COLLAGEN GENE POLYMORPHISMS PREVIOUSLY ASSOCIATED WITH RESISTANCE TO SOFT-TISSUE INJURY ARE MORE COMMON IN COMPETITIVE RUNNERS THAN NON-ATHLETES

Running head: Collagen genes and running

Hannah R. Dines[1], Jennifer Nixon[1], Sarah J. Lockey[2], Adam J. Herbert[3], Courtney Kipps[4], Charles R. Pedlar[4,5], Stephen H. Day[6], Shane M. Heffernan[7], Mark R. Antrobus[1,8], Jon Brazier[1,9], Robert M. Erskine[4,10], Georgina K. Stebbings[1], Elliott C. R. Hall[1] & Alun G. Williams[1,4]

[1]*Sports Genomics Laboratory, Department of Sport and Exercise Sciences, Manchester Metropolitan University, Manchester, UK;*
[2]*Faculty of Health, Education, Medicine and Social Care, Anglia Ruskin University, Chelmsford, UK;*
[3]*School of Health Sciences, Birmingham City University, Birmingham, UK;*
[4]*Institute of Sport, Exercise and Health, University College London, London, UK;*
[5]*Faculty of Sport, Health and Applied Science, St Mary's University, Twickenham, UK;*
[6]*School of Medicine and Clinical Practice, University of Wolverhampton, Wolverhampton, UK;*
[7]*Applied Sports, Technology, Exercise and Medicine Research Centre (A-STEM), College of Engineering, Swansea University, Swansea, UK;*
[8]*Department of Sport, Exercise and Life Sciences, University of Northampton, Northampton, UK;*
[9]*Department of Psychology and Sports Sciences, University of Hertfordshire, Hatfield, UK;*
[10]*School of Sport and Exercise Sciences, Liverpool John Moores University, Liverpool, UK;*

Corresponding author details:
Dr Elliott C R Hall, PhD
Department of Sport and Exercise Sciences,
Manchester Metropolitan University,
Manchester, M1 5GD, United Kingdom
Email: elliotthall@live.co.uk

ABSTRACT

Single nucleotide polymorphisms (SNPs) of collagen genes have been associated with soft-tissue injury and running performance. However, their combined contribution to running performance is unknown. We investigated the association of two collagen gene SNPs with athlete status and performance in 1429 Caucasian participants, including 597 competitive runners (354 men, 243 women) and 832 non-athletes (490 men, 342 women). Genotyping for *COL1A1* rs1800012 (C>A) and *COL5A1* rs12722 (C>T) SNPs was performed by real-time polymerase chain reaction. The numbers of 'injury-resistant' alleles from each SNP, based on previous literature (rs1800012 A allele, rs12722 C allele), were combined as an injury-resistance score (RScore, 0 to 4; higher scores indicate injury-resistance). Genotype frequencies, individually and combined as RScore, were compared between cohorts and investigated for associations with performance via official race times. Runners had 1.34 times greater odds of being rs12722 CC homozygotes than non-athletes (19.7% *vs.* 15.5%, P=0.020) with no difference in rs1800012 genotype distribution (P=0.659). Fewer runners had RScore 0 (18.5% *vs.* 24.7%) and more had RScore 4 (0.6% *vs.* 0.3%) than non-athletes (P<0.001). Competitive performance was not associated with *COL1A1* genotype (P=0.933), *COL5A1* genotype (P=0.613) or RScore (P=0.477). Whilst not associated directly with running performance amongst competitive runners, a higher combined frequency of injury-resistant *COL1A1* rs1800012 A and *COL5A1* rs12722 C alleles in competitive runners than non-athletes suggests these SNPs may be advantageous via a mechanism that supports, but does not directly enhance, running performance.

Key words:

Sports, genomics, collagen, endurance, soft-tissue

INTRODUCTION

A cornerstone of sports genomics research is to identify genetic variants that are more common in athletes than the general population, under the assumption that specific polymorphisms contribute to the attainment of elite performance. Approximately 66% of the variance in athlete status is due to genetic factors (15), where heritable differences in anthropometry, physiological capacity, training adaptation and/or tissue structure may be advantageous to general athletic ability and/or specific sports performance (2) and may also affect injury risk (13). The latter is especially important because injuries result in training interruptions, which are detrimental to physiological adaptation. Indeed, training experience is associated with individual performance in competitive running (21), where two thirds of athletes report that their training has been interrupted by injury (30) and where training age (years) is inversely related to injury incidence (20). This suggests that the success of runners who reach elite level may, in part, be due to avoiding injury. Therefore, whilst polymorphisms associated with elite athlete status are typically assumed to augment training adaptations and/or enhance athletic performance, a lack of causal evidence means it is also possible that elite competitors are genetically more resistant to injury than most.

Ligament and tendon pathologies are amongst the most reported medical conditions in the developed world (5) and have serious implications for athletes (14). Distance running has a greater risk of overuse injury relative to other activities (8,24), with Achilles tendinopathy amongst the most commonly reported (16). Notably, evidence that similar injuries have a heritable component of ~40% (23) suggests that genetically susceptible athletes may incur more frequent suspension of training, with those at a lower risk of injury able to maximise adaptation by sustained adherence to established training principles. Furthermore, it is possible that polymorphisms affecting the structure and function of connective tissues could influence

both performance and injury phenotypes. For example, properties of the muscle-tendon unit (MTU), such as increased stiffness, are associated with a wide range of improved lower limb performance attributes (11) but may also increase injury incidence (13). It follows that inter-individual variability in tissue properties has the potential to affect athletes' likelihood of injury and impact their performance.

