Association of Type XI Collagen Genes with Chronic Achilles Tendinopathy in Independent Populations from South Africa and Australia

Melanie Hay[1], Jon Patricios[2,3], Robert Collins[3,4], Andy Branfield[4], Jill Cook[5], Christopher J Handley[6], Alison V September[1], Michael Posthumus[1], Malcolm Collins[7,1]

[1] MRC/UCT Research Unit for Exercise Science and Sports Medicine, Department of Human Biology, University of Cape Town, Cape Town, South Africa. [2]Morningside Sports Medicine, Johannesburg, South Africa. [3]Section of Sports Medicine, Faculty of Health Sciences, University of Pretoria, Pretoria, South Africa. [4]The Centre for Sports Medicine & Orthopaedics, Johannesburg, South Africa. [5]Department of Physiotherapy, Monash University, Melbourne, Victoria, Australia. [6]School of Public Health and Human Biosciences, La Trobe University, Melbourne, Victoria, Australia. [7]South African Medical Research Council, Cape Town, South Africa

Address for correspondence:

Prof M Collins

UCT/MRC Research Unit for Exercise Science and Sports Medicine

P.O. Box 115, Newlands 7725, South Africa

Tel: +27-21-650 4574

Fax: +27-21-686 7530

Email: malcolm.collins@uct.ac.za

Word Count: 2705

Running title: Type XI collagen genes and Achilles tendinopathy.

ABSTRACT

Background: . Type XI collagen, which is expressed in developing tendons and is encoded by the *COL11A1*, *COL11A2* and *COL2A1* genes, shares structural and functional homology with type V collagen, which plays an important role in collagen fibril assembly. We investigated the association of these three polymorphisms with Achilles tendinopathy (AT) and whether these polymorphisms interact with *COL5A1* to modulate the risk of AT. **Methods**: 184 participants diagnosed with chronic AT (TEN) and 338 appropriately matched asymptomatic controls (CON) were genotyped for the three polymorphisms. **Results**: Although there were no independent associations with AT, the TCT pseudohaplotypte constructed from rs3753841 (T/C), rs1676486 (C/T) and rs1799907 (T/A) was significantly over-represented (p=0.006) in the TEN (25.9%) compared to the CON (17.1%) group. The TCT(AGGG) pseudohaplotypes constructed using these type XI collagen polymorphisms and the functional *COL5A1* rs71746744 (-/AGGG) polymorphism was also significantly over-represented (p<0.001) in the TEN (25.2%) compared to the CON (9.1%) group. **Discussion**: Genes encoding structural and functionally related type XI (*COL11A1* and *COL11A2*) and type V (*COL5A1*) collagens interact with one another to collectively modulate the risk for Achilles tendinopathy. Although there are no immediate clinical applications, the results of this study provide additional evidence that inter-individual variations in collagen fibril assembly might be an important molecular mechanism in the etiology of chronic Achilles tendinopathy.

Key words: *COL11A1*, *COL11A2*, *COL5A1*, type V collagen, tendon, injury

INTRODUCTION

Injury to the mid-substance of the Achilles tendon is a multi-factorial condition, resulting from the poorly understood interactions of several extrinsic and intrinsic risk factors. [1] The genetic profile of an individual is an important intrinsic risk factor. [2] Although DNA sequence variants within several genes have been independently associated with risk of chronic Achilles tendinoapthy [2 3], and one of the most extensively studied genes to date is *COL5A1*. [4 5 6 7] *COL5A1* encodes the α1 chain of type V collagen, which plays an important role in collagen fibril nucleation and the regulation of fibril diameter (fibrillogenesis). [8] Several variants within the 3'-untranslated region (3'-UTR) of *COL5A1* have been associated with Achilles tendinopathy. [5 7] These variants are believed to be functional, altering *COL5A1* mRNA stability within the cytoplasm of the tenocyte. [6] Increased mRNA stability has been associated with tendinoapthy, which is believed to result in increased α1(V) chain and type V collagen production, decreasing the fibril diameter and packing density, and potentially altering the mechanical properties of the tendon. [9]

Type XI collagen, a heterotrimer consisting of α1(XI), α2(XI) and α3(XI) chains encoded by the *COL11A1* (chromosome 1p21), *COL11A2* (chromosome 6p21.3) and *COL2A1* (chromosome 12q13.11) genes respectively [10], shares structural and functional homology with type V collagen. [11] Although predominately expressed in cartilage [11], type XI collagen is also expressed in developing tendons. [12]

Mutations in all three genes have been implicated in various inherited Mendelian connective tissue disorders [13] and polymorphisms within these genes have also

been associated with multifactorial musculoskeletal injuries and connective tissue disorders. [14 15 16-21]

Single nucleotide polymorphisms (SNPs), namely *COL11A1* rs3753841 (T/C), *COL11A1* rs1676486 (C/T) and/or *COL11A2* rs1799907 (T/A), have been associated with lumbar disc herniation (LDH) [18], limbus vertebra in gymnasts [21], lumbar spine stenosis [17], ossification of the posterior longitudinal ligament of the spine (OPLL) [14-16], and rheumatoid arthritis. [19]

Therefore, we investigated the association of the *COL11A1* polymorphisms rs3753841 and rs1676486, as well as the *COL11A2* rs1799907 polymorphism with Achilles Tendinopathy in a South African and an Australian population. A secondary aim of this study was to investigate whether the *COL11A1* and *COL11A2* polymorphisms interact with the previously investigated rs71746744 (-/AGGGG) polymorphism within the *COL5A1* 3'-UTR [7] to modulate the risk of Achilles Tendinopathy.

METHODS

Participants and DNA extraction

A total of 267 unrelated, physically active participants from South Africa, consisting of 161 asymptomatic controls (SA CON) and 106 individuals diagnosed with chronic Achilles tendinopathy (SA TEN) were recruited as previously described. [4 22] The SA TEN group were recruited from Sports Medicine clinical practices in Cape Town and Johannesburg, South Africa. The diagnosis was confirmed by a Sports Physician,

using the previously described inclusion and exclusion criteria. [22] The SA CON participants reported no history of tendon pathology and were recruited from various recreational sporting clubs. A total of 255 participants from Australia, consisting of 177 asymptomatic controls (AUS CON) and 78 patients with chronic Achilles tendinopathy (AUS TEN) were recruited as previously described. [5] The diagnosis of the tendinopathy participants, who were recruited from the Musculoskeletal Research Centre at La Trobe University in Melbourne Australia, was confirmed using the same clinical criteria by a Sports Physiotherapist. [5]

All participants were of self-reported Caucasian European ancestory. They gave written informed consent and completed questionnaires concerning their medical history and involvement in physical activity. This study was approved by the Human Research Ethics Committee at the University of Cape Town, South Africa, and the Human Ethics Committees of La Trobe and Deakin Universities in Melbourne, Australia. Total DNA was extracted from approximately 4.5 ml of venous blood as previously described. [5]

COL11A1 and *COL11A2* Genotyping

The *COL11A1* rs3753841 (T/C) and rs1676486 (C/T) non-synonymous SNPs, as well as the *COL11A2* rs1799907 (IVS6-4, T/A) SNP, were genotyped using flourescence-based Taqman® PCR assays (Applied Biosystems, Foster City, CA, USA). Inventoried allele specific probes and flanking primer sets were used along with a pre-made PCR mastermix containing ampliTaq® DNA polymerase Gold (Applied Biosystems, Foster City, CA, USA) in a final reaction volume of 8 µl. The PCR reactions were conducted on an Applied Biosystems StepOnePlus™ Real-

Time PCR system (Applied Biosystems), using the Applied Biosystems Step-OnePlus™ Real-Time PCR software Version 2.2.2 (Applied Biosystems) following the manufacturers recommended cycling conditions.

