The Role of Novel Genetic Variants and DNA Methylation as Risk Factors for Tendon Pathology in Physically Active Individuals

Submitted for the Degree of

DOCTOR OF PHILOSPHY

At the University of Northampton

May 2016

Rebecca Rickaby

© Rebecca Rickaby 2016.

This thesis is copyright material and no quotation from it may be published without proper acknowledgement.

ACC No.: S1006412
CLASS NO 616.75 RIC

DEDICATION

I dedicate this thesis to my mum, for her support and encouragement in helping me complete this work.

ACKNOWLEDGEMENTS

Throughout my PhD journey I have met many kind and caring people. Not only those who have directly influenced my research, but those who have been there as friends and shown support and wise words of advice in times of need. I would not have got here without you all.

I would like to thank my supervisory team. I huge thank you to my director of studies and first supervisor Dr Stuart Raleigh, who has guided and supported me throughout every step of my PhD. Stuart's dedication to the research has helped to motivate me and keep me focussed throughout. It has not been easy. Without his encouragement I would not have got to this stage. I would also like to thank my second supervisor Prof Bill Ribbans. His clinical insights have been invaluable in aiding my understanding of tendinopathy. Thanks also go to my third supervisor Dr Mike Posthumus from the University of Cape Town.

This thesis was supported by funding secured by Dr Stuart Raleigh from the Rosetrees Trust and the University of Northampton. Therefore, gratitude must be extended to both institutions and again to my director of studies, for selecting me and enabling me to undertake and complete this PhD.

Other important 'thank yous' must go to the School of Health. The school provided me with a great working environment and all I needed to conduct my studies and more. I was also supported by the school at several conferences through grant funding. I am grateful to the School of Health team and especially to Steph Arnold for her help with processing many endless orders of lab

consumables and equipment. Special thanks must go to Val Wright, Lin Chapman and Dan Shaw, for their support and maintenance of the labs and equipment.

Last but not least, I need to thank Dr Louis El Khoury for all his expert help and advice with the practical aspects of my research. Louis was invaluable in facilitating my integration into the lab environment and being on hand when I was leaning things for the first time. Thanks also to my PhD colleagues who I have shared various offices with over the past several years, whom have been there for support both as friends and professional colleagues.

DEDICATION i
ACKNOWLEDGEMENTS ii
TABLE OF CONTENTS iv
LIST OF FIGURES ix
LIST OF TABLES xii
LIST OF SCIENTIFIC OUTPUTS FROM THIS THESIS xv
ABREVIATIONS xvii
ABSTRACT xix

CHAPTER 1: LITERATURE REVIEW

1.1 INTRODUCTION 1
1.2 TENDON ANATOMY AND PHYSIOLOGY 1
1.2.1 Achilles Tendon 8
1.2.2 Patellar Tendon 9
1.3 TENDON PATHOLOGY 10
1.3.1 Epidemiology 12
1.3.2 Achilles Tendon Pathology 13
1.3.2.1 Non-Insertional Achilles Tendinopathy 13
1.3.2.2 Insertional Achilles Tendinopathy 14
1.3.2.3 Achilles Tendon Rupture 15
1.3.3 Patellar Tendinopathy 16
1.3.4 Patellar Tendon Rupture 17
1.3.5 Current Management and Treatment 18
1.4 RISK FACTORS ASSOCIATED WITH TENDON PATHOLOGY 20
1.4.1 Intrinsic Risk Factors 21
1.4.1.1 Age and Gender 22
1.4.1.2 Anatomical Factors 22
1.4.1.3 Systemic Diseases 22
1.4.2 Extrinsic Risk Factors 23
1.4.2.1 Activity Levels 23
1.4.2.2 Environment, Training Errors, Footwear 23
1.4.2.3 Steroids and Fluoroquinolones 23
1.4.3 Genetic Risk Factors 25
1.4.3.1 Repeat Polymorphisms and Single Nucleotide Polymorphisms 26

1.5 COPY NUMBER VARIATION 33
1.5.1 The Effect of CNV on the Genome 36
1.6 EPIGENETICS 38
1.6.1 DNA Methylation 39
1.6.2 DNA Methylation and Disease 40
1.7 AIMS AND OBJECTIVES 42

CHAPTER 2: METHODS
2.1 INVESTIGATING GENETIC ASSOCIATIONS IN COMPLEX DISEASES 44
2.2 PARTICIPANTS 47
2.2.1 Ethical Approval 47
2.2.2 British Achilles Tendon Cohort 47
2.2.3 South African and Australian Achilles Tendon Cohorts 51
2.2.4 Australian Patellar Tendon Tissue Samples 55
2.2.5 DNA Extraction and Quantification 55
2.3 QUANTITATIVE PCR METHODS 57
2.3.1 TaqMan Technology 57
2.3.2 SNP Selection and Genotyping 58
2.3.3 CNV Selection and Genotyping 62
2.3.4 QPCR Efficiency 68
2.4 DNA METHYLATION AND PYROSEQUENCING 69
2.4.1 Bisulphite Treatment 71
2.4.2 PCR Amplification of Target CpG Region 74
2.4.3 Gel Electrophoresis 75
2.4.4 Pyrosequencing 76
2.5 STATISTICS 79
2.5.1 Power Calculations 79
2.5.2 Data Analysis 80

CHAPTER 3: COPY NUMBER VARIANTION WITHIN SELECTED COLLAGEN AND TNC GENES AS RISK FACTORS FOR ACHILLES TENDON PATHOLOGY
3.1 INTRODUCTION 83
3.1.1 COL5A1 83
3.1.2 COL4A1 87

3.1.3 TNC 89
3.1.4 Chapter Aims 91
3.2 METHODS 92
3.2.1 Participants 92
3.2.2 DNA Collection and Extraction 92
3.2.3 Copy Number Genotyping 92
3.2.4 Data Analysis 93
3.3 RESULTS 94
3.3.1 *COL5A1* Copy Number Frequencies 97
3.3.2 *COL4A1* Copy Number Frequencies 100
3.3.3 *TNC* Copy Number Frequencies 104
3.4 CHAPTER SPECIFIC CONCLUSIONS 107

CHAPTER 4: COPY NUMBER VARIATION WITHIN *MMP* AND *ADAM* GENES AS RISK FACTORS FOR ACHILLES TENDON PATHOLOGY
4.1 INTRODUCTION 109
4.1.1 MMP3 109
4.1.2 MMP23B 113
4.1.3 ADAM8 115
4.1.4 Chapter Aims 118
4.2 METHODS 119
4.2.1 Participants 119
4.2.2 DNA Collection and Extraction 119
4.2.3 Copy Number Genotyping 119
4.2.4 Data Analysis 120
4.3 RESULTS 121
4.3.1 *MMP3* Copy Number Frequencies 123
4.3.2 *MMP23B* Copy Number Frequencies 126
4.3.3 *ADAM8* Copy Number Frequencies 128
4.4 CHAPTER SPECIFIC CONCLUSIONS 130

CHAPTER 5: INVESTIGATING THE ROLE OF APOPTOSIS ASSOCIATED GENES AND THE RISK OF ACHILLES TENDON PATHOLOGY
5.1 INTRODUCTION 132
5.1.1 TNFRSF1A 132

5.1.2 CASP8 136
5.1.3 CASP3 138
5.1.4 Chapter Aims 140
5.2 METHODS 141
5.2.1 Participants 141
5.2.2 DNA Collection and Extraction 141
5.2.3 TaqMan SNP Genotyping 141
5.2.4 Copy Number Genotyping 142
5.2.5 Data Analysis 142
5.3 RESULTS 143
5.3.1 *TNFRSF1A* Genotype Frequencies 145
5.3.2 *CASP8* Copy Number Frequencies 147
5.3.3 *CASP3* Genotype Frequencies 149
5.4 CHAPTER SPECIFIC CONCLUSIONS 151

CHAPTER 6: DNA METHYLATION WITHIN THE *ADAMTS4*, *TIMP1* AND *MMP11* GENE PROMOTERS IN PATELLAR TENDINOPATHY
6.1 INTRODUCTION 153
6.1.1 ADAMTS4 153
6.1.2 TIMP1 154
6.1.3 MMP11 155
6.1.4 Chapter Aims 155
6.2 METHODS 156
6.2.1 Participants 156
6.2.2 DNA Collection and Bisulphite Treatment 156
6.2.3 Loci Selection 156
6.2.3.1 *ADAMTS4* Hs_NDUFS2_01_PM Assay 157
6.2.3.2 *TIMP1* Custom Designed Assays 159
6.2.3.3 *MMP11* Hs_CHCHD10_01_PM Assay 161
6.2.4 PCR Amplification and Gel Electrophoresis 163
6.2.5 Pyrosequencing 163
6.2.6 Data Analysis 163
6.3 RESULTS 164
6.3.1 *ADAMTS4* CpG Methylation Status 169
6.3.2 *TIMP1* CpG Methylation Status 171

6.3.3 *MMP11* CpG Methylation Status 173
6.4 CHAPTER SPECIFIC CONCLUSIONS 175

CHAPTER 7: DISCUSSION

7.1 Copy Number and SNP Genotyping Studies 178
7.2 Pyrosequencing Studies 186
7.3 Limitations and Future Work 189
7.4 Concluding Remarks 194

REFERENCES 195

APPENDIX I **Ethical Approval** 206
APPENDIX II **Participant Information Sheet & Consent Form** 212
APPENDIX III **Participant Questionnaires** 218
APPENDIX IV **Protocols** 243
APPENDIX V **QPCR Efficiency** 245
APPENDIX VI **Publications** 256

LIST OF FIGURES

Figure 1.1 The structural organisation of collagen in tendon 3

Figure 1.2 Schematic overview of the tendon ECM showing the network of connected molecules, including collagen and proteoglycans, surrounding and supporting the cell (tenocyte) 5

Figure 1.3 Diagram showing the general structural ECM components of tendon, consisting of major and minor fibrillar collagens and associated proteins including tenascin C and matrix metalloproteinase 3 (MMP3) 7

Figure 1.4 Image of Achilles tendon and associated gastrocnemius and soleus muscle 8

Figure 1.5 Image of the right knee showing the patellar tendon and associated bones (femur, patella, tibia and fibula) and knee ligaments 9

Figure 1.6 MRIs of a normal and tendinopathic Achilles tendon 14

Figure 1.7 MRIs of a normal and abnormal patellar tendon 17

Figure 1.8 Risk Factors for Tendon Injury 21

Figure 1.9 Timeline of Studies in Human Genetic Variation 34

Figure 2.1 Summary of the stages in a real-time qPCR reaction containing TaqMan SNP Genotyping Assay 61

Figure 2.2 Summary of the stages in a duplex real-time qPCR reaction containing TaqMan Copy Number and Reference Assay 66

Figure 2.3 Amplification plot showing the quantitative determination of copy number using the ΔΔCt method 67

Figure 2.4 Overview of the Pyrosequencing cascade 78

Figure 3.1 The relationship between *COL5A1* genotypes, disease risk and the associated effect on connective tissue and collagen fibre assembly ... 85
Figure 3.2 Location of the Hs00180523_cn and Hs02980647_cn Copy Number Assays within the *COL5A1* gene ... 86
Figure 3.3 Location of the Hs00739915_cn and Hs02119212_cn Copy Number Assays within the *COL4A1* gene ... 88
Figure 3.4 Location of the Hs00634176_cn and Hs06903309_cn Copy Number Assays within the *TNC* gene ... 90
Figure 3.5 Typical CN profiles generated from the CopyCaller Software ... 95
Figure 3.6 Scatter plot showing the raw continuous CN values on the y-axis, plotted against the discrete integer CN values on the x-axis ... 96
Figure 4.1 Location of the Hs02908568_cn and Hs02276714_cn Copy Number Assays within the *MMP3* gene ... 112
Figure 4.2 Location of the Hs07485808_cn Copy Number Assay within the *MMP23B* gene ... 114
Figure 4.3 Location of the Hs02429708_cn Copy Number Assay within the *ADAM8* gene ... 117
Figure 4.4 Scatter plot showing the raw continuous CN values on the y-axis, plotted against the discrete integer CN values on the x-axis ... 122
Figure 5.1 TNFR1 Pathway showing the cascade of reactions following the binding of TNF-α to the TNFR1 receptor ... 133
Figure 5.2 Location of the rs4149577 SNP within the *TNFRSF1A* gene ... 135
Figure 5.3 Location of the Hs02601709_cn Copy Number Assay within the *CASP8* gene ... 137
Figure 5.4 Location of the rs1049253 SNP within the *CASP3* gene ... 139

Figure 5.5 Typical allelic discrimination cluster plot for the rs1049253 SNP in the *CASP3* gene 144
Figure 6.1 Location of the Hs_NDUFS2_01_PM CpG Assay within the *ADAMTS4* gene 158
Figure 6.2 Location of the Custom Designed CpG Assays within the *TIMP1* gene 160
Figure 6.3 Location of the Hs_CHCHD10_01_PM CpG Assay within the *MMP11* gene 162
Figure 6.4 Image of *ADAMTS4* PCR product from a 1% agarose gel electrophoresis 164
Figure 6.5 Image of *TIMP1* PCR product from a 1% agarose gel electrophoresis 165
Figure 6.6 Image of *MMP11* PCR product from a 1% agarose gel electrophoresis 165
Figure 6.7 Typical Pyrogram graphs for the *ADAMTS4* gene 166
Figure 6.8 Typical Pyrogram graphs for the *TIMP1* gene 167
Figure 6.9 Typical Pyrogram graphs for the *MMP11* gene 168
Figure 6.10 Comparison of the mean DNA methylation between CON and PT at 6 CpG sites within the promoter of the *ADAMTS4* gene 170
Figure 6.11 Comparison of the mean DNA methylation between the CON and PT groups within the promoter of the *TIMP1* gene 172
Figure 6.12 Comparison of the mean DNA methylation between the CON and PT groups within the promoter of the *MMP11* gene 174

LIST OF TABLES

Table 1.1 Genetic Variants Associated with human Achilles Tendon Pathology 31

Table 2.1 General characteristics of the CON, ATP, TEN, and RUP groups in the UK cohort 50

Table 2.2 General characteristics of the CON, ATP, TEN, and RUP groups in the SA cohort 53

Table 2.3 General characteristics of the CON, ATP and TEN groups in the AUS cohort 54

Table 2.4 Reaction Components for a Typical SNP Genotyping Real-Time PCR 59

Table 2.5 Fluorescence detected and the corresponding genotype identified 61

Table 2.6 TaqMan Copy Number Assays Investigated 63

Table 2.7 Reaction Components for Copy Number Real-Time PCR 64

Table 2.8 PyroMark CpG Assays 70

Table 2.9 The Effect of Bisulfite Treatment 71

Table 2.10 Reaction Components for Bisulfite Treatment 72

Table 2.11 Thermal Cycling Conditions for Bisulfite Treatment 72

Table 2.12 Reaction Components for PyroMark PCR 74

Table 2.13 Thermal Cycling Conditions for PyroMark PCR 75

Table 2.14 Reagents for PCR Product Immobilisation 76

Table 2.15 Steps for Capturing Sepharose Beads using PyroMark Vacuum Workstation 77

Table 3.1a Copy number distribution within the *COL5A1* gene in the cases (ATP/TEN/RUP) and CON groups using the Hs00180523_cn assay 98

Table 3.1b Copy number distribution within the *COL5A1* gene in the cases (ATP/TEN/RUP) and CON groups using the Hs02980647_cn assay 99

Table 3.2a Copy number distribution within the *COL4A1* gene in the cases (ATP/TEN/RUP) and CON groups using the Hs00739915_cn assay 102

Table 3.2b Copy number distribution within the *COL4A1* gene in the cases (ATP/TEN/RUP) and CON groups using the Hs02119212_cn assay 103

Table 3.3a Copy number distribution within the *TNC* gene in the cases (ATP/TEN/RUP) and CON groups using the Hs00634176_cn assay 105

Table 3.3b Copy number distribution within the *TNC* gene in the cases (ATP/TEN/RUP) and CON groups using the Hs06903309_cn assay 106

Table 4.1a Copy number distribution within the *MMP3* gene in the cases (ATP/TEN/RUP) and CON groups using the Hs02276712_cn assay 124

Table 4.1b Copy number distribution within the *MMP3* gene in the cases (ATP/TEN/RUP) and CON groups using the Hs02908568_cn assay 125

Table 4.2 Copy number distribution within the *MMP23B* gene in the cases (ATP/TEN/RUP) and CON groups using the Hs07485808_cn assay 127

Table 4.3 Copy number distribution within the *ADAM8* gene in the cases (ATP/TEN/RUP) and CON groups using the Hs02429708_cn assay 129

Table 5.1 Genotype frequency distribution for the rs4149577 variant within the *TNFRSF1A* gene in the cases (ATP/TEN/RUP) and CON groups 146

Table 5.2 Copy number distribution within the *CASP8* gene in the cases (ATP/TEN/RUP) and CON groups using the Hs02601709_cn assay 148

Table 5.3 Genotype frequency distribution for the rs1049253 variant within the *CASP3* gene in the cases (ATP/TEN/RUP) and CON groups 150

Table 6.1 Mean methylation percentages within CpG sites of the *ADAMTS4* gene promoter in Patellar tendinopathy (PT) cases and asymptomatic controls (CON) 169

Table 6.2 Mean methylation percentages within CpG sites of the *TIMP1* gene promoter (custom assay region A) in Patellar tendinopathy (PT) cases and asymptomatic controls (CON) 171

Table 6.3 Mean methylation percentages within CpG sites of the *MMP11* gene promoter in Patellar tendinopathy (PT) cases and asymptomatic controls (CON) 173

LIST OF SCIENTIFIC OUTPUTS FROM THIS THESIS

PAPERS IN INTERNATIONAL PEER-REVIEWED JOURNALS

Rickaby, R., El Khoury, L., Ribbans, W. J. & Raleigh, S. M. (2015) Variation within three apoptosis associated genes as potential risk factors for Achilles tendinopathy in a British based case-control cohort. *Gene*. 571(2), 167-71.

PRESENTATIONS AT INTERNATION CONGRESSES

Raleigh, S. M., El Khoury, L., Rickaby, R. & Samiric, T. Epigenetic changes within the *ADAMTS4* gene promoter may modify risk of patellar tendinopathy. 20th Annual Congress of the European College of Sport Science, Sustainable Sport, Malmo, June 24th-27th, 2015.

Rickaby, R., El Khoury, L., Ribbans, W. J. & Raleigh, S. M. Variation within the *CASP3* gene and the risk of Achilles tendinopathy in a British case-control cohort. Federation of European Biochemical Societies (FEBS) & European Molecular Biology Organisation (EMBO) 50th Anniversary Conference, Paris, August 30th - September 4th, 2014.

Rickaby, R., Collins, M., Posthumus, M. & Raleigh, S. M. Genomic copy number variation within the *TNC* and *ADAM8* genes as potential risk factors for Achilles tendon pathology. 18th Annual Congress of the European College of Sport Science, Unifying Sport Science, Barcelona, June 26-29th, 2013.

INVITED PRESENTATIONS

Rickaby, R., Posthumus, M., Collins, M., Handley, C. J., Cook, J. & Raleigh, S. M. Copy Number Variation in the *COL5A1* Gene and the Risk of Achilles Tendon Pathology. Rosetrees Trust 25th Anniversary Event, 'Bridging the Gap' for Effective Translational Medical Research, London, October 17th, 2012.

INTERNAL PRESENTATIONS

Rickaby, R., Samiric, T. & Raleigh, S. M. DNA Methylation of the *ADAMTS4* Gene in Jumper's Knee. University of Northampton 10th Annual Postgraduate Research Student Poster Competition, Northampton, May 13th, 2015.

ABBREVIATIONS

ADAM(s)	A disintegrin and metalloproteinase(s)
ADAM8	Gene encoding the A disintegrin and metalloproteinase-8
ADAMTS4	Gene encoding the A disintegrin and metalloproteinase with thrombospondin motifs 4
ANOVA	One-was analysis of variance
ATP	Achilles tendon pathology
AUS	Australian
BMI	Body Mass Index
bp	Base pairs
CASP3	Gene encoding for caspase-3
CASP8	Gene encoding for caspase-8
CN	Copy number
CNV	Copy number variation
CNVs	Copy number variants
COL4A1	Gene encoding the α1 chain of type IV collagen
COL5A1	Gene encoding the α1 chain of type V collagen
CON	Control group
DGV	Database of genomic variants
ECM	Extracellular matrix
GDF(s)	Growth differentiation factor(s)
GDF-5	Growth differentiation factor-5
HR	Homologous recombination
HWE	Hardy-Weinberg Equilibrium
kb	Kilobase

miRNA(s)	Micro RNA(s)
MMP(s)	Matrix metalloproteinase(s)
MMP3	Gene encoding the matrix metalloproteinase-3
MMP11	Gene encoding the matrix metalloproteinase-11
MMP23B	Gene encoding the matrix metalloproteinase-23
NAHR	Nonallelic homologous recombination
NHEJ	Nonhomologous end-joining
PCR	Polymerase chain reaction
PT	Patellar tendinopathy
RUP	Achilles tendon rupture group
SA	South African
SNP(s)	Single nucleotide polymorphism(s)
TEN	Achilles tendinopathy group
TIMP(s)	Tissue inhibitor of metalloproteinase(s)
TIMP1	Gene encoding the tissue inhibitor of metalloproteinase-1
TIMP2	Gene encoding the tissue inhibitor of metalloproteinase-2
TNC	The gene encoding for tenacin-C
TNF-α	Tumour necrosis factor-alpha
TNFR1	Tumour necrosis factor receptor 1
TNFRSF1A	Gene encoding the tumour necrosis factor receptor 1
UK	United Kingdom
UTR	Untranslated region

ABSTRACT

Along with the many health benefits of regular physical activity, there is also the risk of injury. Around 30-50% of all sporting injuries involve damage to tendons and in long-distance runners the lifetime risk of developing Achilles tendinopathy exceeds 50%. Patellar tendinopathy (PT) is likewise common among sporting populations, with prevalence rates higher than 30% in both volleyball and basketball players.

The main aims of this thesis were to investigate whether novel genetic variants (copy number variation (CNV) and single nucleotide polymorphisms (SNPs)) in candidate genes were associated with Achilles tendon pathology (ATP) and to investigate whether DNA methylation status was altered in patellar tendinopathy, within the promoter regions of candidate genes.

Copy number variation was investigated in the *ADAM8*, *CASP8*, *COL4A1*, *COL5A1*, *MMP23B*, *MMP3* and *TNC* genes across all ATP cohorts. The British and South African ATP cohorts were genotyped for the *CASP3* rs1049253 and *TNFRSF1A* rs4149577 single nucleotide polymorphisms. These studies were carried out using fluorescence based TaqMan qPCR. DNA methylation was measured in the *ADAMTS4*, *MMP11* and *TIMP1* genes by Pyrosequencing, using DNA isolated from healthy and tendinopathic patellar tendon tissue.

On the whole, the genetic association studies on copy number variation within this thesis showed that CNV within the loci we investigated does not appear to be a risk factor for ATP. Nevertheless, significant differences in discrete copy number were reported within the *COL5A1* gene between the AUS CON and ATP

groups (p = 0.012 using the Hs00180523_cn assay) and within the *COL4A1* gene between the combined CON and ATP groups (p = 0.014 using the Hs00739915_cn assay). There was also a significant difference in copy number within the *MMP3* gene between the AUS CON and ATP groups (p = 0.035 using the Hs02908568_cn). However, these significant findings were not replicated when data were analysed as continuous copy number, therefore further investigation would be needed to replicate these findings in other cohorts. Similarly, the genetic association studies on the single nucleotide polymorphisms *TNFRSF1A* rs4149577 and *CASP3* rs1049253 found no associations with ATP within combined and individual South African and UK cohorts.

The DNA methylation studies within this thesis showed that the methylation status of the promoter regions of the *ADAMTS4*, *MMP11* and *TIMP1* genes was altered at certain CpG sites in patellar tendinopathy compared to controls. There were significant differences in methylation status reported between the CON and PT groups for CpG site 3 of the *ADAMTS4* gene (p = 0.016), CpG site 4 of the *MMP11* gene (p = 0.045) and CpG site 2 of the *TIMP1* gene (p = 0.012).

The results from this thesis provide evidence that copy number variation within the genes studied may not be as important in the risk of developing ATP. This thesis has also excluded two SNP variants as risk factors for ATP in the cohorts investigated. The preliminary studies on DNA methylation status within the *ADAMTS4*, *MMP11* and *TIMP1* genes suggest that DNA methylation within these genes may be an important risk factor for patellar tendinopathy.

CHAPTER 1: LITERATURE REVIEW

1.1 INTRODUCTION

Tendon pathology is a multifactorial disorder prevalent among sporting populations (Meeuwisse 1994; Maffulli et al. 2003). The increased stress exerted on tendons during exercise can lead to damage in both professional and recreational athletes, due to repetitive mechanical loading (Maffulli et al. 2003; Xu and Murrell 2008; Collins and Raleigh 2009). A variety of risk factors are known to predispose to Achilles tendon pathology (ATP) and research has identified that genetics plays a prominent role in an individual's susceptibility (Raleigh and Collins 2012).

1.2 TENDON ANATOMY AND PHYSIOLOGY

In general, a tendon is considered a connective tissues that connects muscle to bone (Benjamin et al. 2008). Tendons transmit the force generated by muscle to bone and therefore facilitate joint movement (Xu and Murrell 2008; Birch 2007). The release and storage of elastic strain energy is another function of tendons, which helps to aid locomotion (Birch 2007). Tendons have two junctional regions, known as the myotendinous junction and the osteotendinous junction (Reinking 2012). Tendons exhibit high mechanical strength, optimal elasticity and good flexibility to allow movement at the joint, while limiting muscle damage by acting as a buffer to absorb external forces (Sharma and Maffulli 2005; Sakabe and Sakai 2011).

Typically tendons consists of fibroblasts called tenocytes (tendon cells) intermixed with a highly organised extracellular matrix (ECM) abundant in collagen (Cook et

al. 2002; Sakabe and Sakai 2011). Tenocytes are flat, spindle shaped cells that are sparsely distributed among the tendon (Cook et al. 2002). These tenocytes are responsible for the synthesis of all ECM components, namely the ground substance and collagen, and constitute around 90-95% of the cellular component of tendons (Khan et al. 1999; Sharma and Maffulli 2005). Collagen is a principle structural protein component of tendons, with type I collagen being the most prevalent (Mokone et al. 2006). However, numerous other collagen types are present, including type II, III, V, VI, IX, and XI (Benjamin et al. 2008). The main role of collagen within the tendon is to resist tension yet allowing a level of compliance and this is believed to be possible due to its hierarchical architecture (Benjamin et al. 2008). Collagen is arranged in hierarchical levels of increasing complexity, starting with the smallest unit, the collagen fibril (Khan et al. 1999; Karousou et al. 2008) (**Figure 1.1**). The fibrils combine to form fibres, which make up fascicles (Karousou et al. 2008). More specifically, triple-helix polypeptide chains of tropocollagen form microfibrils and microfibrils combine to form fibrils (Benjamin et al. 2008). Fibrils then assemble into fibres (primary bundles), fascicles (secondary bundles), tertiary bundles and ultimately the tendon itself (Sharma and Maffulli 2005). Collagens such as type III and V form heterotypic fibres with type I collagen (Mokone et al. 2006).

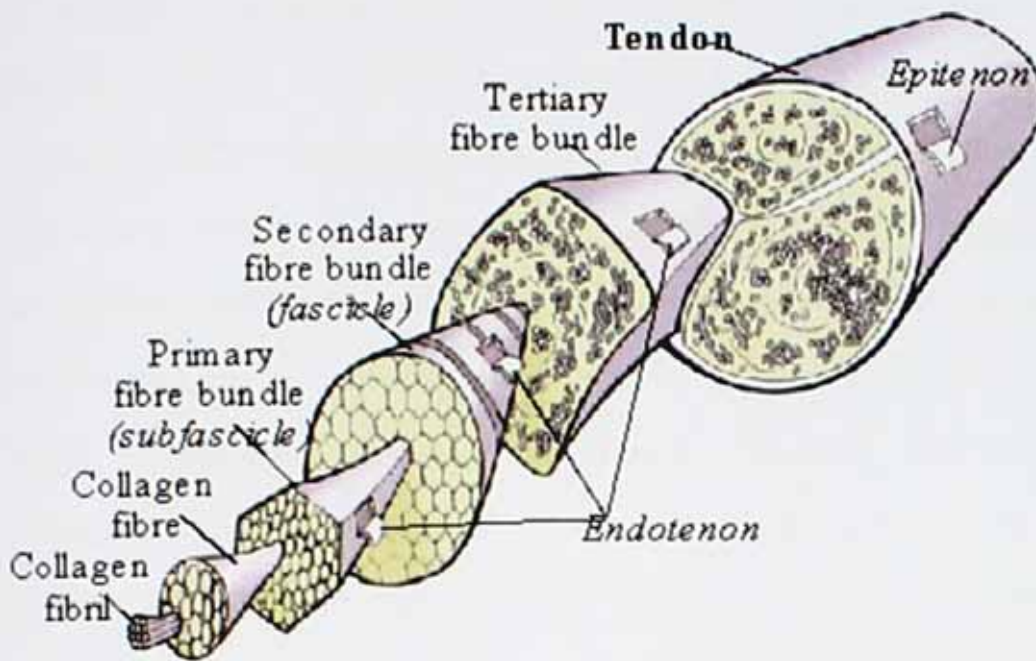
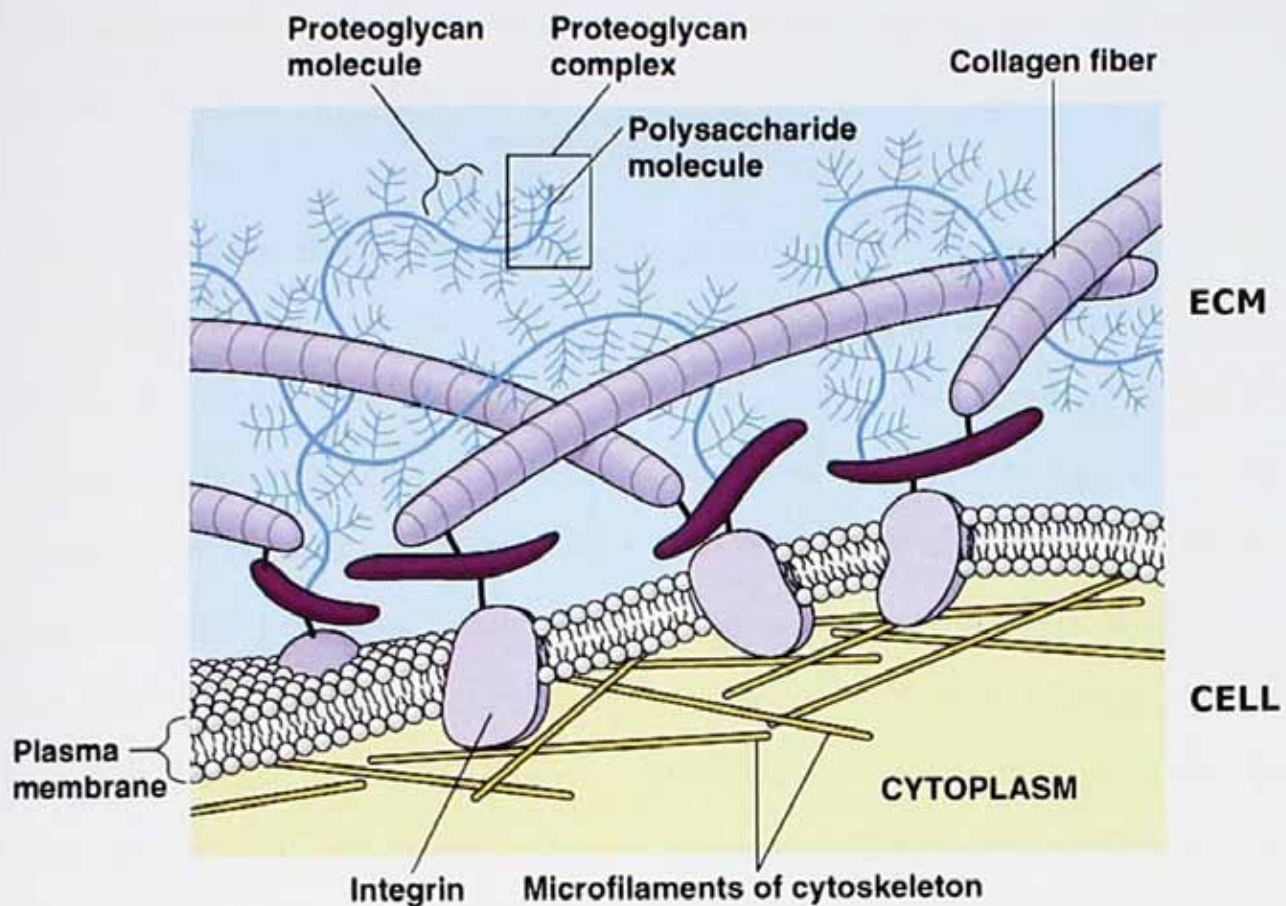


Figure 1.1 The structural organisation of collagen in tendon. Image adapted from Khan et al. (1999).

A thin connective tissue layer called the endotenon is present between the fascicles and fibre bundles, functioning to facilitate sliding between the fascicles (Benjamin et al. 2008) (**Figure 1.1**). This role is essential as it enables tendons to change shape as their associated muscles contract, as well as transmit tension with the changing angles of a joint when it moves. The epitenon, of which the endotenon is an extension of, is a further loose connective tissue sheath that covers the entire tendon and provides vascular, lymphatic and nerve supply (Khan et al. 1999). In comparison to skeletal muscle, oxygen supply to tendons is said to be over 7 times lower (Sharma and Maffulli 2005) due to poorer vascularity (Young et al. 2005). The lower metabolic rate is beneficial for tendons to maintain tension and support loads for long periods, however, it can result in a slower healing in response to injury (Young et al. 2005; Sharma and Maffulli 2005).

In addition to tenocytes and collagen, the ECM consists of elastin, water and a complex network of protein molecules (**Figure 1.2**) (Karousou et al. 2008; Reinking 2012). These are often described as the ground substance and their presence is minimal in normal tendon (Cook et al. 2002). The ground substance of the ECM contains proteoglycans, glycoproteins, glycosaminoglycans and other specialised proteins including enzymes (Sharma and Maffulli 2005; Xu and Murrell 2008). Proteoglycans are glycosylated proteins (protein-polysaccharide complexes) that have both structural and signalling roles within the tendon (**Figure 1.2**) (Xu and Murrell 2008). They are hydrophilic proteins that aid cell migration and the diffusion of water-soluble molecules (Sharma and Maffulli 2005). Their hydrophilic nature enables them to trap water, affecting the viscoelastic properties of tendon and assisting the tendon to resist compressive forces (Xu and Murrell 2008). Indeed, proteoglycans have a number of vital functions within the cell that include water retention, ion transport, nutrient diffusion and the mediation of cell-matrix interactions (Xu and Murrell 2008). They consist of a protein core with attached glycosaminoglycan (GAG) sidechains and tendons contain a range of different proteoglycans (Birch 2007). Proteoglycans found in tendons include decorin, biglycan, fibromodulin, versican, lumican and the large aggregating proteoglycan aggrecan (Birch 2007; Rees et al. 2000). Proteoglycans, such as aggrecan, can form complexes with other proteoglycans and collagen (Xu and Murrell 2008). Proteoglycans have been shown to interact with collagen fibres and regulate collagen fibril diameter through their roles in collagen fibril formation and maturation (Birch 2007; Xu and Murrell 2008).



Copyright © Pearson Education, Inc., publishing as Benjamin Cummings

Figure 1.2 Schematic overview of the tendon ECM showing the network of connected molecules, including collagen and proteoglycans, surrounding and supporting the cell (tenocyte).

Tenascin-C is a large hexametric glycoprotein component of the ECM and part of the tenascin family (**Figure 1.3**) (Midwood et al. 2011). It has been implicated in a number of cellular processes including cell migration, proliferation and apoptosis (Minear et al. 2011). Tenascin-C expression is distinct and tightly controlled, with little or no expression in healthy adult tissues (Midwood et al. 2011). Expression is up-regulated in tissues undergoing mechanical loading and during wound healing, while is down-regulated after tissue repair (Sharma and Maffulli 2005; Midwood et al. 2011). In addition, tenascin-C also has elastic properties and roles

in collagen fibre alignment and orientation, highlighting its structural role in tendon (Sharma and Maffulli 2005).

Matrix metalloproteinases (MMPs) are a large family of over 20 enzymes involved in degradation and remodelling of the ECM (Karousou et al. 2008; Posthumus et al. 2011). They are zinc-containing enzymes, known to degrade a range of collagenous and non-collagenous ECM components as well as being involved in the activation of other MMPs (Posthumus et al. 2011; Raleigh et al. 2009). MMPs, along with their inhibitors, tissue inhibitor of metalloproteinases (TIMPs), play an important role in tendon homeostasis (Xu and Murrell 2008). Matrix metalloproteinase 3 (MMP3) (**Figure 1.3**) is one member of this family MMP3, which can degrade collagenous and non-collagenous ECM proteins (Somerville et al. 2003; Visse and Nagase 2003).

Growth differentiation factors (GDFs) are part of the transforming growth factor beta (TGF-β) superfamily with important functions in tissue growth and homeostasis, including tendon tissue (Posthumus et al. 2010). GDF5 (**Figure 1.3**) is involved in the growth, repair and maintenance of bone and cartilage, as well as musculoskeletal soft tissues including tendon (Foster et al. 2012). In relation to cartilage, GDF-5 functions in chondrogenesis (cartilage development), chondrocyte proliferation and the maintenance and repair of synovial joints (Reynard et al. 2011).

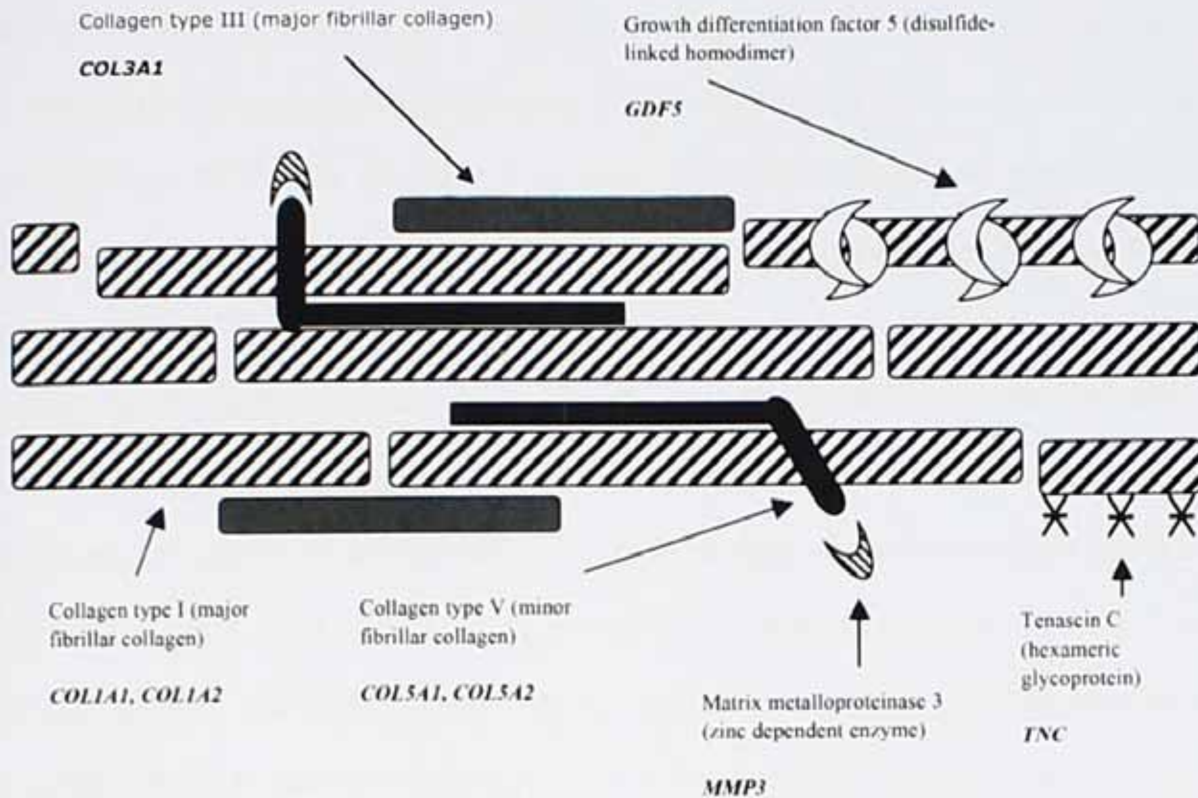


Figure 1.3 Diagram showing the general structural ECM components of tendon, consisting of major and minor fibrillar collagens and associated proteins including tenascin C and matrix metalloproteinase 3 (MMP3). Adapted from Collins and Raleigh (2009) and Foster et al. (2012).

1.2.1 Achilles Tendon

The human Achilles tendon permits plantarflexion and dorsiflexion of the foot at the ankle joint (**Figure 1.4**) (Benjamin et al. 2008). The myotendinous junction of the Achilles tendon is where it originates from the soleus and gastrocnemius muscles, while the osteotendinous junction is where the Achilles tendon inserts into the calcaneus (heel) (Cook et al. 2002; Lesic and Bumbasirevic 2004). In addition to the epitenon described previously, the Achilles tendon is further encased in a single layer of connective tissue cells called the paratenon. Collectively the epitenon and paratenon are often termed the peritendon (Schepsis et al. 2002; Cook et al. 2002). The paratenon, consisting mainly of type I and type III collagen fibrils with some elastic fibres, is a highly vascularised tissue responsible for the main blood supply to the tendon (Khan et al. 1999; Lesic and Bumbasirevic 2004). During normal physiological movement, the Achilles tendon is subjected to high strains and can be considered as an energy storing tendon (Birch 2007).

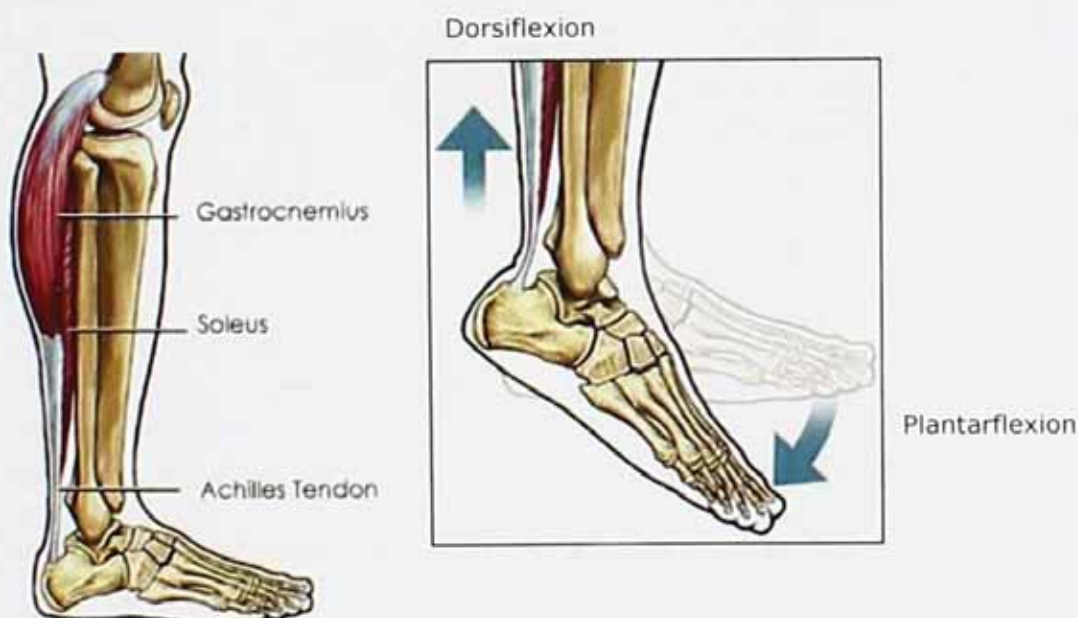


Figure 1.4 Image of Achilles tendon and associated gastrocnemius and soleus muscle (http://www.valleo.com/achilles_tendonitis).

1.2.2 Patellar Tendon

The patellar tendon, in conjunction with the quadriceps tendon, permits extension of the knee and connects the patella (knee cap) to the tibia (**Figure 1.5**) (Tandeter and Shvartzman 1999; Khan et al. 1998). By definition it is a ligament as it connects bone to bone, however, it is typically referred to as the patellar tendon. Supporting knee ligaments including the anterior and posterior cruciate ligaments (ACL and PCL respectively) and the medial and lateral collateral ligaments (MCL and LCL respectively) help to provide stability at the joint (**Figure 1.5**) (Tandeter and Shvartzman 1999).

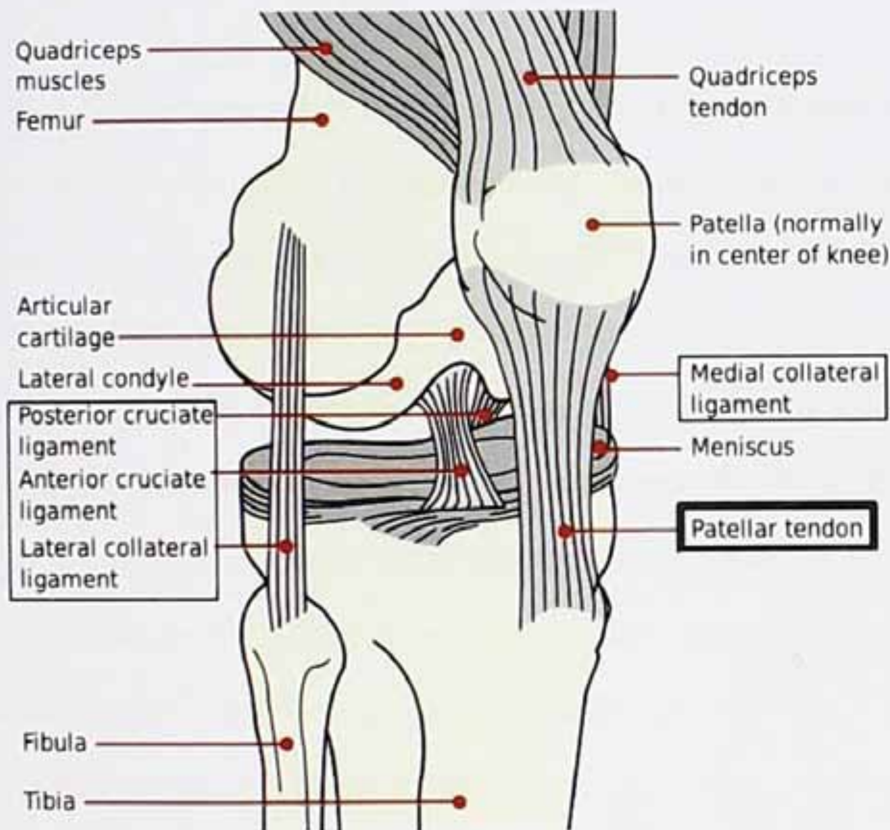


Figure 1.5 Image of the right knee showing the patellar tendon and associated bones (femur, patella, tibia and fibula) and knee ligaments. Ligaments shown are outlined in boxes and include: posterior cruciate ligament (PCL); anterior cruciate ligament (ACL); lateral collateral ligament (LCL) and medial collateral ligament (MCL) (http://orthotape.com/Patella_Knee_support_band_strap.asp).

1.3 TENDON PATHOLOGY

The understanding of the molecular basis to tendinopathy has changed over time (Rees et al. 2013). Before the 1990s, tendinitis was the term used to describe painful tendons, implying discomfort of an inflammatory nature (Christian et al. 2014). This tendinitis model was widely accepted and it seemed that inflammation was the process responsible for tendon pain. Further investigation typically showed a lack of inflammatory cells in tendon, despite other molecular disruption (Cook and Purdam 2009; Rees et al. 2013). However, while inflammation may not be the dominant pathology, it is believed that the inflammatory response may play a role at different stages (Rees et al. 2013).

The term tendinopathy is now widely accepted and used to describe a process of tendon pain classified as degeneration, caused from an abnormal or failed healing response following repetitive mechanical loading (Rees et al. 2013; Christian et al. 2014; Samiric et al. 2009). Apart from the physical symptoms experienced by an individual, at a molecular level the tendon has undergone a number of changes that are atypical of normal tendon turnover (Khan et al. 1998; Parkinson et al. 2010). Microscopically, visible changes can be seen in the appearance of the ECM (Parkinson et al. 2010). Whereas normal tendon consists predominantly of an ECM abundant in highly organised Type I collagen, tendinopathic tendons often displays a disorganisation of the ECM and collagen fibres (Samiric et al. 2009; Parkinson et al. 2010). In the long term, tendinopathy that is not addressed can lead to permanent tissue damage from a viscous cycle of matrix disorganisation, disrepair and subsequent degeneration of tendon tissue (Parkinson et al. 2010).

Furthermore, research now leans towards a multistage model to describe tendinopathy. (Cook and Purdam 2009). The three stages (reactive tendinopathy, tendon disrepair and degenerative tendinopathy) of tendinopathy described by Cook and Purdam (2009) highlight the complexity of tendinopathy and the clinical and molecular changes that can differentiate each stage. In brief, reactive tendinopathy (stage 1) is typified by a non-inflammatory proliferative response that results in thickening of part of the tendon (Cook and Purdam 2009). This is a short-term adaptation in response to overload, in an attempt to reduce stress and/or to resist compression (Cook and Purdam 2009). Tendon disrepair (stage 2) is characterised by a further increase in the number of cells and subsequent protein synthesis of collagens and proteoglycans, in an effort to attempt tendon healing (Cook and Purdam 2009). This disproportionate increase in cellular and protein components leads to disorganisation of the ECM (Cook and Purdam 2009). Degenerative tendinopathy (stage 3) describes the longer-term effects of tendinopathy, resulting in widespread disorganisation of the ECM and areas of cell death from apoptosis (Cook and Purdam 2009). Once a tendon has reached this stage, some of the pathological changes are likely to be permanent (Cook and Purdam 2009).

Throughout this thesis the term tendinopathy will be used to describe the above. When discussing tendinopathy and tendon rupture collectively, the term tendon pathology will be used.

1.3.1 Epidemiology

The epidemiology of tendinopathy varies amongst active and sedentary individuals, as well as in different sports and with regard to different tendons (Sobhani et al. 2013; de Jonge et al. 2011; Christian et al. 2014; Young et al. 2005). In general, tendon related injuries have been said to account for 30-50% of all sporting injuries, with 6-18% of these injuries involving the Achilles tendon (September et al. 2009; Järvinen et al. 2005). Despite being the strongest tendon in the body, the Achilles is prone to damage and rupture (Lesic and Bumbasirevic 2004) and is the most commonly ruptured tendon in humans (Young et al. 2005). There is a 52% lifetime risk of Achilles tendinopathy in elite long-distance runners and an annual incidence rate as high as 9% has been reported in such individuals (de Jonge et al. 2011). Furthermore, some of the highest incidence rates (per 1000 athletes per season) of ankle and foot injuries have been observed in runners, ballet dancers and gymnasts (Sobhani et al. 2013). This is not surprising, since the increased stress on tendons during physical exercise means that tendon injuries are more common in both professional and recreational athletes, typically as a result of repetitive mechanical loading (Maffulli et al. 2003; Xu & Murrell, 2008; Collins & Raleigh, 2009). However, the occurrence of tendon injuries in the general population is also evident and the incidence in both athletes and sedentary individuals is said to be increasing (Young et al. 2005; de Jonge et al. 2011). Indeed, incidence and prevalence rates of Achilles tendinopathy of 1.85 and 2.01 per 1000 respectively have been described in the general population (de Jonge et al. 2011).

Although much of the research in this thesis is concerned with Achilles tendinopathy and rupture, Chapter 6 describes the studies into epigenetics, which

were investigated in the patellar tendon. Patellar tendinopathy is prevalent among athletic populations, particularly in athletes competing in sports involving jumping (Samiric et al. 2009; Christian et al. 2014). Over 30% of sport related injuries are said to be due to patellar tendinopathy (Samiric et al. 2009), with prevalence rates of 44.6% and 31.9% reported in volleyball and basketball players respectively (Christian et al. 2014). However, while the term "jumper's knee" can be justified by the prevalence in sports such as basketball, where repetitive loading of the patellar tendon is apparent, patellar tendinopathy is also common in sedentary individuals (Samiric et al. 2009).

1.3.2 Achilles Tendon Pathology

Achilles tendon pathology (ATP) defines a range of Achilles tendon injuries, broadly classified as Achilles tendinopathy and Achilles tendon rupture. However, further classification identifies complete or partial ruptures, as well as the anatomical location and nature of tendinopathy (Schepsis et al. 2002; Raleigh & Collins, 2012).

1.3.2.1 Non-Insertional Achilles Tendinopathy

Non-insertional Achilles tendinopathy is also termed mid-portion Achilles tendinopathy and describes pain and swelling in the tendon 2-7 cm from insertion into the calcaneus (van Dijk et al. 2011) (**Figure 1.6B**). Tendinosis may also be evident, describing the presence of tendon degeneration with a lack of clinical or histological inflammation, which may be asymptomatic (Lesic and Bumbasirevic 2004; van Dijk et al. 2011).

A **B**

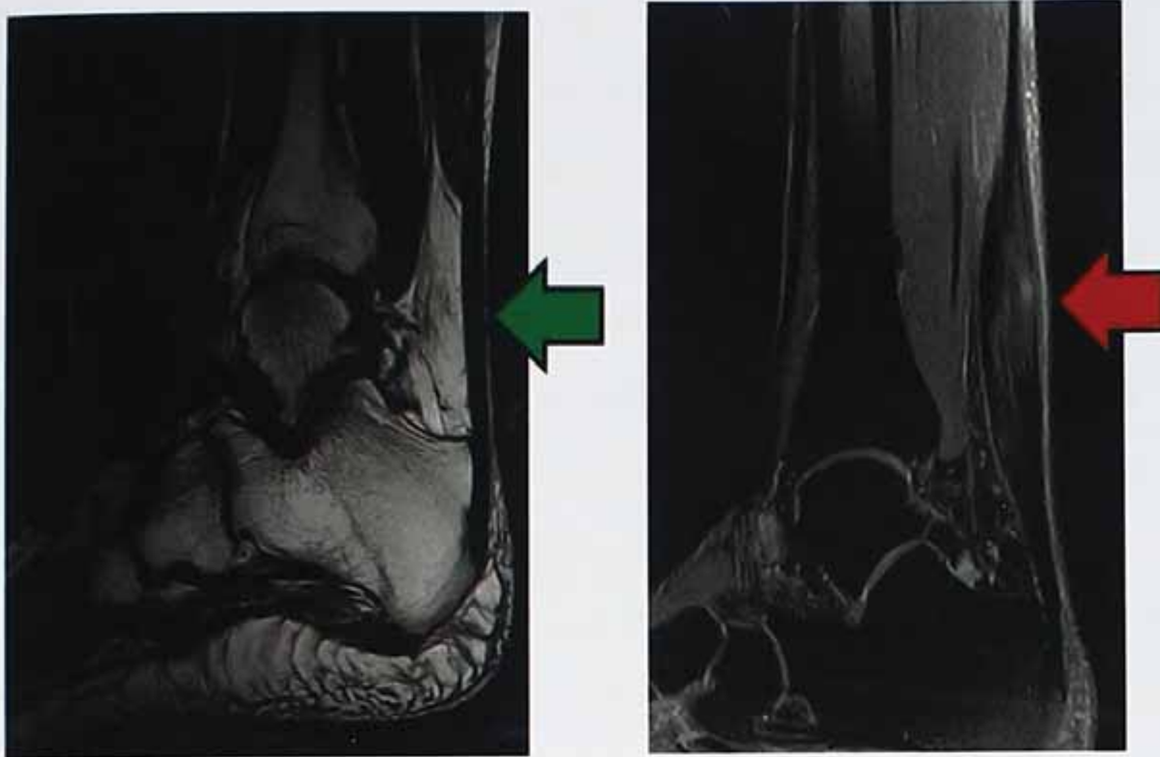


Figure 1.6 MRIs of a normal and tendinopathic Achilles tendon.

A. Magnetic resonance image (MRI) of a normal Achilles tendon. **B**. MRI of a pathological degenerate tendon with non-insertional Achilles tendinopathy, where a change in signal intensity and swelling/thickening can be seen in the mid-portion of the tendon. Images from The Country Clinic, Northampton with permission.

1.3.2.2 Insertional Achilles Tendinopathy

Insertional tendinopathy is described as heel pain and tenderness localised at the Achilles tendon insertion, where the presence of bone spurs and calcification is common (Schepsis et al. 2002; Lesic and Bumbasirevic 2004). Bone spurs may be palpable and these can cause small tears in the tendon at the tendon-bone junction (van Dijk et al. 2011). In addition, insertional Achilles tendinopathy is often associated with retrocalcanael bursitis and/or Haglund's deformity, although the presence of these pathologies alone is not conclusive of a diagnosis of tendinopathy. (Schepsis et al. 2002; Lesic and Bumbasirevic 2004).

1.3.2.3 Achilles Tendon Rupture

The tendon is designed to withstand a certain level of physiological stress exerted on it through normal movement, however, stretch exerted that is greater than 5% elongation causes damage to the collagen fibres themselves and can lead to rupture (Khan et al. 1998). Although Achilles tendon rupture can develop as an end result of Achilles tendinopathy, different molecular changes have been observed that distinguish it from tendinopathy (Jones et al. 2006; Corps et al. 2008; Parkinson et al. 2010). Indeed, apparent sporadic rupture may occur without any symptoms of pain, yet at a structural and molecular level, altered expression levels of certain matrix metalloproteinases has been detected (Jones et al. 2006; Parkinson et al. 2010). For example, Jones et al., (2009) showed that ADAMTS4 (a disintegrin and metalloproteinase with thrombospondin motifs 4) was up-regulated in ruptured Achilles tendons, compared with painful and normal Achilles tendons.

1.3.3 Patellar Tendinopathy

Patellar tendinopathy is a chronic degenerative condition, characterised by pain and tenderness in the knee, that can prevent or impede an individual from participating in physical activity and exercise (Parkinson et al. 2010; Samiric et al. 2009). Typically the pain is localised to the proximal insertion of the patellar tendon to the patella, which intensifies with prolonged activity and flexion of the knee (Christian et al. 2014). It is commonly first recognised when an individual is asked to undergo a declining squat test (Christian et al. 2014). Ultrasonography and MRI can then be used to clinically diagnose patellar tendinopathy, often revealing an abnormal tendon appearance and a wider or thicker portion of tendon at the site of tenderness in the knee (Christian et al. 2014) (**Figure 1.7B**). For professional athletes, as well as recreational athletes, this chronic pathology can mean having to limit training or even end their careers or sporting hobbies (Christian et al. 2014).

