Genetic Susceptibility to ACL Injury: The Role of Collagen Gene Variants in Elite Indian Athletes

Khalid Khader[1$] (ORCID 0000-0003-1092-1498), Shahrumi Reza[2,3$], Tiyasha De[2], Alia Hashmi[4], Deepak Joshi[1] (ORCID 0009-0002-1659-2488), Nitin Mehta[1] (ORCID 0009-0007-9021-6700), Vaisakhan SR[1] (ORCID 0009-0004-5051-3274), Ildus I. Ahmetov[5,6] (ORCID 0000-0002-6335-4020), Mohammed Faruq[2*] (ORCID 0000-0001-8278-8396)

1. Sports Injury Centre, New Delhi
2. Genomics and Molecular Medicine, CSIR-Institute of Genomics and Integrative Biology (CSIR -IGIB), Mall Road, Delhi 110007, India
3. Academy of Scientific and Innovative Research (AcSIR), Sector 19, Kamla Nehru Nagar, Ghaziabad, Uttar Pradesh 201002, India.
4. Jamia Millia Islamia, Delhi
5. Laboratory of Genetics of Aging and Longevity, Kazan State Medical University, 420111 Kazan, Russia
6. Research Institute for Sport and Exercise Sciences, Liverpool John Moores University, Liverpool L3 5AF, United Kingdom

$ equal contribution

* Corresponding Author

Address of Correspondence

Dr. Mohammed Faruq
CSIR-Institute of Genomics and Integrative Biology
Mall Road, Delhi, India 110007

Abstract

Anterior cruciate ligament (ACL) injury is among the most common and disabling conditions in sports, often leading to long recovery periods and an increased risk of re-injury. While biomechanical and environmental factors are well recognized, genetic variation is increasingly understood to influence ligament structure and resilience. Collagen genes are of particular interest as they encode the key structural proteins that maintain ligament integrity. However, data on genetic susceptibility to ACL injury among Indian athletes remain limited. We conducted a case-control genetic association study involving 180 elite Indian athletes, including 90 clinically and MRI-confirmed non-contact ACL injuries and 90 controls with no prior history of knee injury, verified through clinical examination. Genotyping was performed using the Illumina Infinium Global Screening Array-24 v3.0 platform, followed by rigorous quality control and population structure analysis. Four functionally relevant collagen gene variants: *COL1A1* (rs1800012), *COL3A1* (rs1800255), *COL5A1* (rs12722), and *COL12A1* (rs970547) were examined using logistic regression under multiple genetic models, adjusting for age and sex. Variants in *COL3A1* (rs1800255 A allele) and *COL1A1* (rs1800012 A allele) showed significant association with increased risk of ACL injury, while the *COL5A1* (rs12722 C allele) appeared to confer a protective effect. The *COL12A1* (rs970547) variant showed no association. This study provides genetic evidence from Indian elite athletes suggesting that variation in collagen genes associates with ACL injury risk. The findings highlight the contribution of collagen fibrillogenesis to ligament strength and support the future use of genetic markers in individualized injury prevention and rehabilitation strategies.

Keywords

ACL injury; collagen genes; genetic susceptibility; Indian athletes; single-nucleotide polymorphism.

1. Introduction

Anterior cruciate ligament (ACL) injury is among the most frequent and debilitating musculoskeletal problems in sports medicine. Such injuries often require surgical reconstruction and prolonged rehabilitation, and many athletes continue to experience long-term consequences such as early-onset osteoarthritis, even with optimal surgical care and rehabilitation. Non-contact ACL ruptures are particularly concerning because they occur without direct impact, typically resulting from complex interactions among intrinsic and extrinsic factors, including joint biomechanics, neuromuscular control, and environmental conditions [1].

Over the past decade, increasing attention has turned to the role of genetic factors in determining ligament strength and susceptibility to injury. Evidence from twin and family-based studies supports the notion that soft-tissue and ligament injuries have a heritable component, indicating that subtle genetic variation may influence the remodelling capacity, tensile strength, and healing potential of connective tissue, with heritability for ACL rupture estimated at approximately 69% [2]. In this context, genetic profiling offers a promising avenue for understanding why some individuals develop ligament failure despite similar training loads, biomechanics, and external risk exposure.

Collagen-encoding genes have been a major focus of research because collagens represent the primary structural elements of ligaments and provide essential tensile support to withstand high mechanical demands. Type I collagen is the predominant fibrillar protein of the ACL, while types III, V, and XII collagens contribute to fibril organization, cross-linking, and repair during tissue remodelling [3]. Variants that alter the expression or structure of these proteins may disrupt fibrillogenesis or extracellular matrix assembly, weakening the ligament and increasing vulnerability to overload-related failure. Among the key loci investigated, *COL1A1, COL3A1, COL5A1*, and *COL12A1* have consistently emerged in studies of ligament and tendon pathology [4,5]. The *COL1A1* rs1800012 variant affects a Sp1 transcription-factor motif, potentially altering collagen cross-linking and tensile efficiency [6]. The *COL3A1* rs1800255 variant has been implicated in altered tissue remodelling responses to repetitive stress [7], whereas *COL5A1* rs12722 influences mRNA stability and expression of type V collagen, which regulates fibril diameter and packing density [8]. *COL12A1* rs970547, encoding a fibril-associated collagen, may modulate collagen-proteoglycan interactions, affecting the ligament's capacity to adapt to tensile strain [9].