The critical role of collagen in ligament and tendon (33) suggests that injuries to these tissues could be influenced by variation in collagen-encoding genes. Indeed, single nucleotide polymorphisms (SNPs) of the collagen Pro-α1 (I) chain (*COL1A1*) and collagen Pro-α1 (V) chain (*COL5A1*) genes have been associated with collagenous tissue injury. *COL1A1* encodes the collagen Pro-α1 (I) chain protein, a major component of type I collagen involved in the structure of ligament and tendon, and a transversion substitution (rs1800012 C>A) affecting the binding site of the Sp1 transcription factor is associated with greater α1 (I) chain mRNA and protein production (31). Increased COL1A1 expression is proposed as a mechanistic explanation for the reduced risk of ligament injury in carriers of the A allele (17) and rare AA genotype (22,39). The relationship of rs1800012 with tendon injuries is less understood, though the A allele was also associated, as part of a haplotype, with reduced risk of Achilles tendon rupture (22). Together, this suggests that the rs1800012 CC genotype increases injury risk, with one or more copies of the A allele conferring protection.

COL5A1 encodes the collagen Pro-α1 (V) chain protein, a major component of type V collagen which regulates the diameter of collagen fibrils, and the location of a SNP (rs12722 C>T) in the 3' untranslated region suggests it could affect the level, location or timing of *COL5A1* expression (42). The rs12722 T allele has been associated with greater risk of Achilles tendinopathy (34), anterior cruciate ligament (ACL) injury (29,39), and exercise-induced

muscle cramping (36), but also with faster running times (12,38). The link between the T allele and increased *COL5A1* mRNA stability (28) underpins a hypothesis that the TT genotype influences the mechanical properties of collagenous tissue, including a reduction in the tensile strength of ligament and tendon (13). This may explain why the functional T allele and TT genotype have been associated with increased soft-tissue injury susceptibility (34,39), and the CC genotype with injury resistance.

Despite previous associations with injury and running performance, we are not aware of any studies investigating whether the frequency of *COL1A1* rs1800012 and *COL5A1* rs12722 genotypes differs between competitive runners and the general population. The polygenic nature of performance (44) and injury (35) phenotypes suggests that athletes possessing a greater number of favourable alleles from numerous SNPs have a greater chance of attaining elite status, either through enhanced performance and/or injury resistance. Furthermore, the interaction of types I and V collagen is believed to modulate collagen fibril diameter (10), suggesting that genotypes of collagen-encoding genes could also interact and affect tissue properties. Therefore, the aim of the present study was to investigate the individual and combined associations of the *COL1A1* rs1800012 and *COL5A1* rs12722 SNPs with athlete status and performance in competitive runners. Based on previous associations with reduced injury risk, and because injury avoidance reduces training interruptions (21), it was hypothesised that the rs1800012 AA and rs12722 CC genotypes would be (i) individually more frequent amongst athletes than non-athletes, (ii) observed in combination more frequently in athletes than non-athletes and (iii) associated, individually and in combination, with running performance.

METHODS

Experimental Approach to the Problem

The aim of this study was to investigate whether *COL1A1* and *COL5A1* genotypes previously associated with soft-tissue injuries are associated, individually or in combination, with running performance. Genotype analysis was conducted in a cohort of competitive runners to investigate whether genotype and allele frequencies differed from a sample of non-athletes, and whether there was a relationship between athlete genotype and performance according to official retrospective records.

Subjects

This study recruited 1429 Caucasian participants including 597 athletes (mean ± SD; age: 36 (9) years; height: 1.70 (0.24) m; mass: 60.8 (10.7) kg; 354 males and 243 females) and 832 non-athletes (age: 30 (16) years; height: 1.72 (0.17) m; mass: 73.6 (14.3) kg; 490 males and 342 females). All athletes were competitive runners primarily recruited from the London Marathon Pre-Race Registration Expo between 2012 and 2014, in addition to national/regional athletic clubs and organisations in the UK, and were 92.3 British, 1.0% Irish, and 6.7% from other nationalities. Athletes competed in a range of race distances, listing their primary competitive event as 800 m (4.4%), 1500 m (7.0%), 3000 m (2.9%), 5000 m (30.2%), 10000 m (25.6%), half-marathon (8.2%) or marathon (21.7%). Non-athletes were healthy, unrelated and recreationally active Caucasian participants including 89.3% British, 8.3% South African and 2.4% from other nationalities, and were recruited through mail-outs, posters and word of mouth. Genotyping for *COL1A1* rs1800012 and *COL5A1* rs12722 was performed in 1177 (574 athletes, 603 non-athletes) and 1159 (541 athletes, 618 non-athletes) participants, respectively, with 907 participants (518 athletes, 389 non-athletes) genotyped for both SNPs. Manchester Metropolitan University Ethics Committee approved the study, which was conducted in

accordance with the Declaration of Helsinki (2013) with written informed consent from each participant.