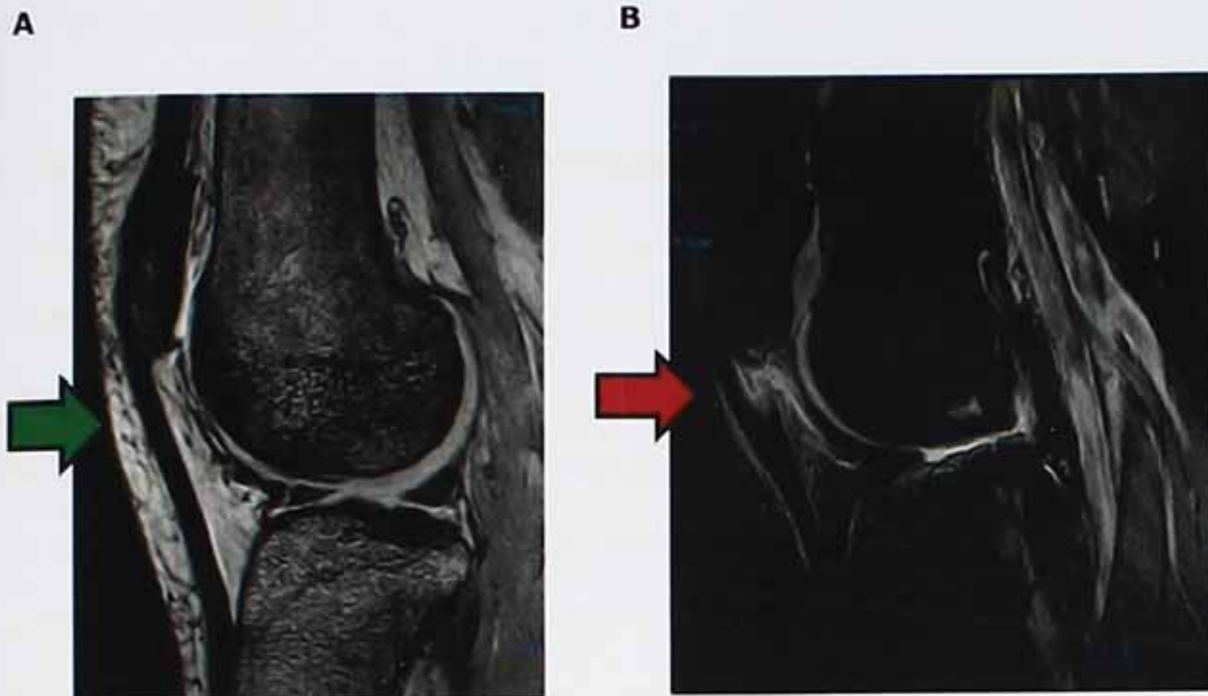


Figure 1.7 MRIs of a normal and abnormal patellar tendon.

A. MRI of a normal patellar tendon. **B**. MRI of a pathological degenerate proximal patellar tendon, showing a change in signal intensity and increased tendon thickness characteristic of patellar tendinopathy. Images from The Country Clinic, Northampton with permission.

1.3.4 Patellar Tendon Rupture

As with Achilles tendon rupture, it evident that while changes in ECM composition and disruption of collagen organisation is concordant in both tendinopathy and rupture, these conditions display differences with each other when compared to normal tendon (Jones et al. 2006; Cook and Purdam 2009). If the tendon is elongated beyond its normal physiological range, then trauma can occur in the form of tears in the collagen fibres (Khan et al. 1998). In relation to patellar tendon rupture, the risk of rupture is high in sports involving jumping, such as basketball (Khan et al. 1998; Christian et al. 2014). This is due to the high levels of stress

exerted on the patellar tendon that can be 6-8 times an individual's body weight (Khan et al. 1998).

1.3.5 Current Management and Treatment

Despite overloading being a key contributing factor of tendinopathy, mechanical loading is in fact a vital treatment strategy for managing the condition (Cook and Purdam 2014). A complete loss of loading and isolation of the tendon without mechanical stimuli can have a detrimental effect on tendon strength and will result in a catabolic effect, as tendon homeostasis is maintained through physiological loading within a healthy range (Killian et al. 2012; Cook and Purdam 2014; Reinking 2012). Mechanical loading maintains tendon homeostasis through ECM remodelling, as many of the ECM components are up-regulated in response to such stimuli (Killian et al. 2012). However, the treatment of tendinopathy with a measured mechanical loading regime can become challenging in athletes competing competitively and especially during the peak season of their chosen sport (Cook and Purdam 2014). Without any decrease in current activity levels, it can be hard for the tendon ECM to be given chance to remodel and repair, therefore appropriate loads are important as to not intensify or trigger reactive tendinopathy (Cook and Purdam 2014).

The modification of training programmes and physical activity (reduction in intensity, frequency and duration) may be necessary initially and the complete cessation of activity should be assessed on an individual basis (Schepsis et al. 2002; Reinking 2012). If required, the correction of any foot alignment problems is important in relation to Achilles tendinopathy, as excessive pronation can increase the stress exerted on tendons like the Achilles (Schepsis et al. 2002).

Eccentric strength training (muscle lengthening under load) has been shown to be beneficial in the treatment of tendinopathy, where eccentric declining squat training is considered an important part of the treatment of patellar tendinopathy (Christian et al. 2014). Indeed, eccentric strength training has been shown to reduce pain in both Achilles and patellar tendinopathy, through the reduction of blood flow to neovessels and the subsequent decline in pain from nerve endings (Christian et al. 2014).

Injection therapies, such as high volume ultrasound-guided injections, have been effective in treating the pain associated with tendinopathy by impeding neovascularisation and may be more effective than eccentric training (Christian et al. 2014). Meanwhile, the injection of steroids such as corticosteroids may in fact cause tendon ruptures (Schepsis et al. 2002). Finally, although considered a last resort to the less invasive treatment measures described above, surgical intervention may be necessary in certain chronic cases of tendinopathy, where conservative therapies have not been effective (Christian et al. 2014; Kader et al. 2002).

1.4 RISK FACTORS ASSOCIATED WITH TENDON PATHOLOGY

ATP has a complex, multifactorial aetiology, which is yet to be fully defined. A range of both intrinsic and extrinsic factors have been identified that increase an individual's risk of developing ATP (Meeuwisse, 1994; Raleigh & Collins, 2012) (**Figure 1.8**). It is the presence of numerous factors together that increase the risk. Following predisposition and susceptibility, an individual may experience an inciting event during training or physical activity, resulting in tendon injury (Raleigh & Collins, 2012). The severity and impact of these injuries on athletes highlights the importance of research in this area.

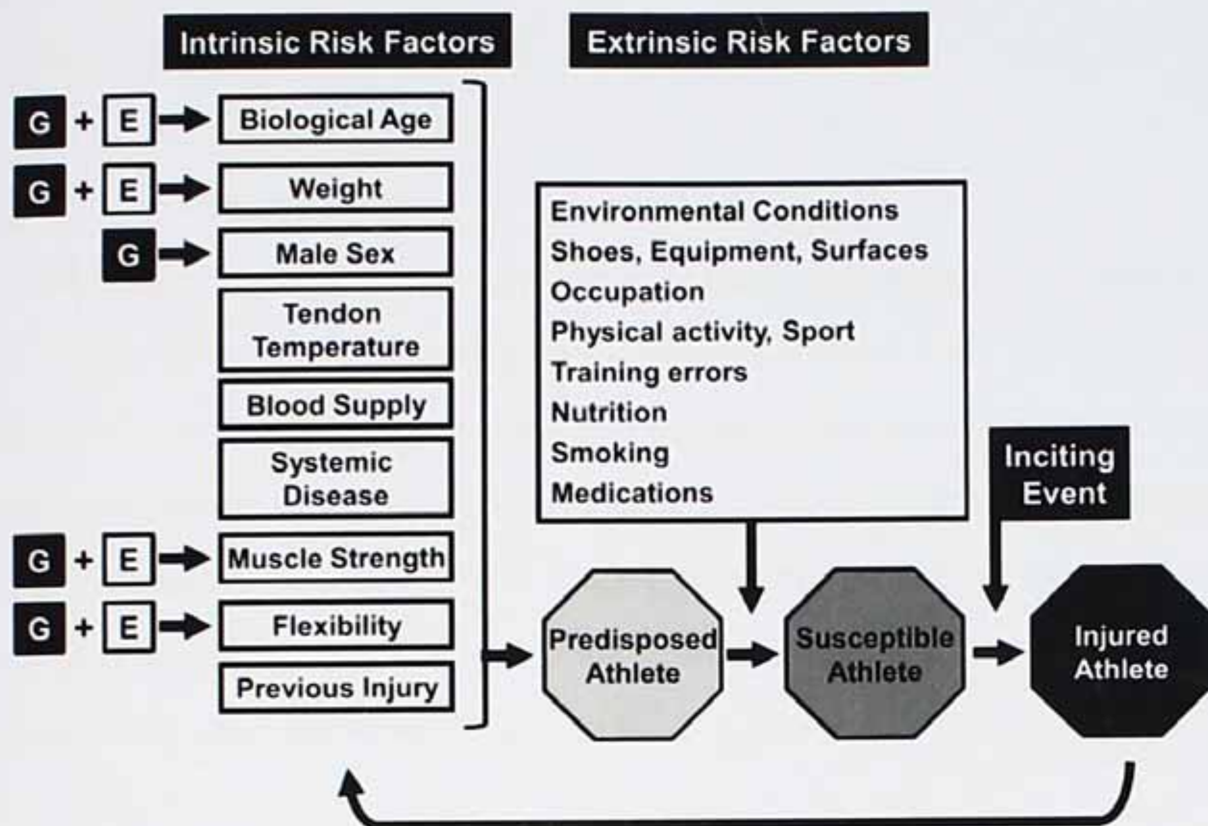


Figure 1.8 Risk Factors for Tendon Injury.

Graphic diagram illustrating the role of intrinsic and extrinsic risk factors in the aetiology of tendon pathology (Raleigh and Collins 2012). Several of the intrinsic risk factors shown are considered multifactorial phenotypes, determined by genetic (G) and environmental (E) factors. The arrow leading from the injured athlete back to intrinsic risk factors illustrates that individuals are predisposed to additional injuries following recovery (Raleigh & Collins, 2012).

1.4.1 Intrinsic Risk Factors

Intrinsic risk factors such as flexibility, biological age and previous injury can predispose individuals to ATP (Raleigh & Collins, 2012) (**Figure 1.8**). Non-modifiable intrinsic factors include increasing age, gender and an individual's genetics. Modifiable intrinsic risk factors include muscle inflexibility and obesity (Reinking 2012). Many intrinsic factors have both an environmental and genetic

component, showing that these phenotypes are themselves multifactorial (Raleigh & Collins, 2012).

1.4.1.1 Age and Gender

Research has shown that older age and male gender appear to be predisposing risk factors for tendon injuries including Achilles tendinopathy and Achilles tendon rupture (Kraemer et al. 2012). Achilles tendon rupture has been shown to be more common in men than women (Vosseller et al. 2013). Research suggests that the lower incidence of Achilles tendon injuries in women is due to the decreased stiffness and greater elongation noted in the female Achilles during plantarflexion and following exercise (Joseph et al. 2014).

1.4.1.2 Anatomical Factors

Anatomic factors affecting the lower limb have also been shown to increase the risk of tendon pathology (Järvinen et al. 2005; Schepsis et al. 2002). These include hyper and hypo-pronation, as well as decreased joint flexibility and muscle weakness (Järvinen et al. 2005).

1.4.1.3 Systemic Diseases

There is some evidence to show that certain systemic diseases, such as rheumatoid arthritis, can manifest with ATP and may affect the risk of tendon injury and/or rupture (Järvinen et al. 2005; Suzuki and Okamoto 2014; Kraemer et al. 2012). Other systemic diseases including gout and ankylosing spondylitis (both types of arthritis) have also been suggested to affect the risk of tendon rupture (Järvinen et al. 2005; Gerster et al. 1996). Despite these reports, the effect of these concurrent systemic diseases on the risk of Achilles tendon injuries

and tendon ruptures in general is considered negligible (Järvinen et al. 2005). Only ~2% of Achilles tendon injuries are thought to be attributed to the presence of an underlying systemic disease, while the major influencing factor is overuse (Järvinen et al. 2005). These diseases were taken into account when recruiting participants and exclusion criteria were implicated where appropriate.

1.4.2 Extrinsic Risk Factors

Further exposure to extrinsic factors (external modifiable factors that impact on an athlete) including training volume, exercise intensity, training errors, environmental conditions, ground surface and medication, can result in predisposed individuals becoming susceptible to ATP (**Figure 1.8**) (Meeuwisse 1994; Raleigh and Collins 2012). Ultimately, the occurrence of an inciting event leads to tendon injury (**Figure 1.8**).

1.4.2.1 Activity Levels

As previously described, tendon injuries are prevalent amongst the athletic population due to the high levels of physical activity and common overuse due to repetitive mechanical loading of the joints (Collins and Raleigh 2009) (**Figure 1.8**). Research has shown that 75% of ruptures to the Achilles tendon are related to sports (Järvinen et al. 2005).

1.4.2.2 Environment, Training Errors, Footwear

Other extrinsic risk factors implicated in the risk of ATP include environment conditions such as cold weather, a hard training surface and the footwear worn (Järvinen et al. 2005; Raleigh and Collins 2012).

1.4.2.3 Steroids and Fluoroquinolones

Use of the antibiotics fluoroquinolones has been shown to associate with tendinopathy, as research suggests that approximately 2-6% of Achilles tendon ruptures in individuals over 60 years old may be credited to quinolone use (Sendzik et al. 2005). Fluoroquinolones appear to cause changes in the ECM and disrupt tendon homeostasis (Khaliq and Zhanel 2003; Sendzik et al. 2005). Changes in collagen fibril size and organisation has been observed in cultured human tendon cells, as well as increases in matrix metalloproteinases and the apoptosis related caspase 3 (Sendzik et al. 2005). Furthermore, neovascularisation, degenerative tendon lesions and necrosis and have also been reported in human patients, as a result of certain fluoroquinolones use (Khaliq and Zhanel 2003). Likewise, the use of steroids, both injectable corticosteroid injections as well oral anabolic steroids taken by competitive athletes, has been shown to increase the risk of tendon rupture (Schepsis et al. 2002; Järvinen et al. 2005). In relation to anabolic steroid use, tendons have been shown to be stiffer and absorb lower loads before rupturing (Schepsis et al. 2002).

1.4.3 Genetic Risk Factors

In 1989, Jozsa and colleagues were the first to show an association between the ABO blood groups and tendon rupture. Specifically, a significant association between blood group O and tendon rupture was demonstrated in a Hungarian population. Over 50% of tendon rupture patients were blood group O and this blood group was overrepresented in comparison with the control population (Jozsa et al. 1989). Later in 1992, the increased frequency of tendon injury in people with blood group O was confirmed (Kujala et al. 1992). Nevertheless, more recent studies have failed to find an association with blood group (Reinking 2012). In fact, current research points towards genes proximal to the *ABO* gene locus on chromosome 9 as causal factors for ATP (Mokone et al. 2005; Mokone et al. 2006; Reinking 2012). The *ABO* gene locus is located on chromosome 9q34 and the initial research on blood groups and tendon injury was what led to interest in other genes located on this chromosome.

Furthermore, family history has been shown to be an important factor in the development of ATP (Kraemer et al. 2012). Indeed, a positive family history of Achilles tendinopathy was shown to result in a fivefold increased risk of developing the condition (Kraemer et al. 2012). Specifically, Achilles tendinopathy patients had a significantly increased number of relatives with Achilles tendon disorders (Kraemer et al. 2012). This evidence supports the findings from the numerous genetic association studies to date investigating genetic variants and the risk of ATP. These are summarised in Table 1.1 and described below. In addition, the familial link with ATP further reinforces the importance of genetics in the aetiology of such conditions and the need for further research in this area.

1.4.3.1 Repeat Polymorphisms and Single Nucleotide Polymorphisms

While more than 99% of the DNA sequence in the human genome is identical between individuals, a small portion of the genome exhibits genetic variation (Shastry 2002). Single nucleotide polymorphisms (SNPs) are a type of small-scale genetic variation, where a single nucleotide base is altered in the DNA sequence of normal individuals (Shastry 2002). Specifically, they occur at a frequency of approximately 1 in 1000 base pairs and the rare allele has a frequency higher than 1% (Brookes 1999). Repeat polymorphisms are another type of small scale genetic variation, also termed microsatellites (September et al. 2007). With this type of polymorphism, repeated base pair sequences of DNA are seen in varying lengths in individuals (Mokone et al. 2005). For example, a dinucleotide repeat polymorphism is a repeated tandem base pair sequence that is present in different lengths in genomes (Mokone et al. 2005).

One of the first genes to be investigated for its role in ATP was *TNC*. Mokone et al. (2005) found that a guanine-thymine (GT) repeat polymorphism in the *TNC* gene was associated with ATP. The *TNC* gene encodes the ECM protein tenascin-C (**Figure 1.3**), which is in close proximity to the *ABO* gene locus on chromosome 9q32-q34 and is expressed in a range of tissues, including tendons (Mokone et al., 2005). Tenascin-C is believed to play a role in the regulation of cell-matrix interactions around tendons, binding to other ECM proteins and cell receptors (Mokone et al., 2005). In healthy adult tissues, tenascin-C expression is shown to be low, however expression is up-regulated in response to injury (Midwood et al. 2011). Furthermore, mechanical loading on tendons is shown to regulate *TNC* gene expression in a dose-dependent manner (Järvinen et al. 1999). Although the

exact mechanisms controlling *TNC* expression are not fully understood, certain cytokines and growth factors including transforming growth factor-β (TGF-β) have been shown to induce synthesis of tenascin-C (Järvinen et al. 1999).

Following this genetic association, in 2006 Mokone et al. showed that the rs12722 SNP within the 3' untranslated region (UTR) of *COL5A1* gene, also located on chromosome 9, was associated with ATP in a South African Caucasian population. The alpha 1 type V collagen protein, encoded by the *COL5A1* gene, is a minor structural protein constituent of tendon (Mokone et al. 2006) and is thought to be involved in the regulation of type I collagen fibril diameter (September et al. 2009). The A2 allele of the rs12722 variant was significantly overrepresented in the CON group, showing a protective benefit of this allele and a decreased risk of Achilles tendinopathy in individuals possessing this variant (Mokone et al. 2006). In 2009, September et al. (2009) also investigated the rs12722 variant in Australian and South African Caucasian populations. They found that the CC genotype was underrepresented in Achilles tendinopathy compared to controls and these individuals were at a decreased risk of developing Achilles tendinopathy (September et al. 2009). In addition, in the South African Achilles tendinopathy group, the TC inferred haplotype (rs12722-rs3196378) was found to be overrepresented (September et al. 2009).

The next gene to be investigated in relation to ATP was matrix metalloproteinase 3 (*MMP3*). Research in 2009 highlighted that several SNPs within the candidate gene *MMP3* were associated with the risk of Achilles tendinopathy (Raleigh et al. 2009). Specifically it was found that the GG genotype of rs679620, the CC genotype of rs591058 and the AA genotype of rs650108 that were found to be

associated with Achilles tendinopathy. The rs679620 variant is a non-synonymous SNP that lies within exon 2 of the *MMP3* gene, while the rs650108 SNP lies within intron 8 of the gene (Raleigh et al. 2009). In addition, the haplotype analysis of the rs679620, rs591058, and rs650108 variants in this study showed that the ATG haplotype was underrepresented in Achilles tendinopathy, when compared to controls (Raleigh et al. 2009).

In 2010, then next genetic association with ATP was discovered. Although the specific role of GDF-5 in tendon is unknown, it has been found to associate with the risk of ATP (Posthumus et al. 2010). The rs143383 variant (T/C) is a functional promoter SNP that lies within the 5′ UTR of the *GDF5* gene and has previously been associated with other musculoskeletal conditions including osteoarthritis (Posthumus et al. 2010). Individuals from an Australian cohort with the TT genotype of the functional *GDF5* rs143383 variant were significantly over-represented in the ATP group compared to controls (P=0.011) and two times as likely to develop ATP. Furthermore, when the Australian and South African cohorts were combined, significance was also observed (P=0.004) (Posthumus et al. 2010).

Two years later in 2012, a new genetic association study investigating genetic variants in genes involved in the apoptosis pathway was published, showing the association of novel SNPs within the candidate gene *CASP8* and the risk of Achilles tendinopathy (Nell et al. 2012). The rs1045485 and rs3834129 functional SNPs (intronic variants within the *CASP8* gene) were independently associated with Achilles tendinopathy, as well as associating with risk of tendinopathy when combined as a haplotype (Nell et al. 2012). A significant difference in genotype

(P=0.0294) was detected between the CON and ATP groups for the *CASP8* rs1045485 variant (Nell et al. 2012). While for the *CASP8* rs3834129 variant, significant differences in genotype and allele frequencies were reported between the two Caucasian cohorts investigated (South African and Australian), as well as between the CON and ATP groups (Nell et al. 2012).

Saunders et al. (2013) were next to publish data showing that SNPs within the *COL27A1* and *TNC* gene were associated with Achilles tendinopathy in Caucasian populations (Saunders et al. 2013). They demonstrated the independent allelic associations of the rs2104772 (P = 0.017) and rs1330363 (P = 0.020) SNPs within the *TNC* gene and the risk of Achilles tendinopathy (Saunders et al. 2013). Furthermore, the GCA haplotype (*COL27A1* rs946053 - *TNC* rs13321 - *TNC* rs2104772) was also found to associate with an increased risk of Achilles tendinopathy (P = 0.019), where an increased frequency of this haplotype was observed in Achilles tendinopathy compared to controls (Saunders et al. 2013). This was the second study to investigate genetic variants within the *TNC* gene and the risk of ATP, following on from the findings from Mokone et al. (2005).

The most recently published genetic association study in relation to ATP found that a variant within the *TIMP2* gene was associated with the risk of ATP (El Khoury et al. 2013). El Khoury et al. (2013) showed that the genotype frequency of the rs4789932 SNP was significantly different in a combined Caucasian population (South African and Australian) between the ATP and CON groups. The CC genotype was significantly overrepresented in the combined CON group, while the CT genotype was significantly overrepresented in the combined ATP group (El Khoury et al. 2013).

To date, no larger scale variants, such as copy number variation (CNV) have been associated with ATP. This type of genetic variation is described in section 1.5.

Table 1.1 Genetic Variants Associated with human Achilles Tendon Pathology

Gene	Variant(s)	Population	Findings	Reference
Tenascin C (*TNC*)	GT repeat polymorphism (intron 7)	South African Caucasian	12 and 14 GT repeat variants associated with ATP.	(Mokone et al. 2005)
	rs2104772 rs1330363	South African and Australian Caucasian	T allele of rs2104772 less frequent in TEN. G allele of rs1330363 more frequent in TEN.	(Saunders et al. 2013)
Type V Collagen α1 (*COL5A1*)	rs12722	South African Caucasian	A2 allele overrepresented in CON.	(Mokone et al. 2006)
Type V Collagen α1 (*COL5A1*)	rs12722 rs3196378	South African and Australian Caucasian	CC genotype of rs12722 underrepresented in TEN. TC inferred haplotype rs12722-rs3196378 associated with TEN in SA.	(September et al. 2009)
Matrix metalloproteinase 3 (*MMP3*)	rs679620 rs591058 rs650108	South African Caucasian	GG genotype of rs679620, CC genotype of rs591058 and AA genotype of rs650108 overrepresented in TEN. Rs679620-rs591058-rs650108 associated as inferred haplotype.	(Raleigh et al. 2009)
	rs679620	British Caucasian	GG genotype overrepresented in RUP.	(El Khoury et al. 2016)
Growth differentiation factor 5 (*GDF5*)	rs143383	South African and Australian Caucasian	TT genotype overrepresented in AUS ATP and combined ATP (SA & AUS).	(Posthumus et al. 2010)
Caspase 8 (*CASP8*)	rs1045485 rs3834129	South African and Australian Caucasian	Significant differences in genotype for rs3834129 & rs1045485 between CON and TEN.	(Nell et al. 2012)

Type XXVII Collagen α1 (*COL27A1*) and Tenascin C (*TNC*)	rs946053 rs13321 rs2104772	South African and Australian Caucasian	rs946053-rs13321-rs2104772 (GCA haplotype) overrepresented in TEN.	(Saunders et al. 2013)
Tissue inhibitor of metalloproteinase (*TIMP2*)	rs4789932	South African and Australian Caucasian	CC genotype overrepresented in CON. CT genotype overrepresented in ATP.	(El Khoury et al. 2013)
		British Caucasian	CT genotype overrepresented in male CON.	(El Khoury et al. 2016)
Fibrillin 2 (*FBN2*)	rs331079	South African and Australian Caucasian	GG genotype overrepresented in TEN.	(El Khoury et al. 2015)

Summary table of the genetic association studies published to date in relation to ATP. The gene is listed along with the type of genetic variation and a brief overview of the findings. The research paper is referenced in the far right column.

1.5 COPY NUMBER VARIATION

The completion of the Human Genome Project in 2003 revealed many important insights into the genome (Ionita-Laza et al. 2009). Variability in the human genome is vast and research has shown it can take many forms (Redon et al. 2006). These range from small scale variants such as SNPs, to large chromosomal anomalies (Redon et al. 2006). Data from the Human Genome Project illustrated that SNPs constituted a major part of genetic variation amongst individuals (Ionita-Laza et al. 2009). As a result, much research was focused on this type of genetic variation and by the end of the 20th century, millions of SNPs had been identified (Dear 2009). With advances in research came a greater understanding of the variation within the genome (**Figure 1.9**) (Beckmann et al. 2007). An important discovery was that missing or additional copies of some DNA regions was typical in healthy human genomes (Dear 2009). These structural differences were classed as copy number variation (CNV). Copy number variants (CNVs) are now considered a common type of genetic variation and account for a substantial amount of genetic variability in humans (Ionita-Laza et al. 2009).

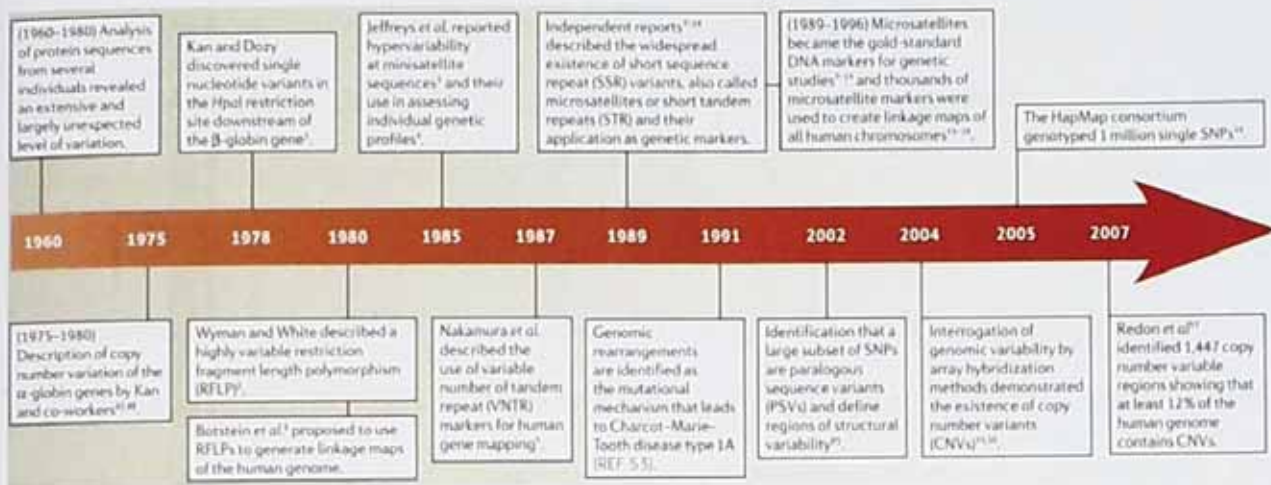


Figure 1.9 Timeline of Studies in Human Genetic Variation (Beckmann et al. 2007).

CNVs are segments of DNA greater than 1kb in size, which show altered copy number when compared to a reference genome (Redon et al. 2006). These structural changes occur in different forms, including insertions, duplications and deletions and are attracting increased interest for their potential role in genetic diversity and disease (Redon et al. 2006; Henrichsen et al. 2009). Copy number variation (CNV) has the potential to influence phenotype via altering gene dosage and disrupting coding sequences of DNA (Yang et al. 2008). An increase in copy number of a gene or gene region could result in increased expression of that gene, while is could similarly silence gene expression if inserted within a coding region (Dear 2009).

Since the 1980s the existence of CNV as a form of genetic variation was largely accepted, however CNV discovery dates back much further than this (Beckmann et al. 2007). In the 1900's the first non-pathological CNV was recognised in humans, showing that males have a single X sex chromosome and a vestigial Y

sex chromosome (Dear 2009). In contrast, trisomy 21 or Down's syndrome, was the first human autosomal aneuploidy identified causing adverse phenotype consequences (Dear 2009). Despite knowledge about the presence of CNV within the genome, the mechanism by which CNV arises is still not fully understood. Research has shown that CNVs may appear to show a Mendelian pattern of inheritance, as opposed to earlier assumptions that CNV arose from independent mutations (McCarroll 2008). Nevertheless, while it appears that a considerable amount of CNV is inherited, these polymorphisms also occur de novo within the genome (Hastings et al. 2009). Furthermore, early knowledge about CNV was restricted to changes visible microscopically, whereas advanced technology has led to identification of a wide range of CNV (Robberecht et al. 2013). While homologous recombination (HR) is an important DNA repair mechanism, it can be the cause of structural chromosomal change, namely CNV (Hastings et al. 2009). Nonallelic homologous recombination (NAHR) is another important mechanism involved in the generation of CNV that can occur in both meiosis and mitosis, through rearrangements of the genome such as chromosomal translocations (Zhang et al. 2009). Indeed, NAHR can occur following unequal crossing-over between two nonallelic DNA sequences (Zhang et al. 2009). Another major mechanisms believed to be involved in the formation of CNV is nonhomologous end-joining (NHEJ) (Zhang et al. 2009; Hastings et al. 2009). Under normal circumstances, NHEJ functions to repair double strand breaks (DSBs) in DNA caused by reactive oxygen species, as well as physiological V(D)J recombination (Zhang et al. 2009).

1.5.1 The Effect of CNV on the Genome

In comparison to SNPs, CNVs appear to have a much higher de novo locus-specific mutation rate (Zhang et al. 2009). While the mutation rate of SNPs is known to be comparatively constant across the genome, CNVs show a diverse and varied mutation rate at different loci (Zhang et al. 2009). It is estimated that CNV locus-specific mutation rates can vary from 1.7×10^{-6} to 1.0×10^{-4} per locus per generation (Zhang et al. 2009). Indeed, more than 10% of the human genome is believed to contain CNVs, highlighting the extent of this type of variation (Redon et al. 2006). The majority of CNV within the genome is believed to be benign and have no phenotypic consequences (Dear 2009). It is expected that CNV causing a phenotypic effect would largely be selected out of the population, however there is evidence showing associations with CNV and complex diseases (Dear 2009; Ionita-Laza et al. 2009).

Research into neurological disorders including autism spectrum disorder, Major Depressive Disorder and schizophrenia has revealed potential associations with CNV (Holt et al. 2012; Degenhardt et al. 2012; Grozeva et al. 2012). However, the association between CNV and schizophrenia described by Grozeva et al. (2012) was not supported by Zhao et al. (2012), who reported a lack of association. CNV is also believed to associate with the autoimmune diseases rheumatoid arthritis (RA) and systemic lupus erythematosus (SLE), where low copy number of the Fc gamma receptor 3B (*FCGR3B*) gene was observed in both RA and SLE (Graf et al. 2012; Yu et al. 2010; Molokhia et al. 2011). Molokhia et al. (2011) reported that low copy number of the *FCGR3B* gene was associated with an increased risk of SLE in an Afro-Caribbean population. In addition, increased copy number of the histamine H4 receptor (HRH4) has also been shown

to increase the risk of SLE, suggesting that CNV may be responsible in part for the increased expression of HRH4 observed in SLE patients (Yu et al. 2010). These seems plausible since altered copy number is known to affect gene expression, alter gene dosage and cause phenotypic variation (Redon et al. 2006).

Graf et al. (2012) investigated CNV within intron 5 of the *FCGR3B* gene and reported that low copy number at this loci was associated with an increased risk of RA in patients recruited from hospitals in Australia (Graf et al. 2012). As the Fc gamma receptor 3B is known to be involved in interactions between immune complexes and immune cells including neutrophils, it is clear to see how altered copy number of the *FCGR3B* gene can have a physiological effect on the genome and alter the risk of such autoimmune diseases discussed above (Graf et al. 2012). RA is a musculoskeletal soft tissue pathology, as is ATP, highlighting an interest in investigating CNV in other musculoskeletal conditions. As predisposition to certain multifactorial diseases appears to be associated with CNV within the human genome (Estivill & Armengol, 2007), it is plausible therefore to hypothesise that CNV may also have a role in ATP. This hypothesis has never been tested.

1.6 EPIGENETICS

Epigenetics is a term used to describe the heritable changes in gene expression that arise without changes in the coding DNA sequence (Reynard and Loughlin 2012; Raleigh 2012; Roach et al. 2005). The epigenetic regulation of the genome occurs by three principal mechanisms, namely histone modification, non-coding RNAs (for example micro RNAs (miRNAs)) and DNA methylation (Rakyan et al. 2011; Reynard and Loughlin 2012). While histone modification and DNA methylation can influence gene transcription, miRNAs can act post-transcriptionally, with all mechanisms having the shared role of regulating gene expression and affecting downstream protein production (Reynard and Loughlin 2012; Auclair and Weber 2012).

In order or all the DNA to fit inside the nucleus of a eukaryotic cell, it needs to be tightly packaged and organised (Das and Tyler 2013). Consequently, DNA is combined with chromatin and wrapped around histone proteins to form a histone octamer (Reynard and Loughlin 2012). This unit forms the structure of a nucleosome, a stable structure within the nucleus supported by hundreds of DNA-protein interactions, that in addition to packaging the DNA, functions to regulate DNA replication, transcription and repair (Das and Tyler 2013). Histone acetylation is an example of histone modification, which regulates gene expression without altering the DNA sequence (Rakyan et al. 2011; Raleigh 2012). Epigenetic regulation of the genome can also occur through the action of miRNAs on mRNA (Bartel 2004; Reynard and Loughlin 2012). At around 20 bp in length, these miRNA molecules are present in the cytoplasm of the cell and can bind to target mRNAs, regulating gene expression post-transcriptionally (Bartel 2004; Reynard and Loughlin 2012). Specifically the interaction of miRNA and mRNA occurs within

the 3-UTR of the target mRNA molecule through complementary base pairing (Reynard and Loughlin 2012). Depending on the level of complementarity, the mRNA may be cleaved and degraded, or translation may be suppressed and result in gene silencing (Reynard and Loughlin 2012; Bartel 2004). The epigenetic mechanism focused on in this thesis is DNA methylation, namely CpG island methylation. This is described in section 1.6.1 below.

1.6.1 DNA Methylation

Of the epigenetic mechanisms detailed above, DNA methylation is the most common epigenetic modification present in eukaryotic DNA (Auclair and Weber 2012). Indeed, DNA methylation of cytosine residues within CpG islands of a gene promoter is predominantly responsible for silencing non-expressed genes within a cell (Roach et al. 2005). This results in distinct cell types with specific functions and is a vital process during embryonic development and cellular differentiation (Roach et al. 2005; Auclair and Weber 2012). The process of DNA methylation typically occurs within CG dinucleotides (CpG sites) and involves the addition of a methyl group to carbon 5 of the cytosine residue to make 5-methylcytosine (Auclair and Weber 2012; Reynard and Loughlin 2012). These sites are collectively termed CpG islands are sparse throughout the genome but are known to be concentrated within the promoter regions of around 30% of genes, where their methylation status can regulate gene transcription by preventing the binding of transcription factors (Reynard and Loughlin 2012). Hypermethylation of a gene promoter is generally associated with the silencing of that gene or suppressed expression (Raleigh 2012). Meanwhile, hypomethylation typically results in increased gene expression (Raleigh 2012; Reynard and Loughlin 2012). As such, the regulation and maintenance of DNA methylation is vital, as abnormal changes

in this epigenetic process can have pathologic consequences (Roach et al. 2005; Auclair and Weber 2012; Rakyan et al. 2011).

1.6.2 DNA Methylation and Disease

While epigenetic profiles of individuals change during development and the differentiation of different tissue types, altered epigenetic states have been associated with disease (Reynard and Loughlin 2012). Although epigenetic variation has largely been investigated in relation to cancer, research in to the epigenetic component of other complex diseases is now emerging (Rakyan et al. 2011). Research into the musculoskeletal disease osteoarthritis is highlighting that altered DNA methylation status is involved in this polygenic disease (Reynard and Loughlin 2012). For example, loss of DNA methylation and altered expression of metalloproteinase genes including *ADAMTS4* has been observed in chondrocytes in osteoarthritis, showing an association between the *ADAMTS4* gene and this musculoskeletal disease (Roach et al. 2005). ADAMTS4 is an enzyme involved in the homeostasis of the ECM and as well as its association with osteoarthritis, higher levels of ADAMTS4 mRNA have been detected in ruptured Achilles tendon compared with normal and tendinopathic tendon (Corps et al. 2008). Furthermore, loss of DNA methylation and altered expression was identified in the promoters of the *MMP3*, *MMP9* and *MMP13* genes in these chondrocytes (Roach et al. 2005). Promoter DNA methylation has also been demonstrated to contribute to the regulation of the *GDF5* gene in osteoarthritis (Reynard et al. 2011).

Many of the proteins involved in tendon biology show altered expression levels in ATP (Ireland et al. 2001). For example, expression of tenascin-C has been shown to increase, while MMP3 expression was shown to decrease in ATP (Ireland et al.

2001; Jelinsky et al. 2011). This highlights the interest in investigating DNA methylation status in tendon-associated genes, to see whether methylation status may be influencing expression levels or modifying the risk of ATP. There is no published research to date on the role of epigenetics as a risk factor for ATP. Therefore, CpG methylation status was an important focus of this thesis.

1.7 AIMS AND OBJECTIVES

This research focuses on the novel genetic and epigenetic basis of sport related tendinopathies. Published research on the genetic predisposition to tendon injury is increasing, further highlighting the importance of this risk factor in relation to tendinopathy. Although there is evidence for a genetic predisposition to ATP, research to date has focused on the involvement of small-scale variants including SNPs. The role of genomic copy number variation in relation to ATP has never been investigated. Based on this, the primary aims of this thesis were:

- To determine whether CNV in a number of candidate genes previously associated with ATP (*CASP8*, *COL5A1*, *MMP3* and *TNC*) was associated with ATP in South African and Australian Caucasian populations.

- To determine whether CNV in a number of novel candidate genes (*ADAM8*, *COL4A1* and *MMP23B*) was associated with ATP South African and Australian Caucasian populations.

- To investigate whether CNV in both previously associated and novel candidate genes was associated with ATP in a newly recruited British UK case-control population.

The additional aims of this thesis were focused around novel SNPs within candidate genes and the risk of ATP and the investigation of DNA methylation status in relation to patellar tendinopathy. Accordingly the further aims of this thesis were:

- To determine whether SNPs in novel candidate genes involved in apoptosis (*TNFRSF1A* and *CASP3*) were associated with ATP in a UK and South African population.

- To determine whether there were differences in promoter DNA methylation profiles within the *ADAMTS4*, *TIMP1* and *MMP11* genes in patellar tendinopathy and healthy patellar tendon tissue.

CHAPTER 2: METHODS

2.1 INVESTIGATING GENETIC ASSOCIATIONS IN COMPLEX DISEASES

The research undertaken in this thesis was based on the candidate gene case-control study approach, where genes and/or pathways were investigated based on their biological role in the tendon, or where genes had been previously associated with tendon pathology. The candidate gene approach is a form of genetic association study, where genetic variation and disease status are investigated to look for potential correlations (Lewis and Knight 2012). For example, a higher observed frequency of a particular genetic variant in a disease population may mean that the variant increases the risk of that disease (Lewis and Knight 2012). Likewise, an increased frequency of a genetic variant in a control population could indicate a protective effect and a reduced risk of disease. Genetic association studies are vital in identifying risk factors for complex diseases, where many polymorphisms in candidate genes have been shown to associate with complex diseases such as autoimmune and psychiatric diseases (Lewis and Knight 2012). While single nucleotide polymorphisms SNPs are the most commonly explored markers in genetic association studies, other types of genetic variants including insertion/deletions, variable-number tandem repeats (VNTRs) and copy number variants CNVs are also investigated (Lewis and Knight 2012).

The investigation of the genetic component of complex diseases is centred on two predominant methods, genome-wide association studies (GWAS) and the candidate gene approach (Amos et al. 2011; Wilkening et al. 2009). While candidate gene association studies were historically the approach of choice when

investigating the genetic susceptibility to a disease or trait, advancing technology has now also highlighted the relevance of the GWAS method (Amos et al. 2011; Wilkening et al. 2009). However, each approach has its strengths and weaknesses and are both relevant methods for association studies (Amos et al. 2011; Patnala et al. 2013). Furthermore, important things to consider when conducting either type of genetic association study include the clear definition of the disease phenotype, as well as the use of a matched control population, which will help to reduce the chance of false-positive results from population stratification (Lewis and Knight 2012).

Candidate gene association studies employ a hypothesis driven approach, where a prior knowledge of a genes function or biological pathway is required (Wilkening et al. 2009). Indeed, the selection of genes is based on previous associations with the disease or trait in question (Patnala et al. 2013). GWAS do not require a prior knowledge of gene function, as the whole genome can be screened and associations may be found in genes with no obvious relevance to the disease being studied (Amos et al. 2011). While it may seem advantageous to be able to discover many associations within genes without a knowledge of their functional relevance, these studies require a very large samples size and an alpha level many times lower than $p < 0.05$, due to the low power of these studies as a result of multiple testing (Amos et al. 2011). This is important to reduce the risk of detecting false positives in this GWAS approach (Amos et al. 2011). In contrast, the candidate gene approach can provide higher statistical power and can also prove beneficial in combination with GWAS (Amos et al. 2011; Patnala et al. 2013). Candidate gene association studies can be a useful method for post-GWAS research, where the identification of the causal variants for a trait can be unearthed from all the

associated genes identified (Wilkening et al. 2009). The candidate gene approach may often be a more cost and time effective method for investigating genetic associations, particularly when a prior knowledge of gene function is employed when selecting candidates (Wilkening et al. 2009). As such, despite the advancement of high throughput genotyping and the screening of many variants at a time, there is a vital place for the candidate gene approach in genetic association studies (Lewis and Knight 2012; Peters 2009). These types of genetic association studies can advance our understanding of complex diseases and identify candidates for further functional/mechanistic studies (Peters 2009). Furthermore, the value of genetic association studies alone is also significant, as certain variants that are associated with a disease are in non-coding regions or are synonymous, resulting in no change to the amino acid sequence of the protein (Peters 2009). In relation to Achilles tendon pathology, the case-control candidate gene approach has been widely used in many studies in multiple cohorts to identify genetic risk factors for this pathology (Raleigh and Collins 2012).

2.2 PARTICIPANTS

2.2.1 Ethical Approval

The research conducted and reported in this thesis forms part of a multicentred, on-going, collaborative study between the University of Northampton (headed by Dr Stuart M Raleigh) and colleagues based at the University of Cape Town (Professor Malcolm Collins), Monash University (Professor Jill Cook) and Seth O'Neill based at the University of Leicester. Ethical approval for these studies was granted from the Research Ethics Committees of the Faculty of Health Sciences at the University of Cape Town (South Africa), La Trobe University (Australia), Monash University (Australia), the University of Northampton (UK) and the University of Leicester (UK). Further information regarding the ethics approval can be found in the Appendix I.

2.2.2 British Achilles Tendon Cohort

A total of one hundred and thirty seven British Caucasian participants diagnosed with ATP were recruited for this case-control genetic association study through the County Clinic in Northampton, UK. One hundred and thirty one asymptomatic Caucasian control (CON) participants (apparently healthy, physically active and without history of tendon/ligament pathology) were recruited from the East Midlands region of the UK. Recruitment was initially undertaken by Dr Louis El Khoury, and later continued by the author. The ATP participants were all non-NHS patients clinically diagnosed by Professor William J Ribbans. Patients with either chronic Achilles tendinopathy (noninsertional or insertional tendinopathy) (TEN) or partial/complete Achilles tendon ruptures (RUP), based on established published criteria (Schepsis et al. 2002; Kader et al. 2002; Mokone et al. 2006),

were included in the study. For all cases, the clinical diagnosis of Achilles tendinopathy was verified by subsequent imaging with MRI and/or ultrasound.

The clinical criteria used for the diagnosis of chronic Achilles tendinopathy were gradual progressive pain over the Achilles tendon area for greater than 6 months, together with at least one of six criteria (Mokone et al. 2006; September et al. 2009). These were: 1) early-morning pain of Achilles tendon area; 2) early-morning stiffness of Achilles tendon area; 3) history of swelling of Achilles tendon area; 4) tenderness to palpation of Achilles tendon; 5) palpable nodular thickening of affected Achilles; 6) movement of the painful area in the Achilles tendon with plantar-dorsi-flexion (positive "shift" test) (Mokone et al. 2006).

Written informed consent was given by all participants according to the Declaration of Helsinki. The participant information sheet and consent form can be seen in Appendix II. Participants who returned signed consent forms were then asked to completed a physical activity/medical history/injury questionnaire (Appendix III) and provide a saliva sample.

An additional 28 British Caucasian non-NHS ATP participants were recruited through physiotherapist Seth O'Neill at the University of Leicester, via adverts at running clubs and through the www.achillestendonresearch.com website. All participants were runners with self-reported tendon pain for more than 3 months. Achilles tendinopathy was diagnosed during clinical examination with the presence of localised tendon pain, with pain on palpation and increased pain with tendon loading (bilateral heel raise, unilateral heel raise, hopping, jumping and running). Ultrasonography was completed on all individuals. These samples were included

in the studies in Chapter 5, but none of the other studies. An overview of the British (both groups combined) participant characteristics can be seen in **Table 2.1**.

Table 2.1 General characteristics of the CON, ATP, TEN, and RUP groups in the UK cohort

	CON (n=131)	ATP (n=165)	p-value	TEN (n=128)	p-value	RUP[b] (n=37)	p-value
Age (years)[a]	41.3 ± 11.3 (122)	46.3 ± 13.6 (164)	0.001	46.5 ± 13.5 (128)	0.001	45.4 ± 14.1 (36)	0.074
Gender (% male)	62.6 (82)	63.0 (104)	1.000	61.7 (79)	0.821	67.6 (25)	0.616
Height (cm)	175.0 ± 10.5 (121)	173.1 ± 9.2 (162)	0.101 0.013[c]	172.6 ± 8.7 (125)	0.056 0.009[c]	174.5 ± 10.8 (37)	0.806 0.300[c]
Weight (kg)	80.5 ± 19.6 (122)	78.9 ± 14.8 (128)	0.485 0.996[d]	77.9 ± 14.4 (99)	0.279 0.244[d]	82.4 ± 15.9 (29)	0.611 0.420[d]
BMI (kg m^{-2})	25.7 ± 5.1 (122)	26.3 ± 4.1 (131)	0.316 0.671[e]	26.1 ± 4.2 (101)	0.534 0.762[e]	27.0 ± 3.8 (30)	0.223 0.462[e]

Differences in the characteristics of CON vs. ATP, CON vs. TEN and CON vs. RUP groups were analysed using ANOVA and χ^2 tests. For full details see Chapter 2: 2.5.2 Data Analysis. Values are expressed as means ± standard deviation (SD) with the total number of participants (n) is in parentheses. Gender is the percentage of males with number of males in parentheses. cm: centimetres; kg: kilograms; m: metres. [a] Age of CON is age at recruitment, while age of ATP, TEN and RUP is age at initial injury. [b] The RUP group includes those that initially presented with rupture and those that later developed ruptures from an initial diagnosis of tendinopathy. [c] P value co-varied for gender. [d] P value co-varied for age, gender and height. [e] P value co-varied for age and gender.

2.2.3 South African and Australian Achilles Tendon Cohorts

One hundred and seventy-five (115 South African (SA) and 60 Australian (AUS)) self-reported Caucasian participants diagnosed with ATP and 247 (98 SA and 149 AUS) asymptomatic CON were recruited, as previously described (Mokone et al. 2006; September et al. 2009). To avoid possible effects of population stratification, the ATP and CON groups were matched for country of birth (September et al. 2009).

SA ATP participants with a current/past clinical history of ATP were recruited from the Medical Practice at the Sports Science Institute of South Africa and other clinical practices within the Cape Town area of South Africa. Initial diagnoses of ATP were made by an experienced clinician using clinical criteria and the diagnoses of all SA ATP participants was reviewed and confirmed by Martin P Schwellnus (Mokone et al. 2005; Mokone et al. 2006). SA CON participants (apparently healthy, physically active and without history of ATP) were recruited from recreational sporting clubs (Mokone et al. 2006). AUS ATP participants were recruited by the Musculoskeletal Research Centre at La Trobe University in Melbourne, Australia and were diagnosed by Jill Cook using clinical criteria (Mokone et al. 2005; Mokone et al. 2006). AUS CON participants (apparently healthy and without history of tendon/ligament pathology) were recruited from Melbourne, Australia (September et al. 2009). As with the British cohort, participants gave written informed consent and completed a physical activity/medical history/injury questionnaire similar to those shown in Appendix II and III. Ethical approval for use of the samples is detailed in Appendix I.

The SA and AUS ATP groups consisted of participants clinically diagnosed with either chronic Achilles tendinopathy or partial/complete Achilles tendon ruptures (defined for the SA cohort only). Cases of Achilles tendinopathy typically presented with gradual progressive pain with early-morning pain/stiffness in the Achilles tendon area. The clinical criteria used for the diagnosis of chronic Achilles tendinopathy are described in detail elsewhere (Mokone et al. 2005; Mokone et al. 2006). In addition, soft-tissue ultrasound examination was performed (all AUS ATP participants and a sub-group of SA ATP participants) to confirm diagnosis. Achilles tendon rupture for the SA cohort was diagnosed with clinically validated criteria, confirmed by examination at time of surgery and/or by ultrasound imaging, magnetic resonance imaging (MRI) or computerised tomography (CT) scan, as previously described (Mokone et al. 2006). Details of the SA and AUS participant characteristics can be seen in **Table 2.2** and **2.3**.

Table 2.2 General characteristics of the CON, ATP, TEN, and RUP groups in the SA cohort

	CON (n=98)	ATP (n=115)	p-value	TEN (n=75)	p-value	RUP[b] (n=40)	p-value
Age (years)[a]	36.8 ± 9.9 (91)	40.5 ± 12.9 (107)	0.025	40.5 ± 13.7 (69)	0.051	40.7 ± 11.4 (38)	0.056
Gender (% male)	66.3 (65)	71.3 (82)	0.327	72.0 (54)	0.401	70.0 (28)	0.451
Height (cm)	175.5 ± 9.7 (95)	176.1 ± 8.8 (104)	0.671 0.614[b]	176.6 ± 8.9 (66)	0.495 0.960[b]	175.3 ± 8.7 (38)	0.887 0.354[b]
Weight (kg)	72.1 ± 12.2 (97)	81.1 ± 15.3 (107)	0.000 0.000[c]	78.4 ± 14.1 (69)	0.003 0.005[c]	86.2 ± 16.2 (38)	0.000 0.000[c]
BMI (kg m^{-2})	23.3 ± 2.8 (95)	26.0 ± 4.0 (104)	0.000 0.000[d]	24.9 ± 3.4 (66)	0.001 0.005[d]	27.9 ± 4.2 (38)	0.000 0.000[d]

Differences in the characteristics of CON vs. ATP, CON vs. TEN and CON vs. RUP groups were analysed using ANOVA and χ^2 tests. For full details see Chapter 2: 2.5.2 Data Analysis. Values are expressed as means ± SD with the total number of participants (n) is in parentheses. Gender is the percentage of males with number of males in parentheses. cm: centimetres; kg: kilograms; m: metres. [a] Age of CON is age at recruitment, while age of ATP, TEN and RUP is age at initial injury. [b] P value co-varied for gender. [c] P value co-varied for age, gender and height. [d] P value co-varied for age and gender.

Table 2.3 General characteristics of the CON, ATP and TEN groups in the AUS cohort

	CON (n=149)	ATP (n=60)	p-value	TEN (n=60)	p-value
Age (years)[a]	38.8 ± 12.0 (146)	40.3 ± 14.0 (59)	0.422	40.3 ± 14.0 (59)	0.422
Gender (% male)	39.6 (59)	66.7 (40)	0.000	66.7 (40)	0.000
Height (cm)	171.7 ± 9.2 (147)	173.4 ± 9.7 (58)	0.236 0.126[b]	173.4 ± 9.7 (58)	0.236 0.126[b]
Weight (kg)	73.4 ± 14.0 (148)	80.4 ± 14.9 (60)	0.002 0.371[c]	80.4 ± 14.9 (60)	0.002 0.371[c]
BMI (kg m^{-2})	24.9 ± 4.0 (147)	26.7 ± 4.2 (58)	0.005 0.383[d]	26.7 ± 4.2 (58)	0.005 0.383[d]

Differences in the characteristics of CON vs. ATP and CON vs. TEN groups were analysed using ANOVA and χ^2 tests. For full details see Chapter 2: 2.5.2 Data Analysis. Values are expressed as means ± SD with the total number of participants (n) is in parentheses. Gender is the percentage of males with number of males in parentheses. cm: centimetres; kg: kilograms; m: metres. [a] Age of CON is age at recruitment, while age of ATP and TEN is age at initial injury. [b] P value co-varied for gender. [c] P value co-varied for age, gender and height. [d] P value co-varied for age and gender.

2.2.4 Australian Patellar Tendon Tissue Samples

In addition to the Achilles tendon cohorts recruited and described above, DNA samples from 20 patellar tendon samples (all male) were used for DNA methylation analysis. These were obtained from La Trobe University, Melbourne, Australia. As previously described by Parkinson et al. (2009), the patellar tendon samples included 10 abnormal samples obtained from individuals during surgical debridement for recalcitrant overuse patellar tendinopathy and 10 healthy samples obtained during anterior cruciate ligament (ACL) reconstruction surgery using a patellar tendon graft (Parkinson et al. 2010). DNA from the patellar tendon tissue samples was isolated by members of staff at La Trobe University using the PureLink Genomic DNA Kit (Invitrogen, Carlsbad, CA, USA) as per the manufacturer's recommendations.

2.2.5 DNA Extraction and Quantification

For the British cohort, DNA was extracted from 2 mL of saliva collected using ORAGENE-DNA collection kits (OG-500) and DNA purification was carried out using the prepIT-L2P DNA extraction kit (DNA Genotek Inc., Ontario, Canada). PrepIT-L2P purifier reagent was added and mixed by vortexing to precipitate impurities and inhibitors. After incubation on ice, samples were centrifuged at 13,000 rpm, resulting in a formed pellet of impurities at the bottom of the micro-centrifuge tubes and samples with reduced turbidity. The supernatant was transferred into clean micro-centrifuge tubes and 90% ethanol added. This precipitated the DNA, forming a visible clot of DNA fibres. Samples were centrifuged again at 13,000 rpm, with the DNA forming as a small pellet. Supernatant was removed and discarded and 70% ethanol was then added. Once removed, any remaining ethanol was allowed to evaporate. DNA pellets were hydrated in TE buffer and left

at room temperature overnight. Samples were vortexed to ensure full rehydration before measurement of concentration and sample storage. The protocol for DNA extraction can be seen in Appendix IV.

DNA concentration (ng/μL) and purity (260/280 ratio) were measured using the Nano Drop 2000 (Thermo Fisher Scientific, USA). The 260/280 ratio is the ratio of absorbance at 260 nm and 280 nm and is used as a measure of nucleic acid purity. A ratio of ~1.8 is typically accepted as pure for DNA, although it is dependent upon the nucleotide composition of the nucleic acid. A ratio of lower than this may indicate the presence of contaminants, such as protein, that absorb strongly at ~280 nm. Working-stock plates of DNA samples were prepared by diluting DNA samples to 10 ng/μL with TE buffer. These working-stock plates were stored at 2-5 °C. Long-stock DNA samples were stored in aliquots at -20 °C. Dr Louis El Khoury assisted with participant recruitment and DNA extraction for the British cohort.

For the South African cohort, DNA was provided as aliquots. Samples had been prepared from 4.5 mL of venous blood, obtained by venipuncture of a forearm vein and collected into an ethylenediaminetetraacetic acid (EDTA) vacutainer tube. Blood samples were stored at 4°C until total DNA extraction, using a procedure described by Lahiri and Nurnberger (1991), with modifications by Mokone et al. (2006). DNA was extracted for the Australian cohort by members of staff from the Musculoskeletal Research Centre at La Trobe University, from approximately 4.5 mL of venous blood, using a sequence extraction technique (FlexiGene DNA Kit, Qiagen P/L, Valencia, California, USA) as per the manufacturer's recommendations.

2.3 QUANTITATIVE PCR METHODS

2.3.1 TaqMan Technology

TaqMan technology was the predominant method used in this thesis for the sections on both CNV and SNP genotyping, using a quantitative PCR (qPCR) platform. TaqMan Copy Number Assays were used for all the studies on copy number and are a well-established and robust method for copy number determination (D'haene et al. 2010). Advantages of this method include the fact that reactions can be performed as real-time duplex PCR reactions. This means that the target CNV region co-amplifies with the reference assay RNaseP within the same well. Indeed, it was found in the present study, that this minimised the time and cost of each PCR analysis. A detailed discussion of the advantages of using qPCR methods compared to alternative genotyping strategies can be found elsewhere (Bodin et al. 2005; D'haene et al. 2010). In addition to rapid genotype calling and cost effectiveness, the CNV assays used were found to be reproducible. In the present research, copy number was determined four times for each DNA sample and for each probe used. Copy number call was identical between each of the four replicates for all probes.

Validation of our qPCR methods for CNV determination was also confirmed by establishing the PCR efficiency for each TaqMan Copy Number probe used. Standard curves of known DNA concentration vs. probe specific fluorescent signal can be seen in Appendix V. For each TaqMan probe used, PCR efficiency averaged >90% and there was never more than a 10% difference between the efficiency of the CNV target probe compared to the efficiency of the reference RNaseP probe used to normalise copy number values. In addition to determining the quality of reproducibility, each qPCR run contained no template controls (NTCs) to check for

spurious calls and possible contamination. Furthermore, all samples used for the qPCR efficiency studies were also crosschecked against the original qPCR runs to determine if any discrepancies in concordance existed in copy number.

2.3.2 SNP Selection and Genotyping

The predicted genotype and allele frequencies of the SNPs within the candidate genes investigated in this thesis were checked on databases including Ensembl (http://www.ensembl.org/index.html), NCBI dbSNP (http://www.ncbi.nlm.nih.gov/SNP/) and SNPper (http://snpper.chip.org/bio/snpper-enter/). It was considered important that the minor allele for each SNP investigated was detected in our sample size, to enable an accurate representation of genotypes as seen in the predicted frequencies. Therefore, SNPs were selected with a reported minor allele frequency (MAF) of > 0.15 in Caucasian populations. A lower MAF may have resulted in underrepresentation of the minor allele, affecting the analysis and interpretation of results. A rational for the selection of each gene and assay is detailed within Chapter 5.

TaqMan SNP Genotyping Assay, TaqMan Universal PCR Master Mix, No AmpErase UNG and genomic DNA (gDNA) were combined per well of a 96-well plate (**Table 2.4**). The total reaction volume per well was 12 μL, giving a final DNA reaction concentration of 1 ng/μL.

Table 2.4 Reaction Components for a Typical SNP Genotyping Real-Time PCR.

Reaction Component	Volume per well (µL)
TaqMan Universal PCR Master Mix, No AmpErase UNG	6.0
TaqMan SNP Genotyping Assay, 40x	0.3
Nuclease-free water	4.5
gDNA	1.2
Total Reaction Volume	12.0

The alleles were determined from genomic DNA using TaqMan SNP Genotyping Assays, according to the manufacturer's instructions. Each TaqMan SNP Genotyping Assay contained sequence-specific forward and reverse primers to amplify the target sequence and two TaqMan probes, which contained a minor groove binder (MGB) and a non-fluorescent quencher (NFQ) at the 3' end (**Figure 2.1** Part 1). One probe was provided labelled with VIC dye at the 5' end of the probe and the other with FAM dye at the 5' end of the probe, to detect the minor and major alleles. TaqMan MGB probes provided a fluorescence signal for the amplification of each allele. **Table 2.5** gives an overview of the fluorescence detected and the allele associated with that probe. Assays were diluted from a 40x concentration to a 1x concentration in each reaction.