Despite considerable global research efforts, findings regarding collagen gene polymorphisms and injury risk have been inconsistent. Such variability likely reflects population-specific allele frequencies, genetic background, sport-specific biomechanics, training loads, and environmental exposures [10]. Indian athletes remain particularly underrepresented, even though India comprises one of the most genetically diverse populations worldwide and has a rapidly evolving athletic ecosystem spanning sports with distinct biomechanical demands. Understanding the genetic architecture of ACL injury in Indian athletes is, therefore, both timely and necessary. Identifying relevant variants in this population may provide insight into the biological underpinnings of ligament vulnerability and support the development of

precision-based injury-risk profiling, individualized rehabilitation, and sport-specific conditioning strategies.

In this study, we investigated four well-characterized collagen gene variants *COL1A1* (rs1800012), *COL3A1* (rs1800255), *COL5A1* (rs12722), and *COL12A1* (rs970547) for their association with non-contact ACL injury in elite Indian athletes [11-13]. Although genome-wide genotyping data were generated, our analysis focused on these biologically plausible markers previously implicated in ligament and soft-tissue pathology. By extracting and examining these SNPs from high-density array data, this study provides initial genetic evidence in an underrepresented Indian athletic cohort and establishes a foundation for future large-scale genetic and functional studies in Indian sports medicine.

2. Materials and Methods

2.1. Study Design and Cohort Description

This hospital-based, case-control study was conducted between January 2024 and May 2025. Participant recruitment and clinical evaluation were carried out at the Sports Injury Centre, Vardhman Mahavir Medical College and Safdarjung Hospital, New Delhi. All molecular and genotyping analyses were performed at the CSIR-Institute of Genomics and Integrative Biology (CSIR-IGIB), New Delhi. The study protocol was approved by the Institutional Ethics Committee (IEC/VMMC/SJH/Cert/04/2024-307), and written informed consent was obtained from all participants before inclusion.

Initially, a total of 224 athletes aged 18-45 years were enrolled, comprising 112 individuals with isolated, non-contact anterior cruciate ligament (ACL) tears confirmed clinically and by magnetic resonance imaging (MRI), and 112 uninjured controls. The control group included athletes with no previous history of knee injury or major musculoskeletal injury and intact ligament structure verified by clinical examination. They were designed as an uninjured comparison group rather than a cohort with alternative injury phenotypes. Therefore, injury-type stratification between ACL and non-ACL injury categories was not applicable in the present study. Each participant completed a structured questionnaire recording demographic data, sports participation history, and mechanism of injury. Individuals with multi-ligament injuries, degenerative joint changes, prior knee surgery, or systemic disorders affecting connective tissue metabolism were excluded. Controls were excluded if they had any musculoskeletal injury or chronic systemic illness. These inclusion and exclusion criteria ensured a clinically homogeneous cohort suitable for genetic association analysis.

The final cohort (90 ACL-injured and 90 non-injured athletes) who passed quality control (described below) represented a broad range of sports commonly played at regional and national levels. The study included athletes from pivoting and contact sports such as football, kabaddi, wrestling, basketball, hockey, and cricket, as well as combat and strength-related disciplines including boxing, judo, taekwondo, and wushu. Additional representation came from athletics and other sport disciplines, reflecting the diversity of the study population. This distribution included several sports associated with substantial biomechanical loading of the knee joint and repetitive movement demands. Nearly all participants were national-level athletes, and approximately 3% had represented India in international competitions (Table 1).

2.2. Sample Collection and DNA Extraction

Peripheral blood samples were collected from all participants using EDTA-coated vacutainers. Genomic DNA was isolated using a modified salting-out method. DNA concentration and purity were measured spectrophotometrically (A260/A280 ratio), and integrity was verified through agarose gel electrophoresis. Of the 224 DNA samples collected, 204 (108 ACL-injured cases and 96 controls) met quality standards and were processed for genotyping.

2.3. Genotyping and Primary Data Processing

Genotyping was performed using the Illumina Infinium Global Screening Array-24 v3.0 on the Illumina iScan platform, which assays approximately 650,000 single-nucleotide polymorphisms (SNPs) distributed across the genome. This array includes well-curated variants from ClinVar, the GWAS Catalog, PharmGKB, and PharmVar, enabling comprehensive coverage of clinically relevant loci. Raw intensity data were imported into GenomeStudio v2.0 for genotype clustering and calling according to Illumina's recommended practices.

Of the 204 genotyped samples, 195 achieved a genotype call rate ≥ 0.95 and were retained for downstream analysis. Variants with a GenTrain score ≥ 0.7 were selected, yielding 612,322 high-quality SNPs for subsequent quality control. Genotype data were exported in PLINK format for statistical and population-based analyses.

2.4. Quality Control and Population Structure Analysis

Quality control was performed using PLINK v1.9. Samples and SNPs with >10% missingness were excluded, and variants deviating from Hardy-Weinberg equilibrium ($p < 1 \times 10^{-6}$) were removed. Strand inconsistencies were corrected, and ambiguous SNPs were filtered using Snpflip. Related samples were identified through identity-by-descent (IBD) analysis and excluded to ensure independence of observations. After filtering, 575,979 high-quality variants across 180 unrelated individuals (90 ACL-injured and 90 non-injured athletes) were retained for analysis. To evaluate population structure, principal component analysis (PCA) was performed in PLINK v1.9. The first two principal components were visualized using the ggplot2 package in R version 4.3.0.

2.5. Genome-Wide Association Analysis

To explore broader genetic contributions, a genome-wide association study (GWAS) was conducted using the same dataset. Logistic regression was performed in PLINK v1.9, adjusting for age, sex, and the top principal components to control for potential confounding and population stratification. The analysis included only high-quality variants that passed all filtering criteria described above.

