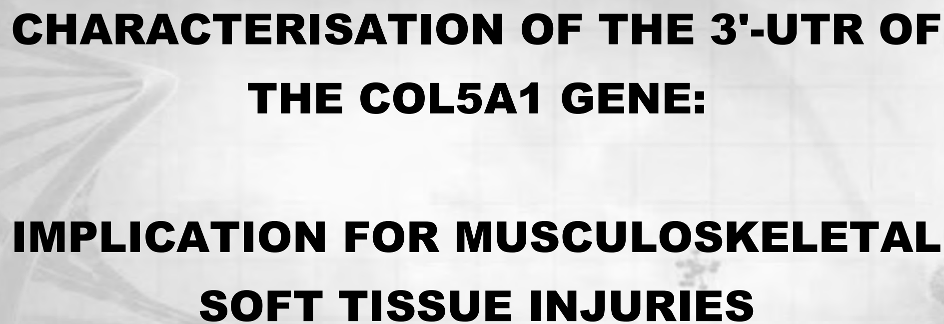


CHARACTERISATION OF THE 3'-UTR OF THE COL5A1 GENE:

IMPLICATION FOR MUSCULOSKELETAL SOFT TISSUE INJURIES

BY
MARY-JESSICA NANCY LAGUETTE

Supervisor:
Prof Malcolm Collins

Co-supervisor:
A/Prof Sharon Prince

This thesis is presented for the degree of
DOCTOR OF PHILOSOPHY
In the Department Of Human Biology
Faculty of Health Sciences
University of Cape Town
South Africa,

November 2014

UCT/MRC Research Unit for Exercise Science and Sports Medicine
Sports Science Institute of South Africa
Boundary Road, Newlands, 7700
Cape Town
South Africa

University of Cape Town

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

University of Cape Town

To my parents, Mary-Jane and Alain Laguette,

TABLE OF CONTENTS

Table of contents ii

Acknowledgements v

List of Scientific outputs from this Thesis vi

Abbreviations viii

List of figures from this dissertation xii

List of tables from this dissertation xvii

Abstract xix

Chapter 1 LITERATURE REVIEW 1

1.1 Introduction 1

1.2 Musculoskeletal Soft Tissue Injuries 3

1.3 Structure of the Achilles tendon 5

1.4 Risk factors and Models of Pathology for Tendinopathy 10

1.5 The Collagen Family and Fibril Forming Collagens 18

1.6 Type V Collagen and the *COL5A1* Gene 27

1.7 *COL5A1* and *MIR60*8 are associated with musculoskeletal soft tissue injuries and/or other phenotypes 34

1.8 Eukaryotic 3'-UTRs 45

1.8.1 Polyadenylation signals 46

1.8.2 The secondary RNA structure 49

1.8.3 MicroRNA binding sites 50

1.8.4 Protein binding sites 57

1.8.5 Characteristics of the *COL5A1* 3'-UTR 60

1.9 Aim of the study 63

Chapter 2 MATERIAL AND METHODS 64

2.1 Participants used for the Cloning of the 3'-UTR 64

2.2 Cloning the *COL5A1* 3'-UTR Reporter Gene Constructs 68

2.3 Generation of Deletion Constructs of the COL5A1 3'-UTR 72

2.4 Generation of Site-Directed Mutants within the *COL5A1* 3'-UTR 73
2.5 Production of Competent Cells and Transformation 78
2.6 Maxi Preparation of Plasmid DNA by Caesium Chloride and Ethidium Bromide Gradient Centrifugation 79
2.7 Sequencing 81
2.8 Annotation of the Template Sequence 83
2.9 Cell Culture 84
2.10 Mycoplasma Testing 84
2.11 Transient Co-Transfections and Luciferase Assays 85
2.12 Messenger RNA Decay Assay 85
2.13 Nuclear and Cytoplasmic Protein Harvest for an RNA Electrophoretic Mobility Shift Assay (EMSA) 87
2.14 Non-radioactive RNA EMSA 89
2.15 Participants for Genetic Association Study 91
2.16 DNA Extraction for Genetic Association Study 94
2.17 *COL5A1* 3'-UTR Genotyping 94
2.18 Preliminary Gene Expression Study using Skin Biopsy 99
2.19 Derivation of Participant's Primary Fibroblast Cell Line from Skin Biopsy 103
2.20 Expression levels of *COL5A1* and *COL1A1* in Primary skin fibroblast cell lines .. 104
2.21 Statistics 106
Chapter 3 RESULTS 108
3.1 Overall Increase in *COL5A1* mRNA Stability in the Tendinopathic Phenotype 109
3.2 Identification of Two Major Allelic Forms of the *COL5A1* 3'-UTR 110
3.3 Identification of a Region of the *COL5A1* 3'-UTR that Confers mRNA Stability ... 116
3.4 The SVWI-38 cells exhibit No Overall Increase in *COL5A1* mRNA Stability in the Tendinopathic Phenotype. 120
3.5 The Newly Annotated Variants (rs71746744, rs16399 and rs1134170) are also Associated with Achilles Tendinopathy 121
3.5.1 Characteristics of Participants 121
3.5.2 *COL5A1* 3'-UTR Genotype Frequencies 127

3.5.3 Inferred Haplotype Study 133

3.5.4 Gene-gene Interactions 135

3.6 Independently, the variants rs12722 and rs71746744 are not responsible for the tendinopathic phenotype.. 138

3.7 Establishment of Primary Skin Fibroblast Cell Lines 142

3.8 A Preliminary Study shows an Increase in the Expression Levels of *COL5A1* and *COL1A1* for Individuals with a T allele at rs12722.. 145

3.9 Differential Binding of RNA Binding Proteins (RBPs) to the C- and T-allelic forms. 149

Chapter 4 DISCUSSION 158

4.1 The tendinopathic phenotype has an overall increase in *COL5A1* mRNA stability and expression. 158

4.2 Identification of C- and T-allelic forms of the *COL5A1* 3'-UTR............................ 162

4.3 Variants rs71746744 (-/AGGG), rs16399 (ATCT/-) and rs1134170 (A/T) are also associated with Achilles tendinopathy 165

4.4 A 57bp Region of the *COL5A1* 3'-UTR Confers mRNA Stability.......................... 167

4.5 Differential Binding of RNA Binding Proteins (RBPs) to the C- and T-allelic forms. 169

4.6 In Perspective 170

4.7 CONCLUSION 178

Bibliography 180

Appendices 208

A.1 Appendix A 208

A.2 Appendix B 248

A.3 Appendix C 252

A.4 Appendix D 255

A.5 Appendix E 261

ACKNOWLEDGEMENTS

I would like to dedicate this thesis to the following individuals:

- My amazing supervisors, Malcolm and Sharon, for their vibrant personalities, heroic patience, excellent guidance and constant support. I have been blessed with the best supervisory duo a student could ever dream of. Thank you for being such amazing role models and teachers! Thank you for the opportunities, your wisdom and your encouragement throughout the years!

- My parents, Mary-Jane and Alain, and family for their emotional and financial support: their enthusiasm, interest and love carried me through the years. You have made many sacrifices to help each of your children achieve career and life goals. You never seize to amaze me. Thank you to my amazing siblings Katty, Julie, William, Bruce and extended family Yann and Dav for their love. I am so privileged to have you in my life.

- My boyfriend, Rick, for the joy and happiness that he brings to my life. I cannot say thank you enough for stepping in each time I needed help, for the stimulating conversations and the encouragements. Thank you for the space to write, the work sessions and the patience. Thank you for the runs and adventures to refuel. An extended thank you to your family for being so lovely and encouraging.

I would like to express my gratitude to the following individuals:

- Thank you Alison for your contagious enthusiasm for the research we do. Thank you for teaching me throughout the years, for your wisdom, for building confidence, getting me involved and for the stimulating discussions.

- Thank you Dr Caroline D'Alton, Kevin, Courtnee and Yoonus for their involvement with aspects of the laboratory work. You are amazing!

- Thank you A/Prof A. September, Dr J. Peres, Dr M. Posthumus and Miss M. Rahim for taking the time to read part of this thesis and for your feedback.

- Thank you to my officemates, past and present, especially Marilize, Kyle and girlfriend Liz for making writing a thesis a pleasant experience (Cake!). Thank you Marilize, Nick and Kevin for their drive and cheers to the finishing line!

- Thank you to EVERY team member of the Prince's Lab and the Biochemistry Lab at ESSM for their support and friendship: this thesis would not be possible without you!

- Thank you to the outstanding supporting Staff at ESSM and HUB, specially Neezaam, Trevino, Fiona, Lesa, Lance and Ayesha for everything!

This thesis was supported, in part, by funds from the National Research Foundation (NRF) of South Africa, the University of Cape Town (UCT) and the Medical Research Council (MRC) of South Africa

LIST OF SCIENTIFIC OUTPUTS FROM THIS THESIS

ARTICLES IN INTERNATIONAL PEER-REVIEWED JOURNALS

1. Abrahams Y., **Laguette M-J.**, Prince S. and 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 May;77(3):204-14

2. **Laguette M-J.**, Abrahams Y., Prince S. and Collins M. Sequence Variants within the 3'-UTR of the *COL5A1* gene alters mRNA stability: Implications for musculoskeletal soft tissue injuries; **Matrix Biol.** 2011 Jun;30(5-6):338-45

PRESENTATIONS AT INTERNATIONAL CONGRESSES

1. **Laguette M-J.**, Abrahams Y., Prince S. and Collins M. Genetic Variants within the *COL5A1* 3'-UTR: Possible Role in the Aetiology of Musculoskeletal Soft Tissue Injuries. **GRC Collagen Conference 2013**, New London, NH, United States of America (Poster Presentation)

2. Collins M., **Laguette M-J.**, Abrahams Y., Posthumus M., Prince S. Functional analysis of the 3'-UTR of the *COL5A1* Gene: Implications for 'exercise-related phenotypes'. **16th Annual Congress of the European College of Sport Science (ECSS)**, Liverpool, United Kingdom, 2011 (Oral Presentation)

3. **Laguette M-J.**, Abrahams Y., Prince S. and Collins M. Sequence variants within the 3'-UTR of the *COL5A1* gene alter mRNA stability: Implications for musculoskeletal soft tissue injuries. **36th FEBS Congress: Biochemistry for Tomorrow's Medicine**, Torino, Italy, 2011 (Oral and Poster Presentation)

4. **Laguette M-J.**, Prince S. and Collins M. Sequence variants within the 3'-UTR of the *COL5A1* gene alter mRNA stability: Implications for musculoskeletal soft tissue injuries. **IOC World Conference on Prevention of Injury and Illness**, Monaco, 2011 (Oral Presentation)

5. **Laguette M-J.**, Prince S. and Collins M. Sequence variants within the 3'-UTR of the *COL5A1* gene alter mRNA stability: Implications for musculoskeletal soft tissue injuries. **The Joint Congress of the African and Southern African Societies of Human Genetics**, Cape Town, 2011 (Poster Presentation)

6. **Laguette M-J.**, Prince S. and Collins M. Sequence Variants within the 3'-Untranslated region of the *COL5A1* gene alters mRNA stability in patients diagnosed with chronic Achilles Tendinopathy. **The International Scientific Tendinopathy Symposium**, Umeå, Sweden, 2010 (Oral Presentation)

PRESENTATIONS AT LOCAL CONGRESSES

1. Collins M., **Laguette M-J.**, Abrahams Y., Clark C. and Prince S. The role of Type V Collagen in Occupational and Recreational Musculoskeletal Soft Tissue Injuries. **42nd Annual Conference of the Physiology Society of Southern African (PSSA)**, Umhlanga, South Africa, 2014 (Oral Presentation).

2. **Laguette M-J.**, Prince S. and Collins M. Polymorphisms within the *COL5A1* 3'-UTR may play an important role in the aetiology of musculoskeletal soft tissue injuries. **Congress of the South African Society of Biochemistry and Molecular Biology 2014**, Cape Town, South Africa. (Oral Presentation)

3. **Laguette M-J.**, Abrahams Y., Prince S. and Collins M. Genetic variants within the *COL5A1* 3'-UTR: Possible Role in the Aetiology of Musculoskeletal Soft Tissue Injuries. **MRC Early Career Scientists Conference 2012**, Cape Town, South Africa (Oral Presentation)

4. **Laguette M-J.**, Abrahams Y., Prince S. and Collins M. Sequence variants within the 3'-UTR of the *COL5A1* gene alter mRNA stability: Implications for musculoskeletal soft tissue injuries. **SASBMB/FASBMB congress 2012**, Drakensberg, South Africa (Oral Presentation)

5. **Laguette M-J.**, Prince S. and Collins M. Sequence variants within the 3'-UTR of the *COL5A1* gene alter mRNA stability: Implications for musculoskeletal soft tissue injuries. **Research Day 2011: Departments of CLS and Human Biology**, UCT, Cape Town. (Oral Presentation)

6. **Laguette M-J.**, Prince S. and Collins M. Sequence variants within the 3'-UTR of the *COL5A1* gene alter mRNA stability: Implications for musculoskeletal soft tissue injuries. **Young Researchers Forum**, Cape Town, 2011 (Poster Presentation)

7. **Laguette M-J.**, Prince S. and Collins M. Sequence Variants within the 3'-UTR region of the *COL5A1* gene alters mRNA stability in patients suffering from Achilles tendinopathy. **The 22nd Congress of the South African Society of Biochemistry and Molecular Biology**, Bloemfontein, 2010 (Oral Presentation)

ABBREVIATIONS

ACL	Anterior cruciate ligament
APS	Ammonium persulphate
ATP	Adenosine triphosphate
AUS	Australia
Ave	Average
BCA	Bicinchoninic acid
BJHS	Benign Joint Hypermobility Syndrome
BMI	Body mass Index
bp	Base pair
BSA	Bovine serum albumin
CASP8	Gene encoding for Caspase-8
CI	Confidence Interval
cm	Centimetre
COL	Collagenous domain
COL1A1	Gene encoding for the α1 chain of type I collagen
COL1A2	Gene encoding for the α2 chain of type I collagen
COL2A1	Gene encoding for the α1 chain of type II collagen
COL3A1	Gene encoding for the α1 chain of type III collagen
COL4A1	Gene encoding for the α1 chain of type IIII collagen
COL4A2	Gene encoding for the α2 chain of type IIII collagen
COL5A1	Gene encoding for the α1 chain of type V collagen
COL5A2	Gene encoding for the α2 chain of type V collagen
COL9A2	Gene encoding for the α2 chain of type IX collagen
COL11A1	Gene encoding for the α1 chain of type XI collagen
COL11A2	Gene encoding for the α2 chain of type XI collagen
COL15A1	Gene encoding for the α1 chain of type XV collagen
COL27A1	Gene encoding for the α1 chain of type XXVII collagen
CON	Physically fit asymptomatic control individuals
CPSF	Cleavage and polyadenylation specific factor
CsCl	Caesium Chloride
C_T	Threshold cycle
D'	Normalised linkage disequilibrium
DCN	Gene encoding for Decorin
DEPC	Diethylpyrocarbonate
DGCR8	Gene encoding for the DiGeorge Syndrome Critical Region/Microprocessor Complex Subunit
DMEM	Dulbecco's modified Eagle's medium
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid

dNTPs	Deoxynucleotide triphosphates
DTT	Dithiothreitol
EAMC	Exercise-associated muscle cramping
ECM	Extracellular matrix
EDS	Ehlers-Danlos syndrome
EDTA	Ethylenediaminetetraacetic acid
EGTA	Ethylene glycol tetra-acetic acid
EIF4B	Eukaryotic translation initiation factor 4B
ELAVL1	Embryonic lethal, abnormal vision, Drosophila-like 1
EMSA	Electrophoretic mobility shift assay
EST	Expressed sequence tag
FACIT	Fibril-associated collagen with interrupted triple helix
FAM	6-carboxyfluorescein
FBN2	Gene encoding for Fibronectin-2
FBS	Fetal bovine serum
FWD	Forward primer
GDF5	Gene encoding for Growth differentiation factor-5
GLY	Glycine
HNRNPA1	Heterogeneous nuclear ribonucleoprotein A1
HRM	High resolution melting
HRP	Horseradish peroxidase enzyme
HWE	Hardy-Weinberg equilibrium
IL-1RN	Gene encoding for the Interleukin-1 receptor antagonist
IL-1β	Gene encoding for the Interleukin-1 β
IL-6	Gene encoding for the Interleukin-6
kb	Kilobase
kg	Kilogram
KHDRBS3	KH domain-containing ribonucleic acid-binding signal transduction-associated protein 3
LD	Linkage disequilibrium
LUC	Luciferase reporter
m	Metre
mA	Milliampere
MIR608	Gene encoding for micro ribonucleic acid Hsa-miR-608
miRNA	Micro ribonucleic acid
miRISC	miRNA-induced silencing complex
ml	Millilitre
mm	Millimetre
MMP3	Gene encoding for Matrix metalloproteinase-3
mRNA	Messenger ribonucleic acid

NC	Non-collagenous domain
ng	Nanogram
NONO	Non-pou domain-containing octamer-binding protein
OD	Optical density
OR	Odds ratio
PABP	Poly(A)-binding protein
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline
PBS/T	0.1% (v/v) Tween 20 in 1x phosphate-buffered saline
PCR	Polymerase chain reaction
PIPES	Piperazine-N,N′-bis(2-ethanesulfonic acid)
pmol	Picamole
poly(A)	Polyadenylation
PTBP1	Polypyrimidine tract-binding protein 1
Q-RT-PCR	Quantitative real-time reverse transcription polymerase chain reaction
RBP	ribonucleic acid binding protein
RE	Restriction endonuclease enzyme
REV	Reverse primer
RFLP	Restriction fragment length polymorphism
RNA	Ribonucleic acid
RNAse	Ribonuclease (ribonucleic acid degrading) enzyme
RNP	Ribonucleoprotein
ROM	Range of motion
rpm	Revolutions per minute
SA	South Africa
SD	Standard Deviation
SDM	Site-directed mutagenesis
SDS	Sodium dodecyl sulphate
SE	Skin explant
siRNA	Small interfering ribonucleic acid
SNP	single nucleotide polymorphism
SOB	Super optimal broth
SOC	Super optimal broth with catabolite repression
SR	Sit and Reach
STREGA	Strengthening the reporting of genetic association studies initiative
STROBE	Strengthening the reporting of observational studies in epidemiology statement
SV40	*Simian virus 40*
TBE	Tris/Borate/EDTA buffer
TE	Tris/EDTA buffer
TEM	Transmission electron micrograph
TEMED	Tetramethylethylenediamine
TEN	Achilles tendinopathy

TFB	Transformation broth
TFB1	Transformation buffer 1
TFB2	Transformation buffer 2
TIMP2	Gene encoding for Tissue inhibitor of MMP-2
TNC	Gene encoding for Tenascin C
tRNA	Transfer ribonucleic acid
UTR	Untranslated region
UV	Ultraviolet light
VIC	4,7,2′-trichloro-7′-phenyl-6-carboxyfluorescein
WT	Wild-type
yrs	Years
µg	Microgram
µl	Microlitre
µM	Micromolar

LIST OF FIGURES FROM THIS DISSERTATION

Figure 1.1 Lateral and posterior view of the lower leg, indicating the position of the Achilles tendon (bold) modified from Asplund and Best (2013).. 7

Figure 1.2 A schematic diagram representing the hierarchical structure of the tendon unit (modified from Wang, 2006)... 9

Figure 1.3 Model representing the role of the intrinsic factors in predisposing the individual to become susceptible to Achilles tendinopathy, modified from Meeuwisse (1994) and Ribbans and Collins (2014). 12

Figure 1.4 Schematic representation of fibril-forming collagen biosynthesis from transcription through mRNA processing, ribosomal protein synthesis followed by post-translational modifications and secretion to fibril formation (modified from Gelse *et al.* 2003). 26

Figure 1.5 A schematic diagram modified from Raleigh and Collins (2009) representing the basic structural unit of the collagen fibril which consists predominately of type I collagen (not drawn to scale).. 28

Figure 1.6 Transmission electron micrographs (TEMs) modified from Wenstrup *et al.* 2011, Wenstrup *et al.* 2006 and Sun *et al.* 2011 showing the effect of haploinsufficiency of mice *col5a1* in (Panel A) dermis and (Panel B) tendon. Panel C, the corneal-stroma-specific *Col5a1*-null mouse model showing the fibril implications in the stroma of the mice. 33

Figure 1.7 Nucleotide sequence of *COL5A1* exon 66 containing the first 5 sequence variants (Panel A) and the latter 2 sequence variants (Panel B) (September *et al.*, 2009). 37

Figure 1.8 Representation of the 3'-UTR of the *COL5A1* gene (9q34) at its terminal end showing exons 65 and 66 with the intronic region in between (intron 65), and the *MIR608* gene(10q24) modified from September *et al.* (2009). 38

Figure 1.9 Diagram illustrating the common association of the T or C allele of the rs12722 in the context of injury risk, EAMC, flexibility and endurance performance as presented in the genetic association studies 42

Figure 1.10 Integral model of the relationship between *COL5A1* genotype (black boxes), connective tissue biochemical and mechanical properties (white boxes), flexibility, disease or injury risk as well as physical activity, modified from Collins and Posthumus (2011).. 43

Figure 1.11 Diagram depicting (A) the three sites important in determining where the poly(A) addition will occur in the 3'-UTR, namely the hexanucleotide poly(A) signal, the cleavage site (also known as the poly(A) site), and the GU-rich downstream element; (B) the mRNA 'closed loop' structure formed from cooperation of the 5'-cap and the 3'-UTR through the poly(A) tail modified from Huntzinger and Izaurralde (2011). 48

Figure 1.12 Diagram illustrating (A) the ability of miRNAs to target multiple genes; (B) many gene targets have several seed matches for multiple miRNA binding in their 3'-UTRs and (C) a complex network of mutual interactions exist between miRNA and their targets (modified from Peter, 2010). 51

Figure 1.13 The biogenesis of miRNA modified from Rutnam *et al.* (2013).. 53

Figure 1.14 Schematic diagram representing the main types of miRNA genetic variations with a potential role in disorders (adapted from Meola *et al*. 2009). 56

Figure 1.15 A diagram illustrating the role of RBPs in the regulation of post-transcriptional gene expression, modified from Glisovic *et al.* (2008).. 59

Figure 1.16 Schematic representation of *COL5A1* 3'-UTR on exon 66 (the open rectangle), showing the numerous variants investigated in genetic association studies (boxed below with arrows). 62

Figure 2.1 Schematic drawing representing the cloning strategy of the human *COL5A1* 3'-UTR (left, in red) into the pGL3-Promoter vector (right, in black), substituting the SV40 late poly (A) signal (blue box) of the firefly Luc+ reporter gene (yellow box). 68

Figure 2.2 Nucleotide sequence of *COL5A1* exon 66 (PubMed accession no. NM_000093) containing the 3'-untranslated region (UTR) (nucleotides 148 to 2688). 71

Figure 2.3 Construction of two deletion constructs of the *COL5A1* 3'-UTR. 72

Figure 2.4 Schematic diagram representing the site-directed mutations (Grey boxes) that were generated in the *COL5A1* 3'-UTR cloned from the TEN and CON participants. 77

Figure 2.5 A typical image of a 6% PAGE gel discriminating restriction fragments from individuals who are homozygous AA (159 and 34 bp, lane 3), homozygous TT (193, lane 4) and heterozygous AT (193, 159 and 34 bp, lane 5) upon digestion of the 193 bp PCR products (Uncut: U, lane 2) of rs1134170 with the *PSh*AI restriction enzyme. 97

Figure 2.6 A typical image of a 6% PAGE gel discriminating restriction fragments from individuals who are homozygous CC (316, 271 and 80bp, lane 3), homozygous TT (351 and 316bp, lane 4) and heterozygous CT (351, 316, 271 and 80 bp, lane 5) upon digestion of the 667 bp PCR product (Uncut: U, lane 2) of rs12722 with the *Bst*UI restriction enzyme. 102

Figure 3.1 Functional analysis of the 2.5 kb *COL5A1* 3'-untranslated region (UTR) in HT1080 cells. 111

Figure 3.2 The seven tightly linked polymorphic sites within the *COL5A1* 3'-untranslated region (UTR) are annotated as black boxes. 113

Figure 3.3 The relative luciferase activity of the pooled control (CON, black bars) and tendinopathy (TEN, grey bars) clones, containing the COL5A1 3'-UTR C- (CON 2.1, CON 3.2 and CON 5.1) or T- (TEN 1.1, TEN 2.2, TEN 3.2 and TEN 5.1) allelic form, in HT1080 cells. 114

Figure 3.4 Schematic representation of the human *COL5A1* 3'-untranslated region (UTR). 115

Figure 3.5 The *COL5A1* 3'-untranslated region (UTR) deletion constructs in HT1080 cells.. 117

Figure 3.6 Functional analysis of the *COL5A1* 3'-untranslated region (UTR) Δ488 deletion constructs in HT1080 cells.. 118

Figure 3.7 Functional analysis of the *COL5A1* 3'-untranslated region (UTR) Δ57 deletion constructs in HT1080 cells. 119

Figure 3.8 Functional analysis of the 2.5 kb *COL5A1* 3'-untranslated region (UTR) in SVWI-38 cells with respect to the C- and T-allelic forms. 120

Figure 3.9 A genetic study investigating the variants rs71746744, rs16399 and rs1134170 (grey arrows). 122

Figure 3.10 Genotype frequency distributions for the *COL5A1* 3'-untranslated region (UTR) polymorphisms, (A) rs71746744 (-/AGGG), (B) rs16399 (ATCT/-) and (C) rs1134170 (A/T), in control (CON) and chronic Achilles tendinopathy (TEN) for the combined SA and AUS (SA+AUS) cohorts. 132

Figure 3.11 The frequency distribution of 8 of the possible 36 inferred haplotypes with a frequency >1% constructed from *COL5A1* 3'-UTR variants rs13946 (C/T), rs12722 (C/T),

rs3196378 (C/A), rs71746744 (-/AGGG), rs16399 (ATCT/-) and rs1134170 (A/T) in the combined Australian and South African Achilles tendinopathy (TEN, white bars) and control (CON, black bars) groups. 134

Figure 3.12 The frequency distribution of 4 of the possible 8 inferred haplotypes with a frequency >1% constructed from *COL5A1* 3'-UTR variants rs71746744 (-/AGGG), rs16399 (ATCT/-) and rs1134170 (A/T) in the combined Australian and South African Achilles tendinopathy (TEN, white bars) and control (CON, black bars) groups. 135

Figure 3.13 Genotype risk score frequency distributions of the Hsa-miR-608 gene (*MIR608*) rs4919510 (C/G) polymorphism and the three *COL5A1* 3'-untranslated region (UTR) polymorphisms, (i) rs71746744 (-/AGGG) polymorphism, (ii) rs16399 (ATCT/-), and (iii) rs1134170 (A/T) in the pooled South African and Australian control (CON, clear bars) and chronic Achilles tendinopathy (TEN, black bars) groups. 137

Figure 3.14 Examination of the effect of the variant rs12722 on the mRNA stability of the *COL5A1* 3'-UTR. 140

Figure 3.15 Examination of the effect of variant rs71746744 on the mRNA stability of the *COL5A1* 3'-UTR (C-allelic form). 141

Figure 3.16 (A, B) High-power photomicrographs showing migrating cells (see arrows) out of skin explants (SE) after approximately a week in culture. 144

Figure 3.17 Relative mRNA abundance of (A) *COL5A1* (clear bars) and (B) *COL1A1* (grey bars) expressed by the skin fibroblasts cultured from each participant (SB-1 to SB-7). (C) The ratio of the relative mRNA abundance of the two genes, *COL1A1*/*COL5A1* (black bars). 147

Figure 3.18 Relative mRNA abundance of (A) *COL5A1* (clear bars) and (B) *COL1A1* expressed (grey bars) in skin fibroblasts pooled with respect to their genotype at the rs12722 locus. (C) The ratio of the relative mRNA abundance of the two genes, *COL1A1*/*COL5A1* (black bars). 148

Figure 3.19 Putative RBPs binding sites (underlined) within the 57bp deletion region (flanking RE sites annotated and sequence underlined) together with the sequence of the RNA biotinylated probes (mapped in grey highlight) used in this experiment for the (A) C-allelic form and (B) the T-allelic form. 152

Figure 3.20 Blot of a non-radioactive RNA EMSA where 13 and 26μg of the cytoplasmic protein extract reacted with 0.125μM of the C- or T-allelic probes spanning the functional 57bp deletion region as described in the results section. 155

Figure 3.21 Blot of a non-radioactive RNA EMSA where (A) 12 and 24 μg of the nuclear protein extract reacted with 0.125μM of the biotinylated C- or T-allelic probes spanning the functional 57bp deletion region (B) Both 12μg of the nuclear and 13μg cytoplasmic protein extract respectively were used with the biotinylated probes in a replicated experiment. 156

Figure 4.1 The most stable predicted secondary structures of a region belonging to the C (left panel) and T (right panel) functional forms of the *COL5A1* 3'-UTR modified from Abrahams *et al*., 2013. 164

Figure 4.2 Schematic diagram showing the three polymorphisms present in the 57 bp deleted region. The polymorphisms rs71698207, rs71746744 and rs11103544 are underlined and identified below the sequence. 168

Figure 4.3 A schematic diagram representing the proposed mechanism of how *COL5A1* may increase (right panel) and decrease (left panel) the risk of chronic Achilles tendinopathy. 173

Figure 4.4 A schematic diagram representing the proposed mechanism by which polymorphisms within *COL5A1* and *COL11A1* potentially affect fibrillogenesis (Hay *et al.*, 2013). 176

LIST OF TABLES FROM THIS DISSERTATION

Table 1.1 Genetic polymorphisms shown to be independently associated with chronic Achilles tendinopathy. 17

Table 1.2 Classification of collagen types modified from Mienaltowski and Birk (2014)....... 20

Table 1.3 Isoforms of type V and XI collagen as well as the genes encoding the α chains. 24

Table 2.1 The *COL5A1* 3'-UTR single nucleotide polymorphisms rs12722 (C/T) and rs13946 (C/T) genotypes of the individual Achilles tendinopathy (TEN) and asymptomatic control (CON) participants included in this study, as well as the consensus genotypes. 65

Table 2.2 General characteristics of the individual Achilles tendinopathy (TEN) and asymptomatic control (CON) participants included in this study, as well as the average or relative values. 66

Table 2.3. Characteristics of the Achilles tendinopathy (TEN) participants as well as features of their pathology. 67

Table 2.4 Primers designed for the site-directed mutagenesis of rs12722 (C/T) within the *COL5A1* 3'-UTR (figure 2.2) of the parental constructs (pGL3-COL5A1-3'UTR) cloned from the TEN and CON participants. 74

Table 2.5 Primers designed for the site-directed mutagenesis of rs71746744 (-/AGGG) within the *COL5A1* 3'-UTR (figure 2.2) of the parental constructs (pGL3-COL5A1-3'UTR) cloned from the TEN and CON participants. 75

Table 2.6 Primers designed to sequence the cloned *COL5A1* 3'-UTR. 82

Table 2.7 Primers designed to sequence the entire plasmids containing the site-directed mutants within rs12772 and rs71746744 of the cloned *COL5A1* 3'-UTR. 83

Table 2.8 Primers used for quantifying relative mRNA abundance in the decay assay. 87

Table 2.9 Custom-made allele specific probes and flanking primer sets primers used in custom designed Fluorescence-based Taqman® PCR assays. 95

Table 2.10 Characteristics and genotype of participants who donated a skin biopsy. 101

Table 2.11 Primers extracted from Luna *et al.* (2011) to quantify relative expression of the *COL5A1* and *COL1A1* genes. 106

Table 3.1 Summary of the identified sequence differences within the *COL5A1* 3'-UTR region (2.5kb) of each clone.. 112

Table 3.2 Descriptive characteristics of the combined South African and Australian Achilles tendinopathic (TEN) and control (CON) participants 123

Table 3.3 Collection of symptoms present in the South African Achilles tendinopathic (TEN) participants where the list includes (i) tenderness to palpation, (ii) early morning stiffness, (iii) a history of swelling, (iv) early morning pain, (v) palpable thickening and (vi) a positive ''shift'' test. 126

Table 3.4 Physiological characteristics of the three rs71746744 (-/AGGG) genotype groups of the combined Australian and South African participants 128

Table 3.5 Physiological characteristics of the three rs16399 (ATCT/-) genotype groups of the combined Australian and South African participants 128

Table 3.6 Physiological characteristics of the three rs1134170 (A/T) genotype groups of the combined Australian and South African participants 129

Table 3.7 The linkage disequilibrium (LD) between the three newly annotated variants rs71746744, rs16399 and rs1134170 as well as with respect to the reported rs12722 variant within the combined Australian and South African participants.. 130

Table 3.8 Genotype frequency distributions the *COL5A1* 3'-untranslated region (UTR) polymorphisms, rs71746744 (-/AGGG), rs16399 (ATCT/-) and rs1134170 (A/T), in control (CON) and chronic Achilles tendinopathy (TEN) groups of South African (SA) and Australian (AUS) cohorts. 131

Table 3.9 Genotype of participants that donated a skin biopsy. 143

Table 3.10 Summary of the size, putative binding site sequence, localisation and function of key RBPs binding putatively within the 57bp deletion region of the 3'-UTR of *COL5A1* gene 153

ABSTRACT

Introduction. *COL5A1* encodes the α1 chain of type V collagen, a minor fibrillar collagen that is an important regulator of collagen fibril assembly. A polymorphism (rs12722, C/T) within the 3'-untranslated region (UTR) of *COL5A1* is associated with chronic Achilles tendinopathy (TEN) and other soft tissue injuries as well as exercise-related phenotypes. These phenotypes are directly or indirectly associated with the mechanical properties of musculoskeletal soft tissue. It has therefore been hypothesised that variants in the *COL5A1* gene, specifically the 3'-UTR, regulate synthesis of the α1(V) chain and type V collagen production. Type V collagen levels in turn regulate fibril architecture and structure and, thereby, mechanical properties of musculoskeletal soft tissues. Although the 3'-UTR of many eukaryotic genes have been shown to play an important regulatory role, the function of the *COL5A1* 3'-UTR is currently unknown.

Aim. The primary aim of this thesis was therefore to determine whether the *COL5A1* 3'-UTR was functional and to identify functional differences between the *COL5A1* 3'-UTR cloned from participants with TEN and healthy asymptomatic control individuals. The secondary aim was to start mapping the functional regions within the 3'-UTR, focusing on regions which are potentially responsible for contributing to the tendinopathic phenotype.

Methods. The entire 2.5 kb *COL5A1* 3'-UTR from 5 TEN, with a 'severe' TEN phenotype, and 5 control participants were cloned downstream of a luciferase reporter gene. These constructs were sequenced and co-transfected into HT1080

cells together with an internal control. Luciferase activity was measured as an indication of mRNA stability. To map the region(s) involved with the tendinopathic phenotype, deletion constructs were generated and assayed. In addition, 160 Caucasian TEN and 342 control participants were genotyped for newly annotated *COL5A1* 3'-UTR markers rs71746744, rs16399 and rs1134170 within the functional region identified. Site-directed mutagenesis of rs71746744 and rs12722 was used to further map the area potentially responsible for the tendinopathic phenotype. Moreover, skin biopsies were taken from participants having a known genotype at rs12722. Primary fibroblast cell lines were cultured, RNA extracted and cDNA obtained for Q-RT-PCR in order to quantify *COL5A1*and *COL1A1* expression levels. Lastly, in preliminary RNA electrophoretic mobility shift assays (EMSAs), biotinylated C- and T-allelic RNA probes for a specific 57bp functional region were incubated with either nuclear or cytoplasmic protein extracts from HT1080 cells in order to investigate putative distinguishing RNA:RBP complex formation.

Results. When all the cloned *COL5A1* 3'-UTRs were sequenced, two major forms were identified. One form (termed the C-allelic form) was predominantly identified in the controls and the other (the T-allelic form) in tendinopathic participants. The C- and T-allelic forms were distinguished by the previously associated rs12722 (C/T) and an additional six polymorphisms which spanned the 3'-UTR namely, rs13946 (C/T), rs3196378 (C/A), rs71746744 (-/AGGG), rs16399 (ATCT/-), rs1134170 (A/T) and rs3128575 (T/C). Although the 3'-UTR of both the C- and T-allelic forms were functional, the luciferase activity of the C-allelic form (69.0 ± 22.0%, N=24) was significantly lower ($p<0.001$) than the T-allelic form (90.6 ± 13.7%, N=30). This suggests an overall increase in mRNA stability for the T-allelic form of *COL5A1* 3'-

UTR. Furthermore, an overall increased relative mRNA expression of both *COL5A1* (1.58 ± 0.89 arbitrary units, $p<0.001$, N=14) and *COL1A1* (1.88 ± 1.10 arbitrary units, $p=0.0015$, N=13) were observed in primary skin fibroblasts donated from participants having a *COL5A1* TT genotype at rs12722 compared with those with the CC genotype (*COL5A1*: 0.57 ± 0.19, N=12 and *COL1A1*: 0.67 ± 0.29, N=12). Furthermore, using deletion constructs, additional elements which may regulate *COL5A1* mRNA stability were identified. Specifically the analysis of constructs containing a 57bp deletion within the 3'-UTR, where rs71746744 (-/AGGG) was removed, revealed that there was no significant difference ($p = 0.440$) in the relative luciferase activity of the C-allele (144.2 ± 28.1%, N=9) and the T-allele (156.5 ± 42.5%, N=18). This suggests that this region was, at least in part, responsible for the observed difference in mRNA stability between the C- and T-allelic forms. In addition to rs71746744 (-/AGGG), this region contained several putative RNA protein binding sites and the downstream flanking sequence contained rs16399 (ATCT/-) and rs1134170 (A/T). All three genetic markers were independently associated with chronic Achilles tendinopathy and in linkage disequilibrium ($D' \geq 0.871$). Specifically the AGGG/AGGG ($p=0.008$, odds ratio (OR)=2.0, 95% confidence interval (CI)=1.2-3.3), -/- ($p=0.015$, OR=1.7, 95% CI=1.1- 2.7) and TT ($p=0.011$, OR=1.8, 95% CI=1.2-2.9) genotypes of rs71746744, rs16399 and rs1134170, respectively, were associated with increased risk of TEN. Although both the rs71746744 and rs12722 variants appeared to be functional, they did not independently contribute to the tendinopathic phenotype. However, a putative RNA-binding protein complex was identified within the 57bp deleted region of the C-allelic, but not the T-allelic, form.

Conclusion. Two major functional forms of the *COL5A1* 3'-UTR were identified. The T-allelic form, which has a high mRNA stability, was predominately associated with the tendinopathic individuals, while the C-allelic form was associated with the asymptomatic controls. A region responsible for differences in the mRNA stability between the C- and T-allelic forms was mapped to a 57bp region. Three polymorphisms within and downstream of this region were shown to be independently associated with chronic Achilles tendinopathy. A putative RNA-binding protein complex was identified within the 57bp region of the C-allelic form. These novel results have important implications for our understanding of the molecular basis of musculoskeletal soft tissue injuries and other exercise-related phenotypes.

CHAPTER 1 LITERATURE REVIEW

1.1 Introduction

The *COL5A1* gene encodes for the α1 chain of type V collagen, a quantitatively minor fibrillar collagen which plays an important role in fibril assembly and lateral growth in skin, tendons, ligaments and other connective tissues (Birk et al., 1990;Wenstrup et al., 2004). Type V collagen also plays an important role in tendon development (Roulet et al., 2007;Wenstrup et al., 2004). Although very little is currently known about its regulation, *COL5A1* is essential for life and its altered expression has a critical impact on both embryonic development and disease (Lincoln et al., 2006;Roulet et al., 2007;Segev et al., 2006). This is further demonstrated by observations that *Col5a1* null mice die *in utero* (Wenstrup et al., 2004;Wenstrup et al., 2006), since it is involved in the formation of a number of vital tissues and organs including, but not limited to, the heart, skin, bones, vertebral column, tendons, ligaments and the cornea of the eye (reviewed by Gelse et al., 2003;Kadler et al., 2007).

Moreover, rare mutations within *COL5A1* cause the classic form (types 1 and 2) of the Ehlers-Danlos syndrome (EDS) which is a severe heritable autosomal dominant connective tissue disorder characterised, amongst other clinical symptoms, by joint hypermobility and major skin and musculoskeletal soft tissue defects (Beighton et al., 1998;Malfait et al., 2010;Wenstrup and De Paepe, 1993). In a murine model, haploinsufficiency of *Col5a1* causes features of the severe EDS phenotype such as the skin and musculoskeletal soft tissue defects (Wenstrup et al., 2000;Wenstrup et

al., 2006). Given that having only one functional copy of this gene can drastically alter the phenotype of these connective tissues, it has been postulated that polymorphisms within regulatory regions, such as the 3'-untranslated region (UTR) of *COL5A1* may alter type V collagen production, fibrillogenesis and the mechanical properties of musculoskeletal soft tissues such as tendons (Collins and Posthumus, 2011). This “normal” biological variation within the mechanical properties of these tissues have been proposed, at least in part, to explain the inter-individual variation in susceptibility to occupational and physical activity associated musculoskeletal soft tissues injuries and other related multifactorial phenotypes (Collins and Posthumus, 2011).

Indeed, a common polymorphism (rs12722, C/T) within the 3'-UTR of *COL5A1* have been associated with several musculoskeletal soft tissue injuries (Abrahams et al., 2013;Burger et al., 2014;Mokone et al., 2006;Posthumus et al., 2009;September et al., 2009), exercise-associated muscle cramps (O'Connell et al., 2013), range of motion measurements (Abrahams et al., 2014;Brown et al., 2011;Collins et al., 2009) and endurance running performance (Abrahams et al., 2014;Brown et al., 2011;Posthumus et al., 2010). The 3’-UTR of eukaryotic genes is described as a zone rich in diverse translational control mechanisms (reviewed by Mazumder et al., 2003) and is an area where polymorphisms can influence the mRNA stability and gene expression (Akai et al., 1999;Conne et al., 2000;Wang et al., 2008). Mazumder *et al.* (2003) described this region of a gene as holding promises for the future identification of potential sites involved in human diseases.

This thesis will examine the proposed function of polymorphisms within the *COL5A1* 3'-UTR on the aetiology of musculoskeletal soft tissue injuries, specifically chronic Achilles tendon overuse injuries (tendinopathy), and other exercise-associated phenotypes at the molecular, as well as cellular levels. This review will focus on key areas relevant to this thesis.

1.2 Musculoskeletal Soft Tissue Injuries

Musculoskeletal soft tissues (tendons, ligaments and skeletal muscle) are common structures that can be injured as a result of participating in competitive and recreational physically activities (reviewed by Collins and Raleigh, 2009;Kannus, 1997), as well as, activities in the workplace; especially where cumulative and repetitive motion is involved (reviewed by Barr and Barbe, 2002;Yassi, 2000). Chronic overuse injuries can occur in the tendons of both the lower and upper limbs (tendinopathy) and are characterized by a painful tendon less capable of sustaining repeated tensile load, a decreased exercise tolerance of the tissue and a reduction in its function (Cook and Purdam, 2009). As much as 50% of all sporting injuries involve tendons and a fifth of these affects the Achilles tendon (reviewed by Jarvinen et al., 2005;Kujala et al., 2005;Rees et al., 2006). Among running-related musculoskeletal injuries (reviewed by Lopes et al., 2012), Achilles tendinopathy is one the three most frequently reported injuries (incidence ranging from 9.1% to 10.9%; prevalence ranging from 6.2% to 9.5%), together with medial tibial stress syndrome, commonly known as shin splints, and plantar fasciosis (jogger's heel). Together with patellofemoral syndrome (runner's knee), Achilles tendinopathy is also reported to be the main musculoskeletal injury among ultra-marathon runners

(prevalence ranging from 2.0% to 18.5%). The prevalence of Achilles tendon injuries appears to be increasing globally, with nearly 825 000 runners in the United States alone estimated to present with Achilles tendon injuries during 2004 (Albers and Hoke, 2003).

In addition, chronic Achilles tendinopathy is not limited to physically active populations but also experienced by sedentary individuals (Jarvinen et al., 2005;Rees et al., 2006). This observation is conceivably linked to the higher prevalence of overweight, obesity and metabolic diseases (World Health Organisation, 2014) within the sedentary population. Body composition and several systemic diseases are risk factors for Achilles tendinopathy (Gaida et al., 2009;Gaida et al., 2010;Magnan et al., 2014). These may operate by causing structural changes in the tendon matrix (Geyer, 2005;Wearing et al., 2013). In addition, although not ascertained in Achilles tendinopathy, changes in tendon properties as a result of periods of inactivity (Kubo et al., 2004), e.g. stress-shielded during periods of bed rest or spaceflight, suggest that sedentary behaviour may compromise the inherent ability of the tendon to sustain a normal or unaccustomed load.

Acute and chronic musculoskeletal soft tissue injuries often result in reduced functional capacity in the workplace and significant loss of athletic performance amongst affected individuals (Ljungqvist et al., 2008). This is because these injuries (1) are often resistant to treatment, (2) recur and/or (3) are risk factors for other orthopedic injuries (reviewed by Collins and Raleigh, 2009;Rees et al., 2006). Although increasing population-wide participation in physical activity is a major health priority to reduce the increasing burden of non-communicable diseases on

society (Bangsbo et al., 2014;Matheson et al., 2013), musculoskeletal soft tissue injuries are a recognised adverse effect of being physically active (Abate et al., 2009;Verhagen, 2012). Prevention of these injuries therefore appears to be a reasonable long-term strategy which will require an understanding of the molecular mechanisms underpinning them. Although they have been described at the clinical level, the biological and molecular mechanisms causing these injuries are however currently poorly understood. An understanding of the basic structure of the Achilles tendon is necessary to understand molecular mechanisms and will be summarised in the following section, followed by a brief summary of the common pathologies that affect this tendon and surrounding tissues.

1.3 Structure of the Achilles tendon

The Achilles (or calcaneal) tendon is the largest and strongest tendon within the human body (Jarvinen et al., 2001). It attaches the gastrocnemius as well as the soleus muscles, commonly known as the calf muscles, to the calcaneus bone posteriorly (the heel bone) (Figure 1.1). The Achilles tendon is responsible for transmitting forces generated in the calf muscles to the calcaneus permitting walking, running and jumping, as well as stabilizing the ankle-joint and maintaining balance (reviewed by Wang, 2006). The point of union of the tendon with the bone is referred to as the enthesis or osteotendinous junction and its attachment to the muscle is termed the myotendinous junction as described by Kannus (2000). This tendon is able to resist extreme forces since it can experience up to 12.5 times of the body weight during bouts of running (reviewed by Wang, 2006). Unlike other tendons, the Achilles tendon does not have a true synovial sheath but it is covered by a paratenon

(reviewed by Asplund and Best, 2013). The paratenon is essentially a loose fibrillar tissue, consisting mainly of types I and III collagen, as well as elastin fibres, which acts as an elastic sleeve and serves the same role as a tendon sheath to reduce friction and facilitate tendon gliding.

Lastly, there is the presence of a protective fibrous sac, the bursa, cushioning in areas where bony protrusions might compress or damage the sliding tendon. The bursa has a synovial membrane that secretes synovial fluid (reviewed by Kannus, 2000). The Achilles tendon has a subtendinous, also known as the retrocalcaneal bursa, and a subcutaneous bursa. Its subtendinous bursa is located between the Achilles tendon and the calcaneus (heel bone), whereas the subcutaneous bursa is positioned between the skin and the posterior aspect of the Achilles tendon, by the enthesis (Figure 1.1) (reviewed by Asplund and Best, 2013). The Achilles tendon is innervated by nerves attached to the nearby muscles or cutaneous nerves such as the sural nerve (Bjur et al., 2005). The insertions and mid-section of the Achilles tendon receive blood through the well vascularised paratenon and the attached muscles. It is unclear whether the blood supply is uniformly distributed throughout the tendon; however the mid-section is believed to have the least vascularisation (reviewed by Theobald et al., 2005).

Tendon tissue is a dynamic structure which is capable of responding to mechanical forces by changing its metabolism as well as its structural and mechanical properties. The structure and physiology of tendons have been extensively reviewed (Kannus, 2000;O'Brien, 1997;Silver et al., 2003;Wang, 2006) and will be briefly described in the following paragraphs.

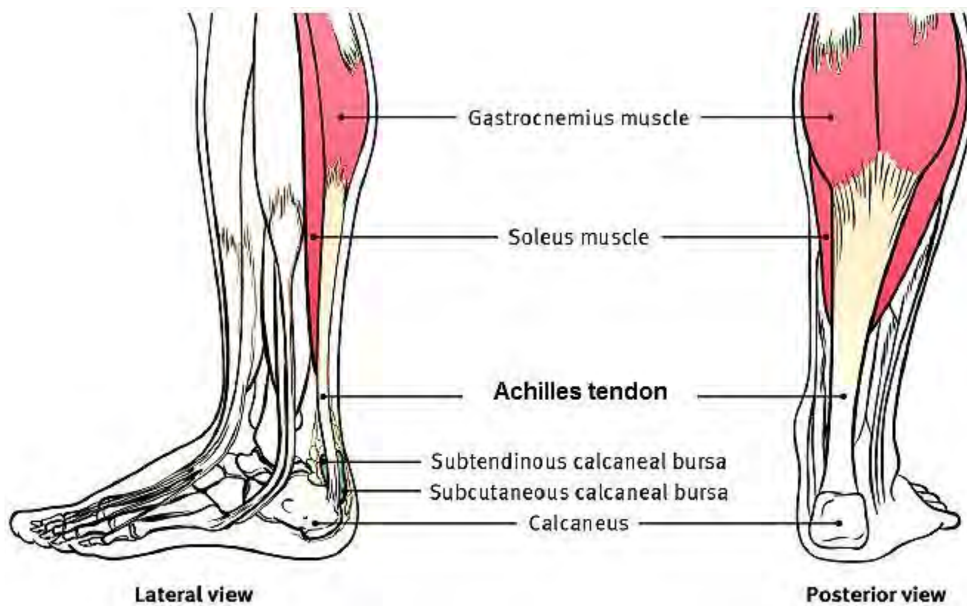


Figure 1.1 Lateral and posterior view of the lower leg, indicating the position of the Achilles tendon (bold) modified from Asplund and Best (2013). The Achilles tendon attaches to the gastrocnemius and the soleus muscles (the calf muscles) and the calcaneus bone (the heel bone). The position of the bursae associated with the tendon is also indicated.

The midsubstance of the tendon is composed predominately of collagen and elastin fibres which make up 65-80% and about 2% of the dry mass of the tendon respectively as well as non-fibre forming proteins embedded in a gel-like medium known as the ground substance (Kannus, 2000). The ground substance is a complex mixture of glycoproteins, proteoglycans, water and mineral salts (O'Brien, 1997). The collagen fibres are made up of collagen fibrils, which are the smallest functional unit of the tendon (O'Brien, 1997), and they in turn consist predominately of type I collagen and other quantitatively minor collagen types (Figure 1.2). The collagen and elastin fibres, as well as the ground substance and non-fibre forming proteins, are produced by tenoblasts (younger cells, highly metabolically active) and tenocytes (mature cells, metabolically active) (Kannus, 2000). In essence, these cells, which lie

between the fibre bundles, are elongated fibroblasts and fibrocytes responsible for the secretion and maintenance of all the components of the extracellular matrix (ECM).

Each collagen fibre is surrounded by a thin connective tissue sheath called the endotenon. The individual fibres form primary (subfascicle), secondary (fascicle) and tertiary bundles each delineated by the endotenon. The tendon, which is surrounded by a layer of connective tissue called the epitenon, is a collection of tertiary bundles (Figure 1.2). The endotenon is a thin reticular network of connective tissue while the epitenon is comprised of relatively dense fibrillar network of collagen. Both of these tissues provide a vascular, lymphatic and nerve supply to the tendon (Wang, 2006).

There are a number of acute or chronic pathologies that can affect the insertional or non-insertional (mid-substance or mid-portion) regions of the Achilles tendon as well as its surrounding tissues and these have been extensively reviewed (Asplund and Best, 2013;Cook et al., 2002;Jarvinen et al., 2005;Magnan et al., 2014). In summary, acute injuries include partial and complete ruptures of the tendon (Jarvinen et al., 2005;Kannus and Natri, 1997) while inflammation of the surrounding tissues includes paratenonitis and bursitis (Jarvinen et al., 2001;Kannus, 1997;Schepsis et al., 2002;Weinfeld, 2014). Chronic degeneration of the tendon (Kannus, 1997) is referred to as tendinosis and can be detected by ultrasonography or magnetic resonance imaging (Asplund and Best, 2013). Because of the different pathologies within and around the tendon, a clinical diagnosis of pain, swelling, stiffness and/or weakness of the Achilles tendon is generally referred to as Achilles tendinopathy (Jarvinen et al., 2001). For the purpose of this dissertation, the term Achilles

tendinopathy (TEN) has been used and implies a painful chronic condition of the mid-portion of the Achilles tendon.

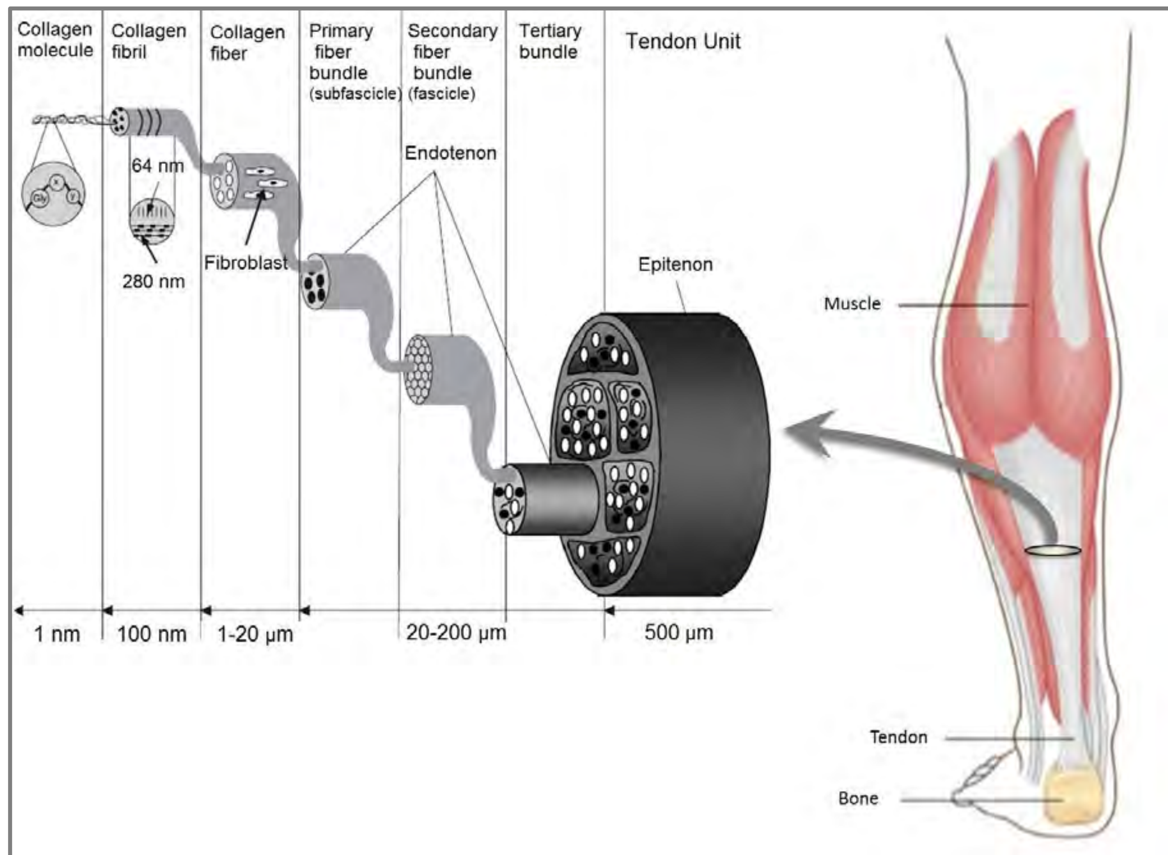


Figure 1.2 A schematic diagram representing the hierarchical structure of the tendon unit (modified from Wang, 2006). The collagen fibrils, made up of highly organised and packed tropocollagen molecules (about 280 nm in size with a *D*-period of about 67 nm – see section 1.5), are assembled into collagen fibres. These in turn will be grouped into primary fibre bundles called subfascicles, and those into secondary fibre bundles called fascicles. Collections of these fascicles will make tertiary bundles a component of the tendon unit. The subfascicle, fascicle and tertiary bundle will each be enveloped by the endotenon, and the final tendon unit wrapped with the epitenon. The relative size of each component of this organisation is indicated in the diagram.

1.4 Risk factors and Models of Pathology for Tendinopathy

The exact biological mechanisms that cause tendinopathies are poorly understood but the following theories have been proposed: the mechanical theory, vascular theory, the 'neurogenic' hypothesis, the iceberg theory and the pathology continuum theory. While the mechanical theory argues that repetitive mechanical loading within the normal thresholds in the tendon may cause fatigue, damage to the fibres and ultimately tendon failure (Arnoczky et al., 2007;Rees et al., 2006;Wren et al., 2003), the vascular theory puts forward that a compromised vascular supply to the tendon impacts negatively on the metabolism of the tendon cells (Fenwick et al., 2002;Langberg et al., 1998). The 'neurogenic' hypothesis asserts that several neurotransmitters and mediators in the tendon play an important role in the aetiology of tendinopathy (Alfredson et al., 1999;Alfredson et al., 2000;Alfredson et al., 2001;Andersson et al., 2008;Bjur et al., 2005).

In the iceberg theory, overload and prolonged repetitive strain of the tendon induce the production of pro-inflammatory mediators such as cytokines, prostaglandins, nitric oxide, growth factors and neuropeptides which in turn induce apoptosis, pain mediators and matrix remodelling agents resulting in a weakened tendon with an increased risk of injuries (Fredberg and Stengaard-Pedersen, 2008). Finally, the continuum model is a unifying model divided into three main states: (1) the 'reactive tendinopathy' which is a temporary adaptive and localised thickening of the tendon, (2) a 'tendon dysrepair' state reminiscent of a failed-healing process and (3) the 'degenerative tendinopathy', incorporating clinical, histological and imaging information at each of these states (Cook and Purdam, 2009).

Several intrinsic and extrinsic (environmental) risk factors have been shown to, or suggested to, associate with tendon injuries which are therefore considered to be multi-factorial (Kannus, 1997;Wren et al., 2001). Although the level of evidence and certainty varies, extrinsic risk factors for Achilles tendinopathy include physical activity, occupation, training errors, training or playing surfaces, footwear, environmental conditions, smoking, medication use (e.g. corticosteroids, fluoroquinolone antibiotics), nutrition and psychological factors (Figure 1.3). Intrinsic risk factors include gender, age, body mass and size, a previous Achilles tendon injury, reduced blood supply to the tendon, tendon temperature, lower limb malalignment, joint laxity and specific systemic diseases (reviewed by Jarvinen et al., 2001;Magnan et al., 2014;Mokone et al., 2005) (Figure 1.3). Genetic polymorphisms have recently been identified as additional intrinsic risk factors for chronic Achilles tendinopathy and other musculoskeletal soft tissue injuries (reviewed by Ribbans and Collins, 2013).

The intrinsic risk factors are believed to predispose individuals to injuries (Figure 1.3) (Bahr and Holme, 2003;Bahr and Krosshaug, 2005;Meeuwisse, 1994). It is argued that the susceptibility to an injury within the predisposed individual is determined by exposure to extrinsic risk factors. Through an inciting event, the susceptible individual may then sustain an injury or become symptomatic. Furthermore, many of the intrinsic risk factors are also in their own right multifactorial phenotypes, which are determined by an interaction of both genetic and non-genetic/environmental factors. As mentioned above and summarized in table 1.1, variants within several genes have been reported to be associated with chronic Achilles tendinopathy.

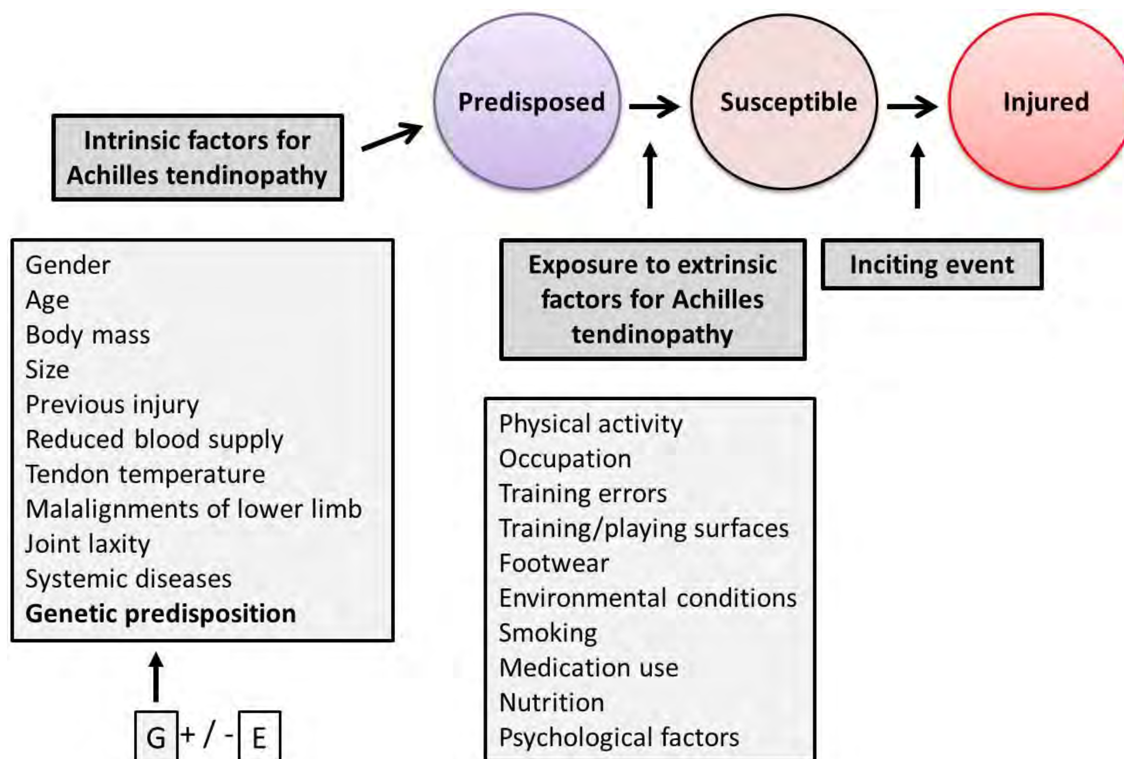


Figure 1.3 Model representing the role of the intrinsic factors in predisposing the individual to become susceptible to Achilles tendinopathy, modified from Meeuwisse (1994) and Ribbans and Collins (2014). Once an individual is predisposed by intrinsic factors, the susceptibility to Achilles tendinopathy is determined by exposure to extrinsic risk factors. Upon experiencing an inciting event, the susceptible individual may then sustain an injury. The intrinsic factors involved are multifactorial phenotypes determined by an interaction of both genetic (G, grey box) with or without environmental factors (E, clear box).

