

Cross-Species Analysis of Genetic Architecture and Polygenic Risk Scores for Non-Contact ACL Rupture in Dogs and Humans

Corresponding Author: Professor Peter Muir

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

Dear Authors

This is an extremely well written and novel manuscript looking at the risk of non-contact ACL rupture in both dogs and humans. I commend you on your collation of such large numbers for GWAS analysis particularly in the dog population where acquisition of these such well profiled samples is so challenging. The addition of the large human datasets is also extremely novel and the results add credence to the heritability of this condition in at risk individuals and may be crucial in developing programmes for future personalised interventions to avoid an ACL tear which would be very impactful for this condition.

I would like to ask the authors if they were able to add the signalment of the dogs and humans into their analysis or is this a future objective if this project?

I would also suggest that the authors should try and simplify the terminology in the text in terms of acronyms and abbreviations as much as possible if they can.

Reviewer #2

(Remarks to the Author)

This manuscript investigates and compares the genetic architecture of non-contact ACL rupture in dogs and humans. They find a highly polygenic architecture of the trait, with similar degrees of polygenicity in both species, suggesting that this trait should be considered as a heritable disease rather than an injury event. Through GWAS, they also find susceptibility genes shared between dogs and humans and evaluate the performance of polygenic risk score prediction mainly in dogs with different breeds. While the results of this study is interesting and such cross-species genetic study in general is valuable, I find some conclusions tend to be over stated and the manuscript could be strengthened by more in-depth analyses. My specific comments are as below.

Some conclusion statements may be too strong. Such as “PRS prediction is accurate at predicting risk of non-contact ACL rupture in dogs” – it’s more useful to state what the prediction accuracy is; “Remarkably similar results were obtained for humans” – however, from the first paragraph in Results, $h^2_I = 0.52$ in dogs but ~ 0.3 in humans.

Related to the comment above, my understanding is that the difference in h^2_I between dogs and humans is in part attributed to the pedigree relationships. There are relatives in dogs but not in human samples. If so, this needs to be noted, and the terminology “SNP-based heritability” should be used in human samples.

The posterior standard error for S is quite large in both canine and human samples. The conclusion that the trait “is under negative selection” should be toned down, e.g., “is likely under negative selection”. In addition to the posterior standard error, they could report the posterior probability that S is negative to support the conclusion.

It is interesting that they found 16 disease genes shared between humans and dogs. However, more in-depth analysis is warranted. For example, is 16 significantly higher than sampling genes at random? Compared to the 50 shared genes in the literature, how many are novel findings? Do they explain similar amount of variance in different species? Do they share any putative causal variants and are those variants had similar effect size or same effect direction?

I wonder if there is any particular reason that they chose (S)BayesR for genetic architecture estimation but not for polygenic risk score prediction, given that (S)BayesR is actually designed to improve polygenic prediction.

A large block of text is used to compare prediction accuracy between prediction methods within dogs or humans. However, it seems for this project the focus should be comparison between dogs and humans. In my opinion, it would be more interesting if they can evaluate cross-species prediction between dogs and humans.

Minor:

Page 4, "we used 4,753 genomic windows". Not clear how the genomic windows are defined when reading up to here. The "30 SNPs per window" in the subsequent sentences should be used earlier.

Page 4, "we performed an analysis of GWAS summary statistics for shared risk genes". Not clear if they perform gene-based tests or simply map the nearest gene to the GWAS loci.

Page 6, "No one has ever undertaken joint analysis of dogs and humans, and this approach is a key innovation." I don't think they can call it a "joint analysis" because they analysed canine and human samples separately.

Page 6, "Our SNP-based estimates of dog non-contact ACL rupture heritability are remarkably like the high value of 0.69 described in a recent human twin study (7)." I think "remarkably like" is too strong.

