Follow us @BoneJointOpen

Freely available online OPEN ACCESS

BJO

KNEE

Three genes associated with anterior and posterior cruciate ligament injury

A GENOME-WIDE ASSOCIATION ANALYSIS

S. K. Kim,
C. Nguyen,
A. L. Avins,
G. D. Abrams

From Stanford University School of Medicine, California, USA

Aims

The aim of this study was to screen the entire genome for genetic markers associated with risk for anterior cruciate ligament (ACL) and posterior cruciate ligament (PCL) injury.

Methods

Genome-wide association (GWA) analyses were performed using data from the Kaiser Permanente Research Board (KPRB) and the UK Biobank. ACL and PCL injury cases were identified based on electronic health records from KPRB and the UK Biobank. GWA analyses from both cohorts were tested for ACL and PCL injury using a logistic regression model adjusting for sex, height, weight, age at enrolment, and race/ethnicity using allele counts for single nucleotide polymorphisms (SNPs). The data from the two GWA studies were combined in a meta-analysis. Candidate genes previously reported to show an association with ACL injury in athletes were also tested for association from the meta-analysis data from the KPRB and the UK Biobank GWA studies.

Results

There was a total of 2,214 cases of ACL and PCL injury and 519,869 controls within the two cohorts, with three loci demonstrating a genome-wide significant association in the meta-analysis: *INHBA, AEBP2,* and *LOC101927869*. Of the eight candidate genes previously studied in the literature, six were present in the current dataset, and only *COL3A1* (rs1800255) showed a significant association (p = 0.006).

Conclusion

Genetic markers in three novel loci in this study and one previously-studied candidate gene were identified as potential risk factors for ACL and PCL injury and deserve further validation and investigation of molecular mechanisms.

Cite this article: ***Bone Jt Open*** **2021;2-6:414–421.**

Keywords: knee, genetics, anterior cruciate ligament, posterior cruciate ligament

Introduction

The anterior cruciate ligament (ACL) and posterior cruciate ligament (PCL) are critical structures within the knee, allowing for tibiofemoral sagittal and rotational plane stability with motion and athletic activity. ACL injuries are common in both recreational and elite athletes and often lead to surgical intervention to reconstruct the ligament. While not occurring at the same rate as ACL injuries, PCL injuries can also lead to significant disability and recovery time following injury.

Less attention has been paid to the genetic risk factors associated with ACL injury, and very little data is available to assess the genetic risk factors associated with PCL injury. Two genetic approaches have been employed in the past to uncover genetic risk factors for ACL injury. In the first, candidate gene studies have tested a small number of DNA polymorphisms in genes known to have biological functions in ACL biology.[1,2] These studies have reported single-nucleotide polymorphisms (SNPs) in eight candidate genes to show an association with ACL rupture in athletes.[1] In the second approach,

Correspondence should be sent to Geoffrey D Abrams; email: geoffa@stanford.edu

doi: 10.1302/2633-1462.26.BJO-2021-0040.R1

Bone Jt Open 2021;2-6:414–421.

a genome-wide association screen (GWAS) was used to screen millions of polymorphisms spanning the entire genome for those showing the strongest association with ACL injury.[2] The advantages of a GWAS are that it reports the strongest signals from across the entire genome, and the criteria for statistical significance are well-developed which aids in reproducibility in validation studies. The main disadvantage of GWA studies is that large cohorts are required to achieve statistical significance to account for the large number of tested polymorphisms (multiple hypothesis correction). The previous genome-wide association study for ACL injury did not return any results that demonstrated significant genetic associations, potentially due small sample size.[2]

The purpose of this study was to perform a screen of the entire genome for polymorphisms associated with either ACL or PCL injury using the Kaiser Permanente Research Board (KPRB) and the UK Biobank datasets. Using our dataset, we then sought to validate candidate genes previously-reported to show an association with ACL injury.

Methods

GWAS for ACL and PCL injury were performed using data from the KPRB and from the v3 release of the UK Biobank. Institutional review board approval for data analysis was not required due to the use of de-identified information being utilized.