Real-time qPCR reactions were run in 96-well plates on a StepOnePlus Real-Time PCR System, running StepOne Software version 2.1 (Applied Biosystems, Foster City, California, USA). During qPCR, the samples were heated at 92°C to denature the gDNA template (**Figure 2.1** Part 2). Following cooling to 60 °C, the TaqMan SNP Genotyping MGB probes annealed to their complementary sequence between

the forward and reverse primer binding sites (**Figure 2.1** Part 2). While the oligonucleotide probe is intact, reporter dye fluorescence is quenched due to the proximity of the quencher dye. The TaqMan Universal PCR Master Mix, No AmpErase UNG contained AmpliTaq Gold DNA Polymerase (Life Technologies), which amplified the target gDNA by extending the sequence-specific primers bound to the DNA template. (**Figure 2.1** Part 3). The AmpliTaq Gold DNA Polymerase cleaved the probes hybridised to their target complimentary sequences, separating the reporter dye from the quencher, causing fluorescence of the reporter (**Figure 2.1** Part 3). The PCR products were amplified after each PCR cycle and this was detected by the increase in fluorescence of the reporter dyes. The fluorescence signal generated indicated which alleles were present in the samples. Fluorescence from a single dye only (FAM or VIC) indicated a homozygote genotype at the loci. The detection of fluorescence from both FAM and VIC dyes indicated a heterozygote genotype at the loci (**Table 2.5**). This process continued with each cycle of the PCR.

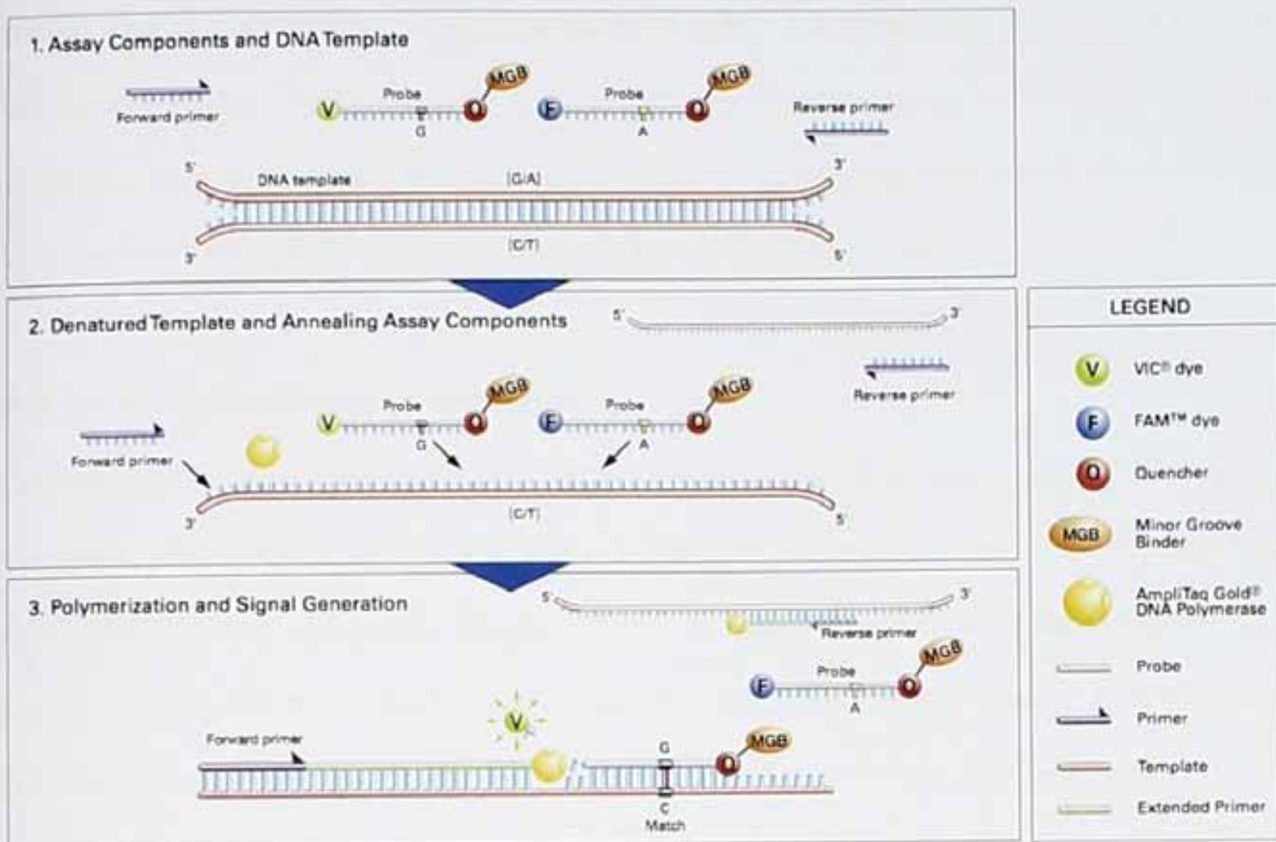


Figure 2.1 Summary of the stages in a real-time qPCR reaction containing TaqMan SNP Genotyping Assay. Image from page 43 of the TaqMan SNP Genotyping Assay Protocol with kind permission of Thermo Fisher Scientific.

Table 2.5 Fluorescence detected and the corresponding genotype identified.

Fluorescence	Genotype
VIC dye fluorescence only	Homozygote for Allele 1
FAM dye fluorescence only	Homozygote for Allele 2
Both VIC and FAM dye fluorescence	Heterozygote Allele1/Allele2

In summary, the cycling conditions were holding at 95 °C for 10 min, followed by 50 cycles of denaturing at 92 °C for 15 sec and annealing/extension at 60 °C for 1 min. Each plate included within and through plate positive controls and no template controls (NTCs) to check reproducibility and to ensure no contamination

within and across plates. Genotype clusters were automatically generated by the software, which distinguished between the homozygote and heterozygote samples. Some samples did not automatically genotype and amplification was checked and samples were manually called where applicable.

2.3.3 CNV Selection and Genotyping

Catalogued CNVs were identified with reference to the Database of Genomic Variants (DGV) (http://projects.tcag.ca/variation/) hosted by The Centre for Applied Genomics, Toronto. TaqMan Copy Number Assays (Applied Biosystems, Foster City, California, USA) were selected using the TaqMan Assay Search Tool (https://www.lifetechnologies.com/uk/en/home/life-science/pcr/real-time-pcr/real-time-pcr-assays/cnv-analysis-using-taqman.html). The criteria used in selecting relevant TaqMan Copy Number Assays for analysis was based on the region being within, or close to, a candidate gene previously associated with ATP and/or in a gene with a known role in tendon biology. Selection was also based on the copy number variation being within a catalogued region reported in the DGV; the probe being in an exonic region of the gene and being in a region where CNV was reported at a frequency >10% in the population studied. Where possible one or more of these selection criteria were met for each assay chosen, although there were times when assays were selected in novel genes without previously reported CNV. Regardless of the selection criteria, a rational for the selection of each gene and target CNV region is detailed within the specific chapters (Chapter 3, 4 & 5). Details of all the TaqMan Copy Number Assays selected and investigated can be seen in **Table 2.6**.

Table 2.6 TaqMan Copy Number Assays Investigated

Gene	Assay ID	Assay Location	Assay Gene Location	Copy Number Variation ID	CNV Location	CNV Size (bp)
ADAM8	Hs02429708_cn	Chr.10:135076357	Exon 23	esv3625037* (gain)	Chr10:135066539 - 135077371	10833
CASP8	Hs02601709_cn	Chr.2:202146600	Intron 11-Intron 12	esv2721643* (del)	Chr2:202146618 - 202149432	2815
COL4A1	Hs00739915_cn	Chr.13:110847375	Intron 22-Exon22	nsv563101* (loss)	Chr13:110818598 - 110856153	37556
COL4A1	Hs02119212_cn	Chr.13:110850912	Exon 21	nsv1052147* (gain)	Chr13:110765635 - 110917782	152148
COL5A1	Hs00180523_cn	Chr.9:137619157	Intron 4-Exon 5	nsv469910* (loss)	Chr9:137598731 - 137642577	43847
COL5A1	Hs02980647_cn	Chr.9:137644529	Exon 14-Intron 14	nsv818731* (gain)	Chr9:137631198 - 137647314	16117
MMP23B	Hs07485808_cn	Chr.1:1567625	Exon 1	esv2659399* (loss)	Chr1:1564645 - 1568392	3749
MMP3	Hs02276714_cn	Chr.11:102712920	Intron 4-Exon 4	n/a (esv3627536)*	(Chr11:102707966 - 102709474)	(1509)
MMP3	Hs02908568_cn	Chr.11:102709884	Exon 7	n/a (esv3627536)*	(Chr11:102707966 - 102709474)	(1509)
TNC	Hs00634176_cn	Chr.9:117793858	Intron 23-Exon 23	nsv831696* (gain)	Chr9:117701582 - 117874150	172570
TNC	Hs06903309_cn	Chr.9:117812454	Intron 15	esv2738965* (del)	Chr9:117811685 - 117813771	2087

Details of the location of each assay, recorded CNV (if applicable) with type of CNV in parenthesis, the location and the size of the CNV. Locations on NCBI build 37. Information compiled from Life Technologies and the DGV. * indicates more than one recorded CNV in this region. n/a indicates no catalogued CNV within the location of the assay investigated, although CNV is catalogued within other regions of the gene (shown in parenthesis).

The TaqMan Copy Number Assays detected the target genomic sequences of interest, while the Reference Assay (RNase P) detected a sequence known to exist in two copies in a diploid genome (Mayo et al. 2010). Specifically, the reference probe detected the human ribonuclease P RNA component H1 (H1RNA) gene (RPPH1) on chromosome 14 (Mayo et al. 2010). The TaqMan Copy Number Reference Assay RNase P is the standard reference assay for human gDNA CN quantification (Mayo et al. 2010).

TaqMan Copy Number Assay, TaqMan Copy Number Reference Assay, TaqMan Universal PCR Master Mix No AmpErase UNG and gDNA were combined per well of a 96-well plate (**Table 2.7**). Total reaction volume per well was 10 μL, giving a final DNA reaction concentration of 1 ng/μL.

Table 2.7 Reaction Components for Copy Number Real-Time PCR.

Reaction Component	Volume per well (μL)
TaqMan Universal PCR Master Mix, No AmpErase UNG	5.0
TaqMan Copy Number Assay, 20x	0.5
TaqMan Copy Number Reference Assay, 20x	0.5
Nuclease-free water	3.0
gDNA	1.0
Total Reaction Volume	10.0

Real-time qPCR reactions were run on a StepOnePlus Real-Time PCR System running StepOne Software version 2.1 (Applied Biosystems, Foster City, California,

USA) in 96-well plates. CN was determined from genomic DNA using TaqMan Copy Number Assay with Reference Assay RNase P, performed as a duplex reaction and according to the manufacturer's instructions. The TaqMan Copy Number Assay contained two primers (forward and reverse) and a FAM dye-labelled nonfluorescent-minor groove binder (NFQ-MGB) probe (**Figure 2.2** Part a). The TaqMan Copy Number Reference Assay RNase P contained two primers and a VIC dye-labelled TAMRA probe (**Figure 2.2** Part a). Both assays are 5′ nuclease assays and were diluted from a 20x concentration to a 1x concentration in each reaction.

Initially, the gDNA template was denatured at 95 °C and assay primers (Copy Number and Reference Assay) annealed to their specific target sequences upon cooling to 60 °C (**Figure 2.2** Part b). Sequencing probes also annealed to their complementary sequences between forward and reverse primer binding sites. While intact, the reporter dye signal of the probes (FAM for Copy Number Assay and VIC for Reference Assay respectively) was quenched, due to the proximity of the quencher (NFQ-MGB /TAMRA for Copy Number Assay and Reference Assay respectively). During each PCR cycle, the target and reference sequences were simultaneously amplified by AmpliTaq Gold DNA Polymerase (Life Technologies). AmpliTaq Gold DNA Polymerase, with 5′ nuclease activity, cleaved the oligonucleotide probes hybridised to each amplicon sequence. Each time a probe was cleaved, the quencher separated from the reporter dye, causing fluorescence to be detected in the real-time PCR reaction (**Figure 2.2** Part c). The increase of amplified PCR products was detected by the increase in fluorescence of the reporter dyes following each PCR cycle.

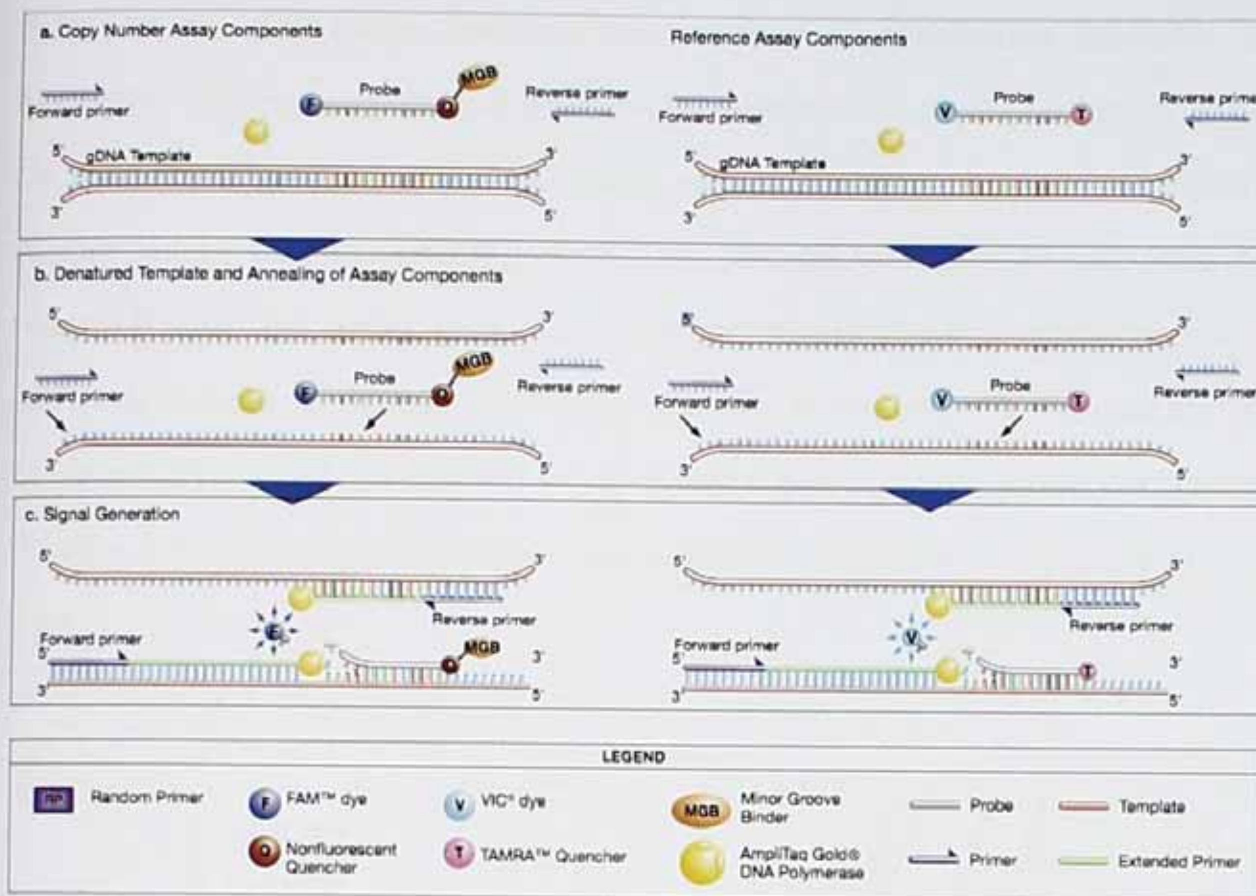


Figure 2.2 Summary of the stages in a duplex real-time qPCR reaction containing TaqMan Copy Number and Reference Assay. Image from page 36 of the TaqMan Copy Number Assay Protocol with kind permission of Thermo Fisher Scientific.

In summary, the cycling conditions were holding at 95 °C for 10 min, followed by 40 cycles of denaturing at 95 °C for 15 sec and annealing/extension at 60 °C for 1 min. Each sample was tested in quadruplicate and all PCR runs included NTCs. Data were exported to CopyCaller Software version 2.0 (Applied Biosystems, Foster City, California, USA) for calculation of CN.

The CN of target sequence was determined by relative quantification (RQ) using the comparative Ct (ΔΔCt) method, where cycle threshold (Ct) is the number of PCR cycles at which PCR product fluorescence is detected (Mayo et al. 2010). CN call was based on the Ct difference (ΔCt) between target and reference (**Figure**

2.3) and the CopyCaller Software used a maximum-likelihood algorithm to determine an integer value for predicted CN (discrete CN value) (Park et al. 2012; Liu et al. 2011) and presented this along with the initial calculated CN, as non-integer values (continuous CN value). This method was used to calculate the relative CN of the target gene normalised to RNase P, without the use of a calibrator sample. CopyCaller Software was used to calculate Confidence and Z-Score quality metrics. All CN calls were reviewed and accepted with a Confidence Value > 80% (although confidence was typically > 99%) and a Z-Score of Z ≤ 2.65.

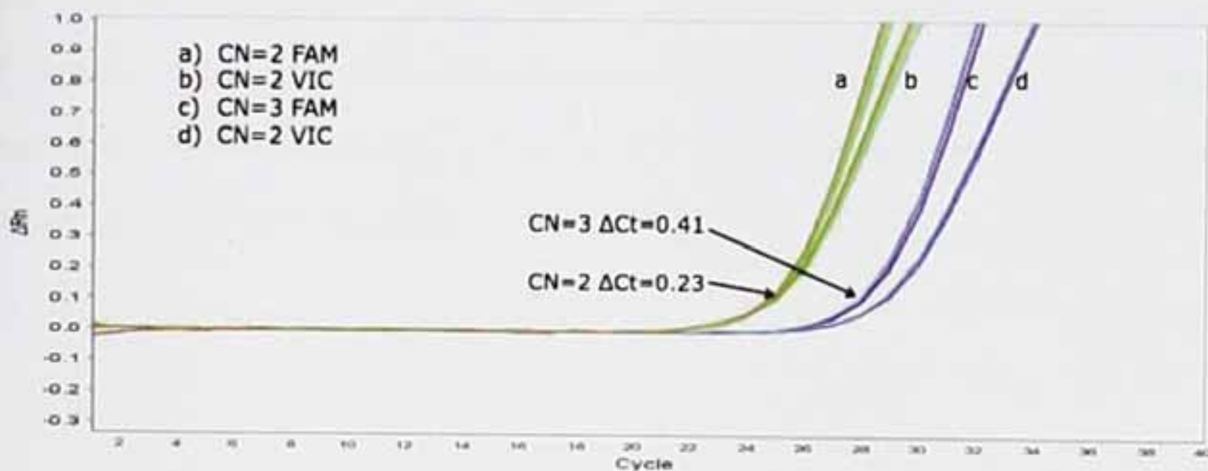


Figure 2.3 Amplification plot showing the quantitative determination of copy number using the ΔΔCt method. CN calls were based on the ΔCt between the target (FAM-labelled) and reference (VIC-labelled) probes. a) shows an example of CN=2 for the target, while c) shows an example of CN=3 for the target, when compared with their corresponding reference assay amplifications (b and d respectively).

2.3.4 QPCR Efficiency

For quality control of CN results, qPCR was carried out using twofold serial dilutions of genomic DNA for each TaqMan Copy Number and Reference Assay. Standard curves were generated, with a linear trend line fit to the data to determine the PCR efficiency for each assay, plotting cycle threshold (Ct) against DNA concentration (log concentration) (D'haene et al. 2010). Efficiency (E) was calculated using the equation $E = 10^{(-1/m)}-1$, where m is the slope of the linear trend line of the standard curve (Liu et al. 2011). Optimal efficiency (100%) would be illustrated when m = -3.32 and would demonstrate the doubling of PCR product with each complete cycle (D'haene et al. 2010). Amplification efficiencies between the range of 90-110% are typically considered acceptable and linearity (represented by the correlation coefficient r^2) should be close to 1 to show linearity across dilution series (D'haene et al. 2010). Details of QPRC efficiencies for each assay can be seen in Appendix V.

2.4 DNA METHYLATION AND PYROSEQUENCING

Candidate genes of interest were viewed in NCBI Gene (www.ncbi.nlm.nih.gov/gene). Predesigned PyroMark CpG Assays were identified using the Qiagen PyroMark CpG Assay search tool (http://www.qiagen.com/products/catalog/assay-technologies/pyrosequencing/pyromark-cpg-assays). For the PyroMark Custom Designed CpG Assays, probes and primers were individually designed using PyroMark Assay Design 2.0 software (Qiagen Valencia, CA, USA). A section of the promoter region before the start site of the target gene was selected and the FASTA sequence downloaded. This was then opened in SnapGene Viewer version 2.2.2 (www.snapgene.com/products/snapgene_viewer/), to be able to view the sequence. The nucleotide sequence was then copied into CpG Island Searcher (www.cpgislands.usc.edu), which detailed CpG islands present in the target region. The selected CpG island identified in CpG Island Searcher was then located in the FASTA sequence. This was copied into PyroMark Assay Design 2.0 software (Qiagen), to design the PCR and sequencing primers for the Pyrosequencing assays. The primer sets were generated by the software based on the sequence entered. Quality scores were assigned to each primer set to indicate the suitability for Pyrosequencing analysis and to assist the selection of an assay. **Table 2.8** details the pre-designed and custom PyroMark CpG Assays investigated.

Table 2.8 PyroMark CpG Assays.

Gene (Chr.)	Assay	PCR and sequencing primers	Biotin labelled primer	Nucleotide dispensation order	Number of CpG sites assayed	Amplicon size (bp)
***ADAMTS4* (Chr.1)**	Hs_NDUFS2_01_PM	-	Reverse	TGTCATGTATGTCGTAGTCG AGAGATAGAGAGTCGTCGTCGTCG	6	244
***MMP11* (Chr.22)**	Hs_CHCHD10_01_PM	-	Reverse	ATATCTGTCGATTAGTTCGT CGTTGTCG	4	141
***TIMP1* (Chr.X)**	Custom Designed Region A	Forward – TTGGGTAAGTTAGT TTAAGGTAGAGTA Reverse – TTCCCACTCCATCC TCATAC Sequencing – CTCCATCCTCATAC TTT	Forward	ATCTCTCTCACATCTCTAGC ATCTCATCTCCGATCAGATC ATCACATCTACTTCTATCAT CACTCATCACTCATCGATCT	3	170
***TIMP1* (Chr.X)**	Custom Designed Region B	Forward – TGGGTTAGAAGGAT GTGAAGG Reverse – AACCCACCCTCTAC AACTT Sequencing - AAGGATGTGAAGGAG	Reverse	AGAGAAGTCATGAGATGAGT GAGAGTAGATGATAGAGATA GTATCGTAGAGTTAGATGTG TAGTCGGTCGAG	3	170

2.4.1 Bisulphite Treatment

Prior to Pyrosequencing analysis, bisulfite conversion was carried out on DNA samples from the patellar tendinopathy cases and controls. Bisulfite treatment of target DNA converted unmethylated cytosine residues into uracil. The methylated cytosines remained unchanged and this resulted in different DNA sequences for methylated and unmethylated DNA (**Table 2.9**).

Table 2.9 The Effect of Bisulfite Treatment.

	Original sequence	After bisulfite treatment	After PCR amplification
Unmethylated DNA	A-C-G-T-C-G-T-C-A	A-**U**-G-T-**U**-G-T-**U**-A	A-T-G-T-T-G-T-T-A
Methylated DNA	A-C-G-T-C-G-T-C-A	A-C-G-T-C-G-T-**U**-A	A-**C**-G-T-**C**-G-T-T-A

Example sequences showing the effect of bisulfite treatment on unmethylated and methylated DNA and the resulting sequences after PCR amplification. Unmethylated cytosines (C) are converted to uracil (**U**) during bisulfite conversion. Subsequent PCR results in only methylated cytosines (**C**) being present in the sequence, as uracil became thymine (T) (Madi et al. 2012).

The conversion of unmethylated cytosines to uracil was critical for the accurate determination of methylation status (Anderson and Brown 2005; Madi et al. 2012). The template DNA was incubated in high bisulfite salt concentrations at a high temperature and low pH. The EpiTect Fast Bisulfite Conversion Kit (Qiagen) was used for conversion and purification of sample DNA. In brief, template genomic DNA was combined with Bisulfite Solution, DNA Protect Buffer and RNase-free water in PCR tubes and mixed thoroughly (**Table 2.10**). Once added to the reaction mix, the DNA Protect Buffer turned from green to blue, indicating the correct pH for the bisulfite conversion.

Table 2.10 Reaction Components for Bisulfite Treatment.

Reaction Component	Volume per well (μL)
DNA	2
RNase-free water	18
Bisulfite Solution	85
DNA Protect Buffer	35
Total Reaction Volume	140

Bisulfite conversion was performed on a Techne TC-512 thermal cycler (Bibby Scientific Ltd, Staffordshire, UK) on a 30 minute programme. The thermal cycler conditions can be seen in **Table 2.11**.

Table 2.11 Thermal Cycling Conditions for Bisulfite Treatment.

Step	Time	Temperature
Denaturation	5 min	95 °C
Incubation	10 min	60 °C
Denaturation	5 min	95 °C
Incubation	10 min	60 °C
Hold	Indefinite	20 °C

Following the bisulfite treatment in the thermal cycler, a series of steps were undertaken to purify the DNA. These steps were necessary to remove salts and chemicals used in the bisulfite conversion process that could inhibit Pyrosequencing. These steps included the addition of different buffers and ethanol

and centrifugation through MinElute DNA spin columns into collections tubes. These steps started with the addition of 310 µL of Buffer BL and 250 µL ethanol (95-100%), vortexing between each addition. The contents were transferred into MinElute DNA spin columns with collection tubes and centrifuged (1 min at 13,000 rpm) to separate the DNA from the leftovers from the bisulfite conversion reactions. Next, 500 µL of Buffer BW (wash buffer) was passed through the spin columns (centrifuge 1 min at 13,000 rpm) and 500 µL of Buffer BD (desulfonation buffer) was added and incubated at room temperature for 15 min. Again the spin columns were centrifuged for 1 min at 13,000 rpm. After two more wash steps (2 x 500 µL Buffer BW centrifuged for 1 min at 13,000 rpm), 250 µL ethanol (95-100%) was added to each spin column and centrifuged. The spin columns were then transferred to new 2 mL collection tubes and spun again to remove any residual liquid. Finally, the spin columns were transferred to 1.5 mL micro-centrifuge tubes and the DNA samples were eluted in 15 µL of Buffer EB (elution buffer), quantified using the Nano Drop 2000 (Thermo Fisher Scientific, USA) and stored at -20 °C.

2.4.2 PCR Amplification of Target CpG Region

The target region of interest was amplified using the PyroMark PCR Kit (Qiagen). Target DNA was combined with the PCR kit components, along with the specific PCR primers supplied with the PyroMark CpG Assay (Qiagen) for the specific gene promoter region of interest (**Table 2.12**). One primer of the set was biotinylated at its 5′ end, which was necessary for the preparation of single-stranded PCR product for subsequent Pyrosequencing. PCR was performed on a Techne TC-512 thermal cycler (Bibby Scientific Ltd, Staffordshire, UK). The thermal cycling conditions for the PyroMark PCR can be seen in **Table 2.13**.

Table 2.12 Reaction Components for PyroMark PCR.

Reaction Component	Volume per well (μL)
PyroMark PCR Master Mix	12.5
CoralLoad Concentrate	2.5
PyroMark PCR Primers	2.5
RNase-free water	5.5
Template DNA	2.0
Total Reaction Volume	25.0

Table 2.13 Thermal Cycling Conditions for PyroMark PCR.

Step	Time	Temp
Initial PCR activation step (activation of HotStar Taq DNA Polymerase)	15 min	95 °C
3-step cycling (45 cycles):		
Denaturation	30 sec	94 °C
Annealing	30 sec	56 °C
Extension	30 sec	72 °C
Final extension	10 min	72 °C

2.4.3 Gel Electrophoresis

Product amplification was checked prior to Pyrosequencing analysis by running samples (along with a 1 kb ladder) on a 1% agarose gel stained with SYBR Safe DNA gel stain (Invitrogen, Carlsbad CA, USA). Typical gel images showing amplicon purity are shown in Chapter 6.

2.4.4 Pyrosequencing

A Qiagen Q24 PyroMark (Qiagen) was used to sequence the target gene promoter regions. The Pyrosequencing method enabled quantification of DNA methylation status at CpG sites in the promoter regions of the selected genes, providing accurate and detailed profiles of DNA methylation patterns. The proportion of methylated DNA at each site was illustrated by the peak heights in the Pyrograms generated by the PyroMark Q24 software.

The first step of Pyrosequencing involved immobilising the biotinylated PCR product onto streptavidin-coated Sepharose beads (Streptavidin Sepharose High-Performance, GE Healthcare, UK). PCR product was combined with streptavidin-coated Sepharose beads, PyroMark Binding Buffer and high-purity water in a 24-well PCR plate sealed with cap strips. Reagent volumes required for this immobilisation step can be seen in **Table 2.14**.

Table 2.14 Reagents for PCR Product Immobilisation.

Reagent	Volume per well (µL)
PyroMark Binding Buffer	40
Streptavidin-coated Sepharose beads	2
High-purity water	28
PCR product	10
Total Volume	**80**

The PCR plate was then agitated constantly for 10 min at 1400 rpm. Immediately following this, the Sepharose beads were captured using the PyroMark Q24 Vacuum Workstation (Qiagen). After being processed through a series of wash steps using the PyroMark Vacuum Workstation (**Table 2.15**), the immobilised PCR product was released into a 24-well PyroMark Q24 plate containing 2.5 µL PyroMark Sequencing Primer (3 µM) and 22.5 µL PyroMark Annealing Buffer (Qiagen) and incubated at 80 °C for 2 min on a pre-heated hot plate. The PyroMark Annealing Buffer enhanced the hybridisation of the PyroMark Sequencing Primer to the single-stranded DNA template.

Table 2.15 Steps for Capturing Sepharose Beads using PyroMark Vacuum Workstation.

Step	Volume (mL)	Time (sec)
70% ethanol (wash)	50	5
PyroMark Denaturation Solution[a]	40	5
PyroMark Wash Buffer[b]	50	10

[a]The PyroMark Denaturation Solution separated the complimentary DNA strand from the biotin-tagged strand, which would be the template used for Pyrosequencing. [b]The PyroMark Wash Buffer ensured that only the single-stranded biotin-tagged DNA was immobilised to the PyroMark Vacuum Tool.

After cooling for 10 min, the PyroMark Q24 plate was loaded into the PyroMark Q24. PyroMark Q24 software was used to generate the nucleotide dispensation order and the volumes of nucleotides (dATPαS, dGTP, dCTP and dTTP), enzyme mix and substrate mix to add to the PyroMark Q24 Cartridge for Pyrosequencing.

The enzyme mix contained: DNA polymerase (incorporated dispensed nucleotides), ATP sulfurylase (converted pyrophosphate to ATP), luciferase (generated light signal) and apyrase (degraded ATP and unincorporated nucleotides). The substrate mix contained: adenosine 5′ phosphosulfate (APS) (for generation of ATP) and luciferin (substrate for luciferase). After each nucleotide incorporation event, pyrophosphate (PPi) is released and converted to ATP by ATP sulfurylase (**Figure 2.4**). The ATP then drives the conversion of luciferin to oxyluciferin by luciferase. The light generated by this reaction is detected as a peak in the Pyrogram (**Figure 2.4**) (Madi et al. 2012).

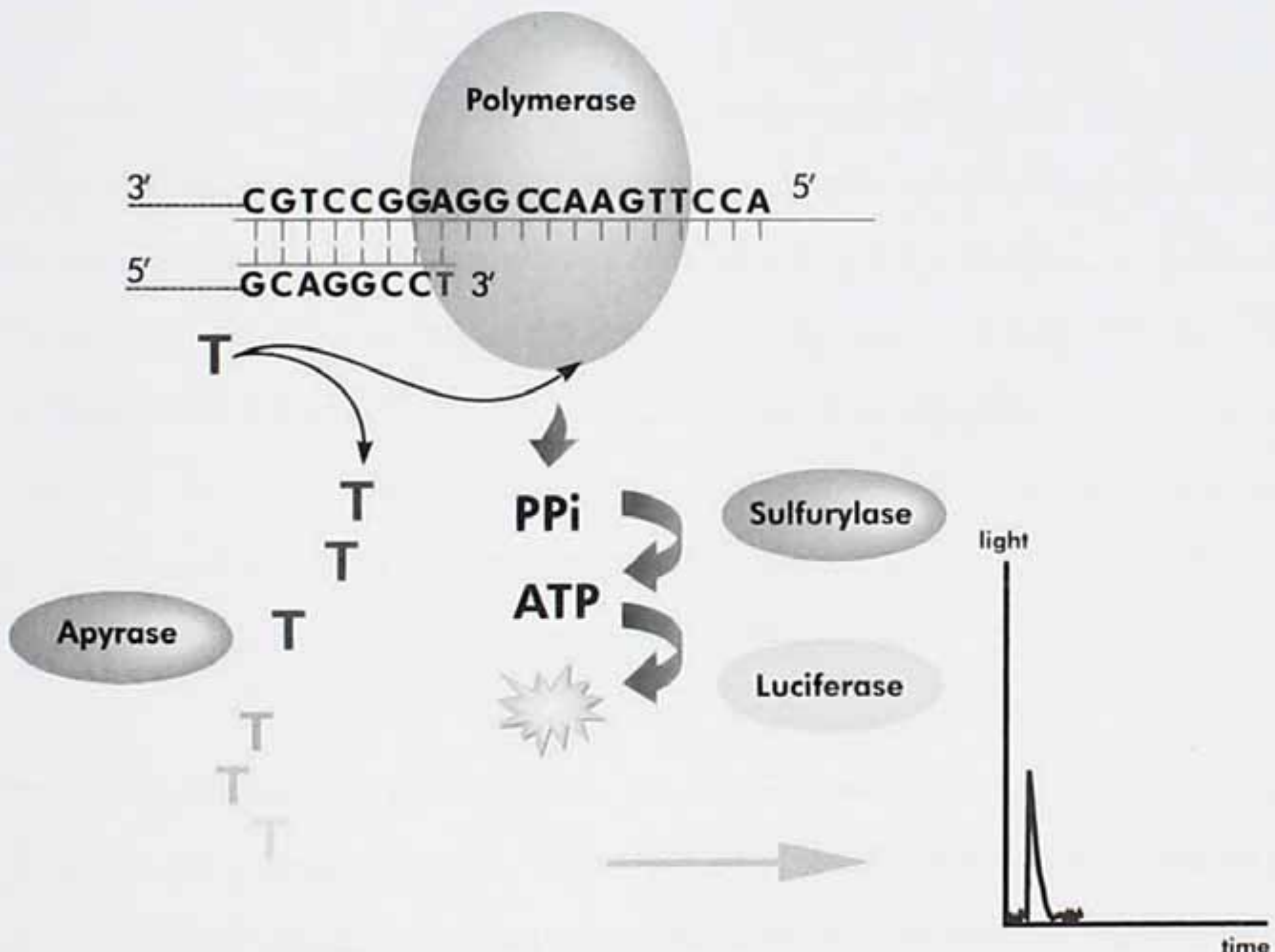


Figure 2.4 Overview of the Pyrosequencing cascade. Figure from PyroMark Gold Q24 Reagents Handbook. ©QIAGEN, all rights reserved.

2.5 STATISTICS

2.5.1 Power Calculations

The statistical power of our SNP analyses was calculated using Quanto version 1.2.4 (http://hydra.usc.edu/gxe) (Gauderman 2002). Assuming a 10% prevalence of ATP, an odds ratio (OR) of 2.2, an allele frequency of 0.45 and a recessive mode of inheritance, our analyses had 80% power to detect associations at the $p < 0.05$ significance level (based on 137 cases and 137 controls). However, when the individual cohorts were combined, our analyses had >90% power to detect associations at the $p < 0.05$ significance level (based on 330 cases and 330 controls).

For our discrete CNV data analyses, we used Sampsize version 0.6 (http://sampsize.sourceforge.net/) for a power calculation (Glaziou 2005). With an assumed CN exposure frequency of 20%, an OR of 2.2 and without a specified mode of inheritance, our cohort size (based on 130 cases and 130 controls) was sufficient for 80% power. When the cohorts were combined, the number of cases exceeded 300, resulting in >99% power to detect associations with the same parameters. Furthermore, lowering the CN exposure frequency to 10% in the model based on a combined cohort, power reached >90%.

For our continuous CNV data analyses, we used G*Power Software version 3.1.9.2 (http://www.gpower.hhu.de/en.html) (Faul et al. 2007). With an effect size of $d = 0.5$ (medium effect), a sample size of 90 cases and 90 controls resulted in > 90% power to detect associations at the $p < 0.05$ significance level. Based on the same parameters, our continuous CNV data analyses on a combined cohort (based on 300 cases and 300 controls) had > 99% power to detect associations.

2.5.2 Data Analysis

All data were analysed using IBM SPSS Statistics, version 20 (IBM Corp. Armonk, NY). Details about specific analyses and findings are discussed in the relevant chapters.

For the SNP genotype and CNV discrete CN data (<2 CN, =2 CN and >2 CN), a Pearson's chi-squared (χ^2) Test or Fisher's Exact Test was used to analyse for any differences between groups. When the cell count in categories did not exceed 5, the p value considered was based on Fisher's exact test, due to not meeting the assumptions for the χ^2 test. For the SNP data, Hardy-Weinberg equilibrium (HWE) was established using a HWE calculator (Michael H. Court, 2005-2008) accessed from www.tufts.edu/~mcourt01/Documents/Court lab - HW calculator.xls. with $p < 0.05$ considered to be a deviation from HWE. A χ^2 test was also used to determine any significant differences between the gender distributions of the groups.

Continuous CN data (non-integer values) were tested for normality using a Shapiro-Wilk/Kolmogorov-Smirnov Test. When considering normally distributed CN data sets, an independent samples t-Test was used to analyse for differences in the continuous CN data, based on the means. Levene's Test for Equality of Variance determined the appropriate p value to be considered. A Levene's Test value of > 0.05 meant equal variance was assumed, while a value < 0.05 meant equal variance was not assumed. If not normally distributed, the non-parametric Mann-Whitney U Test was used to test for differences in the continuous CN data. The CN data was principally analysed as discrete calls, although the continuous

data was also analysed for clarity, as it has been suggested that a degree of rounding error can occur during the process of discriminating CN into discrete integer groups (Park et al. 2012). However, on the whole, we found concordance in our analyses and therefore have confidence that the degree of error in CN calling was minor. Differences in DNA methylation (%) between the patellar tendinopathy and control groups were also analysed using an independent samples t-Test or Mann-Whitney U Test, with significance at $p < 0.05$.

A one-way analysis of variance (ANOVA) was used to determine any significant differences between the characteristics (age, height, weight, BMI) of the cases and controls, co-varied where appropriate. ANOVA was also used to determine any relationship between genotypes (SNP/CNV) and participant characteristics.

Comparisons included controls (CON) vs. the collective ATP cases (UK, SA and AUS cohorts combined), CON vs. tendinopathy (TEN) and CON vs. rupture (RUP). Cohorts (UK, SA and AUS) were also analysed separately and split for gender. Details of the specific analyses can be seen in the relevant chapters. For all analyses significance was accepted when $p < 0.05$. Analyses were also undertaken when a subset of the UK cohort were excluded for certain potentially relevant comorbidity, as described in the literature in Chapter 1 (1.4.1.3 Systemic Diseases) and indicated by the participant in the questionnaire upon recruitment (Appendix III). No changes in significance or non-significance from the original analyses were noted. Therefore, P values shown are based on the whole cohort.

SNPator is a web-based data analysis tool used for SNP data, freely available at http://www.snpator.org/public/new_login/index.php (Morcillo-Suarez et al.

2008). The SNP genotype data analyses were validated using this software resource. GraphPad Prism version 5.0 (GraphPad Software, San Diego, California, USA) was also used for certain data analysis and the production of graphs, which can be seen in the relevant chapters.

CHAPTER 3: COPY NUMBER VARIANTION WITHIN SELECTED COLLAGEN AND TNC GENES AS RISK FACTORS FOR ACHILLES TENDON PATHOLOGY

3.1 INTRODUCTION

While over 90% of the cellular content of tendon consists of tenocytes, the gross structure of tendon consists of an extracellular matrix (ECM) rich in collagens (Kader et al. 2002). The collagen fibrils are bundled together in an organised framework to form fascicles, providing high tensile strength (Kader et al. 2002). Tenascin-C is an ECM glycoprotein expressed particularly in musculoskeletal tissue, which is known to interact with other ECM components and cell surface receptors (Midwood et al. 2011). Variants within a number of collagen genes and within the *TNC* gene have been associated with Achilles tendon pathology (ATP) (Mokone et al. 2005; Mokone et al. 2006; September et al. 2008; September et al. 2009; Abrahams et al. 2013; Saunders et al. 2013). However, the effect of copy number variation (CNV) on the risk of ATP has never been investigated.

3.1.1 COL5A1

The *COL5A1* gene encodes the alpha 1 type V collagen protein, a minor structural protein constituent of tendons (Mokone et al. 2006). Although the function of the type V collagen is not fully defined, it is thought to be involved in the regulation of type I collagen fibril diameter and despite being a minor structural component, may have a more predominant role in developing connective tissues (September et al. 2009).

In 2006, Mokone et al. showed that the *Bst*UI restriction fragment length polymorphism (RFLP) (rs12722) within the 3'-untranslated region (UTR) of the

COL5A1 gene was associated with ATP in a South African Caucasian cohort. The allele frequencies of the *COL5A1 Bst*UI RFLP between the ATP and control (CON) groups were significantly different ($p = 0.006$) (Mokone et al. 2006). The *COL5A1 Bst*UI RFLP was also later found to associate with Achilles tendinopathy in an Australian Caucasian cohort, where the CC genotype was associated with a significantly decreased risk of developing Achilles tendinopathy (September et al. 2009). More recently in 2013, three single nucleotide polymorphisms (SNPs) within the functional *COL5A1* 3'-UTR region were independently associated with Achilles tendinopathy (Abrahams et al. 2013).

While certain mutations in collagen genes including the *COL5A1* gene are known to cause the inherited connective tissue disorder Ehlers-Danlos syndrome (EDS), other polymorphisms mentioned in the previous chapter have been shown to associate with the risk of ATP (**Figure 3.1**) (Collins and Posthumus 2011). As can be seen in the figure below, mutations in the *COL5A1* gene can result in abnormal collagen fibre formation compared to the typical wild-type collagen structure. While the CC genotype of the *COL5A1* rs12722 was associated with a decreased risk of Achilles tendinopathy, the TT genotype was found to be associated with Achilles tendinopathy and increased *COL5A1* mRNA stability (Mokone et al. 2006; Collins and Posthumus 2011; Laguette et al. 2011). It is proposed that the TT genotype results in an increase in type V collagen production where more is incorporated into the collagen fibril, resulting in collagen fibres containing smaller fibrils and a tendon with increased stiffness and reduced tensile strength (Collins and Posthumus 2011).

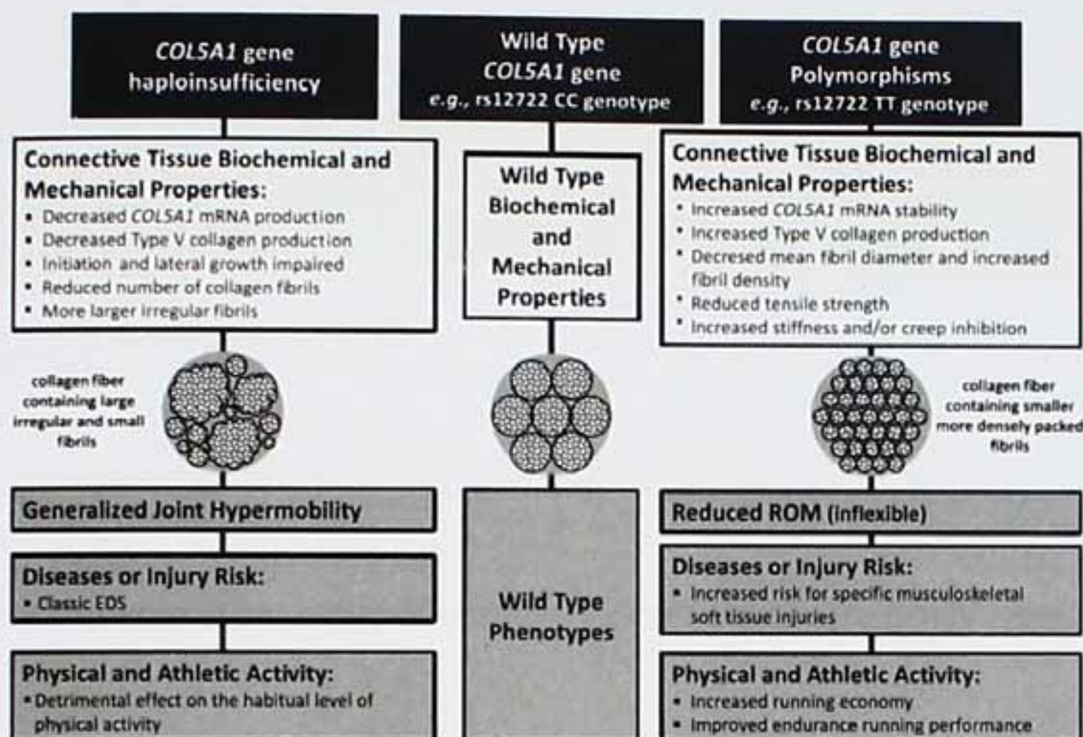
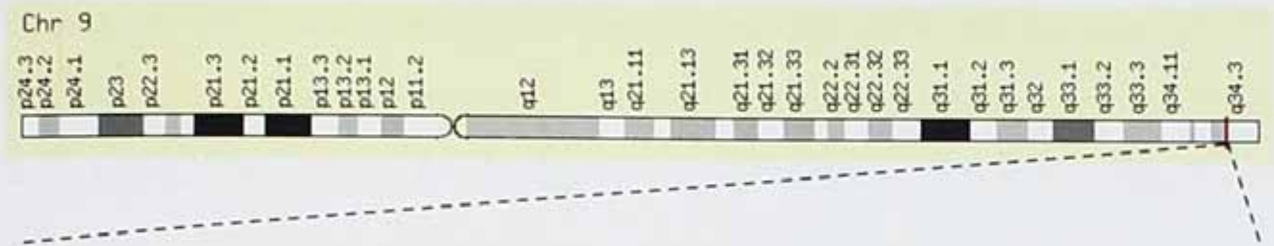


Figure 3.1 The relationship between *COL5A1* genotypes, disease risk and the associated effect on connective tissue and collagen fibre assembly (Collins and Posthumus 2011).

The *COL5A1* gene was selected for CNV investigation as a number of variants within this gene have previously been associated with ATP (Mokone et al. 2006; September et al. 2009; Abrahams et al. 2013). In addition, CNV was reported in the Database of Genomic Variants for this gene (http://dgv.tcag.ca/dgv/app/home). Two TaqMan Copy Number Assays (Hs00180523_cn and Hs02980647_cn) (Life Technologies) were selected within this region, spanning intron 4 - exon 5 and exon 14 - intron 14 respectively (**Figure 3.2**).

Chromosome 9 – *COL5A1*

a



b

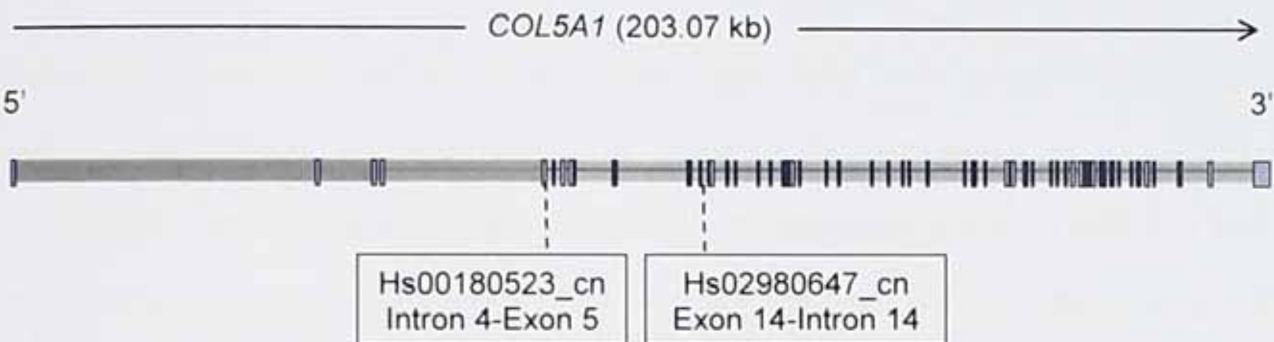


Figure 3.2 Location of the Hs00180523_cn and Hs02980647_cn Copy Number Assays within the *COL5A1* gene

(**a**) Chromosomal location of the *COL5A1* gene (Chr.9:137,533,620-137,736,686). (**b**) Size and exonic structure of the *COL5A1* gene with location of the TaqMan Copy Number Assays (Hs00180523_cn Chr.9:137,619,157 and Hs02980647_cn Chr.9:137,644,529). All locations referenced from NCBI build 37. Compiled from www.ensembl.org and www.lifetechnologies.com.

3.1.2 COL4A1

Another structural component of the tendon ECM is the alpha 1 type IV collagen, encoded by the *COL4A1* gene (Seals and Courtneidge 2003; Jelinsky et al. 2011). While COL4A1 protein is known to be cleaved by several MMPs including MMP2, 3, 9, 10 and 13, both ADAM10 and ADAM15 are also involved in the degradation and remodelling of COL4A1 (Somerville et al. 2003; Seals and Courtneidge 2003).

Mutations in the *COL4A1* gene have been associated with congenital muscular dystrophies (CMDs) in humans and mice (Labelle-Dumais et al. 2011). Specifically, certain mutations within the *COL4A1* gene are known to interfere with the folding and assembly of the COL4A1 protein via the disruption of splicing and the subsequent deletion of amino acids from its sequence (Labelle-Dumais et al. 2011). In human tendinopathy, altered expression levels of the *COL4A1* gene have also been identified, with a 2.7 fold increase in expression compared to normal tendon (Jelinsky et al. 2011).

The *COL4A1* gene was selected for CNV investigation as altered expression of this gene was reported between tendinopathy and normal tendon (Jelinsky et al. 2011), which it was hypothesised could be the result of CNV. Furthermore, CNV was reported in the Database of Genomic Variants for this gene. Two TaqMan Copy Number Assays (Hs00739915_cn and Hs02119212_cn) (Life Technologies) were selected within this region, spanning intron 22 - exon 22 and exon 21 respectively (**Figure 3.3**).

Chromosome 13 - *COL4A1*

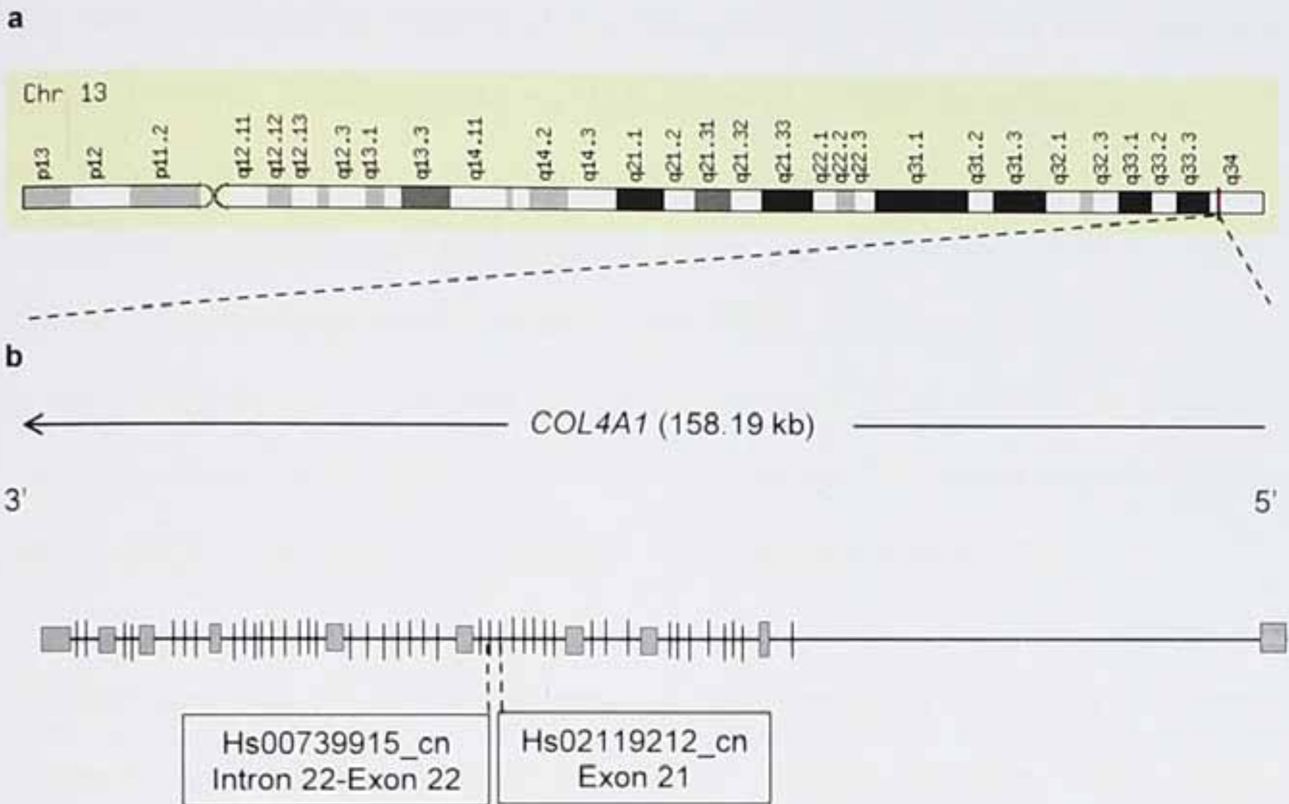


Figure 3.3 Location of the Hs00739915_cn and Hs02119212_cn Copy Number Assays within the *COL4A1* gene

(**a**) Chromosomal location of the *COL4A1* gene (Chr.13:110,801,318-110,959,496). (**b**) Size and exonic structure of the *COL4A1* gene with location of the TaqMan Copy Number Assays (Hs00739915_cn Chr.13:110,847,375 and Hs02119212_cn Chr.13:110,850,912). All locations referenced from NCBI build 37. Compiled from www.ensembl.org and www.lifetechnologies.com.

3.1.3 TNC

The *TNC* gene encodes the ECM glycoprotein tenascin-C, which is expressed in a range of tissues, including tendons (Mokone et al., 2005). In relation to tendon, tenascin-C plays a role in the regulation of cell-matrix interactions around the tendon, by binding to other ECM proteins and cell receptors (Mokone et al., 2005). Tenascin-C expression is low in healthy adult tissue, however, expression is shown to be up-regulated in response to injury (Midwood et al. 2011). In addition, mechanical loading on tendons is shown to regulate *TNC* gene expression in a dose-dependent manner (Järvinen et al. 1999; Saunders et al. 2013).

The *TNC* gene was one of the first genes to be investigated for its role in ATP and research has highlighted some interesting findings (Mokone et al. 2005; September et al. 2012; Saunders et al. 2013). Mokone et al. (2005) found that a guanine-thymine (GT) repeat polymorphism within intron 17 of the *TNC* gene was associated with ATP in a South African population, although the functional effect of this variant is unknown. Furthermore, Saunders et al. (2013) showed that certain SNPs were associated with Achilles tendinopathy in South African and Australian cohorts (Saunders et al. 2013). There was a significant allelic association between the rs2104772 and rs1330363 variants and Achilles tendinopathy (Saunders et al. 2013). Accordingly, two TaqMan Copy Number Assays (Hs06903309_cn and Hs00634176_cn) (Life Technologies) were selected within this region (**Figure 3.4**).

Chromosome 9 – *TNC*

a

b

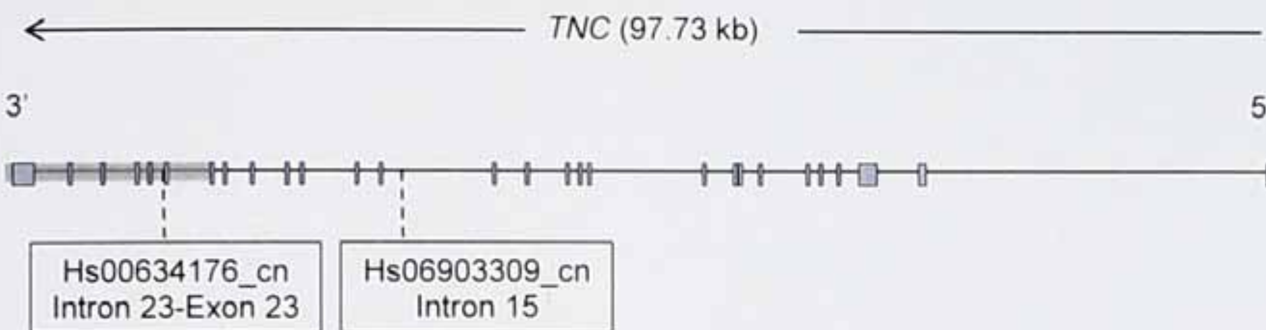


Figure 3.4 Location of the Hs00634176_cn and Hs06903309_cn Copy Number Assays within the *TNC* gene

(**a**) Chromosomal location of the *TNC* gene (Chr.9:117,782,806-117,880,536). (**b**) Size and exonic structure of the *TNC* gene with location of the TaqMan Copy Number Assays (Hs00634176_cn Chr.9:117,793,858 and Hs06903309_cn Chr.9:117,812,454). All locations referenced from NCBI build 37. Compiled from www.ensembl.org and www.lifetechnologies.com.

3.1.4 Chapter Aims

The specific aim of this chapter was to determine whether CNV within the candidate genes *COL5A1* (measured using the Hs00180523_cn and Hs02980647_cn Copy Number Assays), *COL4A1* (using the Hs00739915_cn and Hs02119212_cn Copy Number Assays) and *TNC* (using the Hs00634176_cn and Hs06903309_cn Copy Number Assays) was associated with ATP in a UK, South African and Australian Caucasian population.

3.2 METHODS

The methods below are specific to this chapter and are described in brief. An overview of the detailed methodology is detailed in Chapter 2.

3.2.1 Participants

A total of 290 (122 UK, 111 SA, 57 AUS) Caucasian participants diagnosed with ATP were recruited for this study. A total of 357 (131 UK, 95 SA, 131 AUS) physically active Caucasian control participants (asymptomatic and without history of ATP) were recruited as controls.

3.2.2 DNA Collection and Extraction

For the UK cohort, DNA was extracted from 2 mL of saliva collected using ORAGENE-DNA kits (OG-500) and DNA purification was carried out using the prepIT-L2P DNA extraction kit (DNA Genotek Inc., Ontario, Canada). For the South African cohort, DNA was extracted from 4.5 mL of venous blood, obtained by venipuncture of a forearm vein and collected into an ethylenediaminetetraacetic acid (EDTA) vacutainer tube. For the AUS cohort, DNA was extracted from approximately 4.5 mL of venous blood, using a sequence extraction technique (FlexiGene DNA Kit, Qiagen P/L, Valencia, California, USA). Full details of DNA collection and extraction are described in Chapter 2.

3.2.3 Copy Number Genotyping

Copy number genotyping was carried out using a StepOnePlus Real-Time PCR System (Applied Biosystems, Foster City, California, USA). TaqMan Copy Number Assay and TaqMan Copy Number Reference Assay RNase P were run simultaneously in a duplex real-time PCR reaction. Reactions were performed in

96-well plates using universal PCR cycling conditions. Each reaction comprised of 10 ng of purified genomic DNA, combined with a mastermix containing AmpliTaq Gold DNA Polymerase and probes and primers for both the target copy number and reference assay. Reactions were run in quadruplicate for each DNA sample. NTCs were included within each plate to check for quality control and to check for contamination. The data were exported to CopyCaller Software version 2.0 (Applied Biosystems, Foster City, California, USA) for calculation of copy number.

