGGATGCCCAG GCATCCCAGTTAAAG	ACGTTGGATGAGCCCT TTCCTACTTTTACC	CCATTTAGTTGTCACTCC
3	rs21973355	ACGTTGGATGAGTACC ACACGTGTTCTCTG	ACGTTGGATGATCAGT GCTCTCACACCTAC	GCAGTGTACGTTAGAGTT
3	rs22184669	ACGTTGGATGACCTGT AAGTTACCTCTTTC	ACGTTGGATGCTATGA ATTCATACTCCTGGC	TCATACTCCTGGCATAAAA
3	rs8959368	ACGTTGGATGAGTGC TCCTGACTTCCTCTC	ACGTTGGATGGGTATT TTAGTGAGTGCCTG	GTAGTGCCTGAGAGTGAGA
3	rs23605299	ACGTTGGATGAACAA AAAGGAGGAGAAGCG	ACGTTGGATGTTCCTG AGACGTAGGAGGAG	ATTCTGAGGCGATCGGATTC
3	rs23806044	ACGTTGGATGCCAAA GACAAGCAGGTCTTC	ACGTTGGATGTGCTGC TCAGCAATTCTCTC	CGCAATTCTCTCTTCCTTTTG
3	rs22014075	ACGTTGGATGGCTTGT CAAATAAAACACTG	ACGTTGGATGACCCAT TCCTCTATTGCCTC	CCTCTGAATTGATGTCTTGCC
3	rs8712319	ACGTTGGATGCCCTTC TTGAGAAGTTTCCG	ACGTTGGATGGTCAA ACTCCCCTGAAACAC	CCCCACCCTGAAACACCCCACC
3	rs8647709	ACGTTGGATGGGATC AGATTTAAAAAGGC	ACGTTGGATGGGGCA GTGAGTTCTGGTTTC	ATAAAATGTCTTCCTGAACTTA
3	rs22346338	ACGTTGGATGTTGGA CAGGGGTCCTCAAGT	ACGTTGGATGTGCAG TGTGCTGGCTTTGAG	GGCGCGTGCCAGTGTTCATGGT
3	rs8754330	ACGTTGGATGAACCCT CCTGATGCAAACTC	ACGTTGGATGTGTGG ATTGCTAGTGCACAG	CCTGTTAGAAGCAATGCCCATCC
3	rs21946171	ACGTTGGATGAACATC CTTAGTGTCTCTAC	ACGTTGGATGTGCCCC TTCAAATCACTATG	GAGGAGCAAGCAGTGTTTCTTTA
3	rs8613571	ACGTTGGATGGAGAG AGGGAGCAATATCTA	ACGTTGGATGTGAATC ATGAGGAAAGAAGG	GGACTGAGGAAAGAAGGCCCTAG
3	rs21987554	ACGTTGGATGGGGAA TTGCATGAAGGAGTG	ACGTTGGATGTTCCAG TTGCAGACTCTTTC	TGCAGACTCTTTCTATGTATTTTG
3	rs22383474	ACGTTGGATGCTATAC TCCTCCATGGAACC	ACGTTGGATGCTGCA GAATCTTGGTTCTGG	CCTTCCTGGCCGTCCCTTCTTTGTT
3	rs21998646	ACGTTGGATGTCTCGC TTATCTGCAAGGAC	ACGTTGGATGGGGCA CATTCATGAGTAAAG	CCTACTTCTTAGAATAAGTCTCTGA
3	rs23801537	ACGTTGGATGCGGGA AAAAGATTGGATACTC	ACGTTGGATGCTAGAC CTTTAGACTGATA	CCGACTGATATTTAAAGCAAGAAA T
3	rs21977372	ACGTTGGATGTCAGA GCAGGACTCACTTAG	ACGTTGGATGGGAGT ATTGTAACCTTTCTAC	CTTCACCTTTCTACTAATAAAAACC T
3	rs22380841	ACGTTGGATGACAGT AGCCCTAGATCTAAC	ACGTTGGATGGTGATT CTGGATACATTGCC	GGTTTAATGACTGTCAAATTCTTG AA
3	rs21880470	ACGTTGGATGCTCATT GCCATTCATTCTGC	ACGTTGGATGTCCATG TCAGGACAGAACAC	CCCAAAACAGCCATTTATACATATA AT