Running Performance

Athletes' best individual times from competitive outdoor races approved by the International Amateur Athletics Federation (IAAF) were accessed using an up-to-date collaborative Python programming code (https://bitbucket.org/KzMilligan/personalbest/src/master/www.bitbucket.com) linked to official online records (www.thepowerof10.info). For male and female athletes, times were converted into seconds and compared with the median of the 10 best official race times (Best10) recorded at the time of analysis (November 2020) by individual athletes of the same sex for each respective distance to remove the exaggerated influence of exceptional performances. To control for the fact that participating runners competed in a range of race distances, the percentage difference between each athlete's personal best time (PB) and Best10 for the same event was calculated (PB time/Best10). For example, a PB of 115.0 represents a performance that is 115.0% of (15.0% longer than) Best10 for that distance. For athletes who reported PB times for multiple race distances, the PB time closest to Best10 for that distance (i.e. with the smallest percentage difference) was selected as the athlete's best personal best (BPB) and used for subsequent analysis. Athlete inclusion criteria for the present study was achievement of a BPB performance within 60% of Best10. Information on non-athlete participants was collected by self-reported questionnaire. Accordingly, the non-athletes included for analysis had no experience in any sport above a very low competitive level.

Procedures

Blood (68%), buccal swab (23%) or saliva (9%) samples were obtained via the following protocols. Blood (5 mL) was drawn from a superficial forearm vein into an EDTA tube and stored in sterile tubes at -20°C until processed. Saliva samples were collected into Oragene DNA OG-500 collection tubes (DNA Genotek, Ottawa, Ontario, Canada) according to the manufacturer's protocol and stored at room temperature until processed. Sterile buccal swabs (Omni swab; Whatman, Springfield, Mill, UK) were rubbed against the buccal mucosa of the cheek for ~30 s. Tips were ejected into sterile tubes and stored at -20°C until processed.

DNA isolation and genotyping

DNA isolation was conducted using the Qiagen QIAcube spin protocol (Qiagen, West Sussex, UK) performed using the QIAamp DNA Blood Mini kit according to manufacturer instructions. Real-time polymerase chain reaction (PCR) genotyping was performed using a StepOnePlus (Applied Biosystems, Paisley UK) or Fluidigm EP1 system (Fluidigm, CA, USA).

StepOnePlus protocol

In 10 µl reactions, 5 µl of TaqMan GTXpress Master Mix or Genotyping Master Mix, 4.3 µl of nuclease-free H_2O (Qiagen) and 0.5 µl of TaqMan SNP assay (Applied Biosystems) were mixed with participant DNA solution and pipetted into individual wells of a 96-well plate. Adjustments were made to volumes of H_2O and DNA solution depending on sample origin, i.e. blood, saliva or buccal. Thermocycling conditions using Genotyping Master Mix were 95°C for 10 min then 40 cycles of 15 s at 95°C and 1 min at 60°C. Using GTXpress Master Mix, conditions were 20 s at 95°C then 50 cycles of 3 s at 95°C and 20 s at 60°C. Allele-specific probes identified by VIC/FAM dyes identified alleles of interest on the forward strand for *COL5A1* rs12722 as 5'-CACACCCA[C/T]GCGCCCCG-3' and for *COL1A1* rs1800012 as:

5'-CGCCC[A/C]CATTC-3'. Genotyping calls were performed using StepOnePlus analysis software version 2.3. All samples were genotyped in duplicate and each run contained three positive controls and one negative control. All duplicates were in 100% agreement.

Fluidigm Protocol

Analysis was performed using a Fluidigm EP1 system and Biomark 192.24 GT Dynamic Array Integrated Fluid Circuit (IFC) (Fluidigm, Cambridge, UK). A mix of 2 μL GTXpress Master Mix (Applied Biosystems), 0.2 μL Fast GT Sample Loading Reagent (Fluidigm), 0.2 μL nuclease-free H_2O and 1.6 μL purified DNA (or H_2O as control) was pipetted into each well of the IFC. Additionally, 1.78 μL TaqMan SNP Genotyping assay (Applied Biosystems), 1.78 μL Assay Loading Reagent (Fluidigm) and 0.18 μL ROX (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 before PCR was performed using a real-time FC1 Cycler (Fluidigm) GT 192X24 Fast v1 protocol. Thermocycling was 95°C for 120 s before 45 cycles of 2 s at 95°C and 20 s at 60°C. End-point analysis was completed using Fluidigm SNP Genotyping Analysis software version 4.5.1. All samples were assayed in duplicate and there was 100% genotype agreement between duplicates.

Injury Resistance Score (RScore)

Based on previous literature (17,22,29,34,39), an injury-resistance score (RScore) was calculated for each participant from their rs1800012 and rs12722 genotypes. Specifically, the genotypes associated with injury resistance (rs1800012 AA, rs12722 CC) each possess two copies of the protective allele and were considered most injury resistant. Accordingly, a maximum RScore of 4 was only possible in participants of rs1800012 AA and rs12722 CC genotype. A linear trend was applied to the remaining genotypes so that the heterozygous

genotypes for each SNP (rs1800012 CA, rs12722 TC) were scored 1, and the homozygous genotypes with no injury-resistant alleles (rs1800012 CC, rs12722 TT) scored 0. Accordingly, the five possible RScores were 0, 1, 2, 3 and 4 based on the number of protective alleles possessed when combining both SNPs, and a superior RScore indicates lower risk of injury.

