Gene variants within the *COL1A1* gene are associated with reduced Anterior Cruciate Ligament injury in professional soccer players
Abstract

Objective: To examine the association of COL1A1-1997G/T and +1245G/T polymorphisms, individually and as haplotypes, with ACL ruptures in professional soccer players. Design: Subjects were 91 male professional soccer players with surgically diagnosed primary ACL ruptures. The control group consisted of 143 apparently healthy male professional soccer players, who were without any self-reported history of ligament or tendon injury. Both subjects and healthy controls are from the same soccer teams, of the same ethnicity (Polish, East-Europeans for ≥ 3 generations), a similar age category, and had a comparable level of exposure to ACL injury.

Methods: Genomic DNA was extracted from the oral epithelial cells using GenElute Mammalian Genomic DNA Miniprep Kit (Sigma, Germany). All samples were genotyped using a Rotor-Gene real-time polymerase chain reaction (PCR). Results: Genotype distributions for both polymorphisms met the Hardy-Weinberg expectations in both subjects and controls (p>0.05). Higher frequency of the COL1A1 G-T (-1997G/T and +1245G/T polymorphisms) haplotype was significantly associated with reduced risk for ACL rupture (Hap.score -1.98, p=0.048). The TT genotype was under-represented in the ACL rupture group. However, this result was not statistically significant (p=0.084 Fisher’s exact test, recessive mode: TT vs GT+GG).

Conclusions: Higher frequency of the COL1A1 G-T haplotype is associated with reduced risk of ACL injury in a group of professional soccer players. Consequently, carrying two copies the COL1A1 G-T haplotype may be protective against ACL injury.

Key words: ACL, Sport injuries, Soccer, Athletes, Genetic Polymorphism, COL1A1
1. Introduction

Soccer is an intermittent team sport with high physiological demands. Professional players cover 8,000–12,000 m during a match, with up to 20% of the overall distance corresponding to maximal or near maximal running efforts.\(^1\) \(^2\) Rapid movements such as jumping and tackling are also frequently performed by players during a match.\(^3\) \(^4\) Given the high physiological demands, it is not surprising that soccer is also associated with a relatively large number of injuries. The incidence of soccer-related injuries is estimated to be 10-35 per 1000 hours of exposure in adult male soccer players,\(^5\) \(^6\) with approximately 60-80% of injuries occurring in the lower extremities, most commonly at the knee or ankle.\(^7\)

The anterior cruciate ligament (ACL) rupture is one of the most severe musculoskeletal soft tissue injuries in professional sport.\(^8\) \(^9\) ACL ruptures are complex, multifactorial disorders determined by the interaction of extrinsic and intrinsic risk factors.\(^10\) The familial aggregation observed in ACL injuries\(^11\) \(^12\) has prompted researchers to investigate a possible genetic linkage.

Genetic polymorphisms within the major alpha chains of the collagen type I gene (COL1A1), which is located on chromosome 17q21, have been shown to influence the predisposition for ACL rupture in non-athletic cohorts.\(^13\) In particular, the functional COL1A1 Sp1 binding site polymorphism (COL1A1 Sp1 +1245G/T, rs1800012), initially described in 1996,\(^14\) has been associated with the risk for ACL injury. The uncommon COL1A1 Sp1 TT genotype is significantly underrepresented in participants with ACL ruptures, and therefore a higher frequency of COL1A1 TT genotype was proposed to have a protective effect against ACL rupture.\(^9\) \(^15\) \(^16\) A second polymorphism (G/T) has been identified in the proximal promoter of COL1A1, at position -1997 (rs1107946) relative to the transcription start site, which is in linkage disequilibrium with the Sp1 polymorphism.\(^17\) This COL1A1-1997 G/T polymorphism has been associated with bone mineral density (BMD), and has also been reported to interact with the COL1A1 Sp1 +1245G/T polymorphism to regulate BMD.\(^17\) \(^18\) Furthermore, a functional significance of the COL1A1-1997 G/T polymorphism was demonstrated in the in-vitro regulation of the COL1A1 gene in osteoblasts; the G allele showed a higher transcriptional activity than the T allele.\(^19\) Despite its functionality, thus far, the potential influence of the COL1A1-1997 G/T polymorphism on the
incidence of ACL rupture injury has never been studied.