V. Primer and probe sequences for the candidate gene study detailed in Chapter 5

Appendix 4. Primer and Probe sequences for candidate gene study

Plex #	SNP_ID	Forward Primer (5' - 3')	Reverse Primer (5' - 3')	Probe (5' - 3')
1	rs23942105	ACGTTGGATGAAACGA TGGCAGGTGGTTC	ACGTTGGATGTGAGTTT CCTGTGCTTCCCC	TCCTCTGAGAGCCCA
1	rs24048819	ACGTTGGATGCTGAGC ACCCAGTACCAAG	ACGTTGGATGCGCCTG GTGTGACTCTATTC	CACTGAGGGGTTTCC
1	rs22868921	ACGTTGGATGTACAGA AGACCCTGTGGATG	ACGTTGGATGACAGAG GCCAGGCCCGTCAA	CGCCCCCTCGCACGCC
1	rs24260515	ACGTTGGATGGAACTT AGTCCCCTATACAC	ACGTTGGATGACCCAAT GGACATCTAAGAG	GCGTGCGTGCGTGCAT
1	rs24469329	ACGTTGGATGTACTCG AGCTCACATATCCG	ACGTTGGATGTAGCTCT TCTGTGCTTGCTC	GGCCAGTTCCTCCTCTT
1	rs22255739	ACGTTGGATGCTTTTCC AGCTGTTCCAGTC	ACGTTGGATGAGGGAG CAGCAGGAGAGAC	CCAGGGCCCCCAGGACC
1	rs24535768	ACGTTGGATGCCTGGT GAATAATAAGAGGG	ACGTTGGATGTTAGGG AAGGGTCATCAGAG	CTTGGAGAATTGCATGG
1	rs24730359	ACGTTGGATGTCTCTG GCTTCAGGTTGTTG	ACGTTGGATGAGCTACT GGAGATGATGAGG	GCCAGCGTCAAGTACCAG
1	rs22957764	ACGTTGGATGTATCCC CTGCTCAGCTTTTG	ACGTTGGATGCTGTGTC CTAGCAGGGACAT	CTGCCTGCCCAGCAGTCCC
1	rs9049305	ACGTTGGATGCCTTGG AGATCTCTTGACTG	ACGTTGGATGACTTACA TTGGCTCCAACAG	GTTGCTCCAACAGGGCCCA
1	rs8721918	ACGTTGGATGAGCTTT CTGAAGCCAGAGTC	ACGTTGGATGTGGCCCT GACACCTTCTTG	GACGTCTTGCCAGGAGGGT
1	rs8652327	ACGTTGGATGGTGAAG TAATGGTCTGCTGG	ACGTTGGATGGACCTT GGACTCCATCTCTC	CCCTCATCTCTCCCTGCGGG
1	rs22229903	ACGTTGGATGCTTCCA CGGTGTTGAAAAGC	ACGTTGGATGAGCCGT GTGAGCGTATATTG	TGAGATTGCGAAGACTGCTA
1	rs8867454	ACGTTGGATGTGATCA CCTTTGTCACCTCG	ACGTTGGATGAATTATA GTCTCACCAAGC	TGTATCACCAAGCAATATCCA
1	rs23527990	ACGTTGGATGTATGAC GCCATCTGCTACAC	ACGTTGGATGTTCTAAG CTCCCCTCCTCTG	GCACCAATTTCCGTGGTGCTA
1	rs22876135	ACGTTGGATGTCACGT AATGTCACTTTGGC	ACGTTGGATGCTTTCCT GTGTTGAGAGGTG	TGTGTTGAGAGGTGAGATCTA
1	rs24048798	ACGTTGGATGGGCCAC TGCTGCTAGCTCA	ACGTTGGATGAGTCTGT GCCTTTGGCAGGT	GCAAGGTGGTGAGGCCCCTTGC
1	rs8706610	ACGTTGGATGAGACGG GAATGCATGGAATG	ACGTTGGATGACAACG TTTTGCAAGAGCAG	ACAAAACAGTTCCCTTGTTTAAT
1	rs24546719	ACGTTGGATGGGGTCC GGGAAGTCCAGG	ACGTTGGATGTGCCCTC TCCCCTCCGCTG	GGTTTTGGCCGAGATGGCATCCC
1	rs22578989	ACGTTGGATGTTTGCA TGGATAACTGTGCC	ACGTTGGATGGTGAGT ACTCTGCTTCACA	GGGGTTGCTTCACAGCAGCTAAT
1	rs8957513	ACGTTGGATGCTCAAG	ACGTTGGATGTTGAGC	GGACAAGGTTATGATGTCAGGGT

Plex #	SNP_ID	Forward Primer (5' - 3')	Reverse Primer (5' - 3')	Probe (5' - 3')
		CAAACTCTGTGCTG	ATCAGACTCAAGGG	
1	rs22270010	ACGTTGGATGGTCATG TCTGACTAATGGAG	ACGTTGGATGGAGCTG AGGCTAAAGTAGAG	TTTGAAAGTGAACCTCTTTCCCAA
1	rs22402697	ACGTTGGATGGTCCTT TGTCAGTTTTTTTT	ACGTTGGATGTCTCCCT CTTTCTGTGTCTCTGC	TTGCATCTCTCTCTCTCTATCTCTG
1	rs8907726	ACGTTGGATGCCAAGT CCCGTTATTCTCTC	ACGTTGGATGTGATCG CATCCAACTTGCAG	GGGATCAACCCAGTACTCTCCTGAA
1	rs22214751	ACGTTGGATGAAATCC TTGGGTGAACCACG	ACGTTGGATGTGGATG TCCACCTGTTGAAG	ATCTAGTTGACATTAAAAAATGATT C
1	rs23269193	ACGTTGGATGTTCCTG GTCCTCATCTCATC	ACGTTGGATGGCAACA CCAAACTGCTTGAG	CCCTTCTGCAGAATCATCTACTTCA TC
1	rs24324848	ACGTTGGATGTTTGGC ACCAGATGAACCAG	ACGTTGGATGGTTAGG GAATTATTACTCAG	AATGATTATATTCTTACTTTTTCTCT A
1	rs24535784	ACGTTGGATGGTATTT ATTGTGAGGGATTGG	ACGTTGGATGCTGCCTT TAGACTTCACTGC	GGGTTGGACTTCTCAGGCTCCATAT CT
1	rs8615753	ACGTTGGATGCTCTCTC TCTGTGACTATC	ACGTTGGATGAGCAGC CTATCTCCATTTTG	CCCATTTTGGTTTTTTTTTTAACTTTT T
2	rs22197326	ACGTTGGATGTGGGTG TCGGTAGGAGACAG	ACGTTGGATGAAGCAA GGCTCAGAGCAAAC	CACAGTCCCCCACAC
2	rs23644868	ACGTTGGATGATGCTA GGCCACAAAATAAG	ACGTTGGATGTCAGGG TCAAGTGTTTGAGC	GCCTCATGTCGAGTG
2	rs22816235	ACGTTGGATGGTTAGG AAAGCCCGTGTGTG	ACGTTGGATGAAGGGA ACACAGGTCCCAC	ACTCCCCTTCATGCCC
2	rs22330976	ACGTTGGATGGTATGC AGAAAGTTATGAG	ACGTTGGATGGCTAATT ACCATCTCCCTCG	CTCTCCCTCGGCATTG
2	rs24048807	ACGTTGGATGTAGAGA CCATGCAGACAGAC	ACGTTGGATGTTGATG GGATTGCCATCCAG	GGGATGTCTTCCAGCC
2	rs24048794	ACGTTGGATGAGCCAA CTGCAAAACTCACC	ACGTTGGATGCTGTGT GTATATGTGTGTGC	TGCTCGGGTGTAGCAG
2	rs22378757	ACGTTGGATGCCCCTC TATACCACATAAGC	ACGTTGGATGGCAGCC AGAAAGTTTATCGG	ATCGGCCTTTTTTCTCC
2	rs24448073	ACGTTGGATGTTTCCA CTGCTTCTGCAGCC	ACGTTGGATGACCATTC CAGGCCTACTGAC	CGGCCTACTGACCTGTGC
2	rs24754119	ACGTTGGATGTGGACC CTACAACAGGTAAG	ACGTTGGATGATGAAA TGGGAGGAGAGAAC	TGAGGAGAGAACACTGAG
2	rs24306291	ACGTTGGATGGGTCCC TACAAGAAAATGAG	ACGTTGGATGTTTCTGT GTGTAGCCTCAAC	TTGTAGCCTCAACTACATT
2	rs24770733	ACGTTGGATGAGCACC AAGAAACAAAATG	ACGTTGGATGCCCGTA AGTTGTGATCATGC	ATGACTGTATAGCTGAACT
2	rs23624676	ACGTTGGATGCATCTTT GTGGTCTCAGAAC	ACGTTGGATGCCTTTTC TGTAACTTTTCTC	AACTTTTCTCTTGACATCAT
2	rs8686504	ACGTTGGATGCAGAGA TTTAGATGCCCTGC	ACGTTGGATGATATCTG GACCACCCGTCAC	ACCGCCACCCCCCTGGGGTA
2	rs24278934	ACGTTGGATGCTAGCC AAGGCAGTATGAAG	ACGTTGGATGGTGCAC CATCTCTTTCTACC	CCTAGAGAGAACATTCTAGAT
2	rs24602445	ACGTTGGATGAAAAGG GAGAACCTGCCATC	ACGTTGGATGACGACC CACATGTCTTGCAG	TGGGATAAGAAAATGCCTTAC
2	rs24306964	ACGTTGGATGGTTTTT	ACGTTGGATGAGGCCT	CATCTCCAGAGAGACCTCTTCC