Reviewer #3

(Remarks to the Author)

Although the molecular mechanisms are poorly understood, genetic predisposition to non-contact ACL ruptures in both humans and canines has been established. The authors investigated genetic contribution to non-contact ACL rupture in both species and described these injuries (or disease) as a moderately heritable polygenic disease with shared risk genes in humans and canines.

This is an extremely novel body of work which in my opinion makes a significant contribution to the field. I don't have any major or minor concerns.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

Thank you for your considered response and revision of your manuscript .
I am happy with your response to my comments.

Reviewer #3

(Remarks to the Author)

The authors have addressed the reviewers comments.

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Reviewer #1 Comments

This is an extremely well written and novel manuscript looking at the risk of non-contact ACL rupture in both dogs and humans. I commend you on your collation of such large numbers for GWAS analysis particularly in the dog population where acquisition of these such well profiled samples is so challenging. The addition of the large human datasets is also extremely novel and the results add credence to the heritability of this condition in at risk individuals and may be crucial in developing programmes for future personalised interventions to avoid an ACL tear which would be very impactful for this condition.

I would like to ask the authors if they were able to add the signalment of the dogs and humans into their analysis or is this a future objective if this project?

Authors response

The public human summary statistics used for GWAS are derived from other publications as described in our manuscript. Signalment for these subjects is not available. It is a future goal of this research to perform an updated GWAS using UK Biobank individual data for which sex and age as well as other relevant covariates could be reported, and this is more clearly stated in the manuscript now (lines 299-300).

For the dog GWAS, for the subjects recruited at UW Madison, sex, neutering, age, weight, and coat color were also recorded. The intent is to consider these covariates further in dog GWAS in another paper focused on PRS prediction in veterinary medicine. However, the dog signalment covariates are not directly relevant to the cross-species analysis. Because we wanted to undertake a consistent analysis in humans and dogs, they were not considered in detail in the current report. The manuscript writing has been revised to clarify (lines 335-340).

I would also suggest that the authors should try and simplify the terminology in the text in terms of acronyms and abbreviations as much as possible if they can.

Authors response

Thank you for this helpful comment. The manuscript has been revised to remove several acronyms to address this concern and clarify the writing.

Reviewer #2 Comments

This manuscript investigates and compares the genetic architecture of non-contact ACL rupture in dogs and humans. They find a highly polygenic architecture of the trait, with similar degrees of polygenicity in both species, suggesting that this trait should be considered as a heritable disease rather than an injury event. Through GWAS, they also find susceptibility genes shared between dogs and humans and evaluate the performance of polygenic risk score prediction mainly in dogs with different breeds. While the results of this study is interesting and such cross-species genetic study in general is valuable, I find some conclusions tend to be over stated and the manuscript could be strengthened by more in-depth analyses. My specific comments are as below.

Authors response

Thank you for this helpful comment. In revising the manuscript in response to your constructive comments, the conclusions have been revised to reflect your feedback. Where relevant more in-depth analysis has been undertaken to support the findings. These changes have been highlighted in the revised manuscript.

Some conclusion statements may be too strong. Such as “PRS prediction is accurate at predicting risk of non-contact ACL rupture in dogs” – it’s more useful to state what the prediction accuracy is; “Remarkably similar results were obtained for humans” – however, from the first paragraph in Results, $h^2_l = 0.52$ in dogs but ~ 0.3 in humans.

Authors response

Thank you for this feedback. The manuscript has been revised as suggested (lines 103-106, 227-228).

Related to the comment above, my understanding is that the difference in h^2_l between dogs and humans is in part attributed to the pedigree relationships. There are relatives in dogs but not in human samples. If so, this needs to be noted, and the terminology “SNP-based heritability” should be used in human samples.

