

Human genetic variation, sport and exercise medicine, and Achilles tendinopathy: Role for angiogenesis-associated genes

Masouda Rahim¹, Louis Y. El Khoury², Stuart M. Raleigh³, William J. Ribbans³, Michael Posthumus¹, Malcolm Collins¹, and Alison V. September¹

¹Division of Exercise Science and Sports Medicine, Department of Human Biology, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa;

²School of Biological Sciences, University of Essex, Colchester, United Kingdom;

³Division of Health and Life Sciences, University of Northampton, Northampton, United Kingdom

Running title: *Genetics of Achilles Tendinopathy*

Correspondence:
Dr. Alison V. September
Associate Professor
Division of Exercise Science and Sports Medicine
University of Cape Town
PO Box 115, Newlands, 7725
South Africa

Phone: +27 21 650 4559
Fax: +27 21 686 7530
Email: Alison.September@uct.ac.za

Keywords: Genetic variation; association study; sports medicine; Achilles tendinopathy; vascular endothelial growth factor; kinase insert-domain receptor; polymorphism

Contact Information for Co-Authors

Masouda Rahim

Division of Exercise Science and Sports Medicine, University of Cape Town, PO Box 115,

Newlands, 7725, South Africa

Phone: +27 21 650 3141

Fax: +27 21 686 7530

Email: masouda.rahim@gmail.com

Louis Y. El-Khoury

School of Biological Sciences, University of Essex, Colchester, CO4 3SQ, United Kingdom

Phone: +44 1206 873431

Email: louis.el-khoury@essex.ac.uk

Stuart M. Raleigh

Centre for Physical Activity and Chronic Disease, Institute of Health and Wellbeing,

University of Northampton, Boughton Green Road, Northampton, NN2 7AL, United

Kingdom

Phone: +44 1604 892306

Fax: +44 1604 893667

Email: stuart.raleigh@northampton.ac.uk

William J. Ribbans

Centre for Physical Activity and Chronic Disease, Institute of Health and Wellbeing,

University of Northampton, Boughton Green Road, Northampton, NN2 7AL, United

Kingdom

Phone: +44 1604 795415

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Fax: +44 1604 795416
Email: billribbs@uk-doctors.co.uk

Michael Posthumus
Division of Exercise Science and Sports Medicine, University of Cape Town, PO Box 115,
Newlands, 7725, South Africa
Phone: +27 21 650 4573
Fax: +27 21 686 7530
Email: mposthumus@me.com

Malcolm Collins
Division of Exercise Science and Sports Medicine, University of Cape Town, PO Box 115,
Newlands, 7725, South Africa
Phone: +27 21 650 6235
Fax: +27 21 686 7530
Email: malcolm.collins@uct.ac.za

List of abbreviations

ACL: anterior cruciate ligament

BMI: body mass index

CON: control group

ECM: extracellular matrix

EDTA: ethylenediaminetetraacetic acid

HWE: Hardy-Weinberg equilibrium

KDR: kinase insert-domain receptor

LD: linkage disequilibrium

RFLP: restriction fragment length polymorphism

PCR: polymerase chain reaction

SA: South Africa

SNP: single nucleotide polymorphism

TEN: Achilles tendinopathy

UK: United Kingdom

VEGFA: vascular endothelial growth factor, A isoform

Abstract

Sport and Exercise Medicine is one of the important subspecialties of 21st century healthcare contributing to improving the physical function, health and vitality of populations while reducing the prevalence of lifestyle-related diseases. Moreover, sport and exercise are associated with injuries such as Achilles tendinopathy, a common tendon injury. The angiogenesis-associated signalling pathway plays a key role in extracellular matrix remodelling, with increased levels of angiogenic cytokines reported after cyclic stretching of tendon fibroblasts. We investigated the variants in angiogenesis genes in relation to the risk of Achilles tendinopathy in two population samples drawn independently from South Africa (SA) and United Kingdom (UK). The study sample was comprised of 120 SA and 130 UK healthy controls, and 108 SA and 87 UK participants with Achilles tendinopathy. All participants were genotyped for five functional polymorphisms in the vascular endothelial growth factor (*VEGFA*) (rs699947, rs1570360, rs2010963) and kinase insert-domain receptor (*KDR*) genes (rs1870377, rs2071559). The *VEGFA* A-G-G inferred haplotype was associated with an increased risk of Achilles tendinopathy in the SA group (15% in controls versus 20% in cases, p=0.048) and the combined SA+UK group (14% in controls versus 20% in cases, p=0.009). These new findings implicate the *VEGFA* gene with Achilles tendinopathy risk, while highlighting the potential biological significance of the angiogenesis signalling pathway in the aetiology of Achilles tendinopathy. The evidence suggesting a genetic contribution to the susceptibility of sustaining a tendon injury is growing. We anticipate that high-throughput and multi-omics approaches, building on genomics, proteomics and metabolomics, may soon uncover the pathophysiology of many diseases in the field of Sports and Exercise Medicine, as a new frontier of global precision medicine.

Introduction

Sport and Exercise Medicine is one of the important subspecialties of 21st century healthcare contributing to improving the physical function, health and vitality of the population and working together with other disciplines to counter the increasing prevalence of chronic lifestyle-related disease. However, Sport and Exercise Medicine has not received the commensurate attention it deserves in the field of human genetics and the next generation omics technology applications. In this context, Achilles tendinopathy, one of the common soft tissue injuries, is amenable to research using modern genetics approaches. Tendon injuries constitute approximately 30-50% of all sporting injuries (Järvinen et al., 2005), with annual incidence rates of Achilles tendinopathy reported to be between 7-9% for top-level runners (Hirschmüller et al., 2012).