Polymorphisms within *COL5A1* (Abrahams et al., 2013;Mokone et al., 2006;September et al., 2009), *MIR608* (Abrahams et al., 2013), *TNC* (Mokone et al., 2005;Saunders et al., 2013), *FBN2* (El Khoury et al., In Press), *MMP3* (Raleigh et al., 2008), *TIMP2* (El Khoury et al., 2013), *GDF5* (Posthumus et al., 2010) and *CASP8* (Nell et al., 2012) have been reported to independently associate with chronic Achilles tendinopathy. The association of *COL5A1* and *MIR608* variants with chronic Achilles tendinopathy, other injuries and exercise-associated phenotypes will be

reviewed in detail in Section 1.7. The *TNC* gene, which is located 19.6Mbp upstream of *COL5A1* on chromosome 9, encodes for the hexameric glycoprotein tenascin C. Tenascin C plays an important morphoregulatory role during development, tissue remodeling, and in disease by regulating cell adhesion as well as signaling (Jones and Jones, 2000). The protein is predominately expressed in regions responsible for transmitting high levels of mechanical force in healthy adult tendons (Chiquet and Fambrough, 1984;Chiquet et al., 2003;Jarvinen et al., 1999) and the expression of the *TNC* gene in tendons is regulated in a dose-dependent manner by mechanical loading (Jarvinen et al., 1999;Jarvinen et al., 2003). The protein is also expressed around the collagen fibres and cells of the Achilles tendon (Jarvinen et al., 2003). Functionally distinct isoforms of the protein are expressed in degenerate tendons (Riley et al., 1996). Increases in tenascin C expression have also been reported in chronic Achilles tendinopathy biopsy samples (Alfredson et al., 2003). A GT dinucleotide tandem repeat polymorphism within intron 17 of *TNC* (Mokone et al., 2005) as well as two SNPs, rs1330363 (A/G) and rs2104772 (A/T) (Saunders et al., 2013), are also associated with chronic Achilles tendinopathy. The single nucleotide base change at rs2104772 (T/A) causes an amino acid change (Leu1677Ile) and has been associated with both adult asthma (Matsuda et al., 2005) and childhood rhinoconjunctivitis (Orsmark-Pietras et al., 2008), while rs1330363 (A/G), found in intron 15, has only been associated with childhood rhinoconjunctivitis (Orsmark-Pietras et al., 2008).

Similarly, the GG genotype of rs331079 (C/G) within *FBN2* was recently associated with increased risk of chronic Achilles tendinopathy (El Khoury et al., In Press). *FBN2* encodes for fibrillin-2, a member of a multi-domain protein family. These

proteins are large, cysteine-rich glycoproteins present in the extracellular matrix of extensible tissues (Baldwin et al., 2013;Sakai et al., 1986). Fibrillin-2 may be important for the formation of microfibril filaments, providing strength and flexibility to tendons (Zhang et al., 1994). Mutations within the *FBN2* gene is linked to congenital contractural arachnodactyly, a rare congenital connective tissue disorder characterised by, among other complications, the manifestation of contractures and abnormally slender digits (Baldwin et al., 2013;Sabatier et al., 2014).

During normal tissue homeostasis, matrix remodelling and in disease the turnover of the ECM is regulated by a number of protein families, such as the matrix metalloproteases (MMPs), tissue inhibitors of MMPs (TIMPs), growth factors and other signaling molecules (Ribbans and Collins, 2013). The GG, CC and AA genotypes of the *MMP3* rs679620 (A/G), rs591058 (C/T) and rs650108 (A/G) SNPs, respectively, are all independently associated with increased risk of chronic Achilles tendinopathy (Raleigh et al., 2008). The *MMP3* gene encodes for matrix metalloprotease-3 which can catalytically degrade multiple ECM substrates including: types II, IV, V, IX, X collagens, laminin, fibronectin, proteoglycan, decorin, aggrecan as well as activate other MMPs via propeptide removal (Birkedal-Hansen et al., 1993;Somerville et al., 2003;Visse and Nagase, 2003). The SNP rs679620 (A/G) which is in strong linkage disequilibrium with the other two SNPs, is a non-synonymous polymorphism (Raleigh et al., 2008). Specifically, the A/G changes a positively charged lysine residue to a negatively charged glutamate residue, which may have implications for the enzyme formed.

With respect to the tissue inhibitor of MMP-2 (TIMP2), the CC and CT genotypes of the rs4789932 (C/T) promoter SNP was under- and over-represented with chronic Achilles tendinopathy respectively (El Khoury et al., 2013). Deregulated expression of *TIMP2* RNA (Jones et al., 2006;Karousou et al., 2008) and serum *TIMP2* protein (Pasternak et al., 2010) have been implicated in Achilles tendon rupture samples. Furthermore, the TT genotype of the *GDF5* rs143383 (T/C) variant was significantly associated with increased risk of Achilles tendinopathy in two populations (Posthumus et al., 2010). This functional SNP, where the T allele was correlated with reduced expression (Dodd et al., 2013;Miyamoto et al., 2007;Southam et al., 2007;Syddall et al., 2013) is found in the 5'-UTR of the gene and has also been associated with lumbar disc degeneration (Williams et al., 2011), osteoarthritis (Chapman et al., 2008), congenital hip dysplasia (Dai et al., 2008;Rouault et al., 2010) as well as a range of phenotypic measures such as height, hip axis length and fracture risk (Vaes et al., 2009). GDF5 is a member of the TGF-β superfamily which plays an essential role in tissue growth and differentiation as well as homeostasis (Benke et al., 2013). The Achilles tendons from mutant *GDF-5* deficient mice are weaker and contained 40% less collagen (Mikic et al., 2001).

Caspases are another class of matrix remodelling cysteine proteases (Nunez et al., 1998). Caspase-8 (CASP8) is one of the key executioners of apoptosis (programmed cell death) which is responsible for the proteolytic cleavage of key substrates and other caspases in the pathway (Cohen, 1997). The expression of *CASP8* was found to be elevated in tendinopathy (Millar et al., 2008). An independent association of *CASP8* rs3834129 and rs1045485 with chronic Achilles tendinopathy was recently reported (Nell et al., 2012). Specifically the del/del

genotype of rs3834129 (insertion/deletion of CTTACT) increases the risk of Achilles tendinopathy by a factor of 1.6 while the C allele at the rs1045485 locus reduces the odds of TEN by a factor of 0.6 (Nell et al., 2012).

Lastly, gene-gene interactions have also been reported to modulate the risk for Achilles tendinopathy. For example, interactions between polymorphisms within the 3'-UTR of *COL5A1* and the type XI collagen genes, namely *COL11A1* and *COL11A2*, modulate risk (Hay et al., 2013). In addition, within the *COL27A1-TNC* gene cluster, a genomic region containing both genes has been implicated in influencing risk of chronic Achilles tendinopathy through haplotype analysis (Saunders et al., 2013). Lastly, variants within the interleukin genes, *IL-6*, *IL-1RN* and *IL-1β*, which are inflammatory and matrix remodeling signaling molecules, also interact to modify the risk for chronic Achilles tendinopathy (September et al., 2011).

Table 1.1 Genetic polymorphisms shown to be independently associated with chronic Achilles tendinopathy.

Protein	Gene	Polymorphism(s)	Risk Genotype or Allele	Reference
Type V collagen	*COL5A1*	rs12722 (C/T)	T allele	Mokone *et al.*, 2006 September *et al.*, 2009
Tenascin C	*TNC*	GT dinucleotide tandem repeats in intron 17 (z11654) rs1330363 (A/G) rs2104772 (A/T)	12 or 14 repeats G allele A allele	Mokone *et al.*, 2005 Saunders *et al.*, 2014
Non-protein MicroRNA Hsa-miR-608	*MIR608*	rs4919510 (C/G)	CC genotype	Abrahams *et al.*, 2013
Matrix metalloprotease-3	*MMP3*	rs679620 (A/G), rs591058 (C/T), rs650108 (A/G)	GG CC AA genotypes	Raleigh *et al.*, 2009
Tissue inhibitors of MMP-2	*TIMP2*	rs4789932 (C/T)	CT genotype	El Khoury *et al.*, 2013
Growth differentiation factor-5	*GDF5*	rs143383 (T/C)	TT genotype	Posthumus *et al.*, 2010
Caspase-8	*CASP8*	rs1045485 (T/C) rs3834129 (del/ins)*	T allele del/del genotype	Nell *et al.*, 2012
Fibrillin-2	*FBN2*	rs331079 (C/G)	GG allele	El Khoury *et al.*, In Press

* ‘del’ for deletion and ‘ins’ for insertion of CTTACT

1.5 The Collagen Family and Fibril Forming Collagens

The collagens are a family of twenty-eight structurally and functionally diverse proteins. Many are important structural components of, or are associated with, the building blocks of tendons, ligaments and other musculoskeletal tissues as reviewed by Gelse et al., 2003. All collagens consist of three polypeptides, known as the α chains, wound in a characteristic uninterrupted or interrupted right handed triple helix (reviewed by Kadler et al., 2007;Mienaltowski and Birk, 2014). Repeating Gly-X-Y triplet amino acids within each of the three α-chains, in which every third residue is a glycine and the X and Y amino acids are often proline and hydroxyproline respectively, are responsible for forming the collagen triple helix (reviewed by Gelse et al., 2003;Kadler et al., 2007).

The collagen family is commonly divided into the fibril-forming (fibrillar) and non-fibril-forming (non-fibrillar) collagens (Table 1.2) as reviewed in Gelse et al. (2003). The "classical" fibril-forming collagens are further sub-divided into the major (types I, II and III) and minor (types V and XI) fibril-forming collagens. Although not usually included as classical fibrillar collagens, the more recently identified non-abundant types XXIV and XXVII collagen are nevertheless also fibrillar collagens (Boot-Handford et al., 2003;Koch et al., 2003;Matsuo et al., 2006;Matsuo et al., 2008). The non-fibril-forming collagens, which are beyond the scope of this review, are divided into several groups including fibril-associated collagens (FACIT), FACIT-like collagens, network-forming collagens, transmembrane collagens, multiplexins and other molecules bearing collagenous domains (reviewed by Gelse et al., 2003;Kadler et al., 2007;Mienaltowski and Birk, 2014).

Type V collagen, together within types I, II, III and XI, are the “classical” members of the fibril-forming collagen family, which all have a long uninterrupted triple-helical (COL) domain (Figure 1.4). As the designation indicates, the fibrillar collagen family participates in the formation of highly ordered fibrils (reviewed by Kadler et al., 2007;Mienaltowski and Birk, 2014). Type I, II and III are the most abundant collagen species within the fibril structure, where type I collagen is the predominant macromolecule of connective tissues, including the tendon mid-substance (reviewed by Kannus, 2000). Type II collagen is the predominant protein found within cartilage, but it is also an important structural component of fibrocartilaginous zones of connective tissue and the osteotendinous junction (Fukuta et al., 1998). Type III collagen is also a major fibrillar collagen which forms heterotypic fibrils together with type I collagen, and is important in healing processes and during fibrillogenesis (Banos et al., 2008;Liu et al., 1995). Tissues with elastic properties, such as skin and arteries, also have higher type III collagen content than less elastic tissue as reviewed in Gelse et al. (2003). Type III collagen also form fine supporting meshwork of reticular fibres, also known as the reticulin (Ushiki, 2002).

Table 1.2 Classification of collagen types modified from Mienaltowski and Birk (2014)

Classification	Collagen types	Supramolecular structure
Fibril-forming collagen		
Classical		
Major	I, II, III	Striated fibrils
Minor	V, XI	Striated fibrils, retain N-terminal regulatory domains
Non-classical	XXIV, XXVII	Unknown
Non-fibril forming collagen		
FACIT[a] collagens	IX, XII, XIV	Associated with fibrils, other interactions
FACIT-like collagens	XVI, XIX, XXI, XXII	Interfacial regions, basement membrane zones
Network-forming collagens		
Basement membrane	IV	Chicken wire network with lateral association
Beaded filament-forming	VI	Beaded filaments, networks
Anchoring fibrils	VII	Laterally associated anti-parallel dimers
Hexagonal networks	VIII, X	Hexagonal lattices
Transmembrane collagens	XIII, XVII, XXIII, XXV Gliomedins, ectodysplasin	Transmembrane and shed soluble ecto-domains
Multiplexin collagens (Endostatins)	XV, XVIII	Basement membranes, cleaved C-terminal domains influence angiogenesis
Other molecules with collagenous domains	XXVI, XXVIII Acetylcholinesterase, adiponectin, C1q, collectins, surfactant protein, others	Collagenous domains in primarily non-collagenous molecules

[a] Fibril-associated collagen with interrupted triple helix

Types II and III collagen are both homotrimers consisting of α1(II) and α1(III) collagen chains respectively while type I is a heterotrimer consisting of two α1(I) and one α2(I) chains (reviewed by Gelse et al., 2003). Homotrimers of the α1(I) chains can also be formed (Han et al., 2010). Each collagen pro-α chain namely the pro-α1(I), -α2(I), -α1(II) and -α1(III) is encoded by the *COL1A1*, *COL1A2*, *COL2A1* and *COL3A1* genes respectively.

The biosynthesis of the fibril-forming collagens is reviewed by Gelse et al. (2003) and Mienaltowski and Birk (2014) as a complex multi-step process from gene transcription in the nucleus to aggregation into supramolecular structures. Transcription of the collagen genes may be regulated by the cell type, numerous growth factors and cytokines (Rossert et al., 2000). Many collagen genes have been shown to undergo alternative splicing (Matsuo et al., 2006;McAlinden et al., 2007;Mitchell et al., 2012) and their pre-mRNA undergoes capping at the 5'-end and polyadenylation at the 3'-end.

The mature mRNA is transported out of the nucleus and translated into precursor α chains (reviewed by Mienaltowski and Birk, 2014). In the rough endoplasmic reticulum, a signal peptidase catalyzes the removal of the signal peptide and three pro-α chains trimerise to form the procollagen molecule (Figure 1.4). The fibrillar procollagen molecules, which contain amino-terminal globular (N), triple helical (COL) and carboxy-terminal globular domains (C), are post-translationally modified. Specifically, certain proline and lysine residues are hydroxylated to produce hydroxyproline and hydroxylysine respectively. In addition, procollagen is also glucosylated and galactosylated at specific residues. Furthermore, the globular

structure of the N-terminal and C-terminal domains are stabilised and the procollagens secreted into the extracellular space through the Golgi apparatus. The precursor molecules will be further modified extracellularly, e.g. both globular domains of the procollagen molecule are cleaved outside of the cell and the collagen triple helix is incorporated into the collagen fibril as reviewed by Mienaltowski and Birk (2014).

In contrast, the amino globular domains of the quantitatively minor types V and XI collagens, which form heterotypic fibrils with the major fibrillar collagens, are not removed (reviewed by Fichard et al., 1994). The uncleaved N-terminal extension of type V collagen contributes mainly to a large flexible globular domain (Birk et al., 1990). When incorporated in the forming fibril, this globular domain can fit in the staggered array of molecules protruding through the fibril gaps and the fibril surface, thus thought to regulate fibril growth by sterically preventing the addition of new molecules (Linsenmayer et al., 1993).

Due to structural homologies and similarities between types V and XI collagen, it has been proposed that type XI collagen regulates fibrillogenesis of the type II collagen containing cartilage fibrils in a similar way as previously described for type V collagen (reviewed by Fichard et al., 1994). Although traditionally classified as two distinct proteins, types V and XI collagen are now considered a single collagen type with multiple tissue-specific isoforms or sub-types, including a hybrid isoform (Table 1.3) (Wenstrup et al., 2011). Both collagen types V and XI play a crucial and similar, if not coordinated, role in the regulation of collagen fibrillogenesis during early tendon development (Birk et al., 1990; Wenstrup et al., 2011). For instance, reduced tendon

fibril number and integrity was observed in a mouse model with a compound heterozygous genotype for *Col5a1* and *Col11a1* (Wenstrup et al.,2011). When compared to the mice with a reduction of type V collagen alone, the phenotype was more severe with a reduction in type XI collagen but it was worst with the absence of type XI collagen and a reduction in type V collagen. Moreover, in a mouse study, it was observed that both collagens shared a similar pattern of expression. However, while type V collagen continues to be steadily expressed past early tendon development, type XI appears to be transiently expressed during the early developmental stages (Wenstrup et al., 2011).

The newly identified types XXIV and XXVII collagen have shorter interrupted helical regions than the other typical fibril-forming collagens. They resemble the invertebrate collagens and while they are involved in bone development, to date they have not been linked to tendon biology (Boot-Handford et al., 2003;Koch et al., 2003;Matsuo et al., 2006;Matsuo et al., 2008).

Table 1.3 Isoforms of type V and XI collagen as well as the genes encoding the α chains.

Collagen type	Isoform	Genes encoding the α chains and chromosomal location (Chr)
Type V collagen	$\alpha1(V)_2\alpha2(V)$	*COL5A1* (Chr9q34.2-q34.3), *COL5A2* (Chr2q14-q32)
	$\alpha1(V)_3$	*COL5A1*
	α1(V)α2(V)α3(V)	*COL5A1*, *COL5A2*, *COL5A3* (Chr19p13.2)
Type XI collagen	α1(XI)α2(XI)α1(IIB)*	*COL11A1* (Chr1p21), *COL11A2* (Chr6p21.3), *COL2A1* (Chr12q13.11)
Hybrid	$\alpha1(XI)_2\alpha2(V)$	*COL11A1*, *COL5A2*

*α1(IIB) is a post-translationally modified α1(II) chain

Collagen fibrillogenesis is the process by which fibrils are made and assembled to form the highly organised structures such as the tendon tissue (Birk et al., 1990;Fichard et al., 1994). The assembly and deposition of collagen fibrils, mostly heterotypic, with tissue-specific structures and organisation involves a sequence of events that occur in both intracellular and extracellular compartments (reviewed by Mienaltowski and Birk, 2014). The authors further describe the collagen fibrils as composites of different matrix molecules and they note that the control of hetero-polymeric mixing and trimer type stoichiometry begins within the intracellular compartments. It involves regulation by processing enzymes as well as interaction with these in a spatial and temporal manner during secretion (reviewed by Mienaltowski and Birk, 2014).

As previously mentioned the collagen fibril in tendons and other non-cartilaginous connective tissue consists predominately of type I collagen. Fibril assembly begins in the cell-surface crypts, at the surface of the fibroblast surface (Birk and Trelstad, 1986). As the propeptides are removed by procollagen proteinases, some degree of spontaneous self-assembly of the collagen molecules into protofibrils occurs as reviewed in Mienaltowski and Birk (2014). While the cleavage of the C-propeptides is an essential step for regulating fibril formation, the presence of uncleaved N-terminal propeptides projecting outwards between the gaps of the adjacent type I collagen molecules in the fibril, as in the case with type V collagen, may regulate the diameter of the forming fibrils (Wenstrup et al., 2004). Non-reducible covalent crosslinks will further stabilise the fibrils together (Fichard et al., 1994;Gelse et al., 2003).

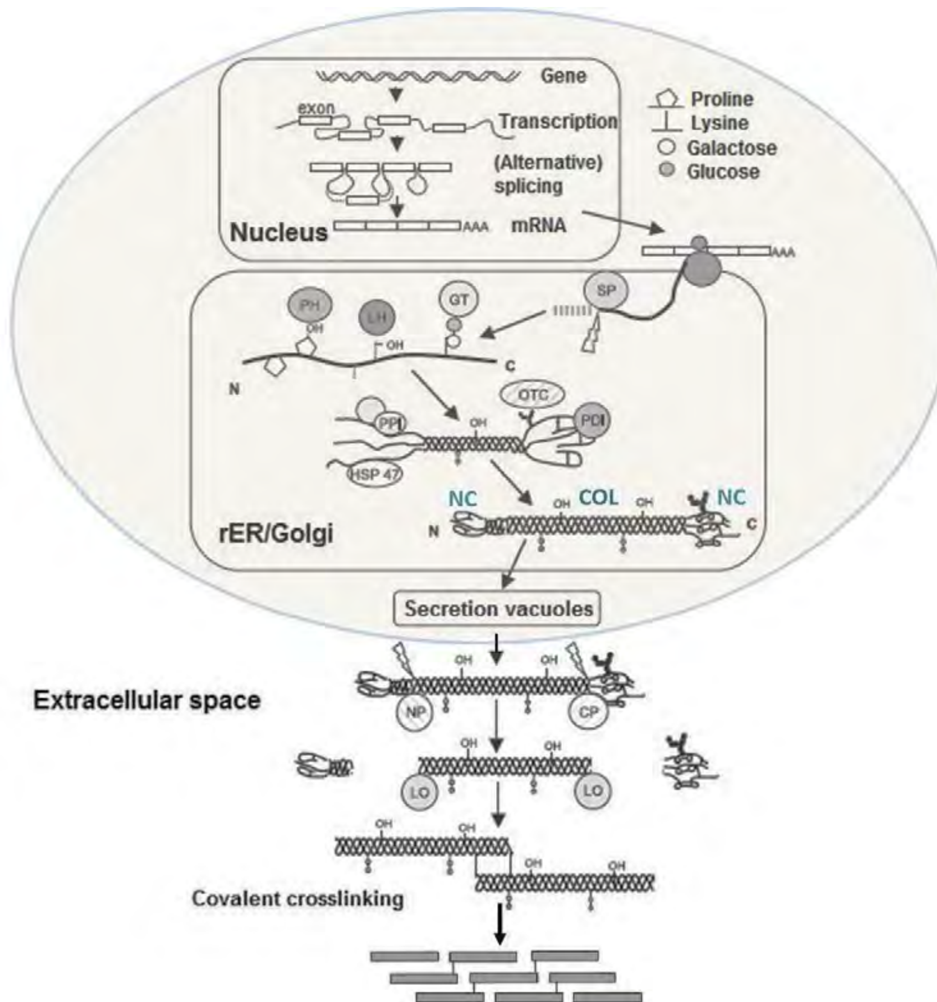


Figure 1.4 Schematic representation of fibril-forming collagen biosynthesis from transcription through mRNA processing, ribosomal protein synthesis followed by post-translational modifications and secretion to fibril formation (modified from Gelse *et al.* 2003). COL: the collagenous domain; NC: the non-collagenous domain; N: amino globular domain and C: carboxyl domain. Typical enzymes involved in the process are indicated. SP: signal peptidase; GT: hydroxylysyl galactosyltransferase and galactosylhydroxylysyl glucosyltransferase; LH: lysyl hydroxylase; PH: prolyl hydroxylase; OTC: oligosaccharyl transferase complex; PDI: protein disulphide isomerase; PPI: peptidyl-prolyl cis-trans-isomerase; NP: procollagen N-proteinase; CP: procollagen C-proteinase; LO: lysyl oxidase; HSP47: heat shock protein 47, colligin1.

1.6 Type V Collagen and the *COL5A1* Gene

The major isoform and most ubiquitous form of type V collagen is a heterotrimer consisting of two α1(V) chains and one α2(V) chain (Roulet et al.,2007; Wenstrup et al.,2011). The pro-α1(V) chain is encoded by the *COL5A1* gene on human chromosome 9q34.2-q34.3. The pro-α2(V) and pro-α3(V) chain are encoded by the *COL5A2* gene (Chromosome 2q14-q32) and *COL5A3* gene (Chromosome 19p13.2) respectively. Except in the case of the type V-type XI collagen isoform (Table 1.3), the α1(V) chain is common to all isoforms of type V collagen and therefore important for the study of expression or regulation of that collagen (Roulet et al., 2007). The α2(V) chains form heterotypic molecules with α1(XI) chains, $\alpha1(XI)_2\alpha2(V)$, but the distribution and functional roles of these less common isoforms are not well described (Birk, 2001;Wenstrup et al., 2011;reviewed in Fichard et al., 1994). Throughout this thesis, for ease of reading, the term 'type V collagen' will refer to its major isoform, $\alpha1(V)_2\alpha2(V)$, unless stated otherwise.

During murine development, *Col5a1* expression, which coincided with the pro-α1(I) chain developmental expression, was detected using *in situ* hybridization in the heart, dorsal aorta wall, branchial arches, mesonephrotic tubules and intestinal mesenchyme (Roulet et al., 2007). Increased expression was detected in discrete regions during later stages of development including tendons, skin, cornea, bones, vertebral column and ligaments, highlighting the important contribution of type V collagen in the development of functional connective tissues (Roulet et al., 2007).

Collagen types I and V interact in heterotypic fibrils within tendons (Figure 1.5) and are co-expressed during embryonic development. The triple helical domain of type V collagen is buried within the heterotypic fibril while, as mentioned previously, the uncleaved N-terminal extension of type V collagen is exposed at the surface. Type V collagen is a dominant regulator of collagen fibrillogenesis (Sun et al., 2011;Wenstrup et al., 2011), specifically, it nucleates fibril assembly, controlling fibril number and initial diameter of the fibril. The protruding N-globular domain is believed to have a regulatory function in this assembly process (reviewed in Birk, 2001).

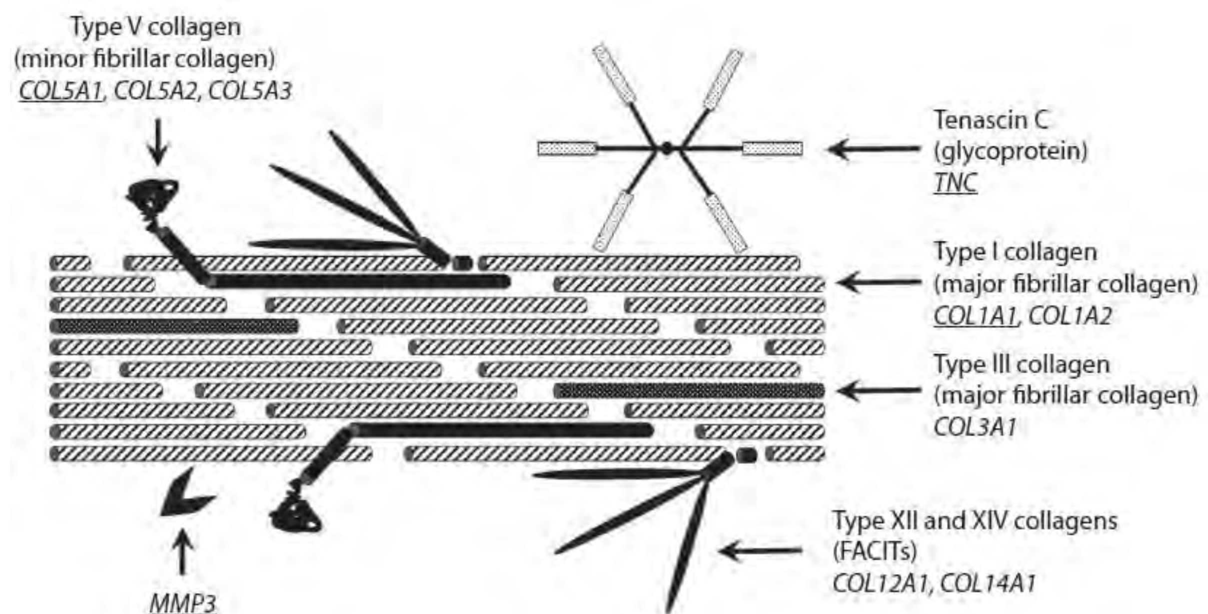


Figure 1.5 A schematic diagram modified from Raleigh and Collins (2009) representing the basic structural unit of the collagen fibril which consists predominately of type I collagen (not drawn to scale). The fibril also contains trace amounts of types III and V collagen. Types I and III collagen are major fibrillar collagens, while type V collagen is a part of the minor fibrillar collagen sub-family. Types XII and XIV collagen are associated with the surface of the fibril and belong to the sub-family of fibril-associated collagens with interrupted triple helices (FACITs). Tenascin C and MMP3 are also shown in this diagram. Other collagen types, proteoglycans and glycoproteins, associated with or structural components of the fibril are not depicted in this diagram. The genes that have been shown to be independently associated with any musculoskeletal soft tissue injuries are underlined.

Rare mutations within *COL5A1* as well as mutations within *COL5A2* cause the classic form (types 1 and 2) of Ehlers-Danlos syndrome (EDS), where EDS is described as a severe heritable connective tissue disorder characterised amongst other clinical symptoms, by fragile hyper-extensible skin, joint hypermobility and musculoskeletal soft tissue defects (Beighton et al., 1998;Malfait et al., 2010). Most patients with classic EDS have been reported to have disease-causing mutations within one copy of the *COL5A1* gene, which result in a loss of function of the mutated gene due to a 50% reduction in the production of type V collagen (haploinsufficiency) (Malfait et al., 2010;Wenstrup et al., 2000). In support of this, although viable, *Col5a1* +/- knockout mice, which only produce half the amount of type V collagen, display similar symptoms observed in patients suffering from the classic form of EDS (Wenstrup et al., 2006). Comparable with EDS patients, the mice had decreased aortic stiffness and tensile strength, as well as hyper-extensible skin with decreased tensile strength (Wenstrup et al., 2006).

Transmission electron micrographs (TEMs) of the dermis of these heterozygous mice displayed two populations of fibrils, where one was circular in appearance and comparable to the wild-type dermis while the other population consisted of large structurally aberrant fibrils (Figure 1.6; Panel A). Lateral and linear growth was also disrupted as seen in the developing dermis at day 10 postnatal (P10) and more severe in the mature dermis (20-weeks) (Wenstrup et al., 2006). Similarly, in the tendon of the heterozygous mice at P30, two populations of fibrils were also observed (Figure 1.6; Panel B) (Wenstrup et al., 2011). TEMs showed an increased amount of small and larger diameter fibrils in the heterozygous mice compared to the wild-type mice. In addition, the cross-sectional outline of the fibrils was irregular in

the heterozygous mice compared to wild-type mice. Investigation of the biomechanical properties of the tendons in the heterozygous mice indicated that they had decreased tissue stiffness, suggesting increased elasticity reminiscent of the joint laxity or hyper-extensibility observed in EDS patients (Wenstrup et al., 2011). Moreover, *Col5a1* null mutant mice died *in utero* at approximately embryonic day 10, indicating the crucial roles that *Col5a1* and type V collagen play in the development of the mesenchymal tissue (Wenstrup et al., 2004). Collagen fibril formation was absent in the type V collagen deficient mice (*Col5a1* -/-).

Type V collagen is critical in the initiation of fibril assembly as displayed by the *in vivo* murine models mentioned above (Wenstrup et al., 2004). In addition, in chick and murine animal models examining the corneal tissues, it was shown that this minor fibrillar collagen regulates fibrillogenesis (Birk et al., 1990;Segev et al., 2006;Sun et al., 2011). A reduced fibril diameter was observed when the relative level of type V collagen was increased in *in vitro* experiments (Birk et al., 1990). In line with these results, larger fibrils were observed in the stroma of conditional type V collagen knockout mice (Figure 1.6; Panel C) (Sun et al., 2011). Recently, genome-wide association studies (GWAS) uncovered a genetic variant in the relative vicinity of the *COL5A1* gene on chromosome 9 which associated with central corneal thickness (Vitart et al., 2010).

Additional studies using *Col5a1* mutant mouse models or targeted deletions of *Col5a1* have shown the analogous functional role played by type V collagen in the fibril assembly of the dermis and tendon tissue (Wenstrup et al., 2006;Wenstrup et al., 2011). Aberrant fibril formation was observed in the corneal stroma, the skin and

the soft tissues when *Col5a1* expression was either increased or decreased in these studies (Wenstrup et al., 2006;Wenstrup et al., 2011). Moreover, in a mouse skin model harbouring targeted deletion of the *Col5a2* gene, the deposition of $\alpha1(V)_3$ homotrimers was favoured over the more abundant isoform of type V collagen, $\alpha1(V)_2\alpha2(V)$ in the skin matrix (Chanut-Delalande et al., 2004). Interestingly, this altered secretion/deposition lead to exclusion of the $\alpha1(V)_3$ homotrimers from incorporation into the heterotypic collagen fibrils and a severely impaired matrix assembly, where disorganised and thin fibrils were observed coupled with a five-fold increase in *Col5a1* expression compared to the WT mice. A transgenic mouse model, with a targeted overexpression of pro-α1(V) chain in the epidermis, displayed thin fibrillar material composed of the collagen V homotrimer underneath the epidermal basement membrane (Bonod-Bidaud et al., 2012). Although not significant, stiffness and rupture stress of the transgenic skin were lower compared with the WT. In addition, the experiments showed that the homotrimer differed in its interaction with ECM molecules, acting as a bridging molecule at the epidermal-dermal interface. The authors proposed that distinct molecular forms of a collagen type can differentially influence the skin ECM architecture as well as its mechanical properties.

Although only trace amounts of type V collagen have been identified in healthy tendons (reviewed in Birk, 2001; Silver et al.,2003), increases in type V collagen content together with a decrease in fibril diameter and changes in the biomechanical properties have been reported with age in the rabbit patellar tendon (Dressler et al., 2002). In addition, increase in types III and V collagen content together with a reduction in the content of type I collagen have been reported in biopsy samples of

degenerative tendons from patients with posterior tibial tendon dysfunction syndrome (Goncalves-Neto et al., 2002;Satomi et al., 2008). Furthermore, in a damage accumulation model using the rat patellar tendon, an upregulation of type I, III and V collagen expression was observed as a result of a cyclic fatigue loading protocol (Andarawis-Puri et al., 2012;Fung et al., 2010).

In another study, the effect of a particular chain of type V collagen on fibrillogenesis was investigated using specific siRNA and primary rat Achilles tenocytes (Lu et al., 2011). This investigation showed that the pro-α1 and pro-α2 chains of type V collagen had, to a certain extent, a different effect on the regulation of the tendon matrix and that an optimal level of the pro-α1 chain in particular was pivotal in regulating fibrillogenesis. Whereas siRNA specific to *Col5a1* decreased the ratios of *Col5a1*, *Col3a1* and decorin gene (*DCN*) to *Col1a1*, siRNA to *Col5a2* increased the ratios of *Col5a1*, *Col3a1* and *DCN* to *Col1a1* (Lu et al., 2011). Importantly, the authors observed that a tissue-engineered tendon treated with *Col5a1* siRNA had abnormal collagen fibril morphology while a tendon treated with *Col5a2* siRNA had collagen fibrils comparable to controls.

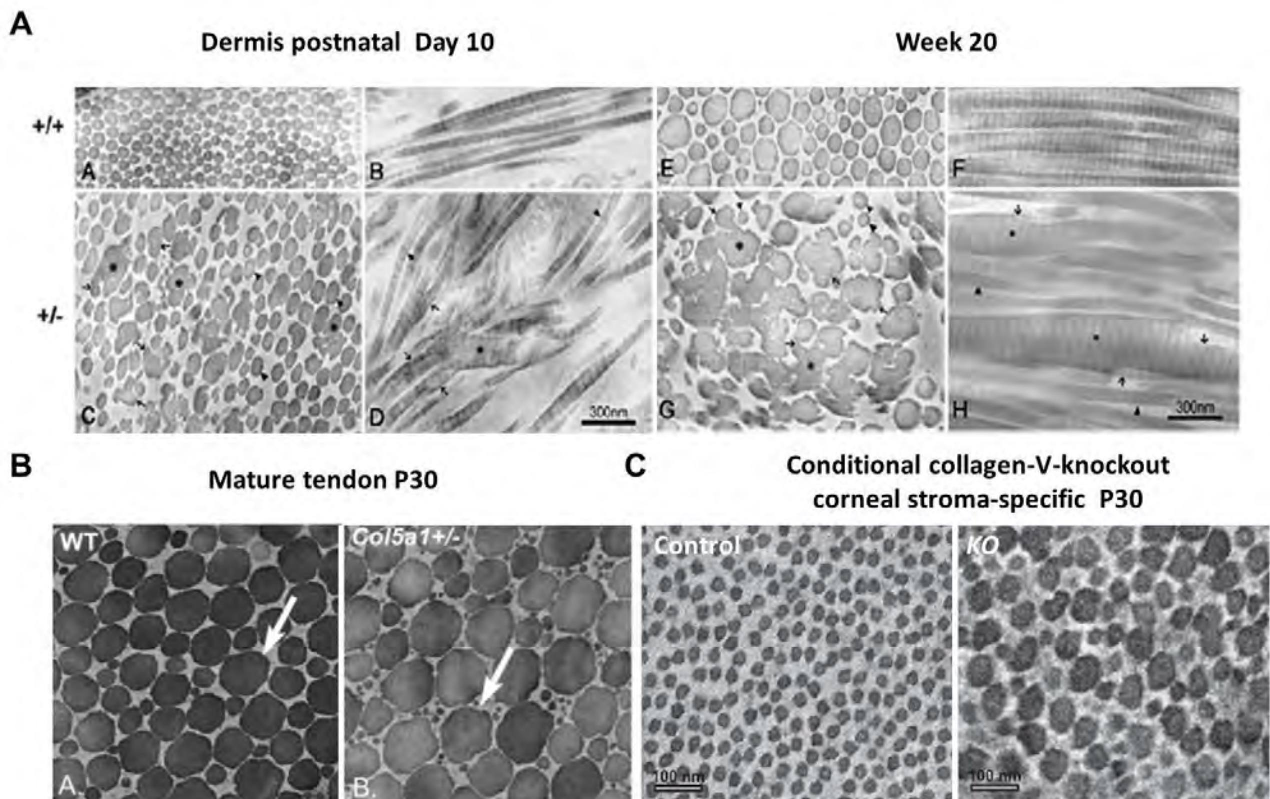


Figure 1.6 Transmission electron micrographs (TEMs) modified from Wenstrup *et al.* 2011, Wenstrup *et al.* 2006 and Sun *et al.* 2011 showing the effect of haploinsufficiency of mice *Col5a1* in (**Panel A**) dermis and (**Panel B**) tendon. **Panel C**, the corneal-stroma-specific *Col5a1*-null mouse model showing the fibril implications in the stroma of the mice. **Panel A** shows the cross-sections (A, C, E, G) next to the longitudinal sections (B, D, F, H) at day 10 and week 20 postnatal. The +/+ or WT indicate presence of both copies of *Col5a1* in the mice while +/- indicate heterozygous mice. The arrowheads show cylindrical fibrils (similar to the WT profile) whereas the asterix indicate aberrant fibrils and the small arrows indicate disruption of the fibril organisation and growth. **Panel B** shows the WT *Col5a1* +/+ mice tendon (left, A) and the haploinsufficient mice, +/- (right, B) at day 30 postnatal. Two populations of fibrils can be observed in the heterozygous mice together with irregular cross-sections (white arrow in B), where a mixture of small and larger fibrils can be seen in comparison to normal circular fibrils of the WT (white arrow in A). **Panel C** shows the increase in size of the fibrils of the stroma when the *Col5a1* gene is knocked out (*KO*) in a conditional knockout mutant, specific to the corneal stroma compared to the control.

1.7 *COL5A1* and *MIR60*8 are associated with musculoskeletal soft tissue injuries and/or other phenotypes

Initial evidence for the involvement of genes in the aetiology of musculoskeletal soft tissue injuries arose from reports of an association between the ABO blood group, specifically blood group O, and increased risk of Achilles and other tendon injuries (reviewed in September *et al.*, 2009). Because the ABO blood group is determined by an enzyme encoded by a single gene located on human chromosome 9q34, a mere 1.4Mb upstream from the *COL5A1* gene, the possible association between *COL5A1* polymorphisms and the risk of Achilles tendon injuries was investigated (Mokone et al., 2006). Of particular interest to the current study are reports that variants within the *COL5A1* 3'-UTR associate independently with several recreational and occupational musculoskeletal soft tissue injuries (Abrahams et al., 2013;Burger et al., 2014;Mokone et al., 2006;Posthumus et al., 2009;September et al., 2009) and other exercise-associated phenotypes (Abrahams et al., 2014;Brown et al., 2011;Brown et al., 2011;Collins et al., 2009;O'Connell et al., 2013;Posthumus et al., 2010).

The terminal exon (exon 66) of the *COL5A1* gene encodes for the last 47 amino acids (141bp) of the carboxy-terminal domain of the α1(V) chain, the stop codon and 2.5kb 3'-UTR (Figure 1.7). Variants, which span the entire 3'-UTR, have previously been investigated for association with chronic Achilles tendinopathy. Although rs13946 (C/T, *Dpn*II RFLP, nucleotide 230 within exon 66) and rs11103544 (C/T, *Mbo*II RFLP, nucleotide 1041) were not independently associated, rs12722 (C/T), also known as the *Bst*UI RFLP, located at nucleotide 414 within exon 66, was

associated with chronic Achilles tendinopathy in a self-reported Caucasian South African and Australian population (Mokone et al., 2006;September et al., 2009) (Figures 1.7 and 1.8). Specifically, individuals with a CC genotype for rs12722 had a significantly decreased risk of developing chronic Achilles tendinopathy compared with those with a T allele (TC or TT genotype) in both the Australian (OR 0.42, 95% CI 0.20 to 0.86, p=0.017) and South African (OR 0.38, 95% CI 0.18 to 0.77, p=0.008) participants. Interestingly, individuals in only the Australian group with the heterozygous TC genotype for rs3196378 (C/A, *Aci*I RFLP, nucleotide 880) had a significantly increased risk of developing chronic Achilles tendinopathy compared with those with homozygous TT or CC genotypes (September et al., 2009).

Polymorphisms rs3196378 and rs11103544 were selected because they are located within putative polymorphic miRNA recognition sequences, 5'-CCACCCCA-3' and 5'-TTTTCTAC-3' (polymorphic nucleotides underlined), respectively, and therefore may potentially be functionally significant (September et al., 2009). Although the miRNA that bind to the putative 5'-TTTTCTAC-3' sequence is currently unknown, the 5'-CCACCCCA-3' sequence binds Hsa-miR-608. Two forms of the 25bp 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 (http://www.ncbi.nlm.nih.gov). The CC genotype of rs4919510 (C/G) within the *MIR608* gene is associated with increased risk of chronic Achilles tendinopathy in the South African and Australian cohorts (Abrahams et al., 2013). The upstream polymorphism rs10858286 (C/T), within intron 65, and the distal 3'-UTR polymorphisms, rs4504708 (G/T) and rs3128575 (A/G) located 2129 and 2135bp downstream of rs12722, respectively, were not significantly associated with Achilles

tendinopathy in the Australian cohort and were therefore not evaluated in the South African cohort (September et al., 2009).

A single study has also investigated a variant, COL5A1_01 G/A, within intron 65 of the equine *COL5A1* gene for an association with superficial digital flexor (SDF) tendinopathy and reported that the AA genotype was significantly associated with SDF tendinopathy (Tully et al., 2014). Other equine *COL5A1* variants which were located upstream of the associated polymorphism were however not independently associated with SDF. Polymorphisms within the 3'-UTR of the equine *COL5A1* gene have not been investigated. Both the investigated human and equine associated variants are nevertheless located within the 3'-end of the gene suggesting that this region of the *COL5A1* gene within both species could contain important elements directly involved in the aetiology of tendinopathy.

A

```
ACCAAGAAAG GCTACCAGAA GACGGTTCTG GAGATCGACA CCCCCAAAGT     50
GGAGCAGGTG CCCATCGTGG ACATCATGTT CAATGACTTC GGTGAAGCGT    100
CACAGAAATT TGGATTTGAA GTGGGGCCGG CTTGCTTCAT GGGCTAGGAG    150
CCGCCGAGCC CGGGCTCCCG AGAGCAACCT CGTGACCTCA GCATGCCATT    200
CGTTCGTGAG TGTCCCGTGC ACGTCCTGAY CCTGGACAGT GAAGGCTTCT    250
                          rs13946 (DpnII RFLP)

CCCTCCCCTC CCACCTGACT TCATCTACGC CTCGGCACCA CGGGGTGTGG    300
GACCCCAGCC CGGAGAGAAC AGAGGGAAGG AGCCGCGCCC CCACCTGGAG    350
CTGAATCACA TGACCTAGCT GCACCCCAGC GCCTGGGCCC GCCCCACGCT    400
CTGTCCACAC CCAYGCGCCC CGGGAGCGGG GCCATGCCTC CAGCCCCCCA    450
       rs12722 (BstUI RFLP)

GCTCGCCYGA CCCATCCTGT TCGTGAATAG GTCTCAGGGG TTGGGGGAGG    500
    rs1134114

GACTGCCAGA TTTGGACACT ATATTTTTTT CTAAATTCAA CTTGAAGATG    550
TGTATTTCCC CTGACCTTCA AAAAATGTTC CAAGGTAAGC CTCGTAAAGG    600
TCATCCCACC ATCACCAAAG CCTCCGTTTT TAACAACCTC CAACACGATC    650
CATTTAGAGG CCAAATGTCA TTCTGCAGGT GCCTTCCCGA TGGATTAAAG    700
GTGCTTATGT TTTTGTGAGT TTTAAGTAAA TATTTGTATT GTATTGTTAT    750
AAATGTTAAG TGTGCCTGGC TTTCAATCAT GCACGGAAAC CCAGTCTCAG    800
TCCCACGGAC AGAATGGGCG AGGCATGGAT TCTGGGTTGC AGTACCGTTC    850
TGATTAGAAA TAGGAAGTCT CCCCACCCCM GCCCTGGCCA AGAACGTGCA    900
                          rs3196378 (AciI RFLP)

ATAAATTGGA AGTTTGCCCC GGGGCAGCAA GAATTTATGC TGCCATTGAA    950
AAGCAGGTAC CAGTGCCCCT TTTCAGACAG TTTTTGATTC GCTCTAGACT   1000
TTTTTTTTTT TTAATAGGGA AAAAATTTGA TAATTTTCTT YTTTCTACAT   1050
                                        rs11103544 (MboII RFLP)

GCACTTAAGA CTAAAACACA GGTTTGGATT AATTTTATTT GCTTCCTTTT   1100
```

B

```
AAAAAGAAAA AAATGGTAAG CAAAAAACCC AAGATAAAGT TTCGAGGACA   2500
TCAGGCCTTT TGAAATACAA TGTCAAATGA CACATTGTAC GKTTTCAAAA   2550
                                             rs4504708

AATCCGCTAG ACATGTCATA AGTTTTAACT GTAATGCCCA GGAAAGGATA   2600
TCTTAAAATA TTCTAAACTT GTGTAACAAA GGAATAATTA ACTGTAAYAG   2650
                                                rs3128575

TTTTTCAATA AATCGAGTTG GGTGTTTCCA CCGT                    2684
```

Figure 1.7 Nucleotide sequence of *COL5A1* exon 66 containing the first 5 sequence variants (**Panel A)** and the latter 2 sequence variants (**Panel B)** (September *et al.*, 2009). The positions of the SNPs are annotated in bold (Y: C or T; M: A or C and K: G or T). Their accession numbers are given with the name of the RFLPs in brackets. The nucleotide position of exon 66 is indicated on the right. The shaded sequences represent the translated region of the *COL5A1* exon 66 (nucleotides 1–147) and exon 1 of an expressed sequence tag (EST) (nucleotides 433–604). The stop codon of exon 66 and the translation start codon of exon 1 are underlined. The dotted line indicates the untranslated region of the EST. The putative recognition sequences of the two microRNAs are double underlined.

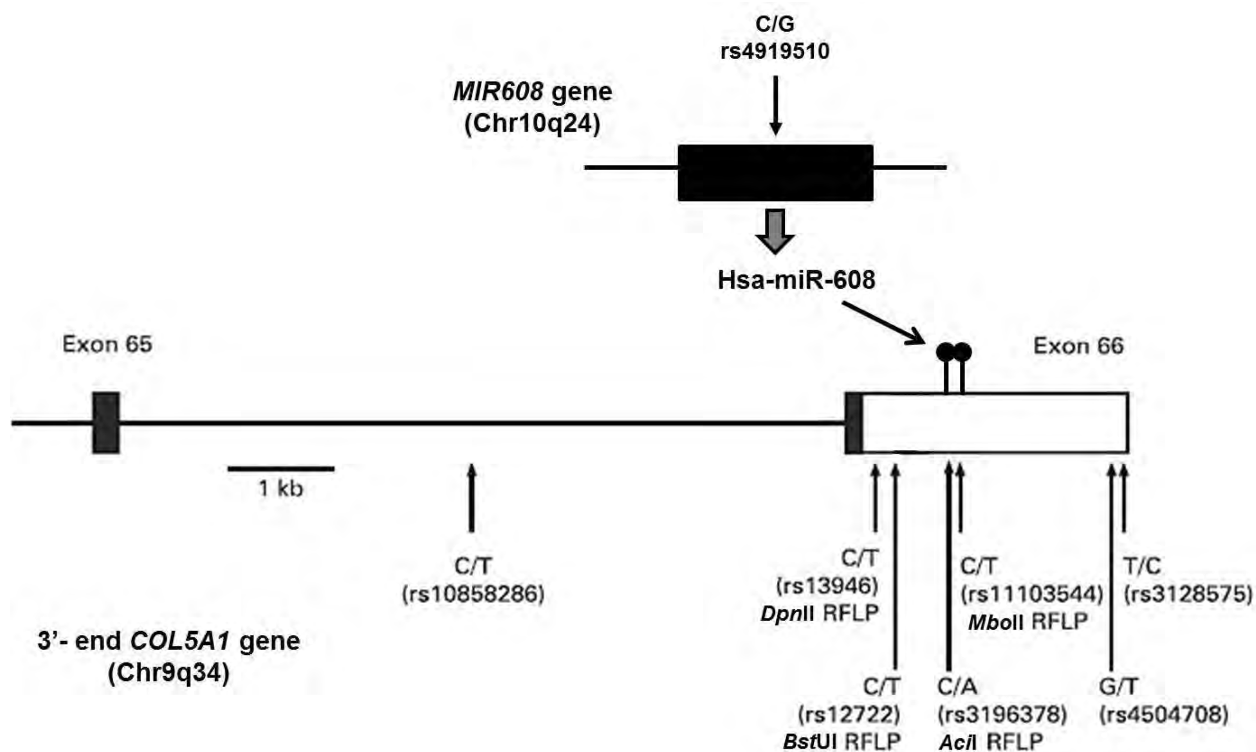


Figure 1.8 Representation of the 3'-UTR of the *COL5A1* gene (9q34) at its terminal end showing exons 65 and 66 with the intronic region in between (intron 65), and the *MIR608* gene(10q24) modified from September *et al.* (2009). The rectangles represent the exons while the straight lines correspond to introns. The shaded rectangles are the translated regions while the clear ones represent the untranslated regions within the exons. The accession number and nucleotide change of each polymorphism described in section 1.7 is annotated on the diagram. The common restriction fragment length polymorphism (RFLP) associated with them is indicated below. SNPs rs3196378 and rs11103544 are each within putative miRNA binding sites.

Exercise-associated muscle cramping (EAMC), which is a common medical condition amongst participants of endurance events, is defined as "painful, spasmodic and involuntary contraction of skeletal muscle that occurs during or immediately after exercise" (Schwellnus et al., 1997). Although a number of hypotheses have been proposed to explain the aetiology of EAMC, pre-race serum creatine kinase activity, a marker of muscle damage, tended to be higher in twenty

athletes who developed EAMC during participation in a 56 km ultra-endurance road race when compared with 29 non-crampers (Schwellnus, 2009;Schwellnus et al., 2011). This suggests that muscle damage, including the connective tissue components of skeletal muscle, may play a role in the aetiology of EAMC. In support of this, the CC genotype of *COL5A1* rs12722 was significantly over-represented in 150 Caucasian male ultra-endurance athletes (ultra-marathon and Ironman triathletes) with no self-reported history of previous (lifelong) EAMC when compared to 116 Caucasian male ultra-endurance athletes with self-reported history of EAMC within the past 12 months (O'Connell et al., 2013).

Altered musculotendinous flexibility, which is defined as ‘‘the ability to move a joint through its complete range of motion (ROM)” (Whaley, 2006), is an intrinsic risk factor for several musculoskeletal soft tissue injuries, including Achilles tendinopathy (Kaufman et al., 1999) and ACL rupture (Uhorchak et al., 2003). Although ROM is a complex phenotype, it has been estimated, using classical twin studies, that up to 70% is determined by inherited components (Maes et al., 1996). As previously mentioned, rare *COL5A1* mutations cause classical forms of EDS, which is characterised by generalised joint hypermobility (Malfait et al., 2010). Furthermore, it has also been proposed that *COL5A1* is associated with benign joint hypermobility syndrome, also an inherited condition (Grahame, 1999;Zweers et al., 2005). Based on these collective reports, it was proposed that variants within the *COL5A1* 3’-UTR were associated with normal joint range of motion (Collins et al., 2009). In support of this hypothesis *COL5A1* rs12722 has been associated with lower limb range of motion, namely sit and reach (SR) measurements (Brown et al., 2011;Brown et al., 2011;Collins et al., 2009). Specifically, SR measurements increased with age in

apparently healthy physically active participants with a CC genotype for rs12722, while there was no correlation in the other two genotypes (Brown et al., 2011). Furthermore, the *COL5A1* rs12722 genotype was associated with SR ROM in older (≥35 years) participants (TT 225 ± 96 mm, TC 245 ± 100 mm, CC 32 ± 108 mm, P=0.017), but not younger participants. Sex and the *COL5A1* genotype accounted for 22.8% of the variance in SR ROM in the older group (Brown et al., 2011). Similarly, a recent study reported that the rs12722 genotype was not associated with sit and reach ROM measurements in a young (25.2 ± 4.0 years) Brazilian cohort (Bertuzzi et al., 2014).

A previous study consisting of a mixed cohort of older participants, with and without a history of chronic Achilles tendinopathy and/or Achilles tendon rupture, reported that the homozygote individuals (TT and CC genotypes) were associated with greater lower limb ROM measurements (Collins et al., 2009). The difference in age and injury profile of the two cohorts could possibly explain this inconsistency between the studies. Similar relationships between the *COL5A1* genotype groups and sit and reach measurements were observed when only the uninjured participants from the mixed cohort were analysed separately (Brown et al., 2011). *COL5A1* rs13946, rs3196378 and rs11103544 were however not associated with the lower limb ROM measurements (Collins et al., 2009). In addition, the CC genotype of rs12722 was also associated with a decreased joint laxity (Bell et al., 2012) in a gender-specific manner consistent with female-specific associations with ACL injury (Posthumus et al., 2009;Posthumus et al., 2010).

Endurance running performance is also a multifactorial phenotype that is strongly associated with range of motion (Craib et al., 1996;Jones, 2002) which has been proposed to affect running economy (Gleim et al., 1990). A relationship between running performance and rs12722 has therefore been investigated (Posthumus et al., 2010). Individuals with the TT genotype had a better running performance when completing the road running stage (42.2km) of the Ironman Triathlon compared to those with the CC and TC genotype (p=0.019, TT= 294.2 ± 52.1 mins, CC= 307.4 ± 48.6 mins, N= 313) (Posthumus et al., 2010). Similarly, athletes with a TT genotype (341 ± 41 min) completed a 56 km road running ultra-marathon significantly faster (p=0.014) than athletes with either a TC or CC genotype (365 ± 39 min). The *COL5A1* genotype and age accounted for 19% of performance variance in these athletes (Brown et al., 2011). Although no direct association between pre-race sit and reach ROM measurements (flexibility) and time to complete the 56-km ultra-marathon race was reported, the *COL5A1* T and C alleles were significantly over-represented in the ''inflexible-fast'' and ''flexible-slow'' athletes respectively (Brown et al., 2011). However, a recent study reported no significant differences between the *COL5A1* genotypes and running economy measurements at 10 km/hr (p=0.232) and 12 km/hr (p=0.259) in a physically active young male Brazilian cohort (Bertuzzi et al., 2014).

Although seemingly unrelated, all four phenotypes (Figure 1.9) are directly or indirectly related to the mechanical properties of musculoskeletal soft tissues (Collins and Posthumus, 2011). Collins and Posthumus (2011) have therefore hypothesised that variants within regulatory regions, such as the 3'-UTR, of the *COL5A1* gene alter collagen fibril architecture and structure and by implications the mechanical

properties of soft tissues. In brief, this hypothesis suggests that since the pro-α1(V) chain is essential to the formation of type V procollagen assembly (Roulet et al., 2007), changes in *COL5A1* expression associated with polymorphisms in regulatory regions may have a direct effect on the amount of type V collagen incorporated into the collagen heterofibril. The authors propose that if this amount is increased it may reduce the average fibril diameter and increase the total numbers of fibres present as indicated in figure 1.10. The smaller, densely packed and organised collagen fibrils may result in reduced ROM and improved endurance running performance but can cause an increased risk for musculoskeletal soft tissue injuries.

Figure 1.9 Diagram illustrating the common association of the T or C allele of the rs12722 in the context of injury risk, EAMC, flexibility and endurance performance as presented in the genetic association studies, where green depicts a positive effect on the phenotype while red depicts a negative effect. All four phenotypes are directly or indirectly related to the mechanical properties of musculoskeletal soft tissues.

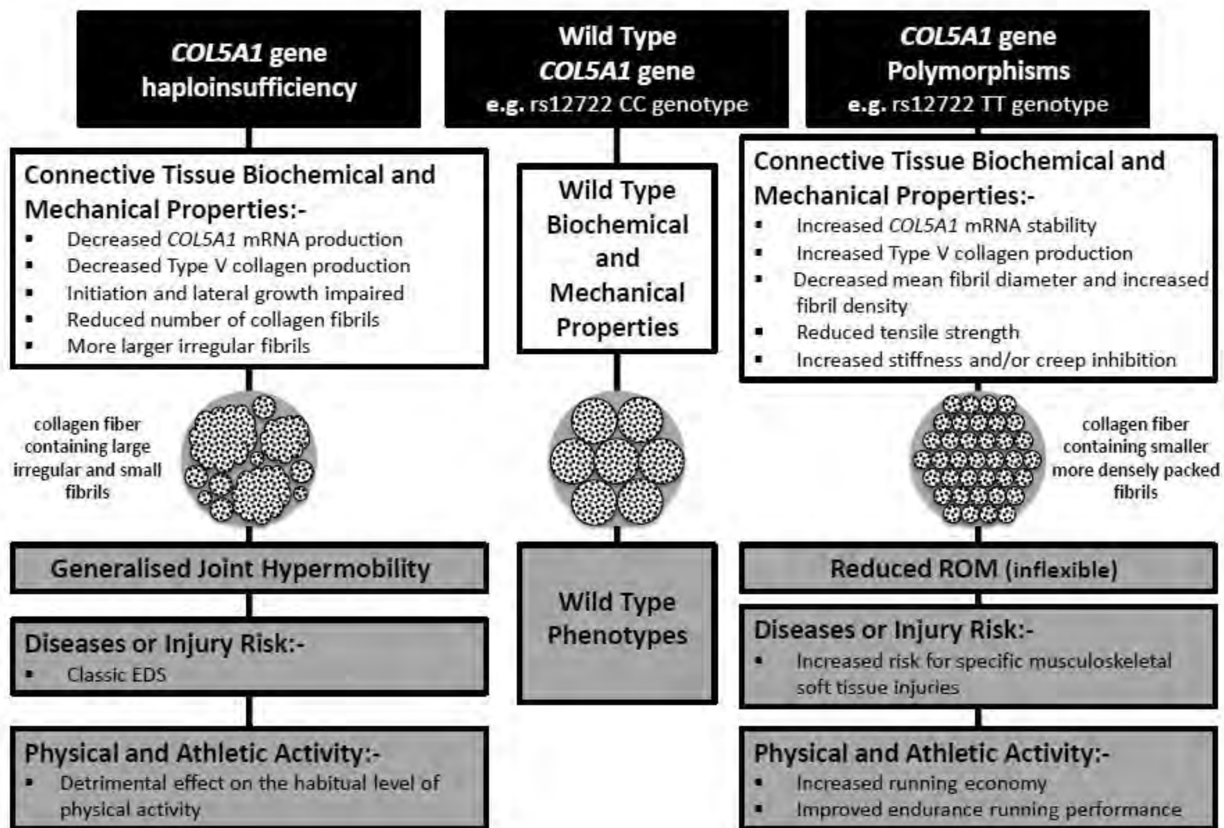


Figure 1.10 Integral model of the relationship between *COL5A1* genotype (black boxes), connective tissue biochemical and mechanical properties (white boxes), flexibility, disease or injury risk as well as physical activity, modified from Collins and Posthumus (2011). **Left panel:** The effects of disease causing *COL5A1* mutations on decreased type V collagen production: abnormal fibrillogenesis (large and small irregular fibrils) and generalized joint hypermobility. These mutations cause Ehlers-Danlos syndrome (EDS) which has a detrimental effect on the habitual level of physical activity. **Middle panel:** The wild type *COL5A1* gene and phenotypes. It is proposed that larger, regularly shaped, stronger and more compliant fibrils are produced from the wild type gene. These may be associated with increased joint range of motion (ROM) and decreased risk for specific musculoskeletal soft tissue injuries but lower endurance running performance. **Right panel:** The effect of functional common polymorphisms within the *COL5A1* gene on proposed increased type V collagen production. Smaller regularly shaped weaker fibrils are produced during fibrillogenesis. These fibrils may have an increased stiffness and/or are creep inhibition, are associated with reduced joint ROM, increased risk for musculoskeletal soft tissue injuries and faster endurance running performance.

Interestingly, despite the tendon dimensions being similar between healthy participants, large and irregular collagen fibrils were observed in the patellar tendons of patients diagnosed with the classic form of Ehlers-Danlos syndrome and having a mutation in the *COL5A1* gene (Nielsen et al., 2014). Reduced (~50%) tendon stiffness and Young's modulus, which relates inversely to flexibility, was reported in these patients compared to healthy participants or individuals suffering from benign joint hypermobility syndrome (BJHS). However, no pathology was linked to the mechanical properties of the BJHS group indicating that the hypermobility phenotype observed in these two conditions may occur through different mechanisms or defects (Nielsen et al., 2014).

To further investigate the hypothesis by Collins and Posthumus (2011), Kubo *et al.* (2013) recently examined the association between the *COL5A1* rs12722 genotypes and knee extensors as well as plantar flexor tendon dimensions and mechanical properties *in vivo*. The authors reported a significantly greater maximal tendon elongation (CC: 24.5 ± 5.4 mm versus TT+TC: 21.1 ± 5.4 mm) and strain (CC: 7.61 ± 1.62 % versus TT+TC: 6.51 ± 1.58 %) as well as lower stiffness (CC: 66.2 ± 19.3 N/mm versus TT+TC: 78.2 ± 18.5 N/mm) for the knee extensors in Japanese male participants with a CC genotype compared to those with a combined TT and CT genotype (Kubo et al., 2013). No differences were however reported for the plantar flexor (Kubo et al., 2013).

In contrast, Foster *et al.* reported no significant differences between the genotypes and patellar tendon dimensions (volume) and functional (elastic modulus) properties assessed *in vivo* during ramp isometric knee extensions within 84 recreationally

active, non-obese Caucasian, males and females between the ages 18 and 39, with no history of injuries to the knee (Foster et al., 2014). Although there were no significant differences between the maximal stiffness of genotype groups (p=0.203), the median maximal stiffness (range) was however lower in individuals with the CC genotype compared to those with a TT genotype (TT: 855.7 N/mm (1,569.1), TC: 707.6 N/mm (1,729.1) and CC: 555.3 N/mm (1,261.5)). This was also reflected in the comparison of the stiffness, elastic modulus and Z-scores (Foster et al., 2014).

Although there have been some contradictory results, there is a growing body of evidence that suggests that polymorphisms within the *COL5A1* 3'-UTR is associated with range of motion measurements, endurance running performance and musculoskeletal soft tissue properties and injuries. Since these studies suggest that the *COL5A1* 3'-UTR might play an important regulatory role in type V collagen production, the main aim of this thesis was therefore to investigate this hypothesis. In preparation for the experimental chapters of this thesis the following section will briefly review some of the main features and functions of eukaryotic 3'-untranslated regions (UTRs), specially focusing on those potentially relevant to the *COL5A1* 3'-UTR.

1.8 Eukaryotic 3'-UTRs

Within eukaryotic genes there are non-coding regions, namely the 5'-UTR, introns and the 3'-UTR, that contain regulatory elements that impact on gene expression (reviewed by Barrett et al., 2012). Interestingly, the average length of the 3'-UTR sequences has increased during evolution, linking their utilisation to organism

complexity (reviewed by Mazumder et al., 2003;Taft et al., 2007). The 3'-UTR is located immediately downstream of the protein coding sequence and determines the fate of mRNA because it is involved in numerous post-transcriptional regulatory processes, including transcript cleavage, mRNA stability, polyadenylation, translation and mRNA localisation as reviewed in Barrett et al.(2012). The 3'-UTR contains several elements, such as polyadenylation (poly(A)) signals, microRNA (miRNA) binding sites and protein binding sites, which are involved in these regulatory processes. In this context, the mRNA secondary structure is also emerging as an important feature.

