

Genetic Dissection of Anterior Cruciate Ligament Rupture in the Dog Model

By

Lauren Ashley Baker

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctorate of Philosophy
(Comparative Biomedical Sciences)

at the

UNIVERSITY OF WISCONSIN-MADISON

2019

Date of final oral examination: 10/08/2019

The dissertation is approved by the following members of the Final Oral Committee:

Peter Muir, Professor, Surgical Sciences, School of Veterinary Medicine

Susannah Sample, Assistant Professor, Surgical Sciences, School of Veterinary Medicine

Lauren Trepanier, Professor, Medical Sciences, School of Veterinary Medicine

Corinne Engelman, Associate Professor, Population Health Sciences

Guilherme Rosa, Professor, Animal Sciences

Christina Kendzierski, Professor, Biostatistics and Medical Informatics

Acknowledgments

I would like to express my gratitude to my advisor, Dr. Peter Muir, for his mentorship and unwavering support during the last 7 years. I am extremely grateful to be afforded the opportunity to grow personally and professionally in the Comparative Orthopedic Research Laboratory and the Comparative Genetics Laboratory at the University of Wisconsin School of Veterinary Medicine. I would also like to express my sincerest thanks to the members of my committee. To Dr. Susannah Sample for always being (literally) at my side and available to listen and provide your expertise and advice as I learned to navigate the world of veterinary research. Thank you to Dr. Corinne Engleman for making me feel so welcome in your laboratory, and for keeping me up to date on the latest and greatest in genetic epidemiology methods. Thank you to Dr. Guilherme Rosa for the hours of teaching that occurred during our many meetings. To Dr. Lauren Trepanier, for your excellent career advice and for being a valued role model. And to Dr. Christina Kendzierski, for all of your support throughout this process.

I would also like to thank all the members of the Comparative Orthopedic Research Laboratory and Comparative Genetics Laboratory. To Zhengling Hao and Jordan Gruehl, for your expertise in molecular techniques without which none of this work would have been possible. To Brett Nemke, for keeping our lab organized and for your friendship and support of my work. Thank you to the interns and residents that have contributed to this work, especially Nicola Volstad, Alexander Piazza, and Christopher Hoffman. And a huge thank you to my colleague and treasured friend, Emily Binversie. You inspire me, and I am truly honored to work with you every day. I would also like to acknowledge the hard work of the undergraduate students and veterinary students that have contributed to this project; specifically, thank you to Allison Stilin, Kore Chan, Nathan Bollig, and Rachel McNally.

I would also like to thank the administration and members in the Computation and Informatics in Biology and Medicine training program, especially Drs. Mark Craven and Louise Pape, for providing me with the opportunity to receive training and support in bioinformatics through the National Library of Medicine training grant T15 LM007359. It has been a pleasure to be a part of this program, and I hope to continue to contribute to this group in the future. For me, never has the learning curve been so steep, and I am so proud of the growth I was able to achieve with your support over the past 3 years.

Lastly, I would like to express my gratitude to my husband, Mike, who has lovingly supported my education through veterinary school and graduate school. Thank you for being my rock, for keeping me grounded, and for reminding me that there is more to life than books. To my daughter, Evelyn, I hope that I inspire you to achieve the goals you set for yourself, whatever they may be.

Abstract

Anterior cruciate ligament rupture (ACLR) is a common condition that occurs in both human beings and the domestic dog. ACLR is characterized by mid-substance fiber rupture, which ultimately progresses to complete ligament rupture and knee instability. In human beings, post-rupture patients are at 50% risk of early-onset osteoarthritis, and this risk is not modified significantly with surgical treatment. The presentation and progression of ACLR in the dog is similar to human ACLR. Medical treatments are not available for either species. The risk of ACLR is multifactorial and complex, and it is widely recognized that a combination of genetic predisposition and environmental factors influence disease risk. Understanding the genetic basis of this disease will provide the opportunity to develop disease-modifying therapy and a predictive genetic test. The dog offers a unique opportunity for genetic study of spontaneous ACLR. Homogeneity within dog breeds improves power to detect regions of the genome that are associated with ACLR through genome-wide association study (GWAS). This dissertation presents research into the genetic epidemiology of ACLR in the Labrador Retriever. We analyze data using three approaches to GWAS to identify genes and regulatory regions that are associated with rupture. Our results suggest roles for genes involved in inflammatory pathways, morphological development, and extracellular matrix proteins. We also discuss the interpretation of GWAS through LD-Score regression, and evaluate several models for genomic prediction of ACLR in the Labrador Retriever. This body of work provides evidence that ACLR in the dog is a highly polygenic complex trait disease. Future work will require larger sample sizes to make robust conclusions about the molecular pathways involved in disease risk. Ultimately this One Health project is expected to lead to advancements in medical knowledge of ACLR in both humans and dogs.

List of Figures	Page
4.1 The Labrador Retriever is a dog breed with a relatively low amount of inbreeding	31
4.2 Linear mixed model GWAS corrects for population structure and identifies 99 ACLR associated loci explaining a large proportion of phenotypic variance	33
4.3 Phenotype variance was explained to a large degree by the associated genomic loci.....	36
4.4 Genetic risk scoring using GWAS associated loci from linear mixed model analysis with GEMMA segregates ACLR disease risk in case and control Labrador Retriever dogs	39
4.1S Genetic risk scoring using GWAS associated loci from linear mixed model analysis with GCTA, GEMMA, and PUMA segregates ACLR disease risk in case and control Labrador Retriever dogs	65
5.1 Measurements of proximal tibial morphology taken from a lateral stifle radiograph.....	75
5.2 Manhattan plot of $\log_{10}$ Bayes factors (BF) of the multivariate phenotype	79
5.1S Scree plot of principal components eigenvectors calculated from Labrador Retriever genotypes	86
5.2S Distribution of tibial plateau angle (TPA) and relative tibial tuberosity width (rTTW) among cases and controls.....	87
6.1 Schematic of data analysis and modeling workflow	97
6.2 Results of 10-fold cross validation with models trained on feature sets from 5-15,000 SNPs.....	103
6.3 Results of holdout validation with models trained on feature sets from 5 to 15,000 SNPs	104
6.1S Percentage of explained variance by each principal component using the genetic relationship matrix of the combined Wisconsin and Hayward et al. (2016) datasets	113

6.2S	Plot of the first three principal components computed using the genomic relationship matrix of the combined Wisconsin and Hayward et al. (2016) datasets	113
8.1	Q-Q plot of GWAS results without correction for population structure	129
9.1	Manhattan plot of genome-wide association analysis of ACLR case and control dogs using BayesRC.....	144

List of Tables	Page
4.1 ACLR associated loci identified by GWAS in the Labrador Retriever; a dog breed with a high prevalence	34
4.2 Genetic risks coring in ACLR case and control Labrador Retriever dogs using GWAS associated SNPs from linear mixed model analysis .	38
4.3 Receiver operating characteristic (ROC) analysis of genetic risk scoring in ACLR case and control Labrador Retriever dogs using GWAS associated SNPs from linear mixed model analysis	38
4.4 Functional annotation clustering of genes in regions associated with anterior cruciate ligament rupture identified by GWAS in the Labrador Retriever.....	40
4.1S Anterior cruciate ligament rupture associated SNPs identified by GWAS in the Labrador Retriever; a dog breed with a high disease Prevalence	48
4.2S Statistical Power and odds ratio correction for anterior cruciate ligament rupture GWAS risk loci identified by GEMMA in the Labrador Retriever.....	62
4.3S Estimation of the expected number of loci to be discovered in a future GWAS of ACLR in the Labrador Retriever using INPower	64
5.1 Summary statistics for individual covariates and phenotypes in ACLR case and control groups.....	78
5.2 SNPs associated with the multivariate phenotype of anterior cruciate ligament rupture (ACLR), tibial plateau angle (TPA), and relative tibial tuberosity width (rTTW)	79
5.3 The posterior probability of effect (PPE) calculated for each SNP and corresponding phenotypes.....	80
6.1 Bayesian model performance for genomic prediction of ACLR in dogs ..	101
6.2 Highest performing models in 10-fold cross validation for prediction of ACLR in Labrador Retriever dogs.....	105
6.3 Highest performing models in holdout validation for prediction of ACLR in Labrador Retriever dogs.....	106
6.4 Highest performing ensemble models in 10-fold cross validation	107

6.5	Highest performing ensemble models in holdout validation	107
7.1	Number of dogs in each breed included in the multibreed reference panel used for imputation.....	116
7.2	Accuracy of imputation of Labrador Retriever SNPs based on comparison to whole genome sequencing data.....	119
8.1	Results of LD Score regression using difference breed populations as references.....	130
9.1	Affected case and matched unaffected control dogs used for RNA sequencing	141
9.2	Candidate genes for ACLR derived from human and dog studies	142
9.3	The number of SNPs assigned to biological priors defined by differential gene expression analysis and candidate genes reported in the literature.....	145
9.4	The top 50 SNP effects from Bayesian mixture model (BayesRC) association analysis	145
9.1S	Differentially expressed genes identified through RNA sequencing of anterior cruciate ligament tissue from ACLR case and control dogs.....	152
9.2S	Differentially expressed genes identified through RNA sequencing of synovium tissue from ACLR case and control dogs.....	157
9.3S	Functional annotation using DAVID identifies a cluster of 141 genes for proteins with a transmembrane helical domain	168

Table of Contents	Page
Acknowledgments	i
Abstract.....	iii
List of Figures	iv
List of Tables	vi
Table of Contents	viii
 Chapter 1: Anterior Cruciate Ligament Rupture: A One Health Investigation ...	 1
 Chapter 2: The Epidemiology of Anterior Cruciate Ligament Rupture in the Dog.	 6
Introduction.....	7
Age	7
Sex.....	8
Obesity	9
Conformation.....	10
Bilateral ACLR.....	11
Genetics	11
 Chapter 3: The Genetics of Anterior Cruciate Ligament Rupture in the Dog	 13
Introduction.....	14
Heritability of ACLR in the dog.....	14
Complex trait genetics	15
Candidate gene analyses	15
Genome-wide association	17
GWAS for genomic prediction	20
 Chapter 4: Genome-wide Association Analysis in Dogs Implicates 99 Loci as Risk Variants for Anterior Cruciate Ligament Rupture	 22
Abstract	23
Introduction.....	24
Materials and Methods.....	26
Ethics statement	26
Canine samples and phenotyping	26
Genome-wide association	27
Genome-wide significance	28
Defining associated loci in the genome	28
Genetic risk score computation	29

Pathway analysis	29
Heritability estimation	30
Results	31
GWAS population of Labrador Retrievers	30
GWAS Identifies 99 regions association with anterior cruciate ligament rupture	31
Risk loci clearly distinguish ACLR cases from controls	38
GWAS pathways are enriched for carbohydrate binding Proteins	41
ACLR in the Labrador Retriever has moderate heritability	42
Discussion	43
Supplementary information	49
 Chapter 5: Multivariate Genome-wide Association Analysis Identifies Novel and Relevant Variants Associated with Anterior Cruciate Ligament Rupture Risk in the Dog Model	 68
Abstract	69
Introduction	71
Methods	74
Recruitment	74
Genotyping	75
Radiographic knee morphology	75
Statistical analysis	76
Results	78
Discussion	81
Conclusions	85
Declarations	86
Supplementary Information	87
 Chapter 6: Bayesian and Machine-Learning Models for Genomic Prediction in a Dog Model of ACL Rupture	 89
Abstract	90
Introduction	91
Materials and Methods	93
Data collection and phenotyping	93
SNP genotyping quality control	94
Bayesian analyses	94
Machine Learning Analyses	96
Results	101
Bayesian analyses	102
Machine learning analyses	102
Ensemble learning	108
Covariate analysis	108
Discussion	109
Supplementary Information	114

Chapter 7: Imputation of Labrador Retriever Genotypes Using a Multi-breed Reference Panel	115
Introduction.....	116
Methods.....	117
Results.....	118
Discussion	121
 Chapter 8: Effect of Linkage Disequilibrium Structure on Linkage Disequilibrium Score Regression for Interpretation of Canine GWAS.....	 123
Abstract	124
Introduction.....	125
Methods.....	128
Results.....	129
Genome-wide association	129
Linkage disequilibrium score regression.....	130
Discussion	131
 Chapter 9: Exploiting Biological Priors for Enhanced GWAS in a Dog Model of ACLR	 135
Abstract	136
Introduction.....	138
Materials and Methods.....	140
Data collection and phenotyping	140
RNA sequencing and differential gene expression analysis ...	141
Association analysis and assignment of biological priors.....	142
Results.....	144
Labrador Retriever dataset.....	144
Differential gene expression	144
Association analysis	145
Discussion	147
Conclusion.....	152
Supplementary Information.....	153
 Chapter 10: Summary and Future Directions.....	 173
 References	 179

Chapter 1

Anterior Cruciate Ligament Rupture: A One Health Investigation

The similarity of disease processes in humans and animals has been recognized for centuries. Comparative medicine, as a formal discipline, has existed since at least the 18th century, though the use of animal models in medicine and surgery dates back to ancient times [Zinnstag 2011]. The concept of *one health* emerged over the last half-century to highlight the indisputable interconnectedness between humans, companion animals, livestock species, wildlife, and the ecosystems in which we co-exist. The One Health Initiative promotes collaboration between public health, human health, veterinary health and other scientific health and environmental professions to promote advances in biomedical research, public health efficacy, and improving medical education and clinical care into the 21st century. The research described in this dissertation encapsulates the value of One Health, such that we can use insights gained from the study of companion animal populations to inform human medical research.

A wide range of common human diseases are also prevalent in animals, particularly long-lived companion animal species. The dog, in particular, has emerged as an excellent animal model for the study of common spontaneous diseases. For example, dogs are known to spontaneously develop auto-immune diseases [Gershwin 2007], endocrinopathies [Fall 2007; De Bruin 2009; Chase 2006], behavioral disorders [Overall 2000], cancers [Rowell 2011], obesity [Bergman 2007], and osteoarthritis [Pascuel-Garrido 2017], among others, all of which bear a striking resemblance to analogous diseases in humans. The dog is also well-suited to genetic research, because not only is the domestic dog the most phenotypically diverse land animal species, but pet dogs also receive long-term preventative and acute health care that approaches the level of care seen in human beings [Rowell, 2011]. The portfolio of known diseases in the dog is only second in size to humans [Rowell, 2011]. Additionally, selection pressure instituted through breed creation has inadvertently selected for high-prevalence of breed-specific diseases [Shearin 2010], which provides a unique opportunity to identify genes that associate with disease in otherwise relatively homogeneous purebred populations.

Anterior cruciate ligament rupture (ACLR) is another example of a common condition that is shared by human beings and their canine counterparts. The anterior cruciate ligament (ACL) is one of the two intraarticular cruciate ligaments in the knee/stifle. Together with the posterior cruciate ligament, the ACL plays an important role in maintaining knee stability, primarily preventing anterior translation of the tibia bone and counteracting rotational forces in either direction. The analogous structure in the dog is properly named the cranial cruciate ligament (CCL). For the sake of clarity, human terminology will be used for both species throughout this dissertation.

ACLR is characterized by mid-substance fiber rupture that may occur as a partial or complete rupture of the ligament that is most commonly associated with knee instability [Tjoumakaris 2011]. In human beings, ACLR most often occurs via a noncontact mechanism, during sudden decelerations, landing, or pivoting movements on a flexed knee [Smith 2012b]. Post-rupture, patients are at 50% risk of early-onset knee osteoarthritis [Lohmander 2007], high risk of contralateral rupture [Paterno 2012], and if surgical repair is pursued, graft rupture [Wright 2011]. The management of osteoarthritis is a lifelong endeavor. Medical treatments provide some relief for symptoms but are not disease-modifying [Gelber 2015]. Overall, the management of ACLR extends far beyond the cost of initial treatment, and has an annual impact of 7.6-17.7 billion dollars per year in the United States alone [Mather 2013].

The risk of non-contact ACLR is multifactorial and complex and it is widely recognized that a combination of genetic predisposition and environmental factors influence disease risk. The genetic basis of ACLR is not well-understood. A family predisposition has been identified [Flynn 2005, Meyer 2014], but genetic heritability has not been calculated. The majority of genetic research of ACLR is comprised of candidate gene studies, which have provided conflicting and limited evidence for association with collagen genes, proteoglycan genes, matrix-metalloproteinase genes, and elastin and fibrillin [Kaynak 2017]. Two separate studies have evaluated a mutation (G1023T; rs1800012) in intron 1 of *Collagen Type 1 alpha 1 (COL1A1)*,

the binding site for the transcription factor Sp1, in relation to ACLR [Khoschnau 2008; Posthumus 2009a]. Although the TT genotype was rare, both studies concluded that the homozygous TT genotype seemed to have a protective effect on development of ACLR. It is proposed that the increased binding of the Sp1 transcription factor increases expression of alpha 1 chain, but it is unknown how this protects against ACLR. In another study, female participants with ACLR were 2.4 times more likely to have the AA homozygous genotype for the SNP rs970547 in exon 65 of *COL12A1* (*collagen type 12 alpha 1*) [Posthumus 2010]. This finding was confirmed in a separate cohort [O'Connell 2015]. This study also reported an interaction between COL5A1 rs12722 and COL12A1 rs970547 such that individuals with the T+A- pseudo-haplotype were at greater risk of experiencing ACLR [O'Connell 2015]. Another candidate gene study identified an association between matrix metalloproteinase genes and ACLR, particularly *MMP12* rs2276109, where the AA genotype is significantly more prevalent among individuals with confirmed non-contact ACLR [Posthumus 2012].

Angiogenesis signaling may also play a role in ACLR risk, as the *VEGFA* rs699947 CC genotype is also significantly over-represented among individuals with non-contact ACLR [Rahim 2014]. Proteoglycans have also been identified as a potential factor influencing the risk of ACLR in human beings. Proteoglycans play important roles in fibrillogenesis and collagen extracellular matrix homeostasis. A candidate gene study identified loci in *ACAN* (aggrecan) and *DCN* (decorin) that were associated with ACL injury susceptibility [Mannion 2014]. A separate study evaluated gene expression in biopsies taken from ruptured ACL and noted that expression of *ACAN* was significantly upregulated in female compared to male patients, suggesting that aggrecan may be playing a role in the observed sex differences regarding susceptibility to ACLR [Johnson 2015], wherein females are at greater risk for ACLR than males.

A recent genome-wide association study of ACLR was not able to identify genome-wide significant loci, though the analysis only included 598 cases [Kim 2017]. Additionally, a whole

exome sequencing approach of 234 ACLR cases and 232 controls was unable to identify significant differences [Gibbon 2018]. Many of these studies note that a major limiting factor for genetic analysis of ACLR is limited sample size.

The dog offers a unique opportunity for studying the genetic basis of spontaneous ACLR. Canine ACLR has a similar clinical presentation and progression to ACLR in human beings. The vast majority of ACLR cases in dogs are noncontact injuries, and >50% of dogs will rupture the contralateral limb within one year of diagnosis in the primary limb [Muir 2011a]. The condition shows distinct breed predisposition, with Newfoundlands, Rottweilers, and Labrador Retrievers leading the pack [Witsberger 2008]. Individual level data and tissues are far more accessible in dogs than human beings, and the homogeneity of purebred dogs suggests that good statistical power can be achieved with relatively modest sample sizes [Karlsson 2008].

The following dissertation presents research into the genetic epidemiology of ACLR in the Labrador Retriever dog. The Labrador Retriever is at high risk of ACLR (5.79% breed prevalence) [Witsberger 2008], and is also the most common breed in the United States, Canada, and the UK. It is my expectation that insights gained from this research will not only provide valuable information to veterinarians who wish to understand how this condition affects their patients, but will also be hypothesis-generating for human medical research, laying the groundwork for further work that will identify the molecular basis of ACLR on a more global level. This One Health project aims to 1) identify targetable pathways for ACLR medical treatment, and 2) provide a foundation for development of prediction methods for the next generation of patients, which would allow the opportunity for medical intervention before patients make their way off of the field and into the orthopedic surgeon's office.

Chapter 2

The Epidemiology of Anterior Cruciate Ligament Rupture in the Dog

A version of this chapter has previously been published:

Baker LA, Muir P. Epidemiology of Cranial Cruciate Ligament Rupture. In: Muir, P. Advances in the Canine Cranial Cruciate Ligament. 2nd Edition. New York, NY. John Wiley and Sons; 2018. p. 109-114.

Introduction

ACLR is one of the most common causes of pelvic limb lameness in dogs [Witsberger 2008]. The condition is a chronic, progressive disease that affects the entire joint, not just the ACL, and results in the eventual rupture of the ligament and end-stage osteoarthritis [Bleedorn 2011; Comerford 2011; Muir 2011a]. A great deal of time and effort has been spent characterizing the epidemiological aspects of canine ACLR. Through this work, we have learned that ACLR is a very complex disease process with multiple contributing genetic and environmental risk factors. Furthermore, it is clear that there is no single combination of risk factors that will invariably lead to ACL degeneration and eventual rupture. Every dog should be considered an individual with a certain amount of inborn genetic risk on which environmental variables act to influence development and progression of disease. The following is a summary of research of risk factors known to contribute to ACL disease. Thorough understanding of these risk factors can aid in identification of dogs with increased susceptibility to the condition and guide patient management.

Age

ACLR is a chronic degenerative process, and so it follows that it is a condition of primarily middle-aged dogs [Whitehair 1993; Witsberger 2008; Reif 2003]. A study examining medical records data of over 1 million dogs reported that dogs 1-4 years of age were significantly less likely than dogs in other age groups to experience ACLR [Witsberger 2008]. While peak age of onset has been reported at 7-10 years [Whitehair 1993], a more recent study of 166 Labrador Retrievers reported peak age of onset of approximately 4 years. Notably, only 6% of these dogs had an age of onset over 8 years, and there were no dogs over 10 years of age. This suggests that older dogs are also at decreased risk for ACLR [Reif 2003]. Notably, there is an interaction with age and other risk factors. For example, large breed dogs tend to

present at younger ages than small breeds [Duval 1999]. Dogs that experience bilateral ACLR tend to be younger than dogs with unilateral ruptures [Cabrera 2008; Grierson 2011].

Sex

It is unclear whether sex differences have an effect on the incidence of ACLR in dogs. Multiple studies have reported an increased prevalence of ACLR in spayed females compared to other sex groups [Whitehair 1993; Lampman 2003; Powers 2005; Adams 2011]. Others have reported no significant differences [Witsberger 2008] or an increased risk in males [Grierson 2011]. The majority of research agrees that neutered dogs of either sex are at increased risk of developing ACLR compared to intact dogs [Whitehair 1993; Duval 1999; Lampman 2003; Slauterbeck 2004; Duerr 2007; Witsberger 2008; Torres de la Riva 2013; Belanger 2017]. A study of 360 Golden Retrievers [Torres de la Riva 2013] reported that early neutering, defined as castration or ovariectomy before age 1 year, significantly increased the likelihood of ACLR. In fact, there was no occurrence of ACLR in studied dogs that were intact (n=122). Occurrence in early neutered males and ovariectomized females was reported at 5.1 and 7.7 percent, respectively. A similar pattern has been reported in the German Shepherd Dog [Hart 2016] and the Labrador Retriever [Hart 2014; Ekenstedt 2017]. The effect of early neutering persisted after adjusting for differences in body condition score [Torres de la Riva 2013]. This is an important observation, as weight gain is often cited as an explanation for increased incidence of orthopedic disease among neutered dogs [Whitehair 1993; Duval 1999; Buote 2009]. If not through weight gain, how does early neutering affect the ACL? It has been suggested that absence of gonadal hormones leads to atypical growth plate closure and thus altered conformation [Torres de la Riva 2013; Hart 2016]. Indeed, early neutering has been reported to be a significant risk factor for development of excessive tibial plateau angles in large breed dogs [Duerr 2007]. Additionally, absence of gonadal hormones may affect the size, shape, or material properties of the ACL itself [Slauterbeck 2004]. A study investigating the

effect of gonadectomy on ACL collagen homeostasis in rabbits found that collagen concentrations were lower and fiber diameters were greater in the absence of gonadal hormones which may predispose them to rupture [Light 2012]. It has also been suggested that supraphysiologic levels of luteinizing hormone (LH) resulting from loss of negative feedback on the hypothalamic-pituitary-gonadal axis may be mediating effects on cruciate ligament, though this hypothesis has not yet been tested [Zwida 2016]. Research confirming a direct role for gonadal hormones on the canine ACL matrix architecture is not currently available.

Obesity

An association between higher body weight and increased risk of ACLR has been reported [Whitehair 1993; Duval et al. 1999]. While this is an important finding, evaluating weight alone does not distinguish between large size dogs and dogs that are truly overweight. Other work has attempted to evaluate body condition more objectively. A study of 755 dogs reported that those with overweight or obese body condition scores were twice as likely to be diagnosed with ACLR compared to dogs of normal weight [Lampman 2003]. A similar study evaluated body weight as a percentage of recommended breed weight and found that dogs in the obese category were nearly 4 times more likely to sustain ACLR compared to dogs of normal weight [Adams 2011]. Mechanical as well as metabolic factors explain the increased risk associated with obesity. From a mechanical standpoint, obese animals experience increased loading in their limbs, which in turn may over-stress the ACL and predispose it to rupture [Whitehair 1993; Adams 2011]. Additionally, there has been interest in the endocrine function of adipose tissue and its role in ACLR pathophysiology. Pro-inflammatory adipokines released by adipose tissue may contribute to the degenerative process underlying ACL disease [Pallu 2010; Adams 2011; Comerford 2011; Koskinen 2011].

Conformation

The anatomy of the stifle joint has been evaluated extensively for risk factors that may be associated with the development of ACLR. Anatomic features investigated include the dimensions of the distal femur [Comerford 2006; Lewis 2008], overall limb alignment [Dismukes 2008; Mostafa 2009] and proximal tibial conformation [Selmi 2001; Macias 2002; Reif 2003; Dennler 2006; Osmond 2006; Schwandt 2006; Guerrero 2007; Cabrera 2008; Inauen 2009; Ragetly 2011; Fuller 2014b; Haynes 2015; Janovec 2017]. Anatomic factors associated with a predisposition to ACLR include a narrow femoral intercondylar notch, excessive or pathologic (ie, outside the 95% confidence intervals for the population) tibial plateau angle, a relatively small proximal tibial width, and distal femoral torsion. While a great number of studies have focused on the effects of tibial plateau angle, the research is conflicting. Theories and *ex vivo* research suggest that increasing tibial plateau angle increases ACL strain [Warzee 2001; Reif 2003; Kowaleski 2005; Sharar 2006; Duerr 2007; Kim 2008; Kipfer 2008; Haynes 2015]. One study reported that a multiple logistic regression model using the combination of tibial plateau angle and femoral anteversion angle was able to discriminate between case and control limbs with an area under the ROC curve of 96% [Ragetly 2011] suggesting that a multivariate approach may provide further insight into the effect of tibial morphology as a risk factor. A follow-up publication by the same group with an updated model in a new cohort reported sensitivity and specificity of 87% and 79%, respectively [Griffon 2017]. Multiple other studies have reported no significant effect of proximal tibial morphology on the risk of ACLR or clinical outcome after surgical treatment [Wilke 2002; Reif 2003; Conzemius 2005; Guerrero 2007; Havig 2007; Guastella 2008; Fuller 2014b]. The wide spectrum of results supporting or negating the effect of conformational variables on the risk of ACLR provides further support for overall complexity of the condition. While bony conformation may contribute to ACLR in some way, there remains no definitive evidence that it is a primary causal factor.

Bilateral ACLR

Perhaps one of the greatest risk factors for developing ACLR is having already been diagnosed with the condition. The incidence of bilateral ruptures diagnosed on initial clinical presentation is in the range of 11-17% [de Bruin 2006; Cabrera 2008; Buote 2009]. The incidence of contralateral rupture after initial diagnosis is between 22 and 54% [de Bruin 2006; Cabrera 2008; Buote 2009; Grierson 2011; Fuller 2014b]. Dogs diagnosed with unilateral ACLR often have signs of moderate to severe osteoarthritis in the stable contralateral stifle [de Bruin et al. 2007; Grierson 2011; Muir 2011a; Fuller 2014a; Chuang 2014; Sample 2017]. The median time to contralateral ACLR rupture has been reported in the range of 405 to 1688 days [Grierson 2011; Muir 2011a; Fuller 2014b]. Stifle synovitis is likely an early event that contributes to the progression of ACLR [Erne 2009; Bleedorn 2011; Muir 2011a; Little 2014; Döring 2018]. Radiographic assessment of osteoarthritis present in stable stifle joints is correlated with arthroscopic assessment of synovitis [Bleedorn 2011; Sample 2017]. Therefore, dogs at greater risk for bilateral rupture can be identified through radiographic assessment. The presence of radiographic synovial effusion and osteophytosis in the stable contralateral stifle at the time of diagnosis is a significant risk factor for development of contralateral ACLR [Chuang 2014; Fuller 2014a], with more severe radiographic change associated with decreased time to contralateral rupture [Chuang 2014]. These results support the practice of obtaining bilateral stifle radiographs when evaluating a dog for unilateral ACLR. Evaluation of both joints for osteoarthritis is valuable for patient management as well as client education and represents the gold standard of care in small animal practice.

Genetics

The aforementioned risk factors primarily affect ACL disease progression. The only risk factor that seems to affect the initiation of ACLR in dogs is genetic influence, i.e. breed. It has long been recognized that certain breeds, such as the Newfoundland, Rottweiler, and Labrador

Retriever, are predisposed to ACLR while others, such as the Greyhound, are almost never diagnosed with the condition [Whitehair 1993; Witsberger 2008]. Much recent work has been done to characterize the genetic basis of ACLR in dogs, including the work outlined in this dissertation. The following chapter summarizes the work outside of our group that had been done to dissect the genetic basis of ACLR.

Chapter 3

The Genetics of Anterior Cruciate Ligament Rupture in the Dog

A version of this chapter has previously been published:

Baker LA, Muir P. Genetics of Cruciate Ligament Rupture. In: Muir, P. Advances in the Canine Cranial Cruciate Ligament. 2nd Edition. New York, NY. John Wiley and Sons; 2018. p. 57-64.

Introduction

In dogs, the most profound risk factor for ACLR disease initiation is breed. Dogs of high-risk breeds tend to present with ACLR at an earlier age and are more likely to present with bilateral rupture [Harasen 2008; Guthrie 2012]. The effect of breed and age on ACLR presentation suggest that genetic factors play a role in ACLR pathogenesis. It is likely that genetic risk factors combine with environmental modifiers such as body condition score or neuter status to affect expression of the ACLR trait. Identification of valid genetic risk factors influencing ACLR may lead to discovery of targets for medical intervention and aid development of a predictive test for the condition. A predictive test of ACLR may be used to screen young dogs that may be predisposed to the condition. This information can then be used to develop selective breeding strategies or counsel clients on environmental modification to minimize the impact of the clinical course of the disease. In order to meet these goals, much recent work has been done to discover the genetic basis of ACLR.

Heritability of ACLR in the dog

Statistically, heritability is defined as the proportion of phenotypic variance that may be explained by genetic variance. It may be loosely defined as the extent to which individual genetic differences contribute to changes in observable phenotype. Heritability is expressed as a number between 0.0 and 1.0. For example, a heritability of 0.5 says that 50% of disease risk is genetic, and the other 50% must come from environmental influence. A disease that is 100% genetic would have a heritability of 1.0.

Prior to our calculation of heritability in the Labrador Retriever, which will be described in the next chapter, heritability of ACLR had been assessed in two high-risk breeds, the Boxer and the Newfoundland. Newfoundlands have the highest prevalence of ACLR compared to any other breed. A study of 411 Newfoundlands (92 cases, 319 controls) estimated narrow sense heritability derived from an 11-generation pedigree at 0.27 [Wilke 2006]. A similar number was

reported in a prospective study of 414 Boxer litters, which placed overall heritability of ACLR at 0.28 [Nielen 2001]. Both of these studies used restricted maximum likelihood (REML) analysis for calculating heritability estimates.

Taken together, this work suggests that canine ACLR is a moderately heritable complex disease. Heritability estimates of this level indicate that one can expect a reasonable reduction in overall prevalence of the condition through genetic screening applied to a selective breeding program [Nielen 2001]. While it has previously been suggested that ACLR has an autosomal recessive mode of inheritance [Wilke 2006], more recent research, including our own, suggests that a polygenic mode of inheritance is more likely [Baird 2014 ab].

Complex trait genetics

Complex traits or diseases are simply defined as phenotypic traits that are determined by both genetic and non-genetic (environmental) factors. The genetic factors that influence complex traits are typically composed of multiple small-effect mutations that contribute in combination to phenotypic variation [Robinson 2014]. These mutations are expected to be large in number, and so their individual contribution to variance on a population level is quite small [Robinson 2014]. The relatively small contribution by individual mutations makes all but the largest effect mutations difficult to detect without very large sample sizes. Complete discovery of genetic variants contributing to ACLR and other complex traits will likely require a combination of complementary analyses including candidate gene analyses, large-scale genome-wide association studies (GWAS), and next generation sequencing.

Candidate gene analyses

For many, the candidate gene approach represents the first logical step for genetic association testing. Prior knowledge of gene function is used to choose candidate genes to test for association based on their potential role in disease etiology. Candidate gene studies have

been successful in identifying large-effect mutations, and mutations associated with Mendelian (monogenic) traits. However, this approach has faced criticism as current knowledge of gene functions is limited, and thus disease associations may be missed when they are not within “obvious” genes [Tabor 2002; Baird 2014a]. Additionally, complex diseases like ACLR are likely composed of many small-effect mutations. Therefore, it is unlikely that associations identified in a candidate gene study would explain a large amount of the genetic variance for the trait. With that being said, associations identified in these studies are still valuable as part of the explanation and should be considered complementary to genome-wide approaches.

In a study of a population of Newfoundland dogs, several genes were selected for analysis based on their known association with joint hypermobility in cattle or primary arthritis formation in humans. This included *Cartilage Oligomeric Matrix Protein (COMP)*, *Matrilin-3 (MATN)*, *Collagen Type 9 alpha 1, 2 and 3 (COL9A1, COL9A2, COL9A3)*, *Fibrillin-1 (FBN1)*, and *Interleukin Receptor 4 (IL4R)*. Single nucleotide polymorphisms (SNPs) were identified in *COMP*, *COL9A1*, *COL9A2*, and *FBN1*. Based on Chi-square analyses there were no significant association between the SNPs and ACLR case-control status, although some suggestion of an association was found between *COL9A1* (located on canine chromosome 12) and ACLR affected status ($P = 0.10$) [Wilke 2005].

Another study evaluated microsatellite markers closely located to the genes *Collagen Type 9 alpha 1, 2 and 3 (COL9A1, COL9A2, and COL9A3)* in a population of Boxers with a high incidence of ACLR [Temwichitr 2007]. Data from this study suggested that the candidate genes were not related to ACLR in the population studied.

More recently, a candidate gene approach was used to evaluate several potential ACLR risk genes in high-risk breeds including the Newfoundland, Labrador Retriever, Rottweiler, and Staffordshire Bull Terrier [Baird et al 2014b]. Genes tested for association included collagen genes: *COL1A1*, *COL1A2*, *COL3A1*, *COL5A1*, *COL5A2*, *COL5A3*, *COL6A1*, *COL6A3*, *COL11A1*, *COL11A2*, *COL14A1*, and *COL24A1*; fibril/elastic fiber formation genes: *FMOD*

(fibromodulin), *DCN* (decorin), *ELN* (elastin), *OPTC* (opticin), *LTBP2* (latent transforming growth factor beta binding protein 2), and *BGN* (biglycan); extracellular matrix genes: *FBN1* (fibrillin 1), *COMP* (cartilage oligomeric matrix protein), and *ACAN* (aggrecan); collagen-formation genes: *SERPINH1* (serpin peptidase inhibitor clade H (heat shock protein 47) member 1 (collagen binding protein 1), *PLOD1* (procollagen-lysine, 2-oxoglutarate 5-dehydrogenase), and *LOX* (lysyl oxidase); collagen cleavage genes: *MMP1* (matrix metalloproteinase 1) and *CTSK* (cathepsin K); and ligament/tendon/limb development gene *SIX1* (SIX homeobox 1).

When breeds were analyzed separately, two significant SNPs were identified: one in the Labrador Retriever (corrected $P < 0.001$) and one in the Rottweiler (corrected $P = 0.02$). The SNP identified in Labrador Retrievers was located within *COL24A1*. The SNP identified in Rottweilers was initially considered located within the elastin gene but was re-mapped to an intergenic region when the data was updated to the newest build of the canine genome. When all breeds were considered together, three significant SNPs were identified. Two SNPs were located in *COL5A1*, and the third was located within *COL1A1*. These results suggest that collagen genes, particularly *COL24A1*, *COL5A1*, and *COL1A1*, play some role in ACLR pathogenesis. It is also important to note that the associations on *COL5A1* and *COL1A1* were shared across all four breeds. This provides evidence that these mutations occurred in the canine genome before breed formation [Karlsson 2008]. Discovery of genetic mutations that are shared across multiple breeds has the most potential for clinical impact.

Genome-Wide Association

Genome Wide Association Studies (GWAS) use genetic markers across the entire set of DNA for an individual (genome) to first identify a chromosomal region that is associated with a disorder, and then identify all genes located in that region. The genes are then organized by function and selected for further investigation according to possible involvement in development of the disease or trait being investigated.

Before the advent of commercial single nucleotide polymorphism (SNP) arrays, GWAS could be performed using microsatellites (MSATs). A microsatellite is a variable repeating segment of DNA, usually found in a non-coding segment. Microsatellites can be highly polymorphic and very informative. However, recent advancements in DNA sequencing technology, computational hardware, and bioinformatics have made SNP arrays the tool of choice for GWAS. As whole genome sequencing (WGS) becomes increasingly financially accessible, reference panels based on WGS will increasingly be used for SNP panel data imputation to higher densities, further enabling discovery of disease-associated variants through GWAS [Wu 2017; Visscher 2017]. Imputation of our SNP panel to a higher density using a multi-breed reference population (not based on WGS) is described in Chapter 7.

Using microsatellites, a GWAS for ACLR was performed using 90 Newfoundland dogs that were selected for the GWAS based on their degree of inter-relatedness and the statistical likelihood that they segregated into homozygous unaffected and homozygous affected animals [Macrossan 2005]. Age and other potential contributors to ACLR risk were not considered in this analysis. A total of 495 MSATs were used to compare genotypes and allele frequencies between ACLR case and control dogs. Four markers (located on 4 chromosomes) were significant after false discovery rate was controlled at the 0.05 level using the Storey and Tibshirani method [Storey 2003; Wilke 2009]. The MSATs were CPH19 located on chromosome 3, FH3702 on chromosome 5, REN147D07 located on chromosome 13, and FH3750 located on chromosome 24. Initial validation of the 4 markers confirmed significance of three; canine chromosomes 3, 5, and 13. Positional candidate genes located on chromosome 3 were sequenced for mutation identification; Versican core protein precursor (*VCAN*) and aggrecan core protein precursor (*cartilage-specific protein core protein, CSPCP*). Proteoglycans are major components of extracellular matrix of cartilage [Schwartz 2002]. A significant association was identified with a SNP found in *CSPCP*, but not *VCAN*, and the trait, suggesting a role for aggrecan in ACLR pathogenesis.

In follow-up to this work, an updated GWAS of ACLR in the Newfoundland was performed using SNP genotyping [Baird 2014b]. Initially, 96 dogs (48 cases, 48 controls) were genotyped using a SNP array with >170,000 SNP markers evenly spaced across the genome. Clinical diagnoses were confirmed by a veterinary surgeon. Control dogs were defined as dogs over the age of 5 years with no history of stifle injury, stifle instability, or pelvic limb lameness. After quality control, the samples were analyzed using the Cochran-Mantel-Haenszel (CMH) test as well as Efficient Mixed-Model Association Expedited (EMMAX) analysis. Both the CMH test and EMMAX were used to correct for relatedness that may exist in the population. The 65 most significant SNPs from this analysis were re-genotyped in a larger group of 271 Newfoundlands (96 from the preliminary GWAS and 175 new) using Sequenom genotyping, and case-control association was re-analyzed. Three main associations were identified on chromosomes 1, 3, and 33. The association on chromosome 1 included SNPs within the *RNF152* gene, the association on chromosome 3 included SNPs within the *SORCS2* gene, and the association on chromosome 33 included SNPs with *SEMA5B*, *DIRC2*, and *ZDHHC23*. *SEMA5B*, *SORCS2*, and *ZDHHC23* all have various roles in the nervous system. There were also other nervous system genes identified in regions that did not maintain statistical significance after correction for multiple testing. This provides evidence for the potential role of neurological pathways in ACLR disease risk. The authors cited the importance of mechanoreceptors for appropriate proprioception, as reduced proprioception may prevent appropriate response to mechanical loading, placing the ACL at increased risk of matrix damage and fiber rupture [Baird 2014b]. These regions did not overlap with the regions identified in the earlier Newfoundland ACLR GWAS [Wilke 2009].

ACLR in dogs was also evaluated in two separate GWAS using the same group of multiple breeds by a research group at Cornell University. In the first study, [Hayward 2016], no genome-wide significant loci were identified. The second analysis [Huang 2017] used a novel iterative mixed model to improve statistical power to detect SNP effects (Farm CPU), and was

able to identify significant loci on canine chromosomes 7, 8, and 9. The locus on chromosome 9 was within a noncoding region of the genome. The closest gene to the locus on chromosome 7 was *DYM* (dimeclin) which encodes a golgi protein important for endochondral bone development. The closest gene to the locus on chromosome 8 was *CLMN* (calmin) which is a protein expressed in the brain during development.

GWAS for Genomic Prediction

While GWAS has most often been used for identification of candidate genes and/or pathways that may be contributing to disease pathogenesis, this approach may also be used to develop genomic prediction algorithms. Here research is less concerned with the biological effect of a mutation tagged by a SNP, and instead focuses on the statistical effect of a particular genotype and its ability to predict a disease outcome, often in combination with other SNP genotypes. A recent study performed GWA separately in a population of 46 Newfoundlands (22 cases and 24 controls) and 333 Labrador retrievers (190 cases and 143 controls) using a mixed linear model approach. GWAS-identified SNPs were used to develop a classification tree, which is a statistical method that evaluates each SNP for its predictive ability and selects the best SNPs to create a predictive model [Wilke 2015]. In Newfoundlands, 19 SNPs were used for diagnostic model assessment. The model selected three SNPs for best prediction. Cross validation of the model yielded an area under the ROC curve of 95.5%, indicating that the model was an excellent classifier in this population. In Labrador Retrievers, 13 SNPs were used in the same procedure. The diagnostic model selected all 13 SNPs for best prediction, with an area under the ROC curve of 88.4%, indicating that the model was also a good classifier in the population, but slightly reduced compared to Newfoundland dogs [Wilke 2015]. The SNPs identified in this GWAS did not overlap with regions identified in other ACLR GWAS in dogs [Wilke 2009; Baird 2014]. While the results of this work are promising, they should not be over-interpreted as the population used for GWAS was also used for diagnostic model development

and validation. A truly predictive model must be tested in new populations that were not used to train the model. Nevertheless, these results speak to the idea that a genetic test for ACLR is possible, as accurate stratification of cases and controls can be achieved. Bayesian and machine learning models for genomic prediction in our Labrador Retriever data is presented in Chapter 6.

Chapter 4

Genome-wide Association Analysis in Dogs Implicates 99 Loci as Risk Variants for Anterior Cruciate Ligament Rupture

Baker LA, Kirkpatrick B, Rosa GJ, Gianola D, Valente B, Sumner JP, Baltzer W, Hao Z, Binversie EE, Volstad N, Piazza A, Sample SJ, Muir P. Genome-wide association analysis in dogs implicates 99 loci as risk variants for anterior cruciate ligament rupture. PloS One. 2017; 12:e0173810.

Acknowledgments: This work was supported by grants from the Morris Animal Foundation (D13CA-020, www.morrisanimalfoundation.org), the Robert Draper Technology Innovation Fund from the University of Wisconsin-Madison and the Melita Grunow Family Professorship in Companion Animal Health. Lauren Baker received support from the National Institutes of Health (T32OD10999) and is supported by an NLM training grant to the Computation and Informatics in Biology and Medicine Training Program (NLM 5T15LM007359). Susannah Sample received support from the National Institutes of Health (K01 OD019743-01A1). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors gratefully acknowledge the help of Dr. Aimee Teo Broman and Dr. Christina Kendzierski for advice regarding statistical analysis of data.

Abstract

Anterior cruciate ligament rupture (ACLR) is common in humans and dogs. A non-contact mechanism is typical and second ACLRs are common. Disease prevalence exceeds 5% in several dog breeds. We provide insight into the genetic etiology of ACLR by genome-wide association study (GWAS) in a high-risk breed using 98 case and 139 control Labrador Retrievers. We identified 129 single nucleotide polymorphisms (SNPs) within 99 risk loci. A chromosome 24 locus containing 9 genes with diverse functions met genome-wide significance. Associated regions ($P < 5.0E-04$) explained approximately half of phenotypic variance. GWAS pathways were enriched for c-type lectins, a gene set that includes aggrecan, and a gene set encoding membrane transport. The genetic contribution to dog ACLR was represented by ~172 loci, with heritability estimated at 0.48. We conclude ACLR is a moderately heritable highly polygenic complex trait. Our results implicate c-type lectin pathways in ACL homeostasis.

Introduction

Anterior cruciate ligament rupture (ACLR) can be devastating and life changing, particularly in young adults, as there is a 78% risk of knee arthritis after ACLR that is not influenced by surgical treatment [von Porat 2004]. The incidence of ACLR is high; it is estimated at 33.3-36.9/100,000 person years [Miyasaka 1991; Gianotti 2009]. About 80,000 ACL graft surgeries are performed per year in the USA, with annual costs of approximately \$1 billion [Griffin 2000]. A non-contact mechanism explains most ACLRs, with rupture typically occurring during landing or pivoting movements [Gianotti 2009]. Risk of contralateral rupture is also high at 11.8% of patients [Wright 2011]. Physiological and anatomic factors, such as intercondylar notch shape and posterior tibial slope influence disease risk and risk of ACLR; subsequent contralateral ACLR is more frequent in females [Sutton 2013]. Although the underlying mechanism has not been fully defined, it is widely accepted that ACLR is a complex disorder with both genetic and environmental contributions to disease risk [September 2007].

ACLR patients are twice as likely to have a close relative with ACLR, when compared with individuals without ACLR [Flynn 2005]. Family members of patients with bilateral ACLR are also at higher risk of ACLR [Harner 1994]. Family history of ACLR and a young age substantially increases risk of both ACL graft and contralateral ACLR [Brophy 2012; Webster 2014]. This suggests that a substantial genetic contribution to ACLR exists, although heritability has not been formally estimated. Candidate gene studies have focused on ligament matrix constituents that could influence structural properties, and genes that influence extracellular matrix remodeling. Associations have been identified with a number of genes including the $\alpha 1$ chain of type I collagen (*COL1A1*), the $\alpha 1$ chain of type V collagen (*COL5A1*) and the $\alpha 1$ chain of type XII collagen (*COL12A1*) [Khoschnau 2008; Posthumus 2009a; Posthumus 2009b]. The *COL1A1* polymorphism is located at the binding site for the Sp1 transcription factor [Khoschnau 2008; Posthumus 2009a]. These associations may influence risk of other orthopaedic

conditions, such as shoulder dislocation and Achilles tendon injury [Khoschnau 2008; Posthumus 2009a]. Recently, interaction between the *COL5A1* and *COL12A1* variants has been confirmed [Posthumus 2009b]. In other candidate gene studies, association between ACLR and polymorphisms in matrix metalloproteinase genes has been identified [Malila 2011; Posthumus 2012; O'Connell 2015], particularly association with a polymorphism in the promotor region of matrix metalloproteinase-3 (*MMP-3*, stromelysin-1) in athletes [Malila 2011]. Associations between polymorphisms in vascular endothelial factor A (*VEGFA*) and kinase insert-domain receptor (*KDR*), genes involved in angiogenesis signaling have also been found [Rahim 2014]. Some of these associations are gender-specific [Posthumus 2009a, Posthumus 2009b, Rahim 2014].

In addition to human genetic studies, genome-wide association study (GWAS) has been undertaken in domestic dogs (*Canis lupus familiaris*) with ACLR. Spontaneous non-contact ACLR is common in domestic dogs [Muir 2011a]. Similar to human beings, familial (breed) susceptibility and second ACLRs are typical [Witsberger 2008; Muir 2011a]. The dog is an important model organism for comparative genomic studies, because selective breeding has created distinct genetically isolated populations (breeds) with extensive linkage disequilibrium (LD) and haplotype blocks that are ~10 to 100 times longer than in humans [Sutter 2004; Lindblad-Toh 2005]. The most important factor for risk of disease initiation in dogs is genetic background (breed) [Witsberger 2008]. Disease prevalence exceeds 5% in several breeds (incidence of ~5,000 cases/100,000 dog years), ~100-fold higher than humans, whereas other breeds, such as the Greyhound, experience a much lower disease incidence [Miyasaka 1991; Witsberger 2008; Gianotti 2009]. Heritability of canine ACLR has been estimated at 0.27 in the Newfoundland and 0.28 in the Boxer [Nielen 2001; Wilke 2006]. Candidate gene research using SNP genotyping in dogs has also implicated genes that regulate extracellular matrix composition and ligament strength, including *COL1A1* and *COL5A1* as risk loci for the trait [Baird 2014a]. GWAS has shown that canine ACLR is a complex trait with loci on CFA 1, 3, and

33 in the Newfoundland [Baird 2014b]. Association signals in the *SORCS2* and *SEMA5B* genes suggest neuronal signaling pathways may also influence risk of ACLR [Baird 2014].

We present a GWAS in the domestic dog to discover additional candidate loci associated with ACLR. To take advantage of the LD structure in dogs, the GWAS was conducted in a single high-risk breed, the Labrador Retriever, the most common purebred dog in the USA. Prevalence of ACLR in this breed is 5.79% [Witsberger 2008]. The GWAS identified 99 candidate loci that influence risk of ACLR. A SNP on CFA 24 was significantly associated with ACLR. Pathway analysis implicated c-type lectins, such as aggrecan, in the disease pathogenesis.

Materials and Methods

Ethics Statement

All procedures were performed in strict accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health and the American Veterinary Medical Association and with approval from the Animal Care Committee of the University of Wisconsin-Madison (protocol #V1070). Informed consent of each owner was obtained before participation in the study.

Canine Samples and Phenotyping

DNA was isolated from client-owned Labrador Retrievers using blood or saliva collection swabs. A four-generation pedigree was collected from each dog to ensure purebred status and identify siblings, which were excluded from the GWAS. Each dog underwent an orthopaedic examination that included assessment of knee stability [Muir 1997]. Radiographs of the affected knee(s) were also assessed in cases. In addition, lateral weight-bearing knee radiographs [Muir 1997] were made to screen phenotype-negative control dogs. While it is not possible to identify

the cruciate ligaments radiographically in the dog, compression of the infrapatellar fat pad in the knee by synovial effusion and knee osteophytosis are degenerative changes typically associated with ACLR [Chuang 2014]. Dogs were considered cases if anterior translation of the tibia was detected clinically and radiographic signs were consistent with ACLR. Labrador Retrievers ≥ 8 years of age have $\sim 0.35\%$ chance of developing ACLR [Reif 2004; Witsberger 2008]. Therefore, control dogs were ≥ 8 years of age with a normal orthopaedic clinical exam and normal knee radiographs. Habitual activity of each dog was documented using a questionnaire.

Genome-Wide Association

Genome-wide SNP genotyping was performed in 98 cases and 139 controls using the Illumina CanineHD BeadChip, which genotypes 173,662 SNPs evenly spaced across the genome. Data underwent quality control filtering using PLINK [Chang 2015]. All samples had a genotyping call rate of $\geq 95\%$. 49,859 SNPs were excluded because minor allele frequency (MAF) was ≤ 0.05 and 7,468 SNPs were excluded because of a low genotyping rate ($\leq 95\%$). 153 SNPs were excluded because of deviation from Hardy-Weinberg equilibrium at $P < 1E-07$. 118,992 SNPs were used for further analysis.

To account for ancestral population structure and family relatedness in the study dogs, single marker linear mixed model (LMM) analysis was performed using GCTA (Genome-wide Complex Trait Analysis) [Yang 2011a] and GEMMA (Genome-wide Efficient Mixed Model Association) [Zhou 2012], software tools optimized for complex trait GWAS. Penalized Unified Multiple-locus Association (PUMA), in which all SNPs are analyzed together, was also used to aid detection of weaker associations often found in complex traits [Hoffman 2013]. We used logistic regression and a 2D-MCP penalty for this analysis [Tang 2014]. In the PUMA analysis, the first 20 eigenvectors were used as covariates in the association analysis to correct for population structure. Eigenvectors were obtained by principal component analysis using GCTA.

