Whole Genome Sequencing Approach to Identifying Genetic Risk Factors Underlying Anterior Cruciate Ligament Injuries in a Twin Family Study

By:

Daneil Feldmann

Student number: FLDDAN002

Supervisors:

Prof Alison V September

Prof Malcolm Collins

A/Prof Emile Chimusa

Submitted to the University of Cape Town

In fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

University of Cape Town

September 2021

Department of Human Biology, Faculty of Health Sciences
Physiological Sciences
University of Cape Town
Sports Science Institute of South Africa
Boundary Road, Newlands
Cape Town
South Africa

DECLARATION

Doctoral Degrees Board
University of Cape Town
Private Bag X3, Rondebosch
7701 South Africa
Tel: +27 21 650 2202

PHD THESIS TITLE: Whole Genome Sequencing Approach to Identifying Genetic Risk Factors Underlying Anterior Cruciate Ligament Injuries in a Twin Family Study

I, **Daneil Feldmann**, hereby:

(a) Grant the University of Cape Town free licence to reproduce the above thesis in whole or in part, for the purpose of research;

(b) Declare that:

i. The above thesis is my own unaided work, both in concept and execution, and that apart from the normal guidance from my supervisors, I have received no assistance except as stated below:

The bioinformatics analyses of the whole genome sequencing data in Chapter 2 of this thesis, was carried out by A/Prof Emile Chimusa, as mentioned in the Acknowledgements.

ii. Neither the substance nor any part of the above thesis has been submitted in the past, or is being, or is to be submitted for a degree at this University or at any other university, except as stated below: N/A

iii. I have applied and been given permission to include published papers in my thesis in terms of Rule GP6.7: Yes

I am now presenting the thesis for examination for the degree of PhD.

ACKNOWLEDGEMENTS

Financial assistance to conduct this research and thesis was funded by the National Research Foundation, and the University of Cape Town, South Africa.

First and foremost, to my supervisor Prof Alison September, who took me on as a PhD student with no prior background in genetics. Thank you for your unwavering patience, and support in my continual learning in the field. Thank you for always encouraging me to persevere, and strive for excellence in all aspects of my research. Most importantly, thank you for inspiring me with your passion and extensive knowledge in the field.

To Prof Malcolm Collins, who played an instrumental role in building my critical thinking skills and approaching my work with an acute lens. Thank you for your invaluable input throughout the writing up of this thesis.

To A/Prof Emile Chimusa, who helped tremendously with the bioinformatics analyses. Thank you for your guidance and assistance in developing a methodology, to bring this research to the high standard presented in this thesis.

My sincerest thanks to Dr Masouda Rahim, for being an incomparable mentor in the beginning stages of my PhD. For 'showing me the ropes' and taking me under your wing. Your support, guidance and kindness were most appreciated. To Dr Nancy Laguette, and Dr Sharon Mkumbuzi. Two outstanding researchers, mentors and most importantly friends. Thank you both for always being available for the smallest of queries, and taking on every request with such interest and professionalism. You both are inspirational women, whom I greatly respect.

Thank you to Dr Jon Patricios in your assistance with the recruitment of participants for the ACL twin study. To our collaborators in Poland, Sweden and Australia. Thank you for providing us with the DNA samples to conduct this research. I would also like to thank all the participants of the study, who gave of their time and contributed their DNA to this research.

To the administrative staff: Ayesha, and Lesa for your friendly smiles, and always being willing to help. To Trevino Larry and Neezaam Karriem, for your support in the laboratory and all technical matters. To my research group, thank you for creating a space to learn and grow, and of course to share many a laugh. An extra special thank you to Sam Beckley, for being my office bestie, and never being too busy to listen to my random ramblings.

Thank you to the examiners for kindly taking the time to review and examine this thesis.

Lastly, to my loving family and Davide. Thank you for always supporting my endeavours, even when not fully comprehendible. For believing in my abilities, when I lacked confidence, and always being ready to celebrate my achievements (even the smallest ones) during my PhD journey.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

A	Adenine
ACAN	Aggrecan
ACL	Anterior cruciate ligament
ACTR5	Actin Related Protein 5
ADIG	Adipogenin
AEBP2	AE Binding Protein 2
ANOVA	Analysis of variance
ARSI	Arylsulfatase Family Member I
AT	Achilles tendon
ATG7	Autophagy Related 7
BGN	Biglycan
BMI	Body mass index
bp	Base pairs
BPI	Bactericidal Permeability Increasing Protein
C	Cytosine
CASP8	Caspase 8
CEP57L1A	Centrosomal Protein 57 Like 1
CI	Confidence interval
cm	Centimetre
COL1A1	Collagen Type I α1 Chain
COL3A1	Collagen Type III α1 Chain
COL5A1	Collagen Type V α1 Chain
COL11A1	Collagen Type XI α1 Chain
COL11A2	Collagen Type XI α2 Chain

COL12A1	Collagen Type XII α1 Chain
COL27A1	Collagen Type XXVII α1 Chain
CNV	Copy number variation
CON	Control
CS	Chondroitin sulfate
DAMP	Damage-associated molecular pattern
DCN	Decorin
DHX35	DEAH-Box Helicase 35
DIRC2	Disrupted in renal carcinoma 2
DNA	Deoxyribonucleic acid
DOCK2	Dedicator Of Cytokinesis 2
DPP6	Dipeptidyl Peptidase Like 6
DS	Dermatan sulfate
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
EPB41	Erythrocyte Membrane Protein Band 4.1
FACIT	Fibril associated collagen with interrupted helices
FAH	Fumarylacetoacetate Hydrolase
FAM83D	Family With Sequence Similarity 83 Member D
FDFT1	Farnesyl-Diphosphate Farnesyltransferase 1
FNB1	Fibrillin-1
FBN2	Fibrillin-2
FDR	False discovery rate
FGF	Fibroblast growth facto
G	Guanine

GAG	Glycosaminoglycan
GDF5	Growth differentiation factor 5
HA	Hyaluronan
HWE	Hardy-Weinberg Equilibrium
IL	Interleukin
IL1B	Interleukin-1β
IL1RN	Interleukin-1 receptor antagonist
IL6	Interleukin-6
IL6R	Interleukin-6 receptor
INHBA	Inhibin Subunit Beta A
Indels	Insertions and deletions
kb	Kilobase
KDR	Kinase insert-domain receptor
kg	Kilogram
KIF26A	Kinesin Family Member 26A
KS	Keratan sulfate
LBP	Lipopolysaccharide Binding Protein
LD	Linkage disequilibrium
LncRNA	Long non-coding RNA
LincRNA	Long non-coding RNAs longer than 200 nucleotides
LINC01250	Long Intergenic Non-Protein Coding RNA 1250
LP	Link protein
LRR	Leucine-rich repeats
LUM	Lumican
MIR608	MicroRNA 608

miRNA	Micro ribonucleic acid
ml	Millilitre
MMP	Matrix metalloproteinase
MMP1	Matrix metalloproteinase 1
MMP2	Matrix metalloproteinase 2
MMP3	Matrix metalloproteinase 3
MMP7	Matrix metalloproteinase 7
MMP8	Matrix metalloproteinase 8
MMP9	Matrix metalloproteinase 9
MMP10	Matrix metalloproteinase 10
MMP12	Matrix metalloproteinase 12
MRI	Magnetic resonance imaging
mRNA	Messenger ribonucleic acid
NCBI	National Centre for Biotechnology Information
NGS	Next generation sequencing
OLFML2B	Olfactomedin Like 2B
OR	Odds ratio
p	Short arm of chromosome or Probability
PCA	Principal components analysis
PCL	Posterior cruciate ligament
PCR	Polymerase chain reaction
POMC	Proopiomelanocortin
PPP1R16B	Protein Phosphatase 1 Regulatory Subunit 16B
PRP	Platelet-rich plasma
q	Long arm of chromosome

RALGAPB	Ral GTPase Activating Protein Non-Catalytic Subunit Beta
RFLP	Restriction fragment length polymorphism
RNF152	Ring finger protein 152
ROR2	Receptor Tyrosine Kinase Like Orphan Receptor 2
rSNPs	Regulatory polymorphisms
RNA	Ribonucleic acid
SA	South Africa
SEMA5B	Semaphorin 5B
SerpinA11	Serpin Family A Member 11
SLC32A1	Solute Carrier Family 32 Member 1
SLRP	Small leucine-rich proteoglycan
SMARCD1	SWI/SNF Related, Matrix Associated, Actin Dependent Regulator Of Chromatin, Subfamily D, Member 1
SNP	Single nucleotide polymorphism
SORCS2	Sortilin related VPS10 domain containing receptor 2
T	Thymine
TGFβ	Transforming growth factor beta
TIMP	Tissue inhibitors of metalloproteinases
TIMP1	Metalloproteinase inhibitor 1
TIMP2	Metalloproteinase inhibitor 2
TIMP3	Metalloproteinase inhibitor 3
TIMP4	Metalloproteinase inhibitor 4
TNC	Tenascin-C
µl	Microlitre
USA	United States of America

UTR	Untranslated region
VEGFA	Vascular endothelial growth factor A isoform
VNTR	Variable nucleotide tandem repeat
WES	Whole exome sequencing
WDR6	WD Repeat Domain 6
WGS	Whole genome sequencing
ZDHHC23	Zinc Finger DHHC-Type Palmitoyltransferase 23

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Background: Predisposition to ACL rupture is multifactorial, resulting from a complex interplay of intrinsic and extrinsic risk factors. Variation in the genome is now considered a key intrinsic risk factor, but the majority of currently implicated loci have been identified through case-control genetic association studies, which are limited by a candidate gene approach and insufficient statistical power. The primary aim of this thesis was to use a whole genome sequencing (WGS) approach within the context of a twin family study to identify novel or previously implicated genetic loci contributing to ACL rupture predisposition (Chapter 2). Additionally, this research aimed to explore prioritised genetic polymorphisms previously associated with ACL rupture and functioning in key biological pathways implicated through the WGS analyses, independently and as a collective, with ACL rupture predisposition in a large combined ACL rupture dataset (Chapter 3 and 4).

Methods: The complete genomes of all family members in two unrelated families, each with affected twins were sequenced. Variants with potential loss of function effect were prioritised, and explored for probable biological function in the ACL rupture risk pathway. Furthermore, identity by descent analysis (IBD) was performed to identify potential disease causing mutations, on chromosomal regions shared between family members, and across families. Enriched biological pathway analyses were further explored to prioritise potential candidate genes. Two biological networks were prioritised which highlighted the angiogenesis and proteoglycan family of proteins. Specific polymorphisms within previously investigated candidate genes were further explored in case-control genetic association studies conducted in a large collective data set, including participants from three independent (Sweden, Poland and Australia) cohorts, combined with previously published South African and Polish data. The anterior cruciate ligament (ACL) rupture group included individuals diagnosed with a clinical

diagnosis of an ACL rupture based on physical examination, and confirmed by either magnetic resonance imaging or arthroscopy. Only ACL ruptures resulting from a non-contact mechanism of injury were included. The control group comprised individuals of similar age to cases with no prior history of ACL injury or other ligament and tendon injuries, and participating in regular sporting activity, which was similar to cases. Participant samples were genotyped for single nucleotide polymorphisms in the *VEGFA* (rs699947 C/A rs1570360 G/A, rs2010963 G/C) and *KDR* (rs2071559 A/G, rs1870377 T/A) genes (Sweden CON: 116 ACL: 95; Poland CON: 149 ACL: 127 and Australia CON: 83 ACL: 342). Additionally, in the *ACAN* (rs2351491 C/T, rs1042631 T/C, rs1516797 T/G), *DCN* (rs516115 T/C) and *BGN* (rs1126499 C/T, rs1042103 G/A) genes (Sweden and Poland). Haplotype analyses were explored (*VEGFA*, *KDR*, *ACAN* and *BGN*) using the individual genotype data. In addition, inferred allele interactions were presented for *VEGFA-KDR*, *ACAN-BGN ACAN-DCN*, *BGN-DCN*, and *VEGFA-DCN* as a proxy for gene-gene interactions within the discrete angiogenesis and proteoglycan gene families, and between genes as a proxy for pathway interactions. For association studies, frequencies were calculated for the genotype, allele, inferred haplotypes and allele interactions, and the distributions compared between the control and ACL rupture participants. The statistical programs in R were used for all the analyses, and a p value < 0.05 was accepted to be significant.

Results: The WGS analyses highlighted six candidate genetic loci in three genes (*COL12A1*, *CATSPER2*, and *KCNJ12*) with predicted loss of function effects in all affected and unaffected family members within the two studied families. Of the three genes, polymorphisms within *COL12A1* were previously associated with ACL rupture predisposition, while *CATSPER2* and *KCNJ12* are two novel genetic loci with no known previous association with predisposition to ACL rupture. The IBD analyses identified several regions shared in each independent family, of which a segment including a long intergenic non-protein coding RNA (lincRNA)

LINC01250 gene in the telomeric region of chromosome 2p25.3 was shared between affected twins in both families, and an affected brother. Furthermore, several functional partners were highlighted. Genetic association analyses of the prioritised polymorphisms in a combined cohort identified an independent association of the *VEGFA* rs2010963 CC genotype and C allele with increased risk (genotype p = 0.0001, FDR p = 0.001, OR 2.16, 95% CI: 1.47-3.19; allele p = 0.0006, FDR p = 0.003, OR 1.29, 95% CI: 1.11-1.49). Furthermore, the association of the *VEGFA* A-A-G and A-G-G inferred haplotypes (rs699947 A/C-rs1570360 G/A-rs2010963 G/C) with reduced risk (p = 0.010, haplo.score: -2.58, OR: 0.85, 95% CI: 0.69-1.05; A-G-G: p = 0.036, haplo.score: -2.09, OR: 0.81, 95% CI: 0.64-1.02) of ACL rupture. Moreover, a reduced interval (rs1570360 G/A-rs2010963 G/C) revealed an association of the *VEGFA* -G-G and -A-G inferred haplotypes with reduced risk (-G-G: p = 0.031, haplo.score: -2.15, OR: 1.00 and -A-G: p = 0.024, haplo.score: -2.25, OR: 0.98, 95% CI: 0.82-1.18) and the -G-C inferred haplotype with increased risk p = 0.012, haplo.score: 2.50, OR: 1.18, 95% CI: 0.99-1.40). The *KDR* genotype and haplotype analyses illustrated that it is highly unlikely that the investigated *KDR* polymorphisms are associated with modulating ACL rupture risk. Inferred allele interactions noted a significant association of the *VEGFA* (rs699947 A/C, rs2010963 G/C) - *KDR* (rs2071559 A/G, rs1870377 T/A) A-G-A-A (p = 0.005, OR: 0.51, 95% CI: 0.30-0.87) and A-G-G-A (p = 0.018, OR: 0.93, 95% CI: 0.54-1.60) combinations with reduced ACL rupture risk. Further, a significant association of the *VEGFA* (rs699947 C/A, rs1570360 G/A, rs20109630 G/C) - *DCN* (rs516115 T/C) A-G-G-T (p = 0.010, OR: 0.53, 95% CI: 0.30-0.91), A-A-G-C (p = 0.010, OR: 0.42, 95% CI: 0.21-0.81) and A-A-G-T (p = 0.046, OR: 0.77, CI: 0.49-1.2) allele combinations with reduced risk was noted for male participants in the collective cohort. No independent or haplotype associations with ACL rupture risk were noted for any of the investigated proteoglycan polymorphisms, in the collective cohort.

Conclusion: Collectively, this work has expanded current knowledge on the genetic regions contributing to ACL rupture predisposition, and further highlights the polygenic nature of multifactorial phenotypes. Employing whole genome sequencing in a twin family context, together with a pathway based approach, novel and previously implicated genetic loci were identified towards the aims of the thesis. The catalogue of candidate *in silico* mutations and modifier genes that clustered in pathophysiological pathways important in ACL rupture, and with implications for therapeutic intervention were identified, and need to be interrogated. Of particular interest are the novel *CATSPER2, KCNJ12* and *LINC01250* genetic loci. Furthermore, additional evidence to support the implication of the *VEGFA* gene in modulating ACL rupture risk is provided, and highlighted is the potential collaboration of members within the angiogenesis and proteoglycan gene family in modulating risk. The studies in Chapter 3 and 4 suggest genetic association studies in single populations are less informative, and instead larger collective cohorts with increased statistical power should be employed. Further to that, rather than investigating single polymorphisms, larger regions of the genome should be explored to determine the potential interacting components contributing to musculoskeletal injury risk. Going forward, characterisation of the functional biological effect of implicated loci may assist in unravelling the underlying mechanisms altering tissue homeostasis, and subsequently an individual's capacity for healing and adaptive response.

CHAPTER 1

LITERATURE REVIEW

1.1 INTRODUCTION AND SCOPE OF THE THESIS

Globally, the burden of non-communicable diseases such as coronary heart disease, type 2 diabetes and cancer has escalated (Wang and Wang, 2020). Along with other lifestyle factors, physical inactivity is a major risk factor contributing to this increased prevalence of these health conditions (Lee et al., 2012). In an effort to mitigate the risk, an increase in exercise participation among sedentary populations has followed, and in parallel an increased incidence of musculoskeletal injuries (Safiri et al., 2020).

Rupture of the anterior cruciate ligament (ACL) is one of the most common debilitating musculoskeletal injuries affecting the lower limb (Hewett et al., 2010; Kaynak et al., 2017; Kiapour and Murray, 2014). The genetic contribution to ACL rupture is the focus of the current thesis, and the phrase "ACL injury" is used interchanably with rupture in the thesis. A review of the literature follows a narrative approach, where **Chapter 1** provides a brief overview on ACL rupture, with information on the gross anatomy and physiology of the ligament (**Section 1.3**), the microstructure (**Section 1.4**), ligament healing and remodelling (**Section 1.5)** an overview of the epidemiology of ACL rupture (**Section 1.6**) and the associated risk factors (**Section 1.7**). As yet, the exact cause and mechanism of ACL rupture is unclear, however growing evidence points to a genetic contribution (**Section 1.8**) with the current implicated genetic loci summarised in **Section 1.9**. Although few studies have utilised a hypothesis-free approach (Baird et al., 2014; Baker et al., 2017; Baker et al., 2018; Caso et al., 2016; Gibbon

et al., 2018; Kim et al., 2017, Kim et al., 2021) using next generation sequencing technologies discussed in **Section 1.11**. Our current understanding of the genes potentially involved in the pathogenesis of ACL rupture have predominately been identified through candidate gene studies. While providing a stepping-stone towards the elucidation of potential underlying mechanisms, candidate gene studies are limited by a hypothesis driven approach, where prior knowledge about the injury phenotype and affected tissue are required. For this reason, the research presented in this thesis harnessed an unbiased, hypothesis-free approach through whole genome sequencing technologies. To build on previous research surrounding the genetic contribution to ACL rupture mechanisms.

In summary, the main aim of this thesis was to employ whole genome sequencing (WGS) technologies in a twin family study, to identify novel or previously implicated genetic polymorphisms underpinning predisposition to ACL rupture (**Chapter 2**). Additionally, the second aim was to further explore functional partners highlighted through WGS analyses, that have previously been implicated in ACL rupture predisposition, in large combined datasets (**Chapters 3 and 4**). The summary and discussion of the main findings of the thesis are presented in **Chapter 5**, alongside the strengths, limitations, and future direction of research.

1.2 CONTRIBUTION TO THE FIELD

1.2.1 Manuscripts

The following manuscript was published in the Journal of Orthopaedic Research. *Investigation of multiple populations highlight VEGFA polymorphisms to modulate anterior cruciate ligament injury*. 2021. Feldmann DC, Rahim M, Suijkerbuijk MAM, Laguette MJ, Cieszczyk P, Ficek K, Huminska-Lisowska K, Häger CK, Stattin E, Nilsson KG, Alvarez-Rumero J, Eynon N, Feller J, Tirosh O, Posthumus M, Chimusa ER, Collins M, September AV.

The following manuscript is currently in peer review (PLoS ONE). *A Whole Genome Sequencing Approach to Anterior Cruciate Ligament Rupture – A Family "Twin" Study*. Feldmann DC, Bope CD, Patricios J, Chimusa ER, Collins M, September AV.

1.2.2 International Conference Presentations

The following abstract title was presented at the 24th Annual Congress of the ECSS (European College of Sports Science) hosted by Charles University in Prague, Czech Republic. *Investigation of three independent populations strengthens the hypothesis that genetic loci within the proteoglycan and angiogenesis associated pathways predispose to anterior cruciate ligament injury*. Feldmann DC, Rahim M, Suijkerbuijk MAM, Cieszczyk P, Ficek K, Huminska-Lisowska K, Häger CK, Stattin E, Nilsson KG, Posthumus M, Collins M, September AV.

1.3 GROSS ANATOMY OF THE ACL

The ACL is located in the knee where is connects the upper and lower limb, together with the posterior cruciate ligament, the medial collateral ligament, and the lateral collateral ligament (**Figure 1.1**). Comprised of two bundles, the anteromedial bundle of the ACL is tight in knee flexion, and the posterolateral bundle convex and tight in knee extension (Kweon et al., 2015). Anatomically, the ACL originates from the fossa of the medial surface of the intercondylar notch of the lateral femoral condyle, and the fibres run obliquely to insert medially onto the intertubercular ridge, between the medial and lateral tibial spines (Mall et al., 2013; Petersen and Tillmann, 2002; Zantop et al., 2006). The ACL has an average length of 38 mm (range of 31 - 38 mm) a width of 11 mm (10 - 12 mm) and an ultimate tensile load of 2160N (Kweon et al., 2015; Lord et al., 2015; Mall et al., 2013). Functionally, the cruciate ligaments are the major stabilisers of the knee joint (Imran and O'Connor, 1998) and the primary function of the ACL is to prevent anterior translation of the tibia relative to the femur (Imran and O'Connor, 1998; Kiapour and Murray, 2014; Noyes, 2009). Secondary functions of the ACL include resisting tibial rotation, medio-lateral movement in the extended knee (Imran and O'Connor, 1998), varus, and valgus stresses (Liu-Ambrose, 2003). In addition to its role in resisting mechanical forces, other functions of the ACL include maintaining joint homeostasis and joint proprioception (Frank, 2004).

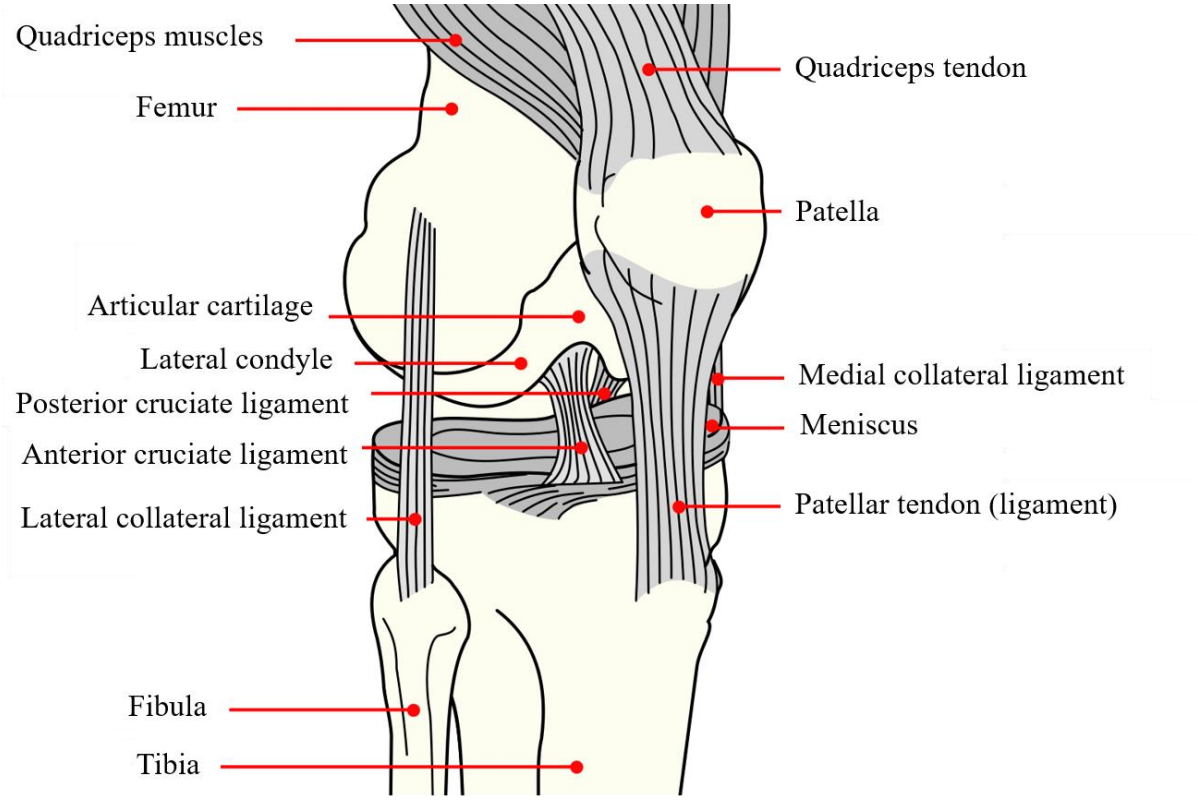


Figure 1.1: Basic anatomy of the knee joint displaying the anterior cruciate ligament (ACL) and the surrounding structures. The femur and tibia, as well as the posterior cruciate ligament, medial collateral ligament and lateral collateral ligament are shown. *Wikimedia Commons, the free media repository*. Retrieved 13:15, April 8, 2021 from [commons.wikimedia.org].

1.3.1 Innervation

The posterior articular branches of the tibial nerve innervate the ACL, with most of the nerve fibres occuring within the ligament vasculature, thus providing a vasomotor function (Kennedy et al., 1982). Receptors of the nerve fibres include Ruffini corpuscles, Vater-Pacini corpuscles, Golgi-like tension receptors, and free-nerve endings (Haus and Halata, 1990). Mechanoreceptors (Ruffini, Pacini, and Golgi-like receptors) collectively monitor proprioceptive information, and provide signalling for knee postural changes (Duthon et al., 2006). Whereas free-nerve endings, function as nociceptors along with releasing neuropeptides with vasoactive functions, suggesting a possible role in maintaining tissue homeostasis or in the remodelling of grafts (Haus and Halata, 1990).

1.3.2 Vascularisation

Blood supply to the cruciate ligaments is via the middle genicular artery. While the overlying connective tissue sheath enclosing the entire ligament contains blood vessels, very few actually penetrate the midsubstance of the ACL (Bray et al., 2002; Frank, 2004). Thus rendering the ACL relatively hypovascular in nature (Everhart et al., 2017) which is thought to contribute to its limited healing capacity post injury (Bray et al., 2002; Duthon et al., 2006) and inferior post-surgical biomechanical properties, in comparison to the native tissue (Gurlek et al., 2017). Furthermore, the distribution of blood vessels along the ACL is not homogenous, with the proximal region displaying greater vascularity compared to the distal region of the ligament (Duthon et al., 2006).

1.4 MICROSTRUCTURE OF THE ACL

The ACL is distinguished by three zones (proximal, middle, and distal portion). The proximal region is less solid, and highly cellular with round and ovoid cells, whereas the middle portion consists of spindle-shaped and fusiform fibroblasts, with a high density of collagen bundles. The distal region is the most solid region of the ACL, and it contains ovoid fibroblasts, chrondroblasts and a low density of collagen bundles (Duthon et al., 2006). Structurally, ligaments and tendons are organised in a hierarchical manner (Hoffmann and Gross, 2007) where collagen units are assembled into microfibrils, followed by fibrils, fibre bundles and fascicles (**Figure 1.2**). Finally, the entire ligament is enclosed in the epiligament (connective tissue sheath) characterised by blood vessels, lymphatics and nerves (Frank, 2004). This hierarchical structure ensures ligaments are flexible under tension, and is further aided by the "crimp" wave patterning of the fibres, which acts as a shock absorber, and allows for longitudinal elongation of the tissue without fibrous damage (Duthon et al., 2006).

Biochemically, ligaments are comprised of approximately two-thirds water and one-third solid components. The ACL is composed predominantly of type I collagen, made-up of heterotrimers of two α1 and one α2 chains. Type I collagen fibres are responsible for the hierarchical structure of ligament tissue, they run parallel to each other and contain cross-links which provide strength against forces acting perpendicular to the fibres (Hoffmann and Gross, 2007). The remainder of the ACL is comprised of types III, VI, V, XI and XIV collagens, proteoglycans, elastin, glycosaminoglycans and various glycoproteins, of which all are synthesised and secreted by ligament fibroblasts (Frank, 2004).

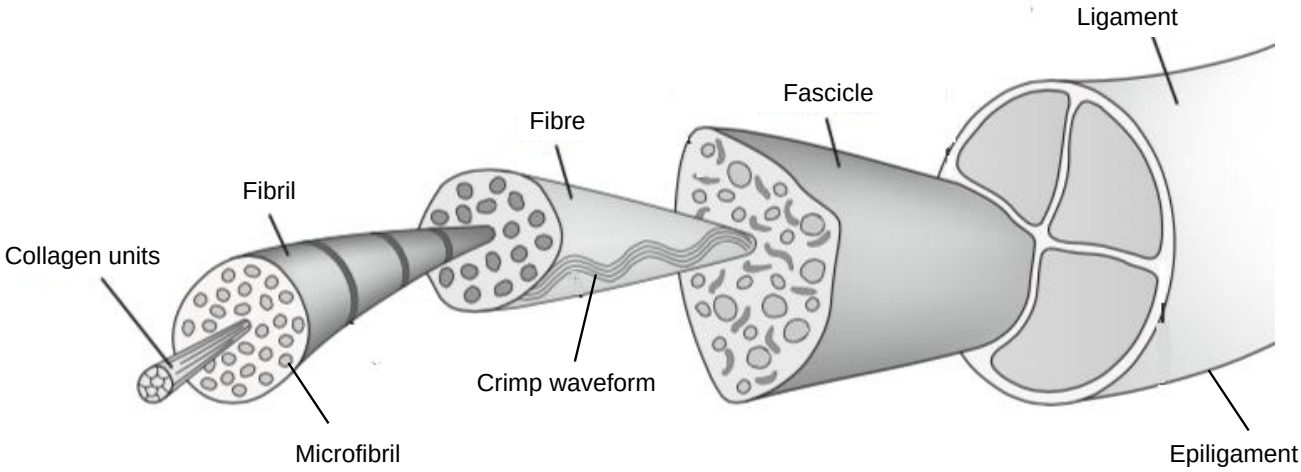


Figure 1.2: The hierarchical structure of ligaments. Adapted from Wikimedia Commons, the free media repository. Retrieved 16:30, June 14, 2021 from [commons.wikimedia.org].

1.5 LIGAMENT HEALING AND REMODELLING

Over time, chronic exposure to mechanical load results in an increase in ligament mass, stiffness and load to failure (Frank et al., 1999). However, when overloaded or exposed to tensions greater than the ligament can sustain, resultant tissue failure and subsequent partial or a complete ligament tear occurs. Post injury, ligament remodelling involves a sequence of overlapping, but distinct events categorised by three consecutive stages: (1) the acute inflammatory stage, (2) the proliferative or regenerative stage, and (3) the tissue remodelling stage.

The acute inflammatory stage starts immediately after an injury has occurred, and continues for 48 to 72 hours. During this stage, blood gathers at the injury site, platelet cells interact with matrix components to initiate clotting, and the secretion of growth factors specific to the inflammatory pathway initiates the wound healing process. Some of these growth factors include platelet-derived growth factor (PDGF), transforming growth factor β (TGFβ), vascular endothelial growth factor A (VEGFA) and fibroblast growth factor (FGF). PDGF and TGFβ attract immune system cells to the injury site and stimulate proliferation, whereas VEGFA promotes angiogenesis, and FGF promotes the growth of cells in collagen formation. The proliferative/repair stage occurs when immune cells secrete growth factors and cytokines that initiate fibroblast proliferation for extracellular matrix remodelling. The initial tissue appears as scar tissue, which is disorganised by more blood vessels, fat cells, fibroblasts and inflammatory cells (Frank, 2004; Shrive et al., 1995). Over the next few weeks, fibroblast cells deposit various types of collagen, proteoglycans, glycoproteins and other matrix proteins. Initial collagen fibrils laid down are smaller in diameter than the native ligament tissue. Following the proliferative stage, remodelling begins. The remodelling stage can extend from

months to years post injury, and is characterised by collagen maturation and changes to the extracellular matrix (ECM) which begins to resemble the native ligament tissue. Despite the long duration remodelling phase of the tissue, long-term differences in matrix structure and function remain, where even fully remodelled scar tissue is grossly, microscopically and functionally different to the native tissue (Frank et al., 1999). Regarding ACL ligament reconstruction, effective healing and remodelling is paramount to ensuring adequate ligamentization of a tendon graft. And it is the fine tuning of key growth factors such as VEGFA, which play an important role in ensuring that effective remodelling of the tendon graft after ACL reconstruction takes place (Yoshikawa et al., 2006).

1.6 EPIDEMIOLOGY OF ACL RUPTURES

1.6.1 Mechanisms of Injury

Rupture of the ACL occurs as a result of excessive torsional or translational trauma to the knee joint, which involves a change in velocity, or the production of a multidirectional force across the joint while weight bearing (Wetters et al., 2016). This results in a loss of mechanical stabilization and kinaesthetic acuity, which in turn affects knee joint control, function, and balance (Ageberg, 2002).

ACL ruptures occur primarily through contact, or non-contact mechanisms (Hewett et al., 2007). Rupture of the ACL through contact mechanisms, occur due to a direct blow to the knee by another person or object, and are most often accompanied by concomitant knee pathology due to the higher-energy force associated with the direct interaction. However, the majority (~70%) of all ACL ruptures occur through non-contact mechanisms (Pfeifer et al., 2018) and

differ from contact mechanisms in that the forces resulting in injury do not come from an external source, but are generated within the individual (Marshall et al., 2007). An additional mechanism of injury, completely different to that of contact or non-contact injuries occurs in skiing (Bere et al., 2011). Termed the 'phantom foot' in recreational skiers, the ski acts as a lever to bend or twist the knee. In professional skiers, an alternate mechanism called the 'slip-catch' accounts for the majority of ACL ruptures sustained (Shea et al., 2014).

1.6.2 Incidence of ACL Rupture

The ACL is one of the most commonly injured ligaments in the knee, with the highest incidence of cases occuring in sporting populations (Johnson, 1983). Sports characterised by pivoting movements, jumping, rapid deceleration during cutting, and sudden changes in direction, such as soccer, rugby, american football, australian football, netball, basketball and hockey, are associated with the highest risk of ACL rupture (Murray, 2009). Moreover, several studies have shown female athletes are at an increased risk of ACL rupture (Beynnon et al., 2014; Joseph et al., 2013). The mechanisms underlying the noted increased risk are currently unknown, however are likely due to multifactorial factors including both intrinsic and extrinsic components (Kaeding et al., 2017) discussed in the next section. However, although males are at a lower risk in comparison to females, due to a greater number of males participating in sport, and higher exposure rates, male athletes account for the majority of documented ACL rupture cases (Gianotti et al., 2009; Griffin et al., 2000).

Epidemiological data describing the incidence rate of ACL ruptures in the South African population has not been published. However, epidemiological data from the United States indicates that between 75 000 to 250 000 new ACL ruptures occur every year (Brophy et al.,

2009; Gornitzky et al., 2016; Gottlob and Baker, 2000; Mall et al., 2014) with more than 175 000 of these undergoing reconstruction (Gottlob and Baker, 2000). For ACL reconstructions occurring between 2005 and 2013 in the US, the total immediate procedure cost for the 229,446 ACL surgeries identified in the database was $2,622,928,663.00, with an average cost of $11,431.57 per procedure (Herzog et al., 2017). In addition to the cost of surgery, the rehabilitation of an injured or ruptured ACL is also a costly and lengthy process, where time out of play in an athletic population can be up to 255 days (Dallalana et al., 2007). And regardless of surgical intervention, the ACL never fully regains pre-injury functionality (Murray, 2009) lending to long term psychological effects, disability, pain and discomfort, reduced knee function, and a 10-fold greater risk of osteoarthritis of the knee (Kiapour and Murray, 2014; Ruiz et al., 2002). Further to that, injury to the ACL rarely occurs in isolation, and is often accompanied by meniscal tears, medial collateral ligament injuries, and chondral lesions (Gianotti et al., 2009; Griffin et al., 2000; Kiapour and Murray, 2014).

Given the incidence, costs and debilitating long-term implications of ACL rupture, prevention efforts rather than rehabilitation is key to reducing the public burden of disease. In this regard, prevention begins with an understanding of the underlying mechanisms of injury, and the associated modifiable and non-modifiable risk factors.

1.7 RISK FACTORS

Multiple extrinsic and intrinsic risk factors have been implicated in the aetiology of ACL rupture, yet the exact cause is currently incompletely understood (Griffin et al., 2000; Meeuwisse, 1994). Extrinsic risk factors include those that are externally controlled, and include type of sport, nature and intensity of physical activity, playing surface, footwear, and the weather conditions. Although summarised in **Figure 1.3**, further detail on these external risk factors is beyond the scope of the current thesis, and can be reviewed in (Pfeifer et al., 2018). On the other hand, intrinsic risk factors are inherent to the individual, and include individual-specific features such as anatomical structure, sex and hormonal constitution, neuromuscular control, biomechanics, physiological make-up, and as previously mentioned genetic variants (Pfeifer et al., 2018). Intrinsic risk factors are further categorised as modifiable (can be altered) and non-modifiable (innate) and will be briefly discussed in the following section.

The available research investigating the relationship between the anatomy of the knee, and ACL rupture risk is currently disputed. Structural components including ACL morphology, tibial and femoral surface geometry, the structural alignment of the lower extremities, and knee-joint laxity have all been explored (Shultz et al., 2012). Morphologically, a smaller ACL (area and volume) is associated with an increased risk of ACL rupture (Chaudhari et al., 2009; Wang et al., 2020). Further to that, a greater tibial slope (the angle between the line perpendicular to the tibial axis, and the posterior inclination of the tibial plateau) is suggested to increase the risk of ACL rupture by some authors (Alentorn-Geli et al., 2014; Rahnemai-Azar et al., 2016). Whereas other have not observed this effect of tibial slope on ACL rupture risk (Blanke et al., 2016). It appears the difficulty in establishing a consensus

between studies stems from the lack of agreement in the degree of slope, and the anatomical location (medial, lateral or posterior) of the tibial slope that incurs a greater risk of injury (Wordeman et al., 2012). Like the contradicting evidence surrounding tibial slope and injury risk, likewise the current evidence proposing a correlation between intercondylar notch width and ACL rupture is conflicted. Some authors suggest a narrow notch results in the impingement of the ACL at the anterior and posterior roof of the notch (Keays et al., 2016; LaPrade and Burnett, 1994; Souryal and Freeman, 1993). Whereas, others do not agree (Shelbourne et al., 1998; Teitz et al., 1997).

Generalised joint hypermobility refers to the hyperextensibility of the synovial joints, where the ability to extend passively or actively past the normal physiological range of motion occurs. A recent review by Sundemo et al. (2019) suggests the current literature indicates that greater general, anterior and internal rotation knee laxity may indeed contribute to an increased risk of ACL rupture (Sundemo et al., 2019). Physiologically, the biomechanics of agonist and antagonist muscle groups in the lower limb have also been identified as indicators of ACL rupture potential. The indicating factor contributing to risk is the ratio of relative hamstring to quadriceps muscle (H/Q) strength. The dominance of the quadriceps group to stabilise the knee increases anterior shear stress on the tibia, and in turn ACL loading. Thus placing the ACL under greater stress (Hewett et al., 2010). Currently, it is found that a lower H/Q ratio is predictive of an increased risk of ACL rupture (Hewett et al., 2010; Letafatkar et al., 2015; Söderman et al., 2001).

As previously mentioned in **Section 1.6.2**, females are at a greater risk of ACL rupture in comparison to their male counterparts. However, the exact molecular mechanisms and risk factors explaining the observed sex discrepancy are not yet fully understood (Balachandar et

al., 2017; Smith et al., 2012). Some of the proposed risk factors include differences in anatomy and physiology. For example, females athletes display different H/Q ratios (Myer et al., 2004) and movement patterns to male athletes, largely influenced by differences in anatomical structure and neuromuscular control (Shultz et al., 2012; Smith et al., 2012). Additionally, the ACL contains oestrogen and progesterone receptor sites (Balachandar et al., 2017) and thus changes in sex hormone concentrations over the menstrual cycle may potentially play a role in the metabolism, composition and biomechanical properties of ligament (Shultz et al., 2012) and therefore influence ACL tissue characteristics, and potential response to loading.

Traditionally, research has adopted a simplistic view in injury prediction models, by which each component is observed independently, with the assumption of a linear behaviour in injury causality (Bittencourt et al., 2016). To some extent, the Meeuwise model (**Figure 1.3**) addressed this limitation, by adopting a multifactorial approach to investigating the nature of injury risk (Meeuwisse, 1994). The model describes how a complex interplay between extrinsic and intrinsic risk factors, lends to an increase in the risk of injury. That is, an athlete becomes predisposed as a result of the numerous intrinsic risk factors acting internally, however alone these factors are generally insufficiently causative. It is only then with further exposure of a predisposed athlete to extrinsic risk factors, that increased susceptibility to injury occurs. Finally, it is with a particular inciting event, that a susceptible individual may suffer a resultant injury (Meeuwisse, 1994). Despite this understanding, because of the sequence of actions occurring prior to an injury, the inciting event typically becomes the focus. However, truly, it is the numerous intrinsic risk factors, which are distant from the outcome, that require further exploration (Meeuwisse, 1994).

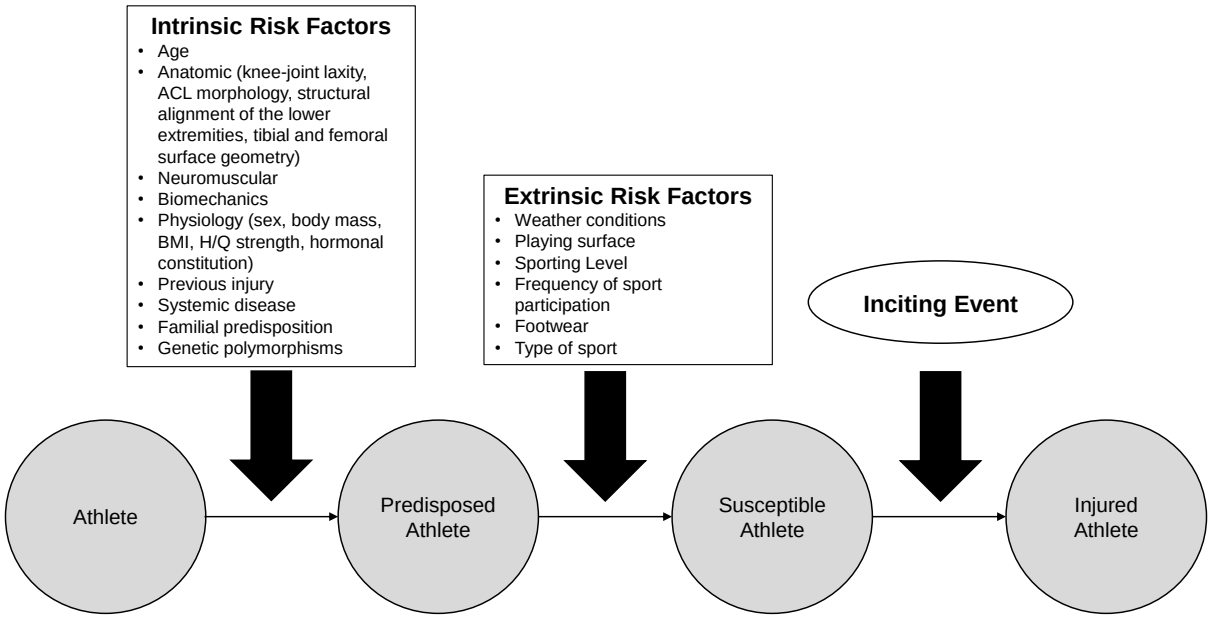


Figure 1.3: The complex interaction between extrinsic and intrinsic risk factors increasing an individual's risk of injury, as well as the inciting event that results in an injury. Figure adapted from Meeuwisse (1994) and risk factors sourced from Smith et al., (2012) and Pfeifer et al. (2018).

Although addressing several previous limitations, the Meeuwise model described is insufficient in acknowledging the complex interaction occurring between injury risk components (Bittencourt et al., 2016). The interplay between intrinsic and extrinsic risk factors resulting in injury, does not occur through the summation of these isolated risk factors in a linear fashion, but rather from the interaction of the participating factors, which may be modified by the occurring interactions, or alternatively by additional unpredictable factors that may emerge (Bittencourt et al., 2016). Aptly coined the 'web of determinants' (Philippe and Mansi, 1998) the complex systems model (**Figure 1.4**) suggests these factors interact with each other in a manner that is non-linear or expected, and small changes in a few determinants can result in unexpected consequences (Coffey, 1998). When considering sports injury risk, the risk determinants interact to result in an altered outcome, which may differ to that observed in an individual factor acting alone. An athlete should therefore be considered as a complex system, where interacting units of the system result in a web of determinants. These complex interactions occur from the system history, and from observable regularities (risk or protective profile). The ensuing injury or adaptive response (the emerged pattern) will then further provide feedback to effect further interactions, and hence may in turn change the web of determinants, depending on the subsequent outcome (Bittencourt et al., 2016).

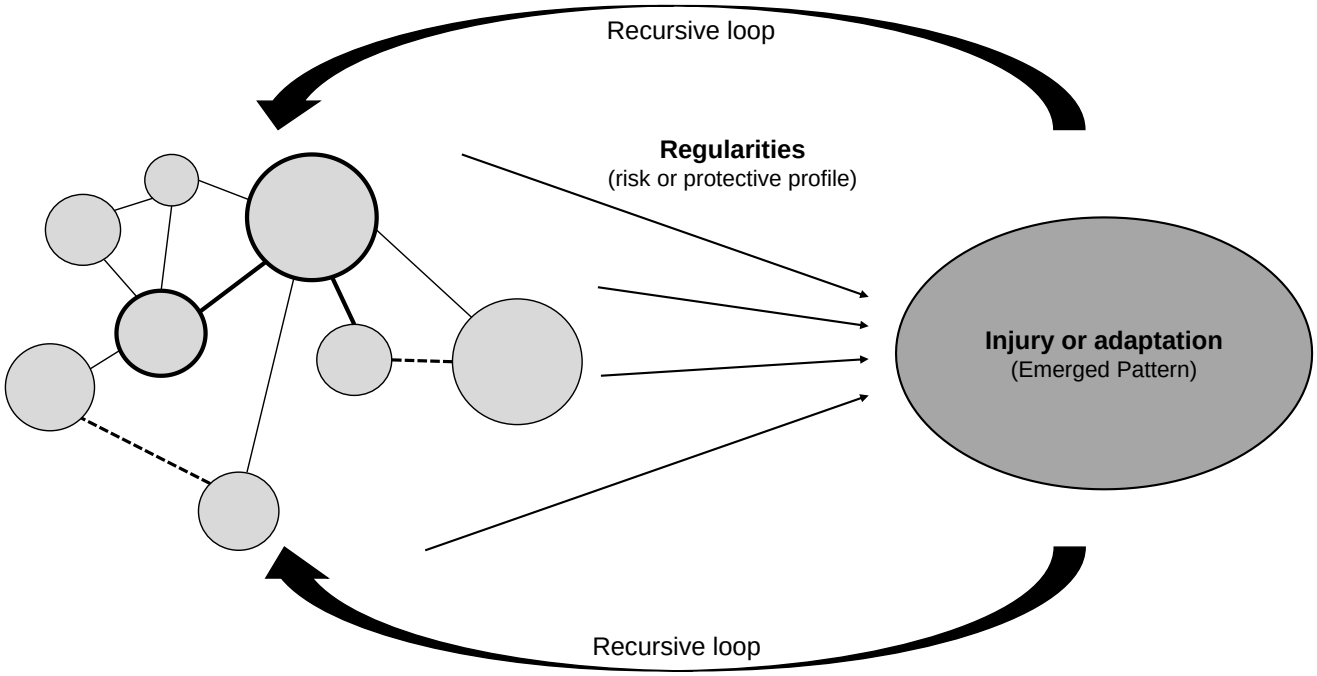


Figure 1.4: Complex systems model for sports injury. The circles represent the web of determinants (variables) comprised of contributing units of varying influence/weight. Circles and lines in darker borders represent those with more interactions and greater weight, whereas those in lighter borders represent less influence. Dotted lines display weaker interactions. Figure adapted from Bittencourt et al. (2016).

1.8 HERITABILITY OF ACL RUPTURE

Heritability estimates are commonly calculated to characterise the genetic contribution to the variation in a clinical phenotype, in comparison to the variation attributed to environmental or other non-genetic factors (Visscher et al., 2008). Expressed as $H^2 = V_g/V_p$, H is the heritability estimate, V_g the variation in genotype, and V_p the variation in phenotype. Heritability estimates range from 0 to 1, where if $H = 1$, this indicates that all variation in a population is due to differences or variation between genotypes, with no environmental impact. On the other hand, a heritability estimate of $H = 0$, indicates no genetic variation, and the phenotypic variation observed is due to the influence of environmental factors (Taylor and Meaney, 2018). It is therefore critical that the heritability of a complex multifactorial phenotype is estimated so that it can inform best study design and sample size estimation. The heritability of ACL rupture has also been explored.

To date, a handful of studies in various populations have described a plausible familial predisposition to ACL rupture. In the 1980's a study first reported a familial predisposition to ACL rupture (Lambert, 1984). A decade later, Harner et al. (1994) reported a genetic basis for ACL rupture predisposition, and later work by Flynn et al. (2005) suggested individuals with an ACL tear were twice as likely to have a first-degree relative with a history of ACL rupture. Most recently, Magnusson et al. (2020) estimated a fairly large ~69% heritability component to ACL rupture using a twin study approach. Twin studies serve as an ideal model to explore the genetic contribution to a trait, as they allow for an individual estimation of the genetic versus environmental contributors of variance, for a given trait (Magnusson et al., 2020). On-going research investigating familial predisposition has highlighted specific anatomical, anthropometric, biomechanical and neuromuscular components potentially underpinning the

genetic predisposition to ACL rupture (Goshima et al., 2014; Hewett et al., 2010b; Kay et al., 2015).

The theme that an individual's genetic architecture influences predisposition to ACL rupture is growing. It is therefore important that the genetic components and specifically genetic polymorphisms be explored to increase our understanding of predisposition, and identify the particular components accounting for the variability in predisposition. A single nucleotide polymorphism (SNP) is a variation at a single position in a DNA sequence, and occurs at a rate of approximately 1 per 1000 base pairs (Mullikin et al., 2000). SNPs are the most common types of genetic variation, and are associated with predisposition to many common human diseases. For this reason, SNPs are commonly used in genetic association studies to explore the genome for variation associated with disease predisposition (Bell, 2002). To date, over 80 loci within numerous genes have been implicated with predisposition to musculoskeletal injuries (Rahim et al., 2019). The majority of which are characterized as SNPs, while others include ins/del and tandem repeat polymorphisms, summarized in Rahim et al. (2019).

1.9 A GENETIC BASIS FOR ACL RUPTURE

To date, a large majority of studies implicating genetic loci in the risk profile of musculoskeletal injuries have followed a candidate gene, case-control approach (Rahim et al., 2016a). Implementing this hypothesis driven approach, genes are selected based on biological function and *a priori* hypothesis that the gene product is involved in injury development (Gibson, 2009). Typically, candidate genes are genotyped in a cohort of cases presenting with the phenotype of interest, and matched controls (Long, 2001). So far, the research has focussed on genes encoding collagenous and noncollagenous matrix components, as well as numerous

key cell signalling factors involved in musculoskeletal biological processes. Together these components regulate cellular homeostasis and tissue-remodelling pathways in response to mechanical loading, and to date over 80 loci have been associated with the risk of musculoskeletal injury. Many of these loci map to regions encoding structural components of connective tissue, while others map to regions involved in the regulation of the ECM and cell signalling pathways. The genetic loci implicated in ACL rupture predisposition are summarised in **Table 1.1**, with more detail described in the following sections below.

1.9.1 Structural Components

Type I collagen (**Figure 1.5**) is the major fibrillar collagen found in ligaments, making up 70-80% of ligament dry weight (Frank, 2004). It is a heterotrimer comprised of two α1(I) and one α2(I) chain, encoded by the *COL1A1* (chr17q21.33) and *COL1A2* (chr7q21.3) genes respectively. Rare mutations in *COL1A1* cause severe connective tissue disorders such as Ehlers-Danlos syndrome (D'Alessio et al., 1991) and osteogenesis imperfecta (Bonod-Bidaud and Ruggiero, 2013). Furthermore, in sports-related musculoskeletal injuries. Two functional *COL1A1* polymorphisms in the Sp1 binding site in intron 1 (rs180012 G/T) and in the proximal promoter (rs1107946 G/T) were associated with the risk of ACL rupture (Ficek et al., 2013; Posthumus et al., 2009a; Stępień-Słodkowska et al., 2016), cruciate ligament injury (Khoschnau et al., 2008) shoulder dislocations (Khoschnau et al., 2008) and Achilles tendon ruptures (Posthumus et al., 2008) in single populations.

However, genetic studies in single populations are characteristically limited by small sample sizes and subsequently power. Therefore making it difficult to establish an association with a phenotype, to be reflective of a biological consequence (Gibson, 2009). In a recent study

combining populations to increase sample size, and therefore statistical power, a significant association of the rs1800012 polymorphism with the risk of ACL rupture was identified in a large combined cohort (Gibbon et al., 2020). The rs1800012 polymorphism is rare in many populations and therefore it was not surprising that large cohorts are required to identifiy a positive association with this DNA variant and may explain differences in association noted at this locus. Functionally, the rs1107946 polymorphism (Jin et al., 2009) and Sp1 binding site polymorphism (rs180012) have been shown to be important in collagen transcription (Mann et al., 2001). The rs1107946 G allele is associated with greater transcription compared to the T allele, and the rs1800012 T allele similarly displays increased levels of *COL1A1* mRNA compared to the G allele (Jin et al., 2009; Mann et al., 2001). Moreover, these polymorphisms are in linkage disequilibrium and the (rs1107946 G/T - rs1800012 G/T) G-T inferred haplotype was associated with altered risk of ACL rupture (Ficek et al., 2013). It may be postulated that an increase in the expression of the *COL1A1* gene, which results in type I collagen homotrimer formation and incorporation into the tissue (Jin et al., 2009), may alter the biomechanical properties of ligament, and in so doing, alters an individual's predisposition for ACL rupture (Gibbon et al., 2020).

In addition to type I collagen, other fibrillar collagens including type III and V (**Figure 1.5**) maintain the structural integrity of tendon and ligament tissue (Wu et al., 2010). Type III in particular plays an important role in adjusting the strength and flexibility of tissues, and in wound healing. Upregulated in the early phases of wound healing, levels of type III collagen remain high for several weeks after injury. After which, type III collagen is replaced by type I, resulting in increased tissue strength (Asling et al., 2009). Type III collagen is encoded by the *COL3A1* (2q32.2) gene and is a homotrimer comprised of three identical α1(III) chains. Disease-causing mutations in *COL3A1* cause severe disorders including vascular type Ehlers–

Danlos syndrome and several types of fibroses (Kuivaniemi and Tromp, 2019). Interestingly, a *COL3A1* exonic rs1800255 (G/A) polymorphism previously linked to the risk of sporadic intracranial aneurysms (Chen et al., 2010a) and pelvic organ prolapse (Kluivers et al., 2009) was associated with altered risk of ACL rupture (O'Connell et al., 2015; Stępień-Słodkowska et al., 2014).

Type V collagen intercalates with type I and III (**Figure 1.5**) and although quantitatively minor, plays an important role in initiating fibrillogenesis and regulating fibril diameter (Birk et al., 1990; Sun et al., 2011). The major isoform of type V collagen comprises of two α1(V) and one α2(V) chains, encoded by the *COL5A1* (chr9q34.3) and *COL5A2* (chr2q32.2) genes respectively. Similar to other collagens, mutations in *COL5A1* and *COL5A2* cause classical Ehlers-Danlos syndrome (Malfait and De Paepe, 2005). Furthermore, two common polymorphisms (rs12722 C/T and rs3196378 C/A) located in the 3'-UTR of *COL5A1* have previously been associated with altered risk of ACL rupture (Posthumus et al., 2009b) chronic Achilles tendinopathy (Mokone et al., 2006; September et al., 2009) and lateral epicondylitis (Altinisik et al., 2015). All these studies suggest a protective role of the C allele. Moreover, Laguette et al. (2011) described two major functional forms of the *COL5A1* 3'-UTR (C- and T-forms) which showed significant differences in mRNA stability (Laguette et al., 2011).

Furthermore, an additional three polymorphisms (rs71746744 -/AGGG, rs16399 ATCT/- and rs1134170 A/T) within or upstream of the *COL5A1* 3'-UTR functional region (Laguette et al., 2011) were independently associated with Achilles tendinopathy (Abrahams et al., 2013). Polymorphisms in the *COL5A1* gene appear to alter the predicted secondary structure of the 3′-UTR, with the evidence suggesting the T functional form of the *COL5A1* 3'-UTR results in increased *COL5A1* mRNA stability (Laguette et al., 2011). A proposed theory put forward by

Collins and Posthumus (2011) suggest an increase in mRNA stability results in an increased type V collagen production among individuals with the *COL5A1* rs12722 TT genotype and that the subsequent increased type V collagen production may confer a structural effect on the fibril architecture (Collins and Posthumus, 2011). They further hypothesise that the resulting change in the collagen fibril diameter and density, in turn can effect the biomechanical properties (ultimate tensile strength, stiffness and creep) of the musculoskeletal soft tissues. And that these changes within the collagen fibril, may potentially lead to an increased risk of musculoskeletal injury (Collins and Posthumus, 2011).

Another minor collagen, similar in structure and functional homology to type V collagen, is type XI collagen. Predominantly expressed in cartilage, type XI collagen is also expressed in developing tendons (Wenstrup et al., 2011) and regulates collagen fibril assembly and diameter (Fichard et al., 1995). Type XI collagen is a heterotrimer consisting of α1 (XI), α2 (XI) and α3 (XI) chains encoded by the *COL11A1* (1p21), *COL11A2* (6p21.3) and *COL2A1* (12q13.11) genes respectively. Mutations in type XI collagen genes cause various inherited Mendelian connective tissue disorders (Majava et al., 2007) and single nucleotide polymorphisms *COL11A1* rs3753841 (T/C), *COL11A1* rs1676486 (C/T) and *COL11A2* rs1799907 (T/A) have been linked to several disorders involving the spine (Koga et al., 1998; Koyama et al., 2012; Maeda et al., 2001; Mio et al., 2007; Videman et al., 2009). In exercise related injury phenotypes, a functional *COL11A1* rs3753841 (T/C), rs1676486 (C/T) and *COL11A2* rs1799907 (T/A) pseudo-haplotype interacted with a functional polymorphism within *COL5A1* to modulate the risk of chronic Achilles tendinopathy (Hay et al., 2013). To current knowledge, type XI collagen polymorphisms are yet to be explored in human ACL rupture risk. However, in a domestic dog model, a *COL11A1* polymorphism (rs8652327 T/C) was associated with cranial cruciate ligament rupture – a condition comparable to human ACL rupture (Baird et al.,

2014). Due to the available evidence in other exercise related phenotypes, and the findings from a canine model, polymorphisms localising to the type XI gene require consideration in ACL rupture predisposition.

Other non-fibrillar collagens, classed as fibril-associated collagens with interrupted triple helices (FACITs) are also present in connective tissue (**Figure 1.5**). Type XII collagen falls within the FACIT family, and primarily localises to type I collagen-rich tissues such as bone, tendon, muscle and ligament (Delbaere et al., 2020). Type XII collagen, encoded by the *COL12A1* gene (6q13-q14.1) is a homotrimer consisting of three α1 (XII) chains. Each chain consists of two helical collagenous domains (COL1 and COL2) flanked by three non-collagenous regions (NC1, NC2, and NC3). Collagen type XII associates with the surface of the collagen fibril (Shaw and Olsen, 1991) at sites of cell-matrix communication (Bader et al., 2009) and has also been shown to interact with small leucine-rich proteoglycans (SLRPs) decorin and fibromodulin (Font et al., 1996). Type XII collagen has numerous functions, some of which include a role in fibrillogenesis, forming inter-fibrillar connections and mediating fibril interaction with other extracellular and cell surface molecules within ligaments, regulating expression in response to mechanical loading and increased content during healing along with type III and V collagen, described in Delbaere et al. (2020). Mutations in *COL12A1* have been observed in patients presenting with a mixed myopathic and Ehlers-Danlos syndrome like phenotype (Hicks et al., 2014; Zou et al., 2014). Moreover, *COL12A1* knockout studies in mice indicate varying muscle weakness, with decreased grip strength and connective tissue disorders (Zou et al., 2014).

In ACL rupture predisposition, a non-synonymous polymorphism (rs970547 G/A) residing in exon 65 of *COL12A1* was associated with altered risk in various study groups. In a South

African cohort, the AA genotype was associated with increased risk in females (Posthumus et al., 2010) which was supported by a recent study in Chinese males (Zhao et al., 2020). Further to that, the alternate AG and GG genotypes were associated with reduced risk in a small sampling group of predominantly male participants from India (Rakesh et al., 2016). Exploration of the interaction effect of collagen genes further highlighted this locus (O'Connell et al., 2015). As inferred pseudo-haplotype analysis constructed from the *COL5A1* rs12722 and *COL12A1* rs970547 polymorphisms, indicate the potential for gene-gene interactions between these loci and altered risk of ACL rupture. In contrast to these findings however, the rs970457 association with altered risk of ACL rupture was not observed in Polish male football players (Ficek et al., 2014) or in Norwegian and Finnish female athletes (Sivertsen et al., 2019).

Finally, a non-classical minor fibrillar collagen, type XXVII has been explored in connective tissue disorders. This homotrimeric collagen consists of three α1 (XXVII) chains, encoded by the *COL27A1* (9q32) gene. Collagen type XXVII differs structurally from the classical fibrillar collagens in numerous respects (Plumb et al., 2011). The major triple helical domains are shorter, and type XXVII collagen has two interruptions in the characteristic Gly-X-Y repeat at conserved locations in the major helical domain. Additionally, type XXVII collagen lacks the N-terminal region, and the N-terminal minor helical domain characteristic of classical fibrillar collagens. Rather, a 'variable' domain and a thrombospondin domain are present (Koch et al., 2003). Mutations in *COL27A1* have been linked to Steel syndrome (Gonzaga-Jauregui et al., 2020) and osteochondrodysplastic phenotypes. In exercise-related phenotypes, an inferred pseudo-haplotype constructed with an intronic *COL27A1* polymorphism (rs946053) and two *TNC* polymorphisms (rs13321 and rs2104772) modulate the risk of Achilles tendinopathy in a combined South African and Australian cohort (Saunders et al., 2013). However, in ACL rupture predisposition, the investigation of three *COL27A1* polymorphisms (rs2567706 A/G,

rs2241671 G/A, and rs2567705 A/G) found no association with altered risk (Gibbon et al. 2018).

In addition to collagenous constituents, other ECM components are critical in the structural maintenance of connective tissue. Fibrillins are large glycoproteins present in the ECM of tendons and ligaments. They are characterised by three homologous molecules (fibrillin-1, fibrillin-2 and fibrillin-3) which assemble into microfibrils in the ECM and establish a scaffold for the deposition of elastin (Sakai and Keene, 2019). Elastic fibres made-up of fibrillins and elastin have three major functions: provide mechanical properties, including elastic recoil and resilience to tissue (Butler et al., 1978) regulate TGFβ activity (Charbonneau et al., 2004; Feng and Derynck, 2005) and participate in cell migration, survival and differentiation (Ito et al., 1997). Mutations identified in the fibrillin-1 encoding gene (*FBN1*) mapped to 15q21.1, are linked to inherited genetic disorders such as Marfan syndrome, Weill-Marchesani syndrome, geleophysic dysplasia, and acromicric dysplasia (Sakai and Keene, 2019). While mutations in fibrillin-2 (*FBN2*) located on 5q23.3, associate with pathologies such as congenital contractual arachnodactyly (Gupta et al., 2002). Interestingly, a rare pathogenic polymorphism in *FBN2* (c.5009T>G; p.Phe1670Cys) was identified in a three-generation family, with early onset of carpal tunnel syndrome (Peeters et al., 2020). In exercise-related phenotypes, a polymorphism in intron 7 of *FBN2* (rs331079 G/T) previously associated with intracranial aneurysms (Ruigrok et al., 2006) was associated with altered risk of ACL rupture and Achilles tendinopathy in participants from Australia and South Africa (El Khoury et al., 2014). Although intronic variants do not influence the primary structure of a protein, they have roles necessary for effective protein expression. Currently, the function of the rs331079 polymorphism has not been described, and thus how it modulates ACL rupture and AT risk is as yet unknown (El Khoury et al., 2014).

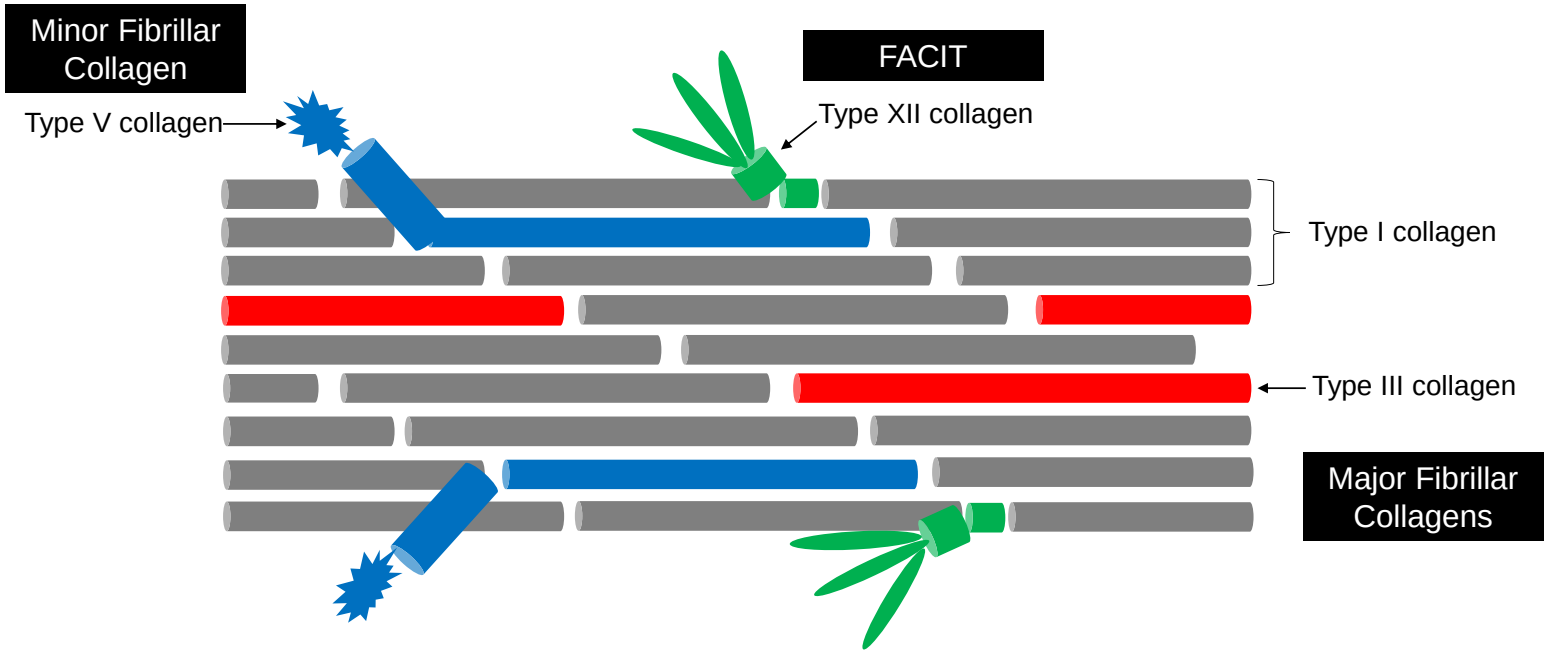


Figure 1.5: Structure of the collagen fibril. The arrangement of Types I, III, V and XII collagen are shown. FACIT: Fibril Associated Collagens with Interrupted Triple Helices. Modified from Fang et al. (2012).

1.9.2 Matrix Metalloproteases

Matrix metalloproteases (MMPs) are a family of zinc-containing enzymes that catalytically degrade various collagenous and non-collagenous components of the ECM, and are crucial in maintaining the homeostatic balance of the ECM (Birkedal-Hansen et al., 1993; Somerville et al., 2003). Several polymorphisms in MMP encoding genes have been associated with connective tissue disorders such as rheumatoid arthritis (Chen et al., 2012) osteoarthritis (Lepetsos et al., 2014) lumbar disc degeneration (Martirosyan et al., 2016) and idiopathic scoliosis (Zhao et al., 2016). Moreover, polymorphisms localising to MMP genes have also been explored in exercise-related phenotypes, including ACL rupture (Lulinska-Kuklik et al., 2019b; Malila et al., 2011; Posthumus et al., 2012) and Achilles tendinopathy (El Khoury et al., 2016; Gibbon et al., 2017; Raleigh et al., 2009).

