Tendon and ligament injury-associated gene variants in elite rugby

Jon J Brazier

2021

Tendon and ligament injury-associated gene variants in elite rugby

Jon J Brazier

A thesis submitted in partial fulfilment of the requirements of the Manchester Metropolitan University for the degree of Doctor of Philosophy

Department of Sport and Exercise Sciences

2021

Acknowledgements

To begin with, I would like to thank my director of studies Dr Alun Williams, for all the time, support and guidance provided throughout this process. His expert opinion and enthusiasm were deeply appreciated, and I will be forever grateful that I had such a superb director of studies. In addition, I would like to thank the supervisory team of; Dr Georgina Stebbings, Dr Robert Erskine, Professor Liam Kilduff and Dr Stephen Day whose expertise and advice helped throughout this thesis. Furthermore, I would like to thank everyone who aided in sample collection; Malcolm Collins, Yannis Pitsiladis, Stuart Raleigh, Mark Bennett, Shane Heffernan, Adam Herbert, Peter Callus and Sarah Lockey.

Secondly, I would like to extend my thanks to all the participants that gave up their time and DNA, without their participation this thesis and my contribution to the Rugbygene project would not have been possible. Thanks, must also go to all the supportive staff of Manchester Metropolitan University who have provided advice and guidance throughout my studies.

I also would like to thank my good friend and fellow research student, Mark Antrobus, whose humour, advice and general attitude to life were very much appreciated throughout this process.

Finally, I wish to thank my family, particularly my wife Emma, my mother, sister, and my children for all their support and encouragement throughout. I dedicate this thesis to my mother, Diana Brazier, whose constant love and support has enabled me to achieve things I never thought possible.

Table of contents

Publications

The following chapters have been published or submitted for publication:

Chapter 1:

Brazier, J, Antrobus, M.R., Stebbings, G, K., Day, S, H., Heffernan, S, M., Cross, M and Williams, A, G (2019) Tendon and ligament injuries in elite rugby: the potential genetic influence, *Sports*, 7 (6). https://doi.org/10.3390/sports7060138

Chapter 3:

Brazier, J., Antrobus, M, R., Herbert, A, J., Callus, P, C., Stebbings, G, K., Day, S, H., Heffernan, S, M., Kilduff, L, P., Bennett, M, A., Erskine, R, M., Raleigh, S, M., Collins, M., Pitsiladis, Y,P., and Williams, A, G (2021- under review) Gene variants previously associated with reduced soft tissue injury risk: part 1 – independent associations with elite status in rugby. *European Journal of Sport Science.*

Chapter 4:

Brazier, J., Antrobus, M, R., Herbert, A, J., Callus, P, C., Khanal, P., Stebbings, G, K., Day, S, H., Heffernan, S, M., Kilduff, L, P., Bennett, M, A., Erskine, R, M., Raleigh, S, M., Collins, M., Pitsiladis, Y, P., and Williams, A, G (2021- under review) Gene variants previously associated with reduced soft tissue injury risk: part 2 - polygenic associations with elite status in rugby. *European Journal of Sport Science.*

Publications associated with aspects of this thesis

Brazier, J., Antrobus, M., Stebbings, G. K., Day, S. H., Callus, P., Erskine, R. M., Bennett, M. A., Kilduff, L. P., Williams, A. G. (2020) Anthropometric and physiological characteristics of elite male rugby athletes. *Journal of Strength and Conditioning Research, 34* (6), 1790–1801.

Antrobus, M. R., **Brazier, J.**, Stebbings, G. K., Day, S. H. Heffernan, S. M., Kilduff, L. P., Erskine, R. M., Williams, A. G. Genetic factors that could affect concussion risk in elite rugby. *Sports* 2021, *9* (2), 19.

Conference proceedings

Brazier, J., Antrobus, M., Herbert, A., Stebbings, G., Day, S., Heffernan, S., Kilduff, L., Cullen, T., Raleigh, S., Erskine, R and Williams, A (2020) Tendon and ligament-related genes associate with elite status in rugby, 25th Annual Congress of the European College of Sports Science, (Seville, Spain).

Brazier, J., Antrobus, M., Herbert, A., Stebbings, G., Day, S., Heffernan, S., Erskine, R., Kilduff, L and Williams, A (2018) Association of MMP3 but not TIMP2 gene variants with elite rugby player status and rugby code. 23rd Annual Congress of the European College of Sports Science, (Dublin, Ireland).

Brazier, J., Antrobus, M., Heffernan, S., Stebbings, G., Day, S., Hennis, P., Erskine, R., Kilduff, L and Williams, A (2017) An investigation into genetic associations with musculoskeletal soft tissue injuries in elite rugby – preliminary findings. Middlesex University Strength and Conditioning Conference, (London, UK).

Related conference proceedings

Stebbings, G. K., Herbert, A. J., **Brazier, J.,** Lockey, S. J., Erskine, R. M., Day, S. H. and Williams, A. G. (2017). No association between tendon-related genes and performance in elite European caucasian marathon runners. BMJ Publishing Group Ltd and British Association of Sport and Exercise Medicine. International Sports Science and Sports Medicine Conference (Newcastle upon Tyne, UK).

Antrobus, M.R., **Brazier, J.,** Herbert, A.J., Stebbings, G.K., Day, S.H., Heffernan, S., Erskine, R.M., Kilduff, L.P. & Williams, A.G. (2019). Association between *MAPT* polymorphism but not *APOE Promoter* and elite rugby athlete status. Proceedings of the 24th Annual Congress of the ECSS, (Dublin, Ireland).

Antrobus, M.R., **Brazier, J.,** Herbert, A.J., Stebbings, G.K., Day, S.H., Heffernan, S., Erskine, R.M., Kilduff, L.P. & Williams, A.G. (2020). Association between *COMT* polymorphism but not *BDNF-AS and NOS3* in elite rugby athletes. Proceedings of the 25th Annual Congress of the ECSS, (Seville, Spain).

List of figures

List of tables

List of abbreviations

Abbreviation	Definition
3’UTR	3’ untranslated re*g*ion
α1	Alpha 1
ACAN	Aggrecan
ACL	Anterior cruciate ligament
ADAM-12	A disintegrin and metalloproteinase 12
ADAMT	A disintegrin and metalloproteinase with thrombospondin motifs
AI	All injured rugby athlete group
ANI	All non-injured rugby athlete group
AUC	Area under the curve
BH	Benjamini-Hochberg corrections
CASP8	Caspase-8
CI	Confidence intervals
Col I	Type I collagen
Col III	Type III collagen
Col V	Type V collagen
Col XII	Type XII collagen
COL1A1	Collagen type I alpha I
COL3A1	Collagen type III alpha I
COL5A1	Collagen type V alpha I
COL12A1	Collagen type XII alpha I
COLGALT1	Collagen beta (1-0) galatosyltransferase 1)
DCN	Decorin
DNA	Deoxyribonucleic acid
ECM	Extra cellular matrix
EDTA	ethylenediamine tetra acetic acid
FAM	Fluorescein dye
FBN2	Fibrillin-2
G	Gauge
GAS	Gene association studies
GDF5	Growth differentiation factor-5
GWAS	Genome-wide association studies
HWE	Hardy-Weinberg equilibrium

KDR	Kinase insert domain receptor
LR	Ligament rupture rugby athlete group
LS	Ligament sprain rugby athlete group
MCL	Medial collateral ligament
MDR	Multifactor dimensionality reduction
MIR608	MicroRNA 608
MMP	Matrix metalloproteinase
MMP3	Matrix metalloproteinase – 3
MRI	Magnetic resonance imaging
mRNA	Messenger ribonucleic acid
NA	Non-athletes
NID1	Nidogen 1
NIL	Non-injured ligament rugby athlete group
NIT	Non-injured tendon rugby athlete group
OR	Odds ratio
p	Petit (short arm of chromosome)
PCR	Polymerase chain reaction
PRISP	Professional rugby injury surveillance project
q	Queue (tail/long arm of chromosome)
RA	Rugby athletes
RL	Rugby league
RLA	Rugby league athletes
RNA	Ribonucleic acid
ROC	Receiver operating characteristic curve
ROX	6-carboxy-X-rhodamine reference dye
rpm	Revolutions per minute
rs	Reference SNP cluster identification number
RU	Rugby union
RUA	Rugby union athletes
RUB	Rugby union backs
RUF	Rugby union forwards
SA	South Africa
SD	Standard deviation
SNP	Single nucleotide polymorphisms
TNC	Tenascin C

TGS	Total genotype score
TIMP	Tissue inhibitors of metalloproteinases
T	Tendinopathy rugby athlete group
TR	Tendon rupture rugby athlete group
VEGF	Vascular endothelial growth factor
VEGFA	Vascular endothelial growth factor A
VIC	Aequorea Victoria dye
w	Cohen's w effect size for Chi square tests
χ^2	Chi square

Abstract

Elite rugby has one of the highest reported injury incidences of any professional sport. Some of the most severe of all these injuries are to the tendon and ligament. The aetiology of these injuries is highly multi-factorial, with a growing body of evidence suggesting that some of the inter-individual variability in injury susceptibility may be due to genetic variation. However, little effort has been devoted to the study of genetic injury traits within rugby athletes. Consequently, the overall aim of the present thesis was to investigate genetic associations with elite athlete status and soft-tissue injury within an elite rugby population. Genotype data was collected from 1572 participants, consisting of 663 elite rugby athletes and 909 non-athletes. *COLGALT1* rs8090, *COL3A1* rs1800255, *COL5A1* rs12722 and rs3196378, *MIR608* rs4919510, *MMP3* rs591058 and rs679620 and *NID1* rs4660148 polymorphisms were independently associated with elite status in rugby. Furthermore, when all polymorphisms (*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) were incorporated into a polygenic profile significant differences existed between elite rugby athletes and non-athletes which persisted across all sub- groups. Additionally, *COL5A1* rs12722, *MMP3* rs679620, *NID1* rs4660148 and *TIMP2* rs4789932 were associated with soft-tissue injury history in elite rugby athletes. These results identify novel genetic associations with elite status and soft-tissue injury in rugby. In conclusion, there appears to be genetic associations with elite athlete status and soft-tissue injury risk, potentially enabling career success. Further research is required to replicate the findings of this thesis in comparable and differing cohorts. Nonetheless, the work presented here has

further enhanced the current understanding of genetic associations with elite status and soft-tissue injury, which may in future have implications on training and injury management in elite rugby athletes.

Chapter 1

Introduction and Literature Review

A portion of this chapter is published in:

Brazier, J, Antrobus, M.R., Stebbings, G, K., Day, S, H., Heffernan, S, M., Cross, M and Williams, A, G (2019) Tendon and ligament injuries in elite rugby: the potential genetic influence, *Sports*, 7 (6). https://doi.org/10.3390/sports7060138

1.1 Introduction to the thesis

Elite athletic status has been shown to be a heritable trait (De Moor et al., 2007). Williams et al. (2014) highlighted that depending on the particular phenotypes required ~ 50% of the variance can be explained by genetic factors, with the remainder attributable to environmental factors. Despite this genetic influence, research has progressed at a modest pace (Williams et al., 2014). Typically, in sports genomics research, athletes from opposite ends of the physical performance spectrum such as elite power/sprint athletes and elite endurance athletes are compared with each other or with non-athletes (Eynon et al., 2009; Eynon et al., 2010; Eynon et al., 2011; Ruiz et al., 2011). The rationale being that by utilising homogenous athlete groups such as elite sprinters, it is plausible that they would possess extreme phenotypes only found at the elite level. Additionally, these elite athletes would participate at a similar standard within the same events/competitions, and it is likely that they would also have similar training routines and volume loads. Thus, enabling a reduction in confounding factors and a clear distinction to be made between cohorts studied.

Elite rugby athletes are an appealing cohort to investigate within sports genomics, as not only do they need to possess distinct physical attributes to succeed, but their elite status can be quantified relatively easily via their level of league competition played and whether they competed at international level for a 'high performance union' (Regulation 16, http://www.worldrugby.org). Elite rugby athletes undergo heavy training loads and are likely to exhibit characteristics near the limits of human physiological capability; indeed, elite rugby has one of the highest incidences of injury in sport, with tendon and ligament injuries some of the most frequent and

severe. Regular participation at the elite level in rugby would mean players have been exposed to one of the highest levels of risk for tendon and ligament injury in any professional sporting environment and, at least to some extent, been able to succeed in that sport despite that high environmental risk. This ability to recover from or withstand musculoskeletal soft tissue injury that is potentially performance-limiting or career-ending, but nevertheless achieve elite status, may be reflected in distinct genetic characteristics. However, to date, the literature regarding rugby genomics is in its infancy, with a paucity of published studies focusing on genetics and performance phenotypes (Heffernan et al., 2016; Bell et al., 2012b; Bell et al., 2010; Goh et al., 2009; Heffernan et al., 2017b; Heffernan et al., 2017a; Hall et al., 2021; Abrahams et al., 2019b; Abrahams et al., 2019a; Mc Fie et al., 2018; Abrahams et al., 2018). Additionally, only one of the aforementioned studies (Heffernan et al., 2017a) has investigated genes previously associated with musculoskeletal soft-tissue injury within an elite rugby population. As such, due to the very high injury rates within elite rugby (Brooks et al., 2005b), as well as the fact rugby athletes perform within a well-defined set of rules and parameters, elite rugby athletes are a highly suitable population to study genetic variation of soft-tissue injury risk and thus is the purpose of this thesis.

1.1.1 Thesis overview

Elite rugby has one of the highest reported injury rates of any professional sport (Brooks and Kemp, 2008a), with the majority being to the tendon, ligament and muscle (Williams et al., 2013). As such, Chapter 1 discusses and reviews the current tendon and ligament injury incidence and severity rates in elite rugby, highlighting the current disparity between reporting procedures of elite rugby union and rugby

league. Section 1.2.3 then details the risk factors for these injuries and how genetics may account in part for the inter-individual variability within these injuries.

Chapter 1.2.4 focusses on tendon and ligament pathologies, distinguishing between tendinopathy and tendon rupture, ligament tear and ligament rupture. Heritability has been estimated at 40% for tendon injury (Hakim et al., 2003) and 70% for ligament injury (Magnusson et al., 2020). Thus, a broad review of the likely genetic influence is performed in section 1.2.5. Section 1.2.6 follows with a specific and detailed evaluation of candidate genes that were studied within this thesis and provides a rationale for their inclusion.

In Chapter 2, the general methodology is described regarding how allele and genotype frequencies of elite athletes and non-athletes were investigated - essential for Chapters 3, 4 and 5. More specifically, Chapter 2 describes the sample collection and DNA isolation techniques, allele and genotyping procedures, athlete injury history data collection protocol, and statistical analyses utilised in the subsequent chapters. Because similar methods were utilised for the experimental chapters (3, 4 and 5), Chapter 2 will be cross-referenced from those latter chapters.

In Chapter 3, the main aim was to investigate whether allele and genotype frequencies of polymorphisms previously associated with tendon and ligament injury (*COLGALT1* (rs8090), *COL1A1* (rs1800012), *COL3A1* (rs1800255), *COL5A1* (rs12722), *COL5A1* (rs3196378) *KDR* (rs1870377), *MIR608* (rs4919510), *MMP3* (rs679620, rs591058 and rs650108), *NID1* (rs4660148), *TIMP2 (*rs4789932) and

VEGFA (rs699947)) differed between elite rugby athletes and non-athlete controls, and between rugby union (RU) sub-group playing positions. To achieve elite status in rugby it is plausible that these athletes may have had less interruption to their competitive careers through injury than their non-elite counterparts. Evidence of this has been recently identified in elite football, where injuries were found to be negatively associated with athlete career progression (Larruskain et al., 2021). Therefore, it was hypothesised that elite rugby athletes would have a lower frequency of alleles and genotypes associated with tendon and ligament injury risk than non-athlete controls.

In Chapter 4, tendon and ligament injury associated polygenic profiles were compared between elite rugby athletes and non-athlete controls, and between RU sub-group playing positions. This was performed via the application of three different total genotype score (TGS) algorithms, inferred haplotype analysis and multifactor dimensionality reduction modelling. It was hypothesised that elite rugby athletes would a have higher TGS and higher frequencies of the 'preferable' inferred haplotype allele combinations, indicating more genetic resistance to tendon and ligament injury.

In Chapter 5, the main aim was to investigate whether polymorphisms previously associated with tendon and ligament injury were associated with a previous history of tendon and ligament injury in elite rugby athletes. It was hypothesised that elite rugby athletes with no history of tendon and ligament injury would have a lower frequency of risk-associated alleles and genotypes compared to athletes with a history of injury.

Finally, Chapter 6 took a retrospective view of the results observed within each chapter, collated and synthesised these findings to discuss how genetic factors could impact tendon and ligament injury risk in elite rugby. This chapter also considers the implications of these findings for future research.

1.2 Literature review

1.2.1 Introduction

Due to the characteristics of the game of rugby, whereby high impact body contact frequently occurs through multiple physical collisions and tackles, musculoskeletal injuries are extremely common (Gabbett, 2004; Hoskins et al., 2006). Rugby union (RU) has one of the highest reported incidences of match injuries within professional sport, regardless of the injury definition used (Brooks and Kemp, 2008a). This is likely in part due to well-established and frequently applied injury surveillance research compared to other collision sports. Rugby league (RL) does not currently have a comparable level of injury surveillance research which limits our understanding somewhat. The majority of injuries in both RU and RL occur during tackles (Gibbs, 1993; Gabbett, 2004; Gabbett et al., 2011; Brooks et al., 2005b; Fuller et al., 2013; Fuller et al., 2017). However, numerous other causes have been documented, including but not limited to rucks, mauls, scrums (Fuller et al., 2007) and via tripping, twisting, slipping, falling, overexertion and overuse (Gabbett and Hodgson, 2003). A meta-analysis by Williams et al. (2013) reported the total incidence of injury (injuries per 1000 player hours) as 81/1000 in matches (~3 injuries per match) and 3/1000 in training in men's professional RU.

The regular occurrence of injury in RU limits competitive success. For example, Williams et al.'s (2016) recent 7-year prospective study assessing playing time loss from injury and team success in elite RU found clear negative associations between injury measures (injury burden and injury days per match) and team success (league points tally and Eurorugby Club Ranking). Thus, reductions in injury incidence and severity could enhance team success.

Due to the high incidence of injury in RU, numerous injury surveillance studies have been conducted during international competitions, particularly during five rugby World Cups from 1995 to 2015 (Best et al., 2005; Fuller et al., 2008; Fuller et al., 2013; Fuller et al., 2017), as well as single and multiple seasons for professional (Brooks et al., 2005b; Kemp et al., 2015; Kemp et al., 2016) and community level rugby (Haseler et al., 2010; Roberts et al., 2013). Although numerous injury surveillance studies have been carried out in RU, only studies from 2007 are consistent with the international consensus statement for epidemiological studies in rugby (Fuller et al., 2007c). Therefore, comparisons with earlier studies are problematic. Unfortunately, this consistency has not existed to the same degree for RL, although recent steps have been taken towards a consensus-driven approach (Fitzpatrick et al., 2018).

Injury data collection is an essential part of trying to understand the risk (incidence and severity) of participation in sport and how that risk changes over time. van Mechelen et al. (1992) designed a four-step model for injury prevention within sport. It involves (i) identifying the extent of the sports injury problem, (ii) identifying the characteristics and mechanisms that contribute to the development of injury, (iii)

introducing measures to reduce future risk and/or severity of injury, (iv) an evaluation of those measures by repeating the first step. A similar risk management model was proposed by Fuller and Drawer (2004), which aimed to identify risk factors and estimates that could be evaluated and then communicated to the sports community. Having a deeper understanding of these areas enables coaches, doctors and strength and conditioning staff to assess current practices in injury prevention, treatment and rehabilitation, and make adjustments accordingly. It also allows governing bodies to identify areas of high risk and to introduce strategies to mitigate such a risk. Finally, longitudinal injury data allows researchers to monitor the impact and effectiveness of any interventions. What is apparent from the research undertaken thus far is that injuries vary considerably in location, diagnosis and profile.

Fundamental understanding of injury mechanisms and differences in inter-individual risk begins with the genome and the biological composition of tissues that depend on co-ordinated expression of selected genes at the protein level. The aim of this literature review therefore is, firstly, to highlight the incidence and severity rates of tendon and ligament injury within elite rugby. Secondly, to discuss the biological composition of tendons and ligaments and how genomics may influence this and subsequent predisposition to injury. The steps necessary to better understand the genomic aspects of injury within elite rugby will then be considered.

1.2.2 Tendon and ligament injury incidence rates and severity in rugby

1.2.2.1 Tendon and ligament injury incidence rates in rugby

Numerous injury surveillance studies have been carried out within professional RL, with muscle/tendon and ligament/joint (non-bone) injuries consistently the two most frequent types of injury (Gibbs, 1993; Gissane et al., 1993; Seward et al., 1993; Stephenson et al., 1996; Gissane et al., 2003; King et al., 2014). In the more recent of these studies, ligament/joint (non-bone) injuries made up 25.2% and muscle/tendon 24.2% of all injuries (Gissane et al., 2003). However, the majority of professional RL studies are dated, have limited application to present day RL and inconsistent methodological approaches and definitions were used. Cross et al. (2018) demonstrated the importance of utilising consistent definitions for injury by showing that incidence of injury with a >24-hour time-loss definition was approximately double that when using a >7 day definition. For example, Gissane et al.'s (1993) injury definition was 'the onset of pain or a disability resulting from either training for or playing rugby league,' while Seward et al.'s (1993) definition was 'that which caused a player to be unavailable for selection in a match, or participation in a training session or any other injury which required medical treatment, other than routine conservative measures.' These differences provide substantially different portrayals of injury risk. When the injury definition is more exclusive and includes only more severe injuries, joint/ligament are most frequent. However, when the definition is more inclusive, muscular, head and neck injuries are most frequent (Hoskins et al., 2006). This has led to much debate on definitions of injury within RL (Orchard and Hoskins, 2007; Hodgson et al., 2007; King et al., 2009). A very recent attempt was made at a consensus-driven approach to standardise epidemiological studies in RL (Fitzpatrick et al., 2018) and these data are probably more valid than those previously reported. Three different ligament injuries were in the top five for incidence: medial collateral ligament (MCL) 3.9/1000 hours, syndesmosis 2.7/1000 hours and ankle lateral ligament 2.6/1000 hours (Fitzpatrick et al., 2018).

In RU, injury incidence rates are easier to identify than RL due to the consensus statement on injury definitions and data collection procedures for studies in RU (Fuller et al., 2007c). However, much like RL, muscle/tendon and ligament/joint (non-bone) injuries are consistently the top two most frequently occurring injury groups in elite level RU (Brooks et al., 2005b; Fuller et al., 2013; Fuller et al., 2017; Williams et al., 2013; Fuller et al., 2008; Moore et al., 2015) with more muscle/tendon injuries in backs than forwards at English Premiership and International level. For ligament/joint (non-bone) injuries, forwards appear to have more frequent occurrence at International level, while backs have more at English Premiership level (Brooks et al., 2005b; Fuller et al., 2008; Fuller et al., 2013). It should be noted, however, that these apparent differences between forwards and backs are based on data provided in the literature but not statistical testing. Table 1.1 summarises the match injury incidence of muscle/tendon and ligament/joint (non-bone) injuries from post-2007 studies where methodologies align with the consensus statement on injury definitions and data collection procedures (Fuller et al., 2007c). It is worth noting that, at World Cup competitions, although muscle/tendon injuries have a high incidence this is mainly due to muscle rather than tendon injuries (Fuller et al., 2008; Fuller et al., 2013; Fuller et al., 2017). It is likely that this also occurs in the English Premiership and Super 14 competitions but the data are not clear.

Table 1.1. Muscle/tendon and ligament/joint (non-bone) injury incidence rates in elite rugby

Study	**Level**	**Injury Type**		**Match Injuries Incidence** (Injuries/1000 player hours (95% CI))		
		Main Group	Sub-group	Forwards	Backs	All
Brooks et al. (2005b)[1,2]	English Premiership clubs[4]	Muscle / Tendon	Strain / tear / rupture	14	20	17
		Ligament / joint (non-bone)	Sprain / rupture	13	15	14
Fuller et al. (2008)	International	Muscle /Tendon	Muscle rupture / tear / strain	18 (12-29)	27 (18-40)	22 (17-30)
			Tendon rupture / tendinopathy	0	1(0.2-8)	0.5 (0.1-4)
		Ligament / joint (non-bone)	Sprain / rupture	25(17-37)	23 (15-36)	25 (18-33)
Fuller et al. (2013)	International	Muscle / Tendon	Muscle rupture / tear / strain	20 (13-30)	20 (13-32)	20 (14-27)
			Tendon rupture / tendinopathy	1 (0.1-7)	5 (2-12)	3 (1-6)
		Ligament / joint (non-bone)	Sprain / rupture	22 (14-33)	18 (11-29)	20 (14-27)
Moore et al. (2015)	International	Muscle / Tendon	Muscle strain / rupture	-	-	34 (23-49)
			Tendinopathy / rupture	-	-	9(4-18)
		Ligament / joint (non-bone)	Sprain / rupture	-	-	43 (31-61)
Fuller et al. (2017)	International	Muscle / Tendon	-	34	39	-
		Ligament / joint (non-bone)	-	25	34	
Williams et al. (2013)[3]	English Premiership clubs[4], Super 14 clubs[4], Vodacom cup 2008 clubs[4] and International	Muscle / Tendon	-	-	-	40 (21-76)
		Ligament / joint (non-bone)	-	-	-	34 (18-65)

[1]Study was before 2007 consensus statement on injury definitions and data collection procedures but uses very similar methods. [2]No confidence intervals were annotated in the study. [3]Meta-analysis of seven studies used for pooled analysis (Brooks et al., 2005a; Brooks et al., 2005b; Brooks et al., 2005c; Fuller et al., 2008; Fuller et al., 2009; Fuller et al., 2010a; Fuller et al., 2013).[4]Top tier of professional rugby competitions in England, Australia, New Zealand and South Africa.

In the English Premiership RU competition across the eight most recently reported seasons from 2011-2019, ligament injuries were consistently in the top five most common injuries (Kemp et al., 2016; Kemp et al., 2017; Kemp et al., 2018; Kemp et al., 2020), with MCL in the top five every season apart from 2015-16 and 2018-19. The Professional Rugby Injury Surveillance Project (PRISP) reports individual injuries such as MCL, hamstring or ankle lateral ligament, rather than grouping all muscle/tendon or ligament/joint (non-bone) injuries together. Outside of the top five injuries, there are no available data on further muscle/tendon and ligament/joint (non-bone) injuries, making more detailed or grouped analysis impossible. Figure 1.1 shows the top five most common match injuries in the English Premiership competition during 2011-2019, highlighting the frequency of ligament injuries.

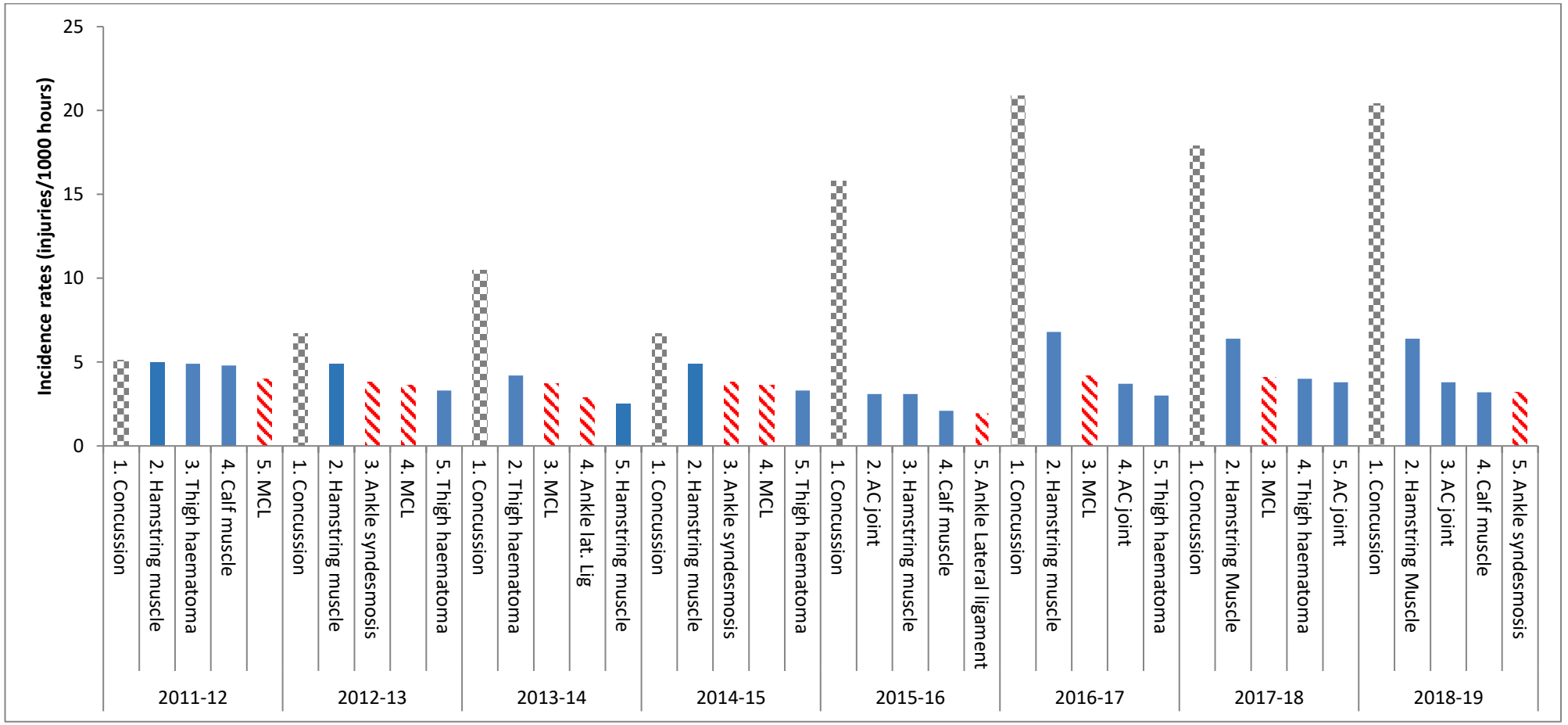

Figure 1.1. Top five most common injuries – English Premiership Rugby. Adapted from the PRISP annual reports 2011-2019 (Kemp et al., 2018; Kemp et al., 2019; Kemp et al., 2020). Key: Lined bars = Ligament injuries; Squared bars = Concussion; Filled bars = Any other injury.

1.2.2.2 Tendon and ligament injury severity and burden in rugby

The current literature is limited regarding the severity (days absence from full training or match play) of injuries at specific anatomical locations in elite RL. From the available data, Gibbs (1994) found ankle ligament tears were the most severe, followed by MCL tears and groin muscle/tendon tears. More recently, Orchard (2004) stated anterior cruciate ligament (ACL) tears were the most severe, followed by shoulder sprains and dislocations and MCL tears. That is supported by Fitzpatrick et al. (2018), although that study calculated severity from date of occurrence until date of return to full training, which differs from RU's consensus statement on injury definitions and data collection procedures (Fuller et al., 2007c) and would increase severity data. These studies suggest that ligament/joint (non-bone) and muscle/tendon injuries are the main causes of RL players missing matches, thus impairing competitive success and player wellbeing.

RU has similar but more consistent findings to RL, with muscle/tendon and ligament/joint (non-bone) injuries making up three of the top five most severe injuries for forwards; ACL, Achilles tendon and MCL injuries caused 988, 726 and 718 days absence, respectively (Brooks et al., 2005b). For backs, three of the top five most severe were hamstring muscle, MCL and ACL injuries causing 1176, 870 and 815 days absence, respectively (Brooks et al., 2005b). Knee injuries in particular (ACL and MCL) resulted in the greatest absence for forwards and backs (Brooks et al., 2005b). At the 2007 RU World Cup, muscle/tendon (mainly muscle) and ligament/joint (non-bone) were the third and fourth most severe injuries, with backs having a higher severity of both (Fuller et al., 2008) (not tested statistically). At the 2011 RU World Cup, ligament/joint (non-bone) and muscle/tendon (mainly tendon)

were the third and fourth most severe injuries with backs again having a higher severity of both (Fuller et al., 2013) (not tested statistically). Fuller et al. (2017) identified knee ligament injuries as the most severe and Achilles tendon injuries as the fourth most severe at the 2015 RU World Cup for all players. In Williams et al.'s (2013) meta-analysis, a similar pattern is seen with ligament/joint (non-bone) injuries the second most severe and muscle/tendon injuries the fourth. Table 1.2 summarises the severity of muscle/tendon and ligament/joint (non-bone) injuries for English Premiership and World Cup competitions. The large variability can be attributed to several factors such as different settings (league or cup tournament), cohort sizes and opportunities for data collection.

Table 1.2. Muscle/tendon and ligament/joint (non-bone) injury severity rates

Study	**Level**	**Injury Type**		**Match injuries Severity** (Days absence (95% CI))		
		Main Group	Sub-group	Forwards	Backs	All
Brooks et al. (2005b)[1, 2]	English Premiership clubs[5]	Muscle / Tendon	Strain / tear / rupture	44	38	23
		Ligament / joint (non-bone)	Sprain / rupture	69	91	46
Fuller et al. (2008)	International	Muscle / Tendon	Muscle rupture / tear / strain	17 (10-25)	21 (9-33)	20 (12-27)
			Tendon rupture / tendinopathy	0	4*	4*
		Ligament / joint (non-bone)	Sprain / rupture	14 (8-20)	18(9-27)	16 (11-21)
Fuller et al. (2013)	International	Muscle / Tendon	Muscle rupture /tear / strain	15 (8-23)	27 (16-38)	21 (14-28)
			Tendon rupture / tendinopathy	4*	36 (0-92)	29 (0-75)
		Ligament / joint (non-bone)	Sprain / rupture	38 (8-68)	42 (12-72)	39 (18-61)
Fuller et al. (2017)[2,3]	International	Ligament / joint (non-bone)	Knee ligament	-	-	1507
			Achilles tendon	-	-	188*
Williams et al. (2013)[4]	English Premiership clubs[5] and International	Muscle / Tendon	-	-	-	15 (5-24)
		Ligament / joint (non-bone)	-	-	-	29 (19-39)

[1]Study was before 2007 consensus statement on injury definitions and data collection procedures in rugby union. [2]No confidence intervals reported. [3]Study only reported injuries causing most days absence rather than mean severity across a main injury group. [4]Meta-analysis with four studies used for pooled analysis (Brooks et al., 2005b; Brooks et al., 2005c; Fuller et al., 2008; Fuller et al., 2013). [5]Top tier of professional rugby competitions in England. * Only one result in category.

For injury burden (days absence/1000 hours), in the English Premiership competition across 2011-19 ligament/joint (non-bone) injuries dominate the top five highest risk match injuries. Three different ligament injuries were usually in the top five highest risk injuries (all except 2015-16, 2017-18, 2018-19 when there were two), with ACL injuries included every season (apart from 2017-18) (Kemp et al., 2015; Kemp et al., 2016; Kemp et al., 2017; Kemp et al., 2018; Kemp et al., 2020). Figure 1.2 features the top five highest risk injuries during 2011-2019.

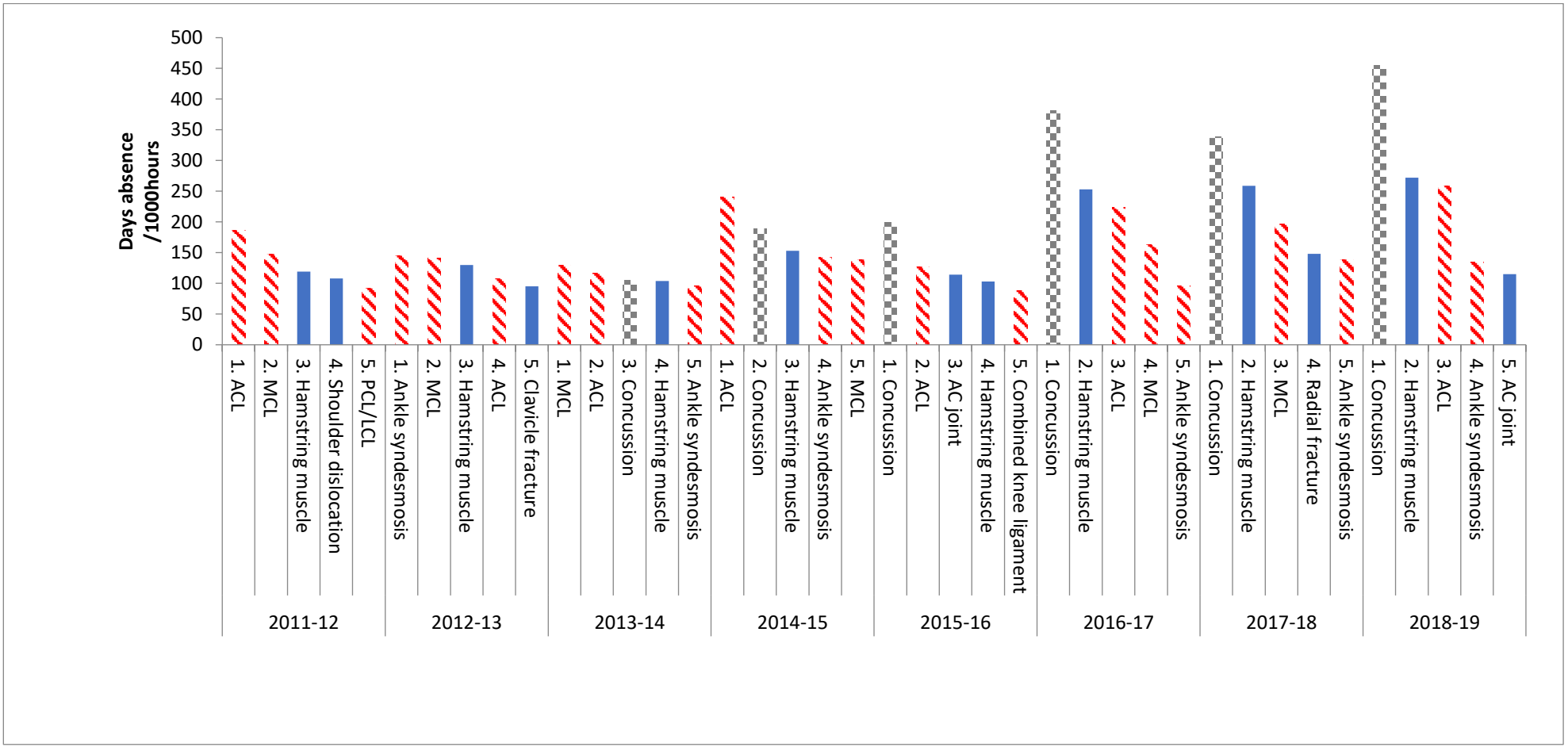

Figure 1.2. Top five highest risk match injuries – English Premiership Rugby. Adapted from the PRISP annual reports 2011-2019 (Kemp et al., 2018; Kemp et al., 2019; Kemp et al., 2020). Key: Lined bars = Ligament injuries; Squared bars = Concussion; Filled bars = Any other injury.

In elite rugby, muscle/tendon and ligament/joint (non-bone) injuries are some of the most severe and frequently occurring injuries players receive and are therefore extremely debilitating to playing squads. Generally in elite RU, there appears to be a trend towards more severe injuries (Kemp et al., 2019). Whether this is due to a more conservative approach to injury management or increased damage caused by larger collisions remains to be established. A deeper understanding of the potential causes and subsequently any preventative measures against these injuries would be of great value to both governing bodies and medical staff.

1.2.3 Risk factors for injury in rugby

From the available literature, it is difficult to state exactly how each muscle/tendon and ligament/joint (non-bone) injury occurred during rugby matches or training. Nevertheless, the most common causes of injury in RL and RU are tackles and physical collisions (Fuller et al., 2007), with the ball carrier generally at highest risk (Quarrie, 2008) though not for concussion (Cross et al., 2017). Further risk factors for injury in rugby identified in previous literature are: playing position (Brooks et al., 2005b; Gabbett et al., 2011), level of play (Williams et al., 2013), training volume and load (Gabbett and Jenkins, 2011; Cross et al., 2016b; Windt et al., 2016), ground conditions and playing surface (Williams et al., 2011), anthropometric characteristics (Fuller et al., 2010b; Gabbett et al., 2012), previous injury (Quarrie, 2001) including concussion (Cross et al., 2016a; Rafferty et al., 2018), physiological characteristics (Gabbett et al., 2012; Williams et al., 2013) and age (Brooks, 2004). The precise mechanisms of tendon and ligament injury are not well understood (Riley, 2004;

Griffin et al., 2006), with multiple factors probably involved (Riley, 2004; Griffin et al., 2006). It has been suggested that interactions between genetic and environmental factors can amplify intrinsic risk factors (anthropometry, physiological characteristics, etc.) and place a predisposed athlete at higher risk of injury once an inciting event occurs (Meeuwise, 1994; Riley, 2004; September et al., 2006). For example, under relatively constant loading conditions the matrix of ligaments and tendons will adapt in response to load (Khan and Scott, 2009). However, substantial variation in the loading pattern such as higher strains or a higher volume of low strains could lead to maladaptation, resulting in degeneration or a failed healing response (Khan and Scott, 2009) and thus injury (Figure 1.3). The tolerable load varies between individuals, resulting in large inter-individual variation in response to ligament and tendon tissue loading. This large inter-individual variation is thought to be partly due to a genetic component (Collins and Raleigh, 2009), meaning some individuals are more predisposed to ligament and tendon injury than others.

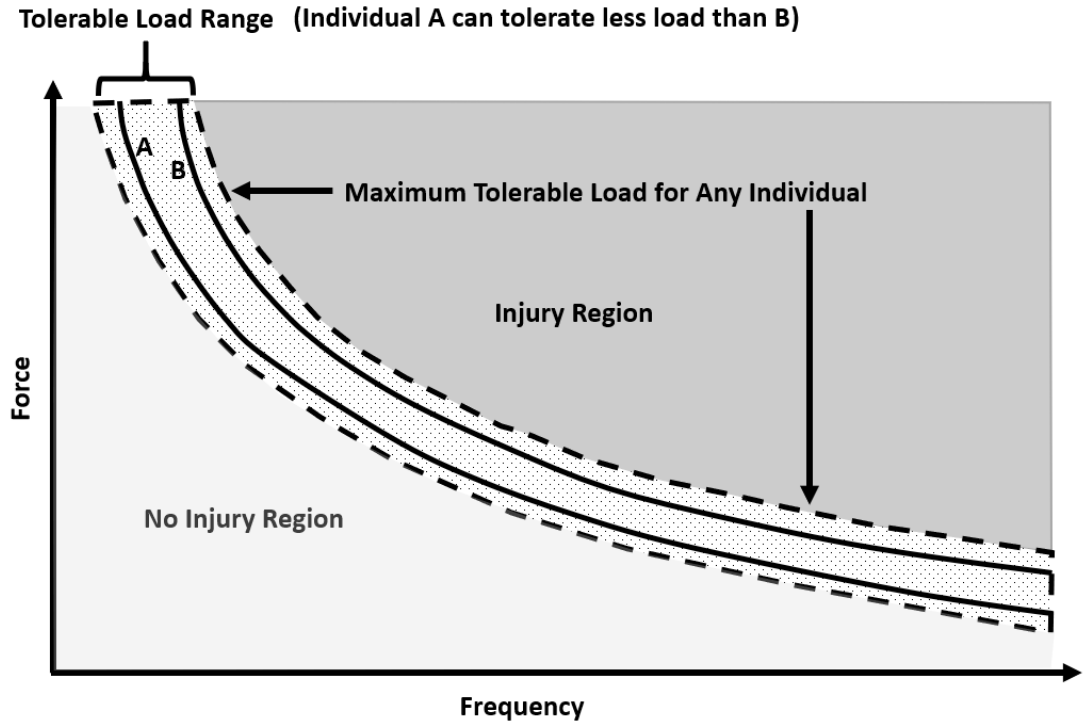

Figure 1.3. Hypothetical curve illustrating the relationship between magnitude (force) and frequency of load, which can injure the tendon or ligament (injury region). The tolerable load range for a given population is indicated by the dashed lines. The tolerable load curves for two hypothetical individuals are indicated by the solid lines (A can tolerate less load than B). Adapted from (September et al., 2012).

1.2.4 Tendon and ligament pathologies

1.2.4.1 Tendinopathy

Tendons, especially the Achilles, are designed to tolerate significant loads. Mechanical loading of tendon leads to an increase in collagen gene expression and an upturn in collagen protein complex synthesis, which is likely regulated by the strain experienced by local tenocytes (Magnusson et al., 2010). The increased collagen formation peaks ~24 hours after substantial mechanical loading, while the degradation of collagen proteins also increases after loading but appears to peak earlier (Magnusson et al., 2010). Thus, maintaining tendon homeostasis is a finely tuned process, and despite tendon's ability to adjust to mechanical loading, overuse will potentially result in injury such as tendinopathy.

