UNIVERSITY OF CAPE TOWN

The regulation of the COL5A1 gene via the 3'-UTR

and its impact on Achilles tendinopathy and other exercise-related phenotypes

Yoonus Abrahams

7/15/2013

Master of Science in Medicine in Cell biology

Supervisors:

Prof. Malcolm Collins
A/Prof. Sharon Prince

An investigation into the regulation of the *COL5A1* gene via its 3'-UTR and its impact on Achilles tendinopathy and other exercise-related phenotypes

By

Yoonus Abrahams

SUBMITTED TO THE UNIVERSITY OF CAPE TOWN
In fulfilment of the requirements for the degree

Master of Science in Medicine in Cell biology

Faculty of Health Science

UNIVERSITY OF CAPE TOWN
July 2013

Supervisors:
Prof. Malcolm Collins
A/Prof. Sharon Prince

Acknowledgements

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

I wish to express my gratitude to the following persons and organisations for their contributions to the completion of this dissertation:

My parents, Jogera and Moegsien Abrahams, for their love, support (both financial and emotional) and guidance.

A/Prof. Sharon Prince for her guidance, expertise, encouragement and quick smile.

Prof. Malcolm Collins for his flawless leadership and friendship.

Nancy Laguette, Kevin O'Connell and Erica-Mari Nell for their technical expertise, motivation and friendship.

Doctors Deeya Ballim and Amaal Abrahams for their advice, friendship and understanding.

Melanie Hay and Kim Stephenson for their commiseration, laughter and encouragement.

To my high school science teachers, Mr Du Plooy, Mr Triem and Ms Potgieter, for bestowing both their knowledge and love of science on me.

To Aretha Cooper, Shannagh Hare, Tarryn Wilmer and the rest of the T-Box family, both past and present, for their willingness to share their experience and passion.

To the SSISA/ESSM Staff and students, for their friendship and encouragement.

To the National Research Foundation, Medical Research Council, University of Cape Town and the Myers and Brunow Scholarship for their financial support.

Finally, to Carl Sagan, Bill Bryson and Richard Feynman for their inspiration, and for instilling the good that science can achieve.

Abstract

Introduction: The *COL5A1* gene encodes for the α1 chain of type V collagen, a minor fibrillar collagen protein which plays a vital role in fibril assembly in tendons, ligaments and other connective tissue. DNA sequence variants within the 3'-untranlated region (UTR) of this gene have previously been associated with the chronic form of the degenerative tendon disorder Achilles tendinopathy. In addition, two major functional forms of the 3'-UTR were previously identified. These distinct forms are determined by seven sequence variants, four of which are independently associated with Achilles tendinopathy. The "at risk" allelic form of the 3'-UTR (the "T-allelic" form) is hypothesised to give rise to a more stable mRNA when compared to the wild type allelic form (the "C-allelic" form). Since the *COL5A1* 3'-UTR contains several putative miRNA binding sites, two of which are polymorphic, we hypothesis that differential miRNA binding to the *COL5A1* 3'-UTR is at least in part responsible for the differences in mRNA stability between the T- and C-allelic forms. Two forms of the mature Hsa-miR-608, which are produced from the polymorphic (SNP rs4919510, C/G) *MIR608* gene, can potentially bind to one of these putative polymorphic miRNA binding sites. The aims of this dissertation therefore were (i) to determine whether the two polymorphic and a third putative miRNA binding sites within the 3'-UTR of the *COL5A1* gene were capable of repressing the *COL5A1* mRNA stability in vitro and to determine whether the two forms of the *COL5A1* 3'-UTR showed a significant difference in response to miRNA regulation and (ii) to determine whether the *MIR608* single nucleotide polymorphism rs4919510 is associated with chronic Achilles tendinopathy.

Methods: Plasmid constructs containing the C- or T-allelic forms of the *COL5A1* 3'-UTR downstream of a luciferase reporter gene were co-transfected with Hsa-miR-608 or Has-miR-125a-5p miRNA mimic or negative siRNA control into HT1080 fibrosarcoma or SVWI-38 SV40 transformed fibroblast cells. In the case of the second putative polymorphic miRNA binding site, termed the *Mbo*II octamer, where commercial miRNA mimic was unavailable, a site-directed mutagenesis approach was undertaken. The rs4919510 polymorphism in the *MIR608* gene was genotyped using genomic DNA from 166 Caucasian chronic Achilles tendinopathic participants (TEN) and 342 apparently healthy Caucasian control participants (CON) using fluorescence based Taqman® PCR.

Results: The rs4919510 variant within the *MIR608* gene is independently associated with chronic Achilles tendinopathy in a combined South African and Australian Caucasian population (p=0.023). Specifically, the CC genotype was significantly over-represented in the TEN (68.5%) compared to the CON (57.1%) participants. Functional analysis of luciferase activity of constructs containing the COL5A1 3'-UTR showed significant reduction when co-

transfected with Hsa-miR-608 mimic, with the polymorphic binding site within the T-allelic form of the *COL5A1* 3'-UTR showing greater repression in the HT1080 fibrosarcoma cell line (p=0.003) but not in the SV40 transformed SVWI-38 fibroblast cell line. Co-transfection with Hsa-miR-125a-5p mimic significantly reduced the luciferase activity of the *COL5A1* 3'-UTR constructs with the T-allelic form showing a significantly greater decrease in mRNA stability than the C-allelic form at 10 pmol (p=0.011) and 20 pmol (p=0.021) but not at the 1 pmol concentration (p=0.999) in HT1080 fibrosarcoma cells. The *Mbo*II octamer mutants show increased mRNA stability when compared to the parental constructs in both the HT1080 (CON Mut 2 (1.43 ±0.06 fold), Mut 3(1.19 ±0.05 fold) TEN Mut 1 (1.87 ±0.06 fold) Mut2 (1.67 ±0.06 fold) Mut 3 (1.77 ±0.03 fold)) and SVWI-38 (CON Mut 2 (1.65 ±0.04 fold), Mut 3(1.17 ±0.07 fold) TEN Mut 1 (1.66 ±0.05 fold) Mut2 (1.25 ±0.05 fold) Mut 3 (1.19 ±0.03 fold)) cell lines (p<0.001), indicating the presence of a negative regulator of the *COL5A1* gene. Finally, the two major functional forms of the 3'-UTR had distinct secondary structures when folded in silico using the Sfold algorithm. We hypothesise that this difference may have implications in the regulation of the *COL5A1* 3'-UTR by *cis*-elements.

Conclusion: This dissertation shows for the first time that there is an association of the CC genotype of the rs4919510 variant in the *MIR608* gene with chronic Achilles tendinopathy. Additionally, this study shows that all three investigated putative miRNA binding were potentially functional. It is however, unlikely that these three regulatory elements within the *COL5A1* 3'-UTR are responsible for the observed increased mRNA stability of the T-allelic form in the HT1080 fibrosarcoma cell line. These results have important implications for our understanding of the molecular basis of musculoskeletal soft tissue injuries and other exercise-related phenotypes. Additional research is however, required to identify and elucidate the mechanisms of these and other regulatory elements.

Declaration

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

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

Signature: ……

Date: …….

Contents

List of Figures

List of Tables

1. Introduction

A number of injuries affect the Achilles tendon and the surrounding tissues such as chronic tendinopathy, complete or partial rupture, bursitis and peritendinitis [1]. Although injuries to the Achilles tendon are common as a result of participation in physical activity and certain occupational activities [2], they also occur in sedentary individuals [3].

Achilles tendon injuries have been reported to account for between 6 and 18% of all tendon injuries [4], [5]. Forty-one percent of patients with chronic tendinopathy eventually present with pathology in the contralateral tendon [6]. It has been estimated that in the United States alone, 825 000 runners would present with chronic Achilles tendinopathy per year [7]. In addition, the prevalence of Achilles tendon injuries appears to be increasing with the annual incidence increasing from 18.2 per 100 000 individuals in 1984 to 37.3 in 1996 in the Danish population [8].

Achilles tendon injuries have been well characterized at the clinical level with both intrinsic and extrinsic risk factors having been described (Table 1.1) [9], [10], [11], [12]. The molecular mechanisms responsible for Achilles tendon injuries are however largely unknown and the known mechanisms will be briefly reviewed in section 1.4 of this introductory chapter. As summarized in Table 1.2, DNA sequence polymorphisms are emerging as important intrinsic risk factors for Achilles tendinopathy. Of particular interest to the current study is a common C/T single nucleotide polymorphism (SNP rs12722; also known as the *Bst*UI RFLP) in the 3'-untranslated region (3'-UTR) of *COL5A1*, which was shown to be associated with chronic Achilles tendinopathy [13], [14] and anterior cruciate ligament (ACL) injuries in females [15]. In all of these studies the CC genotype at rs12722 was shown to be over-represented in the respective asymptomatic controls. An age-related increase in range of motion measurements has also been reported in individuals with the CC genotype [16], [17].

Table 1.1 Intrinsic and extrinsic risk factor for Achilles tendinopathy

Extrinsic factors	Reference	Intrinsic factors	Reference
Occupation	[9]	Age	[9]
Footwear		Gender	
Training errors		Poor nutrition	
Running surface		Systemic disease	
Level of physical activity	[18]	**Genetics**	[13], [14]
Plantar flexor strength	[19]	Fluoroquinolone antibiotic treatment	[20]

The COL5A1 gene encodes for the α1(V) chain of type V collagen [21] and genetic association studies have found four variants within and around putative miRNA binding sites within the 3'-UTR of *COL5A1* which are independently associated with Achilles tendinopathy [14], [22]. As the 3'-UTR of eukaryotic genes has been shown to contain elements which are emerging as important post-transcriptional regulators [23], [24], the focus of this dissertation is to investigate the role the *COL5A1* 3'-UTR has in the regulation of the *COL5A1* gene. These elements, in particular microRNA binding sites, will be reviewed in section 1.6. A brief summary of the structure and function of tendons (sections 1.1 and 1.2), type V collagen and the *COL5A1* gene will follow in subsequent sections.

Table 1.2 Genetic polymorphisms associated with Achilles tendinopathy

Gene	Description	Polymorphism	Association (increased { ↓ } or decreased { ↑ } risk of AT)	Reference
TNC	Tenascin C : ECM glycoprotein	GT repeat, intron 17	13 and 17 repeats ↓ risk 12 and 14 repeats ↑ risk	[25]
COL5A1	**Fibrillar type V collagen pro-α1 subunit**	**rs12722 (T/C)**	**CC genotype ↓ risk**	[13], [14]
MMP-3	Matrix Metalloproteinase 3 : zinc-dependent endopeptidase	rs591058 (T/C)	CC genotype ↑ risk	[26]
		rs679620 (A/G)	GG genotype and G allele ↑ risk	
		rs650108 (G/A)	AA genotype ↑ risk	
TIMP-2	Tissue inhibitor of Metalloproteinase -2	rs4789932 (C/T)	CC genotype ↓ risk CT genotype ↑ risk	[27]
GDF5	Growth and Differentiation factor 5 :cytokine signalling molecule	rs143383 (T/C)	TT genotype ↑ risk	[28]
IL-1β	Interleukin 1 β: pro-inflammatory cytokine	rs16944 (T/C)	TT genotype modulates risk (↓)	[29]
		rs1143627 (T/C)	C allele modulates risk (↓)	
IL-1RN	Interleukin 1 receptor antagonist, competes for binding to the Interleukin 1 receptor with IL-1α and -β	rs2234663 (Variable Number Tandem Repeat)	genotypes not containing A2 allele modulates risk (↓)	[29]
IL-6	Interleukin 6: cytokine with both pro- and anti-inflammatory capabilities	rs1800795 (G/C)	GG genotype and G allele ↑ risk	[29]

1.1 Anatomy of the Achilles tendon

Healthy tendons are dense collagen-containing fibrous structures connecting muscles to bone which act to transmit energy from muscular contraction. Tendon morphology varies depending on the mechanical environment, with tendons exposed to complex loading such as the rotator cuff having diverse orientations of collagen fibres. In contrast, tendons experiencing mainly uniaxial tensile loading, such as flexor tendons and the Achilles tendon have close, parallel collagen fibre orientations [30], [31], [32].

The dry weight of a healthy tendon is composed of approximately 70 – 80% type I collagen with the remaining percentage composed of minor collagens (such as types III, V, XII and XIV), elastin, glycoproteins (such as fibrillin, tenascin C, fibronectin, laminin and thrombospondins) and proteoglycans (such as decorin, hyularan, biglycan and fibromodulin) and cells [33], [34]. Tendons contain relatively few cells with the resident fibroblasts, called tenocytes, regulating extracellular matrix (ECM) turnover [35]. The Achilles tendon is the largest and strongest tendon in the human body, capable of withstanding stress loads of up to 12.5 times the body weight while running. The Achilles tendon originates at both the soleus and gastrocnemius muscles and inserts at the calcaneus (Figure 1.1). It lacks a synovial membrane and is instead covered by the paratenon [5]. The Achilles tendon is composed largely of collagen fibrils, arranged in parallel fibres [12], [35], [36]. This parallel organisation of collagen fibrils, facilitated by proteoglycans and glycoproteins, allow the tendon to resist and transmit the tensile forces produced by the soleus and gastrocnemius muscles [35].

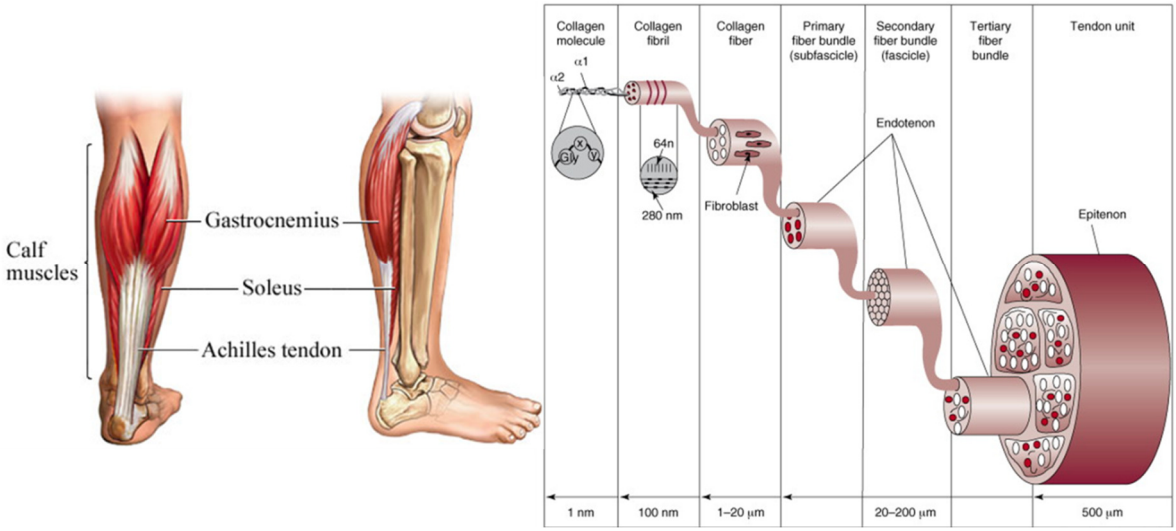


Figure 1.1 Diagram showing Achilles tendon anatomy (left) and collagen fibre hierarchy from collagen to tendon unit (right).
The Achilles tendon is connected to the gastrocnemius and soleus muscles and inserts at the calcaneus (left). The collagen fibrils are packaged into parallel bundles called collagen fibres. These fibres are grouped together and covered by the endotenon into fascicles. These groups of fascicles are bound by the epitenon into the functional tendon unit (Image adapted from Liu et al 2008)[37].

1.2 Collagens and the Collagen fibril

The term "collagen" collectively represents a group of 29 distinct structural macromolecules, all of which are formed as either hetero- or homotrimers consisting of three polypeptides known as α-chains. All collagen molecules have at least one triple helical protein domain which is stabilized by a repeating Gly-X-Y motif where approximately 30% of the X and Y positions are occupied by proline and hydroxyproline amino acid residues [38], [39]. These proline and hydroxyproline residues have their hydrophobic side chains facing outward of the helix while the glycine proton side chains face inwards to reduce steric pressure, thereby stabilizing the helix. The fibrillar collagens (such as types I; II and III) are transcribed as procollagens, with amino- and carboxy-terminal propeptides which flank an uninterrupted triple helical region (Figure 1.2). After processing within the endoplasmic reticulum and Golgi apparatus and secretion into the ECM, these terminal globular domains are cleaved by N- and C-procollagen proteinases, leaving short non-helical regions called telopeptides which flank the helix. Prior to secretion, the procollagen proteins undergo post-translational modification by lysyl oxidases to create aldehyde residues within the triple helical regions as well as their non-helical telopeptide terminal regions [38],

[39]. These modified residues then spontaneously form inter- and intramolecular cross-links with other lysine, hydroxylysine or aldehyde residues causing polymerisation of the individual collagen molecules which self-assemble into the collagen fibril [40].

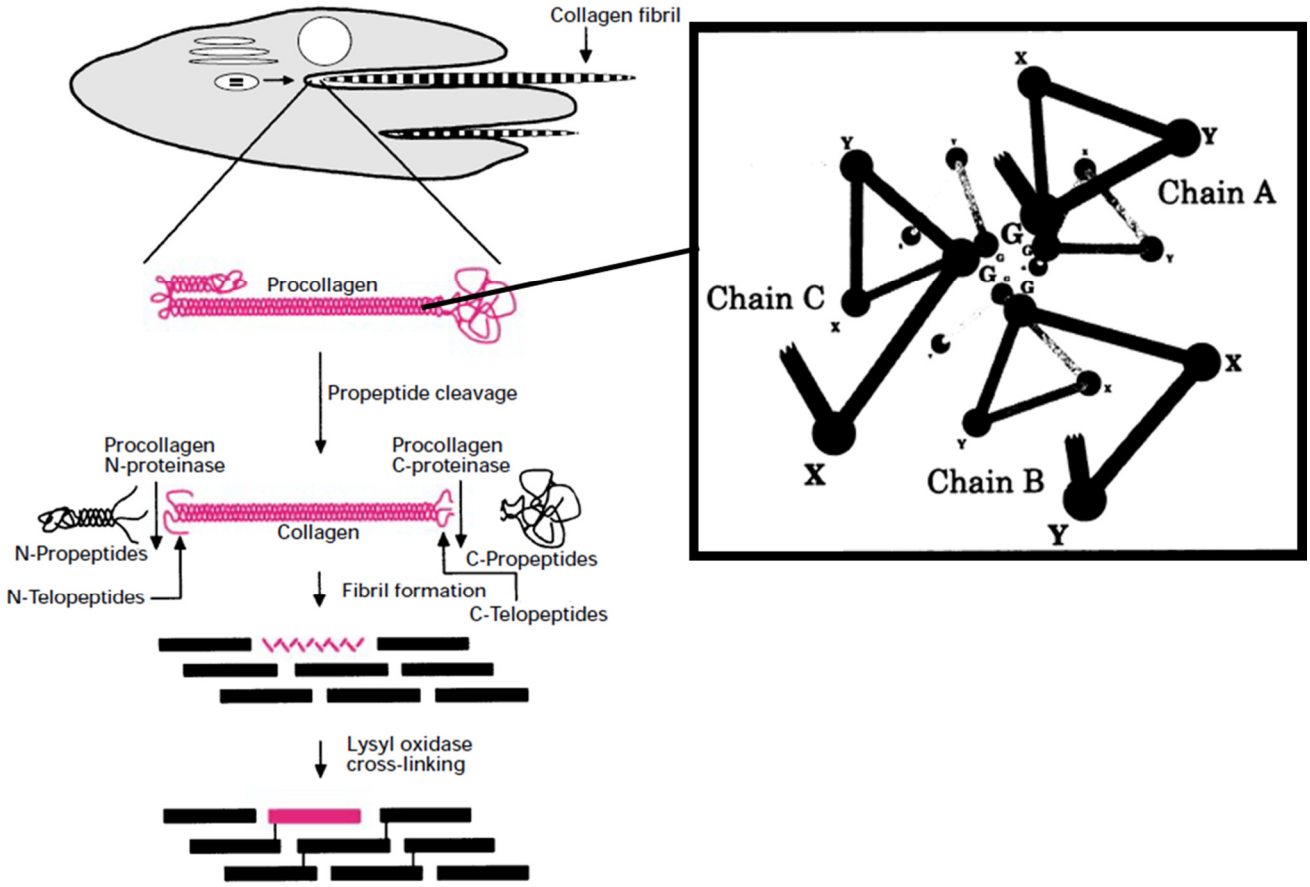


Figure 1.2 Schematic diagram showing collagen fibril synthesis by fibroblast cell.
Collagen peptide chains with the Gly-X-Y motif assemble to form soluble procollagen right-handed triple helices with N and C-terminal propeptides. N- and C-terminal propeptides are cleaved by N- and C-procollagen proteinases respectively and the collagen molecule is then modified by lysyl oxidase to produce aldehyde residues which spontaneously from divalent inter- and intramolecular crosslinks, forming the collagen fibril. Image adapted from Kadler *et al*. (1996) [40] and Van der Rest and Garrone (1991) [39].

The collagen fibril is the smallest functional unit of tendons and ligaments and is composed of the fibrillar collagens types I, III and V as well as fibril associated collagens with interrupted triple helices (FACITs) (Figure 1.3).

FACITs, such as collagen types XII and XIV, associate with the external surface of the fibril and regulate its diameter, interact with the surrounding matrix and arrange fibrils into parallel conformations [38], [40], [41]. The FACIT collagens have numerous functions which is beyond the scope of this dissertation and has been extensively reviewed by Shaw and Olsen (1991) [42].

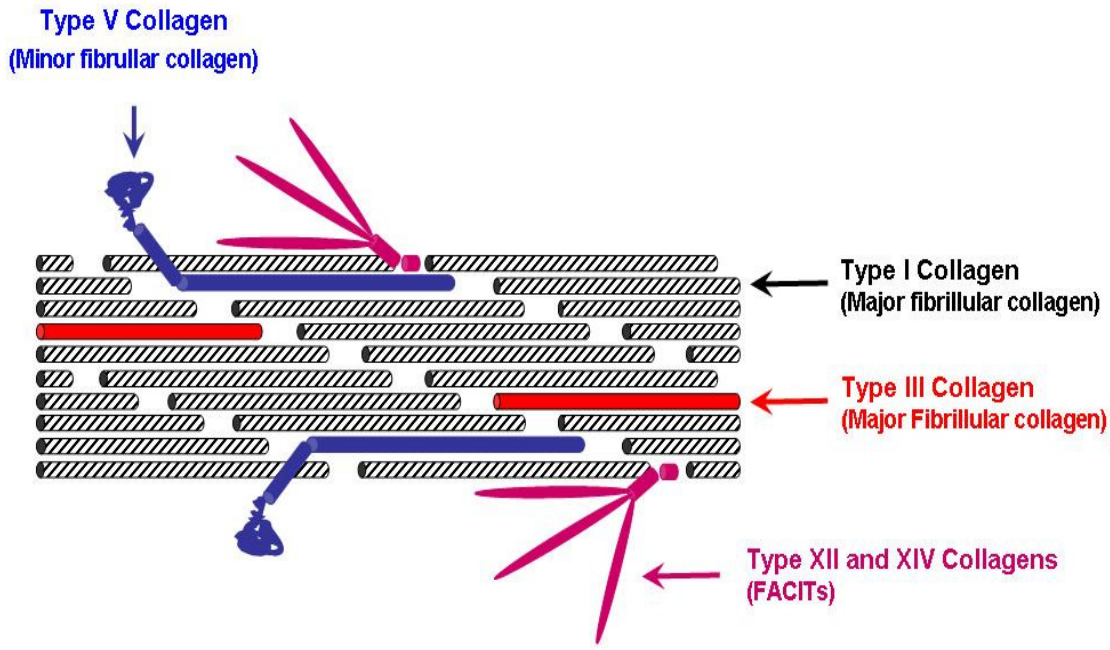


Figure 1.3 Schematic representation of collagen fibre from tendon with associated collagen molecules (adapted from Collins and Raleigh (2009) [12]).
Type I collagen (hatched) is composed solely of helical domains while type V collagen has 2 helical domains within the fibril and 1 globular domain which protrudes to the exterior. FACIT collagens have a short helical domain and interact with the exterior surface of the fibril

1.2.1 Type I collagen

Type I collagen, the major component of tendons is encoded by the *COL1A1* and *COL1A2* genes which encode the α1(I) and α2(I) chains respectively. The predominant form of type I collagen in tendons is a heterotrimer composed of two α1(I) and one α2(I) chains. The orientation of these fibrils gives the tissue its characteristic mechanical properties with a parallel arrangement

noted for tendons exposed mainly to tensile and compressive forces and a multidirectional arrangement noted in the small intestine [43].

1.2.2 Type III collagen

Type III collagen is a quantitatively minor collagen which is expressed as a homotrimer composed of three α1(III) chains encoded by the *COL3A1* gene. Type III collagen expression is markedly increased at the site of tendon rupture and in healing tendons [44]. *In vitro* wounding experiments show that wounded tenocytes expressed increased amounts of type III collagen as opposed to type I [45]. It is believed that increased expression of type III collagen in tendons decreases collagen fibril strength and, over time, increases susceptibility to tendon rupture [9], [44].

1.2.3 Type V collagen

As type V collagen is the focus of this dissertation, it will therefore be discussed in greater detail. It is a quantitatively minor fibrillar collagen most notably present in tendons, ligaments, skin and the cornea with its component chains being encoded by the *COL5A1*, *COL5A2* and *COL5A3* genes. The major isoform which is present in tendons however, is composed of two α1(V) chains and one α2(V) chain. Type V collagen has been shown to play a vital role in the initiation of collagen fibrillogenesis (fibril creation), forming heterotypic fibrils together with collagen types I and III [46]. Although type V collagen is present in relatively low quantities compared to collagen type I, there is significant evidence to suggest that it has an important function in the formation of connective tissue [47].

Type V collagen, unlike most fibrillar collagens, maintains its N-terminal globular peptide domain. During fibrillogenesis, the type V molecules associates with type I molecules in such a way that the globular domain of type V protrudes outwards of the fibril limiting lateral expansion (Figure 1.3) [40]. In an experiment conducted by Birk *et al*. (1990) type I and V collagen molecules were extracted from chick embryos and purified [46]. These molecules were then combined in varying ratios in *in vitro* assembly assays

and the diameters of the resulting fibrils were measured using transmission electron microscopy. The authors concluded that that an increased ratio of type V collagen to collagen type I significantly reduces fibril diameter. In the cornea of the eye where very thin fibrils are required to allow for the transparency of the organ, the heterotypic type I/V bundles have increased type V content relative to tendons and other non-cartilaginous connective tissues [48].

It has previously been shown that tendons with fibrils with large diameters are more resistant to mechanical damage and Lu *et al.* (2011) attempted to engineer *COL5A1*- and *COL5A2*- siRNA producing tenocytes in a Sprague-Dawley rat model [49], [50]. These engineered tenocytes were hypothesised to produce more robust collagen fibrils and injury resistant tendons. The engineered tenocytes did indeed have significantly reduced *COL5A1* and *COL5A2* expression. This reduction however, had deleterious effects on other matrix components such as *COL3A1*, *COL11A1*, and the proteoglycan decorin with the resulting *COL5A1*-siRNA producing tendons with abnormal fibril morphology.

It appears that there is limited biological redundancy in the formation of the collagen fibril. Mutations within many of the genes which encode for collagens such as *COL1A1, COL1A2, COL3A1, COL5A1* and *COL5A2* cause severe connective tissue disorders such as osteogenesis imperfecta and Ehlers-Danlos syndrome (EDS).

1.3 Ehlers-Danlos syndrome

EDS is a heterogeneous soft tissue disorder in which patients exhibit joint hypermobility, skin hyper-extensibility and unusual skin scarring [51]. The classical form of EDS arises from deleterious mutations in the genes which encode the fibrillar collagen with more than 40% of classical EDS arising from haploinsufficiency of the *COL5A1* gene alone [52]. While *Col5a1* mice die *in utero* at embryonic day 10.5 due to cardiac failure [47], [53], experiments conducted by Wenstrup *et al*. (2011) found that *Col5a1* haploinsufficient

mice (*col5a1*$^{+/-}$) survive infancy but show symptoms characteristic of EDS such as skin hyper-extensibility [54].

The fibril architecture of the tendons of *col5a1*$^{+/-}$ mice was however comparable to the wild type fibril architecture. To investigate this protection of the tendon fibril morphology, the authors introduced a mutation in the structurally similar type XI collagen gene in *col5a1*$^{+/-}$ mice. The fibrils of the resulting mice showed a disrupted architecture in tendons. Type XI collagen predominantly forms heterotypic fibrils with type II collagen in cartilaginous tissues [55]. To investigate whether type XI collagen was present in the adult mouse tendon, Wenstrup *et al.* (2011) investigated the expression patterns of collagen types V and XI in a murine model [54]. The authors observed that relatively high levels of *col11a1, col11a2* and *col11a3* are expressed in the developing tendon (embryonic day 18) together with *col5a1* and *col5a2*. This expression however, decreases as the tendon matures with basal levels attained at post-natal day 30. This data indicates a co-operative effect of collagen types V and XI during tendon development but this interaction may be absent during mature tendon healing. The absence of type XI collagen expression with increased expression of the fibril weakening type III collagen may cause a reduction in the strength of the recovered tissue, increasing the risk of tendon pathology.

1.4 Pathology of Achilles tendinopathy

Mechanical loading is generally agreed to be the primary stimulus that causes Achilles tendon injuries and several theories have been proposed to explain how it causes degeneration of the tendon [31]. These include (i) the mechanical theory which states that repeated loading of the tendon within the physiological range eventually leads to matrix degeneration (reviewed by Archambault et al (1995) [56]) and Cook and Purdam (2012) [57]), (ii) the vascular theory which states that certain regions of tendons are susceptible to vascular compromise and neovascularisation and inappropriate vascularisation of the tendon may cause tendinopathy (reviewed by Pufe *et al.* (2005) [58]) and (iii) the neural theory which states that neural cells within and around the tendon may affect the secretion of signalling factors,

such as substance P and calcitonin gene related peptide which affect mast cells within the tendon and cause degeneration of the fibrils (reviewed by Abate *et al.* (2009) [59]). All three theories are proposed to result in the degeneration of the tendon and disruption of the collagen fibrils.

Using both clinical and biochemical data, several injury models have been proposed, of which two, the "iceberg" and "continuum of tendon pathology" models, will be reviewed here. The "iceberg theory" injury model proposed by Fredberg and Stengaard-Pederson, states that repeated heavy loading produces pathological changes to the extracellular matrix or cellular component of the tendon [60]. This progressive damage leads to changes in the micro- and macrostructures of the tendon and causes collagen fibres to slide past one another causing denaturation and inducing inflammation, oedema and pain. In addition, the authors proposed that cytokines and pro-inflammatory molecules such as prostaglandin E2 (PGE2) secreted from inflammatory cells within the tendon and surrounding tissues during loading exercise contribute significantly to tendon degeneration [61].

According to the "iceberg theory" if an individual continues to overload the tendon without sufficient periods of rest, the microtrauma accumulates and the tendon develops asymptomatic tendon pathology. If this individual continues to expose the tendon in question to excessive loading, the tendon abnormalities accumulate to the point where symptoms occur and this is known as the "tip of the iceberg" (Figure 1.4). The strength of this model is that it explains recurrent tendinopathy, as individuals who become symptomatic may temporarily lower the loading of the tendon until symptoms subside but not long enough for the tendon to reach homeostasis and, as such, will fluctuate between symptomatic and asymptomatic states.

According to the model of tendon pathology proposed by Cook and Purdam (2009) (Figure 1.5), tendon pathology occurs as a continuum which may be divided into three major stages. The first stage is reactive tendinopathy, in which there is a non-inflammatory proliferative response within the cells of the tendon [62]. This may occur due to an acute compressive or tensile

overload of the tendon and often results in thickening of the overloaded portion of the tendon. Histologically, the major features of this stage are the rapid increase in large, water sequestering proteoglycans such as aggrecan. This sequestration of water is believed to thicken the overloaded portion of the tendon by increasing the cross-sectional area and decreasing the force per unit area. The integrity of the collagen molecules is mostly maintained at this stage and the tendon has the potential to revert to its normal state if the load is reduced or there is sufficient period between loadings.

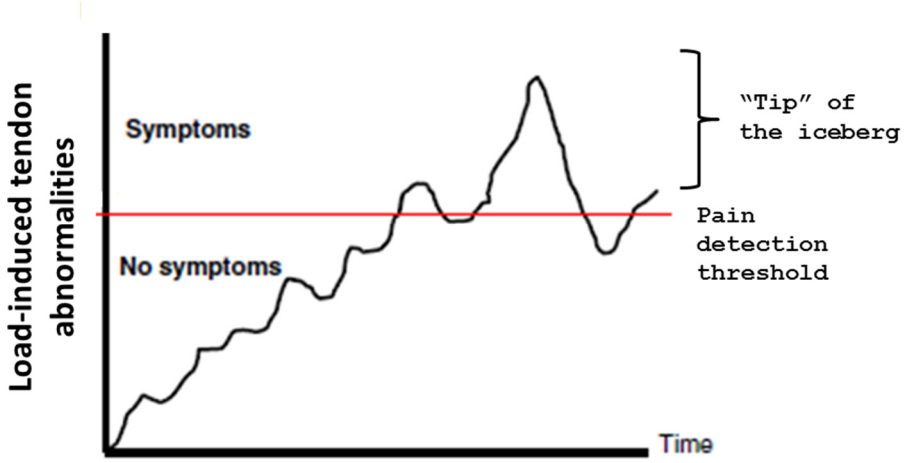


Figure 1.4 Schematic diagram of the Iceberg theory showing asymptomatic tendon pathology before the onset of symptomatic tendinopathy. Adapted from Fredberg and Stengaard-Pederson (2008) [60]

The second stage is tendon disrepair where there is an overall increase in the total number of cells within the tendon including chondrocytes and myofibroblasts, which increase the production of collagen and proteoglycans. The increased proteoglycan-rich ground substance causes separation of the collagen strands and disruption of the matrix. This stage is hard to clinically distinguish but the major symptoms are thickening of the tendon and focal structural changes, which are sometimes accompanied by increased vascularity. Partial reversal of symptoms is possible with proper clinical treatment.

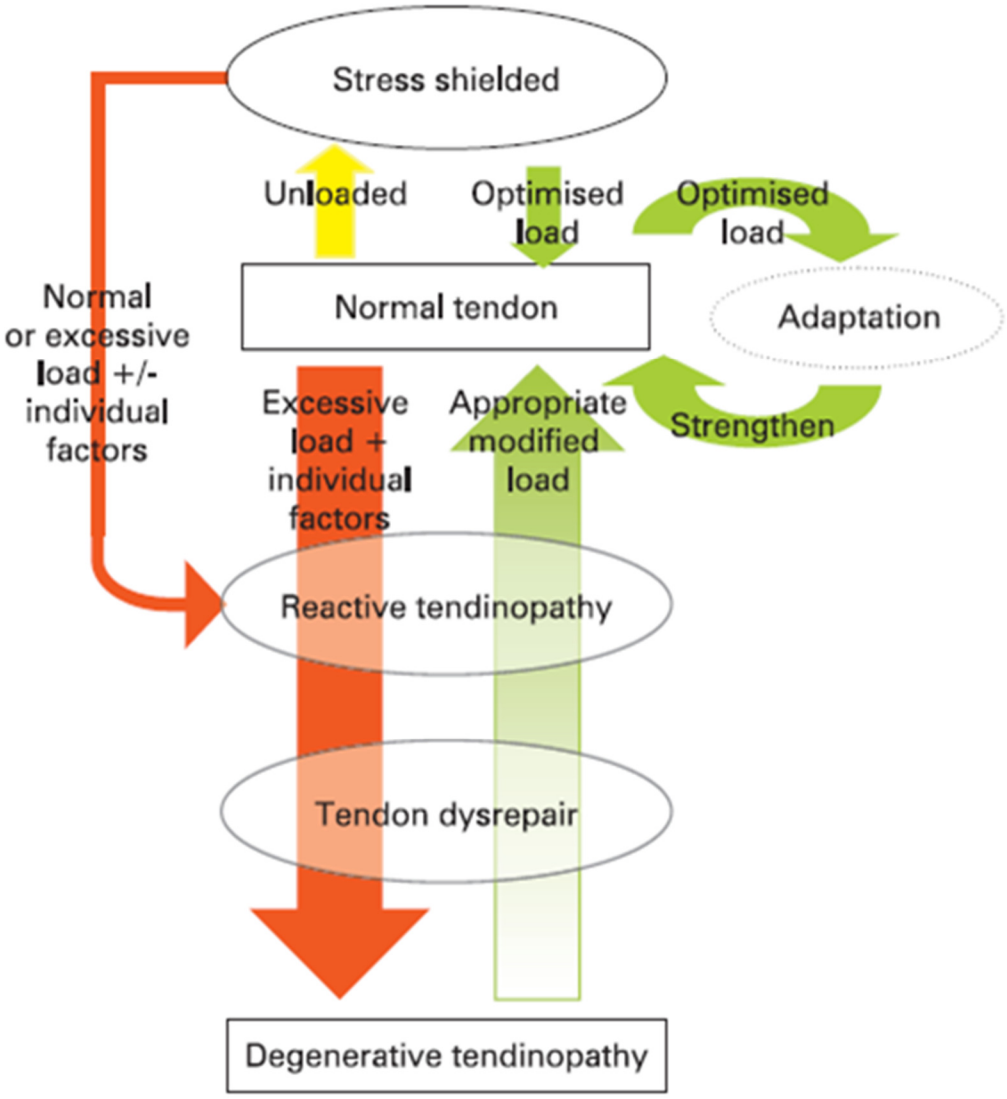


Figure 1.5 Schematic diagram of the pathological continuum of tendinopathy showing transition from healthy tendon to degenerative tendinopathy along with the tendon's adaptive response when excessive load is removed [62]

The final stage is degenerative tendinopathy, when the tendon consists of heterogeneous regions of disrupted matrix and healthy tendon. Histologically, there are large areas of acellularity due to apoptosis and large portions of the matrix are disordered. The most striking feature however, is the disruption of the collagen fibrils. It has been proposed that the nature of the collagen fibrils directly affect the mechanical properties of the tendon and its susceptibility to injury. Although several molecular processes are likely to be involved in the complex aetiology of tendinopathy, it has been

proposed that the nature of the collagen fibrils may directly affect the mechanical properties of the tendon and its susceptibility to injury [63]. Recent evidence has suggested that regulation of the 3'-untranslated region (3'-UTR) of the *COL5A1* gene may play an important role in regulating the mechanical properties of tendons [64].

1.5 3'-UTR regulatory elements

Several different regulatory mechanisms may be present within the 3'-UTR including polyadenylation (poly(A)) signals, RNA-binding protein and miRNA binding sites. These mechanisms affect protein levels by influencing the speed of translation, the rate of mRNA decay and/or affecting the ability of the ribosome translation complex to translate the gene transcripts with high fidelity [65], [66], [67], [68].

1.5.1 Poly(A) signals

The poly(A) signal is a hexanucleotide sequence AAUAAA, which is followed by a GU rich region 20-40 bases downstream [69]. Together, these two regions determine the site of pre-mRNA cleavage and the position at which the adenosine polymer is initiated. The speed at which the polyadenosine tail degrades influences the half-life of the transcript by a process of deadenylation-dependant mRNA decay [70]. The transcript is de-adenylated and decapped at the 5' end before degradation by endo- and exonucleases [66]. The *COL5A1* gene has 3 putative poly(A) signals within its 3'-UTR. Laguette *et al.* (2011) generated a *COL5A1* 3-UTR luciferase construct in which 2 of the poly(A) signals were removed by deleting the final 477 bp of the *COL5A1* 3'-UTR and found that, although there was an overall increase in mRNA stability in both the C- and T-allelic forms, the T-allele still showed increased mRNA stability over it C-allele counterpart [64]. This indicates that the while the poly(A) signals may be important in the regulation of the *COL5A1* 3'-UTR, they do not play a role in the difference in mRNA stability observed between the C- and T-allelic forms.

1.5.2 RNA-binding proteins

RNA-binding proteins (RBPs) are proteins which bind to single- or double-stranded RNA through a RNA recognition motif [71]. This group of proteins have a variety of functions including protection from dsRNA-viral infection, mRNA transport, RNA splicing and influencing mRNA translation [72], [73], [74], [75], [76], [77], [78]. There is, to date, no literature available on the regulation of the *COL5A1* 3'-UTR by RNA-binding proteins.