In ACL rupture predisposition, polymorphisms in four MMP genes (*MMP1* rs1799750 1G/2G, *MMP3* rs679620 A/G, *MMP10* rs486055 C/T, and *MMP12* rs2276109 A/G) investigated in a South African Caucasian cohort, noted an independent association of *MMP12* rs2276109 and inferred haplotypes constructed from the four polymorphisms, with altered risk (Posthumus et al., 2012). Furthermore, *MMP3* rs679620 was further highlighted in ACL rupture predisposition in a Polish cohort, in addition to the *MMP3* rs591058 (T/C) polymorphism (Lulinska-Kuklik et al., 2019b). However, despite the similarities noted at the *MMP3* gene, in a more recent study in Polish participants, the previous associations noted at *MMP1* rs1799750, *MMP10* rs486055 and *MMP12* rs2276109 were not replicated (Lulinska-Kuklik et al., 2020). Thus, the evidence is unclear if polymorphisms mapping to MMPs do indeed contribute to ACL rupture predisposition.

The lack of agreement in study findings may be due to the small sample sizes characteristic of the studies reporting on MMPs and predisposition to ACL rupture. Therefore, to gain a better understanding of the role of MMPs in musculoskeletal injury profiles, there is a need to increase independent sample sets by harnessing larger collaborative efforts.

1.9.3 ECM Regulatory Components

Proteoglycans are glycosylated proteins that consist of a protein core, with one or more covalently attached glycosaminoglycan (GAG) chain(s) at a specific site (Chen and Birk, 2013). They form part of the ground substance of both ligament and tendons and are one of the proteins upregulated during injury of these tissues (Dunkman et al., 2014) and therefore require exploration in context of understanding the aetiology of ligament injury phenotypes. Proteoglycans are categorised into four major classes based on location (cellular and subcellular) overall gene/protein homology, and the presence of specific protein components within their respective protein cores (Iozzo and Schaefer, 2015). Class I proteoglycans consists of intracellular secretory granules, of which there is only one: serglycin, a unique proteoglycan with heparin side chains. Class II is comprised of cell surface proteoglycans that are transmembrane, or GPI-anchored, and class III are pericellular basement membrane zone proteoglycans. Class IV are located extracellularly, and are classified as hyalectan-lectincan, SLRPs and spock (Iozzo and Schaefer, 2015). SLRPs are further divided into five classes, and consist of a protein core containing numerous leucine-rich repeats (LRR) to which one to two chondroitin (CS) or dermatan sulfate (DS), or several keratan sulfate (KS) chains are attached (Iozzo and Schaefer, 2015).

The most abundant SLRPs in ligament are decorin (DCN) and biglycan (BGN) (**Figure 1.6**). Decorin interacts with fibrillar collagens in the connective tissue matrix, and is a key regulator of fibrillogenesis (Chen et al., 2013). It prevents lateral fusion of collagen fibrils and promotes the correct formation of the fibrils (Iozzo, 1997). Furthermore, decorin maintains tissue integrity by binding to fibronectin and thrombospondin, and is involved in collagen degradation (Bhide et al., 2005) cellular growth (Moscatello et al., 1998) and extracellular signalling (Schönherr et al., 2005). Further, decorin is involved in cellular proliferation and the cell cycle (Iozzo et al., 1999) angiogenesis (Järveläinen, et al., 2015) and apoptotic processes (Schönherr et al., 2005) and can act as a damage-associated molecular pattern (DAMP) interacting with the immunity receptors TLR 2 and 4, therefore regulating a sterile inflammatory response within connective tissue (Merline et al., 2011). Interestingly, decorin and biglycan compete for the same site on collagen (Iozzo, 1997) and in the absence of decorin, biglycan is overexpressed and functionally compensates for decorin (Zhang et al., 2009). Mutations in the decorin gene (*DCN* chr12q22) causes congenital stromal corneal dystrophy (Bredrup et al., 2005) and targeted disruption of decorin in mice, results in disorganized collagen fibril formation and skin fragility (Danielson et al., 1997).

Like decorin, biglycan is primarily involved in regulating collagen fibrillogenesis and structure by interacting with collagen fibrils (Chiu et al., 2012). Expressed in skin, cartilage, bone, tendon and ligaments (Corsi et al., 2002) biglycan is an important regulator of tendon development, aging and healing (Dunkman et al., 2014) and plays a role in the inflammatory pathway (Kolb et al., 2001). Furthermore, biglycan functions in ECM remodelling pathways, where together with decorin, acts to regulate TGFβ signalling through a dose-dependent negative feedback loop (Fu et al., 2005). In knockout mouse models, biglycan deficiency induces morphological changes to the collagen fibrils. These changes include irregular fibrils

and alterations in fibril diameter resulting in mechanically inferior tissue (reviewed in Ameye and Young, 2002). It seems the tissue environment governs biglycan function, where under physiological conditions it acts as a structural component. However, under tissue stress or injury, biglycan is released from the ECM in a soluble form, transforms into an endogenous ligand of innate immunity receptors, and interacts with toll-like receptors on macrophages to aid in triggering an inflammatory response (Frey et al., 2013).

Biglycan is encoded by the *BGN* gene located on chrXq28, and contains 8 exons, and 7 possible transcripts. The promoter region has several Sp1 binding sites and several transcription start sites. Furthermore, there are six interleukin (IL)-6 response elements within the promoter, which generally induce biglycan expression when IL-6 is bound (Ungefroren and Krull, 1996). Mutations in the *BGN* gene are observed in X-linked spondyloepimetaphyseal dysplasia (an inherited disorder of bone growth that results in dwarfism, skeletal abnormalities, and visual limitations) (Cho et al., 2016) and in an X-linked syndromic form of thoracic aortic aneurysm and dissection (Meester-Loeys Syndrome) associated with increased TGFβ signalling (Meester et al., 2017).

Other large aggregating hyalectan-lectincan proteoglycans, such as aggrecan and versican are present in lesser amounts in ligament tissue (Ilic et al., 2005). Aggrecan (**Figure 1.6**) is a large structural proteoglycan that forms aggregates via interactions with hyaluronan in the ECM, and is thought to interact with the collagen fibril indirectly (Heinegård, 2009). By binding water through its highly negative charge, aggrecan stabilises the collagen network and maintains the structural integrity of connective tissue by providing viscoelastic properties and resisting compressive forces during loading (Kresse and Schönherr, 2001). Aggrecan is composed of a protein core comprising three globular domains, G1, G2 and G3 (**Figure 1.6**) each of which

have a specific function (Kiani et al., 2002). The G1 domain of the N-terminal binds hyaluronan (HA) via a link protein (LP), forming a very stable G1/hyaluronan/link protein complex that provides mechanical stability to the tissue (Kiani et al., 2002). The function of the highly conserved G2 domain, separated from G1 via an extended interglobular domain, is less clear. It does not interact with HA or LP (Watanabe H et al., 1997) but may have a direct or indirect role in GAG chain attachment (Kiani et al., 2002) and in inhibiting product secretion to produce a mature functional aggrecan molecule (Kiani et al., 2001). The largest domain lies between G2 and G3, and contains a GAG attachment region with an extended stretch of KS and CS chains. The function of the KS region is uncertain, but is thought to be involved in tissue distribution of aggrecan, and play a role in product processing (Kiani et al., 2001). Due to its structure, KS chains hold water within aggrecan, and thus contribute to the load-bearing capacity of aggrecan in connective tissue (Kiani et al., 2002). The CS region forms the bulk of the GAG attachment region with ~100 CS chains. The negatively charged CS chains account primarily for the major function of aggrecan in its ability to hold large amounts of water in the ECM, and thus provide a critical structural role in connective tissue (Kiani et al., 2002). Lastly, the C-terminal G3 domain consists of four structural motifs which include an epidermal growth factor (EGF) repeat, a calcium-binding EGF repeat, a C-type lectin domain, and a complement regulatory protein repeat (Morawski et al., 2012). The G3 domain forms a link between aggrecan and the ECM matrix, by calcium-dependant interactions with simple sugars (Saleque et al., 1993) and ECM molecules such as tenascins, fibulins (Aspberg et al., 1997) and fibrillin-1 (Isogai et al., 2002). Further to that, animal models have shown the G3 domain regulates the attachment of GAG chains in the CS region, and affects the secretion of aggrecan (Kiani et al., 2001).

Several polymorphisms localising to the aggrecan encoding gene, *ACAN* (chr15q26.1) are associated with human conditions such as sickle cell anaemia (Sebastiani et al., 2010) myopia (Yip et al., 2011) osteochondritis dissecans (Stattin et al., 2010) and lumbar disc degeneration (Videman et al., 2009). Versican is the largest proteoglycan within the hyalectan class, and is encoded by the *VCAN* gene mapped to chr5q14.2-q14.3. Versican functions within the tensional regions of ligaments and tendons (Campbell et al., 1996; Waggett et al., 1998) where it plays a role in cell adhesion, migration and inflammation (Wight et al., 2014). Mutations in *VCAN* cause two autosomal dominant eye disorders, Wagner syndrome and erosive vitreo-retinopathy (Mukhopadhyay et al., 2006).

To date, two studies report the association of proteoglycan genomic loci in the predisposition to ACL rupture. In a case-control genetic association study, the *ACAN* rs1516797 (T/G) polymorphism located within exon 4, was associated with altered risk of ACL rupture in a South African Caucasian (Mannion et al., 2014) and Polish (Cięszczyk et al., 2017) cohort. More specifically, the *ACAN* rs1516797 GT genotype was significantly associated with increased risk, in Polish participants (Cięszczyk et al., 2017) with a trend for TT genotype and a reduced risk noted in the South African group (Mannion et al., 2014). Furthermore, the rs1516797 G allele was significantly associated with increased risk, and inferred haplotype analyses across *ACAN* polymorphisms, identified significant associations with altered risk of ACL rupture in the South African group (Mannion et al., 2014).

For the *DCN* gene, a polymorphism residing in an intronic region (rs516115 A/G) was significantly associated with altered risk of ACL rupture. Where the rs516115 GG genotype and G allele conferred reduced risk in female participants from South Africa (Mannion et al., 2014) with a trend for the G allele association noted in Polish participants (Cięszczyk et al.,

2017). Furthermore, the rs516115 AA genotype was significantly associated with increased risk in the South African females (Mannion et al., 2014) with a trend noted in females from Poland (Cięszczyk et al., 2017).

For *BGN* polymorphisms, an independent association of the *BGN* rs1042103 (G/A) A allele was noted with increased risk in Polish males (Cięszczyk et al., 2017) and inferred haplotype analyses across *BGN* (rs1126499, rs1042103) polymorphisms were highlighted with altered risk of ACL rupture in participants from Poland and South Africa (Cięszczyk et al., 2017; Mannion et al., 2014). In other related phenotypes, exploration of the *BGN* rs1126499, *BGN* rs1042103 and *ACAN* rs1516797 polymorphisms revealed an association of the rs1126499 polymorphism with carpal tunnel syndrome in a South African cohort of mixed ancestry (Burger et al., 2014). In addition, the association of the rs1042103 and rs1516797 polymorphisms with risk of lumbar disc degeneration in a male Finnish twin cohort (Videman et al., 2009).

In summary, the proteoglycans play a crucial role in maintaining the structural integrity of ligament tissue (Iozzo and Schaefer, 2015) and based on current research findings, there is evidence to suggest polymorphisms in these genomic regions may impart functional consequence in musculoskeletal injury predisposition (Burger et al., 2014; Cięszczyk et al., 2017; Mannion et al., 2014; Videman et al., 2009). Despite being underpowered due to sample size, the studies conducted by Mannion et al. (2014) and Cięszczyk et al. (2017) highlight regions potentially involved in ACL rupture predisposition, which require further exploration in larger, collective data sets.

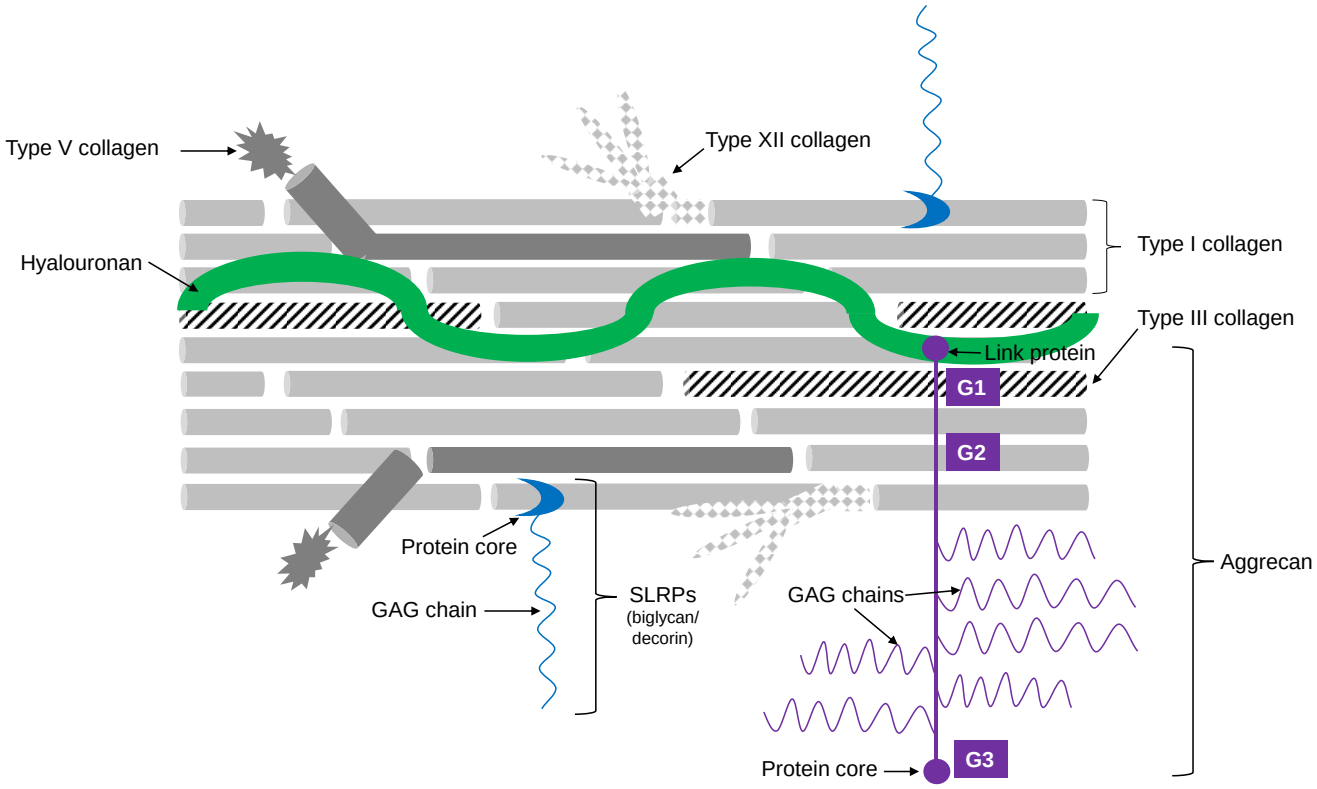


Figure 1.6: Interaction and structural localisation of small leucine-rich proteoglycans biglycan and decorin, and large aggregating hyalectan-lectincan proteoglycan aggrecan, within the collagen fibril. The protein core and glycosaminoglycan chains for the proteoglycans are shown, in addition to the link protein, and three globular domains of aggrecan. Modified from Fang et al. (2012) and Roughley (2006).

1.9.4 Cell Signalling Factors, Angiogenesis and Apoptosis

Ligaments respond to mechanical stimuli by initiating a plethora of matrix remodelling pathways which function to either degrade or synthesise components of the ECM, to maintain homeostasis (Yang et al., 2005) and to induce adaptive changes to the tissue for future load bearing (Cox and Erler, 2011). This process is tightly controlled by numerous cytokines, growth factors and signalling molecules. The interleukins are a group of cytokines upregulated in response to mechanical loading and play an important role in the inflammatory response during ligament and tendon injury (Tsuzaki et al., 2003). Furthermore, growth factors such as TGFβ and growth differentiation factor 5 (GDF5) play a key role in the development, remodelling and proliferation of musculoskeletal soft tissue (Gumucio et al., 2015). As such, polymorphisms localising to the genes encoding for these cytokines and growth factors have been investigated with risk of ligament and tendon injury (Burger et al., 2015; Chen et al., 2015; Laguette et al., 2020; Rahim et al., 2017; September et al., 2011; Suijkerbuijk et al., 2019).

For the interleukins, genes explored include interleukin-1β (*IL1B* rs16944 C/T), interleukin-1 receptor antagonist [*IL1RN* rs2234663 variable nucleotide tandem repeat (VNTR)], interleukin-6 (*IL6* rs1800795 G/C), and interleukin-6 receptor (*IL6R* rs2228145 A/C). IL1B is a pro-inflammatory cytokine secreted by macrophages in injured soft tissue, it upregulates its own and other inflammatory mediator expression such as IL6, which in turn upregulates the expression of proteins involved in ECM remodelling (Tsuzaki et al., 2003). Independent associations were noted for the *IL1B* rs16944 (2q14.1) promoter polymorphism and altered risk of ACL rupture (Rahim et al., 2017) and Achilles tendon rupture (Brown et al., 2017). IL6 functions in the apoptotic pathway, and polymorphisms localising to *IL6* gene have been

associated with various disorders (Fishman et al., 1998; Testa et al., 2006). In ACL rupture predisposition, the *IL6* rs1800795 (7p15.3) polymorphism was associated with reduced risk in a Polish cohort (Lulińska-Kuklik et al., 2019b) with the *IL6R* rs2228145 polymorphism similarly conferring reduced risk of ACL rupture (Suijkerbuijk et al., 2019) and carpal tunnel syndrome (Burger et al., 2015). Moreover, combinations of the *IL1B*, *IL6* and *IL6R* polymorphisms and the previously associated *COL5A1* rs12722 were associated with altered risk of ACL rupture (Suijkerbuijk et al., 2019) and Achilles tendinopathy (September et al., 2011). Suijkerbuijk et al. (2019) took a step further and conducted *in vitro* work to investigate the functional effect of these polymorphisms on the structural ECM components. The findings indicate the modulation of key structural and fibril-associated ECM component expression by the *IL1B* rs16944 and *IL6* rs1800795 polymorphisms, providing support for the current associations noted for these polymorphisms and altered ACL rupture risk.

For the TGFβ family, the exonic *TGFB1* rs1442 G/C and *TGFBR3* rs1805113 (G/A) polymorphisms were associated with altered risk of ACL rupture (Laguette et al., 2020). Furthermore, associations of inferred allele combinations constructured for the *TGFB1* and *TGFBR3* genes (rs1805113 G/A, rs1805117 T/C, and rs1442 G/C) were observed (Laguette et al., 2020). TGFB1 participates in cell-collagen interactions, whereare TGFBR3 plays a role in the TGFB signalling pathway (Laguette et al., 2020). Currently, the functional significance of these loci are not understood, and therefore further study is required to elucidate the biological and functional significance of this region in ligament injury pathogenesis.

GDF5 belongs to the TGFβ superfamily (Egli et al., 2009), and plays a role in the development and repair of musculoskeletal soft tissues, particularly in ligament homeostasis (Chen et al., 2015). Several SNPs in the *GDF5* gene have been explored with the risk of sports related

conditions and injuries such as osteoarthritis, meniscus injury and ACL and tendon rupture (Ge et al., 2014; Raleigh et al., 2013; Southham et al., 2007). Further, a functional polymorphism (rs1413383 A>G), in the 5′-UTR of the *GDF5* gene was associated with the risk of Achilles tendon pathology (Posthumus et al., 2010) and ACL ruptures in a Chinese population (Chen et al., 2015).

Angiogenesis is the formation of new blood vessels from existing vasculature, and is controlled by several growth factors and cytokines. It is hypothesised that angiogenesis can be triggered by several stimuli, which can be biological and/ or mechanical, in an effort to promote remodelling of both tendon and ligament. This has been supported by studies identifying increased pro-angiogenic expression protein profiles such as the A-isoform of vascular endothelial growth factor (VEGFA) in tendon cells (Mousavizadeh et al., 2014; Nakama et al., 2006; Petersen et al., 2004) and within specimens from ruptured tendons and ligaments, degenerative tendons, and ligament reconstructive surgery (Beye et al., 2008; Bray et al., 2003; Pufe et al., 2001; Yoshikawa et al., 2006). When exploring functional partners regulating the extracellular matrix (ECM) of tendon and ligament, the components of angiogenesis is placed centrally within this network (Rahim et al., 2021). VEGFA is the primary regulator of angiogenesis, and its activity is further regulated by binding to its receptor the kinase insert-domain receptor (KDR) (Ferrara et al., 2003).

VEGFA-KDR signalling is linked to the pathogenesis of conditions such as ocular disease (Ferrara and Adamis, 2016) and pre-eclampsia during pregnancy (Koga et al., 2003). In the lymphatic system, the VEGFA pathway is linked to an autosomal dominant congenital form of hereditary lymphedema (Nonne-Milroy disease). Characterized by excessive accumulation of protein-rich extracellular fluid in the interstitial space, due to impaired lymphatic vessel

function, individuals with Nonne-Milroy disease have reported tyrosine kinase-inactivating mutations present in the vascular endothelial growth factor-3 gene (Irrthum et al., 2000). Furthermore, the expression of VEGFA and KDR in non-endothelial tumour cells stimulates the proliferation and survival of endothelial cells, leading to the formation of new blood vessels and promoting tumour expansion (Nishida et al., 2006).

As angiogenesis plays an important role in tissue vascularity and healing capacity, one could hypothesise that dysregulation of the pathway may have negative implications on ligament tissue integrity, and thus potentially contribute to the pathobiology of ACL rupture. Previous work exploring the potential association of angiogenesis genes and predisposition to ACL rupture, have identified associations of polymorphisms in *VEGFA* (6p21.1) and *KDR* (4q12) with altered risk. For *VEGFA*, two polymorphisms in the promoter region were associated with altered risk in a South African Caucasian cohort (Rahim et al., 2014). Where the *VEGFA* rs699947 (C/A) CC genotype conferred increased risk, and the *VEGFA* rs1570360 (G/A) GA genotype was associated with a reduced risk of ACL rupture (Rahim et al., 2014). A third *VEGFA* polymorphism, located in the 5'-UTR region (rs2010963 G/C) was associated with altered risk of ACL rupture in Polish participants, where the rs2010963 CC genotype conferred increased risk (Lulinska-Kuklik et al., 2019a). For the *KDR* gene, a promoter polymorphism (rs2071559 G/A) was associated with altered risk of ACL rupture in a South African Caucasian cohort, with the rs2071559 GA genotype associating with reduced risk in females (Rahim et al., 2014). A later study by the same authors in a South African Mixed population confirmed the rs2071559 association with reduced risk in male participants (Rahim et al., 2018). Additionally, inferred haplotype analyses further highlighted polymorphisms localising to *VEGFA* and *KDR* with altered risk of ACL rupture risk in the South African (Rahim et al., 2014) and Polish (Lulinska-Kuklik et al., 2019a) cohorts.

Additionally, following a pathway-based approach, the *VEGFA* rs699947 and rs2010963 polymorphisms were further highlighted as the most significant predictors of ACL rupture risk, out of a range of potential intrinsic risk factors, including age, sex, body mass index (BMI) and polymorphisms within five (*IL1B*, *IL6*, *IL6R*, *VEGFA* and *KDR*) genes (Rahim et al., 2021). Moreover, genetic polymorphisms localising to *VEGFA* and *KDR* have also been implicated in the risk profiles of Achilles tendinopathy (Rahim et al., 2016b) patellar and shoulder tendinopathy (Salles et al., 2016) as well as osteoarthritis (Wang et al., 2019) and diabetic retinopathy (Lu et al., 2013).

Evidently, angiogenesis is critical for connective tissue homeostasis. However, the previous *VEGFA* and *KDR* association studies are not consistent in implicating the same polymorphisms with altered rupture risk (Lulinska-Kuklik et al., 2019a; Rahim et al., 2014; Rahim et al., 2018). As previously mentioned with findings in other single populations, this may be due to insufficient power of small sample sizes. Therefore to determine if angiogenesis is one of the key biological pathways contributing to ACL rupture risk, it is paramount that these polymorphisms are explored in larger, more comprehensive cohorts in order to ascertain if, and how these genetic loci may modulate an individual's predisposition to ACL rupture.

Apoptosis is the controlled programming of cell death and in connective tissue is necessary to maintain homeostasis in response to mechanical loading (Millar et al., 2009). An essential role player in the apoptotic signalling cascade is Caspase-8 (*CASP8* 2q33-34.21) a cysteine-aspartic acid protease that is upregulated by IL-1β and IL-6 and functions in propagating the cell death pathway (Kruidering and Evan, 2000; Millar et al., 2009). Polymorphisms within *CASP8* have been associated with various musculoskeletal soft tissue injuries (Brown et al., 2017; Nell et al., 2012; Rahim et al., 2017; Seale et al., 2020). Specifically, the functional *CASP8* rs3834129

(ins/del) ins allele was independently associated with altered risk of ACL rupture in male participants (Rahim et al., 2017). In contrast however, the alternate allele; rs3834129 DD genotype was associated with reduced risk of Achilles tendinopathy in a British cohort (Brown et al., 2017) but with increased risk in a combined South African and Australian cohort (Nell et al., 2012). Inferred haplotypes constructed for the two functional *CASP8* rs3834129 and *CASP8* rs1045485 (G/C) polymorphisms (Sun et al., 2007) revealed an association with the del-G haplotype and increased ACL rupture risk in the South African cohort (Rahim et al., 2017). This genetic interval was further highlighted in a Swedish study group, whereby the addition of *CASP8* 3'UTR rs13113 (T/A) polymorphism to form the rs3834129-rs1045485-rs13113 inferred haplotype, in addition to the rs3834129-rs1045485 inferred haplotype were noted to modulate altered risk of ACL rupture (Seale et al., 2020). However, as observed in the independent associations, the rs3834129-rs1045485 alternate allele combination (ins-G) was associated with increased Achilles tendinopathy risk in the cohort from Britain (Brown et al., 2017). The noted alternate risk allele combinations could be due to the nature of these injuries, chronic (Achilles tendinopathy) vs acute (ACL) and potentially indicates the functional significance of the polymorphisms, in that the genetic sequence related with a condition may be specific to the injury phenotype (Seale et al., 2020).

Table 1.1: Genetic polymorphisms within candidate genes previously associated with anterior cruciate ligament rupture.

Gene and chromosomal location	Encoded protein	Protein function	Polymorphism	Location	Reference
Structural Components					
COL1A1 (17q21)	α1 (I) collagen chain	Major fibrillar collagen	rs1107946 (G/T)	Promoter	Ficek et al., 2013; Gibbon et al., 2020; Stępień-Słodkowska et al., 2016
			rs1800012 (G/T)	Intron 1	Ficek et al., 2013; Posthumus et al., 2009a; Wang et al., 2017
COL3A1 (2q31)	α1 (III) collagen chain	Major fibrillar collagen	rs1800255 (G/A)	Exon 30	O'Connell et al., 2015; Stępień-Słodkowska et al., 2014
COL5A1 (9q34)	α1 (V) collagen chain	Minor fibrillar collagen	rs12722 (T/C)	3' -UTR	Lulińska-Kuklik et al., 2018; O'Connell et al., 2015; Posthumus et al., 2009b
COL12A1 (6q13-q14.1)	α1 polypeptide type XII collagen	Fibril-associated collagens with interrupted triple helices collagen	rs970547 (A/G)	Exon 65	O'Connell et al., 2015; Posthumus et al., 2010; Zhao et al., 2020
Matrix Metalloproteinases					
MMP1 (11q22)	Matrix metalloproteinase 1	Degradation of collagenous and non-collagenous ECM components	rs1799750 (1G/2G)	Promoter	Posthumus et al., 2012
MMP3 (11q22)	Matrix metalloproteinase 3		rs3025058 (5A/6A)	Promoter	Malila et al., 2011
			rs679620 (A/G)	Exon 2	Lulinska-Kuklik et al., 2019b; Posthumus et al., 2012
MMP10 (11q22)	Matrix metalloproteinase 10		rs486055 (C/T)	Exon 1	Posthumus et al., 2012
MMP12 (11q22)	Matrix metalloproteinase 12		rs2276109 (A/G)	Promoter	
Regulators of the Extracellular Matrix					
TNC (9q33)	Tenascin-C glycoprotein	Regulates cell-matrix interactions	rs2104772 (T/A)	Exon 17	Gibbon et al., 2018
FBN2 (5q23-q31)	Fibrillin-2	Component of connective tissue microfibrils	rs331079 (G/T)	Intron 7	El Khoury et al., 2014
ACAN (15q26)	Aggrecan	Major proteoglycan of articular cartilage, contributes to load-bearing properties of cartilage	rs2351491 (C/T)	Exon 11	Cięszczyk et al., 2017; Mannion et al., 2014
			rs1042631 (C/T)		
			rs1516797 (T/G)	Intron 12	
BGN (Xq28)	Biglycan	Collagen fibrillogenesis	rs1126499 (C/T)	Exon 4	Cięszczyk et al., 2017; Mannion et al., 2014; Willard et al., 2018
			rs1042103 (G/A)	Exon 8	
DCN (12q21)	Decorin	Collagen fibrillogenesis	rs13312816 (C/T)	Intron 1	Mannion et al., 2014
			rs516115 (A/G)	Intron 3	
LUM (12q21)	Lumican	Collagen fibrillogenesis	rs2268578 (T/C)	Intron 3	

Cell Signalling, Angiogenesis and Apoptosis					
IL-1B (2q14)	Interleukin-Iβ	Role in the inflammatory pathway and ECM degradation	rs16944 (T/C)	Promoter	Rahim et al., 2017
IL-1RN (2q14)	Interleukin-1 receptor antagonist	Antagonist for IL-1α and IL-1β	rs2234663 VNTR	Intron 2	September et al., 2011
IL6 (1q21)	Interleukin-6	Role in apoptosis and the inflammatory pathway	rs1800795 (G/C)	Promoter	Lulińska-Kuklik et al., 2019b
IL6-R (1q21)	Interleukin-6 receptor	Receptor for IL-6	rs2221845 (A/C)	Exon 9	Suijkerbuijk et al., 2019
VEGFA (6p21)	Vascular epithelial growth factor A isoform	Essential regulator of angiogenesis	rs699947 (C/C)	Promoter	Rahim et al., 2014
			rs1570360 (G/A)	Promoter	
			rs2010963 (G/C)	3'-UTR	Lulinska-Kuklik et al., 2019a
KDR (4q11-4q12)	Kinase insert domain receptor	Receptor for VEGFA and mediates VEGFA signalling	rs2071559 (G/A)	Promoter	Rahim et al., 2014; Rahim et al., 2018
			rs1870377 (A/T)	Exon 11	
TGFβ1 (19q13.2)	Transforming growth factor beta induced	Cell-collagen interactions	rs1442 (G/C)	Exon 6	Laguette et al., 2020
TGFβR3 (1p22.1)	Transforming growth factor-β receptor III	Capturing and retaining TGFβ for presentation to signalling receptors	rs1805113 (G/A)	Exon 13	
GDF5 (20q11.22)	Growth differentiation factor 5	Development and repair of soft tissue	rs1413383 (A/G)	5'-UTR	Chen et al., 2015
CASP8 (2q33-q34)	Caspase-8	Initiation caspase	rs3834129 (ins/del)	Promoter	Rahim et al., 2017; Seale et al., 2020

1.10 MATRIX REMODELLING

To date, most of the genetic loci explored include distinct families of genes. However, as previously mentioned, matrix remodelling does not occur in defined functional spaces, but through the interaction of genes and their encoding proteins, within interconnected biological pathways (Collins et al., 2018). Moreover, their collective interaction regulates the expression patterns of ECM components. In this regard, one can hypothesise that mechanical loading, amongst other biological stimuli such a hypoxia or cell density to name a few, can provide a stimulus through which key molecules initiate downstream signalling pathways, lending to multiple interactions that ultimately conclude in extracellular matrix remodelling and repair mechanisms. A comprehensive explanation of all the genes and gene products involved in connective tissue matrix remodelling is beyond the scope of this thesis. Rather a summary of the key genes, particularly those previously implicated with ACL rupture risk, will be discussed in the following paragraph, and visually represented in **Figure 1.7**.

Cytokines are critical in the initial cell signalling pathways triggered post mechanical loading and following injury to the ligament (**Section 1.9.4**). IL-1β is a pro-inflammatory cytokine that in addition to inducing its own expression, initiates numerous downstream signalling cascades through the upregulation of other components including cytokines (IL-6), growth factors (TGFβ) and collagenases (MMPs) (Tsuzki et al., 2003) (**Figure 1.7**). IL-6 is a pleiotropic cytokine that has both pro-inflammatory and anti-inflammatory effects. Known to induce apoptotic cell death (Millar et al., 2009), it also plays a role in modulating TGFβ, and upregulating the expression of VEGFA (Cohen et al., 1966). As previously mentioned, matrix remodelling pathways act in either a stimulatory or an inhibitory fashion to result in the synthesis or degradation of ECM components to maintain homeostasis. In this regard, TGFβ

plays a pivotal role in remodelling pathways, where it modulates: (1) collagen synthesis (Chan et al., 2008), (2) proteoglycan (ACAN, BGN, DCN) production, with BGN and DCN specifically providing a dose-dependent feedback for TGFβ activity (Fu et al., 2005), (3) Tissue inhibitors of metalloproteinases (TIMPs) which inhibit MMPs, and (4) VEGFA in the angiogenesis signalling pathway (Juneja et al., 2006). Furthermore, VEGFA a potent angiogenic factor critical for wound healing and ligamentization post reconstruction, is independently upregulated in response to loading (Pufe et al., 2005) and additionally, modulated through TGFβ (Juneja et al., 2006) and interestingly through DCN activity (Järveläinen, et al., 2015). Furthermore, it is interesting to note the altered expression of several ECM remodelling genes (*COL1A1, COL3A1, MMP, TIMP1* and *MMP13*) during the graft remodelling phases (necrosis, proliferation and ligamentization) of an ACL reconstruction (Yung et al., 2020).

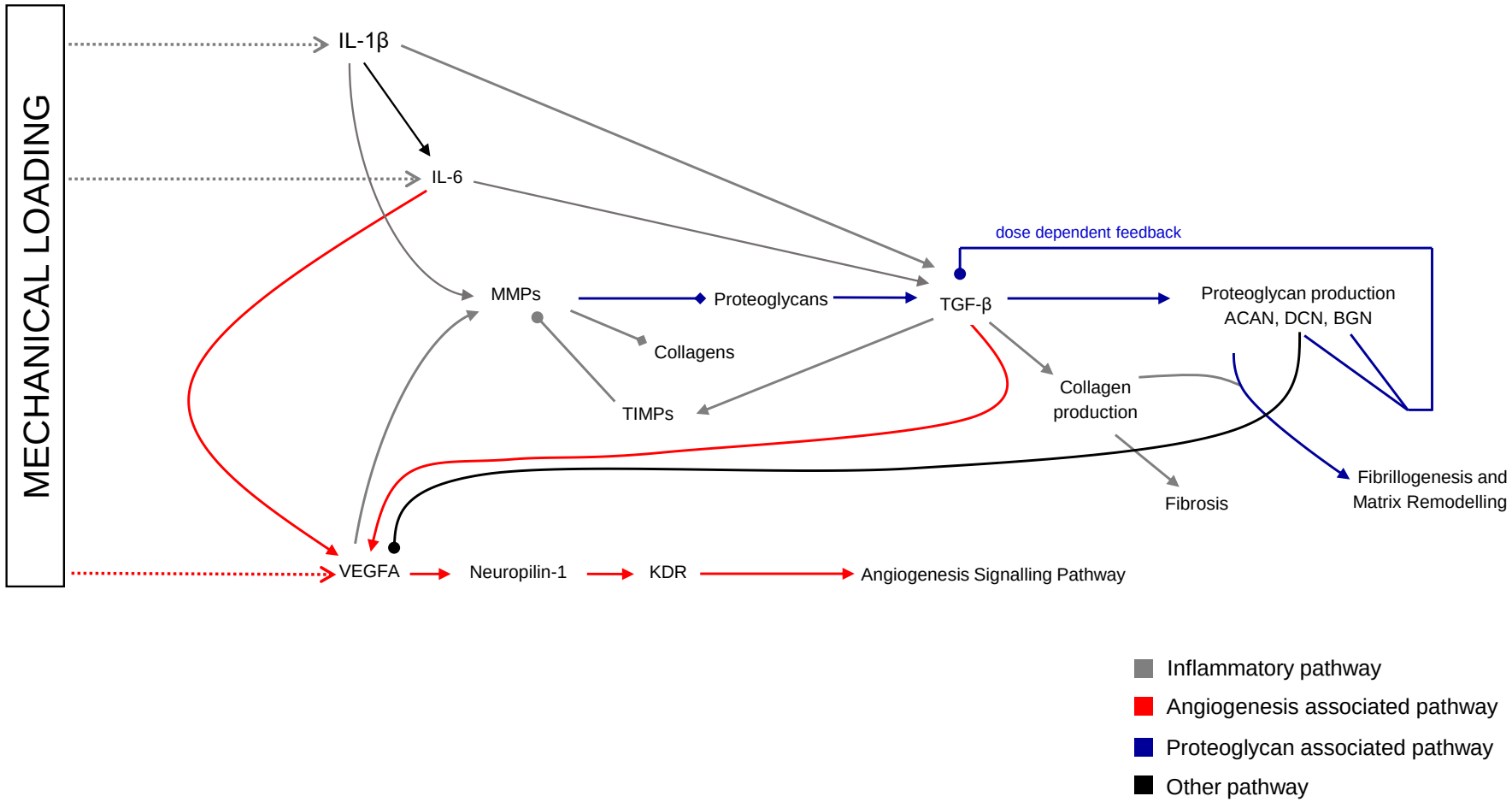


Figure 1.7: Schematic representation of the interconnected pathways, and their respective ECM components in extracellular matrix remodelling. A pointed arrowhead represents activation/upregulation; a circular arrowhead displays inhibition/downregulation, and degradation with a diamond arrowhead. Dotted lines indicate primary signalling factors. Diagram adapted from Rahim et al. (2014).

1.11 SUMMARY OF THE GENETIC APPROACHES FOLLOWED: PAST AND PRESENT

The primary aim of past and current research has involved a multidisciplinary approach to investigate the molecular and biological mechanisms underpinning an individual's genetic predisposition to ACL rupture. As reviewed in the previous section, a case-control candidate gene association study approach has predominately been used to date to implicate over 80 genetic loci with the risk of ACL rupture, and other exercise-related phenotypes. Genes involved in the biological function of connective tissue, i.e. structural, cell signalling and remodelling components of tendon and ligament have been the focus of these studies. However, most of these studies have explored candidate genes in single populations, which as mentioned are limited by small sample sizes, and hence underpowered to accurately detect associations, specifically with rare variants. Further to that, the findings are often contradictory between different populations, with alternate risk alleles modulating risk. To date, only the *COL1A1* rs1800012 polymorphism has consistently been associated with the risk of musculoskeletal injury, confirmed in a large combined analysis previously described in **Section 1.9.1**.

1.11.1 Application of New Technologies

In addition to being underpowered, using a candidate gene approach overlooks regions with potential risk modulating effects in other protein coding regions, as well as deep intronic variants in non-coding regions of the genome. Thus moving forward, employing a hypothesis-free approach through new technologies such as genome-wide association studies (GWAS) and next generation sequencing (NGS) methods, are more advantageous. As these methods provide greater coverage of the potential genomic regions involved in musculoskeletal injury

pathologies, they therefore hold promise in improving our understanding of the missing heritability.

To date, a few studies have employed GWAS technologies to investigate the genetic contribution predisposing to ACL rupture risk. Canine cranial cruciate ligament (CCL) rupture is similar in characteristics and risk factors to human ACL rupture (Baker et al., 2018). A GWAS in 96 Newfoundland dogs identified SNPs in three main regions of interest on chromosome 1 (*RNF152* gene) chromosome 3 (*SORCS2* gene) and chromosome 33 (*SEMA5B*, *DIRC2* and *ZDHHC23* genes) to associate with CCL rupture risk (Baird et al., 2014). The function of these genes and others identified in the study indicates a potential role for neurological pathways in the ACL rupture risk profile (Baird et al., 2014). As the CCL contains many mechanoreceptors (Schultz et al., 1984), altered neuronal signalling may negatively influence mechanotransduction, resulting in reduced proprioception and potentially an increased risk of injury (Baird et al., 2014).

In Labrador Retrievers, a GWAS identified 129 SNPs residing in 99 loci to associate with CCL rupture (Baker et al., 2017). Out of the 99, one loci containing nine genes (*BPI, LBP, RALGAPB, SLC32A1, ADIG, ACTR5, PPP1R16B, FAM83D, DHX35*) with various functions mapping to chromosome 24, met genome-wide significance. Interestingly, two loci were located in non-coding regions on the genome, containing regulatory polymorphisms (rSNPs). Complex phenotypes are caused by alterations in biological networks and not single genes, and rSNPs are able to influence gene expression through various mechanisms, including DNA methylation and long non-coding RNAs (lncRNA) (Schierding et al., 2014). Using a multivariate genome wide association approach, the same authors (Baker et al., 2018) identified three additional loci on chromosome 1, 4 and 23 to associate with CCL risk in the same group

of Labrador Retrievers. These loci reside in *ROR2* (cartilage and bone development), *DOCK2* (immune cell migration) and near a lncRNA. LncRNAs are a subset of non-coding RNAs (RNAs without protein-coding functions) that have gained much interest and attention due to their important role in regulating transcripts during biological and pathological processes (Mattick, 2004).

In human ACL rupture risk, Kim et al. (2017) performed genome-wide association screens for ACL rupture and Achilles tendinopathy/rupture using data from the Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort consisting of 102,979 ethnically diverse individuals. The aims of the study were to identify genetic markers that might be used to predict these injuries, and furthermore to analyse previously associated polymorphisms in the study cohort. Although the authors did not identify any single nucleotide polymorphism with genome wide significance ($p < 5 \times 10^{-8}$) four (rs1937810, rs57104447, rs57224706, rs60713544) and three (rs4067493, rs113435565, rs11960097) SNPs with borderline significance ($p < 10^{-6}$) were shown to be associated with AT and ACL rupture risk, respectively (Kim et al., 2017). Although there is little information on the rs57104447, rs60713544, rs113435565, and rs11960097 polymorphisms, rs1937810, rs57224706 and rs4067493 are intronic polymorphisms mapped to the *MMP7*, *SMARCD1*, and *DPP6* genes respectively [Database of Single Nucleotide Polymorphisms (dbSNP)]. Polymorphisms in *MMP7* are associated with variation in bone mineral density, *SMARCD1* is linked to chromatin organization pathways, and *DPP6* promotes cell surface expression and gating characteristics of voltage-gated potassium channels (Stelzer et al., 2016). Furthermore, moderate to weak evidence for the previously implicated *MIR608* rs4919510 ($p = 5.1 \times 10^{-3}$) polymorphism in Achilles tendon injury risk, was also observed.

The strengths of the GWA study included the investigation of a number of SNPs, in a large ethnically diverse sampling group. However, several limitations were recognised by the authors to suggest possible reasons for the inability to replicate previous candidate gene studies. Firstly, no information on participant activity level/frequency, or type of sport was reported, which is an established risk factor for ACL rupture and Achilles tendinopathy. Secondly, the Achilles tendon injury cases may have been misdiagnosed, and further to that, no differentiation between Achilles tendinopathy, bursitis or rupture was made (Kim et al., 2017). As current research suggests there may be potential differences in the underlying genetic mechanisms contributing to chronic (Achilles tendinopathy) versus acute (ACL and AT rupture) injuries (Gibbon et al., 2020) it is vital to stratify these injuries into separate groups for analyses. Recently, authors from the same research group (Kim et al., 2021) recognised the small sample size in Kim et al. (2017) as an additional limitation, and performed a meta-analysis using the combined data from two GWA studies (2,214 cases and 519,869 controls) to investigate the genetic contribution to ACL and posterior cruciate ligament (PCL) injury. Three novel loci (*INHBA*, *AEBP2*, and *LOC101927869*) that met genome significance were found to associate with ACL and PCL injury.

Furthermore, out of eight candidate genes previously implicated, the *COL3A1* rs1800255 polymorphism was significantly associated with combined ACL/PCL injury risk. INHBA (Inhibin Subunit Beta A) plays a role in osteoblastic cell growth during bone development, and therefore suggests a potential role in bone growth variation, which may contribute to ligament injury risk (Kim et al., 2021). *AEBP2* encodes for AE Binding Protein 2, which acts as an accessory subunit, facilitating histone trimethylation, leading to transcriptional control of an affected target gene (Stelzer et al., 2016). Currently, a proposed mechanism by which both

AEBP2 and *LOC101927869* (uncharacterised) alter ligament injury risk, is unknown (Kim et al., 2021).

Although GWAS's are increasingly useful in identifying regions of the genome potentially involved in modulating altered risk of musculoskeletal injury, they are limited in that they only examine specific regions of interest, while excluding others that may contribute to risk. Therefore, the movement to NGS technologies such as whole exome sequencing (WES) and whole genome sequencing (WGS) which encompasses the entire genome, and allows for greater depth of exploration has ensued. WES entails the enrichment and sequencing of all exons (protein encoding regions) in the genome. Accounting for only 1% of the genome, exons reportedly contain approximately 85% of mutations in the human genome (Glass and Nuara, 2013).

Applying a whole exome sequencing approach, a novel copy number variation (CNV) deletion region on chr6:109466479-109485174, corresponding to *CEP57L1A* was identified in a mother and her daughter, both affected by ACL/PCL agnesia (absence of the ACL/PCL). In a family of twin sibling males surgically diagnosed with ACL non-contact rupture, WES was performed on affected twins and non-affected progenitors (Caso et al., 2016). For SNPs, small insertions and deletions (indels) detection, variant filtering and to prioritize polymorphisms associated with ACL rupture, WES previously carried out on 16 healthy unaffected controls was also included in the study. Shared SNPs among family members obtained by family trios and duos were explored, and the findings indicated family members shared a set of 11 new variants identified in *SerpinA11, ARSI, NOCHT4, EPB41, FDFT1, POMC, KIF26A, OLFML2B, ATG7, FAH* and *WDR6* (Caso et al., 2016).

Furthermore, Gibbon et al. (2018) utilised a WES platform and a tiered filtering strategy of exemplar cases and control samples, to explore the genetic signature of Achilles tendinopathy. Two polymorphisms within the *TNC* gene were found to associate with the risk of Achilles tendinopathy (rs1061494) and ACL rupture (rs2104772) (Gibbon et al., 2018). The authors of this study recognised the sample size limitation, in that although a highly selected groups of exemplar cases and controls were used, the small number of sequenced individuals may have resulted in risk modifiers being undetected. Furthermore, the size of the Achilles tendinopathy group and the female ACL sub-groups were also a limitation.

Although WES has provided useful insight into the protein encoding regions of the genome, by only sequencing the exome, the promoter and exon/intron boundaries, and non-coding regions of the genome are not included. Many of the previously implicated genetic loci in soft tissue injury predisposition are located in such regions, and therefore the potential effects of variants in intronic and regulatory regions are overlooked using WES. However, despite the limitations, new advances in technology such as GWAS and WES, which draw on an unbiased approach to investigating the contribution of genetic variation in musculoskeletal injury phenotypes, have highlighted biological pathways not previously explored in candidate-gene studies. For example, cell communication and neurological pathways, in addition to long non-coding regions of the genome are new areas of interest.

There is therefore great value in employing NGS techniques to open up a spectrum of biological pathways to further explore. From the vast array of literature to date, it is clear that there is immense complexity between the potential interplay of risk factors contributing to an ACL rupture. With various scenarios and external influences affecting the observed emergent pattern, and the subsequent weighting of each determinant (Bittencourt et al., 2016). As described, a

multitude of genetic loci has been implicated in the risk profile of soft tissue injuries. However, how these observed genetic associations connect and interact is not yet known. For example, the current working hypothesis of two collagen-encoding genes, *COL1A1* and *COL5A1* describe a plausible theory as to how polymorphisms in these regions potentially influence ligament biology. However, there remains much paucity in identifying a comprehensive biological predisposition passport for all implicated loci.

The current work therefore aims to build on previous research, and continue exploring the genome to improve on the current understanding of the genetic factors contributing to an individual's predisposition to ACL rupture. By harnessing a whole genome sequencing approach, this thesis aimed to provide a more complete profile of the entire coding and non-coding regions of the genome contributing to ACL rupture predisposition. The research gleaned from the analyses has potential application towards a future clinical setting where it may serve to assist clinicians in developing multifactorial models to identify individuals at an increased risk of ACL rupture, and personalized pre- and rehabilitation programs. Additionlly, by identifying novel candidate genes through WGS techniques, a new spectrum of genes that play a role in pathophysiological pathways important in ACL rupture can be explored, and may potentially provide information for the design of new therapeutics.

1.12 AIMS AND OBJECTIVES OF THE RESEARCH

The primary aim of this thesis was to use a whole genome sequencing approach within the context of a twin family study to identify novel or previously implicated genetic loci contributing to ACL rupture predisposition. This aim was achieved through the following objectives (**Chapter 2**):

- Identified and prioritised predicted pathogenic or potential modifiable variants predisposing to ACL rupture.
- Identified and prioritised genetic sequences or genetic intervals common between affected family members within each family and/or affected members between families.
- Identified and prioritised functional partners for further genetic analyses in large cohorts.

The second aim of this thesis was to explore the identified and prioritised functional partners implicated through the WGS analyses in **Chapter 2,** and previously associated, independently and as a collective, with ACL rupture predisposition in a large combined ACL rupture dataset. For this reason, genes in the angiogenesis signalling pathway (**Chapter 3**) and the proteoglycan pathways (**Chapter 4**) were explored with non-contact ACL rupture risk in a combined cohort, comprised of individuals from three independent populations of different ancestry. To identify if there was a genetic predisposition profile between components of the angiogenesis and proteoglycan encoding genes.

Chapter 3 explored the following objectives:

The investigation of five functional polymorphisms within angiogenesis genes *VEGFA* rs699947 C/A, *VEGFA* rs1570360 G/A, *VEGFA* rs2010963 G/C, *KDR* rs2071559 A/G, and *KDR* rs1870377 T/A with non-contact ACL rupture risk using a candidate gene, case-control genetic association study design in participants from Sweden, Poland and Australia.

The investigation of *VEGFA* rs699947 C/A, *VEGFA* rs1570360 G/A, *VEGFA* rs2010963 G/C, *KDR* rs2071559 A/G, and *KDR* rs1870377 T/A with non-contact ACL rupture risk in a combined cohort comprising pooled genotyping data from the independent cohorts above (Sweden, Poland and Australia) and previously published cohorts from South Africa (Rahim et al., 2014) and Poland (Lulinska-Kuklik et al., 2019a).

The investigation of allele interactions between *VEGFA* (rs699947 C/A, rs1570360 G/A, rs2010963 G/C) and *KDR* (rs2071559 A/G, rs1870377 T/A) polymorphisms, as a proxy for gene-gene interactions with risk of non-contact ACL rupture.

Chapter 4 explored the following objectives:

The investigation of six polymorphisms within proteoglycan genes *ACAN* rs2351491 C/T, *ACAN* rs1042631 T/C, *ACAN* rs1516797 T/G, *BGN* rs1042103 C/T, *BGN* rs1126499 G/A and *DCN* rs516115 T/C with non-contact ACL rupture risk using a candidate gene, case-control genetic association study design on participants from Sweden and Poland.

The investigation of *ACAN* rs2351491 C/T, *ACAN* rs1042631 T/C, *ACAN* rs1516797 T/G, *BGN* rs1042103 C/T, *BGN* rs1126499 G/A and *DCN* rs516115 T/C with non-contact ACL rupture risk in a combined cohort comprising pooled genotyping data from independent cohorts above (Sweden and Poland) and previously published cohorts from South Africa (Mannion et al., 2014) and Poland (Cięszczyk et al., 2017).

The investigation of allele interactions between *ACAN* (rs2351491 C/T, rs1042631 T/C, rs1516797 T/G), *BGN* (rs1042103 C/T, rs1126499 G/A) and *DCN* (rs516115 T/C) polymorphisms, in addition to *VEGFA* (rs699947 C/A, rs1570360 G/A, rs2010963 G/C) and *DCN* (rs516115 T/C) polymorphisms, as a proxy for gene-gene interactions with risk of non-contact ACL rupture.

CHAPTER 2

A WHOLE GENOME SEQUENCING APPROACH TO ANTERIOR CRUCIATE LIGAMENT RUPTURE – A FAMILY "TWIN" STUDY

2.1 INTRODUCTION

As reviewed in **Section 1.7** of Chapter 1, numerous intrinsic risk factors have been implicated in ACL rupture predisposition, with a growing body of evidence now supporting the genetic contribution (Kaynak et al., 2017). Previous research implicating genetic polymorphisms in ACL rupture predisposition, have focused primarily on a case-control candidate gene association approach, with the majority of implicated loci localising to the genes encoding for structural components of the ECM, angiogenesis associated signalling molecules, and regulators of the extracellular matrix (**Chapter 1, Section 1.9**). However, as previously mentioned there is inconsistency in the associated risk loci between small studies in single populations, therefore making it difficult to establish true associations with the potential risk loci (**Chapter 1, Section 1.9**). Furthermore, the recent estimated ~69% heritability component for ACL rupture (Magnusson et al., 2020) suggests that a large heritability component remains unexplored. To some extent, GWAS and NGS technologies have begun to increase our understanding of the genetics underlying ACL rupture predisposition (**Chaper 1, Section 1.11**) but there remains much paucity in the application of these technologies in identifying the full spectrum of genetic loci predisposing to ACL rupture risk.

Therefore, to build on previous research, the study presented in this chapter aimed to employ a WGS approach in two unrelated twin families with a history of ACL ruptures, to identify novel and/or previously implicated biological genomic signatures with predisposition to ACL rupture. To achieve this aim, the main objectives included (i) identifying predicted pathogenic or potential modifiable variants in ACL rupture predisposition, (ii) identifying genetic sequences or genetic intervals common between affected members within families, or affected members between families, and (iii) identifying and prioritise functional partners for further genetic analyses in a large, collective cohort.

2.2 METHODS

2.2.1 Ethics Statement

The study recruited two South African families of Dutch/Irish (Family A) and English/Scottish (Family B) descent. All participants completed questionnaires regarding personal details, medical and sporting history, and family history of ligament and other musculoskeletal soft tissue injuries (**Appendix A, Section 2**). Written informed consent was obtained from all the participants according to the declaration of Helsinki (**Appendix A, Section 2**) and the study was approved by the Research Ethics Committee of the Faculty of Health Sciences within the University of Cape Town, South Africa (HREC 823/2017) (**Appendix A, Section 1**). All reported ACL ruptures were confirmed clinically and radiologically.

2.2.2 Participant Characteristics and Clinical Descriptors

Family A monozygotic twin males (Twin 1 and Twin 2) were recruited at 33 years old. Both individuals sustained two unilateral non-contact ACL ruptures of the right limb. Twin 1 at age 20 and 27 years, and Twin 2 at age 28 and 32 years (**Figure 2.1**). All ACL ruptures occurred while playing club level touch rugby and the number of years of participation were 17 years and 12 years for Twin 1 and 2, respectively. Both individuals had a previous history of other ligament or tendon injury (**Table 2.1**). Twin 1 had sustained injuries to the lateral ligament of the right ankle, and the extensor tendon of the wrist. Twin 2 had sustained an injury to the medial ligament of the right ankle. Their male sibling was recruited at 41 years old and had sustained bilateral non-contact ACL ruptures at 12 and 31 years of age, both while playing rugby union. The number of collective years of participation was 10 years, and the sibling also had a previous history of bilateral injury to the tibialis posterior tendon. The father, recruited at 61 years, had sustained a unilateral non-contact ACL rupture on the right limb at the age of 30 years, while playing amateur level football (38 years of participation). No history of previous other ligament or tendon injuries were recorded. The mother, recruited at age 58, reported participating in dancing for over 10 years at a social level, with no previous ligament or tendon injuries.

Family B monozygotic twin males (Twin 3 and Twin 4) were recruited at the age of 29 years. Twin 3 had sustained a unilateral non-contact ACL rupture of the right limb at the age of 27 years, and Twin 4 three non-contact ACL ruptures (two in right limb, and one in left limb) at ages 21, 25, and 27 years (**Figure 2.1**). Both individuals ruptured their ACL playing provincial level rugby union, which they had participated in for 10 years, and both reported a history of previous other ligament or tendon injury (**Table 2.1**). Twin 3 had sustained injuries to the lateral

ligament of the right ankle, the right shoulder ligaments and the supraspinatus tendon of the right shoulder. Twin 4 had sustained an injury to the medial ligament of the right ankle. Their female sibling and mother (31 and 58 years at recruitment respectively) had no previous history of ligament or tendon injury. The female sibling had participated in karate and swimming activities (10 and 26 years of participation, respectively) and the mother in hockey (4 years) horse riding (15 years) and swimming (8 years of participation). The father was deceased at the time of recruitment, and a medical history was therefore unavailable. However, the family reported no known history of any ligament or tendon injury.

2.2.3 Whole Genome Sequencing

Genomic DNA was isolated from venous blood according to a previously described standard protocol (Lahiri and Nurnberger, 1991) with slight modifications (Mokone et al., 2005) and prepared according to the requirements of the WGS service provider (BGI Genomics, Hong Kong). All samples passed quality control measures stipulated by BGI Genomics. In summary, for each sample the isolated DNA was fragmented, and the fragments selected by Agencourt AMPure XP-Medium kit to an average size of 200 - 400 bp. Fragments were end repaired and 3’ adenylated, and adaptor sequences ligated to the 3’ adenylated fragments. Fragments were then amplified by PCR and underwent a purification step followed by heat denaturation and circularized. Single stranded circular (ssCir) DNA formatted to a library were qualified by quality control measures and sequenced using the BGISEQ-500 at 30X coverage. ssCir DNA molecule nanoballs were loaded into a patterned nanoarray using high density DNA nanochip technology, and pair end 100 bp reads obtained by combinatorial Probe-Anchor Synthesis (cPAS).

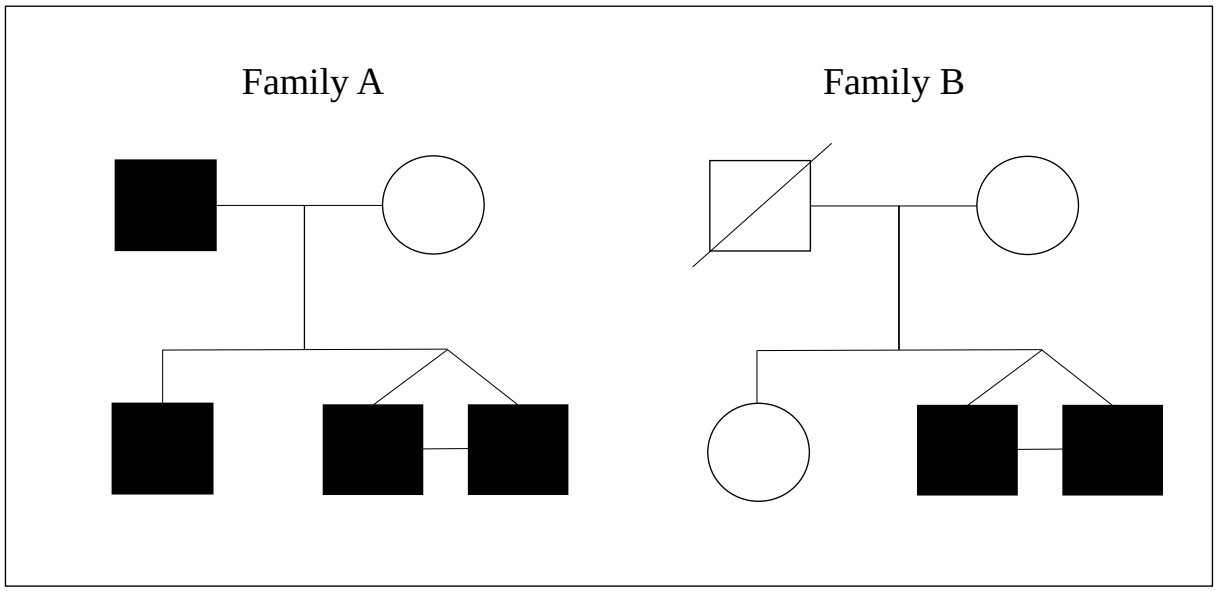


Figure 2.1: Pedigree structure for Family A and Family B with known history of diagnosed non-contact anterior cruciate ligament (ACL) rupture. Circles indicate females, squares, males. Filled symbols indicate participants clinically and radiologically diagnosed with non-contact anterior cruciate ligament rupture, and open symbols uninjured participants. Symbols with a line through them indicate deceased individuals.

Table 2.1: History of other musculoskeletal soft tissue injuries (ligament and tendon) in family members of Family A and Family B

Family	Family Member	History of other injuries (not ACL)	Injury Type
A	Twin	Yes	Lateral ligaments of the right ankle, and wrist extensor tendon
	Twin	Yes	Medial ligaments of the right ankle
	Brother	Yes	Left and right tibialis posterior tendon
	Father	No	-
	Mother	No	-
B	Twin	Yes	Medial ligaments of right ankle
	Twin	Yes	Lateral ligaments of the right ankle, and supraspinatus tendon of right shoulder
	Sister	No	-
	Mother	No	-

2.2.4 Variant Calling

The Burrows-Wheeler Alignment tool (Li and Durbin, 2009; Li et al., 2008) was utilised to reconstruct reads by alignment against the complete human reference genome build hg38. Post alignment was performed using the Picard tool kit (McKenna et al., 2010). This process consisted of sorting, marking duplication reads, and the BAM files were sorted by coordinates, indexed, and read groups. Additionally, read pair information was recalculated to observe any changes by leveraging Picard FixMate Information. Bcftools (Danecek and McCarthy, 2017; Narasimhan et al., 2016) was applied to create a clean version of the BAM files. **Appendix B, Section 1, Figure S1** shows the overall quality control of the bam files. Current variant calling approaches have differing advantages (DePristo et al., 2011; Liu et al., 2012; Liu et al., 2013) and may produce differing variant calls. Here we considered an ensemble approach implemented in VariantMetaCaller (Gézsi et al., 2015) that may find a call consensus in detecting SNPs and short indels.

The information generated from two independent variant caller pipelines: (1) An incremental joint variant discovery implemented in GATK 3.0 HaplotypeCaller (McKenna et al., 2010) which calls samples independently to produce gVCF files and leverages the information from the independent gVCF file to produce a call-set at the genotyping step; (2) bcftools via mpileup (Danecek and McCarthy, 2017; Garrison and Marth, 2012; Narasimhan et al., 2016) was performed to produce an additional genotyped call-set. The best practice specific to each caller was adopted (Cornish and Guda, 2015). Before applying the ensemble approach from the resulting variant sets from the callers above, each resulting VCF file was filtered using the GATK 3.0 tool Variant Filtration. The final call-set was produced from VariantMetaCaller

(Gézsi et al., 2015) a support vector machines approach that combines the hard-filtered VCF files obtained from these 2 variant callers.

2.2.5 Annotation, In Silico Prediction of Mutation and Prioritization

ANNOVAR (Wang et al., 2010) was used to perform gene-based annotation to detect whether the SNPs detected resulted in protein coding changes and to produce a list of the affected amino acids. Population frequency and pathogenicity for each variant was obtained from 1000 Genomes exome (1000 Genomes Project Consortium, 2012), Exome Aggregation Consortium (ExAC) (Karczewski et al., 2017) targeted exon datasets and COSMIC (Forbes et al., 2015). Gene functions were obtained from RefGene (O'Neill et al., 2016) and different functional predictions were obtained from ANNOVAR's library, which contains up to 21 different functional scores including SIFT (Ng and Henikoff, 2003; Ng and Henikoff, 2006) LRT (Fujita et al., 2011) MutationTaster (Lubeck et al., 2012) MutationAssessor (Reva et al., 2007; Reva et al., 2011) FATHMM (Shihab et al., 2014) fathmm-MKL (Shihab et al., 2014) RadialSVM (Dong et al., 2015) LR (Dong et al., 2015) PROVEAN (Dong et al., 2015) MetaSVM (Dong et al., 2015) MetaLR (Dong et al., 2015) CADD (Kircher et al., 2014) GERP++ (Cooper et al., 2005) DANN, M-CAP, Eigen, GenoCanyon, Polyphen2 HVAR (Garber et al., 2009) Polyphen2 HDIV (Adzhubei et al., 2010) PhyloP (Garber et al., 2009) and SiPhy (Garber et al., 2009). Additionally, conservative, and segmental duplication sites, dbSNP code and clinical relevance reported in dbSNP (Sherry et al., 2001) were also included. From the resulting functional annotated dataset, we first filtered variants for rarity, exonic variants, non-synonymous, stop codons, predicted functional significance and deleteriousness (Ng and Henikoff, 2003; Ng and Henikoff, 2006). The resulting functional annotated data set was independently filtered for predicted functional status (of which each predicted functional status

is of "deleterious" (D), "probably damaging" (D), "disease_causing_automatic" (A) or "disease_causing" (D). (Li et al., 2013; Sifrim et al., 2013; Wheeler et al., 2007) from these 21 in silico prediction mutation tools. A casting vote approach was utilized to retain only a variant if it had at least 17 predicted functional status “D” or “A” out of 21. The retained variants were further filtered for rarity, exonic variants, nonsynonymous mutations, yielding a final candidate list of predicted mutant and genetics modifier variants.

2.2.6 Phased and Haplotype Inference

Additional VCF files were accessed from the 1000 Genomes Project Consortium, 2015 and the African Genome Variation Project (AGVP) which recently characterized the admixture across 20 ethno-linguistic groups (**Appendix B, Section 1, Table S1**) from sub-Saharan Africa (Gurdasani et al., 2015). PLINK was used to carry out quality control on the VCF files, and in total 2,504 BAM files from 1000 Genomes Project and 2,428 from AGVP were retained. Based on the initial sample description (population or country labels), the population ethno-linguistic information was used to categorize the obtained data per ethnic group as described (Gudykunst and Schmidt, 1987; Michalopoulos, 2012). The family datasets were merged with the available data from 20 worldwide ethnic groups regardless of depth of coverage.

To increase the accuracy, the resulting VCF file containing 4,932 samples of 20 ethnic groups, was used to further conduct quality control in removing all structured, indel, multi-allelic variants and those with low minor allele frequency (MAF < 0.05) prior to phasing. From the resulting curated data, the haplotypes were first phased and inferred using Eagle (Loh et al., 2016). Further, sites discordance was compared between these haplotype panels and

independently with their original VCF file prior phasing. The only site with phase switch-errors showed discrepancies in MAF, and was removed.

2.2.7 Principal Component Analysis

Principal components analysis (PCA) is now routinely used to detect and quantify the genetic structure of populations. LD pruning was performed on the merged family dataset using PLINK to remove correlated ($r^2 > 0.15$) SNPs in a 1,000-SNP window, advancing by 10 SNPs at a time. The pruned dataset contained 9,487,525 SNPs and 9 individuals with a genotype call rate of 99.9%. VCF files of 2,504 samples were accessed from 1000 Genomes Project Consortium, 2015 and 2,428 samples from the African Genome Variation Project (AGVP) which has recently characterized the admixture across 18 ethno-linguistic groups from sub-Saharan Africa (Gurdasani et al., 2015). A quality control check on these VCF files was conducted using plink and vcftools. Based on the initial sample description (population or country labels), the population ethno-linguistic information was used to categorize the obtained data per ethnic group. The merged family datasets and the 20 ethnic groups' data set yielded 45,096 SNPs, in 4,941 samples in 20 ethnic groups. Smartpca was used to perform PCA on the pruned family dataset and the merged data set of 20 worldwide ethnic groups. Lastly, GENESIS (V0.2.6b) (http://www.bioinf.wits.ac.za/software/genesis) was used for plot visualizations.

2.2.8 Percentage Pathogenic Variants with Previously Implicated Genes

To evaluate the distribution of pathogenic variants across worldwide ethnics and the family dataset within genes known to be associated with the risk of ligament and tendon injuries (**Appendix B, Section 1, Table S2**), the dbSNP database was leveraged to extract SNPs

associated with these genes. The obtained SNPs were extracted from the merged phased data containing 4,941 samples from 20 ethnic groups and the family's data set. The resulting data was split into each ethnic group, and annotation using ANNOVAR were performed as described above. The fraction of pathogenic within a gene was approximated as the count of reported pathogenic SNPs from ANNOVAR, divided by the total count of associated SNPs to the gene from dbSNPs.

2.2.9 Distribution of Minor Allele Frequency and Gene-Specific SNP Frequencies

To examine the extent of how common SNPs are distributed across these 20 ethnic groups in genes known to be associated with the phenotype of interest, the distribution of the minor allele frequency was investigated. To this end, the proportion of minor alleles were categorized into 6 bins (0-0.05, > 0.05-0.1, > 0.1-0.2, > 0.2-0.3, > 0.3-0.4, > 0.4-0.5) with respect to each ethnic group with a disease. The minor allele frequency (MAF) per SNP for each category was computed using Plink software. Furthermore, the fraction of gene-specific SNP frequency for each gene was computed, assuming SNPs upstream and downstream within the associated gene region as annotated from dbSNP database. Minor allele frequency per SNP was aggregated at the gene level as from previous works (Wonkam et al., 2020).

