Gene Variants Previously Associated with Reduced Soft Tissue Injury Risk: Part 1 – Independent Associations with Elite Status in Rugby

Authors; 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]

Affiliations;

[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 M1 5GD, UK

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

[d] Sport and Exercise Science, University of Northampton, Northampton NN1 5PH, UK

[e] Department of Sport and Exercise, School of Health Sciences, Birmingham City University, Birmingham, B15 3TN, UK

[f] Faculty of Science and Engineering, University of Wolverhampton, Wolverhampton WV1 1LY, UK

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

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

[i] School of Health Sciences, Coventry University, Coventry, CV1 5FB, UK

[j] Division of Exercise Science and Sports Medicine, Department of Human Biology, University of Cape Town, Cape Town, 7700, South Africa

[k] FIMS Reference Collaborating Centre of Sports Medicine for Anti-Doping Research, University of Brighton, Brighton, BN2 0YJ, UK

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

*Corresponding author: j.brazier2@herts.ac.uk
University of Hertfordshire
College Lane
Hertfordshire
AL10 9AB
Phone: UK 01707 284000

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 personalized management of injury risk amongst athletes.

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.

Keywords

Genetics, Ligament, Rugby League, Rugby Union, Tendon.

1. Introduction

Elite rugby has one of the highest reported injury incidences of any professional sport.[1] 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.[2] 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.[3] Within rugby league (RL), injury incidence rates have been reported at 172/1000 in matches,[4] with the majority of injuries occurring to the lower limb via strains and sprains.[4] 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.[5] 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,[6] 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.[1]

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.[7] Indeed, it was recently found that ACL rupture was ~69% heritable.[7] 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.[6] Type I collagen is the predominant collagen type in ligaments and tendons accounting for ~90% in ligaments[8] and ~95% in tendons.[9] 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.[8] The diameter and formation of the type I collagen fibril is regulated by types V and III collagen amongst other molecules.[10] 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.[11-15]

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.[16] 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.[16] Alterations to this state of dynamic equilibrium may underpin the degenerative changes seen in the pathological progression of asymptomatic tendons,[17] with imbalances producing collagen disruption.[18] 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.[19, 20] Furthermore, the rs679620 polymorphism has been associated with Achilles tendon[21] and ACL rupture.[22] In addition, the *TIMP2* gene, which is 1 of 4 TIMPs that are natural inhibitors of the MMPs,[22, 23] also has a polymorphism (rs4789932) previously associated with Achilles tendinopathy.[21, 24]

Other genes and pathways have been associated with increased risk of soft tissue injury. One being angiogenesis which is essential during the repair and remodeling of injured tendons and has been implicated in matrix remodeling following mechanical loading.[25] 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.[25] 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.[26, 27] Additionally, several genetic variants recently identifed in a genome-wide assocation study (GWAS) for Achilles tendon and ACL tears and tendinopathy,[28] 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;[29] *NID1* rs4660148 encodes a member of the nidogen family thought to play a role in the development of the ECM;[30]

MIR608 rs4919510 encodes a small non-coding RNA involved in gene silencing and translational repressions.[31]

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.[32] 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.[32] by adding additional participants to the previously studied *COL5A1* (rs12722 and rs3196378) polymorphisms. It was hypothesized 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.[32]

2.2 Procedures

As previously reported,[33] 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.[34]

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 utilized (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,[33] 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[32] 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 subgroups according to their movement patterns. Two sub-groups were defined as forwards (props, hookers, locks, flankers, number eights) and backs (scrum halves, fly halves, centers, wings, full backs).[35] 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%, Fig. 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%, Fig. 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.