1.8.1 Polyadenylation signals

The poly(A) signal is a highly conserved hexanucleotide consensus sequence, AAUAAA (reviewed by Edwalds-Gilbert et al., 1997). The formation of the poly(A) tail is the result of the addition of a series of adenosine bases (about 200 residues) to the 3'-end of an RNA molecule by the poly(A) polymerase through the interaction of numerous factors such as the cleavage and polyadenylation specificity factor (CPSF) (reviewed by Glisovic et al., 2008). This addition will occur at a site, about 11 to 23 nucleotides downstream of the hexanucleotide known as the cleavage site, but proximal to a GU- or U-rich region (Figure 1.11A). These structures are essential for splicing, transcription termination as well as for initiation and termination of translation (reviewed extensively by Barrett et al., 2012;Cooke et al., 1999;Edwalds-Gilbert et al., 1997;Hilleren and Parker, 1999;Yeung et al., 1998). The presence of several polyadenylation sites adds to the versatility of expression from a single gene. For example, alternative poly(A) signals may play an important role in the synthesis

of different isoforms by differential regulation of mRNA splicing (reviewed by Barrett et al., 2012;Edwalds-Gilbert et al., 1997). The human pro-α2(I) collagen gene (COL1A2) uses multiple poly(A) signals present in its 3'-UTR which is thought to correlate to tissue specificity (Myers et al., 1983).

Deadenylation is the initial step in the decay of many mRNAs (reviewed in Ross, 1995). Poly(A) tails are therefore structural and functional elements of eukaryotic mRNA that regulate mRNA stability as well as export to the cytoplasm (Wang et al., 1999;Zarudnaya et al., 2003). The poly(A) tail also provides the mRNA with a binding site for a class of regulatory factors called the poly(A) binding proteins (PABP) () (reviewed extensively in Barrett et al., 2012; Brook and Gray, 2012 and Mangus et al., 2003). PABPs play an essential role in the regulation of gene expression by for example affecting the export, stability, decay and translation of mRNAs (Barrett et al., 2012). They regulate the ultimate length of the poly(A) tail and facilitate the formation of the 'closed loop' structure of mRNAs in the cytoplasm in conjunction with the 5'-end of the transcript thereby promoting translation initiation, termination as well as stability of the mRNA (Figure 1.11B) (Brook and Gray, 2012;Mangus et al., 2003).

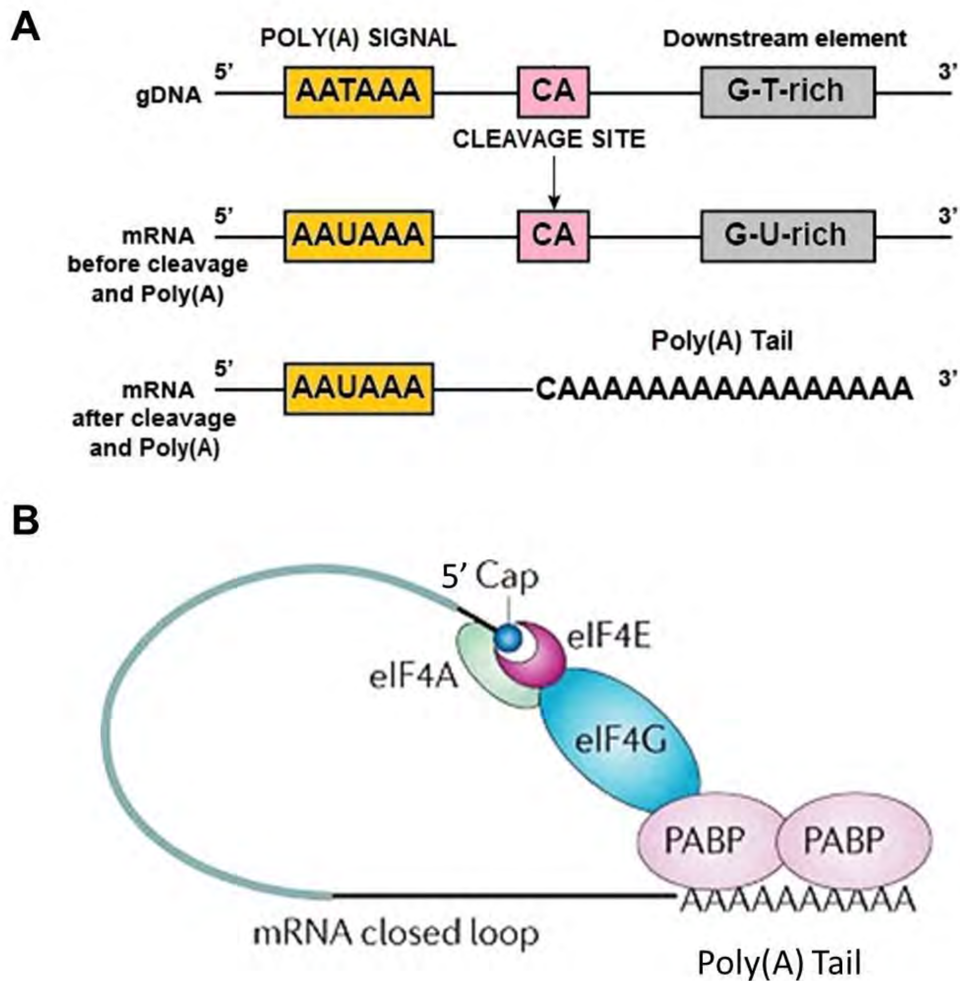


Figure 1.11 Diagram depicting (A) the three sites important in determining where the poly(A) addition will occur in the 3'-UTR, namely the hexanucleotide poly(A) signal, the cleavage site (also known as the poly(A) site), and the GU-rich downstream element; (B) the mRNA 'closed loop' structure formed from cooperation of the 5'-cap and the 3'-UTR through the poly(A) tail modified from Huntzinger and Izaurralde (2011). PABP proteins bind to the poly(A) tail and are essential for the stability of this structure in conjunction with proteins of the ETS transcription factor (eIF) family.

1.8.2 The secondary RNA structure

The secondary structure of mRNA refers to a complex two-dimensional conformation resulting from the folding of the single-stranded RNA molecule (Nebel and Scheid, 2011). It is emerging as an important determinant of translation efficiency. Indeed, polymorphisms that cause alterations in the mRNA secondary structure may disrupt gene expression through affecting accessibility of important *trans*-acting factors to regulatory elements in the mRNA (reviewed by Barrett et al., 2012).

The secondary structure is determined by stem-loops (hairpins), pseudoknots and tetraloops (four-base hairpin loop motifs) (Laing and Schlick, 2010). 'Stems' are formed by complementary canonical Watson and Crick base pairs GC and AU, along with GU wobble base-pairing while the 'hairpin' is a single-stranded region that folds back on itself via regions of complementary base pairs (Laing and Schlick, 2010). The 'internal loop' is the single-stranded region between two stems and a 'junction' is the point of connection between three or more helical stems. Finally, 'pseudoknots' are defined as base pairs that intertwine in single stranded regions (Laing and Schlick, 2010). Dynamic folding is guided by the free energy landscape generated by these base pairing interactions within the primary sequence of the mRNA molecule as well as interactions of the mRNA molecule with the polymerase enzyme, other proteins, solvent and metabolites (Laing and Schlick, 2010;Zhang and Chen, 2002;Zuker et al., 1999). Several computational and experimental approaches are used to predict RNA secondary structures, where prediction software such as Mfold (Zuker, 2003) and Sfold (Ding et al., 2004) have been developed to elucidate RNA secondary architecture *in silico* complementing existing experimental methods such

as RNA crystallography and fluorescence resonance energy transfer (FRET) (Laing and Schlick, 2010).

1.8.3 MicroRNA binding sites

MicroRNAs are a large family of endogenous, evolutionary conserved and non-coding RNAs (19-25 nucleotides long) which are implicated in the regulation of nearly every biological process by modulating gene expression at the post-transcriptional level (reviewed by Ambros, 2004). They may induce gene silencing by binding sites within target genes resulting in mRNA cleavage or repression of protein translation (Lau and Lai, 2005;Matzke and Birchler, 2005). MicroRNAs function through perfect or imperfect base-pairing at the miRNA binding sites mostly located in the 3'-UTR of their target genes (Meola et al., 2009;Rutnam et al., 2013). Importantly, the ‘seed region’, which is composed of 6-7 nucleotides, plays a critical role in the recognition of the target mRNA by displaying a perfect or partial complementarity between the miRNA and its target (Meola et al., 2009;Rutnam et al., 2013). Numerous miRNA target site prediction tools and databases such as miRanda (http://www.microrna.org) or TargetScan (http://targetscan.org) have been developed to determine the presence of a putative binding site within a gene (Bartel, 2009). In addition, it appears that miRNAs may cooperate in complex networks to regulate target genes. For example, one miRNA can act on several genes while many mRNA have ‘seed’ matches allowing the binding of multiple miRNA in their 3’-UTRs (Peter, 2010) (Figure 1.12). In effect, miRNAs are functionally unique and have been quoted as able to operate as a molecular ‘switch’ through precise

regulation of the timing of cellular events as well as via the collaborative inhibition or repression of interconnected genes (Abdellatif, 2012).

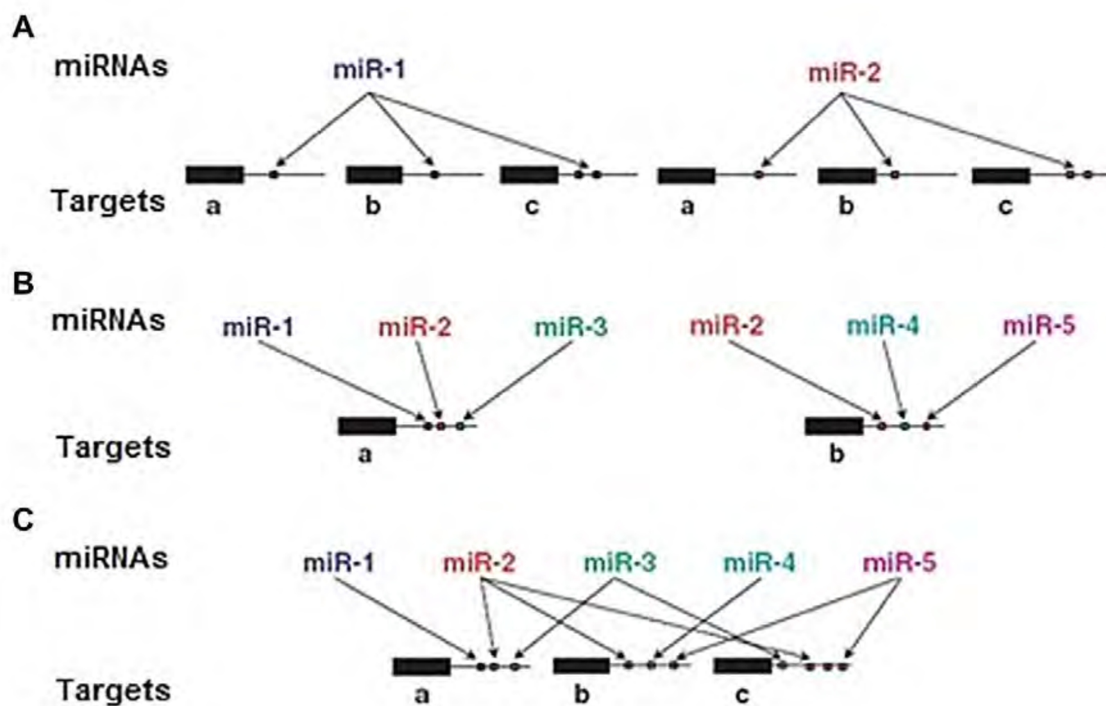


Figure 1.12 Diagram illustrating (A) the ability of miRNAs to target multiple genes; (B) many gene targets have several seed matches for multiple miRNA binding in their 3'-UTRs and (C) a complex network of mutual interactions exist between miRNA and their targets (modified from Peter, 2010).

A miRNA may be transcribed from one miRNA gene code and expressed under the control of its own promoter and regulatory sequences, or may be arranged in clusters of miRNAs and their expression co-regulated in tandem (reviewed by Baskerville and Bartel, 2005). The miRNA can be highly tissue-specific or expressed in a pervasive manner even though at differing levels (Graves and Zeng, 2012). The miRNAs are transcribed in the nucleus as long transcripts, known as primary miRNA transcripts (pri-miRNAs), or may contain multiple miRNAs transcripts (Altuvia et al., 2005). The biogenesis of miRNA and its regulation is beyond the scope of this thesis and has been extensively reviewed (Bartel, 2004;Finnegan and Pasquinelli, 2013;Kim and Nam, 2006). In brief, several proteins are involved in this process, namely members of the Argonaute family, Pol II-dependent transcription and the two RNase III proteins, Drosha and Dicer (Kim and Nam, 2006) (Figure 1.13). The Drosha RNase III protein forms a complex with the double-stranded RNA-binding protein DGCR8 (Gregory et al., 2004). The pri-miRNAs are then processed by this complex into a smaller stem-loop miRNA precursor of about 70 nucleotides called pre-miRNAs (Denli et al., 2004). The pre-miRNAs are then exported across into the cytoplasm by the Exportin-5 complex (Bohnsack et al., 2004;Lund et al., 2004). These pre-miRNAs undergo additional cleaving by Dicer, another RNase III protein, thus producing a 19-25 nucleotide long RNA duplex. These duplexes then enter a complex containing the ribonucleoprotein (RNP) and Argonaute-2 (Ago2) called the miRNA-induced silencing (miRISC) complex (Schwarz and Zamore, 2002;Tang, 2005). Only one strand of the miRNA-duplex termed the mature miRNA is incorporated into the complex, while the other strand, denoted as miRNA*, is removed (Bartel, 2004). Within this complex, miRNAs will bind to their targets and regulate gene expression (Rutnam et al., 2013).

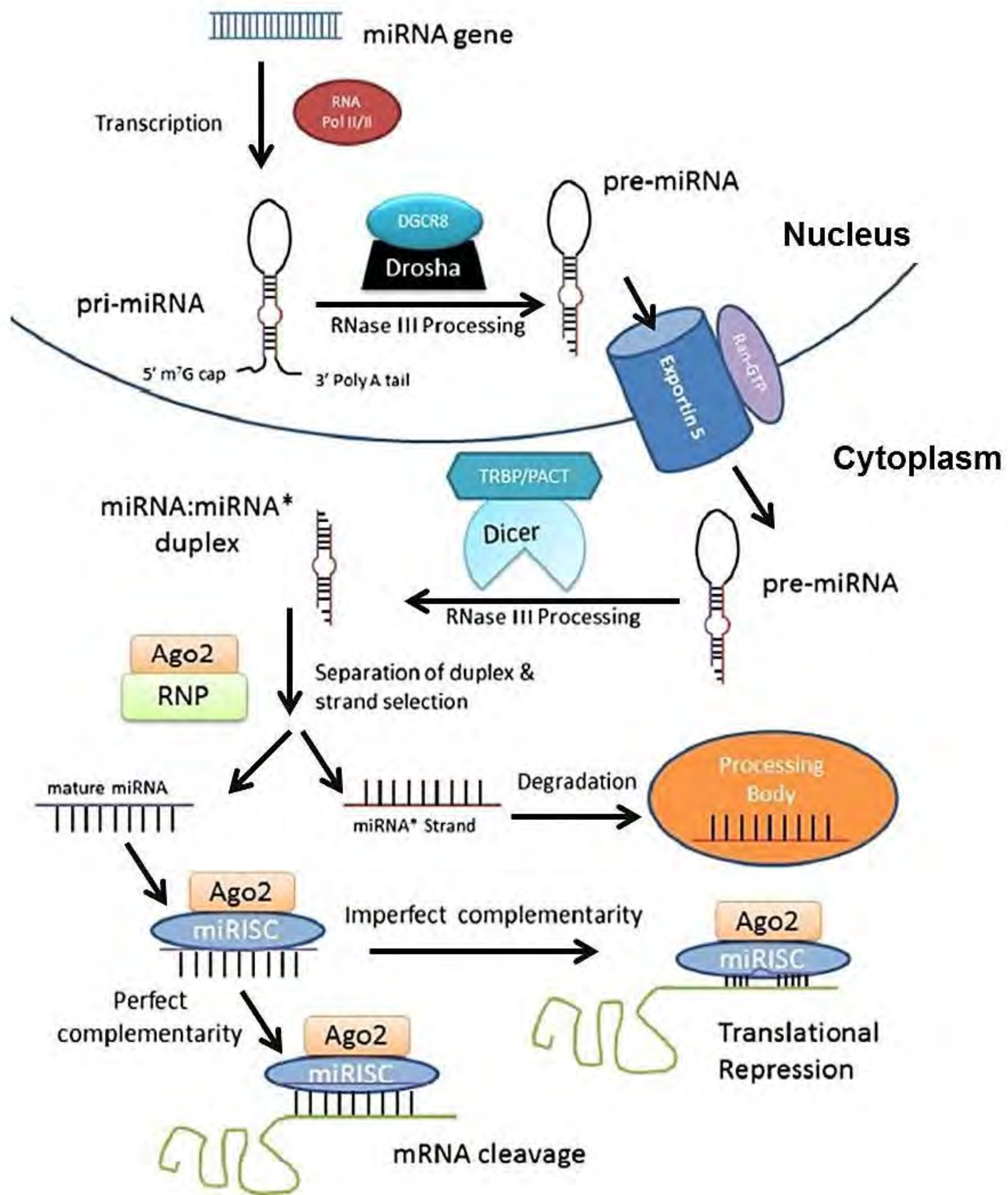


Figure 1.13 The biogenesis of miRNA modified from Rutnam *et al.* (2013). The miRNA genes are transcribed by RNA Pol II/III to make pri-miRNA. The pri-miRNA is cleaved by Drosha complexed with DGCR8 generating pre-miRNA. In turn pre-miRNA is exported by the Exportin-5 complex into the cytoplasm. The Dicer complex processes it to create the miRNA:miRNA* duplex. The RNP complex separates the duplex to create the mature miRNA which will integrate the miRISC complex in order to silence or repress target mRNAs.

Relevant to musculoskeletal soft tissue biology, there is mounting evidence that miRNAs contribute to the formation, maintenance and remodelling of the ECM (extensively reviewed by Edeleva and Shcherbata, 2013;Guller and Russell, 2010;Neves et al., 2014;Piccinini and Midwood, 2014). Piccinini *et al.* (2014) argued that miRNAs influences the ECM either directly, by targeting the mRNA of ECM molecules, or indirectly, by modulating the expression of genes that regulate the synthesis or the degradation of the ECM. This regulation is further mediated through cross-talk between the ECM and the cell adhesion and signalling mediators in the presence of the cell sensory apparatus (Edeleva and Shcherbata, 2013). To date, several members of the ECM have had their mRNA regulated by miRNAs, including the structural components such as the soluble multi-adhesive molecules (e.g. fibronectin), proteoglycans (e.g. versican), and the collagens as well as the enzymatic components such as MMPs (Edeleva and Shcherbata, 2013;Piccinini and Midwood, 2014). Recently, Piccinini *et al.* (2014) generated a comprehensive list of ECM encoding genes that are directly or indirectly regulated by miRNA. Among them were several collagen genes, namely *COL1A1*, *COL1A2*, *COL2A1, COL3A1*, *COL4A1*, *COL4A2*, *COL5A1*, *COL5A2* and *COL15A1*.

Cancer researchers have reported that miRNA let-7g inhibits cell migration expression and colony formation by down-regulating the 3'-UTR of *COL1A2*; thereby suppressing hepatocellular carcinoma metastasis (Ji et al., 2010). Moreover, the over-expression of a different miRNA, miR-301, up-regulated human breast cancer cell migration and invasion through its down-regulation of *COL2A1* while exerting additional pressure through repressing tumour suppressors (Shi et al., 2011). Another miRNA, miR-29c, was significantly down-regulated when screening

nasopharyngeal carcinomas compared to healthy tissue (Sengupta et al., 2008). The authors further observed that miR-29c was able to target multiple collagens, including *COL1A1*, *COL1A2*, *COL3A1*, *COL4A1*, *COL4A2* and *COL15A1*. Decreased levels of miRNA in these tumours could correspond to an increased expression of these ECM molecules and likely increase their invasiveness and metastatic potential (Sengupta et al., 2008).

In cartilage and bone development, miR-145 suppresses chondrogenic differentiation of murine embryonic mesenchymal cells by directly targeting *SOX9* mRNA, a positive master regulator of chondrogenesis. At early stages of chondrogenic differentiation, miR-145 expression decreases which correlates with an increased expression of *COL2A1*, *COL9A2* and *COL11A1* among other ECM molecules (Martinez-Sanchez et al., 2012;Yang et al., 2011). As expected, miR-145 overexpression results in a reduction in the levels of these ECM molecules (Yang et al., 2011). Similarly, miR-675 and miR-1247 show an interaction with *SOX9* and regulated the expression of *COL2A1* (Dudek et al., 2010;Martinez-Sanchez and Murphy, 2013). Lastly, miR-29b was shown to be a negative regulator of the *COL1A1* gene in liver fibrogenesis (Ogawa et al., 2010) while miR-133a (Castoldi et al., 2012;Matkovich et al., 2010) and miR-29b (Castoldi et al., 2012;van Rooij et al., 2008) showed an important role in myocardial fibrosis by targeting *COL1A1* among others.

With respect to disease-causing polymorphisms (mutations) and genetic variants, mechanisms affecting miRNA function in human genetic disorders have been proposed (Meola et al., 2009), mainly through: (1) variations in the miRNA nucleotide

sequences; (2) nucleotide changes in miRNA binding sites; and (3) changes in the nucleotide sequence of genes that participate in the general processes of miRNA processing, regulation and function (Figure 1.14). Of significance to this thesis, Hsa-miR-608 has a binding site in the *COL5A1* 3'-UTR at the polymorphic site rs3196378 (C/A), where the A nucleotide at that locus was preferred for repression compared to the C nucleotide (Abrahams, 2013;Laguette et al., 2011). In addition, as mentioned in section 1.7, the *MIR608* gene is also polymorphic (rs4919510, C/G) and is one of the many variants reported to be independently associated with chronic Achilles tendinopathy (Abrahams et al., 2013).

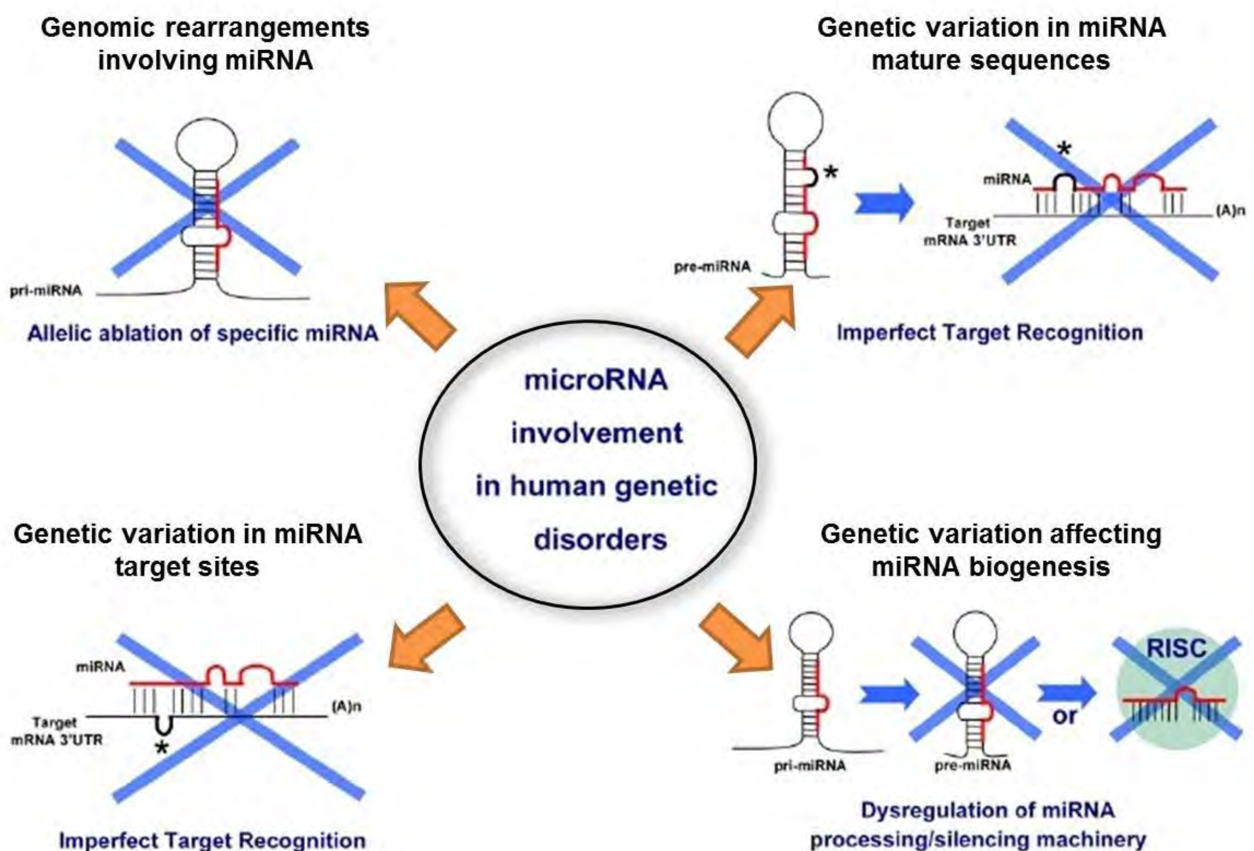


Figure 1.14 Schematic diagram representing the main types of miRNA genetic variations with a potential role in disorders (adapted from Meola *et al*. 2009).

1.8.4 Protein binding sites

Protein binding sites within the 3'-UTR allow the binding of *trans*-acting factors such as RNA binding proteins (RBP) to the mRNA. RBPs are involved in RNA modification and processing, stabilisation and destabilisation, shuttling and cellular localisation of the mRNA (Extensively reviewed in Glisovic et al., 2008;Hogan et al., 2008;Ross, 1995). The members of this vast and diverse protein family are nuclear and cytoplasmic proteins and they bind one or more RNA binding motif or domain in order to control and regulate gene expression (Glisovic et al., 2008).

Several RNA binding motifs have been characterised and used to screen for new RBPs, namely the (1) RNA-binding domain (RBD), also known as the RNA recognition motif (RRM), (2) the K-homology (KH) domain (type I and type II), (3) the RGG (Arg-Gly-Gly) box, (4) the Sm domain, (5) the DEAD/DEAH box, (6) the zinc finger (ZnF), (7) the double stranded RNA-binding domain (dsRBD), (8) the cold-shock domain; (9) the Pumilio/FBF (PUF or Pum-HD) domain, and lastly, (10) the Piwi/Argonaute/Zwille (PAZ) domain (reviewed in Chen and Varani, 2005;Lunde et al., 2007). The diversity of this class of proteins is mostly achieved via the presence of multiple combination of domains, presence of auxiliary functional domains, splice variants, post-translational modifications such as, but not limited to, phosphorylation and methylation and/or spatiotemporal-specific expression (Glisovic et al., 2008).

In addition, RBPs function in numerous cellular processes, but the focus of this thesis is their role in the regulation of post-transcriptional gene expression. Figure 1.15 provides an overview of these processes. RBPs are actively involved in splicing

events and RNA processing. For example, they play an essential role in miRNA biogenesis, e.g. Dicer, as described in section 1.8.3. Furthermore, specific RBPs are responsible for polyadenylation of the mRNA as described in section 1.8.1. Mainly, the CPSF complex together with PAPB proteins, drive the activity of the poly(A) polymerase (Bienroth et al., 1991). Export of the mRNA is also crucial and a three step process mediated by RBPs have been described by Glisovic et al (2008) as follows: (1) the generation of a cargo-carrier complex in the nucleus, followed by (2) translocation of the complex through the nuclear pore complex, and lastly, (3) release of the cargo in the cytoplasm with subsequent recycling of the carrier (Figure 1.15). For example, the RBP TAP (ATP-binding cassette transporter/ sub-family B) is a constitutive transport element binding protein, that is involved in the mRNA export process, in conjunction with other adaptor proteins mediating the affinity to the mRNA (Gruter et al., 1998;Stutz et al., 2000). Moreover, mRNA localisation, translation and mRNA turnover are key processes in the regulation of gene expression where RBPs play a critical role (Glisovic et al., 2008). For example, the RBP zipcode-binding protein-1 (ZBP1) binds the β-actin 3’-UTR and moves with the mRNA into the cytoplasm, blocking the initiation of its translation (Huttelmaier et al., 2005) until the cell lamella in polarised cells, such as in fibroblasts, is reached (Oleynikov and Singer, 2003;Ross et al., 1997).

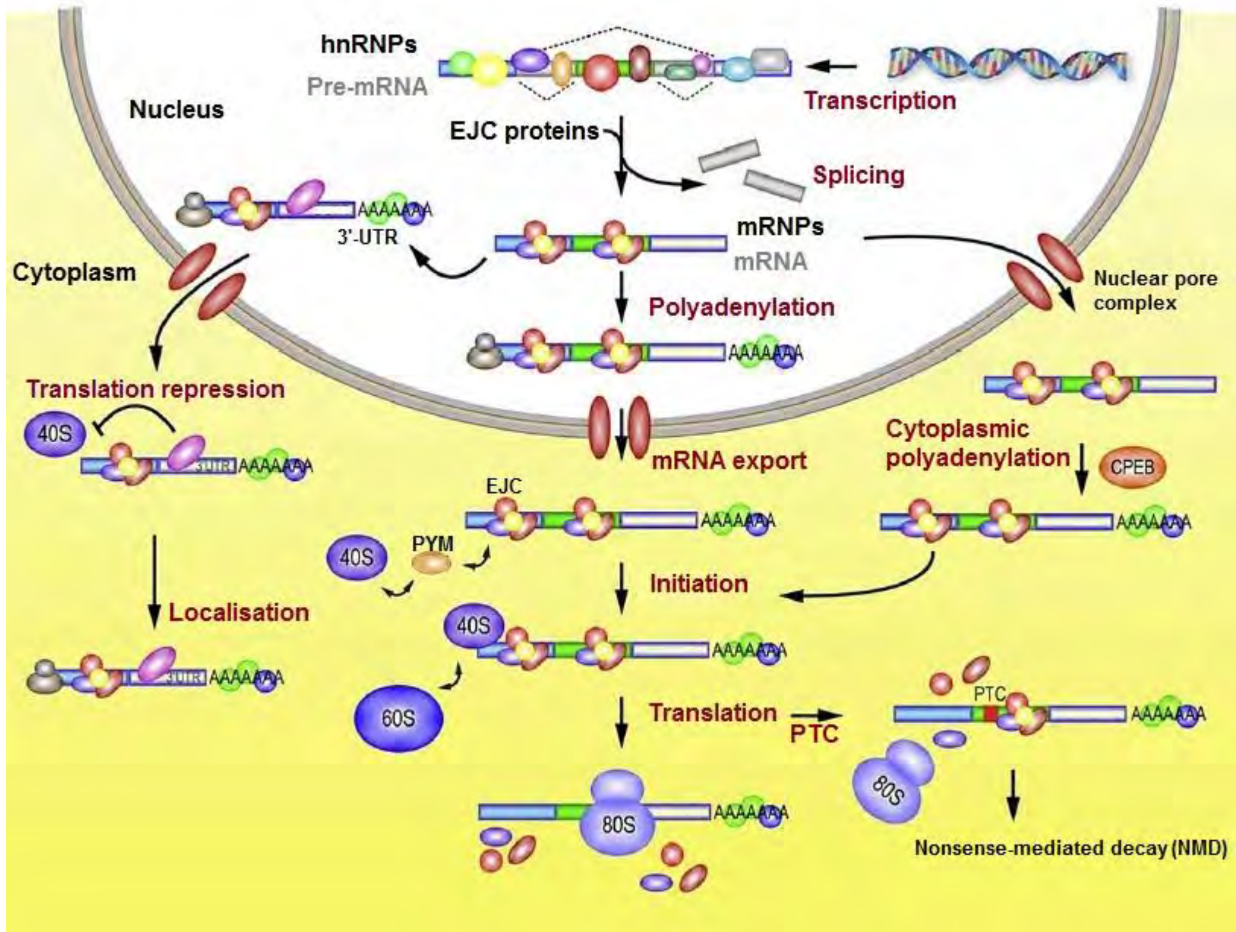


Figure 1.15 A diagram illustrating the role of RBPs in the regulation of post-transcriptional gene expression, modified from Glisovic *et al.* (2008). Many different classes of RBPs interact with various small non-coding RNAs to form ribonucleoprotein (RNP) complexes which take part in various cellular processes. In the presence of heterogeneous nuclear ribonucleoproteins (hnRNPs), the pre-mRNAs are first transcribed by the RNA polymerase II in the nucleus. These then go through several processing steps that will determine the fate of the transcript. During splicing, modifications of the exon-junction complex (EJC) occurs specifically on spliced mRNAs, and this affects the fate of the messenger ribonucleoproteins (mRNPs) in the next steps, for e.g. recruitment of ribosomal subunits for translation initiation, or surveillance of mRNA for nonsense-mediated mRNA decay (NMD). Relevant to this thesis, RBPs bound to the 3'-UTR of an mRNA can repress the initiation of translation and direct the subcellular localisation of the mRNAs. Some mRNAs, upon transport to the cytoplasm, are further modified by the cytoplasmic polyadenylation ribonucleoprotein, CPEB. PTC stands for premature stop codon. PYM is a protein that binds to the EJC.

Importantly, RBPs regulate mRNA turnover and stability. The following *cis*-elements in the 3'-UTR are recognised by trans-acting elements such as the RBPs: poly(A), iron-responsive elements, specific protein recognition sites, long range stem loops and AU-rich elements (Ross, 1995). The RBPs are able to bind these motifs through interactions with other partners influencing the stabilisation or destabilisation of the mRNA (Glisovic et al., 2008). The following examples demonstrate that the collagen genes are also regulated by RBPs. Namely, the RBP αCP was shown to bind to the COL1A1 3'-UTR in a C-rich region and mRNA decay experiments indicated that it stabilised the mRNA (Lindquist et al., 2004). In addition, the three RBPs, heterogeneous nuclear ribonucleoprotein (hnRNP) A1, E1, and K, are positive regulators of collagen synthesis acting on the 3'-UTRs of *COL1A1*, *COL1A2*, and *COL3A1* mRNAs (Thiele et al., 2004).

1.8.5 Characteristics of the *COL5A1* 3'-UTR

The 3'-UTR of the *COL5A1* gene, encoded by exon 66, contains several features that suggest an important role in the regulation of its gene. To the best of my knowledge, there are however no published reports describing a role for the 3'-UTR in *COL5A1* gene expression and indeed, although an essential gene, the regulation of *COL5A1* is poorly understood. In addition to the numerous genetically associated polymorphisms, with unexplored functions (Mokone et al., 2006;September et al., 2009), the following features, which may play important regulatory roles in the expression of *COL5A1* gene, have been identified within its 3'-UTR (Figure 1.16):

(i) Three putative polymorphic (rs1134114, C/T; rs3196378, C/A and rs11103544, T/C) miRNA binding sites, 5'-CCGACCCA-3', 5'- CCACCCCA -3' and 5'-TTTTCTAC-3' for Hsa-miR886-5p, Hsa-miR-608 and an unknown miRNA, respectively (Abrahams, 2013;September et al., 2009).

(ii) Three poly(A) signals (AAUAAA) within the distal 3'-UTR.

(iii) The first exon of an expressed sequence tag (EST; ENSESTG00000033016) which is transcribed in the antisense direction with its second and last exon at the start of *COL5A1*. To date, it is unknown if it encodes for a functional protein or not.

(iv) Several putative non-polymorphic miRNA and RBP binding sites.

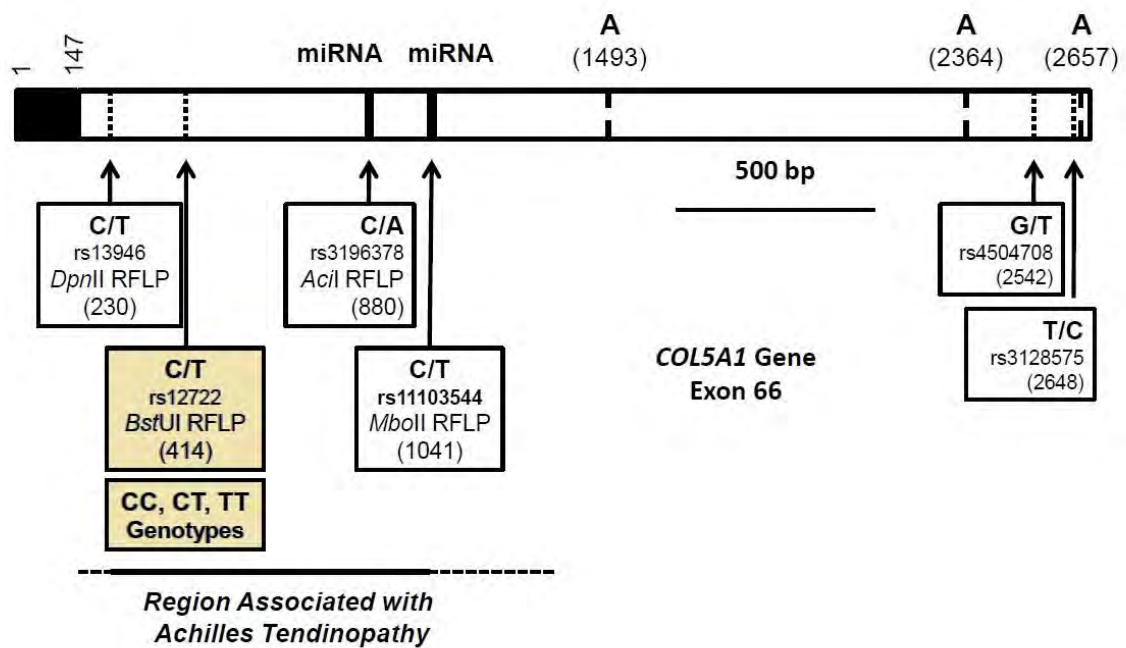


Figure 1.16 Schematic representation of *COL5A1* 3'-UTR on exon 66 (the open rectangle), showing the numerous variants investigated in genetic association studies (boxed below with arrows). The *Bst*UI RFLP (C/T, rs12722) is shown here (yellow) together with the generated genotypes from its allele combinations (CC, CT or TT). The area underlined depicts the zone strongly associated with Achilles tendinopathy and other musculoskeletal soft-tissues injuries. The putative miRNA binding sites and poly(A) signals (shown as A, bold) as well as position on exon 66 is indicated (above the rectangle).

1.9 Aim of the study

The hypothesis of this thesis is that the *COL5A1* 3'-UTR is functional and that common polymorphisms within it may result in alterations in type V collagen production which is predicted to affect fibrillogenesis and by implication the mechanical properties of musculoskeletal soft tissues. Therefore the aim of this thesis was to determine whether the 3'-UTR of *COL5A1* was functional and whether polymorphisms within the 3'-UTR generate functional differences.

The specific objectives of this thesis were therefore:

(i) To determine whether the *COL5A1* 3'-UTR was functional and to identify functional differences between the 3′-UTR from patients with chronic Achilles tendinopathy having the 'at risk' TT rs12722 genotype (*Bst*UI RFLP) and asymptomatic controls having the 'protective' CC genotype at the same locus.

(ii) To identify additional polymorphisms within the 3'-UTR that are associated with chronic Achilles tendinopathy.

(iii) To map the functional regions of the *COL5A1* 3'-UTR, focussing on those regions which are potentially responsible for contributing to the tendinopathic phenotype.

(iv) To measure *COL5A1* (Type V collagen) and *COL1A1* (Type I collagen) gene expression in primary skin biopsies from individuals with a known 'at risk' or 'protective' genotype at rs12722.

(v) Identify any putative regulatory motifs within the functional regions and determine whether there are any binding differences between the control and tendinopathic phenotypes.

CHAPTER 2 MATERIAL AND METHODS

2.1 Participants used for the Cloning of the 3'-UTR

For the purpose of this study, five de-identified individuals diagnosed with chronic Achilles tendinopathy (TEN) within the mid-portion of the tendon, as well as five healthy and physically active de-identified asymptomatic individuals with no self-reported history of tendon injuries (CON) were chosen from the participants included in a previous study (Mokone et al., 2006). By design, the five selected TEN participants all had a TT genotype at *COL5A1* rs12722, the genotype associated with an increased risk of Achilles tendinopathy (Table 2.1). In contrast, the five selected CON participants all had the "protective" CC genotype at this polymorphism. In addition, the TEN participants all had a TT genotype at rs13946 (also known as the *Dpn*II RFLP) which was also examined in the study but not associated with the disease, whereas the CON subjects were either CC (n=3) or TC (n=2) at this polymorphism (Table 2.1). The use of the participants' previously extracted total genomic DNA samples (Mokone et al., 2006) was approved for this study by the Human Research Ethics Committee of the Faculty of Health Sciences at the University of Cape Town (Appendix A).

The two groups were matched for sex (TEN 4 males vs. CON 2 males, p=0.524), age (initial injury, TEN 38.4 ± 16.1 years vs. recruitment, CON 39.7±11.5 years, p=0.910), weight (TEN 75.8±9.7 kg vs. CON 65.0±12.4 kg, p=0.198), height (TEN 178±6 cm vs. CON 175±8 cm, p=0.471) and BMI (TEN 23.8±2.0 kg.m^{-2} vs. CON 21.1±2.2 kg.m^{-2}, p=0.099) (Table 2.2).

Table 2.1 The *COL5A1* 3'-UTR single nucleotide polymorphisms rs12722 (C/T) and rs13946 (C/T) genotypes of the individual Achilles tendinopathy (TEN) and asymptomatic control (CON) participants included in this study, as well as the consensus genotypes.

	Genotype	
Participant	**rs12722**	**rs13946**
TEN-1	TT	TT
TEN-2	TT	TT
TEN-3	TT	TT
TEN-4	TT	TT
TEN-5	TT	TT
Consensus	**TT**	**TT**
CON-1	CC	CC
CON-2	CC	CC
CON-3	CC	CC
CON-4	TC	CC
CON-5	TC	CC
Consensus	**YC***	**CC**

*Y=T or C

Table 2.2 General characteristics of the individual Achilles tendinopathy (TEN) and asymptomatic control (CON) participants included in this study, as well as the average or relative values.

Participants	**Age at recruitment** (years)	**Age at Injury** (years)	**Gender**	**Height** (cm)	**Weight** (kg)	**BMI** (kg/m^2)
TEN-1	49	31	Male	186	87	25.1
TEN-2	27	20	Female	Unknown	Unknown	Unknown
TEN-3	62	61	Male	175	64	20.8
TEN-4	47	32	Male	173	73	24.4
TEN-5	49	48	Male	179	79	24.7
Ave ± SD.	46.8 ± 12.6	38.4 ± 16.1	80%*	178 ± 6	75.8 ± 9.7	23.8 ± 2.0
CON-1	Unknown	-	Female	167	54	19.4
CON-2	28	-	Male	183	78	23.3
CON-3	51	-	Female	171	56	19.1
CON-4	40	-	Female	170	58	20.1
CON-5	Unknown	-	Male	183	79	23.6
Ave ± SD.	39.7 ± 11.5	-	40%*	175 ± 8	65 ± 12.4	21.1 ± 2.2

BMI - Body mass index; Ave – average; SD. - Standard Deviation; cm - centimetres; kg – kilograms; m - metres.

*Expressed as a percentage of males in the group.

Moreover, a “severe” phenotype was an inclusion criterion for the TEN subjects and was determined by most of the following criteria:

- Early age of onset of initial symptoms (3 of 5 were ≤32 years of age at the initial injury)
- Multiple Achilles tendon injuries (2 of 5)
- Bilateral chronic Achilles tendinopathy (3 of 5)

Table 2.3. Characteristics of the Achilles tendinopathy (TEN) participants as well as features of their pathology.

Participants	Age at injury (years) *	Number of Injuries	Tendon Injured	Number of “Severe” Phenotype [a]
TEN-1	31	1	Bilateral	2
TEN-2	20	1	Unknown	1
TEN-3	61	2	Bilateral	2
TEN-4	32	2	Right	2
TEN-5	48	1	Bilateral	1

[a] Age at initial injury < 40 years, multiple injuries and bilateral Achilles tendon injury.
* Average ± standard deviation = 38.4 ± 16.1 years

2.2 Cloning the *COL5A1* 3’-UTR Reporter Gene Constructs

The 2546 bp *COL5A1* 3′-UTR was amplified from previously extracted total genomic DNA (Mokone et al., 2006) from the TEN and CON participants included in this study (section 2.1) using a nested polymerase chain reaction (PCR) and cloned into the pGL3-Promoter vector (Promega Corporation, Madison, WI, USA), substituting the SV40 (*Simian virus 40*) late poly (A) signal of the firefly luc+ reporter gene (Appendix B) (Wang et al., 2008) as illustrated in figure 2.1 and described below.

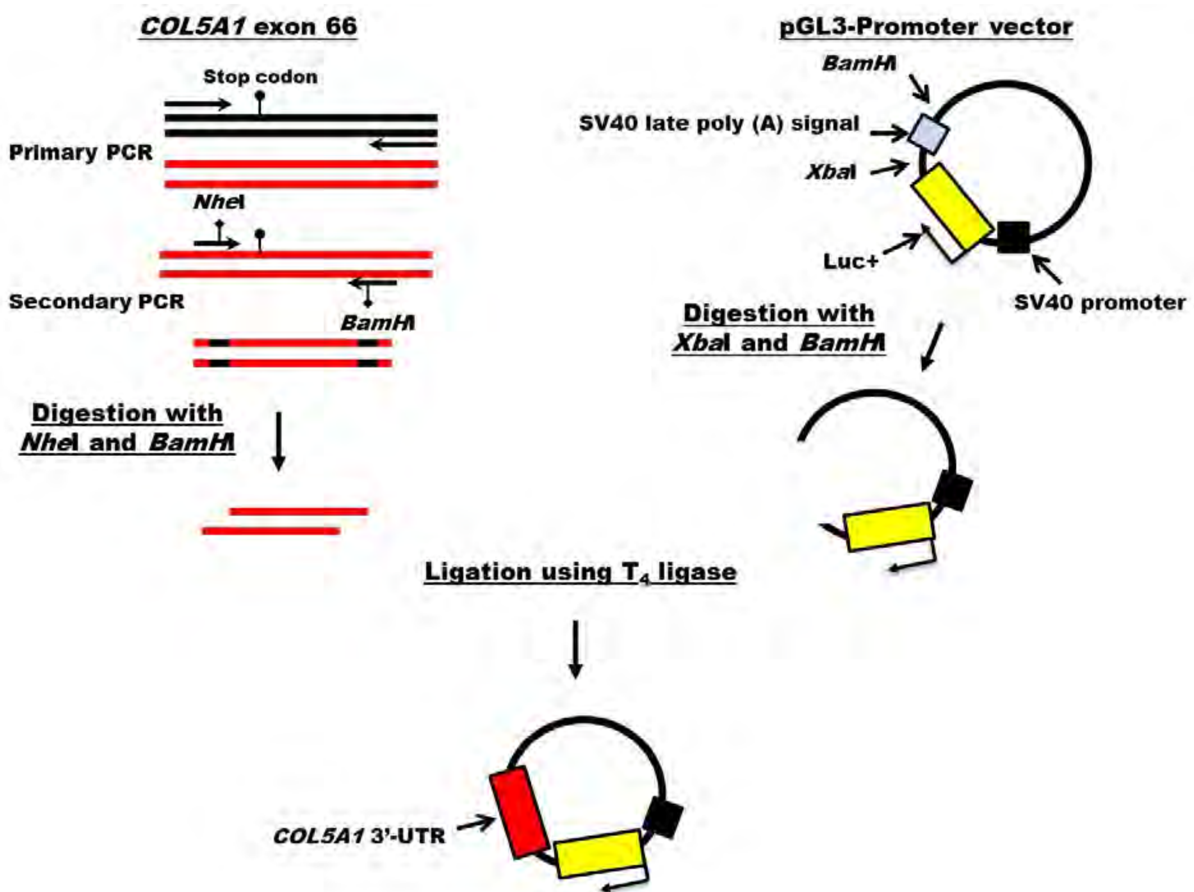


Figure 2.1 Schematic drawing representing the cloning strategy of the human *COL5A1* 3’-UTR (left, in red) into the pGL3-Promoter vector (right, in black), substituting the SV40 late poly (A) signal (blue box) of the firefly Luc+ reporter gene (yellow box).

The *COL5A1* 3′-UTR was PCR amplified using an external set of primers (forward: 5′-GAT CAG CTT CAA TCC TGT GTG TGC-3′, and reverse: 5′-GTT ATC TCC AGA GCT CCT AGC GTC CT-3′) and an internal set of primers in which restriction enzyme (RE) sites were created for cloning (forward: 5′- GCT TGC TTC ATG G**GC TAG C**AG CCG C-3′ with the *Nhe*I RE site (underlined and bold) producing complementary sticky ends to an *Xba*I RE digest, reverse: 5′-GAG TCA CTC **GGA TCC** GTT TAC GGT GG-3′ with the *Bam*HI RE site (underlined and bold). The secondary PCR products were digested with *Nhe*I (Promega Corporation, Madison, WI, USA) and *Bam*HI (New England Biolabs, Ipswich, MA, USA). The digested fragment was purified using QIAEX®II Gel Extraction kit (QIAGEN GmbH, Hilden, Germany) and then ligated using the T_4 DNA Ligase enzyme (Promega Corporation, Madison, WI, USA) into the prepared pGL3-Promoter vector. The later was digested using *Xba*I (New England Biolabs, Ipswich, MA, USA) and *Bam*HI enzymes as shown in figure 2.2.

```
GTGGGGCCGG CTTGCTTCAT GGGCTAGgAG CCGCCGAGCC CGGGCTCCCG AGAGCAACCT 180
                         NheI
CGTGACCTCA GCATGCCATT CGTTCGTGAG TGTCCCGTGC ACGTCCTGAC CCTGGACAGT 240
                                                DpnII (rs13946)
GAAGGCTTCT CCCTCCCCTC CCACCTGACT TCATCTACGC CTCGGCACCA CGGGGTGTGG 300
GACCCCAGCC CGGAGAGAAC AGAGGGAAGG AGCCGCGCCC CCACCTGGAG CTGAATCACA 360
TGACCTAGCT GCACCCCAGC GCCTGGGCCC GCCCCACGCT CTGTCCACAC CCACGCGCCC 420
                                                     BstUI (rs12722)
CGGGAGCGGG GCCATGCCTC CAGCCCCCCA GCTCGCCCGA CCCATCCTGT TCGTGAATAG 480
                                     rs1134114
GTCTCAGGGG TTGGGGGAGG GACTGCCAGA TTTGGACACT ATATTTTTTT CTAAATTCAA 540
CTTGAAGATG TGTATTTCCC CTGACCTTCA AAAAATGTTC CAAGGTAAGC CTCGTAAAGG 600
TCATCCCACC ATCACCAAAG CCTCCGTTTT TAACAACCTC CAACACGATC CATTTAGAGG 660
CCAAATGTCA TTCTGCAGGT GCCTTCCCGA TGGATTAAAG GTGCTTATGT TTTTGTGAGT 720
TTTAAGTAAA TATTTGTATT GTATTGTTAT AAATGTTAAG TGTGCCTGGC TTTCAATCAT 780
GCACGGAAAC CCAGTCTCAG TCCCACGGAC AGAATGGGCG AGGCATGGAT TCTGGGTTGC 840
AGTACCGTTC TGATTAGAAA TAGGAAGTCT CCCCACCCCC GCCCTGGCCA AGAACGTGCA 900
                                  hsa-miR-608
                                         AciI (rs3196378)
ATAAATTGGA AGTTTGCCCC GGGGCAGCAA GAATTTATGC TGCCATTGAA AAGCAGGTAC 960
CAGTGCCCCT TTTCAGACAG TTTTTGATTC GCTCTAGACT TTTTTTTTTT TTAATAGGGA 1020
                                   XbaI    rs71698207   rs71746744
AAAAATTTGA TAATTTTCTT TTTTCTACAT GCACTTAAGA CTAAAACACA GGTTTGGATT 1080
                  MboII (rs11103544) AflII
AATTTTATTT GCTTCCTTTT TCCGCTTTTC TTCCCGCAGA GCCTGATGGG AGAATGTCCA 1140
GGGCAGGGAA ACCACATTTT TTGTAGGTGA TAACTCAATG AAAATTGGTG CTTATTTTTT 1200
ACACTTCTCT CTTGTGGCTC TCTTGTGGTG CTATCTATCT GTTTTAAGGT CTCCTTGAAG 1260
GCGCACTGGG GACCCTGGCC ATGCCTCGTT CTCCCTGCTT TCTTTATCCT GTTATTGCCT 1320
CCACAGTCTG TTGCCAAGGA CTCTAAGATC AATGCACGTC ACTTTCCTTT CCACTGGGCA 1380
GGATAGCCAA GCACACTCCC TCCTGCGCTC TCCCGCCCCG GTGCGTCCAC TCCCGAGGGC 1440
TGTTATGAGG ACTGGGTTGT GCCTACTTGA TTTGAAAACA CACACAAGCa ataaaaaGCC 1500
TCTTCCTGCA TTGTCTGTGG TGTGACCATA GCAGATTATA TTTGGTTCCT GAaTGTTTGT 1560
GGTGCTAATT TCTGTGTTTG TTCCAAGCCG TTCAGTCATG CCATGCGCTG CCTCGGTAGA 1620
TGGAGTAATG TACAATGAAC TCCATGAGTC TCTCCAGGGC TGCCTGCAGC ACGTCTTTTC 1680
CAAGTAGCCT ATTTGGATTC CCATCTCAAA TGTCCTGGAT GCGAGCGTCA GCGGCTCCAG 1740
AGCTCGGGGC GGGTGAGGTC CCCTTTGGGG AACCCTTTCC TGGCCATCGA GGTCGGGGGG 1800
CTGCCGTCTG TGGGCAGGAG GACCCGAGGG GCAGCCAGGA AAGGCGATCT CTTCACTGTG 1860
AAAAGTTGCC CGGGTGCAGC GCCTTTTCCT TCTACCATGG GAAATGCAGG CTGGGCCCTT 1920
GGGGTGAGCC TGCGGGGCTC TGGTGCTGTC CCCGACCCCC ACCACCACCA GAATGCAGTT 1980
CCAGCTTAGG AAGCCACAAA CAAGCCACCC AGGAGGAACA AAACACCGCC AGCGTGGATT 2040
TTCCAAATTT CCCTGGAAAG TAAGTCTCGC TCTTGCCAAA GAAAAGTCTG GCTTGGAGAG 2100
TCTCTGGAGC CCAGGATGCC AGCATGTGCC AATGACTGTC ACCTTCATCT CTTCAAAAGA 2160
AAAGCCATAG CCGAGGACTG TCCCGCGACC CCCGTGGACT GCGTCTAGGT CATGTGATTC 2220
                                        TthIII
TGTTTTCATT TCTCATCCCA TCCAATTTGT CCTTTTCTCC TGTCATTTTC TTCCTCTGTG 2280
GTCCCTTCAA AGTTGTTATA ATTTGTACTG AACTTCAAAA TGTGTCCCGT TCTCCCCAGA 2340
CCACTCTAGC CACAGTATAT TGCaataaaa TTACTTCTTA TATTTGCAGA AaTTCTTTTG 2400
GTGTAATTTT ATTTTTTCCT CTCAATATAT ATAATTGGAC AAACGCTGGC AAAAAGAAAA 2460
AAATGGTAAG CAAAAAACCC AAGATAAAGT TTCGAGGACA TCAGGCCTTT TGAAATACAA 2520
TGTCAAATGA CACATTGTAC GGTTTCAAAA AATCCGCTAG ACATGTCATA AGTTTTAACT 2580
GTAATGCCCA GGAAAGGATA TCTTAAAATA TTCTAAACTT GTGTAACAAA GGAATAATTA 2640
ACTGTAATAG TTTTTCaata aaTCGAGTTG GGTGTTTCCA CCGTAAaCgG ATccGAGTGA 2700
                                                     BamHI
```

(Please see figure legend on the following page)

Figure 2.2 Nucleotide sequence of *COL5A1* exon 66 (PubMed accession no. NM_000093) containing the 3'-untranslated region (UTR) (nucleotides 148 to 2688). The sequences highlighted in grey represent the last 27 nucleotides of the translated region of exon 66 (nucleotides 121-147). The stop codon (TAG) within the exon is double underlined. Nucleotide numbers of exon 66 are annotated on the right side of the sequence. The positions of key polymorphisms within the 3'-UTR are highlighted in black with the text in white. The sequence contains the wild-type nucleotides of the polymorphic sites. The accession numbers of selected polymorphisms are given, including, when appropriate, the name of the restriction fragment length polymorphism. The four unlabelled polymorphic sites are the (i) ATCT at position 1237, rs16399; (ii) A at position 1272, rs1134170; (iii) G at position 2542, rs4504708; and (iv) T at position 2648, rs3128575. The putative recognition sequences of the two micro RNAs are highlighted in grey, while the three putative polyadenylation signals are in white lower case and highlighted in black. Key restriction sites are annotated on the sequences. Mutations generated in the sequences of the 3'-UTR are shown in lower case. These were used to create *NheI* and *BamHI* restriction sites for cloning purposes.

2.3 Generation of Deletion Constructs of the COL5A1 3'-UTR

Two deletion constructs (Figure 2.3) pGL3-COL5A1-3′UTR (Δ488) using RE *Tth*111I (Promega Corporation, Madison, WI, USA) and *Bam*HI, and pGL3-COL5A1-3′UTR (Δ57) using *Xba*I and *Afl*II (New England Biolabs, Ipswich, MA, USA) were generated from the parental constructs (pGL3-COL5A1-3'UTR) cloned from the TEN and CON participants. The parental constructs were digested using the respective restriction enzyme pairs sequentially before being visualized on a 0.7% agarose gel. The DNA fragment generated was gel purified or re-precipitated as described before. The digested constructs were blunt-ended using the DNA Polymerase I Large (Klenow) Fragment (Promega Corporation, Madison, WI, USA) according to the manufacturer's recommendations. This was followed by gel purification or re-precipitation before ligation using the T_4 DNA Ligase enzyme. The ligation products were transformed into competent E. coli bacteria (DH05α strain) as described in section 2.5.

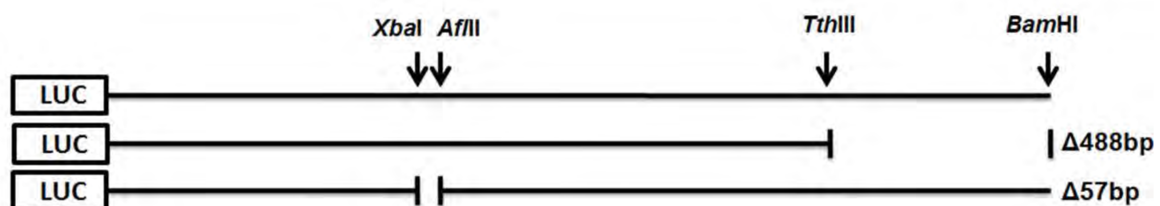


Figure 2.3 Construction of two deletion constructs of the *COL5A1* 3'-UTR. The parental constructs (pGL3-COL5A1-3′UTR, on top) cloned from the TEN and CON participants were digested using the annotated respective enzymes to remove the indicated sequences of the parental *COL5A1*-3'UTR to generate deletion constructs pGL3-COL5A1-3′UTR (Δ488) and pGL3-COL5A1-3′UTR (Δ57).

2.4 Generation of Site-Directed Mutants within the *COL5A1* 3'-UTR

Site-directed mutagenesis (SDM) of rs12722 (C/T) within the *COL5A1* 3'-UTR (Figure 2.2) of the parental constructs (pGL3-COL5A1-3'UTR) cloned from the TEN and CON participants was performed using the QuikChange™ Site-Directed Mutagenesis guidelines (Stratagene, La Jolla, CA, USA). Table 2.4 lists the primers (IDT, Coralville, IA, USA) used to generate the SDM mutants, which were designed as recommended by the QuikChange™ Site-Directed Mutagenesis manual. The PCR reaction was optimized and performed in a total volume of 50 µl containing 50 ng of circular plasmid DNA (pGL3-COL5A1-3'UTR), 20 µM of each of the forward and reverse primers, 1X reaction buffer (20mM Tris-HCl, pH 8.8; 10mM KCl; 10mM $(NH_4)_2SO_4$; 2mM $MgSO_4$; 0.1% Triton® X-100 and 0.1mg/ml nuclease-free BSA), 10 mM of each dNTPs, 4 µl of Dimethyl Sulphoxide (DMSO) (Sigma-Aldrich, Munich, Germany) and 1 µl of *Pfu* DNA polymerase (Promega Corporation, Madison, WI, USA). The cycling conditions included an initial denaturing step at 95°C for 1 min followed by 17 cycles of denaturing at 95°C for 45 s, annealing at 55°C for 1 min and extension at 73°C for 14 mins using an Applied Biosystems 2720 Thermal cycler (Applied Biosystems, Foster City, CA, USA). Two mutants were generated: (i) TEN rs12722_T>C, where the T polymorphic nucleotide of rs12722 was mutated to a C nucleotide within the parental constructs (pGL3-COL5A1-3'UTR) cloned from the TEN participants and (ii) CON rs12722_C>T, where the C polymorphic nucleotide of rs12722 was mutated to a T nucleotide within the parental constructs (pGL3-COL5A1-3'UTR) cloned from the CON participants, as illustrated in figure 2.4.

Table 2.4 Primers designed for the site-directed mutagenesis of rs12722 (C/T) within the *COL5A1* 3'-UTR (figure 2.2) of the parental constructs (pGL3-COL5A1-3'UTR) cloned from the TEN and CON participants.

Primer Name	5'- Position exon 66 [a]	Primer Sequence (5'-3')
FWD TEN rs12722_T>C	400	TCT GTC CAC ACC CA**C** GCG CCC CGG GAG
REV TEN rs12722_T>C	425	TCC CGG GGC GC**G** TGG GTG TGG ACA GAG [b]
FWD CON rs12722_C>T	400	TCT GTC CAC ACC CA**T** GCG CCC CGG GAG
REV CON rs12722_C>T	425	TCC CGG GGC GC**A** TGG GTG TGG ACA GAG [b]

The mutated nucleotides within the primers are highlighted in bold and underlined.
[a] Refer to figure 2.2
[b] reverse complement

For variant rs71746744 (-/AGGG), within the *COL5A1* 3'-UTR (Figure 2.2) of the parental constructs (pGL3-COL5A1-3'UTR) cloned from the TEN and CON participants, mutants were generated using the KAPAHiFi™ DNA Polymerase (Kapa Biosystems, MA, USA) and phosphorylated primers as recommended by the manufacturer (Table 2.5).

Table 2.5 Primers designed for the site-directed mutagenesis of rs71746744 (-/AGGG) within the *COL5A1* 3'-UTR (figure 2.2) of the parental constructs (pGL3-COL5A1-3'UTR) cloned from the TEN and CON participants.

Primer Name	5'-Position exon 66 [a]	Primer Sequence (5'-3')
FWD CON rs71746744_C>T	1016	/5Phos/**AGG G**AA AAA ATT TGA TAA TTT TCT TTT TTC TAC ATG CAC
FWD TEN rs71746744_T>C	1020	/5Phos/AAA AAA TTT GAT AAT TTT CTT TTT TCT ACA TGC AC
REV CON rs71746744	986	/5Phos/**CCC T**AT TAA AAA AAA AAA AAA AAG TCT AGA GCG AAT C [b]

The mutated nucleotides within the primers are highlighted in bold and underlined. 5Phos denotes the 5' position of the phosphorylated nucleotide.

[a] Refer to figure 2.2

[b] reverse complement

For the deletion mutagenesis, the primers were designed with a gap between them (FWD TEN rs71746744_T>C and REV CON rs71746744), while for the insertion, the extra nucleotides were added at the 5'- position of the forward primer (FWD CON rs71746744_C>T and REV CON rs71746744). The PCR reaction was performed in a total volume of 25 µl containing 5 ng of circular plasmid DNA, 0.3 µM of each of the forward and reverse primers, and 12.5 µl of 2x KAPAHiFi™ HotStart ReadyMix (1x concentration containing 0.3 mM of each dNTP, 2.5 mM MgCl2 and 1 U of KAPA HiFi HotStart DNA Polymerase in a proprietary reaction buffer) (Kapa Biosystems, MA, USA). The cycling conditions included an initial denaturing step at 95◦C for 2 mins followed by 16 cycles of denaturing at 98°C for 20 s, annealing at 60°C for 15 s, an extension at 72°C for 3 mins 30 s and a final extension step at 72°C for 5 mins

using an Applied Biosystems 2720 Thermal cycler (Applied Biosystems, Foster City, CA, USA).