Statistical Analysis

Quanto version 1.2.4, was used to determine the statistical power for a given sample size and minor allele frequency. [23] Assuming a minor allele frequency of 0.36, 0.20 and 0.26 for rs3753841, rs1676486 and rs1799907 respectively, a sample size of approximately 180 cases and 360 controls, would be adequate to detect a genetic effect size of at least 2.25, 1.85 and 1.95 respectively, at a power of 80% and a significance level of 5%.

Data were analysed using STATISTICA Version 10.0 (Stat- Soft, Tulsa, OK, USA) and GraphPad Prism version 5.0d for Mac OS X (GraphPad Software, San Diego, CA, USA) programs. 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, SA and combined cohorts. A $\chi 2$-analysis or Fisher's exact test was used to analyse any differences in the genotype frequencies and other categorical data between the groups. Hardy-Weinberg equilibrium (HWE) was established using the program Genepop web version 3.4 (http://genepop.curtin.edu.au/). Linkage disequilibrium (LD) between the polymorphisms within *COL11A1* was calculated using CubeX: cubic exact solution (http://www.oege.org/software/cubex/). [24] The Chaplin case-control haplotype inference package was used to infer haplotypes and pseudohaplotypes. [25] [26] Statistically significant differences were accepted when $p<0.05$.

RESULTS

Participant Characteristics

The SA CON and SA TEN groups were matched for height, sex and country of birth (Table 1). The SA TEN group was however significantly older (p=0.007) and heavier (weight: p=0.002; BMI: p=0.003) than the SA CON group. The mean difference between age of injury and age of recruitment of the SA TEN group was 7.1 ± 8.9 years. There was therefore also a significant difference between the age at recruitment of the SA TEN (48.2 ± 11.6 years, n=99) and the SA CON (36.4 ± 10.8 years, n=154, p<0.001) groups. When adjusted for age of recruitment, weight and BMI were no longer significantly different.

The AUS CON and AUS TEN groups were similarly matched for age and country of birth (Table 1). There was significantly more males in the AUS CON (40.3%) compared to the AUS TEN (71.8%, p<0.001) groups. After adjusting for sex, the groups were matched for height. The mean difference between age of injury and age of recruitment of the AUS TEN group was 9.0 ± 10.1 years. There was therefore a significant difference between the age at recruitment of the AUS TEN (49.6 ± 12.8 years, n=78) and the AUS CON (39.4 ± 12.3 years, n=174, p<0.001) groups. After adjusting for both age at recruitment and sex, the AUS CON and AUS TEN groups were also matched for weight and BMI (Table 1).

Table 1: Characteristics of the South African (SA) and Australian (AUS) chronic Achilles tendinopathy (TEN) groups and their respective control (CON) groups.

	SA CON	SA TEN	p value	AUS CON	AUS TEN	p value
Age (years)	36.4 ± 10.8 (154)	40.9 ± 14.8 (92)	0.007	39.4 ± 12.3 (174)	40.7 ± 14.5 (77)	0.487
Sex (%male)	63.8 (160)	67.6 (105)	0.518	40.3 (176)	71.8 (78)	<0.001
Height (cm)	175 ± 9 (155)	176 ± 9 (92)	0.345	171 ± 9 (175)	174 ± 10 (75)	0.059 0.101 [b]
Weight (kg)	72.2 ± 11.9 (159)	77.2 ± 13.4 (97)	0.002 0.117 [a]	72.7 ± 14.4 (176)	79.7 ± 13.4 (78)	<0.001 0.343 [c]
BMI (kg/cm^2)	23.6 ± 2.8 (151)	24.8 ± 3.3 (81)	0.003 0.126 [a]	24.7 ± 3.9 (175)	26.2 ± 3.5 (75)	0.003 0.385 [c]
Country of birth (% Australian)	n.d.	n.d.	n.d.	80.9 (173)	76.6 (77)	0.497
Country of birth (% South African)	73.6 (159)	74.3 (101)	1.000	n.d.	n.d.	n.d.

Participants were included in the analysis if they were genotyped at least once for rs3753841, rs1676486 or rs1799907. Variables are expressed as mean ± standard deviation, except for sex and country of birth, which are represented as a percentage (%). The number of participants for which data was available is in parenthesis (n). Age of CON groups = age of recruitment; age of TEN groups = age of initial onset of symptoms. Weight, height and BMI values are those reported at recruitment. BMI, body mass index; n.d., not determined.

[a] Covaried for age of recruitment.

[b] Covaried for sex.

[c] Covaried for age of recruitment and sex.

When the SA and AUS groups were combined, there were significantly more males in the TEN group (69.4%) compared to the CON group (51.5%, p<0.001). The TEN group (age of initial injury: 40.8 ± 14.7 years, weight: 78.3 ± 13.4 kg) was also significantly older and heavier than the CON group (age of recruitment: 38.0 ± 11.6 years, p=0.022; weight: 72.5 ± 13.2 kg, adjusted p=0.049). The groups were however matched for height (TEN: 175 ± 10 cm, CON: 173 ± 9 cm, adjusted p=0.419) and BMI (TEN: 25.5 ± 3.5 kg/cm^2, CON: 24.1 ± 3.5 kg/cm^2, adjusted p=0.214).

There were no significant *COL11A1* or *COL11A2* genotype effects on any of the combined SA and AUS participant characteristics (data not shown).

COL11A1 and *COL11A2* genotype and allele frequency distributions

There were no significant differences in any of the genotype or allele frequency distributions between the SA CON and SA TEN or the AUS CON and AUS TEN groups for *COL11A1* rs3753841 (T/C), *COL11A1* rs1676486 (C/T) and *COL11A2* rs1799907 (T/A) (Supplementary Table 1). In addition, there were no significant differences in the genotype distribution of these three polymorphisms between the participants from South Africa and Australia (Supplementary Table 1). For this reason, the two populations were combined. As illustrated in figure 1, there were no significant differences (p>0.100) in the genotype frequency distributions between CON and TEN groups of the combined Australian and South African cohorts when any of the three polymorphisms were analysed. There was also no significant difference in the *COL11A1* rs3753841 (p=0.121, Figure 1C) and rs1676486 (p=0.221, Figure 1D) allele frequency distributions between combined CON and TEN

