Gene variants previously associated with reduced soft tissue injury risk: Part 1 – independent associations with elite status in rugby

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

Published online: 29 Mar 2022.

Submit your article to this journal ⬚

Article views: 714

View related articles ⬚

View Crossmark data ⬚

Gene variants previously associated with reduced soft tissue injury risk: Part 1 – independent associations with elite status in rugby

Jon Brazier [a,b,c], Mark R. Antrobus[a,b,d], Adam J. Herbert[e], Peter C. Callus[a,b], Georgina K. Stebbings[a,b], Stephen H. Day[f], Shane M. Heffernan[g], Liam P. Kilduff[g], Mark A. Bennett[g], Robert M. Erskine [h,l], Stuart. M. Raleigh[i], Malcolm Collins[j], Yannis. P. Pitsiladis[k] and Alun G. Williams[a,b,g,l]

[a]Manchester Metropolitan University Institute of Sport, Manchester, UK; [b]Department of Sport and Exercise Sciences, Musculoskeletal Science and Sports Medicine Research Centre, Faculty of Science and Engineering, Manchester Metropolitan University, Manchester, UK; [c]Department of Psychology, Sport and Geography, University of Hertfordshire, Hatfield, UK; [d]Sport and Exercise Science, University of Northampton, Northampton, UK; [e]Department of Sport and Exercise, School of Health Sciences, Birmingham City University, Birmingham, UK; [f]Faculty of Science and Engineering, University of Wolverhampton, Wolverhampton, UK; [g]Applied Sports Science Technology and Medicine Research Centre (A-STEM), Faculty of Science and Engineering, Swansea University, Swansea, UK; [h]Research Institute for Sport & Exercise Sciences, Liverpool John Moores University, Liverpool, UK; [i]School of Health Sciences, Coventry University, Coventry, UK; [j]Division of Exercise Science and Sports Medicine, Department of Human Biology, University of Cape Town, Cape Town, South Africa; [k]FIMS Reference Collaborating Centre of Sports Medicine for Anti-Doping Research, University of Brighton, Brighton, UK; [l]Institute of Sport, Exercise and Health, University College London, London, UK

ABSTRACT

There is growing evidence of genetic contributions to tendon and ligament pathologies. Given the high incidence and severity of tendon and ligament injuries in elite rugby, we studied whether 13 gene polymorphisms previously associated with tendon/ligament injury were associated with elite athlete status. Participants from the RugbyGene project were 663 elite Caucasian male rugby athletes (RA) (mean (standard deviation) height 1.85 (0.07) m, mass 101 (12) kg, age 29 (7) yr), including 558 rugby union athletes (RU) and 105 rugby league athletes. Non-athletes (NA) were 909 Caucasian men and women (56% female; height 1.70 (0.10) m, mass 72 (13) kg, age 41 (23) yr). Genotypes were determined using TaqMan probes and groups compared using X^2 and odds ratio (OR). *COLGALT1* rs8090 AA genotype was more frequent in RA (27%) than NA (23%; $P =$ 0.006). *COL3A1* rs1800255 A allele was more frequent in RA (26%) than NA (23%) due to a greater frequency of GA genotype (39% vs 33%). For *MIR608* rs4919510, RA had 1.7 times the odds of carrying the CC genotype compared to NA. *MMP3* rs591058 TT genotype was less common in RA (25.1%) than NA (31.2%; $P < 0.04$). For *NID1* rs4660148, RA had 1.6 times the odds of carrying the TT genotype compared to NA. It appears that elite rugby athletes have an inherited advantage that contributes to their elite status, possibly via resistance to soft tissue injury. These data may, in future, assist personalised management of injury risk amongst athletes.

KEYWORDS

Genetics; ligament; rugby league; rugby union; tendon

Highlights

- The elite rugby athletes we studied had differing genetic characteristics to non-athletes regarding genetic variants previously associated with soft-tissue injury risk.
- *COLGALT1* rs8090, *COL3A1* rs1800255, *MIR608* rs4919510, *MMP3* rs591058 and *NID1* rs4660148 were all associated with elite status in rugby.
- We propose that elite rugby athletes might possess an inherited resistance to soft tissue injury, which has enabled them to achieve elite status despite exposure to the high-risk environment of elite rugby.

1. Introduction

Elite rugby has one of the highest reported injury incidences of any professional sport Brooks and Kemp

(2008). This is likely due to a combination of well-established injury surveillance systems and the characteristics of the game, whereby high-impact body collision

frequently occurs, in addition to the high intensity, multispeed and multidirectional nature of play Brazier et al. (2020). Meta-analyses have reported the total incidence of injury (injuries per 1000 player h) as 81/1000 in matches (~3 injuries per match) and 3/1000 in training for elite rugby union (RU) athletes, with the majority being tendon, ligament and muscle injuries of the lower limb Williams, Trewartha, Kemp, and Stokes (2013). Within rugby league (RL), injury incidence rates have been reported at 172/1000 in matches, King, Gissane, Clark, and Marshall (2014) with the majority of injuries occurring to the lower limb via strains and sprains King et al. (2014). Injury incidence and severity also appear to differ between playing positions within RU, with backs having a higher rate of incidence and severity compared to forwards in recent Rugby World Cup competitions Fuller, Taylor, Kemp, and Raftery (2017). Furthermore, some of the most severe (days absence from full training or match play) injuries for both RU and RL are those affecting tendon and ligament, Brazier et al. (2019) and therefore potentially the most debilitating to a player and playing squad. For example, elite RU forwards (from 12 different English Premiership clubs) had 988, 726 and 718 days absence across 2 seasons from anterior cruciate ligament (ACL), Achilles tendon and medial collateral ligament injuries, respectively Brooks and Kemp (2008).