2.6. Targeted Variant Selection

Although genome-wide genotyping data were generated, this study employed a hypothesis-driven candidate variant approach focusing on four collagen gene polymorphisms previously reported to be associated with ligament structure, connective tissue integrity, and ACL injury

susceptibility across multiple populations: *COL3A1* rs1800255 (chromosome 2), *COL12A1* rs970547 (chromosome 6), *COL5A1* rs12722 (chromosome 9), and *COL1A1* rs1800012 (chromosome 17). These variants were selected based on prior literature reporting their functional relevance in collagen fibrillogenesis, extracellular matrix organization, and soft-tissue biomechanics. The selected variants were extracted from the genome-wide dataset for targeted association testing. After all filtering steps, 180 participants (90 ACL-injured and 90 non-injured athletes) were included in the final analysis.

2.7. Statistical Analysis

Allele and genotype frequencies of the four collagen gene variants were compared between ACL-injured athletes and controls under allelic, genotypic, dominant, recessive, and trend models. Association analyses were performed using logistic regression in PLINK1.9 with age and sex included as covariates. An additive genetic model was applied, in which genotypes were coded as 0, 1, or 2 according to the number of effect alleles carried. Odds ratios (ORs) and 95% confidence intervals (CIs) were estimated for all analyses. Multiple testing correction for the primary association analyses was performed using the Benjamini-Hochberg false discovery rate (FDR) method, with $p < 0.05$ considered statistically significant. To explore possible sex-specific genetic effects, separate logistic regression analyses were performed in males and females using the same additive genetic model. Effect alleles were kept consistent across all analyses to allow direct comparison of the direction of associations between the two groups. To provide population-level context, allele frequencies observed in ACL-injured athletes were compared with reference datasets from the 1000 Genomes Project (global and South Asian populations) and the IndiGen dataset (India's national repository cataloging over 55 million genetic variants from 1,029 whole genome sequences). Pairwise SNP-SNP interaction analyses were also performed to assess potential epistatic effects among the four loci. Collectively, these analyses enabled a comprehensive evaluation of the contribution of collagen gene polymorphisms to ACL injury susceptibility among Indian athletes.

3. Results

3.1. Demographic and Clinical Characteristics

The final analysis included 180 elite Indian athletes, comprising 90 individuals with non-contact ACL injuries and 90 uninjured controls. Demographic and sports characteristics of the study population are summarized in Table 1. The median age was 23 years in both groups, with comparable age ranges observed between ACL-injured athletes and controls. The ACL group included 63 males and 27 females, while the control group included 53 males and 37 females. Anthropometric characteristics, including height, weight, and BMI, were broadly similar between the two groups.

3.2. Population Structure and Relatedness

Principal component analysis (PCA) showed that ACL-injured and control athletes were genetically comparable, with no evidence of population stratification between groups. The first two principal components captured most of the genetic variation within the cohort (PC1:

78.08%; PC2: 12.73%), together explaining 90.8% of the total variance, indicating that the major axes of genetic variation were effectively represented. (Supplementary Figure S1).

To minimize the effect of hidden relatedness, only individuals with an identity-by-descent (IBD) value ≤0.25 were retained for downstream analysis. This ensured that subsequent association testing was not confounded by ancestry or sample overlap.

3.3. Genome-Wide Association Study (GWAS)

A genome-wide association analysis was conducted to identify variants associated with non-contact ACL injury susceptibility. After quality control, 575,979 high-confidence SNPs were analysed in 90 ACL-injured and 90 control athletes. Three loci exceeded the suggestive significance threshold ($p < 1 \times 10^{-5}$), mapping near the *UMAD1/GLCCI1-DT*, *GZMM/BSG*, and *ZDHHC14/SNX9* regions. We assessed loci showing moderate association signals (P ~10^{-4}-10^{-5}) to evaluate their potential biological relevance. They mapped near genes involved in extracellular matrix organization and connective tissue biology, including collagen-related genes (*COL6A5, COL6A6, COL4A1,* and *COL23A1*), elastic fibre and matrix organization genes (*FBN1, FBLN2*), integrin family genes (*ITGA8, ITGA9*), extracellular matrix remodelling regulators (*ADAMTS9, TNC*), and genes implicated in tissue repair and signalling pathways (*VEGFA, BMP7, PDGFC,* and *WNT5A*) (Supplementary Table 1). However, none reached genome-wide significance and therefore were not prioritized for targeted association analysis (Supplementary Figure S2). We further examined previously reported ACL injury-associated variants identified in other populations to assess their relevance in the present cohort. Most of these variants did not demonstrate a significant association in our GWAS study (Supplementary Table 2).

3.4. Targeted Analysis of Collagen Gene Variants

We analysed four candidate collagen gene variant**s**, previously implicated in connective-tissue integrity and soft-tissue injury susceptibility: *COL3A1* rs1800255, *COL1A1* rs1800012, *COL5A1* rs12722, and *COL12A1* rs970547. Genotype and allele frequencies were compared between ACL-injured athletes and non-injured controls using chi-square tests. Frequency distributions are shown in Figure 1, with complete data provided in Supplementary Tables 3 and 4.

Significant differences in genotype and allele distributions between ACL-injured athletes and controls were observed for rs1800255 (*COL3A1*), rs1800012 (*COL1A1*), and rs12722 (*COL5A1*) ($P < 0.05$), whereas rs970547 (*COL12A1*) showed no significant association. Compared with controls, ACL-injured athletes showed higher frequencies of the rs1800255-A and rs1800012-A alleles and a lower frequency of the rs12722-C allele.