The underlying biological pathways and molecular mechanisms culminating in acute and overuse injuries are currently under investigation in an effort to improve both preventive and treatment modalities. To date, several intrinsic factors, including genetics, have been identified as risk factors predisposing an individual to musculoskeletal soft tissue injuries (Collins and Raleigh, 2009; Meeuwisse, 1994; Posthumus et al., 2011). The results of genetic-association studies, along with histopathological examination and expression studies, have assisted in elucidating some of the key biological pathways that may potentially be involved in the aetiology of musculoskeletal soft tissue injuries (Collins et al., 2015; Riley, 2008; Xu and Murrell, 2008).

Tendons are metabolically active tissues and are able to respond to stimuli but the rate of metabolism is slow (O'Brien, 1997). It is possible that the metabolic rate may increase in response to mechanical loading to allow for remodelling of the extracellular matrix (ECM) (Wang, 2006). Evidence shows that mechanical stimuli causes an increase in the expression of several cell signalling molecules (Jiang et al., 2012; Legerlotz et al., 2012; Skutek et al.,

2001). In particular, a number of studies have reported a pro-angiogenic expression profile after cyclic strain of tenocytes (Mousavizadeh et al., 2014; Nakama et al., 2006; Petersen et al., 2004) and during remodelling (Petersen et al., 2003). Moreover, an increase in vascularity was observed after mechanical loading and in symptomatic Achilles tendons (Lewis et al., 2009; Molloy et al., 2003; Pufe et al., 2001; Scott et al., 2008) whereas normal healthy tendons are relatively avascular structures (Pufe et al., 2001). A number of cell signalling molecules function in the angiogenesis pathway, such as transforming growth factor- β and fibroblast growth factor, with vascular endothelial growth factor (VEGF) considered to be a critical regulator of angiogenesis.

The *VEGFA* gene codes for the A isoform of the protein. This potent angiogenic factor is able to function in a diverse range of signalling pathways (Evans, 2015), and expression levels may be induced by mechanical loading, hypoxia as well as a number of biochemical factors/molecules (Neufeld et al., 1999; Petersen, Pufe, et al., 2003). Although, healthy tendons display negligible amounts of VEGF, this increases significantly during remodelling (Mousavizadeh et al., 2014; Petersen et al., 2004; Pufe et al., 2001). Furthermore, expression levels of VEGF in degenerative tendons are comparable to foetal tendons (Legerlotz et al., 2012; Pufe et al., 2001). VEGF binds to one of two high-affinity receptors, Flt-1 (fms related tyrosine kinase 1) or KDR (kinase insert-domain receptor), localised on the cell surfaces of vascular endothelial cells. KDR, encoded by the *KDR* gene, is considered the major mediator of the angiogenic, mitogenic, and permeability-enhancing effects of VEGF (Bao et al., 2009; Ferrara et al., 2003) both in normal physiology and in pathological processes (Ferrara et al., 2003).

Recently, genetic variants within the *VEGFA* and *KDR* genes were associated with risk of anterior cruciate ligament (ACL) ruptures (Rahim et al., 2014). Specifically, independent associations as well as haplotype combinations were implicated, thereby suggesting genomic

intervals harbouring potential DNA motifs within these two genes (Rahim et al., 2014). The aim of this study was to investigate the previously associated functional *VEGFA* rs699947, rs1570360, rs2010963 and *KDR* rs1870377 and rs2071559 variants with risk of Achilles tendinopathy in a South African and a British population.

Materials and methods

Participant recruitment

A total of 120 physically active South African Caucasian asymptomatic control (SA CON group) participants and 87 participants with chronic Achilles tendinopathy (SA TEN) were included in this study. These samples were previously recruited (2004-2013), according to the inclusion and exclusion criteria described by Mokone et al. (Mokone et al., 2005, 2006). In addition, a previously recruited (2011-2014) British Caucasian study group (Rickaby et al., 2015) consisting of 130 asymptomatic controls (UK CON group) and 108 participants with chronic Achilles tendinopathy (UK TEN) was included. The participants from each population were matched for country of birth (self-reported) and provided written informed consent. Furthermore, each participant completed questionnaires specifying personal details, injury details and medical and sporting history. For the cases, age was recorded at time of injury and for the controls at time of recruitment.

This study was approved by the Faculty of Health Sciences Research Ethics Committee within the University of Cape Town, South Africa and the Research Ethics Committee of the University of Northampton. This case-control genetic association study is in accordance with the recommendations of the STREGA Statement for reporting the results of genetic association studies (Little et al., 2009).

DNA extraction

For the South African participants, approximately 5ml of venous blood was collected from each patient by venipuncture of the forearm vein, into an EDTA Vacutainer® tube. Samples were stored at -20°C prior to total DNA extraction using the standard protocol as described by Lahiri and Nurnberger (1991) with slight modifications (Mokone et al., 2006). For the British participants, 2ml of saliva was obtained from each participant for DNA extraction as previously described (Rickaby et al., 2015). DNA extractions were performed at the University of Northampton and the samples made available for this study.

SNP selection and genotyping

The five single nucleotide polymorphisms (SNPs) selected for analysis within the *VEGFA* (rs699947, rs1570360, rs2010963) and *KDR* (rs1870377 and rs2071559) genes were previously implicated with risk of anterior cruciate ligament ruptures (Rahim et al., 2014). All samples (SA group n=273 and UK group n=242) were genotyped for the five polymorphisms using standard genotyping protocols. Polymerase chain reaction (PCR) and restriction fragment length polymorphism (RFLP) analysis was used for the *VEGFA* rs699947 (*Bgl*II) (Garza-Veloz et al., 2011) and *KDR* rs1870377 (*Alu*I) SNPs. Details of the primers sequences and PCR conditions are available on request. Restriction fragments were resolved, together with a 100bp size standard, by 2% agarose gel electrophoresis. Catalogued TaqMan™ Genotyping Assays (Applied Biosystems, Foster City, California, USA) were used for the *VEGFA* rs1570360, *VEGFA* rs2010963 and *KDR* rs2071559 SNPs. The PCR reactions were conducted using the Applied Biosystems StepOnePlus™ Real-Time PCR system and the Applied Biosystems StepOnePlus™ Real-Time PCR software v2.2.2 (Applied Biosystems), following manufacturer's recommendations.

