Journal Pre-proof

Variants within the MMP3 and COL5A1 genes associate with soft tissue injury history in elite male rugby athletes

Jon Brazier, Mark R. Antrobus, Peter C. Callus, Adam J. Herbert, Georgina K. Stebbings, Daniel Martin, Stephen H. Day, Liam P. Kilduff, Mark A. Bennett, Robert M. Erskine, Stuart.M. Raleigh, Tom Cullen, Malcolm Collins, Yannis.P. Pitsiladis, Shane M. Heffernan, Alun G. Williams

PII: S1440-2440(25)00153-7

DOI: https://doi.org/10.1016/j.jsams.2025.05.007

Reference: JSAMS 3147

To appear in: *Journal of Science and Medicine in Sport*

Received date: 8 November 2024

Revised date: 15 April 2025

Accepted date: 14 May 2025

Please cite this article as: J. Brazier, M.R. Antrobus, P.C. Callus, et al., Variants within the MMP3 and COL5A1 genes associate with soft tissue injury history in elite male rugby athletes, *Journal of Science and Medicine in Sport* (2025), https://doi.org/10.1016/j.jsams.2025.05.007

Variants within the *MMP3* and *COL5A1* genes associate with soft tissue injury history in elite male rugby athletes.

Jon Brazier [1*], Mark R. Antrobus [2], Peter C. Callus [3], Adam J. Herbert [4], Georgina K. Stebbings [3], Daniel Martin [3], Stephen H. Day [5], Liam P. Kilduff [6], Mark A. Bennett [6], Robert M. Erskine [7, 12], Stuart. M. Raleigh [8a], Tom Cullen [8b], Malcolm Collins [9], Yannis. P. Pitsiladis [10, 11], Shane M. Heffernan [6], and Alun G. Williams [3, 6, 12.]

[1] Department of Psychology, Sport and Geography, University of Hertfordshire, Hatfield AL10 9AB, UK.

[2] Faculty of Wellbeing, Education and Language Studies, The Open University, Walton Hall, Milton Keynes, MK7, 6AA, UK.

[3] Department of Sport and Exercise Sciences, Manchester Metropolitan University Institute of Sport, Manchester Metropolitan University, Manchester, UK.

[4] Research Centre for Life and Sport Sciences, College of Life Sciences, Birmingham City University, Birmingham, UK.

[5] NHS Scotland Academy, NHS Golden Jubilee, Beardmore Street, Clydebank, G81 4HX, UK.

[6] Applied Sports Science Technology and Medicine Research Centre (A-STEM), Faculty of Science and Engineering, Swansea University, Swansea SA1 8EN, UK,

[7] Research Institute for Sport & Exercise Sciences, Liverpool John Moores University, Liverpool L3 3AF, UK.

[8a] Cardiovascular and Lifestyle Medicine Research Group, CSELS, Coventry University, UK.

[8b] Centre for Physical Activity, Sport and Exercise Sciences, Coventry University, Coventry, UK.

[9] Health through Physical Activity, Lifestyle and Sport Research Centre (HPALS), Department of Human Biology, and the International Federation of Sports Medicine (FIMS) Collaborative Centre of Sports Medicine, University of Cape Town, Cape Town, South Africa.

[10] Department of Sport, Physical Education and Health, Hong Kong Baptist University, Hong Kong, Hong Kong SAR.

[11] Department of Movement, Human and Health Sciences, University of Rome 'Foro Italico', Rome, Italy.

[12] Institute of Sport, Exercise and Health, University College London, London WC1E 6BT, UK.

*Corresponding author:

Dr Jon Brazier,
j.brazier2@herts.ac.uk
School of Life and Medical Science
Room S104, Institute of Sport
University of Hertfordshire
Al10 9EU

Twitter @JonBrazier1

Variants within the *MMP3* and *COL5A1* genes associate with soft tissue injury history in elite male rugby athletes.

Abstract

Objectives:

To investigate associations between genetic variants within *COLGALT1, COL1A1, COL3A1, COL5A1, KDR, MIR608, MMP3, NID1, TIMP2* and *VEGFA* and injury history in elite male rugby athletes.

Design: A case-control genetic association study was conducted on 184 elite male rugby athletes.

Methods: Participants were genotyped for 13 genetic polymorphisms previously associated with soft tissue injury using standard PCR assays. Injury data were collected via a self-reported injury-history questionnaire. Single-locus association and Total Genotype Score (TGS) analyses were conducted using χ^2 tests. In addition, multifactor dimensionality reduction and inferred haplotype analysis were used to identify genetic interactions.

Results:

The TT genotype of *MMP3* rs679620 was underrepresented in the non-injured ligament group compared to the ligament sprain and ligament rupture groups (10%, 32%, 25%; $P < 0.04$, respectively). The T allele of *MMP3* rs679620 was overrepresented in the non-injured tendon group compared to the tendinopathy group (50%, 38%; $P < 0.02$). The proportion of C allele carriers of *COL5A1* rs12722 was higher in the tendon rupture group than the non-injured tendon group (96%, 75%; $P < 0.02$). Furthermore, the T-C inferred haplotype frequency of *COL5A1* rs12722 and *COL5A1* rs3196378 was higher in the tendon rupture, ligament sprain and total injured athlete groups compared to their respective non-injured groups ($P < 0.02$).

Conclusion:

This study is the first to identify associations between *MMP3* rs679620 and *COL5A1* rs12722 and soft-tissue injury history in elite male rugby athletes. These findings support the growing evidence that soft-tissue injury could be influenced by an athlete's genetic predisposition.

Practical implications

- Elite rugby athletes carrying the TT genotype of *MMP3* rs679620 have an increased risk of ligament injury.
- Elite rugby athletes carrying the T-C inferred haplotype of *COL5A1* rs12722 and *COL5A1* rs3196378 have an increased risk of soft tissue injury.
- In the future, these findings may help to develop injury prevention protocols in those with high-risk genetic profiles.

Introduction

Compared to other team sports rugby has one of the highest injury incidence and severity rates, most of these are soft tissue injuries which damage the ligaments, tendons and muscles [1]. This is likely due to the intermittent, collision-based nature of the sport. Furthermore, the increasing size and strength of elite rugby athletes [2] is likely contributing to the high injury rates in the modern game, due to associated changes in momentum during collisions and changes of direction.

Analysis of match injury data across 16-seasons of the English Premiership identified sprains and ligament injuries had the highest incidence and highest burden of injury (22.6/1000 h and 30 days, respectively) [1]. The most recent available data from the men's rugby union World Cup (2019) found ligament injuries to be the most common, accounting for 21.7% of all injuries reported during matches [3]. Furthermore, knee ligament injuries were the most severe causing 935 days absence [3]. Tendon injuries, although not as common or severe as ligament injuries, are a considerably debilitating injury for rugby players. Specifically, Achilles tendon (AT) injuries have been in the top five highest burden training injuries across multiple English rugby Premiership seasons (2018/19-2020/21), costing 5.5 days absence per 1000 hours [4]. Achilles tendon injury appears to be particularly debilitating for rugby union forward athletes causing 726 days absence across two seasons [5]. Both ligament and tendon injuries are highly complex, multifactorial disorders that are determined by the interaction of several extrinsic and intrinsic factors [6, 7]. However, the growing body of evidence around the heritability of ligament and tendon injuries [8, 9] has prompted further research into possible genetic aetiology.