1.5.3 MicroRNAs

Micro-ribonucleic acids (miRNAs) are short non-coding RNAs between 18 and 24 nucleotides (nt) long that are capable of repressing protein synthesis at the mRNA level [79]. This is achieved by miRNAs modulating the stability and/or the translational efficiency of target mRNAs by binding to their 3'-UTR [80]. The product of a single miRNA gene is capable of regulating hundreds of different messenger RNAs [81]. Furthermore, the messenger RNAs may have binding sites for several miRNAs in their respective 3'-UTRs. This leads to a complex regulatory system with built-in redundancy and feed-forward mechanisms [82].

MicroRNAs act together with a group of proteins called the RNA-induced silencing (RIS) complex to carry out their function. Argonaute II, the endoribonuclease present in this complex, is responsible for mRNA cleavage [75], [83]. Its enzymatic activity is dependent on the miRNA having a high level of complementarity to its target sequence. mRNA cleavage by perfect base pair complementarity is rare in humans but is the dominant form of mRNA regulation in plants [84]. If the complementarity between the miRNA and its mRNA target is insufficient to induce mRNA cleavage, secondary repression mechanisms such as silencing complex-induced translation interference and miRNA-directed de-adenylation may come into effect [65], [85].

The first repression mechanism relies on the steric interference between the multiple silencing complexes bound to the 3'-UTR and the RNA polymerase II translation complex. This leads to reduced translational efficiency and

increased truncated non-functional product formation. The second repression mechanism, hypothesized to be the dominant mechanism by which miRNA regulates mRNA, causes rapid miRNA-induced de-adenylation of the target mRNA [65]. The target mRNA is then quickly degraded by endogenous exonucleases. Together these processes effectively lower the total target protein concentration.

Both of these mechanisms rely on the target miRNA recognizing its binding site within the target mRNA. The current theory is that this binding occurs through an mRNA-miRNA "pivot" mechanism [86]. This theory states that a transition nucleation bulge occurs when 5 consecutive bases (nt 2-6) of the miRNA seed region binds to the target sequence in a "pivot-pairing" manner. This mechanism raises the important possibility that non-canonical mRNA-miRNA binding sites, which do not require complete miRNA seed complementarity would be allowed as long as a G-bulge occurs at positions 5-6. This would have important implications for the role of mRNA secondary structure with regard to miRNA-mediated regulation and target recognition.

To date, only miR608 has been shown to regulate the *COL5A1* mRNA [64].

1.6 RNA Secondary structure

RNA secondary structure refers to the 2-dimensional conformation that the RNA molecule forms shortly after transcription by the ribosome. This folding is directly influenced by the primary sequence of the RNA molecule which governs the free energy landscape for the folding mechanics [87], [88]. This folding is believed to be more complex than that of DNA due to two main factors. Firstly, RNA base pairing allows guanine to uracil pairing (G>U) increasing the possible RNA conformations and secondly, RNA is prone to spontaneous formation of tertiary structures and pseudo-knots [89].

RNA secondary structure is important for correct regulation of certain processes at the mRNA level. In order to regulate mRNA, *cis*-elements such as RNA-binding proteins and miRNAs bind to unstable, single-stranded regions within the 3'-UTR [90]. In a paper published by Zhao et al. (2005)

non-regulatory “spacer” regions of 3-UTR were deleted and the effect on miRNA binding observed [91]. These regions were found to be vital for correct regulation, and in some instances, when deleted the effects of miRNA-mediated mRNA decay was completely abolished.

One of the current hypotheses explaining this observed phenomena, states that changes in the primary RNA sequence cause alterations in the secondary structure of the mRNA which render the RIS complex unable to bind to the relevant region within the 3’-UTR [90]. The guide-strand of the miRNA molecule within the RIS complex binds to unstable RNA which is free single-stranded, within a bulge or a hairpin loop [91]. Should the binding site become stabilized by base-pairing, the *cis*-element becomes unable to recognize and bind to the respective site [92].

The positions of these unstable regions are constantly in flux due to several factors including the fluid nature of the cytoplasm, changes in temperature, the binding of RNA-binding proteins, the effect of small molecules such as ATP and Mg^{2+} and K^{+} ion concentrations [78], [89], [93]. There are however, favoured conformations of RNA secondary structure which exist *in vivo*. These conformations are present in “conformation wells”, which are points in the free energy landscape in which only a few conformations of the secondary structure are favourable. At a given temperature and under homeostatic conditions, the RNA secondary structure may cycle through only a subset of all the possible permutations [93]. The presence of these conformation wells allows the RNA to spend significant amounts of time in conformations that are suitable for regulation by trans-elements rather than constantly shifting through all possible conformations. Significant alterations to the primary sequence, especially sequence duplications, insertions and deletions, may alter the secondary structure of mRNA thereby shifting the location of local minima within the free energy landscape. These altered landscapes could affect RNA regulation by trans-acting mechanisms such as RNA-binding proteins and miRNAs and may lead to mRNA dysregulation.

1.7 Characterizing the structure and function of the *COL5A1* 3'-UTR

As previously stated, Achilles tendinopathy is a multifactorial disorder and as such, the risk profile will differ significantly between individuals depending on their exposure to the extrinsic and intrinsic factors described in Table 1.1. With respect to the identified intrinsic genetic component, independent genetic associations with variants in several genes such as the ECM glycoprotein tenascin C (*TNC*) [25], matrix metalloprotease 3 (*MMP-3*) [26]and *COL5A1* [13], [14], tissue inhibitor of metalloproteinase-2 (*TIMP-2*) [27] and caspase 8 (*CASP8*) [94] and variation in the expression of these genes or functionality of their products may contribute significantly to the development of tendinopathy.

As previously discussed, much of the studies to date have focused on the association of variants within the 3'-UTR of the *COL5A1* gene with Achilles tendinopathy and other exercise-associated phenotypes and several polymorphisms have been associated with Achilles tendinopathy (See Figure 1.6). The variants rs12722 (C/T), rs3196378 (A/C), rs71746744 ($AGGG_{(n)}$), rs16399 ($ATCT_{(n)}$) and rs1134170 (A/T) were all associated with Achilles tendinopathy in at least one Caucasian population [13], [14], [22]. Variants rs13946 and rs3128575 were also investigated but were not found to be associated with Achilles tendinopathy [14].

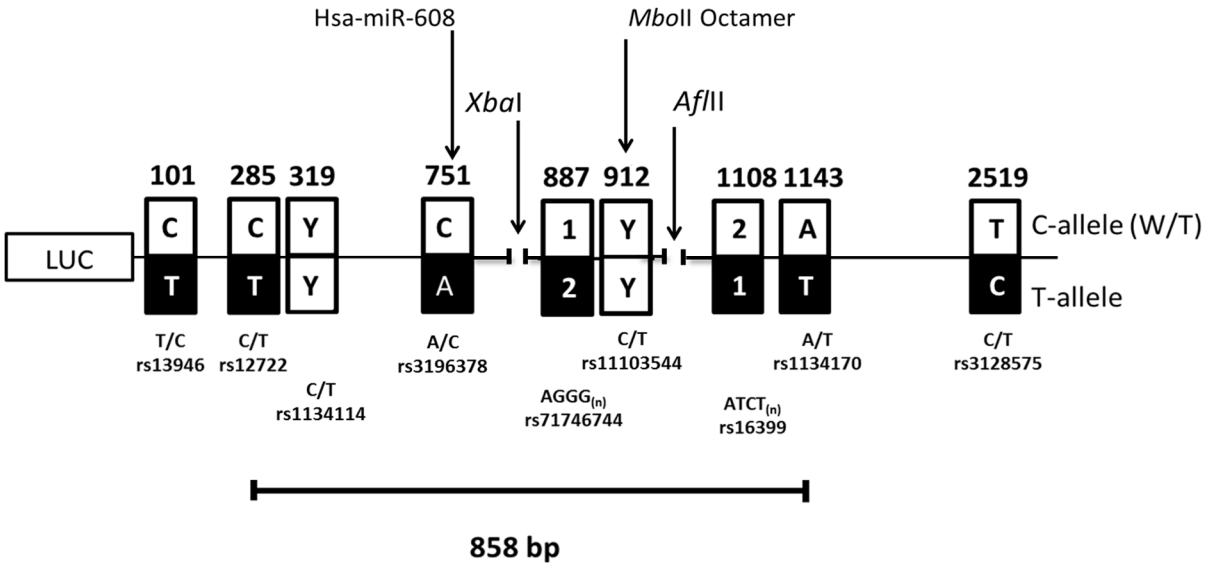


Figure 1.6 Variants within the *COL5A1* 3'-UTR used to determine the allelic form of the *COL5A1* 3'-UTR.
Variants rs12722, rs3196378, rs71746744, rs16399 and rs1134170 were associated with Achilles tendinopathy in at least one population. The 57 bp region containing rs71746744 and rs11103544 is shown along with the positions of restriction enzyme sites for *Xba*I and *Afl*II [13], [14], [22].

Laguette *et al.* (2011) also found that these seven polymorphic regions appeared to be in strong linkage disequilibrium, giving rise to two major allelic forms of the *COL5A1* 3'-UTR [64]. These forms were annotated as the C-allelic form, which represents the wild-type sequence, and the T-allelic form, which was predominantly found in clones derived from symptomatic Achilles tendinopathic individuals and is believed to represent the "at risk" form of the *COL5A1* 3'-UTR. In functional studies, it was found that the T-allele showed an overall increased luciferase activity in a fibrosarcoma cell line. This alludes to increased stability of the T-allelic form of the *COL5A1* mRNA over its C-allelic counterpart. This difference in mRNA stability between the allelic forms was abolished when a 57 bp region containing the *Mbo*II RFLP conserved octamer as well as the $AGGG_{(n)}$ VNTR (rs71746744), was removed.

In addition, two putative polymorphic miRNA binding sites were previously identified within the 3'-UTR [14] (Figure 1.6). The miRNA Hsa-miR-608 is believed to bind the first polymorphic region overlapping with the rs3196378

(C/A) polymorphism in the *COL5A1* 3'-UTR. September *et al.* (2009) found that the CA genotype of this SNP was significantly over-represented in an Australian tendinopathy group compared to the Australian control group [14]. The A-allele was also predominantly found within the tendinopathic clones created by Laguette *et al.* (2011) and it was hypothesized that this SNP alters the binding affinity of Hsa-miR-608 in tendinopathy patients [64].

The second polymorphic miRNA binding site overlaps with the *Mbo*II RFLP polymorphism (rs11103544) and is believed to bind an unknown miRNA. This site has a conserved octamer which is predicted to be a miRNA binding site by the Patrocles miRNA-binding site algorithm [23], [95], [96]. The *Mbo*II RFLP (rs11103544) has not been found to be associated with Achilles tendinopathy and does not contribute to the identification of the C- or T-allelic forms [14].

Collins and Posthumus (2011) hypothesize that over-expression of type V collagen would decrease the maximum load stress of the Achilles tendon by decreasing the average fibril diameter and increasing the total number of fibres present (Figure 1.7) [63]. This may lead to more tightly packed bundles which would decrease the flexibility of the tendon, making it more susceptible to both micro- and macro-injury. This hypothesis is supported by experiments investigating age-related type V increase in rabbit patella tendon [97]. An increased rate of damage to the tendon may alter the production/degradation balance and increase the tendon's susceptibility to degenerative phenotypes such as tendinopathy. These data make *COL5A1*, which encodes a subunit of type V collagen, an excellent gene for further investigation.

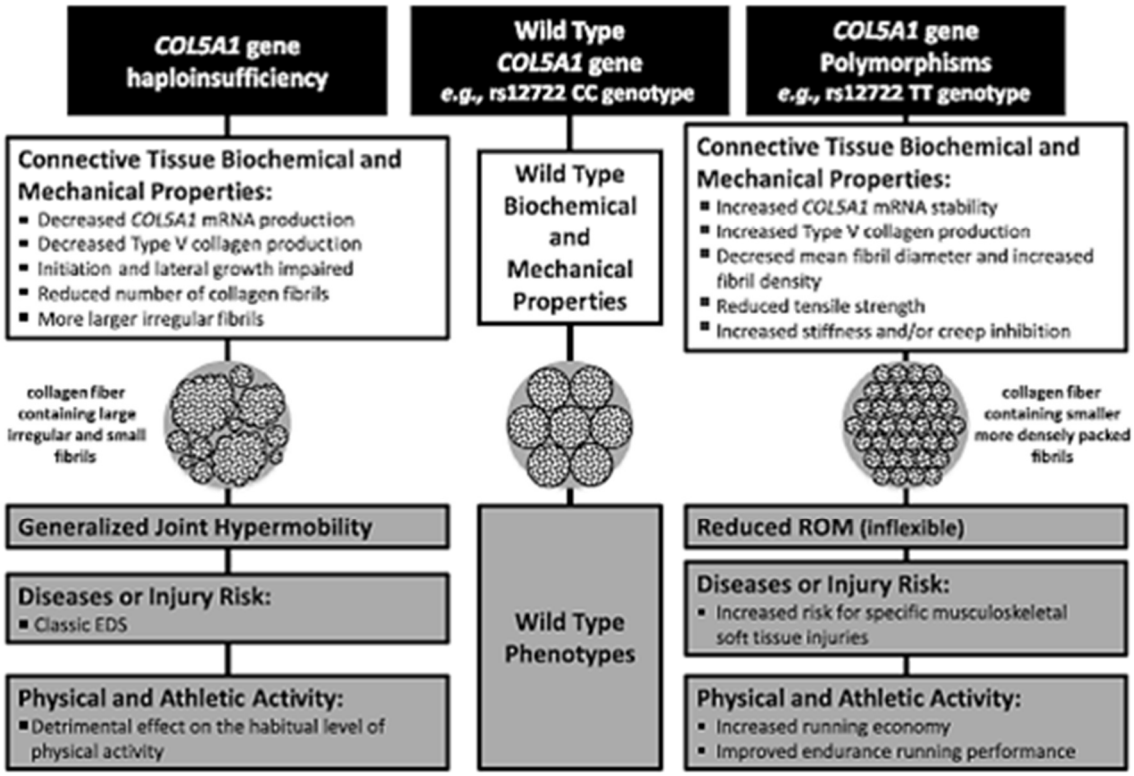


Figure 1.7 Schematic diagram showing the novel hypothesis proposed by Collins and Posthumus (2011).
The left panel shows irregular fibril architecture resulting from *COL5A1* gene haploinsufficiency. Classical EDS symptoms are present at the phenotype level. The centre panel shows the wild-type phenotype with normal fibril architecture. The right panel shows the hypothesised fibril architecture resulting from increased mRNA stability of the T-allelic from of the *COL5A1* gene; the collagen fibrils have a decreased diameter and are more densely packed. This alteration is believed to reduce range-of-motion and increase risk for specific musculoskeletal soft tissue injuries such as Achilles tendinopathy.

1.8 Aims

The aim of this study is to investigate the miRNA regulatory mechanisms within the 3'-UTR of *COL5A1* and specifically, to investigate whether the identified miRNAs are more efficient at regulating one of the two functional allelic forms of the *COL5A1* 3'-UTR. This will be achieved in the following manner:

1.
    a. To investigate the whether the addition of Hsa-miR-608 mimic is sufficient to reduce the luciferase expression of a luciferase construct containing the T- or C-allelic form of the *COL5A1* 3'-UTR using a cell culture co-transfection approach
    b. To determine whether there is a significant difference in the regulation of this miRNA binding site between the C- and T-allelic forms of the *COL5A1* 3'-UTR using luciferase assays.

2.
    a. To investigate the validity of the polymorphic *Mbo*II RFLP octamer as a putative miRNA binding site using a site-directed mutagenesis approach and luciferase assays.
    b. To determine whether there is a significant difference in the regulation of this miRNA binding site between the C- and T-allelic forms of the *COL5A1* 3'-UTR using luciferase assays.

3.
    a. To investigate whether the addition of Hsa-miR-125a-5p mimic is sufficient to reduce the luciferase expression of a luciferase construct containing the T- or C-allelic form of the *COL5A1* 3'-UTR using a cell culture co-transfection approach
    b. To determine whether there is a significant difference in the regulation of this miRNA binding site between the C- and T-allelic forms of the *COL5A1* 3'-UTR using luciferase assays.

4.

a. To determine whether the rs4919510 variant within the *MIR608* gene is associated with Achilles tendinopathy by genotyping 2 Caucasian populations with Fluorescent Taqman® PCR in a case-control study.
b. To investigate gene-gene interaction between rs4919510 and the rs3196378 variant present within the predicted binding site for Hsa-miR-608 in the *COL5A1* 3'-UTR.

5.

To investigate the co-operative effects of the variants within the *COL5A1* 3'-UTR on mRNA secondary structure by creating an *in silico* model of the 3'-UTR using the Sfold online folding algorithm.

2. Materials and Methods

2.1 Control and tendinopathic participants and DNA extraction

Three hundred and twelve asymptomatic control participants (CON) and 143 participants diagnosed with chronic Achilles tendinopathy (TEN) were included in this study. These participants were recruited from South Africa (SA TEN, N = 75 and SA CON = 135) and Australia (AUS TEN, N = 68 and AUS CON = 177). All participants were of self-reported European Caucasian ancestry, gave written informed consent (Appendix A) prior to participation in the study, provided personal particulars and completed a questionnaire regarding medical history and physical activity history (Appendix B). The research project was approved by the Human Research Ethics Committee of the Faculty of Health Science at the University of Cape Town (Appendix C) and the Human Ethics Committee of La Trobe University, Melbourne, Australia (Appendix D).

The South African participants with clinically diagnosed Achilles tendinopathy were recruited from the Sports Medicine Practice at the Sports Science Institute of South Africa in Cape Town, South Africa. The clinical diagnostic criteria for Achilles tendinopathy included progressive pain in the Achilles tendon area for a minimum of 6 months and at least one of the following six criteria: (1) Early morning pain over the Achilles tendon region; (2) early morning stiffness over the Achilles tendon area; (3) a history of swelling of the Achilles tendon area; (4) tenderness to palpation; (5) nodular thickening over the affected tendon or (6) movement of the painful area in the tendon during plantar-dorsi-flexion [98]. The SA TEN diagnoses were reviewed and confirmed by a Sports Physician, using an inclusion and exclusion criteria and a checklist (Appendix E). The SA CON participants reported no history of tendon pathology and were recruited from various recreational sporting clubs. The SA CON group were recruited such that the sex and country of birth were similar to that of the SA TEN group. In addition, in order to avoid the possible confounding effects of age, the age of the SA CON group was matched to that of the age of initial injury onset in the SA TEN group.

The Australian tendinopathy participants were recruited from the Musculoskeletal Research Centre at La Trobe University in Melbourne Australia [14]. A Sports Physiotherapist confirmed the AUS TEN diagnoses, using the same criteria as that of the South African participants. The participants were recruited such that their country of birth was matched between the AUS CON and AUS TEN groups, and the age of recruitment of the AUS CON group matched that of initial onset of injury of the AUS TEN group.

The diagnosis of all Australian tendinopathic participants were confirmed using soft tissue ultrasound imaging of the affected tendon. Diagnosis was confirmed by soft tissue ultrasound imaging of the affected Achilles tendon for a subset (63 of 210) of the South African participants. For the SA participants, approximately 4.5ml of venous blood was collected via venipuncture of a forearm vein into EDTA vacutainer tubes and stored at 4° C until total DNA extraction. Total DNA was extracted from blood using a method described by Lahiri and Nurnberger (1991) [99] and modified by Mokone et al (2006). The DNA was stored at -20° C until sequencing or genotyping. For the Australian participants, approximately 4.5ml of venous blood was collected and DNA was extracted using a sequence extraction technique (FlexiGene DNA Kit, Qiagen P/L, Valencia, California, USA) as described by the manufacturer.

2.2 Cloning of the *COL5A1* 3'-UTR DNA constructs

Ten individuals were selected from the TEN (n=5) and CON (n=5) participants (Table 2.1). The 5 participants in the tendinopathic sub-group represented a "severe" phenotype with early onset of initial symptoms (3 of 5 participants were ≤32 years of age), bilateral Achilles tendinopathy (3 of 5) and a mid-tendon injury (5 of 5). Tendinopathic participants were also known to have the "at risk" TT genotype for chronic Achilles tendinopathy at the *COL5A1* 3'-untranslated region (3'-UTR) single nucleotide polymorphism rs12722 (C/T)[13], [14]. The five CON participants, on the other hand, had the "protective" CC genotype at rs12722.

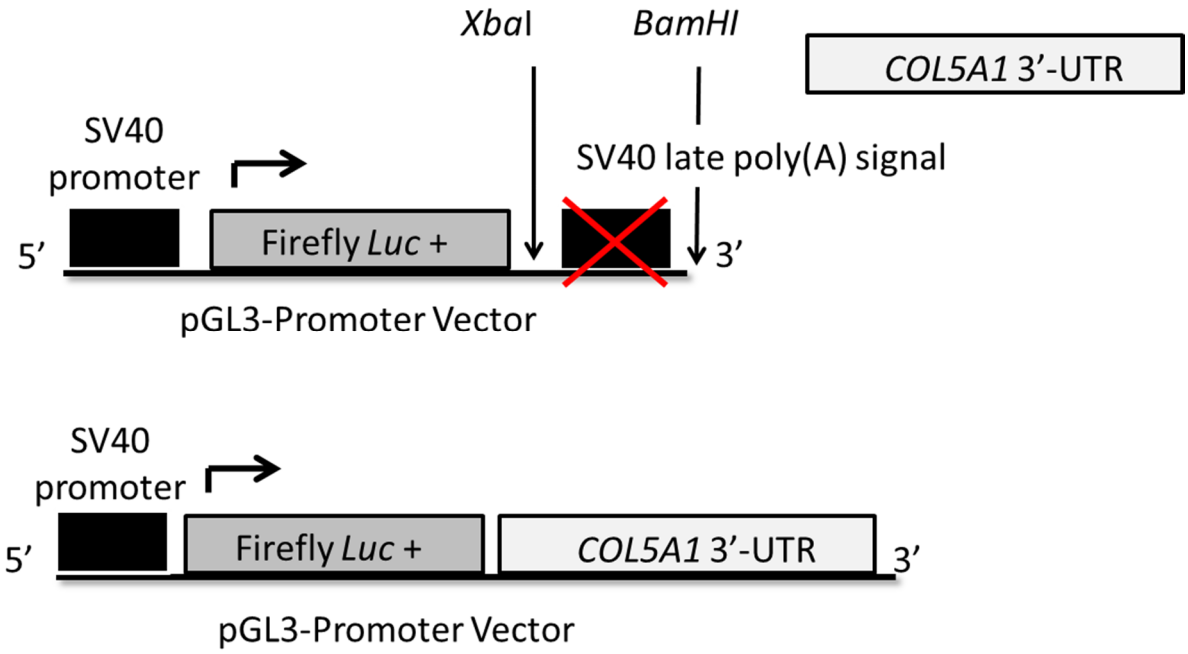


Figure 2.1 Schematic representation of the 2546 bp *COL5A1* 3'-UTR cloned downstream of the firefly luciferase gene (Firefly Luc+) into the pGL3-Promoter vector, substituting the SV40 late poly(A) signal of the reporter gene.
*Xba*I and *Bam*HI restriction enzyme sites were used to clone the *COL5A1* 3'-UTR into the pGL3-Promoter vector. The expression the firefly luciferase gene is driven by a SV40 viral promoter. The plasmid construct is circular but was linearized for schematic purposes (Appendix F)

The pGL3-TEN/CON luciferase vectors were constructed by amplifying a 2546 bp fragment of the 3'- UTR of either tendinopathy or control participants from total genomic DNA via a nested polymerase chain reaction (Appendix G for primers). The amplified fragment was cloned into the pGL3-Promoter vector (Promega Corporation, Madison, Wisconsin, USA) downstream of the firefly luciferase reporter gene using the restriction sites *Bam*HI and *Xba*I to replace the normal SV40 late poly(A) signal (Figure 2.1 - Appendix H)[100].

The pRL-TK vector (Appendix I) contains a constitutive thymidine kinase 3 promoter which drives the expression of a *renilla* luciferase reporter gene. pRL-TK was co-transfected with each pGL3_TEN/CON vector and was used as an internal control to measure transfection efficiency.

Table 2.1 General characteristics of the Achilles tendinopathic (TEN) and control (CON) individuals used to clone the 3'-untranslated region of the *COL5A1* gene.

Participant	*Age at recruitment (yrs.)*	*Gender*	*Height (cm)*	*Weight (kg)*	*BMI (kg/m²)*	*COL5A1 rs12722 genotype*
TEN-1	49	Male	186	87	25.1	TT
TEN-2	27	Female	Unknown	Unknown	Unknown	TT
TEN-3	62	Male	175	64	20.8	TT
TEN-4	47	Male	173	73	24.4	TT
TEN-5	49	Male	179	79	24.7	TT
Mean ± s.d.	46.8 ± 12.6		178.3 ± 5.7	75.8 ± 9.7	23.8 ± 2.0	
CON-1	Unknown	Female	167	54	19.4	CC
CON-2	28	Male	183	78	23.3	CC
CON-3	51	Female	171	56	19.1	CC
CON-4	40	Female	170	58	20.1	CC
CON-5	Unknown	Male	183	79	23.6	CC
Mean ± s.d.	39.7 ± 11.5		174.8 ± 7.6	65 ± 12.4	21.1 ± 2.2	

BMI - Body mass index **s.d. –** standard deviation

DNA constructs were amplified using caesium chloride/ethidium bromide gradient centrifugation. The pGL3-TEN/CON luciferase constructs were transformed into the DH5α strain of *E. coli* and were selected for ampicillin resistance. Positive clones were picked and grown in a 2 litre conical flask containing 400 ml of 100 µg/ml ampicillin Luria broth at 37° C and shaken at 150 rpm for 20 hours. The cell cultures were then centrifuged at 7000 rpm using the JA-10 rotor in a Beckman Coulter Avanti® J-E centrifuge for 10 min at 4^{O}C. The pellets were resuspended in 40 ml of maxiprep solution I. The cells were lysed by the addition of 80 ml of maxiprep solution II containing 1% sodium dodecyl sulphate. The alkaline lysates were neutralized using maxiprep solution III and the resulting macromolecule aggregates were removed by centrifugation at 7000 rpm using the JA-10 rotor in a Beckman Coulter Avanti® J-E centrifuge for 10 min at 4^{O} C. The supernatants were mixed with 90 ml of isopropanol and centrifuged at 7000 rpm using the JA-10 rotor in a Beckman Coulter Avanti® J-E centrifuge for 10 min at 4^{O} C to pellet remaining genetic material. The supernatants were discarded and the pellets resuspended in 2 ml of sterile nuclease free water. The solution was then mixed with 4.1 g of caesium chloride salt and 200 µl of 10mg/ml

ethidium bromide, a DNA intercalating agent. The solution was ultracentrifuged at 53 000 rpm using the Beckman Vti 65.2 Vertical tube rotor for 20 hours at 20^{O} C and the resulting bands of plasmid DNA were removed with 21 gauge needle into a 5 ml syringe. The solution was washed with 1 volume of isobutanol a minimum of 3 times to remove the ethidium bromide and 1 µl of 10 mg/ml RNAse A was added to the solution to remove any RNA contamination. The plasmid DNA samples were dialysed in pH 7.5 Tris-EDTA solution to reduce the caesium chloride concentration. The remaining protein components were removed by phenol-chloroform extraction and the aqueous supernatants were transferred to fresh eppendorf tubes. The solutions were centrifuged at 10 000 rpm using the Biofuge Fresco bench top centrifuge (Heraeus Instruments, Hanau, Germany) for 5 minutes and the supernatants removed and discarded. The plasmid DNA pellets were washed with increasing concentrations of ethanol, air dried for 15 min and resuspended in 100 µl of nuclease free water and nanodropped using the NanoDrop® ND-1000 spectrophotometer for quality and quantity.

The entire *COL5A1* 3'-UTR within the pGL3-TEN/CON luciferase construct clones were sequenced by the the Central Analytical Facility (University of Stellenbosch, Stellenbosch, South Africa) as previously described [64]. BioEdit 7.0.5.2 (www.mbio.ncsu.edu/bioedit/bioedit.html) and several miRNA databases (Patrocles, www.patrocles.org) [95], miRBase (www.mirbase.org) [101], [102] and MicroCosm Targets (www.ebi.ac.uk/enright-srv/microcosm) were used to analyze the obtained sequence information (refer to section 2.8). Two major allelic forms of the *COL5A1* 3'-UTR were distinguished from the sequenced clones which were named the C- (which corresponds to the published wild-type sequence) and T-alleles. The C-form was predominately identified in the CON clones, while the T form was predominately identified in the TEN clones [64]. Seven tightly linked polymorphisms (rs13946, rs12722, rs3196378, rs71746744, rs16399, rs1134170 and rs3128575) distinguished the two major forms of the 3'-UTR (Figure 2.2)

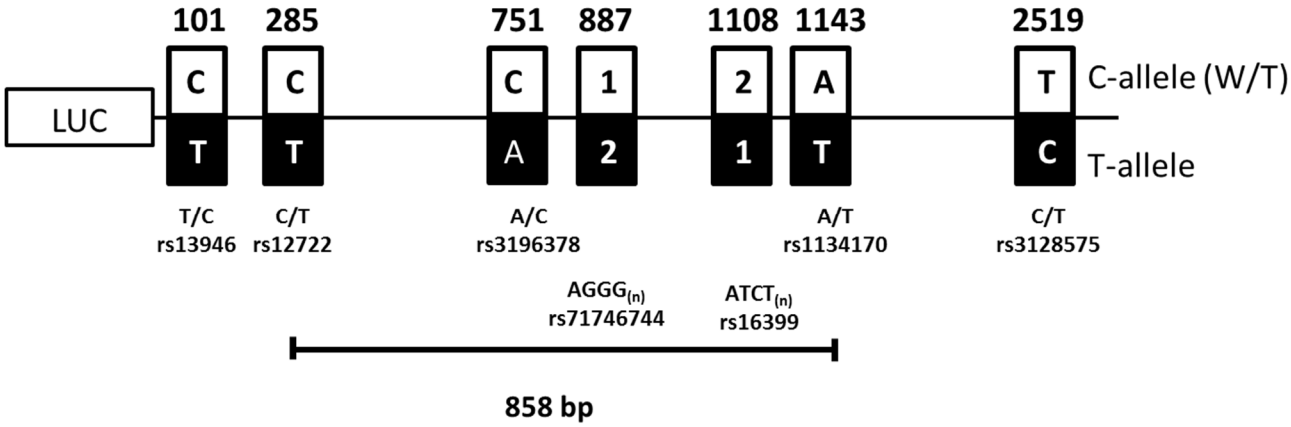


Figure 2.2 Schematic representation of the 2546 bp *COL5A1* 3′-UTR cloned downstream of the firefly luciferase gene (LUC of the pGL3-Promoter vector), substituting the SV40 late poly(A) signal of the reporter gene [64].
The seven tightly linked polymorphic sites within the 3′-UTR, which distinguished between the C- (wild-type) and T-allelic forms, are annotated as white (top) black (bottom) boxes respectively. The accession numbers associated with the polymorphic sites are indicated together with the nucleotide changes. The sequence of the seven polymorphisms for the C- (wild-type) and T alleles was C-C-C-$(AGGG)_1$-$(ATCT)_2$-A-T and T-T-A-$(AGGG)_2$-$(ATCT)_1$-T-C respectively. The nucleotide position of the polymorphisms within wild-type sequence of exon 66 of *COL5A1* is also indicated.

2.3 Site-directed mutagenesis of constructs

One construct of each allelic form of the *COL5A1* 3'-UTR (Con 5.1 and Ten 1.1.) was subjected to site-directed mutagenesis at an *Mbo*II RFLP site which overlapped a region believed to contain a putative miRNA binding site (Figure 2.3) [64]. A *Bgl*II restriction enzyme site was introduced into the putative miRNA binding site mutating the primary sequence from **5'-**TTT TCT TTT TTC TAC ATG CAC TTA AGA C-3' to 5'-TTT TCT TAG ATC TAC ATG CAC TTA AGA C-3' (Figure 2.3) using *pfu* High fidelity polymerase. Parental template DNA was amplified and mutated using the following primers (the putative miRNA binding site is underlined):

Fwd: 5’-GGG AAA AAA TTT GAT AAT TTT CTT AGA TCT ACA TGC ACT TAA GAC-3’

Rev: 5’-GTC TTA AGT GCA TGT AGA TCT AAG AAA ATT ATC AAA TTT TTT CCC-3’

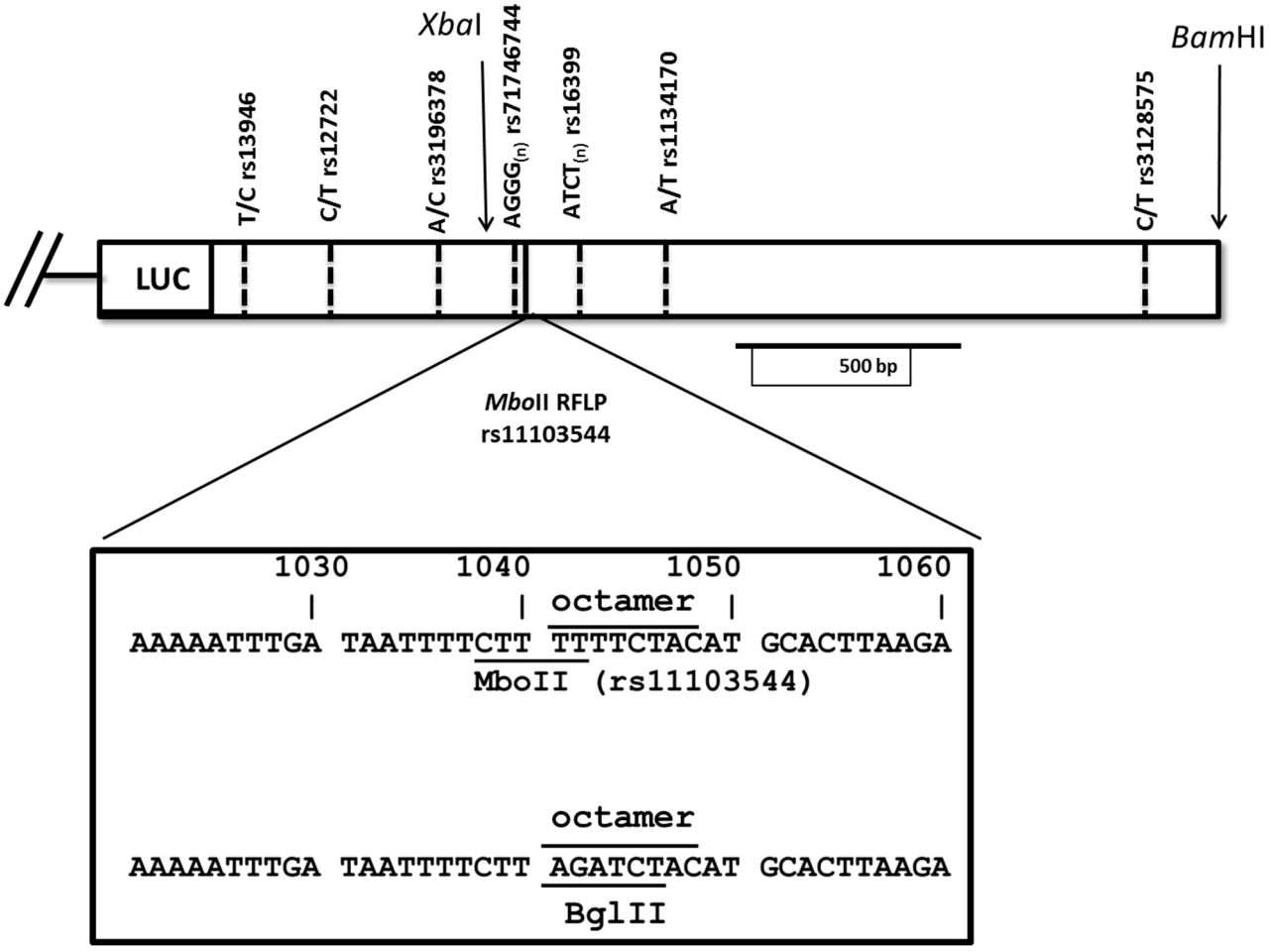


Figure 2.3 Schematic representation of exon 66 of the *COL5A1* gene showing position of the wild-types sequence and mutated nucleotides within the MboII conserved octamer and their position within the 2546 bp sequence of the 3’-UTR of *COL5A1*.

A final concentration of 8% DMSO was used to facilitate DNA strand separation. *Dpn*I methylation-specific restriction digestion was used to remove parental template DNA. Successfully mutated clones were confirmed using *Bgl*II diagnostic enzyme digestion. These clones were digested using the *Bam*HI and *Xba*I restriction enzymes and the resulting 2 kb fragment containing the mutated 3’-UTR region was inserted into the *Bam*HI-*Xba*I sites of corresponding parental construct.

2.4 Plasmid construct sequencing

Different primer pairs (Appendix J) were used to PCR amplify selected areas of the 2 kb cloned *Bam*HI-*Xba*I fragment of the *COL5A1* 3'-UTR containing the *Mbo*II RFLP site-directed mutation. The PCR fragments were gel purified using the QIAEX®II Gel Extraction kit (QIAGEN GmbH, Hilden, Germany), and the subsequent sequencing reactions using the appropriate primers were performed by the Central Analytical Facility (University of Stellenbosch, Stellenbosch, South Africa). BioEdit 7.0.5.2 (www.mbio.ncsu.edu/bioedit/bioedit.html) was used to analyze the obtained sequence information.

2.5 Cell culture and propagation of cell lines

The HT1080 human fibrosarcoma cells (ATCC® no. CCL-121) and SVWI-38 immortalised human embryonic lung fibroblast cells (ATCC® no. CCL-75) [103] were cultured from frozen stocks stored in liquid nitrogen. Cells were maintained in 10 cm cell culture dishes at 37° C in an atmosphere composed of 5% CO_2 and 95% air at 65% humidity. Cells were grown in Dulbecco's Modified Eagle's Medium (DMEM) containing 10% fetal bovine serum (FBS) (Highveld Biological, South Africa) and 1% Penicillin/Streptomycin antibiotic solution. Cells were maintained at 70-90% confluency and passaged every 48-72 hours by treating with 0.05%/0.02% w/v trypsin/EDTA solution. Cells were regularly tested for mycoplasma infection and visually monitored for fungal and microbial infections.

2.6 Transient transfections

Twenty-four hours prior to transfection HT1080 and SVWI38 cells were seeded at 7.5 x 10^4 and 2.0 x 10^5 cells per well respectively, in a 12 well plate and were 60-80% confluent at time of transfection.

2.6.1 DNA transient transfections

Cells were transfected with pRL-TK (50 ng) and pGL3-*COL5A1*_TEN/CON (500 ng) per well using XtremegeneHD (Roche Diagnostics, Roche Applied Science, Mannheim, Germany). The XtremegeneHD transfection reagent was prepared in a ratio of 3:150 of transfection reagent to serum-free culture medium and added to nucleic acid mix in a solution of nuclease free water. Mixture was briefly vortexed and centrifuged at 13 000 rpm using the Biofuge Fresco bench top centrifuge (Heraeus Instruments, Hanau, Germany) for 30 seconds and incubated at room temperature for 15 min. Forty-nine µl of transfection complex was added drop-wise to each well.

2.6.2 RNA:DNA transient co-transfections

Cells were transfected with pRL-TK (50 ng) and pGL3-*COL5A1*_TEN/CON (500 ng) per well using Attractene (QIAGEN GmbH, Hilden, Germany), a non-liposomal reagent which can bind to both DNA and RNA. The reagent-nucleic acid complexes fuse with the plasma membranes of the recipient cells and release their nucleic acid content. Cells were also transfected with 0, 1, 10 or 20 pmol of miR608 or miR125a-5p miScript miRNA Mimic (QIAGEN GmbH, Hilden, Germany). The amount of transfected RNA was kept constant by adding Allstars negative control siRNA** (QIAGEN GmbH, Hilden, Germany). The miRNA mimic is a synthetic RNA which imitates the mature form of the miRNAs tested. Allstars negative control is a short sequence of scrambled siRNA which has been extensively tested to confirm that it does not influence mRNA stability of any currently annotated human gene or of the transfected vectors. Attractene transfection reagent was prepared in a ratio of 9:240 of Attractene to serum-free culture medium and added to the nucleic acid mix in a solution of nuclease free water. Mixture was briefly vortexed and centrifuged at 13 000 rpm using the Biofuge Fresco bench top centrifuge (Heraeus Instruments, Hanau, Germany) for 30 seconds and incubated at room temperature for 15 min. Seventy-

seven µl of transfection complex was added drop-wise to each well (Appendix K for complete protocol).

The miR608 co-transfection into the human fibrosarcoma cell line HT1080 was completed using an earlier protocol in which cells were co-transfected with a 1; 10; 50 or 100 pmol miRNA mimic **OR** Allstars negative control. This protocol however, showed significant levels of cell cytotoxicity and was altered for all future experiments (Appendix L).