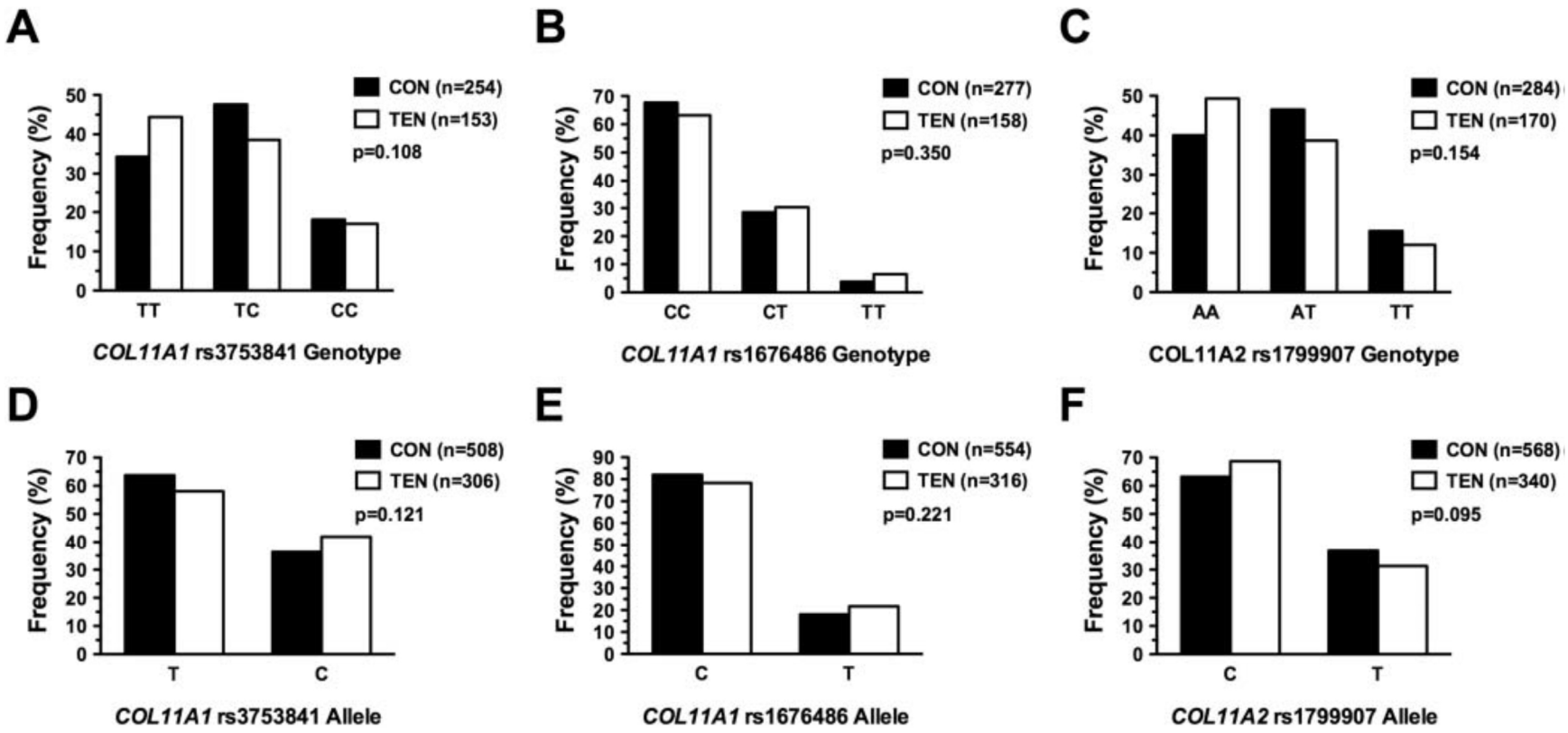


Figure 1

Genotype and allele frequency distributions of the combined South African and Australian CON (black bars) and TEN (clear bars) groups for *COL11A1* rs3753841 (T/C), *COL11A1* rs1676486 (C/T) and *COL11A2* rs1799907 (T/A). **(A)** rs3753841 genotype distribution, CON HWE=0.795, TEN HWE= 0.052. **(B)** rs1676486 genotype distribution, CON HWE=0.082, TEN HWE=0.602. **(C)** rs1799907 genotype frequency distribution CON HWE=1.000, TEN HWE=0.100. **(D)** rs3753841 allele frequency distribution. **(E)** rs1676486 allele frequency distribution. **(F)** rs1799907 allele frequency distribution. The number of genotyped participants (n) for each polymorphism is indicated in parenthesis in graphs A to C. The number of alleles (n, 2 for each participant) for each polymorphism is indicated parenthesis in graphs D to F. The p value is also indicated in each graph.

groups. However a tendency for the minor T allele of *COL11A2* rs1799907 (p=0.095) to be over-represented in the combined TEN group was noted (Figure 1F). Although there was a tendency for some to deviate, all the groups were in HWE (Table 1 and Figure 1).

COL11A1 and *COL11A2* interactions

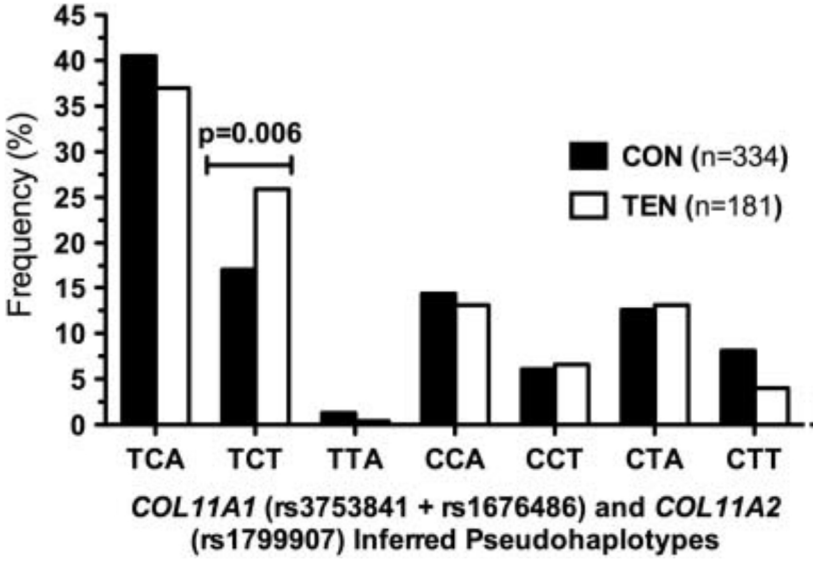


Figure 2

The frequency of seven of the possible eight inferred pseudohaplotypes with a frequency >0.4% constructed from *COL11A1* rs3753841 (T/C), *COL11A1* rs1676486 (C/T) and *COL11A2* rs1799907 (T/A) in the combined Australian and South African Achilles tendinopathy (TEN, clear bars) and control (CON, solid bars) groups. The p value of the significantly different TCT pseudohaplotype is indicated. The total number (n) of pseudohaplotypes within the CON and TEN groups is indicated in parenthesis in the graph.

Polymorphisms *COL11A1* rs3753841 and rs1676486 were in high linkage disequilibrium for the CON (SA D'=0.862; AUS D'=0.911) and TEN (SA D'=1.000; AUS D'=0.898) groups, as well as the combined group (D'=0.932). Since the three polymorphisms investigated in this study are located within type XI collagen genes,

inferred pseudohaplotypes were constructed. Seven of the possible eight pseudohaplotypes, were inferred with a frequency greater than 0.4% from the two *COL11A1* and one *COL11A2* polymorphisms (Figure 2). The TCT pseudohaplotype was significantly over-represented (p=0.006) in the TEN group (25.9%, n=47) compared to the CON group (17.1%, n=57) when the combined populations were investigated (Figure 2). Similar pseudohaplotype distributions were obtained when the SA and AUS cohorts were analysed seperately, with the TCT pseudohaplotype being significantly over-represented in the TEN group of both the SA (CON 19.6% vs TEN 29.7%, p=0.017) and AUS (CON 12.9% vs TEN 21.3%, p=0.041) cohorts (Supplementary Figure 1).

