An investigation of DNA sequence variants in genes that regulate collagen fibrillogenesis and predisposition to musculoskeletal soft tissue injuries

By

Melanie Hay

This dissertation is presented for the degree of MSc (Med)

in Exercise Physiology

In the Department of Human Biology

Faculty of Health Sciences

University of Cape Town

South Africa

April 2013

UCT/MRC Research Unit for Exercise Science and Sports Medicine

Sports Science Institute of South Africa

Boundary Road, Newlands, 7700

Cape Town

South Africa

Plagiarism Declaration

I, Melanie Claire Hay, hereby declare that the work on which this dissertation is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole, nor any part of it, has been, is being, or is to be submitted for another degree in this, or any other university.

I empower the university to reproduce for the purpose of research, either the whole or any portion of the contents, in any manner whatsoever.

Signature:.....

Date:.....

Table of Contents

Acknowledgements

I would firstly like to thank my supervisors; Prof Malcolm Collins, Dr Mike Posthumus and Dr Alison September. It is rare to receive the type of support and guidance that I have received from this incredible team of supervisors. I am eternally grateful for their uncomplaining willingness to work late nights and weekends to help me meet my deadline. My dissertation is immeasurably strengthened by their close attention: Malcolm always saw the big picture and the implications of my results, and was instrumental in helping turn my results into a 'story'. Mike's approach was systematic, and he helped turn a sometimes jumbled mess into a structure that flowed and made sense. Alison, in addition to providing the occasional pep-talk and emotional support when needed, was meticulous in every detail, and didn't let a single flaw or inconsistency slip past her eagle eyes.

I am grateful to Dr Willem van der Merwe, Dr Hayden Hobbs and Dr Dion O'Cuinneagain for help recruiting ACL participants. I would also like to acknowledge Dr George Mokone, Prof Martin Schwellnus and all who were involved in recruiting participants for the Achilles tendinopathy study.

I would like to thank my fellow research "team-mates", Yoonus Abrahams, Shameemah Adams, Marilize Burger, Nancy Laguette, Sasha Mannion, Kevin O'Connell, Masouda Rahim and Colleen Saunders for a wonderful two years. It has been a privilege to be surrounded by such wonderful colleagues and friends.

A special thank you must go to my other student colleagues and friends, Caroline Dalton, Kasha Dickie, Liske Kotze, Theresa Mann, Phoebe Runciman, Stefano Scribani and Phillipa Skowno, who made each day at ESSM a joy. A very special mention goes to Kim Stephenson, who was more than a work friend, but someone from whom I received constant support during tough times.

The funding for this research was supplied by the National Research Foundation (NRF) and the Medical Research Council (MRC).

In additional, without personal funding for my studies, this endeavour would not have been possible. I am deeply appreciative of the financial aid provided by DAAD/NRF in 2011, the DST-NRF in 2012, and UCT and the Ernst and Ethel Ericksen Trust for scholarships in 2011 and 2012, as well as my supervisors Prof Malcolm Collins and Dr Alison September.

Finally, I would like to thank my family, who have been unwavering in their support. My parents, Wendy Hay and Malcolm Hay have pushed me when necessary, comforted me at other times, and constantly encouraged me. My siblings, Kirsten and Alastair have kept me from losing perspective, and my twin sister, Michelle Hay is absolutely my inspiration and my rock. She has been beside me this entire journey, and will be for the rest of my life.

Last but far from least, I'd like to thank my fiancé Morgan Commins, who believed in me, and supported me, and has waited patiently for me to finish this Masters.

Abbreviations

A	Adenine nucleotide
ACL	Anterior cruciate ligament
ADAMTS	A disintegrin and metalloproteinase with thrombospondin motifs
ADAMTS-2	A disintegrin and metalloproteinase with thrombospondin motifs-2 protein
ADAMTS-3	A disintegrin and metalloproteinase with thrombospondin motifs-3 protein
ADAMTS-14	A disintegrin and metalloproteinase with thrombospondin motifs-14 protein
AT	Chronic Achilles tendinopathy
ANOVA	One-way analysis of variance
AUS	Australia
BMI	Body mass index
BMP-1	Bone morphogenetic protein
bp	Base pairs
C	Cytosine nucleotide
CI	Confidence Interval
COL1A1	The gene encoding for the α1 chain of type I collagen
COL5A1	The gene encoding for the α1 chain of type V collagen
COL5A2	The gene encoding for the α2 chain of type V collagen
COL5A3	The gene encoding for the α3 chain of type V collagen
COL11A1	The gene encoding for the α1 chain of type XI collagen
COL11A2	The gene encoding for the α2 chain of type XI collagen
COL12A1	The gene encoding for the α1 chain of type XII collagen
COL14A1	The gene encoding for the α1 chain of type XIV collagen
CON	Control group
DNA	Deoxyribonucleic acid
ECM	Extracellular matrix
EDS	Ehlers-Danlos Syndrome
EDTA	Ethylenediaminetetraacetic acid
ER	Endoplasmic reticulum

FACIT	Fibril Associated Collagen with Interrupted Triple helixes
FDL	Flexor digitorum longus tendon
G	Guanine nucleotide
GAGs	Glycoaminoglycans
GDF5	The gene encoding the growth/differentiation factor 5 protein
GJL	Generalized joint laxity
GPC	Golgi-to-plasma membrane carriers
HWE	Hardy-Weinberg equilibrium
LDH	Lumbar disc herniation
L	Leucine amino acid
LD	Linkage disequilibrium
MCL	Medial collateral ligament
MIR608	The gene encoding the microRNA Hsa-miR-608
MMP	Matrix metalloproteinsase
MMP3	The gene encoding for the matrix metalloproteinase-3
MMP12	The gene encoding for the matrix metalloproteinase-12
mRNA	Messenger RNA
MTJ	Myotendinous junction
mTLD	Mammalian tolloid protein
N	Newtons
NCBI	Nation Centre for Biotechnology Information
NON	Non-contact mechanism of injury sub-group
OLJ	Osteoligamentous junction
OPLL	Ossification of the posterior longitudinal ligament of the spine
OR	Odds ratio
OSMED	Otospondylomegaepiphyseal dysplasia
OTJ	Osteotendinous junction
p	Short arm of chromosome
P	Proline amino acid
PCL	Posterior cruciate ligament

PCR	Polymerase chain reaction
PDI	Protein disulphide oxidase
q	Long arm of chromosome
RER	Rough endoplasmic reticulum
RFLP	Restriction fragment length polymorphism
ROM	Range of motion
S	Serine amino acid
SA	South Africa
SIFT	Sorting intolerant from intolerant
SLRPs	Small leucine-rich proteoglycans
SNP	Single nucleotide polymorphism
SSISA	Sports Science Institute of South Africa
STRP	Short tandem repeat polymorphisms
T	Thymine nucleotide
TEN	Achilles tendinopathy group
TLL-1	Tolloid-like protein
TNC	The gene encoding for tenacin-C
UTR	Untranslated region
WT	Wild type
WZS	Weissenbacher–Zweymuller syndrome
α1(I)	The α1 chain of type I collagen
α1(III)	The α1 chain of type III collagen
α1(V)	The α1 chain of type V collagen
α1(XI)	The α1 chain of type XI collagen
α1(XII)	The α1 chain of type XII collagen
α2(I)	The α2 chain of type I collagen
α2(V)	The α2 chain of type V collagen
α2(XI)	The α2 chain of type XI collagen
α3(V)	The α3 chain of type V collagen

List of Figures

List of Tables

Abstract

Background

Chronic Achilles tendinopathy (AT) and anterior cruciate ligament (ACL) rupture are multifactorial conditions caused by the complex interaction of several intrinsic and extrinsic risk factors. Polymorphisms within several genes, including those that encode collagen molecules, the basic structural component of ligaments and tendons, are intrinsic risk factors for the development of these injuries. Both type V and XI collagens regulate collagen fibril assembly in a concentration-dependent manner; and collagen fibril structure is associated with the mechanical properties of the tissue. Functional polymorphisms within the 3'-untranslated region of the *COL5A1* gene, which encodes the α1 chain of type V collagen, have previously been associated with AT risk and may also modulate the risk of ACL ruptures. In addition, the polymorphic *MIR608* gene encoding for Hsa-miR-608 microRNA plays a role in regulating type V collagen content in tendons and ligaments and has previously been associated with AT. The *MIR608* gene is therefore also a potential candidate gene for modulating the risk of ACL ruptures. Due to the functional homology between type V and type XI collagens, functional polymorphisms within the *COL11A1* and *COL11A2* genes, which encode type XI collagen, could also modulate the risk of AT and ACL injuries.

Aim

The aim of this dissertation was to use a case-control genetic study to investigate the association of polymorphisms within the *COL5A1*, *MIR608*, *COL11A1* and *COL11A2*, genes with AT and/or ACL injuries in Caucasian populations. These aims were explored in three studies:

i) Determine whether the *COL5A1* rs71746744 (-/AGGG) and rs1134170 (A/T) polymorphisms and the *MIR608* rs4919510 (C/G) polymorphism are associated with ACL rupture risk (Chapter 2).

ii) Determine whether the *COL11A1* rs3753841 (T/C) and rs1676486 (C/T) polymorphisms and the *COL11A2* rs1799907 (A/T) polymorphism are associated with ACL rupture risk. A secondary aim was to determine whether the *COL11A1* and *COL11A2* polymorphisms interact with *COL5A1* rs71746744 (-/AGGG) to modulate ACL rupture risk (Chapter 3).

iii) Determine whether the *COL11A1* rs3753841 (T/C) and rs1676486 (C/T) and *COL11A2* rs1799907 (A/T) polymorphisms are associated with AT risk, and investigate whether these polymorphisms interact with each other, or with the *COL5A1* rs71746744 (-/AGGG) polymorphism to modulate the risk of developing AT (Chapter 4).

Methods

In the ACL studies, 215 participants who had undergone reconstructive surgery for ACL rupture (ACL group) and 217 apparently healthy matched controls (CON group) were genotyped for polymorphisms within the *COL5A1* (rs71746744 and rs1134170) and the *MIR608* (rs4919510) genes (Chapter 2); as well as functional polymorphisms within the *COL11A1* (rs3753841 and rs1676486) and *COL11A2* (rs1799907) genes (Chapter 3). In the AT study, 182 participants with clinically diagnosed Achilles tendinopathy (TEN group), and 336 apparently healthy matched controls (CON group) were genotyped for the functional *COL11A1* and *COL11A2* polymorphisms (Chapter 4). In all studies the variants were examined for independent associations as well as for interactions with other investigated variants using two-way gene interaction, inferred haplotype or inferred pseudohaplotype analysis.

Results

The primary finding was that the *COL11A1* rs3753841 polymorphism was significantly associated with ACL ruptures. The TC genotype of *COL11A1* rs3753841 was significantly over-represented in the CON group (54.2%) compared to the ACL group (41.4%, p=0.009, OR=1.7, 95% CI= 1.1-2.5) (Chapter 2). None of the other investigated polymorphisms in the *COL5A1, MIR608, COL11A1* and *COL11A2* genes were independently associated with ACL rupture or AT risk.

The secondary findings were that various allele combinations of the investigated polymorphisms were significantly associated with either injury risk or protection. The inferred -/A haplotype constructed from *COL5A1* rs71746744 (-/AGGG) and rs1134170 (A/T) was significantly under-represented in the ACL rupture group (25.0%, p<0.001) and the sub-group with a confirmed non-contact (NON) mechanism of ACL injury (24.7%, p=0.015) compared to the control group (27.6%), suggesting a protective effect against ACL rupture (Chapter 2). The inferred CT haplotype constructed from *COL11A1* rs3753841 and rs1676486 was significantly under-represented (p=0.044) in the ACL group (17.5%) compared to the CON group (21.3%). In addition, the inferred CTA pseudohaplotype constructed from the *COL11A1* and *COL11A2* polymorphisms was significantly under-represented in the ACL group (p=0.010, 9.5%) and NON sub group (p=0.037, 10.3%) compared to the CON group (14.7%) (Chapter 3). The TCT pseudohaplotype constructed from *COL11A1* rs3753841, *COL11A1* rs1676486 and *COL11A2* rs1799907 was significantly over-represented (p=0.006) in the TEN (25.9%) compared to the CON (17.1%) group. The TCT(AGGG) pseudohaplotypes constructed using these type XI collagen polymorphisms and the functional

COL5A1 rs71746744 (-/AGGG) polymorphism was also significantly over-represented ($p<0.001$) in the TEN (25.2%) compared to the CON (9.1%) group (Chapter 4).

Conclusion

Genes encoding the structurally and functionally related type XI (*COL11A1* and *COL11A2*) and type V (*COL5A1*) collagens may potentially interact with one another to collectively modulate the risk for AT. They also appear to possibly interact to weakly modulate the risk of ACL ruptures, although further work is required to confirm this. Although there are no immediate clinical applications, the results of these studies provide additional evidence suggesting that inter-individual variations in collagen fibril assembly might be an important molecular mechanism in the aetiology of musculoskeletal soft tissue injuries.

Chapter 1: Review of Literature

1.1 Musculoskeletal Soft Tissue Injuries

The musculoskeletal system is responsible for enabling movement and locomotion and comprises the bones (skeleton), muscles, cartilage, tendons, ligaments and joints. The soft tissues of the musculoskeletal system include the muscles, tendons and ligaments, which together, play a role in moving the bones of the skeleton to bring about locomotion. These tissues are vulnerable to injury, especially when exposed to frequent use, or excessive forces. Musculoskeletal soft tissue injuries refer to injuries sustained to tendons, ligaments or muscles. These injuries can be the result of an acute trauma or chronic overuse; and can occur during occupational activities as well as during competitive and recreational sporting activities[1]. Their aetiology is multifactorial and comprises both intrinsic and extrinsic risk factors[2]. Genetic susceptibility is an intrinsic risk factor that is receiving increasing attention: the identification of predisposing genetic variants performs the dual function of providing insights into the molecular mechanisms that underlie these conditions, as well as providing tools in clinical practice to improve prevention, diagnosis and treatment of the injuries[3–5]. Two common sites of injury, that will be the focus of this dissertation, are the anterior cruciate ligament (ACL) of the knee[6, 7] and the Achilles tendon of the heel[1]. In order to understand how injuries to ligaments and tendons come about, it is necessary to first understand the anatomy of these structures in detail.

1.2 Structure of Ligaments and Tendons

1.2.1 Tissue Organization

Tendons and ligaments are the anatomical structures that connect muscle to bone[8], and bone to bone[9], respectively. Tendons are responsible for transferring the forces produced during muscle contraction into movement[8], while ligaments help to passively stabilize joints and guide them through their normal range of motion[9]. Tendons and ligaments are highly organized, hierarchically structured tissues[8–11] (Figure 1.1). The development of organized structure will be dealt with in detail in Section 1.3. Briefly, at the molecular level, individual collagen molecules (tropocollagen) aggregate to form collagen microfibrils that merge to form collagen fibrils[8]. Collagen fibrils are the smallest unit visible by electron microscope, and their diameter is highly regulated as it determines many of the properties of the tissue[8] (Section 1.4).

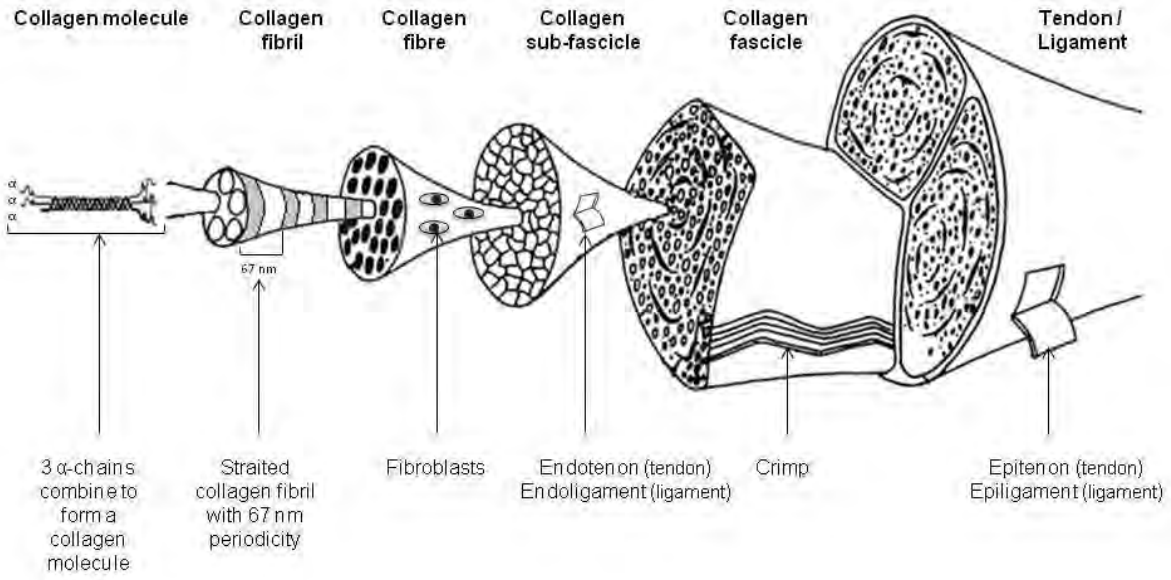


Figure 1.1 The hierarchical organization of tendon and ligaments. Three α-chains combine to form a collagen molecule. Collagen molecules aggregate to from fibrils; fibrils combine to form fibres, fibres combine to form subfascicles, subfascicles combine to form fascicles and fascicles combine to form a tendon or ligament. (Modified from Kastelic J et al., 1978[12] in Andrades, J.A., et al., 2011[13]).

Many collagen fibrils combine to form a collagen fibre, which is the basic unit of tendons and ligaments. The fibres combine to form a primary fibre bundle (subfascicle), which in turn combine to form a secondary fibre bundle (fascicle). Many secondary fibre bundles make up a tertiary bundle, and together, the tertiary bundles make up a tendon or ligament[8]. At each hierarchical stage, starting at the level of the collagen fibre, is a layer of endotenon or endoligament, which binds the fibres or bundles together[8]. Finally, a layer of epitenon or epiligament surrounds the entire tendon or ligament respectively[8].

A tendon or ligament may also be divided into distinct regions based on subtle differences in biochemical composition[14]. The myotendinous junction (MTJ) is the site at which the tendon attaches to muscle. The mid-substance of a tendon or ligament is the 'classic' tendon or ligament tissue composed predominantly of collagen fibres arranged parallel to the direction of force, while the osteotendinous junction (OTJ) or osteoligamentous junction (OLJ) is a region of fibrocartilage that transitions to bone and attaches the tendon or ligament to bone[14]. Regions of fibrocartilage occur at sites where tendons or ligaments experience compression, pass through soft-tissue pulleys, lie adjacent to other ligaments or traverse bony prominences, but their most common occurrence is at the insertions sites of tendons or ligaments to bone (enthesis)[14,15]. Injuries to tendons generally occur at or near the insertion site of a tendon, with the notable exception of the Achilles tendon,

which is more often injured in the midsubstance[14]. There are multiple reasons for increased injury at these sites: (i) they are more highly stressed, (ii) they are exposed to repeated shear and/ or compressive forces, and (iii) they are relatively less vascularised than tendon midsubstance[14].

1.2.2 The Molecular Components of Ligaments and Tendons

Tendons and ligaments contain a proportionately small cellular component, made up of 95% fibroblasts (tenocytes or ligamentocytes) and 5% other cell types[8]. The fibroblasts are responsible for secreting and maintaining the extracellular matrix (ECM), consisting of collagen fibrils and ground substance[8]. Fibroblasts are the living component of the tissue that is able to respond and adapt to mechanical loading forces, thus enabling the tendon or ligament to repair damage or adapt to changes in load over time[16]. Tendons and ligaments are poorly vascularised tissues, and thus have a well-developed anaerobic energy generation capacity[8]. However, this adaptation results in a metabolic activity level that is lower than that of skeletal muscle, resulting in a compromised healing capacity and a mismatch between the rate of muscle and tendon or ligament adaptation[8].

The ECM is the main component of ligaments and tendons and is made up of collagen fibres and ground substance. The collagen fibres are the major protein constituent of tendons and ligaments and form dense bands that are arranged parallel to the direction of force [8,11]. The ground substance consists of proteins such as elastin, proteoglycans, glycoproteins and glycoaminoglycans (GAGs) and roughly 68% water. The water content of the tissue is thought to be responsible for the viscoelastic properties of ligaments and tendons [9].

1.2.3 Collagen Types in Tendons and Ligaments

The most common collagen in tendons and ligaments is type I collagen, however ligaments also contain 9-12 % type III collagen, and small amounts of type IV, V, VI, XII and XIV collagen are found in both[8–10]. The OTJ or OLJ where the tendon or ligament attaches to bone, contains a high proportion of type II collagen, and some type IX and XI collagen[11]. The major fibrillar collagens (type I, II and III) form elongated fibrils that help maintain the structural integrity of fibrous connective tissue such as tendons, ligaments and cartilage[17]. Type V, XI, XXIV and XXVII collagen are minor fibrillar collagens that are less abundant[17,18]. However, type V collagen plays a vital role in tendons and ligaments, where it initiates fibril assembly[19–21] and regulates fibril diameter and lateral growth[22–27]. Type XI collagen performs a similar role to type V collagen[28], but is expressed more in cartilage[29] and at the OLJ[30]. In addition to fibril-forming collagens, there are a number of non-fibrillar collagens with diverse functions[17,18]. FACIT collagens, (fibril associated collagens with interrupted triple helices)

consist of type XI, XII, XIV and XX collagens; and are involved in the linkage of various collagen fibres, as well as forming interactions with molecules of the ECM[31].

1.3 Fibrillogenesis

Fibrillogenesis, or fibril assembly, is the highly regulated, multi-step process, by which individually translated collagen molecules are modified, transported, combined into larger molecules and eventually incorporated into the architecture of the mature tendon or ligament described in Section 1.2 [32]. The mechanical integrity of connective tissues is therefore ultimately dependent on proper regulation of the fibrillogenesis process[32].

Although fibrillogenesis is a continuous process, it is often described as occurring in three steps[32]. The first step occurs in close association with the fibroblasts and involves the synthesis, packaging, secretion and processing of collagen molecules, and their extracellular assembly into collagen intermediates[32] (Section 1.3.1). The second step occurs extracellularly and consists of the linear growth of the fibril intermediates; and the final step involves the lateral growth of fibril intermediates into mature fibrils of larger diameter[32] (Section 1.3.2). The tissue-specific structure and function of different tissue is due to independent regulation of each step, with different temporal and spatial expression of varying combinations of collagens, collagen-associated proteins, leucine-rich repeat family proteoglycans and glycoproteins[32] (Section 1.3.2.1-3). This review will focus mainly on type I collagen fibrillogenesis as it is the predominant collagen type in tendons and ligaments.

1.3.1 Collagen Synthesis and Transport

The initial stage of collagen synthesis is similar for all collagen types and occurs within the fibroblast (Figure 1.2). After gene transcription in the nucleus, collagen α-chain mRNA transcripts are transported to cystolic ribosomes, which are transported to the endoplasmic reticulum (ER) for translation into procollagen α-chains containing a terminal C-propetide region, a helical collagenous section and an N-propeptide terminal region[32]. The α-chains then undergo a series of post-translational modifications, such as hydroxylation of about 100 proline and 5-10 lysine residues to hydroxyproline and hydroxylysine respectively, and glycosylation of some hydroxylysine residues with galactose or glucosylgalactose[33]. In fibril-forming collagens, the formation of the procollagen triple helix from the three α-chains is initiated at the C-propeptide nucleation point by a set of intra- and inter-chain disulfide bonds, catalysed by protein disulfide isomerase (PDI) [34,35]. In fibrillar collagens, the helical domains of each α-chain helical domain are woven together to form a triple helix, beginning at the C-propeptide [17,18,32,34].

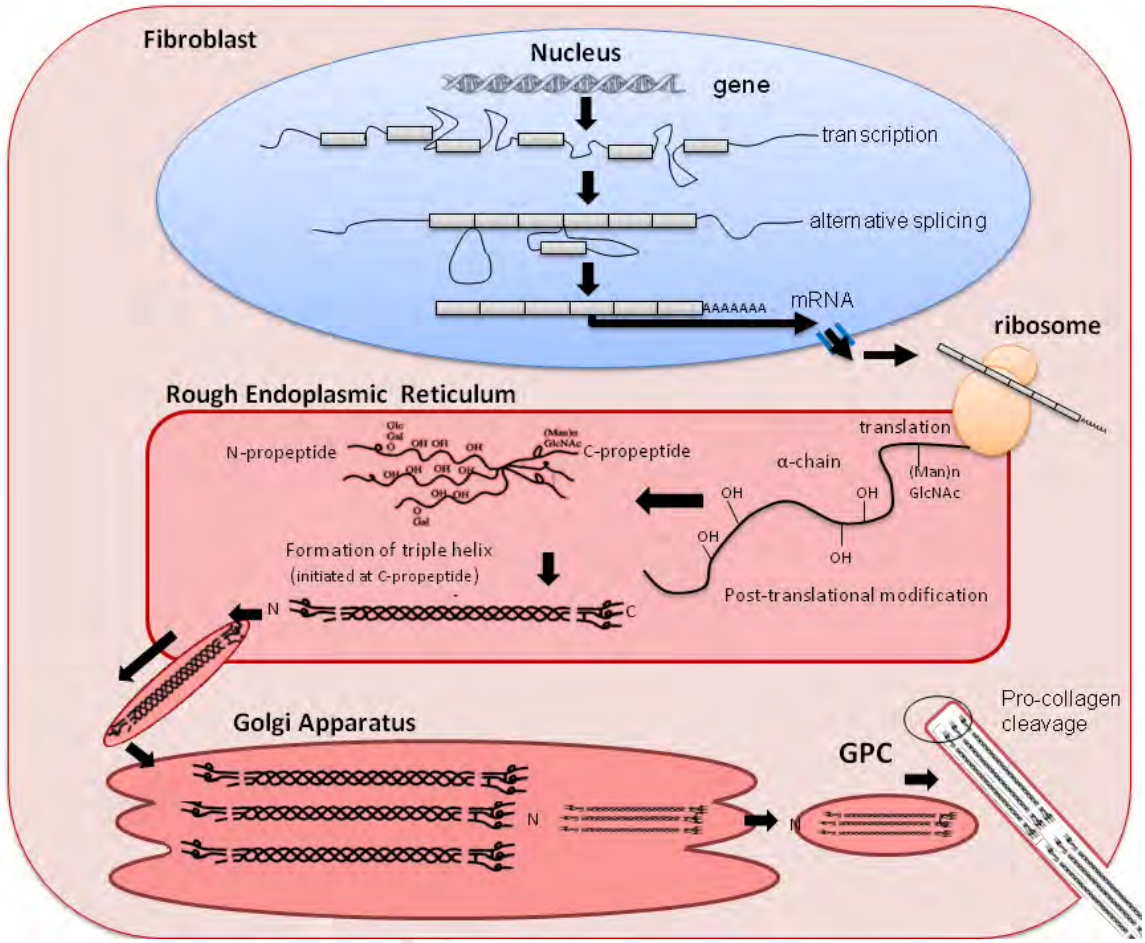


Figure 1.2 Collagen synthesis and transport inside the fibroblast. Collagen genes are transcribed in the nucleus and processed via splicing of introns to form mRNA. The mRNA is transported into the cytoplasm, and then to the ribosome, and attaches to the rough endoplasmic reticulum (RER), where it gets translated into the pro-collagen α-chain. The pro-collagen α-chains are post-translationally modified by hydroxylation and glycosylation. Three pro-collagen α-chains then wind together to form a triple helix, begging from the C-terminal propeptide and propagating along the three chains. The pro-collagen molecules are then transported to the Golgi-apparatus. In the Golgi apparatus, lateral aggregation of collagen molecules occurs, and eventually, individual vesicles bud off from the cisternae and form Golgi-to-plasma membrane carriers (GPCs). The GPCs transport the laterally aggregated pro-collagen molecules to deep fibril-forming channels that are continuous with the extracellular environment, but which extend deep into the cell. Further processing of the C- and N-terminal regions occur extracellularly.

The FACIT collagens differ in their NC1 domains, therefore a cysteine-containing motif in the COL1 domain plays a role in trimer assembly rather than the C-propeptide[36]. The triple helix is stabilized by inter-chain hydrogen interactions between amino acids which repeat in the sequence 'Gly- X- Y', in which glycine occurs every third residue, X is usually proline and Y is usually 4-hydroxyproline[17] .

The majority of collagens form homotrimers that consist of three identical α-chains; however they may also form heterotrimers, which contain more than one type of α-chain in different ratios and combinations[17,18]. Type I collagen, the most abundant collagen type in tendons and ligaments is formed by two proα1(I) chains and one proα2(I) chain[31]. After the triple-helical procollagen is folded, the molecules are transported from the ER to the Golgi apparatus, where the organized lateral aggregation of procollagen into bundles occurs[33]. As the procollagens are transported through the Golgi, their long length results in distensions of the Golgi cisternae[33]. The Golgi cisternae slowly shrink due to vesicle expulsion of lipids, resulting in a narrowing of the Golgi cisternae lumen, and a localized increase in procollagen concentration, which in turns results in lateral aggregation of procollagen into tightly bundled parallel arrays[32]. Eventually, the ends of the cisternae bud off into secretory vacuoles called Golgi-to-plasma membrane carriers (GPCs)[32,37]. The GPCs fuse to an extracellular compartment that extends from deep within the cell to the ECM[33].

The final posttranslational modification of the procollagen molecule is the cleavage of the C and N-terminal propeptides (Figure 1.3). There is some debate as to whether this cleavage occurs within the GPC (Kadler Model[33,37]), in the extracellular compartment (Birk Model[33,38]), or both. The C-terminal propeptide is removed by zinc metalloproteinases such as bone morphogenetic protein (BMP-1), mammalian tolloid (mTLD) and tolloid-like (TLL-1), while the N-propeptide is cleaved by several members of the "A Disintegrin and Metalloproteinase with Thrombospondin motifs" (ADAMTS) family, namely ADAMTS-2, ADAMTS-3 and ADAMTS-14[33]. Removal of the C and N-terminal propeptides causes a 10 000-fold decrease in the solubility of the triple helical region, resulting in a spontaneous self-assembly of the molecules into a rod-like collagen fibril monomers ~300 nm in length and 1.5 nm in diameter, called tropocollagen[32]. Tropocollagen can spontaneously form string-like aggregates called fibrils that gave a characteristic 67 nm banding pattern caused by the quarter-staggered alignment of collagen fibril monomers[31]. The tropocollagen molecules are covalently bonded by cross-links between lysine and hydroxylysine residues, a reaction that is catalyzed by the lysyl oxidase enzyme[11]. These cross-links can form intramolecularly between α chains of the same tropocollagen molecule, or intermolecularly between adjacent tropocollagen molecules[11]. Tropocollagen molecules aggregate extracellularly to form collagen fibrils intermediates ~10-30 μm in length[32].

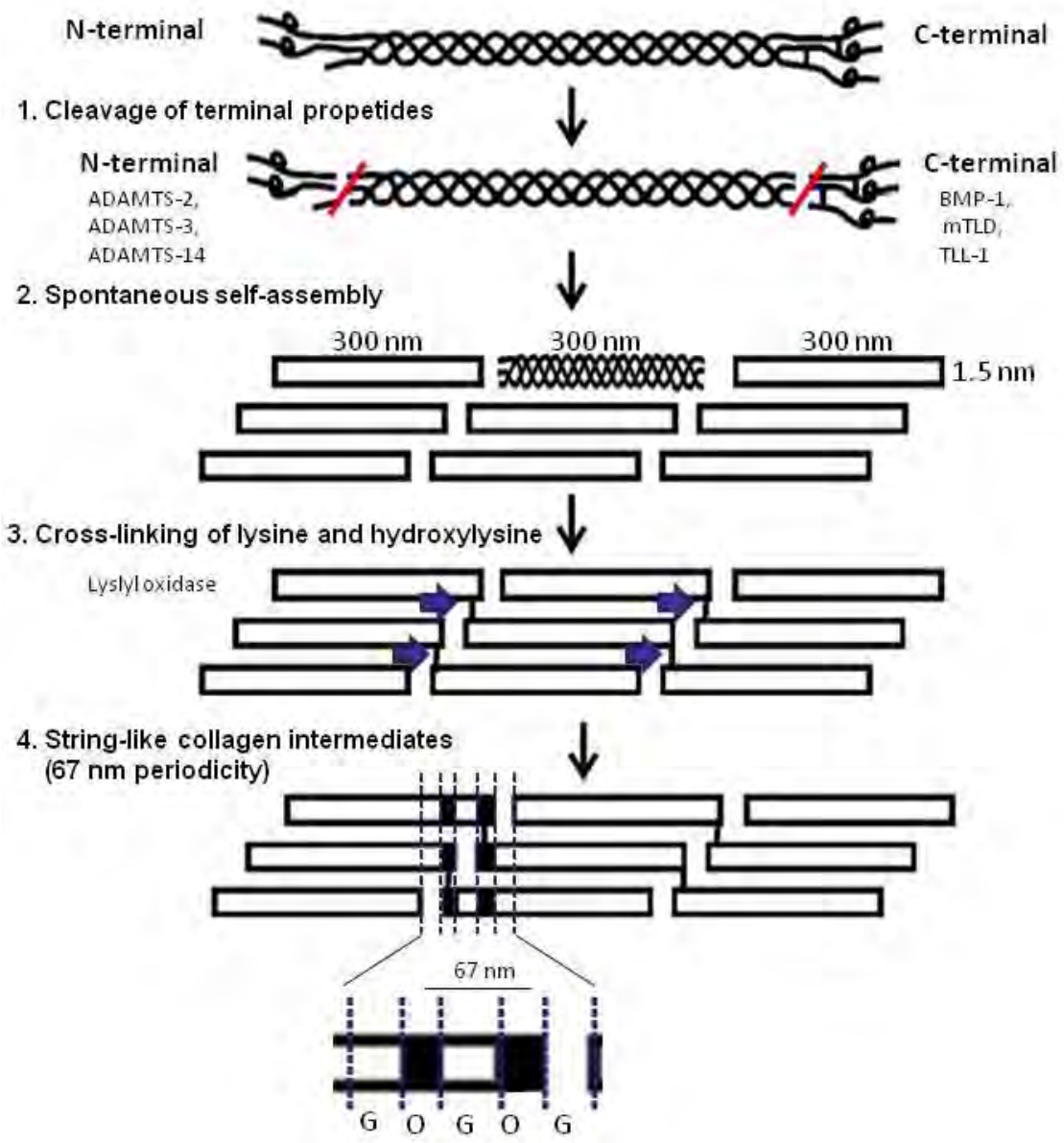


Figure 1.3 Schematic diagram showing the final post-translation modifications of pro-collagen molecules, and the spontaneous assembly, and cross-linking that result in the characteristic 67 nm banding pattern. **1)** The C-terminal propeptides are removed by BMP-1, mTLD and TLL-1, and the N-terminal propeptides are removed by ADAMTS-2, ADAMTS-3 and ADAMTS-14. **2)** The resulting tropocollagen molecules experience a 10 000 fold decrease in solubility and spontaneously from fibril monomers 300 nm in length and 1.5 nm in diameter (tropocollagen). **3)** Adjacent molecules get bonded covalently by the lysyl oxidase enzyme at selected lysine and hydroxylysine residues. **4)** Collagen fibril intermediates ~10-30 µm in length form that have the characteristic 67 nm banding pattern, which occurs as a result of the three-quarter staggered arrangement of adjacent tropocollagen molecules. G, gap; O, overlap.

1.3.2 Collagen Fibril Growth

The second and third steps of linear and lateral growth occur extracellularly, and involve the fusion of collagen intermediates into collagen fibrils of increasing length and diameter[32]. The structure and strength of the tissue is determined by intra- and inter-molecular cross-links between fibrillar collagens and collagen-associated proteins, as well as other ECM molecules such as the leucine-rich repeat family of proteoglycans and glycoproteins[32]. The expression of these molecules is extremely complex, and the temporal and spatial expression of molecules in the correct sequence and combination is required for the proper tissue specific structure and biomechanical properties to develop[32]. The role of many of these molecules as regulators in fibrillogenesis was discovered by knock-out studies in mice[39]. The next section reviews some of the key regulators of linear and lateral fibril growth, with a discussion of the mechanism of regulation, temporal expression of the molecule and the phenotype of the knock-out organism where available.

1.3.2.1 Fibrillar Collagens

1.3.2.1.1 Type III Collagen

Type III collagen often co-localizes with type I collagen and can form heterotypic fibrils with type I collagen[32]. The ratio of type I to type III collagen is thought to regulate the size and uniformity of type I fibrils, with higher type III collagen content resulting in smaller diameter fibrils[32]. This regulation occurs due to partial processing of the transcript in which the N-propeptide is retained and interferes with the incorporation of type I monomers[40]. The expression of type III collagen differs temporally and spatially during development. Early on in development the expression of type III collagen is high, however in mature fibrils, type III collagen is usually confined only to the endotenon, having been removed by protease and collagenase activity[32,33].

1.3.2.1.2 Type V Collagen

Type V collagen is a minor fibrillar collagen that acts as a nucleator for type I collagen during fibril assembly, and regulates fibril diameter in mature tendons via retention of the non-collagenous N-propeptide[23,24] (Figure 1.4). The triple helical domain of type V collagen occurs in the interior of the fibril, while the NH_2-domain projects outwards[24] and interacts sterically or electrostatically with other collagen molecules or with leucine-rich proteoglycans[25]. Upon reaching a critical density, the initiation of new fibrils is favoured over the continued growth of existing fibrils[25]. Type V collagen is essential for life, as *COL5A1 -/-* mice died at embryonic day 10, and had a virtual lack of collagen fibril formation[21]. Long-term dermal fibroblast cultures deficient in type V collagen had normal rates of

collagen synthesis, yet deposited less than half the amount of hydroxyproline as controls, showing a role for type V collagen in fibril nucleation[19]. Increasing the amount of type V collagen results in a progressive decrease in the mean fibril diameter *in vitro*[23], in cultured fibroblasts[25,19] and in mice[21]. In addition, different tissues expressing different levels of type V collagen have different mean diameter fibrils[41]. For example, the chick cornea has a high proportion of type V collagen (20%) and has a very narrow range of small diameter fibrils in comparison to tissues containing predominantly type I collagen[22]. The function of type V collagen in musculoskeletal soft tissue injuries will be described in more detail in the next chapter (Chapter 2) of this dissertation.

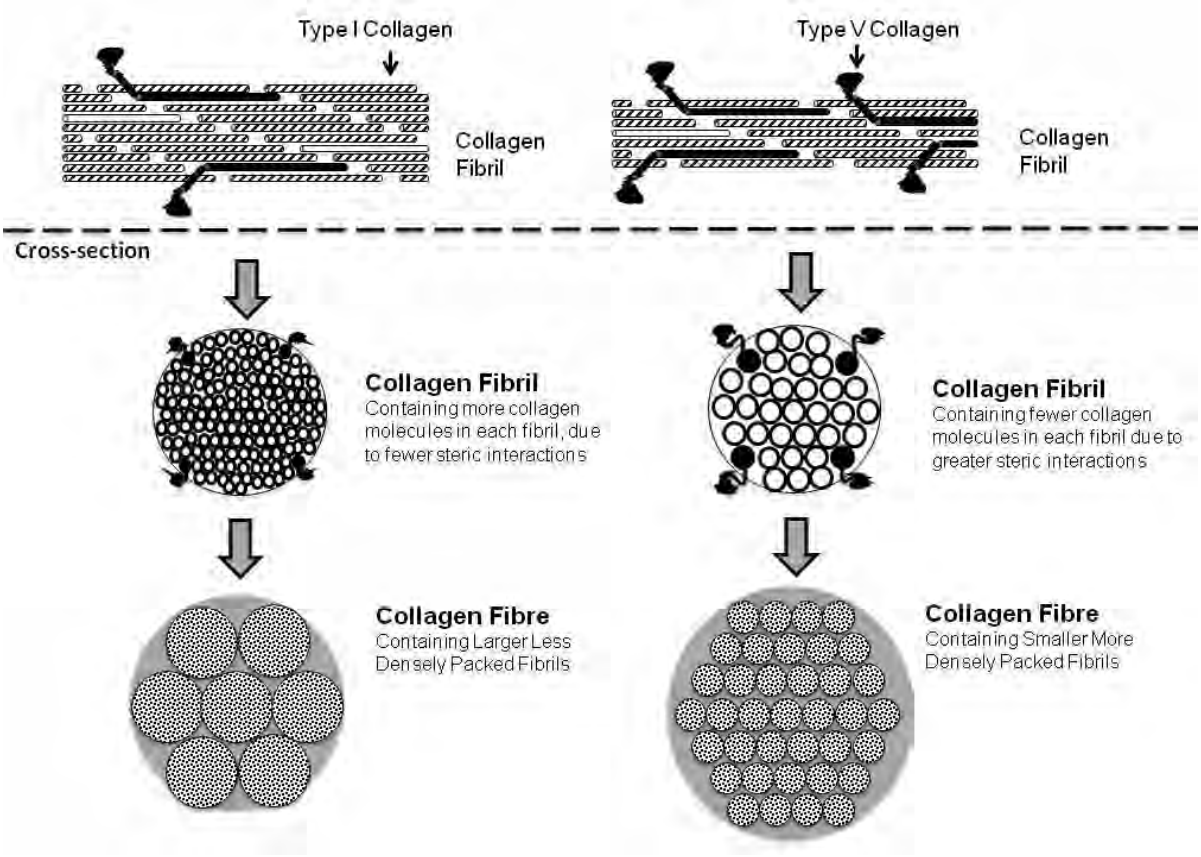


Figure 1.4 Model for type V collagen regulation of fibril diameter. The negatively charged amino-terminal domains of type V collagen project from the fibril surface, and when sufficient numbers have accumulated, they block further accretion of collagen monomers and thereby limit growth in diameter **Left**: Collagen fibrils possessing low levels of type V collagen incorporate a large number of type I collagen molecules and have large diameters. This results in collagen fibres made up of fewer, larger diameter fibrils, that are less densely packed. **Right**: Fibrils with a high percentage of type V collagen, prevent accretion of type I collagen, and therefore have small diameter fibrils. Therefore a greater number of smaller diameter collagen fibrils get packed closely together in collagen fibres.

1.3.2.1.3 Type XI Collagen

Type XI collagen shares structural and functional homology with type V collagen[28], in that it has a collagenous domain that is situated in the core of collagen fibrils[29], and a protruding N-terminal domain that interacts sterically to control fibril diameter[42]. However unlike type V collagen, type XI collagen is expressed mainly in cartilage[43], which consists mostly of type II collagen[29]. In addition, type XI collagen is essential for the interaction of collagens with proteoglycans[28]. Although Type XI protein is predominantly expressed in cartilage [44], it is found in many other non-cartilaginous tissues[45–48]. The function of type XI collagen in tendon and ligament biology will be described in more detail in Chapters 3 and 4 of this dissertation.

1.3.2.2 FACIT Collagens

1.3.2.2.1 Type XII and Type XIV Collagen

Collagens in the FACIT subfamily of collagens consist of multiple collagenous domains, separated by non-collagenous regions[31,36]. Type XII and XIV collagens are large disulphide-linked homotrimeric molecules with three extended N-terminus globular domains that extend outward and serve as attachment points for other molecules[36]. There are two isoforms of type XII collagen; a long (XIIA) form and a short (XIIB) form[49]. The short XIIB-1 isoform is predominantly expressed in tendons and ligaments in response to mechanical loading[49]. During fibrillogenesis, type XII and type XIV collagen have different temporal and spatial expression and perform different functions[50]. Specifically, type XII collagen is involved in integrating and providing stability to the ECM, particularly interfacial regions and regions which experience high stress; while type XIV is expressed mainly during early development, where its down-regulation plays a role in the transition to linear growth in fibrillogenesis [50,51]. Type XII and XIV collagens also facilitate sliding of collagen fibrils along each other by diminishing interactions between adjacent fibrils via low affinity of the NC3 domain to ligands[52]. Procollagen I *N*-proteinase binds to collagen XIV, perhaps sequestering the molecule and preventing cleavage of the Collagen 1 N-propeptide. Presence of uncleaved N-propeptide on collagen I influences fibril morphology and limits collagen aggregation[53]. Type XII collagen has been shown to bind decorin and fibromodulin[54]. And type XIV collagen has been shown to bind to decorin[55] and type VI collagen[56].

1.3.2.3 Small Leucine-Rich Proteoglycans

Several members of the Small Leucine-Rich Proteoglycans (SLRPs), such as decorin, biglycan, lumican and fibromodulin, associate with collagen fibrils in tendon and regulate fibrillogenesis[39]. Proteoglycans are a diverse group of molecules that consist of a core protein and glycosaminoglylcan

side chains (GAGs)[57]. GAGS are linear polysaccharides that bind water and create a hydrated gel that resists compressive forces[57]. Proteoglycan bridges between collagen fibrils affect tendon and ligament biomechanical properties by maintaining hydration and increasing the ability of the tissue to resist and transmit tensile stresses[58].

1.3.2.3.1 Decorin and Biglycan

Decorin and Biglycan are members of the Class I family of proteoglycans[57]. Decorin is a low molecular weight chondroitin sulfate proteoglycan, widely expressed in tendon, that regulates fibril diameter by regulating the lateral fusion of fibrils during fibrillogenesis[59]. Decorin-deficient mice have uncontrolled lateral fusion of collagen fibrils, resulting in large diameter fibrils with irregular outlines[59,60]. In addition to altered fibril morphology, decorin-deficient mice have skin fragility[60] and reduced tendon tensile strength and elasticity[59]. Biglycan has two chondroitin or dermatan sulphate side chains attached to its core protein and is found in human tendon[59], where it tends to colocalize with type VI collagen[61]. Biglycan-deficient mice have smaller diameter collagen fibrils, less densely packed that wild-type fibrils[62], with an irregular cross-section that results in mechanical compromise of tendon[63]. There appears to be compensation by these proteins, whereby biglycan is up-regulated in decorin-deficient mice[59].

1.3.2.3.2 Fibromodulin and Lumican

Fibromodulin, a keratin sulphate proteoglycan, and lumican, a glycoprotein, are members of the class II family of proteogylcans[57]. Both are expressed in tendon where they compete for the same binding site on type I collagen fibrils. [64]. Fibromodulin aids in the maturation of collagen fibrils, from thin diameter immature fibrils to larger diameter thicker fibrils[65–67]. As a result fibromodulin-deficient tendons have small diameter fibres[65–67] with reduced tensile strength[65,66]. However lumican deficient mice have larger diameter fibrils in tail tendon[68]. Lumican and fibromodulin are both expressed early in development; however, although lumican-deficient mice have abnormal fibril phenotypes early in development, they resemble the wild-type by maturation, suggesting an early role for lumican in fibrillogenesis, and a role in maturation for fibromodulin[66]. The regulation of lumican and fibromodulin concentration is orchestrated; whereby lumican expression is increased in the absence of fibromodulin[67]. In the absence of fibromodulin there is a dosage-dependent effect for lumican on tendon strength [65].

This section has reviewed the complex process of fibrillogenesis and discussed several of the key proteins involved in the regulation of tendon and ligament architecture. An aberration in the normal expression or function of these proteins causes changes in collagen architecture that has severe

consequences for the mechanical integrity of connective tissues. The relationship between collagen fibril structure and biomechanical properties will be discussed in the following section (Section 1.4).

1.4 Fibril Diameter and Tendon and Ligament Biomechanical Properties

In a tendon or ligament, the total cross-sectional area is divided into multiple parallel threads[69] (Figure 1.1 and 1.4). This imparts tendons and ligaments with great resilience and elasticity[69]. The increased resilience is a consequence of the hierarchical structure, as tears are diverted and stopped at the surface of each fibril rather than propagating across a single larger fibre[69].

The two physical properties most often used to characterize the biomechanical properties of a tissue are the ultimate tensile stress and the Young modulus[16]. The tensile stress is the maximum force (N) per unit area that a tissue can withstand before rupture[10,70,71]. Strain is the degree to which the tissue deforms under a given stress, and is usually expressed as a percentage of the original length (%). Young's modulus refers to the slope of the stress-strain curve in the elastic region, and gives a measure of elasticity, as it describes the force per area (stress) that is required to cause a given change in length (strain) (Figure 1.5) [10,70].

The collagen bundles along the length of a tendon or ligament exhibit waviness referred to as "crimp" (Section 1.2; Figure 1.2). When a sufficient load is applied to a tendon or ligament, the crimp straightens and elongates to a point of stiffness or restraint [8,10,72] (Figure 1.5). This is referred to the "toe region" in a stress strain curve[72]. If stress is applied beyond this point, stretching of the helical region of the collagen molecules occurs, resulting in further elongation (elastic region)[10,71]. Should the applied load continue to increase, collagen fibre bundles begin to fail, resulting in the accumulation of micro-tears[15]. The accumulation of micro-tears eventually results in an overuse injury, with its accompanying symptoms of pain and oedema[15]. Eventually, as micro-tears accumulate or if the load continues to increase, the entire ligament or tendon may rupture.

The ultimate tensile strength of a tissue depends on the substances of which they are made, and their structure[69]. Generally, a tissue consisting of large diameter fibrils tends to have greater ultimate tensile strength[73]. This is presumably due to an increase in the intrafibrillar covalent cross-links between collagen molecules[73]. In addition, the shearing stress exerted at the interface of the fibril with the extracellular matrix by the application of a tensile load is lower in larger diameter fibrils due to the greater volume to surface area ratio; therefore the larger the fibril diameter the lower the shearing stress[69].

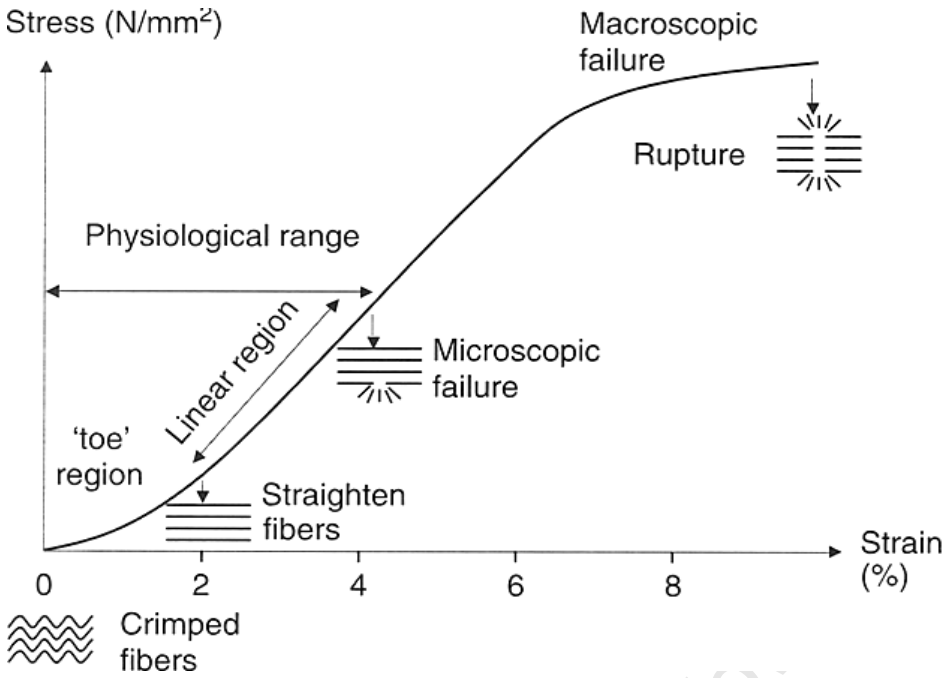


Figure 1.5 Mechanism of internal deformation of tendon or ligament. When strain is applied to a tendon, there is initially a straightening out of the crimp, resulting in the 'toe region' of the stress–strain curve. Strain within the physiological range causes elastic deformation of the tendon. Higher levels of strain result in partial or complete rupture of fibrils. Image modified from Wang, (2006)[72] in Jose *et al*., (2011)[13].

The Young Modulus (elasticity) may be affected by diameter, because smaller fibril diameter increases the surface area per unit mass of the fibrils, thus enhancing the probability of interfibrillar non-covalent cross-links between the collagen fibrils and components of the matrix[71,73]. For this reason, smaller diameter fibrils may reduce flexibility and protect against non-recoverable creep[71].

The structural properties of ligament and tendon connective tissue therefore directly affect the biomechanical properties of a tissue. Specifically, tensile stress, which is the maximum load that the tissue can tolerate before microscopic-tears and macroscopic rupture, is dependent on fibril diameter[73]. In addition, the elastic potential of a tissue is dependent on the number of interfibrillar cross-links, and elasticity is therefore maximised when fibril diameter is smaller and the surface to volume to ratio between the collagen fibril and the adjacent ECM is greater[69,71].

The accumulation of micro-tears in response to excessive or repetitive stress is one of the proposed mechanism of injury in AT[74]. In contrast, ACL ruptures are acute events that occur due to excessive external or internal forces that exceed the tensile strength of the tissue in a discrete incident[75]. The following section will examine some of the extrinsic and intrinsic factors that contribute to these conditions.

1.5 Aetiology of Musculoskeletal Soft Tissue Injuries

Musculoskeletal injuries are multi-factorial disorders which occur due to the complex interaction of intrinsic and extrinsic risk factors (Figure 1.6)[2,75,76]. Intrinsic risk factors, which are usually biological in nature, determine the extent to which an individual is predisposed to an injury[2]. The predisposed athlete then becomes susceptible to injury following exposure to extrinsic factors, which are often environmental or behavioural in nature[2]. However, an inciting event is needed for an injury to occur. Although the aetiology of AT[76–79,74,14] and ACL rupture[75,80] is unique, there are several categories of intrinsic risk factors in common, including: biological age, gender, history of previous injury and anatomical features. Examples of categories of extrinsic risk factors in common include the duration, nature and frequency of physical activity, weather conditions, training surfaces, footwear, nutrition and certain medications[2]. The relative contribution of intrinsic and extrinsic risk factors is likely to differ between overuse injuries such as AT, and acute injuries such as ACL rupture[2].