2.7 Luciferase assay

After a total of 45 hours of transfection, the activity of firefly luciferase by pGL3-CON/TEN and renilla luciferase expressed by pRL-TK was measured independently using the Dual Luciferase Reporter Assay system (Promega Corporation, Madison, Wisconsin, USA) according to the manufacturer's instructions. Cells were washed twice with 1 ml of chilled 1X Phosphate Buffered Saline (PBS) and lysed with 100 µl of 1X Passive Lysis Puffer (PLB) (Promega Corporation, Madison, Wisconsin, USA) for twenty minutes on a shaker at room temperature. Lysates were collected and stored overnight at -80° C, then thawed and briefly vortexed and centrifuged at 13 000 rpm using the Biofuge Fresco bench top centrifuge (Heraeus Instruments, Hanau, Germany) for 30 seconds. Ten µl of supernatant of each sample was plated into a 96-well plate and incubated in a 1X Luciferase Assay Buffer containing 50 µl of firefly luciferase substrate, beetle luciferin. Upon exposure to the substrate, firefly luciferase expressed by pGL-3-TEN/CON DNA constructs will catalyse the beetle luciferin to produce oxyluciferin and light. The emitted light directly correlates with the amount of luciferase present and is used as an indirect measure of mRNA stability. Light emission (luminescence) was quantified with a Luminoskan Ascent luminometer (Thermo Labsystems, Cheshire, UK). Fifty microliters of 1X Stop & Glo Buffer (Promega Corporation, Madison, Wisconsin, USA) was then added to the samples. This reagent contains renilla luciferase substrate which interacts with renilla luciferase to produce luminescence. These readings were analysed with Ascent Software 2.6 (Thermo Labsystems, Cheshire, UK). Firefly luciferase values were normalized to constitutively expressed renilla luciferase values

in each sample. The data obtained from renilla luciferase readings were used to normalize firefly luciferase results.

2.8 Bioinformatic analyses

Several bioinformatics databases including Ensembl, NCBI, Hapmap, dbSNP, 1000genomes and UCSC were investigated prior to initiation of biochemical assays. Highly conserved regions were identified and investigated in greater detail. The *COL5A1* 3'-UTR was annotated with single nucleotide polymorphisms and variable nucleotide tandem repeats using the Hapmap (http://hapmap.ncbi.nlm.nih.gov - Accessed November 1st 2012) dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/) and Ensembl (http://www.ensembl.org) databases. RNA binding protein motifs were identified using the RNA binding protein database RBPDB (http://rbpdb.ccbr.utoronto.ca/ - Accessed November 1st 2012)

2.8.1 Identification of putative miRNA binding sites

Polymorphic miRNA binding sites present within the 3'-UTR were identified using the Patrocles polymorphic miRNA-target database (http://www.patrocles.org). The Patrocles prediction algorithm relies on the evolutionarily conserved heptamer and octamer seed regions to identify potential sites (Appendices M and N). microCOSM Targets, a database curated by the European Bioinformatic Institute (EBI) (a subsidiary of the European Molecular Biology Laboratory - EMBL) was used to identify potential non-polymorphic miRNA binding sites (Appendix O). All chosen miRNAs were investigated using the miRBase database (www.miRBase.org)[101], [102]

2.8.2 Identification of conserved sequences between species

The DNA sequence of the exon 66 of *COL5A1* was aligned to the corresponding region in several animals including mouse, rat, chimpanzee and gorilla using ECR browser (http://ecrbrowser.dcode.org/ Accessed November 1st 2012- Appendix P).

2.8.3 mRNA secondary structures

All secondary structures of the wild-type and mutated C and T functional forms of the *COL5A1* 3'-UTR were generated using the Sfold online RNA folding tool (http://sfold.wadsworth.org) [104], [105]. The Sfold RNA folding algorithm generates RNA secondary structures using a statistical sample from the Boltzmann ensemble of secondary structures. All structures were folded at 37° C and 1 M NaCl in the absence of divalent ions.

2.9 MIR608 genotyping

The participants (143 TEN and 312 CON) were genotyped for the G>C SNP (rs4919510) present in the *MIR608* gene using a custom designed Fluorescence-based Taqman PCR assay (Applied Biosystems, Foster City, CA, USA). Allele-specific probes and flanking primer sets (sequences available on request) were used along with a premade PCR mastermix containing ampliTaq DNA polymerase Gold (Applied Biosystems) in a final reaction volume of 8 µl. The two-step PCR consisted of a 10-min heat activation step (95◦C) followed by 40 cycles of 15 s at 92◦ C and 1 min at 60◦C using the XP Thermal Cycler, Block model XP-G (Bioer Technology Co., Ltd, Tokyo, Japan). End-point fluorescence using a 7900 HT Fast Real-Time PCR System and the SDS Software version 2.3 (Applied Biosystems) was used to determine the genotypes of each polymorphism. Duplicates of three to nine DNA samples were included in each genotyping run as positive controls. The mature miRNA miR608 has the following sequence:

5'-AGG GGT GGT GTT GGG ACA GCT **S**CG T-3', where S is a cytosine or guanine.

2.10 Statistical analyses

All data was analysed with STATISTICA V10.0 (Statsoft Inc., Tulsa, OK, USA) and GraphPad Prism version 5.0d for Mac OS X (GraphPad Software, San Diego, CA, USA, http://www.graphpad.com) programs.

Normally distributed data were analyzed using an independent t-test or a one-way analysis of variance (ANOVA), followed by, if appropriate, a Tukey HSD post-hoc analysis. Data which deviated from normal distribution were analysed using the non-parametric Mann-Whitney test.

A one-way analysis of variance was used to determine any significant differences between the characteristics of the TEN and CON groups within the AUS and SA cohorts. A Chi-squared analysis or Fisher's exact test was used to analyse any differences in the genotype frequencies and other categorical data between the groups. Statistical significant differences were accepted when $P<0.05$. Hardy-Weinberg equilibrium (HWE) was established using the program Genepop web version 3.4 (http://genepop.curtin.edu.au/). Data were presented as mean ± standard deviation or a frequency and results were considered significant at $p<0.05$.
Combined genotype frequencies were also analyzed using the Monte Carlo test (CLUMP program, version 2.0) (Sham & Curtis, 1995). When combined gene-gene interactions and effects were analyzed, statistically significant differences were accepted when $P<0.025$ (Bonferroni correction).

3. Results

3.1 Identification of putative miRNA binding sites

To identify putative miRNA binding sites within the *COL5A1* 3'-UTR, an extensive bioinformatic investigation of 2.5 kb of the wild-type sequence (equivalent to the C-functional form)(Appendix Q) was undertaken [64]. Three major miRNA databases were used in the investigation: (1) the Patrocles database was used to identify polymorphic miRNA binding sites; (2) microCOSM Targets was used to identify potential miRNA binding sites using a less stringent prediction algorithm and (3) miRBase was used to confirm potential miRNA regulators and to identify related miRNAs.

Investigation of the Patrocles database returned three putative heptameric or octameric polymorphic miRNA sites for Hsa-miR886-5p, Hsa-miR-698 and an unknown miRNA (hereafter referred to as the *Mbo*II octamer) (Table 3.1 and Figure 3.1). All 3 sites were identified by their sequence conservation across species and were confirmed by genetic co-expression studies or gene expression analyses.

Table 3.1 Three putative polymorphic miRNA binding sites within the 2.5 kb *COL5A1* 3'-untranslated region (UTR) identified using the Patrocles algorithm

miRNA ID	SNP ID	Description	Target Heptamer/Octamer	Position [a]	Confirmation [b]
Hsa-miR886-5p	rs1134114	C/T	C<u>C</u>GACCCA	327-335	Co-expression
Hsa-miR-608	rs3196378	C/A	CCACCC<u>A</u>	744-751	Co-expression
***Mbo*II RFLP octamer**	rs11103544	T/C	<u>T</u>TTTCTAC	912-919	Gene expression + Hapmap

[a] Nucleotide position within the *COL5A1* 3'-UTR (Appendix Q)
[b] All sites were confirmed by co-expression or gene expression experiments (Appendices M and N). Polymorphic nucleotides are underlined.
RFLP, restriction fragment length polymorphism

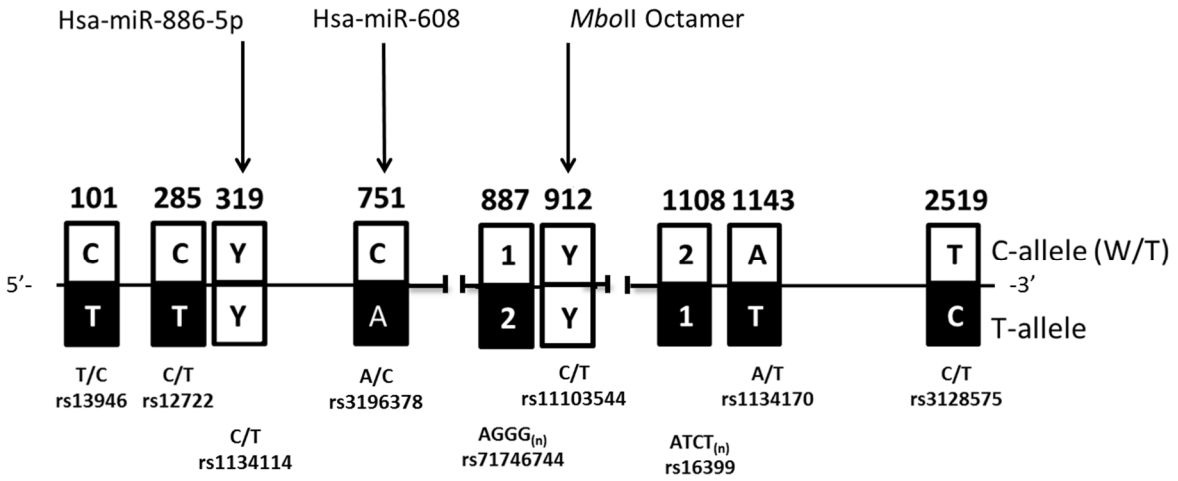


Figure 3.1 Schematic diagram showing positions of the three evolutionarily conserved heptameric or octameric polymorphic miRNA binding sites within the 3'-UTR of *COL5A1*.
The seven polymorphisms which determine the C- (wild-type) and the T-allelic forms of the 3'-UTR are shown in the white and black squares respectively. The positions of the three putative miRNA binding sites are indicated. The Hsa-miR-608 binding site contains single nucleotide polymorphism (SNP) rs3196378. The Hsa-miR-886-5p and MboII octamer binding sites contain SNPs rs1134114 and rs11103544 respectively. The C or T alleles of these two SNPs are associated either form of the 3'-UTR. The nucleotide positions and the rs numbers of each of the polymorphisms are annotated above or below the boxes respectively. Y - C or T; 1 - $(AGGG)_1$ or $(ATCT)_1$; 2 - $(AGGG)_2$ or $(ATCT)_2$.

The microCOSM algorithm identified 33 putative miRNA binding sites within the 3'-UTR of the human *COL5A1* gene (Appendix O). The 10 putative miRNA binding sites with the highest binding probability based on a combination of the miRANDA binding, Poisson binding and Orthologous miRNA binding statistical tests are listed in Table 3.2. The miRANDA binding score is a measure of the binding probability of the miRNA to its target, the Poisson score denotes the probability of a single transcript having multiple miRNA hits while the orthologous miRNA binding score is a measure of the probability of a miRNA family binding to the same conserved site in closely related species from the same orthologous group [106]. Any predicted miRNA binding site which scored a p-value greater than 0.051 for any of the above-stated tests was omitted from further analysis. Except for miR-377* and miR-330-5p, the remaining eight putative miRNA binding sites were located within the 858 bp region of the *COL5A1* 3'-UTR previously associated with chronic Achilles tendinopathy [22].

Table 3.2 The ten putative miRNA binding sites with the greatest binding probability within the 3'-UTR of *COL5A1* as predicted by the MicroCOSM Targets database

Rfam ID	Score	Energy	Base P	Poisson P	Org P	Position [a]
hsa-miR-768-5p	18.261	-29.09	8.08E-03	8.05E-03	8.05E-03	592 to 615
hsa-miR-377*	17.738	-26.18	1.16E-02	1.16E-02	1.16E-02	31 to 52
hsa-miR-125a-5p	17.577	-23.67	1.32E-02	1.32E-02	4.71E-04	346 to 361
hsa-miR-150	17.316	-32.17	2.33E-02	2.30E-02	2.30E-02	357 to 369
hsa-miR-499-5p	17.189	-11.35	2.14E-02	2.11E-02	3.03E-04	577 to 595
hsa-miR-330-5p	17.105	-26.56	4.74E-02	4.63E-02	2.75E-03	176 to 187
hsa-miR-124	17.065	-17.22	1.79E-02	1.77E-02	3.01E-05	537 to 556
hsa-miR-125b	16.999	-24.99	2.19E-02	2.16E-02	4.71E-04	344 to 361
hsa-miR-331-3p	16.871	-25.46	3.93E-02	3.86E-02	3.86E-02	345 to 362
hsa-miR-570	16.788	-17.88	5.05E-02	4.93E-02	1.13E-02	577 to 596

[a] Nucleotide position within the *COL5A1* 3'-UTR (Appendix Q)
Score – calculated using the miRanda algorithm as a measure of base-pair complementarity.
Energy – calculated using Vienna RNA folding routines as a measure of thermodynamic stability.
Base P – a measure of the binding probability calculated using the miRanda score.
Poisson P – a measure of the probability that a single transcript has more than one statistically significant hit according to a Poisson distribution
Org P - an estimation of the probability of the same miRNA family hitting multiple transcripts for different species in an orthologous group
Base P and Poisson P values are calculated using the model proposed by Rehmsmeier *et al.* (2004) [106]

The miRNAs Hsa-miR-608 and the conserved *Mbo*II RFLP octamer were chosen for further investigation (section 3.2) due to the polymorphic nature of their binding sites and their strategic positions within the region of the *COL5A1* 3'-UTR shown to be previously associated with chronic Achilles tendinopathy [13], [14], [22]. The *MIR608* gene also contains a C>G SNP (rs4919510) within the mature miRNA sequence. The association of this polymorphism with Achilles tendinopathy was therefore investigated (section 3.3). In addition to these miRNA binding sites, the Hsa-miR-125a-5p binding site was also chosen due to its proximity to the previously associated rs12722 variant

(C/T) [13], [14] and the commercial availability of its miRNA mimic. Figure 3.2 is schematic diagram showing the positions of the 13 miRNA bindings sites identified in above analyses in relation to the 7 allele differentiating variants.

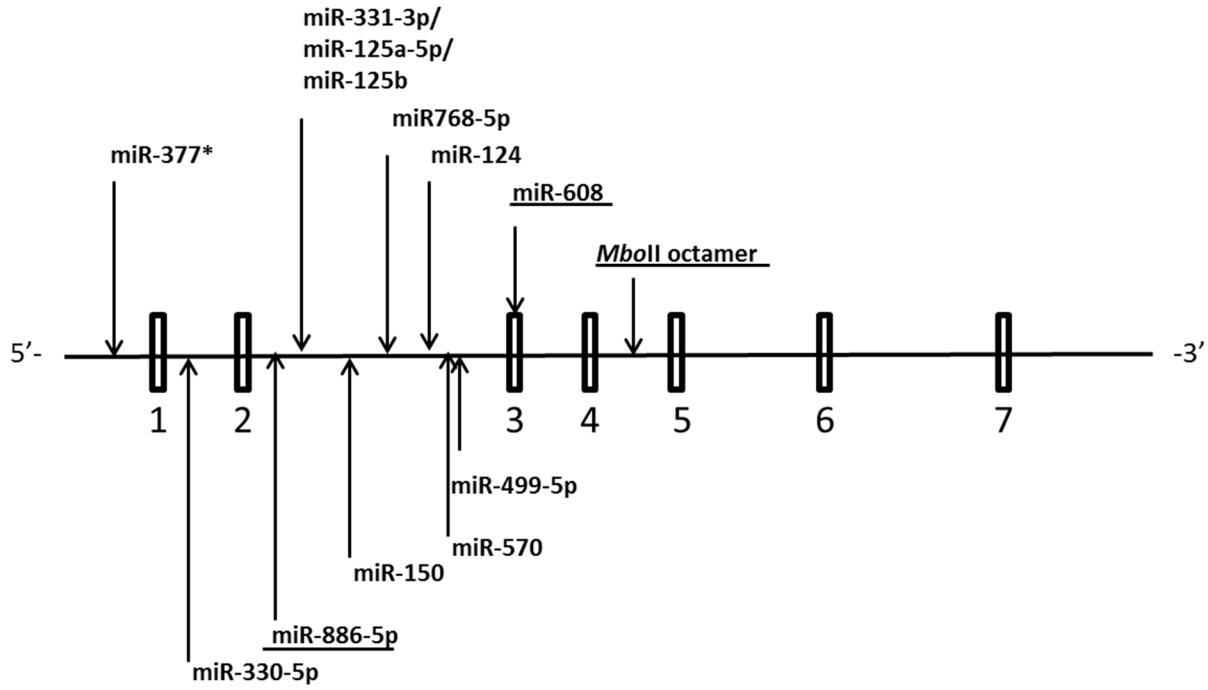


Figure 3.2 Schematic diagram showing positions of the 13 putative miRNA binding sites within the 3'-UTR of *COL5A1* in relation to the seven polymorphisms (numbered rectangles) that determine the C- and T-allelic forms.
The three putative polymorphic miRNA sites predicted by the Patrocles database are underlined. (1) rs13946 (C/T), (2) rs12722 (C/T), (3) rs3196378 (A/C), (4) rs71746744 ($AGGG_{(n)}$), (5) rs16399 ($ATCT_{(n)}$), (6) rs1134170 (A/T) and (7) rs3128575 (C/T)

3.2 miRNA analyses

The next aim of this project was to test whether the miRNA mimics for Hsa-miR-608 and hsa-miR125a-5p were capable of repressing the luciferase expression of the *COL5A1* 3'-UTR constructs and whether these mimics were able to repress the luciferase expression of the C- and T-allelic forms to different degrees [64]. To this end several reporter constructs containing the full length *COL5A1*-3'UTR were transfected with miRNA mimic and assayed for luciferase activity in HT1080 fibrosarcoma and/or SVWI-38 cells. The sequence of the *COL5A1* 3'-UTR within all control (CON 5.1, CON 3.2 and CON2.1) and tendinopathic (TEN 5.1, TEN 3.2, TEN2.2 and TEN 1.1) constructs

used in this study was previously confirmed to correspond to the C- and T-allelic forms respectively [64]. As no miRNA mimic was commercially available for the *Mbo*II octamer, a site-directed mutagenesis approach was undertaken to investigate whether this site was a valid binding site for a negative regulator of the *COL5A1* 3'-UTR.

3.2.1 The co-transfection of Hsa-miR-608 mimic reduces luciferase expression by binding to the polymorphic binding site within in the *COL5A1* 3'-UTR in the HT1080 fibrosarcoma cell line with greater repression observed with the T-allele.

A statistically significant reduction in normalised luciferase activity was noted for both the CON (Clone CON 2.1, Figure 3.3) and TEN (Clone TEN 2.2, Figure 3.3) constructs at all doses (1; 10; 50 and 100 pmol) of miRNA mimic tested. Interestingly, there was no dose-dependent reduction in luciferase activity observed for either CON or TEN constructs.

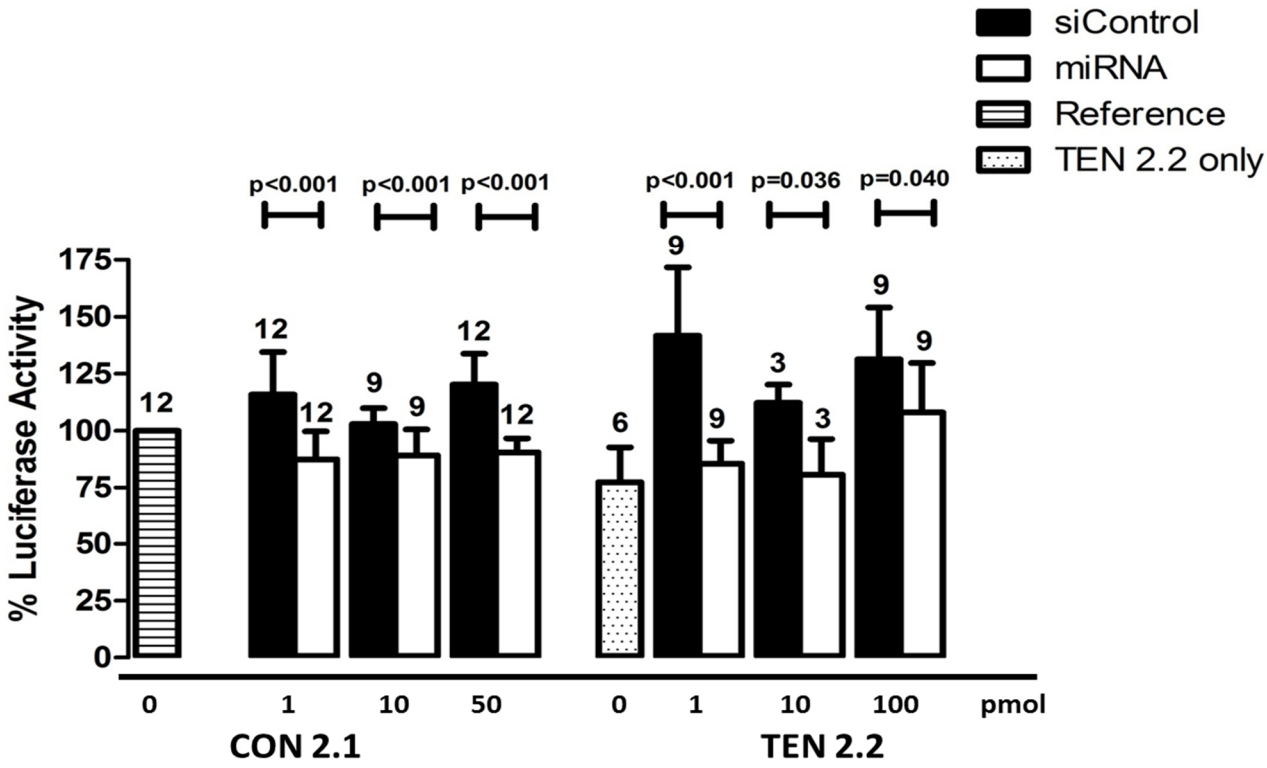


Figure 3.3 Histograms showing normalized luciferase activity for asymptomatic C-allele and symptomatic T-allele constructs with increasing concentrations of Hsa-miR-608 mimic or siControl.
All firefly luciferase activities were normalised to *renilla* luciferase activity, which was included as an internal transfection control. All graphed data points represent the average normalised luciferase activity as a percentage of pGL3-CON 2.1 (reference) normalised luciferase activity. Number of biological replicates (N) is denoted above corresponding bar. Error bars represent standard deviation. Significance was accepted at $p<0.05$.

Two additional clones containing the C- (clones CON 5.1 and CON 3.2) and T-allelic (clones TEN 5.1 and TEN 3.2) forms were assayed with 1 pmol of Hsa-miR-608 mimic or siControl to confirm these findings (Figure 3.4). A significant reduction in normalised luciferase activity was noted when 1 pmol of miRNA mimic was tested with both the C-allelic constructs ($p<0.001$; Figure 3.4 A) and T-allelic constructs (TEN 5.1 - $p<0.001$; TEN 3.2 – $p=0.002$; Figure 3.4 B).

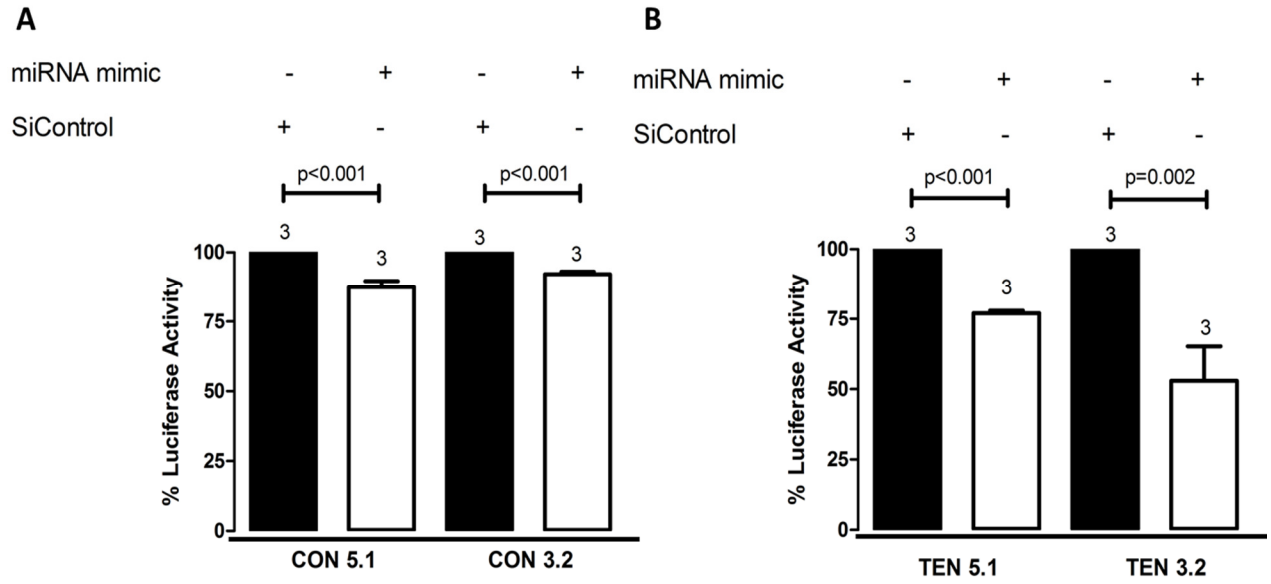


Figure 3.4 Histograms showing normalized luciferase activity for two more asymptomatic C-allele (A) and symptomatic T-allele constructs (B) transfected with 1 pmol of Hsa-miR-608 mimic or siControl.
All firefly luciferase activities were normalised to *renilla* luciferase activity, which was included as an internal transfection control. The normalised luciferase activity of the HT1080 cells transfected with 1pmol miRNA mimic is expressed relative to the normalised luciferase activity of the cells transfected with 1 pmol siControl. Number of biological replicates (N) is denoted above corresponding bar. Error bars represent standard deviation. Significance was accepted at $p<0.05$.

To investigate whether there was a significant difference in Hsa-miR-608's ability to regulate the different allelic forms of the 3'-UTR, the level of reduction in luciferase activity at 1 pmol transfection was pooled for the T-allele constructs (TEN 2.2: 60.8 ± 25.1%, N=3; TEN 3.2: 53.0 ± 12.2%, N=3 and TEN 5.1: 77.1 ± 0.9%, N=3) and C–allele constructs (CON 2.1: 80.1 ± 8.7%, N=3; CON 3.2: 92.1 ± 0.9%, N=3 and CON 5.1: 87.4 ± 2.2%, N=3). To evenly

weight the contribution of each construct, only the highest, lowest and median values of CON 2.1 (N=12) and TEN 2.2 (N=9) at the 1 pmol miRNA mimic transfection were used.

The pooled luciferase activity of the T-allelic constructs was significantly lower (p=0.003) in comparison to the pooled C-allele constructs. There was, on average, a 13.7 ± 7.6% and 36.4 ± 15.6% reduction in luciferase activity for the C- and T-allele constructs respectively after transfection with 1 pmol of miRNA mimic (Figure 3.5). These observations suggest that the co-transfected Hsa-miR-608 mimic represses the T-allelic form of *COL5A1* to a greater extent than the wild type C-allelic form in the HT1080 fibrosarcoma cell line.

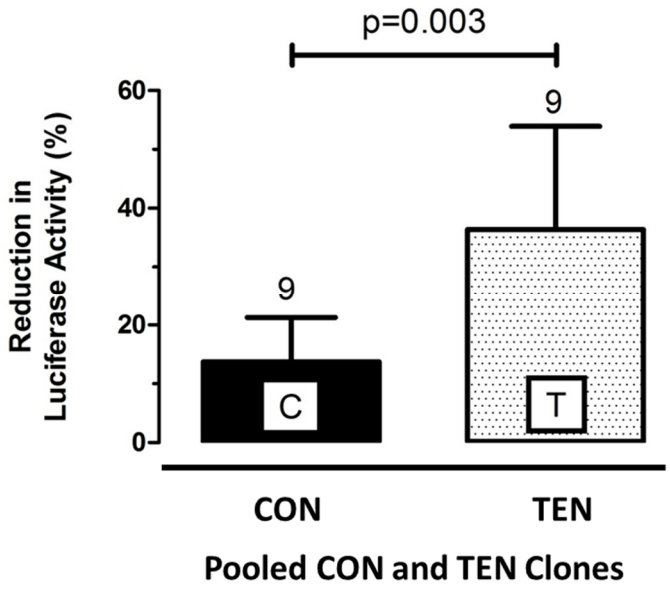


Figure 3.5 Histogram showing averaged luciferase activity of pooled T- and C-allelic constructs co-transfected with 1 pmol of miRNA mimic.
All firefly luciferase activities were normalised to *renilla* luciferase activity, which was included as an internal transfection control. Number of biological replicates (N) is denoted above corresponding bar. Error bars represent standard deviation. Significance was accepted at $p<0.05$.

3.2.2 The addition of Hsa-miR-608 mimic is also capable of repressing luciferase activity in another fibroblast cell line.

The Hsa-miR-608 co-transfection was repeated in a second fibroblast-like cell line, SVWI-38, which was obtained by immortalisation of the normal Wi-38 fibroblasts with the Simian Virus 40 (SV40). To reduce sources of error and the level of cytotoxicity observed during HT1080 transfection, the transfection protocol was modified. The total transfected RNA was maintained at 20 pmol and all luciferase values were normalised with the corresponding 20 pmol siControl luciferase activity (100%).

A significant reduction in luciferase activity was noted for both the C- (Figure 3.6 A, $p<0.001$) and T-allelic (Figure 3.6 B, $p=0.016$) constructs. Specifically, when the SVWI-38 fibroblasts were co-transfected with CON 5.1 and miR-608 mimic, there was a statistically significant reduction in normalised luciferase activity between the 0 pmol (100 ± 0.0%, n=18) and 20 pmol (85.8 ± 14.0%, n=18) miR-608 mimic groups ($p=0.007$), as well as between the 1 pmol (106.2 ± 16.2%, n=18) and 20 pmol groups ($p<0.001$). Similarly within the TEN 1.1 transfected cells, there was a significant decrease in normalised luciferase activity between the 0 pmol (100 ± 0.0%, n=15) and 20 pmol (84.2 ± 17.7%, n=15) miR-608 mimic groups ($p=0.016$).

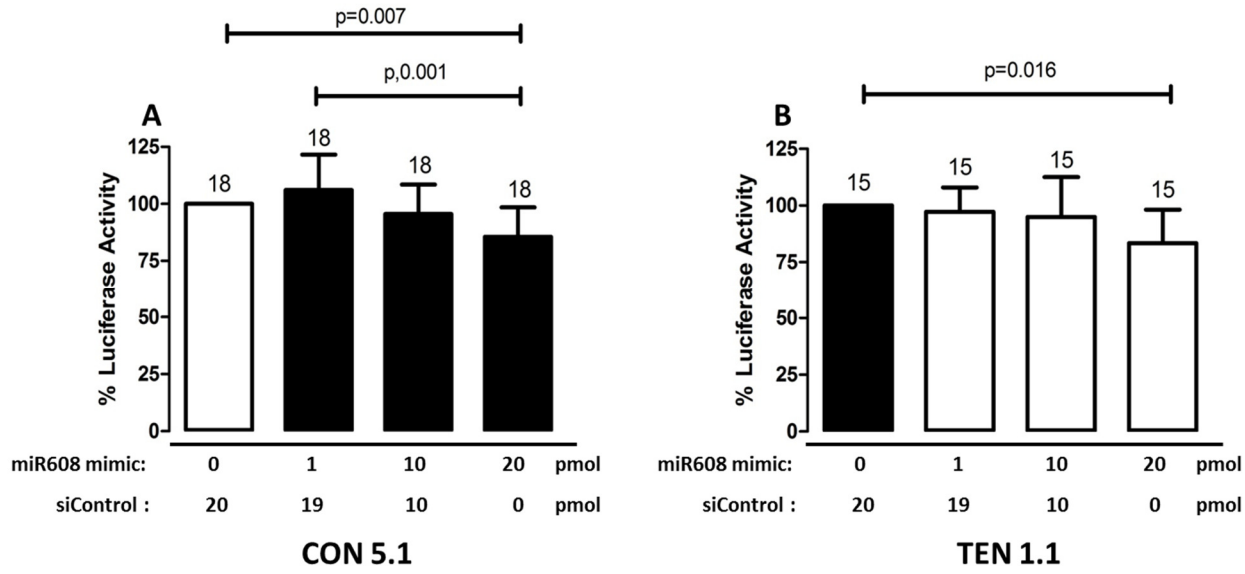


Figure 3.6 Histograms showing averaged luciferase activity of increasing amounts of Hsa-miR-608 mimic co-transfected with (A) C-allelic and (B) T-allelic forms of the *COL5A1* 3'-UTR in SVWI-38 fibroblasts.
All firefly luciferase activities were normalised to *renilla* luciferase activity, which was included as an internal transfection control. Normalised luciferase activities are shown as a percentage of 20 pmol siControl activity (100%). Number of biological replicates (N) is denoted above corresponding bar. Error bars represent standard deviation. Significance was accepted at $p<0.05$.

Since Hsa-miR-608 had a higher affinity for its binding site in the T allelic form of the *COL5A1* 3'-UTR in HT1080 fibrosarcoma cells (Figure 3.5), the relative luciferase activities of the C- and T-allelic constructs were compared at the 1; 10 and 20 pmol Hsa-miR-608 mimic transfections in the SVWI-38 cells. There was however no significant difference in luciferase activity between the allelic forms at the 1 pmol (CON 5.1: 106.2 ± 16.2% vs TEN 1.1: 97.5 ± 12.6%; p= 0.101), 10 pmol (CON 5.1: 95.7 ± 13.7% vs TEN 1.1: 95.4 ± 18.2%; p= 0.962) or 20 pmol level of miRNA mimic (CON 5.1: 85.8 ± 14.0% vs TEN 1.1: 84.2 ± 17.2%; p= 0.763).

3.2.3 The *Mbo*II RFLP conserved octamer is a valid binding site for a negative regulator of the *COL5A1* gene in HT1080 fibrosarcoma cells.

As previously mentioned there was no miRNA mimic commercially available for the MboII octamer and therefore to determine whether this site was functional it was mutated by site-directed mutagenesis and the impact of this investigated in luciferase assays as described earlier. Except for the CON 5.1 mutant clone 1 (Mut 1), there was a significant increase in luciferase activity when CON 5.1 mutant clones 2 and 3, as well as TEN mutant clones 1, 2 and 3 were analysed and compared to their respective parental controls (Figure 3.7 A and B). To determine the anomalous results obtained for CON 5.1 Mut 1 approximately 94% of the *COL5A1* 3'-UTR insert within this construct was sequenced and no spurious mutations were found. As it was the only mutated construct to show a significant reduction in luciferase activity (1 of 6), it was considered an outlier and removed from further analyses.

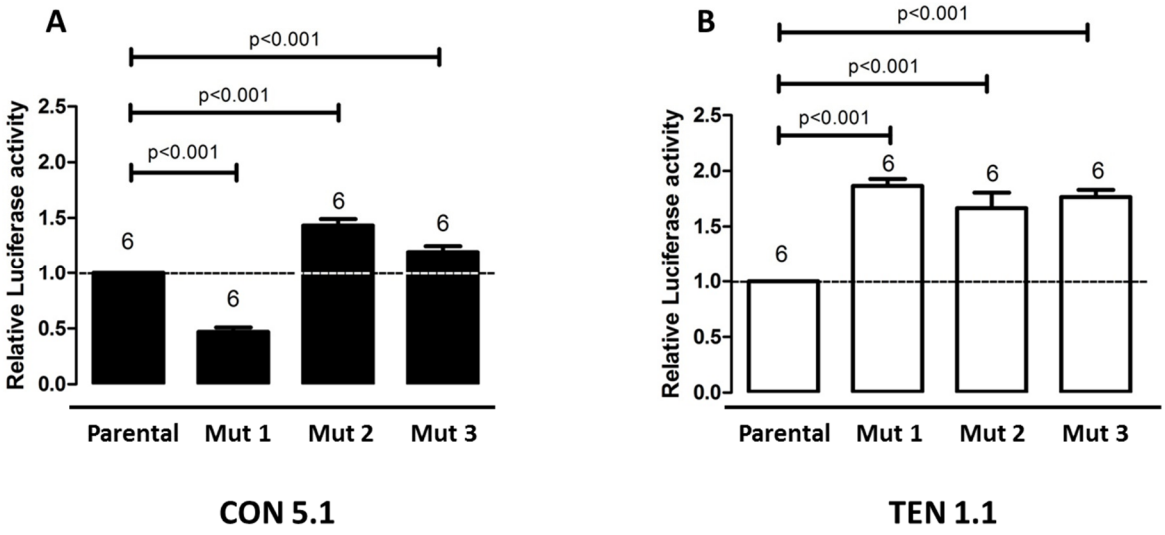


Figure 3.7 Histograms showing normalised luciferase activity of C-allelic (A) and T-allelic (B) constructs with mutated *Mbo*II RFLP octamer sites transfected in the HT1080 fibrosarcoma cells.
All firefly luciferase activities were normalised to *renilla* luciferase activity, which was included as an internal transfection control. Luciferase activity was expressed relative to the normalised luciferase activity of cells transfected with the parental vectors (set to 1.0). Number of biological replicates (N) is denoted above corresponding bar. Error bars represent standard deviation. Significance was accepted at $p<0.05$.

3.2.4 The *Mbo*II RFLP conserved octamer is also a valid binding site in SVWI-38 cells.

To confirm that the results obtained above could be reproduced in another fibroblast cell line, the experiments performed in the previous section were repeated in the SVWI-38 cell line. Indeed, a comparable trend was observed in the HT1080 and SVWi-38 cell lines. Except for the CON 5.1 mutant clone 1, there was a significant increase in luciferase activity when CON 5.1 mutant clones 2 and 3, as well as the TEN mutant clones 1; 2 and 3 were analysed and compared to their respective parental controls (Figure 3.8 A and B). CON 5.1 Mut 1 was omitted from further analyses.

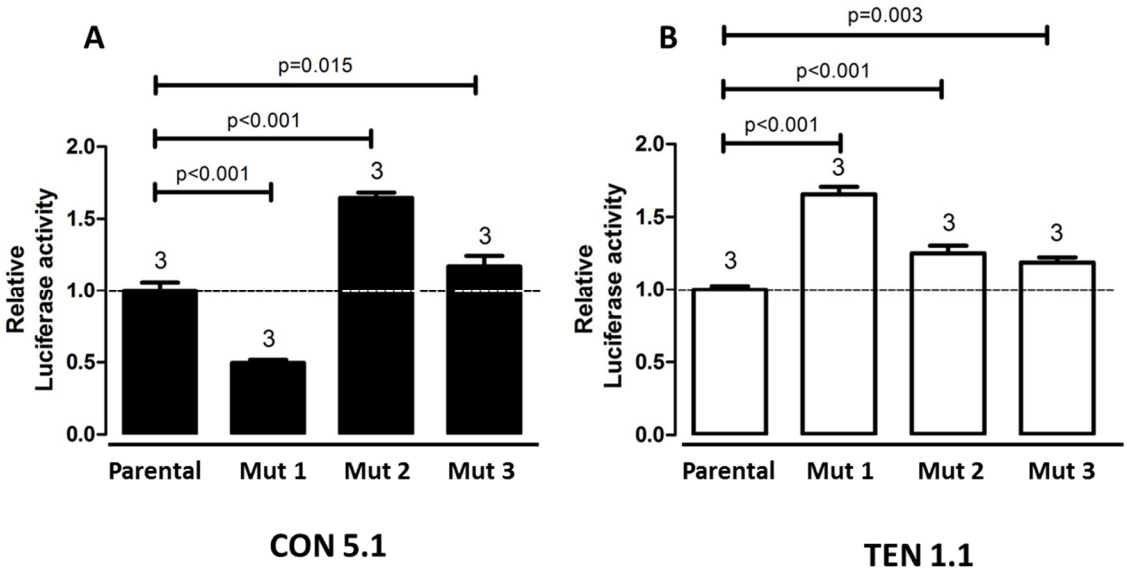


Figure 3.8 Histograms showing normalised luciferase activity of C-allelic (A) and T-allelic (B) constructs with mutated MboII RFLP octamer sites transfected in the SVWI-38 cells.
All firefly luciferase activities were normalised to *renilla* luciferase activity, which was included as an internal transfection. Luciferase activity was expressed relative to the normalised luciferase activity of cells transfected with the parental vectors (set to 1.0). Number of biological replicates (N) is denoted above corresponding bar. Error bars represent standard deviation. Significance was accepted at $p<0.05$.

