Received: 15 November 2019 | Accepted: 3 January 2020

DOI: 10.1002/jor.24585

Journal of Orthopaedic Research®

RESEARCH ARTICLE

Exploring new genetic variants within *COL5A1* intron 4-exon 5 region and TGF-β family with risk of anterior cruciate ligament ruptures

Mary-Jessica N. Laguette[1,2,3] | Kelly Barrow[4] | Firzana Firfirey[1,2,3] | Senanile Dlamini[1,2,3] | Colleen J. Saunders[5,6] | Collet Dandara[4] | Junaid Gamieldien[5] | Malcolm Collins[1,2,3] | Alison V. September[1,2,3]

[1]Division of Exercise Science and Sports Medicine (ESSM), University of Cape Town, Cape Town, South Africa

[2]International Federation of Sports Medicine (FIMS) Collaborative Centre of Sports Medicine, ESSM, University of Cape Town, Cape Town, South Africa

[3]Research Centre for Health Through Physical Activity and Sport, University of Cape Town, Cape Town, South Africa

[4]Department of Human Genetics, University of Cape Town, Cape Town, South Africa

[5]South African National Bioinformatics Institute/MRC Unit for Bioinformatics Capacity, University of the Western Cape, Cape Town, Bellville, South Africa

[6]Division of Emergency Medicine, Department of Surgery, University of Cape Town, Cape Town, South Africa

Correspondence
Prof Alison Victoria September, Division of Exercise Science and Sports Medicine (ESSM), University of Cape Town, P.O. Box 115, Newlands 7725, South Africa.
Email: alison.september@uct.ac.za

Funding information
Department of Science and Technology of South Africa; University of Cape Town Research Council; South African National Research Fund

Abstract

Variants within genes encoding structural and regulatory elements of ligaments have been associated with musculoskeletal soft tissue injury risk. The role of intron 4-exon 5 variants within the α1 chain of type V collagen (*COL5A1*) gene and genes of the transforming growth factor-β (TGF-β) family, *TGFBR3* and *TGFBI*, was investigated on the risk of anterior cruciate ligament (ACL) ruptures. A case-control genetic association study was performed on 210 control (CON) and 249 participants with surgically diagnosed ruptures (ACL), of which 147 reported a noncontact mechanism of injury (NON). Whole-exome sequencing data were used to prioritize variants of potential functional relevance. Genotyping for *COL5A1* (rs3922912 G>A, rs4841926 C>T, and rs3124299 C>T), *TGFBR3* (rs1805113 G>A and rs1805117 T>C), and *TGFBI* (rs1442 G>C) was performed using Taqman SNP genotyping assays. Significant overrepresentation of the G allele of *TGFBR3* rs1805113 was observed in CON vs ACL ($P = .014$) and NON groups ($P = .021$). Similar results were obtained in a female with the G allele (CON vs ACL: $P = .029$; CON vs NON: $P = .016$). The *TGFBI* rs1442 CC genotype was overrepresented in the female ACL vs CON ($P = .013$). Associations of inferred allele combinations were observed in line with the above results. *COL5A1* intron 4-exon 5 genomic interval was not associated with the risk of ACL ruptures. Instead, this novel study is the first to use this approach to identify variants within the TGF-β signaling pathway to be implicated in the risk of ACL ruptures. A genetic susceptibility interval was identified to be explored in the context of extracellular matrix remodeling.

KEYWORDS

betaglycan, collagen type V alpha 1 chain, genetic predisposition, knee ligament injury, transforming growth factor beta induced

[Correction added on 16 Jan 2020, after first online publication: Correspondence information was updated]

1 | INTRODUCTION

Anterior cruciate ligament (ACL) ruptures are severe musculoskeletal injuries within the athletic community with a lengthy recovery time and a high cost of surgical repair.[1] Additionally, the burden of such injuries is compounded by a life-long predisposition for other conditions such as arthritis.[2] Both intrinsic and extrinsic factors, including genetic components, have been implicated in the aetiology of these multifactorial conditions and collectively predispose an individual to the risk of musculoskeletal soft tissue injuries.[3]

Sequence variants within genes that encode structural or regulatory elements of these tissues were previously associated with the risk of musculoskeletal soft tissue injuries.[3] Several genetic variants within the 3′-untranslated region (UTR) of *COL5A1*, the gene which encodes the α1 chain of type V collagen, were associated with a range of injuries including chronic Achilles tendinopathy and ACL ruptures.[3] Specifically, the *COL5A1* rs12722 C>T CC genotype was underrepresented in participants with chronic tendinopathy and in female participants with ACL ruptures in South African (SA) population studies.[4,5] This genetic locus was further refined in an independent study that examined its association with Achilles tendinopathy in an Australian population.[6,7] In addition, the *COL5A1* rs12722 and rs13946 C>T variants have been implicated in modulating risk of lateral epicondylitis commonly known as tennis elbow in a Turkish population.[8] Inferred haplotypes constructed using these variants were also implicated with modulating risk of ACL injuries in recreational Polish skiers,[9] while those constructed from a combination of *COL5A1* rs12722, rs3196378 C>A, and rs71746744 AGGG indel were identified as risk modifiers in chronic Achilles tendinopathy and rupture in a British population.[10]

In vitro studies have shown increased *COL5A1* messenger RNA (mRNA) stability to be associated with the rs12722 TT genotype.[11] Allelic forms of this 3′-UTR may contain sites that confer differences in mRNA stability. Although the biological functions of the *COL5A1* 3′-UTR variants are unknown, this study implied that they are important for the regulation of this gene and modulation of injury risk. Importantly, type V collagen plays a pivotal role in regulating fibrillogenesis, and mutations within the *COL5A1* gene contribute to the classic form of Ehlers-Danlos syndrome (EDS), a severe inheritable connective tissue disorder.[12]

Likewise, the transforming growth factor-β (TGF-β) superfamily and receptors play a central role in musculoskeletal soft tissue development, remodeling, cell differentiation, and proliferation.[13] For instance, TGF-β1 coordinates cartilage and tendon differentiation in limb mesenchyme of mice as well as chick embryonic limb development.[14,15] Moreover, several studies have suggested the importance of TGF-β in collagen remodeling[16,17] and the application of TGF-β1 to injured rabbit ACLs significantly improved their healing response.[18] Of interest, the proteoglycan betaglycan, TGF-β receptor III (TGFβR3), is able to bind to TGF-β ligands and is involved in capturing and retaining TGF-β for presentation to the signaling receptors.[19] In addition, adhesion protein TGF-β induced (TGFβI) plays an important role in cell-collagen interactions.[20] These studies collectively highlight the significance of the TGF-β signaling family, which is comprised of multifunctional regulatory proteins, in the development and healing of connective tissues and fibrosis.