KPRB cohort. Our analysis cohort included 83,414 individuals of European ancestry who were genotyped at 670,572 SNPs. We can infer the genotype of many more polymorphisms using the sequence data from the large number of complete genomes using a process termed imputation. Imputation was performed by pre-phasing the genotypes with Shape-IT v2.r644 and then imputing to a cosmopolitan reference panel consisting of all individuals from the 1,000 Genomes Project using IMPUTE2 v2.2.2 and standard procedures with a cutoff of $R^2 > 0.3$. The final number of SNPs following imputation was 12,365,897. The quality of the imputed data was previously validated.[3]

UK biobank cohort. Genotype data were obtained from the v3 release of the UK Biobank.[4] The UK Biobank electronic healthcare records were available for 438,669 individuals. Genotype data were imputed centrally by the UK Biobank with IMPUTE2 using the Haplotype Reference Consortium and the UK10k + 1000GP3 reference panels.[5] Metrics for quality control were established and then used to filter DNA variants by UK Biobank.[4] Imputed SNPs were excluded if they had an IMPUTE2 info score < 0.4 (the IMPUTE2 info score indicates the accuracy of the SNPs whose genotype was imputed, not directly genotyped).

Database quality control. For both the KPRB and the UK Biobank cohorts, individuals were excluded if they were outliers based on genotyping missingness rate, whose sex inferred from the genotypes did not match their self-reported sex, who withdrew from participation, or who were not of European ancestry. Genotype missingness identifies SNPs where many calls are null indicating that the genotype is not reliable. The purpose of restricting individuals to those with European ancestry is to reduce population stratification in the study; for example, if the risk of ACL and PCL injury among individuals with African ancestry is higher than that for European individuals, then any SNP with an allele frequency that is different between African and European ancestries would appear to be associated with ACL and PCL injury. Overall, these filters resulted in excluding 102,230 individuals (18.9%) and 2,668 individuals (3.1%) (mostly due to the ancestry filter) in the KPRB and UK Biobank cohorts, respectively. Genetic variants were excluded that failed quality control procedures in any of the genotyping batches, that showed a departure from Hardy-Weinberg of $p < 10^{-50}$ or that had a Minor Allele Frequency < 0.001. Determination of genetic ancestry was performed by principal component analysis (PCA) calculated centrally by either KPRB or UK Biobank, as previously described.[4]

Phenotype definitions. In the KPRB cohort, ACL and PCL injury cases were identified based on clinical diagnoses captured in the Kaiser Permanente Northern California electronic health record. International Classification of Diseases, Ninth Revision (ICD-9)[6] or International Classification of Diseases, Tenth Revision (ICD-10)[7] codes were used to identify cases of ACL and PCL injury (Table I). In the UK Biobank cohort, ACL and PCL injury cases were identified from inpatient data (ICD-9, ICD-10) and primary care data (Read v2 or Read v3) (Table I).

Genome-wide association. GWA studies were conducted using PLINK v2.0a2.[2] SNP associations were tested with ACL and PCL injury with a logistic regression model using allele counts for typed and imputed SNPs. The model was adjusted for genetic sex, height, weight, and race/ethnicity using ten principal components. For the UK Biobank, the age of enrolment was also included as an adjustment. The final number of SNPs that was analyzed was 12,365,897 in the KPRB cohort and 17,136,336 in the UK Biobank cohort. To account for inflation due to population stratification, the genomic control parameter (λgc) was calculated ($\lambda gc = 1.001$ for KPRB, and $\lambda gc = 0.927$ for the UK Biobank). The genomic control parameter is used because the p-values are not normally distributed. To account for this, p-values were adjusted for the genomic control in each population.

Results using odds ratios per allele from each cohort were combined by inverse-variance, fixed-effects meta-analysis. Here, meta-analysis refers to a statistical method to combine data from GWAS's performed on two independent cohorts. A p-value < 5×10^{-8} was used as a threshold for genome-wide significance. Power calculations were

Table I. Phenotype definitions.