3.2.4 Data Analysis

Data were analysed using IBM SPSS Statistics version 20 (IBM Corp. Armonk, NY). A one-way analysis of variance (ANOVA) was used to test for any significant differences between the characteristics (age, height, weight, BMI) of the CON and ATP groups, as well as to test for differences between the characteristics and copy number. A Pearson's chi-squared (χ^2) test was used to determine any significant differences between the gender of the CON and ATP groups. A χ^2 or Fisher's Exact Test was used to determine any significant differences in discrete copy number (<2, =2, >2 copies) between the CON and ATP groups, while differences in continuous copy number were analysed with a non-parametric Mann-Whitney U Test. The ATP group was also split to analyse the TEN and RUP groups separately. Data were analysed by gender and reported if significant. Significance was accepted at $p < 0.05$ for all analyses.

3.3 RESULTS

When the characteristics of the CON and ATP cohorts were compared, some significant differences were identified. Weight and BMI were significantly different between the CON and ATP group in the SA cohort and the AUS cohort (both $p < 0.05$). For the SA cohort, the mean weight (kg) of the CON and ATP participants was 72.1 ± 12.2 and 81.1 ± 15.3 respectively. For the AUS cohort, the mean weight of the CON and ATP participants was 73.4 ± 14.0 and 80.4 ± 14.9 respectively. There was a significant difference in age between the CON and ATP group in both the UK (CON, 41.3 ± 11.3; ATP, 46.3 ± 13.6; $p = 0.001$), and the SA (CON, 36.8 ± 9.9; ATP, 40.5 ± 12.9; $p = 0.025$) cohorts. Furthermore, gender distribution (% males) was significantly different between the CON and ATP groups in the AUS cohort ($p < 0.05$). Full details of participant characteristics can be seen in Chapter 2.

Copy number was calculated using CopyCaller Software version 2.0 (Applied Biosystems, Foster City, California, USA) based on the calling algorithm detailed in the methods in Chapter 2. A typical CopyCaller output graph can be seen in **Figure 3.5**.

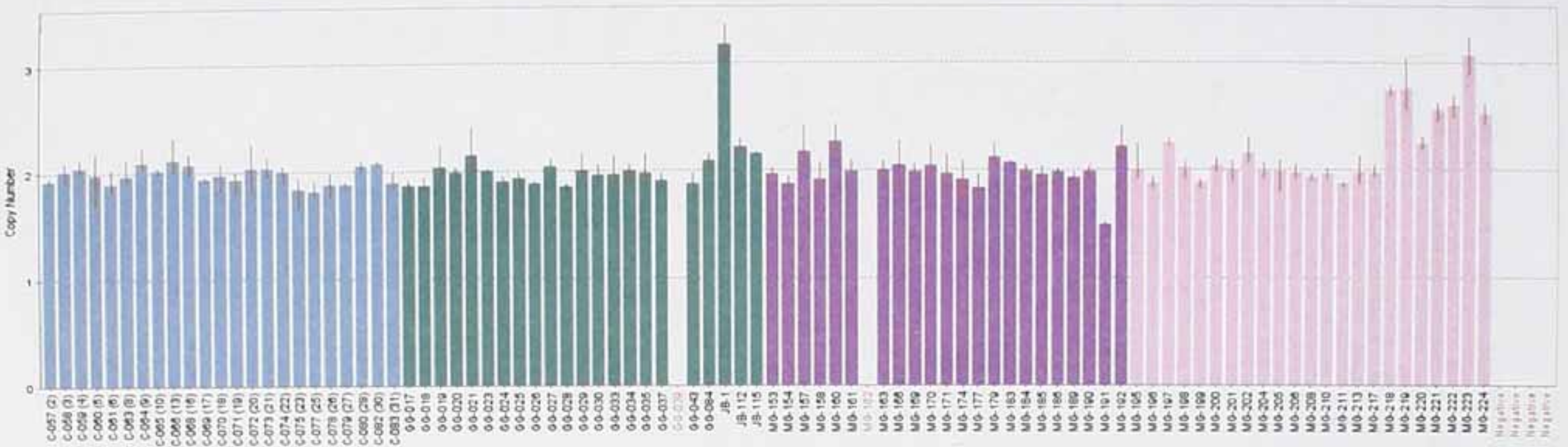


Figure 3.5 Typical CN profiles generated from the CopyCaller Software. Each bar represents a sample (mean of n=4 replicates) as labelled on the x-axis. The height of bars on the y-axis represents the CN calculated by the software, with error bars showing the standard deviation (SD) between replicates. The figure shown is an example of copy numbers calculated using the TaqMan Copy Number Assay Hs06903309_cn within the *TNC* gene.

Data were analysed as both continuous and discrete CN to check calling accuracy. **Figure 3.6** shows the relationship between the continuous CN values and the discrete CN calls. Clusters can be seen, showing how the CopyCaller Software calls each continuous CN value to the nearest discrete CN.

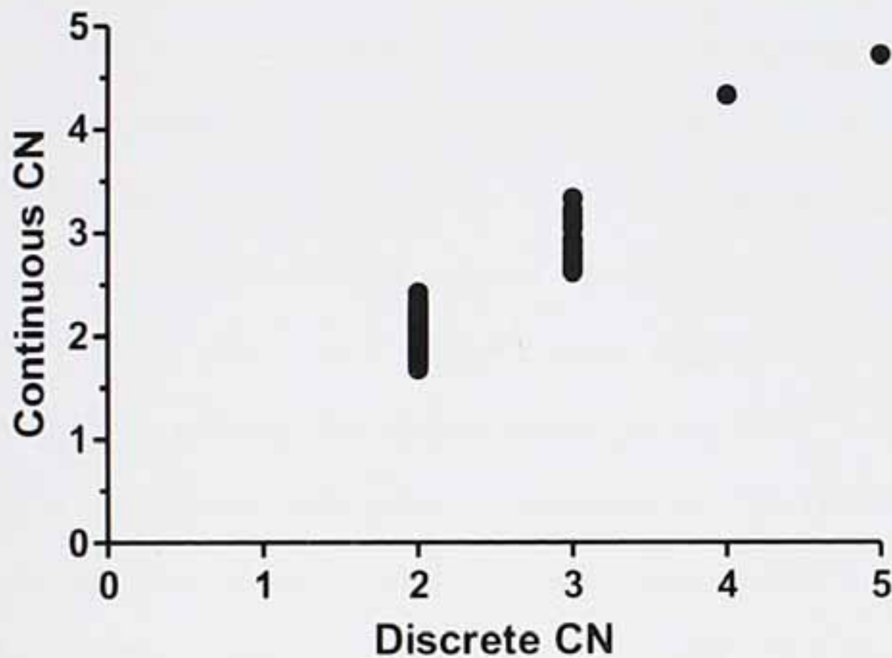


Figure 3.6 Scatter plot showing the raw continuous CN values on the y-axis, plotted against the discrete integer CN values on the x-axis. Graph shown was for the Hs06903309_cn assay within the *TNC* gene with a combined cohort (UK, SA and AUS).

3.3.1 *COL5A1* Copy Number Frequencies

There were no significant differences in discrete copy number for the Hs00180523_cn assay within the *COL5A1* gene for the combined CON vs. ATP (p = 0.131), CON vs. TEN (p = 0.195) or CON vs. RUP (p = 0.460) groups (**Table 3.1a**). When the SA and AUS cohorts where analysed separately, there were no differences for the SA cohort between the CON vs. ATP (p = 1.000), CON vs. TEN (p = 0.702) or CON vs. RUP (p = 0.513) groups. For the AUS cohort, there was a significant difference in copy number between the CON and ATP groups (p = 0.012). Although the total number of AUS ATP participants is relatively small (n = 44), the distribution of copy number was significantly different between the CON (CN<2, n=4; CN=2, n=65; CN>2, n=0) and ATP (CN<2, n=0; CN=2, n=40; CN>2, n=4) groups. This is also shown for the CON vs. TEN analysis, however, this is due to the AUS cohort comprising of only tendinopathy cases and no ruptures. Nevertheless, when the continuous copy number data were also analysed with a Mann-Whitney U Test, there was no significant difference between the AUS CON and ATP groups (p value not shown). This was also the case for all other analyses for the continuous CN data (p values not shown).

For the Hs02980647_cn assay within the *COL5A1* gene, there were no differences in discrete CN for the UK cohort between the CON vs. ATP (p = 1.000), CON vs. TEN (p = 1.000) or CON vs. RUP (p = 1.000) groups (**Table 3.1b**).

The QPCR efficiencies of the *COL5A1* assays both exceed 97%. Efficiency graphs for the Hs00180523_cn and Hs02980647_cn assays can be seen in Appendix V (**Figure V.5** and **Figure V.6** respectively).

Table 3.1a Copy number distribution within the *COL5A1* gene in the cases (ATP/TEN/RUP) and CON groups using the Hs00180523_cn assay.

Cohort		Mean CN	< 2 CN	= 2 CN	> 2 CN	n	p-value
	COL5A1						
	Hs00180523_cn						
COMBINED	CON	1.94 ± 0.27	4	150	1	155	-
	ATP	2.00 ± 0.25	1	146	5	152	0.131
	TEN	1.99 ± 0.25	1	110	4	115	0.195
	RUP	2.02 ± 0.25	0	36	1	37	0.460
SA	CON	1.97 ± 0.18	0	85	1	86	-
	ATP	1.99 ± 0.21	1	106	1	108	1.000
	TEN	1.97 ± 0.18	1	70	0	71	0.702
	RUP	2.02 ± 0.25	0	36	1	37	0.513
AUS	CON	1.91 ± 0.35	4	65	0	69	-
	ATP	2.03 ± 0.34	0	40	4	44	**0.012**
	TEN	2.03 ± 0.34	0	40	4	44	**0.012**
	RUP	-	0	0	0	0	-

Differences in discrete copy number between the CON vs. ATP, CON vs. TEN and CON vs. RUP groups were analysed with a χ^2 or Fisher's Exact Test. Mean CN (± SD) represents the continuous CN data, which were analysed with a Mann-Whitney U Test.

Table 3.1b Copy number distribution within the *COL5A1* gene in the cases (ATP/TEN/RUP) and CON groups using the Hs02980647_cn assay.

Cohort		Mean CN	< 2 CN	= 2 CN	> 2 CN	n	p-value
	COL5A1						
	Hs02980647_cn						
UK	CON	2.00 ± 0.12	0	129	1	130	-
	ATP	1.99 ± 0.08	0	118	0	118	1.000
	TEN	1.98 ± 0.08	0	84	0	84	1.000
	RUP	2.00 ± 0.08	0	34	0	34	1.000

Differences in discrete copy number between the CON vs. ATP, CON vs. TEN and CON vs. RUP groups were analysed with a Fisher's Exact Test. Mean CN (± SD) represents the continuous CN data, which were analysed with a Mann-Whitney U Test.

3.3.2 *COL4A1* Copy Number Frequencies

When the combined cohort (UK and SA) was analysed for differences in discrete CN for the Hs00739915_cn assay, there was a significant differences between both the CON vs. ATP (p = 0.014) and the CON vs. TEN (p = 0.033) groups (**Table 3.2a**). Specifically, the distribution of copy number was significantly different between the combined CON (CN<2, n=5; CN=2, n=208; CN>2, n=2) and ATP (CN<2, n=0; CN=2, n=211; CN>2, n=7) groups. The distribution of copy number was also significantly different between the combined CON and the TEN (CN<2, n=0; CN=2, n=146; CN>2, n=6) groups. In both analyses, the CON participants had a higher frequency of participants with CN<2, while the ATP and TEN cases had a higher frequency of participants with CN>2. However when the continuous CN data was analysed with a Mann-Whitney U Test, there were no significant differences observed. In addition, there was no significant difference in CN for the combined CON vs. RUP groups (p = 0.425).

When the UK and SA cohorts were analysed separately, there were no significant differences for the discrete CN data between the CON vs. ATP (UK, p = 0.061; SA, p = 0.180), the CON vs. TEN (UK, p = 0.083; SA, p = 0.075), nor the CON vs. RUP (UK, p = 0.584; SA, p = 1.000) groups. There were also no significant differences between the continuous CN data (p values not shown).

For the Hs02119212_cn assay within the *COL4A1* gene, there were no differences in discrete CN for the combined cohort between the CON vs. ATP (p = 0.615), CON vs. TEN (p = 0.512) or CON vs. RUP (p = 0.575) groups (**Table 3.2b**). When the UK and SA cohorts were analysed separately, there were no significant differences between the CON vs. ATP (UK, p = 1.000; SA, p = 1.000), CON vs.

TEN (UK, p = 1.000; SA, p = 1.000) nor the CON vs. RUP (UK, p = 1.000; SA, p = 0.506) groups. When the continuous CN data were analysed, there were no significant differences for any comparisons (combined or separate cohorts).

The QPCR efficiencies of the *COL4A1* assays both exceed 94%. Efficiency graphs for the Hs00739915_cn and Hs02119212_cn assays can be seen in Appendix V (**Figure V.3** and **Figure V.4** respectively).

Table 3.2a Copy number distribution within the *COL4A1* gene in the cases (ATP/TEN/RUP) and CON groups using the Hs00739915_cn assay.

Cohort		Mean CN	< 2 CN	= 2 CN	> 2 CN	n	p-value
	COL4A1						
	Hs00739915_cn						
COMBINED	CON	1.96 ± 0.29	5	208	2	215	-
	ATP	2.01 ± 0.31	0	211	7	218	**0.014**
	TEN	2.00 ± 0.27	0	146	6	152	**0.033**
	RUP	2.03 ± 0.38	0	65	1	66	0.425
UK	CON	1.93 ± 0.31	5	121	0	126	-
	ATP	1.96 ± 0.13	0	115	1	115	0.061
	TEN	1.95 ± 0.14	0	85	0	85	0.083
	RUP	2.00 ± 0.12	0	30	0	30	0.584
SA	CON	2.01 ± 0.24	0	87	2	89	-
	ATP	2.06 ± 0.42	0	96	7	103	0.180
	TEN	2.06 ± 0.36	0	61	6	67	0.075
	RUP	2.06 ± 0.24	0	35	1	36	1.000

Differences in discrete copy number between the CON vs. ATP, CON vs. TEN and CON vs. RUP groups were analysed with a Fisher's Exact Test. Mean CN (± SD) represents the continuous CN data, which were analysed with a Mann-Whitney U Test.

Table 3.2b Copy number distribution within the *COL4A1* gene in the cases (ATP/TEN/RUP) and CON groups using the Hs02119212_cn assay.

Cohort		Mean CN	< 2 CN	= 2 CN	> 2 CN	n	p-value
	COL4A1						
	Hs02119212_cn						
COMBINED	CON	1.99 ± 0.13	0	214	2	216	-
	ATP	1.99 ± 0.14	0	226	1	227	0.615
	TEN	1.99 ± 0.12	0	156	0	156	0.512
	RUP	1.98 ± 0.17	0	70	1	71	0.575
UK	CON	1.98 ± 0.12	0	127	1	128	-
	ATP	1.99 ± 0.06	0	120	0	120	1.000
	TEN	1.99 ± 0.06	0	86	0	86	1.000
	RUP	1.98 ± 0.07	0	34	0	34	1.000
SA	CON	2.00 ± 0.15	0	87	1	88	-
	ATP	1.99 ± 0.19	0	106	1	107	1.000
	TEN	1.99 ± 0.17	0	70	0	70	1.000
	RUP	1.98 ± 0.23	0	36	1	37	0.506

Differences in discrete copy number between the CON vs. ATP, CON vs. TEN and CON vs. RUP groups were analysed with a Fisher's Exact Test. Mean CN (± SD) represents the continuous CN data, which were analysed with a Mann-Whitney U Test.

3.3.3 *TNC* Copy Number Frequencies

There were no significant differences in copy number for either loci investigated within the *TNC* gene (Hs00634176_cn and Hs06903309_cn) between the combined CON and ATP groups (Hs00634176_cn assay, p = 0.829; Hs06903309_cn assay, p = 0.676), nor between the combined CON and TEN (Hs00634176_cn assay, p = 0.641; Hs06903309_cn assay, p = 0.275) or CON and RUP (Hs00634176_cn assay, p = 1.000; Hs06903309_cn assay, p = 0.706) groups (**Table 3.3a** and **3.3b**). There were also no differences for the UK, SA and AUS cohorts when analysed separately. Furthermore, when the continuous CN data were analysed, there were no significant differences between the combined or separate cohorts.

The QPCR efficiency of the *TNC* assays averaged >90%. The efficiency graphs for the Hs00634176_cn and Hs06903309_cn assays can be seen in Appendix V (**Figure V.9** and **Figure V.10** respectively).

Table 3.3a Copy number distribution within the *TNC* gene in the cases (ATP/TEN/RUP) and CON groups using the Hs00634176_cn assay.

Cohort		Mean CN	< 2 CN	= 2 CN	> 2 CN	n	p-value
	TNC						
	Hs00634176_cn						
COMBINED	CON	1.98 ± 0.12	2	302	1	305	-
	ATP	1.96 ± 0.15	3	274	0	277	0.829
	TEN	1.96 ± 0.17	3	203	0	206	0.641
	RUP	1.98 ± 0.08	0	71	0	71	1.000
UK	CON	2.00 ± 0.12	0	127	1	128	-
	ATP	1.97 ± 0.08	0	121	0	121	1.000
	TEN	1.97 ± 0.09	0	87	0	87	1.000
	RUP	1.98 ± 0.08	0	34	0	34	1.000
SA	CON	1.97 ± 0.09	0	88	0	88	-
	ATP	1.97 ± 0.10	0	107	0	107	-
	TEN	1.97 ± 0.11	0	70	0	70	-
	RUP	1.98 ± 0.08	0	37	0	37	-
AUS	CON	1.95 ± 0.15	2	87	0	89	-
	ATP	1.91 ± 0.30	3	46	0	49	0.347
	TEN	1.91 ± 0.30	3	46	0	49	0.347
	RUP	-	0	0	0	0	-

Differences in discrete copy number between the CON vs. ATP, CON vs. TEN and CON vs. RUP groups were analysed with a Fisher's Exact Test. Mean CN (± SD) represents the continuous CN data, which were analysed with a Mann-Whitney U Test.

Table 3.3b Copy number distribution within the *TNC* gene in the cases (ATP/TEN/RUP) and CON groups using the Hs06903309_cn assay.

Cohort		Mean CN	< 2 CN	= 2 CN	> 2 CN	n	p-value
	TNC						
	Hs06903309_cn						
COMBINED	CON	2.03 ± 0.25	0	325	12	337	-
	ATP	2.04 ± 0.25	0	254	12	266	0.676
	TEN	2.05 ± 0.27	0	185	11	196	0.275
	RUP	2.01 ± 0.18	0	69	1	70	0.706
UK	CON	2.00 ± 0.11	0	128	1	129	-
	ATP	2.01 ± 0.14	0	120	2	122	0.613
	TEN	2.02 ± 0.16	0	86	2	88	0.567
	RUP	1.98 ± 0.10	0	34	0	34	1.000
SA	CON	2.02 ± 0.19	0	86	2	88	-
	ATP	2.04 ± 0.23	0	98	5	103	0.455
	TEN	2.04 ± 0.23	0	63	4	67	0.404
	RUP	2.03 ± 0.23	0	35	1	36	1.000
AUS	CON	2.08 ± 0.37	0	111	9	120	-
	ATP	2.14 ± 0.46	0	36	5	41	0.350
	TEN	2.14 ± 0.46	0	36	5	41	0.350
	RUP	-	0	0	0	0	-

Differences in discrete copy number between the CON vs. ATP, CON vs. TEN and CON vs. RUP groups were analysed with a Fisher's Exact Test. Mean CN (± SD) represents the continuous CN data, which were analysed with a Mann-Whitney U Test.

3.4 CHAPTER SPECIFIC CONCLUSIONS

For the Hs00180523_cn assay, that measures CNV within the *COL5A1* gene, there was a significant difference in discrete CN between the AUS CON and ATP groups (p = 0.012) (**Table 3.1a**). The distribution of CN was significantly different between the AUS CON (CN<2, n=4; CN=2, n=65; CN>2, n=0) and AUS ATP (CN<2, n=0; CN=2, n=40; CN>2, n=4) groups. Specifically, the CON group had a higher frequency of participants with CN<2 and the ATP group had a higher frequency of participants with CN>2.

We also found significant differences in discrete CN between both the combined CON vs. ATP (p = 0.014) and the combined CON vs. TEN (p = 0.033) groups for the *COL4A1* Hs00739915_cn assay (**Table 3.2a**). In both cases, the CON group had a higher frequency of participants with CN<2, while the ATP and TEN cases had a higher frequency of participants with CN>2.

When the continuous CN data for the *COL5A1* Hs00180523_cn and *COL4A1* Hs00739915_cn assay were analysed, there were no significant differences reported (data not shown). This highlights the importance of analysing data as both discrete and continuous copy number, as potential associations may be missed if only one type of analysis is used. Indeed, the findings from the continuous and discrete CN analyses were not always concordant and it has been suggested that rounding errors can occur during the process of discriminating CN into discrete integer groups (Park et al. 2012).

The discrete CN data for the *COL5A1* Hs00180523_cn and *COL4A1* Hs00739915_cn assays showed that the CON groups had a higher frequency of

participants with CN<2, when compared to the ATP groups. Although the data were not concordant with the continuous CN data, the tentative differences in discrete CN reported in the *COL5A1* and *COL4A1* genes could suggest that lower copy number at these loci may be protective against ATP.

CHAPTER 4: COPY NUMBER VARIATION WITHIN *MMP* AND *ADAM* GENES AS RISK FACTORS FOR ACHILLES TENDON PATHOLOGY

4.1 INTRODUCTION

The matrix metalloproteinases (MMPs) are a large family of metal-dependent proteases involved in the remodelling of the extracellular matrix (ECM) through the degradation of ECM proteins (Birkedal-Hansen et al. 1993; Somerville et al. 2003). Over 20 specific enzymes have been identified in humans and despite differences in structure and function, many share substrate specificities and can cleave a range of collagenous and non-collagenous ECM proteins (Somerville et al. 2003). Although remodelling is vital for ECM homeostasis, the activity of the MMPs must also be controlled (Visse and Nagase 2003). This role is carried out by specific enzymes called tissue inhibitors of metalloproteinases (TIMPs) (Visse and Nagase 2003). The balance of MMPs and TIMPs is important as an imbalance could lead to pathology (Schulze-Tanzil et al. 2011). In addition to the MMPs, the a disintegrin and metalloproteinases (ADAMs) are important enzymes with roles in tendon remodelling that are also regulated by the action of TIMPs (El Khoury et al. 2013). The ADAMs belong to the zinc protease superfamily, as do the MMPs, and both are in the same subgroup (metzincin) based on the primary structure of their catalytic sites (Seals and Courtneidge 2003).

4.1.1 MMP3

MMP3 (stromelysin 1) is one member of the MMP family, encoded by the *MMP3* gene and located on chromosome 11 (Visse and Nagase 2003). MMP3 substrates include several collagens (types II, IV, V, IX, X), the proteoglycans aggrecan and decorin, fibronectin and laminin (Somerville et al. 2003; Raleigh et al. 2009). This

is not an exclusive list, as MMP3 also degrades a number of other collagenous and non-collagenous ECM proteins, as well as activating proMMPs into their functional forms (Somerville et al. 2003; Visse and Nagase 2003). For example, MMP3 is required for the activation of proMMP1 to MMP1 (Visse and Nagase 2003). MMP1 (collagenase 1) can cleave collagens I, II and III, and as type I collagen is the predominant collagen in tendon (Collins and Raleigh 2009), it is clear to see how variation within the *MMP3* gene can have a downstream effect on collagen turnover and the structure of the tendon itself. For example, if levels of the gene were over expressed and more MMP3 protein produced, there could be an increase in collagen turnover. This would disrupt the homeostatic balance, impairing tendon structure and potentially leading to pathology.

In 2009, Raleigh et al. published research showing that three SNPs within the *MMP3* gene were associated with the risk of Achilles tendinopathy in a Caucasian population. Independent associations were found with the GG genotype of rs679620, the CC genotype of rs591058 and the AA genotype of rs650108, which were all overrepresented in the Achilles tendinopathy group compared to controls (Raleigh et al. 2009). This study genotyped controls and individuals with either Achilles tendinopathy or Achilles tendon rupture (partial or complete). When the tendinopathy and rupture groups were grouped and compared collectively against controls, there were no significant differences in genotype or allele distributions. However, when considering the pathologies separately, the authors found that there were significant differences in both genotype ($p = 0.031$) and allele ($p = 0.037$) distributions between the control and Achilles tendinopathy group for the *MMP3* rs679620 SNP (Raleigh et al. 2009).

The rational for selection of the *MMP3* gene for investigation of genomic CNV was based on the fact that SNPs within this gene have previously been associated with Achilles tendinopathy (Raleigh et al. 2009). In addition, altered expression levels of the *MMP3* gene have been observed, with lower levels of MMP3 RNA seen in degenerate Achilles tendon tissue compared to control (Jones et al. 2006; Ireland et al. 2001). There was only one recorded CNV reported within this gene (esv2663139) in the Database of Genomic Variants (http://dgv.tcag.ca/dgv/app/home). Two TaqMan Copy Number Assays were used to genotype two loci within the *MMP3* gene and details of the location of these assays are shown in **Figure 4.1**.

Chromosome 11 – *MMP3*

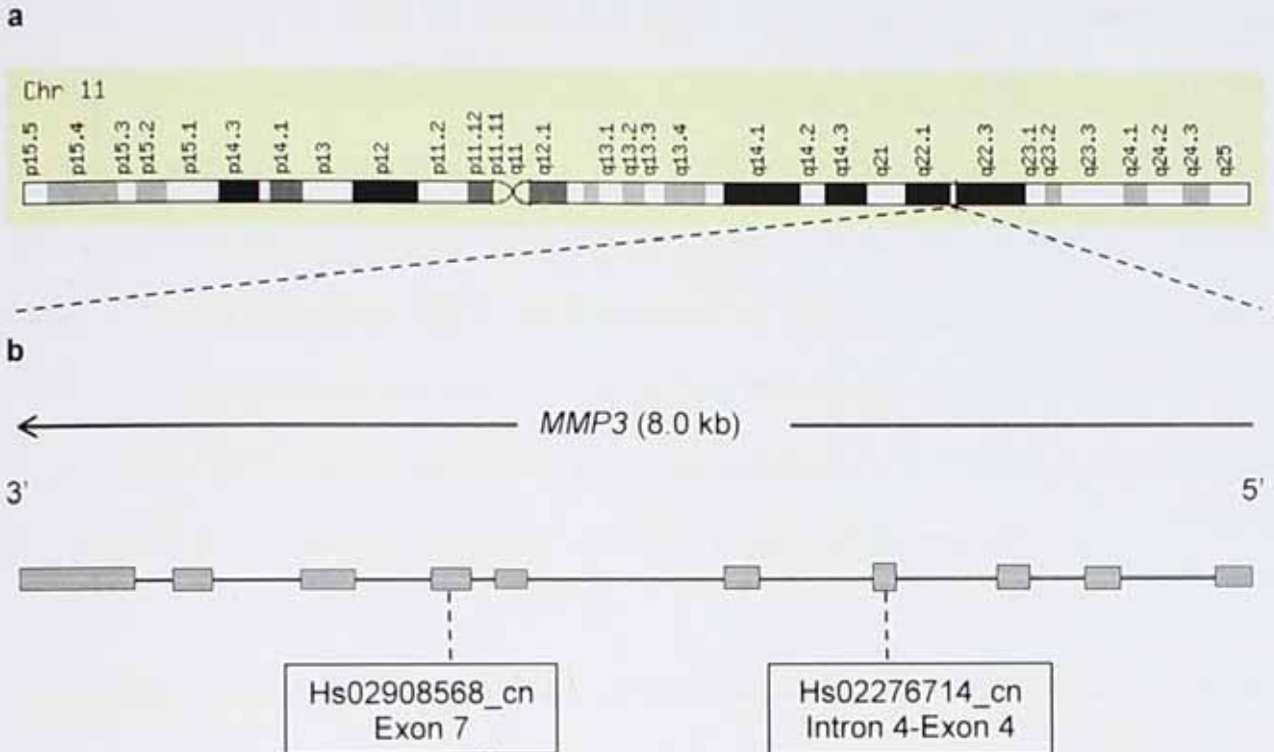


Figure 4.1 Location of the Hs02908568_cn and Hs02276714_cn Copy Number Assays within the *MMP3* gene

(**a**) Chromosomal location of the *MMP3* gene (Chr.11:102,706,532-102,714,534). (**b**) Size and exonic structure of the *MMP3* gene with location of the TaqMan Copy Number Assays (Hs02908568_cn Chr.11:102,709,884 and Hs02276714_cn Chr.11:102,712,920). All locations referenced from NCBI build 37. Compiled from www.ensembl.org and www.lifetechnologies.com.

4.1.2 MMP23B

Initially identified in 1998 by Gururajan et al. and originally named *MMP22*, the *MMP23B* gene is located on chromosome 1 (Gururajan et al. 1998). The *MMP23B* gene is located in a duplicated region of the chromosome, along with the *MMP23A* gene, which is almost identical in nucleotide sequence and was previously named *MMP21* (Somerville et al. 2003; Gururajan et al. 1998). Also known as cysteine array MMPs (CA-MMPs), the two closely related MMP23 enzymes differ in structure to the other MMP family members, with the presence of a cysteine-rich domain followed by an immunoglobulin like domain (Visse and Nagase 2003).

MMP23A/B are known to degrade the collagenous substrate gelatin (Somerville et al. 2003) and are of interest in human tendinopathy as changes in gene expression have been observed in painful compared with normal Achilles tendon ($p = 0.001$), as well as painful compared with ruptured Achilles tendon ($p = 0.002$) (Jones et al. 2006). Specifically, higher levels of MMP23 were detected in painful tendon, compared with normal tendon (Jones et al. 2006; Schulze-Tanzil et al. 2011). Based on the observed changes seen in ATP, CNV within the *MMP23B* gene was investigated to see if this form of genetic variation may have a role in the pathology. Furthermore, there has been CNV reported within the location of the MMP23B gene in the Database of Genomic Variants. One TaqMan Copy Number Assays was used to genotype a region within exon 1 of the *MMP23B* gene. Details of the location of this assay are shown in **Figure 4.2**.

Chromosome 1 - *MMP23B*

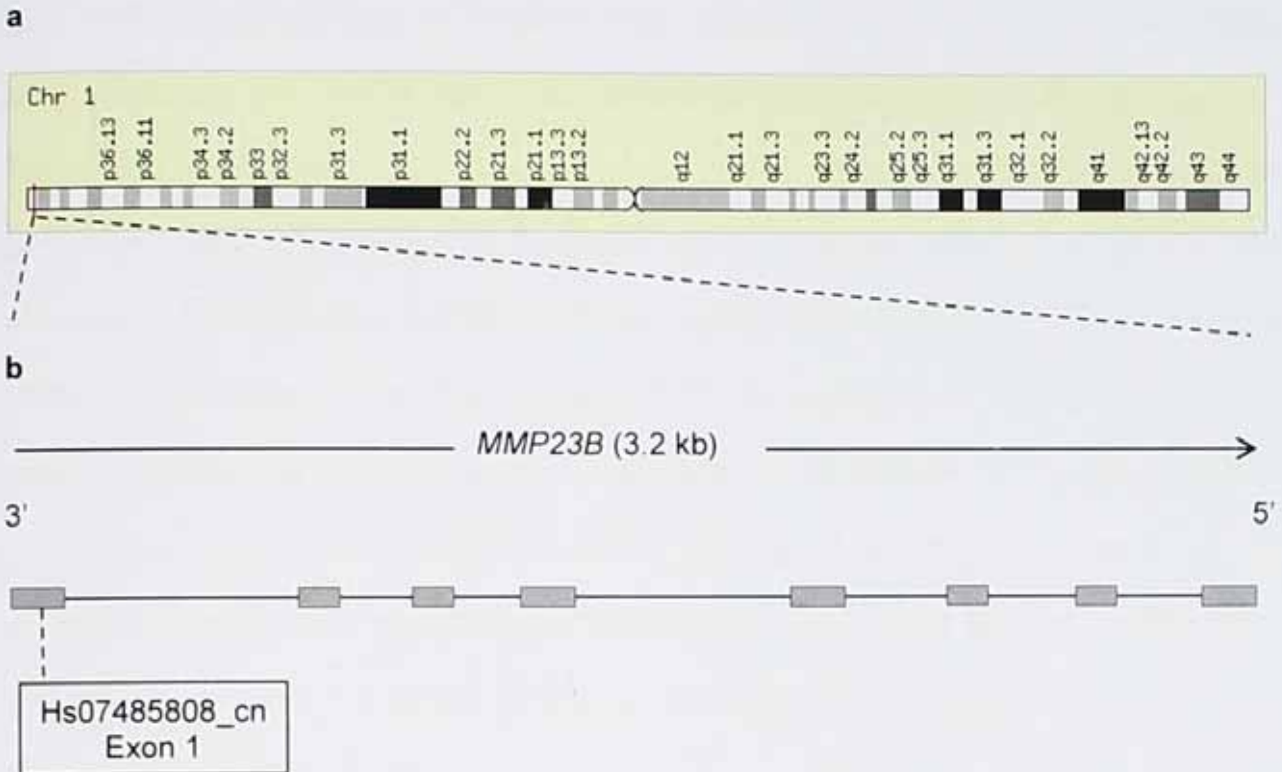


Figure 4.2 Location of the Hs07485808_cn Copy Number Assay within the *MMP23B* gene

(**a**) Chromosomal location of the *MMP23B* gene (Chr.1:1,567,474-1,570,639). (**b**) Size and exonic structure of the *MMP23B* gene with location of the TaqMan Copy Number Assay (Chr.1:1,567,625). All locations referenced from NCBI build 37. Compiled from www.ensembl.org and www.lifetechnologies.com.

4.1.3 ADAM8

The ADAMs are a family of multi-domain transmembrane proteins, that share similarities to the MMPs with the presence of a metalloproteinase domain (Primakoff and Myles 2000; Xu and Murrell 2008). There are 19 identified ADAM genes and although the specific functions of many of the ADAM gene products is unknown, the function of each of the protein domains has been broadly distinguished (Seals and Courtneidge 2003; Primakoff and Myles 2000). The catalytic activity of the metalloproteinase site of the ADAMs is regulated by the prodomain, which while intact, renders the metalloproteinase domain of the protein inactive (Seals and Courtneidge 2003). Cleavage of the prodomain from the protein enables the ADAM protein to be active in its mature form (Seals and Courtneidge 2003).

The ADAMs are known to have numerous functions throughout the body, with roles in cell migration, muscle development and fertilisation being just a few examples (Seals and Courtneidge 2003). Generally, the metalloproteinase domain has protease activity while the disintegrin and cysteine-rich domains have adhesion activity, enabling the enzyme to bind substrates (Primakoff and Myles 2000). The EGF-like domain of the enzyme is understood to be involved in membrane fusion and the phosphorylated cytoplasmic tail is thought to regulate the activity of other ADAMs (Primakoff and Myles 2000).

In relation to Achilles tendinopathy, elevated expression of the *ADAM8* gene has been identified in ruptured Achilles tendon compared to controls (Jones et al. 2006). A 19-fold difference in expression was shown ($p = 0.002$). In addition, Jones et al. (2006) also reported a 20-fold difference ($p = <0.0001$) in gene

expression in ruptured compared with painful tendon, highlighting the changes occurring within the tendon at a molecular level in different forms of ATP (tendinopathy and rupture). Due to this association between expression levels of the *ADAM8* gene and ATP, this gene was selected as a candidate for analysis of CNV. The Database of Genomic Variants also indicated CNV within the *ADAM8* gene. One TaqMan Copy Number Assays was used to genotype a region within exon 23 of the *ADAM8* gene. Details of the location of this assay is shown in **Figure 4.3**.

Chromosome 10 - *ADAM8*

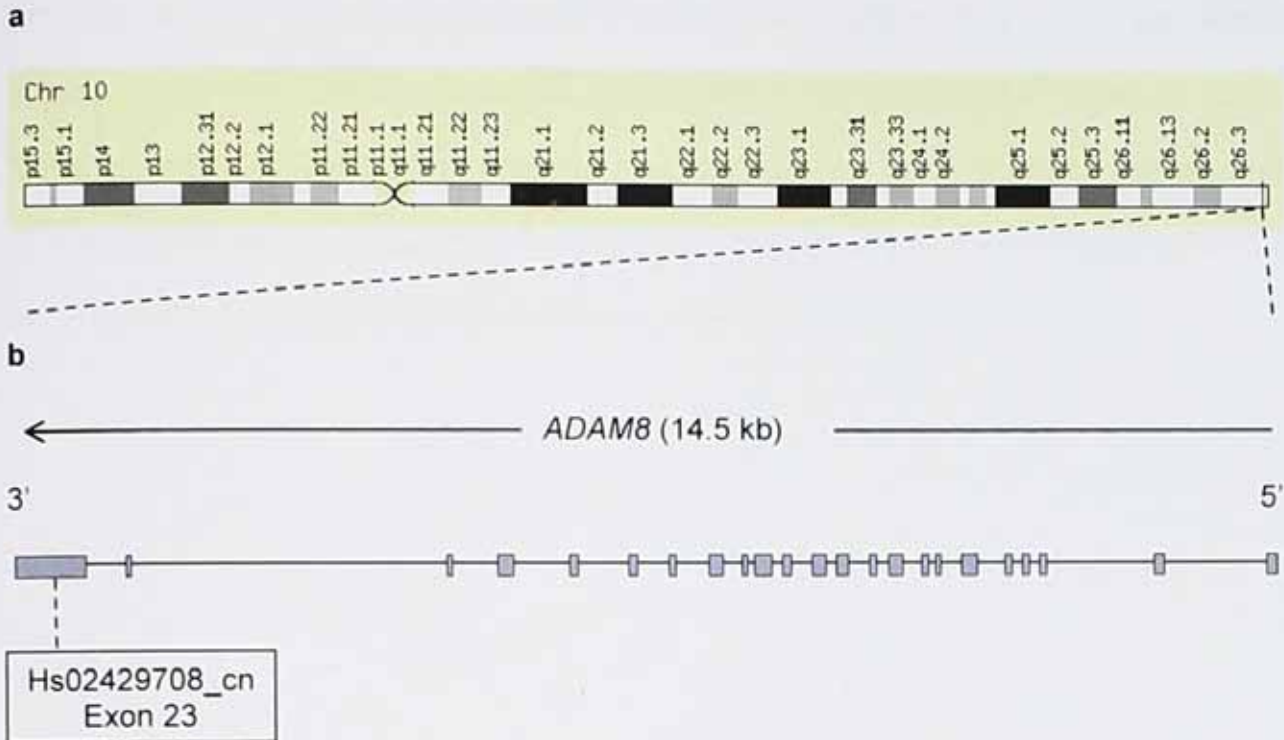


Figure 4.3 Location of the Hs02429708_cn Copy Number Assay within the *ADAM8* gene

(**a**) Chromosomal location of the *ADAM8* gene (Chr.10:135,075,907-135,090,372). (**b**) Size and exonic structure of the *ADAM8* gene with location of the TaqMan Copy Number Assay (Chr.10:135,076,357). All locations referenced from NCBI build 37. Compiled from www.ensembl.org and www.lifetechnologies.com.

4.1.4 Chapter Aims

The specific aim of this chapter was to determine whether CNV in the candidate genes *MMP3* (measured using the Hs02908568_cn and Hs02276714_cn Copy Number Assays), *MMP23B* (using the Hs07485808_cn Copy Number Assay) and *ADAM8* (using the Hs02429708_cn Copy Number Assay) was associated with ATP in UK, South African and Australian Caucasian populations.

4.2 METHODS

For an overview of the detailed methodology used please see Chapter 2. The methods below are specific to this chapter and are described in brief.

4.2.1 Participants

A total of 286 (123 UK, 110 SA, 53 AUS) Caucasian participants diagnosed with ATP were recruited for this study. For controls, 345 (131 UK, 94 SA, 120 AUS) physically active Caucasian participants (asymptomatic and without history of ATP) were recruited.

4.2.2 DNA Collection and Extraction

For the UK cohort, DNA was extracted from 2 mL of saliva. For the South African and Australian cohorts, DNA was extracted from approximately 4.5 mL of venous blood. Full details of DNA collection and extraction are described in Chapter 2.

4.2.3 Copy Number Genotyping

TaqMan Copy Number Assay and TaqMan Copy Number Reference Assay RNase P were run simultaneously in a duplex real-time PCR reaction using a StepOnePlus Real-Time PCR System (Applied Biosystems, Foster City, California, USA). Reactions were performed in 96-well plates using universal PCR cycling conditions. Each reaction contained 10 ng purified genomic DNA combined with a mastermix containing AmpliTaq Gold DNA Polymerase and the probes and primers for both the target copy number and reference assay. Quadruplicate reactions were run for each DNA sample and NTCs were included within each plate. Data were exported to CopyCaller Software version 2.0 (Applied Biosystems, Foster City, California, USA) for calculation of copy number.

4.2.4 Data Analysis

A one-way analysis of variance (ANOVA) was used to test for any significant differences between the characteristics (age, height, weight, BMI) of the CON and ATP groups, as well as to test for differences between the characteristics and copy number. A Pearson's chi-squared (χ^2) test was used to determine any significant differences between the gender of the CON and ATP groups. A Fisher's Exact Test was used to determine any significant differences in discrete copy number (<2, =2, >2 copies) between the CON and ATP groups. A t-Test or Mann-Whitney U Test was used to test for differences in the continuous CN data and were reported if significant. CON and ATP were compared within cohorts and with a combined Caucasian cohort (UK, SA and AUS). The ATP group was also split to analyse the TEN and RUP groups separately and data were analysed by gender and reported if significant. Significance was accepted at $p < 0.05$ for all analyses.

4.3 RESULTS

When the characteristics of the CON and ATP cohort were compared, there was a significant difference in age between the CON and ATP group in the UK ($p = 0.001$), and the SA ($p = 0.025$) cohorts. Weight and BMI were also significantly different between the CON and ATP group in the SA cohort and the AUS cohort (both $p < 0.05$). Gender distribution (% males) was significantly different between the AUS CON and ATP groups ($p < 0.05$). See Chapter 2 for full details of participant characteristics.

The relationship between the continuous CN values and the discrete CN calls for the Hs02276714_cn assay can be seen in **Figure 4.4**. Good clustering is obtained around each whole copy number, with the continuous CN values grouping into distinct clusters for CN = 1, CN = 2 and CN = 3. The CopyCaller software made calls of discrete copy number based on the continuous copy number data.

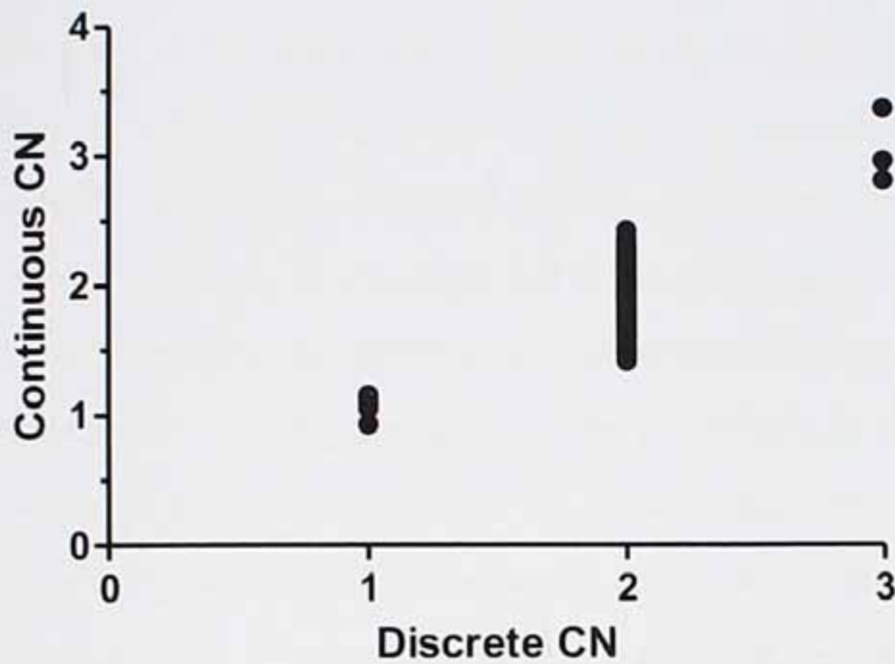


Figure 4.4 Scatter plot showing the raw continuous CN values on the y-axis, plotted against the discrete integer CN values on the x-axis. Graph shown was for the Hs02276714_cn assay within the *MMP3* gene with a combined cohort (UK, SA and AUS).

4.3.1 *MMP3* Copy Number Frequencies

There were no significant differences in copy number for either loci investigated within the *MMP3* gene (Hs02276714_cn and Hs02908568_cn) between the combined CON and ATP groups, nor between the combined CON and TEN or CON and RUP groups. There were also no differences for the UK and SA cohorts when analysed separately (**Table 4.1a** and **4.1b**). There was a significant difference in copy number within the *MMP3* gene (Hs02908568_cn) between the AUS CON and ATP group ($p = 0.035$) and between the AUS CON and TEN group ($p = 0.035$). However, the low sample size (CON, 87; ATP, 33; TEN,33) had reduced power (**Table 4.1b**) and analysis of the continuous CN data with a Mann-Whitney U Test showed a lack of significance ($p = 0.057$) (data not shown).

The QPCR efficiency of the *MMP3* assay was >95%. The efficiency graph for the Hs02276714_cn assay can be seen in Appendix V (**Figure V.7**). Data was not obtained for the Hs02908568_cn assay.

Table 4.1a Copy number distribution within the *MMP3* gene in the cases (ATP/TEN/RUP) and CON groups using the Hs02276712_cn assay.

Cohort		Mean CN	< 2 CN	= 2 CN	> 2 CN	n	p-value
	MMP3						
	Hs02276714_cn						
COMBINED	CON	1.97 ± 0.21	4	305	2	311	-
	ATP	1.98 ± 0.15	0	257	1	258	0.244
	TEN	1.99 ± 0.16	0	186	1	187	0.339
	RUP	1.98 ± 0.12	0	71	0	71	1.000
UK	CON	1.98 ± 0.23	2	125	2	129	-
	ATP	1.99 ± 0.14	0	118	1	119	0.623
	TEN	1.99 ± 0.16	0	85	1	86	0.790
	RUP	2.00 ± 0.10	0	33	0	33	1.000
SA	CON	1.97 ± 0.19	0	93	0	93	-
	ATP	1.98 ± 0.15	0	106	0	106	-
	TEN	1.99 ± 0.15	0	68	0	68	-
	RUP	1.96 ± 0.14	0	38	0	38	-
AUS	CON	1.95 ± 0.19	2	87	0	89	-
	ATP	1.97 ± 0.21	0	33	0	33	1.000
	TEN	1.97 ± 0.21	0	33	0	33	1.000
	RUP	-	0	0	0	0	-

Differences in discrete copy number between the CON vs. ATP, CON vs. TEN and CON vs. RUP groups were analysed with a Fisher's Exact Test. Mean CN (± SD) (± SD) represents the continuous CN data, which were analysed with a Mann-Whitney U Test.

Table 4.1b Copy number distribution within the *MMP3* gene in the cases (ATP/TEN/RUP) and CON groups using the Hs02908568_cn assay.

Cohort		Mean CN	< 2 CN	= 2 CN	> 2 CN	n	p-value
	MMP3						
	Hs02908568_cn						
COMBINED	CON	2.01 ± 0.24	0	272	5	277	-
	ATP	2.00 ± 0.21	0	216	5	221	0.757
	TEN	2.02 ± 0.24	0	151	5	156	0.506
	RUP	1.97 ± 0.10	0	65	0	65	0.588
UK	CON	1.99 ± 0.10	0	101	1	102	-
	ATP	1.98 ± 0.07	0	88	0	88	1.000
	TEN	1.98 ± 0.06	0	60	0	60	1.000
	RUP	1.98 ± 0.08	0	28	0	28	1.000
SA	CON	2.00 ± 0.17	0	87	1	88	-
	ATP	1.97 ± 0.12	0	100	0	100	0.468
	TEN	1.97 ± 0.13	0	63	0	63	1.000
	RUP	1.97 ± 0.11	0	37	0	37	1.000
AUS	CON	2.04 ± 0.38	0	84	3	87	-
	ATP	2.16 ± 0.45	0	28	5	33	**0.035**
	TEN	2.16 ± 0.45	0	28	5	33	**0.035**
	RUP	-	0	0	0	0	-

Differences in discrete copy number between the CON vs. ATP, CON vs. TEN and CON vs. RUP groups were analysed with a Fisher's Exact Test. Mean CN (± SD) (± SD) represents the continuous CN data, which were analysed with a Mann-Whitney U Test.

4.3.2 *MMP23B* Copy Number Frequencies

For the *MMP23B* gene within exon 1, we found no significant differences in copy number between the combined CON and ATP ($p = 0.090$), the CON and TEN ($p = 0.387$) or the CON and RUP ($p = 0.436$) groups. There were also no significant differences when the UK and SA cohorts were analysed separately (**Table 4.2**).

The QPCR efficiency of the *MMP23B* assay was >86%. The efficiency graph for the Hs07485808_cn assay can be seen in Appendix V (**Figure V.8**).

Table 4.2 Copy number distribution within the *MMP23B* gene in the cases (ATP/TEN/RUP) and CON groups using the Hs07485808_cn assay.

Cohort		Mean CN	< 2 CN	= 2 CN	> 2 CN	n	p-value
	MMP23B						
	Hs07485808_cn						
COMBINED	CON	1.96 ± 0.32	6	206	3	215	-
	ATP	1.96 ± 0.19	3	218	0	221	0.090
	TEN	1.95 ± 0.20	3	150	0	153	0.387
	RUP	2.00 ± 0.15	0	68	0	68	0.436
UK	CON	1.97 ± 0.25	2	124	2	128	-
	ATP	1.97 ± 0.12	0	117	0	117	0.249
	TEN	1.96 ± 0.13	0	86	0	86	0.421
	RUP	2.00 ± 0.10	0	31	0	31	1.000
SA	CON	1.94 ± 0.40	4	82	1	87	-
	ATP	1.95 ± 0.24	3	101	0	104	0.456
	TEN	1.93 ± 0.26	3	64	0	67	1.000
	RUP	2.00 ± 0.17	0	37	0	37	0.519

Differences in discrete copy number between the CON vs. ATP, CON vs. TEN and CON vs. RUP groups were analysed with a Fisher's Exact Test. Mean CN (± SD) represents the continuous CN data, which were analysed with a Mann-Whitney U Test.

4.3.3 *ADAM8* Copy Number Frequencies

We found no significant differences in copy number between the combined CON and ATP (p = 0.727), the CON and TEN (p = 0.707) or the CON and RUP (p = 1.000) groups within the *ADAM8* gene. When the UK, SA and AUS cohorts were analysed separately, there were no significant difference between groups (**Table 4.3**).

The QPCR efficiency of the *ADAM8* assay was >92%. The efficiency graph for the Hs02429708_cn assay can be seen in Appendix V (**Figure V.1**).

Table 4.3 Copy number distribution within the *ADAM8* gene in the cases (ATP/TEN/RUP) and CON groups using the Hs02429708_cn assay.

Cohort		Mean CN	< 2 CN	= 2 CN	> 2 CN	n	p-value
	ADAM8						
	Hs02429708_cn						
COMBINED	CON	2.01 ± 0.26	0	299	5	304	-
	ATP	2.00 ± 0.15	0	272	3	275	0.727
	TEN	1.99 ± 0.15	0	202	2	204	0.707
	RUP	2.02 ± 0.16	0	70	1	71	1.000
UK	CON	2.00 ± 0.12	0	128	1	129	-
	ATP	1.99 ± 0.09	0	122	0	122	1.000
	TEN	1.99 ± 0.09	0	88	0	88	1.000
	RUP	2.00 ± 0.10	0	34	0	34	1.000
SA	CON	2.02 ± 0.37	0	88	1	89	-
	ATP	1.99 ± 0.17	0	106	1	107	1.000
	TEN	1.96 ± 0.15	0	70	0	70	1.000
	RUP	2.04 ± 0.20	0	36	1	37	0.503
AUS	CON	2.03 ± 0.27	0	83	3	86	-
	ATP	2.03 ± 0.21	0	44	2	46	1.000
	TEN	2.03 ± 0.21	0	44	2	46	1.000
	RUP	-	0	0	0	0	-

Differences in discrete copy number between the CON vs. ATP, CON vs. TEN and CON vs. RUP groups were analysed with a Fisher's Exact Test. Mean CN (± SD) represents the continuous CN data, which were analysed with a Mann-Whitney U Test.

4.4 CHAPTER SPECIFIC CONCLUSIONS

Despite a range of frequencies reported in the Database of Genomic Variants for the *MMP23B* and *ADAM8* genes (Chapter 2), when genotyped, the majority of samples from all three cohorts gave a copy number of 2, with only a small number of individuals showing either a loss or gain in copy number.

There was no prior data available to predict the frequency of CNV within the *MMP3* gene at the two exonic loci we investigated, as no CNV has been reported in the Database of Genomic Variants at these loci. However, we reported variation in copy number at both these loci. It is recognised that the predicted frequency of a given CNV derived from databases alone might incorporate a degree of error, due to uncertainty into the actual extent of CNV in human populations of different backgrounds (Pinto et al. 2007; Grozeva et al. 2012).

There were significant differences between the participant characteristics between groups (Chapter 2). However, these were not found to influence the relationship between CNV in the loci tested and ATP, as none of the anthropometric data were associated with or interacted with copy number.

The discrete CN data for the *MMP3* Hs02908568_cn assay showed that the AUS ATP group had a higher frequency of participants with CN>2, compared with the AUS CON group. While this analysis was underpowered, it could imply that increased copy number at this locus may increase the risk of ATP.

Although we found no statistically powered associations between CNV in the metalloproteinase genes investigated and ATP, research has shown their

importance in tendinopathy and other musculoskeletal pathologies (Raleigh et al. 2009; El Khoury et al. 2013; Jones et al. 2006; Kerna et al. 2009; Jones and Riley 2005).

CHAPTER 5: INVESTIGATING THE ROLE OF APOPTOSIS ASSOCIATED GENES AND THE RISK OF ACHILLES TENDON PATHOLOGY

5.1 INTRODUCTION

The removal of damaged tenocytes by apoptosis is a normal mechanism in tendon healing (September et al. 2012), however, relatively little is known about genetic variation within genes involved in apoptosis and the risk of ATP. Previous research has been limited to variants within the caspase-8 (*CASP8* rs384129 and rs1045485), nitric oxide synthase 2 (*NOS2* rs2779249) and nitric oxide synthase 3 (*NOS3* rs1799983) genes in two cohorts from the Southern Hemisphere (South Africa and Australia) (Nell et al. 2012).

5.1.1 TNFRSF1A

Tumour necrosis factor receptor 1 (TNFR1), encoded by the *TNFRSF1A* gene, is a cell receptor that can signal apoptosis in response to the pro-inflammatory cytokine tumour necrosis factor-alpha (TNF-α) (Hosaka et al. 2005; Gaida et al. 2012). TNF-α is involved in an intracellular apoptotic signalling pathway that can result in apoptosis through binding to TNF receptors and the subsequent initiation of a cascade of reactions (**Figure 5.1**) (Hosaka et al. 2005).

Figure 5.1 TNFR1 Pathway showing the cascade of reactions following the binding of TNF-α to the TNFR1 receptor. © 2009 QIAGEN, all rights reserved.

TNFR1 can signal cell death in response to TNF-α by recruiting TNF receptor-associated death domain (TRADD), which then interacts with Fas-associated death domain (FADD) (Hosaka et al., 2005). The FADD protein can then activate

procaspase-8 to caspase-8 (Hosaka et al. 2005). TNF-α and its receptors TNFR1 & TNFR2 have been identified in cultured tenocytes (Backman et al. 2013) and TNFR1 mRNA and protein have been identified in human Achilles tendon (Gaida et al. 2012). Furthermore, TNFR1 is known to be highly expressed in tenocytes isolated from Achilles tendinosis (Gaida et al. 2012), but the pathological significance of this is not fully understood. While variants within the *TNFRSF1A* gene, including the rs4149577 variant, have previously been associated with musculoskeletal diseases (ankylosing spondylitis) (Karaderi et al. 2012) and inflammatory demyelinating diseases (neuromyelitis optica) (Park et al. 2013), this variant has not been investigated as a risk factor for ATP. One TaqMan Genotyping Assay was used to genotype the rs4149577 C/T intronic SNP within the *TNFRSF1A* gene. Details of the location of this assay are shown in **Figure 5.2**.

Chromosome 12 – *TNFRSF1A*

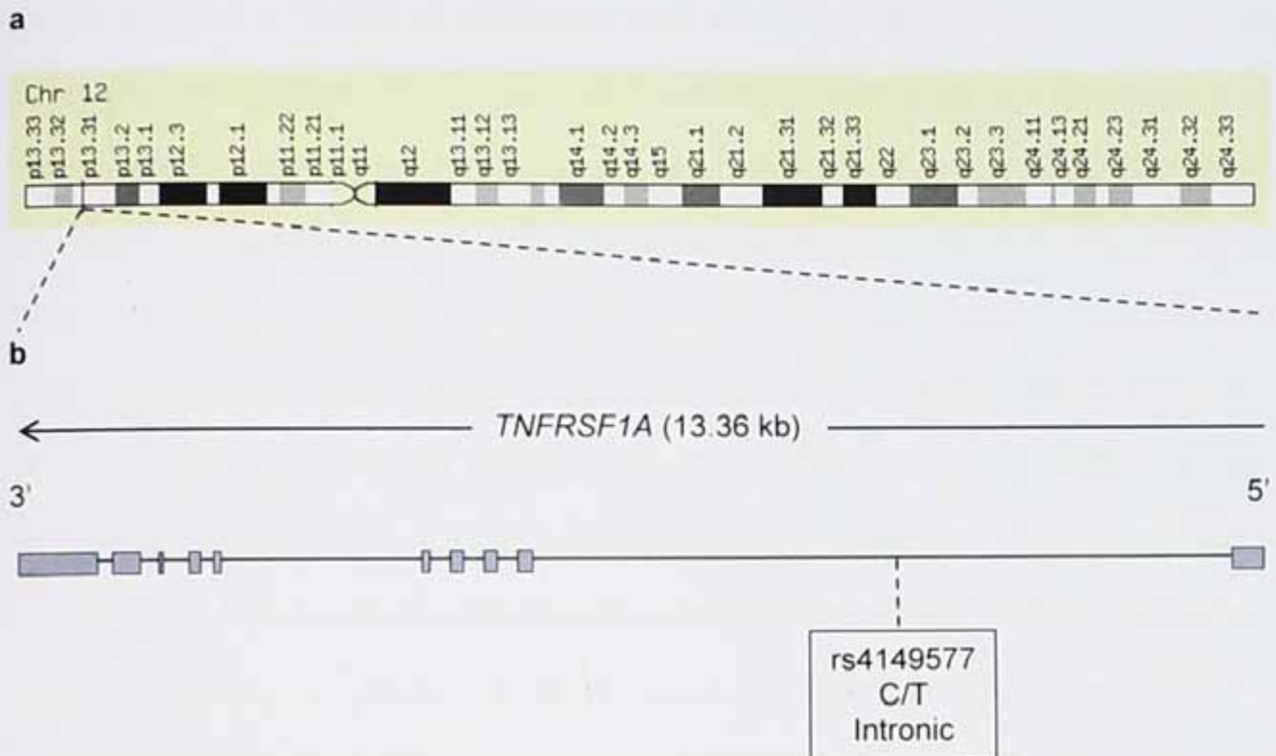


Figure 5.2 Location of the rs4149577 SNP within the *TNFRSF1A* gene

(**a**) Chromosomal location of the *TNFRSF1A* gene (Chr.12:6,437,923-6,451,283). (**b**) Size and exonic structure of the *TNFRSF1A* gene with location of the rs4149577 C/T variant (Chr.12:6,447,522). All locations referenced from NCBI build 37. Compiled from www.ensembl.org and www.lifetechnologies.com.