Anterior cruciate ligament (ACL) tears seem at least twice as likely in individuals with a family history of ACL tear compared to those with no family history [10]. Indeed, a twin study found that the genetic contribution to ACL rupture was ~69% [8]. Most research into the genetics of ligament injury has utilised genetic association studies (GAS) to investigate the influence of single nucleotide polymorphism (SNP) (a DNA sequence variation when a single nucleotide alters between individuals) variation individually and collectively. From these studies, variants in several genes have been associated with altered risk of ligament injury such as; collagen type I alpha I *(COL1A1* rs1800012*)* [11]*,* collagen type III alpha I *(COL3A1* rs1800255*)* [12]*,* collagen type V alpha I *(COL5A1* rs12722; rs3196378*)* [13]. The α1 chains of type I, III and V collagen are coded by these genes, respectively, and are thought to potentially influence collagen repair and regulation [14]. Genetic variants within matrix metalloproteinase-3 (*MMP3* rs591058, rs650108, rs679620*)* have also been associated with risk of ligament injury [15, 16]*.* The matrix metalloproteinase protein family aid in regulating the extracellular matrix (ECM), which affects the biomechanical properties of ligaments and tendon [17]. Additionally, nidogen proteins aid in the development of the ECM [18] and will likely affect its structure and function. The nidogen 1 gene (*NID1* rs4666048) was the strongest associated polymorphism identified in a fixed effect meta-analysis as part of a genome wide association (GWAS) study for ACL rupture [19]. Furthermore, vascular endothelial growth factor A *(VEGFA* rs699947*)* and kinase domain receptor *(KDR* rs1870377*)* genes have previously been associated with ACL and Achilles tendon injury [20, 21]. These genes code for the VEGFA protein, an endothelial cell mitogen that stimulates angiogenesis, which is essential during the remodelling and repair of injured soft tissue, and its receptor protein KDR [22].

Similarly, it has been suggested that there is a genetic component to the aetiology of tendon injury. Indeed, in a twin study of tennis elbow (epicondylitis) in women [9], heritability was estimated at ~40%. Numerous GAS report associations between Achilles tendinopathy and several genetic variants across a variety of genes such as *COL5A1 (*rs12722*)* [23], MicroRNA 608 (*MIR608* rs4919510), a small non-coding RNA that mediates gene silencing and translational repression [24], *MMP3* (rs591058, rs650108, rs679620*)* [25, 26] and tissue inhibitors of metalloproteinases-2 (*TIMP2* rs4789932*),* which aids in the regulation of the ECM with MMP proteins [26]. Furthermore, the Collagen beta(1-O) galactosyltransferase 1 (*COLGALT1* rs8090) gene, which may contribute to the pathogenesis of connective tissue disorders, potentially through aberrant post-translational modifications that impair the function of collagen-modifying enzymes [27], was the strongest associated SNP identified in a fixed-effect meta-analysis as part of a GWAS for Achilles tendon pathology [19]. In addition, Achilles tendon rupture has been associated with variants in the *MMP3* (rs679620) and *TIMP2* (rs4789932) genes [26].

Genetic variation may have a strong influence on tendon and ligament structure and function, which could alter an individual's risk of injury. Inter-individual variability of tendon and ligament properties is likely to cause micro and macro-trauma at differing strain levels among individuals, thus similar injury-inciting events amongst rugby players and/or playing positions may have vastly different outcomes. This could influence individual injury incidence and severity rates, thus affecting time lost from matches and training as well as potentially affecting early retirement. Ficek et al., [28] found potential evidence of this within male professional footballers, identifying that the *COL1A1* G-T haplotype (rs1107946 and rs1800012, respectively) was associated with reduced risk of ACL injury. Further evidence of this was found in male academy soccer players with a maturation status of pre- peak height velocity with variants of *COL5A1* rs12722 and *VEGFA* rs2010963 associated with a higher prevalence of ligament and tendon injuries [29]. However, no association between risk of ACL rupture and several collagen gene variants were found in a cohort of elite female athletes from high-risk team sports [30]. This could indicate potential sex and maturation differences in these associations, so further elucidation is necessary.

Past research has shown that for genetic variants previously associated with soft-tissue injury, elite male rugby athletes possess more 'favourable' variants both individually and collectively compared to a non-athlete population (*COLGALT1* rs8090, *COL3A1* rs1800255, *COL5A1* rs12722 and rs3196378 *MIR608* rs4919510, *MMP3* rs591058 and rs679620 and *NID1* rs4660148; three-SNP model of *COL5A1* rs12722, *COL5A1* rs3196378 and *MIR608* rs4919510; 13-SNP total genotype score model (TGS)) [31, 32] . However, there is no previous research directly investigating genetic associations with the history of soft tissue injury within an elite rugby athlete population. Thus, the aims of this study were to (1) investigate whether gene variants previously associated with soft-tissue injury risk are associated with history of soft-tissue injury in elite rugby athletes and (2) compare polygenic characteristics between athletes with a history of soft-tissue injury and those with no history of injury. It was hypothesised that the genotypes and alleles associated with soft-tissue injury risk would be overrepresented in elite rugby athletes with a history of injury compared to those with no previous history. Furthermore, it was hypothesised that athletes with no history of soft-tissue injury would have a higher TGS [33], i.e. a greater combination of favourable genotypes, than athletes with a history of injury.

Materials and Methods

Participants

This study was conducted in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines for a case-control observational study. Manchester Metropolitan University ethics committee granted approval of this study, which complies with the Declaration of Helsinki (Ethics code 24048). The participants were from the RugbyGene project (led by Manchester Metropolitan University with national and international collaboration which is ongoing [34]) and comprised of elite male rugby athletes of European ancestry (n = 184 (165 rugby union (RU), 19 rugby league (RL); mean (standard deviation) height 1.86 (± 0.07) m, mass 102 (± 12) kg, body mass index 29.5 (± 4.12) kg/m^2, age 26 (± 5) yr) including 37.3% British, 33.0% Italian, 20.0% Irish, and 9.7% of other nationalities, having given written informed consent. For TGS and SNP-SNP epistasis interaction analyses, 177 elite rugby athletes were utilised as 7 athletes did not have a full data set for all 13 polymorphisms investigated. All participants were considered elite rugby players, as they had competed regularly (at least 5 matches) since 1995 (when rugby became professional) in the highest professional league in the UK, Ireland, or South Africa for RU and the highest professional league in the UK for RL [31]. Furthermore, 55% of the RU athletes had competed at international level for a “high performance union” (Regulation 16, http://www.worldrugby.org), and 50% of RL athletes had competed at international level.

Procedures

The procedures are consistent with those reported previously [31]. Blood, buccal swab or saliva samples were attained via the following procedures. Blood (36.7% of all samples) was drawn from a superficial forearm vein into an EDTA tube, saliva (63.2%) samples were collected into Oragene DNA OG-500 collection tubes (DNA Genotek, Ottawa, Ontario, Canada) and sterile buccal swabs (Omni swab; Whatman, Springfield Mill, UK) were rubbed against the buccal mucosa of the cheek for ~30 s.

DNA isolation was performed with the QIAamp DNA Blood Mini kit and standard spin column protocol (Qiagen, West Sussex, UK). Briefly, 200 µL of whole blood/saliva, or one buccal swab, was lysed and incubated, the DNA washed, and the eluate containing isolated DNA stored at 4°C.