Interactions between *COL11A1*, *COL11A2* and *COL5A1*

Type V and type XI collagen both regulate fibrillogenesis, therefore inferred pseudohaplotypes were constructed between the *COL11A1* and *COL11A2* polymorphisms investigated in this study, and the previously investigated *COL5A1* rs71746744 (-/AGGG) polymorphism [7]. Eleven of the possible 16 pseudohaplotypes were inferred with a frequency greater than 2% from the *COL11A1*, *COL11A2* and *COL5A1* polymorphisms (Figure 3). The TCT(AGGG) pseudohaplotype was significantly over-represented (p=0.003) in the TEN group (25.2%, n=46) compared to the CON group (9.1%, n=30) (Figure 3). Similar pseudohaplotype distrubitions were obtained when the SA and AUS cohorts were analysed seperately, with the TCT(AGGG) pseudohaplotype being significantly over-represented in the TEN group of both the SA (CON 10.0% vs TEN 29.4%, p=0.003) and AUS (CON 7.2% vs TEN 20.3%, p=0.008) cohorts (Supplementary Figure 2).

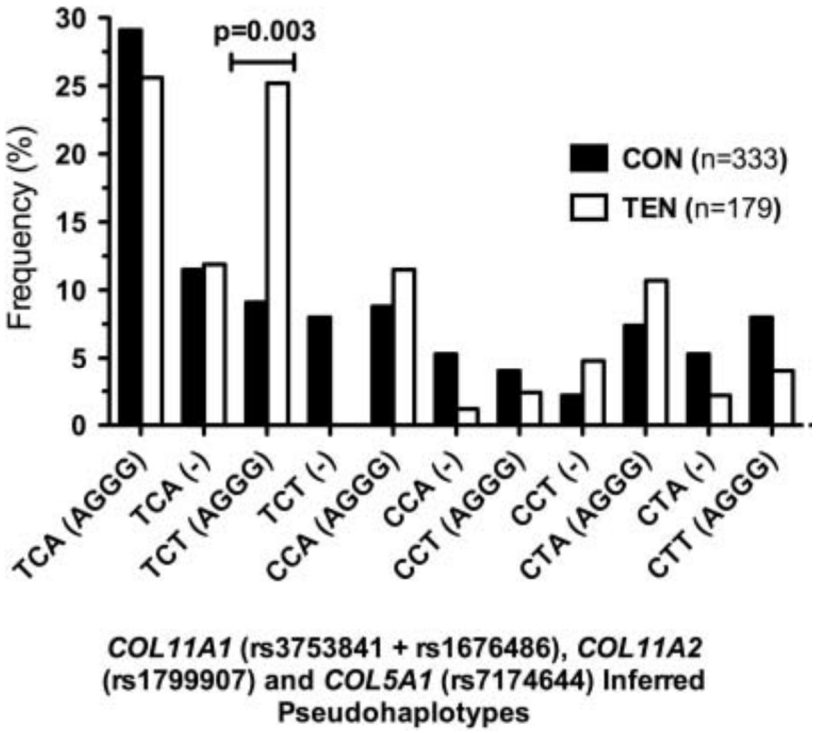


Figure 3

The frequency of eleven of the possible 16 inferred pseudohaplotypes with a frequency >2% constructed from *COL11A1* rs3753841 (T/C), *COL11A1* rs1676486 (C/T), *COL11A2* rs1799907 (T/A) and *COL5A1* rs71746744 (-/AGGG) in the combined Australian and South African Achilles tendinopathy (TEN, clear bars) and control (CON, solid bars) groups. The p value of the significantly different TCT(AGGG) pseudohaplotype is indicated. The total number (n) of pseudohaplotypes within the CON and TEN groups is indicated in parenthesis in the graph.

DISCUSSION

Polymorphisms in several genes, including *COL5A1*, have been identified as intrinsic risk factors for Achilles tendinopathy. [2] *COL5A1* encodes the α1 chain of type V collagen, which plays an important role in collagen fibril assembly; the basic building block of tendons. [8]. Type XI collagen, which is encoded by the *COL11A1*, *COL11A2* and *COL2A1* genes [10], shares structural and functional homology with type V collagen. [11]

None of the three investigated polymorphisms within the *COL11A1* (rs3753841 and rs1676486) and *COL11A2* (rs1799907) genes were independently associated with chronic Achilles tendinopathy in the Australian, South African or combined cohorts. The genotype and allele frequencies calculated for these three SNPs were similar to the reported frequencies presented in the NCBI SNP database (http://www.ncbi.nlm.nih.gov/snp/).

The main finding of this study was the association of the TCT inferred pseudohaplotype, constructed from the three polymorphisms within *COL11A1* (rs3753841 T/C and rs1676486 C/T) and *COL11A2* (rs17999079 T/A), with increased risk of chronic Achilles tendinopathy. The non-synonymous rs3753841 polymorphism within exon 52 of *COL11A1*, results in a predicted deleterious substitution of a leucine for a proline at amino acid position 1323 of the α1(XI) chain. [27] The rs1676486 polymorphism within exon 62, also results in an amino acid substitution, proline to serine at position 1535, which could potentially cause a conformational change in type XI collagen. [18] In addition, the T allele of rs1676486 also appears to be associated with increased mRNA degradation. [18]

Furthermore, the T allele of *COL11A2* rs1799907 (IVS6-4, T/A), which was also implicated in the pseudohaplotypes, produces a distinct isoform of the α2(XI) chain in which several amino acids in the amino terminal acidic domain are deleted. This acidic domain provides potential sites for the interaction of type XI collagen with other molecules and may prevent further deposition of collagen molecules in the fibril. [15] [16] It is therefore reasonable to hypothesise that the risk associated

pseudohaplotype has a biological consequence. Although expressed in the developing tendon, further research is however required to determine the functional significance of the type XI collagen gene polymorphisms in tendinopathy.

Intriguingly, while the minor C allele of *COL11A1* rs3753841 and the minor T allele of *COL11A1* of rs1676486 was over-represented in patients with lumber disc hermiation, the current study however implicated the alternate alleles in the risk-associated pseudohaplotype. Similarly, the functional rare TT genotype of the *COL1A1* Sp1 binding site polymorphism was reported to be associated with risk for several multifactorial disorders, including lumber disc disease [28-31], but was associated with protection in anterior cruciate ligament ruptures in three independent populations. [32-34] These findings suggest that both the alternative alleles within a gene can be associated with increased risk of different multifactorial disorders.

An additional finding of this study was the additive contribution of the AGGG allele of the *COL5A1* -/AGGG polymorphism together with the TCT inferred pseudohaplotype of the *COL11A1 and COL11A2* genes increased risk of chronic Achilles tendinopathy. This gene-gene association is perhaps not surprising taking into account the mounting evidence that the structural and functional homology between the α1(XI) and α1(V) chains encoded by *COL11A1* and *COL5A1*, respectively, facilitate an interchangeability between the two polypeptides resulting in heterotrimer [$\alpha1(XI)_2\alpha2(V)$] formation. [35 36 37 38] It is possible that the reported interaction in this study between the *COL5A1* and *COL11A1* polymorphisms indicate the role of a minor fibrillar collagen consisting of type V/XI α-chains in tendinopathy. Alternatively the traditional type XI collagen heterotimer may functionally replace or compensate

for type V collagen in tendons. The possible role of type XI collagen in pathology of the mature tendon needs to be investigated.