Authors response

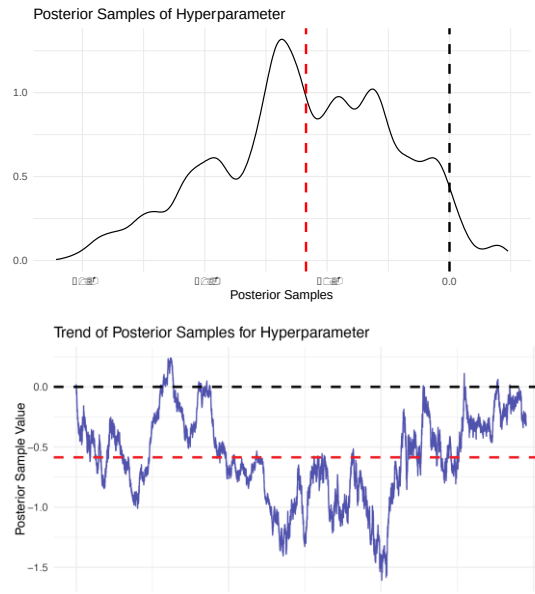
Thank you for this feedback. The manuscript has been revised as suggested (lines 112-122, 229-231).

The posterior standard error for S is quite large in both canine and human samples. The conclusion that the trait “is under negative selection” should be toned down, e.g., “is likely under negative selection”. In addition to the posterior standard error, they could report the posterior probability that S is negative to support the conclusion.

Authors response

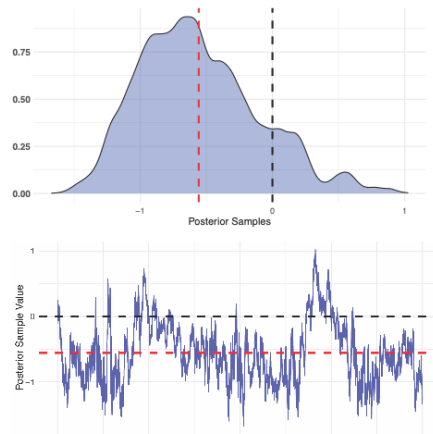
Thank you do much for the helpful comments. We have undertaken additional analysis and reported more statistics to highlight how the S value has a negative tendency distribution in both species. These new results have been added to the manuscript (lines 131-143, Supplementary Figure S4).

In human, the posterior analysis of the hyperparameter S from the Bayesian regression model reveals a strong tendency towards negative values. The mean and median are both negative, with the mean at -0.59 and the median at -0.590 too, and the mode was -0.686. Approximately 97% of the samples fall in the negative range, indicating a predominant negative skew. The standard deviation was 0.360, reflecting moderate variability. The 95% HPD interval ranges from -1.285 to 0.054, confirming that most of the credible interval is in the negative space.



The red dashed line represents mean of posterior samples, and the black dashed line represent 0 values as separating positive and negative spaces.

The analysis of the hyperparameter S in dog populations showed a mean of -0.558 and a median of -0.613 and also a mode of -0.64, indicating a central tendency towards negative values. About 87% of the posterior samples were negative, suggesting a strong negative tendency. The 95% HPD interval ranges from -1.363 to 0.314, with most of the credible interval falling in the negative range.



The red dashed line represents mean of posterior samples, and the black dashed line represent 0 values as separating positive and negative spaces.

It is interesting that they found 16 disease genes shared between humans and dogs. However, more in-depth analysis is warranted. For example, is 16 significantly higher than sampling genes at random? Compared to the 50 shared genes in the literature, how many are novel findings? Do they explain similar amount of variance in different species? Do they share any putative causal variants and are those variants had similar effect size or same effect direction?

Authors response

We added extra analysis and investigated more gene functions (Lines 426-428, Supplementary Information). Fisher's Exact Test was used to determine if whether the 16 shared genes are significantly higher than sampling genes at random by chance. The *P*-value from Fisher's Exact Test was $P=0.0083$. Common significance levels (α) are 0.05 or 0.01. Consequently, we reject the null hypothesis. This suggests that the overlap of the 16 shared genes between the dog and human GWAS results is significant and not likely due to random chance implying meaningful biological connections (Figure 5).