Because neutering has a significant effect on risk of ACLR [Whitehair 1993; Witsberger 2008], it was included as a covariate with the GEMMA, GCTA, and PUMA analyses.

Genome-Wide Significance

We defined genome-wide significance using permutation testing. Use of a Bonferroni correction for the number of SNPs tested is too conservative in dog breeds, as extensive LD means that SNPs are often inherited in haplotype blocks [Lindblad-Toh 2005]. We defined genome-wide significance by randomly permuting the phenotypes and re-running the GWAS LMM 1,000 times. Genome-wide significance was defined by identifying the 5% quantile of the set of minimum P -values from the GWAS permutations. Additionally, we estimated the number of haplotype blocks in the Labrador Retriever SNP data using PLINK [Chang 2015], with LD windows of 500kb, 1Mb, and 5Mb and used the number of haplotype blocks to estimate genome-wide significance by Bonferroni correction of $P < 0.05$. To facilitate further dissection of genetic variants associated with the ACL phenotype, we also identified a larger set of candidate ACLR regions at $P < 5E-04$ [Karlsson 2013]. Although some of the regions included may not be true associations, this would likely weaken rather than strengthen the gene set and pathway analyses, leading to false negatives rather than false positives [Karlsson 2013].

Defining Associated Loci in the Genome

Linkage-disequilibrium (LD) clumping using PLINK [Chang 2015] was used to define regions of association with the ACLR trait from the GWAS results. LD clumping defined regions around SNPs associated at $P < 5E-04$ within 1 Mb of the index SNP ($r^2 > 0.8$ and $P < 0.01$). We also used GCTA to estimate in-breeding coefficient and the amount of phenotypic variance explained by the associated loci, which were defined as SNPs with $r^2 > 0.2$ within 5Mb of the peak SNP in each locus [Yang 2011a; Tang 2014; Karlsson 2013].

For complex trait GWAS with a large number of risk loci, sites that are not discovered are expected to have smaller effect sizes in a second generation GWAS, because those with larger effect sizes will have been identified in the first round of GWAS [Park 2010]. To estimate the number of risk loci that are likely associated with ACLR, we used INPower [Park 2010]. Odds ratios were corrected for the winner's curse before INPower analysis was performed [Ghosh 2008; Park 2010].

Genetic Risk Score Computation

Two approaches were used to calculate the genetic risk scores (GRS), a simple risk alleles count method (cGRS) and a weighted method (wGRS) [Chen 2011]. The wGRS weights each risk allele by the logarithm odds ratio ($\text{Log}(\text{OR})$) for that allele. The wGRS is a linear combination of the number of risk alleles weighted by the $\text{Log}(\text{OR})$ as coefficients [Chen 2011]. The Mann-Whitney U test was used to compare cGRS scores for each LMM in case and control groups. To estimate the total risk captured by the genetic risk scoring for each LMM, we calculated the odds ratios according to the wGRS quartiles [Karlsson 2008]. We also measured the discriminative power attributable to the GRS by plotting receiver operating characteristic (ROC) curves and calculated the area under the curve (AUC) for the Labrador Retriever case and control dogs. AUC 95% confidence intervals were calculated using 2000 stratified bootstrap replicates. An R software package (pROC) was used for these analyses [Robin 2011].

Pathway Analysis

Pathway analysis was performed with two methods. DAVID [Wagget 2006] analyses were run on the ACLR loci identified from our GWAS. ACLR loci were transposed to CanFam 3.1 coordinates (genome.ucsc.edu/cgi-bin/hgLiftOver) with 500kb flanks added to the start and end and gene size correction turned on and stringency set to high. A list of genes from the

LiftOver coordinates was then analyzed. Probability values were evaluated after Benjamini correction with DAVID.

Pathway analysis with INRICH [Lee 2014] was performed on canFam2 intervals using a map file lifted over from the CanFam3.1 Broad Improved Canine Annotation catalog (UCSC Genome Browser). We used 1,000,000 permutations matched for region size, SNP density, and gene number. INRICH reports significance for each gene set and the experiment-wide significance, correcting for the number of gene sets (P_{corr}). We considered $P_{\text{corr}} < 0.05$ to be significant [Karlsson 2013]. We tested gene sets from the KEGG (Kyoto Encyclopedia of Genes and Genomes), Gene Ontology, and MSigDB (Molecular Signatures Database).

Heritability Estimation

Genomic heritability was estimated from SNPs using the BGLR statistical package [Perez 2014]. SNPs with missing genotypes were filtered out using PLINK [Chang 2014]. Heritability estimation was performed using 99,103 SNPs. A genomic best linear unbiased prediction (GBLUP) model was fitted using a SNP-derived genomic relationship matrix, which is equivalent to a non-parametric reproducing kernel Hilbert spaces (RKHS) method [Perez 2014]. Narrow sense genetic heritability was also estimated using a data matrix prepared from pedigrees. To fit the model, 30,000 iterations of the Gibbs sampler were used with burn-in of 5,000 iterations. A correction factor was used to transform the heritability estimate on the observed scale from the regression model to the liability scale for a binary trait [Zhou 2013] and a population prevalence of 0.0579 [Witsberger 2008] was used for this correction.

Results

GWAS Population of Labrador Retrievers

We genotyped 237 Labrador Retrievers using the Illumina CanineHD BeadChip. All dogs had individual call rates of >95%. The final dataset contained 118,992 SNPs from 98 cases and 139 phenotype-negative controls. Median inbreeding coefficient was 0.025 (Figure 4.1). The ratio of females to males in the case and control groups was 0.92 and 0.83 respectively. Of the 114 females, 99 were ovariohysterectomized (0.87). Of the 123 males, 96 were castrated (0.78). Mean age of the dogs in the case and control groups was 6.0 ± 2.5 years and 10.4 ± 1.7 years, respectively.

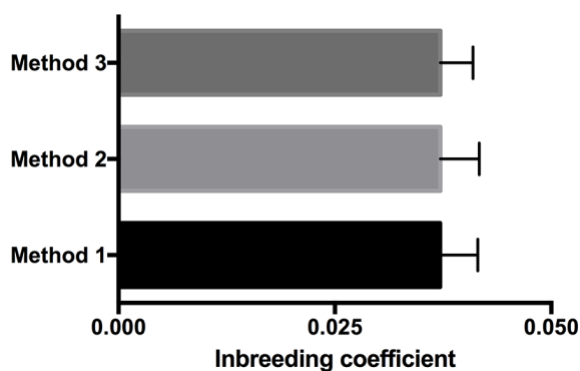


Figure 4.1. The Labrador Retriever is a dog breed with a relatively low amount of inbreeding. Whiskers represent the minimum and maximum values for each analysis method (n = 237 dogs).

GWAS Identifies 99 Regions Associated with Anterior Cruciate Ligament Rupture

We tested for association between ACLR and SNPs with a MAF >0.05 in the Labrador Retriever breed, controlling for cryptic relatedness and population structure using linear mixed model (LMM) analysis with three approaches, including a penalized multiple regression method for improved detection of weak associations. We identified all SNPs with either significant association based on analysis of 1,000 random phenotype permutations to define genome-wide significance ($P < 1.549 \times 10^{-6}$ for GCTA, $P < 6.097 \times 10^{-7}$ for GEMMA and $P < 4.35 \times 10^{-7}$ for PUMA) or suggestive association ($P < 5.00 \times 10^{-4}$; Figure 4.2; Methods) and defined regions of association

through linkage disequilibrium (Table 4.1, Table 4.1S). Control dogs were considered phenotype-negative because of the selection criteria used for recruitment. We identified 21,713, 21,754, and 21,861 haplotype blocks in the Labrador Retriever genome with LD windows of 5Mb, 1Mb, and 5kb respectively, yielding a genome-wide significance estimate of $P < 2.29 \times 10^{-6}$ to $P < 2.30 \times 10^{-6}$.

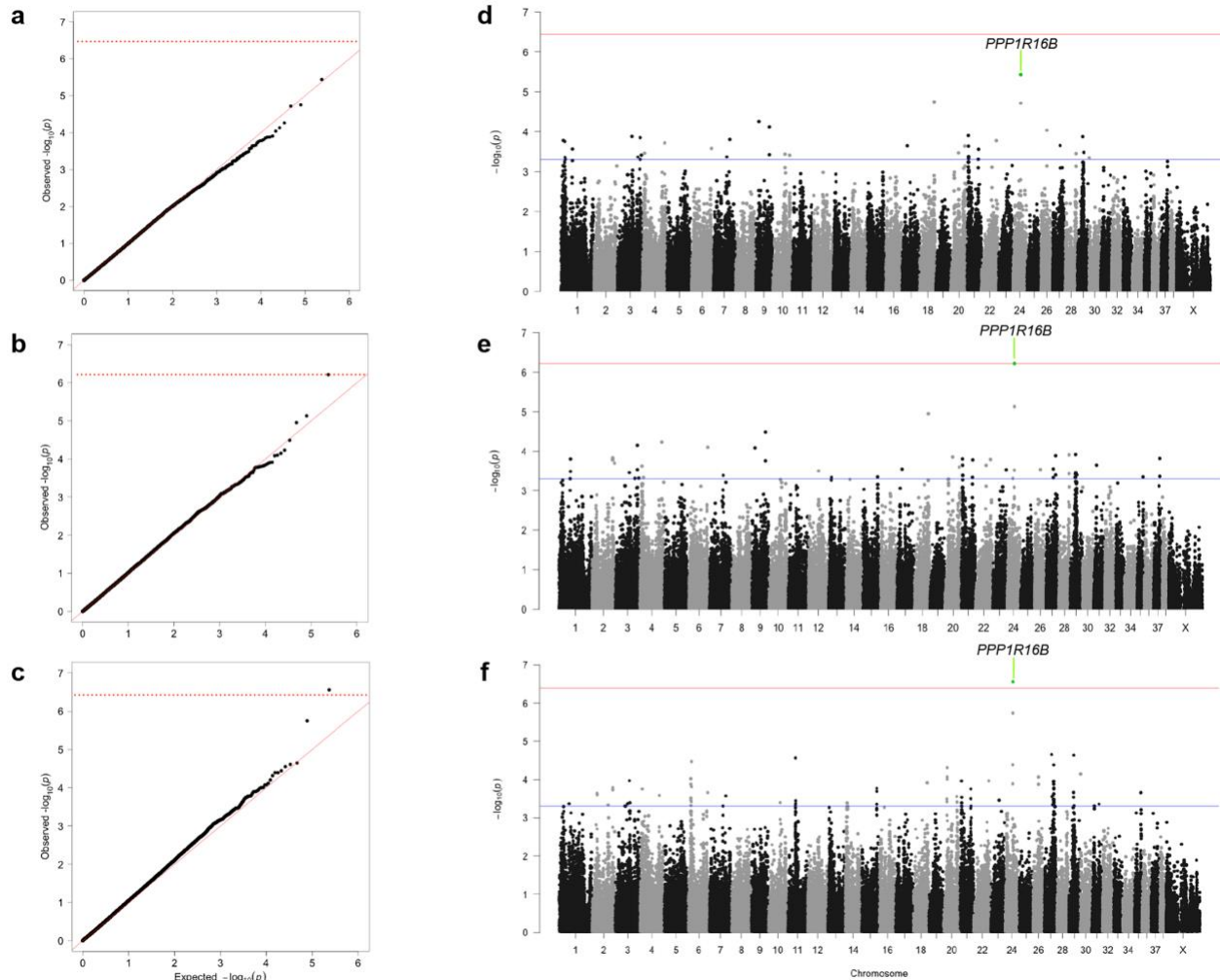


Figure 4.2. Linear mixed model GWAS corrects for population structure and identifies 99 ACLR associated loci explaining a large proportion of phenotypic variance. For each linear mixed model (LMM), the QQ plots show no evidence of population stratification relative to the expected distribution. Permutation testing with each model determined genome-wide significance at (a) $P < 3.63\text{E-}7$ for GCTA (Genome-wide Complex Trait Analysis), $\lambda = 0.987$ (b) $P < 6.097\text{E-}7$ for GEMMA (Genome-wide Efficient Mixed Model Association), $\lambda = 0.994$ and (c) $P < 4.01\text{E-}7$ for PUMA (Penalized Unified Multiple-locus Association, $\lambda = 1.012$). The plots represent analysis of 118,992 SNPs from 98 cases and 139 phenotype-negative controls. (d) With GCTA, 36 loci have $P < 5\text{E-}4$, with the most significant locus located in CFA24, which did not meet genome-wide significance defined by minimum p-values from permutation testing. (e) With GEMMA, 47 loci have $P < 5\text{E-}4$, with the locus on CFA24 meeting genome-wide significance defined by minimum p-values from permutation testing. (f) With PUMA, 65 loci were significant at $P < 5\text{E-}4$ and the locus on CFA24 exceeded genome-wide significance defined by minimum p-values from permutation testing. The single SNP that met genome-wide significance lies within the gene *PPP1R16B*.

Table 4.1. ACLR associated loci identified by GWAS in the Labrador Retriever, a dog breed with a high disease prevalence

SNP	Chr	Position	<i>P</i>	LMM	Risk allele	f(A)	f(U)	OR	Region start-end	Size (kb)	Genes
BICF2G630500368	24	30241088	2.76E-07	1,2,3	G	0.83	0.66	2.56	30241088-30245795	5	<i>BPI</i> , <i>LBP</i> , <i>RALGAPB</i> , <i>SLC32A1</i> , <i>ADIG</i> , <i>ACTR5</i> , <i>PPP1R16B</i> , <i>FAM83D</i> , <i>DHX35</i>
BICF2P1121006	18	54279578	1.11E-05	1,2,3	A	0.63	0.42	2.28	No LD		Many (40 genes)
BICF2S2356299	27	30557856	2.21E-05	2,3	A	0.43	0.27	2.03	No LD		<i>AEBP2</i> , <i>PLEKHA5</i>
BICF2P483191	29	21601273	2.31E-05	1,2,3	C	0.73	0.51	2.54	No LD		<i>C29H8orf34</i> , <i>SULF1</i> , <i>SLCO5A1</i>
BICF2P50610	11	32270617	2.75E-05	3	A	0.29	0.19	1.7	31939564-32270617	331	<i>C11H9orf123</i> , <i>PTPRD</i>
BICF2P890246	9	53427907	3.23E-05	1,2	A	0.16	0.36	2.99	53427907-53432248	4	Many (20 genes)
BICF2S23324965	6	14077648	3.36E-05	3	G	0.68	0.60	1.42	14077648-14092057	14	<i>TRRAP</i> , <i>TMEM130</i> , <i>NPTX2</i> , <i>BAIAP2L1</i> , <i>BRI3</i> , <i>TECPR1</i> , <i>LMTK2</i> , <i>PMS2</i> , <i>EIF2AK1</i> , <i>ANKRD61</i> , <i>USP42</i> , <i>CYTH3</i>
BICF2P544126	24	29772193	4.09E-05	3	G	0.94	0.87	2.28	29772193-29794411	22	<i>CTNBL1</i> , <i>VSTM2L</i> , <i>TTI1</i> , <i>RPRD1B</i> , <i>TGM2</i> , <i>KIAA1755</i> , <i>BPI</i> , <i>LBP</i> , <i>RALGAPB</i> , <i>ADIG</i> , <i>SLC32A1</i> , <i>ACTR5</i> , <i>PPP1R16B</i> , <i>FAM83D</i>
BICF2P526639	27	39217437	4.12E-05	2,3	G	0.23	0.12	2.18	39211186-39217437	6	<i>KLRD1</i> , <i>GABARAPL1</i> , <i>TMEM52B</i> , <i>OLR1</i> , <i>CLEC7A</i> , <i>CLEC1B</i> , <i>CLEC12B</i> , <i>CLEC12A</i> , <i>CLEC2B</i> , <i>KLRF1</i> , <i>CD69</i> , <i>CLEC2D</i> , <i>KLRB1</i>
BICF2P1462185	20	15053718	4.89E-05	3	A	0.85	0.74	1.90	14838270-15053718	215	<i>EDEM1</i> , <i>ARL8B</i>

BICF2P1208798	9	12671217	5.49E-05	1,2	G	0.56	0.36	2.27	No LD		Many (20 genes)
BICF2G630175389	4	84260906	5.87E-05	1,2	A	0.83	0.68	2.28	No LD		<i>CDH10</i>
BICF2S24415473	3	86974042	7.07E-05	1,2	G	0.40	0.26	1.97	86948527-86974042	26	<i>STIM2, TBC1D19, CCKAR, RBPJ, SEL1L3</i>
BICF2G630412697	30	3126573	7.22E-05	1,3	G	0.96	0.86	4.23	No LD		<i>OR4K2, OR4K1, ORFN5</i>
BICF2P498515	6	75848537	7.89E-05	1,2,3	A	0.16	0.06	3.11	No LD		<i>LRRIQ3</i>
BICF2P792911	26	22894961	8.55E-05	1,2,3	G	0.44	0.27	2.14	No LD		<i>ADRBK2, MYO18B, SEZ6L, ASPHD2, HPS4, SRRD, TFP11, TPST2, CRYBB1, CRBA4</i>
BICF2G630810143	6	11130832	9.46E-05	3	A	0.44	0.32	1.72	11130832-11177149	46	<i>UPK3B, LRWD1, ALKBH4, ORAI2, PRKRIP1, SH2B2, CUX1, MYL10, RABL5, FIS1, ZNHIT1, PLOD3</i>
BICF2P564273	3	55250188	1.07E-04	1,2,3	A	0.70	0.52	2.16	No LD		<i>ACAN, HAPLN3, MFGE8, ABHD2, RLBP1, FANCI, POLG, RHCG, TICRR, C3H15orf38, KIF7, PLIN1, PEX11A, WDR93, AMPN, C3H15orf38</i>
TIGRP2P297337	22	58201452	1.08E-04	1,2,3	A	0.44	0.27	2.2	No LD		<i>EFNB2, ARGLU1</i>
BICF2G630658881	21	7582214	1.09E-04	1,2,3	G	0.49	0.32	2.12	7581714-8383209		<i>JRKL, CCDC82, MAML2, MTMR2, CEP57, FAM76B, SESN3</i>

Note: OR odds ratio calculated from PLINK. LMM Linear mixed model 1 – GCTA, 2 – GEMMA, 3 – PUMA. F(A) and F(U) represent the frequency of the risk allele in case and control dogs, respectively. Data represent the twenty most significant loci of 99 associations with canine ACLR. SNP position and genomic regions are based on CanFam 2.0. Genes lists were derived from the SNP locus or LD block with 500kb flanking regions after liftover to CanFam 3.1

With the Labrador Retriever breed, associated regions ($P < 5.0 \times 10^{-4}$) explained approximately half of the phenotypic variance in the ACLR trait (Figure 4.3). For GCTA, 36 loci in 72.7Mb of the genome explained 48.09% of the phenotypic variance. For GEMMA, 47 loci in 82.7Mb of the genome explained 55.88% of the phenotypic variance. For PUMA, 65 loci in 86.58Mb of the genome explained 50.28% of the phenotypic variance in the ACLR trait.

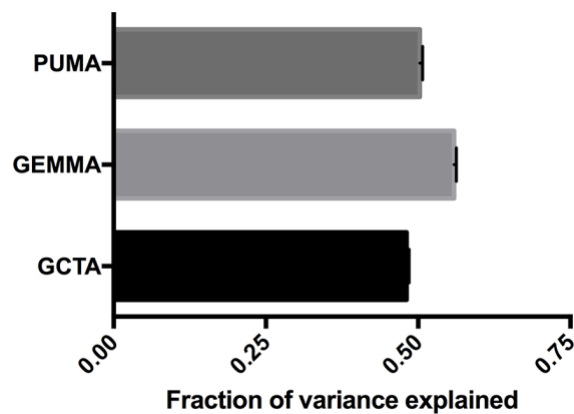


Figure 4.3. Phenotype variance was explained to a large degree by the associated genomic loci. Loci identified by linear mixed model (LMM) analysis were broadly defined as SNPs with $r^2 > 0.5$ within 5Mb of the peak SNP. (a) For GCTA, 36 loci in 72.7Mb of the genome explained 48.09% of the phenotypic variance. (b) For GEMMA, 47 loci in 82.7Mb of the genome explained 55.88% of the phenotypic variance. (c) For PUMA, 65 loci in 86.58Mb of the genome explained 50.28% of the phenotypic variance in the ACLR trait.

We identified 129 SNPs associated with canine ACLR. Using LD clumping, we found that these SNPs reside in 99 loci. Two of these regions were located in uncharacterized or non-coding regions of the genome. A SNP on CFA24 met genome-wide significance for LMM association analysis with GEMMA ($P = 6.10 \times 10^{-7}$) and PUMA ($P = 2.77 \times 10^{-7}$), but not GCTA ($P = 3.63 \times 10^{-6}$). This SNP resides in a 5kb haplotype block with one other SNP. Nine genes are located within the locus defined by 500kb flanking regions including bactericidal/permeability-

increased protein (*BPI*), lipopolysaccharide binding protein (*LBP*), Ral GTPase activation protein beta subunit (*RALGAPB*), adipogenin (*ADIG*), solute carrier family 32, member 1 (*SLC32A1*), ARP5 actin-related protein 5 (*ACTR5*), protein phosphatase 1, regulatory subunit 16B (*PPP1R16B*), family with sequence similarity 83, member D (*FAM83D*), and DEAH (Asp-Glu-Ala-His) box polypeptide 35 (*DHX35*). Although many risk loci contained large numbers of genes, two loci did not (Table 4.1, Table 4.1S), suggesting these SNPs are associated with mutations affecting gene expression (rSNPs).

Power analysis of our GWAS data set using INPower estimates that 172 loci explain the genetic contribution to ACLR in the Labrador Retriever (Table 4.2S). INPower estimates that in a future GWAS, large numbers of dogs will be needed for discovery of future loci (>1,500) (Table 4.3S).

Risk loci clearly distinguish ACLR cases from controls

To evaluate the cumulative effects of associated ACLR risk loci, we used a genetic risk scoring approach using a simple allele count (cGRS) or a weighted approach (wGRS). We found significant differences in the number of risk alleles in cases and controls for GCTA ($P < 2.2 \times 10^{-16}$), GEMMA ($P < 2.2 \times 10^{-16}$), and PUMA ($P < 2.2 \times 10^{-16}$) (Table 4.2), with a shift to increased numbers of risk alleles in the cases (Figure 4.4). When the odds ratios according to the wGRS quartiles for each LMM were calculated, there was a significant increase in ACLR odds ratios with increasing wGRS quartile for all three LMM, using the first wGRS quartile as a reference (Figure 4.4, Figure 4.1S).

Table 4.2. Genetic risk scoring in ACLR case and control Labrador Retriever dogs using GWAS associated SNPs from linear mixed model analysis

LMM	Number of Risk alleles		Significance
	Control	Case	
GCTA	24 (16, 40)	35 (24, 43)	$P<2.2E-16$
GEMMA	33 (22, 43)	45 (29, 55)	$P<2.2E-16$
PUMA	62 (37, 84)	77 (56, 99)	$P<2.2E-16$

Note: Data represent median (range) for allele counting (cGRS) [51]. LMM Linear mixed models used were GCTA, GEMMA, and PUMA. The Mann-Whitney U test was used to determine significance.

AUC differences between cGRS and wGRS were small and we found that there were no significant differences in ROC AUC for cGRS and wGRS for any of the three LMM analyses. For both cGRS and wGRS analyses, GCTA and GEMMA yielded increased ROC AUC values, when compared with PUMA. Overall, cGRS for GEMMA yielded the highest AUC at 0.9634, indicating that this algorithm is most efficient at stratifying case and control dogs (Table 4.3).

Table 4.3. Receiver operating characteristic (ROC) analysis of genetic risk scoring in ACLR case and control Labrador Retriever dogs using GWAS associated SNPs from linear mixed model analysis

LMM	cGRS		wGRS	
	AUC	95% confidence interval	AUC	95% confidence interval
GCTA	0.9487	0.9191-0.9725	0.9464	0.9183-0.9694
GEMMA	0.9634	0.9369-0.9824	0.9601	0.933-0.9801
PUMA	0.8842*	0.8356-0.9158	0.8909*	0.8458-0.9263

Note: LMM Linear mixed models used were GCTA, GEMMA, PUMA. AUC= Area under the ROC curve. Simple risk alleles count method (cGRS) and weighted method (wGRS).

*Significantly different from GCTA and GEMMA ($P<0.005$ for cGRS and $P<0.05$ for wGRS).

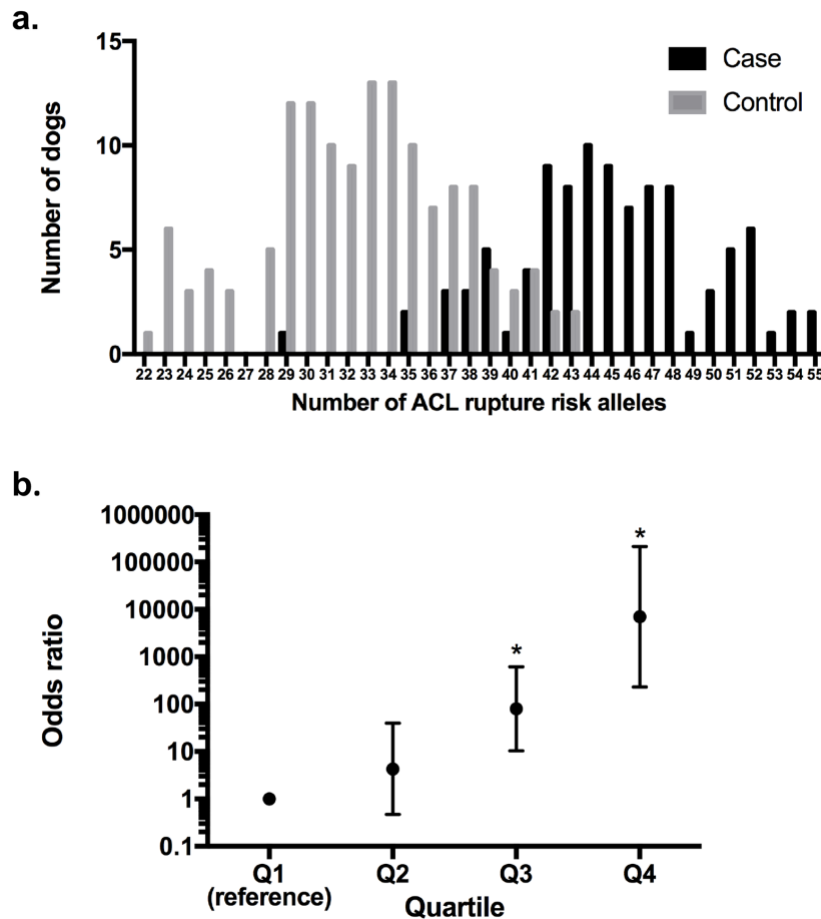


Figure 4.4. Genetic risk scoring using GWAS associated loci from linear mixed model analysis with GEMMA segregates ACLR disease risk in case and control Labrador Retriever dogs. (a) Distribution of the number of ACLR risk loci in case and control groups of Labrador Retriever dogs. The number of risk alleles in cases and controls is significantly different ($P < 2.2 \times 10^{-16}$) (b) ACLR odds ratios of weighted genetic risk scores (wGRS) relative to the first quartile. Vertical bars represent the 95% confidence intervals. * Odds ratio is significantly different from the reference first quartile.

GWAS pathways are enriched for carbohydrate binding proteins

Functional annotation clustering using DAVID revealed significant enrichment of a cluster of 24 genes encoding carbohydrate binding proteins ($P < 1.84 \times 10^{-7}$, $P_{corr} = 1.21 \times 10^{-4}$) that includes aggrecan (*ACAN*), a large structural protein that stabilizes the collagen network in ligament matrix [Mannion 2014]. The majority of these proteins are c-type lectin receptors (CLRs). Enrichment of a cluster of 11 genes encoding proteins with antimicrobial activity was also found ($P < 1.74 \times 10^{-4}$, $P_{corr} = 0.02$) (Table 4.4), including *LBP* and *BPI* which were also present in the locus on CFA 24 that met genome-wide significance. A third cluster of 21 genes involved with nucleosome assembly was also highly significant, but enrichment was likely overinflated as 17 of 18 genes in the cluster are located in a 1Mb region surrounding a single SNP. Using INRICH, we identified enrichment for a single set of genes (*TTR*, *SLC9A5*, *SLC10A1*, *SLC37A4*, *SLC6A1*, *AQP9*, *GABRP*, *GJB1*, *KCNJ3*, *ALB*, *GABRB3*, *P2RX1*, *SLC16A2*) ($P < 5.0 \times 10^{-4}$, $P_{corr} = 0.07$). This pathway primarily consists of genes encoding membrane transport proteins with a wide range of physiological functions including pH regulation, glucose homeostasis, and signal transduction.

Table 4.4. Functional annotation clustering of genes in regions associated with anterior cruciate ligament rupture identified by GWAS in the Labrador Retriever

Cluster Term	Gene	Location start-end (CanFam 3.1)
Carbohydrate binding	<i>ACAN</i>	chr3:51995108-52032255
	<i>HAPLN3</i>	chr3:52033940-52037803
	<i>NPTX2</i>	chr6:10805532-10808481
	<i>CD248</i>	chr18:50991445-50997063
	<i>CLEC17A</i>	chr20:48069197-48079333
	<i>CALR</i>	chr20:49250753-49255568
	<i>MAN2B1</i>	chr20:49445696-49461722
	<i>FCER2</i>	chr20:52449871-52457459
	<i>CLEC4G</i>	chr20:52476065-52480126
	<i>CD209</i>	chr20:52516871-52521506
	<i>KLRK1</i>	chr27:35634936-35658993
	<i>KLRD1</i>	chr27:35698995-35706308
	<i>OLR1</i>	chr27:35828941-35838821
	<i>CLEC7A</i>	chr27:35864121-35867654

	<i>CLEC9A</i>	chr27:35919910-35934027
	<i>CLEC1B</i>	chr27:35938182-35964350
	<i>CLEC12B</i>	chr27:35953661-35961617
	<i>CLEC12A</i>	chr27:36040314-36051214
	<i>CLEC2B</i>	chr27:36120649-36145641
	<i>CD69</i>	chr27:36197203-36205042
	<i>CLEC2D</i>	chr27:36274914-36301802
	<i>KLRF1</i>	chr27:36152112-36157825
	<i>KLRB1</i>	chr27:36355936-36376701
	<i>KLRG1</i>	chr27:36740984-36742709
	<i>M6PR</i>	chr27:36784859-36798029
Antimicrobial	<i>DEFB132</i>	chr24:20589296-20592303
	<i>DEFB128</i>	chr24:20652527-20654682
	<i>DEFB126</i>	chr24:20725869-20727386
	<i>DEFB125</i>	chr24:20743523-20757512
	<i>DEFB118</i>	chr24:20771397-20900359
	<i>DEFB116</i>	chr24:20833054-20840751
	<i>DEFB121</i>	chr24:20927258-20929346
	<i>BPI</i>	chr24:26772292-26801096
	<i>LBP</i>	chr24:26810094-26821440
	<i>HIST1H2BG</i>	chr35:24178933-24179313
	<i>HIST1H2BC</i>	chr35:24989922-25000072

ACLR in the Labrador Retriever has moderate heritability.

Using a Bayesian method, SNP heritability and narrow sense genetic heritability of ACLR was estimated at 0.538 and 0.521 using SNP markers and pedigrees, respectively. After correction to the liability scale for a binary trait, these estimates were 0.493 and 0.476, respectively.

Discussion

By undertaking a within-breed GWAS in the Labrador Retriever, we found 99 regions of association with the trait, suggesting that ACLR is a complex, potentially highly polygenic condition. These loci explained between 48% and 56% of the disease phenotypic variation, depending on which LMM was used for the association analysis, suggesting that inherited factors make an important contribution to the disease in the Labrador Retriever dog model. We estimated narrow sense genomic heritability to be 0.48-0.49, higher values than past estimates in the Newfoundland and Boxer breeds [Nielen 2001; Wilke 2006].

Our studied sample of Labrador Retriever dogs represented typical features of the general population, with an approximately equal number of male and female dogs and a large majority of the dogs being neutered by castration or ovariectomy, respectively. ACLR in dogs is an acquired condition [Muir 2011a; Chuang 2014]. In the present study, ACLR cases were middle-aged dogs typically, with a mean age of 6.0 years. In dogs, loss of sex steroids through neutering is a risk factor for ACLR [Whitehair 1994; Witsberger 2008]. In human beings, ACLR is predisposed to female athletes [Sutton 2013]. Knee laxity in women is lowest in the follicular phase of the menstrual cycle (low estrogen), when ACLR is most common [Beynon 2006; Hewett 2007]. This suggests that the influence of sex steroid levels on ACL laxity in both species may influence accumulation of matrix damage over time and, consequently, risk of rupture.

A large majority of the risk loci discovered in this study differ from a recent GWAS in the Newfoundland breed, which identified loci predominantly on CFA 1, 3, and 33 [Baird 2014]. Although overlap of risk loci between dog breeds needs more investigation, a different genetic architecture in different breeds suggests that this complex trait likely consists of many genetic variants that are concentrated differently in different breeds through population bottlenecks and intense selection [Karlsson 2013]. A similar scenario likely summarizes the genetic basis of

human ACLR [Robinson 2014], although GWAS of human ACLR has not been performed to date.

Because of the high LD within breeds of dog, risk loci often contained large numbers of genes. However, two risk loci appeared to contain rSNPs located in gene deserts in intergenic regions of the genome of >500kb that lack annotated genes or protein-coding sequences [Schierding 2014]. Complex trait disease is caused by disturbance to biological networks, not by isolated genes or proteins. Regulatory SNPs can influence gene expression through a number of mechanisms that include the three dimensional organization of the genome, RNA splicing, transcription factor binding, DNA methylation, and long non-coding RNAs (lncRNA) [Schierding 2014; Huang 2015]. Investigation of SNPs associated with complex trait disease in dogs with potential regulatory function through expressed quantitative trait loci (eQTL) studies or other methods is currently lacking.

One locus consisting of a 5kb haplotype block with one other SNP on CFA 24 met genome-wide significance in the present study. Nine genes were identified in this block with diverse physiological effects on cellular and tissue homeostasis. For example, *ACTR5* plays an important role in chromatin remodeling during transcription, DNA repair, and DNA regulation [Kitayana 2009]. *DHX35* encodes an ATP-ase that plays a role in RNA splicing [Ilagan 2013]. *RALBAPB* as well as *FAM83D* are both important for mitotic regulation [Personnic 2014; Liao 2015]. While a relationship between cellular homeostasis/proliferation and ACLR has not been established, it is feasible that aberrations in the genes that govern these processes could have a wide range of effects that may alter ligament tissue integrity and homeostasis. Other genes in this block include *LBP* and *BPI*, which have an important function regarding immuno-stimulatory capacity of innate immune mechanisms. Notably, our top SNP resides within *PPP1R16B*. *PPP1R16B* encodes a protein that promotes angiogenesis through inhibition of phosphatase and tensin homolog (*PTEN*) [Obeidat 2014]. Angiogenesis-associated signaling is important for

ligament matrix remodeling after mechanical loading, and variations in this cascade have been associated with non-contact ACLR risk [Rahim 2014].

To further investigate the large number of genes we identified within risk loci, we also undertook pathway analysis of our data using two different methods. Pathway analysis using DAVID revealed an association with a cluster of 24 carbohydrate-binding protein genes. The majority of these proteins were c-type lectin receptors (CLRs). CLRs primarily function as pattern recognition receptors; they play roles in activation or suppression of the immune response through recognition of microbial, fungal, or self molecules, including recognition of MHC class 1 [Dambuza 2015; Yokoyama 2003]. Aberrant immune function may play a role in the pathogenesis of canine ACLR, as development of synovial inflammation is an early event in disease pathogenesis [Bleedorn 2011] and is a significant factor influencing disease progression [Chuang 2014]. Other genes in the cluster include aggrecan (*ACAN*) and hyaluronan and proteoglycan link protein 3 (*HAPLN3*), or cartilage link protein. Aggrecan is a large aggregating proteoglycan that interacts with cartilage link protein and hyaluronic acid to form stable aggregates in collagenous tissues [Spicer 2003]. Through binding to fixed charged groups, the proteoglycan aggregate maintains osmotic pressure in collagenous tissues to promote water retention. Tissue hydration is important for efficient distribution of load and for the ability of cells to accomplish repair [Halper 2014]. Equine degenerative suspensory ligament desmitis (DSLDD), a debilitating disorder of horses that leads to collagen disruption and eventual rupture of the suspensory ligament, is associated with a 15-fold increase in aggrecan content of affected ligaments [Plaas 2011]. Moreover, recent work has linked human *ACAN* rs1516797 with the risk of ACL injury in both male and female participants [Mannion 2014]. A separate study revealed *ACAN* gene expression is up-regulated in ACL samples from female compared to male patients that have undergone ACL reconstruction, suggesting a possible etiology for the observed sex differences among patients with ACLR [Johnson 2015]. The precise mechanism

by which *ACAN* up-regulation may lead to ligament weakening is currently unclear, though a structural change appears to be the most likely etiology [Plaas 2011; Johnson 2015].

DAVID also revealed an association with a cluster of 11 genes encoding proteins with antimicrobial activity, including *LBP* and *BPI*, which reside within the locus that met genome-wide significance in this study. These proteins work together to bind lipopolysaccharide, aiding in host defense against gram-negative organisms. DNA from a wide variety of bacterial species has been identified in the synovium of both human and canine arthritis patients [Gérard 2001; Muir 2007]. While a causal link between the presence of bacterial DNA and development of arthritis has not yet been established, this information suggests that interactions between antimicrobial proteins and environmental bacteria may play a role in the pathogenesis of ACLR. Additionally, seven of the genes in this cluster encode beta-defensin proteins. Defensins are antibiotic peptides involved in host defense at epithelial and mesenchymal surfaces. Increased expression of beta defensins has been implicated in the pathogenesis of osteoarthritis in both human beings and mouse models [Varoga 2006].

We also tested genomic regions associated with ACLR for gene set enrichment using INRICH. One pathway, module 415 from the Molecular Signatures Database, was inflated. This pathway included 13 genes, most of which encode membrane transport proteins with various physiological roles. *GJB1* is a member of the large connexin family and encodes connexin 32, a gap junction protein that has been implicated in the regulation of collagen synthesis and the matrix remodeling response to mechanical loading of tendon [Young 2009; Waggett 2006]. Other genes in this module are associated with central nervous system function. *SLC6A1*, *GABRP*, and *GABRB3* are all associated with GABA signaling and mutations in *TTR* and have been associated with sensorimotor polyneuropathy [Dohrn 2013]. Previous work has suggested a role for neurological pathways in susceptibility to ACLR in Newfoundland dogs [Baird 2014b].

ACLR GRSs were calculated for each dog to determine the cumulative effect of ACLR-associated loci on disease risk. While previous work found that wGRS better accounted for genetic risk [Chen 2011], our study found no difference between cGRS and wGRS for any of the LMMs used. This is consistent with the idea that the ACLR phenotype is associated with a large number of genetic loci with small effects. In diseases with genetic loci with large effects, wGRS would more accurately represent the cumulative effect of individual loci on genetic risk. Overall, classification capability of GRS is high, with a cGRS for GEMMA AUC of approximately 96%, indicating that we have clearly captured genetic loci that contribute to ACLR risk in our LMM association analysis. Additionally, the proportion of variance explained by risk SNPs was calculated separately for each algorithm. Of the three, SNPs identified by GEMMA captured the largest proportion of variance. It should be noted that these estimates for classification capability as well as phenotypic variance explained are likely inflated as the same data were used for SNP selection, classification capability, and variance estimation. Future work should include verification of these results by applying these methods to an independent test cohort of case and control dogs.

SNP-based and pedigree-based heritability of ACLR were estimated at 0.49 and 0.48, respectively, using a Bayesian method. These estimates are considerably higher than restricted maximum likelihood (REML) heritability estimates calculated for other breeds of dog [Nielen 2001; Wilke 2006]. It is unclear whether ACLR is truly more heritable in the Labrador Retriever than in other breeds or if the higher value is a reflection of sampling variation or the Bayesian method used. REML estimation of heritability was attempted but the algorithm did not converge, probably because of the size or structure of the data set.

A limitation for this study is sample size. Canine GWAS for a complex trait requires approximately 100 cases and 100 controls to detect a five-fold risk allele [Karlsson 2008], suggesting our study has reasonable power. Only one of 99 regions identified associated with ACLR met genome-wide significance. For complex trait diseases in dogs with high population

prevalence and risk alleles that have relatively small effects, larger data sets with additional dogs may be required to improve statistical power. INPower analysis of our data suggested that >1500 dogs may be required to capture additional loci that contribute to the ACLR phenotype.

In conclusion, detailed analysis of the genetic risk factors involved in the initiation and progression of ACLR will provide a clear understanding of the genetic factors that cause disturbance to biological networks sufficient to lead to ACLR. Our data suggest that genetic risk of ACLR is influenced by multiple genomic loci with small individual additive effects. In our association study, we clearly show that ACLR is a highly polygenic trait. Our results suggest that biological networks that control innate immune mechanisms, aggrecan (*ACAN*) signaling, cellular proliferation, membrane transport, and/or neuronal signaling pathways should be further investigated. The genetic loci we have identified in this study will guide further dissection of genomic variants associated with the ACLR phenotype. Importantly, dogs have long been studied as genetic models of human disease. Here, we highlight how dogs are a genetically amenable model organism for studying orthopaedic complex trait disease and provide a model system that facilitates complex trait dissection. Ultimately, insights gained from this research may also lead to novel treatments and advances in complex trait genomic prediction. Such advances could have a large impact on human and animal health.

Supplementary Information

Table 4.1S | Anterior cruciate ligament rupture associated SNPs identified by GWAS in the Labrador Retriever, a dog breed with a high disease prevalence

SNP	chr	Position	P	LMM	Risk Allele	f(A)	f(U)	OR	Region start-end	Genes
BICF2G630709791	1	10788643	1.64E-04	1	C	0.30	0.16	2.18	10788643-11025688	<i>DNAJA1, RTTN, CD226, DOK6</i>
BICF2S23147946	1	17290917	1.74E-04	1	G	0.28	0.15	2.19	No LD	<i>BCL2, PHLPP1, ZCCHC2, TNFRSF11A, KIAA1468, PPIAP1, PIGN</i>
BICF2P181859	1	17840093	4.32E-04	1	A	0.21	0.09	2.60	No LD	<i>ZCCHC2, TNFRSF11A, KIAA1468, PIGN, RNF152, CDH20</i>
BICF2G630712921	1	19148000	4.94E-04	1	G	0.46	0.32	1.87	18645187-19148000	<i>CDH20, MC4R, PMAIP1, CCBE1</i>
BICF2G630713147	1	19274346	4.93E-04	1	A	0.37	0.24	1.83	19253280-19274346	<i>MC4R, PMAIP1, CCBE1</i>
BICF2P818099	1	39021948	4.29E-04	3	G	0.65	0.54	1.55	No LD	<i>PLAGL1, SF3B5, STX11, UTRN</i>
BICF2S23638642	1	45229667	4.95E-04	2	G	0.18	0.07	3.07	No LD	<i>AKAP12, ZBTB2, C6orf211, RMND1, CCDC170, ESR1, SYNE1</i>
BICF2P206910	1	46405864	3.25E-04	2	C	0.14	0.04	4.05	46405864-46443183	<i>MYCT1, VIP, FBX05, MTRF1L, RGS17</i>
BICF2S22959529	1	46443183	1.58E-04	1,2	C	0.14	0.03	4.78		
BICF2P1054044	2	20002181	2.28E-04	3	G	0.85	0.77	1.74	No LD	<i>MTPAP, MAP3K8, BAMBI</i>
BICF2S23533020	2	20501149	2.47E-04	3	G	0.92	0.84	2.33	20288541-20501149	<i>BAMBI, MPP7</i>
BICF2P720951	2	62500174	4.61E-04	3	A	0.26	0.16	1.82	62252040-62500174	<i>GPR114, CCDC102A, DOK4, COQ9, CIAPIN1, CCL17, CX3CL1, CCL22, PLLP,</i>

										ARL2BP, RSPRY1, CPNE2, NLRC5, HERPUD1, SLC12A3, NUP93, MT1, MT2, MT3, MT4, BBS2, OGFOD1, AMFR, GNAO1
TIGRP2P31530	2	80046637	1.65E-04	2,3	G	0.67	0.55	1.69	80046637-80079758	C1QB, C1QC, C1QA, EPHA8, ZBTB40, WNT4, CDC42, HSPG2, LDLRAD2, USP48, ALP, ECE1
BICF2P247448	2	80068322	1.48E-04	2,3	G	0.68	0.55	1.67		
BICF2P1066899	2	80079758	1.65E-04	2,3	A	0.67	0.55	1.69		
BICF2G630464857	2	87862734	2.00E-04	2	G	0.14	0.04	3.69	No LD	C1orf167, AGTRAP, DRAXIN, MAD2L2, FBXO6, FBXO44, FBXO2, PTCHD2, UBIAD1, MTOR, ANGPTL7, EXOSC10, SRM, MASP2, TARDBP, CASZ1, APITD1, DFFA, PGD, NDN, MAGEL2, MKRN3
BICF2G630108404	3	39424250	4.96E-04	3	G	0.24	0.13	2.06	39424250-39517119	NDN, MSGEL2, MKRN3
BICF2S23148483	3	39469011	4.96E-04	3	G	0.24	0.13	2.06		
BICF2P866702	3	39517119	4.96E-04	3	A	0.24	0.13	2.06		
BICF2P1109077	3	48873558	4.26E-04	3	G	0.80	0.67	1.90	48873558-48880579	MCTP2
BICF2P1431921	3	48878860	4.26E-04	3	G	0.80	0.67	1.90		
BICF2P241884	3	48880579	4.26E-04	3	G	0.80	0.67	1.90		
BICF2P564273	3	55250188	1.07E-04	1,2,3	A	0.70	0.52	2.16	No LD	ACAN, HAPLN3, MFGE8, ABHD2, RLBP1, FANCI, POLG, RHCG, TICRR, C3H15orf38, KIF7,

										<i>PLIN1, PEX11A, WDR93, AMPN, C3H15orf38</i>
TIGRP2P46522	3	57698908	4.01E-04	3	C	0.70	0.54	1.91	57660110-57698908	<i>PDE8A, RS17, CPEB1, AP3B2, FSD2, WHAMM, HOMER2, FAM103A1, BTBD1, TM6SF1, BNC1, SH3GL3</i>
BICF2G630349775	3	77625147	4.26E-04	1,2	A	0.18	0.07	2.90	No LD	<i>TBC1D1, PGM2, RELL1, C4orf19, KIAA1239</i>
BICF2S2342150	3	86948527	2.91E-04	1,2	A	0.40	0.27	1.83	86948527-86974042	<i>STIM2, TBC1D19, CCKAR, RBPJ, SEL1L3</i>
BICF2S24415473	3	86974042	7.07E-05	1,2	G	0.40	0.26	1.97		
BICF2G630359517	3	91944314	3.77E-04	1,2	C	0.51	0.36	1.82	No LD	<i>KCNIP4, PACRGL, SLIT2</i>
BICF2G63058646	4	9120668	1.75E-04	1,2,3	G	0.13	0.04	3.71	No LD	<i>SLC35F3, KCNK1, PCNXL2</i>
BICF2P295392	4	14536870	4.61E-04	2	G	0.11	0.04	3.07	No LD	<i>BICC1, PHYHIPL, FAM13C</i>
BICF2G630168473	4	74924050	3.31E-04	3	G	0.30	0.18	1.96	No LD	<i>WDR70, NUP155, C4H5orf42, NIPBL, Q9N280, RANBP3L</i>
BICF2G630175389	4	84260906	5.87E-05	1,2	A	0.83	0.68	2.28	No LD	<i>CDH10</i>
BICF2G630810143	6	11130832	9.46E-05	3	A	0.44	0.32	1.72	11130832-11177149	<i>UPK3B, LRWD1, ALKBH4, ORAI2, PRKRIP1, SH2B2, CUX1, MYL10, RABL5, FIS1, ZNHIT1, PLOD3</i>
BICF2G630810159	6	11177149	9.46E-05	3	A	0.44	0.32	1.72		
BICF2G630810173	6	11181920	1.33E-04	3	G	0.61	0.51	1.50	11035074-11181920	<i>YWHAG, SRCRB4D, ZP3, UPK3B, LRWD1, ALKBH4, ORAI2, PRKRIP1, SH2B2, CUX1, MYL10, RABL5, FIS1,</i>

										<i>ZNHIT1, PLOD3, MOGAT3</i>
BICF2P170661	6	11439931	3.39E-04	3	A	0.31	0.22	1.57	No LD	<i>CUX1, MYL10, RABL5, FIS1, ZNHIT1, PLOD3, MOGAT3, AP1S1, SERPINE1, MUC3A, ACHE, SLC12A9, ACHE, SRRT, EPHB4</i>
BICF2P1354767	6	11462695	3.02E-04	3	G	0.48	0.38	1.53	11106977-11462695	<i>UPK3B, LRWD1, ORAI2, PRKRIP1, SH2B2, CUX1, MYL10, RABL5, FIS1, ZNHIT1, PLOD3, MOGAT3, Q7YSA1, MUC3A, ACHE, SRRT, SLC12A9, EPHB4</i>
BICF2P205255	6	11484772	4.75E-04	3	G	0.54	0.45	1.43	No LD	<i>CUX1, MYL10, RABL5, FIS1, ZNHIT1, PLOD3, MOGAT3, AP1S1, SERPINE1, MUC3A, ACHE, SLC12A9, ACHE, SRRT, EPHB4</i>
BICF2P1358119	6	13131171	3.74E-04	3	C	0.44	0.34	1.54	13131171-13182379	<i>AZGP1, GJC3, TRIM4, OR2AE1, CYP3A12, ZNF498, ZNF655, FAM200A, ZKSCAN5, ZNF394, ZNF789, ATP5J2, CPSF4, PDAP1, ARPC1B, KPNA7, SMURF1, TRRAP, TMEM130</i>
BICF2S23324965	6	14077648	3.36E-05	3	G	0.68	0.60	1.42	14077648-14092057	<i>TRRAP, TMEM130, NPTX2, BAIAP2L1, BRI3, TECPR1,</i>
BICF2S22961650	6	14092057	1.53E-04	3	G	0.68	0.61	1.35		

										<i>LMTK2, PMS2, EIF2AK1, ANKRD61, USP42, CYTH3</i>
BICF2P498515	6	75848537	7.89E-05	1,2,3	A	0.16	0.06	3.11	No LD	<i>LRRIQ3</i>
BICF2P1072682	7	53407178	4.09E-04	1,2,3	C	0.28	0.14	2.42	No LD	<i>None</i>
BICF2P1090079	7	64389761	1.56E-04	1,3	C	0.47	0.31	1.94	No LD	<i>CDH2, CHST9</i>
BICF2P1208798	9	12671217	5.49E-05	1,2	G	0.56	0.36	2.27	No LD	<i>EFCAB13, ITGB3, MYL4, CDC27, KANSL1, MAPT, SPPL2C, CRHR1, NSF, WNT3</i>
BICF2P890246	9	53427907	3.23E-05	1,2	A	0.16	0.36	2.99	53427907-53432248	<i>SOHLH1, ALL2, LCN9, GLT6D1, LCN1, SURF1, SURF2, SURF4, MED22, C9H9orf96, REXO4, ADAMTS13, CACFD1, SLC2A6, ADAMTSL2, DOPO, SARDH, VAV2, WDR5, RXRA</i>
BICF2P139678	9	53432248	1.75E-04	1,2	A	0.83	0.65	2.71		
BICF2S23113199	10	46246942	3.57E-04	1,3	A	0.63	0.46	1.95	No LD	<i>AFF3, REV1, EIF5B, TXNDC9, LYG1, LYG2, MRPL30, MITD1, LIPT1, TSGA10</i>
BICF2P401973	10	65344772	3.79E-04	1	G	0.84	0.71	2.23	No LD	<i>FAM161A, CCT4, COMMD1, B3GNT2, TMEM17, EHBP1</i>
BICF2P454456	11	32175491	4.77E-04	3	A	0.19	0.13	1.62	31831896-32175491	<i>C11H9orf123, PTPRD</i>
BICF2P50610	11	32270617	2.75E-05	3	A	0.29	0.19	1.70	31939564-32270617	<i>C11H9orf123, PTPRD</i>
BICF2P531097	11	32908558	3.58E-04	3	A	0.44	0.29	1.92	32908558-32922914	<i>PTPRD</i>
BICF2P1290820	11	32922914	4.19E-04	3	G	0.44	0.30	1.91		