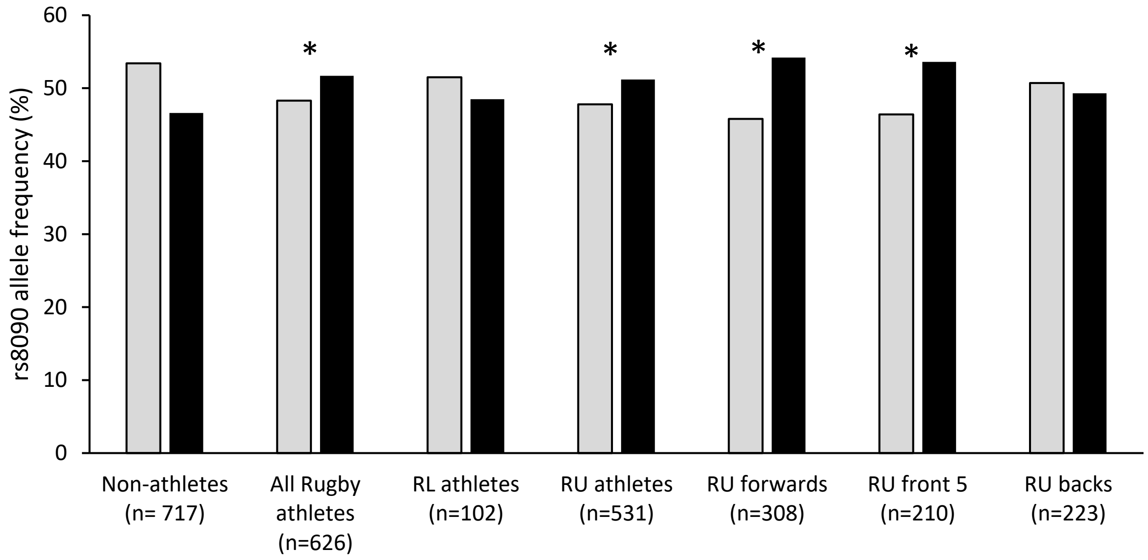

Figure 1. Allele frequency of *COLGALT1* rs8090 for non-athlete and athlete groups. Asterisks (*) indicate a difference in allele frequency between the particular athlete group or sub-group and non-athletes ($P < 0.01$). RL, rugby league; RU, rugby union;. Grey bars = G Allele; Black bars = A allele.

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 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%, $P \leq 0.04$, Fig. 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%, $P < 0.05$, Fig.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.

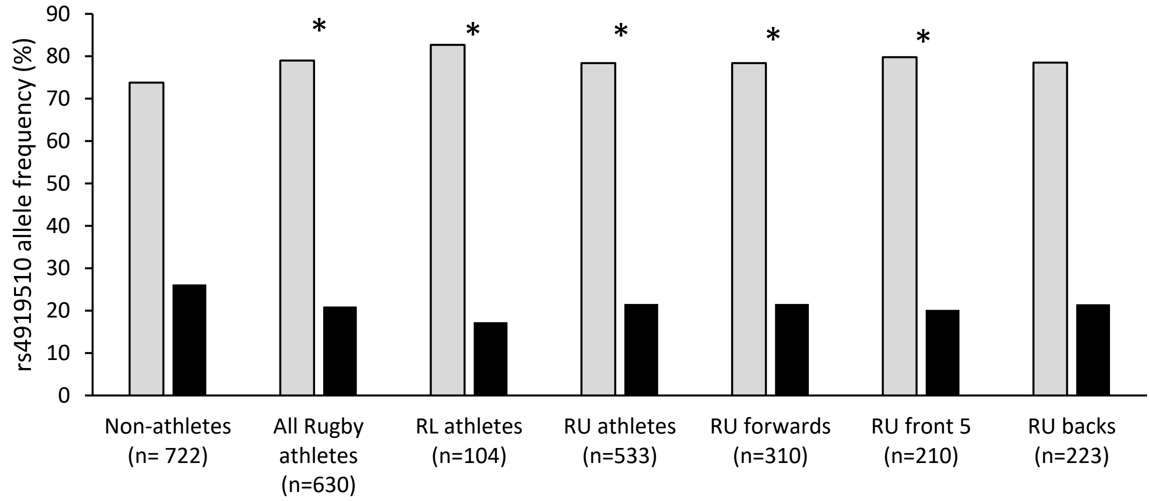

Figure 2. Allele frequency of *MIR608* rs4919510 for non-athlete and athlete groups. Asterisks (*) indicate a difference in allele frequency between the particular athlete group or sub-group and non-athletes ($P \leq 0.03$). RL, rugby league; RU, rugby union. Grey bars = C allele; Black bars = G allele.