Code	Description	Cases, n
KPRB		
ICD-9		
844.2	Sprain of cruciate ligament of knee	1,304
717.83	Old disruption of anterior cruciate ligament	456
717.84	Old disruption of posterior cruciate ligament	26
ICD-10		
S83.509A	Sprain of unspecified cruciate ligament of unspecified knee, initial encounter	9
S83.511A	Sprain of anterior cruciate ligament of right knee, initial encounter	83
S83.512A	Sprain of anterior cruciate ligament of left knee, initial encounter	86
S83.519A	Sprain of anterior cruciate ligament of unspecified knee, initial encounter	7
S83.521A	Sprain of posterior cruciate ligament of right knee, initial encounter	6
S83.522A	Sprain of posterior cruciate ligament of left knee, initial encounter	10
	Total unique cases	1,328
	Total unique controls	82,086
UK Biobank		
ICD-9		
844.2	Sprain of cruciate ligament of knee	15
ICD-10		
S83.5	Sprain and strain involving (anterior)(posterior) cruciate ligament of knee	302
M23.61	Other spontaneous disruption of anterior cruciate ligament of knee	163
M23.62	Other spontaneous disruption of posterior cruciate ligament of knee	6
Read V2		
S542.	Sprain of cruciate ligament of knee	42
S5421	Partial tear, knee, anterior cruciate ligament	80
S5C3.	Complete tear, knee, anterior cruciate ligament	87
N07yD	Old partial tear anterior cruciate ligament	8
N07yE	Old complete tear anterior cruciate ligament	1
7K6PL	Reconstruction of anterior cruciate ligament of knee	155
N07y2	Old anterior cruciate ligament disruption	10
Read V3		
S542.	Sprain of cruciate ligament of knee	84
S5421	Partial tear, knee, anterior cruciate ligament	92
N07yD	Old partial tear anterior cruciate ligament	8
N07yE	Old complete tear anterior cruciate ligament	6
N07y2	Old anterior cruciate ligament disruption	43
S5C3	Complete tear, knee, anterior cruciate ligament	108
	Total unique cases	886
	Total unique controls	437,783

ICD-9, International Classification of Diseases, Ninth Revision; ICD-10, International Classification of Diseases, Tenth Revision; KPRB, Kaiser Permanente Research Board.

conducted with the software using the Genetic Association Study Power Calculator.[8]

The observed p-values were compared to the distribution of p-values expected by chance in a Q-Q plot (Supplemental Figure aa). The black dots deviate from the red line for the lowest observed p-values in the upper right-hand corner, indicating that the observed association signals are significantly stronger than the signals that would be expected by chance. The p-values from every SNP in the meta-analysis are shown in a Manhattan plot (Supplemental Figure ab).

Table II. Validation of candidate gene studies.

			ACL only		ACL/PCL combined		
SNP	**Gene**	**EA**	**OR**	**p-value***	**OR**	**p-value***	**Ref**
rs1516797	ACAN	T	1.070	0.260	1.070	0.130	1
rs516115	DCN	A	0.930	0.340	0.990	0.820	1
rs970547	COL12A1	T	1.090	0.230	1.050	0.130	7
rs2276109	MMP12	T	0.940	0.570	1.020	0.570	8
rs1800255	COL3A1	A	1.130	0.065	1.100	0.0062	9
rs331079	FBN2	G	0.990	0.930	1.050	0.350	10

*Logistic regression.
ACL, anterior cruciate ligament; EA, effect allele; OR, odds ratio; PCL, posterior cruciate ligament; SNP, single nucleotide polymorphism.

Further bioinformatics investigations of the top genome-wide significant loci from the GWAS were conducted. QQ and Manhattan plots were created using qqman. Regional association plots were generated for each locus with LocusZoom.[9] The genomic context of each SNP was investigated using RegulomeDB[10] web tools. ChIP seq data from the ENCODE project was used to determine whether SNPs were located within transcription factor binding sites.[11] Chip seq (Chromatin Immunoprecipitation followed by DNA sequencing) is a technique where transcription factors are immunoprecipitated from chromatin and the bound DNA is sequenced, revealing DNA sites that are bound in vivo. Summary statistics for all SNPs from the GWAS will are available at NIH GRASP.[12]

Testing of previously identified candidate genes. Candidate genes were tested for validation when the relevant polymorphisms were present in the summary statistics from the genome-wide analyses performed using the KPRB and UK Biobank data (Table II). Because only a small number of SNPs are being tested, the threshold for statistical significance can be much lower than the genome-wide threshold that was used for the genome-wide study above ($p < 5 \times 10^{-8}$). It is still important, however, to adjust the p-value threshold to compensate for multiple testing. The Bonferroni method was used to set the p-value threshold at $p = 0.05/6 = 8.3 \times 10^{-3}$.

Ethical approval. This study analyzed stored data from the KPRB and UK Biobank subjects who consented to genomic testing and use of their genomic data, as well as health data from the KPNC and UK Biobank electronic health records. The health and genotype data for the subjects were de-identified. All study procedures were approved by the Institutional Review Board of the Kaiser Foundation Research Institute. Patients were not involved

Table III. Study demographics.