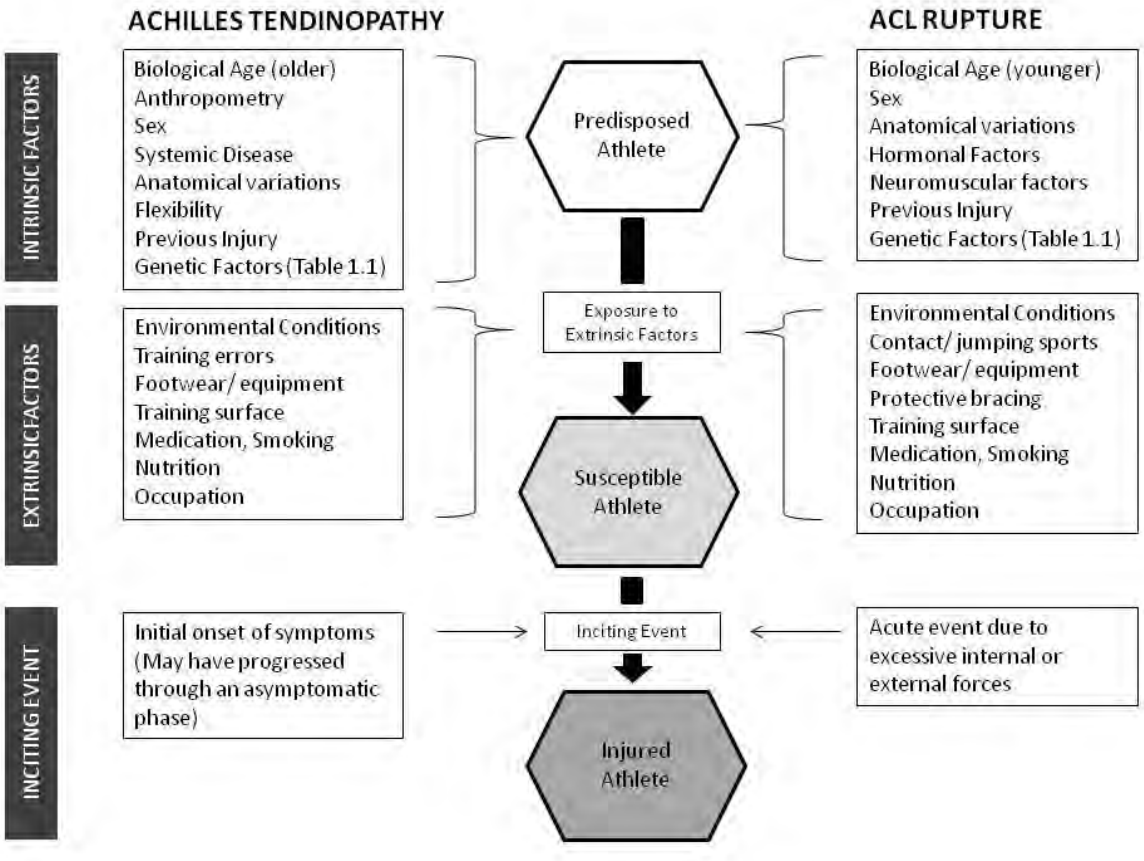


Figure 1.6 A schematic diagram illustrating the complex relationship between intrinsic risk factors, extrinsic risk factors and a specific inciting event in the causation of musculoskeletal soft tissue injuries. Intrinsic risk factors predispose an individual to injury; however extrinsic risk factors, such as behaviour and environment contribute to an individual's susceptibility to becoming injured, and an inciting event is needed to cause an individual to actually become injured. [Adapted from the original model proposed by Meeuwisse[2]].

1.5.1 Genetic Risk Factors for Musculoskeletal Injuries

Many of the intrinsic risk factors such as sex[81], anatomical variations[82,83], flexibility[84,85], systemic disease[86] and hormonal factors[87,88] are at least in part genetically determined[5,89]. In addition, evidence for a genetic predisposition to musculoskeletal injury is provided by familial studies that clearly show that a family history of injury is associated with increased risk[90–92]. Flynn et al showed that participants with an ACL rupture were twice as likely to have a relative with an ACL rupture compared to uninjured individuals[90] and a more recent study on 350 families showed that 11.3% of patients with an ACL rupture had a family history of ACL injury[91]. Similarly, a study of 310 participants revealed that a family history of Achilles tendinopathy increased the risk of developing Achilles tendinopathy five-fold[93]. Discovering which genes are involved in musculoskeletal soft tissue injuries will be important not only to identify at-risk individuals, but will also aid in understanding the molecular mechanisms underlying soft tissue injuries[3,4].

Previous studies on AT and ACL rupture have used a candidate gene approach to investigate associations with certain likely candidate genes and injury risk rather than a whole genome approach, and a number of genes have been associated with AT and ACL rupture to date (Table 1.1). Usually these candidate genes are chosen based on evidence for their biological function. However, musculoskeletal soft tissue injuries are likely to be polygenic in nature, and occur as a consequence of variants at several genetic loci that act alone or in combination to increase susceptibility[94].Therefore, as no single genetic association will adequately capture injury risk, there is an increasing tendency to examine the genetic profile of an individual[95] which take into account genetic variants at a number of different loci[95], often involved in vastly different processes such as cell signalling[96], ECM degradation[97,98], apoptosis[99] or in the case of this dissertation; fibrillogenesis.

Table 1.1 A summary of previous associations between genes and chronic musculoskeletal soft tissue injuries

Gene	Function	Polymorphism	Associated injury	Reference
TNC	Encodes glycoprotein TNC, a structural constituent of tendons and ligaments	GT repeat in intron 17	TEN	Mokone *et al.* (2005)[100]
COL5A1	Encodes the pro-α1 polypeptide of type V collagen	rs12722	TEN	Mokone *et al.* (2006)[101]; September *et al* (2008)[102]
			ACL ruptures in females	**Posthumus *et al.* (2009)**[103]
		rs71746744 rs1134170 rs16399	TEN	Abrahams *et al*., (2013)[104]
MIR608	Encodes the microRNA Hsa-miR-608	rs4919510	TEN	Abrahams *et al*., (2013)[104]
COL1A1	Encodes the pro-α1 polypeptide of type I collagen	rs1800012	TEN ACL ruptures Shoulder dislocations	**Khoschnau *et al*. (2008)**[105]; **Posthumus *et al.* (2009)**[106] **Ficek *et al* (2012)**[107]
COL12A1	Encodes the pro-α1 polypeptide of type XII collagen	rs970547	TEN ACL ruptures in females	**Posthumus *et al*. (2010)**[108]
GDF5	Growth/differentiation factor 5	rs143383	TEN	Posthumus *et al.* (2010)[109]
MMPs	Important in ligament homeostasis and integrity	*MMP3* rs679620 rs591058 rs650108	TEN	Raleigh *et al*. (2009)[98];
		MMP12 rs2276109	ACL ruptures	**Posthumus *et al*. (2012)**[97]
IL-1β*	Encodes the IL-1β protein which causes the upregulation of COX-2 and PGE_2	rs16944 rs1143627	TEN	September *et al*. (2011)[96]
IL-1RN*	Encodes IL-1ra, the receptor agonist of IL-1β	rs2234663	TEN	September *et al*. (2011)[96]
IL-6*	Encodes the IL-6 protein which plays a role in cell apoptosis	rs1800795	TEN	September *et al*. (2011)[96]
CASP8	Encodes a cysteine-aspartic protease which mediates cell apoptosis	rs3834126 rs1045485	TEN	Nell *et al.* (2012)[99]

Abbreviations: ACL, anterior cruciate ligament; CASP8, caspase 8; COX-2, cyclooxygenase-2; IL-1β, interleukin 1β; IL-1RN, interleukin 1 receptor agonist; IL-6, interleukin 6; MMP, matrix metalloproteinase; NO, nitric oxide; NOS 2; nitric oxide synthase 2; PGE_2, prostaglandin E_2; PTGER4, prostaglandin E receptor 4 (EP4); TEN, Achilles tendinopathy; TNC, Tenascin C.

* The *IL-1β*, *IL-1RN* and *IL-6* genes were not associated independently with TEN, but significant associations were found when analysed in combination with each other and *COL5A1*.

References in bold refer to studies that investigated ACL rupture. All other references refer to Achilles tendinopathy.

1.6 Scope of the Dissertation

Based on the previously discussed role of fibril diameter and density on determining the tensile strength and elasticity of ligaments and tendons (Section 1.4), the focus for this dissertation will be on genetic variants in genes encoding the minor fibrillar collagens, type V and type XI collagen (Section 1.3.2.1), because of their role in the regulation of collagen fibril diameter.

1.6.1 The *COL5A1* and *MIR608* Genes

Type V collagen is a minor fibrillar collagen present in relatively small quantities in tendons and ligaments[8,9]. Type V collagen consists of varying combinations of three α-chains, α1(V), α2(V) and α3(V), encoded by the *COL5A1*, *COL5A2* and *COL5A3* genes respectively[110]. There are several different isoforms which differ in α chain composition and ratio[27]. Type V collagen occasionally forms $\alpha1(V)_3$ homotrimers or α1(V)α2(V)α3(V) heterotrimers, however the most common isoform is a heterotrimer consisting of two α1(V) chains and one α2(V) chain[41]. The $\alpha1(V)_2\alpha2(V)$ isoform is capable of forming heterotypic fibrils with type I collagen, where it is thought to aid in collagen nucleation and the initiation of fibril assembly[19–21] (Section 1.3.2), and regulate fibril diameter[22–25,27] (Figure 1.4).

Mutations in either the *COL5A1* or *COL5A2* genes, which encode the α1 or α2 chain of type V collagen respectively, result in the heritable connective tissue disease Ehlers-Danlos Syndrome (EDS)[111,112]. The mutations cause a haplo-insufficiency of type V collagen, the consequence of which leads to the symptoms of the syndrome, which include: hyperextensible skin, hypermobile joints, abnormal scarring and other systemic involvement[111,112].

The *COL5A1* gene on chromosome 9q34[113] spans 203.07 kb and comprises a coding sequence of 8468 bps spread over 66 exons[114] that encode a 1838 amino acid protein. A number of studies have shown an association between injury and performance phenotypes and a common C to T single nucleotide polymorphism (SNP rs12722 or *Bst*UI RFLP) within the 3'-UTR of the *COL5A1* gene[101–103,115–117]. Studies on a Caucasian population showed that individuals homozygous for the C variant of this SNP were less likely to suffer from AT[101,102]. Similarly, a study on ACL injuries showed that CC homozygote females were less likely to suffer ACL ruptures[103]. In a subsequent study investigating the effect of the *COL5A1* rs12722 SNP on range of motion (ROM), the CC genotype was found to protect against an age-related decline in flexibility [116]. Additionally, a study examining *COL5A1* and endurance performance revealed that athletes homozygous for the T variant of the *COL5A1* rs12722 SNP completed the running component of the Ironman triathlon faster than those with a CT or CC variant[117]. The improved running performance was hypothesised to be due to altered biomechanical

properties, which resulted in improved running economy[117]. Taken together, these studies suggest that the CC genotype is protective against musculoskeletal injury and results in decreased flexibility, whereas the TT genotype predisposes to injury, decreases flexibility and improves running performance. This hypothesis has recently been reviewed[110].

To investigate the functional significance of these associations, the 3'- untranslated region (UTR) of the *COL5A1* gene was cloned from tendinopathic and control individuals and compared[118]. Two major functional forms containing 7 tightly linked polymorphisms were identified and found to have significant differences in mRNA stability[118]. The C form, which contains the C-allele of rs12722 and corresponds to the wild-type sequence, was identified in most of the clones generated from asymptomatic controls, while the T form contained the T-allele of rs12722 and was the predominant form identified in the AT patients[118]. Recently, the remainder of these variants have been investigated, and the AGGG/AGGG, -/- ATCT and TT alleles of rs71746744, rs16399 and rs1134170 respectively, which correspond to the T-form cloned from tendinopathic individuals, was associated with increased risk of AT[104]. Using *in silico* statistical folding algorithms, the variant combinations were found to alter mRNA secondary structure which may affect the post-transcriptional regulation of mRNA by altering a miRNA binding site[104]. Changes in mRNA stability can alter mRNA levels and thus affect the rate of translation and protein synthesis[119,120]. This may alter the concentration of type V collagen, with potential consequences for the structure and biomechanical properties of tendons and ligaments[110]. Although the *COL5A1* variants have been investigated for an association with AT, additional variants in the *COL5A1* 3'-UTR remain to be investigated in ACL rupture. Therefore the first aim of this dissertation was to investigate the *COL5A1* rs71746744 and rs1134170 gene variants for an association with ACL rupture (Chapter 2).

The *COL5A1* 3'-UTR contains several putative *cis*-acting elements including a functional Hsa-miR-608 binding site [102,104]. Two forms of the mature Hsa-miR-608, which are produced from the polymorphic (SNP rs4919510, C/G) *MIR608* gene on chromosome 10q24, can potentially bind this miRNA binding site[104]. No studies have analysed the functional significance of the rs4919510 (C/G) SNP on Hsa-miR-608 binding to the *COL5A1* 3'-UTR; however Hsa-miR-608 was show to bind preferentially to the A allele of the rs3196378 (C/A) polymorphism in the *COL5A1* 3'-UTR[118]. Although the effect of the rs4919510 G allele on *COL5A1* 3'-UTR binding is unknown the CC genotype of *MIR608* rs4919510 was recently shown to be associated with increased risk of AT[104]. Therefore an additional aim of this dissertation was to investigate the *MIR608* rs4919510 gene variants for an association with ACL rupture (Chapter 2).

1.6.2 The *COL11A1* and *COL11A2* Genes

Type XI collagen is a quantitatively minor fibrillar collagen comprising three polypeptide chains, α1(XI), α2(XI) and α1(II), encoded by the *COL11A1*, *COL11A2* and *COL2A1* genes respectively. Type XI collagen plays a similar role to type V collagen tendons and ligaments, in that it regulates collagen fibril diameter in cartilage[43], which mostly consists of type II collagen[29]. In addition type XI collagen is essential for the interaction of collagens with proteoglycans[28]. Type XI protein is predominantly expressed in cartilage[44], but is also found in the ocular vitreous, nucleus pulposus of the intevertebral disc and the inner ear, as well as many other non-cartilaginous tissues[45,47,48].

Mutations in *COL11A1, COL11A2* or *COL2A1* cause diseases such as Stickler syndrome[121], Marshall syndrome[122], Otospondylomegaepiphyseal dysplasia (OSMED)[123] and Weissenbacher–Zweymuller syndrome (WZS)[124]. These syndromes are known as type XI collagenopathies[125], and they are phenotypically similar disorders that result in facial anomalies, cleft palate, hearing defects and a spectrum of epiphyseal dysplasias with wide metaphyses and spondylar abnormalities with occasional ocular changes[125]. Stickler syndrome causes joint hypermobility, similar to that seen in Ehlers-Danlos syndrome, a disorder of type V collagen production, however in Stickler syndrome the hyperextensibility occurs only in childhood and decreases in adulthood, often to be replaced by a degenerative arthropathy[121].

The *COL11A1* gene is located on the reverse strand of chromosome 1p21.1 and spans 232.03 Kb. The mRNA transcript is 7286 bp in length, made up of 67 exons that encode an 1806 amino acid protein. A number of multifactorial conditions or syndromes have been investigated for associations with *COL11A1*, such as early-onset arthritis[126], osteoarthritis[127], lumbar disc pathologies[128–130], cleft palate[131,132] and cancer[133]. Two common functional polymorphisms were identified by these previous studies. The rs3753841 (T/C) polymorphism is a T>C single nucleotide substitution that results in an amino acid change from a leucine to a proline at amino acid 1323, and the C- allele has been associated with lumbar disc herniation (LDH)[129]. The rs1676486 (C/T) polymorphism is a C>T nucleotide change that results in a non-synonymous amino acid change from a proline to a serine at position 1535, and may have an effect on *COL11A1* mRNA stability[129], and the T-allele of this variant has been associated with LDH[129] and limbus vertebra in gymnasts[130].

The *COL11A2* gene is located on chromosome 6p21.32 and is 29.82 Kb in length. The mRNA transcript is 6209 bp long spread over 64 exons, and encodes a 1650 amino acid protein. The *COL11A2* gene has been investigated for associations with osteoarthritis[126,127], rheumatoid arthritis[134], osteochondrodysplasias [135], ossification of the posterior longitudinal ligament of the

spine (OPLL) [136–139], lumbar[128,140,141] and intevertebral [142,143] disc disease[144] and cleft palate [131,132]. These studies have identified a gene variant, rs1799907 (A/T), which is located 4 bp upstream of the start of exon 6, and appears to result in alternative splicing of the *COL11A2* gene[137]. The T allele of *COL11A2* rs1799907 was associated with risk of lumbar disc desiccation[128], lumbar spine stenosis[141] and OPLL[136–139].

Although the *COL11A1* and *COL11A2* genes have been associated with a number of musculoskeletal disorders that affect the spine and cartilage[128–130,134,136,137,141,142], these genes have never been investigated for an association with conditions that affect tendons and ligaments, despite the fact that these genes are known to be expressed in tendon in early development[48,145,146]. Therefore, an additional aim of this dissertation was to investigate the *COL11A1* rs3753841 and rs1676486 polymorphisms, and the *COL11A2* rs1799907 polymorphism for associations with ACL rupture (Chapter 3) and AT (Chapter 4).

1.6.3 Interactions Between Type V and Type XI Collagen

The structural and functional properties of type V and type XI collagen are closely related and their primary structures are highly conserved at the gene and protein levels[28]. Homology is greatest between the proα1(XI), proα2(XI) and proα1(V) chains, and this is particularly noticeable in the NC3 domain, which is composed of two sub-domains in these proteins, suggesting a strong topological homology between the N-terminal propeptide extensions of the proα1(XI), proα2(XI) and proα1(V) chains[28]. Both proteins are usually buried within the major collagens, however they contain cell adhesion and heparin binding sites that enable them to interact with other ECM molecules[28].

Proα1(XI) and proα2(V) chains have been shown to form heterotrimers of $[\alpha1(XI)]_2\alpha2(V)]$, *in vivo* [147,148] and *invitro*[28]. The α1(XI) chain is present in collagen V extracts from bone[149] and the human rhabdomyosarcoma cell line A204, which synthesizes type V collagen as its only fibrillar collagen, expresses *COL11A1* and forms $[\alpha1(XI)]_2\alpha2(V)]$ hybrid molecules[147]. Although it was generally thought that the tissue expression patterns of type V and type XI collagen do not overlap, it is becoming increasingly evident that several molecules are in fact heterotypic molecules of type V and XI[28].

Recently Wenstrup et al (2011)[146] provided phenotypic evidence that collagens V and XI may have shared or synergistic roles in developing tendon[146]. Type V and type XI collagen procollagen α-chains transcripts were expressed at similar levels in early development; however collagen V predominated in later developmental stages. When flexor digitorum longus (FDL) tendons from wild-type (WT) mice were compared to (i) *COL5A1 +/-,* (ii) *COL11A1 +/-,* (iii) *COL5A1 +/-, COL11A1+/-* (iv) *COL11A1-/-*

, and (v) *COL5A1 +/-, COL11A1-/-* knock-out mice, there was a decrease in fibril number in all of the genotypes. However while there was almost normal fibril appearance in the *COL5A1+/-* mice, the most severe phenotype was observed in the *COL5A1+/-, COL11A1-/-* compound heterozygotes. This suggests that there is biological redundancy between these two collagen types and that *COL11A1* expression may partly rescue low expression of *COL5A1* in haplo-insufficient mice.

Since there is evidence for an interaction between type V and type XI collagen, and variants in the *COL5A1* gene have previously been associated with both ACL rupture in females [103] and AT[101,102,104], the previously described *COL11A1* and *COL11A2* variants were investigated for interactions with the *COL5A1* rs71746744 variant in determining the risk of ACL rupture (Chapter 3) and Achilles tendinopathy (Chapter 4).

1.7 Dissertation Objectives

1.7.1 Hypothesis

This dissertation hypothesises that genetic variants in the *COL5A1*, *COL11A1* and *COL11A2* genes that encode for type V and XI collagens respectively, may alter injury risk because of their role in determining musculoskeletal soft tissue structural properties. Subtle changes in tissue architecture may alter the mechanical properties of the tissue such that they are less resilient to repeated mechanical loads[110]. In the case of AT, this may make the tissue more susceptible to accumulated micro-trauma due to decreased tensile strength of thin diameter fibrils, and in the case of acute ACL rupture, a less resilient tissue may be more likely to rupture when exposed to an excessive load[110]. The effect size of the genetic variants in the genes encoding for type V or type XI collagen may differ between the injuries, as the tissues differ slightly in structure, and the injuries differ in aetiology. These genes will be investigated together, because of a possible interaction effect.

1.7.2 Objectives

The objectives of this dissertation are:

i) To determine whether the rs71746744 and rs1134170 polymorphisms in the *COL5A1* gene and the rs4919510 polymorphism in the *MIR608* gene are associated with ACL rupture in a South African Caucasian population (Chapter 2).

ii) To determine whether the rs3753841 and rs1676486 polymorphisms in the *COL11A1* gene and the rs1799907 polymorphism in the *COL11A2* gene are associated with ACL rupture in a Caucasian South African population. A secondary aim will be to determine whether the *COL11A1* and *COL11A2* polymorphisms interact with the rs71746744 polymorphism in the *COL5A1* gene to modulate ACL rupture risk (Chapter 3).

iii) To determine whether the rs3753841 and rs1676486 variants in the *COL11A1* gene and the rs1799907 variant in the *COL11A2* gene are associated with AT in a Caucasian South African and Australian population. In addition, to investigate whether these variants interact with each other, or with the rs71746744 variant in *COL5A1,* to modulate the risk of developing AT (Chapter 4).

Chapter 2: Association of Variants in the *COL5A1* 3'-UTR and *MIR608* with Anterior Cruciate Ligament Rupture

2.1 Introduction

The ACL is a major intra-articular ligament of the knee and a common site of acute musculoskeletal injury in young, physically active individuals[75]. The ACL acts as a primary restraint against anterior tibial translation and a secondary restraint against rotary loads[7]. During dynamic movement the ACL works together with the posterior cruciate ligament (PCL) to stabilize the knee, and guide it through its normal range of motion[7,150–152]. There are three major mechanisms of ACL rupture: non-contact, direct contact, and indirect contact[153]. Non-contact injuries comprise approximately 70% of injuries, and occur due to excessive internal forces[75]. The remainder of the injuries are made up of direct contact mechanism of injury, in which there is a collision with a solid object at the injured knee, or indirect contact mechanism in which contact is made at a location other than the knee[154]. A further mechanism of injury is rare and unique to skiing activities, and is referred to as the "phantom foot" mechanism, in which the tail of the ski acts as a lever[155], generating excessive internal tibial torque while the knee is in flexion beyond 90° or in hyperextension[156–158].

Although the incidence of ACL ruptures in the South African population has not been documented, it is known that the incidence of ACL rupture in New Zealand is approximately 36 ruptures per 100 000 citizens per year[159] and Norway reports a similar incidence of 34 ruptures per 100 000 citizens per year[160]. The mean age at which ACL ruptures occur is 27 years[160], however the highest number of injuries occur within the 16-18 year age group[159]. Individuals who suffer from ACL ruptures are forced to decrease activities of daily life while injured[161] and may require reconstructive surgery and extensive rehabilitation in order to return to an active lifestyle. These injuries are costly[162] and can shorten a sporting career[163] and affect academic performance due to absenteeism[164]. In addition, ruptures to the ACL place individuals at a higher risk of other chronic knee pathologies such as osteoarthritis[165,166].

There is no single cause of ACL ruptures, as these conditions, like other musculoskeletal soft tissue injuries, have complex aetiologies[75,80]. Although several intrinsic and extrinsic risk factors have been identified, an individual's genetic profile has been proposed to be a strong intrinsic risk factor[3,167] (Chapter 1, Section 1.5.1). Genetic variants may influence injury risk via subtle changes in biological pathways involved in ligament biology or by altering the structural properties of the tissue. To date

functional variants within the *COL1A1* gene have been shown to be associated with ACL ruptures in Swedish[105], South African[106] and Polish[107] populations. In addition, a variant in the matrix metalloproteinase 12 gene (*MMP12)*, part of a cluster of four MMP genes on chromosome 11, has also been independently associated with protection against ACL rupture, as have haplotypes consisting of variants from each of *MMP10*, *MMP1*, *MMP3* and *MMP12*[97]. Although associated with AT[109], a functional variant within the *GDF5* gene is not associated with ACL ruptures[168].The total incidence of ACL ruptures is higher in males than females, however females have a higher risk of rupture[159,169,170]. Previous studies have shown that genetic factors that predispose to injury may differ between males and fermales[103,108]. In support of this, variants within the *COL5A1* [103] and *COL12A1*[108] genes have been shown to be associated with ACL ruptures in females.

With respect to *COL5A1*, previous studies have associated the common single nucleotide polymorphism (SNP) rs12722 (C/T) in its 3' UTR, with AT[101,102], ACL rupture in females[103], range of motion (ROM)[115,116] and endurance running performance[117]; suggesting a role for this region in multiple sports injury and exercise-related phenotypes[110]. The *COL5A1* gene encodes the α(I)V chain of type V collagen, a minor fibrillar collagen that is vital for nucleation of collagen fibrils and the regulation of collagen fibril diameter (Chapter 1, Section 1.3.2.1 and 1.6.1). Collagen fibril diameter and density directly affect the biomechanical properties of a tissue, such as tensile strength and elasticity (Chapter1, Section 1.4).

The 3'-UTR of eukaryotic genes contain elements such as poly(A) signals[171], protein binding sites[172] and microRNA (miRNA) binding sites, that play an important role in post-transcriptional regulation of the mRNA transcript[173,174]. MicroRNAs[175,176] are non-coding RNAs 18-25 bp in length, that bind to complementary sequences in the 3'-UTR of mRNA transcripts; resulting in them being silenced or targeted for degradation[175,176]. This post-transcriptional regulation of mRNA transcripts has been shown to play an important role in disease aetiology[177].

Collins and Posthumus have thus recently proposed that *COL5A1* 3-UTR variants may alter type V collagen concentration, which in turn alters collagen fibril diameter, consequently affecting the tensile strength and elasticity of the tissue[110]. Recently, two functional forms of the 2.5 kb *COL5A1* 3'-UTR, containing seven tightly linked polymorphisms; rs13948 (C/T), rs12722 (C/T), rs3196378 (C/A), rs71746744 (-/AGGG), rs16399 (ATCT/-), rs1134170 (A/T) and rs3128575 (T/C); were cloned

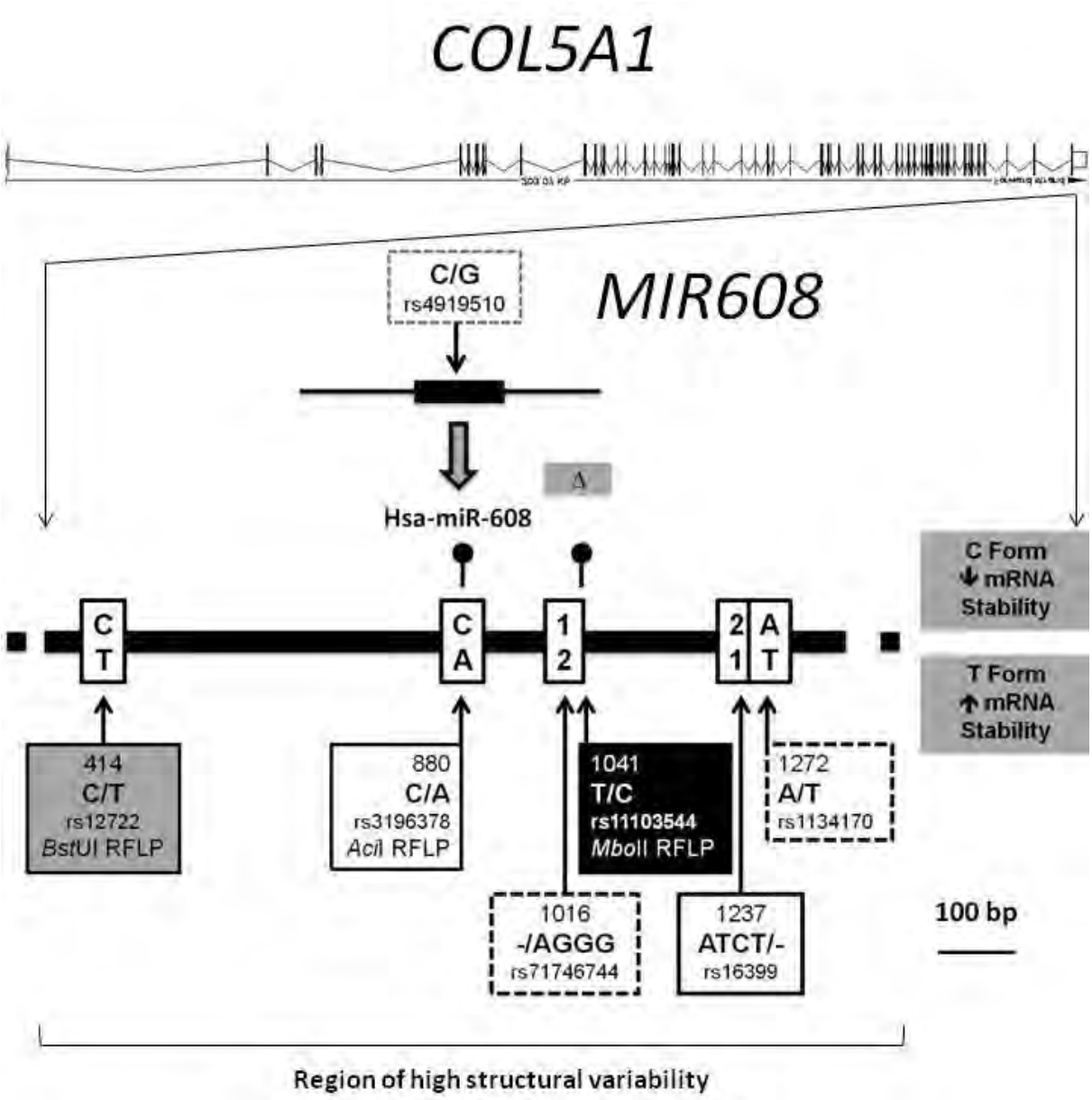


Figure 2.1 Schematic representation of the *COL5A1* gene on chromosome 9q34 and the *MIR608* gene on chromosome 10q24. The 66 exons in *COL5A1* are represented by vertical lines, and introns are represented by adjoining horizontal lines. The 2.5 kb *COL5A1* 3'-UTR is encoded by exon 66. A 858 bp region, containing rs12722 (grey box), as well as, rs3196378, rs71746744, rs16399 and rs1134170 (white boxes) of the 3'-UTR is shown. SNP rs12722 (grey box) has previously been associated with ACL ruptures in females and other exercise associated phenotypes[101–103,115–117,178]. The black box containing rs11103544 is within a putative miRNA binding site, which has not been associated with Achilles tendinopathy[102]. The *MIR608* gene contains SNP rs4919510 and encodes for Hsa-miR-608, which binds to a polymorphic binding site in the *COL5A1* 3'-UTR as indicated in the diagram. The rs71746744 and rs1134170 polymorphisms in *COL5A1* and the rs4919510 SNP in *MIR608* have been selected for investigation for an association with ACL rupture (dashed boxes). The accession numbers and/or RFLP associated with the polymorphism are indicated, together with the nucleotide changes. The nucleotide positions for the polymorphisms within the 3'-UTR are for the wild-type sequence (C-functional form). The two miRNA binding sites are indicated by a black solid circle and a line. The location of a previously described 57 bp deletion (Δ) containing rs71746744 and the second polymorphic miRNA binding site (rs11103544) is indicated[118]. This figure is modified from Abrahams, *et.al*., (2013)[104]
1 - $(AGGG)_1$ or $(ATCT)_1$, 2 - $(AGGG)_2$ or $(ATCT)_2$.

from tendinopathic and control individuals and found to have significant differences in mRNA stability[118] (Figure 2.1). The C form, identified in most of the clones generated from asymptomatic controls contains the polymorphisms in the sequence C-C-C-$(AGGG)_1$-$(ATCT)_2$-A-T and is associated with decreased mRNA stability[118]; while the T form was the predominant form identified in the AT patients, contains the polymorphisms in the sequence T-T-A-$(AGGG)_2$-$(ATCT)_1$-T-C, and was associated with increased mRNA stability[118]. Deletion of a 57 bp region containing a putative polymorphic, rs11103544 (T/C), miRNA binding site and the short tandem repeat polymorphism (STRP) rs71746744 (-/AGGG), abolished the difference in mRNA stability between these two forms[118] (Figure 2.1). SNP rs11103544, within the putative miRNA binding site, has previously been reported not to associate with AT[102]. STRP rs71746744 (-/AGGG), within the 57 bp deleted region, as well as the two upstream polymorphisms, rs16399 (ATCT/-) and rs1134170 (A/T), have recently been shown to be associated with AT[104]. The AGGG/AGGG, -/- and TT genotypes of rs71746744, rs16399 and rs1134170, respectively, were significant over-represented in the tendinopathy group[104]. The variants are thought to alter the secondary structure of the mRNA transcript, changing the accessibility of microRNAs or other as yet unidentified RNA stability proteins to bind to the transcript, and therefore altering the stability of the mRNA[118].

Additionally, an upstream polymorphic (rs3196378, C/A) miRNA binding site for Hsa-miR-608 in the *COL5A1* 3'-UTR has also been shown to be functional[118] (Figure 2.1). The mature 25bp Hsa-miR-608 microRNA is encoded by the *MIR608* gene situated on chromosome 10q24 and is itself polymorphic (rs4919150, C/G). Hsa-miR-608 has the sequence: 5'-AGG GGT GGT GTT GGG ACA GCT SCG T-3', where S is a cytosine (C) or guanine (G). Hsa-miR-608, containing the wild-type C allele, binds preferentially to the A allele of the *COL5A1* rs3196378 (C/A) polymorphism in[118]. Although the effect of the rs4919510 G allele on binding to the 3'-UTR is unknown the CC genotype of rs4919510 was independently associated with AT[104].

Since the *COL5A1*-3'-UTR SNP, rs12722, has previously been shown to associate with ACL ruptures in females[103], the aim of this study was to investigate the association of rs71746744 (-/AGGG) and rs1134170 (A/T), within the potentially functional region of the 3'-UTR, with ACL ruptures in a South African Caucasian population. It was hypothesised that the AGGG/AGGG and TT genotypes of these polymorphisms would be associated with ACL ruptures in females. In addition, the association of the *MIR608* gene variant (rs4919510, C/G) with ACL ruptures was also investigated in this study. It was hypothesised that the CC genotype of this miRNA SNP would be associated with ACL ruptures.

2.2 Materials and Methods

2.2.1 Study Design

A case-control genetic association study was conducted on individuals of self-reported European descent recruited from the South African population. Approval for the study was obtained from the Research Ethics Committee of the Faculty of Health Sciences within the University of Cape Town (reference number 164/2006) (Appendix B, B1.1).

2.2.2 Participants

Two hundred and fifteen Caucasian participants (160 male and 55 female) with surgically confirmed ACL rupture (ACL group) were recruited from the Sports Science Orthopaedic Clinics at the Sports Science Institute of South Africa (SSISA). Prior to participation, participants were required to sign informed written consent forms in accordance with the declaration of Helsinki (Appendix B, B1.2). Participants were provided with information about the study (Appendix B, B1.3) and were required to complete questionnaires (Appendix B, B1.4) containing detailed information on personal and sporting particulars, injury details and medical history. Information on the mechanism of injury was used to identify participants that sustained their injury through non-contact mechanisms (NON, 55.8%, n=120), as these individuals are likely to represent a more susceptible group. The NON group was therefore analysed as a sub-group of the ACL group in all statistical analyses. The remainder of the participants in the ACL group had sustained their injuries via direct contact (10.7%, n=23), indirect contact (15.3%, n=33) or ski (5.6%, n=12) mechanisms. Several of the participants had no clear mechanism of injury and could not be classified (12.6%, n=27).

In addition, 217 apparently healthy, unrelated, physically active Caucasians (134 male and 83 female) without any self-reported history of ligament or tendon injury were recruited as control (CON) participants from sports clubs and the Sports Science Institute (SISSA) wellness centre within the Southern Suburbs region of Cape Town, South Africa. Both the ACL and CON groups completed detailed sports participation particulars. Sports participation of the CON and ACL groups was characterised into contact sports, non-contact jumping sports, non-contact non-jumping sports and skiing sports as previously defined,[90] with slight modification[108]. On average, the female ACL and CON groups were matched for participation in non-contact non-jumping sports as well as skiing sports. The male ACL and CON groups were however only matched for non-contact non-jumping sports. On average, more females and males from the ACL groups participated in both contact sports and non-contact jumping sports in comparison to the CON groups. In addition, this study noted that more males from the ACL group also participated in skiing sports in comparison to the male only CON group.

2.2.3 DNA Extraction

Approximately 4.5ml of venous blood was collected via venipuncture of a forearm vein into EDTA vacutainer tubes and stored at 4 °C until total DNA extraction. Total DNA was extracted from blood using a method described by Lahri and Nurnberger[179] and modified by Mokone et al[100] (See appendix C1 for details). The DNA was stored at -20 °C until polymerase chain reaction (PCR) analysis.

2.2.4 *COL5A1* and *MIR608* Taqman® Genotyping

Genotyping of the rs71746744 STRP and rs1134170 SNP in the *COL5A1* 3'-UTR and the rs4919510 SNP in *MIR608* was performed using custom designed Fluorescence-based Taqman PCR assays (Applied Biosystems, Foster City, CA, USA). Inventoried allele specific probes and flanking primer sets (sequences available from manufacturer on request) were used along with a pre-made PCR mastermix containing ampliTaq® DNA polymerase Gold (Applied Biosystems) in a final reaction volume of 8 µl. For quality control purposes, a minimum of three DNA samples of known genotypes, together with at least four DNA free control samples were included on each PCR plate. The PCR was performed on a StepOnePlus™ real-time PCR machine (Applied Biosystems). The PCR conditions consisted of a 10 minute heat activation step (95°C), followed by 40 cycles of 15 seconds at 92°C and 1 minute at 60°C. Genotypes were determined by end-point fluorescence and analyzed using the StepOne Software Version 2.2.2 (Applied Biosystems) (Figure 2.2).

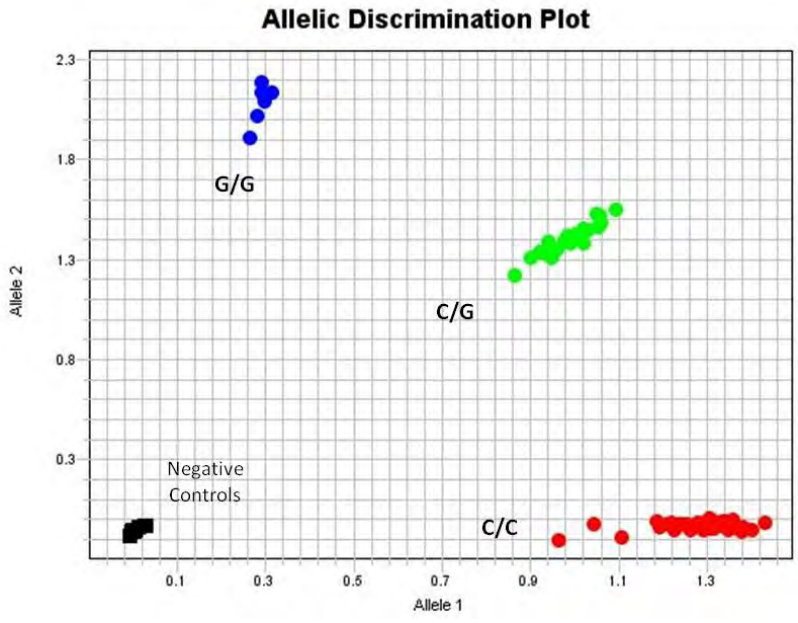


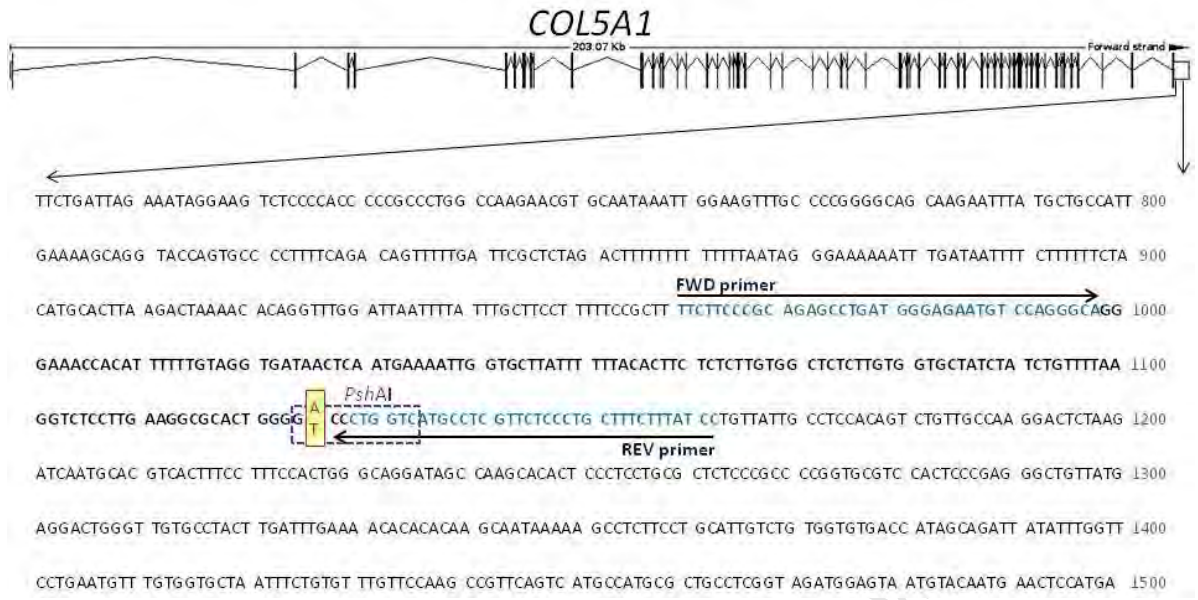
Figure 2.2 An example of a typical end-point fluorescence allelic discrimination plot using StepOne Software version 2.2.2 (Applied Biosystems). This plot shows *MIR608* rs4919510. The primer for Allele 1 (C) is labelled with a VIC probe, and allele 2 (G) is labelled with a FAM probe. The different genotypes cluster according to their fluorescence, resulting in 3 clusters, containing VIC only (C/C), VIC and FAM in equal amounts (C/G) and FAM only (G/G). No-DNA (Negative) controls are included in each experiment to determine the origin of the scatter.

2.2.5 *COL5A1* rs1134170 RFLP Genotyping

Due to difficulties discerning between the AT and AA genotypes of SNP rs1134170 using the TaqMan®-based methods, 220 samples were regenotyped using restriction fragment length polymorphism (RFLP) analysis. The PCR reactions were performed in a total volume of 50 µl containing at least 100 ng of total genomic DNA, 20 pmol of each of the respective forward (5'- CAG AGC CTG ATG GGA GAA TGT CCA GGG CA -3') and reverse (5'- GGA TAA AGA AAG CAG GGA GAA CGA GGC ATG ACC AG -3') primers, 1X reaction buffer (20 mM Tris-HCl, 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 2 mM $MgSO_4$, 0.1 % Triton X-100), 200 µM of each dNTPs and 1 unit of Taq DNA polymerase (New England Biolabs, Ipswich, Massachusetts, USA). The cycling conditions included an initial denaturing step at 95°C for 2 minutes followed by 30 cycles of denaturing at 95°C for 30 seconds, annealing at 65°C for 30 seconds, extension at 68°C for 1.5 minutes and a final extension step at 68°C for 5 minutes using the XP Thermal Cycler Block (Bioer Technology Co., Tokyo, Japan). The PCR products were restricted using 5U *PSh*AI restriction enzyme in 1x NEBuffer 4, supplemented with 100 µg/ ml BSA (New England Biolabs). The products were digested overnight at 37° C to produce 159-bp and 34-bp fragments for the A allele and an uncut 193-bp fragment for the T allele. The digestion products were separated on 2% agarose gels and visualized using SYBER Gold staining (Invitrogen Molecular probes, Eugene, Oregon). The gels were photographed under UV light using a Uvitec photodocumentation system (Uvitec Limited, Cambridge, UK). The O'GeneRuler DNA size standard (Fermentas) was included on all gels to facilitate sizing of the fragments. The genotypes were derived from the sizes of the DNA fragments obtained (Figure 2.3).

It should be noted that not all DNA samples were successfully genotyped for each polymorphism investigated. Although there are numerous possible reasons for failure to amplify, the exact reasons are not known. It is possible that a reagent used during DNA extraction, or present in the buffer in which the sample is suspended, may interfere with primer binding. Alternatively, participants may harbour a unique polymorphism at the primer binding site. Repeated genotyping of a sample that has failed to amplify may produce a false result, and therefore samples were excluded from analysis if they failed to amplify after two additional attempts.

A



B

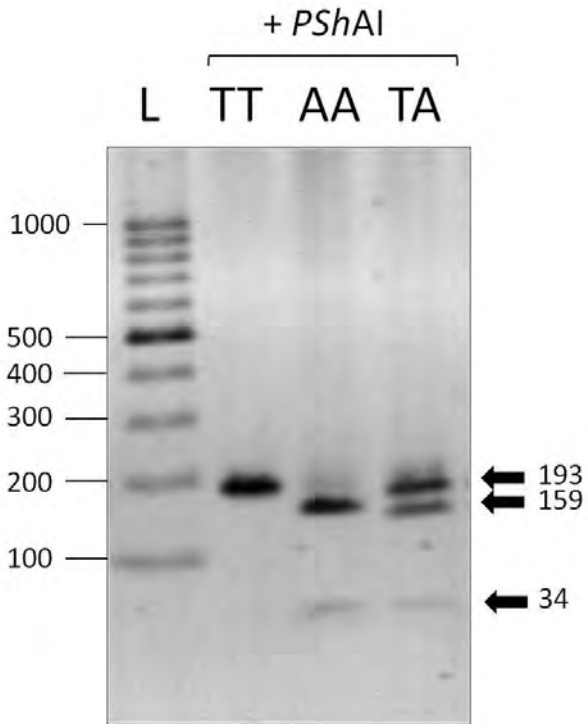


Figure 2.3 A) Schematic representation of the rs1134170 (A/T) polymorphism in the 3'-UTR of the *COL5A1* gene. Exons are represented by vertical lines, and introns are represented by adjoining horizontal lines. An 800 bp region of the 3'-UTR is indicated showing the location of the rs1134170 (A/T) polymorphism. Sequence numbers refer to 3'-UTR position. The 193 bp amplicon used for RFLP analysis is indicated in bold, flanked by the forward (FWD) and reverse (REV) primers (black arrows). The recognition sequence of the *Psh*AI restriction endonuclease is indicated (dashed box). **B)** An example of a typical 2% agarose gel showing the genotypes of the *COL5A1* rs1134170 polymorphism. The genotype of each sample is indicated at the top of each lane. The left lane (**L**) contains the DNA ladder: O'GeneRuler 100 bp DNA Ladder (Fermentas). The *PSh*AI enzyme only restricts in the presence of an A-allele. Therefore the **TT** genotype remains undigested, the **AA** genotype produces a product of 159 bp and 34 bp, and the **AT** genotype produces three fragments of 193 bp, 159 bp and 34bp. [RFLP analysis was used to discern between AT and AA genotypes, and TT genotypes and undigested samples were included as controls.]

2.2.6 Statistical Analysis

Quanto version 1.2.4. was used to determine the statistical power of the study for a given sample size and allele frequencies[180]. Assuming a minor allele frequency of 0.29, 0.28 and 0.19 for rs71746744, rs1134170 and rs4919510 respectively, a sample size of approximately 215 cases and an equal number of controls would be adequate to detect a genetic effect size of at least 1.5, 1.5 and 1.6 respectively, at a power of 80% and a significance level of 5%. GenePop was used to determine whether the SNPs were in Hardy-Weinberg equilibrium (HWE) (http://genepop.curtin.edu.au/genepop_op1.html). Data were analysed using STATISTICA Version 10.0 (Stat- Soft, Tulsa, OK, USA) and GraphPad Prism version 5.0d for Mac OS X (GraphPad Software, San Diego, CA, USA) programs. Continuous variables, such as weight, age, height and BMI were compared between the CON, ACL and NON groups using one-way analysis of variance (ANOVA), and were co-varied where necessary. Pearson's chi-squared tests, or Fischer's tests when required, were used to determine whether there were any significant differences in allele frequencies or other categorical variables between the CON and ACL groups and NON subgroups. Females and males were analysed separately and together to look for sex-specific associations. Chaplin Case-control haplotype inference package (Epstein Software) was used to infer haplotypes and pseudohaplotypes[181,182]. Linkage disequilibrium (LD) between the *COL5A1* rs71746744 STRP and rs1134170 SNP was calculated using Cubex software (http://www.oege.org/cgi-bin/cubex.py)[183]. In all cases, statistical significance was accepted when $p < 0.05$.

2.3 Results

2.3.1 Participant Characteristics

Age, height and weight are self-reported values at the time of the first ACL rupture for the ACL group and NON subgroup, and at recruitment for the CON group. The participants in the CON and ACL groups, as well as the NON subgroup, were matched for age and country of birth (Table 2.1). There were significantly fewer males in the CON group (61.6 %) compared to the ACL group (74.4%) (p=0.004), as well as the NON subgroup (75.8%) (p=0.008). After adjustment for sex, there were no significant differences in height between the CON and ACL (adjusted p=0.368) groups and NON (adjusted p=0.167) subgroup. However, after adjusting for sex, the ACL group and NON subgroup were significantly heavier (ACL, adjusted p=0.005; NON, adjusted p=0.015) than the CON group, with a correspondingly higher BMI (ACL, adjusted p<0.001, NON, adjusted p=0.002). The ACL participants were recruited on average 4.5 ± 8.8 years after the initial injury and had gained on average 2.1 ± 12.4 kg in that time. There were no significant genotype effects on height, weight or BMI, and there were no significant genotype distribution differences between sex and age (Appendix A: Table A1.1).

Table 2.1 Characteristics of the participants within the asymptomatic control group (CON), anterior cruciate ligament rupture group (ACL) and the ACL subgroup with a non-contact mechanism of injury (NON).

	CON	ACL	P-value ‡	NON	P-value §
Age (years)	28.7±11.2 (210)	27.1±11.0 (187)	0.167	26.8±10.5 (114)	0.130
Sex (% male)	61.6 (216)	74.4 (215)	**0.004**	75.8 (120)	**0.008**
Height (cm)	175.2±9.2 (210)	177.6±9.4 (194)	0.368[a]	178.3±9.2 (116)	0.167[a]
Weight (kg)	74.2±15.1 (211)	80.7±16.8 (194)	**0.005**[a]	80.4±15.8 (117)	**0.015**[a]
BMI (kg/m^2)	24.1±3.5 (208)	25.9±4.1 (191)	**<0.001**[a]	25.7±3.9 (115)	**0.002**[a]
Country of birth (% South African)	86.5 (207)	83.2 (196)	0.354	82.9 (117)	0.385

Participants who were genotyped for at least one of rs71746744, rs1134170 or rs4919510 were included in the analysis. Except for sex and country of birth, the variables are expressed as mean ± standard deviation (n= number of participants for which valid data was available). Sex and country of birth are expressed as a percentage. Age, height and weight are self-reported values at the time of the first ACL rupture for the ACL group and NON subgroup, and at recruitment for the CON group. Significant p-values <0.005 are indicated by bold typeface.

‡ CON vs ACL,
§ CON vs NON,
[a] P-value adjusted for sex

2.3.2 Genotype and Allele Frequency Distributions

2.3.2.1 *COL5A1* rs71746744 (AGGG/-)

There were no significant *COL5A1* rs71746744 genotype (p=0.840) or allele (p=0.830) frequency differences between the ACL (n=214) and CON (n=216) groups for the combined participants. Similarly, there were no significant genotype (male, p=0.802; female, p=0.309) or allele (male, p=0.717; female, p=0.343) frequency differences between the male CON (n=133) and male ACL (159) participants, or the female CON (n=83) and female ACL (n=55) participants. When analysing the NON subgroups, there were no significant genotype or allele frequency differences between the combined (p=0.971 and p=1.000, respectively), male (p=0.401 and p=0.643, respectively) or female (p=0.479 and p=0.485, respectively) NON subgroups when compared to their respective CON groups. Except for the male CON group, which was not in HWE (p=0.050), all the other groups and subgroups were in HWE. The previously investigated SNP rs12722 (C/T) was in high LD with rs71746744 for the CON group (D'=0.964; r^2 =0.461), ACL group (D'=0.944, r^2=0.520) and combined groups (D'=0.956, r^2=0.481).

It should be noted that, although not significantly different, the -/- genotype distribution within the female CON subgroup (10.8%, n=9 of 83) was higher than the female ACL group (3.6%, n=2 of 55, p=0.126) and NON subgroup (3.5%, n=1 of 29, p=0.229) (Table 2.2).

2.3.2.2 *COL5A1* rs1134170 (T/A)

There were no significant *COL5A1* rs1134170 genotype (p=0.519) or allele (p=0.339) frequency differences between the ACL group (n=215) and CON (n=216) group for the combined participants (Table 2.3). Similarly, there were no significant genotype (male, p=0.746; female, p=0.733) or allele (male, p=0.519; female, p=0.489) frequency differences between the male CON (n=134) and ACL (n=160) participants, or female CON (n=82) and female ACL (n=55) participants. When analysing the NON subgroups, there were also no significant differences in genotype or allele frequency distribution of the combined (p=0.624 and p=0.395 respectively), male (p=0.781 and p=0.524 respectively) or female (p=0.833 and p=0.621 respectively) NON subgroups when compared to their respective CON groups. The CON group (p=0.030) and male ACL (p=0.022) subgroups were not in HWE. All the other groups and subgroups were in HWE. The previously investigated SNP rs12722 (C/T) was in high LD with rs1134170 for the CON group (D'=0.947; r^2 =0.463), ACL group (D'=0.907, r^2=0.437) and combined participants (D'=0.933, r^2=0.454).

Table 2.2 Genotype and allele frequency distribution of *COL5A1* rs71746744 (AGGG/-) in the asymptomatic control (CON) group, ACL rupture (ACL) group and non-contact (NON) ACL rupture subgroup.

		Genotype					Allele		
	N	AGGG/ AGGG	AGGG/ -	-/ -	P-Value	HWE	AGGG	-	P-Value
All Participants	430								
• **CON group**	216	48.6 (105)	44.4 (96)	6.9 (15)		0.266	70.8 (306)	29.2 (126)	
• **ACL group**	214	48.6 (104)	45.8 (98)	5.6 (12)	0.840	0.071	71.5 (306)	28.5 (122)	0.830
• **NON subgroup**	120	49.2 (59)	43.3 (52)	7.4 (9)	0.971	0.593	70.8 (170)	29.2 (70)	1.000
Male Participants	292								
• **CON group**	133	48.1 (64)	47.4 (63)	4.5 (6)		**0.050**	71.8 (191)	28.2 (75)	
• **ACL group**	159	47.2 (75)	46.5 (74)	6.3 (10)	0.802	0.138	70.5 (224)	29.5 (94)	0.717
• **NON subgroup**	91	48.3 (44)	42.9 (39)	8.8 (8)	0.401	0.877	69.8 (127)	30.2 (55)	0.643
Female Participants	138								
• **CON group**	83	49.4 (41)	39.8 (33)	10.8 (9)		0.548	69.3 (115)	30.7 (51)	
• **ACL group**	55	52.7 (29)	43.6 (24)	3.6 (2)	0.309	0.267	74.6 (82)	25.4 (28)	0.343
• **NON subgroup**	29	51.7 (15)	44.8 (13)	3.5 (1)	0.479	0.363	74.1 (43)	25.9 (15)	0.485

Values are percentages, with sample number (n) displayed in parenthesis. N is the total number of successfully genotyped participants in each group. HWE is the p-value for the exact tests for HWE. Significant values, ($p<0.05$) are shown in bold typeface.