5.1.2 CASP8

Caspases form a family of proteases that are important in the regulation of apoptosis (Hengartner 2000). They can mediate inflammation and apoptosis and are known to have a direct role in tendon apoptosis (Lamkanfi et al. 2007; Nell et al. 2012). Caspase-8 is an enzyme component of the caspase pathway, encoded by the *CASP8* gene, which regulates apoptosis of tendon fibroblasts (September et al. 2012) and can activate downstream effector caspases (**Figure 5.1**) (Hosaka et al. 2005). Regulating the balance between synthesis and degradation is vital for maintaining ECM homeostasis and the removal of damaged tendon fibroblasts during normal tendon turnover is important (September et al. 2012). However, atypical tenocyte apoptosis has been shown in tendinopathy, where elevated expression levels of *CASP8* have been observed (Nell et al. 2012). Previously, two SNPs (rs1045485 and rs3834129) within the *CASP8* gene have been independently associated with ATP (Nell et al. 2012). For the *CASP8* rs1045485 variant, a significant difference in genotype (P=0.0294) was detected between CON and ATP groups (Nell et al. 2012). For the *CASP8* rs3834129 variant, significant differences in genotype and allele frequencies were identified between the two countries (South African and Australian cohorts), as well as between the CON and ATP groups (Nell et al. 2012). However the role of copy number variation within this gene has not been considered in relation to ATP. The *CASP8* gene has many CNV regions recorded on the Database of Genomic Variants (http://dgv.tcag.ca/dgv/app/home). One TaqMan Copy Number Assays was used to genotype a region within intron 11 – intron 12 of the *CASP8* gene. Details of the location of this assay are shown in **Figure 5.3**.

Chromosome 2 – *CASP8*

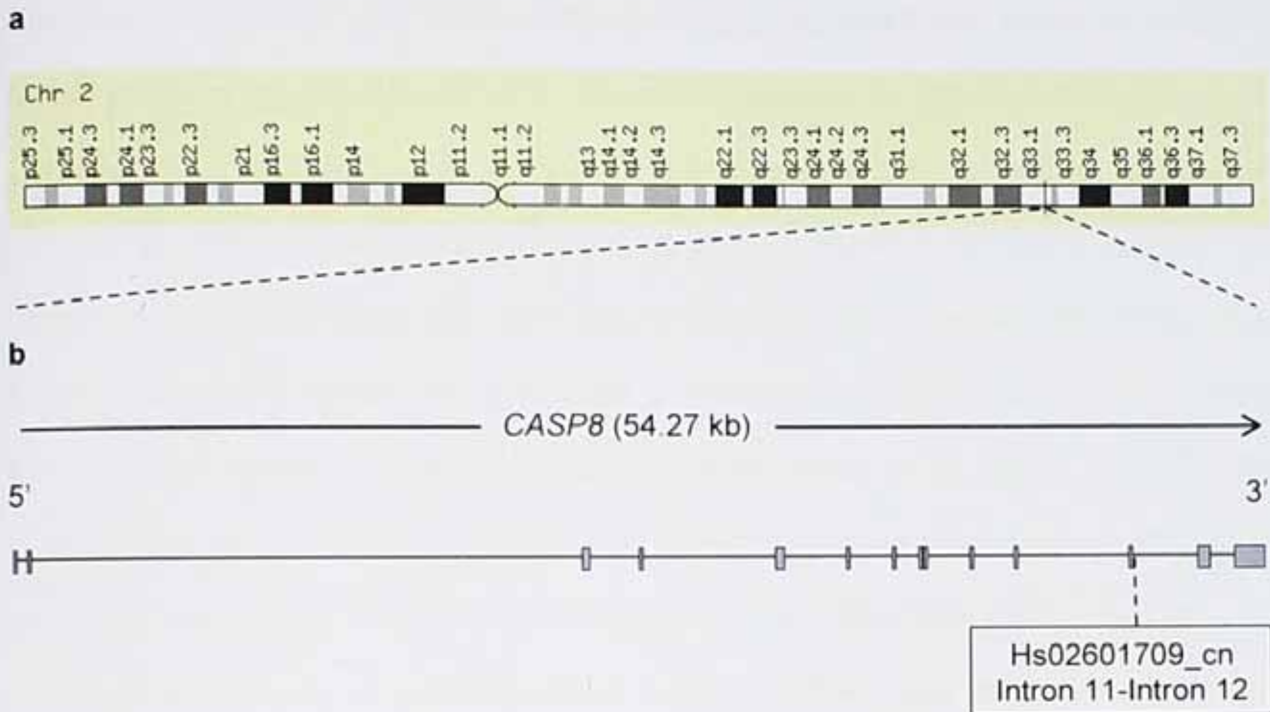


Figure 5.3 Location of the Hs02601709_cn Copy Number Assay within the *CASP8* gene

(**a**) Chromosomal location of the *CASP8* gene (Chr.2:202,098,166-202,152,434). (**b**) Size and exonic structure of the *CASP8* gene with location of the TaqMan Copy Number Assay (Chr.2:202,146,600). All locations referenced from NCBI build 37. Compiled from www.ensembl.org and www.lifetechnologies.com.

5.1.3 CASP3

Caspase-3 is one of three known effector caspases, which can selectively cleave target proteins, such as Bcl-2, after aspartate residues in their primary sequence (Hengartner 2000; Kirsch et al. 1999). Apoptosis can be activated through a number of complex pathways, however, caspase-3 appears to have a critical role in chromatin condensation and DNA fragmentation (Porter and Jänicke 1999). The *CASP3* rs1049253 variant lies within the 3' untranslated region (UTR) of the gene and is known site for microRNA (miRNA) binding (Guan et al. 2013). MiRNAs can bind mRNAs and have an important role in regulating apoptosis associated with carcinogenesis (Guan et al. 2013). Specifically, the rs1049253 variant is known to influence the binding of miR-885-5p to *CASP3* mRNA (Guan et al. 2013). It has been associated with the risk of certain cancers and the CC genotype of this variant has been shown to influence levels of *CASP3* mRNA expression (Guan et al. 2013). The potential role of the *CASP3* rs1049253 variant as a risk factor for ATP is unknown. One TaqMan Genotyping Assay was used to genotype the rs1049253 C/T 3'-UTR SNP within the *CASP3* gene. Details of the location of this assay are shown in **Figure 5.4**.

Chromosome 4 - *CASP3*

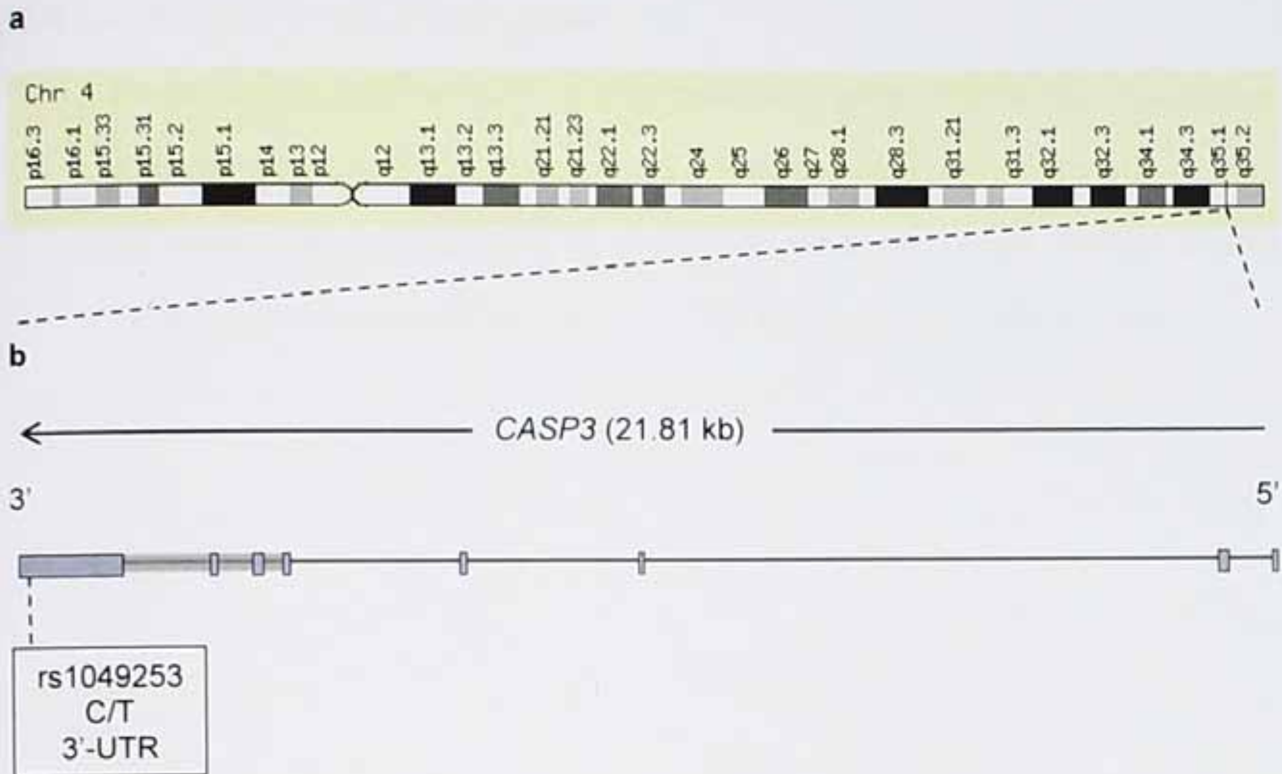


Figure 5.4 Location of the rs1049253 SNP within the *CASP3* gene

(**a**) Chromosomal location of the *CASP3* gene (Chr.4:185,548,850-185,570,663). (**b**) Size and exonic structure of the *CASP3* gene with location of the rs1049253 C/T variant (Chr.4:185,548,951). All locations referenced from NCBI build 37. Compiled from www.ensembl.org and www.lifetechnologies.com.

5.1.4 Chapter Aims

The specific aims of this chapter were:

- To determine whether CNV in the candidate gene *CASP8* was associated with ATP in a UK, South African and Australian Caucasian population.
- To determine whether SNPs in the novel candidate genes *TNFRSF1A* and *CASP3* were associated with ATP in a UK and South African population.

5.2 METHODS

For an overview of the detailed methodology used please see Chapter 2. The methods below are specific to this chapter and are described in brief.

5.2.1 Participants

For this study, 294 (165 UK, 114 SA, 15 AUS) Caucasian participants diagnosed with ATP were recruited. A total of 316 (131 UK, 97 SA, 88 AUS) physically active Caucasian control participants (asymptomatic and without history of ATP) were recruited as controls.

5.2.2 DNA Collection and Extraction

For the UK cohort, DNA was extracted from 2 mL of saliva. For the South African and Australian cohorts, DNA was extracted from approximately 4.5 mL of venous blood. Full details of DNA collection and extraction are described in Chapter 2.

5.2.3 TaqMan SNP Genotyping

Real-time PCR reactions were run with TaqMan SNP Genotyping Assays using a StepOnePlus Real-Time PCR System (Applied Biosystems, Foster City, California, USA). TaqMan SNP Genotyping Assay, TaqMan Universal PCR Master Mix, No AmpErase UNG and genomic DNA (gDNA) were combined per well of a 96-well plate. TaqMan SNP Genotyping Assays contained sequence-specific forward and reverse primers to amplify the target sequence and two TaqMan fluorescent probes, to detect the minor and major alleles. Homozygote and heterozygote samples were automatically determined by the software as the samples grouped into genotype clusters. Some samples were manual called after analysis of the amplification and fluorescence detected.

5.2.4 Copy Number Genotyping

TaqMan Copy Number Assay and TaqMan Copy Number Reference Assay RNase P were run simultaneously in a duplex real-time PCR reaction using a StepOnePlus Real-Time PCR System (Applied Biosystems, Foster City, California, USA). Each reaction contained 10 ng purified genomic DNA combined with a mastermix containing AmpliTaq Gold DNA Polymerase and the probes and primers for both the target copy number and reference assay. Data were exported to CopyCaller Software version 2.0 (Applied Biosystems, Foster City, California, USA) for calculation of copy number.

5.2.5 Data Analysis

A χ^2 or Fisher's Exact Test was used to determine any significant differences in discrete copy number (<2, =2, >2 copies) and SNP genotypes between the CON and ATP groups. Differences in continuous copy number were analysed with a non-parametric Mann-Whitney U Test. The ATP group was also split to analyse the TEN and RUP groups separately and data were analysed by gender and reported if significant. Significance was accepted at $p < 0.05$ for all analyses. For the SNP genotyping data, Hardy-Weinberg equilibrium (HWE) was established using a HWE calculator (Michael H. Court, 2005-2008) accessed from www.tufts.edu/~mcourt01/Documents/Court lab - HW calculator.xls. $P < 0.05$ was considered to be a deviation from HWE.

5.3 RESULTS

As previously reported in preceding chapters, there were significant differences between certain participant characteristics of the CON and ATP cohort. Full details can be seen in Chapter 2 and will be considered in the discussion in Chapter 7.

A typical cluster plot generated by the StepOne Software for TaqMan SNP Genotyping can be seen in **Figure 5.5**. This is the output generated after the completed qPCR cycle. The cluster plot shown is for the *CASP3* rs1049253 SNP.

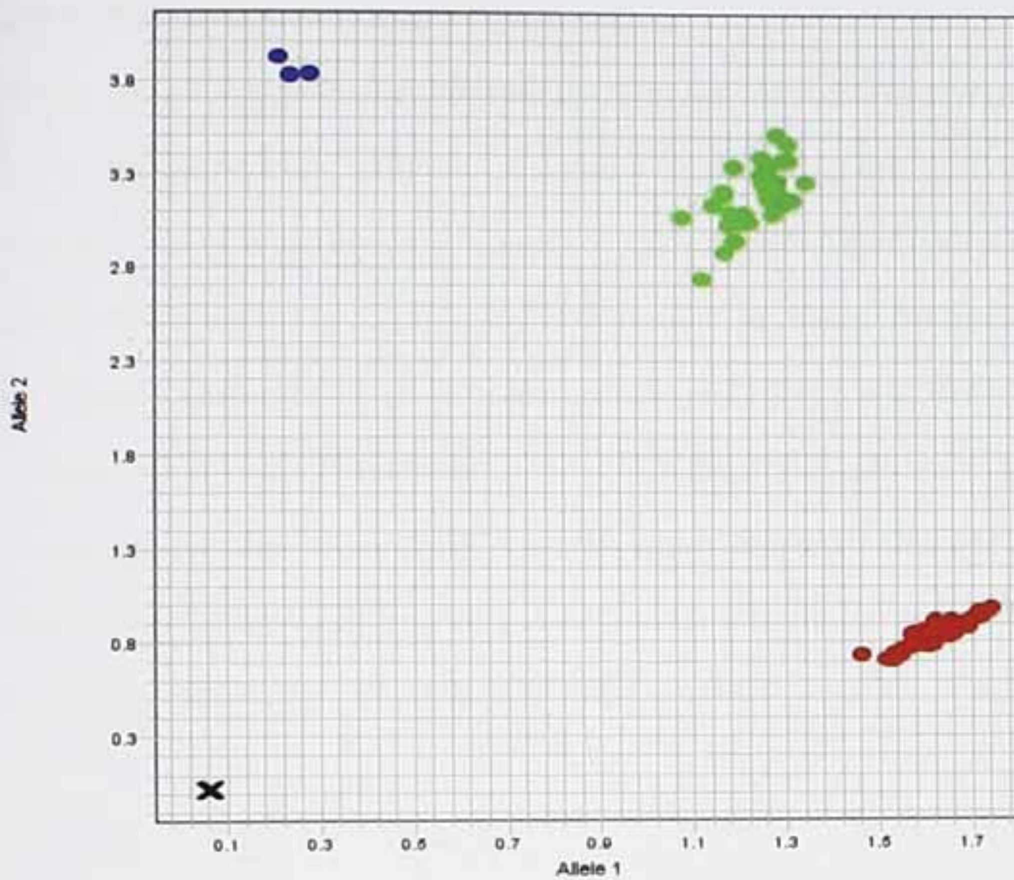


Figure 5.5 Typical allelic discrimination cluster plot for the rs1049253 SNP in the *CASP3* gene. The green clusters represent the heterozygote AG samples. The blue and red clusters represent the homozygotes GG and AA respectively. The black Xs represent the NTCs.

5.3.1 *TNFRSF1A* Genotype Frequencies

There were no significant differences in genotype for the *TNFRSF1A* rs4149577 variant between the combined CON and ATP groups (p = 0.556), nor between the combined CON and TEN (p = 0.648) or CON and RUP groups (p = 0.559) (**Table 5.1**). There were no significant differences in allele frequency between the combined CON and ATP (p = 0.282), CON and TEN (p = 0.405) or CON and RUP groups (p = 0.298). There were also no significant differences in genotype or allele frequency for the UK and SA cohorts when analysed separately (**Table 5.1**). Hardy-Weinberg equilibrium was observed for all frequencies in the combined and individual cohorts for the CON, ATP, TEN and RUP groups.

Table 5.1 Genotype frequency distribution for the rs4149577 variant within the *TNFRSF1A* gene in the cases (ATP/TEN/RUP) and CON groups.

Cohort		TT	CT	CC	n	p-value	MAF	p-value	HWE
	***TNFRSF1A* rs4149577**								
COMBINED	CON	26.9 (61)	48.5 (110)	24.7 (56)	227	-	48.9 (222)	-	0.647
	ATP	30.9 (86)	47.1 (131)	21.9 (61)	278	0.556	45.5 (253)	0.282	0.406
	TEN	31.0 (63)	45.8 (93)	23.2 (47)	203	0.648	46.1 (187)	0.405	0.266
	RUP	30.7 (23)	50.7 (38)	18.7 (14)	75	0.559	44.0 (66)	0.298	0.807
UK	CON	27.5 (36)	48.9 (64)	23.7 (31)	131	-	48.1 (126)	-	0.806
	ATP	32.1 (53)	46.1 (76)	21.8 (36)	165	0.696	44.8 (148)	0.432	0.376
	TEN	29.7 (38)	46.9 (60)	23.4 (30)	128	0.934	46.9 (120)	0.782	0.506
	RUP	40.5 (15)	43.2 (16)	16.2 (6)	37	0.285	37.8 (28)	0.118	0.623
SA	CON	26.0 (25)	47.9 (46)	26.0 (25)	96	-	50.0 (96)	-	0.683
	ATP	29.2 (33)	48.7 (55)	22.1 (25)	113	0.789	46.5 (105)	0.470	0.818
	TEN	33.3 (25)	44.0 (33)	22.7 (17)	75	0.575	44.7 (67)	0.327	0.341
	RUP	21.1 (8)	57.9 (22)	21.1 (8)	38	0.632	50.0 (38)	1.000	0.330

Values expressed as a frequency (%) with number of participants (n) in parenthesis. HWE: Hardy-Weinberg equilibrium. MAF: minor allele frequency. Differences in genotype and allele frequency between the CON vs. ATP, CON vs. TEN and CON vs. RUP groups were analysed with a χ^2 or Fisher's Exact Test.

5.3.2 *CASP8* Copy Number Frequencies

There were no significant differences in copy number for the locus investigated within the *CASP8* gene (Hs02601709_cn) between the combined CON and ATP groups ($p = 0.402$), nor between the combined CON and TEN ($p = 0.653$) or CON and RUP groups ($p = 0.363$) (**Table 5.2**). When the UK, SA and AUS cohorts where analysed separately, there were also no differences between the CON and ATP groups (UK, $p = 0.155$; SA, $p = 0.327$; AUS, $p = 1.000$). This was also the case for the CON vs. TEN (UK, $p = 0.124$; SA, $p = 0.861$; AUS, $p = 1.000$) and CON vs. RUP groups (UK, $p = 0.695$; SA, $p = 0.064$) (**Table 5.2**). Although the $p = 0.065$ is approaching significance, the total number of SA RUP in this group is $n = 28$. There were no rupture cases in the AUS cohort, hence why there is no comparison for the CON vs. RUP groups in this instance.

In addition to the main analyses of copy number using a χ^2 or Fisher's Exact Test for the discrete data, the continuous copy number data were also analysed with a Mann-Whitney U Test, to check concordance. The mean CN ± SD are shown for each group. The p values are not shown, however, there were no significant differences in any of the comparisons of the combined or separate cohorts shown in **Table 5.2**. This supports the findings reported from the discrete copy number analyses.

The QPCR efficiency of the *CASP8* assay exceed 100%. The efficiency graph for the Hs02601709_cn assay can be seen in Appendix V (**Figure V.2**).

Table 5.2 Copy number distribution within the *CASP8* gene in the cases (ATP/TEN/RUP) and CON groups using the Hs02601709_cn assay.

Cohort		Mean CN	< 2 CN	= 2 CN	> 2 CN	n	p-value
	CASP8						
	Hs02601709_cn						
COMBINED	CON	2.15 ± 0.56	6	138	56	200	-
	ATP	2.13 ± 0.54	4	142	43	189	0.402
	TEN	2.14 ± 0.55	3	104	34	141	0.653
	RUP	2.11 ± 0.54	1	38	9	48	0.363
UK	CON	2.09 ± 0.59	2	46	14	62	-
	ATP	1.95 ± 0.44	4	73	9	86	0.155
	TEN	1.94 ± 0.44	3	57	6	66	0.124
	RUP	1.99 ± 0.46	1	16	3	20	0.695
SA	CON	2.35 ± 0.51	0	40	31	71	-
	ATP	2.32 ± 0.58	0	57	31	88	0.327
	TEN	2.38 ± 0.57	0	35	25	60	0.861
	RUP	2.20 ± 0.58	0	22	6	28	0.064
AUS	CON	1.98 ± 0.52	4	52	11	67	-
	ATP	2.03 ± 0.48	0	12	3	15	1.000
	TEN	2.03 ± 0.48	0	12	3	15	1.000
	RUP	-	0	0	0	0	-

Differences in discrete copy number between the CON vs. ATP, CON vs. TEN and CON vs. RUP groups were analysed with a χ^2 or Fisher's Exact Test. Mean CN (± SD) represents the continuous CN data, which were analysed with a Mann-Whitney U Test.

5.3.3 *CASP3* Genotype Frequencies

There were no significant differences in genotype for the *CASP3* rs1049253 variant between the combined CON and ATP groups ($p = 0.804$), nor between the combined CON and TEN ($p = 0.680$) or CON and RUP ($p = 0.745$) groups. There were no significant differences in allele frequency between the combined CON and ATP ($p = 0.966$), CON and TEN ($p = 0.627$) or CON and RUP groups ($p = 0.380$). There were also no significant differences in genotype or allele frequency for the UK and SA cohorts when analysed separately (**Table 5.3**).

Table 5.3 Genotype frequency distribution for the rs1049253 variant within the *CASP3* gene in the cases (ATP/TEN/RUP) and CON groups.

Cohort	***CASP3* rs1049253**	TT	CT	CC	n	p-value	MAF	p-value	HWE
COMBINED	CON	63.7 (142)	32.3 (72)	4.0 (9)	223	-	20.2 (90)	-	0.973
	ATP	64.5 (178)	30.4 (84)	5.1 (14)	276	0.804	20.3 (112)	0.966	0.326
	TEN	62.9 (127)	31.2 (63)	5.9 (12)	202	0.680	21.5 (87)	0.627	0.309
	RUP	68.9 (51)	28.4 (21)	2.7 (2)	74	0.745	16.9 (25)	0.380	0.926
UK	CON	62.6 (82)	34.4 (45)	3.1 (4)	131	-	20.2 (53)	-	0.461
	ATP	61.6 (101)	32.3 (53)	6.1 (10)	164	0.468	22.3 (73)	0.550	0.397
	TEN	62.2 (79)	30.7 (39)	7.1 (9)	127	0.309	22.4 (57)	0.540	0.184
	RUP	59.5 (22)	37.8 (14)	2.7 (1)	37	0.876	21.6 (16)	0.793	0.479
SA	CON	65.2 (60)	29.3 (27)	5.4 (5)	92	-	20.1 (37)	-	0.406
	ATP	68.8 (77)	27.7 (31)	3.6 (4)	112	0.743	17.4 (39)	0.486	0.691
	TEN	64.0 (48)	32.0 (24)	4.0 (3)	75	0.893	20.0 (30)	0.980	1.000
	RUP	78.4 (29)	18.9 (7)	2.7 (1)	37	0.412	12.2 (9)	0.132	0.486

Values expressed as a frequency (%) with number of participants (n) in parenthesis. HWE: Hardy-Weinberg equilibrium. MAF: minor allele frequency. Differences in genotype and allele frequency between the CON vs. ATP, CON vs. TEN and CON vs. RUP groups were analysed with a χ^2 or Fisher's Exact Test.

5.4 CHAPTER SPECIFIC CONCLUSIONS

The *TNFRSF1A* rs4149577, *CASP3* rs1049253 and *CASP8* variants did not associate with ATP in our combined cohort or separate UK/SA/AUS cohorts. This study seems to exclude these three variants from a role in ATP.

Despite TNFR1 having a role in apoptosis and being expressed at significantly higher levels in tenocytes from Achilles tendinosis compared to controls (Gaida et al. 2012), we only investigated a single SNP within this gene (*TNFRSF1A* rs4149577) and it is possible that other variants within the DNA sequence may influence the risk of ATP. For example, the *TNFRSF1A* rs4149577 variant is an intronic SNP that is in linkage disequilibrium (LD) with a number of other variants, including *TNFRSF1A* rs1800693 and rs4149578 (www.ensembl.org) (Karaderi et al. 2012; Park et al. 2013; Swaminathan et al. 2010). These variants have been associated with musculoskeletal and inflammatory diseases and may be worthy of future investigation in ATP (Karaderi et al. 2012; Park et al. 2013; Swaminathan et al. 2010).

Nell *et al.* (2012) previously reported that variants within the *CASP8* gene were associated with Achilles tendinopathy in a South African and Australian cohort. The copy number variant spanning intron 11 - intron 12 of the *CASP8* gene that we investigated here overlaps the rs1045485 SNP investigated by Nell *et al.* (2012), but it is approximately 50 000 base pairs from the rs3834129 SNP. We found no association between CNV at this locus and ATP, in contrast to Nell *et al.* (2012). This may be related to a number of factors. SNPs like the *CASP8* rs1045485 variant may have a more profound impact on the development of ATP, compared to larger-scale variants such as CNV. This claim is supported by the fact

that the *CASP8* rs1045485 SNP is a non-synonymous G>C polymorphism that results in an amino acid change (Asp302His), which may affect interaction with other proteins (September et al. 2012). In contrast, the functional effect of altered copy number at the *CASP8* locus we investigated is presently unknown.

CHAPTER 6: DNA METHYLATION WITHIN THE *ADAMTS4*, *TIMP1* AND *MMP11* GENE PROMOTERS IN PATELLAR TENDINOPATHY

6.1 INTRODUCTION

Patellar tendinopathy (jumper's knee) is a common injury in sporting populations, with prevalence rates of higher than 30% in both basketball and volleyball players (Christian et al. 2014). While commonly affecting athletic populations, it is evident that patellar tendinopathy also occurs in the general population that lead a sedentary lifestyle (Samiric et al. 2009). DNA methylation is one way in which the genome is regulated, so that different cell types can be produced with specific functions (Auclair and Weber 2012). Altered DNA methylation of genes can result in a change in gene expression (Auclair and Weber 2012) and in relation to tendon injury, this could potentially disrupt the tendon environment and increase the risk of injury. The way in which DNA methylation influences gene expression is discussed in the introduction section on epigenetics (1.6.1).

6.1.1 ADAMTS4

ADAMTS4 (a disintegrin and metalloproteinase with thrombospondin motifs 4) is an enzyme involved in the homeostasis of the extracellular matrix (ECM) (Corps et al. 2008) and is able to cleave the proteoglycans aggrecan, versican and brevican (Corps et al. 2008; Jones and Riley 2005). ADAMST4 has also been shown to cleave COMP (cartilage oligomeric matrix protein), indicating that its proteolytic activity is not limited to proteoglycans (Jones and Riley 2005). COMP in a non-collagenous glycoprotein which can bind to type I, II and IX collagens, facilitating collagen-collagen interactions and influencing fibril formation (Verma and Dalal 2013). Therefore it is clear to see how hypomethylation of the *ADAMTS4*

gene and the subsequent increased expression of the gene could contribute to the degradation of cartilage in musculoskeletal pathologies like osteoarthritis, through an increase in the cleavage of COMP. In addition, cartilage is a connective tissue consisting of extracellular matrix and collagen fibrils, as are tendons (Verma and Dalal 2013).

Indeed, a loss of DNA methylation and altered expression of the *ADAMTS4* gene has been observed in chondrocytes in osteoarthritis (Roach et al. 2005), identifying it as an interesting candidate to research in other musculoskeletal conditions like ATP. In relation to tendinopathy, higher levels of ADAMTS4 mRNA have been detected in ruptured Achilles tendon compared with normal and tendinopathic tendon (Corps et al. 2008). However, in one study on patellar tendinopathy, there was no significant difference in ADAMTS4 expression between cases and controls (Parkinson et al. 2010).

6.1.2 TIMP1

Tissue inhibitors of metalloproteinases (TIMPs) are a small family of proteins, encoded by the *TIMP* genes, that can inhibit MMP activity (Jones et al. 2006). TIMPs also inhibit the action of ADAMs and ADAMTSs, however this role appears more limited (El Khoury et al. 2013; Jones et al. 2006). Nevertheless, TIMP3 is believed to be the prominent TIMP inhibitor of ADAMs and ADAMTSs (Jones et al. 2006) and ECM homeostasis is maintained in part from a balance of metalloproteinases and TIMPs (Karousou et al. 2008).

An 8-fold difference in expression of the *TIMP1* gene has been reported in ruptured compared with normal Achilles tendon, as well as a 5-fold increase between

ruptured and painful Achilles tendon (Jones et al. 2006). In addition, research by El Khoury et al. (2013) found that a SNP in the *TIMP2* gene was associated with ATP in a combined Caucasian cohort. The *TIMP2* rs4789932 variant was shown to be genotypically associated, highlighting this variant as a risk factor for ATP (El Khoury et al. 2013).

6.1.3 MMP11

Matrix metalloproteinase 11 (MMP11), also known as stromelysin 3, is an ECM enzyme of the MMP family known to cleave the proteoglycan laminin (Somerville et al. 2003). Differences in expression levels of the *MMP11* gene have also been reported in normal and pathologic Achilles tendon (Jones et al. 2006). In painful compared with normal tendon, a 5-fold increase in expression was observed, while a 6-fold increase was noted in ruptured compared with normal tendon (Jones et al. 2006).

6.1.4 Chapter Aims

The specific aim of this chapter was to determine whether there were differences in promoter DNA methylation profiles within the *ADAMTS4*, *TIMP1* and *MMP11* genes in patellar tendinopathy and healthy patellar tendon tissue.

6.2 METHODS

An overview of the detailed methodology related to this chapter is detailed in Chapter 2. The methods below are specific to this chapter and are described in short.

6.2.1 Participants

DNA from 20 tissue samples from the patellar tendon of male participants were collected for this study. These samples consisted of 10 normal healthy samples (CON) from individuals undergoing ACL reconstruction surgery and 10 abnormal samples from individuals undergoing surgery for patellar tendinopathy (Parkinson et al. 2010). Ethical approval was gained from the relevant research ethics committees, details of which can be seen in Appendix I.

6.2.2 DNA Collection and Bisulphite Treatment

DNA from patellar tendon samples was isolated by members of staff at La Trobe University, using the PureLink Genomic DNA Kit (Invitrogen, Carlsbad, CA, USA). Bisulfite treatment of target DNA converted the unmethylated cytosine residues into uracil. This process was critical for the accurate determination of methylation status. The methylated cytosines remained unchanged and this resulted in different DNA sequences for methylated and unmethylated DNA.

6.2.3 Loci Selection

The purpose of this study was to identify whether DNA methylation within the *ADAMTS4, TIMP1* and *MMP11* genes were altered in tissue from patellar tendinopathy (PT) compared to healthy control tissue (CON). To do this, part of the gene promoter of each gene was sequenced using Pyrosequencing.

6.2.3.1 *ADAMTS4* Hs_NDUFS2_01_PM Assay

The Hs_NDUFS2_01_PM PyroMark CpG Assay (Qiagen) was selected in the promoter of the *ADAMTS4* gene (**Figure 6.1**). The location of the *ADAMTS4* gene is at Chr.1:161,154,098-161,168,846 (14.75 kb in length) in Assembly GRCh37.p13 (www.ensembl.org). The location of Hs_NDUFS2_01_PM is at Chr.1:161,171,810-161,171,848 (39 bp in length) in Assembly GRCh37.p13. This is the precise location in the human genome of the sequence analyzed by the assay. (www.ensembl.org). The last bp of the assay is approximately 3 kb (2964 bp) from the start site of the *ADAMTS4* gene. Data obtained from Gene Cards Human Gene Database (www.genecards.org) and NCBI Gene (www.ncbi.nlm.nih.gov/gene).

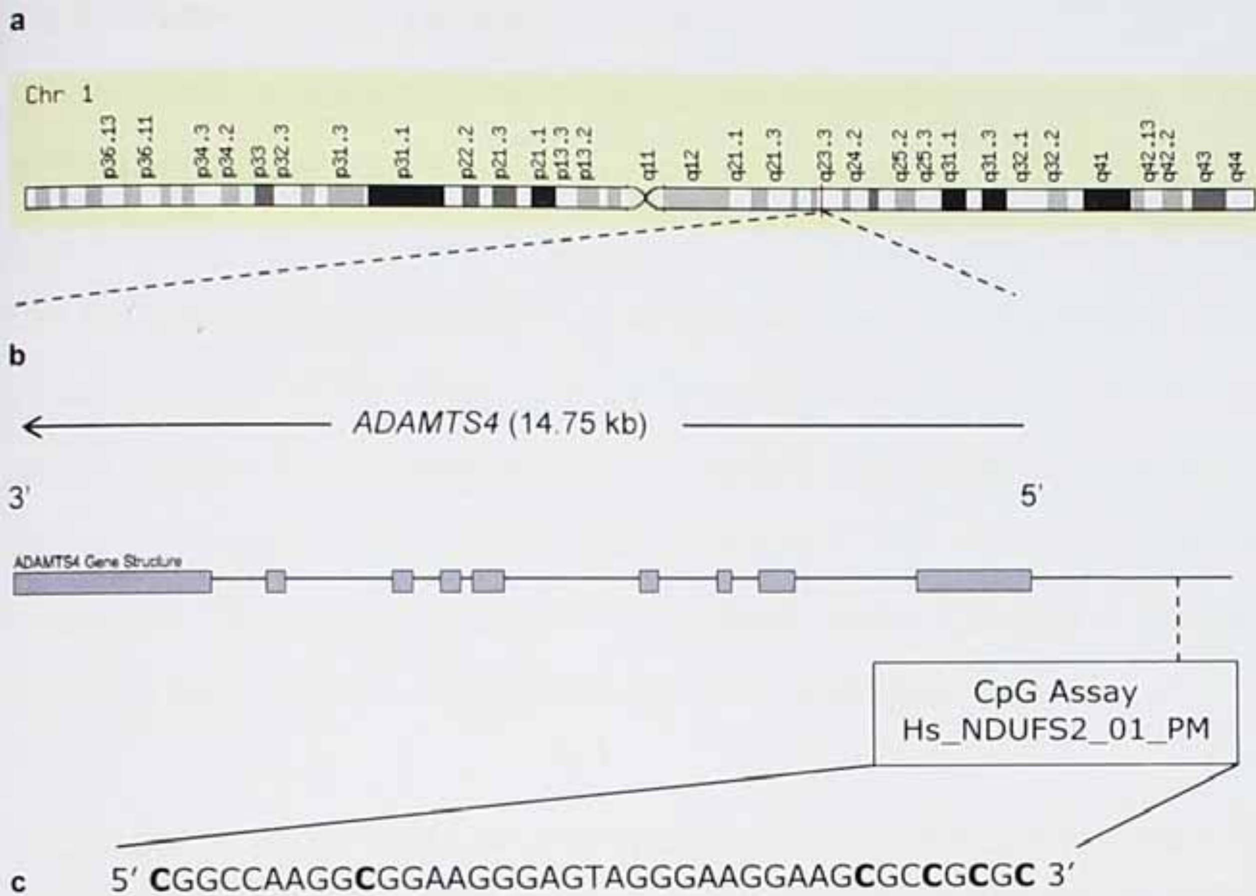


Figure 6.1 Location of the Hs_NDUFS2_01_PM CpG Assay within the *ADAMTS4* gene

(**a**) Chromosomal location of the *ADAMTS4* gene (Chr.1:161,154,098-161,168,846). (**b**) Size and exonic structure of the *ADAMTS-4* gene with location of the Hs_NDUFS2_01_PM PyroMark CpG Assay (Chr.1:161,171,810-161,171,848) in the promoter region, 2964 bp from the start site of the gene. (**c**) Nucleotide sequence of the Hs_NDUFS_01_PM PyroMark CpG Assay with CpG sites in bold. Compiled from www.ensembl.org, www.qiagen.com and www.lifetechnologies.com.

6.2.3.2 *TIMP1* Custom Designed Assays

Two PyroMark CpG Assays (Qiagen) were custom designed in the promoter of the *TIMP1* gene (**Figure 6.2**). The location of the *TIMP1* gene is at Chr.X:47,441,712-47,446,188 (4.48 kb in length) in Assembly GRCh37.p13 (www.ensembl.org). CpG Island Searcher (www.cpgislands.usc.edu) was used to screen a 10 kb region of the promoter of the *TIMP1* gene (47,431,427-47,441,644) to identify CpG islands present. A CpG island of 200 bp in length was identified in CpG Island Searcher, located at Chr.X:47,432,856-47,433,055. This region was used to design the Pyrosequencing assays in PyroMark Assay Design 2.0 software (Qiagen). Below is the nucleotide sequence of the CpG island selected.

ATCTAAGGGTGAGGTGTCTAGGAGGGCATAGAAAGGGCTTGGACCAAGGACAGGCTG
TTGGGCAAGTCAGCCTAAGGCAGAGCAGCTGGGTGGGGCAGGG**AGGACGTCTGAGC**
TGGGATGGGAGTTGGAATGGACCAGAAAAGGCAGATGGGTGATGATTCGCGGG
GAGACTGAGATGGGCCAGAAGGATGTGAAGGAGG/AGGAAAAGTATGAGGAT
GGAGTGGGAAGAGCCAGATGATAGGAGGACTAGGACTCGGGCCAGGAGTTTTA
GGGATGGGTTGGTGGCGGGGGGGCGAGTTTGGCAGGGAAGGCTCTGGGCTGGGC
AGGAAGTTGCAGAGGGTGGGTCAGCCAGGTGCCTGGGAGGCCCTGGAGGAGGAAACT
TTCCCAGGGCCTTTGAGGA

The CpG island identified is indicated in bold, with 100 bp of sequence included either side for designing primers in the PyroMark Assay Design 2.0 software (Qiagen). The / indicates the distinction between the two 100 bp regions that assays were designed for (Region A and Region B respectively). The CpG island is approximately 8.8 kb from the start site of the *TIMP1* gene. Data obtained from Gene Cards Human Gene Database (www.genecards.org) and NCBI Gene (www.ncbi.nlm.nih.gov/gene).

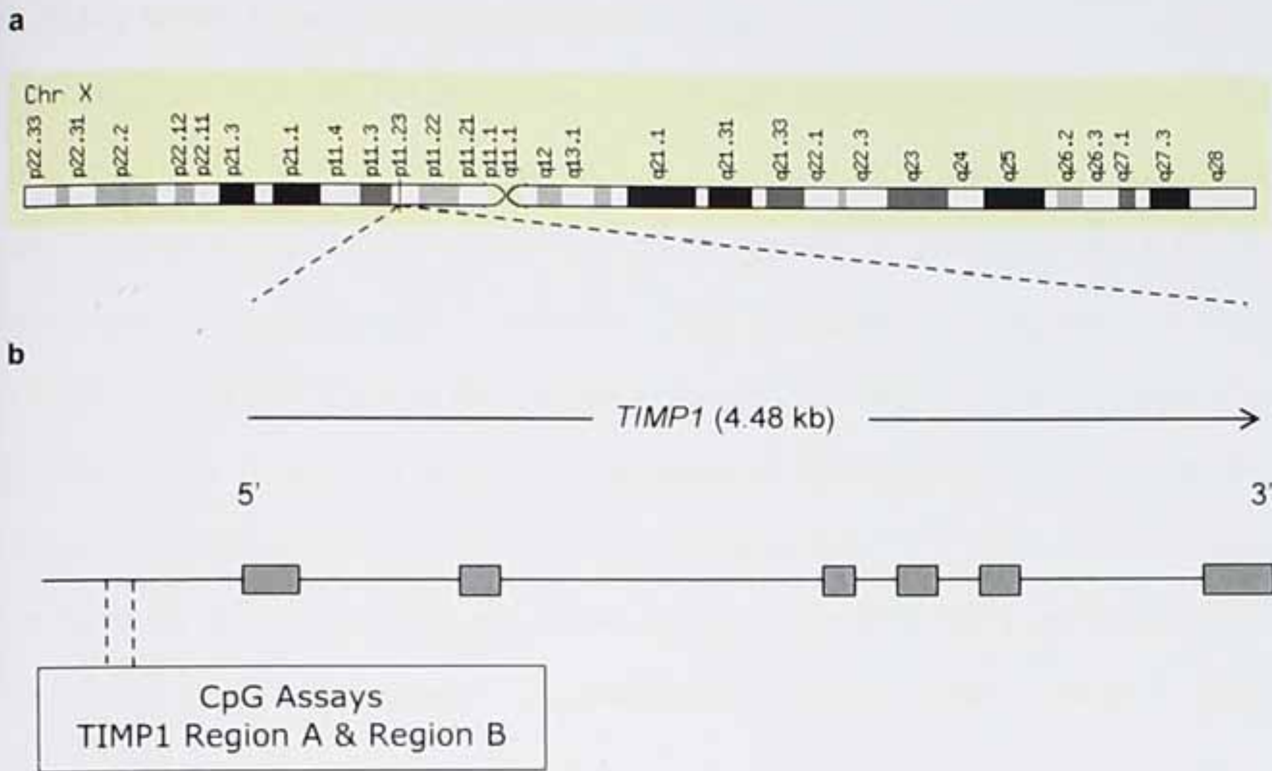


Figure 6.2 Location of the Custom Designed CpG Assays within the *TIMP1* gene

(**a**) Chromosomal location of the *TIMP1* gene (Chr.X:47,441,712-47,446,188). (**b**) Size and exonic structure of the *TIMP1* gene with location of the Custom PyroMark CpG Assays. Custom Designed PyroMark CpG Assays were designed in a CpG island (Chr.X:47,432,856-47,433,055) in the promoter region, 8.8 kb from the start site of the *TIMP1* gene. Compiled from www.ensembl.org, www.qiagen.com and www.lifetechnologies.com.

6.2.3.3 *MMP11* Hs_CHCHD10_01_PM Assay

The Hs_CHCHD10_01_PM PyroMark CpG Assay (Qiagen) was selected in the promoter of the *MMP11* gene (**Figure 6.3**). The location of the *MMP11* gene is at Chr.22:24,110,413-24,126,503 (16.09 kb in length) in Assembly GRCh37.p13 (www.ensembl.org). The location of Hs_CHCHD10_01_PM is at Chr.22:24,110,347-24,110,382 (36 bp in length) in Assembly GRCh37.p13. This is the precise location in the human genome of the sequence analyzed by the assay. (www.ensembl.org). The last bp of Hs_CHCHD10_01_PM is approximately 30 bp from the start site of the *MMP11* gene. Data obtained from Gene Cards Human Gene Database (www.genecards.org) and NCBI Gene (www.ncbi.nlm.nih.gov/gene).

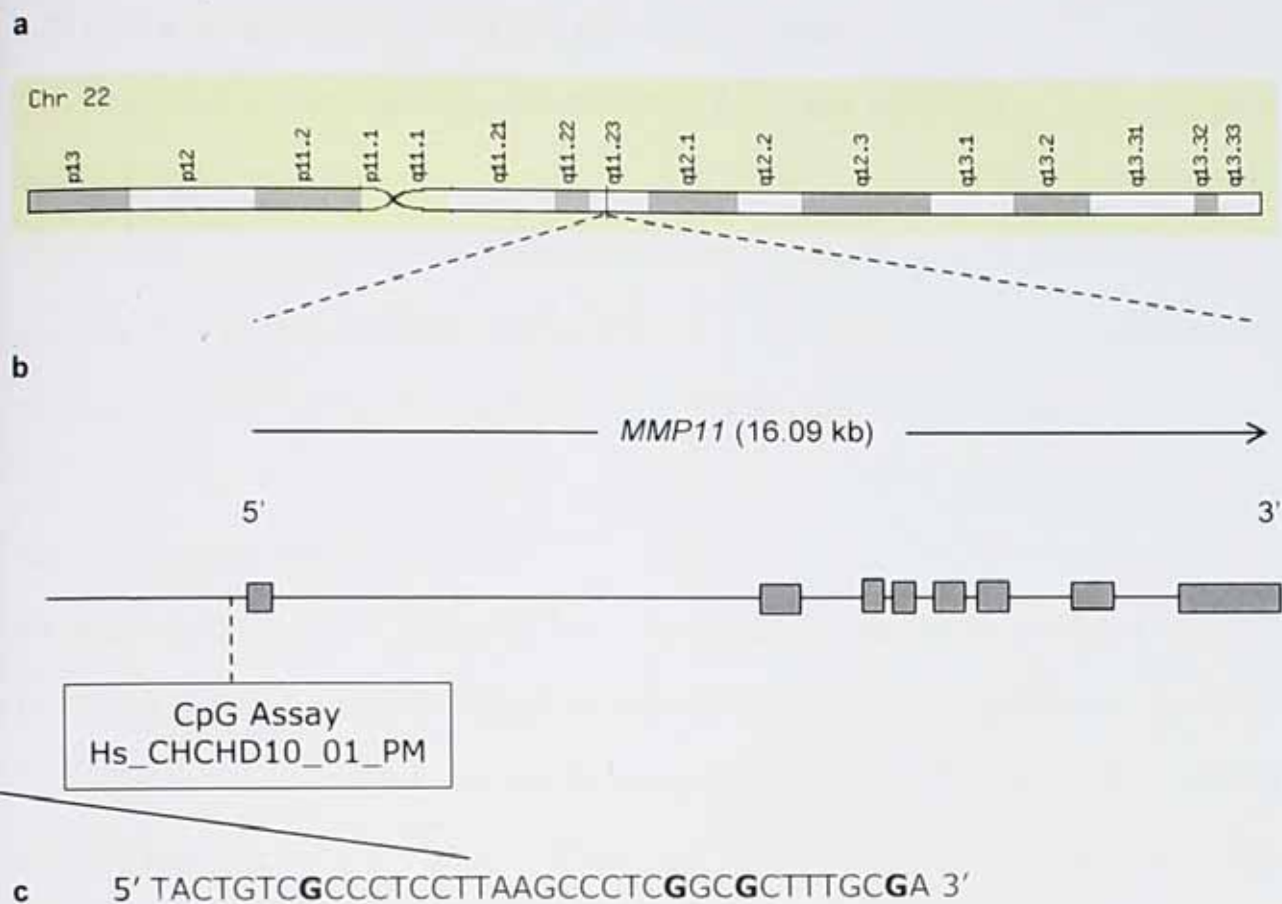


Figure 6.3 Location of the Hs_CHCHD10_01_PM CpG Assay within the *MMP11* gene

(**a**) Chromosomal location of the *MMP11* gene (Chr.22: 24,110,413-24,126,503). (**b**) Size and exonic structure of the *MMP11* gene with location of the Hs_CHCHD10_01_PM PyroMark CpG Assay (Chr.22: 24,110,347-24,110,382) in the promoter region 30 bp from the start site. (**c**) Nucleotide sequence of the Hs_CHCHD10_01_PM PyroMark CpG Assay with CpG sites in bold. Compiled from www.ensembl.org, www.qiagen.com and www.lifetechnologies.com.

6.2.4 PCR Amplification and Gel Electrophoresis

The target regions of interest were amplified using the PyroMark PCR Kit (Qiagen) and PCR was performed on a Techne TC-512 thermal cycler (Bibby Scientific Ltd, Staffordshire, UK). Product amplification was checked prior to Pyrosequencing analysis by running samples (along with a 1 kb ladder) on a 1% agarose gel stained with SYBR Safe DNA gel stain (Invitrogen, Carlsbad CA, USA).

6.2.5 Pyrosequencing

A Qiagen Q24 PyroMark (Qiagen) was used to sequence the target gene regions. The Pyrosequencing method enabled quantification of DNA methylation status at CpG sites in the promoter regions of the selected genes. This provided accurate and detailed profiles of DNA methylation patterns. The peak heights in the Pyrograms generated by the PyroMark Q24 software illustrated the proportion of methylated DNA at each site. See **Figure 6.7**, **6.8** and **6.9**.

6.2.6 Data Analysis

Differences in DNA methylation (%) between the patellar tendinopathy and control groups were analysed using an independent samples t-Test, with significance at $p < 0.05$.

6.3 RESULTS

The CON and PT groups in this study were matched for gender as all participants were male. The groups were similarly matched for age, with no significant difference between the CON (25.1 ± 6.4) and PT (23.2 ± 4.4) groups ($p = 0.449$). Height and weight data were not available.

The PCR reactions conducted prior to Pyrosequencing were successful, as can be seen in the gel images in **Figure 6.4**, **6.5** and **6.6**.

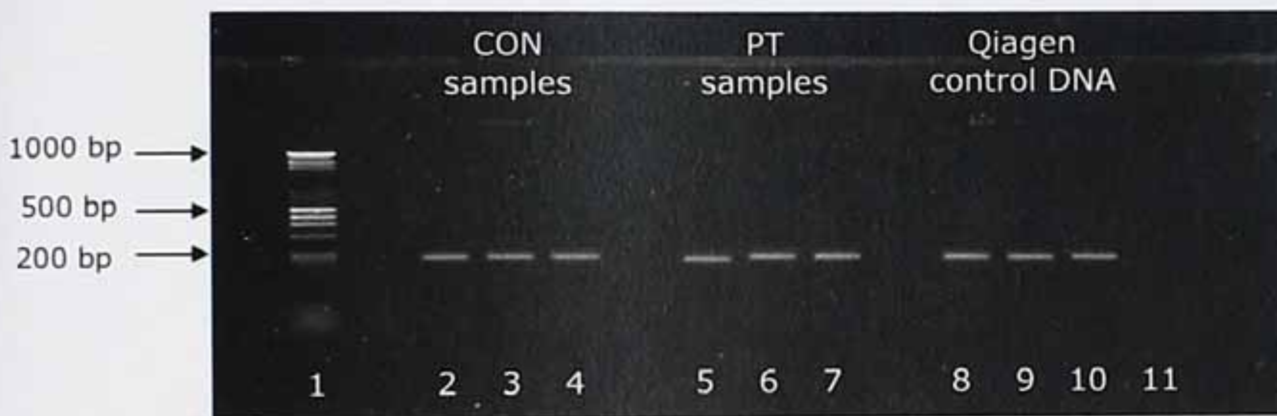


Figure 6.4 Image of *ADAMTS4* PCR product from a 1% agarose gel electrophoresis confirming the amplification of PCR product. Well 1, 1 kb ladder; wells 2-4, CON samples; wells 5-7, PT samples; well 8, hypermethylated control; well 9, hypomethylated control; well 10, bisulfite converted hypomethylated control; well 11, no template control (negative control).

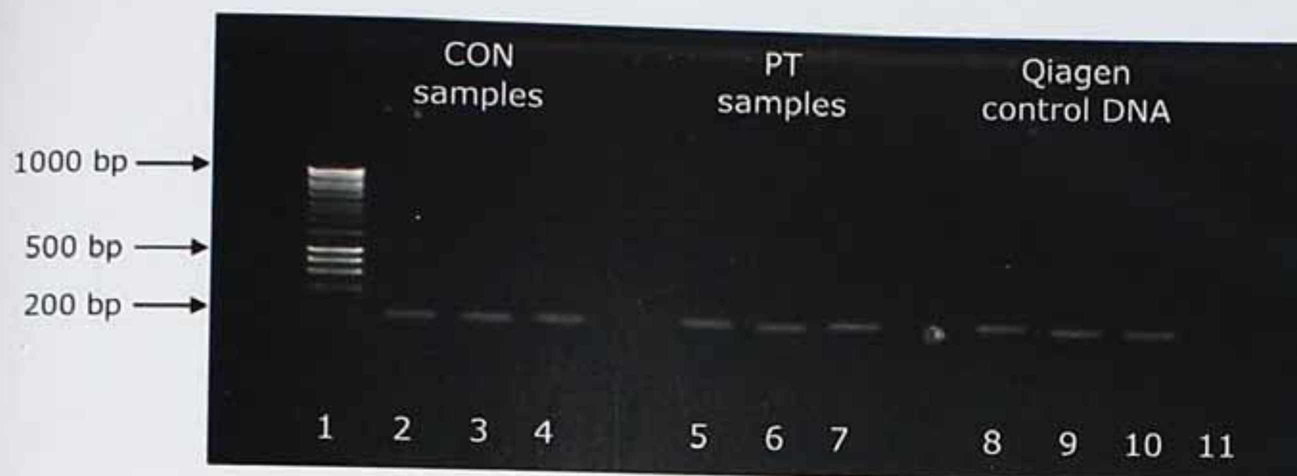


Figure 6.5 Image of *TIMP1* PCR product from a 1% agarose gel electrophoresis confirming the amplification of PCR product. Well 1, 1 kb ladder; wells 2-4, CON samples; wells 5-7, PT samples; well 8, hypermethylated control; well 9, hypomethylated control; well 10, bisulfite converted hypomethylated control; well 11, no template control (negative control).

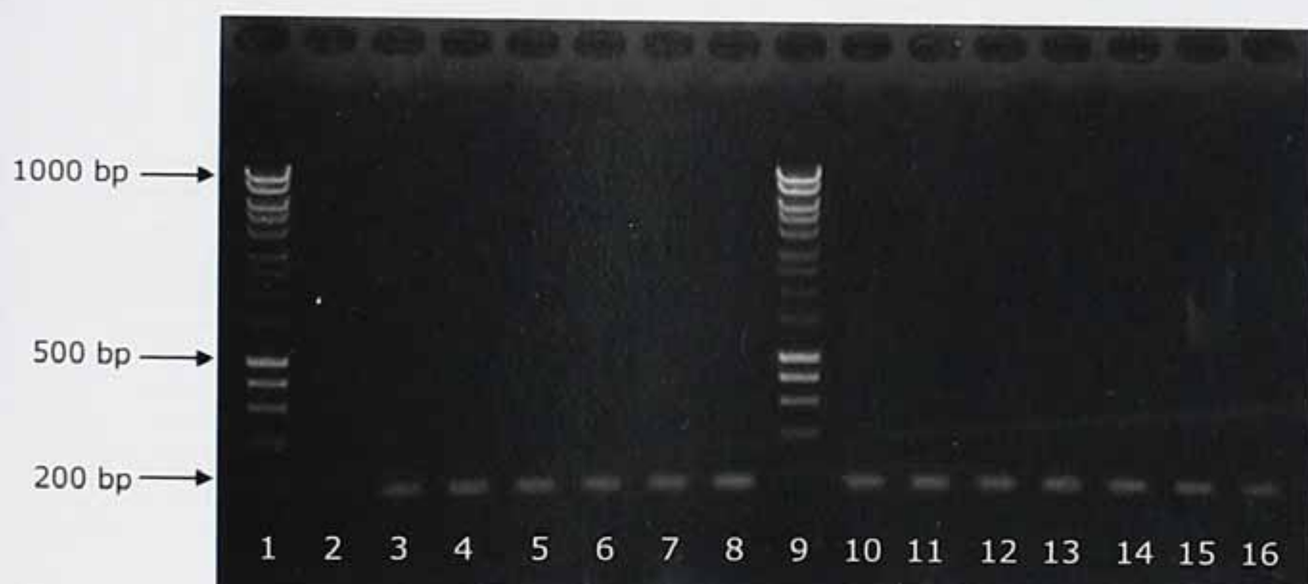


Figure 6.6 Image of *MMP11* PCR product from a 1% agarose gel electrophoresis confirming the amplification of PCR product. Well 1, 1 kb ladder; well 2, no template control (negative control); wells 3-8, CON samples; well 9, 1 kb ladder; wells 10-13, PT samples; well 14, hypermethylated control; well 15, hypomethylated control; well 16, bisulfite converted hypomethylated control.

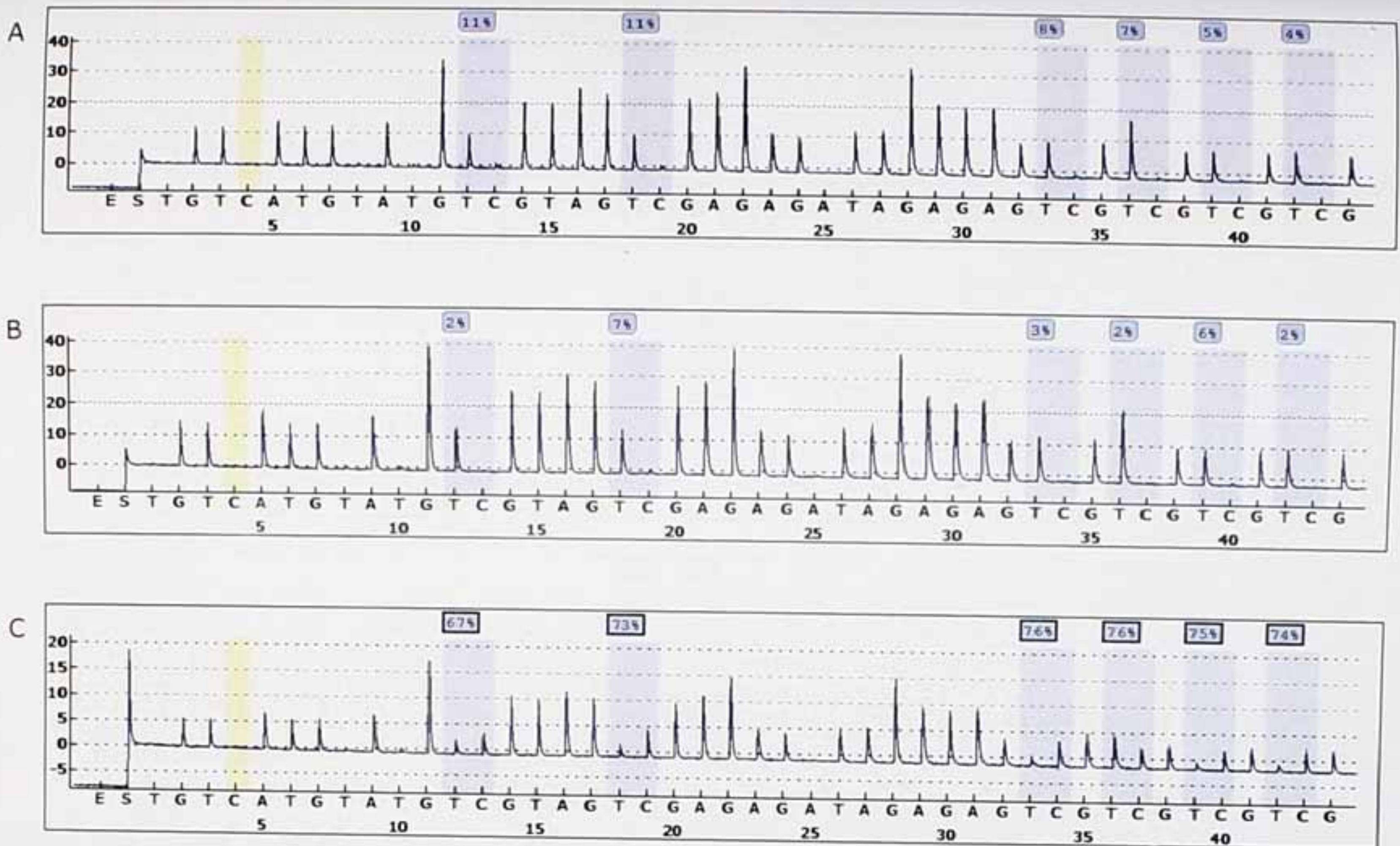


Figure 6.7 Typical Pyrogram graphs for the *ADAMTS4* gene (Hs_NDUFS2_01_PM PyroMark CpG Assay) showing CpG sites shaded in blue with the percentage (%) methylation above each site. A) Pyrogram of PT sample. B) Pyrogram of CON sample. C) Pyrogram of hypermethylated control.