BICF2P65003	12	40691540	3.15E-04	2	G	0.71	0.61	1.61	40691540-41066621	SENP6, MYO6, IMPG1
BICF2G630606359	13	13352804	4.61E-04	2	G	0.69	0.57	1.67	13352804-13503950	NUDCD1, PKHD1L1, EBAG9, SYBU, KCNV1
BICF2S23620879	14	10265645	4.05E-04	3	C	0.5	0.37	1.71	9796003-10265645	COPG2, MEST, CEP41, CPA1, CPA4, SSMEM1, TMEM209, KLHDC10, UBE2H, NRF1, FAM40B, AHCYL2, SMO, TSPAN33, TNPO3, IRF5, KCP, ATP6V1F, FLNC
BICF2G630519882	14	11686985	4.99E-04	3	A	0.49	0.36	1.74	11675474-11686985	SND1, LRRC4, PAX4, ARF5, GCC1, ZNF800, GRM8
BICF2P594418	15	58424953	4.48E-04	2	A	0.19	0.10	2.07	No LD	FAM198B, TMEM144, RXFP1, ETFDH, PPID, C4orf46, FNIP2
BICF2G630422966	15	58852255	4.41E-04	3	A	0.41	0.28	1.80	58852255-58978372	FAM198B, TMEM144, RXFP1, ETFDH, PPID, FNIP2, C4orf46, C4orf45, RAPGEF2
BICF2G630422956	15	58891376	4.41E-04	3	A	0.41	0.28	1.80		
BICF2G630422900	15	58967776	2.04E-04	3	G	0.37	0.24	1.87		
BICF2G630422895	15	58978372	1.69E-04	3	A	0.37	0.24	1.89		
BICF2P880005	17	20749191	2.22E-04	1,2	G	0.44	0.31	1.75	No LD	KLHL29, ATAD2B
BICF2P1121006	18	54279578	1.11E-04	1,2,3	A	0.63	0.42	2.28	No LD	CCS, CCDC87, CTSF, ZDHHC24, PELI3, MRPL11, SLC29A2, B3GNT1, BRMS1, RIN1, CD248, TMEM151A, YIF1A, CNIH2, KLC2, PACS1, SF3B2, GAL3ST3, CATSPER1,

										<i>BANF1, EIF1AD, SART1, EFEMP2, MUS81, FIBP, FOSL1, CCDC85B, CTSW, SNX32, OVOL1, AP5B1, RNASEH2C, KAT5, A1XFH4, SIPA1, PCNXL3, MAP3K11, EHBP1L1, LTBP3, SCYL1</i>
BICF2P888055	20	13815084	3.25E-04	3	A	0.73	0.62	1.71	No LD	<i>GRM7</i>
BICF2P582174	20	14124824	3.76E-04	3	C	0.82	0.73	1.61	14118014-14124824	<i>GRM7</i>
TIGRP2P270462	20	15036973	8.51E-05	3	G	0.85	0.75	1.88	14838270-15053718	<i>EDEM1, ARL8B</i>
BICF2P716829	20	15048191	9.85E-05	3	G	0.85	0.75	1.86		
BICF2P1462185	20	15053718	4.90E-05	3	A	0.85	0.74	1.90		
BICF2P178583	20	30190042	1.41E-04	1,2	G	0.48	0.31	2.09	30039696-30190042	<i>ADAMTS9, PRICKLE2, PSMD6, ATXN7, THOC7, SNTN, C3orf49, SYNPR</i>
BICF2S2328420	20	51150968	2.77E-04	3	G	0.75	0.60	1.96	No LD	<i>ADGRE3, ADGRE2, ZNF333, CLEC17A, NDUFB7, TECR, DNAJB1, GIPC1, PTGER1, PKN1, DDX39A, ADGRE5, PRKACA, C19orf67, PALM3, IL27RA, RFX1, DCAF15, PODNL1, CC2D1A, C20H19orf57, NANOS3, ZSWIM4, MRI1</i>

BICF2P420488	20	52326317	3.84E-04	3	G	0.85	0.75	1.84	51873051-52326317	C19orf57, PRKACA, PALM3, IL27RA, RFX1, DCAF15, PODNL1, CC2D1A, C20H19orf57, NANOS3, ZSWIM4, MRI1, CCDC130, CACNA1A, NACC1, IER2, STX10, TRMT1, LYL1, NFIX, DAND5, RAD23A, CALR, SYCE2, RTBDN, KLF1, DNASE2, GCDH, RNASEH2A, PRDX2, JUNB, HOOK2, BEST2, TNPO2, FBXW9, DHPS, WDR83, WDR83OS, MAN2B1, ZNF791, ACP5, ELOF1, ACP5
TIGRP2P277002	20	55563965	2.25E-04	1,2	A	0.25	0.10	2.98	No LD	INSR, ARHGEF18, PEX11G, ZNF358, MCOLN1, CAMSAP3, XAB2, STXBP2, RETN, TRAPPC5, FCER2, CLEC4G, CD209, EVI5L, LRRC8E, MAP2K7, SNAPC2, CCL25, TIMM44, ELAVL1, FBN3, CERS4, CD320, KANK3, ANGPTL4, RAB11B, MARCH2, HNRNPM, PRAM1, ZNF414, MYO1F
BICF2P111342	21	7150110	1.25E-04	1,2	A	0.31	0.18	2.04	No LD	None

BICF2G630658881	21	7582214	1.09E-04	1,2,3	G	0.49	0.32	2.12	7582214-8382709	JRKL, CCDC82, MAML2, MTMR2, CEP57, FAM76B, SESN3
BICF2G630658668	21	8205285	4.03E-04	3	G	0.49	0.32	2.00		
BICF2G630658620	21	8382709	2.90E-04	3	G	0.50	0.33	2.02		
BICF2G630658768	21	8033283	4.09E-04	1,2	C	0.27	0.14	2.28	8033283-8061623	JRKL, CCDC82, MAML2, MTMR2, CEP57, FAM76B
BICF2G630658756	21	8040746	4.73E-04	1	A	0.26	0.13	2.30		
BICF2G630658723	21	8061623	4.09E-04	1,2	G	0.27	0.14	2.28		
BICF2S2442023	21	43533585	3.19E-04	3	G	0.50	0.34	1.93	43507320-43533585	NUCB2, NCR3LG1, KCNJ11, ABCC8, USH1C, OTOG, MYOD1, KCNC1, U6, TPH1, SERGEF, SAAL1, SAA1, HPS5, GTF2H1
BICF2S2361376	21	43752575	1.76E-04	1,2,3	A	0.60	0.42	1.97	43752575-43808389	USH1C, OTOG, MYOD1, KCNC1, U6, B6EY10, SAAL1, SAA1, HPS5, GTFSH1, LDHC, TSG101, UEVLD, SPTY2D1, TMEM86A, PTPN5
BICF2P321064	21	44627903	1.66E-04	1,2,3	G	0.38	0.24	1.99	No LD	TSG101, UEVLD, SPTY2D1, TMEM86A, PTPN5, ZDHHC13, CSRP3, E2F8, NAV2
TIGRP2P293361	22	42354230	2.27E-04	2	A	0.49	0.36	1.74	No LD	SLITRK5
TIGRP2P297337	22	58201452	1.08E-04	1,2,3	A	0.44	0.27	2.20	No LD	EFNB2, ARGLU1
BICF2G630375268	23	33376383	3.48E-04	3	A	0.74	0.63	1.75	No LD	TMEM108, BFSP2, CDV3, TOPBP1, TF, RAB6B, SLCO2A1
BICF2S23730962	23	53809871	2.93E-04	2	A	0.87	0.78	1.88	No LD	TIPARP, LEKR1, CCNL1, VEPH1, PTX3

BICF2G630502225	24	23992936	4.86E-04	2	G	0.92	0.81	2.84	No LD	ZCCHC3, C20orf96, DEFB132, DEFB128, B0FF14, DEFB126, DEFB125, DEFB118, DEFB116, DEFB121, DB119, DEFB122, B0FF10, BCL-XL, HM13CO4I2, TPX2, MYLK2, DUSP15, TTLL9, PDRG1, XKR7, HCK, TM9SF, C20orf160
BICF2G630500835	24	29648925	1.29E-04	2,3	A	0.79	0.67	1.90	No LD	CTNNBL1, VSTM2L, TTI1, RPRD1B, TGM2, KIAA1755, BPI, LBP, RALGAPB, ADIG, SLC32A1, ACTR5
BICF2P544126	24	29772193	4.09E-05	3	G	0.94	0.87	2.28	29772193-29794411	CTNNBL1, VSTM2L, TTI1, RPRD1B, TGM2, KIAA1755, BPI, LBP, RALGAPB, ADIG, SLC32A1, ACTR5, PPP1R16B, FAM83D
BICF2S24111418	24	29794411	4.09E-05	3	A	0.94	0.87	2.28		
BICF2G630500368	24	30241088	2.76E-07	1,2,3	G	0.83	0.66	2.56	30241088-30245795	BPI, LBP, RALGAPB, SLC32A1, ADIG, ACTR5, PPP1R16B, FAM83D, DHX35
BICF2G630500363	24	30245795	1.82E-06	1,2,3	G	0.80	0.62	2.44		
BICF2G630799191	26	22848912	1.33E-04	3	G	0.61	0.45	1.92	No LD	ADRBK2, MYO18B, SEZ6L, ASPHD2, HPS4, SRRD, TFP11, TPST2, CRYBB1, CRBA4

BICF2P792911	26	22894961	8.55E-05	1,2,3	G	0.44	0.27	2.14	No LD	ADRBK2, MYO18B, SEZ6L, ASPHD2, HPS4, SRRD, TFP11, TPST2, CRYBB1, CRBA4
BICF2S2356299	27	30557856	2.21E-05	2,3	A	0.43	0.27	2.03	No LD	AEBP2, PLEKHA5
BICF2P1332722	27	30603252	2.19E-04	1,2	G	0.78	0.60	2.33	No LD	AEBP2, PLEKHA5
BICF2P1047447	27	31108106	2.80E-04	3	A	0.52	0.38	1.75	No LD	AEBP2, PLEKHA5, CAPZA3, PLCZ1, PIK3C2G
BICF2P487060	27	33778510	4.55E-04	3	C	0.40	0.27	1.81	33778510-33809600	MGST1, SLC15A5, DERA, STRAP, EPS8, PTPRO
BICF2P599881	27	35600038	2.66E-04	3	G	0.85	0.74	1.97	No LD	PLBD1, ATF7IP, NMDE2
BICF2P1410038	27	37697040	3.33E-04	3	C	0.70	0.56	1.82	No LD	BCL2L14, ETV6, TAS2R42, CAFA-T2R67, CAFA-T2R43, CAFA-T2R12, TAS2R10, TAS2R7, CSDA, STYK1, MAGOHB, LY49
BICF2S23535135	27	37814333	1.13E-04	3	A	0.31	0.20	1.83	No LD	ETV6, TAS2R42, CAFA-T2R67, CAFA-T2R43, CAFA-T2R12, TAS2R10, TAS2R7, CSDA, STYK1, MAGOHB, LY49
TIGRP2P355298	27	39134291	1.31E-04	3	G	0.74	0.57	2.16	No LD	KLRK1, KLRD1, GABARAPL1, TMEM52B, OLR1, CLEC7A, CLEC1B, CLEC9A, CLEC12B, CLEC12A, CLEC2B, KLRF1, CD69, CLEC2D, KLRB1
BICF2S23255928	27	39211186	1.10E-04	2,3	A	0.23	0.13	2.07	39211186-39217437	

BICF2P526639	27	39217437	4.12E-05	2,3	G	0.23	0.12	2.18		KLRD1, GABARAPL1, TMEM52B, OLR1, CLEC7A, CLEC1B, CLEC12B, CLEC12A, CLEC2B, KLRF1, CD69, CLEC2D, KLRB1
BICF2S23152419	27	39428263	1.77E-05	3	A	0.79	0.64	2.13	39428263-39445306	CLEC1B, CLEC12A, CLEC2B, KLRF1, CD69, CLEC2D, KLRB1, A2M, KLRG1, M6PR, PHC1, RIMKLB
BICF2P491441	27	39434491	3.34E-04	3	G	0.78	0.63	2.06		
TIGRP2P355396	27	39445306	3.34E-04	3	A	0.78	0.63	2.06		
BICF2P337576	27	39463031	2.35E-04	3	G	0.70	0.55	1.92	39346073-39511019	OLR1, CLEC7A, CLEC1B, CLEC12B, CLEC12A, CLEC2B, KLRF1, CD69, CLEC2D, KLRB1, A2M, KLRG1, M6PR, PHC1, RIMKLB
BICF2P794117	27	39511019	3.76E-04	3	C	0.71	0.56	2.00		
BICF2P155064	27	39526004	4.68E-04	3	G	0.21	0.14	4.78	No LD	CLEC12A, CLEC2B, KLRF1, CD69, CLEC2D, KLRB1, A2M, KLRG1, M6PR, PHC1, RIMKLB
BICF2P992747	27	39580957	1.65E-04	3	A	0.77	0.62	2.07	No LD	CLEC2B, KLRF1, CD69, CLEC2D, KLRB1, A2M, KLRG1, M6PR, PHC1, RIMKLB, MFAP5
BICF2S23652189	27	39644847	3.99E-04	3	G	0.31	0.21	1.64	39606871-39644847	AICDA, APOBEC1, DPPA3
TIGRP2P362234	28	41376035	3.71E-04	2	G	0.87	0.74	2.37	41376035-41377128	MGMT, EBF3, GLRX3
BICF2S23346408	28	41377128	1.25E-04	1,2	A	0.87	0.73	2.53		
BICF2S23713161	29	20562935	4.59E-04	2	G	0.75	0.63	1.74	No LD	PREX2, C29H8orf34

BICF2S23410873	29	20672864	2.18E-04	2,3	C	0.82	0.66	2.35	20672864-20703202	PREX2, C29H8orf34
BICF2P1135545	29	20703202	2.81E-04	2,3	A	0.82	0.67	2.31		
BICF2P483191	29	21601273	2.31E-05	1,2,3	C	0.73	0.51	2.54	No LD	C29H8orf34, SULF1, SLCO5A1
BICF2P139173	29	22067666	3.59E-04	2	A	0.50	0.36	1.79	22050835-22191229	SULF1, SLCO5A1, VTI1B, PROM14, NCOA2, TRAM1
BICF2P361907	29	22191229	3.47E-04	2	G	0.50	0.36	1.75		
BICF2P456086	29	23130206	4.85E-04	3	A	0.88	0.71	3.04	No LD	TRAM1, LACTB2, XKR9, EYA1 TRAM1, LACTB2, XKR9, EYA1
BICF2P662502	29	26040013	3.24E-04	1,2	G	0.90	0.73	2.87	No LD	JPH1, GDAP1, PI15, CRISPLD1
BICF2G630412697	30	3126573	7.22E-05	1,3	G	0.96	0.86	4.23	No LD	OR4K2, OR4K1, ORFN5
BICF2S2356993	31	12920807	2.25E-04	2,3	A	0.25	0.14	2.04	No LD	ROBO2
BICF2P287265	31	30555902	4.38E-04	3	G	0.96	0.87	3.38	No LD	MIS18A, MRAP, URB1, FAM176C, C31H21orf59, SYNJ1, GCFC1, OLIG2, OLIG1, SC5A3, SUB1, IL10RB
BICF2S23054250	35	26868731	2.20E-04	2,3	A	0.49	0.33	1.97	No LD	LRRC16A, SCGN, HIST1H2AA, HIST1H2BA, SLC17A4, SLC17A1, SLC17A2, TRIM38, HFE, HIST1H4E, HIST1H1C, HIST1H1A, HIST1H3C, HIST1H3A, HIST1H2AE, HIST1H2BB, HIST1H1T, HIST1H2BC, HIST1H2AC,

										<i>HIST4H4, HIST2H2AB, HIST1H4F, HIST1H2BG, HIST1H4L, BTN1A1, HMGN4</i>
BICF2P1086740	37	26916351	4.34E-04	2	G	0.41	0.24	2.18	No LD	<i>SMARCAL1, IGFBP2, IGFBP5, TNP1</i>
BICF2P708698	37	26924473	1.53E-04	2	A	0.63	0.45	2.06	No LD	<i>SMARCAL1, IGFBP2, IGFBP5, TNP1</i>

Note: OR odds ratio calculated from PLINK. LMM Linear mixed model 1 – GCTA, 2 – GEMMA, 3 – PUMA. f(A) and f(U) represent the frequency of the risk allele in case and control dogs, respectively. SNP position and genomic regions are based on CanFam 2.0. Gene list was identified using UCSC Genome Browser after LiftOver to CanFam 3.1 and the addition of 500kb flanking regions.

Table 4.2S Statistical Power and odds ratio correction for anterior cruciate ligament rupture GWAS risk loci identified by GEMMA in the Labrador Retriever

SNP	Chr	MAF	Beta	Significance	OR	f(u)	Power	Corrected OR	Corrected Power	INPower Estimated number of loci
BICF2S23638642	1	0.12	0.23	4.95E-04	3.07	0.07	0.99	1.38	0.19	6.1
BICF2S22959529	1	0.08	0.31	1.58E-04	4.78	0.03	0.99	1.83	0.30	3.9
BICF2P247448	2	0.40	-0.18	1.48E-04	1.67	0.55	0.84	1.23	0.22	2.6
BICF2G630464857	2	0.08	0.30	2.00E-04	3.69	0.04	0.98	1.58	0.23	3.7
BICF2P564273	3	0.41	-0.17	1.07E-04	2.16	0.52	0.99	1.47	0.61	3.5
BICF2G630349775	3	0.12	0.24	4.26E-04	2.90	0.07	0.98	1.37	0.18	5.5
BICF2S24415473	3	0.32	0.19	7.07E-05	1.97	0.26	0.96	1.50	0.59	2.0
BICF2G630359517	3	0.42	0.15	3.77E-04	1.82	0.36	0.94	1.20	0.18	6.2
BICF2G63058646	4	0.08	0.30	1.76E-04	3.71	0.04	0.98	1.62	0.25	4.1
BICF2P295392	4	0.07	0.29	4.61E-04	3.07	0.04	0.91	1.39	0.14	5.9
BICF2G630175389	4	0.26	-0.20	5.87E-05	2.28	0.68	0.98	1.69	0.77	2.3
BICF2P498515	6	0.10	0.29	7.89E-05	3.11	0.06	0.98	1.93	0.57	3.4
BICF2P1072682	7	0.20	0.20	4.09E-04	2.42	0.14	0.99	1.30	0.21	4.1
BICF2P1208798	9	0.44	0.21	5.49E-05	2.27	0.36	1.00	1.70	0.87	1.9
BICF2P890246	9	0.28	-0.21	3.23E-05	2.99	0.36	1.00	2.19	1.00	2.5
BICF2P65003	12	0.35	-0.20	3.15E-04	1.61	0.61	0.75	1.16	0.13	2.5
BICF2G630606359	13	0.38	-0.16	4.61E-04	1.67	0.57	0.83	1.16	0.14	4.8
BICF2P594418	15	0.14	0.22	4.48E-04	2.07	0.10	0.84	1.24	0.13	4.6
BICF2P880005	17	0.37	0.17	2.22E-04	1.75	0.31	0.89	1.21	0.19	4.0
BICF2P1121006	18	0.49	-0.21	1.11E-04	2.28	0.42	1.00	1.50	0.66	1.6
BICF2P178583	20	0.38	0.18	1.41E-04	2.09	0.31	0.99	1.36	0.41	3.4
TIGRP2P277002	20	0.16	0.23	2.25E-04	2.98	0.10	1.00	1.45	0.30	3.0
BICF2P111342	21	0.06	0.20	1.25E-04	2.04	0.18	0.95	1.38	0.34	3.1
BICF2G630658881	21	0.39	0.19	1.09E-04	2.12	0.32	0.99	1.45	0.56	2.0
BICF2G630658768	21	0.19	0.21	4.09E-04	2.28	0.14	0.97	1.28	0.19	3.9
BICF2S2361376	21	0.49	0.17	1.76E-04	1.97	0.42	0.98	1.28	0.30	3.8
BICF2P321064	21	0.30	0.18	1.66E-04	1.99	0.24	0.96	1.29	0.27	3.7
TIGRP2P293361	22	0.42	0.17	2.27E-04	1.74	0.36	0.90	1.21	0.20	4.0
TIGRP2P297337	22	0.34	0.19	1.08E-04	2.20	0.27	0.99	1.48	0.57	2.3
BICF2S23730962	23	0.19	-0.22	2.93E-04	1.88	0.78	0.79	1.22	0.15	3.8

BICF2G630502225	24	0.14	-0.22	4.86E-04	2.84	0.81	0.98	1.35	0.25	4.4
BICF2G630500835	24	0.28	-0.17	1.29E-04	1.90	0.67	0.91	1.33	0.33	4.4
BICF2G630500368	24	0.27	-0.26	2.76E-07	2.56	0.66	1.00	2.44	1.00	0.8
BICF2P792911	26	0.34	0.19	8.55E-05	2.14	0.27	0.99	1.53	0.65	2.2
BICF2S2356299	27	0.34	0.17	2.21E-05	2.03	0.27	0.98	1.71	0.84	4.2
BICF2P1332722	27	0.32	-0.18	2.19E-04	2.33	0.60	1.00	1.34	0.38	3.8
BICF2P526639	27	0.17	0.23	4.12E-05	2.18	0.12	0.93	1.71	0.63	3.5
BICF2S23346408	28	0.21	-0.22	1.25E-04	2.53	0.73	0.99	1.53	0.55	2.2
BICF2S23713161	29	0.32	-0.16	4.59E-04	1.74	0.63	0.85	1.18	0.15	6.2
BICF2S23410873	29	0.27	-0.18	2.18E-04	2.35	0.66	0.99	1.34	0.35	4.6
BICF2P483191	29	0.39	-0.20	2.31E-05	2.54	0.51	1.00	2.02	0.98	2.3
BICF2P361907	29	0.42	0.18	3.47E-04	1.75	0.36	0.90	1.19	0.17	2.8
BICF2P662502	29	0.20	-0.19	3.24E-04	2.87	0.73	1.00	1.39	0.37	4.9
BICF2S2356993	31	0.19	0.20	2.25E-04	2.04	0.14	0.91	1.27	0.18	4.7
BICF2S23054250	35	0.39	0.16	2.20E-04	1.97	0.33	0.98	1.26	0.26	4.1
BICF2P1086740	37	0.31	0.17	4.34E-04	2.18	0.24	0.99	1.26	0.23	4.1
BICF2P708698	37	0.48	-0.16	1.53E-04	2.06	0.45	0.99	1.33	0.39	4.1

Note: MAF minor allele frequency, OR odds ratio, and f(u) frequency of risk allele in control dogs calculated from PLINK. Corrected OR was calculated using an approximate conditional likelihood approach. GWAS data were derived using GEMMA. For each risk locus detected by GEMMA, the total number of risk loci was estimated using INPower.

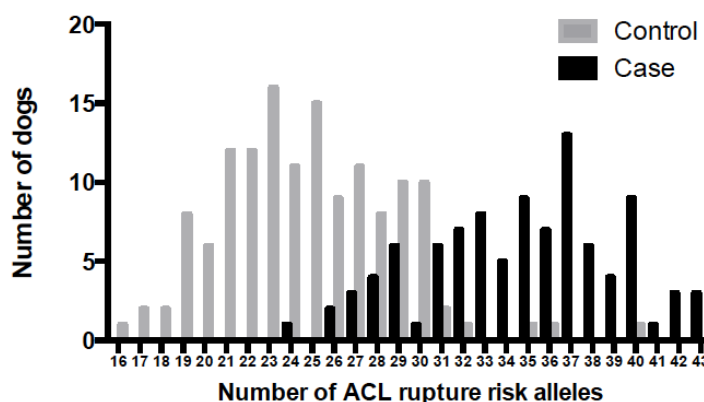
Table 4.3S | Estimation of the expected number of loci to be discovered in a future GWAS of ACLR in the Labrador Retriever using INPower

Sample size	Expected number of loci to be discovered	Probability of detecting at least the expected number of loci
250	0.0	0.00 (0 loci)
500	0.0	0.02 (1 locus)
750	0.1	0.06 (1 locus)
1,000	0.2	0.15 (1 locus)
1,250	0.3	0.29 (1 locus)
1,500	0.6	0.47 (1 locus)
1,750	1.1	0.66 (1 locus)
2,000	1.7	0.51 (2 loci)
2,250	2.6	0.25 (3 loci)
2,500	3.7	0.52 (4 loci)

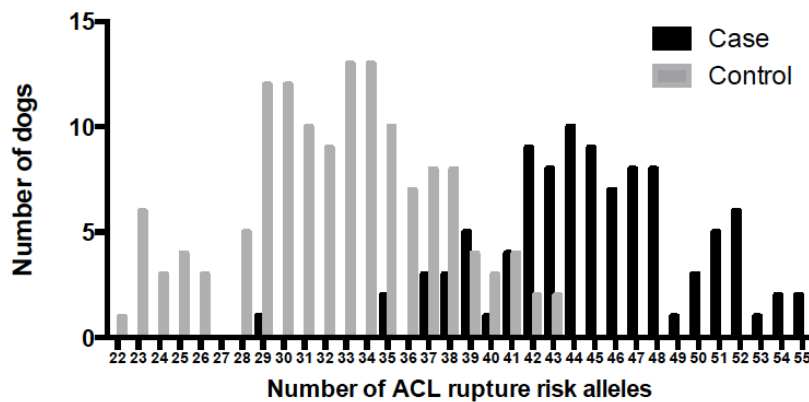
Note: Power calculations were performed with INPower.

Figure 4.1S. Genetic risk scoring using GWAS associated loci from linear mixed model analysis with GCTA, GEMMA, and PUMA segregates ACLR disease risk in case and control Labrador Retriever dogs. Distribution of the number of ACLR risk loci in case and control groups of Labrador Retriever dogs for GCTA (a), GEMMA (b), and PUMA (c). The number of risk alleles in cases and controls is significantly different ($P < 2.2E-16$). ACLR odds ratios of weighted genetic risk scores (wGRS) relative to the first quartile for GCTA (d), GEMMA (e), and PUMA (f). Vertical bars represent the 95% confidence intervals. *Odds ratio is significantly different from the reference first quartile.

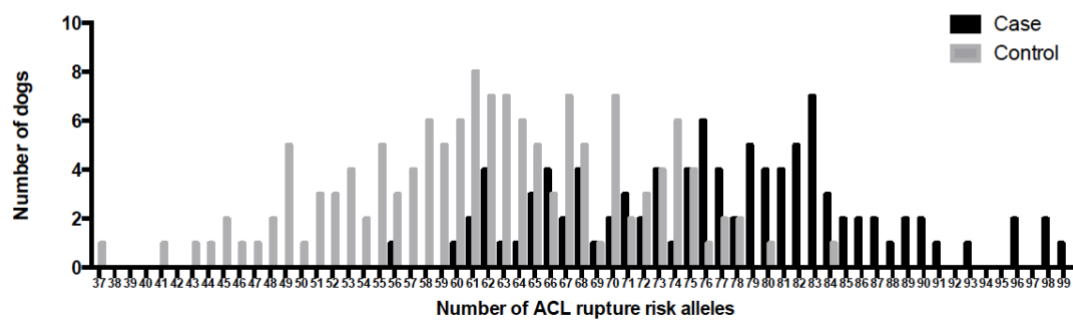
a)



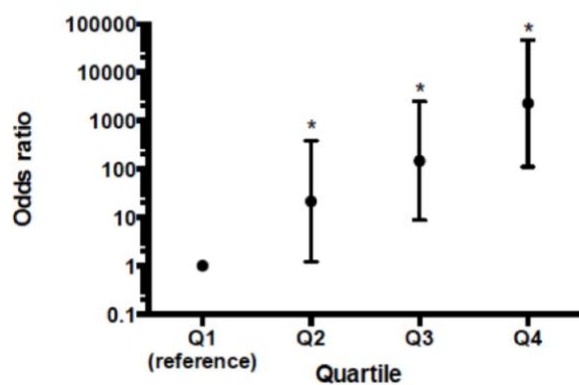
b)



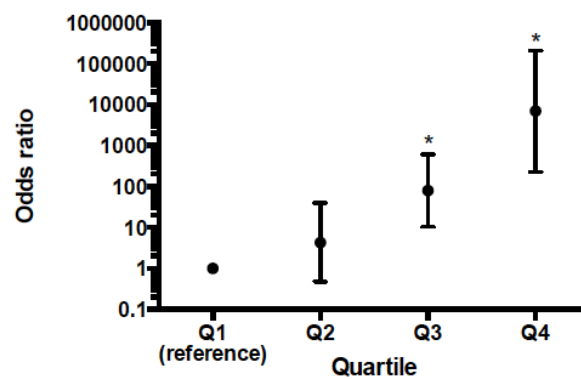
c)



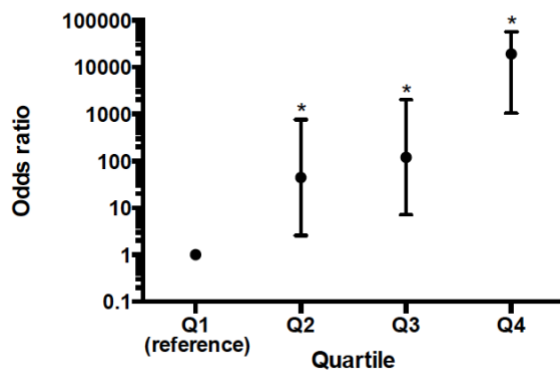
d)



e)



f)



Chapter 5

Multivariate Genome-Wide Association Analysis Identifies Novel and Relevant Variants Associated with Anterior Cruciate Ligament Rupture Risk in the Dog Model

Baker LA, Rosa GJ, Hao Z, Piazza A, Hoffman C, Binversie EE, Sample SJ, Muir P. Multivariate genome-wide association analysis identifies novel and relevant variants associated with anterior cruciate ligament rupture risk in the dog model. BMC Genetics. 2018;19:39.

Acknowledgments: This work was supported by grants from the Morris Animal Foundation (D13CA-020, www.morrisanimalfoundation.org), the Robert Draper Technology Innovation Fund from the University of Wisconsin-Madison and the Melita Grunow Family Professorship in Companion Animal Health. Lauren Baker received support from the National Institutes of Health (T32OD10999) and is supported by a National Library of Medicine training grant to the UW-Madison Computation and Informatics in Biology and Medicine Training Program (NLMT15LM007359). Susannah Sample received support from the National Institutes of Health (K01OD019743-01A1). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

The authors would like to thank the faculty, residents, and students throughout the University of Wisconsin-Madison UW Veterinary Care Hospital and the many individual breeders and pet owners for their help in recruitment of Labrador Retrievers for this study. The authors would also like to thank the American Kennel Club Canine Health Foundation, which provided support by giving access to the AKC pedigree database.

Abstract

Background

Anterior cruciate ligament rupture (ACLR) is a debilitating and potentially life-changing condition in humans, as there is a high prevalence of early-onset osteoarthritis after injury. Identification of high-risk individuals before they become patients is important, as post-treatment lifetime burden of ACLR in the USA ranges from \$7.6 to \$17.7 billion annually. ACLR is a complex disease with multiple risk factors including genetic predisposition. Spontaneous ACLR in the dog is an excellent model for human ACLR, as risk factors and disease characteristics in humans and dogs are similar. In a univariate genome-wide association study (GWAS) of 237 Labrador retrievers, we identified 99 ACLR candidate loci. It is likely that additional variants remain to be identified. Joint analysis of multiple correlated phenotypes is an underutilized technique that increases statistical power, even when only one phenotype is associated with the trait. Proximal tibial morphology has been shown to affect ACLR risk in both humans and dogs. In the present study, tibial plateau angle (TPA) and relative tibial tuberosity width (rTTW) were measured on bilateral radiographs from purebred Labrador Retrievers that were recruited to our initial GWAS. We performed a multivariate genome wide association analysis of ACLR status, TPA, and rTTW.

Results

Our analysis identified 3 loci with moderate evidence of association that were not previously associated with ACLR. A variant on Chr4 associated with both ACLR and TPA resides within *DOCK2*, a gene that has been shown to promote immune cell migration and invasion in synovitis, an important predictor of ACLR. A locus on Chr1 associated with both ACLR and rTTW is located within *ROR2*, a gene important for cartilage and bone development. A third locus on Chr23 associated with only ACLR is located near a long non-coding RNA (lncRNA). lncRNA's are important for regulation of gene transcription and translation.

Conclusions

These results did not overlap with our previous GWAS, for which different methods were used, and supports the need for further work. The results of the present study are highly relevant to ACLR pathogenesis, and identify potential drug targets for medical treatment.

Introduction

Anterior cruciate ligament rupture (ACLR) is common in human beings, particularly in young, athletic individuals [Sanders 2016]. ACLR is a debilitating injury with a long recovery period. There is a high prevalence of early-onset knee osteoarthritis after injury [Lohmander 2004]. Treatment for ACLR is costly; the lifetime burden of ACLR in the US is estimated at \$7.6 billion annually when treated with surgical reconstruction and \$17.7 billion annually if treated with rehabilitation [Mather 2013]. Nearly three quarters of these cases can be classified as non-contact ruptures, typically occurring during landing or pivoting movements [Smith 2012a]. A clear explanation for ACLR without physical contact is not yet available, though it is generally understood that non-contact ACLR is a complex disease caused by a combination of intrinsic (variables that describe the individual) and extrinsic (variables that describe the environment of the individual) risk factors [Smith 2012a, Smith 2012b].

Identification of intrinsic factors that increase individual risk of ACLR is an important epidemiologic goal. Thorough understanding of these risk factors will provide physicians with the means to identify individuals at high risk before they become patients in the orthopaedic surgeon's office. Young women are up to 10x more likely to rupture their ACL compared to men [Gwinn 2000]. Among these women, a higher than average body mass index is a significant risk factor [Uhorchak 2003; Shultz 2015]. Gonadal hormones may play a role in ACLR risk [Shultz 2015]. Certain characteristics of femur and tibia morphology have also been implicated in ACLR pathophysiology. The general hypothesis underlying these risk factors is that bone morphology can alter knee joint biomechanics, potentially placing excessive stress on the ACL. For instance, it has been suggested that a narrow femoral intercondylar notch may impinge on the ACL in certain knee positions [Smith 2012b], and increased posterior tibial slope may increase anterior-directed forces on the tibia leading to greater weight-bearing load on the ACL that in certain situations may exceed its failure strength [Smith 2012b, Hashemi 2008].

Genetics plays an important and deterministic role in directing the development of individual morphology and physiology. Therefore, it is perhaps unsurprising that familial analyses support the existence of genetic influence in ACLR risk. A close family history of ACLR doubles individual risk [Flynn 2005] and increases the risk of ACL graft rupture and contralateral ACLR [Harner 1994; Brophy 2012]. Candidate gene studies have reported variants in genes for collagens, proteoglycans, matrix metalloproteinases, angiogenesis-associated proteins, elastin, and fibrillin [Kaynak 2017]. A recent genome-wide association study (GWAS) in humans did not identify any variants that met genome-wide significance but had a small case sample size of 598 individuals of various ethnic backgrounds [Kim 2017]. Additionally, much of the existing research on ACLR has come from a single population, and further research is needed to confirm or deny reported associations [Kaynak 2017; John 2016]. Because ACLR is a complex disease, the genetic influence on ACLR is likely composed of many variants in multiple genomic loci with small to moderate individual effects [Robinson 2014]. Discovery of these smaller effect loci in human populations will require a combination of very large sample sizes, often in the hundreds of thousands, high-quality phenotyping, and sophisticated statistical methods [Robinson 2014].

One approach to identifying genetic variants that influence complex diseases is to study the trait in a model organism that may improve GWAS utility, such as the dog. For many diseases, presenting clinical signs, pathogenesis, and treatment are extremely similar between humans and their canine counterparts [Shearin 2010]. The advantage of the dog lies in its unique history and genomic architecture [Karlsson 2008]. Purebred dogs are closed populations descendent from a small number of founder individuals [Karlsson 2008]. Selective breeding for visual and behavioral characteristics has also inadvertently selected for heritable diseases [Karlsson 2008; Shearin 2010]. This selective process has created long segments of DNA in linkage disequilibrium (LD) that likely contain disease risk variants [Karlsson 2008]. This combined effect allows GWAS to be performed using fewer SNP markers and smaller sample

sizes than would be required to perform the same experiment in human populations.

Discoveries made in the dog can then be used to inform candidate gene studies in human populations, saving great effort in the way of time and research funding.

The canine knee joint is a long-established model for human knee pathology [Gregory 2012]. Though there are important differences between humans and quadrupedal mammals in weight-bearing, gait, and joint range of motion, studies have shown the relative dimensions of internal knee structures, including the cruciate ligaments, are similar between dogs and humans. In addition, the canine ACL has similar cell density, blood vessel density, and cell shape when compared to the human ACL [Proffen 2012]. ACLR is the most common cause of pelvic limb lameness in the dog [Comerford 2011]. Like human ACLR, the vast majority of ruptures do not involve contact injury and typically occur while the dog is running or playing. As in humans, multiple risk factors have been implicated, including influence of gonadal hormones [Duerr 2007; Witsberger 2008; Torres de la Riva 2013] and high body condition score [Lampman 2003; Adams 2011]. Anatomic risk factors that are associated with ACLR in dogs include narrow femoral intercondylar notch [Aiken 1995; Comerford 2006; Lewis 2008] distal femoral torsion [Ragety 2011], excessive tibial plateau angle (analogous to posterior tibial slope in human beings) [Morris 2001; Mostafa 2009; Su 2015; Janovec 2017] and a relatively narrow tibial tuberosity width [Inauen 2009]. The most important risk factor for disease initiation is genetic influence, i.e. breed. While many breeds may be affected with ACLR, Newfoundlands, Rottweilers, and Labrador Retrievers are at especially high risk [Whitehair 1993; Witsberger 2008]. Heritability of ACLR in dogs is moderate at 0.3-0.5 [Nielen 2001; Wilke 2006; Baker 2017]. Variants associated with ACLR in collagen genes have been reported [Baird 2014a]. A GWAS in the Newfoundland breed identified variants in genes associated with neuronal signaling pathways [Baird 2014b]. Through linear mixed model GWA analysis, the present authors identified 128 SNPs in 99 regions that were associated with ACLR in the Labrador Retriever [Baker 2017]. Gene set and pathway analysis identified enrichment for genes

associated with angiogenesis, innate immune mechanisms, and extracellular matrix proteins, particularly aggrecan, which has been linked to ACLR in humans [Mannion 2014; Johnson 2015].

Given the complex, polygenic nature of ACLR [Baker 2017; Baird 2014b], it is likely that additional variants remain to be identified [Robinson 2014]. InPower analysis [Park 2010] of our Labrador Retriever discovery GWAS suggested that at least 172 loci influence ACLR risk in the dog [Baker 2017]. While improved statistical power is best achieved through larger sample sizes, joint analysis of multiple correlated phenotypes is an underutilized technique that has the potential to increase statistical power to detect the moderate and small effect associations expected with complex traits without the logistical concerns associated with increasing sample size [Robinson 2014]. To date, published GWA analyses in the dog model have all been univariate, i.e. they consider each phenotype independently. Here we present a multivariate genetic association analysis of anatomic variables, tibial plateau angle (TPA) and relative tibial tuberosity width (rTTW), with ACLR case-control status in the dog model. We identified three loci with moderate association, two of which reside in genes that have not been previously linked to ACLR but have an established role in chronic immune-mediated conditions of human beings, including rheumatoid arthritis and progression of osteoarthritis. Both of these genes have been identified as drug targets for other disorders and may be candidates for ACLR treatment.

Methods

Recruitment

Recruitment and quality control have been reported previously [Baker 2017]. Purebred Labrador Retriever dogs were recruited through the University of Wisconsin-Madison UW Veterinary Care Hospital, online advertising, and contact with local and national breed clubs. If available, a pedigree was collected from each dog enrolled in the study to confirm purebred

status. The Labrador Retriever is well-suited to GWAS because it is relatively outbred compared to some rare breeds with smaller population size and has an average LD decay over distance of 785 kb in sequencing data [Gray 2009]. This allows for a better mapping resolution than the many breeds with LD decay over distance >1-2 Mb. Full siblings were excluded from the analysis to avoid bias due to over-represented genotypes in closely related individuals [Anderson 2010]. Control dogs were over 8 years of age [Reif 2004], had stable stifles on examination, and no signs of effusion or osteophytosis on lateral radiographs [Chuang 2014]. Affected dogs were of any age, with examination and radiographic signs consistent with ACLR [Harasen 2002].

Genotyping

Blood or saliva was collected from case and control dogs for DNA extraction. Dogs were genotyped using the Illumina Canine HD BeadChip, which contains approximately 220k single nucleotide polymorphism (SNP) markers on the CanFam3.1 reference genome. Quality control was performed using PLINK [Chang 2015] and previously established protocol [Baker 2017]. Briefly, SNPs were removed for minor allele frequency <0.01, missingness >90%, and Hardy-Weinberg equilibrium of $P < 1 \times 10^{-7}$. All individuals had a genotyping call rate of >95%.

Radiographic Knee Morphology

Bilateral knee radiographs were evaluated to confirm that the limb was not appreciably rotated before measurement. The position of the proximal tibia relative to the proximal fibula was evaluated to assess the degree of tibial rotation along the long axis of the bone. Radiographs with substantial rotational malpositioning were removed from analysis. TPA and rTTW were measured using Horos DICOM image viewing and measurement tools

(<http://www.horosproject.org>). These traits are stable in skeletally mature animals.

Measurements were made according to published techniques [Abel 2003; Inauen 2009] (Figure 5.1). If bilateral radiographs were available and of adequate quality, TPA and rTTW were measured bilaterally and the final value was the average of the two measurements.



Figure 5.1. Measurements of proximal tibial morphology taken from a lateral stifle radiograph. Before measurement, the radiograph was evaluated for appropriate technique and position. A. The most anterior point of the tibial tuberosity. B. The cross point of a circle on line AC where the center is at the most caudal point of the tibial plateau (C) and also crosses the most anterior point of the tibial plateau (D). C. The most posterior point of the tibial plateau. D. The most anterior point of the tibial plateau. E. A reference line extending from the center of the intercondylar eminence proximally to the center of the tarsus distally. F. The cross point of line E and the tibial slope line (DC). Tibial plateau angle (TPA) was measured as the obtuse angle at $EFC - 90^\circ$. Relative tibial tuberosity width (rTTW) was measured as the width of the tibial tuberosity divided by the width of the proximal tibia not including the tibial tuberosity (AB/BC).

Statistical analysis

To avoid losing dogs due to missing information, dogs with unknown weight ($n=7$ cases, 30 controls) or age ($n=5$ cases, 4 controls) were assigned the average weight or age of their case or control group before radiographic evaluation. For cases, this weight was 37.05

kilograms and age was 5.95 years. For controls, this weight was 34.16 kilograms and age was 10.53 years. Age, weight, TPA and rTTW data were tested for normality using the Shapiro-Wilk test. Measured variables were evaluated independently for differences between cases and controls using the Wilcoxon rank-sum test. Data summaries and initial statistical analyses were performed using the R statistical package [R Core Team, 2016].

To correct for population structure and confounding variables in the dataset, principal components of a genetic relationship matrix were estimated from a pruned set of SNPs with LD $r^2 < 0.5$. Catell's scree test was used to determine the number of principal components to retain for analysis [Cangelosi 2007]. Case-control phenotypes were residuals of multiple logistic regression against variables known to influence ACLR in dogs [Witsberger 2008]: age, weight, and neuter status, as well as the first 6 principal components (see Figure 5.1S). Before association analysis, each phenotype was quantile transformed to a standard normal distribution [Shim 2015].

Genome-wide multivariate association was performed using mvBIMBAM software [Stephens 2013; Shim 2015]. MvBIMBAM frames association analysis as a model comparison problem whereby genotypes may be directly associated with a phenotype, indirectly associated with a phenotype (perhaps through another phenotype), or unassociated with a phenotype. The genotype may also be directly associated with more than one phenotype (referred to here as the multivariate phenotype). Bayes Factors are calculated to measure support for each model compared to the null (no association). The model comparison framework allows the user to determine not only whether an association exists between a genotype and phenotype(s), but which phenotypes are responsible for the association [Stephens 2013]. Bayes Factors (Log_{10} scale) were evaluated for evidence of association with the multivariate phenotype. Marginal posterior probabilities of associated SNPs were then evaluated to determine which phenotypes may be influencing the association. Regions with evidence of association were evaluated using

the CanFam3.1 Broad Improved Canine Annotation catalog [Hoeppner 2014] in the UCSC Table Browser [Karolchik 2004].

Results

Genotyping information was available for 336 purebred Labrador Retriever dogs. 114 dogs were excluded due to radiographs of inadequate quality, leaving 222 dogs for GWAS analysis. The final dataset contained 135,482 SNPs from 69 cases and 153 controls. The ratio of females to males in case and control groups was 0.97 and 0.88, respectively. Of the 116 males, 83 were castrated (71.5%). Of the 106 females, 87 were ovariohysterectomized (82.1%). There was no significant difference in the distribution of neutered animals across case and control groups ($\chi^2=0.1583$, $P=0.69$).

Before missing data imputation, average ages of case and control dogs retained for analysis were 6.1 and 10.5 years, respectively, and average weights were 37.67 and 33.96 kg, respectively. A summary of statistical analysis for individual covariates and additional phenotypes is given in Table 5.1. Shapiro-Wilk tests confirmed all variables were not normally distributed. Therefore, Wilcoxon rank-sum tests were used to evaluate variables for differences between case and control groups. By design, dogs in the control group were significantly older than those in the case group. Dogs affected with ACLR were both significantly heavier and had a significantly smaller rTTW than unaffected dogs. Measured TPA did not differ between case and control groups.

Table 5.1. Summary statistics for individual covariates and phenotypes in ACLR case and control groups

Variables	Case	Control	<i>P</i>
Age (years)	6.4 (0.88-12.5)	10.5 (8.0-14.8)	< 2.2E-16*
Weight (kg)	37.1 (27.0 – 58.5)	34.2 (20.8-50.3)	0.00015*
TPA (degrees)	29.0 (20.9-35.0)	28.0 (20.9-37.6)	0.13
rTTW	0.64 (0.46-1.00)	0.70 (0.45-1.00)	0.00032*

All values reflect median values and range (parentheses) in case and control datasets. Values with an asterisk (*) indicate that result was statistically significant. TPA = tibial plateau angle; rTTW = relative tibial tuberosity width. *P*-values are result of Wilcoxon rank-sum tests performed for age, weight, TPA, and rTTW.

The multivariate GWAS provided moderate evidence of association (Bayes Factor >3) for 4 SNPs with the phenotypes (Figure 5.2, Table 5.2). Of these, two SNPs on chromosome 23 were in perfect LD and may be considered a single association. The three regions identified were on chromosomes 1, 4, and 23. The association on chromosome 1 is located within the tyrosine-protein kinase transmembrane receptor 2 gene (*ROR2*), which is involved in early formation of chondrocytes, and is required for cartilage and growth plate development. The association on chromosome 4 is within the dedicator of cytokinesis 2 gene (*DOCK2*), an important regulator of lymphocyte migration. The association on chromosome 23 is in an intergenic region close to a long non-coding RNA of unknown function.

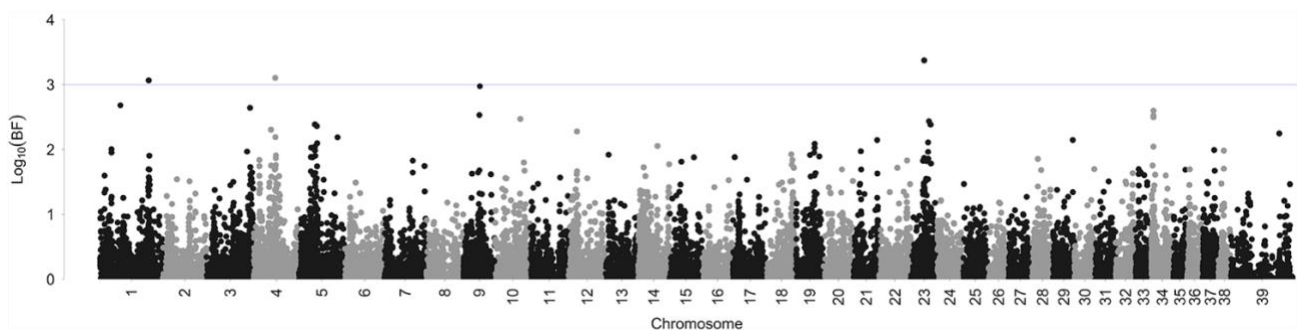


Figure 5.2. Manhattan plot of $\log_{10}$ Bayes factors (BF) of the multivariate phenotype. The multivariate phenotype is the combined effect of tibial plateau angle (TPA), relative tibial tuberosity width (rTTW), and ACLR case-control status. SNPs on chromosomes 1, 4, and 23 showed moderate evidence of association with the multivariate phenotype ($BF > 3$).

Table 5.2. SNPs associated with multivariate phenotype of anterior cruciate ligament rupture (ACLR), tibial plateau angle (TPA), and relative tibial tuberosity width (rTTW)

SNP	Chr	Position	BF _{mult}	BF _{ACLR}	BF _{TPA}	BF _{rTTW}	Gene	Location (bp)	Exon Dist. (kb)
BICF2G630788965	1	95183740	3.06	0.78	-0.23	1.75	<i>ROR2</i>	95124036-95282085	51.9
BICF2P1286728	4	42104780	3.10	0.67	2.18	-0.24	<i>DOCK2</i>	41779969-42177966	0.34
BICF2P1160758	23	26140588	3.37	4.24	-0.16	-0.11	(lncRNA)	26107359-26132719	N/A
TIGRP2P303751	23	26148612							

This table shows Bayes Factors (BF) for multivariate association test as well as univariate tests. BF_{mult} = result for multivariate phenotype; BF_{ACLR} = ACLR only; BF_{TPA} = tibial plateau angle only; BF_{rTTW} = relative tibial tuberosity width only. SNPs on chromosomes 1 and 4 were located within intronic regions of genes. The distance from the SNP location to the nearest gene exon is reported (Exon Dist.). The locus on chromosome 23 resides in a region <8kb from a long non-coding RNA (lncRNA).

An advantage of the mvBIMBAM package is the ability to discern which phenotypes in the multivariate model are responsible for the association. The posterior probability of effect (PPE) calculated for each SNP and the corresponding phenotype are displayed in Table 5.3. All identified SNPs showed strong probability of association with ACLR. In addition to association with the case-control phenotype, the *DOCK2* SNP was also associated with TPA and the *ROR2* SNP was also associated with rTTW. The association on chromosome 23 was primarily associated with ACLR and not TPA or rTTW. Evidence for this can be seen with higher BF_{ACLR} compared to BF_{mult} and relatively low PPE of TPA and rTTW compared to ACLR for this locus. It should be noted that PPE is the sum of the probability of a direct effect and the probability of an indirect effect on the phenotype. In all cases, the PPE was primarily explained by the probability of a direct effect on the individual phenotypes, with little influence from indirect effect.

Table 5.3. The posterior probability of effect (PPE) calculated for each SNP and corresponding phenotype

SNP	PPE _{ACLR}	PPE _{TPA}	PPE _{rTTW}
BICF2G630788965	0.87	0.28	0.99
BICF2P1286728	0.86	0.99	0.48
BICF2P1160758	0.99	0.49	0.54
TIGRP2P303751			

Posterior probability of effect (PPE) is 1 minus the probability of no effect on the phenotype.

Discussion

Through multivariate association analysis of ACLR case and control dogs, TPA, and rTTW, we have identified three loci with moderate evidence of association with ACLR in the Labrador Retriever model. Two of these loci are within genes that have not previously been linked to ACLR in human or dog studies, though they are highly relevant given what is known regarding ACLR pathogenesis. ACLR in dogs is a highly polygenic complex trait [Baker 2017;

Baird 2014b], and our group previously reported 99 risk loci associated with the condition using a dataset containing many of the same dogs used in the current analysis [Baker 2017]. The results of the current study should be considered complimentary, and provides evidence for the benefit of using a multivariate approach to improve power to detect smaller effect associations expected when studying the genetic contribution to complex trait disease.