Statistical Analysis

Genotype distributions and allele frequencies were compared between athletes and non-athletes by χ^2 goodness-of-fit tests with Bonferroni adjustment to control for multiple comparisons. The χ^2 test of independence was used to determine Hardy-Weinberg equilibrium, to investigate whether genotype frequencies differed between males and females in either cohort, and to investigate the association of RScore with athletic status. Odds ratios (OR) were calculated when genotype or RScore frequencies differed significantly between groups. Differences in athlete and non-athlete characteristics were investigated using independent samples t-test. One-way analysis of variance was used to investigate whether athletes' BPB was associated with individual genotypes or RScore. All statistical analyses were performed using R (version 3.5.1). *P* values < 0.05 were considered statistically significant. All data are presented as mean (standard deviation).

RESULTS

Participant characteristics are presented in Table 1. In males, athletes were ~7 years older (36 (9) and 29 (5) years, respectively; $P < 0.001$) and ~13 kg lighter (66.1 (8.9) and 78.9 (12.9) kg, respectively; $P < 0.001$) than non-athletes. In females, athletes were ~3 years older (35 (10) and 32 (18) years, respectively; $P = 0.017$) and ~13 kg lighter (53.1 (8.2) and 66.0 (12.9) kg, respectively; $P < 0.001$) than non-athletes.

Genotype distributions met Hardy-Weinberg equilibrium for *COL1A1* and *COL5A1* in both athletes and non-athletes ($P \geq 0.184$). No differences in genotype frequency were observed between between male and female athletes ($P \geq 0.679$) or non-athletes ($P \geq 0.663$). The *COL5A1* CC genotype was more common in athletes than non-athletes (19.8% *vs.* 15.5%, OR = 1.34, $P = 0.020$, Figure 1), with no difference in *COL1A1* genotype distribution between athletes and non-athletes ($P = 0.659$, Figure 1). In participants genotyped for both SNPs (n = 907), athletes had lower odds of having an RScore 0 (18.5% *vs.* 24.7%, OR = 1.45) and higher odds of having an RScore of 4 (0.6% *vs.* 0.3%, OR = 2.01) than non-athletes ($P < 0.001$, Figure 2). Athletes' BPB was not associated with *COL1A1* genotype ($P = 0.933$), *COL5A1* genotype ($P = 0.613$) or RScore ($P = 0.477$, Table 2).

TABLE 1 ABOUT HERE

TABLE 2 ABOUT HERE

FIGURE 1 ABOUT HERE

FIGURE 2 ABOUT HERE

DISCUSSION

The aim of this study was to investigate whether the injury-resistant *COL1A1* rs1800012 and *COL5A1* rs12722 genotypes were associated with the athletic status of competitive runners, and whether these SNPs are, individually or in combination, associated with performance. The main findings were that athletes had 1.45 greater odds of having at least one injury-resistant allele than non-athletes. However, neither individual SNPs nor RScore were directly related to athlete performance. Athletes had 1.34 greater odds of being *COL5A1* CC genotype than non-athletes, with no difference in *COL1A1* genotype distribution according to athlete status. These findings suggest that competitive runners have ~2 times greater odds of possessing the

combination of injury-resistant *COL1A1* AA and *COL5A1* CC genotypes than non-athletes, and that the *COL5A1* CC genotype may be advantageous to competitive runners via a mechanism independent of performance enhancement *per se*. Nonetheless, the limited number of athletes with the optimal combination of genotypes suggests other factors, including additional beneficial genetic variants, probably contribute more to elite athlete status.

The rs1800012 AA and rs12722 CC genotype combination was more common in athletes than non-athletes. Previous associations of each genotype with lower risk of injury (17,22,29,32,34) informed our hypothesis that possessing both genotypes would offer the greatest injury resistance and be advantageous for competitive running, because fewer training interruptions would allow athletes to maximise physiological adaptation through sustained adherence to established training principles (16). However, less than 1% of participants genotyped for both SNPs, and only 0.6% of athletes, had the optimal combination of rs1800012 and rs12722 genotypes. Therefore, whilst the likelihood of possessing both genotypes is limited by a low rs1800012 minor allele frequency, the majority of competitive runners in the present study had one or more *COL1A1* and/or *COL5A1* 'risk' alleles. Although it appears that possessing favourable genotypes at both rs1800012 and rs12722 may be advantageous to a small proportion of athletes, alternative genotype combinations at these SNPs do not appear to hinder the ability of most distance running athletes. This suggests that the combined influence of these 'injury resistant' *COL1A1* and *COL5A1* genotypes on running athlete status is small.