Genetic variation may have a strong influence on inter-individual differences in tendon and ligament structure and function, which could alter an individual's risk of injury. Magnusson, Turkiewicz, Hughes, Frobell, and Englund (2020) Indeed, it was recently found that ACL rupture was ~69% heritable. Magnusson et al. (2020) Inter-individual variability of tendon and ligament properties is likely to cause microtrauma and macrotrauma at differing strain levels among individuals, thus similar injury-inciting events amongst rugby players may have vastly different outcomes Brazier et al. (2019). Type I collagen is the predominant collagen type in ligaments and tendons accounting for ~90% in ligaments Frank (2004) and ~95% in tendons Riley et al. (1994). The remaining 5-10% consists mainly of type III and V collagen with the other fibril forming or associated collagen types present in trace quantities Frank (2004). The diameter and formation of the type I collagen fibril is regulated by types V and III collagen amongst other molecules Banos, Thomas, and Kuo (2008). The α1 chains of types I, III and V collagen are encoded by the COL1A1, COL3A1 and COL5A1 genes, respectively. Polymorphisms (COL1A1 rs1800012, COL3A1 rs1800255, COL5A1 rs12722 and rs3196378) within these genes have previously been associated with ACL injury Khoschnau et al. (2008); Posthumus

et al. (2009); Stępień-Słodkowska et al. (2015); O'Connell et al. (2015); Brown et al. (2017)

The biomechanical properties of ligaments and tendons are principally a component of the extracellular matrix (ECM), which is in a constant state of dynamic equilibrium between synthesis and degradation Riley (2004). This is controlled in part by the balance of matrix metalloproteinase (MMP) and tissue inhibitors of metalloproteinase (TIMP), as their activities regulate the amount of ECM turnover Riley (2004). Alterations to this state of dynamic equilibrium may underpin the degenerative changes seen in the pathological progression of asymptomatic tendons, Jones et al. (2006) with imbalances producing collagen disruption Dalton, Cawston, Riley, Bayley, and Hazleman (1995). One of the proteins from the MMP family, MMP3, which is encoded by the MMP3 gene has previously had several polymorphisms (rs591058, rs650108 and rs679620) associated with Achilles tendinopathy Raleigh et al. (2009); Gibbon et al. (2017) Furthermore, the rs679620 polymorphism has been associated with Achilles tendon El Khoury, Ribbans, and Raleigh (2016) and ACL rupture Posthumus et al. (2012). In addition, the TIMP2 gene, which is 1 of 4 TIMPs that are natural inhibitors of the MMPs, Posthumus et al. (2012); Visse and Nagase (2003) also has a polymorphism (rs4789932) previously associated with Achilles tendinopathy El Khoury et al. (2016); El Khoury et al. (2013)

Other genes and pathways have been associated with increased risk of soft tissue injury. One being angiogenesis which is essential during the repair and remodelling of injured tendons and has been implicated in matrix remodelling following mechanical loading Petersen et al. (2004). Vascular endothelial growth factor (VEGF) is an endothelial cell mitogen that stimulates angiogenesis with the A isoform, coded by VEGFA, thought to be the most potent. Petersen et al. (2004) The majority of the biological effects of VEGFA are facilitated via its receptor: Kinase insert-domain receptor (KDR). Genes that encode for both of these proteins (VEGFA and KDR) have previously been investigated for their associations with ACL rupture and Achilles tendinopathy, with a VEFGA gene polymorphism (rs699947) associated with both forms of injury Rahim et al. (2014), Rahim et al. (2016) Additionally, several genetic variants recently identifed in a genome-wide assocation study (GWAS) for Achilles tendon and ACL tears and tendinopathy, Kim et al. (2017) are worthy of future study: COLGALT1 rs8090 is potentially important in the aetiology of connective tissue disorders due to post-translational modifications possibly disrupting collagen modifying enzymes; Schegg, Hülsmeier, Rutschmann, Maag, and Hennet (2009) NID1 rs4660148 encodes a member of the

nidogen family thought to play a role in the development of the ECM; Ho, Böse, Mokkapati, Nischt, and Smyth (2008) *MIR608* rs4919510 encodes a small non-coding RNA involved in gene silencing and translational repressions Matzke and Birchler (2005).

Given the association of genetic markers with injury risk, and that regular participation at the elite level would mean that players have been exposed to one of the highest levels of risk for tendon and ligament injury in any professional sporting environment, it is plausible that elite rugby athletes may possess an inherited resistance against soft tissue injury, which has enabled them to achieve elite status despite exposure to the high-risk environment of elite rugby. Indeed, recent research provides some evidence of this, where the injury-protective C alleles and CC genotypes of both the *COL5A1* rs12722 and rs3196378 polymorphisms were, individually and in combination, associated with elite athlete status in rugby Heffernan et al. (2017). This suggests an inherited resistance against soft tissue injury could enhance the chance of career success in certain sports. Therefore, the principle objective of the present study was to investigate whether the following genetic polymorphisms, previously associated with tendon and ligament injury, differed in genotype and allele frequencies between elite rugby athletes and a non-athlete population, and/or between playing positions: *COL1A1* (rs1800012), *COL3A1* (rs1800255), *KDR* (rs1870377), *MMP3* (rs679620, rs591058 and rs650108), *TIMP2* (rs4789932), *VEGFA* (rs699947), *COLGALT1* (rs8090), *MIR608* (rs4919510) and *NID1* (rs4660148). An additional objective was to expand on the previous work of Heffernan et al. Heffernan et al. (2017) by adding additional participants to the previously studied *COL5A1* (rs12722 and rs3196378) polymorphisms. It was hypothesised that elite rugby athletes would possess fewer of the injury-risk genotype/alleles than a non-athlete population.

2. Method

2.1. Participants

Manchester Metropolitan University, the University of Glasgow and the University of Cape Town ethics committees granted approval of this study, which complies with the Declaration of Helsinki. The participants were from the RugbyGene project, comprising elite Caucasian male rugby athletes (n = 663; mean (standard deviation) height 1.85 (0.07) m, mass 101 (12) kg, age 29 (7) yr) including 62.2% British, 13.6% South African, 10.5% Irish, 8.7% Italian and 5% of other nationalities were recruited, having given written informed consent.

Caucasian non-athletes (n = 909, 44% male, height 1.70 (0.10) m, mass 72 (13) kg, age 41 (23) yr) included 94.8% British, 3.5% South African and 1.7% other nationalities, and were eligible for inclusion if they were Caucasian, age ≥18 and had not competed at an elite or sub-elite competitive level in any sport. Rugby players were considered elite if they had competed regularly (∼5 matches) since 1995 in the highest professional league in the UK, Ireland, or South Africa for RU and the highest professional league in the UK for RL. Seven athletes competed in both elite RU and RL and were included in both groups that were analyzed separately. Of the RU athletes, 49.1% had competed at international level for a "high performance union" (Regulation 16, http://www.worldrugby.org), and 42% of RL athletes had competed at international level. It should be noted that for *COL5A1* (rs12722) and *COL5A1* (rs3196378) data from 540 elite male rugby athletes and 565 non-athletes were included from a previous study Heffernan et al. (2017).