Traditionally, ‘tendinitis’ was the preferred term to describe chronic pain in a symptomatic tendon, which implied that inflammatory processes played a central role in the disease aetiology. However, treatment protocols designed to modify inflammation had limited success (Almekinders and Temple, 1998; Bisset et al., 2006) and few or no inflammatory cells were found in symptomatic tissue (Hashimoto et al., 2003; Maffulli et al., 2003). Therefore, the terms ‘tendinosis’ or more generally ‘tendinopathy’ are now preferred (Abate et al., 2009). Tendinopathy is a diverse clinical syndrome associated with swelling, pain, impaired tissue healing and decreased performance (Morrey et al., 2013). There appears to be a continuum between physiology and pathology; as such, overuse (e.g. excessive repetitive loading of the tendon) could be considered the primary cause of disease (Abate et al., 2009).

Biomolecular studies of tendinopathy are relatively sparse although some observations have been made. Increased expression of messenger RNA (mRNA) has been found for type I and III collagens within symptomatic tendons (Pajala et al., 2009; Magnusson et al., 2010). This could reflect decreases in total collagen content (and a biological attempt to compensate) and an increased ratio of type III collagen relative to type I (Riley et al., 1994a; Gonçalves-Neto et al., 2002; Eriksen et al., 2002). This increased proportion of type III collagen within the main fibre bundles appears to reduce fibril diameter (Lapiere et al., 1977), probably weakening the tendon increasing risk of rupture (Magnusson et al., 2002).

Tendinopathies are caused by multiple intrinsic and/or extrinsic risk factors (Table 1.3) (Riley, 2004). Common intrinsic risk factors include age, sex, anatomical factors, hyperthermia and systemic diseases (Nourissat et al., 2013; Magnan et al., 2014), with genetic variation also recently proposed (Ribbans and Collins, 2013). Anatomical factors such as alignment and suboptimal biomechanics could contribute to two-thirds of Achilles tendon disorders among athletes (Kvist, 1994). Low-level highly repetitive strains below the failure threshold, or high strains even without great repetition, cause tendon degeneration (Wang, 2006). Thus, excessive loading during physical training is considered the primary extrinsic determinant of tendon degeneration (Selvanetti et al., 1997). In the presence of intrinsic risk factors such as genetic predisposition, excessive loading may therefore further increase the risk of tendinopathy. There is no direct evidence of what causes tendinopathy in rugby players, although potential causes include: differing ground conditions that change the magnitude and temporal characteristics of the loads experienced; running and certain contact situations that elevate low-level repetitive loading; tackling, scrums and mauls that elicit high strain; excessive training and match volume (insufficient recovery and/or excessive loading).

Table 1.3. Intrinsic and extrinsic risk factors for tendon pathologies

Intrinsic risk factors	Extrinsic risk factors
Age	Environmental conditions
Anthropometry	Shoes, surfaces
Sex	Occupation
Systemic diseases	Physical activity, sport
Anatomical variation	Training errors
Flexibility	Nutrition
Previous injury	Medication, smoking
Genetics	Mechanical loading
Hyperthermia	

Adapted from (Ribbans and Collins, 2013; Nourissat et al., 2013; Magnan et al., 2014)

1.2.4.2 Tendon rupture

Tendon rupture is an acute injury where partial or complete tearing of the tendon occurs. This is observed at the microscopic and macroscopic level, whereas tendinopathy occurs without macroscopic tearing (Nourissat et al., 2015). Partial or complete rupture will inhibit tendon continuity, limiting range of motion and force-generating capabilities. Extrinsic risk factors are thought to dominate tendon rupture incidence, with intrinsic risk factors also considered important (Sharma and Maffulli, 2006). Intrinsic and extrinsic risk factors for tendon rupture are similar to those mentioned for tendinopathy (Table 1.3), although rupture often follows one isolated overloading event (Magnusson et al., 2010; Thornton and Hart, 2011; Kaux et al., 2011; Voleti et al., 2012; Wang et al., 2012; Docheva et al., 2015). In rugby, this is probably through high loading scenarios such as scrums, mauls, sprinting, tackling and landing from jumps. During loading, the crimping formation of the collagen within the tendon is lost, and the collagen responds to the increasing load linearly (Leadbetter, 1992). Tendon strain >4% causes microscopic tearing of fibres and strain >8-10% causes macroscopic failure and rupture (Butler et al., 1978; Leadbetter, 1992).

The aetiology of tendon rupture is not completely understood (Longo et al., 2009). However, it appears to be multi-factorial, typically involving a combination of excessive loading and intrinsic risk factors (Axibal and Anderson, 2013). Histologically, degenerative tendinopathy is the most frequent finding in acute tendon ruptures (2010).

1.2.4.3 Molecular changes in tendinopathy and tendon rupture

Gene expression is altered in symptomatic tendons. Increased mRNA expression has been reported for proteoglycans such as aggrecan and biglycan (Corps et al., 2006), decorin and versican (Karousou et al., 2008), glycoproteins such as tenascin-C and fibronectin (Magnusson et al., 2010), angiogenic factors such as vascular endothelial growth factor (VEGF) (Pufe et al., 2001), collagen type I (Karousou et al., 2008), TIMP 1 and 2 (Karousou et al., 2008) and proteolytic enzymes such as the disintegrin and metalloproteinase (ADAM-12) (Jones et al., 2006) plus several matrix metalloproteinases (MMP 1, 2, 9, 13 and 23) (Jones et al., 2006; Karousou et al., 2008). Conversely, decreased mRNA expression has been reported for TIMP3 and MMP 3, 10 and 12 (Jones et al., 2006). However, the molecular signature of tendinopathy appears quite different from that of tendon rupture. Jones et al. (2006) found lower mRNA expression in ruptured than tendinopathic tendons of ADAMTS 2, 3 and 17, MMP 7, 16, 23, 24 and 28, as well as TIMP 2, 3 and 4 and increased expression of ADAMs 8 and 12, ADAMTS4, TIMP1 and MMP 1, 8, 10, 12, 19 and 25. Such differences in gene expression potentially contribute to disease pathophysiology (Xu and Murrell, 2008).

Alterations in gene expression in symptomatic tendons suggests there is an interaction between genes and environment and thus a genetic component to the aetiology of this disease. Indeed, in a twin study of tennis elbow (epicondylitis) in women (Hakim et al., 2003), heritability was estimated at ~40%. Furthermore, several studies report associations between Achilles tendinopathy and several genetic variants across a variety of genes such as *COL5A1* (Mokone et al., 2006; Abrahams et al., 2013; Khoury El et al., 2014), *MMP3* (Raleigh et al., 2009; El

Khoury et al., 2016), *TNC (*Tenascin C*)* (Mokone et al., 2005; Saunders et al., 2013), *TIMP2* (El Khoury et al., 2013; El Khoury et al., 2016), Fibrillin-2 (*FBN2*) (El Khoury et al., 2015), Caspase-8 (*CASP8*) (Nell et al., 2012), microRNA-608 (*MIR608*) (Abrahams et al., 2013) and growth differentiation factor-5 (*GDF5*) (Posthumus et al., 2010b). Additionally, Achilles tendon rupture has been associated with variants in the *MMP3* and *TIMP2* genes (El Khoury et al., 2016).

1.2.4.4 Anterior cruciate ligament tear and rupture

ACL injuries are among the most frequent knee ligament injuries in sport and usually require reconstruction (Corry, 2000; Majewski et al., 2006). In RU, although ACL injuries are not the most frequent they have been in the top five most severe injuries for seven of the last eight seasons in English Premiership Rugby (2011-2019) (Kemp et al., 2017; Kemp et al., 2019; Kemp et al., 2020), accounting for 259 days absence/1000 hours in 2018-2019 (Kemp et al., 2020). ACL injuries frequently lead to muscle weakness, altered movement, joint effusion, reduced functional performance and have been associated with continuing clinical sequelae such as chondral lesions, meniscal tears and increased risk of early-onset post-traumatic osteoarthritis (von Porat et al., 2004; Nebelung and Wuschech, 2005; Quatman et al., 2011; Chu et al., 2011; Levine et al., 2013).

Dallalana (2007) established that the primary mechanisms of ACL injury in RU are a player being tackled, tackling or in general collisions, accounting for 43%, 29% and 14% of all ACL injuries, respectively (Dallalana, 2007). However, the remaining 14% of ACL injuries occurred through non-contact mechanisms such as twisting and

turning (Dallalana, 2007). More recently, using video to analyse the mechanisms for ACL injury in RU showed that 57% occurred through contact (Montgomery et al., 2018). Two main scenarios were identified: offensive running and being tackled, suggesting that the ball carrier is at increased risk of ACL injury. The remaining 43% were through non-contact mechanisms, mainly sidestepping manoeuvres. There are numerous intrinsic and extrinsic risk factors for ACL injury (Table 1.4), although trauma to the knee is a fundamental requirement (Smith et al., 2012b; Smith et al., 2012a).

Table 1.4. Intrinsic and extrinsic risk factors for ACL injuries

Intrinsic risk factors	**Extrinsic risk factors**
Biological age	Environmental conditions
Body mass index	Shoes, surfaces
Sex	Protective equipment
Neuromuscular and cognitive factors	Physical activity, sport
Hormonal factors	Training errors
Previous injury	Mechanical loading
Anatomical risk factors (Q angle, tibial slope, generalised and specific knee joint laxity, notch width, pelvic tilt, ACL size, foot pronation, decreased relative (to quadriceps) hamstring strength and recruitment)	
Genetics	

Adapted from (Griffin et al., 2006; Alentorn-Geli et al., 2009; Smith et al., 2012b; Smith et al., 2012a)

ACL injuries can either be partial tears or complete ruptures. Like tendon, when load is placed through the ACL the crimping formation of collagen will stretch linearly with increasing load (Woo et al., 2000). Strains >4% cause microscopic tearing of the fibres and >8-10% strain cause macroscopic failure and rupture (Woo et al., 2000). Though high-traumatic strains are a typical cause of ACL rupture, microscopic damage to ligament tissue occurs at relatively low levels of strain (Provenzano et al., 2002). Furthermore, changes at the microscopic level such as extra-cellular matrix (ECM) alterations and cellular damage alter the mechanical properties of

ligaments, thus when a ligament with microstructural alterations has strain applied, rupture can follow (Provenzano et al., 2002). Thus, ACL rupture may occur in the same manner as tendon rupture, with prior degeneration of the tissue before the inciting event.

1.2.4.5 Molecular characteristics of ACL tear and rupture

Over the last ~25 years, numerous studies have examined genetic factors that potentially predispose an individual to ACL injury (Harner et al., 1994; Flynn et al., 2005; Khoschnau et al., 2008; Posthumus et al., 2009c; Posthumus, 2009; Posthumus et al., 2010a; Malila et al., 2011; Posthumus et al., 2012; Stępien-Słodkowska et al., 2013; Ficek et al., 2013; Ficek et al., 2014; Mannion et al., 2014; Rahim et al., 2014; Johnson et al., 2015; O'Connell et al., 2015; Stępień-Słodkowska et al., 2015b; Stępień-Słodkowska et al., 2015a). ACL tears seem at least twice as likely in individuals with a family history of ACL tear compared to those with no family history (Harner et al., 1994; Flynn et al., 2005). Indeed, a recent twin study found that the genetic contribution to ACL rupture was ~69% (Magnusson et al., 2020). The majority of research into the genetics of ACL injury has utilised gene association studies (GAS). From these studies, variants in several genes have been associated with altered risk of ACL injury, as detailed in section 1.2.6.; *COL1A1* (Posthumus et al., 2009c; Ficek et al., 2013; Stępien-Słodkowska et al., 2013), *COL5A1* (Posthumus, 2009; Stępień-Słodkowska et al., 2015b), *COL12A1* (Posthumus et al., 2010a), *COL3A1* (O'Connell et al., 2015; Stępień-Słodkowska et al., 2015a), *MMP3* (Malila et al., 2011), *MMP12* (Posthumus et al., 2012), proteoglycan genes *ACAN* (Aggrecan) and *DCN* (Decorin) (Mannion et al., 2014), angiogenesis-associated

signalling pathway genes *VEGFA* and *KDR* (Rahim et al., 2014) and the ECM gene *FBN2* (El Khoury et al., 2015).

1.2.5 Genetics of tendon and ligaments

Genetic variation may have a strong influence on tendon and ligament structure and function, which could alter an individual's risk of injury. Inter-individual variability of tendon and ligament properties is likely to cause microtrauma and macrotrauma at differing strain levels among individuals, thus similar injury-inciting events amongst rugby players may have vastly different outcomes. Published associations exist between gene variants (of proteins that play structural and functional roles within tendons and ligaments) and susceptibility to injury for tendinopathy (Mokone et al., 2005; Mokone et al., 2006; Raleigh et al., 2009; Posthumus et al., 2010b; Nell et al., 2012; Abrahams et al., 2013; Saunders et al., 2013; El Khoury et al., 2013; Khoury El et al., 2014; El Khoury et al., 2015; El Khoury et al., 2016), tendon rupture (El Khoury et al., 2016) and ACL rupture (Posthumus et al., 2009c; Posthumus, 2009; Posthumus et al., 2010a; Malila et al., 2011; Posthumus et al., 2012; Ficek et al., 2013; Stępien-Słodkowska et al., 2013; Rahim et al., 2014; Stępień-Słodkowska et al., 2015b; Stępień-Słodkowska et al., 2015a; O'Connell et al., 2015; El Khoury et al., 2015). Therefore, due to the high incidence and severity of tendon and ligament injuries within elite rugby, there is a potential future role for genetic screening of players to aid in injury risk management, but the practicalities are yet to be developed. In addition, the literature regarding genetic variants and tendon and ligament injuries is in its infancy, with little replication. Currently, there are no studies examining the genetics of tendon and ligament injuries within elite rugby.

1.2.6 Identifying candidate genes

Traditionally, top-down or unmeasured genotype approaches have been utilised to identify the heritability of phenotypes. While these provide useful estimates for identifying the genetic influence of certain phenotypes, they offer no evidence of the specific genes or polygenic profile that contributes to the phenotype. Furthermore, high-throughput approaches such as genome-wide association studies (GWAS) frequently identify a variety of candidate genes, of which only a small percentage are actually relevant to the phenotype of interest and validating all the identified candidate genes is not always possible (Tranchevent et al., 2016). GWAS also require particularly large sample sizes to be effective and meet the generally accepted significance level of $P < 5 \times 10^{-8}$ to minimise the risk of false positives, but that is not yet feasible in rugby. Thus, there is a need to study candidate genes because an adequately powered GWAS is currently impossible, although judicious use of GWAS results from other relevant populations to identify candidate genes can be fruitful (Heffernan et al., 2017b). A strength of GAS is that selection of candidate genes is based on detailed knowledge of a protein and its role vis-à-vis the phenotype of interest. Once a candidate gene is identified, the next logical step is to find functionally significant polymorphisms, with priority given to non-synonymous (missense) single nucleotide polymorphisms (SNPs) that change an amino acid in a protein or a nonsense variation that creates a premature stop codon, as these are most likely to have substantial biological effects (Tabor et al., 2002). However, polymorphisms in regions of DNA that regulate the expression of genes have recently become more appreciated for their functional roles (Drysdale et al., 2000; Newton-Cheh and Hirschhorn, 2005). Thus, several genes have been identified that may influence injury risk and are worthy of study within elite rugby (Table 1.5).

Table 1.5. Candidate genes, candidate proteins and their abbreviations.

Candidate Protein	**Candidate protein abbreviation**	**Candidate gene**	**Candidate gene abbreviations**
Type I collagen	Col I	*Collagen type I alpha I*	*COL1A1*
Type III collagen	Col III	*Collagen type III alpha I*	*COL3A1*
Type V collagen	Col V	*Collagen type V alpha I*	*COL5A1*
Kinase insert domain receptor	KDR	*Kinase insert domain receptor*	*KDR*
N/A - Non-coding RNA	N/A	*MicroRNA 608*	*MIR608*
Matrix metalloproteinase - 3	MMP3	*Matrix metalloproteinase – 3*	*MMP3*
Tissue inhibitors of metalloproteinases – 2	TIMP2	*Tissue inhibitors of metalloproteinases-2*	*TIMP2*
Vascular endothelial growth factor A	VEGFA	*Vascular endothelial growth factor A*	*VEGFA*
Nidogen 1	NID1	*Nidogen 1*	*NID1*
Collagen beta (1-0) galatosyltransferase 1)	COLGALT1	*Collagen beta (1-0) galatosyltransferase 1)*	*COLGALT1*

1.2.6.1 COL1A1 as a candidate gene

COL1A1 codes for the α1 chain of Col I, which is responsible for the high tensile strength of tendons and ligaments via its strong parallel fibre bundles and cross-linking formation (Thompson and Czernuszka, 1995). Several studies have investigated associations between the Sp1 polymorphism (rs1800012) and a variety of soft tissue injuries; including cruciate ligament ruptures, Achilles tendinopathy and

rupture, shoulder dislocation and tennis elbow (Table 1.6). Individuals of TT genotype appear to be at lower risk of cruciate ligament injury, particularly the ACL (Khoschnau et al., 2008; Posthumus, 2009; Ficek et al., 2013). In contrast, there seems to be no association between tendinopathies or tendon rupture and the Sp1 polymorphism (Posthumus et al., 2009a).

Table 1.6. *COL1A1* rs1800012 genetic association studies with tendon and ligament injuries in humans.

Study	Phenotype	Target population	Participants	Findings
Khoschnau et al. (2008)	Cruciate ligament ruptures, Shoulder dislocations	Sweden	No ethnicity reported. 233 cruciate ligament injury participants, 126 shoulder dislocation participants, 325 female controls	Individuals with TT genotype had a reduced risk of injury for cruciate ligament ruptures and shoulder dislocations compared to GG carriers.
Posthumus et al. (2009)	ACL injuries	SA	Caucasian. 117 ACL rupture participants. 130 controls.	TT genotype underrepresented in ACL injury group compared to controls.
Posthumus et al. (2009a)	Achilles tendinopathy, Achilles tendon ruptures	SA	Caucasian. 85 Achilles tendinopathy participants. 41 participants with partial or complete ruptures. 126 controls.	No differences in genotypes.
Ficek et al. (2013)	ACL injuries	Poland	Caucasian. 91 professional football players with ACL rupture – all non-contact. 143 apparently healthy professional soccer players as controls.	No differences in genotypes. There was an overrepresentation of G-T haplotypes (1997G+1245T) in controls suggesting, carriers may have reduced risk of injury.
Stępien-Słodkowska et al. (2013)	ACL injuries	Poland	No ethnicity reported. 138 male recreational skiers with ACL rupture. 183 apparently healthy male skiers as controls.	Carriers of the GG genotype were at lower risk of ACL injury than carriers of the TT genotype.
Erduran et al. (2014)	Tennis elbow	Turkey	No ethnicity reported. 103 with tennis elbow. 103 apparently healthy controls.	No differences in genotypes.

Key: SA = South Africa

1.2.6.2 COL3A1 as a candidate gene

Col III is an important fibrillar collagen that is similar in structure to Col I. However, Col III is a homotrimeric molecule (three α1 (III) chains) as opposed to the heterotrimeric form of Col I (Gelse et al., 2003; Banos et al., 2008). Col III frequently mixes with Col I to form mixed fibrils and is also plentiful in elastic tissue (von der Mark, 1981). Specifically, it is found in the solid component of tendons and ligaments (Frank, 2004), where it functions with Col I, V and XII to enable normal collagen fibrillogenesis (Liu et al., 1997; Minamitani et al., 2004). The pro-α1 chains of Col III are encoded by the *COL3A1* gene. Four studies have investigated the association between *COL3A1* and ACL rupture (Table 1.7), but none have examined Achilles tendon pathology. Stępień-Słodkowska et al. (2015a) found the AA genotype of the *COL3A1* rs1800255 polymorphism was more common in male recreational Polish skiers with ACL rupture than apparently healthy skiers. Similar evidence was found in Polish professional footballers (O'Connell et al., 2015), and these results have been supported within two separate GWAS (Kim et al., 2017; Kim et al., 2021). Collectively, these results suggest that individuals involved in sport who carry the AA genotype may have increased risk of ACL rupture.

Table 1.7. *COL3A1* rs1800255 genetic association studies with tendon and ligament injuries in humans.

Study	Phenotype	Target population	Participants	Findings
Stephien-Slodkowska et al. (2015a)	ACL rupture	Poland	No ethnicity reported. 138 male recreational skiers with ACL ruptures. 183 male apparently healthy skiers.	The AA genotype was overrepresented in the ACL group compared to controls.
O'Connell et al. (2015)	ACL rupture	SA/ Poland	Caucasian. 333 participants with ACL rupture (242 SA and 91 Poland). 378 apparently healthy controls (235 SA and 143 Poland).	No differences in genotype frequency distributions between the SA ACL group and the SA control group. However, the AA genotype was overrepresented in the Polish ACL group compared to Polish controls. No allele associations for any of the groups.
Kim et al. (2017)	ACL rupture	Mixed	Caucasian, Latin-American, East Asian, African, South-east Asian. 5,148 Achilles tendon injury participants. 97,831 apparently healthy controls. 598 ACL rupture participants. 98,744 apparently healthy controls.	No associations after Benjamini-Hochberg correction for testing multiple hypotheses.
(Kim et al., 2021)	ACL rupture	Mixed	European Ancestry 2,214 ACL and PCL injury participants. 519,869 apparently healthy controls.	A allele was overrepresented in the ACL/PCL group compared to controls.

Key: SA = South Africa

1.2.6.3 COL5A1 as a candidate gene

Probably the most explored gene regarding tendon and ligament injury is *COL5A1* (Table 1.8), which encodes the α1 chains of type V collagen. Col V is a minor fibrillar collagen that is known to associate with type I and III collagen (Gillies and Lieber, 2011). Although Col V is a minor collagen in terms of content, research suggests that it functions as a major collagen in developing connective tissues (Roulet et al., 2007). Mokone et al. (2006) were the first to associate the *COL5A1* gene with Achilles tendon pathology, finding the C allele of the rs12722 polymorphism less common in those with injury. This association was replicated for Achilles tendinopathy (September et al., 2009) and ACL rupture in females (Posthumus, 2009), with the C allele also underrepresented in tennis elbow patients versus controls (Altinisik et al., 2015). These findings suggest the C allele may be protective against tendon and ligament injuries. A recent investigation by the RugbyGene project (Heffernan et al., 2015) found differences in allele and genotype frequencies for the *COL5A1* rs12722 and rs3196378 polymorphisms between elite rugby athletes (rs12722: CC genotype = 21%, C allele = 47%; rs3196378: CC genotype 23%, C allele = 48%) and non-athletes (rs12722: CC genotype: 16%, C allele = 41%; rs3196378: CC genotype = 16%, C allele = 41%, $p \leq 0.02$) (Heffernan et al., 2017a). These findings suggest that elite rugby players may have an inherited resistance against soft-tissue injury.

Table 1.8. *COL5A1* rs12722 genetic association studies with tendon and ligament injuries in humans.

Study	Phenotype	Target population	Participants	Findings
Mokone et al. (2006)	Achilles tendon pathology, Achilles tendinopathy, Achilles tendon rupture	SA	Caucasian. 111 participants with current or past history of Achilles tendon pathology, Including 72 chronic tendinopathy participants, 39 Achilles tendon rupture participants.	The frequency of the A2 (C) allele was higher in the controls compared to the Achilles tendon pathology group. An even stronger protective role was seen for the A2 (C) allele in in controls compared to the chronic tendinopathy patients.
September et al. (2009)	Achilles tendinopathy	SA / Australia	Caucasian. 83 Australian and 93 SA tendinopathy patients, 210 Australian and 132 SA controls.	Individuals with CC genotype in both populations (Australian/SA) had a reduced risk of developing Achilles tendinopathy compared to any other genotypes.
Posthumus et al. (2009b)	ACL injuries	SA	Caucasian. 129 ACL rupture participants, 216 physically active controls with no history of ACL injury.	The CC genotype was underrepresented in the female ACL rupture group, but not in the male.
Stepien-Slodkowska et al. (2015b)	ACL injuries	Poland	No ethnicity reported. 138 male recreational skiers with ACL ruptures, 183 apparently healthy male recreational skiers without any reported history of ligament or tendon injury.	No differences in genotype distribution between groups. Higher frequency of rs12722 C-T and rs13946 C-T polymorphisms haplotype in controls suggests reduced risk of ACL injury.
Altinisik et al. (2015)	Tennis elbow	Turkey	No ethnicity reported. 152 tennis elbow patients, 195 healthy controls.	Individuals with the A2 (C) allele were underrepresented in patient group. Individuals with A1

				allele (T) have an increased risk of developing tennis elbow.
Brown et al. (2017)	Achilles tendinopathy, Achilles tendon rupture	UK	Caucasian. 87 Achilles tendinopathy participants, 25 Achilles tendon rupture participants, 130 asymptomatic controls.	No independent differences found between groups. Three inferred allele combinations from rs12722, rs3196378 and rs71746744 were identified as risk modifiers. The T-C-D combinations was associated with increased risk of Achilles tendon pathology and rupture, the C-A-I combination was associated with increased risk of Achilles tendon pathology, tendinopathy and rupture, the C-C-D combination was associated with decreased risk of Achilles tendon pathology and rupture.
Kim et al. (2017)	Achilles tendinopathy, Achilles tendon rupture		Caucasian, Latin-American, East Asian, African, South-east Asian. 5,148 Achilles tendon injury participants. 97,831 apparently healthy controls. 598 ACL rupture participants. 98,744 apparently healthy controls.	No associations after Benjamini-Hochberg correction for testing multiple hypotheses.

Key: SA = South Africa.

1.2.6.4 KDR as a candidate gene

The angiogenesis pathway is a crucial component for the remodelling of the ECM in response to loading (Petersen et al., 2004). Research has shown an increased expression of angiogenic molecules after mechanical loading on tendon tissue (Petersen et al., 2004; Pufe et al., 2005), suggesting upregulation may be important to maintain ECM homeostasis after loading. Several growth factors and cytokines regulate the angiogenesis pathway, principally VEGF with the A isoform thought to be the most potent (Petersen et al., 2003), which is encoded by *VEGFA*. The biological effects of *VEGFA* are mainly regulated by the KDR protein. Variants within *KDR* have been previously associated with ACL rupture (Rahim et al., 2014; Rahim et al., 2017) but not Achilles tendinopathy (Rahim et al., 2016) (Table 1.9). The A allele of rs1870377 appears to have a possible protective effect as it was overrepresented within control populations compared to ACL rupture groups when included as part of inferred haplotype analysis (Rahim et al., 2014; Rahim et al., 2017). However, when analysed independently no associations were found, suggesting its influence is likely polygenic in nature.

Table 1.9. *KDR* rs1870377 genetic association studies with tendon and ligament injuries in humans.

Study	Phenotype	Target population	Participants	Findings
Rahim et al. (2014)	ACL rupture	SA	Caucasian. 227 ACL rupture participants. 227 apparently healthy controls with no history of ACL injury.	No independent differences found between groups. The G-A haplotype (rs2071559 and rs1870377) was significantly under-represented in the controls compared to the ACL group.
Rahim et al. (2016)	Achilles tendinopathy	SA/UK	Caucasian. 195 chronic Achilles tendinopathy participants (87 from SA, 108 UK). 250 asymptomatic controls (120 SA, 130 from UK).	No independent differences found between groups. No inferred haplotype differences between groups.
Rahim et al. (2017)	ACL rupture	SA	SA coloured ethnic group (unique to Western Cape of SA). 98 ACL rupture participants. 100 physically active asymptomatic controls with no history of tendon or ligament injury.	No independent differences found between groups. The A-A inferred haplotype (rs2071559 and rs1870377) was significantly over-represented in the control group compared to the ACL group.

Key: SA = South Africa.

1.2.6.5 MIR608 as a candidate gene

MicroRNA's (miRNA) are a class of small non-coding RNA's that induce gene silencing and translational repression (Lau and Lai, 2005; 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). This potential variance has been seen in the microRNA 608 (*MIR608)* gene, which was associated with altered risk of Achilles tendinopathy (Abrahams et al., 2013; Brown et al., 2017; Kim et al., 2017). To date, three studies have investigated the link between the *MIR608* rs4919510 polymorphism and Achilles tendon pathology (Table 1.10), with none examining ACL rupture. Abrahams et al. (2013) were the first to explore this area, comparing 160 Achilles tendinopathy participants to 342 controls. They showed that the CC genotype was more frequent in the Achilles tendinopathy group than controls. However, a replication study by Brown et al. (2017) found no differences between equivalent groups, although the CG genotype was associated with decreased risk of Achilles rupture. The latest investigation involved a much broader population and a genome-wide approach; Kim et al. (2017) observed that although *MIR608* rs4919510 did not approach genome-wide significance ($P < 5 \times 10^{-8}$), when covariates such as age, sex and ancestry were not used in analysis a tentative association was identified ($P = 5.1 \times 10^{-3}$). These combined results suggest that *MIR608* may have a role in altering tendon injury risk but the evidence is inconclusive.

Table 1.10. *MIR608* rs4919510 genetic association studies with tendon and ligament injuries in humans.

Study	Phenotype	Target population	Participants	Findings
Abrahams et al. (2013)	Achilles tendinopathy	SA/ Australia	Caucasian. 160 chronic Achilles tendinopathy participants 342 apparently healthy controls.	The CC genotype frequency of rs4919510 was overrepresented compared to the CG and GG genotypes. The combined rs4919510 CC genotype and *COL5A1* rs3196378 CA genotype was overrepresented in the tendon group compared to controls. Furthermore, the rs4919510 CC genotype and the *COL5A1* rs3196378 A allele was overrepresented in the tendon group compared to controls.
Brown et al. (2017)	Achilles tendinopathy and Achilles tendon rupture	UK	Caucasian. 112 Achilles tendon pathology participants (87 chronic Achilles tendinopathy and 25 Achilles tendon rupture. 130 apparently healthy controls.	No differences in genotypes frequency or allele frequency distributions between Achilles tendinopathy and controls. However, the CG genotype of rs4919510 was associated with decreased risk of rupture compared to controls. When inferred allele combinations were analysed for rs4919510 and *COL5A1* rs3196378 no associations found with risk of Achilles tendinopathy.
Kim et al. (2017)	Achilles tendinopathy, Achilles tendon rupture		Caucasian, Latin-American, East Asian, African, South-east Asian. 5,148 Achilles tendon injury participants. 97,831 apparently healthy controls. 598 ACL rupture participants. 98,744 apparently healthy controls.	Moderate-weak evidence of replication ($P = 5.1 \times 10^{-3}$) for Achilles tendinopathy or rupture, but no replication with ACL rupture, after Benjamini-Hochberg correction for testing multiple hypotheses.

Key: SA = South Africa

1.2.6.6 MMP3 as a candidate gene

MMP3, a protein encoded by the *MMP3* gene, has a fundamental role in the regular development, repair and remodelling of connective tissues, by regulating ECM homeostasis via proteolytic activity (Foster, 2012). Several studies have examined the association between polymorphisms rs679620, rs591058 and rs650108 within *MMP3* and Achilles tendon pathologies and ACL ruptures (Table 1.11). These three polymorphisms span most of the *MMP3* gene as they are within all four major haploblocks (one exon SNP rs679620, two intron SNPs rs591058, rs650108) (Foster, 2012). Raleigh et al. (2009) first investigated the three polymorphisms, finding all three independently associated with increased risk of Achilles tendinopathy, specifically the GG genotype of rs679620, CC genotype of rs591058 and AA genotype of 650108. The GG genotype of rs679620 has also been associated with Achilles tendon rupture (El Khoury et al., 2016). Conversely, Posthumus et al. (2012) and Gibbon et al. (2016) found no independent associations between any of these variants and Achilles tendinopathy (Gibbon et al., 2016) or ACL rupture (Posthumus et al., 2012; Gibbon et al., 2016). However, when inferred haplotype was considered, Posthumus et al. (2012) and Gibbon et al. (2016) found they were associated with ACL rupture and Achilles tendinopathy, respectively. Interestingly, Gibbon et al. (2016) found the G (rs679620), C (rs5901058) and G (rs650108) alleles were overrepresented in controls, which contrasts with previous findings Raleigh et al. (2009), but aligns with a recent study of Achilles tendon rupture, ACL tears and tendinopathy in a broader population (Kim et al., 2017). Therefore, the literature appears to suggest the chromosomal region 11q22 has some influence on musculoskeletal injuries, most likely polygenic in nature, and warrants further investigation.

Table 1.11. *MMP3* rs679620, rs591058 and rs650108 genetic association studies with tendon and ligament injuries in humans.

Study	Phenotype	Target population	Participants	Findings
Raleigh et al. (2009)	Achilles tendinopathy and rupture	SA	Caucasian. 114 Achilles tendon pathology patients including 75 with Achilles tendinopathy and 39 with partial or complete rupture. 98 controls.	Independent associations between the GG genotype of rs679620, the CC genotype of rs591058 and the AA genotype of rs650108 and Achilles tendinopathy. The ATG haplotype (rs679620, rs591058 and rs650108) was under-represented in the tendinopathy compared to controls. No associations between *MMP3* variants and Achilles tendon rupture.
Posthumus et al. (2012)	ACL rupture	SA	Caucasian. 129 ACL rupture patients. 216 apparently healthy controls.	No independent associations for rs679620 compared to controls. Haplotypes T-1G-A-A and C-2G-G-G (*MMP10* rs485055, *MMP1* rs1799750, *MMP3* rs679620 and *MMP12* rs2276109) were different between control and ACL groups and controls and non-contact subgroup, respectively.
El Khoury et al. (2016)	Achilles tendinopathy and rupture	UK	Caucasian. 118 Achilles tendon pathology patients including 93 with Achilles tendinopathy and 25 participants with partial or complete rupture. 131 asymptomatic controls.	rs679620 GG genotype overrepresented in Achilles tendon rupture group compared to controls. No association with Achilles tendinopathy.

Gibbon et al. (2016)	Achilles tendinopathy ACL rupture	Australia/ SA	Caucasian. 160 Achilles tendinopathy patients. 195 apparently healthy controls. 234 ACL rupture patients. 232 apparently healthy controls.	No independent differences for rs679620, rs591058 and 650108 between Achilles tendinopathy and controls or between ACL rupture and controls. Haplotype 6a-G-C-G (rs3205058, rs679620, rs591058 and rs650108) overrepresented in the control group compared to the Achilles tendinopathy group when only Australian samples analysed. No genotype or allele frequency differences from inferred haplotypes for ACL injury.
Kim et al. (2017)	Achilles tendinopathy, Achilles tendon rupture		Caucasian, Latin-American, East Asian, African, South-east Asian. 5,148 Achilles tendon injury participants. 97,831 apparently healthy controls. 598 ACL rupture participants. 98,744 apparently healthy controls.	No associations after Benjamini-Hochberg correction for testing multiple hypotheses.

Key: SA = South Africa

1.2.6.7 TIMP2 as a candidate gene

TIMPs are natural inhibitors of MMPs, which they bind with in a 1:1 stoichiometry (Visse and Nagase, 2003). In pathological conditions such as Achilles tendinopathy where irregular MMP activity occurs, alterations in TIMP are important as they directly influence MMP activity (Visse and Nagase, 2003). *TIMP2* rs4789932 was associated with Achilles tendon pathologies in two studies (El Khoury et al., 2013; El Khoury et al., 2016) (Table 1.12). However, they contain opposing findings with the CT genotype associated with Achilles tendon pathology in one (El Khoury et al., 2013), but overrepresented in controls in another (El Khoury et al., 2016). Recently, Kim et al. (2017) reported no association after corrections for testing multiple hypotheses, but possibly adds a little support to the data of El Khoury et al. (2016). Thus, although at present it is unclear which genotype/allele within the *TIMP2* polymorphism affects tendon injury risk, the evidence tentatively suggests that it may play a role.

Table 1.12. *TIMP2* rs4789932 genetic association studies with tendon and ligament injuries in humans.

Study	Phenotype	Target population	Participants	Findings
El Khoury et al. (2013)	Achilles tendinopathy and rupture	SA / Australia.	Caucasian. 173 Achilles tendon pathology participants of which 134 with Achilles tendinopathy and 39 with partial or complete rupture. 248 asymptomatic controls.	Association between *TIMP2* rs4789932 and Achilles tendinopathy. The CC variant was overrepresented within controls, while the CT variant was over-represented within the combined Achilles tendon pathology group. No differences between the rupture group and controls.
El Khoury et al. (2016)	Achilles tendinopathy and rupture	UK	Caucasian. 118 Achilles tendon pathology participants of which 93 had chronic Achilles tendinopathy and 25 participants with Achilles tendon rupture. 131 asymptomatic controls.	Difference in genotype frequency between male Achilles tendon pathology compared to male controls. Further, difference between male ruptures compared to controls. The CT genotype was associated with lower risk of Achilles tendon pathology.
Kim et al. (2017)	Achilles tendinopathy, Achilles tendon rupture		Caucasian, Latin-American, East Asian, African, South-east Asian. 5,148 Achilles tendon injury participants. 97,831 apparently healthy controls. 598 ACL rupture participants. 98,744 apparently healthy controls.	No associations after Benjamini-Hochberg correction for testing multiple hypotheses.

Key: SA = South Africa

1.2.6.8 VEGFA as a candidate gene

Angiogenesis is essential during the repair and remodelling of injured tendons, although it can also potentially reduce mechanical stability due to proteolytic activity in the ECM by invading endothelial cells (Petersen et al., 2004). Vascular endothelial growth factor (VEGF) is an endothelial cell mitogen that stimulates angiogenesis (Ferrara, 1999; Neufeld et al., 1999). It activates endothelial cells and vascular smooth muscle migration and proliferation, as well as enhancing endothelial cell survival and differentiation (Egginton, 2009). VEGF has a number of isoforms (A-D); the most relevant being VEGFA (Egginton, 2009) encoded by the *VEGFA* gene. Variants within *VEGFA* have been associated with ACL rupture (Rahim et al., 2014) and Achilles tendinopathy (Rahim et al., 2016) (Table 1.13). Interestingly, a polymorphism appears to play a different role in acute (ACL rupture) and chronic (Achilles tendinopathy) injury. The CC variant of rs699947 was overrepresented in non-contact ACL ruptures compared to controls, suggesting a role in increased ACL rupture risk (Rahim et al., 2014). Yet the CC variant might protect against Achilles tendinopathy, being underrepresented in a control population versus an Achilles tendinopathy group (Rahim et al., 2016). Further investigation is needed to improve understanding of its role in musculoskeletal injury.

Table 1.13. *VEGFA* rs699947 genetic association studies with tendon and ligament injuries in humans.

Study	Phenotype	Target population	Participants	Findings
Rahim et al. (2014)	ACL rupture	SA	Caucasian. 227 ACL rupture participants. 227 apparently healthy controls with no history of ACL injury.	The CC genotype of rs699947 was overrepresented in participants with non-contact ACL ruptures compared to controls. The rs1570360 GA genotype was overrepresented within controls. The A-A-G haplotype (rs699947, rs1570360 and 2010963) was overrepresented in the control group compared to the non-contact ACL group.
Rahim et al. (2016)	Achilles tendinopathy	SA/UK	Caucasian. 195 chronic Achilles tendinopathy participants (87 from SA, 108 UK). 250 asymptomatic controls (120 SA, 130 from UK).	The CC genotype of rs699947 was overrepresented in the SA control group compared to the SA tendinopathy group. No other independent frequency differences found. The *VEGFA* A-G-G inferred haplotype (rs699947. rs1570360 and rs2010963) was associated with increased risk of tendinopathy in the SA group and the SA and UK combined group.
Rahim et al. (2017)	ACL rupture	SA	SA coloured ethnic group (unique to Western Cape of SA). 98 ACL rupture participants. 100 physically active asymptomatic controls with no history of tendon or ligament injury.	No differences in genotype or allele frequency data for any of the *VEGFA* polymorphisms studied. Further, no associations found from inferred haplotype analysis.

Key: SA = South Africa

1.2.6.9 NID1 as a candidate gene

NID1 rs4660148 was the strongest associated polymorphism identified in a fixed-effect meta-analysis as part of a GWAS for ACL rupture ($P < 5 \times 10^{-5}$) (Kim et al., 2017). To date, apart from being identified within Kim et al's. (2017) study, the *NID1* gene has not been investigated in regard to its influence on any musculoskeletal soft tissue injury. Therefore, the evidence is extremely sparse at present. *NID1* is located on chromosome 1q42.3 and covers 82,352 bases (Birney et al., 2004). It encodes a member of the nidogen family of basement membrane glycoproteins. Nidogens, also known as enactins, are a group of highly conserved sulphated glycoproteins, which have been implicated as playing a major structural role in the basement membrane (Ho et al., 2008). Basement membranes are small but ubiquitous ECM structures that are fundamental for tissue homeostasis and morphogenesis (Mayer et al., 1998). The main components of basement membranes are variants of collagen IV and laminin. These create two independent networks to provide the basic scaffold structure of basement membranes. However, they have a weak affinity for each other so nidogen is required to stabilise and bind them. In addition, it stimulates the integration of perlecan (a major proteoglycan) and a number of other basement membrane proteins (Fox et al., 1991; Timpl and Brown, 1996). Nidogens are thought to play an essential role in the development of the ECM, particularly when tissues are experiencing rapid turnover and growth (Fox et al., 1991). Therefore, they may influence the aetiology of musculoskeletal soft tissue injuries through their functions and interactions within the ECM.

1.2.6.10 COLGAL1 as a candidate gene

The *COLGALT1* rs8090 polymorphism was the strongest associated SNP identified in a fixed-effect meta-analysis as part of a GWAS for Achilles tendon pathology ($P < 6 \times 10^{-4}$) (Kim et al., 2017). Like the *NID1* gene, there is no previous research on *COLGALT1* regarding its effects on the pathophysiology of musculoskeletal injury. *COLGALT1* is located on chromosome 19p13.11, covers 27,569 bases (Birney et al., 2004) and encodes the collagen beta(1-0) galactosyltransferase 1 protein. It initiates collagen glycosylation through its activation of beta(1-0) galactosyltransferase enzymes (Schegg et al., 2009; Liefhebber et al., 2010), which is meditated in the endoplasmic reticulum prior to the triple-helix formation (Schegg et al., 2009). Specifically, hydroxylysine can be modified by the transfer of galactose by galactosyltransferases. These posttranslational modifications may be important in the aetiology of connective tissue disorders and musculoskeletal injuries due to the potential production of defective collagen modifying enzymes. However, the biological significance of collagen glycosylation is still uncertain and further work is required in this area.

1.3 Conclusions/future directions

The exact pathophysiology of tendon and ligament injuries is yet to be fully elucidated, as they are complex multifactorial conditions. There appears to be growing evidence of a genetic influence, although much stronger evidence is needed. The genes mentioned within this text and many other genes should be explored further regarding their relevance to tendon and ligament injuries. This would be particularly useful in a sport such as rugby, due to its high incidence and severity of injury.

To be truly relevant to elite rugby, research must involve appropriate cohorts who possess the extreme phenotypes and behaviours only found at elite level. As discussed in section 1.1, elite rugby athletes are likely to exhibit characteristics near the limits of human physiological capability; indeed, regular participation at the elite level would expose athletes to one of the highest levels of risk for tendon and ligament injury and mean they have been able to succeed despite that high environmental risk. This ability to recover or withstand musculoskeletal soft tissue injury that is potentially performance-limiting or career-ending, yet achieve elite status, may be due to distinct genetic characteristics. Furthermore, large sample sizes are required for genetic research to gain sufficient statistical power and reduce the likelihood of statistical errors. The sample sizes should be in the hundreds and ideally thousands (especially if GWAS or other hypothesis-free approaches are to be used), which is extremely challenging due to the limited number of elite athletes in a given sport. Therefore, large international collaborations are required to achieve this aim within rugby (Heffernan et al., 2015). Additionally, genetic analyses of players in combination with injury data could prove fruitful in explaining some of the currently unexplained inter-individual variability in injury susceptibility and may provide new markers of injury risk within elite rugby. Such findings could then be applied alongside existing non-genetic data to aid the personalised management of playing load and injury risk amongst rugby players.