Genotyping for 13 polymorphisms (see list below) was performed using two protocols. Protocol one: Approximately 20% of the DNA samples were genotyped via real-time PCR using a StepOnePlus (Applied Biosystems, Paisley, UK) as previously described [31], with adjustment of thermocycling conditions depending on reagents utilised. Protocol two: Approximately 80% of the DNA samples were genotyped by combining 2 µL GTXpress Master Mix (Applied Biosystems), 0.2 µL Fast GT Sample Loading Reagent (Fluidigm, Cambridge, UK), 0.2 µL H_2O and 1.6 µL of purified DNA, for samples derived from blood and saliva. Furthermore, 1.78 µL assay (Applied Biosystems), 1.78 µL Assay Loading Reagent (Fluidigm, Cambridge, UK) and 0.18 µL ROX reference dye (Invitrogen, Paisley, UK) were combined per assay inlet. An integrated fluid circuit controller RX (Fluidigm) was used to mix samples and assays using a Load Mix (166x) script. PCR was performed using a real-time FC1 Cycler (Fluidigm, Cambridge, UK) GT 192X24 Fast v1 protocol. The 192X24 microchip plate was then placed into the EP1 Reader (Fluidigm) for end-point analysis. Duplicates of all samples were in 100% agreement for both protocols. For both protocols, the appropriate TaqMan assays were utilised (Applied Biosystems, Paisley, UK) and assay context sequences for each polymorphism are shown in Table 1.

Soft-tissue injury history

Soft tissue injury history in elite rugby athletes was collected utilising a self-reported injury-history questionnaire (Supplementary document). This was developed by the current investigators using prior literature [35] and in consultation with medical practitioners and experienced researchers. Athletes were asked to provide details of their geographic ancestry, playing position, playing history, highest level of play, tendon and ligament injury incidence independent of mechanism, and whether each injury was medically diagnosed. An investigator assisted participants with completion of the questionnaire to maximise accuracy. To reduce athlete-response bias, medical staff and coaches were not present during questionnaire completion.

Calculation of TGS

To quantify the combined influence of the candidate polymorphisms (Table 1) an additive TGS algorithm was utilised [33], based on the assumption of codominance effects of the alleles. The homozygote genotypes with the lower soft tissue injury risk, according to prior literature, were allocated a ‘genotype score’ of 2, heterozygote genotypes were scored 1 and the higher soft tissue injury risk homozygote genotypes were scored 0.

Table 1. Genotype score of each polymorphism.

Gene name	Gene abbreviation	rs number	Assay context sequence	Polymorphism	Genotype score (GS)
Collagen beta(1-O) galactosyltransferase 1	*COLGALT1*	8090	*COLGALT1* rs8090: CTCCC**[A/G]**GTCCC	A/G	AA = 2, GA = 1, GG = 0
Collagen type I alpha I	*COL1A1*	1800012	*COL1A1* rs1800012: CGCCC**[A/C]**CATTC	A/C	AA = 2, AC = 1, CC = 0
Collagen type III alpha I	*COL3A1*	1800255	*COL3A1* rs1800255: GTGGA**[A/G]**CTGGT	A/G	GG = 2, GA = 1, AA = 0
Collagen type V alpha I	*COL5A1*	12722	*COL5A1* rs12722*:* ACCCA[**C/T**]GCGCC	C/T	CC = 2, CT = 1, TT = 0
		3196378	*COL5A1* rs3196378: ACCCC[**A/C**]GCCCT	C/A	CC = 2, CA = 1, AA = 0
Kinase Domain Receptor	*KDR*	1870377	*KDR* rs1870377: ACAGC**[A/T]**TGGCT	A/T	TT = 2, TA = 1, AA = 0
MicroRNA 608	*MIR608*	4919510	*MIR608* rs4919510: CAGCT**[C/G]**CGTTT	G/C	GG = 2, GC = 1, CC = 0
Matrix metalloproteinase-3	*MMP3*	591058	*MMP3* rs679620: TTTTT**[C/T]**GAGGT	T/C	TT = 2, TC = 1, CC = 0
		650108	*MMP3* rs591058: GAAAT**[C/T]**GAGAA	G/A	GG = 2, GA = 1, AA

					= 0
		679620	*MMP3* rs650108: TTAGA**[A/G]**GTAGC	T/<u>C</u>	TT = 2, TC = 1, CC = 0
Nidogen 1	*NID1*	4660148	*NID1* rs4660148: TTTTC**[G/T]**TTGGG	T/<u>G</u>	TT = 2, TG = 1, GG = 0
Tissue inhibitors of metalloproteinases-2	*TIMP2*	4789932	*TIMP2* rs4789932: TATCT**[A/G]**CTGTA	G/<u>A</u>	GG = 2, GA = 1, AA = 0
Vascular endothelial growth factor A	*VEGFA*	699947	*VEGFA* rs699947: TGGCA**[A/C]**GATCT	C/<u>A</u>	CC = 2, CA = 1, AA = 0

Tendon and ligament injury-associated alleles in previous literature are underlined [32]. Assay context sequences with alleles corresponding to VIC/FAM, respectively, highlighted in bold.

TGS model

TGS = (100/26) * $COLGALT1_{rs8090}$ + $COL1A1_{rs1800012}$ + $COL3A1_{rs1800255}$ + $COL5A1_{rs12722}$ + $COL5A1_{rs3196378}$ + $KDR_{rs1870377}$ + $MIR608_{rs4919510}$ + $MMP3_{rs679620}$ + $MMP3_{rs591058}$ + $MMP3_{rs650108}$ + $NID1_{rs4660148}$ + $TIMP2_{rs4789932}$ + $VEGFA_{rs699947}$

A TGS of 100 represents the 'perfect' polygenic profile for soft tissue injury risk and 0 represents the 'worst' possible outcome for the variants examined in this study [32].

Data Analysis

Pearson's Chi-square (χ^2) tests were used to measure Hardy-Weinberg equilibrium and to compare genotype (using three analysis models: additive, recessive, and dominant) and allele frequencies between injured and non-injured groups. The group comparisons were as follows: total injured athletes (TIA) vs total non-injured athletes (TNIA), then sub-group analysis of athletes who had a history of; tendon rupture (TR) vs no tendon injury (NIT); tendinopathy (TD) vs NIT; ligament rupture (LR) vs no ligament injury (NIL); ligament sprain (LS) vs NIL. With 80% statistical power, analyses of total injured athletes compared with total non-injured athletes were able to detect a small-moderate effect size (w) of 0.25 and analyses between smallest injury subgroups (LR v NIL) were able to detect a moderate effect size (w) of 0.32. For each polymorphism, 20 tests were subjected to Benjamini-Hochberg corrections to control for false discovery rate and probability values are reported. Where appropriate, odds ratios (OR) were calculated to estimate effect size. Total genotype scores for all groups were not normally distributed, therefore Mann-Whitney U tests were utilised to compare TGS between injured athlete groups and non-injured. Means and extent of kurtosis were calculated to describe the distribution of TGS within groups. χ^2 tests were used to compare the frequency of injured athletes and non-injured athletes in the top and bottom thirds of TGS scores. Bonferroni adjustment was utilised where appropriate to control for false discovery. We also evaluated the ability of the TGS to correctly distinguish injured from non-injured athletes across all groups by receiver operating characteristic (ROC) curves, calculating the area under the curve (AUC) and 95% confidence intervals (95% CI). Multifactor dimensionality reduction (MDR) software (https://sourceforge.net/projects/mdr/) was used to identify any possible SNP-SNP epistasis interactions. Haplotypes were inferred using SNPStats. SPSS for Windows version 29 (SPSS, Chicago, IL) software was used for analysis. P values < 0.05 were considered statistically significant.