Gene variants within members of this signaling family have been investigated in the context of musculoskeletal soft tissue injuries, including chronic Achilles tendinopathy and ACL ruptures. The 5′-UTR of the growth differentiation factor 5 gene harbors a functional variant (rs1413383 A>G), which is associated with the risk of Achilles tendon pathology in a SA population[21] and ACL ruptures in a Chinese population.[22] However, this variant was not significantly associated with ACL ruptures in a SA population.[23] Furthermore, variant rs1800469 C>T localized to the promoter region of the *TGFB1* gene was not associated with the risk of Achilles tendinopathy.[21] A little is known about the involvement of other members of this superfamily and their receptors within the context of genetic susceptibility to ACL ruptures.

Therefore, this study aimed to investigate the relevance of previously unexplored variants within other regions of the *COL5A1* gene, namely an area spanning its intron 4-exon 5, and genes of the TGF-β superfamily as well as of their receptor, namely *TGFBR3* and *TGFBI*, for associations with ACL ruptures in a SA cohort. Guided by data generated by a recent whole-exome sequencing (WES),[24] this study aimed to uncover previously unstudied but potentially functional variants in the context of ACL injury susceptibility.

2 | METHODS

2.1 | Participant recruitment

A case-control genetic association study was conducted (level III of evidence). Four-hundred and fifty-nine physically active and unrelated S individuals of self-reported European Caucasian ancestry were recruited. Of these, 210 were asymptomatic control participants (CON: 118 males and 92 females) and 249 participants with surgically diagnosed ACL ruptures (ACL: 184 males and 65 females). All ruptures were clinically diagnosed and confirmed by ultrasound, magnetic resonance imaging, arthroscopy, or during surgery. ACL participants were recruited from the Sports Science Orthopaedic Clinic and other hospitals in Cape Town between 2006 and 2016. The participants were also grouped according to a noncontact mechanism of injury (NON subgroup: 147 participants). CON participants, with no history of musculoskeletal soft tissue injuries, were recruited simultaneously from sporting clubs and gyms in Cape Town. Questionnaires were completed specifying personal details, injury, sporting, and medical history. This study was approved by the Human Research Ethics Committee of the Faculty of Health Sciences, University of Cape Town, South Africa (HREC164/2006).

2.2 | DNA extraction

Approximately 5 mL of venous blood was collected from each participant and DNA extracted as previously described.[4] The DNA samples were stored at −20°C until further analysis.

2.3 | Genetic variant prioritization

WES was previously performed on 20 individuals using the SureSelect Human All Exon V5+UTRs (71 Mb) Capture Kit (Agilent Technologies, California) and subjected to paired-end sequencing on the Illumina HiSeq 2000/25000 platform at ×30 coverage (Otogenetics Corporation).[24] An established variant calling pipeline was applied in collaboration with the SA National Bioinformatics Institute, University of the Western Cape.[24] For the current study, 128 variants with an allele frequency difference ≥20% between cases and controls were localized to the genes of interest and were extracted for further interrogation. The genes and upstream/downstream regions were imported from NCBI and functional annotation was accomplished using data extracted from Uniprot,[25] DIANA-Tarbase v7.0,[26] and Ensembl[27] on the human reference genome (GRCh38) in Geneious v2.1.[28]

Within *COL5A1*, a region between intron 4 and exon 5 was investigated based on prior communication from a collaborator (I. Ahmetov; personal communication) and the implication of this region in splicing outcomes for EDS.[29,30] Variant rs3922912 G>A and rs4841926 C>T were thereby selected. From the WES data, rs3124299 C>T was also identified within exon 5 (≥30% allele frequency difference).[24] The latter is adjacent to a laminin G-like domain and two putative MicroRNA binding sites.

Similarly, genes within the TGF-β family were examined. From the WES study,[24] only two genes (*TGFBR3* and *TGFBI*) containing variants with high allele frequency differences within functional regions were prioritized. The rs1805113 G>A was identified in *TGFBR3* located on chromosome 1. It is a synonymous variant situated within exon 13 with a 35% allele frequency difference. It overlaps with a disulphide bond site and is within a zona pellucida domain. Three variants were identified within the 3′-UTR and were in linkage disequilibrium. A 30% allele frequency difference was noted for rs1805117 T>C and it was, therefore, selected for genotyping. One variant, rs1442 G>C, in exon 6 of the *TGFBI* located on chromosome 5 was identified using the criteria above. It is located within a fasciclin-like domain and had a 20% allele frequency difference. All the variants were confirmed using Sanger sequencing (Inqaba Biotec, Pretoria, South Africa).

2.4 | Genotyping of ACL cohort

Informed by the prioritization, all participants were genotyped for *COL5A1* rs3922912 G>A, rs4841926 C>T and rs3124299 C>T, *TGFBR3* rs1805113 G>A and rs1805117 T>C, as well as *TGFBI* rs1442 G>C using Taqman SNP genotyping assays and Taqman Universal genotyping Master mix (Applied Biosystems, Waltham, MA) following the manufacturer's recommendations on the Applied Biosystems QuantStudio 3 Real-Time PCR System (Applied Biosystems). The genotypes were called using the Thermo Fisher Cloud suite (Thermo Fisher Scientific, Waltham). All laboratory work was conducted at the Division of Exercise Science and Sports Medicine, University of Cape Town.

2.5 | Statistical analyses

Power analysis was performed using QUANTO v1.2.4 (http://hydra.usc.edu/gxe). Minor allele frequencies (EUR population) were taken from the 1000 Genomes Project. Assuming allele frequencies between 0.1 and 0.9 for the risk allele of each variant investigated and a dominant model of inheritance, our current sample size would be adequate to detect an allelic odds ratio (OR) of 1.2 and greater at a power of 80% at $P < .05$. Genotypes were analyzed using the programming environment R[31] and SNPassoc package (version 1.9.2).[32] Hardy-Weinberg equilibrium (HWE) was determined. Differences between genotype and allele frequencies were assessed using Pearson's χ^2 and Fisher's exact test. A post hoc analysis considering inheritance models was performed to identify the overrepresented genotype. Sex stratification was performed as it remains an intrinsic risk factor in the aetiology of ACL ruptures. Genotype data were used to generate allele constructs using the haplo.stats package (version 1.7.6) and frequencies compared using haplo.score.[33] All inferred allele constructs occurring at a frequency of less than 1% were excluded. A significance level of $P < .05$ was set for all statistical analyses.