1.4 Aims and objectives

The overall aim of the current thesis was to investigate whether genetic polymorphisms previously associated with tendon and ligament injury, differed in genotype and allele frequencies between elite rugby athletes and a non-athlete

population. Additionally, to establish whether genetic variation exists between elite rugby athletes with a history of tendon and ligament injury and those without. More specifically, the objectives were:

1. To recruit further participants to the biobank of elite rugby union athletes as part of the ongoing RugbyGene project with the aim of evaluating molecular characteristics of elite rugby athletes.
2. To investigate whether genotype and allele frequencies of genes previously associated with reduced tendon and ligament injury risk (*COLGALT1* (rs8090), *COL1A1* (rs1800012), *COL3A1* (rs1800255), *COL5A1* (rs12722), *COL5A1* (rs3196378) *KDR* (rs1870377), *MIR608* (rs4919510), *MMP3* (rs679620, rs591058 and rs650108), *NID1* (rs4660148), *TIMP2 (*rs4789932) and *VEGFA* (rs699947)) differed between elite rugby athletes and non-athlete controls, and between RU playing positions.
3. To investigate if tendon and ligament injury-associated polygenic profiles differ between elite rugby athletes and non-athletes, and between RU playing positions.
4. Finally, to establish whether tendon and ligament injury-associated polymorphisms were associated with a history of previous tendon and ligament injury within elite rugby athletes.

Chapter 2

General Methods

2.0 General methods

Methods are described in detail within this chapter, from which subsequent chapters use some or all of the experimental procedures and/or participants.

2.1 Participants

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, having given written informed consent. Caucasian non-athletes (n = 909, 44% male, height 1.70 (0.10) m, mass 72 (13) kg, age 41 (7) yr) included 94.8% British, 3.5% South African and 1.7% other nationalities. 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. 49.1% of the RU athletes 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) 540 elite male rugby athletes and 565 non-athlete’s data were utilised from a previous study by Heffernan et al. (2017a). Manchester Metropolitan University (MMU) (Ethics code: 12.07.11 (i)), the University of Glasgow and the University of Cape Town ethics committees granted approval of this study, which complies with the Declaration of Helsinki. It should be noted that sample size and participant details will vary for each experimental chapter due to cohort size, genetic data availability and methodological approach. Thus, precise sample sizes and

accompanying descriptive statistics will be specified in each experimental chapter (3, 4 and 5).

2.2 DNA sample collection

Blood (~80% of all samples), saliva (~15%) and buccal swab samples (~5%) were attained via the following procedures. A trained phlebotomist took a 5 mL blood sample from a superficial forearm vein into a 5 mL ethylenediamine tetra acetic acid (EDTA) tube (BD Vacutainer Systems, Plymouth, UK). The samples were aliquoted into 2 mL sterile microcentrifuge tubes (Eppendorf AG, Hamburg, Germany) and put in storage at −20°C until processing. Although whole-blood sampling is the preferred method due to the large quantities of cells containing DNA (Feigelson et al., 2001), less invasive saliva and buccal sample collection methods were also employed where necessary (according to participant preference, or sample collection far from freezer storage), as they also provide sufficient DNA for PCR (Feigelson et al., 2001). Saliva samples were collected into Oragene DNA OG-500 collection tubes (DNA Genotek, Ottawa, Ontario, Canada) according to the manufacturer's protocol (Figure 2.1) and stored at room temperature until processing. Prior to collection participants were instructed to abstain from food and drink for at least 30 min before producing the saliva sample. For buccal cell collection, participants were instructed not to eat or drink for at least 1 hour prior to giving the sample. Sterile buccal swabs (Omni swab; Whatman, Springfield Mill, UK) were rubbed against the buccal mucosa of the cheek for ~30 s. A second swab was collected in the same manner from the opposite cheek. Tips were ejected into sterile 2 mL tubes and stored at −20°C until processing. All samples were coded and labelled to enable participant anonymity and align with the Human Tissue Act (2004).

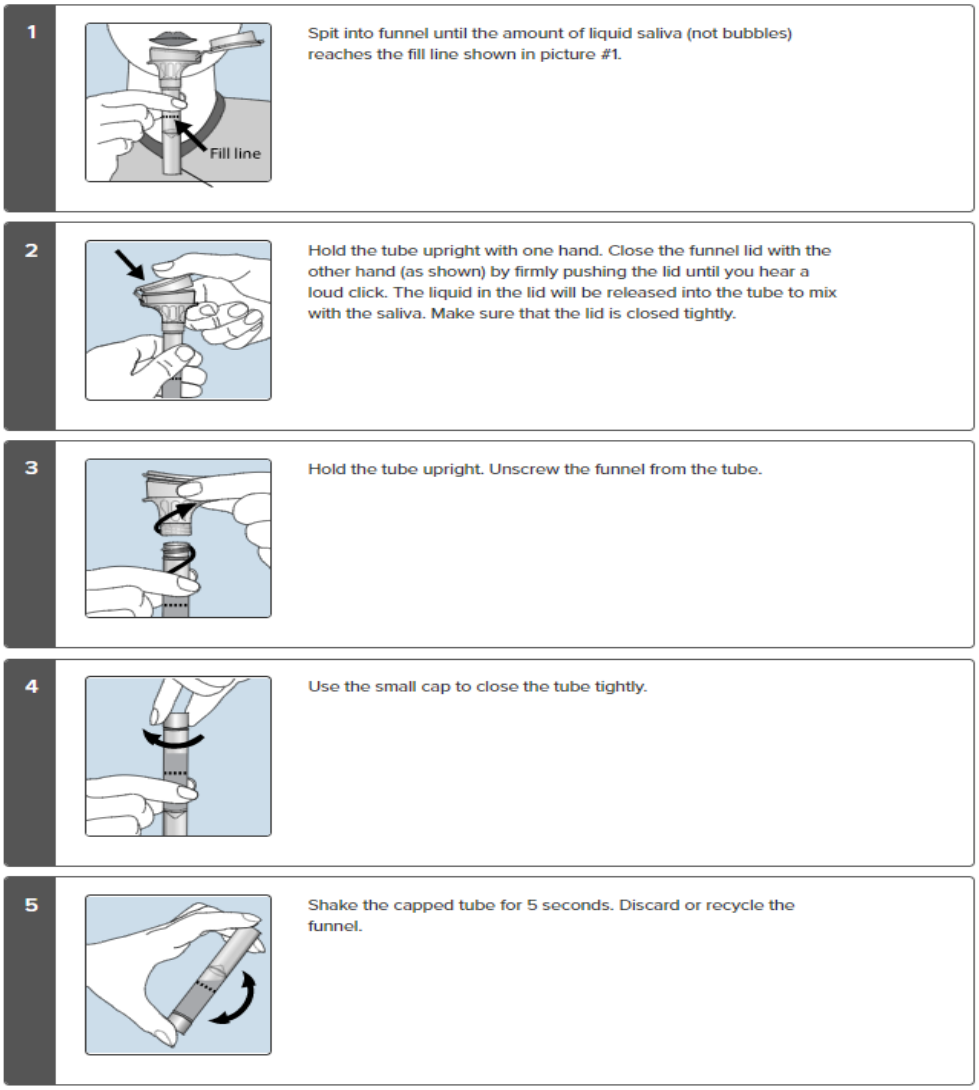

Figure 2.1. Oragene DNA sample collection procedure (DNA Genotek)

2.3 DNA isolation

DNA isolation was performed in the MMU (by the author), University of Glasgow, and University of Cape Town laboratories. Most of the samples were processed and genotyped in the MMU laboratory. There were several differences between protocols summarised below. At MMU (performed by the author) and Glasgow, DNA isolation was performed with the QIAamp DNA Blood Mini kit and standard spin column protocol, following the manufacturer's instructions (Qiagen, West Sussex,

UK). 200 µL of whole blood/saliva, or one buccal swab 600 µL of phosphate-buffered saline), were combined with Qiagen protease and incubated at 56°C for 10 min in the QIAcube incubator. Following brief centrifugation and the addition of ethanol (96%) to buffers, the resultant lysate was centrifugated at 6000 *g* (8000 rpm) for 60 s at room temperature (15-25°C) leaving the DNA sample bound to the spin column silica membrane (Figure 2.2). Wash buffers were passed through the silica membrane during further centrifugation cycles removing proteins, nucleases and other impurities that can inhibit PCR (Willard et al., 1998). Finally, the remaining purified DNA was eluted in a low-salt pH-balanced buffer providing a 100 µL solution containing isolated DNA, which was stored at 4°C pending further analysis. Genomic DNA yield is sample-dependent due to the number of cells contained per sample, however, typical yields from 200 µL of whole blood range between 3-12 µg of DNA, saliva between 5-15 µg and from one buccal swab 0.5-3.5 µg with A_{260}/A_{280} ratios of 1.7-1.9 (Qiagen.com) which are deemed good quality (Glasel, 1995).

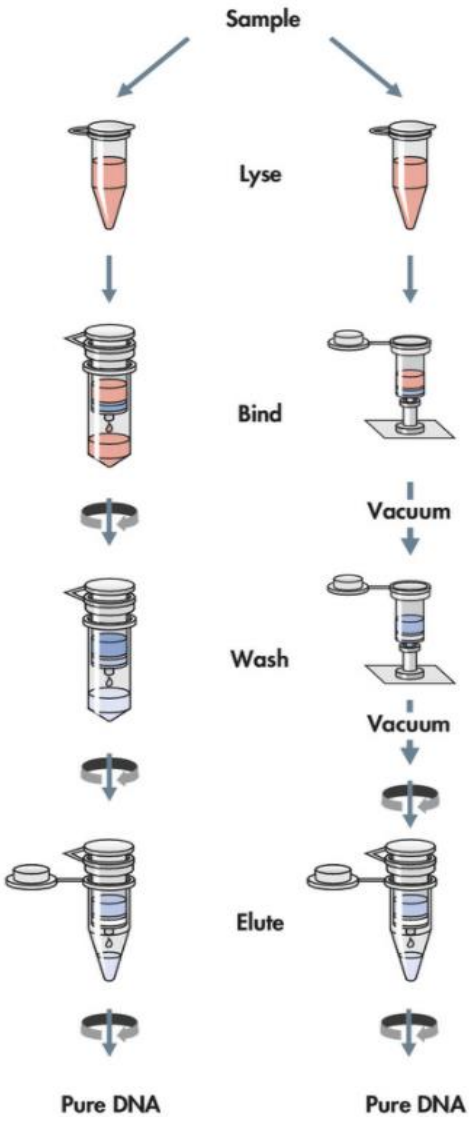

Figure 2.2. QIAamp spin column procedure (Qiagen.com)

In Cape Town, DNA was isolated from whole blood by a different protocol (Lahiri and Nurnberger, 1991). 5 mL blood samples were combined with 5 mL of low salt buffer and 125 µL of Nonidet P-40 (NP-40, Sigma) to lyse the cells. Then samples were centrifuged for 10 min at 2200 rpm at room temperature. Nuclear pellets were then washed, centrifugated and resuspended in a high salt buffer. Samples were then combined with a lysis buffer and incubated for 10 min at 55°C. Samples were then mixed with 100% ethanol until the DNA precipitated. Samples were then centrifuged for 5min at 12000 rpm at 4°C, mixed with 70% ethanol and dried. DNA

hydration buffer was added, and samples were stored at -20°C until further analysis. Genotyping of DNA isolated in Cape Town was performed in Glasgow and MMU.

At Northampton, FlexiGene kits (Qiagen) were utilised to isolate DNA (Figure 2.3). Each sample was combined with lysis buffer and centrifugated at 2000 *g* for 5 min at room temperature. Then supernatant was extracted, nuclear pellets were washed and resuspended in Qiagen protease-containing denaturation buffer. Samples were then incubated at 65°C for 10 min and then centrifuged for 5 min at 12000 rpm. DNA was further precipitated following protein ingestion, isopropanol was added and the sample centrifuged for 3 min at 2000 *g.* The samples were then washed in 70% ethanol and dried. Dehydrated DNA was then resuspended in low-salt pH-balanced buffer and samples stored at -20°C until further analysis.

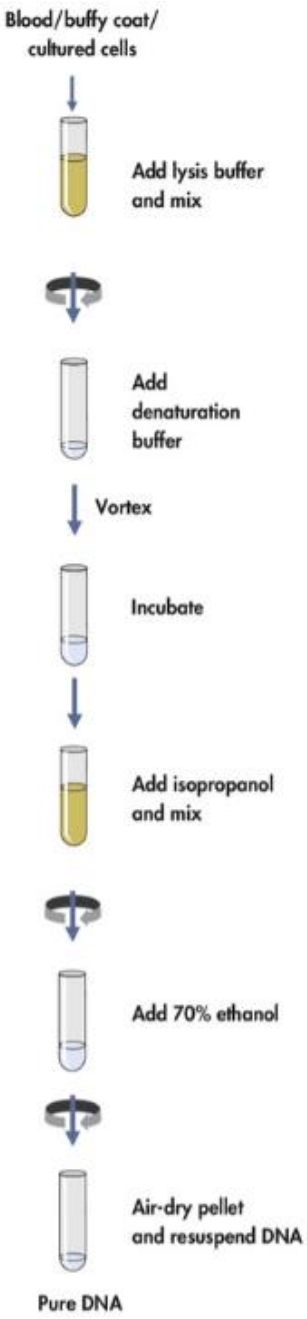

Figure 2.3. FlexiGene Procedure (FlexiGene - Qiagen.com)

2.4 Quantification of DNA

The yield and purity of some samples (i.e. Whether the proteins had been successfully removed) were evaluated using a Biophotometer Plus (Eppendork UK Limited, Stevenage, UK). This was performed periodically over several years to check the DNA was of broadly similar quality and concentration. Briefly, ~12 µL of extracted DNA was pipetted into a cuvette (UVettes, Eppendorf, UK) and exposed

to 260 nm (A_{260}) of ultra violet light (optimal absorption wavelength of DNA), and 280 nm (A_{280}) of ultra violet light (optimal absorption wavelength of protein contaminants) (Glasel, 1995). The ratio of absorbance at 260 nm divided by the reading at 280 nm was determined, with an optimal density range of A_{260}/A_{280} ratio of 1.7-2.0 generally accepted as good quality DNA (Glasel, 1995). After quantification of DNA the remaining sample was stored at 4°C until further analysis.

2.5 Genotyping

Genotyping for *COLGALT1* (rs8090), *COL1A1* (rs1800012), *COL3A1* (rs1800255), *COL5A1* (rs12722), *COL5A1* (rs3196378) *KDR* (rs1870377), *MIR608* (rs4919510), *MMP3* (rs679620, rs591058 and rs650108), *NID1* (rs4660148), *TIMP2 (*rs4789932) and *VEGFA* (rs699947) was performed using two protocols. Protocol one: ~374 elite male rugby athlete samples were genotyped by combining 5 µL Genotyping Master Mix, (Applied Biosystems, Paisley, UK) 4.3 µL H_2O, 0.5 µL SNP-specific TaqMan assay (Applied Biosystems), and 0.2 µL of purified DNA (~9 ng), for samples derived from blood and saliva. For DNA taken from buccal swabs, 5 µL Genotyping Master Mix was combined with 3.5 µL H_2O, 0.5 µL assay mix, and 1 µL DNA solution (~9 ng DNA). Negative controls containing nuclease-free H_2O were included on each 96-well plate in place of the DNA sample. Real-time polymerase chain reaction (PCR) was performed using a StepOnePlus real-time detector (Applied Biosystems) (by the author and fellow researchers). However, some of the previous samples analysed within this thesis for *COL5A1* (rs12722) and *COL5A1* (rs3196378) utilised a Chromo4 real-time system (Bio-Rad, Hertfordshire, UK). In brief, denaturation began at 95°C for 10 min, with 40 cycles of incubation at 92°C for 15 s and then annealing and extension at 60°C for 1 min. For 75 samples using the StepOnePlus real-time detector (Applied Biosystems) adjustment of thermocycling conditions were made because GTXpress Master Mix (Applied Biosystems) was utilised.

Denaturation began at 95°C for 20 s, with 40 cycles of incubation at 95°C for 1 s and then annealing and extension at 60°C for 20 s. Genotyping analysis was performed with StepOnePlus software version 2.3 (Figure 2.4). Duplicates of all samples were in 100% agreement.

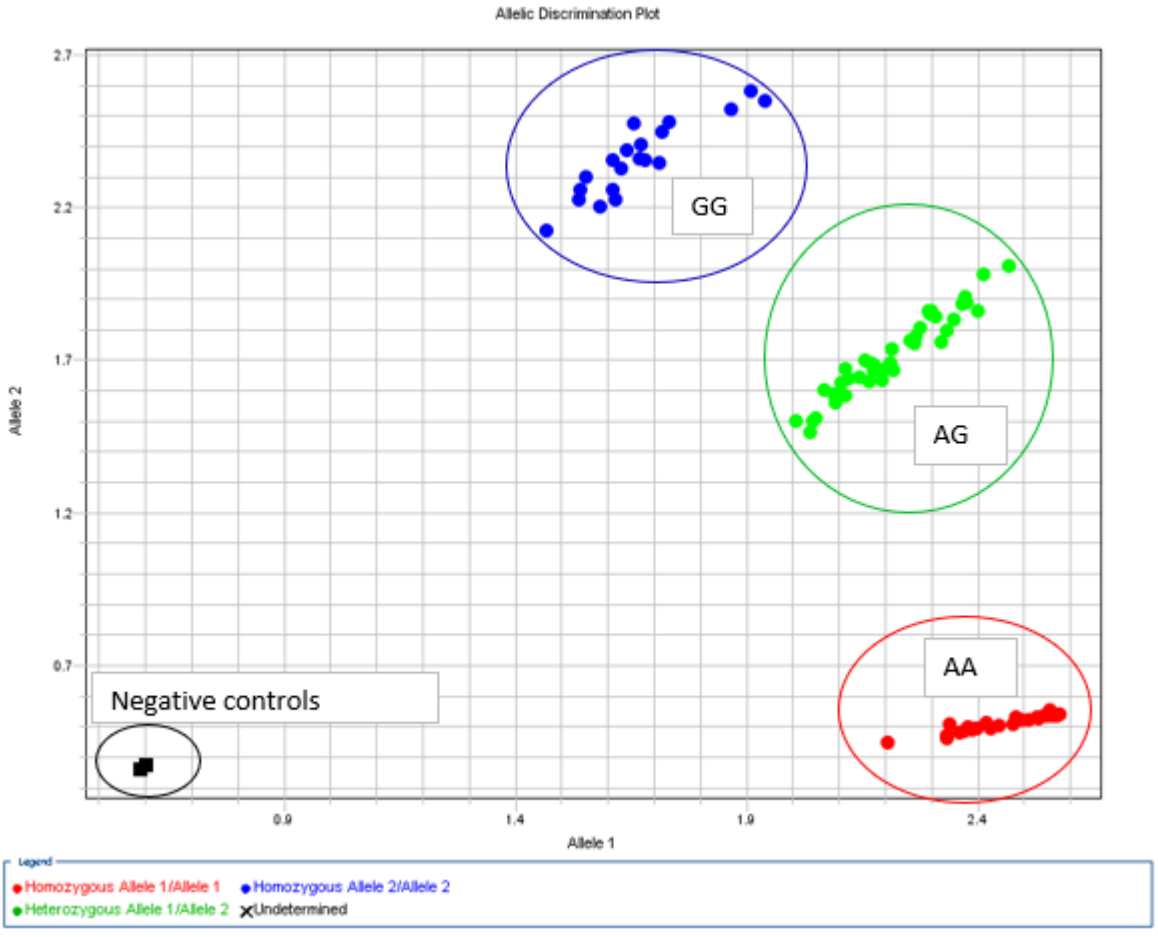

Figure 2.4. Example allelic discrimination plot for *TIMP2* rs4789952 obtained using the StepOnePlus Real-Time PCR system.

Protocol Two: Utilised the Fluidigm EP1 system, ~262 elite male rugby athletes (596 for *MMP3* rs591058, 120 for *COL5A1* rs12722 and rs3196378) and 723 non-athlete controls (909 for *COL1A1* rs1800012 and 885 for *MMP3* rs679620, rs591058 and rs650108) were genotyped for *COLGALT1* (rs8090), *COL1A1* (rs1800012), *COL3A1* (rs1800255), *COL5A1* (rs12722), *COL5A1* (rs3196378), *KDR* (rs1870377), *MIR608* (rs4919510), *MMP3* (rs679620, rs591058 and rs650108),

NID1 (rs4660148), *TIMP2 (*rs4789932) and *VEGFA* (rs699947) by combining 2 µL GTXpress Master Mix (2X) (Applied Biosystems), 0.2 µL 20X 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, saliva and buccal swab. Furthermore, 1.78 µL assay (20X) (Applied Biosystems), 1.78 µL 2X Assay Loading Reagent (Fluidigm) and 0.18 µL ROX reference dye (Invitrogen, Paisley, UK) were combined per assay inlet on the 192x24 microchip plate. Negative controls of nuclease-free H_2O were placed on each 96-well in place of DNA samples. An integrated fluid circuit controller RX (Fluidigm) was used to mix samples and assays using a Load Mix (166x) script. PCR was performed using a real-time FC1 Cycler (Fluidigm) 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 192X24 microchip plate was then placed into the EP1 Reader (Fluidigm) for end-point analysis. Genotyping analysis was performed with the Fluidigm SNP genotyping analysis software (Figure 2.5). Duplicates of all samples were in 100% agreement.

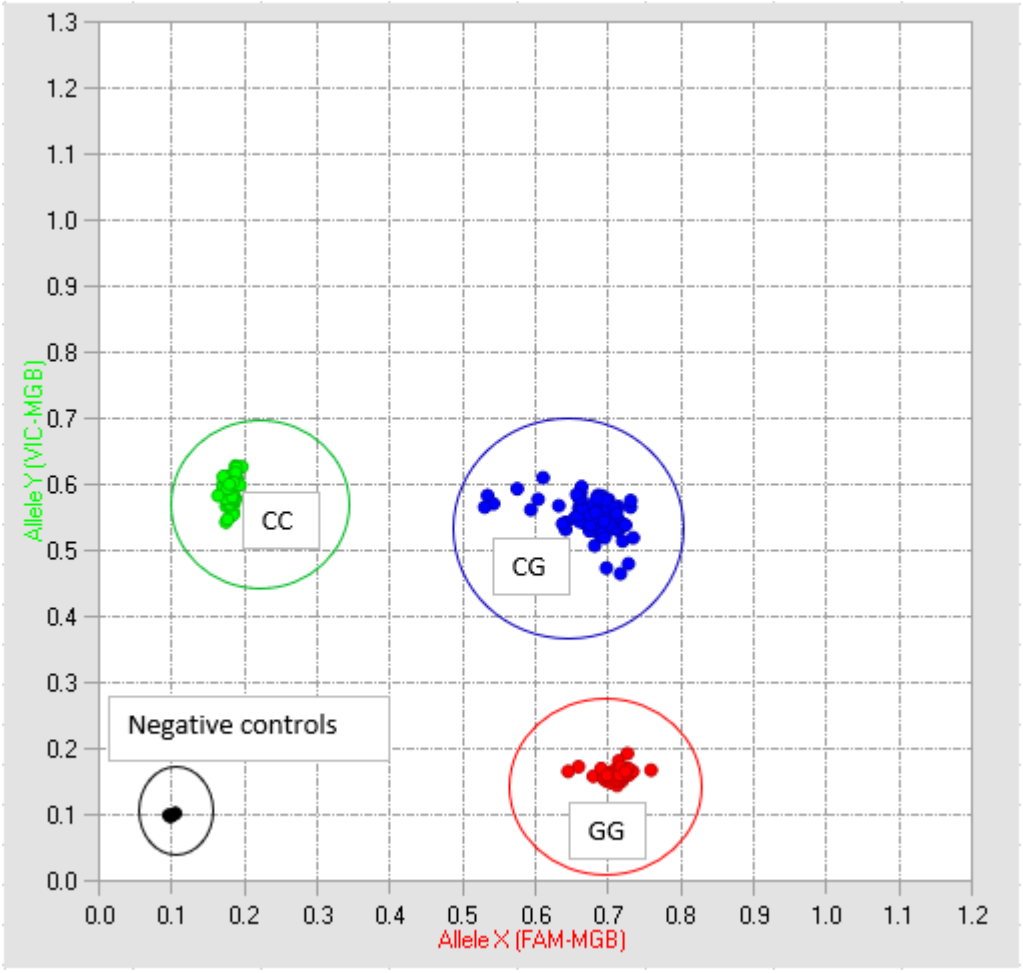

Figure 2.5. Example allelic discrimination plot for *MIR608* rs4919510 obtained using the EP1 Fluidigm PCR system.

2.6 Genotyping assays

Participants were genotyped for 13 polymorphisms: *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) using the appropriate TaqMan assays (Applied Biosystems). The TaqMan assay context sequence for each polymorphism is shown in Table 2.1. Genotypes were determined using the fluorescence-based detection technique of TaqMan real-time PCR. Utilising region-specific forward and reverse primers and 2 allele-specific

TaqMan probes (which are specifically designed to target a particular polymorphism), the wild-type allele is amplified separately from an alternative allele. Each TaqMan probe has a fluorescent reporter dye (VIC or FAM, Applied Biosystems) attached to its 5' end and a quencher dye attached at it 3' end (Malkki and Petersdorf, 2012), which bind to a complimentary allele-specific sequence of the desired SNP emitting a fluorescent dye signal that is recognised by the PCR machine (Figure 2.6).

Table 2.1. TaqMan assay context sequence for each polymorphism with VIC/FAM highlighted in bold respectively.

Polymorphism	VIC	FAM	Primers (5'-3')
COLGALT1 rs8090	**A** allele	**G** allele	CTCCC[**A/G**]GTCCC
COL1A1 rs1800012	**A** allele	**C** allele	CGCCC**[A/C]**CATTC
COL3A1 rs1800255	**A** allele	**G** allele	GTGGA**[A/G]**CTGGT
COL5A1 rs12722	**C** allele	**T** allele	ACCCA*[***C/T***]*GCGCC
COL5A1 rs3196378	**A** allele	**C** allele	ACCCC**[A/C]**GCCCT
KDR rs1870377	**A** allele	**T** allele	ACAGC**[A/T]**TGGCT
MIR608 rs4919510	**C** allele	**G** allele	CAGCT**[C/G]**CGTTT
MMP3 rs591058	**C** allele	**T** allele	GAAAT**[C/T]**GAGAA
MMP3 rs650108	**A** allele	**G** allele	TTAGA**[A/G]**GTAGC
MMP3 rs679620	**C** allele	**T** allele	TTTTT**[C/T]**GAGGT
NID1 rs4660148	**G** Allele	**T** allele	TTTTC**[G/T]**TTGGG
TIMP2 rs4789932	**A** allele	**G** allele	TATCT**[A/G]**CTGTA
VEGFA rs699947	**A** allele	**C** allele	TGGCA**[A/C]**GATCT

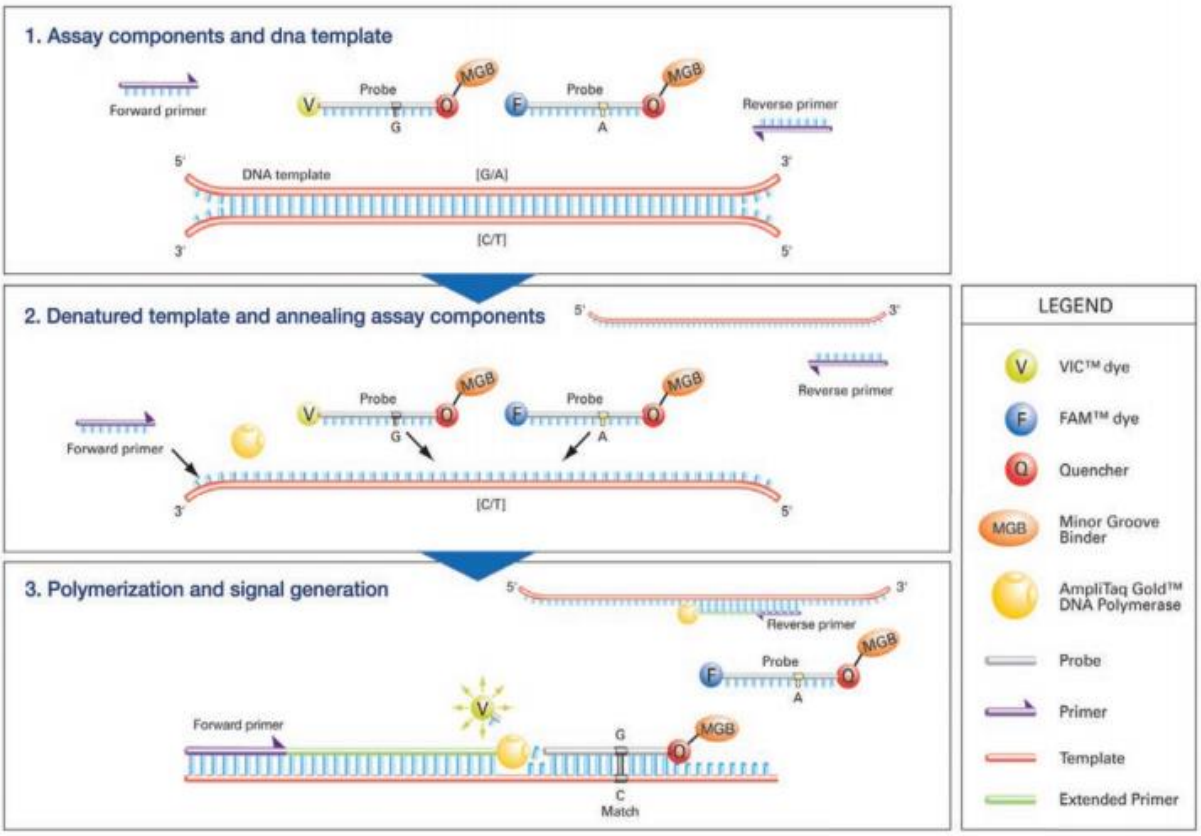

Figure 2.6 The TaqMan SNP genotyping Assay. (1) The four TaqMan SNP genotyping assay components and the target DNA template with the SNP alleles in parentheses. (2) The denatured DNA target and annealing of the assay components. (3) Signal generation leading to specific allele detection. (Thermofisher-TaqMan Assays)

2.7 Positional subgroups

To examine genotype and allele frequencies within the substantial RU cohort, athletes were placed into positional subgroups according to their movement patterns. Two positional subgroups were defined as forwards (props, hookers, locks, flankers, number eights) and backs (scrum halves, fly halves, centres, wings, full backs) (Cahill et al., 2013). The largest RU positional group within the subgroups were the front 5 (props, hookers and locks) with a sample size of 210, so they were also analysed as a discrete group. The RL player cohort size was substantially

smaller, therefore, genotype and allele frequencies between RL-specific subgroups was not performed due to a relatively low statistical power.

2.8 Elite rugby athlete tendon and ligament injury history data collection

To examine tendon and ligament injury history in elite rugby athletes, a self-reported questionnaire was utilised. The questions focussed on tendon and ligament injury specifically. Athletes were asked if they had ever suffered from tendon or ligament injury, and if so to provide specific detail of how this occurred and at what age. They were also asked to provide details as to how these injuries were medically diagnosed and if they were confirmed via a scan e.g MRI. Additional details on any family history of these injuries were also collected. The injury history questions were part of the genetic profile of elite athletes' questionnaire (concerning athlete ethnic background and elite status) which took ~15 min to complete (Appendix 1). The questionnaires were completed by the athletes with a researcher present (the author or other research team member), who clarified responses and encouraged full disclosure. To minimise athlete-response bias, coaches and medical staff were not present and players were informed that those personnel would not have access to their questionnaire responses (Guskiewicz et al., 2005). Retrospective injury data collection in the style of self-reported questionnaires can be susceptible to recall bias due to a reliance on memory (Rothman, 2012), with the level of recall accuracy declining as the level of detail requested increases (Gabbe et al., 2003). However, retrospective studies are time and resource efficient and have been suggested to encourage greater participation from elite team sport populations (Mukherjee, 2015). Therefore, due to the large sample utilised within this thesis and the population of elite team sport athletes, this approach was considered appropriate.

2.9 Calculation of TGS

The current literature regarding genetic associations with soft tissue injury is currently equivocal. Therefore, three different TGS models were utilised for analysis of elite athlete status in chapter 4; TGS model (1) based on prior literature (Table 4.1); TGS model (2) based on elite rugby athlete cohort frequency data from Chapter 3 of this thesis (Table 4.2) and TGS model (3) based on SNPs associated with elite status in rugby in Chapter 3 of this thesis (Table 4.3). However, only TGS model (1) was utilised for injury analysis in chapter 5 as this was deemed the most relevant (Table 4.1).

To quantify the combined influence of the candidate polymorphisms (Table 4.1, 4.2 and 4.3) an additive TGS algorithm was utilised (Williams and Folland, 2008), based on the assumption of codominance effects of the alleles. For bi-allelic polymorphisms, the homozygote genotypes with the 'preferable' soft tissue injury risk were allocated a 'genotype score' of 2, heterozygote genotypes were scored 1 and the other 'non-preferable' homozygote genotypes were scored 0.

TGS based on literature (Model 1) and elite rugby athlete cohort frequency data from Chapter 3 of this thesis (Model 2):

TGS algorithm: TGS = (100/26) * (*COLGALT1*$_{rs8090}$ + *COL1A1*$_{rs1800012}$ + *COL3A1*$_{rs1800255}$ + *COL5A1*$_{rs12722}$ + *COL5A1*$_{rs3196378}$ + *KDR*$_{rs1870377}$ + *MIR608*$_{rs4919510}$ + *MMP3*$_{rs679620}$ + *MMP3*$_{rs591058}$ *MMP3*$_{rs650108}$ + *NID1*$_{rs4660148}$ + *TIMP2*$_{rs4789932}$ + *VEGFA*$_{rs699947}$)

TGS based on SNPs associated with elite status in rugby in Chapter 3 of this thesis (Model 3):

TGS algorithm: TGS = (100/14) * (*COLGALT1*$_{rs8090}$ + *COL3A1*$_{rs1800255}$ + *COL5A1*$_{rs12722}$ + *COL5A1*$_{rs3196378}$ + *MIR608*$_{rs4919510}$ + *MMP3*$_{rs591058}$ + *NID1*$_{rs4660148}$)

Using the TGS algorithm, a TGS of 100 represents the 'perfect' polygenic profile for soft tissue injury risk and 0 represents the 'worst' possible outcome for the candidate genes examined in this study.

2.10 Statistical analysis

Statistical Package for Social Sciences (SPSS) for Windows version 26 (SPSS, Chicago, IL) and Microsoft Excel version 2013 were used for data analysis. Specific statistical analyses for each experimental chapter are described within the respective methods sections.

Before commencing any statistical analyses, data were tested for parametricity when required. Normal distribution was assessed using Shapiro-Wilks test of normality and z-scores were calculated using skewness and kurtosis data. For quality control and to identify genotyping errors in unrelated individuals, the Hardy-Weinberg equilibrium (HWE) was used (Hosking et al., 2004). Pearson's Chi-square tests (χ^2) were utilised to test for HWE. Furthermore, Pearson's χ^2 tests (goodness of fit and test of independence) were used to compare genotype (using three analysis models; additive, recessive, and dominant) and allele frequencies between

rugby athletes and non-athletes and between positional subgroups, and between injured and non-injured athlete groups for all 13 SNPs. *A-priori* power analyses using G*power 3.1.9.7 (Franz Faul, Universität Kiel, Germany) were performed to determine the statistical power to detect genetic associations between groups for elite athlete status and injury (details are specified within the experimental chapter). Odds ratios were calculated to estimate effect size (where appropriate) to highlight the strength of an association. Height and body mass were compared between rugby athletes and non-athletes using independent T-tests.

Mann-Whitney U tests were utilised to compare TGS differences between rugby athlete groups and non-athletes. Means and kurtosis were calculated to examine the distribution of TGS within groups. Pearson's χ^2 tests were used to compare the frequency of athletes and non-athletes in the top and bottom quartile of TGS scores. Similarly, Pearson's χ^2 tests were used to compare the frequency of injured and non-injured athletes in the top and bottom thirds of TGS scores. The ability of the TGS to correctly distinguish elite athletes from non-athletes and between injured athletes from non-injured athletes across all groups was evaluated by receiver operating characteristic (ROC) curves (Zweig and Campbell, 1993). The area under the ROC curve (AUC) was calculated with 95% confidence intervals (95% CI). Multifactor dimensionality reduction (https://sourceforge.net/projects/mdr/) software was used to calculate SNP-SNP epistasis interactions (Moore et al., 2006). Where appropriate, haplotype frequencies were inferred using SNPStats software (Solé et al., 2006). Alpha was set at 0.05 and Benjamini-Hochberg and Bonferroni corrections were implemented as appropriate to control false discovery rate (details are specified within the experimental chapter).

Chapter 3

Variants from multiple genes previously associated with reduced soft tissue injury risk are associated with elite status in rugby

3.1 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 et al., 2013). Within rugby league (RL), injury incidence rates have been reported at 172/1000 in matches (King et al., 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 et al., 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 two 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. 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. 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., 1994b). 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 et al., 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., 2009c; Stępień-Słodkowska et al., 2015a; 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. 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 et al., 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., 2016). Furthermore, the rs679620 polymorphism has been associated with Achilles tendon (El Khoury et al., 2016) and ACL rupture (Posthumus et al., 2012). In addition, the *TIMP2* gene, which is one of four 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). Most 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 identified in a genome-wide association study (GWAS) for Achilles tendon and ACL tears and tendinopathy are worthy of future study: *COLGALT1* rs8090, *NID1* rs4660148 and *MIR608* rs4919510 (Kim et al., 2017).

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., 2017a). This suggests an inherited resistance against soft tissue injury could enhance the chance of career success in certain sports, as previous evidence, although in football, has identified injuries to be negatively associated with athlete career progression (Larruskain et al., 2021). Therefore, the principle objective of this chapter was to investigate the following genetic polymorphisms previously associated with tendon and ligament injury: *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. (2017a) by adding additional participants to the previously studied *COL5A1* (rs12722 and rs3196378) polymorphisms. Thus, this chapter investigated whether these 13 polymorphisms differed in genotype and allele frequencies between elite rugby athletes and a non-athlete population, and/or between playing positions. It was hypothesised that elite rugby athletes would possess fewer of the injury-risk genotype/alleles than a non-athlete population.

3.2 Method

3.2.1 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-athlete controls (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. 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., 2017a).

3.2.2 Procedures

Blood, buccal or saliva samples were attained via the procedures detailed in Chapter 2.2. Briefly, blood was drawn from a superficial forearm vein into an EDTA tube and stored in sterile tubes at −20°C until processing. 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 processing. 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 processing.

DNA isolation was performed in the Manchester, Glasgow, and Cape Town laboratories. In Manchester and Glasgow, DNA isolation was performed with the QIAamp DNA Blood Mini kit and standard spin column protocol, following the manufacturer's instructions (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 Cape Town, DNA was isolated from whole blood by a different protocol (Lahiri and Nurnberger, 1991).

Genotyping for *COLGALT1* (rs8090, G/A), *COL1A1* (rs1800012, C/A), *COL3A1* (rs1800255, G/A), *COL5A1* (rs12722, C/T), *COL5A1* (rs3196378, A/C) *KDR* (rs1870377, A/T), *MIR608* (rs4919510, C/G), *MMP3* (rs679620, C/T, rs591058, C/T and rs650108, A/G), *NID1* (rs4660148, G/T), *TIMP2 (*rs4789932. G/A) and *VEGFA* (rs699947, A/C) was performed using two protocols. In both protocols, the appropriate TaqMan assays were utilised (Applied Biosystems) and assay context sequences for each polymorphism are presented in Table 2.1.

In the first protocol, ~374 elite male rugby athlete samples were genotyped via real-time PCR using a StepOnePlus (Applied Biosystems) as previously described in detail in section 2.5 with adjustment of thermocycling conditions because GTXpress Master Mix (Applied Biosystems, Paisley, UK) was used for 75 samples. Genotyping analysis was performed with StepOnePlus software version 2.3. Duplicates of all samples were in 100% agreement. In the second protocol, ~262 elite male rugby athletes (596 for *MMP3* rs591058) and 723 non-athlete controls (909 for *COL1A1* rs1800012 and 885 for *MMP3* rs679620, rs591058 and rs650108) were genotyped

for *COLGALT1* (rs8090), *COL1A1* (rs1800012), *COL3A1* (rs1800255), *KDR* (rs1870377), *MIR608* (rs4919510), *MMP3* (rs679620, rs591058 and rs650108), *NID1* (rs4660148), *TIMP2 (*rs4789932) and *VEGFA* (rs699947) 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) was used to mix 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 192X24 microchip plate was then placed into the EP1 Reader for end-point analysis. Genotyping analysis was performed with the Fluidigm SNP genotyping analysis software. Duplicates of all samples were in 100% agreement.

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 subgroups were defined as forwards (props, hookers, locks, flankers, number eights) and backs (scrum halves, fly halves, centres, wings, full backs) (Cahill et al., 2013). The largest RU positional group within the subgroups were the front five (props, hookers and locks) with a sample size of 210, so they were also analysed as a discrete group.

3.2.3 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 subgroups (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 (Benjamini and Hochberg, 1995) 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.

3.3 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) (Table 3.1). Athletes (all male) were taller and heavier ($P < 0.05$) than the male non-athletes.

Table 3.1. Genotype and allele distribution of controls and athletes separated by code (RL and RU) and into positional subgroups for RU, presented as genotype/allele counts followed by percentage in parentheses.

