Biological variation in musculoskeletal injuries: current knowledge, future research and practical implications

Malcolm Collins, Alison V September, Michael Posthumus

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

Correspondence to
Professor Malcolm Collins, Division of Exercise Science and Sports Medicine, University of Cape Town, P.O. Box 115, Newlands 7725, South Africa; malcolm.collins@uct.ac.za

ABSTRACT
Evidence from familial and genetic association studies have reported that DNA sequence variants play an important role, together with non-genetic factors, in the aetiology of both exercise-associated and occupational-associated acute and chronic musculoskeletal soft tissue injuries. The associated variants, which have been identified to date, may contribute to the interindividual variation in the structure and, by implication, mechanical properties of the collagen fibril and surrounding matrix within musculoskeletal soft tissues, as well as their response to mechanical loading and other stimuli. Future work should focus on the establishment of multidisciplinary international consortia for the identification of biologically relevant variants involved in modulating injury risk. These consortia will improve the limitations of the published hypothesis-driven genetic association studies, since they will allow resources to be pooled in recruiting large well-characterised cohorts required for whole-genome screening. Finally, clinicians and coaches need to be aware that many direct-to-consumer companies are currently marketing genetic tests directly to athletes without it being requested by an appropriately qualified healthcare professional, and without interpretation alongside other clinical indicators or lifestyle factors. These specific genetic tests are premature and are not necessarily required to evaluate susceptibility to musculoskeletal soft tissue injury. Current practice should rather consider susceptibility through known risk factors such as a positive family history of a specific injury, a history of other tendon and/or ligament injuries and participation in activities associated with the specific musculoskeletal injuries. Potential susceptible athletes may then be individually managed to reduce their risk profile.

INTRODUCTION
Acute and chronic musculoskeletal soft tissues are commonly injured as a result of participation in physical activity and specific occupations. Several extrinsic and intrinsic risk factors have been reported to predispose an individual to musculoskeletal soft tissue injuries. Evidence from familial and predominately case–control genetic association studies, and more recently a single genome-wide association study (GWAS) in canines, have reported that DNA sequence variants play an important role in the aetiology of musculoskeletal soft tissue injuries. These include ACL ruptures (reviewed in ref. 1),[2–8] Achilles tendon injuries (reviewed in refs. 9 and 10),[4] [11] [12] lateral epicondylitis,[13] rotator cuff injuries[14–16] and carpal tunnel syndrome[17–19] in humans, as well as, equine superficial

digital flexor tendinopathy[20] and canine cranial cruciate ligament rupture (CCLR).[21] [22]

We have regularly reviewed the involvement of genetic factors in tendon and ligament injuries;[1] [9] [10] [23] [24] therefore, this review will specifically focus on recent developments and future research directions highlighting the discussions during the recent 'Genomics, Genetics and Exercise Biology Celebratory Symposium in Santorini, Greece', as well as the current and future clinical applications of this area of investigation.

GENETIC RISK FACTORS FOR MUSCULOSKELETAL SOFT TISSUE INJURIES IN HUMANS
Although there are important limitations to a case–control candidate-gene design approach,[25] all the published studies to date reporting an association of DNA sequence variants with musculoskeletal soft tissue injuries in humans, equines and canines have used this approach (reviewed in ref. 1), with the exception of a single study in canines.[22] In this approach, an a priori hypothesis that the protein or non-protein products of the genes are believed to be involved in the aetiology of the injury is used to select the candidates (reviewed in ref. 9). Genes encoding the basic structural components of tendons, ligaments and other connective tissue structures, namely the collagen fibril, have been identified as ideal candidates for evaluation. The collagen fibril will therefore be briefly summarised in the following paragraph.

With the exception of cartilage, the collagen fibril consists predominately of type I collagen. The type I collagen fibril and surrounding matrix also consist of other collagen types, elastin, many non-fibre forming proteins (glycoproteins) and proteoglycans (figure 1) (reviewed in refs. 10 and 24). The cellular content of connective tissue is dispersed between the collagen fibres and is responsible for the turnover and maintenance of the tissue matrix. From an exercise point of view, connective tissue is a dynamic tissue, with its cells able to respond appropriately or inappropriately to tissue loading and other stimuli.[26] An appropriate response will allow for repair, healing and/or adaptation of the tissue. An inadequate or inappropriate response will eventually lead to a pathological state with degradation or even failure of the matrix.

Common DNA sequence variants within genes encoding for structural components or regulators of the collagen fibril and other extracellular matrix proteins, such as collagens, fibrillins, proteoglycans and non-fibre forming proteins (glycoproteins),

Gene	Protein	AT	AR	ACL	Sho Dis	TE	CTS
Matrix Building Blocks							
COL1A1	α1(I) chain of Type I Collagen	X	✔ (?)	✔	✔	X	·
COL3A1	α1(III) chain of Type III Collagen	·	·	✔	·	·	·
COL5A1	α1(V) chain of Type V Collagen	✔	X (?)	✔ (F)	·	✔	✔
COL11A1 + COL11A2	α1(XI) and α1(XI) chains of Type XI Collagen	✔	·	·	·	·	·
COL12A1	α1(XII) chain of Type XII Collagen	X	✔ (?)	✔ (F)	·	·	·
COL14A1	α1(XIV) chain of Type XIV Collagen	X	X (?)	·	·	·	·
COL27A1 - TNC locus	α1(XXVII) chain of Type XXVII Collagen, Tenascin C (Glycoprotein)	✔	✔ (?)	·	·	·	·
ELN	Elastin (Elastic Fibres)	X	·	X	·	·	·
FBN2	Fibrillin-2 (Elastic Fibres)	✔	·	✔	·	·	·
ACAN	Aggrecan (Proteoglycan)	·	·	✔	·	·	X
BGN	Biglycan (Proteoglycan)	·	·	✔ (F)	·	·	✔
LUM - DCN locus	Lumican and Decorin (Proteoglycans)	·	·	✔	·	·	·
FMOD	Fibromodulin (Proteoglycans)	·	·	X	·	·	·
Genentic" Regultors							
MIR608	Non-Protein Coding Gene (miRNA)	✔					