As illustrated in figure 2.4, the generation of two mutants were attempted: (i) CON rs71746744_ins AGGG, where a second AGGG tetranucleotide was inserted into rs71746744 of the parental constructs (pGL3-COL5A1-3'UTR) cloned from the CON participants and (ii) TEN rs71746744_del AGGG, where the second AGGG tetranucleotide was deleted from rs71746744 of the parental constructs (pGL3-COL5A1-3'UTR) cloned from the TEN participants.

Following the SDM PCR reactions, the mutants were subjected to a digestion by the *Dpn*I enzyme (Promega Corporation, Madison, WI, USA) in order to remove the residual methylated parental templates and any hemi-methylated DNA following the manufacturer's recommendations for 1hr at 37°C and heat inactivated at 65°C for 20 mins. Mutants for rs71746744 required an extra ligation step after the above digestion. This was done as described previously. Finally, all the mutants generated were transformed into competent E. coli bacteria (DH05α strain) as described in section 2.5. Maxi preparations of the constructs were then made and purified as described in section 2.6.

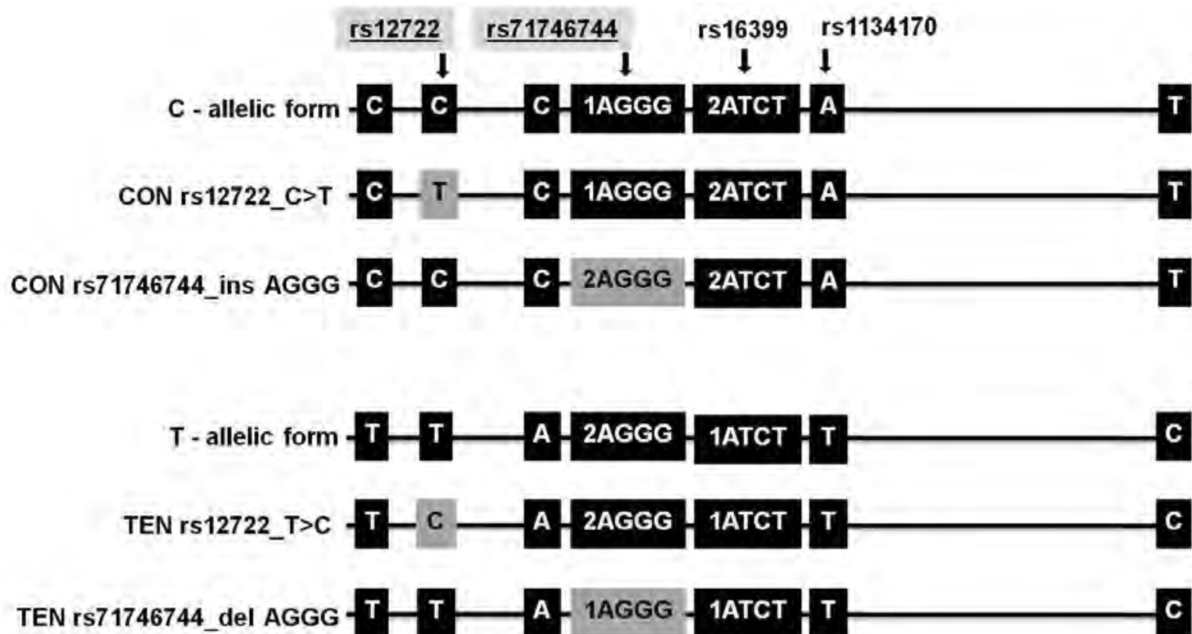


Figure 2.4 Schematic diagram representing the site-directed mutations (Grey boxes) that were generated in the *COL5A1* 3'-UTR cloned from the TEN and CON participants. **Top Panel:** CON rs12772_C>T, where the C polymorphic nucleotide of rs12722 was mutated to a T nucleotide within the parental constructs (pGL3-COL5A1-3'UTR) cloned from the CON participants (C-allelic form). CON rs71746744_ins AGGG, where a second AGGG tetranucleotide was inserted into rs71746744 of the parental constructs (pGL3-COL5A1-3'UTR) cloned from the CON participants. **Bottom Panel:** TEN rs12722_T>C, where the T polymorphic nucleotide of rs12722 was mutated to a C nucleotide within the parental constructs (pGL3-COL5A1-3'UTR) cloned from the TEN participants (T-allelic form). TEN rs71746744_del AGGG, where the second AGGG tetranucleotide was deleted from rs71746744 of the parental constructs (pGL3-COL5A1-3'UTR) cloned from the TEN participants.

2.5 Production of Competent Cells and Transformation

The method described below was initially developed to optimize bacterial transformation efficiency, transformation being the uptake of naked DNA by bacterial cells (Hanahan, 1983). A single E. coli colony (DH5α strain) from a freshly streaked SOB (Super Optimal Broth) agar plate (Appendix B) was resuspended in 5 ml of Transformation broth (TFB) (Appendix B) and shaken overnight at 37ºC. This starter culture was then inoculated into 100 ml of pre-warmed TFB broth in a 1 L conical flask, and shaken at 37ºC until the log phase growth (an Optical Density of 0.35 at 600 nm) was reached. This culture was transferred to an ice-cold centrifuge tube and chilled on ice for at least 15 mins. It was pelleted by centrifugation at 3000 rpm for 10 mins at 4ºC using the rotor JA-14 in the J2-21 centrifuge (Beckman, USA). The supernatant was discarded and the pellet gently resuspended in 21 ml of Transformation buffer 1 (TFB1) solution (Appendix B). The resuspended cells were incubated on ice for 90 mins and centrifuged as described before. The supernatant was decanted once more and the cells resuspended in 4 ml of ice-cold TFB2 solution containing glycerol (Appendix B). Finally, 100 µl aliquots of the above resuspension were stored in the -70ºC until needed.

During transformation, a 100 µl aliquot of competent cells was mixed with 10-50 ng of plasmid DNA by gentle flicking and left on ice for 30 mins. The mixture was moved to a 42ºC pre-heated water bath for 1 min and then put back on ice for 2 mins. Nine hundred microlitres of pre-warmed SOC (Super Optimal broth with Catabolite repression) medium (Appendix B) was added and recovery, as well as expression of the antibiotic resistance gene was allowed by incubation for 30 mins at 37ºC.

Typically 50, 100 and 200 µl aliquots of the transformation mixture were plated onto SOB agar plates, containing 100 µg/ml of Ampicillin (Sigma-Aldrich, Munich, Germany) where appropriate. The plates were inverted and left overnight at 37ºC.

2.6 Maxi Preparation of Plasmid DNA by Caesium Chloride and Ethidium Bromide Gradient Centrifugation

Following an adapted protocol (Sambrook et al., 1989), 50 ul of glycerol stock of DH5α competent cells, transformed with DNA plasmids, was grown overnight in 5 ml of SOC medium and 100 µg/ml Ampicillin. From this starter culture, 1ul was inoculated into 400 ml of SOC medium with 100 µg/ml Ampicillin and grown overnight at 37ºC at 200 revolutions per minute (rpm). The overnight culture was spun at 7000 rpm for 10 mins at 4ºC using the JA-20 rotor in the J-21B centrifuge (Beckman, USA). The supernatant was removed and the pellet was resuspended in 10 ml of Solution I (Appendix B), followed by the addition of 80 ml of Solution II (Appendix B). After several inversions, 45 ml of ice-cold Solution III (Appendix B) was added. The resulting white precipitated solution was spun as described above. The supernatant was removed and filtered into a clean centrifuge tube containing 90 ml of isopropanol (Merck (Pty) Ltd, South Africa). The precipitated DNA was spun down at 7000 rpm for 10 mins using the above centrifuge. The supernatant was discarded and the pellet air-dried briefly. The DNA was resuspended in 2 ml of sterile double distilled water, transferred to 10 ml tubes containing 4.8 g of caesium chloride (CsCl) (Roche Diagnostics, Mannheim, Germany) in 2.1 ml of sterile distilled water. At that point, 200 µl of ethidium bromide (10 mg/ml, Promega Corporation, Madison, WI, USA) was added to each tube. The solution was spun for 10 mins at room

temperature at 3000 rpm using the Labofuge 400R centrifuge (Heraeus, Germany). The top clear red solution was transferred into an ultracentrifuge tube (Beckman Quick-Seal Ultracentrifuge tubes, USA) using a 2 ml syringe fitted with a 21-gauge needle. The tubes were heat-sealed following the manufacturer's instructions and centrifuged for 16 hrs at 53,000 rpm at 20ºC using the VTI 65.2 HR rotor in the L8-79M Ultracentrifuge (Beckman, USA).

On the following day, the lower band representing plasmid DNA was extracted using a syringe fitted with a 21-gauge needle and the extracted band released into a 10 ml tube. The subsequent plasmid DNA was purified using the following protocol (Heilig et al., 1998). Briefly, the ethidium bromide and any RNA contamination was extracted from the plasmid DNA solution using organic solvent isobutanol (Sigma-Aldrich, Munich, Germany) and digested 10 mg/ml RNase A (Sigma-Aldrich, Munich, Germany) respectively. The solution (1 ml) was transferred into a dialysis cassette (Pierce, thermo scientific, MA, USA) and immersed in a beaker containing TE buffer pH 7.5 (see appendix B) and the process of dialysis was allowed to take place overnight at room temperature for the removal of CsCl. Proteinase K (Promega Corporation, Madison, WI, USA) was dissolved in 0.2% Sodium dodecyl sulphate (SDS) (Sigma-Aldrich, Munich, Germany) and added to the DNA solution for an hour at 37ºC followed by a phenol:chloroform (1:1) extraction to remove any protein contamination.

Finally, the plasmid DNA was concentrated by ethanol precipitation (Sambrook et al., 1989). Purity and concentration of the eluted DNA was calculated by nanodrop

analysis using the NanoDrop® ND-1000 spectrophotometer NanoDrop (Coleman Technologies Inc., USA) whereas integrity was observed on a 0.7% agarose gel.

2.7 Sequencing

Different primer pairs were created to PCR amplify selected regions of the cloned *COL5A1* 3′-UTR (section 2.2) and deletion constructs (section 2.3) (Table 2.5). All primers were manufactured by IDT (Coralville, IA, USA). The PCR reactions were performed in a total volume of 60 µl containing at least 50 ng of plasmid DNA, 20 pmol of each of the forward and reverse primers, 1X reaction buffer, 200 µM of each dNTPs and 1 unit of *Taq* DNA polymerase (New England Biolabs, Ipswich, MA, USA). The cycling conditions included an initial denaturing step at 94°C for 3 mins followed by 35 cycles of denaturing at 94°C for 1 min, annealing between 60-65°C depending on primer pairs used for 1 min, extension at 72°C for 1.5 mins and a final extension step at 72°C for 5 mins using the XP Thermal Cycler Block (Bioer Technology, Tokyo, Japan). The PCR products were then gel purified using the QIAEX®II Gel Extraction kit (QIAGEN, Hilden, Germany) as described by the manufacturer and sequenced at the Central Analytical Facility (University of Stellenbosch, Stellenbosch, South Africa) using the appropriate primers.

Furthermore, the insert and vector of all the site-directed mutant clones (section 2.4) were completely sequenced by Inqaba Biotec (Pretoria, South Africa) following maxi-preparation purification. An independent set of primers (Inqaba Biotec, Pretoria, South Africa) was created for this purpose (Table 2.6), in addition to the primer set

below. BioEdit version 7.0.5.2 software (www.mbio.ncsu.edu/bioedit/bioedit.html) was used to analyse the obtained sequence information.

Table 2.6 Primers designed to sequence the cloned *COL5A1* 3'-UTR.

Primer Name	5'- Position exon 66 [a]	Primer Sequence (5'-3')
FWD1	1850 [b]	AAG GTC TTA CCG GAA AAC TCG ACG
REV1	488	CCT GAG ACC TAT TCA CGA AC [c]
FWD2	667	GTC ATT CTG CAG GTG CCT TC
REV2	774	GAA AGC CAG GCA CAC TTA AC [c]
FWD3	1058	AGA CTA AAA ACA CAG GTT TGG
REV3	1230	CAC CAC AAG AGA GCC ACA AG [c]
REV4	1468	AAG TAG GCA CAA CCC AGT CC [c]
FWD4	1546	TTC CTG AAT TGT GGT GC
REV5	1912	AGC CTG CAT TTC CCA TGG [c]
FWD5	2087	TCT GGC TTG GAG AGT CTC TGG
REV6	2278	CAG AGG AAG AAA ATG ACA GG [c]
FWD6	2483	GAT AAA GTT TCG AGG ACA TC
REV7	2284 [a]	CAT AAG TGC GGC GAC GAT AGT CAT GC [c]

[a] Refer to figure 2.2

[b] 5'-position in the pGL3-Promoter vector (Accession no. U47298.2)

[c] reverse complement

Table 2.7 Primers designed to sequence the entire plasmids containing the site-directed mutants within rs12772 and rs71746744 of the cloned *COL5A1* 3'-UTR.

Primer Name	5'-Position Construct [a]	Primer Sequence (5'-3')
SEQ1	640	GAA CGT GAA TTG CTC AAC AG
SEQ2	1280	TGC CAA GAG GTT CCAT CTG C
SEQ3	5040	TGT AGG TCG TTC GCT CCA AG
SEQ4	5680	TCC ATA GTT GCC TGA CTC CC
SEQ5	6240	AGA TGC TTT TCT GTG ACT GG
SEQ6	6880	TTC GCC CTT TGA CGT TGG AG
SEQ7	7235	CTA GCA AAA TAG GCT GTC CC

[a] 5'-Position in the pGL3-Promoter vector sequence (Accession no. U47298.2) with inserted 3'-UTR at position 1938.

2.8 Annotation of the Template Sequence

The NCBI website (http://www.ncbi.nlm.nih.gov/nuccore/89276750) was used to annotate the multiple poly (A) signals present on the 3'-UTR of *COL5A1*. Several miRNA databases (Patrocles, www.patrocles.org) (Hiard et al., 2010), miRBase (www.mirbase.org) (Griffiths-Jones, 2004;Griffiths-Jones et al., 2006) and MicroCosm Targets (www.ebi.ac.uk/enright-srv/microcosm) were used to annotate the putative miRNA binding sites on the sequence.

2.9 Cell Culture

HT1080 fibrosarcoma cells (ATCC® no. CCL-121) (Rasheed et al., 1974), SVWI-38 cells (ATCC® no. CCL-75) (de Haan et al., 1986) as well as primary skin fibroblast cell lines (section 2.17) were maintained in Dulbecco's modified Eagle's medium (DMEM) (Highveld Biological, South Africa) containing 10% foetal bovine serum (FBS) (Highveld Biological, South Africa), 200 units/ml penicillin and 100 µg/ml streptomycin (Complete DMEM). The HT1080 human fibroblast cell line is derived from a fibrosarcoma arising adjacent to the acetabulum, at the hip joint, of a 35 years old Caucasian male individual (Rasheed et al., 1974) while the SVWI-38 is a human embryonic lung fibroblast cell line (WI-38) transformed with the *Simian virus 40* (SV40) (de Haan et al., 1986). The lines were incubated at 37°C (95% air, 5% CO2, 65% humidity). The cells were regularly monitored for mycoplasma and infections.

2.10 Mycoplasma Testing

The cells were grown on a sterile coverslip in a 35 mm culture dish in antibiotic-free medium (DMEM with 10% FBS) for a minimum of 24 hrs. On testing day, the cells were fixed using a mixture of glacial acetic acid and methanol (ratio 1:3) for a few seconds and washed gently with water to remove the fixing solution. Once the coverslip had air-dried for a few minutes, the DNA was stained with 500 µl of Hoechst no. 33258 (0.5 µg/ml) for not more than 30 sec and washed off with water to prevent excessive staining. The coverslip was then mounted on a slide with a drop of mounting fluid (0.02M citric acid, 0.06M $Na_2HPO_4.H_2O$ and 50% glycerol, pH 5.5). The cells were viewed shortly after setting of the slide by fluorescence microscopy.

Mycoplasma positive cells will have staining in the nucleus as well as in the cytoplasm and around the cells.

2.11 Transient Co-Transfections and Luciferase Assays

Transient co-transfections were performed using FuGENE HD (Roche Diagnostics, Mannheim, Germany) according to the manufacturer's instructions. The HT1080 or SVWI-38 cells were plated at 1.0 X 10^5 cells per well in 12-well plates and 24 h later co-transfected with 500 ng of a pGL3-COL5A1-3′UTR vector and 50 ng of the pRL-TK vector containing the thymidine kinase promoter driving the expression of a renilla reporter which was used as an internal control for transfection efficiency. Each experiment was done in triplicate and repeated at least twice. In all the transfection experiments, cells were cultured for 45 hrs and extracts were assayed for firefly and renilla luciferase activity using the dual luciferase assay system (Promega Corporation, Madison, WI, USA). Luciferase activities were measured using the Luminoskan Ascent luminometer (Thermo Labsystems, Franklin, MA, USA). Firefly luciferase values were normalised to the renilla luciferase activity.

2.12 Messenger RNA Decay Assay

HT1080 cells were plated at 0.5 X 10^5 cells per ml in 6 cm dishes in 4 ml of complete DMEM (Highveld Biological, South Africa) at 37°C (95% air, 5% CO2, 65% humidity). Co-transfections using 1 ug of TEN and CON full constructs (pGL3-COL5A1-3′UTR) as well as 100 ng transfection efficiency construct pRL-TK were done using FuGENE HD (Roche Diagnostics, Mannheim, Germany) according to the manufacturer's

instructions. After 30 hrs of co-transfection, one experimental subset was treated with 5μM Actinomycin D for 1 hr to stop *de novo* transcription while the other experimental set was not treated using this inhibitor.

Thereafter, total RNA was extracted using the High Pure RNA isolation kit (Roche Diagnostics, Mannheim, Germany) from the treated and untreated plates at the following time intervals (t): 0, 15, 30 mins and 1 hr. The RNA was quantified and integrity assessed on a 1 % agarose gel using the RiboRuler™RNA ladder and 2X RNA Loading Dye (95% formamide, 0.025% SDS, 0.025% bromophenol blue, 0.025% xylene cyanol, 0.025% ethidium bromide and 0.5 mM EDTA) (Fermentas, thermo scientific, MA, USA) following the manufacturer's recommendations. It was kept at -80°C until cDNA was obtained using the Improm-II™ Reverse Transcriptase kit (Promega Corporation, Madison, WI, USA). Finally the transcript abundance was quantified in real time using the Power SYBR green PCR master (Applied Biosystems, Foster City, CA, USA) and the PCR reactions were carried out in the Applied Biosystems StepOnePlus real-time PCR System (Life Technologies, Applied Biosystems, Foster City, CA, USA) following the manufacturer's recommended cycling conditions. The threshold cycle (C_T) values were used to derive the relative transcript abundance using the comparative C_T method. Primers for the TEN and CON constructs, specific to their genetic differences, were designed as well as primers for pRL-TK to gauge transfection efficiency. In addition, *β-actin* was used as a housekeeping gene for this experiment. All primers were manufactured by IDT (Coralville, IA, USA).

Table 2.8 Primers used for quantifying relative mRNA abundance in the decay assay.

Primer Name	Primer Sequence (5'-3')
FWD for both ALLELIC FORMS	GGA AAG ATC GCC GTG TAA TTC
REV T-ALLELIC FORM	AGA AGC CTT CAC TGT CCA GGA
REV C-ALLELIC FORM	AGA AGC CTT CAC TGT CCA GGG
FWD pRL-TK	GGT AAC GCG GCC TCT TCT TA
REV pRL-TK	GCC AAA CAA GCA CCC CAA TC
FWD β-ACTIN	CGG CAT CGT CAC CAA CTG
REV β-ACTIN	AAC ATG ATC TGG GTC ATC TTC TC

2.13 Nuclear and Cytoplasmic Protein Harvest for an RNA Electrophoretic Mobility Shift Assay (EMSA)

A monolayer of HT1080 cells were split into five 10cm dishes and incubated to 90% confluency. The medium was aspirated off and the cells in each dish were washed twice with 1ml of 0.15M 1X Phosphate-buffered saline (PBS) at room temperature. The cells were harvested by scraping in 800μl PBS, pooled in a 12.5ml falcon tube and were spun for 4 minutes at 400 x *g*. The supernatant was discarded and the pellet was resuspended in 800μl PBS in a 1.5ml microfuge tube on ice. The samples were pulsed in a microfuge at room temperature and the supernatant discarded. The pellet was resuspended in 250μl solution I (20mM PIPES pH 6.8, 1mM EGTA pH 6.8, 1mM magnesium chloride ($MgCl_2$), 10X Protease Inhibitor Cocktail, 1mM Sodium Orthovanadate (Na_3VO_4), 1mM Sodium Fluoride (NaF)) and incubated on ice for 5

minutes. Triton X-100 was added to a final concentration of 0.5% and the cells were gently resuspended and incubated on ice for 5 minutes. Thereafter, the lysed cells were centrifuged at 900 x *g* at 4°C for 5 minutes.

The supernatant, containing the cytoplasmic proteins was collected in 100µl aliquots and stored at -80°C. The nuclear pellet was resuspended in 250µl of solution I and centrifuged at 900 x *g* at 4°C for 5 minutes. The supernatant was discarded and the pellet resuspended in 250µl of solution II (100mM KCl (potassium chloride), 300mM sucrose, 10mM PIPES pH6.8, 3mM $MgCl_2$, 1mM EGTA pH6.8, 10X Protease Inhibitor Cocktail, 1mM Na_3VO_4, 1mM NaF) and 10µl DNase was added to the extract before it was incubated at 30°C for 45 minutes, with occasional flicking. After incubation the extract was centrifuged at 1500 x *g* for 5 minutes at 4°C. The supernatant which contains the nuclear proteins was collected in 100µl aliquots and stored at -80°C.

The protein obtained was quantified using the Pierce bicinchoninic acid (BCA) Protein Assay Kit (Pierce Biotechnology, Rockford, USA). Solution I were used as diluents respectively and a protein standard curve was established from protein standards made up in their diluent using 2mg/ml BSA (bovine serum albumin) in 0.9% NaCl (Sodium chloride) . Proteins were diluted to concentrations ranging from 1:20 – 1:200. In a 96-well microplate (Greiner Bio-One GmbH, Frickenhausen, Germany) 25µl of each standard and protein sample was added per well in duplicate, as well as 200µl of working reagent (15ml of Pierce BCA Protein Assay Reagent A and 0.3ml of Pierce BCA Protein Assay Reagent B at a 50:1 ratio) following the

manufacturer's instructions. Readings were obtained at an OD of 595 nm on the RT-2100C Microplate Reader (Rayto, Shenzhen, China).

2.14 Non-radioactive RNA EMSA

A preliminary RNA EMSA was performed to determine if there are any interactions between RNA binding proteins (RBP) and putative sites on the C- and T-allelic forms of the *COL5A1* mRNA 3'-UTR in the *Xba*I and *Afl*II 57 bp region deleted in section 2.3. Single stranded biotin-labelled C- and T-allelic form RNA probes corresponding to nucleotides 993 to 1059 and 993 to 1063 of the *COL5A1* mRNA 3'-UTR were purchased from IDT (Coralville, IA, USA). Both probes were designed with two modifications; (1) a 2'O-Methyl RNA base was added to the 5' end of both probes to assist in protecting the probes from degradation and (2) a Biotin tag was added to the 3' end to bind to the streptavidin-HRP(horseradish peroxidase) conjugate to allow for detection. The oligonucleotide sequences are as follows:

C-allelic form: 5' – mUrUrU rUrUrU rUrUrA rArUrA rGrGrG rArArA rArArA rUrUrU rGrArU rArArU rUrUrU rCrU/ 3Bio/ - 3'
T-allelic form: 5' – mUrUrU rUrUrU rUrUrA rArUrA rGrGrG rArGrG rGrArA rArArA rArUrU rUrGrA rUrArA rUrUrU rUrCrU /3Bio/ - 3'

All apparatus was washed thoroughly with bleach and DEPC (Diethylpyrocarbonate) treated water prior to use to avoid RNase contamination. Between 4 and 100µg of protein lysate was added to 1X binding buffer (2M KCl and 1M $MgCl_2$) containing DTT at a final concentration of 10µM, 5% glycerol, 10mg/ml tRNA and 0.125µM of C-

or T-allelic biotin-labelled target RNA probes in a final volume of 18µl. The DTT assists in reducing the disulphide bonds of the protein and the tRNA acts as a non-specific competitor. The binding reactions took place for 30 minutes at room temperature. After incubation, 2µl of 10X loading dye (0.25% Bromophenol Blue, 0.25% Xylene Cyanol FF, 15% Ficoll) was added to each sample to a final volume of 20µl. The reaction mixture was loaded onto a pre-electrophoresed 1.5mm thick 8% native polyacrylamide gel (40% Acrylamide, 10X TBE, 10% APS, TEMED) and electrophoresed at 200 V in 0.5X TBE buffer for an hour on ice.

Following electrophoresis the protein samples were transferred to positively charged Hybond nylon membrane (Amersham, UK) at 380mA for 30min on ice. RNA was cross-linked on the membrane for 60 seconds using UV Stratalinker 1800 apparatus (Stratagene, CA, USA). The membrane was then blocked whilst shaking for 15 minutes in blocking buffer (3% BSA in 0.1% (v/v) Tween 20 in 1x PBS (PBS/T)). A 1:300 dilution of streptavidin-HRP conjugate was made in blocking buffer and the membrane was incubated in this solution for 15 minutes whilst shaking.

The membrane was subsequently washed 5X for 5 minutes on a shaker in 1X PBS/T washing buffer and thereafter incubated for 5 minutes in a 1X PBS equilibration buffer. To detect RNA-protein complexes the membrane was placed in 4ml of equal parts chemiluminescent detection reagent component 1 and chemiluminescent detection reagent component 2 purchased from Advansta (CA, USA), and the chemiluminescent signal was captured using Biospectrum Imaging system using Visionworks LS software (UVP, UK).

2.15 Participants for Genetic Association Study

The reporting of the following case-control genetic association study is in compliance with the recommendations drawn by the STREGA (STrengthening the REporting of Genetic Association studies) initiative, which is an extension of the STROBE (STrengthening the Reporting of OBservational Studies in Epidemiology) statement (Little et al., 2009).

Three hundred and forty-two asymptomatic (342) control participants (CON) and 160 with diagnosed chronic Achilles tendinopathy (TEN) were included in this study. The TEN and CON participants were recruited from South Africa and Australia as previously described (Mokone et al., 2005;Mokone et al., 2006;September et al., 2009;September et al., 2011). All participants were of self-reported European Caucasian ancestry.

All participants were given information about the study (Appendix A) and signed an informed consent form (Appendix A) prior to participation in this study, in accordance with the Declaration of Helsinki (World Medical Association, 2013). All participants completed questionnaires about their medical and sporting participation histories. Approval for the study was obtained from the Human Research Ethics Committee at the University of Cape Town and the Human Ethics Committee at La Trobe and Deakin Universities, Melbourne, Australia (Appendix A).

South African Participants

A total of 227 unrelated physically active participants, 149 asymptomatic individuals (CON) and 81 diagnosed with chronic Achilles tendinopathy (TEN) were included in this study. The participants were recruited from medical practices within Cape Town and Johannesburg in South Africa between 2004 and 2012. The TEN participants were all physically active prior to the onset of the condition. The clinical diagnosis, which was performed as described in Mokone et al (2006) (Mokone et al., 2006), of each participant was made and reviewed by experienced sports clinicians.

The stringent clinical diagnostic criteria for TEN included gradual progressive pain over the posterior lower limb in the Achilles tendon area for greater than 6 months, together with at least one out of the following six criteria: (1) early morning pain over the Achilles tendon area, (2) early morning stiffness over the Achilles tendon area, (3) a history of swelling over the Achilles tendon area, (4) tenderness to palpation over the Achilles tendon, (5) palpable nodular thickening over the affected Achilles, or (6) movement of the painful area in the Achilles tendon with plantar-dorsi-flexion (positive ‘‘shift’’ test) (Kader et al., 2002;Paavola et al., 2002;Schepsis et al., 2002). In addition to these criteria, soft-tissue ultrasound examination was performed in a sub-group of subjects to confirm the above diagnosis.

Physically healthy and active control subjects were recruited from various recreational sporting clubs. They had no history of chronic Achilles tendinopathy and were matched to the TEN participants for age at onset and country of birth.

Exclusion criteria included history of current or past fluoroquinolone antibiotic use or previous local corticosteroids injection in the Achilles tendon or the area surrounding the Achilles tendon prior to the onset of symptoms. Furthermore, participants with diagnosed connective tissue disorders or any other systemic diseases believed to be associated with TEN, such as, but not limited to, EDS, benign hypermobility joint syndrome, rheumatoid arthritis, systemic lupus erythematosus, hyperparathyroidism, renal insufficiency, diabetes mellitus and familial hypercholesterolemia were also excluded from the study.

Australian Participants

A total of 272 unrelated physically active participants, 193 asymptomatic individuals (CON) and 79 diagnosed with chronic Achilles tendinopathy (TEN), were included in this study. The participants were recruited at the Musculoskeletal Research Centre of La Trobe University, Melbourne, Australia from 2008 to 2010. The clinical diagnosis for the TEN participants was performed as described for the South African participants (Mokone et al., 2005) and confirmed by soft tissue ultrasound examination (September et al., 2009). In addition, apparently healthy participants, with no prior history of soft tissue pathology, were recruited from Melbourne, Australia. Again, the AUS CON subjects were matched for age of onset and country of birth to the AUS TEN group. Exclusion criteria were identical to those of the South African study group.

2.16 DNA Extraction for Genetic Association Study

Blood was collected from all participants by venipuncture of a forearm vein into an EDTA vacutainer tube. These samples were stored at 4˚C until DNA extraction. For the South African participants, total DNA was extracted from the collected blood using a method described by Lahiri and Nurnberger (1991) (Lahiri and Nurnberger, 1991) with modifications by Mokone et al. (2005) (Mokone et al., 2005) (Appendix C). In the case of the Australian participants, the DNA was extracted using the FlexiGene DNA kit (Qiagen, Valencia, CA, USA) as per the manufacturer's instructions. The extracted DNA was stored at -20˚C until needed.

2.17 *COL5A1* 3'-UTR Genotyping

Using the extracted DNA, the participants were genotyped for the rs71746744 (-/AGGG; 91 TEN and 198 CON), rs16399 (ATCT/-; 120 TEN and 254 CON) and rs1134170 (A/T; 107 TEN and 241 CON) polymorphisms within the 3'-UTR of the *COL5A1* gene. Genotyping was performed using custom-designed Fluorescence-based Taqman® PCR assays (Applied Biosystems, Foster City, CA, USA). Custom-made allele specific probes and flanking primer sets (Table 2.9) 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 two-step PCR consisted of a 10 mins heat activation step (95º C) followed by 40 cycles of 15 sec at 92º C and 1 min at 60º C using the XP Thermal Cycler, Block model XP-G (BIOER Technology, Tokyo, Japan). End-point fluorescence using a 7900 HT Fast Real-Time PCR System and the SDS Software version 2.3 (Applied

Biosystems, Foster City, CA, USA) was used to determine the genotypes of each polymorphism.

Table 2.9 Custom-made allele specific probes and flanking primer sets primers used in custom designed Fluorescence-based Taqman® PCR assays.

Primer Name	Primer Sequence (5'-3')	Reporter Dye	Reporter Sequence (5'-3')
rs1134170_F	CTT GTG GTG CTA TCT ATC TGT TTT AAG GT	VIC	TGG CCA GGG TCC CCA
rs1134170_R	GGA TAA AGA AAG CAG GGA GAA CGA	FAM	TGG CCA GGG ACC CCA
rs16399_F	TTC TCT CTT GTG GCT CTC TTG TG	VIC	CCT TAAA AAC AGA TAG ATA GC
rs16399_R	CCC AGT GCG CCT TCA AG	FAM	AGA CCT TAA AAC AGA TAG C
rs71746744_F	GCC CCT TTT CAG ACA GTT TTT GAT T	VIC	TTT CCC TCC CTA TTA AAA
rs71746744_R	CAA ACC TGT GTT TTA GTC TTA AGT GCA T	FAM	CAA ATT TTT TCC CTA TTA AAA

In addition, 310 (162 TT and 148 AT) of the 348 (89.1%) DNA samples were regenotyped for SNP rs1134170 using the Restriction Fragment Length Polymorphism (RFLP) method. Only two of the AT genotypes were reassigned as an AA genotype (0.6% genotyping error). The fragments containing rs1134170 were amplified using 5'- CAG AGC CTG ATG GGA GAA TGT CCA GGG CA -3' as the forward primer and 5'- GGA TAA AGA AAG CAG GGA GAA CGA GGC ATG ACC AG -3' as the reverse primer.

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 forward and reverse primers, 1X buffer (10 mM Tris-HCl pH 8.4, 50 mM KCl, 1.5 mM MgCl2, pH8.3), 200 µM of each dNTPs and 1 unit of Taq DNA polymerase (New England Biolabs, Ipswich, MA, USA). The cycling conditions included an initial denaturing step at 95°C for 2 mins followed by 30 cycles of denaturing at 95°C for 30 s, annealing at 65°C for 30 s, extension at 68°C for 1.5 mins and a final extension step at 68°C for 5 mins using the XP Thermal Cycler Block (Bioer Technology, Tokyo, Japan).

The PCR products were digested with the *PSh*AI restriction enzyme (New England Biolabs, Ipswich, MA, USA) following the manufacturers instructions, to produce 159-bp and 34-bp fragments for the A allele and an uncut 193-bp fragment for the T allele. The resultant products were separated on 6% nondenaturing polyacrylamide gels (PAGE) and visualized using SYBER Gold staining (Invitrogen Molecular Probes, Eugene, Oregon) (Figure 2.5). The gels were photographed under UV light using a Uvitec photodocumentation system (Uvitec Limited, Cambridge, UK), and the genotypes derived from the sizes of the DNA fragments obtained.

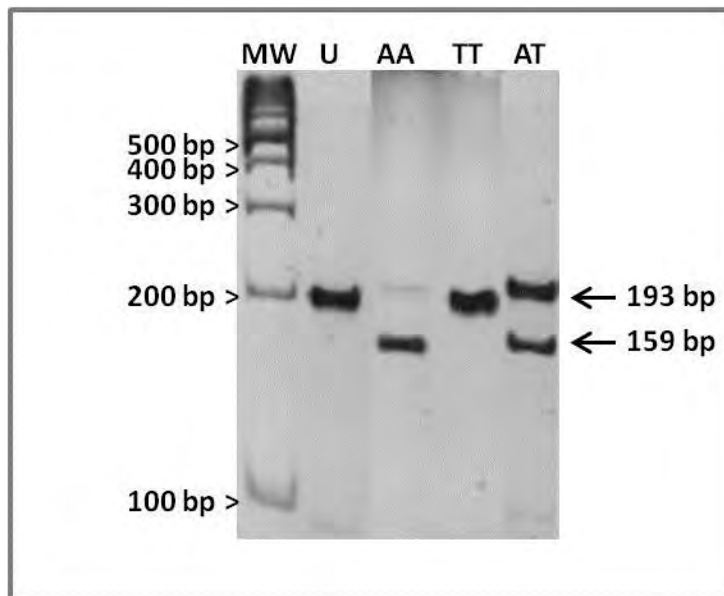


Figure 2.5 A typical image of a 6% PAGE gel discriminating restriction fragments from individuals who are homozygous AA (159 and 34 bp, lane 3), homozygous TT (193, lane 4) and heterozygous AT (193, 159 and 34 bp, lane 5) upon digestion of the 193 bp PCR products (Uncut: U, lane 2) of rs1134170 with the *PSh*AI restriction enzyme. The size of the fragments are indicated on the right while the genotypes are annotated on top. Fragment sizes of the 100bp DNA ladder (MW) (lane 1) are also indicated on the left. Electrophoresis was performed at 120 Volts for 2 hrs.

In addition, high resolution melting (HRM) analysis was performed for rs16399 (ATCT/-) by the Central Analytical Facility (University of Stellenbosch, Stellenbosch, South Africa). The DNA template was quantified using the NanoDrop ND1000 (NanoDrop Technologies, Wilmington, DE, USA) and adjusted to 5 ng/µl. The 20 ul PCR reactions were performed in the ABI Fast 96-well optical plates (Applied Biosystems, Foster City, CA, USA) which contained:1X ABI MeltDoctor HRM Master Mix (Applied Biosystems, Foster City, CA, USA), 6 pmol of each designed primer (Forward: 5'-CAC TTC TCT CTT GTG GCT C-3', Reverse: 5'-CAG TGC GCC TTC

AAG GAG AC-3'), 20 ng of DNA template a final volume of. The HRM-PCR was performed in the StepOne Real-time PCR System (Applied Biosystems, Foster City, CA, USA) with the following conditions: an activation step at 95°C for 10 mins followed by 40 cycles consisting of a denaturing step at 95°C for 15 sec and an annealing step at 60°C for 1 min. This was followed by a melt curve consisting of the sequential steps: a denaturing step at 95°C for 10 sec, an annealing step at 60°C for 1 min, a HRM step at 95°C for 15 sec (ramping rate of 1%) ending with an annealing step at 60°C for 15 sec. In each experiment, DNA samples of known genotypes (confirmed using Sanger sequencing) were included as reference standards.

Data collection and primary analysis including amplification plots were performed using the StepOne Software Version 2.2.1 (Applied Biosystems, Foster City, CA, USA). The high-resolution melt analysis was performed using the High Resolution Melt Software Version 3.0.1 (Applied Biosystems, Foster City, CA, USA). The genotypes of the samples were assigned automatically using the melt curve profiles. The pre-melt and post-melt regions were found between 70.9° C to 71.3° C and 78.0° C to 78.3° C, respectively. Aligned melt curves and difference plots were generated as well as silhouette scores for each sample. Any samples with low amplification or with outlier melt profiles were removed from the HRM analysis.

2.18 Preliminary Gene Expression Study using Skin Biopsy

Participants

Individuals of self-reported European Caucasian ancestry were invited to take part in a component of the study from 2013 to 2014. All participants were given information about the study (Appendix A) and signed an informed consent form (Appendix A) prior to participation in accordance with the Declaration of Helsinki (World Medical Association, 2013) and guidelines on good clinical laboratory practice (Ezzelle et al., 2008). Participants completed questionnaires about their medical and injury history (Appendix A). Approval for the study was obtained from the Human Research Ethics Committee at the University of Cape Town (Appendix A). Participants were included strictly with respect to their genotype at the rs12722 locus and exclusion criteria described in section 2.13 were applied. Following blood collection and DNA extraction, only individuals with the desired genotype were retained and invited to donate a skin biopsy sample (table 2.10). Three participants having the *COL5A1* *Bst*UI RFLP (rs12772) TT genotype ("at risk"), 3 asymptomatic active control participants having the CC genotype as well as an individual having the TC genotype at this locus were recruited.

Genotyping of Participants

Approximately 4.5ml of venous blood was collected from all participants by venipuncture of a forearm vein into an EDTA vacutainer tube. These samples were stored at 4˚C until DNA extraction which was performed as previously described

(Mokone et al., 2005), with minor modifications (section 2.16). Participants were genotyped for *COL5A1* rs12722 as previously described (Mokone et al., 2006) and two investigators independently confirmed the genotypes. Briefly, 667 bp fragments containing the *COL5A1 Bst*UI RFLP (rs12722) were amplified by PCR using the forward Primer, 5'- GAA GAC GTT TCT GGA GGA TC -3', and reverse primer, 5'-GGA GGC ACC TGC AGA ATG AC -3'.

The PCR reactions were performed in a total volume of 60 µl containing at least 100 ng of total genomic DNA, 20 pmol of each of the forward and reverse primers, 1X buffer (10 mM Tris-HCl pH 8.4, 50 mM KCl, 1.5 mM MgCl2, pH8.3), 200 µM of each dNTPs and 1 unit of Taq DNA polymerase (New England Biolabs, Ipswich, MA, USA). The cycling conditions included an initial denaturing step at 94°C for 3 min, followed by 35 cycles of denaturing at 94°C for 1 min, annealing at 53°C for 1 min, extension at 72°C for 1.5 min, and a final extension step at 72°C for 8 min using the XP Thermal Cycler Block (Bioer Technology, Tokyo, Japan).

The PCR products were digested with the *Bst*UI RE enzyme (New England Biolabs, Ipswich, MA, USA) to produce 351 and 316bp fragments for the T allele; 316, 271 and 80bp fragments for the C allele. The resultant fragments were separated, together with a 100bp DNA ladder, on 6% non-denaturing polyacrylamide gels and visualized by SYBER Gold staining (Invitrogen Molecular ProbesTM, Oregon, USA). The gels were photographed under UV light using a Uvitec photodocumentation system (Uvitec Limited, Cambridge, UK) and the sizes of the DNA fragments determined (Figure 2.6). In addition to the primary genetic marker *COL5A1* rs12722,

other polymorphisms important in distinguishing between the two the allelic forms of CON and TEN were also genotyped as described in section 2.17.

Table 2.10 Characteristics and genotype of participants who donated a skin biopsy.

Participant	rs12722	rs71746744	Age at Recruitment (Years)	Gender	Height (cm)	Weight (kg)	BMI (kg/m^2)
SB-1	TT	AGGG/AGGG	47	M	173.0	73.1	24.4
SB-2	TT	AGGG/AGGG	25	M	182.0	75.0	22.6
SB-3	TT	AGGG/AGGG	26	M	183	76	22.7
SB-4	TC	-/AGGG	26	F	174	86	28.4
SB-5	CC	-/AGGG	30	F	163	80	30.1
SB-6	CC	-/-	22	F	174.0	65.0	21.5
SB-7	CC	-/AGGG	24	M	179	96	30.0

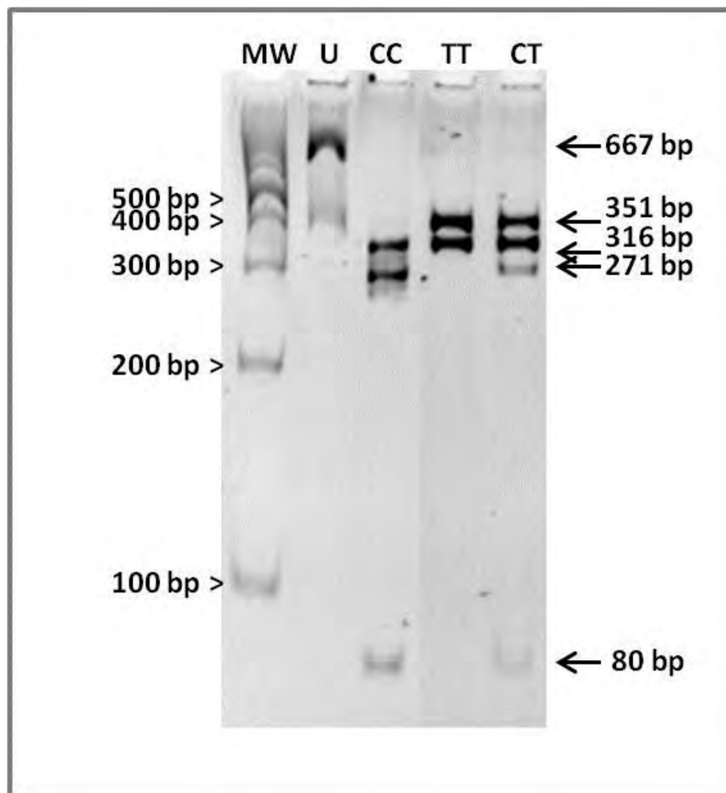


Figure 2.6 A typical image of a 6% PAGE gel discriminating restriction fragments from individuals who are homozygous CC (316, 271 and 80bp, lane 3), homozygous TT (351 and 316bp, lane 4) and heterozygous CT (351, 316, 271 and 80 bp, lane 5) upon digestion of the 667 bp PCR product (Uncut: U, lane 2) of rs12722 with the *Bst*UI restriction enzyme. The size of the fragments are indicated on the right while the genotypes are annotated on top. Fragment sizes of the 100bp DNA ladder (MW) (lane 1) are also indicated on the left. Electrophoresis was performed at 120 Volts for 2 hrs.

Skin Biopsy Procedure

All needle biopsies were performed in a sterile environment by an experienced medical practitioner. The area of the forearm to be biopsied was cleaned thoroughly with 70% isopropyl alcohol. A 23-gauged needle was used to obtain a 2 mm wide superficial skin sample from the epidermis and upper dermis layer, minimizing

scarring for the participant. The biopsy site was dressed and the obtained biopsies were transported in sterile vials containing 3ml of Dulbecco's Modified Eagle's medium (DMEM) (Gibco) supplemented with 10% foetal calf serum (Gibco), 200 units/ml penicillin and 100 µg/ml streptomycin (complete medium). Sterile and aseptic techniques were used at each stage. All biopsies were kept at 4°C and used within 24 hrs.

2.19 Derivation of Participant's Primary Fibroblast Cell Line from Skin Biopsy

Primary fibroblast cell cultures were established from the skin biopsies (Baumgarten, 1993). While working under a laminar flow hood, the dermis and epidermis were dissected from the rest of the skin and washed several times with sterile 1X PBS. Baumgarten's methodology was used to obtain outgrowth of fibroblast cells from the explants. Briefly, the tissue was finely sliced into approximately 1mm by 1mm pieces while covered in complete medium. The fragments were dispensed into several 35mm culture dishes and covered with a sterile coverslip placed on top of them to keep them in place, allowing their attachment to the dish's surface. An additional 2.5ml of complete medium was added and the plates placed in a 37°C humidified (95% air/ 5% CO_2, 65% humidity) incubator and the culture monitored for infections. The culture medium was changed twice weekly until enough cells had grown out of the explants.

As described by Baumgarten *et al*. (1993), once a reasonable amount of the fibroblast explants had migrated out of the explants, the coverslips were inverted and placed cell side up into new 35mm dishes. The original dishes were kept because they also contained our cells of interest. When the fibroblasts, outgrown from the explants, covered 70 to 90% of the culture area (2-3weeks) the medium was removed, the cells were washed twice with PBS, trypsinised (passage one) using 0.25% trypsin/EDTA solution and seeded into a 50ml culture flask. Mycoplasma testing was performed regularly between passages on cells grown in antibiotic-free medium containing 10% foetal calf serum. Subcultures involved expansion of the fibroblast culture into several flasks in complete medium. Stocks of fibroblasts from passage two and three were cryogenically preserved using 10% DMSO (Sigma-Aldrich, Munich, Germany) and stored in liquid nitrogen for further usage.

2.20 Expression levels of *COL5A1* and *COL1A1* in Primary skin fibroblast cell lines

Quantitative Real-Time Reverse Transcription PCR (Q-RT-PCR) was performed to look at expression of *COL5A1* and *COL1A1* in the primary fibroblast cells derived from participants with a defined genotype. The primary fibroblast cells at early passages (Passages 3-6) were seeded at 2 X 10^5 cells per ml in two 6 cm dishes per samples in 4 ml of complete DMEM (Highveld Biological, South Africa). The cells were incubated at 37°C (95% air, 5% CO_2, 65% humidity) overnight or until an 80% confluency was obtained. Thereafter, total RNA was extracted using the High Pure RNA isolation kit (Roche Diagnostics, Mannheim, Germany). The RNA was quantified and integrity assessed on a 1% agarose gel using the RiboRuler™ RNA

ladder and 2X RNA Loading Dye (Fermentas, thermo scientific, MA, USA) following the manufacturer's recommendations. It was kept at -80 °C until cDNA was obtained using the Improm-II™ Reverse Transcriptase kit (Promega Corporation, Madison, WI, USA). Finally the transcript abundance was quantified in real time using the Power SYBR green PCR master (Applied Biosystems, Foster City, CA, USA) and the PCR reactions were carried out in the Applied Biosystems StepOnePlus real-time PCR System (Life Technologies, Applied Biosystems, Foster City, CA, USA) following the manufacturer's recommended cycling conditions followed by a melting curve protocol. The threshold cycle (C_T) values were used to derive the relative transcript abundance using the comparative C_T method. SB-7 is used as a control sample and *β-actin* as the housekeeping gene. The primers used for this experiment, *COL5A1* and *COL1A1* as well as primers for *β-actin* as a housekeeping gene, have been used in the literature in a similar experiment (Luna et al., 2011) (Table 2.6). All primers were manufactured by IDT (Coralville, IA, USA)

Table 2.11 Primers extracted from Luna *et al.* (2011) to quantify relative expression of the *COL5A1* and *COL1A1* genes.

Primer Name	Primer Sequence (5'-3')
FWD COL1A1	AGC CAG CAG ATC GAC AAC A
REV COL1A1	TCT TGT CCT TGG GGT TCT T
FWD COL5A1	GGC TGT GCT ACC AAG AAA GG
REV COL5A1	GAG GTC ACG AGG TTG CTC T
FWD β-ACTIN	CCT CGC CTT TGC CGA TCC G
REV β-ACTIN	GCC GGA GCC GTT GTC GAC G

2.21 Statistics

Data were analysed using STATISTICA version 10.0 (StatSoft, Tulsa, OK, USA) and GraphPad Prism version 5.03 (GraphPad Software, San Diego, CA, USA, www.graphpad.com) programs. For the analysis of Luciferase assays and the expression studies, normally distributed data were analysed using an independent t-test or a one-way analysis of variance (ANOVA), followed by, if appropriate a Tukey HSD post-hoc analysis. The nonparametric Mann–Whitney U test was performed for data deviating from the normal distribution. Data sets are presented as mean ± standard deviation and significance was accepted as $p<0.05$.

For the genetic association study, a one-way analysis of variance was used to determine any significant differences between the characteristics of the TEN and CON groups within the AUS and SA cohorts. The Quanto V.1.2 software (http://hydra.usc.edu/gxe) was used to determine the statistical power of the sample size. Assuming allele frequencies between 0.2 and 0.8 for the "risk" allele of each SNP investigated, our sample size of 100 cases to 200 controls would be adequate to detect an allelic OR of 2.0 and greater at a power of 80% and a significance level of 5%. A Chi^2-analysis or Fisher's exact test was used to analyse any differences in the genotype frequencies and other categorical data between the groups. Statistically significant differences were accepted when $P<0.05$; and $P<0.025$ when combined gene-gene interactions and effects were analysed.

Combined genotype frequencies were also analysed using the Monte Carlo test (CLUMP program, version 2.0) (Sham and Curtis, 1995). Hardy-Weinberg equilibrium (HWE) was established using the program Genepop web version 3.4 (http://genepop.curtin.edu.au/). Linkage disequilibrium (LD) was calculated using CubeX: cubic exact solution (www.oege.org/software/cubex/) (Gaunt et al., 2007). Inferred haplotype analysis (Lin et al., 2008) was performed using the Hapstat case-control haplotype inference software. Statistically significant differences were accepted when $p<0.05$.

CHAPTER 3 RESULTS

Aspects of the data presented in this chapter have been published in the following peer-reviewed articles:

- **Mary-Jessica Laguette**, Yoonus Abrahams, Sharon Prince and Malcolm Collins. Sequence variants within the 3'-UTR of the *COL5A1* gene alter mRNA stability: Implications for musculoskeletal soft tissue injuries. *Matrix Biology*. 2011 Jun;30(5-6):338-45

- Yoonus Abrahams, **Mary-Jessica Laguette**, Sharon Prince and Malcolm Collins. Polymorphisms within the *COL5A1* 3'-UTR that alters mRNA structure and the MIR608 gene are associated with Achilles tendinopathy. *Ann Hum Genet.* 2013 May;77(3):204-14

3.1 Overall Increase in *COL5A1* mRNA Stability in the Tendinopathic Phenotype

As mentioned in Chapter 1, the C/T single nucleotide polymorphism (SNP) rs12722, also referred to as *Bst*UI RFLP, in the 3'-untranslated region (3'-UTR) of *COL5A1* is associated with chronic Achilles tendinopathy (TEN) in two independent populations. Furthermore, DNA was previously obtained from South African (SA) tendinopathic participants with the TT genotype ('at risk') for rs12722 and physically active asymptomatic individuals (CON) with a CC genotype ('protective'). To determine whether the *COL5A1* 3'-UTR was functional, the 2.5 kb *COL5A1* 3'-UTR was amplified from the DNA of four TEN (1, 2, 3 and 5) and three CON (2, 3 and 5) participants and cloned downstream of a luciferase reporter gene. Two separate clones were obtained for CON 2, CON 3, TEN 2 and TEN 3, while only a single clone was obtained for CON 5, TEN 1 and TEN 5.

These constructs were transiently co-transfected into HT1080 cells as described in section 2.9 and the relative normalised luciferase activity was used as an indirect indication of mRNA stability (Figure 3.1A). On average, the TEN clones showed a higher mRNA stability compared with the CON clones ($p<0.001$, Figure 3.1A). There was a 2.1 fold ($p<0.001$) difference in the average normalised luciferase activity of the construct with the lowest activity (CON 5.1: 49.4 ± 12.5%) when compared to the one with the highest activity (TEN 2.2: 104.2 ± 13.2%). It is worth noting that the luciferase activity of the CON 2.1 construct was more similar to the TEN constructs (Figure 3.1A). However, when all the TEN and all the CON constructs were pooled, there was, on average, a significantly ($p<0.001$) higher luciferase activity for the TEN

constructs (Figure 3.1B). These results suggest that the Achilles tendinopathic phenotype is associated with increased *COL5A1* mRNA stability.

3.2 Identification of Two Major Allelic Forms of the *COL5A1* 3'-UTR

To identify any potential sequence differences between the *COL5A1* 3'-UTRs cloned from the TEN and CON participants, the 2.5 kb 3'-UTR of all the clones were sequenced and analyzed. Only 9 polymorphisms were identified within the 3'-UTR (Table 3.1) which produced two major allelic forms termed the C- (which corresponds to the published wild type sequence) and T-allelic forms. Seven (rs13946, rs12722, rs3196378, rs71746744, rs16399, rs1134170 and rs3128575) of the 9 polymorphisms within the C- and T-allelic forms were tightly linked (Table 3.1, Figure 3.2). Importantly, with the exception of CON 3.1 and TEN 3.1, the C- and T-allelic forms, corresponded with the CON and TEN clones respectively (Table 3.1). Furthermore, when the luciferase activity of the individual clones were divided into the allelic groups based on sequence there was a significant increase ($p<0.001$) in the T-form allelic group (TEN 1.1, TEN 2.2, TEN 3.2 and TEN 5.1; 90.6 ± 13.7%) compared to the C-form allelic group (CON 2.1; CON 3.2 and CON 5.1; 69.0 ± 22.0%) (Figure 3.3).

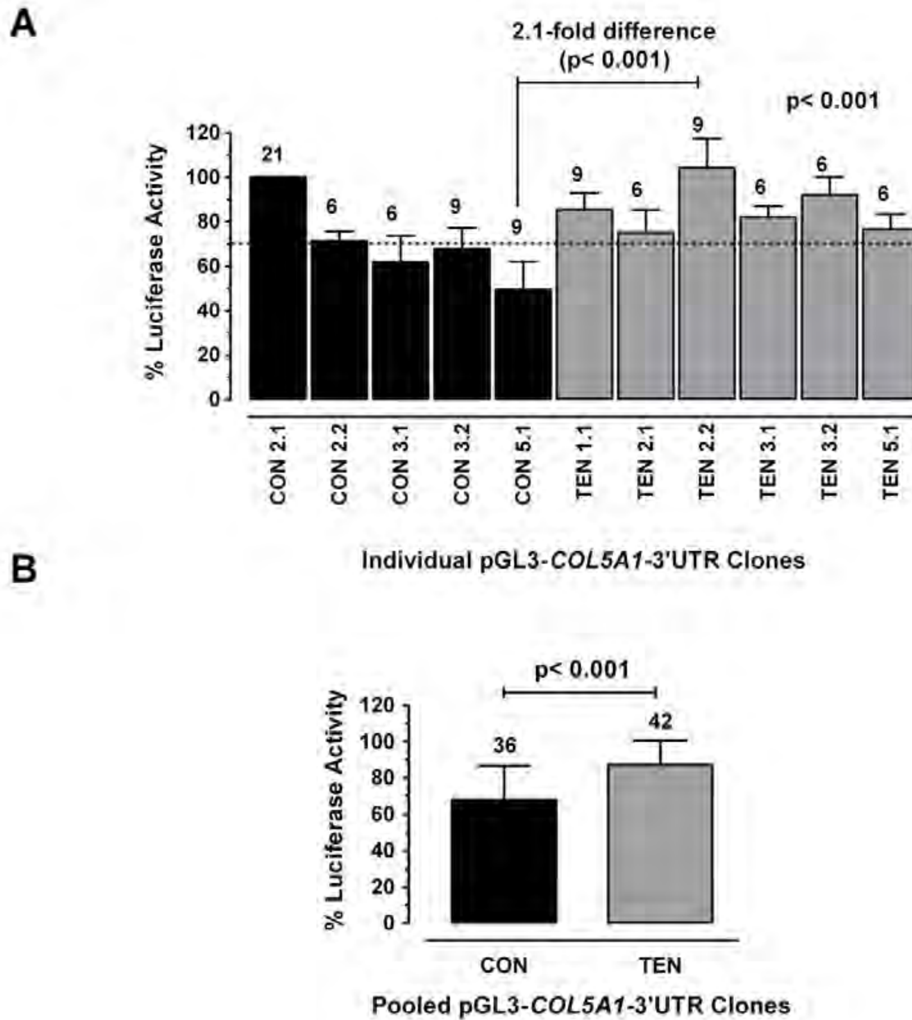


Figure 3.1 Functional analysis of the 2.5 kb *COL5A1* 3'-untranslated region (UTR) in HT1080 cells. (A) The relative luciferase activity of the individual control (CON, black bars) and tendinopathic (TEN, grey bars) clones are expressed as averages ± standard deviations, with the number of assays (n) indicated above each bar. Clones were obtained from 4 TEN (1, 2, 3 and 5) and 3 CON (2, 3 and 5) participants. Two separate clones were obtained from CON 2, CON 3, TEN 2 and TEN 3, while only a single clone was obtained from CON 5, TEN 1 and TEN 5. The constructs (500 ng) and 50 ng of pRL_TK were co-transfected into HT1080 fibroblast cells for 45hrs. Firefly luciferase activity for each clone was normalised with Renilla luciferase activity and expressed relative to CON 2.1. The global p-value as well as the p-values between the highest (TEN 2.2) and lowest (CON 5.1) is indicated. The dotted line represents the average activity of CON 2.2. (B) The relative luciferase activity of the pooled control (CON, black bars) and tendinopathy (TEN, grey bars) clones. Values are expressed as averages ± standard deviations, with the number of assays (n) indicated above each bar.

Table 3.1 Summary of the identified sequence differences within the *COL5A1* 3'-UTR region (2.5kb) of each clone.

POSITION	230	369	414	880	1000	1016	1237	1272	2648
POLYMORPHISM	**C/T**	**C**	**C/T**	**C/A**	**$(T)_n$**	**$(AGGG)_n$**	**$(ATCT)_n$**	**A/T**	**T/C**
SNP	rs13946	-	rs12722	rs3196378	rs71698207[b]	rs71746744[b]	rs16399	rs1134170	rs3128575[C]
RFLP	*Dpn*II	-	*Bst*UI	*Aci*I	-	-	-	-	-
CON 2.1	C	C	C	C	13	1	2	A	T
CON 2.2	C	T[f]	C	C	13	1	2	A	T
CON 3.1[d]	T	C	T	A	13	2	1	T	C
CON 3.2	C	C	C	C[a]	13	1	2	A	T
CON 5.1	C	C	C	C	13	1	2	A	T
TEN 1.1	T	C	T	A	13	2	1	T	C
TEN 2.1	T	C	T	A	12	2	1	T	C
TEN 2.2	T	C	T	A	13	2	1	T	C
TEN 3.1[e]	C	C	C	C	14	1	2	A	T
TEN 3.2	T	C	T	A	13	2	1	T	C
TEN 5.1	T	C	T	A	13	2	1	T	C
C- allelic form (W/T)	**C**		**C**	**C**		**1**	**2**	**A**	**T**
T- allelic form	**T**		**T**	**A**		**2**	**1**	**T**	**C**

Clones were obtained from 3 out of 5 CON participants (2, 3 and 5) and 4 out of 5 TEN participants (1, 2, 3 and 5). Two different clones were obtained from CON 2, CON 3, TEN 2 and TEN 3, while only a single clone was obtained from CON 5, TEN 1 and TEN 5. Position is the nucleotide number of the *COL5A1* exon 66 sequence; n, the number of T's or repeats; SNP, single nucleotide polymorphism; RFLP, restriction fragment length polymorphism. The C- or wild type (W/T) allelic form is the consensus sequence of the major variant isolated from CON constructs, while the T-allelic form is the consensus sequence of the major variant isolated from CON constructs. Deviations of the consensus sequences are highlighted in grey.

[a] Single nucleotide polymorphism (SNP) rs3196378 (*Aci*I RFLP) is an A in CON 3.1 with the 57 bp deletion within the *COL5A1* 3'-UTR.
[b] SNPs rs71698207, rs71746744 and rs11103544 (*Mbo*II RFLP, T in all the constructs), is deleted within the 57 bp deletion within the *COL5A1* 3'-UTR.
[c] SNPs rs3128575 and rs4504708 (G in all the constructs), is deleted within the 488 bp deletion within the *COL5A1* 3'-UTR.
[d] The genotype of participant from which CON 3.1 was cloned was CC and CC at SNPs rs13946 (*Dpn*II RFLP) and rs12722 (*Bst*UI RFLP) respectively.
[e] The genotype of participant from which TEN 3.1 was cloned was TT and TT at SNPs rs13946 (*Dpn*II RFLP) and rs12722 (*Bst*UI RFLP) respectively.
[f] Has not been reported in any public database. Since it was only identified in a single clone, we cannot exclude the posibility that this nucleotide substitution was introduced by the polymerase enzyme machinery during cloning. This substitution however does not appear to have any effect on mRNA stability (refer to CON 2.2 in figure 3.1(A)).

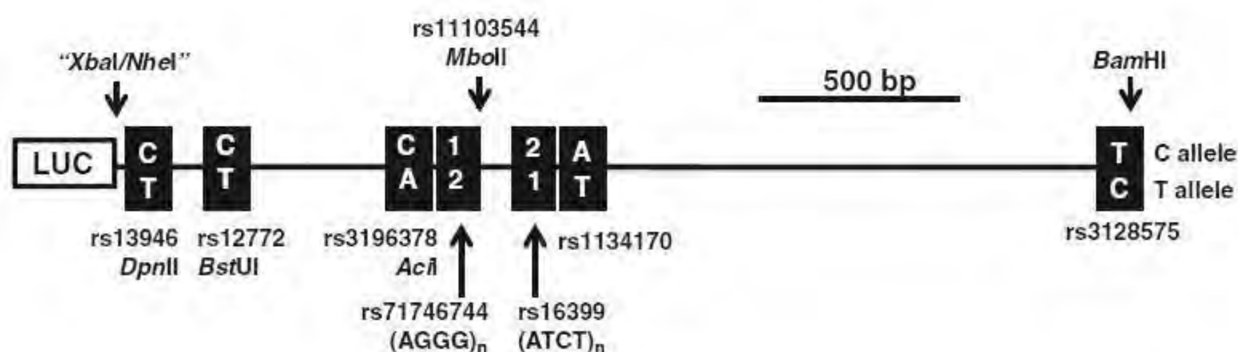


Figure 3.2 The seven tightly linked polymorphic sites within the *COL5A1* 3'-untranslated region (UTR) are annotated as black boxes. The accession numbers and/or restriction fragment length polymorphism (RFLP) associated with the polymorphic sites are indicated together with the nucleotide changes. The sequence of the seven polymorphisms for the C-(wild type) and T-allelic forms were C-C-C-$(AGGG)_1$-$(ATCT)_2$-A-T and T-T-A-$(AGGG)_2$-$(ATCT)_1$-T-C respectively. LUC, luciferase reporter gene.