2.2.10 Identity by Descent and Functional Genomics

Haplotypes are identical by descent if they are identical and inherited from a common ancestor. Tracts of identity by descent (IBD) are broken up by recombination during meiosis, so expected length of IBD depends on the number of generations since the common ancestor for the locus. If the common ancestor lived a great many generations ago, the individuals share very short

tracts of genetic material. Accurate estimation of genomic IBD sharing depends not only on detection of IBD tracts but also on accurate estimation of the ends of those tracts and modelling linkage disequilibrium. Here, the two-family data were used separately, to investigate the overall genomic identity-by-descent (IBD) sharing between pairs of individuals within each family and thus across families. The segments of IBD were obtained using the Refined IBD algorithm (Browning and Browning, 2013). The genomic IBD segments within families were evaluated and the shared segments between the two families compared. The potential functional roles of the genes localised to these shared IBDs and additional selected genes were explored using network and enrichment analysis, to gain insight into potential disease compromised networks. To explore functional partners, GeneCards (Stelzer et al., 2016) premium analytical tool; *GenesLikeMe*, was utilised, where a weighting of 3 (out of 5) was set for each attribute (sequence paralogs, domains, super pathways, expression patterns, phenotypes, compounds, disorders and gene ontologies). From the results, the top 100 inferred functional partners were further enriched for common sub-networks, pathways, biological processes, molecular functions, and association to potential human phenotypes using Enrichr software (Kuleshov et al., 2016). Finally, functional partners were summarised and grouped by protein function.

2.2.11 Network and Enrichment Analysis

From the retained final candidate list of predicted mutant and genetics modifier variants, how each set of the variant genes are layered and interact within a biological network was examined. This was carried out using a comprehensive human Protein-Protein Interaction (PPI) network to identify sub-networks of interactive mutant and genetic modifier variant genes (Warde-Farley et al., 2010; Montojo et al., 2014; Franz et al., 2018). Furthermore, Enrichr software

(Kuleshov et al., 2016) was again utilised to determine how these genes interact in sub-networks, their pathways, biological processes, molecular functions, and association to potential human phenotypes. The most significant pathway enriched for genes in the network were selected from KEGG (Kanehisa and Goto, 2000) Panther (Mi et al., 2016) Biocarta (Nishimura, 2001) and Reactome (Croft et al., 2013) and gene ontologies from the Gene Ontology Consortium (Reference Genome Group of the Gene Ontology Consortium, 2009) were defined for cellular component, biological process, and molecular function.

2.3 RESULTS

2.3.1 Variant Discovery

A total number of 10,560,616 variants were called in the whole genome sequence dataset, of which 1.13% were exonic (distributed as 5.2% nonframeshift deletion, 3% nonframeshift insertion, 19% frameshift deletion, 14% frameshift insertion, 31% nonsynonymous, 24% synonymous, 3% stopgain, 0,08% stoploss and 1.1% unknown), 0.35% ncRNA_exonic, 53% intergenic, 43% intronic, 0.03% splicing, 0.96% UTR3, 0.12% UTR5, 0.64% upstream, 0.66% downstream and 0.46% other.

2.3.2 Genetics Structure

From **Figure 2.2A**, the South African families of Dutch/Irish (Family A) and English/Scottish (Family B) descent were observed to form a cluster close to the European Cluster. Principal component analysis (PCA) of data from Families A and B, showed that the two families **(Figure 2.2B**) clustered separately from each other. Within Family A, the twins were very close

together, with their male sibling in close proximity. However, the father and mother positioned further away from all the progenitors. The individuals in Family B on the other hand, were more closely grouped, with the uninjured mother and female sibling in close proximity to the twins.

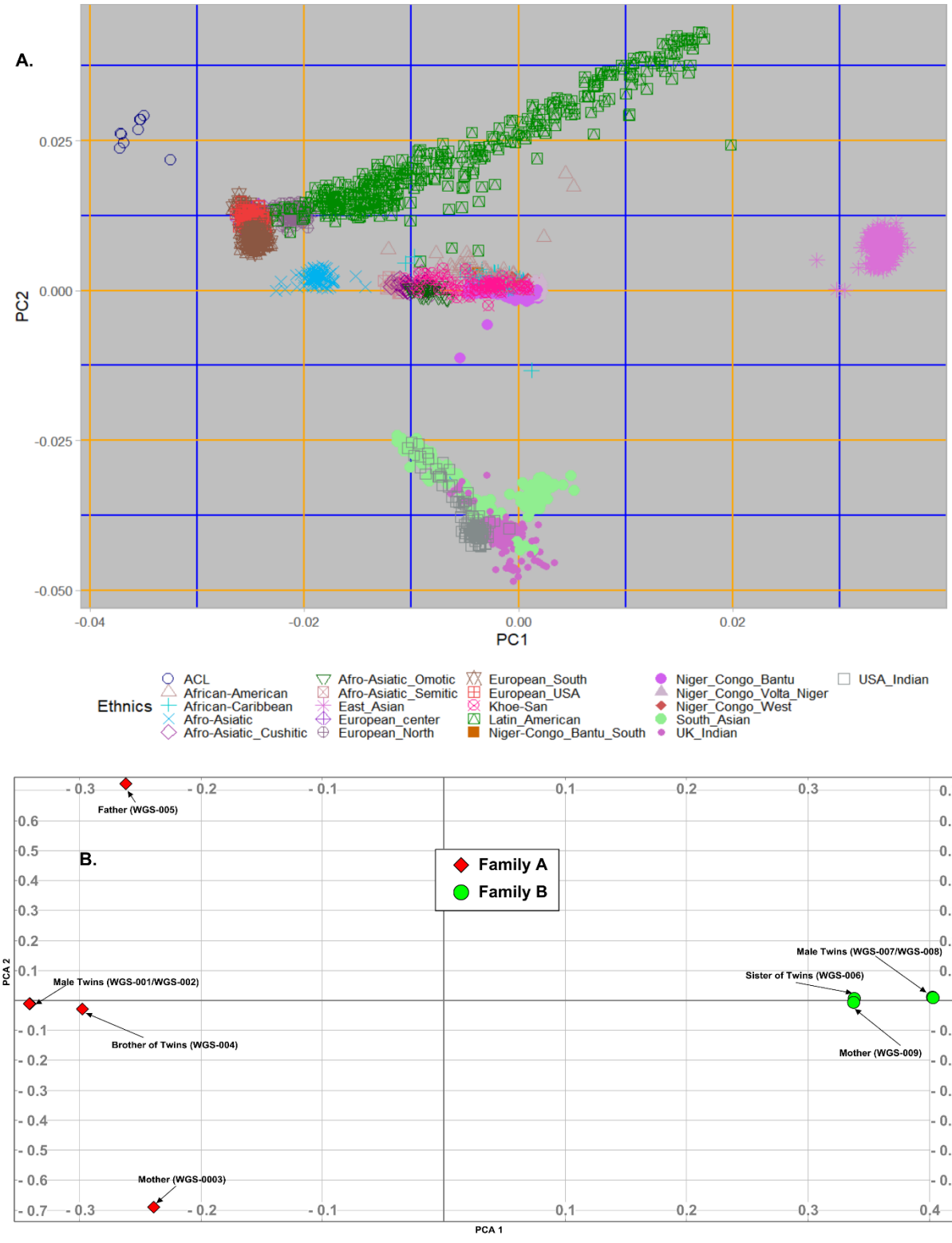


Figure 2.2: Principal component analysis on merged data sets of Family A and B. Family A is depicted in red diamonds, with Family B in green circles.

2.3.3 In-Silico Putative Deleterious Variants

Filtering mutations, 29 genes with predicted mutations were prioritized for Family A (**Table 2.2**) of which polymorphisms within two genes, namely *COL11A1* and *COL12A1* which encode for the α1 chains of types XII and XI collagen respectively, have previously been associated with ACL rupture. For Family B, a list of 18 genes (**Table 2.3**) including the previously associated *COL12A1*, were prioritized. Of the 47 genes prioritized (29 in Family A and 18 in Family B) three genes *COL12A1*, *CATSPER2* and *KCNJ12*, were common to both families (**Table 2.4**). *CATSPER2* and *KCNJ12*, which have not previously been investigated as candidate genes, both encode for ion channels associated with cation and potassium ion channels, respectively. Six non-synonymous SNPs were identified within these three genes, one each in *COL12A1* and *CATSPER2*, and four in *KCNJ12* (**Table 2.4**). From SIFT, PolyPhen-2 and FATHMM_pred prediction tools, the Gly3058Ser substitution at *COL12A1* rs970547 C>T, and the Arg511His substitution at *CATSPER2* rs144399798 C>T were predicted moderately damaging (CADD 20-30) with deleterious loss of function effect (**Table 2.4**). While the Glu139Lys, Gly145Ser, Arg261His, and Ile262Ser substitutions at *KCNJ12* rs76265595, rs75029097, rs77270326 and rs76684759 variants, respectively, were predicted strongly damaging (CADD > 30) with deleterious loss of function effect (**Table 2.4**). The glycine to serine substitutions at rs970547 and rs75029097 result in a change from a non-polar, aliphatic to a polar non-charged amino acid. For the arginine to histidine substitutions at rs144399798 and rs77270326, a basic polar amino acid is substituted for another basic polar amino acid, whereas for the glutamate to lysine substitution at rs76265595, an acidic polar amino acid is substituted for a basic amino acid. Lastly, the isoleucine to serine substitution at rs76684759, results in a non-polar, aliphatic amino acid change to a polar non-charged residue (Bouaoun et al., 2016).

Table 2.2: Candidate list of genes with predicted mutations in Family A

Gene	Family A	Variant	Chromosomal Location
ABCA13	ATP Binding Cassette Subfamily A Member 13	rs77147473	6p12.3
AK2	Adenylate Kinase 2	rs113711467	1p35.1
AK3	Adenylate Knase 3	rs763448502	8p24.1
CAPN2	Calpain 2	rs145466296	1q41
CATSPER2	**Cation Channel Sperm Associated 2**	**rs144399798**	**12q15.3**
COL11A1	Collagen Type XI Alpha 1 Chain	rs139064549	1p21.1
COL12A1	**Collagen Type XII Alpha 1 Chain**	**rs970547**	**6q14.1**
EYS	Eyes Shut Homolog	rs928941618	6q12
F13A1	Coagulation factor XIII A chain	rs3024477	6p25.1
F5	Coagulation Factor V	rs6027	1q24.2
GDF9	Growth Differentiation Factor 9	rs61754582	3q31.1
GUCY2C	Guanylate Cyclase 2C	rs1306674918	11p12.3
KCNJ12	**Potassium Inwardly Rectifying Channel Subfamily J Member 12**	**rs76265595**	**17p11.2**
		rs75029097	
		rs77270326	
		rs77270326	
KIF23	Kinesin Family Member 23	rs148511930	15q23
KMT2C	Lysine Methyltransferase 2C	rs4024419	7q36.1
MASP1	Mannan Binding Lectin Serine Peptidase 1	rs140933134	3q27.3
MLH1	MutL Homolog 1	rs35001569	2p22.2
MPZL3	Myelin Protein Zero Like 3	rs63750449	11q23.3
MYO18A	Myosin XVIIIA	rs142071163	17q11.2
NLRP14	NLR Family Pyrin Domain Containing 14	rs201591312	9p15.4
NPHP4	Nephrocystin 4	rs149541731	1p36.31
OVCH2	Ovochymase 2	rs35641267	11p15.4
PEX6	Peroxisomal Biogenesis Factor 6	rs61759818	5p21.1
PMM1	Phosphomannomutase 1	rs577794661	17q13.2
PNKD	Metallo-Beta-Lactamase Domain Containing	rs750714549	1q35
SI	Sucrase-Isomaltase	rs147259983	3q26.1
SLC7A2	Solute Carrier Family 7 Member 2	rs138434001	7p22
SPG11	SPG11 Vesicle Trafficking Associated, Spatacsin	rs140440674	15q21.1
TTC38	Tetratricopeptide Repeat Domain 38	rs78183930	22q13.31

Genes in **bold** depict predicted pathogenic mutations shared in Family A and B

Table 2.3: Candidate list of genes with predicted mutations in Family B

Gene	Family B	Variant	Chromosomal Location
ABCC1	ATP Binding Cassette Subfamily C Member 1	rs45511401	16p13.11
ABCD4	ATP Binding Cassette Subfamily D Member 4	rs45568335	14q24.3
ACAD9	Acyl-CoA Dehydrogenase Family Member 9	rs115532916	3q21.3
ATP10A	ATPase Phospholipid Transporting 10A	rs140982945	15q12
CATSPER2	**Cation Channel Sperm Associated 2**	**rs144399798**	**15q15.3**
COL12A1	**Collagen Type XII Alpha 1 Chain**	**rs970547**	**6q14.1**
DHODH	Dihydroorotate Dehydrogenase (Quinone)	rs61733129	16q22.2
GKN2	Gastrokine 2	rs62133344	2p13.3
GMPR2	Guanosine Monophosphate Reductase 2	rs34354104	14q12
HNF4G	Hepatocyte Nuclear Factor 4 Gamma	unknown	8q21.13
INHA	Inhibin Subunit Alpha - Encodes a member of the TGFB family	rs139051234	2q35
IPP	Intracisternal A Particle-Promoted Polypeptide	rs142095376	1p34.1
KCNJ12	**Potassium Inwardly Rectifying Channel Subfamily J Member 12**	**rs76265595**	**17p11.2**
		rs75029097	
		rs77270326	
		rs77270326	
PKHD1L1	PKHD1 Like 1	rs75029097	8q23.1
PMS2	PMS1 Homolog 2, Mismatch Repair System Component	rs77270326	7p22.1
RTEL1	Regulator Of Telomere Elongation Helicase 1	rs76684759	20q13.33
SEC24D	SEC24 Homolog D, COPII Coat Complex Component	rs185023340	4q26
SLC26A7	Solute Carrier Family 26 Member 7	rs63750123	8q21.3

Genes in **bold** depict predicted pathogenic mutations shared in Family A and B

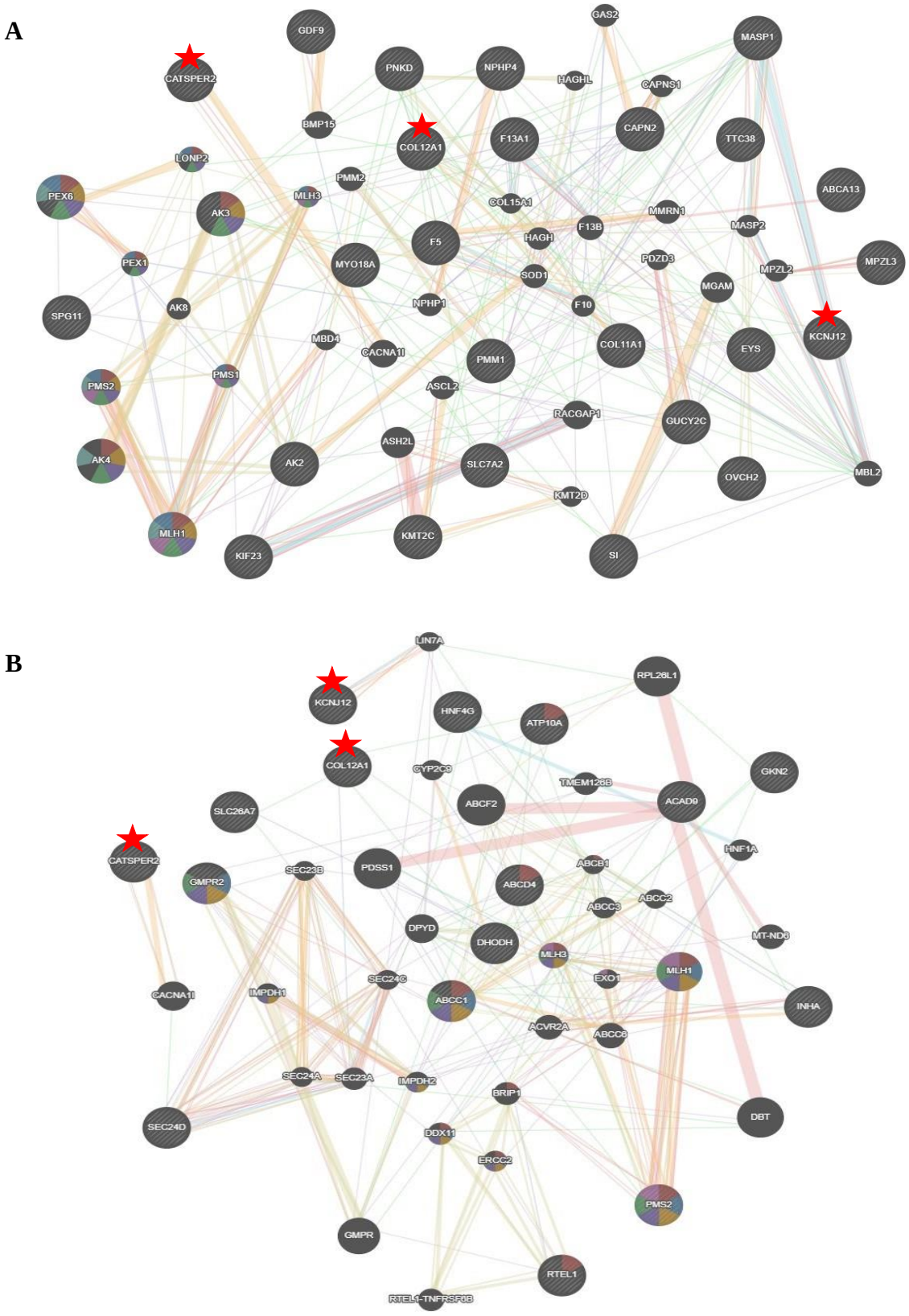
Table 2.4: Predicted functional effect of mutations common to Family A and Family B

Number of Participants	Genotype	Gene	Region	SNP	Protein Change	Functional effect£	ExAC AFR	ExAC EUR
(9)	Homozygous	*COL12A1*	6q14.1	rs970547 C>T	Gly3058Ser	Loss of function of growth plate cartilage chondrocyte morphogenesis	0.0026	0.73
(9)	Homozygous	*CATSPER2*	15q15.3	rs144399798 C>T	Arg511His	Loss of function in ion channel activity and voltage-gated ion channel activity.	0.0083	0.0003
(9)	Homozygous	*KCNJ12*	17p11.2	rs76265595 G>A	Glu139Lys	Loss of function of inward-rectifier potassium channels	0	0.71
				rs75029097 G>A	Gly145Ser	Loss of function of inward-rectifier potassium channels	0	0.32
				rs77270326 G>A	Arg261His	Loss of function of inward-rectifier potassium channels	0.001	0.11
				rs76684759 T>G	Ile262Ser	Loss of function of inward-rectifier potassium channels	0.002	0.017

2.3.4 Pathways and Biological Processes Associated with Genes with Mutational Burdens

To determine potential interactions and functional pathways for the prioritized candidate genes in **Table 2.2 and 2.3**, an interaction network between the predicted pathogenic and genetic modifier variants for Family A and B was generated (**Figures 2.3A-B**). Physical interactions, co-expression, predicted, co-localization, pathways and genetic interactions are depicted in the figure. Furthermore, functions of the genes in the networks are distinguished by colour coding and classified according to metabolic process. Candidate gene sets (**Appendix B, Table S3**) were enriched in pathways relevant to ACL pathophysiology, including complement/coagulation cascades (p = 3.0e-7), purine metabolism (p = 6.0e-7) and mismatch repair (p = 6.9e-5) pathways. Gene-set previously associated with ligament and tendon injury (**Appendix B, Table S3**) were enriched in pathways including PI3K-Akt signalling pathway (p = 8.9e-11), protein digestion and absorption (p = 5.9e-10) and rheumatoid arthritis (p = 1.1e-6).

Additional enriched top significant ($p < 0.05$) pathways, biological processes, molecular function, and human phenotypes associated with the genes previously associated with ligament and tendon injury (**Appendix B, Table S3**) and the candidate list of predicted pathogenic genes and their interacting genes for Family A and Family B (**Appendix B, Table S2 and Figure 2.3**, respectively) are shown in **Appendix B, Table S3**.



See overleaf for figure legend

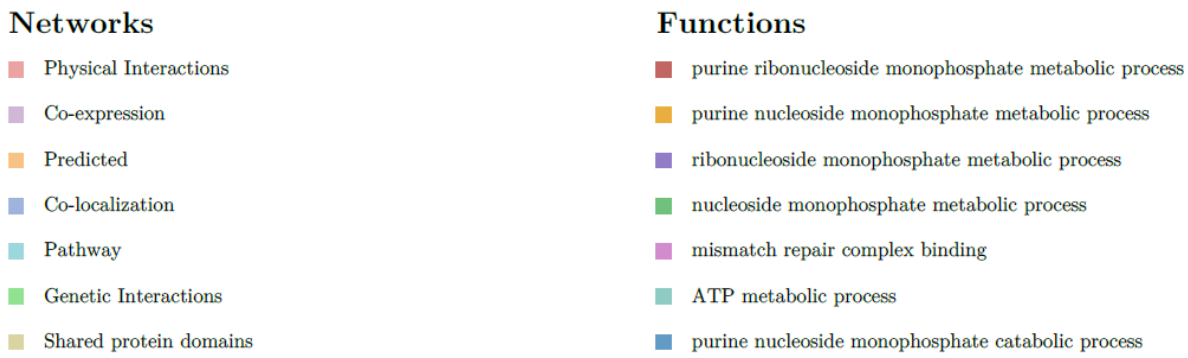


Figure 2.3: Biological sub-network of the candidate mutant genes with interacting genes in Family A (A) and Family B (B). Candidate genes common to Family A and B are highlighted by red stars. Weighting of interaction is depicted by line thickness. All query genes are given the maximum node size, and the size of related genes is inversely proportional to the rank of the gene based on its score assessed by GeneMANIA (Warde-Farley et al., 2010).

2.3.5 Distribution of Pathogenic variants, Minor Allele and Gene-Specific in SNPs Frequencies

The distribution of reported pathogenic polymorphisms across worldwide ethnics and the current family dataset was investigated in genes previously associated with risk of ligament and tendon injuries. *ADIPOQ, COL5A1, COL11A1, COL12A1, DEFB1, FBN2, ITGB3, LUM, MMP1* and *VEGFA* had a considerable higher proportion of pathogenic variants in both families compared to 20 other ethnic groups (**Figure 2.4A**). Moreover, for the *TGFB1, FBN2, VEGFA* and *MMP1* genes, Family B had a higher proportion of pathogenic variants, compared to 20 other ethnic groups and Family A (**Figure 2.4A**). Of these, *COL11A1* was prioritised with predicted mutations in Family A, and *COL12A1* with predicted mutations in both families. Of the 40 genes previously associated with ligament and tendon injury, 29 genes had a greater gene-specific SNP frequency when compared to the other 20 ethnic groups (**Figure 2.4B**). Notably, *CASP8, IL6, MMP8, COL1A1, MMP3, COL12A1* and *MMP1* genes had the highest gene-specific SNP frequency of > 1.6. Again, the previously prioritised *COL12A1* was highlighted, as one of the genes with the highest gene-specific SNP frequency (**Figure 2.4B**).

Furthermore, variation was observed in the distribution of MAF at rare and common variants between the South African families of Dutch/Irish (Family A) and English/Scottish (Family B) descent, and the rest of 20 worldwide ethnic groups (**Figure 2.5**). The South African families of Dutch/Irish (Family A) and English/Scottish (Family B) have a lower frequency of rare variants at MAF between 0.0 and 0.1, and higher frequency of common variants at MAF bin 0.1 - 0.5, in contrast to those from the 20 ethnic groups (**Figure 2.5**).

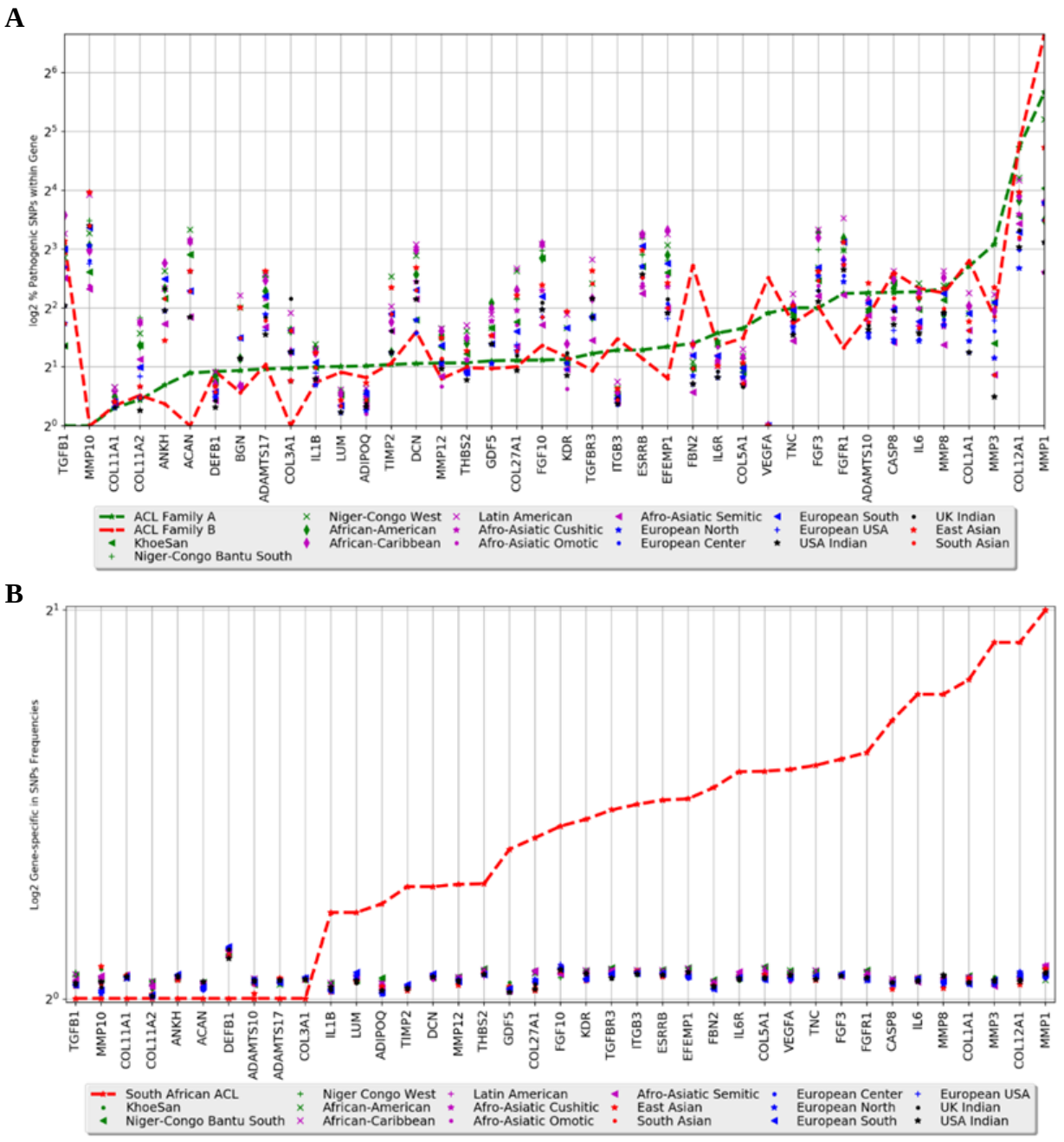


Figure 2.4: A: Gene-specific proportion of pathogenic SNPs across 20 ethnic groups, and Family A and B from 40 genes known to associate with ligament and tendon injury and B: Gene-specific SNPs minor allele frequencies across 20 ethnic groups from 40 genes previously associated with ligament and tendon injury.

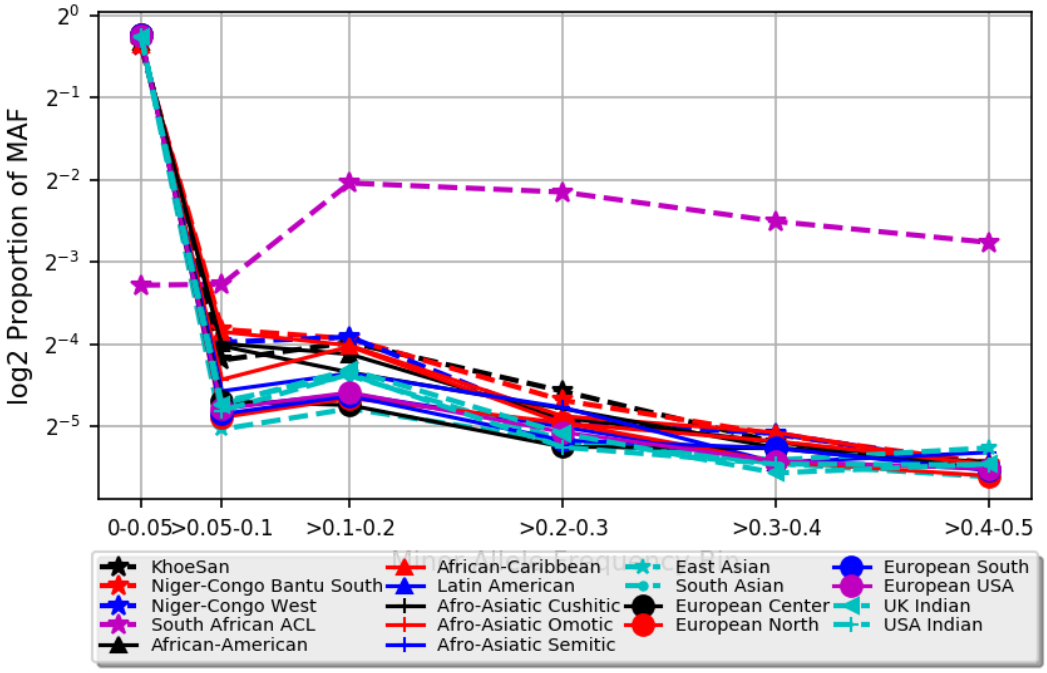


Figure 2.5: Minor allele frequencies at rare and common variants between Family A and Family B, and the rest of 20 worldwide ethnic groups.

2.3.6 Shared Identity by Descent

IBD analysis was leveraged to explore regions of the genome common within and between families. 17 genes in three segments were shared between Twin 1 and Twin 2 of Family A (**Table 2.5**). In addition, both Family A twins shared the *LINC01250* gene on chromosome 2p25.3, which encodes for long intergenic non-protein coding RNA 1250, with their affected brother. The affected father and affected brother shared a segment on chromosome 14q32.33, which consisted of 29 genes, and the unaffected mother shared 20 genes with the affected brother in a segment located at 19q13.42.

For Family B, twin males shared four segments comprising 37 genes (**Table 2.6**) and Twin 3 shared a gene segment on chromosome 13q34 which included one gene, with the unaffected sister. In addition, Family A twins, their affected brother and Family B twins also shared the *LINC01250* gene located on chromosome 2q25.3. And common to both families, both sets of twins shared the 15q11.2 region comprising 8 genes, and the unaffected mother and affected brother from Family A shared the *KIR3DXI* gene with Twin 3 and 4 from Family B.

Table 2.5: Detecting shared identity-by-descent (IBD) segments between family members in Family A.

Sample 1	Sample 2	Chromosomal Location	Start	End		LOD	IBD Segment Length
Twin 1 and Twin 2 (affected)	Brother (affected)	2p25.3	3005297	3057952	*LINC01250*	5.34	3.944
Twin 1 (affected)	Twin 2 (affected)	2q25.3	3005007	3058005	*LINC01250**	6.37	3.945
		15q11.2	22292484	22584320	*GOLGA6L22**, *GOLGA8EP**, *HERC2P2**, *HERC2P7**, *RN7SL545P**, *SPATA31E3P**	5.94	3.59
		21q21.1	16448783	16633125	*MIR125B2*, *MIR99A*, *MIRLET7C* ,*MIR99AHG*	10.67	2.077
Father (affected)	Brother (affected)	14q32.33	105952630	106208389	*ADAM6*, *IGHV1-2*, *IGHVIII-2-1*, *IGHV1-3*, *IGHV4-4*, *IGHV7-4-1*, *IGHV2-5*, *IGHVIII-5-1*, *IGHVIII-5-2*, *IGHV3-6*, *IGHV3-7*, *IGHV3-64D*, *IGHV5-10-1*, *IGHV3-11*, *IGHVIII-11-1*, *IGHV1-12*, *IGHV3-13*, *IGHVIII-13-1*, *IGHV1-14*, *IGHV3-15*, *IGHVII-15-1*, *IGHV3-16*, *IGHVIII-16-1*, *IGHV1-17*, *IGHV1-18*, *IGHV3-19*, *SLC20A1P2*	56.56	1.516
Mother (unaffected)	Brother (affected)	19q13.42	54184163	54537844	*CDC42EP5*, *KIR3DX1**, *LAIR1*, *LAIR2*, *LENG8*, *LENG8-AS1*, *LENG9*, *LILRA4*, *LILRA5*, *LILRA6*, *LILRB2*, *LILRB3*, *LILRB5*, *MBOAT7*, *MIR4752*, *RNU6-1307P*, *RPS9*, *TSEN34*, *TTYH1*, *VN1R104P*	7.76	1.778

Bold depicts genes shared between affected individuals in both families, and ***** depicts genes shared in family members (affected/unaffected) between families.

Table 2.6: Detecting shared identity-by-descent (IBD) segments between family members in Family B.

Sample 1	Sample 2	Chromosomal Location	Start	End	Genes	LOD	IBD Segment Length
Twin 3 (affected)	Twin 4 (affected)	2q25.3	3009692	3047881	*LINC01250**	3.31	2.748
		15q11.2	22370898	22524432	*GOLGA6L22*, GOLGA8EP*, HERC2P2*, HERC2P7*, RN7SL545P*, SPATA31E3P**	4.3	1.918
		19q13.42 19q13.42	54529431 54539272	55093392 55088813	*EPS8L1, FCAR, GP6, GP6-AS1, KIR2DL1, KIR2DL3, KIR2DL4, KIR2DP1, KIR2DS4, KIR3DL1, KIR3DL2, KIR3DL3, KIR3DP1, KIR3DX1*, LILRA1, LILRA2,LILRB1, LILRB1-AS1, LILRB4, LILRP1, LILRP1, LILRP2, MIR8061, NCR1, NLRP2, NLRP7, PPP1R12C, RDH13, RNU6-222P, VN1R105P*	7.4 7.35	1.775 1.751
Twin 3 (affected)	Sister (unaffected)	13q34	111663671	111916421	*LINC00354*	4.31	1.551

Bold depicts genes shared between affected individuals in both families, and ***** depicts genes shared in family members (affected/unaffected) between families.

2.3.7 Functional Partners and Further Enrichment Analysis

From the prioritised list of genes with mutational burdens, and IBD analysis, inferred functional partners of selected genes of interest were explored. Genes (*COL11A1, COL12A1, CATSPER2, KCNJ12, GP6, MIR99A, MIR99AHG, MIR125B, MIRLET7C,* and *LINC0150*) were selected based on potential pathogenicity for ACL rupture, shared regions in affected family members, previously published associated genes, and shared biology/function (**Appendix B, Section 1, Table S4**). The majority of the functional partners identified for Family A and Family B genes of interest included *collagens, proteoglycans, glycoproteins, integrins, laminins, growth factors, interleukins, microRNAs, apoptotic genes* and *protein kinases* (**Appendix B, Section 1, Table S4**). A few genes are involved in ion channel activity and signalling, with others involved in reproduction and fertilization pathways. However, the large majority of functional partners are involved in regulating the synthesis and degradation of ECM components, collagen fibrillogenesis and new blood vessel formation through angiogenesis related pathways (**Appendix B, Section 1, Table S4**).

2.4 DISCUSSION

The study employed a whole genome sequencing approach in family members from two unrelated twin families, with a history of ACL rupture. The findings presented in this chapter addresses potential function-altering variants and genetic modifiers in ACL rupture predisposition. The main findings include (i) the identification of genetic variants in three genes (*COL12A1, CATSPER2, KCNJ12*) that are commonly enriched for deleterious and loss-of-function mutations in a phenotypically defined family of patients, and with evidence of genetic association with different phenotypes, providing support for the complexity of the genetic

architecture of ACL rupture phenotypic variability. In addition, (ii) a shared IBD segment including the *LINC01250* gene in the telomeric region of chromosome 2p25.3 was noted between affected twins in both families, and an affected brother. And (iii) the identification of inferred functional partners that have previously been associated with altered risk of ACL rupture.

Type XII collagen (*COL12A1*) is a fibril-associated collagen belonging to the interrupted triple helices (FACITs) family. In addition to its role in fibrillogenesis (reviewed in **Chapter 1 Section 1.9.1**) type XII collagen provides a molecular bridge between fibrillar collagens, and other matrix molecules facilitating fibril interaction with other extracellular and cell surface molecules within ligaments (Frank, 2004). It is interesting to note that *COL12A1* was one of the genes highlighted to have a higher SNP burden in the two twin families, compared to previous world populations. Additionally, several case-control genetic association studies have previously explored the *COL12A1* rs970547 polymorphism identified in this study, with ACL rupture risk susceptibility (Ficek et al., 2014; O'Connell et al., 2015; Posthumus et al., 2010; Sivertsen et al., 2019; Zhao et al., 2020).

The cation channel sperm associated 2 (*CATSPER2*) and potassium inwardly rectifying channel subfamily J member 12 (*KCNJ12*) genes have not been implicated with risk of susceptibility to ACL rupture, and are therefore novel and noteworthy. However, the *CATSPER2* gene is mostly likely not relevant to ACL rupture risk, based on current lack of biological evidence to ligament or connective tissue phenotypes. Ion channels within human tissue are ubiquitous, and channel defects have been implicated in a wide variety of diseases affecting the nervous, cardiovascular, respiratory, endocrine, urinary system and immune systems (Kim, 2014). Also, potassium and ion-channel related genes, and their dysfunction have been implicated in the

development of chronic painful conditions (Schmidt and Schmidt, 2016; Waxman et al., 2014) and subsequently identified as potential therapeutic targets for painful conditions (Biasiotta et al., 2016). Even more, ion channels modulate membrane ion conductance across all cells and tissues, establishing electrical fields that affect cellular behaviours under normal conditions, during critical periods of development, and in response to tissue injury (Franklin et al., 2017). Further to that, ion channels and transporters are directly involved in the angiogenesis pathway, as they are expressed by vascular endothelial cells (Nilius and Droogmans, 2001) and are thought to contribute to vasodilation, in response to extracellular K+ concentration (Hibino et al., 2010). Therefore, the modified expression and activity of ion channels may be related to vascular alteration in pathological conditions (Biasiotta et al., 2016).

The angiogenesis related pathway plays a significant role in regulating ECM homeostasis, and perturbations of the expression of specific genes functioning in this pathway have also been implicated in contributing to the outcomes of surgical interventions such as ACL reconstruction (**Chapter 1, Section 1.9.4**). These candidate genes therefore represent new gene targets to explore the clinical heterogeneity and pathogenesis of ACL rupture, or in potential drug targeting strategies. New ideas for analgesic drug design are urgently needed, especially given the number of recent high-profile failures with some prospective targets, such as the neurokinin receptor 1 antagonists (Du and Gamper, 2013; Karthaus et al., 2019) which have caused many lead pharmaceutical companies to curb their focus in this area.

No shared IBD segments were identified at the regions where *COL12A1, CATSPER2* and *KCNJ12* are located, suggesting these mutations may not have occurred since the time of the most recent common ancestors and are therefore not founder mutations. Interestingly population structure analysis revealed Family A and Family B clustered separately, but in close

proximity to the European's cluster, possibly as a result of genetic drift in an isolated founder population, that occurred with the Euro-Asia settlement in South Africa since 1652 (Hunt, 2005). Interestingly, the two families shared an IBD segment that included a long intergenic non-protein coding RNA (lincRNA) *LINC01250* gene in the telomeric region of chromosome 2p25.3. It is known that the telomeric regions of any human chromosome harbour structural variants and repetitive nucleotide sequences (Vega et al., 2003) which was further illustrated by the high LD pattern among variants within non-protein genes, such as *LINC01250* (**Appendix B, Section 1, Figure S2B**). LincRNAs function broadly to fine tune target gene expression by the direct modulation of nuclear architecture, in addition to indirectly through transcription or translation activities (Ransohoff et al., 2018). From this observation, one could speculate that structural variation and changes within the *LINC01250* region may contribute to the severity and phenotypic variation of ACL rupture through the modulation of functional genes involved in ligament biology. Further investigation is needed to test this.

Notably, enriched pathways represented by the genes of interest point to relevant pathophysiological mechanisms affecting collagen fibrillogenesis, cell to cell communication, angiogenesis signalling, and homeostasis of the ECM through proteins such as integrins, interleukins, growth factors, glycoproteins and protein kinases (**Appendix B, Section 1, Table S4**) of which some are already therapeutic targets. Interestingly, recent evidence suggests that long noncoding RNA encoding genes are critical in angiogenesis and cell migration and proliferation pathways, with some lncRNA encoding genes vital to wound healing processes (Luan et al., 2018). Furthermore, *COL1A1, COL5A1, COL12A1, ACAN, BGN, DCN, FBN2, VEGFA, KDR* and *TGFB1* inferred functional partners have previously been associated with ACL rupture (**Chapter 1, Table 1.1**) and of these, *COL1A1, COL5A1, COL12A1, FBN2, VEGFA*, and *TGFB1* were also noted with predictive pathogenic SNPs in Family A and B

(**Figure 2.4A**). Differentiation was observed in the distribution of minor allele frequency at rare and common variants, between the family cohort and the rest of 20 worldwide ethnic groups. Possibly, due to (1) genetic drift and population bottleneck following the recent South African apartheid, where interracial marriage was prevented; and (2) family history of ligament and other musculoskeletal soft tissue injuries that might shape the genetic makeup of the two South African families studied. Furthermore, the identification of several genes previously associated with ACL rupture, with a higher proportion of pathogenic polymorphisms and gene-specific in SNP frequency, justifies and indicates that the actionability of these ACL-associated genes may have differing effects on worldwide ethnic groups, supporting the beneficial use of personalised medicine, and enabling a recommendation for ACL-specific clinical actionable genes list.

The study presented has provided novel insights into the genetic architecture of ACL rupture among the investigated family cohort. However, it was limited by the modest sample size of the ACL-family cohort, as larger sample sizes would most likely yield more findings. Furthermore, due to the complexity and multifactorial nature of ACL rupture, the number of affected families available for sequencing is limited. Moreover, the study was performed on isolated ACL family cases, without the possibility of variant segregation studies within the family, or in trio. Additionally, the study was limited by the availability of reliable pathogenicity prediction algorithms for intronic and intergenic variants. Therefore, it is suggested that coding variants are prioritized to facilitate this process. Overall, the findings support the need for intensive familial studies in multiple African versus European descent populations, to unravel the novel genes and those variants that are relevant in clinical practice for diverse populations of differing genetic background. Futhermore, future work should aim to explore the potential effects of these DNA variants on protein structure at an *in silico* level

using molecular dynamic simulations. Additionally, going forward, the association of these prioritised genes and their potential clinical heterogeneity variants within ACL rupture risk, should be investigated by utilising large case-control cohorts from the Going consortium (Pitsiladis et al., 2016); an established consortium to investigate the genetic predisposition underpinning ACL rupture, between different collaborating centres in the Southern and the Northern hemispheres.

2.5 CONCLUSION

In summary, the aim to identify novel and/or previously implicated biological genomic signatures with susceptibility to ACL rupture was achieved using a WGS approach in two South African twin families, with a history of ACL rupture. From the data, a catalogue of candidate *in silico* mutations and modifier genes that clustered in pathophysiological pathways important in ACL rupture, and with implications for therapeutic intervention were identified. Additionally, inferred functional partners previously associated with ACL rupture risk in small, single populations were highlighted for further exploration in larger, collective data sets (**Chapter 3 and 4**). Interestingly, three candidates clustering in potassium and ion-channel gene-families were implicated with *in silico* mutations, which not only play a role in angiogenesis, but their dysfunction is known to be involved in the development of chronic painful conditions and represent key therapeutic targets. This research fills an important gap in knowledge by using a WGS approach focusing on potential deleterious coding variants, important in two unrelated families with a historical record of ACL rupture. Therefore, making significant contributions to the present knowledge of the natural history, and clinical heterogeneity of ACL rupture, with the potential for informing the design of new therapeutics.

CHAPTER 3

INVESTIGATION OF POLYMORPHISMS LOCALISED TO *VEGFA* AND *KDR* WITH ACL RUPTURE PREDISPOSITION, IN A LARGE COLLECTIVE COHORT

The data presented in this chapter is currently in peer review (*Journal of Orthopaedic Research*) **Investigation of multiple populations highlight *VEGFA* polymorphisms to modulate anterior cruciate ligament injury**. Feldmann DC, Rahim M, Suijkerbuijk MAM, Laguette MJ, Cieszczyk P, Ficek K, Huminska-Lisowska K, Häger CK, Stattin E, Nilsson KG, Alvarez-Rumero J, Eynon N, Feller J, Tirosh O, Posthumus M, Chimusa ER, Collins M, September AV.

3.1 INTRODUCTION

Findings from the WGS study in **Chapter 2**, highlighted angiogenesis associated genes as inferred functional partners of prioritised WGS candidates (**Appendix B, Section 1, Table S4**). Furthermore, polymorphisms within angiogenesis associated genes have previously been implicated with altered risk of ACL rupture in single populations, following a case-control genetic association design (as previously reviewed in **Chapter 1, Section 1.9.4**). Vascularisation is critical for ligament healing (**Chapter 1, Section 1.3.2**) and so considering the role of the angiogenesis pathway in vascularisation, polymorphisms localising to specific angiogenesis associated genes were further explored in this chapter, in a large combined analysis, with predisposition to non-contact ACL rupture.

The angiogenesis-associated signalling cascade is controlled by several growth factors, of which VEGFA is the critical regulator. VEGFA activity occurs through binding to its primary receptor KDR, in the angiogenesis signalling pathway (**Chapter 1, Section 1.9.4**). Together, these proteins are considered key modulators of ligament matrix remodelling, with increased expression levels noted post-mechanical loading, after ligament injury, and following ligament reconstructive surgery (**Chapter 1, Section 1.9.4**). So far, the findings from previous genetic association studies investigating *VEGFA* and *KDR* polymorphisms with ACL rupture risk (Lulinska-Kuklik et al., 2019a; Rahim et al., 2014; Rahim et al., 2018) have been conflicted by contrasting genotype associations. Which as previously discussed may be as a result of exploring small sampling groups from single populations, with insufficient statistical power (**Chapter 1, Section 1.9.1**).

On that premise, this chapter aimed to explore the previously associated angiogenesis genes (*VEGFA* and *KDR*) that were implicated through WGS analyses in a large collective data set, with predisposition to non-contact ACL rupture. To achieve this aim, the objectives included (i) the investigation of five functional (Hansen et al., 2010; Lambrechts et al., 2003; Shahbazi et al., 2002; Wang et al., 2007) polymorphisms within angiogenesis genes *VEGFA* (rs699947 C/A, rs1570360 G/A, rs2010963 G/C) and *KDR* (rs2071559 A/G, rs1870377 T/A) with non-contact ACL rupture risk using a candidate gene, case-control genetic association study design in participants from Sweden, Poland and Australia. (ii) The investigation of *VEGFA* and *KDR* polymorphisms in a combined data set, comprising pooled genotyping data from the independent cohorts above (Sweden, Poland and Australia) and previously published cohorts from South Africa (Rahim et al., 2014) and Poland (Lulinska-Kuklik et al., 2019a), and (iii) the investigation of allele interactions across *VEGFA* and *KDR* polymorphisms, as a proxy for gene-gene interactions with risk of non-contact ACL rupture.

The *a priori* hypothesis tested was based on the reported functional consequence of the genotypes, such that (i) the rs699947 CC, rs1570360 GG (Shahbazi et al., 2002) and rs2010963 CC (Awata et al., 2002) genotypes associated with an increased VEGFA expression, would be associated with increased risk, and that (ii) the functional inferred *VEGFA* A-A-G haplotype, associated with a decreased VEGFA expression (Lambrechts et al., 2003) would be associated with a reduced risk of ACL rupture and (iii) the *KDR* G-A inferred haplotype constructed from rs2071559 and rs1870377, associated with reduced *KDR* transcription (Dong et al., 2007) may potentially impair the efficacy of VEGFA binding to KDR (Wang et al., 2007) and thereby influence the downstream signalling effects of VEGFA, and lead to an increased risk of ACL rupture.

3.2 METHODS

3.2.1 Participant Characteristics

This study followed a case-control genetic association, comprised of three populations (Sweden, Poland and Australia). All participants from the individual cohorts completed questionnaires from the respective research centres detailing participant personal details, ancestry, lifestyle habits, occupational details, sporting history (sports played, number of years, playing level, frequency), details of ACL injury, history of other ligament or tendon injury, and medical history. Written informed consent was obtained from all participants according to the declaration of Helsinki and the study approved by the Human Research Ethics Committee of the Faculty of Health Sciences within the University of Cape Town, South Africa (HREC 269/2014, 110/2018, 2013.223, 622/2015) the Regional Ethical Review Board in Umeå, Sweden (dnr. 2011-200-31M) the Bioethics Committee for Clinical Research, the Regional Medical Chamber, Gdansk, Poland (K-8/16) and the Epworth Hospital (HREC approval: 57012) Victoria, Australia (**Appendix A, Section 1**).

The inclusion criteria for the cases included individuals with a clinical diagnoses of an anterior cruciate ligament rupture (ACL) based on physical examination, and confirmed by either magnetic resonance imaging or arthroscopy. Only ACL ruptures resulting from a non-contact mechanism were included. For the control participants, inclusion criteria included: participation in regular sporting activites, participated in similar sports to cases, had no history of ACL injury or other ligament and tendon injuries, and were within a similar age category as the ACL rupture group. All participants participated in regular sporting activities, primarily at a recreational level. The type of sport and years of participation are provided in the

supplementary data for the Swedish (cases and controls), Polish (cases and controls) and Australian (cases only) participants.

The Swedish cohort was comprised of 116 control (SWE CON) and 95 ACL (SWE ACL) participants recruited between 2011 and 2013 from orthopaedic clinics in two hospitals in the cities of Umeå:Västerbotten and Luleå: Norrbotten as previously described (Suijkerbuijk et al., 2019). The Polish cohort comprised 149 control (POL CON) and 127 ACL (POL ACL) participants recruited from the Galen Orthopedics Clinic in Poland between 2008 and 2018. While the Australian cohort consisted of 83 control males only (AUS CON) and 342 ACL (AUS ACL) participants. Australian control participants were recruited from the Genes and Skeletal Muscle Adaptive Response to Training (Gene SMART) cohort (Yan et al., 2017). These individuals were from the student and staff populations from universities and local communities. Whereas the ACL cases were recruited between 2006 and 2018 from Epworth Richmond hospital in Melbourne Australia.

For the previously published South African cohort, recruitment took place between 2006 and 2012. Cases from the Sports Science Orthopaedic Clinic in Cape Town, South Africa and controls from local sporting clubs and gyms within the Cape Town area, as described in (Rahim et al., 2014). Lastly, for the previously published Polish cohort, the participants were recruited between 2009 and 2016 with male case participants recruited from Polish soccer leagues, and controls from similar soccer teams. Female cases were recruited from either soccer teams and included amateur skiers, and the female controls were recruited from sports clubs and wellness centres as described in (Lulinska-Kuklik et al., 2019). All these ACL cases resulted from non-contact mechanisms, as described by authors.

Type of sports participation for all Swedish (**Appendix B, Section 2, Table S6)** and Polish (**Appendix B, Section 2, Table S7**) cases and controls, as well as Australian case participants (**Appendix B, Section 2, Table S8**) was categorized into contact sports, non-contact jumping sports, non-contact non-jumping sports, and skiing sports (with a non contact mechanism of injury), as previously defined (Flynn et al., 2005). Participation years for Swedish participants are described in **Appendix B, Section 2, Table S6**, while data describing participation years for Polish controls and all Australian participants was unavailable.

Level of sport was classified into three groups (elite, national and recreational) as previously categorized for Swedish (Markström et al., 2019) and Australian (Voisin et al., 2014) participants. The majority (76%) of Swedish case participants were recreational, with 6% elite, and the remaining 18% competing at a national level. Similarly, most of the Swedish controls were recreational (83%), while 8% were elite, and 8% competing at a national level (**Appendix B, Section 2, Table S9**). Data describing the level of sport in Polish participants was missing. For the Australian case participants, 98% were recreational athletes, with 2% of participants competing at a national level (**Appendix B, Section 2, Table S9).** Type of sport and years of participation was unavailable for the Australian controls, however participants were deemed moderately trained (VO_{2peak} 35–60 ml. $min^{-1}.kg^{-1}$) (Yan et al., 2017), participating in physical activity at a recreational level (**Appendix B, Section 2, Table S9**). Sports participation data for the South African cohort was previously described in Rahim et al. (2014). Type of sport was categorized into contact, non-contact jumping, non-contact, non-jumping, and skiing sports (with a non contact mechanism of injury) with participation years described. Briefly, female participants were matched for all sporting activity. Male participants were matched for participation in non-contact jumping sports and skiing sports. However, significantly more male cases participated in contact sports and significantly more male controls participated in

non-contact, non-jumping sports (Rahim et al., 2014). Level of sport was not documented. Similarly, data for the published Polish cohort was previously described in Lulinska-Kuklik et al. (2019), where type of sport, level of exposure and sporting level are documented. Briefly, all male participants were matched for type of sport, level and frequency of exposure. Additionally, sports participation data for female case and control participants were comparable as described by the authors (Lulinska-Kuklik et al., 2019).

For the combined cohort (COMB) analyses, genotype data from Swedish, Polish and Australian participants collected in this study, were pooled with the genotype data from previously published study groups from South Africa (CON: 227 ACL: 126) (Rahim et al., 2014) and Poland (CON: 190 ACL: 222) (Lulinska-Kuklik et al., 2019a) with approval from the respective authors. *KDR* polymorphisms were not investigated in the published Polish study group (Lulinska-Kuklik et al., 2019a) thus for the current study, 1,677 samples (COMB CON: 765 and COMB ACL: 912) were investigated for the *VEGFA* polymorphisms, and 1,265 samples (COMB CON: 575 and COMB ACL: 690) for the *KDR* polymorphisms.

3.2.2 DNA Isolation

For the cohort from Sweden, genomic DNA was extracted from venous blood using a rapid non-enzymatic ethanol precipitation as previously described by Lahiri and Nurnberger (Lahiri and Nurnberger, 1991) with slight modifications (Mokone et al., 2006). For Poland, DNA was extracted from oral epithelial cells using a Gen Elute Mam¬malian Genomic DNA Miniprep Kit (Sigma, Germany) according to the manufacturer's recommendations, and for Australia, DNA was isolated from a venous blood aliquot using a sequenced extraction technique

(FlexiGene DNA Kit, Qiagen P/L, Valencia, California, USA) or the MagSep Blood gDNA kit with the epMotion M5073 automated pipetting system (Eppendorf, Germany).

3.2.3 SNP Selection and Genotyping

Single nucleotide polymorphisms were selected based on biological significance of the encoding genes, and previous associations with multifactorial phenotypes. Participants were genotyped for five functional single nucleotide polymorphisms (SNPs) *VEGFA*: rs699947 C/A, rs1570360 G/A rs201963 G/C (**Figure 3.1A**) and *KDR*: rs2071559 G/A and rs1870377 T/A (**Figure 3.1B**) using standard PCR based technology for genotyping as previously described (Rahim et al., 2014). For the Swedish cohort only, restriction fragment length polymorphism (RFLP) analysis was used to genotype *VEGFA* rs699947 (*Bgl*II) and *KDR* rs1870377 (*Alu*I) and custom-designed fluorescence-based TaqMan™ PCR assays (Applied Biosystems, Foster City, CA, USA) used to genotype *VEGFA* rs1570360 (assay ID: C___1647379_10) *VEGFA* rs2010963 (C___8311614_10) and *KDR* rs2071559 (C__15869271_10). TaqMan™ PCR assays were used to genotype all five SNPs in the Polish and Australian cohorts (assay ID rs699947: C___8311602_10 and rs1870377: C__11895315_20). The TaqMan™ genotyping PCR reactions were conducted using the Applied Biosystems QuantStudio™ Real-Time PCR system and the Applied Biosystems QuantStudio™ Real-Time PCR software (Applied Biosystems). The manufacturer's instructions were followed. Negative controls (no DNA) and five repeat samples (known genotypes) were included as quality control measures for each 96-well plate. Genotypes were confirmed by two independent investigators (DF and MR) with an average 98.7% call rate, and laboratory work was conducted at The University of Cape Town.

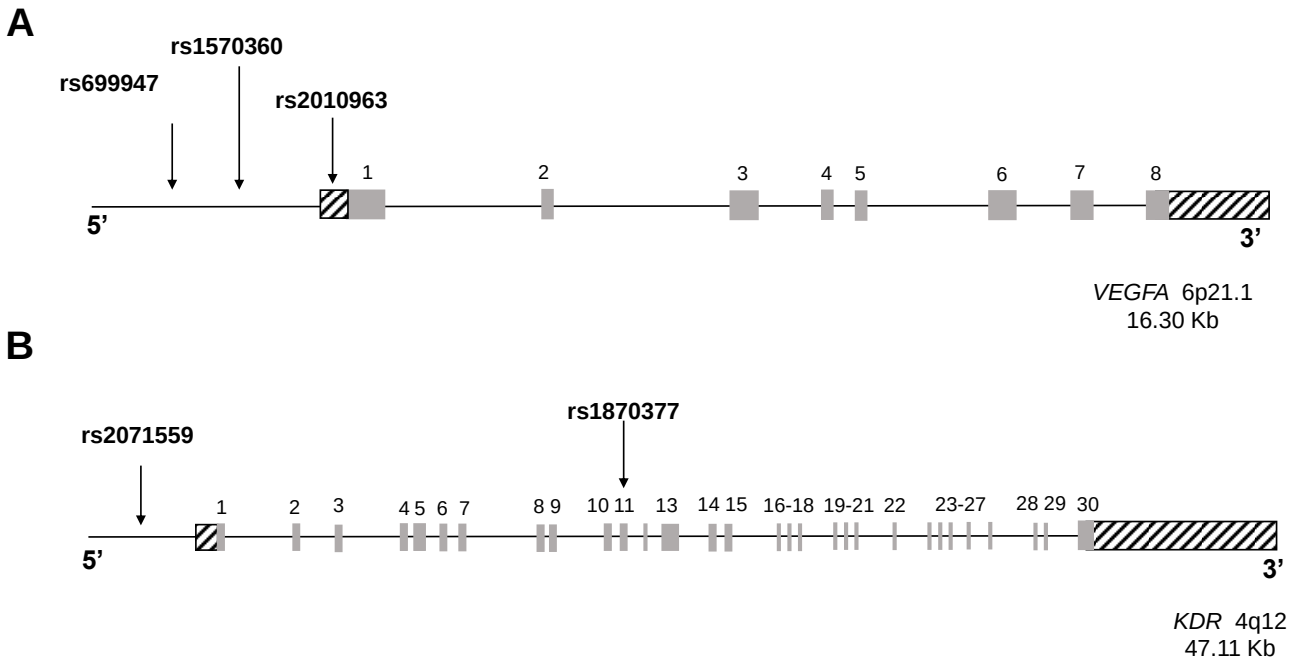


Figure 3.1: Schematic representation of (A) *VEGFA* depicting the rs699947 (C/A), rs1570360 (G/A) and rs2010963 (G/C) polymorphisms, and (B) *KDR* depicting the rs2071559 (A/G) and rs1870377 (T/A) polymorphisms. Exons are shown as grey boxes (with exon number above), introns as horizontal lines, and untranslated regions are depicted as hatched boxes. The size of the genes are indicated in Kb and their chromosomal locations are also indicated. Figures are not drawn to scale. All the information used to construct the figure was obtained from databases hosted by the National Centre for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/) and Ensembl Genome Data Centre (http://www.ensembl.org/index.html).

3.2.4 Statistical Analyses

Power calculations were calculated using QUANTO v1.2.4 (http://biostats.usc.edu/software). For the Swedish cohort, assuming minor allele frequencies between 0.2 and 0.5, a sample size of 85 cases would detect an allelic odds ratio (OR) of 2.0 and greater, at a power of 80% and a significance level of 5%. For the Polish and Australian cohorts, assuming minor allele frequencies between 0.2 and 0.5, a sample size of 118 cases and greater would detect an OR of 1.8 and greater, at a power of 80% and a significance level of 5%.

The statistical program R (R Development Core Team, 2020) was used for all statistical analyses. Participant descriptive statistics and genotype effects were compared using a one-way analysis of variance to determine any significant differences between the mean characteristics of the CON and ACL groups, and any possible genotype effects on age, sex, body mass, height and body mass index (BMI). The R packages *genetics* (Warnes et al., 2012) and *SNPassoc* (González et al., 2014) were used to analyse differences in genotype and allele frequencies between groups, and to calculate Hardy-Weinberg equilibrium (HWE) probabilities. Inferred haplotypes were constructed using the genotype data, and inferred allele interactions for *VEFGA-KDR* explored using the combined allele analyses from the genotype data. The R package *haplo.stats* (Sinnwell and Schaid, 2018) was used, adjusting for confounding variables where appropriate, for the individual and collective cohorts. For Sweden, age and type of sport was a considered confounder, for Poland age, body mass and type of sport, for Australia age and sex, and for the combined cohort, age, BMI and country were included in the adjustment model. The strength of the association is reflected by the score statistic (Haplo.score), where negative values infer decreased risk, and positive scores indicate an increased risk for ACL rupture. Sex is a known intrinsic risk factor for ACL rupture

predisposition, and previous research stratified by sex (Rahim et al., 2017). In the combined cohort, sex descriptive characteristics for the previous published Polish cohort and the Australian ACL group were missing, and therefore stratification by sex was not carried out. Furthermore, due to reduced power, stratification by sex was not included for the single independent cohort analyses. Chi-squared tests were used to compare genotype frequency distributions in the combined cohort analyses. Statistical significance was accepted when $p < 0.05$, and the False Discovery Rate (FDR) procedure used to adjust for multiple comparisons where statistical significance was met, using the method applied for multiple testing under dependency (Benjamini and Hochberg, 1995).

3.3 RESULTS

3.3.1 Participant Characteristics

Participants in the COMB CON group differed significantly in age, body mass and BMI compared to individuals in the COMB ACL group (**Table 3.1**). The COMB ACL (28.6 ± 11.3 years, n = 610) group were significantly younger than the COMB CON (30.5 ± 12.3 years, n = 567, p = 0.008) group. For body mass, the COMB ACL group (78.0 ± 14.9 kg, n = 606) were significantly heavier than the COMB CON (74.3 ± 13.9 kg, n = 560, p < 0.001) group (**Table 3.1**) even when adjusted for age (p = 0.008). And for BMI, the COMB ACL group had a significantly greater mean BMI (25.3 ± 5.0 $kg.m^{-2}$, n = 596) than the participant mean for the COMB CON (23.9 ± 3.1 $kg.m^{-2}$, n = 557, p < 0.001) group, even after adjustment for age and body mass (**Table 3.1**).

Table 3.1: Descriptive characteristics for all participants in the combined control (CON) group and the combined anterior cruciate ligament rupture (ACL) group

	COMB CON	**COMB ACL**	**p-value**
	n = 765	n = 913	
Age (years)	30.5 ± 12.3 (567)	28.6 ± 11.3 (610)	**0.008**
Sex (%male)	65 (575)	54 (690)	0.187
Height (cm)	175.9 ± 9.6 (560)	175.7 ± 10.3 (599)	0.611
Body Mass (kg)	74.3 ± 13.8 (560)	78.0 ± 14.9 (606)	**< 0.001 (0.008)[a]**
BMI (kg.m^{-2})	23.9 ± 3.1 (557)	25.3 ± 5.0 (596)	**< 0.001 (< 0.001)[b]**

Values are expressed as mean ± standard deviation; sex is represented as a percentage. The number of participants (n) with available data for each variable is in parentheses. CON vs ACL, p-values in **bold** typeset indicates significance ($p < 0.05$). P-values are depicted as unadjusted and adjusted in parentheses. [a]: adjusted by age, [b]: adjusted by age and body mass.

Participants in the Swedish cohort were previously described (Suijkerbuijk et al., 2019). In brief, cases and controls were matched for all descriptive charactersitics, except for age. Where participants in the SWE CON (44.7 ± 11.9 years, n = 114, $p < 0.001$) were significantly older than participants in the SWE ACL (37.1 ± 13.3 years, n = 93) group (**Appendix B, Section 2, Table S5**).

In the Polish cohort (**Appendix B, Section 2, Table S5**). The POL ACL group were significantly older (31.4 ± 10.1 years, n =127, $p < 0.001$) and heavier (79.0 ± 14.9 kg, n = 126, $p < 0.001$) than the POL CON (age: 21.0 ± 1.8 years, n = 149; body mass: 72.6 ± 12.0 kg, n =149) group (**Appendix B, Section 2, Table S5**). Additionally, there was a significant difference in BMIs between the POL CON and POL ACL group, however when adjusted for age and body mass, there was no longer a difference between groups (**Appendix B, Section 2, Table S5**).

Furthermore, in the Australian cohort, participants in the AUS CON and AUS ACL groups differed significantly in age, sex, height and body mass (**Appendix B, Section 2, Table S5**). The AUS CON group consisted of only male participants that were significantly older (31.0 ± 8.3 years, n = 82, $p < 0.001$) taller (180.0 ± 0.1 cm, n = 82, $p < 0.001$) and heavier (81.6 ± 12.0 kg, n = 82, $p = 0.022$) than the AUS ACL group (age: 25.2 ± 9.4 years, n = 269; height: 175.2 ± 0.1 cm, n = 268 ; body mass: 77.5 ± 14.7 kg, n = 268). However, when adjusted for (age and sex), and (age, sex and height) no significant difference in height ($p = 0.120$) and body mass ($p = 0.123$) respectively, was noted (**Appendix B, Section 2, Table S5**).

3.3.2 Sports Participation

The SWE CON and SWE ACL groups were matched for participation in non-contact jumping sports (**Appendix B, Section 2, Table S6**). However, a significantly higher ($p < 0.001$) proportion of SWE CON participants (72%, n = 66) participated in non-contact, non-jumping sports compared to the SWE ACL participants (6%, n = 5), and no SWE-CON individuals participated in contact sports compared to 68% (n = 62) of SWE-ACL participants ($p < 0.001$). There were no differences in years of sport participation or sporting level between Swedish controls and cases (**Appendix B, Section 2, Table S6 and Table S9**).

The POL CON and POL ACL groups were matched for participation in contact sports (**Appendix B, Section 2, Table S7**). However, significantly ($p = 0.009$) more controls (44%, n = 62) participated in non-contact, non-jumping sports compared with cases (27%, n = 34) with no controls participating in non-contact jumping sports, compared to 21% (n = 24) of cases (**Appendix B, Section 2, Table S7**).

For the AUS ACL group, 85% (n = 245) of participants participated in contact sports, with 4% (n = 11) participating in non-contact jumping, and 11% (n = 32) in non-contact, non-jumping sports (**Appendix B, Section 2, Table S8**). Furthermore, participants in the AUS CON and AUS ACL groups were matched for sporting level, where the majority (98%, n = 283) of the AUS ACL and all AUS CON participants were recreational athletes (**Appendix B, Section 2, Table S9**).

3.3.3 Genotype Effects

In all participants of the combined cohort, *VEGFA* rs699947 (C/A) was associated with body mass (**Table 3.2**). Participants with the rs699947 CC genotype were significantly heavier (77.7 ± 15.9 kg, n = 268, p = 0.018) than those with the CA (76.5 ± 13.9 kg, n = 579) or AA (74.4 ± 13.1 kg, n = 291) genotypes (**Appendix B, Section 2, Table S10**). Additionally, participants with the *VEGFA* rs1570360 (G/A) GG genotype had significantly greater BMIs (24.8 ± 3.7 $kg.m^{-2}$, n = 513, p = 0.048) than participants with the AA (23.9 ± 2.8 $kg.m^{-2}$, n = 124) genotype (**Appendix B, Section 2, Table S10**). Moreover, *VEGFA* rs2010963 (G/C) was associated with age and body mass (**Table 3.2**). Participants with the rs2010963 GG genotype were significantly younger (28.2 ± 10.8 years, n = 508, p = 0.003) than those with the CG (30.1 ± 12.2 years, n = 502) or CC (32.3 ± 13.4 years, n = 153) genotypes (**Appendix B, Section 2, Table S10**) and furthermore participants with the rs2010963 CG genotype were significantly heavier (77.3 ± 14.9 kg, n = 500, p = 0.047) than individuals with the GG (75.1 ± 14.2 kg, n = 501) genotype (**Appendix B, Section 2, Table S10**).

For the *KDR* polymorphisms, the rs1870377 (T/A) was associated with age (**Table 3.2**). Where participants with the rs1870377 TT genotype were significantly younger (28.6 ± 11.5 years, n = 645, p = 0.0014) than participants with the AT (30.1 ± 12.1 years, n = 421) or AA (32.0 ± 11.3 years, n = 86) genotypes (**Appendix B, Section 2, Table S10**).

Table 3.2: Genotype associations with descriptive measures for the *VEGFA* rs699947 (C/A), rs1570360 (G/A) and rs2010963 (G/C) and *KDR* rs2071559 (A/G) and rs1870377 (T/A) polymorphisms, for all participants in the combined cohort.