Figure 1 A summary of all the genes that encode for the collagen fibril and surrounding matrix, as well as the non-coding RNAs, which are associated with chronic Achilles tendinopathy (AT), Achilles tendon rupture (AR), cruciate ligament/ACL, shoulder dislocation (SD), tennis elbow (TE) and carpal tunnel syndrome (CTS). Genes in which sequence variants have been associated with a specific injury are indicated with a tick (P) and a cross (X) if not associated. Associations that have been confirmed in another population are highlighted in grey. Those genes which require confirmation in larger cohorts are indicated by a question mark in parenthesis (?). Genetic variants only are associated with females by (F). Genes which have not been associated with any of the injuries have not been included in the figure. A schematic diagram of the collagen fibril, which consists predominately of type I collagen (grey tubes). The major isoform of type I collagen is a heterotrimer consisting of two α1(I) and one α2(I) chains, which are encoded for by the *COL1A1* and *COL1A2* genes, respectively. The fibril also contains trace amounts of types III (white tubes), V, XII, XIV (black structures) and other collagen and non-collagen proteins non-included in the diagram. Types III, XII and XIV collagen are all homotrimers consisting of three α1(III), α1(XII) and α1(XIV) chains, respectively. The α-chains are encoded by the *COL3A1*, *COL12A1* and *COL14A1* genes. The major isoform of type V collagen is a heterotrimer consisting of two α1(V) and one α2(V) chains, which are encoded for by the *COL5A1* and *COL5A2* genes, respectively. The proteins are not necessarily drawn to scale.

have been associated with susceptibility to musculoskeletal soft tissue injuries (figure 1). Interestingly, the effect sizes of some of these associations are relatively large for multifactorial disorders (OR>2), suggesting that genetic variants are probably strong modulators of the risk for musculoskeletal soft tissue injuries.[27–29] Sex-specific associations have also been reported for a subset of these variants, especially with ACL ruptures.[30 31] Since tendons and ligaments are dynamic tissues, associations have not been limited to genes coding for structural proteins. Variants within genes encoding proteins involved in the angiogenesis-associated signalling pathways (such as vascular endothelial growth factor and kinase insert-domain receptor with ACL injuries[2]), the apoptosis cell signalling pathways (such as *CASP8* with Achilles tendinopathy),[32] as well as the remodelling pathways (such as the *MMP10*, *MMP*1, *MMP3*, *MMP12* gene locus and *TIMP2*[12 33–35]), have also been associated with musculoskeletal soft tissue injuries.

Although the association of a limited number of the genetic markers have been repeated in other population groups or investigated in different injury types, the current results appear to indicate that there are similarities and differences with respect to the associated DNA sequence variants with acute and chronic injuries, as well as injuries with males and females. The following section will therefore focus on the reported association of variants within some of the collagen genes where these similarities and differences are illustrated.

The *COL1A1* gene encodes for the α1(I) chain of type I collagen, which is a heterotrimer consisting of two α1(I) and one α2(I) chains. However, a homotrimer consisting of three α1(I) chains may also be produced and incorporated into connective tissues together with the heterotrimer.[36] A functional variant located in a transcription factor, known as Sp1, binding site (rs1800012, G/T) within the first intron of *COL1A1* has been associated with cruciate ligament, mainly ACL, ruptures in four studies.[7 37–39] This variant has also been reported to associate with shoulder dislocations[37] and although the sample size is small, similar genotype distributions have been reported for Achilles ruptures.[29 40] It is reasonable to propose that the *COL1A1* variant could possibly predispose individuals to a number of different acute tendon and ligament injuries, but to date it has not been associated with any overuse injuries such as chronic Achilles tendinopathy[40] and lateral epicondylitis.[41]

With the exception of one study, where the *COL1A1* GG genotype was significantly under-represented in the ACL rupture group in recreational skiers,[7] all the other studies have reported similar genotype distributions, with an under-representation of the TT genotype in the acute injury groups (table 1).[37–39] Although a detailed explanation is beyond the

Table 1 DNA sequence variants within the *COL1A1*, *COL5A1*, *COL11A1*, *COL11A2* and *COL12A1* genes, which encode for the α1(I) chain of type I collagen, and the α1(V) chain of type V collagen, α1(XI) chain of type XI collagen, α2(XI) chain of type XI collagen and the α1(XII) chain of type XII collagen, respectively, tested for association with chronic AT, AR, CL/ACL, Sho Dis, TE and CTS

Gene and chromosomal location	Variant	Location	AT	AR	CL/ACL	Sho Dis	TE	CTS	Reference
Individual genes (single variants and variant-variant interactions)									
COL1A1 (17q21)	rs1800012 (G/T)	SP1 binding site within intron 1 at nucleotide 1023	X	↓ TT (?)	↓ TT* ↓ GG†	↓ TT	X	–	7 37–40
	rs1107946 (G/T)	Transcription factor binding site within promoter at -1997	–	–	↑ G-T haplotype	–	–	–	39
	rs1800012 (G/T)	SP1 binding site within Intron 1							
COL5A1 (9q34)	rs13946 (C/T)	3′-UTR	X	X (?)	X	–	↓ TT	↓ TT	13 19 27 31 42 43
	rs12722 (C/T)		↓ CC	X (?)	↓ CC (F)	–	↑ TT	↓ CC†	
	rs71746744 (-/AGGG)		↑ AGGG/AGGG	–	–	–	–	X	
	rs16399 (ATCT/-)		↑ -/-	–	–	–	–	–	
	rs1134170 (A/T)		↑ TT	–	–	–	–	–	
	rs13946 (C/T)	3′-UTR	–	–	↓ T-C†‡ haplotype	–	–	↓ T-W-C‡ haplotype	8 19
	rs12722 (C/T)								
COL11A1 (1p21)	rs3753841 (T/C)	Exon 52, Leu1323Pro	↑ T-C-T haplotype	–	–	–	–	–	44
	rs1676486 (C/T)	Exon 62, Pro1535Ser							
COL11A2 (6p21)	rs1799907 (T/A)	Intron 6							
COL12A1 (6q12-q13)	rs970547 (A/G)	Exon 65, Ser3058Gly	X	–	↑ AA (F) X (M)	–	–	–	30 45
Gene-gene interactions									
COL5A1	rs12722 (C/T)	–	–	–	↑ T-A (F) haplotype	–	–	–	5
COL12A1	rs970547 (A/G)								
COL11A1	rs3753841 (T/C)	–	↑ T-C-T-AGGG haplotype	–	–	–	–	–	44
COL11A2	rs1676486 (C/T)								
COL5A1	rs1799907 (T/A)								