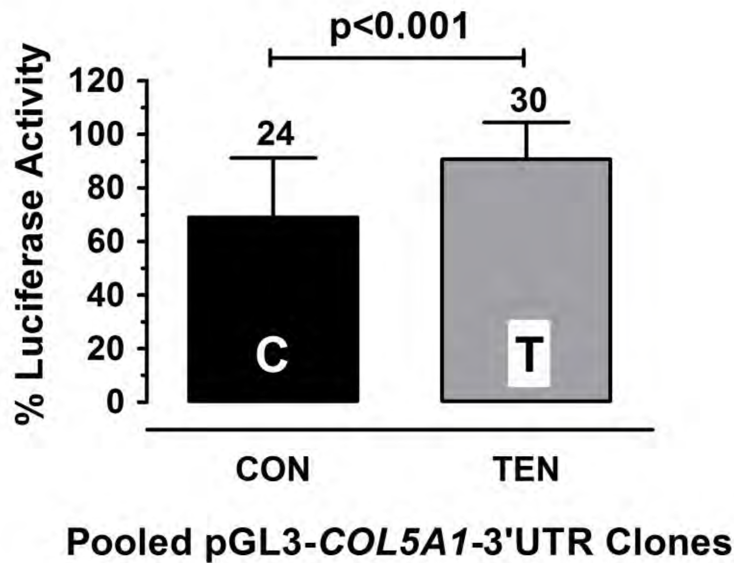


Figure 3.3 The relative luciferase activity of the pooled control (CON, black bars) and tendinopathy (TEN, grey bars) clones, containing the COL5A1 3'-UTR C- (CON 2.1, CON 3.2 and CON 5.1) or T- (TEN 1.1, TEN 2.2, TEN 3.2 and TEN 5.1) allelic form, in HT1080 cells. Values are expressed as averages ± standard deviations, with the number of assays (n) indicated above each bar. Only six CON 2.1 values were included when calculating the pooled CON luciferase activity to avoid skewing the results by a single clone.

Using a range of bioinformatic tools (section 2.8), two putative polymorphic miRNA binding sites, three putative polyadenylation signals and an expressed sequence tag (EST) within the *COL5A1* 3'-UTR were identified (Figure 3.4). The first exon of the EST (ENSESTG00000033016) with an unknown function is expressed in the opposite orientation to the *COL5A1* gene. Each putative miRNA binding site contains a single nucleotide polymorphism and is located within the region previously associated with the exercise-related phenotypes (Abrahams et al., 2013;Brown et al., 2011;Brown et al., 2011;Burger et al., 2014;Collins et al., 2009;Mokone et al., 2006;O'Connell et al., 2013;Posthumus et al., 2009;Posthumus et al., 2010;September et al., 2009). The miRNA target sites identified are CCACCCCA

and TTTTCTAC and they contain the polymorphisms (underlined) rs3196378 and rs11103544 respectively. While the core sequence of CCACCCCA is reported to bind Hsa-miR-608, the miRNA that binds TTTTCTAC is still unknown. The binding of Hsa-miR-608 to the C- and T-allelic forms of *COL5A1* 3'-UTR has previously been reported (Abrahams et al., 2013).

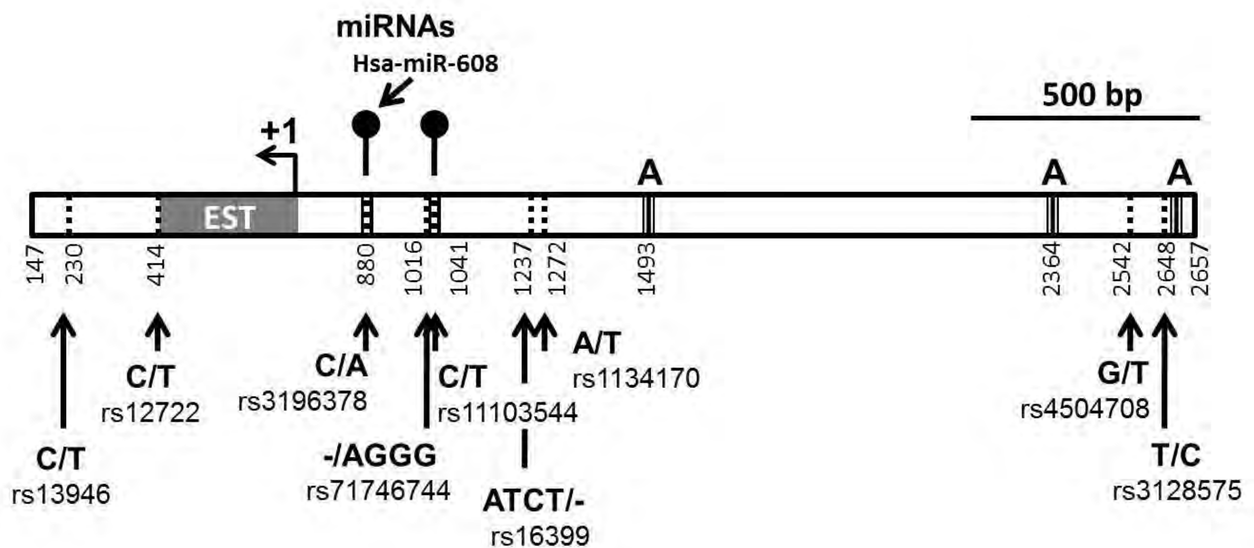


Figure 3.4 Schematic representation of the human *COL5A1* 3'-untranslated region (UTR). Key variants that have been previously investigated (dotted lines) (Abrahams et al., 2013;Mokone et al., 2006;September et al., 2009) and identified within the two functional forms, putative miRNA binding sites (black lines) present within the studied polymorphic sites, an expressed sequence tag (EST) (grey shading, ENSESTG00000033016, function unknown) as well as three putative polyadenylation signals (A, compound lines) are indicated on this diagram. Binding of Has-miR-608 to rs3196378 has been previously studied (Abrahams et al., 2013). The nucleotide number within exon 66 of the *COL5A1* gene and the transcription start site of the EST are indicated.

3.3 Identification of a Region of the *COL5A1* 3'-UTR that Confers mRNA Stability

Having identified putative regulatory elements within the *COL5A1* 3'-UTR, two deletion constructs, based on the availability of convenient restriction sites, were generated to test whether they do indeed contribute to *COL5A1* mRNA stability (Figure 3.5). The first construct (Δ488) lacks the last 488bp of the *COL5A1* 3'-UTR which contains the two distal polymorphisms (rs4504708 and rs3128575), not associated with chronic Achilles tendinopathy (September et al., 2009), and two terminal polyadenylation signal sites (Figure 3.5). A 57bp *Xba*I-*Afl*II fragment, containing the putative miRNA binding site harbouring the *Mbo*II RFLP (rs11103544) (September et al., 2009) as well as polymorphisms rs71698207 and rs71746744, were deleted from the second construct (Δ57).

When the Δ488 constructs were assayed, the average relative luciferase activity of TEN 1.1 (Δ488), which contains the T allele, was significantly higher (121.3 ± 22.5%, p<0.001) than the pooled CON constructs (CON 2.1 and CON 5.1) with a C allele (84.0 ± 14.2%) (Figure 3.6A). Although TEN 2.1 (Δ488) yielded similar results (p=0.007) (data not shown) it was not included in this analysis because it contains 12 instead of 13 T's at rs71698207 (Table 3.1). These results support the finding that the regulatory region associated with the tendinopathic phenotype may reside within the proximal 2kb of the 3'-UTR.

With the exception of CON 2.1, there was a significant increase in the luciferase activity when the Δ488 deletion constructs were compared to their parent full length 3'-UTR constructs (Figure 3.6B). This suggests that the terminal 488bp region of the *COL5A1* 3'-UTR contains novel regulatory elements which however do not contribute to the tendinopathic phenotype.

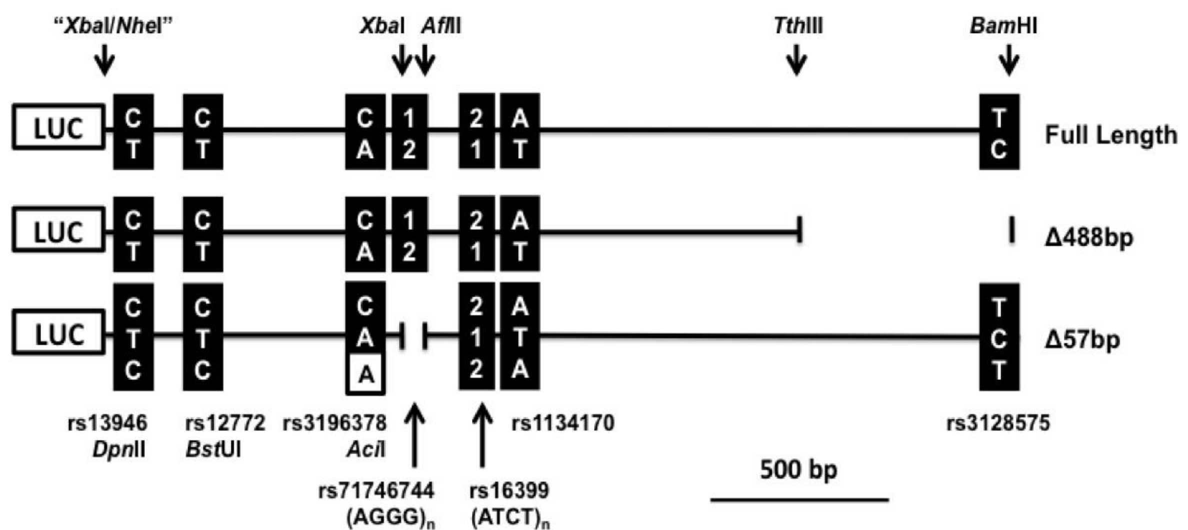


Figure 3.5 The *COL5A1* 3'-untranslated region (UTR) deletion constructs in HT1080 cells. Schematic representations of the full length *COL5A1* 3'- UTR cloned upstream of the firefly luciferase gene (LUC+), together with Δ488bp and Δ57bp deletion constructs. Key restriction enzymes used to generate these clones are indicated. The seven tightly linked polymorphic sites within the 3'-UTR are annotated as black boxes. The accession numbers and/or restriction fragment length polymorphism (RFLP) associated with the polymorphic sites are indicated together with the nucleotide changes. The sequence of the seven polymorphisms for the C (wild type) and T alleles were C-C-C-$(AGGG)_1$-$(ATCT)_2$-A-T and T-T-A-$(AGGG)_2$-$(ATCT)_1$-T-C, respectively. CON 3.2 (Δ57) contains an A (highlighted in the white box) instead of a C nucleotide at position 880, while the sequence of the other six polymorphic sites were correct.

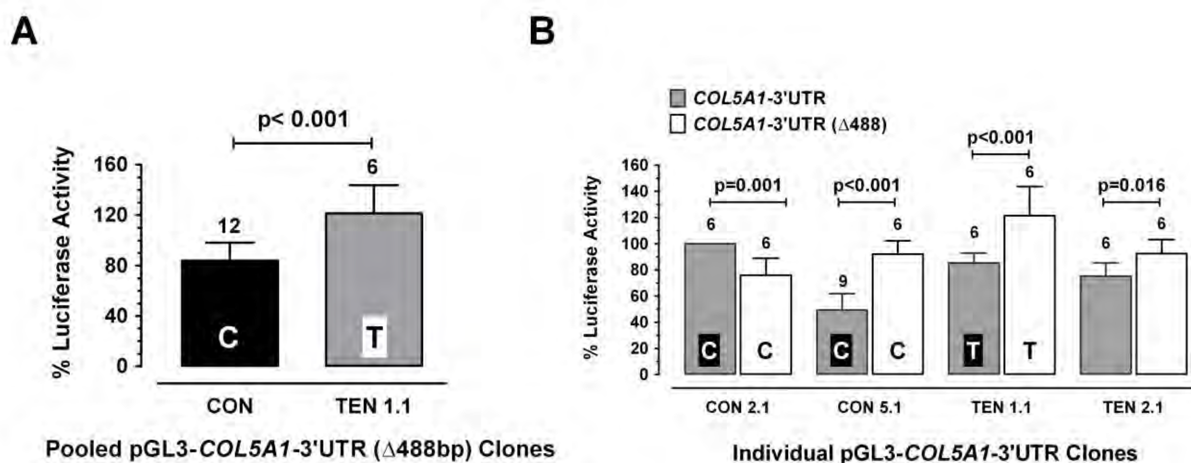


Figure 3.6 Functional analysis of the *COL5A1* 3'-untranslated region (UTR) Δ488 deletion constructs in HT1080 cells. (A) The relative luciferase activity of the pooled control (CON, black bars) and tendinopathy (TEN, grey bars) clones, containing the *COL5A1* 3'-UTR C- or T-alleles, respectively. (B) The relative luciferase activity of the individual full length (grey bars) and Δ488 (clear bars) clones. The CON constructs containing the C (wild type) *COL5A1* 3'-UTR are labelled with a "C", while the TEN constructs with the T allele are labelled with a "T". Values are expressed as averages ± standard deviations, with the number of assays (n) indicated above each bar. Firefly luciferase activity for each clone was normalised with renilla luciferase activity and expressed relative to full length CON 2.1.

Analysis with the Δ57 constructs revealed that there was no significant difference in the relative luciferase activity of CON 5.1 (Δ57), which contains the C allele, (144.2 ± 28.1% p=0.440) and the two pooled TEN deletion constructs (TEN 1.1 and TEN 2.1) with a T allele (156.5 ± 42.5%) (Figure 3.7A). CON 3.2 (Δ57) was not included in this analysis because it contains an A instead of a C nucleotide at position 880 which altered its luciferase activity (Figure 3.7B).

When the Δ57 constructs (CON 5.1, TEN 1.1 and TEN 2.1) were expressed relative to the activity of their respective full length 3'-UTR constructs, they exhibited a significant increase in luciferase activity (Figure 3.7B). These results indicate that the deleted TTTTCTAC binding site, for which there is currently no identified miRNA, may be functional but whether it contributes to the tendinopathic phenotype is as yet unclear.

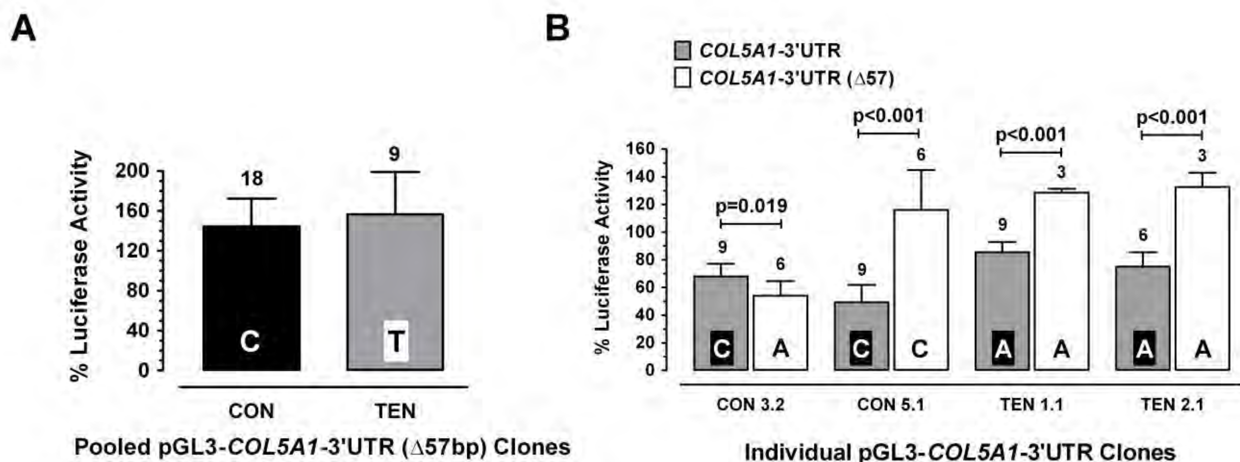


Figure 3.7 Functional analysis of the *COL5A1* 3'-untranslated region (UTR) Δ57 deletion constructs in HT1080 cells.(A) The relative luciferase activity of the pooled control (CON, black bars) and tendinopathy (TEN, grey bars) Δ57 clones, containing the *COL5A1* 3'-UTR C- (CON 5.1) or T- (TEN 1.1 and TEN 2.1) alleles, respectively. CON 3.2., which contains an A nucleotide at position 880, was not included in the analysis. (B) The relative luciferase activity of the individual full length (grey bars) and Δ57 (clear bars) clones. The nucleotide (C or A) at position 880 of the *COL5A1* 3'-UTR is indicated for each clone. Values are expressed as averages ± standard deviations, with the number of assays (n) indicated above each bar. Firefly luciferase activity for each clone was normalised with renilla luciferase activity and expressed relative to full length CON 2.1.

3.4 The SVWI-38 cells exhibit No Overall Increase in *COL5A1* mRNA Stability in the Tendinopathic Phenotype.

Having established the activity of the C- and T-allelic forms in the HT1080 cells, their activity was next examined in another fibroblast cell line. To this end, the constructs CON 2.1, CON 5.1, TEN 1.1 and TEN 2.2 were transiently co-transfected into the SVWI-38 SV40-transformed lung fibroblast cells (Section 2.9) with pRL-TK. Again, the normalized luciferase activity generated by each construct was expressed relative to CON 2.1, and the resulting activities of constructs representing the C- and T-allelic form were pooled as described before. There was no significant difference (p=0.135) in the luciferase activities of the two allelic forms (CON 72.3 ± 18.3% vs TEN 64.7 ± 12.8%) in the SVWI-38 cells (Figure 3.8).

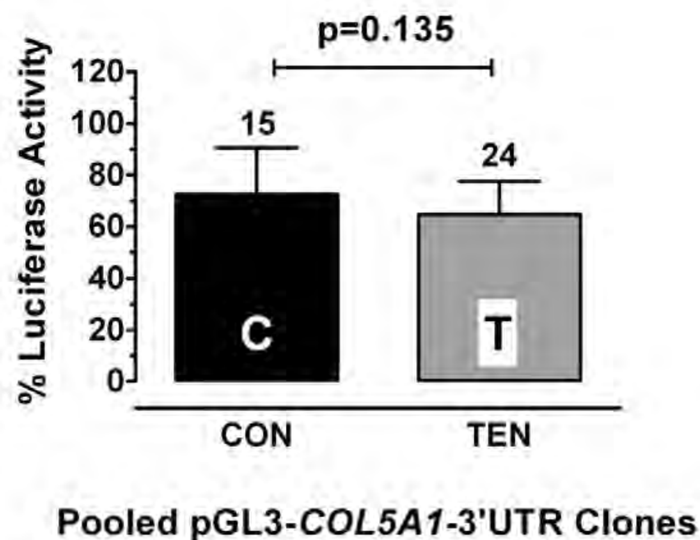


Figure 3.8 Functional analysis of the 2.5 kb *COL5A1* 3'-untranslated region (UTR) in SVWI-38 cells with respect to the C- and T-allelic forms. The graph shows the relative luciferase activity of all the control (CON, black bars) and tendinopathy (TEN, grey bars) clones, containing only *COL5A1* 3'-UTR C- or T-alleles. Values are expressed as averages ± standard deviations, with the number of assays (n) indicated above each bar.

3.5 The Newly Annotated Variants (rs71746744, rs16399 and rs1134170) are also Associated with Achilles Tendinopathy

As previously reported, two major allelic forms, namely the C- and T-alleles of the *COL5A1* 3'-UTR were identified, which show significant differences in mRNA stability (Section 3.2). In addition, deletion of a 57 bp region within the 3'-UTR containing the putative polymorphic miRNA binding site (rs11103544, *Mbo*II RFLP, T/C), as well as an AGGG short tandem repeat polymorphism (STRP rs71746744, -/AGGG) was sufficient to abolish the difference in mRNA stability between the C- and T-functional forms (Section 3.3). Variant rs11103544 was previously shown not to independently associate with Achilles tendinopathy (September et al., 2009). However, the associations of the variant rs71746744, the downstream rs16399 (ATCT/-) and rs1134170 (A/T) variants have not previously been investigated (Figure 3.9).

3.5.1 Characteristics of Participants

In order to examine the genetic association of these three variants with respect to soft tissue injuries, 342 control participants (CON) and 160 individuals with diagnosed chronic Achilles tendinopathy (TEN) from a South African (SA), as well as an Australian (AUS) population group, were included in this study. There were significantly more male participants within the combined SA and AUS TEN groups when compared to the combined CON groups (Table 3.2). When analysed separately, there were no significant differences (p=0.128) in gender between the SA TEN (72.5 %, N=80) and SA CON (62.4 %, N=171) groups. There was, however,

significantly more males ($p<0.001$) within the AUS TEN (73.0 %, N=63) group when compared to the AUS CON group (43.1 %, N=153).

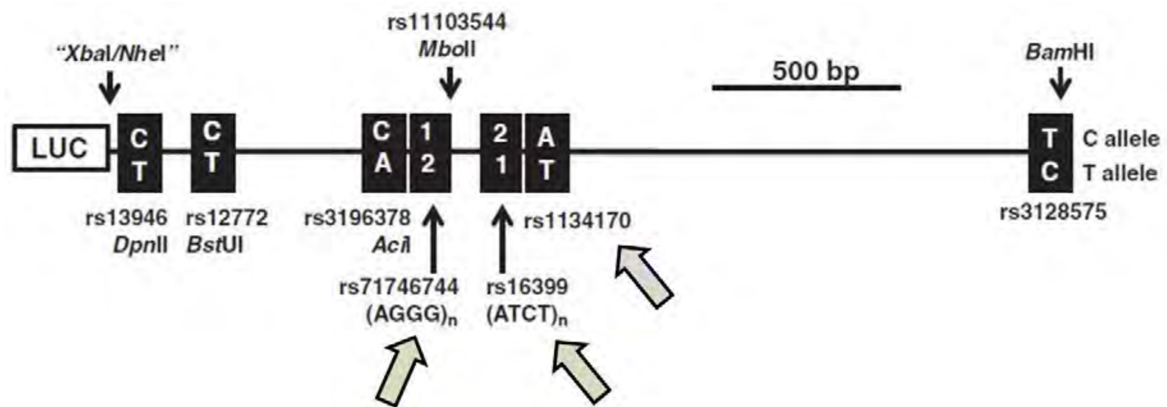


Figure 3.9 A genetic study investigating the variants rs71746744, rs16399 and rs1134170 (grey arrows). The seven tightly linked polymorphic sites within the 3'-UTR are annotated as black boxes. The accession numbers and/or restriction fragment length polymorphism (RFLP) associated with the polymorphic sites are indicated together with the nucleotide changes. The sequence of the seven polymorphisms for the C (wild type) and T alleles were C-C-C-$(AGGG)_1$-$(ATCT)_2$-A-T and T-T-A-$(AGGG)_2$-$(ATCT)_1$-T-C respectively.

Table 3.2 Descriptive characteristics of the combined **South African** and **Australian** Achilles tendinopathic (TEN) and control (CON) participants

	TEN Group (144)	**CON Group** (296)	**P-value**	**Co-varied P-value**
Gender (% Male)	72.7 (143)	52.4 (294)	<0.001	N.D.
Age (yrs) [a]	40.4 ± 14.5 (136)	37.6 ± 11.8 (286)	0.037	N.D.
Height (cm)	175.5 ± 8.8 (132)	173.0 ± 9.4 (290)	0.011	0.671 [b]
Weight (kg)	79.2 ± 14.2 (137)	72.8 ± 13.2 (294)	<0.001	0.057 [c]
BMI (kg/m^2)	25.6 ± 3.9 (132)	24.3 ± 3.6 (286)	<0.001	0.305 [c]

Except for gender, which is presented as a frequency, values are presented as mean ± standard deviations. The number of participants with non-missing data is in parenthesis.

[a] Age of the TEN participants is at the time of their initial injury, while the age of the CON participants is at the time of recruitment.

[b] co-varied for gender.

[c] co-varied for gender and age at recruitment.

yrs, years; kg, kilogram; cm, centimetre, N.D., not determined.

There were no significant differences between the age of the initial injury of the AUS TEN group (40.4 ± 14.7 yrs, N=62) and the age of recruitment of the AUS CON group (38.5 ± 12.6 yrs, N=151). The SA TEN group (age of the initial injury 40.4 ± 14.4 yrs, N=74) was however significantly older (P=0.036) than the SA CON group (36.6 ± 10.8 yrs, N=135). The combined SA and AUS TEN group (age at initial injury) was therefore significantly older than the combined CON group (Table 3.2). The combined TEN groups were recruited on average 9.2 ± 11.02 yrs (SA TEN 8.6 ± 10.74 yrs, N=74 and AUS TEN 9.8 ± 11.38 yrs, N=62) after their initial injury.

When co-varied for gender, there were no significant differences in height between the combined TEN and CON groups (Table 3.2). The SA groups were matched for height (TEN: 176.9 ± 8.7 cm, N=71 vs CON: 174.8 ± 9.5 cm, N=137, p=0.110) and similarly, there were no significant differences (adjusted p=0.152, unadjusted p=0.086) in height between the AUS groups (TEN: 173.8 ± 8.7 cm, N=61 vs CON: 171.5 ± 9.2 cm, N=153).

When adjusted for both age at recruitment and gender, there were no significant differences in the weight and BMI of the combined SA and AUS TEN and CON groups (Table 3.2). Similarly when adjusted for age at recruitment and gender, there were no significant differences (weight: adjusted p=0.318, unadjusted p<0.001 and BMI: p=0.371, unadjusted p=0.005) in the weight and BMI of the AUS TEN (weight: 80.6 ± 14.5 kg, N=63; BMI: 26.6 ± 4.1 kg/m^2, N=61) and CON (weight: 73.4 ± 14.2 kg, N=153; BMI: 24.9 ± 4.0, N=153) groups. Although adjusted for age at recruitment and gender, the SA TEN group (78.1 ± 14.0 kg, N=74) was significantly (adjusted p=0.034 and unadjusted p=0.001) heavier than the CON group (72.1 ± 12.1 kg, N=141). The SA groups were nevertheless matched for BMI (TEN: 24.7 ± 3.4, N=71 kg/m^2 vs CON: 23.6 ± 2.8 kg/m^2, N=133, adjusted p=0.194, unadjusted p=0.010). In terms of the country of birth, 72.1% (N=222) of participants in the SA groups reported that they were born in South Africa, and similarly, 78.1% (N=215) of the participants in the AUS groups were born in Australia.

As indicated in section 2.13, stringent clinical diagnostic criteria were used during the recruitment of participants by an experienced medical practitioner (Mokone et al., 2005;Mokone et al., 2006;September et al., 2009). None of the participants included

in this study had symptoms or signs of EDS (skin hyper extensibility, bruising, recurrent joint effusions, subluxations or dislocations, ocular manifestations or cardiovascular manifestations), hypermobility or benign hypermobility joint syndrome when their medical examinations were reviewed by the medical practitioner (Mokone et al., 2005;Mokone et al., 2006;September et al., 2009).

In the SA TEN participants (N=81) included in this study (i) 76.5% (N=62) were clinically diagnosed with tenderness to palpation, (ii) 58.0% (N=47) with early morning stiffness, (iii) 39.5% (N=31) with a history of swelling, (iv) 35.8% (N=28) with early morning pain, (v) 29.6% (N=23) with palpable thickening, (vi) 23.5% (N=19) tested positive for the ‘‘shift’’ test, and (vii) 93.8% (N=76) had a gradual progressive pain profile. Table 3.3 displays the frequencies of individuals having only one or more of the symptoms listed above. In addition, the diagnosis was confirmed in a subset of the TEN participants (39.5%) by soft-tissue ultrasound examination of the affected Achilles tendon. Among the TEN participants, 41.1% (N=73) had a confirmed bilateral Achilles tendinopathy while 24% (N=18) reported multiple injuries to the tendon. To confirm the pathology and clinical diagnosis of the AUS TEN group, all the subjects were examined using a soft-tissue ultrasound of the affected Achilles tendon (September et al., 2009). Together, 93.8% (N=64) were symptomatic where 79.7% was either of a gradual onset or acute (20.3%). Within the pathology group, 53.1% (N=64) were diagnosed with bilateral chronic Achilles tendinopathy and 40.4% (N=47) reported multiple injuries to the tendon.

Table 3.3 Collection of symptoms present in the **South African** Achilles tendinopathic (TEN) participants where the list includes (i) tenderness to palpation, (ii) early morning stiffness, (iii) a history of swelling, (iv) early morning pain, (v) palpable thickening and (vi) a positive ''shift'' test.

Number of clinical symptoms present	TEN Group (81)
All Six	13.6 (11)
Five	6.2 (5)
Four	7.4 (6)
Three	13.6 (11)
Two	25.9 (21)
One	27.2 (22)
None*	6.2 (5)

*Tendinopathic participants with missing clinical diagnostic data but whose diagnosis was confirmed by soft-tissue ultrasound examination.
Values are presented as a frequency (%), whereas the number of symptomatic TEN participants is in parenthesis.

Running was the predominant (68.1%, N=69) sporting activity resulting in injury in the SA cohort and the groups were matched for the mean number of years participating in running (8.5 ± 8.1 yrs, N=107, CON versus TEN 8.6 ± 10.0 yrs, N=65, p=0.972). However, there was a significant difference in hours of training between the two SA groups (p=0.003, 3.6 ± 3.0 hrs/week (N=103) CON vs. TEN 2.1 ± 2.6 hrs/week (N=57)), where the SA CON group trained for more hours per week. In addition, while the SA TEN group participated in more high impact sports compared to the CON group in the past ((p<0.001), 17.0 ± 16.5 yrs (N=107) CON vs. TEN 32.2 ± 29.7 yrs (N=65)), there was no difference between the two SA groups in the high

impact training performed over the past two years. Although all AUS participants were physically active individuals, the type of sporting activity that they were involved in and the frequency of this activity as well as their hours of training were not recorded.

3.5.2 *COL5A1* 3'-UTR Genotype Frequencies

With the exception of significant differences between the weight and BMI within the three rs16399 (ATCT/-) genotype groups, there were no significant genetic interactions with any of the other physiological variables (age, height, and gender) for any of the three *COL5A1* 3'-UTR variants (Tables 3.4 to 3.6). Participants with an ATCT/ATCT genotype were significantly heavier, with a corresponding larger BMI, than those with ATCT/- or a -/- genotypes (Table 3.5). Similar results were obtained when the pathology and country groups were analyzed separately, except for the AUS TEN individuals with the -/- genotype being taller than the ones with the ATCT/- .genotype (p=0.03) (Appendix D).

Table 3.4 Physiological characteristics of the three **rs71746744** (-/AGGG) genotype groups of the combined Australian and South African participants

	-/-	AGGG/-	AGGG/AGGG	P-value
Gender (% Male)	60.71 (28)	51.28 (117)	60.99 (141)	0.268
Age (yrs) [a]	38.7 ± 12.6 (28)	38.5 ± 13.1 (113)	38.8 ± 12.6 (137)	0.973
Height (cm)	173.8 ± 7.3 (27)	173 ± 10.0 (115)	174.2 ± 9.1 (133)	0.621
Weight (kg)	76.5 ± 17.4 (27)	73.3 ± 14.4 (116)	75.8 ± 12.8 (139)	0.306
BMI (kg/m^2)	25.1 ± 4.4 (27)	24.4 ± 4.1 (113)	24.9 ± 3.5 (133)	0.530

Except for gender, which is presented as a frequency, values are presented as mean ± standard deviations. The number of participants with non-missing data is in parenthesis.
[a] Age at recruitment.
yrs, years; kg, kilogram; cm, centimetre.

Table 3.5 Physiological characteristics of the three **rs16399** (ATCT/-) genotype groups of the combined Australian and South African participants

	ATCT/ATCT	ATCT/-	-/-	P-value
Gender (% Male)	64.9 (37)	54.54 (154)	64.44 (180)	0.154
Age (yrs) [a]	39.6 ± 13.5 (37)	37.9 ± 12.7 (147)	38.1 ± 12.1 (173)	0.774
Height (cm)	172.9 ± 9.2 (36)	173.5 ± 9.6 (149)	174.7 ± 9.2 (173)	0.410
Weight (kg)	81.0 ± 18.6 (36)	72.3 ± 12.5 (151)	75.4 ± 13.5 (178)	0.002
BMI (kg/m^2)	27.0 ± 6.1 (35)	24.0 ± 3.2 (147)	24.6 ± 3.4 (172)	<0.001

Except for gender, which is presented as a frequency, values are presented as mean ± standard deviations. The number of participants with non-missing data is in parenthesis.
[a] Age at recruitment.
yrs, years; kg, kilogram; cm, centimetre.

Table 3.6 Physiological characteristics of the three **rs1134170** (A/T) genotype groups of the combined Australian and South African participants

	AA	AT	TT	P-value
Gender (% Male)	50.00 (4)	55.25 (181)	62.89 (159)	0.338
Age (yrs) [a]	38.8 ± 13.5 (4)	38.4 ± 12.5 (154)	38.8 ± 13.8 (152)	0.971
Height (cm)	170.5 ± 10.9 (4)	173.1 ± 9.4 (175)	174.9 ± 9.2 (152)	0.164
Weight (kg)	75.0 ± 12.5 (4)	74.1 ± 13.8 (179)	75.6 ± 13.8 (156)	0.636
BMI (kg/m^2)	25.7 ± 2.7 (4)	24.7± 4.1 (172)	24.6 ± 3.3 (151)	0.815

Except for gender, which is presented as a frequency, values are presented as mean ± standard deviations. The number of participants with non-missing data is in parenthesis.

[a] Age at recruitment.

yrs, years; kg, kilogram; cm, centimetre.

The genotype distributions of rs71746744, rs16399 and rs1134170 were similar within the SA and AUS cohorts (Table 3.8) and were therefore combined for further analysis. As illustrated by figure 3.10, the AGGG/AGGG (p=0.008, odds ratio (OR)=2.0, 95% confidence interval (CI)=1.2-3.3), -/- ATCT (p=0.015, OR=1.7, 95% CI=1.1- 2.7) and TT (p=0.011, OR=1.8, 95% CI=1.2-2.9) genotype frequencies of rs71746744, rs16399 and rs1134170 respectively were significantly over-represented in the combined Achilles tendinopathy cohorts. Except for rs1134170, the other two polymorphisms were in Hardy-Weinberg equilibrium (HWE). All three polymorphisms were in linkage disequilibrium (LD) (D' ≥ 0.87) (Table 3.7).

Table 3.7 The linkage disequilibrium (LD) between the three newly annotated variants rs71746744, rs16399 and rs1134170 as well as with respect to the reported rs12722 variant within the combined Australian and South African participants.

	rs12722	**rs71746744**	**rs16399**
rs71746744	**0.77** (0.83)	**-**	**-**
rs16399	**0.73** (0.68)	**0.87** (0.91)	**-**
rs1134170	**0.55** (0.68)	**0.90** (0.89)	**0.93** (0.93)

The *D′* value between the pairs of variants is in bold for the control (CON) group while in brackets for the chronic Achilles tendinopathy (TEN) group.

Table 3.8 Genotype frequency distributions the *COL5A1* 3'-untranslated region (UTR) polymorphisms, rs71746744 (-/AGGG), rs16399 (ATCT/-) and rs1134170 (A/T), in control (CON) and chronic Achilles tendinopathy (TEN) groups of South African (SA) and Australian (AUS) cohorts.

		SA		AUS	
		CON	**TEN**	**CON**	**TEN**
rs71746744		**N=101**	**N=51**	**N=97**	**N=40**
	-/-	8.9 (9)	2.0 (1)	15.5 (15)	7.5 (3)
	-/AGGG	48.5 (49)	37.3 (19)	38.1 (37)	30.0 (12)
	AGGG/AGGG	42.6 (43)	60.8 (31)	46.4 (45)	62.5 (25)
	HWE	**0.503**	**0.666**	**0.124**	**0.374**
	P-value [a]	**0.040**		**0.095**	
rs16399		**N=133**	**N=68**	**N=121**	**N=52**
	ATCT/ATCT	9.0 (12)	4.4 (3)	16.5 (20)	3.9 (2)
	ATCT/-	46.6 (62)	41.2 (28)	38.8 (47)	32.7 (17)
	-/-	44.4 (59)	54.4 (37)	44.6 (54)	63.5 (33)
	HWE	**0.552**	**0.536**	**0.114**	**1.000**
	P-value [b]	**0.183**		**0.031**	
rs1134170		**N=131**	**N=66**	**N=110**	**N=41**
	AA	0.0 (0)	0.0 (0)	3.6 (4)	0.0 (0)
	AT	56.5 (74)	40.9 (27)	56.4 (62)	46.3 (19)
	TT	43.5 (57)	59.1 (39)	40.0 (44)	53.7 (22)
	HWE	**<0.001**	**0.056**	**0.002**	**0.087**
	P-value [c]	**0.050**		**0.144**	

Genotypes are expressed as percentages with numbers (N) in parenthesis.
HWE are exact P-values from tests of Hardy-Weinberg equilibrium.
[a] AGGG/AGGG genotype vs. the combined -/AGGG and -/- genotypes
[b] -/- ATCT genotype vs. the combined ATCT/- and ATCT/ATCT genotypes
[c] TT genotype vs. the combined AT and TT genotypes

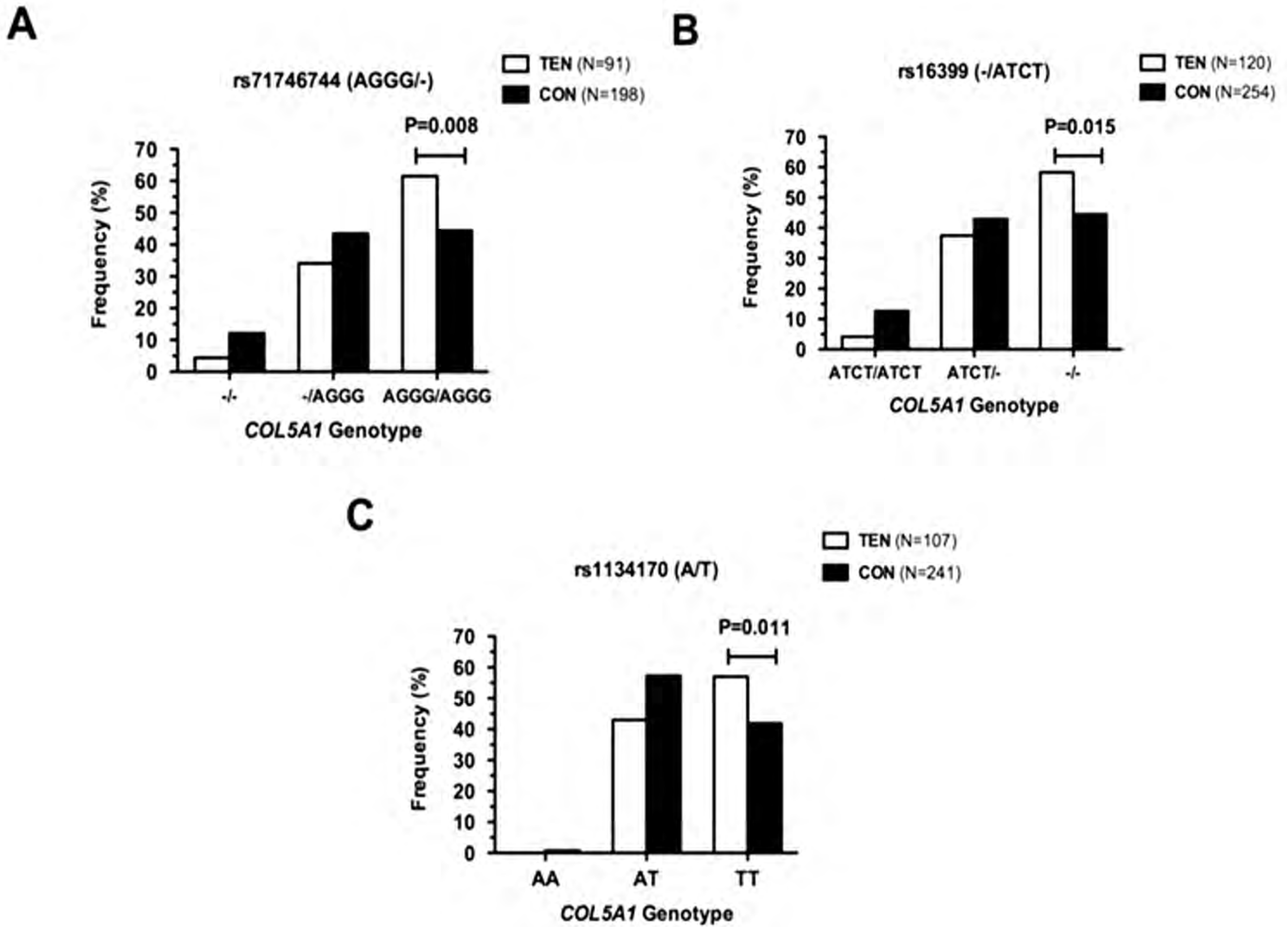


Figure 3.10 Genotype frequency distributions for the *COL5A1* 3'-untranslated region (UTR) polymorphisms, (A) rs71746744 (-/AGGG), (B) rs16399 (ATCT/-) and (C) rs1134170 (A/T), in control (CON) and chronic Achilles tendinopathy (TEN) for the combined SA and AUS (SA+AUS) cohorts. The genotype frequency for the CON and TEN groups are denoted by black and white bars respectively. The P-value for each (A) AGGG/AGGG genotype vs. the combined -/AGGG and -/- genotypes, (B) -/- ATCT genotype vs. the combined ATCT/- and ATCT/ATCT genotypes, (C) TT genotype vs. the combined AT and TT genotypes is displayed on the graph. The total number of participants (N) per group is indicated.

3.5.3 Inferred Haplotype Study

As described in section 3.2 two major functional forms of the *COL5A1* 3'-UTR, namely the C- and T-forms were identified after sequencing the 3'-UTR from CON and TEN participants respectively. The distinct C- and T-functional forms were determined by seven polymorphisms, namely rs13946 (C/T), rs12722 (C/T), rs3196378 (C/A), rs71746744 (-/AGGG), rs16399 (ATCT/-), rs1134170 (A/T) and rs3128575 (T/C) (Figure 3.2). The nucleotide sequence associated with the C-form and T-forms were C-C-C-$(AGGG)_1$-$(ATCT)_2$-A-T and T-T-A-$(AGGG)_2$-$(ATCT)_1$-T-C, respectively. With the exception of rs3128575, which had previously been genotyped in only the SA cohort (September et al., 2009), rs13946, rs12722, and rs3196378 have previously been genotyped in most of the participants within the AUS and SA cohorts (September et al., 2009). In order to confirm the sequencing results, inferred haplotypes were constructed with six of the seven polymorphisms that distinguished the two forms of the 3'-UTR.

Eight of the possible 36 inferred haplotypes had a frequency >1% (Figure 3.11). In agreement with the sequencing data (section 3.2), the T-T-A-$(AGGG)_2$-$(ATCT)_1$-T haplotype was significantly over-represented (p=0.003) in the combined SA and AUS TEN group (42.4%,N=76) when compared to the CON group (39.7%, N=150) (Figure 3.11). In addition, the T-C-C-$(AGGG)_2$-$(ATCT)_1$-T haplotype was also significantly over-represented (p=0.015) in the TEN group (15.7%, N=28) compared with the CON group (11.1%, N=42), mapping the associated region to rs71746744 (-/AGGG), rs16399 (ATCT/-) and rs1134170 (A/T). Four of the possible 8 inferred haplotypes constructed from rs71746744 (-/AGGG), rs16399 (ATCT/-) and rs1134170 (A/T) had

a frequency >1% (Figure 3.12). The AGGG)$_2$-(ATCT)$_1$-T haplotype was significantly over-represented (p=0.007) in the combined SA and AUS TEN group (72.2%,N=130) when compared to the CON group (62.5%, N=236).

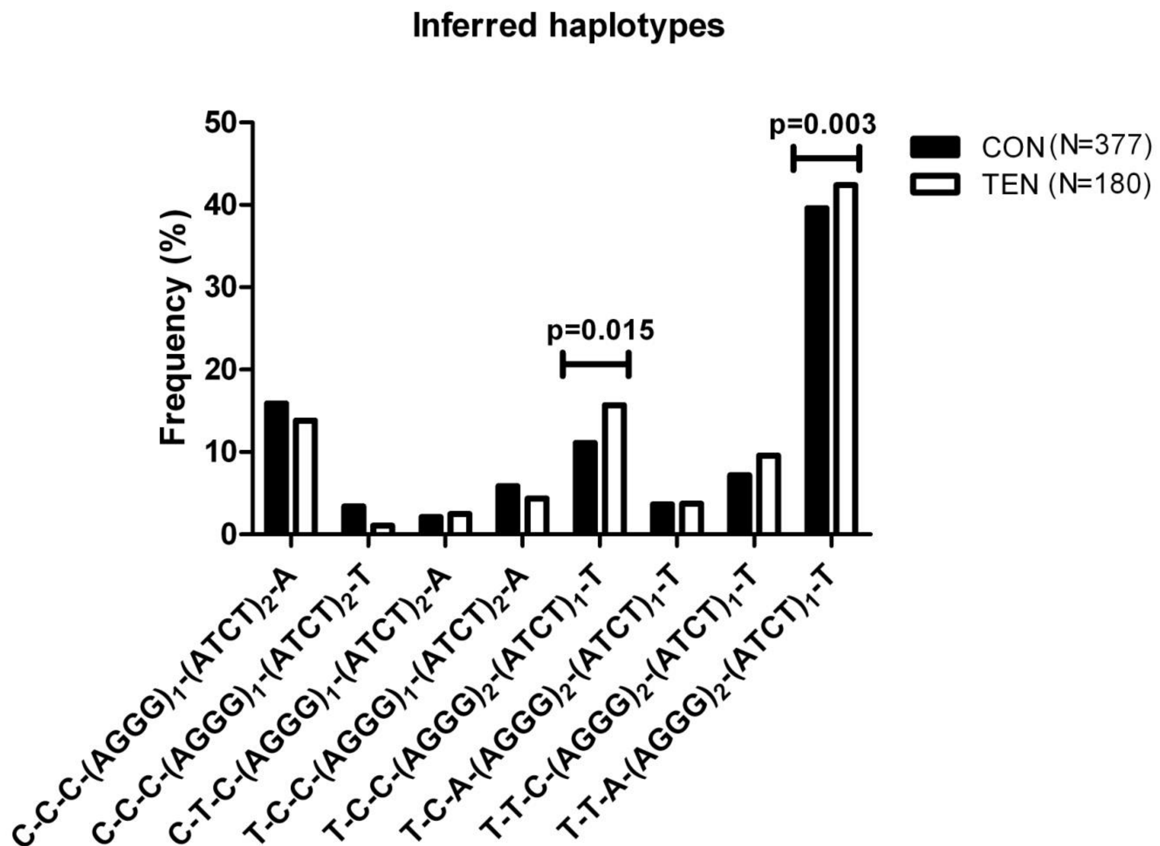


Figure 3.11 The frequency distribution of 8 of the possible 36 inferred haplotypes with a frequency >1% constructed from *COL5A1* 3'-UTR variants rs13946 (C/T), rs12722 (C/T), rs3196378 (C/A), rs71746744 (-/AGGG), rs16399 (ATCT/-) and rs1134170 (A/T) in the combined Australian and South African Achilles tendinopathy (TEN, white bars) and control (CON, black bars) groups. The P-value of the significantly different haplotypes and the total number (N) of haplotypes within the CON and TEN groups are indicated on the graph.

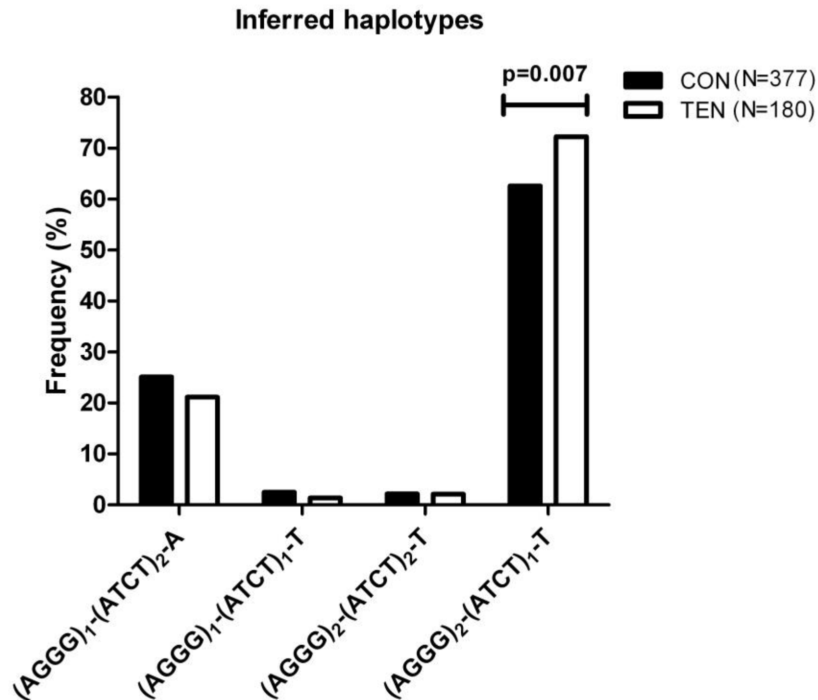


Figure 3.12 The frequency distribution of 4 of the possible 8 inferred haplotypes with a frequency >1% constructed from *COL5A1* 3'-UTR variants rs71746744 (-/AGGG), rs16399 (ATCT/-) and rs1134170 (A/T) in the combined Australian and South African Achilles tendinopathy (TEN, white bars) and control (CON, black bars) groups. The P-value of the significantly different haplotypes and the total number (N) of haplotypes within the CON and TEN groups are indicated on the graph.

3.5.4 Gene-gene Interactions

As previously mentioned in section 3.2, the *COL5A1* 3'-UTR contains a functional polymorphic (rs3196378, C/A) Hsa-miR-608 miRNA binding site (Figure 3.4) (Abrahams et al., 2013). Two forms of the mature Hsa-miR-608 are produced from the polymorphic (SNP rs4919510, C/G) *MIR608* gene on chromosome 10q24 (Abrahams et al., 2013). Although it was previously reported that there is no genotype-genotype association between these two SNPs and chronic Achilles tendinopathy, the CC genotype of *MIR608* was significantly over-represented in the

TEN group (Abrahams et al., 2013). Since, the CC *MIR608* rs4919510, AGGG/AGGG *COL5A1* rs71746744, -/- ATCT *COL5A1* rs16399 and TT *COL5A1* rs1134170 genotypes were all independently associated with increased risk of chronic Achilles tendinopathy, gene-gene interactions were further investigated.

Each of the four individual “at risk” genotypes contributed a score of 2 towards a participants “genotype risk score”. Participants with all four individual Achilles tendinopathy “at risk” genotypes had a total score of 8, while those with none of the risk genotypes had a score of 0. When compared to the CON group, the “genotype risk score” of 8 was significantly over-represented within the TEN group (p=0.004; odds ratio=2.6; 95% confidence interval 1.3 to 4.9), while the “genotype risk score” of 0 was significantly under-represented (p=0.019; odds ratio=3.1; 95% confidence interval 1.2 to 8.5) (Figure 3.13). When the CC *MIR608* rs4919510 “at risk” genotype was excluded from the analyses, the three *COL5A1* 3’-UTR “at risk” genotypes were also significantly over-represented (p=0.006; odds ratio=2.3; 95% confidence interval 1.3 to 4.3) within the TEN participants (60.0%, N=36 of 60) when compared to the CON participants (39.4%, N=61 of 155) (Figure 3.13). In contrast, participants with none of the three “at risk” *COL5A1* genotypes were significantly over-represented (P=0.002; odds ratio=2.7; 95% confidence interval 1.4 to 5.0) within the CON participants (55.5%, N=86 of 155) when compared to the TEN participants (31.7%, N=19 of 60).

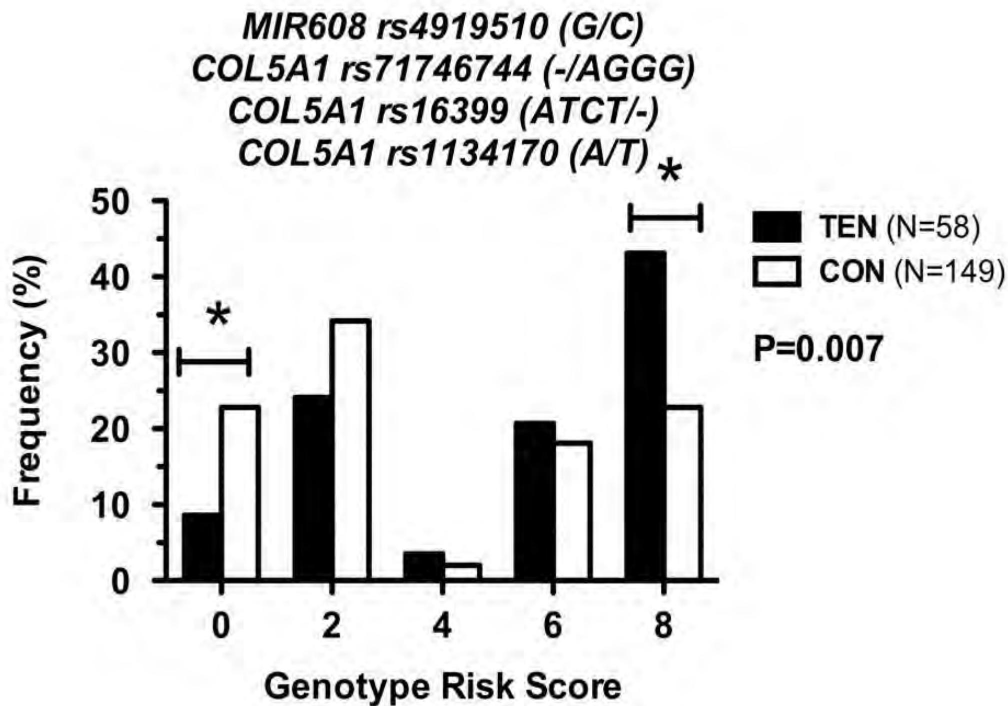


Figure 3.13 Genotype risk score frequency distributions of the Hsa-miR-608 gene (*MIR608*) rs4919510 (C/G) polymorphism and the three *COL5A1* 3'-untranslated region (UTR) polymorphisms, (i) rs71746744 (-/AGGG) polymorphism, (ii) rs16399 (ATCT/-), and (iii) rs1134170 (A/T) in the pooled South African and Australian control (CON, clear bars) and chronic Achilles tendinopathy (TEN, black bars) groups. The 'at risk' genotypes for chronic Achilles tendinopathy at each variant contributed 2 points (rs4919510, CC; rs71746744, AGGG/AGGG; rs16399, -/- ATCT; rs1134170, TT) towards the genotype risk scores while the non-risk genotypes (rs4919510, CG and GG; rs71746744, -/- AGGG and -/AGGG; rs16399, ATCT/- and ATCT/ATCT; rs1134170, AT and AA) contributed 0 points. As indicated by an asterisks, the genotype risk score of 0 was significantly under-represented in the TEN group, p=0.019, OR=3.1 and 95% CI=1.2 to 8.5. The genotype risk score of 8 was however significantly over-represented in the TEN group, p=0.004, OR=2.6 and 95% CI=1.3 to 4.9. The global P-value, which was calculated by combining the 4 and 6 genotype risk scores, is indicated.

3.6 Independently, the variants rs12722 and rs71746744 are not responsible for the tendinopathic phenotype.

Since the *COL5A1* 3’-UTR variants rs12722 (C/T) and rs71746744 (-/AGGG) were independently associated with chronic Achilles tendinopathy (Abrahams et al., 2013;September et al., 2009) (section 3.5.2), site directed mutagenesis of these sites were attempted on the parental T- and C-allelic constructs as described in section 2.4 to determine whether they were functional. In addition since variant rs71746744 (-/AGGG) is within the functional Δ57bp deletion region (section 3.3) and is also tightly linked to the two other independently associated variants, rs16399 (ATCT/-) and rs1134170 (A/T), (Table 3.7), it was decided to only mutate rs717146744. The mutants generated (insert and vector) were sequenced to confirm the desired nucleotide changes. Two mutants were obtained with an altered rs12722 (T/C) variant while only one was obtained with an altered rs71746744 (-/AGGG) variant (Figure 3.14A & 3.15A).

One rs12722_C>T and one rs12722_T>C construct was generated, where a C to a T nucleotide change was introduced at rs12722 in the CON parental construct (pGL3-COL5A1-3’UTR) and a T to a C nucleotide change was introduced at rs12722 in the parental TEN construct as illustrated in figure 3.14A. The site-directed mutants as well as their parental constructs were transfected into HT1080 cells and assayed as described in section 3.1.

There was a significant decrease (p=0.022) in the relative luciferase activity of the mutated rs12722 CON construct (73.5%, n=17) compared to the CON parental

construct (100%, n=17) (Figure 3.14B). In contrast, there was a significant ($p=0.021$) increase in the relative activity of the mutated rs12722 TEN construct (133.6%) when compared to the TEN parental construct (Figure 3.14C). Interestingly, when the relative luciferase activities of the mutated rs12722 TEN and CON constructs were compared (Fold increase of 1.8 ± 0.367,N=17, $p<0.001$; figure 3.14D), the difference in the activity was more pronounced than that reported between the C- and T-allelic form. Since the difference in mRNA stability increased rather than decreased, it appears that the variant rs12722 T/C on its own is not responsible for the tendinopathic phenotype.

There was no significant difference ($p=0.411$) in the relative luciferase activity between the CON rs71746744 (-/AGGG) construct which contains only one AGGG tetranucleotide, (80.0 ± 17.8 %, N=9) and the CON mutant rs71746744 construct which contained two AGGG tetranucleotides (72.6 ± 21.2 %, N= 12) (Figure 3.15A and B). A comparable experiment using the TEN construct could not be performed due to technical difficulties with deleting one of the two AGGG tetranucleotides from the TEN construct (Figure 2.4).

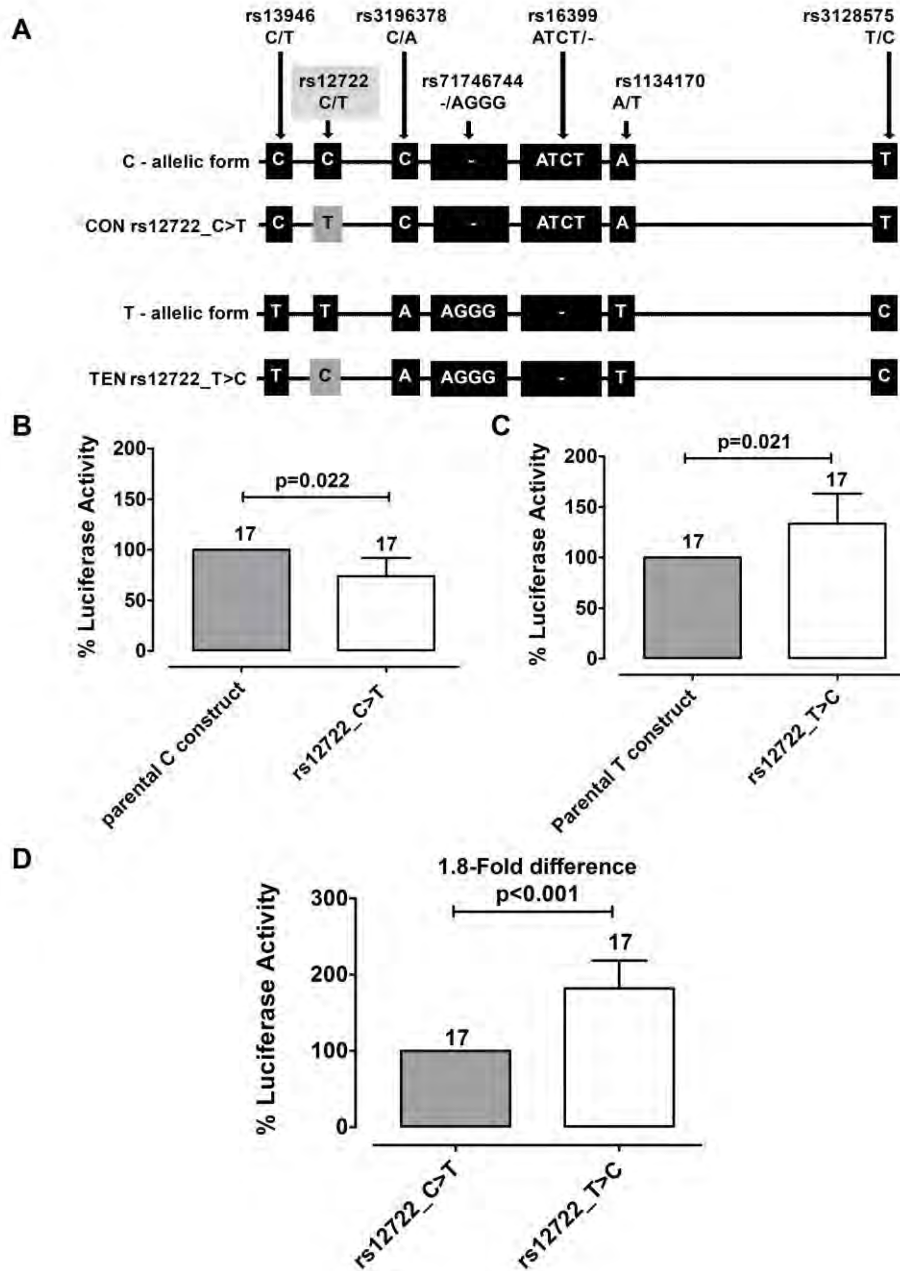


Figure 3.14 Examination of the effect of the variant rs12722 on the mRNA stability of the *COL5A1* 3'-UTR. (A) Schematic diagram representing the site-directed mutants and their parental constructs used in this experiment. The altered sites are shown as grey boxes. (B) The relative luciferase activity of rs12722_C>T compared to the C allele construct. (C) The relative luciferase activity of rs12722_T>C compared to the T allele construct. (D) The fold difference illustrated between rs12722_C>T and rs12722_T>C. The number of assays (n) is indicated below the bar. Values are expressed as mean ± standard deviations. Firefly luciferase activity for each construct was normalised to renilla luciferase activity and expressed relative to full length CON 2.1.