Results

Injury analysis

Twenty-nine athletes reported no relevant soft tissue injury, while 155 athletes had suffered from a relevant form of soft tissue injury. A breakdown of the sub-group injury totals for each genotype can be seen in Supplementary Table 1. Fifty-four athletes had suffered from both a ligament and tendon injury, 13 athletes had both a ligament and tendon rupture; and four athletes had injuries across all four categories (tendon and ligament rupture, tendinopathy and ligament sprain). Six athletes had more than one area of tendinopathy, 18 athletes suffered more than one ligament rupture, and 21 athletes had more than one area of ligament sprain. A breakdown of injury location for each sub-group can be found in Supplementary Table 2.

Genotype and allele frequencies

Genotype frequencies were in Hardy-Weinberg equilibrium for 96 of the 104 possible values across all polymorphisms in the injured and non-injured groups apart from *COLGALT1* rs8090 (LS group), *COL5A1* rs12722 (NIT and TR groups), *COL5A1* rs3196378 (NIT, LS and TIA groups), *MIR608* rs4919510 (TNIA group) and *TIMP2* rs4789932 (TR group) (supplementary table 1).

For *MMP3* rs679620, the TT genotype and T allele were underrepresented, whilst the proportion of C allele carriers were overrepresented in TD (13.7%, 38.3% and 86.3% respectively) compared to NIT (28.1%, 50.0% and 71.9%; $P < 0.04$, Fig 1). However, the TT genotype and T allele were underrepresented, whilst the proportion of C allele carriers were more common in NIL (10%, 36.3% and 90%, respectively) compared to LR (32.4%, 52.1% and 67.6%, $P = 0.03$, Fig 1). Similarly, the TT genotype was underrepresented, whilst the proportion of C allele carriers was more common in NIL (10% and 90%) compared to LS (25.2% and 74.5%, $P = 0.04$). The LR and LS groups had over three times the odds of carrying the TT genotype compared to NIL (LR odds ratio (OR): 4.3, 95% confidence intervals (CI:1.4-13.6); LS OR= 3.0 (1.0-9.2). For allele/genotype frequency data for all SNPs please refer to Supplementary Table 1.

********************Figure 1 here (Please see end of document)***********************************

Figure 1. Allele frequency of *MMP3* rs679620 for all injured and non-injured groups. Asterisks (*) indicate a difference in allele frequency between the injured group and their respective non-injured group ($P < 0.05$).

For *COL5A1* rs12722, the TC genotype and proportion of C-allele carriers were more common in TR (73.1% and 96.2%, respectively) compared to NIT (40.5% and 75.0% $P < 0.02$). The TR group had eight times the odds of carrying the C allele compared to NIT (OR = 8.3 (1.1 - 64.3)). There were no differences in genotype or allele frequencies for *COL5A1* rs12722 between any other groups.

There were no differences in genotype or allele frequencies between any groups for the other 11 genetic variants.

Haplotype and SNP epistasis analysis

The T-C inferred haplotype frequency of *COL5A1* rs12722 and *COL5A1* rs3196378, respectively, was lower in NIT, NIL and TNIA (11.7%, 6.7% and 2.2%, respectively) compared to TR, LS and TIA (37.7%, 17.8% and 16.2%, *P* < 0.02, Figure 2). Additionally, the C-A inferred haplotype frequency of *COL5A1* rs12722 and *COL5A1* rs3196378 was higher in TR and LS than their respective non-injured groups (*P* < 0.05). Furthermore, the C-C inferred haplotype was higher in NIT than TR (*P* < 0.01). There were no inferred haplotype frequency differences for *MMP3* rs591058, rs650108 and rs679620 between any injured and non-injured groups. Multifactor dimensionality reduction analysis could not identify a model to discriminate between any injury group or sub-group and the respective non-injured group (P ≥ 0.35).

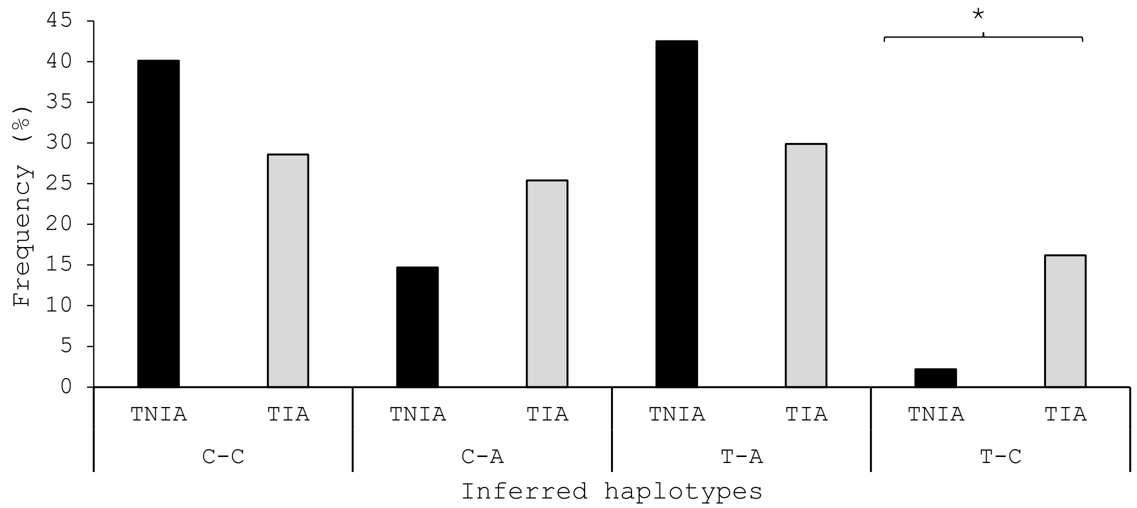


Figure 2. Inferred haplotypes derived from *COL5A1* rs12722 and rs3196378. TNIA = total non-injured athletes, TIA = total injured athletes. *Different from TNIA (*P* < 0.001).

Total genotype score

There was a significant difference between the TGS of LS compared to NIL (*P* = 0.02), with the frequency distribution for LS and NIL shown in Figure 3. There were no differences in TGS between any other injured and non-injured groups (TR vs NIT, TD vs NIT, LR vs NIL and TIA vs TNIA; *P* > 0.05). Mean (standard deviation) TGS and kurtosis statistics are reported in Table 2. When the numbers of injured and non-injured athletes in the upper and lower thirds of the TGS were compared, there were no differences between any injured and non-injured groups (*P* > 0.05). Finally, the TGS was able to distinguish between LS and NIL (AUC = 0.62; 95% CI = 0.52 – 0.72, *P* < 0.02, Figure 4), but could not distinguish between any other groups.