Before association analysis, we evaluated covariates and supplementary phenotypes for differences between cases and controls. The statistically significant difference in age between cases and controls was expected, as age was an important part of the recruitment strategy [Baker 2017]. Additionally, case dogs in the sampled group weighed significantly more than controls. The association between higher body weight and increased risk of ACLR in dogs has been reported [Duval 1999; Inauen 2009]. While this is a notable finding, it is important to recognize that these findings do not distinguish between dogs of large body size and those that are overweight.

Several studies have evaluated the effect of excessive TPA on ACLR risk in dogs. The ACL opposes anterior tibial thrust. A steeper TPA increases anterior tibial thrust, placing increased mechanical stress on the ACL and potentially predisposing the ligament to rupture [Reif 2004]. Results of case control studies of TPA on ACLR risk have been inconsistent with multiple studies declaring an excessively steep TPA as a risk factor [Morris 2001; Mostafa 2009; Su 2015; Janovec 2017] while others were not able to identify a significant effect [Reif 2004; Buote 2009; Fuller 2014b]. The present study did not identify a difference in TPA between affected and unaffected dogs. However, the TPA of affected dogs was steeper than that of unaffected dogs, which is consistent with previous studies that were able to identify significant differences. Though it remains plausible that TPA has some effect on ACLR risk in dogs, it is likely that this effect is either indirect or fairly small. Posterior tibial slope (PTS) in humans is analogous to TPA measured in dogs. Similar to dogs, human patients affected with ACLR have

steeper PTS [Giffin 2004; Brandon 2006; Todd 2010] and this effect may be more pronounced in female patients [Hohmann 2011].

Evaluation of rTTW and ACLR status confirmed results of a previous study that reported the rTTW of dogs affected with ACLR is significantly narrower than the rTTW of unaffected dogs [Inauen 2009]. Dogs with a relatively narrow tibial tuberosity experience greater anterior tibial thrust, again placing greater mechanical stress on the ACL and potentially increasing rupture risk. While our study was able to confirm rTTW as a risk factor for ACLR in dogs, a recent study in small breed dogs did not identify a difference between case and control groups [Janovec 2017]. It may be that risk due to rTTW is limited to large breed dogs, and this possibility should be evaluated in future epidemiological studies. To the authors' knowledge, tibial tuberosity morphology has not been evaluated as a risk factor for ACLR in human beings.

The locus with greatest evidence of association was located on chromosome 23. This locus was primarily associated with ACLR and not TPA or rTTW. This locus is within an intergenic haplotype block that is less than 8kb from the transcription location of a long non-coding RNA (lncRNA) with unknown function. LncRNAs are broadly defined as long transcripts (>200 nucleotides) that do not function in a protein-coding fashion [Quinn 2016]. While some of these transcripts are likely non-functional, a growing list of lncRNAs have specific biological roles in many categories including cellular proliferation, chromatin remodeling, and regulation of gene transcription and translation [Kung 2013, Quinn 2015]. LncRNA expression tends to be highly tissue-specific [Cabili 2011]. The lncRNA identified appears to be primarily expressed in testicular tissue, but not ovarian tissue [Hoeppner 2014]. Expression levels were not evaluated in tissue that might be associated with the joint (e.g. collagenous tissue or synovium). While there is evidence of gonadal influence on ACLR risk in both humans and dogs [Hewett 2007; Witsberger 2008; Beynnon 2008; Torres de la Riva 2013], it is difficult to speculate on potential function of this lncRNA, if any, without further investigation.

The SNP association located on chromosome 4 was within the *DOCK2* gene. *DOCK2* protein interacts with Rac1 to induce lymphocyte migration into tissues, including the synovium. *DOCK2* signaling has been linked to synovitis associated with rheumatoid arthritis in human beings [Whitaker 2015]. Lymphoplasmacytic synovitis precedes ACLR in dogs, and arthritis is typically present at the time of diagnosis [Bleedorn 2011; Little 2014]. Additionally, numbers of lymphocytes are positively correlated with radiographic evidence of degenerative joint disease, indicating that lymphocytic inflammation likely contributes to progression of joint disease [Muir 2011b]. In humans, synovitis and associated lymphocytic inflammation plays a role in post-injury osteoarthritis progression and chronicity [Haynes 2002]. In an animal model of joint pathology, experimentally induced synovitis resulted in weakening of the cruciate ligaments [Goldberg 1982]. In both humans and dogs, surgical correction of joint instability fails to prevent osteoarthritis development and progression, indicating that disease progression has a substantial biochemical component [Lohmander 2007; Simon 2015; Paschos 2017]. It is feasible that aberrant *DOCK2* signaling may contribute to the inflammatory cascade associated with ACLR and play a role in both onset and progression of disease.

The SNP association located on chromosome 1 was within the *ROR2* gene, a protein with multiple roles, any of which may be relevant to ACLR. *ROR2* has a well-established role in bone development, where it is essential for appropriate patterning and chondrocyte expansion during growth [DeChiara 2000]. *ROR2* also plays complimentary roles in both osteoblastogenesis and osteoclastogenesis from mesenchymal stem cells [Maeda 2012; den Hollander 2015]. Aberrant *ROR2* signaling appears to play a role in bone loss associated with rheumatoid arthritis [Holley 2015], and has been shown to be significantly up-regulated in these patients [Song 2015]. *ROR2* plays a more general role in cellular proliferation and migration, and this role has been implicated in several cancers of various tissue types [Rebagay 2012, Ford 2013]. Its role in cellular expansion also has important implications for joint healing and homeostasis [DeChiara 2000] as tissue repair requires proliferation and migration to sites of

healing. Down-regulation of *ROR2* inhibits the regenerative capacity of chondrocytes [DeChiara 2000]. Indeed, *ROR2* is down-regulated in human patients with end-stage osteoarthritis [den Hollander 2015]. These roles indicate that altered *ROR2* signaling during development may affect the patterning and growth of long bones in a way that augments ACLR risk. Later in life, aberrant *ROR2* signaling may influence processes important for healing damaged ligament or cartilage and could contribute to post-rupture progression of osteoarthritis.

The results of the current study did not overlap with those from our previous GWAS of ACLR in Labrador Retrievers [Baker 2017]. The present analysis was substantially different from our previous GWAS in a number of ways, and this is likely to have impacted the results. The current dataset represents dogs that had been in the previous GWAS as well as dogs that have since been added to our dataset. Therefore, while dog breed, geographical location, and recruitment strategy remained constant, the sample analyzed was substantially different from the previous analysis. Additionally, we made further correction for environmental risk factors by regressing the case-control phenotype against age, weight, and neutered status, whereas the previous study only included neutered status as a covariate. It should also be noted that the present multivariate analysis was performed using an algorithm that employs a Bayesian approach, while the previous GWAS was based on traditional linear mixed model analysis of a single case-control phenotype. Ultimately, the difference in results between the two studies highlights the need for validation in a new and larger population, as well as through other methods beyond GWAS including whole genome sequencing and gene expression studies.

Conclusions

A multivariate analysis of ACLR, rTTW, and TPA has identified three novel variants with moderate probability of effect on ACLR. Two of these variants reside within genes that have previously been implicated in other joint pathologies, particularly rheumatoid arthritis and progression of osteoarthritis. This analysis complements our previous discovery GWAS of

ACLR in the dog model, and provides further support for ACLR as a complex, multifactorial disease that may be influenced by aberrant signaling in the immune system and/or developmental variables. While multiple GWAS for ACLR have been performed in the dog model, this study is the first to identify genes directly related to joint morphology and homeostasis, which speaks to the power of the multivariate approach. Both *DOCK2* and *ROR2* represent potential drug targets for therapies targeting the biochemical aspect of ACLR and associated osteoarthritis. CPYPP is a small molecule inhibitor of *DOCK2* that may provide the scaffold for a *DOCK2*-targeting immunosuppressant [Nishikimi 2012]. As an identified oncogene, *ROR2* has been described as a candidate for targeted therapy for several cancers, and the search for drugs that modulate *ROR2* activity is currently underway. Future work to validate these loci as bona fide risk factors for ACLR will provide improved understanding of ACLR pathogenesis as well as the opportunity for medical intervention through targeted drug therapy. Such advances will have an important impact on both human and animal orthopaedic health.

Declarations

Ethics approval and consent to participate

All procedures were performed in strict accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health and the American Veterinary Medical Association and with approval from the Animal Care Committee of the University of Wisconsin-Madison (protocol #V1070, V005463). Informed consent of each owner was obtained before participation in the study.

Supplementary Information

Figure 5.1S. Scree plot of principal components eigenvectors calculated from Labrador Retriever genotypes. The majority of variance is explained by the first 6 principal components, which were chosen to include as covariates in the analysis.

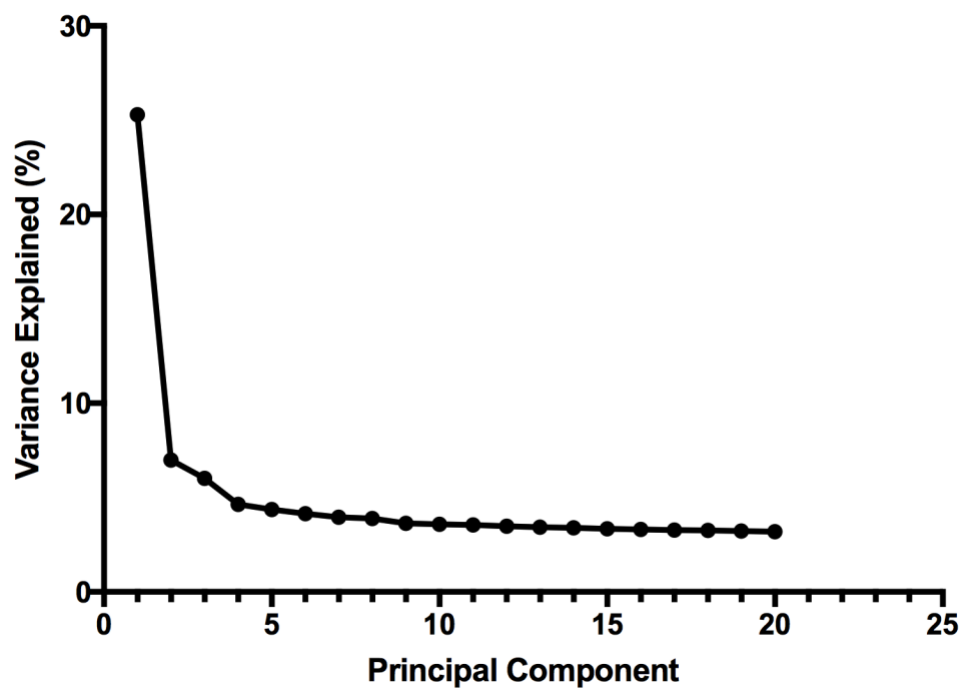
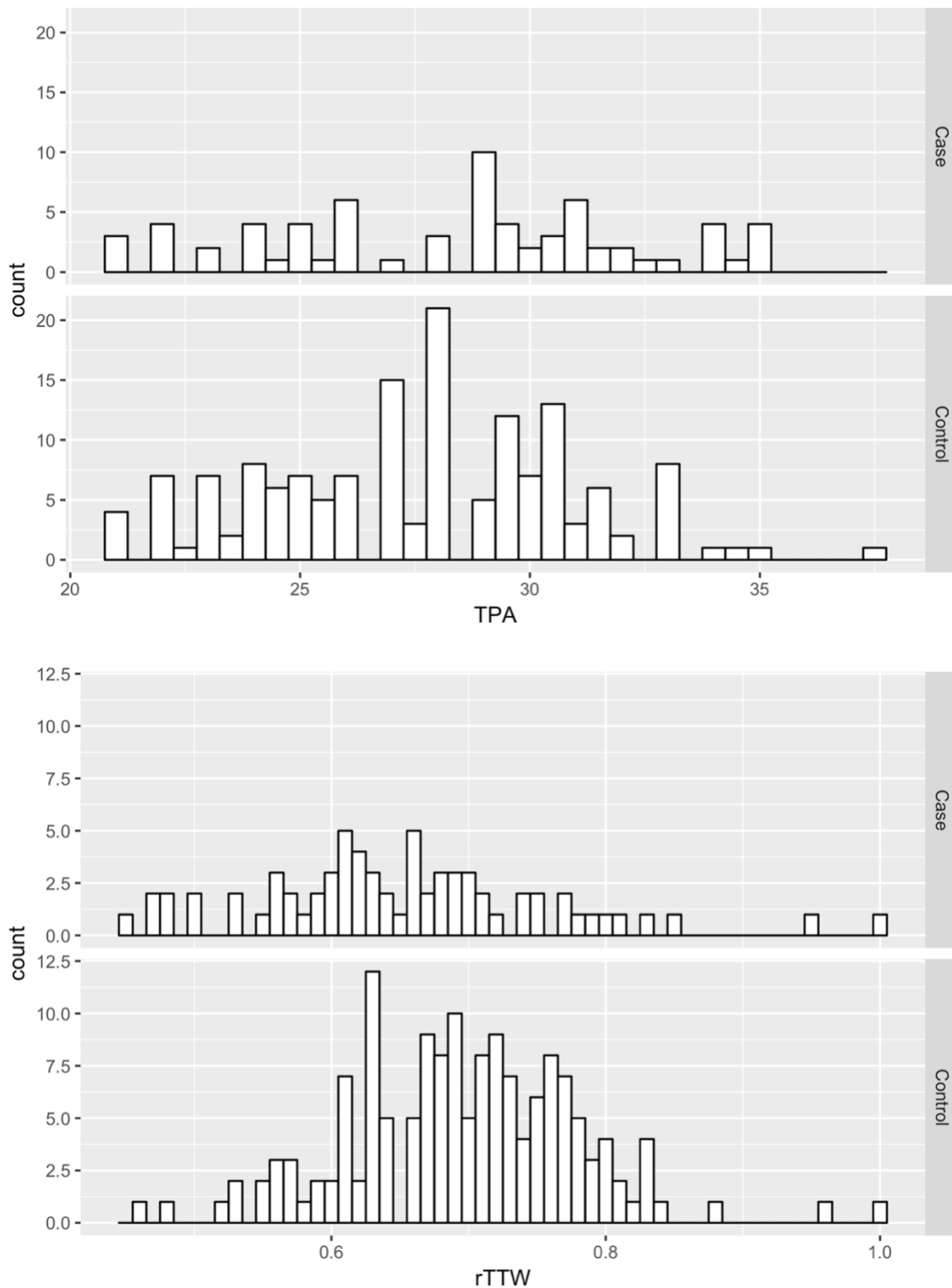


Figure 5.2S. Distribution of tibial plateau angle (TPA) and relative tibial tuberosity width (rTTW) among cases and controls.



Chapter 6

Bayesian and Machine-Learning Models for Genomic Prediction in a Dog Model of ACL Rupture

Baker LA, Chan K, Lopes FB, Bollig N, Todhunter RJ, Rosa GJM, Binversie EE, Sample SJ, Muir P. Bayesian and machine learning models for genomic prediction of anterior cruciate ligament rupture in the canine model. G3: Genes, Genomes, Genetics. 2019. *In Review*.

Abstract

Anterior cruciate ligament rupture (ACLR) is a common, debilitating condition that leads to early-onset osteoarthritis and reduced quality of life. ACLR is a complex trait with both genetic and environmental risk factors. Characterizing the genetic basis of ACLR would provide the ability to identify individuals that have high genetic risk and allow the opportunity for preventative management. Spontaneous ACLR is also common in dogs and shows a similar clinical presentation and progression. Thus, the dog has emerged as an excellent genomic model for human ACLR. Genome-wide association studies (GWAS) in the dog have identified a number of candidate genetic variants, but research in genomic prediction has been limited. In this analysis, we explore several Bayesian and machine learning models for genomic prediction of ACLR in the Labrador Retriever dog. Our work demonstrates the feasibility of predicting ACLR from SNPs in the Labrador Retriever model with and without consideration of non-genetic risk factors. Genomic prediction of a small independent sample approached clinical relevance using gradient boosted trees. This analysis represents the first steps towards development of a predictive algorithm for ACLR in the Labrador Retriever model. Future work may extend this algorithm to other high-risk breeds of dog. The ability to accurately predict individual dogs at high risk for ACLR would identify candidates for clinical trials that would benefit both veterinary and human medicine.

Introduction

Anterior cruciate ligament rupture (ACLR) is a common condition with serious long-term consequences, as up to 50% of affected individuals will develop osteoarthritis (OA) within 10 years of rupture [Lohmander 2007]. This is especially troubling given that the highest incidence is in adolescent athletes [Lohmander 2007], who will experience a significant health burden while they are still very young. The impact of this reality is reflected in the lifetime burden of ACLR in the United States, which is \$7.6 billion annually if surgical treatment is pursued, versus \$17.7 billion if treatment is limited to physical rehabilitation [Mather 2013]. The vast majority of ACLRs occur in the absence of contact injury [Smith 2012a], and surgical reconstruction does not consistently prevent development of OA, which supports the hypothesis that ACLR is at least partially due to biochemical influences. Several risk factors have been identified, including genetic predisposition [Smith 2012a, 2012b, Kaynak 2017]. Understanding the genetic basis of ACLR is important, as it would allow medical professionals to identify those individuals that have a higher inborn risk of rupture. Interventions could then take place to mitigate risk and potentially prevent these people from developing ACLR.

Spontaneous ACLR is also a disease of importance in veterinary medicine, as the condition is diagnosed in 20% of dogs evaluated for lameness at university hospitals [Johnson 1994]. The American public spends greater than \$1 billion annually on treatments for canine ACLR [Wilke 2005]. ACLR in dogs has a similar presentation and progression to ACLR in humans, including development of OA in spite of surgical stabilization [Rayward 2004]. Thus, spontaneous ACLR in dogs has emerged as an excellent model for ACLR in human beings [Gregory 2012; Proffen 2012]. ACLR in dogs has particular value as a genomic model, as the condition has a marked breed-predisposition, and in some breeds, prevalence is ~100 fold greater than in human beings [Witsberger 2008; Gianotti 2009]. Extensive linkage disequilibrium (LD) in dogs facilitates genome-wide association study (GWAS) [Karlsson 2008], and multiple ACLR GWAS in dogs have been undertaken [Baird 2014; Hayward 2016; Baker 2017; Baker

2018]. However, most of this research has focused on biological interpretation of SNPs that reach genome-wide significance, and little has been done to attempt genomic prediction of canine ACLR.

Genomic prediction as a method focuses less on individual SNPs and assumes that all SNP markers may be linked to causal variants, even if their effects are quite small [Meuwissen 2016]. These polygenic effects act in combination to influence risk of disease [Robinson 2014]. The number of genetic variants that are believed to affect complex traits, such as ACLR, has increased ~100 fold during the last 18 years with most estimates suggesting there are thousands of small effect variants distributed across the genome [Meuwissen 2016]. SNPs with measurable effects can be used on their own to estimate genetic risk or combined with measurements of non-genetic risk factors to create absolute risk models that estimate the probability that an individual will develop the disease over time [Chatterjee 2016]. The ability to predict ACLR in dogs would be extremely valuable from a veterinary medical perspective, but also because it would enable prospective research of interventional treatments using spontaneous ACLR in the dog as a model for human ACLR. Insights gained from research in the dog model would lead to advancements in both veterinary and human medical research.

There are multiple methods for genomic prediction. Each method has advantages and disadvantages with respect to model assumptions, and how well the model fits the data. With respect to prediction of complex traits, points to consider when choosing a model include the genetic architecture of the trait in terms of the potential presence of major genes, epistatic interactions, and a polygenic component. In addition, other factors to be considered include marker density and the strength of LD among them, as well as sample size [Hayes 2010; Perez 2014]. Bayesian models lend themselves well to genomic prediction, as they have the ability to incorporate prior information about expected SNP effects, for example allowing SNPs to have varying effect sizes, which makes more sense biologically than assuming all SNPs have the same effect size [Moser 2015; Meuwissen 2016]. Classification-based machine learning

methods have also gained popularity for genomic prediction of binary traits. Here, a GWAS training set is viewed as a supervised classification problem whereby individuals are partitioned into case or control groups, and each group can be described using a combination of SNP inputs that may have one of 3 discrete values corresponding to the number of minor alleles present at each SNP [Botta 2014]. As no single model has been shown to perform best across data sets and traits [Perez 2014], the following analyses were performed to investigate the feasibility of genomic prediction of ACLR in the dog model using several Bayesian and machine learning approaches. We provide insight on which methods appear to be most suitable for genomic prediction of a complex trait disease in purebred dogs, and potential and future directions for development of a predictive genetic test for ACLR.

Materials and Methods

Data collection and phenotyping

Client-owned Labrador Retrievers were recruited from the UW-Madison UW Veterinary Care teaching hospital and through online advertising. All owners gave informed consent to participate in the study. When possible, a four-generation pedigree was obtained to confirm purebred status. Each dog was carefully phenotyped through orthopaedic exam [Muir 1997] and lateral stifle radiographs. ACLR in affected dogs was verified during surgical treatment. Dogs classified as controls were over the age of 8 years, negative for palpable knee laxity, and showed no signs of joint effusion or osteophytosis that would be consistent with ACLR on lateral stifle radiographs [Chuang 2014]. This age cutoff was chosen because Labrador Retrievers 8 years of age and older have approximately a 6% chance of developing ACLR [Reif 2004]. DNA was isolated from saliva or blood samples obtained in accordance with the Guide for the Care and Use of Laboratory Animals and were approved by the Institutional Animal Care and Use Committee of the School of Veterinary Medicine, University of Wisconsin-Madison. SNP

genotyping was performed using the Illumina Canine HD BeadChip, which contains approximately 174,000 SNPs distributed evenly across the canine genome (CanFam3.1). In addition, Bayesian analyses used public data from a recent study that used the same genotyping platform [Hayward 2016] to increase sample size by 287 Labrador Retriever dogs.

SNP genotyping quality control

Genotype data were filtered with PLINK for quality control [Chang 2015]. All samples had a genotyping call rate >95%. SNPs were excluded if minor allele frequency (MAF) was less than or equal to 0.05, if genotyping rate was less than or equal to 95% or if there was deviation from Hardy-Weinberg proportions at $P < 1E-07$.

Bayesian Analyses

Because Bayesian analyses used additional data from another publication which represented dogs that were recruited in a different region, principal components analysis was performed before model fitting to investigate the existence of population stratification in the dataset. Genomic prediction models were fitted using five Bayesian logistic model specifications: Bayesian Ridge Regression, Bayesian LASSO, Bayes A, Bayes B, and Bayes C π [Gianola 2013]. For these methods, the general linear predictor was:

$$\boldsymbol{\eta} = 1\boldsymbol{\mu} + \boldsymbol{X}\boldsymbol{b} + \sum_{j=1}^K \boldsymbol{z}_j \boldsymbol{a}_j + \boldsymbol{e}$$

where $\boldsymbol{\eta}$ is a $n \times 1$ vector of the liability; $\boldsymbol{\mu}$ is an intercept; $\boldsymbol{X}$ is an incidence matrix of the fixed effects in $\boldsymbol{b}$ (weight, sex, neutering, and source of the data ([Baker 2018] or [Hayward 2016])); K is the number of markers fitted; $\boldsymbol{z}_j$ is an $n \times 1$ vector denoting the genotypes of the animals for marker j ; $\boldsymbol{a}_j$ is the effect marker j ; and $\boldsymbol{e}$ is a vector of residual effects. The SNP genotypes were coded as 0, 1, and 2 for AA, AB, and BB, respectively where A is the major allele and B is the

minor allele. The vector of residuals was assumed to be normally distributed. Models were run using the BGLR statistical package [Perez 2014] in R (www.Rproject.org) for a total of 40,000 iterations with the first 30,000 iterations discarded. Each Bayesian model employs different prior assumptions for marker effects. A brief description denoting the difference between the models follows.

Bayesian ridge regression

In Bayesian Ridge Regression (BRR), an independent Gaussian prior with common variance is assigned to each regression coefficient. This scenario assumes that all markers have some effect and shrinkage is applied homogenously across the dataset.

Least Absolute Shrinkage and Selection Operator (LASSO)

Bayesian LASSO regression (Park and Casella 2008), uses a double-exponential or Laplace prior distribution for marker effects. This places a higher mass at zero, meaning it induces a strong shrinkage toward zero. This is a logical application in a situation where most of the many thousands of SNP markers available are assumed to have little or no effect on the trait being tested.

Bayes A

Bayes A [Meuwissen 2001] uses a scaled-t prior distribution for marker effects. Similar to Bayesian LASSO, this places a higher mass at zero, inducing strong shrinkage toward zero. The scaled-t distribution places slightly less emphasis on shrinkage toward zero, allowing more flexibility for marker effects than Bayesian LASSO [de los Campos 2013].

Bayes B

Bayes B assumes that most of the genetic markers have zero effect, so that the distribution can be described as a mixture model where π is the probability that the SNP has no

effect and $(1-\pi)$ is the probability that the SNP contributes to genetic variance [Meuwissen 2001]. Non-null marker effects are assumed to have a scaled-t prior distribution, as in Bayes A. Therefore, the model is fairly stringent, assuming that relatively few markers have non-null effects.

Bayes $C\pi$

Bayes $C\pi$ [Habier 2011] is similar to Bayes B, except that a prior distribution is also assumed for the proportion of null effect markers, and non-null effect markers are assumed to have a Gaussian prior with a common variance. Unlike Bayes B, in Bayes $C\pi$ the probability of a SNP having no effect, π , is treated as an unknown and is inferred from the data. To do this, the inclusion (or exclusion) of each marker in the model is modeled with a specific inclusion parameter (i.e. indicator variable), δ_i , which is equal to 1 if the marker j is fitted in the model and zero otherwise.

Leave-One-Out Cross Validation

Bayesian models used a Leave-One-Out Cross Validation (LOOCV) technique to compare the prediction ability of the different models considered. In LOOCV, a single observation is used for the validation set, and the remaining observations make up the training set. The statistical learning method is fit to the $n-1$ training observations, and a prediction of ACLR case or control status is obtained for the excluded observation. The approach is repeated n times, one for each observation taken as a validation set. The prediction accuracy was assessed by area under the receiver operating characteristic (ROC) curve (AUC).

Machine Learning Analyses

Feature (SNP) selection

Linkage disequilibrium (LD) is extensive in purebred dog populations [Sutter 2004]. In genomic selection, SNPs that are in LD with the risk loci serve as surrogates in the model. In some machine learning applications, however, the strong LD among SNPs may lead to diminished importance of the true risk loci or tag SNPs in the model, as their effects may end up being partially captured by many SNPs. To mitigate this effect, SNPs with LD r^2 greater than 0.7 were pruned using PLINK with a window size of 50 SNPs and overlap of 5 SNPs until no pairs remained. SNPs were selected for inclusion in the training set by one of two filter methods: 1) ranked P -values from a linear mixed model GWAS using Genome-wide Efficient Mixed Model Analysis (GEMMA) [Zhou 2012], where smaller P -values are considered most likely associated with ACLR or 2) ranked SNPs based on the mean difference in allele frequency between cases and controls. SNPs with the largest mean difference were considered to be the most likely associated with ACLR [Hajiloo 2013]. The number of genetic variants believed to affect ACLR in dogs is unknown, though there are likely hundreds to thousands of non-null effect SNPs [Baker 2017; Baker 2018]. Therefore, prediction performance of each model was assessed at several SNP inclusion thresholds from 5 to 15,000 SNPs.

Experimental Design

Exploration into the performance of classification-based data mining methods for predicting ACLR in Labrador Retrievers was divided into two broad experiments: 1) 10-fold cross-validation to provide a robust evaluation of the various methods and 2) holdout validation to confirm findings with prevention of any overfitting to training data. Both experiments were performed with and without LD pruning of SNPs and with and without covariates to illustrate the effects of feature correlation reduction and inclusion of non-genetic risk factors on ACLR prediction (Figure 6.1).

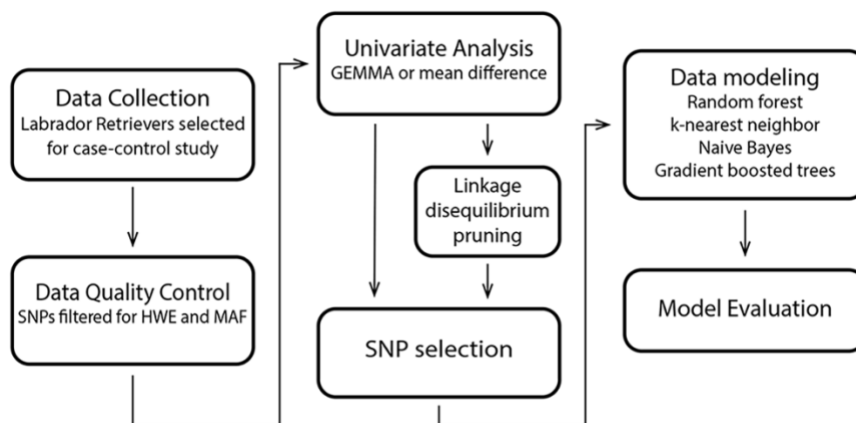


Figure 6.1. Schematic of data analysis and modeling workflow. HWE: Hardy-Weinberg equilibrium; MAF: minor allele frequency; SNP: single nucleotide polymorphism; GEMMA: Genome-wide Efficient Mixed Model Analysis.

10-fold cross-validation was performed to assess performance of each classification-based data mining method. Care was taken in 10-fold cross-validation runs to ensure that feature selection was performed only with consideration to the training set without knowledge of the test set for each fold. Each model was trained three times and the prediction accuracy was averaged across the runs. The classification models were again assessed by prediction performance scored by AUC. In this study, four classification methods were considered. A brief description of each method follows:

Weighted Subspace Random Forest

In Random Forest (RF), a collection (“forest”) of separate tree-structured classifiers each cast a vote for the classification of an input and the majority vote of the trees is chosen as the correct classification [Breiman 2001]. This method has the benefits of being fast and unlikely to over-fit to the dataset. Further, it is easily optimizable and provides variable importance estimates for further feature refinement. One shortcoming of random forest for high-dimensional data is the random selection of features which can fail to consistently select informative features. To address this issue, weighted subspace random forest (wRF) was used in the final validation of the methods. wRF weights each of the SNPs based on correlation of the SNP with

the case or control class. It then calculates probability based on weights and uses it for variable selection [Zhao 2017]. wRF was implemented using the R package 'wsrf' [Zhao 2017]. Models were built using at least 1000 trees and the square root of the total number of features at each tree split.

Gradient boosted Trees

Similar to RF, gradient boosted trees (GBT) uses an ensemble of tree-based classifiers for phenotype prediction. However, instead of creating decision trees independently of the other trees, trees are created conceptually in serial order, with each new tree attempting to minimize the mean squared error of the previous trees [Natekin 2013]. Gradient boosting theoretically provides an advantage over random forest at the cost of greater computational complexity and the need to tune hyperparameters. The R package 'xgboost' [Chen 2016] was used for implementation of gradient boosted trees. Tuning of the hyperparameters was performed using 5-fold cross-validation grid search techniques. The cross-validation function from xgboost was used to determine the number of rounds to run the algorithm. The hyperparameters used were learning rate $\eta = 0.05$, minimum loss reduction $\gamma = 0.3$, maximum tree depth = 10, subsample ratio of columns when constructing trees = 0.8, subsample ratio of training instances = 0.8 and evaluation metric of binary classification error rate with 1000 rounds of training.

Naïve Bayes

One of the first machine learning methods used in bioinformatics, Naïve Bayes (NB) is a classification method based on Bayes' theorem. A training set is used to calculate frequencies of genotypes in case or control individuals, and this information is used to calculate the probability of an unknown individual's classification. NB is known for being simple and computationally efficient, but it is prone to miscalibration when features are high in number, as is the case with SNP datasets [Acikel 2016]. Though it has been theoretically outclassed by

ensemble data mining methods, NB is still an excellent baseline for comparing classifiers [Acikel 2016]. The R package 'e1071' [Meyer 2017] was used for NB implementation.

K-nearest Neighbors

K-nearest neighbors (KNN) is the most simplistic classifier, as it does not build a classifier using the training data. Instead, KNN compares the unknown input with classification of the k -nearest data points and uses the features of these neighbors to classify the unknown input. If multiple classifications are possible, a majority vote is applied [Acikel 2016]. Because this method does not depend on training and tuning, it serves as another baseline method for comparing other classifiers. The R package 'caret' [Kuhn 2008] was used for KNN implementation. Models considered the five closest neighbors for classification decisions.

Ensemble Learning

Ensemble learning methods were applied to determine whether better predictive performance could be obtained when multiple classifiers are considered in aggregate. Two methods of ensemble learning were used, 1) n-Agreement ensemble and 2) supervisory machine learning. LD pruning was applied to the SNP dataset and covariates were included as additional features in all models used for ensemble learning.

When the four machine learning algorithms described above were used with two methods of feature selection, a total of 8 base-level models were considered. n-Agreement ensemble defines an ensemble agreement algorithm at each integer n between 1 and 8, rendering a positive prediction if and only if at least n of the 8 base models agree on a positive prediction. This n-agreement ensemble was applied on each fold within the cross-validation workflow at each integer value of n between 1 and 8.

In the supervisory machine learning approach, predictions from each of the 8 base-level learners were used as features in 1) logistic regression or 2) random forest models. The cross-validation workflow was extended for this ensemble method. After performing 10-fold cross validation to train the base models, the aggregated predictions from these folds were randomly

re-ordered and re-partitioned into 10 new folds. An additional 10-fold cross-validation experiment was performed using the new folds as features in the supervisory models. This process was also extended to holdout validation, where the 8 base models were fit to the entire training set and applied to the independent test set. All models were evaluated by comparing predictions to true phenotypes using AUC.

Covariate analysis

Covariates used in the study were known non-genetic risk factors for ACLR in dogs. They included weight, sex, and neuter status (castration and ovariohysterectomy in males and females, respectively) [Witsberger 2008]. Covariates were incorporated as additional features in each classification method alongside SNPs. Covariates were also evaluated independently as predictors of ACLR using logistic regression. The R package ‘stats’ [R Core Team 2013] was used for implementation of the logistic regression function.

Results

The final dataset for Bayesian analyses included 622 Labrador Retriever dogs (247 cases and 375 controls). Among cases, there were 14 intact females, 25 intact males, 111 ovariohysterectomized females, and 97 castrated males. Among controls, there were 59 intact females, 65 intact males, 130 ovariohysterectomized females, and 121 castrated males. After SNP data quality control, 105,435 SNPs remained for analysis. After principal components analysis was performed, a single principal component could be used to separate dogs recruited at University of Wisconsin-Madison and dogs from Hayward et al. (2016) (Figures 6.1S and 6.2S). To correct for this effect, a categorical variable denoting the source of the data was incorporated into all models, which is equivalent to including a single principal component in the analysis.

Machine learning analyses used a subset of the above dataset containing only dogs recruited to the ACLR GWAS at the University of Wisconsin-Madison. This dataset contained 336 Labrador Retriever dogs (147 cases, 220 controls). Among cases, there were 10 intact females, 16 intact males, 53 ovariectomized females, and 55 castrated males. Among controls, there were 18 intact females, 30 intact males, 74 ovariectomized females, and 80 castrated males. After quality control, ~120,000 SNPs were used in classification-based data mining models depending on the exact subset of samples included. An additional dataset of 31 Labrador Retriever dogs (18 cases and 13 controls) was used for holdout validation.

Bayesian Analyses

The prediction accuracy for the Bayesian models described is shown in **Table 6.1**. Of the models, BRR had the best predictive performance in leave-one-out cross validation.

Table 6.1. Bayesian model performance for genomic prediction of ACLR in dogs

Model	AUC
Bayes A	0.581
Bayes B	0.587
Bayes C π	0.563
LASSO	0.589
BRR	0.614

LASSO = Least Absolute Shrinkage and Selection Operator; BRR= Bayesian ridge regression; AUC= Area under the ROC curve

Machine Learning Analyses

Results of 10-fold cross validation experiments are summarized in Table 6.2. In 10-fold cross-validation, larger numbers of SNPs in a feature set did not improve predictive performance (Figure 6.2). In general, models performed similarly regardless of the model chosen or methods used for feature selection, though models that used mean difference for

SNP selection required fewer SNPs to meet equivalent levels of performance. When LD pruning was not performed and covariates were not considered, the best performing model was GBT with 4,000 SNPs derived from GEMMA analysis (AUC = 0.599). Removal of highly correlated SNPs through LD pruning did not have a significant effect on classifier performance, though the same level of performance was achieved with fewer SNPs. Including covariates as predictors accentuated the performance of the classifiers, both with and without LD pruning. The best performing model overall was KNN with 10 SNPs chosen through mean difference (AUC = 0.609). This result was the same with and without LD pruning.

To corroborate the results of 10-fold cross-validation, holdout validation was performed in an independent test set of Labrador Retrievers (n=18 cases and n=13 controls) with models trained using the complete 336 dog dataset. In general, performing LD pruning and including covariates as predictors led to improvements in classifier performance, though these improvements were somewhat inconsistent (Table 6.3). Adding additional SNPs as features improved model performance for some models, particularly GBT, with peak performance achieved when models included 3000 or more SNPs. This gain in performance is lost at ~10,000 or more SNPs (Figure 6.3). Overall, the best performing classifier was GBT trained on 3000 SNPs chosen by mean difference with LD pruning and covariates (AUC = 0.808).

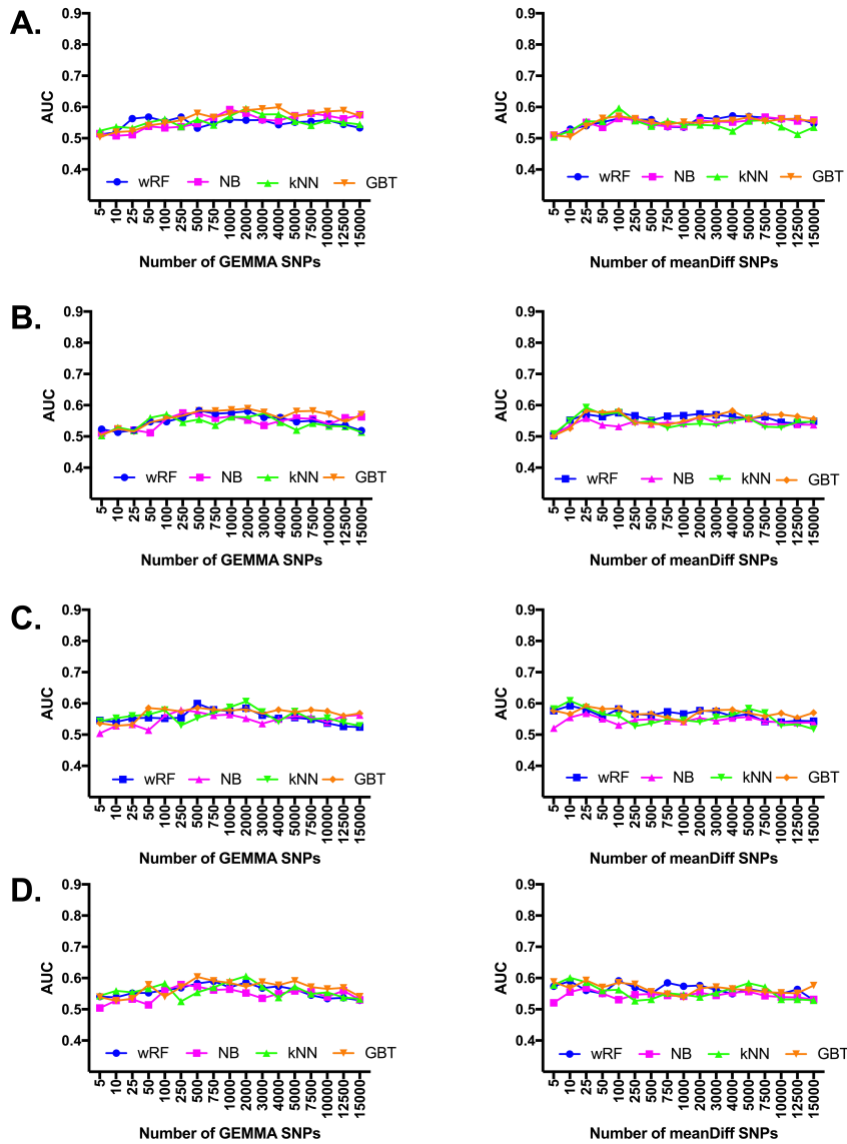


Figure 6.2. Results of 10-fold cross validation with models trained on feature sets from 5 to 15,000 SNPs. Averages for model prediction across all folds over three runs per model are reported. This analysis used $n=134$ cases and $n=202$ controls. A. Model performance without LD pruning or covariates; B. Model performance after LD pruning was performed at $r_2>0.7$. C. Model performance with covariates (weight, sex, neutering) considered as additional features. D. Model performance with LD pruning and covariates. AUC: area under the ROC curve; wRF: weighted subspace random forest; NB: Naïve Bayes; kNN: k -nearest neighbor; GBT: gradient boosted trees.

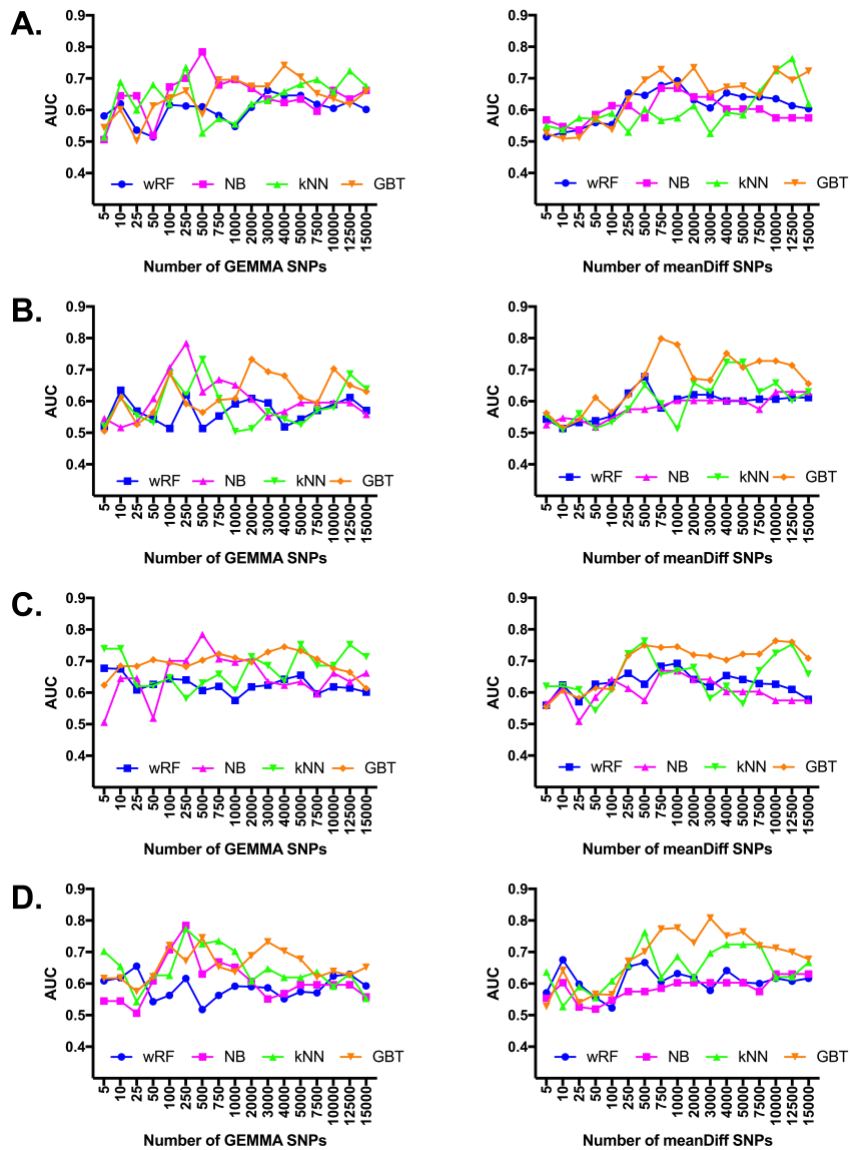


Figure 6.3. Results of holdout validation with models trained on feature sets from 5 to 15,000 SNPs. Holdout validation was performed with a Labrador Retriever training set ($n=134$ cases and $n=202$ controls) and a Labrador Retriever test set ($n=18$ cases and $n=13$ controls). Averages for model prediction over three runs per model are reported. A. Model performance without LD pruning or covariates; B. Model performance after LD pruning was performed at $r_2 > 0.7$. C. Model performance with covariates (weight, sex, neutering) considered as additional features. D. Model performance with LD pruning and covariates. AUC: area under the ROC curve. wRF: weighted subspace random forest; NB: Naïve Bayes; kNN: k -nearest neighbor; GBT: gradient boosted trees.

Table 6.2. Highest performing models in 10-fold cross validation for prediction of ACLR in Labrador Retriever dogs

Model	Feature Selection	No. SNPs	AUC
<i>No SNPs removed for LD; Covariates not considered</i>			
wRF	GEMMA	50	0.568
	meanDiff	4000	0.572
GBT	GEMMA	4000	0.599
	meanDiff	100	0.571
NB	GEMMA	1000	0.592
	meanDiff	7500	0.568
KNN	GEMMA	2000	0.595
	meanDiff	100	0.597
<i>Highly correlated SNPs removed; Covariates not considered</i>			
wRF	GEMMA	500	0.582
	meanDiff	100	0.576
GBT	GEMMA	2000	0.589
	meanDiff	25	0.582
NB	GEMMA	250	0.575
	meanDiff	25	0.563
KNN	GEMMA	3000	0.573
	meanDiff	25	0.593
<i>No SNPs removed for LD; Covariates added to model</i>			
wRF	GEMMA	500	0.600
	meanDiff	10	0.593
GBT	GEMMA	500	0.586
	meanDiff	25	0.591
NB	GEMMA	250	0.579
	meanDiff	25	0.568
KNN	GEMMA	2000	0.606
	meanDiff	10	0.609
<i>Highly correlated SNPs removed; Covariates added to model</i>			
wRF	GEMMA	750	0.590
	meanDiff	100	0.591
GBT	GEMMA	500	0.604
	meanDiff	25	0.593
NB	GEMMA	250	0.579
	meanDiff	25	0.568
KNN	GEMMA	2000	0.606
	meanDiff	10	0.609

wRF = weighted random forest; GBT = gradient boosted trees; NB = Naïve Bayes; KNN = K nearest neighbors; AUC = Area under the ROC curve

Table 6.3. Highest performing models in holdout validation for prediction of ACLR in Labrador Retriever dogs

Model	Feature Selection	No. SNPs	AUC
<i>No SNPs removed for LD; Covariates not considered</i>			
wRF	GEMMA	3000	0.662
	meanDiff	1000	0.692
GBT	GEMMA	4000	0.599
	meanDiff	100	0.571
NB	GEMMA	500	0.784
	meanDiff	750	0.669
KNN	GEMMA	250	0.735
	meanDiff	12500	0.763
<i>Highly correlated SNPs removed; Covariates not considered</i>			
wRF	GEMMA	10	0.635
	meanDiff	500	0.678
GBT	GEMMA	2000	0.733
	meanDiff	750	0.799
NB	GEMMA	250	0.784
	meanDiff	10000	0.630
KNN	GEMMA	500	0.733
	meanDiff	4000	0.724
<i>No SNPs removed for LD; Covariates added to model</i>			
wRF	GEMMA	5	0.677
	meanDiff	1000	0.692
GBT	GEMMA	4000	0.745
	meanDiff	10000	0.764
NB	GEMMA	500	0.784
	meanDiff	750	0.669
KNN	GEMMA	5000	0.752
	meanDiff	500	0.763
<i>Highly correlated SNPs removed; Covariates added to model</i>			
wRF	GEMMA	25	0.655
	meanDiff	10	0.675
GBT	GEMMA	500	0.745
	meanDiff	3000	0.808
NB	GEMMA	250	0.784
	meanDiff	10000	0.630
KNN	GEMMA	250	0.774
	meanDiff	500	0.763

wRF = weighted random forest; GBT = gradient boosted trees; NB = Naïve Bayes; KNN = K nearest neighbors; AUC = Area under the ROC curve

Ensemble Learning

Ensemble learning did not result in gains in performance when compared to base learners in 10-fold cross validation (Table 6.4). Overall, the best performing model was supervisory learning using logistic regression and base models that were trained using 250 SNPs. When the model was extended to holdout validation, there was some improvement in predictive performance, and the best performing model was supervisory learning using logistic regression and base models trained using 500 SNPs (Table 6.5).

Table 6.4. Highest performing ensemble models in 10-fold cross validation

Ensemble	n	No. SNPs	AUC
nAgreement	5	2000	0.602
GLM	N/A	250	0.612
RF	N/A	5000	0.609

GLM = supervisory learning with logistic regression; RF = random forest; n= agreement threshold; AUC = area under the ROC curve

Table 6.5. Highest performing ensemble models in holdout validation

Ensemble	n	No. SNPs	AUC
nAgreement	3	10000	0.790
GLM	N/A	500	0.829
RF	N/A	3000	0.714

GLM = supervisory learning with logistic regression; RF = random forest; n= agreement threshold; AUC = area under the ROC curve

Covariate Analysis

For covariate analysis, a training set of 237 Labrador Retrievers (n=99 cases and n= 138 controls) was used to train a logistic regression model using sex, neuter status, and body weight as predictors. Prediction using a test set of 99 dogs (n=35 cases and n=64 controls) was moderately successful (AUC = 0.630).

Discussion

This work demonstrates that it is feasible to predict ACLR using SNP data from dogs within the Labrador Retriever breed with a sufficient sample size. For all models, the best predictions were achieved when covariates were considered in the analysis. This is reasonable, as the heritability of ACLR in dogs has been estimated between 0.3 and 0.5 [Nielen 2001; Wilke 2006; Baker 2017], which means a substantial proportion of variance for ACLR is explained through environmental effects. When the genomic profile is considered alone, the maximum AUC that can be achieved in a classifying algorithm is dependent upon heritability of the trait and disease prevalence. As the disease prevalence of ACLR in the Labrador Retriever is 0.0579 [Witsberger 2008], the maximum achievable AUC in a model that explains 100% of genetic variance, assuming a heritability of 0.4, is 0.86 [Wray 2010]. Given our relatively small sample size and prior evidence supporting the hypothesis that ACLR is highly polygenic [Baird 2014; Baker 2017], it is unlikely that we can explain 100% of genetic variance, and therefore, while the AUC we were able to achieve using SNP data alone appears relatively poor, it is good given the heritability and prevalence of ACLR in the Labrador Retriever population. Notably, the maximum AUC that can be achieved with a genomic profile that explains one quarter of genetic variance is 0.69, which is much closer to the estimates achieved in this exploratory analysis.

Genomic prediction using Bayesian regression was overall poorly predictive in leave-one-out cross validation. Of note, Bayesian ridge regression, the only tested model that does not place a higher mass at zero under the assumption that few SNPs will have an effect on the trait, performed better than the other Bayesian models. This provides some additional evidence that ACLR in the Labrador Retriever has an underlying genetic architecture that is highly polygenic.

Since canine SNP data is highly correlated due to extensive within breed LD [Sutter 2004; Karlsson 2008], many SNPs that are highly correlated offer the same information to the model. Through the use of LD pruning, highly correlated SNPs are removed from the feature

set, thereby allowing for a greater number of unique SNPs to be considered in the model. LD pruning with a larger window or by the variance inflation factor did not contribute to consistently increased prediction accuracy. We found use of LD pruning of SNPs in feature selection improved the prediction accuracy of some models and required fewer SNPs to reach peak prediction accuracy. Since the LD pruning was performed as a preprocessing step for 10-fold cross-validation, it was done before splitting of the folds. To ensure that these results were not a product of test set overfitting, holdout validation was also performed with and without LD pruning. The holdout validation results corroborated that LD pruning increases model performance as well as the ability for NB, kNN and GBT to predict ACLR with reasonable accuracy after LD pruning. Therefore, future models for genomic prediction in dogs should include LD filtering as part of data quality control.

Our machine learning approach implemented selection was based solely on univariate filtering methods. It is notable that using mean difference for SNP selection allowed similar performance to be achieved using fewer SNPs than ranking via linear mixed model GWAS analysis (GEMMA). This difference in feature selection performance is likely influenced by sample size, as GWAS using 90% of the dataset may have been underpowered. An additional point is that, by definition, mean difference chooses SNPs where there is a larger difference between cases and controls, and so it is logical that ranking SNPs in this way would be an ideal method to choose SNPs for case-control classification.