Table 2.3 Genotype and allele frequency distribution of *COL5A1* rs1134170 (T/A) in the asymptomatic control (CON) group, ACL rupture (ACL) group and non-contact (NON) ACL rupture subgroup.

		Genotype					Allele		
	N	TT	AT	AA	P-Value	HWE	T	A	P-Value
All Participants	431								
• **CON group**	216	48.1 (104)	46.8 (101)	5.1 (11)		**0.030**	71.5 (309)	28.5 (123)	
• **ACL group**	215	52.1 (112)	44.6 (96)	3.3 (7)	0.519	0.011	74.4 (320)	25.6 (110)	0.339
• **NON subgroup**	120	52.5 (63)	44.2 (53)	3.3 (4)	0.624	0.071	74.6 (179)	25.4 (61)	0.395
Male Participants	294								
• **CON group**	134	48.5 (65)	47.0 (63)	4.5 (6)		0.054	72.0 (193)	28.0 (75)	
• **ACL group**	160	51.9 (83)	45.0 (72)	3.1 (5)	0.746	**0.022**	74.4 (238)	25.6 (82)	0.519
• **NON subgroup**	91	52.7 (48)	44.0 (40)	3.3 (3)	0.781	0.118	74.7 (136)	25.3 (46)	0.524
Female Participants	137								
• **CON group**	82	47.6 (39)	46.3 (38)	6.1 (5)		0.280	70.7 (116)	29.3 (48)	
• **ACL group**	55	52.7 (29)	43.6 (24)	3.6 (2)	0.733	0.267	74.5 (82)	25.5 (28)	0.489
• **NON subgroup**	29	51.7 (15)	44.8 (13)	3.5 (1)	0.833	0.363	74.1 (43)	25.9 (15)	0.621

Values are percentages, with sample number (n) displayed in parenthesis. N is the total number of successfully genotyped participants in each group. HWE is the p-value for the exact tests for HWE. Significant values, ($p<0.05$) are shown in bold typeface.

Table 2.4 Genotype and allele frequency distribution of *MIR608* rs4919510 (C/G) in the asymptomatic control (CON) group and ACL rupture (ACL) group and non-contact (NON) ACL rupture subgroup.

		Genotype					Allele		
	N	CC	CG	GG	P-Value	HWE	C	G	P-Value
All Participants	430								
• **CON group**	217	67.3 (146)	27.6 (60)	5.1 (11)		0.149	81.1 (352)	18.9 (82)	
• **ACL group**	213	63.4 (135)	32.4 (69)	4.2 (9)	0.543	0.961	79.6 (339)	20.4 (87)	0.572
• **NON subgroup**	119	64.7 (77)	30.3 (36)	5.0 (6)	0.878	0.509	79.8 (190)	20.2 (48)	0.689
Male Participants	293								
• **CON group**	134	70.9 (95)	23.9 (32)	5.2 (7)		0.064	82.8 (222)	17.2 (46)	
• **ACL group**	159	62.3 (99)	32.1 (51)	5.7 (9)	0.277	0.480	78.3 (249)	21.7 (69)	0.169
• **NON subgroup**	91	61.5 (56)	31.9 (29)	6.6 (6)	0.340	0.406	77.5 (141)	22.5 (41)	0.157
Female Participants	137								
• **CON group**	83	61.5 (51)	33.7 (28)	4.8 (4)		0.950	78.3 (130)	21.7 (36)	
• **ACL group**	54	66.7 (36)	33.3 (18)	0.0 (0)	0.254	0.142	83.3 (90)	16.7 (18)	0.307
• **NON subgroup**	28	75.0 (21)	25.0 (7)	0.0 (0)	0.294	0.450	87.5 (49)	12.5 (7)	0.133

Values are percentages, with sample number (n) displayed in parenthesis. N is the total number of successfully genotyped participants in each group. HWE is the p-value for the exact tests for HWE. Significant values, ($p<0.05$) are shown in bold typeface.

2.3.2.3 *MIR608* rs4919510 (C/G)

There were no significant *MIR608* rs4919510 genotype (p=0.543) or allele (p=0.572) frequency differences between the ACL (n=213) and CON (n=217) groups (Table 2.4). Similarly, there were no significant genotype (male, p=0.277; female, p=0.254) or allele (male, p=0.169; female, p=0.307) frequency differences between the male CON (n=134) and ACL (n=159) participants, or female CON (n=83) and female ACL (n=54) participants. When analysing the NON subgroups, there were also no significant differences in genotype or allele frequency distribution of the combined (p=0.878 and p=0.689 respectively), male (p=0.340 and p=0.157 respectively) or female (p=0.294 and p=0.133 respectively) NON subgroups when compared to their respective CON groups. All of the groups were in HWE.

It should be noted that, although not significantly different, the CC genotype distribution within the female NON subgroup (75.0%, n=21 of 28) was higher than the female CON group (61.5%, n=51 of 83, p=0.294) (Table 2.4). Similarly the GG genotype, although present within the female CON group (4.8%, n=4 of 83), was absent from the female ACL group and NON subgroup (Table 2.4)

2.3.3 Inferred Haplotype and Pseudohaplotypes

As the polymorphisms investigated in this study could theoretically interact with one another to regulate type V collagen production, gene-gene interactions were investigated.

2.3.3.1 Inferred Haplotype Analysis of *COL5A1* rs71746744 and rs1134170

Inferred haplotypes consisting of *COL5A1* rs71746744 and rs1134170 were constructed separately for the male and female CON and ACL groups (Appendix A: Table A1.2.1), as well as NON subgroups (Appendix A: Table A1.2.2). Since rs71746744 and rs1134170 were in high LD for the CON group (D'=0.977; r^2=0.922), ACL group (D'=0.973, r^2=0.809) and combined groups (D'=0.975, r^2=0.866), three of the four possible haplotypes, AGGG/T, AGGG/A and -/A, were inferred with a frequency greater than 1% (Figure 2.4). The AGGG/T inferred haplotype, which was hypothesised to be associated with increased risk for AT, was the major haplotype in all the groups (68.0%). However in agreement with the hypothesis, the opposite -/A inferred haplotype was significantly overrepresented in the male CON (27.6%, n=36 of 134) group compared to the ACL group (25.0%, n=39 of 160) and NON (24.7 %, 22 of 120) subgroup (CON vs ACL, p=0.001; CON vs NON, p=0.026) (Figure 2.4B).

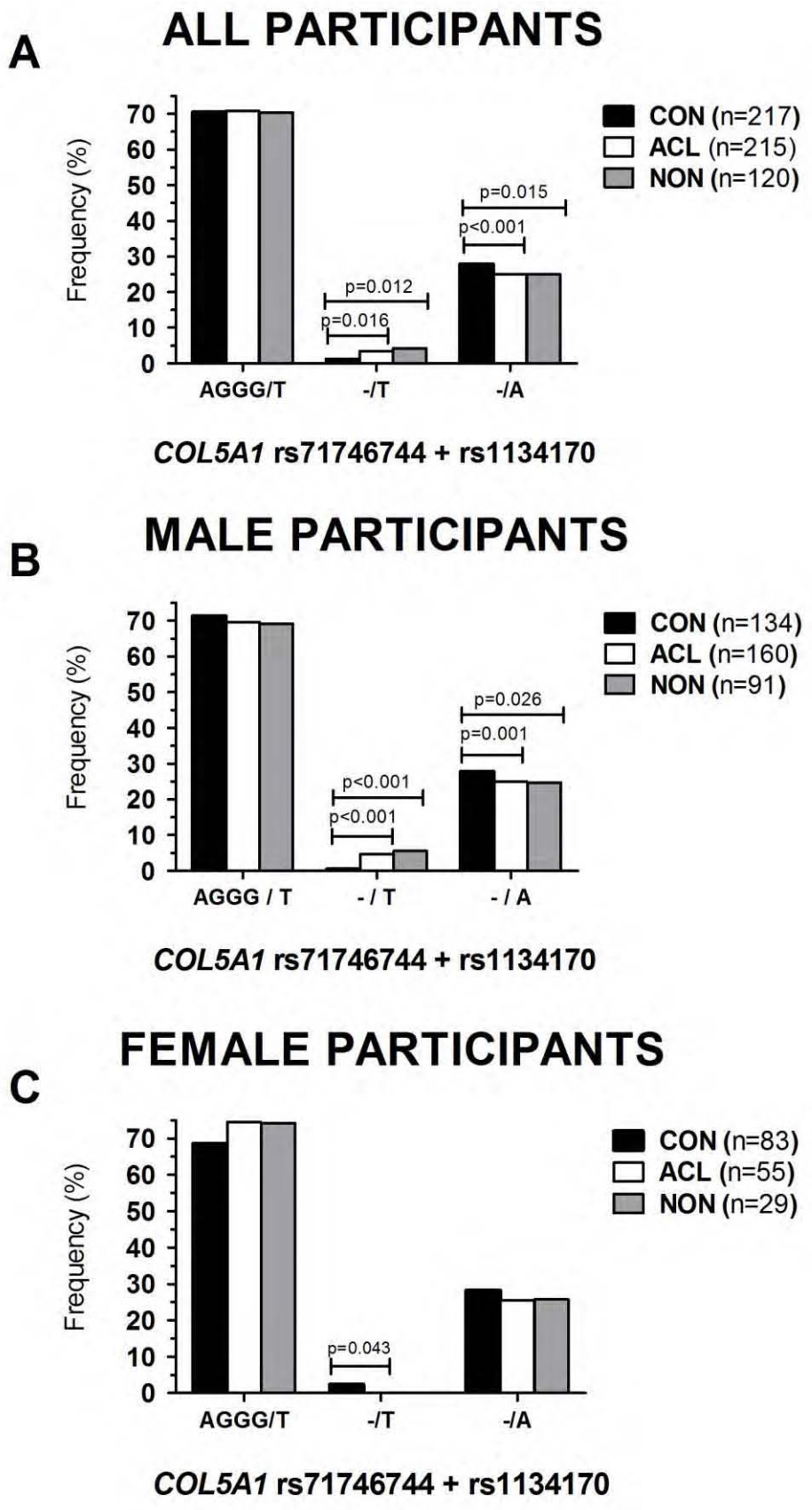


Figure 2.4 Inferred haplotype frequencies for *COL5A1* rs71746744 and rs1134170 in asymptomatic controls (CON, solid black bars), anterior cruciate ligament group (ACL, white bars) and non-contact (NON, grey bars) subgroup, in (**A**) combined participants, (**B**) males and (**C**) females. The number of participants (n) in each group is in parenthesis. Significant p values <0.05 are indicated by a line on the graph.

Although similar frequencies were observed for the female CON group (28.3%, 23 of 81), ACL group (25.5%, 13 of 53) and NON subgroup (25.9%, 7 of 29), the -/A inferred haplotype was not significantly different (CON vs ACL, p=0.206; CON vs NON, p=0.319) (Figure 2.4C). When the male and female groups were analysed together the -/A inferred haplotype was also significantly over-represented in the CON group (27.9%, 60 of 217) compared to the ACL group (25.1%, 53 of 215) (p<0.001) and NON subgroup (25%, 29 of 120) (p=0.015) (Figure 2.4A).

The rare -/T inferred haplotype (<6%) of rs71736744 and rs1134170 was over-represented in the combined ACL (3.5%, 7 of 215) group and male ACL (4.7%, 7 of 160) group compared to the combined CON group (1.2%, 2 of 217) (p=0.016) and male CON group (0.0%, 0 of 134) (p<0.001); however in females the -/T haplotype was over-represented in the CON (2.4%, 2 of 83) group compared to the ACL (0.0%, 0) group (p=0.042). The frequency of the -/T haplotype in NON subgroups were similar to the ACL group, and remained significant for the combined (4.2%, 5 of 120) (p=0.012) and male (5.5%, 5 of 91) (p<0.001) participants. There was however no significant difference in -/T haplotype frequency between the female CON group and NON subgroups (p=0.120).

2.3.3.2 Pseudohaplotype Analysis of *COL5A1* rs71746744 and rs1134170 and *MIR608* rs4919510

As before, inferred pseudohaplotypes were constructed separately for the male and female CON and ACL groups (Appendix A: Table A1.3.1), as well as NON subgroups (Appendix A: Table A1.3.2). Five of the possible eight pseudohaplotypes (AGGG/T/C, AGG/T/G, -/T/C, -/A/C and -/A/G), were inferred with a frequency greater than 1% from the rs71746744 and rs1134170 polymorphisms in *COL5A1* and the rs4919510 polymorphism in *MIR608* (Figure 2.5).

The AGGG/T/C inferred pseudohaplotype, which was hypothesised to be associated with increased risk of AT, was not significantly different in the combined ACL group (p=0.715) or NON subgroups (p=0.370) (Figure 2.5A), however it was significantly (p=0.032) over-represented in the female ACL (63.4%, 34 of 55) group compared to the female CON (54.1%, 44 of 83) group (Figure 2.5C). The AGGG/T/C inferred pseudohaplotype was however significantly (p=0.040) under-represented in the male NON subgroup (53.5%, 48 of 91) compared to the male CON group (57.1%, 76 of 134)(Figure2.5B).

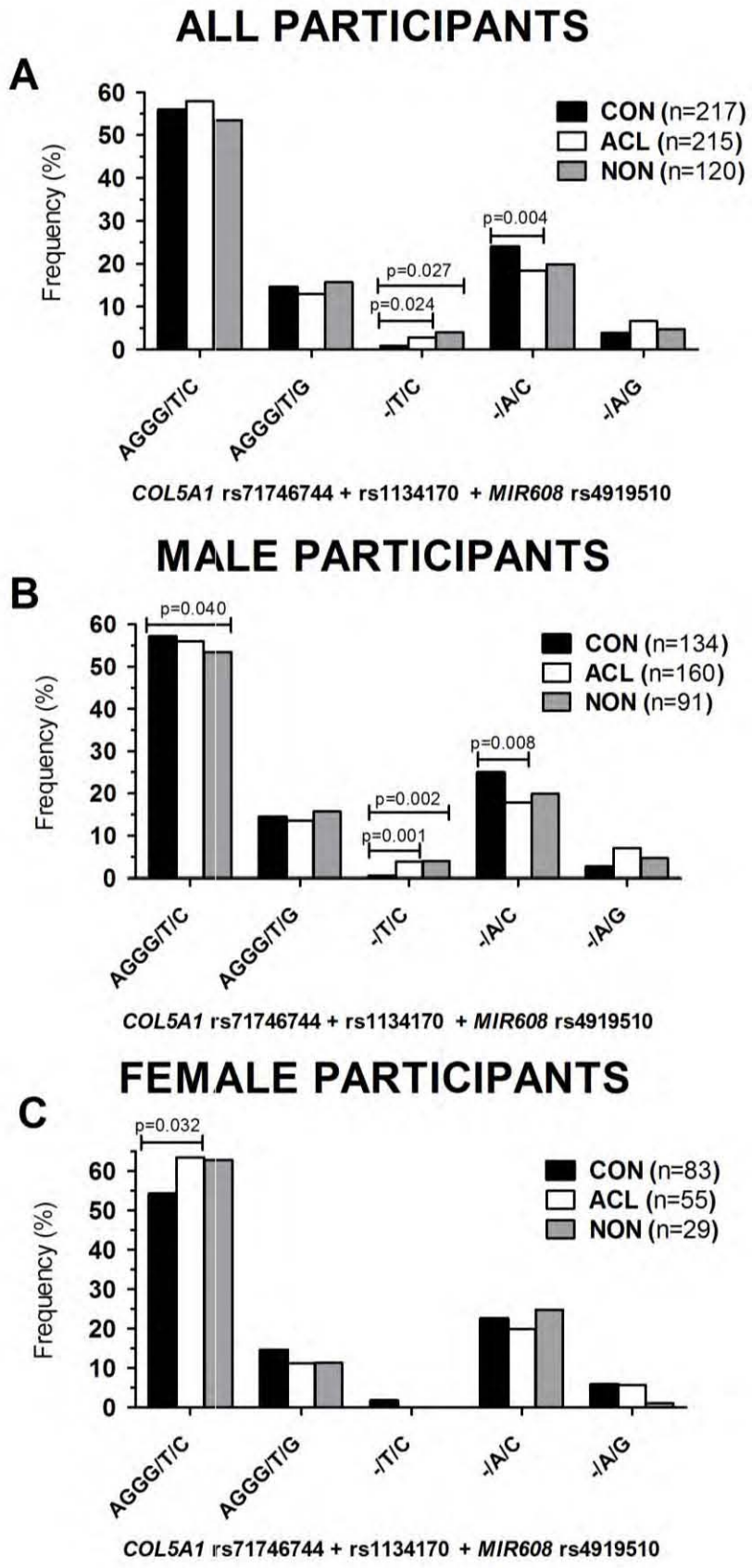


Figure 2.5 Inferred pseudohaplotype frequencies for *COL5A1* rs71746744 and rs1134170 and *MIR608* rs4919510 for (**A**) combined participants (**B**) males and (**C**) females in asymptomatic controls (CON, solid black bars), anterior cruciate ligament group (ACL, white bars) and non-contact (NON, grey bars) subgroup. The number of participants (n) in each group is in parenthesis. Significant p values <0.05 are indicated by a line on the graph.

The -/A/C inferred pseudohaplotype was significantly over-represented in the combined CON (24,0%, 52 of 217) group compared to the ACL (18.4%, 39 of 215) group (p=0.004). The -/A/C haplotype was also significantly over-represented in the male CON (25.2%, 33 of 130) group compared to the male ACL group (17.9%, 28 of 156) (p=0.008) (Figure 2.5B). There was no significant difference between the -/A/C haplotype in the female ACL group (p=0.221) or in the combined (p=0.107), male (p=0.106) or female (0.603) NON subgroups.

Finally the rare -/T/C inferred pseudohaplotype was significantly over-represented in the combined ACL group (2.8%, 6 of 215) and NON subgroup (3.0%, 3 of 120) compared to the CON group (0.8%, 1 of 217) (CON vs ACL, p=0.024; CON vs NON, p=0.027). This relationship was also observed in the male participants, with the -/T/C inferred pseudohaplotype being over-represented in both the male ACL (3.9%, 6 of 160) group and male NON (4.0%, 3 of 91) subgroup compared to the male CON (0.0%, 0 of 134) group (CON vs ACL, p=0.001, CON vs NON, p=0.002).

2.4 Discussion

The common SNP, rs12722 (C/T) in the 3'-UTR of the *COL5A1* gene, has previously been associated with AT[100, 102] and ACL rupture in females[103]. Recently, several additional polymorphisms, rs71746744 (-/AGGG), rs16399 (ATCT/-) and rs1134170 (A/T), in close proximity to the rs12722 SNP were investigated for associations with AT[104]. The AGGG/AGGG -/- ATCT and TT genotypes of *COL5A1* rs71746744 (-/AGGG), rs16399 (ATCT/-) and rs1134170 (A/T) respectively were found to associate with an increased risk of AT[104]. Since rs12722 has previously been associated with ACL rupture in females, this study investigated the *COL5A1* rs71746744 (-/AGGG) and rs1134170 (A/T) SNPs for an association with ACL rupture.

The first main finding of this study was that there were no significant independent associations between the *COL5A1* rs71746744 (-/AGGG), *COL5A1* rs1134170 (A/T) and *MIR608* rs4919510 (G/C) polymorphisms and risk of ACL rupture. In addition, the genotype and allele frequencies remained non-significant when participants were analysed according to sex or mechanism of injury.

Since it has previously been shown that the *COL5A1* rs12722 SNP is associated with increased risk of ACL rupture in females[108], it was hypothesised that the AGGG/AGGG and TT genotypes of rs71746744 and rs1134170 may also be associated with ACL rupture in females. Although there were no significant associations in the female group, it was noted that the rare -/- genotype distribution of rs71746744 within the female CON subgroup (10.8%) was higher than the female ACL (3.6%) and NON (3.5%) subgroups. This trend is consistent with previous findings that showed that the rs71746744 -/- genotype was protective against AT[104]. Further research within a larger cohort of female participants is required to establish if this gene variant is a risk factor for ACL ruptures. In agreement with previously published findings, where rs12722 (C/T) was not associated with ACL ruptures in males[103], the two additional *COL5A1* 3'-UTR polymorphisms analysed in this study were also not associated with ACL ruptures in males. The sample size of the male participants included in this study was large enough to detect associations with a genetic effect size of between 1.6 and 1.7, with a power of 80% and significance level of 5%. The allele and genotype frequency distributions of the two *COL5A1* polymorphisms were similar to previously published values for a Caucasian South African population[104].

The CC genotype of the *MIR608* gene was previously associated with an increased risk of AT[104], however there was no association between the CC genotype and risk of ACL rupture. It should be noted however, that in females, the CC genotype was higher in the NON subgroup and the GG genotype was present in the CON group, but entirely absent in both the ACL and NON subgroups.

Although insignificant, this trend is in agreement with the hypothesis that CC genotype of *MIR608* could be associated with ACL ruptures in females. It has been proposed that Hsa-miR-608 binds to *COL5A1* 3'-UTR, altering mRNA stability, and potentially changing the concentration of type V collagen, with subsequent effects on collagen fibril diameter and biomechanical properties[110]. The allele and genotype frequency distributions of the *MIR608* polymorphisms were similar to previously published values for a Caucasian South African population[104].

There was a departure from HWE for the *COL5A1* rs71746744 and rs1134170 markers in both the ACL and CON groups. The departure may be due to the sample sizes being too small; or it may be a sign that the sample does not represent a random mating population[184,185]. Since the ACL groups were selected based on their clinical diagnosis, the departure may also be a true reflection of a difference in genotype frequency distribution in the injured group.

COL5A1 haplotypes were constructed separately for males and females as previous studies have identified different genetic risk factors in males and females[103,108] . An additional finding of this study was that, although a small and probably not a practically relevant difference, the -/A inferred haplotype constructed from *COL5A1* rs71746744 and rs1134170 was significantly (p=0.001) under-represented in males with ACL ruptures when compared to controls. This finding could indicate that variants within the *COL5A1* 3'-UTR are "weak" modulators of ACL ruptures in males. An association with the -/A haplotype and protection is consistent with previous studies on the *COL5A1* 3'-UTR[118] and AT[104], and agree with the C-form of the *COL5A1* 3'-UTR identified from asymptomatic controls[118] (Figure 2.1).

Inferred pseudohaplotypes consisting of the *COL5A1* rs71746744 (-/AGGG), and rs1134170 (A/T) polymorphisms and the *MIR608* rs4919510 (G/C) SNP were constructed because the *COL5A1* 3'-UTR contains a functional binding site for Hsa-miR-608 [118]. The additive contribution of the C, rather than the G, allele of the *MIR608* (C/T) SNP to a protective effect against ACL rupture when combined with the -/A *COL5A1* inferred haplotype, was however unexpected, and we cannot exclude the possibility that this is a false positive. Much larger samples sizes are required to test whether the *COL5A1* 3'-UTR polymorphisms are weak modulators of risk for ACL ruptures in males.

There was no significant difference in the distribution of the proposed at risk AGGG/T inferred haplotype constructed from *COL5A1* rs71746744 and rs1134170 between the female ACL rupture and control groups. However when the *MIR608* rs4919510 SNP was included in the analysis the

proposed at risk AGGG/T/C inferred haplotype, constructed from the two *COL5A1* and one *MIR608* polymorphism, was significantly (p=0.032) associated with increased risk of ACL ruptures in females.

It has been hypothesised that differences in the secondary structure of the *COL5A1* mRNA alters its stability, either via its accessibility to microRNAs or other as yet unidentified mRNA stability proteins[104,118]. As a result, more α1(V) chains can be translated from the more stable transcript. Synthesis of the α1(V) chain is the rate-limiting step in type V collagen production, and type V collagen is known to regulate collagen fibril diameter and density[19,21,23]. This has potential effects on the biomechanical properties of the tissue as fibril diameter and density have been associated with both the tensile strength and elasticity of a tissue[73]. Thus the *COL5A1* 3'-UTR polymorphisms may exert their effects via changes in biomechanical properties[110]. The significant haplotype association of -/A suggests that the genetic variants may not be significant individually, but in combination, due to the effect that the combined variants have on the mRNA secondary structure.

Although the results of this study largely support previous findings on risk allele combinations for AT[104], there were no significant independent associations detected, despite a sample size that should have been able to detect an effect size of 1.6. The reason for the different effect size of these polymorphisms in AT compared to ACL injury may be due to the difference in the aetiology of these two injuries. AT is an overuse injury that results from the accumulation of microtrauma to the tendon over a long period. Thus tendinopathic tissue often has many pathological changes, concomitant with the altered regulation of many proteins and the dysregulation of numerous biological pathways. One such change measured in degenerative tendons, is an increase in type III and V collagens, and a decrease in type I collagen[186]. However ACL rupture is an acute event, and thus degeneration and altered regulation of proteins, such as type V collagen is less likely. Nonetheless, *COL5A1* may still influence ACL risk via altering the tensile strength of the tissue.

The fact that neither of the *COL5A1* genetic variants was independently associated with ACL rupture in this study does not necessarily mean that these variants do not play a role in its aetiology. It is possible that the *COL5A1* gene does not play an important role in ACL rupture in males, but does play a role in female injury. Alternatively, the effect size of the *COL5A1* variant may be smaller in ACL rupture compared to AT due to its different aetiology, and may therefore require larger samples sizes to detect.

In conclusion, we propose that rs71746744 and rs1134170 gene variants in the *COL5A1* 3'-UTR cause changes in mRNA stability that alter type V collagen production and therefore modify the architecture of the collagen fibres. Individuals with a -/A haplotype will have decreased *COL5A1* mRNA stability, decreased type V collagen synthesis and therefore larger mean diameter fibrils. We propose that as a result of increased fibril diameter, the ligaments of -/A individuals will have increased tensile strength, and thus a decreased risk of rupture when exposed to excessive internal or external forces.

Chapter 3: The *COL11A1* and *COL11A2* Genes and Anterior Cruciate Ligament Rupture

3.1 Introduction

As reviewed in Chapter 1 (Section 1.5.1), an individual's genetic profile has been identified as one of the important intrinsic risk factors for ACL ruptures, and a number of genetic variants have been shown to associate with ACL rupture to date (table 1.1). One such variant, a common C/T single nucleotide polymorphism (SNP rs12722) within the 3' -UTR of *COL5A1*, was shown to associated with ACL ruptures in females[103]. Specifically, the CC genotype was shown to be over-represented in the asymptomatic female controls[103]. Both increased knee hyperextension (genu recurvatum)[187,188] and hamstring flexibility[189] have also been associated with risk of ACL ruptures. Interestingly, the CC genotype of *COL5A1* rs12722, which was under-represented in females with ACL rupture, was also associated with decreased knee hyperextension and general joint laxity in females[190]. This suggests that the *COL5A1* rs12722 genotype influences injury risk via changes in joint laxity (the biomechanical properties of the ligament). In addition, an age-related increase in lower limb range of motion measurements has also been reported in individuals with the *COL5A1* rs12722 CC genotype[116]. These findings suggest that variants within the *COL5A1* gene and other related genes may be important genetic risk factors for ACL ruptures.

As previously discussed, *COL5A1* encodes the α1 chain of type V collagen, which plays an important role in collagen fibril nucleation and the regulation of fibril diameter (fibrillogenesis) [19,20]. In the previous chapter, two additional polymorphisms, rs71746744 (-/AGGG) and rs1134170 (A/T), within the *COL5A1* 3'-UTR, were investigated for an association with ACL ruptures. These variants have previously been associated with AT[102,104] and are believed to be functional, altering *COL5A1* mRNA stability within the cytoplasm of the tenocyte[118]. Increased mRNA stability has been associated with tendinopathy, which is believed to result from increased α1(V) chain and type V collagen production, which results in decreased collagen fibril diameter and packing density, and potentially altered mechanical properties of the tendon[110]. In agreement with the *COL5A1* rs12722 SNP findings, these two additionally analysed *COL5A1* 3'-UTR polymorphisms were not associated with ACL ruptures in males (Chapter 2). Although the trends were as proposed, the association of these additional polymorphisms with ACL ruptures in females was however inconclusive because of the small sample size (Chapter 2).

Type XI collagen shares structural homology with type V collagen[28], which together with types II and IX collagen, forms heterotypic fibrils in cartilage [29,44]. Both type V and type XI collagens regulate collagen fibril diameter[28]. This regulation occurs via their protruding N-terminal domains, which limit the appositional lateral growth of collagen fibrils by blocking further accretion of type I and type II collagen molecules respectively[28]. In addition, both collagens play a role in the interaction of collagens with proteoglycans of the ECM[28]. Until recently, the expression of type XI collagen was thought to be confined to cartilage. However, it has now been demonstrated that type XI collagen is expressed in many non-cartilaginous tissues[28,30,45,47,48,146], as well as in the fibrocartilaginous zone of the mature ACL[30].

Type XI collagen is a quantitatively minor fibrillar collagen comprising three polypeptide chains; α1(XI), α2(XI) and α3(XI), encoded by the *COL11A1*, *COL11A2* and *COL2A1* genes respectively[191]. The α3 (XI) chain was revealed to be an overglycosylated form of the α1(II)-collagen chain, lacking an N-terminal cysteine-rich domain[191]. Mutations in *COL2A1*, *COL11A1* and *COL11A2* have all been implicated in various inherited connective tissue disorders[121–124,192,193], and polymorphisms within these genes have previously been associated with musculoskeletal injuries or connective tissue disorders[128–130,134,136,137,139,141,142].

The *COL11A1* gene has previously been investigated for associations with early-onset arthritis[126], osteoarthritis[127], lumbar disc pathologies[131,132] cleft palate[128] and cancer[133]. However, it has only been significantly associated with lumbar disc desiccation and bulging[128], LDH[129], and limbus vertebra in gymnasts[130]. The *COL11A2* gene has been investigated for associations with osteoarthritis[126,127], rheumatoid arthritis[134], osteochondrodysplasias [135], OPLL [136–139], lumbar[128,140,141] and intevertebral [142,143] disc disease, and cleft palate [131,132]. However the only diseases which showed significant association included: lumbar disc desiccation, bulging and height narrowing[128], rheumatoid arthritis [134] and OPLL[137].

Functional polymorphisms identified by these earlier studies include the rs3753841 (T/C) and rs1676486 (C/T) SNPs of *COL11A1* and the rs1799907 (A/T) SNP of *COL11A2*. Specifically, the C allele of rs3753841 was associated with LDH[129] and the T-allele of the rs1676486 was associated with LDH[129] and limbus vertebra in gymnasts[130]. Additionally, the T allele of *COL11A2* rs1799907 was associated with risk of lumbar disc desiccation[128], lumbar spine stenosis[141] and OPLL[136–139].

The primary aim of this study was therefore to investigate the association of the *COL11A1* rs3753841 (T/C) and rs1676486 (C/T) polymorphisms, as well as, the *COL11A2* rs1799907 (A/T) polymorphism

with ACL ruptures in a South African Caucasian population. A secondary aim of this study was to investigate whether the *COL11A1* and *COL11A2* polymorphisms interact with the previously investigated rs71746744 (-/AGGGG) polymorphism within the *COL5A1* 3'-UTR to modulate the risk of ACL rupture. Although the previous Chapter did not show an independent association with rs71746744 and risk of ACL rupture, deletion of a 57 bp region in the *COL5A1* 3'-UTR containing the rs71746744 STRP abolishes the differences in mRNA stability between tendinopathic and asymptomatic individuals[118]. The rs71746744 variant was therefore selected based on its proposed biological function.

3.2 Methods

3.2.1 Study Design and Participants

Two hundred and fifteen (160 males and 55 females) participants with surgically confirmed ACL ruptures and 216 (133 males and 83 females) apparently healthy, unrelated, physically active control (CON) participants were recruited for this case-control genetic association study as previously described in Section 2.2.2 of Chapter 2. Prior to participation in this study, all participants gave written informed consent (Appendix B, B1.2) and completed the questionnaires (Appendix B, B1.4) as detailed in the previous Chapter. The 120 (91 males and 29 females) participants who had ruptured their ACL via confirmed non-contact (NON) mechanisms were analysed as a separate subgroup in this study.

Approval for the study was obtained from the Research Ethics Committee of the Faculty of Health Sciences within the University of Cape Town (reference number 164/2006) (Appendix B, B1.1).

3.2.2 Polymorphism Selection

Two genes, *COL11A1* and *COL11A2*, encoding for the α1(XI) and α2(XI) subunits of type XI collagen respectively, were chosen for this analysis. Functional polymorphisms that had previously been associated with musculoskeletal conditions were chosen.

COL11A1

The *COL11A1* gene has been mapped to human chromosome 1p21. The rs3753841 (T/C) SNP is a T>C single nucleotide base change in exon 52 of the *COL11A1* gene that results in a non-synonymous amino acid change from a leucine to a proline at position 1323 of the pro α1(XI) protein (Figure 3.1). This polymorphism has been investigated in a number of conditions, and has been significantly associated with susceptibility to LDH[129].

The rs1676486 SNP is a T>C single nucleotide base change in exon 62 of the *COL11A1* gene that results in a non-synonymous amino acid change from a proline to a serine at amino acid position 1535 (Figure 3.1). In addition to the potential affects from the amino acid change, Mio et al showed that the T allele of rs1676486 had significantly lower expression than the C allele, which appears to be due to increased mRNA degradation of the T allele transcript[129]. This variant has been associated with limbus vertebra in gymnasts[130] and LDH[129].

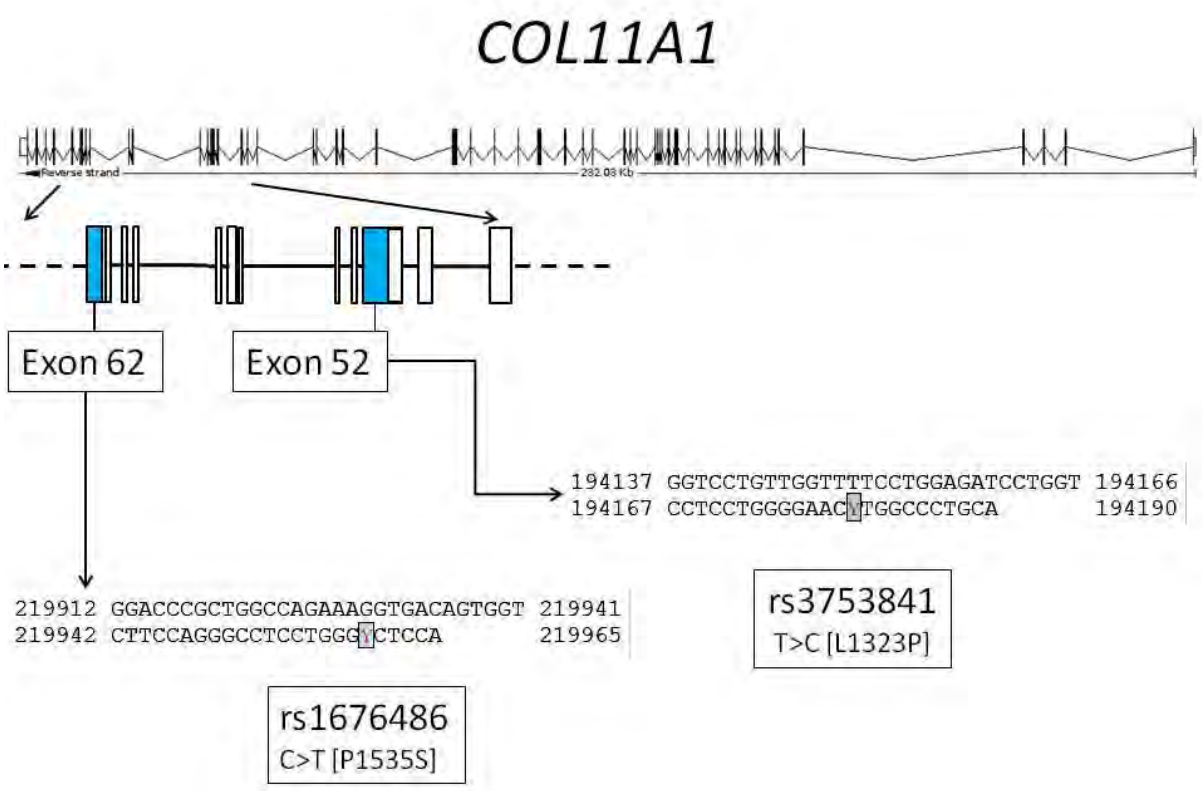


Figure 3.1 Schematic drawing of the *COL11A1* gene, showing the position of the selected rs3753841 and rs1676486 SNPs. The *COL11A1* gene is located on the reverse strand of chromosome 1 and is 232 030 base pairs in length. The entire *COL11A1* transcript is depicted, with exons represented as vertical lines, and introns represented by the adjoining lines. A close-up of the region containing exons 49-62 is shown below, with the position of the SNPs depicted in boxes underneath. The SNP accession number together with its respective base change and resulting amino acid change are indicated in boxes below their location. The sequence and position of the SNPs in *COL11A1* is relative to the ENST00000370096 ensembl transcript. The processed transcript is 7286 bp in length, comprises 67 exons, and encodes a 1806 amino acid protein.

Notation in this study is (major allele/ minor allele).
Although T (leucine) is the major allele of rs3753841 in a Caucasian population, the ancestral allele is C (proline). The ancestral allele of rs1676486 is C (proline), which is also the major allele.

COL11A2

The *COL11A2* gene has been mapped to human chromosome 6p21.3. The IVS6-4 (rs1799907) polymorphism in *COL11A2* is a T-A base change located in intron 5, 4 base pairs upstream of the start of exon 6, that results in alternative slicing (Figure 3.2). The rs1799907 SNP has been significantly associated with lumbar disc desiccation, bulging and height narrowing [128] rheumatoid arthritis [134] and OPLL[49-51] .

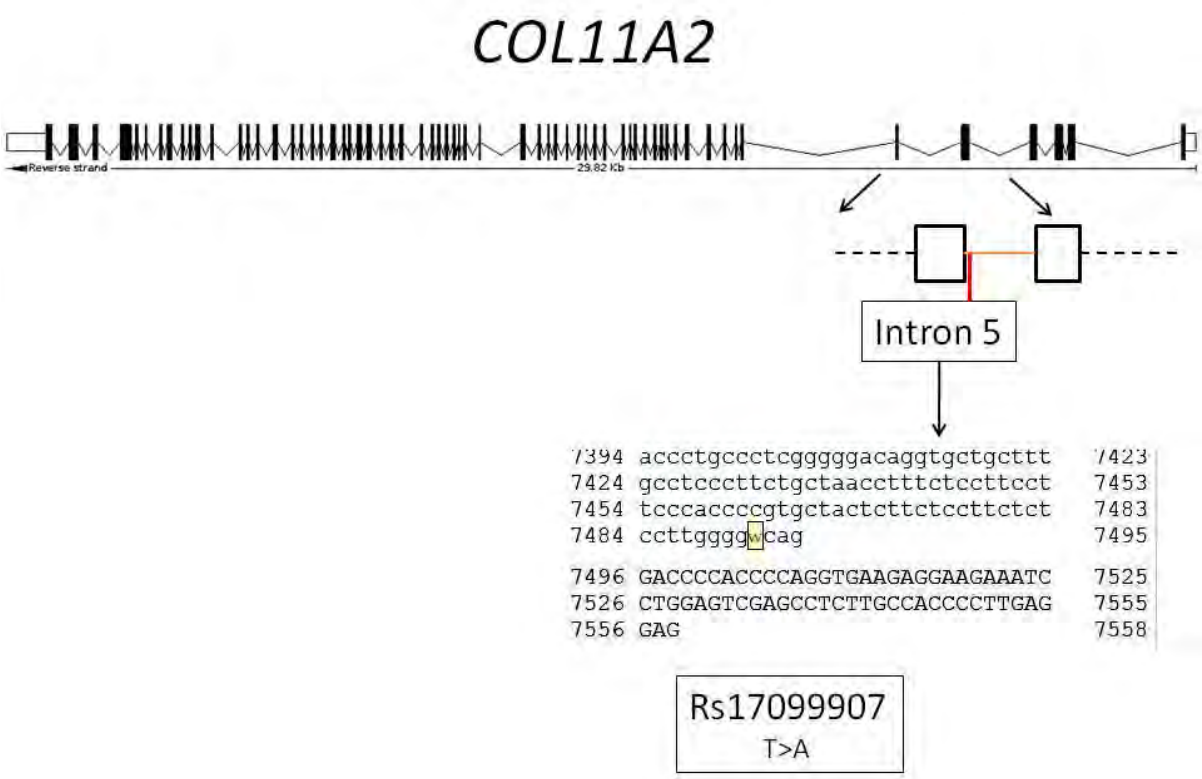


Figure 3.2 Schematic drawing of the *COL11A2* gene showing the position of the rs1799907 SNP. The *COL11A2* gene is located on the reverse strand of chromosome 6 and there are 12 *COL11A2* transcripts. The entire *COL11A2* transcript is depicted with exons represented as vertical lines, and introns represented by the adjoining lines. A close-up of the region containing exons 5-6 and intron 5 is shown below, with the position of the SNP depicted in a box beneath. The SNP accession number together with its respective base change and resulting amino acid change is indicated below. The position of the SNP is relative to the ENST00000448717 ensembl transcript. The gene is 29 819 bp in length. The processed transcript is 6209 bp long, comprising 64 exons that encode a 1650 amino acid protein.

COL5A1

This sample population in this study has previously been genotyped for a number of polymorphisms within the 3'-UTR of the *COL5A1* gene, which has been mapped to human chromosome 9q34[101,102]. In addition to AT, the rs12722 SNP has been associated with ACL rupture in females[103], ROM[115,116] and endurance running performance[117]. There are seven tightly linked polymorphisms within this region: rs13946, rs12722, rs3196378, rs71746744, rs16399, rs1134170 and rs3128575, which are thought to alter mRNA stability[118]. The rs71746744 polymorphism was selected to investigate gene-gene interaction effects between *COL11A1* and *COL11A2* variants because deletion of a 57 bp region containing rs71746744, was shown to abolish the difference in mRNA stability between asymptomatic and AT clones[118] (Refer to Figure 2.1).

3.2.3 Genotyping

Genotyping was performed using fluorescence-based Taqman PCR assays (Applied Biosystems, Foster City, CA, USA). Allele specific probes and flanking primer sets (sequences available from manufacturer on request) were used along with a pre-made PCR mastermix containing ampliTaq® DNA polymerase Gold (Applied Biosystems, Foster City, CA, USA) in a final reaction volume of 8 µl. The PCR reactions were run on StepOnePlus qRT-PCR machine (Applied Biosystems) using the manufacturers recommended cycling conditions (Chapter 2, Section 2.2.4). A small number of samples failed to amplify (Chapter 2, Section 2.2.5).

3.2.4 Statistical Analysis

Quanto version 1.2.4. was used to determine the statistical power of the study for a given sample size and allele frequency [180]. Assuming a minor allele frequency of 0.41, 0.21 and 0.34 for rs3753841, rs1676486 and rs1799907 respectively, a sample size of approximately 215 cases and equal number of controls would be adequate to detect a genetic effect size of at least 1.5, 1.6 and 1.5 respectively, at a power of 80% and a significance level of 5%. GenePop was used to determine whether the SNPs were in HWE (http://genepop.curtin.edu.au/genepop_op1.html). Data were analysed using STATISTICA Version 10.0 (Stat- Soft, Tulsa, OK, USA) and GraphPad Prism version 5.0d for Mac OS X (GraphPad Software, San Diego, CA, USA) programs. Continuous variables, such as weight, age, height and BMI were compared between the CON and ACL groups and NON subgroups using one-way ANOVA and values were covaried for sex where necessary. Pearson's chi-squared tests, or Fischer's tests when required, were used to determine whether there were any significant differences in allele frequencies or other categorical variables between the CON and ACL groups and NON subgroups. Females and males were analysed separately and together to look for sex-specific associations. Chaplin case-control haplotype inference package (Epstein Software) was used to infer haplotypes and pseudohaplotypes[181,182]. LD between the *COL11A1* rs3753841 and rs1676486 SNPs was calculated using Cubex software (http://www.oege.org/cgi-bin/cubex.py)[183]. In all cases, statistical significance was accepted when $p<0.05$.

3.3 Results

3.3.1 Participant Characteristics

The participants in the CON and ACL groups, as well as the NON subgroup, were matched for age and country of birth (Table 3.1). There were significantly more males in the ACL group (74.4%, p=0.004) and the NON subgroup (75.8%, p=0.008) compared to the CON group (61.4%). The participants were matched for height after adjusting for sex. However the ACL group and NON subgroup remained significantly heavier (ACL, adjusted p=0.005; NON, adjusted p=0.015) than the CON group, with a correspondingly higher BMI (ACL, adjusted p<0.001; NON, adjusted p=0.002) when adjusted for sex.

The ACL participants were recruited on average 4.5 ± 8.8 years after the initial injury and had gained 2.1 ± 12.4 kg on average after the rupture. There were no genotype effects on any of the descriptive measures of the participants (Appendix A, Table A2.1).

Table 3.1 Characteristics of the participants within the asymptomatic control group (CON), anterior cruciate ligament rupture group (ACL) and the ACL subgroup with a non-contact mechanism of injury (NON).

	CON	ACL	p-value ‡	NON	p-value §
Age (years)	28.7±11.2 (210)	27.1±11.0 (187)	0.167	26.8±10.5 (114)	0.130
Sex (% male)	61.6 (216)	74.4 (215)	**0.004**	75.8 (120)	**0.008**
Height (cm)	175.±9 (210)	178±9 (194)	0.368 [a]	178±9 (116)	0.168[a]
Weight (kg)	74.2±15.1 (211)	80.7 ±16.8 (194)	**0.005** [a]	80.4±15.8 (117)	**0.015**[a]
BMI (kg/m^2)	23.8±4.2 (210)	26.2±9.2 (196)	**<0.001** [a]	25.5±4.5 (116)	**0.002** [a]
Country of birth (% South African)	86.5 (207)	83.2 (196)	0.354	82.9 (117)	0.385

Participants who were genotyped for at least one of rs3753841, rs1676486 or rs1799907 SNPs were included in the analysis. Most variables are expressed as mean ± standard deviation (n= number of participants for which valid data was available). Sex and country of birth are expressed as a percentage. Age, height and weight are self-reported values at the time of the first ACL rupture for the ACL group and NON subgroup, and at recruitment for the CON group. Significant p-values are indicated by bold typeface.

‡ CON vs ACL
§ CON vs NON
[a] P-value adjusted for sex

3.3.2 Genotype and Allele Frequency Distributions

3.3.2.1 *COL11A1* rs3753841 (T/C)

There was a significant difference (p=0.028) in the *COL11A1* rs3753841 genotype distribution when the ACL (n=215) group was compared to the CON (n=216) group (Table 3.2). The TC genotype was significantly (p=0.009, OR=1.7, 95% CI=1.1 to 2.5) over-represented in the CON group (54.2%) when compared to the ACL group (41.4%). Although the genotype distribution of the NON (n=120) subgroup was similar to the ACL group, there was no significant difference (p=0.162) in the rs3753841 genotype distribution between the CON group and NON subgroup. Similarly there were no significant differences in the allele frequency distributions when the ACL group (p=0.718) or the NON subgroup (p=0.503) were compared to the CON group (Table 3.2).

When the male and female participants were analysed separately, there were no significant differences in the rs3753841 genotype (male p=0.139 and female p=0.175) or allele (male p=0.606 and female p=1.000) frequency distributions when the male (n=160) or female (n=55) ACL groups were compared to their respective CON groups (male, n=133; female n=82) (Table 3.2). Similarly there were also no significant differences in the rs3753841 genotype (male p=0.303 and female p=0.508) or allele (male p=0.493 and female p=1.000) frequency distributions when the male (n=91) or female (n=29) NON subgroup was compared to their respective CON groups (Table 3.2). The genotype and allele frequency distributions of the male and female groups were nevertheless similar to their combined frequency distribution. Except for the ACL group (p=0.046), all the other groups and subgroups were in HWE.

3.3.2.2 *COL11A1* rs1676486 (C/T)

There were no significant *COL11A1* rs1676486 genotype (p=0.212) or allele (p=0.282) frequency differences between the ACL (n=215) and CON (n=216) groups for the combined participants (Table 3.3). Similarly, there were no significant differences in the rs1676486 genotype (male p=0.250 and female p=0.440) or allele distributions (male p=0.161 and female p=1.000) when the male (n=160) and female (n=55) ACL groups were compared to the appropriate CON groups (male, n=133; female, n=82) (Table 3.3). The genotype and allele frequency distributions of the male and female groups were nevertheless similar to their combined frequency distribution.

Table 3.2 Genotype and allele frequency distribution of *COL11A1* rs3753841 (T/C) in the asymptomatic control group (CON) and ACL rupture (ACL) and non-contact (NON) ACL rupture groups.

		Genotype			P-Value		Allele		P-Value
	N	TT	TC	CC		HWE	T	C	
All Participants	431								
• **CON group**	216	31.9 (69)	54.2 (117)	13.9 (30)		0.078	59.0 (255)	40.9 (177)	
• **ACL group**	215	39.5 (85)	41.4 (89)	19.1 (41)	**0.028**	**0.046**	60.2 (259)	39.8 (171)	0.718
• **NON subgroup**	120	40.0 (48)	43.3 (52)	16.7 (20)	0.162	0.361	61.7 (148)	38.3 (92)	0.503
Male Participants	293								
• **CON group**	133	30.8 (41)	54.1 (72)	15.0 (20)		0.203	57.9 (154)	42.1 (112)	
• **ACL group**	160	38.8 (62)	42.5 (68)	18.7 (30)	0.139	0.147	60.0 (192)	40.0 (128)	0.606
• **NON subgroup**	91	39.6 (36)	44.0 (40)	16.5 (15)	0.303	0.496	61.5 (112)	38.5 (70)	0.493
Female Participants	137								
• **CON group**	82	34.1 (28)	53.7 (44)	12.2 (10)		0.248	61.0 (100)	39.0 (64)	
• **ACL group**	55	41.8 (23)	38.2 (21)	20.0 (11)	0.175	0.142	60.9 (67)	39.1 (43)	1.000
• **NON subgroup**	29	41.4 (12)	41.4 (12)	17.2 (5)	0.508	0.514	62.1 (36)	37.9 (22)	1.000

Values are expressed as a frequency (%) with the number of subjects (n) in parentheses. N is the total number of successfully genotyped participants in each group. HWE is the p-value for the exact tests for HWE. Significant p-values are indicated by bold typeface.

When analysing the NON subgroups there were also no significant differences in the distribution of the rs1676486 genotype or allele frequency within all participants (p=0.528 and p=0.513 respectively), male (p=0.631 and p=0.416 respectively) or female (p=0.554 and p=1.000 respectively) NON subgroups compared to their respective CON groups (Table 3.3). Although not significant, there was a tendency for the TT genotype to be over represented in the female NON subgroup (10.3%, 3 of 83), compared to the female CON group (2.4%, 2 of 29). Except for the female NON subgroup (p=0.005), all the other groups and subgroups were in HWE.

Table 3.3 Genotype and allele frequency distribution of *COL11A1* rs1676486 (C/T) in the asymptomatic control group (CON) and ACL rupture (ACL) and non-contact (NON) ACL rupture groups.

		Genotype			P-Value		Allele		P-Value
	N	CC	CT	TT		HWE	C	T	
All Participants	431								
• **CON group**	216	61.6 (133)	34.3 (74)	4.2 (9)		0.746	78.7 (340)	21.3 (92)	
• **ACL group**	215	68.4 (147)	26.5 (57)	5.1 (11)	0.212	0.089	81.6 (351)	18.4 (79)	0.282
• **NON subgroup**	120	66.7 (80)	28.3 (34)	5.0 (6)	0.528	0.348	80.8 (194)	19.2 (46)	0.513
Male Participants	292								
• **CON group**	132	58.3 (77)	36.4 (48)	5.3 (7)		0.892	76.5 (202)	23.5 (62)	
• **ACL group**	160	67.5 (108)	27.5 (44)	5.0 (8)	0.250	0.218	81.2 (260)	18.8 (60)	0.161
• **NON subgroup**	91	63.7 (58)	33.0 (30)	3.3 (3)	0.631	0.711	80.2 (146)	19.8 (36)	0.416
Female Participants	138								
• **CON group**	83	66.3 (55)	31.3 (26)	2.4 (2)		0.598	81.9 (136)	18.1 (30)	
• **ACL group**	55	70.9 (39)	23.6 (13)	5.4 (3)	0.440	0.200	82.7 (91)	17.3 (19)	1.000
• **NON subgroup**	29	75.9 (22)	13.8 (4)	10.3 (3)	0.554	**0.005**	82.8 (48)	17.2 (10)	1.000

Values are expressed as a frequency (%) with the number of subjects (n) in parentheses. N is the total number of successfully genotyped participants in each group. HWE is the p-value for the exact tests for HWE. Significant p-values are indicated by bold typeface.

3.3.2.3 *COL11A2* rs1799907 (A/T)

There were no significant differences in the *COL11A2* rs1799907 genotype (p=0.261) or allele (p=0.737) frequency distributions between ACL (n=215) and CON (n=216) groups (Table 3.4). Similarly there were also no significant differences in the genotype (p=0.286) or allele (p=0.799) frequency distributions when the NON (n=120) subgroup was compared to the CON group (Table 3.4). There were no significant differences in the rs1799907 genotype (male p=0.344 and female p=0.592) or allele (male p=0.448 and female p=0.701) frequency distributions when the male (n=160) or female ACL (n=55) groups were analysed separately and compared to their respective CON (male, n=133, female, n=82) groups (Table 3.4).

Table 3.4 Genotype and allele frequency distribution of *COL11A2* rs1799907 (A/T) in the asymptomatic control group (CON) and ACL rupture (ACL) and non-contact (NON) ACL rupture groups.

		Genotype			P-Value		Allele		P-Value
	N	AA	AT	TT		HWE	A	T	
All Participants	431								
• **CON group**	216	44.4 (96)	43.1 (93)	12.5 (27)		0.546	66.0 (285)	34.0 (147)	
• **ACL group**	215	39.5 (85)	50.1 (109)	9.8 (21)	0.261	0.099	64.9 (279)	35.1 (151)	0.737
• **NON subgroup**	120	39.2 (47)	51.7 (62)	9.2 (11)	0.286	0.138	65.0 (156)	35.0 (84)	0.799
Male Participants	293								
• **CON group**	133	46.6 (62)	42.1 (56)	11.3 (15)		0.663	67.7 (180)	32.3 (86)	
• **ACL group**	160	39.4 (63)	50.6 (81)	10.0 (16)	0.344	0.171	64.7 (207)	35.3 (113)	0.448
• **NON subgroup**	91	37.4 (34)	52.7 (48)	9.9 (9)	0.287	0.178	63.7 (116)	36.3 (66)	0.388
Female Participants	137								
• **CON group**	82	40.2 (33)	45.1 (37)	14.6 (12)		0.757	62.8 (103)	37.2 (61)	
• **ACL group**	55	40.0 (22)	50.9 (28)	9.1 (5)	0.592	0.351	65.4 (72)	34.6 (38)	0.701
• **NON subgroup**	29	44.8 (13)	48.3 (14)	6.9 (2)	0.557	0.491	69.0 (40)	31.0 (18)	0.429

Values are expressed as a frequency (%) with the number of subjects (n) in parentheses. N is the total number of successfully genotyped participants in each group. HWE is the p-value for the exact tests for HWE. Significant p-values are indicated by bold typeface.