2.2. Procedures

As previously reported, Heffernan et al. (2016) blood, buccal or saliva samples were collected and stored at −20°C until processing. DNA isolation was performed using two procedures: Firstly, using 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. In the second procedure, DNA was isolated from whole blood by a different protocol Lahiri and Nurnberger (1991).

Genotyping for *COLGALT1* (rs8090), *COL1A1* (rs1800012), *COL3A1* (rs1800255), *COL5A1* (rs12722 and rs3196378), *KDR* (rs1870377), *MIR608* (rs4919510), *MMP3* (rs679620, rs591058, and rs650108), *NID1* (rs4660148), *TIMP2* (rs4789932) and *VEGFA* (rs699947) was performed using 2 protocols. In both protocols, the appropriate TaqMan assays were utilised (Applied Biosystems, Paisley, UK) and assay context sequences for each polymorphism are presented in the supplementary material (SM) 1. Protocol one: ∼374 elite male rugby athlete samples were genotyped via real-time PCR using a StepOnePlus (Applied Biosystems) as previously described in detail, Heffernan et al. (2016) with adjustment of thermocycling conditions because GTXpress Master Mix (Applied Biosystems) was used for 75 samples. Protocol two: ∼262 elite male rugby athletes (596 for *MMP3* rs591058) and 909 non-athletes were genotyped for the aforementioned polymorphisms except *COL5A1* (rs12722 and rs3196378), 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) and 0.18 μL ROX reference dye (Invitrogen, Paisley, UK) were combined per assay inlet. An integrated fluid circuit controller RX (Fluidigm) mixed samples and assays using a Load Mix (166x) script. PCR was performed using a real-time FC1 Cycler (Fluidigm) GT 192X24 Fast v1 protocol. In brief, denaturation began at 95°C for 120 s followed by 45 cycles of incubation at 95°C for 2 s and then annealing and extension at 60°C for 20 s. The EP1 Reader was used for end-point analysis. Genotyping analysis was performed with the Fluidigm SNP genotyping analysis software. Duplicates of all samples were in 100% agreement. A small number of samples were not successfully genotyped for all of the 13 polymorphisms investigated, hence there is slight variation in sample size for each polymorphism. Furthermore, COL5A1 rs12722 and rs3196378 contained an additional ~30 samples from a previous study Heffernan et al. (2017) which were no longer available for genotyping of the other polymorphisms.

2.2.1. RU forwards, backs and positional roles

To examine genotype and allele frequencies within the substantial RU cohort, athletes were placed into sub-groups according to their movement patterns. Two sub-groups were defined as forwards (props, hookers, locks, flankers, number eights) and backs (scrum halves, fly halves, centres, wings, full backs) Cahill, Lamb, Worsfold, Headey, and Murray (2013). The largest RU positional group within the sub-groups were the front 5 (props, hookers and locks) with a sample size of 210, so they were also analyzed as a discrete group.

2.3. Data analysis

SPSS for Windows version 26 (SPSS, Chicago, IL) software was used to conduct Pearson's Chi-square ($χ^2$) tests to compare genotype (using three analysis models; additive, recessive, and dominant) and allele frequencies between athletes and non-athletes and between RU positional subgroups. With 80% statistical power, analyses of all genetic models in positional subgroups compared with non-athletes (RU forwards, RU Front 5 and RU backs), were able to detect a small effect size (w) of 0.10 and analysis between sub-groups (RU forwards vs RU backs; RU front 5 vs RU backs) were able to detect a small-to-moderate effect size (w) of 0.22. For each

polymorphism, 32 tests were subjected to Benjamini-Hochberg corrections to control false discovery rate and corrected probability values are reported. Where appropriate, odds ratios (OR) were calculated to estimate effect size. Alpha was set at 0.05.

3. Results

Genotype frequencies were in Hardy-Weinberg equilibrium for all polymorphisms in the non-athlete and athlete groups apart from COLGALT1 rs8090 (RL athlete group), MIR608 rs4919510 (non-athlete, rugby athlete, RU and RU front 5 athlete groups), MMP3 rs679620 (RL athlete group), NID1 rs4660148 (non-athlete group) and TIMP2 rs4789932 (non-athlete group) (SM 2: Table 1). Athletes (all male) were taller and heavier ($P < 0.05$) than the male non-athletes.

For COLGALT1 rs8090, the AA genotype, proportion of A-allele carriers and A allele were overrepresented in all athletes (27.3%, 76% and 51.7%, respectively) and RU athletes (27.7%, 76.7%, and 51.2%) compared with non-athletes (23.0%, 70.1%, and 46.6%, Figure 1 and SM 2: Table 1, $P < 0.03$). Furthermore, the AA genotype, proportion of A-allele carriers and A allele were overrepresented in the sub-groups of RU forwards (29%, 79.6% and 54.2%) and RU front 5 (27.6%, 79.5% and 53.6%) compared with non-athletes (23.0%, 70.1%, and 46.6%, Figure 1, $P \leq 0.03$). Compared with non-athletes, RU forwards had 1.8 times the odds of possessing the AA genotype, and 2.1 times the odds of carrying the A allele (SM 3: Table 5). There were no differences in genotype or allele frequencies for COLGALT1 rs8090 for any other groups (RL vs non-athletes, RU backs vs non-athletes, RU forwards vs RU backs and RU front 5 vs RU backs). For $χ^2$, OR and allele/genotype frequency data for all SNPs, please refer to SM 2 and 3.