Every PCR plate (RFLP and TaqMan™ analyses) contained negative controls (no DNA) and five repeat samples (known genotypes) as a quality control measure for reliable genotyping and detection of contamination. Genotypes were confirmed by two independent investigators, with an average genotyping success rate of 95%. Samples that failed twice to amplify during PCR for a particular SNP were considered unsuccessfully genotyped. (Rahim et al., 2014) All genotyping was conducted at the Division of Exercise Science & Sports Medicine within the University of Cape Town.

Statistical analysis

Power calculations were calculated using QUANTO v1.2.4 (<http://hydra.usc.edu/gxe>) to determine sample size for the study. Assuming minor allele frequencies between 0.2 and 0.5, a sample size of 85 cases would be adequate to detect an allelic odds ratio (OR) of 2.0 and greater at a power of 80% and a significance level of 5%. The programming environment R (R Development Core Team, 2010) was used for all data analysis. Basic descriptive statistics were compared using one-way analysis of variance to determine any significant differences between the characteristics of the CON vs. TEN groups. The R packages *genetics* (Warnes et al., 2011) and *SNPassoc* (González et al., 2007) were used to analyse any differences in genotype and allele frequencies between the diagnostic groups (CON/TEN) and to calculate Hardy-Weinberg equilibrium probabilities and linkage disequilibrium. Haplotypes were inferred using the R package *haplo.stats* (Schaid et al., 2002; Sinnwell and Schaid, 2011). Analysis of the participant characteristics revealed potential confounding variables so for this reason the genotype and haplotype analyses were adjusted accordingly. In addition to the analyses within the SA and UK study groups, the genotype and allele frequency distributions were compared between the countries (SA/UK), adjusting for age, weight and diagnostic group and in the combined (SA+UK) groups, adjusting for age, weight and country. Statistical significance was accepted when $p < 0.05$. No adjustments were made

for multiple testing as currently no obvious appropriate method exists (Nyholt, 2004; Perneger, 1998). The Bonferroni correction was considered too conservative for this study, as the tests are on the same group of participants and several of the polymorphisms were in tight linkage disequilibrium (Nyholt, 2004). Furthermore, correction for multiple testing was considered inappropriate as there was an *a priori* hypothesis (Perneger, 1998; Posthumus et al., 2012)

Results

Participant characteristics

The participant characteristics for the SA and UK groups were previously described in earlier studies investigating the same study participants (Mokone et al., 2005, 2006; Rickaby et al., 2015). Briefly, the SA CON and SA TEN participants were unmatched for age, weight and BMI (Table 1). In the UK study group, the CON and TEN participants were matched for all characteristics except age (Table 2). There were no significant genotype effects on age, sex, height, weight or BMI for any of the polymorphisms (Table S1).

Genotype and allele frequencies

The *VEGFA* rs699947 genotype frequency was significantly different ($p=0.003$) between the SA CON and SA TEN groups (Table 3). The CC genotype was significantly over-represented in the SA CON group (32%, $n=35$) compared to the SA TEN group (17%, $n=14$) ($p=0.019$, OR: 2.30, 95% CI: 1.14-4.64). Additionally, the CA genotype was significantly under-represented in the SA CON group (40%, $n=43$) compared to the SA TEN group (60%, $n=49$) ($p=0.005$, OR: 2.32, 95% CI: 1.28-4.17). No other significant differences in genotype or allele frequency distributions were observed between the CON and TEN groups (Table 3) for any of the five polymorphisms investigated in either the South African

or the British participants. The genotype and minor allele frequency distributions were similar to those reported for other European populations (<http://www.ncbi.nlm.nih.gov/snp/>) and for the previous ACL study by our group (Rahim et al., 2014). All the groups were in HWE except for the UK TEN group for the rs1570360 SNP ($p=0.020$) and the rs1870377 SNP ($p=0.015$).

Comparisons in genotype and allele frequency distributions between the countries, showed only the *KDR* rs2071559 genotype frequency distribution was significantly different between the groups ($p=0.026$) (Table 3). When the SA+UK groups were combined no significant differences were observed in genotype or allele frequency distribution (Table 3). The linkage disequilibrium (LD) structure of the investigated SNPs within the *VEGFA* gene, is presented in the supplementary material (Fig S1).

Inferred haplotypes

Of the eight possible haplotype combinations constructed from the *VEGFA* variants (rs699947 C/A – rs1570360 G/A – rs2010963 G/C), only four (C-G-G, C-G-C, A-G-G, A-A-G) were inferred at a frequency greater than 2%. The A-G-G inferred haplotype combination was significantly under-represented (Fig. 1A) in the SA CON group (15%, $n=18$) vs. the SA TEN group (20%, $n=22$) (dominant model, $p=0.048$; haplo.score:1.97). Similarly, when the SA and UK groups were combined the A-G-G haplotype was significantly under-represented in the SA+UK CON group (14%, $n=35$) vs. the SA+UK TEN group (20%, $n=39$) (dominant model, $p=0.009$; haplo.score:2.61) (Fig. 1C). However, no significant differences were observed between the UK CON and UK TEN groups (Fig. 1B).

Haplotypes were also inferred for the rs2071559 A/G and rs1870377 T/A polymorphisms within the *KDR* gene but no significant differences in frequency distribution were observed between any of the groups after adjusting for the potential confounders (data not shown).

Discussion

The angiogenesis-associated signalling pathway is considered a key biological mediator to promote matrix remodelling. Therefore, this study investigated five polymorphisms within the *VEGFA* (rs699947, rs1570360 and rs2010963) and *KDR* (rs2071559 and rs1870377) genes with risk of Achilles tendinopathy in two independent populations. The main finding of this paper was the haplotype association of the three *VEGFA* SNPs with risk of TEN in the SA group and the combined SA+UK group.