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 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.[32] that found associations between *COL5A1* rs12722 and *COL5A1* rs3196378 polymorphisms and elite rugby status. As hypothesized, 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.[28] 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.[29] Specifically, hydroxylysine can be modified by the transfer of galactose by galactosyltransferases.[29] These posttranslational modifications might be important in the aetiology of connective tissue disorders due to the production of defective collagen modifying enzymes.[29] 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.[30] 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.[30] 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. MicroRNAs (miRNA) are a class of small non-coding RNAs that induce gene silencing and translational repression.[31] 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.[36] One of the miRNA family, *MIR608* rs4919510, has been associated with altered risk of Achilles tendinopathy.[15, 28, 37] However, the evidence is not consistent regarding which genotype is injury-protective. Abrahams et al.[37] 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.[28] However, the only other study to date could not replicate these results, but found an association between the CG genotype and Achilles tendon rupture.[15] 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.[13, 14] Furthermore, when covariates were removed, weak supporting evidence was found in a GWAS (P =0.03).[28] More recent evidence on elite female athletes from high-risk team sports found no associations between rs1800255 and ACL rupture.[38] 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,[39] 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 remodeling of connective tissues, controlling ECM homeostasis via proteolytic activity.[40] There is a possible relationship between the CC genotype of rs591058, GG genotype of rs679620, AA genotype of rs650108, and Achilles tendinopathy.[19] However, these findings have not been replicated within a Caucasian population,[20] or for ACL injury.[22] 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.[32] 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 sub-groups 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[32] 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 personalized 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.