Plex #	SNP_ID	Forward Primer (5' - 3')	Reverse Primer (5' - 3')	Probe (5' - 3')
		ACTGGGTTACAGGG	GTCTCTCCTTTTTC	
2	rs22919040	ACGTTGGATGTAAAAA TGGCCTCCTGGTCC	ACGTTGGATGCTTGGCC TCTAGCAACTCAC	GGCGCTAGCAACTCACCTTCTT
2	rs8946670	ACGTTGGATGAAATGA GCAGTGGATCCGAG	ACGTTGGATGAGGTTC AGTACCTGGTCTCC	GTAACTGGTCTCCCGTCGCTGTA
2	rs23906411	ACGTTGGATGTGCTGT GTCTATATACTCTG	ACGTTGGATGCGGTAG GACCCTGAAGAAAA	AATTTTAGAGGTAAAAAAGCATC
2	rs24560587	ACGTTGGATGGATTCC AGGTGAGGCCTCAT	ACGTTGGATGGAGAGC AGGATTAGAATCAG	AACAAGGACAAAGGGGACCCCATA
2	rs24575600	ACGTTGGATGATATCC ACGATGGGCACTTG	ACGTTGGATGTGCTTCC CAGACGAAGAAAG	GCATACGAAGAAAGGGTATCAGAA
2	rs8697617	ACGTTGGATGGCTCTC AGAATTCAGCAACC	ACGTTGGATGGGTGAA CAAAACCATTAACCC	TAAAAAAACCATTAACCCCAAACAC
2	rs22842378	ACGTTGGATGCTTATCT CCATCTCAGGGTC	ACGTTGGATGTATCACG TACCTTCACACCC	CCCACCCCGAGGTCCAGGATAGCC C
2	rs24748955	ACGTTGGATGGAGTAT ATCTGGGAATTTC	ACGTTGGATGGGGCCC ATCTGGAAATATAG	TGCAATAAAATATCTTTCACATATA A
2	rs8946915	ACGTTGGATGGACAGA TGGTATATAAGACAC	ACGTTGGATGGAGCTC TTGTCCACTTAAAC	CTTCAACATTTAGCAATAGTGCAAT T
2	rs8879536	ACGTTGGATGAATATT CCACCACTGGTCTC	ACGTTGGATGCCGTGG AATTAGTGGTGTTC	GCCGTTGAGTGTAAGTTACTCGATT T
2	rs22929544	ACGTTGGATGAAGACA AAAGGAGGAGGCAG	ACGTTGGATGAATCCTC CTTTGTTCCCCTC	CTGTCCCCTCCCCAGACTAGGCATT C
2	rs9222557	ACGTTGGATGGTTTTTT TTGGTTTTGTTTG	ACGTTGGATGCAGGTA ACATATTTTCTCC	TTCTCCATGTCTTTTTATTTTTCATAT A
2	rs22786371	ACGTTGGATGAGAGG AGGTCCCTGTCCCCA	ACGTTGGATGGAAGAA ACAAGTCTGCCCTG	TGGACGAAACAAGTCTGCCCTGGG ACGA
3	rs23627318	ACGTTGGATGGCACAT ACTGCTTCTCTTCC	ACGTTGGATGAGCACC TTTATGCCTCACAC	CCTCACACTCAGGCT
3	rs24654871	ACGTTGGATGCTTAGG GAGGGGAGCTGGG	ACGTTGGATGAGGCAC TGCGGACAGAGAGA	AGAGGGGCGGCCTCC
3	rs24299845	ACGTTGGATGACCCCG TATGGACAATAAGC	ACGTTGGATGAACCTG GAGAAGCAGGGAAC	AGCAGGCACAGGCGT
3	rs23942103	ACGTTGGATGAAACGA TGGCAGGTGGTTC	ACGTTGGATGTGAGTTT CCTGTGCTTCCCC	TGCTTCCCCGGCTCCT
3	rs8721919	ACGTTGGATGAAGCAG CACAAAGCAGTGAC	ACGTTGGATGATGGAG AGGGATGAGAGGAG	TTGGCCCTGACACCTTC
3	rs8861243	ACGTTGGATGGGGAG GTATGAGATCAGAGC	ACGTTGGATGGAATGA ACAGAAACCAGGGC	CATAAGAGTCCTCGGTT
3	rs22930994	ACGTTGGATGTTCCTTC CTTTGGGTCCTTC	ACGTTGGATGCAGGGG ATACTTACCTTGGC	ACTTACCTTGGCTCCTTT
3	rs22597612	ACGTTGGATGAACTCC TCCAACTAAGCCAG	ACGTTGGATGTCAGAT ACTGTGCTAGGTGC	GCTAGGTGCTAGAGACAG
3	rs22217603	ACGTTGGATGAACAGT GACTTACAGGCTCC	ACGTTGGATGCATGTGT GTCTCCTTCAGGC	CTCCCCTTCAGGCTGCCCG
3	rs9222556	ACGTTGGATGATGGCA CCCATTGGACCAGT	ACGTTGGATGGTTGTA GGTTTCTCATACTG	ACCTCATTTTTGCTTTAGG
3	rs24029016	ACGTTGGATGGCATGT	ACGTTGGATGTCAGGA	GGAAAGAGGATGAGAAATG