Higher frequency of the *COL5A1* CC genotype in runners than non-athletes suggests this genotype may be advantageous. Indeed, the CC genotype was previously associated with a lower incidence of exercise-associated muscle cramping in ironman and ultra-marathon athletes (36) and with elite rugby status (25). We hypothesised that the rs12722 CC genotype

would be more common in athletes than non-athletes due an apparent protective effect against soft-tissue injuries (29,32,34), which are common in distance runners (8,24). In non-athletes, one study found greater quadriceps muscle-tendon stiffness in T allele carriers than CC homozygotes (26), whilst another found no association with the volume or elastic modulus of the patellar tendon (19). However, whether *COL5A1* genotypes affect the mechanical properties of athletes' tendons and other tendons involved in running, such as the triceps-surae tendon (18), remains unknown. Without injury records we cannot confirm that overrepresentation of the CC genotype in athletes is due to an injury-protective mechanism. Nevertheless, because exercise-associated muscle cramping (36) and the risk of ligament (29), tendon (34) and skeletal muscle injuries (32) are shown to be lower for CC homozygotes, we support Heffernan and colleagues (25) in proposing that soft-tissue injury resistance is likely to contribute to the association of rs12722 with athlete status.

No association was observed between *COL1A1* and *COL5A1* genotypes or RScore with athletes' BPB, indicating that neither SNP, individually or in combination, influences the competitive performance of distance runners. Despite associations with ligament injury (17), tendon injury (22), muscle strength (6) and bone mineral density (45), we are unaware of literature investigating *COL1A1* rs1800012 and athletic performance. Some (12,38), but not others (9), report associations of *COL5A1* rs12722 with running ability. Specifically, TT homozygotes ran fastest during an ironman triathlon (38) and completed an ultra-marathon faster (12) than CC and CT genotypes, but neither running economy nor VO_2max were related to rs12722 genotype (9). In the ultra-marathon cohort, the T allele was associated with reduced range of motion and better performance (12), with authors suggesting that this supports the relationship between muscle-tendon stiffness and running economy (4) as a mechanistic link to endurance running. Although we found no association of rs12722 with performance, it is

possible that the CC genotype is advantageous for longer duration running events such as an ironman or ultra-marathon, but not shorter events. Indeed, rs12722 was associated with performance during the marathon section of a 226 km ironman triathlon (38) and a 56 km ultra-marathon (12), although not with shorter events in our investigation, nor with non-athletes' running economy during 10 min of treadmill running (9). Theoretically, increased muscle-tendon stiffness in TT homozygotes as proposed by Collins and colleagues (13) may enhance force transmission during ground contact, minimising changes in muscle fibre length and thus energy cost (18), which could offset or delay muscular fatigue. Therefore, *COL5A1* variants might affect the contribution of collagenous tissue to running mechanics (43) in longer duration events.

Whilst not directly associated with performance within runners, our data demonstrate that the *COL5A1* CC genotype is nevertheless more common in competitive runners than non-athletes. Some SNPs have been associated with running performance (40,41), with others demonstrating no relationship (37) or being solely associated with endurance athlete status (1). Athlete status is highly heritable (3,15) and elite athletes in particular are defined as such due to the achievement of elite performance. However, it remains possible that the association of polymorphisms with athlete status may not necessarily be due to performance-enhancing benefits. Indeed, most participants with the optimal combination of injury-protective *COL1A1* and *COL5A1* genotypes (RScore 4) were athletes, but neither genotype was associated directly with performance, and the *COL5A1* CC genotype was more frequent in athletes than non-athletes yet was also unrelated directly to performance. This suggests that the *COL5A1* CC genotype provides an advantage to runners, that is enhanced in combination with the rare *COL1A1* AA genotype, which could be due to the influence of both SNPs on the mechanical properties of collagenous tissue and their relationship to injury resistance. It is also pertinent

to consider the influence of environmental factors on athletic performance and injury resistance, including training periodisation, recovery strategies and nutritional support. Adherence to these established principles also contributes to the attainment of athlete status, and adherence is likely to be strongest when injuries are less frequent and less severe. Thus, a greater frequency of the *COL5A1* CC genotype in competitive runners suggests that the rs12722 SNP may be one of many genes that contribute to the heritability of athlete status (3,15).

This study is the first to investigate the combined association of *COL1A1* and *COL5A1* SNPs with performance and athlete status in runners. However, we acknowledge some limitations. Firstly, a lack of injury records means we are unable to evaluate directly whether genotype and RScore link to the athlete status extended phenotype via the intermediate phenotype of injury resistance. We also recognise that our association data would be complemented by measures of range of motion, flexibility and tendon properties to provide greater context regarding potential underlying mechanisms, particularly considering the previous relationship between *COL5A1* genotype, flexibility and running performance (12,38). Despite restricting our sample to Caucasian participants and recording data on nationality, we recognise that population stratification may influence our results, whereby the genotype differences between groups might also be influenced by geographic ancestry. It is also acknowledged that the candidate-gene approach cannot quantify the polygenic interaction of multiple genes, nor the possibility that epigenetic modifications at specific polymorphisms may also be important to physical performance and/or injury resistance, and we consequently advocate genome- and epigenome-wide association studies in order to address circumvent such constraints. Finally, low frequency of the rs1800012 A allele means that only a small

proportion of the several hundred athletes in our investigation were AA homozygotes, which may limit the ability to detect associations with athlete status and/or performance.

PRACTICAL APPLICATIONS

Competitive runners appear more likely to possess a beneficial combination of injury-resistant *COL1A1* rs1800012 AA and *COL5A1* rs12722 CC genotypes than non-athletes. With neither SNP nor any combination of genotypes associated with running performance directly, our findings suggest that variants of collagen-encoding genes may offer an advantage to runners that is not itself performance-enhancing. In line with previous work, we suggest that this combination of genotypes may be associated with greater injury-resistance in competitive runners and, thus, continued sporting competition via more sustained adherence to established training principles.