Polymorphism	Genotype/Hardy-Weinberg Equilibrium (HWE)	Non-athletes	All Rugby Athletes	RL Athletes	RU Athletes	RU Forwards	RU Front 5	RU Backs
COLGALT1 rs8090								
	GG	214 (29.9)	150 (24.0)*	27 (26.5)	124 (23.3)*	63 (20.4)*	43 (20.5)*	61 (27.4)
	GA	338 (47.1)	305 (48.7)	51 (50.0)	260 (49.0)	156 (50.6)	109 (51.9)	104 (46.6)
	AA	165 (23.0)	171 (27.3)	24 (23.5)	147 (27.7)	89 (29.0)	58 (27.6)	58 (26.0)
	Total	717	626	102	531	308	210	223
	G allele carriers	552 (76.9)	455 (72.6)*	78 (76.5)	384 (72.3)*	219 (71.1)*	152 (72.4)	165 (73.9)
	A allele carriers	503 (70.1)	476 (76.0)*	75 (73.5)	407 (76.7)*	245 (79.6)*	167 (79.5)*	162 (72.6)
	G allele	766 (53.4)	605 (48.3)	105 (51.5)	508 (47.8)	282 (45.8)	195 (46.4)	226 (50.7)
	A allele	668 (46.6)	647 (51.7)*	99 (48.5)	554 (51.2)*	334 (54.2)*	225 (53.6)*	220 (49.3)
	HWE *P* value & χ^2	>0.15, 1.99	>0.25, 0.37	<0.01[#], 7.64	>0.25, 0.19	>0.25, 0.13	>0.25, 0.40	>0.25, 1.00
COL1A1 rs1800012								
	CC	618 (68.0)	426 (67.5)	76 (73.7)	354 (66.2)	205 (65.7)	137 (64.3)	149 (66.8)
	CA	262 (28.8)	191 (30.3)	25 (24.3)	168 (31.4)	100 (32.1)	71 (33.3)	68 (30.5)
	AA	29 (3.2)	14 (2.2)	2 (2.0)	13 (2.4)	7 (2.2)	5 (2.3)	6 (2.7)
	Total	909	631	103	535	312	213	223
	C allele carriers	880 (96.8)	617 (97.8)	101 (98.0)	522 (97.6)	305 (97.7)	208 (97.7)	217 (97.3)
	A allele carriers	291 (32.0)	205 (32.4)	27 (26.2)	181 (33.8)	107 (34.3)	76 (35.7)	74 (33.1)
	C allele	1498 (82.4)	1043 (82.6)	177 (85.9)	876 (81.9)	510 (81.7)	345 (81.0)	366 (82.1)
	A allele	320 (17.6)	219 (17.4)	29 (14.1)	194 (18.1)	114 (18.3)	81 (19.0)	80 (17.9)
	HWE *P* value & χ^2	>0.25, 0.04	>0.15, 1.92	>0.25, <0.01	>0.15, 1.78	>0.15, 1.67	>0.20, 1.44	>0.25, 0.29
COL3A1 rs1800255								
	GG	432 (59.8)	342 (54.3)	66 (64.1)	280 (52.4)	154 (49.7)	106 (50.0)	126 (56.2)

	GA	243 (33.7)	247 (39.2)*	31 (30.1)	219 (41.0)*	133 (42.9)*	94 (44.3)*	86 (38.4)
	AA	47 (6.5)	41 (6.5)	6 (5.8)	35 (6.6)	23 (7.4)	12 (5.7)	12 (5.4)
	Total	722	630	103	534	310	212	224
	G allele carriers	675 (93.4)	589 (93.4)	97 (94.1)	499 (93.4)	287 (92.6)	200 (94.3)	212 (94.6)
	A allele carriers	290 (40.1)	288 (45.7)*	37 (35.9)	254 (47.5)*	156 (50.3)*	106 (50.0)*	98 (43.7)
	G allele	1107 (76.7)	931 (73.9)	163 (79.1)	779 (72.9)	441 (71.1)	306 (72.2)	338 (75.4)
	A allele	337 (23.3)	329 (26.1)	43 (20.9)	289 (27.1)*	179 (28.9)*	118 (27.8)	110 (24.6)
	HWE *P* value & χ^2	>0.10, 2.55	>0.25, 0.16	>0.25, 0.81	>0.25, 0.80	>0.25, 0.62	>0.10, 2.28	>0.25, 0.29
COL5A1 rs12722								
	TT	266 (33.2)	177 (26.7)	26 (24.8)	153 (27.0)	91 (27.8)	55 (24.7)	62 (26.0)
	TC	397 (49.5)	334 (50.4)	54 (51.4)	284 (50.3)	161 (49.2)	116 (52.0)	123 (51.7)
	CC	139 (17.3)	152 (22.9)*	25 (23.8)	128 (22.7)*	75 (22.9)*	52 (23.3)*	53 (22.3)*
	Total	802	663	105	565	327	223	238
	T allele carriers	663 (82.6)	511 (77.1)*	80 (76.2)	437 (77.3)*	252 (77.1)*	171 (76.7)*	185 (77.7)
	C allele carriers	536 (66.8)	486 (73.3)*	79 (75.2)	412 (72.9)*	236 (72.2)	168 (75.3)*	176 (73.9)*
	T allele	929 (57.9)	688 (51.9)	106 (50.5)	590 (52.2)	343 (52.4)	226 (50.7)	247 (51.9)
	C allele	675 (42.1)	638 (48.1)*	104 (49.5)*	540 (47.8)*	311 (47.6)*	220 (49.3)*	229 (48.1)*
	HWE *P* value & χ^2	>0.25, 0.19	>0.25, 0.05	>0.25, 0.08	>0.25, 0.30	>0.25, 0.05	>0.25, 0.37	>0.25, 0.29
COL5A1 rs3196378								
	AA	242 (32.8)	190 (28.8)	32 (30.5)	160 (28.5)	94 (29.0)	59 (26.8)	66 (27.7)
	AC	364 (49.4)	313 (47.4)	52 (49.5)	265 (47.1)	147 (45.4)	110 (50.0)	118 (49.6)
	CC	131 (17.8)	157 (23.8)*	21 (20.0)	137 (24.4)*	83 (25.6)*	51 (23.2)	54 (22.7)
	Total	737	660	105	562	324	220	238
	A allele carriers	606 (82.2)	503 (76.2)*	84 (80.0)	425 (75.5)*	241 (74.4)*	169 (76.8)	184 (77.3)
	C allele carriers	495 (67.2)	470 (71.2)	73 (69.5)	402 (71.5)	230 (70.9)	161 (73.2)	172 (72.3)
	A allele	848 (57.5)	693 (52.5)	116 (55.2)	585 (52.0)	335 (51.7)	228 (51.8)	250 (52.5)
	C allele	626 (42.4)	627 (47.5)*	94 (44.8)	539 (48.0)*	313 (48.3)*	212 (48.2)*	226 (47.5)
	HWE *P* value & χ^2	>0.25, 0.08	>0.20, 1.59	>0.98, <0.01	>0.15, 1.72	>0.10, 2.71	>0.98, <0.01	>0.25, 0.08
KDR rs1870377								

	TT	396 (54.8)	363 (57.3)	57 (54.8)	310 (57.8)	184 (59.0)	118 (55.1)	126 (56.2)
	TA	283 (39.1)	232 (36.7)	39 (37.5)	196 (36.6)	106 (34.0)	80 (37.4)	90 (40.2)
	AA	44 (6.1)	38 (6.0)	8 (7.7)	30 (5.6)	22 (7.0)	16 (7.5)	8 (3.6)
	Total	723	633	104	536	312	214	224
	T allele carriers	679 (93.9)	595 (93.9)	96 (92.3)	506 (94.4)	290 (92.9)	198 (95.5)	216 (96.4)
	A allele carriers	327 (45.2)	270 (45.3)	47 (45.2)	226 (42.1)	128 (41.0)	96 (44.9)	98 (43.8)
	T allele	1075 (74.3)	958 (75.7)	153 (73.6)	816 (76.1)	474 (75.9)	316 (73.8)	342 76.4)
	A allele	371 (25.7)	308 (24.3)	55 (26.4)	256 (23.9)	150 (24.0)	112 (26.2)	106 (23.6)
	HWE *P* value & χ^2	>0.25, 0.49	>0.25, 0.01	>0.25, 0.13	>0.25, 0.02	>0.20, 1.51	>0.25, 0.23	>0.05, 2.82
MIR608 rs4919510								
	CC	407 (56.4)	403 (64) *	72 (69.2)*	336 (63.0)*	194 (62.6)	136 (63.8)*	142 (63.7)
	CG	252 (34.9)	190 (30.2)	28 (26.9)	164 (30.8)	98 (31.6)	68 (31.9)	66 (29.6)
	GG	63 (8.7)	37 (5.8)	4 (3.9)	33 (6.2)	18 (5.8)	9 (4.2)	15 (6.7)
	Total	722	630	104	533	310	213	223
	C allele carriers	659 (91.2)	593 (94.1)*	100 (96.2)	500 (93.8)	292 (94.2)	204 (95.8)*	208 (93.3)
	G allele carriers	315 (43.6)	227 (36.0)*	32 (30.8)*	197 (36.9)*	116 (37.4)*	77 (36.2)*	81 (36.3)*
	C allele	1066 (73.8)	996 (79.0)*	172 (82.7)*	836 (78.4)*	486 (78.4)*	340 (79.8)*	350 (78.5)
	G allele	378 (26.2)	264 (21.0)	36 (17.3)	230 (21.6)	134 (21.6)	86 (20.2)	96 (21.5)
	HWE *P* value & χ^2	0.01[#], 6.78	0.02[#], 5.05	>0.25, 0.37	<0.05[#], 4.39	>0.20, 1.39	>0.25, 0.02	>0.05, 3.42
MMP3 rs591058								
	TT	276 (31.2)	160 (25.1)*	19 (18.4)	143 (26.5)	77 (24.6)	53 (24.7)	66 (29.1)
	TC	419 (47.3)	329 (51.8)	59 (57.3)	273 (50.6)	164 (52.4)	113 (52.6)	109 (48.0)
	CC	190 (21.5)	147 (23.1)	25 (24.3)	124 (22.9)	72 (23.0)	49 (22.8)	52 (22.9)
	Total	885	636	103	540	313	215	227
	T allele carriers	695 (78.5)	489 (76.9)	78 (75.7)	416 (77.0)	241 (76.9)	166 (77.2)	175 (77.0)
	C allele carriers	609 (68.8)	476 (74.8)*	84 (81.6)*	397 (73.5)	236 (75.3)	162 (75.3)	161 (70.9)
	T allele	971 (54.9)	649 (51.0)*	97 (47.1)	559 (51.8)	318 (50.8)	219 (50.9)	241 (53.0)
	C allele	799 (45.1)	623 (49.0)	109 (52.9)	521 (48.2)	308 (49.2)	211 (49.1)	213 (46.9)
	HWE *P* value & χ^2	>0.15, 1.72	>0.25, 0.78	>0.10, 2.30	>0.25, 0.08	>0.25, 0.73	>0.25, 0.57	>0.25, 0.29

MMP3 rs650108								
	GG	493 (55.8)	342 (53.8)	45 (43.7)	299 (55.4)	165 (52.7)	112 (52.3)	134 (59.0)
	GA	322 (36.5)	252 (39.6)	51 (49.5)	204 (37.8)	126 (40.3)	85 (39.7)	78 (34.4)
	AA	68 (7.7)	42 (6.6)	7 (6.8)	37 (6.8)	22 (7.0)	17 (7.9)	15 (6.6)
	Total	883	636	103	540	313	214	227
	G allele carriers	815 (92.3)	594 (93.4)	96 (93.2)	503 (93.1)	291 (92.9)	197 (92.1)	212 (93.4)
	A allele carriers	390 (44.2)	294 (46.2)	58 (56.3)	241 (44.6)	148 (47.3)	102 (47.7)	93 (40.9)
	G allele	1308 (74.1)	936 (73.6)	141 (68.4)	802 (74.3)	456 (72.8)	309 (72.2)	346 (76.2)
	A allele	458 (25.9)	336 (26.4)	65 (31.6)	278 (25.7)	170 (27.2)	119 (27.8)	108 (23.8)
	HWE *P* value & χ^2	>0.10, 2.27	>0.25, 0.24	>0.10, 2.20	>0.25, 0.07	>0.25, 0.10	>0.25, 0.02	>0.25, 0.62
MMP3 rs679620								
	TT	252 (28.7)	151 (23.8)	15 (14.4)	138 (25.7)	71 (22.9)	46 (21.7)	67 (29.4)
	TC	426 (48.5)	328 (51.6)	63 (60.6)	268 (49.8)	161 (51.9)	112 (52.8)	107 (46.9)
	CC	200 (22.8)	156 (24.6)	26 (25.0)	132 (24.5)	78 (25.2)	54 (25.5)	54 (27.7)
	Total	878	635	104	538	310	212	228
	T allele carriers	678 (77.2)	479 (75.4)	78 (75.0)	406 (75.5)	232 (74.8)	158 (74.5)	174 (76.3)
	C allele carriers	626 (71.3)	484 (76.2)	89 (85.6)*	400 (74.3)	239 (77.1)	166 (78.3)	161 (70.6)
	T allele	930 (53.0)	630 (49.6)	93 (44.7)	544 (50.6)	303 (48.9)	204 (48.1)	241 (52.8)
	C allele	826 (47.0)	640 (50.4)	115 (55.3)	532 (49.4)	317 (51.1)	220 (51.9)	215 (47.2)
	HWE *P* value & χ^2	>0.25, 0.60	>0.25, 0.70	<0.025#, 5.28	>0.25, <0.01	>0.25, 0.48	>0.25, 0.72	>0.25, 0.78
NID1 rs4660148								
	GG	347 (48.2)	302 (47.5)	42 (40.4)	263 (48.8)	144 (45.9)	93 (43.1)	119 (53.0)
	GT	325 (45.1)	268 (42.1)	54 (51.9)	217 (40.3)	135 (43.0)	97 (44.9)	82 (36.4)
	TT	48 (6.7)	66 (10.4)*	8 (7.7)	59 (10.9)*	35 (11.1)*	26 (12.0)*	24 (10.6)*
	Total	720	636	104	539	314	216	225
	G allele carriers	672 (93.3)	570 (89.6)*	96 (92.3)	480 (89.0)*	279 (88.9)*	190 (88.0)*	201 (89.3)*
	T allele carriers	373 (51.8)	336 (52.8)	62 (59.6)	276 (51.2)	170 (54.1)	123 (56.9)	106 (47.1)
	G allele	1019 (70.8)	872 (68.6)	138 (66.3)	743 (68.9)	423 (67.4)	283 (65.5)*	320 (71.1)
	T allele	421 (29.2)	400 (31.4)	70 (33.7)	335 (31.1)	205 (32.6)	149 (34.5)	130 (28.9)

	HWE *P* value & χ^2	<0.02[#], 5.95	>0.25, 0.33	>0.05, 2.75	>0.15, 1.95	>0.25, 0.16	>0.25, <0.01	>0.05, 2.87
TIMP2 rs4789932								
	GG	245 (35.1)	212 (33.4)	39 (37.5)	176 (32.7)	95 (30.4)	65 (30.7)	81 (36.0)
	GA	309 (44.2)	303 (47.8)	49 (47.1)	257 (47.9)	151 (48.4)	101 (47.6)	106 (47.1)
	AA	145 (20.7)	119 (18.8)	16 (15.4)	104 (19.4)	66 (21.2)	46 (21.7)	38 (16.9)
	Total	699	634	104	537	312	212	225
	G allele carriers	554 (79.3)	515 (81.2)	88 (84.6)	433 (80.6)	246 (78.8)	166 (78.3)	187 (83.1)
	A allele carriers	454 (64.9)	422 (66.5)	65 (62.5)	361 (67.2)	217 (69.6)	147 (69.3)	144 (64.0)
	G allele	799 (57.2)	727 (57.3)	127 (61.0)	609 (56.7)	341 (54.6)	231 (54.5)	268 (59.6)
	A allele	599 (42.8)	541 (42.7)	81 (39.0)	465 (43.2)	283 (45.4)	193 (45.5)	182 (40.4)
	HWE *P* value & χ^2	0.01[#], 6.63	>0.25, 0.34	>0.25, <0.01	>0.25, 0.34	>0.25, 0.17	>0.25, 0.33	>0.25, 0.11
VEGFA rs699947								
	CC	166 (24.6)	167 (27.8)	21 (21.9)	147 (28.7)	85 (28.3)	58 (28.3)	62 (29.2)
	CA	316 (46.8)	290 (48.3)	51 (53.1)	245 (47.9)	147 (49.0)	100 (48.8)	98 (46.2)
	AA	193 (28.6)	144 (23.9)	24 (25.0)	120 (23.4)	68 (22.7)	47 (22.9)	52 (24.5)
	Total	675	601	96	512	300	205	212
	C allele carriers	482 (71.4)	457 (76.0)	72 (75.0)	392 (76.5)	232 (77.3)	158 (77.1)	160 (75.5)
	A allele carriers	509 (75.4)	434 (72.2)	75 (78.1)	365 (71.2)	215 (71.7)	147 (71.7)	150 (70.8)
	C allele	648 (48.0)	624 (51.9)	93 (48.4)	539 (52.6)	317 (52.8)	216 (52.7)	222 (52.4)
	A allele	702 (52.0)	578 (48.1)	99 (51.6)	485 (47.4)	283 (47.2)	194 (47.3)	202 (47.6)
	HWE *P* value & χ^2	>0.05, 2.61	>0.25, 0.67	>0.25, 0.39	>0.25, 0.83	>0.25, 0.09	>0.25, 0.10	>0.25, 1.14

RU, rugby union; RL, rugby league. The genotype and allele carrier data represent the additive, dominant and recessive models, respectively. Asterisks (*) indicate differences in genotype distribution of the rugby athlete group or subgroups compared to non-athletes ($P \leq 0.05$). Hashtag ([#]) indicate not in HWE.

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. 3.1, Table 3.1 and appendix 2, $P < 0.03$). Furthermore, the AA genotype, proportion of A-allele carriers and A allele were overrepresented in the subgroups 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 3.1, $P \leq 0.03$). 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 (Appendix 2: 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 all χ^2, OR and allele/genotype frequency data for all SNPs, please refer to table 3.1 and appendix 2.

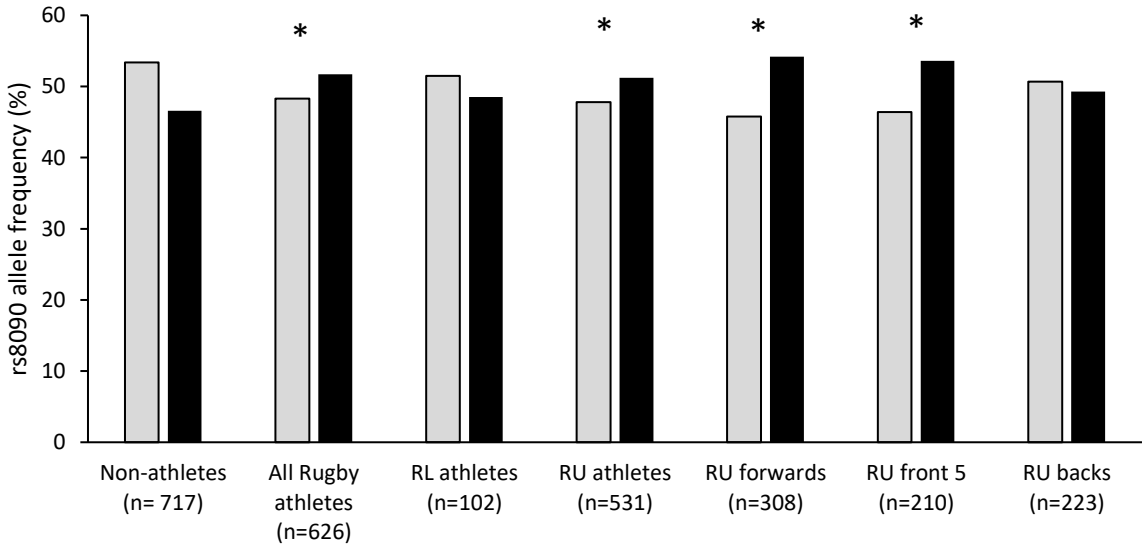

Figure 3.1. Allele frequency of *COLGALT1* rs8090 for non-athlete and athlete groups. Asterisks (*) indicate a difference in allele frequency between the particular athlete group or subgroup and non-athletes ($P < 0.01$). RU, rugby union; RL, rugby league. Grey bars = G Allele; Black bars = A allele.

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 subgroup, GA genotype and A-allele carriage were overrepresented (50.0% and 50.0%) 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 (Appendix 2: 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 subgroup 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%, $P \leq 0.04$, Table. 3.1). Furthermore, the C allele was overrepresented while the number of G-allele carriers was underrepresented in the RU forwards subgroup (78.4% and 37.4%) compared to non-athletes (73.8% and 43.6%, $P < 0.05$, Fig 3.2). 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 (Appendix 2: Table 2). There were no differences in genotype or allele frequencies for *MIR608* rs4919510 for any other groups.

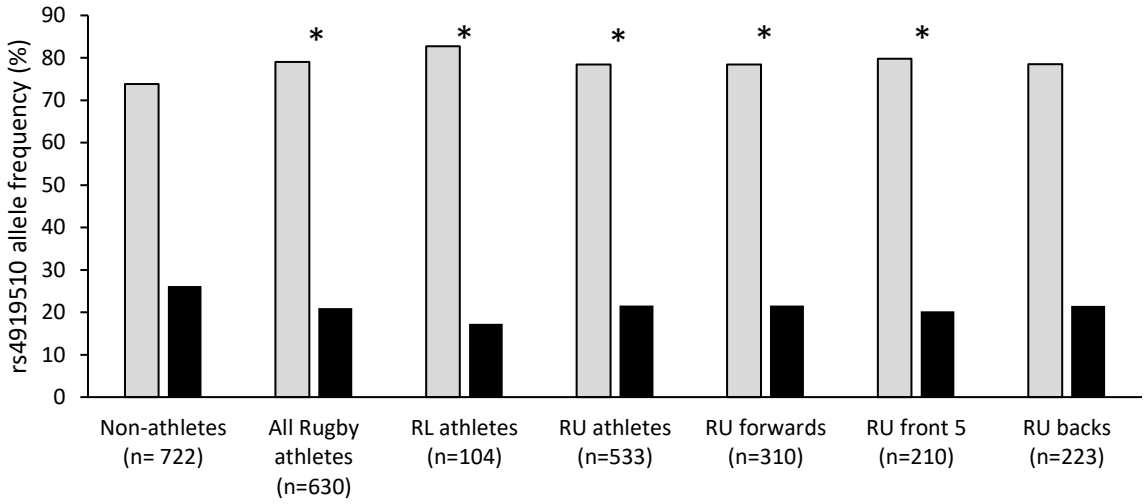

Figure 3.2. Allele frequency of *MIR608* rs4919510 for non-athlete and athlete groups. Asterisks (*) indicate a difference in allele frequency between the particular athlete group or subgroup and non-athletes ($P \leq 0.03$). RU, rugby union; RL, rugby league. Grey bars = C allele; Black bars = G allele.

For *MMP3* rs591058, the TT genotype and T allele were underrepresented, 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 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 (Appendix 2: 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 subgroups 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 subgroup 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 (Appendix 2: Table 5), while in the RU front 5 that increased to 2.0 times the odds (Appendix 2: 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 Table 3.1 and appendix 2 for further details).

3.4 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, these findings with an increased cohort size support the previous work of Heffernan et al. (2017a) 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 the authors' 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, a fixed-effect meta-analysis identified the G allele of *COLGALT1* rs8090 ($P < 6 \times 10^{-4}$) and the G allele of *NID1* rs4660148 ($P < 5 \times 10^{-5}$) as the polymorphisms most strongly associated with Achilles tendon injury and ACL rupture, respectively. While these results were not genome-wide significant ($P < 5 \times 10^{-8}$), the findings suggested 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 the present 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), which is mediated in the endoplasmic reticulum prior to triple-helix formation (Schegg et al., 2009). Specifically, hydroxylysine can be modified by the transfer of galactose by galactosyltransferases. These posttranslational modifications might be important in the aetiology of connective tissue disorders and musculoskeletal injuries due to the production of defective collagen modifying enzymes. The results of this study 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 whether it is functional or linked to functional polymorphisms.

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 subgroups. *NID1* encodes a member of the nidogen family of basement membrane glycoproteins. Nidogens are a group of highly conserved sulphated glycoproteins, implicated as playing a major structural role in the basement membrane (Ho et al., 2008). The main components of basement membranes are variants of collagen IV and laminin, which create two independent networks to provide the basic scaffold structure of basement membranes. However, they have a weak affinity for each other so nidogen is required to stabilise and bind them (Ho et al., 2008). Nidogens are thought to play an essential role in 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 and interactions within the ECM. However, to date this has not been demonstrated experimentally, and the exact mechanisms are still to be elucidated. Nevertheless, 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 (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 (Kim et al., 2017; Abrahams et al., 2013; Brown et al., 2017). However, the evidence is not consistent regarding which genotype is injury-protective. Abrahams et al. (2013) found the CC genotype overrepresented within a tendinopathy group compared to controls, which has been supported by moderate GWAS evidence ($P < 5 \times 10^{-8}$) when covariates were removed from analysis (Kim et al., 2017). However, the only other study to date could not replicate these results but did find an association between the CG genotype and Achilles tendon rupture (Brown et al., 2017). Although the present evidence is equivocal, the data from this study demonstrate a likely benefit from carrying the CC genotype in enabling elite rugby athlete status to be achieved. Indeed, 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 of the *COL3A1* rs1800255 polymorphism within elite rugby athletes, RU athletes, RU forwards and RU front 5 compared to non-athletes. Four studies have previously investigated the association between *COL3A1* rs1800255 and ACL rupture but none have examined tendon pathology. Stępień-Słodkowska et al. (2015a) found the AA genotype was more common in male recreational Polish skiers with ACL rupture than apparently healthy skiers. This was also found in Polish professional

footballers, but not replicated in a wider population (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). The data from this study show elite male rugby athletes carry a larger proportion of 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 may influence injury risk (e.g., via potentially affecting collagen formation and/or structure) is not clear. It is understood that *COL3A1* can influence the tensile strength of collagen (Kluivers et al., 2009), and as such it is possible that there may be some benefit to carrying 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 regular development, repair and remodelling of connective tissues, by controlling ECM homeostasis via proteolytic activity (Foster, 2012). Previous evidence has found a possible relationship between the CC genotype of rs591058, GG genotype of rs679620, AA genotype of rs650108, and risk of Achilles tendinopathy (Raleigh et al., 2009). However, these findings have not been replicated within a Caucasian population. When inferred haplotypes were considered, Posthumus et al. (2012) and Gibbon et al. (2016) found they were associated with ACL rupture and Achilles tendinopathy, respectively. However, Gibbon et al. (2016) found that the C (rs5901058), G (rs679620) and G (rs650108) alleles were overrepresented in controls, opposing the findings of Raleigh et al. (2009). Thus, evidence suggests

that the chromosomal region 11q22 has some influence on musculoskeletal injuries but is perhaps more interactive in nature with currently no clear evidence regarding which should be considered risk genotypes. The present study found the TT genotype and T allele of *MMP3* rs591058 were underrepresented, and the proportion of C-allele carriers overrepresented, in elite rugby athletes compared to non-athletes. Furthermore, a higher proportion of C-allele carriers were observed within RL athletes for *MMP3* rs591058 and rs679620 compared to non-athletes. Indeed, RL athletes had twice the odds of carrying the C allele at rs591058 and almost two and half times the odds of carrying the C allele at rs679620. These findings suggest some likely benefit from carrying the C allele, perhaps allowing for better regulation of ECM homeostasis and thus a more robust athlete.

The hypothesis of an inherited injury resistance within elite rugby athletes was further supported by the findings regarding the *COL5A1* polymorphisms rs12722 and rs3196378. Heffernan et al. (2017a) 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, the results from this study with an additional 123 athletes (totalling 663) extend the previous report as the CC genotype and C allele of both rs12722 and rs3196378 were found to be more common in elite rugby athletes than non-athletes.

As part of this investigation, genotype and allele frequencies were examined for differences between RU athlete subgroups (RU forwards vs RU backs and RU front 5 vs RU backs). There were no differences between subgroups for any

polymorphism studied, which suggests that elite rugby athletes, regardless of their playing position, are likely benefiting from some inherited advantage, possibly through a lower risk of soft tissue injury.

3.5 Conclusion

In conclusion, the first associations between the *COLGALT1* rs8090, *COL3A1* rs1800255, *MIR608* rs4919510, *MMP3* rs591058 and rs679620 and *NID1* rs4660148 polymorphisms and elite status in a large cohort of rugby athletes have been presented. The present study also extended Heffernan et al. (2017a) associations of the *COL5A1* rs12722 and rs3196378 polymorphisms with elite status in rugby using a larger cohort. The potential injury-reducing mechanisms behind the gene variants in this study need much more elucidation. Nevertheless, it is proposed that elite rugby athletes 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. It is very likely that this inherited benefit is highly polygenic in nature and the present study only examined 13 of many polymorphisms that might be relevant, so these findings should be interpreted in that context. Future research should look to identify further genetic variants that may influence soft tissue injury and combine analysis of elite rugby genotype data such as these with injury severity and incidence data from matches and training, to enhance knowledge in this area.

Chapter 4

Tendon and ligament injury-associated polygenic profiles of elite rugby athletes

4.1 Introduction

Chapter 3 identified several genetic variants independently associated with elite status in rugby. The studied genetic variants were previously associated with soft-tissue injury risk and were deemed important to investigate due to rugby having one of the highest incidence of injury compared to other team sports (Williams et al., 2013). This may in part be due to the increasing size and strength of the athletes (Sedeaud et al., 2013), which likely result in greater momentum during collisions, as well as producing higher forces/torques during accelerations and decelerations. Muscle/tendon (50%) and joint/ligament (32.7%) injuries account for >80% of all injuries in rugby (Schwellnus et al., 2014). More recent research that analysed match injury data across 16-seasons of the English Premiership found that ligament injuries had the highest incidence of injury and they also had the highest burden (22.6/1000 hours and 30 days, respectively) (West et al., 2020). Chapter 3 demonstrated differing genetic characteristics between elite rugby athletes and non-athletes across several genetic polymorphisms previously associated with soft tissue injury. Therefore, investigating the polygenic components of these injuries including in the context of rugby code and playing position, which vary in terms of physiological demands, may increase knowledge of any interaction between genetic injury susceptibility and rugby code/playing position. In turn, that might reflect an impact of intrinsic (e.g. anthropometric) and extrinsic (e.g. training load and match exposure) factors on how genetic injury susceptibility manifests itself in injury.

In a recent investigation utilising a twin study approach, anterior cruciate ligament (ACL) rupture was found to be ~69% heritable (Magnusson et al., 2020). It is highly

likely that this is polygenic in nature, rather than due to one genetic variant. Therefore, investigating the polygenic extent of a phenotype is a worthy proposition. Previous work has utilised the total genotype score (TGS) to indicate the magnitude of an individual's genetic predisposition for disease risk and athletic performance (Gómez-Gallego et al., 2010; Williams and Folland, 2008; Ruiz et al., 2009; Eynon et al., 2011). The TGS model established by Williams and Folland (2008) is a genetic algorithm that indicates the proportion of 'preferable' genotypes possessed in relation to a specific phenotype. Individuals with higher TGS results are thought to possess the more optimal polygenic profile for the targeted phenotype. This approach can be applied to determine an individual's genetic predisposition to elite athlete status or injury risk.

The potential polygenic nature of soft tissue injury risk has been demonstrated previously by Raleigh et al. (2009) who found an interaction between the *MMP3* rs679620 and the *COL5A1* rs12722 single nucleotide polymorphisms (SNPs) modified risk of Achilles tendinopathy (AT). When the A and C alleles of *MMP3* rs679620 and *COL5A1* rs12722, respectively, were combined (pseudo-haplotypes), they were associated with absence of AT (P = 0.002) (Raleigh et al., 2009). Similar research by Abrahams et al. (2013) found that the combination of the CC genotype of *MIR608* rs4919510 and the CA genotype of *COL5A1* rs3196378 was overrepresented in an AT cohort compared to uninjured controls. Furthermore, when *MIR608* rs4919510 CC, *COL5A1* rs71746744 AGGG/AGGG, *COL5A1* rs16399 -/- ATCT and *COL5A1* rs1134170 TT genotypes were combined in a genotype risk score, they were overrepresented within the AT cohort (odd ratio = 2.6; 95% confidence interval 1.3-4.9) (Abrahams et al., 2013). Additional supporting

evidence of a polygenic risk of injury was found by Brown et al. (2017) who reported the inferred allele combination of C-C-D of *COL5A1* rs12722, rs3196378 and rs71746744, respectively, was associated with reduced risk of Achilles tendon pathology and rupture. Further inferred haplotype combinations have been associated with Achilles tendinopathy involving the *MMP3* (Raleigh et al., 2009; Gibbon et al., 2016) and *VEGFA* (Rahim et al., 2016) genes. Similarly, inferred haplotype combinations have been associated with risk of anterior cruciate ligament injury involving polymorphisms within the MMP1, *MMP3, MMP10* and *MMP12* genes (Posthumus et al., 2012) and also *COL5A1* (Stępień-Słodkowska et al., 2015b). It therefore appears that a combination of 'preferable' genetic variants reduces the risk of susceptibility to soft tissue injuries.

Given the association of genetic markers with injury risk individually and in combination, 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. Previous work by Heffernan et al. (2017a) found elite rugby athletes carried more of the protective C alleles of the *COL5A1* rs12722 and rs3196378 polymorphisms than non-athletes and were more likely to possess the C-C inferred haplotype and the CC-CC SNP-SNP combination at those loci. Those data suggest an inherited resistance against soft tissue injury could increase the ability to withstand years of a high volume of intense training and match play, and thus contribute to the attainment of elite competitive status. Additionally, Chapter 3 highlighted several further soft tissue injury-associated genetic variants linked to elite status in rugby. Consequently, the objective of the present study was to expand on the work presented in Chapter 3

and that of Heffernan et al. (2017a) by investigating if more complex polygenic profiles, indicative of tendon and ligament injury risk, differ between elite rugby athletes and a non-athlete population. It was hypothesised that elite rugby athletes would have a higher frequency of the 'injury protective' allele/genotype combinations than non-athletes, as well as higher TGS scores.

4.2 Method

4.2.1 Participants

This study was conducted in accordance with the STROBE guidelines for a case-control observational study (von Elm et al., 2007). 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 (RA) (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 (NA) (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. For TGS and SNP-SNP epistasis interaction analyses, 590 elite rugby athletes and 436 non-athletes were utilised.

4.2.2 Procedures

Sample collection and genotyping via real-time PCR were consistent with those reported in Chapter 2 and thesis section 3.2.2.

RU Forwards, Backs and Positional Roles

To examine TGS results within the RU cohort, athletes were placed into subgroups according to their movement patterns. The two subgroups were defined as RU forwards (props, hookers, locks, flankers, number eights) and RU backs (scrum halves, fly halves, centres, wings, full backs) (Cahill et al., 2013).

Calculation of TGS

The current literature regarding genetic associations with soft tissue injury is equivocal, therefore three different TGS models were utilised for analysis; (1) TGS based on prior literature for genetic associations with soft tissue injury (Table 4.1); (2) TGS based on elite rugby athlete frequency data from Chapter 3 regardless of whether there were previously associated with soft tissue injury (Table 4.2); and (3) TGS based on elite rugby athlete frequency data but only for SNPs previously associated with elite status in rugby from Chapter 3 (Table 4.3).

To quantify the combined influence of the candidate polymorphisms (Table 4.1, 4.2 and 4.3) an additive TGS algorithm was utilised (Williams and Folland, 2008), based on the assumption of codominance effects of the alleles. The homozygote genotypes with the ‘preferable’ soft tissue injury risk were allocated a ‘genotype

score' of 2, heterozygote genotypes were scored 1 and the other 'non-preferable' homozygote genotypes were scored 0.

Prior literature based TGS model (1) and elite data-based model (2):

TGS algorithm: TGS = (100/26) * (*COLGALT1*$_{rs8090}$ + *COL1A1*$_{rs1800012}$ + *COL3A1*$_{rs1800255}$ + *COL5A1*$_{rs12722}$ + *COL5A1*$_{rs3196378}$ + *KDR*$_{rs1870377}$ + *MIR608*$_{rs4919510}$ + *MMP3*$_{rs679620}$ + *MMP3*$_{rs591058}$ *MMP3*$_{rs650108}$ + *NID1*$_{rs4660148}$ + *TIMP2*$_{rs4789932}$ + *VEGFA*$_{rs699947}$)

Previously associated elite status TGS model (3):

TGS algorithm: TGS = (100/14) * (*COLGALT1*$_{rs8090}$ + *COL3A1*$_{rs1800255}$ + *COL5A1*$_{rs12722}$ + *COL5A1*$_{rs3196378}$ + *MIR608*$_{rs4919510}$ + *MMP3*$_{rs591058}$ + *NID1*$_{rs4660148}$)

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

Table 4.1. Genotype score of each polymorphism based on literature, and genotype frequencies in elite rugby athletes and non-athletes from Chapter 3 (TGS model 1).

Gene abbreviation	rs number	Polymorphism	Genotype score (GS)	Frequency in non-athletes (%)	Frequency in elite rugby athletes (%)	Evidence
COLGALT1	8090	A/G	AA = 2, GA = 1, GG = 0	23.0, 47.1, 29.9	27.3, 48.7, 24.0	Strongest association within genome-wide association screen (P= 6 x10^{-4}) for ACL rupture. NB Not genome-wide significant (P <5 x 10^{-8}) (Kim et al., 2017).
COL1A1	1800012	A/C	AA = 2, AC = 1, CC = 0	3.2, 28.8, 68.0	2.2, 30.3, 67.5	AA genotype underrepresented in ACL injury group compared to uninjured (Khoschnau et al., 2008; Posthumus, 2009; Stępien-Słodkowska et al., 2013).
COL3A1	1800255	A/G	GG = 2, GA = 1, AA = 0	54.3, 39.2, 6.5	59.8, 33.7, 6.5	AA genotype overrepresented in ACL injury group (Stępień-Słodkowska et al., 2015a; O'Connell et al., 2015).
COL5A1	12722 &	C/T	CC = 2, CT = 1, TT = 0	22.9, 50.4, 26.7	17.3, 49.5, 33.2	CC genotype of rs12722 underrepresented in Achilles tendinopathy injury group compared to uninjured (Mokone et al., 2006; September et al., 2009; Brown et al., 2017) and tennis elbow (Altinisik et al., 2015) groups. CC genotype of rs3196378 underrepresented in Achilles tendinopathy group compared to uninjured (September et al., 2009; Brown et al., 2017).
	3196378	C/A	CC = 2, CA = 1, AA = 0	23.8, 47.4, 28.8	17.8, 49.4, 32.8	

KDR	1870377	A/T	TT = 2, TA = 1, AA = 0	57.3, 36.7, 6.0	54.8, 39.1, 6.1	A allele overrepresented in ACL injury group as part of inferred haplotype compared to uninjured (Rahim et al., 2014).
MIR608	4919510	G/C	GG = 2, GC = 1, CC = 0	5.8, 30.2, 64	8.7, 34.9, 56.4	CC genotype overrepresented in Achilles tendinopathy compared to uninjured (Abrahams et al., 2013).
MMP3	*591058,*	T/C	TT = 2, TC = 1, CC = 0	25.1, 51.8, 23.1	31.2, 47.3, 21.5	CC genotype of rs591058, AA genotype of rs650108 and GG (CC) genotype of rs679620 associated with Achilles tendinopathy (Raleigh et al., 2009). Additionally, rs679620 GG (CC) genotype and G (C) allele overrepresented in Achilles tendon rupture group compared to uninjured (El Khoury et al., 2016).
	650108	G/A	GG = 2, GA = 1, AA = 0	53.8, 39.6, 6.6	55.8, 36.5, 7.7	
	&					
	679620	T/C	TT = 2, TC = 1, CC = 0	23.8, 51.6, 24.6	28.7, 48.5, 22.8	
NID1	4660148	T/G	TT = 2, TG = 1, GG = 0	10.4, 42.1, 47.5	6.7, 45.1, 48.2	Strongest association within genome-wide association screen (P = 5 x10^{-5}) for Achilles tendon injury. NB: Not genome-wide significant (P <5 x 10^{-8}) (Kim et al., 2017).
TIMP2	4789932	G/A	GG = 2, GA = 1, AA = 0	33.4, 47.8, 18.8	35.1, 44.2, 20.7	GG genotype underrepresented in Achilles tendon pathology group compared to uninjured (El Khoury et al., 2013).
VEGFA	699947	C/A	CC = 2, CA = 1, AA = 0	27.8, 48.3, 23.9	24.6, 46.8, 28.6	CC genotype underrepresented in Achilles tendinopathy group compared to uninjured (Rahim et al., 2016).

Tendon and ligament injury-associated alleles in previous literature are underlined.

Table 4.2. Genotype score of each polymorphism based on elite rugby athlete allele frequency data from Chapter 3 (TGS Model 2).

Gene	rs number	Polymorphism	Genotype score (GS)
COLGALT1	8090	A/G	AA = 2, GA = 1, GG = 0
COL1A1	1800012	A/C	CC = 2, AC = 1, AA = 0
COL3A1	1800255	A/G	AA = 2, GA = 1, GG = 0
COL5A1	12722 & 3196378	C/T C/A	CC = 2, CT = 1, TT = 0 CC = 2, CA = 1, AA = 0
KDR	1870377	A/T	TT = 2, TA = 1, AA = 0
MIR608	4919510	G/C	CC = 2, GC = 1, GG = 0
MMP3	591058, 650108 & 679620	T/C G/A T/C	CC = 2, TC = 1, TT = 0 AA = 2, GA = 1, GG = 0 CC = 2, TC = 1, TT = 0
NID1	4660148	T/G	TT = 2, TG = 1, GG = 0
TIMP2	4789932	G/A	GG = 2, GA = 1, AA = 0
VEGFA	699947	C/A	CC = 2, CA = 1, AA = 0

Alleles that are overrepresented in non-athletes compared to rugby athletes are underlined and considered the 'risk' allele.

Table 4.3. Genotype score of each polymorphism based on previous associations in Chapter 3 with elite status in rugby (TGS model 3).

Gene	rs number	Polymorphism	Genotype score (GS)
COLGALT1	8090	A/G	AA = 2, GA = 1, GG = 0
COL3A1	1800255	A/G	AA = 2, GA = 1, GG = 0
COL5A1	12722 & 3196378	C/T C/A	CC = 2, CT = 1, TT = 0 CC = 2, CA = 1, AA = 0
MIR608	4919510	G/C	CC = 2, GC = 1, GG = 0
MMP3	591058	T/C	CC = 2, TC = 1, TT = 0
NID1	4660148	T/G	TT = 2, TG = 1, GG = 0

Alleles that are overrepresented in non-athletes compared to rugby athletes are underlined and considered the 'risk' allele.

4.2.3 Data analysis

TGS data for all groups was not normally distributed, therefore Mann-Whitney U tests were utilised to compare TGS differences between athlete groups and non-athletes. Means and extent of kurtosis were calculated to describe the distribution of TGS within groups. Pearson's Chi-square (χ^2) tests were used to compare the frequency of athletes and non-athletes in the top and bottom quartile of TGS scores. Bonferroni adjustment was utilised for each TGS approach where appropriate to control for false discovery. With 80% statistical power, analyses of all TGS models between all rugby athletes and non-athletes were able to detect a small effect size (w) of 0.12 and analysis between positional subgroups (RU forwards and RU backs) and non-athletes were able to detect a small effect size (w) of 0.15. The ability of the TGS to correctly distinguish elite athletes from non-athletes by receiver operating characteristic (ROC) curves was also evaluated, calculating the area under the curve (AUC) and 95% confidence intervals (95% CI). Multifactor dimensionality reduction (MDR) software (https://sourceforge.net/projects/mdr/) was used to calculate SNP-SNP epistasis interactions (Moore et al., 2006). Haplotypes were inferred using SNPStats (Solé et al., 2006). SPSS for Windows version 26 (SPSS, Chicago, IL) software was used for analysis. *P* values < 0.05 were considered statistically significant.

4.3 Results

TGS analyses

For the TGS based on prior literature (Model 1), there were no differences in TGS between any rugby athlete group and NA (RA vs NA, RU vs NA, RL vs NA, RU

forwards vs NA and RU backs vs NA) ($P \geq 0.076$). Mean (standard deviation) and kurtosis statistics are reported in Table 4.4, frequency distribution for RA and NA shown in Figure 4.1. Similarly, when the numbers of athletes (including discrete groups) and non-athletes in the upper and lower 25% of TGSs were compared, no significant differences were found. Accordingly, the TGS could not distinguish between elite rugby athletes and non-athletes (AUC = 0.519, 95% CI = 0.48 – 0.55; $P = 0.305$), nor between non-athletes and any other athlete group.

Table 4.4. TGS Mean and Kurtosis statistics for the three TGS models.

Group	**Mean (SD) TGS based on prior literature**	**Mean (SE) kurtosis statistic**
Non-athletes	49.8 (11.7)	-0.130, (0.233)
All Rugby Athletes	50.5 (10.6)	-0.063, (0.201)
RU athletes	50.8 (10.6)	-0.085, (0.218)
RL Athletes	48.9 (10.4)	-0.168, (0.488)
RU Forwards	50.3 (10.3)	0.289, (0.285)
RU Backs	51.6 (11.1)	-0.477, (0.336)
Group	**Mean (SD) TGS based on elite rugby athlete frequencies**	**Mean (SE) kurtosis statistic**
Non-athletes	48.7 (10.8)	0.096, (0.233)
All Rugby Athletes	52.1 (10.7)	-0.320, (0.201)
RU Athletes	52.0 (10.8)	-0.025, (0.218)
RL Athletes	52.4 (10.8)	-0.017, (0.488)
RU Forwards	52.4 (11.0)	-0.014, (0.285)
RU Backs	51.5 (10.4)	-0.041, (0.336)
Group	**Mean (SD) TGS based on SNPs previously associated with elite status in rugby**	**Mean (SE) kurtosis statistic**
Non-athletes	42.6 (12.9)	0.296, (0.233)
All Rugby Athletes	47.6 (13.4)	-0.161, (0.201)
RU Athletes	47.7 (13.4)	-0.173, (0.218)
RL Athletes	47.0 (13.3)	0.168, (0.488)
RU Forwards	48.5 (13.4)	-0.195, (0.285)
RU Backs	46.6 (13.4)	-0.227, (0.336)

RU = rugby union; RL = rugby league

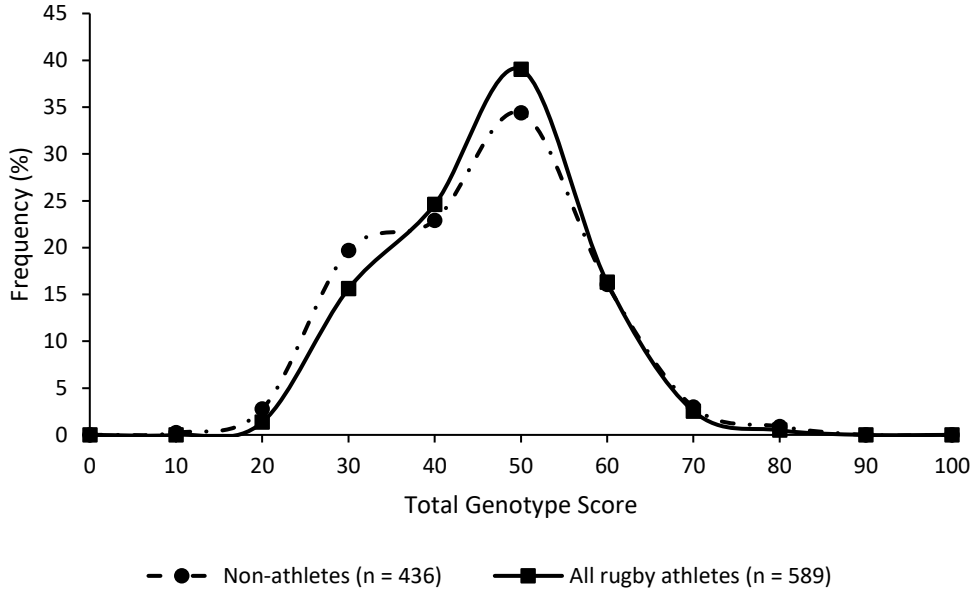

Figure 4.1. Frequency distribution of TGS based on prior literature as calculated from Table 4.1. Mean (SE) kurtosis statistics in all rugby athletes were -0.063 (0.201) and non-athletes was -0.130 (0.233).