3.2.5 The co-transfection of miR-125a-5p mimic reduces luciferase expression by binding to the non-polymorphic binding site within the *COL5A1* 3'-UTR in the HT1080 fibrosarcoma cell line with greater repression observed with the T-allele.

To determine whether miR-125a-5p mimic is capable of repressing the activity of the *COL5A1* 3'-UTR, C-allelic and T-allelic constructs were co-transfected with miR-125a-5p mimic in HT1080 cells. A significant reduction in luciferase activity was noted for both the C- (Figure 3.9 A, p=0.005) and T-allelic (Figure 3.9 B, p<0.001) constructs. Specifically, when the HT1080 cells were co-transfected with CON 5.1 and miR-125a-5p mimic, there was a statistically significant reduction in normalised luciferase activity between the 1 pmol (119.3 ± 22.2%, N=6) and 20 pmol (87.7 ± 12.0%, N=6) miR-608 mimic groups (p=0.004), as well as between the 10 pmol (115.8 ± 2.9%, N=3) and 20 pmol groups (p=0.045). Similarly within the TEN 1.1 transfected cells, there was a significant decrease in normalised luciferase activity between the 0 pmol (100 ± 0.0%, n=15) and 20 pmol (65.0 ± 6.4%, N=6) miR-125a-5p mimic groups (p<0.001), as well as between the 1 pmol (121.4 ± 11.5%, N=6) and 20 pmol (p<0.001) and 10 pmol (81.6 ± 4.0%, N=3) and 20 pmol miR-125a-5p mimic groups (p=0.022).

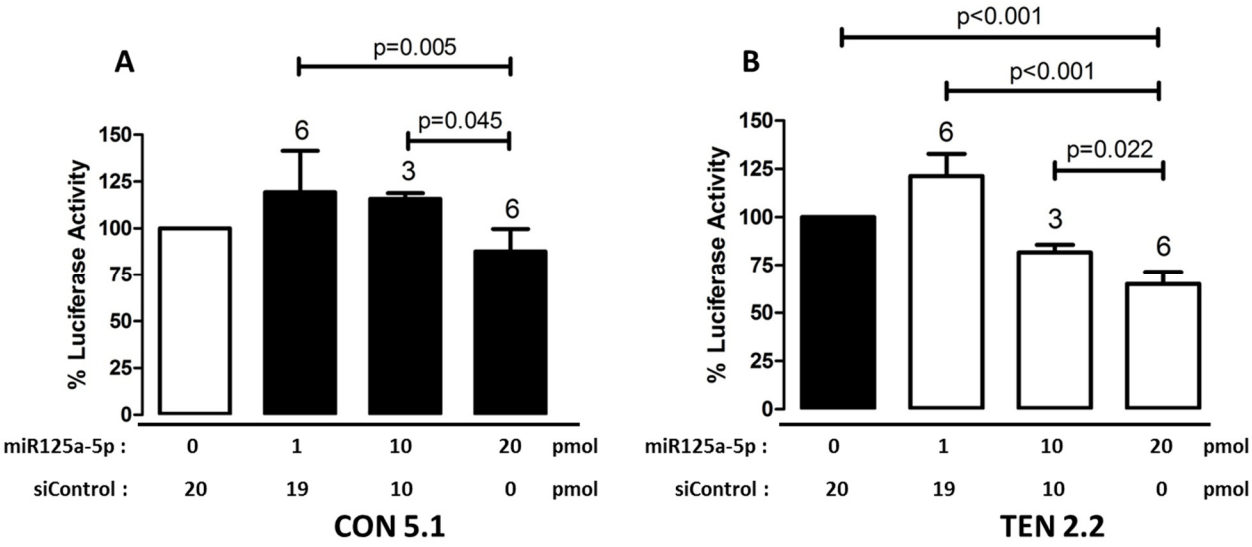


Figure 3.9 Histogram showing averaged luciferase activity of increasing amounts of Hsa-miR-125a-5p mimic co-transfected in HT1080 cells with (A) C- and (B) T-allelic COL5A1 3'-UTR constructs.
All firefly luciferase activities were normalised to *renilla* luciferase activity, which was included as an internal transfection control. Normalised luciferase activities are shown as a percentage of 20 pmol siControl activity (100%). Number of biological replicates (N) is denoted above corresponding bar. Error bars represent standard deviation. Significance was accepted at p<0.05.

The normalised luciferase activities of the C- and T-allelic were compared at 1; 10 and 20 pmol Hsa-miR-125a-5p mimic transfections. There was no significant difference between the allelic forms at the 1 pmol miR-125a-5p mimic transfection (CON 5.1: 119.3 ± 22.2% vs TEN 2.2: 121.4 ± 11.5%; p=0.999). The T-allelic form however, showed a significantly greater reduction in luciferase activity at 10 pmol (CON 5.1: 115.8 ± 3.0% vs TEN 2.2: 81.6 ± 4.0%; p=0.011) and 20 pmol miR-125a-5p mimic transfection (CON 5.1: 87.7 ± 12.0% vs TEN 2.2: 65.0 ± 6.4%; p=0.021)

3.3 The CC genotype of *MIR608* rs4919510 (C/G) is associated with Achilles tendinopathy

As previously reported in section 3.1 the Hsa-miR-608 binding site, which contains a single nucleotide polymorphism (SNP) rs3196378 (AciI RFLP, C/A), is valid. Two forms of the mature Hsa-miR-608, which are produced from the polymorphic *MIR608* gene (SNP rs4919510, C/G) on chromosome 10q24, can potentially bind to this site (http://www.ncbi.nlm.nih.gov). The *MIR608* gene encodes a 25 bp mature miRNA with SNP rs4919510 occurring outside of the target recognition octamer at nucleotide position 22. The aims of this section were therefore to determine (i) whether SNP rs4919510 (C/G) within the *MIR608* gene is associated with chronic Achilles tendinopathy and (ii) to investigate the gene-gene interactions between rs4919510 and the SNP (rs3196378) in its *COL5A1* binding site.

3.3.1 Participant characteristics

3.3.1.1 South African participants

Participants with Achilles tendinopathy (SA TEN) and control (SA CON) participants in the South African population were matched for height, gender and body mass index (BMI) (Table 3.3). The SA TEN group was significantly older (p<0.001) at the age of recruitment with an average age of 48.2 ± 11.3 years (N=71) compared to the SA CON group at 37.0 ± 10.7 years (N=128). The SA TEN participants were however, similarly matched (p=0.191) to the SA CON group at the age of initial injury (Table 3.3). The SA TEN group were significantly heavier (p=0.015) at the time of recruitment, with an average

weight of 77.5 ± 13.7 kg (N=68) compared to 72.9 ± 12.1 kg (N=134) for the SA CON group. When co-varied for age and gender, a significant difference between the SA TEN and SA CON groups was observed (p=0.023).

Table 3.3 Descriptive characteristics of the South African (SA) Achilles tendinopathic (TEN) and control (CON) participants.

	All SA Participants (210)	SA TEN Group (75)	SA CON Group (135)	p-value	Co-varied p-value
Gender (% Male)	67.62 (142)	71.62 (53)	66.42 (89)	0.534	N.D.
Age (yrs) [a]	37.9 ± 14.8 (198)	39.4 ± 10.7 (70)	37.0 ± 10.7 (128)	0.191	N.D.
Weight (kg)	74.4 ± 12.8 (202)	77.5 ± 13.7 (68)	72.9 ± 12.1 (134)	**0.015**	**0.023** [b]
Height (cm)	175.8 ± 9.2 (195)	176.7 ± 8.7 (65)	175.3 ± 9.4 (130)	0.327	N.D.
BMI (kg/m^2)	24.0 ± 3.0 (191)	24.6 ± 3.2 (65)	23.7 ± 2.9 (126)	0.051	N.D.

Except for gender, which is presented as a frequency, values are presented as mean ± standard deviations. The number of participants with non-missing data is in parenthesis.
[a] Age of the TEN participants is at their initial injury, while the age of the CON participants is at recruitment.
[b] co-varied for age and gender.
yrs, years; kg, kilogram; cm, centimetre, N.D. not determined.

3.3.1.2 Australian participants

There was a significantly greater percentage of males (p<0.001) in the Australian (AUS) TEN group (72.1%; N=49) than in the AUS CON group (41.5%; N=73) (Table 3.4). As a result, the AUS TEN group was, on average, heavier (AUS TEN: 80.6 ± 13.9 kg, N=68 vs AUS CON: 73.3 ± 14.6 kg, N=175; p<0.001), taller (AUS TEN: 174.7 ± 10.4 cm, N=65 vs AUS CON: 171.7 ± 9.2 cm, N=174, p=0.036) and had a higher BMI (AUS TEN: 26.4 ± 3.8 kg/m^2, N=65 vs AUS CON: 24.8 ± 4.1 kg/m^2, N=126, p=0.007) than the AUS CON group. When co-varied for age and gender, no significant differences in weight (p=0.358), height (p=0.302) or BMI (p=0.154) were noted. Furthermore, the AUS TEN group was significantly older (48.0 ± 12.9 years, N=68) at the age of recruitment when compared to the AUS CON group (38.2 ± 12.1 years, N=173 p<0.001) but were similarly matched (p=0.703) at the age of initial injury (Table 3.4).

Table 3.4 Descriptive characteristics of the Australian (AUS) Achilles tendinopathic (TEN) and control (CON) participants.

	All AUS Participants (245)	AUS TEN Group (68)	AUS CON Group (177)	p-value	Co-varied p-value
Gender (% Male)	49.80 (122)	72.06 (49)	41.48 (73)	**<0.001**	N.D.
Age (yrs) [a]	38.4 ± 12.6 (240)	38.9 ± 13.9 (67)	38.2 ± 12.1 (173)	0.703	N.D.
Weight (kg)	75.3 ± 14.8 (243)	80.6 ± 13.9 (68)	73.3 ± 14.6 (175)	**<0.001**	0.358 [b]
Height (cm)	172.5 ± 9.6 (239)	174.7 ± 10.4 (65)	171.7 ± 9.2 (174)	**0.036**	0.302 [b]
BMI (kg/m^2)	25.2 ± 4.1 (239)	26.4 ± 3.8 (65)	24.8 ± 4.1 (126)	**0.007**	0.154 [b]

Except for gender, which is presented as a frequency, values are presented as mean ± standard deviations. The number of participants with non-missing data is in parenthesis.
[a] Age of the TEN participants is at their initial injury, while the age of the CON participants is at recruitment.
[b] co-varied for age and gender.
yrs, years; kg, kilogram; cm, centimetre, N.D., not determined.

3.3.1.3 Combined South African and Australian participants

There was a significantly greater percentage of males (p<0.001) in the combined SA+AUS TEN group (71.8%; N=102) than in the combined SA+AUS CON group (52.3%; N=162) (Table 3.5). As a result, the SA+AUS TEN group was heavier (SA+AUS TEN: 79.1 ± 73.1 kg, N=102 vs SA+AUS CON: 73.1 ± 13.6 kg, N=309; p=0.001) and had a higher BMI (SA+AUS TEN: 25.5 ± 3.6 kg/m^2, N=130 vs SA+AUS CON: 24.3 ± 3.7 kg/m^2, N=304, p=0.047) than the SA+AUS CON group. The TEN and CON groups were however, similarly matched for height (unadjusted p=0.056; gender adjusted p=0.724). The SA+AUS TEN group was significantly older (48.1 ± 11.8 years, N=143) at the age of recruitment when compared to the SA+AUS CON group (37.7 ± 11.5 years, N=301 p<0.001), but were similarly matched (p=0.255) to the AUS CON group at the age of initial injury (Table 3.5). When co-varied for gender and age at recruitment a significant difference was noted for weight (p=0.004) but not BMI (p=0.069).

Table 3.5 Descriptive characteristics of the combined South African and Australian Achilles tendinopathic (TEN) and control (CON) participants.

	All Combined Participants (455)	Combined TEN Group (143)	Combined CON Group (312)	p-value	Co-varied p-value
Gender (% Male)	54.07 (246)	71.83 (102)	52.26 (162)	**<0.001**	N.D.
Age (yrs) [a]	38.2 ± 12.5 (438)	39.2 ± 14.4 (137)	37.7 ± 11.5 (301)	0.255	N.D.
Weight (kg)	74.9 ± 13.9 (445)	79.1 ± 13.9 (136)	73.1 ± 13.6 (309)	**0.001**	**0.004 [b]**
Height (cm)	174.0 ± 9.5 (434)	175.7 ± 9.6 (130)	173.3 ± 9.4 (304)	0.056	N.D
BMI (kg/m^2)	24.7 ± 3.7 (430)	25.5 ± 3.6 (130)	24.3 ± 3.7 (304)	**0.047**	0.069 [b]

Except for gender, which is presented as a frequency, values are presented as mean ± standard deviations. The number of participants with non-missing data is in parenthesis.
[a] Age of the TEN participants is at their initial injury, while the age of the CON participants is at recruitment.
[b] co-varied for age and gender.
yrs, years; kg, kilogram; cm, centimetre, N.D., not determined.

3.3.1.4 Genotype effects on physiological variables

There were no significant genotype effects on gender (p=0.541), age (p=0.093), weight (p=0.981), height (p=0.197) or BMI (p=0.604) for rs4919510 in the combined SA+AUS population (Table 3.6). As there was no significant genetic association with each individual population, only genotype effect in the combined population was considered.

Table 3.6 Physiological characteristics of the three MIR608 rs4919510 (G/C) genotype groups of the combined Australian and South African participants.

	CC	CG	GG	p-value
Gender (% Male)	56.36 (155)	61.39 (97)	63.16 (12)	0.541
Age (yrs) [a]	42.0 ± 12.9 (269)	39.4 ± 12.0 (154)	39.1 ± 11.0 (17)	0.093
Weight (kg)	75.0 ± 14.6 (272)	74.8 ± 13.0 (155)	75.3 ± 10.3 (18)	0.981
Height (cm)	173.5 ± 9.4 (264)	174.5 ± 9.8 (152)	177.1 ± 10.0 (18)	0.197
BMI (kg/m^2)	24.8 ± 3.9 (260)	24.6 ± 3.4 (152)	24.0 ± 2.8 (18)	0.604

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

3.3.2 Genotype Distributions

There was no significant difference in genotype frequencies in the genotype groups between the TEN and CON participants when the SA (p=0.186) and AUS (p=0.081) populations were analysed separately (Figure 3.10). Since the genotype frequencies of the TEN and CON groups between the SA and AUS populations were similar, the two populations were combined for further analysis to increase statistical power. As illustrated in Figure 3.11, the CC genotype frequency of the combined SA and AUS TEN groups (TEN CC genotype: 68.5%) was significantly over-represented (CC vs CG+CG, p=0.023, odds ratio (OR)=1.6, 95% confidence interval (CI): 1.1 – 2.5) when compared to the combined CON group (CON CC genotype: 57.1%). The genotype distributions of rs4919510 were in Hardy-Weinberg equilibrium in all groups (SA CON: p=0.815; SA TEN: p=1.000; AUS CON: p= 0.147; AUS TEN: p=0.413; SA+AUS CON: p=0.345 SA+AUS TEN: p=0.815).

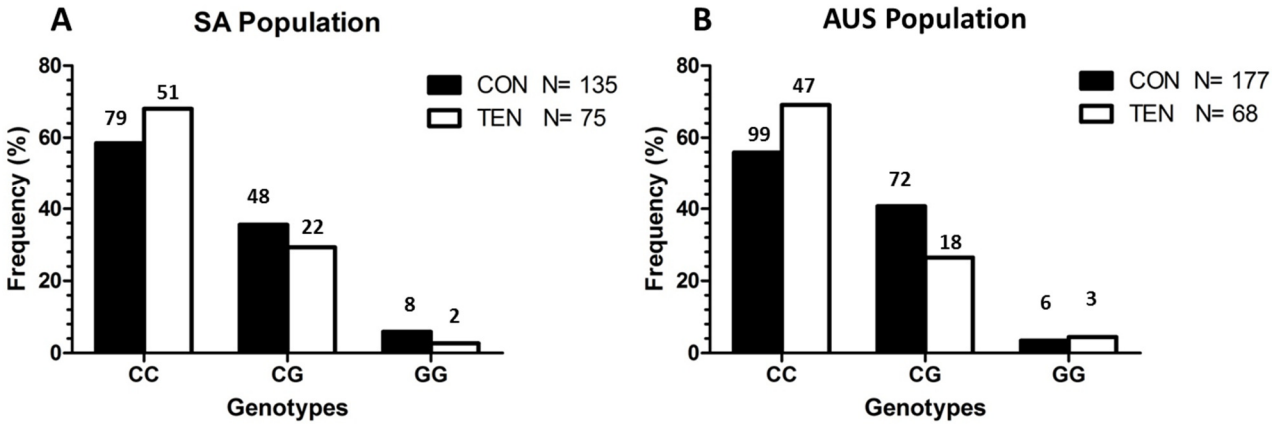


Figure 3.10 Genotype frequency distributions of *MIR608* rs4919510 (C/G) in the (A) South African and (B) Australian populations Achilles tendinopathy (TEN) and asymptomatic control (CON) participants.
The genotype frequency for the CON and TEN groups are denoted by black and white bars respectively. (A) SA: CC vs. G allele, p= 0.186. (B) AUS: CC vs. G allele, p= 0.081. The number of participants (N) in each genotype group is denoted above corresponding bar.

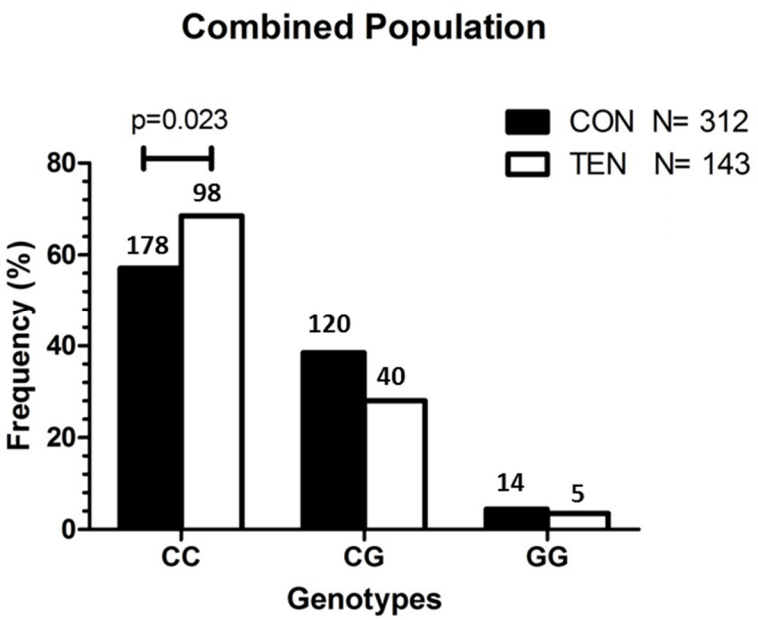


Figure 3.11 Genotype frequency distributions of MIR608 rs4919510 (C/G) in the combined South African and Australian populations Achilles tendinopathy (TEN) and asymptomatic control (CON) groups.
The genotype frequency for the CON and TEN groups are denoted by black and white bars respectively. As indicated, the CC genotype is significantly over-represented in the CON group when compared to the G allele (GC + GG genotype). The number of participants (N) in each genotype group is denoted above corresponding bar. Significance was accepted at $p<0.05$.

3.3.3 Genotype-Genotype Interactions

As previously mentioned, the Hsa-miR-608 binding site within the *COL5A1* 3'-UTR is also polymorphic [14]. The combined genotype frequencies of *MIR608* SNP rs4919510 and its corresponding *COL5A1* binding site SNP rs3196378 (C/A, AciI RFLP) were therefore investigated. The A nucleotide of rs3196378 was identified within the T-functional form of the *COL5A1* 3'-UTR, which was predominately cloned from TEN subjects [64]. There were no significant differences between the global P values (Monte Carlo tests) when the combined rs4919510 and rs3196378 genotype frequencies were compared between the pooled TEN and CON groups (Table 3.7). The combined *MIR608* CC and *COL5A1* rs3196378 CA genotypes were however significantly over-represented in TEN (42.3%) when compared to the CON (30.9%) groups ($P=0.022$, OR=1.6, 95% CI = 1.1 to 2.5). The *MIR608* CC and *COL5A1* rs3196378

AA combined genotype distributions were, however, similar between the TEN (10.9%) and CON (9.5%) groups. This similarity was due to the combined *MIR608* CC and *COL5A1* CC genotypes being under-represented within the AUS TEN cohort, but not the SA TEN cohort (Table 3.7). The combined *MIR608* CC genotype and *COL5A1* rs3196378 A allele (CA and AA genotypes) was significantly over-represented in the TEN group (53.2%, N = 73) when compared to the CON groups (40.4%, N = 111) (P=0.016, OR=1.7, 95% CI = 1.1 to 2.5).

Table 3.7 Combined genotype frequency distributions of the MIR608 gene rs4919510 (C/G) single nucleotide polymorphism (SNP) and the *COL5A1* 3'-UTR SNP rs3196378 (C/A) within the Hsa-miR-608 binding site in control (CON) and chronic Achilles tendinopathy (TEN) groups of South African (SA) and Australian (AUS) cohorts, as well as the combined SA and AUS (SA+AUS) cohorts.

MIR608	***COL5A1* 3'-UTR**	**SA CON** (N=100)	**SA TEN** (N=74)	**AUS CON** (N=175)	**AUS TEN** (N=63)	**SA+AUS CON** (N=275)	**SA+AUS TEN** (N=137)	**TEN/CON**
CC	**CC**	13.0 (13)	12.2 (9)	15.4 (27)	15.9 (10)	14.5 (40)	13.9 (19)	0.96
CC	**CA**	28.0 (28) [a]	40.5 (30) [a]	32.6 (57) [c]	44.4 (28) [c]	30.9 (85) [d,e]	42.3 (58) [d,e]	1.37
CC	**AA**	11.0 (11) [b]	14.9 (11) [b]	8.6 (15)	6.4 (4)	9.5 (26) [e]	10.9 (15) [e]	1.15
CG	**CC**	12.0 (12)	6.8 (5)	8.0 (14)	0.0 (0)	9.5 (26)	3.6 (5)	0.38
CG	**CA**	20.0 (20)	14.9 (11)	25.7 (45)	23.8 (15)	23.6 (65)	19.0 (26)	0.81
CG	**AA**	10.0 (10)	8.1 (6)	6.9 (12)	4.8 (3)	8.0 (22)	6.6 (9)	0.83
GG	**CC**	2.0 (2)	1.4 (1)	0.6 (1)	0.0 (0)	1.1 (3)	0.7 (1)	0.64
GG	**CA**	2.0 (2)	1.4 (1)	1.7 (3)	4.7 (3)	1.8 (5)	2.9 (4)	1.61
GG	**AA**	2.0 (2)	0.0 (0)	0.6 (1)	0.0 (0)	1.1 (3)	0.0 (0)	0.00

Genotype pairs are expressed as percentages with numbers (N) in parenthesis. TEN/CON, SA+AUS TEN/SA+AUS CON.
[a] SA TEN/SA CON = 1.45
[b] SA TEN/SA CON = 1.35
[c] AUS TEN/AUS CON = 1.36
[d] SA+AUS TEN vs SA+AUS CON (*MIR608* CC genotype + *COL5A1* CA genotype), P=0.022, odds ratio = 1.6, 95% confidence interval = 1.1 to 2.5.
[e] SA+AUS TEN vs SA+AUS CON (*MIR608* CC genotype + *COL5A1* A allele), P=0.016, odds ratio = 1.7, 95% confidence interval = 1.1 to 2.5.

3.4 mRNA secondary structure

3.4.1 The secondary structures of the T- and C-allelic forms of the *COL5A1* 3'-UTR differ significantly across an 858 bp region associated with AT

The secondary structures of the T- and C-allelic forms of the *COL5A1* 3'-UTR were generated using the statistical RNA folding algorithm (Sfold) to identify regions of high and low RNA stability. The position of stable regions differed significantly between the two allelic forms (Figure 3.12). The most striking differences exist in an 858 bp region containing five of the seven allele determining variants, namely rs12722; rs3196378; rs71746744; rs16399 and rs11634170 as well as the three putative miRNA binding sites investigated in this study (Inserts A and B). The $AGGG_{(n)}$ short tandem repeat polymorphism (STRP) (rs71746744), which has been associated with AT, appears to be directly involved in the secondary structure of the *Mbo*II octamer binding site (Figure 3.13, top inserts). A significant re-orientation of the positions of unstable regions occur in the region containing the miR-608 binding site, the $AGGG_{(n)}$ STRP and the *Mbo*II octamer (insert A). This region maintains a similar orientation in all 10 of the most stable iterations of the C-allelic form (CCC12AT), but is only present in 2 of the 10 structures (4^{th} and 5^{th} most stable) of the T-allelic form (TTA21TC).

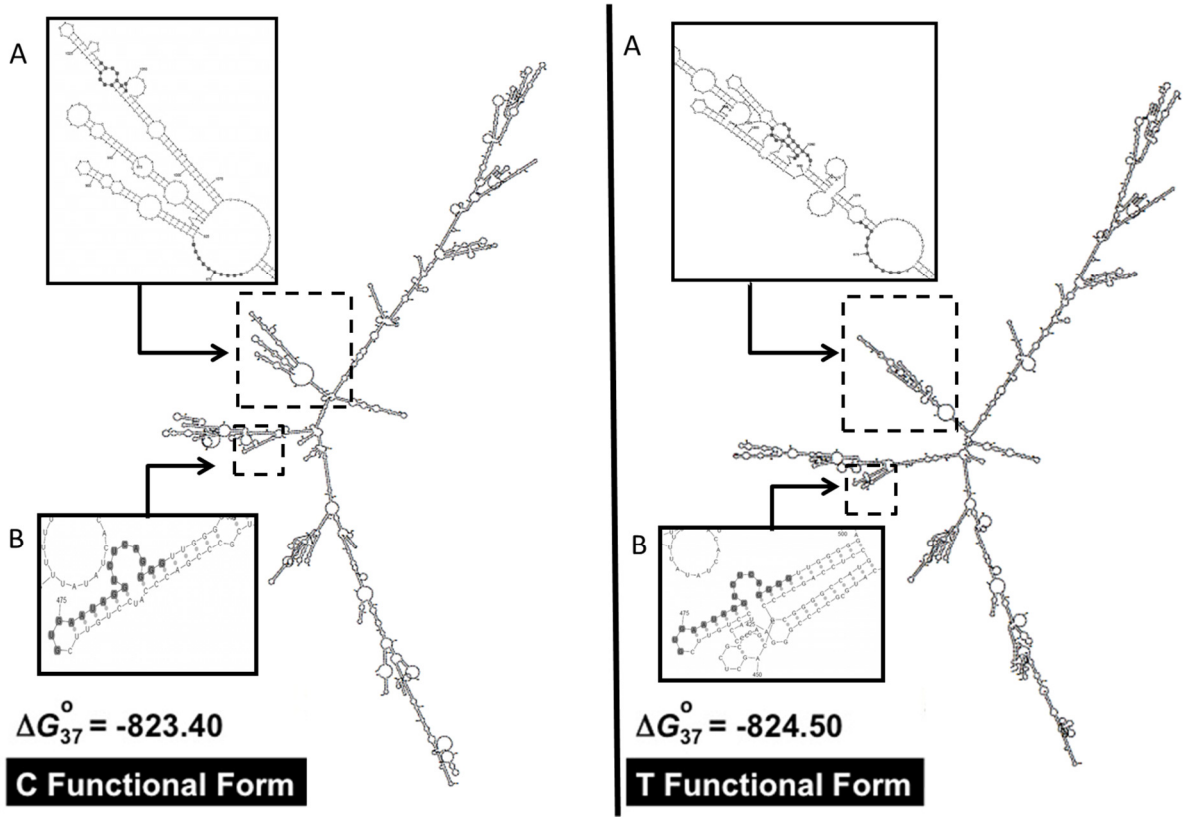


Figure 3.12 The most stable predicted secondary structures of the C (left panel) and T (right panel) functional forms of the COL5A1 3'-UTR.
The region, which contains the putative miR-608 binding site, the *Mbo*II octamer and the AGGG short tandem repeat polymorphism (STRP) (rs71746744), is indicated with box A. Box B indicated the region which contains the predicted binding site for Hsa-miR-125a-5p. Regions A and B of the C (left insert) and T (right insert) functional forms of *COL5A1* 3'-UTR is expanded in the inserts. Predicted binding sites and sequence variations are highlighted in the inserts. Nucleotide positions within the 3'-UTR are also indicated.

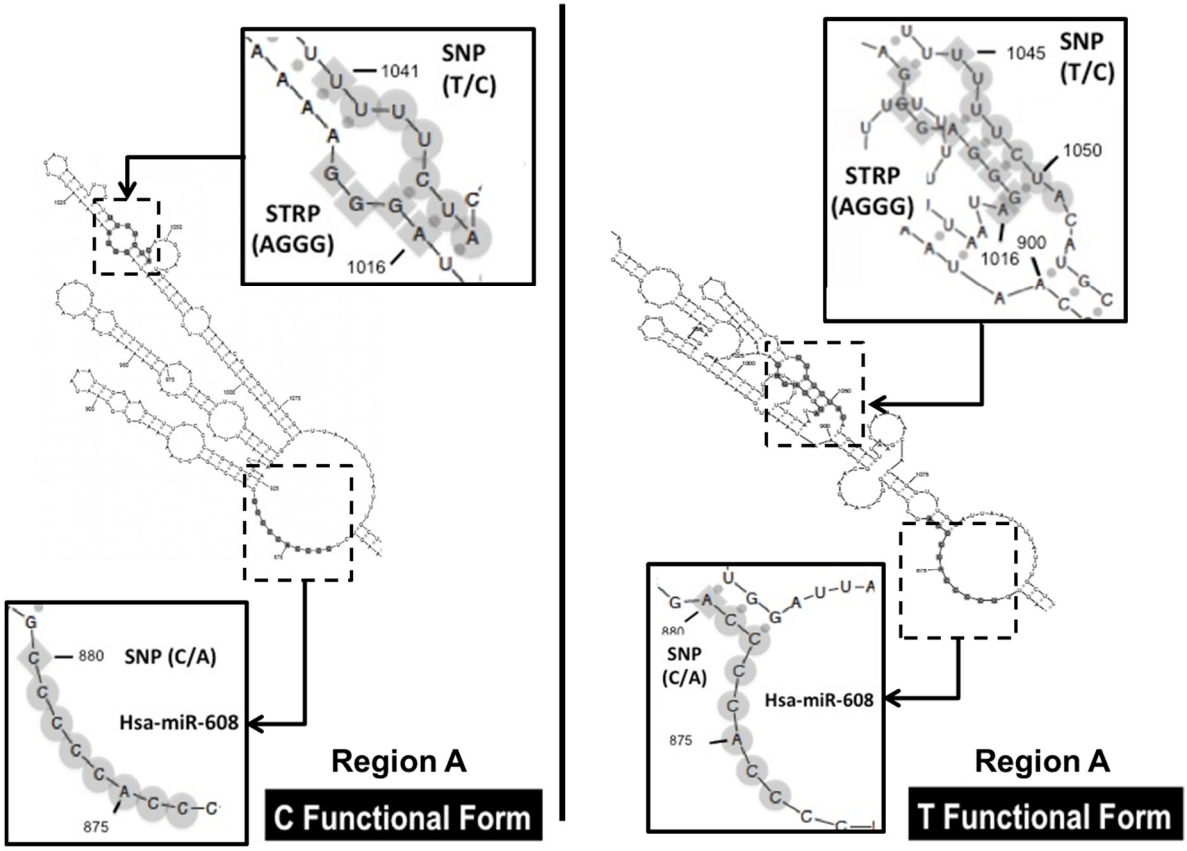


Figure 3.13 The most stable predicted secondary structures of region A (refer to Figure 3.12) of the C (left panel) and T (right panel) functional forms of the COL5A1 3'-UTR.
This region contains both polymorphic miRNA binding sites, the AGGG variable nucleotide tandem repeat (STRP) (rs71746744), single nucleotide polymorphism (SNP) rs11103544 (T/C) and SNP rs3196378 (C/A). The region to which Hsa-miR-608 (bottom inserts) and the *Mbo*II octamer (top inserts) binds are expanded in the boxed inserts. The one and two copies of the AGGG STRP are highlighted with grey diamonds in the top inserts. The miRNA binding sites are highlighted with grey circles. The SNPs within these binding sites are indicated with grey diamonds. Nucleotide positions within the 3'-UTR are also indicated.

When the *COL5A1* 3'-UTR was subjected to *in silico* mutagenesis, analysis of the secondary structure of the region present in insert A produced 3 distinct groups. The first group composed of mutations which produced 7 or more C-form structures predominantly had a C at rs3196378 (6 of 8) and a single copy of the $AGGG_{(n)}$ STRP at rs71746744 (7 of 8). The second group which produced an intermediate number of C forms (5 and 6) had varied combinations of mutations between the T- and C-allelic sequences. The final group which produced 4 or less C form structures were predominantly had an A at rs3196378 (6 of 8), 2 copies of the $AGGG_{(n)}$ STRP at rs71746744 (7 of 8), a single copy of the $ATCT_{(n)}$ at rs16399 (6 of 8) and a T at rs1134170 (6 of 8).

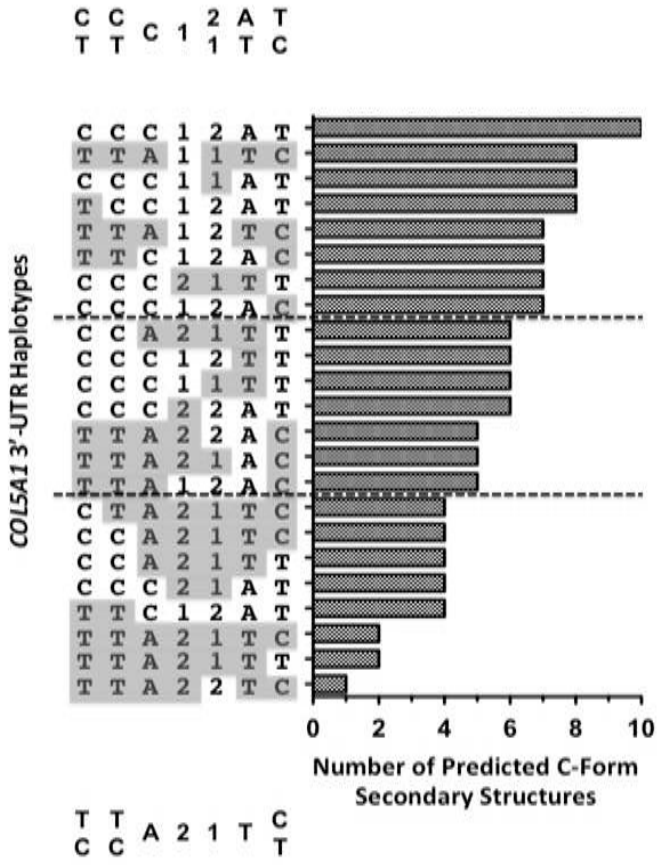


Figure 3.14 The number of predicted secondary structures similar to the C-form for each of the COL5A1 3'-UTR haplotypes generated during in silico site-directed mutagenesis.

The haplotypes consist of the seven polymorphic sites; rs13946 (C/T), rs12722 (C/T), rs3196378 (C/A), rs71746744 (-/AGGG), rs16399 (ATCT/-), rs1134170 (A/T) and rs3128575 (T/C); that determine the distinct C and T functional forms. The nucleotide sequence associated with the C form (C-C-C-1$_{AGGG}$-2$_{ATCT}$-A-T) is highlighted in white, while the sequence associated with the T form (T-T-A-2$_{AGGG}$-1$_{ATCT}$-T-C) is highlighted in grey. The algorithm generates RNA secondary structures using a statistical sample from the Boltzmann ensemble of secondary structures. All structures were folded at 37° C and 1M NaCl in the absence of divalent ions. The ΔG values for the 10 most stable structures are indicated. The secondary structures which are similar to the C functional form of the *COL5A1* 3'-UTR were identified for each haplotype and indicated. The preferred sequences within the *COL5A1* 3'-UTR polymorphisms that produced 7 or more C-form structures (above the top dashed line) is indicated above the histogram, while the preferred sequences within the *COL5A1* 3'-UTR polymorphisms that produced 4 or less C-form structures (below the bottom dashed line) is indicated below the histogram. A single sequence was indicated at a variant only if it was present in 6 or more of the haplotypes.

Of interest was that a single copy of the $AGGG_{(n)}$ STRP (TTA11TC) was sufficient to alter the structure of insert A to the C form in 8 of the 10 most stable structures (Figure 3.14).

Analysis of the secondary structure of the miR-125a-5p binding site (insert B) showed that the hairpin structure and position of the internal bulge was maintained in all 10 of the most stable secondary structures for both the C- and T-allelic form (Figure 3.15).

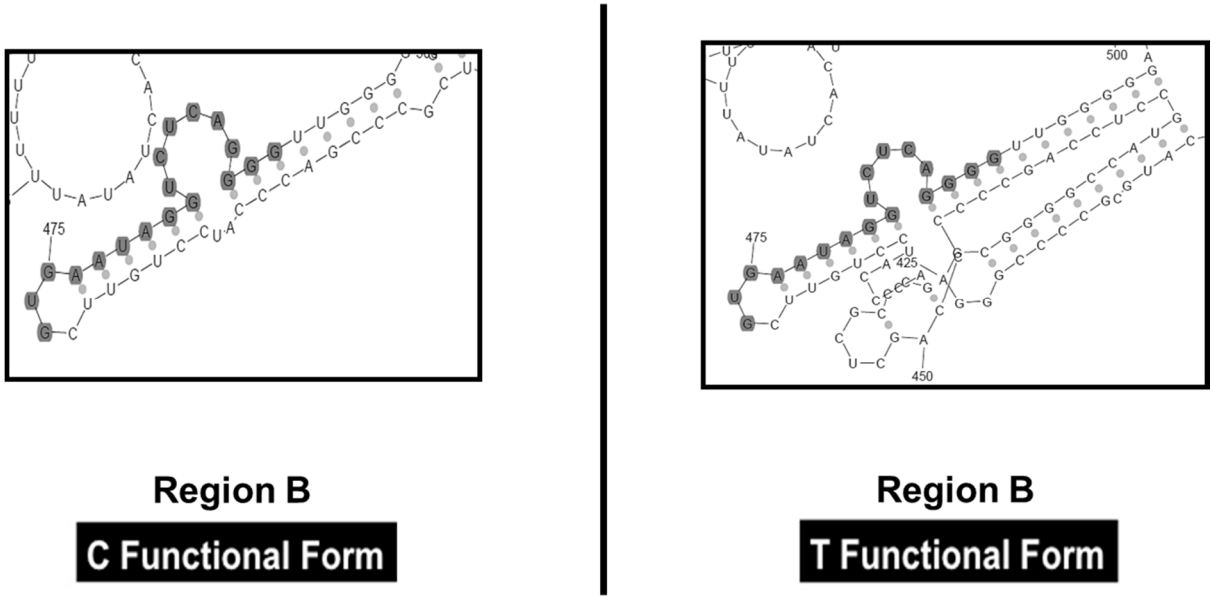


Figure 3.15 The most stable predicted secondary structures of region B (refer to Figure 3.12) of the C (left panel) and T (right panel) functional forms of the *COL5A1* 3'-UTR.
This region contains the putative non-polymorphic binding site for Hsa-miR-125a-5p. The bases to which Hsa-miR-125a-5p binds are highlighted with grey diamonds. The seed binding bases (uGGGGAC) for miR125a-5p are present within an unstable internal bulge region. Nucleotide positions within the 3'-UTR are also indicated.