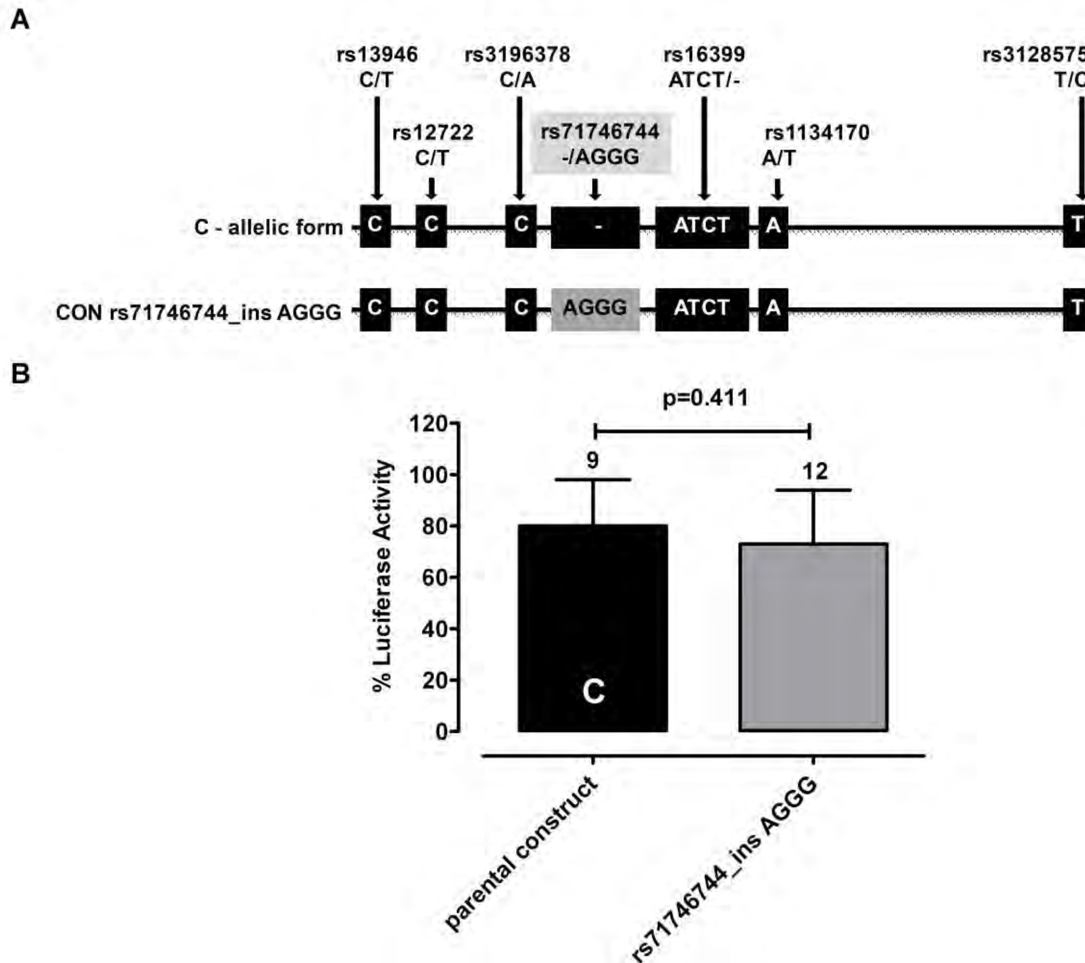


Figure 3.15 Examination of the effect of variant rs71746744 on the mRNA stability of the *COL5A1* 3'-UTR (C-allelic form). (A) Schematic diagram representing the site-directed mutant and its parental construct. The altered sites are shown as grey boxes. (B) The relative luciferase activity (%) of the mutant and its parental construct were plotted. The luciferase activity of rs71746744_ins AGGG is shown as the grey bar on the right while that of the C-allelic construct is shown in black on the left. The number of assays (n) is indicated above the bar. Values are expressed as mean ± standard deviations. Firefly luciferase activity for each construct was normalised to renilla luciferase activity and expressed relative to full length CON 2.1.

3.7 Establishment of Primary Skin Fibroblast Cell Lines

Since the *in vitro* experiments and genetic association studies described above implicated the 3'-UTR in the regulation of the *COL5A1* gene, the next set of experiments were performed to determine whether these findings could be replicated *in vivo*. Due to ethical considerations about obtaining tenocytes, especially from healthy participants, these experiments were carried out on primary fibroblasts with a known *COL5A1* genotype derived from skin biopsies. Briefly, primary fibroblast cultures were established from a superficial skin sample which was obtained from each participant with known *COL5A1* 3'-UTR genotypes at rs12722, rs71746744, rs16399 and, rs1134170 (section 2.16; Table 3.9). Primary fibroblasts could be seen migrating out of the skin explants after approximately one week (Figure 3.16). To increase the yield of fibroblasts, the explants were placed under coverslips and a week later the coverslips were inverted and placed cell-side up in new plates with fresh complete DMEM.

Once a moderate amount of cells were obtained, the explants were removed and the fibroblasts were allowed to proliferate until they were dominant in the population of cells present in the dish. The fibroblasts appeared predominantly as large, irregular and elongated in shape or were typically spindle-shaped. The plates were then treated with trypsin/EDTA and passaged for the first time and when they reached 80% confluence, they were either passaged a second time or frozen down for later use. When cultured at higher confluence, the cells aligned in an orderly parallel manner.

Table 3.9 Genotype of participants that donated a skin biopsy.

Participant	**rs12722 (T/C)**	**rs71746744 (-/AGGG)**	**rs16399 (ATCT/-)**	**rs1134170 (T/A)**
SB-1	TT	AGGG/AGGG	-/-	TT
SB-2	TT	AGGG/AGGG	-/-	TT
SB-3	TT	AGGG/AGGG	-/-	TT
SB-4	TC	-/AGGG	-/-	AT
SB-5	CC	-/AGGG	-/-	AT
SB-6	CC	-/-	ATCT/-	AA
SB-7	CC	-/AGGG	-/-	TT

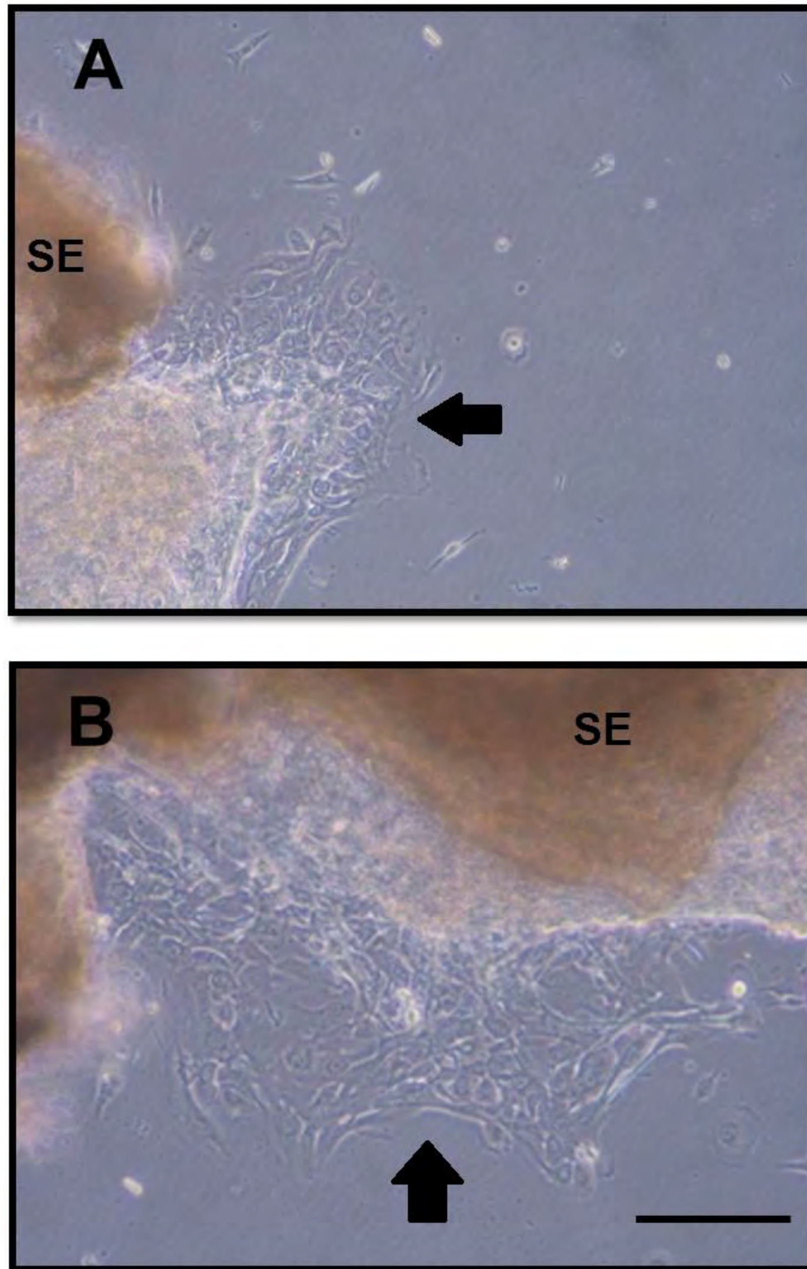


Figure 3.16 (A, B) High-power photomicrographs showing migrating cells (see arrows) out of skin explants (SE) after approximately a week in culture. The scale bar represents 100µm.

3.8 A Preliminary Study shows an Increase in the Expression Levels of *COL5A1* and *COL1A1* for Individuals with a T allele at rs12722.

Fibrillogenesis requires both Type V and Type I collagen for initiation and elongation of the fibril. This raised the question as to whether the levels of *COL5A1* and *COL1A1* were altered in individuals with a T allele at rs12722. To address this, total RNA was extracted from the established primary fibroblast cultures and subjected to Quantitative Real-Time Reverse Transcription PCR (Q-RT-PCR) (donor's cells SB-1 to SB-7).

Figure 3.17A and B respectively shows the relative mRNA abundance of *COL5A1* (clear bars) and *COL1A1* (grey bars) with respect to the rs12722 genotype of each participant (indicated at the base of the bar graphs). Two independent biological replicates were used and each experiment was performed in triplicate. The third graph (Figure 3.17C) displays the ratio of the relative mRNA expression of *COL1A1* and *COL5A1* (COL1A1/COL5A1). An inter-individual variation was observed for both *COL5A1* and *COL1A1* (global $p < 0.001$, $N \geq 30$).

When compared to all donor fibroblasts, the relative mRNA abundance of *COL5A1* was significantly the highest in SB-2 (TT genotype, 2.25 ± 0.85, $p<0.024$, N=5). However, there was no significant difference in the relative mRNA abundance of *COL5A1* between SB-1 and SB-2, both donor fibroblasts having the TT genotype. The *COL5A1* transcript abundance was significantly higher for SB-1 (1.75 ± 0.68, N=4) compared to SB-3 ($p=0.049$, N=5), SB-5 ($p=0.005$, N=6) and SB-6 ($p=0.011$, N=6). Similarly, compared to all donor fibroblasts the expression of *COL1A1* (Figure

3.17B) was significantly higher for SB-2 (3.34 ± 0.45, p<0.001, N=4). The mRNA expression of *COL1A1* was significantly elevated in the SB-1 donor fibroblasts from a participant with a TT genotype (1.47 ± 0.48, p<0.001, N=4) compared to SB-5 donor fibroblasts from a CC genotype (0.45 ± 0.21, N=6). However, there were no significant differences in the *COL1A1* to *COL5A1* ratio (Figure 3.17C) within the individual samples except for participant SB-2 (p=0.008, N≥4).

When the relative expression levels of the transcripts was pooled in terms of the genotype at the rs12722 locus an interesting pattern was observed (Figure 3.18A, B). The TT genotype displayed a higher *COL5A1* (1.58 ± 0.89, p<0.001, N=14) and *COL1A1* (1.88 ± 1.10, p=0.0015, N=13) mRNA abundance compared to that of the CC genotype (*COL5A1*: 0.57 ± 0.19, N=12 and *COL1A1*: 0.67 ± 0.29, N=12) regardless of whether SB-2 was included or excluded from the analysis (p=0.010 for *COL5A1* and p=0.003 for *COL1A1*, N=9). No significant differences were observed between the ratio of *COL1A1* to *COL5A1* as shown in figure 3.18C.

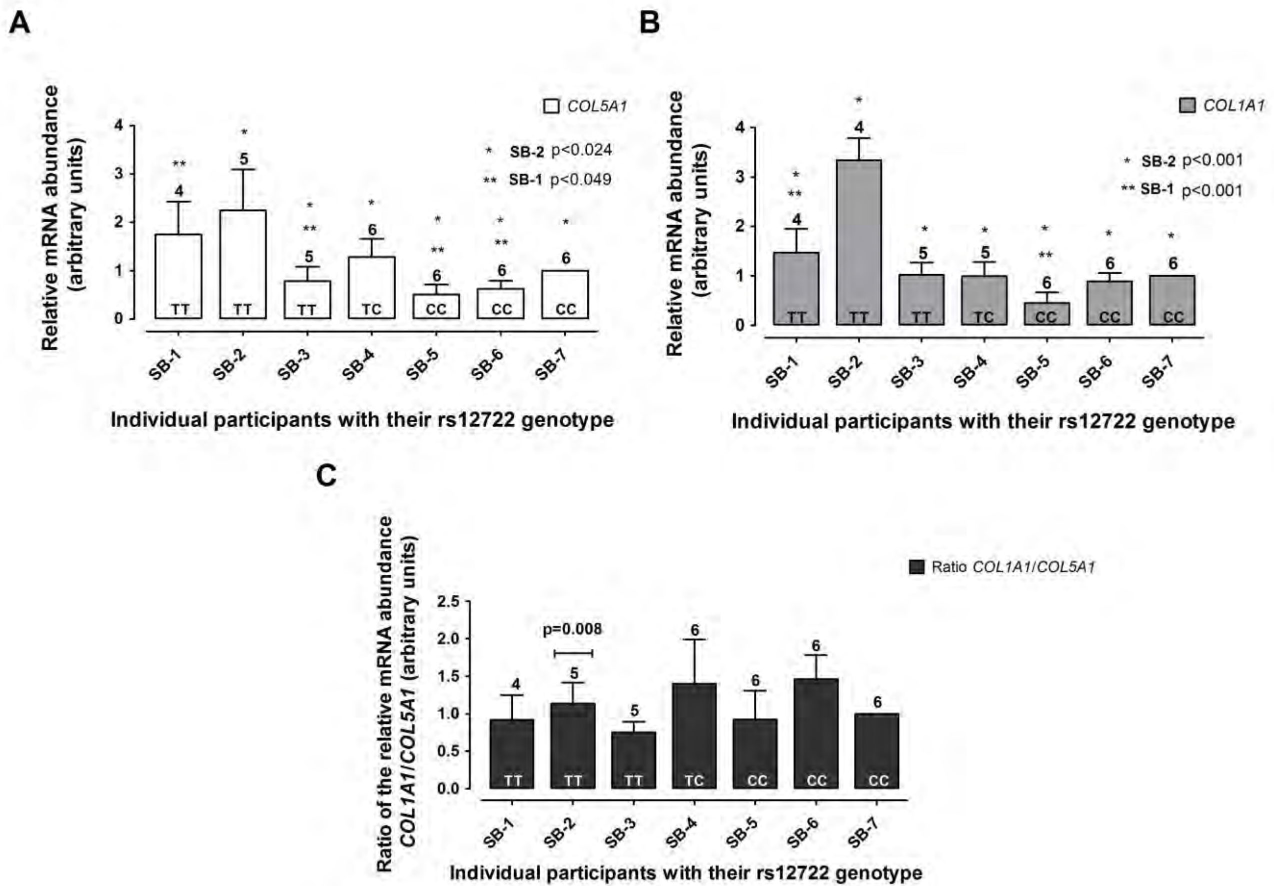


Figure 3.17 Relative mRNA abundance of (A) *COL5A1* (clear bars) and (B) *COL1A1* (grey bars) expressed by the skin fibroblasts cultured from each participant (SB-1 to SB-7). (C) The ratio of the relative mRNA abundance of the two genes, *COL1A1*/*COL5A1* (black bars). The relative mRNA abundance ($2^{-\Delta\ (\Delta CT)}$) is given in arbitrary units. The genotype at the rs12722 locus of each individual is indicated at the base of each bar graph. The number of samples included is shown on top of the bars and the significance is indicated on the graphs. SB-7 (CC at rs12722) is used as a control sample and *β-actin* as the housekeeping gene. The comparative C_T method was used to derive the relative mRNA abundance from the threshold cycle (C_T) values obtained.

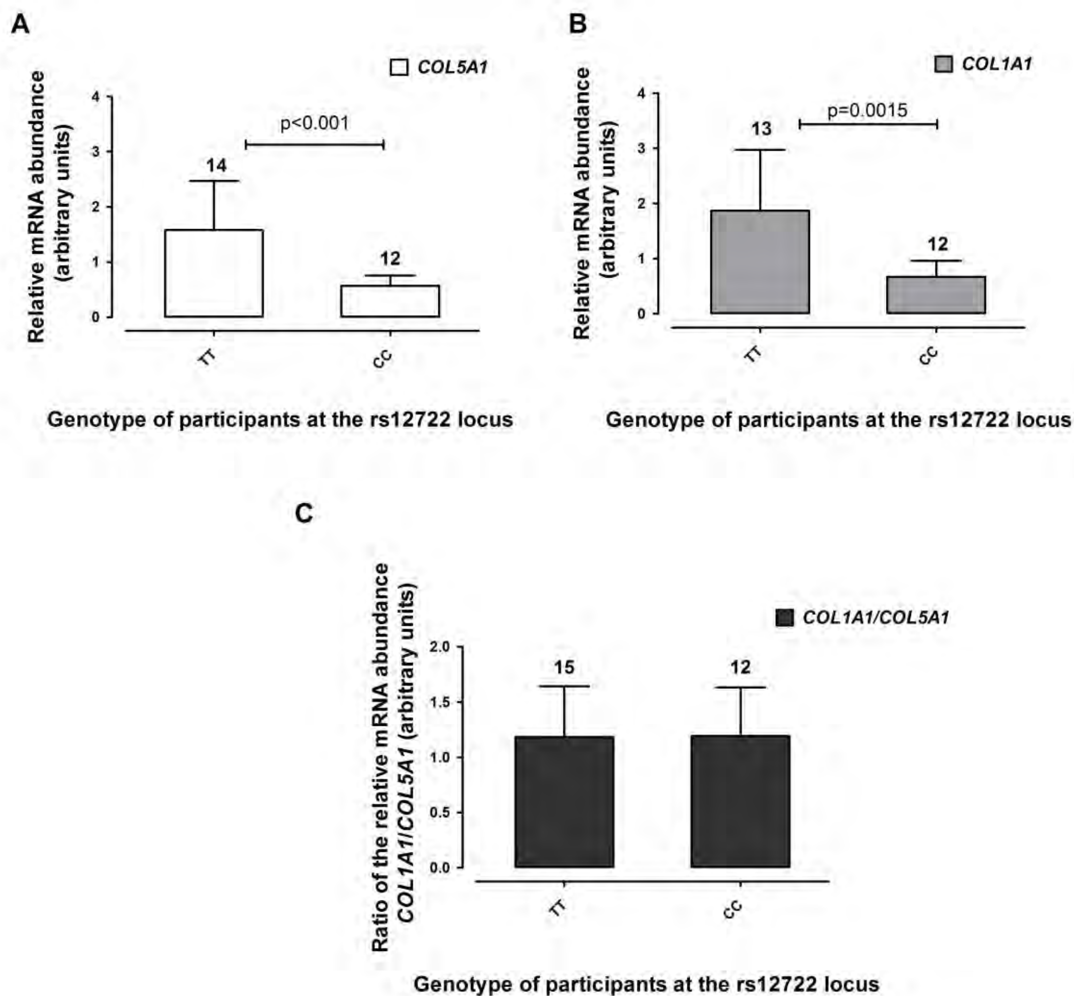


Figure 3.18 Relative mRNA abundance of (A) *COL5A1* (clear bars) and (B) *COL1A1* expressed (grey bars) in skin fibroblasts pooled with respect to their genotype at the rs12722 locus. (C) The ratio of the relative mRNA abundance of the two genes, *COL1A1*/*COL5A1* (black bars). The relative mRNA abundance ($2^{-\Delta\ (\Delta CT)}$) is given in arbitrary units. The number of samples included is shown on top of the bars and the significance indicated on the graphs. SB-7 (CC at rs12722) is used as a control sample and *β-actin* as the housekeeping gene. The comparative C_T method was used to derive the relative mRNA abundance from the threshold cycle (C_T) values obtained.

3.9 Differential Binding of RNA Binding Proteins (RBPs) to the C- and T-allelic forms.

As reported in section 3.3, a 57bp *Xba*I-*Afl*II deletion region within the *COL5A1* 3'-UTR may contain site(s) responsible for the functional differences between the C- and T-allelic forms. Previous work in our laboratory excluded the involvement of two polymorphic miRNA binding sites (Abrahams, 2013;Laguette et al., 2011). However, when this region was screened using a RNA binding protein (RBP) database (http://rbpdb.ccbr.utoronto.ca/) putative sites were identified for NONO, KHDRBS3, HNRNPA1, EIF4B, ELAVL1 and PTBP1 in both allelic forms (See Table 3.10 and Figure 3.19).

The non-pou domain-containing octamer-binding protein (NONO) is a 54kDa nuclear protein involved in RNA biogenesis and DNA repair (Krietsch et al., 2012). The KH domain-containing RNA binding signal transduction-associated protein 3 (KHDRBS3) is a 55kDa nuclear protein which may act as a shuttle between the nucleus and cytoplasm (Lei et al., 2011). The heterogeneous nuclear ribonucleoprotein A1 (HNRNPA1) is 35kDa which affects mRNA stability during collagen synthesis (Thiele et al., 2004) and is involved in numerous cellular processes through its ability to bind to both DNA and RNA sequences (Chen et al., 2010). The eukaryotic translation initiation factor 4B (EIF4B) is an 80kDa protein important for mRNA biogenesis (Meyer et al., 1995) and capable of binding targets in the cytoplasm thereby directing translation initiation (Shahbazian et al., 2010). The (embryonic lethal, abnormal vision, Drosophila)-like 1 (ELAVL1) is a 36kDa protein which can increased mRNA stability, resulting in greater translation and gene

expression (Viiri et al., 2013). The polypyrimidine tract-binding protein 1 (PTBP1) is 57KDa in size and is involved in mRNA biogenesis, namely alternative splicing (Romanelli et al., 2013).

Interestingly, compared to the C-allelic form which had 1 binding site for NONO, the T-allelic form had two binding sites for this RBP (Figure 3.19B). The current study investigated the possibility that one or more of the putative RBP sites identified may be responsible for the functional differences between the C- and T-allelic forms using RNA electromobility shift assays (EMSAs). Briefly, this assay involved incubating biotinylated C- and T-allelic RNA probes with either nuclear or cytoplasmic protein extracts from HT1080 cells. Protein-RNA complexes and free RNA probe were resolved using native polyacrylamide gel electrophoresis and following crosslinking, signals were detected using a streptavidin-HRP conjugate and the chemiluminescent signal emitted captured. While DNA EMSAs are well established in our laboratory, we had not previously performed RNA EMSAs and therefore it had to be optimised, as described below.

Protein-RNA interactions are influenced by the tertiary structure of the protein and in RNA EMSAs, both the RNA and protein(s) must be correctly folded to allow proper binding (Butcher and Pyle, 2011). In addition, RNA is particularly vulnerable to rapid degradation and care must be taken to prevent contamination of the reaction with RNases (RNA degrading enzymes). Several strategies were therefore explored to harvest protein lysates suitable for RNA EMSAs and the most critical factors are described here. Rather than a mechanical lysis, a gentle chemical lysis was employed where re-suspension buffers excluded the reducing reagent DTT but

included 1mM EGTA pH 6.8, phosphatase inhibitors as well as protease inhibitors (detailed in section 2.13). Moreover RNase-free techniques were implemented at every step of the RNA EMSA protocol. To optimise complex formation in the binding assays, a titration with decreasing amounts of tRNA, the non-specific competitor, was performed. Based on the results obtained, 10mg/ml tRNA was chosen for all future experiments. Lastly, to determine the optimal quantity of protein lysates required, the RNA probes were reacted with increasing amounts of protein extracts. Based on the results from these experiments, 12 and 24μg nuclear extracts and 13 and 26μg cytoplasmic extracts were used (detailed in section 2.14).

It is important to note that due to inherent differences the T-allelic probe was 4 nucleotides longer than the C-allelic form. Furthermore, due to limitations of the size of the probes that can be used in these RNA EMSAs, probes were generated that only spanned five of the putative RBP sites (see grey areas in Figure 3.19).

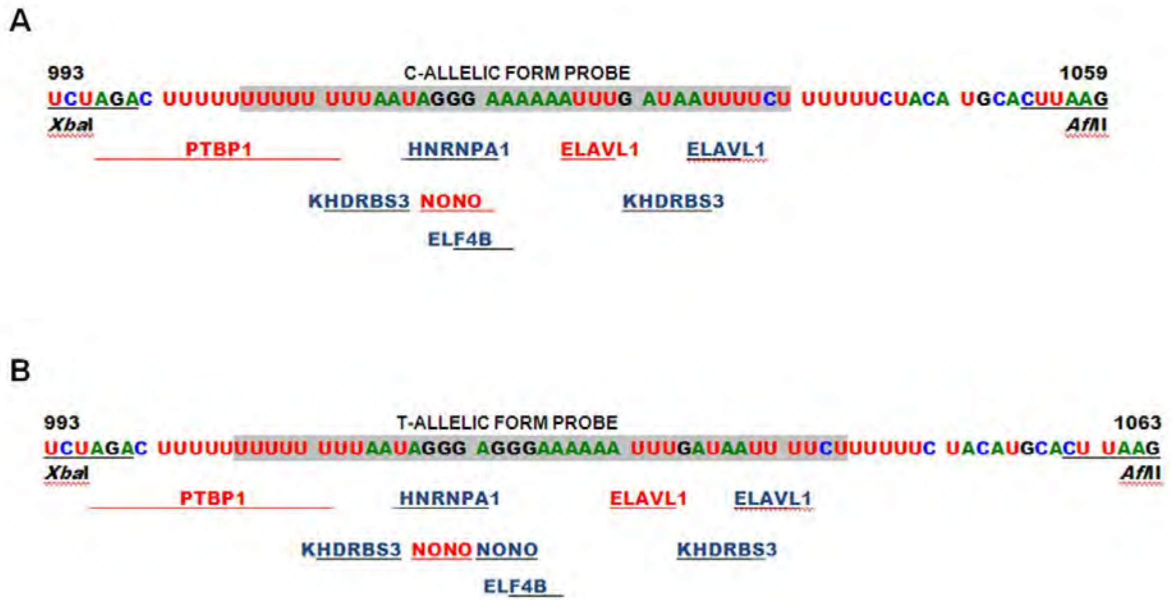


Figure 3.19 Putative RBPs binding sites (underlined) within the 57bp deletion region (flanking RE sites annotated and sequence underlined) together with the sequence of the RNA biotinylated probes (mapped in grey highlight) used in this experiment for the (A) C-allelic form and (B) the T-allelic form. The position is the nucleotide number of the *COL5A1* exon 66 sequence.

Table 3.10 Summary of the size, putative binding site sequence, localisation and function of key RBPs binding putatively within the 57bp deletion region of the 3'-UTR of *COL5A1* gene

RNA Binding Protein	Molecular Mass (kDa)	Binding site sequence	Primary Localization	Key Function
NONO	54	AGGGA	Nuclear	RNA biogenesis
KHDRBS3	55	UUUAAU/ GAUAAU	Nuclear	Nuclear and cytoplasmic shuttle protein
HNRNPA1	35	UAGGGA	Nuclear	RNA biogenesis
EIF4B	80	GGAA	Cytoplasmic	Translation initiation
ELAVL1	36	AUUU	Nuclear and cytoplasmic	RNA biogenesis
PTBP1	57	AGAC$(U)_{11}$	Nuclear	RNA biogenesis

NONO, Non-pou domain-containing octamer-binding protein; KHDRBS3, KH domain-containing RNA-binding signal transduction-associated protein 3; HNRNPA1, heterogeneous nuclear ribonucleoprotein A1; EIF4B, Eukaryotic translation initiation factor 4B; ELAVL1, Embryonic lethal, abnormal vision, Drosophila-like 1; PTBP1, Polypyrimidine tract-binding protein 1

The results show that the signal for the shorter C-allelic unbound probe was consistently more intense and that it migrated faster through the gel than the unbound T-allelic probe used at the same concentration. A large complex was observed when 13 and 26µg of the cytoplasmic extract was incubated with 0.125µM of the C- and T-allelic probe respectively (indicated by * in figures 3.20 and 3.21B). This complex was also detected in lanes containing cytoplasmic extract only i.e. in the absence of the RNA probes, but not in lanes containing nuclear protein extracts with or without probes (Figures 3.20 and 3.21). Together these results suggest that the cytoplasm of HT1080 cells contain endogenous biotinylated proteins such as carboxylases which is consistent with other reports (Bina, 2006;Wolinsky and Driskell, 2005).

Importantly, a signal showing a complex was detected when nuclear extracts were incubated with the C-allelic RNA probe (indicated by ** in figure 3.21A). This signal was also present when the C-allelic probe was incubated with cytoplasmic extract but it was much fainter (Figure 3.21B). This signal (**) was absent in the presence of the T-allelic probe. These results indicate that a complex may be forming in the presence of the C-allelic probe selectively. Noteworthy, when the T-allelic RNA probe was incubated with either nuclear or cytoplasmic protein extracts, a signal was detected at the level of the wells (indicated by $^{\#}$ in figure 3.20 and 3.21A, B). It is tempting to speculate that this represents a true complex because the signal for the unbound T-allelic probe was less intense in these lanes. Furthermore, this “complex” was also present when the C-allelic probe was incubated with both protein lysates but the signal was much less intense which corresponded with a more intense signal

for the unbound C-allelic probe. These results suggest that a large complex may be forming in the presence of the T-allelic probe.

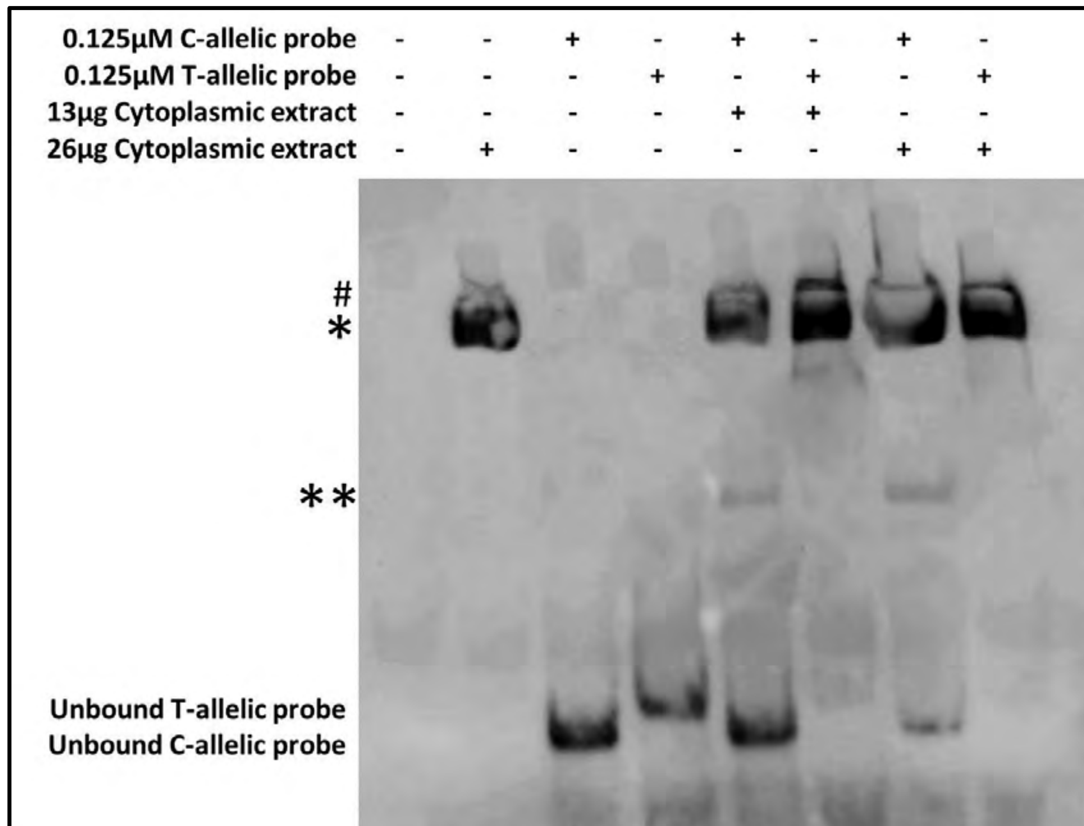


Figure 3.20 Blot of a non-radioactive RNA EMSA where 13 and 26µg of the cytoplasmic protein extract reacted with 0.125µM of the C- or T-allelic probes spanning the functional 57bp deletion region as described in the results section. The symbol *, ** and # denoted the position of RNA-Protein complexes on the blot.

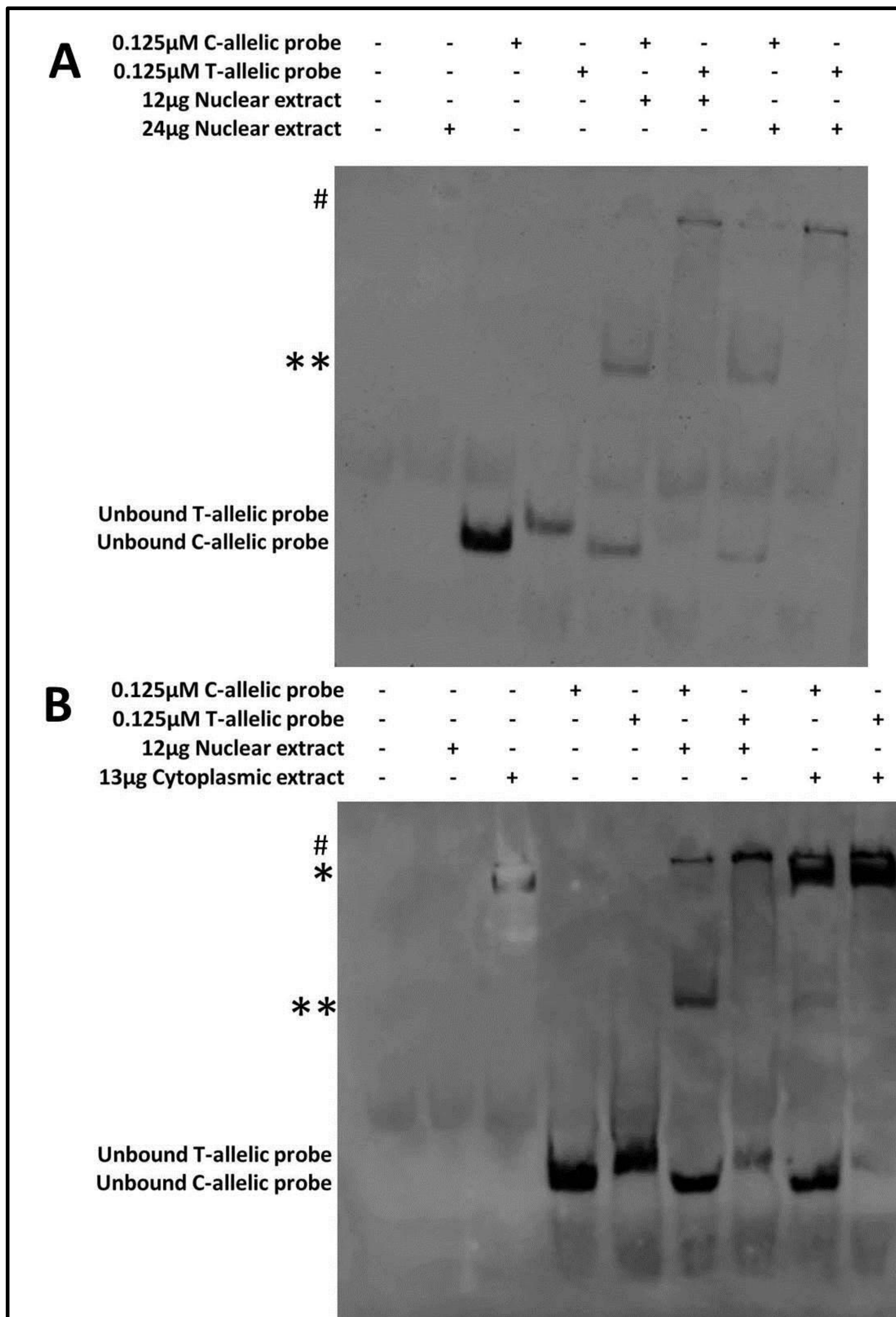


Figure 3.21 Blot of a non-radioactive RNA EMSA where (A) 12 and 24 µg of the nuclear protein extract reacted with 0.125µM of the biotinylated C- or T-allelic probes spanning the functional 57bp deletion region (B) Both 12µg of the nuclear and 13µg cytoplasmic protein extract respectively were used with the biotinylated probes in a replicated experiment. The symbol *, ** and # denoted the position of RNA-Protein complexes on the blot as described in the text.

CHAPTER 4 DISCUSSION

Research in our laboratory has previously reported the association of a common C/T polymorphism (SNP rs12722; *Bst*UI RFLP) within the *COL5A1* 3'-UTR with musculoskeletal soft tissue injuries including chronic Achilles tendinopathy (Mokone et al., 2006;September et al., 2009), ACL injuries in females (Posthumus et al., 2009), exercise-associated muscle cramps (O'Connell et al., 2013), Carpel Tunnel syndrome in occupational workers (Burger et al., 2014), as well as complex exercise-related phenotypes, including range of motion measurements (Brown et al., 2011;Brown et al., 2011;Collins et al., 2009) and running endurance performance (Brown et al., 2011;Posthumus et al., 2010). Although the 3'-UTR of eukaryotic genes contain important regulatory elements shown to be involved in the aetiology of many diseases (Conne et al., 2000), the biological function of the *COL5A1* 3'-UTR is unknown. This thesis reports novel findings on the role of the *COL5A1* 3'-UTR and its possible involvement in musculoskeletal soft tissue injuries and exercise-related phenotypes.

4.1 The tendinopathic phenotype has an overall increase in *COL5A1* mRNA stability and expression.

Expression from both copies of the *COL5A1* gene is required for normal fibrillogenesis of the heterotypic collagen fibrils which contain, among other proteins, types I, III and V collagen (Birk, 2001). Mutations resulting in the loss of function of one *COL5A1* copy are responsible for a significant proportion of the heritable

connective tissue disorder, classic Ehlers-Danlos syndrome (Malfait et al., 2010). The 50% reduction in the amount of type V collagen produced in these patients result in abnormal fibrillogenesis, which is characterised by an increase in the average fibre diameter with a greater variability in the width and shape of the fibrils. Similar findings have been reported in *Col5a1* +/- mice, while *Col5a1* -/- mice die in utero at E10.5 (Wenstrup et al., 2006), displaying a direct relationship between the transcription of the gene, the α1(V) chain synthesis, type V collagen as well as heterotypic fibril organisation. Based on the dosage sensitivity demonstrated in these studies, it is speculated that relatively small changes in *COL5A1* mRNA stability within the normal physiological range could result in inter-individual variation in fibrillogenesis. This could lead to altered susceptibility to musculoskeletal soft tissue injuries, as well as to variations in flexibility and endurance running performance.

The primary aim of this thesis was therefore to determine whether the *COL5A1* 3'-UTR was functional and whether it differed between participants with Achilles tendinopathy (TT genotype at rs12722) and asymptomatic controls (CC genotype at rs12722). To this end, a well described approach to measuring the effect of the 3'-UTR on mRNA stability was employed in two fibroblast cell lines (Ogawa et al., 2010;Ross, 1995;Wang et al., 2008). In this thesis, experiments performed in the HT1080 fibroblast cell line show an overall increase in *COL5A1* mRNA stability in the tendinopathic phenotype. A fold difference was obtained which is consistent with reported values for other extracellular matrix genes which range from 1.5 to 4 fold (Akhtar et al., 2010;Du et al., 2010;Liu et al., 2009;Ogawa et al., 2010;Thiele et al., 2004). The HT1080 cells are commonly used for transfection assays with reporter genes and have previously been used to study the expression of pGL3-Promoter

constructs in which the 3′-UTR of *COL1A1*, *COL1A2* and *COL3A1* were cloned downstream of the luciferase reporter (Thiele et al., 2004).

When the above experiments with the *COL5A1 3'-UTR* were however performed in the SV40 transformed WI-38 lung fibroblast cell line, no significant difference was observed in mRNA stability between the asymptomatic controls and the tendinopathic phenotype. The reason(s) for the different results in the two cells lines is currently unknown but it is worth noting that these cells were derived from different sources which may account for these differences. While the HT1080 human fibroblast cell line was derived from a fibrosarcoma arising adjacent to the acetabulum (at the hip joint) of a 35 years old Caucasian male individual (Rasheed et al., 1974), the SVWI-38 is a human embryonic lung fibroblast cell line (WI-38) transformed with the *Simian virus 40* (SV40) (de Haan et al., 1986). It is also worth noting that although specific expression of the type V collagen genes have not been investigated in this SV40 transformed fibroblast cell line, *COL1A2* gene expression is transcriptionally repressed in these cells (Parker et al., 1992). To avoid such limitations in the future, prospective experiments should be perform in more relevant cells such as a primary fibroblast or tenocyte cell line.

Importantly, consistent with the data obtained from *in vitro* luciferase assays in the HT1080 cells, results from an *ex vivo* assay also showed that there was a higher mRNA abundance of *COL5A1* in primary fibroblasts from individuals with an 'at risk' genotype (rs12722, TT) compared to that with a 'protective' genotype (rs12722, CC). Indeed, a similar fold difference of approximately 2.8-fold mRNA abundance of *COL5A1* was obtained between the two genotypes in both experimental approaches.

Studies in murine models and EDS patients with haploinsufficiency of *COL5A1* display severe phenotypes such as aberrant skin as well as musculoskeletal soft tissue defects (Wenstrup et al., 2000;Wenstrup et al., 2006). Type V collagen as mentioned in Chapter 1 is important in the development of the skin and, although the skin inherently has a different organisation compared to the tendon, symptoms of EDS manifest in the skin as well. Therefore, making use of fibroblasts derived from skin explants is relevant as an alternative when examining the possible functional effects of genetic variations in the *COL5A1* gene.

Of particular interest, was the observation that *COL1A1* expression followed a similar trend to that seen for *COL5A1* in the *ex vivo* study. While the *COL1A1* genotype of the donor fibroblasts was not established in the current study, functional variants within the first intron of the *COL1A1* gene have been shown to alter its expression at an inter-individual level in osteoarthritis (Jin et al., 2009;Mann et al., 2001). It would therefore be important for future studies to investigate if this is also the case in Achilles tendinopathy. The possibility of gene-gene interactions between the *COL5A1* and *COL1A1* genes cannot be excluded since they are both intimately involved in fibrillogenesis. Future genetic and *ex vivo* studies with the functional variants of *COL1A1* and a larger sample size are necessary to explore whether altered *COL1A1* gene expression is a possible confounder.

Future experiments should also focus on expanding the current primary fibroblast study in order to control for other interacting variants and explore the effect of specific genotype combinations on expression of *COL5A1* and related genes, both at the mRNA and the protein level. Moreover, due to the variability present between

individuals, it is important to repeat these experiments in larger studies to compensate for this variation. Furthermore after careful ethical consideration, it would be ideal to ascertain COL5A1 mRNA and protein expression in tenocytes from tendinopathic and healthy individuals with respect to the 3'-UTR genotype at the different loci.

Unfortunately, *in vitro* mRNA decay assays performed to confirm the above results were unsuccessful due to technical reasons (Appendix E) and further work is thus required to optimise this assay.

4.2 Identification of C- and T-allelic forms of the *COL5A1* 3'-UTR

The current study attempted to establish if there were differences in the sequence of the *COL5A1* 3'-UTR in control and Achilles tendinopathic individuals and identified two major allelic forms referred to as the C- and T-allelic forms. While the C-allelic form corresponds to the wild type sequence and was identified in most of the clones generated from asymptomatic controls, the T-allelic form was generally identified in the Achilles tendinopathic patients. Importantly, the mRNA stability of the T-allelic form was significantly higher than the C-allelic form suggesting that more pro-α1(V) chain may be synthesised from the T-allelic form. As mentioned above, the preliminary mRNA expression study suggests that the relative mRNA abundance of *COL5A1* may be increased in the presence of the TT genotype at rs12722 which is a contributor of the T-allelic form. Further work is however required to determine whether these results translate into increased type V collagen production and altered fibrillogenesis.

Although 49 polymorphisms have been identified within the *COL5A1* 3'-UTR (www.ncbi.nlm.nih.gov, accessed on 18 January 2011), the allele frequencies of 36 of these polymorphisms is currently unknown and most are probably likely to be rare. Two of these variants, rs71746744 (-/AGGG) and rs16399 (ATCT/-), nevertheless contributed to the major *COL5A1* allelic form and, as shown in this thesis, have a relatively high minor allele frequency in this population. Only seven of the remaining 13 polymorphisms, with allele frequency data, had a minor allele frequency of >20%, of which five, rs13946 (C/T), rs12722 (C/T), rs3196378 (C/A), rs1134170 (A/T) and rs3128575 (T/C), were also identified within the *COL5A1* 3'-UTR allelic forms. This thesis demonstrate that the two major *COL5A1* 3'-UTR allelic forms were determined by seven tightly linked polymorphisms. Based on (1) the available information for the identified *COL5A1* 3'-UTR polymorphisms within the public databases, (2) the limited number of clones generated and sequenced within this study, and (3) the strategy of only cloning the "extreme" phenotype and specific genotypes for the single nucleotide polymorphisms rs13946 and rs12722, it is not surprising that only the wild type (C-allelic form) and a single allele (T-allelic form) associated with tendinopathy were identified. The sequence of the seven polymorphisms for the C- (wild type) and T-allelic form were C-C-C-$(AGGG)_1$-$(ATCT)_2$-A-T and T-T-A-$(AGGG)_2$-$(ATCT)_1$-T-C respectively. Sequencing of a larger population would be required to identify other potentially functional forms of the 3'-UTR.

Using an *in silico* approach, it was demonstrated that the C- and T-functional forms had clearly different predicted structural differences in their most stable state (Abrahams et al., 2013) (Figure 4.1). Furthermore, analyses of the 2.5kb *COL5A1* 3'-UTR indicated that the distinct predicted secondary structures of the C- and T-

functional forms were due to the sequence differences within the seven polymorphic sites, with the rs71746744 (-/AGGG) variant modulating most of the differences in their structure.

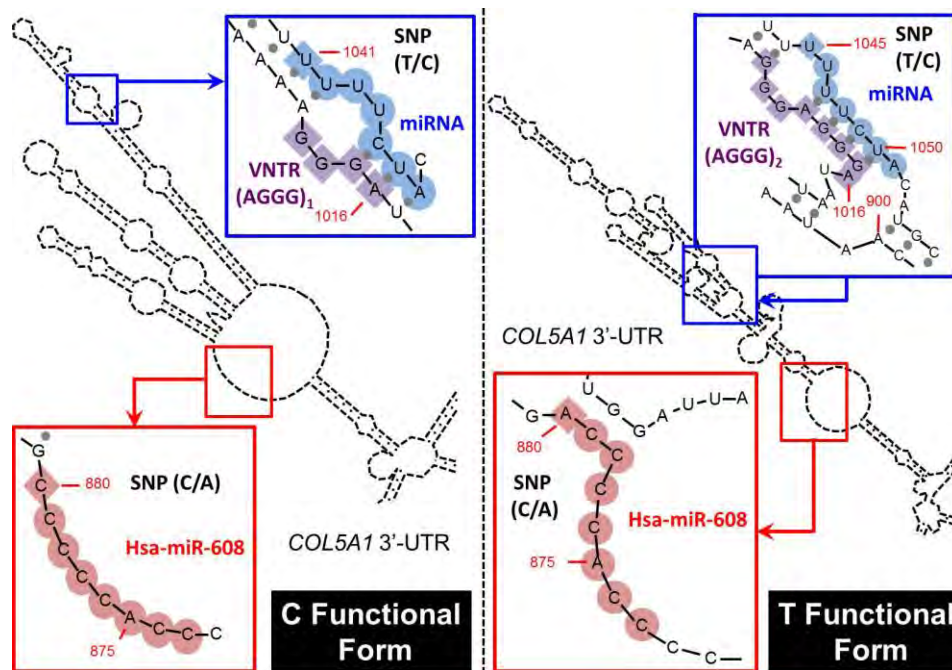


Figure 4.1 The most stable predicted secondary structures of a region belonging to the C (left panel) and T (right panel) functional forms of the *COL5A1* 3'-UTR modified from Abrahams *et al*., 2013. This region contains both polymorphic miRNA binding sites, single nucleotide polymorphism (SNP) rs11103544 (T/C) (in blue) and SNP rs3196378 (C/A) (in red), and the AGGG variable nucleotide tandem repeat (STRP) (rs71746744, in purple). The region to which Hsa-miR-608 (red box, bottom inserts) and the second unknown miRNA (blue box, top inserts) binds are expanded in the boxed inserts. Nucleotide positions within the 3'-UTR are also indicated. The secondary structures were generated using the Sfold online RNA folding tool (available at http://sfold.wadsworth.org).

4.3 Variants rs71746744 (-/AGGG), rs16399 (ATCT/-) and rs1134170 (A/T) are also associated with Achilles tendinopathy

A key novel finding of this study was that three additional sequence variants, namely rs71746744 (-/AGGG), rs16399 (ATCT/-) and rs1134170 (A/T) are also independently associated with Achilles tendinopathy. Specifically, the AGGG/AGGG, -/- and TT genotypes of rs71746744 (-/AGGG), rs16399 (ATCT/-) and rs1134170 (A/T) respectively were significantly over-represented in the tendinopathic participants. There was approximately a two-fold increased risk of developing chronic Achilles tendinopathy with any one of these three genotypes, which are tightly linked (D' ≥ 0.87) and located within a 256bp region of the *COL5A1* 3'-UTR. All three *COL5A1* 3'-UTR variants assayed in this study are between nucleotide positions 1016 and 1272 downstream of rs12722 and linked to this variant as well (D' ≥ 0.55).

In addition, since a *MIR608* polymorphism (SNP rs4919510, C/G) was also independently associated to Achilles tendinopathy (Abrahams et al., 2013) and the *COL5A1* 3'-UTR contains a functional Hsa-miR-608 binding site, genotype risk score frequency distributions were calculated consisting of the independently associated *MIR608* and the three *COL5A1* 3'-UTR, rs71746744, rs16399 and rs1134170, polymorphisms. The highest risk score, consisting of all the 'at risk' genotypes, was significantly over-represented in the symptomatic compared to asymptomatic participants, with or without *MIR608* included in the analysis. This finding supports the polygenic nature of musculoskeletal soft tissue injuries and other related phenotypes (Collins and Raleigh, 2009).

Allelic interaction of the variants in the *COL5A1* 3'-UTR was studied and inferred haplotypes were constructed with (1) six of the seven polymorphisms and (2) the three recently genotyped variants, described above, which distinguished the two forms of the 3'-UTR. Interestingly, in both cases, the inferred haplotype representing the T-allelic form was significantly over-represented reinforcing what was observed when the clones were sequenced. Furthermore, a recent cross-sectional study has determined that the *COL5A1* rs71746744 was independently associated with pre-race joint range of motion and running performance during the Two Oceans 56-km ultramarathon road race in 2009 and 2011 (Abrahams et al., 2014). The AGGG/AGGG genotype was significantly over-represented in the fastest and inflexible athletes compared with those with either the -/AGGG or -/- genotype. Together, these findings strengthen the associations linking the *COL5A1* 3'-UTR to an involvement in the biomechanical properties of soft tissues. Research has primarily focussed on the association of rs12722 (C/T, BstUI RFLP) with various musculoskeletal soft tissue injuries and exercise-associated phenotypes in Caucasian population groups. The novel independent association of variants rs71746744 (-/AGGG), rs16399 (ATCT/-) and rs1134170 (A/T) with chronic Achilles tendinopathy reported in this thesis should be investigated in all the other phenotypes as well as other ethnic groups.

4.4 A 57bp Region of the *COL5A1* 3'-UTR Confers mRNA Stability

The *COL5A1* 3'-UTR contains several putative polyadenylation signals (Ross, 1995) and miRNA binding sites (Meola et al., 2009) and to begin to identify the regulatory elements within the two major *COL5A1* 3'-UTR allelic forms, deletion constructs (Δ488bp and Δ57bp) were generated to include or exclude some of these. These constructs were cloned downstream of a luciferase reporter and luciferase activity measured as an indication of their impact on *COL5A1* mRNA stability. A deletion construct (Δ488bp) in which two of the three putative polyadenylation signal sites were removed had significantly higher luciferase activity in the T-allelic form compare to the C-allelic form. Generally, however there was a significant increase in luciferase activity of the Δ488bp constructs for both allelic forms when compared to their respective parental full length constructs. This suggests that this deleted region of the 3'-UTR contain novel regulatory elements which do not appear to contribute to the tendinopathic phenotypes. Further work is required to identify these regulatory elements and their role in regulating *COL5A1* expression.

When the Δ57bp constructs for the C- and T-allelic forms were compared there were no differences in their luciferase activity. This suggests that this region must contain regulatory elements responsible for the different impact of the C- and T-allelic forms on *COL5A1* mRNA stability. The Δ57bp region contains the following polymorphisms: (1) rs71698207 which did not contribute to the major T- or C-allelic form and was not investigated at the genetic level (Table 3.1), (2) the associated rs71746744 which is a contributor to the major allelic forms and (3) one of the putative polymorphic miRNA binding sites (rs11103544, *Mbo*II RFLP, T/C) (Figure

4.2). Although the putative miRNA binding site within the 57bp region is polymorphic, we have previously shown that the SNP within this site (rs11103544) is not associated with chronic Achilles tendinopathy (September et al., 2009). In addition, this SNP together with rs71698207 was not one of the major sequence variants that differentiated between the C- and T-functional forms of the *COL5A1* 3'-UTR. A site-directed mutagenesis experiment indicated that this site does not have implications in the tendinopathic phenotype and that possibly other regulatory sites in the region might be responsible for the increase in mRNA stability for the T-allelic form (Abrahams, 2013).

Figure 4.2 Schematic diagram showing the three polymorphisms present in the 57 bp deleted region. The polymorphisms rs71698207, rs71746744 and rs11103544 are underlined and identified below the sequence.

Moreover, in the current study, the upstream rs12722 (T/C) variant, which has been consistently associated with various musculoskeletal soft tissue injuries, and rs71746744 (-/AGGG) within the functional Δ57bp deletion region were mutated independently using site-directed mutagenesis in order to determine whether the variants were individually functional. Although some effect was observed in the luciferase activity of the rs12722 mutants, none of the two variants were independently responsible for the difference in mRNA stability observed between the

C- and T-allelic forms. These results suggest that there might be complex interactions at play in the 3’-UTR where not a single element or region within the area is responsible for the observations. It is also possible that the two unexamined variants, rs16399 (ATCT/-) and rs1134170 (A/T), may contribute to the phenotype but future experiments need to investigate this possibility. Moreover, to study the possible coordinated role of different sites within the 3’-UTR on mRNA stability, the effect of modifying the untested variants one at a time or combining different variants should be examined by site-directed mutagenesis.

4.5 Differential Binding of RNA Binding Proteins (RBPs) to the C- and T-allelic forms.

Although we initially focused on the functional polymorphic miRNA binding sites within and downstream of the Δ57bp region, they did not appear to significantly contribute to the tendinopathic phenotype. In an attempt to explore other regulatory units in the deleted 57bp region, putative sites for RNA binding proteins were investigated in the region. This study reports differential RNA-RBP complex formation within this region between the C- and T-allelic forms of the *COL5A1* 3′-UTR. The putative RBPs believed to bind differentially within the 57bp region of the 3′-UTR are primarily involved in translation initiation, RNA biogenesis and/or in shuttling the RNA between the nucleus and cytoplasm (Glisovic et al., 2008;Hogan et al., 2008;Ross, 1995). A distinct signal was detected when both the nuclear and cytoplasmic extract was assayed with the C-allelic RNA probe but not the T-allelic RNA probe. This indicates that a protein(s), present in both the cytoplasm and the nucleus, preferentially binds the C-allelic RNA probe i.e. is able to distinguish

between the two forms, accounting for the tendinopathic phenotype. Further work is needed to determine the specificity of the different RBP-RNA binding complexes observed and the identity of the RBP(s) responsible for these findings. It is imperative that these include competition assays with non-biotinylated RNA probes. Further studies should aim at repeating the experiments in a relevant cell line such as the primary skin fibroblast cell lines or tenocytes cell lines. In addition, the identity of the putative RBP(s) involved in the complex formation should be determined by the use of antibodies against the RBPs detected via *in silico* analysis of the region and/or by systematically mutating the putative RBP binding sites. In light of the mRNA expression and mRNA stability studies, it would be interesting to determine if the RBP complex associated with the C-allelic probe leads to increased *COL5A1* mRNA degradation i.e. whether it is responsible for the decrease in C-allelic mRNA stability compared to the T-allelic form.

4.6 In Perspective

As reviewed in chapter 1, musculoskeletal soft tissue injuries and other exercise-associated phenotypes are multi-factorial and multigenic in nature. Numerous intrinsic and extrinsic factors interact together to determine the biological variations within these phenotypes. Specifically the common C to T rs12722 polymorphism (*Bst*UI RFLP) within the *COL5A1* 3’-UTR has previously been associated with a number of sports and occupational injuries as well as, performance-related phenotypes (Abrahams et al., 2013;Brown et al., 2011;Brown et al., 2011;Burger et al., 2014;Mokone et al., 2006;O'Connell et al., 2013;Posthumus et al., 2009;Posthumus et al., 2010;September et al., 2009). Although seemingly unrelated,

these phenotypes are all directly or indirectly associated with the mechanical properties of musculoskeletal soft tissue. It has therefore been hypothesised that the mechanical properties of musculoskeletal tissue are determined, at least in part, by variants within the *COL5A1* gene (Collins and Posthumus, 2011). Indeed, the reported associations imply that there is an increased type V collagen production among individuals with a *COL5A1* rs12722 TT genotype, which results in structural and architectural changes within the collagen fibril. Furthermore, these changes result in altered mechanical properties of musculoskeletal soft tissues, which in turn associate with increased risk of specific injuries and muscle damage, as well as, reduced joint ROM (flexibility), and increased endurance running ability (Collins and Posthumus, 2011).

The novel findings of this thesis supported the hypothesis of Collins and Posthumus (2011). The *COL5A1* rs12722 TT genotype was associated with increased expression of the *COL5A1* gene in primary human skin fibroblasts, while the *COL5A1* rs12722 T allele was associated with increased mRNA stability. In addition two major functional forms of the *COL5A1* 3'-UTR were identified and three additional polymorphisms, rs71746744 (-/AGGG), rs16399 (ATCT/-) and rs1134170 (A/T), within the 3'-UTR were shown to be independently associated with chronic Achilles tendinopathy. Specifically, the AGGG/AGGG, -/- and TT genotypes of rs71746744, rs16399 and rs1134170 respectively, were associated with increased risk. As illustrated in figure 4.3, the four independently associated polymorphisms appear to alter the predicted secondary structure of the 3'-UTR (Abrahams et al., 2013). It is proposed that changes in the secondary structure may affect the post-transcriptional regulation of the mRNA increasing the overall mRNA stability and

gene expression of *COL5A1*; and by implication, type V collagen production. Since type V collagen regulates fibrillogenesis, this may in turn affect the mechanical properties of connective tissue. As a result, the tissue with reduced risk for specific injuries and possibly muscle damage, as well as, increased joint ROM (flexibility), and decreased endurance running ability, will contain larger less densely packed fibrils. The tissue containing smaller more densely packed fibrils on the other hand will be at increased risk of specific injuries and muscle damage, as well as, reduced joint ROM (flexibility), and increased endurance running ability (Collins and Posthumus, 2011;Ribbans and Collins, 2013) (Figure 4.3).

Although the molecular mechanisms underlying these observations are not yet understood this thesis started investigating possible regulatory mechanisms and suggests future directions for research. Evidence in mouse skin matrix (Chanut-Delalande et al., 2004), from *Col5a2* targeted deletion experiments, highlights a conceivable manner in which this could occur. An increased stability of COL5A1 mRNA may lead to an overexpression of the α1(V) chain and a resulting increase in the $\alpha1(V)_3$ homotrimers. As displayed in their experiments, this may favour the assembly of the latter as opposed to the major isoform of type V collagen, $\alpha1(V)_2\alpha2(V)$, and lead to atypical assembly of heterotypic collagen fibrils. Transgenic mouse lines overexpressing the human pro-α1(V) chain in the epidermis (Bonod-Bidaud et al.,2012) displayed an accumulation of unmasked $\alpha1(V)_3$ fibrils at the epidermis–dermis interface. Ultrastructural modifications and changes in biomechanical properties were observed, although not statistically significant. To examine this possibility further, prospective studies should measure the mRNA expression of *COL5A2* and the ratios between COL5A1/COL5A2 mRNA expressions

as well as protein expression, that is, the α1(V) and α2(V) chain synthesis and ratios thereof.

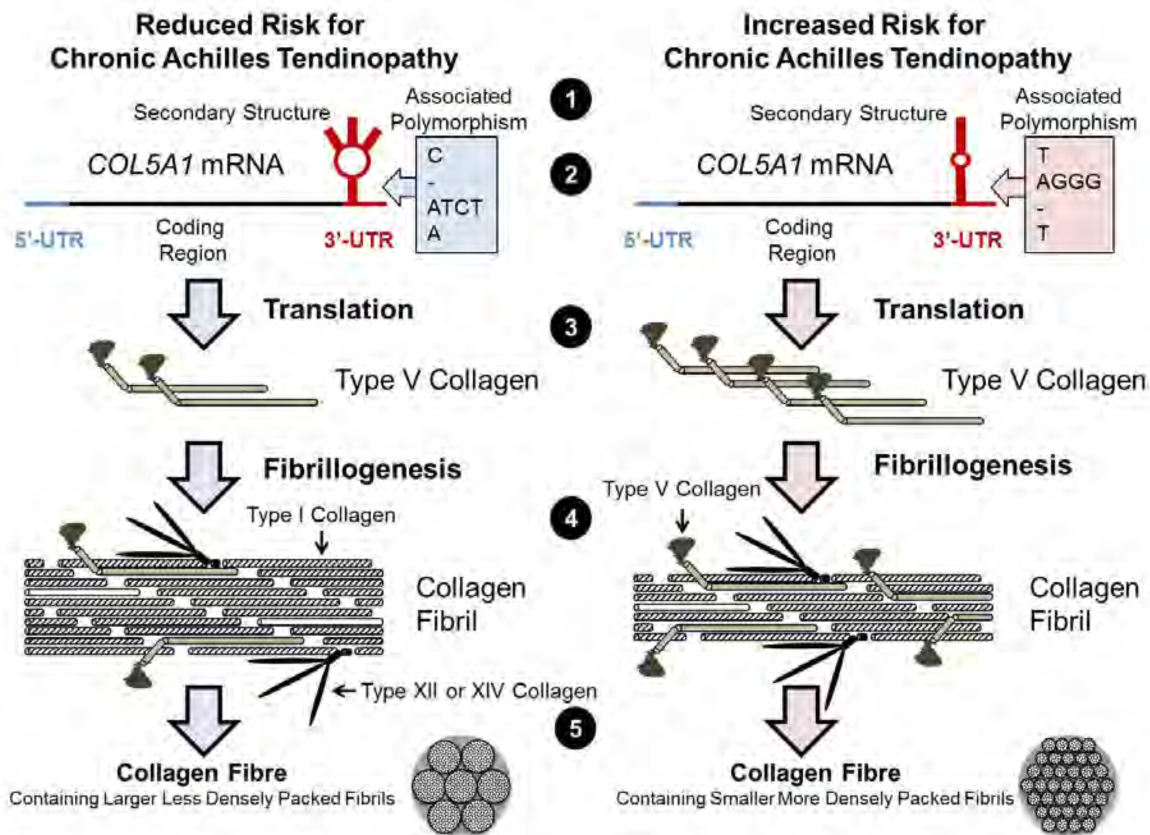


Figure 4.3 A schematic diagram representing the proposed mechanism of how *COL5A1* may increase (right panel) and decrease (left panel) the risk of chronic Achilles tendinopathy. Firstly (1) the four polymorphisms rs12722 (C/T), rs71746744 (-/AGGG), rs16399 (ATCT/-) and rs1134170 (A/T) within the 3'-UTR are independently associated with tendinopathy and (2) appear to alter the predicted secondary structure of the 3'-UTR. (3) The post-transcriptional regulation of the mRNA affects the mRNA stability as well as the gene expression of *COL5A1* and by implication will affect type V collagen production and (4) since type V collagen regulates fibrillogenesis, this may affect the mechanical properties of tendons. As a result, (5) the tendon with reduced risk of injury will contain larger less densely packed fibrils while the tendon with increased risk may contain smaller more densely packed fibrils. Modified from Collins and Posthumus (2011) and Ribbans and Collins (2013).