Similarly, when males and females were analysed separately, there were no significant differences in the rs1799907 genotype (male p=0.287 and female p=0.557) or allele (male p=0.388 and female p=0.429) frequency distributions when the male (n=91) or female (n=29) NON subgroups were compared to their respective CON groups (Table 3.4). The genotype and allele frequency distributions of the male and female groups were nevertheless similar to the combined groups.

The genotype distribution for all groups and subgroups were in HWE.

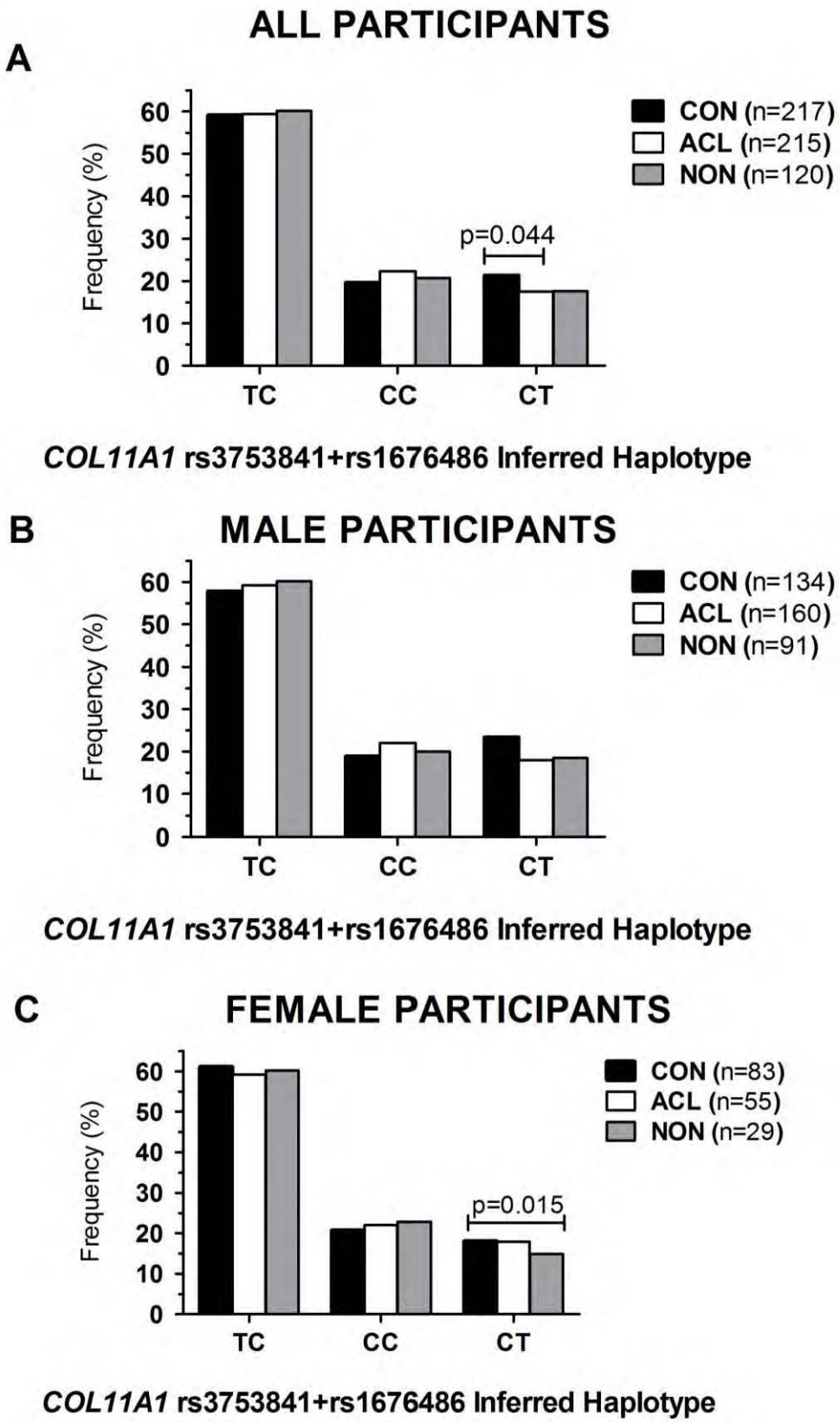


Figure 3.3 Inferred haplotype frequencies for *COL11A1* rs3753841 and rs1676486 for (**A**) combined (**B**) male and (**C**) female participants. Asymptomatic controls (CON) are depicted by solid black bars, anterior cruciate ligament (ACL) rupture groups are depicted by white bars and the non-contact (NON) subgroup is depicted by solid grey bars. The number of participants in each group is in parenthesis (n). Any significant differences ($p < 0.05$) between groups are indicated by lines on the graphs.

3.3.3 Haplotype Analysis of *COL11A1* rs3753841 and *COL11A1* rs1676486

The *COL11A1* rs3753841 and rs1676486 polymorphisms were in high LD for the CON group (D'=1.000; r^2=0.393), ACL group (D'=0.923, r^2=0.291) and combined groups (D'=0.963, r^2=0.341). Because previous studies have shown that the genetic variants that predispose to ACL risk may differ depending on sex[103,108], inferred haplotypes were constructed separately for the male and female CON and ACL groups (Appendix A: Table A2.2.1), as well as NON subgroups (Appendix A: Table A2.2.2).

Three of the possible four haplotypes (TC, CC and CT), were inferred with a frequency greater than 1% from the two *COL11A1* polymorphisms (Figure 3.3) Although the inferred haplotype frequency distribution between the combined groups were similar to the male and female groups, the *COL11A1* CT inferred haplotype was only significantly under-represented in (i) the combined ACL group (17.5%, 37 of 215) compared to the combined CON group (21.3%, 46 of 217) (p=0.044) (Figure 3.3A), and (ii) in the female NON subgroup (14.9%, 4 of 29) compared to the female CON group (18.1%, 14 of 83) (p=0.015) (Figure 3.3C).

3.3.4 Pseudohaplotype Analysis of *COL11A1* and *COL11A2*

Since the *COL11A1* (chromosome 1p21) and *COL11A2* (chromosome 6p21.3) genes encode for type XI collagen α-chains, inferred pseudohaplotypes were constructed from *COL11A1* rs3753841, *COL11A1* rs1676486 and *COL11A2* rs1799907. Once again, inferred haplotypes were constructed separately for the male and female CON and ACL groups (Appendix A: Table A2.3.1), as well as NON subgroups (Appendix A: Table A2.3.2).

Six pseudohaplotypes (TCA, TCT, CCA, CCA, CTT, CTA and CTT) of the possible eight were inferred with a frequency greater than 1% from the two *COL11A1* and one *COL11A2* SNPs (Figure 3.4). The CTA inferred pseudohaplotype was significantly (ACL vs CON, p=0.010 and NON vs CON, p=0.037) over-represented in the CON group (14.7%, 31 of 217) compared to the ACL group (9.5%, 20 of 215) and NON subgroup (10.3%, 12 of 120) (Figure 3.4A).

Similarly, in the male participants, the CTA inferred pseudohaplotype was also significantly (ACL vs CON, p=0.015 and NON vs CON, p=0.044) over-represented in the CON (15.9%, 21 of 134) group compared to the ACL group (10.4%, 16 of 160) and NON subgroup (11.6%, 10 of 91) (Figure 3.4B). There was no significant difference in inferred pseudohaplotype frequency distribution in the ACL group or NON subgroup when the females were analysed separately (Figure 3.4C). The inferred pseudohaplotype frequency distributions between the male and female CON group, ACL group and NON subgroups were similar.

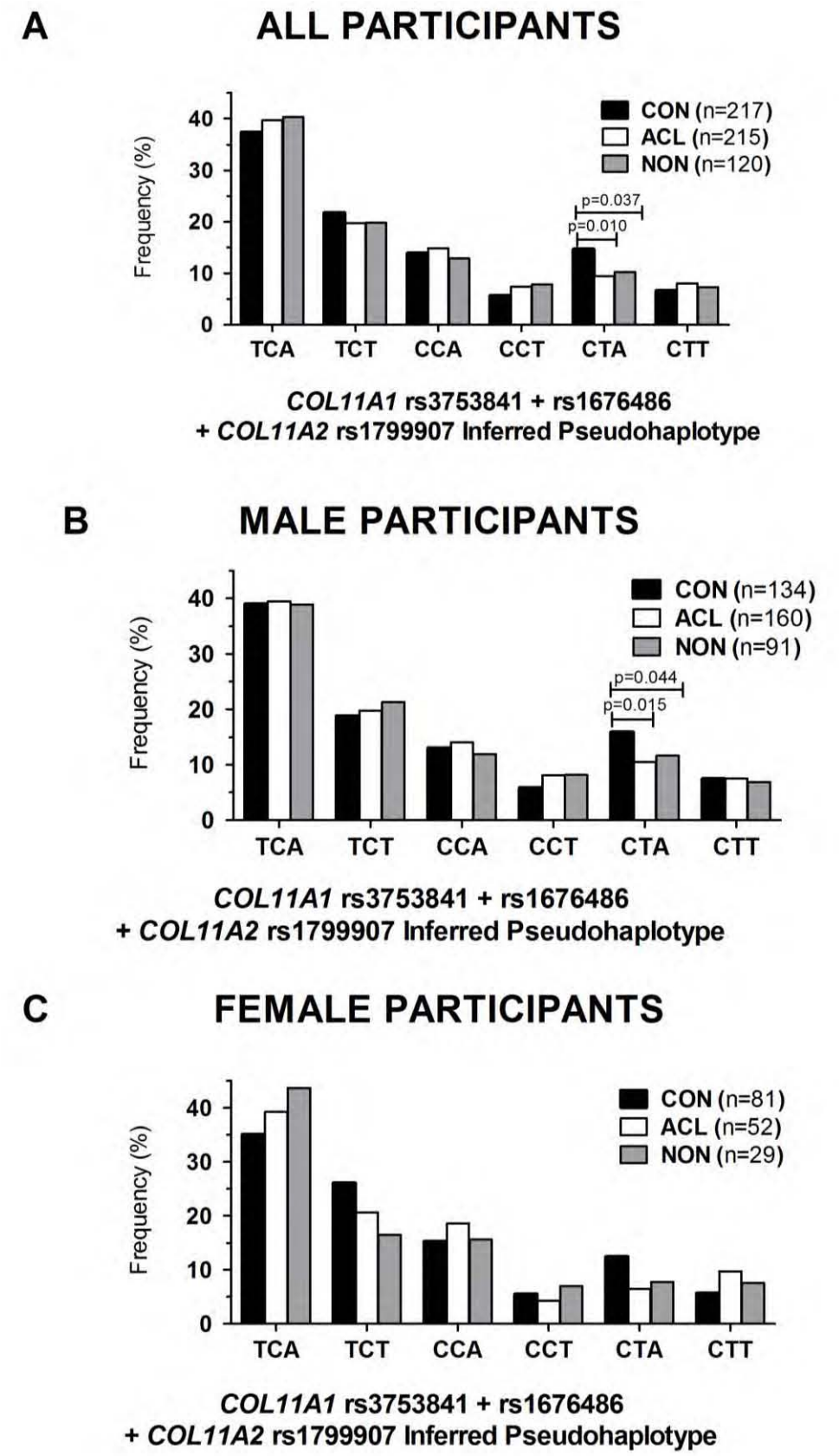


Figure 3.4 Inferred pseudohaplotype frequencies for *COL11A1* rs3753841 and rs1676486 and *COL11A2* rs1799907 polymorphisms in (**A**) combined (**B**) male and (**C**) female participants. Asymptomatic controls (CON) are depicted by solid black bars, anterior cruciate ligament (ACL) rupture groups are depicted by white bars and the non-contact (NON) subgroup is depicted by solid grey bars. Any significant differences ($p < 0.05$) between groups are indicated by lines on the graphs.

3.3.5 Pseudohaplotype Analysis of *COL11A1*, *COL11A2* and *COL5A1*

Due to the possibility for interactions between type V and type XI collagen, gene-gene interactions were investigated between polymorphisms within *COL11A1, COL11A2* and *COL5A1*. The *COL5A1* rs71746744 STRP was selected for this analysis because deletion of a 57 bp region in the *COL5A1* 3'-UTR containing the rs71746744 STRP abolishes the differences in mRNA stability between tendinopathic and asymptomatic individuals[118]. The rs12722 SNP in the *COL5A1* 3'-UTR has previously been associated with ACL rupture in females[103], therefore inferred pseudohaplotypes were once again constructed separately for the male and female CON and ACL groups (Appendix A: Table A2.4.1), as well as NON subgroups (Appendix A: Table A2.4.2).

Eleven of the sixteen pseudohaplotypes were inferred with a frequency greater than 1% from the *COL11A1* rs3753841 (T/C), *COL11A1* rs1676486 (C/T), *COL11A2* rs1799907 (A/T) and *COL5A1* rs71446744 (AGGG/-) polymorphisms (Figure 3.5). None of the expected inferred pseudohaplotypes were significantly different between the CON and ACL groups and NON subgroups when combined (Figure 3.5A), male (Figure 3.5B) or female (Figure 3.5C) participants were analysed. In the combined male and female analysis the following two inferred pseudohaplotypes were significantly different between groups: (i) TCT(AGGG) was significantly over-represented (p=0.030) in the CON (16.2%, 35 of 217) group compared to the NON (8.1%, 9 of 120) subgroup, and (ii) CTA(AGGG) was significantly over-represented (p=0.021) in the CON (12.9%, 28 of 217) group compared to the ACL (9.2%, 19 of 215) group. There was a trend (p=0.051) for the TCA(AGGG) inferred pseudohaplotype to be under-represented in the CON group (Figure 3.5A).

When only the male participants were analysed the following three inferred pseudohaplotypes were significantly different between groups: (i) TCT(AGGG) was significantly over-represented (p=0.018) in the CON (16.1%, 21 of 134) group compared to the ACL (9.0%, 14 of 160) group, (ii) TCT(-) was significantly over-represented in the ACL (10.1% 16 of 160) group and NON (14.6%, 13 of 91) subgroup compared to the CON (2.6%, 3 of 134) group (CON vs ACL, p=0.031; CON vs NON, p=0.008), and (iii) CTA(AGGG) was significantly over-represented (p=0.029) in the CON (13.6%, 18 of 134) group compared to the ACL (9.8%, 15 of 160) group (Figure 3.5B). There were however no significant differences in the inferred pseudohaplotype distribution between the groups when only the females were analysed. However the frequencies were similar to those observed in males (Figure 3.5C).

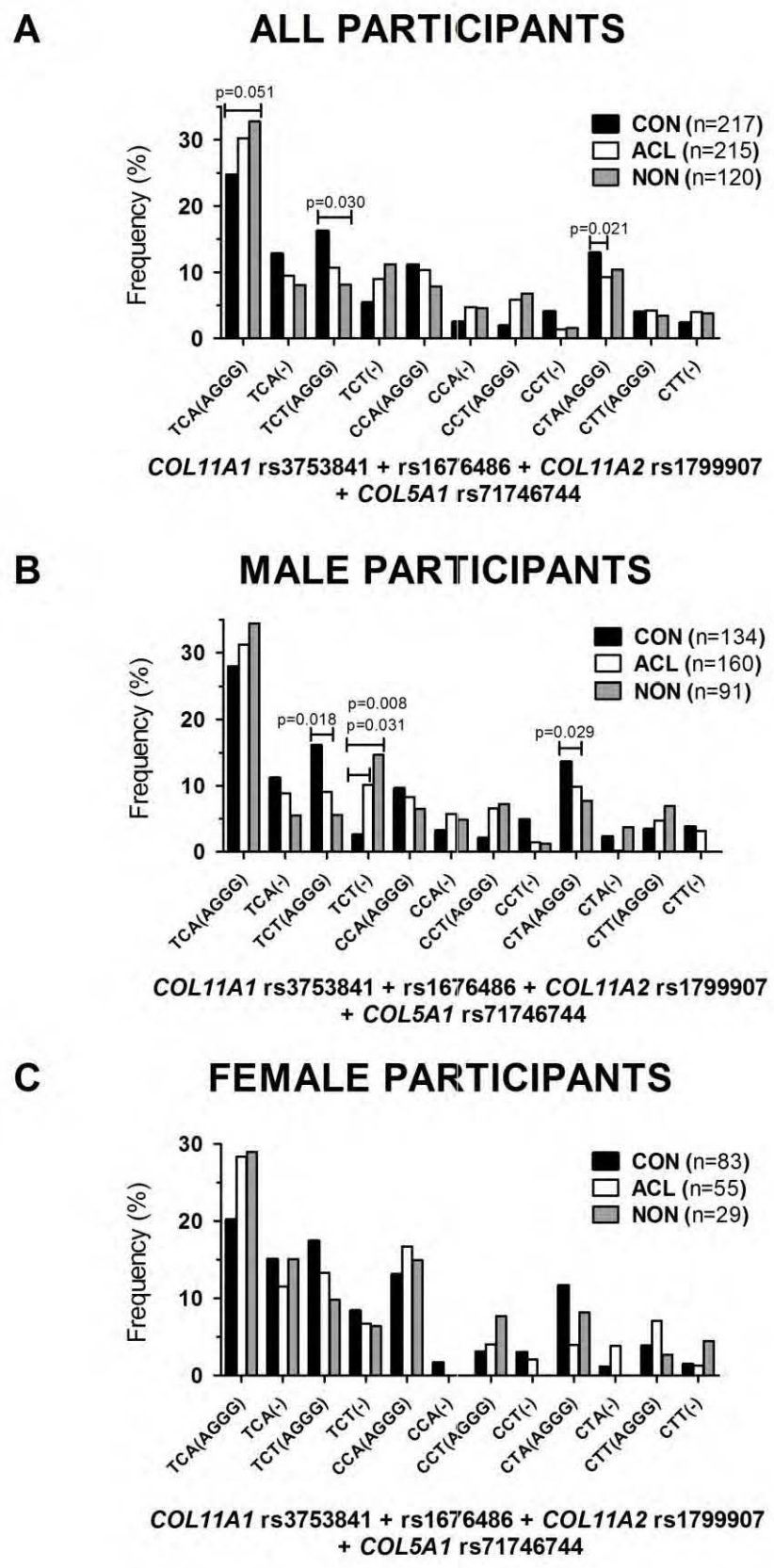


Figure 3.5 Inferred pseudohaplotype frequencies for *COL11A1* rs3753841 and rs1676486, *COL11A2* rs1799907 and *COL5A1* rs71746744 in (**A**) all (**B**) male and (**C**) female participants. Asymptomatic controls (CON) are depicted by solid black bars, anterior cruciate ligament (ACL) rupture groups are depicted by white bars and the non-contact (NON) subgroup is depicted by solid grey bars. Any significant differences ($p < 0.05$) between groups are indicated by lines on the graphs.

3.4 Discussion

Familial heritability studies have suggested that genetic factors play an important role in susceptibility to ACL rupture[90,92]. The rs12722 (C/T) SNP in the *COL5A1* gene, which encodes the α1(V) chain of type V collagen, has previously been associated with ACL ruptures in females[103]. Although there were trends in the females, no significant association was found between the rs71746744 (AGGG/-) STRP or rs1134170 (T/A) SNP and ACL rupture (Chapter 2), despite their close proximity to the rs12722 SNP in the *COL5A1* 3'-UTR.

Type V and type XI collagen are structurally and functionally similar minor fibrillar collagens[28] that initiate collagen fibril nucleation during early fibrillogenesis and regulate fibril diameter[23,26,28,42,43]. The major type V collagen isoform [$\alpha1(V)_2\alpha2(V)$] is encoded by the *COL5A1* and *COL5A2* genes respectively, and type XI collagen is encoded by the *COL11A1*, *COL11A2* and *COL2A1* genes[28]. Until recently, the expression of these two proteins in tissues was thought to be mutually exclusive; with type V collagen expressed with type I collagen in tendon, ligament and skin[28], and type XI collagen expressed with type II collagen in cartilage[29,43]. However, evidence now suggests that these collagens have overlapping expression during the development of some tissues, and that the α1 chains are interchangeable and capable of forming heterotrimers in some cases[147–149,194]. Type XI collagen is expressed in addition to type V collagen at the bone-ligament interface of the ACL[30]. The *COL11A1* rs3753841 and rs1676486 and *COL11A2* rs1799907 polymorphisms were therefore investigated for an association with ACL rupture. In addition, the previously investigated rs71746744 AGGG/- STRP in the *COL5A1* 3'-UTR (Chapter 2) was investigated for an interaction effect with *COL11A1* and *COL11A2*.

The first main finding of this study was that there was a significant difference in the genotype frequency distribution of the *COL11A1* rs3753841 (T/C) polymorphism between asymptomatic controls (CON) and individuals with ACL rupture. Specifically, heterozygosity for the rs3753841 T/C polymorphism was associated with protection (p=0.009, OR=1.67, 95% CI= 1.14-2.45). The association of genetic heterozygosity with protection from disease phenotypes is not unprecedented, although it may be more difficult to interpret biologically[195]. There is no obvious biological explanation for the association of the TC genotype of rs3753841 with ACL injuries. This association might merely indicate other variants within the *COL11A1* gene and chromosomal region may be associated with ACL ruptures. Further work is required to investigate this possibility. The *COL11A1* rs1676486 and *COL11A2* rs1799907 polymorphisms were however not associated with ACL rupture.

The genotype and allele frequencies for the ACL and CON groups for all three SNPs were similar to the genotype and allele frequencies reported for a European population in the National Centre for Biotechnology Information (NCBI) database (http://www.ncbi.nlm.nih.gov/snp/). There was a departure from HWE for the *COL11A1* rs3753841 and rs1676486 markers in the ACL and CON groups. Departures from HWE suggest that perhaps the sample size of this study is too small or the individuals investigated in this study do not represent a randomly mating population[185]. The ACL group was selected based on clinical criteria and geographical ancestry and the CON group was selected to match the age, geographical ancestry and physical activity of the ACL group. The observed departure from HWE may thus be explained by the fact that the recruitment of the ACL and CON groups was highly selective.

The second major finding of this study was that the *COL11A1* CT inferred haplotype, constructed from the non-synonymous rs3753841 and rs1676486 SNPs, was under-represented in the ACL group, suggesting a protective effect. As previously mentioned the *COL11A1* rs3753841 (T/C) SNP results in an amino acid change from a leucine to a proline at position 1323 (L1323P) of the α1(XI) chain that is predicted to be deleterious, while the rs1676486 SNP results in a change from proline to a serine at position 1535 (P1535S) that is very near to a predicted disulphide bond at amino acid position 1532. The resultant α1 (XI) chain from the CT haplotype therefore differs in structure in at least two amino acid residues from the wild type TC allele. This change in the amino acid sequence of the α(I)XI chain may have an effect on the structure and function of type XI collagen. Further work is required to test this hypothesis. *COL11A1* mRNA stability has also been shown to be altered by the rs1676486 SNP[129]. The possible role of this SNP in regulating type XI collagen production will be discussed in greater detail in the following chapters of this dissertation.

In order to investigate the interaction between the two *COL11A1* SNPs and the rs1799907 SNP in *COL11A2*, Inferred pseudohaplotypes were constructed. The *COL11A1-COL11A2* CTA inferred pseudohaplotype was significantly over-represented in the CON group compared to the ACL group. Therefore, in combination, the C, T and A alleles of rs3753841, rs1676486 and rs1799907 respectively, appear to be protective. The *COL11A2* rs1799907 (T/A) SNP occurs in intron 5, and results in alternative splicing of the *COL11A2* gene, resulting in protein products of slightly different length. In the conventional type XI isoform, [α1(XI) α2(XI) α1(II)], the α1(XI) and α2(XI) chains combine in a 1:1 ratio. It is possible that the alternatively spliced variant from either the A or the T allele of rs1799907 interacts better with either the C or T allele of rs3753841[137].

The third major finding of this study was that the rs71746744 STRP in *COL5A1* does not appear to interact with the rs3753841 and rs1676486 SNPs in *COL11A1* and the rs1799907 SNP in *COL11A2* to

modulate ACL rupture risk in the hypothesised manner. A complex picture emerged when the inferred pseudohaplotypes were constructed from *COL11A1* rs3753841 and rs1676486, *COL11A2* rs1799907 and *COL5A1* rs71746744. Three major inferred pseudohaplotype, namely TCT(-), CTA(AGGG) and CTA(AGGG), were significantly different between the CON and ACL groups and NON subgroups in the combined and male groups. The biological significance of these associations is not apparently clear. It is also possible that these are merely false positive findings due to the relative small sample size of this study. Further work in larger cohorts is required to investigate the association of these inferred haplotypes with ACL ruptures.

Although there was no evidence for any sex-specific effects of the type XI collagen genes in modulating the risk of ACL ruptures in this study, the small female sample size is a limitation of the current study. In particular this study was under powered to conclusively determine whether there were any type XI and type V collagen gene interactions in the modulation of ACL ruptures in females.

In conclusion, the functional *COL11A1* rs3753841 and rs1676486 polymorphisms, as well as the *COL11A2* rs1799907 polymorphism appear to play a role in susceptibility to ACL rupture. Specifically, the CTA inferred pseudohaplotype constructed from these SNPs appears to be protective; however the mechanism of this interaction is not understood and requires further investigation.

Chapter 4: The *COL11A1* and *COL11A2* Genes and Chronic Achilles Tendinopathy

4.1 Introduction

The Achilles tendon is the largest and strongest tendon within the human body and originates at the gastrocnemius and soleus muscles of the calf, and inserts at the calcaneus bone (heel) [1]. There are a number of injuries that affect the Achilles tendon or surrounding tissues, such as bursitis, achillodinia, peritendinitis, tendinosis, tendinitis and rupture[196]. The focus of this specific study is on injuries within the midsubstance of the tendon structure, specifically chronically painful overuse Achilles tendon injuries. Since all the participants included in this dissertation with a painful overuse Achilles tendon injury were diagnosed clinically, the term chronic Achilles tendinopathy (AT) will be utilized, as it makes no assumptions about the underlying pathology.

AT constitutes between 6 to 18% of reported tendon injuries[79] and is among the most common sporting injuries, especially within a growing and active generation of older individuals[5]. In fact, the prevalence of AT and Achilles tendon ruptures appears to be increasing worldwide[197,198]. With time, AT often worsens, and involvement of the previously healthy contralateral tendon is common[199]. Unfortunately, AT can become a chronic condition, requiring management, and in severe cases requiring surgery[200–202].

Although Achilles tendinopathies are well studied at the clinical level, they remain poorly understood at the molecular level[14,15,78,200,203,204]. Histopathological findings typically reveal a degenerative rather than inflammatory process[77,205,206], and the tendinopathic tissue is often characterized by collagen fibril degeneration and microtears, decreased fibril diameter, tenoctye dedifferentiation, hypo-or acellularity and increased glycoaminoglycan content[77,205].

AT is a multifactorial disorder, resulting from the numerous interactions of extrinsic and intrinsic risk factors (Chapter 1, Section 1.5.1). The genetic profile of an individual has recently been identified as one of the important intrinsic risk factors[4,5,93,95,167]. Previous studies have reported independent associations of several genes with AT (Chapter 1, Table 1.1)[5]. The genetic predisposition to AT is therefore polygenic, resulting from the interaction of risk alleles in a number of genes that make small cumulative contributions to the overall risk[3,4,95]. One of the most extensively investigated genes associated with AT is *COL5A1*[101,102,104], the gene encoding the α1 chain of type V collagen. The

COL5A1 gene is also associated with ACL rupture in females[103] and other exercise-associated phenotypes[115–117,178].

Type V collagen plays an important role in collagen fibril nucleation, and fibril diameter regulation during fibrillogenesis (Chapter 1, Section 1.3.2.1). As mentioned in Chapter 2, there are seven tightly linked polymorphisms in the *COL5A1* 3'-UTR; five of which (rs12722 (C/T), rs3196378 (C/A), rs71746744 (-/AGGG), rs16399 (ATCT/-) and rs1134170 (A/T)) are thought to alter the secondary structure of the mRNA transcript[104], and cause the observed changes in mRNA stability between tendinopathic and asymptomatic individuals[118]. The increase in *COL5A1* mRNA stability in tendinopathic individuals may increase the amount of α1(V) protein that gets translated, thus decreasing the fibril diameter, and potentially altering biomechanical properties of the tissue[110].

As discussed in the previous chapter (Chapter 3), type XI collagen shares structural homology with type V collagen[28], which together with types II and IX collagen forms heterotypic fibrils in cartilage [29,44]. Like type V collagen, type XI collagen is also a vital collagen fibril nucleator and regulator of collagen fibril diameter[42,43]. Recently, a synergistic relationship between *COL5A1* and *COL11A1* was demonstrated in developing tendon in mice[146], and the expression of *COL11A1* has been detected in developing human tendon [48,145].

In the previous chapter (Chapter 3), two non-synonymous SNPs (rs3753841 T/C and rs1676486 C/T) in the *COL11A1* gene and an intronic SNP (rs1799907 A/T) in the *COL11A2* gene were investigated for an association with ACL rupture. Heterozygosity for the rs3753841 (T/C) polymorphism was associated with protection from ACL rupture. In addition, the CTA inferred pseudohaplotype consisting of the C, T, and A alleles of rs3753841, rs1676486 and rs1799907 respectively was associated with protection from ACL rupture. Although the *COL11A1* and *COL11A2* polymorphisms were investigated for interactions with the rs71746744 -/AGGG STRP in the 3'-UTR of the *COL5A1* gene, the findings were inconclusive due to the small female sample size of the study (Chapter 3).

The aim of this study was to investigate whether the rs3753841 (T/C) and rs1676486 (C/T) polymorphisms in *COL11A1* and the rs1799907 (A/T) polymorphism in *COL11A2* are associated with AT in a South African and Australian population. Due to the potential for the functional *COL11A1*, *COL11A2* and *COL5A1* SNPs to interact during tendon fibrillogenesis, a secondary aim of this study was to investigate whether the *COL11A1* and *COL11A2* variants interact with the rs71746744 (-/AGGGG) polymorphism in *COL5A1* to modulate the risk of AT.

4.2 Methods

4.2.1 Study Design

A case-control genetic association study was conducted on individuals of self-reported European ancestry recruited from South Africa (SA) and Australia (AUS). This study forms part of a broader study to determine the genetic risk factors that contribute to AT. Approval for the study was obtained from the Research Ethics Committee of the Faculty of Health Sciences within the University of Cape Town (reference number 172/2005) (Appendix B, B2.1) and Human Ethics Committees of La Trobe and Deakin Universities, Melbourne, Australia.

4.2.2 Participants

South African Participants

A total of 264 unrelated, physically active participants from South Africa, consisting of 160 asymptomatic controls (SA CON group) and 104 individuals diagnosed with chronic Achilles tendinopathy (SA TEN group) were included in this study. The participants with clinically diagnosed Achilles tendinopathy were recruited from the Sports Medicine practice at the Sports Science Institute of South Africa and other collaborating clinical practices in Cape Town and Johannesburg, South Africa. The SA TEN diagnoses were reviewed and confirmed by a Sports Physician, using an inclusion and exclusion criteria and a checklist (Appendix B, B2.2). The SA CON participants reported no history of tendon pathology and were recruited from various recreational sporting clubs.

The SA CON group were recruited such that the sex and country of birth were similar to that of the SA TEN group. In addition, the age of the SA CON group was matched to that of the age of initial injury onset in the SA TEN group in order to avoid the possible confounding effects of age. As previously noted of this population, the majority of participants in the TEN group had participated in long-distance running[100]. Therefore the CON group was matched for the number of years of participation in running activities. Although the TEN participants completed significantly less running training (hr/week) than the CON group during the past 2 years, the TEN group had participated for significantly more years in high-impact sports. There was no significant difference in the hours of training in high-impact sports during the past 2 years between the two groups.

Australian Participants

A total of 254 participants from Australia, consisting of 176 asymptomatic controls (AUS CON group) and 78 patients with chronic Achilles tendinopathy (AUS TEN group) were included in this study. The tendinopathy participants were recruited from the Musculoskeletal Research Centre at La Trobe

University in Melbourne Australia[102]. A Sports Physiotherapist confirmed the AUS TEN diagnoses, using criteria similar to that of the SA study. The participants were recruited such that their country of birth was matched between the AUS CON and AUS TEN groups, and the age of recruitment of the AUS CON group matched that of initial onset of injury of the AUS TEN group.

In accordance with the Declaration of Helsinki, all participants signed informed consent forms (Appendix B, B2.3). SA participants were given information regarding the study (Appendix B, B2.4), gave their written informed consent and completed questionnaires concerning their medical history and involvement in physical activity (Appendix B, B2.5). AUS participants received similar forms and information to those used in the SA study, however unfortunately the physical activity of the AUS participants was not documented.

4.2.3 DNA Extraction

For the SA participants, approximately 4.5 ml of venous blood was collected via venipuncture of a forearm vein into EDTA vacutainer tubes and stored at 4 °C until total DNA extraction. Total DNA was extracted from blood using a method described by Lahiri and Nurnberger (1991)[179] and modified by Mokone et al[100] (See Appendix C1 for details). The DNA was stored at -20 °C until PCR analysis. For the AUS participants, approximately 4.5 ml of venous blood was collected and DNA was extracted using a sequence extraction technique (FlexiGene DNA Kit, Qiagen P/L, Valencia, California, USA) as per the manufacturer's instructions.

4.2.4 Genotyping

The rs3753841 (T/C) and rs1676486 (C/T) polymorphisms in *COL11A1* (Chapter 3, Figure 3.1) and the rs1799907 (A/T) polymorphism in *COL11A2* (Chapter 3, Figure 3.2) were genotyped for this study. The rs71746744 (-/AGGG) STRP (Chapter 2, Figure 2.1) had previously been genotyped, and was included in this study for interaction effects. Genotyping was performed using fluorescence-based Taqman PCR assays (Applied Biosystems, Foster City, CA, USA). Allele specific probes and flanking primer sets (sequences available from manufacturer on request) were used along with a pre-made PCR mastermix containing ampliTaq® DNA polymerase Gold (Applied Biosystems, Foster City, CA, USA) in a final reaction volume of 8 µl. The PCR reactions were run on StepOnePlus qRT-PCR machine (Applied Biosystems) using the manufacturers recommended cycling conditions (previously described in Section 2.2.4). Genotypes were determined by end-point fluorescence and analyzed using the StepOne Software Version 2.2.2 (Applied Biosystems) (example provided in Figure 2.2). A small number of samples failed to amplify (Chapter 2, Section 2.2.5).

4.2.5 Statistical Analysis

Quanto version 1.2.4. was used to determine the statistical power of the study for a given sample size, allele frequency and odds ratio[180]. Assuming a minimum allele frequency of 0.42, 0.22 and 0.31 in the pooled CON group for rs3753841, rs1676486 and rs1799907 respectively, a sample size of approximately 150 cases would be adequate to detect a allelic odds ratio of at least 1.6, 1.7 and 1.6 respectively, at a power of 80% and a significance level of 5%. GenePop was used to determine whether the SNPs were in HWE (http://genepop.curtin.edu.au/genepop_op1.html). Statistical analysis was conducted using STATISTICA 10 (StatSoft Inc., Tulsa, OK, USA). Continuous variables, such as weight, age, height and BMI, were compared using one-way ANOVA, co-varying for sex and age of recruitment where required. Pearson's chi-squared tests, or Fischer's exact tests when required, were used to determine whether there were any significant differences in allele frequencies or other categorical variables between the CON and TEN groups. LD between the *COL11A1* rs3753841 and rs1676486 SNPs was calculated using Cubex software (http://www.oege.org/cgi-bin/cubex.py)[183]. Combined genotype frequencies were analyzed using the Monte Carlo test (CLUMP program, version 2.0)[207].The odds ratios and 95% confidence intervals were determined using GraphPad Prism version 5.0d (GraphPad Software, San Diego, California, USA). Chaplin case-control haplotype inference package (Epstein Software) was used to infer haplotypes (and pseudohaplotypes)[181,182]. In all cases, statistical significance was accepted when $p<0.05$. The results were not adjusted for multiple testing, as no obvious appropriate method exists to date.

4.3 Results

4.3.1 Participant Characteristics

The AUS CON and AUS TEN groups were similarly matched for age (p=0.487), and country of birth (p=0.792) (Table 4.1). There was a significant difference in the sex composition of the AUS CON (40.3% males) and AUS TEN (71.8% males) groups, with a larger proportion of females in the AUS CON group (p=0.010). After adjusting for sex, the groups were matched for height (p=0.101). The mean difference between age of injury and age of recruitment of the AUS TEN group was 9.0 ± 10.1 years. There was therefore a significant difference between the age at recruitment of the AUS TEN (49.6 ± 12.8 years) and the AUS CON (39.4 ± 12.3 years) groups (p<0.001). After adjusting for both age at recruitment and sex, the AUS CON and AUS TEN groups were matched for weight (p=0.343) and BMI (p=0.384) (Table 4.1).

The SA CON and SA TEN groups were matched for height (p=0.395), sex (p=0.553) and country of birth (p=0.941) (Table 4.1). However, there were significant differences in age (p=0.007), as well as weight (p=0.003) and BMI (p=0.003), with the SA TEN group being older and heavier than the SA CON group. The mean difference between age of injury and age of recruitment of the SA TEN group was 8.3 ± 9.8 years. There was therefore also a significant difference between the age at recruitment of the SA TEN (48.4 ± 11.4 years) and the SA CON (36.4 ± 10.8 years) groups (p<0.001). When weight and BMI were adjusted for age of recruitment, weight (p=0.150) and BMI (p=0.125) were no longer significantly different. There were no significant genotype effects on participant characteristics (Appendix A, Table A3.1.).

Table 4.1 Characteristics of the AUS TEN and SA TEN groups and their respective AUS CON and SA CON groups

	AUS CON	AUS TEN	p-value	Covaried p-value	SA CON	SA TEN	p-value	Covaried p-value
Age (years)	39.4±12.3 (174)	40.7±14.5 (77)	0.487		36.4±10.8 (154)	40.9±14.8 (92)	**0.007**	
Sex (%male)	40.3 (176)	71.8 (78)	**0.010**		63.8 (160)	67.3 (104)	0.553	
Height (cm)	171.5±9.3 (175)	174.0±10.3 (75)	0.059	0.101 ‡	175.0±9.1 (155)	176.0 ±8.9 (91)	0.395	
Weight (kg)	72.7±14.4 (176)	79.7±13.4 (78)	**<0.001**	0.343 §	72.2±11.9 (159)	77.1±13.4 (96)	**0.003**	0.150 #
BMI (kg/cm^2)	24.7±3.9 (175)	26.2±3.5 (75)	**0.003**	0.384 §	23.6±2.8 (151)	24.8±3.3 (81)	**0.003**	0.125 #
Country of birth (% Australia)	80.9 (173)	76.6 (77)	0.792		ND	ND	ND	
Country of birth (% South Africa)	ND	ND	ND		73.6 (159)	74.0 (100)	0.941	

Participants were included in the analysis if they were genotyped at least once for rs3753841, rs1676486 or rs1799907. Variables are expressed as mean ± standard deviation, except for sex and country of birth, which are represented as a percentage (%). Number of participants for which data was available is in parenthesis (n). Age of CON groups = age of recruitment; age of TEN groups = age of initial onset of symptoms. Weight, height and BMI values are those reported at recruitment. Significant differences ($p<0.05$) are highlighted in bold typeset. BMI, body mass index, ND, not determined. ‡ Covaried for sex. # Covaried for age of recruitment. § Covaried for age of recruitment and sex.

4.3.2 Genotype and Allele Frequency Distributions

4.3.2.1 *COL11A1* rs3753841 (T/C)

There were no significant differences in genotype (p=0.235) or allele (p=0.429) frequency distribution between AUS TEN (n=56) and AUS CON (n=104) participants for *COL11A1* rs3753841 (Table 4.2). Similarly, there were no significant differences between SA CON (n=150) and SA TEN (96) groups for genotype (p=0.321) and allele (p=0.143) frequency distributions (Table 4.2). Since the genotype frequency distributions for both cohorts were similar (p=0.226), they were combined for further analysis (CON, n=254; TEN, n=152). As illustrated in Figure 4.1A, there was no significant difference in the genotype (p=0.102) frequency distribution between CON and TEN groups of the combined AUS and SA cohorts. There was however a tendency (p=0.087) for the T- allele to be over-represented in the TEN group (Figure 4.1B). Except for the AUS TEN (p=050) group, all the other groups were in HWE.

Table 4.2 Genotype and allele frequency distribution of the functional SNPs rs3753841 and rs1676486 in *COL11A1* and the rs1799907 SNP in *COL11A2*, in the AUS TEN, AUS CON, SA TEN and SA CON groups.

		rs3753841 genotype					rs3753841 allele		
	N	**TT**	**TC**	**CC**	**P-value**	**HWE**	**T-allele**	**C-allele**	**P-value**
Australia	160								
AUS CON	104	29.8 (31)	49.0 (51)	21.2 (22)		0.904	54.3 (113)	45.7 (95)	
AUS TEN	56	41.1 (23)	35.7 (20)	23.2 (13)	0.235	**0.050**	58.9 (66)	41.1 (46)	0.429
South Africa	246								
SA CON	150	37.3 (56)	46.7 (70)	16.0 (24)		0.786	60.7 (182)	39.3 (118)	
SA TEN	96	46.9 (45)	40.6 (39)	12.5 (12)	0.321	0.441	67.2 (129)	32.8 (63)	0.143

		rs1676486 genotype					rs1676486 allele		
	N	**CC**	**CT**	**TT**	**P-value**	**HWE**	**C-allele**	**T- allele**	**P-value**
Australia	205								
AUS CON	138	62.3 (86)	31.2 (43)	6.5 (9)		0.264	77.9 (215)	22.1 (61)	
AUS TEN	67	67.2 (45)	29.8 (20)	3.0 (2)	0.537	0.901	82.1 (110)	17.9 (24)	0.326
South Africa	229								
SA CON	139	64.0 (89)	29.5 (41)	6.5 (9)		0.165	78.8 (219)	21.2 (59)	
SA TEN	90	68.9 (62)	26.7 (24)	4.4 (4)	0.688	0.405	82.2 (148)	17.8 (32)	0.367

		rs1799907 genotype					rs1799907 allele		
	N	**AA**	**AT**	**TT**	**P-value**	**HWE**	**A-allele**	**T-allele**	**P-value**
Australia	192								
AUS CON	126	52.4 (66)	34.9 (44)	12.7 (16)		0.055	69.8 (176)	30.2 (76)	
SA TEN	66	39.4 (26)	45.4 (30)	15.2 (10)	0.227	0.782	62.1 (82)	37.9 (50)	0.126
South Africa	261								
SA CON	158	46.8 (74)	41.8 (66)	11.4 (18)		0.576	67.7 (214)	32.3 (102)	
SA TEN	103	40.8 (42)	46.6 (48)	12.6 (13)	0.629	0.901	64.1 (132)	35.9 (74)	0.389

Values are percentages (%), with sample number of participants (n) displayed in parenthesis. N is the total number of successfully genotyped participants in each group. HWE is the p-value for the exact tests for HWE. Significant values, ($p<0.05$) are shown in bold typeface.

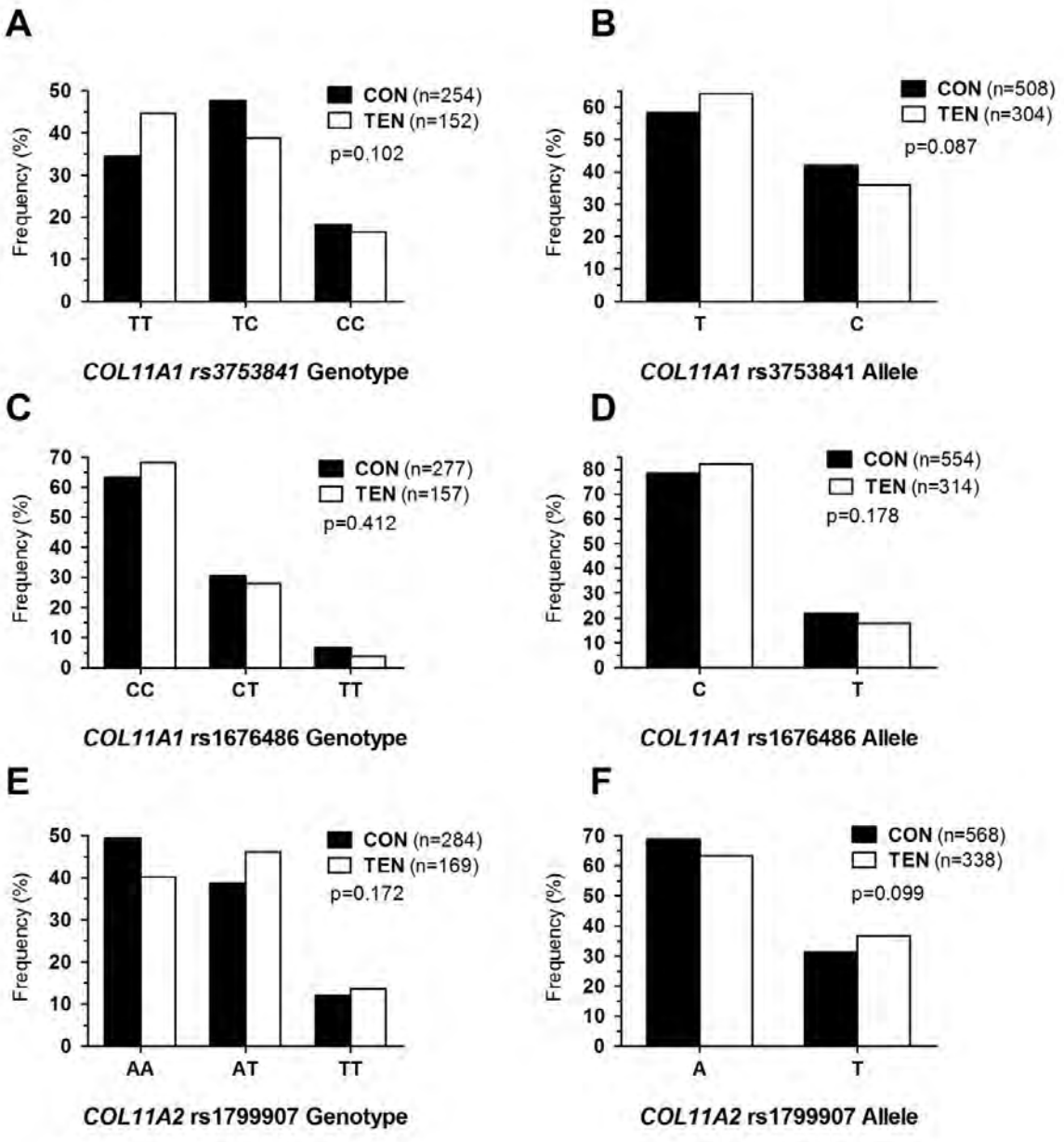


Figure 4.1 Genotype and allele frequency distributions of the combined Australian and South African CON and TEN participants for *COL11A1* rs3753841 (T/C), *COL11A1* rs1676486 (C/T) and *COL11A2* rs1799907 (A/T). Frequencies are expressed as percentages. The CON group is represented by the black bars, and the TEN groups are represented by white bars. The number of participants (n) is in parenthesis, followed by the p-value. (**A**) rs3753841 genotype distribution, CON, HWE=0.729, TEN, HWE= 0.054. (**B**) rs3753841 allele frequency distribution (**C**) rs1676486 genotype distribution, CON, HWE=0.076, TEN HWE=0.584. (**D**) rs1676486 allele frequency distribution, (**E**) rs1799907 genotype frequency distribution CON, HWE=0.092, TEN HWE=0.933) and (**F**) rs1799907 allele frequency distribution.

4.3.2.2 *COL11A1* rs1676486 (C/T)

There were no significant differences in genotype (p=0.537) or allele (p=0.326) frequency distribution between AUS TEN (n=67) and AUS CON (n=138) participants for *COL11A1* rs1676486 (Table 4.2). Similarly, there were no significant differences between SA CON (139) and SA TEN (90) groups for genotype (p=0.688) and allele (p=0.367) frequency distributions (Table 4.2). Since the genotype frequency distributions for both cohorts were similar (p=0.943), they were combined for further analysis (CON, n=277; TEN, n=157). There were no significant differences between the genotype (p=0.412) and allele (p=0.178) frequency distributions of the CON and TEN groups when the AUS and SA cohorts were combined (Figure 4.1C and Figure 4.1D). All groups were in HWE.

4.3.2.3 *COL11A2* rs1799907 (A/T)

There were no significant differences in genotype (p=0.227) or allele (p=0.126) frequency distribution for *COL11A2* rs1799907 between AUS TEN (n=66) and AUS CON (n=126) participants (Table 4.2). Similarly there were no significant differences in genotype (p=0.629) or allele (p=0.389) frequency between SA CON (n=158) and SA TEN (n=103) groups (Table 4.2). Since the genotype frequency distributions for both cohorts were similar (p=0.327), they were combined for further analysis (CON, n=284; TEN, n=169). There was no significant difference between the genotype (p=0.172) frequency distribution between CON and TEN groups when the AUS and SA cohorts were combined (Figure 4.1E). There was however a tendency (p=0.099) for the minor T allele to be over-represented in the TEN group (Figure 4.1F). All groups were in HWE.

4.3.3 Genotype Interactions Between *COL11A1* and *COL11A2*

LD between the *COL11A1* rs3753841 and rs1676486 polymorphisms was similar for the CON group (D' =0.903; r^2 =0.342), TEN group (D'=0.962, r^2=0.328) and combined groups (D'=0.923, r^2=0.338) and was shown to be tightly linked. Since the three SNPs investigated in this study could theoretically interact with one another to regulate or affect type XI collagen production, gene-gene interactions were investigated (Table 4.3).

There was no significant difference (Monte Carlo test) between genotype frequency distributions when the combined *COL11A1* rs3753841 and rs1676486 SNPs for the pooled South African and Australian TEN and CON groups were analysed. There was however a tendency towards significance in the *COL11A1* rs3753841 and *COL11A2* rs1799907 interaction (T1, p=0.051) and *COL11A1* rs1676486 and *COL11A2* rs1799907 interaction (T1, p=0.060), therefore specific biologically meaningful genotype interactions were analysed further. The TT-TT genotype combination of rs3753841 and rs17999079 was significantly over-represented in the TEN group (9.3%) compared to

the CON group (3.2%) (p=0.019, OR=3.1, 95% CI=1.2-7.9). The interaction of the *COL11A1* rs1676486 CC genotype together with a *COL11A2* rs1799907 AT or TT genotype was also significantly over-represented in the TEN group (47.9%) compared to the CON group (31.7%) (p=0.002, OR= 2.0, 95% CI= 1.3-3.1).

Table 4.3 Two-way combined genotype frequencies distributions for *COL11A1* rs3753841, rs1676486 and *COL11A2* rs1799907

***COL11A1* rs3753841 and *COL11A1* rs1676486**				
rs3753841	**rs1676486**	**CON (**n=233**)**	**TEN (**n=141**)**	**TEN/CON**
TT	**CC**	33.5 (78)	42.6 (60)	1.27
TT	**CT**	1.3 (3)	0.7 (1)	0.55
TT	**TT**	0.0 (0)	0.0 (0)	1.00
CT	**CC**	23.9 (54)	23.4 (33)	0.98
CT	**CT**	22.8 (53)	15.6 (22)	0.69
CT	**TT**	0.9 (2)	0.0 (0)	ND
CC	**CC**	4.3 (10)	2.8 (4)	0.66
CC	**CT**	7.7 (18)	11.4 (16)	1.47
CC	**TT**	6.4 (15)	3.6 (5)	0.55
***COL11A1* rs3753841 and *COL11A2* rs1799907**				
rs3753841	**rs1799907**	**CON (**n=216 **)**	**TEN (**n=140**)**	**TEN/CON**
TT	**AA**	17.1 (37)	17.1 (24)	1.00
TT	**AT**	15.7 (34)	20.0 (28)	0.76
TT	**TT**	3.2 (7)	9.3 (13)	**2.87**[a]
CT	**AA**	24.1 (52)	13.6 (19)	0.56
CT	**AT**	16.7 (36)	20.7 (29)	1.24
CT	**TT**	6.0 (13)	2.9 (4)	0.48
CC	**AA**	6.5 (14)	8.6 (12)	1.32
CC	**AT**	7.9 (17)	6.4 (9)	0.82
CC	**TT**	2.8 (6)	1.4 (2)	0.51
***COL11A1* rs1676486 and *COL11A2* rs1799907**				
rs1676486	**rs1799907**	**CON (**n=230**)**	**TEN (**n=144**)**	**TEN/CON**
CC	**AA**	31.7 (73)	22.2 (32)	0.70
CC	**AT**	25.2 (58)	36.1 (52)	**1.43**[b]
CC	**TT**	6.5 (15)	11.8 (17)	**1.81**[b]
CT	**AA**	15.7 (36)	14.6 (21)	0.93
CT	**AT**	10.0 (23)	9.7 (14)	0.97
CT	**TT**	4.8 (11)	2.1 (3)	0.44
TT	**AA**	1.3 (3)	1.4 (2)	0.77
TT	**AT**	3.9 (9)	1.4 (2)	*0.36
TT	**TT**	0.9 (2)	0.7 (0)	0.79

MonteCarlo tests performed with 2000 simulations.
Monte Carlo Test: rs3753841 and rs1676486, P-values for: T1=0.226, T2=0.145, T3=0.078, T4=0.141
Monte Carlo Test: rs3753841 and rs1799907, P-values for: T1=0.051, T2=**0.038**, T3=**0.015**, T4=0.073
Monte Carlo Test: rs1676486 and rs1799907, P-values for: T1=0.060, T2=**0.046**, T3=**0.022**, T4=0.052
ND, no data. Significant associations (p<0.05) are indicated in bold typeset.

[a] p=0.019, OR=3.1, 95 % CI=1.2-7.9

[b] p=0.002, OR= 2.0, 95 % CI= 1.3-3.1

4.3.4 Inferred Haplotype and Pseudohaplotypes Analysis of *COL11A1* and *COL11A2*

Since significant gene-gene interactions were observed between the *COL11A1* and *COL11A2* genes in the previous section (Section 4.3.3), inferred haplotypes and pseudohaplotypes were constructed between the SNPs to confirm these findings (Appendix A, Table A3.2).

4.3.4.1 *COL11A1* rs3753841 and *COL11A1* rs1676486

Three haplotypes (TC, CC and CT), of the possible four, were inferred with a frequency greater than 1.2% from the *COL11A1* SNPs rs3753841 and rs1676486 (Figure 4.2A). The major *COL11A1* TC inferred haplotype was significantly over-represented in the TEN group (63.3%, n=106) compared to the CON group (57.8%, n=172) (p=0.029).

4.3.4.2 *COL11A1* rs3753841 and *COL11A2* rs1799907

All four possible pseudohaplotypes were inferred with a frequency greater than 1% from the *COL11A1* rs3753841 and *COL11A2* rs1799907 SNPs (Figure 4.2B). In agreement with the interactions reported in Table 4.3, the TT inferred pseudohaplotype was significantly over-represented in the TEN group (26.9%, n=48) compared to the CON group (16.9%, n=84) (p=0.005).