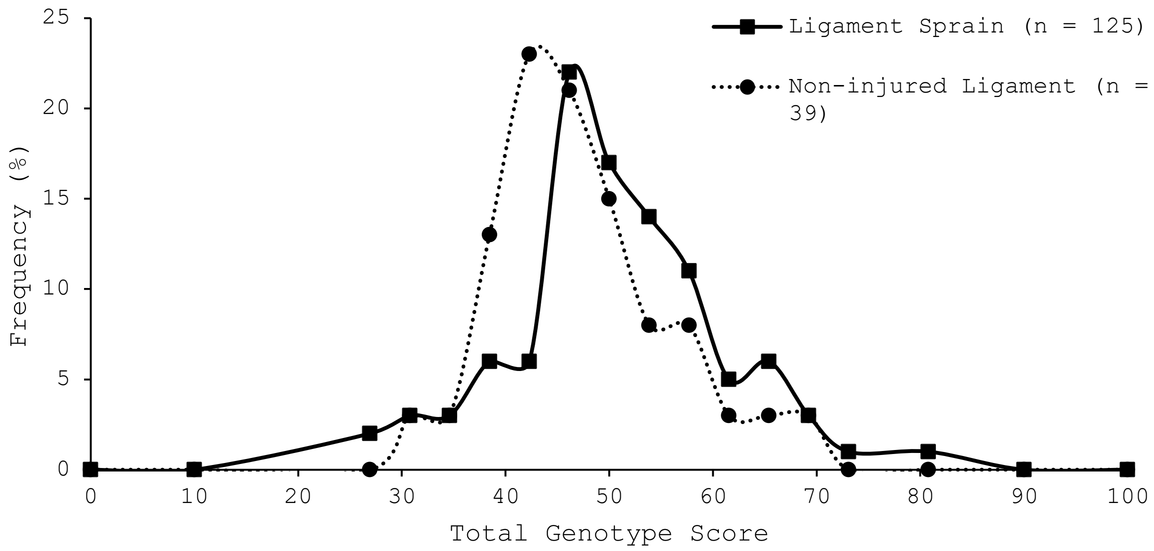


Figure 3. Frequency distribution of TGS. There was a significant difference in TGS between LS and NIL (*P* < 0.02).

Table 2. TGS mean and kurtosis.

Group	Mean (SD) TGS	Mean (SE) kurtosis statistic
Total non-injured athletes	47.4 (8.0)	1.6 (0.9)

Total injured athletes	49.8 (10.2)	-0.0 (0.4)
Non-injured tendon	49.7 (10.3)	0.1 (0.5)
Tendon rupture	49.3 (9.6)	0.3 (0.9)
Tendinopathy	48.2 (8.5)	-0.2 (0.7)
Non-injured ligament	47.1 (8.2)	0.6 (0.7)
Ligament rupture	50.3 (11.3)	-0.7 (0.6)
Ligament sprain	50.3 (9.8)	0.4 (0.4)

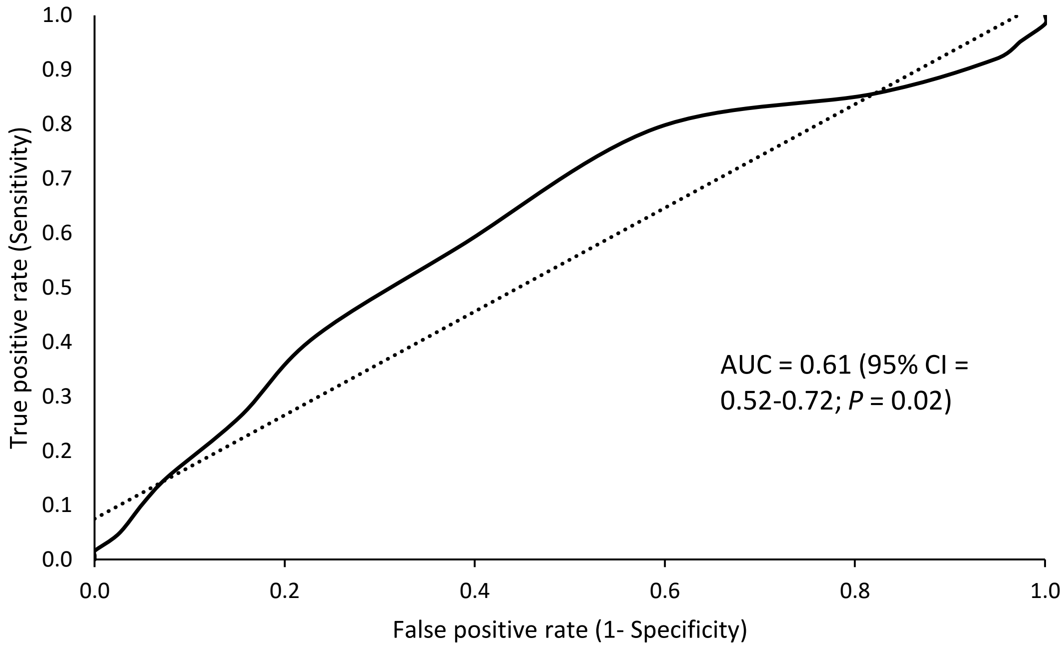


Figure 4. Receiver operating characteristic curve (ROC) summarizing the ability of TGS to classify ligament sprain from non-injured ligament rugby athletes. AUC indicates the area under the curve (95% confidence intervals).

Discussion

This study is the first to identify associations between *MMP3* rs679620 and *COL5A1* rs12722 and soft-tissue injury history in elite male rugby athletes, thus indicating a likely inherited advantage from carrying protective genetic variants involved in collagen and ECM structure and function. Furthermore, the T-C inferred haplotype frequency of *COL5A1* rs12722 and *COL5A1* rs3196378 was higher in the tendon rupture, ligament sprain and total injured athlete groups compared to their respective non-injured groups. As such, the TGS was able to distinguish between the non-injured ligament group and the ligament sprain group. These combined findings suggest a likely polygenic influence on soft tissue injury risk in rugby. As hypothesized, elite male rugby athletes with a history of soft-tissue injury mostly carried more of the apparent injury-risk genotype/alleles than non-injured athletes, although this was not consistent for all polymorphisms.

The TT genotype and T allele of *MMP3* rs679620 were overrepresented in the non-injured tendon athlete group compared to the tendinopathy group which aligns with previous findings [25, 36]. Conversely, the TT genotype and T allele were underrepresented in the non-injured ligament group compared to the ligament rupture group. The TT genotype was also underrepresented in the ligament sprain group. Indeed, the ligament rupture and ligament sprain groups had over three times the odds of carrying the TT genotype compared to non-injured athletes. Similar findings have been seen within sporting populations for non-contact ACL ruptures [16] and knee injuries [29]. The mRNA expression profile of MMP3 appears to contrast between ACL injury and Achilles tendinopathy [37, 38]. Therefore, the opposing findings within this study could be due to underlying differences in the pathophysiology of tendon and ligament injuries. Alterations in ECM homeostasis are thought to play a role in soft tissue injury risk. Specifically, MMP3, which regulates ECM homeostasis via proteolytic activity, is considered to be an essential regulator of matrix degradation and remodelling [38]. The present findings suggests that the rs679620 missense polymorphism may, along with other variants, play a role in tendon and ligament injury risk, possibly via the regulation of ECM homeostasis.

Previously, the C allele of *COL5A1* rs12722 was identified as protective from ligament injury in females [13] and Achilles tendinopathy in males and females [23]. However, we found the proportion of C allele carriers to be higher in the tendon rupture group, with the TT genotype overrepresented in the non-injured tendon group. This aligns with Hall et al., [29] who found male academy soccer players carrying the CC genotype had a higher prevalence of musculoskeletal and ligament injury than T allele carriers. Furthermore, when *COL5A1* rs12722 and rs3196378 were combined, the T-C inferred haplotype was overrepresented in the tendon rupture, ligament sprain and total injured athlete groups compared to their respective non-injured groups. The *COL5A1* 3′ untranslated region, where rs12722 and rs3196378 are located, has been shown to affect mRNA stability [39], which may lead to altered COL5A1 mRNA secondary structure – possibly influencing type V collagen production. Although this proposed mechanism has still to be fully elucidated, our data, which align in part with previous findings, suggests variants within *COL5A1* may influence the soft-tissue injury incidence of elite rugby athletes.