For COL3A1 rs1800255, GA genotype and carriage of the A allele were more common in all athletes (39.2% and 45.7%, respectively) compared to non-athletes (33.7% and 40.1%, $P < 0.04$). Furthermore, GA genotype, A-allele carriage and the A allele were overrepresented in RU athletes (41.0%, 47.5% and 27.1%) and RU forwards (42.9%, 50.3% and 28.9%) compared to non-athletes (33.7%, 40.1% and 23.3%, $P < 0.02$). For the RU front 5 sub-group, GA genotype and A-allele carriage were overrepresented (both 50%) compared to non-athletes (33.7% and 40.1%, $P < 0.02$). In addition, RU forwards had 1.5 times the odds of carrying the A allele compared to non-athletes (SM 3: Table 5). There were no differences in COL3A1 rs1800255 genotype or allele frequencies between any other groups.

For COL5A1 rs12722, the CC genotype, proportion of C-allele carriers and C allele were overrepresented in
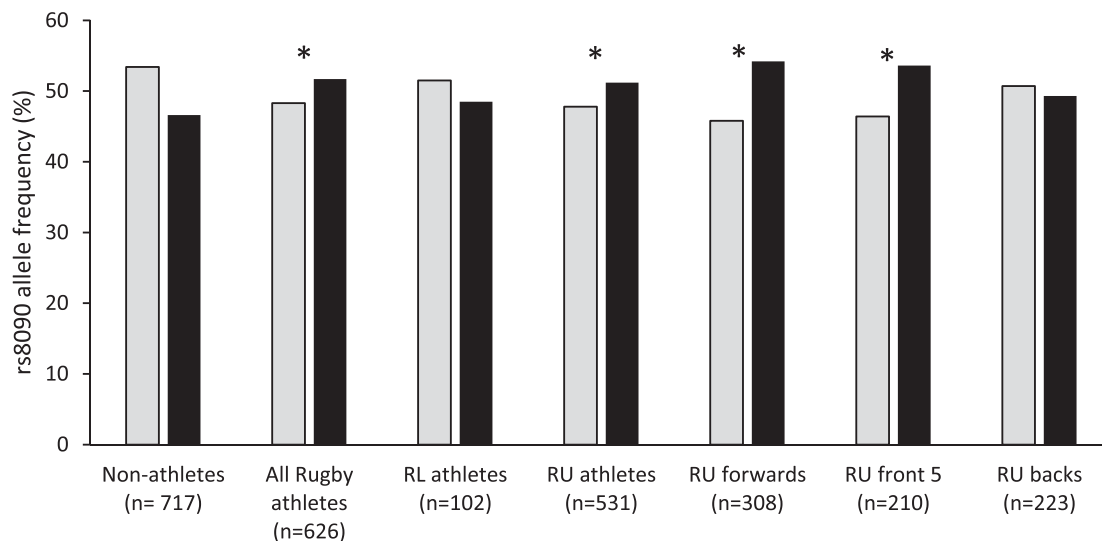
Figure 1. Allele frequency of *COLGALT1* rs8090 for non-athlete and athlete groups. Asterisks (*) indicate a difference in allele frequency between the particular athlete group or sub-group and non-athletes (*P* < 0.01). RL, rugby league; RU, rugby union;. Grey bars = G Allele; Black bars = A allele.

all athletes (22.9%, 73.3% and 48.1%, respectively), RU athletes (22.7%, 72.9% and 47.8%), RU backs (22.3%, 73.9% and 48.1%) and RU front 5 (23.3%, 75.3% and 49.3%) compared to non-athletes (17.3%, 66.8% and 42.1%, *P* < 0.01). Furthermore, the CC genotype and C allele were overrepresented in RU forwards (22.9% and 47.6%) compared to non-athletes (17.3% and 42.1%, *P* < 0.01). In RL athletes, the C allele was overrepresented (49.5%) compared to non-athletes (42.1%). There were no differences in genotype or allele frequencies for *COL5A1* rs12722 for any other groups.

For *COL5A1* rs3196378, the CC genotype and C allele were overrepresented, while the proportion of A allele carriers were underrepresented in all athletes (23.8%, 47.5% and 76.2%), RU athletes (24.4%, 48.0% and 75.5%) and RU forwards (25.6%, 48.3% and 74.4%) compared to non-athletes (17.8%, 42.4% and 82.2%, *P* < 0.01). Additionally, the C allele was overrepresented in the sub-group of RU front 5 (48.2%) compared to non-athletes (42.4%). There were no differences in genotype or allele frequencies for *COL5A1* rs3196378 for any other groups.

For *MIR608* rs4919510, the CC genotype and C allele were overrepresented, whilst the number of G-allele carriers were underrepresented in all athletes (64.0%, 79.0%, 36%, respectively), RL athletes (69.2%, 82.7% and 30.8%), RU athletes (63.0%, 78.4% and 36.9%) and RU front 5 (63.8%, 79.8% and 36.2%) compared to non-athletes (56.4%, 73.8%, 43.6%, *P* ≤ 0.04, Figure 2). Furthermore, the C allele was overrepresented while the number of G-allele carriers was underrepresented in the RU forwards sub-group (78.4% and 37.4%) compared to non-athletes (73.8% and 43.6%, *P* < 0.05, Figure 2).

Additionally, G-allele carriers were underrepresented in RU backs (36.3%) compared to non-athletes (43.6%, *P* < 0.05). All athletes had 1.7 times the odds of carrying the CC genotype, and RL athletes had 2.8 times the odds, compared to non-athletes (SM 3: Table 2). There were no differences in genotype or allele frequencies for *MIR608* rs4919510 for any other groups.

For *MMP3* rs591058, the TT genotype and T allele were overrepresented, whilst the proportion of C-allele carriers was overrepresented in all athletes (25.1%, 51.0% and 74.8%) compared to non-athletes (31.2%, 54.9% and 68.8%, *P* < 0.04). Furthermore, C-allele carriers were overrepresented in RL athletes (81.6%) compared to non-athletes (68.8%, *P* < 0.05). There were no differences in genotype or allele frequencies for any other groups. For *MMP3* rs679620, C-allele carriers were overrepresented in RL athletes (85.6%) compared to non-athletes (71.3%, *P* < 0.04). Compared to non-athletes, RL athletes had 2.4 times the odds of carrying the C allele (SM 3: Table 3). There were no differences in genotype or allele frequencies for *MMP3* rs591058 or rs679620 for any other groups.