4. Discussion

A common C/T polymorphism (SNP rs12722) within the *COL5A1* 3'-UTR has previously been associated with several complex exercise-related phenotypes, including chronic Achilles tendinopathy [13], [14], ACL injuries in females [15], range of motion measurements [16], [17], running endurance performance [107], [108] and exercise associated muscle cramps [109]. Three additional polymorphisms within the *COL5A1* 3'-UTR, namely rs71746744 (-/AGGG), rs16399 (ATCT/-) and rs1134170 (T/A), have also been reported to independently associate with chronic Achilles tendinopathy [22]. Another study reported two major functional forms of the *COL5A1* 3'-UTR, namely the C- and T-forms, which were predominantly found in Achilles tendinopathy and asymptomatic control participants, respectively [64]. Importantly, the authors showed that the C- and T-forms exhibited significant differences in their mRNA stability. The aim of this dissertation was therefore to investigate the basis for these differences by exploring the role of putative miRNA binding sites within the 3'-UTR of *COL5A1* on mRNA stability. In particular, a region of the *COL5A1* 3'-UTR previously associated with Achilles tendinopathy, containing three miRNA binding sites was investigated using luciferase activity as an indirect measure of mRNA stability. Briefly, in order to test whether the putative miRNA binding sites were valid negative regulatory elements, the T- and C-allelic 3'-UTRs which were previously cloned downstream of a luciferase reporter gene, were transfected into fibrosarcoma or fibroblast-like cells [64]. This approach was similar to that used by Thiele *et al.* (2004) to indirectly measure the mRNA stability of the fibrillar collagen types I and III in a pGL3-Promoter construct [110]. The first main novel findings of this dissertation were: (i) the co-transfection of Hsa-miR-608 mimic represses luciferase activity of the C-and T-allelic constructs of the 3'-UTR of *COL5A1* and preferentially represses the T-allelic form in the HT1080 fibrosarcoma cell line but not in the SVWI-38 transformed fibroblast line; (ii) the region containing the predicted polymorphic miRNA binding site which overlaps with the *Mbo*II RFLP is also valid and contains a negative regulatory element which represses the *COL5A1* 3'-UTR in both HT1080 and SVWI-38 cell lines; (iii) the co-transfection of Hsa-miR-125a-5p miRNA mimic

represses luciferase activity of both C- and T-allelic constructs of the *COL5A1* 3'-UTR but preferentially represses the T-allele in the HT1080 cell line.

The addition of Hsa-miR-608 mimic was sufficient to cause a significant decrease in the luciferase activity of the C- and T-allelic constructs in the HT1080 cell line. There was however, no dose-dependent reduction in luciferase activity with increasing amounts of transfected Hsa-miR-608 mimic. This may be due to increased cytotoxicity of transfecting large amounts of miRNA mimic which may have negatively affected other protein targets as well as depleting the RNA-induced silencing complexes (RISC) machinery within the cells. Additionally, the cells may become saturated with miRNA mimic, obscuring any dose-dependent features in the data set. To address this, two C-allelic and two T-allelic constructs containing the *COL5A1* 3'-UTR derived from different individuals were tested at the lowest functional concentration of miRNA mimic. It is important to note that while the two C-allelic constructs have the same DNA sequences they have been reported to display minor variations in their luciferase activity and this was also found to be true for the two T-allelic constructs [64]. To prevent skewing of the data, the luciferase activity obtained for the siControl on each construct was set to 100% and the luciferase activity obtained for the miRNA mimic was then calculated as a percentage of their individual siControl before pooling. The results from the current study show that the Hsa-miR-608 mimic repressed the T-allele to a greater extent resulting in a corresponding decreased expression of luciferase. These results suggest that Hsa-miR-608 may preferentially bind the T-allele due to the A nucleotide of rs3196378 within the Hsa-miR-608 binding site or as a result of a change in mRNA secondary structure due to other variants present within the 3'-UTR. This experiment was repeated in the SV40 transformed fibroblast cell line, SVWI-38, using a modified protocol to prevent depletion of the RIS-complexes. Hsa-miR-608 mimic was once again found to be capable of reducing luciferase activity, in a dose-dependent manner. There was however, no preferential repression of the T-allele at any of the tested miRNA mimic concentrations. The reasons for the cell-type specific

preferential repression of the T-allele are currently unknown and require further research.

Investigations into the impact of the *Mbo*II octamer regulatory element revealed that it may play an important role in mRNA stability since mutating this element significantly increased luciferase activity in both HT1080 and SVWI-38 cell lines. An exception was the results obtained for the CON 5.1 Mut 1 construct which showed a statistically significant decrease in luciferase activity. When approximately 94 percent of the 2.5 kb 3'-UTR insert of this construct was sequenced no spurious mutations were identified. As this clone was the only one to show a significant reduction in luciferase activity and mutations in the promoter or luciferase gene regions of the vector could not be ruled out, it was omitted from all further analyses. Interestingly, the *Mbo*II octamer is located within a 57 bp region which, when deleted, was previously shown to abolish the difference in mRNA stability between the T- and C-forms [64]. It was not, however, possible to conclude that the *Mbo*II octamer is the sole regulatory element responsible for the abolishment of the difference in mRNA stability observed in the 57bp deletion constructs. Bioinformatic analyses were therefore performed to identify putative RNA-binding protein motifs present within and around the *Mbo*II octamer as well as in the surrounding 57 bp region that may affect mRNA stability and the results obtained are summarised in Table 4.1.

Table 4.1 RNA-binding protein binding sites within the Δ57 bp region of the *COL5A1* 3'-UTR

RNA binding protein: Gene Symbol	RNA binding protein: Name	Binding element: Description	Relative Score (%)	Start	End
PTBP1/ HNRNP 1	Polypyrimidine tract binding protein 1/ Heterogeneous Ribonucleoprotein polypeptide I	AGACUUUUUU UUUUU	96	996	1010
HNRNPA 1	Heterogeneous Ribonucleoprotein A1	UAGGGA	100	1015	1020
NONO	Non-POU domain containing octamer binding protein	AGGGA	100	1016	1020
EIF4B	Eukaryotic translation initiation factor 4B	GGAA	100	1018	1021
KHDRBS 3	Heterogeneous Ribonucleoprotein G	UUUAAU	93	1010	1015
			95	1029	1034
ELAVL1	Embryonic lethal, abnormal vision, Drosophila-like 1	AUUU	86	1025	1028
			86	1033	1036

The KHDDRBS3, PTBP1 and hnRNPA1 genes encode heterogeneous nuclear ribonucleoproteins which bind to pre-mRNAs in the nucleus. The hnRNPA1 and NONO proteins have previously been shown to act as nuclear-cytoplasmic shuttle proteins, splicing factors and regulators of gene expression [72], [111], [112], [113], [114], [115], [116]. Eukaryotic initiation factor 4B (eIF4B) facilitates the binding of mRNA to the pre-initiation complex and acts in concert with eIF4F to bind to and unwind the secondary structure of the mRNA cap structure in the 5'-UTR. It has also been shown to stimulate the ATPase and RNA helicase activities of eIF4A [117], [118]. Members of the ELAVL protein family have also been found to bind to AU rich elements to destabilize mRNAs and influence gene expression [119], [120], [121]. The regulatory abilities of RNA-binding proteins vary to a great extent and their role in the regulation of the *COL5A1* gene should be investigated in greater detail (Appendix R). It is worth noting that in addition to the above RNA-binding protein motifs, several other putative miRNA binding sites within the *COL5A1* 3'-UTR were identified by the bioinformatic analyses. Since these may be involved in the regulation of the *COL5A1* mRNA stability further

research is needed to test the validity of these sites and their relation to Achilles tendinopathy and other exercise-related phenotypes.

In this dissertation, the co-transfection of Hsa-miR-125a-5p mimic is sufficient to reduce the luciferase expression of the C- and T-allelic constructs in the HT1080 cell line. A dose-dependent reduction in mRNA stability was noted for both the C- and T-allelic constructs, with the T-allelic construct showing statistically greater reduction in mRNA stability at the 10 and 20 pmol concentrations of Hsa-miR-125a-5p mimic. This data is contrary to expectation as the sequence for the miR-125a-5p binding site is identical in both allelic forms. Additionally, the effects of the other variants present within the 3'-UTR of COL5A1 may alter the accessibility of the target site and cause the T-allelic form to be more susceptible to repression by Hsa-miR-125a-5p. It is also possible that an external factor such as RNA-binding proteins are acting on the *COL5A1* 3'-UTR to change local secondary structures and alter its susceptibility to miRNA regulation, independent of the sequence of the miR-125a-5p binding site.

.

In summary, contrary to expectations, at least 2 of the 3 regulatory sites tested in this dissertation showed a preferential reduction in mRNA stability in the more stable T-allelic form of the *COL5A1* 3'-UTR in the HT1080 cell line. These results could not however be reproduced in the SVWI-38 cells since no statistical differences between the two allelic forms of the *COL5A1* 3'-UTR were observed. It is worth noting that while there are obvious limitations to using a non-tendon cell line such as the HT1080, it is an established fibroblast cell line which is generally agreed to represent one of the best models for our initial study. Furthermore, HT1080 cells are commonly used for transfection assays with reporter genes and have previously been used to study the expression of other pGL3 promoter-collagen 3'-UTR constructs [110]. Our research group is, however, planning to confirm the findings of this study in other fibroblast cell lines including primary fibroblasts derived from control and tendinopathic participants. Primary fibroblast cell lines should be established from tendon or other

tissue biopsies from asymptomatic control subjects and patients with chronic Achilles tendinopathy in order to compare *COL5A1* 3'-UTR genotypes and *COL5A1* mRNA and type V collagen protein levels.

Furthermore, this dissertation has identified several other putative miRNA binding sites within the *COL5A1* 3'-UTR which may regulate the mRNA stability of the *COL5A1* mRNA and further research is needed to test the validity of these sites and their relation to Achilles tendinopathy and other exercise-related phenotypes.

The second main novel finding of this dissertation was that the polymorphic *MIR608* gene (rs4919510, C/G) was also associated with chronic Achilles tendinopathy. The CC genotype of this polymorphism was significantly over-represented within the tendinopathic participants. Each allele of the *MIR608* gene encodes for a distinct mature miRNA, Hsa-miR-608, which binds to a polymorphic (rs3196378, *Aci*I RFLP, C/A) *cis*-acting element within the *COL5A1* 3'-UTR [14], [64]. Since the A allele of rs3196378 within the Hsa-miR-608 binding site was identified within the T-functional form of the *COL5A1* 3'-UTR [64], the combined genotype frequencies of *MIR608* rs4919510 and rs3196378 within the miRNA binding site was investigated. It was hypothesized that the combined *MIR608* rs4919510 CC and *COL5A1* miRNA binding site rs3196378 AA genotypes should be over-represented in the chronic Achilles tendinopathy group. Although the combined *MIR608* CC and *COL5A1* rs3196378 AA genotype distributions were similar between the AUS TEN and AUS CON groups, the combined *MIR608* CC genotype and *COL5A1* rs3196378 A allele (CA and AA genotypes) were significantly over-represented in all the TEN participants when compared to all the CON participants. Overall, an odds ratio of 1.6 (95% CI= 1.1 – 2.5) was found for rs4919510 and Achilles tendinopathy, which increased to 1.7 (95% CI= 1.1 – 2.5) when a CA or AA genotype was present at the rs3196378 *COL5A1* miRNA binding locus.

The genetic component of multifactorial conditions such as Achilles tendinopathy are complex, with interactions between several gene families

such as members of the TGF-β superfamily (*GDF5* and *TGF-β*), [28], [122], inflammatory cytokines and receptors (*IL-1β, IL-6* and *IL-1R*) [29], members of the apoptotic pathway (*CASP8*) [94], extracellular matrix degradation pathway (*MMP-3* and *TIMP-2*) [26], [27] and the fibrillar collagen family (*COL5A1*) [13], [14]. This makes elucidating the major etiological mechanisms difficult and requires multi-genic analyses to accurately determine risk.

The third main novel finding of this dissertation was the clear structural differences in the most stable C- and T-functional forms of the *COL5A1* 3'-UTR. The predicted secondary structure of the region containing the miRNA binding sites of interest was distinctly different in these two forms. Sfold analyses revealed that sequence differences within only seven polymorphic sites spanning the entire 2.5 kb *COL5A1* 3'-UTR determined the distinct predicted secondary structures of the C- and T-functional forms. All seven of these polymorphic sites contribute to the predicted structures associated with the C- and T-functional forms although the weight of their contributions varies. The secondary structure of the Hsa-miR125a-5p binding site was maintained in all the investigated structures of both the C- and T-allelic forms. This was of particular interest because the T-allelic form showed significantly greater repression by the Hsa-miR125a-5p mimic *in vitro* (Figures 3.11 and 3.17). In this study, all secondary structures were folded at 37°C and 1 M NaCl as previously described [123]. Divalent cation (Mg^{2+}) concentrations are only considered relevant for the tertiary folding of RNA and were ignored for the purpose of this study [124]. One limitation of this study is the use of an *in silico* prediction of the secondary structures of the 3'-UTR. For example, Sfold which was used in this study, is unable to predict pseudoknots and the accuracy of the prediction also diminishes as the size of the RNA increases [89]. To address this issue, the 858 bp subsection of the mRNA containing the variants rs12722, rs3196378, rs71746744, rs16399 and rs1134170 and the 3 miRNA binding sites was folded and analysed for unstable regions in each allelic form. The seed-binding octamers of all 3 miRNA binding sites were found to be partially or completely within unstable single stranded regions of the mRNA (Appendix S). For both allelic forms, the stability status of the miRNA binding sites was unchanged by removal of the

57 bp region. The algorithms used in the prediction of RNA secondary structures are imperfect and several obstacles must be overcome to accurately predict the structure of RNA *in vivo*. A significant amount of research has however been conducted to accurately determine the energies that stabilize helices and hence *in silico* secondary structure prediction remains a powerful tool [88], [123], [125].

In conclusion, Laguette *et al.* (2011) found that the T-allelic form of the *COL5A1* gene showed an increased mRNA stability relative to the wild-type C-allelic form. Collins and Posthumus (2011) hypothesized that this increase in mRNA stability may alter the relative levels of type V collagen in ligaments and tendons thereby altering the fibril architecture and increasing the risk of injury [63], [64]. This dissertation aimed to investigate whether 3 miRNA binding sites in the *COL5A1* 3'-UTR could contribute to the difference in mRNA stability observed between the two allelic forms. The results however indicate that they do not contribute to increased stability of the T-allelic form of the *COL5A1* 3'-UTR. This dissertation has however, raised interesting questions about the regulation of the *COL5A1* gene at the mRNA level, which may have important implications in the prevention and treatment of Achilles tendinopathy and other musculoskeletal soft tissue injuries.

Bibliography

[1] G. Puddu, E. Ippolito, and F. Postacchini, “A classification of Achilles tendon disease.,” *The American journal of sports medicine*, vol. 4, no. 4, pp. 145–50, 1976.

[2] G. Riley, “The pathogenesis of tendinopathy. A molecular perspective,” *Rheumatology*, vol. 43, no. 2, p. 131, 2004.

[3] J. S. Young, S. M. Kumta, and N. Maffulli, “Achilles tendon rupture and tendinopathy: management of complications.,” *Foot and ankle clinics*, vol. 10, no. 2, pp. 371–82, Jun. 2005.

[4] G. Riley, “Tendinopathy—from basic science to treatment,” *Nature Clinical Practice Rheumatology*, vol. 4, no. 2, pp. 82–89, 2008.

[5] M. F. Mazzone and T. McCue, “Common conditions of the achilles tendon.,” *American family physician*, vol. 65, no. 9, pp. 1805–10, May 2002.

[6] M. Paavola, P. Kannus, T. Jarvinen, K. Khan, and T. Järvinen, “Achilles tendinopathy,” *The Journal of Bone and*, vol. 7, pp. 2062–2076, 2002.

[7] D. Albers and B. Hoke, “Techniques in Achilles tendon rehabilitation,” *Techniques in Foot & Ankle Surgery*, vol. 2, no. 3, pp. 208–219, 2003.

[8] S. Houshian, T. Tscherning, and P. Riegels-Nielsen, “The epidemiology of Achilles tendon rupture in a Danish county.,” *Injury*, vol. 29, no. 9, pp. 651–4, Nov. 1998.

[9] G. W. Hess, “Achilles tendon rupture: a review of etiology, population, anatomy, risk factors, and injury prevention.,” *Foot & ankle specialist*, vol. 3, no. 1, pp. 29–32, Feb. 2010.

[10] M. Magra and N. Maffulli, “Genetic aspects of tendinopathy,” *Journal of Science and Medicine in Sport*, vol. 11, no. 3, pp. 243–247, 2008.

[11] A. V. September, G. Mokone, M.P. Schwellnus, and M. Collins, “Genetic risk factors for Achilles tendon injuries,” *International SportMed Journal*, vol. 7, no. 3, pp. 201–215, 2006.

[12] M. Collins and S. M. Raleigh, “Genetic risk factors for musculoskeletal soft tissue injuries,” *Med Sport Sci*, vol. 54, pp. 136–49, 2009.

[13] G. Mokone, M. Schwellnus, T. Noakes, and M. Collins, “The COL5A1 gene and Achilles tendon pathology,” *Scandinavian journal of medicine & science in sports*, vol. 16, no. 1, pp. 19–26, 2006.

[14] A.V September and J. L. Cook, “Variants within the COL5A1 gene are associated with Achilles tendinopathy in two populations,” *British journal of …*, vol. 43, no. 5, pp. 357–65, May 2009.

[15] M. Posthumus, A. V. September, W. van der Merwe, M. P. Schwellnus, and M. Collins, “The COL5A1 gene is associated with increased risk of anterior cruciate ligament ruptures in female participants,” *The American journal of sports medicine*, vol. 37, no. 11, p. 2234, 2009.

[16] J. C. Brown, C.-J. Miller, M. P. Schwellnus, and M. Collins, “Range of motion measurements diverge with increasing age for COL5A1 genotypes.,” *Scandinavian journal of medicine & science in sports*, vol. 21, no. 6, pp. e266–72, Dec. 2011.

[17] M. Collins, G. Mokone, A.V. September, L. Van Der Merwe, and M.P. Schwellnus, “The COL5A1 genotype is associated with range of motion measurements,” *Scandinavian journal of medicine & science in sports*, vol. 19, no. 6, pp. 803–810, 2009.

[18] U. Fredberg and L. Bolvig, “Significance of ultrasonographically detected asymptomatic tendinosis in the patellar and achilles tendons of elite soccer players: a longitudinal study.,” *The American journal of sports medicine*, vol. 30, no. 4, pp. 488–91, 2002.

[19] N. N. Mahieu, E. Witvrouw, V. Stevens, D. Van Tiggelen, and P. Roget, “Intrinsic risk factors for the development of achilles tendon overuse injury: a prospective study.,” *The American journal of sports medicine*, vol. 34, no. 2, pp. 226–35, Feb. 2006.

[20] P. Van der Linden, J. Van de Lei, H. Nab, A. Knol, and B. H. C. Stricker, “Achilles tendinitis associated with fluoroquinolones,” *British journal of clinical pharmacology*, vol. 48, pp. 433–437, 1999.

[21] F. Malfait, R. J. Wenstrup, and A. De Paepe, “Clinical and genetic aspects of Ehlers-Danlos syndrome, classic type.,” *Genetics in medicine : official journal of the American College of Medical Genetics*, vol. 12, no. 10, pp. 597–605, Oct. 2010.

[22] Y. Abrahams, M.-J. Laguette, S. Prince, and M. Collins, “Polymorphisms within the COL5A1 3’-UTR that alters mRNA structure and the MIR608 gene are associated with Achilles tendinopathy.,” *Annals of human genetics*, vol. 77, no. 3, pp. 204–14, May 2013.

[23] X. Xie, J. Lu, E. Kulbokas, and T. Golub, “Systematic discovery of regulatory motifs in human promoters and 3-UTRs by comparison of several mammals,” *Nature*, vol. 434, no. 7031, pp. 338–345, 2005.

[24] B. Mazumder, V. Seshadri, and P. L. Fox, “Translational control by the 3’-UTR: the ends specify the means,” *Trends in biochemical sciences*, vol. 28, no. 2, pp. 91–98, 2003.

[25] G. G. Mokone, M. Gajjar, A. V. September, M. P. Schwellnus, J. Greenberg, T. D. Noakes, and M. Collins, “The guanine-thymine dinucleotide repeat polymorphism within the tenascin-C gene is associated with Achilles tendon injuries,” *The American Journal of Sports Medicine*, vol. 33, no. 7, p. 1016, 2005.

[26] S. M. Raleigh, L. van der Merwe, W. J. Ribbans, R. K. Smith, M. P. Schwellnus, and M. Collins, “Variants within the MMP3 gene are associated with Achilles tendinopathy: possible interaction with the COL5A1 gene.,” *British journal of sports medicine*, vol. 43, no. 7, pp. 514–20, Jul. 2009.

[27] L. El Khoury, M. Posthumus, M. Collins, C. J. Handley, J. Cook, and S. M. Raleigh, “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 / Sports Medicine Australia*, pp. 10–15, Mar. 2013.

[28] M. Posthumus, M. Collins, J. Cook, C. J. Handley, W. J. Ribbans, R. K. W. Smith, M. P. Schwellnus, and S. M. Raleigh, “Components of the transforming growth factor-beta family and the pathogenesis of human Achilles tendon pathology--a genetic association study.,” *Rheumatology (Oxford, England)*, vol. 49, no. 11, pp. 2090–7, Nov. 2010.

[29] A. V September, E.-M. Nell, K. O’Connell, J. Cook, C. J. Handley, L. van der Merwe, M. Schwellnus, and M. Collins, “A pathway-based approach investigating the genes encoding interleukin-1β, interleukin-6 and the interleukin-1 receptor antagonist provides new insight into the genetic susceptibility of Achilles tendinopathy.,” *British journal of sports medicine*, vol. 45, no. 13, pp. 1040–7, Oct. 2011.

[30] S. Woo and R. Debski, “Injury and repair of ligaments and tendons,” *Annual review of Biomedical Engineering* pp. 83–118, 2000.

[31] K. M. Khan, J. L. Cook, F. Bonar, P. Harcourt, and M. Astrom, “Histopathology of common tendinopathies. Update and implications for clinical management.,” *Sports medicine (Auckland, N.Z.)*, vol. 27, no. 6, pp. 393–408, Jun. 1999.

[32] L. E. Defrate, A. van der Ven, P. J. Boyer, T. J. Gill, and G. Li, “The measurement of the variation in the surface strains of Achilles tendon grafts using imaging techniques.,” *Journal of biomechanics*, vol. 39, no. 3, pp. 399–405, Jan. 2006.

[33] R. Iozzo and A. Murdoch, “Proteoglycans of the extracellular environment: clues from the gene and protein side offer novel perspectives in molecular diversity and function.,” *The FASEB journal*, vol. 10, no. April, 1996.

[34] K. a Derwin, L. J. Soslowsky, J. H. Kimura, and a H. Plaas, “Proteoglycans and glycosaminoglycan fine structure in the mouse tail tendon fascicle.,” *Journal of orthopaedic research : official publication of the Orthopaedic Research Society*, vol. 19, no. 2, pp. 269–77, Mar. 2001.

[35] L. Józsa and P. Kannus, “Histopathological findings in spontaneous tendon ruptures.,” *Scandinavian journal of medicine & science in sports*, vol. 7, no. 2, pp. 113–8, Apr. 1997.

[36] M. Järvinen, L. Józsa, P. Kannus, T. L. Järvinen, M. Kvist, and W. Leadbetter, “Histopathological findings in chronic tendon disorders.,” *Scandinavian journal of medicine & science in sports*, vol. 7, no. 2, pp. 86–95, Apr. 1997.

[37] Y. Liu, H. Ramanath, and D. Wang, “Tendon tissue engineering using scaffold enhancing strategies,” *Trends in biotechnology*, no. March, pp. 201–209, 2008.

[38] K. Gelse, E. Pöschl, and T. Aigner, “Collagens–structure, function, and biosynthesis,” *Advanced Drug Delivery Reviews*, vol. 55, no. 12, pp. 1531–1546, 2003.

[39] M. Van der Rest and R. Garrone, “Collagen family of proteins,” *The FASEB Journal*, vol. 5, no. 13, p. 2814, 1991.

[40] K. E. Kadler, D. F. Holmes, J. a Trotter, and J. a Chapman, “Collagen fibril formation.,” *The Biochemical journal*, vol. 316 (Pt 1, pp. 1–11, May 1996.

[41] K. E. Kadler, C. Baldock, J. Bella, and R. P. Boot-Handford, “Collagens at a glance,” *Journal of Cell Science*, vol. 120, no. 12, p. 1955, 2007.

[42] L. M. Shaw and B. R. Olsen, “FACIT collagens: diverse molecular bridges in extracellular matrices,” *Trends in Biochemical Sciences*, vol. 16, no. null, pp. 191–194, Jan. 1991.

[43] V. Ottani and M. Raspanti, “Collagen structure and functional implications,” *Micron*, vol. 32, pp. 251–260, 2001.

[44] H. A Eriksen, A. Pajala, J. Leppilahti, and J. Risteli, “Increased content of type III collagen at the rupture site of human Achilles tendon.,” *Journal of orthopaedic research : official publication of the Orthopaedic Research Society*, vol. 20, no. 6, pp. 1352–7, Dec. 2002.

[45] N. Maffulli, S. W. Ewen, S. W. Waterston, J. Reaper, and V. Barrass, “Tenocytes from ruptured and tendinopathic achilles tendons produce greater quantities of type III collagen than tenocytes from normal achilles tendons. An in vitro model of human tendon healing.,” *The American journal of sports medicine*, vol. 28, no. 4, pp. 499–505, 2000.

[46] D. E. Birk, J. Fitch, J. Babiarz, K. Doane, and T. Linsenmayer, “Collagen fibrillogenesis in vitro: interaction of types I and V collagen regulates fibril diameter,” *Journal of Cell Science*, vol. 95, no. 4, p. 649, 1990.

[47] R. J. Wenstrup, J. B. Florer, E. W. Brunskill, S. M. Bell, I. Chervoneva, and D. E. Birk, “Type V collagen controls the initiation of collagen fibril assembly.,” *The Journal of biological chemistry*, vol. 279, no. 51, pp. 53331–7, Dec. 2004.

[48] V. Vitart, G. Benčić, C. Hayward, J. \vSkunca Herman, J. Huffman, S. Campbell, K. Bućan, P. Navarro, G. Gunjaca, J. Marin, and others, “New loci associated with central cornea thickness include COL5A1, AKAP13 and AVGR8,” *Human molecular genetics*, vol. 19, no. 21, p. 4304, 2010.

[49] D. A. Parry, G. R. Barnes, and A. S. Craig, “A comparison of the size distribution of collagen fibrils in connective tissues as a function of age and a possible relation between fibril size distribution and mechanical properties.,” *Proceedings of the Royal Society of London. Series B, Containing papers of a Biological character. Royal Society (Great Britain)*, vol. 203, no. 1152, pp. 305–21, Dec. 1978.

[50] P. Lu, G. R. Zhang, X. H. Song, X. H. Zou, L. L. Wang, and H. W. Ouyang, “Col V siRNA engineered tenocytes for tendon tissue engineering.,” *PloS one*, vol. 6, no. 6, p. e21154, Jan. 2011.

[51] E. J. Lawrence, “The clinical presentation of Ehlers-Danlos syndrome.,” *Advances in neonatal care : official journal of the National Association of Neonatal Nurses*, vol. 5, no. 6, pp. 301–14, Dec. 2005.

[52] R. J. Wenstrup, J. B. Florer, M. C. Willing, C. Giunta, B. Steinmann, F. Young, M. Susic, and W. G. Cole, “COL5A1 haploinsufficiency is a common molecular mechanism underlying the classical form of EDS.,” *American journal of human genetics*, vol. 66, no. 6, pp. 1766–76, Jun. 2000.

[53] J. Lincoln, J. B. Florer, G. H. Deutsch, R. J. Wenstrup, and K. E. Yutzey, “ColVa1 and ColXIa1 are required for myocardial morphogenesis and heart valve development.,” *Developmental dynamics : an official publication of the American Association of Anatomists*, vol. 235, no. 12, pp. 3295–305, Dec. 2006.

[54] R. J. Wenstrup, S. M. Smith, J. B. Florer, G. Zhang, D. P. Beason, R. E. Seegmiller, L. J. Soslowsky, and D. E. Birk, “Regulation of collagen fibril nucleation and initial fibril assembly involves coordinate interactions with collagens V and XI in developing tendon.,” *The Journal of biological chemistry*, vol. 286, no. 23, pp. 20455–65, Jun. 2011.

[55] A. Fichard, J. Kleman, and F. Ruggiero, “Another look at collagen V and XI molecules,” *Matrix biology*, vol. 14, pp. 515–531, 1995.

[56] J. M. Archambault, J. P. Wiley, and R. C. Bray, “Exercise loading of tendons and the development of overuse injuries. A review of current literature.,” *Sports medicine (Auckland, N.Z.)*, vol. 20, no. 2, pp. 77–89, Aug. 1995.

[57] J. L. Cook and C. Purdam, “Is compressive load a factor in the development of tendinopathy?,” *British journal of sports medicine*, vol. 46, no. 3, pp. 163–8, Mar. 2012.

[58] T. Pufe, W. J. Petersen, R. Mentlein, and B. N. Tillmann, “The role of vasculature and angiogenesis for the pathogenesis of degenerative tendons disease.,” *Scandinavian journal of medicine & science in sports*, vol. 15, no. 4, pp. 211–22, Aug. 2005.

[59] M. Abate, K. Gravare-Silbernagel, C. Siljeholm, A. Di Iorio, D. De Amicis, V. Salini, S. Werner, R. Paganelli, and K. G. Silbernagel, “Pathogenesis of tendinopathies: inflammation or degeneration,” *Arthritis Res Ther*, vol. 11, no. 3, p. 235, Jan. 2009.

[60] U. Fredberg and K. Stengaard-Pedersen, “Chronic tendinopathy tissue pathology, pain mechanisms, and etiology with a special focus on inflammation.,” *Scandinavian journal of medicine & science in sports*, vol. 18, no. 1, pp. 3–15, Feb. 2008.

[61] J. Zhang and J. H. C. Wang, “Production of PGE2 increases in tendons subjected to repetitive mechanical loading and induces differentiation of tendon stem cells into non-tenocytes,” *Journal of Orthopaedic Research*, vol. 28, no. 2, pp. 198–203, 2010.

[62] J. L. Cook and C. R. Purdam, “Is tendon pathology a continuum? A pathology model to explain the clinical presentation of load-induced tendinopathy.,” *British journal of sports medicine*, vol. 43, no. 6, pp. 409–16, Jun. 2009.

[63] M. Collins and M. Posthumus, “Type V collagen genotype and exercise-related phenotype relationships: a novel hypothesis.,” *Exercise and sport sciences reviews*, vol. 39, no. 4, pp. 191–8, Oct. 2011.

[64] M.-J. Laguette, Y. Abrahams, S. Prince, and M. Collins, “Sequence variants within the 3’-UTR of the COL5A1 gene alters mRNA stability: implications for musculoskeletal soft tissue injuries.,” *Matrix biology : journal of the International Society for Matrix Biology*, vol. 30, no. 5–6, pp. 338–45, Jun. 2011.

[65] L. Wu, J. Fan, and J. G. Belasco, “MicroRNAs direct rapid deadenylation of mRNA,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 11, p. 4034, 2006.

[66] M. R. Fabian and N. Sonenberg, “The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC.,” *Nature structural & molecular biology*, vol. 19, no. 6, pp. 586–93, Jun. 2012.

[67] Q. Jing, S. Huang, S. Guth, T. Zarubin, A. Motoyama, J. Chen, F. Di Padova, S.-C. Lin, H. Gram, and J. Han, “Involvement of microRNA in AU-rich element-mediated mRNA instability.,” *Cell*, vol. 120, no. 5, pp. 623–34, Mar. 2005.

[68] W. Hu and J. Coller, “What comes first: translational repression or mRNA degradation? The deepening mystery of microRNA function.,” *Cell research*, vol. 22, no. 9, pp. 1322–4, Sep. 2012.

[69] D. F. Colgan and J. L. Manley, “Mechanism and regulation of mRNA polyadenylation,” *Genes & Development*, vol. 11, no. 21, pp. 2755–2766, Nov. 1997.

[70] C. Beelman and R. Parker, “Degradation of mRNA in eukaryotes,” *Cell*, vol. 81, pp. 179–183, 1995.

[71] A. Grishok, A. E. Pasquinelli, D. Conte, N. Li, S. Parrish, I. Ha, D. L. Baillie, A. Fire, G. Ruvkun, and C. C. Mello, “Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing,” *Cell*, vol. 106, no. 1, pp. 23–34, 2001.

[72] D. Katz and N. Theodorakis, “AU-A, an RNA-binding activity distinct from hnRNP A1, is selective for AUUUA repeats and shuttles between the nucleus and the cytoplasm,” *Nucleic acids …*, vol. 22, no. 2, pp. 238–46, Jan. 1994.

[73] S. Francia, F. Michelini, A. Saxena, D. Tang, M. de Hoon, V. Anelli, M. Mione, P. Carninci, and F. d’Adda di Fagagna, “Site-specific DICER and DROSHA RNA products control the DNA-damage response.,” *Nature*, vol. 488, no. 7410, pp. 231–5, Aug. 2012.

[74] R. W. Walters, S. S. Bradrick, and M. Gromeier, “Poly (A)-binding protein modulates mRNA susceptibility to cap-dependent miRNA-mediated repression,” *RNA*, vol. 16, no. 1, p. 239, 2010.

[75] K. Okamura, A. Ishizuka, and H. Siomi, “Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways,” *Genes &*, pp. 1655–1666, 2004.

[76] A. J. Pollard, A. R. Krainer, S. C. Robson, and G. N. Europe-Finner, “Alternative splicing of the adenylyl cyclase stimulatory G-protein G alpha(s) is regulated by SF2/ASF and heterogeneous nuclear ribonucleoprotein A1 (hnRNPA1) and involves the use of an unusual TG 3’-splice Site.,” *The Journal of biological chemistry*, vol. 277, no. 18, pp. 15241–51, May 2002.

[77] M. Niepmann, “Hepatitis C virus RNA translation.,” *Current topics in microbiology and immunology*, vol. 369, pp. 143–66, Jan. 2013.

[78] I. M. Silverman, F. Li, and B. D. Gregory, “Genomic era analyses of RNA secondary structure and RNA-binding proteins reveal their significance to post-transcriptional regulation in plants,” *Plant Science*, vol. 205–206, pp. 55–62, May 2013.

[79] D. P. Bartel, “MicroRNAs:: Genomics, Biogenesis, Mechanism, and Function,” *Cell*, vol. 116, no. 2, pp. 281–297, 2004.

[80] W.-Y. Choi, A. J. Giraldez, and A. F. Schier, “Target protectors reveal dampening and balancing of Nodal agonist and antagonist by miR-430.,” *Science (New York, N.Y.)*, vol. 318, no. 5848, pp. 271–4, Oct. 2007.

[81] L. P. Lim, N. C. Lau, P. Garrett-Engele, A. Grimson, J. M. Schelter, J. Castle, D. P. Bartel, P. S. Linsley, and J. M. Johnson, “Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs,” *Mol. Cell. Biol*, vol. 19, no. 1991, pp. 6379–6395, 1999.

[82] K. A. O’Donnell, E. A. Wentzel, K. I. Zeller, C. V. Dang, and J. T. Mendell, “c-Myc-regulated microRNAs modulate E2F1 expression,” *Nature*, vol. 435, no. 7043, pp. 839–843, 2005.

[83] G. Jannot, M. Boisvert, I. Banville, and M. Simard, “Two molecular features contribute to the Argonaute specificity for the microRNA and RNAi pathways in C. elegans,” *Rna*, pp. 829–835, 2008.

[84] B. Zhang, Q. Wang, and X. Pan, “MicroRNAs and their regulatory roles in animals and plants,” *Journal of cellular physiology*, vol. 210, no. 2, pp. 279–289, 2007.

[85] J. Doench and P. Sharp, “Specificity of microRNA target selection in translational repression,” *Genes & development*, vol. 18, no. 5, pp. 504–511, Mar. 2004.

[86] S. W. Chi, G. J. Hannon, and R. B. Darnell, “An alternative mode of microRNA target recognition.,” *Nature structural & molecular biology*, vol. 19, no. 3, pp. 321–7, Mar. 2012.

[87] W. Zhang and S.-J. Chen, “RNA hairpin-folding kinetics.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 4, pp. 1931–6, Feb. 2002.

[88] M. Zuker, D. Mathews, and D. Turner, “Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide,” *RNA biochemistry and biotechnology*, 1999.

[89] C. Laing and T. Schlick, “Computational approaches to 3D modeling of RNA.,” *Journal of physics. Condensed matter : an Institute of Physics journal*, vol. 22, no. 28, p. 283101, Jul. 2010.

[90] D. Long, R. Lee, P. Williams, C. Y. Chan, V. Ambros, and Y. Ding, “Potent effect of target structure on microRNA function.,” *Nature structural & molecular biology*, vol. 14, no. 4, pp. 287–94, Apr. 2007.

[91] Y. Zhao, E. Samal, D. Srivastava, and others, “Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis,” *Nature*, vol. 436, no. 7048, pp. 214–220, 2005.

[92] D. E. Draper, “Themes in RNA-protein recognition.,” *Journal of molecular biology*, vol. 293, no. 2, pp. 255–70, Oct. 1999.

[93] E. A. Dethoff, J. Chugh, A. M. Mustoe, and H. M. Al-Hashimi, “Functional complexity and regulation through RNA dynamics.,” *Nature*, vol. 482, no. 7385, pp. 322–30, Feb. 2012.

[94] E.-M. Nell, L. van der Merwe, J. Cook, C. J. Handley, M. Collins, and A. V September, “The apoptosis pathway and the genetic predisposition to Achilles tendinopathy.,” *Journal of orthopaedic research : official publication of the Orthopaedic Research Society*, vol. 30, no. 11, pp. 1719–24, Nov. 2012.

[95] S. Hiard, C. Charlier, W. Coppieters, M. Georges, and D. Baurain, “Patrocles: a database of polymorphic miRNA-mediated gene regulation in vertebrates.,” *Nucleic acids research*, vol. 38, no. Database issue, pp. D640–51, Jan. 2010.

[96] B. P. Lewis, C. B. Burge, and D. P. Bartel, “Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets.,” *Cell*, vol. 120, no. 1, pp. 15–20, Jan. 2005.

[97] M. R. Dressler, D. L. Butler, R. Wenstrup, H. a Awad, F. Smith, and G. P. Boivin, “A potential mechanism for age-related declines in patellar tendon biomechanics.,” *Journal of orthopaedic research : official publication of the Orthopaedic Research Society*, vol. 20, no. 6, pp. 1315–22, Nov. 2002.

[98] M. Paavola, P. Kannus, T. A. H. Järvinen, T. L. N. Järvinen, L. Józsa, and M. Järvinen, “Treatment of tendon disorders. Is there a role for corticosteroid injection?,” *Foot and ankle clinics*, vol. 7, no. 3, pp. 501–13, Sep. 2002.

[99] D. K. Lahiri and J. I. Nurnberger, “A rapid non-enzymatic method for the preparation of HMW DNA from blood for RFLP studies.,” *Nucleic acids research*, vol. 19, no. 19, p. 444, Oct. 1991.

[100] G. Wang, J. M. Van Der Walt, G. Mayhew, Y. J. Li, S. Zuchner, W. K. Scott, E. R. Martin, and J. M. Vance, “Variation in the miRNA-433 Binding Site of FGF20 Confers Risk for Parkinson Disease by Overexpression of [alpha]-Synuclein,” *The American Journal of Human Genetics*, vol. 82, no. 2, pp. 283–289, 2008.

[101] S. Griffiths-Jones, “The microRNA Registry.,” *Nucleic acids research*, vol. 32, no. Database issue, pp. D109–11, Jan. 2004.

[102] S. Griffiths-Jones, “miRBase: microRNA sequences and annotation.,” *Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.]*, vol. Chapter 12, p. Unit 12.9.1–10, Mar. 2010.

[103] J. de Haan, W. Gevers, and M. Parker, “Effects of Sodium Butyrate on the Synthesis and Methylation of DNA in Normal Cells and Their Transformed Counterparts,” *Cancer research*, pp. 713–716, 1986.

[104] Y. Ding, C. Y. Chan, and C. E. Lawrence, “RNA secondary structure prediction by centroids in a Boltzmann weighted ensemble.,” *RNA (New York, N.Y.)*, vol. 11, no. 8, pp. 1157–66, Aug. 2005.

[105] Y. Ding, C. Y. Chan, and C. E. Lawrence, “Sfold web server for statistical folding and rational design of nucleic acids.,” *Nucleic acids research*, vol. 32, no. Web Server issue, pp. W135–41, Jul. 2004.

[106] M. Rehmsmeier, P. Steffen, M. Höchsmann, and R. Giegerich, “Fast and effective prediction of microRNA/target duplexes,” *Rna*, no. 2003, pp. 1507–1517, 2004.

[107] M. Posthumus, M. P. Schwellnus, and M. Collins, “The COL5A1 gene: a novel marker of endurance running performance.,” *Medicine and science in sports and exercise*, vol. 43, no. 4, pp. 584–9, Apr. 2011.

[108] J. C. Brown, C.-J. Miller, M. Posthumus, M. P. Schwellnus, and M. Collins, “The COL5A1 gene, ultra-marathon running performance, and range of motion.,” *International journal of sports physiology and performance*, vol. 6, no. 4, pp. 485–96, Dec. 2011.