COLGALT1 rs8090 and *NID1* RS4660148 were the strongest SNPs associated with Achilles tendon pathology and ACL rupture in a fixed-effect meta-analysis within a GWAS [19], though they did not reach genome-wide significance ($P > 6 \times 10^{-5}$; $P > 5 \times 10^{-6}$, respectively). *COL3A*1 rs1800255 and *MIR608* rs4919510, have also previously been linked with ligament and tendon injuries [12, 24]. However, no associations were observed between these variants and tendon or ligament injury within this study. *COL3A1* encodes type III collagen, which plays a key role in type I collagen fibrillogenesis and is influenced by *MIR608* [24]. Elevated collagen III levels during fibrillogenesis reduce collagen I content, resulting in smaller, disorganised fibrils and diminished tensile strength [17]. Additionally, *COLGALT1* has been shown to affect collagen I function, with mouse models linking its mutation to musculoskeletal abnormalities [40]. Thus, it was considered that these variants would influence soft tissue injury risk. Additionally, prior research has found all four variants to be associated with elite status within a male rugby population [31], suggesting that they may influence performance-related traits such as muscular strength, particularly relevant for rugby athletes. However, based on our present findings they appear to have limited effect on soft tissue injury.

COL1A1 rs1800012 and *KDR* rs1870377 have previously been associated with ligament injury [11, 20], while TIMP2 rs4789932 and *VEGFA* rs699947 have been linked to Achilles tendon pathology and ligament injury [20, 21, 26]. Based on this prior evidence, all four variants were investigated for potential associations with soft-tissue injury. However, no significant associations were identified. These results may reflect inconsistencies in the existing literature, as only findings related to *COL1A1* rs1800012 have been consistently replicated. Therefore, further research is required to clarify the functional relevance of these polymorphisms.

When polygenic analysis was performed, there was a higher mean TGS in the ligament sprain group compared to the non-injured ligament group. Additionally, ROC analysis could significantly discriminate the non-injured ligament group from the ligament sprain group. This suggests that rugby athletes with no history of ligament sprain do not appear to carry 'preferable' soft-tissue injury associated polygenic profiles. A possible reason for this could be due to the equivocal evidence base of the prior literature regarding the 'risk' allele of each SNP, with only four (*COL1A1* rs1800012, *COL3A1* rs1800255, *COL5A1* rs12722 and *MMP3* rs679620) of the thirteen polymorphisms studied having had their 'risk' allele replicated in a comparable cohort. There were no other differences in TGS between any other injured group and their respective non-injured groups. Furthermore, when the top and bottom thirds of the TGS were compared, there were no differences found between any groups. The results could in part be down to the relatively small sample sizes of the individual TGS groups, reducing their statistical power. However, it is most likely that at this present time the evidence base for allocating 'risk' alleles for TGS analysis on soft-tissue injury is limited due to the lack of consistent findings across polymorphisms.

The present study is not without limitations. The retrospective nature of injury data collection is susceptible to recall bias due to reliance on memory. However, retrospective studies are time and resource efficient and promote greater participation from elite sporting populations, thus it was deemed appropriate for this cohort. Furthermore, this study did not account for mechanism of injury such as via contact or non-contact which could influence the results and should be considered when interpreting the present data. Finally, the aetiology of soft-tissue injury is extremely complex including multiple intrinsic and extrinsic factors. Therefore, the data provided in this study should be regarded within that context, particularly due to the relatively small number of genetic variants studied.

The practical application of genetic data for injury risk management remains limited at this time. While findings from this and other studies strongly suggest a heritable component to soft tissue injury susceptibility, the specific genetic variants, both individually and in combination, have yet to be fully identified and understood. Moreover, mechanistic evidence detailing how these variants influence injury risk is currently sparse. Nonetheless, as additional variants are discovered and functional pathways are better characterised, such data may eventually contribute to more personalised approaches to injury prevention and management among athletes.

Conclusion

We have presented the first genetic associations between *MMP3* rs679620 and *COL5A1* rs12722 and soft-tissue injury history in elite male rugby athletes. Furthermore, the T-C inferred haplotype of *COL5A1* rs12722 and *COL5A1* rs3196378, respectively, was also associated with soft-tissue injury. The functionality of these genetic variants needs further elucidation to identify how soft-tissue injury risk may be affected. Nevertheless, the current data suggest soft-tissue injury could be influenced by an athlete's genetic predisposition. This study provides further insight into the detailed aetiology of soft tissue injuries within elite male rugby and may, in future, be worthy of consideration for managing the interindividual variability of injury risk in rugby.

Acknowledgments

The authors wish to thank all athletes, their respective scientific support staff for their time and willingness to participate in the research.

Declaration of interest statement:

The authors declare that they have no competing interests.

References

1. West SW, Starling L, Kemp S, Williams S, Cross M, Taylor A, et al. Trends in match injury risk in professional male rugby union: a 16-season review of 10 851 match injuries in the English Premiership (2002–2019): the Professional Rugby Injury Surveillance Project. British journal of sports medicine. 2021;55(12):676-82.
2. Cunniffe B, Proctor W, Baker JS, Davies B. An evaluation of the physiological demands of elite rugby union using Global Positioning System tracking software. J Strength Cond Res. 2009 Jul;23(4):1195-203.
3. Fuller CW, Taylor A, Douglas M, Raftery M. Rugby World Cup 2019 injury surveillance study. South African journal of sports medicine. 2020;32(1):v32i1a8062.
4. Kemp S, Hudson, S., Anstiss, T., Brooks, J., Bryan, R., Cross, M., Jones, B., Henderson, L., Locke, D., Lee, M., Rossiter, M., West, S., McKay, C., Williams, S., Stokes, K. England Professional Rugby Injury Surveillance Project Season Report 2022-23. 2024:1-37.
5. Brooks JH, Fuller CW, Kemp SP, Reddin DB. Epidemiology of injuries in English professional rugby union: part 1 match injuries. British journal of sports medicine. 2005 Oct;39(10):757-66.
6. Nourissat G, Berenbaum F, Duprez D. Tendon injury: from biology to tendon repair. Nature reviews Rheumatology. 2015 Apr;11(4):223-33.
7. 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. 2012 Mar;4(2):155-61.