3 | RESULTS

3.1 | Participant characteristics

Participant characteristics were compared between the CON (n = 210) and ACL groups (n = 249) as well as between the CON and NON subgroups (n = 162) (Table S1). ACL and NON participants were matched for mean age in comparison to CON. There were significantly more men in the ACL group (73.9%, n = 184, $P < .001$) and NON subgroup (71%, n = 115, $P = .003$) when compared to CON (56.2%, n = 118). The ACL group (177.4 ± 9.9 cm, $P = .003$) and NON subgroup (177.1 ± 10.4 cm, $P = .020$) were also significantly taller than the CON group (174.6 ± 9.7 cm). Participants within the ACL group (80.7 ± 16.5 kg, $P < .001$) and NON subgroup (79 ± 16.1 kg, $P < .001$) were heavier than CON (73.4 ± 14.4 kg). The body mass index (BMI) was higher in both the ACL group (25.3 ± 4.4 kg/m^2, $P < .001$) and NON subgroup (25.0 ± 4.4 kg/m^2, $P = .015$) in comparison to CON (23.9 ± 3.3 kg/m^2). When adjusted for sex as a confounder, height (CON vs ACL/NON) and BMI (CON vs NON) differences were no longer significant. More CON participants (83%) were born in South Africa (73.9%, $P = .035$ and 77.2%, $P = .023$, respectively) compared to ACL group or NON subgroup. However, all participants were of the same self-reported ancestral background.

Male and female ACL and NON were matched to the CON group for sports participation in contact sports, noncontact jumping sports, noncontact nonjumping sports, and skiing sports (Tables S2 and S3).

Descriptive characteristics of CON and ACL participants were grouped according to genotype (Table S4). For the *COL5A1* rs4841926, *COL5A1* rs3124299, *TGFBR3* rs1805113, and the *TGFBI* rs1442 loci, there were no genotype effects on age, weight, height,

BMI, sex, and country of birth (COB). For the *COL5A1* rs3922912 G>A, there were no genotype effects on height, weight, BMI, sex and COB, except for age ($P = .033$). Individuals with the AA genotype were younger (26.1 ± 8.9 years) than individuals with either the AG (26.8 ± 10.4 years) or GG genotype (27.3 ± 9.9 years). At the *TGFBR3* rs1805117 T>C locus, there were no genotype effects on age, weight, BMI, sex, and COB. However, there were significant genotype effects on height ($P = .026$). Individuals with the CC genotype were taller (184.9 ± 7.0 cm) than individuals with either the TC (175.2 ± 10.1 cm) or TT genotype (height, 176.2 ± 9.8 cm). When covaried for sex, observed genotyping effects were no longer significant.

3.2 | Genotype and allele frequencies

3.2.1 | *COL5A1*

When variants rs3922912 G>A, rs4841926 C>T, and rs3124299 C>T within the intron 4-exon 5 genomic interval of *COL5A1* were examined, no significant differences were noted for the genotype nor allele frequency distributions in any of the group comparisons (Tables 1 and 2; Table S5). For variants rs3922912 and rs3124299, all three groups were in HWE. However, the *COL5A1* rs4841926 CC genotype was absent in our studied population groups.

3.2.2 | *TGFBR3*

There was a significant difference in rs1805113 G>A genotype ($P = .033$) frequency distribution between the CON and ACL groups (Table 1). More specifically, the GG genotype was overrepresented in the CON group (24%, n = 50) compared to the ACL group (15%, n = 38) ($P = .010$, OR: 0.48, 95% CI: 0.286-0.848) (Table 1). Furthermore, a significant difference was observed in the allele frequency distribution between the CON and ACL ($P = .014$) as well as between the CON and NON groups ($P = .021$) where the G allele was underrepresented in the ACL and NON groups. All three groups were in HWE. After stratifying for sex, no significant differences in frequency distributions were observed between any of the group comparisons when only male participants were evaluated (Table S5). However, when only females were evaluated, a significant difference between the genotype ($P = .036$) and allele frequency distributions ($P = .016$) was noted between CON and NON subgroups. Specifically, the GG genotype was overrepresented in the female CON group (26%, n = 24) compared to the NON subgroup (13%, n = 6) ($P = .017$, OR: 0.30, 95% CI: 0.11-0.80). Lastly, a trend was observed for females between CON (n = 92) and ACL (n = 65) overall genotype frequencies ($P = .054$).

For the *TGFBR3* rs1805117 T>C locus, no significant differences in genotype or allele frequency were observed between all groups in all participants and after stratifying for sex (Tables 1 and 2 and Table S5). All three groups were in HWE. Interestingly, the CC genotype was absent within the female groups (Table 2).

3.2.3 | TGFBI

For the rs1442 G>C locus, no significant differences in genotype ($P = .331$) or allele frequency distributions ($P = .231$) were observed between the CON group compared to the ACL ($P = .389$) or NON (p = 0.182) groups when all participants were evaluated (Table 1). All three groups were in HWE. However, after stratifying for sex, significant differences were noted when only females were compared (Table 2) in both the genotype ($P = .036$) and allele ($P = .021$) frequencies between the female CON and ACL groups. Specifically, the CC genotype was underrepresented in the CON (13%, n = 12; $P = .013$) vs ACL groups (29%, n = 19; OR: 0.31, 95% CI: 0.12-0.77). In addition, the C allele was significantly underrepresented in the CON group (38%, n = 70) vs the ACL group (52%, n = 68; OR: 0.57, 95% CI: 0.352-0.922).

3.3 | Inferred haplotypes

3.3.1 | *COL5A1*

Inferred haplotypes were constructed for *COL5A1* gene using genotyping data from the three variants studied (rs3922912 G>A-rs4841926 C>T, and rs3124299 C>T) within the region of interest, intron 4-exon 5 (20,151 bp) (Figure S1). However, no significant differences or trends in their distribution were noted for the *COL5A1* haplotypes, even when the participants were stratified for sex.

3.3.2 | *TGFBR3*

Inferred haplotypes were constructed for the *TGFBR3* gene using the rs1805113 G>A and rs1805117 T>C genotype data (Figure 1). The most frequently occurring inferred haplotype A-T, (CON: 47.9%, n = 101; ACL: 52.6%, n = 131) was significantly overrepresented in the ACL group ($P = .043$; OR: 1.18, 95% CI: 0.866-1.612). While the alternate haplotype G-C haplotype was significantly overrepresented ($P = .009$; OR: 0.58, 95% CI: 0.346-0.965) in the CON group (CON: 16.0%, n = 34; ACL: 9.2%, n = 23).

When only male participants (CON: n = 118, ACL: n = 184) were compared, no significant differences were noted in the frequency distributions (Figure 1b). However, when females were compared (Figure 1c; CON: n = 92, ACL: n = 65), the A-T inferred haplotype was significantly overrepresented ($P = .027$; OR: 1.46, 95% CI: 0.840-2.547) in the ACL group (CON: 46.0%, n = 54; ACL = 56.6%, n = 104). The alternate G-C inferred haplotype (CON: 16.6%, n = 15, ACL: 8.1%, n = 5) was also found to be significantly overrepresented in the CON group ($P = .033$; OR: 0.51, 95% CI: 0.188-1.379).