[109] K. O’Connell, M. Posthumus, M. P. Schwellnus, and M. Collins, “Collagen genes and exercise-associated muscle cramping.,” *Clinical journal of sport medicine : official journal of the Canadian Academy of Sport Medicine*, vol. 23, no. 1, pp. 64–9, Jan. 2013.

[110] B.-J. Thiele, A. Doller, T. Kähne, R. Pregla, R. Hetzer, and V. Regitz-Zagrosek, “RNA-binding proteins heterogeneous nuclear ribonucleoprotein A1, E1, and K are involved in post-transcriptional control of collagen I and III synthesis.,” *Circulation research*, vol. 95, no. 11, pp. 1058–66, Nov. 2004.

[111] Y. Huang, L. Lin, X. Yu, G. Wen, X. Pu, H. Zhao, C. Fang, J. Zhu, S. Ye, L. Zhang, and Q. Xiao, “Functional Involvements of Heterogeneous Nuclear Ribonucleoprotein A1 in Smooth Muscle Differentiation from Stem Cells in Vitro and in Vivo.,” *Stem cells (Dayton, Ohio)*, Jan. 2013.

[112] J. Liu, B. Huang, Y. Xiao, H.-M. Xiong, J. Li, D.-Q. Feng, X.-M. Chen, H.-B. Zhang, and X.-Z. Wang, “Aberrant expression of splicing factors in newly diagnosed acute myeloid leukemia.,” *Onkologie*, vol. 35, no. 6, pp. 335–40, Jan. 2012.

[113] Y. Park, J.-M. Lee, M.-Y. Hwang, G. Son, and D. Geum, “NonO binds to the CpG island of oct4 promoter and functions as a transcriptional activator of oct4 gene expression.,” *Molecules and cells*, vol. 35, no. 1, pp. 61–9, Jan. 2013.

[114] E. Kowalska, J. a Ripperger, D. C. Hoegger, P. Bruegger, T. Buch, T. Birchler, A. Mueller, U. Albrecht, C. Contaldo, and S. a Brown, “NONO couples the circadian clock to the cell cycle.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 5, pp. 1592–9, Jan. 2013.

[115] R. Peng, B. T. Dye, I. Pérez, D. C. Barnard, A. B. Thompson, and J. G. Patton, “PSF and p54nrb bind a conserved stem in U5 snRNA.,” *RNA (New York, N.Y.)*, vol. 8, no. 10, pp. 1334–47, Oct. 2002.

[116] W. W. Zhang, L. X. Zhang, R. K. Busch, J. Farrés, and H. Busch, “Purification and characterization of a DNA-binding heterodimer of 52 and 100 kDa from HeLa cells.,” *The Biochemical journal*, vol. 290 Pt 1, pp. 267–72, Feb. 1993.

[117] S. E. Walker, F. Zhou, S. F. Mitchell, V. S. Larson, L. Valasek, A. G. Hinnebusch, and J. R. Lorsch, “Yeast eIF4B binds to the head of the 40S ribosomal subunit and promotes mRNA recruitment through its N-terminal and internal repeat domains.,” *RNA (New York, N.Y.)*, vol. 19, no. 2, pp. 191–207, Mar. 2013.

[118] A. Parsyan, Y. Svitkin, D. Shahbazian, C. Gkogkas, P. Lasko, W. C. Merrick, and N. Sonenberg, “mRNA helicases: the tacticians of translational control.,” *Nature reviews. Molecular cell biology*, vol. 12, no. 4, pp. 235–45, Apr. 2011.

[119] E. Hollams, K. Giles, and A. Thomson, “MRNA stability and the control of gene expression: implications for human disease,” *Neurochemical*, vol. 27, no. 10, pp. 957–980, 2002.

[120] C. Mauch, A. Hatamochi, and K. Scharffetter, “Regulation of collagen synthesis in fibroblasts within a three-dimensional collagen gel,” *Experimental cell research*, vol. 279, no. 22, pp. 822–829, 1988.

[121] J. Guhaniyogi and G. Brewer, “Regulation of mRNA stability in mammalian cells,” *Gene*, vol. 265, no. 1–2, pp. 11–23, 2001.

[122] M. Rickert, H. Wang, P. Wieloch, H. Lorenz, E. Steck, D. Sabo, and W. Richter, “Adenovirus-mediated gene transfer of growth and differentiation factor-5 into tenocytes and the healing rat Achilles tendon.,” *Connective tissue research*, vol. 46, no. 4–5, pp. 175–83, Jan. 2005.

[123] T. Xia, J. SantaLucia, M. E. Burkard, R. Kierzek, S. J. Schroeder, X. Jiao, C. Cox, and D. H. Turner, “Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs.,” *Biochemistry*, vol. 37, no. 42, pp. 14719–35, Oct. 1998.

[124] Z.-J. Tan and S.-J. Chen, “Predicting ion binding properties for RNA tertiary structures.,” *Biophysical journal*, vol. 99, no. 5, pp. 1565–76, Sep. 2010.

[125] J. SantaLucia and D. H. Turner, “Measuring the thermodynamics of RNA secondary structure formation.,” *Biopolymers*, vol. 44, no. 3, pp. 309–19, Jan. 1997.

Appendix

Buffers and Solutions

Trypsin-EDTA
8g NaCl
1.26g Na_2HPO_4
0.2g KCl
0.2g KH_2PO_4
0.5g Trypsin
0.2g EDTA

Make up to 1 litre with dH_2O
bring to pH 7.4
Filter sterilise through 0.2µM filter
Store at 4° C

10x Tris-Borate-EDTA (TBE) electrophoresis buffer
0.89M Tris
0.89M Boric acid
0.02M EDTA

Make up to 1 litre with dH_2O
For use, dilute to 1x

10% Sodium dodecyl sulphate (SDS)
100g SDS

Make up to 1 litre with dH_2O
Bring to pH 7.2
Store at room temperature

10x Phosphate Buffered Saline (PBS)
80g NaCl
26.8g $Na_2HPO_4.12H_2O$
2g KCl
2.4g KH_2PO_4

Make up to 1 litre, pH to 6.9 and autoclave
For use, dilute to 1x

Penicillin/Streptomycin
900mg penicillin
1500mg streptomycin

Make up to 150ml in 1x PBS
Filter through 0.2µM filter
Store at -20° C

Luria Broth
10g/l Bacto-tryptone
5g/l yeast extract
5g/l NaCl

Luria Agar
100g/l Bacto-tryptone
5g/l yeast extract
10g/l NaCl
15g/l agar

Bring to pH 7

DEPC-treatment
0.1% DEPC in distilled water
Stir for 30 minutes
Soak pipette tips and microfuge tips in DEPC-treated water overnight
Remove as much water as possible and autoclave

Maxiprep Solution I
4.5g of glucose
10ml 0.5M EDTA pH 8.0
12.5ml 1M Tris-Cl pH 8.0

Make up to 500ml with ddH_2O and filter sterilise

Maxiprep Solution II
8ml 10M NaOH
20ml 20% Sodium Dodecyl Sulphate solution

Make up to 400ml with autoclaved ddH_2O

Maxiprep Solution III
147.2g Potassium Acetate (KAc)
57.5ml Glacial Acetic acid

Make up to 500 ml with ddH_2O and autoclave

Appendix A : Achilles tendinopathy informed consent form

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

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

INFORMED CONSENT

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

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

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

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

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

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

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

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

Name of Participant: _____

Signature: _____ **Date:**_____

Name of Researcher:_____

Signature: _____ **Date:**_____

Appendix B: Achilles tendinopathy Questionnaire

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

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

Instructions

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

Please complete all twelve sections A to L

Subject Number: _____

Version 3
(June 2011)

Subject No: _____

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

Subject No: _____

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

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

Subject No: _____

Section E. General Personal Medical History

Do you currently suffer from any of these medical conditions:

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

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

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

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

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

If female:

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

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

Subject No: _____

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

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

Subject No: _____

Section E. General Personal Medical History

Do you currently suffer from any of these medical conditions:

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

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

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

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

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

If female:

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

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

Subject No: _____

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

Subject No: _____

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

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

Subject No: _____

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

Subject No: _____

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

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

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

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

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

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

Also indicate if you partially or completely tore the muscle.

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

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

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

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

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

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

Approximate number of days: .....

Approximate date of muscle injury?

Subject No: _____

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

Subject No: _____

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

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

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

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

Please tick which **tendon/s** you have injured? (next column on the tright)

Also indicate (tick) if your injured tendon was longsatnding pain (tendinopathy) or an acute tear/rupture

	Tendon		Longstanding Pain (Tendinopathy) Left Right	Acute Tear/ Rupture Left Right
	Foot and ankle:	☐ Achilles tendon	☐☐	☐☐
		☐ Tibialis posterior	☐☐	☐☐
		☐ Plantar fascia	☐☐	☐☐
	Knee:	☐ Patellar tendon	☐☐	☐☐
	Elbow and wrist:	☐ Wrist extensor tendon	☐☐	☐☐
	Shoulder:	☐ Rotator cuff	☐☐	☐☐
	Other: _____		☐☐	☐☐

Please tick which **ligament/s** you have injured? (next column on the tright)

Also indicate if your sprained or completely tore the ligamant.

Ligament	Sprain Left Right	Complete Tear Left Right
☐ Shoulder ligaments	☐☐	☐☐
☐ Elbow ligaments	☐☐	☐☐
☐ Wrist ligaments	☐☐	☐☐
☐ Finger ligaments	☐☐	☐☐
☐ Knee (ACL)	☐☐	☐☐
☐ Knee (MCL)	☐☐	☐☐
☐ Knee (PCL)	☐☐	☐☐
☐ Knee (LCL)	☐☐	☐☐
☐ Ankle lateral ligaments	☐☐	☐☐
☐ Ankle medial ligaments	☐☐	☐☐
☐ Spinal ligaments	☐☐	☐☐
☐ Other: _____	☐☐	☐☐

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

Subject No: _____

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

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

Subject No: _____

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

Subject No: _____

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

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

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

Appendix C: University of Cape Town Ethics Approval

UNIVERSITY OF CAPE TOWN

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

28 April 2011

Sent via internal mail

HREC REF: 158/2011

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

Dear A/PROF COLLINS,

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

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

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

Approval is granted until 30 April 2012

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

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

Please quote the HREC. REF in all your correspondence.

Yours sincerely

Lesley Henley

A/PROF MARC BLOCKMAN
pp CHAIRPERSON, FHS HUMAN ETHICS

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

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

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

Appendix D: University of La Trobe Ethics approval

UNIVERSITY OF CAPE TOWN

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

08 May 2008

REC REF: 086/2005

A/Prof M Collins
Human Biology
Sports Science Institute

Dear A/Prof Collins

PROJECT TITLE: THE GENETIC BASIS OF TENDINOPATHY

Thank you for your letter to the Research Ethics Committee dated 24 April 2008.

Addendum to undertake further genetic analysis is approved.
We note that future analysis will be performed on de-identified samples.

Please would you submit an annual progress report which includes a description of the current status of the research and your publication as outlined in this letter.

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

Please quote the REC. REF in all your correspondence.

Yours sincerely

PROFESSOR M BLOCKMAN
pp CHAIRPERSON, HSF HUMAN ETHICS

lemjedi

Appendix E: Diagnosis of Achilles Tendinopathy Checklist

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

CLINICAL DIAGNOSIS OF ACHILLES TENDINOPATHY

SUBJECT NAME/NUMBER/CODE: _____

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

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

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

Date: _____ / _____ / 20____

Investigator: _____

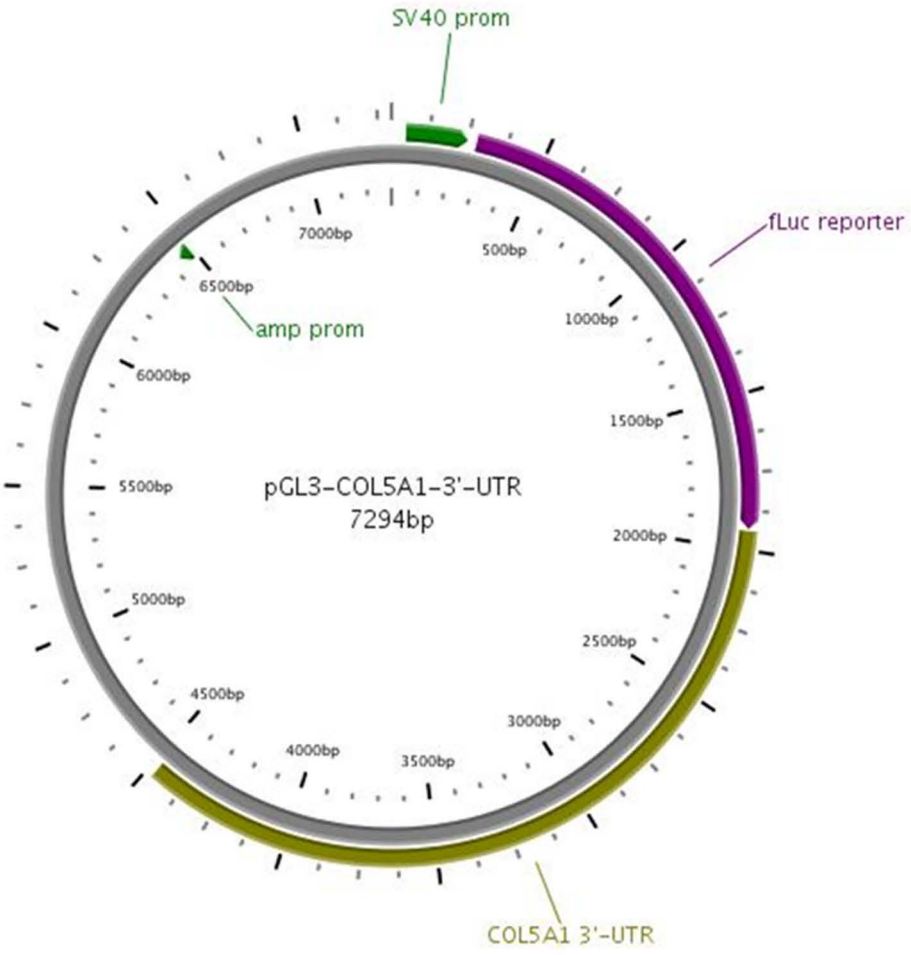
Signature: _____

References:

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

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

Appendix F: pGL3-TEN/CON DNA construct with 3'-UTR insert.



Appendix G: Nested PCR primer sets

Primer set 1 :

FWD: 5’-GCTTGCTTCATGGGCTAGCAGCCGC-3’

REV: 5’-GAGTCACTCGGATCCGTTTACGGTGG-3’

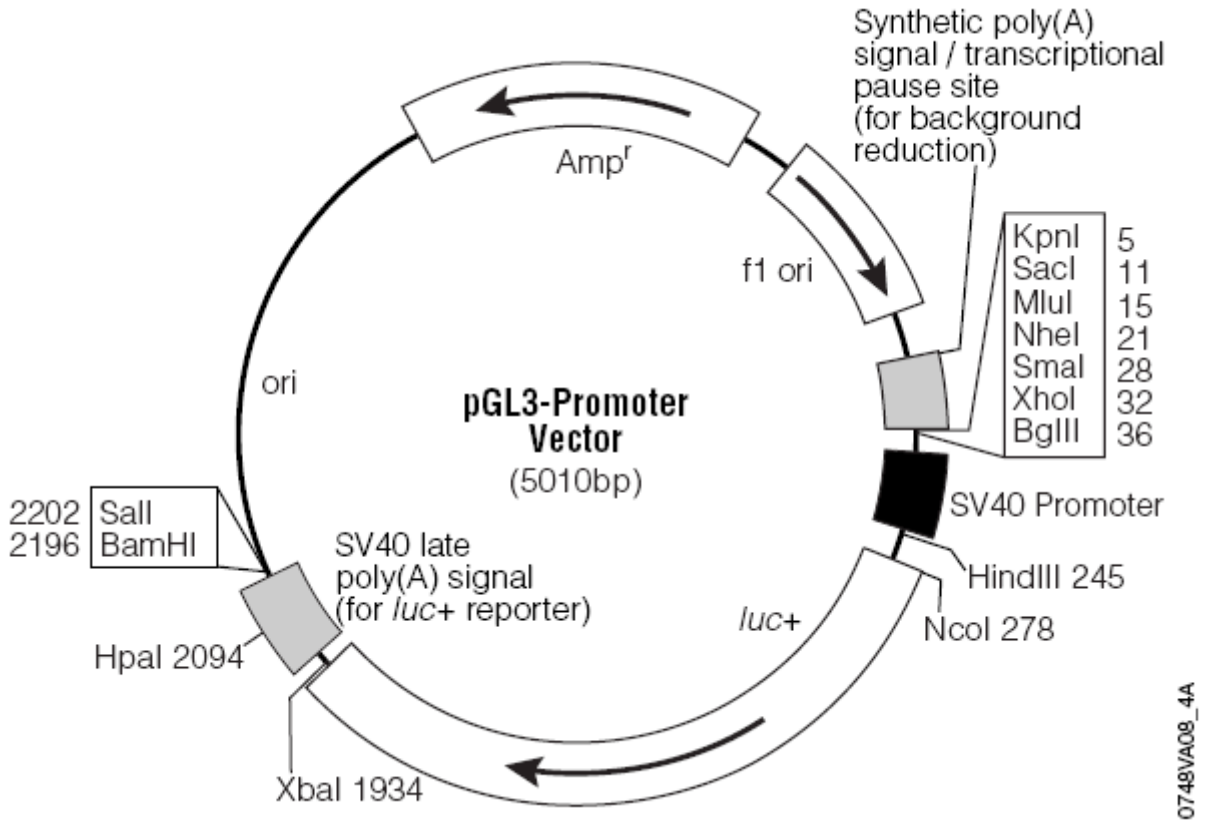
N.B. – the forward primer in set 1 contains a *Nhe*I restriction site (bold, underlined), the reverse primer in set 1 contains a *Bam*HI restriction site. These two site were used to clone the *COL5A1* 3’-UTR into the pGL3-Promoter vector

Primer set 2 :

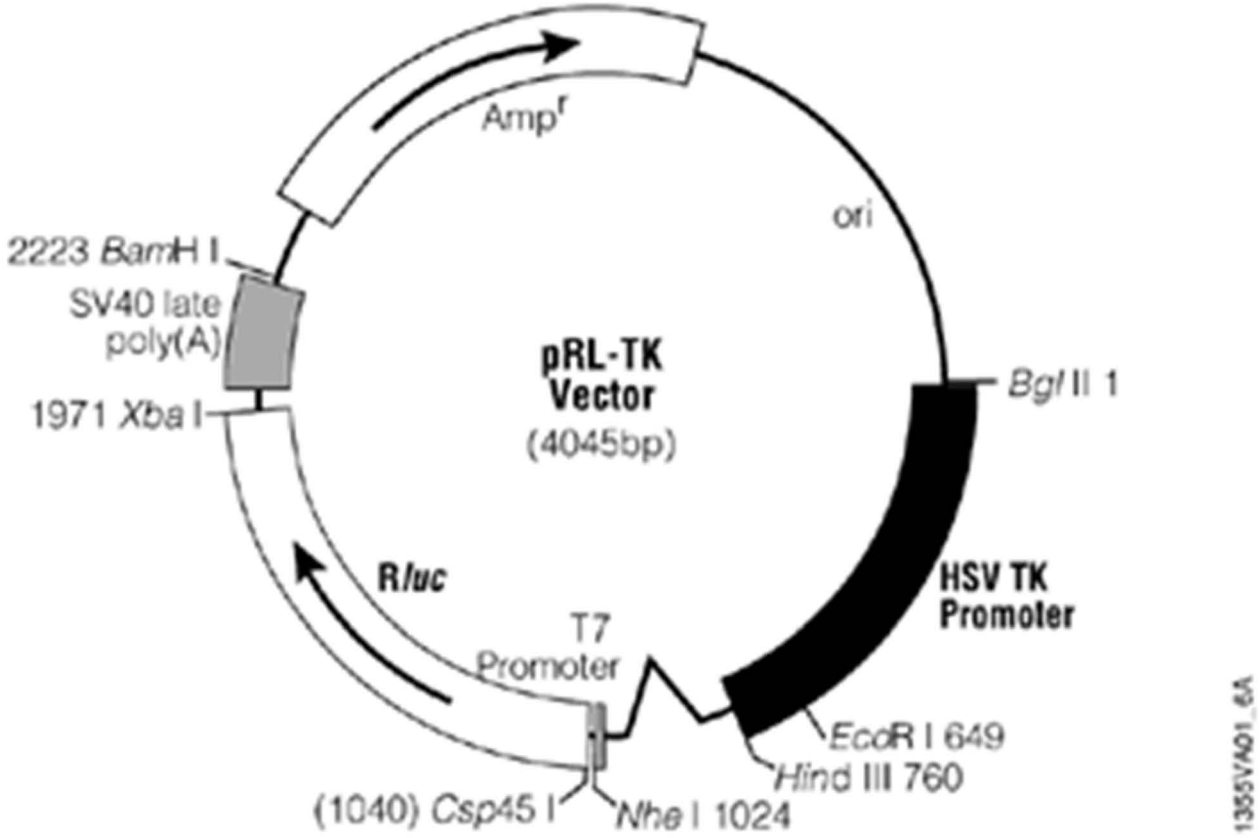
FWD: 5’-GATCAGCTTCAATCCTGTGTGTGC-3’

REV: 5’-GTTATCTCCAGAGCTCCTAGCGTCCT-3’

Appendix H: pGL3-Promoter Vector with restriction sites.



Appendix I: Thymidine Kinase



Appendix J: The primers used to sequence the 2.5 kb *COL5A1* 3'-UTR cloned into pGL3-Promoter vector.

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

[b]5'-position within the pGL3-Promoter vector (accession number U47298.2)

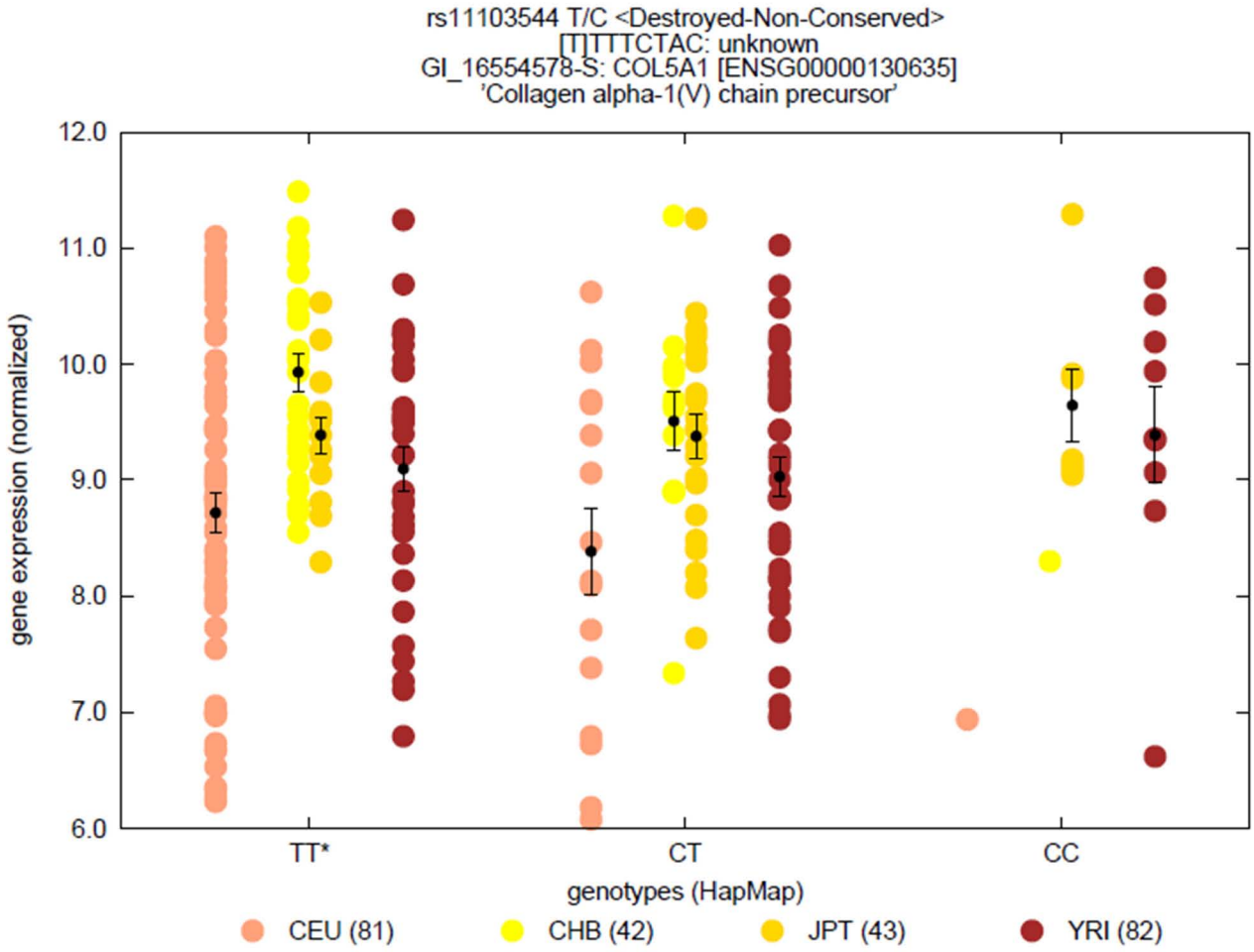
Appendix K: Attractene Reagent Transfection Protocol (Hsa-miR-608 HT1080)

1. Twenty four hours prior to transfection, 7.5 x 10^4 cells were seeded per well in a 12-well plate. Cells were seeded in 1 ml of Dulbecco's Modified Eagle Medium (DMEM) (Highveld Biological, South Africa) containing 10% GIBCO Fetal Bovine Serum (FBS) (Highveld Biological, South Africa) and 1% Penicillin/Streptomycin antibiotic solution.
2. Cells were incubated under normal growth conditions at 37 ºC, 5% CO_2 and 65% humidity.
3. The next day, when cells reached confluency of 60-80%, 1500 ng of construct DNA along with 150 ng of r*enilla* luciferase construct DNA dissolved in sterile H_2O was added to a tube and made up to a volume of 240 µl with DMEM without serum or antibiotics.
4. The reaction mixture was briefly vortexed and centrifuged to mix sample DNA.
5. A corresponding volume of Hsa-miR-608 microRNA mimic or scrambled siRNA control was added to their respective tubes to absolute values of 0; 3; 30; 150; 300 pmol.
6. Reaction mixtures were briefly vortexed and centrifuged to mix sample DNA and RNA
7. 9 µl of Attractene transfection reagent (QIAGEN®) was added to each eppendorf.
8. Reaction mixtures were briefly vortexed and centrifuged at high speed.
9. Samples were allowed to incubate at room temperature for 15 minutes to allow for complex formation.
10. While complex formation was taking place, medium was gently aspirated from cells and 1000 µl of fresh DMEM containing serum and antibiotics was added to each well.
11. Seventy seven (77) µl of the transfection complexes were added drop-wise onto the cells and gently swirled.
12. Cells were incubated at 37 ºC and 5% CO_2 for 45 hours before protein harvesting.

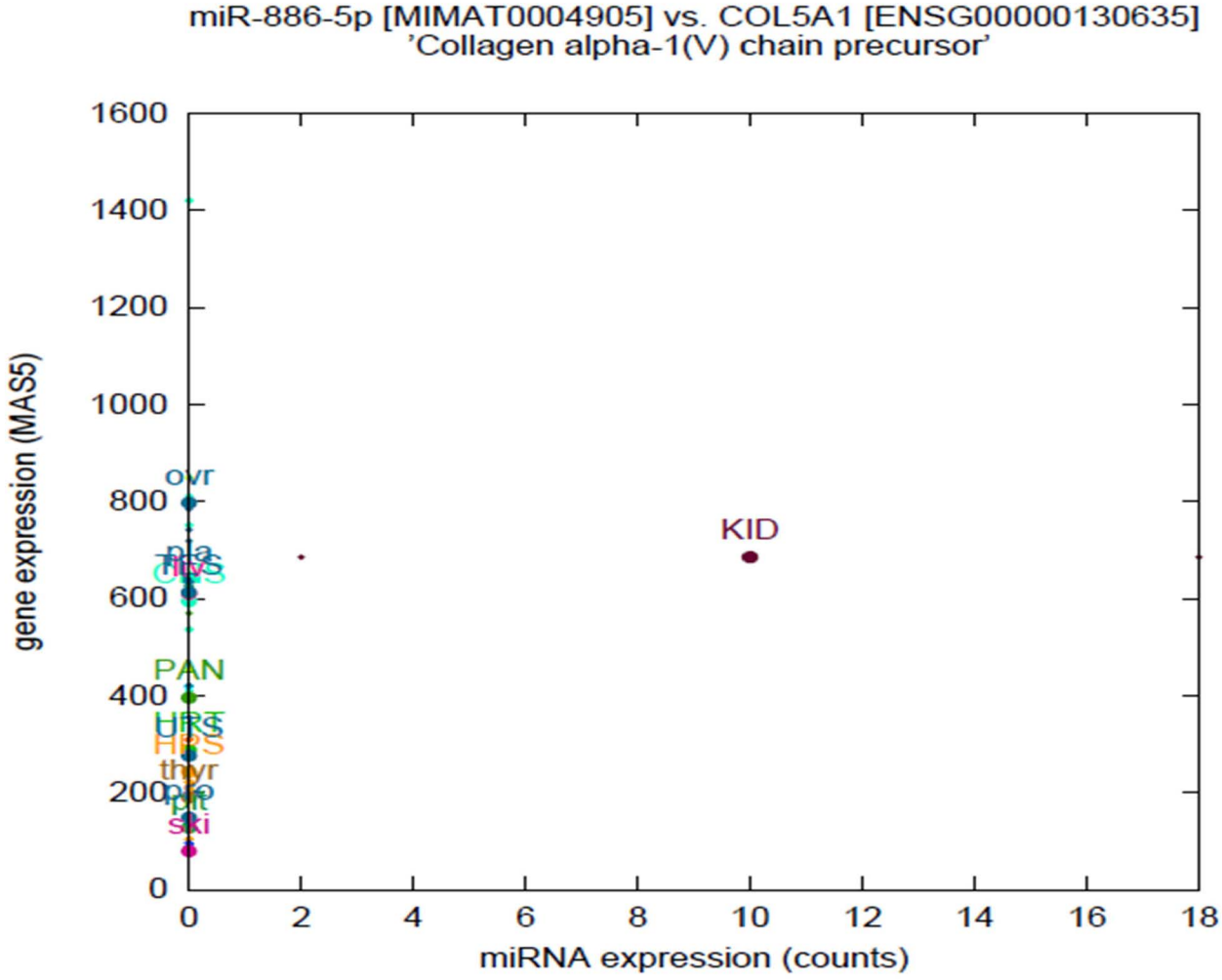
Appendix L: Attractene Reagent Transfection Protocol (Altered)

1. Twenty four hours prior to transfection, 7.5 x 10^4 cells were seeded per well in a 12-well plate. Cells were seeded in 1 ml of Dulbecco's Modified Eagle Medium (DMEM) (Highveld Biological, South Africa) containing 10% GIBCO Fetal Bovine Serum (FBS) (Highveld Biological, South Africa) and 1% Penicillin/Streptomycin antibiotic solution.
2. Cells were incubated under normal growth conditions at 37 ºC, 5% CO_2 and 65% humidity.
3. The next day, when cells reached confluency of 60-80%, 1500 ng of construct DNA along with 150 ng of r*enilla* luciferase construct DNA dissolved in sterile H_2O was added to a tube and made up to a volume of 240 µl with DMEM without serum or antibiotics.
4. The reaction mixture was briefly vortexed and centrifuged to mix sample DNA.
5. A corresponding volume of microRNA mimic was added to the respective tubes to absolute values of 0; 3; 30 or 60 pmol of miRNA mimic. Where needed, siRNA negative control was then added to the solution to bring the total amount of RNA up to 60 pmol.
6. Reaction mixtures were briefly vortexed and centrifuged to mix sample DNA and RNA
7. 9 µl of Attractene transfection reagent (QIAGEN®) was added to each eppendorf.
8. Reaction mixtures were briefly vortexed and centrifuged at high speed.
9. Samples were allowed to incubate at room temperature for 15 minutes to allow for complex formation.
10. While complex formation was taking place, medium was gently aspirated from cells and 1000 µl of fresh DMEM containing serum and antibiotics was added to each well.
11. Seventy seven (77) µl of the transfection complexes were added drop-wise onto the cells and gently swirled.
12. Cells were incubated at 37 ºC and 5% CO_2 for 45 hours before protein harvesting.

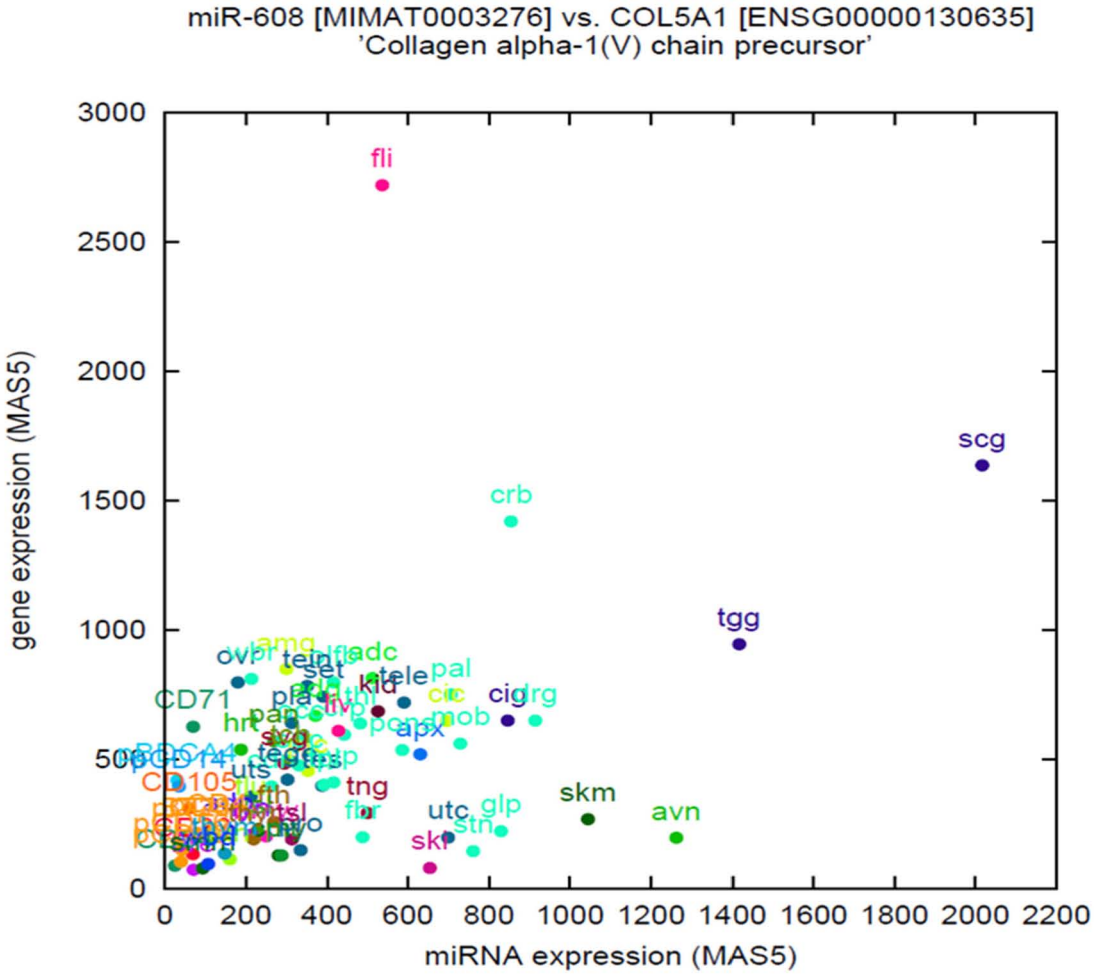
Appendix M: Confirmation of rs11103544 using gene expression and HapMap genotypes



Appendix N: Confirmation of polymorphic miRNA binding sites Hsa-miR-886-5p (rs1134114) and Hsa-miR-608 (rs3196378) by target gene and miRNA gene-expression



target gene expression taken from SymAtlas 2.0
miRNA expression taken from Landgraf et al. (2007)



target gene expression taken from SymAtlas 2.0
miRNA expression taken from host gene SEMA4G [ENSG00000095539]
'Semaphorin-4G precursor'

Figure Legend

Code	Organism	System
ADP	Mus musculus	adipose tissue
CNS	Homo sapiens	central nervous system
CNS	Mus musculus	central nervous system
EMB	Mus musculus	embryo
HRT	Homo sapiens	heart
HPS	Homo sapiens	hematopoietic system
HPS	Mus musculus	hematopoietic system
IMS	Homo sapiens	immune system
INT	Mus musculus	intestine
KID	Homo sapiens	kidney
KID	Mus musculus	kidney
NS	Homo sapiens	nervous system
NS	Mus musculus	nervous system
PAN	Homo sapiens	pancreas
PAN	Mus musculus	pancreas
TES	Homo sapiens	testis
TES	Mus musculus	testis
UTS	Homo sapiens	uterus
eye	Mus musculus	eye