8. Magnusson K, Turkiewicz A, Hughes V, Frobell R, Englund M. High genetic contribution to anterior cruciate ligament rupture: Heritability ~69. British journal of sports medicine. 2020 Dec 7.
9. Hakim AJ, Cherkas LF, Spector TD, MacGregor AJ. Genetic associations between frozen shoulder and tennis elbow: a female twin study. Rheumatology (Oxford, England). 2003 Jun;42(6):739-42.
10. 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 Jan;33(1):23-8.
11. Khoschnau S, Melhus H, Jacobson A, Rahme H, Bengtsson H, Ribom E, et al. Type I collagen alpha1 Sp1 polymorphism and the risk of cruciate ligament ruptures or shoulder dislocations. The American journal of sports medicine. 2008 Dec;36(12):2432-6.
12. Stępień-Słodkowska M, Ficek K, Maciejewska-Karłowska A, Sawczuk M, Ziętek P, Król P, et al. Overrepresentation of the COL3A1 AA genotype in Polish skiers with anterior cruciate ligament injury. Biology of sport. 2015 Jun;32(2):143-7.
13. 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. 2009 Nov;37(11):2234-40.
14. Brazier J, Antrobus M, Stebbings GK, Day SH, Heffernan SM, Cross MJ, et al. Tendon and Ligament Injuries in Elite Rugby: The Potential Genetic Influence. Sports (Basel, Switzerland). 2019 Jun 4;7(6).
15. Lulińska-Kuklik E, Rahim M, Moska W, Maculewicz E, Kaczmarczyk M, Maciejewska-Skrendo A, et al. Are MMP3, MMP8 and TIMP2 gene variants associated with anterior cruciate ligament rupture susceptibility? Journal of Science and Medicine in Sport. 2019 2019/07/01/;22(7):753-7.
16. Simunic-Briski N, Vrgoc G, Knjaz D, Jankovic S, Dembic Z, Lauc G. MMP3 single-nucleotide polymorphisms are associated with noncontact ACL injuries in competing high-level athletes. Journal of Orthopaedic Research. 2024;42(1):109-14.
17. Riley G. The pathogenesis of tendinopathy. A molecular perspective. Rheumatology (Oxford, England). 2004 Feb;43(2):131-42.
18. Ho MS, Böse K, Mokkapati S, Nischt R, Smyth N. Nidogens-Extracellular matrix linker molecules. Microscopy research and technique. 2008 May;71(5):387-95.
19. Kim SK, Roos TR, Roos AK, Kleimeyer JP, Ahmed MA, Goodlin GT, et al. Genome-wide association screens for Achilles tendon and ACL tears and tendinopathy. PloS one. 2017;12(3):e0170422.
20. Rahim M, Gibbon A, Hobbs H, van der Merwe W, Posthumus M, Collins M, et al. The association of genes involved in the angiogenesis-associated signaling pathway with risk of anterior cruciate ligament rupture. Journal of orthopaedic research : official publication of the Orthopaedic Research Society. 2014 Dec;32(12):1612-8.
21. Rahim M, El Khoury LY, Raleigh SM, Ribbans WJ, Posthumus M, Collins M, et al. Human Genetic Variation, Sport and Exercise Medicine, and Achilles Tendinopathy: Role for Angiogenesis-Associated Genes. Omics : a journal of integrative biology. 2016 Sep;20(9):520-7.
22. Petersen W, Pufe T, Zantop T, Tillmann B, Tsokos M, Mentlein R. Expression of VEGFR-1 and VEGFR-2 in degenerative Achilles tendons. Clinical orthopaedics and related research. 2004 Mar(420):286-91.
23. Mokone GG, Schwellnus MP, Noakes TD, Collins M. The COL5A1 gene and Achilles tendon pathology. Scand J Med Sci Sports. 2006 Feb;16(1):19-26.
24. Abrahams Y, Laguette MJ, Prince S, Collins M. Polymorphisms within the COL5A1 3'-UTR that alters mRNA structure and the MIR608 gene are associated with Achilles tendinopathy. Annals of human genetics. 2013 May;77(3):204-14.
25. Raleigh SM, van der Merwe L, Ribbans WJ, Smith RK, Schwellnus MP, Collins M. Variants within the MMP3 gene are associated with Achilles tendinopathy: possible interaction with the COL5A1 gene. British journal of sports medicine. 2009 Jul;43(7):514-20.

26. El Khoury L, Ribbans WJ, Raleigh SM. MMP3 and TIMP2 gene variants as predisposing factors for Achilles tendon pathologies: Attempted replication study in a British case-control cohort. Meta gene. 2016 Sep;9:52-5.
27. Schegg B, Hülsmeier AJ, Rutschmann C, Maag C, Hennet T. Core glycosylation of collagen is initiated by two beta(1-O)galactosyltransferases. Molecular and cellular biology. 2009 Feb;29(4):943-52.
28. Ficek K, Cieszczyk P, Kaczmarczyk M, Maciejewska-Karłowska A, Sawczuk M, Cholewinski J, et al. Gene variants within the COL1A1 gene are associated with reduced anterior cruciate ligament injury in professional soccer players. J Sci Med Sport. 2013 Sep;16(5):396-400.
29. Hall ECR, Baumert P, Larruskain J, Gil SM, Lekue JA, Rienzi E, et al. The genetic association with injury risk in male academy soccer players depends on maturity status. Scandinavian Journal of Medicine & Science in Sports. 2022;32(2):338-50.
30. Sivertsen EA, Haug KBF, Kristianslund EK, Trøseid AS, Parkkari J, Lehtimäki T, et al. No Association Between Risk of Anterior Cruciate Ligament Rupture and Selected Candidate Collagen Gene Variants in Female Elite Athletes From High-Risk Team Sports. The American journal of sports medicine. 2019 Jan;47(1):52-8.
31. Brazier J, Antrobus MR, Herbert AJ, Callus PC, Stebbings GK, Day SH, et al. Gene variants previously associated with reduced soft tissue injury risk: Part 1 – independent associations with elite status in rugby. European Journal of Sport Science. 2023;23(5):726-35.
32. Brazier J, Antrobus MR, Herbert AJ, Callus PC, Khanal P, Stebbings GK, et al. Gene variants previously associated with reduced soft-tissue injury risk: Part 2 - Polygenic associations with elite status in Rugby. Eur J Sport Sci. 2023 Aug;23(8):1779-88.
33. Williams AG, Folland JP. Similarity of polygenic profiles limits the potential for elite human physical performance. The Journal of physiology. 2008 Jan 1;586(1):113-21.
34. Heffernan SM, Kilduff LP, Day SH, Pitsiladis YP, Williams AG. Genomics in rugby union: A review and future prospects. European Journal of Sport Science. 2015;15(6):460-8.
35. Varley I, Hughes DC, Greeves JP, Stellingwerff T, Ranson C, Fraser WD, et al. RANK/RANKL/OPG pathway: genetic associations with stress fracture period prevalence in elite athletes. Bone. 2015 Feb;71:131-6.
36. Nie G, Wen X, Liang X, Zhao H, Li Y, Lu J. Additional evidence supports association of common genetic variants in MMP3 and TIMP2 with increased risk of chronic Achilles tendinopathy susceptibility. Journal of Science and Medicine in Sport. 2019 2019/10/01/;22(10):1074-8.
37. Beye JA, Hart DA, Bray RC, McDougall JJ, Salo PT. Injury-induced changes in mRNA levels differ widely between anterior cruciate ligament and medial collateral ligament. The American journal of sports medicine. 2008 Jul;36(7):1337-46.
38. Jones GC, Corps AN, Pennington CJ, Clark IM, Edwards DR, Bradley MM, et al. Expression profiling of metalloproteinases and tissue inhibitors of metalloproteinases in normal and degenerate human achilles tendon. Arthritis and rheumatism. 2006 Mar;54(3):832-42.
39. Laguette MJ, Abrahams Y, Prince S, Collins M. Sequence variants within the 3'-UTR of the COL5A1 gene alters mRNA stability: implications for musculoskeletal soft tissue injuries. Matrix biology : journal of the International Society for Matrix Biology. 2011 Jun;30(5-6):338-45.
40. Geister KA, Lopez-Jimenez AJ, Houghtaling S, Ho TH, Vanacore R, Beier DR. Loss of function of Colgalt1 disrupts collagen post-translational modification and causes musculoskeletal defects. Disease models & mechanisms. 2019 Jun 17;12(6).

Acknowledgments/Funding Information/Ethical Compliance

The authors wish to thank all athletes, their respective scientific support staff for their time and willingness to participate in the research.

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Manchester Metropolitan University ethics committee granted approval of this study, which complies with the Declaration of Helsinki (Ethics code 24048).