In addition, since type V and type XI collagen shares structural and functional homologies, Hay *et al.* (2013) investigated whether functional polymorphisms within the genes encoding type XI (*COL11A1* and *COL11A2*) and type V (*COL5A1*) collagens interact with one another to collectively modulate the risk for chronic Achilles tendinopathy. Specifically, the authors reported the association of the TCT(AGGG) inferred pseudohaplotype, constructed from the *COL11A1* (rs3753841 T/C and rs1676486 C/T) , *COL11A2* (rs17999079 T/A) and the *COL5A1* rs71746744 (-/AGGG) genes, respectively, with an increased risk of chronic Achilles tendinopathy (Hay et al., 2013). Interestingly, the rs1676486 polymorphism within exon 62 of *COL11A1*, which is associated with lumbar disc herniation in a Japanese population, results in an amino acid substitution (Pro1535Ser), which could hypothetically cause a conformational change in type XI collagen (Mio et al., 2007). More importantly, the T allele of rs1676486 was also associated with increased rate of mRNA degradation in their experiments.

Type XI collagen is produced during tendon development but whether it is produced in the mature healthy, injured or in the healing tendon is currently unclear (Wenstrup et al., 2011). Indeed, Wenstrup *et al.* (2011) demonstrated coordinate roles for collagens V and XI in the regulation of fibril nucleation and assembly during tendon development. In light of the findings presented in this thesis, the combined findings by Hay *et al.* (2013) and Mio *et al.* (2007) as well as the documented homology in structure and function of the two collagen types (Fichard et al., 1994;Wenstrup et al., 2011), it is tempting to speculate, granting the multigenic and multifactorial nature of soft tissue injuries, that type V and XI collagen may interact in modulating the risk for these injuries. Hence, type XI collagen was integrated into the hypothesis described

above (Collins and Posthumus, 2011;Hay et al., 2013) (Figure 4.4). As previously mentioned, the *COL5A1* rs71746744 (-/AGGG) and *COL11A1* rs1676486 (C/T) polymorphisms are part of an inferred pseudohaplotype that is associated with chronic Achilles tendinopathy. The *COL5A1* rs71746744 deletion (-) allele and the *COL11A1* rs1676486 T allele are associated with increased mRNA degradation as described (Figure 4.4; left panel). The altered mRNA stability associated with these polymorphisms is believed to result in altered α1(V) and α1(XI) chains and types V and XI collagen production. Since types V and XI collagen regulate collagen fibril assembly and diameter (fibrillogenesis), this may alter the mechanical properties of tendons. Given the inverse relationship between the types V and XI collagen content of the fibril and its diameter during tendon development (Wenstrup et al., 2011), thinner and more densely packed collagen fibrils, which are associated with chronic Achilles tendinopathy (Collins and Posthumus, 2011), are expected due to the increased production of types V and XI collagen (right panel). The association of these type XI collagen gene polymorphisms, as well as their interaction with the *COL5A1* 3'-UTR polymorphisms, with the other injury and performance-related phenotypes remains to be investigated.

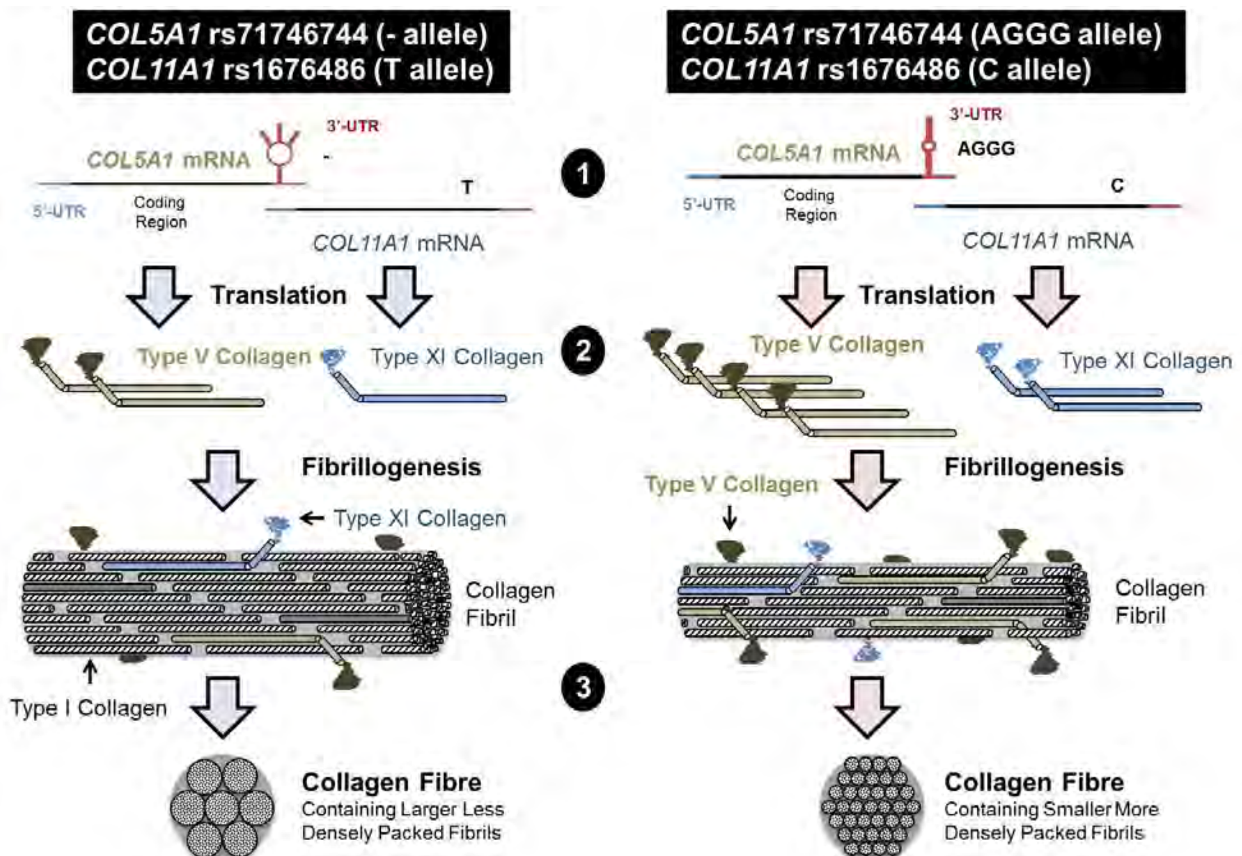


Figure 4.4 A schematic diagram representing the proposed mechanism by which polymorphisms within *COL5A1* and *COL11A1* potentially affect fibrillogenesis (Hay *et al.*, 2013). (1) The *COL5A1* rs71746744 and *COL11A1* rs1676486 polymorphisms are part of an inferred pseudohaplotype that is associated with chronic Achilles tendinopathy where the *COL5A1* rs71746744 (-/AGGG) deletion (-) allele and the *COL11A1* rs1676486 (C/T) T allele are associated with increased mRNA degradation (left panel). The scenario depicting decreased mRNA degradation is indicated in the right panel. (2) 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 (3) Types V and XI collagen regulate collagen fibril assembly and diameter (fibrillogenesis), hence altering the mechanical properties of tendons. Of note, there is an inverse relationship between the types V and XI collagen content of the fibril and its diameter during development as described in Wenstrup *et al.* (2011). Thinner, more densely packed collagen fibrils will be produced due to the increased production of types V and XI collagen (right panel) where thinner fibrils are associated with chronic Achilles tendinopathy (Collins and Posthumus, 2011).

As reviewed in chapter 1, the exact biological mechanisms that cause tendinopathies are poorly understood. Several theories have however been proposed, which include the mechanical theory (Arnoczky et al., 2007;Wren et al., 2003), vascular theory (Fenwick et al., 2002;Langberg et al., 1998), the 'neurogenic' hypothesis (Alfredson et al., 1999;Andersson et al., 2008;Bjur et al., 2005), the iceberg theory (Fredberg and Stengaard-Pedersen, 2008) and the pathology continuum theory (Cook and Purdam, 2009). Although alterations in the structure and mechanical properties of the collagen fibril instinctively appear to support the models that propose that mechanical overuse of tendon is the primary inciting event in tendinopathy (the mechanical and iceberg theories), biological variation in the structure of the collagen fibril is likewise consistent with any of the theories that propose that the cellular responses within the tendon are the primary initiating event. Tenocytes and other soft tissue cells are able to detect and respond to the mechanical load applied to the collagen fibrils in the extracellular matrix of tendons (Arnoczky et al., 2002;Wall and Banes, 2005). Due to the physical link of the ECM to the cytoskeletal proteins in the tenocytes via integrin and other transmembrane receptors, these cells are able to respond appropriately or pathologically to mechanical loading (Khan and Scott, 2009). Type V collagen has specifically been reported to be involved in cellular remodelling in response to mechanical forces (Nakatani et al., 2002) as well as in apoptosis and angiogenesis (Luparello and Sirchia, 2005;Merwin et al., 1990;Nardo et al., 2014;Souza et al., 2010), demonstrating its importance in the cellular microenvironment.

4.7 CONCLUSION

Common polymorphisms within *COL5A1* 3'-UTR are believed to result in alterations in type V collagen production. Changes in the expression of type V collagen is hypothesised to affect fibrillogenesis and by implication the mechanical properties of musculoskeletal soft tissues. Variants within the 3'-UTR have been associated with recreational and occupational musculoskeletal soft tissue injuries, as well as other exercise-associated phenotypes. Since the function of the *COL5A1* 3'-UTR is unknown and the region contains several putative regulatory elements, the primary aim of this thesis was to determine whether the *COL5A1* 3'-UTR was functional and whether this genomic region contains common polymorphisms responsible for measurable functional variations that may regulate its gene mRNA stability and expression. This is relevant because changes in mRNA expression of *COL5A1* may lead to changes in the α1(V) chain protein production and the synthesis of type V collagen.

Two major functional forms of the *COL5A1* 3'-UTR were identified in this thesis. In addition three additional polymorphisms, namely, rs71746744 (-/AGGG), rs16399 (ATCT/-) and rs1134170 (A/T), downstream of rs12722 (T/C), were identified within the functional forms and shown to independently associate with chronic Achilles tendinopathy. The C-functional form, which corresponded to the wild type sequence, was predominately identified in asymptomatic controls with the "protective" CC rs12722 genotype and was associated with decreased mRNA stability and *COL5A1* gene expression. The T-functional form was however predominately identified in the

chronic Achilles tendinopathic patients with the ‘at risk’ TT genotype and was associated with increased mRNA stability and *COL5A1* gene expression.

The major functional region of the *COL5A1* 3’-UTR, responsible for the tendinopathic phenotype, was mapped to a 57bp region containing rs71746744 (-/AGGG) and several putative RNA binding protein motifs. Preliminary analysis of this region showed differential RNA-RBP complex formation within this region between the C- and T-allelic forms of the *COL5A1* 3′-UTR. Finally, this thesis provides molecular evidence for the proposed hypothesis that polymorphisms within the *COL5A1* 3’-UTR alter fibril architecture and structure and, thereby, mechanical properties of musculoskeletal soft tissues.

BIBLIOGRAPHY

Abate, M., Silbernagel, K. G., Siljeholm, C., Di Iorio, A., De Amicis, D., Salini, V., Werner, S. and Paganelli, R. (2009). Pathogenesis of Tendinopathies: Inflammation Or Degeneration? *Arthritis Res. Ther.* **11,** 235.

Abdellatif, M. (2012). Differential Expression of microRNAs in Different Disease States. *Circ. Res.* **110,** 638-650.

Abrahams, Y. (2013). *The Regulation of the COL5A1 Gene Via the 3'-UTR and its Impact on Achilles Tendinopathy and Other Exercise-Related Phenotypes.* Unpublished Masters Dissertation. University of Cape Town.

Abrahams, S., Posthumus, M. and Collins, M. (2014). A Polymorphism in a Functional Region of the COL5A1 Gene: Association with Ultraendurance-Running Performance and Joint Range of Motion. *Int. J. Sports Physiol. Perform.* **9,** 583-590.

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. *Ann. Hum. Genet.* **77,** 204-214.

Akai, J., Kimura, A. and Hata, R. I. (1999). Transcriptional Regulation of the Human Type I Collagen Alpha2 (COL1A2) Gene by the Combination of Two Dinucleotide Repeats. *Gene* **239,** 65-73.

Akhtar, N., Rasheed, Z., Ramamurthy, S., Anbazhagan, A. N., Voss, F. R. and Haqqi, T. M. (2010). MicroRNA-27b Regulates the Expression of Matrix Metalloproteinase 13 in Human Osteoarthritis Chondrocytes. *Arthritis Rheum.* **62,** 1361-1371.

Albers, D. and Hoke, B. (2003). Techniques in Achilles Tendon Rehabilitation. *Tech Foot Ankle Surg* **2,** 208.

Alfredson, H., Forsgren, S., Thorsen, K., Fahlstrom, M., Johansson, H. and Lorentzon, R. (2001). Glutamate NMDAR1 Receptors Localised to Nerves in Human Achilles Tendons. Implications for Treatment? *Knee Surg. Sports Traumatol. Arthrosc.* **9,** 123-126.

Alfredson, H., Ljung, B. O., Thorsen, K. and Lorentzon, R. (2000). In Vivo Investigation of ECRB Tendons with Microdialysis Technique--no Signs of Inflammation but High Amounts of Glutamate in Tennis Elbow. *Acta Orthop. Scand.* **71,** 475-479.

Alfredson, H., Lorentzon, M., Backman, S., Backman, A. and Lerner, U. H. (2003). cDNA-Arrays and Real-Time Quantitative PCR Techniques in the Investigation of Chronic Achilles Tendinosis. *J. Orthop. Res.* **21,** 970-975.

Alfredson, H., Thorsen, K. and Lorentzon, R. (1999). In Situ Microdialysis in Tendon Tissue: High Levels of Glutamate, but Not Prostaglandin E2 in Chronic Achilles Tendon Pain. *Knee Surg. Sports Traumatol. Arthrosc.* **7,** 378-381.

Altuvia, Y., Landgraf, P., Lithwick, G., Elefant, N., Pfeffer, S., Aravin, A., Brownstein, M. J., Tuschl, T. and Margalit, H. (2005). Clustering and Conservation Patterns of Human microRNAs. *Nucleic Acids Res.* **33,** 2697-2706.

Ambros, V. (2004). The Functions of Animal microRNAs. *Nature* **431,** 350-355.

Andarawis-Puri, N., Sereysky, J. B., Sun, H. B., Jepsen, K. J. and Flatow, E. L. (2012). Molecular Response of the Patellar Tendon to Fatigue Loading Explained in the Context of the Initial Induced Damage and Number of Fatigue Loading Cycles. *J. Orthop. Res.* **30,** 1327-1334.

Andersson, G., Danielson, P., Alfredson, H. and Forsgren, S. (2008). Presence of Substance P and the Neurokinin-1 Receptor in Tenocytes of the Human Achilles Tendon. *Regul. Pept*. **150**, 81-87.

Arnoczky, S. P., Lavagnino, M. and Egerbacher, M. (2007). The Mechanobiological Aetiopathogenesis of Tendinopathy: Is it the Over-Stimulation Or the Under-Stimulation of Tendon Cells? *Int. J. Exp. Pathol.* **88,** 217-226.

Arnoczky, S. P., Tian, T., Lavagnino, M., Gardner, K., Schuler, P. and Morse, P. (2002). Activation of Stress-Activated Protein Kinases (SAPK) in Tendon Cells Following Cyclic Strain: The Effects of Strain Frequency, Strain Magnitude, and Cytosolic Calcium. *J. Orthop. Res.* **20,** 947-952.

Asplund, C. A. and Best, T. M. (2013). Achilles Tendon Disorders. *BMJ* **346:**f1262.

Bahr, R. and Holme, I. (2003). Risk Factors for Sports Injuries--a Methodological Approach. *Br. J. Sports Med.* **37,** 384-392.

Bahr, R. and Krosshaug, T. (2005). Understanding Injury Mechanisms: A Key Component of Preventing Injuries in Sport. *Br. J. Sports Med.* **39,** 324-329.

Baldwin, A. K., Simpson, A., Steer, R., Cain, S. A. and Kielty, C. M. (2013). Elastic Fibres in Health and Disease. *Expert Rev. Mol. Med.* **15,** e8.

Bangsbo, J., Junge, A., Dvorak, J. and Krustrup, P. (2014). Executive Summary: Football for Health - Prevention and Treatment of Non-Communicable Diseases Across the Lifespan through Football. *Scand. J. Med. Sci. Sports* **24 Suppl 1,** 147-150.

Banos, C. C., Thomas, A. H. and Kuo, C. K. (2008). Collagen Fibrillogenesis in Tendon Development: Current Models and Regulation of Fibril Assembly. *Birth Defects Res. C. Embryo. Today* **84,** 228-244.

Barr, A. E. and Barbe, M. F. (2002). Pathophysiological Tissue Changes Associated with Repetitive Movement: A Review of the Evidence. *Phys. Ther.* **82,** 173-187.

Barrett, L. W., Fletcher, S. and Wilton, S. D. (2012). Regulation of Eukaryotic Gene Expression by the Untranslated Gene Regions and Other Non-Coding Elements. *Cell Mol. Life Sci.* **69,** 3613-3634.

Bartel, D. P. (2004). MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell* **116,** 281-297.

Bartel, D. P. (2009). MicroRNAs: Target Recognition and Regulatory Functions. *Cell* **136,** 215-233.

Baskerville, S. and Bartel, D. P. (2005). Microarray Profiling of microRNAs Reveals Frequent Coexpression with Neighboring miRNAs and Host Genes. *RNA* **11,** 241-247.

Baumgarten, I. (1993). A Comparison of Metabolic Pathway Dynamics in Man and Other Mammals. Cape Technikon Theses & Dissertations. **96**.

Beighton, P., De Paepe, A., Steinmann, B., Tsipouras, P. and Wenstrup, R. J. (1998). Ehlers-Danlos Syndromes: Revised Nosology, Villefranche, 1997. Ehlers-Danlos National Foundation (USA) and Ehlers-Danlos Support Group (UK). *Am. J. Med. Genet.* **77,** 31-37.

Bell, R. D., Shultz, S. J., Wideman, L. and Henrich, V. C. (2012). Collagen Gene Variants Previously Associated with Anterior Cruciate Ligament Injury Risk are also Associated with Joint Laxity. *Sports Health.* **4,** 312-318.

Benke, K., Agg, B., Szilveszter, B., Tarr, F., Nagy, Z. B., Polos, M., Daroczi, L., Merkely, B. and Szabolcs, Z. (2013). The Role of Transforming Growth Factor-Beta in Marfan Syndrome. *Cardiol. J.* **20,** 227-234.

Bertuzzi, R., Pasqua, L. A., Bueno, S., Lima-Silva, A. E., Matsuda, M., Marquezini, M. and Saldiva, P. H. (2014). Is the COL5A1 rs12722 Gene Polymorphism Associated with Running Economy? *PLoS One* **9,** e106581.

Bienroth, S., Wahle, E., Suter-Crazzolara, C. and Keller, W. (1991). Purification of the Cleavage and Polyadenylation Factor Involved in the 3'-Processing of Messenger RNA Precursors. *J. Biol. Chem.* **266,** 19768-19776.

Bina, M. (2006). Gene Mapping, Discovery, and Expression: Methods in Molecular Biology. **338,** 21-29.

Birk, D. E. (2001). Type V Collagen: Heterotypic Type I/V Collagen Interactions in the Regulation of Fibril Assembly. *Micron* **32,** 223-237.

Birk, D. E., Fitch, J. M., Babiarz, J. P., Doane, K. J. and Linsenmayer, T. F. (1990). Collagen Fibrillogenesis in Vitro: Interaction of Types I and V Collagen Regulates Fibril Diameter. *J. Cell. Sci.* **95 (Pt 4),** 649-657.

Birk, D. E. and Trelstad, R. L. (1986). Extracellular Compartments in Tendon Morphogenesis: Collagen Fibril, Bundle, and Macroaggregate Formation. *J. Cell Biol.* **103,** 231-240.

Birkedal-Hansen, H., Moore, W. G., Bodden, M. K., Windsor, L. J., Birkedal-Hansen, B., DeCarlo, A. and Engler, J. A. (1993). Matrix Metalloproteinases: A Review. *Crit. Rev. Oral Biol. Med.* **4,** 197-250.

Bjur, D., Alfredson, H. and Forsgren, S. (2005). The Innervation Pattern of the Human Achilles Tendon: Studies of the Normal and Tendinosis Tendon with Markers for General and Sensory Innervation. *Cell Tissue Res.* **320,** 201-206.

Bohnsack, M. T., Czaplinski, K. and Gorlich, D. (2004). Exportin 5 is a RanGTP-Dependent dsRNA-Binding Protein that Mediates Nuclear Export of Pre-miRNAs. *RNA* **10,** 185-191.

Bonod-Bidaud C., Roulet M., Hansen U., Elsheikh A., Malbouyres M., Ricard-Blum S., Faye C., Vaganay E., Rousselle P., Ruggiero F. (2012) *In Vivo* Evidence for a Bridging Role of a Collagen V Subtype at the Epidermis–Dermis Interface. *J Invest Dermatol.* **132(7),**1841-1849.

Boot-Handford, R. P., Tuckwell, D. S., Plumb, D. A., Rock, C. F. and Poulsom, R. (2003). A Novel and Highly Conserved Collagen (Pro(Alpha)1(XXVII)) with a Unique Expression Pattern and Unusual Molecular Characteristics Establishes a New Clade within the Vertebrate Fibrillar Collagen Family. *J. Biol. Chem.* **278,** 31067-31077.

Brook, M. and Gray, N. K. (2012). The Role of Mammalian Poly(A)-Binding Proteins in Co-Ordinating mRNA Turnover. *Biochem. Soc. Trans.* **40,** 856-864.

Brown, J. C., Miller, C. J., Posthumus, M., Schwellnus, M. P. and Collins, M. (2011). The COL5A1 Gene, Ultra-Marathon Running Performance, and Range of Motion. *Int. J. Sports Physiol. Perform.* **6,** 485-496.

Brown, J. C., Miller, C. J., Posthumus, M., Schwellnus, M. P. and Collins, M. (2011). The COL5A1 Gene, Ultra-Marathon Running Performance, and Range of Motion. *Int. J. Sports Physiol. Perform.* **6,** 485-496.

Brown, J. C., Miller, C. -., Schwellnus, M. P. and Collins, M. (2011). Range of Motion Measurements Diverge with Increasing Age for COL5A1 Genotypes. *Scand. J. Med. Sci. Sports* **21,** e266-e272.

Burger, M., de Wet, H. and Collins, M. (2014). The COL5A1 Gene is Associated with Increased Risk of Carpal Tunnel Syndrome. *Clin. Rheumatol.* **34(4)**:767-74

Butcher, S. E. and Pyle, A. M. (2011). The Molecular Interactions that Stabilize RNA Tertiary Structure: RNA Motifs, Patterns, and Networks. *Acc. Chem. Res.* **44,** 1302-1311.

Castoldi, G., Di Gioia, C. R., Bombardi, C., Catalucci, D., Corradi, B., Gualazzi, M. G., Leopizzi, M., Mancini, M., Zerbini, G., Condorelli, G. et al. (2012). MiR-133a Regulates Collagen 1A1: Potential Role of miR-133a in Myocardial Fibrosis in Angiotensin II-Dependent Hypertension. *J. Cell. Physiol.* **227,** 850-856.

Chapman, K., Takahashi, A., Meulenbelt, I., Watson, C., Rodriguez-Lopez, J., Egli, R., Tsezou, A., Malizos, K. N., Kloppenburg, M., Shi, D. et al. (2008). A Meta-Analysis of European and Asian Cohorts Reveals a Global Role of a Functional SNP in the 5' UTR of GDF5 with Osteoarthritis Susceptibility. *Hum. Mol. Genet.* **17,** 1497-1504.

Chanut-Delalande, H., Bonod-Bidaud, C., Cogne, S., Malbouyres, M., Ramirez, F., Fichard, A., & Ruggiero, F. (2004). Development of a Functional Skin Matrix Requires Deposition of Collagen V Heterotrimers. *Mol. Cell. Bio.* **24(13),** 6049–6057.

Chen, M., Zhang, J. and Manley, J. L. (2010). Turning on a Fuel Switch of Cancer: hnRNP Proteins Regulate Alternative Splicing of Pyruvate Kinase mRNA. *Cancer Res.* **70,** 8977-8980.

Chen, Y. and Varani, G. (2005). Protein Families and RNA Recognition. *FEBS J.* **272,** 2088-2097.

Chiquet, M. and Fambrough, D. M. (1984). Chick Myotendinous Antigen. II. A Novel Extracellular Glycoprotein Complex Consisting of Large Disulfide-Linked Subunits. *J. Cell Biol.* **98,** 1937-1946.

Chiquet, M., Renedo, A. S., Huber, F. and Fluck, M. (2003). How do Fibroblasts Translate Mechanical Signals into Changes in Extracellular Matrix Production? *Matrix Biol.* **22,** 73-80.

Cohen, G. M. (1997). Caspases: The Executioners of Apoptosis. *Biochem. J.* **326 (Pt 1),** 1-16.

Collins, M., Mokone, G. G., September, A. V., van der Merwe, L. and Schwellnus, M. P. (2009). The COL5A1 Genotype is Associated with Range of Motion Measurements. *Scand. J. Med. Sci. Sports*. **19(6)**:803-10.

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

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

Conne, B., Stutz, A. and Vassalli, J. D. (2000). The 3' Untranslated Region of Messenger RNA: A Molecular 'Hotspot' for Pathology? *Nat. Med.* **6,** 637-641.

Cook, J. L., Khan, K. M. and Purdam, C. (2002). Achilles Tendinopathy. *Man. Ther.* **7,** 121-130.

Cook, J. L. and Purdam, C. R. (2009). Is Tendon Pathology a Continuum? A Pathology Model to Explain the Clinical Presentation of Load-Induced Tendinopathy. *Br. J. Sports Med.* **43,** 409-416.

Cooke, C., Hans, H. and Alwine, J. C. (1999). Utilization of Splicing Elements and Polyadenylation Signal Elements in the Coupling of Polyadenylation and Last-Intron Removal. *Mol. Cell. Biol.* **19,** 4971-4979.

Craib, M. W., Mitchell, V. A., Fields, K. B., Cooper, T. R., Hopewell, R. and Morgan, D. W. (1996). The Association between Flexibility and Running Economy in Sub-Elite Male Distance Runners. *Med. Sci. Sports Exerc.* **28,** 737-743.

Dai, J., Shi, D., Zhu, P., Qin, J., Ni, H., Xu, Y., Yao, C., Zhu, L., Zhu, H., Zhao, B. et al. (2008). Association of a Single Nucleotide Polymorphism in Growth Differentiate Factor 5 with Congenital Dysplasia of the Hip: A Case-Control Study. *Arthritis Res. Ther.* **10,** R126.

de Haan, J. B., Gevers, W. and Parker, M. I. (1986). Effects of Sodium Butyrate on the Synthesis and Methylation of DNA in Normal Cells and their Transformed Counterparts. *Cancer Res.* **46,** 713-716.

Denli, A. M., Tops, B. B., Plasterk, R. H., Ketting, R. F. and Hannon, G. J. (2004). Processing of Primary microRNAs by the Microprocessor Complex. *Nature* **432,** 231-235.

Ding, Y., Chan, C. Y. and Lawrence, C. E. (2004). Sfold Web Server for Statistical Folding and Rational Design of Nucleic Acids. *Nucleic Acids Res.* **32,** W135-41.

Dodd, A. W., Syddall, C. M. and Loughlin, J. (2013). A Rare Variant in the Osteoarthritis-Associated Locus GDF5 is Functional and Reveals a Site that can be Manipulated to Modulate GDF5 Expression. *Eur. J. Hum. Genet.* **21,** 517-521.

Dressler, M. R., Butler, D. L., Wenstrup, R., Awad, H. A., Smith, F. and Boivin, G. P. (2002). A Potential Mechanism for Age-Related Declines in Patellar Tendon Biomechanics. *J. Orthop. Res.* **20,** 1315-1322.

Du, B., Ma, L. M., Huang, M. B., Zhou, H., Huang, H. L., Shao, P., Chen, Y. Q. and Qu, L. H. (2010). High Glucose Down-Regulates miR-29a to Increase Collagen IV Production in HK-2 Cells. *FEBS Lett.* **584,** 811-816.

Dudek, K. A., Lafont, J. E., Martinez-Sanchez, A. and Murphy, C. L. (2010). Type II Collagen Expression is Regulated by Tissue-Specific miR-675 in Human Articular Chondrocytes. *J. Biol. Chem.* **285,** 24381-24387.

Edeleva, E. V. and Shcherbata, H. R. (2013). Stress-Induced ECM Alteration Modulates Cellular microRNAs that Feedback to Readjust the Extracellular Environment and Cell Behavior. *Front. Genet.* **4,** 305.

Edwalds-Gilbert, G., Veraldi, K. L. and Milcarek, C. (1997). Alternative Poly(A) Site Selection in Complex Transcription Units: Means to an End? *Nucleic Acids Res.* **25,** 2547-2561.

El Khoury, L., Posthumus, M., Cook, J., Handley, C., Collins, M. and Raleigh, S. M. (In Press). *ELN* and *FBN2* Gene Variants as Risk Factors for Two Sports Related Musculoskeletal Injuries. *Int J Sports Med.*

El Khoury, L., Posthumus, M., Collins, M., Handley, C. J., Cook, J. and Raleigh, S. M. (2013). Polymorphic Variation within the ADAMTS2, ADAMTS14, ADAMTS5, ADAM12 and TIMP2 Genes and the Risk of Achilles Tendon Pathology: A Genetic Association Study. *J. Sci. Med. Sport* **16,** 493-498.

Ezzelle, J., Rodriguez-Chavez, I. R., Darden, J. M., Stirewalt, M., Kunwar, N., Hitchcock, R., Walter, T. and D'Souza, M. P. (2008). Guidelines on Good Clinical Laboratory Practice: Bridging Operations between Research and Clinical Research Laboratories. *J. Pharm. Biomed. Anal.* **46,** 18-29.

Fenwick, S. A., Hazleman, B. L. and Riley, G. P. (2002). The Vasculature and its Role in the Damaged and Healing Tendon. *Arthritis Res.* **4,** 252-260.

Fichard, A., Kleman, J. and Ruggiero, F. (1994). Another Look at Collagen V and XI Molecules. *Matrix Biol.* **14,** 515 - 531.

Finnegan, E. F. and Pasquinelli, A. E. (2013). MicroRNA Biogenesis: Regulating the Regulators. *Crit. Rev. Biochem. Mol. Biol.* **48,** 51-68.

Foster, B. P., Morse, C. I., Onambele, G. L. and Williams, A. G. (2014). Human COL5A1 rs12722 Gene Polymorphism and Tendon Properties in Vivo in an Asymptomatic Population. *Eur. J. Appl. Physiol.* **114,** 1393-1402.

Fredberg, U. and Stengaard-Pedersen, K. (2008). Chronic Tendinopathy Tissue Pathology, Pain Mechanisms, and Etiology with a Special Focus on Inflammation. *Scand. J. Med. Sci. Sports* **18,** 3-15.

Fukuta, S., Oyama, M., Kavalkovich, K., Fu, F. H. and Niyibizi, C. (1998). Identification of Types II, IX and X Collagens at the Insertion Site of the Bovine Achilles Tendon. *Matrix Biol.* **17,** 65-73.

Fung, D. T., Wang, V. M., Andarawis-Puri, N., Basta-Pljakic, J., Li, Y., Laudier, D. M., Sun, H. B., Jepsen, K. J., Schaffler, M. B. and Flatow, E. L. (2010). Early Response to Tendon Fatigue Damage Accumulation in a Novel in Vivo Model. *J. Biomech.* **43,** 274-279.

Gaida, J. E., Alfredson, H., Kiss, Z. S., Bass, S. L. and Cook, J. L. (2010). Asymptomatic Achilles Tendon Pathology is Associated with a Central Fat Distribution in Men and a Peripheral Fat Distribution in Women: A Cross Sectional Study of 298 Individuals. *BMC Musculoskelet. Disord.* **11,** 41-2474-11-41.

Gaida, J. E., Ashe, M. C., Bass, S. L. and Cook, J. L. (2009). Is Adiposity an Under-Recognized Risk Factor for Tendinopathy? A Systematic Review. *Arthritis Rheum.* **61,** 840-849.

Gaunt, T. R., Rodriguez, S. and Day, I. N. (2007). Cubic Exact Solutions for the Estimation of Pairwise Haplotype Frequencies: Implications for Linkage Disequilibrium Analyses and a Web Tool 'CubeX'. *BMC Bioinformatics* **8,** 428.

Gelse, K., Pöschl, E. and Aigner, T. (2003). Collagens—structure, Function, and Biosynthesis. *Adv. Drug Deliv. Rev.* **55,** 1531-1546.

Geyer, M. (2005). Achillodynia. *Orthopade* **34,** 677-681.

Gleim, G. W., Stachenfeld, N. S. and Nicholas, J. A. (1990). The Influence of Flexibility on the Economy of Walking and Jogging. *J. Orthop. Res.* **8,** 814-823.

Glisovic, T., Bachorik, J. L., Yong, J. and Dreyfuss, G. (2008). RNA-Binding Proteins and Post-Transcriptional Gene Regulation. *FEBS Lett.* **582,** 1977-1986.

Goncalves-Neto, J., Witzel, S. S., Teodoro, W. R., Carvalho-Junior, A. E., Fernandes, T. D. and Yoshinari, H. H. (2002). Changes in Collagen Matrix Composition in Human Posterior Tibial Tendon Dysfunction. *Joint Bone Spine* **69,** 189-194.

Grahame, R. (1999). Joint Hypermobility and Genetic Collagen Disorders: Are they Related? *Arch. Dis. Child.* **80,** 188-191.

Graves, P. and Zeng, Y. (2012). Biogenesis of Mammalian microRNAs: A Global View. *Genomics Proteomics Bioinformatics* **10,** 239-245.

Gregory, R. I., Yan, K. P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N. and Shiekhattar, R. (2004). The Microprocessor Complex Mediates the Genesis of microRNAs. *Nature* **432,** 235-240.

Griffiths-Jones, S. (2004). The microRNA Registry. *Nucleic Acids Res.* **32,** D109-11.

Griffiths-Jones, S., Grocock, R. J., van Dongen, S., Bateman, A. and Enright, A. J. (2006). miRBase: microRNA Sequences, Targets and Gene Nomenclature. *Nucleic Acids Res.* **34,** D140-4.

Gruter, P., Tabernero, C., von Kobbe, C., Schmitt, C., Saavedra, C., Bachi, A., Wilm, M., Felber, B. K. and Izaurralde, E. (1998). TAP, the Human Homolog of Mex67p, Mediates CTE-Dependent RNA Export from the Nucleus. *Mol. Cell* **1,** 649-659.

Guller, I. and Russell, A. P. (2010). MicroRNAs in Skeletal Muscle: Their Role and Regulation in Development, Disease and Function. *J. Physiol.* **588,** 4075-4087.

Han, S., Makareeva, E., Kuznetsova, N. V., DeRidder, A. M., Sutter, M. B., Losert, W., Phillips, C. L., Visse, R., Nagase, H. and Leikin, S. (2010). Molecular Mechanism of Type I Collagen Homotrimer Resistance to Mammalian Collagenases. *J. Biol. Chem.* **285,** 22276-22281.

Hanahan, D. (1983). Studies on Transformation of Escherichia Coli with Plasmids. *Journal of Molecular Biology,* **166,** 557-580.

Hay, M., Patricios, J., Collins, R., Branfield, A., Cook, J., Handley, C. J., September, A. V., Posthumus, M. and Collins, M. (2013). Association of Type XI Collagen Genes with Chronic Achilles Tendinopathy in Independent Populations from South Africa and Australia. *Br. J. Sports Med.* **47,** 569-574.

Heilig, J. S., Elbing, K. and Brent, C. (1998). Large Scale Preparation of Plasmid DNA. In *Current Protocols in Molecular Biology* (ed. F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith and K. struhl), pp. 1.7.7-1.7.16. New York, USA: John Wiley & Sons Inc.

Hiard, S., Charlier, C., Coppieters, W., Georges, M. and Baurain, D. (2010). Patrocles: A Database of Polymorphic miRNA-Mediated Gene Regulation in Vertebrates. *Nucleic Acids Res.* **38,** D640-51.

Hilleren, P. and Parker, R. (1999). mRNA Surveillance in Eukaryotes: Kinetic Proofreading of Proper Translation Termination as Assessed by mRNP Domain Organization? *RNA* **5,** 711-719.

Hogan, D. J., Riordan, D. P., Gerber, A. P., Herschlag, D. and Brown, P. O. (2008). Diverse RNA-Binding Proteins Interact with Functionally Related Sets of RNAs, Suggesting an Extensive Regulatory System. *PLoS Biol.* **6,** e255.

Huttelmaier, S., Zenklusen, D., Lederer, M., Dictenberg, J., Lorenz, M., Meng, X., Bassell, G. J., Condeelis, J. and Singer, R. H. (2005). Spatial Regulation of Beta-Actin Translation by Src-Dependent Phosphorylation of ZBP1. *Nature* **438,** 512-515.

Jarvinen, T. A., Jozsa, L., Kannus, P., Jarvinen, T. L., Hurme, T., Kvist, M., Pelto-Huikko, M., Kalimo, H. and Jarvinen, M. (2003). Mechanical Loading Regulates the Expression of Tenascin-C in the Myotendinous Junction and Tendon but does Not Induce De Novo Synthesis in the Skeletal Muscle. *J. Cell. Sci.* **116,** 857-866.

Jarvinen, T. A., Jozsa, L., Kannus, P., Jarvinen, T. L., Kvist, M., Hurme, T., Isola, J., Kalimo, H. and Jarvinen, M. (1999). Mechanical Loading Regulates Tenascin-C Expression in the Osteotendinous Junction. *J. Cell. Sci.* **112 Pt 18,** 3157-3166.

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

Jarvinen, T. A., Kannus, P., Paavola, M., Jarvinen, T. L., Jozsa, L. and Jarvinen, M. (2001). Achilles Tendon Injuries. *Curr. Opin. Rheumatol.* **13,** 150-155.

Ji, J., Zhao, L., Budhu, A., Forgues, M., Jia, H. L., Qin, L. X., Ye, Q. H., Yu, J., Shi, X., Tang, Z. Y. et al. (2010). Let-7g Targets Collagen Type I Alpha2 and Inhibits Cell Migration in Hepatocellular Carcinoma. *J. Hepatol.* **52,** 690-697.

Jin, H., van't Hof, R. J., Albagha, O. M. and Ralston, S. H. (2009). Promoter and Intron 1 Polymorphisms of COL1A1 Interact to Regulate Transcription and Susceptibility to Osteoporosis. *Hum. Mol. Genet.* **18,** 2729-2738.

Jones, A. M. (2002). Running Economy is Negatively Related to Sit-and-Reach Test Performance in International-Standard Distance Runners. *Int. J. Sports Med.* **23,** 40-43.

Jones, F. S. and Jones, P. L. (2000). The Tenascin Family of ECM Glycoproteins: Structure, Function, and Regulation during Embryonic Development and Tissue Remodeling. *Dev. Dyn.* **218,** 235-259.

Jones, G. C., Corps, A. N., Pennington, C. J., Clark, I. M., Edwards, D. R., Bradley, M. M., Hazleman, B. L. and Riley, G. P. (2006). Expression Profiling of Metalloproteinases and Tissue Inhibitors of Metalloproteinases in Normal and Degenerate Human Achilles Tendon. *Arthritis Rheum.* **54,** 832-842.

Kader, D., Saxena, A., Movin, T. and Maffulli, N. (2002). Achilles Tendinopathy: Some Aspects of Basic Science and Clinical Management. *Br. J. Sports Med.* **36,** 239-249.

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

Kannus, P. (1997). Etiology and Pathophysiology of Chronic Tendon Disorders in Sports. *Scand. J. Med. Sci. Sports* **7,** 78-85.

Kannus, P. (2000). Structure of the Tendon Connective Tissue. *Scand. J. Med. Sci. Sports* **10,** 312-320.

Kannus, P. and Natri, A. (1997). Etiology and Pathophysiology of Tendon Ruptures in Sports. *Scand. J. Med. Sci. Sports* **7,** 107-112.

Karousou, E., Ronga, M., Vigetti, D., Passi, A. and Maffulli, N. (2008). Collagens, Proteoglycans, MMP-2, MMP-9 and TIMPs in Human Achilles Tendon Rupture. *Clin. Orthop. Relat. Res.* **466,** 1577-1582.

Kaufman, K. R., Brodine, S. K., Shaffer, R. A., Johnson, C. W. and Cullison, T. R. (1999). The Effect of Foot Structure and Range of Motion on Musculoskeletal Overuse Injuries. *Am. J. Sports Med.* **27,** 585-593.

Khan, K. M. and Scott, A. (2009). Mechanotherapy: How Physical Therapists' Prescription of Exercise Promotes Tissue Repair. *Br. J. Sports Med.* **43,** 247-252.

Kim, V. N. and Nam, J. (2006). Genomics of microRNA. *Trends in Genetics* **22,** 165-173.

Koch, M., Laub, F., Zhou, P., Hahn, R. A., Tanaka, S., Burgeson, R. E., Gerecke, D. R., Ramirez, F. and Gordon, M. K. (2003). Collagen XXIV, a Vertebrate Fibrillar Collagen with Structural Features of Invertebrate Collagens: Selective Expression in Developing Cornea and Bone. *J. Biol. Chem.* **278,** 43236-43244.

Krietsch, J., Caron, M. C., Gagne, J. P., Ethier, C., Vignard, J., Vincent, M., Rouleau, M., Hendzel, M. J., Poirier, G. G. and Masson, J. Y. (2012). PARP Activation Regulates the RNA-Binding Protein NONO in the DNA Damage Response to DNA Double-Strand Breaks. *Nucleic Acids Res.* **40,** 10287-10301.

Kubo, K., Akima, H., Ushiyama, J., Tabata, I., Fukuoka, H., Kanehisa, H. and Fukunaga, T. (2004). Effects of 20 Days of Bed Rest on the Viscoelastic Properties of Tendon Structures in Lower Limb Muscles. *Br. J. Sports Med.* **38,** 324-330.

Kujala, U. M., Sarna, S. and Kaprio, J. (2005). Cumulative Incidence of Achilles Tendon Rupture and Tendinopathy in Male Former Elite Athletes. *Clin. J. Sport Med.* **15,** 133-135.

Laguette, M. J., Abrahams, Y., Prince, S. and Collins, M. (2011). Sequence Variants within the 3'-UTR of the COL5A1 Gene Alters mRNA Stability: Implications for Musculoskeletal Soft Tissue Injuries. *Matrix Biol.* **30,** 338-345.

Lahiri, D. K. and Nurnberger, J. I.,Jr. (1991). A Rapid Non-Enzymatic Method for the Preparation of HMW DNA from Blood for RFLP Studies. *Nucleic Acids Res.* **19,** 5444.

Laing, C. and Schlick, T. (2010). Computational Approaches to 3D Modeling of RNA. *J. Phys. Condens Matter* **22,** 283101.

Langberg, H., Bulow, J. and Kjaer, M. (1998). Blood Flow in the Peritendinous Space of the Human Achilles Tendon during Exercise. *Acta Physiol. Scand.* **163,** 149-153.

Lau, N. C. and Lai, E. C. (2005). Diverse Roles for RNA in Gene Regulation. *Genome Biol.* **6,** 315.

Lei, K. F., Liu, B. Y., Wang, Y. F., Chen, X. H., Yu, B. Q., Guo, Y. and Zhu, Z. G. (2011). SerpinB5 Interacts with KHDRBS3 and FBXO32 in Gastric Cancer Cells. *Oncol. Rep.* **26,** 1115-1120.

Lin, D. Y., Hu, Y. and Huang, B. E. (2008). Simple and Efficient Analysis of Disease Association with Missing Genotype Data. *Am. J. Hum. Genet.* **82,** 444-452.

Lincoln, J., Florer, J. B., Deutsch, G. H., Wenstrup, R. J. and Yutzey, K. E. (2006). ColVa1 and ColXIa1 are Required for Myocardial Morphogenesis and Heart Valve Development. *Dev. Dyn.* **235,** 3295-3305.

Lindquist, J. N., Parsons, C. J., Stefanovic, B. and Brenner, D. A. (2004). Regulation of Alpha1(I) Collagen Messenger RNA Decay by Interactions with alphaCP at the 3'-Untranslated Region. *J. Biol. Chem.* **279,** 23822-23829.

Linsenmayer, T. F., Gibney, E., Igoe, F., Gordon, M. K., Fitch, J. M., Fessler, L. I. and Birk, D. E. (1993). 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,** 1181-1189.

Little, J., Higgins, J. P., Ioannidis, J. P., Moher, D., Gagnon, F., von Elm, E., Khoury, M. J., Cohen, B., Davey-Smith, G., Grimshaw, J. et al. (2009). STrengthening the REporting of Genetic Association Studies (STREGA)--an Extension of the STROBE Statement. *Eur. J. Clin. Invest.* **39,** 247-266.

Liu, S. H., Yang, R. S., al-Shaikh, R. and Lane, J. M. (1995). Collagen in Tendon, Ligament, and Bone Healing. A Current Review. *Clin. Orthop. Relat. Res.* **(318),** 265-278.

Liu, X., Yu, J., Jiang, L., Wang, A., Shi, F., Ye, H. and Zhou, X. (2009). MicroRNA-222 Regulates Cell Invasion by Targeting Matrix Metalloproteinase 1 (MMP1) and Manganese Superoxide Dismutase 2 (SOD2) in Tongue Squamous Cell Carcinoma Cell Lines. *Cancer. Genomics Proteomics* **6,** 131-139.

Ljungqvist, A., Schwellnus, M. P., Bachl, N., Collins, M., Cook, J., Khan, K. M., Maffulli, N., Pitsiladis, Y., Riley, G., Golspink, G. et al. (2008). International Olympic Committee Consensus Statement: Molecular Basis of Connective Tissue and Muscle Injuries in Sport. *Clin. Sports Med.* **27,** 231-9, x-xi.

Lopes, A., Hespanhol, L., Jr., Yeung, S. and Costa, L. (2012). What are the Main Running-Related Musculoskeletal Injuries? **42,** 891-905.

Lu, P., Zhang, G. R., Song, X. H., Zou, X. H., Wang, L. L. and Ouyang, H. W. (2011). Col V siRNA Engineered Tenocytes for Tendon Tissue Engineering. *PLoS One* **6,** e21154.

Luna, C., Li, G., Qiu, J., Epstein, D. L. and Gonzalez, P. (2011). Cross-Talk between miR-29 and Transforming Growth Factor-Betas in Trabecular Meshwork Cells. *Invest. Ophthalmol. Vis. Sci.* **52,** 3567-3572.

Lund, E., Guttinger, S., Calado, A., Dahlberg, J. E. and Kutay, U. (2004). Nuclear Export of microRNA Precursors. *Science* **303,** 95-98.

Lunde, B. M., Moore, C. and Varani, G. (2007). RNA-Binding Proteins: Modular Design for Efficient Function. *Nat. Rev. Mol. Cell Biol.* **8,** 479-490.

Luparello, C. and Sirchia, R. (2005). Type V Collagen Regulates the Expression of Apoptotic and Stress Response Genes by Breast Cancer Cells. *J. Cell. Physiol.* **202,** 411-421.

Maes, H. H., Beunen, G. P., Vlietinck, R. F., Neale, M. C., Thomis, M., Vanden Eynde, B., Lysens, R., Simons, J., Derom, C. and Derom, R. (1996). Inheritance of Physical Fitness in 10-Yr-Old Twins and their Parents. *Med. Sci. Sports Exerc.* **28,** 1479-1491.

Magnan, B., Bondi, M., Pierantoni, S. and Samaila, E. (2014). The Pathogenesis of Achilles Tendinopathy: A Systematic Review. *Foot and Ankle Surgery* **20,** 154-159.

Malfait, F., Wenstrup, R. J. and De Paepe, A. (2010). Clinical and Genetic Aspects of Ehlers-Danlos Syndrome, Classic Type. *Genet. Med.* **12,** 597-605.

Mangus, D. A., Evans, M. C. and Jacobson, A. (2003). Poly(A)-Binding Proteins: Multifunctional Scaffolds for the Post-Transcriptional Control of Gene Expression. *Genome Biol.* **4,** 223.

Mann, V., Hobson, E. E., Li, B., Stewart, T. L., Grant, S. F., Robins, S. P., Aspden, R. M. and Ralston, S. H. (2001). A COL1A1 Sp1 Binding Site Polymorphism Predisposes to Osteoporotic Fracture by Affecting Bone Density and Quality. *J. Clin. Invest.* **107,** 899-907.

Martinez-Sanchez, A., Dudek, K. A. and Murphy, C. L. (2012). Regulation of Human Chondrocyte Function through Direct Inhibition of Cartilage Master Regulator SOX9 by microRNA-145 (miRNA-145). *J. Biol. Chem.* **287,** 916-924.

Martinez-Sanchez, A. and Murphy, C. L. (2013). miR-1247 Functions by Targeting Cartilage Transcription Factor SOX9. *J. Biol. Chem.* **288,** 30802-30814.

Matheson, G. O., Klugl, M., Engebretsen, L., Bendiksen, F., Blair, S. N., Borjesson, M., Budgett, R., Derman, W., Erdener, U., Ioannidis, J. P. et al. (2013). Prevention and Management of Noncommunicable Disease: The IOC Consensus Statement, Lausanne 2013. *Clin. J. Sport Med.* **23,** 419-429.

Matkovich, S. J., Wang, W., Tu, Y., Eschenbacher, W. H., Dorn, L. E., Condorelli, G., Diwan, A., Nerbonne, J. M. and Dorn, G. W.,2nd. (2010). MicroRNA-133a Protects Against Myocardial Fibrosis and Modulates Electrical Repolarization without Affecting Hypertrophy in Pressure-Overloaded Adult Hearts. *Circ. Res.* **106,** 166-175.

Matsuda, A., Hirota, T., Akahoshi, M., Shimizu, M., Tamari, M., Miyatake, A., Takahashi, A., Nakashima, K., Takahashi, N., Obara, K. et al. (2005). Coding SNP in Tenascin-C Fn-III-D Domain Associates with Adult Asthma. *Hum. Mol. Genet.* **14,** 2779-2786.

Matsuo, N., Tanaka, S., Gordon, M. K., Koch, M., Yoshioka, H. and Ramirez, F. (2006). CREB-AP1 Protein Complexes Regulate Transcription of the Collagen XXIV Gene (Col24a1) in Osteoblasts. *J. Biol. Chem.* **281,** 5445-5452.

Matsuo, N., Tanaka, S., Yoshioka, H., Koch, M., Gordon, M. K. and Ramirez, F. (2008). Collagen XXIV (Col24a1) Gene Expression is a Specific Marker of Osteoblast Differentiation and Bone Formation. *Connect. Tissue Res.* **49,** 68-75.

Matzke, M. A. and Birchler, J. A. (2005). RNAi-Mediated Pathways in the Nucleus. *Nat. Rev. Genet.* **6,** 24-35.

Mazumder, B., Seshadri, V. and Fox, P. L. (2003). Translational Control by the 3'-UTR: The Ends Specify the Means. *Trends Biochem. Sci.* **28,** 91-98.

McAlinden, A., Liang, L., Mukudai, Y., Imamura, T. and Sandell, L. J. (2007). Nuclear Protein TIA-1 Regulates COL2A1 Alternative Splicing and Interacts with Precursor mRNA and Genomic DNA. *J. Biol. Chem.* **282,** 24444-24454.

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

Meola, N., Gennarino, V. A. and Banfi, S. (2009). microRNAs and Genetic Diseases. *Pathogenetics* **2,** 7.

Merwin, J. R., Anderson, J. M., Kocher, O., Van Itallie, C. M. and Madri, J. A. (1990). Transforming Growth Factor Beta1 Modulates Extracellular Matrix Organization and Cell-Cell Junctional Complex Formation during in Vitro Angiogenesis. *J. Cell. Physiol.* **142,** 117-128.

Meyer, K., Petersen, A., Niepmann, M. and Beck, E. (1995). Interaction of Eukaryotic Initiation Factor eIF-4B with a Picornavirus Internal Translation Initiation Site. *J. Virol.* **69,** 2819-2824.

Mienaltowski, M. J. and Birk, D. E. (2014). Structure, Physiology, and Biochemistry of Collagens. *Adv. Exp. Med. Biol.* **802,** 5-29.

Mikic, B., Schalet, B. J., Clark, R. T., Gaschen, V. and Hunziker, E. B. (2001). GDF-5 Deficiency in Mice Alters the Ultrastructure, Mechanical Properties and Composition of the Achilles Tendon. *J. Orthop. Res.* **19,** 365-371.

Millar, N. L., Wei, A. Q., Molloy, T. J., Bonar, F. and Murrell, G. A. (2008). Heat Shock Protein and Apoptosis in Supraspinatus Tendinopathy. *Clin. Orthop. Relat. Res.* **466,** 1569-1576.

Mio, F., Chiba, K., Hirose, Y., Kawaguchi, Y., Mikami, Y., Oya, T., Mori, M., Kamata, M., Matsumoto, M., Ozaki, K. et al. (2007). A Functional Polymorphism in COL11A1, which Encodes the Alpha 1 Chain of Type XI Collagen, is Associated with Susceptibility to Lumbar Disc Herniation. *Am. J. Hum. Genet.* **81,** 1271-1277.

Mitchell, A. L., Judis, L. M., Schwarze, U., Vaynshtok, P. M., Drumm, M. L. and Byers, P. H. (2012). Characterization of Tissue-Specific and Developmentally Regulated Alternative Splicing of Exon 64 in the COL5A1 Gene. *Connect. Tissue Res.* **53,** 267-276.

Miyamoto, Y., Mabuchi, A., Shi, D., Kubo, T., Takatori, Y., Saito, S., Fujioka, M., Sudo, A., Uchida, A., Yamamoto, S. et al. (2007). A Functional Polymorphism in the 5' UTR of GDF5 is Associated with Susceptibility to Osteoarthritis. *Nat. Genet.* **39,** 529-533.

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

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

Mokone, G. G., Schwellnus, M. P., Noakes, T. D. and Collins, M. (2006). The COL5A1 Gene and Achilles Tendon Pathology. *Scand. J. Med. Sci. Sports* **16,** 19-26.

Myers, J. C., Dickson, L. A., de Wet, W. J., Bernard, M. P., Chu, M. L., Di Liberto, M., Pepe, G., Sangiorgi, F. O. and Ramirez, F. (1983). Analysis of the 3' End of the Human Pro-Alpha 2(I) Collagen Gene. Utilization of Multiple Polyadenylation Sites in Cultured Fibroblasts. *J. Biol. Chem.* **258,** 10128-10135.

Nakatani, T., Marui, T., Hitora, T., Doita, M., Nishida, K. and Kurosaka, M. (2002). Mechanical Stretching Force Promotes Collagen Synthesis by Cultured Cells from Human Ligamentum Flavum Via Transforming Growth Factor-Beta1. *J. Orthop. Res.* **20,** 1380-1386.

Nardo, T., Micalizzi, G., Vicinanza, R., De Iuliis, F., Taglieri, L. and Scarpa, S. (2014). Adhesion to Type V Collagen Enhances Staurosporine-Induced Apoptosis of Adrenocortical Cancer Cells. *Tumour Biol.* **35(10)**:9949-55

Nebel, M. E. and Scheid, A. (2011). Analysis of the Free Energy in a Stochastic RNA Secondary Structure Model. *IEEE/ACM Trans. Comput. Biol. Bioinform* **8,** 1468-1482.

Nell, E. M., van der Merwe, L., Cook, J., Handley, C. J., Collins, M. and September, A. V. (2012). The Apoptosis Pathway and the Genetic Predisposition to Achilles Tendinopathy. *J. Orthop. Res.* **30,** 1719-1724.

Neves, V. J., Fernandes, T., Roque, F. R., Soci, U. P., Melo, S. F. and de Oliveira, E. M. (2014). Exercise Training in Hypertension: Role of microRNAs. *World J. Cardiol.* **6,** 713-727.

Nielsen, R. H., Couppe, C., Jensen, J. K., Olsen, M. R., Heinemeier, K. M., Malfait, F., Symoens, S., Paepe, A. D., Schjerling, P., Magnusson, S. P. et al. (2014). Low Tendon

Stiffness and Abnormal Ultrastructure Distinguish Classic Ehlers-Danlos Syndrome from Benign Joint Hypermobility Syndrome in Patients. *FASEB J.* **28(11)**:4668-76.

Nunez, G., Benedict, M. A., Hu, Y. and Inohara, N. (1998). Caspases: The Proteases of the Apoptotic Pathway. *Oncogene* **17,** 3237-3245.

O'Brien, M. (1997). Structure and Metabolism of Tendons. *Scand. J. Med. Sci. Sports* **7,** 55-61.

O'Connell, K., Posthumus, M., Schwellnus, M. P. and Collins, M. (2013). Collagen Genes and Exercise-Associated Muscle Cramping. *Clin. J. Sport Med.* **23,** 64-69.

Ogawa, T., Iizuka, M., Sekiya, Y., Yoshizato, K., Ikeda, K. and Kawada, N. (2010). Suppression of Type I Collagen Production by microRNA-29b in Cultured Human Stellate Cells. *Biochem. Biophys. Res. Commun.* **391,** 316-321.

Oleynikov, Y. and Singer, R. H. (2003). Real-Time Visualization of ZBP1 Association with Beta-Actin mRNA during Transcription and Localization. *Curr. Biol.* **13,** 199-207.

Orsmark-Pietras, C., Melen, E., Vendelin, J., Bruce, S., Laitinen, A., Laitinen, L. A., Lauener, R., Riedler, J., von Mutius, E., Doekes, G. et al. (2008). Biological and Genetic Interaction between Tenascin C and Neuropeptide S Receptor 1 in Allergic Diseases. *Hum. Mol. Genet.* **17,** 1673-1682.

Paavola, M., Kannus, P., Jarvinen, T. A., Khan, K., Jozsa, L. and Jarvinen, M. (2002). Achilles Tendinopathy. *J. Bone Joint Surg. Am.* **84-A,** 2062-2076.

Parker, M. I., Smith, A. A., Mundell, K., Collins, M., Boast, S. and Ramirez, F. (1992). The Abolition of Collagen Gene Expression in SV40-Transformed Fibroblasts is Associated with Trans-Acting Factor Switching. *Nucleic Acids Res.* **20,** 5825-5830.

Pasternak, B., Schepull, T., Eliasson, P. and Aspenberg, P. (2010). Elevation of Systemic Matrix Metalloproteinases 2 and 7 and Tissue Inhibitor of Metalloproteinase 2 in Patients with a History of Achilles Tendon Rupture: Pilot Study. *Br. J. Sports Med.* **44,** 669-672.

Peter, M. E. (2010). Targeting of mRNAs by Multiple miRNAs: The Next Step. *Oncogene* **29,** 2161-2164.

Piccinini, A. M. and Midwood, K. S. (2014). Illustrating the Interplay between the Extracellular Matrix and microRNAs. *Int. J. Exp. Pathol.* **95,** 158-180.

Posthumus, M., Collins, M., Cook, J., Handley, C. J., Ribbans, W. J., Smith, R. K., Schwellnus, M. P. and Raleigh, S. M. (2010). Components of the Transforming Growth Factor-Beta Family and the Pathogenesis of Human Achilles Tendon Pathology--a Genetic Association Study. *Rheumatology (Oxford)* **49,** 2090-2097.

Posthumus, M., Schwellnus, M. P. and Collins, M. (2010). The Col5a1 Gene: A Novel Marker of Endurance Running Performance. *Med. Sci. Sports Exerc.* **43(4)**:584-9.

Posthumus, M., September, A. V., O'Cuinneagain, D., van der Merwe, W., Schwellnus, M. P. and Collins, M. (2009). The COL5A1 Gene is Associated with Increased Risk of Anterior Cruciate Ligament Ruptures in Female Participants. *Am. J. Sports Med.* **37,** 2234-2240.

Posthumus, M., September, A. V., O'Cuinneagain, D., van der Merwe, W., Schwellnus, M. P. and Collins, M. (2010). The Association between the COL12A1 Gene and Anterior Cruciate Ligament Ruptures. *Br. J. Sports Med.* **44,** 1160-1165.

Raleigh, S. M., Van der Merwe, L., Ribbans, W. J., Smith, R. K., Schwellnus, M. P. and Collins, M. (2008). Variants within the MMP3 Gene are Associated with Achilles Tendinopathy: Possible Interaction with the COL5A1 Gene. *Br. J. Sports Med.* **43(7)**:514-20

Rasheed, S., Nelson-Rees, W. A., Toth, E. M., Arnstein, P. and Gardner, M. B. (1974). Characterization of a Newly Derived Human Sarcoma Cell Line (HT-1080). *Cancer* **33,** 1027-1033.

Rees, J. D., Wilson, A. M. and Wolman, R. L. (2006). Current Concepts in the Management of Tendon Disorders. *Rheumatology (Oxford)* **45,** 508-521.

Ribbans, W. J. and Collins, M. (2013). Pathology of the Tendo Achillis: Do our Genes Contribute? *Bone Joint J.* **95-B,** 305-313.

Riley, G. P., Harrall, R. L., Cawston, T. E., Hazleman, B. L. and Mackie, E. J. (1996). Tenascin-C and Human Tendon Degeneration. *Am. J. Pathol.* **149,** 933-943.

Romanelli, M. G., Diani, E. and Lievens, P. M. (2013). New Insights into Functional Roles of the Polypyrimidine Tract-Binding Protein. *Int. J. Mol. Sci.* **14,** 22906-22932.

Ross, A. F., Oleynikov, Y., Kislauskis, E. H., Taneja, K. L. and Singer, R. H. (1997). Characterization of a Beta-Actin mRNA Zipcode-Binding Protein. *Mol. Cell. Biol.* **17,** 2158-2165.

Ross, J. (1995). mRNA Stability in Mammalian Cells. *Microbiol. Rev.* **59,** 423-450.

Rossert, J., Terraz, C. and Dupont, S. (2000). Regulation of Type I Collagen Genes Expression. *Nephrol. Dial. Transplant.* **15 Suppl 6,** 66-68.

Rouault, K., Scotet, V., Autret, S., Gaucher, F., Dubrana, F., Tanguy, D., El Rassi, C. Y., Fenoll, B. and Ferec, C. (2010). Evidence of Association between GDF5 Polymorphisms and Congenital Dislocation of the Hip in a Caucasian Population. *Osteoarthritis Cartilage* **18,** 1144-1149.

Roulet, M., Ruggiero, F., Karsenty, G. and LeGuellec, D. (2007). A Comprehensive Study of the Spatial and Temporal Expression of the col5a1 Gene in Mouse Embryos: A Clue for Understanding Collagen V Function in Developing Connective Tissues. *Cell Tissue Res.* **327,** 323-332.

Rutnam, Z. J., Wight, T. N. and Yang, B. B. (2013). miRNAs Regulate Expression and Function of Extracellular Matrix Molecules. *Matrix Biol.* **32,** 74-85.

Sabatier, L., Djokic, J., Hubmacher, D., Dzafik, D., Nelea, V. and Reinhardt, D. P. (2014). Heparin/Heparan Sulfate Controls Fibrillin-1, -2 and -3 Self-Interactions in Microfibril Assembly. *FEBS Lett.* **588,** 2890-2897.

Sakai, L. Y., Keene, D. R. and Engvall, E. (1986). Fibrillin, a New 350-kD Glycoprotein, is a Component of Extracellular Microfibrils. *J. Cell Biol.* **103,** 2499-2509.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). *
Molecular Cloning: A Laboratory Manual.* New York, USA: Cold Spring Harbor Laboratory Press.

Satomi, E., Teodoro, W. R., Parra, E. R., Fernandes, T. D., Velosa, A. P., Capelozzi, V. L. and Yoshinari, N. H. (2008). Changes in Histoanatomical Distribution of Types I, III and V Collagen Promote Adaptative Remodeling in Posterior Tibial Tendon Rupture. *Clinics (Sao. Paulo)* **63,** 9-14.

Saunders, C. J., van der Merwe, L., Posthumus, M., Cook, J., Handley, C. J., Collins, M. and September, A. V. (2013). Investigation of Variants within the COL27A1 and TNC Genes and Achilles Tendinopathy in Two Populations. *J. Orthop. Res.* **31,** 632-637.