Acknowledgments

The authors wish to thank all athletes, their respective scientific support staff and the control participants 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. Brooks JHM, Kemp SPT. Recent trends in rugby union injuries. *Clin Sports Med* 2008;**27**:51.
2. Brazier J, Antrobus M, Stebbings GK, et al. Anthropometric and physiological characteristics of elite male rugby athletes. *J Strength Cond Res*. 2020;**34**:1790-1801. doi:10.1519/JSC.0000000000002827
3. Williams S, Trewartha G, Kemp S, Stokes K. A meta-analysis of injuries in senior men's professional rugby union. *Sports Med*. 2013;**43**:1043-1055. doi:10.1007/s40279-013-0078-1
4. King D, Gissane C, Clark T, Marshall S. The incidence of match and training injuries in rugby League: A pooled data analysis of published studies. *Int J Sports Sci Coach*. 2014;**9**:417-431.
5. Fuller CW, Taylor A, Kemp SPT, Raftery M. Rugby world cup 2015: World rugby injury surveillance study. *Br J Sports Med*. 2017;**51**:51-57. doi:10.1136/bjsports-2016-096275
6. Brazier J, Antrobus M, Stebbings GK, et al. Tendon and ligament injuries in elite rugby: The potential genetic influence. *Sports (Basel, Switzerland)*. 2019;**7**:138. doi:10.3390/sports7060138
7. Magnusson K, Turkiewicz A, Hughes V, Frobell R, Englund M. High genetic contribution to anterior cruciate ligament rupture: Heritability ~69. *Br J Sports Med*. 2020;doi:10.1136/bjsports-2020-102392
8. Frank CB. Ligament structure, physiology and function. *J Musculoskelet Neuronal Interact*. 2004;**4**:199.
9. Riley GP, Harrall RL, Constant CR, Chard MD, Cawston TE, Hazleman BL. Glycosaminoglycans of human rotator cuff tendons: changes with age and in chronic rotator cuff tendinitis. *Ann Rheum Dis*. 1994;**53**:367-376. doi:10.1136/ard.53.6.367
10. Banos CC, Thomas AH, Kuo CK. Collagen fibrillogenesis in tendon development: current models and regulation of fibril assembly. *Birth Defects Res C Embryo today : reviews*. 2008;**84**:228-244. doi:10.1002/bdrc.20130
11. Khoschnau S, Medicinska och farmaceutiska v, Uppsala u, et al. Type I collagen alpha1 Sp1 polymorphism and the risk of cruciate ligament ruptures or shoulder dislocations. *Am J Sports Med*. 2008;**36**:2432.
12. Posthumus M, September AV, Keegan M, et al. Genetic risk factors for anterior cruciate ligament ruptures: COL1A1 gene variant. *Br J Sports Med*. 2009;**43**:352-356. doi:10.1136/bjsm.2008.056150
13. Stępień-Słodkowska M, Ficek K, Maciejewska-Karłowska A, et al. Overrepresentation of the COL3A1 AA genotype in Polish skiers with anterior cruciate ligament injury. *Biol sport*. 2015;**32**:143-147. doi:10.5604/20831862.1144416
14. O'Connell K, Knight H, Ficek K, et al. Interactions between collagen gene variants and risk of anterior cruciate ligament rupture. *Eur J Sport Sci*. 2015;**15**:341-350. doi:10.1080/17461391.2014.936324
15. Brown KL, Seale KB, El Khoury LY, et al. Polymorphisms within the COL5A1 gene and regulators of the extracellular matrix modify the risk of Achilles tendon pathology in a British case-control study. *J Sports Sci*. 2017/08/03 2017;**35**:1475-1483. doi:10.1080/02640414.2016.1221524
16. Riley G. The pathogenesis of tendinopathy. A molecular perspective. *Rheumatology*. 2004;**43**:131-142. doi:10.1093/rheumatology/keg448
17. Jones GC, Corps AN, Pennington CJ, et al. Expression profiling of metalloproteinases and tissue inhibitors of metalloproteinases in normal and degenerate human Achilles tendon. *Arthritis Rheum*. 2006;**54**:832-842. doi:10.1002/art.21672
18. Dalton S, Cawston TE, Riley GP, Bayley IJ, Hazleman BL. Human shoulder tendon biopsy samples in organ culture produce procollagenase and tissue inhibitor of metalloproteinases. *Ann Rheum Dis*. 1995;**54**:571-577. doi:10.1136/ard.54.7.571
19. Raleigh SM, Van Der Merwe L, Ribbans WJ, Smith RKW, Schwellnus MP, Collins M. Variants within the MMP3 gene are associated with Achilles tendinopathy: Possible interaction with the COL5A1 gene. *Br J Sports Med*. 2009;**43**:514-520. doi:10.1136/bjsm.2008.053892
20. Gibbon A, Hobbs H, van der Merwe W, et al. The MMP3 gene in musculoskeletal soft tissue injury risk profiling: A study in two independent sample groups. *J Sports Sci*. 2017;**35**:655-662. doi:10.1080/02640414.2016.1183806