Vascular endothelial growth factor is a potent angiogenic factor, with a key role in the formation of new vasculature (Ferrara et al., 2003). The importance of VEGF in vasculogenesis and angiogenesis was demonstrated by Carmeliet et al. using a mouse gene knockout model, where loss of a single *VEGF* allele led to embryonic lethality (Carmeliet et al., 1996). Consequently, an abundance of research has focused on the role of this growth factor in normal physiology, pathological conditions and as a part of therapeutic strategies. To date, VEGF has been implicated in several pathological conditions including cancer, macular degeneration and rheumatoid arthritis (Evans, 2015). It is hypothesised that angiogenesis also occurs after mechanical loading in an effort to promote remodelling of the tendon extracellular matrix. This is further supported by studies identifying increased VEGF levels after mechanical loading of tenocytes (Mousavizadeh et al., 2014; Nakama et al., 2006; Petersen et al., 2004).

In the present study, the *VEGFA* A-G-G (rs699947 C/A – rs1570360 G/A – rs2010963 G/C) inferred haplotype was associated with increased risk of TEN. This association was noted in both the SA and the combined SA+UK groups which provides further support implicating the *VEGFA* locus in modulating musculoskeletal injury susceptibility (Rahim et al., 2014). The *VEGFA* A-G-G haplotype combination is associated with decreased transcription of the *VEGFA* gene and reduced VEGF plasma levels (Lambrechts et al., 2003).

Hence it is possible that angiogenesis is limited in an already hypovascular structure thus restricting the extent of matrix remodelling that is able to occur. Tang et al. (2016) hypothesised that low VEGF activity may be partly responsible for the slow or weak healing capacity of injured tendons. Furthermore, the authors demonstrated that VEGF gene therapy via an adeno-associated viral type-2 (AAV2) vector resulted in marked improvements in tendon healing and remodelling in an animal model (Tang et al., 2016). Taken collectively, investigation into the role of VEGF in modulating both tendon capacity and injury susceptibility is warranted.

A recent study implicated the A-A-G inferred haplotype, with decreased risk of non-contact ACL ruptures (Rahim et al., 2014). The explanation for the contrasting haplotype associations observed between TEN and ACL ruptures is not completely understood. It is interesting to note that both these implicated haplotypes are associated with reduced VEGF plasma levels (Lambrechts et al., 2003). One could therefore hypothesise that because the vascularity of tendons and ligaments differ (Benjamin et al., 2006; Fenwick et al., 2002), the biological impact of low levels of VEGF on neovascularisation related to remodelling and capacity would potentially be different between these functionally different tissues. Alternatively, the results may be indicating that these polymorphisms are potentially in linkage with a true risk variant. Thus, it is key that these genomic regions are further explored. Specifically, these associations need to be investigated at a functional level to determine the biological significance of these variants, and of angiogenesis in the repair model for TEN and ACL ruptures.

It is not surprising that the haplotype implicated in TEN risk modulation is reflected in the independent association of the rs699947 SNP genotype in the South African study group. Moreover, this promoter (-2578C/A) polymorphism is known to be functional, with the C allele correlating to increased VEGF protein production, and the AA genotype linked with

1
2
3 lower transcriptional activity (Shahbazi et al., 2002) and increased risk in several phenotypes
4
5 (Andraweera et al., 2012; Lambrechts et al., 2009; Seo et al., 2005). Therefore, there is
6
7 evidence implicating the *VEGFA* gene locus and specific variants at a functional and genetic
8
9 level with risk modulation.
10

11 Although, no independent or haplotype associations were observed for the *KDR* gene,
12
13 this remains a viable candidate for further investigation particularly due to its importance in
14
15 the VEGF signalling pathway. KDR is one of two main receptors that VEGFA binds to and is
16
17 responsible for the main downstream signalling effects of the ligand (Ferrara et al., 2003).
18
19 Moreover, *KDR* null mutant mice were observed to be embryonically lethal (Shalaby et al.,
20
21 1995). Therefore, it may be worth exploring additional genomic loci within this candidate
22
23 gene as the two polymorphisms (rs2071559 and rs1870377) investigated in the current study
24
25 are situated in the promoter and exon 11, respectively.
26
27
28

29 A limitation of the study was that SA participants were not matched for age, weight or
30
31 BMI. We were unable to determine the weight at time of injury for the cases due to
32
33 incomplete data but several participants reported an increase in weight due to injury-related
34
35 inactivity. Additionally, the older age of the cases may have influenced the weight of the
36
37 cases. In the UK participants, the CON and TEN groups were also unmatched for age and for
38
39 this reason, all statistical analyses were adjusted for the confounders. Although, the UK TEN
40
41 group deviated from HWE at the rs1570360 and rs1870377 loci, the genotype and minor
42
43 allele frequency distributions were similar to those reported for other European populations.
44
45 The main limitation of this study was the small sample size and therefore the SA and UK
46
47 study groups were also combined for the statistical analyses. A minimum sample size of 85
48
49 cases was adequate to detect only allelic odds ratios of 2.0 and greater, whereas the combined
50
51 SA+UK group sample set was sufficient to detect allelic odds ratios of 1.6 and greater.
52
53
54
55
56
57
58
59
60

Conclusions

Our observations collectively suggest that a genomic interval within the *VEGFA* gene was associated with risk of TEN. This gene requires further exploration to understand the potential role of *VEGFA* and specifically angiogenesis in the aetiology of TEN. Investigation into genes involved in the angiogenesis-associated signalling pathway may aid in the development of a risk profile for tendinopathy. In line with this, our future work is aiming towards employing next-generation technologies, specifically whole exome sequencing, to further explore these regions of interest in an effort to further characterise these genomic signatures.