To place these findings in a broader population context, effect allele frequencies in the ACL-injured cohort were compared with reference data from the 1000 Genomes Project (global and South Asian populations) and IndiGen. The rs1800255-A allele in *COL3A1* was notably more common in the ACL-injured cohort (0.528) than in the global (0.287), South Asian (0.218), and IndiGen (0.28) reference populations. A similar trend was observed for the rs1800012-A allele in *COL1A1*, which showed a frequency of 0.317 in the ACL cohort compared with 0.097

in global populations, 0.091 in South Asians, and 0.06 in IndiGen. In contrast, the rs12722-C allele in *COL5A1* appeared less frequent in the ACL-injured cohort (0.411) relative to the global (0.552), South Asian (0.651), and IndiGen (0.572) populations. No clear population-level difference was observed for rs970547 in *COL12A1* (Supplementary Table 5).

3.5. Variants Associated with Increased ACL Injury Risk

The *COL3A1* rs1800255 variant showed the strongest association with ACL injury. The A allele was more common in cases than controls, yielding an OR of 1.75 (95% CI: 1.20-2.56; P = 0.0036) under the additive model. Allelic analysis further supported this association (χ^2 = 13.08, P = 0.0003), with significant associations also observed under recessive (P = 0.0009) and trend (P = 0.0019) models. These findings indicate an increased risk of ACL injury associated with the rs1800255 A allele.

Similarly, *COL1A1* rs1800012 was significantly associated with ACL injury. The A allele frequency was higher in cases than controls, corresponding to an OR of 1.68 (95% CI: 1.13-2.49; P = 0.010). Allelic testing supported this association (χ^2 = 13.98, P = 0.00018), indicating increased ACL injury susceptibility associated with the rs1800012 A allele. Detailed effect estimates for each SNP are reported in Table 2, with a visual summary provided in Figure 2.

3.6. Variant Associated with Protective Effect

Unlike *COL3A1* and *COL1A1*, *COL5A1* rs12722 showed a protective association with ACL injury. The C allele was more common in controls than cases, corresponding to an OR of 0.68 (95% CI: 0.47-0.99; P = 0.042) in logistic regression analysis. This association was also supported under allelic (P = 0.0083) and dominant (P = 0.0079) models. These findings indicate a potential protective effect of the rs12722 C allele against ACL injury susceptibility. Detailed association results across all genetic models are provided in Supplementary Table 6.

3.7. Sex-stratified association analysis

Sex-stratified analyses were performed to evaluate potential differences in genetic associations between male and female athletes. rs1800255 showed a significant association with ACL injury susceptibility in both females (OR = 2.13, P = 0.019) and males (OR = 1.63, *P* = 0.038), with a consistent direction of effect across sexes. Similarly, rs1800012 demonstrated a comparable direction of association in both groups and reached statistical significance in males (OR = 1.67, *P* = 0.035), while a similar trend was observed in females (OR = 1.79, *P* = 0.095). In contrast, rs970547 and rs12722 exhibited opposite effect directions between males and females; however, these associations did not reach statistical significance following stratification. Overall, sex-stratified analyses did not reveal strong evidence of sex-specific genetic effects within the study cohort (Supplementary tables 7 and 8).

3.8. Gene-Gene Interaction

Pairwise SNP-SNP interaction analysis among *COL3A1* rs1800255, *COL1A1* rs1800012, and *COL5A1* rs12722 revealed no significant epistatic effects, suggesting independent contributions of these loci to injury susceptibility (Supplementary Table 9). Combined interpretation across all models indicated that variants in *COL3A1* and *COL1A1* confer

increased risk of ACL rupture, while *COL5A1* rs12722 exerts a protective effect. The direction of association remained consistent across allelic and additive models, highlighting the importance of collagen synthesis and fibril architecture in determining ligament resilience.

4. Discussion

This study investigated genetic susceptibility to non-contact ACL injury in elite Indian athletes using a combined approach incorporating exploratory genome-wide analysis and targeted evaluation of collagen gene variants. While the genome-wide analysis did not identify variants reaching the conventional threshold for genome-wide significance ($p < 5 \times 10^{-8}$), targeted analysis identified significant associations involving *COL3A1* rs1800255, *COL1A1* rs1800012, and *COL5A1* rs12722. Specifically, the *COL3A1* rs1800255-A allele and *COL1A1* rs1800012-A allele were associated with increased ACL injury risk, whereas the *COL5A1* rs12722-C allele demonstrated a protective effect. No significant association was observed for *COL12A1* rs970547.

The genome-wide analysis was exploratory in nature and aimed to identify potential novel loci associated with ACL injury susceptibility. Although no variants reached genome-wide significance, three loci exceeded the suggestive significance threshold and may represent candidate regions warranting further investigation. Among these, the region near the *BSG* gene is of particular interest because of its established role in extracellular matrix remodelling and MMP regulation [14]. However, these findings remain preliminary. Several loci demonstrating moderate association signals (P ~10^{-4}-10^{-5}) mapped near genes involved in extracellular matrix (ECM) organization and connective tissue biology, including collagen-related genes such as *COL6A5, COL6A6, COL4A1*, and *COL23A1*, as well as genes involved in elastic fibre architecture and matrix organization, including *FBN1* and *FBLN2* [15-18]. In addition, integrin family genes (*ITGA8, ITGA9*) and ECM remodelling regulators such as *ADAMTS9* and *TNC* were identified [19-22]. We also observed signals near genes associated with tissue repair and mechanotransduction pathways, including VEGFA, BMP7, PDGFC, and WNT5A, which have recognized roles in ligament healing and adaptation to mechanical stress [23-26]. Given the modest sample size and the polygenic nature of ligament-injury susceptibility, the absence of genome-wide significant signals was anticipated and is consistent with observations reported for other complex musculoskeletal traits [27-29]. Moreover, stringent correction for ~1 million independent tests reduces the detection probability for small-effect variants, particularly in highly specialized athletic populations where recruitment is inherently challenging. Importantly, population-structure analyses did not indicate ancestry-related confounding, providing confidence that the observed signals reflect biological variation rather than population stratification.