For *NID1* rs4660148, the TT genotype was overrepresented while G-allele carriers were underrepresented in all athletes (10.4% and 89.6%, respectively), RU athletes (10.9% and 89.0%), and the sub-groups RU forwards (11.1% and 88.9%), RU front 5 (12.0% and 88.0%), RU backs (10.6% and 89.3%) compared to non-athletes (6.7% and 93.3%, *P* < 0.05). In addition, the G allele was also underrepresented in the sub-group of RU front 5 (65.5%) compared to non-athletes (70.8%, *P* < 0.05). RU forwards had 1.7 times the odds of carrying the TT genotype (SM 3: Table 5), while in the RU front 5 that
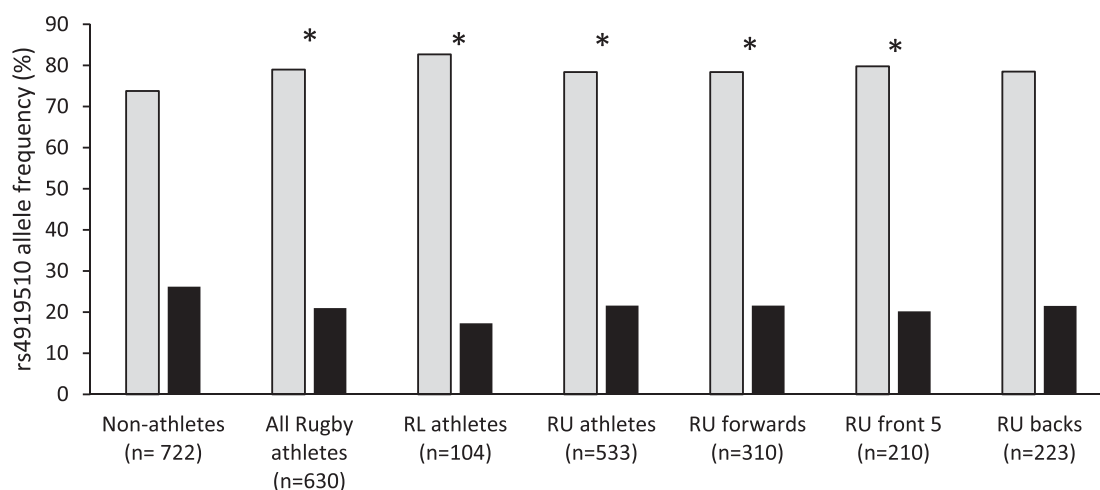
Figure 2. Allele frequency of *MIR608* rs4919510 for non-athlete and athlete groups. Asterisks (*) indicate a difference in allele frequency between the particular athlete group or sub-group and non-athletes (*P* ≤ 0.03). RL, rugby league; RU, rugby union. Grey bars = C allele; Black bars = G allele.

increased to 2.0 times the odds (SM 3: Table 6), compared to non-athletes. There were no differences in genotype or allele frequencies for *NID1* rs4660148 for any other groups.

For *COL1A1* (rs1800012), *KDR* (rs1870377), *MMP3* (rs650108), *TIMP2 (*rs4789932) and *VEGFA* (rs699947) there were no differences in genotype or allele frequencies between any groups. There were no differences in genotype or allele frequencies between RU forwards and RU backs, or between RU front 5 and RU backs, for any of the polymorphisms investigated (See SM 2 and 3 for further details).

4. Discussion

The present study is the first to identify associations between *COLGALT1* rs8090, *COL3A1* rs1800255, *MIR608* rs4919510, *MMP3* rs591058 and rs679620 and *NID1* rs4660148 polymorphisms and athlete status in a large cohort of elite rugby athletes. Furthermore, our findings with a larger cohort support the previous work of Heffernan et al. Heffernan et al. (2017) that found associations between *COL5A1* rs12722 and *COL5A1* rs3196378 polymorphisms and elite rugby status. As hypothesised, elite rugby athletes mostly carried more of the apparent injury-protective genotype/alleles than non-athletes, although this was not consistent for all polymorphisms.

To the best of our knowledge, this is the first study to investigate *COLGALT1* rs8090 and *NID1* rs4660148 in elite athletes. Both polymorphisms were previously identified via a GWAS for Achilles tendon and ACL tears and tendinopathy Kim et al. (2017). In that study, the G alleles of *COLGALT1* rs8090 ($P < 6 \times 10^{-4}$) and *NID1* rs4660148 (*P*

$< 5 \times 10^{-5}$) were most strongly associated with Achilles tendon injury and ACL rupture, respectively. While these results were not genome-wide significant, further investigation was warranted. The protective A allele and AA genotype of *COLGALT1* rs8090 are overrepresented in elite rugby athletes compared to non-athletes in our study, persisting within RU athletes, RU forwards and RU front 5. Indeed, RU forwards had over twice the odds of carrying the A allele than non-athletes. COLGALT1 initiates collagen glycosylation through its activation of beta (1-0) galactosyltransferase enzymes Schegg et al. (2009). Specifically, hydroxylysine can be modified by the transfer of galactose by galactosyltransferases Schegg et al. (2009). These posttranslational modifications might be important in the aetiology of connective tissue disorders due to the production of defective collagen modifying enzymes Schegg et al. (2009). Our results suggest that elite rugby athletes may have some inherited benefit via this pathway that may make them less susceptible to soft tissue injury. Further research into this polymorphism is warranted to establish its functionality.

For *NID1* rs4660148, the TT genotype was overrepresented and the G (risk) allele carriers underrepresented in elite rugby athletes compared to non-athletes, and this association continued across RU athletes and all RU sub-groups. *NID1* encodes a member of the nidogen family of basement membrane glycoproteins Ho et al. (2008). Nidogens are implicated as playing a major structural role in the basement membrane and thus the development of the ECM, particularly when tissues are experiencing rapid turnover and growth Ho et al. (2008). Therefore, they may influence the aetiology of musculoskeletal soft tissue injuries through their

functions within the ECM. It appears that the TT genotype of *NID1* rs4660148 is beneficial for rugby athletes to achieve elite status, possibly through superior resistance to soft tissue injury.