The era of “*precision medicine*” is becoming a large focus underpinning research (Poland et al. 2011; Higdon et al. 2015). The technologies of the *omics* era has revolutionised our capacity to unravel these complex phenotypes. Through Big Data emerging from multi-omics research streams at genomics, proteomics, transcriptomics, epigenomics and bioinformatics levels we can advance the depth and breadth of research in the field of Sports and Exercise Medicine, with a view to improving global health, preventive medicine and quality of life.

Acknowledgements

The authors would like to thank Mr George Mokone and Dr Martin Schwellnus for SA participant recruitment. MR was funded by the University of Cape Town and the National Research Foundation. MP was funded by the Thembakazi Trust. This research was funded in part by the University of Cape Town Research Council, the National Research Foundation and the Department of Science and Technology/ National Research Foundation Collaborative Grant (NRF92550). Authors MC and AS have filed patents on the application of specific

sequence variations (not included in this manuscript) related to risk assessment of anterior cruciate ligament ruptures and Achilles tendinopathy.

Author disclosure statement

The authors declare no competing financial interests.

References

Andraweera PH, Dekker GA, Thompson SD, et al. (2012). The interaction between the maternal BMI and angiogenic gene polymorphisms associates with the risk of spontaneous preterm birth. *Mol Hum Reprod* 18, 459-465.

Bao P, Kodra A, Tomic-Canic M, et al. (2009). The role of vascular endothelial growth factor in wound healing. *J Surg Res* 153, 347–358.

Benjamin M, Toumi H, Ralphs JR, et al. (2006). Where tendons and ligaments meet bone: attachment sites ('entheses') in relation to exercise and/or mechanical load. *J Anat* 208, 471–490.

Carmeliet P, Ferreira V, Brier G, et al. (1996). Abnormal blood vessel development and lethality in embryos lacking a single VEGF allele. *Nature* 380, 435–439.

Collins M and Raleigh SM. (2009). Genetic risk factors for musculoskeletal soft tissue injuries. *Med Sport Sci* 54, 136–149.

Collins M, September AV and Posthumus M. (2015). Biological variation in musculoskeletal injuries: current knowledge, future research and practical implications. *Br J Sports Med* 49, 1497–1503.

Evans I. (2015). An Overview of VEGF-Mediated Signal Transduction. In: Fiedler L, editor(s). *VEGF Signaling*. Caversham, Reading, UK: Humana Press, p. 91–120.

Fenwick SA, Hazleman BL and Riley GP. (2002). The vasculature and its role in the damaged and healing tendon. *Arthritis Res* 4, 252–260.

Ferrara N, Gerber H-P and LeCouter J. (2003). The biology of VEGF and its receptors. *Nat Med* 9, 669–676.

Garza-Veloz I, Castruita-De la Rosa C, Cortes-Flores R, et al. (2011). No association between polymorphisms/haplotypes of the vascular endothelial growth factor gene and preeclampsia. *BMC Pregnancy Childbirth* 11, 35.

González JR, Armengol L, Solé X, et al. (2007). SNPAssoc: an R package to perform whole genome association studies. *Bioinformatics* 23, 644–645.

Higdon R, Earl RK, Stanberry L, et al. (2015) The promise of multi-omics and clinical data integration to identify and target personalized healthcare approaches in autism spectrum disorders. *OMICS* 19(4), 197-208.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Hirschmüller A, Frey V, Konstantinidis L, et al. (2012). Prognostic value of Achilles tendon Doppler sonography in asymptomatic runners. *Med Sci Sports Exerc* 44, 199–205.

Järvinen TAH, Kannus P, Maffulli N and Khan KM. (2005). Achilles tendon disorders: Etiology and epidemiology. *Foot Ankle Clin* 10, 255–266.

Jiang C, Shao L, Wang Q and Dong Y. (2012). Repetitive mechanical stretching modulates transforming growth factor- β induced collagen synthesis and apoptosis in human patellar tendon fibroblasts. *Biochem Cell Biol* 90, 667–674.

Lahiri DK and Nurnberger JI. (1991). A rapid non-enzymatic method for the preparation of HMW DNA from blood for RFLP studies. *Nucleic Acids Res* 19, 5444.

Lambrechts D, Poesen K, Fernández-Santiago R, et al. (2009). Meta-analysis of vascular endothelial growth factor variations in amyotrophic lateral sclerosis: increased susceptibility in male carriers of the -2578AA genotype. *J Med Genet* 46, 840–846.

Lambrechts D, Storkebaum E, Morimoto M, et al. (2003). VEGF is a modifier of amyotrophic lateral sclerosis in mice and humans and protects motoneurons against ischemic death. *Nat Genet* 34, 383–394.

Legerlotz K, Jones ER, Screen HRC and Riley GP. (2012). Increased expression of IL-6 family members in tendon pathology. *Rheumatology* 51, 1161–1165.

Lewis JS, Raza SA, Pilcher J, et al. (2009). The prevalence of neovascularity in patients

clinically diagnosed with rotator cuff tendinopathy. BMC Musculoskelet Disord 10, 163.

Little J, Higgins JPT, Ioannidis JPA, et al. (2009). STrengthening the REporting of Genetic Association studies (STREGA) - an extension of the STROBE statement. Eur J Clin Invest 39, 247–266.

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

Mokone GG, Gajjar M, September AV, et al. (2005). The guanine-thymine dinucleotide repeat polymorphism within the tenascin-C gene is associated with Achilles tendon injuries. Am J Sports Med 33, 1016–1021.

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

Molloy T, Wang Y and Murrell G. (2003). The roles of growth factors in tendon and ligament healing. Sports Med 33, 381–394.

Mousavizadeh R, Khosravi S, Behzad H, et al. (2014). Cyclic strain alters the expression and release of angiogenic factors by human tendon cells. PLoS One 9, e97356.