3.4 | *TGFBR3* and *TGFBI* gene-gene interactions

Inferred allele combinations were constructed for the *TGFBR3* and *TGFBI* genes using the rs1805113 G>A, rs1805117 T>C, and rs1442

TABLE 1 Genotype and minor allele frequency distributions of *COL5A1* rs3922912 G>A, *COL5A1* rs4841926, C>T, *COL5A1* rs3124299 C>T, *TGFBR3* rs1805113 G>A, *TGFBR3* rs1805117 T>C, and *TGFBI* rs1442 G>C variants in all participants in the control group, anterior cruciate ligament rupture group, and subgroup with a noncontact mechanism of injury within this South African cohort

ALL		CON	ACL	*P* value[a]	NON[b]	*P* value
COL5A1 rs3922912	n	209	223		141	
	AA	17 (36)	15 (33)	.309	15 (21)	.765
	AG	47 (98)	54 (121)		53 (75)	
	GG	36 (75)	31 (69)		32 (45)	
	A	41 (170)	58 (187)	.730	41 (117)	.876
	HWE	0.669	0.101		0.301	
COL5A1 rs4841926	n	210	246		162	
	CC	…	…	.809	…	.476
	CT	44 (92)	43 (105)		40 (65)	
	TT	56 (118)	57 (141)		60 (97)	
	C	22 (92)	21 (105)	.746	20 (65)	.587
	HWE	…	…		…	
COL5A1 rs3124299	n	210	248		161	
	CC	32 (67)	26 (64)	.301	27 (44)	.906
	CT	50 (104)	56 (139)		54 (87)	
	TT	19 (39)	18 (45)		19 (30)	
	T	43 (182)	46 (229)	.424	46 (147)	.551
	HWE	1.000	0.055		0.340	
TGFBR3 rs1805113	n	210	248		161	
	AA	24 (51)	32 (80)	**.033**	34 (54)	.138
	AG	52 (109)	52 (130)	**(.010)**[c]	51 (82)	
	GG	24 (50)	15 (38)		16 (25)	
	G	50 (209)	42 (206)	**.014**[d]	41 (132)	**.021**[e]
	HWE	0.679	0.241		0.625	
TGFBR3 rs1805117	n	210	249		162	
	TT	65 (136)	72 (180)	.115	73 (119)	.110
	TC	34 (71)	25 (63)		25 (40)	
	CC	1 (3)	2 (6)		2 (3)	
	C	18 (77)	15 (75)	.211	14 (46)	.137
	HWE	0.068	0.806		1.00	
TGFBI rs1442	n	209	249		162	
	CC	18 (37)	23 (58)	.331	24 (39)	.389
	CG	53 (111)	50 (125)		51 (82)	
	GG	29 (61)	27 (66)		25 (41)	
	C	44 (185)	48 (241)	.231	49 (160)	.182
	HWE	0.327	1.00		1.00	

Note: Relative frequencies are expressed as % with the number of participants (n) in parentheses. The *P* value for Hardy-Weinberg equilibrium exact test is provided and significance indicated in bold typeset ($P < .05$).

Abbreviations: ACL, anterior cruciate ligament; CI, confidence interval; CON, control; HWE, Hardy-Weinberg equilibrium; NON, noncontact; OD, odds ratio.

[a]CON vs ACL.

[b]CON vs NON.

[c]AA vs GG genotype CON vs ACL ($P = .0097$, OR: 0.48, 95% CI: 0.286-0.848).

[d]Allele frequency distributions CON vs ACL ($P = .014$, OR: 0.72, 95% CI: 0.547-0.940).

[e]Allele frequency distributions CON vs NON ($P = .021$, OR: 0.70, 95% CI: 0.517-0.950).

G>C genotyping data to study possible gene-gene interactions in conferring risk (Figure 2). The G-C-G allele combination was significantly overrepresented in the CON (11.2%, n = 24) compared to the ACL (5.5%, n = 85) group ($P = .0067$; OR: 0.44, 95% CI: 0.214-0.888).

When stratified by sex, the same G-C-G inferred allele combination was significantly overrepresented in the CON (14.1%, n = 13) compared to the ACL (4.6%, n = 3) group when only females were compared (CON: n = 92, ACL: n = 65; $P = .011$; OR: 0.428, 95% CI: 0.130-1.41) (Figure 2b). In contrast, the alternate A-T-C inferred allele combination was significantly overrepresented ($P = .0010$; OR: 2.60, 95% CI: 1.04-6.54) in the ACL group (CON: 16.7%, n = 15; ACL: 34.2%, n = 22) (Figure 2c).

4 | DISCUSSION

ACL ruptures are debilitating multifactorial injuries, where a number of extrinsic and intrinsic factors interact with risk modulation. The aim of this study was to investigate previously unexplored variants in

TABLE 2 Genotype and minor allele frequency distributions of *COL5A1* rs3922912 G>A, *COL5A1* rs4841926 C>T, *COL5A1* rs3124299 C>T, *TGFBR3* rs1805113 G>A, *TGFBR3* rs1805117 T>C, and *TGFBI* rs1442 G>C variants in female participants in the control group, anterior cruciate ligament rupture group, and subgroup with a noncontact mechanism of injury within this South African cohort

Female		CON	ACL	*P* value[a]	NON[b]	*P* value
COL5A1 rs3922912	n	91	58		41	
	AA	22 (20)	22 (13)	.422	22 (9)	.280
	AG	41 (37)	50 (29)		54 (22)	
	GG	37 (34)	28 (16)		24 (10)	
	A	42 (77)	47 (55)	.404	49 (40)	.351
	HWE	0.132	1.000		0.760	
COL5A1 rs4841926	n	92	64		47	
	CC	…	…	.279	…	.192
	CT	48 (44)	39 (25)		36 (17)	
	TT	52 (48)	61 (39)		64 (30)	
	C	24 (44)	20 (25)	.407	18 (17)	.288
	HWE	…	…		…	
COL5A1 rs3124299	n	92	65		47	
	CC	34 (31)	23 (15)	.193	21 (10)	.177
	CT	46 (42)	60 (39)		62 (29)	
	TT	21 (19)	17 (11)		17 (8)	
	T	43 (80)	47 (61)	.566	48 (45)	.525
	HWE	0.526	0.138		0.150	
TGFBR3 rs1805113	n	*92*	*65*		*47*	
	AA	22 (20)	34 (22)	.054	40 (19)	**.036**
	AG	52 (48)	54 (35)		47 (22)	**(.017)**[c]
	GG	26 (24)	12 (8)		13 (6)	
	G	52 (96)	39 (51)	**.029**[d]	36 (34)	**.016**[e]
	HWE	0.834	0.435		1.000	
	n	*92*	*65*		*47*	
TGFBR3 rs1805117	TT	63 (58)	75 (49)	.099	79 (37)	.055
	TC	37 (34)	25 (16)		21 (10)	
	CC	…	…		…	
	C	18 (34)	12 (16)	.160	11 (10)	.118
	HWE	…	…		…	
TGFBI rs1442	n	91	65		47	
	GG	36 (33)	25 (16)	**.036**	26 (12)	.224
	CG	51 (46)	46 (30)	**(.013)**[f]	51 (24)	
	CC	13 (12)	29 (19)		23 (11)	
	C	38 (70)	52 (68)	**.021**[g]	49 (46)	.122
	HWE	0.658	0.619		1.000	

Note: Relative frequencies are expressed as % with the number of participants (n) in parentheses. The *P* value for Hardy-Weinberg equilibrium (HWE) exact test is provided and significance indicated in bold typeset ($P < .05$).