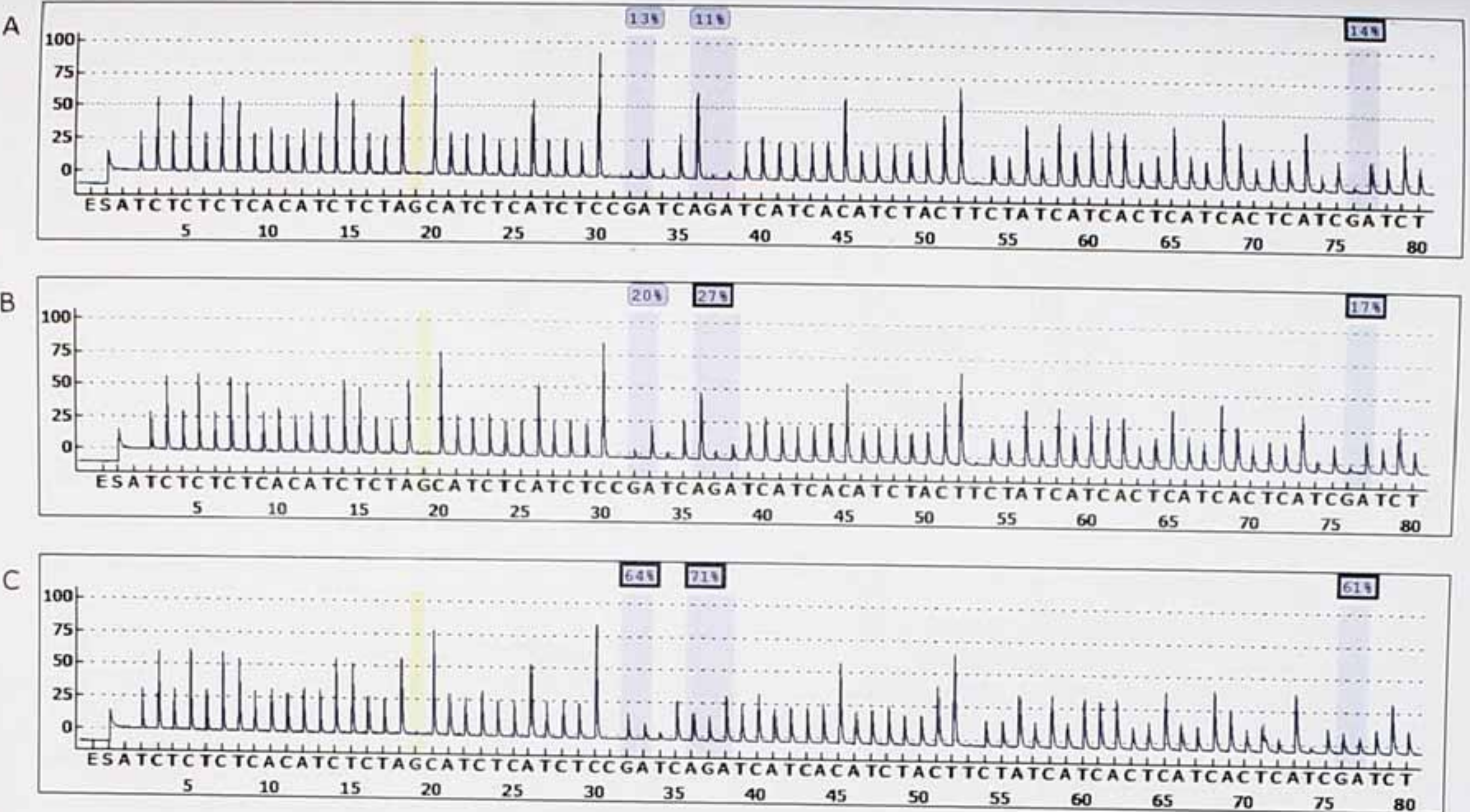


Figure 6.8 Typical Pyrogram graphs for the *TIMP1* gene (Custom Designed PyroMark CpG Assay) showing CpG sites shaded in blue with the percentage (%) methylation above each site. A) Pyrogram of PT sample. B) Pyrogram of CON sample. C) Pyrogram of hypermethylated control.

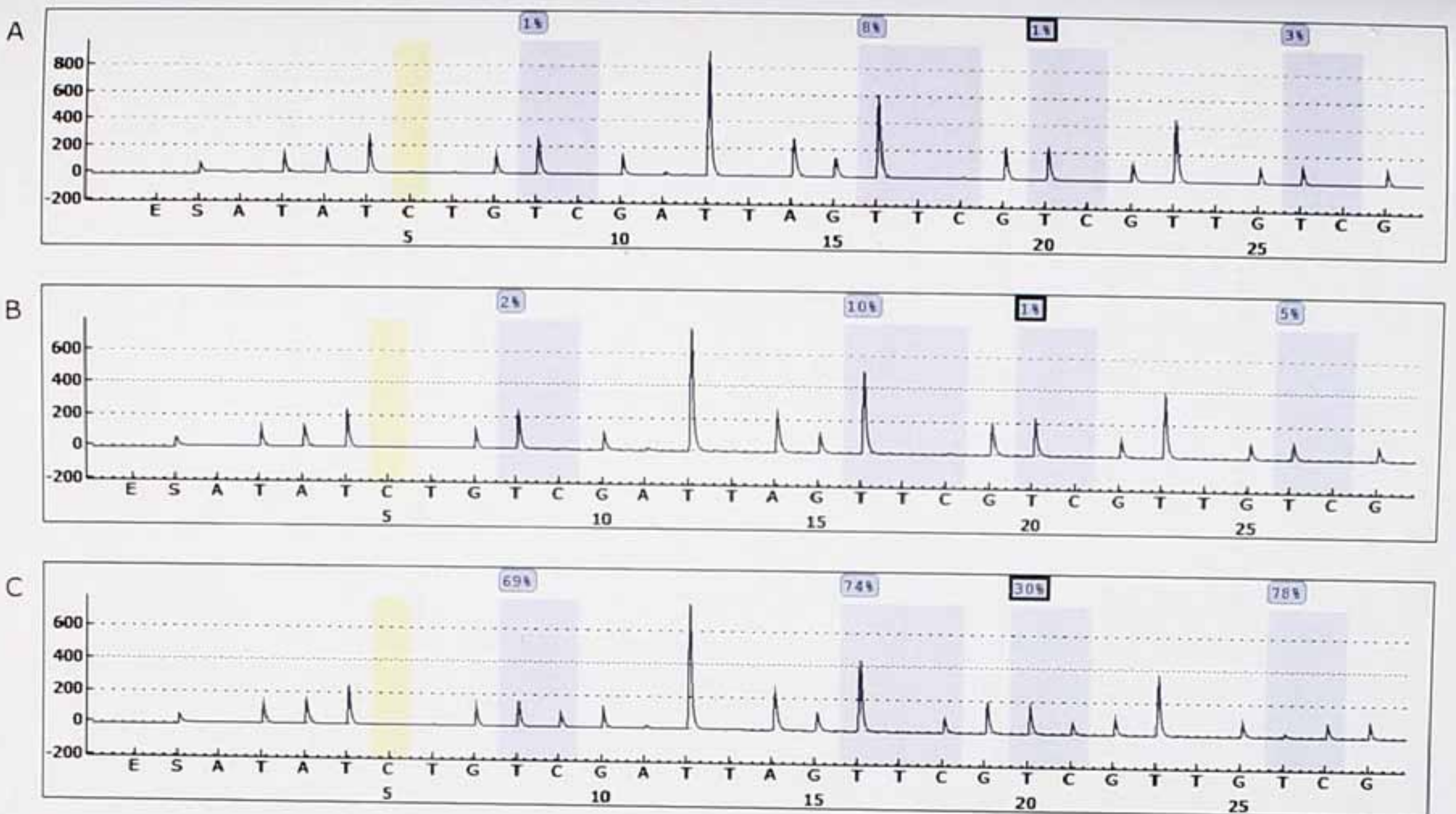


Figure 6.9 Typical Pyrogram graphs for the *MMP11* gene (Hs_CHCHD10_01_PM PyroMark CpG Assay) showing CpG sites shaded in blue with the percentage (%) methylation above each site. A) Pyrogram of PT sample. B) Pyrogram of CON sample. C) Pyrogram of hypermethylated control.

6.3.1 *ADAMTS4* CpG Methylation Status

Methylation status was measured within 6 CpG sites of the *ADAMTS4* gene promoter (**Figure 6.7**). While five of the sites showed no significant difference in mean CpG site methylation, one site was significantly different between the CON and PT groups (**Table 6.1**). Mean methylation in the PT samples (5.1 ± 0.6, n=10) was significantly higher (p = 0.016) than mean methylation of the CON samples (2.8 ± 0.6, n=9) (**Table 6.1** and **Figure 6.10**).

Table 6.1 Mean methylation percentages within CpG sites of the *ADAMTS4* gene promoter in Patellar tendinopathy (PT) cases and asymptomatic controls (CON).

ADAMTS4	CON	PT	p-value
CpG 1	2.6 ± 0.6 (10)	5.2 ± 1.3 (10)	0.086
CpG 2	5.4 ± 0.9 (10)	6.3 ± 1.0 (10)	0.524
CpG 3	2.8 ± 0.6 (9)	5.1 ± 0.6 (10)	**0.016**
CpG 4	4.0 ± 0.5 (9)	3.4 ± 0.9 (10)	0.594
CpG 5	3.2 ± 0.8 (9)	2.6 ± 0.7 (10)	0.555
CpG 6	2.1 ± 0.4 (9)	2.7 ± 0.5 (10)	0.409
Average	3.4 ± 0.4 (10)	4.2 ± 0.7 (10)	0.294

Differences in mean DNA methylation between the CON and PT groups were analysed with an independent samples t-Test. Values represent percentages (%) ± standard error of the mean (SEM).

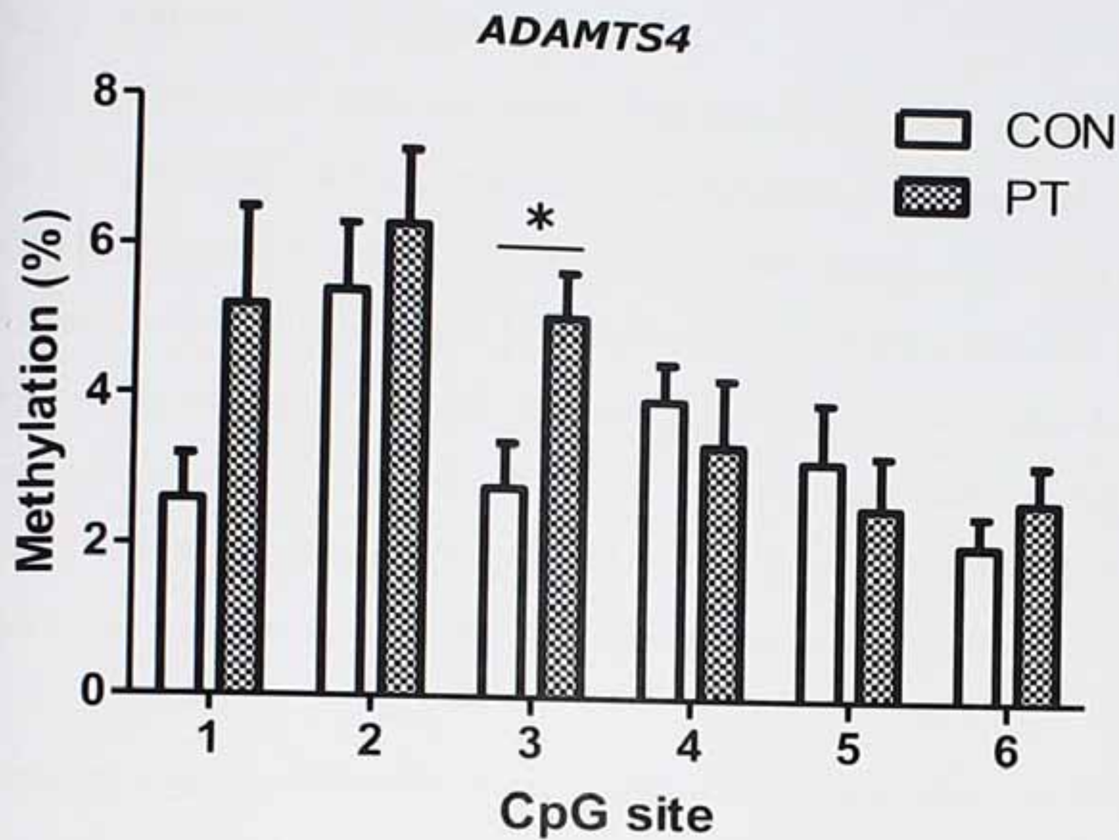


Figure 6.10 Comparison of the mean DNA methylation between CON and PT at 6 CpG sites within the promoter of the *ADAMTS4* gene. Bars represent mean (%) ± standard error of the mean (SEM). CON, white bars and PT, shaded bars.*, p = 0.016.

6.3.2. *TIMP1* CpG Methylation Status

For the *TIMP1* gene promoter, methylation status was measured within 3 CpG sites (**Figure 6.8**) using a Custom Designed PyroMark CpG Assay (Region A). The Custom Designed PyroMark Assay for Region B did not sequence and is therefore not mentioned in these results. There was a significant difference (p = 0.012) in methylation at one CpG site between the CON and PT groups (**Table 6.2**). Mean methylation in the PT samples (20.2 ± 2.0, n=9) was significantly higher than the mean methylation of the CON samples (12.6 ± 1.7, n=8) (**Table 6.2** and **Figure 6.11**). No data was obtained for the third CpG site pyrosequenced.

Table 6.2 Mean methylation percentages within CpG sites of the *TIMP1* gene promoter (custom assay region A) in Patellar tendinopathy (PT) cases and asymptomatic controls (CON).

***TIMP1* Region A**	CON	PT	p-value
CpG 1	14.6 ± 1.3 (10)	15.9 ± 1.8 (9)	0.556
CpG 2	12.6 ± 1.7 (8)	20.2 ± 2.0 (9)	**0.012**
CpG 3	-	-	-
Average	14.0 ± 1.4 (10)	18.1 ± 1.9 (9)	0.092

Differences in mean DNA methylation between the CON and PT groups were analysed with an independent samples t-Test. Values represent percentages (%) ± standard error of the mean (SEM).

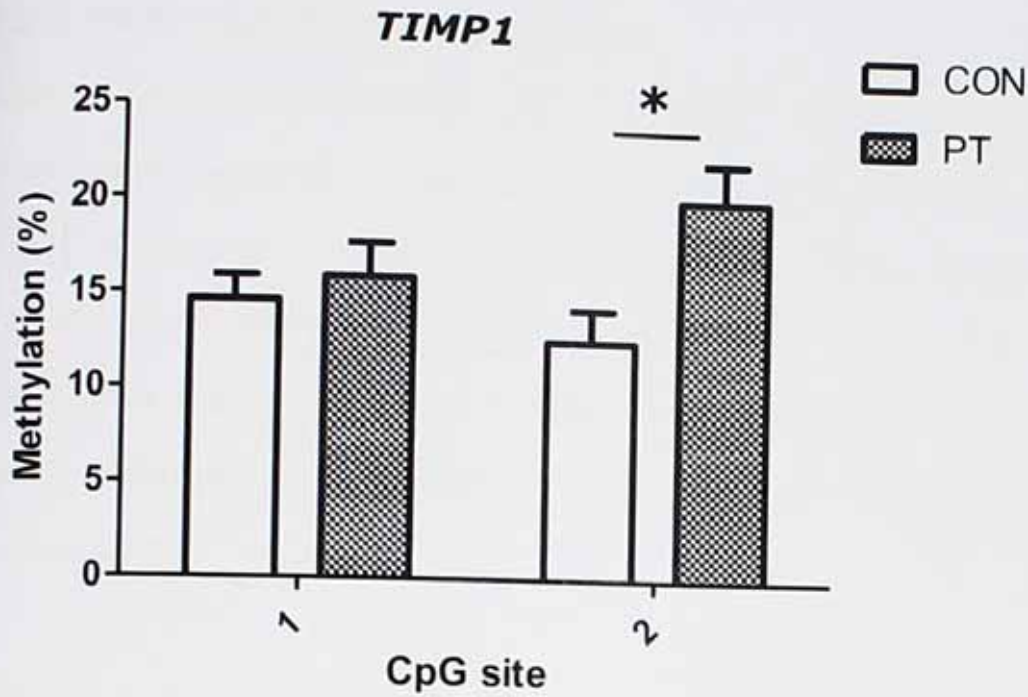


Figure 6.11 Comparison of the mean DNA methylation between the CON and PT groups within the promoter of the *TIMP1* gene. Bars represent mean (%) ± standard error of the mean (SEM). CON, white bars and PT, shaded bars.*, p = 0.012.

6.3.3 *MMP11* CpG Methylation Status

When the methylation status was measured within 4 CpG sites of the *MMP11* gene promoter (**Figure 6.9**), the difference in methylation at one site between the CON and PT groups reached statistical significance (p = 0.045) (**Table 6.3**). At this CpG site, mean methylation in the PT samples (4.8 ± 0.4, n=10) was found to be significantly higher than the mean methylation of the CON samples (3.6 ± 0.4, n=10) (**Table 6.3** and **Figure 6.12**). No data was obtained for the third CpG site pyrosequenced.

Table 6.3 Mean methylation percentages within CpG sites of the *MMP11* gene promoter in Patellar tendinopathy (PT) cases and asymptomatic controls (CON).

MMP11	CON	PT	p-value
CpG 1	2.1 ± 0.4 (10)	2.4 ± 0.3 (10)	0.578
CpG 2	9.1 ± 0.7 (10)	9.2 ± 0.9 (10)	0.933
CpG 3	-	-	-
CpG 4	3.6 ± 0.4 (10)	4.8 ± 0.4 (10)	**0.045**
Average	4.9 ± 0.3 (10)	5.5 ± 0.5 (10)	0.381

Differences in mean DNA methylation between the CON and PT groups were analysed with an independent samples t-Test. Values represent percentages (%) ± standard error of the mean (SEM).

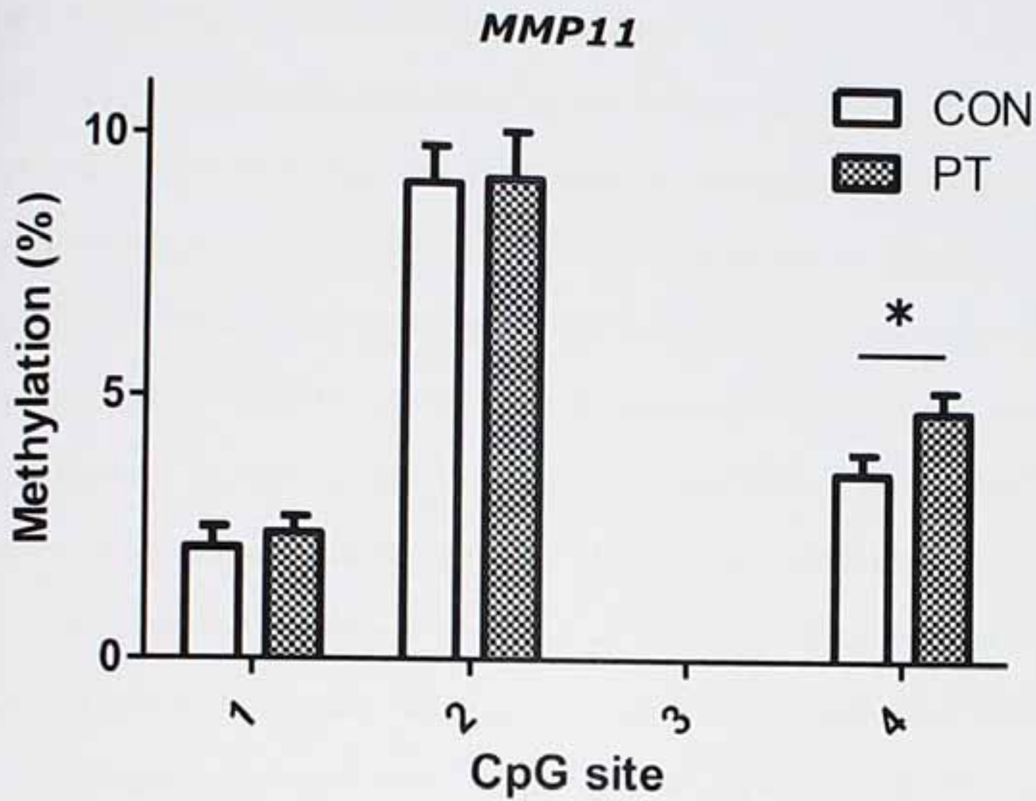


Figure 6.12 Comparison of the mean DNA methylation between the CON and PT groups within the promoter of the *MMP11* gene. Bars represent mean (%) ± standard error of the mean (SEM). CON, white bars and PT, shaded bars.*, p = 0.045.

6.4 CHAPTER SPECIFIC CONCLUSIONS

There was a significant difference (p = 0.016) in DNA methylation between the CON and PT group at CpG site 3 of the *ADAMTS4* gene (**Table 6.1**). Specifically, mean DNA methylation was significantly higher in the PT group compared to the CON group (**Figure 6.10**). Higher expression levels of the *ADAMTS4* gene have previously been detected in ruptured Achilles tendon compared with normal and tendinopathic tendon (Corps et al. 2008), showing that the *ADAMTS4* gene may have a role in tendon pathology. Our data, showing higher mean methylation at one CpG site, is in contrast to previous work into musculoskeletal conditions, where CpG sites as far as 800 bp upstream of the *ADAMTS4* gene were demethylated in osteoarthritis samples (Roach et al. 2005). However, increased accumulation of the proteoglycan aggrecan has been observed in tendinopathy, along with a decrease in levels of ADAMTS4 (Smith et al. 2008). These findings highlight interest in the results from this study, as the increased DNA methylation observed within the *ADAMTS4* gene promoter in the PT group may be resulting in a reduction in the levels of ADAMTS4 protein. As such, this could be causing a build-up of ADAMTS4 substrates, including aggrecan, disrupting the structural integrity of the tendon.

This study found that DNA methylation at one CpG site within the *TIMP1* gene was significantly different (p = 0.012) between the CON and PT groups. Specifically, mean methylation in the PT samples (20.2 ± 2.0, n=9) was found to be significantly higher than in the CON samples (12.6 ± 1.7, n=8). (**Figure 6.11**). Altered expression of the *TIMP1* gene has been shown in ruptured compared with normal Achilles tendon (Jones et al. 2006) and epigenetic mechanisms such as CpG methylation may well have an important role in tendon pathology by affecting

levels of gene expression. While an increase in *TIMP1* gene expression has been reported in ruptured compared with normal Achilles tendon (Jones et al. 2006), the data from this study may infer that the increased DNA methylation is causing a reduction in TIMP1 protein. This could lead to a build-up of substrates that TIMP1 would typically cleave and a subsequent disruption of tendon homeostasis, as is seen in tendon pathology.

This study also found a significant difference in DNA methylation at one CpG site within the *MMP11* gene between the CON and PT groups ($p = 0.045$). Mean methylation in the PT samples (4.8 ± 0.4, n=10) was found to be significantly higher than the CON samples (3.6 ± 0.4, n=10). (**Figure 6.12**). It is postulated that these differences in DNA methylation may influence gene expression and the risk of tendon pathology. Indeed, although the mechanisms are not fully defined, differences in expression levels of the *MMP11* gene have been reported in pathologic Achilles tendon (Jones et al. 2006). Increased expression of the *MMP11* gene was observed in both painful compared with normal tendon, as well as ruptured compared with normal tendon (Jones et al. 2006). Conversely, hypermethylation of a gene promoter is typically associated with reduced expression of that gene (Roach et al. 2005). This suggests that a reduction in *MMP11* gene expression may in fact be apparent in the PT group, as a result of the increased DNA methylation measured at one CpG site. One may speculate that a reduction in *MMP11* gene expression, followed by a subsequent reduction in MMP11 protein, could alter the tendon homeostasis and lead to an accumulation of proteoglycan substrates that it would usually cleave (Somerville et al. 2003).

The preliminary results from this study suggest that methylation status of the *ADAMTS4*, *TIMP1* and *MMP11* genes may be important in relation to the risk of patellar tendinopathy. While significant differences in methylation status have been identified between the CON and PT groups, further work is needed to establish the functional relevance of these findings and whether they may influence gene expression directly.

CHAPTER 7: DISUSSION

7.1 Copy Number and SNP Genotyping Studies

Chapter 3 focused on CNV within the *COL5A1*, *COL4A1* and *TNC* genes. For the *COL5A1* Hs00180523_cn assay, there was a significant difference in discrete CN between the AUS CON and ATP groups (p = 0.012). However, although significant, this analysis was underpowered given the sample size (CON, n=69; ATP, n=44). Furthermore, this significance was not reported when the continuous CN data were analysed. Indeed, it has been speculated that CNV based on continuous calling can avoid the potential rounding errors encountered when pooling values into discrete groups (Park et al. 2012). As such, all analyses on copy number were undertaken on both the continuous and discrete data. There were no other significant differences in copy number for either the Hs00180523_cn or Hs02980647_cn assays within the *COL5A1* gene. While it appears that CNV within the *COL5A1* gene does not seem to play a role in the risk of ATP, Mokone et al. (2006) showed that the rs12722 SNP within the 3' untranslated region (UTR) of *COL5A1* gene was associated with ATP. It may be that CNV within other loci of the gene may be associated but have yet to be investigated. It is also conceivable that smaller-scale variants like the *COL5A1* rs12722 variant may play a more prominent role in the aetiology of ATP.

For the *COL4A1* gene, although no significant differences were found for all analyses on the Hs02119212_cn assay, there was a significant difference in discrete CN for the Hs00739915_cn assay between the combined CON and ATP (p = 0.014) and combined CON and TEN (p = 0.033) groups. This is an interesting finding and given the sample sizes, the analysis was sufficiently powered to detect significance. The *COL4A1* Hs00739915_cn assay spans intron 22 – exon 22 of the

gene. While the functional relevance of altered CN at this region is unknown, it may be that altered gene dosage may have an effect on the tendon and disrupt homeostasis of the ECM through a change in gene expression (Dear 2009; Yang et al. 2008). For example, it might be envisaged that an increase in COL4A1 protein could lead to disorganisation of the collagen fibril and an imbalance in the other collagens, potentially leading to tendon pathology from the change in structural integrity. Furthermore, mutations within the *COL4A1* gene have been shown to cause congenital muscular dystrophies in humans and mice (Labelle-Dumais et al. 2011). These heritable musculoskeletal conditions have serious health consequences including muscle weakness, myopathy, muscle necrosis and in severe cases, respiratory insufficiency (Labelle-Dumais et al. 2011). This shows the interest in investigating this gene in relation to other musculoskeletal conditions like ATP. In addition, altered expression levels of the *COL4A1* gene have also been identified in human tendinopathy, with an increase in expression observed compared to normal tendon (Jelinsky et al. 2011). Nevertheless, when the copy number data were analysed for the *COL4A1* Hs00739915_cn assay as continuous CN, the significance was not shown. This suggests that further replication of this study would be needed in additional samples and cohorts and highlights the importance of difference analysis methods in CNV studies.

The CNV data on the *TNC* gene presented in chapter 3 shows that there was limited variation in the distribution of copy number (ie. CN < 2 or CN > 2) for the Hs00634176_cn assay within intron 23 - exon 23 across cases and controls of all cohorts. Namely, there was no significant different in copy number between the combined CON and ATP ($p = 0.829$), nor between the UK CON vs. ATP ($p = 1.000$) and AUS CON vs. ATP ($p = 0.347$). This was also the case for analyses concerning

the CON vs. TEN and CON vs. RUP subgroups. Statistical analysis on the discrete CN data for the SA cohort was not possible, as all participants had discrete CN values of CN = 2. When these data were analysed as continuous CN, there was no significant differences (data not shown) and little variation between continuous CN values (SA CON, 1.97 ± 0.09; SA ATP, 1.97 ± 0.10; SA TEN, 1.97 ± 0.11; SA RUP, 1.98 ± 0.08). Only 1 UK control participant was reported to have a CN > 2 at this locus. Similarly, only 2 AUS control participants and 3 AUS ATP participants had a CN < 2. As such, no significant differences were found for analyses on both the discrete and continuous CN data between cases controls of individual cohorts and when combined. While CNV is reported within the Database of Genomic Variants at this locus, we have observed little variation within the cohorts we investigated. There were also no significant differences in copy number for the Hs06903309_cn assay within intron 15 of the *TNC* gene. Despite this, there was a higher frequency of participants with a CN > 2. Across all cohorts, a total of 12 control and 12 ATP participants had a CN > 2. No control or ATP participants had a CN < 2 at this locus. The *TNC* gene was the first tendon related gene that was found to be associated with ATP (Mokone et al. 2005), however, it appears that CNV within the two loci investigated in the *TNC* gene are not risk factors for ATP. It is important to note that although CNV may be present within a gene, it does not necessarily result in a phenotypic consequence (Dear 2009). As CNV accounts for a large proportion of the structural variation within the genome, most of the variants have little or no phenotypic effect (Redon et al. 2006; Dear 2009). If all the CNV within the genome did have an effect, they would be classed as mutations and most individuals would be affected with certain disease traits. It seems that CNVs that have a deleterious or harmful effect on the genome are largely selected out of the population (Dear 2009). Nevertheless a large quantity of CNV still exists

within the genome and research has shown the importance of this large-scale variation in relation to complex diseases (Estivill and Armengol 2007). Furthermore, as well as disease associated CNV, CNV may also offer a phenotypic advantage and benefit to the genome in some way (Zhang et al. 2009; Hastings et al. 2009).

In chapter 4, CNV was investigated within the *MMP3*, *MMP23B* and *ADAM8* genes. When the combined CON vs. ATP, CON vs. TEN and CON vs. RUP groups were compared, there were no significant differences in copy number for either loci investigated (intron 4–exon 4 and exon 7) within the *MMP3* gene (assays Hs02276714_cn and Hs02908568_cn respectively). This lack of significance was also apparent when the UK and SA cohorts were analysed separately. However, when we measured CNV within exon 7 of the *MMP3* gene (Hs02908568_cn assay) in the AUS cohort, there was a significant difference in copy number within the between the AUS CON and ATP group ($p = 0.035$) and between the AUS CON and TEN group ($p = 0.035$). While analysis of the continuous copy number data showed a lack of significance and the low sample size (CON, 87; ATP, 33; TEN, 33) reduced the power of these analyses, this preliminary finding still warrants further investigation. The Database of Genomic Variants contained no current data on CNV within the *MMP3* gene at the two exonic loci we investigated, although CNV is present within other regions of the gene. Despite this, we identified variation in copy number at both of these loci (intron 4–exon 4 and exon 7). As new CNVs are continuously being discovered and the extent of CNV within the genome is not yet fully defined, the predicted frequency of a given CNV derived from databases alone might incorporate a degree of error (Pinto et al. 2007; Grozeva et al. 2012). As such, a lack or reported CNV within a gene region does

not rule it out as a plausible candidate for investigation. Indeed, we identified the presence of CNV in a region previously un-catalogued. Despite the preliminary findings on CNV within exon 7 of the *MMP3* gene, it appears that small-scale genetic variants, namely SNPs, play a more prominent role in the risk of ATP. In 2009, the rs679620, rs591058 and rs650108 SNPs within the *MMP3* gene were all found to associate with the risk of Achilles tendinopathy (Raleigh et al. 2009). The rs679620 variant is a non-synonymous SNP that lies within exon 2 of the *MMP3* gene. This means that the rs679620 SNP has an effect of the protein sequence, due to a change in amino acid. Due to the functional effect of this variant, it is likely that this could disrupt the balance of MMP3 protein within the tendon and modify the risk of ATP.

Chapter 4 also investigated CNV within the *MMP23B* gene. We found no significant differences in copy number between the combined CON and ATP ($p = 0.090$), the CON and TEN ($p = 0.387$) or the CON and RUP ($p = 0.436$) groups within exon 1 of the gene. In addition, there were no significant differences when the UK and SA cohorts were analysed separately. CNV within exon 23 of the *ADAM8* gene was also not found to associate with ATP. No significant differences in copy number were found between the combined CON and ATP ($p = 0.727$), the CON and TEN ($p = 0.707$) or the CON and RUP ($p = 1.000$). When the UK, SA and AUS cohorts were analysed separately, the lack of significance was concordant within these groups. Both the assays within the *MMP23B* and *ADAM8* gene lie within coding exonic regions, yet the functional relevance of CNV within these loci is not known. It appears that in ATP, they do not influence disease risk. In contrast to the lack of CNV reported in the Database of Genomic Variants for the *MMP3* gene, the *MMP23B* and *ADAM8* genes both contained CNV within the loci detected by the

assays. However, the majority of samples from our three cohorts gave a copy number of 2, with only a small number of individuals showing either a loss or gain in copy number. Interestingly, higher levels of MMP23 protein were detected in painful tendon compared with normal tendon (Jones et al. 2006; Schulze-Tanzil et al. 2011), showing that this gene is of interest in ATP. Nevertheless, it may be that CNV within other regions of the *MMP23B* gene (other than the region within exon 1 that we investigated here) may be responsible, or that other genetic variants may be involved. This may also be relevant for the *ADAM8* gene, as despite the increased expression identified in ruptured Achilles tendon compared to controls (Jones et al. 2006), we found CNV within exon 23 of the gene to not be associated with ATP. Other genetic variants may be at play here, influencing the expression of these genes.

Chapter 5 detailed the investigations into the *TNFRSF1A* rs4149577 and *CASP3* rs1049253 variants, as well as CNV within intron 11 – intron 12 of the *CASP8* gene. Neither the SNP variants within the *TNFRSF1A* and *CASP3* genes, nor CNV within the *CASP8* gene, were associated with ATP in our combined cohort. This was also the case when the UK, SA and AUS cohorts were analysed separately and when the ATP group was subdivided to the distinct TEN and RUP groups. Importantly, this thesis only investigated a single SNP within the *TNFRSF1A* gene (rs4149577). Despite the lack of association reported at this locus, TNFR1 does have a functional role in apoptosis and has been shown to be expressed at significantly higher levels in tenocytes from Achilles tendinosis compared to controls (Gaida et al. 2012). Therefore it is possible that other genetic variants within the DNA sequence, not yet investigated, may influence the risk of ATP. Moreover, the *TNFRSF1A* rs4149577 variant is an intronic SNP that is in linkage

disequilibrium (LD) with a number of other variants that have been associated with musculoskeletal and inflammatory diseases (www.ensembl.org) (Karaderi et al. 2012; Park et al. 2013; Swaminathan et al. 2010). Examples of these include the *TNFRSF1A* rs1800693 and rs4149578 variants, that have been associated with neuromyelitis optica and ankylosing spondylitis respectively (Karaderi et al. 2012; Park et al. 2013; Swaminathan et al. 2010). The second SNP variant to be covered in chapter 5 was the rs1049253 variant within the *CASP3* gene. While there were no significant differences identified between the cases and controls, the *CASP3* rs1049253 variant has been associated with certain types of cancer (Guan et al. 2013). Interestingly, the rs1049253 variant lies within the 3′ UTR of the *CASP3* gene and the CC genotype of this variant has been shown to alter levels of *CASP3* mRNA expression (Guan et al. 2013). In addition, the location of the rs1049253 variant is a known miRNA binding site, where miRNAs are able to bind to the mRNAs and regulate apoptosis (Guan et al. 2013). Regardless of the functional significance of this SNP, it seems that it is not a risk factor for the development of ATP.

As detailed within chapter 5, there was no significant difference in copy number within intron 11 – intron 12 of the *CASP8* gene between the combined CON and ATP group ($p = 0.402$), the UK CON vs. ATP ($p = 0.155$) group, the SA CON vs. ATP ($p = 0.327$) or the AUS CON vs. ATP group ($p = 1.000$). A lack of significance was also reported when the ATP groups were subdivided into TEN and RUP. SNP variants within the *CASP8* gene have previously been associated with Achilles tendinopathy (Nell et al. 2012), which is one reason why the *CASP8* gene was selected for CNV investigation in this thesis. Furthermore, the CNV we investigated with the Hs02601709_cn assay overlaps the location of the rs1045485 SNP

investigated by Nell et al. (2012). It appears that the *CASP8* rs1045485 variant may have a more profound impact on the development of ATP, compared with CNV at the locus we investigated. Indeed, while the functional effect of altered copy number within the locus we investigate is currently unknown, the *CASP8* rs1045485 variant is known to be a non-synonymous polymorphism that results in an amino acid change (September et al. 2012). This may affect interaction with other proteins (September et al. 2012) and may explain why this variant was associated with Achilles tendinopathy (Nell et al. 2012).

As touched upon in the chapter specific conclusions in this thesis and discussed in the introduction, CN may have a direct effect of altering gene dosage and the subsequent downstream expression level of a gene (Dear 2009). For the studies on the *COL5A1* and *COL4A1* genes, differences in CN were observed between the CON and ATP groups. This was noted for a single locus within each gene spanning intron 4 - exon 5 for the *COL5A1* and intron 22 - exon 22 for the *COL4A1*. Similarly CN within exon 7 of the *MMP3* gene was shown to be significantly different between the AUS ATP and AUS CON group. While these significant results were not reflected in the continuous CN data and were in small sample sizes, the differences identified in discrete CN in the *COL5A1*, *COL4A1* and *MMP3* genes between the ATP and CON groups suggests that altered copy number at these loci may modify the risk of ATP. Indeed, a disruption of the amount and types of collagen in tendon, as well as altered levels of regulatory enzymes like MMP3, are apparent in ATP (Jones et al., 2006).

7.2 Pyrosequencing Studies

Chapter 6 described the pyrosequencing studies investigating DNA methylation status within the promoter regions of the *ADAMTS4*, *TIMP1* and *MMP11* genes. CpG site methylation was investigated in the promoter of the *ADAMTS4* gene, at a location approximately 3 kb from the start site of the gene. The assay used measured the methylation status of 6 CpG sites. At one CpG site, DNA methylation between the CON and PT group was found to be significantly different ($p = 0.016$). Specifically, mean DNA methylation was significantly higher in the PT group compared to the CON group. Higher expression levels of the *ADAMTS4* gene have previously been reported in ruptured Achilles tendon compared with normal and tendinopathic tendon (Corps et al. 2008) and it is reasonable to postulate that this may be the result of epigenetic mechanisms such as DNA methylation. While higher levels of DNA methylation (hypermethylation) is generally considered to silence gene expression (Raleigh 2012), it may be that the higher methylation status observed within this region 3 kb from the start of the gene, may have an indirect effect on gene expression through the disruption of transcription factor binding (Thirunavukkarasu et al. 2006). For example, a number of transcription factors are known to bind within the promoter of the *ADAMTS4* gene, including p53 and PPAR-gamma (peroxisome proliferator-activated receptor gamma) (www.genecards.org).

DNA methylation status was also investigated at 3 CpG sites within the *TIMP1* gene, in a region 8.8 kb from the start site of the gene. As such, DNA methylation status at one CpG site within the *TIMP1* gene was found to be significantly different between the CON and PT groups ($p = 0.012$). Mean methylation in the PT samples was found to be significantly higher than in the CON samples. Interestingly,

altered expression of the *TIMP1* gene has been shown in ruptured compared with normal Achilles tendon (Jones et al. 2006), suggesting that epigenetic mechanisms such as CpG methylation may well have an important role in tendon pathology. As with the *ADAMTS4* gene discussed above, altered DNA methylation may interfere with the binding of transcription factors within the promoter of the *TIMP1* gene. This is supported by the presence of numerous binding sites within the *TIMP1* gene promoter, where transcription factors such as Sp1 (specific protein 1) and c-Fos (www.genecards.org).

Lastly, chapter 6 reported the DNA methylation status measured at 4 CpG sites within the promoter region of the *MMP11* gene. These CpG sites were in a region approximately 30 bp from the start of the *MMP11* gene. There was a weak but significant difference ($p = 0.045$) in DNA methylation at one CpG site within the *MMP11* gene between the CON and PT groups. Mean methylation in the PT samples was found to be significantly higher than the CON samples. These differences in DNA methylation may have an influence on gene expression and the risk of tendon pathology, although functional studies would be necessary in order to prove this. Indeed, differences in expression levels of the *MMP11* gene have been reported in pathologic Achilles tendon (Jones et al. 2006) and increased expression of the *MMP11* gene was observed in both painful compared with normal tendon, as well as ruptured compared with normal tendon (Jones et al. 2006). If epigenetic mechanisms, such as DNA methylation, are responsible for such changes in gene expression, one may speculate that transcription factors may not be able to bind or access the genome as readily, due to increased CpG site methylation within the promoter of the gene. Nevertheless, further work would be needed to try to establish this.

These preliminary investigations into DNA methylation status in patellar tendinopathy suggest that methylation status of CpG sites within the *ADAMTS4*, *TIMP1* and *MMP11* genes may be important in relation to the risk of this condition. It is plausible that the accumulation of the proteoglycan aggrecan observed in tendinopathy (Smith et al. 2008) may be explained by a reduction in ADAMTS4 due to hypermethylation at CpG sites within the *ADAMTS4* gene promoter. Similarly, the hypermethylation observed at a CpG site within the promoter of the *TIMP1* gene could be causing a decrease in expression of the gene and elevated levels of substrates that TIMP1 would usually cleave. Furthermore, the increased DNA methylation reported in thesis, at a CpG site within the *MMP11* gene, could be causing a reduction in *MMP11* gene expression. This may result in lower levels of MMP11 protein, an accumulation of proteoglycan substrates that MMP11 would typically cleave (Somerville et al. 2003) and consequently disrupt the tendon homeostasis. Nevertheless, while significant differences in methylation status have been identified between the control and patellar tendinopathy participants, the functional relevance of these findings is currently unknown and can only be speculated. Additional research would be necessary to establish this.

7.3 Limitations and Future Work

The research detailed in this thesis has advanced the understanding of the genetic risk factors for both Achilles and patellar tendon pathology, whilst also revealing a number of possible avenues for future research. The studies contained within this thesis have investigated novel genes, as well as genes previously associated with tendon pathology. The work has explored SNP variants, along with being the first to investigate the role of copy number variation in relation to ATP. In addition, the investigation of DNA methylation status by Pyrosequencing is an exciting new area researched in this thesis.

As described in the methods in Chapter 2, the SNP and CNV studies in this thesis were undertaken in a sample size sufficient to reach 80% power when cohorts were analysed separately, while >90% power was achieved when the UK, SA and AUS Caucasian cohorts were combined. However, a larger sample size would allow greater confidence in results, particularly for those where cohorts were stratified into subgroups for analyses. A larger sample size would also reduce the risk of false positive and false negative results. As the three cohorts investigated were all Caucasian, further work replicating these studies in other populations would also be advantageous.

The overview of participant characteristics for the UK, SA and AUS cohorts are detailed in the methods in Chapter 2 and mentioned in the results sections of Chapters 3, 4 and 5. There were some significant differences in participant characteristics reported between the CON and ATP cohorts. However, analyses showed that these differences were not found to interact with copy number or genotype in the CNV and SNP studies respectively.

All participants for the ATP studies, both those with Achilles tendon pathology and the control subjects, self-reported as being physically active. Some participants were involved in physical activity and sports recreationally, while others participated in sports at a competitive or professional level. With this in mind, it is known that different sporting/physical activities, as well as the intensity and duration of exercise, affects an individual's risk of developing tendon pathology (Raleigh & Collins, 2012; Järvinen et al. 2005; Christian et al. 2014). Further work with larger sample sizes may look at analyses where cohorts are stratified for activity level or sport.

Whilst the studies within this thesis on genomic copy number have generally revealed that CNV does not appear to be an important risk factor for ATP in the genes investigated, it does warrant further exploration in other genes. This work was the first to address the potential role of CNV in relation to ATP and has addressed a number of important issues including the most appropriate way of analysing and interpreting the copy number data. Copy number data can be reported as the raw continuous data to decimal places. This data is taken directly from the CopyCaller software prior to the calculation of discrete CN using the maximum-likelihood algorithm (Park et al. 2012). While it may appear to be more biologically relevant to analyse CN data based on whole integer gene copies, it may, in certain situations, mean that borderline CN calls are not accurately represented (Park et al. 2012; Plagnol 2009). Specifically, accuracy may be affected when determining values on the borderline between rounding up or down to a whole copy number value (Park et al. 2012; Plagnol 2009). However, a

distinct advantage of interpreting and analysing the continuous CN data is that a smaller sample size is required to reach statistical power.

To overcome any potential rounding errors, all CN data were analysed as both continuous and discrete CN values. Despite some minor discrepancies when the cohorts were stratified into smaller subgroups, we found concordance in our CN analyses and thus have confidence that the degree of error in CN calling was negligible. As discussed in the method (section 2.3.1), TaqMan probe based qPCR is a robust and accurate method for CNV determination. As with all methods, there are other ways of measuring and determining copy number within a sample (Walker et al. 2009; Machado et al. 2013; Fode et al. 2011). Whilst Pyrosequencing was used for measuring methylation status in this research, it may also be an appropriate method for analysing copy number (Fode et al. 2011). The Paralogue Ratio Test (PRT) is another method for the measurement of CN that may be of interest (Walker et al. 2009). However, as described elsewhere, the running costs of these methods and the analysis time required would be significantly greater (Walker et al. 2009; Fode et al. 2011). In addition, while models exist to analyse the combined effect of multiple SNPs as haplotypes, such models are not apparent for CNV analyses. Additional investigations into the effect of CNV on the risk of ATP may look to consider the combined effect of CNV at different loci, if plausible to do so.

The mapping of common CNV regions is a relatively new area in comparison to the vast knowledge of SNPs (Ionita-Laza et al. 2009). CNV discovery is an on-going process, meaning the number and frequency of CNV catalogued within the human genome is liked to increase. This thesis examined the genetic association

of candidate CNV regions with ATP, utilising the catalogued CNV data from the Database of Genomic Variants. Whilst at the time of research the *MMP3* gene had no reported CNV within the two loci investigated in this work, it is likely that future mapping may highlight that CNV is present within this region and indeed much more of the genome. Indeed, the development of new technologies and better detection of CNVs is expected to advance CNV research (Ionita-Laza et al. 2009). Furthermore, despite there being no mapped CNV within the *MMP3* gene at the loci in this thesis, this work identified differences in copy number in both the ATP cases and controls. Although not significantly different, the presence of participants with a CN < 2 or CN >2 suggests that there are CNV regions yet to be mapped in the *MMP3* gene.

The epigenetic work reported in this thesis identified differences in DNA methylation status between the patellar tendinopathy (PT) and control groups at selected CpG sites in the promoter regions of the novel candidate genes *ADAMTS4*, *MMP11* and *TIMP1*. Studying further CpG sites within these genes would be valuable, enabling a fuller analysis of the whole gene promoter of each gene. Additionally, investigation of DNA methylation status within a larger cohort would be beneficial for additional analyses.

This work investigated a range of genetic variants known to be involved in the maintenance of the tendon ECM. Certain genes had been previously associated with tendon pathology, while others were novel candidate genes. Future candidate genes worthy of investigation would be the scleraxis (*SCX*) gene and the tenomodulin (*TeM*) gene. Scleraxis (SCX) is a transcription factor involved in tenogenesis and tendon development (Killian et al. 2012). SCX can be detected in

progenitor tendon cells as well as mature tenocytes and is considered as a tendon marker (Shukunami et al. 2006; Schweitzer et al. 2001). Interestingly, increased expression of scleraxis has been shown in response to mechanical load, leading to changes in ECM production and structure of the tendon (Jones et al. 2013; Killian et al. 2012). Furthermore, SCX positively regulates tenomodulin (*TeM*) gene expression (Shukunami et al. 2006). TeM is a transmembrane protein expressed specifically in tendon and other connective tissues (Léjard et al. 2007) and a lack of TeM has been shown to affect tenocyte differentiation and collagen fibril structure (Docheva et al. 2005).

A variety of different genetic variants have been investigated in this thesis, ranging from small-scale SNPs, to larger genomic CNV, to CpG site methylation. Future research uncovering the genetic risk factors for tendon pathology will likely continue to centre on SNPs, given the interesting findings to date concerning these types of genetic variants. As more genetic variants are found to associate with ATP and others are excluded from a role in the risk of ATP, it is envisaged that a comprehensive and detailed understanding of the genetic predisposition of ATP will be unearthed. This may help to identify at risk individuals and enable preventative strategies to be implemented to modify their injury risk. Furthermore, the preliminary findings within this thesis on DNA methylation status within the *ADAMTS4*, *MMP11* and *TIMP1* genes highlights the interest in researching other forms of genetic variation in relation to ATP, namely epigenetic mechanisms.

7.4 Concluding Remarks

It is envisioned that the continued research and identification of novel genetic variants associated with ATP will help to develop a clinical risk model for this complex pathology, where a greater understanding is gained on an individual's risk of injury. As such, preventative measures will be able to be implemented sooner, with the aim of reducing the risk of ATP.

REFERENCES

Abrahams, Y., Laguette, M.-J., Prince, S., Collins, M. (2013) Polymorphisms within the COL5A1 3'-UTR that alters mRNA structure and the MIR608 gene are associated with Achilles tendinopathy. *Annals of Human Genetics*. **77**(3), 204–14.

Amos, W., Driscoll, E., Hoffman, J.I. (2011) Candidate genes versus genome-wide associations: which are better for detecting genetic susceptibility to infectious disease? *Proceedings. Biological sciences / The Royal Society*. **278**(1709), 1183–8.

Anderson, C.L., Brown, C.J. (2005) Epigenetic predisposition to expression of TIMP1 from the human inactive X chromosome. *BMC genetics*. **6**(1), 48.

Auclair, G., Weber, M. (2012) Mechanisms of DNA methylation and demethylation in mammals. *Biochimie*. **94**(11), 2202–11.

Backman, L.J., Eriksson, D.E., Danielson, P. (2013) Substance P reduces TNF-α-induced apoptosis in human tenocytes through NK-1 receptor stimulation. *British journal of sports medicine*.

Bartel, D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*. **116**(2), 281–97.

Beckmann, J.S., Estivill, X., Antonarakis, S.E. (2007) Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. *Nature Reviews. Genetics*. **8**(8), 639–46.

Benjamin, M., Kaiser, E., Milz, S. (2008) Structure-function relationships in tendons: a review. *Journal of Anatomy*. **212**(3), 211–28.

Birch, H.L. (2007) Tendon matrix composition and turnover in relation to functional requirements. *International Journal of Experimental Pathology*. **88**(4), 241–8.

Birkedal-Hansen, H., Moore, W.G.I., Bodden, M.K., Windsor, L.J., Birkedal-Hansen, B., DeCarlo, A., Engler, J.A. (1993) Matrix Metalloproteinases : A Review. *Critical Reviews in Oral Biology Medicine*. **4**(2), 197–250.

Bodin, L., Beaune, P.H., Loriot, M.-A. (2005) Determination of cytochrome P450 2D6 (CYP2D6) gene copy number by real-time quantitative PCR. *Journal of biomedicine & biotechnology*. **2005**(3), 248–53.

Brookes, A.J. (1999) The essence of SNPs. *Gene*. **234**(2), 177–186.

Christian, R.A., Rossy, W.H., Sherman, O.H. (2014) Patellar tendinopathy - recent developments toward treatment. *Bulletin of the Hospital for Joint Disease (2013)*. **72**(3), 217–24.

Christian, R.A., Rossy, W.H., Sherman, O.H. (2014) Patellar tendinopathy - recent developments toward treatment. *Bulletin of the Hospital for Joint Disease (2013)*. **72**(3), 217–24.

Collins, M., Posthumus, M. (2011) Type V collagen genotype and exercise-related phenotype relationships: a novel hypothesis. *Exercise and sport sciences reviews*. **39**(4), 191–8.

Collins, M., Raleigh, S.M. (2009) Genetic risk factors for musculoskeletal soft tissue injuries. *Medicine and Sport Science*. **54**, 136–149.

Cook, J., Khan, K., Purdam, C. (2002) Achilles tendinopathy. *Manual Therapy*. **7**(3), 121–130.

Cook, J.L., Purdam, C.R. (2009) Is tendon pathology a continuum? A pathology model to explain the clinical presentation of load-induced tendinopathy. *British journal of sports medicine*. **43**(6), 409–16.

Cook, J.L., Purdam, C.R. (2014) The challenge of managing tendinopathy in competing athletes. *British journal of sports medicine*. **48**(7), 506–9.

Corps, A.N., Jones, G.C., Harrall, R.L., Curry, V.A., Hazleman, B.L., Riley, G.P. (2008) The regulation of aggrecanase ADAMTS-4 expression in human Achilles tendon and tendon-derived cells. *Matrix Biology*. **27**(5), 393–401.

Das, C., Tyler, J.K. (2013) Histone exchange and histone modifications during transcription and aging. *Biochimica et biophysica acta*. **1819**(3-4), 332–42.

Dear, P.H. (2009) Copy-number variation: the end of the human genome? *Trends in Biotechnology*. **27**(8), 448–54.

Degenhardt, F., Priebe, L., Herms, S., Mattheisen, M., Mühleisen, T.W., Meier, S., Moebus, S., Strohmaier, J., Groß, M., Breuer, R., Lange, C., Hoffmann, P., Meyer-Lindenberg, A., Heinz, A., Walter, H., Lucae, S., Wolf, C., Müller-Myhsok, B., Holsboer, F., Maier, W., et al. (2012) Association between copy number variants in 16p11.2 and major depressive disorder in a German case-control sample. *American Journal of Medical Genetics. Part B*. **159B**(3), 263–273.

D'haene, B., Vandesompele, J., Hellemans, J. (2010) Accurate and objective copy number profiling using real-time quantitative PCR. *Methods (San Diego, Calif.)*. **50**(4), 262–70.

van Dijk, C.N., van Sterkenburg, M.N., Wiegerinck, J.I., Karlsson, J., Maffulli, N. (2011) Terminology for Achilles tendon related disorders. *Knee Surgery, Sports Traumatology and Arthroscopy*. **19**(5), 835–41.

Docheva, D., Hunziker, E.B., Fässler, R., Brandau, O. (2005) Tenomodulin is necessary for tenocyte proliferation and tendon maturation. *Molecular and cellular biology*. **25**(2), 699–705.

Estivill, X., Armengol, L. (2007) Copy number variants and common disorders: filling the gaps and exploring complexity in genome-wide association studies. E. M. C. Fisher, ed. *PLoS Genetics*. **3**(10), 1787–99.

Faul, F., Erdfelder, E., Lang, A.-G., Buchner, A. (2007) G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*. **39**(2), 175–191.

Fode, P., Jespersgaard, C., Hardwick, R.J., Bogle, H., Theisen, M., Dodoo, D., Lenicek, M., Vitek, L., Vieira, A., Freitas, J., Andersen, P.S., Hollox, E.J. (2011) Determination of beta-defensin genomic copy number in different populations: a comparison of three methods. *PloS one*. **6**(2), e16768.

Foster, B.P., Morse, C.I., Onambele, G.L., Ahmetov, I.I., Williams, A.G. (2012) Genetic Variation , Protein Composition and Potential Influences on Tendon Properties in Humans. *The Open Sports Medicine Journal*. **6**, 8–21.

Gaida, J.E., Bagge, J., Purdam, C., Cook, J., Alfredson, H., Forsgren, S. (2012) Evidence of the TNF-α system in the human Achilles tendon: expression of TNF-α and TNF receptor at both protein and mRNA levels in the tenocytes. *Cells, tissues, organs*. **196**(4), 339–52.

Gauderman, W.J. (2002) Sample size requirements for matched case-control studies of gene-environment interaction. *Statistics in medicine*. **21**(1), 35–50.

Gerster, J.C., Landry, M., Rappoport, G., Rivier, G., Duvoisin, B., Schnyder, P. (1996) Enthesopathy and tendinopathy in gout: computed tomographic assessment. *Annals of the rheumatic diseases*. **55**(12), 921–923.

Glaziou P (2005) *Sampsize – sample size and power* [Online] Available from: http://sampsize.sourceforge.net/. [Accessed: 10.11.14].

Graf, S.W., Lester, S., Nossent, J.C., Hill, C.L., Proudman, S.M., Lee, A., Rischmueller, M. (2012) Low copy number of the FCGR3B gene and rheumatoid arthritis: a case-control study and meta-analysis. *Arthritis Research & Therapy*. **14**(1).

Grozeva, D., Conrad, D.F., Barnes, C.P., Hurles, M., Owen, M.J., O'Donovan, M.C., Craddock, N., Kirov, G. (2012) Independent estimation of the frequency of rare CNVs in the UK population confirms their role in schizophrenia. *Schizophrenia Research*. **135**, 1–7.

Guan, X., Liu, Z., Liu, H., Yu, H., Wang, L.-E., Sturgis, E.M., Li, G., Wei, Q. (2013) A functional variant at the miR-885-5p binding site of CASP3 confers risk of both index and second primary malignancies in patients with head and neck cancer. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*. **27**(4), 1404–1412.

Gururajan, R., Grenet, J., Lahti, J.M., Kidd, V.J. (1998) Isolation and characterization of two novel metalloproteinase genes linked to the Cdc2L locus on human chromosome 1p36.3. *Genomics*. **52**(1), 101–106.

Hastings, P.J., Lupski, J.R., Rosenberg, S.M., Ira, G. (2009) Mechanisms of change in gene copy number. *Nature reviews. Genetics*. **10**(8), 551–64.

Hengartner, M.O. (2000) The biochemistry of apoptosis. *Nature*. **407**(6805), 770–6.

Henrichsen, C.N., Chaignat, E., Reymond, A. (2009) Copy number variants, diseases and gene expression. *Human Molecular Genetics*. **18**(1), 1–8.

Holt, R., Sykes, N.H., Conceição, I.C., Cazier, J.-B., Anney, R.J., Oliveira, G., Gallagher, L., Vicente, A., Monaco, A.P., Pagnamenta, A.T. (2012) CNVs leading to fusion transcripts in individuals with autism spectrum disorder. *European Journal of Human Genetics*, 1–7.

Hosaka, Y., Kirisawa, R., Ueda, H., Yamaguchi, M., Takehana, K. (2005) Differences in tumor necrosis factor (TNF)alpha and TNF receptor-1-mediated intracellular signaling factors in normal, inflamed and scar-formed horse tendons. *The Journal of veterinary medical science / the Japanese Society of Veterinary Science*. **67**(10), 985–91.

Ionita-Laza, I., Rogers, A.J., Lange, C., Raby, B.A., Lee, C. (2009) Genetic association analysis of copy-number variation (CNV) in human disease pathogenesis. *Genomics*. **93**(1), 22–6.