The CC genotype of *MIR608* rs4919510 was overrepresented and G-allele carriers underrepresented in elite rugby athletes across all groups (rugby athletes, RL athletes, RU athletes, RU forwards, RU front 5 and RU backs) compared to non-athletes, suggesting some inherited advantage to attaining elite rugby athlete status. Micro-RNAs (miRNA) are a class of small non-coding RNAs that induce gene silencing and translational repression Matzke and Birchler (2005). Allele-specific polymorphisms within miRNA target sites influence the tissue-specific miRNA regulation of hundreds of genes, which implies that their genetic variation may be a prevalent cause of inter-individual phenotypic variability Kim and Bartel (2009). One of the miRNA family, *MIR608* rs4919510, has been associated with altered risk of Achilles tendinopathy Brown et al. (2017), Kim et al. (2017), Abrahams, Laguette, Prince, and Collins (2013) However, the evidence is not consistent regarding which genotype is injury-protective. Abrahams et al. Abrahams et al. (2013) found the CC genotype overrepresented within a tendinopathy group versus uninjured, supported by moderate GWAS evidence ($P < 5 \times 10^{-8}$) when covariates were removed Kim et al. (2017). However, the only other study to date could not replicate these results, but found an association between the CG genotype and Achilles tendon rupture Brown et al. (2017). Indeed, our data demonstrate a likely benefit from carrying the CC genotype in achieving elite rugby athlete status. Notably, RL athletes were almost 3 times more likely to carry the CC genotype than non-athletes.

Further evidence of a possible genetic role in elite athlete status is provided by the overrepresentation of the GA genotype and a higher proportion of A-allele carriers at *COL3A1* rs1800255 within elite rugby athletes, RU athletes, RU forwards and RU front 5 versus non-athletes. Four studies previously investigated the association between *COL3A1* rs1800255 and ACL rupture but none examined tendon pathology. A higher frequency of the AA genotype was identified in recreational skiers and professional footballers compared to non-injured groups Stępień-Słodkowska et al. (2015), O'Connell et al. (2015) Furthermore, when covariates were removed, weak supporting evidence was found in a GWAS ($P = 0.03$) Kim et al. (2017). More recent evidence on elite female athletes from high-risk team sports found no associations between rs1800255 and ACL rupture Sivertsen et al. (2019). Our data show more elite male rugby athletes carry the purported risk A allele than non-athletes,

which may put them at increased risk of ACL injury. Indeed, RU forwards had 1.5 times greater odds of carrying the A allele than non-athletes. However, how the A allele might influence injury risk (e.g. affecting collagen formation and/or structure) is not clear. It seems *COL3A1* can influence the tensile strength of collagen, Kluivers et al. (2009) and as such there may be some benefit of the A allele for elite athlete status, although perhaps not for ACL injury protection.

The *MMP3* gene encodes the protein MMP3 which has a fundamental role in the development, repair and remodelling of connective tissues, controlling ECM homeostasis via proteolytic activity Foster, Morse, Onambele, Ahmetov, & Williams, 2012). There is a possible relationship between the CC genotype of rs591058, GG genotype of rs679620, AA genotype of rs650108, and Achilles tendinopathy Raleigh et al. (2009). However, these findings have not been replicated within a Caucasian population, Gibbon et al. (2017) or for ACL injury Posthumus et al. (2012). We observed a higher proportion of C-allele carriers in elite rugby athletes compared to non-athletes for *MMP3* rs591058. Furthermore, a higher proportion of C-allele carriers were observed within RL athletes for *MMP3* rs591058 and rs679620. Indeed, RL athletes had twice the odds of carrying the C allele at rs591058 and almost 2.5 times the odds of carrying the C allele at rs679620. Our findings suggest some likely benefit from carrying the C allele, perhaps via better ECM regulation and thus a more robust athlete.

Our hypothesis of an inherited injury resistance within elite rugby athletes was further supported by our findings regarding the *COL5A1* polymorphisms rs12722 and rs3196378. Heffernan et al. Heffernan et al. (2017) previously discussed the possible injury-protective nature of carrying the C allele for both polymorphisms and found them to be overrepresented in elite rugby athletes. Including the previous 540 athletes, our results with an additional 123 athletes extend the previous report as we continue to observe the CC genotype and C allele of both rs12722 and rs3196378 to be more common in elite rugby athletes than non-athletes.

As part of this investigation, we compared genotype and allele frequencies of RU athlete sub-groups (e.g. forwards vs backs). We found no differences between subgroups for any polymorphism, suggesting elite rugby athletes, regardless of playing position, are equally likely to benefit from any inherited lower risk of soft tissue injury.

Seven of the 91 analysis groups for the present study were not in Hardy-Weinberg Equilibrium, which is a potential limitation. However, statistical tests with alpha at 0.05 would be expected to produce four or

five deviations from Hardy-Weinberg equilibrium, so seven is only a little in excess of that. Furthermore, due to our 100% duplication process during genotyping, and broad agreement between our observed genotype frequencies and those of the general population with similar geographic ancestry (e.g. https://www.ncbi.nlm.nih.gov/snp/), we are confident in the internal validity of our data. An additional limitation is that the present study is focused on elite rugby athletes, so these results might not be replicated in other populations and equivalent sport-specific investigations are encouraged.

5. Conclusion

In conclusion, we have presented the first associations between the *COLGALT1* rs8090, *COL3A1* rs1800255, *MIR608* rs4919510, *MMP3* rs591058 and rs679620 and *NID1* rs4660148 polymorphisms and elite status in rugby athletes. We have also extended Heffernan et al.'s Heffernan et al. (2017) associations of the *COL5A1* rs12722 and rs3196378 polymorphisms with elite status in rugby using a larger cohort. The potential injury-reducing mechanisms need more elucidation. Nevertheless, we propose that elite rugby athletes might possess an inherited resistance to soft tissue injury, assisting them to achieve elite status in the high-risk environment of elite rugby. These data may, in future, assist personalised management of injury risk amongst athletes. It is very likely that any inherited resistance to injury is polygenic, therefore, in Part 2, we investigate total genotype scores, inferred haplotypes and epistasis interactions to assess this possibility.