Abbreviations: ACL, anterior cruciate ligament; CI, confidence interval; CON, control; HWE, Hardy-Weinberg equilibrium; NON, noncontact; OD, odds ratio.

[a]CON vs ACL.

[b]CON vs NON.

[c]AA vs GG genotype CON vs NON ($P = .017$, OR: 0.30, 95% CI: 0.11-0.80).

[d]Allele frequency distributions CON vs ACL ($P = .029$, OR: 1.69, 95% CI: 1.045-2.737).

[e]Allele frequency distributions CON vs NON ($P = .016$, OR: 1.92, 95% CI: 1.122-3.323).

[f]CC vs GG genotype CON vs ACL ($P = .013$, OR: 0.31, 95% CI: 0.12-0.77).

[g]Allele frequency distributions CON vs ACL ($P = .021$, OR: 0.57, 95% CI: 0.352-0.922).

the context of ACL injury risk in a SA cohort. To this aim, variants were annotated and prioritized for genotyping within the intron 4-exon 5 genomic intervals of the *COL5A1* and the TGF-β encoding genes, *TGFBR3* and *TGFBI*, using a combination of prior knowledge, de novo generated WES data[24] and a number of bioinformatics tools. The following loci were examined: *COL5A1* rs3922912 G>A, *COL5A1* rs4841926 C>T, *COL5A1* rs3124299 C>T, *TGFBR3* rs1805113 G>A, *TGFBR3* rs1805117 T>C, and *TGFBI* rs1442 G>C.

This is the first study to identify genetic associations implicating *TGFBR3* and *TGFBI* genes in ACL injury risk modulation. An independent association of *TGFBR3* rs1805113 G allele with a decreased risk of ACL injury was noted when all participants and when only females were compared. The *TGFBR3* rs1805113 A>G-rs1805117 T>C inferred haplotype distributions further supported these findings and highlighted a specific genetic interval to potentially be explored for functionality. The A-T inferred haplotype

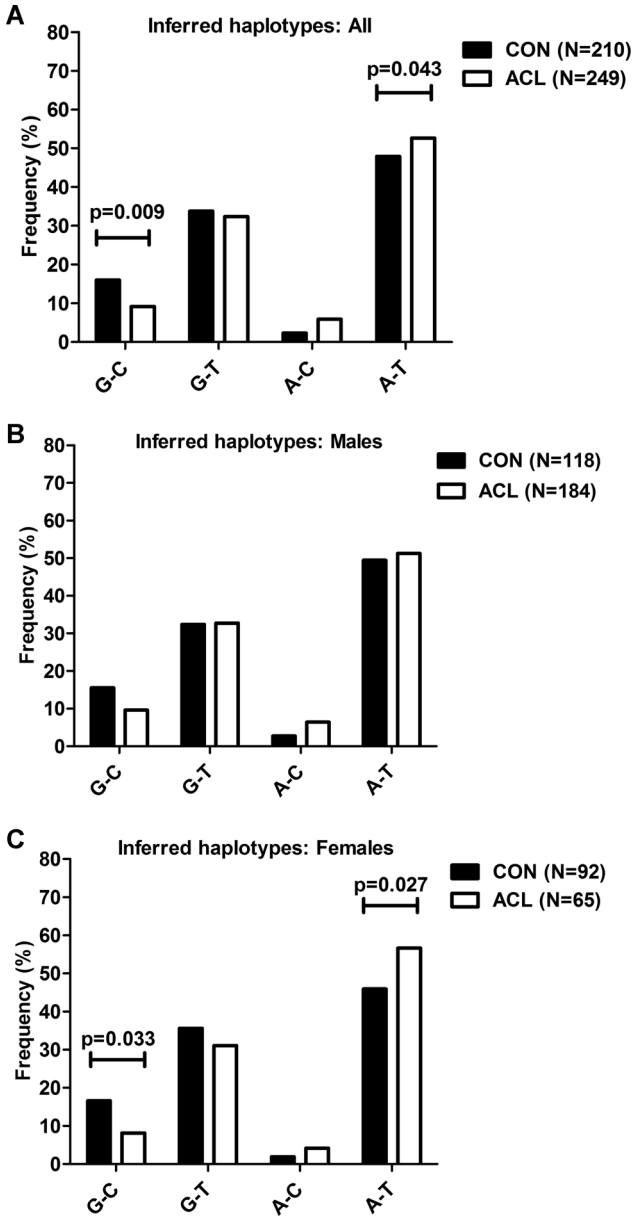

FIGURE 1 Inferred haplotype frequency distributions from the *TGFBR3* rs1805113 G>A, rs1805117 T>C in the control group (CON; black bars) and the anterior cruciate ligament rupture group (ACL; white bars) for (A) all the participants combined, (B) the male participants, and (C) the female participants. Statistically significant differences in haplotype frequency between the groups are depicted on the graph with the *P* values shown above the bars. The number of participants (N) in each group is in parentheses

was associated with increased ACL injury susceptibility when all participants and when only females were evaluated (1.2- and 1.5-fold increased risk, respectively). Conversely, the alternate G-C inferred haplotype was associated with a 0.6- and 0.5-fold decreased risk when all participants or when only females were compared, respectively.

Although we do not currently have an understanding of the functional significance of these specific loci within *TGFBR3*, it is interesting to note that rs1805117 is a "tag" variant for rs1805115 and rs1805116, situated within the 3′UTR and each feature a 30% frequency difference between cases and controls in the WES data.[24] This locus was implicated in a Korean hepatocellular carcinoma population study[34] and was also associated with the risk of pulmonary emphysema in a family-based study.[35] It was proposed that there may be a differential codon usage at this interval which could lead to tissue-specific codon-mediated translational control of TGFβR3.[34] This polymorphism was, however, not associated with Marfan syndrome or related phenotype,[36] but interestingly, dermal TGFβR3 protein expression was positively correlated with the Marfan syndrome phenotype.[37] Furthermore, bioinformatic annotations revealed that rs1805113 was adjacent to sites crucial for the formation of disulphide bonds and glycosylation. The genetic interval contains a large "topological associating domain" characterized by highly conserved large chromatin interaction regions[38] and is within a zona pellucida domain.[39] The latter plays an important role in polymerization of extracellular proteins acting as a module for assembly into filaments and/or matrices.[39] This region may play a fundamental role in regulating expression and the function of the receptor. We hypothesize that potentially the genomic interval flagged in this study requires further functional analyses to characterize the biological and clinical significance of this region in the pathogenesis of ligament injury.