X, not associated; ↓, decreased risk; ↑, increased risk; (?), potentially associated or not associated but needs confirmation in larger cohorts.
*Other mechanisms of injury, except the phantom foot mechanism of injury in skiing.
†The phantom foot mechanism of injury in skiing.
‡Two additional informative rare adjacent variants, namely rs14776422 (C/T) and rs55748801 (G/A), which have not been reported in a white population, were identified within the population group investigated in the CTS study. Since the genotyping method was unable to distinguish between these two variants, the CG wild-type alleles of these variants were designated as a W, while the three alternative combinations, CA, TG and TA, were designated as an M. The T-C haplotype in white populations is therefore equivalent to the T-W-C in other population groups, which contain the two additional rare variants.
AR, Achilles tendon rupture; AT, Achilles tendinopathy; CL/ACL, cruciate ligament/anterior cruciate ligament rupture; CTS, carpal tunnel syndrome; F, females; Sho Dis, shoulder dislocation; TE, tennis elbow; UTR, untranslated region.

scope of this review and has previously been discussed,[10] this exception could be due to the difference in injury mechanisms (the phantom foot mechanism of injury in skiing) and highlights the importance of accurately documenting and stratifying for mechanism of the injury. Further research where the mechanism of injury is clearly documented and considered, as well as the association of *COL1A1* variants with acute and chronic musculoskeletal soft tissue injuries, is required.

To date, only one study which investigated the association of the functional *COL1A1* rs1800012 (G/T) variant with ACL ruptures was reported in undefined populations in Turkey,[41] while all the other publications have been in white populations.[7 37–40] In fact, almost all of the ACL genetic risk factors have been identified in white populations. Since the frequency distribution of the *COL1A1* rs1800012 (G/T) variant and other variants across the human genome varies between population groups, the reported associations cannot necessarily be extrapolated to other population groups. The major problem with the *COL1A1* rs1800012 variant is the reported low frequency of the minor T allele (0.5% in Asian and 4.7% in Nigerian and Kenyan) and the TT genotype in various population groups. Very large data sets will therefore be required to understand the significance of this locus. Investigators should therefore be mindful of this observation and select appropriate informative musculoskeletal soft tissue injury genetic markers for the population group

under investigation at this specific *COL1A1* locus or in other proposed loci.

Like the *COL1A1* variant, the role of the functional 3′-untranslated region (UTR) within the *COL5A1* gene in the aetiology of musculoskeletal soft tissue injuries is also of interest. The association of up to five variants, namely rs13946 (C/T), rs12722 (C/T), rs71746744 (-/AGGG), rs16399 (ATCT/-) and/or rs1134170 (A/T), which span a proximal region of the *COL5A1* 3′-UTR, has been tested in several acute and chronic injuries (table 1). These variants define two major functional forms of the 3′-UTR, which we have previously termed the C-allelic and T-allelic forms (figure 2B).[42 46] Although not all five of the variants have been genotyped in the studies, considering all the published investigations to date, the C-allelic form appears nevertheless to be under-represented in chronic Achilles tendinopathy,[27 42] lateral epicondylitis[13] and carpal tunnel syndrome.[19] The associated 3′-UTR region does not exactly overlap between all these injuries[19] and could reflect the complexity of the biological role the *COL5A1* 3′-UTR could play in the tissue-specific regulation of type V collagen production. Further work is, however, required to test this hypothesis. Interestingly, the two tested *COL5A1* 3′-UTR variants do not appear to be associated with Achilles tendon rupture; however, these findings should be interpreted with caution because of the sample size limitation (table 1).[43] The CC genotype of rs12722,
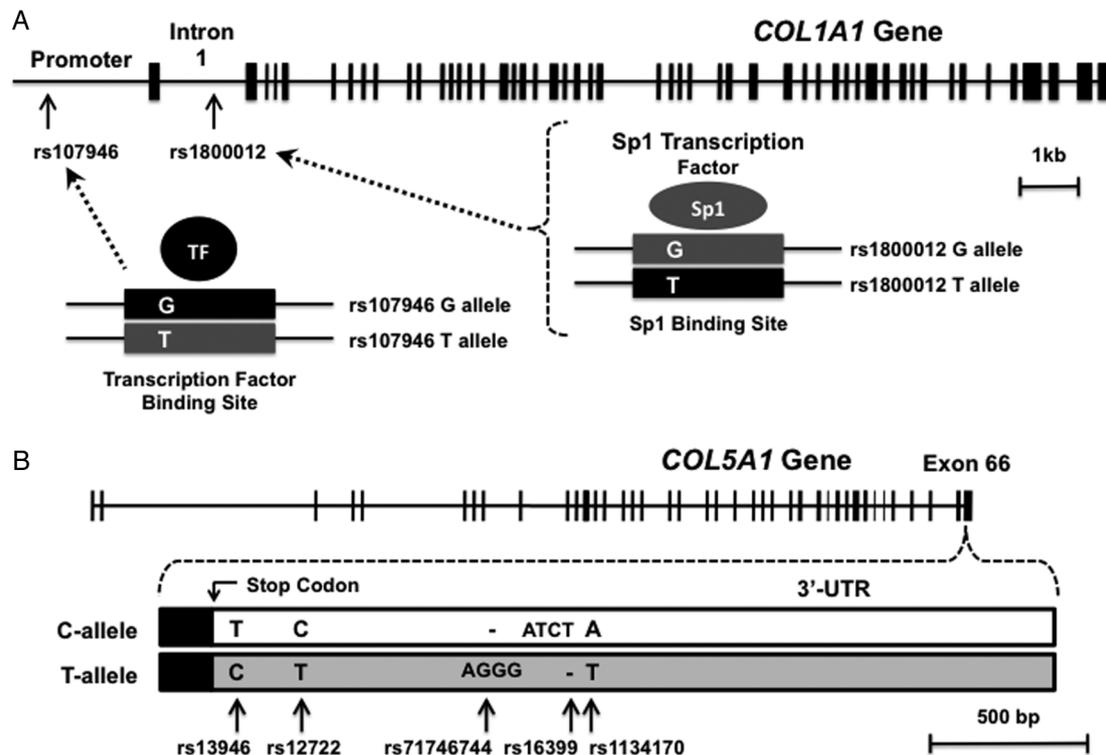
Figure 2 A schematic representation of the exon (vertical lines) and intron (horizontal lines) boundaries of the human (A) *COL1A1* and (B) *COL5A1* genes. The functional variants within the promoter region (rs107946, G/T) and first intron (rs1800012, G/T) of *COL1A1* are indicated. Both these variants are located in transcription factor binding sites, namely an unknown transcription factor binding site in the promoter and an Sp1 binding site in intron 1, which have been expanded. The variant within each binding site affects the affinity of the binding of the respective transcription factor. Increased binding and transcription have been reported to the alleles in black (ie, G allele of rs107946 and the T allele of rs1800012). An increased α1(I) chain is produced from these alleles. Two forms of *COL5A1* exon 66, which encodes for the terminal amino acids of the α1(V) chain (black box), the stop codon and the 3'-untranslated region (UTR) (clear and grey boxes), are indicated. The four variants, rs13946 (C/T), rs12722 (C/T), rs71746744 (-/AGGG), rs16399 (ATCT/-) and rs1134170 (A/T), reported to be associated with musculoskeletal soft tissue injuries, are indicated. The variants in the 3'-UTR indicated in grey are generally associated with musculoskeletal soft tissue injuries and determine one of the two major functional forms of the 3'-UTR (T allele) which has been reported to have increased mRNA stability. All the information used to construct this figure was obtained from Ensembl (http://www.ensembl.org).