Polymorphism	Age	Sex	Body Mass	BMI	Height
VEGFA rs699947 (C/A)	0.629	0.836	**0.018**	0.458	0.278
VEGFA rs1570360 (G/A)	0.626	0.622	0.098	**0.048**	0.522
VEGFA rs2010963 (G/C)	**0.0003**	0.669	**0.047**	0.166	0.908
KDR rs2071559 (A/G)	0.301	0.413	0.101	0.364	0.598
KDR rs1870377 (T/A)	**0.014**	0.122	0.075	0.441	0.227

In all Swedish participants, *KDR* rs2071559 (A/G) was significantly associated with body mass and BMI (**Appendix B, Section 2, Table S11**). Participants with the rs2071559 AA genotype were heavier (76.8 ± 16.6 kg, n = 42) than participants with the GG genotype (69.3 ± 9.7 kg, n = 49, p = 0.034) and had greater BMIs (25.4 ± 3.5 $kg.m^{-2}$, n = 42) than those with the GG genotype (23.7 ± 2.6 $kg.m^{-2}$, n = 49, p = 0.018). No other genotype effects were noted for sex, age, height, body mass or BMI.

In all participants from Poland, *KDR* rs1870377 was associated with sex and body mass (**Appendix B, Section 2, Table S12**). Where the rs1870377 AA genotype was significantly under-represented in Polish male participants (56%, n = 25) compared to the AT genotype (81%, n = 114, p = 0.023) and individuals with the AT genotype were heavier (77.6 ± 15.3 kg, n = 113) than those with the TT (73.5 ± 11.4 kg, n = 129) or AA genotypes (71.7 ± 11.8 kg, n = 25). No further significant genotype effects were noted on sex, age, height, body mass and

BMI (**Appendix B, Section 2, Table S12**). In the Australian cohort, *VEGFA* rs1570360 (G/A) was associated with height (**Appendix B, Section 2, Table S13**), where participants with the rs1570360 AA genotype were significantly taller (178.9 ± 7.7 cm, n = 42, p = 0.035) than individuals with the AG (174.9 ± 9.4 cm, n = 144) genotype (**Appendix B, Section 2, Table S13**).

3.3.4 Genotype and Allele Frequencies

In the combined analysis, no significant differences in the genotype or allele frequencies between the COMB CON and COMB ACL groups for the *VEGFA* rs699947 C/A (genotype p = 0.745 and allele p = 0.771) and *VEGFA* rs1570360 G/A (genotype p = 0.922 and allele p = 0.713) polymorphisms (**Figure 3.2 and Appendix B, Section 2, Table S14**) were noted. However, the *VEGFA* rs2010963 (G/C) CC genotype was significantly over-represented (p = 0.0001, FDR p = 0.001, OR 2.16, 95% CI: 1.47-3.19) in the COMB ACL group (18%) compared to the COMB CON (11%) group (**Figure 3.2C**). Moreover, the rs2010963 CC genotype was still significantly different between cases and controls when the SWE-ACL group was removed from the analysis (**Appendix B, Section 2, Figure S4**). Furthermore, rs2010963 C allele was significantly over-represented (p = 0.0006, FDR p = 0.003, OR 1.29, 95% CI: 1.11-1.49) in the COMB ACL group (40%) compared to the COMB CON (34%) group (**Appendix B, Section 2, Table S14**). For the investigated *KDR* polymorphisms, no significant differences in the genotype or allele frequency distributions were observed between the COMB CON and COMB ACL groups for the *KDR* rs2071559 G/A (genotype p = 0.573 and allele p = 0.688) or *KDR* rs1870377 T/A (genotype p = 0.167 and allele p = 0.097) polymorphisms (**Figure 3.3 and Appendix B, Section 2, Table S14**). In addition, all polymorphisms were in HWE.

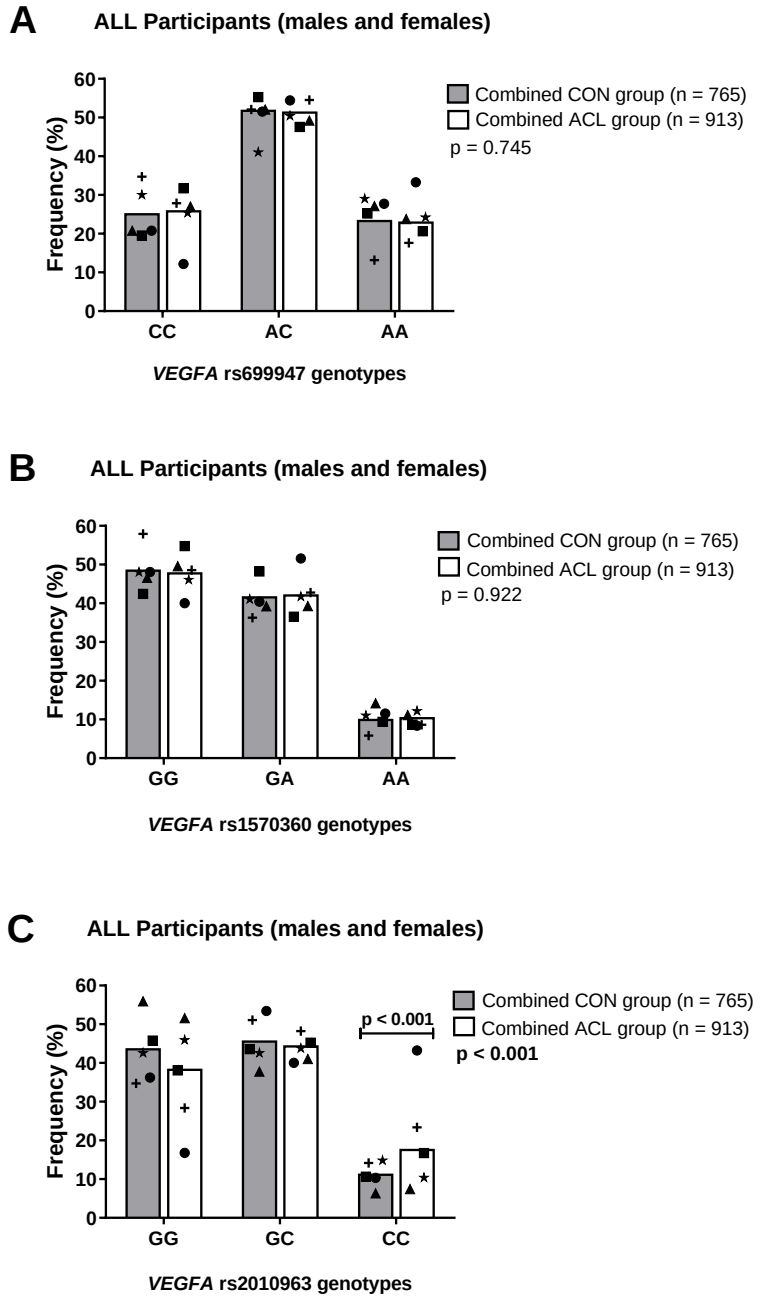


Figure 3.2: Genotype frequency distributions for the *VEGFA* (A) rs699947 (A/C), (B) rs1570360 (G/A) and (C) rs2010963 (G/C) polymorphisms in the combined control (CON) group (grey bars) and the combined anterior cruciate ligament rupture (ACL) group (white bars) for all participants. Genotype frequency distribution differences between the individual Swedish (circles), Polish (triangles), Australian (stars), published South African (squares) (Rahim et al., 2014) and published Polish (crosses) (Lulinska-Kuklik et al., 2019a) cohorts are: CON groups; rs699947 p = 0.008, rs1570360 p = 0.049 and rs2010963 p = 0.006 and ACL groups; rs699947 p = 0.027, rs1570360 p = 0.149 and rs2010963 p < 0.001. Statistically significant differences in genotype frequencies between the combined CON and ACL groups are depicted on the graph. The chi-squared global p-value for comparison of the combined CON (n=765) and combined ACL (n=913) group is depicted in the figure.

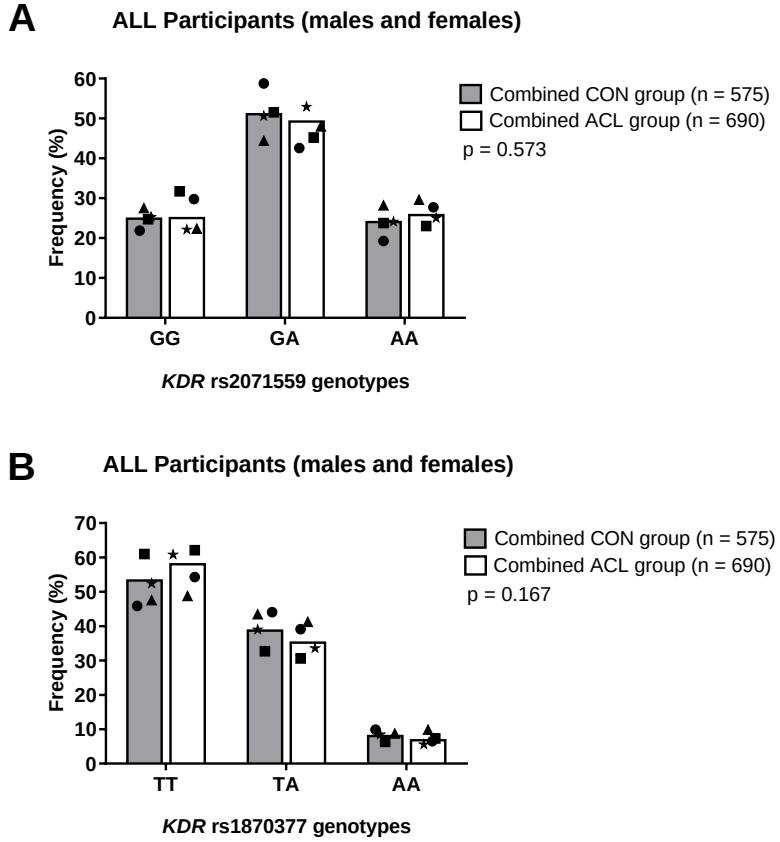


Figure 3.3: Genotype frequency distributions for the *KDR* (A) rs2071559 (A/G) and (B) rs1870377 (T/A) polymorphisms in the combined control (CON) group (grey bars) and the combined anterior cruciate ligament rupture (ACL) group (white bars) for all participants. Genotype frequency distribution differences between the individual Swedish (circles), Polish (triangles), Australian (stars), and published South African (squares) (Rahim et al., 2014) cohorts are: CON groups; rs2071559 p = 0.487, rs1870377 p = 0.134 and ACL groups; rs2071559 p = 0.223, rs1870377 p = 0.242). The chi-squared global p-value for comparison of the combined CON and ACL group is depicted in the figure.

In Swedish participants, no significant differences in the genotype or allele frequencies were noted between the SWE CON and SWE ACL groups for the *VEGFA* rs699947 (genotype p = 0.762 and allele p = 0.179) *VEGFA* rs1570360 (genotype p = 0.642 and allele p = 0.669) *KDR* rs2071559 (genotype p = 0.108 and allele p = 1.000) and *KDR* rs1870377 (genotype p = 0.752 and allele p = 0.228) polymorphisms (**Appendix B, Section 2, Table S15**). However, for the *VEGFA* rs2010963 (G/C) polymorphism, genotype frequency was significantly ($p < 0.001$, FDR $p < 0.001$) different between the SWE-CON and SWE-ACL groups (**Appendix B, Section 2, Table S15**). The GG genotype was significantly under-represented (p = 0.001, OR: 2.8, 95% CI: 1.45-5.41) in the SWE-ACL group (17%) compared to the SWE-CON group (36%) and the CC genotype was significantly over-represented ($p < 0.001$, OR: 6.6, 95% CI: 3.20-13.55) in the SWE-ACL group (43%) compared to the SWE-CON group (10%). In addition, the rs2010963 C allele was significantly over-represented ($p < 0.001$, FDR $p < 0.001$, OR: 2.9, 95% CI: 1.92-4.42) in the SWE-ACL group (63%) compared to the SWE-CON (37%) group (**Appendix B, Section 2, Table S15**).

In Polish participants, no significant genotype or allele frequency differences were noted between the POL CON and POL ACL groups, for *VEGFA* rs699947 C/A (genotype p = 0.158 and allele p = 0.297) *VEGFA* rs1570360 G/A (genotype p = 0.497 and allele p = 0.513) *VEGFA* rs2010963 G/C (genotype p = 0.782 and allele p = 0.491) *KDR* rs2071559 A/G (genotype p = 0.915 and allele p = 0.488) or *KDR* r1870377 T/A (genotype p = 0.268 and allele p = 1.000) (**Appendix B, Section 2, Table S16**).

Furthermore, in the Australian participants, no significant differences in the genotype or allele frequencies were noted between the AUS CON and AUS ACL groups for any of the investigated polymorphisms. Specifically, *VEGFA* rs699947 C/A (genotype p = 0.489 and

allele p = 1.000) *VEGFA* rs1570360 G/A (genotype p = 0.776 and allele p = 0.712) *VEGFA* rs2010963 G/C (genotype p = 0.299 and allele p = 0.644) *KDR* rs2071559 A/G (genotype p = 0.895 and allele p = 1.000) and *KDR* r1870377 T/A (genotype p = 0.289 and allele p = 0.712) (**Appendix B, Section 2, Table S17**).

The frequency distributions were significantly different between the Swedish, Polish, Australian, published South African and published Polish control groups for *VEGFA* rs699947 (p = 0.008) *VEGFA* rs1570360 (p = 0.049) and *VEGFA* rs2010963 (p = 0.006) polymorphisms (**Figure 3.2**). For the ACL groups, a significant difference between the cohorts was noted for the rs699947 (p = 0.027) and rs2010963 (p < 0.001) polymorphisms, whereas the frequencies were similar for the rs1570360 polymorphism (p = 0.149). For *KDR* rs2071559 and rs1870377 SNPs, all genotype frequency distributions were similar between the independent cohort control, and ACL groups (**Figure 3.3**). Moreover, the genotype frequency distributions for the combined control and ACL groups were comparable with the HapMap central European (CEU) reported frequencies (Yates et al., 2019) for all the investigated *VEGFA* and *KDR* polymorphisms.

3.3.5 Inferred Haplotypes

Inferred haplotypes were constructed using the genotype data for the three *VEGFA* (rs699947 C/A, rs1570360 G/A, rs2010963 G/C) and two *KDR* (rs2071599, rs1870377) polymorphisms. Only haplotypes inferred at a frequency greater than 2% were compared. For the combined cohort the *VEGFA* A-A-G haplotype was significantly (p = 0.010, haplo.score: -2.58, OR: 0.85, 95% CI: 0.69-1.05) under-represented in the combined ACL (23%) compared to the combined CON (28%) group and similarly, the A-G-G haplotype was significantly (p = 0.036,

haplo.score: -2.09, OR: 0.81, 95% CI: 0.64-1.02) under-represented in the combined ACL (12%) compared to the combined CON (16%) group (**Figure 3.4A**). Additionally, in a reduced interval analysis, (*VEGFA* rs1570360 G/A - rs2010963 G/C) both the -G-G and -A-G inferred haplotypes were significantly under-represented (-G-G: p = 0.031, haplo.score: -2.15, OR: 1.00 and -A-G: p = 0.024, haplo.score: -2.25, OR: 0.98, 95% CI: 0.82-1.18) in the combined ACL group, compared to the combined CON group (**Figure 3.4C**). While the -G-C haplotype was significantly over-represented (p = 0.012, haplo.score: 2.50, OR: 1.18, 95% CI: 0.99-1.40) in the combined ACL group, when compared to the combined CON group (**Figure 3.4C**). Four inferred haplotypes were constructed from the two *KDR* (rs2071559 A/G, rs1870377 T/A) polymorphisms. However, no significant differences were observed in the haplotype frequency distributions between the COMB CON and COMB ACL group, in all participants (**Figure 3.4D**).

In the Swedish cohort, the *VEGFA* A-A-G and A-G-G inferred haplotypes constructed from rs699947 (A/C) rs1570360 (G/A) and rs2010963 (G/C) were significantly under-represented in the SWE-ACL group compared to the SWE-CON group (A-A-G: p = 0.005, haplo.score: -2.79, SWE-CON 31% vs SWE-ACL 6% and A-G-G: p = 0.048, haplo.score: -1.97, SWE-CON 18% vs SWE-ACL 7%). Additionally, the A-G-C haplotype was significantly over-represented in the SWE-ACL group in comparison to the SWE-CON (A-G-C: p = 0.002, haplo.score: 3.03, SWE-CON group 3% vs SWE-ACL group 19%) group (**Appendix B, Section 2, Figure S3A**). For the inferred *KDR* haplotypes, no significant (p = 0.333) differences between the SWE CON and SWE ACL groups were noted (**Appendix B, Section 2, Figure S3D**). Furthermore, no significant differences in the inferred *VEGFA* (p = 0.788) or *KDR* (p = 0.367) haplotype frequencies were noted between the POL CON and POL ACL groups, or for the inferred

VEGFA (p = 0.869) or *KDR* (p = 0.616) haplotype frequencies between the AUS CON and AUS ACL groups (**Appendix B, Section 2, Figure S3**).

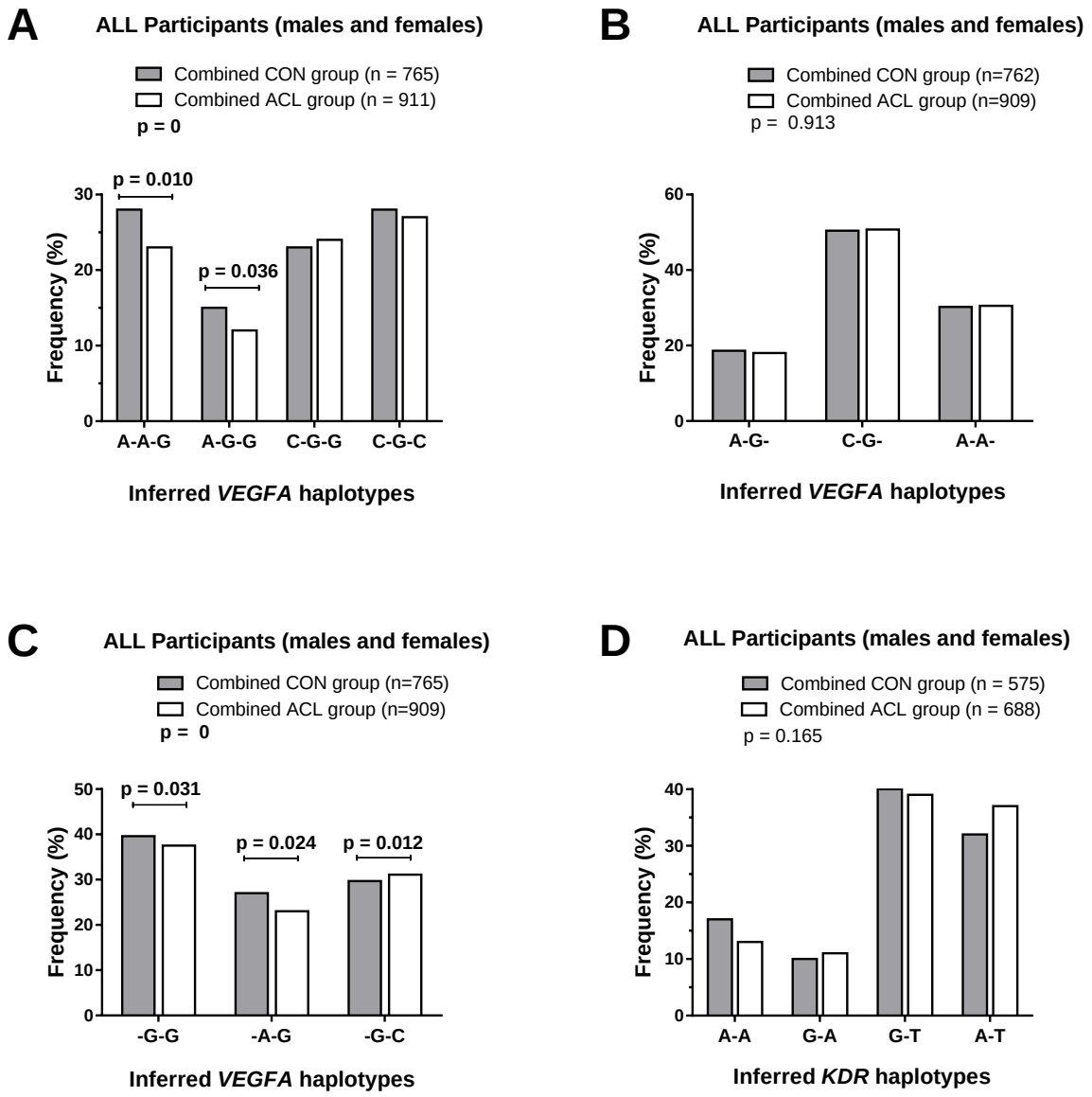


Figure 3.4: Haplotype frequency distributions for the inferred (A) *VEGFA* (rs699947 C/A, rs1570360 G/A rs2010963 G/C), (B) *VEGFA* (rs699947 C/A, rs1570360 G/A), (C) *VEGFA* (rs1570360 G/A, rs2010963 G/C) and (D) *KDR* (rs2071559 A/G, rs1870377 T/A) polymorphisms in the control group (CON; grey bars) and the anterior cruciate ligament rupture group (ACL; white bars) for all participants in the combined Swedish, Polish, Australian, published South African (Rahim et al., 2014) and published Polish (Lulinska-Kuklik et al., 2019a) cohort. Statistically significant differences in haplotype frequency between the combined CON and ACL groups are depicted on the graph. The global p-value for comparison of the combined CON and ACL group is depicted in the figure and the number of participants (n) in each group is in parentheses.

3.3.6 Gene-Gene Interaction Analysis

Allele interactions were explored as a proxy for gene-gene interactions between *VEGFA* (rs699947 A/C, rs2010963 G/C) and *KDR* (rs2071559 A/G, rs1870377 T/A) polymorphisms (**Figure 3.5**). A significant ($p < 0.001$) under-representation of the inferred A-G-A-A ($p = 0.005$, OR: 0.51, 95% CI: 0.30-0.87) and A-G-G-A ($p = 0.018$, OR: 0.93, 95% CI: 0.54-1.60) allele combinations were noted in the COMB ACL (4% and 5%, respectively) group, compared to the COMB CON (9% and 6%, respectively) group.

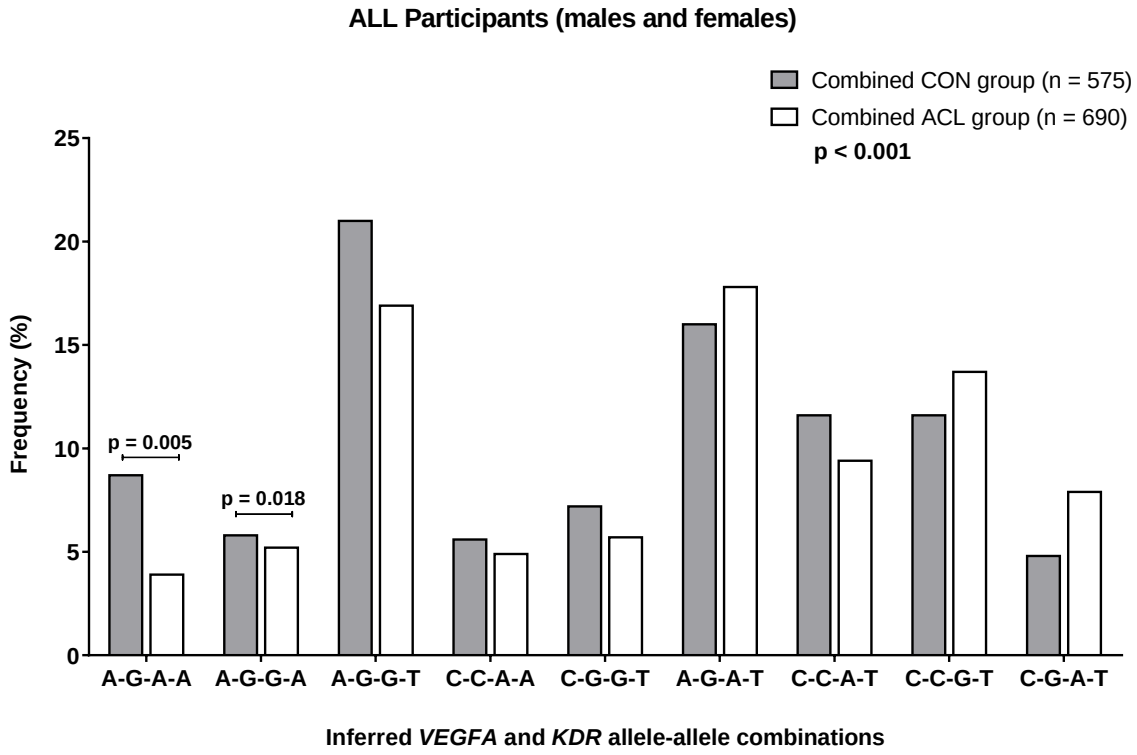


Figure 3.5: Inferred allele-allele frequency distributions for *VEGFA* (rs699947 C/A, rs2010963 G/C) and *KDR* (rs2071559 A/G, rs1870377 T/A) combinations in the control group (CON; grey bars) and the anterior cruciate ligament rupture group (ACL; white bars) for all participants in the Swedish, Polish, Australian, and published South African (Rahim et al., 2014) cohorts. Statistically significant differences in haplotype frequency between the groups are depicted on the graph with an asterisk (*). The number of participants (n) in each group is in parentheses.

3.4 DISCUSSION

It was interesting to note that further exploration of the data from the WGS study highlighted the angiogenesis network of proteins. Taking into account the collective published evidence, part of the elucidation of ACL rupture risk may be through increasing our understanding of the factors contributing to the disrupture of the vasculature. Unravelling this process in tissues such as ligaments, may increase our knowledge of both injury and healing mechanisms (Hauser and Dolan, 2011) in addition to graft ligamentization after ACL reconstructive surgery (Yoshikawa et al., 2006). VEGFA and KDR are two key proteins initiating angiogenesis, and this study aimed to investigate five functional polymorphisms within their encoding genes, *VEGFA* and *KDR* with predisposition to ACL rupture in a large collective data set.

The main findings from the collective cohort of 1,677 samples, included the (i) association of the *VEGFA* rs2010963 CC genotype and C allele with increased risk, (ii) the association of the *VEGFA* A-A-G and A-G-G inferred haplotypes (rs699947 A/C-rs1570360 G/A-rs2010963 G/C) with a reduced risk, which are both in alignment with the *a priori* hypothesis and (iii) the *KDR* genotype and haplotype analyses, illustrated that it is highly unlikely that the investigated *KDR* polymorphisms are associated with modulating non-contact ACL rupture risk predisposition.

Furthermore, when analysing the independent cohorts, associations seemed to be isolated to single cohorts as noted in the cohort from Sweden. This was not a surprise observation, and the reason why the present study had aimed to test the associations in a large combined data set. It was also interesting to note that the *VEGFA* A-G-G inferred haplotype was previously implicated with increased Achilles tendinopathy risk in a combined South African Caucasian

and British study group (Rahim et al., 2016b). This alternate risk allele association between ligament and tendon injury is a thought-provoking finding, and one could hypothesise these differences relate to tissue vascularity and differences in biomechanical properties which are influenced by the regional and load sensitive expression of ECM components (Huisman et al., 2014). Although both tendons and ligaments are poorly vascularized, ligaments have a slightly improved blood supply (Fenwick et al., 2002) and it may be that a predisposition to a lower VEGFA expression may affect these tissues differently, because of their specific healing-load dynamics. The *VEGFA* rs2010963 CC genotype has been associated with increased *VEGFA* gene expression (Awata et al., 2002). Increased VEGFA upregulates the expression of matrix metalloproteinases which are key mediators of ECM turnover during ligament repair (Funahashi et al., 2011). However, research has shown that the overexpression of VEGFA reduces the biomechanical strength of the tendon graft in the early stages of an ACL ligament reconstruction, but is essential in the later phases of ligamentization (Yoshikawa et al., 2006b). Furthermore, in addition to the rs2010963 CC genotype, the rs699947 CC, rs15706630 GG genotypes were also associated with increase in VEGFA expression (Shahbazi et al., 2002) and therefore the alternate alleles with reduced VEGFA expression. The data from this study and previous work (Lulinska-Kuklik et al., 2019; Rahim et al., 2014) suggest that *VEGFA* is associated with altered risk of ACL injury, but there remains much complexity around *VEGFA* gene expression, and its effects on the remodelling continuum.

The *"Goldilocks affect"* hypothesis is therefore still plausible (Nell et al., 2012), whereby a finely tuned homeostatic feedback regulation of turnover of ligament ECM components is required for optimal repair responses. More recently, Willard et al. (2020) and Suijkerbuijk et al. (2019) have shown functional evidence linking a genetic contribution to the expression of ECM components in a susceptibility model.

VEGFA activity relies on a key receptor: KDR, and binding of the receptor is necessary for downstream angiogenesis signalling. No associations were identified for any of the *KDR* polymorphisms or inferred haplotypes in the combined cohort analysis, or in any of the independent participant groups (Swedish, Polish, and Australian). These observations are therefore in agreement with the previously reported lack of independent associations noted for *KDR* and the risk of ACL rupture in a Caucasian study group (Rahim et al., 2014). In contrast, the previously reported inferred *KDR* haplotype association with increased risk of ACL rupture (Rahim et al., 2014) was not observed in this study, and may represent a type I error. KDR is vital for effective VEGFA signalling and altered KDR binding affinity affects the ability of VEGFA to function efficiently. The functional evidence illustrates that the *KDR* rs2071559 G allele is associated with reduced KDR transcription (Wang et al., 2007) resulting in lower levels of KDR and consequently reduced VEGF activity (Hansen et al., 2010). The *KDR* rs1870377 A allele specifically results in an altered binding site for *VEGFA*, which may impair VEGFA binding efficacy (Wang et al., 2007) and can thereby influence downstream VEGFA signalling. This study explored interactions between *VEGFA-KDR*, and the associations noted were exclusively resulting from the alleles inherited at the *VEGFA* locus, with risk being independent of the *KDR* alleles inherited. Furthermore, from the collective data set, the inferred haplotypes identified for the *VEGFA* locus highlights a reduced genetic interval between rs1570360 G/A and rs2010963 G/C to harbour potential functional motifs related to ACL rupture predisposition.

Significant genotype effects were noted on body mass for the *VEGFA* rs699947, and BMI for the *VEGFA* rs1570360 polymorphisms. These findings mirror those observed in a South African population of mixed ancestry (Rahim et al., 2018) where participants in the combined cohort with the rs699947 CC genotype were heavier, and those with the rs1570360 GG

genotype had higher BMIs. These genotypes are associated with increased VEGFA expression (**Section 3.1**) and interestingly, studies have shown that in comparison to normal and underweight individuals, VEGFA serum levels are higher in individuals that are obese (Loebig et al., 2010).

This study represents a large study investigating genetic loci in non-contact ACL rupture predisposition following a case-control genetic association design. One of the limitations of the study was matching participants for confounders, specifically type of sport and frequency of exposure. The authors acknowledge that confounders between the individual cohorts differ and should be considered when pooling cohorts from independent research centres. Going forward, sports participation data in particular should be more uniformly collected, to facilitate the identification of potential confounders and to allow for there adjustment in the analyses. Furthermore, the cases and control numbers were not balanced within the Australian cohort and in particular the control participants sampling would need to increase to facilitate understanding of the frequencies of the variants in this group with reference to the case group. However, when the collective cohort was evaluated, the numbers were more balanced between the groups. In addition, the participants recruited for the independent cohorts were selected from both hospitals and general population sampling.

The data suggest greater variability in the frequency distributions at the functional *VEGFA* polymorphisms between the cohorts. This is most likely not a surprising observation for a biologically relevant genetic interval implicated in modulating a complex phenotype such as ACL ruptures. Future collaborative work requires rigidity in matching participants for sports participation data including type of sport, years of participation, and level of play to determine exposure, as non-genetic factors such as sporting types and frequency of participation do play

a large role in the risk of these injuries. Prospective cohort analyses is required to evaluate the clinical relevance of these genotypes and their contribution to the potential biomechanical properties of ligament and tendon.

3.5 CONCLUSION

VEGFA and KDR are key components in the angiogenesis pathway and altered forms of their genes have downstream effects on their expression. Exploring cohorts from different geographical regions have assisted in supporting the growing body of evidence implicating the *VEGFA* locus in ACL rupture predisposition. Characterisation of the functional biological effects of VEGFA, may in future assist in unravelling the influence of an individual's soft tissue vasculature on the healing and regeneration capacity of ligaments and tendons. Future research should aim to employ a collaborative approach to explore the functional significance of these loci in ligament biology, and large initiatives such as the Athlome Consortium (PMID: 26715623) are required towards this end.

CHAPTER 4

FURTHER EXPLORATION OF POLYMORPHISMS WITHIN THE *ACAN, BGN* AND *DCN* GENES WITH ACL RUPTURE PREDISPOSITION, IN A LARGE COLLECTIVE COHORT

4.1 INTRODUCTION

Similarly to the exploration of angiogenesis genes described in **Chapter 3, Section 3.1**, the WGS study in **Chapter 2** also highlighted proteoglycan encoding genes as inferred functional partners of prioritised WGS candidates (**Appendix B, Section 1, Table S4**). Furthermore, polymorphisms within proteoglycan genes have previously been implicated in ACL rupture risk, carpal tunnel syndrome, and in lumbar disc degeneration (as previously reviewed in **Chapter 1, Section 1.9.3**). Taking the collective evidence, polymorphisms localising to specific proteoglycan genes were further explored in this chapter, in a large, combined analysis with predisposition to ACL rupture.

Mechanical loading of ligament tissue initiates a remodelling response, which is orchestrated via several interconnected pathways to facilitate extracellular matrix homeostasis (**Chapter 1, Section 1.10**). These pathways function as a collective to bring about adaptations to ligament tissue, and the surrounding structure by regulating the synthesis and degradation of the ECM components, which ultimately contribute to maintaining tissue integrity and thereby the biomechanical properties and resilience of ligamentous tissue for future loading (Hauser et al., 2013).

Current research has explored polymorphisms in some of the genes encoding key components of the signalling and remodelling pathways with predisposition to ACL rupture (**Chapter 1, Section 1.9**). Of these, the proteoglycan encoding genes play an important role in maintaining the structural integrity of ligaments by regulating fibrillogenesis, and in regulating the synthesis and degradation (directly or indirectly) of ligament components in remodelling pathways (**Chapter 1, Section 1.9.3**). Interestingly, ECM homeostasis is also regulated through the interactions between key components within proteoglycans encoding genes, and between genes encoding proteoglycans and genes functioning in the angiogenesis associated pathways as described in detail in **Chapter 1, Section 1.10.**

Previous associations with SNPs in proteoglycan genes (*ACAN*, *DCN* and *BGN*) and ACL rupture risk, have been in single populations following a case-control genetic association design (**Chapter 1, Section 1.9.3)**. As these loci are relatively unexplored in the context of ligament rupture predisposition, it is paramount that they are investigated in additional populations of varied ancestry, employing larger sample groups with increased statistical power. In order to improve our understanding of these genetic regions as biologically relevant, in the risk profile of ligament rupture.

The study presented in this chapter therefore aimed to investigate proteoglycan encoding genes (*ACAN*, *DCN* and *BGN*) highlighted in the WGS study, with predisposition to non-contact ACL rupture in a large collective data set. To achieve this aim, the following objectives were included (i) the investigation of the previously implicated *ACAN* (rs2351491 C/T, rs1042631 T/C, rs1516797 T/G) *DCN* rs516115 T/C and *BGN* (rs1126499 C/T, rs1042103 G/A) polymorphisms with predisposition to ACL rupture using a candidate gene, case-control genetic association study design in participants from Sweden and Poland. (ii) The investigation

of *ACAN*, *DCN* and *BGN* polymorphisms in a combined data set, comprising pooled genotype data from the independent cohorts above (Sweden and Poland) and previously published cohorts from South Africa (Mannion et al., 2014) and Poland (Cięszczyk et al., 2017) and (iii) the investigation of allele interactions between *ACAN-BGN*, *ACAN-DCN*, *BGN-DCN* and *VEGFA-DCN* polymorphisms, as a proxy for gene-gene interactions with predisposition to ACL rupture.

Based on previous genetic associations identified in similar study groups, the *a priori* hypothesis was that (i) the *ACAN* rs1516797 GT genotype and G allele would be associated with an increased risk of ACL rupture. (ii) The inferred *ACAN* T-C-T haplotype would be associated with a reduced risk, while the T-C-G haplotype with an increased risk. (iii) The *DCN* rs516115 GG and AA genotypes would be associated with a reduced and increased risk respectively, and (iv) the *BGN* rs1042103 A allele would be associated with an increased risk, with the *BGN* C-G inferred haplotype associated with a reduced risk of ACL rupture.

4.2 METHODS

4.2.1 Participant Characteristics

This study followed a case-control genetic association design, investigating the Swedish and Polish cohorts from **Chapter 3**. All participant characteristics for the Swedish (SWE CON: 116, SWE ACL: 95) and Polish (POL CON: 149, POL ACL: 127) cohorts, in addition to ethical approval are described in detail in **Chapter 3, Section 3.2.1**. For the combined cohort analyses, genotype data from the current Swedish and Polish cohorts were pooled with genotype data from previously published South African (CON: 234, ACL: 126) (Mannion et al., 2014) and Polish (CON: 284 ACL: 229) (Cięszczyk et al., 2017) cohorts, with approval from the respective authors. As previously described (**Chapter 3, Section 3.2.1**) only participants who had ruptured their ACL through non-contact mechanisms were included in the study, and the same criteria was applied for the additional cohorts from South Africa (Mannion et al., 2014) and Poland (Cięszczyk et al., 2017).

ACAN rs2351491 and rs1042631 polymorphisms were not previously investigated in the published Polish study group (Cięszczyk et al., 2017). Therefore, for the combined analyses of these two SNPs, the collective genotype data included that of the Swedish, Polish, and published South African (Mannion et al., 2014) cohorts to total 492 COMB CON and 348 COMB ACL participants. For *ACAN* rs1516797, *BGN* rs1126499, *BGN* rs1042103 and *DCN* rs516115 polymorphisms, the genotype data from all the cohorts [Swedish, Polish, published South African (Mannion et al., 2014) and published Polish (Cięszczyk et al., 2017)] was combined to total 776 COMB CON, and 577 COMB ACL samples.

4.2.2 DNA Isolation

The isolation of DNA from the Swedish and Polish participant samples was performed as previously described in **Chapter 3, Section 3.2.2**.

4.2.3 SNP Selection and Genotyping

Similar to the angiogenesis single nucleotide polymorphisms selected in **Chapter 3, Section 3.2.3**. Proteoglycan SNPs were also selected based on biological significance of the encoding genes, and previous associations with multifactorial phenotypes. All Swedish and Polish participant samples were genotyped as described in **Chapter 3, Section 3.2.3,** using custom-designed fluorescence-based TaqMan™ PCR assays (Applied Biosystems, Foster City, CA, USA) as per manufacturer's instructions. Six SNPs within three proteoglycan genes were investigated: *ACAN* rs2351491 T/C (assay ID: C__25474736_10), *ACAN* rs1042631 C/T (C___8722250_1_), *ACAN* rs1516797 T/G (C____331789_10), *BGN* rs1126499 C/T (C___2617574_1_), *BGN* rs1042103 G/A (C___8898142_10) and *DCN* rs516115 T/C (C___2309580_10) (**Figure 4.1A-C**). In addition, the same stringency of controls were included in all PCR assays as described in **Chapter 3, Section 3.2.3**.

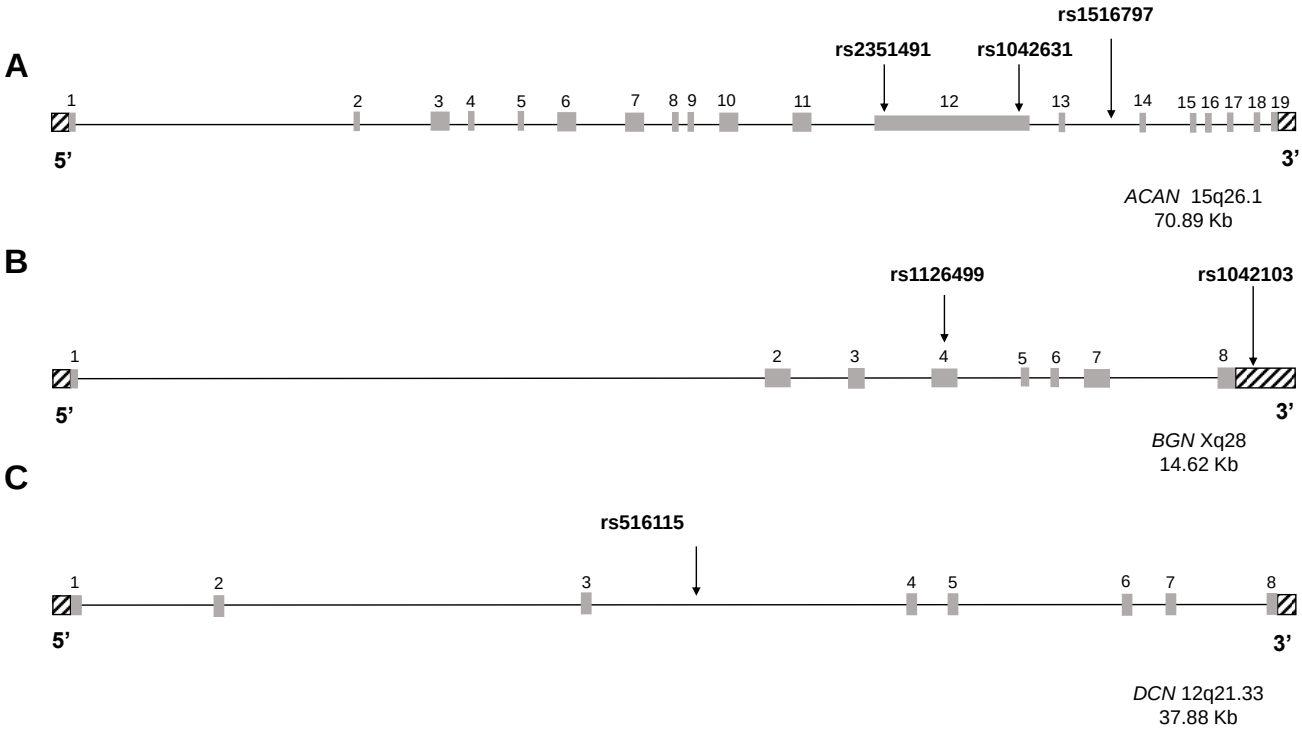


Figure 4.1: Schematic representation of the (A) *ACAN* depicting the rs2351491 (C/T), rs1042631 (T/C) and rs1516797 (T/G) polymorphisms; (B) *BGN* depicting the rs1126499 (C/T) and rs1042103 (G/A) polymorphisms, and (C) *DCN* depicting the rs516115 (T/C) polymorphism. Exons are shown as grey boxes (with exon number above), introns as horizontal lines, and untranslated regions are depicted as hatched boxes. The size of the genes are indicated in Kb and their chromosomal locations are indicated. Figures are not drawn to scale. All the information used to construct the figure was obtained from databases hosted by the National Centre for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/) and Ensembl Genome Data Centre (http://www.ensembl.org/index.html).

4.2.4 Statistical Analyses

For power calculations and all statistical analyses refer to **Chapter 3, Section 3.2.4**. The *BGN* gene is located on the X-chromosome. Therefore genotype and allele frequencies, as well as genotype effects on descriptive participant characteristics, were compared separately for males and females. In the combined cohort, analyses were conducted in all participants (males and females) and in males and females separately. Due to reduced power, stratification by sex was not carried out in the independent cohorts.

4.3 RESULTS

4.3.1 Participant Characteristics

Participants in the COMB CON group differed in sex, body mass and BMI compared to participants in the COMB ACL group (**Table 4.1**). There were significantly more males in the COMB ACL group (67%) compared to the COMB CON (57%) group. The COMB ACL group were significantly ($p < 0.001$) heavier (78.5 ± 14.9 kg, n = 337) than the COMB CON (74.0 ± 13.7 kg, n = 478) group (**Table 4.1**) even when adjusted for sex ($p = 0.001$). Although on average participants in the CON and ACL groups had normal BMIs, the COMB ACL group had significantly ($p < 0.001$, adjusted $p < 0.001$) higher BMIs (25.4 ± 5.9, n = 328) than participants in the COMB CON (23.7 ± 3.0, n = 475) group (**Table 4.1**). The descriptive participant characteristics and the sports participation data for the Swedish (n = 211) and Polish (n = 276) cohorts were previously described in **Chapter 3, Section 3.3**.

Table 4.1: Descriptive characteristics for all participants (males and females) in the combined control (CON) group and the combined anterior cruciate ligament rupture (ACL) group.

	COMB CON	**COMB ACL**	**p-value**
	n = 776	n = 577	
Age (years)	30.2 ± 12.9 (485)	31.3 ± 11.9 (340)	0.270
Sex (%male)	57% (776)	67% (577)	**< 0.001**
Height (cm)	175.3 ± 9.9 (478)	176.1 ± 10.7 (331)	0.284
Body Mass (kg)	73.04 ± 13.7 (478)	78.5 ± 14.9 (337)	**<0.001** (**0.001**)[a]
BMI (kg.m^{-2})	23.7 ± 3.0 (475)	25.4 ± 5.9 (328)	**<0.001** (**<0.001**)[b]

Values are expressed as mean ± standard deviation; sex is represented as a percentage. The number of participants (n) with available data for each variable is in parentheses. CON vs ACL, p-values in **bold** typeset indicates significance ($p < 0.05$). P-values are depicted as unadjusted and adjusted in parentheses. a: adjusted by sex, b: adjusted by sex and body mass.

4.3.2 Genotype Effects

In the combined analysis, *ACAN* rs2351491 was associated with age (**Table 4.2**). Participants with the rs2351491 CC genotype were significantly younger (27.5 ± 10.3 years, n = 105, p = 0.0013) than participants with the TT (30.6 ± 12.1 years, n = 331) and CT (31.8 ± 13.2 years, n = 378) genotypes (**Appendix B, Section 3, Table S18)**. In male participants, the *BGN* rs1126499 and rs1042103 polymorphisms were associated with height and age, respectively (**Table 4.2**). Males with the rs1126499 CC genotype were significantly taller (181.7 ± 6.9 cm, n = 250, p = 0.048) than males with the TT (180.2 ± 9.3 cm, n = 234) genotype. In addition, male participants with the rs1042103 GG genotype were significantly older (29.3 ± 12.1 years, n = 300, p = 0.049) than those with the AA (27.9 ± 10.4 years, n = 197) genotype (**Appendix B, Section 3, Table S18**). No significant genotype effects were noted for any of the *BGN* polymorphisms investigated in female participants (**Table 4.2**).

In all Swedish participants, *DCN* rs516115 (T/C) was associated with sex, height and body mass (**Appendix B, Section 3, Table S19**). The rs516115 CC genotype was significantly under-represented (p = 0.018) in the male participants (1%, n = 1) compared to the female participants (9%, n = 11). For height, participants with the rs516115 CC genotype were significantly shorter in stature (164.0 ± 6.7 cm, n = 12, p = 0.001) than those with either the CT (174.7 ± 9.5 cm, n = 61) or TT (172.1 ± 9.0 cm, n = 119) genotypes. Additionally, participants with the rs516115 CT genotype were significantly heavier (76.5 ± 11.3 kg, n = 62, p = 0.006) than those with the CC (65.1 ± 7.7 kg, n = 11) or TT (71.6 ± 13.9 kg, n = 119) genotypes. In male participants, the *BGN* rs1126499 (C/T) polymorphism was associated with body mass (**Appendix B, Section 3, Table S19**). Where males with the CC genotype were

significantly heavier (86.2 ± 10.3 kg, n = 41, p = 0.033) than those with the TT (81.6 ± 7.5 kg, n = 34) genotype. In Polish participants, no significant genotype effects were noted on age, sex, height, body mass or BMI for any of the *ACAN* or *DCN* polymorphisms investigated in all participants, or on age, height, body mass or BMI for the *BGN* polymorphisms, investigated in the male and female participants seperately (**Appendix B, Section 3, Table S20**).

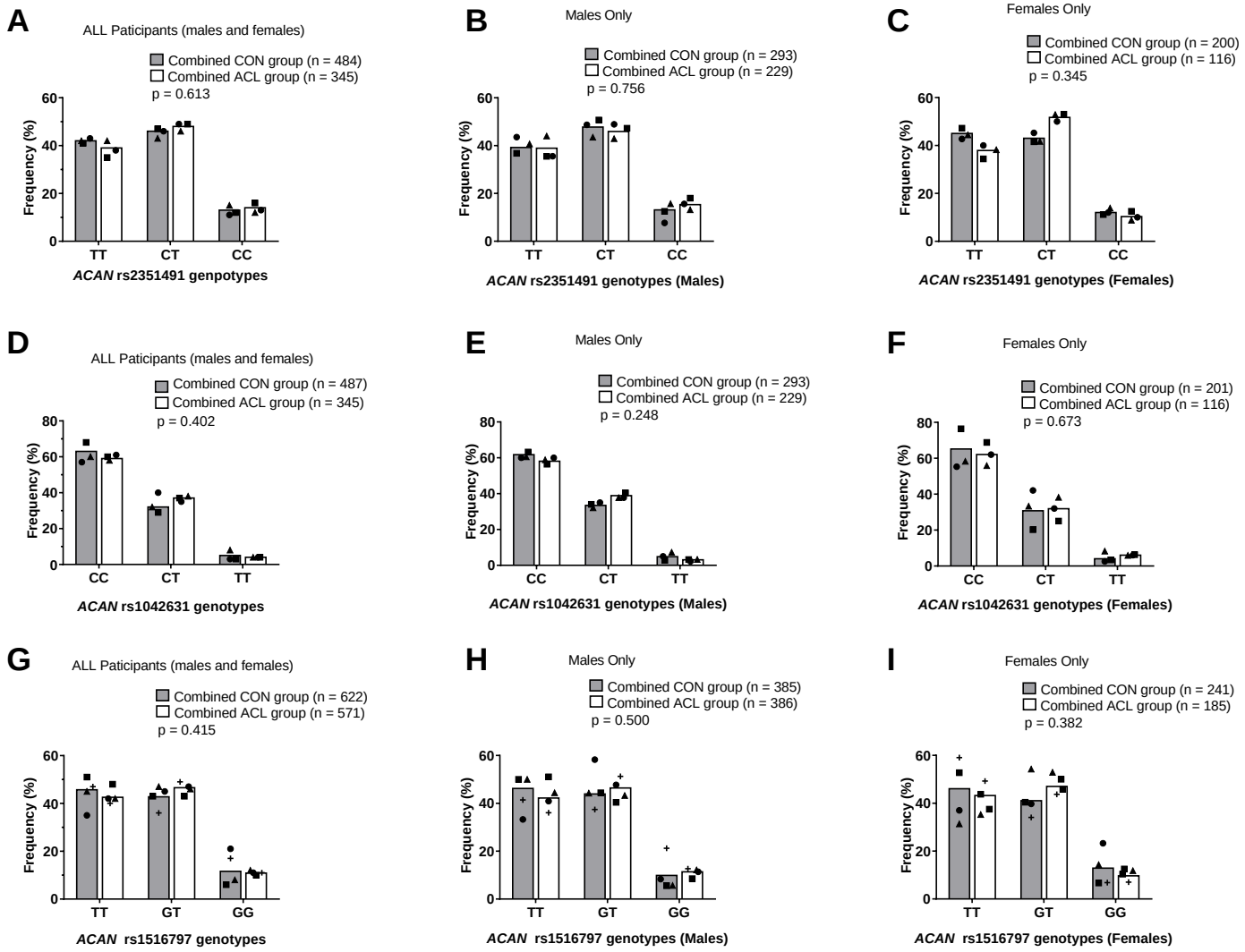
Table 4.2: Genotype associations with descriptive measures for the *ACAN* (rs2351491 C/T, rs1042631 T/C, rs1516797 T/G) and *DCN* (rs516115 T/C) polymorphisms in all participants (males and females) and the *BGN* (rs1126499 C/T, rs1042103 G/A) polymorphisms in male and female participants separately, in the combined cohort. P-values in bold typeset indicates significance ($p < 0.05$).

Polymorphism	Age	Sex	Body Mass	Height	BMI
ACAN rs2351491	**0.013**	0.445	0.533	0.620	0.638
ACAN rs1042631	0.619	0.455	0.495	0.994	0.458
ACAN rs1516797	0.731	0.907	0.129	0.264	0.334
DCN rs516115	0.183	0.432	0.454	0.080	0.723
Males *BGN* rs1126499	0.900	-	0.576	**0.048**	0.293
Males *BGN* rs1042103	**0.049**	-	0.378	0.274	0.399
Females *BGN* rs1126499	0.480	-	0.935	0.839	0.824
Females *BGN* rs1042103	0.760	-	0.293	0.289	0.244

4.3.2 Genotype and Allele Frequencies

In the combined cohort, no significant differences in the genotype or allele frequencies were noted between the COMB CON and COMB ACL groups, for the investigated *ACAN* 2351491 C/T in all participants (genotype p = 0.613 and allele p = 0.378) males (genotype p = 0.756 and allele p = 0.649) and females (genotype p = 0.345 and allele p = 0.488), *ACAN* 1042631 C/T in all participants (genotype p = 0.402 and allele p = 0.395) males (genotype p = 0.245 and allele p = 0.649) and females (genotype p = 0.673 and allele p = 0.474) and *ACAN* 1516797 T/G in all participants (genotype p = 0.415 and allele p = 0.544) males (genotype p = 0.500 and allele p = 0.418) and female (genotype p = 0.382 and allele p = 0.942) participants (**Figure 4.2** and **Appendix B, Section 3, Table S21**). Furthermore, no significant differences in the genotype or allele frequencies were noted between the COMB CON and COMB ACL groups for *DCN* rs516115 T/C in all participants (genotype p = 0.716 and allele p = 0.484) males (genotype p = 0.989 and allele p = 0.912) and female (genotype p = 0.291 and allele p = 0.125) participants (**Figure 4.3** and **Appendix B, Section 3, Table S21**).

Additionally, no significant differences in the genotype or allele frequency distributions were noted for *BGN* rs1126499 C/T in males (genotype p = 0.243 and allele p = 0.099) and in females (genotype p = 0.308 and allele p = 0.165) or for *BGN* rs1042103 G/A in males (genotype p = 0.893 and allele p = 0.879) and female (genotype p = 0.383 and allele p = 0.894) participants (**Figure 4.4** and **Appendix B, Section 3, Table S21**). All investigated polymorphisms for the COMB CON and COMB ACL groups were in HWE.



See overleaf for figure legend.

Figure 4.2: Genotype frequency distributions for the *ACAN* rs2351491 (C/T) polymorphism in (A) all participants, (B) males and (C) females, for *ACAN* rs1042631 (T/C) in (D) all, (E) males and (F) females; and *ACAN* rs1516797 (T/G) in (G) all, (H) males, and (I) females in the combined control (CON) group (grey bars) and the combined anterior cruciate ligament rupture (ACL) group (white bars). Genotype frequency distribution differences between the individual Swedish (circles), Polish (triangles), published South African (squares) (Mannion et al., 2014) and published Polish (crosses) (Cięszczyk et al., 2017) cohorts are: CON groups; rs2351491 All participants $p = 0.789$, Males $p = 0.621$, Females $p = 0.974$, rs1042631 All $p = 0.054$, Males $p = 0.574$, and Females $p = 0.020$, and rs1516797 All $p = 0.0002$, Males $p = 0.574$, and Females $p = 0.008$, and ACL groups; rs2351491 All participants $p = 0.756$, Males $p = 0.0002$, Females $p = 0.979$, rs1042631 All $p = 0.756$, Males $p < 0.001$, and Females $p = 0.853$, and rs1516797 All $p = 0.998$, Males $p = 0.427$, and Females $p = 0.835$. Statistically significant differences in genotype frequencies between the combined CON and ACL groups are depicted on the graph. The chi-squared global p-value for comparison of the combined CON and combined ACL groups is depicted in the figure.

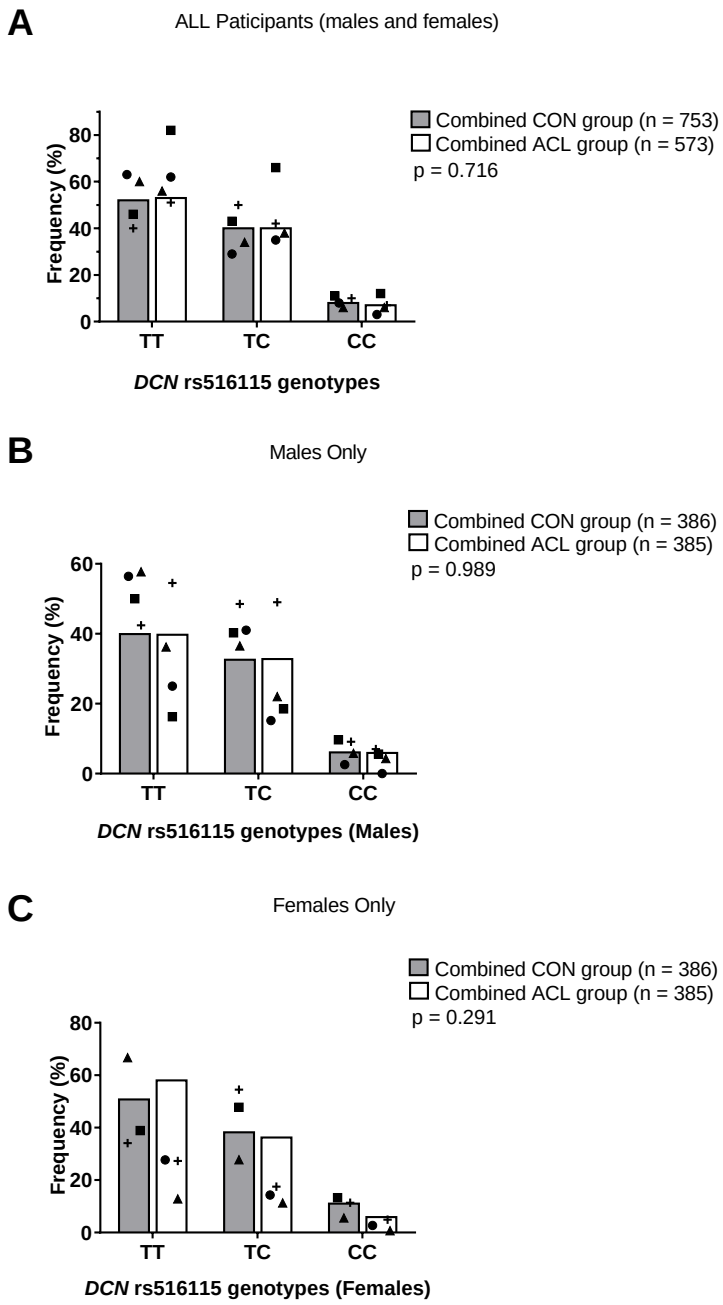


Figure 4.3: Genotype frequency distributions for the *DCN* rs516115 (T/C) polymorphism in (A) all participants, (B) males and (C) females in the combined control (CON) group (grey bars) and the combined anterior cruciate ligament rupture (ACL) group (white bars) for all participants. Genotype frequency distribution differences between the individual Swedish (circles), Polish (triangles), published South African (squares) (Mannion et al., 2014) and published Polish (crosses) (Cięszczyk et al., 2017) cohorts are: CON groups; rs516115 All participants p = 0.001, Males p = 0.297, and Females p = 0.001, and ACL groups; rs516115 All p = 0.298, Males = 0.025, and Females = 0.384. Statistically significant differences in genotype frequencies between the combined CON and ACL groups are depicted on the graph. The chi-squared global p-value for comparison of the combined CON and combined ACL groups is depicted in the figure.

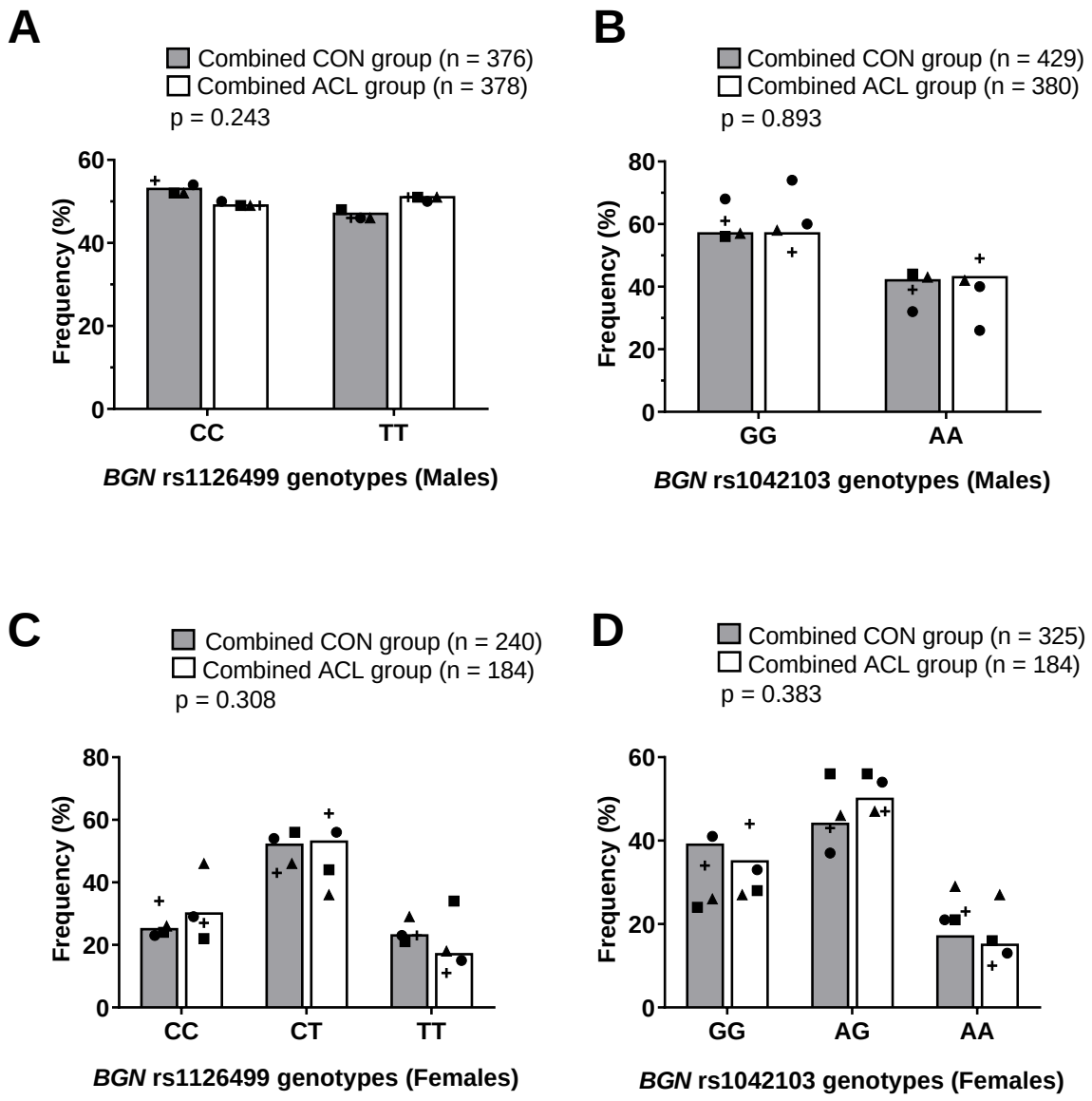


Figure 4.4: Genotype frequency distributions for the *BGN* (A) rs1126499 Males (C/T), (B) *BGN* rs1042103 Males (G/A), *BGN* (C) rs1126499 Females (C/T), and (D) *BGN* rs1042103 Females (G/A) polymorphisms in the combined control (CON) group (grey bars) and the combined anterior cruciate ligament rupture (ACL) group (white bars) for all participants. Genotype frequency distribution differences between the individual Swedish (circles), Polish (triangles), published South African (squares) (Mannion et al., 2014) and published Polish (crosses) (Cięszczyk et al., 2017) cohorts are: CON groups; rs1126499 Males p = 0.001, rs1126499 Females p = 0.748, rs1042103 Males p = 0.612, and rs1042103 Females p = 0.666, and ACL groups; rs1126499 Males p = 0.002, rs1126499 Females p = 0.029, rs1042103 Males p = 0.055, and rs1042103 Females p = 0.247. Statistically significant differences in genotype frequencies between the combined CON and ACL groups are depicted on the graph. The chi-squared global p-value for comparison of the combined CON and combined ACL groups is depicted in the figure.

In all participants from Sweden, no significant differences were noted in the genotype or allele frequency distributions between the SWE CON and SWE ACL groups for the *ACAN* rs2351491 C/T (genotype p = 0.646 and allele p = 0.473) and *ACAN* rs1042631 C/T (genotype p = 0.887 and allele p = 0.726) polymorphisms (**Appendix B, Section 3, Table S22**). However, there was a trend for the *ACAN* rs1516797 (T/G) GG genotype (p = 0.059) and G allele (p = 0.083) to be under-represented in the SWE ACL group, compared to the SWE CON (GG: 11% vs 21%; G: 34% vs 43%) group (**Appendix B, Section 3, Table S22**). Furthermore, no significant differences in the genotype or allele frequencies between the SWE CON and SWE ACL groups (genotype p = 0.144 and allele p = 0.721) were noted for *DCN* rs516115 T/C in all participants (**Appendix B, Section 3, Table S22**). For *BGN* rs1042103 G/A, no significant differences in the genotype or allele frequencies (genotype p = 0.366 and allele p = 0.483) were noted in males. However, a trend (p = 0.062) for the rs1042103 AG genotype to be over-represented in the SWE ACL (54%) group, compared to the SWE CON (37%) group was noted in females (**Appendix B, Section 3, Table S22**). No significant differences in the genotype or allele frequency distributions were noted in male or female participants for the *BGN* rs1126499 C/T polymorphism (**Appendix B, Section 3, Table S22**). All *ACAN* and *DCN* investigated polymorphisms in all participants, and *BGN* polymorphisms in female participants in the SWE CON and SWE ACL groups were in HWE (**Appendix B, Section 3, Table S22**).

In all Polish participants, no significant differences in the genotype or allele frequency distributions between the POL CON and POL ACL groups were noted for the *ACAN* rs2351491 C/T (genotype p = 0.549 and allele p = 0.653) and *ACAN* rs1516797 T/G (genotype p = 0.738 and allele p = 0.356) polymorphisms (**Appendix B, Section 3, Table S22**). However, for the *ACAN* rs1042631 (C/T) polymorphism, a significant difference in the genotype frequencies (global p = 0.010, FDR p = 0.040) between the groups was noted. Specifically, the rs1042631

TT genotype was significantly under-represented (p = 0.020, OR: 0.17, 95% CI: 0.03 - 0.90) in the POL ACL (4%) group, compared to the POL CON group (8%) and the rs1042631 CT genotype significantly over-represented (p = 0.019, OR: 2.29, 95% CI: 1.14 - 4.60) in the POL ACL (38%) group, compared to the POL CON (32%) group in all participants (**Appendix B, Section 3, Table S22**).

For *DCN* rs516115 T/C, no significant differences in the genotype or allele frequency distributions were noted (genotype p = 0.799 and allele p = 0.609) in all participants (**Appendix B, Section 3, Table S22**). Furthermore, no significant differences in the genotype or allele frequency distributions were noted for *BGN* rs1126499 C/T in males (genotype p = 0.732 and allele p = 0.297). However, a trend (p = 0.086) was noted for rs1126499 T allele to be under-represented in the female ACL group (36%, n = 24) compared to the female CON group (51%, n = 36). No significant differences in the genotype or allele frequency distributions were noted for *BGN* rs1042103 G/A in male (genotype p = 0.273 and allele p = 0.837) or female (genotype p = 0.242 and allele p = 0.237) participants (**Appendix B, Section 3, Table S22**). All *ACAN* and *DCN* investigated polymorphisms in all participants, and *BGN* polymorphisms in female participants in the POL CON and POL ACL groups were in HWE (**Appendix B, Section 3, Table S22**).

The frequency distributions were significantly different between the Swedish, Polish, and published South African female control groups for *ACAN* rs1042631 (p = 0.020) and all participants and females for rs1516797 (p = 0.0002). For the ACL groups, the frequency distributions were significantly different in males for the rs2351491 (p = 0.0002) and rs1042631 (p < 0.001) polymorphisms (**Figure 4.2**). For *DCN* rs516115, a significant difference in the frequency distributions were noted for all participants (p = 0.001) and females (p = 0.001) in

the control groups, and for the males (p = 0.025) in the Swedish, Polish, published South African and published Polish ACL groups. Furthermore, for the *BGN* SNPs, a significant difference in the frequency distributions between the male control groups were noted for rs1126499 (p = 0.001). Additionally, the frequency distributions were different between the male ACL groups for rs1126499 (p = 0.002) and female ACL (p = 0.029) groups (**Figure 4.4**) and the frequency distributions for the combined control and ACL groups were comparable with the HapMap central European (CEU) reported frequencies (Yates et al., 2019) for all the investigated polymorphisms.

4.3.3 Inferred Haplotypes

Inferred haplotypes were constructed using the genotype data for the three *ACAN* (rs2351491 C/T, rs1042631 C/T and rs1516797 T/G) and two *BGN* (rs1126499 C/T, rs1042103 G/A) polymorphisms. Only haplotypes inferred at a frequency greater than 2% were compared.