Although we are mindful of the theoretical biological effects of the amino acid substitutions implicated by the risk associated pseodohaplotype within the α1(XI) and α2(XI) chains, it is nevertheless tempting to speculate that the association between the *COL11A1*, *COL11A2* and *COL5A1* genes and increased risk of chronic Achilles tendinoathy is due to their collective effect on (i) mRNA stability, (ii) potential effects on types V and XI collagen production, regulation of collagen fibril diameter, and (iv) altered biomechanical properties of the collagen fibril that is reflected at the tissue level (Figure 4). [9] It is therefore reasonable to conceive that individuals who have these functional alleles as reflected from the peodohaplotypes could possibly be producing a functionally altered types XI and type V collagen which collectively are responsible for the altered biomechanical property of the tendon collagen fibrils. As previously mentioned although type XI collagen is classified as a cartilage protein, it is produced in the developing tendon. [11 12] To our knowledge there is no evidence that the protein is produced in mature tendons. It is therefore possible that the proposed biological consequences of the reported association in this study could be due to altered protein profiling: (i) the expression of type XI collagen in the mature, diseased and/or healing tendon, (ii) its expression only in the fibrocartilaginous regions of the mature tendon and (iii) the result of its expression and function during tendon development. Irrespective of the mechanism(s), further research is required to replicate these findings in larger independent populations and to explore the functional mechanisms underlying the complex genetic associations with the type XI

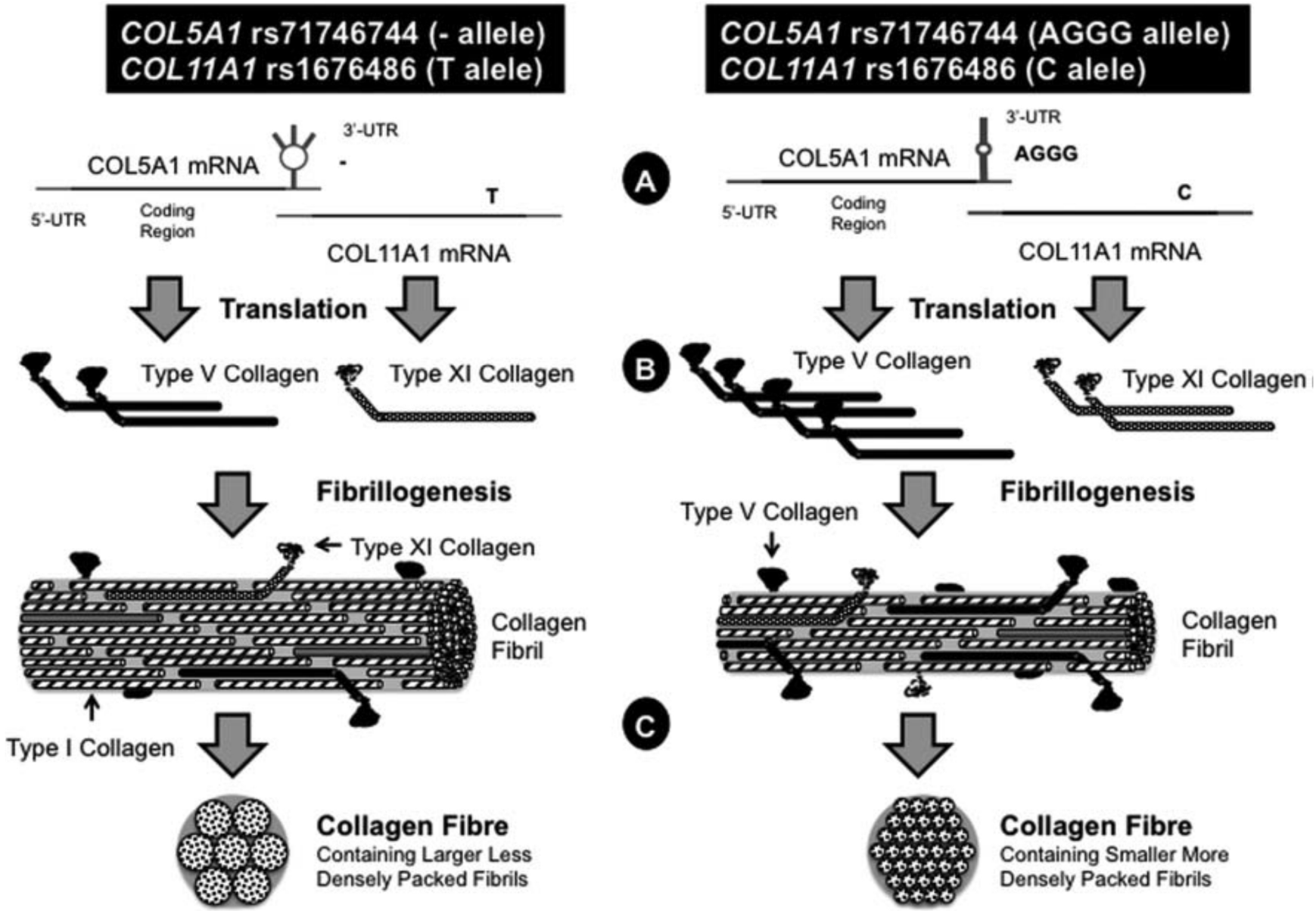

COL5A1 rs71746744 (- allele)
COL11A1 rs1676486 (T alele)
3'-UTR
COL5A1 mRNA
-
5'-UTR
Coding Region
T
COL11A1 mRNA
Translation
Type V Collagen
Type XI Collagen
Fibrillogenesis
Type XI Collagen
Collagen Fibril
Type I Collagen
Collagen Fibre
Containing Larger Less Densely Packed Fibrils
A
B
C
COL5A1 rs71746744 (AGGG allele)
COL11A1 rs1676486 (C alele)
3'-UTR
COL5A1 mRNA
AGGG
5'-UTR
Coding Region
C
COL11A1 mRNA
Translation
Type V Collagen
Type XI Collagen
Fibrillogenesis
Type V Collagen
Collagen Fibril
Collagen Fibre
Containing Smaller More Densely Packed Fibrils

Figure 4

A hypothetical schematic diagram illustrating the proposed mechanism of how polymorphisms within *COL5A1* and *COL11A1* potentially affect fibrilogenesis. Although there is no evidence that type XI collagen is produced in the mature, diseased and/or healing tendon, it is produced and functional during tendon development. [12] **(A)** The *COL5A1* rs71746744 (-/AGGG) and *COL11A1* rs1676486 (C/T) polymorphisms are part of an inffered pseudohaplotype that is associated with chronic Achilles tendinopathy. The *COL5A1* rs71746744 - allele [6 7] and the *COL11A1* rs1676486 T allele [18] are both believed to associated with increased mRNA degradation. Increased mRNA degradation is indicated in the left panel, while decreased mRNA degradation is indicated in the right panel). **(B)** The altered mRNA stability associated with these polymorphisms is believed to result in altered α1(V) and α1(XI) chain and types V and XI collagen production (decreased in left panel and increased in the right panel). **(C)** Types V and XI collagen regulates collagen fibril assembly and diameter (fibrillogenesis) and thus the mechanical properties of tendons. There is an inverse relationship between the types V and XI collagen content of the fibril and its diameter. Thinner more densely packed collagen fibrils are produced due to the increased production of types V and XI collagen (Right Panel). It has previously been proposed that the thinner fibrils are associated with chronic Achilles tendinopathy. [9]

collagen encoding genes and their interactions with the type V collagen encoding gene.