4.3.4.3 *COL11A1* rs1676486 and *COL11A2* rs1799907

Similarly all four possible pseudohaplotypes were inferred with a frequency greater than 1% from the *COL11A1* rs1676486 and *COL11A2* rs1799907 SNPs (Figure 4.2C). In agreement with the interactions reported in Table 4.3, the CT inferred pseudohaplotype was significantly over-represented in the TEN group (32.5%, n=59) compared to the CON group (23.1%, n=76) (p=0.010). The CA, and TT inferred pseudohaplotype were on the other-hand under-represented in the TEN group (CA, 49.5%, n=90; TT, 4.0%, n=7) compared to the CON group (CA, 55.2%, n=182; TT, 8.2%, n=27) (CA, p=0.038; TT, p=0.027).

4.3.4.4 *COL11A1* rs3753841 and *COL11A1* rs1676486 and *COL11A2* rs1799907

Six pseudohaplotypes (TCA, TCT, CCA, CCA, CTT, CTA and CTT), of the possible eight, were inferred with a frequency greater than 1.2% from the two *COL11A1* and one *COL11A2* SNPs (Figure 4.2D). As expected from the top three panels of Figure 4.2, (TC_, T_T and _CT), The TCT pseudohaplotype was significantly (p=0.005) over-represented in the TEN group (26.2%, n=47) compared to the CON group (17.2%, n=58).

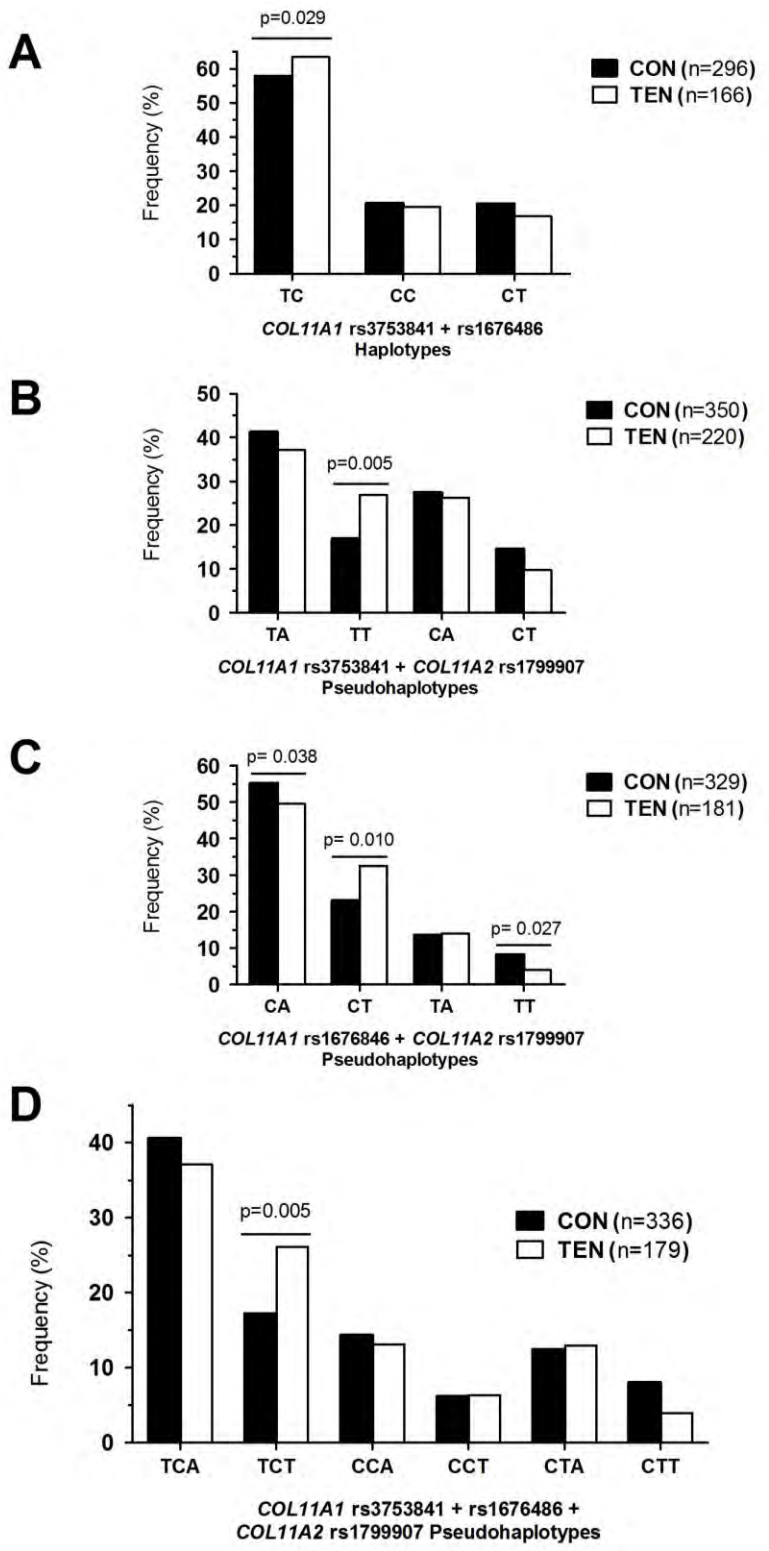


Figure 4.2 Inferred haplotypes and pseudohaplotypes for *COL11A1* rs3753841 and rs1676486 and *COL11A2* rs1799907. The AUS and SA Achilles tendinopathy (TEN, clear bars) and control (CON, black solid bars) participants were pooled for this analysis. **A**) *COL11A1* rs3753841 and rs1676486, **B**) *COL11A1* rs3753841 and *COL11A2* rs1799907, **C**) *COL11A1* rs1676486 and *COL11A2* rs1799907, **D**) *COL11A1* rs3753841 and rs1676486 and *COL11A2* rs1799907. Significant differences (P<0.05) are indicated with a line on the graph.

4.3.5 Genotype Interactions Between *COL11A1*, *COL11A2* and *COL5A1*

Since type V and type XI collagen both regulate fibrillogenesis, the *COL11A1* and *COL11A2* SNPs investigated in this study were investigated for gene-gene interactions with the previously investigated *COL5A1* rs71746744 (AGGG/-) STRP[104] (Table 4.4). There were significant differences (Monte Carlo test) in two-way genotype frequency distributions between *COL5A1* rs71746744 and *COL11A1* rs3753841 (T1, p=0.001), and *COL5A1* rs71746744 and rs1676486 (T1, p=0.002) when the pooled Australian and South African TEN and CON groups were analysed. Interestingly there was no significant difference in two-way genotype frequency for *COL5A1* rs71746744 and *COL11A2* rs1799907 between the TEN and CON groups (T1, p=0.291).

Since the previous suggestions of the chapter suggested that the *COL11A1* rs3753841 T and the *COL11A1* rs1676486 C alleles were associated with AT, the interactions of these alleles with the *COL5A1* rs71746744 AGGG/AGGG genotype were further investigated (Table 4.4). The AGGG/AGGG-TT genotype combination of rs71746744 and rs3753841 was significantly over-represented in the TEN group (27.3 %) compared to the CON group (11.4 %) (p=0.005, OR=2.9, 95 % CI=1.4-5.9). Similarly the *COL5A1* rs71746744 AGGG/AGGG genotype combined with the C allele (CC or CT genotypes) of *COL11A1* rs1676486 was also significantly over-represented in the TEN group (64.1 %) compared to the CON group (38.5 %) (p<0.001, OR=2.9, 95 % CI=1.6-5.0).

4.3.6 Pseudohaplotype Analysis of *COL11A1*, *COL11A2* and *COL5A1*

Since significant gene-gene interactions were observed between the types XI and V collagen genes in the previous section, inferred haplotypes and pseudohaplotypes were constructed between the SNPs to confirm these findings (Appendix A, Table A3.3).

4.3.6.1 *COL11A1* rs3753841 and *COL5A1* rs71746744

All four possible pseudohaplotypes were inferred with a frequency greater than 1% from the *COL11A1* rs3753841 and *COL5A1* rs71746744 polymorphisms (Figure 4.3A). In agreement with the interactions reported in Table 4.4, the T/AGGG inferred pseudohaplotype was significantly over-represented in the TEN group (49.3%, n=80) compared to the CON group (37.2%, n=104) (p=0.004) (Figure 4.3A). The T/- and C/- inferred pseudohaplotype were on the other hand significantly under-represented in the TEN group (C/-, 6.4%, n=10; T/-, 14.7%, n=24) compared to the CON group (C/-, 12.6 %, n=35; T/-, 20.8%, n=58)(C/-, p<0.001; T/-, p=0.021).

Table 4.4 Two-way combined genotype frequencies distributions for *COL5A1*rs71746744 and *COL11A1* rs3753841, rs1676486 and *COL11A2* rs1799907

***COL5A1* rs71746744 and *COL11A1* rs3753841**				
rs71746744	**rs3753841**	**CON (n=158)**	**TEN (n=77)**	**TEN/CON**
AGGG/AGGG	**TT**	11.4 (18)	27.3 (21)	**2.39**[a]
AGGG/AGGG	**CT**	22.2 (35)	23.4 (18)	0.95
AGGG/AGGG	**CC**	7.6 (12)	14.3 (11)	1.88
-/AGGG	**TT**	17.7 (28)	18.2 (14)	1.03
-/AGGG	**CT**	20.9 (33)	13.0 (10)	0.62
-/AGGG	**CC**	9.5 (15)	0 (0)	ND
-/-	**TT**	3.8 (6)	0 (0)	ND
-/-	**CT**	5.7 (9)	1.3 (1)	0.22
-/-	**CC**	1.3 (2)	2.6 (2)	2.00
***COL5A1* rs71746744 and *COL11A2* rs1676486**				
rs71746744	**rs1676486**	**CON (n=177)**	**TEN (n=78)**	**TEN/CON**
AGGG/AGGG	**CC**	24.9 (44)	39.7 (31)	**1.59**[b]
AGGG/AGGG	**CT**	13.6 (24)	24.4 (19)	**1.79**[b]
AGGG/AGGG	**TT**	3.4 (6)	1.3 (1)	0.38
-/AGGG	**CC**	26.6 (47)	26.9 (21)	1.01
-/AGGG	**CT**	17.5 (31)	3.9 (3)	0.22
-/AGGG	**TT**	2.8 (5)	0 (0)	ND
-/-	**CC**	9.6 (17)	1.3 (1)	0.13
-/-	**CT**	1.1 (2)	2.6 (2)	2.36
-/-	**TT**	0.6 (1)	0 (0)	ND
***COL5A1* rs71746744 and *COL11A2* rs1799907**				
rs71746744	**rs1799907**	**CON (n=162)**	**TEN (n=83)**	**TEN/CON**
AGGG/AGGG	**AA**	19.1 (31)	27.7 (23)	1.45
AGGG/AGGG	**AT**	19.1 (31)	28.9 (24)	1.51
AGGG/AGGG	**TT**	6.2 (10)	4.8 (4)	0.77
-/AGGG	**AA**	24.7 (40)	18.1 (15)	0.73
-/AGGG	**AT**	18.5 (30)	13.3 (11)	0.72
-/AGGG	**TT**	2.5 (4)	2.4 (2)	0.96
-/-	**AA**	3.7 (6)	1.2 (1)	0.32
-/-	**AT**	3.7 (6)	3.6 (3)	0.97
-/-	**TT**	2.5 (4)	0 (0)	ND

MonteCarlo tests performed with 2000 simulations.

Monte Carlo Test: rs71746744 and rs3753841, P-values for: T1=**0.001**, T2<**0.001**, T3=**0.011**, T4=**0.006**

Monte Carlo Test: rs71746744and rs1676486, P-values for: T1=**0.002**, T2=**0.001**, T3=**0.013**, T4<**0.001**

Monte Carlo Test: rs71746744 and rs1799907, P-values for: T1=0.291, T2=0.191, T3=0.300, T4=0.176

ND, no data. Significant associations (p<0.05) are indicated in bold typeset.

[a] p=0.005, OR=2.9, 95 % CI=1.4-5.9

[b] p<0.001, OR=2.9, 95 % CI=1.6-5.0

4.3.6.2 *COL11A1* rs1676486 and *COL5A1* rs71746744

All four possible pseudohaplotypes were inferred with a frequency greater than 1% from the *COL11A1* rs1676486 and *COL5A1* rs71746744 polymorphisms (Figure 4.3B). As expected the C/AGGG inferred pseudohaplotype was significantly over-represented in the TEN (63.5 %, n=106) group compared to the CON (49.8%, n=142) group (p<0.001) (Figure 4.3B). The C/- and T/- inferred pseudohaplotypes were on the other hand significantly under-represented in the TEN group (C/-, 18.6%, n=31; T/-, 2.3%, n=3) compared to the CON group (C/-, 28.4%, n=81; T/-, 5.1 %, n=14) (C/-, p=0.016; T/-, p=0.042).

4.3.6.3 *COL11A2* rs1799907 and *COL5A1* rs71746744

All four possible pseudohaplotypes were inferred with a frequency greater than 1% from the *COL11A2* rs1799907 and *COL5A1* rs71746744 polymorphisms (Figure 4.3C). The T/AGGG inferred pseudohaplotype was significantly over-represented in the TEN group (29.6%, n=51) compared to the CON group (21.3%, n=65)(p=0.016). The A/AGGG inferred pseudohaplotype was significantly under-represented in the TEN group (13.7%, n=23) compared to the CON group (23.5%, n=72) (p=0.002) (Figure 4.3C).

4.3.6.4 *COL11A1* rs3753841 and rs1676486 and *COL11A2* rs1799907 and *COL5A1* rs71746744

Eleven of the possible 16 pseudohaplotypes were inferred with a frequency greater than 2% from the combined *COL11A1*, *COL11A2* and *COL5A1* polymorphisms (Figure 4.3D). As expected from Figure 4.3A-C, the TCT(AGGG) pseudohaplotype was significantly over-represented (p<0.001) in the TEN group (25.2%, n=46) compared to the CON group (9.1%, n=30). In addition, the CTA(-) (p=0.037) and CCA(-) (p=0.043) inferred pseudohaplotypes were significantly over-represented in the CON group (CTA(-), 5.3%, n=17; CCA(-), 5.3%, n=18) compared to the TEN group (CTA(-), 2.2%, n=4, CCA(-), 1.2%, n=2).

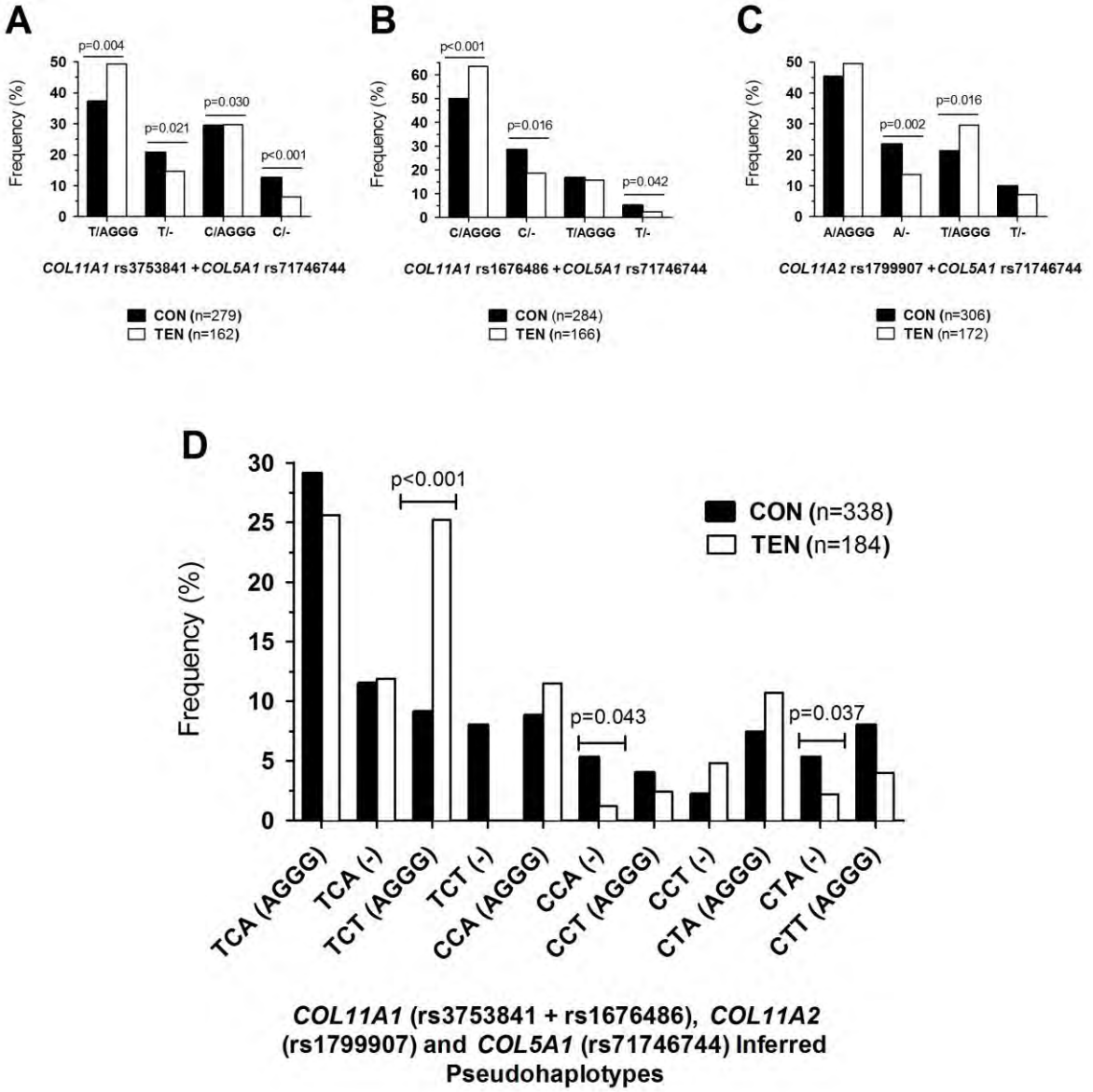


Figure 4.3 Inferred pseudohaplotypes for *COL5A1* rs71746744 and *COL11A1* rs3753841, rs1676486 and *COL11A2* rs1799907 for the pooled Australian and South African control (CON, black solid bars) and Achilles tendinopathy (TEN, clear bars) groups. **A**) *COL11A1* rs3753841 and *COL5A1* rs71746744. **B**) *COL11A1* rs1676486 and *COL5A1* rs71746744. **C**) *COL11A2* rs1799907 and *COL5A1* rs71746744. **D**) *COL11A1* rs3753841 and *COL11A1* rs1676486 and *COL11A2* rs1799907 and *COL5A1* rs71746744. Significant values (P<0.05) are indicated by a line on the graph.

4.4 Discussion

The first main finding of this study was that none of the three newly investigated polymorphisms within the *COL11A1* (rs3753841 and rs1676486) and *COL11A2* (rs1799907) genes were independently associated with AT in the AUS, SA or combined populations. The minor allele frequency of rs3753841 (0.42) rs1676486 (0.22) and rs1799907 (0.20) in the CON group agrees closely with reported minor allele frequencies (0.36, 0.20 and 0.30) respectively, for individuals of European ancestry available on the NCBI database dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP, accessed 12/02/2013).

Departure from HWE was noted in some of the groups, suggesting that perhaps the AUS individuals in this cohort do not represent a randomly mating population, or that the sample size of this study is too small[184,185]. The AUS TEN group was selected based on clinical criteria and geographical ancestry and the AUS CON group was selected to match the age, geographical ancestry and physical activity of the AUS TEN group. The observed departure from HWE may thus be explained by the fact that the recruitment of the CON and TEN groups was highly selective.

The second main finding of this study was that, although there were no independents associations, gene-gene interactions between the *COL11A1* and *COL11A2* genes were associated with AT. Specially, the T and C alleles of the *COL11A1* rs3753841 and rs1676486 SNPs, respectively, as well as the T allele of the *COL11A2* rs1799907 SNP, were associated with increased risk of AT. Consistent results were obtained when investigating combined genotype interactions, haplotype or pseudohaplotype analysis.

As previously mentioned in Chapter 1 (Section 1.6.2) and Chapter 3 (Section 3.4), the non-synonymous *COL11A1* rs3753841 (T/C) and rs1676486 (C/T) SNPs result in L1323P and P1535S amino acid substitutions respectively. The P1535S substitution occurs very near to a predicted disulphide bond at position 1532[129] and could potentially lead to a conformational change in the type XI heterotrimer. Alternatively, the combined effect of the two amino acid changes could alter the folding of the α1 chain, or affect the ability of the α1(XI) chain to form the type XI collagen trimer or interact with other collagens or ECM molecules.

However the effect of the *COL11A1* rs1676486 SNP may be due to quantitative rather than qualitative differences in type XI collagen. Previous investigations into the biological significance of the *COL11A1* SNPs have focussed on rs1676486, where the T-allele of rs1676486 was shown to produce unstable *COL11A1* mRNA transcripts that are degraded faster than the C-allele[129].

Furthermore, the T allele of *COL11A2* rs1799907 (IVS6-4, T/A), which was also implicated in the pseudohaplotypes, produces a distinct isoform of the α2(XI) chain in which several amino acids in the amino terminal acidic domain are deleted. This acidic domain provides potential sites for the interaction of type XI collagen with other molecules and may prevent further deposition of collagen molecules in the fibril[137]. It is therefore reasonable to hypothesise that the risk associated pseudohaplotype has a biological consequence. Although expressed in the developing tendon, further research is however required to determine the functional significance of the type XI collagen gene polymorphisms in tendinopathy.

Intriguingly, while the minor C-allele of *COL11A1* rs3753841 and the minor T-allele of *COL11A1* rs1676486 was over-represented in patients with LDH, the current study described in this chapter found that the T-allele of rs3753841 and the C-allele of rs1676486 were associated with risk of AT. A similar phenomenon has been observed previously for the Sp1 binding site polymorphism (rs1800012) in *COL1A1*; in which the TT genotype was associated with protection against ACL rupture in three independent populations[105–107], but also resulted in increased risk for osteoporosis[208–210] and lumbar disc disease[211,212]. These findings suggest that both the alternative alleles within a gene can be associated with increased risk of different multifactorial disorders.

The third finding of this study was the additive contribution of the AGGG allele of the *COL5A1* -/AGGG polymorphism, together with the TCT inferred pseudohaplotype of the *COL11A1* and *COL11A2* genes, to an increased risk of AT. This gene-gene association is perhaps not surprising taking into account the mounting evidence that the structural and functional homology between the α1(XI) and α1(V) chains encoded by *COL11A1* and *COL5A1* respectively, facilitate an interchangeability between the two polypeptides, resulting in heterotrimer [$\alpha1(XI)_2\alpha2(V)$] formation[147–149,194]. It is possible that the reported interaction in this study between the *COL5A1* and *COL11A1* polymorphisms indicate the role of a minor fibrillar collagen consisting of type V/XI α-chains in tendinopathy. Alternatively the traditional type XI collagen heterotrimer may be able to functionally replace or compensate for type V collagen in tendons. The possible role of type XI collagen in pathology of the mature tendon needs to be investigated.

As previously mentioned, the rapidly degraded T-allele of *COL11A1* rs1676486 was more common in the asymptomatic control population, and the wild-type, more stable C-allele was more common in the tendinopathic population. AT has been associated with increased *COL5A1* mRNA stability[118] and therefore decreased collagen fibril diameters[110]. It is therefore tempting to speculate that the

interaction between this allele and *COL5A1* is due to the effect they have on (i) mRNA stability, (ii) potential effects on type V and XI protein levels, (iii) their subsequent effects on collagen fibril diameter and (iv) altered biomechanical properties of the collagen fibril at the tissue level. It is therefore reasonable to conceive that individuals who have these functional alleles as reflected from the pseudohaplotypes could possibly be producing a functionally altered type XI and type V collagen, which collectively are responsible for the altered biomechanical property of the tendon collagen fibrils.

As previously mentioned, although type XI collagen is classified as a cartilage protein, it is produced in the developing tendon[28,48,145,146]. To our knowledge there is no evidence that the protein is produced in mature tendons. It is therefore possible that the proposed biological consequences of the reported association in this study could be due to altered protein profiling in: (i) the mature, diseased and/or healing tendon, (ii) the fibrocartilaginous regions of the mature tendon and (iii) tendon development. Irrespective of the mechanism(s), further research is required to replicate these findings in larger independent populations and to explore the functional mechanisms underlying the complex genetic associations with the type XI collagen encoding genes and their interactions with the type V collagen encoding gene.

In conclusion, the functional variants within the type XI collagen genes investigated in this study were not independently associated with AT. This study does however provide evidence suggesting that the genes that encode for the structural and functionally related type XI (*COL11A1* and *COL11A2*) and type V (*COL5A1*) collagen interact with one another to collectively modulate the risk for AT. Although expressed in the developing tendon, the role of type XI collagen in the pathology of the mature tendon requires future investigation.

Chapter 5: Conclusions

Both ACL rupture and AT are multifactorial conditions that result from the complex interaction of intrinsic and extrinsic risk factors[2]. Recently, genetic factors have been identified as important intrinsic risk factors in both these conditions[90,92,93]. However, the two conditions differ in their aetiology: ACL rupture is an acute injury that results from the sudden application of excessive internal or external forces that exceed the strength of the tissue[75,153,189]; while AT is a largely degenerative condition that results from overuse with a failure of the tendon to repair and adapt[14,15,77,78,200,203]. Because the aetiology of these conditions differ, it is likely that the molecular mechanisms that underlie these conditions also differ and the genetic variants that predispose to these conditions may therefore be unique. However, as the molecular components and structure of ligaments and tendons are very similar, genetic variants that affect these structures, and therefore the integrity of these tissues, are more likely to be implicated in both. Thus, investigating the genes that are important molecular or regulatory components of both ligaments and tendons might identify genetic variants that predispose individuals to both ACL ruptures and AT. Moreover, this molecular approach may also identify the similarities and differences in the aetiology of ACL ruptures and AT.

The CC genotype of the rs12722 (C/T) polymorphism in the *COL5A1* 3'-UTR was previously associated with reduced risk of AT[101,102] and reduced risk of ACL ruptures in females [103]. Subsequently, three polymorphisms, namely rs71746744 (-/AGGG), rs16399 (ATCT/-) and rs1134170 (A/T), also within the *COL5A1* 3'-UTR, have independently been associated with AT[104]. Specifically the AGGG/AGGG, -/- and TT genotypes of these three polymorphisms were associated with increased risk of AT[104]. Furthermore, the *COL5A1* 3' -UTR contains several putative *cis*-acting elements including a functional Hsa-miR-608 binding site [102,104]. Two forms of the mature Hsa-miR-608, which are produced from the polymorphic (rs4919510, C/G) SNP in the *MIR608* gene on chromosome 10q24, can potentially bind this miRNA binding site[104]. Recently the CC genotype of *MIR608* rs4919510 was shown to be associated with increased risk of AT[104]. Therefore, the first objective of this dissertation was to determine whether the rs71746744 (-/AGGG) and rs1134170 (A/T) polymorphisms in the *COL5A1* gene and the rs4919510 (C/G) polymorphism in the *MIR608* gene were associated with ACL rupture in a South African Caucasian study sample (Chapter 2).

The *COL5A1* gene codes for a component of type V collagen; a protein that regulates the fibril diameter and density of collagen fibrils in a number of connective tissues including tendon[19–21,25,26,213] and cartilage[42,43]. Fibril diameter is positively associated with the tensile strength of a tissue,

and inversely associated with elasticity (Chapter 1, Section 1.4). Type XI collagen is involved in a similar biological process and shares functional homology with type V collagen[28]. Type XI collagen is expressed in developing tendons[48,145,146] and is encoded by the *COL11A1*, *COL11A2* and *COL2A1* genes. These genes are also implicated in several musculoskeletal disorders that affect connective tissue structure and biomechanical properties[121,123,146,192,193,214]. Since the *COL11A1* and *COL11A2* genes have not previously been investigated for an association with acute ACL ruptures and/or AT, the second (Chapter 3) and third (Chapter 4) objectives were therefore to determine whether the rs3753841 and rs1676486 polymorphisms in the *COL11A1* gene, and the rs1799907 polymorphism in the *COL11A2* gene are associated with ACL rupture (Chapter 3) or AT (Chapter 4) in a Caucasian population. A secondary aim of these studies was to determine whether these specific *COL11A1* and *COL11A2* polymorphisms collectively interact with the rs71746744 polymorphism in the *COL5A1* gene to modulate the risk of sustaining an ACL rupture (Chapter 3) or AT injury (Chapter 4).

5.1 Novel Findings of this Dissertation

This dissertation identified the following significant associations:

- None of the *COL5A1* (rs71746744 and rs1134170) and *MIR608* (rs4919510) polymorphisms investigated in this dissertation were independently associated with ACL rupture risk. A novel finding of this dissertation was that the inferred -/A haplotype constructed from the *COL5A1* 3'-UTR rs71746744 (-/AGGG) and rs1134170 (A/T) polymorphisms was however significantly under-represented in all ACL participants (25.1%, p<0.001) and in the male (25.0%, p=0.001) ACL rupture group compared to their respective control groups (all 27.9%, male 27.6%). This suggests a "weak" protective effect of these variants against ACL rupture (Chapter 2). Although the female frequency distribution was similar to that of the combined male and female and male only groups, it was not significant due to its small sample size. The significantly associated inferred pseudohaplotypes constructed from the two *COL5A1* polymorphisms and the *MIR608* rs4919510 (C/G) polymorphism were difficult to interpret. Additional work with larger samples sizes is required to investigate these interactions.

- The TC genotype of the *COL11A1* rs3753841 polymorphism was significantly under-represented within the ACL (41.4%) rupture group, compared to the CON (54.2%) group (p=0.009, OR=1.6, 95% CI=1.1 to 2.5) (Chapter 3), suggesting that this genotype is protective. The biological mechanism of this protection is unknown. None of the other type XI collagen gene polymorphisms, *COL11A1* rs1676486 and *COL11A2* rs1799907, were independently associated with ACL rupture risk. An additional novel finding of this dissertation was that the inferred

COL11A1 CT haplotype constructed from rs3753841 and rs1676486 was significantly under-represented (p=0.044) in the ACL group (17.4%) compared to the CON group (21.3%), suggesting a protective effect. In addition, the inferred CTA pseudohaplotype constructed from the two *COL11A1* polymorphisms and *COL11A2* rs1799907 (A/T) was also significantly under-represented (p=0.010) in the ACL group (9.5%) compared to the CON (14.7%) group, suggesting a protective effect against ACL rupture (Chapter 3).

- The major novel finding of this dissertation was that the TCT inferred pseudohaplotype constructed from *COL11A1* rs3753841, *COL11A1* rs1676486 and *COL11A2* rs1799907 was significantly over-represented (p=0.006) in the TEN (25.9%) group compared to the CON (17.1%) group. The TCT(AGGG) inferred pseudohaplotype constructed using these type XI collagen polymorphisms and the functional *COL5A1* rs71746744 (-/AGGG) polymorphism was also significantly over-represented (p<0.001) in the TEN (25.2%) group compared to the CON (9.1%) group, suggesting a "strong" effect (Chapter 4). In addition, the CTA(-) haplotype was significantly over-represented (p=0.037) in the CON (5.3%) compared to the TEN (2.2%) group.

In summary, five of the six polymorphisms investigated in this study were implicated in modulating the risk of ACL ruptures; either through an independent association (*COL11A1* rs3753841) and/or an inferred haplotype or pseudohaplotype (*COL5A1* rs71746744, *COL5A1* rs1134170, *COL11A1* rs1676486 and *COL11A2* rs1799907) association. Furthermore, all three polymorphisms investigated in the AT study (*COL11A1* rs3753841, *COL11A1* rs1676486 and *COL11A2* rs1799907) were implicated in modulating the risk of AT through an inferred pseudohaplotype association. The *COL5A1* rs71746744, *COL5A1* rs1134170 and *MIR608* rs4919510 polymorphisms have previously been implicated in the risk of AT[104]. There was consistency and agreement (Table 5.1) between the genotype and/or alleles associated with risk of both ACL rupture and AT. This dissertation therefore highlights similarities in the aetiology of acute ACL ruptures and AT. The potential biological significance of these allele combinations will be discussed in the following sections.

Table 5.1 A Summary of reported findings of the *COL5A1, MIR608, COL11A1* and *COL11A2* polymorphisms implicated in the risk of ACL rupture and Achilles tendinopathy

	ACL Rupture	Achilles tendinopathy	
	Protection	**Protection**	**Risk**
***COL5A1* rs71746744 (-/AGGG)**	- allele	- allele	AGGG/AGGG genotype [1]
***COL5A1* rs1134170 (A/T)**	A allele		TT genotype [1]
***MIR608* rs4919510 (C/G)**	Not associated		CC genotype [1]
***COL11A1* rs3753841 (C/T)**	CT genotype C allele	C allele	T allele
***COL11A1* rs1676486 (T/C)**	T allele	T allele	C allele
***COL11A2* rs1799907 (A/T)**	A allele	A allele	T allele

[1] 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*. doi:10.1111/ahg.12013

5.2 Potential Biological Significance of the *COL5A1*, *COL11A1* and *COL11A2* Associations and Interactions

5.2.1 *COL5A1*

This dissertation investigated the *COL5A1* rs71746744 (-/AGGG) and rs1134170 (A/T) polymorphisms for an association with ACL rupture (Chapter 2). As previously mentioned (Section 1.6.1), the *COL5A1* gene encodes the α1 chain of type V collagen, a minor fibrillar collagen that is vital for nucleation of collagen fibrils and the regulation of collagen fibril diameter[23,26]. With the exception of Ehlers Danlos syndrome, where a 50% reduction in the production of type V collagen causes the production of large irregular cauliflower shaped fibrils with reduced tensile strength[21], the greater the concentration of type V collagen in a tissue, the smaller the collagen fibril diameter[23,26], and smaller diameter collagen fibrils are associated with lower tensile strength[73]. Thus polymorphisms which alter the amount of type V collagen produced will alter the structure of the collagen fibrils, with consequent effects on the biomechanical properties of the tissue[110,215].

As previously discussed, the common rs12722 (C/T) SNP in the *COL5A1* 3'-UTR has previously been associated with AT[101,102] and ACL rupture in females[103]. It has also been associated with ROM[115,116], general joint laxity and knee hyperextension (genu recuvarum)[190], and endurance running performance[117]; suggesting a role for this region in multiple sports injury and exercise-related phenotypes[110]. The rs12722 SNP is one of seven tightly linked polymorphisms in the *COL5A1* 3'-UTR:

rs13946 (C/T), rs12722 (C/T), rs3196378 (C/A), rs71746744 (del/AGGG), rs16399 (ACTC/del), rs1134170 (A/T) and rs3128575 (T/C). Two allelic forms of the *COL5A1* 3-UTR have been identified: the C form, which corresponds to the wild-type sequence, was identified in most of the clones generated from asymptomatic controls, and contains the polymorphisms in the sequence C-C-C-$(AGGG)_1$-$(ATCT)_2$-A-T, while the T form was the predominant form identified in the AT patients, and contains the polymorphisms in the sequence T-T-A-$(AGGG)_2$-$(ATCT)_1$-T-C[118]. The T-form was found to have increased mRNA stability compared to the C-form[118] and the AGGG/AGGG, del/del and TT genotypes of rs71746744, rs16399 and rs1134170 respectively, were each independently associated with AT in a South African and Australian study population[104].

When two of these *COL5A1* 3'-UTR variants, rs71746744 (del/AGGG) and rs1134170 (A/T) were investigated, neither of the polymorphisms were independently associated with ACL rupture (Chapter 2). It is probable that the hypothesised association of these variants with ACL ruptures in females could not be determined due to the small sample size of the female participants and the relatively high frequency of the proposed associated alleles of these polymorphisms (rs71746744 AGGG, 0.71; rs1134170 T, 0.72). The previous association of the rs12722 SNP with female ACL ruptures[103] was more informative in a smaller sample size due to the relatively low frequency (0.43) of the minor C allele. Further work with a larger sample size is thus required to test the association of these polymorphisms with ACL rupture in females. In agreement with the previously published study of the rs12722 SNP and ACL rupture, these additional 3'-UTR polymorphisms were not associated with ACL ruptures in males[103].

However, in agreement with the hypothesis, when inferred haplotypes were constructed from these two *COL5A1* 3'-UTR polymorphisms, the -/A haplotype was under-represented in the combined and male ACL rupture groups compared to their respective control groups. Although not significant due to a small sample size, the inferred haplotype distributions were similar in the females. The finding that males with the *COL5A1* -/A haplotype were protected from ACL rupture is a novel finding, as a previous study on the *COL5A1* rs12722 SNP found no association in males[103]. This is however not surprising as haplotype analyses potentially have a greater power to statistically implicate a genetic interval in an association study compared to the power of individual polymorphisms at the same loci[216,217]. This dissertation is therefore the first to suggest that *COL5A1* may be a 'weak' modulator of ACL rupture risk in males. This would be in agreement with the proposed function of the *COL5A1* 3'-UTR polymorphisms on injury risk discussed in the next paragraph. Further work is however required to test this hypothesis.

A molecular mechanism for the role of the *COL5A1* 3'-UTR polymorphisms on injury risk has been proposed (Figure 5.1). Polymorphisms such rs71746744 and rs1134170 in the *COL5A1* 3'-UTR (and others) collectively determine the secondary structure of the *COL5A1* 3'UTR mRNA transcript. The different combinations result in two allelic forms with gross structural differences that are thought to influence the stability of the mRNA transcript by altering a miRNA binding site or some other mRNA stability protein[104].

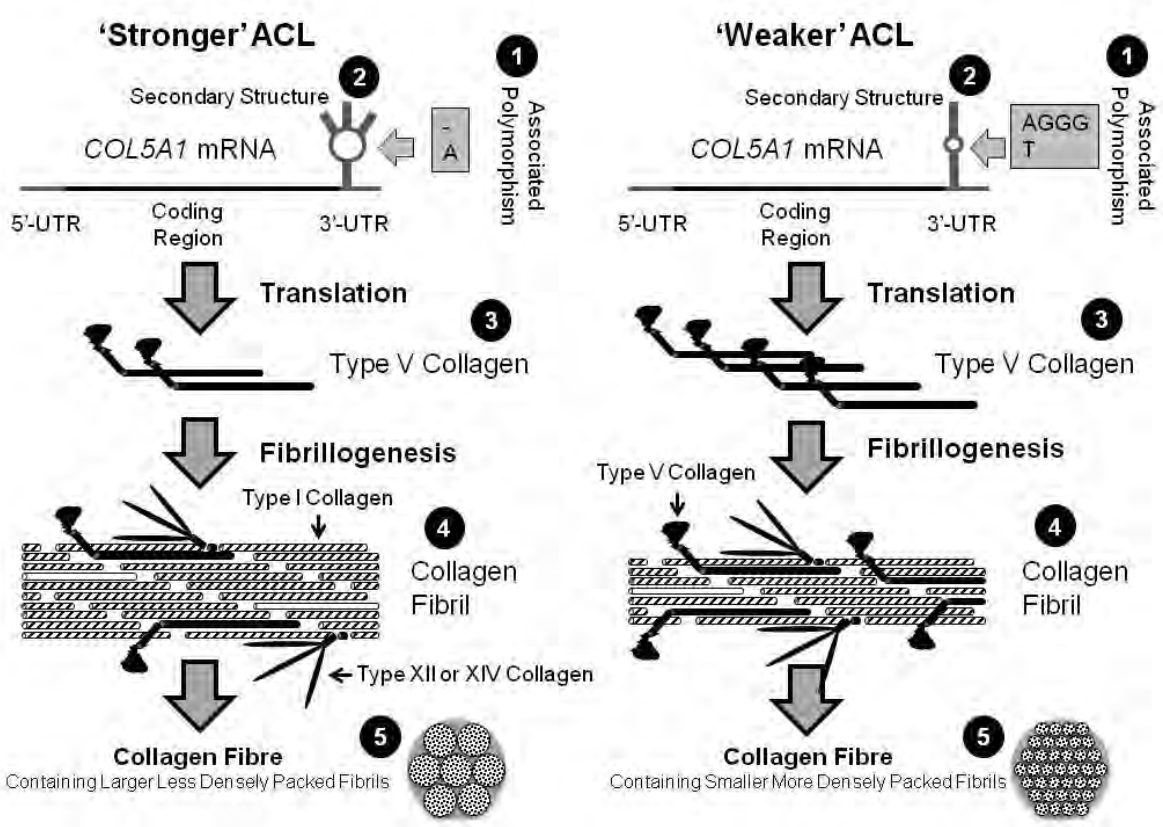


Figure 5.1 A hypothetical schematic diagram illustrating the proposed mechanism of how polymorphisms within the *COL5A1* 3'-UTR potentially affect fibrillogenesis in the anterior cruciate ligament (ACL). (**1**) The *COL5A1* rs71746744 (-/AGGG) and *COL5A1* rs1134170 (A/T) polymorphisms are part of an inferred haplotype that is associated with ACL rupture. (**2**) The *COL5A1* rs71746744 (-) allele and the *COL5A1* rs1134170 A allele are both believed to associated with increased mRNA degradation, due to changes in the secondary structure of the mRNA. Increased mRNA degradation is indicated in the left panel, while decreased mRNA degradation is indicated in the right panel. (**3**) The altered mRNA stability associated with these polymorphisms is believed to result in altered α1(V) chain and type V collagen production (decreased in left panel and increased in the right panel). (**4**) Types V collagen regulates collagen fibril assembly and diameter (fibrillogenesis) and thus the mechanical properties of ligaments. (**5**) There is an inverse relationship between the types V collagen content of the fibril and its diameter. Thinner more densely packed collagen fibrils are produced due to the increased production of types V collagen by the more stable AGGG/T inferred haplotype (Right Panel). Figure modified from Ribbans, W. J., and M. Collins. (2013)Pathology of the tendo Achillis: Do our genes contribute?, The bone & joint journal, 95.3: 305[5].

It is hypothesised that the -/A haplotype has lower mRNA stability than the AGGG/T haplotype. The decreased mRNA stability of the -/A haplotype is predicted to decrease the rate of protein translation, resulting in decreased concentrations of type V collagen. Decreased type V collagen production would subsequently affect collagen fibril diameter, creating larger collagen fibrils with increased tensile strength and decreased elasticity. The increased tensile strength of the collagen fibrils, coupled with the decreased joint flexibility may protect against rupture of the ACL.

5.2.2 *MIR608*

Hsa-miR-608 was shown to bind to the *COL5A1* 3'-UTR and the CC genotype of the rs4919510 (C/G) SNP within the *MIR608* gene was shown to independently associate with risk of AT[104]. Therefore the rs4919510 (C/G) polymorphism in the *MIR608* gene was also investigated for interactions with the *COL5A1* SNPs. The *MIR608* rs4919510 SNP was not independently associated with ACL rupture, however the AGGG/T/C combination of *COL5A1* rs71746744 (-/AGGG), rs1134170 (A/T) and *MIR608* rs4919510 (C/G) was over-represented in the female ACL group compared to the CON group, which is consistent for the risk alleles in AT[104]. Other inferred pseudohaplotypes which do not fit into our current working hypothesis were also identified. These could be a type I error due to the small female sample size. Additional work is required to investigate these polymorphisms in larger sample sizes.

5.2.3 *COL11A1* and *COL11A2*

Polymorphisms in the *COL11A1* and *COL11A2* genes were investigated for associations with risk of developing ACL rupture (Chapter 3) and AT (Chapter 4). As previously mentioned in Section 1.6.2, type XI collagen is a quantitatively minor fibrillar collagen comprising three polypeptide chains, α1(XI), α2(XI) and α1(II), encoded by the *COL11A1*, *COL11A2* and *COL2A1* genes respectively. Type XI collagen plays a similar role to type V collagen in tendons and ligaments, in that it regulates collagen fibril diameter in cartilage[43], which mostly consists of type II collagen[29]. Despite the fact that type XI plays a similar role to type V collagen in regulating collagen fibril diameter[28], and is expressed in developing tendon[48,145,146,213], this is the first study to our knowledge that investigated the *COL11A1* and *COL11A2* for associations with tendon or ligament injuries.

Similar to the findings noted for the *COL5A1* polymorphism analyses conducted in Chapter 2, the *COL11A1* and *COL11A2* SNPs were not independently associated with ACL rupture risk. However, when inferred haplotypes or inferred pseudohaplotypes were constructed, the C, T and A alleles of *COL11A1* rs3753841 (T/C), rs1676486 (C/T) and rs1799907 (A/T) were associated with protection

against ACL rupture, and the complementary T, C and T alleles were associated with risk of AT. This finding suggests a common molecular mechanism is involved in both AT and ACL rupture.

The ancestral allele of the rs3753841 (T/C) polymorphism in exon 52 of the *COL11A1* gene is a cytosine (C) nucleotide; however the major allele (highest frequency) in the Caucasian populations is a thymine (T) nucleotide. The notation for polymorphisms investigated in this dissertation has been (major allele/minor allele). The change from the ancestral C allele to the T allele results in an amino acid change from a proline (codon: C<u>C</u>T) to a leucine (codon: C<u>T</u>T) at position 1323; a substitution that has been predicted to be deleterious by SIFT (Sorting Intolerant From Tolerant)[218]. The SIFT algorithm assesses the effects of single amino acid changes on protein function based on the assumption that sequence homology will be conserved for important amino acid positions[218]. The ancestral (or minor) C-allele has previously been associated with LDH[129]; however in this dissertation the major T allele, which corresponds to the deleterious amino acid change, was associated with risk for AT and the C allele was associated with protection from ACL rupture. The deleterious T allele may increase risk via changes in the structure of the α1(XI) chain, which may affect the assembly of the type XI collagen trimer, or alternatively may affect the ability of the type XI molecule to interact with other collagens or ECM components. Further work is however required to investigate this hypothesis.

The rs1676486 (C/T) polymorphism in exon 62 results in a non-synonymous amino acid change from a proline (codon: <u>C</u>CT) to a serine (codon: <u>T</u>CT) at position 1535. The P1535S occurs very near to a predicted disulphide bond at position 1532[129] and could potentially lead to a conformational change in the protein. Further work is however required to investigate any functional consequences of this amino acid substitution. The T-allele of this variant has been associated with LDH[129] and limbus vertebra in gymnasts[130].

The TC haplotype of rs3753841 and rs1676486 was associated with AT risk and the CT haplotype was associated with protection from ACL rupture. The haplotypes were therefore more informative in capturing the ACL risk profile than either of the individual polymorphisms. As previously mentioned in Chapter 3 (Section 3.4) and Chapter 4 (Section 4.4), this haplotype association could be due to a cumulative effect of the two non-synonymous amino acid changes on protein folding or on the ability of the α1(XI) chain to form the type XI collagen trimer or interact with other collagens or ECM molecules. This genetic interval defined by the associated haplotype therefore requires further exploration to identify the risk associated sequence motif within this gene region so that one can begin to explore the patho-biological significance of these sequences.

Intriguingly, while the minor C allele of *COL11A1* rs3753841 and the minor T allele of *COL11A1* rs1676486 were over-represented in patients with LDH, the current study implicated the alternate alleles in the risk-associated pseudohaplotype. Similarly, the functional rare TT genotype of the *COL1A1* Sp1 binding site polymorphism was reported to be associated with risk for several multifactorial disorders, including osteoporosis[208–210,212] and lumber disc disease[211,212], but was associated with protection against ACL ruptures in three independent populations[105–107]. These findings suggest that both the alternative alleles within a gene can be associated with increased risk of different multifactorial disorders.

Although both *COL11A1* polymorphisms cause amino acid substitutions, as mentioned in Chapter 3 and 4, the effect of the *COL11A1* rs1676486 SNP may be due to quantitative rather than qualitative differences in type XI collagen. The T allele of rs1676486 was shown to produce unstable *COL11A1* mRNA transcripts that are degraded faster than the C allele[129]. It has been hypothesised that the 4856–4865 nucleotides (caaaaaatct) in *COL11A1* mRNA closely match the consensus for a mRNA stability motif, "g/tanaaaag/tcc/t" approximately 200 bp away[129]. The sequence variation may therefore affect the mRNA stability motif and disrupt the *cis*-element critical for mRNA stability. Alternatively, the sequence variation might induce a conformational change in the mRNA that could decrease mRNA stability or increase the sensitivity of the transcript to RNases[129].

There is some evidence that the change in mRNA stability caused by the rs1676486 SNP does affect protein levels[129]. In a study on LDH, the less stable rs1676486 T allele was associated with increased risk, and although protein levels were not measured in the study, there was an inverse relationship between *COL11A1* expression and severity of disc degeneration. This led the authors to hypothesise that the decrease in the *COL11A1* transcript associated with the T-allele led to the decrease in type XI collagen protein levels [129].

Moreover, since the C allele of rs1676486 was associated with AT, this suggests that increased type XI collagen levels are associated with increased injury risk. This is comparable to the proposed mechanisms of increased *COL5A1* mRNA stability leading to increased type V collagen and increased risk of AT (Figure 5.1). This will be further explored in the section on interactions between *COL5A1* and *COL11A1*.

The *COL11A2* rs1799907 (A/T) SNP occurs in intron 5, and results in alternative splicing of the *COL11A2* gene, resulting in protein products of slightly different length[137]. The T allele of *COL11A2* rs1799907 was associated with increased risk of AT when combined with the *COL11A1* TC haplotype

(Chapter 4), and the A-allele was associated protection from ACL rupture when combined with the CT haplotype of *COL11A1* (Chapter 3). In contrast to the investigated *COL11A1* SNPs, where the risk and protection alleles were reversed compared to previous studies[129,130], the association of the T allele of *COL11A2* rs1799907 with risk agrees with previous studies that associated the T allele with lumbar disc desiccation[128], lumbar spine stenosis[141] and OPLL[136–139].

The rs1799907 polymorphism may exert its physiological effect via the alteration it causes in the highly acidic variable domain encoded by exons 6-8 at the amino terminal of the α2(XI) chain[137]. This variable region provides potential sites for the interaction of type XI collagen with other molecules and may prevent further deposition of collagen fibrils[145]. Exon 7, which is often skipped in the T allele, and included in the A allele, contains 8 acidic residues of the 21 amino acids, none of which are basic, suggesting that this region in particular is involved in cross-linking to each other or other molecules[137]. Therefore the A allele may exert is protective effect because it is better able to interact with components of the ECM.

An explanation for how the T allele of rs1799907 interacts with the T or C allele of rs3753841 or rs1676486 is not known, but the missing exons could potentially alter the way that the type XI collagen fibril forms, and the amino acid substitutions could have small effects on the proα(XI) chain. In the conventional type XI isoform, [α1(XI) α2(XI) α1(II)], the α1(XI) and α2(XI) chains combine in a 1:1 ratio. It is possible that the alternatively spliced variant from either the A or the T allele of rs1799907 interacts better with either the C or T allele of rs3753841.

The significant associations of the *COL11A1* and *COL11A2* inferred haplotypes and pseudohaplotypes with AT and ACL rupture risk suggest that type XI collagen plays a role in the aetiology of both conditions. It is therefore reasonable to hypothesise that the associated pseudohaplotype has a biological consequence. It should be noted that the observed statistical association was greater in the AT study than ACL rupture investigation. Both *COL11A1* and *COL11A2* are expressed in the developing tendon[48,145,146], however, further research is required to determine the functional significance of these type XI collagen gene polymorphisms in mature tendon and ligament, as this will facilitate our understanding of their biological significance in tendinopathy and ligament rupture injuries.

5.2.4 *COL5A1*, *COL11A1* and *COL11A2*

Many biological systems have built-in redundancy whereby two or more genes perform the same or similar functions, and inactivation of one of these genes has little or no effect on the biological phenotype[219]. Type V and type XI collagens are highly homologous proteins that perform the similar

function of regulating collagen fibril diameter[28]. In addition, type V and type XI collagens are regulated in a similar manner[194] and can form type V/XI heterotimers[28,147–149,214].

Recently, phenotypic evidence for biological redundancy between type V and type XI collagen in the development of certain tissues was suggested[146,214]. A study on developing tendon in mice revealed that the *COL11A1* gene can compensate for *COL5A1* haploinsufficiency[146]. Deletion of the *COL11A1* gene in addition to *COL5A1 +/-* haploinsufficiency resulted in a far more severe phenotype, with larger, more irregularly shaped fibrils and lower tensile strength, than that of the *COL5A1+/-* knock-out mice alone. Thus, the *COL5A1* +/- haplotype was most severe when compensation by *COL11A1* was prevented. The *COL11A1* gene can therefore functionally "replace" the *COL5A1* gene functions in certain circumstances.

A second example of biological redundancy between type V and XI collagens comes from a study on *cho* mice which harbour a deletion of the *COL11A1* gene[220]. Fernandes et al (2007) showed that lack of the α1(XI) chain, which is vital to the formation of type XI collagen, did not result in a total lack of fibril formation[214]. Rather, in the absence of α1(XI) chains, the *cho* mice synthesized alternate type XI chain assemblies consisting of α1(V)α2(XI)α1(II)[214]. The ability of the α1(V) chain to 'stand in' for the α1(XI) chain in *cho* mice is further evidence of biological redundancy between the *COL5A1* and *COL11A1* genes.

Because the proα1(XI) and proα1(V) chains of type V and type XI collagen may be interchangeable, the previous association of the *COL5A1* gene with AT[101,102] served as a major motivation for investigating the *COL11A1* gene. An aim of this dissertation was therefore to investigate if there was any modulating effect between these two type XI collagen genes on injury risk; and in particular, to identify if there was any potential interactions between the previously associated *COL5A1* polymorphisms and the newly investigated *COL11A1* and *COL11A2* SNPs on injury risk.

In ACL rupture, the interactions between *COL11A1* rs3753841, rs1676486, and *COL11A2* rs1799907 and *COL5A1* rs71746744 did not make biological sense. However, based on previous work[103], the *COL5A1* rs71746744 was proposed to be associated with ACL rupture in females (Chapter 2). The results from this dissertation also suggested that the *COL5A1* 3'-UTR variants may "weakly" modulate the risk of ACL ruptures in males. Larger sample sizes are therefore required to investigate type XI collagen polymorphism interactions with *COL5A1* 3'-UTR polymorphisms in modulating risk for ACL ruptures.

In AT however, the combination of the AGGG allele of rs71746744 and the TCT inferred pseudohaplotype constructed from *COL11A1* rs3753841, rs1676486 and *COL11A2* rs1799907 was "strongly" associated with increased risk of injury. The possible biological significance of this combination of alleles will be discussed in greater detail, together with the development of a model and hypothesis in Section 5.3.

5.3 An Interaction Between Type V/XI Collagen, Collagen Fibril Diameter and Musculoskeletal Soft Tissue Injuries: A Hypothesis

Recently, Collins and Posthumus proposed a novel hypothesis for how the *COL5A1* gene may be associated with musculoskeletal injuries and exercise-related phenotypes via alterations in architecture of collagen fibrils with consequent effects on the fibril's biomechanical properties[110]. Genetic variants in the *COL5A1* 3'-UTR are thought to alter the mRNA stability of *COL5A1* , changing the amount of transcript available to be translated into protein[110]. The α1(V) chain of type V collagen is the rate-limiting step of type V collagen synthesis, and therefore an increase in the α1(V) chain may result in an increase in type V collagen. The increase in type V collagen may result in smaller diameter fibrils[23,26].