Plex #	SNP_ID	Forward Primer (5' - 3')	Reverse Primer (5' - 3')	Probe (5' - 3')
		GCAGAATGTCCTTC	GCCGTAAGGTTTTC	
3	rs22861513	ACGTTGGATGGCCATC ACAGACACAGTTTC	ACGTTGGATGTCTCTGC CATCCTCTATACC	TTAGCACACACACAACCCAA
3	rs22738485	ACGTTGGATGAACTTC TTCAGGAGCTCTCG	ACGTTGGATGTCTGTG GCTGCAGTTATGAC	TGCAGTTATGACGTGGTGTT
3	rs22190862	ACGTTGGATGCATCTG CCCTGTATGCTGAG	ACGTTGGATGAATGGA AAGGGTTTTGCTTG	TGAAAGTACTAATGTTTCCTT
3	rs24566255	ACGTTGGATGCAGTGC TATTGACAGTAGCC	ACGTTGGATGATACCAC ATCTGCTTTATCC	CCAATCTGCTTTATCCGTTCAT
3	rs8824603	ACGTTGGATGACTATG GCTCCAAGGAGAAG	ACGTTGGATGTGTATTC AATGGCACTGCCC	CCCTATGTGTCATTCTTCTGCT
3	rs24278947	ACGTTGGATGATTGGA AGTGTCAGAAGGTG	ACGTTGGATGCTTATAA GAGTTCTGCTGCC	GTCTGCCAAATCTTCTGTAATG
3	rs22239815	ACGTTGGATGATTGCA TCAAAGTTGGCCAC	ACGTTGGATGCTGCAG ATGCAAAAGAACTC	GGGCAAAAGAACTCAAACAAAT
3	rs22184922	ACGTTGGATGCGCCTT CATCCTTGAGTTTC	ACGTTGGATGGGTGTT CTCATCACTGATGG	CGGATGGAAAATCACAAGATGA
3	rs8680883	ACGTTGGATGAACAGG ACGTCTTGGTCCC	ACGTTGGATGGTGGGA CCTTTGTATGCATC	GTACCTTTGTATGCATCTCCCTCC
3	rs8945015	ACGTTGGATGGAAGCT CTTGCAGTGTTGTC	ACGTTGGATGTCACCTT CTGTGTGCTAGAC	CTTTGGCACTTCGAACCACTGCTT
3	rs24736591	ACGTTGGATGTGTTTA CATGTTGGAACAC	ACGTTGGATGATCTTGA GGCTGATGAAGTC	TGTTGACTTCAGTTTAACTAATGG
3	rs9095802	ACGTTGGATGACACTC ACCTTGTCCCCATC	ACGTTGGATGAGGGAA AGATGGGATCCCAG	GGGGTCCCAGGGCCCCTGGGGCTT
3	rs24522316	ACGTTGGATGGAACTT TTGCAGCAGTGCCG	ACGTTGGATGTTCTCGC CCCCGCTGGCAC	GATGGCACCCGCCACCGCGTCCTC CC
3	rs24535705	ACGTTGGATGAAACGC GGGTCTAACTTAGG	ACGTTGGATGTTTCTCT GAGCATGCACCTG	ATGACCCATCCCCACACATGTGAAA A
3	rs8621422	ACGTTGGATGTGGATG TCATCTGTGTCTGG	ACGTTGGATGACACTG ACTTTGGGCACAAC	AACTTCAAAATGAGATTTTAAGAGT T
3	rs24029017	ACGTTGGATGAGGACA TTCTGCACATGCAC	ACGTTGGATGTGTACAC AGACACACCGATG	CCACACAGACACACCGATGCACAG ATA
3	rs22378728	ACGTTGGATGGGGCAC GTGACTACAGAATC	ACGTTGGATGCACTGG TAAAACTTTCATTGC	ATTGCAATATCAATTTCTTTGTAAA CA
3	rs22309803	ACGTTGGATGGAGCGT GGAGAAAGAGTAAG	ACGTTGGATGCCTTTGT CAGGCTAATGCAG	ACCGTTGTCAGGCTAATGCAGTGA TAG
4	rs23269200	ACGTTGGATGTCCCCA TGGCAGCGATGAC	ACGTTGGATGACCCTGC ACTTCTCTCTCAC	CTCTCTCACCCCCCC
4	rs8686503	ACGTTGGATGGTTGAT TGGCTGAGAGGTTG	ACGTTGGATGCTCATCC TCACCTTTGCTTC	TTGCTTCCCCTCCTC
4	rs22863806	ACGTTGGATGAGGGAT GACAGGCCCAGTG	ACGTTGGATGATCCCTC GAAGACCCTCAG	AGGCCCCATACGTCC
4	rs8891835	ACGTTGGATGTTCCTG GTCAACCTGGTGC	ACGTTGGATGAGTAAT GCCCAGAACTCACC	GGATGGATGTCCGGG
4	rs23639382	ACGTTGGATGTGCCCA GAAACTGCACATAG	ACGTTGGATGTCAGGC ACTTAGTGAGTAGG	AGGGTCACTTCACTGT
4	rs23589114	ACGTTGGATGTGACCC	ACGTTGGATGTAGAGT	GTTGTCCTGCACCCCAC