Ireland, D., Harrall, R., Curry, V., Holloway, G., Hackney, R., Hazleman, B., Riley, G. (2001) Multiple changes in gene expression in chronic human Achilles tendinopathy. *Matrix biology : journal of the International Society for Matrix Biology*. **20**(3), 159–69.

Järvinen, T.A., Jozsa, L., Kannus, P., Järvinen, T.L., Kvist, M., Hurme, T., Isola, J., Kalimo, H., Järvinen, M. (1999) Mechanical loading regulates tenascin-C expression in the osteotendinous junction. *Journal of Cell Science*. **112 Pt 18**, 3157–66.

Järvinen, T.A.H., Kannus, P., Maffulli, N., Khan, K.M. (2005) Achilles tendon disorders: etiology and epidemiology. *Foot and ankle clinics*. **10**(2), 255–66.

Jelinsky, S.A., Rodeo, S.A., Li, J., Gulotta, L. V, Archambault, J.M., Seeherman, H.J. (2011) Regulation of gene expression in human tendinopathy. *BMC Musculoskeletal Disorders*. **12**(86), 1–12.

Jones, E.R., Jones, G.C., Legerlotz, K., Riley, G.P. (2013) Cyclical strain modulates metalloprotease and matrix gene expression in human tenocytes via activation of TGFβ. *Biochimica et biophysica acta*. **1833**(12), 2596–607.

Jones, G.C., Corps, A.N., Pennington, C.J., Clark, I.M., Edwards, D.R., Bradley, M.M., Hazleman, B.L., Riley, G.P. (2006) Expression profiling of metalloproteinases and tissue inhibitors of metalloproteinases in normal and degenerate human achilles tendon. *Arthritis and rheumatism*. **54**(3), 832–42.

Jones, G.C., Riley, G.P. (2005) ADAMTS proteinases: a multi-domain, multi-functional family with roles in extracellular matrix turnover and arthritis. *Arthritis research & therapy*. **7**(4), 160–9.

de Jonge, S., van den Berg, C., de Vos, R.J., van der Heide, H.J.L., Weir, A., Verhaar, J.A.N., Bierma-Zeinstra, S.M.A., Tol, J.L. (2011) Incidence of midportion Achilles tendinopathy in the general population. *British journal of sports medicine*. **45**(13), 1026–8.

Joseph, M.F., Lillie, K.R., Bergeron, D.J., Cota, K.C., Yoon, J.S., Kraemer, W.J., Denegar, C.R. (2014) Achilles Tendon Biomechanics in Response to Acute Intense Exercise. *Journal of Strength and Conditioning Research*. **28**(5), 1181–1186.

Jozsa, L., Balint, J.B., Kannus, P., Reffy, A., Barzo, M. (1989) Distribution of blood groups in patients with tendon rupture. An analysis of 832 cases. *The Journal of Bone and Joint Surgery*. **71**(2), 272–4.

Kader, D., Saxena, A., Movin, T., Maffulli, N. (2002) Achilles tendinopathy: some aspects of basic science and clinical management. *British journal of sports medicine*. **36**(4), 239–49.

Karaderi, T., Pointon, J.J., Wordsworth, T.W., Harvey, D., Appleton, L.H., Cohen, C.J., Farrar, C., Harin, A., Brown, M.A., Wordsworth, B.P. (2012) Evidence of genetic association between TNFRSF1A encoding the p55 tumour necrosis factor receptor, and ankylosing spondylitis in UK Caucasians. *Clinical and experimental rheumatology*. **30**(1), 110–3.

Karousou, E., Ronga, M., Vigetti, D., Passi, A., Maffulli, N. (2008) Collagens, proteoglycans, MMP-2, MMP-9 and TIMPs in human achilles tendon rupture. *Clinical orthopaedics and related research*. **466**(7), 1577–82.

Kerna, I., Kisand, K., Tamm, A.E., Lintrop, M., Veske, K., Tamm, A.O. (2009) Missense single nucleotide polymorphism of the ADAM12 gene is associated with radiographic knee osteoarthritis in middle-aged Estonian cohort. *Osteoarthritis and Cartilage*. **17**(8), 1093–1098.

Khaliq, Y., Zhanel, G.G. (2003) Fluoroquinolone-associated tendinopathy: a critical review of the literature. *Clinical Infectious Diseases*. **36**(11), 1404–10.

Khan, K.., Maffulli, N., Coleman, B.., Cook, J.., Taunton, J.. (1998) Patellar tendinopathy: some aspects of basic science and clinical management. *British Journal of Sports Medicine*. **32**, 346–355.

Khan, K.M., Cook, J.L., Bonar, F., Harcourt, P., Astrom, M. (1999) Histopathology of common tendinopathies. Update and implications for clinical management. *Sports Medicine*. **27**(6), 393–408.

El Khoury, L., Posthumus, M., Collins, M., Handley, C.J., Cook, J., Raleigh, S.M. (2013) Polymorphic variation within the ADAMTS2, ADAMTS14, ADAMTS5, ADAM12 and TIMP2 genes and the risk of Achilles tendon pathology: A genetic association study. *Journal of Science and Medicine in Sport*. **16**, 493–498.

El Khoury, L., Posthumus, M., Collins, M., van der Merwe, W., Handley, C., Cook, J., Raleigh, S.M. (2015) ELN and FBN2 gene variants as risk factors for two sports-related musculoskeletal injuries. *International Journal of Sports Medicine*. **36**(4), 333–7.

El Khoury, L., Ribbans, W.J., Raleigh, S.M. (2016) MMP3 and TIMP2 gene variants as predisposing factors for Achilles tendon pathologies: Attempted replication study in a British case-control cohort. *Meta Gene*. **9**, 52–5.

Killian, M.L., Cavinatto, L., Galatz, L.M., Thomopoulos, S. (2012) The role of mechanobiology in tendon healing. *Journal of Shoulder and Elbow Surgery*. **21**(2), 228–37.

Kirsch, D.G., Doseff, A., Chau, B.N., Lim, D.S., de Souza-Pinto, N.C., Hansford, R., Kastan, M.B., Lazebnik, Y.A., Hardwick, J.M. (1999) Caspase-3-dependent cleavage of Bcl-2 promotes release of cytochrome c. *The Journal of biological chemistry*. **274**(30), 21155–61.

Kraemer, R., Wuerfel, W., Lorenzen, J., Busche, M., Vogt, P.M., Knobloch, K. (2012) Analysis of hereditary and medical risk factors in Achilles tendinopathy and Achilles tendon ruptures: a matched pair analysis. *Archives of orthopaedic and trauma surgery*. **132**(6), 847–53.

Kujala, U.M., Järvinen, M., Natri, A., Lehto, M., Nelimarkka, O., Hurme, M., Virta, L., Finne, J. (1992) ABO blood groups and musculoskeletal injuries. *Injury*. **23**(2), 131–3.

Labelle-Dumais, C., Dilworth, D.J., Harrington, E.P., de Leau, M., Lyons, D., Kabaeva, Z., Manzini, M.C., Dobyns, W.B., Walsh, C.A., Michele, D.E., Gould, D.B. (2011) COL4A1 mutations cause ocular dysgenesis, neuronal localization defects, and myopathy in mice and walker-warburg syndrome in humans. *PLoS Genetics*. **7**(5).

Laguette, M.-J., Abrahams, Y., Prince, S., Collins, M. (2011) Sequence variants within the 3′-UTR of the COL5A1 gene alters mRNA stability: implications for musculoskeletal soft tissue injuries. *Matrix Biology*. **30**(5-6), 338–45.

Lamkanfi, M., Festjens, N., Declercq, W., Vanden Berghe, T., Vandenabeele, P. (2007) Caspases in cell survival, proliferation and differentiation. *Cell death and differentiation*. **14**(1), 44–55.

Léjard, V., Brideau, G., Blais, F., Salingcarnboriboon, R., Wagner, G., Roehrl, M.H.A., Noda, M., Duprez, D., Houillier, P., Rossert, J. (2007) Scleraxis and NFATc regulate the expression of the pro-alpha1(I) collagen gene in tendon fibroblasts. *The Journal of biological chemistry*. **282**(24), 17665–75.

Lesic, A., Bumbasirevic, M. (2004) Disorders of the Achilles tendon. *Current Orthopaedics*. **18**(1), 63–75.

Lewis, C.M., Knight, J. (2012) Introduction to Genetic Association Studies. *Cold Spring Harbor Protocols*. **2012**(3), pdb.top068163–pdb.top068163.

Liu, M.M., Agrón, E., Chew, E., Meyerle, C., Ferris, F.L., Chan, C.-C., Tuo, J. (2011) Copy number variations in candidate genes in neovascular age-related macular degeneration. *Investigative ophthalmology & visual science*. **52**(6), 3129–35.

Machado, L.R., Bowdrey, J., Ngaimisi, E., Habtewold, A., Minzi, O., Makonnen, E., Yimer, G., Amogne, W., Mugusi, S., Janabi, M., Aderaye, G., Mugusi, F., Viskaduraki, M., Aklillu, E., Hollox, E.J. (2013) Copy Number Variation of Fc Gamma Receptor Genes in HIV-Infected and HIV-Tuberculosis Co-Infected Individuals in Sub-Saharan Africa. W. He, ed. *PloS one*. **8**(11), 1–8.

Madi, T., Balamurugan, K., Bombardi, R., Duncan, G., McCord, B. (2012) The determination of tissue-specific DNA methylation patterns in forensic biofluids using bisulfite modification and pyrosequencing. *Electrophoresis*. **33**(12), 1736–45.

Maffulli, N., Wong, J., Almekinders, L.C. (2003) Types and epidemiology of tendinopathy. *Clinics in Sports Medicine*. **22**(4), 675–92.

Mayo, P., Hartshorne, T., Li, K., McMunn-Gibson, C., Spencer, K., Schnetz-Boutaud, N. (2010) CNV analysis using TaqMan copy number assays. *Current Protocols in Human Genetics*. Supplement 67:2.13.1–2.13.10.

McCarroll, S.A. (2008) Extending genome-wide association studies to copy-number variation. *Human Molecular Genetics*. **17**(R2), R135–42.

Meeuwisse, W.H. (1994) Assessing causation in sport injury: A multifactorial model. *Clinical journal of sport medicine*. **4**(3), 166–70.

Midwood, K.S., Hussenet, T., Langlois, B., Orend, G. (2011) Advances in tenascin-C biology. *Cellular and molecular life sciences : CMLS*. **68**(19), 3175–99.

Minear, M.A., Crosslin, D.R., Sutton, B.S., Connelly, J.J., Nelson, S.C., Gadson-Watson, S., Wang, T., Seo, D., Vance, J.M., Sketch, M.H., Haynes, C., Goldschmidt-Clermont, P.J., Shah, S.H., Kraus, W.E., Hauser, E.R., Gregory, S.G. (2011) Polymorphic variants in tenascin-C (TNC) are associated with atherosclerosis and coronary artery disease. *Human Genetics*. **129**(6), 641–54.

Mokone, G.G., Gajjar, M., September, A. V, Schwellnus, M.P., Greenberg, J., Noakes, T.D., Collins, M. (2005) The guanine-thymine dinucleotide repeat polymorphism within the tenascin-C gene is associated with achilles tendon injuries. *The American journal of sports medicine*. **33**(7), 1016–21.

Mokone, G.G., Schwellnus, M.P., Noakes, T.D., Collins, M. (2006) The COL5A1 gene and Achilles tendon pathology. *Scandinavian journal of medicine & science in sports*. **16**(1), 19–26.

Molokhia, M., Fanciulli, M., Petretto, E., Patrick, A.L., McKeigue, P., Roberts, A.L., Vyse, T.J., Aitman, T.J. (2011) FCGR3B copy number variation is associated with systemic lupus erythematosus risk in Afro-Caribbeans. *Rheumatology*. **50**(7), 1206–10.

Morcillo-Suarez, C., Alegre, J., Sangros, R., Gazave, E., de Cid, R., Milne, R., Amigo, J., Ferrer-Admetlla, A., Moreno-Estrada, A., Gardner, M., Casals, F., Pérez-Lezaun, A., Comas, D., Bosch, E., Calafell, F., Bertranpetit, J., Navarro, A. (2008) SNP analysis to results (SNPator): a web-based environment oriented to statistical genomics analyses upon SNP data. *Bioinformatics (Oxford, England)*. **24**(14), 1643–4.

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

Park, J.H., Lee, S., Yu, H.G., Kim, J.-I., Seo, J.-S. (2012) Copy number variation of age-related macular degeneration relevant genes in the Korean population. *PloS ONE*. **7**(2), e31243.

Park, T.-J., Kim, H.J., Kim, J.-H., Bae, J.S., Cheong, H.S., Park, B.-L., Shin, H.D. (2013) Associations of CD6, TNFRSF1A and IRF8 polymorphisms with risk of inflammatory demyelinating diseases. *Neuropathology and applied neurobiology*. **39**(5), 519–30.

Patnala, R., Clements, J., Batra, J. (2013) Candidate gene association studies: a comprehensive guide to useful in silico tools. *BMC genetics*. **14**, 39.

Parkinson, J., Samiric, T., Ilic, M.Z., Cook, J., Feller, J.A., Handley, C.J. (2010) Change in proteoglycan metabolism is a characteristic of human patellar tendinopathy. *Arthritis and rheumatism*. **62**(10), 3028–35.

Peters, S.P. (2009) Reporting and evaluating genetic association studies. *Respiratory research*. **10**, 109.

Pinto, D., Marshall, C., Feuk, L., Scherer, S.W. (2007) Copy-number variation in control population cohorts. *Human Molecular Genetics*. **16**(R2), 168–173.

Plagnol, V. (2009) Association tests and software for copy number variant data. *Human Genomics*. **3**(2), 191–4.

Porter, A.G., Jänicke, R.U. (1999) Emerging roles of caspase-3 in apoptosis. *Cell death and differentiation*. **6**(2), 99–104.

Posthumus, M., Collins, M., Cook, J., Handley, C.J., Ribbans, W.J., Smith, R.K.W., Schwellnus, M.P., 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, England)*. **49**(11), 2090–7.

Posthumus, M., Collins, M., van der Merwe, L., O'Cuinneagain, D., van der Merwe, W., Ribbans, W.J., Schwellnus, M.P., Raleigh, S.M. (2011) Matrix metalloproteinase genes on chromosome 11q22 and the risk of anterior cruciate ligament (ACL) rupture. *Scandinavian journal of medicine & science in sports*.

Primakoff, P., Myles, D.G. (2000) The ADAM gene family: surface proteins with adhesion and protease activity. *Trends in genetics : TIG*. **16**(2), 83–7.

Rakyan, V.K., Down, T.A., Balding, D.J., Beck, S. (2011) Epigenome-wide association studies for common human diseases. *Nature Reviews: Genetics*. **12**(8), 529–41.

Raleigh, S.M. (2012) Epigenetic regulation of the ACE gene might be more relevant to endurance physiology than the I/D polymorphism. *Journal of applied physiology (Bethesda, Md. : 1985)*. **112**(6), 1082–3.

Raleigh, S.M., Collins, M. (2012) *Gene Variants that Predispose to Achilles Tendon Injuries : An Update on Recent Advances*. Andrej Cretnik, ed. InTech.

Raleigh, S.M., van der Merwe, L., Ribbans, W.J., Smith, R.K.W., Schwellnus, M.P., Collins, M. (2009) Variants within the MMP3 gene are associated with Achilles tendinopathy: possible interaction with the COL5A1 gene. *British Journal of Sports Medicine*. **43**(7), 514–20.

Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W., Cho, E.K., Dallaire, S., Freeman, J.L., González, J.R., Gratacòs, M., Huang, J., Kalaitzopoulos, D., Komura, D., MacDonald, J.R., Marshall, C.R., et al. (2006) Global variation in copy number in the human genome. *Nature*. **444**(7118), 444–54.

Rees, J.D., Stride, M., Scott, A. (2013) Tendons - time to revisit inflammation. *British Journal of Sports Medicine*. 1-7.

Rees, S.G., Flannery, C.R., Little, C.B., Hughes, C.E., Caterson, B., Dent, C.M. (2000) Catabolism of aggrecan, decorin and biglycan in tendon. *The Biochemical Journal*. **350**, 181–8.

Reinking, M. (2012) Tendinopathy in athletes. *Physical Therapy in Sport*. **13**(1), 3–10.

Reynard, L.N., Bui, C., Canty-Laird, E.G., Young, D.A., Loughlin, J. (2011) Expression of the osteoarthritis-associated gene GDF5 is modulated epigenetically by DNA methylation. *Human Molecular Genetics*. **20**(17), 3450–60.

Reynard, L.N., Loughlin, J. (2012) Genetics and epigenetics of osteoarthritis. *Maturitas*. **71**(3), 200–4.

Roach, H.I., Yamada, N., Cheung, K.S.C., Tilley, S., Clarke, N.M.P., Oreffo, R.O.C., Kokubun, S., Bronner, F. (2005) Association between the abnormal expression of matrix-degrading enzymes by human osteoarthritic chondrocytes and demethylation of specific CpG sites in the promoter regions. *Arthritis and Rheumatism*. **52**(10), 3110–24.

Robberecht, C., Voet, T., Zamani Esteki, M., Nowakowska, B.A., Vermeesch, J.R. (2013) Nonallelic homologous recombination between retrotransposable elements is a driver of de novo unbalanced translocations. *Genome Research*. **23**(3), 411–8.

Sakabe, T., Sakai, T. (2011) Musculoskeletal diseases--tendon. *British Medical Bulletin*. **99**, 211–25.

Samiric, T., Parkinson, J., Ilic, M.Z., Cook, J., Feller, J.A., Handley, C.J. (2009) Changes in the composition of the extracellular matrix in patellar tendinopathy. *Matrix biology : journal of the International Society for Matrix Biology*. **28**(4), 230-6.

Saunders, C.J., van der Merwe, L., Posthumus, M., Cook, J., Handley, C.J., Collins, M., September, A. V (2013) Investigation of variants within the COL27A1 and TNC genes and Achilles tendinopathy in two populations. *Journal of orthopaedic research : official publication of the Orthopaedic Research Society*. **31**(4), 632-7.

Schepsis, A.A., Jones, H., Haas, A.L. (2002) Achilles Tendon Disorders in Athletes. *The American Journal of Sports Medicine*. **30**(2), 287-305.

Schulze-Tanzil, G., Al-Sadi, O., Wiegand, E., Ertel, W., Busch, C., Kohl, B., Pufe, T. (2011) The role of pro-inflammatory and immunoregulatory cytokines in tendon healing and rupture: new insights. *Scandinavian Journal of Medicine & Science in Sports*. **21**(3), 337-51.

Schweitzer, R., Chyung, J.H., Murtaugh, L.C., Brent, A.E., Rosen, V., Olson, E.N., Lassar, A., Tabin, C.J. (2001) Analysis of the tendon cell fate using Scleraxis, a specific marker for tendons and ligaments. *Development (Cambridge, England)*. **128**(19), 3855-66.

Seals, D.F., Courtneidge, S.A. (2003) The ADAMs family of metalloproteases : multidomain proteins with multiple functions. . (616), 7-30.

Sendzik, J., Shakibaei, M., Schäfer-Korting, M., Stahlmann, R. (2005) Fluoroquinolones cause changes in extracellular matrix, signalling proteins, metalloproteinases and caspase-3 in cultured human tendon cells. *Toxicology*. **212**(1), 24-36.

September, A. V, Cook, J., Handley, C.J., van der Merwe, L., Schwellnus, M.P., Collins, M. (2009) Variants within the COL5A1 gene are associated with Achilles tendinopathy in two populations. *British journal of sports medicine*. **43**(5), 357-65.

September, A. V, Posthumus, M., Collins, M. (2012) Application of genomics in the prevention, treatment and management of Achilles tendinopathy and anterior cruciate ligament ruptures. *Recent Patents on DNA & Gene Sequences*. **6**(3), 216-23.

September, A. V, Posthumus, M., van der Merwe, L., Schwellnus, M., Noakes, T.D., Collins, M. (2008) The COL12A1 and COL14A1 genes and Achilles tendon injuries. *International Journal of Sports Medicine*. **29**(3), 257-63.

September, A. V, Schwellnus, M.P., Collins, M. (2007) Tendon and ligament injuries: the genetic component. *British Journal of Sports Medicine*. **41**(4), 241-6.

Sharma, P., Maffulli, N. (2005) Tendon injury and tendinopathy: healing and repair. *The Journal of Bone and Joint Surgery*. **87**(1), 187-202.

Shastry, B.S. (2002) SNP alleles in human disease and evolution. *Journal of Human Genetics*. **47**(11), 561-6.

Shukunami, C., Takimoto, A., Oro, M., Hiraki, Y. (2006) Scleraxis positively regulates the expression of tenomodulin, a differentiation marker of tenocytes. *Developmental Biology*. **298**(1), 234-47.

Smith, M.M., Sakurai, G., Smith, S.M., Young, A.A., Melrose, J., Stewart, C.M., Appleyard, R.C., Peterson, J.L., Gillies, R.M., Dart, A.J., Sonnabend, D.H., Little, C.B. (2008) Modulation of aggrecan and ADAMTS expression in ovine tendinopathy induced by altered strain. *Arthritis and Rheumatism*. 58(4), 1055–66.

Sobhani, S., Dekker, R., Postema, K., Dijkstra, P.U. (2013) Epidemiology of ankle and foot overuse injuries in sports: A systematic review. *Scandinavian Journal of Medicine & Science in Sports*. **23**(6), 669–86.

Somerville, R.P.T., Oblander, S. a, Apte, S.S. (2003) Matrix metalloproteinases: old dogs with new tricks. *Genome Biology*. **4**(6), 216.

Suzuki, T., Okamoto, A. (2014) Case Report Marked Multiple Tendinitis at the Onset of Rheumatoid Arthritis in a Patient with Heterozygous Familial Hypercholesterolemia : Ultrasonographic Observation. *Case Reports in Rheumatology*.**2014**, 1-6.

Swaminathan, B., Matesanz, F., Cavanillas, M.L., Alloza, I., Otaegui, D., Olascoaga, J., Cénit, M.C., de las Heras, V., Barcina, M.G., Arroyo, R., Alcina, A., Fernandez, O., Antigüedad, A., Urcelay, E., Vandenbroeck, K. (2010) Validation of the CD6 and TNFRSF1A loci as risk factors for multiple sclerosis in Spain. *Journal of Neuroimmunology*. **223**(1-2), 100–3.

Tandeter, H., Shvartzman, P. (1999) Acute Knee Injuries: Use of Decision Rules for Selective Radiograph Ordering. *American Family Physician*. **60**(9), 2599–2608.

Thirunavukkarasu, K., Pei, Y., Moore, T.L., Wang, H., Yu, X.-P., Geiser, A.G., Chandrasekhar, S. (2006) Regulation of the human ADAMTS-4 promoter by transcription factors and cytokines. *Biochemical and Biophysical Research Communications*. **345**(1), 197–204.

Verma, P., Dalal, K. (2013) Serum cartilage oligomeric matrix protein (COMP) in knee osteoarthritis: a novel diagnostic and prognostic biomarker. *Journal of Orthopaedic Research*. **31**(7), 999–1006.

Visse, R., Nagase, H. (2003) Matrix metalloproteinases and tissue inhibitors of metalloproteinases: Structure, function, and biochemistry. *Circulation Research*. **92**(8), 827–839.

Vosseller, J.T., Ellis, S.J., Levine, D.S., Kennedy, J.G., Elliott, A.J., Deland, J.T., Roberts, M.M., O'Malley, M.J. (2013) Achilles tendon rupture in women. *Foot & Ankle International*. **34**(1), 49–53.

Walker, S., Janyakhantikul, S., Armour, J.A.L. (2009) Multiplex Paralogue Ratio Tests for accurate measurement of multiallelic CNVs. *Genomics*. **93**(1), 98–103.

Wilkening, S., Chen, B., Bermejo, J.L., Canzian, F. (2009) Is there still a need for candidate gene approaches in the era of genome-wide association studies? *Genomics*. **93**(5), 415–419.

Xu, Y., Murrell, G.A.C. (2008) The basic science of tendinopathy. *Clinical Orthopaedics and Related Research*. **466**(7), 1528–38.

Yang, T.-L., Chen, X.-D., Guo, Y., Lei, S.-F., Wang, J.-T., Zhou, Q., Pan, F., Chen, Y., Zhang, Z.-X., Dong, S.-S., Xu, X.-H., Yan, H., Liu, X., Qiu, C., Zhu, X.-Z., Chen, T., Li, M., Zhang, H., Zhang, L., Drees, B.M., et al. (2008) Genome-wide copy-number-variation study identified a susceptibility gene, UGT2B17, for osteoporosis. *American Journal of Human Genetics*. **83**(6), 663–74.

Young, J.S., Kumta, S.M., Maffulli, N. (2005) Achilles tendon rupture and tendinopathy: management of complications. *Foot and Ankle Clinics*. **10**(2), 371-82.

Yu, B., Shao, Y., Li, P., Zhang, J., Zhong, Q., Yang, H., Hu, X., Chen, B., Peng, X., Wu, Q., Chen, Y., Guan, M., Wan, J., Zhang, W. (2010) Copy number variations of the human histamine H4 receptor gene are associated with systemic lupus erythematosus. *The British Journal of Dermatology*. **163**(5), 935-40.

Zhang, F., Gu, W., Hurles, M.E., Lupski, J.R. (2009) Copy number variation in human health, disease, and evolution. *Annual Review of Genomics and Human Genetics*. **10**, 451-81.

APPENDIX I

Ethical Approval

UNIVERSITY OF CAPE TOWN

Health Sciences Faculty
Research Ethics Committee
Room E52-24 Groote Schuur Hospital Old Main Building
Observatory 7925
Telephone [021] 406 6338 • **Facsimile** [021] 406 6411
e-mail: lamees.emjedi@uct.ac.za

02 October 2008

REC REF: 289/2004

A/Prof M Collins
Human Biology

Dear A/Prof Collins

PROJECT TITLE: THE GENETIC BASIS OF ACHILLES TENDON PATHOLOGY IN AN AUSTRALIAN POPULATION

Thank you for your letter to the Research Ethics Committee dated 29th September 2008.

It is a pleasure to inform you that the Ethics Committee has **approved** the amendment with reference to the above-mentioned study.

Please can we have, if not yet submitted an updated report for 289/2004.

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
pp **CHAIRPERSON, HSF HUMAN ETHICS**

lemjedi

THE UNIVERSITY OF NORTHAMPTON
School of Health

Thornby 2
Park Campus
Boughton Green Road
Northampton NN2 7AL
Tel: 01604 892816
Fax: 01604 791954

Dr Stuart Raleigh
Division of Sport, Exercise and Life Sciences
Brampton Building
Park Campus

Dear Stuart

Re: The Genetic Basis of Tendinopathy

Thank you for your application and associated paperwork for ethics approval for the above research project in collaboration with Professor Collins at the University of Cape Town.

I can now confirm that the ethics panel of the School of Health met on 1st October 2008 and has approved an extension to your original project involving genotyping Australian DNA samples.

May I take this opportunity to wish you well with your research.

Best wishes

Ali Ewing
Principal Lecturer

cc. Professor Carol Phillips
Associate Dean

To: **Seth O'Neill**

Subject: Ethical Application Ref: **so59-b950**

(Please quote this ref on all correspondence)

17/01/2014 09:21:07

Medical(&(Social(Care(Educa0on

Project Title: **A Study testing the genetic profile of people with Achilles tendinopathy**

Thank you for submitting your application which has been considered.

This study has been given ethical approval, subject to any conditions quoted in the attached notes.

Any significant departure from the programme of research as outlined in the application for research ethics approval (such as changes in methodological approach, large delays in commencement of research, additional forms of data collection or major expansions in sample size) must be reported to your Departmental Research Ethics Officer.

Approval is given on the understanding that the University Research Ethics Code of Practice and other research ethics guidelines and protocols will be compiled with

〈 http://www2.le.ac.uk/institution/committees/research-ethics/code-of-practice

〈 http://www.le.ac.uk/safety/

LA TROBE UNIVERSITY

MEMORANDUM

RESEARCH SERVICES

To:	**Dr Tom Samiric, School of Human Biosciences, Faculty of Health Sciences**
From:	**Secretary, La Trobe University Human Ethics Committee**
Subject:	**Review of Human Ethics Committee Application No. 12-086**
Title:	**Epigenetic investigation of normal and abnormal human patellar tendons**
Date:	**31 October 2012**

Thank you for your recent correspondence in relation to the research project referred to above. The project has been assessed as complying with the *National Statement on Ethical Conduct in Human Research*. I am pleased to advise that your project has been granted ethics approval and you may commence the study.

The project has been approved from the date of this letter until 31 December 2015.

Please note that your application has been reviewed by a sub-committee of the University Human Ethics Committee (UHEC) to facilitate a decision about the study before the next Committee meeting. This decision will require ratification by the full UHEC at its next meeting and the UHEC reserves the right to alter conditions of approval or withdraw approval. You will be notified if the approval status of your project changes. The UHEC is a fully constituted Ethics Committee in accordance with the National Statement on Ethical Conduct in Research Involving Humans- March 2007 under Section 5.1.29.

The following standard conditions apply to your project:

- **Limit of Approval.** Approval is limited strictly to the research proposal as submitted in your application while taking into account any additional conditions advised by the UHEC.

- **Variation to Project.** Any subsequent variations or modifications you wish to make to your project must be formally notified to the UHEC for approval in advance of these modifications being introduced into the project. This can be done using the appropriate form: *Ethics - Application for Modification to Project* which is available on the Research Services website at *http://www.latrobe.edu.au/research-services/ethics/HEC_human.htm*. If the UHEC considers that the proposed changes are significant, you may be required to submit a new application form for approval of the revised project.

- **Adverse Events.** If any unforeseen or adverse events occur, including adverse effects on participants, during the course of the project which may affect the ethical acceptability of the project, the Chief Investigator must immediately notify the UHEC Secretary on telephone (03) 9479 1443. Any complaints about the project received by the researchers must also be referred immediately to the UHEC Secretary.

- **Withdrawal of Project.** If you decide to discontinue your research before its planned

  completion, you must advise the UHEC and clarify the circumstances.

- **Annual Progress Reports.** If your project continues for more than 12 months, you are required to submit an *Ethics - Progress/Final Report Form* annually, **on or just prior to 12 February**. The form is available on the Research Services website (see above address). Failure to submit a Progress Report will mean approval for this project will lapse. An audit may be conducted by the UHEC at any time.

- **Final Report.** A Final Report (see above address) is required within six months of the completion of the project or by **30 June 2016.**

If you have any queries on the information above or require further clarification please contact me through Research Services on telephone (03) 9479-1443, or e-mail at: **humanethics@latrobe.edu.au**.

On behalf of the University Human Ethics Committee, best wishes with your research!

Ms Barbara Doherty
Administrative Officer (Research Ethics)
University Human Ethics Committee
Research Compliance Unit / Research Services
La Trobe University Bundoora, Victoria 3086
P: (03) 9479 – 1443 / F: (03) 9479 - 1464
http://www.latrobe.edu.au/research-services/ethics/HEC_human.htm

THE UNIVERSITY OF
NORTHAMPTON

The Graduate School
The University of Northampton
Boughton Green Road
Northampton
NN2 7AL

Dr Stuart Raleigh
School of Health

4th August 2015

Dear Dr Raleigh

I can confirm that the University of Northampton Research Ethics Committee considered and approved the following research proposals:

Project title	Research student	Research supervisors	Approval date
'Novel Genetic Factors that Predispose Humans to Achilles Tendon Pathologies'	Louis El Khoury	Dr Stuart Raleigh, Prof. Bill Ribbans, Dr Mike Posthumus	16 June 2011
'The possible role of copy number variation as a predisposing factor for Achilles tendon pathology'	Rebecca Rickaby	Dr Stuart Raleigh, Prof. Bill Ribbans, Dr Mike Posthumus	13 September 2012
Extension to 'Novel Genetic Factors that Predispose Humans to Achilles Tendon Pathologies'	Louis El Khoury, Rebecca Rickaby	Dr Stuart Raleigh, Prof. Bill Ribbans	4 March 2014

Approval was given by the institutions collecting the samples for them to be stored and used for future related research projects.

If I can be of any further assistance, please contact me.

Yours sincerely

David Watson
Postgraduate Research Manager
(Research Ethics Committee Officer)

Main switchboard 01604 735500
Study enquiries 0800 358 2232
Study email study@northampton.ac.uk
Web www.northampton.ac.uk

Park Campus
Boughton Green Road
Northampton
NN2 7AL

Avenue Campus
St George's Avenue
Northampton
NN2 6JD

A charity providing undergraduate and postgraduate degrees and diplomas

Vice Chancellor Professor Nick Petford BSc, PhD, DSc, FGS, ADM

APPENDIX II

Participant Information Sheet & Consent Form

PARTICIPANT INFORMATION

Background

Thank you for showing an interest in this new study that is going to be conducted at The University of Northampton. The study is being carried out by Dr Stuart Raleigh who is a Reader and a researcher in Molecular Biology, Professor William Ribbans who is a consultant orthopaedic surgeon and Mr Louis El Khoury who is a researcher and PhD student. All three are based at the University of Northampton's School of Health and Professor Ribbans is also based at the County Clinic, Northampton. The contact details of the team are found towards the end of this leaflet. Please feel free to contact them at any time if you require any additional information.

What is the study about?

We are trying to establish whether certain individuals carry particular versions of genes (known as alleles) that may increase their risk of Achilles tendon problems. We are particularly interested to learn more about the types of Achilles tendon problems called Achilles tendinopathy (swollen, painful tendon) and Achilles tendon rupture. The study is about Achilles tendon problems in humans – we never use animals – and we would greatly value you as a volunteer.

Why are we interested?

We are interested because many people with relatively active lifestyles as well as professional athletes experience Achilles tendon problems during or after sports activities. This can sometimes lead to long term disability. In some cases the injury needs surgery and some individuals are reluctant, or have difficulty returning to a sport, or physical activity, they once enjoyed. This research might eventually lead to a way of reducing the number of people who get tendon problems in the future or to better treatment for those already affected.

What's involved if I participate?

We would like individuals to donate a small sample of their spit (saliva) for DNA analysis. We only want about 2 ml of your saliva (the same as one teaspoonful!) which can be taken from simply spitting into a tube. The procedure would take about 3 minutes to perform. You don't even have to attend a clinic. We can send you the tube to spit into. We would also like you to answer some background questions and details about your health and any exercise that you do. If you would like to take part, please complete, sign, and return the consent from in the paid post envelope and we will send you a questionnaire and a saliva collection tube. The total time taken to collect the saliva sample and complete the questionnaire is about 10-15 minutes.

What will happen to my saliva sample?

Your saliva sample will ONLY be used for Achilles tendon studies. Once we receive your sample, it will be labelled with a unique number only and stored in a secure freezer at the University of Northampton's Park campus. We will isolate DNA from your saliva sample. We will then compare the genes between various groups of individuals to establish whether samples from people who have had Achilles tendon problems contain different versions compared to people who don't have tendon problems. We would like to investigate many genes from your sample and in order for us to do this

we would like to retain your sample indefinitely. There may be occasions when we might want to send a small amount of your sample to scientists that we do joint studies with. This would only be done if they could help us with our study. However, if this does happen the sample they receive will only have the number on it and it would not be possible for you to be identified.

Will my sample be used for studies investigating any other disease?
No. The genes that we are interested in analysing are only relevant to musculoskeletal soft tissue injuries (tendon injuries). It will not be possible for us to investigate other illnesses such as: cancer and cardiovascular diseases.

Will the information I provide be kept private?
Yes. We will not send any of your private contact details to any other party. Your personal contact details will only be seen by the three members of the study team. Indeed your personal contact details will be stored securely in a locked filing cabinet at the university and on a password protected computer. The data will not be on any networked computer drives. To make things even more secure your saliva sample and the DNA that we isolate from it will only be labelled with a number. It would not be possible for any person, other than the study team, to identify you from the number that your sample will be allocated.

How long will you be keeping my personal information?
We would like to retain you personal information for an indefinite amount of time. Nevertheless, if you want to end your participation in this project at any time we will immediately destroy your personal information as well as your DNA sample. However any data obtained from the analysis of your sample up until that point will remain part of the research and will be included in any published work.

Why do you need information about my previous medical history?
We need to ask these questions because certain ailments that you may have had in the past might be important to Achilles tendon studies. If we don't have the data we could make a false conclusion in our study.

Why do you need to know about my lifestyle?
This is because it may be lifestyle factors that are most important in causing people to have Achilles problems. Providing this information will help us to establish whether this is true or not.

Why do you need to know about my ethnicity?
This is because certain genetic factors might be different in people of different ethnicity. Having details about a person's ethnicity can help us build a more accurate assessment of the role of genes in causing any Achilles tendon problems. Without details on ethnicity the data we generate might be flawed.

Why do you need my name and contact details?
This is for two main reasons. Firstly we may want to invite you to be part of some future study and might send you a letter. Of course, it would be entirely up to you whether or not you replied to any future invites. You can always send us a request not contact you for any future studies. Secondly, very rarely, samples that are provided for a study can be accidently damaged during the laboratory procedure. If this does happen we would like to think that we could contact you again for a replacement sample. If we were not able to contact you then the valuable sample that you kindly donated would have been lost from the investigation and this might lead to less accurate scientific results.

What happens if I want to be involved but later change my mind?
You simply contact us and let us know. We will then immediately destroy any sample and information that you have donated to us.

Will I be able to obtain information about my own DNA sample?
The study that we are conducting is aimed at understanding how genetic factors affect the Achilles tendon by studying a group of people. It will be difficult to predict any risk on an individual basis. This is why this study is not designed to test individual genetic status. For this reason we do not plan to give back individual results to participants. If you require addition information about this, please contact us on the numbers or e-mails provided on the last page.

How can I obtain information about the progress of the research?
After roughly 18 months, information about the preliminary results of the study will be posted on the University of Northampton School of Health website. For more information please visit:
(www.northampton.ac.uk/info/20038/school-of-health)

What will happen to the data obtained from this study?
We would like to publish this in scientific journals. The publications would never reveal any participant identity or contact details.

What benefit do I get from being involved in this study?
Many scientific studies rely on sample donations from the general population. By taking part in this study you will help to further our understanding of why some people, particularly active sportspeople, develop Achilles tendon problems. Eventually the findings may help people to be treated for the condition. Furthermore, it might be possible that in the future we could identify those most at risk of developing the problem. If this is the case we could educate people about the suitability of sports that they do and advise them on their risk. The finding may also be used to enhance our understanding of how human tendons work under stressed conditions.

We thank you for reading the above information. If you feel that you would like to be involved then please complete and return the consent form on the next page. Please remember that if you have any questions then please contact the study team who will be very pleased to talk you through any aspect of this study.

Contact Details

Louis El Khoury
MX05
University of Northampton
Boughton Green Road
Northampton – UK
NN2 7AL
Tel: 01604 892512
louis.elkhoury@northampton.ac.uk

Dr Stuart Raleigh
School of Health
Kelmarsh Building
Boughton Green Road
Northampton – UK
NN2 7AL
Tel: 01640 892306
stuart.raleigh@northampton.ac.uk

Prof William Ribbans
The County Clinic
57 Billing Road
Northampton – UK
NN1 5DB
Tel: 01604 795414
wjribbans@uk-consultants.co.uk

THE UNIVERSITY OF NORTHAMPTON

CONSENT FORM

If you would like to participate in this study then please complete the consent form below. If you have any further questions relating to the study, or if there is something that you wish to discuss with us then please contact us at the addresses provided above.

Please tick the box that corresponds to your choice.

	Yes	No
I have read the study information sheet and understand what is involved.	☐	☐
I understand that my sample will be analysed anonymously and my contact details kept confidential. My contact details will only be used by the researcher if my sample is accidently damaged. It would be up to me whether I provided another sample.	☐	☐
I understand that I can withdraw my participation at any time and that my sample and information would be destroyed upon my request.	☐	☐
I am willing to participate in this project.	☐	☐
I would be interested in receiving information about future studies. My agreement to this would mean my contact details would be retained securely and for this purpose alone.	☐	☐

Signed: Date:

Name (Print)
Address

Phone Number E-mail:

We appreciate your time and your willingness to participate in this study.
Participant number (to be entered by the research team)

☐

APPENDIX III

Participant Questionnaires

Achilles Tendon Datasheet - Controls

We greatly appreciate your interest in participating in this project. Please complete as much of this questionnaire as possible keeping in mind the following points:

1. This questionnaire is printed on both sides of every sheet. Please pay attention at answering the questions on the back of each sheet.
2. Throughout the questionnaire you might come across scientific terms followed by small numbers such as: ACL[5]. These terms are explained in the glossary on the last page of this questionnaire.
3. Please feel free to skip any question you are having difficulty with.
4. Please mail us back the questionnaire and a team member will contact you to go over any question you are having difficulty with.
5. If you have any concern about the aspect of this project please feel free to contact us on the numbers or e-mails below.

Rebecca Rickaby
Tel: 01604 892893
rebecca.rickaby@northampton.ac.uk

Louis El Khoury
Tel: 01604 892512
louis.elkhoury@northampton.ac.uk

Dr Stuart Raleigh
Tel: 01604 892306
stuart.raleigh@northampton.ac.uk

Prof William Ribbans
Tel: 01604 795414
wjribbans@uk-consultants.co.uk

ACHILLES TENDON STUDY – Control

We would greatly appreciate you providing us with your personal details. However if you prefer not to fill in this section, please proceed to the next page.

SECTION 1: PERSONAL DETAILS			
Study Number To be completed by research team so leave blank			
Surname		First Name	
Address		Contact Telephone Number	
e-mail		Occupation	

<table>
<tr><td>Date of birth
Age</td><td colspan="5">DD/MM/Year Years and months</td></tr>
<tr><td>Height</td><td colspan="2">Cms:</td><td>Ft and Ins:</td><td>Gender</td><td>male female</td></tr>
<tr><td>Weight
(in
underwear)</td><td colspan="2">Kgs:</td><td>Stone & Lbs:</td><td>BMI
To be completed by the research team so leave blank.</td><td></td></tr>
<tr><td>Ethnic group</td><td colspan="5">Black/African White Asian
Mixed Ancestry Chinese Other</td></tr>
<tr><td>Nationality</td><td colspan="3"></td><td>Dominant Hand</td><td>Left Right</td></tr>
<tr><td>Country of Birth</td><td colspan="5"></td></tr>
<tr><td rowspan="3">Smoker</td><td>Yes (Current)</td><td colspan="3">Yes (Ex smoker)</td><td>No, never</td></tr>
<tr><td colspan="2">If yes, Number of years ________</td><td colspan="3">If stopped, when ________________</td></tr>
<tr><td colspan="5">If yes, number per day __________</td></tr>
<tr><td rowspan="2">Do you know your blood group?</td><td>Yes</td><td colspan="4">A B AB O</td></tr>
<tr><td>No</td><td colspan="4">Rh Pos Rh Neg</td></tr>
</table>

If you participate or have participated in different sports or physical activities, please complete section 2 on Sporting and Recreational Details.

SECTION 2: SPORTING AND RECREATIONAL DETAILS			
Type of sport/ physical activity you have participated in	**Sport 1**	**Sport 2**	**Sport 3**
Current or past participation	Current Past	Current Past	Current Past
Year started participation			
Years involved in the sport			
Years in competitive sport			
Have you been involved in the sport at a professional level? If so, please state how many years.			
Level of Participation: e.g. Leisure/Club/Regional/International			
Average hours of training per week - in the last 12 months			
Type of sport/ physical activity you have participated in	**Sport 4**	**Sport 5**	**Sport 6**
Current or past participation	Current Past	Current Past	Current Past
Year started participation			
Years involved in the sport			
Years in competitive sport			
Have you been involved in the sport at a professional level? If so, please state how many years.			
Level of Participation: e.g. Leisure/Club/Regional/International			
Average hours of training per week - in the last 12 months			

If you Run/Jog, please specify the distance crossed every week: ....

SECTION 3: GENERAL MEDICAL DETAILS

Do you suffer from any Connective Tissue and Rheumatological Diseases and Disorders[1]?

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

Has any member of your family (Blood relatives) suffered from any Achilles tendon injury/problem?	Yes ☐ No ☐ Unknown ☐	If Yes, please specify the family member (eg Mother, Son) and type of injury ____________ Rupture ☐ Swelling ☐ Other ☐	
Do you suffer from elevated blood cholesterol?	Yes ☐ No ☐	Do any other members of your family suffer from elevated blood cholesterol?	Yes ☐ No ☐ Unknown ☐
Have you been diagnosed with any of the following diseases?	☐ Diabetes mellitus ☐ Adrenal disorders ☐ Thyroid disorders ☐ Amyloidosis ☐ Renal disease ☐ Other endocrine and metabolic disease (Specify) ____________ ____________ ____________		

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[2] (cortisone tablets)?	Yes ☐ No ☐ Unknown ☐	☐ 3 months ☐ 12 months	☐ 6 months ☐ 24 or more months
Have you ever been given an injection with corticosteroids?	Yes ☐ No ☐ Unknown ☐	☐ 3 months ☐ 12 months	☐ 6 months ☐ 24 or more months
Have you ever been given an injection of corticosteroids in or around a tendon?	Yes ☐ No ☐ Unknown ☐	☐ Once ☐ 3 times	☐ Twice ☐ >3 times
Have you ever used anabolic steroids[3]?	Yes ☐ No ☐ Unknown ☐	☐ 3 months ☐ 12 months	☐ 6 months ☐ 24 or more months
Have you ever used fluoroquinolone antibiotics? (see list below)	Yes ☐ No ☐ Unknown ☐	☐ 3 months ☐ 12 months	☐ 6 months ☐ 24 or more months

If yes, please select from the list below:		
☐ CIPROFLOXACIN	☐ MOXIFLOXACIN	☐ NORFLOXACIN
☐ CIPROXIN	☐ AVELOX	☐ UTINOR
☐ LEVOFLOXACIN	☐ NALIDIXIC ACID	☐ OFLOXACIN
☐ TAVANIC	☐ URIBEN	☐ TARVID

What medication, if any, are you currently using? (please list)	
Do you suffer from any allergy? (please list)	Yes ☐ No ☐ Unknown ☐

SECTION 4: HISTORY OF LIGAMENT AND TENDON INJURIES/PROBLEMS

Question	Response			
Have you ever had a ligament[4] injury/problem in the past?	Yes ▯ No ▯ Unknown ▯			
If yes, please specify which ligaments? (You may tick more than one block, please select either L (left) or R (right)) The Abreviated medical terms are explained in the glossary on the last page.		L R		L R
	Knee (ACL)[5]	▯ ▯	Knee (PCL)[7]	▯ ▯
	Knee (MCL)[6]	▯ ▯	Knee (LCL)[8]	▯ ▯
	Ankle lateral ligaments	▯ ▯	Ankle medial ligaments	▯ ▯
	Spinal ligaments	▯ ▯	Finger ligaments	▯ ▯
	Shoulder ligaments	▯ ▯	Wrist ligaments	▯ ▯
	Elbow ligaments	▯ ▯	Other ligaments	▯ ▯
To your knowledge, have any other members of your family suffered from any ligament injury/problem?	Yes ▯ No ▯ Unknown ▯	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[9] in the past?	Yes ▯ No ▯ Unknown ▯			
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 injury/problem?	Yes ▯ No ▯ Unknown ▯	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 injuries?	▯ Acute shoulder dislocation ▯ Chronic shoulder instability ▯ Chronic ankle instability ______________________________ ______________________________	

Thank you for taking time to complete this questionnaire. Your participation is greatly appreciated. Please go on to sign the consent document that you have been sent.

Please return your **saliva sample**, **completed questionnaire** and **consent form** to us as described in the "participant information" sheet.

If you are unsure about any aspect of this project or require further information please contact the study team using the contact numbers, or emails, found on the front page of this questionnaire.

Glossary

1. **Rheumatologic disease**: A type of disease involving inflammation of muscles, joints, and other tissues
2. **Corticosteroid**: a group of steroid hormones used to treat inflammation
3. **Anabolic steroid**: a synthetic steroid hormone that resembles testosterone in promoting the growth of muscle. Such hormones are used medicinally to treat some forms of weight loss and, by some athletes and others to enhance physical performance
4. **Ligament**: A short band of tough, flexible, fibrous connective tissue that connects two bones or cartilages or holds together a joint
5. **ACL**: Anterior cruciate ligament
6. **MCL**: medial collateral ligament
7. **PCL**: Posterior cruciate ligament
8. **LCL**: Lateral collateral ligament
9. **Tendon**: A flexible but inelastic cord of strong fibrous collagen tissue attaching a muscle to a bone

Achilles Tendon Datasheet - Cases

We greatly appreciate your interest in participating in this project. Please complete as much of this questionnaire as possible keeping in mind the following points:

6. This questionnaire is printed on both sides of every sheet. Please pay attention at answering the questions on the back of each sheet.
7. Throughout the questionnaire you might come across scientific terms followed by small numbers such as: ACL[5]. These terms are explained in the glossary on the last page of this questionnaire.
8. Please feel free to skip any question you are having difficulty with.
9. Please mail us back the questionnaire and a team member will contact you to go over any question you are having difficulty with.
10. If you have any concern about the aspect of this project please feel free to contact us on the numbers or e-mails below.

Louis El Khoury
Tel: 01604 892512
louis.elkhoury@northampton.ac.uk

Dr Stuart Raleigh
Tel: 01604 892306
stuart.raleigh@northampton.ac.uk

Prof William Ribbans
Tel: 01604 795414
wjribbans@uk-consultants.co.uk

ACHILLES TENDON DATASHEET – Case

We would greatly appreciate you providing us with your personal details. However if you prefer not to fill in this section, please proceed to the next page.

SECTION 1: PERSONAL DETAILS			
Study Number To be completed by research team so leave blank			
Surname		First Name	
Address		Contact Telephone Number	
e-mail		Occupation	

<table>
<tr><td>Date of birth
Age</td><td colspan="4">DD/MM/Year Years and months</td></tr>
<tr><td>Height</td><td>Cms:</td><td>Ft and Ins:</td><td>Gender</td><td>male female</td></tr>
<tr><td>Weight (currently and at first injury)</td><td>Kgs:</td><td>Stone & Lbs:</td><td>BMI
To be completed by the research team so leave blank.</td><td></td></tr>
<tr><td>Ethnic group</td><td>Black/African
Mixed Ancestry</td><td>White
Chinese</td><td colspan="2">Asian
Other</td></tr>
<tr><td>Nationality</td><td colspan="2"></td><td>Dominant Hand</td><td>Left Right</td></tr>
<tr><td>Country of Birth</td><td colspan="4"></td></tr>
<tr><td rowspan="3">Smoker</td><td>Yes (Current)</td><td colspan="2">Yes (Ex smoker)</td><td>No, never</td></tr>
<tr><td>If yes, Number of years ________</td><td colspan="3">If stopped, when ______________</td></tr>
<tr><td colspan="4">If yes, number per day _________</td></tr>
<tr><td rowspan="2">Do you know your blood group?</td><td>Yes</td><td>A</td><td>B</td><td>AB O</td></tr>
<tr><td>No</td><td>Rh Pos</td><td colspan="2">Rh Neg</td></tr>
</table>

If you participate or have participated in different sports, please complete section 2 on Sporting and Recreational Details.

SECTION 2: SPORTING AND RECREATIONAL DETAILS			
Type of sport(s) you have participated in	**Sport 1**	**Sport 2**	**Sport 3**
Current or past participation	Current Past	Current Past	Current Past
Year started participation			
Years involved in the sport			
Years in competitive sport			
Have you been involved in the sport at a professional level? If so, please state how many years.			
Level of Participation: e.g. Leisure/Club/Regional/International			
Average hours of training per week - in the last 12 months			
Type of sport(s) you have participated in	**Sport 4**	**Sport 5**	**Sport 6**
Current or past participation	Current Past	Current Past	Current Past
Year started participation			
Years involved in the sport			
Years in competitive sport			
Have you been involved in the sport at a professional level? If so, please state how many years.			
Level of Participation: e.g. Leisure/Club/Regional/International			
Average hours of training per week - in the last 12 months			

If you Run/Jog, please specify the distance crossed every week: ....

SECTION 3: GENERAL MEDICAL DETAILS

Do you suffer from any Connective Tissue and Rheumatological Diseases and Disorders[1]?

☐ Ankylosing Spondylitis
☐ Aspartylglycosaminuria (AGU)
☐ Behcet's Syndrome
☐ Crohn's Disease
☐ Discoid Lupus Erythematosus
☐ Ehlers-Danlos syndrome (EDS)
☐ Eosinophilic Fascitis
☐ Giant Cell (Temporal) Arthritis
☐ Gout
☐ Hypertensive Vasculitis
☐ Lipid Storage Diseases
☐ Marfan Syndrome
☐ Menkes Kinky Hair Syndrome
☐ Mucopolysaccharidoses
☐ Myopathies and Dystrophies
☐ Ochronosis (Homocystinuria)
☐ Osteoarthritis
☐ Osteogenesis imperfecta (OI)
☐ Polyarteritis Nodosa
☐ Polymyalgia Rheumatica
☐ Polymyositis & Dermatomyositis
☐ Pseudogout
☐ Reactive Arthritis
☐ Reiter's Syndrome
☐ Relapsing Polychrondritis
☐ Rheumatoid Arthrits
☐ Scleroderma
☐ Sjogren's Syndrome
☐ Systemic Lupus Erythematosus (SLE)
☐ Systemic Sclerosis
☐ Wegener's Granulomatosis
☐ No, I don't suffer of any of these diseases

Have any other members of your family (Blood relatives) suffered from any Achilles tendon injury/problem?	Yes ☐ No ☐ Unknown ☐	If Yes, please specify the family member (eg Mother, Son) and type of injury ____________ Rupture ☐ Swelling ☐ Other ☐	
Do you suffer from elevated blood cholesterol?	Yes ☐ No ☐	Do any other members of your family suffer from elevated blood cholesterol?	Yes ☐ No ☐ Unknown ☐
Have you been diagnosed with any of the following diseases?	☐ Diabetes mellitus ☐ Adrenal disorders ☐ Thyroid disorders ☐ Amyloidosis ☐ Renal disease ☐ Other endocrine and metabolic disease (Specify) ____________ ____________ ____________		

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[2] (cortisone tablets)?	Yes ▯ No ▯ Unknown ▯	▯ 3 months ▯ 12 months	▯ 6 months ▯ 24 or more months
Have you ever been given an injection with corticosteroids?	Yes ▯ No ▯ Unknown ▯	▯ 3 months ▯ 12 months	▯ 6 months ▯ 24 or more months
Have you ever been given an injection of corticosteroids in or around a tendon?	Yes ▯ No ▯ Unknown ▯	▯ Once ▯ 3 times	▯ Twice ▯ >3 times
Have you ever used anabolic steroids[3]?	Yes ▯ No ▯ Unknown ▯	▯ 3 months ▯ 12 months	▯ 6 months ▯ 24 or more months
Have you ever used fluoroquinolone antibiotics? (see list below)	Yes ▯ No ▯ Unknown ▯	▯ 3 months ▯ 12 months	▯ 6 months ▯ 24 or more months
If yes, please select from the list below:			
▯ CIPROFLOXACIN ▯ CIPROXIN ▯ LEVOFLOXACIN ▯ TAVANIC	▯ MOXIFLOXACIN ▯ AVELOX ▯ NALIDIXIC ACID ▯ URIBEN	▯ NORFLOXACIN ▯ UTINOR ▯ OFLOXACIN ▯ TARVID	
What medication, if any, are you currently using? (please list)			
Do you suffer from any allergy? (please list)	Yes ▯ No ▯ Unknown ▯		

SECTION 4: HISTORY OF LIGAMENT AND TENDON INJURIES/ PROBLEMS

Have you ever had a ligament[4] injury/problem in the past?	Yes ▯ No ▯ Unknown ▯			
If yes, please specify which ligaments? (You may tick more than one block, please select either L (left) or R (right)) The abbreviated medical terms are explained in the glossary on the last page.		L R		L R
	Knee (ACL)[5]	▯ ▯	Knee (PCL)[7]	▯ ▯
	Knee (MCL)[6]	▯ ▯	Knee (LCL)[8]	▯ ▯
	Ankle lateral ligaments	▯ ▯	Ankle medial ligaments	▯ ▯
	Spinal ligaments	▯ ▯	Finger ligaments	▯ ▯
	Shoulder ligaments	▯ ▯	Wrist ligaments	▯ ▯
	Elbow ligaments	▯ ▯	Other ligaments	▯ ▯
To your knowledge, have any other members of your family suffered from any ligament injury/problem?	Yes ▯ No ▯ Unknown ▯	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 had a tendon[9] injury/problem in the past?	Yes ▯ No ▯ Unknown ▯		
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:..		

<table>
<tr><td>To your knowledge, have any other members of your family suffered from any tendon injury/problem?</td><td>Yes ▯ No ▯

Unknown ▯</td><td>If Yes, please specify the family member
▯ Mother
▯ Father
▯ Sibling
▯ Son / daughter
▯ Other family member:..........................

Condition: Please choose tendon injury from the list above
....</td></tr>
<tr><td>Have you ever suffered from any of the following injuries?</td><td colspan="2">▯ Acute shoulder dislocation
▯ Chronic shoulder instability
▯ Chronic ankle instability
____
______________________________</td></tr>
</table>

SECTION 5: DETAILS OF ACHILLES TENDON INJURY

How many times have you had tendon injuries/problems? *Sudden onset is within a few seconds or minutes ^Gradual onset is over days or weeks	Event	Date	Right or Left	Acute or Chronic Injury	Sudden* or Gradual^ Onset	Effect of injury in first month following injury	Grade of injury currently
	1.					☐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 ☐None of the above ☐Not sure	☐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 ☐None of the above ☐Not sure
	2.					☐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 ☐None of the above ☐Not sure	☐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 ☐None of the above ☐Not sure
	3.					☐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 ☐None of the above ☐Not sure	☐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 ☐None of the above ☐Not sure

	4.					☐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 ☐None of the above ☐Not sure	☐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 ☐None of the above ☐Not sure
	5.					☐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 ☐None of the above ☐Not sure	☐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 ☐None of the above ☐Not sure

Thank you for taking time to complete this questionnaire. Your participation is greatly appreciated.

Please can you return your **saliva sample**, and the **completed questionnaire** to us as described in the 'participant information' sheet.

If you are unsure about any aspect of this project or require further information please contact the study team using the contact numbers, or emails, found on the front page of this questionnaire.