Appendix O: 33 putative miRNA targets predicted by MicroCOSM algorithm

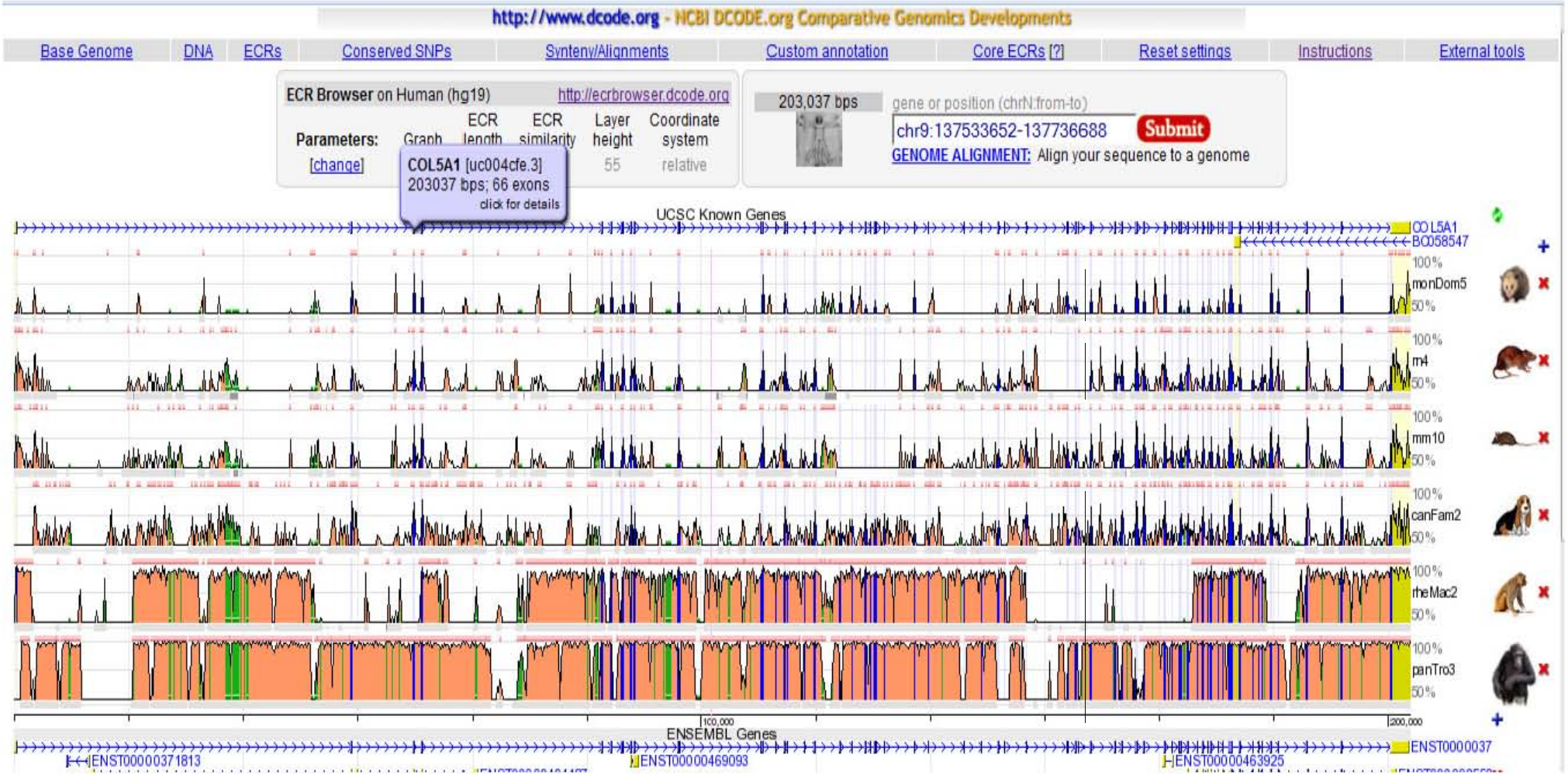
Rfam ID	Score	Energy	Base P	Poisson P	Org P	Start	End
hsa-miR-768-5p	18.261	-29.09	8.08E-03	8.05E-03	8.05E-03	453	479
hsa-miR-377*	17.738	-26.18	1.16E-02	1.16E-02	1.16E-02	-3	17
hsa-miR-125a-5p	17.577	-23.67	1.32E-02	1.32E-02	4.71E-04	302	325
hsa-miR-150	17.316	-32.17	2.33E-02	2.30E-02	2.30E-02	311	333
hsa-miR-499-5p	17.189	-11.35	2.14E-02	2.11E-02	3.03E-04	540	560
hsa-miR-330-5p	17.105	-26.56	4.74E-02	4.63E-02	2.75E-03	130	151
hsa-miR-124	17.065	-17.22	1.79E-02	1.77E-02	3.01E-05	502	521
hsa-miR-125b	16.999	-24.99	2.19E-02	2.16E-02	4.71E-04	306	325
hsa-miR-331-3p	16.871	-25.46	3.93E-02	3.86E-02	3.86E-02	305	326
hsa-miR-570	16.788	-17.88	5.05E-02	4.93E-02	1.13E-02	530	550
hsa-miR-625	16.659	-19.88	5.01E-02	4.88E-02	4.88E-02	69	89
hsa-miR-362-5p	16.635	-26.54	2.49E-02	2.46E-02	2.46E-02	304	327
hsa-miR-211	16.577	-23.71	5.75E-02	5.59E-02	9.23E-03	142	163
hsa-miR-506	16.34	-17.1	3.22E-02	3.17E-02	3.17E-02	502	521
hsa-miR-887	16.154	-20.46	1.35E-02	1.34E-02	1.34E-02	21	42
hsa-miR-517*	16.154	-21.14	3.79E-02	3.72E-02	3.72E-02	474	495
hsa-miR-204	16.049	-21.11	8.65E-02	8.29E-02	1.45E-03	142	163
hsa-miR-886-5p	15.974	-24	3.73E-02	3.67E-02	4.65E-03	278	299

hsa-miR-541	15.943	-24.78	7.54E-02	7.26E-02	5.78E-03	208	229
hsa-miR-425*	15.943	-16.84	2.82E-02	2.78E-02	2.78E-02	504	526
hsa-miR-27a*	15.943	-21.85	5.83E-02	5.67E-02	1.66E-02	205	226
hsa-miR-154	15.943	-15.3	4.48E-02	4.38E-02	1.52E-05	181	202
hsa-miR-219-1-3p	15.732	-11.82	6.29E-02	6.10E-02	8.99E-03	358	379
hsa-miR-560	15.679	-20.68	2.59E-02	2.56E-02	3.55E-04	40	59
hsa-miR-450a	15.626	-14.02	4.31E-02	4.22E-02	7.35E-03	413	434
hsa-miR-769-5p	15.626	-19.81	8.08E-02	7.77E-02	8.02E-03	303	321
hsa-miR-629	15.597	-19.59	8.17E-02	7.85E-02	6.74E-03	408	428
hsa-miR-299-5p	15.415	-14.95	4.22E-02	4.13E-02	1.37E-02	406	427
hsa-miR-205	15.415	-20.12	9.09E-02	8.69E-02	1.15E-03	61	81
hsa-miR-192	15.385	-17.1	4.85E-02	4.74E-02	9.64E-03	300	320
hsa-miR-215	15.385	-17.06	4.71E-02	4.60E-02	4.60E-02	300	320
hsa-miR-574-3p	14.993	-13.83	4.96E-02	4.84E-02	2.61E-03	28	49
hsa-miR-585	14.371	-11.83	9.68E-02	9.23E-02	1.06E-03	100	118

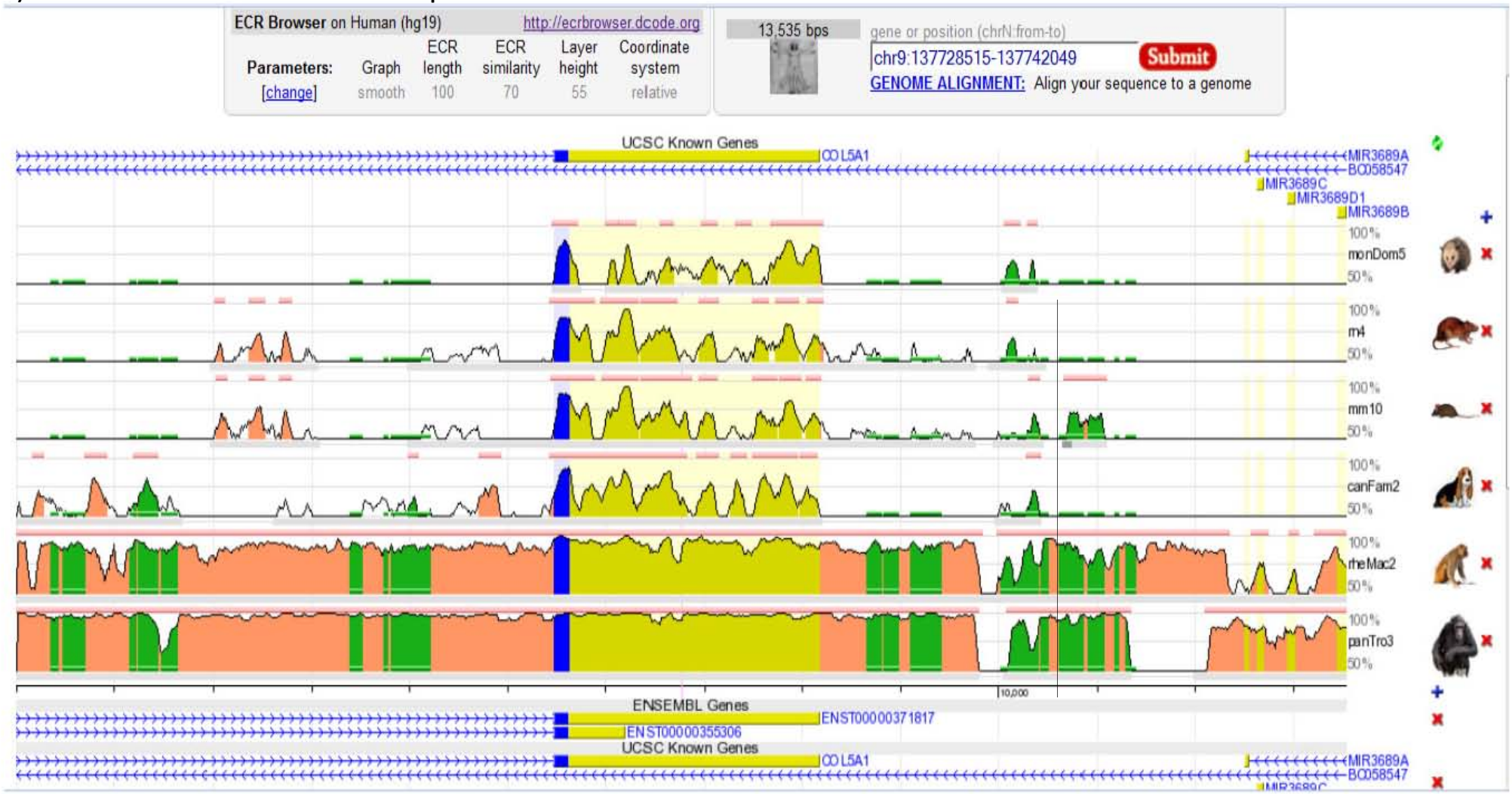
N.B – The start and end positions of the miRNA binding sites indicated in the table represent the wild-type sequence of exon 66 of the *COL5A1* gene and not the position within the pGL3_COL5A1_3'_UTR clones.

Appendix P: Sequence Conservation of the *COL5A1* gene

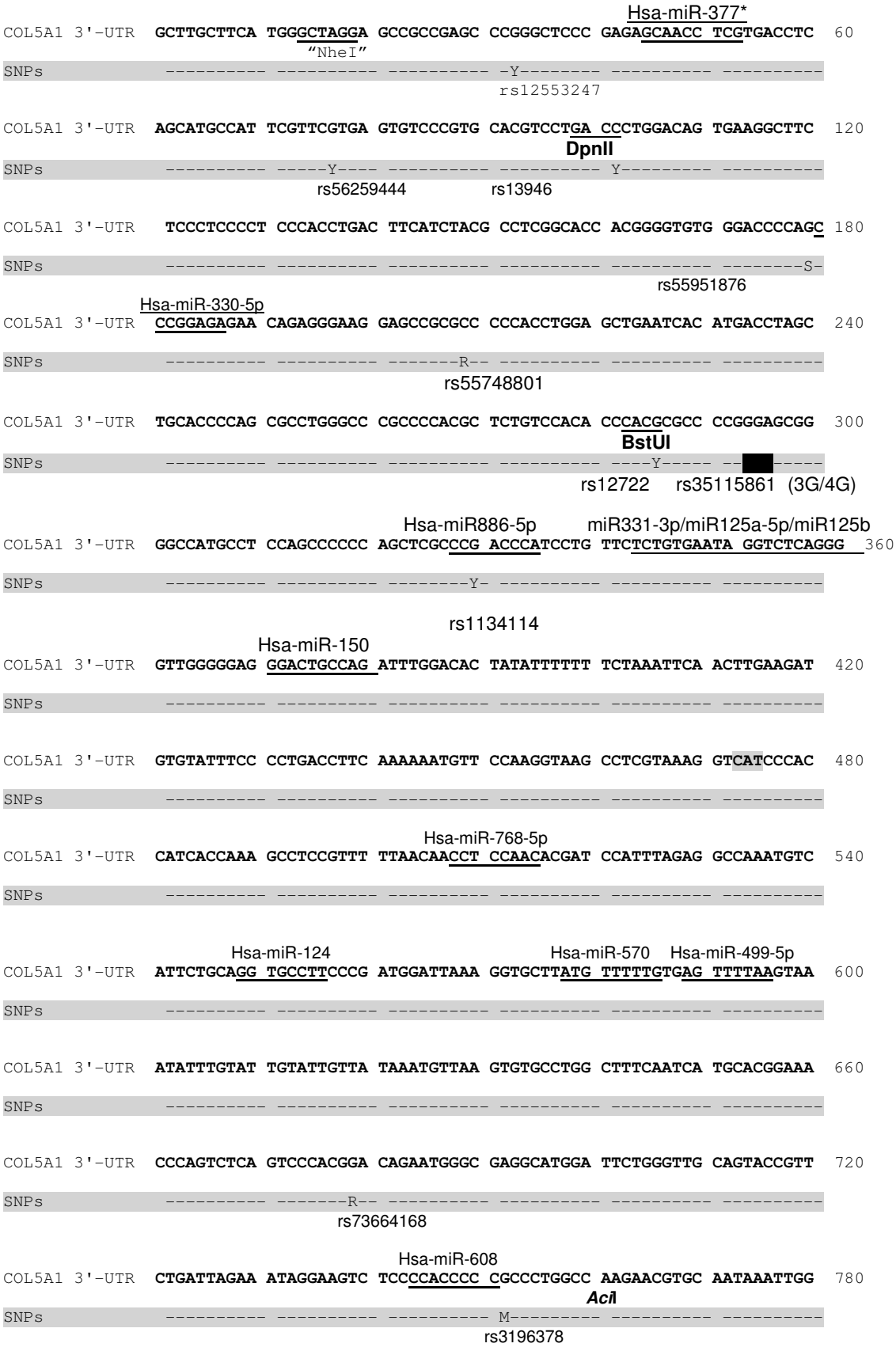
A) Conservation of the *COL5A1* gene across species

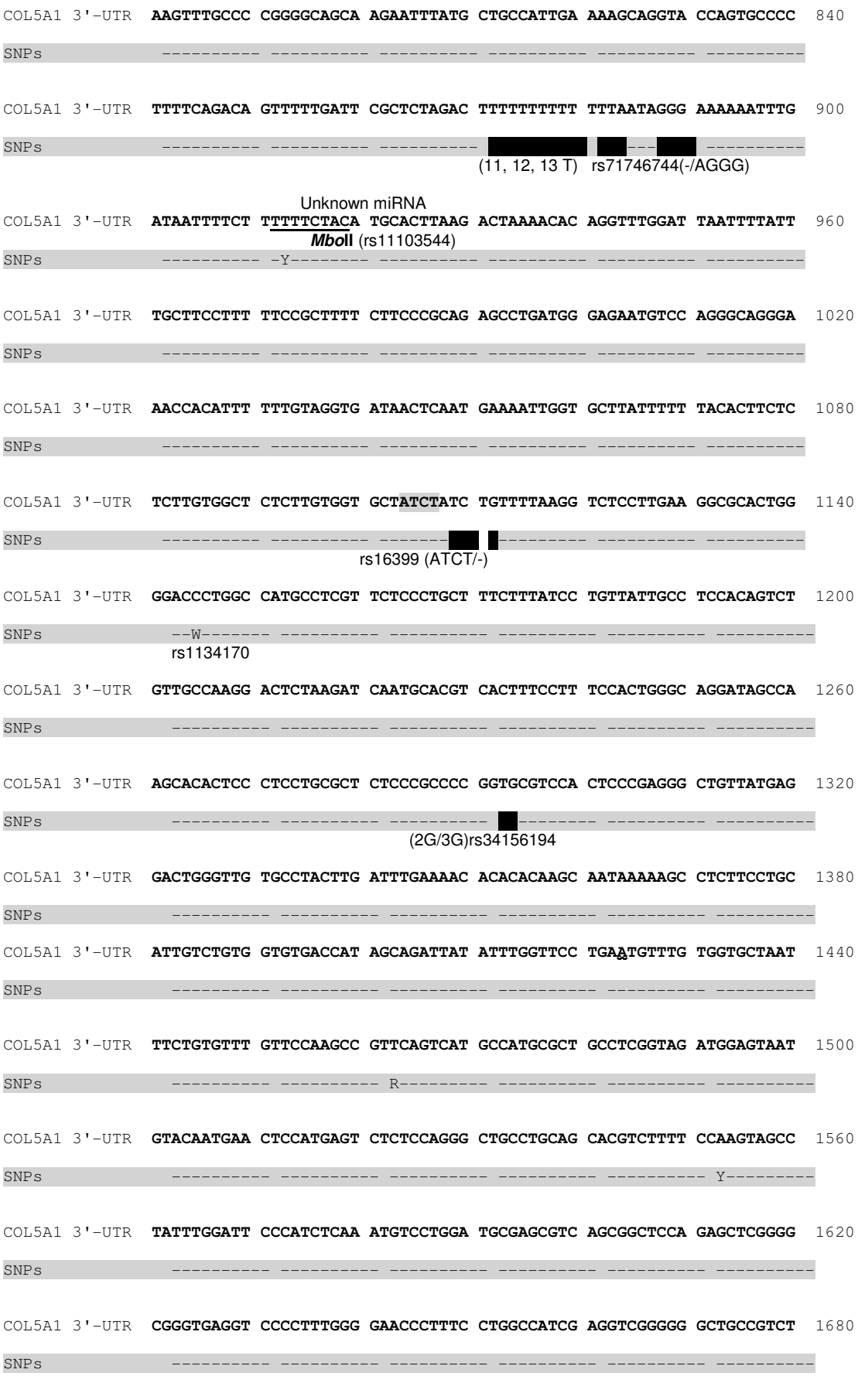


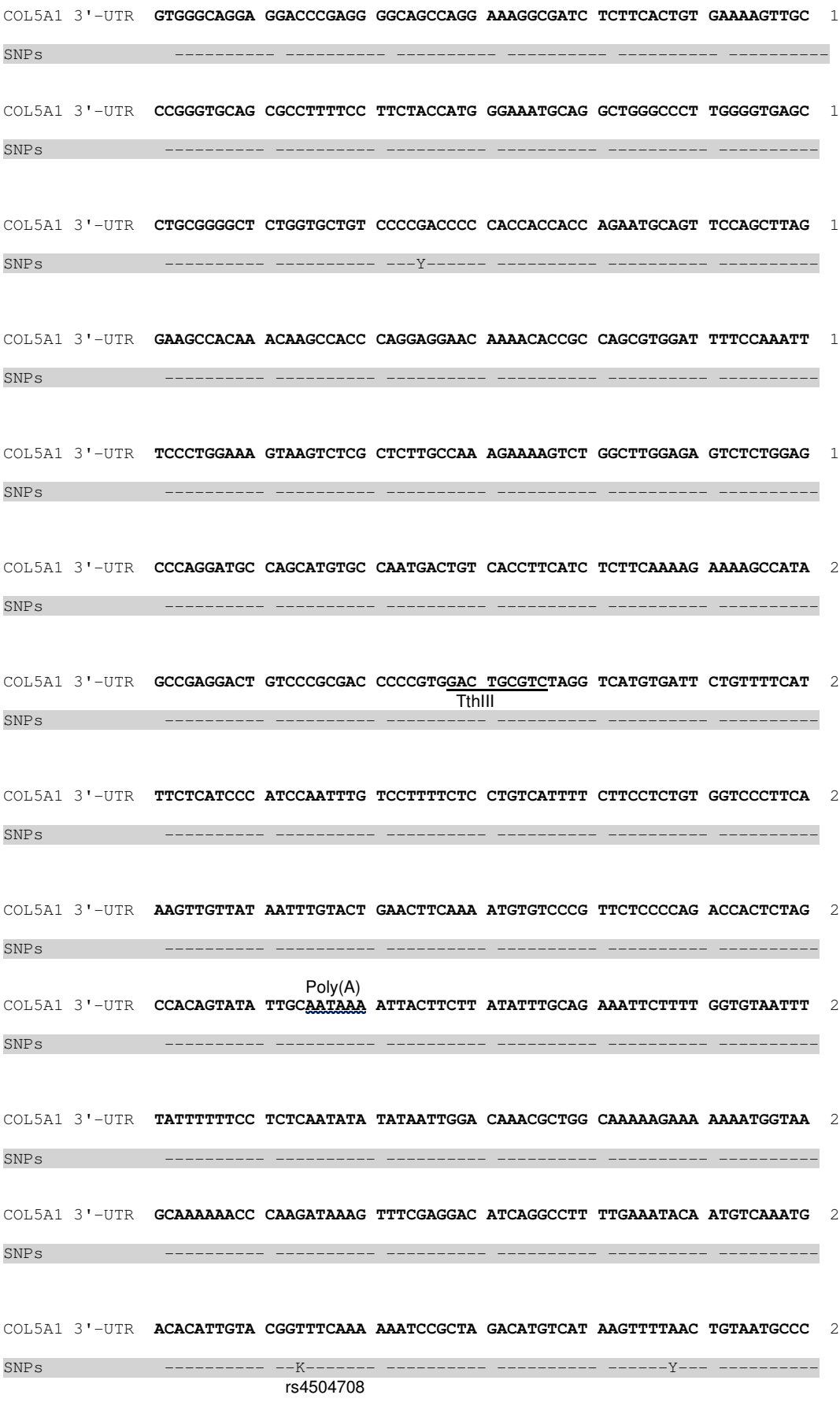
B) Conservation of the *COL5A1* 3'-UTR across species

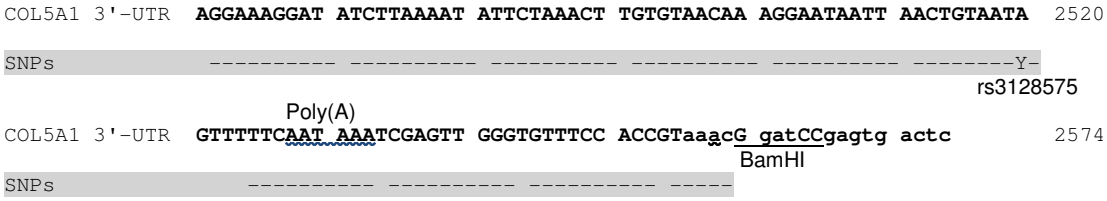


Appendix Q: Annotated sequence of the *COL5A1* 3'-UTR




COL5A1 3'-UTR AAGTTTGCCC CGGGGCAGCA AGAATTTATG CTGCCATTGA AAAGCAGGTA CCAGTGCCCC 840
SNPs ---------- ---------- ---------- ---------- ---------- ----------
COL5A1 3'-UTR TTTTCAGACA GTTTTTGATT CGCTCTAGAC TTTTTTTTTT TTTAATAGGG AAAAAATTTG 900
SNPs ---------- ---------- ---------- ---------- ----------
(11, 12, 13 T) rs71746744(-/AGGG)
Unknown miRNA
COL5A1 3'-UTR ATAATTTTCT TTTTTCTACA TGCACTTAAG ACTAAAACAC AGGTTTGGAT TAATTTTATT 960
MboII (rs11103544)
SNPs ---------- -Y-------- ---------- ---------- ---------- ----------
COL5A1 3'-UTR TGCTTCCTTT TTCCGCTTTT CTTCCCGCAG AGCCTGATGG GAGAATGTCC AGGGCAGGGA 1020
SNPs ---------- ---------- ---------- ---------- ---------- ----------
COL5A1 3'-UTR AACCACATTT TTTGTAGGTG ATAACTCAAT GAAAATTGGT GCTTATTTTT TACACTTCTC 1080
SNPs ---------- ---------- ---------- ---------- ---------- ----------
COL5A1 3'-UTR TCTTGTGGCT CTCTTGTGGT GCTATCTATC TGTTTTAAGG TCTCCTTGAA GGCGCACTGG 1140
SNPs ---------- ---------- -------- --------- ---------- ----------
rs16399 (ATCT/-)
COL5A1 3'-UTR GGACCCTGGC CATGCCTCGT TCTCCCTGCT TTCTTTATCC TGTTATTGCC TCCACAGTCT 1200
SNPs --W------- ---------- ---------- ---------- ---------- ----------
rs1134170
COL5A1 3'-UTR GTTGCCAAGG ACTCTAAGAT CAATGCACGT CACTTTCCTT TCCACTGGGC AGGATAGCCA 1260
SNPs ---------- ---------- ---------- ---------- ---------- ----------
COL5A1 3'-UTR AGCACACTCC CTCCTGCGCT CTCCCGCCCC GGTGCGTCCA CTCCCGAGGG CTGTTATGAG 1320
SNPs ---------- ---------- ---------- --------- ---------- ----------
(2G/3G)rs34156194
COL5A1 3'-UTR GACTGGGTTG TGCCTACTTG ATTTGAAAAC ACACACAAGC AATAAAAAGC CTCTTCCTGC 1380
SNPs ---------- ---------- ---------- ---------- ---------- ----------
COL5A1 3'-UTR ATTGTCTGTG GTGTGACCAT AGCAGATTAT ATTTGGTTCC TGAATGTTTG TGGTGCTAAT 1440
SNPs ---------- ---------- ---------- ---------- ---------- ----------
COL5A1 3'-UTR TTCTGTGTTT GTTCCAAGCC GTTCAGTCAT GCCATGCGCT GCCTCGGTAG ATGGAGTAAT 1500
SNPs ---------- ---------- R--------- ---------- ---------- ----------
COL5A1 3'-UTR GTACAATGAA CTCCATGAGT CTCTCCAGGG CTGCCTGCAG CACGTCTTTT CCAAGTAGCC 1560
SNPs ---------- ---------- ---------- ---------- ---------- Y---------
COL5A1 3'-UTR TATTTGGATT CCCATCTCAA ATGTCCTGGA TGCGAGCGTC AGCGGCTCCA GAGCTCGGGG 1620
SNPs ---------- ---------- ---------- ---------- ---------- ----------
COL5A1 3'-UTR CGGGTGAGGT CCCCTTTGGG GAACCCTTTC CTGGCCATCG AGGTCGGGGG GCTGCCGTCT 1680
SNPs ---------- ---------- ---------- ---------- ---------- ----------


COL5A1 3'-UTR GTGGGCAGGA GGACCCGAGG GGCAGCCAGG AAAGGCGATC TCTTCACTGT GAAAAGTTGC 1740
SNPs ---------- ---------- ---------- ---------- ---------- ----------
COL5A1 3'-UTR CCGGGTGCAG CGCCTTTTCC TTCTACCATG GGAAATGCAG GCTGGGCCCT TGGGGTGAGC 1800
SNPs ---------- ---------- ---------- ---------- ---------- ----------
COL5A1 3'-UTR CTGCGGGGCT CTGGTGCTGT CCCCGACCCC CACCACCACC AGAATGCAGT TCCAGCTTAG 1860
SNPs ---------- ---------- ---Y------ ---------- ---------- ----------
COL5A1 3'-UTR GAAGCCACAA ACAAGCCACC CAGGAGGAAC AAAACACCGC CAGCGTGGAT TTTCCAAATT 1920
SNPs ---------- ---------- ---------- ---------- ---------- ----------
COL5A1 3'-UTR TCCCTGGAAA GTAAGTCTCG CTCTTGCCAA AGAAAAGTCT GGCTTGGAGA GTCTCTGGAG 1980
SNPs ---------- ---------- ---------- ---------- ---------- ----------
COL5A1 3'-UTR CCCAGGATGC CAGCATGTGC CAATGACTGT CACCTTCATC TCTTCAAAAG AAAAGCCATA 2040
SNPs ---------- ---------- ---------- ---------- ---------- ----------
COL5A1 3'-UTR GCCGAGGACT GTCCCGCGAC CCCCGTGGAC TGCGTCTAGG TCATGTGATT CTGTTTTCAT 2100
TthIII
SNPs ---------- ---------- ---------- ---------- ---------- ----------
COL5A1 3'-UTR TTCTCATCCC ATCCAATTTG TCCTTTTCTC CTGTCATTTT CTTCCTCTGT GGTCCCTTCA 2160
SNPs ---------- ---------- ---------- ---------- ---------- ----------
COL5A1 3'-UTR AAGTTGTTAT AATTTGTACT GAACTTCAAA ATGTGTCCCG TTCTCCCCAG ACCACTCTAG 2220
SNPs ---------- ---------- ---------- ---------- ---------- ----------
Poly(A)
COL5A1 3'-UTR CCACAGTATA TTGCAATAAA ATTACTTCTT ATATTTGCAG AAATTCTTTT GGTGTAATTT 2280
SNPs ---------- ---------- ---------- ---------- ---------- ----------
COL5A1 3'-UTR TATTTTTTCC TCTCAATATA TATAATTGGA CAAACGCTGG CAAAAAGAAA AAAATGGTAA 2340
SNPs ---------- ---------- ---------- ---------- ---------- ----------
COL5A1 3'-UTR GCAAAAAACC CAAGATAAAG TTTCGAGGAC ATCAGGCCTT TTGAAATACA ATGTCAAATG 2400
SNPs ---------- ---------- ---------- ---------- ---------- ----------
COL5A1 3'-UTR ACACATTGTA CGGTTTCAAA AAATCCGCTA GACATGTCAT AAGTTTTAAC TGTAATGCCC 2460
SNPs ---------- --K------- ---------- ---------- ------Y--- ----------
rs4504708


COL5A1 3'-UTR AGGAAAGGAT ATCTTAAAAT ATTCTAAACT TGTGTAACAA AGGAATAATT AACTGTAATA 2520
SNPs ---------- ---------- ---------- ---------- ---------- --------Y-
rs3128575
Poly(A)
COL5A1 3'-UTR GTTTTTCAAT AAATCGAGTT GGGTGTTTCC ACCGTaaacG gatCCgagtg actc 2574
BamHI
SNPs ---------- ---------- ---------- -----

Appendix R: RBP motifs within the wild-type C-allelic form of the *COL5A1* 3'-UTR

Score	Relative Score	RBP Name	Start	End	Matching sequence
8.6471013	100%	a2bp1	191	195	GCAUG
8.6471013	100%	a2bp1	823	827	GCAUG
8.6471013	100%	a2bp1	2122	2126	GCAUG
6.93292615	94%	ACO1	237	242	CAGUGA
7.35090902	100%	ACO1	961	966	CAGUGC
5.256645851	81%	EIF4B	69	72	GGAC
5.256645851	81%	EIF4B	234	237	GGAC
7.96844717	87%	EIF4B	297	303	GUGGGAC
5.256645851	81%	EIF4B	300	303	GGAC
6.4668404	100%	EIF4B	325	328	GGAA
5.256645851	81%	EIF4B	500	503	GGAC
5.256645851	81%	EIF4B	514	517	GGAC
6.4668404	100%	EIF4B	785	788	GGAA
5.256645851	81%	EIF4B	807	810	GGAC
6.4668404	100%	EIF4B	863	866	GGAA
6.4668404	100%	EIF4B	908	911	GGAA
6.4668404	100%	EIF4B	1018	1021	GGAA
6.4668404	100%	EIF4B	1147	1150	GGAA
5.256645851	81%	EIF4B	1270	1273	GGAC
5.256645851	81%	EIF4B	1338	1341	GGAC
5.256645851	81%	EIF4B	1449	1452	GGAC
6.4668404	100%	EIF4B	1769	1772	GGAA
5.256645851	81%	EIF4B	1820	1823	GGAC
6.4668404	100%	EIF4B	1838	1841	GGAA
6.4668404	100%	EIF4B	1900	1903	GGAA
7.590657808	80%	EIF4B	1986	1992	UUAGGAA
6.4668404	100%	EIF4B	1989	1992	GGAA
6.4668404	100%	EIF4B	2015	2018	GGAA
6.4668404	100%	EIF4B	2055	2058	GGAA
5.256645851	81%	EIF4B	2175	2178	GGAC
5.256645851	81%	EIF4B	2196	2199	GGAC
5.256645851	81%	EIF4B	2437	2440	GGAC
5.256645851	81%	EIF4B	2496	2499	GGAC
6.4668404	100%	EIF4B	2591	2594	GGAA
6.4668404	100%	EIF4B	2631	2634	GGAA
3.82636396	86%	ELAVL1	108	111	AUUU
3.82636396	86%	ELAVL1	114	117	AUUU
3.82636396	86%	ELAVL1	510	513	AUUU
3.82636396	86%	ELAVL1	523	526	AUUU
3.82636396	86%	ELAVL1	554	557	AUUU
4.40359056	100%	ELAVL1	626	629	GUUU
3.82636396	86%	ELAVL1	652	655	AUUU

4.40359056	100%	ELAVL1	709	712	GUUU
4.40359056	100%	ELAVL1	719	722	GUUU
3.82636396	86%	ELAVL1	732	735	AUUU
4.40359056	100%	ELAVL1	912	915	GUUU
3.82636396	86%	ELAVL1	933	936	AUUU
4.40359056	100%	ELAVL1	980	983	GUUU
3.82636396	86%	ELAVL1	1025	1028	AUUU
3.82636396	86%	ELAVL1	1033	1036	AUUU
4.40359056	100%	ELAVL1	1072	1075	GUUU
3.82636396	86%	ELAVL1	1082	1085	AUUU
3.82636396	86%	ELAVL1	1087	1090	AUUU
3.82636396	86%	ELAVL1	1156	1159	AUUU
3.82636396	86%	ELAVL1	1194	1197	AUUU
4.40359056	100%	ELAVL1	1241	1244	GUUU
3.82636396	86%	ELAVL1	1470	1473	AUUU
3.82636396	86%	ELAVL1	1540	1543	AUUU
4.40359056	100%	ELAVL1	1555	1558	GUUU
3.82636396	86%	ELAVL1	1568	1571	AUUU
4.40359056	100%	ELAVL1	1576	1579	GUUU
3.82636396	86%	ELAVL1	1691	1694	AUUU
3.82636396	86%	ELAVL1	2038	2041	AUUU
3.82636396	86%	ELAVL1	2047	2050	AUUU
4.40359056	100%	ELAVL1	2222	2225	GUUU
3.82636396	86%	ELAVL1	2228	2231	AUUU
3.82636396	86%	ELAVL1	2245	2248	AUUU
3.82636396	86%	ELAVL1	2265	2268	AUUU
3.82636396	86%	ELAVL1	2301	2304	AUUU
3.82636396	86%	ELAVL1	2382	2385	AUUU
3.82636396	86%	ELAVL1	2406	2409	AUUU
3.82636396	86%	ELAVL1	2411	2414	AUUU
4.40359056	100%	ELAVL1	2489	2492	GUUU
4.40359056	100%	ELAVL1	2542	2545	GUUU
4.40359056	100%	ELAVL1	2572	2575	GUUU
4.40359056	100%	ELAVL1	2650	2653	GUUU
4.40359056	100%	ELAVL1	2674	2677	GUUU
11.4274791	81%	ELAVL2	1083	1091	UUUUAUUUG
14.058233	100%	ELAVL2	2407	2415	UUUUAUUUU
12.7846426	100%	ELAVL2	2407	2415	UUUUAUUUU
7.3693752	100%	FUS	57	60	GGUG
7.3693752	100%	FUS	91	94	GGUG
7.3693752	100%	FUS	294	297	GGUG
7.3693752	100%	FUS	678	681	GGUG
7.3693752	100%	FUS	700	703	GGUG
7.3693752	100%	FUS	1166	1169	GGUG

7.3693752	100%	FUS	1187	1190	GGUG
7.3693752	100%	FUS	1227	1230	GGUG
7.3693752	100%	FUS	1420	1423	GGUG
7.3693752	100%	FUS	1519	1522	GGUG
7.3693752	100%	FUS	1561	1564	GGUG
7.3693752	100%	FUS	1752	1755	GGUG
7.3693752	100%	FUS	1873	1876	GGUG
7.3693752	100%	FUS	1923	1926	GGUG
7.3693752	100%	FUS	1942	1945	GGUG
7.3693752	100%	FUS	2400	2403	GGUG
7.3693752	100%	FUS	2671	2674	GGUG
9.8958058	100%	HNRNPA1	1015	1020	UAGGGA
4.09404991	88%	KHDRBS3	530	535	UCUAAA
3.826116535	82%	KHDRBS3	593	598	CGUAAA
3.739441123	80%	KHDRBS3	629	634	UUUAAC
4.092388301	87%	KHDRBS3	694	699	AUUAAA
3.874209753	83%	KHDRBS3	725	730	AGUAAA
4.277494655	91%	KHDRBS3	748	753	UAUAAA
4.325587873	92%	KHDRBS3	900	905	AAUAAA
4.370796073	93%	KHDRBS3	1010	1015	UUUAAU
4.44927358	95%	KHDRBS3	1029	1034	GAUAAU
4.142143128	89%	KHDRBS3	1060	1065	ACUAAA
4.418889291	94%	KHDRBS3	1078	1083	AUUAAU
3.81791863	82%	KHDRBS3	1169	1174	GAUAAC
4.325587873	92%	KHDRBS3	1490	1495	AAUAAA
4.265828835	91%	KHDRBS3	1564	1569	GCUAAU
4.200710743	90%	KHDRBS3	1624	1629	AGUAAU
4.603995645	98%	KHDRBS3	2297	2302	UAUAAU
4.325587873	92%	KHDRBS3	2364	2369	AAUAAA
4.152617525	89%	KHDRBS3	2402	2407	UGUAAU
4.603995645	98%	KHDRBS3	2430	2435	UAUAAU
4.12277259	88%	KHDRBS3	2483	2488	GAUAAA
3.739441123	80%	KHDRBS3	2574	2579	UUUAAC
4.152617525	89%	KHDRBS3	2580	2585	UGUAAU
4.044295083	86%	KHDRBS3	2602	2607	CUUAAA
4.09404991	88%	KHDRBS3	2612	2617	UCUAAA
4.652088863	100%	KHDRBS3	2633	2638	AAUAAU
3.787534341	81%	KHDRBS3	2637	2642	AUUAAC
4.152617525	89%	KHDRBS3	2643	2648	UGUAAU
4.325587873	92%	KHDRBS3	2657	2662	AAUAAA
3.826116535	82%	KHDRBS3	2682	2687	CGUAAA
6.33985	100%	KHSRP	212	215	GUCC
6.33985	100%	KHSRP	223	226	GUCC
6.33985	100%	KHSRP	403	406	GUCC

6.33985	100%	KHSRP	800	803	GUCC
6.33985	100%	KHSRP	1136	1139	GUCC
6.33985	100%	KHSRP	1425	1428	GUCC
6.33985	100%	KHSRP	1712	1715	GUCC
6.33985	100%	KHSRP	1758	1761	GUCC
6.33985	100%	KHSRP	1948	1951	GUCC
6.33985	100%	KHSRP	2180	2183	GUCC
6.33985	100%	KHSRP	2249	2252	GUCC
6.33985	100%	KHSRP	2281	2284	GUCC
6.33985	100%	KHSRP	2324	2327	GUCC
6.6279899	100%	MBNL1	133	136	UGCU
6.17832025	93%	MBNL1	397	400	CGCU
6.6279899	100%	MBNL1	702	705	UGCU
6.6279899	100%	MBNL1	938	941	UGCU
6.17832025	93%	MBNL1	990	993	CGCU
6.6279899	100%	MBNL1	1090	1093	UGCU
6.17832025	93%	MBNL1	1103	1106	CGCU
6.6279899	100%	MBNL1	1189	1192	UGCU
6.6279899	100%	MBNL1	1229	1232	UGCU
6.6279899	100%	MBNL1	1296	1299	UGCU
6.17832025	93%	MBNL1	1406	1409	CGCU
6.6279899	100%	MBNL1	1563	1566	UGCU
6.17832025	93%	MBNL1	1606	1609	CGCU
6.6279899	100%	MBNL1	1944	1947	UGCU
6.17832025	93%	MBNL1	2068	2071	CGCU
6.17832025	93%	MBNL1	2444	2447	CGCU
6.17832025	93%	MBNL1	2555	2558	CGCU
8.9484945	100%	NONO	323	327	AGGGA
8.9484945	100%	NONO	498	502	AGGGA
8.9484945	100%	NONO	1016	1020	AGGGA
8.9484945	100%	NONO	1145	1149	AGGGA
8.7178165	100%	PABPC1	570	574	AAAAA
8.7178165	100%	PABPC1	571	575	AAAAA
8.7178165	100%	PABPC1	1020	1024	AAAAA
8.7178165	100%	PABPC1	1021	1025	AAAAA
8.7178165	100%	PABPC1	1493	1497	AAAAA
7.15446438	84%	PABPC1	1996	2002	ACAAACA
8.7178165	100%	PABPC1	2451	2455	AAAAA
8.7178165	100%	PABPC1	2457	2461	AAAAA
8.7178165	100%	PABPC1	2458	2462	AAAAA
8.7178165	100%	PABPC1	2459	2463	AAAAA
8.7178165	100%	PABPC1	2472	2476	AAAAA
8.7178165	100%	PABPC1	2473	2477	AAAAA
8.7178165	100%	PABPC1	2547	2551	AAAAA

8.7178165	100%	PABPC1	2548	2552	AAAAA
8.244610744	96%	PTBP1	996	1010	AGACUUUUUUUUUU
7.2294196	100%	Pum2	551	554	UGUA
7.2294196	100%	Pum2	735	738	UGUA
7.2294196	100%	Pum2	740	743	UGUA
7.2294196	100%	Pum2	1162	1165	UGUA
7.2294196	100%	Pum2	1629	1632	UGUA
7.2294196	100%	Pum2	2304	2307	UGUA
7.2294196	100%	Pum2	2402	2405	UGUA
7.2294196	100%	Pum2	2536	2539	UGUA
7.2294196	100%	Pum2	2580	2583	UGUA
7.2294196	100%	Pum2	2622	2625	UGUA
7.2294196	100%	Pum2	2643	2646	UGUA
4.84132568	100%	RBM4	334	337	CGCG
4.84132568	100%	RBM4	414	417	CGCG
4.84132568	100%	RBM4	2184	2187	CGCG
5.2682554	100%	RBMX	15	18	CCAG
4.99861593	94%	RBMX	62	65	CCAU
4.40271173	83%	RBMX	159	162	CCCG
4.40271173	83%	RBMX	167	170	CCCG
4.99861593	94%	RBMX	196	199	CCAU
4.40271173	83%	RBMX	214	217	CCCG
4.6667232	88%	RBMX	261	264	CCAC
4.6667232	88%	RBMX	288	291	CCAC
5.2682554	100%	RBMX	305	308	CCAG
4.40271173	83%	RBMX	309	312	CCCG
4.6667232	88%	RBMX	341	344	CCAC
5.2682554	100%	RBMX	376	379	CCAG
4.40271173	83%	RBMX	388	391	CCCG
4.6667232	88%	RBMX	394	397	CCAC
4.6667232	88%	RBMX	405	408	CCAC
4.6667232	88%	RBMX	411	414	CCAC
4.40271173	83%	RBMX	419	422	CCCG
4.99861593	94%	RBMX	432	435	CCAU
5.2682554	100%	RBMX	440	443	CCAG
5.2682554	100%	RBMX	448	451	CCAG
4.40271173	83%	RBMX	456	459	CCCG
4.99861593	94%	RBMX	462	465	CCAU
5.2682554	100%	RBMX	506	509	CCAG
4.6667232	88%	RBMX	606	609	CCAC
4.99861593	94%	RBMX	609	612	CCAU
4.99861593	94%	RBMX	650	653	CCAU
4.40271173	83%	RBMX	686	689	CCCG
5.2682554	100%	RBMX	791	794	CCAG