For the TGS based on data from Chapter 3 in elite rugby athletes (Model 2), there was a significant difference between the TGS of RA compared to NA ($P = 8 \times 10^{-7}$), and this pattern was seen across all subgroups compared to NA (RU vs NA: $P = 4 \times 10^{-6}$, RL vs NA: $P = 0.002$, RU forwards vs NA: $P = 1 \times 10^{-5}$, RU backs vs NA: $P = 0.002$). Mean (standard deviation) and kurtosis statistics are reported in Table 4.4 and frequency distribution for RA and NA shown in Figure 4.2. When the top and bottom TGS quartiles were compared between athletes (including discrete groups) and non-athletes, athletes had a significantly higher frequency in the top 25% as well as a lower frequency within the bottom 25% TGS (Top 25%: RA = 54% vs NA = 31%; Bottom 25%: RA = 46% vs NA = 69%; $\chi^2 = 16.4$, $P = 5 \times 10^{-5}$). This persisted across all subgroups compared to non-athletes. Furthermore, the TGS could distinguish between elite rugby athletes and non-athletes (AUC = 0.59; 95% CI:

0.55-0.63; P = 9 x 10^{-7}, Figure 4.3). This pattern occurred across all subgroups compared to non-athletes.

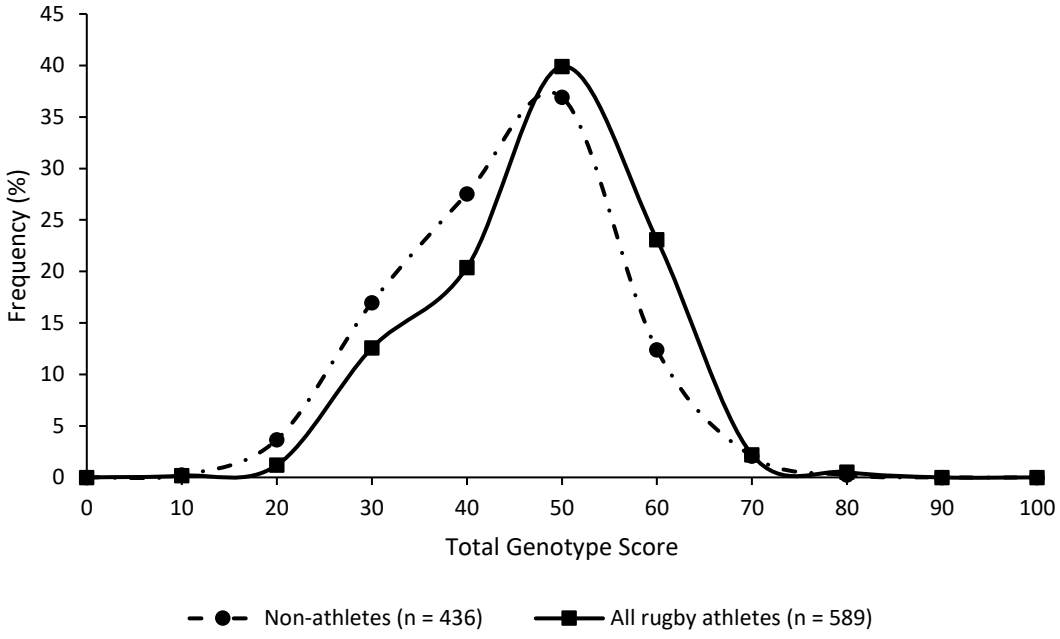

Figure 4.2. Frequency distribution of TGS based on elite rugby athlete data as calculated from Table 4.2. Mean (SE) kurtosis statistic in all rugby athletes was -0.320 (0.201) and non-athletes was 0.096 (0.233).

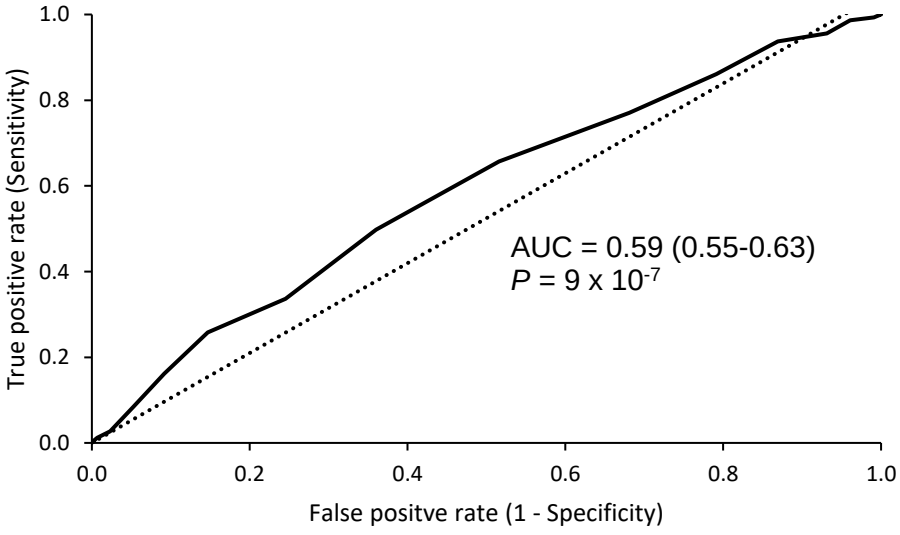

Figure 4.3. Receiver operating characteristic curve (ROC) summarizing the ability of TGS to classify elite status in rugby from non-athletes AUC indicates the area under the curve (95% confidence intervals).

For the TGS based on previous associations with elite status from Chapter 3 (Model 3), there was a significant difference between the TGS of RA compared to NA (P = 8 x 10^{-10}) and this pattern was seen across all subgroups compared to NA (RU vs NA: P = 2 x 10^{-9}, RL vs NA: P = 0.002, RU forwards vs NA: P = 8 x 10^{-9}, RU backs vs NA: P = 5 x 10^{-5}). Mean (standard deviation) and kurtosis statistics are reported in Table 4.4 and frequency distribution for RA vs NA is shown in Figure 4.4. When the frequency of athletes (including subgroups) and non-athletes in the upper and lower 25% TGSs were compared, athletes had a significantly higher frequency in the top 25% as well as a lower frequency within the bottom 25% (Top 25%: RA = 75% vs NA = 50%; Bottom 25%: RA = 25% vs NA = 50%; χ^2 = 12.1, P = 0.001). This was also found across all rugby union subgroups compared to non-athletes. Consequently, the TGS could distinguish between elite rugby athletes and non-athletes (AUC = 0.61; 95% CI: 0.58-0.65; P = 1 x 10^{-9}) and this persisted across all subgroups compared to non-athletes.

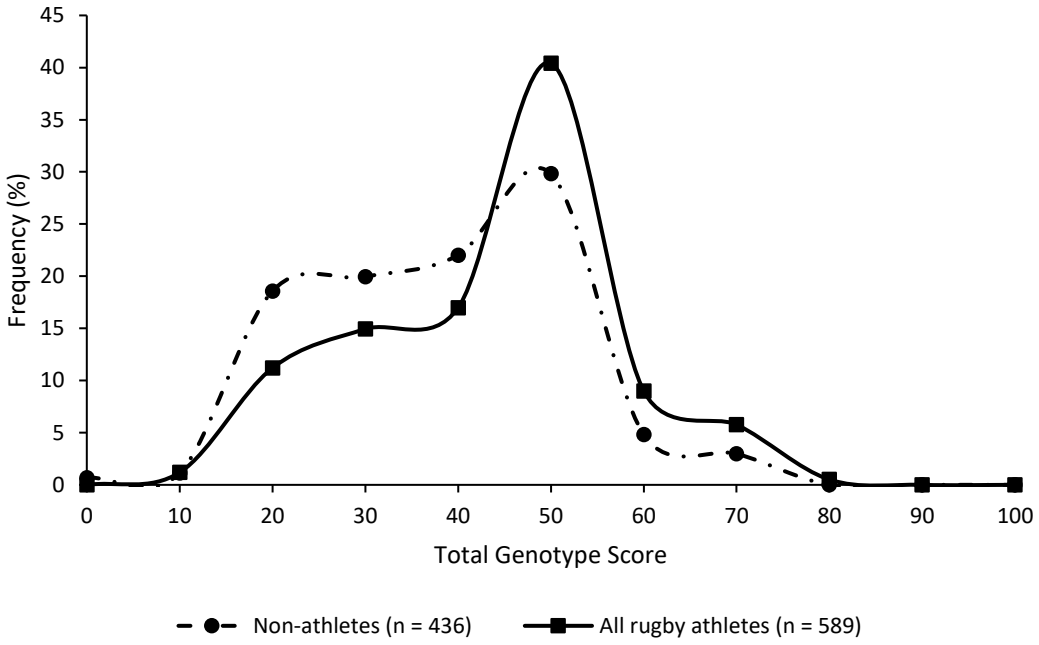

Figure 4.4. Frequency distribution of TGS based on SNPs previously associated with elite status in rugby as calculated from Table 4.3. Mean (SE) kurtosis statistic in All rugby athletes was -0.161 (0.201) and Non-athletes was 0.296 (0.233).

Haplotype and SNP epistasis analysis

Multifactor dimensionality reduction analysis found the *COL5A1* rs12722, *COL5A1* rs3196378 and *MIR608* rs4919510 polymorphisms produced the best model for predicting elite athlete status. There was a greater frequency of the CC-CC-CC genotype combination in RA (9.8%), RU (10.6%), RU backs (9.6%) and to the greatest extent RU forwards (11.3%; OR = 2.3; 95% CI = 1.3-3.9) than in NA (5.3%; all comparisons $P < 0.001$). Accordingly, the T-A inferred haplotype frequency constructed from *COL5A1* rs12722 and *COL5A1* rs3196378, was higher in NA than all athlete groups ($P < 0.022$; Figure 4.5). However, there were no differences for the *MMP3* T-G-A inferred haplotype frequency constructed from rs591058, rs650108 and rs679620, between athlete and non-athlete groups.

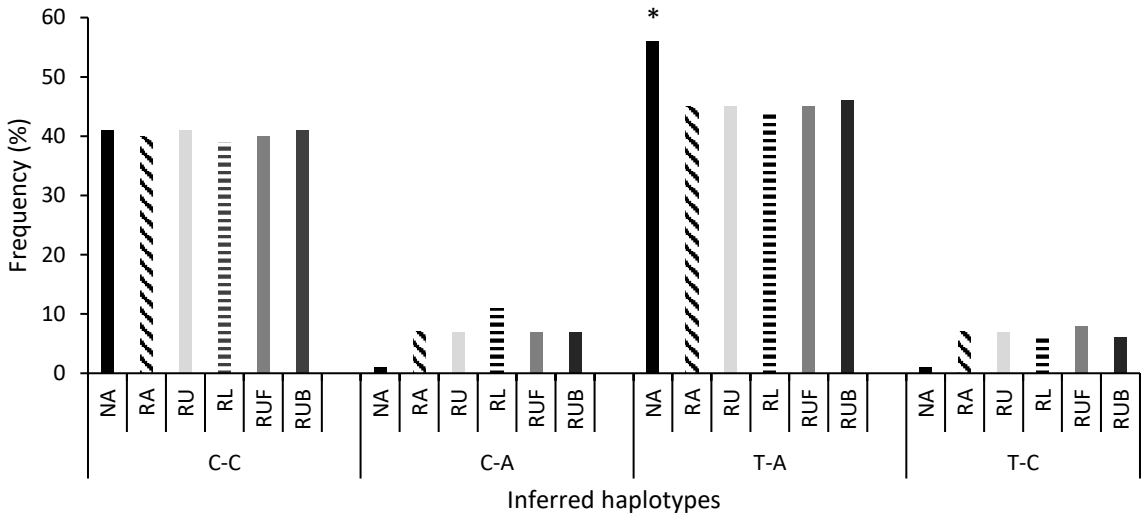

Figure 4.5. Inferred haplotype frequencies derived from *COL5A1* rs12722 and rs3196378. *different from all other groups ($P < 0.022$). RUF = Rugby union forwards, RUB = Rugby union backs.

4.4 Discussion

The present study investigated soft tissue injury associated genetic variants combined in polygenic profiles for determining elite status in rugby. For the 13 soft tissue injury associated polymorphisms used in TGS model 2 on elite rugby athlete data, differences were found between elite rugby athletes and non-athletes and across all subgroups (RA vs NA, RU vs NA, RL vs NA, RU forwards vs NA and RU backs vs NA), indicating that a likely inherited resistance to soft-tissue injury may enable career success. Furthermore, when seven polymorphisms previously associated with elite status in rugby were combined (Model 3), differences were found in TGS between elite rugby athletes and non-athletes and this was consistent for all subgroups. This highlights that a more precise polygenic approach utilising only polymorphisms previously associated with elite status may be more effective. Indeed, multifactor dimensionality reduction analysis found that a three SNP model

of *COL5A1* rs12722, *COL5A1* rs3196378 and *MIR608* rs4919510 produced the best model for predicting elite athlete status with a greater frequency of the CC-CC-CC genotype combination in elite rugby athletes which persisted across all subgroups.

The TGS model based on prior literature (Model 1) showed no difference between elite rugby athletes and non-athletes. These findings would suggest that elite rugby athletes do not carry 'preferable' soft-tissue injury associated polygenic profiles compared to non-athletes. A possible reason for these findings is the equivocal evidence within the prior literature regarding the identified 'risk' allele of each SNP. Of the 13 SNPs investigated only three (*COL1A1* rs1800012, *COL3A1* rs1800255 and *COL5A1* rs12722) have had their identified 'risk' alleles replicated in a separate Caucasian cohort (reviewed in Brazier et al. (2019)). Based on this TGS model, elite rugby athletes had a higher frequency (not all significant) of 'protective' alleles for seven of the SNPs compared to non-athletes, while for the remaining six polymorphisms elite rugby athletes were found to have a higher frequency (not all significant) of the 'risk' alleles compared to non-athletes. Therefore, a TGS model based on equivocal prior literature had potentially reduced the discriminatory power of the TGS. Hence, two further TGS models were employed within this study.

When the TGS algorithm was based on data on elite rugby athletes from Chapter 3 (Model 2), higher mean TGS were found between elite rugby athletes across all groups (RA, RU, RL, RU forwards and RU backs) and non-athletes. This aligns with previous research on elite athlete status with higher means and more optimal performance associated TGS found in elite cyclists and track and field athletes (Ruiz

et al., 2009; Eynon et al., 2011). Elite rugby athletes were chosen for this study due to their relatively consistent exposure to a high environmental soft-tissue injury risk compared to other sporting populations. However, the consistent higher mean TGS across rugby codes and playing positions compared to non-athletes suggests that variation within physical characteristics, environmental risks and training/match loads in elite rugby do not affect the genetic association with athlete status. Indeed, this was more apparent when the top and bottom TGS quartiles were compared, with elite rugby athletes more common in the top quartile and less common in the bottom quartile, which persisted across all groups. This suggests there may be a 'preferable' polygenic soft tissue injury associated profile for achieving elite status in rugby. For six of the included SNPs within this study, elite rugby athletes had higher frequencies (not all significant) of the perceived 'risk' allele. However, the evidence behind the 'risk' alleles and their mechanisms is relatively inconclusive and still needs to be fully elucidated. As such, it could be that some of the variants observed in elite rugby athletes may be of benefit to achieving elite athlete status via a mechanism other than injury protection.

The chances of an athlete carrying 'preferable alleles' related to soft tissue injury risk decrease as the number of polymorphisms included in the TGS increase. Furthermore, it has previously been reported that the inclusion of non-associated SNPs in a TGS model can reduce its accuracy (Yvert et al., 2016). Of the 13 SNPs investigated using the elite data-based TGS model (2), seven have been previously associated with elite status in Chapter 3. Hence, it is possible that the six non-associated polymorphisms included may have little or no influence on soft-tissue injury risk and/or ability to achieve elite athlete status. Therefore, a further TGS

model to improve the accuracy was utilised including only SNPs previously associated with elite athlete status (Model 3). When the seven SNPs previously associated with our athletes were combined, elite rugby athletes had a higher mean TGS than non-athletes and this pattern continued across all subgroups. Furthermore, the seven SNP TGS model (3) provided a better discriminating accuracy for elite status than the 13-SNP model (2) (AUC = 0.61; $P = 1 \times 10^{-9}$, AUC = 0.59; $P = 9 \times 10^{-7}$, respectively).

The likelihood that the three different TGS models included polymorphisms that possibly have no influence on soft tissue injury risk cannot be rejected. Furthermore, all three TGS models gave equal weighting to all polymorphisms, as it was assumed allelic effects would be codominant and thus each polymorphism would have an equal additional effect. It is unlikely that this is the case, as indicated by our MDR analysis that identified a three-SNP model of *COL5A1* rs12722, *COL5A1* rs3196378 and *MIR608* rs4919510 as best able to predict elite athlete status. Indeed, the CC-CC-CC genotype combination was more common in all athlete groups apart from RL, compared to non-athletes. Interactions between *COL5A1* and *MIR608* and risk of Achilles tendinopathy have previously been reported (Abrahams et al., 2013). Furthermore, the C alleles of *COL5A1* rs12722 and *COL5A1* rs3196378 have been identified as ‘protective’ from Achilles tendinopathy and ligament injury (September et al., 2009; Posthumus, 2009; O'Connell et al., 2015), which suggests that the C allele may be beneficial in protecting against tendon and ligament injuries. Conversely, the CC genotype of *MIR608* rs4919510 was found to be associated with Achilles tendinopathy (Abrahams et al., 2013), however these results have not been replicated. To date,

how these polymorphisms might influence injury risk has not been demonstrated mechanistically. Nevertheless, it appears that a combination of *COL5A1* rs12722, *COL5A1* rs3196378 and *MIR608* rs4919510 may be beneficial for rugby athletes to achieve elite status, possibly through a greater resistance to soft-tissue injury. This is further supported by the higher *COL5A1* T-A inferred haplotype frequency constructed from rs12722 and *COL5A1* rs3196378 in non-athletes than all athlete groups.

4.5 Conclusion

This study presents the first polygenic models that discriminate between elite rugby athlete status and non-athletes based on polymorphisms previously associated with soft tissue injury risk. The present study observed that rugby athletes possess polygenic profiles that appear to have enabled them to attain elite status despite training and competing in the high injury risk environment of rugby. However, this study is not without limitations, particularly regarding the limited evidence to support the true 'risk' alleles of each polymorphism. Future research should seek to identify additional polymorphisms that have strong associations with soft-tissue injury risk, because including them in a polygenic model will enhance discriminatory ability. Additionally, analysing genetic data with injury data from rugby matches and training would further enhance this area. Understanding the genetics of soft tissue injury risk would be an important advance towards injury risk stratification in sport and a possible future use of polygenic profiles in screening and management of injury risk.

Chapter 5

Tendon and ligament injury-associated gene variants and history of soft tissue injury in elite rugby athletes

5.1 Introduction

Chapters 1, 3 and 4 have previously discussed the intermittent and collision-based nature of rugby that likely influences injury risk. Indeed, as previously discussed in Chapter 1, rugby has one of the highest incidence and severity rates compared to other team sports (Williams et al., 2013; Whitehouse et al., 2016). Most of these injuries are to ligaments, tendon and muscle (Williams et al., 2013; Schwellnus et al., 2014).

At the most recent rugby union world cup (2019), ligament injuries were found to be the most common accounting for 21.7% of all injuries reported (Fuller et al., 2020). Furthermore, knee ligament injuries were the most severe causing 935 days absence (Fuller et al., 2020). A similar pattern is seen within English premiership rugby with anterior cruciate ligament (ACL) and/or medial collateral ligament injuries consistently in the top 5 most severe injuries across seasons 2011-19, with ACL injuries costing 259 days per 1000 hours to be missed from matches in the 2018-19 season (Kemp et al., 2020). Tendon injuries although not as common and severe as ligament injuries are a very debilitating injury for rugby players. Specifically, Achilles tendon (AT) injuries were in the top 5 highest burden training injuries across the 2018-19 English rugby premiership season, costing 4.2 days absence per 1000 hours (Kemp et al., 2020). Achilles' tendon injury appears to be particularly debilitating for rugby union forward athletes causing 726 days absence across two seasons (Brooks et al., 2005b). Both ligament and tendon injuries are highly complex, multifactorial disorders which are determined by the interaction of several extrinsic and intrinsic factors (Nourissat et al., 2015; Smith et al., 2012b; Smith et

al., 2012a). However, the growing body of evidence around the familial aggregation seen in ligament and tendon injuries (Flynn et al., 2005; Harner et al., 1994; Magnusson et al., 2020; Hakim et al., 2003) has prompted further research into possible genetic linkage.

Over the last ~25 years, numerous studies have examined genetic factors that potentially predispose an individual to ligament injury (Harner et al., 1994; Flynn et al., 2005; Magnusson et al., 2020; Khoschnau et al., 2008; Posthumus, 2009; Posthumus et al., 2012; Posthumus et al., 2009c; Stępien-Słodkowska et al., 2013; Ficek et al., 2013; Ficek et al., 2014; Rahim et al., 2014; O'Connell et al., 2015; Stępień-Słodkowska et al., 2015b; Stępień-Słodkowska et al., 2015a; Kim et al., 2017). ACL tears seem at least twice as likely in individuals with a family history of ACL tear compared to those with no family history (Harner et al., 1994; Flynn et al., 2005). Indeed, a recent twin study found that the genetic contribution to ACL rupture was ~69% (Magnusson et al., 2020). Most research into the genetics of ligament injury has utilised GAS. From these studies, variants in several genes have been associated with altered risk of ligament injury; *COL1A1* (Posthumus et al., 2009c; Ficek et al., 2013; Stępien-Słodkowska et al., 2013; Khoschnau et al., 2008), *COL3A1* (O'Connell et al., 2015; Stępień-Słodkowska et al., 2015a), *COL5A1* (Posthumus, 2009; Stępień-Słodkowska et al., 2015b), *MMP3* (Malila et al., 2011), *MMP12* (Posthumus et al., 2012), and the angiogenesis-associated signalling pathway genes *VEGFA* and *KDR* (Rahim et al., 2014). Similarly, it has been suggested that there is a genetic component to the aetiology of tendon injury. Indeed, in a twin study of tennis elbow (epicondylitis) in women (Hakim et al., 2003), heritability was estimated at ~40%. Furthermore, several GAS report associations

between Achilles tendinopathy and several genetic variants across a variety of genes such as *COL5A1* (Mokone et al., 2006; Abrahams et al., 2013; Khoury El et al., 2014), *MMP3* (Raleigh et al., 2009; El Khoury et al., 2016), *TIMP2* (El Khoury et al., 2013; El Khoury et al., 2016), and microRNA-608 (*MIR608*) (Abrahams et al., 2013). Additionally, Achilles tendon rupture has been associated with variants in the *MMP3* and *TIMP2* genes (El Khoury et al., 2016).

Genetic variation may have a strong influence on tendon and ligament structure and function, which could alter an individual's risk of injury. This could influence individual injury incidence and severity rates, thus affecting time lost from matches and training as well as potentially affecting early career retirement. Ficek et al. (2013) found potential evidence of this within professional footballers, identifying the *COL1A1* G-T haplotype (rs1107946 and rs1800012, respectively) to be associated with reduced risk of ACL injury. However, no association between risk of ACL rupture and several collagen gene variants were found in a cohort of elite female athletes from high-risk team sports (Sivertsen et al., 2019), so further elucidation is necessary.

Therefore, investigating the genetic components of soft-tissue injury within an elite rugby population is a worthy proposition as it may enhance the current knowledge and understanding of rugby-specific injury aetiology. Past research has shown that elite rugby athletes possess more 'favourable' genetic variants previously associated with soft-tissue injury both individually and collectively compared to a non-athlete population (Heffernan et al., 2017a) (Chapters 3 and 4). However, there

is no previous research directly investigating genetic characteristics of soft tissue injury within an elite rugby athlete population. Thus, the aims of this study were firstly to investigate whether gene variants previously associated with soft tissue injury risk are associated with history of soft tissue injury in elite rugby athletes. Secondly, to compare polygenic characteristics between athletes with a history of soft tissue injury and those with no history of injury. It was hypothesised that the soft tissue injury risk associated genotypes and alleles would be overrepresented in elite rugby athletes with a history of injury compared to those with no previous history. Furthermore, it was hypothesised that athletes with no history of soft tissue injury would have a higher TGS than athletes with a history of injury.

5.2 Method

5.2.1 Participants

This study was conducted in accordance with the STROBE guidelines for a case-control observational study (von Elm et al., 2007). 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 = 133; mean (standard deviation) height 1.86 (0.07) m, mass 102 (12) kg, age 26 (5) yr) including 111 RU athletes and 22 RL athletes who were 41.8% Italian, 25.5% Irish, 23.4% British, 5.0% South African, and 4.3% of other nationalities, having given written informed consent. For TGS and

SNP-SNP epistasis interaction analyses, 117 elite rugby athletes were utilised as 16 athletes did not have a full data set for all 13 polymorphisms.

5.2.2 Procedures

Sample collection and genotyping via real-time PCR were consistent with those reported in Chapter 2 and thesis sections 3.2.2 and 4.2.2.

Soft-tissue injury history

Soft tissue injury history in elite rugby athletes was collected utilising a self-reported injury-history questionnaire. Athletes were asked to provide details of their geographic ancestry, playing position, playing history, highest level of play, tendon and ligament injury incidence, mechanism of injury and whether each injury was medically diagnosed (See Appendix 1). The questionnaire took ~15 min to complete, and an investigator assisted participants to maximise accuracy.

Calculation of TGS

To quantify the combined influence of the candidate polymorphisms (Table 4.1) an additive TGS algorithm was utilised (Williams and Folland, 2008), based on the assumption of codominance effects of the alleles. The homozygote genotypes with the lower soft tissue injury risk, according to the prior literature (cf. TGS model 1 in Chapter 4) were allocated a ‘genotype score’ of 2, heterozygote genotypes were

scored 1 and the higher soft tissue injury risk homozygote genotypes were scored 0.

TGS model (1)

TGS = (100/26) * (*COLGALT1*$_{rs8090}$ + *COL1A1*$_{rs1800012}$ + *COL3A1*$_{rs1800255}$ + *COL5A1*$_{rs12722}$ + *COL5A1*$_{rs3196378}$ + *KDR*$_{rs1870377}$ + *MIR608*$_{rs4919510}$ + *MMP3*$_{rs679620}$ + *MMP3*$_{rs591058}$ + *MMP3*$_{rs650108}$ + *NID1*$_{rs4660148}$ + *TIMP2*$_{rs4789932}$ + *VEGFA*$_{rs699947}$)

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

5.2.3 Data Analysis

Pearson's Chi-square (χ^2) tests were used to compare genotype (using three analysis models: additive, recessive, and dominant) and allele frequencies between injured athletes and non-injured athletes across all injury groups (tendon rupture vs no tendon injury, tendinopathy vs no tendon injury, ligament rupture vs no ligament injury, ligament sprain vs no ligament injury and all injured vs all non-injured). With 80% statistical power, analyses of all injured compared with non-injured athletes were able to detect a small-moderate effect size (w) of 0.27 and analysis between injury subgroups were able to detect a moderate effect size (w) of 0.33. For each polymorphism, 20 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. Total genotype score for all groups was not normally distributed, therefore Mann-Whitney U tests

were utilised to compare TGS between injured athlete groups and non-injured. Means and extent of kurtosis were calculated to describe the distribution of TGS within groups. Pearson's χ^2 tests were used to compare the frequency of injured athletes and non-injured athletes in the top and bottom thirds of TGS scores. Bonferroni adjustment was utilised where appropriate to control for false discovery. With 80% statistical power, analyses of all TGS models between all injured athletes and non-injured athletes were able to detect a moderate effect size (w) of 0.33 and analysis between injury subgroups and non-injured athletes were able to detect a moderate effect size (w) of 0.38. The ability of the TGS to correctly distinguish injured athletes from non-injured athletes across all groups was evaluated by receiver operating characteristic (ROC) curves (Zweig and Campbell, 1993), calculating the area under the curve (AUC) and 95% confidence intervals (95% CI). Multifactor dimensionality reduction (MDR) software (https://sourceforge.net/projects/mdr/) was used to identify SNP-SNP epistasis interactions (Moore et al., 2006). Haplotypes were inferred using SNPStats (Solé et al., 2006). SPSS for Windows version 26 (SPSS, Chicago, IL) software was used for analysis. *P* values < 0.05 were considered statistically significant.

5.3 Results

Genotype frequencies were in Hardy-Weinberg equilibrium for all polymorphisms in the injured and non-injured athlete groups apart from *COL5A1* rs12722 (non-injured tendon and ligament sprain groups) and *COL5A1* rs3196378 (non-injured tendon, tendinopathy, ligament sprain and all injured groups) (Table 5.1).

Table 5.1. Genotype and allele distribution of injured and non-injured athletes separated by specific injury, presented as genotype/allele counts followed by percentage in parentheses.

Polymorphism	Genotype/ Hardy-Weinberg Equilibrium (HWE)	Non-injured tendon (NIT)	Tendon rupture (TR)	Tendinopathy (TY)	Non-injured ligament (NIL)	Ligament rupture (LR)	Ligament sprain (LS)	All non-injured (ANI)	All injured (AI)
COLGALT1 rs8090									
	GG	21 (24.4)	5 (29.4)	9 (27.3)	11 (30.6)	14 (27.5)	19 (22.4)	6 (24.0)	28 (25.9)
	GA	46 (53.5)	10 (58.8)	17 (51.5)	20 (55.6)	27 (52.9)	46 (54.1)	15 (60.0)	56 (51.9)
	AA	19 (22.1)	2 (11.8)	7 (21.2)	5 (13.9)	10 (19.6)	20 (23.5)	4 (16.0)	24 (22.2)
	Total	86	17	33	36	51	85	25	108
	G allele carriers	67 (77.9)	15 (88.2)	26 (78.8)	31 (86.1)	41 (80.3)	65 (76.5)	21 (84.0)	81 (75.0)
	A allele carriers	65 (75.6)	12 (70.6)	24 (72.7)	25 (69.4)	37 (72.5)	66 (77.6)	19 (76.0)	80 (74.1)
	G allele	88 (51.2)	20 (58.8)	35 (53.0)	42 (58.3)	55 (53.9)	84 (49.4)	27 (54.0)	112 (57.1)
	A allele	84 (48.8)	14 (41.2)	31 (47.0)	30 (41.7)	47 (46.1)	86 (50.6)	23 (46.0)	84 (42.9)
	HWE *P* value & χ^2	>0.25, 0.43	>0.25, 0.78	>0.75, 0.04	>0.25, 0.73	>0.50, 0.22	>0.25, 0.58	>0.25, 1.08	>0.50, 0.16
COL1A1 rs1800012									
	CC	59 (69.4)	16 (76.2)	24 (70.6)	25 (67.6)	38 (71.7)	59 (67.0)	17 (70.8)	78 (70.3)
	CA	22 (25.9)	5 (23.8)	9 (26.5)	9 (24.3)	13 (24.5)	25 (28.4)	7 (29.2)	28 (25.2)
	AA	4 (4.7)	0 (0.0)	1 (2.9)	0 (0.0)	2 (3.8)	4 (4.5)	0 (0.0)	5 (4.5)
	Total	85	21	34	37	53	88	24	111
	C allele carriers	81 (95.3)	21 (100.0)	33 (97.1)	34 (91.9)	51 (96.2)	84 (95.5)	24 (100.0)	106 (95.5)
	A allele carriers	26 (30.6)	5 (23.8)	10 (29.1)	9 (24.3)	15 (28.3)	29 (32.9)	7 (29.2)	33 (29.7)
	C allele	140 (82.4)	37 (88.1)	57 (83.9)	59 (86.8)	89 (83.9)	143 (81.3)	41 (85.4)	184 (82.9)
	A allele	30 (17.6)	5 (11.9)	11 (16.1)	9 (13.2)	17 (16.1)	33 (18.7)	7 (14.6)	38 (17.1)
	HWE *P* value & χ^2	>0.25, 1.02	>0.50, 0.38	>0.75, 0.20	>0.25, 0.79	>0.50, 0.42	>0.50, 0.40	>0.25, 0.70	>0.10, 1.37

COL3A1 rs1800255									
	GG	46 (54.1)	12 (57.1)	16 (47.0)	20 (55.6)	27 (51.9)	46 (53.5)	12 (48.0)	60 (54.5)
	GA	33 (38.8)	7 (33.3)	16 (47.0)	14 (38.9)	22 (42.3)	35 (40.7)	12 (48.0)	42 (38.2)
	AA	6 (7.1)	2 (9.5)	2 (5.8)	2 (5.6)	3 (5.8)	5 (5.8)	1 (4.0)	8 (7.3)
	Total	85	21	34	36	52	86	25	110
	G allele carriers	79 (92.9)	19 (90.5)	32 (94.1)	34 (94.4)	49 (94.2)	81 (94.2)	24 (96.0)	102 (92.7)
	A allele carriers	39 (48.9)	9 (42.9)	18 (52.9)	16 (44.4)	25 (48.1)	40 (46.5)	13 (52.0)	50 (45.5)
	G allele	125 (73.5)	31 (73.8)	48 (70.6)	54 (75.0)	76 (73.1)	127 (73.8)	36 (72.0)	162 (73.6)
	A allele	45 (26.5)	11 (26.2)	20 (29.4)	18 (25.0)	28 (26.9)	45 (26.2)	14 (28.0)	58 (26.4)
	HWE *P* value & χ^2	>0.99, 0.001	>0.50, 0.40	>0.25, 0.60	>0.75, 0.05	>0.50, 0.29	>0.50, 0.24	>0.25, 0.91	>0.75, 0.03
COL5A1 rs12722									
	TT	22 (25.9)	3 (15.8)	12 (35.3)	6 (17.1)	17 (33.3)	26 (30.2)*	5 (20.8)	30 (27.5)
	TC	32 (37.6)	12 (63.2)	14 (41.2)	17 (48.6)	24 (47.1)	32 (37.2)	10 (41.7)	46 (42.2)
	CC	31 (36.5)	4 (21.1)	8 (23.5)	12 (34.3)	10 (19.6)*	28 (32.6)	9 (37.5)	33 (30.3)
	Total	85	19	34	35	51	86	24	109
	T allele carriers	54 (63.5)	15 (78.9)	26 (76.4)	23 (65.7)	41 (80.4)	58 (67.4)	15 (62.5)	76 (69.7)
	C allele carriers	63 (74.1)	16 (84.2)	22 (64.7)	29 (85.9)	34 (66.6)*	60 (69.8)*	19 (79.2)	79 (72.5)
	T allele	76 (44.7)	18 (47.4)	38 (55.9)	29 (41.4)	58 (56.9)	84 (48.8)	20 (41.7)	106 (48.6)
	C allele	94 (55.3)	20 (52.6)	30 (44.1)	41 (58.6)	44 (43.1)*	88 (51.2)	28 (58.3)	112 (51.4)
	HWE *P* value & χ^2	>0.05[#], 4.84	>0.10, 1.35	>0.75, 0.92	>0.10, 2.5	>0.75, 0.08	>0.05[#], 5.61	>0.25, 0.49	>0.10, 2.63
COL5A1 rs3196378									
	AA	33 (38.8)	4 (22.2)	13 (39.4)	10 (29.4)	18 (36.0)	36 (41.9)	7 (29.2)	42 (39.3)
	AC	29 (34.1)	9 (50.0)	10 (30.3)	16 (47.1)	19 (38.0)	27 (31.4)	11 (45.8)	36 (33.6)
	CC	23 (27.1)	5 (27.8)	10 (30.3)	8 (23.5)	13 (26.0)	23 (26.7)	6 (25.0)	29 (27.1)
	Total	85	18	33	34	50	86	24	107
	A allele carriers	62 (72.9)	13 (72.2)	23 (69.7)	26 (76.5)	37 (74.0)	63 (73.3)	18 (75.0)	78 (72.9)

	C allele carriers	52 (61.2)	14 (77.7)	20 (60.6)	24 (70.6)	32 (64.0)	50 (58.1)	17 (70.8)	65 (60.7)
	A allele	95 (55.9)	17 (47.2)	36 (54.6)	36 (52.9)	55 (55.0)	99 (57.6)	25 (52.1)	120 (56.1)
	C allele	75 (44.1)	19 (52.8)	30 (45.5)	32 (47.1)	45 (45.0)	73 (42.5)	23 (47.9)	94 (43.9)
	HWE *P* value & χ^2	>0.01[#], 8.10	>0.99, 0.0001	>0.05[#], 4.99	>0.50, 0.10	>0.10, 2.70	>0.01[#], 10.99	>0.50, 0.16	>0.01[#], 10.75
KDR rs1870377									
	TT	48 (55.2)	11 (52.4)	20 (57.1)	20 (24.1)	28 (52.8)	48 (54.5)	14 (53.8)	61 (54.5)
	TA	32 (36.8)	9 (42.9)	11 (31.4)	13 (35.1)	20 (37.7)	33 (37.5)	10 (38.5)	41 (36.6)
	AA	7 (8.0)	1 (4.8)	4 (11.4)	4 (10.8)	5 (9.4)	7 (8.0)	2 (7.7)	10 (8.9)
	Total	87	21	35	37	53	88	26	112
	T allele carriers	80 (91.9)	20 (95.2)	31 (88.6)	33 (89.2)	48 (90.6)	81 (92.0)	24 (92.3)	102 (91.1)
	A allele carriers	39 (44.8)	10 (47.6)	15 (42.9)	17 (45.9)	25 (47.2)	40 (45.5)	12 (46.2)	51 (45.5)
	T allele	128 (73.6)	31 (73.8)	51 (72.9)	53 (71.6)	76 (71.7)	129 (73.3)	38 (73.1)	163 (72.8)
	A allele	46 (26.4)	11 (26.2)	19 (27.1)	21 (29.4)	30 (28.3)	47 (26.7)	14 (26.9)	61 (27.2)
	HWE *P* value & χ^2	>0.50, 0.26	>0.50, 0.25	>0.10, 1.48	>0.25, 0.68	>0.50, 0.26	>0.10, 0.16	>0.90, 0.01	>0.25, 0.65
MIR608 rs4919510									
	CC	40 (48.2)	6 (30.0)	18 (51.4)	16 (45.7)	19 (37.3)	44 (50.6)	10 (41.7)	53 (48.6)
	CG	30 (36.1)	12 (60.0)	13 (37.1)	14 (40.0)	25 (49.0)	33 (37.9)	9 (37.5)	44 (40.4)
	GG	13 (15.7)	2 (10.0)	4 (11.4)	5 (14.3)	7 (13.7)	10 (11.5)	5 (20.8)	12 (11.0)
	Total	83	20	35	35	51	87	24	109
	C allele carriers	70 (84.3)	18 (90.0)	31 (88.6)	30 (85.7)	44 (86.3)	77 (88.5)	19 (79.2)	97 (88.9)
	G allele carriers	43 (81.8)	14 (70.0)	17 (48.6)	19 54.3)	32 (62.7)	43 (49.4)	14 (58.3)	56 (51.4)
	C allele	110 (66.3)	24 (60.0)	49 (70.0)	46 (65.7)	63 (61.8)	121 (69.5)	29 (60.4)	150 (68.8)
	G allele	56 (33.7)	16 (40.0)	21 (30.0)	24 (34.3)	39 (38.2)	53 (30.5)	19 (39.6)	68 (31.2)
	HWE *P* value & χ^2	>0.10, 3.05	>0.25, 1.25	>0.25, 0.49	>0.50, 0.44	>0.75, 0.07	>0.75, 0.95	>0.25, 1.12	>0.50, 0.39
MMP3 rs591058									
	TT	21 (24.4)	4 (19.0)	6 (17.1)	7 (19.4)	13 (24.5)	19 (21.6)	4 (16.0)	27 (24.1)

	TC	37 (43.0)	8 (38.1)	16 (45.7)	17 (47.2)	21 (39.6)	37 (42.0)	12 (48.0)	47 (42.0)
	CC	28 (32.6)	9 (42.9)	13 (37.1)	12 (33.3)	19 (35.8)	32 (36.4)	9 (36.0)	38 (33.9
	Total	86	21	35	36	53	88	25	112
	T allele carriers	58 (67.4)	12 (57.1)	22 (62.9)	24 (66.7)	34 (64.2)	56 (63.6)	16 (64.0)	74 (66.1)
	C allele carriers	65 (75.6)	17 (80.9)	29 (82.8)	29 (80.6)	40 (75.5)	69 (78.4)	21 (84.0)	85 (75.9)
	T allele	79 (45.9)	16 (38.1)	28 (40.0)	31 (43.1)	47 (44.3)	75 (42.6)	20 (40.0)	101 (45.1)
	C allele	93 (54.1)	26 (61.9)	42 (60.0)	41 (56.9)	59 (55.7)	101 (57.4)	30 (60.0)	123 (54.9)
	HWE *P* value & χ^2	>0.10, 1.54	>0.25, 0.77	>0.75, 0.08	>0.75, 0.05	>0.10, 2.06	>0.10, 1.73	>0.99, >0.01	>0.10, 2.60
MMP3 rs650108									
	GG	40 (47.6)	8 (38.1)	13 (38.2)	15 (41.7)	24 (46.2)	39 (45.3)	10 (40.0)	51 (46.8)
	GA	38 (45.2)	9 (42.9)	16 (47.1)	17 (47.2)	22 (42.3)	39 (45.3)	13 (52.0)	47 (43.1)
	AA	6 (7.1)	4 (19.0)	5 (14.7)	4 (11.1)	6 (11.5)	8 (9.4)	2 (8.0)	11 (10.1)
	Total	84	21	34	36	52	86	25	109
	G allele carriers	78 (92.9)	17 (80.9)	29 (85.3)	32 (88.9)	46 (88.5)	78 (90.7)	23 (92.0)	98 (89.9)
	A allele carriers	44 (52.4)	13 (61.9)	21 (61.8)	21 (58.3)	28 (53.8)	47 (54.6)	15 (60.0)	58 (53.2)
	G allele	118 (70.2)	25 (59.5)	42 (61.8)	47 (65.3)	70 (67.3)	117 (68.0)	33 (66.0)	149 (68.4)
	A allele	50 (29.8)	17 (40.5)	26 (38.2)	25 (34.7)	34 (32.7)	55 (32.0)	17 (34.0)	69 (31.7)
	HWE *P* value & χ^2	>0.25, 0.56	>0.50, 0.25	>0.95, >0.01	>0.75, 0.06	>0.75, 0.07	>0.50, 0.15	>0.25, 0.62	>0.95, >0.01
MMP3 rs679620									
	TT	20 (23.5)	3 (14.3)	2 (5.7)*	3 (8.6)	13 (24.5)*	17 (19.3)*	2 (8.3)	23 (20.5)*
	TC	36 (42.4)	8 (38.1)	16 (45.7)	17 (48.6)	21 (39.6)	36 (40.9)	12 (50.0)	46 (41.1)
	CC	29 (34.1)	10 (47.6)	17 (48.6)	15 (42.8)	19 (35.8)	35 (39.8)	10 (41.7)	43 (38.4)
	Total	85	21	35	35	53	88	24	112
	T allele carriers	56 (65.9)	11 (52.4)	18 (51.4)	20 (57.1)	34 (64.2)	53 (60.2)	14 (58.3)	69 (61.7)
	C allele carriers	65 (76.5)	18 (85.7)	33 (94.3)*	32 (91.4)	40 (75.5)*	71 (80.7)*	22 (91.7)	89 (79.5)*
	T allele	76 (44.7)	14 (33.3)	20 (28.6)*	23 (32.9)	47 (44.3)*	70 (39.8)*	16 (33.3)	92 (41.1)*
	C allele	94 (55.3)	28 (66.7)	50 (71.4)	47 (67.1)	59 (55.7)	106 (60.2)	32 (66.7)	132 (58.9)

	HWE *P* value & χ^2	>0.10, 1.75	>0.50, 0.43	>0.25, 0.50	>0.50, 0.36	>0.10, 2.06	>0.10, 1.88	>0.50, 0.38	>0.10, 2.58
NID1 rs4660148									
	GG	46 (52.9)	8 (38.1)	17 (48.6)	19 (51.4)	29 (54.7)	41 (46.6)	12 (46.2)	56 (50.0)
	GT	36 (41.4)	8 (38.1)	15 (42.9)	17 (45.9)	17 (32.1)	35 (39.8)	13 (50.0)	44 (39.3)
	TT	5 (5.7)	5 (23.8)*	3 (8.6)	1 (2.7)	7 (13.2)*	12 (13.6)*	1 (3.8)	12 (10.7)*
	Total	87	21	35	37	53	88	26	112
	G allele carriers	82 (94.3)	16 (76.2)*	32 (91.4)	36 (97.3)	46 (86.8)*	76 (86.3)*	25 (96.2)	100 (89.3)*
	T allele carriers	51 (58.6)	13 (61.9)	18 (51.4)	18 (48.6)	24 (45.3)	47 (53.4)	14 (53.8)	56 (50.0)
	G allele	128 (73.6)	24 (57.1)	49 (70.0)	55 (74.3)	75 (70.8)	117 (66.5)	37 (71.2)	156 (69.6)
	T allele	46 (26.4)	18 (42.9)*	21 (30.0)	19 (25.7)	31 (29.2)	59 (33.5)*	15 (28.8)	68 (30.4)
	HWE *P* value & χ^2	>0.50, 0.35	>0.25, 1.04	>0.90, 0.01	>0.10, 1.54	>0.10, 2.68	>0.25, 1.02	>0.25, 1.24	>0.25, 0.56
TIMP2 rs4789932									
	GG	23 (27.1)	6 (30.0)	11 (32.4)	14 (38.9)	13(25.5)	22 (25.6)	10 (40.0)	29 (26.6)*
	GA	41 (48.2)	12 (60.0)	19 (55.9)	14 (38.9)	27 (52.9)	47 (54.6)*	10 (40.0)	58 (53.2)
	AA	21 (24.7)	2 (10.0)	4 (11.8)	8 (22.2)	11 (21.6)	17 (19.8)	5 (20.0)	22 (20.2)
	Total	85	20	34	36	51	86	25	109
	G allele carriers	64 (75.3)	18 (90.0)	30 (88.2)	28 (77.8)	40 (78.4)	69 (80.2)	20 (80.0)	87 (79.8)
	A allele carriers	62 (72.9)	14 (70.0)	23 (67.6)	22 (61.1)	38 (74.5)	64 (74.4)	15 (60.0)	80 (73.4)*
	G allele	87 (51.2)	24 (60.0)	41 (60.3)	42 (58.3)	53 (52.0)	91 (52.9)	30 (60.0)	116 (53.2)
	A allele	83 (48.8)	16 (40.0)	27 (39.7)	30 (41.7)	49 (48.0)	81 (47.1)	20 (40.0)	102 (46.8)
	HWE *P* value & χ^2	>0.75, 0.10	>0.25, 1.25	>0.25, 0.95	>0.10, 1.44	>0.50, 0.19	>0.25, 0.80	>0.25, 0.69	>0.25, 0.51
VEGFA rs699947									
	CC	26 (30.6)	10 (55.6)	8 (23.5)	9 (25.0)	15 (31.2)	30 (34.9)	6 (24.0)	34 (31.8)
	CA	38 (44.7)	6 (33.3)	21 (61.8)	20 (55.6)	23 (47.9)	36 (41.9)	14 (56.0)	50 (46.7)
	AA	21 (24.7)	2 (11.1)	5 (14.7)	7 (19.4)	10 (20.8)	20 (23.3)	5 (20.0)	23 (21.5)
	Total	85	18	34	36	48	86	25	107

	C allele carriers	64 (75.3)	16 (88.9)	29 (85.3)	29 (80.6)	38 (79.2)	66 (76.7)	20 (80.0)	84 (78.5)
	A allele carriers	59 (69.4)	8 (44.4)	26 (76.5)	27 (75.0)	33 (68.8)	56 (65.1)	19 (76.0)	73 (68.2)
	C allele	90 (52.9)	26 (72.2)	37 (54.4)	38 (52.8)	53 (55.2)	96 (55.8)	26 (52.0)	118 (55.1)
	A allele	80 (47.1)	10 (27.8)	31 (45.6)	34 (47.2)	43 (44.8)	76 (44.2)	24 (48.0)	96 (44.9)
	HWE *P* value & χ^2	>0.25, 0.90	>0.25, 0.52	>0.10, 2.04	>0.25, 0.47	>0.75, 0.04	>0.10, 1.97	>0.50, 0.37	>0.50, 0.32

The genotype and allele carrier data represent the additive, dominant model and recessive models, respectively. * difference in genotype distribution of injured athlete group or injury subgroup versus non-injured ($P \leq 0.05$). # not in HWE.