This dissertation adds to this model (Figure 5.2). As previously mentioned, the rapidly degraded mRNA produced from the T allele of *COL11A1* rs1676486 was more common in the asymptomatic control population, and the wild-type, more stable mRNA produced from the C allele was more common in the tendinopathic population. Thus we speculate that the rs1676486 C allele results in higher expression of α1(XI), resulting in smaller, more densely packed fibrils when combined with the stable T-allelic form of *COL5A1* (containing rs71746744 AGGG). Taking this into account, when the *COL5A1* AGGG and *COL11A1* C risk associated alleles are inherited together, they potentially result in a synergistic effect on total risk. We hypothesise that the more stable mRNA transcripts may result in greater translation, higher protein levels and smaller diameter collagen fibrils, which will consequently alter the biomechanical properties of the tendon and ligament tissue. It is also tempting to speculate further from this model, that the rapidly degraded mRNA from the *COL11A1* rs1676486 T allele is able to attenuate the effect of the more stable mRNA from the *COL5A1* T-form, because the excess *COL5A1* mRNA counteracts and replaces lower *COL11A1* mRNA levels (Figure 5.2).

It is possible that the reported interaction in this study between the *COL5A1* and *COL11A1* polymorphisms indicate the role of a minor fibrillar collagen consisting of type V/XI α-chains in tendinopathy. Alternatively the traditional type XI collagen heterotrimer may be able to functionally replace or compensate for type V collagen in tendons.

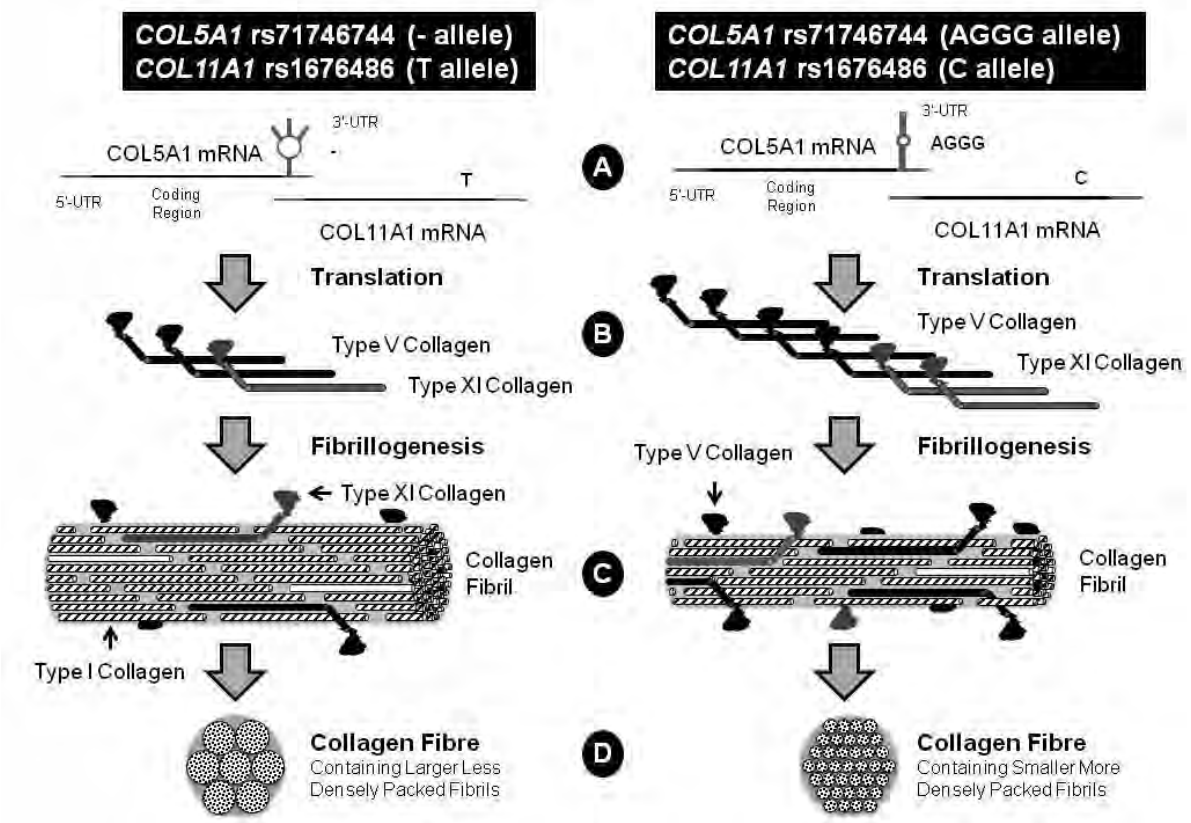


Figure 5.2 A hypothetical schematic diagram illustrating the proposed mechanism of how polymorphisms within *COL5A1* and *COL11A1* potentially affect fibrillogenesis. (**A**) The *COL5A1* rs71746744 (-/AGGG) and *COL11A1* rs1676486 (C/T) polymorphisms are part of an inferred pseudohaplotype that is associated with AT. The *COL5A1* rs71746744 (-) allele and the *COL11A1* rs1676486 T allele are both believed to associated with increased mRNA degradation. Increased mRNA degradation is indicated in the left panel, while decreased mRNA degradation is indicated in the right panel. (**B**) The altered mRNA stability associated with these polymorphisms is believed to result in altered α1(V) and α1(XI) chain and types V and XI collagen production (decreased in left panel and increased in the right panel). (**C**) Types V and XI collagen regulates collagen fibril assembly and diameter (fibrillogenesis). Thus low concentrations of type V/XI collagen result in large diameter fibrils (left panel), and high concentrations results in small diameter fibrils (right panel). (**D**) The diameter and packing density of the collagen fibrils ultimately affect the mechanical properties of tendons. Therefore larger diameter fibrils have greater tensile strength and reduced elasticity (left panel), while small diameter fibrils have decreased tensile strength and increased elasticity (right Panel).

An assumption of this model is that type V and type XI collagen are both expressed and able to interact in tendon and ligament tissue. As previously mentioned type XI is produced in the developing tendon[11 12], however to our knowledge there is no evidence that the protein is produced in mature tendons. The proposed biological consequences of the reported association in this study could thus be the result of its expression and function during tendon development. Alternatively, this association may be due to altered protein profiling in (i) the expression of type XI collagen in the mature, diseased and/or healing tendon, or (ii) its expression only in the fibrocartilaginous regions of the mature tendon. Both of these alternatives should be further investigated by determining the levels of type XI or type V/XI heterotrimers in healthy and diseased mature tendon as well as in the midsubstance and insertion regions.

In summary, the association of these allele combinations with musculoskeletal injury risk may be the biological consequence of their cumulative effect on mRNA stability, with potential effects on protein levels and collagen fibril diameter. Collagen fibril diameter will ultimately affect the mechanical properties of the tissue and hence influence an individual's risk of sustaining a potential injury.

5.4 Limitations

The sample sizes of the groups within this dissertation were not sufficiently powered to detect small genetic effects, the female ACL rupture group in particular. Rather, the studies were powered to detect relatively large genetic effects of 1.5-1.7. This was a limitation, as polygenic multifactorial conditions such as ACL rupture and AT are generally caused by numerous variants, many with a small effect size[94,221,222]. However, this sample size was previously more than adequate to identify the independent association of several polymorphisms with musculoskeletal soft tissue injuries[101–104]. Many of the genes, such as *COL11A1* and *COL11A2*, which are hypothesised to modulate or interact with previously strongly associated variants, such as *COL5A1* 3'-UTR polymorphisms, probably have a small effect size, and therefore much larger sample sizes would be needed in order to increase the power of future studies.

Departure from HWE was present in the ACL and AT cohorts. As previously discussed, this may be due to the sample sizes being too small; however it may also be a sign that the sample does not represent a random mating population[185]. Since the cases for ACL rupture and AT were selected based on their clinical diagnosis, the departure may also be a true reflection of a difference in genotype frequency distribution in the injured group.

- *Achilles Tendinopathy Cohort*

Although an effort was made to match the AUS and SA Achilles tendinopathy and control participants, this was not always possible. The SA cohort, consisting of SA TEN and SA CON groups, all completed physical activity questionnaires, and were grouped for similar levels of activity [100,101]. The physical activity of the AUS CON and AUS TEN groups was unfortunately not captured, and therefore activity between the groups cannot be quantitatively compared. Care was taken however to recruit physically active individuals in both populations, and repeatable results were obtained between the two cohorts.

There are several pathologies that affect the Achilles tendon and surrounding tissues[76]; however this study included only those participants that had been diagnosed with chronic Achilles tendinopathy. Diagnosis was made using clinical criteria and/or soft-tissue ultrasound as previously described [101,102]. The majority of participants injured the mid-portion of their tendon; however a portion of the participants were injured at the insertion site. Although care was taken to only include participants with a well-defined Achilles tendon injury[101,102], we cannot exclude the possibility that the TEN group represented a number of related, but similar, pathologies; and it is possible that these different tendinopathies differ slightly in their genetic predisposition. However, the genetic associations reported in this study would most likely be strengthened if participants with a similar pathology with a different aetiology were excluded.

Since type XI collagen is expressed mainly in the OTJ, the investigated SNPs may have a stronger association with insertional Achilles tendinopathy than midsubstance injury and are ideal candidates for investigating this subset of Achilles tendon injuries. Similarly, type XI collagen polymorphisms would make ideal candidates for association studies of other tendon pathologies, such as patellar tendinopathy, which is characterized by degeneration at the bone-tendon interface where type XI collagen is more likely to be expressed[223].

Another limitation is that the TEN and CON participants differed in several other potentially contributing physical characteristics. The TEN group for both the SA and AUS cohorts were significantly heavier than the CON participants and excess weight has previously been implicated in the development of AT[224]. Future studies should accurately document weight at injury, since many of the TEN group reported gaining weight after the initial injury.

- *Anterior Cruciate Ligament Cohort*

In the ACL cohorts, it was difficult to recruit a large sample of female cases, even though the condition is more likely to affect females[103,108]. Many previous studies on ACL rupture risk have mainly identified significant genetic associations, specifically with *COL5A1*, in the female participants[103,108]. Therefore a larger sample of female participants needs to be recruited in future studies to investigate the association of type V and XI collagen genes with ACL rupture.

5.5 Future Work

As mentioned throughout this dissertation, the findings of this study could be improved upon by increased sample sizes. The barriers to including more participants in this dissertation were purely pragmatic: the number of participants was determined by the number of consenting participants that could be recruited from collaborating doctors during the time-frame of this dissertation. Unfortunately, the more susceptible female ACL rupture group partake in less high-risk sporting activities and therefore injure less frequently than males, and are more difficult to recruit.

However, research into the genetic determinants of musculoskeletal injury is ongoing and the intention is to continue recruiting participants in order to detect small genetic effects in the future. In order to achieve this, additional collaborations are being made with local doctors and international collaborators. Polymorphisms that show the potential to be involved in the aetiology of musculoskeletal injury, such as those identified in this dissertation (*COL5A1*, rs71746744 and rs1134170; *MIR608*, rs4919510; *COL11A1,* rs3753841 and rs1676486 and *COL11A2* rs1799907), will be re-analysed in a larger cohort in the future.

In addition, information was collected on extrinsic risk factors from all recruited participants, and that data is captured for future analysis. Unfortunately, at this stage the sample size is still too small to study complex interactions. However, it is an aim of the research group to continue to refine our understanding of the interactions of extrinsic and intrinsic risk factors in future studies.

To investigate the hypothesized effects of these polymorphisms on tendon and ligament structure and biomechanical properties, imaging work and biomechanical testing should be undertaken. In order to overcome the barriers of tendon and ligament heterogeneity and the confounding effects of age, physical activity and other extrinsic factors on tissue structure and tensile properties, the effect on fibril diameter and biomechanical properties may initially be best determined using a tissue engineering approach, in which the effect of polymorphisms of interest on fibril diameter and biomechanical properties can be tested in controlled conditions.

This dissertation has also revealed that *COL11A1* and *COL11A2* may play a far more important role in tendons and ligaments than previously suspected. The presence of type XI collagen as well as the proportions and combinations of type V and type XI collagen chains should therefore be investigated. The analysis many involve a combination of molecular biology techniques such as qRT-PCR to detect gene expression, Western Blot analysis to detect the presence and proportions of the different type V and type XI α-chains and *in situ* hybridization and immunofluorescence microscopy to confirm their presence and location in tendons and ligaments.

5.6 Clinical Significance

The aetiology of musculoskeletal soft tissue injuries remains unclear, and the molecular mechanisms that contribute to these conditions are likely to be intricate, involving the cumulative biological effect of genes in many different biological pathways.

The clinical significance of the identification of genes that predispose to tendon and ligament injuries is three-fold: (1) it is valuable to predict in advance whether an individual has an increased risk of injury as a result of repetitive loading during either sporting or occupational activity; (2) understanding the molecular mechanism may be useful for determining which therapeutic interventions may be effective in treatment of the injury, and (3), the genetic risk factors may be important in defining the prognosis of the injury for the individual[102]. Therefore, although the identification of an interaction between the genes encoding the minor type V/XI collagens has no immediate clinical applications, it is important as it provides additional evidence that inter-individual variations in collagen fibril assembly might be an important molecular mechanism in the aetiology of musculoskeletal soft tissue injuries.

5.7 Final Remarks

In conclusion, this dissertation further explored previous findings that implicated the *COL5A1* gene in musculoskeletal soft tissue injuries. In addition, we found that genetic variants in *COL11A1*, *COL11A2* and *COL5A1* can interact to increase or modulate risk. Because of functional redundancy between type V and type XI collagen, it is useful to examine variants in the genes encoding both proteins, as these genes may have compensatory or cumulative effects on musculoskeletal injury risk.

References

1. Järvinen, T. A., Kannus, P., Paavola, M., Järvinen, T.L., Józsa, L. & Järvinen, M. Achilles tendon injuries. *Curr. Opin. Rheumatol.* **13,** 150–155 (2001).

2. Meeuwisse, W. H. Assessing Causation in Sport Injury: A Multifactorial Model. *Clin. J. Sport Med.* **4,** 166–170 (1994).

3. Collins, M. & Raleigh, S. M. Genetic Risk Factors for Musculoskeletal Soft Tissue Injuries *Med. Sport Sci.* (Collins, M.) **54,** 136–149 (KARGER, 2009).

4. Collins, M. Genetic risk factors for soft-tissue injuries 101: a practical summary to help clinicians understand the role of genetics and 'personalised medicine'. *Br. J. Sports Med.* **44,** 915–917 (2009).

5. Ribbans, W. J. & Collins, M. Pathology of the tendo Achillis: Do our genes contribute? *Bone Jt. J.* **95-B,** 305–313 (2013).

6. Amis, A. A. & Dawkins, G. P. Functional Anatomy of the Anterior Cruciate Ligament Fibre Bundle Actions Related to Ligament Replacements and Injuries. *J. Bone Jt. Surg.* **73-B,** 260–267 (1991).

7. Duthon, V. B., Barea, C., Abrassart, S., Fasel, J. H., Fritschy, D., & Ménétrey, J. Anatomy of the anterior cruciate ligament. *Knee Surg. Sports Traumatol. Arthrosc.* **14,** 204–213 (2006).

8. Kannus, P. Structure of the tendon connective tissue. *Scand. J. Med. Sci. Sports* **10,** 312–320 (2000).

9. Frank, C. B. Ligament structure, physiology and function. *J. Musculoskelet. Neuronal Interact.* **4,** 199 (2004).

10. Kirkendall, D. T. & Garrett, W. E. Function and biomechanics of tendons. *Scand. J. Med. Sci. Sports* **7,** 62–66 (1997).

11. O'Brien, M. Structure and metabolism of tendons. *Scand. J. Med. Sci. Sports* **7,** 55–61 (1997).

12. Kastelic, J., Galeski, A. & Baer, E. The multicomposite structure of tendon. *Connect. Tissue Res.* **6,** 11–23 (1978).

13. Andrades, J. A., Claros, S., -Palomo, P. , J. M., Zamora-Navas, P., Guerado, E., , M., Araque, M. C. & Becerra, J. in *Regen. Med. Tissue Eng. - Cells Biomater.* (Eberli, D.) (2011). ISBN: 978-953-307-663-8, InTech, DOI: 10.5772/20889. at http://www.intechopen.com/books/regenerative-medicine-and-tissue-engineering-cells-and-biomaterials/skeletal-regeneration-by-mesenchymal-stem-cells-what-else

14. Riley, G. Tendinopathy--from basic science to treatment. *Nat. Clin. Pract. Rheumatol.* **4,** 82–89 (2008).

15. Riley, G. The pathogenesis of tendinopathy. A molecular perspective. *Rheumatol.* **43,** 131–142 (2004).

16. Eastwood, M., McGrouther, D. A. & Brown, R. A. Fibroblast responses to mechanical forces. *Proc. Inst. Mech. Eng. [H]* **212,** 85–92 (1998).

17. Kadler, K. E., Baldock, C., Bella, J. & Boot-Handford, R. P. Collagens at a glance. *J. Cell Sci.* **120,** 1955 –1958 (2007).

18. Gordon, M. K. & Hahn, R. A. Collagens. *Cell Tissue Res.* **339,** 247–257 (2009).

19. Wenstrup, R. J., Florer, J. B., Cole, W. G., Willing, M. C. & Birk, D. E. Reduced type I collagen utilization: a pathogenic mechanism in *COL5A1* haplo-insufficient Ehlers-Danlos syndrome. *J. Cell. Biochem.* **92,** 113–124 (2004).

20. Wenstrup, R. J., Florer, J. B., Brunskill, E. W., Bell, S. M., Chervoneva, I. & Birk, D.E. Type V Collagen Controls the Initiation of Collagen Fibril Assembly. *J. Biol. Chem.* **279,** 53331–53337 (2004).

21. Wenstrup, R. J. Florer, J. B., Davidson, J. M., Phillips, C. L., Pfeiffer, B.J., Menezes, D. W., Chervoneva, I. & Birk, D. E. Murine Model of the Ehlers-Danlos Syndrome: *COL5A1* haploinsufficieny disrupts collagen fibril assembly at multiple stages. *J. Biol. Chem.* **281,** 12888–12895 (2006).

22. Fitch, J. M., Gross, J., Mayne, R., Johnson-Wint, B. & Linsenmayer, T. F. Organization of collagen types I and V in the embryonic chicken cornea: monoclonal antibody studies. *Proc. Natl. Acad. Sci.* **81,** 2791 (1984).

23. Birk, D. E., Fitch, J. M., Babiarz, J. P., Doane, K. J. & Linsenmayer, T. F. Collagen fibrillogenesis in vitro: interaction of types I and V collagen regulates fibril diameter. *J. Cell Sci.* **95 (Pt 4),** 649–657 (1990).

24. Linsenmayer, T. F., Gibney, E., Igoe, F., Gordon, M. K., Fitch, J. M., Fessler, L. I. & Birk, D.E. Type V collagen: molecular structure and fibrillar organization of the chicken alpha 1 (V) NH2-terminal domain, a putative regulator of corneal fibrillogenesis. *J. Cell Biol.* **121,** 1181 (1993).

25. Marchant, J. K., Hahn, R. A., Linsenmayer, T. F. & Birk, D. E. Reduction of type V collagen using a dominant-negative strategy alters the regulation of fibrillogenesis and results in the loss of corneal-specific fibril morphology. *J. Cell Biol.* **135,** 1415 (1996).

26. Birk, D. E. Type V collagen: heterotypic type I/V collagen interactions in the regulation of fibril assembly. *Micron* **32,** 223–237 (2001).

27. Chanut-Delalande, H. Fichard, A., Bernocco, S., Garrone, R., Hulmes, D. J. & Ruggiero, F. Control of heterotypic fibril formation by collagen V is determined by chain stoichiometry. *J. Biol. Chem.* **276,** 24352 (2001).

28. Fichard, A., Kleman, J. P. & Ruggiero, F. Another look at collagen V and XI molecules. *Matrix Biol. J. Int. Soc. Matrix Biol.* **14,** 515–531 (1995).

29. Mendler, M., Eich-Bender, S. G., Vaughan, L., Winterhalter, K. H. & Bruckner, P. Cartilage contains mixed fibrils of collagen types II, IX, and XI. *J. Cell Biol.* **108,** 191–197 (1989).

30. Sagarriga Visconti, C., Kavalkovich, K., Wu, J., & Niyibizi, C. Biochemical Analysis of Collagens at the Ligament–Bone Interface Reveals Presence of Cartilage-Specific Collagens. *Arch. Biochem. Biophys.* **328,** 135–142 (1996).

31. Gelse, K., Pöschl, E. & Aigner, T. Collagens--structure, function, and biosynthesis. *Adv. Drug Deliv. Rev.* **55,** 1531–1546 (2003).

32. Zhang, G., Young, B. B., Ezura, Y., Favata, M., Soslowsky, L.J., Chakravarti, S. & Birk, D. E. Development of tendon structure and function: regulation of collagen fibrillogenesis. *J Musculoskelet Neuronal Interact* **5,** 5–21 (2005).

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

34. Engel, J. & Prockop, D. J. The zipper-like folding of collagen triple helices and the effects of mutations that disrupt the zipper. *Annu. Rev. Biophys. Biophys. Chem.* **20,** 137–152 (1991).

35. Doege, K. J. & Fessler, J. H. Folding of carboxyl domain and assembly of procollagen I. *J. Biol. Chem.* **261,** 8924–8935 (1986).

36. Van der Rest, M. & Dublet, B. Type XII and type XIV collagens: interfibrillar constituents of dense connective tissues. *Semin. Cell Dev. Biol.* **7,** 639–648 (1996).

37. Canty, E. G., Lu, Y., Meadows, R. S., Shaw, M. K., Holmes, D. F. & Kadler, K. E. Coalignment of plasma membrane channels and protrusions (fibripositors) specifies the parallelism of tendon. *J. Cell Biol.* **165,** 553–563 (2004).

38. Birk, D. E. & Trelstad, R. L. Extracellular compartments in tendon morphogenesis: collagen fibril, bundle, and macroaggregate formation. *J. Cell Biol.* **103,** 231–240 (1986).

39. Kalamajski, S. & Oldberg, Å. The role of small leucine-rich proteoglycans in collagen fibrillogenesis. *Matrix Biol.* **29,** 248–253 (2010).

40. Romanic, A. M., Adachi, E., Kadler, K. E., Hojima, Y. & Prockop, D. J. Copolymerization of pNcollagen III and collagen I. pNcollagen III decreases the rate of incorporation of collagen I into fibrils, the amount of collagen I incorporated, and the diameter of the fibrils formed. *J. Biol. Chem.* **266,** 12703–12709 (1991).

41. Linsenmayer, T. F., Fitch, J. M., Schmid, T. M. Zak, N. B., Gibney, E., Sanderson, R. D. & Mayne, R. Monoclonal antibodies against chicken type V collagen: production, specificity, and use for immunocytochemical localization in embryonic cornea and other organs. *J. Cell Biol.* **96,** 124 (1983).

42. Holmes, D. F. & Kadler, K. E. The 10+4 microfibril structure of thin cartilage fibrils. *Proc. Natl. Acad. Sci.* **103,** 17249–17254 (2006).

43. Keene, D. R., Oxford, J. T. & Morris, N. P. Ultrastructural localization of collagen types II, IX, and XI in the growth plate of human rib and fetal bovine epiphyseal cartilage: type XI collagen is restricted to thin fibrils. *J. Histochem. Cytochem.* **43,** 967–979 (1995).

44. Burgeson, R. E. & Hollister, D. W. Collagen heterogeneity in human cartilage: identification of several new collagen chains. *Biochem. Biophys. Res. Commun.* **87,** 1124–1131 (1979).

45. Bernard, M. Yoshioka, H., Rodriguez, E., Van der Rest, M., Kimura, T., Ninomiya, Y., Olsen, B. R. & Ramirez, F. Cloning and sequencing of pro-alpha 1 (XI) collagen cDNA demonstrates that type

XI belongs to the fibrillar class of collagens and reveals that the expression of the gene is not restricted to cartilagenous tissue. *J. Biol. Chem.* **263,** 17159–17166 (1988).

46. Imamura, Y., Scott, I. C. & Greenspan, D. S. The pro-alpha3(V) collagen chain. Complete primary structure, expression domains in adult and developing tissues, and comparison to the structures and expression domains of the other types V and XI procollagen chains. *J. Biol. Chem.* **275,** 8749–8759 (2000).

47. Sandberg, M. M., Hirvonen, H. E., Elima, K. J. & Vuorio, E. I. Co-expression of collagens II and XI and alternative splicing of exon 2 of collagen II in several developing human tissues. *Biochem. J.* **294 (Pt 2),** 595–602 (1993).

48. Lui, V. C., Kong, R. Y., Nicholls, J., Cheung, A. N. & Cheah, K. S. The mRNAs for the three chains of human collagen type XI are widely distributed but not necessarily co-expressed: implications for homotrimeric, heterotrimeric and heterotypic collagen molecules. *Biochem. J.* **311 (Pt 2),** 511–516 (1995).

49. Kania, A. M. Reichenberger, E., Baur, S. T., Karimbux, N. Y., Taylor, R. W., Olsen, B. R. & Nishimura, I. Structural variation of type XII collagen at its carboxyl-terminal NC1 domain generated by tissue-specific alternative splicing. *J. Biol. Chem.* **274,** 22053–22059 (1999).

50. Young, B. B., Zhang, G., Koch, M. & Birk, D. E. The roles of types XII and XIV collagen in fibrillogenesis and matrix assembly in the developing cornea. *J. Cell. Biochem.* **87,** 208–220 (2002).

51. Ansorge, H. L., Meng, X., Zhang, G., Veit, G., Sun, M., Klement, J.F., Beason, D. P., Soslowsky, L. J., Koch, M. & Birk, D.E. Type XIV Collagen Regulates Fibrillogenesis: premature collagen fibril growth and tissue dysfunction in null mice. *J. Biol. Chem.* **284,** 8427–8438 (2008).

52. Nishiyama, T., McDonough, A. M., Bruns, R. R. & Burgeson, R. E. Type XII and XIV collagens mediate interactions between banded collagen fibers in vitro and may modulate extracellular matrix deformability. *J. Biol. Chem.* **269,** 28193–28199 (1994).

53. Hulmes, D. J. S., Kadler, K. E., Mould, A. P., Hojima, Y., Holmes, D. F., Cummings, C., Chapman, J. A. & Prockop, D. J. Pleomorphism in type I collagen fibrils produced by persistence of the procollagen N-propeptide. *J. Mol. Biol.* **210,** 337–345 (1989).

54. Font, B., Eichenberger, D., Rosenberg, L. M. & Van Der Rest, M. Characterization of the interactions of type XII collagen with two small proteoglycans from fetal bovine tendon, decorin and fibromodulin. *Matrix Biol.* **15,** 341–348 (1996).

55. Font, B., Aubert-Foucher, E., Goldschmidt, D., Eichenberger, D. & van der Rest, M. Binding of collagen XIV with the dermatan sulfate side chain of decorin. *J. Biol. Chem.* **268,** 25015–25018 (1993).

56. Brown, J. C., Mann, K., Wiedemann, H. & Timpl, R. Structure and binding properties of collagen type XIV isolated from human placenta. *J. Cell Biol.* **120,** 557–567 (1993).

57. Varki A, Cummings RD, Esko JD, Freeze, H., Stanley, P. Bertozzi, C. R. Hart, G. W. & Etzler, M. E. *Essentials of Glycobiology*. (Cold Spring Harbor Laboratory Press, 2009).

58. Cribb, A. M. & Scott, J. E. Tendon response to tensile stress: an ultrastructural investigation of collagen: proteoglycan interactions in stressed tendon. *J. Anat.* **187,** 423 (1995).

59. Zhang, G., Ezura, Y., Chervoneva, I., Robinson, P. S., Beason, D. P., Carine, E. T., Soslowsky, L. J., Iozzo, R. V. & Birk, D. E. Decorin regulates assembly of collagen fibrils and acquisition of biomechanical properties during tendon development. *J. Cell. Biochem.* **98,** 1436–1449 (2006).

60. Danielson, K. G., Baribault, H., Holmes, D. F., Graham, H., Kadler, K. E. & Iozzo, R. V. Targeted disruption of decorin leads to abnormal collagen fibril morphology and skin fragility. *J. Cell Biol.* **136,** 729–743 (1997).

61. Lechner, B. E., Lim, J. H., Mercado, M. L. & Fallon, J. R. Developmental regulation of biglycan expression in muscle and tendon. *Muscle Nerve* **34,** 347–355 (2006).

62. Corsi, A., Xu, T., Chen, X. D., Boyde, A., Liang, J., Mankani, M., Sommer, B., Iozzo, R. V., Eichstetter, I., Robey, P. G., Bianco, P. & Young, M. F. Phenotypic effects of biglycan deficiency are linked to collagen fibril abnormalities, are synergized by decorin deficiency, and mimic Ehlers-Danlos-like changes in bone and other connective tissues. *J. Bone Miner. Res.* **17,** 1180–1189 (2002).

63. Ameye, L., Aria, D., Jepsen, K., Oldberg, A., Xu, T. & Young, M. F. Abnormal collagen fibrils in tendons of biglycan/fibromodulin-deficient mice lead to gait impairment, ectopic ossification, and osteoarthritis. *Faseb J.* **16,** 673–680 (2002).

64. Svensson, L., Närlid, I. & Oldberg, A. Fibromodulin and lumican bind to the same region on collagen type I fibrils. *Febs Lett.* **470,** 178–182 (2000).

65. Jepsen, K. J., Wu, F., Peragallo, J. H., Paul, J., Roberts, L., Ezura, Y., Oldberg, A., Birk, D. E. & Chakravarti, S. A Syndrome of Joint Laxity and Impaired Tendon Integrity in Lumican- and Fibromodulin-deficient Mice. *J. Biol. Chem.* **277,** 35532–35540 (2002).

66. Ezura, Y., Chakravarti, S., Oldberg, A., Chervoneva, I. & Birk, D. E. Differential expression of lumican and fibromodulin regulate collagen fibrillogenesis in developing mouse tendons. *J. Cell Biol.* **151,** 779–788 (2000).

67. Svensson, L., Aszódi, A., Reinholt, F. P., Fässler, R., Heinegård, D. & Oldberg, A. Fibromodulin-null mice have abnormal collagen fibrils, tissue organization, and altered lumican deposition in tendon. *J. Biol. Chem.* **274,** 9636–9647 (1999).

68. Chakravarti, S., Magnuson, T., Lass, J. H., Jepsen, K. J., LaMantia, C. & Carroll, H. Lumican Regulates Collagen Fibril Assembly: Skin Fragility and Corneal Opacity in the Absence of Lumican. *J. Cell Biol.* **141,** 1277–1286 (1998).

69. Ottani, V., Raspanti, M. & Ruggeri, A. Collagen structure and functional implications. *Micron. 1993* **32,** 251–260 (2001).

70. Woo, S. L., Debski, R. E., Zeminski, J., Abramowitch, S. D., Saw, S. S. & Fenwick, J. A. Injury and repair of ligaments and tendons. *Annu. Rev. Biomed. Eng.* **2,** 83–118 (2000).

71. Silver, F. H., Freeman, J. W. & Seehra, G. P. Collagen self-assembly and the development of tendon mechanical properties. *J. Biomech.* **36,** 1529–1553 (2003).

72. Wang, J. H.-C. Mechanobiology of tendon. *J. Biomech.* **39,** 1563–1582 (2006).

73. Parry, D. A. D., Barnes, G. R. G. & Craig, A. S. A comparison of the size distribution of collagen fibrils in connective tissues as a function of age and a possible relation between fibril size distribution and mechanical properties. *Proc. R. Soc. Lond. B Biol. Sci.* **203,** 305 (1978).

74. Rees, J. D. Current concepts in the management of tendon disorders. *Rheumatology* **45,** 508–521 (2006).

75. Griffin, L. Y., Albohm, M. J., Arendt, E. A., Bahr, R., Beynnon, B. D., Demaio, M., Dick, R. W., Engebretsen, L., Garrett, W. E. Jr., 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. Jr., Shultz, S. J., Silvers, H., Slauterbeck, J., Taylor, D. C., Teitz, C. C., Wojtys, E. M. & Yu, B. Understanding and preventing noncontact anterior cruciate ligament injuries: a review of the Hunt Valley II meeting, January 2005. *Am. J. Sports Med.* **34,** 1512–1532 (2006).

76. Järvinen, T. A. H., Kannus, P., Maffulli, N. & Khan, K. M. Achilles Tendon Disorders: Etiology and Epidemiology. *Foot Ankle Clin.* **10,** 255–266 (2005).

77. Järvinen, M., Józsa, L., Kannus, P., Järvinen, T. L., Kvist, M. & Leadbetter, W. Histopathological findings in chronic tendon disorders. *Scand. J. Med. Sci. Sports* **7,** 86–95 (2007).

78. Fredberg, U. & Stengaard-Pedersen, K. Chronic tendinopathy tissue pathology, pain mechanisms, and etiology with a special focus on inflammation. *Scand. J. Med. Sci. Sports* **18,** 3–15 (2008).

79. Mazzone, M. F. & McCue, T. Common conditions of the Achilles tendon. *Am. Fam. Physician* **65,** 1805–1810 (2002).

80. Posthumus, M., lins, M., September, A. & Schwellnus, M. The Intrinsic Risk Factors for ACL Ruptures: An Evidence-Based Review. *Phys. Sportsmed.* **39,** 62–73 (2011).

81. Kousta, E., Papathanasiou, A. & Skordis, N. Sex determination and disorders of sex development according to the revised nomenclature and classification in 46 XX individuals. *Horm. Athens Greece* **9,** 218–131 (2010).

82. Peeters, M. W., Thomis, M. A., Loos, R. J., Derom, C. A., Fagard, R., Claessens, A. L., Vlietinck, R. F. & Beunen, G. P. Heritability of somatotype components: a multivariate analysis. *Int. J. Obes.* **31,** 1295–1301 (2007).

83. Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., Madden, P. A., Heath, A. C., Martin, N. G., Montgomery, G. W., Goddard, M. E. & Visscher, P. M. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42,** 565–569 (2010).

84. Battié, M. C., Levalahti, E., Videman, T., Burton, K. & Kaprio, J. Heritability of lumbar flexibility and the role of disc degeneration and body weight. *J. Appl. Physiol. Bethesda Md 1985* **104,** 379–385 (2008).

85. Maes, H. H. M., Beunen, G. P., Vlietinck, R. F., Neale, M. C., Thomis, M., Vanden Eynde, B., Lysens, R., Simons, J., Derom, C. & Derom, R. Inheritance of physical fitness in 10-yr-old twins and their parents. *Med. Amp Sci. Sports Amp Exerc.* **28,** 1479–1491 (1996).

86. Falconer, D. S. The inheritance of liability to certain diseases, estimated from the incidence among relatives. *Ann. Hum. Genet.* **29,** 51–76 (1965).

87. Gennari, L., Merlotti, D., De Paola, V., Calabrò, A., Becherini, L., Martini, G. & Nuti, R. Estrogen Receptor Gene Polymorphisms and the Genetics of Osteoporosis: A HuGE Review. *Am. J. Epidemiol.* **161,** 307–320 (2005).

88. Dunning, A. M., Dowsett, M., Healey, C. S., Tee, L., Luben, R. N., Folkerd, E., Novik, K. L., Kelemen, L., Ogata, S., Pharoah, P. D., Easton, D. F., Day, N. E. & Ponder, B. A. Polymorphisms Associated With Circulating Sex Hormone Levels in Postmenopausal Women. *Jnci J. Natl. Cancer Inst.* **96,** 936–945 (2004).

89. Tucker, R. & Collins, M. What makes champions? A review of the relative contribution of genes and training to sporting success. *Br. J. Sports Med.* **46,** 555–561 (2012).

90. Flynn, R. K., Pedersen, C. L., Birmingham, T. B., Kirkley, A., Jackowski, D. & Fowler, P. J. The Familial Predisposition Toward Tearing the Anterior Cruciate Ligament: A Case Control Study. *Am. J. Sports Med.* **33,** 23–28 (2005).

91. Goshima, K., Kitaoka, K., Nakase, J., Takahashi, R. & Tsuchiya, H. Clinical evidence of a familial predisposition to anterior cruciate ligament injury. *Br. J. Sports Med.* **45,** 350 (2011).

92. Harner, C. D., Paulos, L. E., Greenwald, A. E., Rosenberg, T. D. & Cooley, V. C. Detailed analysis of patients with bilateral anterior cruciate ligament injuries. *Am. J. Sports Med.* **22,** 37–43 (1994).

93. Kraemer, R., Wuerfel, W., Lorenzen, J., Busche, M., Vogt, P. M. & Knobloch, K. Analysis of hereditary and medical risk factors in Achilles tendinopathy and Achilles tendon ruptures: a matched pair analysis. *Arch. Orthop. Trauma Surg.* **132,** 847–853 (2012).

94. Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., McCarthy, M. I., Ramos, E. M., Cardon, L. R., Chakravarti, A., Cho, J. H., Guttmacher, A. E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C. N., Slatkin, M., Valle, D., Whittemore, A. S., Boehnke, M., Clark, A. G., Eichler, E. E., Gibson, G., Haines, J. L., Mackay, T. F., McCarroll, S. A. & Visscher, P.M. Finding the missing heritability of complex diseases. *Nature* **461,** 747–753 (2009).

95. Posthumus, M., Saunders, C., September, A. V. & Collins, M. The polygenic profiles in participants with Achilles tendinopathy and controls. *Br. J. Sports Med.* **45,** 369–369 (2011).

96. September, A. V., Nell, E. M., O'Connell, K., Cook, J., Handley, C. J., van der Merwe, L., Schwellnus, M. & Collins, M. A pathway-based approach investigating the genes encoding interleukin-1β, interleukin-6 and the interleukin-1 receptor antagonist provides new insight into the genetic susceptibility of Achilles tendinopathy. *Br. J. Sports Med.* **45,** 1040–1047 (2011).

97. Posthumus, M. Collins, M., van der Merwe, L., O'Cuinneagain, D., van der Merwe, W., Ribbans, W. J., Schwellnus, M. P., Raleigh, S. M. Matrix metalloproteinase genes on chromosome 11q22 and the risk of anterior cruciate ligament (ACL) rupture. *Scand. J. Med. Sci. Sports* **22,** 523–533 (2012).

98. Raleigh, S. M., van der Merwe, L., Ribbans, W. J., Smith, R. K., Schwellnus, M. P. & Collins, M. Variants within the MMP3 gene are associated with Achilles tendinopathy: possible interaction with the *COL5A1* gene. *Br. J. Sports Med.* **43,** 514–520 (2008).

99. Nell, E. M., van der Merwe, L., Cook, J., Handley, C. J., Collins, M., & September, A. V. The apoptosis pathway and the genetic predisposition to Achilles tendinopathy. *J. Orthop. Res.* **30,** 1719–1724 (2012).

100. Mokone, G. G., Gajjar, M., September, A. V., Schwellnus, M. P., Greenberg, J, Noakes, T. D. & Collins, M. The Guanine-Thymine Dinucleotide Repeat Polymorphism Within the Tenascin-C Gene Is Associated With Achilles Tendon Injuries. *Am. J. Sports Med.* **33,** 1016–1021 (2005).

101. Mokone, G. G., Schwellnus, M. P., Noakes, T. D. & Collins, M. The *COL5A1* gene and Achilles tendon pathology. *Scand. J. Med. Sci. Sports* **16,** 19–26 (2006).

102. September, A. V., Cook, J., Handley, C. J., van der Merwe, L., Schwellnus, M. P. & Collins, M. Variants within the *COL5A1* gene are associated with Achilles tendinopathy in two populations. *Br. J. Sports Med.* **43,** 357–365 (2009).

103. Posthumus, M., September, A. V., O'Cuinneagain, D., van der Merwe, W., Schwellnus, M. P. & Collins, M. The *COL5A1* gene is associated with increased risk of anterior cruciate ligament ruptures in female participants. *Am. J. Sports Med.* **37,** 2234–2240 (2009).

104. Abrahams, Y., Laguette, M. J., Prince, S. & Collins, M. Polymorphisms within the *COL5A1* 3′ -UTR that alters mRNA structure and the *MIR608* Gene are associated with Achilles tendinopathy. *Ann. Hum. Genet.* (2013). doi:10.1111/ahg.12013

105. Khoschnau, S., Melhus, H., Jacobson, A., Rahme, H., Bengtsson, H., Ribom, E., Grundberg, E., Mallmin, H. & Michaëlsson, K. Type I Collagen 1 Sp1 Polymorphism and the risk of cruciate ligament ruptures or shoulder dislocations. *Am. J. Sports Med.* **36,** 2432–2436 (2008).

106. Posthumus, M., September, A. V., Keegan, M., O'Cuinneagain, D., Van der Merwe, W., Schwellnus, M. P. & Collins, M. Genetic risk factors for anterior cruciate ligament ruptures: *COL1A1* gene variant. *Br. J. Sports Med.* **43,** 352–356 (2009).

107. 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. & Eynon, N. Gene variants within the *COL1A1* gene are associated with reduced anterior cruciate ligament injury in professional soccer players. *J. Sci. Med. Sport* (2012). doi:10.1016/j.jsams.2012.10.004

108. Posthumus, M., September, A. V., O'Cuinneagain, D., van der Merwe, W., Schwellnus, M. P. & Collins, M. The association between the *COL12A1* gene and anterior cruciate ligament ruptures. *Br. J. Sports Med.* **44,** 1160–1165 (2009).

109. Posthumus, M., Collins, M., Cook, J., Handley, C. J., Ribbans, W. J., Smith, R. K., Schwellnus, M. P. & Raleigh, S. M. Components of the transforming growth factor-beta family and the pathogenesis of human Achilles tendon pathology- a genetic association study. *Rheumatol.* **49,** 2090–2097 (2010).

110. Collins, M. & Posthumus, M. Type V Collagen Genotype and Exercise-Related Phenotype Relationships: A Novel Hypothesis. *Exerc. Sport Sci. Rev.* **39**, 191-198 (2011).

111. Malfait, F., Wenstrup, R. J. & De Paepe, A. Clinical and genetic aspects of Ehlers-Danlos syndrome, classic type. *Genet. Med.* **12,** 597–605 (2010).

112. Mitchell, A. L., Schwarze, U., Jennings, J. F. & Byers, P. H. Molecular mechanisms of classical Ehlers-Danlos syndrome (EDS). *Hum. Mutat.* **30,** 995–1002 (2009).

113. Caridi, G., Pezzolo, A., Bertelli, R., Gimelli, G., Di Donato, A., Candiano, G. & Ghiggeri, G. M. Mapping of the human *COL5A1* gene to chromosome 9q34.3. *Hum. Genet.* **90,** 174–176 (1992).

114. Takahara, K., Hoffman, G. G. & Greenspan, D. S. Complete structural organization of the human alpha 1 (V) collagen gene (*COL5A1*): divergence from the conserved organization of other characterized fibrillar collagen genes. *Genomics* **29,** 588–597 (1995).

115. Collins, M., Mokone, G. G., September, A. V., van der Merwe, L. & Schwellnus, M. P. The *COL5A1* genotype is associated with range of motion measurements. *Scand. J. Med. Sci. Sports* **19,** 803–810 (2009).

116. Brown, J. C., Miller, C. J., Schwellnus, M. P. & Collins, M. Range of motion measurements diverge with increasing age for *COL5A1* genotypes. *Scand. J. Med. Sci. Sports* e266-e272 (2011).

117. Posthumus, M., Schwellnus, M. P. & Collins, M. The *COL5A1* Gene: A novel marker of endurance running performance. *Med. Sci. Sports Exerc.* **43,** 584–589 (2011).

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

119. Conne, B., Stutz, A. & Vassalli, J. D. The 3' untranslated region of messenger RNA: A molecular `hotspot' for pathology? *Nat Med* **6,** 637–641 (2000).

120. Fabian, M. R., Sonenberg, N. & Filipowicz, W. Regulation of mRNA translation and stability by microRNAs. *Annu. Rev. Biochem.* **79,** 351–379 (2010).

121. Snead, M. P. & Yates, J. R. Clinical and Molecular genetics of Stickler syndrome. *J. Med. Genet.* **36,** 353–359 (1999).

122. Shanske, A. L., Bogdanow, A., Shprintzen, R. J. & Marion, R. W. The Marshall syndrome: report of a new family and review of the literature. *Am. J. Med. Genet.* **70,** 52–57 (1997).

123. Giedion, A., Brandner, M., Lecannellier, J., Muhar, U., Prader, A., Sulzer, J., Zweymüller, E. Oto-spondylo-megaepiphyseal dysplasia (OSMED). *Helv. Paediatr. Acta* **37,** 361–380 (1982).

124. Weissenbacher, G. & Zweymueller, E. [Simultaneous occurrance of the Pierre Robin syndrome and fetal chondrodysplasia]. *Monatsschrift für kinderheilkd.* **112,** 315–317 (1964).

125. Spranger, J. The type XI collagenopathies. *Pediatr. Radiol.* **28,** 745–750 (1998).

126. Jakkula, E., Melkoniemi, M., Kiviranta, I., Lohiniva, J., Räinä, S. S., Perälä, M., Warman, M. L., Ahonen, K., Kröger, H., Göring, H. H. & Ala-Kokko, L. The role of sequence variations within the genes encoding collagen II, IX and XI in non-syndromic, early-onset osteoarthritis. *Osteoarthr. Cartil. Oars Osteoarthr. Res. Soc.* **13,** 497–507 (2005).

127. Ikeda, T., Mabuchi, A., Fukuda, A., Kawakami, A., Ryo, Y., Yamamoto, S., Miyoshi, K., Haga, N., Hiraoka, H., Takatori, Y., Kawaguchi, H., Nakamura, K. & Ikegawa, S. Association analysis of single nucleotide polymorphisms in cartilage-specific collagen genes with knee and hip osteoarthritis in the Japanese population. *J. Bone Miner. Res.* **17,** 1290–1296 (2002).

128. Videman, T., Saarela, J., Kaprio, J., Näkki, A., Levälahti, E., Gill, K., Peltonen, L. & Battié, M. C. Associations of 25 structural, degradative, and inflammatory candidate genes with lumbar disc desiccation, bulging, and height narrowing. *Arthritis Rheum.* **60,** 470–481 (2009).

129. Mio, F., Chiba, K., Hirose, Y., Kawaguchi, Y., Mikami, Y., Oya, T., Mori, M., Kamata, M., Matsumoto, M., Ozaki, K., Tanaka, T., Takahashi, A., Kubo, T., Kimura, T., Toyama, Y. & Ikegawa, S. A Functional Polymorphism in *COL11A1*, Which Encodes the α1 Chain of Type XI Collagen, is Associated with Susceptibility to Lumbar Disc Herniation. *Am. J. Hum. Genet.* **81,** 1271–1277 (2007).

130. Koyama, K., Nakazato, K., Min, S., Gushiken, K., Hatakeda, Y., Seo, K. & Hiranuma, K. *COL11A1* gene is associated with limbus vertebra in gymnasts. *Int. J. Sports Med.* **33,** 586–590 (2012).

131. Jugessur, A., Shi, M., Gjessing, H. K., Lie, R. T., Wilcox, A. J., Weinberg, C. R., Christensen, K., Boyles, A. L., Daack-Hirsch, S., Nguyen, T. T., Christiansen, L., Lidral, A. C. & Murray, J. C. Maternal genes and facial clefts in offspring: a comprehensive search for genetic associations in two population-based cleft studies from Scandinavia. *Plos One* **5,** e11493 (2010).

132. Nikopensius, T., Jagomägi, T., Krjutskov, K., Tammekivi, V., Saag, M., Prane, I., Piekuse, L., Akota, I., Barkane, B., Krumina, A., Ambrozaityte, L., Matuleviciene, A., Kucinskiene, Z. A., Lace, B., Kucinskas, V., Metspalu, A. Genetic variants in *COL2A1*, *COL11A2*, and *IRF6* contribute risk to nonsyndromic cleft palate. *Birt. Defects Res. A. Clin. Mol. Teratol.* **88,** 748–756 (2010).

133. Park, H. J., Choe, B. K., Kim, S. K., Park, H. K., Kim, J. W., Chung, J. H., Hong, I. K., Chung, D. H. & Kwon, K. H. Association between collagen type XI α1 gene polymorphisms and papillary thyroid cancer in a Korean population. *Exp. Ther. Med.* **2**, 1111-1116 (2011).

134. Lee, H. S., Lee, A. T., Criswell, L. A., Seldin, M. F., Amos, C. I., Carulli, J. P., Navarrete, C., Remmers, E. F., Kastner, D. L., Plenge, R. M., Li, W. & Gregersen, P. K. Several regions in the major histocompatibility complex confer risk for anti-CCP-antibody positive rheumatoid arthritis, independent of the DRB1 locus. *Mol. Med.* **14,** 293–300 (2008).

135. Vikkula, M., Mariman, E. C., Lui, V. C., Zhidkova, N. I., Tiller, G. E., Goldring, M. B., van Beersum S. E., de Waal Malefijt, M. C., van den Hoogen, F. H. & Ropers, H. H. Autosomal dominant and recessive osteochondrodysplasias associated with the *COL11A2* locus. *Cell* **80,** 431–437 (1995).

136. Maeda, S., Koga, H., Matsunaga, S., Numasawa, T., Ikari, K., Furushima, K., Harata, S., Takeda, J., Sakou, T., Komiya, S. & Inoue, I. Gender-specific haplotype association of collagen alpha 2 (XI) gene in ossification of the posterior longitudinal ligament of the spine. *J. Hum. Genet.* **46,** 1–4 (2001).

137. Maeda, S., Ishidou, Y., Koga, H., Taketomi, E., Ikari, K., Komiya, S., Takeda, J., Sakou, T. & Inoue I. Functional Impact of Human Collagen α2(XI) Gene Polymorphism in Pathogenesis of Ossification of the Posterior Longitudinal Ligament of the Spine. *J. Bone Miner. Res.* **16,** 948–957 (2001).

138. Horikoshi, T., Maeda, K., Kawaguchi, Y., Chiba, K., Mori, K., Koshizuka, Y., Hirabayashi, S., Sugimori, K., Matsumoto, M., Kawaguchi, H., Takahashi, M., Inoue, H., Kimura, T., Matsusue, Y., Inoue, I., Baba, H., Nakamura, K. & Ikegawa, S. A large-scale genetic association study of ossification of the posterior longitudinal ligament of the spine. *Hum. Genet.* **119,** 611–616 (2006).

139. Koga, H., Sakou, T., Taketomi, E., Hayashi, K., Numasawa, T., Harata, S., Yone, K., Matsunaga, S., Otterud, B., Inoue,I. & Leppert, M. Genetic mapping of ossification of the posterior longitudinal ligament of the spine. *Am. J. Hum. Genet.* 62, 1460–1467 (1998).

140. Karppinen, J., Daavittila, I., Solovieva, S., Kuisma, M., Taimela, S., Natri, A., . Haapea, M., Korpelainen, R., Niinimäki, J., Tervonen, O., Ala-Kokko, L., & Männikkö, M. Genetic factors are associated with modic changes in endplates of lumbar vertebral bodies. *Spine* **33,** 1236–1241 (2008).

141. N Noponen-Hietala, E Kyllonen & M Mannikko. Sequence variations in the collagen IX and XI genes are associated with degenerative lumbar spinal stenosis. *Ann. Rheum. Dis.* **62,** 1208–1214 (2003).

142. Solovieva, S., Lohiniva, J., Leino-Arjas, P., Raininko, R., Luoma, K., Ala-Kokko, L., & Riihimäki, H. Intervertebral disc degeneration in relation to the COL9A3 and the IL-1ss gene polymorphisms. *Eur. Spine J.* **15,** 613–619 (2006).

143. Virtanen, I. M., Karppinen, J., Taimela, S., Ott, J., Barral, S., Kaikkonen, K., Heikkilä, O., Mutanen, P., Noponen, N., Männikkö, M., Tervonen, O., Natri, A. & Ala-Kokko, L. Occupational and genetic risk factors associated with intervertebral disc disease. *Spine* **32,** 1129–1134 (2007).

144. Eskola, P. J., Kjaer, P., Daavittila, I. M., Solovieva, S., Okuloff, A., Sorensen, J. S., Wedderkopp, N., Ala-Kokko, L., Männikkö, M. & Karppinen, J. I. Genetic risk factors of disc degeneration among 12-14-year-old Danish children: a population study. *Int. J. Mol. Epidemiol. Genet.* **1,** 158–165 (2010).

145. Zhidkova, N. I., Justice, S. K. & Mayne, R. Alternative mRNA processing occurs in the variable region of the pro-alpha 1(XI) and pro-alpha 2(XI) collagen chains. *J. Biol. Chem.* **270,** 9486–9493 (1995).

146. Wenstrup, R. J., Smith, S. M., Florer, J. B., Zhang, G., Beason, D. P., Seegmiller, R. E., Soslowsky, L. J. & Birk, D. E. Regulation of Collagen Fibril Nucleation and Initial Fibril Assembly Involves Coordinate Interactions with Collagens V and XI in Developing Tendon. *J. Biol. Chem.* **286,** 20455–20465 (2011).

147. Kleman, J.-P., Hartmann, D. J., Ramirez, F. & Rest, M. The human rhabdomyosarcoma cell line A204 lays down a highly insoluble matrix composed mainly of alpha1 type-XI and alpha2 type-V collagen chains. *Eur. J. Biochem.* **210,** 329–335 (1992).

148. Mayne, R., Brewton, R. G., Mayne, P. M. & Baker, J. R. Isolation and characterization of the chains of type V/type XI collagen present in bovine vitreous. *J. Biol. Chem.* **268,** 9381–9386 (1993).

149. Niyibizi, C. & Eyre, D. R. Identification of the cartilage alpha 1(XI) chain in type V collagen from bovine bone. *Febs Lett.* **242,** 314–318 (1989).

150. Zantop, T., Petersen, W., Sekiya, J. K., Musahl, V. & Fu, F. H. Anterior cruciate ligament anatomy and function relating to anatomical reconstruction. *Knee Surg. Sports Traumatol. Arthrosc.* **14,** 982–992 (2006).

151. Bicer, E. K., Lustig, S., Servien, E., Selmi, T. A. S. & Neyret, P. Current knowledge in the anatomy of the human anterior cruciate ligament. *Knee Surg. Sports Traumatol. Arthrosc.* **18,** 1075–1084 (2009).

152. Kopf, S., Musahl, V., Tashman, S., Szczodry, M., Shen, W., & Fu, F. H. A systematic review of the femoral origin and tibial insertion morphology of the ACL. *Knee Surg. Sports Traumatol. Arthrosc.* **17,** 213–219 (2009).

153. Marshall, S.W., Padua, D., McGrath, M. Incidence of ACL injuries. In: Hewett TE, Schultz SJ, Griffin LY (eds.) American Orthopaedic Society for Sports Medicine. *Understanding and preventing noncontact ACL injuries*. (Human Kinetics, 2007).