In addition, this study is the first to identify an independent association between sex-linked injury risk and the C allele of *TGFBI* rs1442. There was a 0.3-fold overrepresentation of the CC genotype in the ACL participants vs controls in this study. *TGFBI* rs1442 is a synonymous SNP and mutations in this gene have been associated with several autosomal dominant corneal dystrophies characterized by severe visual impairment with progressive accumulation of TGFβI and amyloid-like deposits in the corneal matrix.[40] Bioinformatics analyses revealed that rs1442 is adjacent to sites important for calcium binding and within a fasciclin-like domain which is a cell adhesion domain common to many secreted and membrane-anchored proteins.[41] Interestingly, in this specific gene, this type of domain is involved in the molecular mechanisms underlying corneal dystrophies via binding of integrins[41] as well as in the circulating TGFβI antiangiogenic and antitumorigenic potential.[42]

Previous independent research has implicated the 3′UTR of the *COL5A1* gene to be associated with risk modulation of ACL ruptures, Achille tendinopathy, carpal tunnel syndrome, tennis elbow, and other exercise-related phenotypes.[3] The intron 4-exon 5 region was investigated based on prior communication (I. Ahmetov; personal communication) and the implication of this region in different splicing outcomes for EDS.[29,30] Furthermore, the variant prioritization pipeline using the WES data identified the intron 4-exon 5 genomic interval of *COL5A1* as an area that warrants further investigation. However, no independent associations were noted between any of the intron 4-exon 5 region *COL5A1* rs3922912, rs4841926, and rs3124299 variants and ACL injury risk. Similarly, the evaluation of the *COL5A1* inferred haplotypes (rs3922912, rs4841926, and rs3124299) did not implicate these three variants in ACL injury risk.

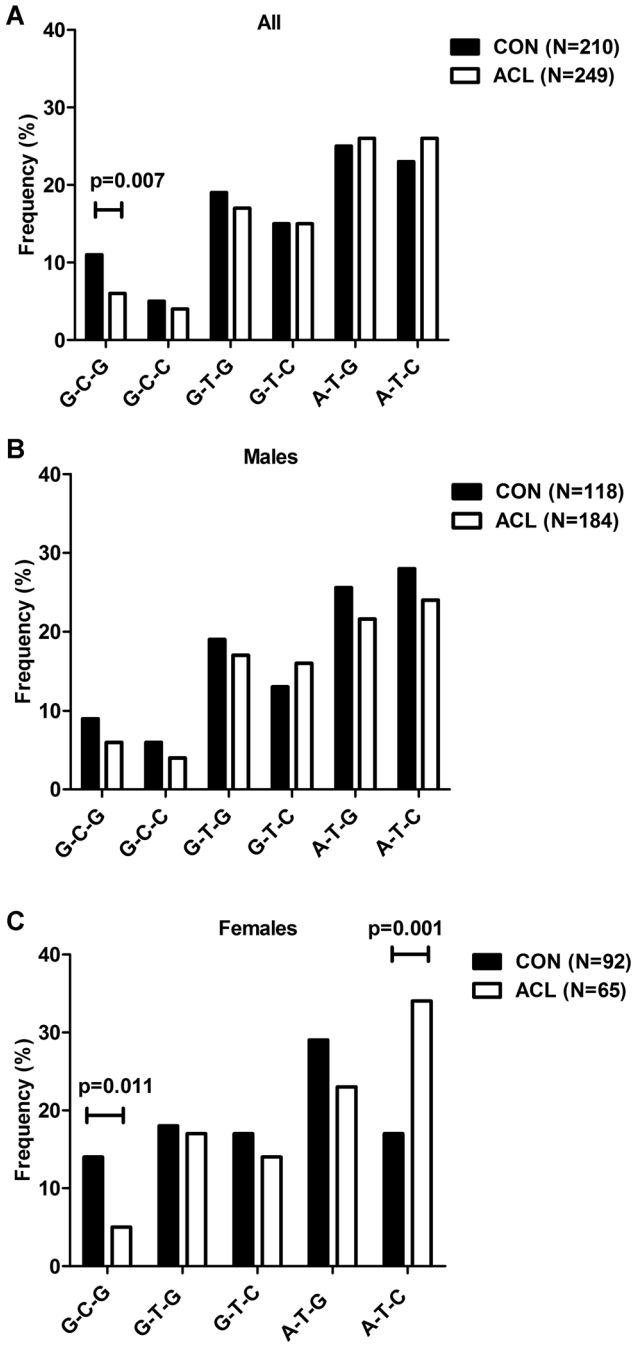

FIGURE 2 Inferred allele combinations from the *TGFBR3* rs1805113 G>A, *TGFBR3* rs1805117 T>C, and *TGFBI* rs1442 G>C variants in the control group (CON; black bars) and the anterior cruciate ligament rupture group (ACL; white bars) for (A) all the participants combined, (B) the male participants, and (C) the female participants. Statistically significant differences in inferred allele combination frequency between the groups are depicted on the graph with the *P* values shown above the bars. The number of participants (N) in each group is in parentheses

There is evidence to suggest that TGFβR3 is a membrane proteoglycan receptor modulating TGF-β access to signaling receptors and enhances the TGF-β pathway.[43] When TGFβR3 is expressed, TGFβR1 and TGFβR2 bind to sites on TGFβR3 competing with the normal signaling complex TGFβR1/TGFβR2 which inhibits Smad2/3 signaling.[44] Deletion of the TGFβR3 cytoplasmic domain results in loss of TGFβR3 clustering which can no longer compete with the TGFβR1/TGFβR2 signaling complexes,[44] thereby causing a dysregulation in the activation of the Smad2/3 signaling pathway. Downstream effects of this dysregulation include the lack of collagen deposition, collagen proliferation, and matrix turnover during early chondrogenesis. This dysregulation has similar effects on osteoblasts during bone development and healing.[45] On the other hand, the *TGFBI* gene is an Arg-Gly-Asp sequence motif (RGD)-containing protein which may be important for cell-collagen interactions; this motif is present in many extracellular matrix proteins modulating cell adhesion and can serve as a ligand recognition sequence for several integrins.[46] In addition, both proteins bind to a vastly diverse set of cell-type-specific integrins and are present in fibrillary networks, where TGFβI is shown to bind to tenascin-C, fibronectin, laminin, biglycan, decorin, and various collagen types such as types I, VI, or XII collagen.[46]

It is for these collective reasons that we investigated the allele combinations between *TGFBR3* and *TGFBI* with ACL risk susceptibility as a proxy of gene-gene interactions. The G-C-G allele combination was found to be overrepresented by 0.4-fold in control vs ACL participants and when female participants were examined separately. Of note, in females only and in line with the independent association analyses, the A-T-C allele combination was linked to a threefold increased risk of ACL. These analyses, therefore, provide preliminary data to suggest that collective interactions between these genes may affect ACL injury risk susceptibility.