21. 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;**9**:52-55. doi:10.1016/j.mgene.2016.03.007
22. Posthumus M, Collins M, van der Merwe L, et al. Matrix metalloproteinase genes on chromosome 11q22 and the risk of anterior cruciate ligament (ACL) rupture. *Scand J Med Sci Sports*. 2012;**22**:523-533. doi:10.1111/j.1600-0838.2010.01270.x
23. Visse R, Nagase H. Matrix Metalloproteinases and Tissue Inhibitors of Metalloproteinases: Structure, Function, and Biochemistry. *Circ Res*. 2003;**92**:827-839. doi:10.1161/01.RES.0000070112.80711.3D
24. El Khoury L, Posthumus M, Collins M, Handley CJ, Cook J, Raleigh SM. Polymorphic variation within the ADAMTS2, ADAMTS14, ADAMTS5, ADAM12 and TIMP2 genes and the risk of Achilles tendon pathology: A genetic association study. *J Sci Med Sport*. 2013;**16**:493-498. doi:10.1016/j.jsams.2013.02.006
25. Petersen W, Pufe T, Zantop T, Tillmann B, Tsokos M, Mentlein R. Expression of VEGFR-1 and VEGFR-2 in degenerative Achilles tendons. *Clin Orthop Relat Res*. 2004;**420**:286-291. doi:10.1097/01.blo.0000126228.78639.76
26. Rahim M, Gibbon A, Hobbs H, et al. The association of genes involved in the angiogenesis-associated signaling pathway with risk of anterior cruciate ligament rupture. *J Orthop Res*. 2014;**32**:1612-1618. doi:10.1002/jor.22705
27. Rahim M, El Khoury L, Raleigh SM, et al. Human genetic variation, sport and exercise medicine, and Achilles tendinopathy:Role for angiogenesis-associated genes. *OMICS*. 2016;**20**:520-527.
28. Kim SK, Roos TR, Roos AK, et al. Genome-wide association screens for Achilles tendon and ACL tears and tendinopathy. *PLoS One*. 2017;**12**:e0170422. doi:10.1371/journal.pone.0170422
29. Schegg B, Hülsmeier AJ, Rutschmann C, Maag C, Hennet T. Core glycosylation of collagen is initiated by two β(1-O)galactosyltransferases. *Mol Cell Biol*. 2009;**29**:943. doi:10.1128/MCB.02085-07
30 Ho MSP, Böse K, Mokkapati S, Nischt R, Smyth N. Nidogens—Extracellular matrix linker molecules. *Microsc Res Tech*. 2008;**71**:387-395. doi:10.1002/jemt.20567
31 Matzke MA, Birchler JA. RNAi-mediated pathways in the nucleus. *Nat Rev Genet*. 2005;**6**:24-35. doi:10.1038/nrg1500
32 Heffernan SM, Kilduff LP, Erskine RM, et al. COL5A1 gene variants previously associated with reduced soft tissue injury risk are associated with elite athlete status in rugby. *BMC Genomics.* 2017;**18** (Suppl 8):820.
33. Heffernan SM, Kilduff LP, Erskine RM, et al. Association of ACTN3 R577X but not ACE I/D gene variants with elite rugby union player status and playing position. *Physiol Genomics.* 2016;**48:**196-201.
34. Lahiri DK, Nurnberger JJI. A rapid non-enzymatic method for the preparation of HMW DNA from blood for RFLP studies. *Nucleic Acids Res*. 1991;**19**:5444-5444. doi:10.1093/nar/19.19.5444
35. Cahill N, Lamb K, Worsfold P, Headey R, Murray S. The movement characteristics of English Premiership rugby union players. *J Sports Sci*. 2013;**31**:229-237. doi:10.1080/02640414.2012.727456
36. Kim J, Bartel DP. Allelic imbalance sequencing reveals that single-nucleotide polymorphisms frequently alter microRNA-directed repression. *Nat Biotechnol*. 2009;**27**:472-477. doi:10.1038/nbt.1540
37. 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. *Ann Hum Genet*. 2013;**77**:204-214. doi:10.1111/ahg.12013
38. Sivertsen EA, Haug KBF, Kristianslund EK, 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. *Am J Sports Med*. 2019;**47**:52-58. doi:10.1177/0363546518808467
39. Kluivers KB, Dijkstra JR, Heniks JCM, Lince SL, Vierhout ME, Kempen LCLTv. COL3A1 2209G>A is a predictor of pelvic organ prolapse. *Int Urogynecol J Pelvic Floor Dysfunct*. 2009;**20**:1113-1118. doi:10.1007/s00192-009-0913-y

40. Foster B, Morse CI, Onambele GL, Ahmetov II and Williams AG. Genetic Variation, Protein Composition and Potential Influences on Tendon Properties in Humans. *Open Sports Med J*. 2012;**6**:8-21. doi:10.2174/1874387001206010008