In contrast to the exploratory GWAS findings, targeted genotyping of *COL1A1, COL3A1, COL5A1*, and *COL12A1* variants demonstrated statistically meaningful associations in our ACL injury group. Comparison with population reference cohorts indicated that the direction of effect observed in the case-control analysis is consistent with population-level differences in allele frequencies, although these comparisons should be interpreted cautiously, as formal statistical testing using individual-level genotype data was not performed. This apparent discrepancy underscores the complementary nature of broad, unbiased GWAS models and

hypothesis-driven candidate-gene approaches. While GWAS primarily interrogates additive effects across the genome, hypothesis-driven analyses permit evaluation under allelic, dominant, recessive, and genotypic models, thereby improving sensitivity to modest but biologically plausible effects. Taken together, these findings support a polygenic model in which multiple collagen-related loci exert incremental influences on ligament structure and injury susceptibility.

Among the investigated variants, *COL3A1* rs1800255 demonstrated the most consistent signal across allelic and genotypic models. The A allele was associated with an approximately 1.75-fold increased risk of ACL rupture, suggesting a meaningful contribution to ligament injury risk. Type III collagen, encoded by *COL3A1*, contributes to ligament elasticity and tissue remodelling. Disruption in fibrillar architecture can reduce tensile resilience and increase vulnerability to mechanical overload. Similar associations have been reported in Polish, Korean, and meta-analytic cohorts [30,31], though some studies noted population-specific differences [32,33]. The current findings confirm that *COL3A1* rs1800255 associates with ACL injury risk in an Indian genetic background, indicating that the risk effect of this allele is not ancestry-restricted, even if its magnitude may vary across populations. Similarly, *COL1A1* rs1800012-A allele showed a significant association in injured athletes. This variant alters an Sp1 binding site, leading to dysregulated α1:α2 chain ratios in type I collagen fibrils and reduced tensile strength [34]. Although previous studies have shown conflicting results [35,36], our findings align with those reporting increased injury susceptibility with the A allele [37,38]. Given that type I collagen constitutes the primary load-bearing component of the ACL, even slight perturbations in fibril composition may have functional consequences under repetitive high-load conditions. Unlike *COL3A1* and *COL1A1*, *COL5A1* rs12722 showed a protective effect, with the C allele significantly enriched in controls. Type V collagen regulates fibril diameter and cross-linking of type I collagen; hence, alleles that improve fibril organization enhance ligament stiffness and recovery. The protective effect of *COL5A1* rs12722 has been widely replicated across multiple populations [39,40]. Our findings provide the first evidence in Indian athletes, confirming that *COL5A1* acts as a resilience-conferring locus even in non-Caucasian groups. The C allele likely contributes to improved fibril organization and tensile strength of the ligament, explaining its protective association. No significant association was detected for *COL12A1* rs970547. As *COL12A1* primarily contributes to extracellular matrix organization rather than to direct tensile strength, its effect on ACL injury susceptibility may be modest or context-dependent, consistent with previous reports [41]. Sex-stratified analyses showed generally consistent directions of effect between males and females, although statistical significance varied after stratification. No robust evidence of sex-specific genetic effects was observed. Similarly, SNP-SNP interaction analysis did not identify significant epistatic effects, suggesting that these variants may contribute independently rather than through strong interaction effects.

Multiple intermediate phenotypes, including generalized joint laxity and altered connective tissue mechanical properties, influence ACL injury susceptibility. Increased generalised joint hypermobility (GJH), knee hyperextension, and instability in translational and rotational axes have been associated with increased ACL rupture risk [42,43]. Previous studies have proposed that collagen gene variants influence these phenotypes through effects on extracellular matrix

organization and collagen fibrillogenesis, processes that affect ligament structure and biomechanical function [44,45]. Variants in *COL1A1, COL5A1*, and *COL12A1* have additionally been associated with connective tissue properties, altered tibial rotation, and ligament mechanical behaviour [46-48]. Although joint laxity and biomechanical measurements were not assessed in the present study, the observed associations involving *COL1A1*, *COL3A1*, and *COL5A1* are consistent with this proposed pathway linking collagen variation to ligament function and injury risk. These findings further suggest that the influence of collagen-related variants may extend beyond ACL injury alone. *COL1A1*, *COL3A1*, *COL5A1*, and *COL12A1* contribute broadly to connective tissue organization and have been implicated in tendon injuries, joint instability, and related musculoskeletal phenotypes, indicating a broader role in connective tissue resilience and mechanically induced injury susceptibility.

This study provides the first genetic evidence of susceptibility to non-contact ACL injury in elite Indian athletes, a population underrepresented in sports genomics research. The observed associations involving *COL3A1*, *COL1A1*, and *COL5A1* support the role of collagen-related pathways in ligament biomechanics and connective tissue integrity, reinforcing the biological importance of extracellular matrix organization in ACL injury susceptibility. Strengths of the study include a clinically well-defined cohort and the integration of genome-wide and targeted approaches; however, the modest sample size limited genome-wide discovery power. Larger multi-centre longitudinal studies integrating genetic, biomechanical, and imaging data will be important for validating these findings and advancing personalized approaches for ACL injury risk assessment, prevention, and rehabilitation in sports medicine.

Acknowledgement

We appreciate the participation of athletes in the study.

Declaration of interest

The authors declare no conflicts of interest.