ACKNOWLEDGEMENTS

The authors thank all subjects for their participation.

REFERENCES

1. Ahmetov I, Kulemin N, Popov D, Naumov V, Akimov E, Bravy Y, Egorova E, Galeeva A, Generozov E, Kostryukova E, Larin A, Mustafina Lj, Ospanova E, Pavlenko A, Starnes L, Żmijewski P, Alexeev D, Vinogradova O, Govorun V. Genome-wide association study identifies three novel genetic markers associated with elite endurance performance. Biol Sport. 2015 Mar;32(1):3-9.

2. Ahmetov II, Fedotovskaya ON. Current Progress in Sports Genomics. Adv Clin Chem. 2015;70:247-314. doi: 10.1016/bs.acc.2015.03.003.

3. Andero J, Saulière G, Marck A, Toussaint JF. A Medal in the Olympics Runs in the Family: A Cohort Study of Performance Heritability in the Games History. Front Physiol. 2018 Sep 18;9:1313. doi: 10.3389/fphys.2018.01313. PMID: 30283357; PMCID: PMC6157334.

4. Arampatzis A, De Monte G, Karamanidis K, Morey-Klapsing G, Stafilidis S, Brüggemann GP. Influence of the muscle-tendon unit's mechanical and morphological properties on running economy. J Exp Biol. 2006 Sep;209(Pt 17):3345-57.

5. Baar K. Minimizing Injury and Maximizing Return to Play: Lessons from Engineered Ligaments. Sports Med. 2017 Mar;47(Suppl 1):5-11.

6. Baumert P; G-REX Consortium, Stewart CE, Lake MJ, Drust B, Erskine RM. Variations of collagen-encoding genes are associated with exercise-induced muscle damage. Physiol Genomics. 2018 Sep 1;50(9):691-693.

7. Bell RD, Shultz SJ, Wideman L, Henrich VC. Collagen gene variants previously associated with anterior cruciate ligament injury risk are also associated with joint laxity. Sports Health. 2012 Jul;4(4):312-8.

8. Bertola IP, Sartori RP, Corrêa DG, Zotz TG, Gomes AR. Profile of injures prevalence in athletes who participated in SESC Triathlon Caiobá-2011. Acta Ortop Bras. 2014;22(4):191-6.

9. Bertuzzi R, Pasqua LA, Bueno S, Lima-Silva AE, Matsuda M, Marquezini M, Saldiva PH. Is the COL5A1 rs12722 gene polymorphism associated with running economy? PLoS One. 2014 Sep 4;9(9):e106581.

10. Birk DE, Fitch JM, Babiarz JP, Doane KJ, Linsenmayer TF. Collagen fibrillogenesis in vitro: interaction of types I and V collagen regulates fibril diameter. J Cell Sci. 1990 Apr;95 (Pt 4):649-57.

11. Brazier J, Maloney S, Bishop C, Read PJ, Turner AN. Lower Extremity Stiffness: Considerations for Testing, Performance Enhancement, and Injury Risk. J Strength Cond Res. 2019 Apr;33(4):1156-1166.

12. Brown JC, Miller CJ, Posthumus M, Schwellnus MP, Collins M. The COL5A1 gene, ultra-marathon running performance, and range of motion. Int J Sports Physiol Perform. 2011 Dec;6(4):485-96.

13. Collins M, Posthumus M. Type V collagen genotype and exercise-related phenotype relationships: a novel hypothesis. Exerc Sport Sci Rev. 2011 Oct;39(4):191-8.

14. Collins M, September AV, Posthumus M. Biological variation in musculoskeletal injuries: current knowledge, future research and practical implications. Br J Sports Med. 2015 Dec;49(23):1497-503.

15. De Moor MH, Spector TD, Cherkas LF, Falchi M, Hottenga JJ, Boomsma DI, De Geus EJ. Genome-wide linkage scan for athlete status in 700 British female DZ twin pairs. Twin Res Hum Genet. 2007 Dec;10(6):812-20.

16. Ericsson KA. Training history, deliberate practice and elite sports performance: an analysis in response to Tucker and Collins review--what makes champions? Br J Sports Med. 2013 Jun;47(9):533-5.

17. Ficek K, Cieszczyk P, Kaczmarczyk M, Maciejewska-Karłowska A, Sawczuk M, Cholewinski J, Leonska-Duniec A, Stepien-Slodkowska M, Zarebska A, Stepto NK, Bishop DJ, Eynon N. 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.

18. Fletcher JR, Esau SP, MacIntosh BR. Changes in tendon stiffness and running economy in highly trained distance runners. Eur J Appl Physiol. 2010 Nov;110(5):1037-46.

19. Foster BP, Morse CI, Onambele GL, Williams AG. Human COL5A1 rs12722 gene polymorphism and tendon properties in vivo in an asymptomatic population. Eur J Appl Physiol. 2014;114(7):1393-402.

20. Francis P, Whatman C, Sheerin K, Hume P, Johnson MI. The Proportion of Lower Limb Running Injuries by Gender, Anatomical Location and Specific Pathology: A Systematic Review. J Sports Sci Med. 2019 Feb 11;18(1):21-31.