154. Marshall, S. W. Recommendations for defining and classifying anterior cruciate ligament injuries in epidemiologic studies. *J. Athl. Train.* **45,** 516–518 (2010).

155. Ettlinger, C. F., Johnson, R. J. & Shealy, J. E. A Method to Help Reduce the Risk of Serious Knee Sprains Incurred in Alpine Skiing. *Am. J. Sports Med.* **23,** 531–537 (1995).

156. Fischer, J. F., Leyvraz, P. F. & Bally, A. A dynamic analysis of knee ligament injuries in alpine skiing. *Acta Orthop. Belg.* **60,** 194–203 (1994).

157. Geyer, M. & Wirth, C. J. [A new mechanism of injury of the anterior cruciate ligament]. *Unfallchirurg* **94,** 69–72 (1991).

158. Hame, S. L., Oakes, D. A. & Markolf, K. L. Injury to the anterior cruciate ligament during alpine skiing: a biomechanical analysis of tibial torque and knee flexion angle. *Am. J. Sports Med.* **30,** 537–540 (2002).

159. Gianotti, S. M., Marshall, S. W., Hume, P. A. & Bunt, L. Incidence of anterior cruciate ligament injury and other knee ligament injuries: a national population-based study. *J. Sci. Med. Sport Sports Med. Aust.* **12,** 622–627 (2009).

160. Granan, L.-P., Bahr, R., Steindal, K., Furnes, O. & Engebretsen, L. Development of a national cruciate ligament surgery registry: the Norwegian National Knee Ligament Registry. *Am. J. Sports Med.* **36,** 308–315 (2008).

161. Barenius, B., Forssblad, M., Engström, B. & Eriksson, K. Functional recovery after anterior cruciate ligament reconstruction, a study of health-related quality of life based on the Swedish National Knee Ligament Register. *Knee Surg. Sports Traumatol. Arthrosc.* **21**, 914-927 (2012).

162. De Loës, M., Dahlstedt, L. J. & Thomée, R. A 7-year study on risks and costs of knee injuries in male and female youth participants in 12 sports. *Scand. J. Med. Sci. Sports* **10,** 90–97 (2000).

163. Brooks, J. H. M. Epidemiology of injuries in English professional rugby union: part 1 match injuries. *Br. J. Sports Med.* **39,** 757–766 (2005).

164. Freedman, K. B., Glasgow, M. T., Glasgow, S. G. & Bernstein, J. Anterior cruciate ligament injury and reconstruction among university students. *Clin. Orthop.* 208–212 (1998).

165. Øiestad, B. E., Holm, I., Engebretsen, L., Aune, A. K., Gunderson, R., & Risberg, M. A. The prevalence of patellofemoral osteoarthritis 12 years after anterior cruciate ligament reconstruction. *Knee Surg. Sports Traumatol. Arthrosc.* (2012).

166. Lohmander, L. S., Englund, P. M., Dahl, L. L. & Roos, E. M. The Long-term Consequence of Anterior Cruciate Ligament and Meniscus Injuries: Osteoarthritis. *Am. J. Sports Med.* **35,** 1756–1769 (2007).

167. September, A. V., Schwellnus, M. P., Collins, M. & Gibson, W. Tendon and ligament injuries: the genetic component * COMMENTARY. *Br. J. Sports Med.* **41,** 241–246 (2007).

168. Raleigh, S. M., Posthumus, M., O'Cuinneagain, D., van der Merwe, W. & Collins, M. The GDF5 Gene and Anterior Cruciate Ligament Rupture. *Int. J. Sports Med.* **34**, 364-367 (2013).

169. Toth, A. P. & Cordasco, F. A. Anterior cruciate ligament injuries in the female athlete. *J. Gend.-Specif. Med.* **4,** 25–34 (2001).

170. Agel, J., Arendt, E. A. & Bershadsky, B. Anterior cruciate ligament injury in national collegiate athletic association basketball and soccer: a 13-year review. *Am. J. Sports Med.* **33,** 524–530 (2005).

171. Colgan, D. F. & Manley, J. L. Mechanism and regulation of mRNA polyadenylation. *Genes Dev.* **11,** 2755–2766 (1997).

172. Ross, J. mRNA stability in mammalian cells. *Microbiol. Rev.* **59,** 423–450 (1995).

173. Mazumder, B., Seshadri, V. & Fox, P. L. Translational control by the 3′-UTR: the ends specify the means. *Trends Biochem. Sci.* **28,** 91–98 (2003).

174. Xie, X., Lu, J., Kulbokas, E. J., Golub, T. R., Mootha, V., Lindblad-Toh, K., Lander, E. S., & Kellis, M. Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals. *Nature* **434,** 338–345 (2005).

175. Lau, N. C. & Lai, E. C. Diverse roles for RNA in gene regulation. *Genome Biol.* **6,** 315 (2005).

176. Matzke, M. A. & Birchler, J. A. RNAi-mediated pathways in the nucleus. *Nat. Rev. Genet.* **6,** 24–35 (2005).

177. Sayed, D. & Abdellatif, M. MicroRNAs in development and disease. *Physiol. Rev.* **91,** 827–887 (2011).

178. Brown, J. C., Miller, C.-J., Posthumus, M., Schwellnus, M. P. & Collins, M. The *COL5A1* gene, ultra-marathon running performance, and range of motion. *Int. J. Sports Physiol. Perform.* **6,** 485–496 (2011).

179. Lahiri, D. K. & Nurnberger, J. I., Jr. A rapid non-enzymatic method for the preparation of HMW DNA from blood for RFLP studies. *Nucleic Acids Res.* **19,** 5444 (1991).

180. Gauderman, W. J. Sample Size Requirements for Association Studies of Gene-Gene Interaction. *Am. J. Epidemiol.* **155,** 478–484 (2002).

181. Epstein, M. P. & Satten, G. A. Inference on Haplotype Effects in Case-Control Studies Using Unphased Genotype Data. *Am. J. Hum. Genet.* **73,** 1316–1329 (2003).

182. Satten, G. A. & Epstein, M. P. Comparison of prospective and retrospective methods for haplotype inference in case-control studies. *Genet. Epidemiol.* **27,** 192–201 (2004).

183. Gaunt, T. R., Rodríguez, S. & Day, I. N. Cubic exact solutions for the estimation of pairwise haplotype frequencies: implications for linkage disequilibrium analyses and a web tool 'CubeX'. *BMC Bioinformatics* **8,** 428 (2007).

184. Turnpenny P, Ellard S. in *Emerys Elem. Med. Genet.* 123–136 (Elsevier Churchill Livingstone, 2005).

185. Gibson, W. T. Genetic association studies for complex traits: relevance for the sports medicine practitioner. *Br. J. Sports Med.* **43,** 314–316 (2008).

186. Gonçalves-Neto, J., Witzel, S. S., Teodoro, W. R., Carvalho-Junior, A. E., Fernandes, T. D., & Yoshinari, H. H. Changes in collagen matrix composition in human posterior tibial tendon dysfunction. *Joint Bone Spine* **69,** 189–194 (2002).

187. Loudon, J. K., Jenkins, W. & Loudon, K. L. The relationship between static posture and ACL injury in female athletes. *J. Orthop. Sports Phys. Ther.* **24,** 91–97 (1996).

188. Ramesh, R., Von Arx, O., Azzopardi, T., & Schranz, P. J. The risk of anterior cruciate ligament rupture with generalised joint laxity. *J. Bone Jt. Surg. - Br. Vol.* **87-B,** 800–803 (2005).

189. Boden, B. P., Dean, G. S., Feagin, J. A., Jr & Garrett, W. E., Jr. Mechanisms of anterior cruciate ligament injury. *Orthopedics* **23,** 573–578 (2000).

190. Bell, R. D., Shultz, S. J., Wideman, L. & Henrich, V. C. Collagen Gene Variants Previously Associated With Anterior Cruciate Ligament Injury Risk Are Also Associated With Joint Laxity. *Sports Heal. Multidiscip. Approach* **4,** 312–318 (2012).

191. Morris, N. P. & Bächinger, H. P. Type XI collagen is a heterotrimer with the composition (1 alpha, 2 alpha, 3 alpha) retaining non-triple-helical domains. *J. Biol. Chem.* **262,** 11345–11350 (1987).

192. Griffith, A. J., Sprunger, L. K., Sirko-Osadsa, D. A., Tiller, G. E., Meisler, M. H., & Warman, M. L. Marshall Syndrome Associated with a Splicing Defect at the *COL11A1* Locus. *Am. J. Hum. Genet.* **62,** 816–823 (1998).

193. Annunen, S., Körkkö, J., Czarny, M., Warman, M. L., Brunner, H. G., Kääriäinen, H., Mulliken, J. B., Tranebjaerg, L., Brooks, D. G., Cox, G. F., Cruysberg, J. R., Curtis, M. A., Davenport, S. L., Friedrich, C. A., Kaitila, I., Krawczynski, M. R., Latos-Bielenska, A., Mukai, S., Olsen, B. R., Shinno, N., Somer, M., Vikkula, M., Zlotogora, J., Prockop, D. J., & Ala-Kokko, L. Splicing Mutations of 54-bp Exons in the *COL11A1* Gene Cause Marshall Syndrome, but Other Mutations Cause Overlapping Marshall/Stickler Phenotypes. *Am. J. Hum. Genet.* **65,** 974–983 (1999).

194. Brown, K. E., Lawrence, R. & Sonenshein, G. E. Concerted modulation of alpha 1(XI) and alpha 2(V) collagen mRNAs in bovine vascular smooth muscle cells. *J. Biol. Chem.* **266,** 23268–23273 (1991).

195. Van Helden, P. Not too much and not too little. *EMBO Rep.* **13,** 942 (2012).

196. Benazzo, F., Mosconi, M., Pio, A. & Combi, F. Hindfoot Tendinopathies in Athletes in *Tendinopathy Athletes* (Woo, S. L.-Y., Renstrm, P. A. F. H. & Arnoczky, S. P.) (Blackwell Publishing Ltd). 184–202 (2008)

197. Houshian, S., Tscherning, T. & Riegels-Nielsen, P. The epidemiology of Achilles tendon rupture in a Danish county. *Injury* **29,** 651–654 (1998).

198. Tumilty, S. Achilles tendon rupture: rising incidence in New Zealand follows international trends. *Phys. Ther. Rev.* **12,** 59–65 (2007).

199. Paavola, M., Kannus, P., Paakkala, T., Pasanen, M. & Järvinen, M. Long-term prognosis of patients with achilles tendinopathy. An observational 8-year follow-up study. *Am. J. Sports Med.* **28,** 634–642 (2000).

200. Cook, J. L. & Purdam, C. R. Is tendon pathology a continuum? A pathology model to explain the clinical presentation of load-induced tendinopathy. *Br. J. Sports Med.* **43,** 409–416 (2008).

201. Paavola, M., Kannus, P., Orava, S., Pasanen, M., & Järvinen, M. Surgical treatment for chronic Achilles tendinopathy: a prospective seven month follow up study. *Br. J. Sports Med.* **36,** 178–182 (2002).

202. Kader, D., Saxena, A., Movin, T., & Maffulli, N. Achilles tendinopathy: some aspects of basic science and clinical management. *Br. J. Sports Med.* **36,** 239–249 (2002).

203. Riley, G. Chronic tendon pathology: molecular basis and therapeutic implications. *Expert Rev. Mol. Med.* **7,** 1–25 (2005).

204. Fu, S.-C., Rolf, C., Cheuk, Y.-C., Lui, P. P. & Chan, K.-M. Deciphering the pathogenesis of tendinopathy: a three-stages process. *Sports Med. Arthrosc. Rehabil. Ther. Technol. Smartt* **2,** 30 (2010).

205. Aström, M. & Rausing, A. Chronic Achilles tendinopathy. A survey of surgical and histopathologic findings. *Clin. Orthop.* 151–164 (1995).

206. Clancy, W. Failed healing responses. *Sports-Induc. Inflamm. Clin. Basic Sci. Concepts Park Ridge Il Am. Orthop. Soc. Sports Med.* (1989).

207. Sham, P. C. & Curtis, D. Monte Carlo tests for associations between disease and alleles at highly polymorphic loci. *Ann. Hum. Genet.* **59,** 97–105 (1995).

208. Jin, H., Evangelou, E., Ioannidis, J. P. A. & Ralston, S. H. Polymorphisms in the 5' flank of COL1A1 gene and osteoporosis: meta-analysis of published studies. *Osteoporos. Int.* **22,** 911–921 (2010).

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

210. Mann, V. & Ralston, S. H. Meta-analysis of COL1A1 Sp1 polymorphism in relation to bone mineral density and osteoporotic fracture. *Bone* **32,** 711–717 (2003).

211. Tilkeridis, C., Bei, T., Garantziotis, S., & Stratakis, C. A. Association of a COL1A1 polymorphism with lumbar disc disease in young military recruits. *J. Med. Genet.* **42,** e44–e44 (2005).

212. Pluijm, S. M. F., Van Essen, H. W., Bravenboer, N., Uitterlinden, A. G., Smit, J. H., Pols, H. A. P., & Lips, P. Collagen type I 1 Sp1 polymorphism, osteoporosis, and intervertebral disc degeneration in older men and women. *Ann. Rheum. Dis.* **63,** 71–77 (2004).

213. Fessler, L. I., Shigaki, N. & Fessler, J. H. Isolation of a new procollagen V chain from chick embryo tendon. *J. Biol. Chem.* **260,** 13286–13293 (1985).

214. Fernandes, R. J., Weis, M., Scott, M. A., Seegmiller, R. E. & Eyre, D. R. Collagen XI chain misassembly in cartilage of the chondrodysplasia (cho) mouse. *Matrix Biol.* **26,** 597–603 (2007).

215. Dressler, M. R., Butler, D. L., Wenstrup, R., Awad, H. A., Smith, F., & Boivin, G. P. A potential mechanism for age-related declines in patellar tendon biomechanics. *J. Orthop. Res.* **20,** 1315–1322 (2002).

216. Ott, J. & Rabinowitz, D. The effect of marker heterozygosity on the power to detect linkage disequilibrium. *Genetics* **147,** 927–930 (1997).

217. Chapman, N. H. & Wijsman, E. M. Genome screens using linkage disequilibrium tests: optimal marker characteristics and feasibility. *Am. J. Hum. Genet.* **63,** 1872–1885 (1998).

218. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4,** 1073–1081 (2009).

219. Nowak, M. A., Boerlijst, M. C., Cooke, J. & Smith, J. M. Evolution of genetic redundancy. *Nature* **388,** 167–171 (1997).

220. Li, Y., Lacerda, D. A., Warman, M. L., Beier, D. R., Yoshioka, H., Ninomiya, Y., Oxford, J. T., Morris, N. P., Andrikopoulos, K., Ramirez, F., Wardell, B. B., Lifferth, G.D., Teuscher, C., Woodward, S.R., Taylor, B.A., Seegmiller, R.E., & Olsen, B. R. A fibrillar collagen gene, *COL11A1*, is essential for skeletal morphogenesis. *Cell* **80,** 423–430 (1995).

221. So, H.-C., Gui, A. H. S., Cherny, S. S. & Sham, P. C. Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. *Genet. Epidemiol.* **35,** 310–317 (2011).

222. Mittag, F., Büchel, F., Saad, M., Jahn, A., Schulte, C., Bochdanovits, Z., Simón-Sánchez, J., Nalls, M. A., Keller, M., Hernandez, D. G., Gibbs, J. R., Lesage, S., Brice, A., Heutink, P., Martinez, M., Wood, N. W., Hardy, J., Singleton, A. B., Zell, A., Gasser, T., & Sharma, M. Use of support vector machines for disease risk prediction in genome-wide association studies: Concerns and opportunities. *Hum. Mutat.* **33,** 1708–1718 (2012).

223. Panni, A. S., Tartarone, M. & Maffulli, N. Patellar tendinopathy in athletes. Outcome of nonoperative and operative management. *Am. J. Sports Med.* **28,** 392–397 (2000).

224. Gaida, J. E., Cook, J. L. & Bass, S. L. Adiposity and tendinopathy. *Disabil. Rehabil.* **30,** 1555–1562 (2008).

APPENDIX

Appendix A- Supplementary Tables

A1 Chapter 2- Supplementary Tables

Table A1.1 Genotype Effects on descriptive measures for *COL5A1* rs71746744 (-/AGGG) and rs1134170 (A/T) and *MIR608* rs4919510 (C/G) in a Caucasian South African population.

	***COL5A1* rs71746744**	***COL5A1* rs1134170**	***MIR608* rs4919510**
Age	0.578	0.479	0.252
Height	0.215	0.218	0.528
Weight	0.338	0.465	0.080
BMI	0.292	0.222	0.591
Gender	0.406	0.808	0.330

P-values are tabulated.

Table A1.2.1 Inferred haplotype Analysis of *COL5A1* rs71746744 and rs1134170 in all participants, male participants and female participants for asymptomatic controls (CON) and ACL rupture group (ACL).

Gene-gene interactions	Inferred allele combinations	CON	ACL	LR	p-value	Model
All Participants						
rs71746744 + rs1134170	**AGGG/T**	70.5 (152)	70.9 (152)	1.566	0.211	Dom
	AGGG/A	0.5(1)	0.5 (1)	0.001	0.921	Rec
	-/T	1.2 (2)	3.5 (7)	5.787	**0.016**	Dom
	-/A	27.9 (60)	25.1 (53)	11.483	**<0.001**	Rec
Male participants						
rs71746744 + rs1134170	**AGGG/T**	71.6 (95)	69.6 (111)	0.973	0.324	Rec
	AGGG/A	0.4 (0)	0.7 (1)	0.198	0.656	Dom
	-/T	0.4 (0)	4.7 (7)	13.234	**<0.001**	Dom
	-/A	27.6 (36)	25.0 (39)	10.306	**0.001**	Rec
Female Participants						
rs71746744 + rs1134170	**AGGG/T**	68.6 (56)	74.6 (40)	2.367	0.124	Dom
	AGGG/A	0.6 (0)	0 (0)	1.032	0.310	Dom/Mult
	-/T	2.4 (2)	0 (0)	4.105	**0.043**	Dom/Mult
	-/A	28.3 (23)	25.4 (13)	1.560	0.206	Rec

CON group and ACL rupture group are represented as a frequency (%), with the number of participants (n) in parenthesis. LR is the estimated risk. ND, not determined due to small sample size. Significant p-values ($p<0.05$) are indicated by bold type-face.

Best fit model was determined by AIC values. Models: Rec= reccesive; Dom= dominant; Mult= multiplicative, Gen= general.

Table A1.2.2 Inferred haplotype Analysis of *COL5A1* rs71746744 and rs1134170 in all participants, male participants and female participants for asymptomatic controls (CON) and non-contact ACL rupture subgroup (NON).

Gene-gene interactions	Inferred allele combinations	CON	NON	LR	p-value	Model
ALL Participants						
rs71746744 + rs1134170	**AGGG/T**	70.5 (152)	70.4 (84)	0.026	0.872	Dom
	AGGG/A	0.5 (1)	0.4 (1)	0.005	0.943	Rec/Dom/Mult
	-/T	1.2 (2)	4.2 (5)	6.360	**0.012**	Dom
	-/A	27.9 (60)	25.0 (29)	5.914	**0.150**	Rec
Male Participants						
rs71746744 + rs1134170	**AGGG/T**	71.6 (95)	69.2 (62)	0.296	0.586	Mult
	AGGG/A	0.4 (0)	0.6 (0)	0.074	0.785	Dom/Mult
	-/T	0.4 (0)	5.5 (5)	13.156	**<0.001**	Dom
	-/A	27.6 (36)	24.7 (22)	4.947	**0.026**	Rec
Female participants						
rs71746744 + rs1134170	**AGGG/T**	68.6 (56)	74.1 (21)	1.516	0.218	Dom
	AGGG/A	0.6 (0)	0 (0)	0.609	0.435	Dom
	-/T	2.4 (2)	0 (0)	2.422	0.120	Dom/Mult
	-/A	28.3 (23)	25.9 (7)	0.991	0.319	Rec

CON group and non-contact (NON) ACL rupture subgroup are represented as a frequency (%), with the number of participants (n) in parenthesis. LR is the estimated risk. ND, not determined due to small sample size. Significant p-values ($p<0.05$) are indicated by bold type-face.

Best fit model was determined by AIC values. Models: Rec= reccesive; Dom= dominant; Mult= multiplicative, Gen= general.

Table A1.3.1 Inferred pseudohaplotypes for the *COL5A1* rs71746744 and rs1134170 and *miR-608* rs4919510 polymorphisms in the control (CON), and anterior cruciate ligament rupture (ACL) groups in all participants, male participants and female participants.

Gene-gene interactions	Inferred allele combinations	CON	ACL	LR	p-value	Model
All Participants						
rs71746744 + rs1134170 + rs4919510	**AGGG/T/C**	55.9 (121)	57.9 (124)	0.133	0.715	Rec
	AGGG/T/G	14.6 (31)	13.0 (27)	0.171	0.679	Rec
	AGGG/A/C	0.3 (0)	0.4 (0)	0.004	0.948	Rec
	AGGG/A/G	0.02 (0)	0.1 (0)	0.001	0.973	Rec/Dom/Mult
	-/T/C	0.8 (1)	2.8 (6)	5.114	**0.024**	Dom
	-/T/G	0.3 (0)	0.7 (1)	1.862	0.172	Dom
	-/A/C	24.0 (52)	18.4 (39)	8.315	**0.004**	Rec
	-/A/G	3.8 (8)	6.6 (14)	1.255	0.263	Dom
Male Participants						
rs71746744 + rs1134170 + rs4919510	**AGGG/T/C**	57.1 (76)	56.0 (89)	2.604	0.106	Dom
	AGGG/T/G	14.5 (19)	13.6 (21)	1.013	0.314	Rec
	AGGG/A/C	0.4 (0)	0.6 (0)	0.067	0.796	Dom/Mult
	AGGG/A/G	0 (0)	0.1 (0)	<0.001	0.989	Rec
	-/T/C	0.4 (0)	3.8 (6)	10.396	**0.001**	Dom
	-/T/G	0 (0)	0.9 (1)	0.008	0.928	Rec
	-/A/C	24.9 (33)	17.9 (28)	6.984	**0.008**	Rec
	-/A/G	2.7 (3)	7.1 (11)	3.618	0.057	Dom
Female Participants						
rs71746744 + rs1134170 + rs4919510	**AGGG/T/C**	54.1 (44)	63.4 (34)	4.617	**0.032**	Dom
	AGGG/T/G	14.5 (12)	11.1 (6)	0.431	0.511	Mult
	AGGG/A/C	0 (0)	0 (0)	IF	IF	IF
	AGGG/A/G	0.6 (0)	0 (0)	1.060	0.303	Mult
	-/T/C	1.6 (1)	0 (0)	3.282	0.070	Dom
	-/T/G	0.8 (0)	0 (0)	1.543	0.214	Dom
	-/A/C	22.6 (18)	19.9 (10)	1.501	0.221	Rec
	-/A/G	5.8 (4)	5.8 (3)	0.360	0.549	Rec

CON group and ACL rupture group are represented as a frequency (%), with the number of participants (n) in parenthesis. LR is the estimated risk. ND, not determined due to small sample size. Significant p-values ($p<0.05$) are indicated by bold type-face.

Best fit model was determined by AIC values. Models: rec= Reccesive; Dom= dominant; Mult= multiplicative, Gen= general.

IF- could not be computed due to an inversion error.

TableA1.3.2 Inferred pseudohaplotypes for the *COL5A1* rs71746744 and rs1134170 and *miR-608* rs4919510 polymorphisms in all participants, male participants and female participants for the asymptomatic control (CON) group, and non-contact anterior cruciate ligament rupture (NON) subgroup.

Gene-gene interactions	Inferred allele combinations	CON	NON	LR	p-value	Model
All Participants						
rs71746744 + rs1134170 + rs4919510	**AGGG/T/C**	55.9 (121)	56.0 (67)	0.805	0.370	Dom
	AGGG/T/G	14.6 (31)	14.4 (17)	0.869	0.351	Rec
	AGGG/A/C	0.3 (0)	0.0 (0)	0.826	0.363	Dom/Mult
	AGGG/A/G	0.0 (0)	0.4 (0)	0.197	0.657	Dom
	-/T/C	0.8 (1)	3.00 (3)	4.904	**0.027**	Dom
	-/T/G	0.3 (0)	1.2 (1)	2.510	0.113	Dom
	-/A/C	24.0 (52)	20.9 (25)	2.601	0.107	Rec
	-/A/G	3.8 (8)	4.1 (4)	0.378	0.538	Rec
Male Participants						
rs71746744 + rs1134170 + rs4919510	**AGGG/T/C**	57.1 (76)	53.5 (48)	4.234	0.040	Dom
	AGGG/T/G	14.5 (19)	15.7 (14)	1.748	0.186	Rec
	AGGG/A/C	0.4 (0)	0 .0(0)	0.730	0.393	Dom
	AGGG/A/G	0.0 (0)	0.6 (0)	<0.001	0.980	Rec
	-/T/C	0.4 (0)	4.0 (3)	9.140	**0.002**	Dom
	-/T/G	0 (0)	1.5 (1)	IF	IF	IF
	-/A/C	24.9 (33)	20.0 (18)	2.607	0.106	Rec
	-/A/G	2.7 (3)	4.7 (4)	1.014	0.314	Dom
Female Participants						
rs71746744 + rs1134170 + rs4919510	**AGGG/T/C**	54.1 (44)	62.8 (18)	3.222	0.073	Dom
	AGGG/T/G	14.5 (12)	11.3 (3)	1.093	0.296	Rec
	AGGG/A/C	0.0 (0)	0.0 (0)	IF	IF	IF
	AGGG/A/G	0.6 (0)	0.0 (0)	0.632	0.427	Dom/Mult
	-/T/C	1.6 (1)	0.0 (0)	1.624	0.203	Dom/Mult
	-/T/G	0.8 (0)	0.0 (0)	0.947	0.330	Dom/Mult
	-/A/C	22.6 (18)	24.8 (7)	0.230	0.631	Rec
	-/A/G	5.8 (4)	1.1 (0)	1.882	0.170	Dom/Mult

CON group and ACL rupture group are represented as a frequency (%), with the number of participants (n) in parenthesis. LR is the estimated risk. ND, not determined due to small sample size. Significant p-values ($p<0.05$) are indicated by bold type-face.

Best fit model was determined by AIC values. Models: Rec= reccesive; Dom= dominant; Mult= multiplicative, Gen= general.

IF- could not be computed due to an inversion error.

A2 Chapter 3- Supplementary Tables

Table A2.1 Genotype Effects on descriptive measures for *COL11A1* rs3753841 and rs1676486 and *COL11A2* rs1799907 in a Caucasian South African population.

	***COL11A1* rs3753841**	***COL11A1* rs1676486**	***COL11A2* rs1799907**
Age	0.904	0.782	0.018
Height	0.403	0.450	0.290
Weight	0.078	0.318	0.611
BMI	0.961	0.548	0.053
Gender	0.868	0.575	0.808

P-values are tabulated.

Table A2.2.1 Inferred haplotypes for the *COL11A1* rs3753841 and rs1676486 polymorphisms in the control (CON), and anterior cruciate ligament rupture (ACL) groups for all participants, male participants and female participants.

Gene-gene interactions	Inferred allele combinations	CON	ACL	LR	p-value	Model
All participants						
rs3753841 + Rs1676486	**TC**	59.1 (128)	59.4 (127)	1.235	0.266	Dom
	TT	0.0 (0)	0.9 (1)	0.038	0.846	Rec
	CC	19.6 (42)	22.3 (47)	1.392	0.238	Rec
	CT	21.3 (46)	17.4 (37)	4.044	**0.044**	Dom
Male Participants						
rs3753841 + Rs1676486	**TC**	57.8 (77)	59.2 (94)	0.745	0.388	Rec
	TT	0.0 (0)	0.8 (1)	1.495	0.221	Mult
	CC	18.8 (25)	22.0 (35)	2.339	0.126	Rec
	CT	23.3 (31)	18.00 (28)	3.589	0.058	Dom
Female Participants						
rs3753841 + Rs1676486	**TC**	61.1 (50)	59.8 (32)	1.084	0.298	Dom
	TT	0.0 (0)	1.1 (0)	ND	ND	ND
	CC	20.8 (17)	23.0 (12)	0.391	0.532	Dom
	CT	18.1 (14)	16.1 (8)	1.084	0.298	Rec

CON group and ACL rupture group are represented as a frequency (%), with the number of participants (n) in parenthesis. LR is the estimated risk. ND, not determined due to small sample size. Significant p-values ($p<0.05$) are indicated by bold type-face.

Best fit model was determined by AIC values. Models: Rec= reccesive; Dom= dominant; Mult= multiplicative, Gen= general.

IF- could not be computed due to an inversion error.

Table A2.2.2 Inferred haplotypes for the *COL11A1* rs3753841 and rs1676486 polymorphisms in the control (CON) group, and non-contact (NON) anterior cruciate ligament rupture subgroup for all participants, male participants and female participants.

Gene-gene interactions	Inferred allele combinations	CON	NON	LR	p-value	Model
All participants						
rs3753841 + Rs1676486	**TC**	59.1 (126)	60.2 (72)	0.335	0.563	Rec
	TT	0.0 (0)	1.5 (1)	0.007	0.931	Rec
	CC	19.5 (42)	20.7 (24)	0.062	0.803	Dom/mult
	CT	21.4 (45)	17.6 (21)	2.274	0.132	Dom
Male Participants						
rs3753841 + Rs1676486	**TC**	57.8 (77)	60.2 (54)	0.453	0.501	Rec
	TT	0.0 (0)	1.4 (1)	0.006	0.940	Rec
	CC	18.8 (25)	20.0 (18)	0.122	0.727	Rec
	CT	23.3 (31)	18.4 (16)	1.408	0.235	Mult
Female Participants						
rs3753841 + Rs1676486	**TC**	61.1 (50)	60.3 (17)	0.099	0.753	Dom
	TT	0.0 (0)	1.9 (0)	ND	ND	ND
	CC	20.8 (17)	22.9 (6)	0.180	0.672	Dom
	CT	18.1 (14)	14.9 (4)	8.297	0.015	Gen

CON group and NON ACL rupture subgroup are represented as a frequency (%), with the number of participants (n) in parenthesis. LR is the estimated risk. ND, not determined due to small sample size. Significant p-values ($p<0.05$) are indicated by bold type-face.

Best fit model was determined by AIC values. Models: Rec= reccesive; Dom= dominant; Mult= multiplicative, Gen= general.

Table A2.3.1 Inferred pseudohaplotypes for the *COL11A1* rs3753841 and rs1676486 and *COL11A2* rs1799907 polymorphisms in the control (CON), and anterior cruciate ligament rupture (ACL) groups for all participants, male participants and female participants.

Gene-gene interactions	Inferred allele combinations	CON	ACL	LR	p-value	Model
All participants						
rs3753841 + rs1676486 + rs1799907	**TCA**	37.4 (81)	39.7 (85)	0.084	0.772	Mult
	TCT	21.7 (47)	19.7 (42)	0.597	0.440	Rec
	TTA	0.0 (0)	0.9 (1)	0.010	0.921	Rec
	TTT	0.0 (0)	0.0 (0) (0)	ND	ND	ND
	CCA	13.9 (30)	14.9 (31)	0.280	0.597	Mult
	CCT	5.7 (12)	7.4 (15)	0.949	0.330	Dom
	CTA	14.7 (31)	9.5 (20)	6.697	**0.010**	Rec
	CTT	6.6 (14)	8.0 (17)	0.035	0.852	Rec/Dom
Male Participants						
rs3753841 + rs1676486 + rs1799907	**TCA**	39.0 (52)	39.5 (63)	0.063	0.802	Rec
	TCT	18.8 (25)	19.7 (31)	0.863	0.353	Dom
	TTA	0.0 (0)	0.8 (1)	ND	ND	ND
	TTT	0.0 (0)	0.0 (0)	ND	ND	ND
	CCA	13.0 (17)	14.00 (22)	0.173	0.678	Dom/Mult
	CCT	5.8 (7)	8.1 (12)	1.153	0.283	Mult
	CTA	15.9 (21)	10.4 (16)	5.907	**0.015**	Rec
	CTT	7.4 (9)	7.5 (12)	0.121	0.728	Dom
Female Participants						
rs3753841 + rs1676486 + rs1799907	**TCA**	35.1 (29)	39.3 (21)	0.351	0.554	Rec
	TCT	26.0 (21)	20.6 (11)	0.707	0.400	Mult
	TTA	0.0 (0)	0.0 (0)	ND	ND	ND
	TTT	0.0 (0)	0.0 (0)	ND	ND	ND
	CCA	15.2 (12)	18.6 (10)	0.304	0.582	Mult
	CCT	5.5 (4)	4.3 (2)	0.342	0.559	Rec
	CTA	12.4 (10)	6.5 (3)	1.141	0.285	Mult
	CTT	5.6 (4)	9.7 (5)	0.718	0.397	Dom

CON group and ACL rupture group are represented as a frequency (%), with the number of participants (n) in parenthesis. LR is the estimated risk. ND, not determined due to small sample size. Significant p-values ($p<0.05$) are indicated by bold type-face.

Best fit model was determined by AIC values. Models: Rec= reccesive; Dom= dominant; Mult= multiplicative, Gen= general.

Table A2.3.2 Inferred pseudohaplotypes for the *COL11A1* rs3753841 and rs1676486 and *COL11A2* rs1799907 polymorphisms in the control (CON) group, and non-contact (NON) anterior cruciate ligament rupture subgroup for all participants, male participants and female participants.

Gene-gene interactions	Inferred allele combinations	CON	NON	LR	p-value	Model
All participants						
rs3753841 + rs1676486 + rs1799907	**TCA**	37.4 (81)	40.3 (48)	0.383	0.536	Rec
	TCT	21.7 (47)	19.8 (23)	0.381	0.537	Rec
	TTA	0.0 (0)	1.5 (1)	ND	ND	ND
	TTT	0.0 (0)	0.0 (0)	ND	ND	ND
	CCA	13.9 (30)	12.9 (15)	0.874	0.350	Rec
	CCT	5.7 (12)	7.8 (9)	1.077	0.299	Rec
	CTA	14.7 (31)	10.3 (12)	4.360	**0.037**	Rec
	CTT	6.6 (14)	7.4 (8)	1.217	0.270	Rec
Male Participants						
rs3753841 + rs1676486 + rs1799907	**TCA**	39.0 (52)	38.9 (35)	0.030	0.863	Dom
	TCT	18.8 (25)	21.3 (19)	1.374	0.241	Dom
	TTA	0.0 (0)	1.4 (1)	ND	ND	ND
	TTT	0.0 (0)	0.0 (0)	ND	ND	ND
	CCA	13.0 (17)	11.9 (10)	2.937	0.087	Rec
	CCT	5.8 (7)	8.2 (7)	1.086	0.297	Dom
	CTA	15.9 (21)	11.6 (10)	4.050	**0.044**	Rec
	CTT	7.4 (9)	6.8 (6)	0.986	0.321	rec
Female Participants						
rs3753841 + rs1676486 + rs1799907	**TCA**	35.1 (29)	43.7 (12)	1.162	0.281	Rec
	TCT	26.0 (21)	16.5 (4)	1.335	0.248	Mult
	TTA	0.0 (0)	1.9 (0)	ND	ND	ND
	TTT	0.0 (0)	0.0 (0)	ND	ND	ND
	CCA	15.2 (12)	15.6 (4)	0.131	0.717	Rec
	CCT	5.5 (4)	7.0 (2)	0.201	0.654	Rec
	CTA	12.4 (10)	7.7 (2)	0.697	0.404	Rec
	CTT	5.6 (4)	7.6 (2)	0.255	0.614	Rec

CON group and NON ACL rupture subgroup are represented as a frequency (%), with the number of participants (n) in parenthesis. LR is the estimated risk. ND, not determined due to small sample size. Significant p-values ($p<0.05$) are indicated by bold type-face.

Best fit model was determined by AIC values. Models: Rec= reccesive; Dom= dominant; Mult= multiplicative, Gen= general.

Table A2.4.1 Inferred pseudohaplotypes for the *COL11A1* rs3753841 and rs1676486 and *COL11A2* rs1799907 polymorphisms in the control (CON) group, and anterior cruciate ligament rupture (ACL) group for all participants, male participants and female participants.

Gene-gene interactions	Inferred allele combinations	CON	NON	LR	p-value	Model
All Participants						
	TCA(AGGG)	24.7 (53)	30.2 (65)	0.839	0.360	Mult
	TCA(-)	12.8 (27)	9.5 (20)	0.708	0.400	Mult
COL11A1	**TCT(AGGG)**	16.2 (35)	10.7 (22)	3.410	0.065	Rec
rs3753841	**TCT(-)**	5.4 (11)	9.0 (19)	0.868	0.351	Dom
+	**TTA(AGGG)**	0.0 (0)	0.9 (1)	0.008	0.928	Rec
COL11A1	**TTA(-)**	0 (0)	0 (0)	ND	ND	ND
rs1676486	**TTT(AGGG)**	0 (0)	0 (0)	ND	ND	ND
+	**TTT(-)**	0 (0)	0 (0)	ND	ND	MD
COL11A2	**CCA(AGGG)**	11.1 (24)	10.3 (22)	0.344	0.557	Mult
rs1799907	**CCA(-)**	2.4 (5)	4.7 (10)	0.452	0.501	Rec
+	**CCT(AGGG)**	1.9 (4)	5.9 (12)	2.057	0.152	Mult
COL5A1	**CCT(-)**	4.1 (8)	1.3 (2)	0.321	0.571	Rec
rs71746744	**CTA(AGGG)**	12.9 (28)	9.2 (19)	5.302	0.021	Rec
	CTA(-)	2.0 (4)	0.0 (0)	3.350	0.067	Dom/mult
	CTT(AGGG)	4.0 (8)	4.2 (9)	0.852	0.356	Rec
	CTT(-)	2.4 (5)	4.0 (8)	0.503	0.478	Dom
Male Participants						
	TCA(AGGG)	27.9 (37)	31.9 (49)	0.066	0.798	Mult
	TCA(-)	11.2 (14)	8.9 (14)	0.450	0.502	Rec
COL11A1	**TCT(AGGG)**	16.2 (21)	9.0 (14)	5.602	0.018	Rec
rs3753841	**TCT(-)**	2.6 (3)	10.1 (16)	4.650	0.031	Mult
+	**TTA(AGGG)**	0.0 (0)	0.6 (0)	0.004	0.947	Rec
COL11A1	**TTA(-)**	0.0 (0)	0.0 (0)	ND	ND	ND
rs1676486	**TTT(AGGG)**	0.0 (0)	0.0 (0)	ND	ND	ND
+	**TTT(-)**	0.0 (0)	0.0 (0)	ND	ND	ND
COL11A2	**CCA(AGGG)**	9.6 (12)	8.2 (13)	0.318	0.573	Dom
rs1799907	**CCA(-)**	3.2 (4)	5.8 (9)	0.520	0.471	Rec
+	**CCT(AGGG)**	1.1 (1)	6.6 (10)	2.754	0.097	Mult
COL5A1	**CCT(-)**	4.9 (6)	1.4 (2)	0.300	0.584	Rec
rs71746744	**CTA(AGGG)**	13.6 (18)	9.8 (15)	4.795	0.029	Rec
	CTA(-)	2.4 (3)	0.2 (0)	3.516	0.061	Dom/Mult
	CTT(AGGG)	3.5 (4)	4.7 (7)	0.576	0.448	Rec
	CTT(-)	3.8 (5)	3.2 (5)	0.365	0.546	Rec
Female Participants						
	TCA(AGGG)	20.2 (16)	28.3 (15)	1.382	0.240	Mult
	TCA(-)	15.1 (12)	11.5 (6)	0.760	0.383	Dom
COL11A1	**TCT(AGGG)**	17.4 (14)	13.3 (7)	0.112	0.738	Mult
rs3753841	**TCT(-)**	8.4 (6)	6.7 (3)	1.239	0.266	Mult
+	**TTA(AGGG)**	0.0 (0)	1.1 (0)	ND	ND	ND
COL11A1	**TTA(-)**	0.0 (0)	0.0 (0)	ND	ND	ND
rs1676486	**TTT(AGGG)**	0.0 (0)	0.0 (0)	ND	ND	ND
+	**TTT(-)**	0.0 (0)	0.0 (0)	ND	ND	ND
COL11A2	**CCA(AGGG)**	13.1 (10)	16.7 (9)	0.796	0.372	Rec
rs1799907	**CCA(-)**	1.7 (1)	0.0 (0)	0.171	0.679	Dom/Mult
+	**CCT(AGGG)**	3.1 (2)	4.1 (2)	0.161	0.688	Rec
COL5A1	**CCT(-)**	3.0 (2)	2.1 (1)	1.562	0.211	Dom/Mult
rs71746744	**CTA(AGGG)**	11.7 (9)	4.0 (2)	1.358	0.244	Mult
	CTA(-)	1.1 (0)	3.8 (2)	0.315	0.575	Dom/Mult
	CTT(AGGG)	3.8 (3)	7.0 (3)	0.274	0.600	Rec
	CTT(-)	1.5 (1)	1.3 (0)	1.433	0.231	Dom

CON group and ACL rupture group are represented as a frequency (%), with the number of participants (n) in parenthesis. LR is the estimated risk. ND, not determined due to small sample size. Significant p-values ($p<0.05$) are indicated by bold type-face.

Best fit model was determined by AIC values. Models: Rec= reccesive; Dom= dominant; Mult= multiplicative, Gen= general.

Table A2.4.2 Inferred pseudohaplotypes for the *COL11A1* rs3753841 and rs1676486 and *COL11A2* rs1799907 polymorphisms in the control (CON) group, and non-contact (NON) anterior cruciate ligament rupture subgroup for all participants, male participants and female participants.

Gene-gene interactions	Inferred allele combinations	CON	NON	LR	p-value	Model
All Participants						
	TCA(AGGG)	24.7 (53)	32.7 (39)	3.820	0.051	Rec
	TCA(-)	12.8 (27)	8.1 (9)	1.206	0.272	Dom
COL11A1	**TCT(AGGG)**	16.2 (35)	8.2 (9)	4.710	**0.030**	Rec
rs3753841	**TCT(-)**	5.4 (11)	11.2 (13)	1.931	0.165	Mult
+	**TTA(AGGG)**	0.0 (0)	1.4 (1)	ND	ND	ND
COL11A1	**TTA(-)**	0 (0	0.0 (0)	ND	ND	ND
rs1676486	**TTT(AGGG)**	0 (0	0.0 (0)	ND	ND	ND
+	**TTT(-)**	0 (0	0.1 (0)	ND	ND	ND
COL11A2	**CCA(AGGG)**	11.1 (24)	7.9 (9)	0.130	0.718	Rec
rs1799907	**CCA(-)**	2.4 (5)	4.5 (5)	0.178	0.673	Rec
+	**CCT(AGGG)**	1.9 (4)	6.8 (8)	1.767	0.184	Dom
COL5A1	**CCT(-)**	4.1 (8)	1.6 (1)	0.227	0.634	Rec
rs71746744	**CTA(AGGG)**	12.9 (28)	10.4 (12)	3.359	0.067	Rec
	CTA(-)	2.0 (4)	0.0 (0)	1.310	0.252	Dom
	CTT(AGGG)	4.0 (8)	3.4 (4)	0.419	0.517	Rec
	CTT(-)	2.4 (5)	3.8 (4)	0.345	0.557	Dom
Male Participants						
	TCA(AGGG)	27.9 (37)	34.4 (31)	2.215	0.137	Rec
	TCA(-)	11.2 (14)	5.6 (5)	0.731	0.392	Dom
COL11A1	**TCT(AGGG)**	16.2 (21)	5.6 (5)	3.392	0.066	Rec
rs3753841	**TCT(-)**	2.6 (3)	14.6 (13)	7.099	**0.008**	Mult
+	**TTA(AGGG)**	0.0 (0)	0.8 (0)	ND	ND	ND
COL11A1	**TTA(-)**	0.0 (0)	0.0 (0)	ND	ND	ND
rs1676486	**TTT(AGGG)**	0.0 (0)	0.0 (0)	ND	ND	ND
+	**TTT(-)**	0.0 (0)	0.0 (0)	ND	ND	ND
COL11A2	**CCA(AGGG)**	9.6 (12)	6.5 (5)	1.633	0.201	Rec
rs1799907	**CCA(-)**	3.2 (4)	4.9 (4)	0.214	0.643	Rec
+	**CCT(AGGG)**	1.1 (1)	7.3 (6)	2.300	0.129	Dom
COL5A1	**CCT(-)**	4.9 (6)	1.3 (1)	0.223	0.637	Rec
rs71746744	**CTA(AGGG)**	13.6 (18)	7.7 (7)	3.102	0.078	Rec
	CTA(-)	2.4 (3)	3.8 (3)	1.519	0.218	Dom/Mult
	CTT(AGGG)	3.5 (4)	6.9 (6)	0.268	0.605	Rec
	CTT(-)	3.8 (5)	0.0 (0)	0.200	0.655	Rec
Female Participants						
	TCA(AGGG)	20.2 (16)	29.0 (8)	1.586	0.208	Mult
	TCA(-)	15.1 (12)	15.0 (4)	0.309	0.578	Dom
COL11A1	**TCT(AGGG)**	17.4 (14)	9.8 (2)	1.339	0.247	Rec
rs3753841	**TCT(-)**	8.4 (6)	6.4 (1)	0.670	0.413	Mult
+	**TTA(AGGG)**	0.0 (0)	1.9 (0)	ND	ND	ND
COL11A1	**TTA(-)**	0.0 (0)	0.0 (0)	ND	ND	ND
rs1676486	**TTT(AGGG)**	0.0 (0)	0.0 (0)	ND	ND	ND
+	**TTT(-)**	0.0 (0)	0.0 (0)	ND	ND	ND
COL11A2	**CCA(AGGG)**	13.1 (10)	14.9 (4)	0.373	0.541	Rec
rs1799907	**CCA(-)**	1.7 (1)	0.0 (0)	0.651	0.420	Dom/Mult
+	**CCT(AGGG)**	3.1 (2)	7.7 (2)	0.280	0.596	Dom
COL5A1	**CCT(-)**	3.0 (2)	0.0 (0)	1.385	0.239	Dom/Mult
rs71746744	**CTA(AGGG)**	11.7 (9)	8.2 (2)	0.653	0.419	Mult
	CTA(-)	1.1 (0)	0.0 (0)	ND	ND	ND
	CTT(AGGG)	3.8 (3)	2.7 (0)	0.107	0.743	Rec
	CTT(-)	1.5 (1)	4.4 (1)	0.741	0.389	Dom

CON group and ACL rupture group are represented as a frequency (%), with the number of participants (n) in parenthesis. LR is the estimated risk. ND, not determined due to small sample size. Significant p-values ($p<0.05$) are indicated by bold type-face.

Best fit model was determined by AIC values. Models: Rec= reccesive; Dom= dominant; Mult= multiplicative, Gen= general.

A3 Chapter 4- Supplementary Tables

Table A3.1 Genotype Effects on descriptive measures for *COL11A1* rs3753841 and rs1676486 and *COL11A2* rs1799907 in a Caucasian South African and Caucasian Australian population.

	COL11A1 rs3753841	*COL11A1* rs1676486	*COL11A2* rs1799907	*COL5A1* rs71746744
Age	0.062	0.866	0.586	0.861
Height	0.817	0.636	0.948	0.625
Weight	0.808	0.529	0.946	0.419
BMI	0.885	0.143	0.997	0.661
Gender	0.936	0.771	0253	0.342

P-values are tabulated.

Table A3.2 Inferred allele combinations for the *COL11A1* rs3753841, *COL11A1* rs1676486 and *COL11A2* rs1799907 polymorphisms in the control (CON), and Achilles tendinopathy (TEN) groups.

Gene-gene interactions	Inferred allele combinations	CON	TEN	LR	p-value	Model
rs3753841 vs rs1676486	**TC**	57.8 (172)	63.3 (106)	4.751	**0.029**	Rec
	TT	1.2 (3)	<0.5 (0)	1.195	0.274	Mult
	CC	20.5 (61)	19.5 (32)	2.158	0.142	Rec
	CT	20.5 (60)	16.8 (28)	1.997	0.158	Dom
rs3753841 vs rs1799907	**TA**	41.2 (132)	37.17 (67)	0.539	0.463	Dom
	TT	16.9 (84)	26.89 (48)	7.830	**0.005**	Mult
	CA	27.4 (88)	26.21 (88)	3.212	0.073	Dom
	CT	14.5 (46)	9.72 (17)	0.983	0.321	Mult
rs1676486 vs rs1799907	**CA**	55.2 (182)	49.5 (90)	4.300	**0.038**	Rec
	CT	23.1 (76)	32.5 (59)	6.576	**0.010**	Mult
	TA	13.5 (44)	14.0 (25)	0.413	0.520	Rec
	TT	8.2 (27)	4.0 (7)	3.185	**0.027**	Dom
rs3753841 vs rs1676486 vs rs1799907	**TCA**	40.6 (137)	37.1 (67)	0.744	0.388	Dom
	TCT	17.2 (58)	26.2 (47)	7.790	**0.005**	Mult
	TTA	1.2 (4)	0.4 (0)	1.176	0.278	Mult
	TTT	0.0 (0)	0.0 (0)	N/A	N/A	N/A
	CCA	14.4 (48)	13.1 (24)	1.913	0.167	Rec
	CCT	6.1 (20)	6.3 (11)	1.284	0.257	Rec
	CTA	12.4 (42)	13.0 (23)	0.450	0.502	Mult
	CTT	8.0 (27)	3.9 (7)	3.694	0.055	Dom

CON and TEN are represented as a frequency, with the number of participants (n) in parenthesis. LR is the estimated risk. ND, not determined due to small sample size. ***** represents a significant p-value ($p<0.05$).

Best fit model was determined by AIC values. Models: Rec= reccesive; Dom= dominant; Mult= multiplicative, Gen= general.

Table A3.3 Inferred allele combinations for the *COL11A1* rs3753841, *COL11A1* rs1676486 and *COL11A2* rs1799907 and *COL5A1* rs71746744 polymorphisms in the control (CON), and Achilles tendinopathy (TEN) groups.

Gene-gene interactions	Inferred allele combinations	CON	TEN	LR	p-value	Model
COL11A1 rs3753841 + ***COL5A1*** rs71746744	**T/AGGG**	37.2 (104)	49.2 (80)	8.322	**0.004**	Mult
	T/-	20.8 (58)	14.7 (24)	5.306	**0.021**	Rec
	C/AGGG	29.4 (82)	29.7 (48)	7.041	**0.030**	Gen
	C/-	12.6 (35)	6.4 (10)	17.007	**<0.001**	Gen
COL11A1 rs1676486 + ***COL5A1*** rs71746744	**C/AGGG**	49.8 (142)	63.5 (106)	11.128	**<0.001**	Dom
	C/-	28.4 (81)	18.6 (31)	5.716	**0.016**	Rec
	T/AGGG	16.7 (47)	15.6 (26)	0.514	0.474	Mult
	T/-	5.1 (14)	2.3 (3)	4.145	**0.042**	Mult
COL11A2 rs1799907 + ***COL5A1*** rs71746744	**A/AGGG**	45.3 (139)	49.5 (86)	0.448	0.503	Dom
	A/-	23.5 (72)	13.7 (23)	10.055	**0.002**	Mult
	T/AGGG	21.3 (65)	29.6 (51)	5.788	**0.016**	Dom
	T/-	9.8 (30)	7.2 (12)	6.927	0.405	Rec
COL11A1 rs3753841 + ***COL11A1*** rs1676486 + ***COL11A2*** rs1799907 + ***COL5A1*** rs71746744	**TCA(AGGG)**	29.2 (98)	25.6 (47)	0.731	0.393	Dom
	TCA(-)	11.6 (39)	11.9 (21)	3.604	0.058	Mult
	TCT(AGGG)	9.1 (30)	25.2 (46)	13.386	**<0.001**	Mult
	TCT(-)	8.0 (27)	0 (0)	1.531	0.216	Rec
	TTA(AGGG)	0 (0)	0 (0)	ND	ND	ND
	TTA(-)	1.2 (3)	0 (0)	ND	ND	ND
	TTT(AGGG)	0 (0)	0 (0)	ND	ND	ND
	TTT(-)	0 (0)	0.5 (0)	ND	ND	ND
	CCA(AGGG)	8.8 (29)	11.5 (21)	0.443	0.506	Rec
	CCA(-)	5.3 (18)	1.2 (2)	4.148	**0.043**	Mult
	CCT(AGGG)	4.0 (13)	2.4 (4)	0.629	0.428	Dom
	CCT(-)	2.2 (7)	4.8 (8)	0.396	0.529	Rec
	CTA(AGGG)	7.3 (24)	10.7 (19)	0.085	0.770	Dom
	CTA(-)	5.3 (17)	2.2 (4)	4.336	**0.037**	Mult
	CTT(AGGG)	7.9 (26)	4.0 (7)	2.928	0.087	Dom
	CTT(-)	0 (0)	0 (0)	ND	ND	ND

CON and TEN are represented as a frequency, with the number of participants (n) in parenthesis. LR is the estimated risk. ND, not determined due to small sample size. Significant values are indicated by bold typeface.

Best fit model was determined by AIC values. Models: Rec= reccesive; Dom= dominant; Mult= multiplicative, Gen= general.

Appendix B- Forms

B1.1 ACL Rupture: Ethics Approval

UNIVERSITY OF CAPE TOWN

Health Sciences Faculty
Research Ethics Committee
Room E52-24 Groote Schuur Hospital Old Main Building
Observatory 7925
Telephone [021] 406 6338 • Facsimile [021] 406 6411
e-mail: [illegible]

26 April 2006

REC REF: 164/2006

Dr M Collins
Human Biology

Dear Dr Collins

PROJECT TITLE: THE COL5A1 AND TNC GENES AND THEIR ASSOCIATION WITH ANTERIOR CRUCIATE LIGAMENT INJURIES

Thank you for submitting your study to the Research Ethics Committee for review.

It is a pleasure to inform you that the Ethics Committee has **formally approved** the above-mentioned study on the 21 April 2006.

This serves to confirm that the University of Cape Town Research Ethics Committee complies to the Ethics Standards for Clinical Research with a new drug in patients, based on the Medical Research Council (MRC-SA), Food and Drug Administration (FDA-USA), International Convention on Harmonisation Good Clinical Practice (ICH GCP) and Declaration of Helsinki guidelines.

The Research Ethics Committee granting this approval is in compliance with the ICH Harmonised Tripartite Guidelines E6: Note for Guidance on Good Clinical Practice (CPMP/ICH/135/95) and FDA Code Federal Regulation Part 50, 56 and 312.

Please quote the REC. REF in all your correspondence.