Author Contributions

Khalid Khader: substantial contribution to research design, sample collection, analysis, interpretation of results, and drafting of manuscript. **Shahrumi Reza:** substantial contribution to research design, performing experiment, analysis, interpretation of results, and drafting of manuscript. **Tiyasha De:** contribution in performing experiment. **Alia Hashmi**: contribution in performing experiments, interpretation of results, and revision and editing of manuscript. **Deepak Joshi:** substantial contribution to research design and sample collection. **Dr Nitin Mehta:** substantial contribution to research design and sample collection. **Vaisakhan SR:** substantial contribution to research design and sample collection. **Ildus I. Ahmetov:** Revision of manuscript. **Mohammed Faruq:** funding, substantial contribution to research design, acquisition, analysis, interpretation of results, revising the paper critically, and approval of the submitted and final versions.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References:

1. Griffin LY, Albohm MJ, Arendt EA, Bahr R, Beynnon BD, DeMaio M, et al. Understanding and preventing noncontact anterior cruciate ligament injuries: A review of the Hunt Valley II meeting, January 2005. *Am J Sports Med.* 2006;34(9):1512–1532.
2. Magnusson K, Turkiewicz A, Hughes V, Frobell R, Englund M. High genetic contribution to anterior cruciate ligament rupture: Heritability ~69. *Br J Sports Med.* 2021;55:385-389.
3. Ricard-Blum S. The collagen family. *Cold Spring Harb Perspect Biol.* 2011;3(1):a004978.
4. Massidda M, Flore L, Scorcu M, Monteleone G, Tiloca A, Salvi M, et al. Collagen gene variants and anterior cruciate ligament rupture in Italian athletes: A preliminary report. *Genes.* 2023;14(7):1418.
5. Rahim M, Gibbon A, Collins M, September AV. Genetics of musculoskeletal soft tissue injuries: Current status, challenges, and future directions. In: *Sports, Exercise, and Nutritional Genomics.* Elsevier; 2019:317–339.
6. Khoschnau S, Melhus H, Jacobson A, Rahme H, Bengtsson H, Ribom E, et al. Type I collagen α1 Sp1 polymorphism and the risk of cruciate ligament ruptures or shoulder dislocations. *Am J Sports Med.* 2008;36(12):2432–2436.
7. Stępień-Słodkowska M, Ficek K, Maciejewska-Karłowska A, Sawczuk M, Ziętek P, Król P, et al. Overrepresentation of the *COL3A1* AA genotype in Polish skiers with anterior cruciate ligament injury. *Biol Sport.* 2014;32(2):143–147.
8. Collins M, Mokone GG, September AV, Van Der Merwe L, Schwellnus MP. The *COL5A1* genotype is associated with range of motion measurements. *Scand J Med Sci Sports.* 2009;19(6):803–810.
9. Lv ZT, Wang W, Zhao DM, Huang JM. *COL12A1* rs970547 polymorphism does not alter susceptibility to anterior cruciate ligament rupture: A meta-analysis. *Front Genet.* 2021;12:665861.
10. Bulbul A, Ari E, Apaydin N, Ipekoglu G. The impact of genetic polymorphisms on anterior cruciate ligament injuries in athletes: A meta-analytical approach. *Biology (Basel).* 2023;12(12):1526.
11. Posthumus M, September AV, Keegan M, O'Cuinneagain D, Van Der Merwe W, Schwellnus MP, et al. Genetic risk factors for anterior cruciate ligament ruptures: *COL1A1* gene variant. *Br J Sports Med.* 2009;43(5):352–356.
12. Guo R, Ji Z, Gao S, Aizezi A, Fan Y, Wang Z, et al. Association of *COL5A1* gene polymorphisms and musculoskeletal soft tissue injuries: A meta-analysis based on 21 observational studies. *J Orthop Surg Res.* 2022;17(1):129.

13. Ginevičienė V, Urnikytė A. Association of *COL12A1* rs970547 polymorphism with elite athlete status. *Biomedicines.* 2022;10(10):2495.
14. Braundmeier AG, Fazleabas AT, Nowak RA. Extracellular matrix metalloproteinase inducer expression in the baboon endometrium: menstrual cycle and endometriosis. *Reproduction*. 2010;140(6):911-920. doi:10.1530/REP-09-0481.
15. Shoulders MD, Raines RT. Collagen structure and stability. *Annu Rev Biochem*. 2009;78:929-958. doi:10.1146/annurev.biochem.77.032207.120833.
16. Ricard-Blum S. The collagen family. *Cold Spring Harb Perspect Biol*. 2011;3(1):a004978. Published 2011 Jan 1. doi:10.1101/cshperspect.a004978
17. Ramirez F, Sakai LY, Dietz HC, Rifkin DB. Fibrillin microfibrils: multipurpose extracellular networks in organismal physiology. *Physiol Genomics*. 2004;19(2):151-154. doi:10.1152/physiolgenomics.00092.2004.
18. de Vega S, Iwamoto T, Yamada Y. Fibulins: multiple roles in matrix structures and tissue functions. *Cell Mol Life Sci*. 2009;66(11-12):1890-1902. doi:10.1007/s00018-009-8632-6.
19. Humphries JD, Byron A, Humphries MJ. Integrin ligands at a glance. *J Cell Sci*. 2006;119(Pt 19):3901-3903. doi:10.1242/jcs.03098.
20. Hynes RO. Integrins: bidirectional, allosteric signaling machines. *Cell*. 2002;110(6):673-687. doi:10.1016/s0092-8674(02)00971-6.
21. Kelwick R, Desanlis I, Wheeler GN, Edwards DR. The ADAMTS (A Disintegrin and Metalloproteinase with Thrombospondin motifs) family. *Genome Biol*. 2015;16(1):113. Published 2015 May 30. doi:10.1186/s13059-015-0676-3.
22. Midwood KS, Chiquet M, Tucker RP, Orend G. Tenascin-C at a glance. *J Cell Sci*. 2016;129(23):4321-4327. doi:10.1242/jcs.190546.
23. Vavken P, Saad FA, Fleming BC, Murray MM. VEGF receptor mRNA expression by ACL fibroblasts is associated with functional healing of the ACL. *Knee Surg Sports Traumatol Arthrosc*. 2011;19(10):1675-1682. doi:10.1007/s00167-011-1443-y.
24. Salazar VS, Gamer LW, Rosen V. BMP signalling in skeletal development, disease and repair. *Nat Rev Endocrinol*. 2016;12(4):203-221. doi:10.1038/nrendo.2016.12.
25. Hee CK, Dines JS, Solchaga LA, Shah VR, Hollinger JO. Regenerative tendon and ligament healing: opportunities with recombinant human platelet-derived growth factor BB-homodimer. *Tissue Eng Part B Rev*. 2012;18(3):225-234. doi:10.1089/ten.TEB.2011.0603.
26. Hu L, Chen W, Qian A, Li YP. Wnt/β-catenin signaling components and mechanisms in bone formation, homeostasis, and disease. *Bone Res*. 2024;12(1):39. Published 2024 Jul 10. doi:10.1038/s41413-024-00342-8
27. Uffelmann E, Huang QQ, Munung NS, De Vries J, Okada Y, Martin AR, et al. Genome-wide association studies. *Nat Rev Methods Primers.* 2021;1(1):59.
28. Kim SK, Nguyen C, Avins AL, Abrams GD. Three genes associated with anterior and posterior cruciate ligament injury: A genome-wide association analysis. *Bone Jt Open.* 2021;2(6):414–421.