Recently, a transcription analysis that included COL1A1 -1997G/T, and COL1A1 Sp1 +1245G/T polymorphisms in the 5’ flank of COL1A1 revealed that the levels of transcription are influenced by haplotype (i.e., a combination of alleles at adjacent locations on the chromosome that are transmitted together), rather than by the genotype, at individual polymorphic sites, indicating that COL1A1 polymorphisms interact with each other to form a haplotype that regulates transcription.\(^{21}\) That leads us to hypothesis that interaction between two or more polymorphisms within the COL1A1 gene may influence the predisposition for ACL injury.

Soccer players are obviously more exposed to ACL rupture injury than the general population. Therefore, the aim of this study was to examine the association of the COL1A1 -1997G/T and COL1A1 Sp1 +1245G/T polymorphisms in the COL1A1 gene, individually and as haplotypes, with ACL ruptures in professional male soccer players. We hypothesized that the: 1) COL1A1 Sp1+1245G/T and the COL1A1 -1997G/T polymorphisms would be individually associated with the incidence of ACL rupture and; 2) The interaction between the COL1A1 Sp1+1245G/T and the COL1A1-1997G/T polymorphisms will form a haplotype that predisposes athletes to a greater risk of ACL rupture.

2. Methods

The study was approved by the Pomeranian Medical University Ethics Committee, Poland and written informed consent was obtained from each participant according to the declaration of Helsinki. A total of 91 professional male soccer players (age=23±3 years) with surgically-diagnosed primary ACL ruptures who qualified for ligament reconstruction were recruited for this study. All players had non-contact ACL injuries. For the obvious reason that soccer teams are homogenous in term of gender, we have recruited only male subjects. Soccer players were all participants in the Polish 1\(^{st}\) division professional soccer league, with an overall training time of 14-18 hours per week (7-9 training sessions a week, 2 hours each training session). Subjects were treated in the Galen Orthopaedics, Bieruń, Poland. The control group consist of 143 apparently healthy male professional soccer players (age=25.2±2.6 years), who were without any self-
reported history of ligament or tendon injury. Both the ACL rupture group and the healthy controls were from the same soccer teams, of the same ethnicity (as self-reported, all Polish, East-Europeans for ≥ 3 generations), of similar age (ACLR group=23±3, control group=24±5), and had a comparable level of exposure to risk of ACL injury (same volume and intensity of training and match play).

We followed recent recommendations for genotype-phenotype association studies.\textsuperscript{22, 23} Genomic DNA was extracted from the oral epithelial cells using GenElute Mammalian Genomic DNA Miniprep Kit (Sigma, Germany) according to manufacturer’s protocol. Allelic discrimination of COL1A1 Sp1 +1245G/T (rs1800012) and -1997G/T (rs1107946) polymorphic sites was performed using a TaqMan Pre-Designed SNP Genotyping Assays (Applied Biosystems, USA), including primers and fluorescently labelled (FAM and VIC) probes for the detection of alleles. All samples were genotyped on a Rotor-Gene real-time polymerase chain reaction (PCR) instrument (Corbett, Australia). Thermal cycler conditions were as follows: an initial step at 95°C for 5 min, followed by 45 cycles of denaturation at 94°C for 15 s and anneal/extend at 60°C for 1 min. We used positive and negative controls for the detection of both polymorphisms. The results were scored by two experienced and independent investigators who were blind to the participants’ data.