We added and highlighted extra comparative analysis by reviewing the literatures regarding the functionality of these 16 shared genes (Supplementary Information file). It is worth to note that there are only a handful of articles related to genes specifically playing role in ACL rupture so hopefully our discovery paper can open a new avenue for more research in this regard. So, all these 16 genes are new, and we believe could be risk genes related to ACL rupture pathogenesis. Many of the genes have regulatory effects on skeletal muscle suggestion neuromuscular control of locomotion may be an underappreciated mechanism leading to ACL fatigue injury and eventual rupture (lines 144-156, 280-283).

1. **STK32B**. Encodes a protein kinase involved in transferring phosphate molecules to serine and threonine residues. Mutations or deletions in this gene can cause skeletal dysplasia. In the context of ACL rupture, STK32B might affect the structural integrity of skeletal muscle and connective tissues due to its role in phosphorylation processes essential for cellular functions and tissue stability.
2. **ARID1B**. Alterations in this gene can lead to low muscle tone and other health issues. About 80% of individuals with ARID1B-related syndrome exhibit low muscle tone, which may influence muscle strength and joint stability. This could increase susceptibility to ACL injuries due to weakened muscle support.
3. **LOXHD1**. Associated with mechanotransduction defects, which can affect how muscles respond to mechanical stress and strain. Disruptions in mechanotransduction could impair muscle adaptation to mechanical forces, potentially contributing to ACL rupture susceptibility.
4. **RSBN1L**. May be involved in the maintenance of muscle mass, as it is expressed in skeletal muscle tissue. Proper maintenance of muscle mass is crucial for supporting joint stability and function, which can influence the risk of ACL rupture.
5. **NIBAN2**. Plays a role in cell survival under stress conditions, potentially impacting muscle tissues through stress response pathways. Enhanced stress response might influence muscle resilience and recovery, indirectly affecting risk of ACL rupture.
6. **TEX9**. Involved in energy-related processes within skeletal muscle cells. Efficient energy metabolism is essential for muscle contraction and overall muscle health, which can impact muscle strength and joint stability, influencing ACL rupture risk.
7. **GRIN2A**. Modulates excitatory neurotransmission, critical for muscle contraction. Proper neurotransmission is vital for effective muscle function and coordination, which can affect muscle support and stability around the ACL.
8. **MBTPS1**. Regulates lipid metabolism, stress responses, and the unfolded protein response (UPR) in the endoplasmic reticulum. Disruptions in these processes can affect muscle cell function and stress responses, potentially impacting muscle health and joint stability.

9. **HSDL1**. Involved in osteoclast function, osteoblast differentiation, and bone formation. These processes are critical for skeletal development and bone health. Disruptions can affect bone density and strength, indirectly influencing the risk of ACL injuries.
10. **DNAAF1**. Affects the assembly of dynein arms and regulates cilia and microvilli. Since cilia and microvilli are involved in cellular signaling and mechanotransduction, their dysfunction could impact muscle cell function and stability, potentially contributing to risk of ACL rupture.
11. **TAF1C**. A transcription factor involved in the TFIID complex, essential for transcription of protein-coding genes. Its role in skeletal muscle development and function is still under investigation, but disruptions could impact muscle growth and repair, influencing risk of ACL rupture.
12. **PTPRM**. A protein tyrosine phosphatase involved in cell signaling that regulates various cellular processes including growth and differentiation. Its role in skeletal muscle might impact muscle development and joint stability, potentially affecting risk of ACL rupture.
13. **RAB31**. Plays a role in TGF- β -induced actin remodeling and cytoskeleton rearrangement. Proper cytoskeleton dynamics are crucial for muscle cell function and stability, and disruptions could impact muscle homeostasis and neuromuscular control.
14. **PTPRT**. A non-receptor-type protein tyrosine phosphatase that interacts with insulin receptors and substrates. Its role in cellular signaling could affect muscle cell function and repair mechanisms, which may influence risk of ACL rupture.
15. **DSCAM**. Involved in cell adhesion processes. Proper cell adhesion is important for maintaining tissue integrity and stability, including in muscle and connective tissues. Disruptions could affect muscle and joint stability, potentially impacting risk of ACL rupture.
16. **RNF165**. Also known as Ark2C, is an E3 ubiquitin ligase that enhances bone morphogenetic protein signaling. Efficient bone morphogenetic protein signaling mediated by RNF165 might contribute to the maintenance of ligament strength and resilience. Deficiencies or dysregulation in this signaling could lead to weaker ligaments and an increased susceptibility to ACL rupture.