which is a marker of the functional 3'-UTR C allele, was associated with reduced risk of ACL ruptures in females.[31]

Additionally, the AA genotype of *COL12A1* rs970547 (A/G) within exon 65, which results in a serine to glycine substitution at amino acid 3058 of the α1(XII) chain of type XII collagen (S3058G), was associated with an increased risk of ACL ruptures among female participants.[30] In agreement, this *COL12A1* polymorphism was not associated with ACL ruptures in Polish male soccer players.[47] Although some of the sex-specific independent associations were not repeated, a significant interaction was noted between the *COL5A1* rs12722 (T/C) and *COL12A1* rs970547 (A/G) polymorphisms, where the T-A inferred allele pair was associated with increased risk of ACL rupture in females in two separate cohorts (South African and Polish).[5] These results indicate a possible sex-specific association of genetic markers with muscular skeletal soft tissue injuries and the differences between acute and chronic injuries. These differences need to be considered in future research and any potential clinical applications.

In addition to protein coding RNAs, the human genome also codes for several categories of non-protein coding (non-coding) RNAs. These non-coding RNAs are predominately involved in cellular processes regulating the flow of information from DNA to proteins.[48] One of these non-coding RNA categories are small RNA molecules known as microRNAs, miRNA, which in

humans play an important role in the post-transcriptional regulation of gene expression, predominantly by binding to the 3'-UTRs of protein-coding genes to inhibit translation.[49 50] The majority of all human protein coding genes are under the control of miRNAs. These miRNAs have been reported to regulate development of tissues and are believed to be involved in the aetiology of many diseases.[49 50] In support of the important role of miRNAs in pathogenesis, the CC genotype of the polymorphic (rs499510, C/G) *MIR608* gene was implicated with increased risk of chronic Achilles tendinopathy in two populations.[42] The *MIR608* gene encodes a small miRNA, which binds to recognition sequences within the 3'-UTRs of *COL5A1* and other genes to inhibit translation. This is the first non-coding gene shown to be associated with musculoskeletal soft tissue injuries and adds another level of complexity to the poorly understood aetiology of these injuries.

GENETIC RISK FACTORS IN OTHER SPECIES

The cranial cruciate ligament in canines and the superficial digital flexor tendon in equines are equivalent to the ACL and Achilles tendon in humans, respectively. Variants within genes encoding for proteins involved in ligament strength and stability, as well as extracellular matrix formation, have been associated with susceptibility to CCLR. These genes include, among others, *COL5A1*, *COL1A1*, *COL3A1*, *COL11A1* and *FBN1*.[21]

One GWAS has identified three main chromosomal regions associated with CCLR on chromosomes 1, 3 and 33. Interestingly, the results of this study implicate the involvement of neurological pathways in susceptibility to CCLR.[22] The involvement of these pathways in ACL injuries requires investigation. Similar to the Achilles tendon, variants within the equine *TNC* and *COL5A1* genes are associated with risk of superficial digital flexor tendinopathy in national hunt thoroughbred racehorses.[20] Since the same genes have been identified across species, these results strengthen the hypothesis that genetic variants modulate the risk of musculoskeletal soft tissue injuries in humans.