ORCID

Jon Brazier 🔟 http://orcid.org/0000-0002-4104-9447
Robert M. Erskine 🔟 http://orcid.org/0000-0002-5705-0207

References

Abrahams, Y., Laguette, M. J., Prince, S., & Collins, M. (2013). Polymorphisms within the COL5A1 3′-UTR that alters mRNA structure and the MIR608 gene are associated with Achilles tendinopathy. *Annals of Human Genetics*, *77*, 204–214. doi:10.1111/ahg.12013

Banos, C. C., Thomas, A. H., & Kuo, C. K. (2008). Collagen fibrillogenesis in tendon development: Current models and regulation of fibril assembly. *Birth Defects Res C Embryo Today: Reviews*, *84*, 228–244. doi:10.1002/bdrc.20130

Brazier, J., Antrobus, M., Stebbings, G. K., Day, S. H., Callus, P., Erskine, R. M., … Williams, A. G. (2019). Tendon and ligament injuries in elite rugby: The potential genetic influence. *Sports (Basel, Switzerland)*, *7*, 138. doi:10.3390/sports7060138.

Brazier, J., Antrobus, M., Stebbings, G. K., Day, S. H., Callus, P., Erskine, R. M., … Williams, A. G. (2020). Anthropometric and physiological characteristics of elite male rugby athletes. *Journal of Strength and Conditioning Research*, *34*, 1790–1801. doi:10.1519/JSC.0000000000002827

Brooks, J. H. M., & Kemp, S. P. T. (2008). Recent trends in rugby union injuries. *Clinics in Sports Medicine*, *27*, 51. doi:10.1016/j.csm.2007.09.001

Brown, K. L., Seale, K. B., El Khoury, L. Y., Posthumus, M., Ribbans, W. J., Raleigh, S. M., … September, A. V. (2017). Polymorphisms within the COL5A1 gene and regulators of the extracellular matrix modify the risk of Achilles tendon pathology in a British case-control study. *Journal of Sports Sciences*, 2017/08/03 *35*, 1475–1483. doi:10.1080/02640414.2016.1221524

Cahill, N., Lamb, K., Worsfold, P., Headey, R., & Murray, S. (2013). The movement characteristics of English Premiership rugby union players. *Journal of Sports Sciences*, *31*, 229–237. doi:10.1080/02640414.2012.727456

Dalton, S., Cawston, T. E., Riley, G. P., Bayley, I. J., & Hazleman, B. L. (1995). Human shoulder tendon biopsy samples in organ culture produce procollagenase and tissue inhibitor of metalloproteinases. *Annals of the Rheumatic Diseases*, *54*, 571–577. doi:10.1136/ard.54.7.571

El Khoury, L., Posthumus, M., Collins, M., Handley, C. J., Cook, J., & Raleigh, S. M. (2013). Polymorphic variation within the ADAMTS2, ADAMTS14, ADAMTS5, ADAM12 and TIMP2 genes and the risk of Achilles tendon pathology: A genetic association study. *Journal of Science and Medicine in Sport*, *16*, 493–498. doi:10.1016/j.jsams.2013.02.006

El Khoury, L., Ribbans, W. J., & Raleigh, S. M. (2016). MMP3 and TIMP2 gene variants as predisposing factors for Achilles tendon pathologies: Attempted replication study in a British case–control cohort. *Meta Gene*, *9*, 52–55. doi:10.1016/j.mgene.2016.03.007.

Foster, B., Morse, C. I., Onambele, G. L., Ahmetov, I. I., & Williams, A. G. (2012). Genetic variation, protein composition and potential influences on tendon properties in humans. *Open Sports Med J*, *6*, 8–21. doi:10.2174/1874387001206010008

Frank, C. B. (2004). Ligament structure, physiology and function. *Journal of Musculoskeletal & Neuronal Interactions*, *4*, 199.

Fuller, C. W., Taylor, A., Kemp, S. P. T., & Raftery, M. (2017). Rugby world cup 2015: World rugby injury surveillance study. *British Journal of Sports Medicine*, *51*, 51–57. doi:10.1136/bjsports-2016-096275.

Gibbon, A., Hobbs, H., van der Merwe, W., Raleigh, S. M., Cook, J., Handley, C. J., … September, A. V. (2017). The MMP3 gene in musculoskeletal soft tissue injury risk profiling: A study in two independent sample groups. *Journal of Sports Sciences*, *35*, 655–662. doi:10.1080/02640414.2016.1183806.

Heffernan, S. M., Kilduff, L. P., Erskine, R. M., Day, S. H., McPhee, J. S., McMahon, G. E., … Williams, A. G. (2016). Association of ACTN3 R577X but not ACE I/D gene variants with elite rugby union player status and playing position. *Physiological Genomics*, *48*, 196–201.

Heffernan, S. M., Kilduff, L. P., Erskine, R. M., Day, S. H., Stebbings, G. K., Cook, C. J., … Williams, A. G. (2017). COL5A1 gene variants previously associated with reduced soft tissue injury risk are associated with elite athlete status in rugby. *BMC Genomics*, *18*(Suppl 8), 820. doi:10.1186/s12864-017-4187-3

Ho, M. S. P., Böse, K., Mokkapati, S., Nischt, R., & Smyth, N. (2008). Nidogens—extracellular matrix linker molecules. *Microscopy Research and Technique*, *71*, 387–395. doi:10.1002/jemt.20567

Jones, G. C., Corps, A. N., Pennington, C. J., Clark, I. M., Edwards, D. R., Bradley, M. M., … Riley, G. P. (2006). Expression profiling of metalloproteinases and tissue inhibitors of metalloproteinases in normal and degenerate human Achilles tendon. *Arthritis and Rheumatism*, *54*, 832–842. doi:10.1002/art.21672.

Khoschnau, S., Melhus, H., Jacobson, A., Rahme, H., Bengtsson, H., Ribom, E., … Michaëlsson, K. (2008). Type I collagen alpha1 Sp1 polymorphism and the risk of cruciate ligament ruptures or shoulder dislocations. *American Journal of Sports Medicine*, *36*, 2432.

Kim, J., & Bartel, D. P. (2009). Allelic imbalance sequencing reveals that single-nucleotide polymorphisms frequently alter microRNA-directed repression. *Nature Biotechnology*, *27*, 472–477. doi:10.1038/nbt.1540.