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

"What are the new findings"

- Functional polymorphisms within the type XI collagen genes (*COL11A1* and *COL11A2*) were not independently associated with chronic Achilles tendinopathy.
- These functional type XI collagen polymorphisms interact with a functional polymorphism within the type V collagen (*COL5A1*) gene to modulate the risk for Achilles tendinopathy in two independent populations.
- Since types XI and V collagen are structurally and functionally related, we propose that these polymorphisms interact to regulate collagen fibril assembly.

"How might it impact on clinical practice in the near future"

- Although there are no immediate clinical applications, the results of this study provide additional evidence that inter-individual variations in collagen fibril, the basic building blocks of tendons, assembly might be an important molecular mechanism in the etiology of chronic Achilles tendinopathy.
- Genetic risk factors could one day be included into multifactorial models to determine an individuals risk for Achilles tendinopathy.

- Genetic risk factors could however never to be used in isolation to diagnose or predict these injuries.

ACKNOWLEDGEMENTS

This study was supported in part by funds from the National Research Foundation of South Africa, University of Cape Town, and the South African Medical Research Council. The authors would like to thank Dr. G.G. Mokone for recruiting the SA participants, Prof. M.P. Schwellnus for the clinical diagnoses of the South African chronic Achilles tendinopathy participants, as well as N. Kariem for his laboratory assistance. M.P. was supported by the Thembakazi Trust. Authors M.C. and A.S. have filed patents on the application of specific sequence variations (not included in this manuscript) related to risk assessment of AT and anterior cruciate ligament injuries.

REFERENCES

1 Riley G. The pathogenesis of tendinopathy. A molecular perspective. *Rheumatology (Oxford)* 2004;**43**:131–42.

2 Ribbans WJ, Collins M. Pathology of the Tendo Achilles. Do our Genes Contribute? *Bone Joint J* In Press.

3 September AV, Posthumus M, Collins M. Application of genomics in the prevention, treatment and management of Achilles tendinopathy and anterior cruciate ligament ruptures. *Recent Pat DNA Gene Seq* 2012;**6**:216–23.

4 Mokone GG, Schwellnus MP, Noakes TD, Collins M. The COL5A1 gene and Achilles tendon pathology. *Scand J Med Sci Sports* 2006;**16**:19–26.

5 September AV, Cook J, Handley CJ, Van Der Merwe L, Schwellnus MP, Collins M. Variants within the COL5A1 gene are associated with Achilles tendinopathy in two populations. *Br J Sports Med* 2009;**43**:357–65.

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

7 Abrahams Y, Laguette M-J, Prince S, Collins M. Polymorphisms within the COL5A1 3'-UTR That Alters mRNA Structure and the MIR608 Gene are Associated with Achilles Tendinopathy. *Ann Hum Genet* Published Online First: 24 January 2013. doi:10.1111/ahg.12013

8 Wenstrup RJ, Florer JB, Brunskill EW, Bell SM, Chervoneva I, Birk DE. Type V collagen controls the initiation of collagen fibril assembly. *J Biol Chem* 2004;**279**:53331–7.

9 Collins M, Posthumus M. Type v collagen genotype and exercise-related phenotype relationships: a novel hypothesis. *Exerc Sport Sci Rev* 2011;**39**:191–8.

10 Fang M, Jacob R, McDougal O, Oxford JT. Minor fibrillar collagens, variable regions alternative splicing, intrinsic disorder, and tyrosine sulfation. *Protein Cell* 2012;**3**:419–33.

11 Fichard A, Kleman JP, Ruggiero F. Another look at collagen V and XI molecules. *Matrix Biol* 1995;**14**:515–31.

12 Wenstrup RJ, Smith SM, Florer JB, Zhang G, Beason DP, Seegmiller RE, *et al.* Regulation of collagen fibril nucleation and initial fibril assembly involves coordinate interactions with collagens V and XI in developing tendon. *J Biol Chem* 2011;**286**:20455–65.

13 Majava M, Hoornaert KP, Bartholdi D, Bouma MC, Bouman K, Carrera M, *et al.* A report on 10 new patients with heterozygous mutations in the COL11A1 gene and a review of genotype–phenotype correlations in type XI collagenopathies. *American Journal of Medical Genetics Part A* 2007;**143**:258–64.

14 Koga H, Sakou T, Taketomi E, Hayashi K, Numasawa T, Harata S, *et al.* Genetic mapping of ossification of the posterior longitudinal ligament of the spine. *Am J Hum Genet* 1998;**62**:1460–7.

15 Maeda S, Ishidou Y, Koga H, Taketomi E, Ikari K, Komiya S, *et al.* Functional impact of human collagen alpha2(XI) gene polymorphism in pathogenesis of ossification of the posterior longitudinal ligament of the spine. *J Bone Miner Res* 2001;**16**:948–57.

16 Maeda S, Koga H, Matsunaga S, Numasawa T, Ikari K, Furushima K, *et al.* Gender-specific haplotype association of collagen alpha2 (XI) gene in ossification of the posterior longitudinal ligament of the spine. *J Hum Genet* 2001;**46**:1–4.

17 Noponen-Hietala N, Kyllönen E, Männikkö M, Ilkko E, Karppinen J, Ott J, *et al.* Sequence variations in the collagen IX and XI genes are associated with degenerative lumbar spinal stenosis. *Annals of the Rheumatic Diseases* 2003;**62**:1208–14.

18 Mio F, Chiba K, Hirose Y, Kawaguchi Y, Mikami Y, Oya T, *et al.* A functional polymorphism in COL11A1, which encodes the alpha 1 chain of type XI collagen, is associated with susceptibility to lumbar disc herniation. *Am J Hum Genet* 2007;**81**:1271–7.

19 Lee H-S, Lee AT, Criswell LA, Seldin MF, Amos CI, Carulli JP, *et al.* Several regions in the major histocompatibility complex confer risk for anti-CCP-antibody

positive rheumatoid arthritis, independent of the DRB1 locus. *Mol Med* 2008;**14**:293–300.

20 Videman T, Saarela J, Kaprio J, Näkki A, Levalahti E, Gill K, *et al.* Associations of 25 structural, degradative, and inflammatory candidate genes with lumbar disc desiccation, bulging, and height narrowing. *Arthritis Rheum* 2009;**60**:470–81.

21 Koyama K, Nakazato K, Min S, Gushiken K, Hatakeda Y, Seo K, *et al.* COL11A1 gene is associated with limbus vertebra in gymnasts. *Int J Sports Med* 2012;**33**:586–90.

22 Mokone GG, Gajjar M, September AV, Schwellnus MP, Greenberg J, Noakes TD, *et al.* The guanine-thymine dinucleotide repeat polymorphism within the tenascin-C gene is associated with achilles tendon injuries. *Am J Sports Med* 2005;**33**:1016–21.