FUNCTIONAL SIGNIFICANCE OF GENETIC RISK FACTORS FOR MUSCULOSKELETAL SOFT TISSUE INJURIES

It is proposed that genetic variants underpinning the genetic loci associated with musculoskeletal soft tissue injuries contribute to the normal interindividual biological variation in the structure and, by implication, mechanical properties of tendons, ligaments and other connective tissues. It is also reasonable to assume that these variants also contribute to the biological variation in the tissue responses to mechanical loading and other stimuli. In addition, many of the intrinsic risk factors for musculoskeletal soft tissue injuries are in their own right multifactorial phenotypes determined, to a lesser or greater extent, by both genetic and environmental factors.[9 51] For example, the heritability of flexibility, a common intrinsic risk factor, has been estimated to be between 64% and 70%.[52 53] It is therefore unlikely, based on the current evidence, that the associated genetic variants are independent risk factors for musculoskeletal soft tissue injuries. Instead, they most likely modulate risk by contributing to the normal biological variations within many of the anatomical, neuromuscular and other risk factors. The matrix of connective tissues adapts in response to load under normal physiological conditions; however, there are a spectrum of loads, ranging from high forces at low frequencies to low forces at high frequencies. Inappropriate responses to these loads may therefore result in maladaptation, eventually resulting in either an acute or chronic injury. Owing to normal biological variation, we propose that each athlete's tolerable load will be unique, thus resulting in a wide interindividual variation in the response of musculoskeletal soft tissues to load or other stimuli. A proportion of individuals will therefore intrinsically have an increased risk for injuries while others may have a reduced risk for such injuries. Molecular genetic tools integrated with other scientific disciplines should collectively be applied by researchers to elucidate the biological mechanisms of common recreational and occupational injuries.[1] An increase in the correct understanding of the underlying biochemical and mechanical properties of these injuries will assist clinicians and trainers in personalising the optimal training load for the individual, allowing them to reach optimal performance while reducing the risk of injury.

The following section will review our current understanding on how the associated genetic variants contribute to the biological variation in connective tissue responses to loading and other stimuli. The function of many of the associated variants is currently unknown. Some of the associated variants are located within exons and are known to potentially alter the amino acid sequence and, therefore, potentially the function of the encoded protein. These amino acid substitutions within the structural, regulatory and/or signalling proteins can contribute to normal biological variation and susceptibility to injury. This hypothesis is supported by the published work on the *COL1A1*, *CO51A1*, *COL11A1* and *COL11A2* associated variants, which are located

in regulatory regions of the genes, which has suggested that variations within the amounts of protein produced within the tissue during development, tissue maintenance and in response to load (eg, types I, V and XI collagen) are an important contributing factor for the biological variation in function of connective tissues.

Increased binding of the transcription factor, Sp1, to the rare T allele of *COL1A1* intronic polymorphism (rs1800012, G/T) has been reported to produce more α1(I) chains, which was proposed to result in increased production and incorporation of the α1(I)$_3$ type I collagen homotrimer into connective tissues (figure 2A).[36] In addition, an interaction between the T allele of this intronic polymorphism and the G allele of a second polymorphism (rs1107946, G/T) within a regulatory element of the *COL1A1* promoter at position 1997 in reducing the risk of ACL ruptures was reported.[39] Since both alleles are known to increase the expression of *COL1A1*,[36] this interaction may potentially result in further increased expression of the α1(I) chain. Exactly how the formation of the proposed type I collagen homotrimers affects the mechanical properties of connective tissues is currently unknown.

Type V collagen, a minor fibrillar collagen relative to type I collagen, interacts with types I and III collagen and plays a role in regulating the formation and diameter of the collagen fibril (figure 1).[54] Since the 3′-UTR of eukaryotic genes plays an important role in post-transcriptional gene regulation,[55 56] we have previously identified two major functional allelic forms of the 3′-UTR of *COL5A1*, which encodes the α1(V) chain of type V collagen, with the c-allelic form having a reduced, mRNA stability compared to the T form (figure 2B).[46 42] As reviewed above, versions of the T-allelic form are associated with several musculoskeletal soft tissue injuries (table 1).[13 19 27 31 42]

On the basis of the genetic association and functional findings of the *COL5A1* 3′-UTR, we proposed that the relative content of type V collagen in tendons, ligaments and other connective tissues alters the fibril diameters and packing density within these tissues, which in turn will alter the mechanical properties of the tissue and susceptibility to injury.[57] Type XI collagen, which is structurally and functionally similar to type V collagen, has been suggested to interact with type V collagen in regulating the formation of the collagen fibril during tendon development.[58] Interestingly we have recently reported that gene-gene interactions between *COL11A1* and *COLL11A2*, which encode for the α1(XI) and α2(XI) chains of type XI collagen, respectively, as well as interactions between *COL11A1*, *COLL11A2* and *COL5A1*, modulate the risk of chronic Achilles tendinopathy in two populations (table 1).[44] On the basis of the proposed functions of the types V and XI collagen gene variants, it was hypothesised that the gene-gene interactions may further modulate the risk of chronic Achilles tendinopathy by affecting mRNA stability, type V and/or XI collagen production, collagen fibril diameter and packing density, as well as change the biomechanical properties of the collagen fibril.[44] Further work is required to test this hypothesis.

Finally, the importance of the collagen fibril in musculoskeletal soft tissue injuries is further highlighted by the fact that lethal and/or severe disease causing mutations, which occur regardless of environmental exposure and independent of other non-genetic factors, have been identified within many of the collagen and non-collagen genes that encode for proteins within or associated with the fibril (reviewed in ref. 10). This suggests that limited biological redundancy is tolerated within the collagen fibril and that common DNA sequence variants within these genes may therefore strategically impact on the biological

consequences of environmental exposures, such as load, and therefore injury.

FUTURE WORK

The establishment of international consortia is an essential 'next step' in the identification of genetic risk factors for musculoskeletal soft tissue injuries. This will aim to improve the limitations of the previously published hypothesis-driven genetic association studies and allow resources to be pooled in recruiting large well-characterised independent cohorts required for whole-genome screening methods to identify the landscape of biologically relevant pathways across all populations involved in the aetiology of musculoskeletal soft tissue injuries.

Multidisciplinary teams consisting of clinicians, such as orthopaedic surgeons, sports and exercise medicine physicians, radiologists, as well as applied and basic scientists, should be included in the consortia. The latter group should consist of, among other disciplines, human molecular geneticists, molecular and cellular biologists, exercise physiologists, biostatisticians and bioinformaticians. It is important that bioinformatician are involved in the study design and planning when next-generation sequencing methods are employed. To this end, an international consortium, which originally included researchers from South Africa and Poland,[5] is currently being expanded to include additional groups from Australia, Qatar, Russia and Italy. The aim of the consortium will be to identify DNA sequence variants that modify the risk of ACL injuries and it is currently the only consortium within the 'ATHLOME PROJECT' that will specifically investigate exercise-associated injuries. Initially, the consortium plans to screen the current known loci for ACL injury susceptibility in larger data sets in an attempt to identify if these loci remain susceptibility loci across all populations using the hypothesis-driven candidate gene case–control study design. These replication studies will be followed by the ACL sequencing project and eventually expand to include other next-generation methodologies.