Variable	Case	Control	p-value
KPRB			
Female (% injured)	930 (15.6)	58,656 (84.4)	0.991*
Male (% injured)	674 (15.6)	42,527 (84.4)	
Height, inches (SD)	67.1 (3.99)	66.7 (4.0)	0.0002†
Weight, lbs (SD)	170.6 (38.3)	170.4 (39.3)	0.960†
UK Biobank			
Female (% injured)	285 (0.13)	225,388 (99.87)	< 0.001*
Male (% injured)	400 (0.21)	186,681 (99.79)	
Height, cm (SD)	171.5 (8.78)	168.5 (9.23)	< 0.001†
Weight, kg (SD)	80.6 (15.2)	78.3 (15.9)	0.0001†
Age at enrolment, yrs (SD)	51.6 (7.78)	57.0 (8.01)	0.0001†

*Chi-squared test.
†Independent samples *t*-test.
KPRB, Kaiser Permanente Research Board; SD, standard deviation.

in this research, except as anonymized subjects in the two cohorts.

Results

Identification of DNA variants associated with ACL and PCL injury. GWA analyses for ACL and PCL injury were performed with the KPRB (83,414 individuals) and UK Biobank (438,669 individuals) cohorts using sex, weight, height and age at enrolment as adjustments (Table III). For KPRB, there were 1,328 cases (1.59%) of ACL and PCL injury and 82,086 (98.41%) controls (Table I). For UK Biobank, there were 886 cases (0.201%) and 437,926 controls (99.799%) (Table I). There were three SNPs with genome-wide significant associations with ACL and PCL injury using $p = 5 \times 10^{-8}$ as a cut-off (Figure 1; Table IV). rs144051132 is located on chromosome 7 in the 3' region of the *Inhibin, β A* gene (*INHBA*), which encodes a signalling molecule in the TGF-beta family that is a growth/differentiation factor during development, including growth of osteoblastic cells involved in bone development (Figure 1a).[13] rs186727643 is located on chromosome 12 in the 3' region of the *AE binding protein* gene (*AEBP2*) which is a DNA-binding transcriptional repressor (Figure 1b). rs188099931 is located on chromosome 21 in the 3' region of the long non-coding RNA gene (*LOC101927869*) of unknown function (Figure 1c).

None of the SNPs affect either protein coding or are known to be associated with changes in expression of a nearby gene. However, ChIP seq experiments show that rs144051132 near the INHBA gene is located in the binding site for the FOS transcription factor, directly within the eight nucleotide motif bound by the transcription factor.[11] Figure 2 shows the canonical nucleotide motif bound by FOS; the fifth position is changed from a T to an A by rs144051132. These observations suggest a model whereby the A allele of rs144051132 interferes with binding of the FOS transcription factor leading to alteration of expression of a nearby gene (likely INHBA), which may in turn lead to increased risk of ACL and PCL injury.

Some of the medical codes used to define the cases of ACL and PCL injury combined both injuries together. To interrogate whether the three SNPs show an association with ACL injury specifically, the genetic association test was repeated using ICD-9 and ICD-10 codes from the KPRB cohort that were specific for ACL injury (Table I). There were 473 cases specifically diagnosed as ACL jury out of 1328 total cases. Both the *INHBA* (rs144051132) and *AEBP2* (rs186727643) SNP showed a significant association with ACL injury ($p < 0.05$o, logistic regression) and the *LOC101927869* (rs188099931) SNP showed an association that was borderline significant ($p = 0.053$, logistic regression) (Table IV). These results suggest that the three loci are associated with ACL injury. Whether or not the three loci are also associated with PCL injuries is unclear as there were not enough PCL-specific cases to test.

Validation of previous candidate gene studies. Previous studies have reported nine SNPs in eight candidate genes to show an association with ACL rupture using $p < 0.05$ as a cutoff (Table II).[1] Of these SNPs, six were contained in our dataset. We first attempted to replicate the previous results for these candidate SNPs using the subset of cases in our dataset that were specific to ACL injury (567 cases) (Table II). None of the six SNPs were nominally significant for association with ACL injury. Next, we examined the candidate genes in our full dataset that contains 2,214 cases of combined ACL and PCL injuries, as the increased sample size would improve the statistical power of the analysis. This time, we found that *COL3A1* (rs1800255) was validated with a p-value = 0.0062 (logistic regression) (Table II). The remaining five candidate genes did not show significant association in our combined ACL and PCL data set.