Kim, S. K., Roos, T. R., Roos, A. K., Kleimeyer, J. P., Ahmed, M. A., Goodlin, G. T., … Dragoo, J. L. (2017). Genome-wide association screens for Achilles tendon and ACL tears and tendinopathy. *PLoS One*, *12*, e0170422. doi:10.1371/journal.pone.0170422.

King, D., Gissane, C., Clark, T., & Marshall, S. (2014). The incidence of match and training injuries in rugby league: A pooled data analysis of published studies. *International Journal of Sports Science & Coaching*, *9*, 417–431. doi:10.1260/1747-9541.9.2.417

Kluivers, K. B., Dijkstra, J. R., Heniks, J. C. M., Lince, S. L., Vierhout, M. E., & Kempen, L. (2009). COL3A1 2209g > A is a predictor of pelvic organ prolapse. *International Urogynecology Journal and Pelvic Floor Dysfunction*, *20*, 1113–1118. doi:10.1007/s00192-009-0913-y

Lahiri, D. K., & Nurnberger, J. J. I. (1991). A rapid non-enzymatic method for the preparation of HMW DNA from blood for RFLP studies. *Nucleic Acids Research*, *19*, 5444–5444. doi:10.1093/nar/19.19.5444

Magnusson, K., Turkiewicz, A., Hughes, V., Frobell, R., & Englund, M. (2020). High genetic contribution to anterior cruciate ligament rupture: Heritability ∼69. *British Journal of Sports Medicine*, doi:10.1136/bjsports-2020-102392

Matzke, M. A., & Birchler, J. A. (2005). RNAi-mediated pathways in the nucleus. *Nature Reviews Genetics*, *6*, 24–35. doi:10.1038/nrg1500

O'Connell, K., Knight, H., Ficek, K., Leonska-Duniec, A., Maciejewska-Karlowska, A., Sawczuk, M., … Collins, M. (2015). Interactions between collagen gene variants and risk of anterior cruciate ligament rupture. *European Journal of Sport Science*, *15*, 341–350. doi:10.1080/17461391.2014.936324

Petersen, W., Pufe, T., Zantop, T., Tillmann, B., Tsokos, M., & Mentlein, R. (2004). Expression of VEGFR-1 and VEGFR-2 in degenerative Achilles tendons. *Clinical Orthopaedics and Related Research*, *420*, 286–291. doi:10.1097/00003086-200403000-00040

Posthumus, M., Collins, M., van der Merwe, L., O'Cuinneagain, D., van der Merwe, W., Ribbans, W. J., … Raleigh, S. M. (2012). Matrix metalloproteinase genes on chromosome 11q22 and the risk of anterior cruciate ligament (ACL) rupture. *Scandinavian Journal of Medicine & Science in Sports*, *22*, 523–533. doi:10.1111/j.1600-0838.2010.01270.x

Posthumus, M., September, A. V., Keegan, M., O'Cuinneagain, D., Van der Merwe, W., Schwellnus, M. P., & Collins, M. (2009). Genetic risk factors for anterior cruciate ligament ruptures: COL1A1 gene variant. *British Journal of Sports Medicine*, *43*, 352–356. doi:10.1136/bjsm.2008.056150.

Rahim, M., El Khoury, L. Y., Raleigh, S. M., Ribbans, W. J., Posthumus, M., Collins, M., & September, A. V. (2016). Human genetic variation, sport and exercise medicine, and Achilles tendinopathy:Role for angiogenesis-associated genes. *OMICS*, *20*, 520–527. doi:10.1089/omi.2016.0116

Rahim, M., Gibbon, A., Hobbs, H., van der Merwe, W., Posthumus, M., Collins, M., & September, A. V. (2014). The association of genes involved in the angiogenesis-associated signaling pathway with risk of anterior cruciate ligament rupture. *Journal of Orthopaedic Research*, *32*, 1612–1618. doi:10.1002/jor.22705

Raleigh, S. M., Van Der Merwe, L., Ribbans, W. J., Smith, R. K. W., Schwellnus, M. P., & Collins, M. (2009). Variants within the MMP3 gene are associated with Achilles tendinopathy: Possible interaction with the COL5A1 gene. *British Journal of Sports Medicine*, *43*, 514–520. doi:10.1136/bjsm.2008.053892

Riley, G. (2004). The pathogenesis of tendinopathy. A molecular perspective. *Rheumatology*, *43*, 131–142. doi:10.1093/rheumatology/keg448

Riley, G. P., Harrall, R. L., Constant, C. R., Chard, M. D., Cawston, T. E., & Hazleman, B. L. (1994). Glycosaminoglycans of human rotator cuff tendons: Changes with age and in chronic rotator cuff tendinitis. *Annals of the Rheumatic Diseases*, *53*, 367–376. doi:10.1136/ard.53.6.367

Schegg, B., Hülsmeier, A. J., Rutschmann, C., Maag, C., & Hennet, T. (2009). Core glycosylation of collagen is initiated by two β(1-O)galactosyltransferases. *Molecular and Cellular Biology*, *29*, 943. doi:10.1128/MCB.02085-07

Sivertsen, E. A., Haug, K. B. F., Kristianslund, E. K., Trøseid, A.-M. S., Parkkari, J., Lehtimäki, T., … Bahr, R. (2019). No association between risk of anterior cruciate ligament rupture and selected candidate collagen gene variants in female elite athletes from high-risk team sports. *American Journal of Sports Medicine*, *47*, 52–58. doi:10.1177/0363546518808467

Stępień-Słodkowska, M., Ficek, K., Maciejewska-Karłowska, A., Sawczuk, M., Ziętek, P., Król, P., … Cięszczyk, P. (2015). Overrepresentation of the COL3A1 AA genotype in Polish skiers with anterior cruciate ligament injury. *Biology of Sport*, *32*, 143–147. doi:10.5604/20831862.1144416

Visse, R., & Nagase, H. (2003). Matrix metalloproteinases and tissue inhibitors of metalloproteinases: Structure, function, and biochemistry. *Circulation Research*, *92*, 827–839. doi:10.1161/01.RES.0000070112.80711.3D

Williams, S., Trewartha, G., Kemp, S., & Stokes, K. (2013). A meta-analysis of injuries in senior men's professional rugby union. *Sports Medicine*, *43*, 1043–1055. doi:10.1007/s40279-013-0078-1