A poorly defined or heterogeneous musculoskeletal soft tissue injury profile is a common reason why genetic associations are often not repeated in independent follow-up studies. The inclusion of only homogeneous clinically well-defined injuries as cases in genetic association studies is therefore an important consideration. The selection of appropriately matched controls is also as important as the selection of the cases. Current results suggest that the control population consists of the survivors who are protected from injury and are more than likely a very select group.[27] Care will therefore be taken by all members of the consortium to use the same criteria to accurately phenotype, with respect to ancestry, sporting and occupational details, the injury profile and mechanism of injury, other injury history and family history, as well as other appropriate medical history and medication use.

Besides the hypothesis-driven or hypothesis independent case–control approach to identify genetic risk factors, researchers also need to investigate the role of these risk factors in the development of musculoskeletal soft tissue injuries in well-designed retrospective and prospective studies. Again, large sample sizes and well-characterised participants need to be included in these studies.

The whole-genome screening method, such as GWAS, whole-genome sequencing (WGS), exon sequencing and targeted sequencing, also needs to be carefully considered. The enriched targeted exome sequencing approach, for example, has better coverage than WGS and is therefore a better technique for identifying rare variants. Since <2% of the human genome contains protein-coding genes, important regulatory elements involved in the aetiology of musculoskeletal soft tissue injuries outside of the exons[39] and non-coding genes, such as miRNAs,[42] will be missed using exome sequencing. For this reason, careful considerations are needed in selecting the most appropriate enriched sequencing platform.

Future research should also consider transcriptomics, which allows for a snapshot of the whole transcriptome within the tissue of interest rather than a predetermined subset of genes extracted from blood or buccal smears. Techniques include total RNA, mRNA, targeted RNA and non-coding RNA sequencing. The role of different isoforms, novel transcripts, gene fusions and small regulatory RNAs in the aetiology of musculoskeletal soft tissue injuries can be explored using these approaches.

There is growing evidence that physiological traits responding to exercise depend on DNA sequence variants as well as on epigenetics.[59] It is reasonable to assume that environmental factors, such as exercise, may also work through epigenetic mechanisms to modulate the risk of musculoskeletal soft tissue injuries.[9] In support of this, an age-dependent increase in the hypermethylation of the COL1A1 promoter in periodontal ligaments has been reported.[60] Future work should therefore consider epigenomic techniques, also within the tissue of interest, such as methylation sequencing and ChIP sequencing, which can be used to identify protein-DNA and protein-RNA interactions.

Finally, the functional variants within or around the genes that modify susceptibility to injury need to be identified and the role of these functional variants and the molecular mechanisms that cause injury needs to be elucidated. This can only be done if human molecular genetics is integrated with the other biological disciplines such as molecular and cell biology.[42 46]

CLINICAL RELEVANCE

The overarching aim of all research focused on identifying the genetic susceptibility underlying musculoskeletal soft tissue injuries is really to understand the molecular mechanisms underpinning injury susceptibility. Before we can design tests to inform injury susceptibility, we need to understand the cumulative effect of all the risk parameters in the injury model. More important, it is pivotal that we explore the potential biological consequences of these various risk parameters before we can design effective clinical management strategies. For these reasons, research into musculoskeletal soft tissue injury susceptibility requires collaboration from a multidisciplinary network of basic scientists and clinicians.

Diagnostic and prognostic genetic testing has successfully been applied to classical Mendelian disorders, such as osteogenesis imperfecta and Ehlers-Danlos syndrome, where genetic mutations are the major determinants of the conditions. Genetic testing will, however, never be diagnostic or prognostic in nature for multifactorial disorders, such as musculoskeletal soft tissue injuries, which are caused by poorly understood complex interactions between environmental factors (eg, load) and an individual's genetic background. Clinicians should therefore refrain from using terms such as diagnostic, prognostic or predictive when considering any clinical application of genetic profiling for injury risk. These tests would only be able to inform injury susceptibility in the context of considering all other clinical physical activity as well as current and past injury risk factors.

Although scientists are only starting to understand its contribution, numerous companies are marketing direct-to-consumer (DTC) genetic tests for common sports injuries. These DTC tests are not injury-specific and premature since the genetic data

are incomplete. The genetic data are also not interpreted together with clinical indicators and lifestyle factors to identify an altered risk for injury by an appropriately qualified healthcare professional. Although beyond the scope of this review, important medical ethical issues are also not considered by these companies, especially when testing professional athletes.[9]

Further, clinicians should be reminded that specific genetic tests are not required to evaluate susceptibility to musculoskeletal injury. Current practice should rather consider susceptibility through known risk factors such as a positive family history of a specific injury, a history of other tendon and/or ligament ruptures and participation in sports associated with the specific musculoskeletal injuries. Athletes who are potential at higher risk than can then be managed personally by, for example, the prescription of the appropriate prehabilitation exercises to reduce risk.

CONCLUSION

Musculoskeletal soft tissue injuries are caused by a complex interaction of a number of biological and non-biological factors. There is evidence suggesting that there is in part a genetic contribution to an individual's risk of sustaining an injury. To date, several variants at known genetic loci have been implicated in risk. These associated variants may contribute to the interindividual variation in the structure and, by implication, mechanical properties of musculoskeletal soft tissues, as well as their response to mechanical loading and other stimuli.

The establishment of international consortia for the identification of genetic risk factors for these injuries will improve the limitations of the published hypothesis-driven genetic association studies. They will also allow resources to be pooled in recruiting large well-characterised cohorts required for whole-genome screening methods to identify biologically relevant variants involved in modulating injury risk.