No significant differences in the haplotype frequencies between the COMB CON and COMB ACL groups were noted for inferred *ACAN* haplotypes, when all participants (**Figure 4.5A**) or when males only (**Figure 4.5B**) and females only (**Figure 4.5C**) were considered. Furthermore, no significant differences in the haplotype frequencies were noted between the COMB CON and COMB ACL groups for the the inferred *BGN* haplotypes, when males (**Figure 4.6A**) and females (**Figure 4.6B**) were considered separately. In all Swedish participants, no significant differences in the haplotype frequencies were noted between the SWE CON and SWE ACL groups for inferred *ACAN* haplotypes (**Appendix B, Section 3, Figure S5A**). Furthermore, no significant differences in haplotype frequencies for inferred *BGN* haplotypes were noted between the SWE CON and SWE ACL groups, in male (**Appendix B, Section 3, Figure S5B**)

or female (**Appendix B, Section 3, Figure S5C**) participants. In participants from Poland, no significant differences in the haplotype frequencies were noted between the POL CON and POL ACL groups for inferred *ACAN* haplotypes (**Appendix B, Section 3, Figure S5D**). However, the inferred *BGN* C-A haplotype was significantly ($p = 0.012$, haplo.score: 2.51, OR: 3.62, 95% CI: 0.76-17.21) over-represented in the female POL ACL group (18%) compared to the female POL CON (3%) group, and the *BGN* T-G haplotype significantly ($p = 0.009$, haplo.score: -2.61, OR: 0.19, 95% CI: 0.03-1.03) under-represented in the female POL ACL (4%) group, compared to the female POL CON (16%) group (**Appendix B, Section 3, Figure S5F**).

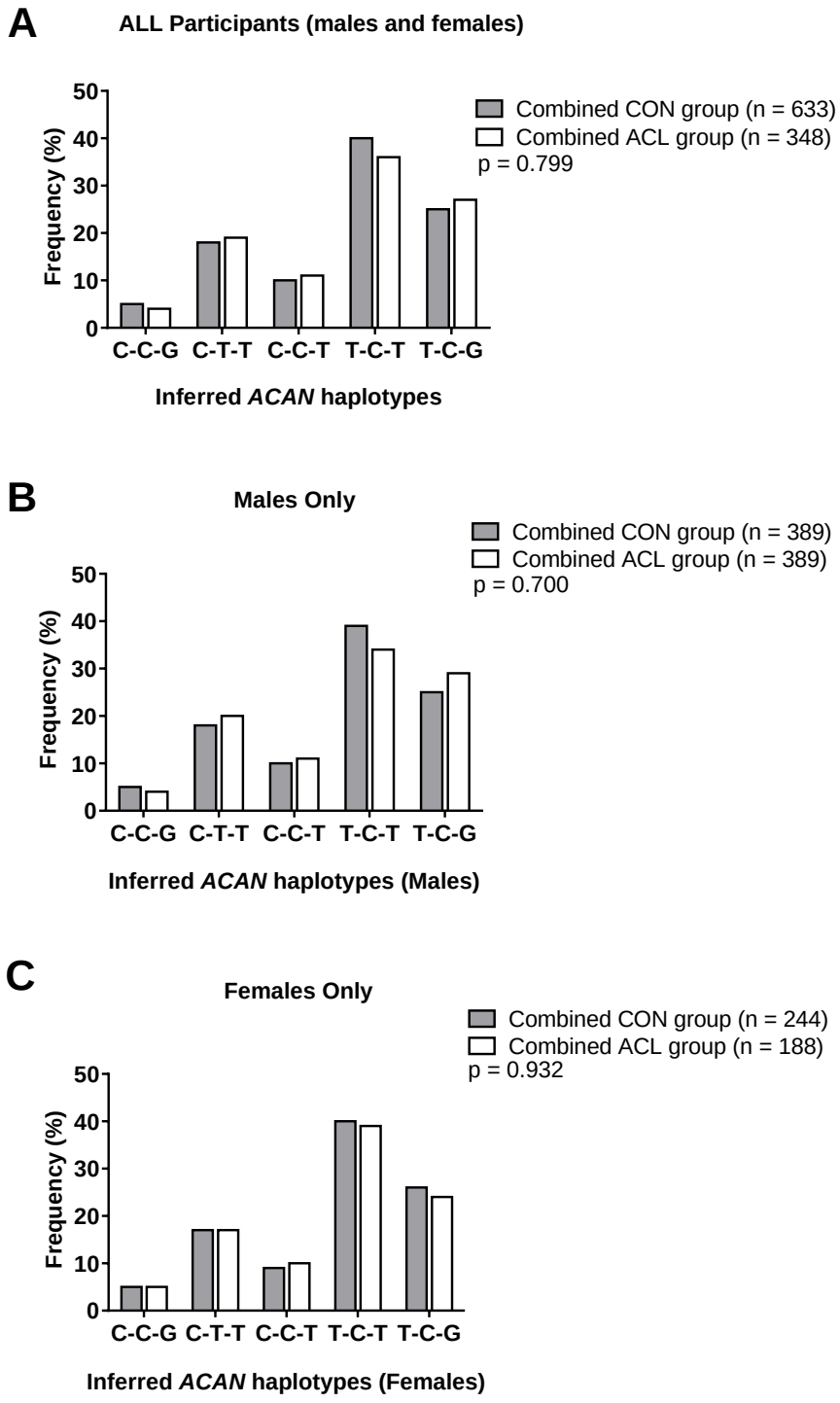


Figure 4.5: Haplotype frequency distributions for the inferred *ACAN* (rs2351491 C/T, rs1042631 T/C and rs1516797 T/G) polymorphisms in (A) all participants, (B) Males and (C) Females in the combined control group (CON; grey bars) and the anterior cruciate ligament rupture group (ACL; white bars) in the combined Swedish, Polish, and published South African (Mannion et al., 2014) cohort. The global p-value for comparison of the combined CON and ACL group is depicted in the figure and the number of participants (n) in each group is in parentheses.

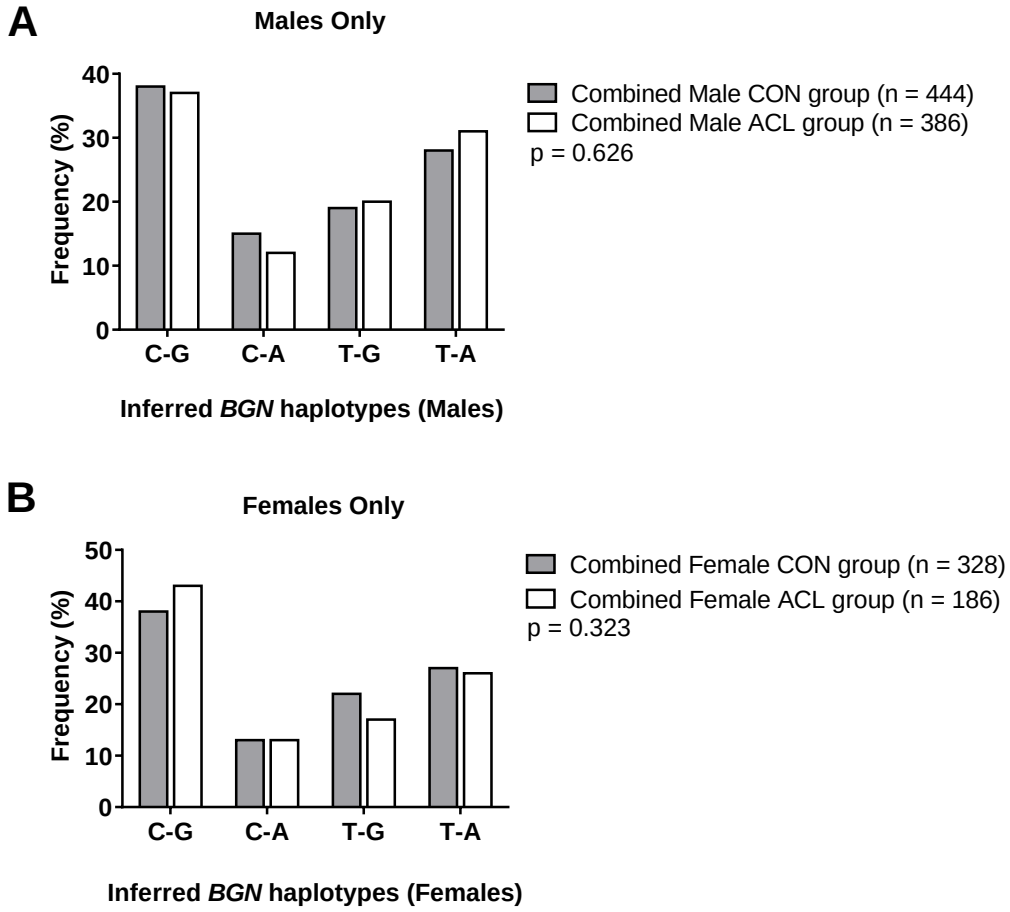


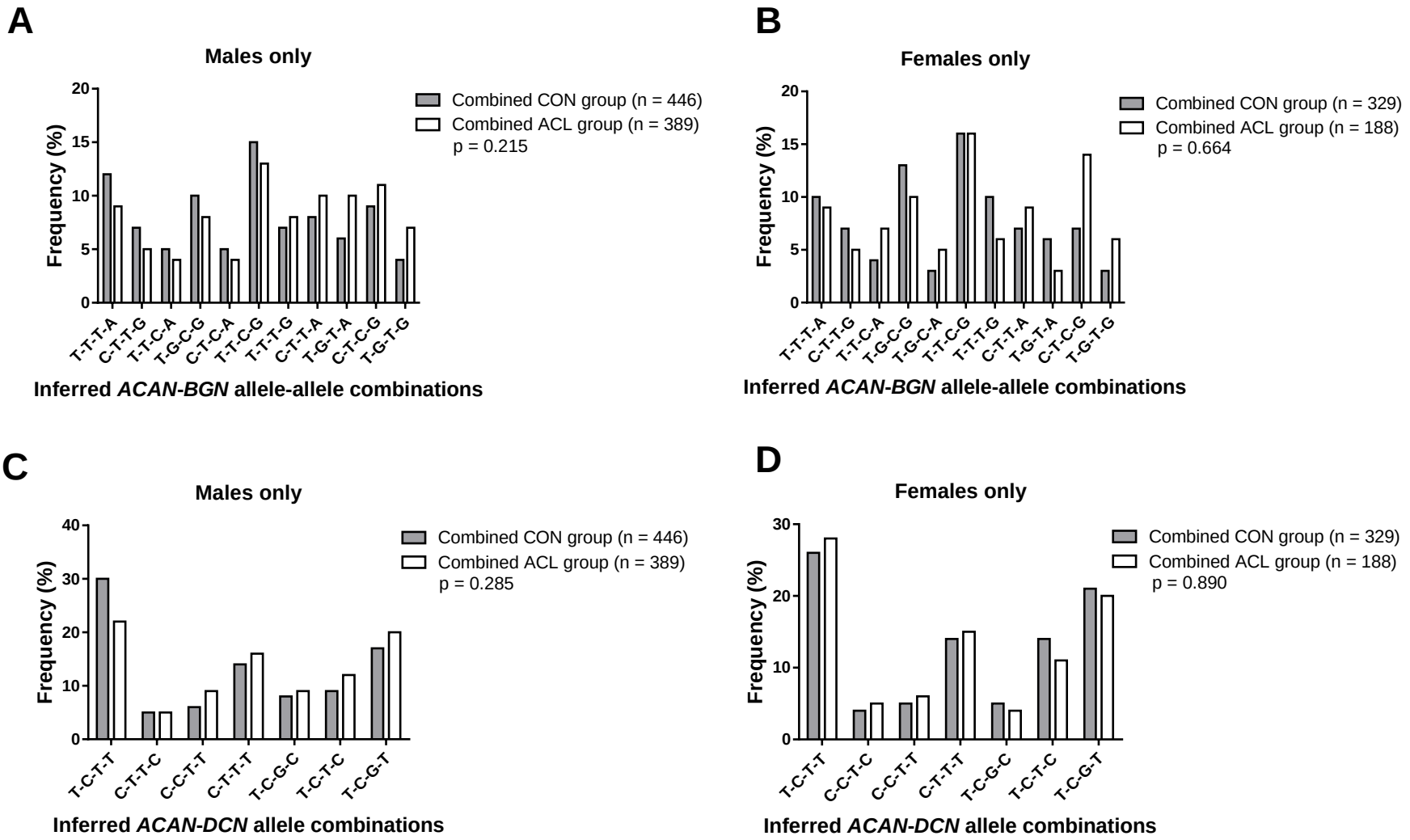
Figure 4.6: Haplotype frequency distributions for the inferred *BGN* (rs1126499 C/T – rs1042103 G/A) polymorphisms in (A) Males and (B) Females in the control group (CON; grey bars) and the anterior cruciate ligament rupture group (ACL; white bars) in the combined Swedish, Polish, published South African (Mannion et al., 2014) and published Polish (Cięszczyk et al., 2017) cohort. The global p-value for comparison of the combined CON and ACL group is depicted in the figure and the number of participants (n) in each group is in parentheses.

4.3.4 Gene-Gene Interaction Analysis

Allele interactions were explored as a proxy for gene-gene interactions using the genotype data. Gene-gene interaction analyses between the proteoglycan encoding genes revealed no significant differences between groups for distributions in the allele combinations of *ACAN* (rs2351491 C/T, rs1516797 T/G) and *BGN* (rs1126499 C/T, rs1042103 G/A) (**Figure 4.7A**: Males and **Figure 4.7B**: Females; *ACAN* (rs2351491 C/T, rs1042631 C/T, rs1516797 T/G) and *DCN* (rs516115 T/C) (**Figure 4.7C**: Males and **Figure 4.7D**: Females; *BGN* (rs1126499 C/T, rs1042103 G/A) and *DCN* (rs516115 T/C) (**Figure 4.8A**: Males and **Figure 4.8B**: Females) and risk of ACL rupture.

Gene-gene interactions between the genes in the angiogenesis associated pathway, and the proteoglycan encoding genes, showed significant ($p < 0.001$) differences in the distribution of allele combinations for *VEGFA* (rs699947 C/A, rs1570360 G/A, rs20109630 G/C) and *DCN* (rs516115 T/C) polymorphisms (**Figure 4.9A**) with an over-representation of the inferred A-G-G-T ($p = 0.010$, OR: 0.53, 95% CI: 0.30-0.91), A-A-G-C ($p = 0.010$, OR: 0.42, 95% CI: 0.21-0.81) and A-A-G-T ($p = 0.046$, OR: 0.77, CI: 0.49-1.2) combinations as noted in the COMB male CON (15%, 10% and 22%, respectively) group, compared to the COMB male ACL (9%, 4% and 19%, respectively) group. A significant ($p < 0.001$) association was also noted for *VEGFA-DCN* allele combinations in the COMB female cohort (**Figure 4.7B**). However, posthoc analysis revealed no significant associations.

.



See overleaf for figure legend.

Figure 4.7: Inferred allele-allele frequency distributions for *ACAN* (rs2351491 C/T, rs1516797 T/G) and *BGN* (rs1126499 C/T, rs1042103 G/A) combinations; and *ACAN* (rs2351491 C/T, rs1042631 C/T, rs1516797 T/G) and *DCN* (rs516115 T/C) combinations in the control group (CON; grey bars) and the anterior cruciate ligament rupture group (ACL; white bars) for male (A&C) and female (B&D) participants in the combined Swedish, Polish, published Polish (Cięszczyk et al., 2017) and South African (Rahim et al., 2014) cohorts. The number of participants (n) in each group is in parentheses.

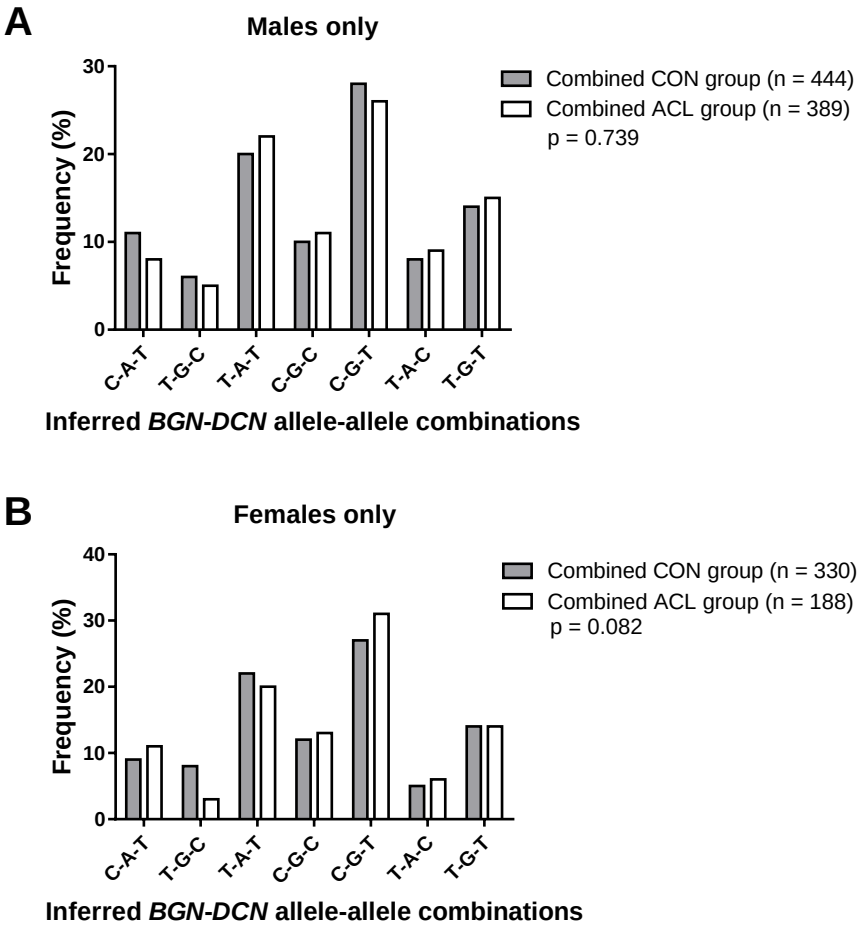


Figure 4.8: Inferred allele-allele frequency distributions for *BGN* (rs1126499 C/T, rs1042103 G/A) and *DCN* (rs516115 T/C) combinations in the control group (CON; grey bars) and the anterior cruciate ligament rupture group (ACL; white bars) for male (A) and female (B) participants in the Swedish, Polish, published Polish (Cięszczyk et al., 2017) and South African (Rahim et al., 2014) cohorts. The number of participants (n) in each group is in parentheses.

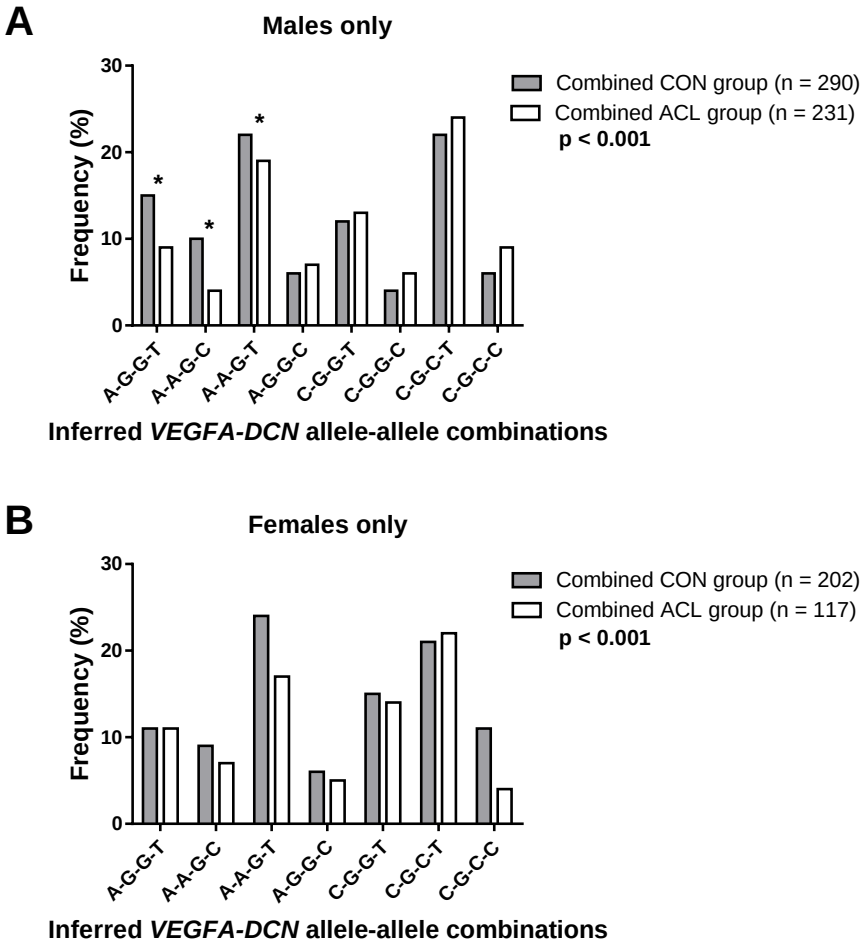


Figure 4.9: Inferred allele-allele frequency distributions for *VEGFA* (rs699947 C/A, rs1570360 G/A, rs2010963 G/C) and *DCN* (rs516115 T/C) combinations in the control group (CON; grey bars) and the anterior cruciate ligament rupture group (ACL; white bars) for A: Male and B: Female participants in the Swedish, Polish, and published South African (Mannion et al., 2014; Rahim et al., 2014) cohorts. Statistically significant differences in allele-allele frequencies between the groups are depicted on the graph with an asterisk (*). The number of participants (n) in each group is in parentheses

4.4 DISCUSSION

Analyses of the functional partners identified in the WGS study, highlighted several links with the proteoglycan encoding genes as potentially important in ACL rupture predisposition. Furthermore, polymorphisms within proteoglycan genes were previously associated with ACL rupture cases in a South African Caucasian (Mannion et al., 2014) and Polish cohort (Cięszczyk et al., 2017). From these collective findings, and in light of the key functions mediated by aggrecan, decorin, and biglycan in ligamentous tissue, the aim of this study was to investigate the previously associated proteoglycan polymorphisms for an association with the risk of ACL rupture in a large cohort, which included data from previously genotyped cohorts from South Africa (Mannion et al., 2014) and Poland (Cięszczyk et al., 2017) and including two additional independent populations from Sweden and Poland. Furthermore, the study aimed to explore gene-gene interaction within the proteoglycan family of genes, and between the angiogenesis associated and proteoglycan genes highlighted in **Chapter 2**'s WGS study, with predisposition to ACL rupture.

The novel finding from the collective cohort included the association of *VEGFA-DCN* allele combinations with altered risk of ACL rupture in male participants. No associations were noted for any of the polymorphisms investigated in the proteoglycan genes, or within any of the proteoglycan gene-gene allele combinations in the combined analysis. However, several population specific findings were identified, including the significant association of (i) the *ACAN* rs1042631 C/T polymorphism with altered risk in all Polish participants, (ii) *ACAN* inferred haplotypes with altered risk in Swedish participants, and (iii) *BGN* inferred haplotypes with altered risk of ACL rupture in Polish females. Further, trends were noted with the (iv)

ACAN rs1516797 GG genotype and G allele with altered risk in all Swedish participants, and (v) the *BGN* rs1042103 AG genotype with altered risk in Swedish females.

The family of proteoglycans play a critical role in maintaining the structural and functional integrity of soft tissues, and do so in part through interacting with various proteins, including collagens, growth factors and cell-signalling molecules (**Chapter 1, Section 1.9.3**). Decorin, a small leucine rich proteoglycan, plays a role in the angiogenesis associated pathway via its stimulatory/inhibitory effect on VEGFA (Järveläinen et al., 2015). Thus, the interaction effect noted between *DCN* and *VEGFA* on ACL rupture risk modulation within the collective cohort, was an interesting observation. Genetic variation within these genes may alter or disrupt how these proteins interact, leading to downstream effects on angiogenesis related signalling. This finding highlights the value of enriched pathway based analyses, conducted in **Chapter 2**, as an approach to further understand the underlying mechanisms modulating injury risk.

Aggrecan is one of the large aggregating proteoglycans, with a primary role in maintaining the structural integrity of connective tissue, and stabilising the collagen network (**Chapter 1, Section 1.9.3**). In the current study, three polymorphisms mapping to the *ACAN* gene were investigated for an association with non-contact ACL rupture risk. In the combined analysis, no associations with any of the investigated *ACAN* polymorphisms and ACL rupture risk were observed. However, several population specific associations were noted in the independent Swedish and Polish cohorts, which were unique to these single cohorts, and not consistent with previous reported associations. The observed differences between small studies in single populations is not surprising, as it suggests possible associations based on geographical location. Rather, emphasis should be placed on exploring genetic polymorphisms in large combined cohorts such as that presented in this study, to truly ascertain which polymorphisms,

and the genes they map to, are involved in the underlying mechanisms contributing to soft tissue injuries, such as ACL rupture.

Structurally, aggrecan is comprised of three domains; G1, G2 and G3, each with a specific function (**Chapter 1, Section 1.9.3**). The *ACAN* rs1042631 polymorphism, encoded by exon 12, overlaps a large region comprising of chondroitin sulfate (CS) chains between the G2 and G3 domains. This region is primarily responsible for aggrecan's ability to hold large amounts of water in the ECM, and contributes to its critical structural role in ligament (Kiani et al., 2002). The intronic *ACAN* rs1516797 polymorphism investigated overlaps the G3 domain, encoded by exons 13-18 (Valhmu et al., 1995). The main role of this region is to form interactions between aggrecan and the ECM, and to regulate the attachment of GAG chains in the CS region (**Chapter 1, Section 1.9.3**). Thus genetic variation in the intervals spanning these regions, may alter critical binding sites in the structural domains of aggrecan, and thus alter its function in ligament tissue.

Bigylcan is one of small leucine rich proteoglycans that plays a role in the inflammatory pathway, connective tissue development, aging, healing and interacts directly with collagen fibrils to regulate fibrillogenesis (**Chapter 1, Section 1.9.3**). In the combined analysis, no significant independent, or haplotype associations were noted for the *BGN* polymorphisms investigated in male and female participants.

Decorin is the most abundant proteoglycan in the matrix of ligamentous tissue, and serves many functions within connective tissues of which some include an important role in collagen fibrillogenesis, collagen degradation, cellular growth and ECM signalling (**Chapter 1, Section 1.9.3**). Previously, the *DCN* rs516115 polymorphism was associated with ACL rupture risk in

females from South Africa (Mannion et al., 2014). However in the current study, this association was not observed in the combined analysis (n = 1326 females) or in the individual female groups from Sweden and Poland. Notably, the small female sampling group from South Africa (n = 61) may have resulted in a type I error. Interesting to note, was the association of *DCN* rs516115 with sex, body mass and height in Swedish participants. Height is a polygenic trait, with a heritability estimate of ~0.8 (Visscher et al., 2010). Numerous genetic loci have been implicated in explaining variability in height through GWAS technologies, and *ACAN* was shown to associate with height in individuals of European ancestry (Weedon et al., 2008). To date however, no association of *DCN* with body mass or height is evident in the literature. Therefore, this finding may not reflect a true association, but may be due to male participants being on average taller, and of a greater body mass than the female participants.

The small sex-specific sampling groups in the individual cohorts limited the current study. Furthermore, incomplete matching for sports participation and frequency of exposure to contact and non-contact jumping sports was an additional limitation. The observed link between the investigated genes and anthropometric characteristics may indicate potential confounders, which could be masking the true genetic effect on an individual's predisposition to ACL rupture. For this reason, increased sample sizes are required for the individual cohorts, when exploring proteoglycans and the association of the encoding genes with predisposition to ACL rupture.

4.5 CONCLUSION

This study did not identify individual polymorphism associations in the combined cohort, in support of proteoglycan encoding genes in non-contact ACL rupture predisposition. However, it did highlight polymorphisms within *DCN* and *VEGFA* genes in collectively contributing to an individual's predisposition to ACL rupture. This was a novel association, and further highlights the potential interactive value between the members of the proteoglycan gene family, and genes functioning in the angiogenesis network. This interplay can potentially affect blood vessel formation, and possibly influence ligamentization in the healing ACL graft.

The study also further highlights the limitations of genetic association studies in single populations, as noted by the lack of previously identified associations within the combined analysis. Important to note going forward, it may not be that informative to investigate a single polymorphism with risk for multifactorial phenotypes, such as ACL rupture. Instead, larger regions of the genome should be explored, to identify potential interacting polymorphisms predisposing to altered risk. Perhaps it is the interaction of connected pathways, highlighted from the gene-gene interaction results, rather than individual pathways where further research should focus.

CHAPTER 5

SUMMARY AND FUTURE DIRECTION

5.1 SUMMARY

Although research in the field of sporting injuries is growing, the exact aetiology and biological mechanisms underlying these complex phenotypes are currently unknown. A highly multifaceted and poorly understood interaction between extrinsic and intrinsic risk factors is at play, with various determinants modifying the emergent outcome (**Chapter 1, Figure 1.4**) (Bittencourt et al., 2016). Increasingly, genetics is being recognised as a key intrinsic risk factor contributing to musculoskeletal injury predisposition, and emerging evidence suggests a multitude of genomic loci are associated with the risk profile of numerous soft tissue injuries (Rahim et al., 2019).

Considering the previous research implicating genomic loci in injury risk, and the findings from this thesis, one can deduce that a complex interaction between genes and gene products is occurring at the molecular level to influence genetic predisposition to injury. Thus far, over 80 loci residing in genes encoding for components critical to maintaining the structure of soft tissues, ECM activities, and cell signalling have been implicated in musculoskeletal injury risk (Rahim et al., 2019). Some of these genes have been highlighted to play an interactive role in regulating ECM components (**Chapter 1, Figure 1.7**). To adequately unravel how these genetic loci potentially interact to modulate injury risk, a complex systems approach as described by Bittencourt et al. (2016) should be adopted. That is, not only should the potential individual effect of a genetic polymorphism be considered, but rather how these genetic loci are interacting

biologically to modulate an individual's genetic predisposition to injury risk. Implementing this approach requires the recognition that a complex system is a dynamic, open system that is non-linear due to recursive loops, and complex interactions between components, which alone differ from that of their interactive outcome (Plsek and Greenhalgh, 2001).

To date the majority of the implicated polymorphisms in ACL rupture predisposition have been identified through case-control genetic association studies, however recent studies harnessing NGS technologies such GWAS and WES have furthered current knowledge on the regions of the genome potentially influencing one's predisposition to altered risk (**Chapter 1, Section 1.11**). Further to that end, the whole genome sequencing approach described in this thesis, has put forward interesting regions of the genome not previously thought to contribute to the pathobiology of soft tissue injury, such as ion channels and long intergenic non-coding RNAs.

One could therefore propose that a network of genetic risk determinants exists, which may not present as an obvious direct interaction of components, but rather an indirect relationship where gene products (coding and non-coding) and proteins interact in a dynamic, unpredictable manner, and are possibly influenced by other external factors. Rickles et al. (2007) describe how a complex systems approach is characterised by recursive loops, which in a sports injury context can be interpreted as an unpredictable change in the system after an injury has occurred, ultimately altering the initial injury predictors in a manner that no longer presents with the same relationship to the outcome (Rickles et al., 2007). From a genetics perspective, a similar approach in thinking can be adopted. For example, increased levels of VEGFA are observed following a ligament or tendon injury (Beye et al., 2008; Pufe et al., 2001) and post mechanical loading (Petersen et al., 2004). Furthermore, reduced proteoglycan levels were observed in ruptured human ACL tissue in comparison to controls (Young et al., 2011). Yet, it is unclear

how these changes in expression may further influence downstream expression patterns, and the interaction of other genes important in wound healing and remodelling pathways. We are beginning to realise that indeed it may be the biological interaction of a multitude of genes and gene products currently implicated in soft tissue injury predisposition, that are modulating injury risk. Supporting the need to utilise a complex systems approach, to fully understand the potential mechanisms of interaction between these factors.

5.2 NOVEL FINDINGS

This research contributes to the current knowledge base surrounding the genetic risk factors contributing to ACL rupture predisposition.

5.2.1 Six Candidate Genetic Loci in Three Genes have been Highlighted

Whole genome sequencing in a twin-family context identified six loci in three genes: *COL12A1*, *CATSPER2* and *KCNJ12* with predicted loss of function effects, common within the studied families. Alone, the genetic contribution to multifactorial phenotypes such as ACL rupture is not causative, but in combination with other intrinsic and extrinsic risk factors contributes to an individual's overall predisposition to sustaining an ACL rupture.

Of the three genes identified, *CATSPER2* and *KCNJ12* are novel candidate genetic risk loci, not previously associated with ACL rupture predisposition. CATSPER is a sperm-specific ion channel localised to the sperm flagellum, where it facilitates the entry of calcium into sperm, and is critical for sperm motility and hence male fertility (Hildebrand et al., 2010). Organised as a heterotetramer complex, it is comprised of four separate pore-forming subunits;

CATSPER1, CATSPER2 (encoded by *CATSPER2*), CATSPER 3 and CATSPER 4, and three auxiliary subunits (Singh and Rajender, 2015). Disease causing mutations (deletion and CNVs) in *CATSPER2* have been associated with male infertility (Avidan et al., 2003; Zhang et al., 2007). However, to current knowledge, the *CATSPER2* rs144399798 (C/T) polymorphism identified in this work is not reported in the literature. This polymorphism resides in exon 2 (Sherry et al., 2001) of the *CATSPER2* gene, with the resultant protein change resulting in a loss of function in ion channel. However, to existing knowledge, this sperm specific ion channel has not been implicated with ligament or connective tissue biology, and therefore based on current knowledge it is an unlikley candidate to be explored for ACL rupture risk.

Potassium ion channels constitute the largest and most diverse class of voltage-gated ion channels, and play a crucial role in excitable cells by repolarizing the membrane following an action potential (Mobasheri et al., 2012). Potassium channels are classified into three classes: tandem pore domain (K2P), voltage gated (Kv) and inwardly rectifying (Kir) (Hibino et al., 2010). The *KCNJ12* gene encodes for the potassium inwardly rectifying channel subfamily J member 12 (Kir2.2) protein, and variants within the *KCNJ12* gene have been associated with periodic paralyses, a condition of recurrent episodes of flaccid limb muscle weakness (Fan et al., 2018). During knee loading activity, neuromuscular control is essential to maintaining knee joint stability, and insufficient control (strength, co-activation and recruitment) is associated with ACL rupture (Lloyd et al., 2005; Shultz et al., 2012). Kir2.2 protein is expressed in skeletal muscle (Hibino et al., 2010) and therefore, one may postulate that variation in the *KCNJ12* gene may lead to functional alterations of the Kir2.2 ion channel, influencing muscle activation patterns during loading, and subsequently altering an individuals risk for ACL rupture.

5.2.2 A Long Intergenic Non-Coding RNA (LINC01250) Gene Common within Twin Families Affected by ACL Rupture

Identity by descent analysis identified several regions shared between members within and across the two families. Of these, affected twins from both families, and an affected male sibling shared the long intergenic non-protein coding RNA gene (*LINC01250*). LincRNAs are autonomously transcribed non-coding RNAs longer than 200 nucleotides, that do not overlap annotated coding genes. They fall within the class of long non-coding RNA (lncRNA) and share similar features with other transcripts of the family (Ransohoff et al., 2018). The expression of lncRNAs tends to be highly tissue-specific (Cabili et al., 2011) with defined biological roles in several categories, including cellular proliferation, chromatin remodelling, and in the regulation of gene transcription, translation and RNA stabilisation (Kung et al., 2013).

Interestingly, lncRNAs have been implicated in canine cranial cruciate ligament rupture risk (**Chapter 1, Section 1.11**) and in the pathogenesis of periodontitis; an inflammatory disease of the human periodontal ligament (Zhou et al., 2021). *LINC01250* appears to be primarily expressed in the brain and nervous system, but also in skeletal muscle and the reproductive and secretory systems (Stelzer et al., 2016). However, little evidence is available identifying genes that share expression patterns, or biological pathways with *LINC01250* therefore making it difficult to speculate on the potential function of this lincRNA in ACL rupture predisposition, without further investigation.

5.2.3 Previously Implicated Genes in ACL Rupture Predisposition, Highlighted through a WGS Pathway Based Approach

An enriched pathway based approach highlighted several genes (*ACAN, BGN, COL1A1, COL5A1, COL12A1, DCN, FBN2, GDF5, IL6, KDR, TGFB1, TNC,* and *VEFGA*) as key inferred functional partners of prioritised WGS candidates. Moreover, these genes were previously associated with altered risk of ACL rupture. ACL ruptures are multifactorial in nature, with the likely cumulative contribution of interacting biologically relevant pathways and their respective components (Collins et al., 2018). Adopting a pathway-based approach as described in this thesis, in addition to methodologies such as multivariate risk modelling (Rahim et al., 2021) the most informative genetic markers, with potential biological impact in connective tissue can be prioritised. The identified candidates can then be further explored with predisposition to ACL rupture in large collective data sets. Towards this end, the studies presented in **Chapter 3** and **4** of this thesis explored polymorphisms localised to *ACAN*, *BGN*, *DCN*, *VEGFA* and *KDR* with predisposition for ACL rupture in large, combined cohorts including participants from independent cohorts, combined with previously published data sets.

5.2.4 Genetic Loci in Angiogenesis Genes Associated with Altered ACL Rupture Risk in a Large Combined Cohort

Evidence to support the previous implication of genetic loci in the angiogenesis-associated pathway with ACL rupture risk in small single population studies, was provided in a large collective cohort consisting of 912 cases and 765 control participants (**Chapter 3**). Observations at the *VEGFA* locus in the combined cohort point to both independent and inferred haplotype associations. Furthermore, the findings indicate a refinement of the *VEGFA* genetic interval that most likely associates with injury risk, and the probable lack of influence of the *KDR* loci in modulating an individual's predisposition for ACL rupture risk.

Importantly, the findings from **Chapter 3** and **4** suggest genetic association studies are more informative in large, collective samples. With smaller single populations providing insight into unique population specific associations. To that end, recent work employing large datasets has identified genetic risk loci underlying an individual's predisposition to ACL rupture (Gibbon et al., 2020) and rotator cuff tear (Tashjian et al., 2021).

5.2.5 Gene-Gene Interaction Analyses Identified VEGFA-DCN Inferred Allele Combinations to Modulate ACL Rupture Risk in a Large, Collective Cohort

Components of the ECM interact to bring about matrix remodelling and connective tissue homeostasis, which is critical for ligament healing (described in **Section 1.10**, **Chapter 1**). The fine-tuning of *VEGFA* expression, modulated in part through DCN activity, is essential for effective wound healing, and ligamentization post ACL reconstruction. The novel implication of the *VEGFA* (rs699947 C/A, rs1570360 G/A, rs20109630 G/C) and *DCN* (rs516115 T/C) inferred allele combinations A-G-G-T, A-A-G-C, and A-A-G-T with altered ACL rupture risk in male participants (combined cohort) is therefore noteworthy, as it suggests a potential collective effort of ECM remodelling components in contributing to ACL rupture predisposition.

5.2.6 Proteoglycan Genes Not Found to Modulate ACL Rupture Risk in a Large, Collective Cohort

Interestingly, none of the previously implicated proteoglycan polymorphisms were associated with altered risk of ACL rupture in a large, collective cohort. This provides important evidence to support the investigation of loci implicated in small single populations, in large datasets with increased statistical power. To ensure noted associations are reflective of a true biological consequence.

5.3 FUTURE DIRECTION

Genetic association studies have provided much information on the intrinsic risk factors contributing to musculoskeletal injury predisposition, however these associations cannot be considered independently but rather as a phased approach, where the functional hypothesis is also explored.

So far, a proposed working hypothesis for polymorphisms located in *COL1A1* and the 3'UTR of *COL5A1* collagen genes is present (**Section 1.9** of **Chapter 1**). Where in brief, variation in the genetic sequences encoding these proteins results in a change in collagen structure, ultimately altering tissue biomechanics and increasing injury risk. In other matrix remodelling genes, associations with the *VEGFA* gene explored in this work are in agreement with the reported biological function of the alleles (**Chapter 3**). Furthermore, recent studies have provided functional evidence, linking a genetic contribution to the expression of ECM components, in musculoskeletal injury predisposition (Suijkerbuijk et al., 2019; Willard et al., 2020). Where polymorphisms within genes of the inflammatory pathway (*ILB* and *IL6*) were found to influence *COL5A1* and *BGN* gene expression in a genetic risk dependent manner, and in doing so, modulate predisposition to ACL rupture (Suijkerbuijk et al., 2019). Further to that, increased *BGN*, *COL5A1*, and *VEGFA* gene expression was associated with a combined *BGN* and *COL5A1* reduced genetic risk profile, in musculoskeletal injury predisposition (Willard et al., 2020).

However, despite this growing knowledge, there is still limited functional evidence for the plethora of implicated genomic regions and polymorphisms with musculoskeletal injury risk. Future work should therefore build on this premise, to further elucidate the biological

implications of associated loci in injury predisposition, and how these genetic regions interact in the pathobiology of soft tissue injuries.

Additionally, future work should consider epigenetic factors such as DNA methylation, as a potential mechanism contributing to musculoskeletal injury risk. To date, few studies have explored the impact of epigenetic factors in the risk profile of musculoskeletal injury phenotypes. In a sports injury model, DNA methylation patterns of ECM remodelling genes; *ADAMTS4* (El Khoury et al., 2018) and *MMP11* (Rickaby et al., 2019) were found to differ between injured and non-injured patients with patella tendinopathy. Furthermore, demethylation of the *ADAMTS4* (Cheung et al., 2009) and *MMP13* (Bui et al., 2012) promoter regions have been observed in osteoarthritic cartilage. The epigenome is modified by environmental factors, and therefore gaining a better understanding of these factors presents an avenue to potentially alter musculoskeletal injury risk (El Khoury et al., 2018).

5.4 STRENGTHS AND LIMITATIONS

5.4.1 Strengths of the Research

A strength of this work was the inclusion of a family based approach with the addition of monozygotic twins. Family-based association studies allow for improved control of confounding bias that occurs due to population stratification, and therefore improves the identification of true risk loci (Thomas and Witte, 2002). Additionally, the study of twins concordant for a particular phenotype of interest provides a valuable tool to explore genetic factors contributing to risk. Currently there is limited research investigating the risk factors

associated with ACL rupture in twins (Pelkowski et al., 2020) and therefore this body of work contributes to the current knowledge gap in this field.

Another strength of the thesis was the use of a whole genome sequencing approach to determine the underlying genetic risk factors contributing to ACL rupture. Utilising a WGS design and a directed bioinformatics pipeline, novel genetic intervals and risk loci were identified that otherwise would have been overlooked if an alternative approach (GWAS or WES) was used. The candidate genetic risk loci (*KCNJ12* and *LINC01250*) identified, are potentially new areas of interest in connective tissue injury phenotypes, and therefore provides important direction for future work. Further to that, using a pathway based approach, genes previously associated with altered risk of ACL rupture were noted as functional partners of some of the prioritised WGS genes. This is a strength of the thesis, as it provides an important step in knowledge of the interactions between regions of the genome implicated in ACL rupture risk.

An additional strength of this thesis was the use of a collective cohort to explore previously associated angiogenesis and proteoglycan encoding genes. To the author's knowledge, the studies described in **Chapter 3** (angiogenesis genes) and **Chapter 4** (proteoglycan genes) are one of the larger studies investigating these loci in ACL rupture predisposition using a case-control genetic association approach. Pooling the genotype data from independent cohorts to form a large combined analysis provided an improved estimate of the overall risk (Lambrechts et al., 2003) contributing to ACL rupture. This work furthers the current understanding of the key contributing polymorphisms within previously implicated regions of interest, involved in ECM remodelling pathways. Therefore providing additional evidence to support the need for more functional work exploring these regions in their biological context.

An additional strength of this work was only including case participants with an ACL rupture sustained through non-contact mechanisms, in the independent cohorts from Sweden, Poland and Australia, in addition to the previously published work contributing to the collective cohort. This aided in removing potential cases of rupture through mechanisms not strongly influenced by genetic load, and specifically supports the hypothesis for a biological link between the implicated genes and predisposition to ACL rupture risk.

5.4.2 Limitations

Although providing novel findings, and supporting the current research describing regions of the genome likely influencing an individuals predisposition to ACL rupture, the studies conducted within this thesis were not without limitations. The lack of functional data for the implicated variants in the WGS study prevents any immediate conclusions on the functional significance of these loci. The potential effects on protein structure resulting from the loss of function sequence alterations highlighted in the WGS study (**Chapter 2, Table 2.4**) needs to be explored at an *in silico* level. In addition, modelling analyses of the impact of the predicted loss of function sequence changes on altering functional partners could also be explored. Bioinformatic modelling can inform potential theories, on how these DNA sequence changes could potentially influence the underlying mechanisms, leading to ligament rupture.

For the case-control genetic association studies, the inclusion of four independent populations was primarily a strength of the research. These single populations were however limited by small sample size, and therefore statistical power. And so moving forward, it would be more informative to increase the single population specific cohort sizes to ensure an observed association is reflective of a biological consequence. Furthermore, within each cohort, emphasis surrounding balance between cases and controls is needed to ensure an understanding of the frequencies of the polymorphisms in each group. The insufficient number of controls in the Australian cohort was a limitiation here. However, this limitation was mitigated in the combined analysis where the numbers were more balanced between cases and controls.

An additional limitation was that participants recruited for the Swedish, Polish and Australian cohorts, were selected on both a hospital and general population basis. Furthermore, all study controls self-reported as having no history of knee ligament or tendon injuries, this however was not confirmed using diagnostic tools, and therefore it cannot be ruled out that these individuals did not have pre-existing or previous injuries. Further to that, each research centre (of the individual population cohorts) used their own questionnaire but there was similarity in the type of questions used. These questions included personal details on descriptive characteristics, ancestry, lifestyle habits, occupational details, sporting history (number of years, playing level, frequency), details of ACL injury, history of other ligament or tendon injury, and medical history. Each research centre was mindful in their attempts to group their cases and controls for the potential confounders such as population ancestry, sporting level (if used in the questionnaire), sports played, number of years of participation and frequency of playing hours (if used in the questionnaire). However, some of the questions related to sports activity profiling, differed in details between research centres and this made it difficult to compare all the specific variables related to sporting profile across the combined cohort. However, these data points were evaluated between cases and controls of the individual research centres to identify potential confounders and to see how well cases and controls were grouped.

For female participants the details concerning the menstrual cycle at the time of ACL rupture was not included. As female sex hormone concentrations over the menstrual cycle may influence ligament tissue characteristics, and hence loading response (**Chapter 1, Section 1.7**). Fluctuating hormone levels throughout the female cycle may pose varying risk to ACL rupture for female athletes, independent of the genetic contribution. It is acknowledged that confounders between the individual cohorts differ and should be considered when pooling

cohorts from independent research centres. Going forward, sports participation data in particular should be more uniformly collected, to facilitate the identification of potential confounders and to allow for there adjustment in the analyses.

To accommodate for unmatching of participant descriptive characteristics within the individual independent cohorts, statistical analyses were adjusted for the respective confounders, such that for Sweden age and type of sport was a considered confounder, for Poland age, body mass and type of sport, for Australia age and sex, and for the combined cohort, age, BMI and country were included in the adjustment model.

5.5 CONCLUDING REMARKS

In final, this body of work provides evidence to support a WGS approach to explore the underlying mechanisms contributing to non-contact ACL rupture. Moreover, the research supports a twin family study as a feasible means to explore genetic sequences and intervals common within individuals presenting with the ACL phenotype. To current knowledge, this research is one of the first to employ whole genome sequencing in the investigation of potential regions of the human genome modulating ACL rupture predisposition. Utilising WGS technologies this research achieved the aim set out to identify novel or previously implicated genetic loci contributing to ACL rupture, by providing preliminary evidence implicating ion channels and long non-coding RNAs as potential contributors underlying ACL rupture predisposition. Additionally, previously implicated genes were identified as key functional partners of prioritised WGS candidates. Furthermore, the second aim to explore the identified functional partners previously associated with altered risk of ACL rupture, independently and as a collective in a large combined dataset was achieved using a case-control genetic association

design, including the collective data from three independent sampling groups, and two previously published datasets. The key findings support the implication of *VEGFA* in altering an individual's predisposition to ACL rupture, and furthermore suggest potential interactions between genes in the angiogenesis-associated pathway and proteoglycan encoding genes in collectively contributing to risk of ACL rupture.

The next step is to explore the potential effect of the mutations identified from the WGS work on protein structure, using protein dynamics simulation. Similarly, there is a need to explore the potential *in silico* effects of genetic variation between some of the sequences identified in the WGS data, and the polymorphisms explored in **Chapters 3** and **4**. Further to that, the mutations identified in the cohorts need to be further explored for prevalence as a risk factor in the general population for the ACL phenotype. Lastly, there is a need for further functional work, to fully understand the biological role of the implicated polymorphisms underlying ligament rupture.

There is much complexity around the knee, and in addition to ligament tissue, the musculature around the knee joint is equally important in stabilising knee movement. It is therefore imperitive that we start developing improved *in silico* modeling to test potential genetic hypotheses and effects on possible hip and knee movements towards injury elucidation using a complex systems approach in thinking (Bittencourt et al., 2016). Through collaboration, we can begin to connect the implicated genetic regions with a potential biological role, and improve our understanding of the varying weighting of their influence, and how their degree of interaction and regulation by external factors ultimately modulates musculoskeletal injury risk.

REFERENCE LIST

Abrahams, Y., M-J. Laguette, S. Prince, and M. Collins. 2013. Polymorphisms within the COL5A1 3 ′ -UTR that alters mRNA structure and the *MIR608* gene are associated with Achilles tendinopathy. *Ann Hum Genet*. 77:204-214.

Adzhubei, I.A., S. Schmidt, L. Peshkin, V.E. Ramensky, A. Gerasimova, P. Bork, A.S. Kondrashov, and S.R. Sunyaev. 2010. A method and server for predicting damaging missense mutations. *Nature methods*. 7:248-249.

Ageberg, E. 2002. Consequences of a ligament injury on neuromuscular function and relevance to rehabilitation - Using the anterior cruciate ligament-injured knee as model. *J Electromyogr Kinesiol*. 12:205-212.

Alentorn-Geli, E., J. Mendiguchia, K. Samuelsson, V. Musahl, J. Karlsson, R. Cugat, and G.D. Myer. 2014. Prevention of anterior cruciate ligament injuries in sports. Part I: systematic review of risk factors in male athletes. *Knee Surg Sports Traumatol Arthrosc*. 22:3-15.

Altinisik, J., G. Meric, M. Erduran, O. Ates, A.E. Ulusal, and D. Akseki. 2015. The BstUI and DpnII Variants of the COL5A1 Gene Are Associated With Tennis Elbow. *Am J Sports Med*. 43:1784-1789.

Ameye, L., and M.F. Young. 2002. Mice deficient in small leucine-rich proteoglycans: novel in vivo models for osteoporosis, osteoarthritis, Ehlers-Danlos syndrome, muscular dystrophy, and corneal diseases. *Glycobiology*. 12:107R-116R.

Asling, B., J. Jirholt, P. Hammond, M. Knutsson, A. Walentinsson, G. Davidson, L. Agreus, A. Lehmann, and M. Lagerström-Fermer. 2009. Collagen type III alpha I is a gastro-oesophageal reflux disease susceptibility gene and a male risk factor for hiatus hernia. *Gut*. 58:1063-1069.

Aspberg, A., R. Miura, S. Bourdoulous, M. Shimonaka, D. Heinegård, M. Schachner, E. Ruoslahti, and Y. Yamaguchi. 1997. The c-type lectin domains of lecticans, a family of aggregating chondroitin sulfate proteoglycans, bind tenascin-R by protein–protein interactions independent of carbohydrate moiety. *Proc Natl Acad Sci*. 94:10116-10121.

Avidan, N., H. Tamary, O. Dgany, D. Cattan, A. Pariente, M. Thulliez, N. Borot, L. Moati, A. Barthelme, L. Shalmon, T. Krasnov, E. Ben-Asher, T. Olender, M. Khen, I. Yaniv, R. Zaizov, H. Shalev, J. Delaunay, M. Fellous, D. Lancet, and J.S. Beckmann. 2003. *CATSPER2*, a human autosomal nonsyndromic male infertility gene. *Eur J Hum Genet*. 11:497-502.

Awata, T., K. Inoue, S. Kurihara, T. ohkubo, M. Watanabe, K. Inukai, I. Inoue, and S. Katayama. 2002. A common polymorphism in the 5'-untranslated region of the *VEGF* gene is associated with diabetic retinopathy in type 2 diabetes. *Diabetes*. 51:1635-1639.

Bader, H.L., D.R. Keene, B. Charvet, G. Veit, W. Driever, M. Koch, and F. Ruggiero. 2009. Zebrafish collagen XII is present in embryonic connective tissue sheaths (fascia) and basement membranes. *Matrix Biol*. 28:32-43.

Baird, A.E.G., S.D. Carter, J.F. Innes, W. Ollier, and A. Short. 2014. Genome-wide association study identifies genomic regions of association for cruciate ligament rupture in Newfoundland dogs. *Anim Genet*. 45:542-5449.

Baker, L., B. Kirkpatrick, G. Rosa, D. Gianola, B. Valente, J. Sumner, W. Baltzer, Z. Hao, E. Binversie, N. Volstad, A. Piazza, S. Sample, and P. Muir. 2017. Genome-wide association analysis in dogs implicates 99 loci as risk variants for anterior cruciate ligament rupture. *PLoS Medicine*. 12:1-19.

Baker, L.A., G.J.M. Rosa, Z. Hao, A. Piazza, C. Hoffman, E.E. Binversie, S.J. Sample, and P. Muir. 2018. Multivariate genome-wide association analysis identifies novel and relevant variants associated with anterior cruciate ligament rupture risk in the dog model. *BMC Genet*. 19:39.

Balachandar, V., J.L. Marciniak, O. Wall, and C. Balachandar. 2017. Effects of the menstrual cycle on lower-limb biomechanics, neuromuscular control, and anterior cruciate ligament injury risk: a systematic review. *Muscles Ligaments Tendons J*. 7:136-146.

Barr, A.E., and M.F. Barbe. 2002. Pathophysiological tissue changes associated with repetitive movement: a review of the evidence. *Phys Ther*. 82:173-187.

Bell, J.I. 2002. Single nucleotide polymorphisms and disease gene mapping. *Arthritis Res*. 4:S273-S278.

Benjamini, Y., and Y. Hochberg. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Statist. Soc. B*. 57:289-300.

Bere, T., T.W. Flørenes, T. Krosshaug, H. Koga, L. Nordsletten, C. Irving, E. Muller, R.C. Reid, V. Senner, and R. Bahr. 2011. Mechanisms of Anterior Cruciate Ligament Injury in World Cup Alpine Skiing: A Systematic Video Analysis of 20 Cases. *Am J Sports Med*. 39:1421-1429.

Berns, G.S., M.L. Hull, and H.A. Patterson. 1992. Strain in the anteromedial bundle of the anterior cruciate ligament under combination loading. *J Orthop Res*. 10:167-176.

Beye, J.A., D.A. Hart, R.C. Bray, J.J. McDougall, and P.T. Salo. 2008. Injury-induced changes in mRNA levels differ widely between anterior cruciate ligament and medial collateral ligament. *Am J Sports Med*. 36:1337-1346.

Beynnon, B.D., P.M. Vacek, M.K. Newell, T.W. Tourville, H.C. Smith, S.J. Shultz, J.R. Slauterbeck, and R.J. Johnson. 2014. The Effects of Level of Competition, Sport, and Sex on the Incidence of First-Time Noncontact Anterior Cruciate Ligament Injury. *Am J Sports Med*. 42:1806-1812.

Bhide, V.M., C.A. Laschinger, P.D. Arora, W. Lee, L. Hakkinen, H. Larjava, J. Sodek, and C.A. McCulloch. 2005. Collagen phagocytosis by fibroblasts is regulated by decorin. *J Biol Chem*. 280:23103-23113.

Biasiotta, A., D. D'Arcangelo, F. Passarelli, E.M. Nicodemi, and A. Facchiano. 2016. Ion channels expression and function are strongly modified in solid tumors and vascular malformations. *J Transl Med*. 14:1-15.

Birk, D.E., J.M. Fitch, J.P. Babiarz, K.J. Doane, and T.F. Linsenmayer. 1990. Collagen fibrillogenesis in vitro: interaction of types I and V collagen regulates fibril diameter. *J Cell Sci*. 95:649-657.

Birkedal-Hansen, H., W.G.I. Moore, and M.K. Bodden. 1993. Matrix Metalloproteinases: A Review. *Crit Rev Oral Biol Med*. 4:197-250.

Bittencourt, N., W.H. Meeuwisse, L.D. Mendonça, A. Nettel-Aguirre, J.M. Ocarino, and S.T. Fonseca. 2016. Complex systems approach for sports injuries: moving from risk factor identification to injury pattern recognition-narrative review and new concept. *Br J Sport Med*. 50:1309-1314.

Blanke, F., A.M. Kiapour, M. Haenle, J. Fischer, M. Majewski, S. Vogt, and C. Camathia. 2016. Risk of noncontact anterior cruciate ligament injuries is not associated with slope and concavity of the tibial plateau in recreational alpine skiers: A Magnetic resonance imaging-based case-control study of 121 patients. *Am J Sports Med*. 44:1508-1514.

Bonod-Bidaud, C., and F. Ruggiero. 2013. Inherited connective tissue disorders of collagens: Lessons from targeted mutagenesis. *In* Genetic manipulation of DNA and protein. D. Figurski, editor. IntechOpen, London.

Bouaoun, L., D, Sonkin, M. Ardin, M. Hollstein, G. Byrnes, J. Zavadil, and M. Olivier. 2016. TP53 Variations in Human Cancers: New Lessons from the IARC TP53 Database and Genomics Data. *Hum Mutat*. 37:865-76.

Bray, R.C., C.A. Leonard, and P.T. Salo. 2002. Vascular physiology and long-term healing of partial ligament tears. *J Orthop Res*. 20:984-989.

Bredrup, C., P.M. Knappskog, J. Majewski, and E.B. Rødahl, H. . 2005. Congenital stromal dystrophy of the cornea caused by a mutation in the decorin gene. *Invest Ophthalmol Vis Sci*. 46:420-426.

Brophy, R.H., R.W. Wright, and M.J. Matava. 2009. Cost Analysis of Converting from Single-Bundle to Double-Bundle Anterior Cruciate Ligament Reconstruction. *Am J Sports Med*. 37:683-687.

Brown, K.L., K.B. Seale, L.Y. El Khoury, M. Posthumus, W.J. Ribbans, S.M. Raleigh, M. Collins, and A.V. September. 2017. 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 Sports Sci*. 35:1475-1483.

Browning, B.L., and S.R. Browning. 2013. Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics*. 194:459-471.

Bui, C., M.J. Barter, J.L. Scott, Y. Xu, M. Galler, L.N. Reynard, A.D. Rowan, and D.A. Young. 2012. cAMP response element-binding (CREB) recruitment following a specific CpG demethylation leads to the elevated expression of the matrix metalloproteinase 13 in human articular chondrocytes and osteoarthritis. *FASEB J*. 26:3000-3011.

Burger, M.C., H. de Wet, and M. Collins. 2014. The *BGN* and *ACAN* genes and carpal tunnel syndrome. *Gene*. 551:160-166.

Burger, M.C., H. De Wet, and M. Collins. 2015. Interleukin and growth factor gene variants and risk of carpal tunnel syndrome. *Gene*. 564:67-72.

Butler, D.L., E.S. Grood, F.R. Noyes, and R.F. Zernicke. 1978. Biomechanics of ligaments and tendons. Exercise and sport sciences reviews. *Exerc Sport Sci Rev*. 6:125-181.

Cabili, M.N., C. Trapnell, L. Goff, M. Koziol, B. Tazon-Vega, A. Regev, and J.L. Rinn. 2011. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes* Dev. 25:1915-27.

Campbell, M.A., A.M. Tester, C.J. Handley, G.J. Checkley, G.L. Chow, A.E. Cant, A.D. Winter, and W.E. Cain. 1996. Characterization of a Large Chondroitin Sulfate Proteoglycan Present in Bovine Collateral Ligament. *Arch Biochem Biophys*. 329:181-190.

Caso, E., A. Maestro, C.C. Sabiers, M. Godino, Z. Caracuel, J. Pons, F.J. Gonzalez, R. Bautista, M.G. Claros, J. Caso-Onzain, E. Viejo-Allende, P.V. Giannoudis, S. Alvarez, P. Maietta, and E. Guerado. 2016. Whole-exome sequencing analysis in twin sibling males with an anterior cruciate ligament rupture. *Injury*. 47:S41-S50.

Chan, K.M., S.C. Fu, Y.P. Wong, W.C. Hui, Y.C. Cheuk, and M.W. Wong. Expression of transforming growth factor beta isoforms and their roles in tendon healing. 2008. *Wound Repair Regen*. 16:399-407.

Charbonneau, N.L., R.N. Ono, G.M. Corson, D.R. Keene, and L.Y. Sakai. 2004. Fine tuning of growth factor signals depends on fibrillin microfibril networks. *Birth Defects Res C Embryo Today*. 72:37-50.

Chaudhari, A.M.W., E.A. Zelman, D.C. Flanigan, C.C. Kaeding, and H.N. Nagaraja. 2009. Anterior cruciate ligament-injured subjects have smaller anterior cruciate ligaments than matched controls: a magnetic resonance imaging study. *Am J Sports Med*. 37:1282-1287.

Chen, B., B. Li, Y.J. Qi, K. Tie, and L.B. Chen. 2015. Association study between growth differentiation factor 5 polymorphism and non-contact anterior cruciate ligament rupture in Chinese Han population. *Int J Clin Exp Med*. 8:22484-22490.

Chen, J., Y. Zhu, Y. Jiang, H. Yu, K. Sun, W. Song, L. Luan, K. Lou, Y. Li, P. Jiang, Q. Pang, and R. Hui. 2010a. A functional variant of the collagen type III alpha1 gene modify risk of sporadic intracranial aneurysms. *Hum Genet*. 131:1137-1143.

Chen, R., E.V. Davydov, M. Sirota, and A.J. Butte. 2010b. Non-synonymous and synonymous coding SNPs show similar likelihood and effect size of human disease association. *PloS One* 5:e13574.

Chen, S., and D.E. Birk. 2013. The regulatory roles of small leucine-rich proteoglycans in extracellular matrix assembly. *FEBS J*. 280:20-37.

Chen, Y., N.B. Nixon, P.T. Dawes, and D.L. Mattey. 2012. Influence of variations across the MMP-1 and -3 genes on the serum levels of MMP-1 and -3 and disease activity in rheumatoid arthritis. *Genes Immun*. 13:29-37.

Cheung, K.S., K. Hashimoto, N. Yamada, and H.I. Roach. 2009. Expression of ADAMTS-4 by chondrocytes in the surface zone of human osteoarthritic cartilage is regulated by epigenetic DNA de-methylation. *Rheumatol Int*. 29:525-534.

Chiu, R., W. Li, R.P. Herber, S.J. Marshall, M. Young, and S.P. Ho. 2012. Effects of biglycan on physico-chemical properties of ligament-mineralized tissue attachment sites. *Arch Oral Biol*. 57:177-187.

Cho, S.Y., J.S. Bae, N.K.D. Kim, F. Forzano, K.M. Girisha, C. Baldo, F. Faravelli, T.J. Cho, D. Kim, K.Y. Lee, S. Ikegawa, J.S. Shim, A.R. Ko, N. Miyake, G. Nishimura, A. Superti-Furga, J. Spranger, O.H. Kim, W.Y. Park, and D. Jin. 2016. *BGN* Mutations in X-Linked Spondyloepimetaphyseal Dysplasia. *Am J Hum Genet*. 98:1243-1248.

Cięszczyk, P., K. Willard, P. Gronek, P. Zmijewski, G. Trybek, J. Gronek, M. Weber-Rajek , P. Stastny, M. Petr, E. Lulińska-Kuklik, K. Ficek, E. Kemeryte-Riaubiene, E. Maculewicz, and A.V. September. 2017. Are genes encoding proteoglycans really associated with the risk of anterior cruciate ligament rupture? *Biol Sport*. 31:97-103.

Coffey, D.S. 1998. Self-organization, complexity and chaos: the new biology for medicine. *Nat Med*. 4:882-885.

Cohen, T., D. Nahari, L.W. Cerem, G. Neufeld, and B.Z. Levi. 1996. Interleukin 6 induces the expression of vascular endothelial growth factor. *J Biol Chem*. 271:736–741.

Collins, M., and M. Posthumus. 2011. Type V collagen genotype and exercise-related phenotype relationships. *Exerc Sport Sci Rev*. 39:191-198.

Collins, M., and S.M. Raleigh. 2009. Genetic risk factors for musculoskeletal soft tissue injuries. *Med Sport Sci*. 54:136-149.

Collins, M., A.V. September, and M. Posthumus. 2018. Genetic influences on anterior cruciate ligament injury. *In* The anterior cruciate ligament. C. Prodromos, editor. Elsevier PA, Philadelphia. 8–12.

Cornish, A., and C. Guda. 2015. A Comparison of Variant Calling Pipelines Using Genome in a Bottle as a Reference. *BioMed research international*:456479.

Corsi, A., T. Xu, X.D. Chen, A. Boyde, J. Liang, M. Mankani, B. Sommer, R.V. Iozzo, I. Eichstetter, P.G. Robey, P. Bianco, and M.F. Young. 2002. Phenotypic effects of biglycan deficiency are linked to collagen fibril abnormalities, are synergized by decorin deficiency, and mimic Ehlers-Danlos-like changes in bone and other connective tissues. *J Bone Miner Res*. 17:1180-1189.

Cox, T.R., and J.T. Erler. 2011. Remodeling and homeostasis of the extracellular matrix: implications for fibrotic diseases and cancer. *Dis Model Mech*. 4:165-178.

Croft, D., A.F. Mundo, R. Haw, M. Milacic, J. Weiser, G. Wu, M. Caudy, P. Garapati, M. Gillespie, M.R. Kamdar, and B. Jassal. 2013. The Reactome pathway knowledgebase. *Nucleic acids research*. 42: D472-D477.

D'Alessio, M., F. Ramirez, B.D. Blumberg, M.K. Wirtz, V.H. Rao, M.D. Godfrey, and D.W. Hollister. 1991. Characterization of a *COL1A1* splicing defect in a case of Ehlers-Danlos syndrome type VII: further evidence of molecular homogeneity. *Am J Hum Genet*. 49:400-406.

Dallalana, R.J., J.H.M. Brooks, S.P.T. Kemp, and A.M. Williams. 2007. The epidemiology of knee injuries in English professional rugby union. *Am J Sports Med*. 35:818-830.

Danecek, P., and S.A. McCarthy. 2017. BCFtools/csq: haplotype-aware variant consequences. *Bioinformatics*. 33(13):2037-2039.

Danielson, K.G., H. Baribault, D.F. Holmes, H. Graham, K.E. Kadler, and R.V. Iozzo. 1997. Targeted disruption of decorin leads to abnormal collagen fibril morphology and skin fragility. *J Cell Biol*. 136:729-743.

Delbaere, S., T. Dhooge, D. Syx, F. Petit, N. Goemans, A. Destrée, O. Vanakker, R. De Rycke, S. Symoens, and F. Malfait. 2020. Novel defects in collagen XII and VI expand the mixed myopathy/Ehlers-Danlos syndrome spectrum and lead to variant-specific alterations in the extracellular matrix. *Genet Med*. 22:112-123.

DePristo, M.A., E. Banks, Poplin, R., K.V. Garimella, J.R. Maguire, C. Hartl, Philippakis, A. A., G. del Angel, M.A. Rivas, M. Hanna, A. McKenna, T.J. Fennell, A.M. Kernytsky, A.Y. Sivachenko, K. Cibulskis, S.B. Gabriel, D. Altshuler, and M.J. Daly. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 43:491–498.

Dong, C., P. Wei, X. Jian, R. Gibbs, E. Boerwinkle, K. Wang, and X. Liu. 2015. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum Mol Genet*. 24:2125-2137.

Dong, D.M., M. Yao, B. Liu, C.Y. Sun, Y.Q. Jiang, and Y.S. Wang. 2007. Association between the -1306C/T polymorphism of matrix metalloproteinase-2 gene and lumbar disc disease in Chinese young adults. *Eur Spine J*. 16:1958-1961.

Du, X., and N. Gamper. 2013. Potassium channels in peripheral pain pathways: expression, function and therapeutic potential. *Curr Neuropharmacol*. 11:621-640.

Dunkman, A.A., M.R. Buckley, M.J. Mienaltowski, S.M. Adams, S.J. Thomas, L. Satchell, A. Kumar, L. Pathmanathan, D.P. Beason, R.V. Iozzo, D.E. Birk, and L.J. Soslowsky. 2014. The tendon injury response is influenced by decorin and biglycan. *Ann of Biomed Eng*. 42:619–630.

Duthon, V.B., C. Barea, S. Abrassart, J.H. Fasel, D. Fritschy, and J. Ménétrey. 2006. Anatomy of the anterior cruciate ligament. *Knee Surg Sports Traumatol Arthrosc*. 14:204-213.

Egli, R.J., L. Southam, J.M. Wilkins, I. Lorenzen, M. Pombo-Suarez, A. Gonzalez, A. Carr, K. Chapman, and J. Loughlin. 2009. Functional analysis of the osteoarthritis susceptibility-associated GDF5 regulatory polymorphism. *Arthritis Rheum*. 60: 2055-2064.