Overall, all models were poorly predictive of ACLR in 10-fold cross validation. Although there were individual folds or specific test sets where prediction accuracy was improved, when the averages of all folds were taken together, these outliers were masked by poor performance of the other folds. This indicated that it was possible to select SNPs for greater predictive performance with specific testing and training sets, but refinement of feature selection is needed. More sophisticated methods, such as wrapper or embedded methods with multivariate analysis of feature relationships may provide greater performance for selecting consequential

features. In addition, other modeling techniques, such as recursive backwards elimination of SNPs based upon gain and mean decrease in accuracy in GBT and random forest, respectively, may provide another method for identifying informative SNPs. In addition, there are many tuning parameters for LD pruning in PLINK and for the ensemble learning methods that could further improve predictive performance. However, without larger data sets, it is difficult to determine a broadly applicable set of tuning parameters that does not lead to overfitting on the current data set. Further work on identifying a set of broadly informative features (SNPs) would also advance biological knowledge of the ACLR disease mechanism through review of associated genes.

Genomic prediction of a small independent test set of dogs showed good prediction using the GBT and NB algorithms. This result was enhanced when highly correlated SNPs were removed from the dataset, achieving genomic prediction >75%. Genomic prediction with >75% accuracy is approaching clinical relevance and could enable selective breeding of dogs and can identify individual dogs with high genetic risk for which preemptive clinical management of environmental risk could be a focus.

When a model considering covariates alone was considered, prediction accuracy was reduced. However, considering covariates along with SNP genotype enhanced the predictive capability of the models considered. Two of the ACLR risk factors that were included in this study are modifiable variables (dog weight and whether a dog was neutered). Ideally, a genomic prediction algorithm would identify high-risk dogs without these variables, so that clinical action could be taken to reduce risk. For example, the link between neutering and ACLR may only refer to dogs who are neutered before one year of age, which is common clinical practice [Torres de la Riva 2013]. Neutering could then be delayed for dogs at high risk of ACLR. Age of neutering was not recorded for the present data. A similar approach could apply to taking extra care to maintain a healthy weight. This is an important consideration for future models, which

should try to capture as much genetic variance as possible so the model will rely less on covariates for predictive accuracy.

In this study, adding an additional decision-making layer through an ensemble learning approach did not lead to an appreciable gain in predictive performance. This is reasonable, as the underlying base models which generated the predictions used as features in the ensemble method did not substantially differ among themselves, so the aggregate of their predictions was not likely to provide a better estimate than the base model predictions alone.

There were several limitations to this study. The sample size used for this research limits the predictive capacity of the models tested, especially when applied to Bayesian regression where sample sizes in the thousands are often needed to accurately estimate SNP effects [de los Campos 2013]. Although model prediction accuracy for ACLR may be clinically significant in Wisconsin Labrador Retrievers, increasing sample size and improving feature selection may further improve performance and validate use of classification-based data mining methods for ACLR prediction within the breed.

In conclusion, genomic prediction of ACLR risk in the Labrador Retriever breed can be achieved with clinically relevant accuracy. This manuscript comprises the first attempt at such a feat. Future prediction models in dog populations should use a training set with a large sample size, implement LD pruning as a part of data quality control, and mean difference in feature selection. A prediction model for ACLR in dogs would allow for selective breeding against ACLR and also provide the opportunity for a precision medicine approach to clinical management of high-risk dogs. One goal of this research would be to develop generalized models that can accurately predict ACLR in all high-risk breeds, such as the Labrador Retriever, Rottweiler, and Newfoundland [Witsberger 2008]. Genomic prediction across ancestral populations (breeds) is likely to be much more challenging. The ultimate goal of this work is to develop the dog as a spontaneous disease model for human ACLR research. This work comprises a part of that goal, as the ability to accurately assess genetic risk for ACLR in the dog would also provide

opportunities for clinical trials of disease-modifying therapy that would benefit both canine and human health.

Supplemental Information

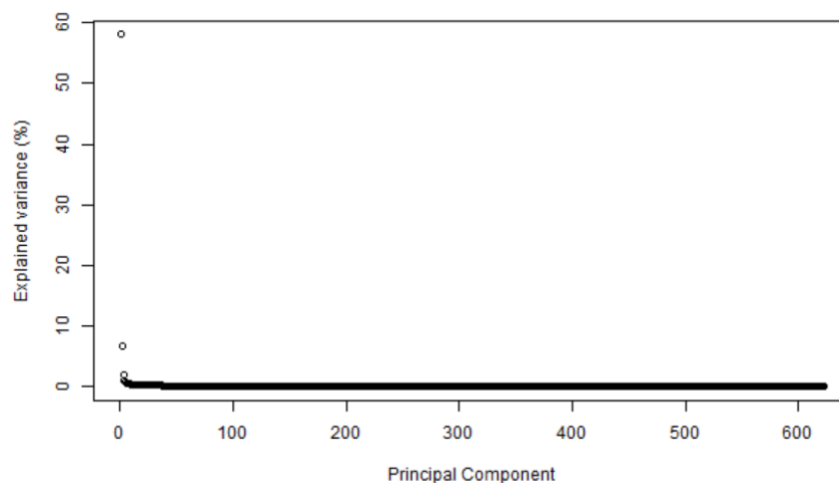


Figure 6.1S. Percentage of explained variance by each principal component computed using the genetic relationship matrix of the combined Wisconsin and Hayward et al. (2016) datasets.

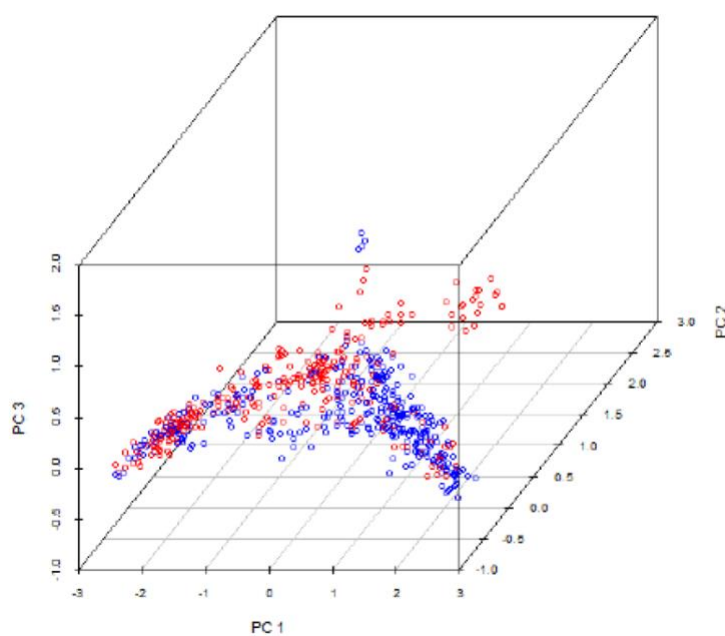


Figure 6.2S. Plot of the first three principal components computed using the genomic relationship matrix of the combined Wisconsin and Hayward et al. (2016) datasets.

Chapter 7

Imputation of Labrador Retriever Genotypes Using a Multi-breed Reference Panel

Introduction

For many years, the most popular single nucleotide polymorphism (SNP) panel for genotyping in dogs has been the Illumina CanineHD Whole-Genome Genotyping BeadChip, which contains approximately 174,000 (recently updated to 220,000) SNPs evenly spaced across the genome. In 2018, ThermoFisher released the Axiom Canine HD array that offers genotypes from approximately 710,000 SNPs. A higher density panel is advantageous, especially for more outbred populations such as the Labrador Retriever, where linkage disequilibrium (LD) is not as extensive, and could therefore benefit from improved SNP coverage of the genome. As individual whole-genome sequencing becomes increasingly accessible, canine reference populations will become available, and would offer millions of SNPs for use in genetic research.

Research funding is limited and thus it is important to balance cost-effectiveness with state-of-the-art methods. Genome-wide association studies (GWAS) require large sample sizes to identify disease associations, especially when the effect of individual mutations is small. SNP panels remain the most cost-effective solution for genotyping large numbers of animals for GWAS. Genotype imputation is a computational method that predicts large numbers of genetic variants from a smaller subset of known genotypes [Li 2009], providing the opportunity to capture additional genetic diversity while keeping genotyping costs to a minimum. GWAS SNPs can be imputed to a higher density using an appropriate reference population and any of a number of available algorithms developed for this purpose. The following describes validation of imputation of Labrador Retriever data from the Illumina CanineHD BeadChip (~170,000 markers) to the Axiom HD Canine Array (>700,000 markers) using a multiple breed reference population and the software program Beagle 5.0 [Browning 2016].

Methods

Imputation validation utilized whole-genome sequencing (WGS) data from n=22 Labrador Retrievers. 173,662 SNPs, present on the Illumina Canine HD BeadChip, were extracted from WGS data to create a test set for the imputation method. The reference panel used for imputation consisted of n=646 dogs of multiple breeds (Table 7.1). A multi-breed reference panel was chosen based on results from Friedenberg et al. (2016), which showed peak concordance with a multi-breed reference compared to a breed-specific reference panel. Each reference dog was genotyped using a pre-commercial version of the ThermoFisher Axiom Canine HD Array, and were genotyped at 992,360 unique SNP markers. Predictive phasing of the reference data was performed using Beagle 5.0 and default settings. The reference SNPs were also extracted from the WGS data for the 22 test Labrador retrievers, to be used for assessing accuracy of the imputation.

Table 7.1. Number of dogs in each breed included in the multibreed reference panel used for imputation.

Breed	Number of dogs
Afghan Hound	9
Akita	12
Basset Hound	11
Beagle	12
Bichon Frise	23
Border Collie	36
English Bulldog	12
Cavalier King Charles Spaniel	12
Chow Chow	12
Doberman Pinscher	12
German Shepherd	12
German Shorthaired Pointer	11
Golden Retriever	58
Great Dane	12
Irish Water Spaniel	11
Italian Greyhound	12
Labrador Retriever	96
Leonberger	22
English Mastiff	12
Miniature Schnauzer	47

Miniature Poodle	12
Newfoundland	20
Old English Sheepdog	12
Pembroke Welsh Corgi	12
Pomeranian	12
Portuguese Water Dog	12
Rottweiler	12
Saluki	11
Scottish Terrier	12
Shetland Sheepdog	10
Siberian Husky	12
Soft-coated Wheaten Terrier	10
Staffordshire Bull Terrier	10
West Highland White Terrier	32
Yorkshire Terrier	12
Total	646

The test dataset was imputed using Beagle 5.0 with the multibreed reference, a window size of 3 cM with a 1 cM overlap, and effective population size of 100. Default values were used for all other parameters. The effective population size of the Labrador Retriever was based on results from two studies, one that states the effective population size is 114 [Calboli 2008], and another more recent study that states the effective population size is 82 [Wiener 2017]. It is important to note that both of these studies were performed in populations of Labrador Retrievers in the United Kingdom, while the Labrador Retrievers in this study were derived from the United States; there is not a published estimate of effective population size for U.S. Labrador Retrievers.

To evaluate accuracy, imputed bi-allelic genotypes from the 22 Labrador Retrievers with WGS data were compared to bi-allelic genotypes extracted from WGS data. If the complete imputed genotype matched the complete WGS genotype, the SNP genotype was scored as correct. Accuracy of imputation was calculated per chromosome as number of genotypes imputed correctly divided by the number of genotypes compared.

Results

Accuracy of the imputation on a per chromosome basis is reported in Table 7.2. Overall, Accuracy exceeded 90% for all autosomes, and the vast majority of autosomes (36/38) achieved accuracy of 96% or higher. After validation, Labrador Retriever datasets were imputed using Beagle 5.0 and the above parameters.

Table 7.2. Accuracy of imputation of Labrador Retriever SNPs based on comparison to whole-genome sequencing data.

Chromosome	Number of SNPs compared	Imputation Accuracy (%)
1	39255	97.6
2	26006	97.7
3	31771	97.4
4	30628	97.6
5	29856	97.6
6	24688	97.3
7	27319	97.8
8	24316	97.0
9	18341	97.0
10	21985	97.3
11	23202	97.5
12	25984	97.6
13	22465	97.8
14	20611	97.4
15	20975	97.4
16	19422	97.3
17	21960	97.6
18	17797	97.0
19	18960	97.4
20	18045	97.3
21	16702	97.1
22	22044	97.4
23	18252	97.7
24	16110	97.4
25	17367	97.3
26	12405	96.4
27	15958	97.4
28	14206	97.3
29	14974	97.3
30	14199	97.1
31	8935	94.4
32	9000	91.2
33	11564	97.0
34	14642	97.5
35	10504	97.7
36	11751	97.2
37	11079	97.2
38	9744	96.2
X	14864	85.5

Discussion

We were able to successfully impute our Labrador Retriever dataset, originally genotyped using the Illumina Canine HD BeadChip with 174,000 SNP markers, to a higher density array of 858,657 SNP markers. Using a test set of 22 Labrador Retrievers, concordance between imputed SNP genotypes and SNP genotypes derived from WGS data was higher than previously reported results in another canine population [Friedenberg 2016].

Using multiple breeds as a reference instead of a single breed that is matched to the dataset seems counterintuitive. Previous work on imputation of canine SNPs found higher concordance with use of a multibreed reference panel [Friedenberg 2016], and this is consistent with imputation performed in other species [Rowan 2019]. Because alleles with low minor allele frequency (MAF) in the targeted breed may have higher MAF in another breed, including information from multiple breeds can improve imputation of those alleles in the target breed [Li 2011a, Rowan 2019]. Another benefit is the ability to use an overall larger reference population when data from multiple breeds is included in the reference. Because accuracy using a multibreed reference was acceptable, we did not attempt to impute our data using only Labrador Retrievers as a reference, and so cannot speculate further on whether this effect holds true with the current dataset.

The vast majority of autosomes had very high concordance (>96%). Three chromosomes (31, 32, and X) imputed with lower accuracy. The X chromosome had the lowest accuracy at 85.5%. Lower accuracy of imputation in the X chromosome was also found in Friedenberg et al. (2016). We did not evaluate the reasons for decreased accuracy in these 3 chromosomes. Notably, chromosomes 31 and 32 had fewer SNPs compared than many of the other chromosomes, indicating that they may have had lower coverage in the WGS data. Imputation in the X chromosome is notoriously difficult and may require special methods to improve accuracy [König 2014].

This method of imputation was applied to our full Labrador Retriever dataset and will be used for future research in the Comparative Orthopaedic and Genetics Laboratory at the University of Wisconsin School of Veterinary Medicine.

Chapter 8

Effect of Breed Linkage Disequilibrium Structure on Linkage Disequilibrium Score Regression for Interpretation of Canine GWAS

Abstract

Genomic inflation may exist in genome-wide association (GWA) data due to population stratification, which is a source of type 1 error, or it may be present due to polygenic inheritance, which is not a source of error. Current methods to detect and correct for genomic inflation do not distinguish between these two sources, attributing nearly all genomic inflation to error. Linkage disequilibrium (LD) score regression is a method used to distinguish between polygenicity and unaccounted population substructure when inflated test statistics are observed in a GWAS. This method has been applied to GWAS test statistics to show that polygenicity may be responsible for the majority of genomic inflation when the studied trait is expected to be polygenic. The purpose of this study was to evaluate whether LD score regression could be applied to GWAS performed in canine populations, where population substructure is known to exist, and high quality breed-specific genomic reference populations are not available. Whole-genome sequencing (WGS) data was used to calculate LD scores separately for three breeds; one that matched the population used for GWAS, a closely related breed, and a distantly-related breed. When a matching breed reference was used, LD score regression attributed the majority of inflation to polygenicity. LD score regression attributed more genomic inflation to bias when scores were calculated from closely related and distantly-related breed references. These results support the utility of LD score regression in canine GWAS and highlight the need for breed-specific reference populations in future work.

Introduction

Many diseases, including common cancers and auto-immune diseases, can be described as complex traits where disease risk is influenced by inherited genetic factors as well as environmental variables. From a genetics perspective, complex traits tend to be highly polygenic. Over the past decade, genome-wide association studies (GWAS) have been successful in identifying many of the genetic variants that influence development of complex trait diseases, and there is evidence to suggest that hundreds or even thousands of individual variants are likely playing a role [Visscher 2017; Moser 2015]. The effect sizes of the vast majority of these variants are very small to modest and act in combination to influence overall risk of disease [Robinson 2014]. These small effect sizes can be difficult to detect in GWAS, and this is typically remedied with improved study design and very large sample sizes.

An important step in GWA analysis is correction for bias due to population stratification. Population stratification occurs when both disease prevalence and allele frequencies differ among subpopulations that exist in a GWAS dataset, often when researcher performing the analysis may not be aware of the differences. This can lead to identification of spurious associations that appear to be linked to disease risk but are merely due to the presence of population structure, or bias in the dataset [Pritchard 2001]. The presence of bias can be evaluated visually by plotting observed and expected test statistics on a Q-Q plot, or quantitatively through calculation of genomic inflation factor (λ). The assumption underlying these tests is that few variants are truly associated with the tested trait, so the vast majority of test statistics should adhere to the null distribution. When they do not, this “genomic inflation” of the test statistics is corrected through any of multiple available methods [Wu 2011]. However, a problem exists when considering variant associations with a trait that is expected to be polygenic. When hundreds to thousands of variants are expected to be truly associated with a trait, this polygenic signal will look nearly identical to genomic inflation due to bias [Yang 2011b]. This can lead to overcorrection of test statistics for error in the dataset that does not truly exist,

further decreasing statistical power to detect small effect variants associated with complex trait disease.

Linkage disequilibrium (LD) score regression is a method that can be used to distinguish between polygenicity and unaccounted population substructure when inflated test statistics are observed in a GWAS. It is based on the idea that variants in LD with more genetic variation are more likely to tag small effect causal variants, and will have inflated test statistics that reflect this, whereas inflation due to population substructure only will not correlate with LD [Bulik-Sullivan 2014]. Briefly, LD scores are calculated for each SNP in the dataset using a reference panel of whole-genome sequenced individuals. The LD score measures the amount of genetic variation that is tagged by the SNP. When GWAS statistics are regressed against the LD scores of the SNPs in the GWAS, the LD score intercept minus one represents the average contribution of bias to inflation in the dataset [Bulik-Sullivan 2014]. This method has been applied successfully in multiple GWAS for complex trait diseases, which have confirmed that polygenic signal accounts for the majority of genomic inflation in complex trait GWAS [Okbay 2016; Hyde 2016].

Over the past several years, the dog has emerged as an excellent model organism to study the genetic contribution to spontaneous complex diseases. For many diseases, presenting clinical signs, pathogenesis, and treatment are extremely similar between humans and their canine counterparts [Shearin 2010]. Selective breeding for visual and behavioral characteristics has also inadvertently selected for heritable diseases [Karlsson 2008]. This selective process has created long ranges of DNA in LD that likely harbor disease risk variants [Karlsson 2008]. This combined effect allows GWAS to be performed using fewer SNP markers and smaller sample sizes than would be required to perform the same experiment in human populations. Discoveries made in the dog can then be used to inform candidate gene studies in human populations.

Individual dog breeds are maintained as isolated populations, and LD structure varies widely between them [Gray 2009]. Within breeds, popular sire effects, line breeding, and selection for within-breed behavioral or physical characteristics also have an effect on population substructure [Björnerfeldt 2008]. Because of this, careful attention has been paid to correcting, and potentially overcorrecting, for apparent population stratification in canine GWAS studies. LD score regression could be usefully applied to canine GWAS results to distinguish between inflation due to bias or polygenic effect. Unfortunately, the current canine genome (CanFam3.1) is based on a single individual Boxer dog, and publicly available, sufficiently dense genetic reference information for other breeds does not currently exist. This means LD score regression is out of reach for the vast majority of canine GWAS that have used other breeds, because the reference used to calculate LD scores and the population targeted in the GWAS should have the same LD structure [Bulik-Sullivan 2014]. It is unclear, however, how differing LD structure between breeds may affect LD scores of SNPs and whether groups of related breeds may use the same reference, or if breed-specific references are needed.

The purpose of this study was to evaluate the utility of LD score regression for interpretation of genomic inflation present in canine GWAS results using whole genome sequencing (WGS) data from three dog breeds with differing LD structure; the same breed used for GWAS (Labrador Retriever - LR), a closely related breed (Golden Retriever - GR), and a distantly-related breed (Rottweiler - RW). We hypothesized that the LD score intercept would vary when calculated with references from each breed, and that breeds with smaller effective population size and greater LD distance (GR, RW) would attribute a greater proportion of genomic inflation to polygenic effect. We discovered that while the LD score intercept did vary by breed, more genetic variance was attributed to bias when breeds other than the GWAS breed were used to calculate the LD scores. This result highlights the need for high-quality breed-specific reference populations for canine genetic research.

Methods

Data for case-control GWAS was initially collected for Baker et al. (2017). Briefly, this dataset consists of 367 purebred Labrador Retriever dogs that have been phenotyped for anterior cruciate ligament rupture (ACLR). A previously published GWAS using 237 dogs from this dataset confirmed ACLR is a highly polygenic complexly inherited disease in dogs [Baker 2017]. An additional 130 dogs had been added to the dataset since publication. Dogs were genotyped using the Illumina CanineHD BeadChip, which has ~174K SNP markers evenly spaced across the canine genome. Standard quality control [Baker 2017] and GWA analyses were performed using PLINK [Chang 2015]. For the purpose of this study, a chi-square test was used for GWA analysis. This test was chosen because it does not attempt to correct for population structure in the dataset, ensuring genomic inflation would be present whether due to polygenic signal or bias.

Illumina paired-end short read whole genome sequencing was performed on dogs of each breed group and aligned to the CanFam3.1 reference genome. Standard quality control measures were applied in accordance with the standards set by the Dog 10K Genomes Project. The Dog 10K Genomes Project is an international effort established to compile high quality whole genome sequencing data from 10,000 dogs. In total, n=8 LR, n=6 GR, and n=6 RW were recruited and sequenced. Data from an additional n=14 LR, n=14 GR, and n=4 RW were acquired from the Dog 10K Genomes project. Together, breed reference populations consisted of 22 LR, 20 GR, and 10 RW dogs. Genomic variant call format (gVCF) files for each dog were joint called separately for each chromosome using GenotypeGVCFs in the Genome Analysis Toolkit (GATK) software package [Van der Auwera 2013]. All variants except for biallelic variants were filtered out and files were converted to PLINK format using VCFtools [Danacek 2011].

LD scores were calculated separately for each breed with the LDSC software package [Bulik-Sullivan 2014] using a window size of 3Mb. This window size was chosen based on

recommendation from the authors of the LDSC software, stating that the window size used to calculate LD scores should cover the size of all or most real LD blocks in the genome. The average LD block size in the dog is approximately 2Mb, although this varies by breed with block size in some rare breeds reported at ~ 4Mb [Gray 2009]. Average LD decay over distance, defined as the genetic distance required for calculated r^2 to decrease to <0.2 , has been reported at 785kb in LR, and 1.4 Mb in GR [Gray 2009]. This measure has not been reported for RW, although the RW breed has a smaller population size and therefore is likely to have more extensive LD due to inbreeding [Gray 2009]. Therefore, a window size of 3Mb was chosen as a reasonable number to cover the average LD block size as well as LD blocks in upper end of the distribution.

LD score regression was performed separately using scores calculated from each breed reference and summary statistics from the above GWAS. LDSC automatically calculates heritability. To ensure heritability estimates calculated through LDSC were comparable to reported estimates, the regression was run with the option to correct to the liability scale using the reported population prevalence of ACLR in Labrador Retrievers (0.0579) [Witsberger 2008]. LDSC calculates both the LD score intercept and the genomic inflation factor (λ). These values were compared to determine how much genomic inflation the algorithm was attributing to polygenicity versus bias. $\lambda > 1.00$ indicates that genomic inflation is present in the dataset. The LD score intercept minus 1 is the average contribution from bias. Therefore, if $\lambda > 1.00$ and the LD score intercept is equal to or close to 1.00, the majority of inflation in the test statistics can be explained by polygenic effect.

Results

Genome-wide Association

Some of the Labrador Retrievers in the GWAS dataset also had available WGS data. These dogs (n=8) were used for the reference and were removed from the GWAS dataset,

leaving a total of 359 dogs for analysis. The final GWAS dataset included 145 cases and 214 controls. After quality control, 136,141 SNPs remained for GWA analysis. A Q-Q plot created from GWAS results confirmed evidence of genomic inflation in the dataset (Figure 8.1). The genomic inflation factor calculated from the GWAS test statistics was $\lambda=1.31$. Because the GWA analysis was performed primarily to create valid test statistics for LD score regression, it consisted only of a chi-squared test without correction for population structure or other confounding factors. This protocol is far below current standards for GWAS analysis. Therefore, while theoretically significant markers were identified, these markers were not evaluated further due to the high likelihood of spurious associations.

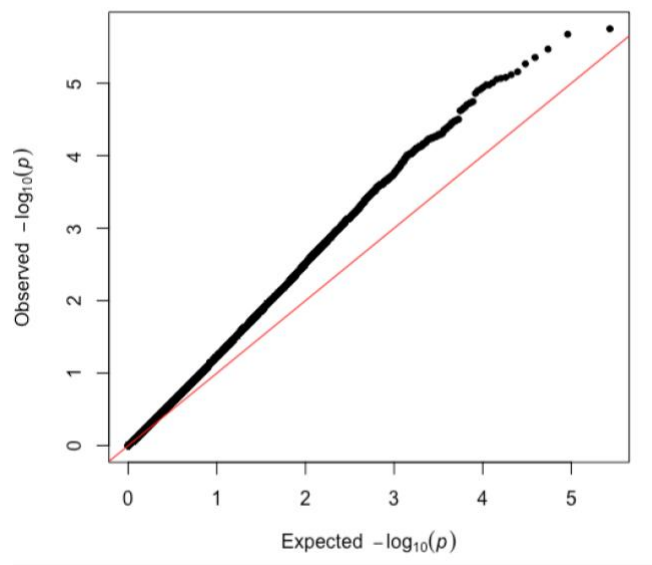


Figure 8.1. Q-Q plot of GWAS results without correction for population structure. The observed P-values consistently deviate from the expected line, indicating the presence of genomic inflation.

Linkage Disequilibrium Score Regression

LDSC automatically removes monomorphic SNPs from the reference data and only uses SNPs with minor allele frequency >5% for calculation of LD scores. In all cases, sufficient

numbers of SNPs remained for score calculation. As expected, LD score regression produced different LD score intercepts when different breed references were used (Table 8.1). The reported ratio is $(\text{intercept}-1)/(\text{mean}(\chi^2)-1)$, which measures the proportion of inflation in the test statistic that is attributed to causes other than polygenicity. When a breed reference that matched the GWAS population was used (LR), LD score regression attributed the majority of genomic inflation to polygenic effect. When a separate but closely related breed was used (GR), LD score regression attributed more genomic inflation to bias. When a distantly related breed was used (RW), the majority of genomic inflation was attributed to bias.

Table 8.1. Results of LD Score regression using different breed populations as references

Reference Breed	LD Score Intercept (SE)	Ratio (SE)
LR	1.04 (0.05)	0.11 (0.15)
GR	1.14 (0.07)	0.44 (0.22)
RW	1.28 (0.05)	0.88 (0.16)
$\lambda = 1.31$		

LR= Labrador Retriever; GR = Golden Retriever; RW = Rottweiler. LD score regression attributes less genomic inflation to polygenicity as breeds become more distantly related to the GWAS population. Genomic inflation factor (λ) is reported here for comparison.

Estimates of heritability calculated by LDSC were nonsensical. Heritability should be reported as a value between 0 and 1. For scores calculated with LR and GR references, reported heritability was >1.0 , with standard errors >1.0 . While the heritability estimated from the RW reference was a more reasonable number (0.48), the standard error was still >1.0 . This is likely due to the very small sample sizes used for GWAS and LD score calculation. The LDSC developers note in the online program documentation that heritability estimates are likely to be noisy with GWAS sample sizes <5000 [Bulik-Sullivan 2017].

Discussion

Population stratification is an important confounding factor that may lead to identification of spurious associations if it is not corrected for during GWA analysis. Current methods used to assess data for the presence of population stratification are based on the assumption that very few markers are likely to associate with the tested trait. This assumption is violated when GWAS is performed for a complex trait, which would generally be expected to be highly polygenic. This can lead well-meaning scientists to over-correct GWAS test statistics in order to account for bias that does not truly exist in the data, increasing type 2 error. LD score regression has been successfully used in human GWAS to distinguish between population stratification and polygenicity when genomic inflation of GWAS test statistics is present.

Purebred domestic dogs are an important model organism for studying the genetic contribution to complex trait diseases. The unique genetic architecture within and between purebred dog populations can make gene-mapping more efficient in the canine model system but can also be a potential source of bias. Evidence from several recent studies suggests that complex trait diseases in dogs are also polygenic [Baker 2017; Karlsson 2014; Tang 2014]. LD score regression could offer critical insight to these studies, all of which carefully corrected for inflation that was assumed to be due to population effects. Unfortunately, sufficiently dense breed-specific reference panels that would be necessary for LD score calculation are not currently available. The purpose of this study was to evaluate whether LD score regression could be applied in canine GWAS, and how genomic structural differences between breeds may affect the analysis. The sample sizes used in this study were far too small to yield robust, accurate estimates. Therefore, these results should be considered highly exploratory, and the reader should avoid over interpretation of conclusions from this work.

In the interest of avoiding overinterpretation, the potentially significant associations identified in the GWAS that was performed for the purpose of this study were not evaluated further. While currently accepted standards for GWAS may be based on some of the assumptions above, the presence of population structure in canine datasets is well-known

[Quignon 2007; Björnerfeldt 2008]. While future work should be careful to avoid overcorrection for apparent population effects, some correction for bias is likely to be necessary. Genotyping data that is intended for GWAS should undergo careful analysis to determine whether and the extent to which population substructure exists in the dataset. Corrections should then be applied on a case-by-case basis, rather than applying across-the-board correction that is based in assumption.

It should be noted that when LD score regression was calculated using a breed reference that matched the breed used for GWAS, the LD score intercept was close to 1.00, and the majority of inflation in the dataset was attributed to polygenicity. This result suggests it is possible our GWAS for ACLR in Labrador Retrievers [Baker 2017] was overcorrected for population effects. This is an interesting result, but for reasons stated above, it should be interpreted with caution. It would be worth confirming this result by performing a GWAS with data that is corrected for environmental variables as well as dataset-specific population structure. LD score regression should then be performed using LD scores that are calculated from a larger sample size that would provide more robust and believable estimates.

We hypothesized that LD score regression would attribute a greater proportion of genomic inflation to polygenicity when scores are calculated from breeds with smaller population size and longer ranges of LD. Instead, we found the opposite: as breeds become more evolutionarily distant from the LR, a greater proportion of polygenicity is attributed to bias. This result was at first surprising, but ultimately makes intuitive sense. The error that was made when formulating the above hypothesis was that LD blocks were assumed to be in generally the same locations, but more extensive in breeds with longer range LD compared to LR. Under this assumption, a SNP marker would tag more variation when it exists within a larger LD block. In reality, LD blocks are not necessarily in the same locations across breeds. A publication by Sutter et al. (2004) compares LD block structure between breeds of dogs and shows large differences in the same region across breeds, even between closely related LR and GR (see

Figure 2 in [Sutter 2004]). Therefore, a SNP in a fixed location could easily tag less variation when scores are calculated from a breed with differing LD structure. As a result, when LDSC sees a SNP with a large test statistic that is not tagging a correspondingly large amount of variation, it interprets the association as more likely due to bias than polygenic effect. In future work, this hypothesis could be confirmed by comparing LD scores of individual SNPs across breeds, and how those scores correlate with LD block structure. The results from the current analysis suggest that breed-specific reference panels that match the breed used for GWAS are absolutely necessary for LD score calculation.

An important source of error in this analysis is the small sample sizes that were used to calculate LD scores. Evidence of this error can be seen in the nonsensical heritability estimates from LDSC, which is likely a result of small sample size used for LD score calculation as well as GWAS. The numbers of dogs used for GWAS exceed numbers used in other successful canine GWAS [Karlsson 2014; Tang 2014], though accurate estimates of heritability often require larger sample sizes. Appropriate sample size and statistical power in genetic studies of inbred populations is an ongoing area of study in our laboratory. While sample sizes in future work should certainly be larger than those used for this exploratory analysis, the number of dogs necessary is unclear. It is worth noting that the developers of LDSC obtained accurate LD scores using a 1000 Genomes Project panel of several hundred individuals [Bulik-Sullivan 2014]. Therefore, it is likely that appropriate sample sizes are within the recruitment goals of the Dog 10K Genomes Project.

Overall, the results of this work indicate that LD score regression can indeed be usefully applied to canine GWAS for complex trait disease. Results of this work point to possible overcorrection of test statistics in published work [Baker 2017]. LD scores calculated from a closely related and distantly related breed had a large effect on results of LD score regression. This highlights the need for breed-specific reference panels in future work.

Chapter 9

Exploiting Biological Priors for Enhanced GWAS in a Dog Model of ACLR

Abstract

Anterior cruciate ligament rupture (ACLR) is a common condition that disproportionately affects young people, 50% of whom will develop knee osteoarthritis (OA) within 10 years of rupture. ACLR has both genetic and environmental risk factors. The genetic basis of ACLR remains unexplained. Spontaneous ACLR in the dog has a similar disease presentation and progression. Breed predisposition supports a genetic influence. The dog is a valuable genomic model for ACLR, as extensive linkage disequilibrium in dogs facilitates genome-wide association study (GWAS). Biologically relevant priors can be assigned in Bayesian mixture model (BMM) analysis to aid locus discovery. The objective of this study is to leverage the dog model with BMM analysis to identify novel and relevant genetic variants associated with ACLR. RNA sequencing was performed on ACL and synovium tissues from four ACLR affected and four matched control dogs. After correction for multiple testing, 186 and 374 differentially expressed genes (DEGs) were identified between ACLR case and control samples in ACL and synovium tissue, respectively. Biological priors were incorporated into GWAS analysis by assigning SNPs within differentially expressed genes to separate mixture classes using the BMM algorithm BayesRC. SNPs were also assigned to a separate class if they were within candidate genes that had been identified in published literature. Genome-wide association identified SNPs with the largest effects within or near genes associated with regulation of the actin cytoskeleton and the extracellular matrix. The results of the current analysis are consistent with previous work published by our laboratory and others, and also highlight new genes in biological pathways that have not previously been associated with ACLR. These results confirm the complex nature of ACLR and provide further support for the hypothesis that extracellular matrix proteins may play a role in predisposition to ACLR. Understanding the genetic basis of ACLR will identify genes or molecular pathways that can be targeted with novel or existing therapies to slow or prevent disease progression. The current investigation used the spontaneous dog ACLR model with a powerful statistical method and has identified relevant

genetic associations that mirror those found in human beings, laying the groundwork for development of disease-modifying therapies for both species.

Introduction

To date, the vast majority of genome-wide association studies (GWAS) have used single-marker models (e.g. linear mixed models) for statistical analysis. This is with good reason, as the linear mixed model (LMM) is relatively simple to perform, efficient, and well-suited to correcting for cryptic and known population structure in the dataset, which is a common statistical challenge in GWAS. Complex traits and diseases are widely regarded as polygenic, with a genetic architecture that implicates hundreds or even thousands of genome-disease associations [Robinson 2014]. Individually, these associations explain a very small proportion of phenotypic variance, and thus typically have at best moderate and more often weak associations with the phenotype of interest. These associations are not likely to meet the stringent P -value thresholds that are set after correction for multiple testing. Thus, only the single nucleotide polymorphisms (SNPs) with the largest effects on the phenotype may be identified, and together these SNPs explain a proportionally small fraction of estimated heritability, a phenomenon commonly referred to as the “missing heritability” problem [Génin 2019].

An alternative to this approach is to estimate the combined effect of all SNPs in the dataset using a Bayesian statistical approach. A Bayesian approach might treat all SNP associations as random effects drawn from a normal distribution, and allows for an unbiased estimate of variance explained by the SNPs [Moser 2015]. This approach can be tailored further to GWAS of complex phenotypes by treating SNP effects as drawn from a mixture of normal distributions corresponding to differing SNP effect sizes, including a distribution for SNPs with zero effect. BayesR [Erbe 2012; Moser 2015] is one such implementation that models SNP effects using four normal distributions with variance ranging from zero to 1% of total genetic variance, which more accurately models the effect size distribution expected from a complex phenotype. BayesR has been shown to be equal or superior to the LMM for both prediction modeling and QTL mapping [Moser 2015; Kemper 2015].

Another advantage of the Bayesian approach to GWAS is the ease with which prior biological information can be incorporated into the model [Stephens 2009]. Most single marker models, including the LMM, assume each SNP has an equal probability of having an effect on the phenotype of interest. In fact, SNPs that are within or near candidate genes may have a higher probability of being associated with the phenotype. Bayesian models allow the user to set a higher prior probability of affect to these SNPs. While there is some subjectivity to assigning prior probabilities, this is an improvement from the arguably arbitrary way biological knowledge is often used to interpret results after GWAS analysis [Stephens 2009; MacLeod 2016; Gallagher 2018]. BayesRC [MacLeod 2016] is a modification to BayesR to allow the incorporation of prior biological knowledge. To do this, SNPs are assigned to separate classes, defined by the user, based on whether the classes differ in the prior likelihood that they contain variants that are associated with the phenotype. This method improves the power and precision to detect associated variants when compared to BayesR [MacLeod 2016].

ACLR in the dog is an established genomic model for human ACLR. ACLR is generally regarded as a complex phenotype in both species with environmental as well as genetic risk factors. There are several advantages to the dog as a genomic model for ACLR that have been discussed previously [Baker 2017; Baker 2018], including higher disease prevalence [Witsberger 2008], established heritability of the disease [Nielen 2002; Wilke 2005; Baker 2017], and within breed homogeneity and extensive linkage disequilibrium (LD) [Karlsson 2008]. While GWAS have been performed [Wilke 2009; Baird 2014; Hayward 2016; Baker 2017; Baker 2018], the associations identified have not been repeatable from one study to the next. Our previous work on the genetic basis of ACLR in the Labrador Retriever [Baker 2017; Baker 2018] supports the hypothesis that ACLR is highly polygenic, and that most, if not all, SNP effects are relatively small. While we have identified some reasonable candidates, the majority of the identified associations do not have clear relevance to ACLR.

The purpose of this study was to incorporate knowledge of ACLR candidate genes with BayesRC GWAS to identify and prioritize genetic variants with clear relevance to the disease process and evaluate the repeatability of associations previously reported in the literature. To do this, we identified candidate genes through RNA sequencing and differential gene expression analysis as well as published literature and used them to assign SNPs to biological priors. We discovered associations in genes from molecular pathways that were not previously implicated in ACLR pathogenesis and repeated associations that have been previously reported in studies of ACLR in both human beings and dogs.

Materials and Methods

Data collection and phenotyping

All procedures were performed in strict accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health and the American Veterinary Medical Association and with approval from the Animal Care Committee of the University of Wisconsin-Madison (protocols V1070, V5463). Informed consent of each owner was obtained before participation in the study. Recruitment and quality control have been reported in previous publications by our laboratory [Baker 2017; Baker 2018]. Client-owned Labrador Retrievers were recruited from the University of Wisconsin-Madison UW Veterinary Care teaching hospital, online advertising, and through local and national breed clubs. If available, a pedigree was collected from each dog to confirm purebred status. All cases were diagnosed by a board-certified veterinary surgeon. In the vast majority of cases, a ruptured cruciate ligament was confirmed during stabilization surgery. Control dogs were over the age of 8 years [Reif 2003] with a normal orthopedic exam and stifle radiographs with no evidence of ACLR (joint effusion or osteophytosis) [Chuang 2014]. Age, weight, and whether the dog was spayed or neutered was recorded. DNA was extracted from blood or saliva. Dogs were genotyped using the Illumina Canine HD BeadChip (220,000 SNPs) and imputed to the higher

density Axiom Canine HD array (710,000 SNPs) using Beagle 5.0 [Browning 2018] and a multibreed reference panel containing 646 dogs of 35 breeds using the method outlined previously [Friedenberg 2016]. Quality control was performed using PLINK 2 [Chang 2015]. SNPs were removed from the dataset if they had minor allele frequency (MAF) <0.01, genotyping <90%, and did not conform to Hardy-Weinberg proportions at a *P*-value less than 1E-07. Because BayesRC does not tolerate missing genotypes, SNPs with any missing genotypes were also removed from the dataset.

RNA sequencing and differential gene expression analysis

Anterior cruciate ligament and synovium tissue biopsies were collected from 4 ACLR affected cases and 4 unaffected control dogs. It was important to examine both ligament and synovium, as the synovitis has been shown to precede ACLR in the dog [Bleedorn 2011] and may play a role in disease progression and development of osteoarthritis [Comerford 2011]. Cases and controls were matched as closely as possible on breed, sex, neutered status, age, and weight (Table 9.1). Medications that the dogs were taking at the time of sample collection were considered. Tissue from cases was collected during knee stabilization surgery. Tissue from unaffected control dogs was collected from dogs undergoing pelvic limb amputation or euthanasia for reasons unrelated to this study. Library preparation and sequencing was performed at the University of Wisconsin-Madison Biotechnology Center (Madison, Wisconsin). RNA sequencing libraries were constructed and 150bp paired-end sequencing was performed using the Illumina Hi-Seq 2500 sequencing platform. Read quality was evaluated using FastQC [Andrews 2010]. Reads were aligned to the CanFam 3.1 genome using STAR [Dobin 2013] and mapped paired-end reads were counted using RSEM (RNA-seq by expectation maximization) [Li 2011b]. Differential expression analysis was performed separately for each tissue using EdgeR [Robinson 2010; McCarthy 2012]. Significance was determined after adjusting *P*-values for multiple testing using a Benjamini-Hochberg correction [Reiner 2003]. Lists of differentially

expressed genes were submitted for pathway analysis using DAVID version 6.8 and high stringency [Huang 2009a; Huang 2009b].

Table 9.1. Affected case and matched unaffected control dogs used for RNA-sequencing

Cases				Matched Controls			
Breed	Sex	Age (yr)	Weight (Kg)	Breed	Sex	Age (yr)	Weight (Kg)
Golden Retriever	CM	8.83	30.5	Golden Retriever	CM	14.9	NA
Golden Retriever	CM	5.64	44.0	Golden Retriever	CM	3.9	34.0
Labrador Retriever	CM	9.70	36.0	Labrador Retriever	CM	12.7	28.5
Labrador Retriever	CM	13.27	36.0	Labrador Retriever	CM	13.5	35.0

CM = castrated male.

Association Analysis and Assignment of Biological Priors

The BayesRC method is described in [MacLeod 2016]. Briefly, a Markov Chain Monte Carlo (MCMC) approach is used to estimate SNP effects, using a mixture of four normal distributions including a zero-effect distribution. The three effect distributions are $N(0, 0.0001\sigma_{2g})$, $N(0, 0.001\sigma_{2g})$, and $N(0, 0.01\sigma_{2g})$, with σ_{2g} representing the additive genetic variance. A mostly uninformative Dirichlet prior ($\alpha = [1, 1, 1, 1]$) is used to define the prior proportion of SNPs in each distribution. As the algorithm runs, it uses the data to estimate the probability that each SNP belongs within distribution 1, 2, 3, and 4, and updates the proportions each iteration. To incorporate biological information, each SNP variant is assigned to a class (of 2 or more classes) where each class represents some biological information. For example, SNPs within DEGs from tissue A could be in one class, and SNPs within DEGs from tissue B could be in another class, and so on.

To incorporate information about genes relevant to the ACLR disease process, we defined three classes based on differential gene expression analysis: DEG from ligament, DEG from synovium, and DEG identified in both tissues. A fourth class represented candidate genes that have been reported in the literature to be associated with ACLR or tendinopathy in human or dog studies (Table 9.2). Because some Labrador Retrievers in the current dataset were

present in the datasets of our previously published work [Baker 2017; Baker 2018], associations identified in our previous studies were not included in the candidate gene class to avoid introducing bias. A fifth class was defined for SNPs that were not within DEGs or candidate genes. SNPs were assigned to classes if they were within the boundaries of a gene $\pm$ 25kb. The size of the flanking region was conservatively defined by calculating the average haplotype block size in our data using PLINK, which was 19.43kb with a maximum haplotype block size of 200kb. Gene boundaries were based on CanFam 3.1 from Ensembl release 97 using the python package PyEnsembl v. 1.7.5. The BayesRC algorithm was run for a total of 200,000 iterations with a burn-in period of 100,000 iterations. Final mean SNP effects were evaluated based on the absolute value of the reported SNP effect. SNP effects were assigned to genes if they were within the gene boundary $\pm$ 25Kb.

Table 9.2. Candidate genes for ACLR derived from human and dog studies

Gene	Human	Dog
<i>ACAN</i>	Mannion et al. (2013), Johnson et al. (2015)	Wilke et al. (2009)
<i>BGN</i>	Mannion et al. (2013)	
<i>COL1A1</i>	Khoschnau et al. (2008), Posthumus et al. (2009a), Ficek et al. (2013), Stepień-Słodkowska et al. (2013)	Baird et al. (2014a)
<i>COL3A1</i>	O'Connell et al. (2015), Stepień-Słodkowska et al. (2015a)	Baird et al. (2014a)
<i>COL5A1</i>	Posthumus et al. (2009b), Stepień-Słodkowska et al. (2015b)	Baird et al. (2014a)
<i>COL5A2</i>		Baird et al. (2014a)
<i>COL6A1</i>	O'Connell et al. (2015)	
<i>COL9A1</i>		Wilke et al. (2005)
<i>COL11A1</i>		Baird et al. (2014a)
<i>COL11A2</i>	Saunders et al. (2016)	
<i>COL12A1</i>	Posthumus et al. (2010); Ficek et al. (2014)	
<i>COL24A1</i>		Baird et al. (2014a)
<i>COLGALT1</i>	Kim et al. (2017)	
<i>DCN</i>	Mannion et al. (2013)	
<i>DIRC2</i>		Baird et al. (2014a)
<i>ELN</i>	Kouhry et al. (2015)	
<i>FBN1</i>		Baird et al. (2014a)
<i>FBN2</i>	Khoury et al. (2015)	

<i>FMOD</i>	Mannion et al. (2013)	
<i>GDF5</i>	Raleigh et al. (2015)	
<i>HIF1A</i>	Rahim et al. (2014)	
<i>IL1B</i>	Rahim et al. (2016)	
<i>IL6</i>	Lulinksa-Kuklik et al. (2019)	
<i>ITGB3</i>	Saunders et al. (2015)	
<i>KDR</i>	Rahim et al. (2014)	
<i>LOX</i>	Saunders et al. (2016)	Baird et al. (2014a)
<i>LTBP2</i>		Baird et al. (2014a)
<i>LUM</i>	Mannion et al. (2013)	
<i>MMP1</i>	Posthumus et al. (2012)	
<i>MMP3</i>	Collins et al. (2009), Posthumus et al. (2012)	
<i>MMP10</i>	Posthumus et al. (2012)	
<i>MMP12</i>	Posthumus et al. (2012)	
<i>NGFB</i>	Rahim et al. (2014)	
<i>RNF152</i>		Baird et al. (2014b)
<i>SEMA5B</i>		Baird et al. (2014b)
<i>SORCS2</i>		Baird et al. (2014b)
<i>TNC</i>	Collins et al. (2009), Gibbon et al. (2018)	
<i>VCAN</i>		Wilke et al. (2011)
<i>VEGFA</i>	Rahim et al. (2014)	
<i>WISP2</i>	Johnson et al. (2015)	
<i>ZDHHC23</i>		Baird et al. (2014b)

Results

Labrador Retriever Dataset

The final dataset included 397 (156 ACLR affected and 241 unaffected control) purebred Labrador Retriever dogs. Of these, 55 were intact males, 30 were intact females, 161 were castrated males, and 151 were ovariohysterectomized females. The distribution of neutering was significantly different between male and female dogs ($\chi^2 = 4.62$, $P = 0.03$). Quality control filtering removed 156,014 SNPs due to low MAF, 1,802 SNPs were removed because they did not adhere to Hardy-Weinberg equilibrium proportions, and 257,187 SNPs were removed for missing genotypes; 443,227 SNPs remained for analysis.

Differential Gene Expression Analysis

FastQC analysis determined that all samples were of good quality. After adjustment for multiple testing and without imposing a threshold for log fold change, we identified 200 genes

from anterior cruciate ligament tissue and 444 genes from synovium tissue that were significantly differentially expressed between case and control dogs (Supplementary Tables 9.1S and 9.2S). To ease interpretation of results, only transcript ID's that could be matched to a known gene were included in the assignment of biological priors. This left a total of 180 DEGs from ACL and 373 DEGs from synovium for prior assignment. There was no significant pathway enrichment identified for DEGs from ligament tissue. A significant annotation cluster was identified in DEG's from synovium tissue that included 141 genes that belonged to the UniProt Knowledgebase category "Transmembrane helix" ($P_{\text{adj}}=1.8\text{E-}06$) (Supplementary Table 9.3S).

Association Analysis

12,209 SNPs were assigned to biological priors based on DEG's identified through RNA sequencing and candidate genes (Table 9.3). SNP effects were averaged over five runs. Overall, an average of 3,728 SNPs (0.8%) had some estimated effect, with the remainder of SNPs assigned to the zero effect distribution. Enhanced GWAS results are visually represented in a Manhattan plot (Figure 9.1). On average, 37 SNPs were assigned to the $0.01\sigma_g$ distribution, 361 SNPs were assigned to the $0.001\sigma_g$ distribution, and 3,330 SNPs were assigned to the $0.0001\sigma_g$ distribution. The 50 largest SNP effects and their distance to genes are reported in Table 9.4.

Figure 9.1. Manhattan plot of genome-wide association analysis of ACLR case and control dogs using BayesRC.

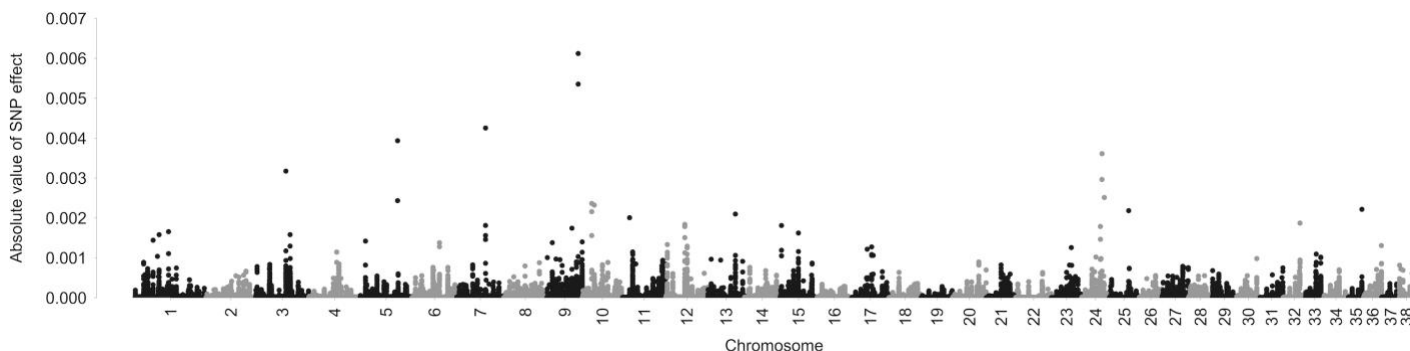


Table 9.3. The number of SNPs assigned to biological priors defined by differential gene expression analysis and candidate genes reported in the literature.