When the variants were observed independently as well as when in inferred allele constructs, several of the significant findings observed in this study were noted in female participants only. Indeed, studies have shown that female athletes are three times more likely to rupture their ACL than male athletes.[47–49] Although the sample size is relatively small, this observation is not isolated and is supported by other genetic association studies on ACL ruptures.[5,49] Although the biological reason underpinning this sex-specific association is unknown, it has been suggested that the presence of intrinsic factors in females including increased quadriceps angle and increased posterior tibial slope[48] as well as hormonal differences may be responsible.[47] It is, therefore, reasonable that we start collecting evidence to better characterize the susceptibility underpinning sex-linked associations.

In addition, some significant differences, although small, were noted when participant characteristics were analyzed with respect to sex, weight, and BMI. However, no genotyping effects were observed on descriptive characteristics for the *TGFBR3* rs1805113 and *TGFBI* rs1442, while genotyping effects on age and height for *COL5A1* rs3922912 and *TFGBR3* rs1805117, respectively, were no longer significant when covarying for sex as a confounder. In addition, the cohort was well matched for sporting activities. The differences in the reported haplotype frequencies may in some cases be small, thus caution should be taken to avoid overinterpretation. The outcome of this novel study should be investigated in larger cohorts and in

independent populations. Exploration in different musculoskeletal soft tissue injuries should be considered.

5 | CONCLUSION

This genetic association study aimed to capture a genetic signature for musculoskeletal soft tissue injury susceptibility within (a) intron 4-exon 5 of *COL5A1*, and (b) the TGF-β encoding gene family which was not previously linked to musculoskeletal soft tissue injury susceptibility. This novel study implicates *TGFBR3* and *TGFBI* in modulating ACL injury risk. TGFβR3 plays a role in the TGF-β signaling pathway which regulates the process of embryonic tendon development and chondrogenesis while TGFβI participate in cell-collagen interactions. Further functional exploration is necessary to understand the biological implications for ligament rupture and to identify potential underlying mechanisms. The application of the WES data together with bioinformatics tools to prioritize potential functional variants was valuable towards shedding light on genes which had not previously been implicated in the aetiology of musculoskeletal soft tissue injuries such as ACL ruptures.

ACKNOWLEDGMENTS

The authors would like to thank Melanie Hay, Dr. Masouda Rahim, Dr. Michael Posthumus, and Dr. Dion O'Cuinneagain for assistance with participant recruitment, as well as Dr. Andrea Gibbon for her assistance with bioinformatics. This study was supported in part by the University of Cape Town Research Council, in part by the National Research Fund, and in part by the Department of Science and Technology. MC and AVS have filed patents on the application of specific sequence variations (not included in this study) related to risk assessment of ACL ruptures and Achilles tendinopathy.

AUTHOR CONTRIBUTIONS

M-JNL, KB, FF, SD, CJS, MC, and AVS substantially contributed to research design, or acquisition, analysis or interpretation of data; and M-JNL, KB, CJS, CD, JG, MC, and AVS drafted the paper or revised it critically. All authors approved the submitted and final versions.

ORCID

Mary-Jessica N. Laguette http://orcid.org/0000-0001-9979-117X

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;6:216-223.
2. Hall M, Bryant AL, Wrigley TV, et al. Does meniscal pathology alter gait knee biomechanics and strength post-ACL reconstruction? *Knee Surg Sports Traumatol Arthrosc*. 2016;24:1501-1509.
3. Rahim M, Gibbon A, Collins M, September AV. Chapter fifteen—genetics of musculoskeletal soft tissue injuries: current status, challenges, and future directions. In: Barh D, Ahmetov I, eds. *Sports, Exercise, and Nutritional Genomics*. Cambridge, MA: Academic Press; 2019:317-339.
4. Mokone GG, Schwellnus MP, Noakes TD, Collins M. The COL5A1 gene and Achilles tendon pathology. *Scand J Med Sci Sports*. 2006;16:19-26.
5. Posthumus M, September AV, O'Cuinneagain D, van der Merwe W, Schwellnus MP, Collins M. The COL5A1 gene is associated with increased risk of anterior cruciate ligament ruptures in female participants. *Am J Sport Med*. 2009;37:2234-2240.
6. September AV, Cook J, Handley CJ, van der Merwe L, Schwellnus MP, Collins M. Variants within the COL5A1 gene are associated with Achilles tendinopathy in two populations. *Brit J Sport Med*. 2009;43:357-365.
7. Abrahams Y, Laguette MJ, Prince S, Collins M. Polymorphisms within the COL5A1 3′-UTR that alters mRNA structure and the MIR608 gene are associated with Achilles tendinopathy. *Ann Hum Genet*. 2013;77:204-214.
8. Altinisik J, Meric G, Erduran M, Ates O, Ulusal AE, Akseki D. The BstUI and DpnII variants of the COL5A1 gene are associated with tennis elbow. *Am J Sport Med*. 2015;43:1784-1789.
9. 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;45:103-111.
10. Brown KL, Seale KB, El Khoury LY, et al. Polymorphisms within the COL5A1 gene and regulators of the extracellular matrix modify the risk of Achilles tendon pathology in a British case-control study. *J Sport Sci*. 2017;35:1475-1483.
11. Laguette MJ, Abrahams Y, Prince S, Collins M. Sequence variants within the 3′-UTR of the COL5A1 gene alters mRNA stability: implications for musculoskeletal soft tissue injuries. *Matrix Biol*. 2011;30:338-345.
12. Connizzo BK, Freedman BR, Fried JH, Sun M, Birk DE, Soslowsky LJ. Regulatory role of collagen V in establishing mechanical properties of tendons and ligaments is tissue dependent. *J Orthop Res*. 2015;33:882-888.
13. Gumucio JP, Sugg KB, Mendias CL. TGF-beta superfamily signaling in muscle and tendon adaptation to resistance exercise. *Exerc Sport Sci Rev*. 2015;43:93-99.
14. Kawakami Y, Rodriguez-Leon J, Izpisua Belmonte JC. The role of TGFbetas and Sox9 during limb chondrogenesis. *Curr Opin Cell Biol*. 2006;18:723-729.
15. Lorda-Diez CI, Montero JA, Martinez-Cue C, Garcia-Porrero JA, Hurle JM. Transforming growth factors beta coordinate cartilage and tendon differentiation in the developing limb mesenchyme. *J Biol Chem*. 2009;284:29988-29996.
16. Hoffmann A, Gross G. Tendon and ligament engineering in the adult organism: mesenchymal stem cells and gene-therapeutic approaches. *Int Orthop*. 2007;31:791-797.
17. Mendias CL, Gumucio JP, Lynch EB. Mechanical loading and TGF-β change the expression of multiple miRNAs in tendon fibroblasts. *J Appl Physiol (1985)*. 2012;113:56-62.
18. Kondo E, Yasuda K, Yamanaka M, Minami A, Tohyama H. Effects of administration of exogenous growth factors on biomechanical properties of the elongation-type anterior cruciate ligament injury with partial laceration. *Am J Sports Med*. 2005;33:188-196.
19. Tekari A, Luginbuehl R, Hofstetter W, Egli RJ. Transforming growth factor beta signaling is essential for the autonomous formation of cartilage-like tissue by expanded chondrocytes. *PLoS One*. 2015;10:e0120857. https://doi.org/10.1371/journal.pone.0120857
20. Ween MP, Oehler MK, Ricciardelli C. Transforming growth factor-beta-induced protein (TGFBI)/(beta ig-H3): a matrix protein with dual functions in ovarian cancer. *Int J Mol Sci*. 2012;13:10461-10477.