Finally, clinicians need to be aware that the genetic tests for sports injuries, which are currently available DTC by several companies, are premature and that specific genetic tests are not necessary or informative to evaluate susceptibility to musculoskeletal soft tissue injury. Current practice should rather consider susceptibility through known risk factors such as a positive family history of a specific injury, a history of other tendon and/or ligament ruptures and participation in sports associated with the specific musculoskeletal injuries. These susceptible athletes may then be individually managed to reduce their risk profile.

REFERENCES

1 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.
2 Rahim M, Gibbon A, Hobbs H, *et al*. The association of genes involved in the angiogenesis-associated signaling pathway with risk of anterior cruciate ligament rupture. *J Orthop Res* 2014;32:1612–18.
3 Mannion S, Mtintsilana A, Posthumus M, *et al*. Genes encoding proteoglycans are associated with the risk of anterior cruciate ligament ruptures. *Br J Sports Med* 2014;48:1640–6.
4 Khoury LE, Posthumus M, Collins M, *et al*. ELN and FBN2 gene variants as risk factors for two sports-related musculoskeletal injuries. *Int J Sports Med* 2015;36:333–7.
5 O'Connell K, Knight H, Ficek K, *et al*. Interactions between collagen gene variants and risk of anterior cruciate ligament rupture. *Eur J Sport Sci* 2015;15:341–50.
6 Stępień-Słodkowska M, Ficek K, Maciejewska-Karłowska A, *et al*. Overrepresentation of the COL3A1 AA genotype in Polish skiers with anterior cruciate ligament injury. *Biol Sport* 2015;32:143–7.
7 Stępien-Słodkowska M, Ficek K, Eider J, *et al*. The+1245 g/t polymorphisms in the collagen type I alpha 1 (col1a1) gene in polish skiers with anterior cruciate ligament injury. *Biol Sport* 2013;30:57–60.
8 Stępień-Słodkowska M, Ficek K, Kaczmarczyk M, *et al*. The variants within the COL5A1 gene are associated with reduced risk of anterior cruciate ligament injury in skiers. *J Hum Kinet* 2015;45:103–11.
9 Ribbans WJ, Collins M. Pathology of the tendo Achillis: do our genes contribute? *Bone Joint J* 2013;95-B:305–13.
10 O'Connell K, Saunders CJ, Collins M. Collagen gene sequence variants in exercise-related traits. *Cent Eur J Sport Sci Med* 2013;1:3–17.
11 Saunders CJ, van der Merwe L, Cook J, *et al*. Extracellular matrix proteins interact with cell-signaling pathways in modifying risk of achilles tendinopathy. *J Orthop Res* 2015;33:898–903.
12 Khoury LE, Collins M, Handley CJ, *et al*. Polymorphic variation within the ADAMTS2, ADAMTS14, ADAMTS5, ADAM12 and TIMP2 genes and the risk of Achilles tendon pathology: a genetic association study. *J Sci Med Sport* 2013;16:493–8.
13 Altinisik J, Meric G, Erduran M, *et al*. The BstUI and DpnII variants of the COL5A1 Gene are associated with Tennis Elbow. *Am J Sports Med* 2015;43:1784–9.
14 Teerlink CC, Cannon-Albright LA, Tashjian RZ. Significant association of full-thickness rotator cuff tears and estrogen-related receptor-β (ESRRB). *J Shoulder Elbow Surg* 2015;24:e31–5.
15 Motta GDR, Amaral MV, Rezende E, *et al*. Evidence of genetic variations associated with rotator cuff disease. *J Shoulder Elbow Surg* 2014;23:227–35.
16 Peach CA, Zhang Y, Dunford JE, *et al*. Cuff tear arthropathy: evidence of functional variation in pyrophosphate metabolism genes. *Clin Orthop Relat Res* 2007;462:67–72.
17 Burger MC, de Wet H, Collins M. The BGN and ACAN genes and carpal tunnel syndrome. *Gene* 2014;551:160–6.
18 Burger MC, de Wet H, Collins M. Interleukin and growth factor gene variants and risk of carpal tunnel syndrome. *Gene* 2015;564:67–72.
19 Burger M, de Wet H, Collins M. The COL5A1 gene is associated with increased risk of carpal tunnel syndrome. *Clin Rheumatol* 2015;34:767–74.
20 Tully LJ, Murphy AM, Smith RKW, *et al*. Polymorphisms in TNC and COL5A1 genes are associated with risk of superficial digital flexor tendinopathy in National Hunt Thoroughbred racehorses. *Equine Vet J* 2014;46:289–93.
21 Baird AEG, Carter SD, Innes JF, *et al*. Genetic basis of cranial cruciate ligament rupture (CCLR) in dogs. *Connect Tissue Res* 2014;55:275–81.
22 Baird AEG, Carter SD, Innes JF, *et al*. Genome-wide association study identifies genomic regions of association for cruciate ligament rupture in Newfoundland dogs. *Anim Genet* 2014;45:542–9.
23 September AV, Schwellnus MP, Collins M. Tendon and ligament injuries: the genetic component. *Br J Sports Med* 2007;41:241–6.

What are the findings?

► Genetic and non-genetic factors play an important role in the aetiology of both acute and chronic musculoskeletal soft tissue injuries.

► The associated variants may contribute to the interindividual variation in the structure and functional properties of musculoskeletal soft tissues, as well as to their response to mechanical loading and other stimuli.

► The establishment of international consortia able to recruit large cohorts for the identification of injury genetic risk factors and the implantation whole-genome screening methods will improve the limitations of the published hypothesis-driven genetic association studies.

► Direct-to-consumer genetic tests for sports injuries are premature.

► In clinical practice personalised medicine can already be considered in the management of musculoskeletal soft tissue injuries without any genetic test.

24   Collins M, Raleigh SM. Genetic risk factors for musculoskeletal soft tissue injuries. *Med Sport Sci* 2009;54:136–49.

25   Gibson WT. Genetic association studies for complex traits: relevance for the sports medicine practitioner. *Br J Sports Med* 2009;43:314–16.