Plex #	SNP_ID	Forward Primer (5' - 3')	Reverse Primer (5' - 3')	Probe (5' - 3')
		TCTTGCCAAAGCAG	GACCAACTGGTCTC	
4	rs24466233	ACGTTGGATGACAGAG GCAGAGGAAGGAG	ACGTTGGATGGTGAAC GAGTGCGAGCTCAT	CGCATGCTGGCGGTGTG
4	rs23242714	ACGTTGGATGGGGATT TTCCACCCAGAAAC	ACGTTGGATGCTCTTTG TCAAGTCTGCTGG	CGGAGCCGGATAGAAGA
4	rs24288752	ACGTTGGATGGGCTTC AGAATAAATCAGAG	ACGTTGGATGGGGAGA AAACATTGGTAGAG	TCTACTGCTCTGACACTG
4	rs24448071	ACGTTGGATGTTTCCA CTGCTTCTGCAGCC	ACGTTGGATGACCATTC CAGGCCTACTGAC	CCCTCTACTGACCTGTGCC
4	rs22823788	ACGTTGGATGATCTAA AGGCCTCTCCTTCC	ACGTTGGATGCCAGCG CCAGAGAACCTG	CCCAGAGAACCTGAGAGAC
4	rs22378703	ACGTTGGATGATTGGG CATGCTGAGGTTTG	ACGTTGGATGGCCTAC ATGCTCAGGATTAG	CATACAATGGGTGGTTTTT
4	rs24607603	ACGTTGGATGAATACC CGCCATTTGCTTCG	ACGTTGGATGTGTTTGT CCTTCTCCGATTG	TTTCACTCAGCATAATACCCT
4	rs8680882	ACGTTGGATGTGGAGA GGTTTAGCAAAGGG	ACGTTGGATGATGCAT ACAAAGGTCCCACG	GGCCAAGGTCCCACGGGCCTC
4	rs22308173	ACGTTGGATGGCCTAT ATGGATGAGGAGAC	ACGTTGGATGTCGGCA ACCAGTTCAACATC	CATGGGAACCTTGCAGAATGA
4	rs24546410	ACGTTGGATGTAATGC CCACCTTCTCTCTC	ACGTTGGATGTCATTCC CTCCTGAACCCAC	CCCTCTGGCCGCCCTTCTCTCC
4	rs24052183	ACGTTGGATGTCTGTT ATGTTCCCAGGCAC	ACGTTGGATGCACTCTC ATTAGATGTGCCC	TGCCCTATTTATTTCTGCATAC
4	rs24278933	ACGTTGGATGCCTTCT GGAACTCTAACAAG	ACGTTGGATGGAGTCA AAGATGTTTGTCCC	CTCTGTTTGTCCCATTCAATTCA
4	rs22372241	ACGTTGGATGAACCAG GATCTGGTTTTTTG	ACGTTGGATGATTTGG ATCCTGTCTCACTC	CTCGCTGATGAGTCTGGCTAAAG
4	rs22803160	ACGTTGGATGCTGGCA CTTAGGAGTCTAGG	ACGTTGGATGCTGGAA ACCTAGCAACTGAG	GGGTACCCCCGCTCCATCCCATCG
4	rs24221465	ACGTTGGATGTTCCTT GACCAGCAGAACAC	ACGTTGGATGTGGCAG TCCATCACCTTAAC	TAAGTCCATCACCTTAACTAAGAA
4	rs22320220	ACGTTGGATGAGAGG ACCACGTGGAGAAAG	ACGTTGGATGCTGTCA GTTTTCTGTAGTTTC	TAACATGTAGTGTTTAAATATGGT
4	rs8898239	ACGTTGGATGTAAGTA ACCTACCACAGGAC	ACGTTGGATGGATTGCT TTGAGAGGTCCAC	CCCAACCTGGCCCAATGGGTCTAAC
4	rs22957760	ACGTTGGATGGAAATC CCCAGGACAAAAGC	ACGTTGGATGTCCTTTG AGCTGCCACTTAC	CATTATCCTTCCTCTCAGGAAACCT A
4	rs24278005	ACGTTGGATGTGATAA GCTCTGTGAAAGGG	ACGTTGGATGGCTCGCT CTTTTACTTTTGG	ATCTTTTACTTTTGGTAAGGATAAC A
4	rs8679272	ACGTTGGATGGGAAGT ATAGATATAAGGAAG	ACGTTGGATGAAGCCC CAGTGTTGACTTAG	TGACTTAGATTAGTTGTATTACAAT G
4	rs22555948	ACGTTGGATGGTGGTC CATTCTAGTGGAAG	ACGTTGGATGCCTTAAA CACAAGTTGGACC	GCCAGATTACATTCCCAATCTGAAA CT
4	rs23537789	ACGTTGGATGGGTGCT GGGTAGATTTCTTC	ACGTTGGATGCCTACAA AGTGAAGAATGGC	GATTAATTTTAAGTTCATGACTTTCT AA
4	rs24568755	ACGTTGGATGAGCTGC TCTGATCTCTGAAG	ACGTTGGATGGAGGGG CCTGATTAAGATTC	GGACTGATTAAGATTCATTATTTCT GAT
5	rs24048809	ACGTTGGATGTTGATG	ACGTTGGATGTAGAGA	CCCGAGGAGCACAAA