Class	Definition	Number of SNPs
1	DEGs from ligament	2,614
2	DEGs from synovium	7,850
3	DEGs from both ligament and synovium	703
4	Candidate genes	1,042
5	SNPs not assigned to biological priors	431,018

DEG = differentially expressed gene, SNP = single nucleotide polymorphism

Table 9.4. The top 50 SNP effects from Bayesian mixture model (BayesRC) association analysis

Chromosome:Location	Class	SNP Effect	Gene	Distance (bp)
chr9:53865770	1	0.006	<i>FNBP1</i>	9,144
chr9:53871457	1	0.005	<i>FNBP1</i>	3,457
chr7:49455960	5	-0.004	None	0
chr5:64359450	1	0.004	<i>ACSF3</i>	21,679
chr24:34970050	5	0.004	None	0
chr3:51975977	4	-0.003	<i>ACAN</i>	19,131
chr24:34842049	5	-0.003	<i>SULF2</i>	0
chr24:38868995	5	-0.003	Noncoding transcript	467
chr5:64356666	1	0.002	<i>ACSF3</i>	19,554
chr10:16199198	5	0.002	None	N/A
chr10:20536317	3	-0.002	<i>FBLN1</i>	0
chr35:23442178	2	0.002	<i>LRRC16A</i>	0
chr25:33171736	2	-0.002	<i>ADAM28</i>	23,934
chr10:16138674	5	0.002	ENSCAFG00000031351	19,408
chr13:47443664	4	-0.002	<i>KDR</i>	0
chr11:12024840	4	0.002	<i>LOX</i>	0
chr32:28999310	3	0.002	<i>RPL34</i>	4,462
chr32:28996275	3	0.002	<i>RPL34</i>	0
chr12:32821658	4	-0.002	<i>COL9A1</i>	0
chr7:49422990	5	-0.002	None	N/A
chr15:2674794	3	0.002	<i>ZMPSTE24</i>	0
chr24:32151336	4	0.002	<i>WISP2</i>	0
chr12:32792729	4	-0.002	<i>COL9A1</i>	0
chr9:43053330	1	-0.002	<i>FLOT2</i>	0
chr1:57856871	3	-0.002	<i>NEPN</i>	0
chr15:31879223	4	-0.002	<i>DCN</i>	0
chr3:59287831	4	-0.002	<i>SORCS2</i>	0

chr1:41687833	3	0.002	<i>AKAP12</i>	0
chr7:49415778	5	-0.002	None	N/A
chr10:16118097	5	0.002	None	N/A
chr12:32815853	4	-0.002	<i>COL9A1</i>	0
chr24:32144138	4	-0.002	<i>WISP2</i>	0
chr7:49426351	5	-0.001	None	N/A
chr1:31289631	3	-0.001	<i>ABRACL</i>	22,185
chr5:9055237	3	-0.001	<i>PKNOX2</i>	0
chr9:60820860	3	-0.001	<i>DAB2IP</i>	0
chr9:9241561	4	0.001	<i>ITGB3</i>	17,351
chr6:47617815	4	-0.001	<i>COL11A1</i>	0
chr12:2604649	4	-0.001	<i>HLA-DPB1</i>	
chr36:30472199	4	0.001	<i>COL3A1</i>	16,392
chr3:59282629	4	-0.001	<i>SORCS2</i>	0
chr12:36850837	4	-0.001	<i>COL12A1</i>	12,221
chr6:47604275	4	-0.001	<i>COL11A1</i>	0
chr17:34324882	3	-0.001	<i>ARID5A</i>	0
chr23:34306524	5	-0.001	<i>DZIP1L</i>	0
chr12:36852892	4	-0.001	<i>COL12A1</i>	14,276
chr12:36767540	4	-0.001	<i>COL12A1</i>	0
chr17:26672332	1	0.001	<i>FAM98A</i>	0
chr15:2655972	3	0.001	<i>COL9A2</i>	0
chr3:52046039	4	-0.001	<i>HAPLN3</i>	0

SNP = single nucleotide polymorphism. Class refers to the biological prior the SNP was assigned to (Table 9.3). SNPs with a distance of zero are located within the listed gene.

Discussion

Incorporation of external biological information using the Bayesian mixture model program, BayesRC, provides a more objective approach to prioritize SNPs based on biological probability of effect in GWAS analysis. This is in contrast to the subjective decisions that are often made when evaluating GWAS results [Thompson 2013]. Here, we were able to identify associations within or near many relevant candidate genes for ACLR. Many of the largest effect SNPs were within or near genes that were either differentially expressed between ACLR case

and control dogs, or were candidate genes that have been reported previously in both human and canine studies of the genetic predisposition to ACLR.

To assign biological priors, we first performed RNA sequencing in ligament and synovium tissue from ACLR affected and matched control dogs. In order to evaluate whether particular types of genes or pathways were represented among differentially expressed genes, DEG lists were analyzed using functional annotation in DAVID. While we did not identify any functions or pathways that were statistically significant in ligament, a large gene cluster denoting proteins with a transmembrane helical domain was identified from synovium. The significance of this cluster is unclear, as a large portion of genes in the genome encode a transmembrane structure [Sonhammer 1998], so it is perhaps unsurprising that many of the identified DEGs may belong to this category. The post-rupture intra-articular environment is incredibly dynamic, with increased cytokine levels of many types and their associated signaling cascades, which could easily lead to an increased expression of membrane proteins as a response. It is notable that many of the genes in this cluster are cytokine receptors. Cytokine signaling cascades have been implicated in the high rate of early-onset osteoarthritis following ACLR, which occurs regardless of whether joint stability was surgically restored [Cameron 1994, Marks 2005], and therefore these findings may indeed have some clinical significance. Future research warrants further investigation into the details of the genes that were significantly differentially expressed between ACLR cases and controls, and whether their differential expression may be related to a response unique to ACLR patients.

Association analysis using biological priors identified several regions of the genome that show association with ACLR in the dog model. A Manhattan plot of GWAS results identifies five regions with the largest effects on chromosomes 9, 7, 5, 3, and 24. The region on chromosome 7 is in a gene desert and may be highlighting a region of the genome that has some regulatory capacity. All remaining regions highlight a specific gene; *FNBP1* on chromosome 9, *ACSF3* on chromosome 5, *ACAN* on chromosome 3, and *SULF2* on chromosome 24. Of these 4 genes,

FNBP1 and *ACSF3* were differentially expressed in ligament, *ACAN* was a candidate gene, and *SULF2* was not included in any of the biological priors.

Formin binding protein 1 (*FNBP1*), which is also known as *FBP17*, is involved in regulation of the actin filament assembly for the actin cytoskeleton, which is important for cellular migration and the maintenance of cell shape [Higgs 2005, Apenstrom 2009]. Specifically, *FNBP1* has been reported to associate with actin regulatory proteins cortactin, dynamin, and Arp2/3 to enhance degradation of extracellular matrix and promote cellular migration (in breast cancer) [Suman 2018]. ACL fibroblasts are known to undergo cytoskeletal reorganization after a strain event to align in longitudinal orientation with the strain [Lee 2005]. Interestingly, this reorganization is also associated with an increase in production of extracellular matrix [Lee 2005]. In this study, *FNBP1* was expressed more highly in cases than controls, which suggests that extracellular matrix degradation may have been taking place along with actin reorganization. This process may be reflective of the proliferative and remodeling phases that occur in the ACL in response to injury [Murray 2000], which have also been reported in the dog [Hayashi 2004]. It could be that there is a heritable difference in response to injury between dogs that rupture their ACL and those that do not.

Acyl-CoA synthetase family member 3, (*ACSF3*), is an enzyme that converts malonic acid to malonyl-CoA, which is the first step in fatty acid synthesis. This protein is localized in mitochondria, and plays an integral role in maintaining mitochondrial metabolism, as malonate is an inhibitor to mitochondrial respiration [Bowman 2019]. In this study, *ACSF3* was expressed more highly in cases than in controls, which suggests that malonate may be produced as a result of ACL injury, necessitating increased expression of *ACSF3* for its clearance. Mutations in *ACSF3* result in malonic and methylmalonic aciduria, which is associated with developmental delay, cardiomyopathy, hypoglycemia, and seizures [Bowman 2019]. A connection between *ACSF3* or malonate and ligament or tendon homeostasis has not been reported.

Aggrecan (*ACAN*) is a large, aggregating proteoglycan that plays a vital role in maintaining moisture in the extracellular matrix of collagenous tissues, including ligamentous tissue. Maintenance of moisture in the extracellular matrix is important to provide space and lubricant between collagen bundles, to distribute force across the tissue, and to protect the tissue from injury and aid healing [Illic 2005]. An association between *ACAN* and ACLR has been reported multiple times in human [Mannion 2013; Johnson 2015] and dog [Wilke 2009]. Aggrecan has also been reported to play a role in degenerative suspensory ligament disease (DSLDD) in horses [Plaas 2011]. While an association between aggrecan and ACLR has been reported before in a GWAS from a subset of this dataset [Baker 2017], we chose to keep *ACAN* in the list of candidate genes because of its connection to degenerative ligament disease across species, and because the previously reported association in the Labrador Retriever was rather weak ($P=1.07E-04$). Because of this, care should be taken not to overinterpret this association, however, the strength of the association in this study as well as the body of evidence that exists to support it leads us to consider this association as additional evidence of a role for aggrecan in the pathobiology of ACLR.

Heparan sulphate 6-O endosulfatase 2 (*SULF2*) is an enzyme with selectivity for glucosamine 6-sulphate, and functions in degradation of sulfated macromolecules, such as proteoglycans [Morimoto-Tomita 2002]. *Sulf* enzymes are important for regulation of overall balance of cartilage matrix synthesis and degradation [Otsuki 2010]. Both *SULF1* and *SULF2* have been reported as overexpressed in osteoarthritic cartilage [Otsuki 2008], though they were not significantly differentially expressed in this study. *SULF2* was not assigned to a biological prior, and therefore this association is derived from genetic data only. *SULF1* and *SULF2* knockout mice are normal at birth but develop early-onset osteoarthritis in their knee joints at 6 months of age with reduced glycosaminoglycan content and lower cellularity in articular cartilage [Otsuki 2008]. Dogs with ACLR also develop early osteoarthritic changes that are

typically present at the time of diagnosis [Chuang 2014], and early-onset osteoarthritis is common in human ACLR patients [Lohmander 2007]. An association with *SULF2* and ACLR specifically has not been previously reported, therefore this gene is a novel candidate gene for ACLR.

While their effect sizes were smaller, several notable candidate genes are present in the 50 largest-effect SNPs from this analysis, including collagen genes *COL9A1*, *COL11A1*, *COL12A1*, *COL3A1* and *COL9A2*. All of these genes were biological priors in the candidate gene class, as various associations have been reported previously in human and dogs (Table 9.2). These associations have not been previously validated in either species. *COL9A2* was also differentially expressed in both ligament and synovium. *COL9A2* encodes the alpha-2 chain of type IX collagen, which is considered specific to cartilage. Other notable associations include several other genes that also have a role in actin cytoskeleton homeostasis, including *LRRC16A* (also known as *CARMIL1*) [Edwards 2013], *KDR* [Luykenaar 2009], *LOX* [Payne 2006], *FLOT2* [Langhorst 2007], *AKAP12* [Benz 2019], *ABRACL* [Wang 2019], *PKNOX2* (also known as *PREP2*) [Haller 2004], *ITGB3* [Urbinati 2012], and *SORCS2* [Deinhardt 2011], which has also been reported as associated with ACLR in Newfoundland dogs [Baird 2014b]. This pattern in the association results suggests that variable actin dynamics may play a role in genetic predisposition to ACLR.

In [MacLeod 2016], it is noted that the Dirichlet prior in BayesRC may have greater influence on the posterior if the number of variants in one class is low relative to the rest of the dataset. The authors suggested that classes should have 1000 variants or greater to allow the data to have strong influence on the posterior parameters, especially when there is greater uncertainty, for example, when candidate genes from reported literature are used for prior assignment. To avoid this, we made sure that >1000 SNPs were assigned to the candidate gene class used in this study (Class 4, Table 9.3). However, there were fewer than 1000 SNPs

in Class 3, which represented genes that were differentially expressed in both ligament and synovium tissues. This was known prior to BayesRC analysis, and the choice was made to maintain this class for two reasons: 1) these genes were differentially expressed in both ACL and synovium tissues, so there is inherently less uncertainty around their candidacy and 2) because they were differentially expressed in both tissues, it seemed important to designate them separately from genes that were differentially expressed in only one tissue. Ten of the top 50 SNP effects were assigned to Class 3, which is a relatively high number given $703/433,227 = 0.16\%$ of SNPs were assigned to Class 3, and these SNPs represent 20% of the top SNP effects. It is not clear whether the effect of these SNPs is due to true association with the disease, or potential bias from prior assignment, and these results should be interpreted with this in mind.

Conclusion

Incorporation of *a priori* biological information into Bayesian mixture model analysis using BayesRC in a dog model of ACLR was able to replicate some associations that were previously reported in human and dog studies, especially in collagen genes, and also identified novel genetic associations with ACLR. The actin cytoskeleton is the basis for cellular organization and shape and is integral for a cell's capacity to migrate. This is the first study to suggest a role for the actin cytoskeleton in risk of ACLR. Additionally, while *SULF2* has been implicated in onset and progression of osteoarthritis, which is an important sequela to ACLR, this is the first publication to report an association between *SULF2* and ACLR itself. These associations provide additional support for the hypothesis that the genetic basis of ACLR is complex and highly polygenic. Several associations reported in human studies have been reported here, in the dog, which highlights the value of One Health medicine [Gyles 2016], and the dog in particular as a valuable model for genomic research.

Supplementary Table 9.1S. Differentially expressed genes identified through RNA sequencing of anterior cruciate ligament tissue from ACLR case and control dogs

Ensembl ID	Gene Name	LogFC	P	Adjusted P
ENSCAFG00000000804	<i>FBLN1</i>	-4.865298521	1.90E-07	2.60E-03
ENSCAFG000000002533	<i>SEC61B</i>	1.322113110	6.27E-07	4.30E-03
ENSCAFG000000013476	<i>PDK3</i>	1.128768499	9.63E-06	2.13E-02
ENSCAFG000000000419	<i>AKAP12</i>	-3.44879101	9.71E-06	2.13E-02
ENSCAFG000000005334	<i>GPR180</i>	1.549014410	1.62E-05	2.13E-02
ENSCAFG000000008716		2.221296240	1.63E-05	2.13E-02
ENSCAFG000000000418	<i>MDM2</i>	1.134421201	1.74E-05	2.13E-02
ENSCAFG000000000931	<i>NUS1</i>	1.028586889	1.80E-05	2.13E-02
ENSCAFG000000003945	<i>DENND2A</i>	-1.763494852	1.84E-05	2.13E-02
ENSCAFG000000004017	<i>ZNF438</i>	-1.367547562	1.86E-05	2.13E-02
ENSCAFG000000010860	<i>ERGIC2</i>	0.839007356	2.01E-05	2.13E-02
ENSCAFG000000018117	<i>CHRD1</i>	-3.708275754	2.01E-05	2.13E-02
ENSCAFG000000000352	<i>XPOT</i>	1.038498199	2.02E-05	2.13E-02
ENSCAFG000000001402	<i>KDEL3</i>	1.964656167	2.98E-05	2.62E-02
ENSCAFG000000005283	<i>CLIP4</i>	-2.619954713	3.05E-05	2.62E-02
ENSCAFG000000014713	<i>REX1BD</i>	2.462479119	3.06E-05	2.62E-02
ENSCAFG000000008975	<i>THBS4</i>	-3.156416687	3.62E-05	2.67E-02
ENSCAFG000000013416	<i>MTX2</i>	0.979339824	3.79E-05	2.67E-02
ENSCAFG000000003663	<i>GSN</i>	-2.72132796	3.88E-05	2.67E-02
ENSCAFG000000000868	<i>SMOC2</i>	3.684567283	3.89E-05	2.67E-02
ENSCAFG0000000031126	<i>HSPA13</i>	0.983470216	4.12E-05	2.69E-02
ENSCAFG000000010549	<i>GLRX2</i>	1.143524606	4.59E-05	2.79E-02
ENSCAFG000000008713	<i>NUCB2</i>	1.183109508	4.69E-05	2.79E-02
ENSCAFG000000017626	<i>FAXDC2</i>	-1.046463422	5.34E-05	2.91E-02
ENSCAFG000000009210	<i>R3HCC1</i>	2.644837866	5.47E-05	2.91E-02
ENSCAFG000000008467	<i>EDIL3</i>	3.684285905	5.52E-05	2.91E-02
ENSCAFG000000017318	<i>EBF1</i>	-1.216185198	5.73E-05	2.91E-02
ENSCAFG000000005738	<i>SLC30A6</i>	1.023971250	6.38E-05	2.98E-02
ENSCAFG000000010871	<i>LRP5</i>	-1.516455192	6.46E-05	2.98E-02
ENSCAFG000000012072	<i>SEC22A</i>	1.039821053	6.68E-05	2.98E-02
ENSCAFG000000014698	<i>CRLF1</i>	2.617791261	7.00E-05	2.98E-02
ENSCAFG000000008236		2.580785888	7.16E-05	2.98E-02
ENSCAFG000000006124	<i>ALG5</i>	0.764205656	7.18E-05	2.98E-02
ENSCAFG000000004456	<i>NDUFB6</i>	0.818638387	7.56E-05	3.00E-02
ENSCAFG000000003679	<i>C15H16orf87</i>	1.111308853	8.24E-05	3.00E-02
ENSCAFG000000005875	<i>FAM98A</i>	0.885355314	8.61E-05	3.00E-02
ENSCAFG000000011286	<i>OSTC</i>	1.005118927	9.73E-05	3.00E-02
ENSCAFG000000001246	<i>PTCH1</i>	-1.520847408	1.00E-04	3.00E-02

ENSCAFG00000030025	<i>FGL2</i>	-2.175822637	1.02E-04	3.00E-02
ENSCAFG00000009847	<i>PCIF1</i>	-0.857346142	1.03E-04	3.00E-02
ENSCAFG00000014799		-1.555158201	1.05E-04	3.00E-02
ENSCAFG00000009254	<i>GALNT5</i>	1.670353869	1.07E-04	3.00E-02
ENSCAFG00000010582	<i>BOC</i>	-1.629248107	1.07E-04	3.00E-02
ENSCAFG00000007424	<i>SWAP70</i>	-1.137616406	1.08E-04	3.00E-02
ENSCAFG00000001550	<i>FOXRED2</i>	-2.399843154	1.09E-04	3.00E-02
ENSCAFG00000007260	<i>MTFR1</i>	1.072563043	1.10E-04	3.00E-02
ENSCAFG00000028869	<i>GPX8</i>	1.521969379	1.12E-04	3.00E-02
ENSCAFG00000009652	<i>TWF1</i>	0.877094457	1.13E-04	3.00E-02
ENSCAFG00000013708	<i>KANSL1</i>	-0.711070306	1.14E-04	3.00E-02
ENSCAFG00000012256	<i>GOLT1B</i>	1.230423645	1.14E-04	3.00E-02
ENSCAFG00000000163	<i>STAT6</i>	-0.833992182	1.17E-04	3.00E-02
ENSCAFG00000014687	<i>TMEM59L</i>	3.065963559	1.17E-04	3.00E-02
ENSCAFG00000032515	<i>ARPP19</i>	0.840004442	1.18E-04	3.00E-02
ENSCAFG00000006546	<i>ARID5A</i>	-1.434106604	1.20E-04	3.00E-02
ENSCAFG00000009732	<i>TSPAN2</i>	2.460603086	1.21E-04	3.00E-02
ENSCAFG00000002392	<i>ZSCAN18</i>	-1.720740163	1.24E-04	3.03E-02
ENSCAFG00000029673		0.818162127	1.28E-04	3.06E-02
ENSCAFG00000002816	<i>MPP6</i>	2.873557958	1.30E-04	3.06E-02
ENSCAFG00000023142	<i>INSYN2B</i>	-5.937195434	1.33E-04	3.06E-02
ENSCAFG00000016848		-1.871508093	1.35E-04	3.06E-02
ENSCAFG00000017740	<i>TBX4</i>	-2.734153411	1.36E-04	3.06E-02
ENSCAFG00000010307	<i>EIF4E</i>	0.758733732	1.44E-04	3.18E-02
ENSCAFG00000010924	<i>PKNOX2</i>	-2.035401793	1.52E-04	3.31E-02
ENSCAFG00000011666	<i>HSPA8</i>	1.427846873	1.57E-04	3.34E-02
ENSCAFG00000010696	<i>ADD3</i>	-1.387568508	1.59E-04	3.34E-02
ENSCAFG00000009568	<i>CLIC6</i>	4.027412221	1.65E-04	3.34E-02
ENSCAFG00000007392	<i>CENPK</i>	3.450862099	1.66E-04	3.34E-02
ENSCAFG00000030429	<i>PDCD2</i>	0.815213510	1.69E-04	3.34E-02
ENSCAFG00000001798	<i>SLC30A9</i>	0.712676126	1.73E-04	3.34E-02
ENSCAFG00000008101	<i>PLOD2</i>	2.353752870	1.74E-04	3.34E-02
ENSCAFG00000010821	<i>TMTC1</i>	-2.114669129	1.75E-04	3.34E-02
ENSCAFG00000011405	<i>TXNIP</i>	-1.702062816	1.75E-04	3.34E-02
ENSCAFG00000007009	<i>TMEM68</i>	1.000476571	1.80E-04	3.37E-02
ENSCAFG00000005677	<i>ARAP1</i>	-1.163438177	1.91E-04	3.51E-02
ENSCAFG00000000008	<i>TXNL4A</i>	0.917872917	1.92E-04	3.51E-02
ENSCAFG00000003683	<i>ORC6</i>	2.227295941	2.00E-04	3.55E-02
ENSCAFG00000014750	<i>ERO1A</i>	1.027776103	2.00E-04	3.55E-02
ENSCAFG00000023017	<i>SLC35E3</i>	1.199408460	2.04E-04	3.55E-02
ENSCAFG00000028843	<i>MAP1LC3C</i>	4.798976230	2.05E-04	3.55E-02

ENSCAFG00000011659	<i>ATP9A</i>	-2.44242906	2.10E-04	3.55E-02
ENSCAFG00000007153	<i>RAB2A</i>	0.709283281	2.11E-04	3.55E-02
ENSCAFG00000013087	<i>PATZ1</i>	-0.967171053	2.14E-04	3.55E-02
ENSCAFG00000016793	<i>ZNF646</i>	-1.059361913	2.15E-04	3.55E-02
ENSCAFG00000007731	<i>ATP1B3</i>	1.165518109	2.20E-04	3.59E-02
ENSCAFG00000004823		0.857286938	2.25E-04	3.60E-02
ENSCAFG00000008224	<i>ANKRA2</i>	0.814271570	2.28E-04	3.60E-02
ENSCAFG00000007666	<i>EIF4G2</i>	1.030988976	2.29E-04	3.60E-02
ENSCAFG00000013774	<i>MAP3K14</i>	-1.463889136	2.35E-04	3.63E-02
ENSCAFG00000015978	<i>UGDH</i>	1.780835299	2.37E-04	3.63E-02
ENSCAFG00000030867	<i>HPCAL4</i>	2.072374947	2.41E-04	3.63E-02
ENSCAFG00000000105	<i>MALT1</i>	-0.971445358	2.45E-04	3.63E-02
ENSCAFG00000007878	<i>SLC43A3</i>	-1.032525583	2.46E-04	3.63E-02
ENSCAFG00000005637		0.645946107	2.49E-04	3.63E-02
ENSCAFG00000029011	<i>CD164</i>	0.813321513	2.51E-04	3.63E-02
ENSCAFG00000002774	<i>ZNF462</i>	-1.50267696	2.52E-04	3.63E-02
ENSCAFG00000002999	<i>ZMPSTE24</i>	0.732205956	2.60E-04	3.71E-02
ENSCAFG00000031469		-1.791450258	2.63E-04	3.72E-02
ENSCAFG00000017665	<i>VMP1</i>	1.339660928	2.68E-04	3.73E-02
ENSCAFG00000011859	<i>SPC25</i>	1.464729622	2.70E-04	3.73E-02
ENSCAFG00000006638	<i>SNAI2</i>	1.308879576	2.85E-04	3.91E-02
ENSCAFG00000028508	<i>CD59</i>	1.643212524	2.96E-04	3.95E-02
ENSCAFG00000018638	<i>KSR1</i>	-1.508558569	2.97E-04	3.95E-02
ENSCAFG00000028589		-1.858620544	2.97E-04	3.95E-02
ENSCAFG00000001800	<i>TMEM263</i>	1.232378725	3.01E-04	3.96E-02
ENSCAFG00000003905	<i>AMD1</i>	0.858416427	3.11E-04	3.96E-02
ENSCAFG00000018879		0.958044776	3.14E-04	3.96E-02
ENSCAFG00000006175	<i>CDC25B</i>	-1.785199129	3.14E-04	3.96E-02
ENSCAFG00000004204	<i>PSMC2</i>	0.621414864	3.15E-04	3.96E-02
ENSCAFG00000024922	<i>ATP5PF</i>	0.793536737	3.15E-04	3.96E-02
ENSCAFG00000015815	<i>ADSS</i>	1.031765794	3.20E-04	3.98E-02
ENSCAFG00000017798	<i>MORF4L2</i>	0.931875811	3.25E-04	3.98E-02
ENSCAFG00000019134	<i>PPL</i>	-3.313895119	3.26E-04	3.98E-02
ENSCAFG00000037738		-1.96757667	3.36E-04	4.00E-02
ENSCAFG00000018842	<i>PHF12</i>	-0.6946492	3.36E-04	4.00E-02
ENSCAFG00000028827	<i>UBE2V2</i>	0.790754345	3.40E-04	4.00E-02
ENSCAFG00000019901	<i>SLC7A5</i>	2.323713647	3.40E-04	4.00E-02
ENSCAFG00000014222	<i>IFT80</i>	0.797618495	3.41E-04	4.00E-02
ENSCAFG00000017792		0.932383622	3.49E-04	4.06E-02
ENSCAFG00000012605	<i>ZNF687</i>	-1.016218362	3.56E-04	4.10E-02
ENSCAFG00000000677	<i>LRP12</i>	1.188723435	3.63E-04	4.14E-02

ENSCAFG00000013826	<i>SEC23A</i>	1.372167964	3.66E-04	4.15E-02
ENSCAFG00000001507	<i>NRF1</i>	-0.87001426	3.77E-04	4.19E-02
ENSCAFG00000008487	<i>PDE3B</i>	-2.384856882	3.85E-04	4.19E-02
ENSCAFG00000017556	<i>TEX14</i>	-1.528913712	3.86E-04	4.19E-02
ENSCAFG00000009187	<i>LOXL2</i>	3.316328965	3.88E-04	4.19E-02
ENSCAFG00000008738	<i>TM9SF3</i>	0.641813657	3.93E-04	4.19E-02
ENSCAFG00000007266	<i>HECTD2</i>	1.176405837	3.94E-04	4.19E-02
ENSCAFG00000007074	<i>IMPAD1</i>	1.119269053	3.95E-04	4.19E-02
ENSCAFG00000003476	<i>BLVRA</i>	1.088794138	3.97E-04	4.19E-02
ENSCAFG00000012672	<i>SCFD1</i>	0.771797602	3.99E-04	4.19E-02
ENSCAFG00000005520	<i>SNRPB2</i>	0.628310836	4.02E-04	4.19E-02
ENSCAFG00000007761	<i>IRF2</i>	-0.76941207	4.04E-04	4.19E-02
ENSCAFG00000013773	<i>BCR</i>	-0.888297127	4.10E-04	4.22E-02
ENSCAFG00000012449	<i>USP53</i>	-1.626748582	4.19E-04	4.22E-02
ENSCAFG00000017515	<i>NRDE2</i>	-0.705657071	4.23E-04	4.22E-02
ENSCAFG00000003198	<i>FKBP15</i>	-0.741145301	4.24E-04	4.22E-02
ENSCAFG00000005820	<i>ACSS3</i>	-1.318294285	4.25E-04	4.22E-02
ENSCAFG00000018147	<i>SS18</i>	0.757596900	4.29E-04	4.22E-02
ENSCAFG00000018588	<i>L3MBTL4</i>	-2.360012768	4.30E-04	4.22E-02
ENSCAFG00000019849	<i>ACSF3</i>	-1.196254746	4.34E-04	4.22E-02
ENSCAFG00000029086	<i>ARF4</i>	0.855031976	4.34E-04	4.22E-02
ENSCAFG00000011167	<i>PAPSS1</i>	1.283284707	4.44E-04	4.24E-02
ENSCAFG00000015564	<i>JKAMP</i>	0.869160357	4.46E-04	4.24E-02
ENSCAFG00000006917	<i>MAP3K1</i>	-1.067656349	4.47E-04	4.24E-02
ENSCAFG00000010877	<i>ABCA9</i>	-3.769425108	4.52E-04	4.24E-02
ENSCAFG00000029752	<i>C11H5orf15</i>	1.024215326	4.53E-04	4.24E-02
ENSCAFG00000002856	<i>CYCS</i>	1.163527903	4.60E-04	4.24E-02
ENSCAFG00000009978	<i>ETS2</i>	-1.408580212	4.61E-04	4.24E-02
ENSCAFG00000017192	<i>HMMR</i>	1.180278706	4.61E-04	4.24E-02
ENSCAFG00000006988	<i>MRPL15</i>	0.629832258	4.70E-04	4.25E-02
ENSCAFG00000015346	<i>STAT5A</i>	-0.558222386	4.71E-04	4.25E-02
ENSCAFG00000009031	<i>CMYA5</i>	-1.733577253	4.72E-04	4.25E-02
ENSCAFG00000016173	<i>THRA</i>	-1.066742116	4.78E-04	4.26E-02
ENSCAFG00000011099	<i>AMZ2</i>	1.185912741	4.79E-04	4.26E-02
ENSCAFG00000031859	<i>FKBP7</i>	1.577556967	4.96E-04	4.36E-02
ENSCAFG00000019645	<i>VAMP7</i>	0.680766688	4.98E-04	4.36E-02
ENSCAFG00000016452		0.885333496	5.00E-04	4.36E-02
ENSCAFG00000019762	<i>SURF4</i>	0.980846795	5.04E-04	4.36E-02
ENSCAFG00000018617	<i>RAVER2</i>	-1.693768558	5.06E-04	4.36E-02
ENSCAFG00000010550	<i>ZBTB8OS</i>	0.699167816	5.12E-04	4.38E-02
ENSCAFG00000011598		1.365035428	5.23E-04	4.45E-02

ENSCAFG00000012909	<i>ZNF592</i>	-1.036116116	5.34E-04	4.50E-02
ENSCAFG00000000281	<i>ABRACL</i>	0.985115102	5.40E-04	4.50E-02
ENSCAFG00000017296	<i>RNF145</i>	1.048101251	5.40E-04	4.50E-02
ENSCAFG00000023022		-2.421793943	5.45E-04	4.50E-02
ENSCAFG00000014358	<i>CASK</i>	1.578716228	5.45E-04	4.50E-02
ENSCAFG00000035640		-3.113588697	5.54E-04	4.54E-02
ENSCAFG00000002994	<i>NT5E</i>	1.329662017	5.68E-04	4.60E-02
ENSCAFG00000016168	<i>BID</i>	0.795879107	5.68E-04	4.60E-02
ENSCAFG00000031507		-3.18148177	5.83E-04	4.69E-02
ENSCAFG00000006680	<i>LRIG1</i>	-1.465260818	5.87E-04	4.69E-02
ENSCAFG00000019738	<i>ATP5PB</i>	0.649997250	5.91E-04	4.69E-02
ENSCAFG00000014600	<i>BRCA1</i>	1.880013822	5.96E-04	4.69E-02
ENSCAFG00000005732	<i>BAHCC1</i>	-1.3261622	5.96E-04	4.69E-02
ENSCAFG00000017616	<i>LARP6</i>	-1.606357343	6.03E-04	4.72E-02
ENSCAFG00000009232	<i>AMIGO2</i>	-4.821536265	6.20E-04	4.81E-02
ENSCAFG00000007200	<i>SLC20A1</i>	1.368397610	6.21E-04	4.81E-02
ENSCAFG00000019827	<i>AMIGO1</i>	-2.082186402	6.39E-04	4.86E-02
ENSCAFG00000015300	<i>SPPL2A</i>	0.799344451	6.40E-04	4.86E-02
ENSCAFG00000019963	<i>FNBP1</i>	-0.852334276	6.57E-04	4.86E-02
ENSCAFG00000019251	<i>CREBBP</i>	-0.747136444	6.59E-04	4.86E-02
ENSCAFG00000014924	<i>MYOC</i>	-5.434488744	6.61E-04	4.86E-02
ENSCAFG00000030902	<i>TRIM59</i>	1.264135843	6.61E-04	4.86E-02
ENSCAFG00000019440	<i>MECP2</i>	-0.767921111	6.63E-04	4.86E-02
ENSCAFG00000030087		2.973130881	6.65E-04	4.86E-02
ENSCAFG00000006162	<i>SMAD9</i>	0.798065077	6.67E-04	4.86E-02
ENSCAFG00000020228	<i>DHX38</i>	-0.568740556	6.71E-04	4.86E-02
ENSCAFG00000003875	<i>PLK4</i>	2.881650357	6.72E-04	4.86E-02
ENSCAFG00000013495	<i>OLR1</i>	3.747433434	6.72E-04	4.86E-02
ENSCAFG00000007306	<i>MAPRE1</i>	1.062858751	6.75E-04	4.86E-02
ENSCAFG00000009391	<i>PTPN5</i>	5.844480640	6.80E-04	4.88E-02
ENSCAFG00000012717	<i>QSOX1</i>	1.337477246	6.86E-04	4.88E-02
ENSCAFG00000031178	<i>CNIH1</i>	0.979393573	6.92E-04	4.88E-02
ENSCAFG00000016479	<i>TMEM184A</i>	-3.131257073	6.92E-04	4.88E-02
ENSCAFG00000031992	<i>SAR1B</i>	0.814297204	6.97E-04	4.88E-02
ENSCAFG00000020290	<i>DAB2IP</i>	-1.067275241	7.00E-04	4.88E-02
ENSCAFG00000013160	<i>NPL</i>	-1.362785474	7.05E-04	4.88E-02
ENSCAFG00000015037	<i>DCUN1D5</i>	0.889160835	7.05E-04	4.88E-02
ENSCAFG00000032183	<i>SGPP1</i>	0.523551290	7.11E-04	4.89E-02

Supplementary Table 9.2S. Differentially expressed genes identified through RNA sequencing of synovium tissue from ACLR case and control dogs

Ensembl ID	Gene Name	logFC	P	Adjusted P
ENSCAFG00000009569	<i>SPP1</i>	6.213698843	5.56E-07	3.79E-03
ENSCAFG00000012598	<i>SLAMF7</i>	3.671933543	8.99E-07	3.79E-03
ENSCAFG00000012614	<i>KCNK2</i>	4.586032792	9.90E-07	3.79E-03
ENSCAFG00000030867	<i>HPCAL4</i>	3.531466567	1.11E-06	3.79E-03
ENSCAFG00000014432		7.073379802	2.34E-06	5.75E-03
ENSCAFG00000016883	<i>ITGAX</i>	2.848214280	2.52E-06	5.75E-03
ENSCAFG00000028568	<i>JCHAIN</i>	6.631592776	3.56E-06	6.11E-03
ENSCAFG00000003757	<i>OSTM1</i>	0.991227213	4.38E-06	6.11E-03
ENSCAFG00000016556	<i>AQP9</i>	3.464606398	4.40E-06	6.11E-03
ENSCAFG00000017342	<i>GPR65</i>	3.025758701	4.75E-06	6.11E-03
ENSCAFG00000011698	<i>TLR7</i>	2.847801092	4.91E-06	6.11E-03
ENSCAFG00000012330	<i>TMEM206</i>	1.225133256	5.45E-06	6.22E-03
ENSCAFG00000002660		2.966687022	6.12E-06	6.40E-03
ENSCAFG00000013504	<i>CLEC7A</i>	4.133138932	6.65E-06	6.40E-03
ENSCAFG000000031806		7.911716282	7.02E-06	6.40E-03
ENSCAFG00000015017	<i>ATP6V0A1</i>	1.214559362	7.53E-06	6.40E-03
ENSCAFG000000031273		5.354536483	8.00E-06	6.40E-03
ENSCAFG00000013940	<i>CLEC4E</i>	3.212744124	8.42E-06	6.40E-03
ENSCAFG00000010608	<i>CFAP58</i>	6.199051557	9.28E-06	6.40E-03
ENSCAFG00000017616	<i>LARP6</i>	-2.502719134	9.45E-06	6.40E-03
ENSCAFG00000029357	<i>NCEH1</i>	1.120173547	9.81E-06	6.40E-03
ENSCAFG00000005685	<i>PAK5</i>	5.795350091	1.14E-05	7.04E-03
ENSCAFG000000031204	<i>RGS10</i>	2.514888716	1.24E-05	7.04E-03
ENSCAFG000000031013	<i>SAMSN1</i>	3.032346558	1.24E-05	7.04E-03
ENSCAFG00000002061	<i>CD72</i>	3.505614240	1.29E-05	7.04E-03
ENSCAFG00000008408	<i>TPD52</i>	1.659460400	1.38E-05	7.28E-03
ENSCAFG00000013495	<i>OLR1</i>	6.019565158	1.49E-05	7.55E-03
ENSCAFG00000001246	<i>PTCH1</i>	-1.831673152	1.78E-05	8.70E-03
ENSCAFG00000013673	<i>CXHXorf21</i>	2.331615606	2.18E-05	1.02E-02
ENSCAFG000000028603		3.163340197	2.23E-05	1.02E-02
ENSCAFG00000012671	<i>TRIM63</i>	2.098492778	2.36E-05	1.04E-02
ENSCAFG00000011041	<i>CSF2RA</i>	2.341655865	3.38E-05	1.28E-02
ENSCAFG00000002515	<i>MACC1</i>	4.405367725	3.46E-05	1.28E-02
ENSCAFG000000031786		8.711951968	3.47E-05	1.28E-02
ENSCAFG00000006175	<i>CDC25B</i>	-2.142961031	3.53E-05	1.28E-02
ENSCAFG000000028667	<i>CBLN3</i>	2.236913637	3.65E-05	1.28E-02
ENSCAFG00000017692	<i>GLA</i>	1.054104266	3.70E-05	1.28E-02
ENSCAFG000000028509		7.174385771	3.83E-05	1.28E-02

ENSCAFG00000013763	<i>LTF</i>	2.594059393	3.91E-05	1.28E-02
ENSCAFG00000008026		2.579873726	3.96E-05	1.28E-02
ENSCAFG00000004406	<i>FBLN2</i>	-2.792105657	4.03E-05	1.28E-02
ENSCAFG00000006691	<i>SLC46A3</i>	1.691301109	4.17E-05	1.28E-02
ENSCAFG00000004595		5.140371591	4.30E-05	1.28E-02
ENSCAFG00000032078		8.737113956	4.30E-05	1.28E-02
ENSCAFG00000031733		7.230974876	4.33E-05	1.28E-02
ENSCAFG00000007012	<i>SPIC</i>	6.183153325	4.45E-05	1.28E-02
ENSCAFG00000004265	<i>THNSL1</i>	1.655225448	4.57E-05	1.28E-02
ENSCAFG00000012206	<i>ACKR3</i>	-2.722790918	4.57E-05	1.28E-02
ENSCAFG00000000281	<i>ABRACL</i>	1.201527108	4.57E-05	1.28E-02
ENSCAFG00000031201		2.923961378	5.96E-05	1.62E-02
ENSCAFG00000015593		2.678330827	6.09E-05	1.62E-02
ENSCAFG00000010997	<i>CD80</i>	2.092379802	6.25E-05	1.62E-02
ENSCAFG00000031397		3.250093129	6.28E-05	1.62E-02
ENSCAFG00000003945	<i>DENND2A</i>	-1.557794039	6.64E-05	1.65E-02
ENSCAFG00000002318	<i>ARL4A</i>	-1.856899059	6.82E-05	1.65E-02
ENSCAFG00000018482	<i>CERS4</i>	-1.770924313	6.86E-05	1.65E-02
ENSCAFG00000028762	<i>KLRG1</i>	2.992506318	6.92E-05	1.65E-02
ENSCAFG00000013314	<i>NCF2</i>	2.183350391	7.18E-05	1.65E-02
ENSCAFG00000007203	<i>GGH</i>	1.654501638	7.21E-05	1.65E-02
ENSCAFG00000029297	<i>ANKRD22</i>	3.638889183	7.29E-05	1.65E-02
ENSCAFG00000010711	<i>ATP6V1A</i>	1.484354604	7.40E-05	1.65E-02
ENSCAFG00000001027	<i>FAM91A1</i>	1.100024849	7.55E-05	1.65E-02
ENSCAFG00000012286	<i>STRADB</i>	1.403366871	7.70E-05	1.65E-02
ENSCAFG00000030196	<i>DRAM2</i>	1.320128888	7.73E-05	1.65E-02
ENSCAFG00000009101	<i>ZHX3</i>	-1.177659593	7.93E-05	1.66E-02
ENSCAFG00000001502	<i>CARD10</i>	-2.144674886	8.03E-05	1.66E-02
ENSCAFG00000003694	<i>PRDM1</i>	2.732698177	8.16E-05	1.66E-02
ENSCAFG00000028847		6.327953766	8.24E-05	1.66E-02
ENSCAFG00000015726	<i>KMO</i>	1.933268978	8.36E-05	1.66E-02
ENSCAFG00000011188	<i>GALNT3</i>	2.566406481	8.79E-05	1.72E-02
ENSCAFG00000023924	<i>CD151</i>	-0.935065824	9.19E-05	1.73E-02
ENSCAFG00000003662	<i>TBC1D9</i>	1.827712756	9.23E-05	1.73E-02
ENSCAFG00000029467		7.178086654	9.24E-05	1.73E-02
ENSCAFG00000001591	<i>TREM1</i>	2.888787101	9.68E-05	1.77E-02
ENSCAFG00000013325	<i>PHEX</i>	-4.294824759	9.72E-05	1.77E-02
ENSCAFG00000002935	<i>VRK2</i>	0.819577208	9.92E-05	1.79E-02
ENSCAFG00000029252	<i>RND3</i>	-1.604965962	1.05E-04	1.87E-02
ENSCAFG00000003859		-3.251139406	1.11E-04	1.96E-02
ENSCAFG00000012903	<i>RUNX3</i>	2.149604373	1.13E-04	1.97E-02

ENSCAFG00000032751	<i>CLEC2B</i>	2.987859098	1.15E-04	1.97E-02
ENSCAFG00000012883	<i>ARHGDIB</i>	1.340153868	1.17E-04	1.97E-02
ENSCAFG00000014074	<i>CRYAB</i>	-2.004377773	1.20E-04	1.97E-02
ENSCAFG00000012569	<i>CD84</i>	2.327042299	1.20E-04	1.97E-02
ENSCAFG00000038896		2.749665480	1.23E-04	1.97E-02
ENSCAFG00000017192	<i>HMMR</i>	1.270231437	1.23E-04	1.97E-02
ENSCAFG00000030258	<i>IGHM</i>	7.435020427	1.25E-04	1.97E-02
ENSCAFG00000016245		2.845557434	1.27E-04	1.97E-02
ENSCAFG00000013478	<i>EDEM3</i>	1.263358563	1.29E-04	1.97E-02
ENSCAFG00000000419	<i>AKAP12</i>	-2.408411931	1.31E-04	1.97E-02
ENSCAFG00000002171	<i>SYK</i>	1.698324508	1.31E-04	1.97E-02
ENSCAFG00000010491	<i>LPCAT1</i>	1.503404334	1.32E-04	1.97E-02
ENSCAFG00000014799		-1.501886779	1.32E-04	1.97E-02
ENSCAFG00000025025	<i>TRIM6</i>	1.671293334	1.34E-04	1.98E-02
ENSCAFG00000018465	<i>PARP8</i>	0.997311184	1.38E-04	2.00E-02
ENSCAFG00000005038	<i>MAP3K19</i>	6.363575755	1.39E-04	2.00E-02
ENSCAFG00000029125	<i>SIRPB2</i>	2.294016598	1.43E-04	2.03E-02
ENSCAFG00000025113	<i>CLEC12A</i>	2.859536357	1.44E-04	2.03E-02
ENSCAFG00000010198	<i>RAB7B</i>	1.333157193	1.48E-04	2.06E-02
ENSCAFG00000031529	<i>CLEC2D</i>	4.991378001	1.49E-04	2.06E-02
ENSCAFG00000032474		3.459286360	1.52E-04	2.07E-02
ENSCAFG00000001712	<i>LRRC19</i>	2.655672911	1.53E-04	2.07E-02
ENSCAFG00000009906	<i>VSIG10</i>	-1.013423393	1.59E-04	2.13E-02
ENSCAFG00000014349	<i>TBC1D14</i>	0.820903131	1.61E-04	2.14E-02
ENSCAFG00000031853		6.533256849	1.66E-04	2.18E-02
ENSCAFG00000020345	<i>PTGFR</i>	1.989957165	1.71E-04	2.19E-02
ENSCAFG00000014879	<i>LRRC25</i>	2.361388721	1.71E-04	2.19E-02
ENSCAFG00000016114	<i>ADA2</i>	2.169768782	1.74E-04	2.19E-02
ENSCAFG00000010095	<i>ATP6V1B2</i>	1.006204618	1.75E-04	2.19E-02
ENSCAFG00000001932	<i>CNTFR</i>	-3.126359708	1.75E-04	2.19E-02
ENSCAFG00000024111		6.351106634	1.76E-04	2.19E-02
ENSCAFG00000001816	<i>ATP8A1</i>	1.345746061	1.78E-04	2.19E-02
ENSCAFG00000004944	<i>PCDH9</i>	3.547100238	1.79E-04	2.19E-02
ENSCAFG00000004854	<i>ATP6V0B</i>	1.115852038	1.83E-04	2.19E-02
ENSCAFG00000032057		7.252134433	1.84E-04	2.19E-02
ENSCAFG00000015176		6.721131662	1.84E-04	2.19E-02
ENSCAFG00000020290	<i>DAB2IP</i>	-1.210249499	1.87E-04	2.19E-02
ENSCAFG00000006200	<i>PLXNC1</i>	2.740794473	1.87E-04	2.19E-02
ENSCAFG00000007307		-4.158353586	1.90E-04	2.20E-02
ENSCAFG00000032358		9.302051606	1.94E-04	2.20E-02
ENSCAFG00000009721	<i>ALCAM</i>	1.544928122	1.95E-04	2.20E-02

ENSCAFG00000016584	<i>HK3</i>	3.973427471	1.95E-04	2.20E-02
ENSCAFG00000029367	<i>COL28A1</i>	-7.108851229	1.96E-04	2.20E-02
ENSCAFG00000015441	<i>LRMDA</i>	1.516775601	1.99E-04	2.22E-02
ENSCAFG00000000647	<i>SERAC1</i>	1.009119435	2.05E-04	2.24E-02
ENSCAFG00000000398	<i>LRP11</i>	1.035341500	2.06E-04	2.24E-02
ENSCAFG00000002323	<i>MGAT4A</i>	1.627362102	2.07E-04	2.24E-02
ENSCAFG00000029821	<i>IFNE</i>	4.900089889	2.08E-04	2.24E-02
ENSCAFG00000028917	<i>SNX10</i>	1.914326555	2.10E-04	2.25E-02
ENSCAFG00000025192	<i>CASP8</i>	1.283884770	2.12E-04	2.26E-02
ENSCAFG00000012621	<i>LY9</i>	2.059033234	2.20E-04	2.32E-02
ENSCAFG00000011677	<i>EPB41</i>	1.154487787	2.27E-04	2.35E-02
ENSCAFG00000023369		6.695293095	2.31E-04	2.35E-02
ENSCAFG00000000100	<i>SEC11C</i>	1.340915129	2.33E-04	2.35E-02
ENSCAFG00000000804	<i>FBLN1</i>	-2.339621938	2.35E-04	2.35E-02
ENSCAFG00000030838	<i>PTGES</i>	-2.230404183	2.35E-04	2.35E-02
ENSCAFG00000018852	<i>TNFRSF17</i>	4.481479689	2.38E-04	2.35E-02
ENSCAFG00000009883	<i>TENT5C</i>	2.400632866	2.38E-04	2.35E-02
ENSCAFG00000007617	<i>ST8SIA4</i>	2.647811526	2.38E-04	2.35E-02
ENSCAFG00000000159	<i>NEMP1</i>	0.941952875	2.39E-04	2.35E-02
ENSCAFG00000000931	<i>NUS1</i>	0.809258502	2.43E-04	2.37E-02
ENSCAFG00000018540	<i>MYO1F</i>	1.997604391	2.46E-04	2.37E-02
ENSCAFG00000032195		8.502290486	2.47E-04	2.37E-02
ENSCAFG00000002918	<i>RUFY3</i>	0.858826968	2.48E-04	2.37E-02
ENSCAFG00000009056	<i>ADAM28</i>	2.985633950	2.49E-04	2.37E-02
ENSCAFG00000011458	<i>LRMP</i>	1.807997812	2.55E-04	2.38E-02
ENSCAFG00000018492	<i>CRLF3</i>	1.021977310	2.57E-04	2.38E-02
ENSCAFG00000009829	<i>CD2</i>	3.453788939	2.59E-04	2.38E-02
ENSCAFG00000015000	<i>TACC3</i>	1.535507489	2.59E-04	2.38E-02
ENSCAFG00000005586	<i>AP3M2</i>	1.708612865	2.63E-04	2.38E-02
ENSCAFG00000007248	<i>GNPTAB</i>	1.097038136	2.63E-04	2.38E-02
ENSCAFG00000010582	<i>BOC</i>	-1.501817831	2.68E-04	2.38E-02
ENSCAFG00000001966	<i>SH3RF3</i>	1.782370601	2.68E-04	2.38E-02
ENSCAFG00000031104	<i>CX3CR1</i>	2.704939383	2.68E-04	2.38E-02
ENSCAFG00000013544	<i>HSPB1</i>	-1.16336307	2.69E-04	2.38E-02
ENSCAFG00000004762	<i>SLC25A19</i>	0.915199894	2.71E-04	2.38E-02
ENSCAFG00000010413	<i>TMEM54</i>	-1.852637699	2.71E-04	2.38E-02
ENSCAFG00000003388	<i>SH3D21</i>	-1.795234059	2.72E-04	2.38E-02
ENSCAFG00000010821	<i>TMTC1</i>	-1.94810964	2.79E-04	2.42E-02
ENSCAFG00000007954	<i>TNKS1BP1</i>	-1.453417186	2.83E-04	2.43E-02
ENSCAFG00000015177	<i>SELL</i>	3.433430108	2.91E-04	2.47E-02
ENSCAFG00000024646	<i>ERAP2</i>	2.063998482	2.93E-04	2.47E-02

ENSCAFG00000001675	<i>CDKN2A</i>	3.379241308	2.93E-04	2.47E-02
ENSCAFG000000030602		5.272179495	2.94E-04	2.47E-02
ENSCAFG000000002774	<i>ZNF462</i>	-1.451829208	2.98E-04	2.47E-02
ENSCAFG000000006243	<i>GNPDA1</i>	1.319626777	3.01E-04	2.47E-02
ENSCAFG000000013773	<i>BCR</i>	-0.92466753	3.02E-04	2.47E-02
ENSCAFG000000023972		1.761585127	3.02E-04	2.47E-02
ENSCAFG000000004692	<i>DCLRE1C</i>	0.838857894	3.03E-04	2.47E-02
ENSCAFG000000007893	<i>RBP4</i>	-2.212294766	3.06E-04	2.48E-02
ENSCAFG000000011086	<i>SLC16A6</i>	2.041481228	3.09E-04	2.48E-02
ENSCAFG000000008123	<i>TENM3</i>	1.205120317	3.10E-04	2.48E-02
ENSCAFG000000030180	<i>IL18</i>	1.774505844	3.24E-04	2.56E-02
ENSCAFG000000002138	<i>IL18R1</i>	2.449359135	3.25E-04	2.56E-02
ENSCAFG000000011879	<i>MOSPD2</i>	1.144568468	3.26E-04	2.56E-02
ENSCAFG000000015489	<i>SCG3</i>	2.658549720	3.30E-04	2.58E-02
ENSCAFG000000010241	<i>PXN</i>	-0.694933654	3.32E-04	2.58E-02
ENSCAFG000000011828	<i>KLHL6</i>	2.014021954	3.36E-04	2.60E-02
ENSCAFG000000033004		2.195463823	3.41E-04	2.63E-02
ENSCAFG000000004011	<i>TBXAS1</i>	1.905871844	3.47E-04	2.66E-02
ENSCAFG000000000103	<i>ZNF532</i>	-1.13516722	3.51E-04	2.66E-02
ENSCAFG000000031753		6.454972430	3.51E-04	2.66E-02
ENSCAFG000000029313	<i>CLDN1</i>	-2.254564035	3.54E-04	2.66E-02
ENSCAFG000000030902	<i>TRIM59</i>	1.470517183	3.55E-04	2.66E-02
ENSCAFG000000016515	<i>PEAR1</i>	-1.769244482	3.64E-04	2.69E-02
ENSCAFG000000019060	<i>ABR</i>	0.860432453	3.64E-04	2.69E-02
ENSCAFG000000010949	<i>SLC37A2</i>	2.720502222	3.67E-04	2.70E-02
ENSCAFG000000031669	<i>CNRIP1</i>	-1.686883442	3.70E-04	2.71E-02
ENSCAFG000000002590		2.527738934	3.78E-04	2.74E-02
ENSCAFG000000032325		9.184766128	3.79E-04	2.74E-02
ENSCAFG000000009516	<i>LY86</i>	2.031288824	3.82E-04	2.75E-02
ENSCAFG000000011177	<i>SLC7A7</i>	1.541660577	3.85E-04	2.76E-02
ENSCAFG000000015914	<i>PELP1</i>	-0.913705653	3.96E-04	2.82E-02
ENSCAFG000000016864	<i>ELMSAN1</i>	-1.109595934	3.97E-04	2.82E-02
ENSCAFG000000004694	<i>PTCH2</i>	-2.523753328	4.05E-04	2.86E-02
ENSCAFG000000033303		-1.348261088	4.10E-04	2.87E-02
ENSCAFG000000011657	<i>EEF2</i>	3.641094216	4.13E-04	2.87E-02
ENSCAFG000000028982		-1.989659723	4.14E-04	2.87E-02
ENSCAFG000000011950	<i>SLC18A2</i>	-3.620140891	4.18E-04	2.87E-02
ENSCAFG000000001361	<i>APOBEC3Z3</i>	1.724474755	4.18E-04	2.87E-02
ENSCAFG000000001738	<i>GRM8</i>	5.054399100	4.18E-04	2.87E-02
ENSCAFG000000014790		1.291436406	4.21E-04	2.87E-02
ENSCAFG000000029700	<i>LPAR6</i>	1.165671054	4.24E-04	2.87E-02