El Khoury, L., M. Posthumus, M. Collins, W. van der Merwe, C.J. Handley, and J. Cook. 2014. *ELN* and *FBN2* gene variants as risk factors for two sports-related musculoskeletal injuries. *Int J Sports Med*. 36:333-337.

El Khoury, L., R. Rickaby, T. Samiric, and S.M. Raleigh. 2018. Promoter methylation status of the TIMP2 and ADAMTS4 genes and patellar tendinopathy. *J Sci Med Sport*. 21:378-382.

El Khoury, L., W.J. Ribbans, and S.M. Raleigh. 2016. *MMP3* and *TIMP2* gene variants as predisposing factors for Achilles tendon pathologies: Attempted replication study in a British case-control cohort. *Meta Gene*. 9:52-55.

Ensembl. Gene *BGN*: ENSG00000182492. Accessed April 2021.

Everhart, J.S., J.H. Sojka, C.C. Kaeding, A.L. Bertone, and D.C. Flanigan. 2017. The ACL injury response: A collagen-based analysis. *Knee*. 24:601-607.

Fan, C., M. Kuhn, A.P. Mbiol, J. Groome, V. Winston, S. Biskup, F. Lehmann-Horn, and K. Jurkat-Rott 2018. Kir2.2 p.Thr140Met: a genetic susceptibility to sporadic periodic paralysis. *Acta Myol*. 37:193-203.

Fang, M., R. Jacob, O. McDougal, and J.T. Oxford. 2012. Minor fibrillar collagens, variable regions alternative splicing, intrinsic disorder, and tyrosine sulfation. *Protein Cell*. 3:419-433.

Feng, X.H., and R. Derynck. 2005. Specificity and versatility in tgf-beta signaling through Smads. *Annu Rev Cell Dev Biol*. 21:659-693.

Fenwick, S.A., B.L. Hazleman, and G.P. Riley. 2002. The vasculature and its role in the damaged and healing tendon. *Arthritis Res*. 4:252-260.

Ferrara, N., and A.P. Adamis. 2016. Ten years of anti-vascular endothelial growth factor therapy. *Nat Rev Drug Discov*. 15:385-403.

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

Ficek, K., P. Cieszczyk, M. Kaczmarczyk, A. Maciejewska-Karłowska, M. Sawczuk, and J. Cholewinski. 2013. Gene variants within the *COL1A1* gene are associated with reduced anterior cruciate ligament injury in professional soccer players. *J Sci Med Sport*. 16:396-400.

Ficek, K., M. Stepien-Slodkowska, M. Kaczmarczyk, A. Maciejewska-Karlowska, M. Sawczuk, J. Cholewinski, A. Leonska-Duniec, A. Zarebska, P. Cieszczyk, and P. Zmijewski. 2014. Does the A9285G Polymorphism in Collagen Type XII α1 Gene Associate with the Risk of Anterior Cruciate Ligament Ruptures? *Balkan J Med Genet*. 17:41-46.

Fichard, A., J.P. Kleman, and F. Ruggiero. 1995. Another look at collagen V and XI molecules. *Matrix Biol*. 14:515-531.

Fiorio Pla, A., and L. Munaron. 2014. Functional properties of ion channels and transporters in tumour vascularization. *Philos Trans R Soc Lond B Biol Sci*. 369:20130103.

Fishman, D., G. Faulds, R. Jeffery, V. Mohamed-Ali, J.S. Yudkin, S. Humphries, and P. Woo. 1998. The effect of novel polymorphisms in the interleukin-6 (*IL-6*) gene on IL-6 transcription and plasma IL-6 levels, and an association with systemic-onset juvenile chronic arthritis. *J Clin Invest*. 102:1369-1376.

Fleming, B.C., P.A. Renstrom, B.D. Beynnon, B. Engstrom, G.D. Peura, G.J. Badger, and R.J. Johnson. 2001. The effect of weightbearing and external loading on anterior cruciate ligament strain. *J Biomech*. 34:163-170.

Flynn, R.K., C.L. Pedersen, T.B. Birmingham, A. Kirkley, D. Jackowski, and P.J. Fowler. 2005. The familial predisposition toward tearing the anterior cruciate ligament: a case control study. *Am J Sports Med*. 33:23-28.

Font, B., D. Eichenberger, L.M. Rosenberg, and M. van der Rest. 1996. Characterization of the interactions of type XII collagen with two small proteoglycans from fetal bovine tendon, decorin and fibromodulin. *Matrix Biol*. 15:341-348.

Forbes, S.A., D. Beare, P. Gunasekaran, K. Leung, N. Bindal, H. Boutselakis, M. Ding, S. Bamford, C. Cole, S. Ward, C.Y. Kok, M. Jia, T. De, J.W. Teague, M.R. Stratton, U. McDermott, and P.J. Campbell. 2015. COSMIC: Exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res*. 43:D805-D811.

Frank, C., N. Shrive, H. Hiraoka, N. Nakamura, Y. Kaneda, and D. Hart. 1999. Optimization of the biology of soft tissue repair. *J Sci Med Sport*. 2:190-210.

Frank, C.B. 2004. Ligament structure, physiology and function. *J of Musculoskelet Neuronal Interact*. 4:199-201.

Franklin, B.M., S.R. Voss, and J.L. Osborn. 2017. Ion channel signaling influences cellular proliferation and phagocyte activity during axolotl tail regeneration. *Mech Dev*. 146:42-54.

Frey, H., N. Schroeder, T. Manon-Jensen, R.V. Iozzo, and L. Schaefer. 2013. Biological interplay between proteoglycans and their innate immune receptors in inflammation. *FEBS J*. 280:2165–2179.

Fu, S-C., Y-P. Wong, Y-C. Cheuk, K-M. Lee, and K-M. Chan. 2005. TGF-β1 Reverses the Effects of Matrix Anchorage on the Gene Expression of Decorin and Procollagen Type I in Tendon Fibroblasts. *Clin. Orthop. Relat. Res*. 431: 226-232.

Fujita, A., K. Kojima, A.G. Patriota, J.R. Sato, P. Severino, S. Miyano, and J. Dopazo. 2011. A fast and robust statistical test based on likelihood ratio with Bartlett correction to identify Granger causality between gene sets. *Bioinformatics*. 27:2349-2351.

Funahashi, Y., C.J. Shawber, A. Sharma, E. Kanamaru, Y.K. Choi, and J. Kitajewski. 2011. Notch modualtes VEGF action in endothelial cells by inducing Matrix Metalloprotease activity. *Vasc Cell*. 3:2.

Garber, M., M. Guttman, M. Clamp, M.C. Zody, N. Friedman, and X. Xie. 2009. Identifying novel constrained elements by exploiting biased substitution patterns. *Bioinformatics*. 25:54-62.

Garrison, E., and G. Marth. 2012. Haplotype-based variant detection from short-read sequencing. *arXiv e-prints*. 1207:3907.

Ge, W., J. Mu, and C. Huang. 2014. The GDF5 SNP is associated with meniscus injury and function recovery in male Chinese soldiers. *Int J Sports Med*. 35: 625-628.

Gézsi, A., B. Bolgár, P. Marx, P. Sarkozy, C. Szalai, and P. Antal. 2015. VariantMetaCaller: automated fusion of variant calling pipelines for quantitative, precision-based filtering. *BMC Genomics*. 16:875.

Ghatak, S., E.V. Maytin, J.A. Mack, V.C. Hascall, I. Atanelishvili, R. Moreno Rodriguez, R.R. Markwald, and S. Misra. 2015. Roles of Proteoglycans and Glycosaminoglycans in Wound Healing and Fibrosis. *Int J Cell Biol*. 834893.

Gianotti, S.M., S.W. Marshall, and P.A. Hume. 2009. Incidence of anterior cruciate ligament injury and other knee ligament injuries: a national population-based study. *J Sci Med Sport*. 12:622-627.

Gibbon, A., H. Hobbs, W. Van der Merwe, S.M. Raleigh, J. Cook, and C.J.e.a. Handley. 2017. The *MMP3* gene in musculoskeletal soft tissue injury risk profiling: A study in two independent sample groups. *J Sports Sci*. 35:655-662.

Gibbon, A., S.M. Raleigh, W.J. Ribbans, M. Posthumus, M. Collins, and A.V. September. 2020. Functional *COL1A1* variants are associated with the risk of acute musculoskeletal soft tissue injuries. *J Orthop Res*. 38:2290-2298.

Gibbon, A., C.J. Saunders, M. Collins, J. Gamieldien, and A.V. September. 2018. Defining the molecular signature of Achilles tendinopahty and anterior cruciate ligament ruptures: A whole-exome sequencing approach. *PLoS One*. 13:e0205860.

Gibson, W. 2009. Genetic association studies for complex traits: relevance for the sports medicine practitioner. *Br J Sport Med*. 43:314-316.

Glass, D., and A.A. Nuara. 2013. Whole exome sequencing. *Dermat*. 24:33-34.

Gonzaga-Jauregui, C., G. Yesil, H. Nistala, A. Gezdirici, Y. Bayram, K.C. Nannuru, D. Pehlivan, B. Yuan, J. Jimenez, Y. Sahin, I.S. Paine, Z.C. Akdemir, S. Rajamani, J. Staples, J. Dronzek, K. Howell, J.M. Fatih, S. Smaldone, A.E. Schlesinger, N. Ramírez, A.S. Cornier, M.A. Kelly, R. Haber, S.M. Chim, K. Nieman, N. Wu, J. Walls, W. Poueymirou, C. Siao, V.R. Sutton, M.S. Williams, J.E. Posey, R.A. Gibbs, S. Carlo, D.H. Tegay, A.N. Economides, J.R. Lupski, and R.G.C. Baylor-Hopkins Center for Mendelian Genetics. 2020. Functional biology of the Steel syndrome founder allele and evidence for clan genomics derivation of *COL27A1* pathogenic alleles worldwide. *Eur J Hum Genet*. 28:1243-1264.

González, J.R., L. Armengol, X. Solé, E. Guinó, J.M. Mercader, X. Estivill, and V. Moreno. 2014. *SNPassoc*: An R package to perform whole genome association studies. *Bioinformatics*. 23:644-645.

Gornitzky, A.L., A. Lott, Yellin, J. L., P.D. Fabricant, J.T. Lawrence, and T.J. Ganley. 2016. Sport-Specific Yearly Risk and Incidence of Anterior Cruciate Ligament Tears in High School Athletes: A Systematic Review and Meta-analysis. *Am J Sports Med*. 44:2716-2723.

Goshima, K., K. Kitaoka, J. Nakase, and H. Tsuchiya. 2014. Familial predisposition to anterior cruciate ligament injury. *AP-SMART*. 1:62-66.

Gottlob, C.A., and C.L.J. Baker. 2000. Anterior cruciate ligament reconstruction: socioeconomic issues and cost effectiveness. *Am J Orthop*. 29:472-476.

Griffin, L.Y., M.J. Albohm, and E.A. Arendt. 2000. Understanding and preventing noncontact anterior cruciate ligament injuries: a review of the Hunt Valley II Meeting, Januaury 2005. *Am J Sports Med*. 34:1512-1532.

Gudykunst, W.B., and K.L. Schmidt. 1987. Language and ethnic identity: An overview and prologue. *J Lang Soc Psychol*. 6:157-170.

Gumucio, J.P., K.B. Sugg, and C.L. Mendias. 2015. TGF-beta superfamily signaling in muscle and tendon adaptation to resistance exercise. Exerc Sport Sci Rev. 43:93-99.

Gupta, P.A., E.A. Putnam, S.G. Carmical, I. Kaitila, B. Steinmann, A. Child, C. Danesino, K. Metcalfe, S.A. Berry, E. Chen, C.V. Delorme, M. Thong, L.C. Adès, and D.M. Milewicz. 2002. Ten novel *FBN2* mutations in congenital contractural arachnodactyly: Delineation of the molecular pathogenesis and clinical phenotype. *Hum Mutat*. 19:39-48.

Gurdasani, D., T. Carstensen, F. Tekola-Ayele, L. Pagani, I. Tachmazidou, K. Hatzikotoulas, S. Karthikeyan, L. Iles, M.O. Pollard, A. Choudhury, G.R. Ritchie, Y. Xue, J. Asimit, R.N. Nsubuga, E.H. Young, C. Pomilla, K. Kivinen, K. Rockett, A. Kamali, A.P. Doumatey, and M.S. ...Sandhu. 2015. The African Genome Variation Project shapes medical genetics in Africa. *Nature*. 517:327-332.

Gurlek, A.C., B. Sevinc, E. Bayrak, and C. Erisken. 2017. Synthesis and characterization of polycaprolactone for anterior cruciate ligament regeneration. *Mater Sci Eng C Mater Biol Appl*. 71:820-826.

Hansen, T.F., F.B. Sørensen, K.-L.G. Spindler, D.A. Olsen, R.F. Andersen, J. Lindebjerg, and A. Jakobsen. 2010. Microvessel density and the association with single nucleotide polymorphisms of the vascular endothelial growth factor receptor 2 in patients with colorectal cancer. *Virchows Arch*. 456:251-260.

Harner, C.D., L.E. Paulos, A.E. Greenwald, T.D. Rosenberg, and V.C. Cooley. 1994. Detailed analysis of patients with bilateral anterior cruciate ligament injuries. *Am J Sports Med*. 22:37-43.

Haslauer, C.M., B.L. Proffen, V.M. Johnson, and M.M. Murray. 2014. Expression of modulators of extracellular matrix structure after anterior cruciate ligament injury. *Wound Repair Regen*. 22:103-110.

Haus, J., and Z. Halata. 1990. Innervation of the anterior cruciate ligament. *Int Orthop*. 14:293-296.

Hauser, R.A., and E.E. Dolan. 2011. Ligament Injury and Healing: An Overview of Current Clinical Concepts. *Journal of Prolotherapy*. 3:836-846.

Hauser, R.A., E.E. Dolan, H.J. Phillips, A.C. Newlin, R.E. Moore, and B.A. Woldin. 2013. Ligament Injury and Healing: A Review of Current Clinical Diagnostics and Therapeutics. *OJTR*. 6:1-20.

Hay, M., J. Patricios, and R. Collins. 2013. Association of type XI collagen genes with chronic Achilles tendinopathy in independent populations from South Africa and Australia. *Br J Sports Med*. 47:569-574.

Heinegård, D. 2009. Proteoglycans and more--from molecules to biology. *Int J Exp Pathol*. 90:575–586.

Herzog, M.M., S.W. Marshall, J.L. Lund, V. Pate, and J.T. Spang. 2017. Cost of Outpatient Arthroscopic Anterior Cruciate Ligament Reconstruction Among Commercially Insured Patients in the United States, 2005-2013. *Orthop J Sports Med*. 5: 2325967116684776.

Hewett, T.E., K.R. Ford, B.J. Hoogenboom, and G.D. Myer. 2010a. Understanding and preventing ACL injuries: Current biomechanical and epidemiologic considerations - update 2010. *NAJSPT*. 5:234.

Hewett, T.E., T.R. Lynch, G.D. Myer, K.R. Ford, R.C. Gwin, and R.S. Heidt, Jr. 2010b. Multiple risk factors related to familial predisposition to anterior cruciate ligament injury: fraternal twin sisters with anterior cruciate ligament ruptures. *Br J Sports Med*. 44:848-855.

Hewett, T.E., S.J. Shultz, and L.Y. Griffin. 2007. Understanding and Preventing Noncontact ACL Injuries. Champaign, IL: Human Kinetics.

Hibino, H., A. Inanobe, K. Furutani, S. Murakami, I. Findlay, and Y. Kurachi. 2010. Inwardly rectifying potassium channels: their structure, function, and physiological roles. *Physiol Rev*. 90:291-366.

Hicks, D., G.T. Farsani, S. Laval, J. Collins, A. Sarkozy, E. Martoni, A. Shah, Y. Zou, M. Koch, C.G. Bönnemann, M. Roberts, H. Lochmüller, K. Bushby, and V. Straub. 2014. Mutations in the collagen XII gene define a new form of extracellular matrix-related myopathy. *Hum Mol Genet*. 23:2353-2363.

Hildebrand, M.S., M.R. Avenarius, M. Fellous, Y. Zhang, N.C. Meyer, J. Auer, C. Serres, K. Kahrizi, H. Najmabadi, J.S. Beckmann, and R.J. Smith. 2010. Genetic male infertility and mutation of CatSper ion channels. *Eur J Hum Genet*. 18:1178-1184.

Hoffmann, A., and G. Gross. 2007. Tendon and ligament engineering in the adult organism: mesenchymal stem cells and gene-therapeutic approaches. *Int Orthop*. 31:791-797.

Hu, H.H., D.Q. Chen, Y.N. Wang, Y.L. Feng, G. Cao, N.D. Vaziri, and Y.Y. Zhao. 2018. New insights into TGF-β/Smad signaling in tissue fibrosis. *Chem Biol Interact*. 292:76-83.

Humphrey, J.D., E.R. Dufresne, and M.A. Schwartz. 2014. Mechanotransduction and extracellular matrix homeostasis. *Nat Rev Mol Cell Biol*. 15:802-812.

Hunt, J. 2005. Dutch South Africa: Early Settlers at the Cape 1652-1708. H.-A. Campbell, editor. University of Pennsylvania Press, Philadelphia. 13-35.

Ilic, M.Z., P. Carter, A. Tyndall, J. Dudhia, and C.J. Handley. 2005. Proteoglycans and catabolic products of proteoglycans present in ligament. *Biochem J*. 385:381-388.

Imran, A., and J.J. O'Connor. 1998. Control of knee stability after ACL injury or repair:interaction between hamstrings contraction and tibial translation. *Clin Biomech*. 13:153-162.

Iozzo, R.V. 1997. The family of the small leucine-rich proteoglycans: key regulators of matrix assembly and cellular growth. *Crit Rev Biochem Mol Biol*. 32:141-174.

Iozzo, R.V., D.K. Moscatello, D.J. McQuillan, and I. Eichsteter. 1999. Decorin is a biological ligand for the epidermal growth factor receptor. *J Biol Chem*. 274.

Iozzo, R.V., and L. Schaefer. 2015. Proteoglycan form and function: A comprehensive nomenclature of proteoglycans. *Matrix Biol*. 42:11-55.

Irrthum, A., M.J. Karkkainen, K. Devriendt, K. Alitalo, and M. Vikkula. 2000. Congenital hereditary lymphedema caused by a mutation that inactivates VEGFR3 tyrosine kinase. *Am J Hum Genet*. 67:295-301.

Isogai, Z., A. Aspberg, D.R. Keene, R.N. Ono, D.P. Reinhardt, and L.Y. Sakai. 2002. Versican interacts with fibrillin-1 and links extracellular microfibrils to other connective tissue networks. *J Biol Chem*. 277:4565-4572.

Ito, S., S. Ishimaru, and S.E. Wilson. 1997. Inhibitory effect of type 1 collagen gel containing alpha-elastin on proliferation and migration of vascular smooth muscle and endothelial cells. *Cardiovasc Surg*. 5:176-183.

Järveläinen, H., A. Sainio, and T.N. Wight. 2015. Pivotal Role for Decorin in Angiogenesis. *Matrix Biol*. 43:15-26.

Jin, H., R.J. van't Hof, O.M.E. Albagha, and S.H. Ralston. 2009. Promoter and intron 1 polymorphisms of *COL1A1* interact to regulate transcription and susceptibility to osteoporosis. *Hum Mol Genet*. 18:2729-2738.

Johnson, R.J. 1983. The anterior cruciate ligament problem. *Clin Orthop Relat Res*. 172:14-18.

Joseph, A.M., C.L. Collins, N.M. Henke, E.E. Yard, S.K. Fields, and R.D. Comstock. 2013. A multisport epidemiologic comparison of anterior cruciate ligament injuries in high school athletics. *J Athl Train*. 48:810-817.

Kaeding, C.C., B. Léger-St-Jean, and R.A. Magnussen. 2017. Epidemiology and Diagnosis of Anterior Cruciate Ligament Injuries. *Clin Sports Med*. 36:1-8.

Kanehisa, M., and S. Goto. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 28:27-30.

Karczewski, K.J., B. Weisburd, B. Thomas, M. Solomonson, D.M. Ruderfer, D. Kavanagh, T. Hamamsy, M. Lek, K.E. Samocha, B.B. Cummings, and D. Birnbaum. 2017. The Exome Aggregation Consortium, Daly, M. J., & MacArthur, D. G. (2017). The ExAC browser: displaying reference data information from over 60 000 exomes. *Nucleic Acids Res*. 45:D840-D845.

Karthaus, M., X. Schiel, C.H. Ruhlmann, and L. Celio. 2019. Neurokinin-1 receptor antagonists: review of their role for the prevention of chemotherapy-induced nausea and vomiting in adults. *Expert Rev Clin Pharmacol*. 12:661-680.

Kay, J., D. de Sa, J. Karlsson, V. Musahl, and O.R. Ayeni. 2015. Anterior Cruciate Ligament Rupture: A Family Affair. *Orthop J Sports Med*. 3:2325967115616783.

Kaynak, M., F. Nijman, J. van Meurs, M. Reijman, and D.E. Meuffels. 2017. Genetic Variants and Anterior Cruciate Ligament Rupture: A Systematic Review. *Sports Med*. 47:1637-1650.

Keays, S.L., R. Keays, and P.A. Newcombe. 2016. Femoral intercondylar notch width size: a comparison between siblings with and without anterior cruciate ligament injuries. *Knee Surg Sports Traumatol Arthrosc*. 24:672-679.

Kennedy, J.C., I.J. Alexander, and K.C. Hayes. 1982. Nerve supply of the human knee and its functional importance. *Am J Sports Med*. 10:329-335.

Kerr, J.F.R., A.H. Wyllie, and A.R. Currie. 1972. Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics. *Br J Cancer*. 26:239-257.

Khoschnau, S., H. Melhus, A. Jacobson, H. Rahme, H. Bengtsson, and E. Ribom. 2008. Type I Collagen α 1 Sp1 Polymorphism and the Risk of Cruciate Ligament Ruptures of Shoulder Dislocations. *Am J Sports Med*. 36 2432-2436.

Kiani, C., V. Lee, L. Cao, L. Chen, Y. Wu, Y. Zhang, M.E. Adams, and B.B. Yang. 2001. Roles of aggrecan domains in biosynthesis, modification by glycosaminoglycans and product secretion. *Biochem J*. 354:199-207.

Kiani, C., L. Chen, Y.J. Wu, A.J. Yee, and B.B. Yang. 2002. Structure and function of aggrecan. *Cell Res*. 12:19-32.

Kiapour, A.M., and M.M. Murray. 2014. Basic science of anterior cruciate ligament injury and repair. *Bone Joint Res*. 3:20-31.

Kim, J. 2014. Channelopathies. *Korean J Pediatr*. 57:1-18.

Kim, S.H., J.W. Mok, H.S. Kim, and C.K. Joo. 2008. Association of -31T>C and -511 C>T polymorphisms in the interleukin 1 beta (IL1B) promoter in Korean keratoconus patients. *Mol Vis*. 14:2109-2116.

Kim, S.K., C. Nguyen, A.L. Avins, and G.D. Abrams. Three genes associated with anterior and posterior cruciate ligament injury:a genome-wide association analysis. 2021. *Bone Jt Open*. 2:414-421.

Kim, S.K., T. Roos, A. Roos, J. Kleimeyer, M. Ahmed, G. Goodlin, M. Fredericson, J. Loannidis, A. Avins, and J. Dragoo. 2017. Genome-wide association screens for Achilles tendon and ACL tears and tendinopathy. *PLoS Medicine*. 12:1-16.

Kircher, M., D.M. Witten, P. Jain, B.J. O'Roak, G.M. Cooper, and J. Shendure. 2014. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Gen*. 46:310-315.

Kluivers, K.B., J.R. Dijkstra, J.C.M. Hendriks, S.L. Lince, M.E. Vierhout, and L.C. van Kempen. 2009. *COL3A1* 2209G>A is a predictor of pelvic organ prolapse. *Int Urogynecol J Pelvic Floor Dysfunct*. 20:1113-1118.

Koch, M., F. Laub, P. Zhou, R.A. Hahn, S. Tanaka, R.E. Burgeson, D.R. Gerecke, F. Ramirez, and M.K. Gordon. 2003. Collagen XXIV, a vertebrate fibrillar collagen with structural features of invertebrate collagens: selective expression in developing cornea and bone. *J Biol Chem*. 278:43236-43244.

Koga, H., T. Sakou, E. Taketomi, K. Hayashi, T. Numasawa, S. Harata, K. Yone, S. Matsunaga, B. Otterud, I. Inoue, and M. Leppert. 1998. Genetic mapping of ossification of the posterior longitudinal ligament of the spine. *Am J Hum Genet*. 62:1460-1467.

Koga, K., Y. Osuga, O. Yoshino, Y. Hirota, X. Ruimeng, T. Hirata, S. Takeda, T. Yano, O. Tsutsumi, and Y. Taketani. 2003. Elevated Serum Soluble Vascular Endothelial Growth Factor Receptor 1 (sVEGFR-1) Levels in Women with Preeclampsia. *J Clin Endocrinol Metab*. 88:2348–2351.

Kolb, M., P.J. Margetts, P.J. Sime, and J. Gauldie. 2001. Proteoglycans decorin and biglycan differentially modulate TGF-beta-mediated fibrotic responses in the lung. *Am J Physiol Lung Cell Mol Physiol*. 280:1327-1334.

Koyama, K., K. Nakazato, S. Min, K. Gushiken, Y. Hatakeda, K. Seo, and K. Hiranuma. 2012. *COL11A1* gene is associated with limbus vertebra in gymnasts. *Int J Sports Med*. 33:586-590.

Kresse, H., and E. Schönherr. 2001. Proteoglycans of the extracellular matrix and growth control. *J Cell Physiol*. 189:266-274.

Kruidering, M., and G.I. Evan. 2000. Caspase-8 in apoptosis: the beginning of “the end”? *IUBMB Life*. 50:85-90.

Kuc, I.M., and P.G. Scott. 1997. Increased diameters of collagen fibrils precipitated in vitro in the presence of decorin from various connective tissues. *Connect Tissue Res*. 36:2287-2296.

Kuivaniemi, H., and G. Tromp. 2019. Type III collagen (*COL3A1*): Gene and protein structure, tissue distribution, and associated diseases. *Gene*. 707:151-171.

Kuleshov, M.V., M.R, Jones, A.D. Rouillard, N.F. Fernandez, Q. Duan Z. Wang, S. Koplev, S.L. Jenkins, K.M. Jagodnik, A. Lachmann, M.G. McDermott, C.D. Monteiro, G.W. Gundersen, and A. Ma'ayan. 2016. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res*. 44:W90–W97.

Kung, J.T., D. and Colognori, J.T. Lee. Long noncoding RNAs: past, present, and future. 2013. *Genetics*. 193:651-69.

Kweon, C., E.S. Lederman, and A. Chhabra. 2015. Anatomy and Biomechanics of the Posterior Cruciate Ligaments and Their Surgical Implications. In Posterior Cruciate Ligament Injuries: A Practical Guide to Management. G.C. Fanelli, editor. Springer, Dordrecht, London. 17-27.

Laguette, M.N., Y. Abrahams, S. Prince, and M. Collins. 2011. Sequence variants within the 3′-UTR of the *COL5A1* gene alters mRNA stability : Implications for musculoskeletal soft tissue injuries. *Matrix Biol*. 30:338-345.

Laguette, M.N., K. Barrow, F. Firfirey, S. Dlamini, C.J. Saunders, C. Dandara, J. Gamieldien, M. Collins, and A.V. September. 2020. 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*. 38:1856-1865.

Lahiri, D.K., and J.J.I. Nurnberger. 1991. A rapid non-enzymatic method for the preparation of HMW DNA from blood for RFLP studies. *Nucleic Acids Res*. 19:5444.

Lambert, K.L. 1984. The syndrome of the torn anterior cruciate ligament. *Adv Orthop*. 7:e314.

Lambrechts, D., E. Storkebaum, M. Morimoto, J. Del-Favero, F. Desmet, and S.L. Marklund. 2003. VEGF is a modifier of amyotrophic lateral sclerosis in mice and humans and protects motoneurons against ischemic death. *Nat Genet*. 34:383-394.

LaPrade, R.F., and Q.M. Burnett. 1994. Femoral intercondylar notch stenosis and correlation to anterior cruciate ligament injuries. *Am J Sports Med*. 22:198-203.

Lepetsos, P., A. Pampanos, E. Kanavakis, M. Tzetis, D. Korres, A.G. Papavassiliou, and N. Efstathopoulos. 2014. Association of MMP-1-1607 1G/2G (rs1799750) polymorphism with primary knee osteoarthritis in the Greek population. *J Orthop Res*. 32:1155–1160.

Letafatkar, A., R. Rajabi, E.E. Tekamejani, and H. Minoonejad. 2015. Effects of perturbation training on knee flexion angle and quadriceps to hamstring cocontraction of female athletes with quadriceps dominance deficit: Pre–post intervention study. *The Knee*. 22:230-236.

Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 25:1754-1760.

Li, H., J. Ruan, and R. Durbin. 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res*. 18:1851-1858.

Li, M.X., J.S. Kwan, S.Y. Bao, W. Yang, S.L. Ho, Y.Q. Song, and P.C. Sham. 2013. Predicting mendelian disease-causing non-synonymous single nucleotide variants in exome sequencing studies. *PLoS genetics*. 9:e1003143.

Liu-Ambrose, T. 2003. The anterior cruciate ligament and functional stability of the knee joint. *BCMJ*. 45:495-499.

Liu, Q., Y. Guo, J. Li, J. Long, B. Zhang, and Y. Shyr. 2012. Steps to ensure accuracy in genotype and SNP calling from Illumina sequencing data. *BMC genomics*. 13 Suppl 8: S8.

Liu, X., S. Han, Z. Wang, J. Gelernter, and B.Z. Yang. 2013. Variant callers for next-generation sequencing data: a comparison study. *PloS one*. 8:e75619.

Liu, Y., Y. Li, M.E. March, K. Nguyen, K. Xu, F. Wang, Y. Guo, B. Keating, J. Glessner, J. Li, T.J. Ganley, J. Zhang, M.A. Deardorff, X. Xu, and H. Hakonarson. 2015. Copy number variation in *CEP57L1* predisposes to congenital absence of bilateral ACL and PCL ligaments. *Hum Genomics*. 9:31.

Lloyd, D.G., T.S. Buchanan, and T.F. Besier. 2005. Neuromuscular biomechanical modeling to understand knee ligament loading. *Med Sci Sports Exerc*. 37:1939-1947.

Loebig, M., J. Klement, A. Schmoller, S. Betz, N. Heuck, U. Schweiger, and K.M. Oltmanns. 2010. Evidence for a relationship between VEGF and BMI independent of insulin sensitivity by glucose clamp procedure in a homogenous group healthy young men. *Plos One*. 5:1–5.

Loh, P.R., P.F. Palamara, and A.L. Price. 2016. Fast and accurate long-range phasing in a UK Biobank cohort. *Nat Gen*. 48:811-816.

Long, A. 2001. Encyclopedia of Genetics. S. Brenner, and J.H. Miller, editors. Academic Press. San Diego, London. 2257p.

Lord, B., J. Grice, G. Cox, S. Yasen, and A. Wilson. 2015. (iii) Anterior cruciate ligament reconstruction – evolution and current concepts. *Orthop Trauma*. 29:12-23.

Lu, Y., Y. Ge, Y. Shi, J. Yin, and Z. Huang. 2013. Two polymorphisms (rs699947, rs2010963) in the *VEGFA* gene and diabetic retinopathy: an updated meta-analysis. *BMC Ophthalmology*. 13:56.

Luan, A., M.S. Hu, T. Leavitt, E.A. Brett, K.C. Wang, M.T. Longaker, and D.C. Wan. 2018. Noncoding RNAs in Wound Healing: A New and Vast Frontier. *Adv Wound Care*. 7:19-27.

Lubeck, E., A.F. Coskun, T. Zhiyentayev, M. Ahmad, and L. Cai. 2012. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods*. 9:743-748.

Lulinska-Kuklik, E., A. Gibbon, and M.E.A. Kaczmarczyk. 2020. Matrix Metalloproteinase Genes (*MMP1, MMP10, MMP12*) on Chromosome 11q22 and the Risk of Non-Contact Anterior Cruciate Ligament Ruptures. *Genes*. 11:766.

Lulinska-Kuklik, E., K. Leznicka, K. Huminska-Lisowska, W. Moska, M. Michalowska-Sawczyn, Z. Ossowski, E. Maculewicz, P. Cieszczyk, M. Kaczmarczyk, W. Ratkowski, K. Ficek, P. Zmijewski, and A. Leonska-Duniec. 2019a. The *VEGFA* gene and anterior cruciate ligament rupture risk in the Caucasian population. *Biol Sport*. 36:3-8.

Lulińska-Kuklik, E., E. Maculewicz, W. Moska, and K. Ficek, M. Kaczmarczyk, M. Michałowska-Sawczyn, K. Humińska-Lisowska, M. Buryta, J. Chycki, P. Cięszczyk, P. Żmijewski, A. Rzeszutko, M. Sawczuk, P.Stastny, M.Petr, and A. Maciejewska-Skrendo. 2019. Are *IL1B, IL6* and *IL6R* Gene Variants Associated with Anterior Cruciate Ligament Rupture Susceptibility? *J Sports Sci Med*. 18:137-145.

Lulińska-Kuklik, E., M. Rahim, D. Domańska-Senderowska, K. Ficek, M. Michałowska-Sawczyn, W. Moska, M. Kaczmarczyk, M. Brzeziański, E. Brzeziańska-Lasota, P. Cięszczyk, and A.V. September. 2018. Interactions between *COL5A1* Gene and Risk of the Anterior Cruciate Ligament Rupture. *J Hum Kinet*. 62:65-71.

Lulinska-Kuklik, E., M. Rahim, W. Moska, E. Maculewicz, M. Kaczmarczyk, A. Maciejewska-Skrendo, K. Ficek, P. Cieszczyk, A.V. September, and M. Sawczuk. 2019b. Are *MMP3, MMP8* and *TIMP2* gene variants associated with anterior cruciate ligament rupture susceptibility? *J Sci Med Sport*. 22:753-757.

Maeda, S., Y. Ishidou, H. Koga, E. Taketomi, K. Ikari, S. Komiya, J. Takeda, T. Sakou, and I. Inoue. 2001. Functional impact of human collagen alpha2(XI) gene polymorphism in pathogenesis of ossification of the posterior longitudinal ligament of the spine. *Bone Miner Res*. 16:948-957.

Magnusson, K., A. Turkiewicz, V. Hughes, R. Frobell, and M. Englund. 2020. High genetic contribution to anterior cruciate ligament rupture: Heritability ~69%. *Br J Sports Med*. 55:385-389.

Majava, M., K.P. Hoornaert, D. Bartholdi, M.C. Bouma, K. Bouman, M. Carrera, K. Devriendt, J. Hurst, G. Kitsos, D. Niedrist, M.B. Petersen, D. Shears, I. Stolte-Dijkstra, J.M. Van Hagen, L. Ala-Kokko, M. Männikkö, and G.R. Mortier. 2007. A report on 10 new patients with heterozygous mutations in the *COL11A1* gene and a review of genotype-phenotype correlations in type XI collagenopathies. *Am J Med Genet*. 143:258-264.

Malfait, F., and A. De Paepe. 2005. Molecular genetics in classic Ehlers–Danlos syndrome. *Am J Med Genet*. 139:17-23.

Malila, S., P. Yuktanandana, S. Saowaprut, W. Jiamjarasrangsi, and S. Honsawek. 2011. Association between matrix metalloproteinase-3 polymorphism and anterior cruciate ligament ruptures. *Genet Mol Res*. 10:4158-4165.

Mall, N.A., P.N. Chalmers, M. Moric, M.J. Tanaka, B.J. Cole, B.R.J. Bach, and G.A. Paletta, Jr. . 2014. Incidence and trends of anterior cruciate ligament reconstruction in the United States. *Am J Sports Med*. 42:2363-2370.

Mall, N.A., A.S. Lee, B.J. Cole, and N.N. Verma. 2013. The Functional and Surgical Anatomy of the Anterior Cruciate Ligament. *Oper Tech Sports Med*. 21:2-9.

Mann, V., E.E. Hobson, and B. Li. 2001. A *COL1A1* Sp1 binding site polymorphism predisposes to osteoporotic fracture by affecting bone density and quality. *J Clin Invest*. 107:899-907.

Mannion, S., A. Mtintsilana, M. Posthumus, W. van der Merwe, H. Hobbs, M. Collins, and A.V. September. 2014. Genes encoding proteoglycans are associated with the risk of anterior cruciate ligament ruptures. *Br J Sports Med*. 48:1640-1646.

Markolf, K.L., D.M. Burchfield, M.M. Shapiro, M.F. Shepard, G.A. Finerman, and J.L. Slauterbeck. 1995. Combined knee loading states that generate high anterior cruciate ligament forces. *J Orthop Res*. 13:930-935.

Markström, J. L., H. Grip, L. Schelin, and C.K. Häger. 2019. Dynamic knee control and movement strategies in athletes and non-athletes in side hops: Implications for knee injury. *Scand J Med Sci Sports*. 29:1181-1189.

Marshall, S., D. Padua, and M. McGrath. 2007. Incidence of ACL injuries. *In* Understanding and preventing noncontact ACL injuries. T. Hewett, S. Shultz, and L. Griffin, editors. Human Kinetics, Champaign, IL. 5-30.

Martirosyan, N.L., A.A. Patel, A. Carotenuto, M.Y. Kalani, E. Belykh, C.T. Walker, M.C. Preul, and N. Theodore. 2016. Genetic alterations in intervertebral disc disease. *Front Surg*. 3:59.

Masabumi, S. 2013. Vascular endothelial growth factor and its receptor system: physiological functions in angiogenesis and pathological roles in various diseases. *J Biochem*. 153:13-19.

Mattick, J.S. 2004. RNA regulation: a new genetics? *Nat Rev Genet*. 5:316-323.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M.A. DePristo. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 20:1297–1303.

Meester, J.A.N., G. Vandeweyer, I. Pintelon, M. Lammens, L. Van Hoorick, S. De Belder, K. Waitzman, L. Young, L.W. Markham, J. Vogt, J. Richer, L. M. Beauchesne, S. Unger, A. Superti-Furga, M. Prsa, R. Dhillon, E. Reyniers, H.C. Dietz, W. Wuyts, G. Mortier, A. Verstraeten, L. Van Laer, and B.L. Loeys. 2017. Loss-of-function mutations in the X-linked biglycan gene cause a severe syndromic form of thoracic aortic aneurysms and dissections. *Genet Med*. 19:386-395.

Meeuwisse, W.H. 1994. Assessing causation in sport injury: a multifactorial model. *Clin J Sport Med*. 4:166-170.

Merline, R., K. Moreth, J. Beckmann, M.V. Nastase, J. Zeng-Brouwers, J.G. Tralhao, P. Lemarchand, J. Pfeilschifter , R.M. Schaefer, R.V. Iozzo, and L. Schaefer. 2011. Signaling by the matrix proteoglycan decorin controls inflammation and cancer through PDCD4 and MicroRNA-21. *Sci Signal*. 4:75.

Mi, H., X. Huang, A. Muruganujan, H. Tang, C. Mills, D. Kang, and P.D. Thomas. 2016. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res*. 45:D183-D189.

Michalopoulos, S. 2012. The Origins of Ethnolinguistic Diversity. *Am Econ Rev*. 102:1508-1539.

Millar, N.L., A.Q. Wei, T.J. Molloy, F. Bonar, and G.A. Murrell. 2009. Cytokines and apoptosis in supraspinatus tendinopathy. *J Bone Joint Surg Am*. 91:417-424.

Mio, F., K. Chiba, Y. Hirose, Y. Kawaguchi, Y. Mikami, T. Oya, M. Mori, M. Kamata, M. Matsumoto, K. Ozaki, T. Tanaka, A. Takahashi, T. Kubo, T. Kimura, Y. Toyama, and S. Ikegawa. 2007. A functional polymorphism in *COL11A1*, which encodes the alpha 1 chain of type XI collagen, is associated with susceptibility to lumbar disc herniation. *Am J Hum Genet*. 81:1271-1277.

Mobasheri, A., R. Lewis, A. Ferreira-Mendes, A. Rufino, C. Dart, and R. Barrett-Jolley. 2012. Potassium channels in articular chondrocytes. *Channels*. 6:416-425.

Mokone, G.G., M. Gajjar, A.V. September, M.P. Schwellnus, J. Greenberg, and T.D. Noakes 2005. The guanine-thymine dinucleotide repeat polymorphism within the tenascin-C gene is associated with achilles tendon injuries. *Am J Sports Med*. 33:1016-1021.

Mokone, G.G., M.P. Schwellnus, T.D. Noakes, and M. Collins. 2006. The *COL5A1* gene and Achilles tendon pathology. *Scand J Med Sci Sports*. 16 19-26.

Morawski, M., G. Brückner, T. Arendt, and R.T. Matthews. 2012. Aggrecan: beyond the cartilage and into the brain. *Int J Biochem Cell Biol*. 44:690-693.

Moscatello, D.K., M. Santra, D.M. Mann, D.J. McQuillan, A.J. Wong, and R.V. Iozzo. 1998. Decorin suppresses tumor cell growth by activating the epidermal growth factor receptor. *J Clin Invest*. 101:106-142.

Mousavizadeh, R., S. Khosravi, H. Behzad, R.G. McCormack, V. Duronio, and A. Scott. 2014. Cyclic strain alters the expression and release of angiogenic factors by human tendon cells. *PLoS One*. 9:e97356.

Mukhopadhyay, A., K. Nikopoulus, A. Maugeri, A.P.M. de Brouwer, C.E. van Nouhuys, C.J.F. Boon, R. Perveen, H.A.A. Zegers, D. Wittebol-Post, P.R. van den Biesen, S.D van der Velde-Visser, H.G. Brunner, G.C.M. Black, C.B. Hoyng, and F.P.M. Cremers. 2006. Erosive vitreoretinopathy and Wagner disease are caused bi intronicmutations in CSPG2/Versican that result in an imbalance of splice variants. *Investig Ophthalmol Vis Sci*. 47:3565-3572.

Mullikin, J.C., S.E. Hunt, C.G. Cole, B.J. Mortimore, C.M. Rice, J. Burton, L.H. Matthews, R. Pavitt, R.W. Plumb, S.K. Sims, R.M. Ainscough, J. Attwood, J.M. Bailey, K. Barlow, R.M. Bruskiewich, P.N. Butcher, N.P. Carter, Y. Chen, C.M. Clee, P.C. Coggill, J. Davies, R.M. Davies, E. Dawson, M.D. Francis, A.A. Joy, R.G. Lamble, C.F. Langford, J. Macarthy, V. Mall, A. Moreland, E.K. Overton-Larty, M.T. Ross, L.C. Smith, C.A. Steward, J.E. Sulston, E.J. Tinsley, K.J. Turney, D.L. Willey, G.D. Wilson, A.A. McMurray, I. Dunham, J. Rogers, and D.R. Bentley. 2000. An SNP map of human chromosome 22. *Nature*. 407.

Murray, M.M. 2009. Current status and potential of primary ACL repair. *Clin Sports Med*. 28:51-61.

Myer, G.D., K.R. Ford, and T.E. Hewett. 2004. Rationale and clinical techniques for anterior cruciate ligament injury prevention among female athletes. *J Athl Train*. 39:352-364.

Narasimhan, V., P. Danecek, A. Scally, Y. Xue, C. Tyler-Smith, and R. Durbin. 2016. BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics*. 32:1749-1751.

Nell, E.-M., L. van der Merwe, J. Cook, C.J. Handley, M. Collins, and A.V. September. 2012. The apoptosis pathway and the genetic predisposition to Achilles tendinopathy. *J Orthop Res*. 30:1719-1724.

Ng, P.C., and S. Henikoff. 2003. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res*. 31:3812-3814.

Ng, P.C., and S. Henikoff. 2006. Predicting the effects of amino acid substitutions on protein function. *Annu Rev of Genomics and Hum Genet*. 7:61-80.

Nilius, B., and G. Droogmans. 2001. Ion channels and their functional role in vascular endothelium. *Physiol Rev*. 81:1415-59.

Nishida, N., H. Yano, T. Nishida, T. Kamura, and M. Kojiro. 2006. Angiogenesis in cancer. *Vasc Health Risk Manag*. 2:213-219.

Nishimura, D. 2001. BioCarta. Biotech Software & Internet Report. *The Computer Software Journal for Scient*. 2:117-120.

Noyes, F.R. 2009. The function of the human anterior cruciate ligament and analysis of single- and double-bundle graft reconstructions. *Sports Health*. 1:66-75.

O'Connell, K., H. Knight, K. Ficek, A. Leonska-Duniec, A. Maciejewska-Karlowska, and M. Sawczuk. 2015. Interactions between collagen gene variants and risk of anterior cruciate ligament rupture. *Eur J Sport Sci*. 29:1-10.

O'Neill, K., S. Pujar, S.H. Rangwala, D. Rausch, L.D. Riddick, C. Schoch, A. Shkeda, S.S. Storz, H. Sun, F. Thibaud-Nissen, I. Tolstoy, R.E. Tully, A.R. Vatsan, C. Wallin, D. Webb, W. Wu, M.J. Landrum, A. Kimchi, T. Tatusova, M. DiCuccio, P. Kitts, T.D. Murphy, and K.D. Pruitt. 2016. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*. 44:D733-D745.

Peeters, S., A. Decramer, S.A. Cain, P. Houpt, F. Verstreken, J. Noyez, C. Hermans, W. Jacobs, M. Lammens, E. Fransen, A.A. Kumar, G. Vandeweyer, B. Loeys, W. Van Hul, C. Baldock, E. Boudin, and G. Mortier. 2020. Delineation of a new fibrillino-2-pathy with evidence for a role of *FBN2* in the pathogenesis of carpal tunnel syndrome. *J Med Genet*. jmedgenet-2020-107085.

Pelkowski, J.N., K. DeMatas, R.C. McCoy, and C.J. Ortiguera. 2020. Seeing Double: A Case of ACL Tears in Monozygotic Twin Female Athletes Within 48 Hours. *Cureus*. 12:e7244.

Perera, R.S., P.H. Poruwalage Harsha Dissanayake, U. Senarath, L.S. Wijayaratne, A.L. Karunanayake, V. Harshadeva, and W. Dissanayake. 2017. Single Nucleotide Variants of Candidate Genes in Aggrecan Metabolic Pathway Are Associated with Lumbar Disc Degeneration and Modic Changes. *PLoS One*. 12:e0169835.

Petersen, W., and B. Tillmann. 1999. Structure and vascularization of the cruciate ligaments of the human knee joint. *Anat Embryol*. 2003:325-334.

Petersen, W., and B. Tillmann. 2002. Anatomy and function of the anterior cruciate ligament. *Orthopade*. 31:710-718.

Petersen, W., D. Varoga, T. Zantop, J. Hassenpflug, R. Mentlein, and T. Pufe. 2004. Cyclic strain influences vascular endothelial growth factor (VEGF) and the hypoxia inducible factor 1 alpha (HIF-la) in tendon fibroblasts. *J Orthop Res*. 22:847-853.

Pfeifer, C.E., P.F. Beattie, R.S. Sacko, and A. Hand. 2018. Risk factors associated with non-contact anterior cruciate ligament injury: a systematic review. *Int J Sports Phys Ther*. 13:575-587.

Philippe, P., and O. Mansi. 1998. Nonlinearity in the epidemiology of complex health and disease processes. *Theor Med Bioeth*. 19:591-607.

Pitsiladis, Y.P., M. Tanaka, N. Eynon, C. Bouchard, K.N. North, A.G. Williams, M. Collins, C.N Moran, S.L. Britton, N. Fuku, E.A. Ashley, V. Klissouras, A. Lucia, I.I. Ahmetov, E. de Geus, M. Alsayrafi, Athlome Project Consortium. 2016. Athlome Project Consortium: a concerted effort to discover genomic and other "omic" markers of athletic performance. *Physiol Genomics*. 48:183-90.

Plsek, P., and T. Greenhalgh. 2001. Complexity science: the challenge of complexity in health care. *BMJ*. 323:625-628.

Plumb, D.A., L. Ferrara, T. Torbica, L. Knowles, A. Mironov, K.E. Kadler, M.D. Briggs, and R.P. Boot-Handford. 2011. Collagen XXVII organises the pericellular matrix in the growth plate. *PLoS one*. 6:e29422.

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

Posthumus, M., A.V. September, D.O. Cuinneagain, W. Van Der Merwe, M.P. Schwellnus, and M. Collins. 2010. The association between the COL12A1 gene and anterior cruciate ligament ruptures. *Br J Sports Med*. 44:1160-1165.

Posthumus, M., A.V. September, and M. Keegan. 2009a. Genetic risk factors for anterior cruciate ligament ruptures: *COL1A1* gene variant. *Br J Sports Med*. 43:352-356.

Posthumus, M., A.V. September, D. O'Cuinneagain, W. Van der Merwe, and M.P. Schwellnus. 2009b. The *COL5A1* Gene Is Associated With Increased Risk of Anterior Cruciate Ligament Ruptures in Female Participants. *Am J Sports Med*. 37:2234-2240.

Posthumus, M., A.V. September, M.P. Schwellnus, and M. Collins. 2008. Investigation of the Sp1-binding site polymorphism within the *COL1A1* gene in participants with Achilles tendon injuries and controls. *J Sci Med Sport*. 12:184-189.

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

Pufe, T., W.J. Petersen, R. Mentlein, and B.N. Tillmann. 2005. The role of vasculature and angiogenesis for the pathogenesis of degenerative tendons disease. *Scand J Med Sci Sports*. 15:211-222.

Rahim, M., M. Collins, and A.V. September. 2016a. Genes and musculoskeletal soft-tissue injuries. *In* Genetics and Sports. M. Posthumus and M. Collins, editors. Karger South Africa, Cape Town. 69-91.

Rahim, M., L. El Khoury, S.M. Raleigh, W.J. Ribbans, M. Posthumus, M. Collins, and A.V. September. 2016b. Human Genetic Variation, Sport and Exercise Medicine, and Achilles Tendinopathy: Role for Angiogenesis-Associated Genes. *OMICS*. 20:520-527.

Rahim, M., A. Gibbon, M. Collins, and A.V. September. 2019. Genetics of musculoskeletal soft tissue injuries: current status, challenges, and future directions. *In* Sports, Exercise, and

Nutritional Genomics. D. Barh and I. Ahmetov, editors. Academic Press Cambridge, MA. 317-339.

Rahim, M., A. Gibbon, H. Hobbs, W. van der Merwe, M. Posthumus, M. Collins, and A.V. September. 2014. The association of genes involved in the angiogenesis-associated signaling pathway with risk of anterior cruciate ligament rupture. *J Orthop Res*. 32:1612-1618.

Rahim, M., H. Hobbs, W. van der Merwe, M. Posthumus, M. Collins, and A.V. September. 2018. Investigation of angiogenesis genes with anterior cruciate ligament rupture risk in a South African population. *J Sports Sci*. 36:551-557.

Rahim, M., M. Lacerda, M. Collins, M. Posthumus, and A.V. September. 2021. Risk modelling further implicates the angiogenesis pathway in anterior cruciate ligament ruptures. *Eur J Sport Sci*.1-8.

Rahim, M., S. Mannion, B. Klug, H. Hobbs, W. van der Merwe, M. Posthumus, M. Collins, and A.V. September. 2017. Modulators of the extracellular matrix and risk of anterior cruciate ligament ruptures. *J Sci Med Sport*. 20:152-158.

Rahnemai-Azar, A.A., Z. Yaseen, C.F. van Eck, J.J. Irrgang, F.H. Fu, and V. Musahl. 2016. Increased lateral tibial plateau slope predisposes male college football players to anterior cruciate ligament injury. *J Bone Joint Surg Am*. 98:1001-1006.

Rakesh, J., S. Prabhakar, M.S. Dhillon, A. Anand, and G. Minhas. 2016. Association of ACL tears and single nucleotide polymorphisms in the collagen12A1 gene in the Indian population - a preliminary case-control study. *Muscles, Ligaments and Tendons J*. 6:253-257.

Raleigh, S.M., L. van der Merwe, W.J. Ribbans, R.K.W. Smith, M.P. Schwellnus, and M. Collins. 2009. Variants within the *MMP3* gene are associated with Achilles tendinopathy: possible interaction with the *COL5A1* gene. *Br J Sports Med*. 43:514-520.

Raleigh, S.M., M. Posthumus, D. O'Cuinneagain, W. van der Merwe, and M. Collins. 2013. The GDF5 gene and anterior cruciate ligament rupture. *Int J Sports Med.* 34: 364-367.

Ransohoff, J.D., Y. Wei, and P.A. Khavari. 2018. The functions and unique features of long intergenic non-coding RNA. *Nat Rev Mol Cell Biol*. 19:143-157.

Reva, B., Y. Antipin, and C. Sander. 2007. Determinants of protein function revealed by combinatorial entropy optimization. *Genome Biol*. 8:R232.

Reva, B., Y. Antipin, and C. Sander. 2011. Predicting the functional impact of protein mutations: Application to cancer genomics. *Nucleic Acids Res*. 39:37-43.

Rickaby, R., L.Y. El Khoury, T. Samiric, S.M. Raleigh. 2019. Epigenetic status of the human mmp11 gene promoter is altered in patellar tendinopathy. *J Sports Sci Med*. 18:155-159.

Rickles, D., P. Hawe, and A. Shiell. 2007. A simple guide to chaos and complexity. *J Epidemiol Community Health*. 61:933–937.

Roughley, P.J. 2006. The structure and function of cartilage proteoglycans. *Eur Cell Mater*. 12:92-101.

Ruigrok, Y.M., G.J.E. Rinkel, R. van't Slot, M. Wolfs, S. Tang, and C. Wijmenga. 2006. Evidence in favor of the contribution of genes involved in the maintenance of the extracellular matrix of the arterial wall to the development of intracranial aneurysms. *Hum Mol Genet*. 15:3361-8.

Ruiz, A.L., M. Kelly, and R.W. Nutton. 2002. Arthroscopic ACL reconstruction: A 5-9 year follow-up. *Knee*. 9:197-200.

Ruoslaht, E., and Y. Yamaguchi. 1991. Proteoglycans as modulators of growth factor activities. *Cell*. 64:867-869.

Safiri, S., A.-A. Kolahi, M. Cross, C. Hill, E. Smith, K. Carson-Chahhoud, M.A. Mansournia, A. Almasi-Hashiani, A. Ashrafi-Asgarabad, J. Kaufman, M. Sepidarkish, S.K. Shakouri, D. Hoy, A.D. Woolf, L. March, G. Collins, and R. Buchbinder. 2021. Prevalence, Deaths, and Disability-Adjusted Life Years Due to Musculoskeletal Disorders for 195 Countries and Territories 1990–2017. *Arthritis Rheumatol*. 73:702-714.

Sakai, L.Y., and D.R. Keene. 2019. Fibrillin protein pleiotropy: Acromelic dysplasias. *Matrix Biol*. 80:6-13.

Saleque, S., N. Ruiz, and K. Drickamer. 1993. Expression and characterization of a carbohydrate-binding fragment of rat aggrecan. *Glycobiology*. 3:185-190.

Salles, J.I., M.E.L. Duarte, J.M. Guimarães, L. Lopes, J.V. Cardoso, D.P. Aguiar, J.O.M. Neto, D.E. Machado, and J.A. Perini. 2016. Vascular Endothelial Growth Factor Receptor-2 Polymorphisms Have Protective Effect against the Development of Tendinopathy in Volleyball Athletes. *PLoS One*. 11:e0167717.

Saunders, C.J., L. van der Merwe, M. Posthumus, J. Cook, C.J. Handley, M. Collins, and A.V. September. 2013. Investigation of variants within the *COL27A1* and *TNC* genes and Achilles tendinopathy in two populations. *J Orthop Res*. 31:632-637.

Schierding, W., W.S. Cutfield, and J.M. O'Sullivan. 2014. The missing story behind Genome Wide Association Studies: single nucleotide polymorphisms in gene deserts have a story to tell. Front Genet. 5:39.

Schmidt, A.P., and S.R.G. Schmidt. 2016. Behavior of ion channels controlled by electric potential difference and of Toll-type receptors in neuropathic pain pathophysiology. *Revista Dor*. 177:13-45.

Schönherr E, Sunderkötter C, Iozzo RV, and S. L. 2005. Decorin, a novel player in the insulin-like growth factor system. *J Biol Chem*. 280:15767-15772.

Schultz, R.A., D.C. Miller, C.S. Kerr, and L. Micheli. 1984. Mechanoreceptors in human cruciate ligaments. A histological study. *J Bone Joint Surg Am*. 66:1072-1076.

Seale, K.B., M.C. Burger, M. Posthumus, C.K. Häger, E. Stattin, K.G. Nilsson, M. Collins, and A.V. September. 2020. The Apoptosis Pathway and *CASP8* Variants Conferring Risk for Acute and Overuse Musculoskeletal Injuries. *J Orthop Res*. 38:680-688.

Sebastiani, P., N. Solovieff, S.W. Hartley, J.N. Milton, A. Riva, D.A. Dworkis, E. Melista, E.S. Klings, M.E. Garrett, M.J. Telen, A. Ashley-Koch, C.T. Baldwin, and M.H. Steinberg. 2010. Genetic modifiers of the severity of sickle cell anemia identified through a genome-wide association study. *Am J Hematol*. 85:29-35.

September, A.V., J. Cook, C.J. Handley, L. Van der Merwe, M.P. Schwellnus, and M. Collins. 2009. Variants within the *COL5A1* gene are associated with Achilles tendinopathy in two populations. *Br J Sports Med*. 43:357-365.

September, A.V., E.-M. Nell, K. O'Connell, J. Cook, C.J. Handley, L. Van der Merwe, M. Schwellnus, and M. Collins. 2011. A pathway-based approach investigating the genes encoding interleukin-1β, interleukin-6 and the interleukin-1 receptor antagonist provides new insight into the genetic susceptibility of Achilles tendinopathy. *Br J Sport Med*. 45:1040-1047.

Shahbazi, M., A.A. Fryer, V. Pravica, I.J. Brogan, H.M. Ramsay, I.V. Hutchinson, and P.N. Harden. 2002. Vascular Endothelial Growth Factor Gene Polymorphisms Are Associated with Acute Renal Allograft Rejection. *J Am Soc Nephrol*. 13:260-264.

Shaw, L.M., and B.R. Olsen. 1991. FACIT collagens: diverse molecular bridges in extracellular matrices. *Trends Biochem Sci*. 16:191-194.

Shea, K.G., N. Archibald-Seiffer, E. Murdock, N.L. Grimm, J.C. Jacobs, W. Jr, S., and H. Van Houten. 2014. Knee Injuries in Downhill Skiers: A 6-Year Survey Study. *Orthop J Sports Med*. 2:2325967113519741.

Shelbourne, K.D., T.J. Davis, and T.E. Klootwyk. 1998. The relationship between intercondylar notch width of the femur and the incidence of anterior cruciate ligament tears: a prospective study. *Am J Sports Med*. 26:402-408.

Sherry, S.T., M.H. Ward, M. Kholodov, J. Baker, L. Phan, E.M. Smigielski, and K. Sirotkin. 2001. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res*. 29:308-311.

Shihab, H.A., J. Gough, M. Mort, D.N. Cooper, I.N.M. Day, and T.R. Gaunt. 2014. Ranking non-synonymous single nucleotide polymorphisms based on disease concepts. *Hum Genomics*. 8:11.

Shrive, N., D. Chimich, L. Marchuk, J. Wilson, R. Brant, and C. Frank. 1995. Soft-tissue "flaws" are associated with material properties of the healing rabbit medial collateral ligament. *J Orthop Res*. 13:923-929.

Shultz, S.J., R.J. Schmitz, A. Benjaminse, A.M. Chaudhari, M. Collins, and D.A. Padua. 2012. ACL Research Retreat VI: an update on ACL injury risk and prevention. *J Athl Train*. 47:591-603.

Sifrim, A., D. Popovic, L.C. Tranchevent, A. Ardeshirdavani, R. Sakai, P. Konings, J.R. Vermeesch, J. Aerts, B. De Moor, and Y. Moreau. 2013. eXtasy: variant prioritization by genomic data fusion. *Nat Methds*. 10:1083-1084.

Singh, A.P., and S. Rajender. 2015. CatSper channel, sperm function and male fertility. *Reprod Biomed Online*. 30:28-38.

Sinnwell, J., and D. Schaid. 2018. Haplo.stats: A package for statistical analysis of haplotypes with traits and covariates when linkage phase is ambiguous: R package (version 1.5.4).

Sivertsen, E.A., K. Haug, E.K. Kristianslund, A.S. Trøseid, J. Parkkari, T. Lehtimäki, N. Mononen, K. Pasanen, and R. Bahr. 2019. No Association Between Risk of Anterior Cruciate Ligament Rupture and Selected Candidate Collagen Gene Variants in Female Elite Athletes From High-Risk Team Sports. *Am J Sports Med*. 47:52-58.

Skutek, M., M. van Griensven, J. Zeichen, N. Brauer, and U. Bosch. 2001. Cyclic mechanical stretching enhances secretion of Interleukin 6 in human tendon fibroblasts. *Knee Surg Sports Traumatol Arthrosc*. 9:322-326.

Smith, H.C., P. Vacek, R.J. Johnson, J.R. Slauterbeck, J. Hashemi, S. Shultz, and B.D. Beynnon. 2012. Risk factors for anterior cruciate ligament injury: a review of the literature-part 2: hormonal, genetic, cognitive function, previous injury, and extrinsic risk factors. *Sports Health*. 4:155-161.

Söderman, K., H. Alfredson, T. Pietilä, and S. Werner. 2001. Risk factors for leg injuries in female soccer players: A prospective ivestigation during one out-door season. Knee Surgery, *Sport Traumatol Arthrosc*. 9:313-321.

Somerville, R.P.T., S.A. Oblander, and S.S. Apte. 2003. Matrix metalloproteinases: old dogs with new tricks. *Genome biol*. 4:216.

Souryal, T.O., and T.R. Freeman. 1993. Intercondylar notch size and anterior cruciate ligament injuries in athletes. *Am J Sports Med*. 21:535-9.

Southam, L., J. Rodriguez-Lopez, J.M. Wilkins, M. Pombo-Suarez, S. Snelling, J.J. Gomez-Reino, K. Chapman, A. Gonzalez, J. Loughlin. 2007. An SNP in the 5'-UTR of GDF5 is associated with osteoarthritis susceptibility in Europeans and with in vivo differences in allelic expression in articular cartilage. *Hum Mol Genet*. 16: 2226-2232.

Stattin, E., F. Wiklund, K. Lindblom, P. Önnerfjord, B. Jonsson, Y. Tegner, T. Sasaki, A. Struglics, S. Lohmander, N. Dahl, D. Heinegård, and A. Aspberg. 2010. A Missense Mutation in the Aggrecan C-type Lectin Domain Disrupts Extracellular Matrix Interactions and Causes Dominant Familial Osteochondritis Dissecans. *Am J Hum Genet*. 86:126-137.

Stelzer, G., R. Rosen, I. Plaschkes, S. Zimmerman, M. Twik, S. Fishilevich, T. Iny Stein, R. Nudel, I. Lieder, Y. Mazor, S. Kaplan, D. Dahary, D. Warshawsky, Y. Guan - Golan, A. Kohn, N. Rappaport, M. Safran, and D. Lancet. 2016. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analysis. *Curr Protoc Bioinformatics*. 54:1.30.31-31.30.33 www.genecards.org.

Stępień-Słodkowska, M., K. Ficek, M. Kaczmarczyk, A. Maciejewska, M. Sawczuk, J. Eider, M. Sygit, A. Leońska-Duniec, P. Ziętek, K. Sygit, and P. Cieszczyk. 2016. Influence of biological factors on injuries occurrence in the Polish population. *Ann Agric Environ Med*. 23:315-318.

Stępień-Słodkowska, M., K. Ficek, A. Maciejewska-Karłowska, M. Sawczuk, P. Ziętek, and P. Król. 2014. Overrepresentation of the *COL3A1* AA genotype in Polish skiers with anterior cruciate ligament injury. *Biol Sport*. 32:143-147.

Strocchi, R., V. de Pasquale, P. Gubellini, A. Facchini, M. Marcacci, R. Buda, S. Zaffagnini, and A. Ruggeri. 1992. The human anterior cruciate ligament: histological and ultrastructural observations. *J Anat*. 180:515-519.

Subhash, C.J., T. Chen, Y.M. Farhat, and C. Veillette. 2016. In Vitro Effect of Transforming Growth Factor-β1 (TGF-β1) on Gene Expression in Human Flexor Digitorum Profundus Tendon Cells. *Orthop Muscular Syst*. 5:3.

Suijkerbuijk, M.A.M., M. Ponzetti, M. Rahim, M. Posthumus, C.K. Häger, E. Stattin, K.G. Nilsson, A. Teti, D.E. Meuffels, B.J.C. van der Eerden, M. Collins, and A.V. September. 2019. Functional polymorphisms within the inflammatory pathway regulate expression of extracellular matrix components in a genetic risk dependent model for anterior cruciate ligament injuries. *J Sci Med Sport*. 22:1219-1225.

Sun, M., S. Chen, S.M. Adams, J.B. Florer, H. Liu, W.W. Kao, R.J. Wenstrup, and D.E. Birk. 2011. Collagen V is a dominant regulator of collagen fibrillogenesis: dysfunctional regulation of structure and function in a corneal-stroma-specific Col5a1-null mouse model. *J Cell Sci*. 124:4096-4105.

Sun, T., Y. Gao, W. Tan, S. Ma, Y. Shi, J. Yao, Y. Guo, M. Yang, X. Zhang, Q. Zhang, C. Zeng, and D. Lin. 2007. A six-nucleotide insertion-deletion polymorphism in the CASP8 promoter is associated with susceptibility to multiple cancers. *Nat Genet*. 39:605-613.

Sundemo, D., E. Hamrin Senorski, L. Karlsson, A. Horvath, B. Juul-Kristensen, J. Karlsson, O.R. Ayeni, and K. Samuelsson. 2019. Generalised joint hypermobility increases ACL injury risk and is associated with inferior outcome after ACL reconstruction: a systematic review. *BMJ Open Sport Exerc Med*. 5:e000620.

Taylor, C., and J.F. Meaney. "Heritability". Encyclopedia Britannica, 5 Apr. 2018, https://www.britannica.com/science/heritability. Accessed 22 July 2021.

Tashjian, R.Z., S.K. Kim, M.D. Roche, K.B. Jones, and C.C. Teerlink. 2021. Genetic variants associated with rotator cuff tearing utilizing multiple population-based genetic resources. *J Shoulder Elbow Surg*. 30:520-531.

Teitz, C.C., B.K. Lind, and B.M. Sacks. 1997. Symmetry of the femoral notch width index. *Am J Sports Med*. 25:687-690.

Testa, R., F. Olivieri, A.R. Bonfigli, C. Sirolla, M. Boemi, F. Marchegiani, M. Marra, S. Cenerelli, R. Antonicelli, A. Dolci, G. Paolisso, and C. Franceschi. 2006. Interleukin-6-174 G > C polymorphism affects the association between IL-6 plasma levels and insulin resistance in type 2 diabetic patients. *Diabetes Res Clin Pract*. 71:299-305.

Thomas, D.C., and J.S. Witte. 2002. Point: population stratification: a problem for case-control studies of candidate-gene associations? *Cancer Epidemiol Biomarkers Prev*. 11:505-512.

Tsuzaki, M., G. Guyton, W. Garret, J.M Archambault, W. Herzog, L. Almekinders, D. Bynum, X. Yang, and A.J. Banes. 2003. IL-1 beta induces COX2, MMP-1, -3 and -13, ADAMTS-4, IL1 beta and IL-6 in human tendon cells. *J Orthop Res*. 21:256-264.

Ungefroren, H., and N.B. Krull. 1996. Transcriptional regulation of the human biglycan gene. *J Biol Chem*. 271:15787-15795.

Valhmu, W.B., G.D. Palmer, P.A. Rivers, S. Ebara, J.F. Cheng, S. Fischer, and A. Ratcliffe. 1995. Structure of the human aggrecan gene: exon-intron organization and association with the protein domains. *Biochem J*. 309:535-542.

Vega, L., M. Mateyak, and V. Zakian. 2003. Getting to the end: telomerase access in yeast and humans. *Nat Rev Mol Cell Biol*. 4:948-959.

Videman, T., J. Saarela, and J. Kaprio. 2009. Associations of 25 structural, degradative, and inflammatory candidate genes with lumbar disc desiccation, bulging, and height narrowing. *Arthritis Rheum*. 60 470-481.

Visscher, P.M., B. McEvoy, and J. Yang. 2010. From Galton to GWAS: quantitative genetics of human height. *Genet Res*. 92:371-379.

Visscher, P.M., W.G. Hill, and N.R. Wray. 2008. Heritability in the genomics era--concepts and misconceptions. *Nat Rev Genet*. 9:255-266.

Voisin, S., P. Cieszczyk, V.P. Pushkarev, D.A. Dyatlov, B.F. Vashlyayev, V.A. Shumaylov, A. Maciejewska-Karlowska, M. Sawczuk, L. Skuza, Z. Jastrzebski, D.J. Bishop, and N. Eynon. 2014. EPAS1 gene variants are associated with sprint/power athletic performance in two cohorts of European athletes. *BMC genomics*. 15:382.