Nakama LH, King KB, Abrahamsson S and Rempel DM. (2006). VEGF, VEGFR-1, and CTGF cell densities in tendon are increased with cyclical loading: An in vivo tendinopathy model. J Orthop Res 24, 393–400.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Neufeld G, Cohen T, Gengrinovitch S and Poltorak Z. (1999). Vascular endothelial growth factor (VEGF) and its receptors. FASEB J 13, 9–22.

Nyholt DR. (2004). A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. Am J Hum Genet 74, 765–769.

O’Brien M. (1997). Structure and metabolism of tendons. Scand J Med Sci Sports 7, 55–61.
Perneger TV. (1998). What’s wrong with Bonferroni adjustments. BMJ 316, 1236–1238.

Petersen W, Pufe T, Zantop T, et al. (2003). Hypoxia and PDGF have a synergistic effect that increases the expression of the angiogenetic peptide vascular endothelial growth factor in Achilles tendon fibroblasts. Arch Orthop Trauma Surg 123, 485–488.

Petersen W, Unterhauser F, Pufe T, et al. (2003). The angiogenic peptide vascular endothelial growth factor (VEGF) is expressed during the remodeling of free tendon grafts in sheep. Arch Orthop Trauma Surg. 123, 168–174.

Petersen W, Varoga D, Zantop T, et al. (2004). Cyclic strain influences vascular endothelial growth factor (VEGF) and the hypoxia inducible factor 1 alpha (HIF-1a) in tendon fibroblasts. J Orthop Res 22, 847–853.

Poland GA, Ovsyannikova IG, Kennedy RB, Haralambieva IH, Jacobson RM. (2011) Vaccinomics and a new paradigm for the development of preventive vaccines against viral infections. OMICS 15(9), 625-636.

Posthumus M, Collins M, van der Merwe L, et al. (2012). Matrix metalloproteinase genes on chromosome 11q22 and the risk of anterior cruciate ligament (ACL) rupture. *Scand J Med Sci Sport* 22, 523–533.

Posthumus M, Collins M, September AV and Schwellnus MP. (2011). The intrinsic risk factors for ACL ruptures: an evidence-based review. *Phys Sportsmed* 39, 62–73.

Pufe T, Petersen W, Tillmann B and Mentlein R. (2001). The angiogenic peptide vascular endothelial growth factor is expressed in foetal and ruptured tendons. *Virchows Arch* 439, 579–585.

R Development Core Team. (2010). R: a language and environment for statistical computing. R Foundation for Statistical Computing, www.r-project.org.

Rahim M, Gibbon A, Hobbs H, et al. (2014). The association of genes involved in the angiogenesis-associated signaling pathway with risk of anterior cruciate ligament rupture. *J Orthop Res* 32, 1612–1618.

Rickaby R, El Khoury L, Ribbans WJ and Raleigh SM. (2015). Variation within three apoptosis associated genes as potential risk factors for Achilles tendinopathy in a British based case–control cohort. *Gene* 571, 167–171.

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

Schaid DJ, Rowland CM, Tines DE, et al. (2002). Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am J Hum Genet* 70, 425–434.

Scott A, Lian Ø, Bahr R, et al. (2008). VEGF expression in patellar tendinopathy: a preliminary study. *Clin Orthop Relat Res* 466, 1598–1604.

Seo JS, Lee S-S, Kim SUIY, et al. (2005). Influence of VEGF gene polymorphisms on the severity of ankylosing spondylitis. *Rheumatology (Oxford)* 44, 1299–1302.

Shahbazi M, Fryer AA, Pravica V, et al. (2002). Vascular endothelial growth factor gene polymorphisms are associated with acute renal allograft rejection. *J Am Soc Nephrol* 13, 260–264.

Shalaby F, Rossant J, Yamaguchi TP, et al. (1995). Failure of blood-island formation and vasculogenesis in Flk-1-deficient mice. *Nature* 376, 62–66.

Sinnwell J and Schaid D. (2011). Haplo.stats: a package for statistical analysis of haplotypes with traits and covariates when linkage phase is ambiguous. R package (version 1.5.4).

Skutek M, van Griensven M, Zeichen J, et al. (2001). Cyclic mechanical stretching enhances secretion of Interleukin 6 in human tendon fibroblasts. *Knee Surg Sports Traumatol Arthrosc* 9, 322–326.

Tang JB, Wu YF, Cao Y, et al. (2016). Basic FGF or VEGF gene therapy corrects

1
2
3 insufficiency in the intrinsic healing capacity of tendons. Sci Rep 6, 20643.
4
5

6
7 Wang JHC. (2006). Mechanobiology of tendon. J Biomech 39, 1563–1582.
8
9

10
11 Warnes G, Gorjanc G, Leisch F and Al. E. (2011). Genetics: a package for population
12 genetics. R Package (version 1.3.6).
13
14
15
16
17

18 Xu Y and Murrell GAC. (2008). The basic science of tendinopathy. Clin Orthop Relat Res
19 466, 1528–1538.
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

Table 1: Characteristics of the South African (SA) asymptomatic control (CON) group and the Achilles tendinopathy (TEN) group.

	SA CON (<i>n</i> = 120)	SA TEN (<i>n</i> = 108)	p-value ^a
Age* (years)	37.3 ± 10.4 (115)	42.9 ± 13.6 (89)	0.001
Sex (% male)	63.3 (120)	70.5 (105)	0.294
Height (cm)	175.0 ± 9.5 (116)	176.5 ± 8.7 (93)	0.239
Weight (kg)	71.4 ± 12.1 (119)	78.8 ± 18.2 (98)	<0.001 (0.023‡)
Body mass index (kg/m ²)	23.3 ± 2.7 (116)	25.3 ± 4.3 (93)	<0.001
Country of birth (% South African)	76.3 (118)	79.8 (104)	0.526

Values are expressed as mean ± standard deviation; sex and country of birth are represented as a percentage.
The number of participants (n) with available data for each variable is in parentheses.
* Age shows self-reported values at the time of onset of symptoms for the TEN group and at time of recruitment for the CON group
^a CON vs. TEN, p-values in bold typeset indicates significance (p<0.05)
‡ p-value adjusted for age and sex

Table 2: Characteristics of the British (UK) asymptomatic control (CON) group and the Achilles tendinopathy (TEN) group.