Plex #	SNP_ID	Forward Primer (5' - 3')	Reverse Primer (5' - 3')	Probe (5' - 3')
		GGATTGCCATCCAG	CCATGCAGACAGAC	
5	rs9117174	ACGTTGGATGATCCTG CATGAGGTAGGAAG	ACGTTGGATGCCTTAGC TCATAAAGCAATC	TCTTTGCCTCAACACT
5	rs22881468	ACGTTGGATGTTACCC GATCTCCTTTAGGC	ACGTTGGATGCTCTACA CTCTCTCTTACCC	CCCTGGCTTCAAGGGC
5	rs23637297	ACGTTGGATGGCTTTT GTTCTGGAGATTGG	ACGTTGGATGCACAGA ACACAGCAGAAAAC	ACAAGGTTCATAGCGAA
5	rs22738509	ACGTTGGATGATTTCTC TACCCAGAAGCCC	ACGTTGGATGAGAAAG CTCTGGAGAAGCAC	AACGCGGACAAGGAGGA
5	rs23648407	ACGTTGGATGGTGTGA GTGAAGAGTACCAG	ACGTTGGATGGTTGGC ATGAAGAATCCCAC	CAAGAATCCCACCATGAT
5	rs9008179	ACGTTGGATGCTTTTG CAGACTTTTTGTGAG	ACGTTGGATGCTTGATA ACTGAGGACCTGC	TGATGGCAGGAGATTATT
5	rs8950487	ACGTTGGATGCAAGGT TATCGATGAGCTGG	ACGTTGGATGCTGACAT CATCGCTGTACTG	CCCTGCTGTACTGAGCCAC
5	rs24767776	ACGTTGGATGAGAATC ATCTCAGCCATGCC	ACGTTGGATGGTGGAT TCTTGGAGCCTTTC	AGGAGCCTTTCTTAGGGAG
5	rs23625894	ACGTTGGATGAGGCAG AGACATAGGCAGAG	ACGTTGGATGATCCTG GAGACCTGGGATCG	CCCCCGGGATCGAATCCCAC
5	rs22214750	ACGTTGGATGGACTGG AGAGAAAAGTCTTC	ACGTTGGATGCCTTATA TTTGGTCTCTGGC	TTCTATCACGGTATGAATTG
5	rs23651397	ACGTTGGATGTTCCTCT GCGTGATGGTTCC	ACGTTGGATGTAAAAC ATGGGCCTGTCCTG	GGAGAAGAATTCCATGAAGG
5	rs22106265	ACGTTGGATGTTGTAG GGACTGAGTTACCG	ACGTTGGATGATCGGA GAAGAGAGAAACGG	CCCGGGGAGAAATCTTCAACC
5	rs24546447	ACGTTGGATGTTGGCC CCCTACATCCTTC	ACGTTGGATGAAGGAG ACATGGGCATCAAG	AGAAAGGGTGATCGGGTAAGT
5	rs22919923	ACGTTGGATGAACCCC CCCTAGTTTGATTC	ACGTTGGATGTTAAGCC TGCCTCAGTGTTC	GGTATAGCCAAATGCTTTCTCA
5	rs8703991	ACGTTGGATGCCAGCT TCAGGTAAAAACAG	ACGTTGGATGAGCAAC ACTCAGTACTTCAC	GAGTTACCAATGTGTAAGAAGT
5	rs22413365	ACGTTGGATGACACTC TGTGTGTGTATGGG	ACGTTGGATGGCAATTT TGTAGATTTCTCTG	GTAGATTTCTCTGTTAATCTCAT
5	rs23989232	ACGTTGGATGGGAGTC GCTAACCAATTCTC	ACGTTGGATGATCTGGT TCACCTGGAGTTC	GGGTGAGGCAAGGTAAGTATTTC
5	rs8587007	ACGTTGGATGCTCCAT GGGAGAAAGGGTAA	ACGTTGGATGCTATCAT TCCTTACTATCCC	TTACGATCCCAAAATCTACCCATA
5	rs8907730	ACGTTGGATGCGAGAG GAACACATATGGAG	ACGTTGGATGAATTTG GTGTGGACGTTGGC	GGATGTTTGCTTTTTATAAACCAA
5	rs24346522	ACGTTGGATGAGAGA GACGCTGTACAAATG	ACGTTGGATGAGGAAA ATTTCTCCTGGGAC	CTCATGGGACATGGTATTAAAAAT
5	rs22255738	ACGTTGGATGCTGTTC CCTTTAGCAACGTC	ACGTTGGATGAAGGTG AGTAAAGGGTCAGG	GAATGGAACTTGTGAGGACTAGAG T
5	rs8893358	ACGTTGGATGTGGCGC TCCTGTAAGTATTG	ACGTTGGATGTGTCATA CTCCACACCTTTC	CTTTAATTCTGTTTTTTTTTCTTTT
5	rs8785162	ACGTTGGATGCCCGTG TGAGTAGAATTTTG	ACGTTGGATGTTTGGA AAGCTGCTCAACTC	AAGATGCTCAACTCAAATCAAGTA AA
5	rs23209310	ACGTTGGATGGGATAA	ACGTTGGATGAACTAG	GTTCCTCTTTCTGTTTTAGGGTATTC