Waggett, A.D., J.R. Ralphs, A.P.L. Kwan, D. Woodnutt, and M. Benjamin. 1998. Characterization of collagens and proteoglycans at the insertion of the human achilles tendon. *Matrix Biol*. 16:457-470.

Wang, C., H. Li, K. Chen, B. Wu, and H. Liu. 2017. Association of polymorphisms rs1800012 in *COL1A1* with sports-related tendon and ligament injuries: a meta-analysis. *Oncotarget*. 8:27627-27634.

Wang, H., S.J. Shultz, S.E. Ross, R.A. Henson, D.H. Perrin, and R.J. Schmitz. 2020. ACL size and notch width between ACLR and healthy individuals: A pilot study. *Sports Health*. 12:61-65.

Wang, K., M. Li, and H. Hakonarson. 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 38:e164.

Wang, S., C. Jiang, and K. Zhang. 2019. Significantly dysregulated genes in osteoarthritic labrum cells identified through gene expression profiling. *Mol Med Rep*. 20:1716-1724.

Wang, Y., and J. Wang. 2020. Modelling and prediction of global non-communicable diseases. BMC public health. 20:822.

Wang, Y., Y. Zheng, W. Zhang, H. Yu, K. Lou, and Y. Zhang. 2007. Polymorphisms of *KDR* Gene Are Associated With Coronary Heart Disease. *J Am Coll Cardiol*. 50:760-767.

Warde-Farley, D., S.L. Donaldson, O. Comes, K. Zuberi, R. Badrawi, P. Chao, M. Franz, C. Grouios, F. Kazi, C.T. Lopes, A. Maitland, S. Mostafavi, J. Montojo, Q. Shao, G. Wright, G.D. Bader, and Q. Morris. 2010. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res*. 38:W214-W220.

Warnes, G., G. Gorjanc, F. Leisch, and M. Man. 2012. Genetics: A package for opulation genetics: R Package (version 1.3.6).

Watanabe, H., S.C. Cheung, N. Itano, K. Kimata, and Y. Yamada. 1997. Identification of hyaluronan-binding domains of aggrecan. *J Biol Chem*. 272:28057-28065.

Waxman, S.G., I.S.J. Merkies, M.M. Gerrits, S.D. Dib-Hajj, G. Lauria, J.J. Cox, J.N. Wood, C.G. Woods, J.P.H. Drenth, and C.G. Faber. 2014. Sodium channel genes in pain-related disorders: phenotype–genotype associations and recommendations for clinical use. *Lancet Neurol*. 13:1152-1160.

Weedon, M.N., H. Lango, C.M. Lindgren, C. Wallace, D.M. Evans, M. Mangino, R.M. Freathy, J.R. Perry, S. Stevens, A.S. Hall, N.J. Samani, B. Shields, I. Prokopenko, M. Farrall, A. Dominiczak, Diabetes Genetics Initiative, Wellcome Trust Case Control Consortium, T. Johnson, S. Bergmann, J.S. Beckmann, …Frayling, T. M. 2008. Genome-wide association analysis identifies 20 loci that influence adult height. *Nat Genet*. 40:575-583.

Wenstrup, R.J., S.M. Smith, J.B. Florer, G. Zhang, D.P. Beason, R.E. Seegmiller, L.J. Soslowsky, and D.E. Birk. 2011. Regulation of collagen fibril nucleation and initial fibril assembly involves coordinate interactions with collagens V and XI in developing tendon. *J Biol Chem*. 286:20455-20465.

Wetters, N., A.E. Weber, T.H. Wuerz, D.L. Schub, and B.R. Mandelbaum. 2016. Mechanism of Injury and Risk Factors for Anterior Cruciate Ligament Injury. *Oper Tech Sports Med*. 24:2-6.

Wheeler, D.L., T. Barrett, D.A. Benson, S.H. Bryant, K. Canese, V. Chetvernin, D.M. Church, M. DiCuccio, R. Edgar, S. Federhen, L.Y. Geer, Y. Kapustin, O. Khovayko, D. Landsman, D.J. Lipman, T.L. Madden, D.R. Maglott, J. Ostell, V. Miller, K.D. Pruitt, G.D. Schuler, E. Sequeira, S.T. Sherry, K. Sirotkin, A. Souvorov, G. Starchenko, R.L. Tatusov, T.A. Tatusova, L. Wagner, and E. Yaschenko. 2007. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*. 35:D5-D12.

Wight, T.N., M.G. Kinsella, S.P. Evanko, S. Potter-Perigo, and M.J. Merrilees. 2014. Versican and the regulation of cell phenotype in disease. *Biochim Biophys Acta*. 1840:2441-2451.

Willard, K., M.-J.N. Laguette, L.A. de Souza Rios, C. D'Alton, M. Nel, S. Prince, M. Collins, and A.V. September. 2020. Altered expression of proteoglycan, collagen and growth factor genesin a TGF-β1 stimulated genetic risk model for musculoskeletal softtissue injuries. *J Sci Med Sport*. 23:695-700.

Wonkam, A., E.R. Chimusa, K. Mnika, G.D. Pule, V.J. Ngo Bitoungui, N. Mulder, D. Shriner, C.N. Rotimi, and A. Adeyemo. 2020. Genetic modifiers of long-term survival in sickle cell anemia. *Clin Transl Med*. 10:e152.

Wordeman, S.C., C.E. Quatman, C.C. Kaeding, and T.E. Hewett. 2012. In vivo evidence for tibial plateau slope as a risk factor for anterior cruciate ligament injury: A systematic review and meta-analysis. *Am J Sports*. 40:1673-81.

Wu, J.J., M.A. Weis, L.S. Kim, and D.R. Eyre. 2010. Type III collagen, a fibril network modifier in articular cartilage. *J Biol Chem*. 285:18537-18544.

Yang, G., H.-J. Im, and J.H.C. Wang. 2005. Repetitive mechanical stretching modulates IL-1 induced COX-2, MMP-1 expression, and PGE2 production in human patellartendon fibroblasts. *Gene*. 363:166-172.

Yang, J., Z. Hu, Y. Xu, J. Shen, J. Niu, X. Hu, J. Guo, Q. Wei, X. Wang, and H. Shen. 2004. Interleukin-1B gene promoter variants are associated with an increased risk of gastric cancer in a Chinese population. *Cancer Lett*. 215:191-198.

Yates, A.D., P. Achuthan, W. Akanni, J. Allen, J. Allen, J. Alvarez-Jarreta, M.R. Amode, I.M. Armean, A.G. Azov, R. Bennett, J. Bhai, K. Billis, S. Boddu, J.C. Marugán, C. Cummins, C. Davidson, K. Dodiya, R. Fatima, A. Gall, C.G. Giron, L. Gil, T. Grego, L. Haggerty, E. Haskell, T. Hourlier, O.G. Izuogu, S.H. Janacek, T. Juettemann, M. Kay, I. Lavidas, T. Le, D. Lemos,

J.G. Martinez, T. Maurel, M. McDowall, A. McMahon, S. Mohanan, B. Moore, M. Nuhn, D.N. Oheh, A. Parker, A. Parton, M. Patricio, M.P. Sakthivel, A.I. Abdul Salam, B.M. Schmitt, H. Schuilenburg, D. Sheppard, M. Sycheva, M. Szuba, K. Taylor, A. Thormann, G. Threadgold, A. Vullo, B. Walts, A. Winterbottom, A. Zadissa, M. Chakiachvili, B. Flint, A. Frankish, S.E. Hunt, G. IIsley, M. Kostadima, N. Langridge, J.E. Loveland, F.J. Martin, J. Morales, J.M. Mudge, M. Muffato, E. Perry, M. Ruffier, S.J. Trevanion, F. Cunningham, K.L. Howe, D.R. Zerbino, and P. Flicek. 2019. Ensembl 2020. *Nucleic Acids Res*. 48:D682-D688.

Yip, S.P., K.H. Leung, P.W. Ng, W.Y. Fung, P.C. Sham, and M.K. Yap. 2011. Evaluation of proteoglycan gene polymorphisms as risk factors in the genetic susceptibility to high myopia. *Invest Ophthalmol Vis Sci*. 52:6396-6403.

Yoshikawa, T., H. Tohyama, T. Katsura, E. Kondo, Y. Kotani, H. Matsumoto, Y. Toyama, and K. Yasuda. 2006. Effects of local administration of vascular endothelial growth factor on mechanical characteristics of the semitendinosus tendon graft after anterior cruciate ligament reconstruction in sheep. *Am J Sports Med*. 34:1918-1925.

Young, B.B., G. Zhang, M. Koch, and D.E. Birk. 2002. The roles of types XII and XIV collagen in fibrillogenesis and matrix assembly in the developing cornea. *J Cell Biochem*. 87:208-220.

Young, K., T. Samiric, J. Feller, and J. Cook. 2011. Extracellular matrix content of ruptured anterior cruciate ligament tissue. *Knee*. 18:242-246.

Yu, B., and W.E. Garrett. 2007. Mechanisms of non-contact ACL injuries. *Br J Sport Med*. Suppl 1:i47–51.

Yung, P.S., Y.M. Lee, S.C. Fu, C.H. Chen, C.G. Rolf, and K.M. Chan. 2020. Differential MMP 1 and MMP 13 expression in proliferation and ligamentization phases of graft remodeling in anterior cruciate ligament reconstruction. Connect Tissue Res. 1-8.

Zantop, T., W. Petersen, J.K. Sekiya, V. Musahl, and F.H. Fu. 2006. Anterior cruciate ligament anatomy and function relating to anatomical reconstruction. *Knee Surg Sports Traumatol Arthrosc*. 14:982-992.

Zhang, G., S. Chen, S. Goldoni, B.W. Calder, H.C. Simpson, R.T. Owens, D.J. McQuillan, M.F. Young, R.V. Iozzo, and D.E. Birk. 2009. Genetic evidence for the coordinated regulation of collagen fibrillogenesis in the cornea by decorin and biglycan. *J Biol Chem*. 284:8888-97.

Zhang, Y., M. Malekpour, N. Al-Madani, K. Kahrizi, M. Zanganeh, N.J. Lohr, M. Mohseni, F. Mojahedi, A. Daneshi, H. Najmabadi, and R.J. Smith. 2007. Sensorineural deafness and male infertility: a contiguous gene deletion syndrome. *J Med Genet*. 44:233-240.

Zhao, D., Q. Zhang, Q. Lu, C. Hong, T. Luo, Q. Duan, S. Shu, J. Lv, and W. Zhao. 2020. Correlations Between the Genetic Variations in the *COL1A1, COL5A1, COL12A1*, and β-fibrinogen Genes and Anterior Cruciate Ligament Injury in Chinese Patientsa. *J Athl Train*. 55:515-521.

Zhao, J., M. Yang, and M. Li. 2016. Association of *IL-6* and *MMP-3* gene polymorphisms with susceptibility to adolescent idiopathic scoliosis: A meta-analysis. *J Genet*. 95:573-579.

Zhou, M., H. Hu, Y. Han, J. Li, Y. Zhang, S. Tang, Y. Yuan, and X. Zhang. 2021. Long non-coding RNA 01126 promotes periodontitis pathogenesis of human periodontal ligament cells via miR-518a-5p/HIF-1α/MAPK pathway. *Cell Prolif*. 54:e12957.

Zou, Y., D. Zwolanek, Y. Izu, S. Gandhy, G. Schreiber, K. Brockmann, M. Devoto, Z. Tian, Y. Hu, G. Veit, M. Meier, J. Stetefeld, D. Hicks, V. Straub, N.C. Voermans, D.E. Birk, E.R. Barton, M. Koch, and C.G. Bönnemann. 2014. Recessive and dominant mutations in *COL12A1* cause a novel EDS/myopathy overlap syndrome in humans and mice. *Hum Mol Genet*. 23:2339-2352.

APPENDIX A

SECTION 1

ETHICAL APPROVAL

UNIVERSITY OF CAPE TOWN
Faculty of Health Sciences
Human Research Ethics Committee

Room E52-24 Old Main Building
Groote Schuur Hospital
Observatory 7925
Telephone [021] 406 6338 • **Facsimile** [021] 406 6411
Email: shuretta.thomas@uct.ac.za
Website: http://www.health.uct.ac.za/fhs/research/humanethics/forms

30 May 2014

HREC REF: 269/2014

Dr M Posthumus
Human Biology
Sport Science Institute

Dear Dr Posthumus

PROJECT TITLE: ***THE IDENTIFICATION OF GENETIC SUSCEPTIBILITY- A 20 YEAR FOLLOW UP OF PATIENTS WITH ACL INJURY***

Thank you for your response to the Faculty of Health Sciences Human Research Ethics Committee dated 26 May 2014.

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

Approval is granted for one year until the 30th June 2015

Please submit a progress form, using the standardised Annual Report Form if the study continues beyond the approval period. Please submit a Standard Closure form if the study is completed within the approval period.
(Forms can be found on our website: www.health.uct.ac.za/fhs/research/humanethics/forms)

We acknowledge that the student, Lee Hill is also involved in this study.

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

Please quote the HREC reference no in all your correspondence.

Yours sincerely

PROFESSOR M BLOCKMAN
CHAIRPERSON, FHS HUMAN ETHICS
Federal Wide Assurance Number: FWA00001637.
Institutional Review Board (IRB) number: IRB00001938
This serves to confirm that the University of Cape Town Human Research Ethics Committee complies to the Ethics Standards for Clinical Research with a new drug in patients, based on the Medical Research Council (MRC-SA), Food and Drug Administration (FDA-USA), International Convention on Harmonisation Good Clinical Practice (ICH GCP) and Declaration of Helsinki guidelines.

UNIVERSITY OF CAPE TOWN
Faculty of Health Sciences
Human Research Ethics Committee

Room E53-46 Old Main Building
Groote Schuur Hospital
Observatory 7925
Telephone [021] 406 6492
Email: sumayah.arlefdien@uct.ac.za
Website: www.health.uct.ac.za/fhs/research/humanethics/forms

13 February 2018

HREC REF: 110/2018

A/Prof A September
Department of Human Biology
Division of ESSM
Newlands

Dear A/Prof September

PROJECT TITLE: THE IDENTIFICATION OF GENETIC RISK FACTORS UNDERLYING ANTERIOR CRUCIATE LIGAMENT INJURIES IN A POLISH POPULATION

Thank you for submitting your study to the Faculty of Health Sciences Human Research Ethics Committee (HREC) for review.

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

Approval is granted for one year until the 28 February 2019.

Please submit a progress form, using the standardised Annual Report Form if the study continues beyond the approval period. Please submit a Standard Closure form if the study is completed within the approval period.
(Forms can be found on our website: www.health.uct.ac.za/fhs/research/humanethics/forms)

Please quote the HREC REF in all your correspondence.

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

Please note that for all studies approved by the HREC, the principal investigator **must** obtain appropriate institutional approval, where necessary, before the research may occur.

Yours sincerely

PROFESSOR M BLOCKMAN
CHAIRPERSON, FHS HUMAN RESEARCH ETHICS COMMITTEE

Federal Wide Assurance Number: FWA00001637.
Institutional Review Board (IRB) number: IRB00001938
This serves to confirm that the University of Cape Town Human Research Ethics Committee complies to the Ethics Standards for Clinical Research with a new drug in patients, based on the Medical

PO Royal Melbourne Hospital
Parkville Victoria 3050
Telephone 61 3 9342 8530
Facsimile 61 3 9342 8548
Email: research@mh.org.au
Website: http://research.mh.org.au
ABN 73 802 706 972

MELBOURNE HEALTH

OFFICE FOR RESEARCH

MELBOURNE HEALTH HUMAN RESEARCH ETHICS COMMITTEE

ETHICAL APPROVAL OF A RESEARCH PROJECT

Dr Phong Tran
Orthopaedic Department
Western Hospital
Gordon Street
FOOTSCRAY VIC 3011

19th February 2014

Dear Dr Tran,

MH Project Number: 2013.223

Project Title: The relationship between genetics functional movement patterns and anterior cruciate ligament injury

HREC Approval Date: 19th February 2014

I am pleased to advise that the above project has received ethical approval.

Participating Sites:

- Western Health

Approved Documents:

- Project Protocol Version 2 dated September 2013
- Western Health Participant Information and Consent Form – ACL Genetic Study dated September 2013
- Western Health Participant Information and Consent Form – Control Genetic Study dated September 2013
- Genetic Basis of Ligament Injury Questionnaire
- Telephone Transcript for ACL Research Study
- Flyer for Advertisement

Site Specific Assessment:

Please note: Please forward this HREC approval certificate to the Director of Research at Western Health together with your Research Governance-Site Specific Assessment application. You cannot commence this study until you have completed all the requirements of the Site Specific Assessment and have received written approval to conduct your research project at Western Health.

Conditions of Ethics Approval:

In order to comply with the National Statement on Ethical Conduct in Human Research 2007, Guidelines for Good Clinical Research Practice and Melbourne Health Research Policies and Guidelines you are required to:

- Submit a copy of this letter to the Radiation Safety Officer (RSO) at Western Health, for addition of the project to the Licence for Research Involving Human Volunteers held by the Department of Human Services Radiation Safety Section Radiation Safety Licence (if your project involves exposure to ionising radiation). Note: You cannot commence the project until you have received notification from the RSO that the project has been added to the Licence;
- Notify the HREC of the actual start date of the project;
- *Submit to the HREC for approval any proposed amendments to the project including any proposed changes to the Protocol, Participant Information and Consent Form/s and the Investigator Brochure;*
- *Notify the HREC of any adverse events in accordance with the Melbourne Health Guidelines for Monitoring and Reporting of Safety in Clinical Trials Involving Therapeutic Products and Other Clinical Research, July 2009;*
- *Notify the HREC of any unforseen events;*
- *Notify the HREC of your inability to continue as Principal Investigator or any other change in research personnel involved in the project;*
- *Notify the HREC if a decision is taken to end the study prior to the expected date of completion or failure to commence the study within 12 months of the HREC approval date;*
- *Notify the HREC of any other matters which may impact the conduct of the project.*

Reporting

You are required to submit to the HREC:

- *An Annual Progress Report every 12 months (or more frequently as requested by the reviewing HREC) for the duration of the project. This report is due on the anniversary of HREC approval. Continuation of ethics approval is contingent on submission of an annual report in a timely manner; and*
- *A comprehensive Final Report upon completion of the project.*

The HREC may conduct an audit of the project at any time.

Please refer to the Office for Research website to access forms such as the Amendment Form, Annual Report/Final Report Form, Guidelines for Monitoring and Reporting of Safety in Clinical Trials Guidelines and Adverse Event Report Forms, and other information and news concerning research at Melbourne Health:
http://www.mh.org.au/www/342/1001127/displayarticle/1001352.html

A list of those HREC members present at the review of this project can be obtained from the above website.

Yours sincerely

Ms. Jessica Turner
Manager - Human Research Ethics Committee

UNIVERSITY OF CAPE TOWN
Faculty of Health Sciences
Human Research Ethics Committee

Room E52-24 Old Main Building
Groote Schuur Hospital
Observatory 7925
Telephone [021] 406 **6626**
Email: shuretta.thomas@uct.ac.za
Website: www.health.uct.ac.za/fhs/research/humanethics/forms

18 May 2016

HREC REF: 622/2015

A/Prof A September
Sport Science Institute
Human Biology

Dear A/Prof September

PROJECT TITLE: ***THE IDENTIFICATION OF GENETIC RISK FACTORS UNDERLYING ANTERIOR CRUCIATE LIGAMENT INJURIES IN AN AUSTRALIAN POPULATION***

Thank you for your response to the Faculty of Health Sciences Human Research Ethics Committee dated 21 April 2016.

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

Approval is granted for one year until the 30th May 2017.

Please submit a progress form, using the standardised Annual Report Form if the study continues beyond the approval period. Please submit a Standard Closure form if the study is completed within the approval period.
(Forms can be found on our website: www.health.uct.ac.za/fhs/research/humanethics/forms)

Please quote the HREC REF in all your correspondence.

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

Please note that for all studies approved by the HREC, the principal investigator **must** obtain appropriate institutional approval before the research may occur.

Yours sincerely

PROFESSOR M BLOCKMAN
CHAIRPERSON, FHS HUMAN RESEARCH ETHICS COMMITTEE
Federal Wide Assurance Number: FWA00001637.
Institutional Review Board (IRB) number: IRB00001938
This serves to confirm that the University of Cape Town Human Research Ethics Committee complies to the Ethics Standards for Clinical Research with a new drug in patients, based on the Medical Research Council (MRC-SA), Food and Drug Administration (FDA-USA), International Convention on Harmonisation Good Clinical Practice (ICH GCP), South African Good Clinical Practice Guidelines (DoH

UNIVERSITY OF CAPE TOWN
Faculty of Health Sciences
Human Research Ethics Committee

Room E53-46 Old Main Building
Groote Schuur Hospital
Observatory 7925
Telephone [021] 406 6626
Email: shuretta.thomas@uct.ac.za
Website: www.health.uct.ac.za/fhs/research/humanethics/forms

04 December 2017

HREC REF: 823/2017

A/Prof Alison September
Human Biology
Sport Science Institute

Dear A/Prof September

PROJECT TITLE: WHOLE GENOME SEQUENCING APPROACH TO IDENTIFYING GENETIC RISK FACTORS UNDERLYING ANTERIOR CRUCIATE LIGAMENT INJURIES IN A TWIN FAMILY STUDY (PhD-candidate- Ms D Feldmann)

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

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

Approval is granted for one year until the 30 December 2018.

Please submit a progress form, using the standardised Annual Report Form if the study continues beyond the approval period. Please submit a Standard Closure form if the study is completed within the approval period.
(Forms can be found on our website: www.health.uct.ac.za/fhs/research/humanethics/forms)

Please quote the HREC REF in all your correspondence.

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

Please note that for all studies approved by the HREC, the principal investigator **must** obtain appropriate institutional approval, where necessary, before the research may occur.

The HREC acknowledge that the student, Daneil Feldmann will also be involved in this study.

Yours sincerely

PROFESSOR M BLOCKMAN
CHAIRPERSON, FHS HUMAN RESEARCH ETHICS COMMITTEE
Federal Wide Assurance Number: FWA00001637.
Institutional Review Board (IRB) number: IRB00001938

SECTION 2

PARTICIPANT QUESTIONAIRES AND INFORMED CONSENT

Department of Human Biology

Faculty of Health Sciences
University of Cape Town
Private Bag X3
Observatory 7935
South Africa
Phone: +27-21-406 6235
Fax: +27-21-448 7226

THE GENETIC BASIS OF EXERCISE-INDUCED ANTERIOR CRUCIATE LIGAMENT INJURY

APPENDIX 1B: PARTICIPANT INFORMATION SHEET

Although there is a high incidence of anterior cruciate ligament (ACL) injuries as a result of participation in exercise and sporting activities, the mechanism and cause of this injury is still poorly understood. There is scientific evidence to suggest that there is a genetic component to exercise-induced ACL injuries. In an attempt to determine whether this is true, we at the Division of Exercise Science and Sports Medicine within the Department of Human Biology of the University of Cape Town, are interested in studying whether certain genes are associated with ACL injuries.

In order to accomplish this, we need to analyse the genetic material, called the DNA. The DNA in our body makes all our proteins and you have protein-coding DNA (exons) and non-protein coding DNA (introns). This project will screen both the exon and intron regions of the genes, by using a whole genome sequencing study (WGS) design which basically gives us the sequence of your entire DNA.

Aims of the research:

1. Analyse the entire DNA sequence to determine if there are any variations which could contribute to ACL injuries.
2. To assist physicians in developing multifactorial models to identify individuals at an increased risk of ACL injury.

Your possible involvement:

Should you agree to participate, you would be asked to do the following:

- Complete a questionnaire relating to your personal particulars, sporting details, and medical history. This information will be anonymous and only a coding system will be used to identify you.
- Donate a 5ml (1 teaspoon) blood sample from a vein in your arm. This will be used for the extraction and analysis of genetic material (DNA).

The DNA will only be used for scientific research purposes. We will perform analyses on your entire genome by sequencing your DNA.

To ensure complete confidentiality of your specific genetic information, the following procedures will be adopted: 1) all the blood samples will be labelled on collection using a numerical coding system that is linked to your details on a master list that will be placed in a sealed envelope, 2) this sealed master list will then be kept in a secure facility and in a separate location, 3) only the principle investigator and co-investigators will have access to this master list, 4) the master list will only be opened if a sample needs to be destroyed, should a participant request this. All data will be analysed anonymously and DNA samples will be stored or destroyed as indicated by you on completion of the study. Please be aware that you are free to request that your DNA sample be destroyed before the completion of the study, and furthermore you may withdraw, with no penalties or consequences, at any point during the study.

Furthermore, due to ethical reasons, we cannot disclose any individual genetic results from this study, however a summary of the overall findings will be communicated to all the participants on completion of the research.

Potential Risks of the study:

- The completion of a questionnaire or a physical examination is not associated with any risk. Questionnaire and other clinical data (paper and electronic) will be kept confidential and secure, and will not be made available to any party other than the research team without the consent of the individual participants.
- The potential risks to participants of blood collection are minimal and are related to 1) blood sample collection technique, and 2) the volume of blood. The potential risks associated with blood collection technique from the ante-cubital veins are: infection, delayed healing, haematoma, physical pain, mental discomfort and injury to a nerve or a vessel. These risks are small and will be minimized by the use of trained phlebotomists, use of sterile techniques and the use of disposable, single-use materials.
- Your personal and genetic information will be kept secure, anonymous and will only be used for research

Potential Benefits of the study:

The research questions that will be addressed by this study have been identified to have a direct impact on improving the understanding of ACL injuries. This could impact on the diagnosis, medical treatment and/or physical activity modifications. The anticipated benefits of this study are that the results will further our understanding of the possible cause/s of ACL injuries and the genetic alterations linked to the injury.

We look forward to working with you, and are most appreciative of your contribution to this novel medical research. If you have any questions or concerns regarding any aspect of the study, please do not hesitate to contact us at:

A/Prof. Alison V September, PhD
(021) 650 4559
alison.september@uct.ac.za

Dr Masouda Rahim, PhD
(+27) 82 666 6269
masouda.rahim@gmail.com

Miss Daneil Feldmann, PhD candidate
(+27) 76 416 3026
daneilfeldmann@gmail.com

Department of Human Biology

Faculty of Health Sciences
University of Cape Town
Private Bag X3
Observatory 7935
South Africa
Phone: +27-21-406 6235
Fax: +27-21-448 7226

GENETIC BASIS OF ANTERIOR CRUCIATE LIGAMENT INJURY

Appendix 2a: QUESTIONNAIRE

A. PERSONAL DETAILS			
Date of birth		Age	
Height (cm)		Gender	Male ☐ Female ☐
Weight (kg)	Pre-Injury:	Current:	
Ethnic group (Only required and used for the genetic aspect of the research)	Black/African ☐ White ☐ Indian ☐ Mixed Ancestry (Coloured) ☐ Asian ☐ Other ☐		
Ancestry: Tribal or national background (eg Xhosa, Dutch, Zulu, German, Italian)	Father		Unknown ☐
	Mother		Unknown ☐
Country of Birth			
Dominant Hand	Left ☐ Right ☐ Ambi ☐	Dominant Leg	Left ☐ Right ☐ Ambi ☐
Smoker	Yes (Current) ☐	Yes (Ex smoker) ☐	No, never ☐
	If yes, Number of years _____	If stopped, when _____	
	If yes, number per day _____		

B. OCCUPATIONAL DETAILS	
What is your current occupation?	
What was your occupation prior to injuring your ligament?	
Prior to injury, did your occupation involve lower limb activity?	□ Yes □ No
If yes please indicate which legs.	Right leg □ Both legs □ Left leg □ None □

(If you participate or have participated in more than 6 sports, please complete additional Sporting Details Questionnaires, Part B)

C. SPORTING DETAILS **Please record your sporting activities in order of importance**			
Type of sport(s) you have participated in (please name)	**Main sport 1**	**Other sport 2**	**Other sport 3**
Current or past participation	Current □ Past □	Current □ Past □	Current □ Past □
Year started participation			
Number of years involved in the sport			
Position played prior to injury (if appropriate)			
Playing level prior to injury (if appropriate)			
Number of years played prior to the injury.			

Type of sport(s) you have participated in (please name)	**Other sport 4**	**Other sport 5**	**Other sport 6**
Current or past participation	Current □ Past □	Current □ Past □	Current □ Past □

Year started participation			
Number of years involved in the sport			
Position played prior to injury (if appropriate)			
Playing level prior to injury (if appropriate)			
Number of years played prior to the injury.			

D. ANTERIOR CRUCIATE LIGAMENT INJURY DETAILS	
Date of ACL injury?	
Which side was injured?	□ Left □ Right □ Both
To what extent was your ligament ruptured?	□ Complete □ Partial □ None □ Unknown
Investigation done to confirm the diagnosis	□ MRI □ Surgery
How bad is your pain today? (mark line: e.g.l----------**I**----------l)	l----------l No pain Pain as bad as it can be

How was the ACL ruptured?	□ Direct impact (directly to the injured knee) □ Twisting and bending with indirect contact (i.e. contact elsewhere on the body) □ Twisting and bending without contact (no external contact) □ Skiing □ Other (please specify)……
Please describe the exactly how the injury occurred (If uncertain please state that you do not know)	…… …… ……
What was the initial treatment? (You may tick more than one block.)	□ Ice application □ Compression □ Immobilisation □ Medication □ Other…….
What was the final treatment?	□ Surgery □ Rehabilitation □ Other……..
What are your current symptoms? (You may tick more than one block.)	□ Pain □ Swelling □ Instability □ Weakness □ Other……

What is your current sports participation?	□ None □ Limited to non-weight bearing exercise □ Limited, not to same level as pre-injury □ Full participation
If you are able to recall, what were the weather and pitch conditions like at the time of injury?	□ Wet and soft ground □ Dry, but soft ground □ Dry and firm ground □ Wet, but firm ground □ Other……..
Associated injuries?	□ Meniscal tear □ MCL tear □ Other ligament tear □ Bone bruising □ Other……..

E. HISTORY OF OTHER LIGAMENT AND TENDON INJURIES IN THE PAST						
Have you ever injured a ligament in the past?	Yes □			No □		
If yes, please specify which ligaments? (You may tick more than one block, please select either L (left) or R (right))		L	R		L	R
	Knee (ACL)			Wrist ligaments		
	Knee (MCL)			Finger ligaments		
	Ankle lateral ligaments			Knee (PCL)		
	Spinal ligaments			Knee (LCL)		
	Shoulder ligaments			Ankle medial ligaments		

	Elbow ligaments			Other ligaments		

To your knowledge, have any other members of your family suffered from any ligament injury?	Yes □ No □	If Yes, please specify the family member □ Mother □ Father □ Sibling □ Son / daughter □ Other family member…….. and condition: Please choose ligament injury from the list above ……

Have you ever injured a tendon in the past?	Yes □		No □	
If yes, please specify which tendon? (You may tick more than one block, please select either L (left) or R (right)			**L**	**R**
	Foot and ankle	Achilles tendon		
		Tibialis posterior		
		Plantar fascia		
	Knee	Patellar tendon		
	Elbow and wrist	Wrist extensor tendons		
	Shoulder	Subscapularis		
		Supraspinatus		
		Infraspinatus		
		Teres minor		
	Other			

To your knowledge, have any other members of your family suffered from any tendon pathology?	Yes □ No □	If Yes, please specify the family member □ Mother □ Father □ Sibling □ Son / daughter □ Other family member:…….. Condition: Please choose tendon injury from the list above ……

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

F. MEDICAL HISTORY		
Do you currently suffer from any of these medical conditions:		
□ High Blood Pressure □ Emphysema □ Malignant disease (cancer) If Yes, what type? _____	□ Angina/Heart Attack □ Rheumatoid arthritis □ Elevated Blood Cholesterol □ Diabetes mellitus □ Renal disease	□ Asthma □ Osteoarthritis (wear & tear) □ Adrenal disorders □ Thyroid disorders □ Amyloidosis

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

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

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

If female:			
At what age did you start menstruating? (years)			
Are you currently using any type of contraception?	□ Yes □ No		
If Yes, what type of contraception are you using?	□ Pill □ Injection □ IUD		
Are you currently?	□ Pre-menopausal (±12 cycles per year at intervals of 23– 33 days & bleeding lasts 3-7 days) □ Menopausal (cycles are irregular and less frequent) □ Post-menopausal (no longer menstruating)		
Family History			
Do any other members of your family suffer from elevated blood cholesterol?	Yes □ No □		If Yes, which relative? □ Mother □ Father □ Sibling □ Son / daughter □ Other relative:……
Is there any history of arthritis in your family?	Yes □ No □		If Yes, which relative? □ Mother □ Father □ Sibling □ Son / daughter □ Other relative:…….. & What type of arthritis? Rheumatoid □ Osteoarthritis □ Other □
Drug and Allergy History	If yes, how long ago (or how many times, where applicable) did you use the medication?		
Have you ever used oral corticosteroids (cortisone tablets)?	Yes ☐	☐ 3 months	☐ 6 months
	No ☐	☐ 12 months	☐ 24 or more months
	Yes ☐	☐ 3 months	☐ 6 months

Have you ever been given an injection with corticosteroids?	No ☐	☐ 12 months	☐ 24 or more months
Have you ever been given an injection of corticosteroids in or around a tendon?	Yes ☐	☐ Once	☐ Twice
	No ☐	☐ 3 times	☐ >3 times
Have you ever used anabolic steroids?	Yes ☐	☐ 3 months	☐ 6 months
	No ☐	☐ 12 months	☐ 24 months
Have you ever used fluoroquinoline antibiotics?	Yes ☐	☐ 3 months	☐ 12 months
	No ☐	☐ 6 months	☐ 24 or more months
	If yes, please select from the list below:		
☐ ADCO-CIPRIN ☐ AVELON ☐ BACTIDRON ☐ CIFLOC ☐ CIFRAN ☐ CIPLA-CIPROFLOXACIN ☐ CIPLOXX ☐ CIPRO-HEXAL ☐ Other	☐ CIPROBAY ☐ CIPROGEN ☐ CPL ALLIANCE CIPROFLOXACIN ☐ DYNAFLOC ☐ FLOXIN ☐ MAXAQUIN ☐ NOROXIN ☐ ORPIC	☐ SANDOZ CIPROFLOXACIN ☐ TAFLOC ☐ TARIVID ☐ TAVANIC ☐ TEQUIN ☐ UNIQUIN ☐ UTN-400 ☐ ZANOCIN	
What medication, if any, are you currently using? (please list)			
What allergies do you have? (please list)			

Faculty of Health Sciences
University of Cape Town
Private Bag X3
Observatory 7935
South Africa
Phone: +27-21-406 6235
Fax: +27-21-448 7226

THE GENETIC BASIS OF EXERCISE-INDUCED ANTERIOR CRUCIATE LIGAMENT INJURY

APPENDIX 1A: INFORMED CONSENT

I, the undersigned, have been fully informed about the research study on the possible genetic contribution to anterior cruciate ligament injuries, which will be carried out by investigators within the Division of Exercise Science and Sports Medicine, within the department of Human Biology at the University of Cape Town. I am aware that this project will investigate the entire genome (protein coding and non-coding regions) by using a whole genome sequencing platform, which will provide the researches with an individual's entire DNA sequence.

I have agreed to donate five millilitres of venous blood (approximately a teaspoon) which will be used for the extraction and analysis of genetic material (DNA). I have also agreed to complete personal particulars, sporting participation, and medical history questionnaires and understand that all the information that is collected during the study will be treated with the strictest confidentiality, and will only be used for scientific research purposes. I also understand that my name and personal particulars will be not released under any circumstances, and that all data will be analysed anonymously. I agree to participate in the study voluntarily, and I have been informed that I will be free to withdraw from the study with no penalty at any time, if I so wish. I understand that my DNA sample will be (please select one):

□ **STORED:** I agree that my blood or tissue sample can be stored **until the completion of the project** but I can choose to request at any time that my stored sample be destroyed. My sample will be identified with a special study code that will remain linked to my name and contact details.

OR

□ **DESTROYED:** Please destroy my blood or tissue sample as soon as the current research project has been completed.

- I understand that my DNA will be screened (analysed) for genetic variations relating to anterior cruciate ligament injuries.
- I understand that withdrawing from research participation at any time will not incur any penalties.
- I understand that I will not receive any individual results, but the overall general outcomes of the study will be disclosed once the study is completed.
- I am aware that there is no direct benefit to myself, however if a genetic predisposition for anterior cruciate ligament injuries can be established, then future generations will be able to establish their risk for this condition. This may allow better prevention and treatment options in the future.
- I have read (or where appropriate, have had read to me) and understand the information about this study, and any questions I have asked have been answered to my satisfaction.
- I agree to participate in the study, realising that I have the right to request that my DNA sample be destroyed at any time.
- I may be contacted by any of the clinicians and researchers involved on this project in the event that additional information is required.
- I agree that research data provided by me or with my permission during the project may be included in a thesis, presented at conferences and published in journals on the condition that either my name not any other identifying information is used.

What if I get injured?

This research study is covered by an insurance policy taken out by the University of Cape Town if you suffer any injury while taking part in the study. The insurer will pay for all reasonable medical costs required to treat your bodily injury, according to the SA Good Clinical Practice Guidelines 2006 (or

latest version), which are based on the Association of the British Pharmaceutical Industry Guidelines. The insurer will pay without you having to prove that the research was responsible for your bodily injury. You may ask the study doctor for a copy of these guidelines.

The insurer will not pay for harm if, during the study, you:

- Use medicines or other substances that are not allowed
- Do not follow the study doctor's instructions
- Do not tell the study doctor that you have a bad side effect from the study medicine
- Do not take reasonable care of yourself and your study medicine

If you are harmed and the insurer pays for the necessary medical costs, usually you will be asked to accept that insurance payment as full settlement of the claim for medical costs. However, accepting this offer of insurance cover does not mean you give up your right to make a separate claim for other losses based on negligence, in a South African court.

It is important to follow the study doctor's instructions and to report straightaway if you have a side effect from the study medicine.

References

Department of Health. Ethics in Health Research: Principles, Processes and Structures. Second Edition. Department of Health, Republic of South Africa, Pretoria, 2015.

FULL NAME OF PARTICIPANT: _____

PARTICIPANT'S SIGNATURE: _____

DATE: _____

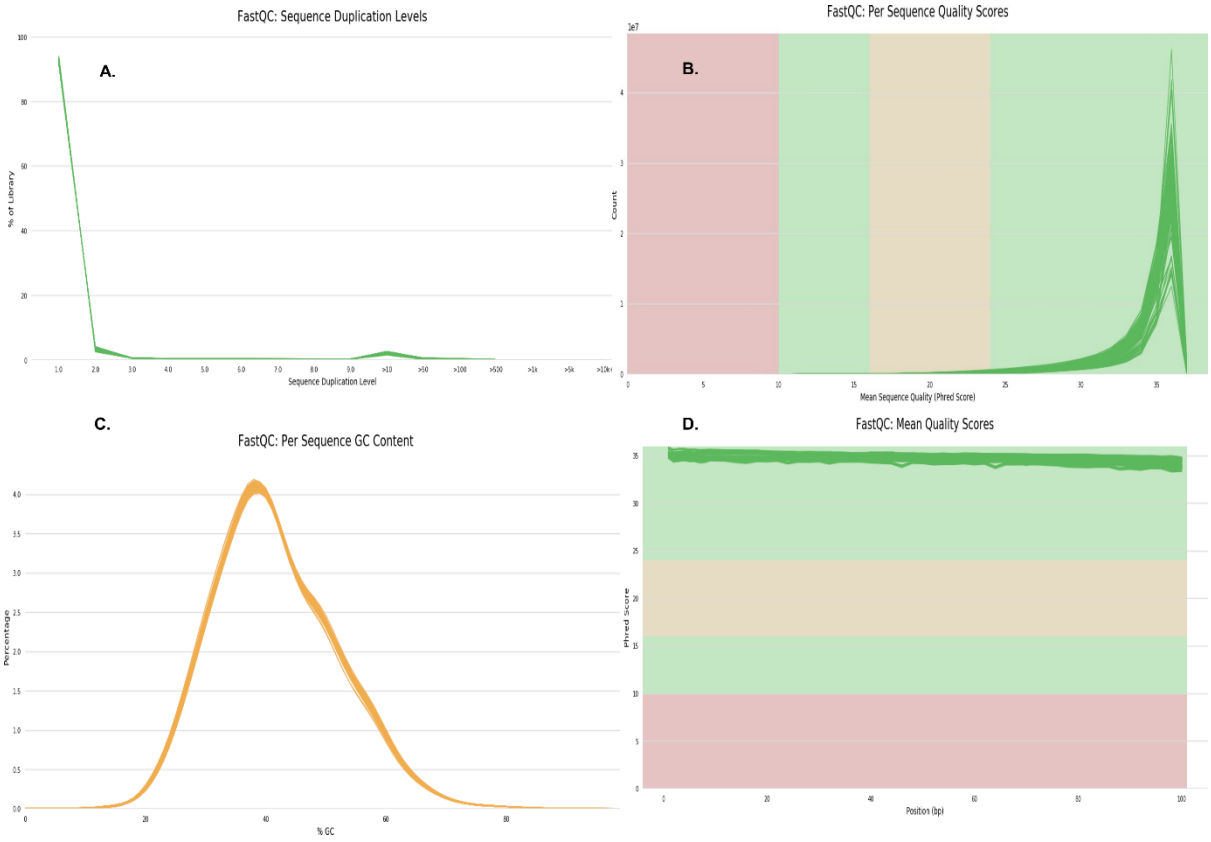
INVESTIGATOR: _____

INVESTIGATOR'S SIGNATURE: _____

APPENDIX B

SECTION 1

SUPPLEMENTARY DATA - CHAPTER 2



Supplementary Figure 1: Overall quality control data of the bam files

Supplementary Table 1: Data obtained from 1000 Genomes Project (1KGP) (Consortium et al., 2012) and the African Genome Variation Project (AGVP) (Gurdasani et al., 2015) used for analysis.

Population label	Ethnic group	Population description	Total Samples
AFR	Afro-Asiatic Semitic	Amhara:Ethiopia	22
	African-American	Americans of African Ancestry in SW USA (ASW)	60
	African-Caribbean	African Caribbeans in Barbado (ACB)	96
	Afro-Asiatic	Al-Gharbiyah, NA, Monufia, Kafrel-Sheikh, Mansoura, Alexandria, Dakahlia, Samanoud, Al-Buhayrah, Minya, AlSharqia, El-Mahalla all from Egypt	99
	Afro-Asiatic Cushitic	Oromo, Somali from Ethiopia	47
	Afro-Asiatic Omotic	Wolayta from Ethiopia	24
	KhoeSan	Khoe-San:Khoesan	84
	Niger Congo Bantu	Baganda, Banyarwanda, Barundi, RwandeseU-gandan, Banyankole:Uganda Bakiga, Mutan- zania, Basoga, other uganda gwas unknown, Mutooro, Batooro, Nyanjiro (Tanzania) from Uganda and Luhya in Webuye, Kenya (LWK)	2158
	Niger-Congo Bantu South	Zulu	98
	Niger-Congo Volta Niger	Esan in Nigeria (ESN), Yoruba in Ibadan, Nigeria (YRI)	205
	Niger-Congo West	Gambian in Western Divisions in the Gambia (GWD), Mende in Sierra Leone (MSL)	198
AMR	Latin American	Puerto Ricans from Puerto Rico (PUR), Colombians from Medellin, Colombia (CLM), Peruvians from Lima, Peru (PEL), Mexican Ancestry from Los Angeles USA (MXL)	347
EUR	European Center	British in England and Scotland (GBR)	91
	European North	Finnish in Finland (FIN)	99
	European South	Iberian Population in Spain (IBS), Toscani in Italia (TSI)	214
	European USA	Utah Residents with Northern and Western European Ancestry (CEU)	99
EAS	East Asian	Southern Han Chinese (CHS), Chinese Dai in Xishuangbanna, China (CDX), Kinh in Ho Chi Minh City, Vietnam (KHV), Han Chinese in Beijing, China (CHB), Japanese in Tokyo, Japan (JPT)	504
SAS	South Asian	Punjabi from Lahore, Pakistan (PJL), Bengali from Bangladesh (BEB)	180
	UK Indian	Sri Lankan Tamil from the UK (STU), Indian Telugu from the UK (ITU)	204
	USA Indian	Gujarati Indian from Houston, Texas (GIH)	103
Total			4,932

Supplementary Table 2: Genes previously associated with ligament and tendon injury.

Gene Name	Encoded Protein	Chromosomal Location	Variant
ACAN	Aggrecan	15q26.1	rs2351491 C/T, rs1042631 C/T, rs1516797 T/G
ADAMTS10	Metallopeptidase with Thrombospondin Type 1 Motif 10	19p13.2	rs62621197 C/T
ADAMTS17	Metallopeptidase with Thrombospondin Type 1 Motif 17	15q26.3	rs72755233 G/A
ADIPOQ	Adiponectin	3q27.3	rs1501299 G/T
ANKH	Progressive ankylosis protein homolog	5p15.2	rs3045 A/G
BGN	Biglycan	Xq28	rs1126499 C/T, rs1042103 G/A
CASP8	Caspase-8	2q33-q34	rs3834129 ins/del, rs1045485 G/C
COL1A1	Pro-α polypeptide of collagen type I	17q21.33	rs1107946 G/T
COL3A1	α1 III collagen chain	2q31	rs1800255 G/A
COL5A1	Pro-α polypeptide of type V collagen	9q34.2-q34.3	rs12722 T/C
COL11A1	α1 XI collagen chain	1p21	rs3753841 T/C, rs1676486 C/T
COL11A2	α2 XI collagen chain	6p21.3	rs1799907 T/A
COL12A1	Pro-α1 polypeptide type XII collagen	6q12-q13	rs970547 A/G rs1800012 G/T
COL27A1	α1 XXVII collagen chain	9q32	rs946053 G/T
DCN	Decorin	12q22	rs516115 A/G
DEFB1	Beta-defensin 1	8p23.1	rs1800972 C/T
EFEMP1	GF-containing fibulin-like extracellular matrix protein 1	2p16.1	rs3791679 A/G
ESRRB	Steroid hormone receptor ERR2	14q24.3	rs1676303 C/T
FBN2	Fibrillin-2	5q23-q31	rs331079 G/T
FGF10	Fibroblast growth factor 10	5p12	rs11750845 C/T, rs1011814 T/C
FGF3	Fibroblast growth factor 3	11q13.3	rs12574452 G/A
FGFR1	Fibroblast growth factor receptor 1	8p11.23	rs13317 T/C
GDF5	Growth differentiation factor 5	20q11	rs143383 T/C
IL1B	Interleukin-Iβ	2q14	rs16944 T/C
IL1RN	Interleukin-1 receptor antagonist	2q14.2	rs2234663
IL6	Interleukin-6	1q21	rs1800795 G/C
IL6R	Interleukin-6 receptor	1q21	rs2221845 A/C
ITGB3	Integrin beta-3	17q21.32	-
KDR	Kinase insert domain receptor	4q11-4q12	rs2071559 A/G, rs2305948 G/A, rs1870377 T/A
LUM	Lumican	12q21.33	rs2268578 T/C
MIR608	microRNA 608	10q24.31	rs4919510 C/G
MMP1	Matrix metalloproteinase 1	11q22.3	rs1799750 1G/2G
MMP3	Matrix metalloproteinase 3	11q22.3	rs679620 A/G
MMP8	Matrix metalloproteinase 8	11q22.3	rs11225395 C/T
MMP10	Matrix metalloproteinase 10	11q22.3	rs486055 C/T
MMP12	Matrix metalloproteinase 12	11q22.3	rs2276109 A/G
THBS2	Thrombospondin-2	6q27	rs9406328 C/T
TIMP2	Metalloproteinase inhibitor 2	17q25	rs4789932 C/T
TNAP	Tissue-Nonspecific Alkaline Phosphatase	1p36.12	rs4654760 C/T
TNC	Tenascin-C glycoprotein	9q33	rs1330363 A/G, rs2104772 T/A, rs13321 G/C, rs1138545 G/A, rs3789870 C/T, rs7021589 A/G, rs10759753 T/C, rs72758637 G/C, rs7035322 G/T
VEGFA	Vascular endothelial growth factor A	6p21.1	rs699947 C/C, rs1570360 G/A, rs2010963 G/C

Supplementary Table 3: The table displays the top significant pathways, GO biological process, Molecular Function and Human Phenotypes associated with the genes previously associated with ligament and tendon injury, and the candidate list of predicted pathogenic genes and their interacting genes for Family A and Family B combined and independently.

Enrichment	P-value	Adjusted P-value	Database
Pathway			
Genes Previously Associated with Ligament and Tendon Injury			
PI3K-Akt signalling pathway	8.9e-11	2.8e-8	KEGG-Human 2019
Protein digestion and absorption	5.9e-10	9.0e-8	
Rheumatoid arthritis	1.1e-6	1.0e-4	
Extracellular matrix organization	1.9e-32	2.9e-29	Reactome 2016
Degradation of the extracellular matrix	7.1e-13	5.4e-10	
Collagen formation	6.4e-12	3.3e-9	
Assembly of collagen fibrils and other multimeric structures	1.4e-11	5.6e-9	
Integrin signalling pathway	2.7e-8	3.1e-6	Panther 2016
Family A and B			
Complement and coagulation cascades	3.0e-7	1.0e-4	KEGG-Human 2019
Purine metabolism	6.0e-7	9.2e-4	
Mismatch repair	6.9e-5	7.1e-3	
Lectin Induced Complement Pathway	1.2e-5	2.7e-3	Biocarta 2016
Metabolism of nucleotides	1.0e-5	9.1e-4	Reactome 2016
Lectin pathway of complement activation	1.0e-5	6.2e-4	
Common Pathway of Fibrin Clot Formation	6.0e-5	3.1e-2	
Family A			
Complement and coagulation cascades	3.0e-9	9.0e-7	KEGG-Human 2019
Mismatch repair	3.9e-5	3.0e-3	
Family B			
ABC transporters	8.0e-10	2.5e-7	KEGG-Human 2019
Mismatch repair	2.1e-7	3.2e-5	
Go Biological Processes			
Genes Previously Associated with Ligament and Tendon Injury			
Extracellular matrix organization	1.5e-28	7.8e-25	GO Biological Process 2018
Extracellular matrix disassembly	4.6e-14	1.1e-10	
Collagen fibril organization	1.7e-13	2.9e-10	
Family A and B			
Establishment of protein localization to peroxisome	1.4e-6	7.3e-3	GO Biological Process 2018
Peroxisomal transport	3.0e-6	5.7e-3	
Complement activation, lectin pathway	8.0e-6	8.2e-3	
Mismatch repair	2.0e-6	6.7e-3	
Family A			
Establishment of protein localization to peroxisome	8.0e-7	4.1e-3	GO Biological Process 2018
Peroxisomal transport	1.3e-6	3.3e-3	
Nucleobase-containing small molecule interconversion	2.2e-6	3.7e-3	
Family B			
Cargo loading into COPII-coated vesicle	1.2e-8	6.1e-5	GO Biological Process 2018
Antigen processing and presentation of peptide antigen via MHC class I	5.6e-7	9.4e-4	
Mismatch repair	9.5e-7	1.2e-3	

Molecular Function			
Genes Previously Associated with Ligament and Tendon Injury			
Growth factor receptor binding	7.3e-10	8.5e-7	GO Molecular Function 2018
Metalloendopeptidase activity	2.2e-7	1.3e-4	
Integrin binding	1.4e-6	4.1e-4	
Growth factor receptor binding	7.3e-10	8.5e-7	
Family A and B			
Adenylate kinase activity	1.4e-6	1.6e-3	GO Molecular Function 2018
Family A			
Adenylate kinase activity	2.0e-9	2.5e-6	GO Molecular Function 2018
Nucleotide kinase activity	5.0e-7	3.1e-4	
Satellite DNA binding	1.7e-4	2.5e-2	
Histone methyltransferase activity	1.3e-5	2.9e-3	
Family B			
ATPase-coupled anion transmembrane transporter	1.2e-8	2.7e-6	GO Molecular Function 2018
Human Phenotype			
Genes Previously Associated with Ligament and Tendon Injury			
Osteoarthritis	2.2e-12	4.0e-9	Human Phenotype Ontology
Family A and B			
Autosomal recessive inheritance	5.0e-10	8.1e-8	Human Phenotype Ontology
Family A			
Abnormality of the common coagulation pathway	1.5e-6	8.7e-5	Human Phenotype Ontology
Autosomal recessive inheritance	1.9e-7	8.4e-5	
Family B			
Autosomal recessive inheritance	6.9e-7	1.2e-4	Human Phenotype Ontology

Supplementary Table 4: Inferred functional partners and enriched pathways for genes of interest in Family A and B

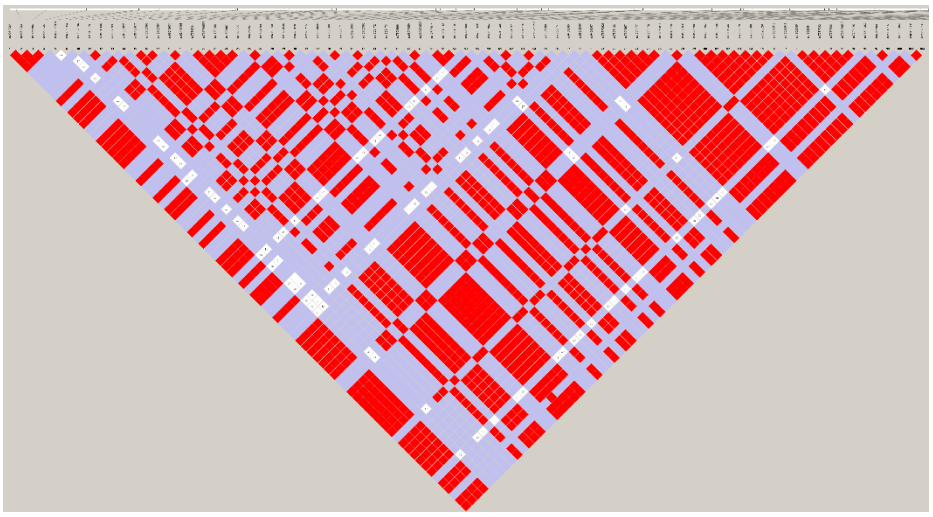
Gene of Interest	Top 100 inferred functional partners	Enrichment pathways	Summary of inferred partners
COL11A1 rs139064549 (G/C)	*BMP1,BRAF,CAPN2,CDC42,COL10A1,**COL11A2,COL12A1**,COL13A1,COL14A1,COL15A1,COL16A1,COL17A1,COL18A1,COL19A1,**COL1A1**,COL1A2,COL20A1,COL21A1,COL22A1,COL23A1,COL24A1,COL25A1,**COL27A1,**COL2A1,COL3A1,COL4A1,COL4A2,COL4A3,COL4A4,COL4A5,COL4A6,**COL5A1**,COL5A2,COL5A3,COL6A1,COL6A2,COL6A3,COL7A1,COL8A1,COL8A2,COL9A1,COL9A2,COL9A3,EGFR,ELN,FBN1,**FBN2**,FN1,GRB2,ITGA1,ITGA10,ITGA11,ITGA2,ITGA2B,ITGA3,ITGA4,ITGA5,ITGA6,ITGA7,ITGA8,ITGA9,ITGAD,ITGAE,ITGAL,ITGAM,ITGAV,ITGAX,ITGB1,ITGB2,**ITGB3**,ITGB4,ITGB5,ITGB6,ITGB7,ITGB8,**KDR**,KRAS,LAMA1,LAMA2,LAMA3,LAMA4,LAMA5,LAMB1,LAMB2,LAMB3,LAMC1,LAMC2,LAMC3,MAPK1,MAPK3,PDGFA,PRKCA,PXN,RHOA,SPP1,**TGFB1**,TGFB2,TGFB3,**TNC**,VCAN*	• ECM-receptor interaction • Regulation of the actin cytoskeleton • Focal adhesion • PI3K-Akt signalling pathway • ECM organization	The majority of the genes (39%) were collages, with integrin and laminin encoding genes making up 26 and 11% of the genes respectively. Proteoglycans made-up 2% of the genes, with 5% of the genes encoding for growth factors. 4% were glycoprotein encoding genes and 3% protein kinase encoding genes.
COL12A1 rs970547 (C/T)	*HSPG2,**ACAN**,ACTA1,AKT1,**BGN**,BMP1,BMP2,BMP4,BRAF,CAPN3,CASP3,COL10A1,**COL11A1,COL11A2**,COL13A1,COL14A1,COL15A1,COL17A1,COL18A1,COL19A1,**COL1A1**,COL1A2,COL21A1,COL22A1,COL23A1,COL24A1,COL25A1,**COL27A1**,COL2A1,COL3A1,COL4A1,COL4A2,COL4A3,COL4A4,COL4A5,**COL5A1**,COL5A2,COL5A3,COL6A1,COL6A2,COL6A3,COL7A1,COL8A1,COL8A2,COL9A1,COL9A2,COL9A3,CREB1,DAG1,**DCN**,DMD,**EFEMP1**,EFEMP2,EGFR,ELN,FBN1,**FBN2**,FGFR3,FN1,**GDF5**,GNAS,HAPLN1,IGF1,ITGA11,ITGA2B,ITGA3,ITGA5,ITGA6,ITGA7,ITGAV,ITGB1,ITGB2,**ITGB3**,ITGB4,**KDR**,KRAS,LAMA1,LAMA2,LAMA3,LAMA4,LAMA5,LAMB1,LAMB2,LAMB3,LAMC1,LAMC2,LMNA,LTBP4,MMP13,MMP9,PDPK1,PTEN,PTK2,PP1,SRC, **TGFB1**,TGFB2, TGFB3, TIMP1,VCAN*	• ECM-receptor interaction • Focal adhesion • ECM organization • Collagen formation • Collagen biosynthesis and modifying enzymes • Assembly of collagen fibrils • ECM proteoglycans	36% of the genes code for collagens with 11% for both integrin and laminin encoding proteins. Proteoglycan encoding genes made-up 7%, with growth factors contributing 9% of the genes. Genes encoding protein kinases and glycoproteins both made-up 5% of the genes.
CATSPER2 rs144399798 (C/T)	*ACR,ADAM2,ADAM20,ADAM21,ADAM30,ADCY10,AURKB,B4GALT1,B4GALT7,BRCA1,BRSK2,CACNA1C,CACNA1D,CACNA1F,CACNA1S,CACNB4,CACNG7,CALM2,CATSPER1,CATSPER3,CATSPERB,CATSPERD,CATSPERG,CD9,CNGA3,FOS,GABRA1,GABRA4,GABRG3,GNA14,GNAS,GNAZ,GNB1,GNB4,GNG7,GNRH1,HCN2,HVCN1,ITPR1,IZUMO1,IZUMO2,IZUMO4,JUN,KCNK10,KCNU1,KCNV2,KCTD3,KCTD6,MAP2K1,MAP2K3,MAP2K4,MAP3K11,MAPK3,MAPK8,MAPK9,MERTK,MYH14,MYH2,MYH9,MYO1C,MYO6,MYO7A,MYO9B,NOS1,NOS3,NPR1,OVGP1,PDE4B,PLA2G10,PLA2G2A,PLA2G4A,PLA2G6,PLCD1,PLCD4,PLCZ1,PPP1R12A,PRKAR1A,PRKCE,PRKCQ,PRKG1,PTK2,RORA,RORB,SLC22A3,SPAM1,STK11,STK33,STK38,STK39,STRC,TAS2R16,TAS2R7,TET1,TSSK6,TXK,ZAN,ZP1,ZP2,ZP3ZP4*	• Gonadotropin-releasing hormone signalling pathway • Vascular smooth muscle contraction • Fertilization • Reproduction pathways	20% of the genes are involved in fertilization pathways, 19% of the genes encode for protein kinases, and the remaining are G-proteins, transmembrane proteins, glycoproteins, and proteins involved in calcium signalling and channel activity.

KCNJ12 rs76265595 (G/A) rs75029097 (G/A) rs77270326 (G/A)	*ABCC8,ABCC9,ADCY1,ADCY2,ADCY3,ADCY4,ADCY5,ADCY6,ADCY7,ADCY8,ADCY9,AKAP9,ATP2A2,ATP2A3,ATP2C1,ATP6AP1,ATP6AP2,ATP6V0D1,B2M,CACNA1C,CACNA1D,CACNA1F,CACNA1S,CACNA2D1,CACNB1,CACNB2,CACNB3,CACNB4,CACNG4,CACNG7,CALM1,CALM2,CALM3,CAMK2B,CAMK2D,CAMK4,CAV3,CYP17A1,GABRA6,GNAI1,GNAI2,GNAI3,GNAS,GNB1,GNB2,GNB3,GNG10,GNG12,GNG2,GNG3,GNG4,GNG5,GNG7,GNG8,GNGT1,GNGT2,GUCY1A1,GUCY1A2,HCN4,HSPD1,ITPR1,ITPR2,JUN,KCNA1,KCNA5,KCND2,KCND3,KCNE3,KCNG2,KCNG4,KCNH2,KCNH4,KCNJ1,KCNJ10,KCNJ11,KCNJ14,KCNJ15,KCNJ16,KCNJ2,KCNJ3,KCNJ4,KCNJ5,KCNJ6,KCNJ8KCNJ9,KCNK3,KCNMB1,KCNQ1,KCNQ2,KCNQ3,KRAS,MAPK1,PLCB3,PRKACA,PRKACB,PRKCA,RYR1,RYR2,SCN5A,VCL*	• Circadian entrainment • Cholinergic synapse • Oxytocin signalling pathway • GABAergic synapse • Morphine addiction • Inwardly rectifying K+ channels • Activation of G protein gated K+ channels	The majority of the genes encode for proteins involved in ion channel activity (46%). Other genes encode for G proteins (17%) and adenylate cyclase's (9%).
GP6	*AKT1,AKT2,AKT3,BSG,CD36,CD40LG,CD44,CD47,CDKN2A,CLEC1B,**COL1A1**, COL1A2,COL4A1,COL4A2,CSF2,EDIL3,F2,F2R,F3,F5,F8,F9,FCER1G,FGA,FGB, FGG,FN1,FYN,GATA3,GNA13,GNAQ,GNAS,GP1BA,GP1BB,GP5,GP9,HDAC2,IFNAR1,IGF2,IL2,IL4,**IL6**,IL6ST,ITGA1,ITGA2,ITGA2B,ITGA3,ITGAV,ITGB1,ITGB2, **ITGB3**,JAK2,JAM3,KIT,KRAS,LUM,LYN,MAPK1,MERTK,NCOA3,NOS3,NRAS,PDPK1,PIK3CA,PIK3CB,PIK3CG,PIK3R1,PLA2G4A,PLAT,PLAU,PLCG2,PLG,PRKCZ, PROC,PROS1,PSEN1,PTGIR,PTGS1,PTPN11,PTPN6,SDC1,SELE,SELL,SELP,SERPINC1,SERPINE1,SHC1,SLC7A11,SNAI2,SOD1,SPP1,SRC,STIM1,**TGFB1**,THBD,THBS1,TNF,TP53,**VEGFA**,VWF*	• ECM-receptor interaction and organization • Hemostasis • Cell surface interactions at the vascular wall	Collagen genes comprised only 4%, while glycoproteins and protein kinases made-up 12 and 10% of the protein encoding genes respectively. Further, integrins (8%) and interleukins (4%) contributed to the functional genes. The remaining genes coded for coagulation factors, fibrinogens, proteoglycans and growth factors.
MIR99A	*APC,BRCA1,CASP3,CCND1,CD44,CDKN1A,CDKN1B,CDKN2A,EGFR,ERBB2,HRAS,KRAS,MAPK1,MIR100,MIR101.1,MIR103A1,MIR106B,MIR10A,MIR10B,MIR122,MIR124.1,MIR125A,MIR125B1,MIR126,MIR129.1,MIR135A1,MIR135B,MIR141,MIR143,MIR145,MIR146A,MIR150,MIR155,MIR15A,MIR15B,MIR17,MIR181A1,MIR183,MIR18A,MIR192,MIR193B,MIR195,MIR199A1,MIR19A,MIR200A,MIR200B,MIR200C,MIR205,MIR20A,MIR21,MIR210,MIR214,MIR221,MIR222,MIR223,MIR224,MIR23A,MIR23B,MIR25,MIR26B,MIR27A,MIR27B,MIR28,MIR29A,MIR29B1,MIR29C,MIR30A, MIR30B,MIR30C1,MIR30E,MIR31,MIR324,MIR330,MIR331,MIR335,MIR342,MIR34A,MIR34C,MIR373,MIR423,MIR483,MIR91,MIR92A1,MIR96,MIRLET7A1,MIRLET7B,MIRLET7D,MIRLET7E,MIRLET7G,MIRLET7I,MMP9,MTOR,MYC,NOTCH1,PIK3CA,PTEN,PTGS2,STAT3,TP5,**VEGFA***	• Interleukin signalling pathway • VEGF signalling pathay • Angiogenesis	75% of the inferred genes encoded for miRNAs. The remaining genes encode for glycoproteins, protein kinases and growth factors.

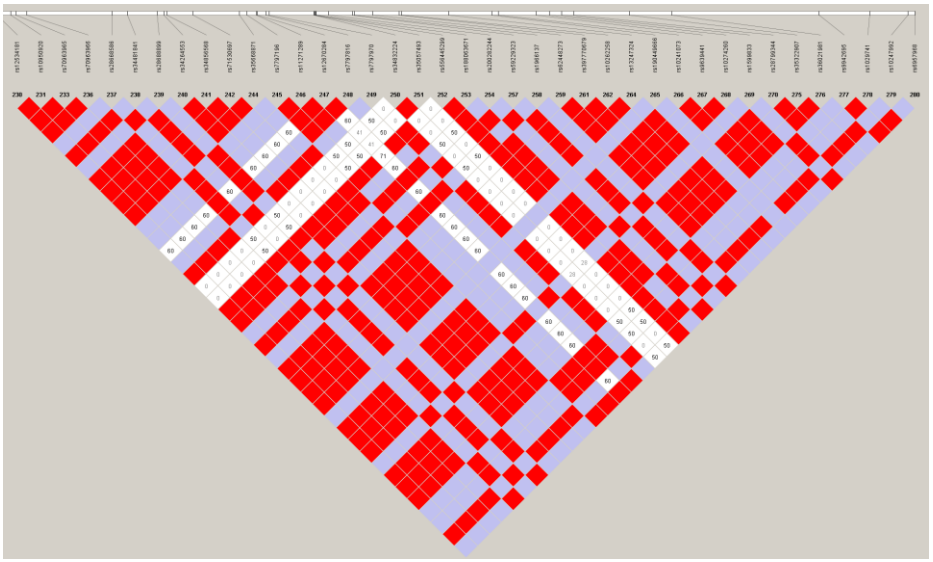
MIR99AHG	*AANAT,ABCB1,ABL1,ADGRG4,ADPGK,AKT1,ALX3,ANKRD1,APAF1,APOL6,ARHGAP20,ASB2,ASXL1,BAALC,BAD,BAX,BCL2L1,BCR,CBFA2T2,CBFB,CBL,CCNA1,CCNA2,CD14,CD33,CD34,CDK4,CDKN1A,CEBPA,CHIC2,CHID1,CREBBP,CSF1R,CSF2,CSF3,CSF3R,DNMT3A,ELAVL3,ELF4,ERG,ETV6,FES,FLT3,FUS,GATA2,HOXA9,HRAS,IDH1,IDH2,IL3,ITGAM,JAK2,KAT6A,KIT,KMT2A,KRAS,LIF,LYN,MAML3,MAP2K1,MAPK1,MCL1,MECOM,MEIS1,MIR100HG,MLF1,MLF2,MNDA,MPO,MYC,MYH11,NPM1,NRAS,NSD1,NTRK2,NUP214,NUP98,PCM1,PICALM,PIM1,PRDM16,PTPN11,RAF1,RARA,RUNX1,RUNX1T1,RUNX2,RUNX3,SEM1,SNAP91,SPI1,STAT3,STAT5A,STAT5B,TERT,TET2,**TGFB1**,TP53,U2AF1,WT1*	• VEGF signalling pathway • Interleukin signalling pathway	Transcription factors and co-activators made-up 23% of the protein encoding genes, with 14% encoding for protein kinases. Apoptosis related genes were encoded by 5% of the genes.
MIR125B2	*MIR125B1,ABCB1,ABCC1,AKT1,AKT2,APC,BAK1,BCL2,BRCA1,CASP3,CCND1,CCNE1,CD44,CDK6,CDKN1A,CDKN1B,CDKN2A,CTNNB1,DNMT1,DNMT3B,E2F1,E2F2,E2F3,EGF,EGFR,ERBB2,FGF2,HRAS,IGF1,IGF1R,IKBKB,MAP2K1,MAP2K2,MAPK1,MAPK3,MAPK8,MDM2,MIR103A2,MIR10B,MIR1.1,MIR125A,MIR126,MIR128.2,MIR130B,MIR133B,MIR137,MIR140,MIR155,MIR15B,MIR181A2,MIR181C,MIR185,MIR195,MIR197,MIR199A1,MIR199A2,MIR199B,MIR21,MIR221,MIR222,MIR23A,MIR23B,MIR25,MIR27B,MIR30E,MIR34A,MIR451A,MIR483,MIR494,MIR73,MIR92,MIR96,MIR99A,MIRLET7C,MTOR,MYEF2,NFKB1,NRAS,PDGFRA,PIK3CA,PIK3CB,PIK3CD,PIK3CG,PIK3R1,PIK3R2,PIK3R3,PPP3R1,PRKCA,PRKCB,PTEN,RAF1,RASGRF1,RHOA,ROCK1,STAT3,**TGFB1**,TNF,TP53,**VEGFA**,WNT3A*	• VEGF signalling pathway • Angiogenesis • Hypoxia response via HIF activation • MiRNAs in cancer • PI3K-Akt signalling pathway	Most of the genes (38%) encoded for miRNAs, with 18 and 5% encoding for protein kinases and growth factors respectively. The remaining genes encoded transcription factors, hormones, angiogenesis related genes.
MIRLET7C	*ABCB1,ABCC1,APC,ATM,BCL2,BCL2L11,BRCA1,CASP3,CCND1,CD44,CDC42,CDK6,CDKN1A,CDKN1B,CDKN2A,DNMT1,DNMT3B,E2F1,E2F3,EGFR,ERBB2,EZH2,HDAC1,HRAS,KRAS,MAP2K1,MAPK1,MAPK3,MCL1,MDM2,MIR100,MIR101.1,MIR101.2,MIR106B,MIR125A,MIR126,MIR128.2,MIR143,MIR145,MIR146A,MIR148A,MIR148B,MIR150,MIR155,MIR15A,MIR16.1,MIR17,MIR181A1,MIR183,MIR195,MIR199A1,MIR199B,MIR200A,MIR200B,MIR200C,MIR203A,MIR21,MIR210,MIR221,MIR222,MIR223,MIR23A,MIR23B,MIR27A,MIR27B,MIR29A,MIR29B1,MIR29B2,MIR29C,MIR30A,MIR31,MIR335,MIR34A,MIR34B,MIR34C,MIR451A,MIR4763,MIR9.1,MIR99A,MIRLET7A1,MIRLET7A2,MIRLET7A3,MIRLET7B,MIRLET7BHG,MIRLET7D,MIRLET7E,MIRLET7F1,MIRLET7F2,MIRLET7G,MMP9,MTOR,MYC,NOTCH1,PIK3CA,PIK3R1,PTEN,SIRT1,STAT3,TP53,**VEGFA***	• VEGF signalling pathway • Angiogenesis • Interleukin signalling pathway	Most of the genes (59%) encode for miRNAs. The remaining genes encode for glycoproteins, protein kinases and growth factors. Furthermore, a few of the genes are involved in apoptotic pathways.
LINC01250	*no data found*		

Genes in **bold** script: highlighted as previously associated with musculoskeletal injury predisposition.