21. Fredericson M, Misra AK. Epidemiology and aetiology of marathon running injuries. Sports Med. 2007;37(4-5):437-9.

22. Gibbon A, Raleigh SM, Ribbans WJ, Posthumus M, Collins M, September AV. Functional COL1A1 variants are associated with the risk of acute musculoskeletal soft tissue injuries. J Orthop Res. 2020 Oct;38(10):2290-2298.

23. Hakim AJ, Cherkas LF, Spector TD, MacGregor AJ. Genetic associations between frozen shoulder and tennis elbow: a female twin study. Rheumatology (Oxford). 2003 Jun;42(6):739-42.

24. Hauret KG, Bedno S, Loringer K, Kao TC, Mallon T, Jones BH. Epidemiology of Exercise- and Sports-Related Injuries in a Population of Young, Physically Active Adults: A Survey of Military Servicemembers. Am J Sports Med. 2015 Nov;43(11):2645-53.

25. Heffernan SM, Kilduff LP, Erskine RM, Day SH, Stebbings GK, Cook CJ, Raleigh SM, Bennett MA, Wang G, Collins M, Pitsiladis YP, Williams AG. COL5A1 gene variants previously associated with reduced soft tissue injury risk are associated with elite athlete status in rugby. BMC Genomics. 2017 Nov 14;18(Suppl 8):820.

26. Kirk EA, Moore CW, Chater-Diehl EJ, Singh SM, Rice CL. Human COL5A1 polymorphisms and quadriceps muscle-tendon mechanical stiffness in vivo. Exp Physiol. 2016 Dec 1;101(12):1581-1592.

27. Kubo K, Miyazaki D, Shimoju S, Tsunoda N. Relationship between elastic properties of tendon structures and performance in long distance runners. Eur J Appl Physiol. 2015 Aug;115(8):1725-33.

28. 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 Biol. 2011 Jun;30(5-6):338-45.

29. Lulińska-Kuklik E, Rahim M, Domańska-Senderowska D, Ficek K, Michałowska-Sawczyn M, Moska W, Kaczmarczyk M, Brzeziański M, Brzeziańska-Lasota E, Cięszczyk P, September AV. Interactions between *COL5A1* Gene and Risk of the Anterior Cruciate Ligament Rupture. J Hum Kinet. 2018 Jun 13;62:65-71.

30. Lysholm J, Wiklander J. Injuries in runners. Am J Sports Med. 1987 Mar-Apr;15(2):168-71.

31. Mann, V., Hobson, E. E., Li, B., Stewart, T. L., Grant, S. F., Robins, S. P., Aspden, R. M., & Ralston, S. H. (2001). A COL1A1 Sp1 binding site polymorphism predisposes to osteoporotic fracture by affecting bone density and quality. *The Journal of clinical investigation*, *107*(7), 899–907.

32. Massidda M, Bachis V, Corrias L, Piras F, Scorcu M, Calò CM. Influence of the COL5A1 rs12722 on musculoskeletal injuries in professional soccer players. J Sports Med Phys Fitness. 2015 Nov;55(11):1348-53.

33. McKee TJ, Perlman G, Morris M, Komarova SV. Extracellular matrix composition of connective tissues: a systematic review and meta-analysis. Sci Rep. 2019 Jul 22;9(1):10542.

34. 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.

35. Nell EM, van der Merwe L, Cook J, Handley CJ, Collins M, September AV. The apoptosis pathway and the genetic predisposition to Achilles tendinopathy. J Orthop Res. 2012 Nov;30(11):1719-24.

36. O'Connell K, Posthumus M, Schwellnus MP, Collins M. Collagen genes and exercise-associated muscle cramping. Clin J Sport Med. 2013 Jan;23(1):64-9.

37. Papadimitriou ID, Lockey SJ, Voisin S, Herbert AJ, Garton F, Houweling PJ, Cieszczyk P, Maciejewska-Skrendo A, Sawczuk M, Massidda M, Calò CM, Astratenkova IV, Kouvatsi A, Druzhevskaya AM, Jacques M, Ahmetov II, Stebbings GK, Heffernan S, Day SH, Erskine R, Pedlar C, Kipps C, North KN, Williams AG, Eynon N. No association between ACTN3 R577X and ACE I/D polymorphisms and endurance running times in 698 Caucasian athletes. BMC Genomics. 2018 Jan 3;19(1):13.

38. Posthumus M, Schwellnus MP, Collins M. The COL5A1 gene: a novel marker of endurance running performance. Med Sci Sports Exerc. 2011 Apr;43(4):584-9.

39. Posthumus M, September AV, Keegan M, O'Cuinneagain D, Van der Merwe W, Schwellnus MP, Collins M. Genetic risk factors for anterior cruciate ligament ruptures: COL1A1 gene variant. Br J Sports Med. 2009 May;43(5):352-6.

40. Rivera MA, Fahey TD, López-Taylor JR, Martínez JL. The Association of Aquaporin-1 Gene with Marathon Running Performance Level: a Confirmatory Study Conducted in Male Hispanic Marathon Runners. Sports Med Open. 2020 Mar 20;6(1):16.