Demographics of ACL and PCL injuries. There was a higher incidence of cases of ACL and PCL injury in the KPRB

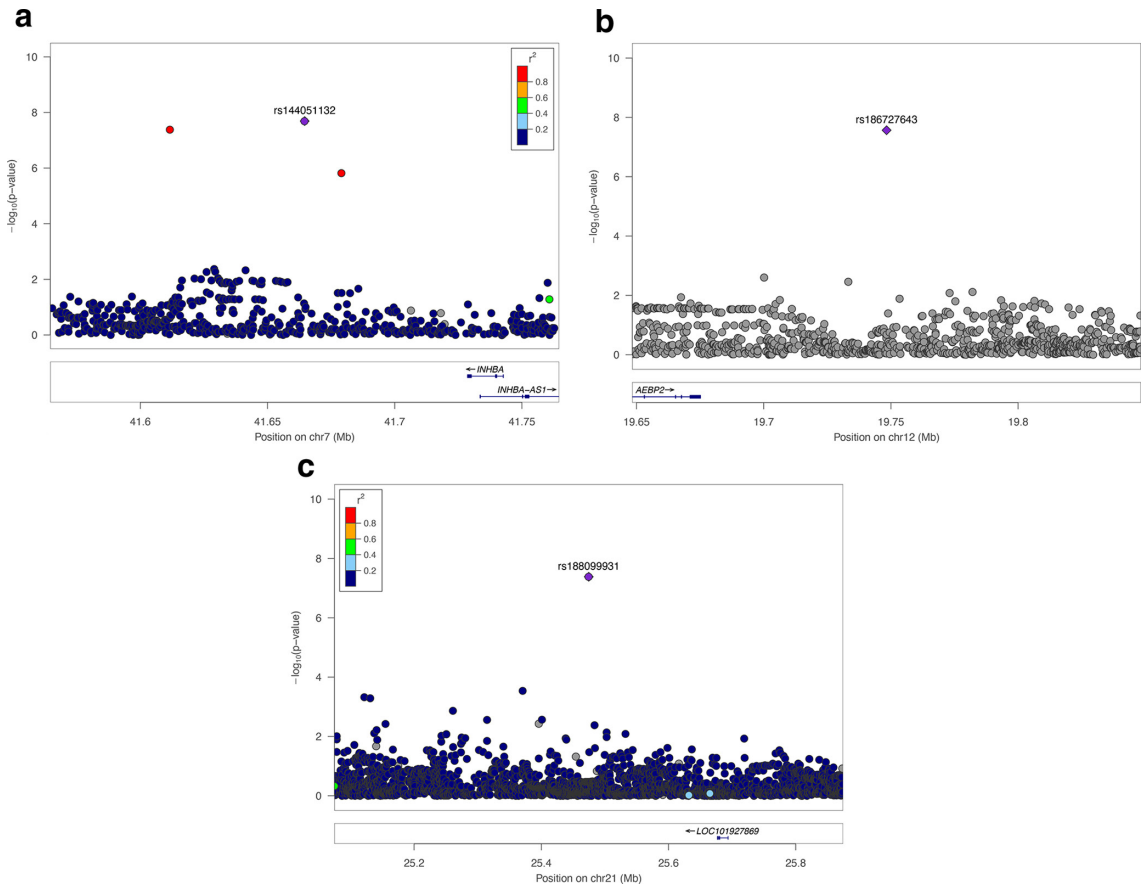

Fig. 1

Regional-association plots. Tested SNPs are arranged by genomic position around the lead SNP (purple diamond). The y-axis indicates -$\log_{10}$ p-values for association with ACL and PCL injury for each SNP. The color of dots of the flanking SNPs indicates their linkage disequilibrium (R^2) with the lead SNP as indicated by the heat map color key. a) Regional-association plot for rs144051132 with ACL and PCL injury, which is located in the 3' region of *INHBA*. b) Regional-association plot for rs186727643 with ACL and PCL injury, which is located in the 3' region of the *AEBP2* gene. c) Regional-association plot for rs188099931 with ACL and PCL injury, which is located in the 3' region of the *LOC101927869* gene.

Table IV. Summary statistics.