23 Gauderman WJ. Sample size requirements for association studies of gene-gene interaction. *Am J Epidemiol* 2002;**155**:478–84.

24 Gaunt TR, Rodríguez S, Day IN. Cubic exact solutions for the estimation of pairwise haplotype frequencies: implications for linkage disequilibrium analyses and a web tool 'CubeX'. *BMC Bioinformatics* 2007;**8**:428.

25 Epstein MP, Satten GA. Inference on haplotype effects in case-control studies using unphased genotype data. *Am J Hum Genet* 2003;**73**:1316–29.

26 Satten GA, Epstein MP. Comparison of prospective and retrospective methods for haplotype inference in case-control studies. *Genet Epidemiol* 2004;**27**:192–201.

27 Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 2009;**4**:1073–

81.

28 Mann V, Ralston SH. Meta-analysis of COL1A1 Sp1 polymorphism in relation to bone mineral density and osteoporotic fracture. *Bone* 2003;**32**:711–7.

29 Mann V, Hobson EE, Li B, Stewart TL, Grant SF, Robins SP, *et al.* A COL1A1 Sp1 binding site polymorphism predisposes to osteoporotic fracture by affecting bone density and quality. *Journal of Clinical Investigation* 2001;**107**:899–907.

30 Skorupski P, Król J, Starega J, Adamiak A, Jankiewicz K, Rechberger T. An alpha-1 chain of type I collagen Sp1-binding site polymorphism in women suffering from stress urinary incontinence. *Am J Obstet Gynecol* 2006;**194**:346–50.

31 Tilkeridis C, Bei T, Garantziotis S, Stratakis CA. Association of a COL1A1 polymorphism with lumbar disc disease in young military recruits. *J Med Genet* 2005;**42**:e44.

32 Khoschnau S, Melhus H, Jacobson A, Rahme H, Bengtsson H, Ribom E, *et al.* Type I collagen alpha1 Sp1 polymorphism and the risk of cruciate ligament ruptures or shoulder dislocations. *Am J Sports Med* 2008;**36**:2432–6.

33 Posthumus M, September AV, Keegan M, O'cuinneagain D, Van Der Merwe W, Schwellnus MP, *et al.* Genetic risk factors for anterior cruciate ligament ruptures: COL1A1 gene variant. *Br J Sports Med* 2009;**43**:352–6.

34 Ficek K, Cięszczyk P, Kaczmarczyk M, Maciejewska-Karłowska A, Sawczuk M, Cholewinski J, *et al.* Gene variants within the COL1A1 gene are associated with reduced anterior cruciate ligament injury in professional soccer players. *J Sci Med Sport* Published Online First: 15 November 2012. doi:10.1016/j.jsams.2012.10.004

35 Brown KE, Lawrence R, Sonenshein GE. Concerted modulation of alpha 1(XI)

and alpha 2(V) collagen mRNAs in bovine vascular smooth muscle cells. *J Biol Chem* 1991;**266**:23268–73.

36 Niyibizi C, Eyre DR. Identification of the cartilage α 1 (XI) chain in type V collagen from bovine bone. *FEBS Lett* 1989;**242**:314–8.

37 Kleman JP, Hartmann DJ, Ramirez F, van der Rest M. The human rhabdomyosarcoma cell line A204 lays down a highly insoluble matrix composed mainly of alpha 1 type-XI and alpha 2 type-V collagen chains. *Eur J Biochem* 1992;**210**:329–35.

38 Mayne R, Brewton RG, Mayne PM, Baker JR. Isolation and characterization of the chains of type V/type XI collagen present in bovine vitreous. *J Biol Chem* 1993;**268**:9381–6.

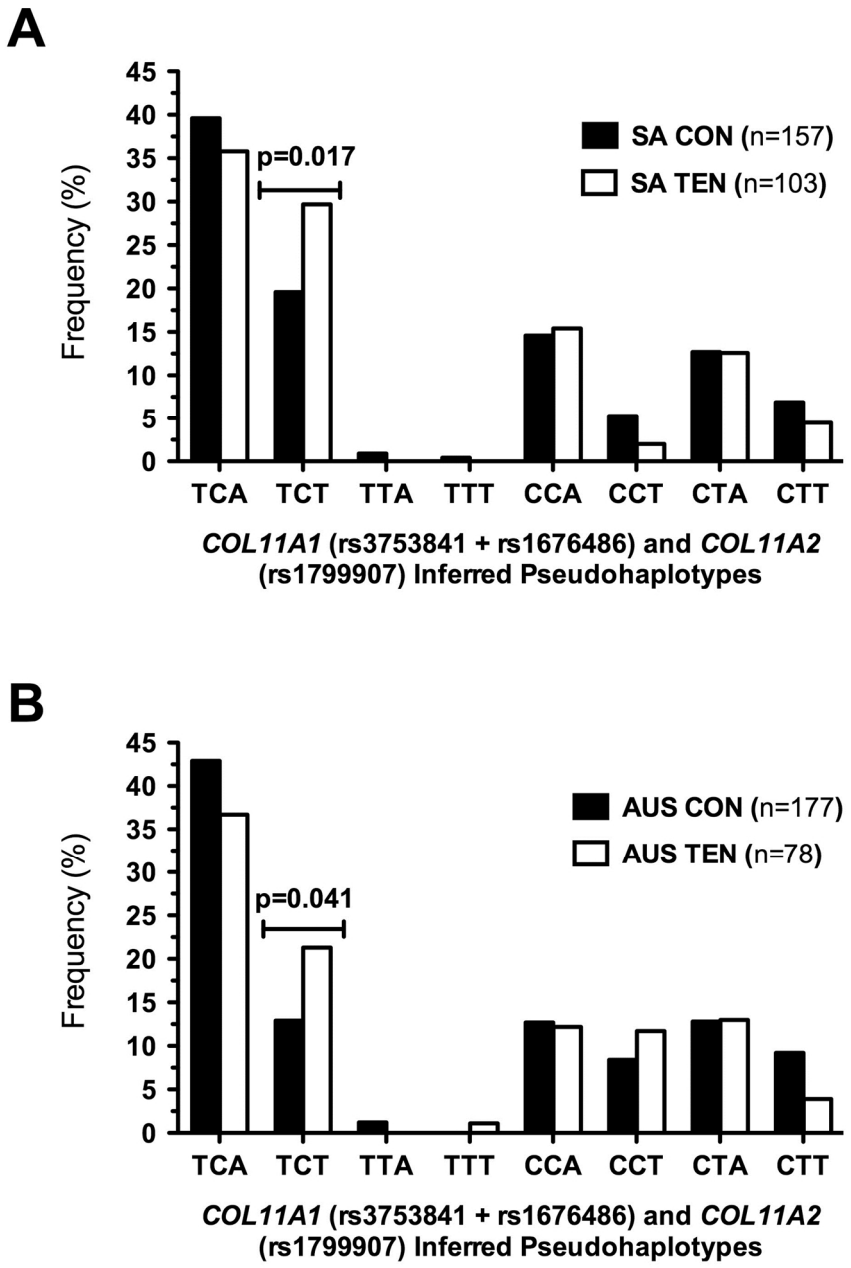
Supplementary Table 1. Genotype and allele frequency distribution of the functional single nucleotide polymorphisms (SNPs) rs3753841 and rs1676486 within *COL11A1* and rs1799907 within *COL11A2*, of the South African (SA) and Australian (AUS) chronic Achilles tendinopathy (TEN) groups and their respective control (CON) groups.