CRediT Author Statement

Jon Brazier – Conceptualization, Methodology, Investigation, Data Curation, Formal Analysis, Writing - Original Draft Preparation, Writing – Review and Editing.

Mark R. Antrobus - Conceptualization, Methodology, Investigation, Data Curation, Formal Analysis, Writing – Review and Editing.

Peter C. Callus – Investigation, Formal Analysis, Writing – Review and Editing.

Adam J. Herbert - Investigation, Formal Analysis, Writing – Review and Editing.

Georgina K. Stebbings - Conceptualization, Methodology, Investigation, Writing – Review and Editing

Daniel Martin - Investigation, Formal Analysis, Writing – Review and Editing.

Stephen H. Day - Conceptualization, Methodology, Writing – Review and Editing

Liam P. Kilduff - Methodology, Resources, Writing – Review and Editing.

Mark A. Bennett - Methodology, Resources, Writing – Review and Editing.

Robert M. Erskine - Conceptualization, Methodology, Investigation, Writing – Review and Editing.

Stuart. M. Raleigh - Resources, Writing – Review and Editing.

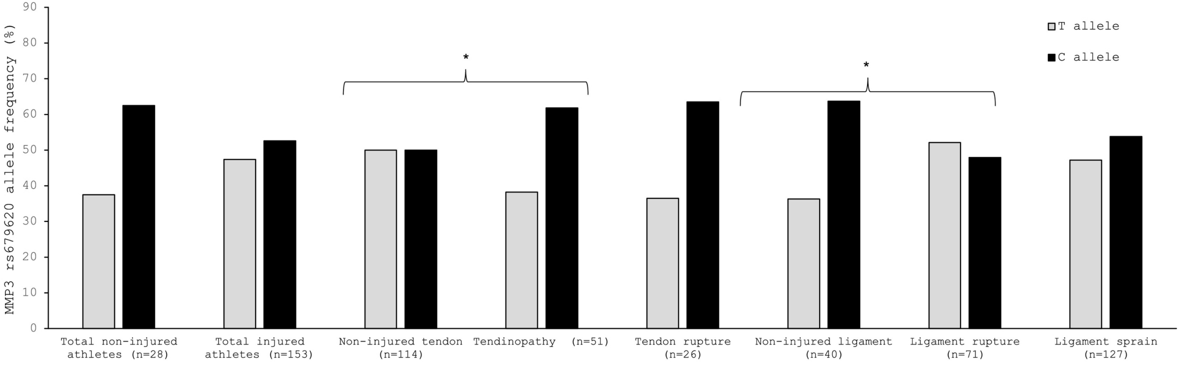
Tom Cullen - Resources, Writing – Review and Editing.

Malcolm Collins - Conceptualization, Methodology, Writing – Review and Editing

Yannis. P. Pitsiladis - Conceptualization, Methodology, Writing – Review and Editi

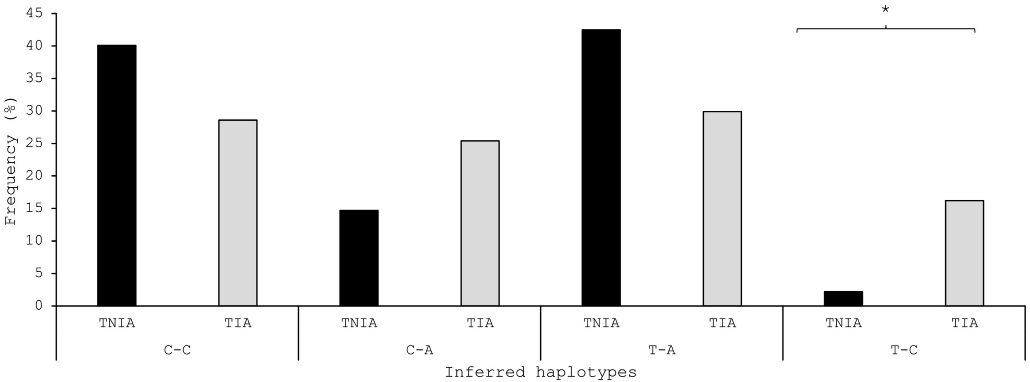
Shane M. Heffernan - Conceptualization, Methodology, Investigation, Writing – Review and Editing.

Alun G. Williams - Conceptualization, Methodology, Investigation, Data Curation, Formal Analysis, Writing – Review and Editing.



MMP3 rs679620 allele frequency (%)
90
80
70
60
50
40
30
20
10
0
T allele
C allele
*
*
Total non-injured athletes (n=28)
Total injured athletes (n=153)
Non-injured tendon (n=114)
Tendinopathy (n=51)
Tendon rupture (n=26)
Non-injured ligament (n=40)
Ligament rupture (n=71)
Ligament sprain (n=127)

Figure 1



45
40
35
30
25
20
15
10
5
0
Frequency (%)
*
TNIA
TIA
TNIA
TIA
TNIA
TIA
TNIA
TIA
C-C
C-A
T-A
T-C
Inferred haplotypes

Figure 2

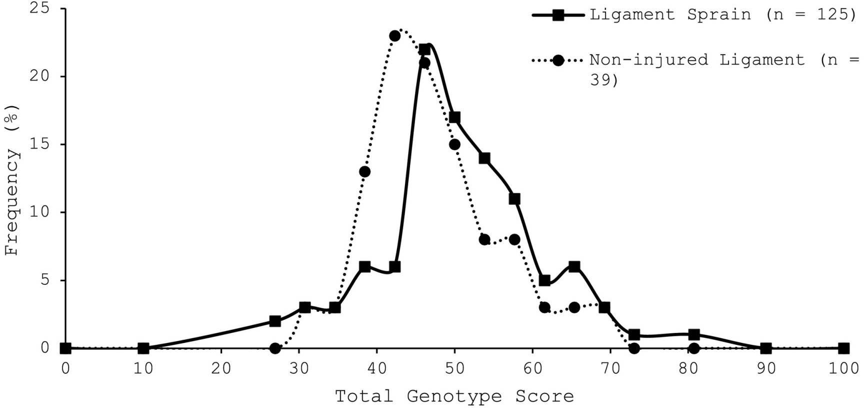

Ligament Sprain (n = 125)
Non-injured Ligament (n = 39)
Frequency (%)
25
20
15
10
5
0
0
10
20
30
40
50
60
70
80
90
100
Total Genotype Score

Figure 3

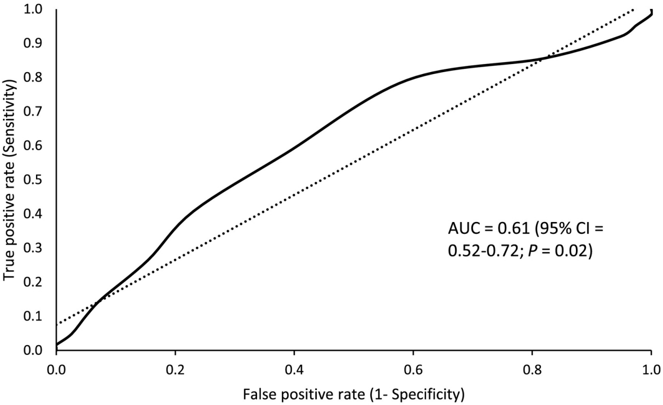

1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
True positive rate (Sensitivity)
0.0
0.2
0.4
0.6
0.8
1.0
False positive rate (1- Specificity)
AUC = 0.61 (95% CI =
0.52-0.72; P = 0.02)

Figure 4