Genotype and allele frequency were analysed using $\chi^2$ or Fisher exact tests. Allelic based odds ratios (OR) with 95% confidence intervals (95%CI) were calculated using logistic regression analysis. A post hoc power calculation was detected for each gene variant. The genotypes between cases and controls were compared in three ways: first, in a general test of association in the 2-by-3 table of phenotype-by-genotype, then two different modes of inheritance of minor allele were assumed: dominant, in which homozygotes and heterozygotes for the minor allele were pooled and compared to homozygotes for the major allele and recessive, in which homozygotes and heterozygotes for the major alleles were pooled and compared to homozygotes for the minor allele. In addition, the programming language and environment R \url{http://www.r-project.org} was used for Hardy-Weinberg and linkage disequilibrium (LD) testing (package genetics) and for the
haplotype analysis (package haplo.stats). The haplo.em, haplo.group and haplo.score functions of
the haplo.stats package were used to infer haplotype frequencies and to test the association
between reconstructed haplotypes and the risk of ACL rupture assuming three possible haplotype
effects: additive, dominant and recessive. Hap.score is the statistical score for haplotypes
reflecting the strength of association; the positive value of hap.score indicates increased risk of
ACL injury for a particular haplotype, while the negative value indicates reduced risk. The pair-
wise linkage disequilibrium between -1997G/T and +1245G/T was estimated by D’ and r². For all
tests, significance was set at p<0.05.

3. Results
Both the ACL rupture group and the healthy controls are at a similar age category (ACL rupture
group=23±3, control group=24±5, p=0.093), with a similar height (ACL rupture
group=179±4.6cm, control group=178±5.2 cm, p=0.13), and a similar body mass (ACL rupture
group=75±3.9 kg, control group=74.3±4.7 kg, p=0.22). The genotype and allele frequencies for
the COLIA1-1997G/T and COLIA1 Sp1 +1245G/T polymorphisms of COLIA1 are shown in
Table 1. The genotype distributions for both polymorphisms met the Hardy-Weinberg
expectations in both groups (ACL rupture and controls) (p>0.2).

There were no significant differences in genotype distribution and allele frequencies of the
COLIA1 -1997G/T (rs1107946) and +1245G/T (rs1800012) polymorphisms between the ACL
rupture group and the control group using the 2-by-3 general test of association (Table 1).
However, none of the 91 participants with ACL rupture harboured the TT genotype of the
COLIA1 +1245G/T (rs1800012) polymorphisms, whereas six (4.2%) TT homozygotes were
present in the control group. There was a trend toward under-representation of the TT genotype in
the ACL rupture group (p=0.084 Fisher’s exact test, recessive mode: TT vs GT+GG). A post hoc
power calculation revealed that assuming confidence level is at 95%, the statistical power to
detect the differences in TT genotype between ACL rupture group and the control group is 47%.
If the difference was two times higher (8.4% vs. 0%) the statistical power would increase to 82%.
Likewise, there were also no significant differences in the dominant or the recessive tests for the
COL1A1 -1997G/T polymorphism.

COL1A1 -1997G/T and +1245G/T were found to be in linkage disequilibrium (Table 2). A total of 3 reconstituted haplotypes with estimated frequency >0.05 were found, and only those were evaluated for an association with ACL rupture. The G-G (-1997G, +1245G) haplotype was the most common (frequency 67.9%). Two other haplotypes, G-T (frequency 16.9%) and T-G (frequency 15.2%), had similar frequencies. The rare T-T haplotype (frequency 0%) was not present in any of the subjects. We then tested the association between these haplotypes and ACL rupture assuming three haplotype effects: additive (considering the count of a particular haplotype as 0, 1 and 2), dominant (heterozygous or homozygous carrier of a particular haplotype versus otherwise) and recessive (homozygous for a particular haplotype versus otherwise). Under the recessive mode of inheritance, haplotype analysis yielded a mild significant association with ACL rupture (p=0.048, Table 3), as two copies of the G-T haplotype conferred decreased risk of this injury.

4. Discussion

We examined the association between COL1A1 -1997G/T (rs1107946) and COL1A1 Sp1+1245G/T (rs1800012) and incidents of ACL ruptures in professional soccer players. In contrast to our initial hypotheses, the COL1A1 Sp1+1245G/T and the COL1A1 -1997G/T polymorphisms were not individually associated with the incidence of ACL rupture. However, a novel finding in the present study was that the G-T haplotype (COL1A1-1997G, +1245T) is significantly underrepresented in the ACL rupture group compared with healthy controls (p=0.048), suggesting that harbouring this particular haplotype may have a protective effect against ACL rupture injury.