Genotype and allele frequencies

For *COL5A1* rs12722, the CC genotype, proportion of C-allele carriers and C allele were overrepresented in NIL (34.3%, 85.9% and 58.6%, respectively) compared to LR (19.6%, 66.6% and 43.1%, $P < 0.02$). Furthermore, C-allele carriers were overrepresented, whilst the TT genotype was underrepresented in NIL (86.9% and 17.1%) compared to LS (69.8% and 30.2%, $P < 0.02$). NIL had 3.5 times the odds of carrying the CC genotype compared to LR. There were no differences in genotype or allele frequencies for *COL5A1* rs12722 between any other groups. For allele/genotype frequency data for all SNPs please refer to Table 5.1, and for χ^2 and OR refer to Appendix 3.

For *MMP3* rs679620, the T allele was overrepresented and C-allele carriers underrepresented in NIT (23.5%, 44.7% and 76.5%, respectively) compared to TY (5.7%, 28.6 and 94.3%, $P < 0.03$, Figure 5.1). However, the TT genotype and T allele were underrepresented, whilst the proportion of C-allele carriers were overrepresented in NIL (8.6%, 32.9% and 91.4%) compared to LR (24.5%, 44.3% and 75.5%, $P < 0.03$). Furthermore, the TT genotype was underrepresented, whilst the number of C-allele carriers were overrepresented in NIL (8.6% and 91.4%) compared to LS (19.3% and 80.7%, $P < 0.01$). Additionally, the TT genotype and T allele were underrepresented, whilst the proportion of C-allele carriers were overrepresented in ANI (8.3%, 33.3% and 91.7%) compared to AI (20.5%, 41.1% and 79.5%, $P < 0.03$). The NIL athletes had 3.5 times the odds of carrying the C allele compared to LR, while TY had over 5 times the odds of carrying the C allele

compared to NIT. There were no differences in genotype or allele frequencies for *MMP3* rs679620 between any other groups.

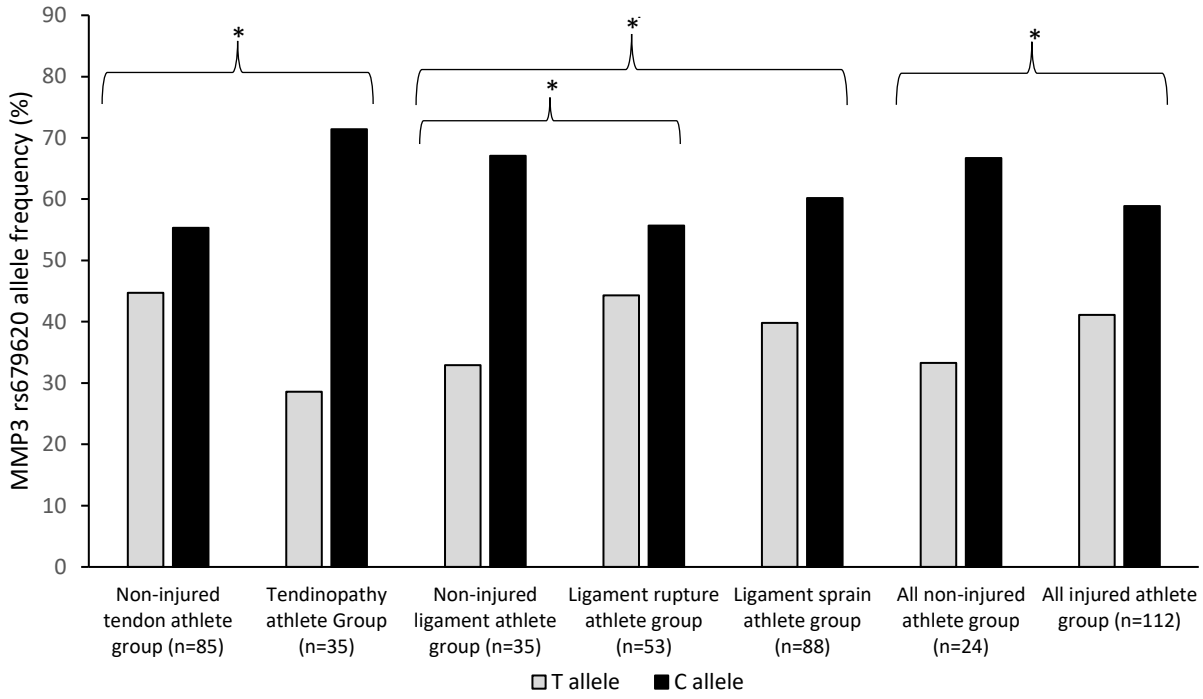

Figure 5.1 Allele frequency of *MMP3* rs679620 for all injured and non-injured athlete groups. Asterisks (*) indicate a difference in allele frequency between the injured athlete group and their respective non-injured athlete group.

For *NID1* rs4660148, the TT genotype and T allele were underrepresented, whilst the number of G-allele carriers were overrepresented in NIT (5.7%, 26.4% and 94.3%, respectively) compared to TR (23.8%, 42.9% and 76.2%, $P < 0.03$). For NIL, the TT genotype was underrepresented, whilst the number of G-allele carriers were overrepresented (2.7% and 97.3%) compared to LR (13.2% and 86.8%, $P < 0.01$) and LS (13.6% and 86.3%, $P < 0.01$). Additionally, the T allele was underrepresented in NIL (29.2%) compared to LS (33.5%, $P < 0.03$). For ANI the

TT genotype was underrepresented, whilst the proportion of G-allele carriers were overrepresented (3.8% and 96.2%) compared to AI (10.7% and 89.3%, $P < 0.01$). The TR, LR and LS athlete groups had over 5 times the odds of carrying the TT genotype compared to their respective non-injured athlete groups, whilst AI had 3 times the odds of carrying the TT genotype compared to ANI. There were no differences in genotype or allele frequencies for *NID1* rs4660148 between any other groups.

For *TIMP2* rs4789932, the GA genotype was underrepresented in the NIL (38.9%) compared to LS (54.6%, $P \leq 0.05$). Furthermore, the GG genotype was overrepresented, whilst the number of A-allele carriers was underrepresented in ANI (40.0% and 60.0%) compared to AI (26.6% and 73.4%, $P \leq 0.05$). There were no differences in genotype or allele frequencies for *TIMP2* rs4789932 between any other groups.

For *COLGALT1* (rs8090), *COL1A1* (rs1800012), *COL3A1* (rs1800255), *COL5A1* (rs3196378) *KDR* (rs1870377), *MIR608* (rs4919510), *MMP3* (rs591058 and rs650108) and *VEGFA* (rs699947) there were no differences in genotype or allele frequencies between any groups (See Table 5.1 for further details).

Haplotype and SNP epistasis analysis

The C-C inferred haplotype frequency of *COL5A1* rs12722 and *COL5A1* rs3196378, respectively, was higher in NIL and ANI compared to LR, LS and AI ($P < 0.01$, Figure

5.2). Additionally, the T-A inferred haplotype frequency of *COL5A1* rs12722 and *COL5A1* rs3196378 was higher in NIT than TR ($P < 0.01$). The T-G-A inferred haplotype frequency of *MMP3* rs591058, rs650108 and rs679620 was lower in NIL than LR and LS ($P < 0.01$). Furthermore, the C-A-G inferred haplotype frequency of *MMP3* rs591058, rs650108 and rs679620 was lower in NIT than TR ($P < 0.01$). Multifactor dimensionality reduction analysis could not identify a model to discriminate between any injury group and the respective non-injured group ($P \geq 0.07$).

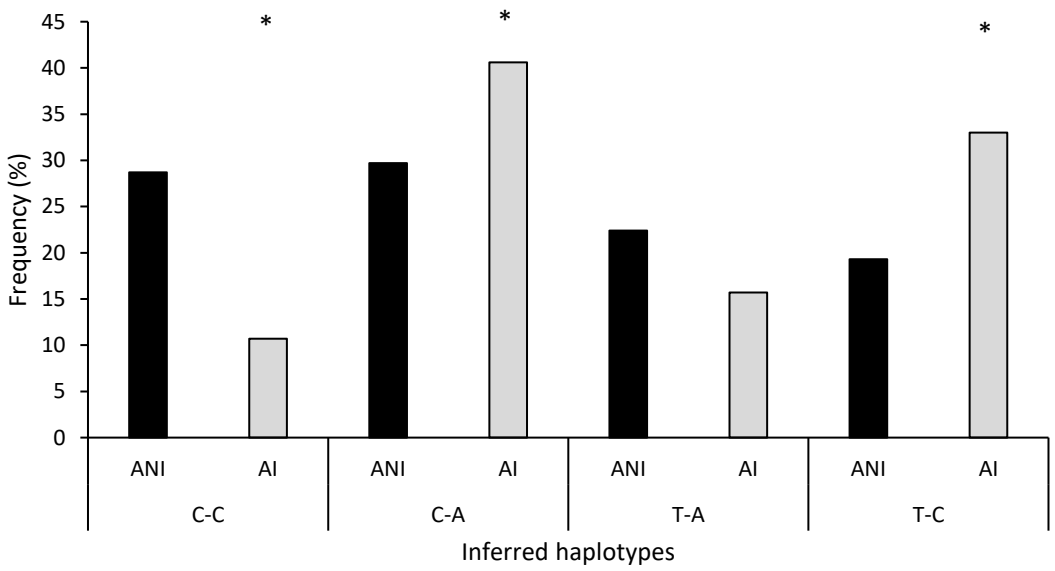

Figure 5.2. Inferred haplotype frequencies derived from *COL5A1* rs12722 and rs3196378. ANI = All non-injured athletes, AI = All injured athletes. *different from ANI ($P < 0.02$).

Total genotype score

There were no differences in TGS between injured and non-injured athlete groups (TR vs NIT, TY vs NIT, LR vs NIL, LS vs NIL, TY, TR and LR vs NIT, TR and LR

and AI vs ANI) (P >0.05). Mean (standard deviation) and kurtosis statistics are reported in Table 5.2, frequency distribution for AI and ANI in Figure 5.3. However, when the numbers of injured and non-injured athletes in the upper and lower third TGSs were compared, ANI had a significantly lower frequency in the lower third of TGS than AI (ANI = 4%, AI = 14%; $\chi^2 = 7.8$, $P = 0.005$). This pattern continued for NIL compared to LR (NIL = 6%, LR = 20%; $\chi^2 = 5.1$, P = 0.024), and also when comparing injured to non-injured when disregarding the relatively common ligament sprains (4% vs.18%; $\chi^2 = 20.4$, $P = 6 \times 10^{-6}$). Conversely, NIT had a higher frequency in the lower third of TGS than TR, which had no athletes present (NIT = 10%, TR = 0%). Finally, the TGS was able to distinguish between TR and NIT (AUC = 0.67; 95% CI = 0.52 – 0.82, $P = 0.045$, Figure 5.4), but could not distinguish between any other groups.

Table 5.2. TGS mean and kurtosis

Group	Mean (SD)	Mean (SE) kurtosis
NIT	45.6 (11.0)	1.18, (0.54)
TR	51.4 (9.6)	1.51 (1.15)
TY	45.7 (7.2)	-0.58 (0.86)
NIL	47.5 (8.0)	-0.30 (0.79)
LR	48.3 (13.2)	0.29 (0.36)
LS	48.9 (10.7)	1.61 (0.55)
ANI	48.3 (8.2)	-0.66 (0.94)
AI	48.9 (11.1)	0.90 (0.49)
Non-injured (excluding ligament sprains)	49.9 (9.1)	1.12 (0.33)
Injured (excluding ligament sprains)	47.8 (11.6)	0.70 (0.59)

NIT = Non-injured tendon athletes, TR = Tendon rupture athletes, TY = Tendinopathy athletes, NIL = Non-injured ligament athletes, LR = Ligament rupture athletes, LS = Ligament sprain athletes, ANI = All non-injured athletes, AI = All injured athletes.

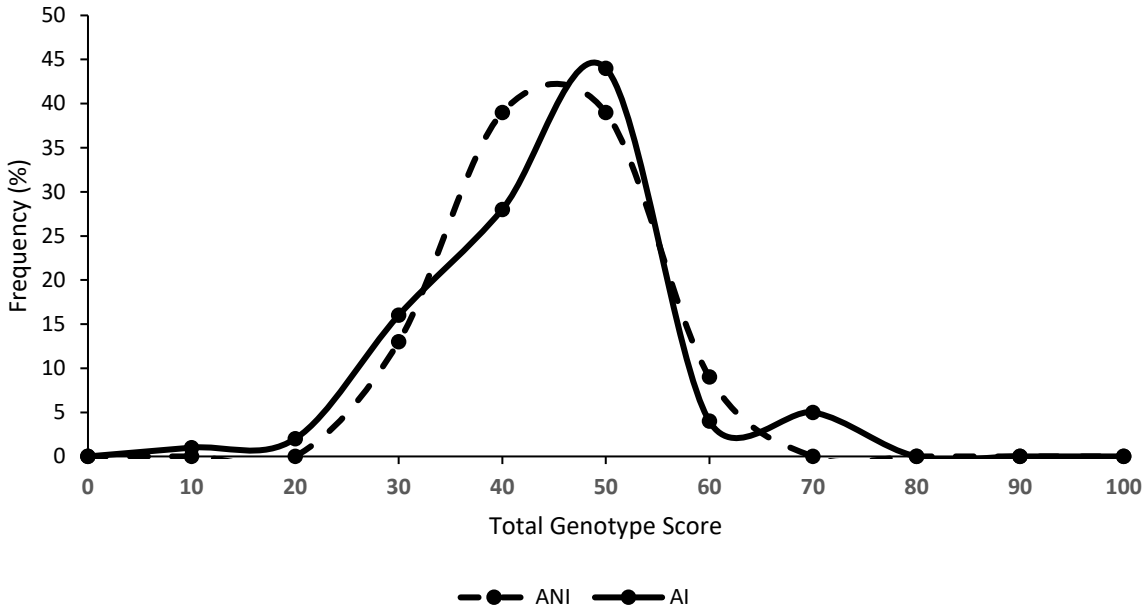

Figure 5.3. Frequency distribution of TGS. There was no difference in mean TGS between ANI and AI. ANI had a lower frequency in the bottom third TGS than AI (ANI = 4%, AI = 14%; $\chi^2 = 7.8$, $P = 0.005$).

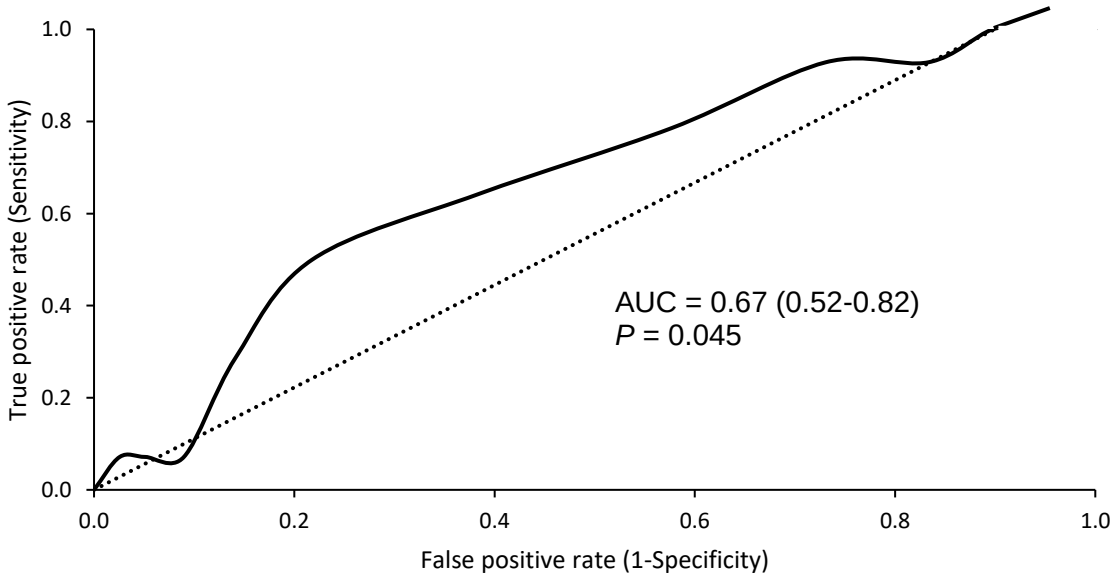

Figure 5.4. Receiver operating characteristic curve (ROC) summarizing the ability of TGS to classify tendon rupture from non-injured tendon rugby athletes. AUC indicates the area under the curve (95% confidence intervals).

5.4 Discussion

The present observations are the first to identify associations between *COL5A1* rs12722, *MMP3* rs679620, *NID1* rs4660148 and *TIMP2* rs4789932 and soft-tissue injury in elite rugby athletes. Thus, indicating a likely inherited advantage from carrying protective genetic variants involved in collagen and ECM structure and function. Furthermore, the injury protective C-C inferred haplotype frequency of *COL5A1* rs12722 and *COL5A1* rs3196378, respectively, were higher in non-injured athlete groups compared to injured athletes. Similarly, the T-G-A inferred haplotype frequency of *MMP3* rs591058, rs650108 and rs679620, respectively, was lower in the non-injured ligament athlete group compared to the ligament rupture and ligament sprain athlete groups. Finally, the TGS was able to distinguish between the non-injured tendon athlete group and the tendon rupture athlete group. These combined findings suggest a likely polygenic influence on soft tissue injury risk in rugby. As hypothesized, elite rugby athletes with a history of soft-tissue injury mostly carried more of the apparent injury-risk genotype/alleles than non-injured athletes, although this was not consistent for all polymorphisms. However, the null hypothesis must be accepted for the secondary aim as no differences were found between injured and non-injured athlete groups for TGS.

The results add further support to the potential genetic influence on soft-tissue injury risk. Previously, the C allele of *COL5A1* rs12722 was identified as protective due to the T allele being associated with ligament injury (Posthumus, 2009; O'Connell et

al., 2015; Bell et al., 2012a) and Achilles tendinopathy (September et al., 2009; Mokone et al., 2006). This aligns with our data with the TT genotype overrepresented in both the ligament rupture and ligament sprain athlete groups and the non-injured ligament athlete group having a higher frequency of C allele carriers. Indeed, the non-injured ligament athletes had 3.5 times the odds of carrying the CC genotype compared to the ligament rupture athlete group. Thus, athletes carrying the CC genotype may have benefited from a reduction in ligament injury across their career maximising their potential training and match time. Furthermore, when *COL5A1* rs12722 and rs3196378 were combined, the C-C inferred haplotype was overrepresented in the non-injured ligament and all non-injured athlete groups compared to the ligament rupture, ligament sprain and all injured athlete groups. However, the T-A inferred haplotype frequency of *COL5A1* rs12722 and *COL5A1* rs3196378, respectively was higher in the non-injured tendon athlete group compared to the tendon rupture group which may suggest a differing underlying molecular mechanism between the injuries. The *COL5A1* 3' untranslated region where rs12722 and rs3196378 are located has been shown to affect mRNA stability (Laguette et al., 2011). Both soft tissue injury 'risk' alleles for rs12722 and rs3196378 (T and A, respectively) were associated with greater HsamiR-608 stability, which may lead to altered COL5A1 protein secondary structure – possibly influencing type V collagen production (Abrahams et al., 2013). Although this proposed mechanism has still to be fully elucidated, our data concur with previous findings that suggest a benefit from carrying the C allele at both rs12722 and rs3196378 has likely played a part in the soft-tissue injury incidence of elite rugby athletes.

The TT genotype and T allele of *MMP3* rs679620 were overrepresented in the non-injured tendon athlete group compared to the tendinopathy athlete group which aligns with previous findings (Raleigh et al., 2009). Conversely, the TT genotype and T allele were underrepresented in the non-injured ligament and all non-injured athlete groups compared to the ligament rupture, ligament sprain and all injured athlete groups, respectively. However, prior literature has not established an association between rs679620 and ligament injury, indeed, the mRNA expression profile of MMP3 appears to contrast between ACL injury and Achilles tendinopathy (Beye, 2008; Jones et al., 2006). Therefore, the opposing findings could be due to underlying differences in the aetiology of these injuries. Furthermore, the T-G-A inferred haplotype of *MMP3* rs591058, rs650108 and rs679620 was underrepresented in the non-injured ligament athlete group compared to the ligament rupture and ligament sprain athlete groups which again contrasts to previous findings on tendinopathy (Raleigh et al., 2009), although not replicated in another (Gibbon et al., 2016). Alterations in extracellular matrix (ECM) homeostasis are thought to play a role in soft tissue injury risk (Riley, 2005a). MMP3 regulates ECM homeostasis via proteolytic activity and is considered an essential regulator of matrix degradation and remodelling (Somerville et al., 2003). It is controlled in part by TIMP which are natural regulators of MMP. However, presently there are no experimental data to demonstrate which alleles/genotypes may or may not be functional. The current study observed a higher frequency of the *TIMP2* rs4789932 GA genotype in the ligament sprain and all injured athlete groups compared to their respective non-injured athlete groups which aligns with prior evidence (El Khoury et al., 2013). Collectively these data suggest that the *MMP3* rs591058, rs650108 and rs679620 and *TIMP2* rs4789932 polymorphisms may along with other variants play

a role in tendon and ligament injury risk, possibly via the regulation of ECM homeostasis.

The *NID1* gene encodes members of the nidogen family of basement membrane glycoproteins. Nidogens are thought to play an essential role in cell interactions with the ECM, especially when tissues are experiencing rapid turnover and growth (Fox et al., 1991). Thus, this may influence the aetiology of soft-tissue injuries via their functions with the ECM. The G allele of *NID1* rs4660148 was previously identified as the polymorphism most strongly associated with ACL rupture ($P < 5 \times 10^{-5}$) as part of a genome wide association screen (Kim et al., 2017) (though it did not reach genome-wide significance). However, there have been no further investigations into this polymorphism and its relationship with soft-tissue injury, so the exact mechanisms are still to be elucidated. In contrast to previous literature, the present study found the TT genotype of *NID1* rs4660148 to be underrepresented and the proportion of G allele carriers to be overrepresented in the non-injured tendon, non-injured ligament and all non-injured athlete groups compared to the tendon rupture, ligament rupture, ligament sprain and all injured athlete groups. Indeed, the tendon rupture, ligament rupture and ligament sprain athlete groups had over 5 times the odds of carrying the TT genotype compared to their respective non-injured athlete groups. Therefore, further investigation is needed to clarify the risk allele/genotype of rs4660148, but the evidence does suggest that it may play a role in influencing soft-tissue injury risk.

When the polygenic profiles of elite rugby athletes were examined, the TGS found no differences between athletes with a history of soft-tissue injury and non-injured athletes for all groups after controlling for multiple comparisons. This suggests that non-injured athletes may not carry 'preferable' soft-tissue injury associated polygenic profiles compared to athletes with a history of injury. As previously discussed in section 4.4, a possible reason for this could be due to the equivocal evidence base of the prior literature regarding the 'risk' allele of each SNP. That being said, when the top and bottom thirds of the TGS were compared, the non-injured ligament, all non-injured and the combined non-injured tendinopathy, tendon rupture and ligament rupture athlete groups had significantly lower frequencies in the bottom third compared to the ligament rupture, all injured and the combined tendinopathy, tendon rupture and ligament rupture groups. Conversely, the tendon rupture group had no athletes at all in the bottom third and a higher frequency in the top third compared to the non-injured tendon athlete group. Additionally, ROC analysis could significantly discriminate the non-injured tendon athlete group from the tendon rupture athlete group. The contrasting results could in part be down to the relatively small sample sizes of the individual TGS groups, reducing their statistical power.

The present study is not without limitations. The retrospective nature of injury data collection is susceptible to recall bias due to reliance on memory (Rothman, 2012). However, retrospective studies are time and resource efficient and promote greater participation from elite sporting populations, thus it was deemed appropriate for this cohort. Furthermore, the sample size is relatively small and only allowed us to identify small-moderate effects.

5.5 Conclusion

This chapter has presented the first associations between *COL5A1* rs12722, *MMP3* rs679620, *NID1* rs4660148 and *TIMP2* rs4789932 and soft-tissue injury in elite rugby athletes. Furthermore, the C-C and T-G-A inferred haplotypes of *COL5A1* rs12722 and *COL5A1* rs3196378 and *MMP3* rs591058, rs650108 and rs679620, respectively, were also associated with soft-tissue injury. The functionality of these genetic variants needs further elucidation to identify how soft-tissue injury risk may be affected. Nevertheless, the current data suggest that elite rugby athletes with no history of soft-tissue injury appear to have inherited more resistance to soft tissue injury than their injured peers. The current study provides further insight into the detailed aetiology of musculoskeletal soft tissue injuries within elite rugby and may, in future, be worthy of consideration for managing the interindividual variability of injury risk in rugby.

Chapter 6

General discussion

6.1 Overview

Elite rugby has one of the highest reported injury incidences of any professional sport (Brooks and Kemp, 2008). Most of these injuries are to the tendon, ligament, and muscle (Williams et al., 2013). Following a review of the existing literature in Chapter 1, it is evident that tendon and ligament injuries are multifactorial and affected by both intrinsic and extrinsic risk factors. Furthermore, previous research has identified a genetic influence on soft-tissue injury risk, with heritability of 40% and 69% for tendon and ligament injury, respectively (Hakim et al., 2004; Magnusson et al., 2020). Indeed, several genetic variants have been previously associated with tendinopathy (Mokone et al., 2005; Mokone et al., 2006; Raleigh et al., 2009; Posthumus et al., 2010b; Nell et al., 2012; Abrahams et al., 2013; Saunders et al., 2013; El Khoury et al., 2013; Khoury El et al., 2014; El Khoury et al., 2015; El Khoury et al., 2016), tendon rupture (El Khoury et al., 2016) and ACL rupture (Posthumus et al., 2009c; Posthumus, 2009; Posthumus et al., 2010a; Malila et al., 2011; Posthumus et al., 2012; Ficek et al., 2013; Stępien-Słodkowska et al., 2013; Rahim et al., 2014; Stępień-Słodkowska et al., 2015b; Stępień-Słodkowska et al., 2015a; O'Connell et al., 2015; El Khoury et al., 2015). However, limited replication has been found. It was postulated in Chapter 1 that to achieve elite athlete status within rugby, athletes may have an inherited resistance to soft-tissue injury which enabled career success due to limited interruptions to their competitive career. Previous evidence in football highlighted how injuries are negatively associated with athlete career progression and thus hinder chances of achieving elite status (Larruskain et al., 2021). Consequently, the current thesis attempted to

investigate the potential genetic influence on elite athlete status and soft-tissue injury risk in rugby. More specifically, the objectives were:

1. To recruit further participants to the biobank of elite rugby union athletes as part of the ongoing RugbyGene project with the aim of evaluating molecular characteristics of elite rugby athletes.
2. To investigate whether genotype and allele frequencies of genetic variants previously associated with reduced tendon and ligament injury risk (*COLGALT1* (rs8090), *COL1A1* (rs1800012), *COL3A1* (rs1800255), *COL5A1* (rs12722), *COL5A1* (rs3196378) *KDR* (rs1870377), *MIR608* (rs4919510), *MMP3* (rs591058, rs650108, and rs679620), *NID1* (rs4660148), *TIMP2 (*rs4789932) and *VEGFA* (rs699947)) differed between elite rugby athletes and non-athlete, and between RU subgroup playing positions.
3. To investigate if tendon and ligament injury-associated polygenic profiles differed between elite rugby athletes and non-athletes, and between RU subgroup playing positions.
4. Finally, to establish whether tendon and ligament injury-associated polymorphisms were associated with a history of previous tendon and ligament injury in elite rugby athletes.

6.2 Main experimental findings

The main findings of the present thesis are discussed in detail in the subsequent sections of this chapter. Briefly, the data presented in Chapter 3 are the first to

identify associations between *COLGALT1* rs8090, *COL3A1* rs1800255, *MIR608* rs4919510, *MMP3* rs591058 and rs679620 and *NID1* rs4660148 polymorphisms and elite athlete status in a large cohort of elite rugby athletes. Additionally, the findings in Chapter 3, with an increased cohort size support the previous work of Heffernan et al. (2017a) which found associations between *COL5A1* rs12722 and rs3196378 polymorphisms and elite status in rugby. 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. This enabled the first two objectives of the thesis to be met. Completion of the third objective revealed a significant polygenic influence on achieving elite athlete status. Indeed, ROC analysis found the TGS models utilising a data led approach (Models 2 and 3) could discriminate between elite rugby athletes and non-athletes. However, this was not the case when modelling TGS solely on prior literature (Model 1) (Chapter 4). Furthermore, Chapter 4 identified 3-SNP epistasis interactions between *COL5A1* rs12722, *COL5A1* rs3196378 and *MIR608* rs4919510, which produced the best model for predicting elite athlete status with a greater frequency of the CC-CC-CC genotype combination in elite rugby athletes which persisted across all subgroups. This was further supplemented by finding the injury risk-associated T-A inferred haplotype frequency of *COL5A1* rs12722 and *COL5A1* rs3196378, respectively, was higher in non-athletes than all athlete groups. Chapter 4's findings, thus suggest a likely polygenic influence on elite athlete status. Finally, objective 4 was achieved in Chapter 5, which identified the first associations between *COL5A1* rs12722, *MMP3* rs679620, *NID1* rs4660148 and *TIMP2* rs4789932 and soft-tissue injury in elite rugby athletes. Additionally, the injury-protective C-C inferred haplotype frequency of *COL5A1* rs12722 and rs3196378, was higher in athletes with no history

of soft-tissue injury than those with such a history. Furthermore, the T-G-A inferred haplotype frequency of *MMP3* rs591058, rs650108 and rs679620, was lower in the non-injured ligament athlete group compared to the ligament rupture and ligament sprain athlete groups. These findings indicate a likely polygenic influence on soft-tissue injury risk.

6.2.1 *COL5A1* rs12722 and rs3196378

In Chapter 3, elite rugby athletes and all RU subgroups, had a higher frequency of *COL5A1* rs12722 CC genotype, C allele and C-allele carriers than non-athletes. Furthermore, for the rs3196378 polymorphism, the CC genotype and C allele were overrepresented while the proportion of A allele carriers was underrepresented in all rugby athletes and some subgroups (RU athletes and RU forwards) compared to non-athletes. This suggests that possessing the C allele of both polymorphisms is beneficial for reaching elite status, possibly via a reduction in soft-tissue injury risk, which has enabled less interruption in training and matches. These findings were supported in part by those in Chapter 5, where rs12722 and rs3196378 were explored in relation to tendon and ligament injury. Rugby athletes with no history of ligament injury possessed higher frequencies of the rs12722 CC genotype, C allele and had a higher proportion of C-allele carriers compared to rugby athletes with a history of ligament rupture. Furthermore, the proportion of rs12722 C-allele carriers was higher, and the frequency of the risk-associated TT genotype was lower in athletes with no history of ligament injury compared to athletes with a past ligament sprain. However, no differences were found between any injured groups for rs3196378. Prior literature has identified the C allele of rs12722 as being protective

due to the T allele being associated with ligament injury (Posthumus, 2009; O'Connell et al., 2015; Bell et al., 2012a) and Achilles tendinopathy (Mokone et al., 2006; September et al., 2009). Additionally, although there have been inconsistent results (September et al., 2009; Brown et al., 2017), it is thought the C allele of rs3196378 would also have a protective role in soft-tissue injury (Laguette et al., 2011; Abrahams et al., 2013).

Col V, the α1 chains of which are encoded by the *COL5A1* gene, is a minor fibrillar collagen in terms of content, but evidence suggests that it functions as a major collagen in developing connective tissues (Roulet et al., 2007). Col V and Col I fibrils co-polymerise to form heterotypic fibres, a process thought to regulate and organise the diameter and structure of these fibres (Birk et al., 1988; Birk et al., 1990). However, this evidence is based on avian cornea, and Col V represents 15-20% of the fibrillary collagen in cornea, while it is only 2-5% in tendon (McLaughlin et al., 1989). Therefore, due to it only accounting for a very small fraction of the total collagen content within tendons, it is unlikely that regulation would be identical to corneal regulation. However, due to the existence of Col V mutations within Ehlers-Danlos syndrome patients, a disease characterised by muscle hypotonia, joint laxity, and joint hypermobility (Beighton et al., 1998), this form of collagen is nevertheless deemed important for tissue structure and function and may influence soft tissue injury risk. Concerning possible molecular mechanisms, the *COL5A1* 3' untranslated region where rs12722 and rs3196378 are located affects mRNA stability (Laguette et al., 2011). Both soft tissue injury 'risk' alleles for rs12722 and rs3196378 (T and A, respectively) were associated with greater Hsa-miR-608 stability, which may lead to altered COL5A1 protein secondary structure – possibly

influencing Col V production (Abrahams et al., 2013). Although this proposed mechanism has still to be fully explained, our data in line with previous findings suggest carrying the C allele of both rs12722 and rs3196378 reduces soft-tissue injury incidence of elite rugby athletes. Indeed, when *COL5A1* rs12722 and rs3196378 were combined in Chapter 4, the injury-risk T-A inferred haplotype was overrepresented in non-athletes compared to all rugby athletes and their subgroups. Similarly, in Chapter 5, the protective C-C inferred haplotype had a higher frequency in the non-injured ligament group and the group of all non-injured athletes compared to the ligament rupture and ligament sprain groups and the group of all injured athletes. Conversely, in Chapter 5, the injury risk T-A inferred haplotype frequency was higher in athletes with no history of tendon injury compared to those with a history of tendon rupture. This may suggest a differing underlying molecular mechanism between ligament and tendon injury, however, there were relatively few participants in the tendon rupture group (n = 19), so these are preliminary. Overall, in line with previous research, it appears the *COL5A1* rs12722 C allele may be beneficial to achieving elite athlete status in rugby and may provide some form of protection against soft-tissue injury risk. Additionally, when in combination with the rs3196378 C allele it may offer additional inherited advantage.

6.2.2 *NID1* rs4660148

The G allele of *NID1* rs4660148 was previously identified as the polymorphism most strongly associated with ACL rupture as part of a GWAS (Kim et al., 2017). Although it did not achieve genome-wide significance ($P > 5 \times 10^{-6}$), the findings suggested it may play a role in the aetiology of ACL rupture. The results in Chapter 3 indicate it

may also play a role in achieving elite status, where rugby athletes and all rugby union subgroups carried a higher frequency of the potentially protective TT genotype than non-athletes. However, in Chapter 5, in contrast to the GWAS, all non-injured athlete groups (non-injured tendon, ligament and all non-injured) had a lower frequency of the TT genotype than the tendon rupture, ligament rupture, ligament sprain and all injured athlete groups. Indeed, the tendon rupture, ligament rupture and ligament sprain athlete groups had over 5 times the odds of carrying the TT genotype compared to their respective non-injured athlete groups. At present the evidence for *NID1* rs4660148 in regard to soft-tissue injury risk is extremely sparse with investigations only carried out by Kim et al. (2017) and the present thesis, which have opposing findings. As discussed in section 1.2.6.9, *NID1* encodes a member of the nidogen family of basement membrane glycoproteins, which are thought to play an essential role in the development of the ECM, particularly when tissues are experiencing rapid turnover and growth (Ho et al., 2008). Thus, their interactions and functions with the ECM may influence the aetiology of soft-tissue injury. Based on the findings within this thesis, carrying the TT genotype appears to be advantageous for elite status, although, it also appears to increase risk of soft-tissue injury. Therefore, it is plausible that rs4660148 may influence several phenotypes, though the mechanisms for this are presently unclear. Future research should seek to determine the functions of rs4660148 to establish its role in injury risk management and elite status.

6.2.3 *MMP3* rs591058, rs650108, and rs679620

MMP3 rs591058, rs650108, and rs679620 were investigated for associations with elite athlete status and soft-tissue injury, having previously been associated with Achilles tendinopathy and ACL rupture (Raleigh et al., 2009; Posthumus et al., 2012; Gibbon et al., 2016). In Chapter 3, elite rugby athletes carried a higher proportion of the previously proposed risk rs591058 C allele (Raleigh et al., 2009) than non-athletes. This was even more apparent in rugby league athletes. Furthermore, rugby league athletes carried a higher proportion of the risk rs679620 C allele. No other differences were found for rs591058, rs650108, and rs679620 between any other groups. The findings from Chapter 3 would suggest that the C alleles of rs591058 and rs679620, perceived as risk alleles (Raleigh et al., 2009), but without replication in a Caucasian cohort, may offer advantages for elite status. MMP3 has a fundamental role in the regular development, repair and remodelling of connective tissues, by regulating ECM homeostasis via proteolytic activity (Foster, 2012). It is plausible that these mechanisms may influence tendon and ligament strength and/or stiffness, potentially influencing performance phenotypes.

In Chapter 5, the frequency of the TT genotype and T allele of rs679620 was higher in athletes with no history of tendon injury compared to athletes with a history of tendinopathy, supporting previous findings from Raleigh et al. (2009). Conversely, the frequency of the TT genotype and T allele was lower in the non-injured ligament group and the group of all non-injured athletes. However, previous literature has not identified an independent association between rs679620 and ligament injury. Furthermore, *MMP3* mRNA expression appears to increase for ACL injury, whilst decreases have been found for Achilles' tendinopathy (Beye, 2008; Jones et al., 2006). Thus, contrasting findings within Chapter 5 could be due to fundamental

biological differences in the aetiology of tendon and ligament injuries. Further evidence of possible differences between tendon and ligament injury were identified in Chapter 5. The frequency of the T-G-A inferred haplotype of *MMP3* rs591058, rs650108 and rs679620, respectively, was lower in athletes with no history of ligament injuries compared to athletes with a history of ligament rupture and ligament sprain, which opposes prior findings on tendinopathy (Raleigh et al., 2009). However, Chapter 5, also identified a higher frequency of the C-G-G inferred haplotype of *MMP3* rs591058, rs650108 and rs679620 in the non-injured ligament athlete group compared to athletes with a history of ligament rupture and/or sprain which aligns with previous tendinopathy research but in a smaller cohort (Gibbon et al., 2016). Precisely how *MMP3* rs591058, rs650108 and rs679620 may influence soft-tissue injury or elite athlete status is still to be fully elucidated. However, it has been proposed that the non-synonymous polymorphism rs679620 could affect the downstream function of the mature MMP3 enzyme and its stimulation (Beyzade et al., 2003), via a change of amino acid and its interaction with additional amino acids. Specifically, the G allele (GAA codon) encodes a glutamate residue, while the A allele (AAA codon) encodes a lysine residue (Riva and Kohane, 2004). The G allele may increase MMP3 activation through adaptation in its function with other amino acids within the propeptide region (Foster, 2012). This may lead to altered ECM structure and/or function, potentially influencing the tendon or ligament. As such, further research is warranted to investigate these potential mechanisms.

6.2.4 *COLGALT1* rs8090, *COL3A1* rs1800255 and *MIR608* rs4919510

COLGALT1 rs8090 was the most strongly associated SNP identified in a fixed-effect meta-analysis for Achilles tendon pathology as part of a GWAS (Kim et al., 2017). Although it did not achieve genome-wide significance ($P > 6 \times 10^{-5}$), it was deemed worthy of investigation in this thesis. Indeed, the AA genotype, A allele and proportion of A allele carriers were higher in elite rugby athletes than non-athletes and this was more apparent in RU athletes, particularly the forwards. A similar pattern was found with *COL3A1* rs1800255 and *MIR608* rs4919510 which were previously associated with ligament and tendon injury, respectively (Stępień-Słodkowska et al., 2015a; O'Connell et al., 2015; Abrahams et al., 2013). The GA and CC genotypes of *COL3A1* rs1800255 and *MIR608* rs4919510, respectively, were more frequent in rugby athletes, particularly the RU forwards. Interestingly the RL subgroup had nearly 3 times the odds of carrying the rs4919510 CC genotype compared to non-athletes. However, no associations were observed between *COLGALT1* rs8090, *COL3A1* rs1800255, *MIR608* rs4919510 and tendon and ligament injury in Chapter 5. These findings suggest that the mechanisms via which these aforementioned SNPs may aid elite status, may have limited influence on protection against soft-tissue injury. Col III is encoded in part by *COL3A1* and thought to play a fundamental role in collagen type I fibrillogenesis, (Banos et al., 2008), which *MIR608* is also thought to influence (Abrahams et al., 2018). Increases in Col III during fibrillogenesis lead to a concomitant decrease in the amount of Col I within the collagen (Riley et al., 1994b) and a larger portion of small diameter and disorganised fibrils (Riley, 2005b). Therefore, it has been postulated that an increased proportion of Col III reduces the tensile strength of tendon/ligament tissue. Additionally, *COLGALT1* has been shown to influence the function of Col I and evidence from mice models found that a mutant phenotype with disorganised

muscle fibres and musculoskeletal defects was associated with *COLGALT1* (Geister et al., 2019). It is therefore plausible to suggest that *COLGALT1* rs8090, *COL3A1* rs1800255, and *MIR608* rs4919510 may be more influential on performance phenotypes such as muscular strength, which is a common attribute of rugby athletes, particularly forwards (Brazier et al., 2020), than soft-tissue injury phenotypes directly.