COL11A1		**rs3753841 genotype**					**rs3753841 allele**		
	n	**TT**	**TC**	**CC**	**p value**	**HWE**	**T**	**C**	**p value**
South Africa	247								
SA CON	150	37.3 (56)	46.7 (70)	16.0 (24)		0.861	60.7 (182)	39.3 (118)	
SA TEN	97	46.4 (45)	40.2 (39)	13.4 (13)	0.367	0.367	66.5 (129)	33.5 (65)	0.215
Australia	160								
AUS CON	104	29.8 (31)	49.0 (51)	21.2 (22)		1.000	54.3 (113)	45.7 (95)	
AUS TEN	56	41.1 (23)	35.7 (20)	23.2 (13)	0.235	0.054	58.9 (66)	41.1 (46)	0.479
COL11A1		**rs1676486 genotype**					**rs1676486 allele**		
	n	**CC**	**CT**	**TT**	**p value**	**HWE**	**C**	**T**	**p value**
South Africa	230								
SA CON	139	64.0 (89)	29.5 (41)	6.5 (9)		0.195	78.8 (219)	21.2 (59)	
SA TEN	91	68.1 (62)	27.5 (25)	4.4 (4)	0.571[a]	0.481	81.9 (149)	18.1 (33)	0.475
Australia	205								
AUS CON	138	62.3 (86)	31.2 (43)	6.5 (9)		0.318	77.9 (215)	22.1 (61)	
AUS TEN	67	67.2 (45)	29.9 (20)	3.0 (2)	0.538[a]	1.000	82.1 (110)	17.9 (24)	0.365
COL11A2		**rs1799907 genotype**					**rs1799907 allele**		
	n	**AA**	**AT**	**TT**	**p value**	**HWE**	**A**	**T**	**p value**
South Africa	262								
SA CON	158	46.8 (74)	41.8 (66)	11.4 (18)		0.580	67.7 (214)	32.3 (102)	
SA TEN	103	40.4 (42)	47.1 (49)	12.5 (13)	0.588	1.000	63.9 (133)	36.1 (75)	0.396
Australia	192								
AUS CON	126	52.4 (66)	34.9 (44)	12.7 (16)		0.059	69.8 (176)	30.2 (76)	
SA TEN	66	39.4 (26)	45.4 (30)	15.2 (10)	0.227	0.796	62.1 (82)	37.9 (50)	0.138

Values are percentages, with sample number of participants (n) displayed in parenthesis. HWE is the p-value for the exact tests for Hardy-Weinberg Equilibrium. [a] CC vs CT + TT.

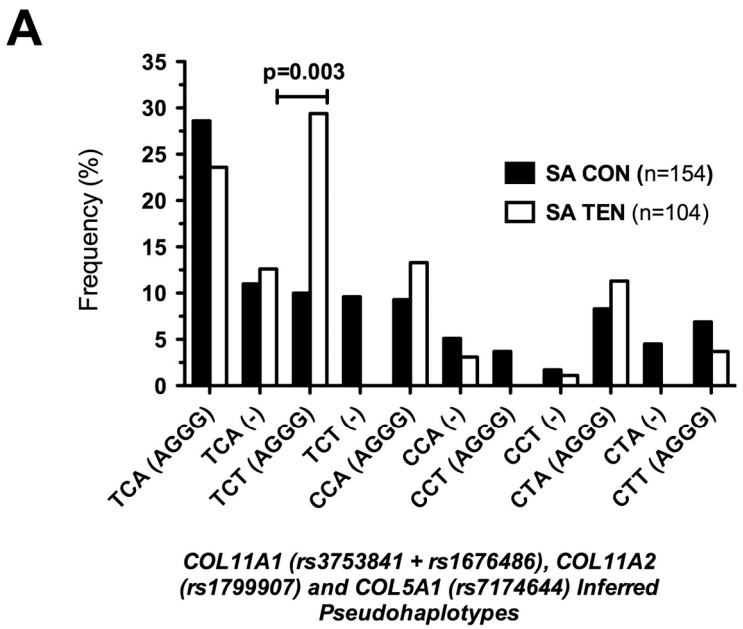
SA CON vs AUS CON: rs3753841 p=0.370, rs1676486 p=0.953 and rs1799907 p=0.500.

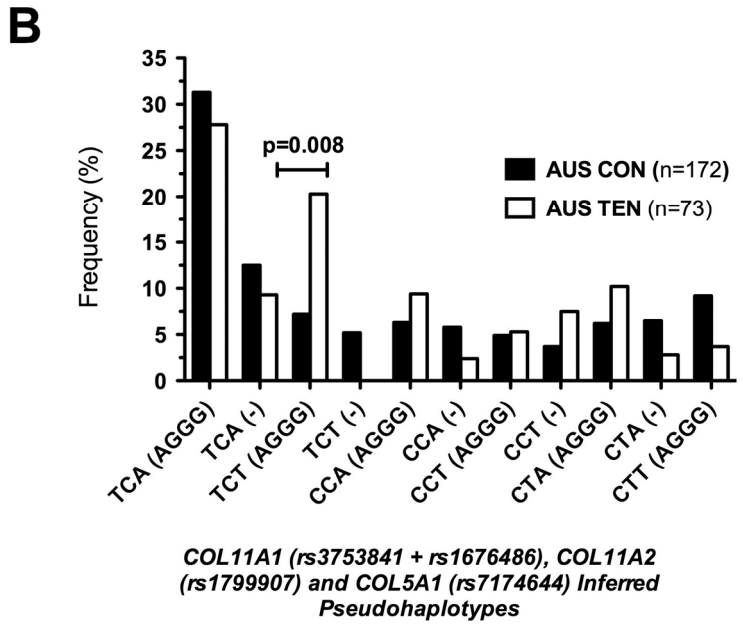
SA TEN vs AUS TEN: rs3753841 p=0.298, rs1676486 p=01.000 and rs1799907 p=0.885.



Supplementary Figure 1

The frequency of the eight inferred pseudohaplotypes constructed from *COL11A1* rs3753841 (C/T), *COL11A1* rs1676486 (T/C), *COL11A2* rs1799907 (T/A) and *COL5A1* rs71746744 (-/AGGG) in the **(A)** South African (SA) and **(B)** Australian (AUS) chronic Achilles tendinopathy (TEN, clear bars) and control (CON, solid bars) groups. The p value of the significantly different TCT pseudohaplotype is indicated in each graph. The total number (n) of inferred pseudohaplotypes within the CON and TEN groups is indicated in parenthesis in the graph.





Supplementary Figure 2

The frequency of eleven of the possible 16 inferred pseudohaplotypes with a frequency >2% constructed from *COL11A1* rs3753841 (C/T), *COL11A1* rs1676486 (T/C), *COL11A2* rs1799907 (T/A) and *COL5A1* rs71746744 (-/AGGG) in the **(A)** South African (SA) and **(B)** Australian (AUS) chronic Achilles tendinopathy (TEN, clear bars) and control (CON, solid bars) groups. The p value of the significantly different TCT(AGGG) pseudohaplotype is indicated in each graph. The total number (n) of inferred pseudohaplotypes within the CON and TEN groups is indicated in parenthesis in the graphs.