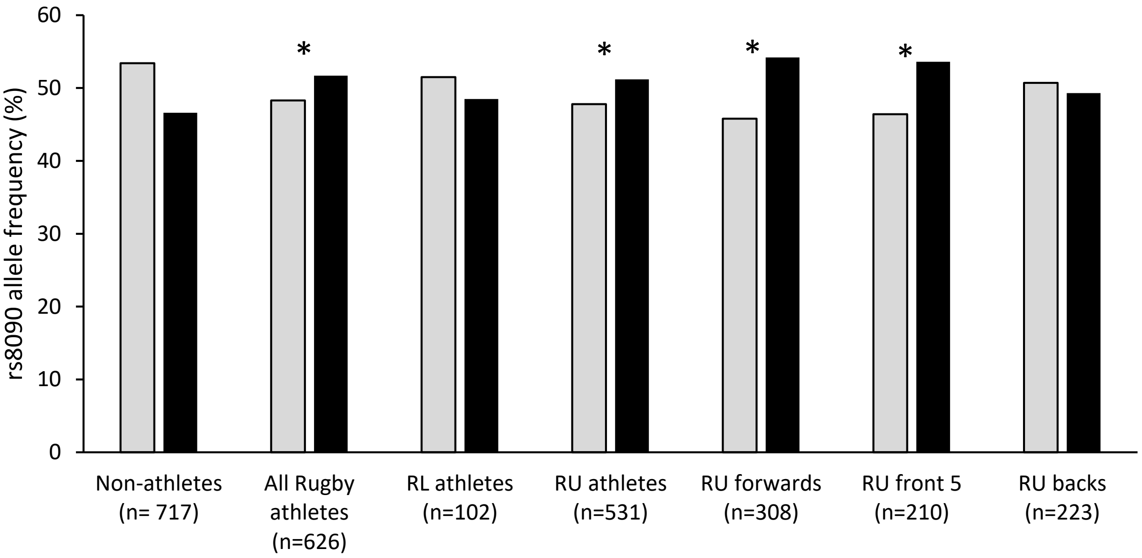

Figure 1. Allele frequency of *COLGALT1* rs8090 for non-athlete and athlete groups. Asterisks (*) indicate a difference in allele frequency between the particular athlete group or sub-group and non-athletes ($P < 0.01$). RL, rugby league; RU, rugby union;. Grey bars = G Allele; Black bars = A allele.

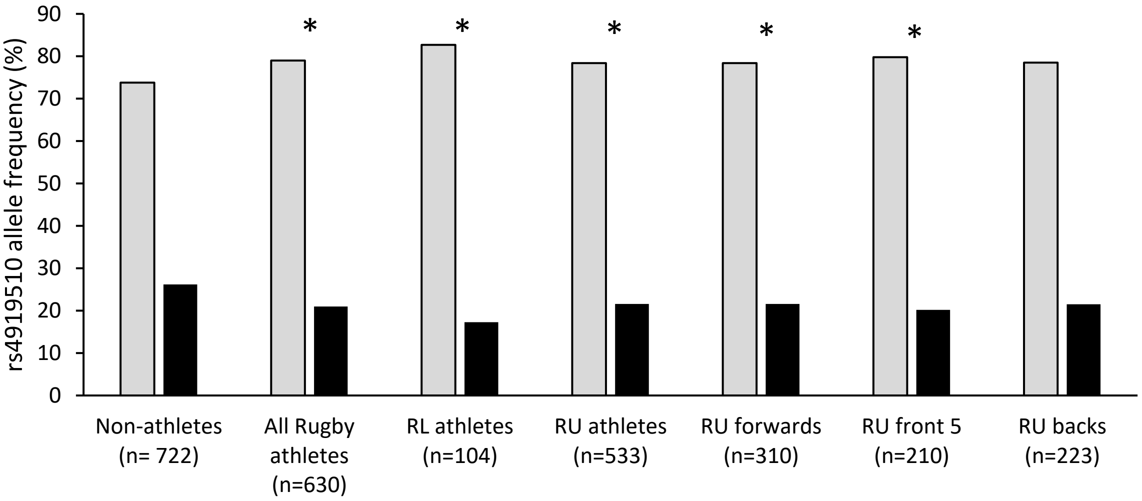

Figure 2. Allele frequency of *MIR608* rs4919510 for non-athlete and athlete groups. Asterisks (*) indicate a difference in allele frequency between the particular athlete group or sub-group and non-athletes ($P \leq 0.03$). RL, rugby league; RU, rugby union. Grey bars = C allele; Black bars = G allele.

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.

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 ($P \leq 0.03$). RL, rugby league; RU, rugby union. Grey bars = C allele; Black bars = G allele.