Yours sincerely

DR. M. BLOCKMAN
CHAIRPERSON, HSF HUMAN ETHICS

lcmjedi

B1.2 ACL Rupture: Informed Consent

Department of Human Biology
UCT/MRC RESEARCH UNIT FOR EXERCISE SCIENCE & SPORTS MEDICINE
Faculty of Health Sciences, University of Cape Town
Private Bag, Rondebosch 7700, South Africa
Tel: + 27 21 650 4561
Fax: + 27 21 686 7530

GENETIC BASIS OF EXERCISE-INDUCED LIGAMENT INUJURY

INFORMED CONSENT

I, the undersigned, have been fully informed about the UCT/MRC Research Unit for Exercise Science and Sports Medicine within the Department of Human Biology of the University of Cape Town's study on the genetic basis of exercise induced chronic ligament pathology. I have agreed to donate five millilitres of venous blood or a Buccal mouthwash sample, which will be used for the extraction and analysis of genetic material (DNA). I have also agreed to complete personal particulars, sporting participation, medical history, stretching and warm up questionnaires and understand that all the information that is collected during the study will be treated with the strictest confidentiality and will only be used for scientific research purposes. I also understand that my name and personal particulars will be not released under any circumstances and that all data will be analysed anonymously.

I agree to participate in the study and I have been informed that I will be free to withdraw from the study at any time if I so wish. I understand that my DNA sample will be destroyed on completion of the study on the genetic basis of ligament pathology. I also understand that I will be free to request that my DNA sample be destroyed before the completion of the study.

I understand that the DNA will be genotyped (analysed) for variations (polymorphisms) within genes relating to the genetic basis of ligament injuries. I understand that whilst there is no direct benefit to myself, if a genetic predisposition for ligament injuries can be established, then future generations will be able to establish their risk for this condition. This may allow better prevention and treatment options in the future. I understand that I will receive the overall results of the study. I have read (or where appropriate, have had read to me) and understand the information about this study, and any questions I have asked have been answered to my satisfaction. I agree to participate in the study, realising that I have the right to request that my DNA sample be destroyed at any time. I agree that research data provided by me or with my permission during the project may be included in a thesis, presented at conferences and published in journals on the condition that either my name not any other identifying information is used.

FULL NAME OF SUBJECT: _____

SUBJECT'S SIGNATURE: _____

DATE: _____

INVESTIGATOR : _____

INVESTIGATOR'S SIGNATURE: _____

B1.3 ACL Rupture: Participant Information

Department of Human Biology
UCT/MRC RESEARCH UNIT FOR EXERCISE SCIENCE & SPORTS MEDICINE
Faculty of Health Sciences, University of Cape Town
Private Bag, Rondebosch 7700, South Africa
Tel: + 27 21 650 4561
Fax: + 27 21 686 7530

THE GENETIC BASIS OF EXERCISE-INDUCED LIGAMENT INJURY

Although there is a high incidence of ligament injuries as a result of participation in exercise and sporting activities, the cause(s) of these injuries are poorly understood. Some researchers have suggested that there is a genetic component to exercise-induced ligament injuries. In an attempt to determine whether there is a genetic basis for ligament pathology, we are interested in studying whether certain genes are associated with ligament injuries. This project is being done in collaboration with the UCT/MRC Research Unit for Exercise Science and Sports Medicine within the Department of Human Biology of the University of Cape Town.

You will be required to visit the Sports Science Institute of South Africa (SSISA) in Boundary Road, Newlands. During the visit, which should take 15 minutes, you will be asked to donate 5 ml (1 teaspoon) of a blood sample for DNA analysis. You will also be required to complete personal particulars, sporting details, medical history and stretching and warm up questionnaires.

All the information retrieved from this study will be treated with the strictest confidentiality and will be used only for scientific research purposes. Your name and personal particulars will not be released under any circumstances and all data will be analysed anonymously. Your DNA sample will be destroyed on completion of the study on the genetic basis of ligament injury. You are also free to request that your DNA sample be destroyed before the completion of the study.

If you are part of the ligament pathology group, we would appreciate it if you could help us by recruiting two other people of same (or similar) age whom you know and who has trained without suffering any ligament or tendon injuries for the control group.

We will keep you informed about the outcomes of this study and look forward to working together with you. If you have any questions about this study, please feel free to contact us at:-

Melanie Hay, BSc (Med) (Hons)
(021) 650-4569/ 082 781 0102
melanie.hay@uct.ac.za

Dr Alison September, PhD
(021) 650 4559
alison.september@uct.ac.za

Dr Mike Posthumus, PhD
(021) 650 4572
michael.posthumus@uct.ac.za

Prof. Malcolm Collins, PhD
(021) 650 4574
malcolm.collins@uct.ac.za

B1.4 ACL Rupture: Questionnaires

Department of Human Biology
UCT/MRC Research Unit for Exercise Science & Sports Medicine
Faculty of Health Sciences, University of Cape Town
Private Bag, Rondebosch 7700, South Africa
Tel: + 27 21 650 4561
Fax: + 27 21 686 7530

GENETIC BASIS OF LIGAMENT INJURY QUESTIONNAIRES

A. PERSONAL PARTICULARS			
Surname			
First Name			
Postal Address		Code	
E-mail address		Phone (day time)	
Date of birth	Y Y Y Y / M M / D D	Cell	
Height (cm)		Gender	Male ☐ Female ☐
Weight (kg)	Pre-Injury:	Current:	
Ethnic group (Only Required and Used for Research Purposes)	Black/African ☐ White ☐ Indian ☐ Mixed Ancestry (Coloured) ☐ Asian ☐ Other ☐		
Ancestry: Tribal or national background (eg Xhosa, Dutch, Zulu, German, Italian)	Father Unknown ☐		
	Mother Unknown ☐		
Country of Birth			
Dominant Hand	Left ☐ Right ☐ Ambi ☐	Dominant Leg	Left ☐ Right ☐ Ambi ☐
Smoker	Yes (Current) ☐	Yes (Ex smoker) ☐	No, never ☐
	If yes, Number of years _____	If stopped, when _____	
	If yes, number per day _____		

(If you participate or have participated in more than 6 sports, please complete additional Sporting Details Questionnaires, Part B)

B. SPORTING DETAILS **Please record your sporting activities in order of importance**			
Type of sport(s) you have participated in (please name)	**Main sport 1**	**Other sport 2**	**Other sport 3**
Current or past participation	Current ☐ Past ☐	Current ☐ Past ☐	Current ☐ Past ☐
Year started participation			
Number of years involved in the sport			
Position played prior to injury (if appropriate)			
Playing level prior to injury (if appropriate)			
Number of years played prior to the injury.			

Type of sport(s) you have participated in (please name)	**Other sport 4**	**Other sport 5**	**Other sport 6**
Current or past participation	Current ☐ Past ☐	Current ☐ Past ☐	Current ☐ Past ☐
Year started participation			
Number of years involved in the sport			
Position played prior to injury (if appropriate)			
Playing level prior to injury (if appropriate)			
Number of years played prior to the injury.			

C. ANTERIOR CRUTIATE LIGAMENT INJURY DETAILS	
Date of ACL injury?	
Which side was injured?	☐ Left ☐ Right ☐ Both
To what extent was your ligament ruptured?	☐ Complete ☐ Partial ☐ None ☐ Unknown
Investigation done to confirm the diagnosis	☐ MRI ☐ Surgery
How bad is your pain today? (mark line: e.g.I----------I----------I)	I----------I No pain Pain as bad as it can be
How was the ACL ruptured? **(please also explain exactly how the injury occurred)**	☐ Direct impact ☐ Twisting and bending with contact ☐ Twisting and bending without contact ☐ Other non-contact ☐ Other……
What was the initial treatment? (You may tick more than one block.)	☐ Ice application ☐ Compresion ☐ Immobilisation ☐ Medication ☐ Other……
What was the final treatment?	☐ Surgery ☐ Reabilitation ☐ Other……
What are your current symptoms? (You may tick more than one block.)	☐ Pain ☐ Swelling ☐ Instability ☐ Weakness ☐ Other……
What is your current sports participation?	☐ None ☐ Limited to non-weight bearing exercise ☐ Limited, not to same level as pre-injury ☐ Full participation

If you are able to recall, what were the weather and pitch conditions like at the time of injury?	☐ Wet and soft ground ☐ Dry, but soft ground ☐ Dry and firm ground ☐ Wet, but firm ground ☐ Other……
Associated injuries?	☐ Meniscal tear ☐ MCL tear ☐ Other ligament tear ☐ Bone bruising ☐ Other……

D. HISTORY OF OTHER LIGAMENT AND TENDON INJURIES IN THE PAST				
Have you ever injured a ligament in the past?	Yes ☐ No ☐			
If yes, please specify which ligaments? (You may tick more than one block, please select either L (left) or R (right))		L R		L R
	Knee (ACL)	☐ ☐	Wrist ligaments	☐ ☐
	Knee (MCL)	☐ ☐	Finger ligaments	☐ ☐
	Ankle lateral ligaments	☐ ☐	Knee (PCL)	☐ ☐
	Spinal ligaments	☐ ☐	Knee (LCL)	☐ ☐
	Shoulder ligaments	☐ ☐	Ankle medial ligaments	☐ ☐
	Elbow ligaments	☐ ☐	Other ligaments	☐ ☐
To your knowledge, have any other members of your family suffered from any ligament injury?	Yes ☐ No ☐	If Yes, please specify the family member ☐ Mother ☐ Father ☐ Sibling ☐ Son / daughter ☐ Other family member…… and condition: Please choose ligament injury from the list above ……		
Have you ever injured a tendon in the past?	Yes ☐ No ☐			

<table>
<tr><td rowspan="9">If yes, please specify which tendon?
(You may tick more than one block, please select either L (left) or R (right))</td><td rowspan="3">Foot and ankle:</td><td></td><td>L R</td></tr>
<tr><td>Achilles tendon</td><td>☐ ☐</td></tr>
<tr><td>Tibialis posterior</td><td>☐ ☐</td></tr>
<tr><td>Plantar fascia</td><td>☐ ☐</td></tr>
<tr><td>Knee:</td><td>Patellar tendon</td><td>☐ ☐</td></tr>
<tr><td>Elbow and wrist:</td><td>Wrist extensor tendons</td><td>☐ ☐</td></tr>
<tr><td rowspan="4">Shoulder:</td><td>Subscapularis</td><td>☐ ☐</td></tr>
<tr><td>Supraspinatus</td><td>☐ ☐</td></tr>
<tr><td>Infraspinatus</td><td>☐ ☐</td></tr>
<tr><td>Teres minor</td><td>☐ ☐</td></tr>
<tr><td colspan="3">Other:.....</td></tr>
<tr><td>To your knowledge, have any other members of your family suffered from any tendon pathology?</td><td>Yes ☐ No ☐</td><td colspan="2">If Yes, please specify the family member
☐ Mother
☐ Father
☐ Sibling
☐ Son / daughter
☐ Other family member:.....
Condition: Please choose tendon injury from the list above
.....</td></tr>
<tr><td>Have you ever suffered from any of the following joint capsule injuries?</td><td colspan="3">☐ Acute shoulder dislocation
☐ Chronic shoulder instability
☐ Chronic ankle instability
☐ Other: _____
_____</td></tr>
</table>

E. MEDICAL HISTORY

Do you currently suffer from any of these medical conditions:

☐ High Blood Pressure	☐ Angina/Heart Attack	☐ Asthma
☐ Emphysema	☐ Rheumatoid arthritis	☐ Osteoarthritis (wear & tear)
☐ Malignant disease (cancer)	☐ Elevated Blood Cholesterol	☐ Adrenal disorders
If Yes, what type? _____	☐ Diabetes mellitus	☐ Thyroid disorders
	☐ Renal disease	☐ Amyloidosis

Do you currently suffer from any other Connective Tissue & Rheumatological Diseases & Disorders?	Yes ☐ No ☐	If Yes, please select from the list below

List of some Connective Tissue and/or Rheumatic Diseases and Disorders

☐ Ankylosing Spondylitis	☐ Lipid Storage Diseases	☐ Pseudogout
☐ Aspartylglycosaminuria (AGU)	☐ Marfan Syndrome	☐ Reactive Arthritis
☐ Behcet's Syndrome	☐ Menkes Kinky Hair Syndrome	☐ Reiter's Syndrome
☐ Crohn's Disease	☐ Mucopolysaccharidoses	☐ Relapsing Polychrondritis
☐ Discoid Lupus Erythematosus	☐ Myopathies and Dystrophies	☐ Scleroderma
☐ Ehlers-Danlos syndrome (EDS)	☐ Ochronosis (Homocystinuria)	☐ Sjogren's Syndrome
☐ Eosinophilic Fascitis	☐ Osteogenesis imperfecta (OI)	☐ Systemic Lupus Erythematosus (SLE)
☐ Giant Cell (Temporal) Arthritis	☐ Polyarteritis Nodosa	☐ Systemic Sclerosis
☐ Gout	☐ Polymyalgia Rheumatica	☐ Wegener's Granulomatosis
☐ Hypersentive Vasulatis	☐ Polymyositis & Dermatomyositis	☐ Other_____

What surgical operations have you had? (please list and give dates)	Operation	Date

If female:

At what age did you start menstruating? (years)	
Are you currently using any type of contraception?	☐ Yes ☐ No
If Yes, what type of contraception are you using?	☐ Pill ☐ Injection ☐ IUD

Are you currently?	☐ Pre-menopausal (±12 cycles per year at intervals of 23–33 days & bleeding lasts 3-7 days) ☐ Menopausal (cycles are irregular and less frequent) ☐ Post-menopausal (no longer menstruating)	
Family History		
Do any other members of your family suffer from elevated blood cholesterol?	Yes ☐ No ☐	If Yes, which relative? ☐ Mother ☐ Father ☐ Sibling ☐ Son / daughter ☐ Other relative:……
Is there any history of arthritis in your family?	Yes ☐ No ☐	If Yes, which relative? ☐ Mother ☐ Father ☐ Sibling ☐ Son / daughter ☐ Other relative:…… & What type of aethritis? Rheumatoid ☐ Osteoarthritis ☐ Other ☐

Drug and Allergy History	If yes, how long ago (or how many times, where applicable) did you use the medication?	
Have you ever used oral corticosteroids (cortisone tablets)?	Yes ☐ No ☐	☐ 3 months ☐ 6 months ☐ 12 months ☐ 24 or more months
Have you ever been given an injection with corticosteroids?	Yes ☐ No ☐	☐ 3 months ☐ 6 months ☐ 12 months ☐ 24 or more months
Have you ever been given an injection of corticosteroids in or around a tendon?	Yes ☐ No ☐	☐ Once ☐ Twice ☐ 3 times ☐ >3 times
Have you ever used anabolic steroids?	Yes ☐ No ☐	☐ 3 months ☐ 6 months ☐ 12 months ☐ 24 or more months
Have you ever used fluoroquinoline antibiotics?	Yes ☐ No ☐	☐ 3 months ☐ 6 months ☐ 12 months ☐ 24 or more months

If yes, please select from the list below:		
☐ ADCO-CIPRIN	☐ CIPROBAY	☐ SANDOZ CIPROFLOXACIN
☐ AVELON	☐ CIPROGEN	☐ TAFLOC
☐ BACTIDRON	☐ CPL ALLIANCE CIPROFLOXACIN	☐ TARIVID
☐ CIFLOC	☐ DYNAFLOC	☐ TAVANIC
☐ CIFRAN	☐ FLOXIN	☐ TEQUIN
☐ CIPLA-CIPROFLOXACIN	☐ MAXAQUIN	☐ UNIQUIN
☐ CIPLOXX	☐ NOROXIN	☐ UTN-400
☐ CIPRO-HEXAL	☐ ORPIC	☐ ZANOCIN
☐ Other_____		

What medication, if any, are you currently using? (please list)	
What allergies do you have? (please list)	

F. OCCUPATIONAL DETAILS	
What is your current occupation?	
What was your occupation prior to injuring your ligament?	
Prior to injury, did your occupation involve lower limb activity?	☐ Yes ☐ No
If yes please indicate which legs.	Right leg ☐ Both legs ☐ Left leg ☐ None ☐

B2.1 Achilles Tendinopathy: Ethics Approval

UNIVERSITY OF CAPE TOWN

Faculty of Health Sciences
Human Research Ethics Committee
Room E52-24 Groote Schuur Hospital Old Main Building
Observatory 7925
Telephone [021] 406 6626 • Facsimile [021] 406 6411
e-mail: shuretta.thomas@uct.ac.za

28 April 2011

Sent via internal mail

HREC REF: 158/2011

A/PROF M COLLINS,
HUMAN BIOLOGY
SPORT SCIENCE INSTITUTE
3RD FLOOR

Dear A/PROF COLLINS,

PROJECT TITLE: THE IDENTIFICATION OF GENETIC RISK FACTORS UNDERLYING ACHILLES TENDON INJURIES: A REPEATABILITY STUDY.

Thank you for submitting your new study to the Faculty of Health Sciences Human Research Ethics Committee

It is a pleasure to inform you that the Ethics Committee has formally approved the above-mentioned study.

Approval is granted until 30 April 2012

Please submit an annual progress report (FHS016) if the research continues beyond the expiry date. Please submit a brief summary of findings if you complete the study within the approval period so that we can close our file.

Please note that the ongoing ethical conduct of the study remains the responsibility of the principal investigator.

Please quote the HREC. REF in all your correspondence.

Yours sincerely

A/PROF MARC BLOCKMAN
CHAIRPERSON, FHS HUMAN ETHICS

Federal Wide Assurance Number: FWA00001637.
Institutional Review Board (IRB) number: IRB00001938

This serves to confirm that the University of Cape Town Human Research Ethics Committee complies to the Ethics Standards for Clinical Research with a new drug in patients, based on the Medical Research Council (MRC-SA), Food and Drug Administration (FDA-USA), International Convention on Harmonisation Good Clinical Practice (ICH GCP) and Declaration of Helsinki guidelines.

The Human Research Ethics Committee granting this approval is in compliance with the ICH Harmonised Tripartite Guidelines E6: Note for Guidance on Good Clinical Practice (CPMP/ICH/135/95) and FDA Code Federal Regulation Part 50, 56 and 312.

B2.2 Inclusion and Exclusion Criteria

Inclusion Criteria:

A clinician reviewed the clinical diagnosis for every participant. Briefly, participants had to have experienced gradual, progressive pain in the region of the Achilles tendon for more than 6 months. Furthermore, at least one other of the following criteria had to be fulfilled regarding the Achilles tendon region:

i) Early morning pain
ii) Early morning stiffness
iii) History of swelling
iv) Tenderness to palpation
v) Palpable nodular thickening
vi) Movement of the affected area with plantar dorsiflexion (positive shift test).

Exclusion Criteria:

Participants with known connective tissue disease or systemic disease known to be linked to tendon pathology were excluded from this study. Such conditions include: Ehlers-Danlos syndrome, benign hypermobility joint syndrome, rheumatoid arthritis, systemic lupus erthyematosus, hyperparathyroidism, renal insufficiency, diabetes mellitus and familial hypercholesterolaemia.

Furthermore, individuals that made use of fluoroquinolone antibiotics or corticosteroid injections prior to injury were also excluded, due to the known increased risk of tendon rupture with use of these drugs.

Department of Human Biology
UCT/MRC RESEARCH UNIT FOR EXERCISE SCIENCE & SPORTS MEDICINE
Faculty of Health Sciences, University of Cape Town
Private Bag, Rondebosch 7700, South Africa
Tel: + 27-21-650-4561 Fax: + 27-21-686-7530

CLINICAL DIAGNOSIS OF ACHILLES TENDINOPATHY

SUBJECT NAME/NUMBER/CODE: _____

Clinical criteria [1;2]	Present
Gradual progressive pain over the posterior lower leg - Achilles tendon area (> 6 weeks)	
Early morning pain	
Early morning stiffness	
History of swelling over the Achilles tendon area	
Tenderness to palpation over the Achilles tendon	
Palpable nodular thickening over the affected Achilles	
Positive "shift" test (movement of the nodular area with plantar-/dorsi-flexion)	

Other criteria	Present
Confirmation of the diagnosis by ultrasound *	
Confirmation of the diagnosis by MRI *	
Confirmation of the diagnosis by CT scan *	

*: One of these criteria must be present to confirm the diagnosis

Date: _____ / _____ / 20____

Investigator: _____

Signature: _____

References:

1. Schepsis AA, Jones H, Haas AL. Achilles tendon disorders in athletes. *Am.J Sports Med* 2002;**30**:287-305.

2. Kader D, Saxena A, Movin T, Maffulli N. Achilles tendinopathy: some aspects of basic science and clinical management. *Br.J Sports Med* 2002;**36**:239-49.

B2.3 Achilles Tendinopathy Informed Consent

Department of Human Biology
UCT/MRC RESEARCH UNIT FOR EXERCISE SCIENCE & SPORTS MEDICINE
Faculty of Health Sciences, University of Cape Town
Private Bag, Rondebosch 7700, South Africa
Tel: + 27-21-650-4561 Fax: + 27-21-686-7530

IDENTIFICATION OF GENETIC RISK FACTORS UNDERLYING ACHILLES TENDON INJURIES: A REPEAT STUDY

INFORMED CONSENT

I, (the participant), have been fully informed about this study on the genetic basis of Achilles tendinopathy to be conducted by the UCT/MRC Research Unit for Exercise Science and Sports Medicine at the University of Cape. I have agreed to donate five millilitres of venous blood or a Buccal mouthwash/swab sample, which will be used for the extraction and analysis of genetic material (DNA). I agree that the blood sample will be taken by a nurse, physician or phlebotomist. I agree to perform a range of motion test (sit-and-reach test or a single leg raise test) to determine my flexibility. I have also agreed to complete personal particulars, sporting participation, personal and family medical history, muscle cramping, as well as, stretching and warm up questionnaires and understand that all the information that is collected during the study will be treated with the strictest confidentiality and will only be used for scientific research purposes. I also understand that my name and personal particulars will not be released under any circumstances and that all data will be analysed anonymously.

I am also prepared to visit the Sports Science Institute of South Africa (SSISA) in Boundary Road, Newlands for a clinical examination in order to confirm my diagnosis (at no cost to myself). If requested, I am prepared to visit the Sports Science Institute of South Africa (SSISA) in Boundary Road, Newlands to undergo an ultrasound or MRI in order to confirm my diagnosis (at no cost to myself). If requested, I am also prepared to visit the Sports Science Institute of South Africa (SSISA) in Boundary Road, Newlands for measurements to determine musculo-tendinous stiffness. I give permission that the study investigators may access my medical records (doctor/physiotherapist) in order to confirm my diagnosis.

I agree to participate in the study and I have been informed that I will be free to withdraw from the study at any time if I so wish. I understand that my DNA sample will be destroyed on completion of the study to identify genetic risk factors associated with Achilles tendon pathology. I also understand that I will be free to request that my DNA sample be destroyed before the completion of the study.

The potential risks associated with blood collection technique from the ante-cubital veins are: infection, delayed healing, haematoma, physical pain, mental discomfort and injury to a nerve or a vessel. These risks are small and will be minimized by the use of trained phlebotomists, use of sterile techniques and the use of disposable, single use materials.

I understand that the DNA will be genotyped (analysed) for variations (polymorphisms) within the type V collagen gene (*COL5A1*), the Tenacin-C gene (*TNC*), the matrix metalloproteinase-3 gene (*MMP3*), the growth and differentiation factor-5 gene (*GDF5*), as well as additional genes, which may become relevant during the course of the investigation.

I understand that whilst there is no direct benefit to myself, if a genetic predisposition for Achilles tendon injuries can be established, then future generations will be able to establish their risk for this condition. This may allow better prevention and treatment options in the future. I understand that I will receive the overall results of the study. I have read (or where appropriate, have had read to me) and understand the information about this study, and any questions I have asked have been answered to my satisfaction. I agree to participate in the study, realising that I have the right to request that my DNA sample be destroyed at any time. I agree that research data provided by me or with my permission during the project may be included in a thesis, presented at conferences and published in journals on the condition that neither my name nor any other identifying information is used.

Any questions regarding this project may be directed to **Dr Alison September** on telephone number **021 650 4559** or e-mail **alison.september@uct.ac.za** or to **Prof Malcolm Collins** on telephone number **021 650 4574** or e-mail **malcolm.collins@uct.ac.za**.

If you have any complaints or queries that the investigator has not been able to answer to your satisfaction, you may contact the Faculty of Health Sciences Human Research Ethics Committee at the University of Cape Town **Prof Marc Blockman** on telephone number **021 406 6452**.

Name of Participant: _____

Signature: _____ **Date:**_____

Name of Researcher:_____

Signature: _____ **Date:**_____

B2.4. Achilles Tendinopathy Participant Information

Department of Human Biology
UCT/MRC RESEARCH UNIT FOR EXERCISE SCIENCE & SPORTS MEDICINE
Faculty of Health Sciences, University of Cape Town
Private Bag, Rondebosch 7700, South Africa
Tel: + 27-21-650-4561 Fax: + 27-21-686-7530

THE IDENTIFICATION OF GENETIC RISK FACTORS UNDERLYING ACHILLES TENDON INJURIES: A REPEAT STUDY

INFORMATION SHEET

Although there is a high incidence of tendon overuse injuries as a result of participation in exercise and sporting activities, the cause(s) of these injuries are poorly understood. Some researchers have suggested that there is a genetic component to exercise-induced tendon injuries. In an attempt to determine whether there is a genetic basis for tendon pathology, we are interested in studying whether certain genes are associated with chronic tendinopathies. This project is being done in the UCT/MRC Research Unit for Exercise Science and Sports Medicine within the Department of Human Biology at the University of Cape Town.

You will be required to visit the Sports Science Institute of South Africa (SSISA) in Boundary Road, Newlands. During the visit, which should take at least 1 hour, you will be asked to donate 5 ml (1 teaspoon) of a blood sample for DNA analysis. You will also be required to (1) perform a range of motion test to determine your flexibility, and (2) complete personal particulars, sporting participation, personal and family medical history, muscle cramping, as well as, stretching and warm up questionnaires. At a later stage, some participants will be asked to visit a doctor (radiologist) for a tendon scan at no cost to themselves.

All the information retrieved from this study will be treated with the strictest confidentiality and will be used only for scientific research purposes. Your name and personal particulars will not be released under any circumstances and all data will be analysed anonymously. Your DNA sample will be destroyed on completion of the study on the genetic basis of tendon pathology. You are also free to request that your DNA sample be destroyed before the completion of the study.

You will not be reimbursed or compensated if you participated in this study. In addition you will not receive personal genetic results. You will however be informed about the overall results of the study and your personal flexibility results.

The University of Cape Town (UCT) has an appropriate insurance policy to cover payment for any trial-related injury.

If you are part of the tendon pathology group, we would appreciate it if you could help us by recruiting two other people of same (or similar) age whom you know and who has trained without suffering any tendon injuries for the control group.

We will keep you informed about the outcomes of this study and look forward to working together with you. If you have any questions about this study, please feel free to contact us at:-

Dr Alison September, PhD	Prof. Malcolm Collins, PhD	Prof. Martin Schwellnus, MBChB, MD
(021) 650 4559	(021) 650 4574	(021) 650 4576
alison.september@uct.ac.za	malcolm.collins@uct.ac.za	martin.schwellnus@uct.ac.za

B2.5 Achilles Tendinopathy Questionnaire

Department of Human Biology
UCT/MRC Research Unit for Exercise Science & Sports Medicine
Faculty of Health Sciences, University of Cape Town
Private Bag, Rondebosch 7700, South Africa
Tel: + 27-21-650-4561 Fax: + 27-21-686-7530

The identification of genetic susceptibility loci underlying Achilles tendon pathology: A Repeat Study.

Instructions

Please answer each question by filling in the details in the allocated space or checking one or more of the option boxes.

Please complete all twelve sections A to L

Subject Number: _____

Version 3
(June 2011)

Subject No: _____

Section A: Personal details			
Surname			
First Name			
Postal Address		Postal/ Zip Code	
E-mail address			
Alternate E-mail address		Phone (day time)	code number
Date of birth	yyyy-mm-dd	Cell (Mobile)	
Height	cm	Sex	Male ☐ Female ☐
Weight	kg	Age	yrs
Ethnic group (Only Required and Used for Research Purposes)	Black/African ☐ White ☐ Indian ☐ Mixed Ancestry (Coloured) ☐ Asian ☐ Other ☐		
Ancestry: Tribal or national background (eg Xhosa, Dutch, Zulu, German, Italian)	Father: Unknown ☐		
	Mother: Unknown ☐		
Country of Birth			
Dominant Hand	Left ☐ Right ☐ Both ☐	Dominant Leg	Left ☐ Right ☐ Both ☐
Current Occupation			
What **percentage** of your **working** day is spent in the following activities?	Sitting: _____ % Standing: _____ % Walking (Lower body activity) _____ % Manual Labour (upper and body activity) _____ %		
Occupation prior to muscle injury?			
Prior to injury, did your occupation involve lower or upper limb activity?	Yes ☐ No ☐		
If yes please indicate which arms/legs.	Right arm ☐ Left arm ☐ Both arms ☐ Right leg ☐ Left leg ☐ Both legs ☐		

2

Subject No: _____

Section B. Sporting Details **Please record your sporting activities in order of importance** **Use an additional form if you participate(d) in more than 6 sports**			
Type of sport(s) you have participated in (please name)	**Main sport 1**	**Other sport 2**	**Other sport 3**
Current or past participation	Current ☐ Past ☐	Current ☐ Past ☐	Current ☐ Past ☐
Year started participation			
Number of years involved in the sport			
Years in competitive sport			
Professional or amateur			
Hours of training per week (last 3 months)			
Hours of training per week (3-12 months)			
Hours of training per week (12-24 months)			

Type of sport(s) you have participated in (please name)	**Other sport 4**	**Other sport 5**	**Other sport 6**
Current or past participation	Current ☐ Past ☐	Current ☐ Past ☐	Current ☐ Past ☐
Year started participation			
Years involved in the sport			
Years in competitive sport			
Professional or amateur			
Hours of training per week (last 3 months)			
Hours of training per week (3-12 months)			
Hours of training per week (12-24 months)			

Subject No: _____

Section E. General Personal Medical History

Do you currently suffer from any of these medical conditions:

☐ High Blood Pressure	☐ Angina/Heart Attack	☐ Asthma
☐ Emphysema	☐ Rheumatoid arthritis	☐ Osteoarthritis (wear & tear)
☐ Malignant disease (cancer)	☐ Elevated Blood Cholesterol	☐ Adrenal disorders
If Yes, what type?	☐ Diabetes mellitus	☐ Thyroid disorders
_____	☐ Renal disease	☐ Amyloidosis

Do you currently suffer from any other Connective Tissue, Rheumatological Or Muscle Diseases & Disorders?	Yes ☐ No ☐	If Yes, please select from the list below

List of some Connective Tissue and/or Rheumatic Diseases and Disorders

☐ Ankylosing Spondylitis	☐ Lipid Storage Diseases	☐ Pseudogout
☐ Aspartylglycosaminuria (AGU)	☐ Marfan Syndrome	☐ Reactive Arthritis
☐ Behcet's Syndrome	☐ Menkes Kinky Hair Syndrome	☐ Reiter's Syndrome
☐ Crohn's Disease	☐ Mucopolysaccharidoses	☐ Relapsing Polychrondritis
☐ Discoid Lupus Erythematosus	☐ Myopathies and Dystrophies	☐ Scleroderma
☐ Ehlers-Danlos syndrome (EDS)	☐ Ochronosis (Homocystinuria)	☐ Sjogren's Syndrome
☐ Eosinophilic Fascitis	☐ Osteogenesis imperfecta (OI)	☐ Systemic Lupus Erythematosus (SLE)
☐ Giant Cell (Temporal) Arthritis	☐ Polyarteritis Nodosa	☐ Systemic Sclerosis
☐ Gout	☐ Polymyalgia Rheumatica	☐ Wegener's Granulomatosis
☐ Hypersentive Vasulatis	☐ Polymyositis & Dermatomyositis	☐ Rhabdomyolysis
☐ Muscular dystrophy	☐ Myopathy	☐ Other_____

What surgical operations have you had? (please list and give dates)	Operation	Date

If female:

At what age did you start menstruating? (years)	
Are you currently using any type of contraception?	☐ Yes ☐ No
If Yes, what type of contraception are you using?	☐ Pill ☐ Injection ☐ IUD

Are you currently?	☐ Pre-menopausal (±12 cycles per year at intervals of 23– 33 days & bleeding lasts 3-7 days) ☐ Menopausal (cycles are irregular and less frequent) ☐ Post-menopausal (no longer menstruating)

Subject No: _____

Section C. Flexibility training history	
Do you perform flexibility training (regular stretching exercises)?	Yes ☐ No ☐
If **YES,** please complete the rest of the flexibility training history section below:- If **NO**, continue completing the questionnaire from section D.	
On average, how many days a week do you perform a stretching session?	days/week
On average, how times a day do you perform a stretching session?	times/day
Please tick which muscle groups do you include in your stretching session?	☐ Hamstrings ☐ Quadriceps ☐ Calf (gastrocnemius) ☐ Calf (soleus) ☐ Groin (inner thigh) ☐ Upper body limbs ☐ Other: _____
Please tick when you stretch? (before, during and/or after exercising. You can tick more than one box)	☐ Before Exercise ☐ During Exercise ☐ After Exercise
When you stretch an individual muscle group, on average, **how long do you hold the stretch** for?	seconds
When you stretch an individual muscle group, on average, **how many times do you stretch the muscle for**?	☐ Once ☐ Twice ☐ 3 times ☐ 4 times ☐ 5 times ☐ 6 or more times

Section D. Lifestyle and habits history			
Please indicate your smoking status	Current smoker ☐	Ex smoker ☐	Never smoked ☐
If you answered yes, (past or current smoker) please complete the section on the right	Number of years of smoking:	If stopped, how many years ago:	
	What is (was) the average number of cigarettes per day:		
On average, how much alcohol do you drink per week (tots, glasses) of spirits, wine or beer?	_____ glasses beer/cider per week _____ glasses wine per week _____ tots of spirits per week		

Subject No: _____

Section E. General Personal Medical History

Do you currently suffer from any of these medical conditions:

☐ High Blood Pressure	☐ Angina/Heart Attack	☐ Asthma
☐ Emphysema	☐ Rheumatoid arthritis	☐ Osteoarthritis (wear & tear)
☐ Malignant disease (cancer)	☐ Elevated Blood Cholesterol	☐ Adrenal disorders
If Yes, what type?	☐ Diabetes mellitus	☐ Thyroid disorders
_____	☐ Renal disease	☐ Amyloidosis

Do you currently suffer from any other Connective Tissue, Rheumatological Or Muscle Diseases & Disorders?	Yes ☐ No ☐	If Yes, please select from the list below

List of some Connective Tissue and/or Rheumatic Diseases and Disorders

☐ Ankylosing Spondylitis	☐ Lipid Storage Diseases	☐ Pseudogout
☐ Aspartylglycosaminuria (AGU)	☐ Marfan Syndrome	☐ Reactive Arthritis
☐ Behcet's Syndrome	☐ Menkes Kinky Hair Syndrome	☐ Reiter's Syndrome
☐ Crohn's Disease	☐ Mucopolysaccharidoses	☐ Relapsing Polychrondritis
☐ Discoid Lupus Erythematosus	☐ Myopathies and Dystrophies	☐ Scleroderma
☐ Ehlers-Danlos syndrome (EDS)	☐ Ochronosis (Homocystinuria)	☐ Sjogren's Syndrome
☐ Eosinophilic Fascitis	☐ Osteogenesis imperfecta (OI)	☐ Systemic Lupus Erythematosus (SLE)
☐ Giant Cell (Temporal) Arthritis	☐ Polyarteritis Nodosa	☐ Systemic Sclerosis
☐ Gout	☐ Polymyalgia Rheumatica	☐ Wegener's Granulomatosis
☐ Hypersentive Vasulatis	☐ Polymyositis & Dermatomyositis	☐ Rhabdomyolysis
☐ Muscular dystrophy	☐ Myopathy	☐ Other_____

What surgical operations have you had? (please list and give dates)	Operation	Date

If female:

At what age did you start menstruating? (years)	
Are you currently using any type of contraception?	☐ Yes ☐ No
If Yes, what type of contraception are you using?	☐ Pill ☐ Injection ☐ IUD

Are you currently?	☐ Pre-menopausal (±12 cycles per year at intervals of 23– 33 days & bleeding lasts 3-7 days) ☐ Menopausal (cycles are irregular and less frequent) ☐ Post-menopausal (no longer menstruating)

5

Subject No: _____

Section F. Family Medical History		
Have any of your blood (biological) relatives ever had the following? **Please tick yes or no. If yes, please tick the relationship of that person to you (You may tick more than one of the relationship blocks).**		
Description		**If Yes, please indicate the relationship**
Chronic **Achilles** tendon injury	Yes ☐ No ☐	☐ Father ☐ Mother ☐ Brother ☐ Sister ☐ Child ☐ Grandfather ☐ Grandmother
Achilles tendon rupture	Yes ☐ No ☐	☐ Father ☐ Mother ☐ Brother ☐ Sister ☐ Child ☐ Grandfather ☐ Grandmother
Any other (not Achilles) tendon injury/rupture	Yes ☐ No ☐	☐ Father ☐ Mother ☐ Brother ☐ Sister ☐ Child ☐ Grandfather ☐ Grandmother
Any ligament injury	Yes ☐ No ☐	☐ Father ☐ Mother ☐ Brother ☐ Sister ☐ Child ☐ Grandfather ☐ Grandmother
Exercise associated muscle cramps	Yes ☐ No ☐	☐ Father ☐ Mother ☐ Brother ☐ Sister ☐ Child ☐ Grandfather ☐ Grandmother
Night muscle cramps	Yes ☐ No ☐	☐ Father ☐ Mother ☐ Brother ☐ Sister ☐ Child ☐ Grandfather ☐ Grandmother
Do any other members of your family suffer from elevated blood cholesterol?	Yes ☐ No ☐	☐ Father ☐ Mother ☐ Brother ☐ Sister ☐ Child ☐ Grandfather ☐ Grandmother
Is there any history of arthritis in your family?	Yes ☐ No ☐	☐ Father ☐ Mother ☐ Brother ☐ Sister ☐ Child ☐ Grandfather ☐ Grandmother
Heart Disease	Yes ☐ No ☐	☐ Father ☐ Mother ☐ Brother ☐ Sister ☐ Child ☐ Grandfather ☐ Grandmother
Diabetes	Yes ☐ No ☐	☐ Father ☐ Mother ☐ Brother ☐ Sister ☐ Child ☐ Grandfather ☐ Grandmother

Subject No: _____

Section G. History of Medication Use			
What medication, if any, are you currently using? (please list)	Name of medication		Years taken
Have you ever used oral corticosteroids (cortisone tablets)? (If **yes**, how long ago?)	Yes ☐ No ☐	☐ 3 months ☐ 12 months	☐ 6 months ☐ 24 or more months
Have you ever been given an injection with corticosteroids? (If **yes**, how long ago?)	Yes ☐ No ☐	☐ 3 months ☐ 12 months	☐ 6 months ☐ 24 or more months
Have you ever used fluoroquinoline antibiotics? (refer to the following list)	Yes ☐ No ☐	☐ 3 months ☐ 12 months	☐ 6 months ☐ 24 or more months

List of some fluoroquinoline antibiotics (may be used in treatment of chlamydia, pneumonia, acute bronchitis, urinary tract infections, skin and soft tissue infection):		
ADCO-CIPRIN	CIPROBAY	SANDOZ CIPROFLOXACIN
AVELON	CIPROGEN	TAFLOC
BACTIDRON	CPL ALLIANCE CIPROFLOXACIN	TARIVID
CIFLOC	DYNAFLOC	TAVANIC
CIFRAN	FACTIVE	TEQUIN
CIPLA-CIPROFLOXACIN	FLOXIN	UNIQUIN
CIPLOXX	MAXAQUIN	UTIN-400
CIPRO-HEXAL	NOROXIN	ZANOCIN
	ORPIC	

Subject No: _____

Section H. Muscle Cramping	
Have you **ever** in your athletic career suffered from **muscle cramping** (painful, spontaneous, sustained spasm of a muscle) during or immediately (within 6 hours) after exercise (in training or competition)	Yes ☐ No ☐
If **YES**, please complete the rest of the muscle cramping section below:- If **NO**, continue completing the questionnaire from section I.	
For how many years have you suffered from cramping?	(years)
Did you suffer from cramping during or after exercise in the **last 12 months**?	Yes ☐ No ☐
With what **type of exercise** is your cramping associated (You can tick more than one form of exercise)?	☐ Swimming ☐ Cycling ☐ Running
In the **last 10 races or training sessions**, how many times have you experienced cramping?	Races: _____/10 Training sessions: _____/10
What treatment/s have you had that **successfully relieved** an acute cramp? (can tick more than one)	☐ Stretching ☐ Resting ☐ Drinking fluid ☐ Ice application ☐ Massage ☐ Magnesium ☐ Salt (tablets or solution) ☐ Other (Specify: _____)
At **what point in the race or training run** do you usually first experience cramping?	☐ First quarter ☐ Second quarter ☐ Third quarter ☐ Fourth quarter ☐ After the race ☐ No pattern
In which **muscles** do you usually cramp (please list the muscle by the one which cramps most frequently (as 1) and the others after that (2-4)?	☐ Calves ☐ Hamstrings ☐ Quadriceps (thigh) ☐ Foot muscles ☐ Other (Specify: _____)
Have you **ever** suffered from cramping in your **whole body** (arms and legs)?	Yes ☐ No ☐
Have you **ever** been **admitted to hospital** following cramping?	Yes ☐ No ☐
Have you **ever** been **confused or in a coma** during or after a cramping episode?	Yes ☐ No ☐
Have you ever had **"dark urine"** in the 3 days following a cramping episode?	Yes ☐ No ☐
If you cramp, **how long** does the cramp usually last for (min)?	(minutes)
If you cramp, how **severe** is the cramp usually? (please tick).	☐ Mild: < 5 minutes and you are able to continue exercising ☐ Moderate: 5-15 minutes and you are able to continue exercising ☐ Severe: >15 minutes or if you have to STOP exercising

8

Subject No: _____

SECTION I. Past History of Skeletal Muscle Injury (Muscle Strain/Tear)

Please complete this section for each muscle injured. If you have had more than one muscle injury additional forms will be available.

Have you ever injured a muscle in the past? Yes ☐ No ☐

If **YES**, please complete the rest of Skeletal Muscle Injury section below:-

If **NO**, continue completing the questionnaire from section J.

If yes, please specify which muscle?
(You may tick more than one block, please select either L (left) or R (right))

Also indicate if you partially or completely tore the muscle.

Partial tear refers to tearing of a few muscle fibres with minor swelling, possible loss of strength and restriction of movement.

Complete tear refers to a tear extending across the whole muscle resulting in complete loss of muscle fuction (loss of strength, movement and ability to contract the muscle).

Muscle Group	Muscle (L-left, R-right)	Partial Tear L	Partial Tear R	Complete Tear L	Complete Tear R
Quadriceps	Vastus Lateralis	☐	☐	☐	☐
	Vastus Medialis	☐	☐	☐	☐
	Vastus Intermedius	☐	☐	☐	☐
	Rectus Femoris	☐	☐	☐	☐
Hamstring	Semitendinosus	☐	☐	☐	☐
	Semimembranosus	☐	☐	☐	☐
	Biceps femoris long	☐	☐	☐	☐
	Biceps femoris short	☐	☐	☐	☐
Hip adductor (groin)	Adductor longus	☐	☐	☐	☐
	Adductor magnu	☐	☐	☐	☐
	Adductor brevi	☐	☐	☐	☐
Calf	Gastrocnemius	☐	☐	☐	☐
	Plantaris	☐	☐	☐	☐
	Soleus	☐	☐	☐	☐
Chronic compartment syndrome of the lower leg	Anterior	Left ☐	Right ☐		
	Lateral	Left ☐	Right ☐		
	Superficial posterior	Left ☐	Right ☐		
	Deep posterior	Left ☐	Right ☐		
Other: (Please Specify)					

How was the muscle injured? **(please also explain exactly how the injury occurred)**

- ☐ Contact with another player
- ☐ Contact with another object (e.g. equipment)
- ☐ No contact (sprinting)
- ☐ No contact (landing)
- ☐ No contact (kicking)
- ☐ No contact (falling)
- ☐ No contact (jumping)
- ☐ No contact (Other)
- ☐ Other: (Please Specify) .....

After sustaining the muscle injury approximately how many days were you off from training or competition?

Approximate number of days: .....

Approximate date of muscle injury?

9

Subject No: _____

Investigation done to confirm the diagnosis	☐ Ultrasound ☐ MRI ☐ CT scan ☐ None	
To your knowledge, have any other members of your family suffered from any muscle pathology?	Yes ☐ No ☐	**If Yes,** please specify the family member ☐ Mother ☐ Father ☐ Sibling ☐ Son / daughter ☐ Other family member: ..... **Condition: Please choose muscle injury from the list above** .....
What was the initial treatment (first 5 days)? **(You may tick more than one block.)**	☐ Rest ☐ Ice application ☐ Compression ☐ Elevation ☐ Immobilisation ☐ Medication (analgesics - pain killers) ☐ Medication (anti-inflammatory drugs) ☐ Other: (Please Specify)	
What was the final treatment? **(You may tick more than one block.)**	☐ Rehabilitation (stretching) ☐ Rehabilitation (strengthening) ☐ Rehabilitation (other) ☐ Strapping/taping ☐ Surgery ☐ *Other: (Please Specify)*	
Following this injury please indicate whether you were able to return to sports (indicate category).	☐ No return to any sport **Return to sport but ...** ☐ Limited to non-sprinting exercise ☐ Limited to non-jumping exercise ☐ Limited, not to same level as pre-injury ☐ Return to full participation in sport	
If you are able to recall, what were the weather and pitch conditions like at the time of injury?	☐ Wet and soft ground ☐ Dry, but soft ground ☐ Dry and firm ground ☐ Wet, but firm ground ☐ *Other: (Please Specify)*	
Associated injuries (Injuries sustained at the same time as the muscle injury)?	☐ Other muscle injury ☐ Tendon injury ☐ Ligament Injury ☐ Bone bruising ☐ *Other: (Please Specify)*	

10

Subject No: _____

Section J. Past History of Tendon, Ligament or Joint Capsule Injury

Please complete this section for each injury. If you have had more than one past injury additional forms will be available.

Have you **ever** in your suffered from **a tendon or ligament injury** (pain, swelling, stiffness) in any tendon (including Achilles tendon, knee tendons, and shoulder tendons) or ligaments (partial or complete tear)?	Yes ☐ No ☐

If **YES,** please complete the rest of the section below:-
If **NO**, continue completing the questionnaire from section L.

	Tendon		Longstanding Pain (Tendinopathy) Left Right	Acute Tear/ Rupture Left Right
Please tick which **tendon/s** you have injured? (next column on the tright) Also indicate (tick) if your injured tendon was longsatnding pain (tendinopathy) or an acute tear/rupture	Foot and ankle:	☐ Achilles tendon ☐ Tibialis posterior ☐ Plantar fascia	☐☐ ☐☐ ☐☐	☐☐ ☐☐ ☐☐
	Knee:	☐ Patellar tendon	☐☐	☐☐
	Elbow and wrist:	☐ Wrist extensor tendon	☐☐	☐☐
	Shoulder:	☐ Rotator cuff	☐☐	☐☐
	Other: _____		☐☐	☐☐

	Ligament	Sprain Left Right	Complete Tear Left Right
Please tick which **ligament/s** you have injured? (next column on the tright) Also indicate if your sprained or completely tore the ligamant.	☐ Shoulder ligaments ☐ Elbow ligaments ☐ Wrist ligaments ☐ Finger ligaments ☐ Knee (ACL) ☐ Knee (MCL) ☐ Knee (PCL) ☐ Knee (LCL) ☐ Ankle lateral ligaments ☐ Ankle medial ligaments ☐ Spinal ligaments ☐ Other: _____	☐☐ ☐☐ ☐☐ ☐☐ ☐☐ ☐☐ ☐☐ ☐☐ ☐☐ ☐☐ ☐☐ ☐☐	☐☐ ☐☐ ☐☐ ☐☐ ☐☐ ☐☐ ☐☐ ☐☐ ☐☐ ☐☐ ☐☐ ☐☐

Have you ever suffered from any of the following joint capsule injuries?	☐ Acute shoulder dislocation ☐ Chronic shoulder instability ☐ Chronic ankle instability ☐ Other: _____

11

Subject No: _____

SECTION K. MEDICAL DETAILS OF TENDON INJURIES				
Symptoms				
How many times have you had tendon injuries? [1]Sudden onset is within a few seconds or minutes [2]Gradual onset is over days or weeks	Tendon Injured	Date of Injury	Acute or Chronic Injury	Sudden[1] or Gradual[2] Onset
	1			
	2			
	3			
	4			
	5			

Please complete a **separate form**, Part K only, for each Tendon Injury you have had	
Injury Number (1,2,3,4,or 5)	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ _____
Which tendon did you injure?	Rotator cuff tendon ☐ • Supraspinatus ☐ • Infraspinatus ☐ • teres minor ☐ Patellar tendon ☐ Wrist extensor tendons ☐ Achilles tendon ☐
Which side was injured?	☐ Left ☐ Right ☐ Both
Which region of your tendon was injured? Please indicate on a diagram. (Only if applicable)	☐ Upper 1/3 ☐ Middle 1/3 ☐ Lower 1/3
To what extent was your Tendon ruptured?	☐ Complete ☐ Partial ☐ None
How were you injured? (e.g. sport, walking)	
Grade of injury **at the time** of injury	☐ pain only after exercise ☐ pain during exercise, but did not cause you to alter training ☐ pain during exercise, which causes you to alter training ☐ pain which causes you to stop training ☐ no pain ☐ not sure ☐ Other (Specify _____)
Grade of injury **currently**	☐ pain only after exercise ☐ pain during exercise, but did not cause you to alter training. ☐ pain during exercise, which causes you to alter training ☐ pain which causes you to stop training ☐ no pain ☐ not sure ☐ Other (Specify _____)

Subject No: _____

Which of the following symptoms were present **before** the injury	☐ Pain (less than 1 week) ☐ Stiffness ☐ Pain (1-4 weeks) ☐ Swelling ☐ Pain (> 4 weeks) ☐ None
Which of the following symptoms were present **after** the injury	☐ Pain (less than 1 week) ☐ Stiffness ☐ Pain (1-4 weeks) ☐ Swelling ☐ Pain (> 4 weeks) ☐ None
If you have or had chronic tendon pain, what seems to alleviate the pain?	
Diagnosis	
Which type of Tendon Disease were you diagnosed with e.g. Rupture, Tendinitis, etc.	
Diagnosed by (Please indicate the name and contact number of the clinician who diagnosed you)	☐ Doctor _____ ☐ Physiotherapist _____ ☐ Biokineticist _____ ☐ Podiatrist _____ ☐ Other _____
If you had a tendon rupture. How was it treated?	☐ Surgically ☐Non-surgically
If applicable, who was the surgeon?	Surgeon _____Phone _____
If applicable, what diagnostic imaging was performed?	☐ Ultrasound ☐ MRI ☐ CT Other _____
If applicable, who did the imaging?	Clinician _____Phone_____

13

Subject No: _____

Section L. Details of Any Other Chronic (Longstanding) Current Injury

Please complete this section for each injury. If you have had more than one past injury additional forms will be available.

What was the approximate date when you first became aware of the injury?	Month Year
Please indicate which side of your body is injured (if applicable)	☐ Right ☐ Left
Please indicate which anatomical area is currently injured	☐ Head ☐ Neck ☐ Face ☐ Front chest ☐ Back chest ☐ Shoulder ☐ Upper arm ☐ Elbow ☐ Forearm ☐ Wrist ☐ Finger ☐ Lower back ☐ Hip ☐ Thigh ☐ Hamstring ☐ Quadriceps ☐ Knee ☐ Shin ☐ Achilles ☐ Ankle ☐ Foot Other (Specify: _____)
Please indicate the type of structure that was injured	☐ Muscle ☐ Tendon ☐ Bone ☐ Ligament ☐ Joint Other (Specify): _____)
Please indicate in which sport (discipline) the injury occurred	☐ Running ☐ Soccer ☐ Rugby ☐ Cricket ☐ Hockey Other (Specify: _____)
Please indicate the severity of the injury (tick one box please)	☐ I only experience symptoms after exercise - Grade 1 ☐ I experience symptoms during exercise, but it does not interfere with exercise - Grade 2 ☐ I experience symptoms during exercise that may interfere with my training/ competition - Grade 3 ☐ I am so painful that I may not be able to train or compete - Grade 4
Please indicate how your injury was treated to date (you can tick more than one)?	☐ Rest ☐ Stretches ☐ Physiotherapy ☐ Surgery ☐ Strengthening exercises ☐ Equipment change ☐ Tablets ☐ Cortisone injection ☐ Other injection ☐ Orthotics Other (Specify: _____)

Appendix C- Detailed Methods and Solutions

C.1 DNA Extraction from Blood

Briefly, the blood samples were allowed to thaw at room temperature, and then transferred to 15ml polypropylene tubes. Two volumes of TKM1 buffer (10mM Tris-HCl pH 7.6, 10mM KCl, 10mM MgCl2 and 2mM EDTA) containing 2.5% Nonidet P-40 (Sigma, St. Louis, MO, USA) was added to lyse the red blood cells, followed by a 10 min incubation at room temperature. The white blood cells were pelleted by centrifugation at 1200 x g at room temperature for 10 minutes and washed with one volume of TKM1 buffer. The pellet was resuspended and pelleted in the same manner until the pellet was clean. The washed pellets were resuspended in 800 ml of TKM2 buffer (10mMTris-HCl pH 7.6, 10mM KCl, 10mM MgCl2, 0.4M NaCl2and 2mM EDTA) to which 50 ml of 10% sodium dodecylsulfate was added, and incubated for 60 minutes at 55 °C or until the pellets had dissolved. One hundred and fifty microlitres of 5M NaClO4 and 500 µl of chloroform were added to each sample, which was then mixed by vortexing for 15–20 seconds. The samples were transferred to 1.5ml microfuge tubes and the protein precipitated by centrifugation at 13 000 r.p.m. (15 000 x g) for 5 min at room temperature. Five hundred microlitres of the top aqueous phases was transferred to new microfuge tubes containing 1ml of absolute ethanol, mixed and the DNA pelleted by centrifugation at 13 000 r.p.m.(15 000 x g) for 2 min at room temperature. The precipitated DNA was air dried and resuspended in 200 µl of TE buffer (10mM Tris-HCl, 1mM EDTA, pH 8.0) and incubated for 1 h at 65 °C or over-night at room temperature.

C.2 Solutions

10% SDS (200ml)

20 g SDS
200 ml dH_2O
autoclave

1 x TE Buffer (pH 8.0) (100 ml)

0.121 Tris base
0.037 g EDTA
dH_2O to 100 ml
HCl to correct pH
autoclave

5M $NaClO_4$ (100 ml)

61.2 g $NaClO_4$
dH_2O to 100 ml
autoclave

TKM1 buffer

10mM Tris-HCl pH 7.6
10mM KCl
10mM $MgCl_2$
2mM EDTA

TKM1 Buffer +NP-40

10mM Tris-HCl pH 7.6
10mM KCl
10mM $MgCl_2$
2mM EDTA
2.5% Nonidet P-40

TKM2 buffer

10mMTris-HCl pH 7.6
10mM KCl
10mM $MgCl_2$
0.4M $NaCl_2$
2mM EDTA