4.6667232	88%	RBMX	803	806	CCAC
4.6667232	88%	RBMX	873	876	CCAC
4.40271173	83%	RBMX	878	881	CCCG
4.40271173	83%	RBMX	918	921	CCCG
4.99861593	94%	RBMX	943	946	CCAU
5.2682554	100%	RBMX	960	963	CCAG
4.40271173	83%	RBMX	1113	1116	CCCG
5.2682554	100%	RBMX	1138	1141	CCAG
4.6667232	88%	RBMX	1152	1155	CCAC
4.99861593	94%	RBMX	1279	1282	CCAU
4.6667232	88%	RBMX	1321	1324	CCAC
4.6667232	88%	RBMX	1371	1374	CCAC
4.40271173	83%	RBMX	1412	1415	CCCG
4.40271173	83%	RBMX	1417	1420	CCCG
4.6667232	88%	RBMX	1427	1430	CCAC
4.40271173	83%	RBMX	1432	1435	CCCG
4.99861593	94%	RBMX	1526	1529	CCAU
4.99861593	94%	RBMX	1601	1604	CCAU
4.99861593	94%	RBMX	1642	1645	CCAU
5.2682554	100%	RBMX	1654	1657	CCAG
4.99861593	94%	RBMX	1701	1704	CCAU
5.2682554	100%	RBMX	1737	1740	CCAG
4.99861593	94%	RBMX	1784	1787	CCAU
4.40271173	83%	RBMX	1823	1826	CCCG
5.2682554	100%	RBMX	1835	1838	CCAG
4.40271173	83%	RBMX	1869	1872	CCCG
4.99861593	94%	RBMX	1895	1898	CCAU
4.40271173	83%	RBMX	1951	1954	CCCG
4.6667232	88%	RBMX	1959	1962	CCAC
4.6667232	88%	RBMX	1962	1965	CCAC
4.6667232	88%	RBMX	1965	1968	CCAC
5.2682554	100%	RBMX	1968	1971	CCAG
5.2682554	100%	RBMX	1981	1984	CCAG
4.6667232	88%	RBMX	1994	1997	CCAC
4.6667232	88%	RBMX	2005	2008	CCAC
5.2682554	100%	RBMX	2009	2012	CCAG
5.2682554	100%	RBMX	2029	2032	CCAG
5.2682554	100%	RBMX	2111	2114	CCAG
5.2682554	100%	RBMX	2119	2122	CCAG
4.99861593	94%	RBMX	2165	2168	CCAU
4.40271173	83%	RBMX	2182	2185	CCCG
4.40271173	83%	RBMX	2191	2194	CCCG
4.99861593	94%	RBMX	2238	2241	CCAU
4.40271173	83%	RBMX	2326	2329	CCCG

5.2682554	100%	RBMX	2336	2339	CCAG
4.6667232	88%	RBMX	2341	2344	CCAC
4.6667232	88%	RBMX	2350	2353	CCAC
5.2682554	100%	RBMX	2588	2591	CCAG
4.6667232	88%	RBMX	2678	2681	CCAC
8.6696024	100%	RBMY1A1	1174	1178	CUCAA
8.6272192	100%	RBMY1A1	1174	1178	CUCAA
7.65633614	88%	RBMY1A1	1483	1487	CACAA
8.6696024	100%	RBMY1A1	1705	1709	CUCAA
8.6272192	100%	RBMY1A1	1705	1709	CUCAA
7.65633614	88%	RBMY1A1	1995	1999	CACAA
8.6696024	100%	RBMY1A1	2421	2425	CUCAA
8.6272192	100%	RBMY1A1	2421	2425	CUCAA
7.5622424	86%	sap-49	713	718	UUGUGA
8.7846348	100%	sap-49	1520	1525	GUGUGA
4.1881759	90%	SFRS1	29	32	UGGA
4.1881759	90%	SFRS1	50	53	UGGA
4.1881759	90%	SFRS1	68	71	UGGA
4.1881759	90%	SFRS1	111	114	UGGA
4.62028767	100%	SFRS1	146	149	AGGA
4.1881759	90%	SFRS1	233	236	UGGA
4.62028767	100%	SFRS1	328	331	AGGA
4.1881759	90%	SFRS1	346	349	UGGA
4.1881759	90%	SFRS1	513	516	UGGA
4.1881759	90%	SFRS1	691	694	UGGA
4.1881759	90%	SFRS1	826	829	UGGA
4.62028767	100%	SFRS1	862	865	AGGA
4.1881759	90%	SFRS1	907	910	UGGA
4.1881759	90%	SFRS1	1075	1078	UGGA
4.62028767	100%	SFRS1	1337	1340	AGGA
4.62028767	100%	SFRS1	1380	1383	AGGA
4.62028767	100%	SFRS1	1448	1451	AGGA
4.1881759	90%	SFRS1	1621	1624	UGGA
4.1881759	90%	SFRS1	1694	1697	UGGA
4.1881759	90%	SFRS1	1716	1719	UGGA
4.62028767	100%	SFRS1	1816	1819	AGGA
4.62028767	100%	SFRS1	1819	1822	AGGA
4.62028767	100%	SFRS1	1837	1840	AGGA
4.62028767	100%	SFRS1	1988	1991	AGGA
4.62028767	100%	SFRS1	2011	2014	AGGA
4.62028767	100%	SFRS1	2014	2017	AGGA
4.1881759	90%	SFRS1	2035	2038	UGGA
4.1881759	90%	SFRS1	2054	2057	UGGA
4.1881759	90%	SFRS1	2094	2097	UGGA

4.1881759	90%	SFRS1	2105	2108	UGGA
4.62028767	100%	SFRS1	2113	2116	AGGA
4.62028767	100%	SFRS1	2174	2177	AGGA
4.1881759	90%	SFRS1	2195	2198	UGGA
4.1881759	90%	SFRS1	2436	2439	UGGA
4.62028767	100%	SFRS1	2495	2498	AGGA
4.62028767	100%	SFRS1	2590	2593	AGGA
4.62028767	100%	SFRS1	2595	2598	AGGA
4.62028767	100%	SFRS1	2630	2633	AGGA
4.567833375	89%	SFRS13A	7	13	AAAGGCU
4.53689248	88%	SFRS13A	314	320	AGAGAAC
4.6654368	91%	SFRS13A	596	602	AAAGGUC
4.365802746	85%	SFRS13A	656	662	AGAGGCC
4.357615916	85%	SFRS13A	697	703	AAAGGUG
4.236719035	82%	SFRS13A	889	895	CAAGAAC
4.213726735	82%	SFRS13A	1335	1341	CAAGGAC
4.631611151	90%	SFRS13A	1840	1846	AAAGGCG
4.52870565	88%	SFRS13A	2078	2084	AAAGAAA
4.52870565	88%	SFRS13A	2156	2162	AAAGAAA
4.52870565	88%	SFRS13A	2453	2459	AAAGAAA
4.715930809	92%	SFRS13A	2593	2599	AAAGGAU
4.50571335	88%	SFRS13A	2628	2634	AAAGGAA
6.63255189	93%	SFRS9	146	150	AGGAG
6.63255189	93%	SFRS9	328	332	AGGAG
7.08652094	100%	SFRS9	1337	1341	AGGAC
5.830231484	82%	SFRS9	1390	1394	AGCAC
7.08652094	100%	SFRS9	1448	1452	AGGAC
5.830231484	82%	SFRS9	1668	1672	AGCAC
6.63255189	93%	SFRS9	1816	1820	AGGAG
7.08652094	100%	SFRS9	1819	1823	AGGAC
6.63255189	93%	SFRS9	2011	2015	AGGAG
7.08652094	100%	SFRS9	2174	2178	AGGAC
7.08652094	100%	SFRS9	2495	2499	AGGAC
8.883996	87%	SNRPA	29	35	UGGAGAU
5.205931269	81%	Vts1	54	60	GCAGGUG
5.53207395	86%	Vts1	426	432	GCGGGGC
5.205931269	81%	Vts1	675	681	GCAGGUG
5.508778455	85%	Vts1	1749	1755	GCGGGUG
5.603586804	87%	Vts1	1906	1912	GCAGGCU
5.42581595	84%	Vts1	1910	1916	GCUGGGC
5.53207395	86%	Vts1	1932	1938	GCGGGGC
5.73001591	89%	Vts1	2445	2451	GCUGGCA
6.33890598	100%	YBX1	1402	1407	CCUGCG
6.33890598	100%	YBX1	1929	1934	CCUGCG

5.09233667	80%	YBX1	2198	2203	ACUGCG
7.945691304	84%	ybx2-a	1345	1350	AAGAUC
7.89298258	84%	ybx2-a	2022	2027	AACACC
5.8842352	94%	YTHDC1	191	196	GCAUGC
5.27877629	84%	YTHDC1	501	506	GACUGC
5.62897815	90%	YTHDC1	777	782	UCAUGC
5.62897815	90%	YTHDC1	1596	1601	UCAUGC
6.23570757	100%	YTHDC1	1971	1976	GAAUGC
5.27877629	84%	YTHDC1	2197	2202	GACUGC
5.98045052	95%	YTHDC1	2582	2587	UAAUGC
10.3039431	100%	ZRANB2	583	588	AGGUAA

B) RBP motifs within the wild-type T-allelic form of the *COL5A1* 3'-UTR

Score	Relative Score	RBP Name	Start	End	Matching sequence
8.6471013	100%	a2bp1	191	195	GCAUG
8.6471013	100%	a2bp1	823	827	GCAUG
8.6471013	100%	a2bp1	2122	2126	GCAUG
6.93292615	94%	ACO1	237	242	CAGUGA
7.35090902	100%	ACO1	961	966	CAGUGC
5.25664585	81%	EIF4B	69	72	GGAC
5.25664585	81%	EIF4B	234	237	GGAC
7.96844717	87%	EIF4B	297	303	GUGGGAC
5.25664585	81%	EIF4B	300	303	GGAC
6.4668404	100%	EIF4B	325	328	GGAA
5.25664585	81%	EIF4B	500	503	GGAC
5.25664585	81%	EIF4B	514	517	GGAC
6.4668404	100%	EIF4B	785	788	GGAA
5.25664585	81%	EIF4B	807	810	GGAC
6.4668404	100%	EIF4B	863	866	GGAA
6.4668404	100%	EIF4B	908	911	GGAA
6.4668404	100%	EIF4B	1022	1025	GGAA
6.4668404	100%	EIF4B	1151	1154	GGAA
5.25664585	81%	EIF4B	1338	1341	GGAC
5.25664585	81%	EIF4B	1449	1452	GGAC
6.4668404	100%	EIF4B	1769	1772	GGAA
5.25664585	81%	EIF4B	1820	1823	GGAC
6.4668404	100%	EIF4B	1838	1841	GGAA
6.4668404	100%	EIF4B	1900	1903	GGAA
7.59065781	80%	EIF4B	1986	1992	UUAGGAA
6.4668404	100%	EIF4B	1989	1992	GGAA
6.4668404	100%	EIF4B	2015	2018	GGAA
6.4668404	100%	EIF4B	2055	2058	GGAA
5.25664585	81%	EIF4B	2175	2178	GGAC
5.25664585	81%	EIF4B	2196	2199	GGAC
5.25664585	81%	EIF4B	2437	2440	GGAC
5.25664585	81%	EIF4B	2496	2499	GGAC
6.4668404	100%	EIF4B	2591	2594	GGAA
6.4668404	100%	EIF4B	2631	2634	GGAA
3.82636396	86%	ELAVL1	108	111	AUUU
3.82636396	86%	ELAVL1	114	117	AUUU
3.82636396	86%	ELAVL1	510	513	AUUU
3.82636396	86%	ELAVL1	523	526	AUUU
3.82636396	86%	ELAVL1	554	557	AUUU
4.40359056	100%	ELAVL1	626	629	GUUU
3.82636396	86%	ELAVL1	652	655	AUUU
4.40359056	100%	ELAVL1	709	712	GUUU

4.40359056	100%	ELAVL1	719	722	GUUU
3.82636396	86%	ELAVL1	732	735	AUUU
4.40359056	100%	ELAVL1	912	915	GUUU
3.82636396	86%	ELAVL1	933	936	AUUU
4.40359056	100%	ELAVL1	980	983	GUUU
3.82636396	86%	ELAVL1	1029	1032	AUUU
3.82636396	86%	ELAVL1	1037	1040	AUUU
4.40359056	100%	ELAVL1	1076	1079	GUUU
3.82636396	86%	ELAVL1	1086	1089	AUUU
3.82636396	86%	ELAVL1	1091	1094	AUUU
3.82636396	86%	ELAVL1	1160	1163	AUUU
3.82636396	86%	ELAVL1	1198	1201	AUUU
4.40359056	100%	ELAVL1	1241	1244	GUUU
3.82636396	86%	ELAVL1	1470	1473	AUUU
3.82636396	86%	ELAVL1	1540	1543	AUUU
4.40359056	100%	ELAVL1	1555	1558	GUUU
3.82636396	86%	ELAVL1	1568	1571	AUUU
4.40359056	100%	ELAVL1	1576	1579	GUUU
3.82636396	86%	ELAVL1	1691	1694	AUUU
3.82636396	86%	ELAVL1	2038	2041	AUUU
3.82636396	86%	ELAVL1	2047	2050	AUUU
4.40359056	100%	ELAVL1	2222	2225	GUUU
3.82636396	86%	ELAVL1	2228	2231	AUUU
3.82636396	86%	ELAVL1	2245	2248	AUUU
3.82636396	86%	ELAVL1	2265	2268	AUUU
3.82636396	86%	ELAVL1	2301	2304	AUUU
3.82636396	86%	ELAVL1	2382	2385	AUUU
3.82636396	86%	ELAVL1	2406	2409	AUUU
3.82636396	86%	ELAVL1	2411	2414	AUUU
4.40359056	100%	ELAVL1	2489	2492	GUUU
4.40359056	100%	ELAVL1	2542	2545	GUUU
4.40359056	100%	ELAVL1	2572	2575	GUUU
4.40359056	100%	ELAVL1	2650	2653	GUUU
4.40359056	100%	ELAVL1	2674	2677	GUUU
11.4274791	81%	ELAVL2	1087	1095	UUUUAUUUG
14.058233	100%	ELAVL2	2407	2415	UUUUAUUUU
12.7846426	100%	ELAVL2	2407	2415	UUUUAUUUU
7.3693752	100%	FUS	57	60	GGUG
7.3693752	100%	FUS	91	94	GGUG
7.3693752	100%	FUS	294	297	GGUG
7.3693752	100%	FUS	678	681	GGUG
7.3693752	100%	FUS	700	703	GGUG
7.3693752	100%	FUS	1170	1173	GGUG
7.3693752	100%	FUS	1191	1194	GGUG

7.3693752	100%	FUS	1231	1234	GGUG
7.3693752	100%	FUS	1420	1423	GGUG
7.3693752	100%	FUS	1519	1522	GGUG
7.3693752	100%	FUS	1561	1564	GGUG
7.3693752	100%	FUS	1752	1755	GGUG
7.3693752	100%	FUS	1873	1876	GGUG
7.3693752	100%	FUS	1923	1926	GGUG
7.3693752	100%	FUS	1942	1945	GGUG
7.3693752	100%	FUS	2400	2403	GGUG
7.3693752	100%	FUS	2671	2674	GGUG
9.8958058	100%	HNRNPA1	1015	1020	UAGGGA
4.09404991	88%	KHDRBS3	530	535	UCUAAA
3.82611654	82%	KHDRBS3	593	598	CGUAAA
3.73944112	80%	KHDRBS3	629	634	UUUAAC
4.0923883	87%	KHDRBS3	694	699	AUUAAA
3.87420975	83%	KHDRBS3	725	730	AGUAAA
4.27749466	91%	KHDRBS3	748	753	UAUAAA
4.32558787	92%	KHDRBS3	900	905	AAUAAA
4.37079607	93%	KHDRBS3	1010	1015	UUUAAU
4.44927358	95%	KHDRBS3	1033	1038	GAUAAU
4.14214313	89%	KHDRBS3	1064	1069	ACUAAA
4.41888929	94%	KHDRBS3	1082	1087	AUUAAU
3.81791863	82%	KHDRBS3	1173	1178	GAUAAC
4.32558787	92%	KHDRBS3	1490	1495	AAUAAA
4.26582884	91%	KHDRBS3	1564	1569	GCUAAU
4.20071074	90%	KHDRBS3	1624	1629	AGUAAU
4.60399565	98%	KHDRBS3	2297	2302	UAUAAU
4.32558787	92%	KHDRBS3	2364	2369	AAUAAA
4.15261753	89%	KHDRBS3	2402	2407	UGUAAU
4.60399565	98%	KHDRBS3	2430	2435	UAUAAU
4.12277259	88%	KHDRBS3	2483	2488	GAUAAA
3.73944112	80%	KHDRBS3	2574	2579	UUUAAC
4.15261753	89%	KHDRBS3	2580	2585	UGUAAU
4.04429508	86%	KHDRBS3	2602	2607	CUUAAA
4.09404991	88%	KHDRBS3	2612	2617	UCUAAA
4.65208886	100%	KHDRBS3	2633	2638	AAUAAU
3.78753434	81%	KHDRBS3	2637	2642	AUUAAC
4.32558787	92%	KHDRBS3	2657	2662	AAUAAA
3.82611654	82%	KHDRBS3	2682	2687	CGUAAA
6.33985	100%	KHSRP	212	215	GUCC
6.33985	100%	KHSRP	223	226	GUCC
6.33985	100%	KHSRP	403	406	GUCC
6.33985	100%	KHSRP	800	803	GUCC
6.33985	100%	KHSRP	1140	1143	GUCC

6.33985	100%	KHSRP	1271	1274	GUCC
6.33985	100%	KHSRP	1425	1428	GUCC
6.33985	100%	KHSRP	1712	1715	GUCC
6.33985	100%	KHSRP	1758	1761	GUCC
6.33985	100%	KHSRP	1948	1951	GUCC
6.33985	100%	KHSRP	2180	2183	GUCC
6.33985	100%	KHSRP	2249	2252	GUCC
6.33985	100%	KHSRP	2281	2284	GUCC
6.33985	100%	KHSRP	2324	2327	GUCC
6.6279899	100%	MBNL1	133	136	UGCU
6.17832025	93%	MBNL1	397	400	CGCU
6.6279899	100%	MBNL1	702	705	UGCU
6.6279899	100%	MBNL1	938	941	UGCU
6.17832025	93%	MBNL1	990	993	CGCU
6.6279899	100%	MBNL1	1094	1097	UGCU
6.17832025	93%	MBNL1	1107	1110	CGCU
6.6279899	100%	MBNL1	1193	1196	UGCU
6.6279899	100%	MBNL1	1233	1236	UGCU
6.6279899	100%	MBNL1	1296	1299	UGCU
6.17832025	93%	MBNL1	1406	1409	CGCU
6.6279899	100%	MBNL1	1563	1566	UGCU
6.17832025	93%	MBNL1	1606	1609	CGCU
6.6279899	100%	MBNL1	1944	1947	UGCU
6.17832025	93%	MBNL1	2068	2071	CGCU
6.17832025	93%	MBNL1	2444	2447	CGCU
6.17832025	93%	MBNL1	2555	2558	CGCU
8.9484945	100%	NONO	323	327	AGGGA
8.9484945	100%	NONO	498	502	AGGGA
8.9484945	100%	NONO	1016	1020	AGGGA
8.9484945	100%	NONO	1020	1024	AGGGA
8.9484945	100%	NONO	1149	1153	AGGGA
8.7178165	100%	PABPC1	570	574	AAAAA
8.7178165	100%	PABPC1	571	575	AAAAA
8.7178165	100%	PABPC1	1024	1028	AAAAA
8.7178165	100%	PABPC1	1025	1029	AAAAA
8.7178165	100%	PABPC1	1493	1497	AAAAA
7.15446438	84%	PABPC1	1996	2002	ACAAACA
8.7178165	100%	PABPC1	2451	2455	AAAAA
8.7178165	100%	PABPC1	2457	2461	AAAAA
8.7178165	100%	PABPC1	2458	2462	AAAAA
8.7178165	100%	PABPC1	2459	2463	AAAAA
8.7178165	100%	PABPC1	2472	2476	AAAAA
8.7178165	100%	PABPC1	2473	2477	AAAAA
8.7178165	100%	PABPC1	2547	2551	AAAAA

8.7178165	100%	PABPC1	2548	2552	AAAAA
8.24461074	96%	PTBP1	996	1010	AGACUUUUUUUUUU
7.2294196	100%	Pum2	551	554	UGUA
7.2294196	100%	Pum2	735	738	UGUA
7.2294196	100%	Pum2	740	743	UGUA
7.2294196	100%	Pum2	1166	1169	UGUA
7.2294196	100%	Pum2	1629	1632	UGUA
7.2294196	100%	Pum2	2304	2307	UGUA
7.2294196	100%	Pum2	2402	2405	UGUA
7.2294196	100%	Pum2	2536	2539	UGUA
7.2294196	100%	Pum2	2580	2583	UGUA
7.2294196	100%	Pum2	2622	2625	UGUA
7.2294196	100%	Pum2	2643	2646	UGUA
4.84132568	100%	RBM4	334	337	CGCG
4.84132568	100%	RBM4	2184	2187	CGCG
5.2682554	100%	RBMX	15	18	CCAG
4.99861593	94%	RBMX	62	65	CCAU
4.40271173	83%	RBMX	159	162	CCCG
4.40271173	83%	RBMX	167	170	CCCG
4.99861593	94%	RBMX	196	199	CCAU
4.40271173	83%	RBMX	214	217	CCCG
4.6667232	88%	RBMX	261	264	CCAC
4.6667232	88%	RBMX	288	291	CCAC
5.2682554	100%	RBMX	305	308	CCAG
4.40271173	83%	RBMX	309	312	CCCG
4.6667232	88%	RBMX	341	344	CCAC
5.2682554	100%	RBMX	376	379	CCAG
4.40271173	83%	RBMX	388	391	CCCG
4.6667232	88%	RBMX	394	397	CCAC
4.6667232	88%	RBMX	405	408	CCAC
4.99861593	94%	RBMX	411	414	CCAU
4.40271173	83%	RBMX	419	422	CCCG
4.99861593	94%	RBMX	432	435	CCAU
5.2682554	100%	RBMX	440	443	CCAG
5.2682554	100%	RBMX	448	451	CCAG
4.40271173	83%	RBMX	456	459	CCCG
4.99861593	94%	RBMX	462	465	CCAU
5.2682554	100%	RBMX	506	509	CCAG
4.6667232	88%	RBMX	606	609	CCAC
4.99861593	94%	RBMX	609	612	CCAU
4.99861593	94%	RBMX	650	653	CCAU
4.40271173	83%	RBMX	686	689	CCCG
5.2682554	100%	RBMX	791	794	CCAG
4.6667232	88%	RBMX	803	806	CCAC

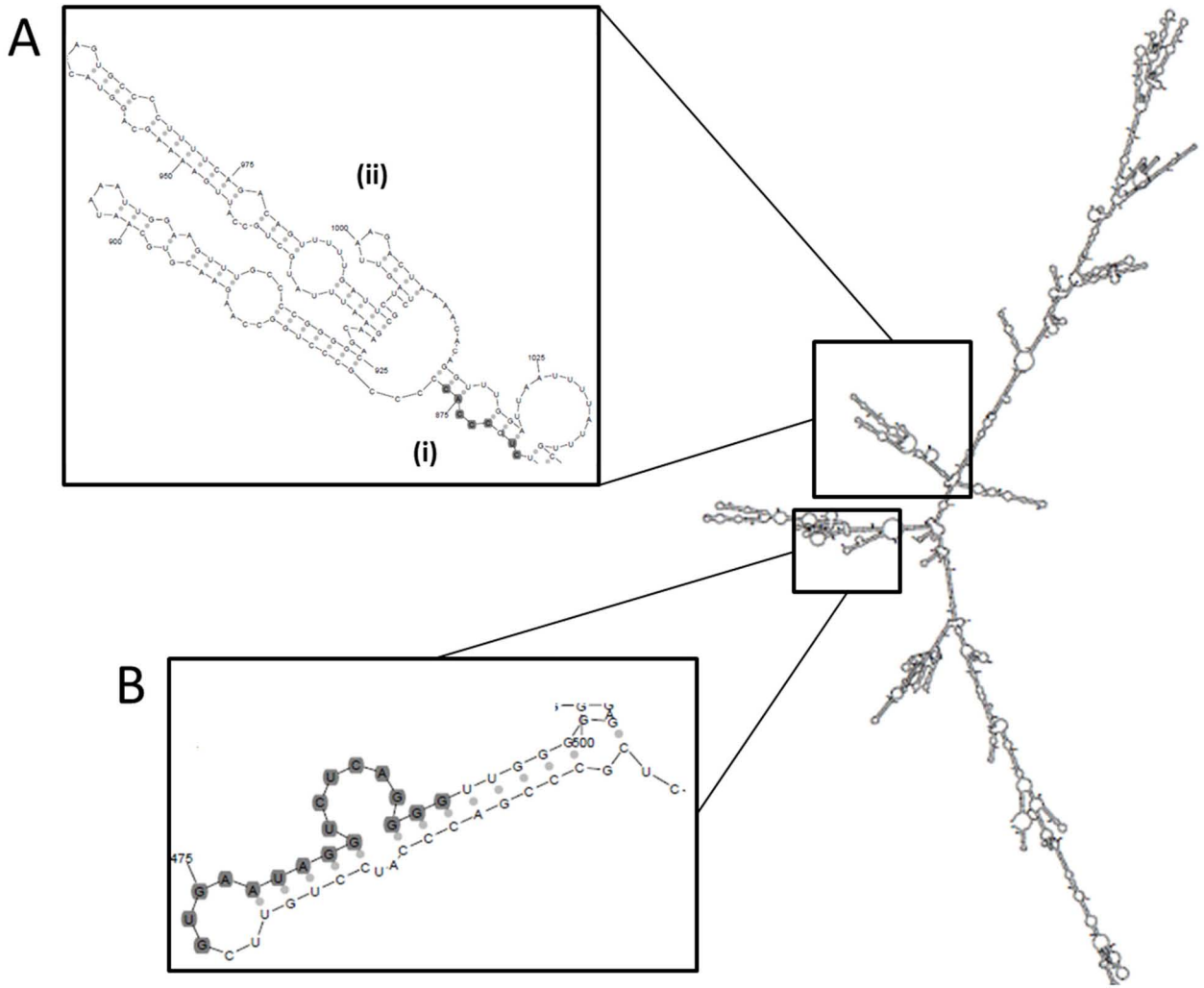
4.6667232	88%	RBMX	873	876	CCAC
5.2682554	100%	RBMX	878	881	CCAG
4.40271173	83%	RBMX	918	921	CCCG
4.99861593	94%	RBMX	943	946	CCAU
5.2682554	100%	RBMX	960	963	CCAG
4.40271173	83%	RBMX	1117	1120	CCCG
5.2682554	100%	RBMX	1142	1145	CCAG
4.6667232	88%	RBMX	1156	1159	CCAC
4.99861593	94%	RBMX	1279	1282	CCAU
4.6667232	88%	RBMX	1321	1324	CCAC
4.6667232	88%	RBMX	1371	1374	CCAC
4.40271173	83%	RBMX	1412	1415	CCCG
4.40271173	83%	RBMX	1417	1420	CCCG
4.6667232	88%	RBMX	1427	1430	CCAC
4.40271173	83%	RBMX	1432	1435	CCCG
4.99861593	94%	RBMX	1526	1529	CCAU
4.99861593	94%	RBMX	1601	1604	CCAU
4.99861593	94%	RBMX	1642	1645	CCAU
5.2682554	100%	RBMX	1654	1657	CCAG
4.99861593	94%	RBMX	1701	1704	CCAU
5.2682554	100%	RBMX	1737	1740	CCAG
4.99861593	94%	RBMX	1784	1787	CCAU
4.40271173	83%	RBMX	1823	1826	CCCG
5.2682554	100%	RBMX	1835	1838	CCAG
4.40271173	83%	RBMX	1869	1872	CCCG
4.99861593	94%	RBMX	1895	1898	CCAU
4.40271173	83%	RBMX	1951	1954	CCCG
4.6667232	88%	RBMX	1959	1962	CCAC
4.6667232	88%	RBMX	1962	1965	CCAC
4.6667232	88%	RBMX	1965	1968	CCAC
5.2682554	100%	RBMX	1968	1971	CCAG
5.2682554	100%	RBMX	1981	1984	CCAG
4.6667232	88%	RBMX	1994	1997	CCAC
4.6667232	88%	RBMX	2005	2008	CCAC
5.2682554	100%	RBMX	2009	2012	CCAG
5.2682554	100%	RBMX	2029	2032	CCAG
5.2682554	100%	RBMX	2111	2114	CCAG
5.2682554	100%	RBMX	2119	2122	CCAG
4.99861593	94%	RBMX	2165	2168	CCAU
4.40271173	83%	RBMX	2182	2185	CCCG
4.40271173	83%	RBMX	2191	2194	CCCG
4.99861593	94%	RBMX	2238	2241	CCAU
4.40271173	83%	RBMX	2326	2329	CCCG
5.2682554	100%	RBMX	2336	2339	CCAG

4.6667232	88%	RBMX	2341	2344	CCAC
4.6667232	88%	RBMX	2350	2353	CCAC
5.2682554	100%	RBMX	2588	2591	CCAG
4.6667232	88%	RBMX	2678	2681	CCAC
8.6696024	100%	RBMY1A1	1178	1182	CUCAA
8.6272192	100%	RBMY1A1	1178	1182	CUCAA
7.65633614	88%	RBMY1A1	1483	1487	CACAA
8.6696024	100%	RBMY1A1	1705	1709	CUCAA
8.6272192	100%	RBMY1A1	1705	1709	CUCAA
7.65633614	88%	RBMY1A1	1995	1999	CACAA
8.6696024	100%	RBMY1A1	2421	2425	CUCAA
8.6272192	100%	RBMY1A1	2421	2425	CUCAA
7.5622424	86%	sap-49	713	718	UUGUGA
8.7846348	100%	sap-49	1520	1525	GUGUGA
4.1881759	90%	SFRS1	29	32	UGGA
4.1881759	90%	SFRS1	50	53	UGGA
4.1881759	90%	SFRS1	68	71	UGGA
4.1881759	90%	SFRS1	111	114	UGGA
4.62028767	100%	SFRS1	146	149	AGGA
4.1881759	90%	SFRS1	233	236	UGGA
4.62028767	100%	SFRS1	328	331	AGGA
4.1881759	90%	SFRS1	346	349	UGGA
4.1881759	90%	SFRS1	513	516	UGGA
4.1881759	90%	SFRS1	691	694	UGGA
4.1881759	90%	SFRS1	826	829	UGGA
4.62028767	100%	SFRS1	862	865	AGGA
4.1881759	90%	SFRS1	907	910	UGGA
4.1881759	90%	SFRS1	1079	1082	UGGA
4.62028767	100%	SFRS1	1337	1340	AGGA
4.62028767	100%	SFRS1	1380	1383	AGGA
4.62028767	100%	SFRS1	1448	1451	AGGA
4.1881759	90%	SFRS1	1621	1624	UGGA
4.1881759	90%	SFRS1	1694	1697	UGGA
4.1881759	90%	SFRS1	1716	1719	UGGA
4.62028767	100%	SFRS1	1816	1819	AGGA
4.62028767	100%	SFRS1	1819	1822	AGGA
4.62028767	100%	SFRS1	1837	1840	AGGA
4.62028767	100%	SFRS1	1988	1991	AGGA
4.62028767	100%	SFRS1	2011	2014	AGGA
4.62028767	100%	SFRS1	2014	2017	AGGA
4.1881759	90%	SFRS1	2035	2038	UGGA
4.1881759	90%	SFRS1	2054	2057	UGGA
4.1881759	90%	SFRS1	2094	2097	UGGA
4.1881759	90%	SFRS1	2105	2108	UGGA

4.62028767	100%	SFRS1	2113	2116	AGGA
4.62028767	100%	SFRS1	2174	2177	AGGA
4.1881759	90%	SFRS1	2195	2198	UGGA
4.1881759	90%	SFRS1	2436	2439	UGGA
4.62028767	100%	SFRS1	2495	2498	AGGA
4.62028767	100%	SFRS1	2590	2593	AGGA
4.62028767	100%	SFRS1	2595	2598	AGGA
4.62028767	100%	SFRS1	2630	2633	AGGA
4.56783337	89%	SFRS13A	7	13	AAAGGCU
4.53689248	88%	SFRS13A	314	320	AGAGAAC
4.6654368	91%	SFRS13A	596	602	AAAGGUC
4.36580275	85%	SFRS13A	656	662	AGAGGCC
4.35761592	85%	SFRS13A	697	703	AAAGGUG
4.23671904	82%	SFRS13A	889	895	CAAGAAC
4.21372674	82%	SFRS13A	1335	1341	CAAGGAC
4.63161115	90%	SFRS13A	1840	1846	AAAGGCG
4.52870565	88%	SFRS13A	2078	2084	AAAGAAA
4.52870565	88%	SFRS13A	2156	2162	AAAGAAA
4.52870565	88%	SFRS13A	2453	2459	AAAGAAA
4.71593081	92%	SFRS13A	2593	2599	AAAGGAU
4.50571335	88%	SFRS13A	2628	2634	AAAGGAA
6.63255189	93%	SFRS9	146	150	AGGAG
6.63255189	93%	SFRS9	328	332	AGGAG
7.08652094	100%	SFRS9	1337	1341	AGGAC
5.83023148	82%	SFRS9	1390	1394	AGCAC
7.08652094	100%	SFRS9	1448	1452	AGGAC
5.83023148	82%	SFRS9	1668	1672	AGCAC
6.63255189	93%	SFRS9	1816	1820	AGGAG
7.08652094	100%	SFRS9	1819	1823	AGGAC
6.63255189	93%	SFRS9	2011	2015	AGGAG
7.08652094	100%	SFRS9	2174	2178	AGGAC
7.08652094	100%	SFRS9	2495	2499	AGGAC
8.883996	87%	SNRPA	29	35	UGGAGAU
5.20593127	81%	Vts1	54	60	GCAGGUG
5.53207395	86%	Vts1	426	432	GCGGGGC
5.20593127	81%	Vts1	675	681	GCAGGUG
5.50877846	85%	Vts1	1749	1755	GCGGGUG
5.6035868	87%	Vts1	1906	1912	GCAGGCU
5.42581595	84%	Vts1	1910	1916	GCUGGGC
5.53207395	86%	Vts1	1932	1938	GCGGGGC
5.73001591	89%	Vts1	2445	2451	GCUGGCA
6.33890598	100%	YBX1	1402	1407	CCUGCG
6.33890598	100%	YBX1	1929	1934	CCUGCG
5.09233667	80%	YBX1	2198	2203	ACUGCG

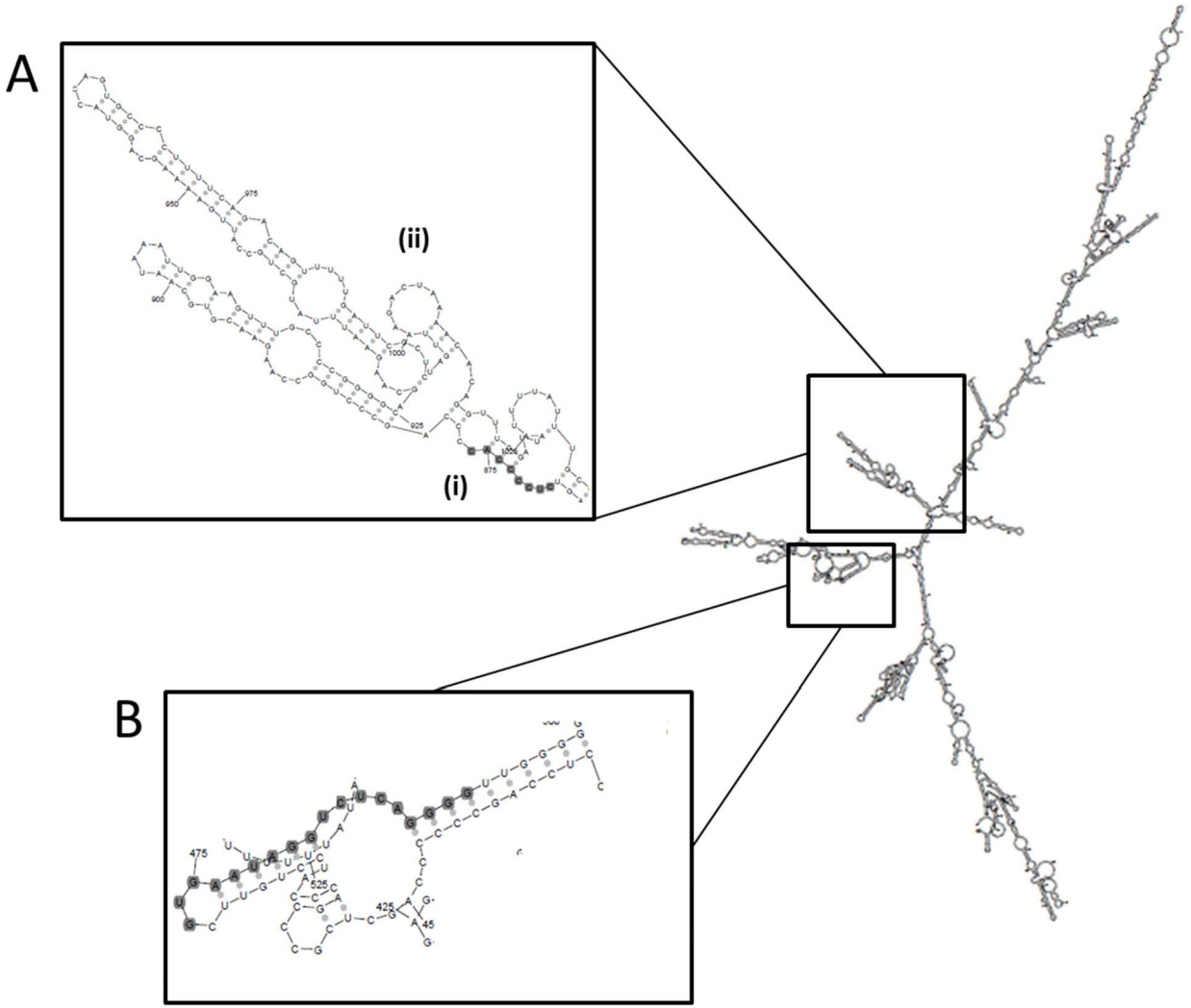
7.9456913	84%	ybx2-a	1345	1350	AAGAUC
7.89298258	84%	ybx2-a	2022	2027	AACACC
5.8842352	94%	YTHDC1	191	196	GCAUGC
5.27877629	84%	YTHDC1	501	506	GACUGC
5.62897815	90%	YTHDC1	777	782	UCAUGC
5.62897815	90%	YTHDC1	1596	1601	UCAUGC
6.23570757	100%	YTHDC1	1971	1976	GAAUGC
5.27877629	84%	YTHDC1	2197	2202	GACUGC
5.98045052	95%	YTHDC1	2582	2587	UAAUGC
10.3039431	100%	ZRANB2	583	588	AGGUAA

Appendix S: RNA secondary structure of C-allelic 57 bp deletion sequence



Appendix T- (a) Diagram showing secondary structure of 2.5 kb Δ57 bp of the C-allelic form of the *COL5A1* 3'-UTR. Insert A shows region containing miR-608 binding site (i), MboII octamer and the AGGG VNTR have been removed (ii). Insert B contains miR125a-5p binding site .

B) RNA secondary structure of T-allelic 57 bp deletion sequence



Appendix T- (b) Diagram showing secondary structure of 2.5 kb Δ57 bp of the T-allelic form of the *COL5A1* 3'-UTR. Insert A shows region containing miR-608 binding site (i), the *Mbo*II octamer and the AGGG VNTR have been removed (ii). Insert B contains miR125a-5p binding site .