A. Meta-analysis							**Meta-analysis**	
Chr	**BP**	**SNP**	**Gene**	**EA**	**EA freq UK Biobank**	**EA freq KPRB**	**OR**	**p-value***
7	41664597	rs144051132	INHBA	A	0.0045	0.0018	2.8784	< 0.001
12	19748303	rs186727643	AEBP2	T	0.00201	0.001274	4.4986	< 0.001
21	25474591	rs188099931	LOC101927869	G	0.004689	0.003741	2.5655	< 0.001
B. KPRB and UKB GWAS					**UK Biobank GWAS**		**KPRB GWAS**	
		SNP	**Gene**	**EA**	**OR (95% CI)**	**p-value***	**OR (95% CI)**	**p-value***
		rs144051132	INHBA	A	2.61 (1.45 to 4.72)	<0.001	3.08 (1.88 to 5.03)	< 0.001
		rs186727643	AEBP2	T	2.4 (0.28 to 21.1)	0.390	4.67 (2.70 to 8.08)	< 0.001
		rs188099931	LOC101927869	G	2.59 (1.48 to 4.54)	0.001	2.55 (1.64 to 3.88)	< 0.001
C. ACL specific validation					**KPRB ACL specific**			
		SNP	**Gene**	**EA**	**OR (95% CI)**	**p-value***		
		rs144051132	INHBA	A	2.8 (1.2 to 6.3)	0.012		
		rs186727643	AEBP2	T	4.3 (1.7 to 10.5)	0.001		
		rs188099931	LOC101927869	G	2.1 (0.98 to 4.4)	0.053		

*Logistic regression.

ACL, anterior cruciate ligament; BP, base pair; CI, confidence interval; EA, effect allele; GWAS, genome-wide association screen; KPRB, Kaiser Permanente Research Board; OR, odds ratio; SNP, single nucleotide polymorphism.

Fig. 2

Positional weight matrix for the FOS transcription factor. Shown is the canonical sequence bound by the FOS transcription factor as a positional weight matrix. The red box indicates the position affected by rs144051132, where T in the reference sequence is replaced by an A nucleotide. An A nucleotide at position 5 is predicted to lower FOS binding and is also associated with increased risk for ACL and PCL injury.

cohort (1.5%) than in the UK Biobank cohort (0.2%). The electronic records for both cohorts extend for the entire lifetime of the patient if reported by the patient and recorded by the physician. The difference in the incidence of ACL and PCL injury likely reflects a bias in how this injury is diagnosed in the San Francisco Bay area, USA, versus the UK, although there may also be a slight difference in the true incidence of injury between the two cohorts.

In both cohorts, tall height slightly increased the risk of injury. In the UK Biobank but not the KPRB cohort, males had a significantly greater risk and being heavier had a slightly increased risk of ACL and PCL injury.

Discussion

Genetic markers for ACL and PCL injury. This study provides new information about possible genetic associations with ACL and PCL injury risk. Three loci were associated with ACL and PCL injury in a meta-analysis from two GWAS's performed in this study (*INHBA, AEBP2,* and *LOC101927869*). *INHBA* is involved in the growth of osteoblastic cells during bone development, suggesting that variation in bone growth underlies this gene's role in ligament injury. How *AEBP2* and *LOC101927869* impact ligament injury is currently unclear.

Some of the medical codes used in our analysis do not distinguish between ACL and PCL injuries, and so it is unclear if the association are for both injuries or specific to just one. Six candidate genes that were previously identified to be associated with ACL injuries in athletes were tested, and only COL3A1 was confirmed to be associated with ACL and PCL injury in the current dataset.

Individuals harboring risk alleles for these four SNPs have an increased risk for ACL and PCL injury (Table IV). Although the risk alleles from the GWA studies are relatively rare (0.1% to 0.4% allele frequency), they confer an increased risk for ACL and PCL injury of between 2.5- to four-fold (Table IV). Genetic testing could provide key information to uninjured athletes and soldiers about their risk for ACL and PCL injury, allowing them to take extra precautions to avoid cruciate ligament injury, such as mandated participation in ACL injury prevention programs. The genetic information could also be used by medical professionals to make more informed decisions regarding ACL and PCL injury diagnosis, management, and return to play.