In the present study, no association was found between the COL1A1 -1997G/T polymorphism and incidence of ACL ruptures. The -1997G/T polymorphism has been associated with bone mineral density (BMD) in several studies and was found to be in high degree of linkage disequilibrium (LD) with the +1245G/T loci. In general, promoter polymorphisms are
suspected to affect gene transcription activity, and thereby gene functions. That, together with the assumption that the -1997G/T polymorphism has never been tested with respect to ACL ruptures, led us to consider this polymorphism separately, as a candidate to influence the risk of sustaining an ACL rupture. Our results suggest that the -1997G/T polymorphism is not associated with incidents of ACL rupture; however it contributes to the combined influence of the COL1A1-1997G, +1245T haplotype on the incidence of ACL ruptures.

Higher frequency distribution of the COL1A1 +1245TT genotype has been reported to be associated with a substantially lower risk of cruciate ligament ruptures in both Swedish and South African participants. Combined analysis of these two studies suggests that the TT genotype has a possible preventive role not only in ACL rupture, but also in other soft tissue injuries. The TT genotype frequency, when compared to combined control groups (4.1%) was about ten times less frequent in subjects with all other soft tissue injuries, namely, cruciate ligament ruptures, shoulder dislocation, and Achilles tendon rupture (2 cases out of 517, 0.4%). In keeping with these observations, the COL1A1 Sp1 (+1245G/T) polymorphism has been proposed as a functional variant which modulates Sp1 binding and COL1A1 gene regulation, increasing the production of collagen α1(I) chain relative to α2(I) and reducing bone strength. Similar to previous studies, we found lower TT genotype frequency among our ACL rupture athletes (0%) compared to controls (6/143 participants, 4.2%). This difference was however only a trend (p=0.084), probably owing to the relatively low sample size, and, more importantly, the low minor allele frequency (MAF) in both subjects and controls (combined MAF=16.8%), which makes meaningful (MAF>40%) association harder to detect.

We have used a haplotype-based approach and reconstructed two-locus haplotypes that were also tested for an association with the risk of ACL rupture. It was assumed that predisposition for ACL rupture might be a polygenic trait, and therefore haplotype in two candidate polymorphisms would provide more information on the complex relationship between DNA sequence variation and traits than any single polymorphism. In fact, studies have shown that haplotype analyses are more powerful than marker-by-marker analyses when the genotypes are in LD with the causative locus. The G-T haplotype (estimated frequency among subjects
14.3%) was significantly lower in our ACL rupture group compared to controls, suggesting that participants with two copies of this haplotype have a decreased risk for ACL injury. Additionally, haplotype frequencies and score (D’ and r²) in our ACL rupture group (data not shown) were similar to those in a large cohort of The Rotterdam Study.\(^{19}\)

A possible functional explanation for the aforementioned results was recently provided by Jin et al.,\(^{21}\) who found a significantly higher transcriptional activity of \textit{COL1A1} gene for the haplotype G-inde-T (-1997G/T, -1663indeT, +1245G/T) compared to all other haplotypes, by using a functional luciferase gene reporter analysis. This study suggests that higher transcriptional activity of \textit{COL1A1} results in an unusual ratio of alpha-1(I) chains relative to alpha-2(I). However, it remains unclear if the higher transcription activity of \textit{COL1A1} has either a negative or positive effect on the incidence of soft tissue injuries, including ACL ruptures.\(^{21, 27}\) In the present study, we were unable to report the association between the -1663indeT polymorphism and ACL rupture due to study limitations. It has been shown, however, that the -1663indeT polymorphism, that corresponds to a deletion of a T within a tract of eight T residues, has a minor, if any, effect on the transcription activity of \textit{COL1A1} gene. In fact, haplotype analysis of the \textit{COL1A1} -1997G/T and -1663indeT polymorphisms revealed that haplotypes containing the G allele (-1997 G/T polymorphism) yielded high transcriptional activity regardless the -1663 allele status.\(^{20}\)