	UK CON (<i>n</i> = 130)	UK TEN (<i>n</i> = 87)	p-value ^a
Age* (years)	41.6 ± 11.6 (123)	45.2 ± 14.4 (87)	0.044
Sex (% male)	63.1 (130)	58.6 (87)	0.509
Height (cm)	174.9 ± 10.4 (123)	172.3 ± 8.5 (84)	0.057
Weight* (kg)	80.2 ± 19.5 (124)	77.1 ± 15.0 (62)	0.263
Body mass index* (kg/m ²)	25.9 ± 4.5 (123)	25.9 ± 4.1 (62)	0.983
Country of birth (% British)	96.0 (124)	96.5 (86)	1.000

Values are expressed as mean ± standard deviation; sex and country of birth are represented as a percentage.

The number of participants (n) with available data for each variable is in parentheses.

* Age, weight and body mass index (BMI) are self-reported values at the time of onset of symptoms for the TEN group and at time of recruitment for the CON group

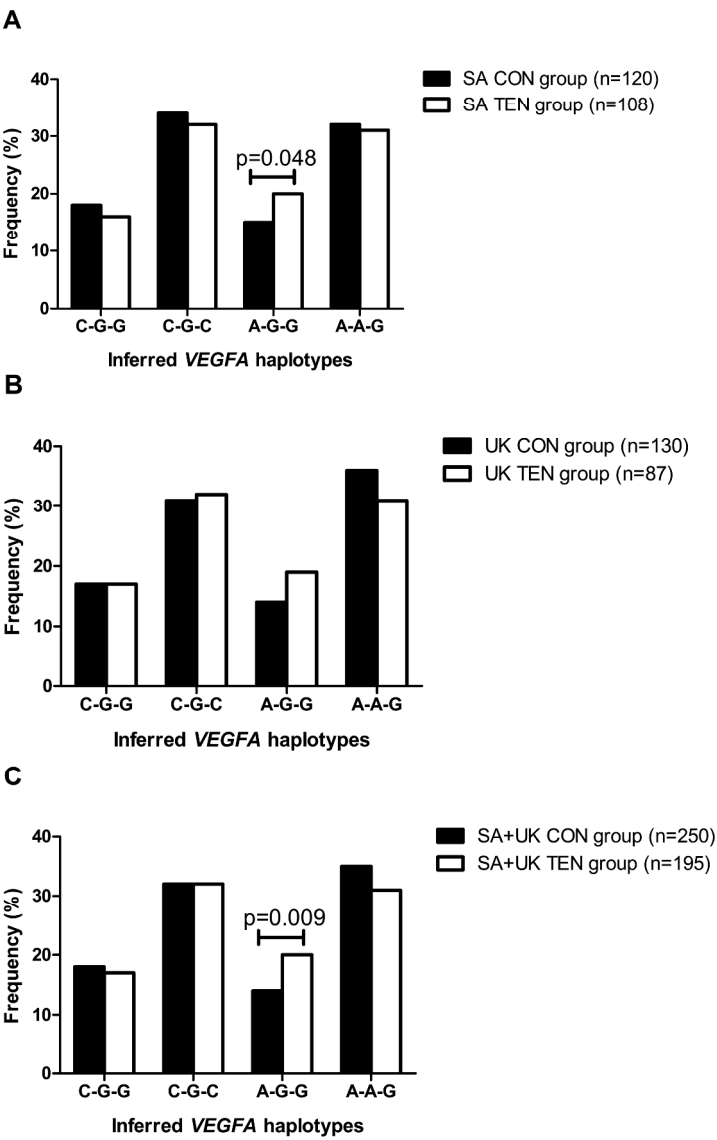
^a CON vs. TEN, p-values in bold typeset indicates significance (p<0.05)

Table 3: Genotype and minor allele frequency distributions, and p-values for Hardy-Weinberg exact test of the polymorphisms investigated in the *VEGFA* and *KDR* genes in all participants (male and female) in the control (CON) and the Achilles tendinopathy (TEN) groups within the South African (SA) and British (UK) populations.

		South African (SA)		British (UK)		p-values			
		CON	TEN	CON	TEN	Country*	Diagnostic group		
							SA*	UK†	SA+UK*
VEGFA rs699947	n	108	81	121	85				
	CC	32 (35)	17 (14)	28 (34)	25 (21)	0.456	0.003	0.864	0.051
	CA	40 (43)	61 (49)	46 (55)	51 (43)				
	AA	28 (30)	22 (18)	26 (32)	25 (21)				
	A allele	48 (103)	52 (85)	49 (119)	50 (85)	0.777	0.357	0.869	0.594
	HWE	0.089	0.117	0.380	1.000				
VEGFA rs1570360	n	105	75	111	82				
	GG	44 (46)	44 (34)	40 (44)	52 (43)	0.162	0.788	0.178	0.564
	GA	45 (47)	47 (37)	42 (47)	32 (26)				
	AA	11 (12)	9 (7)	18 (20)	16 (13)				
	A allele	34 (71)	33 (51)	39 (87)	32 (52)	0.423	0.970	0.130	0.193
	HWE	1.000	0.651	0.242	0.020				
VEGFA rs2010963	n	114	82	118	85				
	GG	42 (47)	42 (34)	49 (58)	46 (39)	0.778	0.883	0.863	0.733
	GC	49 (56)	50 (41)	38 (45)	42 (36)				
	CC	10 (11)	9 (7)	13 (15)	12 (10)				
	C allele	34 (78)	34 (55)	32 (75)	33 (56)	0.636	0.889	0.805	0.942
	HWE	0.412	0.506	0.217	0.806				
KDR rs2071559	n	115	83	119	85				
	GG	35 (40)	28 (23)	26 (31)	22 (19)	0.026	0.438	0.752	0.425
	GA	42 (48)	43 (36)	52 (62)	55 (47)				
	AA	24 (27)	29 (24)	22 (26)	22 (19)				
	A allele	44 (102)	51 (84)	48 (114)	50 (85)	0.528	0.218	0.676	0.347
	HWE	0.139	0.181	0.475	0.391				
KDR rs1870377	n	108	79	116	82				
	TT	56 (60)	51 (40)	62 (72)	57 (47)	0.475	0.693	0.184	0.689
	TA	38 (41)	43 (34)	32 (37)	29 (24)				
	AA	7 (7)	6 (5)	6 (7)	13 (11)				
	A allele	25 (55)	28 (44)	22 (51)	28 (46)	0.619	0.606	0.167	0.144
	HWE	1.000	0.467	0.436	0.015				