Validation of candidate gene studies. Previous studies have tested candidate genes for association with ACL injury, based on the known role of those genes with ligament and bone physiology.[9,14–18] Stepien-Slodkowska et al[18] reported an association of rs1800255 in *COL3A1* with ACL injury in 321 Polish skiers. *COL3A1* encodes type III collagen, which is involved in the repair of connective tissue such as ACL and PCL. This association was supported from our meta-analysis data involving 2,071 cases. Although statistically significant, it should be noted that the association of rs1800255 with ACL and PCL injury was not high on the list from the entire genome in the meta-analysis; specifically, rs1800255 ranked 53,951 among all of the SNPs in the meta-analysis. By contrast, the three SNPs presented in this study are the top signals in the genome.

Other than COL3A1, the other five candidate genes from previous studies were not validated in the meta-analysis data (Table II). Power calculations indicate a 90% chance of statistical significance if the genotype relative risk of the candidate gene is at least 1.20 and the allele frequency is at least 5%. One explanation for the lack of validation is that the previous studies looked at cases of ACL in athletes, whereas our study looked at individuals from the general population.[14–18] Nevertheless, evidence from many studies suggests that candidate gene associations need to be independently replicated, otherwise their credibility is low.[19,20]

Limitations. Our analysis found only three genome-wide significant signals, possibly because ACL and PCL injury may be poorly documented in these cohorts. This type of misclassification error would mostly tend to dilute the strength of any signals, if present. Alternatively, it could be that the heritability of ACL and PCL injury is low. Another limitation is that the phenotypes were defined

from codes contained in electronic health records, and thus we have no information regarding the clinical scenarios surrounding the event. This would include whether patients had prior ACL and PCL injuries that were not captured, and the force and/or impact velocity of the inciting event. Additionally, the cohort included people regardless of whether or not they participated in a sport. For example, we were unable to discriminate if the ACL and PCL injuries identified in this study were related to participation in sports or from other causes, such as falls or motor vehicle accidents. Furthermore, there was a difference in the incidence of cruciate ligament injuries in the KPRB versus the UK Biobank cohort, potentially leading to more clinically applicability in the USA population. While the reasons for this difference in incidence are unknown, this may be due to increased healthcare utilization (specifically MRI to diagnose cruciate ligament injuries) in the USA versus the UK. Lastly, this study only evaluated individuals from the European ancestry group, and the effect in other ethnicities is unknown.

Future studies. It will be important to replicate these results in independent cohorts, especially for athletes and soldiers. Additional studies are warranted to illuminate the underlying biological mechanism for these genes with ACL and PCL injury. These future studies may provide further evidence for using these genetic polymorphisms as diagnostic markers to help predict which athletes harbor a higher risk for incidence of ACL and PCL injury. Follow-up experiments could look at whether the genetic markers affect other aspects of ACL and PCL injury, such as bone/ligament anatomy, length of recovery time, or response to different types of treatment.

Take home message

- Genetic markers in three novel loci in this study and one previously-studied candidate gene were identified as potential risk factors for anterior cruciate ligament and posterior cruciate ligament injury.
- The genetic markers could inform physicians and athletes about risk for injury.

Supplementary material

Figure showing p-values expected by chance in a Q-Q plot, and the p-values from every SNP in the meta-analysis are shown in a Manhattan plot.