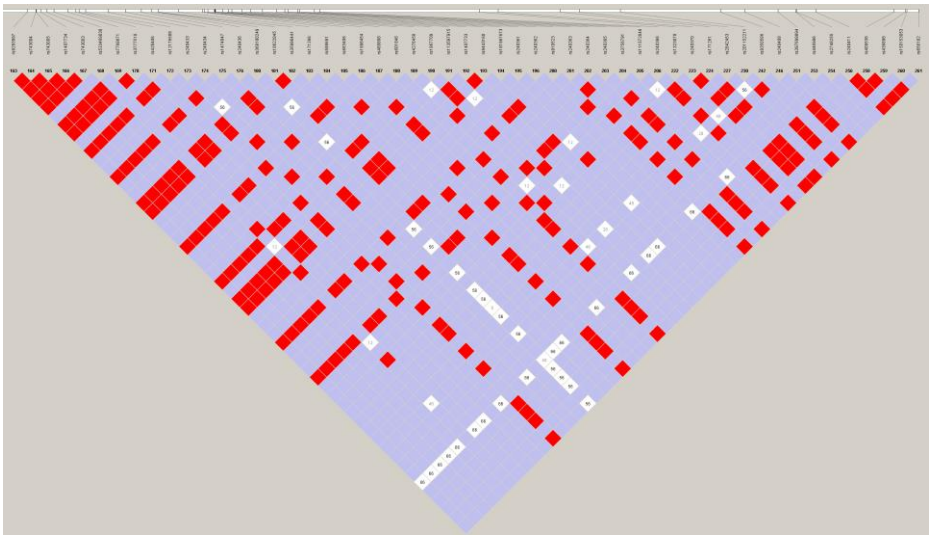
A



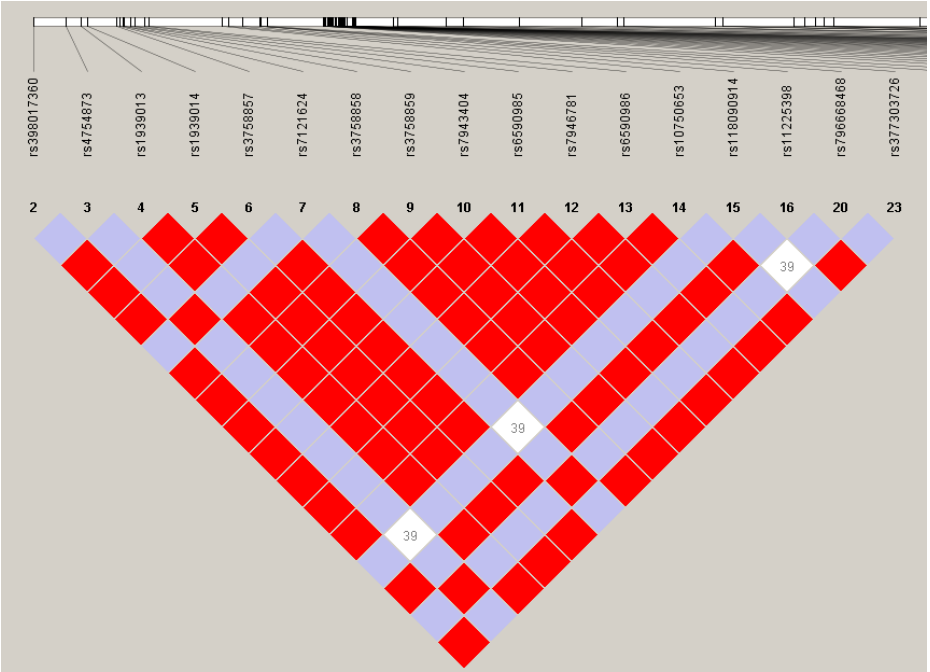
B



C



D



Supplementary Figure 2: Linkage disequilibrium for (A) *CATSPER2*, (B) *LINC01250*, (C) *COL12A1*, and (D) *MMP1*.

SECTION 2

SUPPLEMENTARY DATA - CHAPTER 3

Supplementary Table 5: Descriptive Characteristics for all Swedish, Polish, and Australian participants.

	SWE CON (n = 116)	**SWE ACL (n = 95)**	**p-value**
Age (years)	44.7 ± 11.9 (114)	37.1 ± 13.3 (93)	**<0.001**
Sex (%male)	34 (116)	47 (95)	0.067
Height (cm)	172.3 ± 10.1 (108)	172.8 ± 8.6 (85)	0.732
Body Mass (kg)	72.1 ± 13.6 (107)	74.1 ± 13.2 (86)	0.302
BMI (kg.m^{-2})	24.4 ± 2.9 (107)	24.7 ± 3.0 (84)	0.523
	POL CON (n = 149)	**POL ACL (n = 127)**	**p-value**
Age (years)	21.0 ± 1.8 (149)	31.4 ± 10.1 (127)	**<0.001**
Sex (%male)	75 (149)	80 (127)	0.680
Height (cm)	177.9 ± 9.7 (149)	176.5 ± 12.7 (124)	0.280
Body Mass (kg)	72.6 ± 12.0 (149)	79.0 ± 14.9 (126)	**<0.001** (**0.003**)[a]
BMI (kg.m^{-2})	22.8 ± 2.4 (149)	25.1 ± 4.1 (123)	**<0.001** (0.307)[b]
	AUS CON (n = 83)	**AUS ACL (n = 343)**	**p-value**
Age (years)	31.0 ± 8.3 (82)	25.2 ± 9.4 (269)	**<0.001**
Sex (%male)	100 (83)	42 (342)	**<0.001**
Height (cm)	180.0 ± 0.1 (82)	175.2 ± 0.1 (268)	**<0.001** (0.120)[c]
Body Mass (kg)	81.6 ± 12.0 (82)	77.5 ± 14.7 (268)	**0.022** (0.123)[d]
BMI (kg.m^{-2})	25.2 ± 3.3 (82)	25.1 ± 3.7 (268)	0.896

Values are expressed as mean ± standard deviation; sex is represented as a percentage. The number of participants (n) with available data for each variable is in parentheses. CON vs ACL, p-values in bold typeset indicates significance ($p < 0.05$). P-values are depicted as unadjusted and adjusted in parentheses. a: age, b: age and body mass, c: age and sex, and d: age, sex and height.

Supplementary Table 6: Sports participation according to the type of sport played within the control (CON) group and the anterior cruciate ligament (ACL) rupture group, for participants in the Swedish population.

	CON (n= 92)	ACL (n=91)	p-value
Contact sports[a]			
Participants (%)	0 (0)	68 (62)	**<0.001**
Participation (years)	0	11.8 ± 10.5 (45)	n/a
Non-contact jumping sports[b]			
Participants (%)	28 (26)	26 (24)	0.869
Participation (years)	21.2 ± 17.4 (21)	15.4 ± 13.7 (19)	0.247
Non-contact non-jumping sports[c]			
Participants (%)	72 (66)	6 (5)	**<0.001**
Participation (years)	17.8 ± 13.8 (61)	12.3 ± 10.5 (4)	0.440

Participants in each group are represented as a frequency (%), while the years of participation are represented as a mean ± standard deviation. The number of participants (n) is in parentheses
CON vs. ACL, p-values in bold typeset indicates significance ($p < 0.05$)
[a]Football, floorball.
[b]Cross country skiing, downhill skiing, handball, martial arts, beach volleyball, gymnastics.
[c]Running, walking, swimming, weight lifting, cycling, dancing

Supplementary Table 7: Sports participation according to the type of sport played within the control (CON) group and the anterior cruciate ligament (ACL) rupture group, for participants in the Polish cohort.

	CON (n= 142)	**ACL (n=117)**	**p-value**
Contact sports[a]			
Participants (%)	56 (80)	52 (61)	0.532
Non-contact jumping sports[b]			
Participants (%)	0 (0)	21 (24)	**< 0.001**
Non-contact non-jumping sports[c]			
Participants (%)	44 (62)	27 (32)	**0.009**

Participants in each group are represented as a frequency (%) while the number of participants (n) is in parentheses.
CON vs. ACL, p-values in bold typeset indicates significance ($p < 0.05$).
[a]Football, boxing, karate, American football
[b]Skiing, volleyball, basketball
[c]Running, dancing, gym, cycling, handball, aerobics, swimming, crossfit, motocross.

Supplementary Table 8: Sports participation according to the type of sport played within the anterior cruciate ligament (ACL) rupture group, for participants in the Australian population.

	ACL (n = 288)
Contact sports[a]	
Participants (%)	85 (245)
Non-contact jumping sports[b]	
Participants (%)	4 (11)
Non-contact non-jumping sports[c]	
Participants (%)	11 (32)

Participants in each group are represented as a frequency (%) while the number of participants (n) is in parentheses. CON vs. ACL, p-values in bold typeset indicates significance ($p < 0.05$).
[a]Australian rules, netball, soccer, basketball.
[b]Skiing, wakeboarding, trampoline, dodgeball, skateboarding, jiu jitsu, gymnastics, long jump, cheerleading, judo.
[c]Motorcross, weight training, cycling, track and field, lacrosse, athletics, tennis.

Supplementary Table 9: Sports participation according to the level of sport played within the control (CON) group and the anterior cruciate ligament (ACL) rupture group, for participants in the Swedish and Australian cohort.

SWEDEN			
	CON (n = 96)	ACL (n = 85)	p-value
Elite	8 (8)	6 (5)	0.577
National	8 (8)	18 (15)	0.075
Recreational	83 (80)	76 (65)	0.268
AUSTRALIA			
	CON (n = 83)	ACL (n = 290)	p-value
National	0 (0)	2 (7)	0.356
Recreational	100 (83)	98 (283)	0.356

Supplementary Table 10: Genotype effects on participant characteristics in the combined cohort for *VEGFA* (A: rs699947, B: rs1570360 and C: rs2010963) and *KDR (*D: rs2071559 and E: rs1870377) polymorphisms.

A. Genotype effects *VEGFA* rs699947 C>A

	A/A	**C/A**	**C/C**	**p-value**
Age (years)	29.4 ± 11.5 (293)	29.4 ± 11.9 (583)	28.6 ± 11. 1 (268)	0.629
Sex (% male)	63 (293)	63 (583)	65 (268)	0.836
Height (cm)	175.1 ± 10.8 (290)	175.9 ± 9.7 (576)	176.4 ± 9.3 (265)	0.278
Body mass (kg)	74.4 ± 13.1 (291)	76.5 ± 13.9 (579)	77.7 ± 15.9 (268)	**0.018**
Body mass index (kg/m^2)	24.4 ± 5.8 (289)	24.7 ± 3.3 (572)	24.9 ± 4.1 (265)	0.458

B. Genotype effects *VEGFA* rs1570360 G>A

	A/A	**A/G**	**G/G**	**p-value**
Age (years)	28.4 ± 9.8 (125)	29.5 ± 12.0 (483)	29.5 ± 11.8 (526)	0.626
Sex (% male)	66 (126)	61 (493)	63 (528)	0.622
Height (cm)	176.2 ± 9.0 (126)	175.6 ± 9.8 (476)	176.3 ± 9.5 (514)	0.522
Body mass (kg)	74.2 ± 12.3 (125)	75.8 ± 14.0 (477)	77.1 ± 15.2 (521)	0.098
Body mass index (kg/m^2)	23.9 ± 2.8 (124)	24.5 ± 3.4 (473)	24.8 ± 3.7 (513)	**0.048**

C. Genotype effects *VEGFA* rs2010963 G>C

	G/G	**C/G**	**C/C**	**p-value**
Age (years)	28.2 ± 10.8 (508)	30.1 ± 12.2 (502)	32.3 ± 13.4 (153)	**0.0003**
Sex (% male)	64 (510)	63 (512)	60 (154)	0.669
Height (cm)	175.8 ± 10.3 (501)	175.9 ± 9.7 (494)	175.6 ± 9.9 (150)	0.908
Body mass (kg)	75.1 ± 14.2 (501)	77.3 ± 14.9 (500)	76.3 ± 14.1 (151)	**0.047**
Body mass index (kg/m^2)	24.4 ± 4.9 (497)	24.9 ± 3.8 (492)	24.6 ± 3.0 (150)	0.166

D. Genotype effects *KDR* rs2071559 G>A

	G/G	**A/G**	**A/A**	**p-value**
Age (years)	28.8 ± 10.9 (295)	30.0 ± 12.1 (575)	29.4 ± 12.1 (293)	0.301
Sex (% male)	66 (299)	61 (584)	67 (293)	0.143
Height (cm)	176.9 ± 9.8 (294)	175.4 ± 10.2 (564)	175.6 ± 9.7 (288)	0.101
Body mass (kg)	77.3 ± 14.9 (295)	75.9 ± 14.2 (569)	75.7 ± 14.5 (289)	0.364
Body mass index (kg/m^2)	24.6 ± 3.7 (293)	24.8 ± 4.9 (561)	24.4 ± 3.5 (286)	0.598

E. Genotype effects *KDR* rs1870377 T>A

	T/T	**T/A**	**A/A**	**p-value**
Age (years)	28.6 ± 11.5 (645)	30.1 ± 12.1 (421)	32.0 ± 11.3 (86)	**0.014**
Sex (% male)	63 (650)	64 (428)	53 (87)	0.122
Height (cm)	176.0 ± 9.6 (638)	176.1 ± 10.5 (414)	173.5 ± 9.0 (84)	0.075
Body mass (kg)	75.8 ± 14.3 (645)	76.8 ± 17.8 (412)	75.3 ± 13.3 (85)	0.441
Body mass index (kg/m^2)	24.1 ± 3.4 (637)	24.8 ± 5.6 (409)	24.9 ± 3.2 (84)	0.227

Supplementary Table 11: Genotype effects on participant characteristics in the Swedish cohort for *VEGFA* (A: rs699947, B: rs1570360 and C: rs2010963) and *KDR (*D: rs2071559 and E: rs1870377) polymorphisms.

A. Genotype effects *VEGFA* rs699947 C>A

	A/A	**C/A**	**C/C**	**p-value**
Age (years)	41.1 ± 13.5 (57)	41.3 ± 13.0 (98)	38.4 ± 12.7 (32)	0.539
Sex (% male)	41 (58)	39 (101)	29 (42)	0.920
Height (cm)	170.9 ± 8.6 (53)	172.9 ± 9.6 (94)	172.3 ± 7.8 (30)	0.468
Body mass (kg)	70.9 ± 11.9 (54)	74.1 ± 12.2 (94)	73.5 ± 13.6 (30)	0.322
Body mass index (kg/m^2)	24.2 ± 2.3 (53)	26.4 ± 3.0 (93)	24.6 ± 3.3 (30)	0.637

B. Genotype effects *VEGFA* rs1570360 G>A

	A/A	**A/G**	**G/G**	**p-value**
Age (years)	38.4 ± 11.1 (20)	40.5 ± 12.8 (88)	41.5 ± 13.9 (88)	0.597
Sex (% male)	35 (20)	43 (91)	36 (88)	0.624
Height (cm)	170.8 ± 7.3 (19)	172.8 ± 9.7 (84)	171.9 ± 8.8 (78)	0.621
Body mass (kg)	70.9 ± 10.5 (19)	73.8 ± 12.6 (83)	72.1 ± 14.2 (79)	0.572
Body mass index (kg/m^2)	24.7 ± 2.3 (18)	24.6 ± 2.9 (83)	24.5 ± 3.1 (78)	0.975

C. Genotype effects *VEGFA* rs2010963 G>C

	G/G	**C/G**	**C/C**	**p-value**
Age (years)	43.1 ± 13.2 (58)	40.3 ± 12.8 (97)	41.1 ± 13.4 (53)	0.188
Sex (% male)	34 (58)	42 (100)	43 (53)	0.354
Height (cm)	171.9 ± 9.3 (51)	172.9 ± 9.4 (92)	172.3 ± 9.8 (50)	0.491
Body mass (kg)	71.5 ± 14.6 (51)	74.0 ± 12.9 (93)	72.5 ± 13.2 (50)	0.244
Body mass index (kg/m^2)	24.6 ± 2.4 (49)	24.6 ± 3.3 (92)	24.4 ± 2.6 (50)	0.960

D. Genotype effects *KDR* rs2071559 G>A

	G/G	**A/G**	**A/A**	**p-value**
Age (years)	38.9 ± 12.6 (51)	41.9 ± 12.9 (106)	42.3 ± 14.1 (47)	0.313
Sex (% male)	32 (53)	40 (107)	48 (48)	0.270
Height (cm)	171.0 ± 9.7 (49)	172.9 ± 9.0 (98)	173.0 ± 9.7 (44)	0.455
Body mass (kg)	69.3 ± 9.7 (49)	72.9 ± 12.6 (100)	76.8 ± 16.6 (42)	**0.034**
Body mass index (kg/m^2)	23.7 ± 2.6 (49)	24.5 ± 2.6 (98)	25.4 ± 3.5 (42)	**0.018**

E. Genotype effects *KDR* rs1870377 T>A

	T/T	**T/A**	**A/A**	**p-value**
Age (years)	40.1 ± 12.6 (99)	41.8 ± 13.9 (84)	44.3 ± 11.6 (17)	0.378
Sex (% male)	45 (101)	38 (85)	29 (17)	0.401
Height (cm)	173.5 ± 9.2 (93)	172.1 ± 9.7 (77)	170.6 ± 8.7 (16)	0.379
Body mass (kg)	72.7 ± 12.8 (96)	73.1 ± 13.9 (75)	72.8 ± 14.6 (16)	0.985
Body mass index (kg/m^2)	24.3 ± 2.6 (93)	25.5 ± 3.2 (75)	24.8 ± 3.4 (16)	0.763

Supplementary Table 12: Genotype effects on participant characteristics in the Polish cohort for *VEGFA* (A: rs699947, B: rs1570360 and C: rs2010963) and *KDR (*D: rs2071559 and E: rs1870377) polymorphisms.

A. Genotype effects *VEGFA* rs699947 C>A

	A/A	**C/A**	**C/C**	**p-value**
Age (years)	26.2 ± 9.8 (68)	25.7 ± 8.7 (135)	25.5 ± 7.6 (63)	0.883
Sex (% male)	69 (68)	75 (135)	75 (63)	0.666
Height (cm)	175.5 ± 14.8 (68)	177.7 ± 9.4 (133)	177.9 ± 10.6 (63)	0.222
Body mass (kg)	73.7 ± 11.3 (68)	74.9 ± 11.6 (134)	77.7 ± 18.4 (63)	0.483
Body mass index (kg/m^2)	24.7 ± 10.6 (68)	23.7 ± 2.7 (132)	24.4 ± 5.1 (63)	0.351

B. Genotype effects *VEGFA* rs1570360 G>A

	A/A	**A/G**	**G/G**	**p-value**
Age (years)	25.6 ± 8.9 (34)	25.9 ± 9.2 (104)	25.3 ± 8.2 (127)	0.851
Sex (% male)	68 (34)	74 (104)	76 (127)	0.647
Height (cm)	176.4 ± 10.0 (34)	177.7 ± 9.5 (103)	178.3 ± 9.5 (125)	0.158
Body mass (kg)	72.1 ± 10.9 (34)	74.9 ± 12.0 (103)	76.9 ± 15.5 (127)	0.220
Body mass index (kg/m^2)	23.1 ± 2.6 (34)	23.6 ± 2.6 (102)	24.1 ± 4.1 (125)	0.581

C. Genotype effects *VEGFA* rs2010963 G>C

	G/G	**C/G**	**C/C**	**p-value**
Age (years)	25.7 ± 9.1 (143)	26.2 ± 8.7 (104)	24.4 ± 5.9 (18)	0.713
Sex (% male)	71 (143)	76 (104)	83 (18)	0.404
Height (cm)	176.1 ± 12.3 (142)	178.6 ± 9.6 (102)	180.5 ± 10.1 (18)	0.059
Body mass (kg)	73.7 ± 12.1 (142)	77.7 ± 15.6 (104)	77.8 ± 13.4 (18)	0.951
Body mass index (kg/m^2)	24.1 ± 7.6 (141)	24.2 ± 4.4 (102)	23.8 ± 2.8 (18)	0.106

D. Genotype effects *KDR* rs2071559 G>A

	G/G	**A/G**	**A/A**	**p-value**
Age (years)	24.9 ± 8.1 (67)	25.8 ± 7.9 (123)	26.9 ± 10.3 (77)	0.398
Sex (% male)	84 (67)	71 (123)	70 (77)	0.109
Height (cm)	180.1 ± 9.3 (67)	176.5 ± 13.1 (122)	176.3 ± 8.8 (75)	0.199
Body mass (kg)	78.2 ± 16.9 (67)	75.1 ± 12.9 (122)	74.4 ± 12.0 (77)	0.726
Body mass index (kg/m^2)	24.0 ± 4.8 (67)	24.5 ± 8.2 (121)	23.8 ± 2.9 (75)	0.068

E. Genotype effects *KDR* rs1870377 T>A

	T/T	**A/T**	**A/A**	**p-value**
Age (years)	24.9 ± 8.1 (129)	26.3 ± 9.2 (114)	26.9 ± 8.9 (25)	0.395
Sex (% male)	71 (129)	81 (114)	56 (25)	**0.023**
Height (cm)	177.3 ± 9.5 (129)	177.8 ± 12.8 (113)	173.3 ± 10.4 (24)	0.194
Body mass (kg)	73.5 ± 11.4 (129)	77.6 ± 15.3 (113)	71.7 ± 11.8 (25)	**0.024**
Body mass index (kg/m^2)	23.3 ± 2.6 (129)	25.0 ± 8.9 (112)	23.7 ± 2.6 (24)	0.099

Supplementary Table 13: Genotype effects on participant characteristics in the Australian cohort for *VEGFA* (A: rs699947, B: rs1570360 and C: rs2010963) and *KDR (*D: rs2071559 and E: rs1870377) polymorphisms.

A. Genotype effects *VEGFA* rs699947 C>A

	A/A	**A/C**	**C/C**	**p-value**
Age (years)	25.6 ± 7.4 (87)	27.2 ± 10.6 (170)	26.7 ± 9.1 (92)	0.446
Sex (% male)	57.5 (106)	52.6 (205)	52.2 (111)	0.669
Height (cm)	176.6 ± 9.0 (87)	176.1 ± 9.5 (169)	176.5 ± 9.5 (92)	0.931
Body mass (kg)	77.7 ± 13.9 (87)	78.9 ± 13.7 (169)	78.6 ± 15.6 (92)	0.782
Body mass index (kg/m^2)	24.8 ± 3.7 (87)	25.4 ± 3.2 (169)	25.1 ± 3.9 (92)	0.542

B. Genotype effects *VEGFA* rs1570360 G>A

	A/A	**A/G**	**G/G**	**p-value**
Age (years)	25.3 ± 7.0 (42)	27.3 ± 11.1 (145)	26.6 ± 8.5 (158)	0.509
Sex (% male)	64.0 (50)	52.2 (174)	51.8 (195)	0.282
Height (cm)	178.9 ± 7.7 (42)	174.9 ± 9.4 (144)	176.8 ± 9.6 (158)	**0.015**
Body mass (kg)	77.8 ± 11.8 (42)	77.9 ± 13.9 (144)	79.0 ± 14.9 (158)	0.748
Body mass index (kg/m^2)	24.3 ± 3.1 (42)	25.4 ± (3.6)	25.2 ± 3.4 (158)	0.196

C. Genotype effects *VEGFA* rs2010963 G>C

	G/G	**C/G**	**C/C**	**p-value**
Age (years)	25.4 ± 7.6 (158)	28.0 ± 11.2 (151)	27.0 ± 9.2 (39)	0.051
Sex (% male)	55.4 (193)	53.0 (181)	51.0 (47)	0.824
Height (cm)	176.6 ± 9.1 (158)	175.8 ± 9.4 (150)	176.9 ± 10.4 (39)	0.646
Body mass (kg)	77.4 ± 13.7 (158)	79.6 ± 14.7 (150)	78.3 ± 15.1 (39)	0.381
Body mass index (kg/m^2)	24.7 ± 3.4 (158)	25.7 ± 3.6 (150)	24.8 ± 3.3 (39)	0.058

D. Genotype effects *KDR* rs2071559 G>A

	G/G	**A/G**	**A/A**	**p-value**
Age (years)	26.8 ± 7.9 (84)	27.3 ± 10.4 (178)	24.5 ± 8.9 (87)	0.343
Sex (% male)	55.2 (96)	54.5 (222)	50.4 (105)	0.748
Height (cm)	177.2 ± 9.8 (84)	175.7 ± 9.8 (86)	175.7 ± 9.8 (86)	0.571
Body mass (kg)	80.3 ± 14.5 (84)	78.3 ± 13.5 (178)	77.2 ± 15.5 (86)	0.334
Body mass index (kg/m^2)	25.4 ± 3.0 (84)	25.2 ± 3.5 (178)	24.9 ± 4.2 (86)	0.610

E. Genotype effects *KDR* rs1870377 T>A

	T/T	**T/A**	**A/A**	**p-value**
Age (years)	26.3 ± 10.2 (210)	26.5 ± 8.2 (116)	30.6 ± 8.8 (22)	0.128
Sex (% male)	56.2 (249)	0 (146)	53.8 (26)	0.415
Height (cm)	176.6 ± 9.8 (209)	175.9 ± 8.9 (116)	175.2 ± 7.7 (22)	0.711
Body mass (kg)	78.8 ± 14.9 (209)	175.9 ± 8.9 (116)	79.6 ± 10.1 (22)	0.629
Body mass index (kg/m^2)	25.2 ± 3.7 (209)	24.9 ± 3.5 (116)	25.9 ± 2.6 (22)	0.493

Supplementary Table 14: Minor allele frequency distributions for the *VEGFA* rs699947 (C/A), rs1570360 (G/A), rs2010963 (G/C) and *KDR* rs2071559 (A/G) and rs1870377 (T/A) polymorphisms, in all participants in the combined cohort.

	COMB CON	COMB ACL	p-value
***VEGFA* rs699947 (C/A)**			
C allele	51 (757)	51 (925)	0.771
A allele	49 (731)	49 (873)	
***VEGFA* rs1570360 (G/A)**			
G allele	69 (1039)	69 (1216)	0.713
A allele	31 (459)	31 (554)	
***VEGFA* rs2010963 (G/C)**			
G allele	66 (1005)	60 (1090)	**0.0006** (**0.003**)[a]
C allele	34 (513)	40 (716)	
***KDR* rs2071559 (A/G)**			
A allele	50 (561)	50 (691)	0.688
G allele	50 (571)	50 (679)	
***KDR* rs1870377 (T/A)**			
T allele	73 (818)	76 (1022)	0.097
A allele	27 (308)	24 (330)	

Allele frequencies are expressed as a percentage with the number of participants (n) in parentheses
CON vs. ACL[a] (False Discovery Rate, Benjamini-Hochberg)
P-values in bold typeset indicate significance ($p < 0.05$)

Supplementary Table 15: Genotype and minor allele frequency distributions, and p-values for Hardy-Weinberg exact test of the investigated polymorphisms in all participants (male and female) in the control (CON) and the anterior cruciate ligament (ACL) rupture group within the Swedish cohort.

	SWE CON	SWE ACL	p-value[a]
***VEGFA* rs699947 (C/A)**			
n	*101*	*90*	
C/C	20.8 (21)	12.2 (11)	0.762
A/C	51.5 (52)	54.4 (49)	
A/A	27.7 (28)	33.3 (30)	
A allele	53 (108)	61(109)	0.179
HWE	0.842	0.269	
***VEGFA* rs1570360 (G/A)**			
n	*104*	*95*	
G/G	48.1 (50)	40.0 (38)	0.642
G/A	40.4 (42)	51.6 (49)	
A/A	11.5 (12)	8.4 (8)	
A allele	32 (66)	34 (65)	0.669
HWE	0.499	0.252	
***VEGFA* rs2010963 (G/C)**			
n	*116*	*95*	
G/G	36.2 (42)	16.8 (16)	**< 0.001 (< 0.001)[b]**
G/C	53.4 (62)	40.0 (38)	
C/C	10.3 (12)	43.2 (41)	
C allele	37 (86)	63 (120)	**< 0.001 (< 0.001)[b]**
HWE	0.163	0.187	
***KDR* rs2071559 (A/G)**			
n	*114*	*94*	
G/G	21.9 (25)	29.8 (28)	0.108
A/G	58.8 (67)	42.6 (40)	
A/A	19.3 (22)	27.7 (26)	
A allele	49 (111)	49 (92)	1.000
HWE	0.091	0.152	
***KDR* rs1870377 (T/A)**			
n	*111*	*92*	
T/T	45.9 (51)	54.3 (50)	0.752
T/A	44.1 (49)	39.1 (36)	
A/A	9.9 (11)	6.5 (6)	
A allele	32 (71)	26 (48)	0.228
HWE	1.000	1.000	

Genotype and allele frequencies are expressed as a percentage with the number of participants (n) in parentheses.
CON vs. ACL[a] (adjusted p-values for age and type of sport).
CON vs. ACL[b] (False Discovery Rate, Benjamini-Hochberg).
P-values in bold typeset indicate significance ($p < 0.05$).

Supplementary Table 16: Genotype and minor allele frequency distributions, and p-values for Hardy-Weinberg exact test of the investigated polymorphisms in all participants (male and female) in the control (CON) and the anterior cruciate ligament (ACL) rupture group within the Polish cohort.

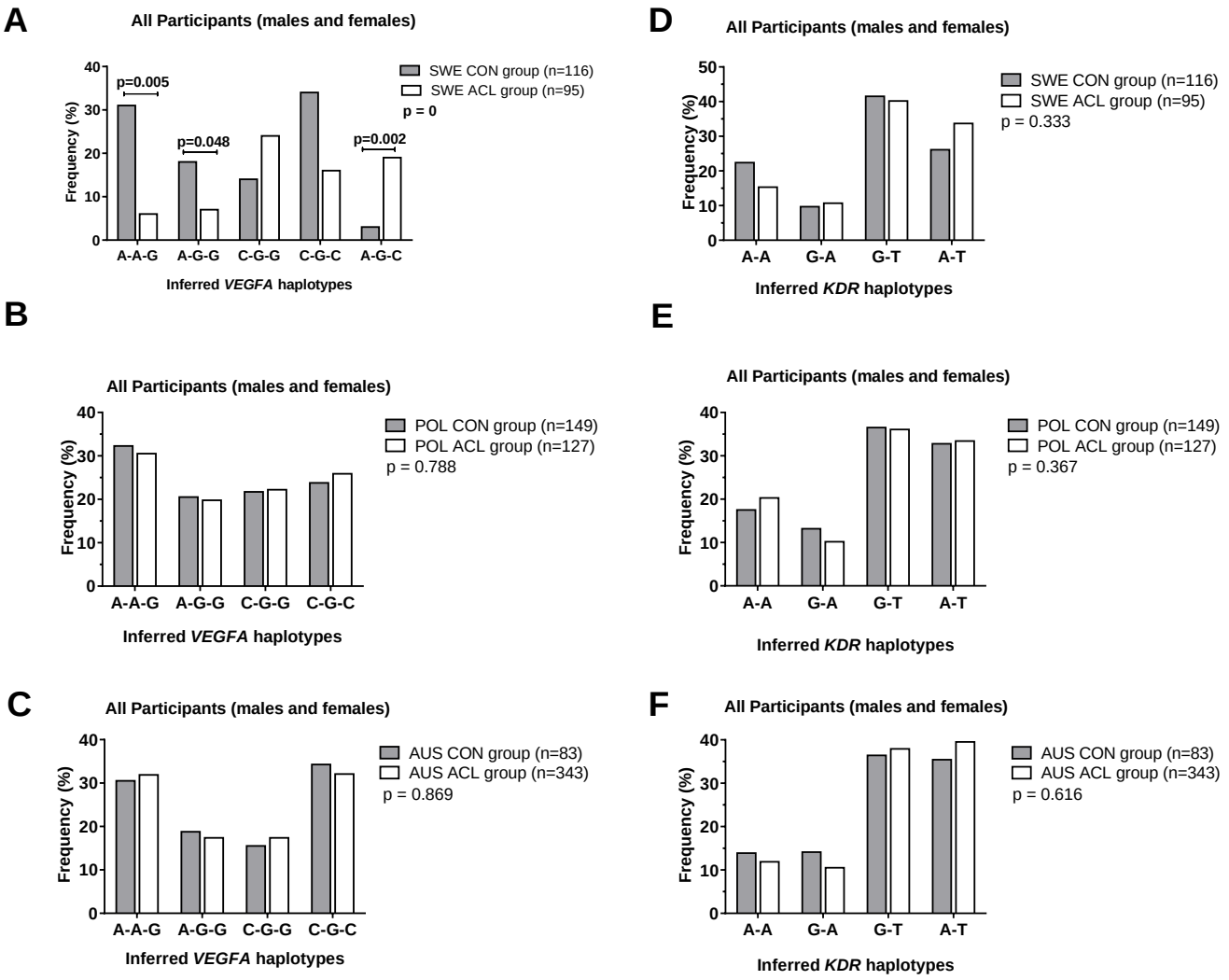
	POL CON	POL ACL	p-value[a]
***VEGFA* rs699947 (C/A)**			
n	*144*	*122*	
C/C	20.8 (30)	27.0 (33)	0.158
A/C	52.1 (75)	49.2 (60)	
A/A	27.1 (39)	23.8 (29)	
A allele	53 (153)	48 (118)	0.297
HWE	0.620	0.858	
***VEGFA* rs1570360 (G/A)**			
n	*148*	*117*	
G/G	46.6 (69)	49.6 (58)	0.497
G/A	39.2 (58)	39.3 (46)	
A/A	14.2 (21)	11.1 (13)	
A allele	34 (100)	31 (72)	0.513
HWE	0.142	0.391	
***VEGFA* rs2010963 (G/C)**			
n	*143*	*122*	
G/G	55.9 (80)	51.6 (63)	0.782
G/C	37.8 (54)	41.0 (50)	
C/C	6.3 (9)	7.4 (9)	
C allele	25 (72)	28 (68)	0.491
HWE	1.000	1.000	
***KDR* rs2071559 (A/G)**			
n	*142*	*125*	
G/G	27.5 (39)	22.4 (28)	0.915
A/G	44.4 (63)	48.0 (60)	
A/A	28.2 (40)	29.6 (37)	
A allele	50 (143)	54 (134)	0.488
HWE	0.182	0.720	
***KDR* rs1870377 (T/A)**			
n	*147*	*121*	
T/T	47.6 (70)	48.8 (59)	0.268
T/A	43.5 (64)	41.3 (50)	
A/A	8.8 (13)	9.9 (12)	
A allele	31 (90)	31 (74)	1.000
HWE	0.848	0.830	

Genotype and allele frequencies are expressed as a percentage with the number of participants (n) in parentheses.
CON vs. ACL[a] (adjusted p-values for age, body mass and type of sport).
P-values in bold typeset indicate significance ($p < 0.05$).

Supplementary Table 17: Genotype and minor allele frequency distributions, and p-values for Hardy-Weinberg exact test of the investigated polymorphisms in all participants (male and female) in the control (CON) and the anterior cruciate ligament (ACL) rupture group within the Australian cohort.

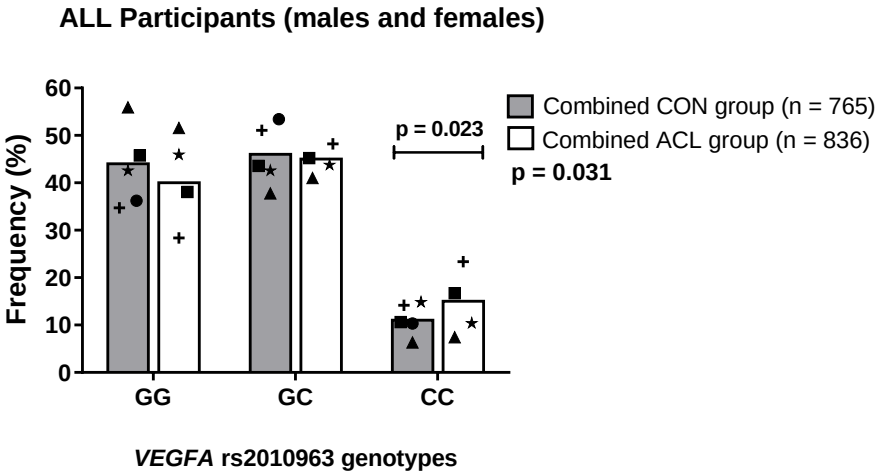
	AUS CON	AUS ACL	p-value[a]
***VEGFA* rs699947 (C/A)**			
n	*83*	*339*	
C/C	30 (25)	25 (86)	0.489
A/C	41 (34)	50 (171)	
A/A	29 (24)	24 (82)	
A allele	49 (82)	49 (335)	1.000
HWE	0.123	0.914	
***VEGFA* rs1570360 (G/A)**			
n	*83*	*336*	
G/G	48 (40)	46 (155)	0.776
G/A	41 (34)	42 (140)	
A/A	11 (9)	12 (41)	
A allele	31 (52)	33 (222)	0.712
HWE	0.619	0.323	
***VEGFA* rs2010963 (G/C)**			
n	*83*	*338*	
G/G	46 (38)	46 (155)	0.299
G/C	40 (33)	44 (148)	
C/C	14 (12)	10 (35)	
C allele	34 (57)	32 (218)	0.644
HWE	0.329	1.000	
***KDR* rs2071559 (A/G)**			
n	*83*	*339*	
G/G	30 (25)	25 (86)	0.895
A/G	41 (34)	50 (171)	
A/A	29 (24)	24 (82)	
A allele	49 (82)	49 (335)	1.000
HWE	0.123	0.914	
***KDR* rs1870377 (T/A)**			
n	*83*	*336*	
T/T	48 (40)	46 (155)	0.289
T/A	41 (34)	42 (140)	
A/A	11 (9)	12 (41)	
A allele	31 (52)	33 (222)	0.712
HWE	0.619	0.323	

Genotype and allele frequencies are expressed as a percentage with the number of participants (n) in parentheses.
CON vs. ACL[a] (adjusted p-values for age, and sex).
P-values in bold typeset indicate significance ($p < 0.05$).



See overleaf for figure legend.

Supplementary Figure 3: Inferred haplotype frequency distributions for *VEGFA* rs699947 C/A, rs1570360 G/A, rs2010963 G/C in the (A) Swedish, (B) Polish and (C) Australian cohorts and *KDR* rs2071559 A/G, rs1870377 T/A in the (D) Swedish, (E) Polish and (F) Australian cohorts in the control group (CON; grey bars) and the anterior cruciate ligament rupture group (ACL; white bars). Statistically significant differences in haplotype frequency between the groups are depicted on the graph. The number of participants (n) in each group is in parentheses.



Supplementary Figure 4: Genotype frequency distributions for the *VEGFA* rs2010963 (G/C) polymorphism in all participants in the combined control (CON) group (grey bars) and the combined anterior cruciate ligament rupture (ACL) group (white bars) with omission of the Swedish ACL group. Statistically significant differences in genotype frequencies between the combined CON and ACL groups are depicted on the graph. The chi-squared global p-value for comparison of the combined CON (n=765) and combined ACL (n=836) group is depicted in the figure.

SECTION 3

SUPPLEMENTARY DATA - CHAPTER 4

Supplementary Table 18: Genotype effects on participant characteristics in the combined cohort for *ACAN* (A: rs2351491, B: rs1042631 and C: rs1516797) *DCN (*D: rs516115) and *BGN* (E: rs1126499 males, F: rs1042103 males, G: rs1126499 females, H: rs1042103 females) polymorphisms.

A. Genotype effects *ACAN* rs2351491 C/T

	T/T	**C/T**	**C/C**	**p-value**
Age (years)	30.6 ± 12.1 (331)	31.8 ± 13.2 (378)	27.5 ± 10.3 (105)	**0.013**
Sex (% male)	60 (336)	62 (385)	67 (108)	0.445
Height (cm)	175.6 ± 10.9 (325)	175.5 ± 9.7 (369)	176.5 ± 9.6 (104)	0.620
Body mass (kg)	75.6 ± 14.7 (327)	74.8 ± 14.5 (373)	7.5 ± 13.9 (104)	0.533
Body mass index (kg/m^2)	24.6 ± 5.8 (322)	24.2 ± 3.5 (367)	24.5 ± 3.2 (103)	0.638

B. Genotype effects *ACAN* rs1042631 C/T

	T/T	**C/T**	**C/C**	**p-value**
Age (years)	29.9 ± 11.9 (36)	31.2 13.5 (275)	30.5 ± 11.9 506	0.619
Sex (% male)	58 (36)	65 (282)	61 (514)	0.455
Height (cm)	175.6 ± 9.3 (35)	175.7 ± 9.3 237	175.6 ± 10.8 493	0.994
Body mass (kg)	73.4 ± 10.3 (36)	74.5 ± 13.4 273	75.9 ± 15.3 498	0.495
Body mass index (kg/m^2)	23.8 ± 2.3 (35)	24.1 ± 3.1 270	24.6 ± 5.3 490	0.458

C. Genotype effects *ACAN* rs1516797 T/G

	T/T	**G/T**	**G/G**	**p-value**
Age (years)	30.5 ± 12.9 (358)	30.5 ± 11.9 (363)	32.1 ± 12.7 (85)	0.731
Sex (% male)	64 (527)	65 (532)	63 (134)	0.907
Height (cm)	176.2 ± 10.7 (357)	175.5 ± 9.7 (354)	174.0 ± 9.8 (79)	0.264
Body mass (kg)	75.4 ± 13.4 (359)	76.0 ± 15.7 (356)	71.9 ± 13.4 (81)	0.129
Body mass index (kg/m^2)	24.4 ± 5.4 (353)	24.6 ± 3.9 (352)	23.6 ± 2.8 (79)	0.334

D. Genotype effects *DCN* rs516115 T/C

	T/T	**C/T**	**C/C**	**p-value**
Age (years)	30.9 ± 12.5 (438)	30.9 ± 12.6 (310)	29.8 ± 11.7 (63)	0.183
Sex (% male)	60 (693)	63 (532)	60 (101)	0.432
Height (cm)	175.4 ± 10.4 (428)	176.1 ± 9.8 (304)	174.0 ± 10.9 (63)	0.080
Body mass (kg)	74.7 ± 1409 (429)	75.7 ± 1305 (310)	76.4 ± 16.7 (62)	0.454
Body mass index (kg/m^2)	24.4 ± 5.4 (425)	24.3 ± 3.0 (303)	24.0 ± 4.0 (61)	0.723

E. Genotype effects Males *BGN* rs1126499 C/T

	C/C	T/T	p-value
Age (years)	29.2 ± 11.9 (251)	29.3 ± 11.4 (238)	0.900
Height (cm)	181.7 ± 6.9 (250)	180.2 ± 9.3 (234)	**0.048**
Body mass (kg)	82.7 ± 12.6 (249)	82.0 ± 12.2 (236)	0.576
Body mass index (kg/m^2)	25.0 ± 3.6 (246)	25.5 ± 6.5 (233)	0.293

F. Genotype effects Males *BGN* rs1042103 G/A

	G/G	A/A	p-value
Age (years)	29.9 ± 12.1 (300)	27.9 ± 10.4 (197)	**0.049**
Height (cm)	181.4 ± 6.7 (294)	180.6 ± 9.9 (197)	0.274
Body mass (kg)	82.6 ± 11.3 (296)	82.1 ± 1401 (196)	0.378
Body mass index (kg/m^2)	25.1 ± 3.2 (291)	25.5 ± 7.2 (195)	0.399

G. Genotype effects Females *BGN* rs1126499 C/T

	T/T	C/T	C/C	p-value
Age (years)	33.5 ± 14.1 (68)	33.8 ± 13.3 (152)	31.6 ± 13.7 (82)	0.480
Height (cm)	166.9 ± 5.8 (65)	166.6 ± 5.8 (147)	166.3 ± 5.6 (80)	0.839
Body mass (kg)	63.2 ± 10.1 (67)	63.6 ± 8.0 (151)	63.7 ± 8.9 (79)	0.923
Body mass index (kg/m^2)	23.2 ± 2.9 (65)	22.9 ± 2.6 (147)	23.0 ± 3.2 (79)	0.824

H. Genotype effects Females *BGN* rs1042103 G/A

	G/G	A/G	A/A	p-value
Age (years)	32.9 ± 12.9 (111)	32.8 ± 13.4 (144	34.4 ± 15.3 (52)	0.760
Height (cm)	166.6 ± 5.6 (110)	166.2 ± 5.6 (135)	167.7 ± 6.7 (52)	0.289
Body mass (kg)	63.4 ± 8.1 (109)	63.1 ± 7.6 (141)	65.3 ± 11.9 (52)	0.293
Body mass index (kg/m^2)	22.9 ± 2.8 (109)	22.9 ± 2.7 (135)	23.6 ± 3.1 (52)	0.244

Supplementary Table 19: Genotype effects on participant characteristics in the Swedish cohort for *ACAN* (A: rs2351491, B: rs1042631 and C: rs1516797) *DCN (*D: rs516115) and *BGN* (E: rs1126499 males, F: rs1042103 males, G: rs1126499 females, H: rs1042103 females) polymorphisms.

A. Genotype effects *ACAN* rs2351491 C/T

	T/T	**C/T**	**C/C**	**p-value**
Age (years)	40.4 ± 12.7 (84)	43.1 ± 13.2 (98)	36.6 ± 12.9 (23)	0.067
Sex (% male)	39 (84)	41 (84)	42 (84)	0.997
Height (cm)	172.3 ± 9.7 (81)	172.4 ± 9.3 (89)	173.7 ± 9.3 (21)	0.817
Body mass (kg)	73.6 ± 13.6 (80)	72.4 ± 13.3 (90)	74.1 ± 14.0 (21)	0.743
Body mass index (kg/m^2)	24.6 ± 2.9 (80)	24.5 ± 2.8 (88)	24.4 ± 3.5 (21)	0.967

B. Genotype effects *ACAN* rs1042631 C/T

	T/T	**C/T**	**C/C**	**p-value**
Age (years)	38.6 ± 14.9 8)	42.8 ± 13.3 77)	40.5 ± 12.9 (122)	0.406
Sex (% male)	38 (8)	39 (79)	41 (124)	0.925
Height (cm)	173.3 ± 10.4 (8)	171.8 ± 8.2 (72)	172.9 ± 10.1 (113)	0.652
Body mass (kg)	72.8 ± 8.9 (8)	71.3 ± 14.1 (73)	74.1 ± 13.2 (112)	0.772
Body mass index (kg/m^2)	24.3 ± 2.7 (8)	24.3 ± 2.9 (71)	24.6 ± 2.9 (112)	0.954

C. Genotype effects *ACAN* rs1516797 T/G

	T/T	**G/T**	**G/G**	**p-value**
Age (years)	42.4 ± 13.7 (76)	40.9 ± 13.2 (91)	38.9 ± 12.2 (33)	0.453
Sex (% male)	38 (78)	45 (93)	33 (33)	0.438
Height (cm)	172.4 ± 9.1 (74)	173.4 ± 9.1 (84)	170.8 ± 10.9 (28)	0.448
Body mass (kg)	72.3 ± 12.9 (74)	75.0 ± 13.6 (83)	68.7 ± 11.9 (29)	0.066
Body mass index (kg/m^2)	24.5 ± 2.9 (73)	27.8 ± 3.0 (83)	23.5 ± 2.7 (28)	0.120

D. Genotype effects *DCN* rs516115 T/C

	T/T	**C/T**	**C/C**	**p-value**
Age (years)	40.8 ± 13.0 (129)	42.5 ± 13.6 (65)	40.3 ± 11.8 (12)	0.684
Sex (% male)	38 (132)	50 (66)	8 (12)	**0.018**
Height (cm)	172.1 ± 9.0 (119)	174.7 ± 9.5 (61)	164.0 ± 6.7 (12)	**0.001**
Body mass (kg)	71.6 ±13.9 (119)	76.5 ± 11.3 (62)	65.1 ± 7.7 (11)	**0.006**
Body mass index (kg/m^2)	24.2 ± 2.9 (118)	24.9 ± 2.7 (61)	24.6 ± 2.5 (11)	0.255

E. Genotype effects Males *BGN* rs1126499 C/T

	C/C	**T/T**	**p-value**
Age (years)	42.8 ± 11.5 (42)	42.4 ± 12.5 (39)	0.874
Height (cm)	181.7 ± 6.7 (42)	179.5 ± 6.4 (33)	0.156
Body mass (kg)	86.2 ± 10.3 (41)	81.6 ±7.5 (34)	**0.033**
Body mass index (kg/m^2)	26.1 ± 2.8 (41)	25.4 ± 2.7 (33)	0.247

F. Genotype effects Males *BGN* rs1042103 G/A

	G/G	A/A	p-value
Age (years)	42.8 ± 12.1 (56)	40.9 ±11.9 (23)	0.737
Height (cm)	181.1 ± 6.9 (52)	180.6 ± 5.3 (20)	0.796
Body mass (kg)	84.3 ± 9.7 (53)	82.6 ± 8.7 (19)	0.741
Body mass index (kg/m^2)	25.8 ± 3.0 (52)	25.3 ± 2.1 (19)	0.850

G. Genotype effects Females *BGN* rs1126499 C/T

	T/T	C/T	C/C	p-value
Age (years)	45.2 ± 13.6 (23)	39.6 ± 16.6 (65)	38.7 ± 14.6 (31)	0.180
Height (cm)	167.8 ± 5.7 (22)	167.2 ± 5.7 (60)	165.2 ± 5.0 (29)	0.176
Body mass (kg)	64.9 ± 11.5 (22)	65.5 ± 8.4 (62)	65.6 ± 10.0 (28)	0.958
Body mass index (kg/m^2)	24.1 ± 2.9 (22)	23.4 ± 2.4 (60)	24.0 ± 3.2 (28)	0.519

H. Genotype effects Females *BGN* rs1042103 G/A

	G/G	A/G	A/A	p-value
Age (years)	38.7 ± 11.9 (46)	41.3 ± 14.6 (53)	42.2 ± 15.9 (21)	0.522
Height (cm)	167.0 ± 5.2 (45)	166.4 ± 5.3 (46)	167.9 ± 6.9 (21)	0.996
Body mass (kg)	65.5 ± 8.5 (44)	65.6 ± 7.6 (48)	65.5 ± 13.9 (21)	0.472
Body mass index (kg/m^2)	23.4 ± 2.5 (44)	23.8 ± 2.7 (46)	24.3 ± 3.2 (21)	0.571

Supplementary Table 20: Genotype effects on participant characteristics in the Polish cohort for *ACAN* (A: rs2351491, B: rs1042631 and C: rs1516797) *DCN (*D: rs516115) and *BGN* (E: rs1126499 males, F: rs1042103 males, G: rs1126499 females, H: rs1042103 females) polymorphisms.

A. Genotype effects *ACAN* rs2351491 C/T

	T/T	**C/T**	**C/C**	**p-value**
Age (years)	26 ± 8.5 (113)	26.0 ± 9.3 (119)	24.8 ± 8.2 (37)	0.741
Sex (% male)	74 (113)	72 (119)	78 (37)	0.758
Height (cm)	177.7 ± 12.9 (111)	176.6 ± 10.0 (118)	178.6 ± 8.8 (37)	0.693
Body mass (kg)	76.2 ±13.1 (113)	74.8 ±14.7 (119)	76.2 ± 12.9 (36)	0.653
Body mass index (kg/m^2)	24.6 ± 8.5 (111)	23.9 ± 3.9 (118)	23.9 ± 3.1 (36)	0.596

B. Genotype effects *ACAN* rs1042631 C/T

	T/T	**C/T**	**C/C**	**p-value**
Age (years)	27.7 ± 10.9 (16)	25.3 ± 9.0 (94)	25.9 ± 8.2 (159)	0.569
Sex (% male)	69 (16)	73 (94)	75 (159)	0.860
Height (cm)	176.9 ± 9.2 (16)	176.6 ± 9.2 (94)	177.8 ± 12.5 (156)	0.236
Body mass (kg)	74.4 ±10.5 (16)	73.7 ± 11.9 (93)	76.8 ± 15.1 (159)	0.504
Body mass index (kg/m^2)	23.7± 1.9 (16)	23.6 ± 2.9 (93)	24.5 ± 7.8 (156)	0.695

C. Genotype effects *ACAN* rs1516797 T/G

	T/T	**G/T**	**G/G**	**p-value**
Age (years)	25.6 ± 9.2 (116)	26.0 ± 8.5 (123)	26.4 ± 7.5 (26)	0.893
Sex (% male)	8 (116)	70 (123)	65 (26)	0.114
Height (cm)	178.3 ± 12.6 (115)	176.6 ± 10.2 (121)	176.1 ± 8.9 (26)	0.523
Body mass (kg)	76.5 ± 12.0 (115)	75.4 ± 15.5 (123)	73.2 ± 13.6 (26)	0.754
Body mass index (kg/m^2)	24.5 ± 8.4 (114)	24.1 ± 4.1 (121)	23.5 ± 3.3 (26)	0.414

D. Genotype effects *DCN* rs516115 T/C

	T/T	**C/T**	**C/C**	**p-value**
Age (years)	25.5 ± 8.3 (153)	26.9 ± 9.3 (95)	24.1 ± 7.9 (15)	0.302
Sex (% male)	73 (153)	73 (95)	80 (15)	0.822
Height (cm)	177.1 ± 12.3 (152)	177.5 ± 9.8 (93)	175.5 ± 8.2 (15)	0.707
Body mass (kg)	75.7 ± 15.2 (152)	74.7 ± 11.7 (95)	77.6 ± 14.6 (15)	0.549
Body mass index (kg/m^2)	24.4 ± 7.9 (151)	23.6 ± 2.6 (93)	25.1 ± 3.7 (15)	0.825

E. Genotype effects Males *BGN* rs1126499 C/T

	C/C	**T/T**	**p-value**
Age (years)	24.3 ± 6.9 (96)	25.9 ± 8.4 (89)	0.154
Height (cm)	182.1 ± 7.5 (95)	179.8 ± 12.5 (88)	0.125
Body mass (kg)	80.6 ± 13.5 (96)	79.4 ± 10.3 (88)	0.501
Body mass index (kg/m^2)	24.3 ± 4.0 (95)	25.3 ± 9.4 (87)	0.356

F. Genotype effects Males *BGN* rs1042103 G/A

	G/G	A/A	p-value
Age (years)	24.6 ± 7.2 (112)	26.0 ± 8.4 (83)	0.195
Height (cm)	181.8 ± 7.2 (110)	180.4 ± 13.1 (83)	0.331
Body mass (kg)	79.7 ± 9.8 (111)	81.3 ± 15.2 (83)	0.370
Body mass index (kg/m^2)	24.1 ± 2.5 (109)	25.7 ± 10.2 (83)	0.114

G. Genotype effects Females *BGN* rs1126499 C/T

	T/T	C/T	C/C	p-value
Age (years)	26.4 ± 12.7 (16)	27.3 ± 8.9 (28)	28.3 ± 11.8 (24)	0.858
Height (cm)	167.0 ± 6.2 (16)	166.0 ± 5.7 (27)	167.0 ± 5.7 (24)	0.779
Body mass (kg)	60.5 ± 6.9 (16)	61.9 ± 5.9 (28)	63.3 ± 8.9 (24)	0.494
Body mass index (kg/m^2)	21.7 ± 2.4 (16)	22.4 ± 2.4 (27)	22.8 ± 3.6 (24)	0.532

H. Genotype effects Females *BGN* rs1042103 G/A

	G/G	A/G	A/A	p-value
Age (years)	28.3 ± 12.6 (23)	26.6 ± 8.6 (33)	28.4 ± 12.9 (15)	0.801
Height (cm)	165.9 ± 5.9 (23)	166.2 ± 5.8 (32)	168.0 ± 6.5 (15)	0.539
Body mass (kg)	60.1 ± 6.8 (23)	62.0 ± 5.9 (33)	64.9 ± 9.4 (15)	0.120
Body mass index (kg/m^2)	21.9 ± 2.8 (23)	22.4 ± 2.6 (32)	23.1 ± 3.4 (15)	0.465

Supplementary Table 21: Minor allele frequency distributions for the *ACAN* rsrs2351491 (C/T), rs1042631 (C/T), rs1516797 (T/G), and *DCN* rs516115 (T/C) polymorphisms in all participants (males and females), and the *BGN* rs1126499 (C/T) and rs1042103 (G/A) polymorphisms in male and female participants separately in the control (CON) and the anterior cruciate ligament (ACL) rupture group within the combined cohort.

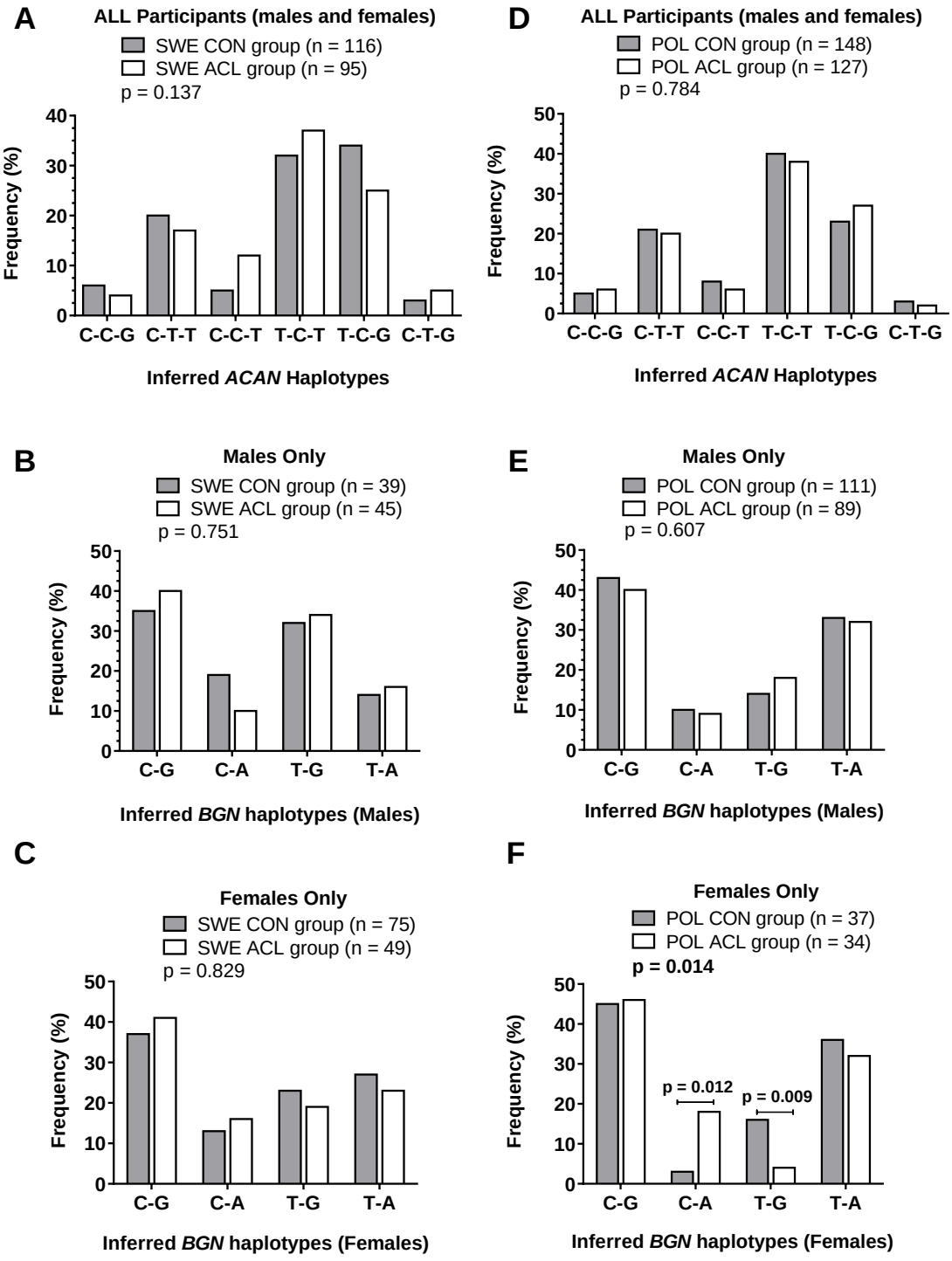
	COMB CON	COMB ACL	p-value	COMB Male CON	COMB Male ACL	p-value	COMB Female CON	COMB Female ACL	p-value
***ACAN* rs2351491 (C/T)**									
C allele	35 (342)	38 (259)	0.378	37 (209)	38 (175)	0.649	33 (133)	36 (84)	0.488
T allele	65 (626)	62 (431)		63 (361)	62 (283)		67 (265)	64 (148)	
***ACAN* rs1042631 (C/T)**									
C allele	79 (774)	78 (536)	0.395	79 (452)	78 (355)	0.649	80 (322)	78 (181)	0.474
T allele	21 (200)	22 (154)		21 (122)	22 (103)		22 (78)	22 (51)	
***ACAN* rs1516797 (T/G)**									
T allele	67 (834)	66 (752)	0.544	67 (515)	65 (505)	0.418	66 (319)	67 (247)	0.942
G allele	33 (410)	34 (390)		33 (249)	35 (267)		34 (161)	33 (123)	
***DCN* rs516115 (T/C)**									
T allele	72 (1081)	73 (837)	0.484	72 (614)	72 (551)	0.912	72 (467)	76 (286)	0.125
C allele	28 (425)	27 (309)		28 (240)	28 (219)		28 (185)	24 (90)	
***BGN* rs1126499 (C/T)**									
C allele	-	-	-	53 (400)	49 (370)	0.099	51 (246)	56 (207)	0.165
T allele	-	-	-	47 (352)	51 (386)		49 (234)	44 (161)	
***BGN* rs1042103 (G/A)**									
G allele	-	-	-	58 (494)	57 (434)	0.879	61 (395)	60 (223)	0.894
A allele	-	-	-	42 (364)	43 (326)		39 (255)	40 (147)	

Allele frequencies are expressed as a percentage with the number of participants (n) in parentheses.
P-values in bold typeset indicate significance ($p < 0.05$).

Supplementary Table 22: Genotype and minor allele frequency distributions, and p-values for Hardy-Weinberg exact test of the investigated *ACAN* and *DCN* polymorphisms in all participants (males and females), and *BGN* polymorphisms in male and female participants separately in the control (CON) and the anterior cruciate ligament (ACL) rupture group within the Swedish and Polish cohorts.

		SWE CON	**SWE ACL**	**P-Value**	**POL CON**	**POL ACL**	**p-value**
ALL							
***ACAN* rs2351491 (C/T)**	*n*	*114*	*95*		*144*	*125*	
	T/T	43 (49)	38 (36)	0.646[a]	42 (60)	42 (53)	0.549
	C/T	46 (53)	49 (47)		43 (62)	46 (57)	
	C/C	11 (12)	13 (12)		15 (22)	12 (15)	
	C allele	66 (151)	63 (119)	0.473	37 (106)	35 (87)	0.653
	HWE	0.834	0.664		0.374	1.000	
***ACAN* rs1042631 (C/T)**	*n*	*116*	*95*		*145*	*124*	
	C/C	57 (66)	61 (58)	0.887[a]	60 (87)	58 (72)	**0.010**[b] (**0.040**)[d]
	C/T	40 (46)	35 (33)		32 (47)	38(47)	
	T/T	3 (4)	4 (4)		8 (11)	4 (5)	
	T allele	23 (54)	22 (41)	0.726	24 (69)	23 (57)	0.839
	HWE	0.305	1.000		0.249	0.612	
***ACAN* rs1516797 (T/G)**	*n*	*112*	*92*		*141*	*124*	
	T/T	35 (39)	42 (39)	0.059[a]	45 (64)	42 (52)	0.738[b]
	G/T	45 (50)	47 (43)		47 (66)	46 (57)	
	G/G	21 (23)	11 (10)		8 (11)	12 (15)	
	G allele	43 (96)	34 (63)	0.083	31 (88)	35 (87)	0.356
	HWE	0.341	0.819		0.332	1.000	
***DCN* rs516115 (T/C)**	*n*	*115*	*95*		*140*	*123*	
	T/T	63 (73)	62 (59)	0.144[a]	60 (84)	56 (69)	0.799[b]
	T/C	29 (33)	35 (33)		34 (48)	38 (47)	
	C/C	08 (9)	03 (3)		6 (8)	6 (7)	
	C allele	22 (51)	21 (39)	0.721	23 (64)	25 (61)	0.609
	HWE	0.099	0.754		0.810	1.000	
MALES							
***BGN* rs1126499 (C/T)**	*n*	*37*	*44*		*104*	*82*	
	C/C	54(20)	50(22)	0.882[c]	52 (56)	49(40)	0.732[b]
	C/T	-	-		-	-	
	T/T	46(17)	50(22)		46(47)	51(42)	
	T allele	46(34)	50(44)	0.638	46(94)	51(84)	0.297
	HWE	-	-		-	-	
***BGN* rs1042103 (G/A)**	*n*	*37*	*42*		*109*	*86*	
	G/G	68(25)	74(31)	0.366[c]	57(62)	58(50)	0.273[b]
	A/G	-	-		-	-	
	A/A	32(12)	26(11)		43(47)	42(36)	
	A allele	32(24)	26(22)	0.483	43(94)	42(72)	0.837
	HWE	-	-		-	-	
FEMALES							
***BGN* rs1126499 (C/T)**	*n*	*74*	*48*		*35*	*33*	
	C/C	23(17)	29(14)	0.892[a]	26 (9)	46 (15)	0.124[b]
	C/T	54(40)	56(27)		46 (16)	36 (12)	
	T/T	23(17)	15(7)		29 (10)	18 (6)	
	T allele	50(74)	43(41)	0.295	51 (36)	36 (24)	0.086
	HWE	0.642	0.386		0.735	0.261	
***BGN* rs1042103 (G/A)**	*n*	*75*	*48*		*37*	*34*	
	G/G	41 (31)	33(16)	0.062[a]	38 (14)	27 (9)	0.242[b]
	A/G	37(28)	54(26)		46 (17)	47 (16)	
	A/A	21(16)	12(6)		16 (6)	27 (9)	
	A allele	40(60)	40(38)	1.000	39 (29)	50 (34)	0.237
	HWE	0.057	0.545		1.000	0.739	

Genotype and allele frequencies are expressed as a percentage with the number of participants (n) in parentheses
CON vs. ACL[a] (adjusted p-values for age and type of sport)
CON vs. ACL[b] (adjusted p-values for age, body mass and type of sport)
CON vs. ACL[c] (adjusted p-values for height and type of sport)
CON vs. ACL[d] (False Discovery Rate, Benjamini-Hochberg)
P-values in bold typeset indicate significance ($P < 0.05$)



Supplementary Figure 5: Inferred haplotype frequency distributions for the *ACAN* (rs2351491 C/T-rs1042631 T/C-rs1516797 T/G) and *BGN* (rs1126499 C/T-rs1042103 G/A) polymorphisms in the control group (CON; grey bars) and the anterior cruciate ligament rupture group (ACL; white bars) for the A to C: Swedish and D to E: Polish cohorts. Statistically significant differences in haplotype frequency between the groups are depicted on the graph. The number of participants (n) in each group is in parentheses.