6.2.5 *COL1A1* rs1800012, *KDR* rs1870377, *TIMP2* rs4789932 and *VEGFA* rs699947

COL1A1 rs1800012 and *KDR* rs1870377 have previously been associated with ligament injury (Khoschnau et al., 2008; Posthumus et al., 2009c; Rahim et al., 2014). *TIMP2* rs4789932 and *VEGFA* rs699947 have been associated with tendinopathy and tendon rupture, and ligament injury and tendinopathy, respectively (El Khoury et al., 2013; El Khoury et al., 2016; Rahim et al., 2014; Rahim et al., 2016). Therefore, all four were investigated for associations with elite athlete status (Chapter 3) and soft-tissue injury (Chapter 5). No associations were found between *COL1A1* rs1800012, *KDR* rs1870377, *TIMP2* rs4789932 and *VEGFA* rs699947 and elite athlete status, nor between *COL1A1* rs1800012, *KDR* rs1870377 and *VEGFA* rs699947 and soft-tissue injury. A possible reason for this could be due to the equivocal prior literature, with only *COL1A1* rs1800012 findings being replicated. However, *TIMP2* rs4789932 was associated with soft-tissue injury. The GA genotype was higher in the ligament sprain group and the group of all injured athletes compared to non-injured athletes which aligns with previous evidence (El Khoury et al., 2013) but opposes others (El Khoury et al., 2016). Currently, there is

no experimental data to demonstrate if rs4789932 is functional or precisely how it may influence soft-tissue injury. Nonetheless, as discussed in section 1.2.6.7, TIMP are natural regulators of MMP, so it may play an indirect role maintaining ECM homeostasis. Any alteration in ECM homeostasis is proposed to influence soft-tissue injury risk (Riley, 2005a). Thus, the GA genotype may affect ECM stability and be precipitous for soft-tissue injury, which would align with the findings of Chapter 5.

6.2.6 Polygenic profiling

Polygenic profiling was performed to determine the combined influence of all 13 of the aforementioned gene polymorphisms on elite athlete status and soft tissue injury. However, when TGS modelling was based on prior literature (Model 1), no differences were found between elite rugby athletes and non-athletes (Chapter 4), which contrasts with previous investigations on elite athlete status using other SNPs (Ruiz et al., 2009; Eynon et al., 2011; Santiago et al., 2010; Ben-Zaken et al., 2015). A likely reason for this is the paucity of evidence in the prior literature supporting the identified 'risk' allele of each polymorphism. Of the 13 polymorphisms investigated in this thesis, only *COL1A1* rs1800012, *COL3A1* rs1800255 and *COL5A1* rs12722 have had their 'risk' alleles identified repeatedly in separate Caucasian cohorts (Brazier et al., 2019). Consequently, two further TGS models were utilised in Chapter 4. Firstly, a data-led TGS model (Model 2) was employed that utilised elite rugby athlete genotype frequency data from Chapter 3. Based on these data, the perceived 'risk' allele was higher in elite rugby athletes for six of the included polymorphisms (not all significant). However, as previously discussed within the

current Chapter, it is plausible that some of the investigated polymorphisms may provide advantage to achieving elite status via other mechanisms than reducing soft tissue injury. The second additional TGS model utilised (Model 3) only included the seven polymorphisms which had been associated with elite status in Chapter 3 (*COLGALT1* rs8090, *COL3A1* rs1800255, *COL5A1* rs12722, *COL5A1* rs3196378, *MIR608* rs4919510, *MMP3* rs591058, *NID1* rs4660148). It has previously been found that including non-associated polymorphisms in TGS can reduce the model's accuracy (Yvert et al., 2016), so it was considered useful to restrict the polymorphisms included in one model. Both additional TGS models found differences between elite rugby athletes and non-athletes which persisted across all rugby union subgroups (Chapter 4). Elite rugby athletes also had a higher frequency in the top quartile and a lower frequency in the bottom quartile TGS for both TGS models (Models 2 and 3) compared to non-athletes. Furthermore, both additional TGS models (2 and 3) were able to discriminate between elite rugby athletes and non-athletes (Chapter 4). These current findings are potentially an important advancement in identifying the genetic characteristics of elite status in rugby and as more relevant polymorphisms are included the sensitivity of the discrimination will improve.

To the authors knowledge, there appears to have been only one study so far on the polygenic analysis of soft-tissue injury risk in athletic populations (Goodlin et al., 2014). As this was a small pilot study, analysis was not performed on genetic risk scores and injury history, meaning it contains no comparable data to the present thesis. In Chapter 5, no differences were seen in TGS (Model 1) between athletes with a history of soft-tissue injury and those with no history of injury. However, when

the top and bottom thirds of the TGS were compared, the non-injured ligament, all non-injured and the combined non-injured tendinopathy, tendon rupture and ligament rupture athlete groups had a significantly lower frequency in the bottom third compared to their respective injured athlete groups. In contrast, the tendon rupture group had no athletes at all in the bottom third and a higher frequency in the top third compared to athletes with no history of tendon injury. Furthermore, the TGS could discriminate athletes with no history of tendon injury from the tendon rupture athlete group. It should be noted that the tendon rupture group only contained 21 athletes which limits its statistical power and increases the chances of type 2 error.

As outlined in Chapter 4, the evidence for most polymorphisms studied within this thesis needs further clarification and therefore, it is likely that the TGS model's utilised in Chapter 4 and 5 included polymorphisms that have no influence on soft-tissue injury risk or elite status. Additionally, as also discussed in Chapter 4, the TGS models employed gave equal weighting to all polymorphisms, furthermore it was assumed allelic effects would be codominant and thus each polymorphism would have an equal additional effect. However, this approach is potentially limited, as it is highly likely that the contribution of each polymorphisms genotype to soft-tissue injury is variable. Indeed, when MDR analysis was performed in Chapter 4, the best model to predict elite athlete status was found to be a 3-SNP model of *COL5A1* rs12722, *COL5A1* rs3196378 and *MIR608* rs4919510. The CC-CC-CC genotype combination was more frequent in all elite rugby groups apart from rugby league athletes compared to non-athletes. These findings align in part, with those in Chapters 3, 4 and 5, with the CC genotype of *COL5A1* rs12722 and *COL5A1* rs3196378, more frequent in elite athletes individually and within an inferred

haplotype compared to non-athletes (Chapter 3 and 4, respectively). As well as the injury risk TT genotype of the *COL5A1* rs12722 being underrepresented within rugby athletes with no history of ligament injury compared to those with a history of ligament sprain and/or rupture (Chapter 5), which was previously discussed in detail in section 6.2.1. The MDR findings of Chapter 4 however, were not replicated for soft-tissue injury in Chapter 5, with MDR analysis unable to identify a model to predict injury with a sufficiently powerful cross-validation statistic. That being said, the combined findings in Chapter 4 and 5, although not consistent, do provide evidence that achieving elite status and having a reduced soft-tissue injury risk are likely polygenic in nature. Future research should seek to determine further polymorphisms associated with the phenotypes investigated in Chapters 4 and 5. Equally importantly, research should also aim to quantify the relative contribution of existing and newly uncovered polymorphisms to enable the polygenic influence on elite status and soft-tissue injury risk to be captured more accurately.

6.3 Study limitations

The present thesis aimed to investigate the genetic influence on elite status and soft-tissue injury within an elite rugby population. A strength of this thesis was the large (n = 663) relatively homogenous cohort of elite level rugby athletes, which to the authors knowledge, is one of the largest samples studied to date for genetic associations within a single sport. Furthermore, it is the first study to investigate genetic association of soft-tissue injury within an elite rugby population. Although these were strengths of the thesis, recruiting and gaining access to elite level rugby athletes was extremely challenging and required international research

collaborations. Genetics research requires very large sample sizes before conclusions can be made to ascertain the individual and collective variant/s effect on a phenotype (Gauderman, 2002). Consequently, although the present thesis sample of ~660 elite rugby athletes is large, increasing the cohort size into the thousands would enhance confidence in the findings. This could also enable larger analyses such as GWAS to be performed with a realistic opportunity of producing results of genome-wide significance that are not excluded following corrections for testing multiple hypotheses. It is also worth noting that the sample size available in this thesis for analyses of soft-tissue injury (n = 138) was smaller than for athlete status. This was due in part to difficulties in gaining sufficient time with the athletes for detailed injury data to be collected, despite the time-efficient retrospective data collection (section 2.8) (Mukherjee, 2015). Nevertheless, the sample size was sufficient to identify small-moderate effects (section 5.2.4).

Although case-control genetic association studies have a variety of strengths as discussed in section 1.2.6, there has been ample criticism of the approach due to limited replication and underpowered studies. A survey of 600 positive associations between disease and gene variants demonstrated this limitation - of 166 associations studied three or more times, only six were replicated (Munafò and Flint, 2004). This is reflected in the equivocal prior literature on genetic associations with soft-tissue injury discussed within this thesis (sections 1.2.6 and 6.2.6). Furthermore, case-control association studies typically only target a very small number of genetic variants that will only explain a limited amount of the variance in heritability of a trait (Wang et al., 2013). Thus, a limitation of this thesis is that only 13 polymorphisms were studied out of a possible 40 million polymorphic DNA sites

in the human genome (Bouchard, 2015). Nonetheless, all polymorphisms chosen within this thesis were based on *a priori* hypotheses, combining prior relevant genetic associations with a detailed review of the role and function of the encoded protein (section 1.2.6). It must also be remembered, of course, that allelic association with a trait or disease does not necessarily infer cause. As such, other than superficially, this thesis does not provide insight into the potential mechanisms via which the studied polymorphisms might contribute to elite status and/or soft tissue injury.

6.4 Conclusion

The present thesis investigated 13 soft-tissue injury associated genetic variants within a unique population of elite rugby athletes, in relation to athlete status and soft-tissue injury. Consequently, this thesis has extended the findings from prior literature that genetic variation exists between elite athletes and non-athletes. Moreover, the present thesis is the first to identify genetic associations with soft-tissue injury within an elite rugby population. It is likely that elite status and soft-tissue injury risk are highly polygenic, and therefore the work from the present thesis should be built upon in future research to advance the field of genomics in sport. Further advances could go some way to providing accurate genetic information to enable the individualisation of training and injury management programmes within elite sport.

6.5 Directions for future research

Several genetic associations have been identified within the present thesis with elite athlete status (*COLGALT1* rs8090, *COL3A1* rs1800255, *COL5A1* rs12722 and rs3196378, *MIR608* rs4919510, *MMP3* rs591058 and rs679620 and *NID1* rs4660148) and soft tissue injury in rugby (*COL5A1* rs12722, *MMP3* rs679620, *NID1* rs4660148 and *TIMP2* rs4789932). However, only *COL5A1* rs12722 consistently demonstrated through Chapters 3, 4 and 5, that the injury protective C allele was potentially beneficial for elite status as well as reducing soft tissue injury in rugby. All other polymorphisms, had inconsistent findings between elite status and soft tissue injury, as discussed in section 6.2. For example, the perceived injury-protective TT genotype of *NID1* rs4660148 was more common in elite rugby athletes than non-athletes but was less common in non-injured rugby athlete groups than their counterparts with a history of soft tissue injury. This could suggest that *NID1* rs4660148 may influence several biological mechanisms that not only affect soft-tissue injury risk but also other phenotypes related to elite status. As such, future research should seek to replicate the genetic associations identified within this thesis, both for elite status and soft tissue injury risk, within other elite sporting populations. Furthermore, the proposed injury-protective/risk alleles for the candidate genes investigated in this thesis are lacking experimental data that would reveal their exact function. Thus, functional studies that can ascertain the potential genetic pathogenesis of soft-tissue injury would be an important focus for future research.

As previously highlighted, a potential limitation of this thesis was the time-limited retrospective nature of the injury data collection. Prospective study designs are generally considered to me more reliable when investigating injury epidemiology

(Mukherjee, 2015). Thus, if access and time to elite athletes are not limiting factors, prospective approaches should be employed, as this will enable more accurate details on exposure time, injury outcomes and a more precise estimation of the risk and incidence of injury. Ideally, future research should seek to combine the genetic analysis of elite rugby athletes, as presented within this thesis, with the meticulously collected rugby injury databases such as PRISP (Kemp et al., 2020). This would allow for a much more detailed investigation of the potential inter-individual variability in injury risk of specific injuries in elite rugby.

Finally, the work presented in Chapters 4 and 5 reveal a likely polygenic influence on elite athlete status and soft tissue injury. It is important to note, that only 13 polymorphisms were included in these polygenic profiles, with some polymorphisms included likely having little to no effect. Therefore, researchers are encouraged to examine further polygenic profiles, using combinations of those reported in this thesis and others reported elsewhere, as well as those yet to be identified. Furthermore, although the candidate gene approach is the most extensively used to quantify the genetic influence on sport-related phenotypes, more complex approaches involving genome-wide technologies will enable further progress to be made. It is generally accepted that elite status and injury risk are highly polygenic in nature, and therefore hypothesis-free genome-wide investigations will enable the whole genome to be searched rather than focussing on individual genes. This could potentially identify many genetic variants which, when combined, could explain substantial proportions of the variance in athlete status and injury risk. However, very large cohorts are required for these complex genomic approaches to become statistically viable, which is challenging due to the limited number of elite athletes

that exist by definition. Therefore, international collaborations such as the Athlome Consortium (Pitsiladis et al., 2016) are required to make this achievable.

Appendices

Appendix 1

Elite athlete questionnaire

The Genetic Profile of Elite Athletes: Questionnaire

Thank you for your interest in our research study. Please answer the following questions about your ethnic origin, athlete status, and your training, diet and injury history.

SECTION A: Questions concerning your ethnic background.

Participant ID code: _____ Date of birth: _____

Gender (please tick): Male ☐ / Female ☐ Height (in metres): _____

Nationality (as on passport, e.g. British): _____ Body weight (in kg): _____

What is your ethnic group? Please tick the appropriate box.

A) White: English ☐ Scottish ☐ Welsh ☐ N. Irish ☐ Irish ☐

French ☐ South African ☐ New Zealander ☐ Australian ☐ Other ☐

If other, please state here: _____

B) Mixed: White & Black British ☐ White & Black Caribbean ☐ White & Black African ☐ White & Asian ☐ White & Latin American ☐ Other ☐

If other, please state here: _____

C) Asian: British ☐ Indian ☐ Pakistani ☐ Chinese ☐ Japanese ☐ Other ☐

If other, please state here: _____

D) Black: British ☐ Caribbean ☐ African ☐ Other ☐

If other, please state here: _____

E) Latin American: Brazilian ☐ Argentinian ☐ Mexican ☐ Colombian ☐ Other ☐

If other, please state here: _____

F) Pacific Islands: Samoa ☐ Fiji ☐ Tonga ☐ PNG ☐ Other ☐

If other, please state here: _____

G) Other ethnic background: ☐ Please state here: _____

I do not wish to state my ethnic origin ☐

Using the ethnic groups above as a guide, please tell us the ethnic origin of your:

Mother: _____ Don't know: ☐

Father: _____ Don't know: ☐

Mother's mother: _____ Don't know: ☐

Mother's father: _____ Don't know: ☐

Father's mother: _____ Don't know: ☐

Father's father: _____ Don't know: ☐

Blood donation

We would like to take a small (10 mL) blood sample from a vein in your arm. Before doing so, please answer the following safety questions.

1. Have you ever been infected with a blood-borne disease? _____ Yes ☐ No ☐
2. Are you anaemic or receiving treatment for anaemia or iron deficiency? __ Yes ☐ No ☐

If you have answered YES to any of these questions and/or you would prefer not to provide a blood sample, a saliva sample may be provided instead.

PLEASE TURN OVER

SECTION B: Questions concerning your athlete status.

1. What is/was your main playing position (if team sport) or your main event (if individual sport). If multiple, please state preferred position/event.

   _____
   _____

2. Please state the number of seasons you have competed as a professional: _____

3. Please state all the professional clubs you have competed for so far in your career:

   _____
   _____

4. Please state the highest level that you have competed, including number of caps earned e.g. England under 16s (4 caps), 18s (10 caps), senior (21 caps): _____
   _____

5. Have you any other athletic achievements? If so please state highest achievements and include relevant details:

   _____
   _____

SECTION C: Questions concerning your training.

1. Typically, how many hours do you train a week?

_____

2. Typically, what is your average running distance per week?

_____

PLEASE TURN OVER

SECTION D: Questions concerning your injury history.

10. Have you ever ruptured your tendon? Yes ☐ If yes, which tendon? No ☐

e.g. Achilles

11. If Yes, please give details of how this occurred and at what age.

Activity	**Age**
e.g. sprinting	*e.g. 20*

12. Have you ever suffered from prolonged tendon pain during exercise that does not go away for weeks? Yes ☐ If yes, which tendon? No ☐

e.g. Achilles

13. Have you ever been told that you have had tendinopathy? Yes ☐ If yes, which tendon? No ☐

e.g. Achilles

14. Yes ☐ No ☐

If Yes, was it confirmed by a scan, e.g. MRI or ultrasound?

15. Does anyone in your close family suffer from tendinopathy? Yes ☐ No ☐ Don't know ☐

If yes, which tendon?

e.g. Achilles

16. Has anyone in your close family ever ruptured a tendon? Yes ☐ No ☐ Don't know ☐

If yes, which tendon?

e.g. Achilles

PLEASE TURN OVER

			If yes, which ligament?	
17.	Have you ever fully ruptured a ligament?	Yes ☐	*e.g. ACL*	No ☐

		Contact	**Non-contact**	**Age**
18.	If Yes, please give details of how this occurred and at what age.	*e.g. tackled from the side*	*e.g. landing from a jump*	*e.g. 20*

19.	Have you ever been told that you have had a ligament sprain/tear?	Yes ☐	If yes, which ligament? *e.g. ACL*	No ☐

20.	Yes ☐	No ☐	

If Yes, was it confirmed by a scan, e.g. MRI or ultrasound?

21. Has anyone in your close family ever ruptured a ligament? Yes ☐ No ☐ Don't know ☐

If yes, which ligament?

e.g. ACL

PLEASE TURN OVER

22. Have you ever been concussed or knocked out? Yes ☐ No ☐

23. If Yes, how many times have you been concussed or knocked out?

_____ times

24. What were you doing at the time of the injury(ies)? E.g. rugby tackle, boxing, road accident.

_____
_____

25. If Yes, how long was your recovery period, until the day when you had no signs and symptoms and were free to train and play fully? (tick, multiple times if necessary, any recovery periods that apply for the different occasions)

<7 days ☐ 7-10 days ☐ 10-20 days ☐

20-40 days ☐ 40-60 days ☐ >60 days ☐

26. If Yes, was/were your concussion(s) or knock-out(s) diagnosed by a medical professional? (tick, multiple times if necessary, any that apply) Yes ☐ No ☐

27. Does anyone in your close family (parents, siblings or grandparents) suffer from a neurological condition, such as:

Dementia, Alzheimer's disease, chronic traumatic encephalopathy (CTE), cognitive impairment, movement disorders, psychiatric disorders, motor neuron disease

Yes ☐ No ☐ Don't know ☐

Who and which condition(s)?

e.g. grandfather, dementia

_____

_____

PLEASE TURN OVER

THIS PAGE HAS DELIBERATELY BEEN LEFT BLANK

Appendix 2

Supplementary results for Chapter 3: Electronic folder link

Please find Chapter 3's supplementary results - χ^2, Benjamini-Hochberg and odds ratio within the following electronic folder (This is open access to anyone with the link below and a Microsoft account – it may ask for a Microsoft email to access):

https://herts365-my.sharepoint.com/:f:/g/personal/jb17afc_herts_ac_uk/Eg7p_gHi7bhGqCM3klGhlh8B5FCuoImSMBDF3gzE1VQImw?e=JxspYI

Password: Tendonligament21

Appendix 3

Supplementary results for Chapter 5: Electronic folder link

Please find Chapter 5's supplementary results - χ^2, Benjamini-Hochberg and odds ratio within the following electronic folder (This is open access to anyone with the link below and a Microsoft account – it may ask for a Microsoft email to access):

https://herts365-my.sharepoint.com/:f:/g/personal/jb17afc_herts_ac_uk/EtwbzCJ8zV9LvTXIWT-MQbcBP-cWbJHso47YfY_IO10Dog?e=cLVDgu

Password: Tendonligament21

References

Abate, M., Silbernagel, K. G., Siljeholm, C., Di Iorio, A., De Amicis, D., Salini, V., Werner, S. and Paganelli, R. (2009) 'Pathogenesis of tendinopathies: inflammation or degeneration?' *Arthritis Research and Therapy*, 11(3) pp. 235-235.

Abrahams, S., Mc Fie, S., Patricios, J., Suter, J., Posthumus, M. and September, A. V. (2018) 'An association between polymorphisms within the APOE gene and concussion aetiology in rugby union players.' *Journal of Science and Medicine in Sport*, 21(2), Feb, 2017/06/25, pp. 117-122.

Abrahams, S., Mc Fie, S., Patricios, J., Suter, J., September, A. V. and Posthumus, M. (2019a) 'Toxic tau: The TAU gene polymorphisms associate with concussion history in rugby union players.' *Journal of Science and Medicine in Sport*, 22(1), Jan, 2018/12/18, pp. 22-28.

Abrahams, S., McFie, S., Lacerda, M., Patricios, J., Suter, J., September, A. V. and Posthumus, M. (2019b) 'Unravelling the interaction between the DRD2 and DRD4 genes, personality traits and concussion risk.' *BMJ Open Sport and Exercise Medicine*, 5(1) 2019/03/01, p. e000465.

Abrahams, Y., Laguette, M. J., Prince, S. and 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(3) pp. 204-214.

Alentorn-Geli, E., Myer, G. D., Silvers, H. J., Samitier, G., Romero, D., Lázaro-Haro, C. and Cugat, R. (2009) 'Prevention of non-contact anterior cruciate ligament injuries in soccer players. Part 1: Mechanisms of injury and underlying risk factors.' *Knee Surgery, Sports Traumatology, Arthroscopy : official journal of the ESSKA*, 17(7) pp. 705-729.

Almekinders, L. C. and Temple, J. D. (1998) 'Etiology, diagnosis, and treatment of tendonitis: an analysis of the literature.' *Medicine and Science in Sports and Exercise*, 30(8) pp. 1183-1190.

Altinisik, J., Meric, G., Erduran, M., Ates, O., Ulusal, A. E. and Akseki, D. (2015) 'The BstUI and DpnII variants of the COL5A1 gene are associated with tennis elbow.' *American Journal of Sports Medicine*, 43(7) pp. 1784-1789.

Axibal, D. P. and Anderson, J. G. (2013) 'Multiple tendon ruptures of unknown etiology.' *Foot and Ankle Secialist*, 6(5) p. 380.

Banos, C. C., Thomas, A. H. and Kuo, C. K. (2008) 'Collagen fibrillogenesis in tendon development: current models and regulation of fibril assembly.' *Birth Defects Research. Part C, Embryo today : reviews*, 84(3) pp. 228-244.

Beighton, P., Paepe, A. D., Steinmann, B., Tsipouras, P. and Wenstrup, R. J. (1998) 'Ehlers-Danlos syndromes: Revised nosology, Villefranche, 1997.' *American Journal of Medical Genetics*, 77(1) pp. 31-37.

Bell, R. D., Shultz, S. J., Wideman, L. and Henrich, V. C. (2012a) 'Collagen gene variants previously associated with anterior cruciate ligament injury risk are also associated with joint laxity.' *Sports Health: A Multidisciplinary Approach*, 4(4) pp. 312-318.

Bell, W., Colley, J. P., Evans, W. D., Darlington, S. E. and Cooper, S. M. (2012b) 'ACTN3 genotypes of Rugby Union players: distribution, power output and body composition.' *Annals of Human Biology*, 39(1) p. 19.

Bell, W., Colley, J. P., Gwynne, J. R., Glazier, P., Evans, W. D. and Darlington, S. E. (2010) 'ACE ID genotype and leg power in rugby union players.' *The Journal of Sports Medicine and Physical Fitness*, 50(3) p. 350.

Ben-Zaken, S., Meckel, Y., Nemet, D. and Eliakim, A. (2015) 'Genetic score of power-speed and endurance track and field athletes.' *Scand J Med Sci Sports*, 25(2), Apr, 2013/11/29, pp. 166-174.

Benjamini, Y. and Hochberg, Y. (1995) 'Controlling the false discovery rate: A practical and powerful approach to multiple testing.' *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1) pp. 289-300.

Best, J. P., McIntosh, A. S. and Savage, T. N. (2005) 'Rugby world cup 2003 injury surveillance project.' *British Journal of Sports Medicine*, 39(11) pp. 812-817.

Beye, J. A. (2008) 'Injury-induced changes in mRNA levels differ widely between anterior cruciate ligament and medial collateral ligament.' *The American Journal of Sports Medicine*, 36(7) pp. 1337-1346.

Beyzade, S., Zhang, S., Wong, Y.-k., Day, I. N. M., Eriksson, P. and Ye, S. (2003) 'Influences of matrix metalloproteinase-3 gene variation on extent of coronary atherosclerosis and risk of myocardial infarction.' *Journal of the American College of Cardiology*, 41(12) pp. 2130-2137.

Birk, D. E., Fitch, J. M., Babiarz, J. P. and Linsenmayer, T. F. (1988) 'Collagen type I and type V are present in the same fibril in the avian corneal stroma.' *The Journal of Cell Biology*, 106(3) p. 999.

Birk, D. E., Fitch, J. M., Babiarz, J. P., Doane, K. J. and Linsenmayer, T. F. (1990) 'Collagen fibrillogenesis in vitro: interaction of types I and V collagen regulates fibril diameter.' *Journal of Cell Science*, 95 (Pt 4) p. 649.

Birney, E., Andrews, T. D., Bevan, P., Caccamo, M., Chen, Y., Clarke, L., Coates, G., Cuff, J., Curwen, V., Cutts, T., Down, T., Eyras, E., Fernandez-Suarez, X. M., Gane, P., Gibbins, B., Gilbert, J., Hammond, M., Hotz, H.-R., Iyer, V., Jekosch, K., Kahari, A., Kasprzyk, A., Keefe, D., Keenan, S., Lehvaslaiho, H., McVicker, G., Melsopp, C., Meidl, P., Mongin, E., Pettett, R., Potter, S., Proctor, G., Rae, M., Searle, S., Slater, G., Smedley, D., Smith, J., Spooner, W., Stabenau, A., Stalker, J., Storey, R., Ureta-Vidal, A., Woodwark, K. C., Cameron, G., Durbin, R., Cox, A., Hubbard, T. and Clamp, M. (2004) 'An overview of Ensembl.' *Genome Research*, 14(5) pp. 925-928.

Bisset, L., Beller, E., Jull, G., Brooks, P., Darnell, R. and Vicenzino, B. (2006) 'Mobilisation with movement and exercise, corticosteroid injection, or wait and see for tennis elbow: randomised trial.' *British Medical Journal*, 333(7575) pp. 939-941.

Bouchard, C. (2015) 'Exercise genomics-a paradigm shift is needed: a commentary.' *British Journal Of Sports Medicine*, 49(23) pp. 1492-1496.

Brazier, J., Antrobus, M., Stebbings, G. K., Day, S. H., Heffernan, S. M., Cross, M. J. and Williams, A. G. (2019) 'Tendon and ligament injuries in elite rugby: the potential genetic influence.' *Sports (Basel, Switzerland)*, 7(6) p. 138.

Brazier, J., Antrobus, M., Stebbings, G. K., Day, S. H., Callus, P., Erskine, R. M., Bennett, M. A., Kilduff, L. P. and Williams, A. G. (2020) 'Anthropometric and physiological characteristics of elite male rugby athletes.' *Journal of Strength and Conditioning Research*, 34(6) pp. 1790-1801.

Brooks, J. H. (2004) *The epidemiology of injuries in professional rugby.* Ph.D. University of Leicester.

Brooks, J. H., Fuller, C. W., Kemp, S. P. and Reddin, D. B. (2005a) 'A prospective study of injuries and training amongst the England 2003 rugby world cup squad.' *British Journal of Sports Medicine*, 39(5) pp. 288-293.

Brooks, J. H. M. and Kemp, S. P. (2008) 'Recent trends in rugby union.' *Clinical Sports Medicine.*, 27(1) pp. 51-73.

Brooks, J. H. M., Fuller, C. W., Kemp, S. P. T. and Reddin, D. B. (2005b) 'Epidemiology of injuries in English professional rugby union: Part 1 match injuries.' *British Journal of Sports Medicine*, 39(10) pp. 757-766.

Brooks, J. H. M., Fuller, C. W., Kemp, S. P. T. and Reddin, D. B. (2005c) 'Epidemiology of injuries in English professional rugby union: Part 2 training injuries.' *British Journal of Sports Medicine*, 39(10) pp. 767-775.

Brown, K. L., Seale, K. B., El Khoury, L. Y., Posthumus, M., Ribbans, W. J., Raleigh, S. M., Collins, M. and 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*, 35(15), pp. 1475-1483.

Butler, D. L., Grood, E. S., Noyes, F. R. and Zernicke, R. F. (1978) 'Biomechanics of ligaments and tendons.' *Exercise and Sport Sciences Reviews*, 6 p. 125.

Cahill, N., Lamb, K., Worsfold, P., Headey, R. and Murray, S. (2013) 'The movement characteristics of English Premiership rugby union players.' *Journal of Sports Sciences*, 31(3) pp. 229-237.

Chu, C. R., Beynnon, B. D., Buckwalter, J. A., Garrett, J. W. E., Katz, J. N., Rodeo, S. A., Spindler, K. P. and Stanton, R. A. (2011) 'Closing the gap between bench and bedside research for early arthritis therapies (EARTH): report from the AOSSM/NIH U-13 Post-Joint Injury Osteoarthritis Conference II.' *The American Journal of Sports Medicine*, 39(7) pp. 1569-1578.

Collins, M. and Raleigh, S. M. (2009) 'Genetic risk factors for musculoskeletal soft tissue injuries.' *Medicine and Sport Science*, 54 p. 136.

Corps, A. N., Robinson, A. H. N., Movin, T., Costa, M. L., Hazleman, B. L. and Riley, G. P. (2006) 'Increased expression of aggrecan and biglycan mRNA in Achilles tendinopathy.' *Rheumatology*, 45(3) pp. 291-294.

Corry, I. (2000) 'Injuries of the sporting knee.' *British Journal of Sports Medicine*, 34(5) pp. 395-395.

Cross, M., Kemp, S., Smith, A., Trewartha, G. and Stokes, K. (2016a) 'Professional rugby union players have a 60% greater risk of time loss injury after concussion: a 2-season prospective study of clinical outcomes.' *British Journal of Sports Medicine*, 50(15) pp. 926-931.

Cross, M., Williams, S., Kemp, S. P. T., Fuller, C., Taylor, A., Brooks, J., Trewartha, G. and Stokes, K. (2018) 'Does the reliability of reporting in injury surveillance studies depend on injury definition?' *Orthopaedic Journal of Sports Medicine*, 6(3) doi: 10.1177/2325967118760536

Cross, M. J., Williams, S., Trewartha, G., Kemp, S. P. T. and Stokes, K. A. (2016b) 'The influence of in-season training loads on injury risk in professional rugby union.' *International Journal of Sports Physiology and Performance*, 11(3) pp. 350-355.

Cross, M. J., Tucker, R., Raftery, M., Hester, B., Williams, S., Stokes, K. A., Ranson, C., Mathema, P. and Kemp, S. (2017) 'Tackling concussion in professional rugby union: a case-control study of tackle-based risk factors and recommendations for primary prevention.' *British Journal of Sports Medicine*, doi:10.1136/bjsports-2017-097912.

Dallalana, R. J. (2007) 'The epidemiology of knee injuries in English professional rugby union.' *The American Journal of Sports Medicine*, 35(5) pp. 818-830.

Dalton, S., Cawston, T. E., Riley, G. P., Bayley, I. J. and 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(7) pp. 571-577.

De Moor, M. H., Spector, T. D., Cherkas, L. F., Falchi, M., Hottenga, J. J., Boomsma, D. I. and De Geus, E. J. (2007) 'Genome-wide linkage scan for athlete status in 700 British female DZ twin pairs.' *Twin Research and Human Genetics*, 10(6), Dec, 2008/01/09, pp. 812-820.

DNA Genotek Collection Instructions - DNA . [Online] [Accessed April 2nd, 2021].https://www.dnagenotek.com/ROW/support/collection-instructions/oragene-dna/OG-500andOG-600.html.

Docheva, D., Müller, S. A., Majewski, M. and Evans, C. H. (2015) 'Biologics for tendon repair.' *Advanced Drug Delivery Reviews*, 84 pp. 222-239.

Drysdale, C. M., McGraw, D. W., Stack, C. B., Stephens, J. C., Judson, R. S., Nandabalan, K., Arnold, K., Ruano, G. and Liggett, S. B. (2000) 'Complex promoter and coding region β2-adrenergic receptor haplotypes alter receptor expression and predict in vivo responsiveness.' *Proceedings of the National Academy of Sciences of the United States of America*, 97(19) pp. 10483-10488.

Egginton, S. (2009) 'Invited review: activity-induced angiogenesis.' *European Journal of Physiology*, 457(5) pp. 963-977.

El Khoury, L., Ribbans, W. J. and 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 pp. 52-55.

El Khoury, L., Posthumus, M., Collins, M., Handley, C. J., Cook, J. and 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(6) pp. 493-498.

El Khoury, L., Posthumus, M., Collins, M., van der Merwe, W., Handley, C., Cook, J. and Raleigh, S. (2015) 'ELN and FBN2 gene variants as risk factors for two sports-related musculoskeletal injuries.' *International Journal of Sports Medicine*, 36(4) pp. 333-337.

Erduran, M., Altinisik, J., Meric, G., Ates, O., Ulusal, A. E. and Akseki, D. (2014) 'Is Sp1 binding site polymorphism within COL1A1 gene associated with tennis elbow?' *Gene*, 537(2) pp. 308-311.

Eriksen, H. A., Pajala, A., Leppilahti, J. and Risteli, J. (2002) 'Increased content of type III collagen at the rupture site of human Achilles tendon.' *Journal of Orthopaedic Research*, 20(6) pp. 1352-1357.

Eynon, N., Ruiz, J. R., Meckel, Y., Morán, M. and Lucia, A. (2011) 'Mitochondrial biogenesis related endurance genotype score and sports performance in athletes.' *Mitochondrion*, 11(1) pp. 64-69.

Eynon, N., Alves, A. J., Meckel, Y., Yamin, C., Ayalon, M., Sagiv, M. and Sagiv, M. (2010) 'Is the interaction between HIF1A P582S and ACTN3 R577X determinant for power/sprint performance?' *Metabolism*, 59(6), pp. 861-865.

Eynon, N., Oliveira, J., Meckel, Y., Sagiv, M., Yamin, C., Sagiv, M., Amir, R. and Duarte, J. A. (2009) 'The guanine nucleotide binding protein beta polypeptide 3 gene C825T polymorphism is associated with elite endurance athletes.' *Experimental Physiology*, 94(3), pp. 344-349.

Feigelson, H. S., Rodriguez, C., Robertson, A. S., Jacobs, E. J., Calle, E. E., Reid, Y. A. and Thun, M. J. (2001) 'Determinants of DNA yield and quality from buccal cell samples collected with mouthwash.' *Cancer Epidemiology Biomarkers, and Prevention*, 10(9), pp. 1005-1008.

Ferrara, N. (1999) 'Molecular and biological properties of vascular endothelial growth factor.' *Journal of Molecular Medicine*, 77(7) pp. 527-543.

Ficek, K., Stepien-Slodkowska, M., Kaczmarczyk, M., Maciejewska-Karlowska, A., Sawczuk, M., Cholewinski, J., Leonska-Duniec, J., Zarebska, A., Cieszczyk, P. and Zmijewski, P. (2014) 'Does the A9285g polymorphism in collagen type XII α1 gene associate with the risk of anterior cruciate ligament ruptures?' *Balkan Journal of Medical Genetics*, 17(1) pp. 41-46.

Ficek, K., Cieszczyk, P., Kaczmarczyk, M., Maciejewska-Karłowska, A., Sawczuk, M., Cholewinski, J., Leonska-Duniec, A., Stepien-Slodkowska, M., Zarebska, A., Stepto, N. K., Bishop, D. J. and Eynon, N. (2013) 'Gene variants within the COL1A1 gene are associated with reduced anterior cruciate ligament injury in professional soccer players.' *Journal of Science and Medicine in Sport*, 16(5) pp. 396-400.

Fitzpatrick, A. C., Naylor, A. S., Myler, P. and Robertson, C. (2018) 'A three-year epidemiological prospective cohort study of rugby league match injuries from the European Super League.' *Journal of Science and Medicine in Sport*, 21(2) pp. 160-165.

FlexiGene DNA Kit. [Online] [Accessed April 2nd, 2021]. https://www.qiagen.com/gb/products/discovery-and-translational-research/dna-rna-purification/dna-purification/genomic-dna/flexigene-dna-kit/#productdetails/.

Flynn, R. K., Pedersen, C. L., Birmingham, T. B., Kirkley, A., Jackowski, D. and Fowler, P. J. (2005) 'The familial predisposition toward tearing the anterior cruciate ligament: a case control study.' *The American Journal of Sports Medicine*, 33(1) p. 23.

Foster, B., Morse, C.I., Onambele, G.L., Ahmetov, I.I and Williams, A.G. (2012) 'Genetic variation, protein composition and potential influences on tendon properties in humans.' *The Open Sports Medicine Journal*, 6(1) pp. 8-21.

Fox, J. W., Mayer, U., Nischt, R., Aumailley, M., Reinhardt, D., Wiedemann, H., Mann, K., Timpl, R., Krieg, T. and Engel, J. (1991) 'Recombinant nidogen consists of three globular domains and mediates binding of laminin to collagen type IV.' *The EMBO Journal*, 10(11) p. 3137.

Frank, C. B. (2004) 'Ligament structure, physiology and function.' *Journal of Musculoskeletal and Neuronal Interactions*, 4(2) p. 199.

Fuller, C., Taylor, A., Douglas, M. and Raftery, M. (2020) 'Rugby world cup 2019 injury surveillance study.' *South African Journal of Sports Medicine*, 32(1), 05/04,

Fuller, C. W. and Drawer, D. (2004) 'The application of risk management in sport.' *Sports Medicine*, 34(6) pp. 349-356.

Fuller, C. W., Clarke, L. and Molloy, M. G. (2010a) 'Risk of injury associated with rugby union played on artificial turf.' *Journal of Sports Sciences*, 28(5) pp. 563-570.

Fuller, C. W., Sheerin, K. and Targett, S. (2013) 'Rugby world cup 2011: International Rugby Board injury surveillance study.' *British Journal of Sports Medicine*, 47(18) pp. 1184-1191.

Fuller, C. W., Laborde, F., Leather, R. J. and Molloy, M. G. (2008) 'International Rugby Board rugby world cup 2007 injury surveillance study.' *British Journal of Sports Medicine*, 42(6) pp. 452-459.

Fuller, C. W., Taylor, A., Kemp, S. P. T. and Raftery, M. (2017) 'Rugby world cup 2015: World rugby injury surveillance study.' *British Journal of Sports Medicine*, 51(1) pp. 51-57.

Fuller, C. W., Brooks, J. H. M., Cancea, R. J., Hall, J. and Kemp, S. P. T. (2007) 'Contact events in rugby union and their propensity to cause injury.' *British Journal of Sports Medicine*, 41(12) pp. 862-867.

Fuller, C. W., Raftery, M., Readhead, C., Targett, S. G. R. and Molloy, M. G. (2009) 'Impact of the International Rugby Board's experimental law variations on the incidence and nature of match injuries in southern hemisphere professional rugby union.' *South African Medical Journal*, 99(4) p. 232.

Fuller, C. W., Ashton, T., Brooks, J. H. M., Cancea, R. J., Hall, J. and Kemp, S. P. T. (2010b) 'Injury risks associated with tackling in rugby union.' *British Journal of Sports Medicine*, 44(3) pp. 159-167.

Fuller, C. W., Molloy, M. G., Bagate, C., Bahr, R., Brooks, J. H., Donson, H., Kemp, S., McCrory, P., McIntosh, A., Meeuwisse, W. H., Quarrie, K. L., Raftery, M. and Wiley, P. (2007c) 'Consensus statement on injury definitions and data collection procedures for studies of injuries in rugby union.' *British Journal of Sports Medicine*, 41(5) pp. 328-331.

Gabbe, B. J., Finch, C. F., Bennell, K. L. and Wajswelner, H. (2003) 'How valid is a self reported 12 month sports injury history?' *British Journal of Sports Medicine*, 37(6) pp. 545-547.

Gabbett, T. J. (2004) 'Incidence of injury in junior and senior rugby league players.' *Sports Medicine*, 34(12) pp. 849-859.

Gabbett, T. J. and Hodgson, P. (2003) 'Incidence of injury in semi-professional rugby league players.' *British Journal of Sports Medicine*, 37(1) pp. 36-43.

Gabbett, T. J. and Jenkins, D. G. (2011) 'Relationship between training load and injury in professional rugby league players.' *Journal of Science and Medicine in Sport*, 14(3) pp. 204-209.

Gabbett, T. J., Jenkins, D. G. and Abernethy, B. (2011) 'Physical collisions and injury in professional rugby league match-play.' *Journal of Science and Medicine in Sport*, 14(3) pp. 210-215.

Gabbett, T. J., Ullah, S. and Finch, C. F. (2012) 'Identifying risk factors for contact injury in professional rugby league players-application of a frailty model for recurrent injury.' *Journal of Science and Medicine in Sport*, 15(6) p. 496.

Gauderman, W. J. (2002) 'Sample size requirements for association studies of gene-gene interaction.' *American Journal of Epidemiology*, 155(5), pp. 478-484.

Geister, K. A., Lopez-Jimenez, A. J., Houghtaling, S., Ho, T. H., Vanacore, R. and Beier, D. R. (2019) 'Loss of function of Colgalt1 disrupts collagen post-translational modification and causes musculoskeletal defects.' *Disease Models and Mechanisms*, 12(6), doi: 10.1242/dmm.037176

Gelse, K., Pöschl, E. and Aigner, T. (2003) 'Collagens—structure, function, and biosynthesis.' *Advanced Drug Delivery Reviews*, 55(12) pp. 1531-1546.

Gibbon, A., Hobbs, H., van der Merwe, W., Raleigh, S. M., Cook, J., Handley, C. J., Posthumus, M., Collins, M. and September, A. V. (2016) 'The MMP3 gene in musculoskeletal soft tissue injury risk profiling: A study in two independent sample groups.' *Journal of Sports Sciences*, pp. 1-8.

Gibbs, N. (1993) 'Injuries in professional rugby league: a three year prospective study of the South Sydney Professional Rugby League Football Club.' *American Journal Of Sports Medicine.*, 21(5) pp. 696-700.

Gibbs, N. (1994) 'Common rugby league injuries. Recommendations for treatment and preventative measures.' *Sports Medicine*, 18(6) pp. 438-450.

Gillies, A. R. and Lieber, R. L. (2011) 'Structure and function of the skeletal muscle extracellular matrix.' *Muscle and Nerve*, 44(3) pp. 318-331.

Gissane, C., Jennings, D. C. and Standing, P. (1993) 'Incidence of injury in rugby league football.' *Physiotherapy Journal*, 79(5) pp. 305-310.

Gissane, C., Jennings, D., Kerr, K. and White, J. (2003) 'Injury rates in rugby league football: impact of change in playing season.' *The American Journal of Sports Medicine*, 31(6) pp. 954-958.

Glasel, J. A. (1995) 'Validity of nucleic acid purities monitored by 260nm/280nm absorbance ratios.' *Biotechniques*, 18(1), pp. 62-63.

Goh, K. P., Chew, K., Koh, A., Guan, M., Wong, Y. S. and Sum, C. F. (2009) 'The relationship between ACE gene ID polymorphism and aerobic capacity in Asian rugby players.' *Singapore Medical Journal*, 50(10) p. 997.