Schepsis, A. A., Jones, H. and Haas, A. L. (2002). Achilles Tendon Disorders in Athletes. *Am. J. Sports Med.* **30,** 287-305.

Schwarz, D. S. and Zamore, P. D. (2002). Why do miRNAs Live in the miRNP? *Genes Dev.* **16,** 1025-1031.

Schwellnus, M. P. (2009). Cause of Exercise Associated Muscle Cramps (EAMC)--Altered Neuromuscular Control, Dehydration Or Electrolyte Depletion? *Br. J. Sports Med.* **43,** 401-408.

Schwellnus, M. P., Allie, S., Derman, W. and Collins, M. (2011). Increased Running Speed and Pre-Race Muscle Damage as Risk Factors for Exercise-Associated Muscle Cramps in a 56 km Ultra-Marathon: A Prospective Cohort Study. *Br. J. Sports Med.* **45,** 1132-1136.

Schwellnus, M. P., Derman, E. W. and Noakes, T. D. (1997). Aetiology of Skeletal Muscle 'Cramps' during Exercise: A Novel Hypothesis. *J. Sports Sci.* **15,** 277-285.

Segev, F., Heon, E., Cole, W. G., Wenstrup, R. J., Young, F., Slomovic, A. R., Rootman, D. S., Whitaker-Menezes, D., Chervoneva, I. and Birk, D. E. (2006). Structural Abnormalities of the Cornea and Lid Resulting from Collagen V Mutations. *Invest. Ophthalmol. Vis. Sci.* **47,** 565-573.

Sengupta, S., den Boon, J. A., Chen, I. H., Newton, M. A., Stanhope, S. A., Cheng, Y. J., Chen, C. J., Hildesheim, A., Sugden, B. and Ahlquist, P. (2008). MicroRNA 29c is Down-Regulated in Nasopharyngeal Carcinomas, Up-Regulating mRNAs Encoding Extracellular Matrix Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 5874-5878.

September, A. V., Cook, J., Handley, C. J., van der Merwe, L., Schwellnus, M. P. and Collins, M. (2009). Variants within the COL5A1 Gene are Associated with Achilles Tendinopathy in Two Populations. *Br. J. Sports Med.* **43(5)**:357-65.

September, A. V., Nell, E. M., O'Connell, K., Cook, J., Handley, C. J., van der Merwe, L., Schwellnus, M. and Collins, M. (2011). A Pathway-Based Approach Investigating the Genes Encoding Interleukin-1beta, 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.

Shahbazian, D., Parsyan, A., Petroulakis, E., Topisirovic, I., Martineau, Y., Gibbs, B. F., Svitkin, Y. and Sonenberg, N. (2010). Control of Cell Survival and Proliferation by Mammalian Eukaryotic Initiation Factor 4B. *Mol. Cell. Biol.* **30,** 1478-1485.

Sham, P. C. and Curtis, D. (1995). Monte Carlo Tests for Associations between Disease and Alleles at Highly Polymorphic Loci. *Ann. Hum. Genet.* **59,** 97-105.

Shi, W., Gerster, K., Alajez, N. M., Tsang, J., Waldron, L., Pintilie, M., Hui, A. B., Sykes, J., P'ng, C., Miller, N. et al. (2011). MicroRNA-301 Mediates Proliferation and Invasion in Human Breast Cancer. *Cancer Res.* **71,** 2926-2937.

Silver, F. H., Freeman, J. W. and Seehra, G. P. (2003). Collagen Self-Assembly and the Development of Tendon Mechanical Properties. *J. Biomech.* **36,** 1529-1553.

Somerville, R. P., Oblander, S. A. and Apte, S. S. (2003). Matrix Metalloproteinases: Old Dogs with New Tricks. *Genome Biol.* **4,** 216.

Southam, L., Rodriguez-Lopez, J., Wilkins, J. M., Pombo-Suarez, M., Snelling, S., Gomez-Reino, J. J., Chapman, K., Gonzalez, A. and Loughlin, J. (2007). An SNP in the 5'-UTR of GDF5 is Associated with Osteoarthritis Susceptibility in Europeans and with in Vivo Differences in Allelic Expression in Articular Cartilage. *Hum. Mol. Genet.* **16,** 2226-2232.

Souza, P., Rizzardi, F., Noleto, G., Atanazio, M., Bianchi, O., Parra, E. R., Teodoro, W. R., Carrasco, S., Velosa, A. P., Fernezlian, S. et al. (2010). Refractory Remodeling of the Microenvironment by Abnormal Type V Collagen, Apoptosis, and Immune Response in Non-Small Cell Lung Cancer. *Hum. Pathol.* **41,** 239-248.

Stutz, F., Bachi, A., Doerks, T., Braun, I. C., Seraphin, B., Wilm, M., Bork, P. and Izaurralde, E. (2000). REF, an Evolutionary Conserved Family of hnRNP-Like Proteins, Interacts with TAP/Mex67p and Participates in mRNA Nuclear Export. *RNA* **6,** 638-650.

Sun, M., Chen, S., Adams, S. M., Florer, J. B., Liu, H., Kao, W. W., Wenstrup, R. J. and Birk, D. E. (2011). Collagen V is a Dominant Regulator of Collagen Fibrillogenesis: Dysfunctional Regulation of Structure and Function in a Corneal-Stroma-Specific Col5a1-Null Mouse Model. *J. Cell. Sci.* **124,** 4096-4105.

Syddall, C. M., Reynard, L. N., Young, D. A. and Loughlin, J. (2013). The Identification of Trans-Acting Factors that Regulate the Expression of GDF5 Via the Osteoarthritis Susceptibility SNP rs143383. *PLoS Genet.* **9,** e1003557.

Tang, G. (2005). siRNA and miRNA: An Insight into RISCs. *Trends Biochem. Sci.* **30,** 106-114.

Theobald, P., Benjamin, M., Nokes, L. and Pugh, N. (2005). Review of the Vascularisation of the Human Achilles Tendon. *Injury* **36,** 1267-1272.

Thiele, B. J., Doller, A., Kahne, T., Pregla, R., Hetzer, R. and Regitz-Zagrosek, V. (2004). RNA-Binding Proteins Heterogeneous Nuclear Ribonucleoprotein A1, E1, and K are Involved in Post-Transcriptional Control of Collagen I and III Synthesis. *Circ. Res.* **95,** 1058-1066.

Tully, L. J., Murphy, A. M., Smith, R. K., Hulin-Curtis, S. L., Verheyen, K. L. and Price, J. S. (2014). Polymorphisms in TNC and COL5A1 Genes are Associated with Risk of Superficial Digital Flexor Tendinopathy in National Hunt Thoroughbred Racehorses. *Equine Vet. J.* **46,** 289-293.

Uhorchak, J. M., Scoville, C. R., Williams, G. N., Arciero, R. A., St Pierre, P. and Taylor, D. C. (2003). Risk Factors Associated with Noncontact Injury of the Anterior Cruciate Ligament: A Prospective Four-Year Evaluation of 859 West Point Cadets. *Am. J. Sports Med.* **31,** 831-842.

Ushiki, T. (2002). Collagen Fibers, Reticular Fibers and Elastic Fibers. A Comprehensive Understanding from a Morphological Viewpoint. *Arch. Histol. Cytol.* **65,** 109-126.

Vaes, R. B., Rivadeneira, F., Kerkhof, J. M., Hofman, A., Pols, H. A., Uitterlinden, A. G. and van Meurs, J. B. (2009). Genetic Variation in the GDF5 Region is Associated with Osteoarthritis, Height, Hip Axis Length and Fracture Risk: The Rotterdam Study. *Ann. Rheum. Dis.* **68,** 1754-1760.

Van Rooij, E., Sutherland, L. B., Thatcher, J. E., DiMaio, J. M., Naseem, R. H., Marshall, W. S., Hill, J. A. and Olson, E. N. (2008). Dysregulation of microRNAs After Myocardial Infarction Reveals a Role of miR-29 in Cardiac Fibrosis. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 13027-13032.

Verhagen, E. (2012). Prevention of Running-Related Injuries in Novice Runners: Are we Running on Empty? *Br. J. Sports Med.* **46,** 836-837.

Viiri, J., Amadio, M., Marchesi, N., Hyttinen, J. M., Kivinen, N., Sironen, R., Rilla, K., Akhtar, S., Provenzani, A., D'Agostino, V. G. et al. (2013). Autophagy Activation Clears ELAVL1/HuR-Mediated Accumulation of SQSTM1/p62 during Proteasomal Inhibition in Human Retinal Pigment Epithelial Cells. *PLoS One* **8,** e69563.

Visse, R. and Nagase, H. (2003). Matrix Metalloproteinases and Tissue Inhibitors of Metalloproteinases: Structure, Function, and Biochemistry. *Circ. Res.* **92,** 827-839.

Vitart, V., Bencic, G., Hayward, C., Skunca Herman, J., Huffman, J., Campbell, S., Bucan, K., Navarro, P., Gunjaca, G., Marin, J. et al. (2010). New Loci Associated with Central Cornea Thickness Include COL5A1, AKAP13 and AVGR8. *Hum. Mol. Genet.* **19,** 4304-4311.

Wall, M. E. and Banes, A. J. (2005). Early Responses to Mechanical Load in Tendon: Role for Calcium Signaling, Gap Junctions and Intercellular Communication. *J. Musculoskelet. Neuronal Interact.* **5,** 70-84.

Wang, G., van der Walt, J. M., Mayhew, G., Li, Y. J., Zuchner, S., Scott, W. K., Martin, E. R. and Vance, J. M. (2008). Variation in the miRNA-433 Binding Site of FGF20 Confers Risk for Parkinson Disease by Overexpression of Alpha-Synuclein. *Am. J. Hum. Genet.* **82,** 283-289.

Wang, J. H. (2006). Mechanobiology of Tendon. *J. Biomech.* **39,** 1563-1582.

Wang, Z., Day, N., Trifillis, P. and Kiledjian, M. (1999). An mRNA Stability Complex Functions with Poly(A)-Binding Protein to Stabilize mRNA in Vitro. *Mol. Cell. Biol.* **19,** 4552-4560.

Wearing, S. C., Hooper, S. L., Grigg, N. L., Nolan, G. and Smeathers, J. E. (2013). Overweight and Obesity Alters the Cumulative Transverse Strain in the Achilles Tendon Immediately Following Exercise. *J. Bodyw Mov. Ther.* **17,** 316-321.

Weinfeld, S. B. (2014). Achilles Tendon Disorders. *Med. Clin. North Am.* **98,** 331-338.

Wenstrup, R. and De Paepe, A. (1993). Ehlers-Danlos Syndrome, Classic Type. In *GeneReviews* (ed. R. A. Pagon, T. D. Bird, C. R. Dolan and K. Stephens). Seattle (WA): University of Washington, Seattle. All rights reserved.

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

Wenstrup, R. J., Florer, J. B., Cole, W. G., Willing, M. C. and Birk, D. E. (2004). Reduced Type I Collagen Utilization: A Pathogenic Mechanism in COL5A1 Haplo-Insufficient Ehlers-Danlos Syndrome. *J. Cell. Biochem.* **92,** 113-124.

Wenstrup, R. J., Florer, J. B., Davidson, J. M., Phillips, C. L., Pfeiffer, B. J., Menezes, D. W., Chervoneva, I. and Birk, D. E. (2006). Murine Model of the Ehlers-Danlos Syndrome. col5a1 Haploinsufficiency Disrupts Collagen Fibril Assembly at Multiple Stages. *J. Biol. Chem.* **281,** 12888-12895.

Wenstrup, R. J., Florer, J. B., Willing, M. C., Giunta, C., Steinmann, B., Young, F., Susic, M. and Cole, W. G. (2000). COL5A1 Haploinsufficiency is a Common Molecular Mechanism Underlying the Classical Form of EDS. *Am. J. Hum. Genet.* **66,** 1766-1776.

Wenstrup, R. J., Smith, S. M., Florer, J. B., Zhang, G., Beason, D. P., Seegmiller, R. E., Soslowsky, L. J. and Birk, D. E. (2011). 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.

Whaley, M. (2006). *ACSM's Guidelines for Exercise Testing and Prescription*. 7th Ed. Baltimore, MA, USA: Lippincott, Williams and Wilkins. 85-86.

Williams, F. M., Popham, M., Hart, D. J., de Schepper, E., Bierma-Zeinstra, S., Hofman, A., Uitterlinden, A. G., Arden, N. K., Cooper, C., Spector, T. D. et al. (2011). GDF5 Single-Nucleotide Polymorphism rs143383 is Associated with Lumbar Disc Degeneration in Northern European Women. *Arthritis Rheum.* **63,** 708-712.

Wolinsky, I. and Driskell, J. A. (2005). *Sports Nutrition: Vitamins and Trace Elements,* 2nd Ed. Taylor & Francis.

World Health Organisation. (2014). Obesity and overweight Fact Sheet, N°311. **2014**.

World Medical Association. (2013). World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. *JAMA* **310,** 2191-2194.

Wren, T. A., Lindsey, D. P., Beaupre, G. S. and Carter, D. R. (2003). Effects of Creep and Cyclic Loading on the Mechanical Properties and Failure of Human Achilles Tendons. *Ann. Biomed. Eng.* **31,** 710-717.

Wren, T. A., Yerby, S. A., Beaupre, G. S. and Carter, D. R. (2001). Influence of Bone Mineral Density, Age, and Strain Rate on the Failure Mode of Human Achilles Tendons. *Clin. Biomech. (Bristol, Avon)* **16,** 529-534.

Yang, B., Guo, H., Zhang, Y., Chen, L., Ying, D. and Dong, S. (2011). MicroRNA-145 Regulates Chondrogenic Differentiation of Mesenchymal Stem Cells by Targeting Sox9. *PLoS One* **6,** e21679.

Yassi, A. (2000). Work-Related Musculoskeletal Disorders. *Curr. Opin. Rheumatol.* **12,** 124-130.

Yeung, G., Choi, L. M., Chao, L. C., Park, N. J., Liu, D., Jamil, A. and Martinson, H. G. (1998). Poly(A)-Driven and Poly(A)-Assisted Termination: Two Different Modes of Poly(A)-Dependent Transcription Termination. *Mol. Cell. Biol.* **18,** 276-289.

Zarudnaya, M. I., Kolomiets, I. M., Potyahaylo, A. L. and Hovorun, D. M. (2003). Downstream Elements of Mammalian Pre-mRNA Polyadenylation Signals: Primary, Secondary and Higher-Order Structures. *Nucleic Acids Res.* **31,** 1375-1386.

Zhang, H., Apfelroth, S. D., Hu, W., Davis, E. C., Sanguineti, C., Bonadio, J., Mecham, R. P. and Ramirez, F. (1994). Structure and Expression of Fibrillin-2, a Novel Microfibrillar Component Preferentially Located in Elastic Matrices. *J. Cell Biol.* **124,** 855-863.

Zhang, W. and Chen, S. J. (2002). RNA Hairpin-Folding Kinetics. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 1931-1936.

Zuker, M. (2003). Mfold Web Server for Nucleic Acid Folding and Hybridization Prediction. *Nucleic Acids Res.* **31,** 3406-3415.

Zuker, M., Mathews, D. H. and Turner, D. H. (1999). Algorithms and Thermodynamics for RNA Secondary Structure Prediction: A Practical Guide.**70,**11-43

Zweers, M. C., Kucharekova, M. and Schalkwijk, J. (2005). Tenascin-X: A Candidate Gene for Benign Joint Hypermobility Syndrome and Hypermobility Type Ehlers-Danlos Syndrome? *Ann. Rheum. Dis.* **64,** 504-505.

APPENDICES

A.1 Appendix A

Ethical approval and recruitment forms:

1. Approval letters from the human research ethics committee

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 6338 • **Facsimile** [021] 406 6411
e-mail: shuretta.thomas@uct.ac.za

17 January 2013

HREC REF: 649/2012

Prof M Collins
Human Biology
Sport Science Institute

Dear Prof Collins

PROJECT TITLE: DETERMINING THE FUNCTIONAL ROLE OF VARIANTS WITHIN THE EXTRACELLULAR MATRIX GENES ON MUSCULOSKELETAL SOFT TISSUE INJURIES, USING PRIMARY HUMAN FIBROBLAST CELL LINES

Thank you for responding to the issues raised by the Faculty of Health Sciences Human Research Ethics Committee in your letter datec on 15th January 2013.

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

Approval is granted for one year till the 30th January 2014

Please submit a progress form, using the standardised Annual Report Form if the study continues beyond the approval period. Please submit a Standard Closure form if the study is completed within the approval period.
(Forms can be found on our website: www.health.uct.ac.za/research/humanethics/forms)

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

PROFESSOR M 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.

s.thomas

HUMAN RESEARCH
ETHICS COMMITTEE
12 MAY 2014
HEALTH SCIENCES FACULTY
UNIVERSITY OF CAPE TOWN

UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

FACULTY OF HEALTH SCIENCES
Human Research Ethics Committee

FHS016: Annual Progress Report / Renewal

HREC office use only (FWA00001637; IRB00001938) This serves as notification of annual approval, including any documentation described below.			
☑ Approved	Annual progress report	Approved until/next renewal date	30.01.2015
☐ Not approved	See attached comments		
Signature Chairperson of the HREC		Date Signed	12/05/2014

Comments to PI from the HREC

Principal Investigator to complete the following:

1. Protocol information

Date form submitted	25.03.14		
HREC REF Number	649/2012	Current Ethics Approval was granted until	30.01.14
Protocol title	Determining the functional role of variants within the extracellular matrix genes on musculoskeletal soft tissue injuries, using primary human fibroblast cell lines.		
Protocol number (if applicable)	N/A		
Are there any sub-studies linked to this study?	No		
If yes, could you please provide the HREC Ref's for all sub-studies? **Note:** A separate FHS016 must be submitted for each sub-study.			
Principal Investigator	Prof Malcolm Collins		
Department / Office Internal Mail Address	Room 5.14,Anatomy Building, Faculty of Health Sciences, University of Cape Town, Anzio Road, Observatory 7700,Cape Town, South Africa		

1.1 Does this protocol receive US Federal funding?	☐ Yes	✓ No
1.2 If the study receives US Federal Funding, does the annual report require full committee approval?	☐ Yes	✓ No

(Note: Please complete the Closure form (FHS010) if the study is completed within the approval period)

UNIVERSITY OF CAPE TOWN

HUMAN RESEARCH
ETHICS COMMITTEE
27 AUG 2014
UNIVERSITY OF CAPE TOWN

FACULTY OF HEALTH SCIENCES
Human Research Ethics Committee

FHS016: Annual Progress Report / Renewal

HREC office use only (FWA00001637; IRB00001938) This serves as notification of annual approval, including any documentation described below.			
☑ Approved	Annual progress report	Approved until/next renewal date	30.10.2015
☐ Not approved	See attached comments		
Signature Chairperson of the HREC		Date Signed	27/8/2014

Comments to PI from the HREC
New candidate (?) approved for investigator

Principal Investigator to complete the following:

1. Protocol information

Date form submitted	25 August 2014		
HREC REF Number	HREC 086/2005	Current Ethics Approval was granted until	
Protocol title	The Genetic Basis of Tendinopathy		
Protocol number (if applicable)	n/a		

Are there any sub-studies linked to this study?	X Yes ☐ No

If yes, could you please provide the HREC Ref's for all sub-studies? Note: A separate FHS016 must be submitted for each sub-study.

HREC 224/2013; 139/2009; 159/2011; 649/2012;

HREC 427/2014

Principal Investigator	Prof Malcolm Collins & Dr Alison September
Department / Office Internal Mail Address	ESSM, SSISA Building, 1 Boundary Road, Newlands, 7700

1.1 Does this protocol receive US Federal funding?	☐ Yes	X No
1.2 If the study receives US Federal Funding, does the annual report require full committee approval?	☐ Yes	X No

(Note: Please complete the Closure form (FHS010) if the study is completed within the approval period)

UNIVERSITY OF CAPE TOWN

HUMAN RESEARCH ... 27 AUG 2014 HEALTH SCIENCES FACULTY

FACULTY OF HEALTH SCIENCES
Human Research Ethics Committee

FHS016: Annual Progress Report / Renewal

HREC office use only (FWA00001637; IRB00001938) This serves as notification of annual approval, including any documentation described below.			
☑ Approved	Annual progress report	Approved until/next renewal date	30.10.2015
☐ Not approved	See attached comments		
Signature Chairperson of the HREC		Date Signed	27/8/2014

Comments to PI from the HREC
New Candidate Genes approved for investigation

Principal Investigator to complete the following:

1. Protocol information

Date form submitted	25 August 2014		
HREC REF Number	HREC ~~086/2004~~ 384/2004	Current Ethics Approval was granted until	
Protocol title	The Genetic Basis of Tendon Pathology in an Australian Population		
Protocol number (if applicable)	n/a		
Are there any sub-studies linked to this study?	X Yes ☐ No		

If yes, could you please provide the HREC Ref's for all sub-studies? Note: A separate FHS016 must be submitted for each sub-study HREC 427/2014 HREC 224/2013; 649/2012;

Principal Investigator	Prof Malcolm Collins & Dr Alison September
Department / Office Internal Mail Address	ESSM, SSISA Building, 1 Boundary Road, Newlands, 7700

1.1 Does this protocol receive US Federal funding?	☐ Yes	X No
1.2 If the study receives US Federal Funding, does the annual report require full committee approval?	☐ Yes	X No

11 February 2014 Page 1 of 6 FHS016

(Note: Please complete the Closure form (FHS010) if the study is completed within the approval period)

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: lamees.emjedi@uct.ac.za

08 May 2008

REC REF: 086/2005

A/Prof M Collins
Human Biology
Sports Science Institute

Dear A/Prof Collins

PROJECT TITLE: THE GENETIC BASIS OF TENDINOPATHY

Thank you for your letter to the Research Ethics Committee dated 24 April 2008.

Addendum to undertake further genetic analysis is approved.
We note that future analysis will be performed on de-identified samples.

Please would you submit an annual progress report which includes a description of the current status of the research and your publication as outlined in this letter.

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

Please quote the REC. REF in all your correspondence.

Yours sincerely

PROFESSOR M BLOCKMAN
CHAIRPERSON, HSF HUMAN ETHICS

lemjedi

2. Recruitment information sheets

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 CHRONIC TENDON PATHOLOGY

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 collaboration with the UCT/MRC Research Unit for Exercise Science and Sports Medicine within the Department of Human Biology and the Division of Human Genetics within the Department of Clinical Laboratory Sciences 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 30 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. 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.

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. Malcolm Collins, PhD
(021) 650 4574
mcollins@sports.uct.ac.za

Prof. Martin Schwellnus, MBChB, MD
(021) 650 4576
mschwell@sports.uct.ac.za

Colleen Saunders, MSc student
(021) 650 4569
csanders@sports.uct.ac.za

The University of Cape Town is committed to policies of equal opportunity and affirmative action which are essential to its mission of promoting critical inquiry and scholarship

Department of Human Biology
University of Cape Town
South Africa

16 July 2004

STUDY ON THE GENETIC BASIS OF TENDON PATHOLOGY

The Musculoskeletal Research Centre at La Trobe University in Melbourne, Australia, in collaboration with the UCT/MRC Research Unit for Exercise Science and Sports Medicine at the University of Cape Town (UCT) in Cape Town, South Africa, are currently studying the genetic basis of chronic tendon pathology in the Australian population.

Studies have suggested that some individuals have a genetic predisposition to tendon injury and that genes (those traits which you inherit form your parents), such as Type V collagen (COL5A1) and Tenascin C (TNC), which encode for important components of tendons are associated with tendon pathology.

The aim of this study is to determine whether the TNC, the COL5A1 or other similar genes are associated with chronic tendon pathology.

In order to participate in this study, you will be required to donate five millilitres of venous blood after giving written consent. The blood sample will be used for the extraction and analysis of genetic material (DNA). The extracted DNA will be sent to UCT in South Africa for analysis. The samples will be shipped to and analysed by UCT anonymously. The DNA will be genotyped (analysed) for variations (polymorphisms) within the *COL5A1*, *TNC* and other candidate genes believed to be associated with tendon pathology. You will be required to complete a number of questionnaires regarding personal particulars, sporting participation and medical history, as well as a stretching and warm up questionnaire. In addition, you will be required to visit a doctor (radiologist) at a later stage for a tendon(s) scan free of charge. All the information collected during the study will be treated with the strictest confidentiality and will only be used for scientific research purposes. Your name and personal particulars will not be released under any circumstances and all the data obtained will be analysed anonymously.

If you would like to participate in the study and/or obtain any additional information, please contact **Dr Jill Cook,** on phone: **9479 5789**, or e-mail: **J.Cook@latrobe.edu.au**.

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

Dear

While our research group has been successful in identifying several genes associated with musculoskeletal soft tissue injuries and exercise-related traits, such as Achilles tendon ruptures or endurance running performance, much more work needs to be done in order to understand the relevance of these genetic associations at the functional level. Two of the associated genes, *COL5A1* and *COL6A1*, are involved in fibrillogenesis, the process by which collagen fibrils, the basic building block of tendons, ligaments and other connective tissues, are formed. It is proposed that naturally occurring inter-individual variations within these genes may alter normal fibrillogenesis and thereby affect the normal biomechanical properties of these tissues.

The aim of this study is to determine if these variants have functional effects on the genes in which they are identified, and the impact that they may have on the gene, since this would provide plausible mechanistic explanations for the associations found. In order to achieve this, tissue samples with known variants will be cultured and analysed. This study is being performed by Kevin O'Connell and Mary-Jessica Laguette who are both PhD students in the UCT/MRC Research Unit for Exercise Science and Sports Medicine, Department of Human Biology at the University of Cape Town.

You may be required to visit the Sports Science Institute of South Africa (SSISA) in Boundary Road, Newlands on two separate occasions. During the first visit, which should take no longer than 30 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 participation, as well as personal and family medical history questionnaires. The DNA will be genotyped (analysed) for DNA sequence variations (polymorphisms) within candidate genes for musculoskeletal soft tissue injuries, such as the *COL5A1* and *COL6A1* genes.

You may be invited by the study investigators to participate in additional components of this study should your connective tissue genotypes match those required for the growth of primary fibroblast cell lines. You may be asked to visit the Sports Science Institute of South Africa (SSISA) in Boundary Road, Newlands on a second occasion in order to donate a single skin punch biopsy (3-4 mm), which will be used to culture primary skin fibroblast cell lines.

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 and/or tissue sample will be destroyed on completion of the study on determining the functional role of variants within the extracellular matrix genes on

The University of Cape Town is committed to policies of equal opportunity and affirmative action which are essential to its mission of promoting critical inquiry and scholarship

musculoskeletal soft tissue injuries, using primary human fibroblast cell lines. You are also free to request that your DNA and/or tissue 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.

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:-

Kevin O'Connell, PhD student
(021) 650 4569
kevin.oconnell@uct.ac.za

Mary-Jessica Laguette, PhD student
(021) 650 4569
nancylaguette@gmail.com

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

3. Informed consent forms

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 CHRONIC TENDON PATHOLOGY

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 and the Division of Human Genetics within the Department of Clinical Laboratory Sciences at the University of Cape Town's study on the genetic basis of exercise induced chronic tendon pathology. I have agreed to donate five millilitres of venous blood 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 not be released under any circumstances and that all data will be analysed anonymously.

If requested, I am also prepared to visit a doctor (radiologist) at a later stage for a tendon scan at no cost to myself (please delete this sentence if not applicable). If requested, I am also prepared to visit the SSISA for measurements to determine musculo-tendinous stiffness.

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 tendon pathology. I also understand that I will be free to request that my DNA sample be destroyed before the completion of the study.

FULL NAME OF SUBJECT: ______________________

SUBJECT'S SIGNATURE: ______________________

DATE: ______________________

INVESTIGATOR: ______________________

INVESTIGATOR'S SIGNATURE: ______________________

The University of Cape Town is committed to policies of equal opportunity and affirmative action which are essential to its mission of promoting critical inquiry and scholarship

13TH JULY 2004

GENETIC BASIS OF TENDON PATHOLOGY

INFORMED CONSENT

I, (the participant), have been fully informed about this study on the genetic basis of tendon pathology to be conducted by the Musculoskeletal Research Centre at La Trobe University in Melbourne, Australia, in conjunction with the UCT/MRC Research Unit for Exercise Science and Sports Medicine at the University of Cape Town in Cape Town, South Africa.

I have agreed to donate five millilitres of venous blood, which will be used for the extraction and analysis of genetic material (DNA), and will be taken by a registered medical physician or nurse. The DNA will only be used for scientific research purposes relating to the genetic basis of tendon pathology only. 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 all data will be analysed anonymously and my DNA sample will be destroyed on completion of the study.

I understand that the DNA extracted form the donated blood sample will be sent to UCT in South Africa for analysis. I understand that the DNA samples will be shipped to and analysed by UCT anonymously. I understand that the DNA will be genotyped (analysed) for variations (polymorphisms) within the COL5A1 and TNC genes, as well as other similar genes relating to the genetic basis of tendon pathology only.

In addition, I am also prepared to visit a doctor (radiologist) at a later stage for a tendon(s) scan at no cost to myself.

I understand that whilst there is no direct benefit to myself, if a genetic predisposition for tendinopathy 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 have read (or, where appropriate, have had read to me) and understood 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 anytime

and, further, to demand that data arising from my participation is not used in the research project provided that this right is exercised within four weeks of the completion of my participation in the project. 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 the Chief Investigator: **Dr Jill Cook**, of the **School of Physiotherapy** on telephone number **9479 5789**.

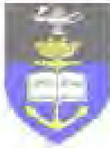
If you have any complaints or queries that the investigator has not been able to answer to your satisfaction, you may contact the Ethics Liaison Officer, Human Ethics Committee, La Trobe University, Victoria, 3086, (ph: 03 9479 1443, e-mail: humanethics@latrobe.edu.au).

Name of Participant: ______________________________

Signature: ______________________________ **Date:** ______________

Name of Researcher: ______________________________

Signature: ______________________________ **Date:** ______________



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

DETERMINING THE FUNCTIONAL ROLE OF VARIANTS WITHIN THE EXTRACELLULAR MATRIX GENES ON MUSCULOSKELETAL SOFT TISSUE INJURIES, USING PRIMARY HUMAN FIBROBLAST CELL LINES.

SKIN PUNCH BIOPSY 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 to identify the functional role of variants within the extracellular matrix (connective tissue) genes on musculoskeletal soft tissue injuries, using primary human fibroblast cell lines. A single skin punch biopsy (3-4 mm), will be used to culture primary skin fibroblast cell lines. A fibroblast is a type of cell that synthesizes the structural framework (extracellular matrix and collagen) for tissues. The punch biopsy will be done by a trained medical practitioner. 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. Names and personal particulars will be not released under any circumstances and all data will be analysed anonymously.

I agree to participate in the study and to donate a single skin punch biopsy as described above. I have been informed that I will be free to withdraw from the study at any time if I so wish. The tissue samples will be destroyed on completion of the study to identify the functional role of variants within the extracellular matrix (connective tissue) genes on musculoskeletal soft tissue injuries, using primary human fibroblast cell lines. I also understand that I will be free to request that my tissue sample be destroyed before the completion of the study.

The potential risks associated with the skin punch biopsy technique from the forearm are: infection, delayed healing, 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 medical practitioner, use of sterile techniques and the use of disposable, single use materials.

The tissue sample will be cultured and analysed using a variety of molecular techniques to determine the functional effects of DNA sequence variations (polymorphisms) within candidate genes for musculoskeletal soft tissue injuries, such as the *COL5A1* and *COL6A1* genes. Any remaining tissue will be discarded appropriately.

There is no direct benefit to myself however, this may allow better prevention and treatment options for musculoskeletal soft tissue injuries in the future. I understand that I will receive the overall results of the study. I

The University of Cape Town is committed to policies of equal opportunity and affirmative action which are essential to its mission of promoting critical inquiry and scholarship
Discovery Health

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 tissue 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 the Principle Investigator: **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**.

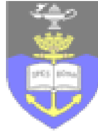
FULL NAME OF SUBJECT: ______________________________

SUBJECT'S SIGNATURE: ______________________________

DATE: ______________________________

INVESTIGATOR : ______________________________

INVESTIGATOR'S SIGNATURE: ______________________________



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

DETERMINING THE FUNCTIONAL ROLE OF VARIANTS WITHIN THE EXTRACELLULAR MATRIX GENES ON MUSCULOSKELETAL SOFT TISSUE INJURIES, USING PRIMARY HUMAN FIBROBLAST CELL LINES.

BLOOD COLLECTION AND DISCARDED TISSUE 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 to identify the functional role of DNA (genetic material) sequence variants within the extracellular matrix (connective tissue) genes on musculoskeletal soft tissue injuries, using primary human fibroblast cell lines. A fibroblast is a type of cell that synthesizes the structural framework (extracellular matrix and collagen) for tissues. Participants are required to donate five millilitres of venous blood or a Buccal mouthwash/swab sample, which will be used for the extraction and analysis of DNA. The blood sample will be taken from a forearm (ante-cubital) vein by a nurse, physician or phlebotomist. The DNA will be genotyped (analysed) for DNA sequence variations (polymorphisms) within candidate genes for musculoskeletal soft tissue injuries, such as the *COL5A1* and *COL6A1* genes, to investigate their effects on connective tissue, such as tendons and ligaments. Participants are required to complete personal particulars, sporting participation, as well as personal and family medical history questionnaires. 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. Names and personal particulars will be not released under any circumstances and all data will be analysed anonymously.

I agree to participate in the study. I agree to donate a blood sample for the extraction and analysis of my DNA. I agree to complete all study related questionnaires as outlined above. 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 and/or tissue sample will be destroyed on completion of the study to identify the functional role of variants within the extracellular matrix (connective tissue) genes on musculoskeletal soft tissue injuries, using primary human fibroblast cell lines. I also understand that I will be free to request that my DNA sample and/or tissue sample be destroyed before the completion of the study.

The potential risks associated with blood collection technique from the forearm (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.

The University of Cape Town is committed to policies of equal opportunity and affirmative action which are essential to its mission of promoting critical inquiry and scholarship

The DNA will be genotyped (analysed) for DNA sequence variations (polymorphisms) within candidate genes for musculoskeletal soft tissue injuries, such as the *COL5A1* and *COL6A1* genes.

If I am undergoing surgery, I agree to allow any of my tissue that is excised during surgery and to be discarded for incineration, to be collected for this study. This collected tissue will be used to culture primary fibroblast cell lines, which will be used in a number of molecular techniques to identify the functional role of variants within candidate genes for musculoskeletal soft tissue injuries, such as the *COL5A1* and *COL6A1* genes. Any remaining tissue will be discarded appropriately.

I give permission to be contacted by the study investigators to be invited to participate in additional components of this study should my connective tissue genotypes match those required for the growth of primary fibroblast cell lines.

I understand that there is no direct benefit to myself however, this may allow better prevention and treatment options for musculoskeletal soft tissue injuries 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 and/or tissue 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 the Principle Investigator: **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**.

FULL NAME OF SUBJECT: ______________________________

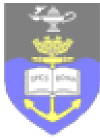
SUBJECT'S SIGNATURE: ______________________________

DATE: ______________________________

INVESTIGATOR : ______________________________

INVESTIGATOR'S SIGNATURE: ______________________________

4. Questionnaire used for the data collection for Achilles Tendinopathy



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 TENDON 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	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 ________		

The University of Cape Town is committed to policies of equal opportunity and affirmative action which are essential to its mission of promoting critical inquiry and scholarship

(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. FLEXIBILITY TRAINING HISTORY	
Do you perform flexibility training (regular stretching exercises)?	Yes ☐ No ☐
<u>If YES,</u> please complete the rest of the flexibility training history section below:- If NO, continue completing the questionnaire from the top of page 5 (Equipment use history).	
On average, how many <u>days a week</u> 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

D. TENDON INJURY - MEDICAL DETAILS

Symptoms

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

Please complete a **separate form**, Part D 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 ____________________________)
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 ____________

E. 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 □			
If yes, please specify which tendon? (You may tick more than one block, please select either L (left) or R (right))			L R	
	Foot and ankle:	Achilles tendon	□ □	
		Tibialis posterior	□ □	
		Plantar fascia	□ □	
	Knee:	Patellar tendon	□ □	
	Elbow and wrist:	Wrist extensor tendons	□ □	
	Shoulder:	Subscapularis	□ □	
		Supraspinatus	□ □	
		Infraspinatus	□ □	
		Teres minor	□ □	
	Other:..			

To your knowledge, have any other members of your family suffered from any tendon pathology?	Yes ☐ No ☐	If Yes, please specify the family member ☐ Mother ☐ Father ☐ Sibling ☐ Son / daughter ☐ Other family member:........................... Condition: Please choose tendon injury from the list above ....
Have you ever suffered from any of the following joint capsule injuries?	☐ Acute shoulder dislocation ☐ Chronic shoulder instability ☐ Chronic ankle instability ☐ Other: ________________________________ ________________________________	

F. 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________________

<table>
<tr><td rowspan="6">What surgical operations have you had? (please list and give dates)</td><td>Operation</td><td>Date</td></tr>
<tr><td></td><td></td></tr>
<tr><td></td><td></td></tr>
<tr><td></td><td></td></tr>
<tr><td></td><td></td></tr>
<tr><td></td><td></td></tr>
<tr><td colspan="3">If female:</td></tr>
<tr><td>At what age did you start menstruating? (years)</td><td colspan="2"></td></tr>
<tr><td>Are you currently using any type of contraception?</td><td colspan="2">☐ Yes ☐ No</td></tr>
<tr><td>If Yes, what type of contraception are you using?</td><td colspan="2">☐ Pill ☐ Injection ☐ IUD</td></tr>
<tr><td>Are you currently?</td><td colspan="2">☐ 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)</td></tr>
<tr><td colspan="3">Family History</td></tr>
<tr><td>Do any other members of your family suffer from elevated blood cholesterol?</td><td>Yes ☐ No ☐</td><td>If Yes, which relative?
☐ Mother
☐ Father
☐ Sibling
☐ Son / daughter
☐ Other relative:</td></tr>
<tr><td>Is there any history of arthritis in your family?</td><td>Yes ☐ No ☐</td><td>If Yes, which relative?
☐ Mother
☐ Father
☐ Sibling
☐ Son / daughter
☐ Other relative:
& What type of aethritis?
Rheumatoid ☐
Osteoarthritis ☐
Other ☐</td></tr>
</table>

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 the Achilles 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)		

G. OCCUPATIONAL DETAILS	
What is your current occupation?	
What was your occupation prior to injuring your tendon?	
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 ☐

5. Questionnaire used for data collection for primary human fibroblast cell lines

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

DETERMINING THE FUNCTIONAL ROLE OF VARIANTS WITHIN THE EXTRACELLULAR MATRIX GENES ON MUSCULOSKELETAL SOFT TISSUE INJURIES, USING PRIMARY HUMAN FIBROBLAST CELL LINES.

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

Section A	Personal Details	Page 2
Section B	Sporting Details	Page 3
Section C	Flexibility Training History	Page 4
Section D	Lifestyle and habits history	Page 4
Section E	General Personal Medical History	Page 5
Section F	Family Medical History	Page 6
Section G	History of Medication Use	Page 7
Section H	Muscle Cramping	Page 8
Section I	Past History of Skeletal Muscle Injury	Page 9-11
Section J	History of Tendon, Ligament or Joint Capsule Injury	Pages 12
Section K	Medical Details of Tendon Injuries	Pages 13-14
Section L	History if Any Other Chronic Current Injury	Pages 15

Subject Number: ________

The University of Cape Town is committed to policies of equal opportunity and affirmative action which are essential to its mission of promoting critical inquiry and scholarship

Version 1
(March 2010)

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 ☐		

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 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	

4

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)

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

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) If **YES**, please complete the rest of the muscle cramping section below:- If **NO**, continue completing the questionnaire from section I.	Yes ☐ No ☐
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

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.

	Muscle Group	Muscle (L-left, R-right)	Partial Tear L	Partial Tear R	Complete Tear L	Complete Tear R
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).	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:					
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					
After sustaining the muscle injury approximately how many days were you off from training or competition?	Approximate number of days:					

9

Approximate date of muscle injury?		
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..	
What was the final treatment? **(You may tick more than one block.)**	☐ Rehabilitation (stretching) ☐ Rehabilitation (strengthening) ☐ Rehabilitation (other) ☐ Strapping/taping ☐ Surgery ☐ Other..	
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..	

Associated injuries (Injuries sustained at the same time as the muscle injury)?	☐ Other muscle injury ☐ Tendon injury ☐ Ligament Injury ☐ Bone bruising ☐ Other…………………………………………………………

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 ligament.	☐ 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: ____________________________________

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 ______________________________)

Which of the following symptoms were present **before** the injury	▯ Pain (less than 1 week) ▯ Pain (1-4 weeks) ▯ Pain (> 4 weeks)	▯ Stiffness ▯ Swelling ▯ None
Which of the following symptoms were present **after** the injury	▯ Pain (less than 1 week) ▯ Pain (1-4 weeks) ▯ Pain (> 4 weeks)	▯ Stiffness ▯ Swelling ▯ 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 ___________

Subject No: __________

Section L. Details of Any Other Chronic (Longstanding) <u>Current</u> 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 ☐ Elbow ☐ Hamstring ☐ Neck ☐ Forearm ☐ Quadriceps ☐ Face ☐ Wrist ☐ Knee ☐ Front chest ☐ Finger ☐ Shin ☐ Back chest ☐ Lower back ☐ Achilles ☐ Shoulder ☐ Hip ☐ Ankle ☐ Upper arm ☐ Thigh ☐ Foot Other (Specify: ______________________)
Please indicate the type of structure that was injured	☐ Muscle ☐ Ligament ☐ Tendon ☐ Joint ☐ Bone 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 ☐ Tablets ☐ Stretches ☐ Cortisone injection ☐ Physiotherapy ☐ Other injection ☐ Surgery ☐ Orthotics ☐ Strengthening exercises ☐ Equipment change Other (Specify: ______________________)

6. Diagnostic criteria forms

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 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: _______/_______/______

Investigator: Prof M Schwellnus

Signature: ____________________________

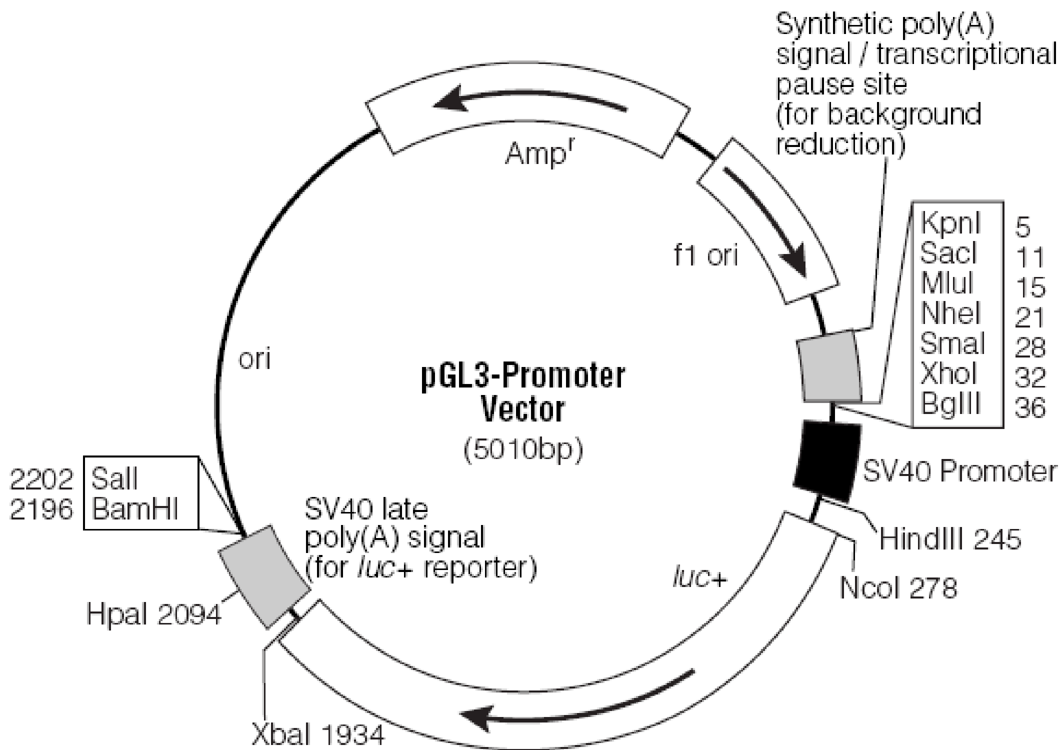
References:

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.

The University of Cape Town is committed to policies of equal opportunity and affirmative action which are essential to its mission of promoting critical inquiry and scholarship

A.2 Appendix B

1. The pGL3-Promoter vector



The pGL3-Promoter Vector (5010 bp) map (Promega Corporation, Madison, Wisconsin, USA)

2. Reagents and solutions used for bacterial work and extraction of plasmid DNA.

SOB (Super Optimal Broth) (For plating):

2% w/v Bacto-Tryptone

0.5% w/v Bacto-Yeast Extract

10mM NaCl

2.5mM KCl

10mM $MgCl_2$

10mM $MgSO_4$

1.5% Agar

H_2O to 2L; Autoclaved

SOC (Super Optimal Broth with Catabolic repression) (Growth medium):

2% w/v Bacto-Tryptone

0.5% w/v Bacto-Yeast Extract

10mM NaCl

2.5mM KCl

10mM $MgCl_2$

10mM $MgSO_4$

20mM Glucose

H_2O to 2L; Autoclaved

TFB broth

2% w/v Difco Bacto-Tryptone

0.5% w/v Difco Bacto-Yeast Extract

0.4% $MgSO_4$

10mM KCl

H_2O to 1L; Autoclaved

TFB 1(Transformation buffer 1) solution

5ml 1M RbCl

0.495g $MnCl_2.4H_2O$

0.147g KOAc

750 mM $CaCl_2.2H_2O$

15 ml of 50% Glycerol

Adjust to pH 5.8 with glacial Acetic acid

H_2O to 50ml; Filter sterilized

TFB 2 (Transformation buffer 2) solution

100 mM Mops (pH 7.0 with NaOH)

1M RbCl

750 mM $CaCl_2$

15 ml of 50% Glycerol

H_2O to 50ml; Filter sterilized

Solution I (GTE)

50mM Glucose

25 mM Tris (pH 8.0)

10mM EDTA

H_2O to 200ml; Filter sterilized

Solution II (Made fresh on the day)

0.2 M NaOH

1% SDS

H_2O to 50ml; Filter sterilized

Solution III

60ml of 5M KOAc (pH 4.8)

11.5ml of Glacial Acetic acid

H_2O to 100ml; Autoclaved

TE buffer pH 7.5

10 mM Tris-HCl (pH 7.5)

1 mM EDTA (pH 8.0)

H_2O to 1 L; Autoclaved

A.3 Appendix C

DNA extraction from whole blood for the genetic study:

1. Draw 5ml of blood into an EDTA vacutainer tube (Purple top).
2. Blood can be stored at 4°C up to 1 week before the DNA is extracted.
3. Transfer the blood to a sterile 15ml polypropylene tube.
4. Add 2volumes (10ml) of TKM1 buffer containing 2.5% NP40.
5. Mix by inverting several times and incubate at room temperature for 10mins in order to enhance the haemolysis of the red blood cells.
6. Centrifuge at 3000rpm (1200Xg) at room temperature for 10mins.
7. Decant the supernatant leaving the white pellet at the bottom of the tube.
8. Add 1 volume (5ml) of TKM1 buffer (without NP40) and invert and vortex the solution.
9. Centrifuge at 3000rpm (1200Xg) at room temperature for 10mins.
10. Decant the supernatant leaving the white pellet at the bottom of the tube.
11. Repeat spteps 7-10 until the pellet in the bottom of the tubes is clean and white.
12. Add 800µl of TKM2 buffer and 50µl of the 10% SDS solution.
13. Vortex and then mix using a blue pipette tip in order to assist in the lyses of the white blood cells.
14. Incubate for 60mins at 55°C in a water bath.
15. Make sure the pellet is totally dissolved before the next steps.
16. Add 150µl of 5M NaClO4 then 500µl of the molecular biology grade chloroform and vortex the solution.
17. Transfer the solution to a sterile 1.5ml microfuge tubes.
18. Centrifuge at 1300rpm at room temperature for 5mins.

19. Carefully transfer 500µl of the top aqueous phase to a new sterile microfuge tube.
20. Add 1ml of absolute ethanol and invert until DNA precipitates.
21. Centrifuge at 1300rpm at room temperature for 5-10mins
22. Carefully tip off the supernatant leaving the DNA pellet at the bottom of the tube.
23. Allow the DNA pellet to air dry completely and add 200µl of 1X TE buffer.
24. Incubate the tubes at 65°C for 15 min in a heating block.
25. Store DNA at 4°C.

Reagents:

TKM1 buffer pH7.6

10mM Tris-HCL

10mM KCl

10mM MgCl2.6H2O

2mM EDTA

Make up 1 Volume which includes 2.5% NP40 and 1 Volume without NP40

H2O to 500ml; Autoclaved.

TKM2 buffer pH7.6

10mM Tris-HCL

10mM KCl

10mM MgCl2.6H2O

2mM EDTA

0.4 M NaCl

H2O to 200ml; Autoclaved.

10% SDS

20ml SDS dissolved in autoclaved H_2O to 200ml

Filter sterilised.

1X TE buffer pH 8.0

10mM Tris-HCL

1mM EDTA

H2O to 100ml; Autoclaved.

5M $NaClO_4$

61.2 grams (5M, MW 122.4)

H2O to 100ml; Autoclaved.

Protocol by Lahiri and Nurnberger (1991) with modifications by Mokone *et al.* (2005)

A.4 Appendix D

1. Additional tables of the physiological characteristics per genotype groups for the overall South African participants

Physiological characteristics of the three **rs71746744** (AGGG/-) genotype groups of the South African participants

	-/-	AGGG/-	AGGG/AGGG	P-value
Gender (% Male)	80.00 (10)	57.35 (68)	63.01(73)	0.366
Age (yrs) [a]	37.9 ± 12.1 (10)	41.6 ± 13.7 (65)	40.5 ± 12.1 (70)	0.676
Height (cm)	174.3 ± 7.4 (10)	174.6 ± 10.0 (66)	174.6 ± 9.6(66)	0.996
Weight (kg)	78.1 ± 13.6 (10)	72.2 ± 12.7 (67)	74.3 ± 12.3 (70)	0.325
BMI (kg/m^2)	25.6 ± 3.8 (10)	23.6 ± 2.8 (64)	24.3 ± 3.1 (66)	0.096

Except for gender, which is presented as a frequency, values are presented as mean ± standard deviations. The number of participants with non-missing data is in parenthesis.
[a] Age at recruitment.
yrs, years; kg, kilogram; cm, centimetre.

Physiological characteristics of the three **rs16399** (ATCT/-) genotype groups of the South African participants

	ATCT/ATCT	ATCT/-	-/-	P-value
Gender (% Male)	80.00 (15)	63.33 (90)	70.21 (94)	0.350
Age (yrs) [a]	38.6 ± 10.6 (15)	40.1 ± 12.6 (85)	41.2 ± 12.2 (89)	0.696
Height (cm)	175.4 ± 8.2 (15)	175.3 ± 9.4 (85)	176.2 ± 9.5 (91)	0.817
Weight (kg)	80.6 ± 13.6 (15)	72.6 ± 12.4 (87)	74.6 ± 13.4 (87)	0.084
BMI (kg/m^2)	25.9 ± 3.1 (14)	23.5 ± 2.9 (83)	23.9 ± 3.0 (86)	0.028

Except for gender, which is presented as a frequency, values are presented as mean ± standard deviations. The number of participants with non-missing data is in parenthesis.
[a] Age at recruitment.
yrs, years; kg, kilogram; cm, centimetre.

Physiological characteristics of the three **rs1134170** (A/T) genotype groups of the South African participants

	AA	AT	TT	P-value
Gender (% Male)	-	66.34 (101)	69.15 (94)	0.675
Age (yrs) [a]	-	40.1 ± 12.0 (97)	41.0 ± 11.9 (89)	0.618
Height (cm)	-	175.3 ± 8.9 (97)	176.2 ± 9.5 (87)	0.269
Weight (kg)	-	73.3 ± 12.3 (100)	75.4 ± 13.6 (90)	0.520
BMI (kg/m^2)	-	23.7± 3.2 (94)	24.1 ± 3.0 (86)	0.430

Except for gender, which is presented as a frequency, values are presented as mean ± standard deviations. The number of participants with non-missing data is in parenthesis.
[a] Age at recruitment.
yrs, years; kg, kilogram; cm, centimetre.

2. Additional tables of the physiological characteristics per genotype groups for the overall Australian participants

Physiological characteristics of the three **rs71746744** (AGGG/-) genotype groups of the Australia participants

	-/-	AGGG/-	AGGG/AGGG	P-value
Gender (% Male)	50.00 (18)	42.86 (49)	59.42 (69)	0.203
Age (yrs) [a]	41.7 ± 13.4 (18)	39.9 ± 13.6 (49)	43.3 ± 14.1 (68)	0.426
Height (cm)	173.6 ± 7.4 (17)	170.9 ± 9.7 (49)	173.7 ± 8.7(68)	0.220
Weight (kg)	75.6 ± 19.7 (17)	74.8 ± 16.5 (49)	77.4 ± 13.2 (70)	0.644
BMI (kg/m^2)	24.9 ± 4.8 (17)	25.6 ± 5.2 (49)	25.6 ± 3.8 (68)	0.822

Except for gender, which is presented as a frequency, values are presented as mean ± standard deviations. The number of participants with non-missing data is in parenthesis.
[a] Age at recruitment.
yrs, years; kg, kilogram; cm, centimetre.

Physiological characteristics of the three **rs16399** (ATCT/-) genotype groups of the Australia participants

	ATCT/ATCT	ATCT/-	-/-	P-value
Gender (% Male)	54.55 (22)	42.19 (64)	58.14 (86)	0.148
Age (yrs) [a]	44.0 ± 14.1 (22)	39.8 ± 13.7 (64)	42.1 ± 12.6 (85)	0.369
Height (cm)	171.1 ± 9.7 (21)	171.1 ± 9.5 (64)	173.1 ± 8.6 (86)	0.356
Weight (kg)	81.33 ± 21.7 (21)	72.0 ± 12.7 (64)	76.3 ± 13.7 (87)	0.027
BMI (kg/m^2)	27.8 ± 7.0 (21)	24.5 ± 3.5 (64)	25.4 ± 3.7 (86)	0.009

Except for gender, which is presented as a frequency, values are presented as mean ± standard deviations. The number of participants with non-missing data is in parenthesis.
[a] Age at recruitment.
yrs, years; kg, kilogram; cm, centimetre.

Physiological characteristics of the three **rs1134170** (A/T) genotype groups of the Australia participants

	AA	AT	TT	P-value
Gender (% Male)	50.0 (4)	42.0 (81)	53.8 (65)	0.359
Age (yrs) [a]	38.75 ± 13.5 (4)	41.0 ± 13.9 (81)	43.2 ± 13.1 (64)	0.554
Height (cm)	170.5 ± 10.9 (4)	170.4 ± 9.5 (79)	173.1 ± 8.5 (65)	0.971
Weight (kg)	75.0 ± 12.5 (4)	75.2 ± 15.4 (80)	75.8 ± 14.2 (66)	0.199
BMI (kg/m^2)	25.7 ± 2.7 (4)	25.2 ± 3.7 (65)	25.8 ± 4.7 (79)	0.657

Except for gender, which is presented as a frequency, values are presented as mean ± standard deviations. The number of participants with non-missing data is in parenthesis.
[a] Age at recruitment.
yrs, years; kg, kilogram; cm, centimetre.

3. **Additional tables of the physiological characteristics per genotype groups for the SA tendinopathic participants specifically**

Physiological characteristics of the three **rs71746744** (AGGG/-) genotype groups of the South African tendinopathic (SA TEN) participants

	-/-	AGGG/-	AGGG/AGGG	P-value
Gender (% Male)	100 (1)	52.63 (19)	67.74 (31)	0.416
Age (yrs) [a]	44.0 (1)	53.1 ± 11.8 (19)	46.6 ± 11.0 (29)	0.153
Height (cm)	177.0 (1)	174.1 ± 10.1 (18)	174.8 ± 8.9 (26)	0.934
Weight (kg)	82.0 (1)	73.7 ± 14.6 (18)	78.2 ± 12.3 (28)	0.492
BMI (kg/m^2)	26.2 (1)	24.12 ± 2.9 (18)	25.2 ± 3.0 (26)	0.471

Except for gender, which is presented as a frequency, values are presented as mean ± standard deviations. The number of participants with non-missing data is in parenthesis.
[a] Age at recruitment.
yrs, years; kg, kilogram; cm, centimetre.

Physiological characteristics of the three **rs16399** (ATCT/-) genotype groups of the South African tendinopathic (SA TEN) participants

	ATCT/ATCT	ATCT/-	-/-	P-value
Gender (% Male)	100 (3)	64.29 (28)	77.78 (36)	0.271
Age (yrs) [a]	42.7 ± 6.1 (3)	48.7 ± 12.2 (27)	47.8 ± 11.3 (33)	0.687
Height (cm)	181.83 ± 4.4 (3)	175.3 ± 9.2 (25)	177.8 ± 8.4 (31)	0.340
Weight (kg)	92.2 ± 11.7 (3)	73.5 ± 14.4 (25)	79.9 ± 13.2 (33)	0.043
BMI (kg/m^2)	27.8 ± 2.4 (3)	23.8 ± 3.5 (25)	24.9 ± 3.3 (31)	0.106

Except for gender, which is presented as a frequency, values are presented as mean ± standard deviations. The number of participants with non-missing data is in parenthesis.
[a] Age at recruitment.
yrs, years; kg, kilogram; cm, centimetre.

Physiological characteristics of the three **rs1134170** (A/T) genotype groups of the South African (SA TEN) participants

	AA	**AT**	**TT**	**P-value**
Gender (% Male)	-	70.37 (27)	76.32 (38)	0.591
Age (yrs) [a]	-	48.3 ± 11.0 (27)	47.1 ± 11.8 (35)	0.674
Height (cm)	-	175.9 ± 8.5 (25)	177.6 ± 8.7 (32)	0.441
Weight (kg)	-	73.9 ± 14.0 (26)	80.7 ± 13.7 (34)	0.065
BMI (kg/m^2)	-	23.8 ± 3.5 (25)	25.3 ± 3.4 (32)	0.114

Except for gender, which is presented as a frequency, values are presented as mean ± standard deviations. The number of participants with non-missing data is in parenthesis.
[a] Age at recruitment.
yrs, years; kg, kilogram; cm, centimetre.

4. Additional tables of the physiological characteristics per genotype groups for the AUS tendinopathic participants specifically

Physiological characteristics of the three **rs71746744** (AGGG/-) genotype groups of the Australian (AUS TEN) participants

	-/-	**AGGG/-**	**AGGG/AGGG**	**P-value**
Gender (% Male)	100 (3)	66.67 (12)	75.00 (24)	0.494
Age (yrs) [a]	44.7 ± 12.0 (3)	53.4 ± 11.9 (12)	52.0 ± 12.8 (24)	0.561
Height (cm)	174.0 ± 6.56 (3)	170.6 ± 10.4 (12)	175.5 ± 7.8 (22)	0.305
Weight (kg)	76.0 ± 16.4 (3)	84.5 ± 19.5 (12)	79.9 ± 11.5 (24)	0.565
BMI (kg/m^2)	25.0 ± 4.4 (3)	28.9 ± 5.2 (12)	26.0 ± 3.8 (22)	0.152

Except for gender, which is presented as a frequency, values are presented as mean ± standard deviations. The number of participants with non-missing data is in parenthesis.
[a] Age at recruitment.
yrs, years; kg, kilogram; cm, centimetre.

Physiological characteristics of the three **rs16399** (ATCT/-) genotype groups of the Australian (AUS TEN) participants

	ATCT/ATCT	ATCT/-	-/-	P-value
Gender (% Male)	100 (2)	58.82 (17)	75.76 (33)	0.300
Age (yrs) [a]	56.0 ± 5.7 (2)	46.8 ± 14.6 (17)	49.0 ± 11.4 (33)	0.582
Height (cm)	174.8 ± 0.4 (2)	169.1 ± 11.2 (17)	175.8 ± 6.9 (32)	0.040
Weight (kg)	107.0 ± 29.7 (2)	76.6 ± 14.6 (17)	81.0 ± 12.1 (33)	0.016
BMI (kg/m^2)	35.0 ± 9.6 (2)	26.7 ± 3.7 (17)	26.3 ± 3.6 (32)	0.012

Except for gender, which is presented as a frequency, values are presented as mean ± standard deviations. The number of participants with non-missing data is in parenthesis.
[a] Age at recruitment.
yrs, years; kg, kilogram; cm, centimetre.

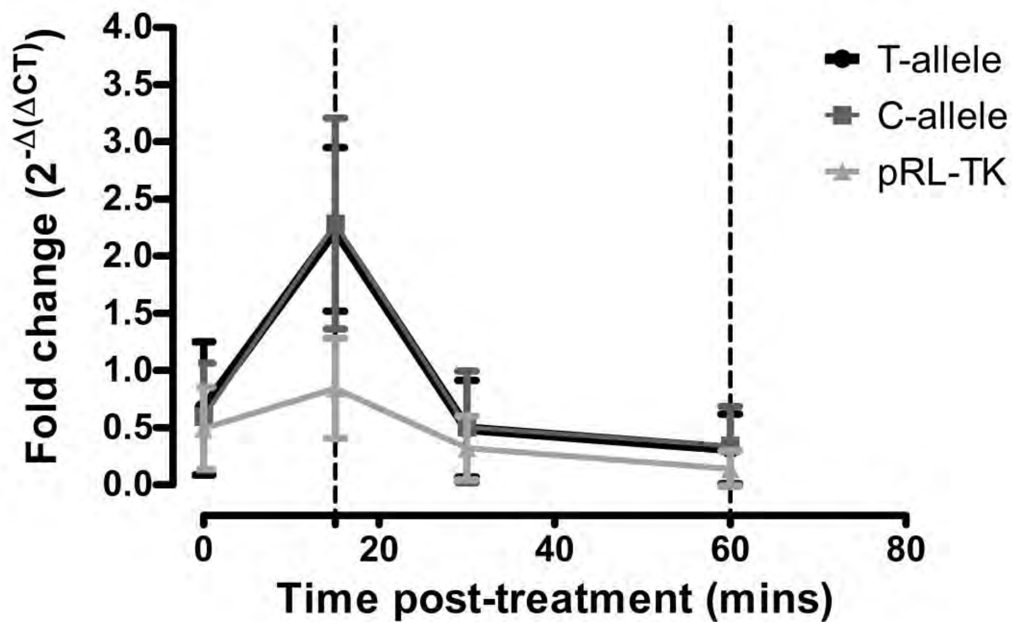
Physiological characteristics of the three **rs1134170** (A/T) genotype groups of the Australian (AUS TEN) participants

	AA	AT	TT	P-value
Gender (% Male)	-	57.89 (19)	77.27 (22)	0.184
Age (yrs) [a]	-	49.2 ± 14.5 (19)	49.4 ± 10.8 (22)	0.959
Height (cm)	-	170.6 ± 10.3 (18)	176.1 ± 6.8 (21)	0.053
Weight (kg)	-	83.3 ± 18.1 (19)	81.4 ± 11.0 (22)	0.687
BMI (kg/m^2)	-	28.5 ± 5.1 (18)	26.3 ± 3.1 (21)	0.112

Except for gender, which is presented as a frequency, values are presented as mean ± standard deviations. The number of participants with non-missing data is in parenthesis.
[a] Age at recruitment.
yrs, years; kg, kilogram; cm, centimetre.

A.5 Appendix E

Messenger RNA decay assay



Messenger RNA decay assay displaying the fold change of transcripts of the C and T-allelic form within the pGL3-COL5A1-3′UTR, as well as that of pRL-TK over time post-treatment. The solid black line represents the decay of the T-allelic form; the solid dark grey line represents the decay of the C-allelic form while the pale grey solid line represents the decay of transcripts of pRL-TK (transfection efficiency control). The broken black lines delimit the area used to calculate the area under the curve (AUC).