29. Sahai H, Khurshid A. Formulae and tables for the determination of sample sizes and power in clinical trials for testing differences in proportions for the two-sample design: A review. *Stat Med.* 1996;15(1):1–21.
30. Seong KH, Kim KJ, Park HJ, Lee BH, Chang E, Park DH. Collagen gene variants and anterior cruciate ligament rupture in the Korean population. *Exerc Sci.* 2024;33(4):409–414.
31. Kim SK, Nguyen C, Avins AL, Abrams GD. Three genes associated with anterior and posterior cruciate ligament injury: A genome-wide association analysis. *Bone Jt Open.* 2021;2(6):414–421.
32. O'Connell K, Knight H, Ficek K, Leonska-Duniec A, Maciejewska-Karlowska A, Sawczuk M, et al. Interactions between collagen gene variants and risk of anterior cruciate ligament rupture. *Eur J Sport Sci.* 2015;15(4):341–350.
33. İpekoğlu G, Çakir Hİ, Bozkurt N, Bülbül A, SavkiN ST, Apaydin N. Genetic factors and anterior cruciate ligament injury risk in professional football players: *COL3A1* (rs1800255) and *COL5A1* (rs12722) polymorphisms. *Turk J Kinesiol.* 2023;9(3):195–203.
34. Perini JA, Lopes LR, Guimarães JAM, Goes RA, Pereira LFA, Pereira CG, et al. Influence of type I collagen polymorphisms and risk of anterior cruciate ligament rupture in athletes: A case–control study. *BMC Musculoskelet Disord.* 2022;23(1):154.
35. Mirghaderi SP, Salimi M, Kheirollahi M, Mortazavi SMJ, Akbari-Aghdam H. Anterior cruciate ligament injury and its postoperative outcomes are not associated with polymorphism in *COL1A1* rs1107946 (G/T): A case–control study in Middle Eastern elite athletes. *J Orthop Surg Res.* 2022;17(1):462.
36. Wang C, Li H, Chen K, Wu B, Liu H. Association of polymorphism rs1800012 in *COL1A1* with sports-related tendon and ligament injuries: A meta-analysis. *Oncotarget.* 2017;8(16):27627–27634.
37. Gibbon A, Raleigh SM, Ribbans WJ, Posthumus M, Collins M, September AV. Functional *COL1A1* variants are associated with the risk of acute musculoskeletal soft tissue injuries. *J Orthop Res.* 2020;38(10):2290–2298.
38. Piri MM, Cerit M, Anılır M, Polat T, Karaduman AA, Szabo A, et al. Effects of Exercise Addiction and the *COL1A1* Gene rs1800012 Polymorphism on Injury Susceptibility in Elite Female Volleyball Players. Genes. 2025;16(11):1300.
39. Pabalan N, Tharabenjasin P, Phababpha S, Jarjanazi H. Association of *COL5A1* gene polymorphisms and risk of tendon–ligament injuries among Caucasians: A meta-analysis. *Sports Med Open.* 2018;4(1):46.
40. Lulińska-Kuklik E, Rahim M, Domańska-Senderowska D, Ficek K, Michałowska-Sawczyn M, Moska W, et al. Interactions between *COL5A1* gene and risk of anterior cruciate ligament rupture. *J Hum Kinet.* 2018;62(1):65–71.
41. Kaynak M, Nijman F, Van Meurs J, Reijman M, Meuffels DE. Genetic variants and anterior cruciate ligament rupture: A systematic review. *Sports Med.* 2017;47(8):1637–1650.
42. Brazier J, Antrobus MR, Herbert AJ, et al. Gene variants previously associated with reduced soft tissue injury risk: Part 1 - independent associations with elite