Plex #	SNP_ID	Forward Primer (5' - 3')	Reverse Primer (5' - 3')	Probe (5' - 3')
		CCTCGATAGCCATC	AGGCTTTCACCGAG	C
5	rs8780512	ACGTTGGATGCCCAGC ATGACAAGTATTTC	ACGTTGGATGACCCTTT TTCATTCTTTTC	TCATGAAAAACTAAGTGTATAATTC AG
5	rs22165230	ACGTTGGATGCCGGCG CTCAGTAACCTG	ACGTTGGATGTGTCGTC GGAGTACTTGTAG	CCCCCCACCATGCCATCCACGCGGC TGG
5	rs24737225	ACGTTGGATGGCAGTT CAGTCATTCAGAGG	ACGTTGGATGGAGGTT TCAACCTAAACCAG	ACAAGAATTTCCTATTTAGAATCCC AAC
5	rs22195359	ACGTTGGATGTACAGA CATCCCCAGAGTTG	ACGTTGGATGTGTGGT GTCACTGTGTGATG	ATAGCTGACATGATTCCACATGTGT ATG
6	rs24619748	ACGTTGGATGAGACTT TAGTGCTTCCTCAG	ACGTTGGATGTCCAGG ATGCCTTCCTTGAC	CCTTGACAACCACCC
6	rs24732865	ACGTTGGATGATACGC CTTATCTGTGACGG	ACGTTGGATGTTACCTT CGAGGGTCATTCC	GGCATCTCCAGCCCC
6	rs24011145	ACGTTGGATGAGGTAA GTGGTTGTCTTGGG	ACGTTGGATGTCTCATC TGCGTGTGCCTTG	TGGCTCCTCTGTGTA
6	rs9024664	ACGTTGGATGGAATGC ACATCACACAGACC	ACGTTGGATGACTGAG AGCTCTGTGGGTTC	GGGTTCTCAGCGGGA
6	rs24038028	ACGTTGGATGTGTCAC CGAGAGTGTCATTG	ACGTTGGATGACTAGT AACCATGTTCTTG	ACAATGACATCGCTCA
6	rs22106266	ACGTTGGATGGGTCTG GGTCGGCCGGCG	ACGTTGGATGGCACGC CGATCCTGCTGC	TGTCCTGCTGCTCCGCA
6	rs24615406	ACGTTGGATGTTCATA TTCGGGCTCTCTGG	ACGTTGGATGAGCGTG GTGCTCTCCAAAGA	TCTCCAAAGACGCTTTA
6	rs24278945	ACGTTGGATGTTGGCT AGCTGCCATTTCTG	ACGTTGGATGTGTGTTG CACGATGATCACC	CGATGGACTTAGATGCT
6	rs8830091	ACGTTGGATGGCCTCC TTTTTTCCAGGATG	ACGTTGGATGTCTGTAC CCTCTGCACTTAG	GAGGCCTTGCCTGTTGGG
6	rs24278946	ACGTTGGATGAGGACG TTGGGTATTAGGAG	ACGTTGGATGTTTACAA TGTGCGGCACCTC	CCGAGGCACCTCCTGAACA
6	rs24380667	ACGTTGGATGCTCTAT ATGAATATTTTACAAG	ACGTTGGATGTCCATTT TTTAAGAGCCGAC	GAGCCGACTCTTAATAGAT
6	rs8950488	ACGTTGGATGCTGACA TCATCGCTGTACTG	ACGTTGGATGCAAGGT TATCGATGAGCTGG	TGGATGTGAAACCAGAGGG
6	rs8561539	ACGTTGGATGCTGAAG TTCTTGGAGTTGCC	ACGTTGGATGGGAATA TTGGCTTTGGCAAC	CCTGGATTAAATACTGACCC
6	rs8927359	ACGTTGGATGTATGGC AATTAAAATCACC	ACGTTGGATGTTCCATA GGGAGCAGATGGC	GGCTTCAGAGGCCTCAAGGG
6	rs24566251	ACGTTGGATGGAGCGT TAATTGCTTGCAGG	ACGTTGGATGGCAGTC TCTTGAGTTCTTAG	TCCTCTTGGGAATCCACATAT
6	rs9095803	ACGTTGGATGAGAAG GCCCAACAGCTCCAG	ACGTTGGATGTTCCCAT CAGGGTGAACGTG	GGCATGGCCCCACAGGGAAAA
6	rs24261266	ACGTTGGATGGCAACA ACCATAAAAGATG	ACGTTGGATGAGGAAT TGTGGGAAGTCCAG	GAGTCCAGGGAAAAAGGTAAT
6	rs23232135	ACGTTGGATGCCATAG AGTCATCTTGCAGG	ACGTTGGATGTACACCT CCTTGACGTTCAC	CCAGAATCACGAAGAAATAACC
6	rs8926268	ACGTTGGATGTTTCCTA AAAGGGACCTGCC	ACGTTGGATGCATTACT CTTGTTACTCTGC	CACGCTTGTTACTCTGCACTTAC
6	rs8898240	ACGTTGGATGGGTCTT	ACGTTGGATGCTTGCCT	TCGCTTGCTTTGAGAGGTCCACC