Genotype and allele frequencies are expressed as a percentage with the number of participants (n) in parentheses. p-values are for differences in genotype and minor allele

frequencies between countries (SA/UK) and diagnostic groups (CON/TEN), adjusted for one another. Differences between the diagnostic groups in the South African, British and combined (SA+UK) groups are shown in the table. * indicates p-values are adjusted for age and weight. † indicates p-values are adjusted for age and bold typeset indicates significance ($p<0.05$) p-values for the exact test of Hardy-Weinberg equilibrium (HWE) for each of the groups are included in the table.



129x197mm (600 x 600 DPI)

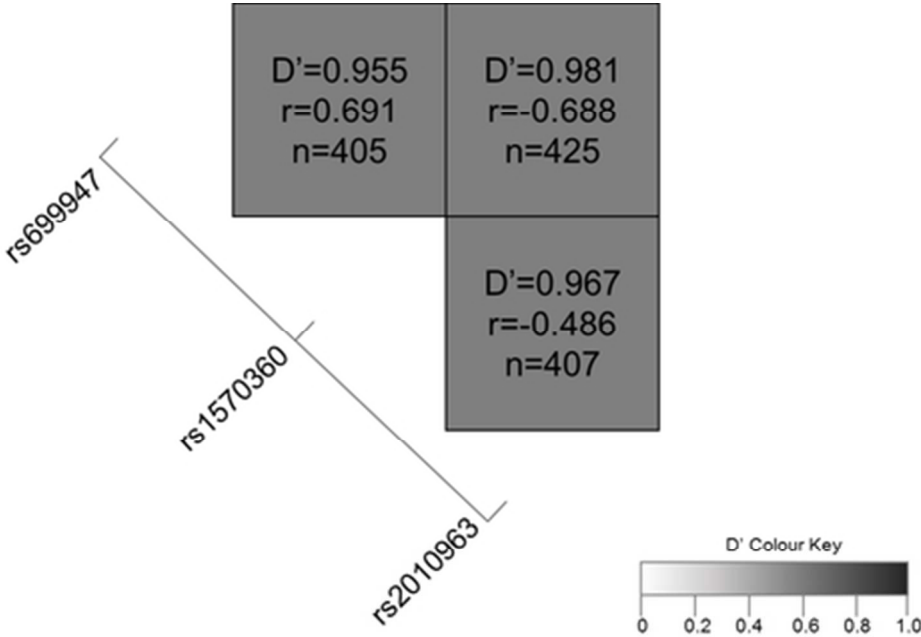
Figure 1: Inferred haplotype frequency distributions for the *VEGFA* rs699947, rs1570360 and rs2010963 polymorphisms in the control group (CON; black bars) and the Achilles tendinopathy group (TEN; white bars) for (A) the South African (SA) participants, (B) the British (UK) participants and (C) the combined SA and UK participants. Statistically significant differences in haplotype frequency between the groups are depicted on the graph, with p-values adjusted for age and weight. The number of participants (n) in each group is in parentheses.

Supplementary data

Table S1: Genotype effects on descriptive measures for the *VEGFA* rs699947, *VEGFA* rs1570360, *VEGFA* rs2010963, *KDR* rs2071559 and *KDR* rs1870377 polymorphisms in the South African (SA) and British (UK) Achilles tendinopathy groups

	<i>VEGFA</i> rs699947		<i>VEGFA</i> rs1570360		<i>VEGFA</i> rs2010963		<i>KDR</i> rs2071559		<i>KDR</i> rs1870377	
	<i>SA</i>	<i>UK</i>	<i>SA</i>	<i>UK</i>	<i>SA</i>	<i>UK</i>	<i>SA</i>	<i>UK</i>	<i>SA</i>	<i>UK</i>
Age	0.799	0.227	0.706	0.405	0.473	0.857	0.706	0.553	0.323	0.894
Sex	0.080	0.540	0.652	0.472	0.506	0.139	0.836	0.868	0.489	0.687
Height	0.020	0.883	0.953	0.619	0.102	0.361	0.685	0.644	0.119	0.805
Weight	0.062	0.126	0.624	0.608	0.717	0.599	0.727	0.265	0.547	0.241
BMI	0.595	0.088	0.464	0.611	0.982	0.120	0.568	0.136	0.980	0.196

Values expressed are p-values adjusted for group, critical age, sex, critical body mass index (BMI).
p-values in bold typeset indicate significance (p<0.05)



20x13mm (600 x 600 DPI)

Figure S1: Linkage disequilibrium (LD) structure mapping across the three *VEGFA* SNPs (rs699947, rs1570360 and rs2010963) on chromosome 6p12. The LD values (D') were calculated using the combined genotyping data of the CON and TEN groups for the South African (SA) and British (UK) study groups. The D' values between pairs of markers are in each cell and the strength of the linkage is indicated by the cell colour, corresponding to the colour key. The cells depict the D' values, sample size (n) and the correlation coefficient (r) for each association.