41. Stebbings GK, Williams AG, Herbert AJ, Lockey SJ, Heffernan SM, Erskine RM, Morse CI, Day SH. TTN genotype is associated with fascicle length and marathon running performance. Scand J Med Sci Sports. 2018 Feb;28(2):400-406.

42. Tabor HK, Risch NJ, Myers RM. Candidate-gene approaches for studying complex genetic traits: practical considerations. Nat Rev Genet. 2002 May;3(5):391-7.

43. Werkhausen A, Cronin NJ, Albracht K, Paulsen G, Larsen AV, Bojsen-Møller J, Seynnes OR. Training-induced increase in Achilles tendon stiffness affects tendon strain pattern during running. PeerJ. 2019 Apr 24;7:e6764.

44. Williams AG, Folland JP. Similarity of polygenic profiles limits the potential for elite human physical performance. J Physiol. 2008 Jan 1;586(1):113-21.

45. Zhang LQ, Liu H, Huang XF. Relation of JAGGED 1 and collagen type 1 alpha 1 polymorphisms with bone mineral density in Chinese postmenopausal women. Int J Clin Exp Pathol. 2014 Sep 15;7(10):7142-7.

FIGURE LEGENDS

Figure 1. Genotype distributions of **(i)** *COL1A1* rs1800012 and **(ii)** *COL5A1* rs12722 in non-athletes and athletes. *CC genotype 4.2% more common and TT genotype 2.6% less common in athletes than non-athletes ($P = 0.022$).

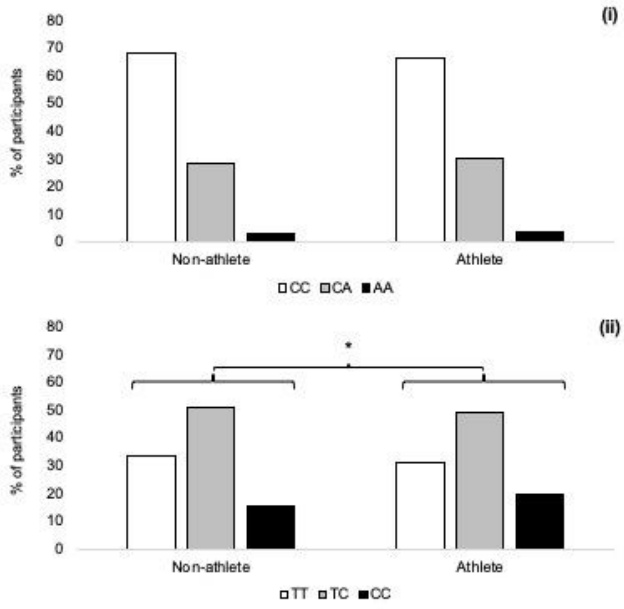
Figure 2. RScore distribution in non-athletes and athletes. *RScores 1, 2, 3 and 4 more common in athletes than non-athletes ($P = 0.001$).

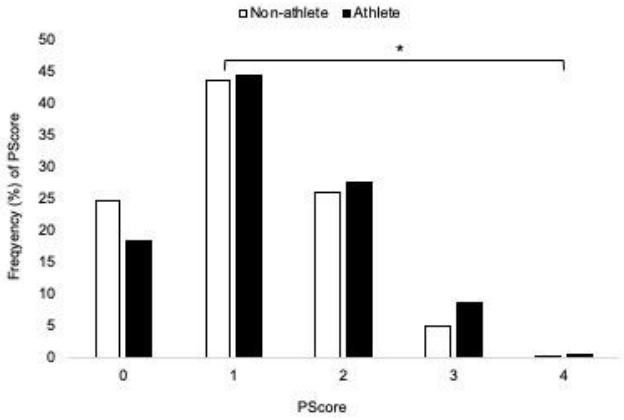
Table 1. Participant characteristics of athletes and non-athletes. Values are presented as mean ± standard deviation.

	Age (years)	**Height (m)**	**Mass (kg)**
Athletes			
Male	36 (9)	1.75 (0.24)	66.1 (8.9)
Female	35 (10)	1.62 (0.22)	53.1 (8.2)
Non-athletes			
Male	29 (15)	1.78 (0.18)	78.9 (12.9)
Female	32 (18)	1.64 (0.11)	66.0 (12.9)

Table 2. Best personal best (BPB) performance according to individual *COL1A1* and *COL5A1* genotypes and injury-resistance score (RScore). Values are percentage (%) of Top10 presented as mean ± standard deviation.

***COL1A1* rs1800012**				
CC		CA		AA
122.3 ± 8.9		122.1 ± 9.3		121.7 ± 7.7
***COL5A1* rs12722**				
TT		TC		CC
122.7 ± 10.2		122.7 ± 8.6		121.7 ± 8.6
RScore				
0	1	2	3	4
124.3 ± 10.5	121.5 ± 8.5	122.6 ± 8.5	123.6 ± 9.0	114.6 ± 2.2


(i)
80
70
60
50
40
30
20
10
0
% of participants
Non-athlete
Athlete
CC CA AA
(ii)
*
80
70
60
50
40
30
20
10
0
% of participants
Non-athlete
Athlete
TT TC CC


Non-athlete
Athlete
*
Freqyency (%) of PScore
50
45
40
35
30
25
20
15
10
5
0
0
1
2
3
4
PScore