Plex #	SNP_ID	Forward Primer (5' - 3')	Reverse Primer (5' - 3')	Probe (5' - 3')
		CCAGTTAGACCCAT	GAGTTTGGTCTAC	
6	rs8861242	ACGTTGGATGTAGAAA GTCACCACCTCCAC	ACGTTGGATGTTAGGG ATGGAGAGAAACGC	TATAGAGAGAAACGCATAATGAA
6	rs24553222	ACGTTGGATGACAGTA AAGACACGAGTGCG	ACGTTGGATGATAAAG AGCCATTCCCCCTG	CCCTGATATTAAAATCTCCTGATT
6	rs24048817	ACGTTGGATGACAACC TTCTGGACTCCATC	ACGTTGGATGTCTACTC AGTGGTCTTGCTC	AAGGATCCTGGGGAGGCTGCTTAC
6	rs24029015	ACGTTGGATGCTGGCT AAAGCATGCCACTG	ACGTTGGATGCCACAC AGAATGTCAGGAAG	GGAGAGGAGAAGAAAAGGAGGAA G
6	rs22595730	ACGTTGGATGGGCAAG CCACAGTTATGAAG	ACGTTGGATGGGGTAG TCTGTAGGAAAATG	GGGAGTCTGTAGGAAAATGAATTG C
6	rs22165232	ACGTTGGATGGCCCGG CGCTCAGTAACCT	ACGTTGGATGTGTCGTC GGAGTACTTGTAG	ACAACCATGCCATCCACGCGGCTG GG
6	rs8523506	ACGTTGGATGCCCAAG TGATTTAGGAGAAA	ACGTTGGATGCACAAT ACTAACAAGTTAGA	ACAAAATGAGAAAACAAATAAACA AA
6	rs22378754	ACGTTGGATGACACCG TCTCAAATCTGACC	ACGTTGGATGGGGATT CCCTTACGAAAAG	GAAAAGAAATTTGAATGTAATTTA AATC
7	rs24528134	ACGTTGGATGTCGGCG TCCACGTAGGAGA	ACGTTGGATGACCAAG GTTGCAAAGGCTGG	GAGCACCCCCCTTGT
7	rs24048796	ACGTTGGATGCTGTGT GTATATGTGTGTGC	ACGTTGGATGAGCCAA CTGCAAAACTCACC	CCACCCCCACTCCCTT
7	rs22356416	ACGTTGGATGATGGAA GCACTCACAGCTGG	ACGTTGGATGTTCTGCT TCTAGGGAGCCCG	TGAGAGAGGACGTGT
7	rs22217604	ACGTTGGATGAGGAGC TCCTTTCACTTACG	ACGTTGGATGATTATCG TGAGGAGACGGAG	ACCCTGCCTTGTCTGC
7	rs24864610	ACGTTGGATGCTATCG CTGTGAGGTTATGC	ACGTTGGATGACGTTCT CACCTTTCACCAC	GGGTAGCCTCACTGTC
7	rs8680884	ACGTTGGATGGTGGGA CCTTTGTATGCATC	ACGTTGGATGAACAGG ACGTCTTGGTCCC	GGTCCCTCCGGAAGCCA
7	rs24758506	ACGTTGGATGAGGGAC TTTTGGACATATTC	ACGTTGGATGTCTTTTC CACTGTGGCAGTC	TTCTCCTTGAATTCCTCC
7	rs23546710	ACGTTGGATGGCAATG ATGGCACTGTTCTG	ACGTTGGATGTTGTGTT CCATTACAGAGCC	GGTCAGAGCCATCTCCAC
7	rs24535614	ACGTTGGATGTTCATG AGAGGCAGAGAGAG	ACGTTGGATGAGGATT GAGTCCTGCATTGG	GTGAGGAGCCTGTTTCTC
7	rs24578426	ACGTTGGATGAAAGGT CCCGGGCGCAGGTA	ACGTTGGATGTGTCTTT TCCCCACAGGGTC	CCTTGTCCTCCCGGGCCCA
7	rs24770735	ACGTTGGATGCAGCTA TACAGTCATGTAATG	ACGTTGGATGGACTGC TCAAGAAACTACCC	ACTACCCGTAAGTTGTGAT
7	rs22917971	ACGTTGGATGGGGTAA TTCTTGGCACACTC	ACGTTGGATGTGAACA GGTTTAAGCCCCAG	ATCCCCAGGGTTATATAGTC
7	rs22289709	ACGTTGGATGTTTCAG GTTGAGCCCGAATG	ACGTTGGATGCCAAGG ATGCACTATGGATG	GGATGCTATCAAAGTATACTG

VI. Q-PCR data – Shapiro Wilk normality test

Appendix 5. Normality test results for the Q-PCR datasets as described in Chapters 7 & 8

	MMP2	Biglycan	Trap	Cathepsin K	COL 1A1	COL 1A2	IL6	DIRC2	PDIA5	PARP-9	MMP1
D/N	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
I/C	0.00	0.2	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.00

KEY: D/N disease vs normal tissue, I/C = TNFα induced vs control data.

A value less than 0.05 means that the data is non-parametric and does not follow a Gaussian distribution

VII. Q-PCR data – primer efficiencies

Appendix 6. Efficiency data for the Q-PCR primers used in Chapters 7 & 8.

Gene	Threshold	Slope	R^2	Efficiency
MMP-2	4.57	-2.92	0.99	120%
Biglycan	0.52	-2.96	0.99	118%
TRAP	3.98	-3.60	0.99	90%
Cathepsin K	0.60	-3.10	0.97	110%
COL1a1	0.40	-3.05	0.97	113%
COl1a2	0.33	-3.24	0.95	104%
IL-6	1.51	-2.80	0.97	128%
DIRC2	1.47	-3.21	0.91	105%
PDIA5	1.6	-3.18	0.93	106%
ATIC	1.00	-3.20	0.99	105%
PARP-9	0.84	-3.54	0.98	92%
HIRIP	0.52	-3.44	0.99	95%
MMP-1	0.20	-3.78	0.90	84%
				Average efficiency: 105%