ENSCAFG00000014924	<i>MYOC</i>	-6.300208962	4.26E-04	2.87E-02
ENSCAFG00000014213	<i>POU2AF1</i>	6.757271994	4.27E-04	2.87E-02
ENSCAFG00000018158	<i>ZNF521</i>	-1.091048329	4.37E-04	2.92E-02
ENSCAFG00000016603	<i>ITGAL</i>	2.370717796	4.40E-04	2.92E-02
ENSCAFG00000000848	<i>DLA-DMA</i>	2.793791543	4.47E-04	2.94E-02
ENSCAFG00000009847	<i>PCIF1</i>	-0.740891032	4.47E-04	2.94E-02
ENSCAFG00000002796	<i>CCDC88A</i>	1.055071618	4.49E-04	2.94E-02
ENSCAFG00000006720	<i>AMOTL2</i>	-1.691677432	4.50E-04	2.94E-02
ENSCAFG00000002519	<i>ITGB8</i>	3.475364680	4.60E-04	2.94E-02
ENSCAFG00000002528	<i>TGFB1</i>	1.201935991	4.60E-04	2.94E-02
ENSCAFG00000002834		1.999514452	4.64E-04	2.94E-02
ENSCAFG00000017697		-1.900240042	4.65E-04	2.94E-02
ENSCAFG00000005334	<i>GPR180</i>	1.053472712	4.66E-04	2.94E-02
ENSCAFG00000013806	<i>M6PR</i>	0.978347582	4.68E-04	2.94E-02
ENSCAFG00000009802	<i>CYLD</i>	0.945614844	4.68E-04	2.94E-02
ENSCAFG00000020300	<i>LPAR3</i>	3.706304549	4.68E-04	2.94E-02
ENSCAFG00000005613	<i>SLC4A7</i>	1.134362291	4.72E-04	2.94E-02
ENSCAFG00000011751	<i>CD86</i>	2.370986364	4.72E-04	2.94E-02
ENSCAFG00000008456		0.845816910	4.75E-04	2.94E-02
ENSCAFG00000007417	<i>MAN2A1</i>	1.165508178	4.87E-04	2.99E-02
ENSCAFG00000013485	<i>KLF7</i>	-1.150227552	4.87E-04	2.99E-02
ENSCAFG00000000812	<i>DLA-DQA1</i>	4.132343700	4.90E-04	2.99E-02
ENSCAFG00000032183	<i>SGPP1</i>	0.548888979	4.95E-04	3.00E-02
ENSCAFG00000014095	<i>DERL3</i>	4.255267613	4.95E-04	3.00E-02
ENSCAFG00000005608	<i>EDEM1</i>	1.096294587	4.99E-04	3.01E-02
ENSCAFG00000011727	<i>MILR1</i>	1.918381860	5.03E-04	3.02E-02
ENSCAFG00000015599		2.359014093	5.04E-04	3.02E-02
ENSCAFG00000030412	<i>BCL2L11</i>	0.891243316	5.10E-04	3.04E-02
ENSCAFG00000016173	<i>THRA</i>	-1.076557184	5.14E-04	3.05E-02
ENSCAFG00000013805	<i>CCR1</i>	1.981163756	5.17E-04	3.05E-02
ENSCAFG00000010623	<i>SORCS3</i>	4.304894070	5.28E-04	3.10E-02
ENSCAFG00000010910	<i>FEZ1</i>	-2.705216709	5.31E-04	3.10E-02
ENSCAFG00000004491	<i>LCP1</i>	1.957337720	5.31E-04	3.10E-02
ENSCAFG00000018711	<i>RANBP3L</i>	2.124913867	5.33E-04	3.10E-02
ENSCAFG00000012205	<i>INPP5F</i>	0.898687940	5.49E-04	3.16E-02
ENSCAFG00000000649	<i>ATP6V1C1</i>	0.803412846	5.51E-04	3.16E-02
ENSCAFG00000032590	<i>IL21R</i>	4.050064604	5.51E-04	3.16E-02
ENSCAFG00000001093	<i>FAM49B</i>	1.263853967	5.53E-04	3.16E-02
ENSCAFG00000001786	<i>ACO1</i>	0.875965116	5.60E-04	3.18E-02
ENSCAFG00000003452	<i>KLHL32</i>	2.778430796	5.82E-04	3.29E-02
ENSCAFG00000013317	<i>LMX1A</i>	-3.740683651	5.90E-04	3.31E-02

ENSCAFG00000003397	<i>TFEC</i>	2.512629099	5.95E-04	3.31E-02
ENSCAFG000000032313	<i>FGL1</i>	-4.09879967	5.95E-04	3.31E-02
ENSCAFG000000008959	<i>RIPK2</i>	0.774117837	5.96E-04	3.31E-02
ENSCAFG000000007693	<i>LPXN</i>	1.832647088	5.96E-04	3.31E-02
ENSCAFG000000007882	<i>PCSK1</i>	2.674606864	6.21E-04	3.42E-02
ENSCAFG000000000086	<i>CDH20</i>	8.406890367	6.21E-04	3.42E-02
ENSCAFG000000003611	<i>PRR12</i>	-0.955322493	6.24E-04	3.42E-02
ENSCAFG000000006055	<i>ATP2C1</i>	0.664942989	6.29E-04	3.43E-02
ENSCAFG000000007361	<i>EPB41L4A</i>	-1.750559981	6.37E-04	3.46E-02
ENSCAFG000000012410	<i>CRTC3</i>	-0.914653146	6.42E-04	3.48E-02
ENSCAFG000000005847	<i>ADAM9</i>	0.969598387	6.54E-04	3.53E-02
ENSCAFG000000031415		6.933636703	6.59E-04	3.54E-02
ENSCAFG000000031078		6.006381826	6.66E-04	3.56E-02
ENSCAFG000000004284	<i>LURAP1</i>	-1.838642803	6.75E-04	3.59E-02
ENSCAFG000000004379	<i>FNDC3A</i>	0.813367824	6.77E-04	3.59E-02
ENSCAFG000000008617	<i>GLRB</i>	-1.797406204	6.79E-04	3.59E-02
ENSCAFG000000005003	<i>NCKAP5</i>	2.763915999	6.88E-04	3.59E-02
ENSCAFG000000011908	<i>SHTN1</i>	1.500444979	6.88E-04	3.59E-02
ENSCAFG000000014564		3.310855789	6.88E-04	3.59E-02
ENSCAFG000000010948	<i>PREX1</i>	1.357161169	6.90E-04	3.59E-02
ENSCAFG000000029632		4.821331743	6.96E-04	3.61E-02
ENSCAFG000000013924	<i>PLCD3</i>	-1.208003979	7.00E-04	3.61E-02
ENSCAFG000000005354	<i>KLF6</i>	-1.880448476	7.02E-04	3.61E-02
ENSCAFG000000015584		5.234599930	7.07E-04	3.62E-02
ENSCAFG000000018672	<i>VTN</i>	-3.573501745	7.14E-04	3.64E-02
ENSCAFG000000018546	<i>NNT</i>	1.021580536	7.15E-04	3.64E-02
ENSCAFG000000005465	<i>DZIP1</i>	-0.95979865	7.17E-04	3.64E-02
ENSCAFG000000039260		-1.918793875	7.21E-04	3.64E-02
ENSCAFG000000017735	<i>GSKIP</i>	1.006033769	7.29E-04	3.67E-02
ENSCAFG000000009364	<i>TFG</i>	2.102334946	7.42E-04	3.71E-02
ENSCAFG000000001028	<i>RNF217</i>	1.116101839	7.43E-04	3.71E-02
ENSCAFG000000006668	<i>KBTBD8</i>	1.248902907	7.48E-04	3.73E-02
ENSCAFG000000004316	<i>GSAP</i>	1.176339521	7.51E-04	3.73E-02
ENSCAFG000000003626	<i>IL15</i>	0.962650314	7.66E-04	3.79E-02
ENSCAFG000000005155	<i>MFSD11</i>	0.931126767	7.73E-04	3.79E-02
ENSCAFG000000010536	<i>CARMIL1</i>	-1.495286372	7.78E-04	3.79E-02
ENSCAFG000000006249	<i>NDFIP1</i>	0.852280495	7.80E-04	3.79E-02
ENSCAFG000000018209	<i>EPN2</i>	-1.169322669	7.85E-04	3.79E-02
ENSCAFG000000007590	<i>ZNF507</i>	-0.768255287	7.86E-04	3.79E-02
ENSCAFG000000000741	<i>SYBU</i>	-1.76832498	7.86E-04	3.79E-02
ENSCAFG000000010550	<i>ZBTB8OS</i>	0.644458470	7.87E-04	3.79E-02

ENSCAFG00000009255	<i>PTPN22</i>	2.555657364	7.88E-04	3.79E-02
ENSCAFG00000008310	<i>ZFHX4</i>	-2.015368428	7.94E-04	3.80E-02
ENSCAFG00000012787		-2.63524956	7.97E-04	3.80E-02
ENSCAFG00000025145	<i>ZNF777</i>	-0.945788633	7.98E-04	3.80E-02
ENSCAFG00000009257	<i>MED13L</i>	-0.805663293	8.02E-04	3.80E-02
ENSCAFG00000008669	<i>TMEM144</i>	1.811930465	8.07E-04	3.81E-02
ENSCAFG00000000480	<i>NIPAL2</i>	2.155259815	8.18E-04	3.83E-02
ENSCAFG00000032359	<i>GIMAP2</i>	1.674323773	8.19E-04	3.83E-02
ENSCAFG00000007755	<i>MAT2A</i>	0.905125684	8.19E-04	3.83E-02
ENSCAFG00000008833	<i>ACP2</i>	0.680472397	8.28E-04	3.86E-02
ENSCAFG00000009379	<i>CRY2</i>	-1.212960736	8.36E-04	3.87E-02
ENSCAFG00000000814		3.950721966	8.39E-04	3.87E-02
ENSCAFG00000013961	<i>TP63</i>	2.887275122	8.40E-04	3.87E-02
ENSCAFG00000016746	<i>FAM155B</i>	-2.059335542	8.47E-04	3.89E-02
ENSCAFG00000029121	<i>FAM78A</i>	1.392007847	8.50E-04	3.89E-02
ENSCAFG00000023704	<i>SLAMF6</i>	4.279752763	8.62E-04	3.94E-02
ENSCAFG00000017626	<i>FAXDC2</i>	-0.769473352	8.76E-04	3.96E-02
ENSCAFG00000032102	<i>DLA-DMB</i>	2.685522565	8.78E-04	3.96E-02
ENSCAFG00000032369	<i>ZDHHC22</i>	1.137467124	8.79E-04	3.96E-02
ENSCAFG00000034404		1.870436481	8.79E-04	3.96E-02
ENSCAFG00000009498	<i>ADA</i>	-1.345181833	8.86E-04	3.97E-02
ENSCAFG00000008347	<i>NUGGC</i>	3.199922659	8.88E-04	3.97E-02
ENSCAFG00000006047	<i>LETM2</i>	-0.88837382	8.93E-04	3.97E-02
ENSCAFG00000002576	<i>HIVEP3</i>	1.602055106	8.93E-04	3.97E-02
ENSCAFG00000023843		8.945197701	9.00E-04	3.99E-02
ENSCAFG00000009152	<i>HDAC7</i>	-1.413010818	9.15E-04	4.04E-02
ENSCAFG00000015482	<i>TOP3B</i>	-0.81971959	9.17E-04	4.04E-02
ENSCAFG00000028636	<i>IGFBP6</i>	-2.825321894	9.25E-04	4.06E-02
ENSCAFG00000024010	<i>TLR1</i>	1.566012233	9.33E-04	4.08E-02
ENSCAFG00000009036	<i>PLCB2</i>	1.803494924	9.37E-04	4.09E-02
ENSCAFG00000007579	<i>PIK3CB</i>	0.958859642	9.43E-04	4.10E-02
ENSCAFG00000030297	<i>GRB2</i>	0.867612766	9.51E-04	4.11E-02
ENSCAFG00000003685	<i>CREM</i>	1.049570764	9.54E-04	4.11E-02
ENSCAFG00000005999	<i>NHLRC3</i>	1.148953215	9.54E-04	4.11E-02
ENSCAFG00000008351	<i>TLR2</i>	1.529756728	9.61E-04	4.13E-02
ENSCAFG00000017181	<i>MAT2B</i>	0.973740676	9.68E-04	4.14E-02
ENSCAFG00000018703	<i>SASH3</i>	1.721082385	9.70E-04	4.14E-02
ENSCAFG00000019766	<i>CEPT1</i>	0.776628264	9.73E-04	4.14E-02
ENSCAFG00000009407	<i>LPCAT2</i>	1.756714080	9.86E-04	4.18E-02
ENSCAFG00000011334	<i>RCC1L</i>	-0.778257461	9.89E-04	4.18E-02
ENSCAFG00000001480	<i>TRIOBP</i>	-0.975666623	1.00E-03	4.22E-02

ENSCAFG00000008520	<i>GPR171</i>	3.185886802	1.01E-03	4.24E-02
ENSCAFG00000012014	<i>SH3BP4</i>	-1.336205269	1.02E-03	4.25E-02
ENSCAFG00000013191	<i>NRP2</i>	1.182057627	1.02E-03	4.25E-02
ENSCAFG00000017712	<i>IGSF6</i>	2.005602154	1.02E-03	4.25E-02
ENSCAFG00000002762	<i>EML6</i>	2.209226310	1.03E-03	4.25E-02
ENSCAFG00000003709	<i>CD37</i>	1.618904538	1.03E-03	4.25E-02
ENSCAFG00000012934	<i>MAP3K3</i>	-0.695822133	1.03E-03	4.25E-02
ENSCAFG00000030892	<i>PERP</i>	-3.465986664	1.03E-03	4.25E-02
ENSCAFG00000030284		7.746740090	1.04E-03	4.27E-02
ENSCAFG00000007143	<i>ASAHI</i>	1.143954981	1.05E-03	4.27E-02
ENSCAFG00000003663	<i>GSN</i>	-1.96568204	1.05E-03	4.27E-02
ENSCAFG00000014222	<i>IFT80</i>	0.686198223	1.05E-03	4.27E-02
ENSCAFG00000002233	<i>NPR2</i>	-0.812039114	1.06E-03	4.28E-02
ENSCAFG00000006830	<i>MSR1</i>	2.199623954	1.06E-03	4.28E-02
ENSCAFG00000011077	<i>ADARB1</i>	-1.016605532	1.07E-03	4.29E-02
ENSCAFG00000011353	<i>DENND1B</i>	1.112480125	1.07E-03	4.29E-02
ENSCAFG00000014295	<i>SMC4</i>	1.009362538	1.08E-03	4.31E-02
ENSCAFG00000014164	<i>MDGA2</i>	4.543286536	1.08E-03	4.31E-02
ENSCAFG00000002420	<i>MEOX2</i>	-1.673905672	1.09E-03	4.34E-02
ENSCAFG00000000157	<i>DCC</i>	3.488890293	1.10E-03	4.34E-02
ENSCAFG00000018169	<i>CCL14</i>	-2.888577651	1.10E-03	4.34E-02
ENSCAFG00000018857	<i>GPC4</i>	-1.504300283	1.10E-03	4.34E-02
ENSCAFG00000029541	<i>ATF5</i>	-1.069695439	1.10E-03	4.34E-02
ENSCAFG00000006150	<i>PANK2</i>	0.858186447	1.11E-03	4.34E-02
ENSCAFG00000010003	<i>MOB2</i>	-1.01718631	1.11E-03	4.34E-02
ENSCAFG00000032319		3.165475904	1.12E-03	4.36E-02
ENSCAFG00000016689	<i>CLTB</i>	-0.905903874	1.14E-03	4.36E-02
ENSCAFG00000011782	<i>GFRA1</i>	2.559935085	1.14E-03	4.36E-02
ENSCAFG00000040020		4.424728541	1.14E-03	4.36E-02
ENSCAFG00000013697	<i>C17H1orf162</i>	1.990612908	1.15E-03	4.36E-02
ENSCAFG00000032328		7.943324878	1.15E-03	4.36E-02
ENSCAFG00000011286	<i>OSTC</i>	0.766024958	1.15E-03	4.36E-02
ENSCAFG00000003855	<i>LDAH</i>	0.620361687	1.15E-03	4.36E-02
ENSCAFG00000020411	<i>NEGR1</i>	4.740391257	1.15E-03	4.36E-02
ENSCAFG00000014255		-1.088579211	1.15E-03	4.36E-02
ENSCAFG00000009615	<i>STK4</i>	0.774839996	1.15E-03	4.36E-02
ENSCAFG00000012556	<i>HTRA1</i>	1.984362820	1.15E-03	4.36E-02
ENSCAFG00000012489	<i>CASTOR1</i>	-1.477718249	1.16E-03	4.36E-02
ENSCAFG00000008477	<i>BLNK</i>	1.425747712	1.16E-03	4.37E-02
ENSCAFG00000020251	<i>NOB1</i>	-0.710541181	1.18E-03	4.41E-02
ENSCAFG00000002131	<i>ERMP1</i>	1.029399516	1.19E-03	4.43E-02

ENSCAFG00000005210	<i>IL2RA</i>	2.676936303	1.19E-03	4.43E-02
ENSCAFG00000003252	<i>AOAH</i>	2.665948188	1.19E-03	4.43E-02
ENSCAFG000000033317		4.065095303	1.20E-03	4.44E-02
ENSCAFG000000011474	<i>FCMR</i>	3.061666908	1.20E-03	4.45E-02
ENSCAFG000000005109	<i>CXCR4</i>	2.978250701	1.21E-03	4.47E-02
ENSCAFG000000004355	<i>LIMS2</i>	-1.443418543	1.21E-03	4.47E-02
ENSCAFG000000031795		5.862947677	1.22E-03	4.48E-02
ENSCAFG000000010684	<i>LINS1</i>	0.717028060	1.22E-03	4.48E-02
ENSCAFG000000037363		2.601006832	1.24E-03	4.52E-02
ENSCAFG000000010157	<i>BACE2</i>	-0.747302629	1.24E-03	4.52E-02
ENSCAFG000000000876	<i>PARVG</i>	1.910027110	1.25E-03	4.53E-02
ENSCAFG000000013278		-2.011793109	1.25E-03	4.53E-02
ENSCAFG000000017315	<i>SEL1L</i>	1.042945265	1.25E-03	4.53E-02
ENSCAFG000000007034	<i>LYN</i>	1.216517139	1.27E-03	4.56E-02
ENSCAFG000000002999	<i>ZMPSTE24</i>	0.616130990	1.27E-03	4.56E-02
ENSCAFG000000004615	<i>APOC1</i>	3.256107273	1.27E-03	4.56E-02
ENSCAFG000000006909	<i>LMO1</i>	1.436784793	1.30E-03	4.65E-02
ENSCAFG000000010177	<i>KDM1B</i>	0.759674389	1.31E-03	4.66E-02
ENSCAFG000000029740	<i>MSRB3</i>	-0.956781659	1.31E-03	4.66E-02
ENSCAFG000000013774	<i>MAP3K14</i>	-1.082900598	1.32E-03	4.66E-02
ENSCAFG000000017259	<i>MAGT1</i>	0.819404867	1.32E-03	4.66E-02
ENSCAFG000000029752	<i>C11H5orf15</i>	0.880069263	1.32E-03	4.66E-02
ENSCAFG000000023802	<i>ARMCX1</i>	-1.162969916	1.33E-03	4.66E-02
ENSCAFG000000007503	<i>SBNO1</i>	0.757325106	1.33E-03	4.66E-02
ENSCAFG000000003198	<i>FKBP15</i>	0.654482450	1.34E-03	4.68E-02
ENSCAFG000000003004	<i>PPT1</i>	1.140460162	1.34E-03	4.68E-02
ENSCAFG000000029976	<i>DNAJB9</i>	1.162154252	1.35E-03	4.69E-02
ENSCAFG000000024944		3.163963024	1.35E-03	4.69E-02
ENSCAFG000000029534	<i>BLOC1S2</i>	0.668503758	1.36E-03	4.69E-02
ENSCAFG000000009506	<i>ABI3BP</i>	2.689304831	1.36E-03	4.69E-02
ENSCAFG000000003056	<i>PLEKHA8</i>	1.148424698	1.36E-03	4.69E-02
ENSCAFG0000000031724	<i>CACNG5</i>	-2.831629768	1.37E-03	4.70E-02
ENSCAFG000000014175	<i>ITGA4</i>	2.015414652	1.37E-03	4.70E-02
ENSCAFG000000004887		-1.384185562	1.37E-03	4.70E-02
ENSCAFG000000031832	<i>CCDC115</i>	0.635666577	1.39E-03	4.75E-02
ENSCAFG000000024087	<i>CGAS</i>	2.002637282	1.40E-03	4.75E-02
ENSCAFG000000002941	<i>ME1</i>	0.756090318	1.40E-03	4.75E-02
ENSCAFG000000013782	<i>CCRL2</i>	1.914764347	1.40E-03	4.75E-02
ENSCAFG0000000032614		2.204900771	1.41E-03	4.75E-02
ENSCAFG000000012780	<i>CD3G</i>	2.776625264	1.41E-03	4.75E-02
ENSCAFG000000031403		7.628902642	1.41E-03	4.75E-02

ENSCAFG00000005419	<i>PNP</i>	-0.86489843	1.42E-03	4.76E-02
ENSCAFG000000034279		-1.54729037	1.43E-03	4.78E-02
ENSCAFG000000008401	<i>P2RX7</i>	1.397050037	1.44E-03	4.78E-02
ENSCAFG000000012781	<i>XPR1</i>	0.732539865	1.44E-03	4.78E-02
ENSCAFG000000030935		8.623120221	1.44E-03	4.78E-02
ENSCAFG000000009592	<i>NELL2</i>	5.727098265	1.44E-03	4.78E-02
ENSCAFG000000031101	<i>CD48</i>	1.655372263	1.45E-03	4.79E-02
ENSCAFG000000012892	<i>ERP27</i>	3.061820381	1.46E-03	4.79E-02
ENSCAFG000000018063	<i>ACSL4</i>	0.985470788	1.46E-03	4.79E-02
ENSCAFG000000008279	<i>PTPRJ</i>	1.573381321	1.46E-03	4.79E-02
ENSCAFG000000003184	<i>CEP68</i>	-1.088652083	1.49E-03	4.82E-02
ENSCAFG000000006546	<i>ARID5A</i>	-1.054388677	1.49E-03	4.82E-02
ENSCAFG000000012789	<i>PTPRO</i>	1.505502333	1.49E-03	4.82E-02
ENSCAFG000000011099	<i>AMZ2</i>	1.034499541	1.50E-03	4.82E-02
ENSCAFG000000012293	<i>ADCY4</i>	-1.426991836	1.50E-03	4.82E-02
ENSCAFG000000031177	<i>ST6GALNAC6</i>	-0.954971176	1.51E-03	4.82E-02
ENSCAFG000000003383	<i>ADAM17</i>	0.950596829	1.51E-03	4.82E-02
ENSCAFG000000011395	<i>NCF1</i>	1.675863852	1.51E-03	4.82E-02
ENSCAFG000000030001		6.875948787	1.52E-03	4.82E-02
ENSCAFG000000009912	<i>ERG</i>	-1.162903042	1.52E-03	4.82E-02
ENSCAFG000000001215	<i>GFRA3</i>	-1.926455064	1.52E-03	4.82E-02
ENSCAFG000000009129		0.890292581	1.52E-03	4.82E-02
ENSCAFG000000019698	<i>PALM</i>	-2.489277872	1.53E-03	4.82E-02
ENSCAFG000000015182	<i>GLB1L</i>	1.349657740	1.53E-03	4.82E-02
ENSCAFG000000007167	<i>CHD7</i>	0.797599143	1.53E-03	4.82E-02
ENSCAFG000000015982	<i>CDC42BPA</i>	-0.866542301	1.53E-03	4.82E-02
ENSCAFG000000005738	<i>SLC30A6</i>	0.690589889	1.53E-03	4.82E-02
ENSCAFG000000031478		0.798736773	1.54E-03	4.82E-02
ENSCAFG000000017669	<i>BTK</i>	1.539784611	1.54E-03	4.82E-02
ENSCAFG000000010603	<i>RGS1</i>	2.147254304	1.54E-03	4.82E-02
ENSCAFG000000031968	<i>C5H11orf52</i>	2.111748866	1.54E-03	4.82E-02
ENSCAFG000000014603	<i>TRIM9</i>	2.862697852	1.55E-03	4.82E-02
ENSCAFG000000012086	<i>CTSS</i>	2.149518683	1.57E-03	4.87E-02
ENSCAFG000000011265	<i>PTPRC</i>	1.815104855	1.59E-03	4.95E-02
ENSCAFG000000006502	<i>TPP1</i>	1.550571291	1.60E-03	4.95E-02
ENSCAFG000000016231	<i>FUT8</i>	1.238204511	1.61E-03	4.99E-02

Supplementary Table 9.3S. Functional annotation using DAVID identifies a cluster of 141 genes for proteins with a transmembrane helical domain.

UniProt Knowledgebase Keyword: Transmembrane Helix
<i>P</i> _{adj} = 1.8E-06
<i>ADAM17</i>
<i>ADAM28</i>
<i>ADAM9</i>
<i>ATP6V0A1</i>
<i>ATP6V0B</i>
<i>ATP8A1</i>
<i>ATP2C1</i>
<i>BOC</i>
<i>CCRL2</i>
<i>CXCR4</i>
<i>CX3CR1</i>
<i>CLEC12A</i>
<i>CLEC2D</i>
<i>CLEC7A</i>
<i>CD151</i>
<i>CD2</i>
<i>CD37</i>
<i>CD3G</i>
<i>CD48</i>
<i>CD72</i>
<i>CD80</i>
<i>CD84</i>
<i>CD86</i>
<i>DCC</i>
<i>DRAM2</i>
<i>FCMR</i>
<i>GPR171</i>
<i>GPR180</i>
<i>GPR65</i>
<i>GIMAP2</i>
<i>LRP11</i>
<i>MDGA2</i>
<i>NIPAL2</i>
<i>NDFIP1</i>

<i>PERP</i>
<i>SEC11C</i>
<i>SEL1L</i>
<i>SLAMF6</i>
<i>SLAMF7</i>
<i>ST6GALNAC6</i>
<i>ST8SIA4</i>
<i>TNFRSF17</i>
<i>VSIG10</i>
<i>ACP2</i>
<i>ALCAM</i>
<i>ADCY4</i>
<i>ANKRD22</i>
<i>AQP9</i>
<i>ACKR3</i>
<i>BACE2</i>
<i>CDH20</i>
<i>CACNG5</i>
<i>CERS4</i>
<i>CCR1</i>
<i>CEPT1</i>
<i>C11H5orf15</i>
<i>CLDN1</i>
<i>CSF2RA</i>
<i>DERL3</i>
<i>ERMP1</i>
<i>FAXDC2</i>
<i>FNDC3A</i>
<i>FUT8</i>
<i>GRM8</i>
<i>GLRB</i>
<i>GPC4</i>
<i>IGSF6</i>
<i>ITGA4</i>
<i>ITGAL</i>
<i>ITGAX</i>
<i>ITGB8</i>

IL18R1
IL2RA
IL21R
KLRG1
KMO
LRRC19
LRRC25
LETM2
LY9
LRMP
LPAR3
LPAR6
LPCAT1
LPCAT2
MSR1
MAGT1
MFSD11
DLA-DMA
DLA-DMB
DLA-DQA1
M6PR
MILR1
MOSPD2
NPR2
NRP2
NNT
NEMP1
OSTC
OSTM1
OLR1
PTCH1
PTCH2
PHEX
PEAR1
PLXNC1
GALNT3
KCNK2

<i>PTGES</i>
<i>PTGFR</i>
<i>PTPRC</i>
<i>PTPRJ</i>
<i>PTPRO</i>
<i>PCDH9</i>
<i>P2RX7</i>
<i>SELL</i>
<i>SLC16A6</i>
<i>SLC18A2</i>
<i>SLC25A19</i>
<i>SLC30A6</i>
<i>SLC37A2</i>
<i>SLC4A7</i>
<i>SLC46A3</i>
<i>SLC7A7</i>
<i>SORCS3</i>
<i>SGPP1</i>
<i>TENM3</i>
<i>TLR2</i>
<i>TLR7</i>
<i>TLR1</i>
<i>TGFBR1</i>
<i>TMTC1</i>
<i>TMEM144</i>
<i>TMEM206</i>
<i>TMEM54</i>
<i>TREM1</i>
<i>TRIM59</i>
<i>VRK2</i>
<i>XPR1</i>
<i>ZDHHC22</i>
<i>ZMPSTE24</i>

Chapter 10

Summary and Future Directions

Personalized medicine refers to the concept that medical counseling, intervention, and treatment can be tailored to the individual, through identification of subpopulations that differ in susceptibility to disease and their ability to respond to treatment. This also allows preventative measures to be targeted to a specific group of high-risk individuals, which produces an ideal outcome, reduction in morbidity and mortality, while also saving time and resources [Hong 2012].

Personalized veterinary medicine is an especially nascent field, though it is becoming more widely recognized that veterinarians and our patients have much to offer the broader scientific community, and we have much to gain from the adoption of practices, from electronic medical records to next generation sequencing, that allow integration of bioinformatics approaches into veterinary practice [Lloyd 2016; Mealy 2019]. The ultimate goal of this project, which will continue beyond the scope of this dissertation, is to provide a personalized medical approach for our veterinary ACLR patients, and to use the information we gain from the study of dogs with ACLR to better understand the biological underpinnings of ACLR in human beings. The work that we have undertaken so far has led us closer to this goal, but there is much work to be done.

When we began this project in 2012, the prevailing view was that complex diseases like ACLR in the dog would probably be associated with a smaller number of genetic variants that have large effects, and these large-effect variants would be relatively easy to identify with modest sample sizes [Lequarre 2011; Rowell 2011; Ostrander 2012]. If there is one hypothesis that has been strongly supported by each of our studies, it is that ACLR in the dog is a highly polygenic complex disease, and, barring possible effects from rare variants, structural variation, or epigenetic changes that a SNP chip cannot identify, there does not appear to be a large-effect variant that is associated with this disease in the Labrador Retriever. Rather, our research supports the notion that many thousands of variants, each with a small individual effects on phenotypic variance, explain genetic predisposition to ACLR in the dog. This is similar to the

reported genetic architectures of complex diseases in human beings [Visscher 2017]. It is possible to identify these small-effect variants, however, and they can be exploited to understand the disease process and potentially predict individuals who are at high risk for developing the disease.

We have now performed three separate GWA studies, each using slightly different statistical approaches with the goal of maximizing power in our modest dataset to identify genetic variants that point to key molecular pathways underlying disease risk. So far, specific findings have not been repeatable from one study to the next, though our population of Labrador Retrievers are similar. While our dataset has grown larger as we continue to recruit dogs to the project, this indicates that we likely remain underpowered to detect with certainty the many small effect variants that contribute to ACLR phenotypic variance. Rather than focus on individual variants, we focus on the molecular patterns that are consistent from one study to the next.

Genes involved in the inflammatory cascade of the innate immune system appear to be playing some role in ACLR pathophysiology. In Chapter 3 [Baker 2017], we identified two functional annotation clusters, one involved in innate immunity and the other was specific to c-type lectins, many of which were pattern recognition receptors that are also important in the innate immune system. In Chapter 4 [Baker 2018], we identified a variant within *DOCK2*, which is important for allowing lymphocytes to migrate into tissues. Lymphoplasmacytic synovitis is present at the time of ACLR diagnosis in the dog [Bleedorn 2011]. In Chapter 9, several of the genes within the top 50 gene effects also have roles in the immune system, including *HLA-DPB1*, which is expressed on antigen-presenting cells including lymphocytes. The hypothesis that immunopathological mechanisms may be in part responsible for weakening and eventual rupture of the ACL in dogs is not new. The synovial immune response is thought to promote progressive degradation of intra-articular structures, including the cruciate ligaments [Doom 2008], though our work is one of the first to provide evidence that genetics may play a role in

initiating or promoting this immune response in the dog. In human beings there is less evidence that synovitis and accompanying joint degeneration is present at the time of ACLR, however, ACLR is associated with post-traumatic osteoarthritis (PTOA). The response of the synovium, cartilage, and other tissues to ACLR appears to trigger a pro-inflammatory cascade that is generally considered to be irreversible and eventually leads to PTOA and end-stage joint disease [Maerz 2018]. PTOA is considered a highly polygenic complex disease in humans [Peffer 2017], and a recent GWAS of PTOA in human beings identified a variant in pseudogene *LSP1P3* (lymphocyte specific protein 1 pseudogene 3) [Yau 2017], perhaps supporting a role for genetic regulation of the immune system in this process.

Proteins involved in the extracellular matrix may also play a role in ACLR. Two other key genes in the c-type lectin class identified in Chapter 3 [Baker 2017] include *ACAN* and *HAPLN3*, which are important proteoglycans for maintaining moisture in collagenous tissues [Mannion 2013]. A much stronger association with a variant within *ACAN* and ACLR was identified again using BayesRC in Chapter 9. The connection between *ACAN* and ACLR in the dog is especially interesting as *ACAN* has also been connected to ACLR in humans [Mannion 2013; Johnson 2015] and degenerative ligament disease in horses [Plaas 2011]. We also identified variants in other extracellular matrix proteins, including many collagen genes, using BayesRC analysis in Chapter 9. Several papers in both humans and dogs have reported associations between variants within collagen genes and ACLR (see Table 9.2).

Through this work, actin dynamics has also emerged as another molecular process that may play a role in risk of ACLR. In [Baker 2018] we identified an association with the gene *ROR2*, which is important for patterning during embryonic limb development, and later plays a role in cellular migration and the ability for tissues to heal. *ROR2* does this through regulation of actin-cytoskeletal dynamics [Bai 2011; Roarty 2015]. Though we did not repeat the specific association with *ROR2* in our study described in Chapter 9, we did identify several genes that also play a role in actin-cytoskeletal dynamics, including the largest-effect gene identified,

FNBP1. The role of actin cytoskeleton dynamics in the cruciate ligament is not well-characterized. The actin cytoskeleton plays a role in collagen deposition in the extracellular matrix in tendon [Canty 2006], and there is evidence of a well-organized actin cytoskeleton network that is closely tied to elastic properties of embryonic tendons [Schiele 2015]. Tendons and ligaments are similar tissues, so it is likely that actin plays a similar role in ligament, including the ACL, though there does not appear to be definitive evidence of this.

The largest limiting factor in this work has been sample size. Due to extensive linkage disequilibrium in the dog, the sample sizes required for GWAS of complex diseases like ACLR in the dog model are probably much smaller than the hundreds of thousands of subjects needed for comparable human diseases, but it is not yet known what the ideal number would be. Given the highly polygenic nature of this disease, and previous InPower analysis [Baker 2017], this number is likely to be in the range of 1,000-2,000 dogs. A recently-funded wave of recruitment to this project is currently underway, which will bring our sample size into this range with approximately 1,000 dogs. In human GWAS there exists a sample size threshold after which discoveries stabilize and begin to grow in size, explaining greater and greater proportions of estimated heritability [Simons 2018]. This larger sample size will allow us to create an updated estimate of heritability of ACLR in the Labrador Retriever. As our sample size grows, it is our expectation that SNPs with the largest effect sizes will begin to stabilize. As we are better able to model SNP effect sizes, our prediction estimates like those discussed in Chapter 6 are also likely to improve and stabilize, which will lay the groundwork for a clinically useful genetic test. ACLR is the most common cause of pelvic limb lameness in the dog, and >\$1 billion is spent annually on treatment of this dog's with ACLR [Wilke 2005]. Thus, such a test would have a broad impact on veterinary medicine, and it would have an incredible economic impact, because it would provide the opportunity to alter non-genetic risk factors (e.g. neutering before 1 year of age) to reduce the likelihood that disease will develop at all. It will also provide the opportunity

for owners and breeders to make breeding decisions regarding high-risk dogs that could reduce population prevalence of ACLR.

Ultimately, the discoveries we make in the dog will need to be validated in human populations. Such work could be done by evaluating GWAS summary statistics to identify associations that are shared between species. Additionally, candidate gene studies could be used to evaluate risk genes that appear to have a high likelihood of effect in both species. Given the similarities between the clinical presentation and progression of ACLR in dogs and humans, we expect that the discoveries made through this project will identify molecular pathways that can be exploited for the development of novel disease-modifying therapies that will either prevent or slow progression of disease. These discoveries would have broad impact to improve biology and understanding of ACLR in both human and veterinary medicine, highlighting the importance of *one health* research.

References

- Abel SB, Hammer DL, Shott S. Use of the proximal portion of the tibia for measurement of the tibial plateau angle in dogs. *American Journal of Veterinary Research*. 2003;64(9):1117-23.
- Acikel C, Son YA, Celik C, Gul H. Evaluation of potential novel variations and their interactions related to bipolar disorders: analysis of genome-wide association study data. *Neuropsychiatric disease and treatment*. 2016;12:2997-3004.
- Adams P, Bolus R, Middleton S, Moores AP, Grierson J. Influence of signalment on developing cranial cruciate rupture in dogs in the UK. *Journal of Small Animal Practice*. 2011;52(7):347-52.
- Aiken SW, Kass PH, Toombs JP. Intercondylar notch width in dogs with and without cranial cruciate ligament injuries. *Veterinary and Comparative Orthopaedics and Traumatology*. 1995;8(3):128-32.
- Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. *Nature Protocols*. 2010;5(9):1564.
- Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
- Arendt E, Dick R. Knee injury patterns among men and women in collegiate basketball and soccer: NCAA data and review of literature. *The American Journal of Sports Medicine*. 1995;23(6):694-701.
- Aspenström P. Formin-binding proteins: modulators of formin-dependent actin polymerization. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*. 2010;1803(2):174-82.
- Bai SW, Herrera-Abreu MT, Rohn JL, Racine V, Tajadura V, Suryavanshi N, Bechtel S, Wiemann S, Baum B, Ridley AJ. Identification and characterization of a set of conserved and new regulators of cytoskeletal organization, cell morphology and migration. *BMC Biology*. 2011;9(1):54.
- Baird AE, Carter SD, Innes JF, Ollier W, Short A. Genome-wide association study identifies genomic regions of association for cruciate ligament rupture in Newfoundland dogs. *Animal genetics*. 2014a;45:542-9.
- Baird AE, Carter SD, Innes JF, Ollier WE, Short AD. Genetic basis of cranial cruciate ligament rupture (CCLR) in dogs. *Connective Tissue Research*. 2014b;55(4):275-81.
- Baker LA, Kirkpatrick B, Rosa GJ, Gianola D, Valente B, Sumner JP, Baltzer W, Hao Z, Binversie EE, Volstad N, Piazza A. Genome-wide association analysis in dogs implicates 99 loci as risk variants for anterior cruciate ligament rupture. *PLoS One*. 2017;12(4):e0173810.

- Baker LA, Rosa GJ, Hao Z, Piazza A, Hoffman C, Binversie EE, Sample SJ, Muir P. Multivariate genome-wide association analysis identifies novel and relevant variants associated with anterior cruciate ligament rupture risk in the dog model. *BMC Genetics*. 2018;19(1):39.
- Belanger JM, Bellumori TP, Bannasch DL, Famula TR, Oberbauer AM. Correlation of neuter status and expression of heritable disorders. *Canine Genetics and Epidemiology*. 2017;4(1):6.
- Benz PM, Ding Y, Stingl H, Loot AE, Zink J, Wittig I, Popp R, Fleming I. AKAP12 deficiency impairs VEGF-induced endothelial cell migration and sprouting. *Acta Physiologica*. 2019:e13325.
- Bergman RN, Kim SP, Hsu IR, Catalano KJ, Chiu JD, Kabir M, Richey JM, Ader M. Abdominal obesity: role in the pathophysiology of metabolic disease and cardiovascular risk. *The American Journal of Medicine*. 2007;120(2):S3-8.
- Beynon BD, Johnson RJ, Braun S, Sargent M, Bernstein IM, Skelly JM, Vacek PM. The relationship between menstrual cycle phase and anterior cruciate ligament injury: a case-control study of recreational alpine skiers. *The American Journal of Sports Medicine*. 2006;34(5):757-64.
- Beynon BD, Shultz SJ. Anatomic alignment, menstrual cycle phase, and the risk of anterior cruciate ligament injury. *Journal of Athletic Training*. 2008;43(5):541-2.
- Björnerfeldt S, Hailer F, Nord M, Vilà C. Assortative mating and fragmentation within dog breeds. *BMC Evolutionary Biology*. 2008;8(1):28.
- Bleedorn JA, Greuel EN, Manley PA, Schaefer SL, Markel MD, Holzman G, Muir P. Synovitis in dogs with stable stifle joints and incipient cranial cruciate ligament rupture: a cross-sectional study. *Veterinary Surgery*. 2011;40(5):531-43.
- Botta V, Louppe G, Geurts P, Wehenkel L. Exploiting SNP correlations within random forest for genome-wide association studies. *PLoS one*. 2014;9(4):e93379.
- Bowman CE, Wolfgang MJ. Role of the malonyl-CoA synthetase ACSF3 in mitochondrial metabolism. *Advances in Biological Regulation*. 2018;71:34-40.
- Brophy RH, Schmitz L, Wright RW, Dunn WR, Parker RD, Andrich JT, McCarty EC, Spindler KP. Return to play and future ACL injury risk after ACL reconstruction in soccer athletes from the Multicenter Orthopaedic Outcomes Network (MOON) group. *The American Journal of Sports Medicine*. 2012;40(11):2517-22.
- Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Patterson N, Daly MJ, Price AL, Neale BM, Schizophrenia Working Group of the Psychiatric Genomics Consortium. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*. 2015;47(3):291-97.
- Bulik-Sullivan BK. LDSC Frequently Asked Questions [Internet]. Github. 2017 [cited August 2019]. Available from: <https://github.com/bulik/ldsc/wiki/FAQ>.

- Buote N, Fusco J, Radasch R. Age, tibial plateau angle, sex, and weight as risk factors for contralateral rupture of the cranial cruciate ligament in Labradors. *Veterinary Surgery*. 2009;38(4):481-9.
- Brandon ML, Haynes PT, Bonamo JR, Flynn MI, Barrett GR, Sherman MF. The association between posterior-inferior tibial slope and anterior cruciate ligament insufficiency. *Arthroscopy: The Journal of Arthroscopic & Related Surgery*. 2006;22(8):894-9.
- Breiman L. Random forests. *Machine Learning*. 2001;45(1):5-32.
- Browning BL, Browning SR. Genotype imputation with millions of reference samples. *The American Journal of Human Genetics*. 2016;98(1):116-26.
- Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes & Development*. 2011;25(18):1915-27.
- Cabrera SY, Owen TJ, Mueller MG, Kass PH. Comparison of tibial plateau angles in dogs with unilateral versus bilateral cranial cruciate ligament rupture: 150 cases (2000–2006). *Journal of the American Veterinary Medical Association*. 2008;232(6):889-92.
- Canty EG, Starborg T, Lu Y, Humphries SM, Holmes DF, Meadows RS, Huffman A, O'Toole ET, Kadler KE. Actin filaments are required for fibroblast-mediated collagen fibril alignment in tendon. *Journal of Biological Chemistry*. 2006;281(50):38592-8.
- Carter P, Tydall A, Dudhia J, Handley CJ. Proteoglycans and catabolic products of proteoglycans present in ligament. *Biochemical Journal*. 2005;385(2):381-8.
- Calboli FC, Sampson J, Fretwell N, Balding DJ. Population structure and inbreeding from pedigree analysis of purebred dogs. *Genetics*. 2008;179(1):593-601.
- Cameron ML, Fu FH, Paessler HH, Schneider M, Evans CH. Synovial fluid cytokine concentrations as possible prognostic indicators in the ACL-deficient knee. *Knee Surgery, Sports Traumatology, Arthroscopy*. 1994;2(1):38-44.
- Cangelosi R, Goriely A. Component retention in principal component analysis with application to cDNA microarray data. *Biology Direct*. 2007;2(1):2.
- Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4(1):7.
- Chase K, Sargan D, Miller K, Ostrander EA, Lark KG. Understanding the genetics of autoimmune disease: two loci that regulate late onset Addison's disease in Portuguese Water Dogs. *International Journal of Immunogenetics*. 2006;33(3):179-84.
- Chatterjee N, Shi J, García-Closas M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nature Reviews Genetics*. 2016;17(7):392-406.
- Chen H, Poon A, Yeung C, Helms C, Pons J, Bowcock AM, Kwok PY, Liao W. A genetic risk score combining ten psoriasis risk loci improves disease prediction. *PLoS One*. 2011;6(4):e19454.

- Chen T, He T, Benesty M, Khotilovich V, Tang Y. Xgboost: extreme gradient boosting. R package version 0.4-2. 2015 Aug 1:1-4.
- Chuang C, Ramaker MA, Kaur S, Csomos RA, Kroner KT, Bleedorn JA, Schaefer SL, Muir P. Radiographic risk factors for contralateral rupture in dogs with unilateral cranial cruciate ligament rupture. *PLoS One*. 2014;9(9):e106389.
- Collins M, Raleigh SM. Genetic risk factors for musculoskeletal soft tissue injuries. In: Collins, M. *Genetics and Sports*. Karger Publishers; 2009:136-149.
- Comerford EJ, Tarlton JF, Avery NC, Bailey AJ, Innes JF. Distal femoral intercondylar notch dimensions and their relationship to composition and metabolism of the canine anterior cruciate ligament. *Osteoarthritis and Cartilage*. 2006;14(3):273-8.
- Comerford EJ, Smith K, Hayashi K. Update on the aetiopathogenesis of canine cranial cruciate ligament disease. *Veterinary and Comparative Orthopaedics and Traumatology*. 2011;24(02):91-8.
- Conzemius MG, Evans RB, Besancon MF, Gordon WJ, Horstman CL, Hoefle WD, Nieves MA, Wagner SD. Effect of surgical technique on limb function after surgery for rupture of the cranial cruciate ligament in dogs. *Journal of the American Veterinary Medical Association*. 2005;226(2):232-6.
- Dambuza IM, Brown GD. C-type lectins in immunity: recent developments. *Current Opinion in Immunology*. 2015;32:21-7.
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G. The variant call format and VCFtools. *Bioinformatics*. 2011;27(15):2156-8.
- De Bruin C, Meij BP, Kooistra HS, Hanson JM, Lamberts SW, Hofland LJ. Cushing's disease in dogs and humans. *Hormone Research in Paediatrics*. 2009;71:140-3.
- De Bruin T, De Rooster H, Bosmans T, Duchateau L, van Bree H, Gielen I. Radiographic assessment of the progression of osteoarthritis in the contralateral stifle joint of dogs with a ruptured cranial cruciate ligament. *Veterinary Record*. 2007;161(22):745-50.
- DeChiara TM, Kimble RB, Poueymirou WT, Rojas J, Masiakowski P, Valenzuela DM, Yancopoulos GD. Ror2, encoding a receptor-like tyrosine kinase, is required for cartilage and growth plate development. *Nature Genetics*. 2000;24(3):271.
- Deinhardt K, Kim T, Spellman DS, Mains RE, Eipper BA, Neubert TA, Chao MV, Hempstead BL. Neuronal growth cone retraction relies on proneurotrophin receptor signaling through Rac. *Sci. Signal.*. 2011 Dec 6;4(202):ra82-.
- de Los Campos G, Hickey JM, Pong-Wong R, Daetwyler HD, Calus MP. Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics*. 2013;193(2):327-45.

- den Hollander W, Ramos YF, Bomer N, Elzinga S, van der Breggen R, Lakenberg N, de Dijcker WJ, Suchiman HE, Duijnisveld BJ, Houwing-Duistermaat JJ, Slagboom PE. Transcriptional associations of osteoarthritis-mediated loss of epigenetic control in articular cartilage. *Arthritis & Rheumatology*. 2015;67(8):2108-16.
- Dennler R, Kipfer NM, Tepic S, Hassig M, Montavon PM. Inclination of the patellar ligament in relation to flexion angle in stifle joints of dogs without degenerative joint disease. *American Journal of Veterinary Research*. 2006;67(11):1849-54.
- Dismukes DI, Fox DB, Tomlinson JL, Cook JL, Essman SC. Determination of pelvic limb alignment in the large-breed dog: a cadaveric radiographic study in the frontal plane. *Veterinary Surgery*. 2008;37(7):674-82.
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15-21.
- Dohrn MF, Röcken C, De Bleecker JL, Martin JJ, Vorgerd M, Van den Bergh PY, Ferbert A, Hinderhofer K, Schröder JM, Weis J, Schulz JB. Diagnostic hallmarks and pitfalls in late-onset progressive transthyretin-related amyloid-neuropathy. *Journal of Neurology*. 2013;260(12):3093-108.
- Doom M, De Bruin T, De Rooster H, van Bree H, Cox E. Immunopathological mechanisms in dogs with rupture of the cranial cruciate ligament. *Veterinary Immunology and Immunopathology*. 2008;125(1-2):143-61.
- Döring AK, Junginger J, Hewicker-Trautwein M. Cruciate ligament degeneration and stifle joint synovitis in 56 dogs with intact cranial cruciate ligaments: correlation of histological findings and numbers and phenotypes of inflammatory cells with age, body weight and breed. *Veterinary Immunology and Immunopathology*. 2018;196:5-13.
- Duerr FM, Duncan CG, Savicky RS, Park RD, Egger EL, Palmer RH. Risk factors for excessive tibial plateau angle in large-breed dogs with cranial cruciate ligament disease. *Journal of the American Veterinary Medical Association*. 2007;231(11):1688-91.
- Duval JM, Budsberg SC, Flo GL, Sammarco JL. Breed, sex, and body weight as risk factors for rupture of the cranial cruciate ligament in young dogs. *Journal of the American Veterinary Medical Association*. 1999;215(6):811-4.
- Edwards M, Liang Y, Kim T, Cooper JA. Physiological role of the interaction between CARMIL1 and capping protein. *Molecular Biology of the Cell*. 2013;24(19):3047-55.
- Ekenstedt KJ, Minor KM, Rendahl AK, Conzemius MG. DNM1 mutation status, sex, and sterilization status of a cohort of Labrador retrievers with and without cranial cruciate ligament rupture. *Canine Genetics and Epidemiology*. 2017;4(1):2.
- Erbe M, Hayes BJ, Matukumalli LK, Goswami S, Bowman PJ, Reich CM, Mason BA, Goddard ME. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. *Journal of Dairy Science*. 2012;95(7):4114-29.