I wonder if there is any particular reason that they chose (S)BayesR for genetic architecture estimation but not for polygenic risk score prediction, given that (S)BayesR is actually designed to improve polygenic prediction.

Authors response

We did use it for polygenic risk score prediction and this is stated in the method section (lines 474-478).

A large block of text is used to compare prediction accuracy between prediction methods within dogs or humans. However, it seems for this project the focus should be comparison between dogs and humans. In my opinion, it would be more interesting if they can evaluate cross-species prediction between dogs and humans.

Authors response

From a clinical perspective, we respectfully disagree with this comment if you are suggesting that a dog training set can be used to predict a human test set or vice versa. Species comparison can easily be made at the gene level, but not at the SNP level. We think that the main value of the cross-species genomic analysis is to identify ACL rupture risk genes, particularly those that

are shared between the two species. Such prior information can then be used to enrich GWAS analysis in both species to increase the power of association for the chosen association analysis. This is an area that we want to pursue further in future work to improve statistical approaches to cross-species pleiotropy analysis (lines 299-300, 328-330).

Minor comments

Page 4, “we used 4,753 genomic windows”. Not clear how the genomic windows are defined when reading up to here. The “30 SNPs per window” in the subsequent sentences should be used earlier.

Authors response

Thank you for this helpful comment. The text has been revised to clarify as recommended (lines 123-125).

Page 4, “we performed an analysis of GWAS summary statistics for shared risk genes”. Not clear if they perform gene-based tests or simply map the nearest gene to the GWAS loci.

Authors response

Thank you for this comment. The text has been revised to clarify as recommended (lines 144-147).

Page 6, “No one has ever undertaken joint analysis of dogs and humans, and this approach is a key innovation.” I don’t think they can call it a “joint analysis” because they analysed canine and human samples separately.

Authors response

Thank you for this perspective. The relevant sentence of the discussion has been revised to reflect this comment and more clearly consider what was done in the analysis (lines 223-225).

Page 6, “Our SNP-based estimates of dog non-contact ACL rupture heritability are remarkably like the high value of 0.69 described in a recent human twin study (7).” I think “remarkably like” is too strong.

Authors response

Thank you for this comment. The relevant sentence has been revised in response to this comment (lines 227-228).

Reviewer #3 Comments

Although the molecular mechanisms are poorly understood, genetic predisposition to non-contact ACL ruptures in both humans and canines has been established. The authors investigated genetic contribution to non-contact ACL rupture in both species and described these injuries (or disease) as a moderately heritable polygenic disease with shared risk genes in humans and canines.

This is an extremely novel body of work which in my optioning makes a significant contribution to the field. I don't have any major or minor concerns.

Authors response

We sincerely appreciate your positive and encouraging feedback regarding our work on the genetic predisposition to non-contact ACL ruptures in both humans and dogs. Your recognition of the novelty and significance of our findings means a great deal to us. We are grateful for your kind words and are pleased that you find our research to be a valuable contribution to the field.

Reviewer #1 Comments

Thank you for your considered response and revision of your manuscript. I am happy with your response to my comments.

Authors response

Thank you for the thoughtful review of our paper.

Reviewer #3 Comments

The authors have addressed the reviewer's comments.

Authors response

Thank you for careful review of our paper.