Glossary

10. **Rheumatologic disease**: A type of disease involving inflammation of muscles, joints, and other tissues
11. **Corticosteroid**: a group of steroid hormones used to treat inflammation
12. **Anabolic steroid**: a synthetic steroid hormone that resembles testosterone in promoting the growth of muscle. Such hormones are used medicinally to treat some forms of weight loss and, by some athletes and others to enhance physical performance
13. **Ligament**: A short band of tough, flexible, fibrous connective tissue that connects two bones or cartilages or holds together a joint
14. **ACL**: Anterior cruciate ligament
15. **MCL**: medial collateral ligament
16. **PCL**: Posterior Cruciate Ligament
17. **LCL**: Lateral collateral ligament
18. **Tendon**: A flexible but inelastic cord of strong fibrous collagen tissue attaching a muscle to a bone

SECTIONS 7, 8, 9 AND 10 TO BE FILLED OUT BY CLINICIAN

SECTION 7: ASSESSMENT OF ACHILLES TENDON INJURY BY CLINICIAN		
Event Number (1,2,3,4,or 5)		
FINDING	YES	NO
Symptoms > 6/12		
Early morning pain over the Achilles area		
Early morning stiffness over the Achilles area		
History of swelling over the Achilles area		
Tenderness to palpation over the Achilles area		
Palpable nodular thickening over the Achilles tendon		
Positive Shift test: Movement of painful/swollen area with DF/PF of ankle		
Other comments:		

SECTION 8: DESCRIPTION OF ACHILLES TENDON INJURY Please complete a separate form for each Tendon Injury			Event Number:	
			Right	Left
Rupture:				
	Full			
	Partial			
	Insertional avulsion			
Tendinopathy:				
	Non-insertional:			
		Peritendonitis		
		Tendinosis		
		Peritendonitis + Tendinosis		
		Other		
		Unknown		
	Insertional:			
		Haglund's		
		Lateral Calcaneal Ridge		
		Intra-tendinous calcification		
		Achilles Bursitis - superficial		
		Achilles Bursitis -deep		
		Tendinosis		

SECTION 9: IMAGING OF ACHILLES TENDON INJURY Please complete a separate form for each Tendon Injury			Event Number:
Type	Date	Location	Details

SECTION 10: DETAILS O F TREATMENT Please complete a separate form for each Tendon Injury				Event Number:
Date	Type	Location	Duration	Outcome

APPENDIX IV

Protocols

prepIT•L2P

Quick reference guide:

Laboratory protocol for manual purification of DNA from 0.5 mL of sample

Purification steps
1. Mix the sample in the DNA Genotek kit by inversion and gentle shaking for a few seconds.
2. Incubate the sample at 50°C in a water incubator for a minimum of 1 hour or in an air incubator for a minimum of 2 hours.
3. Transfer 500 µL of the sample to a microcentrifuge tube.
4. Add 20 µL of PT-L2P and mix by vortexing for a few seconds.
5. Incubate on ice for 10 minutes.
6. Centrifuge at room temperature (RT) for 5 minutes at 15,000 x *g*.
7. Carefully transfer the majority of the clear supernatant with a pipette to a fresh microcentrifuge tube. **Discard the pellet.**
8. Add 600 µL of RT 95% to 100% ethanol to the clear supernatant. Mix gently by inversion 10 times.
9. Let the sample stand at RT for 10 minutes to allow the DNA to fully precipitate.
10. Place the tube into the centrifuge with a known orientation. Centrifuge at RT for 2 minutes at 15,000 x *g*.
11. Carefully pipette off the supernatant and discard it. Take care to avoid disturbing the DNA pellet.
12. Add 250 µL of 70% ethanol and let stand at RT for 1 minute. **Completely remove the ethanol, without disturbing the pellet.**
13. Add 100 µL of TE solution and vortex the sample for at least 5 seconds.
14. Incubate overnight at RT or at 50°C for 1 hour vortexing occasionally.
15. Storage: In aliquots at -20°C for long-term storage (recommended) or at 4°C for up to 2 months.

PD-PR-006 Issue 10/2012-04
© 2012 DNA Genotek Inc., a subsidiary of OraSure Technologies, Inc. all rights reserved.

APPENDIX V

QPCR Efficiency

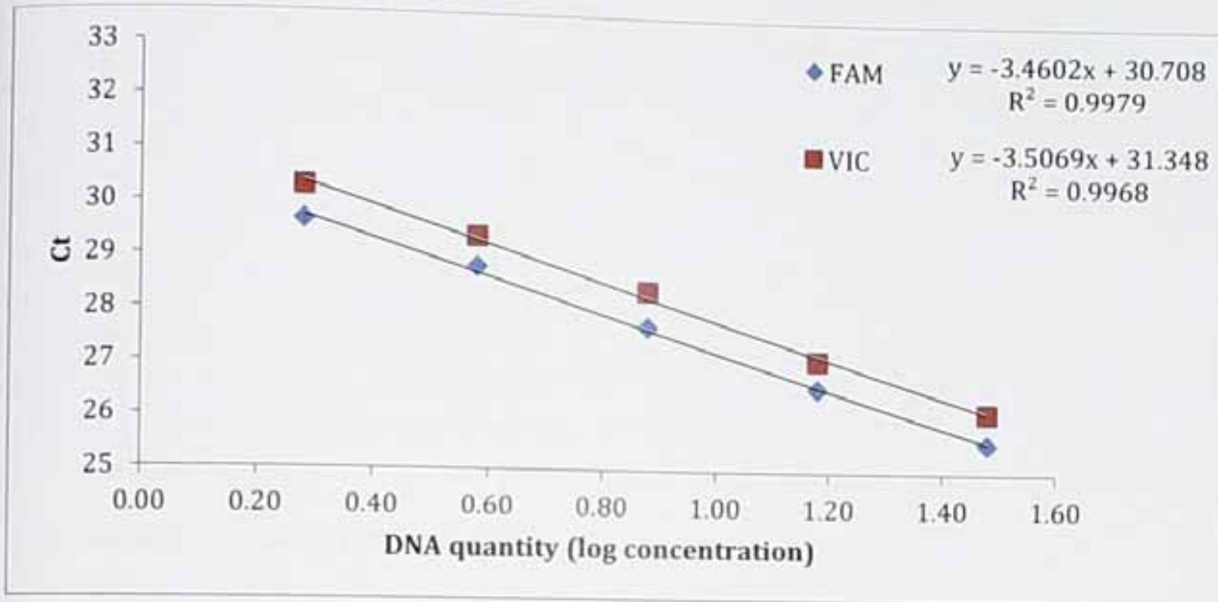


Figure V.1 *ADAM8* Hs02429708_cn

Standard curve for the *ADAM8* Hs02429708_cn (FAM) and Reference Assay RNase P (VIC). A linear trend line was fit to the data to determine the PCR efficiency for each assay, plotting cycle threshold (Ct) against DNA concentration (log concentration). Efficiency (E) was calculated using the equation $E = 10^{(-1/m)}-1$, where m is the slope of the linear trend line. Assays gave efficiencies of 95% and 93%, for FAM and VIC respectively.

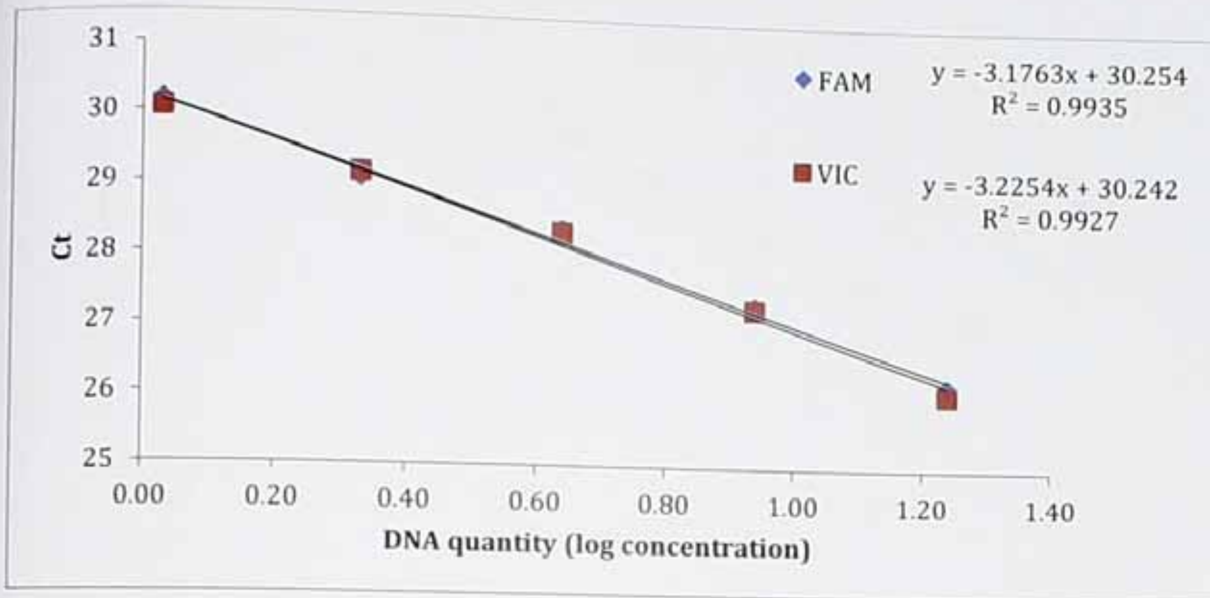


Figure V.2 *CASP8* Hs02601709_cn

Standard curve for the *CASP8* Hs02601709_cn (FAM) and Reference Assay RNase P (VIC). A linear trend line was fit to the data to determine the PCR efficiency for each assay, plotting cycle threshold (Ct) against DNA concentration (log concentration). Efficiency (E) was calculated using the equation $E = 10^{(-1/m)}-1$, where m is the slope of the linear trend line. Assays gave efficiencies of 106% and 104%, for FAM and VIC respectively.

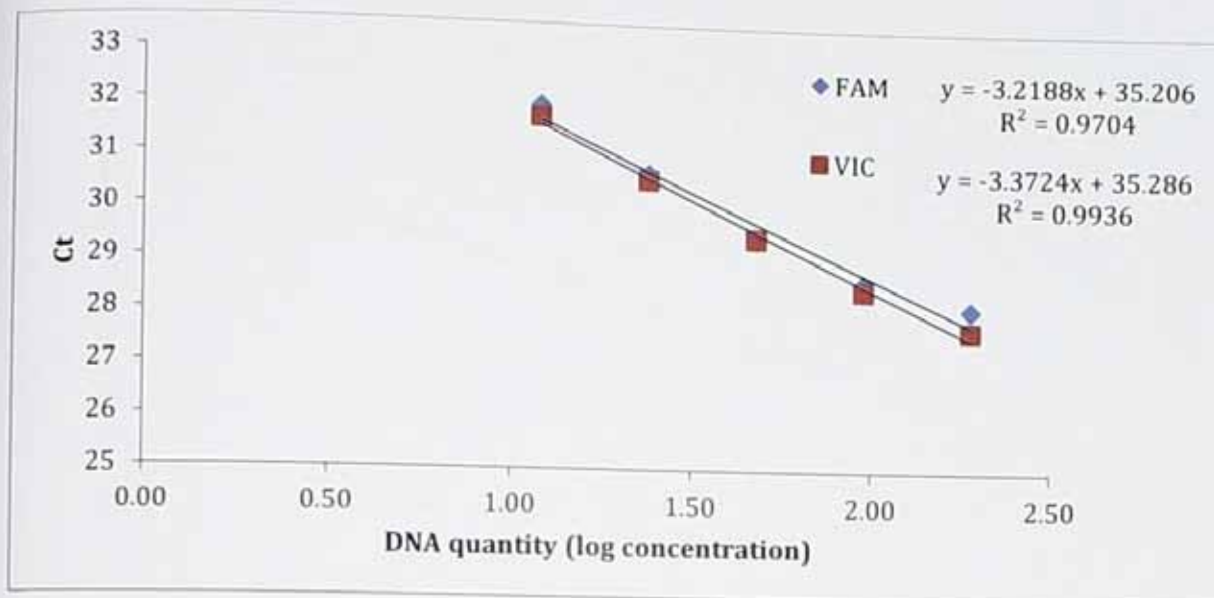


Figure V.3 *COL4A1* Hs00739915_cn

Standard curve for the *COL4A1* Hs00739915_cn (FAM) and Reference Assay RNase P (VIC). A linear trend line was fit to the data to determine the PCR efficiency for each assay, plotting cycle threshold (Ct) against DNA concentration (log concentration). Efficiency (E) was calculated using the equation $E = 10^{(-1/m)}-1$, where m is the slope of the linear trend line. Assays gave efficiencies of 104% and 98%, for FAM and VIC respectively.

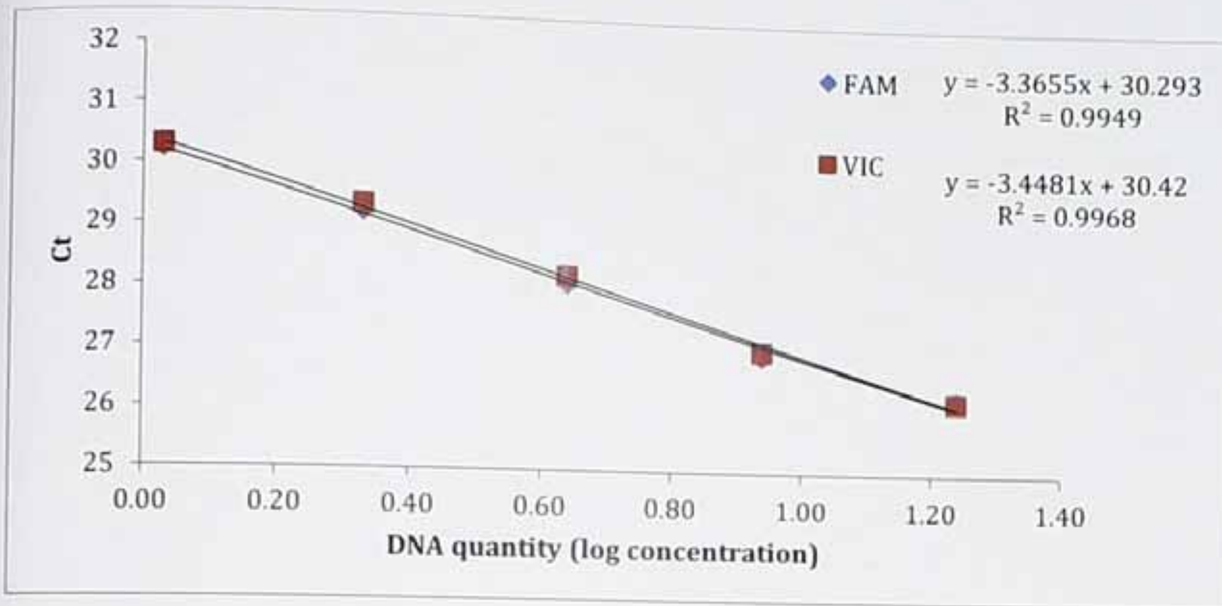


Figure V.4 *COL4A1* Hs02119212_cn

Standard curve for the *COL4A1* Hs02119212_cn (FAM) and Reference Assay RNase P (VIC). A linear trend line was fit to the data to determine the PCR efficiency for each assay, plotting cycle threshold (Ct) against DNA concentration (log concentration). Efficiency (E) was calculated using the equation $E = 10^{(-1/m)} - 1$, where m is the slope of the linear trend line. Assays gave efficiencies of 98% and 95%, for FAM and VIC respectively.

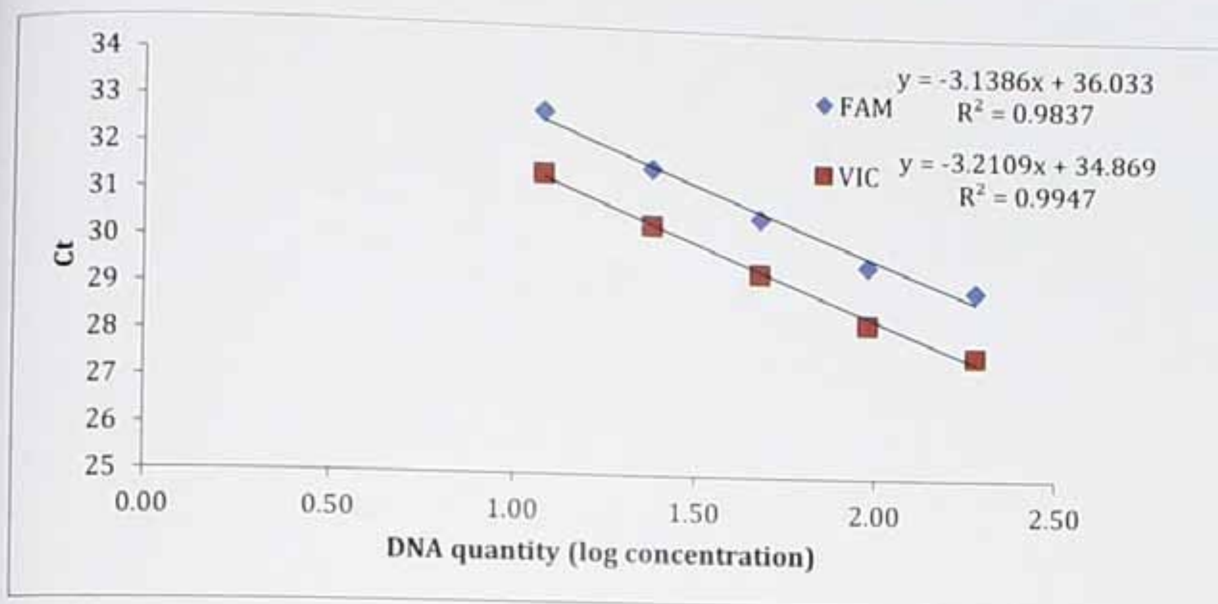


Figure V.5 *COL5A1* Hs00180523_cn

Standard curve for the *COL5A1* Hs00180523_cn (FAM) and Reference Assay RNase P (VIC). A linear trend line was fit to the data to determine the PCR efficiency for each assay, plotting cycle threshold (Ct) against DNA concentration (log concentration). Efficiency (E) was calculated using the equation $E = 10^{(-1/m)} - 1$, where m is the slope of the linear trend line. Assays gave efficiencies of 108% and 105%, for FAM and VIC respectively.

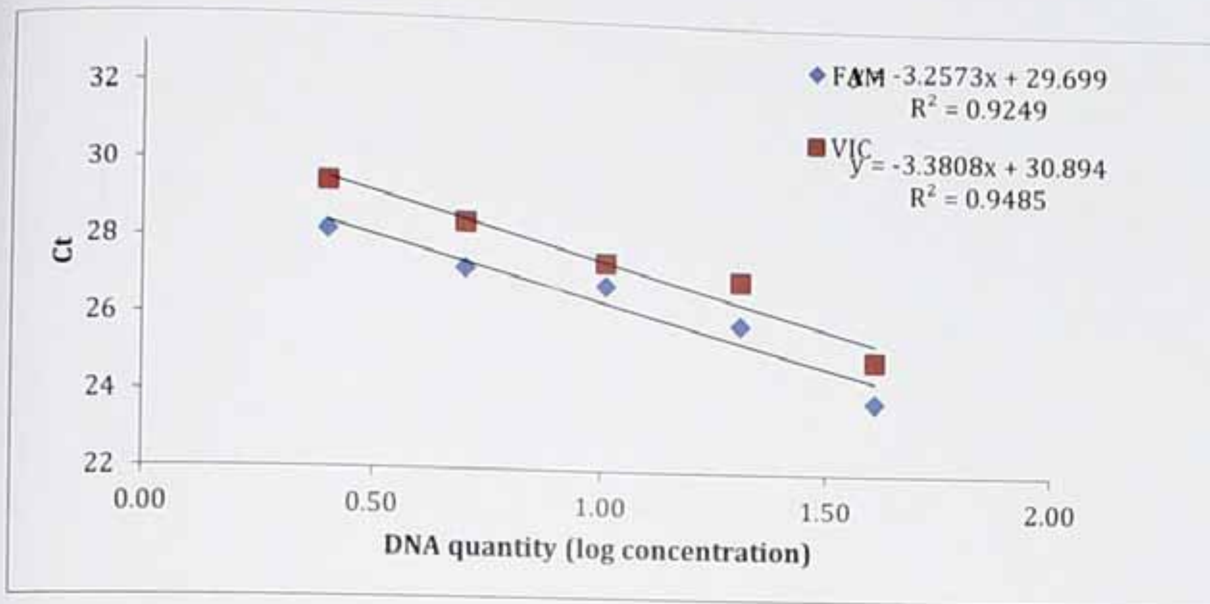


Figure V.6 *COL5A1* Hs02980647_cn

Standard curve for the *COL5A1* Hs02980647_cn (FAM) and Reference Assay RNase P (VIC). A linear trend line was fit to the data to determine the PCR efficiency for each assay, plotting cycle threshold (Ct) against DNA concentration (log concentration). Efficiency (E) was calculated using the equation $E = 10^{(-1/m)}-1$, where m is the slope of the linear trend line. Assays gave efficiencies of 103% and 98%, for FAM and VIC respectively.

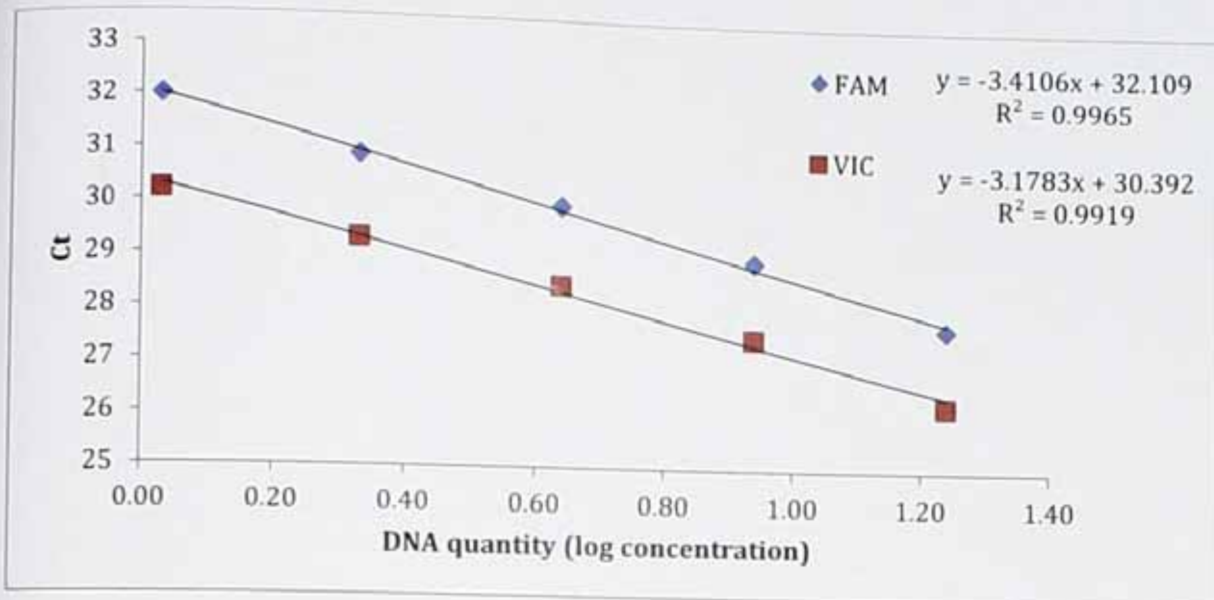


Figure V.7 *MMP3* Hs02276714_cn

Standard curve for the *MMP3* Hs02276714_cn (FAM) and Reference Assay RNase P (VIC). A linear trend line was fit to the data to determine the PCR efficiency for each assay, plotting cycle threshold (Ct) against DNA concentration (log concentration). Efficiency (E) was calculated using the equation $E = 10^{(-1/m)}-1$, where m is the slope of the linear trend line. Assays gave efficiencies of 96% and 106%, for FAM and VIC respectively.

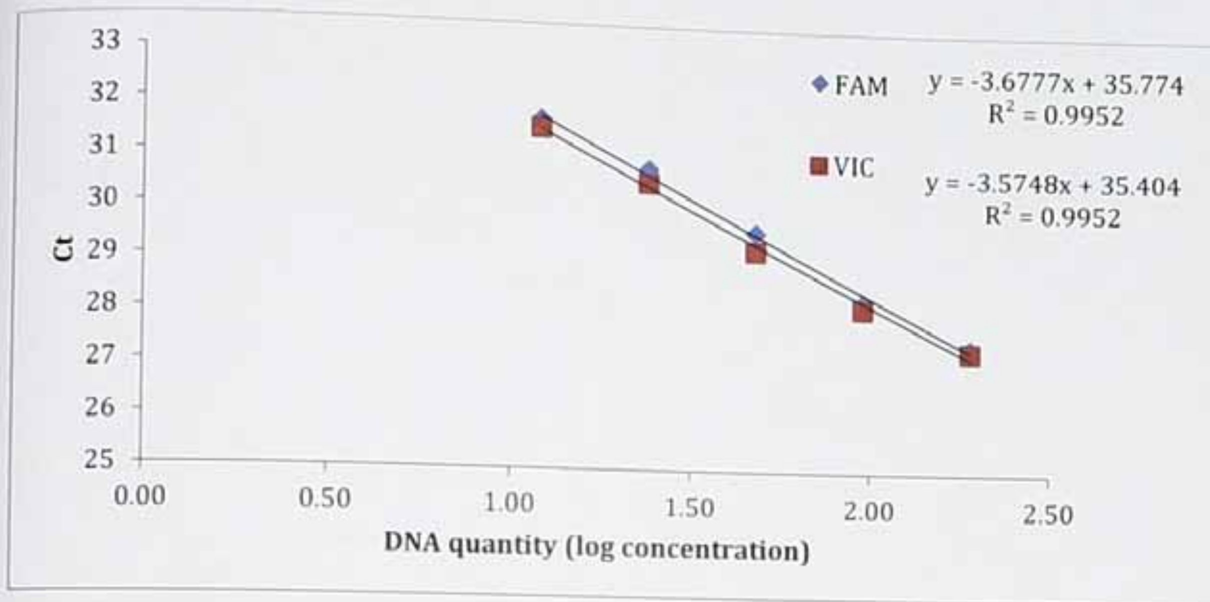


Figure V.8 *MMP23B* Hs07485808_cn

Standard curve for the *MMP23B* Hs07485808_cn (FAM) and Reference Assay RNase P (VIC). A linear trend line was fit to the data to determine the PCR efficiency for each assay, plotting cycle threshold (Ct) against DNA concentration (log concentration). Efficiency (E) was calculated using the equation $E = 10^{(-1/m)} - 1$, where m is the slope of the linear trend line. Assays gave efficiencies of 87% and 90%, for FAM and VIC respectively.

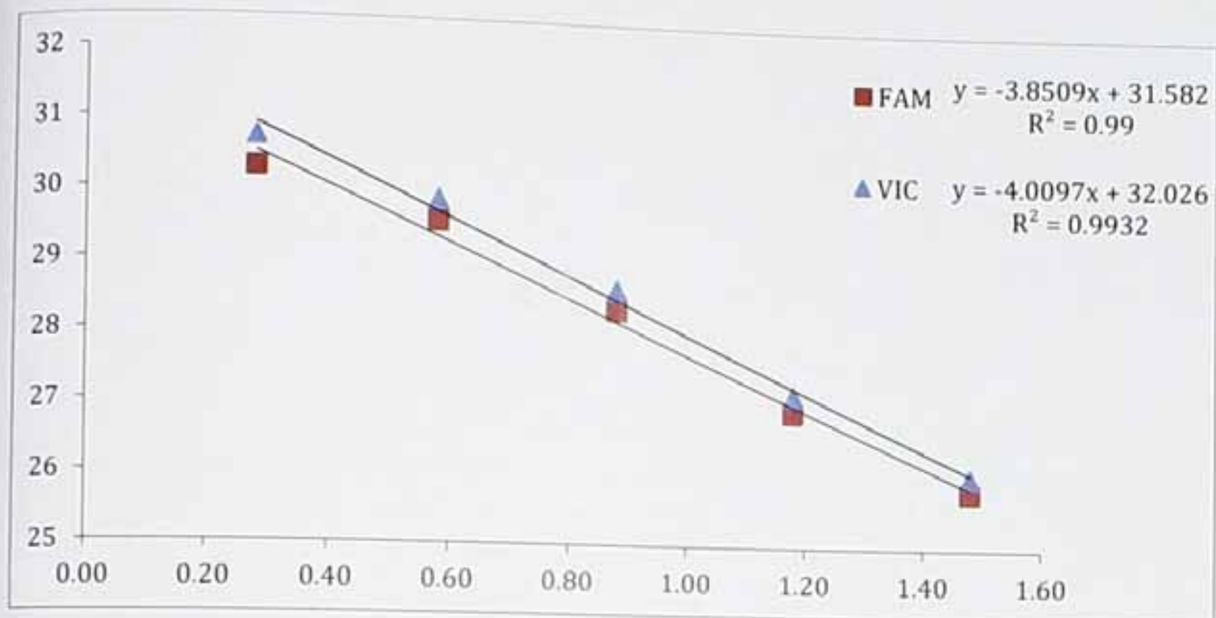


Figure V.9 *TNC* Hs00634176_cn

Standard curve for the *TNC* Hs00634176_cn (FAM) and Reference Assay RNase P (VIC). A linear trend line was fit to the data to determine the PCR efficiency for each assay, plotting cycle threshold (Ct) against DNA concentration (log concentration). Efficiency (E) was calculated using the equation $E = 10^{(-1/m)}-1$, where m is the slope of the linear trend line. Assays gave efficiencies of 82% and 78%, for FAM and VIC respectively.

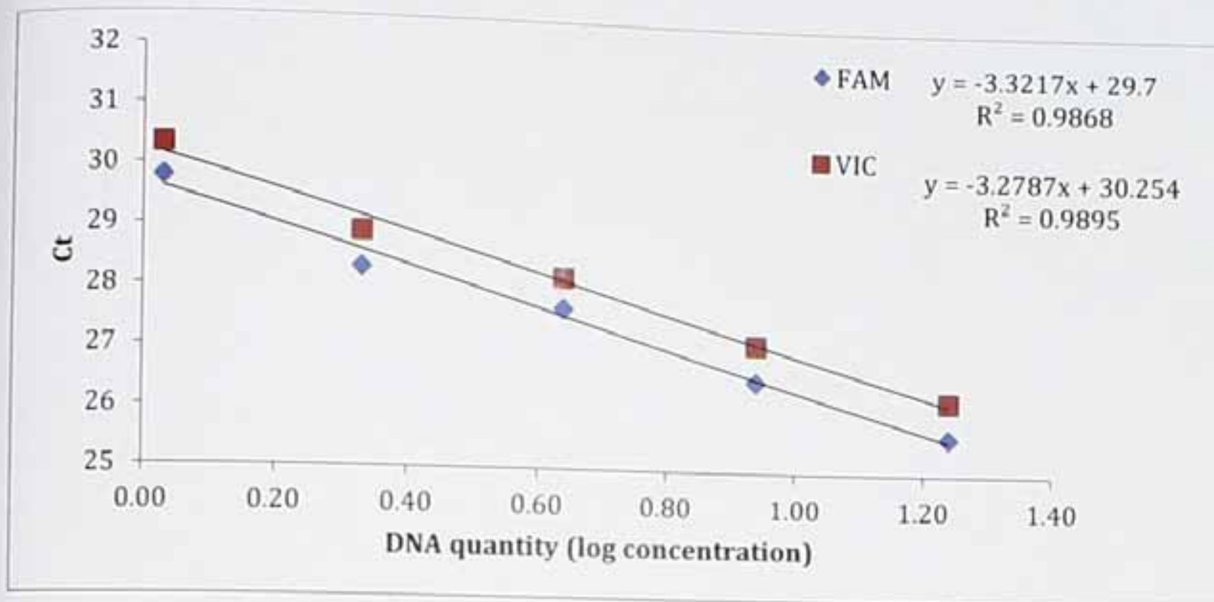


Figure V.10 *TNC* Hs06903309_cn

Standard curve for the *TNC* Hs06903309_cn (FAM) and Reference Assay RNase P (VIC). A linear trend line was fit to the data to determine the PCR efficiency for each assay, plotting cycle threshold (Ct) against DNA concentration (log concentration). Efficiency (E) was calculated using the equation $E = 10^{(-1/m)}-1$, where m is the slope of the linear trend line. Assays gave efficiencies of 100% and 102%, for FAM and VIC respectively.

APPENDIX VI

Publications

THE UNIVERSITY OF NORTHAMPTON | Rosetrees Trust — Helping humanity through medical research | UNIVERSITY OF CAPE TOWN — IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD | MRC

Copy Number Variation in the *COL5A1* Gene and the Risk of Achilles Tendon Pathology

Rebecca Rickaby[1], Michael Posthumus[2], Malcolm Collins[2,3], Christopher J Handley[4], Jill Cook[5] and Stuart M Raleigh[1]

[1]The Division of Health and Life Sciences, University of Northampton, Northampton, UK, [2]Department of Human Biology, UCT/ MRC Research Unit for Exercise Science and Sports Medicine, Faculty of Health Sciences, University of Cape Town, [3]South African Medical Research Council, Cape Town, South Africa, [4]School of Human Biosciences and the Musculoskeletal Research Centre, La Trobe University, Melbourne, Victoria, Australia and [5]Centre of Physical Activity and Nutritional Research, School of Exercise and Nutrition Sciences, Deakin University

INTRODUCTION

Achilles tendon Pathology is a multifactorial condition with numerous identified risk factors [1]. Risk factors include training errors, biological age previous injury and genetic risk factors [2]. Despite being one of the strongest tendons in the human body, the Achilles tendon is prone to damage and rupture [3]. Furthermore, due to the increased stress on tendons during physical exercise, tendon injuries are more common in both professional and recreational athletes [4]. Research to date has also shown associations with single nucleotide polymorphisms (SNPs) within genes encoding proteins with functional roles in tendon biology [2]. In 2005, a guanine-thymine (GT) repeat polymorphism in the *TNC* gene was found to associate with ATP [5]. Later, SNPs within the candidate genes *COL5A1*, *MMP3* and *GDF-5* were shown to associate with ATP [2]. However, predisposition to certain diseases may also be associated with copy number variation (CNV) [6]. Copy number variants (CNVs) are segments of DNA greater than 1kb in size, which show altered copy number when compared to a reference genome [7]. These structural changes can occur in different forms, including insertions, duplications and deletions [7]. CNV has yet to be investigated in relation to ATP, which is the aim of this study.

METHODS

Recruitment

A Caucasian population consisting of a South African (SA) cohort (114 clinically diagnosed cases of ATP and 96 controls) and an Australian (AUS) cohort (59 clinically diagnosed cases of ATP and 152 controls) were recruited for this case-control genetic association study. The Research Ethics Committees of the Faculty of Health Sciences at the University of Cape Town, South Africa, La Trobe University, Australia, Deakin University, Australia and the University of Northampton, United Kingdom approved this study. For the SA cohort, DNA was extracted from whole blood using the procedure described by Lahiri and Nurnberg [8] and modified by Mokone et. al. [5]. For the Australian cohort, DNA was extracted from whole blood using a sequence extraction technique (Flexigene DNA kit, Qiagen P/L, Valencia, California, USA) as per the manufacturer's recommendations.

CNV selection and PCR

Catalogued CNVs were identified using the Database of Genomic Variants (DGV) (http://projects.tcag.ca/variation/) hosted by The Centre for Applied Genomics, Toronto. Applied Biosystems™ fluorescence-based pre-designed TaqMan® Copy Number Assays (Applied Biosystems™, Foster City, California, USA) were selected in the candidate gene *COL5A1* using TaqMan® Copy Number Assay Search tool (Applied Biosystems™) (https://bioinfo.appliedbiosystems.com/genome-database/copy-number-variation.html). See Figure 1. PCR was performed using a StepOnePlus™ Real-Time PCR System (Applied Biosystems™). Each sample was run in quadruplicate, containing TaqMan® Copy Number Assay and Reference Assay RNase P, alongside negative controls. Copy numbers were calculated using CopyCaller® Software version 2.0 (Applied Biosystems™). See Figure 2. Results were accepted with calling confidence >80%.

Statistical analyses

Data were analysed using SPSS version 17.0 (SPSS Science Inc, Chicago, IL). A chi-squared (χ^2) analysis was used to determine any significant differences in copy number frequencies in cases and controls ($P<0.05$).

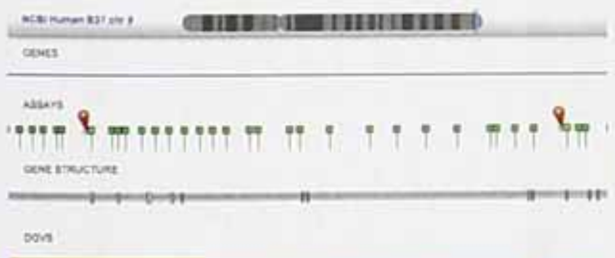


Figure 1. Hs00180523_cn and Hs02980647 TaqMan® Copy Number Assays identified using TaqMan® Copy Number Assay Search tool. Assays selected within exons of the *COL5A1* gene in regions with reported CNV from the DGV.

RESULTS

Table 1. General characteristics of the SA and AUS controls (CON) and ATP cases (ATP).

	SA CON (n=96)	SA ATP (n=114)	AUS CON (n=152)	AUS ATP (n=59)
Age (years)	37.1 ± 10.0 (91)	40.2 ± 12.3 (107)	38.5 ± 11.9 (149)	40.3 ± 14.1 (58)
Gender (% male)	66.3 (95)	73.0 (111)	39.7 (151)	67.8 (59)
Weight (kg)	175.4 ± 9.7 (93)	175.9 ± 8.7 (103)	171.7 ± 9.3 (150)	173.8 ± 9.5 (57)
Height (cm)	72.0 ± 12.3 (95)	80.8 ± 14.9 (106)	73.3 ± 13.9 (151)	80.4 ± 15.0 (59)
BMI (kg/m²)	23.3 ± 2.8 (93)	26.0 ± 3.9 (103)	24.8 ± 4.0 (150)	26.6 ± 4.1 (57)

Values are means ± SD or frequency (%). CON age is age of recruitments, ATP age is age of initial injury.

Figure 2. Typical profile of copy number estimation in CopyCaller® Software version 2.0 (Applied Biosystems™) in South African controls and ATP cases.

Table 2. *COL5A1* Copy number (CN) frequencies in SA and AUS controls and ATP cases

Hs00180523_cn		<2 CN	=2 CN	>2 CN	N	P value
SA	CON	1	90	0	91	-
	ATP	0	108	1	109	0.362
AUS	CON	2	81	3	86	-
	ATP	0	50	0	50	0.221
SA & AUS	CON	3	171	3	177	-
	ATP	0	158	1	159	0.169

DISCUSSION

This is the first study to investigate the relationship between CNV in the *COL5A1* gene and the risk of ATP. Results indicate no association between CN in ATP cases or controls. There was no significant difference in CN between CON and ATP, with *P* values of 0.362 and 0.221 in SA and AUS respectively. The combined Causasian population also showed no significant difference in CN in CON and ATP ($P = 0.169$). Whilst the *COL5A1* gene has known SNP and CNV regions, studies to date have only shown associations with SNPs and the risk of ATP [2]. The investigation of CNV and the risk for ATP is a new area of research. Further advances may reveal associations in the future, aiding the understanding of the genetic basis to ATP. Our studies are now extending to a second CNV region within the *COL5A1* gene (Hs02980647_cn). See Figure 1.

ACKNOWLEDGEMENTS

The authors would like to thank Dr G.G. Mokone for assistance with the recruitment of the SA cohort and Professor M.P. Schwellnus for the clinical diagnoses of the SA participants. The authors are grateful to Louis El Khoury for advice, ideas and technical assistance and Professor W.J. Ribbans for his expert insights into the clinical aspects of tendinopathy. The work was supported by the Rosetrees Trust, The University of Northampton, the National Research Foundation (NRF) of South Africa (FA2005021700015 and FA2007032700010 to M.C.), the University of Cape Town and the South African Medical Research Council (MRC).

REFERENCES

1. Meeuwisse, W.H. (1994) Assessing causation in sport injury: A multifactorial model. Clinical Journal of Sport Medicine, 4 (1), pp.166-70.
2. Raleigh, S.M. & Collins, M. (2012) Gene variants that Predispose to Achilles Tendon Injuries : An Update on Recent Advances. In: Cretnik, A. (ed.) Achilles Tendon. InTech, pp.27-40.
3. Leeni, A. & Rundesteinn, M. (2004) Disorders of the Achilles tendon. Current Orthopaedics, 18 (1), pp.63-75.
4. Maffulli, N., Wong, J. & Almekinders, L.C. (2003) Types and epidemiology of tendinopathy. Clinics in Sports Medicine, 22 (4), pp.675-92.
5. Mokone, G.G., Gajjar, M., September, A.V., Schwellnus, M.P., Greenberg, J., Noakes, T.D. & Collins, M. (2005) The guanine-thymine dinucleotide repeat polymorphism within the tenascin-C gene is associated with achilles tendon injuries. The American Journal of Sports Medicine, 33 (7), pp.1016-21.
6. Estivill, X. & Armengol, L. (2007) Copy number variants and common disorders: filling the gaps and exploring complexity in genome-wide association studies. PLoS Genetics, 3 (10), pp.1787-99.
7. Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W., Cho, E.K., Dallaire, S., Freeman, J.L., Gonzalez, J.R., Gratacos, M., Huang, J., Kalaitzopoulos, D., Komura, D., MacDonald, J.R., Marshall, C.R., Mei, R., Montgomery, L., Nishimura, K., Okamura, K., Shen, F., Somerville, M.J., Tchinda, J., Valsesia, A., Woodwark, C., Yang, F., Zhang, Junjun, Zerjal, T., Zhang, Jane, Armengol, Lluis, Conrad, D.F., Estivill, X., Tyler-Smith, C., Carter, N.P., Aburatani, H., Lee, C., Jones, K.W., Scherer, S.W. & Hurles, M.E. (2006) Global variation in copy number in the human genome. Nature, 444 (7118), pp.444-54.
8. Lahiri, D. & Nurnberger, J. (1991) A Rapid Nonenzymatic Method for the Preparation of HMW DNA from Blood for RFLP Studies. Nucleic Acids Research, 19 pp.5444-5444.

LA TROBE UNIVERSITY AUSTRALIA | DEAKIN UNIVERSITY AUSTRALIA

THE UNIVERSITY OF NORTHAMPTON

UNIFYING SPORT SCIENCE

Rosetrees Trust
Helping humanity through medical research

GENOMIC COPY NUMBER VARIATION WITHIN THE *TNC* AND *ADAM8* GENES AS POTENTIAL RISK FACTORS FOR ACHILLES TENDON PATHOLOGY

Rebecca Rickaby[1], Malcolm Collins[2,3], Michael Posthumus[2] and Stuart M Raleigh[1]

[1] Division of Health and Life Sciences, University of Northampton, Northampton, UK, [2] Department of Human Biology, UCT/MRC Research Unit for Exercise Science and Sport Medicine, Faculty of Health Sciences, University of Cape Town, [3] South African Medical Research Council, Cape Town, South Africa.

Introduction

Achilles tendon pathology (ATP) is a multifactorial disorder prevalent among sporting populations [1,2]. A variety of risk factors are known to predispose to ATP and recent work has identified that genetics has a prominent role [3]. A guanine-thymine (GT) dinucleotide repeat polymorphism within the *TNC* gene has been associated the risk of ATP [4]. Research has shown associations with single nucleotide polymorphisms (SNPs) within genes encoding proteins with functional roles in tendon biology [3]. SNPs within the candidate genes *COL5A1*, *MMP3* and *GDF-5* have been shown to associate with ATP [3]. Furthermore, a higher level of *ADAM8* mRNA expression has been observed in ruptures compared with normal tendons [5]. Predisposition to certain diseases may also be associated with copy number variation (CNV) [6]. Copy number variants (CNVs) are segments of DNA greater than 1kb in size, which show altered copy number when compared to a reference genome [7]. These structural changes can occur in different forms, including insertions, duplications and deletions [7]. To date, the influence of genomic CNV as a predisposing factor for ATP has not been investigated.

Methods

We investigated whether genomic CNV within the *TNC* and *ADAM8* genes were additional risk factors for ATP in a South African Caucasian population consisting of 113 cases and 94 controls. The Research Ethics Committees of the Faculty of Health Sciences at the University of Cape Town, South Africa and the University of Northampton, United Kingdom approved this study. DNA was extracted from whole blood using the procedure described by Lahiri and Nurnberg [8] and modified by Mokone et. al. [4]. TaqMan® copy number probes to intron 15 (Hs06903309), the intron 23- exon 23 boundary (Hs00634176) of the *TNC* gene and exon 23 of the *ADAM8* (Hs02429708) gene were selected (Applied Biosystems™, Foster City, California, USA) for investigation (Fig 1.). Quantitative PCR was performed using a StepOnePlus™ Real-Time PCR System (Applied Biosystems™). Each sample was run as a multiplex PCR and contained a TaqMan® copy number assay along with a reference *RNase P* probe (each diploid genome was assumed to contain two copies of the *RNase P* gene) for normalisation. Copy numbers were calculated using CopyCaller® Software (version 2.0, Applied Biosystems™), see figures 2 and 3 respectively. Results were only accepted when the call confidence exceeded 80% and data were analysed using SPSS version 20 (SPSS Science Inc, Chicago, IL). Significant differences ($P<0.05$) in mean copy number (CN) between cases (ATP) and controls (CON) were analysed using *t*-tests. Analysis of variance (ANOVA) was performed to evaluate the potential effect of anthropometrical variables (age, height, weight and BMI) on genomic CNV. Data are shown in Table 1.

UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

Figure 1. Structure of a) *TNC* and b) *ADAM8* genes showing exons (vertical boxes) and intronic sequences (horizontal bars). Approximate locations of the CNV probes used are shown with arrows.

Results

Figure 2. Representative amplification plot for the quantitative determination of copy number (CN). CN calls were generated automatically based on the Ct value of the probe, in this case, *TNC* (FAM-labelled) relative to the internal control *RNase P* (VIC-labelled) using a StepOne plus™ real time PCR machine (Applied Biosystems™). Differences between the PCR amplification efficiencies of the *ADAM8* (Hs02429708) and *TNC* (Hs06903309 and Hs00634176) probes with respect to the *RNase P* gene probe were 1.6%, 6.3% and 4.0% respectively.

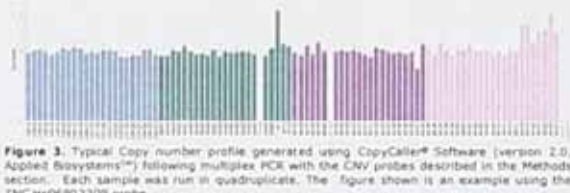


Figure 3. Typical Copy number profile generated using CopyCaller® Software (version 2.0, Applied Biosystems™) following multiplex PCR with the CNV probes described in the Methods section. Each sample was run in quadruplicate. The figure shown is an example using the *TNC* Hs06903309 probe.

Table 1. Mean CN in SA cases and controls

Gene (Assay ID)		Mean CN	N	*P* value
TNC (Hs06903309)	CON	2.03 ± 0.20	88	-
	ATP	2.02 ± 0.21	103	0.953
TNC (Hs00634176)	CON	1.98 ± 0.10	89	-
	ATP	1.98 ± 0.11	107	0.887
ADAM8 (Hs02429708)	CON	2.01 ± 0.34	89	-
	ATP	1.98 ± 0.14	107	0.346

Conclusions

This is the first study to investigate the relationship between gene copy number variation as a potential risk factor for ATP. For *TNC* we found no significant difference in mean CN between ATP cases and controls for either probe tested. Similarly for *ADAM8*, we found no significant difference in CN between ATP cases and controls. In summary, our preliminary findings suggest that CNV at the three loci investigated does not predispose to ATP. We are currently investigating additional CNV regions in additional candidates.

Acknowledgements

The authors would like to thank Dr G.G. Mokone, Professor M.P. Schwellnus, Professor W.J. Ribbans and Louis El Khoury. This research was supported by the Rosetrees Trust, The University of Northampton, the National Research Foundation (NRF) of South Africa, the University of Cape Town and the South African Medical Research Council (MRC).

References

[illegible]

THE UNIVERSITY OF NORTHAMPTON

FEBS EMBO
Paris 2014 hosted by SFBBM
30 August - 4 September

Variation within the *CASP3* gene and the risk of Achilles tendinopathy in a British case-control cohort

Rebecca Rickaby[1], Louis El Khoury[1], William J Ribbans[1] and Stuart M Raleigh[1]
[1]*The Centre for Physical Activity and Chronic Disease, The Institute of Health and Wellbeing, University of Northampton, Northampton, UK*

1. Introduction

Achilles tendon pathology (ATP) is a degenerative condition with known genetic risk factors[1]. Excessive tenocyte apoptosis has been observed in tendinopathy and components of the apoptosis pathway have previously been implicated in the aetiology of ATP[2]. Caspases are a large family of cysteine proteases that play a key role in the execution and regulation of apoptosis[3]. Caspase-3 is one of three known effector caspases, which can selectively cleave target proteins, such as Bcl-2, after aspartate residues in their primary sequence[3].

The rs1049253 single nucleotide polymorphism (SNP) lies within a microRNA (miRNA) binding site in the 3' untranslated region (UTR) of *CASP3*[4] (**Figure 1**). This SNP is associated with the risk of certain cancers and the CC genotype is associated with lower levels of *CASP3* mRNA[4]. MiRNAs can bind mRNAs and have an important role in regulating apoptosis associated with carcinogenesis[4]. Our aim was to determine whether *CASP3* rs1049253 was associated with ATP in a British cohort.

Figure 1. Schematic image of Chromosome 4 showing the location, size and exonic structure of the *CASP3* gene, along with the location of *CASP3* rs1049253 C/T in the 3'-UTR of the gene (Chr 4:185548951).

2. Methods

We recruited 261 (130 ATP cases and 131 asymptomatic controls) British Caucasian participants for this genetic association study. ATP cases were clinically diagnosed with insertional tendinopathy (INS), noninsertional tendinopathy (NON), Achilles tendon rupture or mixed pathology. Written informed consent was obtained and all participants completed a physical activity/medical history/injury questionnaire. The study was approved by the Research Ethics Committee of the University of Northampton, UK. DNA was extracted from 2 mL of saliva collected using ORAGENE·DNA collection kits (OG-500) and DNA purification was carried out using the prepIT·L2P DNA extraction kit (DNA Genotek Inc., Ontario, Canada).

TaqMan assay technology was used to genotype all participants using real-time PCR, with 10 ng of DNA and positive and negative controls included in each PCR run. The TaqMan Genotyping Assay contained FAM and VIC reporter dye labelled probes, with ROX dye as the passive reference. Genotypes were automatically called using StepOne Software, version 2.1 (Applied Biosystems, Foster City, California, USA) (**Figure 2**).

Data were analysed using IBM SPSS Statistics, version 20 (IBM Corp. Armonk, NY). A Pearson's chi-squared (X^2) or Fisher's exact test was used to analyse differences in genotype and allele frequencies for the rs1049253 variant. We compared the collective ATP group against controls. We also conducted sub-analyses for the different types of tendinopathy. Rupture and mixed pathology cases were excluded from sub-analyses. Hardy-Weinberg equilibrium (HWE) was established and $p < 0.05$ was considered to be a deviation from HWE.

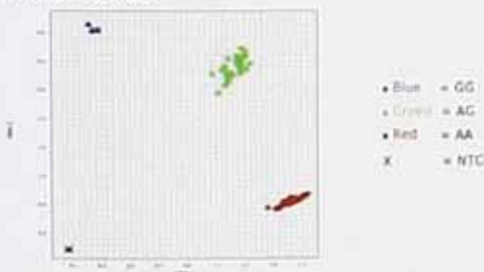


Figure 2. Typical allelic discrimination plot for *CASP3* rs1049253.

3. Results

Table 1a. Genotype and allele frequency distribution of *CASP3* rs1049253 within cases (ATP/INS/NON) and controls (CON).

***CASP3* rs1049253**	**CON** n=131	**ATP** n=130	**INS** n=29	**NON** n=47
TT	62.6 (82)	61.5 (80)	72.4 (21)	55.3 (26)
CT	34.4 (45)	33.1 (43)	20.7 (6)	38.3 (18)
CC	3.1 (4)	5.4 (7)	6.9 (2)	6.4 (3)
p-value		0.643	0.218	0.456
MAF	20.2 (53)	21.9 (57)	17.2 (10)	25.5 (24)
p-value		0.635	0.605	0.284
HWE	0.461	0.700	0.139	0.961

Table 1b. Genotype and allele frequency distribution of *CASP3* rs1049253 within male cases (ATP/INS/NON) and male controls (CON).

***CASP3* rs1049253**	**Male CON** n=82	**Male ATP** n=80	**Male INS** n=15	**Male NON** n=27
TT	68.3 (56)	58.8 (47)	73.3 (11)	55.6 (15)
CT	30.5 (25)	32.5 (26)	13.3 (2)	33.3 (9)
CC	1.2 (1)	8.8 (7)	13.3 (2)	11.1 (3)
p-value		0.072	**0.036**	0.064
MAF	16.5 (27)	25.0 (40)	20.0 (6)	27.8 (15)
p-value		0.058	0.976	0.219
HWE	0.326	0.233	**0.024**	0.379

Values are expressed as a frequency (%) with number of participants (n) in parenthesis. Minor allele frequency (MAF) and Hardy-Weinberg equilibrium (HWE) are shown.

4. Discussion

We found no significant difference in genotype ($p = 0.643$) or allele ($p = 0.635$) frequencies between the ATP group and controls (**Table 1a**). However, we did find a genotypic association ($p = 0.036$) between male insertional tendinopathy cases (INS) and male controls (CON) (**Table 1b**). Specifically, the CC genotype appeared to increase the risk of insertional tendinopathy in males ($p = 0.045$, OR = 12.46, 95% CI 1.05-147.46). Furthermore, *CASP3* rs1049253 was not in HWE in the male INS group ($p = 0.024$). There were no significant differences between male ATP/NON and CON, nor with female ATP/INS/NON and female CON (data not shown). It it important that these data are viewed with caution due to the relatively small sample size, which might also explain the deviation from HWE observed in the male INS group. Therefore replication in a larger cohort would be necessary to increase confidence.

Our preliminary data infer a possible role for the rs1049253 variant as a risk factor for insertional tendinopathy in British males. Although additional research is needed, these results could further implicate the involvement of the apoptosis pathway in the development of ATP and could justify the inclusion of this variant in a risk assessment model for ATP. Such models alone will not predict or diagnose ATP, due to the multifactorial nature of this condition. Nevertheless, in combination with the alteration of modifiable risk factors such as training, could hold potential in lowering the risk of injury[5].

5. Acknowledgements

This research was supported by a grant to Dr Stuart M. Raleigh from the Rosetrees Trust and The University of Northampton. The authors would like to thank Professor Malcolm Collins and Dr Mike Posthumus at the University of Cape Town's UCT/MRC Research unit for Exercise Science and Sport Medicine for discussions concerning the questionnaire design. Authors S.M.R. and W.J.R. have filed patents on the application of specific sequence variations related to risk assessment of Achilles tendinopathy and anterior cruciate ligament injuries. Attendance at FEBS EMBP Paris 2014 was supported by a general travel grant from the Biochemical Society to R.R.

6. References

1. Collins M, Raleigh SM. 2009. Genetic risk factors for musculoskeletal soft tissue injuries. Medicine and Sport Science. 54:136-149.
2. Nell EM, van der Merwe L, Cook J, et al. 2012. The apoptosis pathway and the genetic predisposition to Achilles tendinopathy. Journal of Orthopaedic Research. 30:1719-1724.
3. Hengartner MO. 2000. The biochemistry of apoptosis. Nature. 407:770-776.
4. Guan X, Liu Z, Liu H, et al. 2013. A functional variant at the miR-885-5p binding site of CASP3 confers risk of both index and second primary malignancies in patients with head and neck cancer. FASEB Journal. 27:1404-1412.
5. September AV, Posthumus M, Collins M. 2012. Application of genomics in the prevention, treatment and management of Achilles tendinopathy and anterior cruciate ligament ruptures. Recent Patents on DNA & Gene Sequences. 6:216-223.

Rosetrees Trust
Helping humanity through medical research

Biochemical Society
Advancing Molecular Bioscience

THE UNIVERSITY OF NORTHAMPTON

Institute of Health and Wellbeing

Rosetrees Trust
Supporting the best in medical research

DNA METHYLATION OF THE *ADAMTS4* GENE IN JUMPER'S KNEE

Rebecca Rickaby[a], Tom Samiric[b] and Stuart M Raleigh[a]
[a]The Centre for Physical Activity and Chronic Disease, The Institute of Health and Wellbeing, University of Northampton, Northampton, UK
[b]Department of Physiology, Anatomy and Microbiology, La Trobe University, Melbourne, Australia

Introduction

Jumper's knee (patellar tendinopathy) is a common injury in sporting populations, with prevalence rates higher than 30% in both basketball and volleyball players (Christian *et al.*, 2014).

ADAMTS4 (a disintegrin and metalloproteinase with thrombospondin motifs 4) is an enzyme involved in the homeostasis of the tendon environment (extracellular matrix) (Corps *et al.*, 2008).

Figure 1. Rafael Nadal has suffered from patellar tendinopathy in both knees (http://desktophdwallpaperss.com/rafael-nadal-new-desktop-wallpapers-and-hd-photos-download-free/)

Figure 2. Image of right knee showing the patellar tendon and associated knee ligaments (http://orthotape.com/Patella_Knee_support_band_strap.asp)

DNA methylation is one way in which the genome (entire genetic material of an organism) is regulated, so that different cell types can be produced with specific functions (Auclair & Weber, 2012). Loss of DNA methylation and altered expression of the *ADAMTS4* gene has been observed in osteoarthritis (Roach *et al.* 2005), identifying it as an interesting candidate to research in musculoskeletal conditions.

Aims & Method

The aim of this study was to identify whether DNA methylation within the *ADAMTS4* gene was altered in tissue from patellar tendinopathy (PT) compared to healthy control tissue (CON). To do this, part of the gene promoter was sequenced using a method called Pyrosequencing.

Twenty (10 PT and 10 CON) tissue samples from the patellar tendon (Figure 2) of male participants were collected for this study. Differences in DNA methylation between PT and CON were analysed using a *t*-test with significance at $p<0.05$. Ethics approval was gained from the relevant research ethics committees.

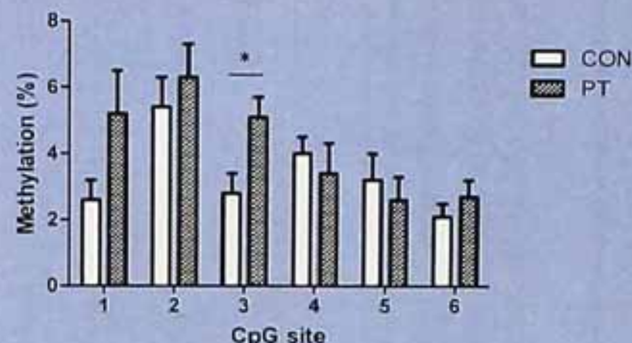


Figure 3. Comparison of the mean DNA methylation between CON and PT at 6 CpG sites within the promoter of the *ADAMTS4* gene. Bars represent mean (%) ± standard error of the mean (SEM). CON, white bars and PT, shaded bars.*, $p = 0.016$.

Results & Discussion

We found a significant difference ($p = 0.016$) in DNA methylation between the CON and PT group at CpG site 3, where mean DNA methylation was significantly higher in the PT group (Figure 3).

Biologically, altered DNA methylation of genes can result in a change in gene expression (Auclair & Weber, 2012). In relation to tendon injury, this could potentially disrupt the tendon environment and increase the risk of injury. Indeed, higher expression levels of *ADAMTS4* have previously been detected in ruptured Achilles tendon compared with normal and tendinopathic tendon (Corps *et al.*, 2008).

This is a new and exciting area of genetics which we are continuing to investigate in other genes. It further highlights the complex nature of tendon injuries and the need for more research in this field.

References

Christian, R.A., Rossy, W.H. & Sherman, O.H. (2014) Patellar tendinopathy: Recent developments toward treatment. *Bulletin of the Hospital for Joint Disease*, **72**(3), 217-224.

Corps, A.N., Jones, G.C., Harrall, R.L., Curry, V.A., Hazelman, B.L. & Riley, G.P. (2008) The regulation of aggrecanase ADAMTS-4 expression in human Achilles tendon and tendon-derived cells. *Matrix Biology*, **27**, 393-401.

Auclair, G. & Weber, M. (2012) Mechanisms of DNA methylation and demethylation in mammals. *Biochimie*, **94**, 2202-2211.

Roach, H.I., Yamada, N., Cheung, K.S.C., Tilley, S., Clarke, N.M.P, Oreffo, R.O.C., Kokubun, S. & Bronner, F. (2005) Association between the abnormal expression of matrix-degrading enzymes by human osteoarthritic chondrocytes and demethylation of specific CpG sites in the promoter regions. *Arthritis & Rheumatism*, **52**(10), 3110-3124.

Article redacted:

Rickaby, R., El Khoury, L., Ribbans, W.J. and Raleigh, S.M., 2015. Variation within three apoptosis associated genes as potential risk factors for Achilles tendinopathy in a British based case–control cohort. Gene, 571(2), pp.167-171.
https://doi.org/10.1016/j.gene.2015.06.010