- Erne JB, Goring RL, Kennedy FA, Schoenborn WC. Prevalence of lymphoplasmacytic synovitis in dogs with naturally occurring cranial cruciate ligament rupture. *Journal of the American Veterinary Medical Association*. 2009;235(4):386-90.
- Fall T, Hamlin HH, Hedhammar Å, Kämpe O, Egenvall A. Diabetes mellitus in a population of 180,000 insured dogs: incidence, survival, and breed distribution. *Journal of Veterinary Internal Medicine*. 2007;21(6):1209-16.
- Ficek K, Cieszczyk P, Kaczmarczyk M, Maciejewska-Karłowska A, Sawczuk M, Cholewinski J, Leonska-Duniec A, Stepien-Słodkowska M, Zarebska A, Stepto NK, Bishop DJ. Gene variants within the COL1A1 gene are associated with reduced anterior cruciate ligament injury in professional soccer players. *Journal of Science and Medicine in Sport*. 2013;16(5):396-400.
- Flynn RK, Pedersen CL, Birmingham TB, Kirkley A, Jackowski D, Fowler PJ. The familial predisposition toward tearing the anterior cruciate ligament: a case control study. *The American Journal of Sports Medicine*. 2005;33(1):23-8.
- Ford CE, Qian Ma SS, Quadir A, Ward RL. The dual role of the novel Wnt receptor tyrosine kinase, ROR2, in human carcinogenesis. *International Journal of Cancer*. 2013;133(4):779-87.
- Friedenberg SG, Meurs KM. Genotype imputation in the domestic dog. *Mammalian Genome*. 2016;27(9-10):485-94.
- Fuller MC, Hayashi K, Bruecker KA, Holsworth IG, Sutton JS, Kass PH, Kantrowitz BJ, Kapatkin AS. Evaluation of the radiographic infrapatellar fat pad sign of the contralateral stifle joint as a risk factor for subsequent contralateral cranial cruciate ligament rupture in dogs with unilateral rupture: 96 cases (2006–2007). *Journal of the American Veterinary Medical Association*. 2014a;244(3):328-38.
- Fuller MC, Kapatkin AS, Bruecker KA, Holsworth IG, Kass PH, Hayashi K. Comparison of the tibial mechanical joint orientation angles in dogs with cranial cruciate ligament rupture. *The Canadian Veterinary Journal*. 2014b;55(8):757.
- Gallagher MD, Chen-Plotkin AS. The post-GWAS era: from association to function. *The American Journal of Human Genetics*. 2018;102(5):717-30.
- Gelber AC. Conventional medical therapy for osteoarthritis: current state of the evidence. *Current Opinion in Rheumatology*. 2015;27(3):312-7.
- Génin E. Missing heritability of complex diseases: case solved?. *Human Genetics*. 2019;4:1-1.
- Gérard HC, Wang Z, Wang GF, El-Gabalawy H, Goldbach-Mansky R, Li Y, Majeed W, Zhang H, Ngai N, Hudson AP, Schumacher HR. Chromosomal DNA from a variety of bacterial species is present in synovial tissue from patients with various forms of arthritis. *Arthritis & Rheumatism*. 2001;44(7):1689-97.
- Gershwin LJ. Veterinary autoimmunity: Autoimmune diseases in domestic animals. *Annals of the New York Academy of Sciences*. 2007;1109(1):109-16.

- Ghosh A, Zou F, Wright FA. Estimating odds ratios in genome scans: an approximate conditional likelihood approach. *The American Journal of Human Genetics*. 2008;82(5):1064-74.
- Gianola D. Priors in whole-genome regression: the Bayesian alphabet returns. *Genetics*. 2013;194(3):573-96.
- Gianotti SM, Marshall SW, Hume PA, Bunt L. Incidence of anterior cruciate ligament injury and other knee ligament injuries: a national population-based study. *Journal of Science and Medicine in Sport*. 2009;12(6):622-7.
- Gibbon A, Saunders CJ, Collins M, Gamielidien J, September AV. Defining the molecular signatures of Achilles tendinopathy and anterior cruciate ligament ruptures: A whole-exome sequencing approach. *PLoS One*. 2018;13(10):e0205860.
- Giffin JR, Vogrin TM, Zantop T, Woo SL, Harner CD. Effects of increasing tibial slope on the biomechanics of the knee. *The American Journal of Sports Medicine*. 2004;32(2):376-82.
- Goldberg VM, Burstein A, Dawson M. The influence of an experimental immune synovitis on the failure mode and strength of the rabbit anterior cruciate ligament. *The Journal of Bone and Joint Surgery*. 1982;64(6):900-6.
- Gray MM, Granka JM, Bustamante CD, Sutter NB, Boyko AR, Zhu L, Ostrander EA, Wayne RK. Linkage disequilibrium and demographic history of wild and domestic canids. *Genetics*. 2009;181(4):1493-505.
- Gregory MH, Capito N, Kuroki K, Stoker AM, Cook JL, Sherman SL. A review of translational animal models for knee osteoarthritis. *Arthritis*. 2012;2012: 764621.
- Grierson J, Asher L, Grainger K. An investigation into risk factors for bilateral canine cruciate ligament rupture. *Veterinary and Comparative Orthopaedics and Traumatology*. 2011;24(03):192-6.
- Griffin LY, Agel J, Albohm MJ, Arendt EA, Dick RW, Garrett WE, Garrick JG, Hewett TE, Huston L, Ireland ML, Johnson RJ. Noncontact anterior cruciate ligament injuries: risk factors and prevention strategies. *Journal of the American Academy of Orthopaedic Surgeons*. 2000;8(3):141-50.
- Griffon DJ, Cunningham D, Gordon-Evans WJ, Tanaka R, Bruecker KA, Boudrieau RJ. Evaluation of a scoring system based on conformation factors to predict cranial cruciate ligament disease in Labrador retrievers. *Veterinary Surgery*. 2017;46(2):206-12.
- Guastella DB, Fox DB, Cook JL. Tibial plateau angle in four common canine breeds with cranial cruciate ligament rupture, and its relationship to meniscal tears. *Veterinary and Comparative Orthopaedics and Traumatology*. 2008;21(02):125-8.
- Guerrero TG, Geyer H, Hässig M, Montavon PM. Effect of conformation of the distal portion of the femur and proximal portion of the tibia on the pathogenesis of cranial cruciate ligament disease in dogs. *American Journal of Veterinary Research*. 2007;68(12):1332-7.

- Guthrie JW, Keeley BJ, Maddock E, Bright SR, May C. Effect of signalment on the presentation of canine patients suffering from cranial cruciate ligament disease. *Journal of Small Animal Practice*. 2012;53(5):273-7.
- Gwinn DE, Wilckens JH, McDevitt ER, Ross G, Kao TC. The relative incidence of anterior cruciate ligament injury in men and women at the United States Naval Academy. *The American Journal of Sports Medicine*. 2000;28(1):98-102.
- Gyles C. One Medicine, One Health, One World. *The Canadian Veterinary Journal*. 2016;57(4):345-6.
- Habier D, Fernando RL, Kizilkaya K, Garrick DJ. Extension of the Bayesian alphabet for genomic selection. *BMC Bioinformatics*. 2011;12(1):186.
- Hajiloo M, Damavandi B, HooshSadat M, Sangi F, Mackey JR, Cass CE, Greiner R, Damaraju S. Breast cancer prediction using genome wide single nucleotide polymorphism data. *BMC bioinformatics*. 2013;14(13):S3.
- Haller K, Rambaldi I, Daniels E, Featherstone M. Subcellular localization of multiple PREP2 isoforms is regulated by actin, tubulin, and nuclear export. *Journal of Biological Chemistry*. 2004;279(47):49384-94.
- Halper J. Proteoglycans and diseases of soft tissues. In: Halper, J. *Progress in heritable soft connective tissue diseases*. Springer, Dordrecht. 2014: 49-58.
- Hammer Lab. PyEnsembl [Internet]. Github. 2017 [cited August 2019]. Available from: <https://github.com/openvax/pyensembl>.
- Harasen G. Diagnosing rupture of the cranial cruciate ligament. *The Canadian Veterinary Journal*. 2002;43(6):475-6.
- Harasen G. Canine cranial cruciate ligament rupture in profile: 2002–2007. *The Canadian Veterinary Journal*. 2008 Feb;49(2):193-4.
- Harner CD, Paulos LE, Greenwald AE, Rosenberg TD, Cooley VC. Detailed analysis of patients with bilateral anterior cruciate ligament injuries. *The American Journal of Sports Medicine*. 1994;22(1):37-43.
- Hart BL, Hart LA, Thigpen AP, Willits NH. Neutering of German Shepherd Dogs: associated joint disorders, cancers and urinary incontinence. *Veterinary Medicine and Science*. 2016;2(3):191-9.
- Hashemi J, Chandrashekar N, Gill B, Beynnon BD, Slauterbeck JR, Schutt Jr RC, Mansouri H, Dabezies E. The geometry of the tibial plateau and its influence on the biomechanics of the tibiofemoral joint. *The Journal of Bone and Joint Surgery*. 2008;90(12):2724.
- Havig ME, Dyce J, Kowaleski MP, Reynolds LR, Budsberg SC. Relationship of tibial plateau slope to limb function in dogs treated with a lateral suture technique for stabilization of cranial cruciate ligament deficient stifles. *Veterinary Surgery*. 2007;36(3):245-51.

- Hayashi K, Frank JD, Dubinsky C, Hao Z, Markel MD, Manley PA, Muir P. Histologic changes in ruptured canine cranial cruciate ligament. *Veterinary Surgery*. 2003;32(3):269-77.
- Hayes BJ, Pryce J, Chamberlain AJ, Bowman PJ, Goddard ME. Genetic architecture of complex traits and accuracy of genomic prediction: coat colour, milk-fat percentage, and type in Holstein cattle as contrasting model traits. *PLoS Genetics*. 2010;6(9):e1001139.
- Haynes MK, Hume EL, Smith JB. Phenotypic characterization of inflammatory cells from osteoarthritic synovium and synovial fluids. *Clinical Immunology*. 2002;105(3):315-25.
- Haynes KH, Biskup J, Freeman A, Conzemius MG. Effect of tibial plateau angle on cranial cruciate ligament strain: an ex vivo study in the dog. *Veterinary Surgery*. 2015;44(1):46-9.
- Hayward JJ, Castelhana MG, Oliveira KC, Corey E, Balkman C, Baxter TL, Casal ML, Center SA, Fang M, Garrison SJ, Kalla SE. Complex disease and phenotype mapping in the domestic dog. *Nature Communications*. 2016;7:10460.
- Hewett TE, Zazulak BT, Myer GD. Effects of the menstrual cycle on anterior cruciate ligament injury risk: a systematic review. *The American Journal of Sports Medicine*. 2007;35(4):659-68.
- Higgs HN. Formin proteins: a domain-based approach. *Trends in Biochemical Sciences*. 2005;30(6):342-53.
- Hoepfner MP, Lundquist A, Pirun M, Meadows JR, Zamani N, Johnson J, Sundström G, Cook A, FitzGerald MG, Swofford R, Mauceli E. An improved canine genome and a comprehensive catalogue of coding genes and non-coding transcripts. *PLoS one*. 2014;9(3):e91172.
- Hoffman GE, Logsdon BA, Mezey JG. PUMA: a unified framework for penalized multiple regression analysis of GWAS data. *PLoS Computational Biology*. 2013;9(6):e1003101.
- Hohmann E, Bryant A, Reaburn P, Tetsworth K. Is there a correlation between posterior tibial slope and non-contact anterior cruciate ligament injuries?. *Knee Surgery, Sports Traumatology, Arthroscopy*. 2011;19(1):109-14.
- Holley RJ, Tai G, Williamson AJ, Taylor S, Cain SA, Richardson SM, Merry CL, Whetton AD, Kiely CM, Canfield AE. Comparative quantification of the surfaceome of human multipotent mesenchymal progenitor cells. *Stem Cell Reports*. 2015;4(3):473-88.
- Hong H, Xu L, Su Z, Liu J, Ge W, Shen J, Fang H, Perkins R, Shi L, Tong W. Pitfall of genome-wide association studies: Sources of inconsistency in genotypes and their effects. *Journal of Biomedical Science and Engineering*. 2012;5(10):557-73.
- Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research*. 2009a Nov 25;37(1):1-3.
- Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*. 2009b;4(1):44.

- Huang Q. Genetic study of complex diseases in the post-GWAS era. *Journal of Genetics and Genomics*. 2015;42(3):87-98.
- Hyde CL, Nagle MW, Tian C, Chen X, Paciga SA, Wendland JR, Tung JY, Hinds DA, Perlis RH, Winslow AR. Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nature Genetics*. 2016;48(9):1031.
- Ilagan JO, Chalkley RJ, Burlingame AL, Jurica MS. Rearrangements within human spliceosomes captured after exon ligation. *RNA*. 2013;19(3):400-12.
- Inauen R, Koch D, Bass M, Haessig M. Tibial tuberosity conformation as a risk factor for cranial cruciate ligament rupture in the dog. *Veterinary and Comparative Orthopaedics and Traumatology*. 2009;22(01):16-20.
- Janovec J, Kyllar M, Midgley D, Owen M. Conformation of the proximal tibia and cranial cruciate ligament disease in small breed dogs. *Veterinary and Comparative Orthopaedics and Traumatology*. 2017;30(03):178-83.
- John R, Dhillon MS, Sharma S, Prabhakar S, Bhandari M. Is there a genetic predisposition to anterior cruciate ligament tear? A systematic review. *The American Journal of Sports Medicine*. 2016;44(12):3262-9.
- Johnson JS, Morscher MA, Jones KC, Moen SM, Klonk CJ, Jacquet R, Landis WJ. Gene expression differences between ruptured anterior cruciate ligaments in young male and female subjects. *Journal of Bone and Joints Surgery*. 2015;97(1):71-9.
- Karlsson EK, Lindblad-Toh K. Leader of the pack: gene mapping in dogs and other model organisms. *Nature Reviews Genetics*. 2008;9(9):713.
- Karlsson EK, Sigurdsson S, Ivansson E, Thomas R, Elvers I, Wright J, Howald C, Tonomura N, Perloski M, Swofford R, Biagi T. Genome-wide analyses implicate 33 loci in heritable dog osteosarcoma, including regulatory variants near CDKN2A/B. *Genome Biology*. 2013;14(12):R132.
- Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, Kent WJ. The UCSC Table Browser data retrieval tool. *Nucleic Acids Research*. 2004;32:D493-6.
- Kaynak M, Nijman F, van Meurs J, Reijman M, Meuffels DE. Genetic variants and anterior cruciate ligament rupture: a systematic review. *Sports Medicine*. 2017;47(8):1637-50.
- Kemper KE, Reich CM, Bowman PJ, Vander Jagt CJ, Chamberlain AJ, Mason BA, Hayes BJ, Goddard ME. Improved precision of QTL mapping using a nonlinear Bayesian method in a multi-breed population leads to greater accuracy of across-breed genomic predictions. *Genetics Selection Evolution*. 2015;47(1):29.
- Khoschnau S, Melhus H, Jacobson A, Rahme H, Bengtsson H, Ribom E, Grundberg E, Mallmin H, Michaëlsson K. Type I collagen $\alpha 1$ Sp1 polymorphism and the risk of cruciate ligament ruptures or shoulder dislocations. *The American Journal of Sports Medicine*. 2008;36(12):2432-6.

- El Khoury L, Posthumus M, Collins M, van der Merwe W, Handley C, Cook J, Raleigh SM. ELN and FBN2 gene variants as risk factors for two sports-related musculoskeletal injuries. *International Journal of Sports Medicine*. 2015;36(04):333-7.
- Kim SE, Pozzi A, Kowaleski MP, Lewis DD. Tibial osteotomies for cranial cruciate ligament insufficiency in dogs. *Veterinary Surgery*. 2008;37(2):111-25.
- Kim SK, Roos TR, Roos AK, Kleimeyer JP, Ahmed MA, Goodlin GT, Fredericson M, Ioannidis JP, Avins AL, Dragoo JL. Genome-wide association screens for Achilles tendon and ACL tears and tendinopathy. *PLoS One*. 2017;12(3):e0170422.
- Kipfer NM, Tepic S, Damur DM, Guerrero T, Hässig M, Montavon PM. Effect of tibial tuberosity advancement on femorotibial shear in cranial cruciate-deficient stifles. *Veterinary and Comparative Orthopaedics and Traumatology*. 2008;21(05):385-90.
- Kitayama K, Kamo M, Oma Y, Matsuda R, Uchida T, Ikura T, Tashiro S, Ohyama T, Winsor B, Harata M. The human actin-related protein hAtp5: nucleo-cytoplasmic shuttling and involvement in DNA repair. *Experimental Cell Research*. 2009;315(2):206-17.
- König IR, Loley C, Erdmann J, Ziegler A. How to include chromosome X in your genome-wide association study. *Genetic Epidemiology*. 2014;38(2):97-103.
- Koskinen A, Juslin S, Nieminen R, Moilanen T, Vuolteenaho K, Moilanen E. Adiponectin associates with markers of cartilage degradation in osteoarthritis and induces production of proinflammatory and catabolic factors through mitogen-activated protein kinase pathways. *Arthritis Research & Therapy*. 2011;13(6):R184.
- Kowaleski MP, Apelt D, Mattoon JS, Litsky AS. The effect of tibial plateau leveling osteotomy position on cranial tibial subluxation: an in vitro study. *Veterinary Surgery*. 2005;34(4):332-6.
- Kuhn M. Building predictive models in R using the caret package. *Journal of Statistical Software*. 2008;28(5):1-26.
- Kung JT, Colognori D, Lee JT. Long noncoding RNAs: past, present, and future. *Genetics*. 2013;193(3):651-69.
- Lampman TJ, Lund EM, Lipowitz AJ. Cranial cruciate disease: current status of diagnosis, surgery, and risk for disease. *Veterinary and Comparative Orthopaedics and Traumatology*. 2003;16(03):122-6.
- Lee CH, Shin HJ, Cho IH, Kang YM, Kim IA, Park KD, Shin JW. Nanofiber alignment and direction of mechanical strain affect the ECM production of human ACL fibroblast. *Biomaterials*. 2005;26(11):1261-70.
- Lee PH, O'Dushlaine C, Thomas B, Purcell SM. INRICH: interval-based enrichment analysis for genome-wide association studies. *Bioinformatics*. 2012;28(13):1797-9.
- Lequarré AS, Andersson L, André C, Fredholm M, Hitte C, Leeb T, Lohi H, Lindblad-Toh K, Georges M. LUPA: a European initiative taking advantage of the canine genome

- architecture for unravelling complex disorders in both human and dogs. *The Veterinary Journal*. 2011;189(2):155-9.
- Lewis BA, Allen DA, Henrikson TD, Lehenbauer TW. Computed tomographic evaluation of the canine intercondylar notch in normal and cruciate deficient stifles. *Veterinary and Comparative Orthopaedics and Traumatology*. 2008;21(02):119-24.
- Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 2011b;12(1):323.
- Li L, Li Y, Browning SR, Browning BL, Slater AJ, Kong X, Aponte JL, Mooser VE, Chisoe SL, Whittaker JC, Nelson MR. Performance of genotype imputation for rare variants identified in exons and flanking regions of genes. *PLoS One*. 2011a;6(9):e24945.
- Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. *Annual Review of Genomics and Human Genetics*. 2009;10:387-406.
- Liao W, Liu W, Liu X, Yuan Q, Ou Y, Qi Y, Huang W, Wang Y, Huang J. Upregulation of FAM83D affects the proliferation and invasion of hepatocellular carcinoma. *Oncotarget*. 2015;6(27):24132.
- Light VA, Montgomery RD, Akingbemi BT. Sex hormone regulation of collagen concentrations in cranial cruciate ligaments of sexually immature male rabbits. *American Journal of Veterinary Research*. 2012;73(8):1186-93.
- Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, Clamp M, Chang JL, Kulbokas III EJ, Zody MC, Mauceli E. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature*. 2005;438(7069):803.
- Little JP, Bleedorn JA, Sutherland BJ, Sullivan R, Kalscheur VL, Ramaker MA, Schaefer SL, Hao Z, Muir P. Arthroscopic assessment of stifle synovitis in dogs with cranial cruciate ligament rupture. *PLoS One*. 2014;9(6):e97329.
- Lloyd KK, Khanna C, Hendricks W, Trent J, Kotlikoff M. Precision medicine: an opportunity for a paradigm shift in veterinary medicine. *Journal of the American Veterinary Medical Association*. 2016;248(1):45-8.
- Lohmander LS, Östenberg A, Englund M, Roos H. High prevalence of knee osteoarthritis, pain, and functional limitations in female soccer players twelve years after anterior cruciate ligament injury. *Arthritis & Rheumatism*. 2004;50(10):3145-52.
- Lohmander LS, Englund PM, Dahl LL, Roos EM. The long-term consequence of anterior cruciate ligament and meniscus injuries: osteoarthritis. *The American Journal of Sports Medicine*. 2007;35:1756-69.
- Lulińska-Kuklik E, Maculewicz E, Moska W, Ficek K, Kaczmarczyk M, Michałowska-Sawczyn M, Humińska-Lisowska K, Buryta M, Chycki J, Cięszczyk P, Żmijewski P. Are IL1B, IL6 and IL6R Gene Variants Associated with Anterior Cruciate Ligament Rupture Susceptibility?. *Journal of Sports Science & Medicine*. 2019;18(1):137.

- Luykenaar KD, El-Rahman RA, Walsh MP, Welsh DG. Rho-kinase-mediated suppression of KDR current in cerebral arteries requires an intact actin cytoskeleton. *American Journal of Physiology-Heart and Circulatory Physiology*. 2009 Apr;296(4):H917-26.
- Macias C, McKee WM, May C. Caudal proximal tibial deformity and cranial cruciate ligament rupture in small-breed dogs. *Journal of Small Animal Practice*. 2002;43(10):433-8.
- MacLeod IM, Bowman PJ, Vander Jagt CJ, Haile-Mariam M, Kemper KE, Chamberlain AJ, Schrooten C, Hayes BJ, Goddard ME. Exploiting biological priors and sequence variants enhances QTL discovery and genomic prediction of complex traits. *BMC Genomics*. 2016;17(1):144.
- Macrossan PE, Kinghorn BP, Wilke VL, Rothschild MF. Selective genotyping for determination of a major gene associated with cranial cruciate ligament disease in the Newfoundland dog. In: *Proceedings of the Association for the Advancements of Animal Breeding and Genetics*. 2005;16:346-49.
- Maeda K, Kobayashi Y, Udagawa N, Uehara S, Ishihara A, Mizoguchi T, Kikuchi Y, Takada I, Kato S, Kani S, Nishita M. Wnt5a-Ror2 signaling between osteoblast-lineage cells and osteoclast precursors enhances osteoclastogenesis. *Nature Medicine*. 2012;18(3):405.
- Maerz T, Sherman E, Newton M, Yilmaz A, Kumar P, Graham SF, Baker KC. Metabolomic serum profiling after ACL injury in rats: A pilot study implicating inflammation and immune dysregulation in post-traumatic osteoarthritis. *Journal of Orthopaedic Research*. 2018;36(7):1969-79.
- Marks PH, Donaldson ML. Inflammatory cytokine profiles associated with chondral damage in the anterior cruciate ligament-deficient knee. *Arthroscopy: The Journal of Arthroscopic & Related Surgery*. 2005;21(11):1342-7.
- McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research*. 2012;40(10):4288-97.
- Malila S, Yuktanandana P, Saowaprut S, Jiamjarasrangsi W, Honsawek S. Association between matrix metalloproteinase-3 polymorphism and anterior cruciate ligament ruptures. *Genetics and Molecular Research*. 2011;10(4):4158-65.
- Mannion S, Mtintsilana A, Posthumus M, van der Merwe W, Hobbs H, Collins M, September AV. Genes encoding proteoglycans are associated with the risk of anterior cruciate ligament ruptures. *British Journal of Sports Medicine*. 2014;48(22):1640-6.
- Mather III RC, Koenig L, Kocher MS, Dall TM, Gallo P, Scott DJ, Bach Jr BR, Spindler KP, MOON Knee Group. Societal and economic impact of anterior cruciate ligament tears. *Journal of Bone and Joint Surgery*. 2013;95(19):1751.
- Mealey KL, Martinez SE, Villarino NF. Personalized medicine: going to the dogs?. *Human Genetics*. 2019;138(5):467-81.
- Meuwissen T, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*. 2001;157(4):1819-29.

- Meuwissen T, Hayes B, Goddard M. Genomic selection: a paradigm shift in animal breeding. *Animal Frontiers*. 2016;6(1):6-14.
- Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F. e1071: misc functions of the Department of Statistics, Probability Theory Group (Formerly: e 1071), TU Wien. R package version 1.6–8. 2017.
- Miyasaka KC. The incidence of knee ligament injuries in the general population. *American Journal of Knee Surgery*. 1991;1:43-8.
- Morimoto-Tomita M, Uchimura K, Werb Z, Hemmerich S, Rosen SD. Cloning and characterization of two extracellular heparin-degrading endosulfatases in mice and humans. *Journal of Biological Chemistry*. 2002;277(51):49175-85.
- Morris E, Lipowitz AJ. Comparison of tibial plateau angles in dogs with and without cranial cruciate ligament injuries. *Journal of the American Veterinary Medical Association*. 2001;218(3):363-6.
- Moser G, Lee SH, Hayes BJ, Goddard ME, Wray NR, Visscher PM. Simultaneous discovery, estimation and prediction analysis of complex traits using a Bayesian mixture model. *PLoS Genetics*. 2015;11(4):e1004969.
- Mostafa AA, Griffon DJ, Thomas MW, Constable PD. Morphometric characteristics of the pelvic limbs of Labrador Retrievers with and without cranial cruciate ligament deficiency. *American journal of Veterinary Research*. 2009;70(4):498-507.
- Muir, P. Physical examination of lame dogs. *Compendium on Continuing Education for the Practising Veterinarian*. 1997;19:1149-61.
- Muir P, Oldenhoff WE, Hudson AP, Manley PA, Schaefer SL, Markel MD, Hao Z. Detection of DNA from a range of bacterial species in the knee joints of dogs with inflammatory knee arthritis and associated degenerative anterior cruciate ligament rupture. *Microbial Pathogenesis*. 2007;42:47-55.
- Muir P, Schwartz Z, Malek S, Kreines A, Cabrera SY, Buote NJ, Bleedorn JA, Schaefer SL, Holzman G, Hao Z. Contralateral cruciate survival in dogs with unilateral non-contact cranial cruciate ligament rupture. *PLoS One*. 2011a;6(10):e25331.
- Muir P, Kelly JL, Marvel SJ, Heinrich DA, Schaefer SL, Manley PA, Tewari K, Singh A, Suresh M, Hao Z, Plisch E. Lymphocyte populations in joint tissues from dogs with inflammatory stifle arthritis and associated degenerative cranial cruciate ligament rupture. *Veterinary Surgery*. 2011b;40(6):753-61.
- Murray MM, Martin SD, Martin TL, Spector M. Histological changes in the human anterior cruciate ligament after rupture. *Journal of Bone and Joint Surgery*. 2000;82(10):1387.
- Myer GD, Heidt RS, Waits C, Finck S, Stanfield D, Posthumus M, Hewett TE. Sex comparison of familial predisposition to anterior cruciate ligament injury. *Knee Surgery, Sports Traumatology, Arthroscopy*. 2014;22(2):387-91.

- Natekin A, Knoll A. Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics*. 2013;7:21.
- Nielen AL, Knol BW. Genetic and epidemiological investigation of a birth cohort of boxers. *Tijdschrift Voor Diergeneeskunde*. 2003;128(19):586-90.
- Nishikimi A, Uruno T, Duan X, Cao Q, Okamura Y, Saitoh T, Saito N, Sakaoka S, Du Y, Suenaga A, Kukimoto-Niino M. Blockade of inflammatory responses by a small-molecule inhibitor of the Rac activator DOCK2. *Chemistry & Biology*. 2012;19(4):488-97.
- Obeidat M, Li L, Ballermann BJ. TIMAP promotes angiogenesis by suppressing PTEN-mediated Akt inhibition in human glomerular endothelial cells. *American Journal of Physiology-Renal Physiology*. 2014;307(5):F623-33.
- O'Connell K, Knight H, Ficek K, Leonska-Duniec A, Maciejewska-Karlowska A, Sawczuk M, Stepień-Słodkowska M, O'Cuinneagain D, van der Merwe W, Posthumus M, Cieszczyk P. Interactions between collagen gene variants and risk of anterior cruciate ligament rupture. *European Journal of Sport Science*. 2015;15(4):341-50.
- Okbay A, Baselmans BM, De Neve JE, Turley P, Nivard MG, Fontana MA, Meddens SF, Linnér RK, Rietveld CA, Derringer J, Gratten J. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nature Genetics*. 2016;48(6):624.
- Osmond CS, Marcellin-Little DJ, Harrysson OL, Kidd LB. Morphometric assessment of the proximal portion of the tibia in dogs with and without cranial cruciate ligament rupture. *Veterinary Radiology & Ultrasound*. 2006;47(2):136-41.
- Ostrander EA, Beale HC. Leading the way: finding genes for neurologic disease in dogs using genome-wide mRNA sequencing. *BMC Genetics*. 2012;13(1):56.
- Otsuki S, Taniguchi N, Grogan SP, D'Lima D, Kinoshita M, Lotz M. Expression of novel extracellular sulfatases Sulf-1 and Sulf-2 in normal and osteoarthritic articular cartilage. *Arthritis Research & Therapy*. 2008;10(3):R61.
- Overall KL. Natural animal models of human psychiatric conditions: assessment of mechanism and validity. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*. 2000;24(5):727-76.
- Pallu S, Francin PJ, Guillaume C, Gegout-Pottie P, Netter P, Mainard D, Terlain B, Presle N. Obesity affects the chondrocyte responsiveness to leptin in patients with osteoarthritis. *Arthritis Research & Therapy*. 2010;12(3):R112.
- Park JH, Wacholder S, Gail MH, Peters U, Jacobs KB, Chanock SJ, Chatterjee N. Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nature Genetics*. 2010;42(7):570.
- Park T, Casella G. The Bayesian LASSO. *Journal of the American Statistical Association*. 2008;103(482):681-6.

- Pascual-Garrido C, Guilak F, Rai MF, Harris MD, Lopez MJ, Todhunter RJ, Clohisy JC. Canine hip dysplasia: a natural animal model for human developmental dysplasia of the hip. *Journal of Orthopaedic Research*. 2018;36(7):1807-17.
- Paschos NK. Anterior cruciate ligament reconstruction and knee osteoarthritis. *World Journal of Orthopedics*. 2017;8(3):212-17.
- Paterno MV, Rauh MJ, Schmitt LC, Ford KR, Hewett TE. Incidence of contralateral and ipsilateral anterior cruciate ligament (ACL) injury after primary ACL reconstruction and return to sport. *Clinical Journal of Sports Medicine*. 2012;22(2):116.
- Payne SL, Hendrix MJ, Kirschmann DA. Lysyl oxidase regulates actin filament formation through the p130Cas/Crk/DOCK180 signaling complex. *Journal of Cellular Biochemistry*. 2006;98(4):827-37.
- Peffers MJ, Balaskas P, Smagul A. Osteoarthritis year in review 2017: genetics and epigenetics. *Osteoarthritis and Cartilage*. 2018;26(3):304-11.
- Pérez P, de Los Campos G. Genome-wide regression and prediction with the BGLR statistical package. *Genetics*. 2014;198(2):483-95.
- Personnic N, Lakisic G, Gouin E, Rousseau A, Gautreau A, Cossart P, Bierne H. A role for Ral GTP ase-activating protein subunit β in mitotic regulation. *The FEBS Journal*. 2014;281(13):2977-89.
- Plaas A, Sandy JD, Liu H, Diaz MA, Schenkman D, Magnus RP, Bolam-Bretl C, Kopesky PW, Wang VM, Galante JO. Biochemical identification and immunolocalization of aggrecan, ADAMTS5 and inter-alpha-trypsin-inhibitor in equine degenerative suspensory ligament desmitis. *Journal of Orthopaedic Research*. 2011;29(6):900-6.
- Posthumus M, September AV, Keegan M, O'Cuinneagain D, Van der Merwe W, Schwellnus MP, Collins M. Genetic risk factors for anterior cruciate ligament ruptures: COL1A1 gene variant. *British Journal of Sports Medicine*. 2009a;43(5):352-6.
- Posthumus M, September AV, O'Cuinneagain D, van der Merwe W, Schwellnus MP, Collins M. The COL5A1 gene is associated with increased risk of anterior cruciate ligament ruptures in female participants. *The American Journal of Sports Medicine*. 2009b;37(11):2234-40.
- Posthumus M, September AV, O'Cuinneagain D, van der Merwe W, Schwellnus MP, Collins M. The association between the COL12A1 gene and anterior cruciate ligament ruptures. *British Journal of Sports Medicine*. 2010;44(16):1160-5.
- Posthumus M, Collins M, Van Der Merwe L, O'cuinneagain D, Van Der Merwe W, Ribbans WJ, Schwellnus MP, Raleigh SM. Matrix metalloproteinase genes on chromosome 11q22 and the risk of anterior cruciate ligament (ACL) rupture. *Scandinavian Journal of Medicine & Science in Sports*. 2012;22(4):523-33.
- Powers MY, Martinez SA, Lincoln JD, Temple CJ, Arnaiz A. Prevalence of cranial cruciate ligament rupture in a population of dogs with lameness previously attributed to hip

- dysplasia: 369 cases (1994–2003). *Journal of the American Veterinary Medical Association*. 2005;227(7):1109-11.
- Pritchard JK, Donnelly P. Case–control studies of association in structured or admixed populations. *Theoretical Population Biology*. 2001;60(3):227-37.
- Proffen BL, McElfresh M, Fleming BC, Murray MM. A comparative anatomical study of the human knee and six animal species. *The Knee*. 2012;19(4):493-9.
- Proffen BL, Murray MM. In vivo models of ACL injury (central defect, porcine, ovine, canine). In: *The ACL handbook*. New York, NY. Springer. 2013: 139-66.
- Quignon P, Herbin L, Cadieu E, Kirkness EF, Hédan B, Mosher DS, Galibert F, André C, Ostrander EA, Hitte C. Canine population structure: assessment and impact of intra-breed stratification on SNP-based association studies. *PLoS One*. 2007;2(12):e1324.
- Quinn JJ, Chang HY. Unique features of long non-coding RNA biogenesis and function. *Nature Reviews Genetics*. 2016;17(1):47.
- R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2016. URL <http://www.R-project.org>
- Ragetly CA, Evans R, Mostafa AA, Griffon DJ. Multivariate analysis of morphometric characteristics to evaluate risk factors for cranial cruciate ligament deficiency in Labrador retrievers. *Veterinary Surgery*. 2011;40(3):327-33.
- Rahim M, Gibbon A, Hobbs H, van der Merwe W, Posthumus M, Collins M, September AV. The association of genes involved in the angiogenesis-associated signaling pathway with risk of anterior cruciate ligament rupture. *Journal of Orthopaedic Research*. 2014;32(12):1612-8.
- Rahim M, Mannion S, Klug B, Hobbs H, van der Merwe W, Posthumus M, Collins M, September AV. Modulators of the extracellular matrix and risk of anterior cruciate ligament ruptures. *Journal of science and medicine in sport*. 2017 Feb 1;20(2):152-8.
- Raleigh SM, Posthumus M, O’Cuinneagain D, van der Merwe W, Collins M. The GDF5 gene and anterior cruciate ligament rupture. *International Journal of Sports Medicine*. 2013;34(04):364-7.
- Rayward RM, Thomson DG, Davies JV, Innes JF, Whitelock RG. Progression of osteoarthritis following TPLO surgery: a prospective radiographic study of 40 dogs. *Journal of Small Animal Practice*. 2004;45(2):92-7.
- Rebagay G, Yan S, Liu C, Cheung NK. ROR1 and ROR2 in human malignancies: potentials for targeted therapy. *Frontiers in Oncology*. 2012;2:34.
- Reif U, Probst CW. Comparison of tibial plateau angles in normal and cranial cruciate deficient stifles of Labrador retrievers. *Veterinary Surgery*. 2003;32(4):385-9.
- Reiner A, Yekutieli D, Benjamini Y. Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics*. 2003;19(3):368-75.

- Roarty K, Shore AN, Creighton CJ, Rosen JM. Ror2 regulates branching, differentiation, and actin-cytoskeletal dynamics within the mammary epithelium. *Journal of Cell Biology*. 2015;208(3):351-66.
- Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Müller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12(1):77.
- Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139-40.
- Robinson MR, Wray NR, Visscher PM. Explaining additional genetic variation in complex traits. *Trends in Genetics*. 2014;30(4):124-32.
- Rowan TN, Hoff JL, Crum TE, Taylor JF, Schnabel RD, Decker JE. A Multi-Breed Reference Panel and Additional Rare Variation Maximizes Imputation Accuracy in Cattle. *bioRxiv*. 2019:517144.
- Rowell JL, McCarthy DO, Alvarez CE. Dog models of naturally occurring cancer. *Trends in Molecular Medicine*. 2011;17(7):380-8.
- Sample SJ, Racette MA, Hans EC, Volstad NJ, Holzman G, Bleedorn JA, Schaefer SL, Waller III KR, Hao Z, Block WF, Muir P. Radiographic and magnetic resonance imaging predicts severity of cruciate ligament fiber damage and synovitis in dogs with cranial cruciate ligament rupture. *PLoS one*. 2017;12(6):e0178086.
- Sanders TL, Maradit Kremers H, Bryan AJ, Larson DR, Dahm DL, Levy BA, Stuart MJ, Krych AJ. Incidence of anterior cruciate ligament tears and reconstruction: a 21-year population-based study. *The American Journal of Sports Medicine*. 2016;44(6):1502-7.
- Saunders CJ, Dashti MJ, Gamielidien J. Semantic interrogation of a multi knowledge domain ontological model of tendinopathy identifies four strong candidate risk genes. *Scientific Reports*. 2016;6:19820.
- Schiele NR, Von Flotow F, Tochka ZL, Hockaday LA, Marturano JE, Thibodeau JJ, Kuo CK. Actin cytoskeleton contributes to the elastic modulus of embryonic tendon during early development. *Journal of Orthopaedic Research*. 2015;33(6):874-81.
- Schierding WS, Cutfield WS, O'Sullivan JM. The missing story behind Genome Wide Association Studies: single nucleotide polymorphisms in gene deserts have a story to tell. *Frontiers in Genetics*. 2014;5:39.
- Schwandt CS, Bohorquez-Vanelli A, Tepic S, Hassig M, Dennler R, Vezzoni A, Montavon PM. Angle between the patellar ligament and tibial plateau in dogs with partial rupture of the cranial cruciate ligament. *American Journal of Veterinary Research*. 2006;67(11):1855-60.
- Schwartz NB, Domowicz M. Chondrodysplasias due to proteoglycan defects. *Glycobiology*. 2002;12(4):57R-68R.

- Selmi AL, Filho JP. Rupture of the cranial cruciate ligament associated with deformity of the proximal tibia in five dogs. *Journal of Small Animal Practice*. 2001;42(8):390-3.
- September AV, Schwellnus MP, Collins M. Tendon and ligament injuries: the genetic component. *British Journal of Sports Medicine*. 2007;41(4):241-6.
- Shahar R, Milgram J. Biomechanics of tibial plateau leveling of the canine cruciate-deficient stifle joint: a theoretical model. *Veterinary Surgery*. 2006;35(2):144-9.
- Shearin AL, Ostrander EA. Leading the way: canine models of genomics and disease. *Disease Models & Mechanisms*. 2010;3:27-34.
- Shim H, Chasman DI, Smith JD, Mora S, Ridker PM, Nickerson DA, Krauss RM, Stephens M. A multivariate genome-wide association analysis of 10 LDL subfractions, and their response to statin treatment, in 1868 Caucasians. *PLoS One*. 2015;10(4):e0120758.
- Shultz SJ, Schmitz RJ, Benjaminse A, Collins M, Ford K, Kulas AS. ACL research retreat VII: An update on anterior cruciate ligament injury risk factor identification, screening, and prevention: March 19–21, 2015; Greensboro, NC. *Journal of Athletic Training*. 2015;50(10):1076-93.
- Simon D, Mascarenhas R, Saltzman BM, Rollins M, Bach BR, MacDonald P. The relationship between anterior cruciate ligament injury and osteoarthritis of the knee. *Advances in Orthopedics*. 2015;2015:928301.
- Simons YB, Bullaughey K, Hudson RR, Sella G. A population genetic interpretation of GWAS findings for human quantitative traits. *PLoS Biology*. 2018;16(3):e2002985.
- Slauterbeck JR, Pankratz K, Xu KT, Bozeman SC, Hardy DM. Canine ovariohysterectomy and orchiectomy increases the prevalence of ACL injury. *Clinical Orthopaedics and Related Research*. 2004;429:301-5.
- Smith HC, Vacek P, Johnson RJ, Slauterbeck JR, Hashemi J, Shultz S, et al. Risk factors for anterior cruciate ligament injury: a review of the literature – part 1: neuromuscular and anatomic risk. *Sports Health*. 2012a;4:69–78.
- Smith HC, Vacek P, Johnson RJ, Slauterbeck JR, Hashemi J, Shultz S, et al. Risk factors for anterior cruciate ligament injury: a review of the literature – part 2: hormonal, genetic, cognitive function, previous injury, and extrinsic risk factors. *Sports Health*. 2012b;4:155–61.
- Song YJ, Li G, He JH, Guo Y, Yang L. Bioinformatics-based identification of microRNA-regulated and rheumatoid arthritis-associated genes. *PLoS One*. 2015;10(9):e0137551.
- Spicer AP, Joo A, Bowling RA. A hyaluronan binding link protein gene family whose members are physically linked adjacent to chondroitin sulfate proteoglycan core protein genes: the missing links. *Journal of Biological Chemistry*. 2003;278(23):21083-91.
- Stephens M. A unified framework for association analysis with multiple related phenotypes. *PLoS One*. 2013;8(7):e65245.

- Stephens M, Balding DJ. Bayesian statistical methods for genetic association studies. *Nature Reviews Genetics*. 2009;10(10):681.
- Stępień-Słodkowska M, Ficek K, Eider J, Leońska-Duniec A, Maciejewska-Karłowska A, Sawczuk M, Zarębska A, Jastrzębski Z, Grenda A, Kotarska K, Cięższyk P. The +1245g/t polymorphisms in the collagen type I alpha 1 (COL1A1) gene in polish skiers with anterior cruciate ligament injury. *Biology of Sport*. 2013;30(1):57-60.
- Stępień-Słodkowska M, Ficek K, Maciejewska-Karłowska A, Sawczuk M, Ziętek P, Król P, Zmijewski P, Pokrywka A, Cięższyk P. Overrepresentation of the COL3A1 AA genotype in Polish skiers with anterior cruciate ligament injury. *Biology of sport*. 2015a;32(2):143.
- Stępień-Słodkowska M, Ficek K, Kaczmarczyk M, Maciejewska-Karłowska A, Sawczuk M, Leońska-Duniec A, Stępiński M, Ziętek P, Król P, Chudecka M, Cięższyk P. The variants within the COL5A1 gene are associated with reduced risk of anterior cruciate ligament injury in skiers. *Journal of Human Kinetics*. 2015b;45(1):103-11.
- Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences*. 2003;100(16):9440-5.
- Su L, Townsend KL, Au J, Wittum TE. Comparison of tibial plateau angles in small and large breed dogs. *The Canadian Veterinary Journal*. 2015;56(6):610.
- Suman P, Mishra S, Chander H. High expression of FBP17 in invasive breast cancer cells promotes invadopodia formation. *Medical Oncology*. 2018;35(5):71.
- Sutter NB, Eberle MA, Parker HG, Pullar BJ, Kirkness EF, Kruglyak L, Ostrander EA. Extensive and breed-specific linkage disequilibrium in *Canis familiaris*. *Genome Research*. 2004;14(12):2388-96.
- Sutton KM, Bullock JM. Anterior cruciate ligament rupture: differences between males and females. *Journal of the American Academy of Orthopaedic Surgeons*. 2013;21(1):41-50.
- Sonnhammer EL, Von Heijne G, Krogh A. A hidden Markov model for predicting transmembrane helices in protein sequences. *Proceedings from the International Conference of Intelligent Systems Molecular Biology*. 1998;6:175-82.
- Tabor HK, Risch NJ, Myers RM. Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nature Reviews Genetics*. 2002;3(5):391-7.
- Tang R, Noh HJ, Wang D, Sigurdsson S, Swofford R, Perloski M, Duxbury M, Patterson EE, Albright J, Castelhana M, Auton A. Candidate genes and functional noncoding variants identified in a canine model of obsessive-compulsive disorder. *Genome Biology*. 2014;15(3):R25.
- Temwichitr J, Hazewinkel HA, Van Hagen MA, Leegwater PA. Polymorphic microsatellite markers for genetic analysis of collagen genes in suspected collagenopathies in dogs. *Journal of Veterinary Medicine Series A*. 2007;54(9):522-6.

- Thompson JR, Gögele M, Weichenberger CX, Modenese M, Attia J, Barrett JH, Boehnke M, De Grandi A, Domingues FS, Hicks AA, Marroni F. SNP prioritization using a Bayesian probability of association. *Genetic Epidemiology*. 2013;37(2):214-21.
- Tjoumakaris FP, Donegan DJ, Sekiya JK. Partial tears of the anterior cruciate ligament: diagnosis and treatment. *The American Journal of Orthopedics*. 2011;40(2):92
- Todd MS, Lalliss S, Garcia ES, DeBerardino TM, Cameron KL. The relationship between posterior tibial slope and anterior cruciate ligament injuries. *The American Journal of Sports Medicine*. 2010;38(1):63-7.
- Torres de la Riva G, Hart BL, Farver TB, Oberbauer AM, Messam LL, Willits N, Hart LA. Neutering dogs: effects on joint disorders and cancers in golden retrievers. *PLoS one*. 2013;8(2):e55937.
- Uhorchak JM, Scoville CR, Williams GN, Arciero RA, Pierre PS, Taylor DC. Risk factors associated with noncontact injury of the anterior cruciate ligament. *The American Journal of Sports Medicine*. 2003;31(6):831-42.
- Urbinati C, Ravelli C, Tanghetti E, Belleri M, Giacomuzzi E, Monti E, Presta M, Rusnati M. Substrate-Immobilized HIV-1 Tat Drives VEGFR2/ α v β 3–Integrin Complex Formation and Polarization in Endothelial Cells. *Arteriosclerosis, Thrombosis, and Vascular Biology*. 2012;32(5):e25-34.
- Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Banks E. From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics*. 2013;43:11.10.1-33.
- Van Steenbeek FG, Hytönen MK, Leegwater PA, Lohi H. The canine era: the rise of a biomedical model. *Animal genetics*. 2016 Oct;47(5):519-27.
- Varoga D, Paulsen FP, Kohrs S, Grohmann S, Lippross S, Mentlein R, Tillmann BN, Goldring MB, Besch L, Pufe T. Expression and regulation of human β -defensin-2 in osteoarthritic cartilage. *The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland*. 2006;209(2):166-73.
- Vazquez AI, de los Campos G, Klimentidis YC, Rosa GJ, Gianola D, Yi N, Allison DB. A comprehensive genetic approach for improving prediction of skin cancer risk in humans. *Genetics*. 2012;192(4):1493-502.
- Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, Yang J. 10 years of GWAS discovery: biology, function, and translation. *The American Journal of Human Genetics*. 2017;101(1):5-22.
- Von Porat AR, Roos EM, Roos H. High prevalence of osteoarthritis 14 years after an anterior cruciate ligament tear in male soccer players: a study of radiographic and patient relevant outcomes. *Annals of the Rheumatic Diseases*. 2004;63(3):269-73.

- Waggett AD, Benjamin M, Ralphs JR. Connexin 32 and 43 gap junctions differentially modulate tenocyte response to cyclic mechanical load. *European journal of cell biology*. 2006;85(11):1145-54.
- Wang D, Liu H, Ren C, Wang L. High expression of ABRACL is associated with tumorigenesis and affects clinical outcome in gastric cancer. *Genetic Testing and Molecular Biomarkers*. 2019;23(2):91-7.
- Warzee CC, Dejardin LM, Arnoczky SP, Perry RL. Effect of tibial plateau leveling on cranial and caudal tibial thrusts in canine cranial cruciate–deficient stifles: An in vitro experimental study. *Veterinary Surgery*. 2001;30(3):278-86.
- Webster KE, Feller JA, Leigh WB, Richmond AK. Younger patients are at increased risk for graft rupture and contralateral injury after anterior cruciate ligament reconstruction. *The American Journal of Sports Medicine*. 2014;42(3):641-7.
- Whitaker JW, Boyle DL, Bartok B, Ball ST, Gay S, Wang W, Firestein GS. Integrative omics analysis of rheumatoid arthritis identifies non-obvious therapeutic targets. *PLoS One*. 2015;10(4):e0124254.
- Whitehair JG, Vasseur PB, Willits NH. Epidemiology of cranial cruciate ligament rupture in dogs. *Journal of the American Veterinary Medical Association*. 1993;203(7):1016-9.
- Wiener P, Sánchez-Molano E, Clements DN, Woolliams JA, Haskell MJ, Blott SC. Genomic data illuminates demography, genetic structure and selection of a popular dog breed. *BMC Genomics*. 2017;18(1):609.
- Wilke VL, Conzemius MG, Besancon MF, Evans RB, Ritter M. Comparison of tibial plateau angle between clinically normal Greyhounds and Labrador Retrievers with and without rupture of the cranial cruciate ligament. *Journal of the American Veterinary Medical Association*. 2002;221(10):1426-9.
- Wilke VL, Robinson DA, Evans RB, Rothschild MF, Conzemius MG. Estimate of the annual economic impact of treatment of cranial cruciate ligament injury in dogs in the United States. *Journal of the American Veterinary Medical Association*. 2005;227(10):1604-7.
- Wilke VL, Conzemius MG, Kinghorn BP, Macrossan PE, Cai W, Rothschild MF. Inheritance of rupture of the cranial cruciate ligament in Newfoundlands. *Journal of the American Veterinary Medical Association*. 2006;228(1):61-4.
- Wilke VL, Zhang S, Evans RB, Conzemius MG, Rothschild MF. Identification of chromosomal regions associated with cranial cruciate ligament rupture in a population of Newfoundlands. *American Journal of Veterinary Research*. 2009;70(8):1013-7.
- Wilke VL, Zaldivar-Lopez S, Ekenstedt K, Evans R, Conzemius M. Genotype influences risk of cranial cruciate ligament disease in the Newfoundland and Labrador retriever breeds. *Journal of Veterinary Medicine and Research*. 2015; 2(3): 1028.
- Witsberger TH, Villamil JA, Schultz LG, Hahn AW, Cook JL. Prevalence of and risk factors for hip dysplasia and cranial cruciate ligament deficiency in dogs. *Journal of the American Veterinary Medical Association*. 2008;232(12):1818-24.

- Wray NR, Yang J, Goddard ME, Visscher PM. The genetic interpretation of area under the ROC curve in genomic profiling. *PLoS Genetics*. 2010;6(2):e1000864.
- Wright RW, Magnussen RA, Dunn WR, Spindler KP. Ipsilateral graft and contralateral ACL rupture at five years or more following ACL reconstruction: a systematic review. *The Journal of Bone and Joint Surgery*. 2011;93(12):1159-65.
- Wu C, DeWan A, Hoh J, Wang Z. A comparison of association methods correcting for population stratification in case-control studies. *Annals of Human Genetics*. 2011;75(3):418-27.
- Wu Y, Zheng Z, Visscher PM, Yang J. Quantifying the mapping precision of genome-wide association studies using whole-genome sequencing data. *Genome Biology*. 2017;18(1):86.
- Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics*. 2011a;88(1):76-82.
- Yang J, Weedon MN, Purcell S, Lettre G, Estrada K, Willer CJ, Smith AV, Ingelsson E, O'connell JR, Mangino M, Mägi R. Genomic inflation factors under polygenic inheritance. *European Journal of Human Genetics*. 2011b;19(7):807.
- Yang J, Zeng J, Goddard ME, Wray NR, Visscher PM. Concepts, estimation and interpretation of SNP-based heritability. *Nature Genetics*. 2017;49(9):1304.
- Yau MS, Yerges-Armstrong LM, Liu Y, Lewis CE, Duggan DJ, Renner JB, Torner J, Felson DT, McCulloch CE, Kwok CK, Nevitt MC. Genome-Wide Association Study of Radiographic Knee Osteoarthritis in North American Caucasians. *Arthritis & Rheumatology*. 2017;69(2):343-51.
- Young NJ, Becker DL, Fleck RA, Goodship AE, Patterson-Kane JC. Maturational alterations in gap junction expression and associated collagen synthesis in response to tendon function. *Matrix Biology*. 2009;28(6):311-23.
- Zhang CH. Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*. 2010;38(2):894-942.
- Zhao H, Williams GJ, Huang JZ. Wsrfr: an R package for classification with scalable weighted subspace random forests. *Journal of Statistical Software*. 2017;77(i03):1.
- Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics*. 2012;44(7):821-4.
- Zhou X, Carbonetto P, Stephens M. Polygenic modeling with Bayesian sparse linear mixed models. *PLoS Genetics*. 2013;9(2):e1003264.
- Zinsstag J, Schelling E, Waltner-Toews D, Tanner M. From one medicine to one health and systemic approaches to health and well-being. *Preventive Veterinary Medicine*. 2011;101(3-4):148-56.

Zwida K, Kutzler MA. Non-reproductive long-term health complications of gonad removal in dogs as well as possible causal relationships with post-gonadectomy elevated luteinizing hormone (LH) concentrations. *Journal of Etiology and Animal Health*. 2016;1(002).