References

1. **September AV**, **Posthumus M**, **Collins M**. Application of genomics in the prevention, treatment and management of Achilles tendinopathy and anterior cruciate ligament ruptures. *Recent Pat DNA Gene Seq.* 2012;6(3):216–223.
2. **Kim SK**, **Roos TR**, **Roos AK**, **et al**. Genome-wide association screens for Achilles tendon and ACL tears and tendinopathy. *PLoS One.* 2017;12(3):e0170422.
3. **Jorgenson E**, **Makki N**, **Shen L**, **et al**. A genome-wide association study identifies four novel susceptibility loci underlying inguinal hernia. *Nat Commun.* 2015;6:10130.
4. **Bycroft C**, **Freeman C**, **Petkova D**, **et al**. The UK Biobank resource with deep phenotyping and genomic data. *Nature.* 2018;562(7726):203–209.
5. **Howie B**, **Marchini J**, **Stephens M**. Genotype imputation with thousands of genomes. *G3.* 2011;1(6):457–470.
6. **Centers for Disease Control and Prevention**. https://www.cdc.gov/nchs/icd/icd9.htm (date last accessed 23 June 2021).
7. **Centers for Disease Control and Prevention**. https://www.cdc.gov/nchs/icd/icd10.htm (date last accessed 23 June 2021).
8. **Skol AD**, **Scott LJ**, **Abecasis GR**, **Boehnke M**. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet.* 2006;38(2):209–213.
9. **Pruim RJ**, **Welch RP**, **Sanna S**, **et al**. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics.* 2010;26(18):2336–2337.
10. **Boyle AP**, **Hong EL**, **Hariharan M**, **et al**. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* 2012;22(9):1790–1797.
11. **ENCODE Project Consortium**. An integrated encyclopedia of DNA elements in the human genome. *Nature.* 2012;489(7414):57–74.
12. **US Department of Health & Human Services**. COVID-19 is an emerging, rapidly evolving situation. https://grasp.nhlbi.nih.gov/FullResults.aspx (date last accessed 26 May 2021).
13. **Hashimoto M**, **Shoda A**, **Inoue S**, **et al**. Functional regulation of osteoblastic cells by the interaction of activin-A with follistatin. *J Biol Chem.* 1992;267(7):4999–5004.
14. **Mannion S**, **Mtintsilana A**, **Posthumus M**, **et al**. Genes encoding proteoglycans are associated with the risk of anterior cruciate ligament ruptures. *Br J Sports Med.* 2014;48(22):1640–1646.
15. **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. *Br J Sports Med.* 2010;44(16):1160–1165.
16. **Posthumus M**, **Collins M**, **van der Merwe L**, **et al**. Matrix metalloproteinase genes on chromosome 11q22 and the risk of anterior cruciate ligament (ACL) rupture. *Scand J Med Sci Sports.* 2012;22(4):523–533.
17. **Khoury LE**, **Posthumus M**, **Collins M**, **et al**. ELN and FBN2 gene variants as risk factors for two sports-related musculoskeletal injuries. *Int J Sports Med.* 2015;36(4):333–337.
18. **Stepien-Slodkowska M**, **Ficek K**, **Maciejewska-Karlowska A**, **et al**. Overrepresentation of the COL3A1 AA genotype in Polish skiers with anterior cruciate ligament injury. *Biol Sport.* 2015;32(2):143–147.
19. **Siontis KC**, **Patsopoulos NA**, **Ioannidis JP**. Replication of past candidate loci for common diseases and phenotypes in 100 genome-wide association studies. *Eur J Hum Genet.* 2010;18(7):832–837.
20. **Ioannidis JP**, **Tarone R**, **McLaughlin JK**. The false-positive to false-negative ratio in epidemiologic studies. *Epidemiology.* 2011;22(4):450–456.

Author information:
- S. K. Kim, PhD, Professor Emeritus
- C. Nguyen, BS, Reseach Assistant
  Developmental Biology, Stanford University School of Medicine, Stanford, California, USA.
- A. L. Avins, MD, MPH, PhD, Researcher, Kaiser Permanente Northern California, Division of Research, Oakland, California, USA.
- G. D. Abrams, MD, Assistant Professor, Orthopaedic Surgery, Stanford University School of Medicine, Stanford, California, USA.

Author contributions:
- S. K. Kim: Designed the study, Analyzed and interpreted the data, Wrote the manuscript.
- C. Nguyen: Analyzed the data, Reviewed the manuscript.
- A. L. Avins: Reviewed the manuscript.
- G. D. Abrams: Designed the study, Interpreted the data, Wrote and reviewed the manuscript,

Funding statement:
- The author or one or more of the authors have received or will receive benefits for personal or professional use from a commercial party related directly or indirectly to the subject of this article.

ICMJE COI statement:
- S. K. Kim reports employment and stock/stock options by AxGen, which are unrelated to this article.

Data sharing:
- Access to data used in this study may be obtained by application to the KPRB via kp.org/researchbank/researchers. A subset of the GERA cohort consented for public use can be found at NIH/dbGaP: phs000674.v3.p3.

Acknowledgements:
- We are deeply indebted to the UK Biobank for providing access to a rich data source, and to the access team for assistance with using the data (Application 17847; "GWAS for risk for sports injuries"). We thank Erik Ingelsson for sharing the database at Stanford containing UK Biobank genotype data, and Chris Chang for help with PLINK. The authors thank the Kaiser Permanente Northern California KPRB team for access to data and assistance in data management. Data used in this study were provided by the Kaiser Permanente Research Bank (KPRB) from the KPRB collection, which includes the Kaiser Permanente Research Programme on Genes, Environment, and Health (RPGEH).

Ethical review statement:
- This study was reviewed by the Kaiser Permanente Research Board.