Gómez-Gallego, F., Ruiz, J. R., Buxens, A., Altmäe, S., Artieda, M., Santiago, C., González-Freire, M., Verde, Z., Arteta, D., Martínez, A., Tejedor, D., Lao, J. I., Arenas, J. and Lucia, A. (2010) 'Are elite endurance athletes genetically predisposed to lower disease risk?' *Physiological Genomics*, 41(1) pp. 82-90.

Gonçalves-Neto, J., Witzel, S. S., Teodoro, W. R., Carvalho-Junior, A. E., Fernandes, T. D. and Yoshinari, H. H. (2002) 'Changes in collagen matrix composition in human posterior tibial tendon dysfunction.' *Joint Bone Spine*, 69(2) pp. 189-194.

Goodlin, G. T., Roos, A. K., Roos, T. R., Hawkins, C., Beache, S., Baur, S. and Kim, S. K. (2014) 'Applying personal genetic data to injury risk assessment in athletes.' *PLoS One*, 10(4), doi: 10.1371/journal.pone.0122676.

Griffin, L. Y., Albohm, M. J., Arendt, E. A., Bahr, R., Beynnon, B. D., DeMaio, M., Dick, R. W., Engebretsen, L., Garrett, W. E., Hannafin, J. A., Hewett, T. E., Huston, L. J., Ireland, M. L., Johnson, R. J., Lephart, S., Mandelbaum, B. R., Mann, B. J., Marks, P. H., Marshall, S. W., Myklebust, G., Noyes, F. R., Powers, C., Shields, C., Shultz, S. J., Silvers, H., Slauterbeck, J., Taylor, D. C., Teitz, C. C., Wojtys, E. M. and Yu, B. (2006) 'Understanding and preventing noncontact anterior cruciate ligament injuries: A review of the Hunt Valley II meeting, January 2005.' *The American Journal of Sports Medicine*, 34(9) pp. 1512-1532.

Guskiewicz, K. M., Marshall, S. W., Bailes, J., McCrea, M., Cantu, R. C., Randolph, C. and Jordan, B. D. (2005) 'Association between recurrent concussion and late-life cognitive impairment in retired professional football players.' *Neurosurgery*, 57(4), pp. 719-726; discussion 719-726.

Hakim, A. J., Cherkas, L. F., Spector, T. D. and MacGregor, A. J. (2003) 'Genetic associations between frozen shoulder and tennis elbow: a female twin study.' *Rheumatology*, 42(6) pp. 739.

Hakim, A. J., Cherkas, L. F., Grahame, R., Spector, T. D. and MacGregor, A. J. (2004) 'The genetic epidemiology of joint hypermobility: a population study of female twins.' *Arthritis and Rheumatism*, 50(8) pp. 2640-2644.

Hall, E. C. R., Almeida, S. S., Heffernan, S. M., Lockey, S. J., Herbert, A. J., Callus, P., Day, S. H., Pedlar, C. R., Kipps, C., Collins, M., Pitsiladis, Y. P., Bennett, M. A., Kilduff, L. P., Stebbings, G. K., Erskine, R. M. and Williams, A. G. (2021) 'Genetic polymorphisms related to vo2max adaptation are associated with elite rugby union status and competitive marathon performance.' *International Journal of Sports Physiology and Performance*, pp. 1-7.

Harner, C. D., Paulos, L. E., Greenwald, A. E., Rosenberg, T. D. and Cooley, V. C. (1994) 'Detailed analysis of patients with bilateral anterior cruciate ligament injuries.' *The American Journal of Sports Medicine*, 22(1) pp. 37-43.

Haseler, C. M., Carmont, M. R. and England, M. (2010) 'The epidemiology of injuries in English youth community rugby union.' *British Journal of Sports Medicine*, 44(15) pp. 1093-1099.

Hashimoto, T., Nobuhara, K. and Hamada, T. (2003) 'Pathologic evidence of degeneration as a primary cause of rotator cuff tear.' *Clinical Orthopaedics and Related Research*, (415) pp. 111.

Heffernan, S. M., Kilduff, L. P., Day, S. H., Pitsiladis, Y. P. and Williams, A. G. (2015) 'Genomics in rugby union: A review and future prospects.' *European Journal of Sport Science*, 15(6) pp. 460-468.

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

Heffernan, S. M., Stebbings, G. K., Kilduff, L. P., Erskine, R. M., Day, S. H., Morse, C. I., McPhee, J. S., Cook, C. J., Vance, B., Ribbans, W. J., Raleigh, S. M., Roberts, C., Bennett, M. A., Wang, G., Collins, M., Pitsiladis, Y. P. and Williams, A. G. (2017b) 'Fat mass and obesity associated (FTO) gene influences skeletal muscle phenotypes in non-resistance trained males and elite rugby playing position.' *BMC Genomics*, 18(1), doi: 10.1186/s12863-017-0470-1.

Heffernan, S. M., Kilduff, L. P., Erskine, R. M., Day, S. H., McPhee, J. S., McMahon, G. E., Stebbings, G. K., Neale, J. P., Lockey, S. J., Ribbans, W. J., Cook, C. J., Vance, B., Robert, C., Bennett, M. A., Wang, G., Collins, M., Pitsiladis, Y. P. and Williams, A. G. (2016) 'Association of ACTN3 R577X but not ACE I/D gene variants with elite rugby player status and playing position.' *Physiological Genomics*, 48(3) pp. 196-121.

Hess, G. W. (2010) 'Achilles tendon rupture: a review of etiology, population, anatomy, risk factors, and injury prevention.' *Foot and Ankle Specialist,* 3(1) p. 29.

Ho, M. S. P., Böse, K., Mokkapati, S., Nischt, R. and Smyth, N. (2008) 'Nidogens-Extracellular matrix linker molecules.' *Microscopy Research and Technique*, 71(5) pp. 387-395.

Hodgson, L., Gissane, C., Gabbett, T. J. and King, D. A. (2007) 'For debate: consensus injury definitions in team sports should focus on encompassing all injuries.' *Clinical journal of Sport Medicine*, 17(3) p. 188.

Hosking, L., Lumsden, S., Lewis, K., Yeo, A., McCarthy, L., Bansal, A., Riley, J., Purvis, I. and Xu, C. F. (2004) 'Detection of genotyping errors by Hardy-Weinberg equilibrium testing.' *European Journal of Human Genetics*, 12(5), pp. 395-399.

Hoskins, W., Pollard, H., Hough, K. and Tully, C. (2006) 'Injury in rugby league.' *Journal of Science and Medicine in Sport*, 9(1) pp. 46-56.

How TaqMan Assays Work - UK //www.thermofisher.com/uk/en/home/life-science/pcr/real-time-pcr/real-time-pcr-learning-center/real-time-pcr-basics/how-taqman-assays-work.html (accessed August 6t^{h}, 2021).

Johnson, J. S., Morscher, M. A., Jones, K. C., Moen, S. M., Klonk, C. J., Jacquet, R. and Landis, W. J. (2015) 'Gene expression differences between ruptured anterior cruciate ligaments in young male and female subjects.' *The Journal of Bone and Joint Surgery,* 97(1) pp. 71.

Jones, G. C., Corps, A. N., Pennington, C. J., Clark, I. M., Edwards, D. R., Bradley, M. M., Hazleman, B. L. and Riley, G. P. (2006) 'Expression profiling of metalloproteinases and tissue inhibitors of metalloproteinases in normal and degenerate human Achilles tendon.' *Arthritis and Rheumatism*, 54(3) pp. 832-842.

Karousou, E., Ronga, M., Vigetti, D., Passi, A. and Maffulli, N. (2008) 'Collagens, proteoglycans, MMP-2, MMP-9 and TIMPs in human achilles tendon rupture.' *Clinical Orthopaedics and Related Research*, 466(7) pp. 1577-1582.

Kaux, J.-F., Forthomme, B., Goff, C. L., Crielaard, J.-M. and Croisier, J.-L. (2011) 'Current opinions on tendinopathy.' *Journal of Sports Science & Medicine*, 10(2) pp. 238-253.

Kemp, S., Brooks, J., Cross, M., Morrow, P., Williams, S., Anstiss, T., Smith, A., Taylor, A., Trewartha, G., Widdowson, J. and Stokes, K. (2015) 'England professional rugby injury surveillance project 2013-14 season report.' [Online] [Accessed on 17th August 2016] http://www.englandrugby.com/mm/Document/General/General/01/30/80/08/EnglandProfessionalRugbyInjurySurveillanceProjectReport2013_2014_Neutral.pdf

Kemp, S., West, S., Brooks, J., Cross, M., Williams, S., Anstiss, T., Smith, A., Bryan, R., Hibbins-Butler, R., O'Leary, B. and Stokes, K. (2018) 'English Professional Rugby Injury Surveillance Project 2016-2017 Season Report.' [Online] [Accessed on 29th August 2018] https://www.englandrugby.com/mm/Document/General/General/01/32/91/95/InjurySurveillanceReport2016-17_English.pdf

Kemp, S., Brooks, J., Cross, M., Morrow, P., Williams, S., Anstiss, T., Smith, A., Taylor, A., Palmer, C., Bryan, R., Trewartha, G. and Stokes, K. (2016) 'England professional rugby injury surveillance project 2014-15 season report.' [Online] [Accessed on 17th August 2016] http://www.englandrugby.com/mm/Document/General/General/01/31/72/86/InjurySurveillanceReport_2014-15_SINGLE_22Mar16_English.pdf

Kemp, S., West, S., Brooks, J., Cross, M., Williams, S., Anstiss, T., Smith, A., Bryan, R., Henderson, L., Locke, D., Reynolds, A. and Stokes, K. (2019) '*English Professional*

Rugby Injury Surveillance Project 2017-2018 Season Report.' [Online] [Accessed on 20th August 2019].https://www.englandrugby.com//dxdam/96/960006d9-269d-4250-a15f-d9e62f8bfe70/PRISP_1718.pdf.

Kemp, S., West, S., Anstiss, T., Brooks, J., Bryan, R., Cross, M., Henderson, L., Locke, D., O'Leary, B., Smith, A., Williams, S. and Stokes, K. (2020) 'England Professional Rugby Injury Surveillance Project Season Report 2018-19.' [Online] [Accessed on 15th September 2020].https://www.englandrugby.com/dxdam/89/8994c1ba-17df-47a3-a4c6-8878ed6728d0/England%20PRISP%2018.19.pdf.

Kemp, S., Brooks, J., West, S., Cross, M., Morrow, P., Williams, S., Anstiss, T., Smith, A., Taylor, A., Palmer, C., Bryan, R., Hibbins-Butler, R. and Stokes, K. (2017) 'English Professional Rugby Injury Surveillance Project 2015-2016 season report.' [Online] [Accessed on 31 March 2017] http://www.englandrugby.com/mm/Document/General/General/01/32/25/17/1516_PRISP_Annual_Report_FINAL(withcontentspage)_English.pdf.

Khan, K. M. and Scott, A. (2009) 'Mechanotherapy: how physical therapists' prescription of exercise promotes tissue repair.' *British Journal of Sports Medicine*, 43(4) pp. 247-252.

Khoschnau, S., Melhus, H., Jacobson, A., Rahme, H., Bengtsson, H., Ribom, E., Grundberg, E., Mallmin, H. and Michaelsson, K. (2008) 'Type I collagen alpha1 Sp1 polymorphism and the risk of cruciate ligament ruptures or shoulder dislocations.' *American Journal of Sports Medicine*, 36(12) p. 2432.

Khoury El, L., Posthumus, M., Collins, M., Ribbans, W. and Raleigh, S. (2014) 'The COL5A1 gene and risk of achilles tendon pathology in a british cohort.' *British Journal of Sports Medicine*, 48(18) p. Sul 2 A54.

Kim, J. and Bartel, D. P. (2009) 'Allelic imbalance sequencing reveals that single-nucleotide polymorphisms frequently alter microRNA-directed repression.' *Nature Biotechnology*, 27(5) pp. 472-477.

Kim, S. K., Nguyen, C., Avins, A. L. and Abrams, G. D. (2021) 'Three genes associated with anterior and posterior cruciate ligament injury : a genome-wide association analysis.' *Bone Joint Open*, 2(6), Jun, 2021/06/26, pp. 414-421.

Kim, S. K., Roos, T. R., Roos, A. K., Kleimeyer, J. P., Ahmed, M. A., Goodlin, G. T., Fredericson, M., Ioannidis, J. P. A., Avins, A. L. and Dragoo, J. L. (2017) 'Genome-wide association screens for Achilles tendon and ACL tears and tendinopathy.' *PLOS ONE*, 12(3), doi: 10.1371/journal.pone.0170422.

King, D., Gissane, C., Clark, T. and 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 and Coaching*, 9(2) pp. 417-431.

King, D. A., Gabbett, T. J., Gissane, C. and Hodgson, L. (2009) 'Epidemiological studies of injuries in rugby league: Suggestions for definitions, data collection and reporting methods.' *Journal of Science and Medicine in Sport*, 12(1) pp. 12-19.

Kluivers, K. B., Dijkstra, J. R., Heniks, J. C. M., Lince, S. L., Vierhout, M. E. and Kempen, L. C. L. v. (2009) 'COL3A1 2209G>A is a predictor of pelvic organ prolapse.' *International Urogynecology Journal*, 20(9) pp. 1113-1118.

Kvist, M. (1994) 'Achilles tendon injuries in athletes.' *Sports Medicine*, 18(3) pp. 173.

Laguette, M.-J., Abrahams, Y., Prince, S. and Collins, M. (2011) 'Sequence variants within the 3′-UTR of the COL5A1 gene alters mRNA stability: Implications for musculoskeletal soft tissue injuries.' *Matrix Biology*, 30(5) pp. 338-345.

Lahiri, D. K. and 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(19) pp. 5444-5444.

Lapiere, C. M., Nusgens, B. and Pierard, G. E. (1977) 'Interaction between collagen type I and type III in conditioning bundles organization.' *Connective Tissue Research*, 5(1) p. 21.

Larruskain, J., Lekue, J. A., Martin-Garetxana, I., Barrio, I., McCall, A. and Gil, S. M. (2021) 'Injuries are negatively associated with player progression in an elite football academy.' *Science and Medicine in Football*, pp. 1-10.

Lau, N. C. and Lai, E. C. (2005) 'Diverse roles for RNA in gene regulation.' *Genome Biology,* 6(4) pp. 315.

Leadbetter, W. B. (1992) 'Cell-matrix response in tendon injury.' *Clinics in Sports Medicine*, 11(3) p. 533.

Levine, J. W., Kiapour, A. M., Quatman, C. E., Wordeman, S. C., Goel, V. K., Hewett, T. E. and Demetropoulos, C. K. (2013) 'Clinically relevant injury patterns after an anterior cruciate ligament injury provide insight into injury mechanisms.' *The American Journal of Sports Medicine*, 41(2) pp. 385-395.

Liefhebber, J. M. P., Punt, S., Spaan, W. J. M. and van Leeuwen, H. C. (2010) 'The human collagen betagalactosyltransferase, GLT25D1, is a soluble endoplasmic reticulum localized protein.' *BMC Cell Biology*, 11 p. 33.

Liu, X., Wu, H., Byrne, M., Krane, S. and Jaenisch, R. (1997) 'Type III collagen is crucial for collagen I fibrillogenesis and for normal cardiovascular development.' *Proceedings of the National Academy of Sciences of the United States of America*, 94(5) pp. 1852-1856.

Longo, U. G., Ronga, M. and Maffulli, N. (2009) 'Acute ruptures of the achilles tendon.' *Sports Medicine and Arthroscopy Review,* 17(2) p. 127.

Maffulli, N., Wong, J. and Almekinders, L. C. (2003) 'Types and epidemiology of tendinopathy.' *Clinics in Sports Medicine*, 22(4) p. 675.

Magnan, B., Bondi, M., Pierantoni, S. and Samaila, E. (2014) 'The pathogenesis of Achilles tendinopathy: a systematic review.' *Foot and Ankle Surgery*, 20(3) pp. 154-159.

Magnusson, K., Turkiewicz, A., Hughes, V., Frobell, R. and Englund, M. (2020) 'High genetic contribution to anterior cruciate ligament rupture: Heritability ~69.' *British Journal of Sports Medicine*,

Magnusson, S. P., Langberg, H. and Kjaer, M. (2010) 'The pathogenesis of tendinopathy: balancing the response to loading.' *Nature Reviews Rheumatology*, 6(5) pp. 262-268.

Magnusson, S. P., Qvortrup, K., Larsen, J. O., Rosager, S., Hanson, P., Aagaard, P., Krogsgaard, M. and Kjaer, M. (2002) 'Collagen fibril size and crimp morphology in ruptured and intact Achilles tendons.' *Matrix Biology*, 21(4) pp. 369-377.

Majewski, M., Susanne, H. and Klaus, S. (2006) 'Epidemiology of athletic knee injuries: A 10-year study.' *The Knee*, 13(3) pp. 184-188.

Malila, S., Yuktanandana, P., Saowaprut, S., Jiamjarasrangsi, W. and Honsawek, S. (2011) 'Association between matrix metalloproteinase-3 polymorphism and anterior cruciate ligament ruptures.' *Genetics and Molecular Research*, 10(4) pp. 4158-4165.

Malkki, M. and Petersdorf, E. W. (2012) 'Genotyping of single nucleotide polymorphisms by 5' nuclease allelic discrimination.' *Methods in Molecular Biology,* 882*,* pp. 173-182.

Mannion, S., Mtintsilana, A., Posthumus, M., van der Merwe, W., Hobbs, H., Collins, M. and September, A. V. (2014) 'Genes encoding proteoglycans are associated with the risk of anterior cruciate ligament ruptures.' *British Journal of Sports Medicine*, 48(22) pp. 1640-1646.

Matzke, M. A. and Birchler, J. A. (2005) 'RNAi-mediated pathways in the nucleus.' *Nature Reviews Genetics*, 6(1) pp. 24-35.

Mayer, U., Kohfeldt, E. and Timpl, R. (1998) 'Structural and genetic analysis of laminin-nidogen interaction.' *Annals of the New York Academy of Sciences*, 857, pp. 130-142.

Mc Fie, S., Abrahams, S., Patricios, J., Suter, J., Posthumus, M. and September, A. V. (2018) 'The association between COMT rs4680 and 5-HTTLPR genotypes and concussion history in South African rugby union players.' *Journal of Sports Science*, 36(8), pp. 920-933.

McLaughlin, J. S., Linsenmayer, T. F. and Birk, D. E. (1989) 'Type V collagen synthesis and deposition by chicken embryo corneal fibroblasts in vitro.' *Journal of Cell Science*, 94 (Pt 2) pp. 371.

Meeuwise, W. H. (1994) 'Assessing causation in sport injury: a multifactorial model.' *Clinical Journal of Sport Medicine*, 4(3) pp. 166-170.

Minamitani, T., Ikuta, T., Saito, Y., Takebe, G., Sato, M., Sawa, H., Nishimura, T., Nakamura, F., Takahashi, K., Ariga, H. and Matsumoto, K.-i. (2004) 'Modulation of collagen fibrillogenesis by tenascin-X and type VI collagen.' *Experimental Cell Research*, 298(1) pp. 305-315.

Mokone, G. G., Schwellnus, M. P., Noakes, T. D. and Collins, M. (2006) 'The COL5A1 gene and Achilles tendon pathology.' *Scandinavian Journal of Medicine & Science in Sports*, 16(1) pp. 19-26.

Mokone, G. G., Gajjar, M., September, A. V., Schwellnus, M. P., Greenberg, J., Noakes, T. D. and Collins, M. (2005) 'The guanine-thymine dinucleotide repeat polymorphism within the tenascin-c gene is associated with achilles tendon injuries.' *The American Journal of Sports Medicine*, 33(7) pp. 1016-1021.

Montgomery, C., Blackburn, J., Withers, D., Tierney, G., Moran, C. and Simms, C. (2018) 'Mechanisms of ACL injury in professional rugby union: a systematic video analysis of 36 cases.' *British Journal of Sports Medicine*, 52(15), pp. 994-1001.

Moore, I. S., Ranson, C. and Mathema, P. (2015) 'Injury risk in international rugby union: Three-year injury surveillance of the welsh national team.' *Orthopaedic Journal of Sports Medicine*, 3(7), doi: 2325967115596194.

Moore, J. H., Gilbert, J. C., Tsai, C.-T., Chiang, F.-T., Holden, T., Barney, N. and White, B. C. (2006) 'A flexible computational framework for detecting, characterizing, and interpreting statistical patterns of epistasis in genetic studies of human disease susceptibility.' *Journal of Theoretical Biology*, 241(2) pp. 252-261.

Morrey, M. E., Dean, B. J. F., Carr, A. J. and Morrey, B. F. (2013) 'Tendinopathy: Same disease different results—why?' *Operative Techniques in Orthopaedics*, 23(2) pp. 39-49.

Mukherjee, S. (2015) 'Retrospective designs in sports injury surveillance studies: All is not lost.' *Sports and Exercise Medicine*, 1(5) pp. 164-166.

Munafò, M. R. and Flint, J. (2004) 'Meta-analysis of genetic association studies.' *Trends in Genetics*, 20(9) pp. 439-444.

Nebelung, W. and Wuschech, H. (2005) 'Thirty-five years of follow-up of anterior cruciate ligament-deficient knees in high-level athletes.' *Arthroscopy,* 21(6) pp. 696.

Nell, E. M., van der Merwe, L., Cook, J., Handley, C. J., Collins, M. and September, A. V. (2012) 'The apoptosis pathway and the genetic predisposition to Achilles tendinopathy.' *Journal of Orthopaedic Research*, 30(11) pp. 1719-1724.

Neufeld, G., Cohen, T., Gengrinovitch, S. and Poltorak, Z. (1999) 'Vascular endothelial growth factor (VEGF) and its receptors.' *FASEB Journal,* 13(1) pp. 9.

Newton-Cheh, C. and Hirschhorn, J. N. (2005) 'Genetic association studies of complex traits: design and analysis issues.' *Mutation research*, 573(1-2) pp. 54.

Nourissat, G., Berenbaum, F. and Duprez, D. (2015) 'Tendon injury: from biology to tendon repair.' *Nature Reviews Rheumatology*, 11(4) pp. 223-233.

Nourissat, G., Houard, X., Sellam, J., Duprez, D. and Berenbaum, F. (2013) 'Use of autologous growth factors in aging tendon and chronic tendinopathy.' *Frontiers in bioscience*, 5 pp. 911-21

O'Connell, K., Knight, H., Ficek, K., Leonska-Duniec, A., Maciejewska-Karlowska, A., Sawczuk, M., Stepien-Slodkowska, M., O'Cuinneagain, D., van der Merwe, W., Posthumus, M., Cieszczyk, P. and Collins, M. (2015) 'Interactions between collagen gene variants and risk of anterior cruciate ligament rupture.' *European Journal of Sport Science*, 15(4) pp. 341-350.

Orchard, J. (2004) 'Missed time through injury and injury management at an NRL club.' *Sport Health*, 22(1) pp. 11-19.

Orchard, J. and Hoskins, W. (2007) 'For debate: consensus injury definitions in team sports should focus on missed playing time.' *Clinical Journal of Sport Medicine*, 17(3) pp. 192.

Pajala, A., Melkko, J., Leppilahti, J., Ohtonen, P., Soini, Y. and Risteli, J. (2009) 'Tenascin-C and type I and III collagen expression in total Achilles tendon rupture. An immunohistochemical study.' *Histology and Histopathology*, 24(10) p. 1207.

Petersen, W., Unterhauser, F., Pufe, T., Zantop, T., Südkamp, N. P. and Weiler, A. (2003) 'The angiogenic peptide vascular endothelial growth factor (VEGF) is expressed during the remodeling of free tendon grafts in sheep.' *Archives of Orthopaedic and Trauma Surgery*, 123(4) pp. 168-174.

Petersen, W., Pufe, T., Zantop, T., Tillmann, B., Tsokos, M. and Mentlein, R. (2004) 'Expression of VEGFR-1 and VEGFR-2 in degenerative Achilles tendons.' *Clinical Orthopaedics And Related Research*, (420) pp. 286-291.

Pitsiladis, Y. P., Tanaka, M., Eynon, N., Bouchard, C., North, K. N., Williams, A. G., Collins, M., Moran, C. N., Britton, S. L., Fuku, N., Ashley, E. A., Klissouras, V., Lucia, A., Ahmetov, II, de Geus, E., Alsayrafi, M. and Athlome Project, C. (2016) 'Athlome Project

Consortium: a concerted effort to discover genomic and other "omic" markers of athletic performance.' *Physiological Genomics*, 48(3) pp. 183-190.

Posthumus, M. (2009) 'The COL5A1 gene is associated with increased risk of anterior cruciate ligament ruptures in female participants.' *The American Journal of Sports Medicine*, 37(11) pp. 2234-2240.

Posthumus, M., September, A. V., Schwellnus, M. P. and Collins, M. (2009a) 'Investigation of the Sp1-binding site polymorphism within the COL1A1 gene in participants with Achilles tendon injuries and controls.' *Journal of Science and Medicine in Sport*, 12(1) pp. 184-189.

Posthumus, M., September, A., O'Cuinneagain, D., Van Der Merwe, W., Schwellnus, M. and Collins, M. (2009b) 'The COL5A1 gene is associated with increased risk of anterior cruciate ligament rupture in female participants.' *Medicine & Science in Sports & Exercise*, 37(11), pp. 2234-40

Posthumus, M., September, A. V., O'Cuinneagain, D., van der Merwe, W., Schwellnus, M. P. and Collins, M. (2010a) 'The association between the COL12A1 gene and anterior cruciate ligament ruptures.' *British Journal of Sports Medicine*, 44(16) pp. 1160-1165.

Posthumus, M., September, A. V., Keegan, M., O'Cuinneagain, D., Van Der Merwe, W., Schwellnus, M. P. and Collins, M. (2009c) 'Genetic risk factors for anterior cruciate ligament ruptures: COL1A1 gene variant.' *British Journal of Sports Medicine*, 43(5) pp. 352-356.

Posthumus, M., Collins, M., Cook, J., Handley, C. J., Ribbans, W. J., Smith, R. K. W., Schwellnus, M. P. and Raleigh, S. M. (2010b) 'Components of the transforming growth factor-β family and the pathogenesis of human achilles tendon pathology-a genetic association study.' *Rheumatology*, 49(11) pp. 2090-2097.

Posthumus, M., Collins, M., van der Merwe, L., O'Cuinneagain, D., van der Merwe, W., Ribbans, W. J., Schwellnus, M. P. and 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(4) pp. 523-533.

Provenzano, P. P., Heisey, D., Hayashi, K., Lakes, R. and Ray Vanderby, Jr. (2002) 'Subfailure damage in ligament: a structural and cellular evaluation.' *Journal of Applied Physiology*, 92(1) pp. 362-371.

Pufe, T., Petersen, W., Tillmann, B. and Mentlein, R. (2001) 'The angiogenic peptide vascular endothelial growth factor is expressed in foetal and ruptured tendons.' *Virchows Archiv*, 439(4) pp. 579-585.

Pufe, T., Petersen, W. J., Mentlein, R. and Tillmann, B. N. (2005) 'The role of vasculature and angiogenesis for the pathogenesis of degenerative tendons disease.' *Scandinavian Journal of Medicine & Science in Sports*, 15(4) pp. 211-222.

QIAamp DNA Blood Mini Kit. [Online] [Accessed April 2nd, 2021]. https://www.qiagen.com/no/products/discovery-and-translational-research/dna-rna-purification/dna-purification/genomic-dna/qiaamp-dna-blood-mini-kit/?clear=true#orderinginformation.

Quarrie, K. L. (2001) 'The New Zealand rugby injury and performance project. VI. A prospective cohort study of risk factors for injury in rugby union football.' *British Journal of Sports Medicine*, 35(3) pp. 157-166.

Quarrie, K. L. (2008) 'Tackle injuries in professional Rugby Union.' *The American Journal of Sports Medicine*, 36(9) pp. 1705-1716.

Quatman, C. E., Kiapour, A., Myer, G. D., Ford, K. R., Demetropoulos, C. K., Goel, V. K. and Hewett, T. E. (2011) 'Cartilage pressure distributions provide a footprint to define female anterior cruciate ligament injury mechanisms.' *The American Journal of Sports Medicine*, 39(8) pp. 1706.

Rafferty, J., Ranson, C., Oatley, G., Mostafa, M., Mathema, P., Crick, T. and Moore, I. S. (2018) 'On average, a professional rugby union player is more likely than not to sustain a concussion after 25 matches.' *British Journal of Sports Medicine*, 53(15), pp.969-973.

Rahim, M., Hobbs, H., van der Merwe, W., Posthumus, M., Collins, M. and September, A. V. (2017) 'Investigation of angiogenesis genes with anterior cruciate ligament rupture risk in a South African population.' *Journal of Sports Sciences*, 36(5), pp.551-557.

Rahim, M., Gibbon, A., Hobbs, H., van der Merwe, W., Posthumus, M., Collins, M. and September, A. V. (2014) 'The association of genes involved in the angiogenesis-associated signaling pathway with risk of anterior cruciate ligament rupture: Angiogenesis and acl rupture risk.' *Journal of Orthopaedic Research*, 32(12) pp. 1612-1618.

Rahim, M., El Khoury, L., Raleigh, S. M., Ribbans, W. J., Posthumus, M., Collins, M. and September, A. V. (2016) 'Human genetic variation, sport and exercise medicine, and Achilles tendinopathy:Role for angiogenesis-associated genes.' *OMICS: A Journal of Integrative Biology*, 20(9) pp. 520-527.

Raleigh, S. M., Van Der Merwe, L., Ribbans, W. J., Smith, R. K. W., Schwellnus, M. P. and 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(7) pp. 514-520.

Ribbans, W. J. and Collins, M. (2013) 'Pathology of the tendo Achillis, do our genes contribute?' *Bone & Joint Journal*, 95B(3) pp. 305-313.

Riley, G. (2004) 'The pathogenesis of tendinopathy. A molecular perspective.' *Rheumatology*, 43(2) pp. 131-142.

Riley, G. (2005a) 'Chronic tendon pathology: molecular basis and therapeutic implications.' *Expert Reviews in Molecular Medicine*, 7(5) p. 1.

Riley, G. P. (2005b) 'Gene expression and matrix turnover in overused and damaged tendons.' *Scandinavian Journal of Medicine & Science in Sports*, 15(4) pp. 241-251.

Riley, G. P., Harrall, R. L., Constant, C. R., Chard, M. D., Cawston, T. E. and Hazleman, B. L. (1994a) 'Tendon degeneration and chronic shoulder pain: changes in the collagen composition of the human rotator cuff tendons in rotator cuff tendinitis.' *Annals of the Rheumatic Diseases*, 53(6) pp. 359-366.

Riley, G. P., Harrall, R. L., Constant, C. R., Chard, M. D., Cawston, T. E. and Hazleman, B. L. (1994b) 'Glycosaminoglycans of human rotator cuff tendons: changes with age and in chronic rotator cuff tendinitis.' *Annals of the Rheumatic Diseases*, 53(6) pp. 367-376.

Riva, A. and Kohane, I. S. (2004) 'A SNP-centric database for the investigation of the human genome.' *BMC Bioinformatics*, 5(1) pp. 33-33.

Roberts, S. P., Trewartha, G., England, M., Shaddick, G. and Stokes, K. (2013) 'Epidemiology of time-loss injuries in English community-level rugby union.' *British Medical Journal Open*, 3(11) p. e003998. doi: 10.1136/bmjopen-2013-003998

Rothman, K. J. (2012) *Epidemiology : an introduction.* New York, NY: Oxford University Press.

Roulet, M., Ruggiero, F., Karsenty, G. and LeGuellec, D. (2007) 'A comprehensive study of the spatial and temporal expression of the col5a1 gene in mouse embryos: a clue for understanding collagen V function in developing connective tissues.' *Cell and Tissue Research*, 327(2) pp. 323-332.

Ruiz, J. R., GÃ³mez-Gallego, F. l., Santiago, C., GonzÃ¡lez-Freire, M., Verde, Z., Foster, C. and Lucia, A. (2009) 'Is there an optimum endurance polygenic profile?' *The Journal of Physiology*, 587(7) pp. 1527-1534.

Ruiz, J. R., Eynon, N., Meckel, Y., Fiuza-Luces, C., Santiago, C., Gómez-Gallego, F., Oliveira, J. and Lucia, A. (2011) 'GNB3 C825T Polymorphism and elite athletic status: A replication study with two ethnic groups.' *International Journal of Sports Medicine*, 32(2), Feb, 2010/11/27, pp. 151-153.

Santiago, C., Ruiz, J. R., Muniesa, C. A., González-Freire, M., Gómez-Gallego, F. and Lucia, A. (2010) 'Does the polygenic profile determine the potential for becoming a world-class athlete? Insights from the sport of rowing.' *Scandinavian Journal of Medicine and Science in Sports*, 20(1), Feb, 2009/05/09, pp. e188-194.

Saunders, C. J., van der Merwe, L., Posthumus, M., Cook, J., Handley, C. J., Collins, M. and September, A. V. (2013) 'Investigation of variants within the COL27A1 and TNC genes and Achilles tendinopathy in two populations.' *Journal of Orthopaedic Research*, 31(4) pp. 632-637.

Schegg, B., Hülsmeier, A. J., Rutschmann, C., Maag, C. and Hennet, T. (2009) 'Core Glycosylation of Collagen Is Initiated by Two β(1-O)Galactosyltransferases.' *Molecular and Cellular Biology*, 29(4) p. 943.

Schwellnus, M. P., Thomson, A., Derman, W., Jordaan, E., Readhead, C., Collins, R., Morris, I., Strauss, O., Van der Linde, E. and Williams, A. (2014) 'More than 50% of players sustained a time-loss injury (>1 day of lost training or playing time) during the 2012 Super Rugby Union Tournament: a prospective cohort study of 17,340 player-hours.' *British Journal of Sports Medicine*, 48(17) pp. 1306-1315.

Sedeaud, A., Vidalin, H., Tafflet, M., Marc, A. and Toussaint, J. F. (2013) 'Rugby morphologies:"bigger and taller", reflects an early directional selection.' *Journal of Sport Medicine and Physical Fitness*, 53(2) pp. 185-191.

Selvanetti, A., Cipolla, M. and Puddu, G. (1997) 'Overuse tendon injuries: Basic science and classification.' *Operative Techniques in Sports Medicine*, 5(3) pp. 110-117.

September, A. V., Posthumus, M. and Collins, M. (2012) 'Application of genomics in the prevention, treatment and management of Achilles tendinopathy and anterior cruciate ligament ruptures.' *Recent Patents on DNA & Gene Sequences*, 6(3) p. 216.

September, A. V., Mokone, G. G., Schwellnus, M. P. and Collins, M. (2006) 'Genetic risk factors for Achilles tendon injuries.' *International Sports Medicine Journal*, 7(3) pp. 201-215.

September, A. V., Cook, J., Handley, C. J., Van Der Merwe, L., Schwellnus, M. P. and Collins, M. (2009) 'Variants within the COL5A1 gene are associated with Achilles tendinopathy in two populations.' *British Journal of Sports Medicine*, 43(5) pp. 357-365.

Seward, H., Orchard, J., Hazard, H. and Collinson, D. (1993) 'Football injuries in Australia at the élite level.' *The Medical Journal of Australia*, 159(5) p. 298.

Sharma, P. and Maffulli, N. (2006) 'Biology of tendon injury: healing, modeling and remodeling.' *Journal of Musculoskeletal and Neuronal Interactions*, 6(2) p. 181.

Sivertsen, E. A., Haug, K. B. F., Kristianslund, E. K., Trøseid, A.-M. S., Parkkari, J., Lehtimäki, T., Mononen, N., Pasanen, K. and 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.' *The American Journal of Sports Medicine*, 47(1) pp. 52-58.

Smith, H. C., Vacek, P., Johnson, R. J., Slauterbeck, J. R., Hashemi, J., Shultz, S. and Beynnon, B. D. (2012a) 'Risk factors for anterior cruciate ligament injury: A review of the literature—Part 2: Hormonal, genetic, cognitive function, previous injury, and extrinsic risk factors.' *Sports Health*, 4(2) pp. 155-161.

Smith, H. C., Vacek, P., Johnson, R. J., Slauterbeck, J. R., Hashemi, J., Shultz, S. and Beynnon, B. D. (2012b) 'Risk factors for anterior cruciate ligament injury: A review of the literature — Part 1: Neuromuscular and anatomic risk.' *Sports Health*, 4(1) pp. 69-78.

Solé, X., Guinó, E., Valls, J., Iniesta, R. and Moreno, V. (2006) 'SNPStats: a web tool for the analysis of association studies.' *Bioinformatics*, 22(15) pp. 1928-1929.

Somerville, R. P. T., Oblander, S. A. and Apte, S. S. (2003) 'Matrix metalloproteinases: old dogs with new tricks.' *Genome biology*, 4(6) p. 216.

Stephenson, S., Gissane, C. and Jennings, D. (1996) 'Injury in rugby league: a four year prospective survey.' *British Journal of Sports Medicine*, 30(4) pp. 331-334.

Stępien-Słodkowska, M., Ficek, K., Eider, J., Leońska-Duniec, A., Maciejewska-Karłowska, A., Sawczuk, M., Zarębska, A., Jastrzębski, Z., Grenda, A., Kotarska, K. and Cięszczyk, P. (2013) 'The +1245g/t polymorphisms in the collagen type I alpha 1 (col1a1) gene in polish skiers with anterior cruciate ligament injury.' *Biology of Sport*, 30(1) p. 57.

Stępień-Słodkowska, M., Ficek, K., Maciejewska-Karłowska, A., Sawczuk, M., Ziętek, P., Król, P., Zmijewski, P., Pokrywka, A. and Cięszczyk, P. (2015a) 'Overrepresentation of the COL3A1 AA genotype in Polish skiers with anterior cruciate ligament injury.' *Biology of Sport*, 32(2) pp. 143-147.

Stępień-Słodkowska, M., Ficek, K., Kaczmarczyk, M., Maciejewska-Karłowska, A., Sawczuk, M., Leońska-Duniec, A., Stępiński, M., Ziętek, P., Król, P., Chudecka, M. and Cięszczyk, P. (2015b) 'The variants within the COL5A1 gene are associated with reduced risk of anterior cruciate ligament injury in skiers.' *Journal of Human Kinetics*, 45(1) pp. 103-111.

Tabor, H. K., Risch, N. J. and Myers, R. M. (2002) 'Candidate-gene approaches for studying complex genetic traits: practical considerations.' *Nature Reviews Genetics*, 3(5) pp. 391-397.

Thompson, J. I. and Czernuszka, J. T. (1995) 'The effect of two types of cross-linking on some mechanical properties of collagen.' *Bio-Medical Materials and Engineering*, 5(1) pp. 37.

Thornton, G. M. and Hart, D. A. (2011) 'The interface of mechanical loading and biological variables as they pertain to the development of tendinosis.' *Journal of Musculoskeletal and Neuronal Interactions*, 11(2) pp. 94.

Timpl, R. and Brown, J. C. (1996) 'Supramolecular assembly of basement membranes.' *BioEssays : news and reviews in molecular, cellular and developmental biology*, 18(2) p. 123.

Tranchevent, L.-C., Ardeshirdavani, A., ElShal, S., Alcaide, D., Aerts, J., Auboeuf, D. and Moreau, Y. (2016) 'Candidate gene prioritization with Endeavour.' *Nucleic Acids Research*, 44(W1) pp. W117-W121.

van Mechelen, W., Hlobil, H. and Kemper, H. C. (1992) 'Incidence, severity, aetiology and prevention of sports injuries. A review of concepts.' *Sports Medicine*, 14(2) p. 82.

Visse, R. and Nagase, H. (2003) 'Matrix Metalloproteinases and Tissue Inhibitors of Metalloproteinases: Structure, function, and biochemistry.' *Circulation Research: Journal of the American Heart Association*, 92(8) pp. 827-839.

Voleti, P. B., Buckley, M. R. and Soslowsky, L. J. (2012) 'Tendon healing: repair and regeneration.' *Annual Review of Biomedical Engineering*, 14 pp. 47-71.

von der Mark, K. (1981) 'Localization of collagen types in tissues.' *International Review of Connective Tissue Research*, 9 pp. 265-324.

von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., Vandenbroucke, J. P., for the, S. I. and Initiative, S. (2007) 'The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for reporting observational studies.' *Preventive Medicine*, 45(4) pp. 247-251.

von Porat, A., Roos, E. M., Roos, H., Medicinska, f., Department of Health, S., Department of, O., Ortopedi, L., Division, III, Faculty of, M., Sektion, III, Department of Clinical Sciences, L., Lunds, u., Fysioterapi, Physiotherapy, Institutionen för, h., Lund, U. and Institutionen för kliniska vetenskaper, L. (2004) 'High prevalence of osteoarthritis 14 years after an anterior cruciate ligament tear in male soccer players: a study of radiographic and patient relevant outcomes.' *Annals of the Rheumatic Diseases*, 63(3) pp. 269-273.

Wang, G., Padmanabhan, S., Wolfarth, B., Fuku, N., Lucia, A., Ahmetov, II, Cieszczyk, P., Collins, M., Eynon, N., Klissouras, V., Williams, A. and Pitsiladis, Y. (2013) 'Genomics of elite sporting performance: what little we know and necessary advances.' *Advances in Genetics*, 84, pp. 123-149.

Wang, J. H. C. (2006) 'Mechanobiology of tendon.' *Journal of Biomechanics*, 39(9) pp. 1563-1582.

Wang, J. H. C., Guo, Q. and Li, B. (2012) 'Tendon biomechanics and mechanobiology--a minireview of basic concepts and recent advancements.' *Journal of Hand Therapy*, 25(2) pp. 133.

West, S. W., Starling, L., Kemp, S., Williams, S., Cross, M., Taylor, A., Brooks, J. H. M. and Stokes, K. A. (2020) 'Trends in match injury risk in professional male rugby union: a 16-season review of 10 851 match injuries in the English Premiership (2002–2019): The Professional Rugby Injury Surveillance Project.' *British Journal of Sports Medicine*,55(12), pp.676-682.

Whitehouse, T., Orr, R., Fitzgerald, E., Harries, S. and McLellan, C. P. (2016) 'The epidemiology of injuries in Australian professional rugby union 2014 super rugby competition.' *Orthopaedic Journal of Sports Medicine*, 4(3). doi: 10.1177/2325967116634075.

Willard, J. M., Lee, D. A. and Holland, M. M. (1998) 'Recovery of DNA for PCR amplification from blood and forensic samples using a chelating resin.' *Methods in Molecular Biology*, 98, pp. 9-18.

Williams, A. G. and Folland, J. P. (2008) 'Similarity of polygenic profiles limits the potential for elite human physical performance.' *The Journal of Physiology*, 586(1) pp. 113-121.

Williams, A. G., Day, S. H., Lockey, S. J., Heffernan, S. M. and Erskine, R. M. (2014) 'Genomics as a practical tool in sport - have we reached the starting line?' *Cellular and Molecular Exercise Physiology*, 3(1), doi:10.7457/CMEP.V3I1.

Williams, S., Hume, P. A. and Kara, S. (2011) 'A Review of Football Injuries on Third and Fourth Generation Artificial Turfs Compared with Natural Turf.' *Sports Medicine*, 41(11) pp. 903-923.

Williams, S., Trewartha, G., Kemp, S. and Stokes, K. (2013) 'A meta-analysis of injuries in senior men's professional rugby union.' *Sports Medicine*, 43(10) pp. 1043-1055.

Williams, S., Trewartha, G., Kemp, S. P. T., Brooks, J. H. M., Fuller, C. W., Taylor, A. E., Cross, M. J. and Stokes, K. A. (2016) 'Time loss injuries compromise team success in Elite Rugby Union: a 7-year prospective study.' *British Journal of Sports Medicine*, 50(11) pp. 651-656.

Windt, J., Gabbett, T. J., Ferris, D. and Khan, K. M. (2016) 'Training load--injury paradox: is greater preseason participation associated with lower in-season injury risk in elite rugby league players?' *British Journal of Sports Medicine*, 51(8) p. 645.

Woo, S. L. Y., Debski, R. E., Zeminski, J., Abramowitch, S. D., Saw, S. S. and Fenwick, J. A. (2000) 'Injury and repair of ligaments and tendons.' *Annual Review of Biomedical Engineering*, 2(1) pp. 83-118.

Xu, Y. and Murrell, G. A. C. (2008) 'The basic science of tendinopathy.' *Clinical Orthopaedics and Related Research*, 466(7) pp. 1528-1538.

Yvert, T., Miyamoto-Mikami, E., Murakami, H., Miyachi, M., Kawahara, T. and Fuku, N. (2016) 'Lack of replication of associations between multiple genetic polymorphisms and endurance athlete status in Japanese population.' *Physiological Reports*, 4(20), e13003. doi: 10.14814/phy2.13003.

Zweig, M. H. and Campbell, G. (1993) 'Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine.' *Clinical Chemistry*, 39(4) pp. 561-577.