26   Khan KM, Scott A. Mechanotherapy: how physical therapists' prescription of exercise promotes tissue repair. *Br J Sports Med* 2009;43:247–52.

27   September AV, Cook J, Handley CJ, *et al*. Variants within the *COL5A1* gene are associated with Achilles tendinopathy in two populations. *Br J Sports Med* 2009;43:357–65.

28   Mokone GG, Gajjar M, September AV, *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.

29   Collins M, Posthumus M, Schwellnus MP. The COL1A1 gene and acute soft tissue ruptures. *Br J Sports Med* 2010;44:1063–4.

30   Posthumus M, September AV, O'cuinneagain D, *et al*. The association between the *COL12A1* gene and anterior cruciate ligament ruptures. *Br J Sports Med* 2010;44:1160–5.

31   Posthumus M, September AV, O'cuinneagain D, *et al*. The *COL5A1* gene is associated with increased risk of anterior cruciate ligament ruptures in female participants. *Am J Sports Med* 2009;37:2234–40.

32   Nell EM, van der Merwe L, Cook J, *et al*. The Apotosis Pathway and the Genetic Predisposition to the Achilles Tendinopathy. *J Orthop Res* 2012;30:1719–24.

33   Posthumus M, Collins M, van der Merwe L, *et al*. Matrix metalloproteinase genes on chromosome 11q22 and the risk of anterior cruciate ligament (ACL) rupture. *Scand J Med Sci Sports* 2012;22:523–33.

34   Malila S, Yuktanandana P, Saowaprut S, *et al*. Association between matrix metalloproteinase-3 polymorphism and anterior cruciate ligament ruptures. *Genet Mol Res* 2011;10:4158–65.

35   Raleigh SM, van der Merwe L, Ribbans WJ, *et al*. Variants within the MMP3 gene are associated with Achilles tendinopathy: possible interaction with the COL5A1 gene. *Br J Sports Med* 2009;43:514–20.

36   Jin H, van't Hof RJ, Albagha OME, *et al*. Promoter and intron 1 polymorphisms of *COL1A1* interact to regulate transcription and susceptibility to osteoporosis. *Hum Mol Genet* 2009;18:2729–38.

37   Khoschnau S, Melhus H, Jacobson A, *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.

38   Posthumus M, September AV, Keegan M, *et al*. Genetic risk factors for anterior cruciate ligament ruptures: COL1A1 gene variant. *Br J Sports Med* 2009;43:352–6.

39   Ficek K, Cięszczyk P, Kaczmarczyk M, *et al*. Gene variants within the *COL1A1* gene are associated with reduced anterior cruciate ligament injury in professional soccer players. *J Sci Med Sport* 2013;16:396–400.

40   Posthumus M, September AV, Schwellnus MP, *et al*. Investigation of the Sp1-binding site polymorphism within the COL1A1 gene in participants with Achilles tendon injuries and controls. *J Sci Med Sport* 2009;12:184–9.

41   Erduran M, Altinisik J, Meric G, *et al*. Is Sp1 binding site polymorphism within *COL1A1* gene associated with tennis elbow? *Gene* 2014;537:308–11.

42   Abrahams Y, Laguette M-J, Prince S, *et al*. Polymorphisms within the *COL5A1* 3'-UTR that alters mRNA structure and the MIR608 gene are associated with Achilles tendinopathy. *Ann Hum Genet* 2013;77:204–14.

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

44   Hay M, Patricios J, Collins R, *et al*. Association of type XI collagen genes with chronic Achilles tendinopathy in independent populations from South Africa and Australia. *Br J Sports Med* 2013;47:569–74.

45   September AV, Posthumus M, van der Merwe L, *et al*. The *COL12A1* and *COL14A1* genes and Achilles tendon injuries. *Int J Sports Med* 2008;29:257–63.

46   Laguette M-J, Abrahams Y, Prince S, *et al*. 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.

47   Ficek K, Stepien-Slodkowska M, Kaczmarczyk M, *et al*. Does the A9285G Polymorphism in Collagen Type XII α1 gene associate with the risk of anterior cruciate ligament ruptures? *Balkan J Med Genet* 2014;17:41–6. .

48   Taft RJ, Pang KC, Mercer TR, *et al*. Non-coding RNAs: regulators of disease. *J Pathol* 2010;220:126–39.

49   Zhang B, Wang Q, Pan X. MicroRNAs and their regulatory roles in animals and plants. *J Cell Physiol* 2007;210:279–89.

50   Sayed D, Abdellatif M. MicroRNAs in development and disease. *Physiol Rev* 2011;91:827–87.

51   Tucker R, Collins M. Athletic performance and risk of injury: can genes explain all? *Dialogues Cardiovasc Med* 2012;17:31–9.

52   Battié MC, Levalahti E, Videman T, *et al*. Heritability of lumbar flexibility and the role of disc degeneration and body weight. *J Appl Physiol* 2008;104:379–85.

53   Maes HH, Beunen GP, Vlietinck RF, *et al*. Inheritance of physical fitness in 10-yr-old twins and their parents. *Med Sci Sports Exerc* 1996;28:1479–91.

54   Birk DE. Type V collagen: heterotypic type I/V collagen interactions in the regulation of fibril assembly. *Micron* 2001;32:223–37.

55   Mazumder B, Seshadri V, Fox PL. Translational control by the 3'-UTR: the ends specify the means. *Trends Biochem Sci* 2003;28:91–8.

56   Xie X, Lu J, Kulbokas EJ, *et al*. Systematic discovery of regulatory motifs in human promoters and 3'-UTRs by comparison of several mammals. *Nature* 2005;434:338–45.

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

58   Wenstrup RJ, Smith SM, Florer JB, *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.

59   Voisin S, Eynon N, Yan X, *et al*. Exercise training and DNA methylation in humans. *Acta Physiol (Oxf)* 2015;213:39–59.

60   Ohi T, Uehara Y, Takatsu M, *et al*. Hypermethylation of CpGs in the promoter of the COL1A1 gene in the aged periodontal ligament. *J Dent Res* 2006;85:245–50.