status in rugby. *Eur J Sport Sci*. 2023;23(5):726-735. doi:10.1080/17461391.2022.2053752.
43. Myer GD, Ford KR, Paterno MV, Nick TG, Hewett TE. The effects of generalized joint laxity on risk of anterior cruciate ligament injury in young female athletes. *Am J Sports Med*. 2008;36(6):1073-1080. doi:10.1177/0363546507313572.
44. Laguette MJ, Abrahams Y, Prince S, Collins M. Sequence variants within the 3'-UTR of the COL5A1 gene alters mRNA stability: implications for musculoskeletal soft tissue injuries. *Matrix Biol*. 2011;30(5-6):338-345. doi:10.1016/j.matbio.2011.05.001.
45. Izu Y, Birk DE. Collagen XII mediated cellular and extracellular mechanisms in development, regeneration, and disease. *Front Cell Dev Biol*. 2023;11:1129000. Published 2023 Mar 2. doi:10.3389/fcell.2023.1129000.
46. Bell RD, Shultz SJ, Wideman L, Henrich VC. Collagen gene variants previously associated with anterior cruciate ligament injury risk are also associated with joint laxity. *Sports Health*. 2012;4(4):312-318. doi:10.1177/1941738112446684.
47. Małecki K, Fabiś-Strobin A, Sałacińska K, et al. Clinical significance of polymorphisms of genes encoding collagen (COL1A1, COL5A1) and their correlation with joint laxity and recurrent patellar dislocation in adolescents. *Sci Rep*. 2023;13(1):22300. Published 2023 Dec 15. doi:10.1038/s41598-023-49378-6.
48. Beckley S, Dey R, Stinton S, et al. The Association of Genetic Variants Within the Type XII Collagen and Tenascin C Genes with Knee Joint Laxity Measurements. *Genes (Basel)*. 2025;16(2):164. Published 2025 Jan 27. doi:10.3390/genes16020164

608

609

Table 1. Demographic and sports characteristics of ACL-injured athletes and non-injured control athletes included in the study		
Parameters	**ACL Injury Cases (N=90)**	**Non-ACL Controls (N=90)**
Age (years), Median(range)	23 (32-18)	23 (29-18)
N(n Male)	90 (63)	90 (53)
N (n Female)	90 (27)	90 (37)
Height (cm), mean ± SD	169.7 ± 6.7	168.6 ± 7.0
Weight (kg), mean ± SD	60.4 ± 8.1	59.2 ± 7.0
BMI (kg/m²), Median (range)	20.7 (27.3-15.6)	20.7 (27.3-16)
Sports (n)		
GYMNASTICS	4	1
KABADDI	9	9
HANDBALL	3	0
HOCKEY	6	7
FOOTBALL	15	8
POLE VAULT	1	0
CRICKET	6	6
MARATHON	1	0
ATHLETICS	3	6
WRESTLER	10	13
BOXING	2	8
BASKETBALL	7	7
VOLLEYBALL	1	2
BADMINTON	3	10
JAVELLIN	2	1
PARAGLIDING	1	0
WUSHU	3	1
MMA	1	1
ATHLETICS 400	1	0
KHOKHO	1	0
TAEKWANDO	1	1
RUGBY	1	0
ARCHERY	0	2
JUDO	0	3
Data are presented as mean ± standard deviation (SD), median (range), or number of participants (n), as indicated. ACL injury cases consisted of athletes with clinically and MRI-confirmed non-contact anterior cruciate ligament (ACL) injury, whereas controls included athletes with no history of ACL or major knee injury. Sports categories indicate the number of athletes participating in each discipline within each group. BMI, body mass index; N, total number of participants; n, subgroup count. Height and weight values were available at the time of recruitment and used to calculate BMI.		

610

Table 2. Association of collagen gene variants with anterior cruciate ligament (ACL) injury in Indian athletes.

Chr	SNP	Gene	Effect Allele	OR	95% CI	p-value	FDR
2	rs1800255	COL3A1	A	**1.75**	1.20-2.56	**0.0036**	**0.014**
17	rs1800012	COL1A1	A	**1.68**	1.13-2.49	**0.010**	**0.020**
9	rs12722	COL5A1	C	**0.68**	0.47–0.99	**0.042**	0.050
6	rs970547	COL12A1	C	0.78	0.50–1.21	0.273	0.273

Odds ratios (ORs) and 95% confidence intervals (CIs) were estimated using logistic regression adjusted for age and sex under an additive genetic model. The effect allele represents the allele tested for association. OR >1 indicates increased ACL injury risk, whereas OR <1 indicates a protective effect. P-values were corrected for multiple testing using the Benjamini–Hochberg false discovery rate (FDR) method.

Figure Legends

Figure 1a. Allele Frequencies in ACL vs Non-ACL Athletes. Comparison of allele frequency distribution for *COL3A1* (rs1800255), *COL5A1* (rs12722), *COL1A1* (rs1800012), and *COL22A1* (rs970547) variants between ACL-injured and non-injured athletes. Frequencies are presented as percentages of total alleles in each group.

Figure 1b. Genotype Frequencies in ACL vs Non-ACL Athletes. Distribution of genotypes for *COL3A1* (rs1800255), *COL5A1* (rs12722), *COL1A1* (rs1800012), and *COL22A1* (rs970547) polymorphisms in athletes with anterior cruciate ligament (ACL) injury and uninjured controls (non-ACL). Data are expressed as percentage frequency. Statistical differences between groups were assessed using the χ^2 test. Asterisks indicate statistically significant differences between ACL and non-ACL groups ($p < 0.05$); ns denotes non-significant results.

Figure 2. Forest plot showing associations between collagen gene variants and ACL injury susceptibility in Indian athletes. Odds ratios (ORs) and 95% confidence intervals (CIs) for the association of COL3A1 rs1800255, COL1A1 rs1800012, COL5A1 rs12722, and COL12A1 rs970547 with ACL injury risk under the additive model. The vertical dashed line indicates the null effect threshold (OR = 1). OR >1 indicates increased ACL injury risk, whereas OR <1 indicates a protective effect. COL3A1 rs1800255 and COL1A1 rs1800012 showed increased risk associations, while COL5A1 rs12722 demonstrated a protective effect; no significant association was observed for COL12A1 rs970547.