21. Posthumus M, Collins M, Cook J, et al. Components of the transforming growth factor-beta family and the pathogenesis of human Achilles tendon pathology--a genetic association study. *Rheumatology (Oxford)*. 2010;49:2090-2097.
22. Chen B, Li B, Qi YJ, Tie K, Chen LB. Association study between growth differentiation factor 5 polymorphism and non-contact anterior cruciate ligament rupture in Chinese Han population. *Int J Clin Exp Med*. 2015;8:22484-22490.
23. Raleigh S, Posthumus M, O'Cuinneagain D, van der Merwe W, Collins M. The GDF5 gene and anterior cruciate ligament rupture. *Int J Sports Med*. 2013;34:364-367.
24. Gibbon A, Saunders CJ, Collins M, Gamieldien J, September AV. Defining the molecular signatures of Achilles tendinopathy and anterior cruciate ligament ruptures: a whole-exome sequencing approach. *PLoS One*. 2018;13:13. https://doi.org/10.1371/journal.pone.0205860
25. Bateman A, Martin MJ, O'Donovan C, et al. UniProt: a hub for protein information. *Nucleic Acids Research*. 2015;43:D204-D212.
26. Vlachos IS, Paraskevopoulou MD, Karagkouni D, et al. DIANA-TarBase v7.0: indexing more than half a million experimentally supported miRNA:mRNA interactions. *Nucleic Acids Research*. 2015;43:D153-D159.
27. Aken BL, Ayling S, Barrell D, et al. The Ensembl gene annotation system. *Database (Oxford)*. 2016;2016:1-19. https://doi.org/10.1093/database/baw093
28. Kearse M, Moir R, Wilson A, et al. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 2012;28:1647-1649.
29. Takahara K, Schwarze U, Imamura Y, et al. Order of intron removal influences multiple splice outcomes, including a two-exon skip, in a COL5A1 acceptor-site mutation that results in abnormal pro-alpha1(V) N-propeptides and Ehlers-Danlos syndrome type I. *Am J Hum Genet*. 2002;71:451-465.
30. Symoens S, Syx D, Malfait F, et al. Comprehensive molecular analysis demonstrates type V collagen mutations in over 90% of patients with classic EDS and allows to refine diagnostic criteria. *Hum Mutat*. 2012;33:1485-1493.
31. R Core Team. 2014. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2014. https://www.R-project.org/
32. Gonzalez JR, Armengol L, Sole X, et al. SNPassoc: an R package to perform whole genome association studies. *Bioinformatics*. 2007;23:644-645.
33. Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA. Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am J Hum Genet*. 2002;70:425-434.
34. Kim JH, Yu SJ, Park BL, et al. TGFBR3 polymorphisms and its haplotypes associated with chronic hepatitis B virus infection and age of hepatocellular carcinoma occurrence. *Dig Dis*. 2011;29:278-283.
35. Hersh CP, Hansel NN, Barnes KC, et al. Transforming growth factor-β receptor-3 is associated with pulmonary emphysema. *Am J Respir Cell Mol Biol*. 2009;41:324-331.
36. Singh KK, Schmidtke J, Keyser B, Arslan-Kirchner M. TGFBR3 variation is not a common cause of Marfan-like syndrome and Loeys-Dietz-like syndrome. *J Negat Results Biomed*. 2012;11:9.
37. Groeneveld ME, Bogunovic N, Musters RJP, et al. Betaglycan (TGFBR3) up-regulation correlates with increased TGF-beta signaling in Marfan patient fibroblasts in vitro. *Cardiovasc Pathol*. 2018;32:44-49.
38. Dixon JR, Selvaraj S, Yue F, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*. 2012;485:376-380.
39. Jovine L, Darie CC, Litscher ES, Wassarman PM. Zona pellucida domain proteins. *Annu Rev Biochem*. 2005;74:83-114.
40. Reinboth B, Thomas J, Hanssen E, Gibson MA. Beta ig-h3 interacts directly with biglycan and decorin, promotes collagen VI aggregation, and participates in ternary complexing with these macromolecules. *J Biol Chem*. 2006;281:7816-7824.
41. Kim JE, Jeong HW, Nam JO, et al. Identification of motifs in the fasciclin domains of the transforming growth factor-beta-induced matrix protein betaig-h3 that interact with the alphavbeta5 integrin. *J Biol Chem*. 2002;277:46159-46165.
42. Son HN, Nam JO, Kim S, Kim IS. Multiple FAS1 domains and the RGD motif of TGFBI act cooperatively to bind αvβ3 integrin, leading to anti-angiogenic and anti-tumor effects. *Biochim Biophys Acta*. 2013;1833:2378-2388.
43. Lopez-Casillas F, Wrana JL, Massague J. Betaglycan presents ligand to the TGF beta signaling receptor. *Cell*. 1993;73:1435-1444.
44. Tazat K, Hector-Greene M, Blobe GC, Henis YI. TβRIII independently binds type I and type II TGF-beta receptors to inhibit TGF-beta signaling. *Mol Biol Cell*. 2015;26:3535-3545.
45. Thapa N, Kang KB, Kim IS. Beta ig-h3 mediates osteoblast adhesion and inhibits differentiation. *Bone*. 2005;36:232-242.
46. Mosher DF, Johansson MW, Gillis ME, Annis DS. Periostin and TGF-beta-induced protein: two peas in a pod? *Crit Rev Biochem Mol Biol*. 2015;50:427-439.
47. Hewett TE, Myer GD, Ford KR. Anterior cruciate ligament injuries in female athletes: part 1, mechanisms and risk factors. *Am J Sports Med*. 2006;34:299-311.
48. Sutton KM, Bullock JM. Anterior cruciate ligament rupture: differences between males and females. *J Am Acad Orthop Sur*. 2013;21:41-50.
49. Renstrom P, Ljungqvist A, Arendt E, et al. Non-contact ACL injuries in female athletes: an International Olympic Committee current concepts statement. *Brit J Sport Med*. 2008;42:394-412.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Laguette M-JN, Barrow K, Firfirey F, et al. Exploring new genetic variants within *COL5A1* intron 4-exon 5 region and TGF-β family with risk of anterior cruciate ligament ruptures. *J Orthop Res*. 2020;38:1856–1865. https://doi.org/10.1002/jor.24585