We believe that the results of our study are overall valid, as we strictly followed the latest genotype:phenotype study recommendations,\(^{22, 23}\) and all of the following criteria have been met: both cases and controls (athletes) clearly presented the main study phenotype (i.e., being a professional soccer players) and where equally exposed to risk of ACL rupture, participants within each cohort were both age and ethnically-matched, genetic assessment was accurate and unbiased. Further, genotype distributions were in Hardy-Weinberg equilibrium (HWE) in both cases and controls and were similar to control genotype frequencies in other studies.\(^{17, 19}\)

This study has some limitations. In general, genetic association studies must be interpreted with caution, since there is a non-trivial possibility of false positive results attributable to chance, particularly in studies involving multiple gene-trait analyses. Furthermore, the present study is hampered by small sample size, which reflects the low number of professional soccer players,
from the same origin, who had non-contact ACL injuries.

5. Conclusion

The COL1A1 G-T haplotype was associated with a reduced risk of ACL injury in a group of professional soccer players. Consequently, carriers of two copies of this haplotype may have a reduced risk of ACL injury. Although functional analyses of COL1A1 5’ flank suggest that the G-T haplotype is responsible for enhanced transcriptional activity of the COL1A1 gene, further investigation, including evaluation of Type I/ Type 3 collagen ratio, and the possible effect of microRNAs, is required to explain the relationship between COL1A1 expression and susceptibility to ACL injury.

6. Practical implications

- The results of the present study will assist to understand which genetic profiles contribute to higher ACL injury risk.
- Discovering the complex relationship between gene variants and ACL rupture among athletes may assist clinicians and coaches to optimize training, and to reduce the risk for ACL rupture.
- Since it is suggested that ACL rupture is a polygenic trait, our results suggest that identifying the genetic profile associated with ACL ruptures via haplotype analysis have become a worthy alternative to single-locus analysis.

Acknowledgments

There has been no financial assistance with the project.
References


### Table 1. Association between the presence of either $COL1A1$ -1997G/T (rs1107946) or $COL1A1$+1245G/T (rs1800012) polymorphisms and incidence of ACL rupture

<table>
<thead>
<tr>
<th>SNP</th>
<th>Group</th>
<th>HWE</th>
<th>Genotype distribution</th>
<th>p value</th>
<th>Allele frequency</th>
<th>p value</th>
<th>OR (95%CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1997G/T</td>
<td>ACLR n=91</td>
<td>0.288</td>
<td>GG: 60 (65.9%)</td>
<td>0.228</td>
<td>G: 82.4</td>
<td>0.246</td>
<td>1.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GT: 30 (33.0%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TT: 1 (1.1%)</td>
<td>pD 0.182</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control n=143</td>
<td>0.726</td>
<td>GG: 107 (74.8%)</td>
<td></td>
<td>G: 86.4</td>
<td></td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GT: 33 (23.1%)</td>
<td>pD 0.563</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TT: 3 (2.1%)</td>
<td>pD 1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+1245G/T</td>
<td>ACLR n=91</td>
<td>0.201</td>
<td>GG: 65 (71.4%)</td>
<td>0.138</td>
<td>G: 85.7</td>
<td>0.232†</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GT: 26 (28.6%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TT: 0 (0%)</td>
<td>pR 0.084</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control n=143</td>
<td>0.578</td>
<td>GG: 96 (67.1%)</td>
<td></td>
<td>G: 81.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GT: 41 (28.7%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TT: 6 (4.2%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a) P values correspond to genotype distribution and allele frequency.

b) OR correspond to the odds ratio for the incidence of ACL rupture.

ACLR=anterior cruciate ligament rupture; HWE =Hardy-Weinberg Equilibrium; OR=odds ratio; CI=confident interval

pD and pR are two-sided Fisher’s exact test probabilities for dominant (TT+GT vs GG) and recessive (TT vs GT+GG) modes of inheritance of the minor allele (-1997T, +1245T), respectively.
Table 2. Pair-wise linkage disequilibrium (LD) and inferred haplotype frequencies for *COLIA1* -1997G/T and +1245G/T

<table>
<thead>
<tr>
<th>Gene</th>
<th>SNP1</th>
<th>SNP2</th>
<th>D’</th>
<th>r²</th>
<th>Haplotypes</th>
<th>Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>COLIA1</em></td>
<td>-1997G/T</td>
<td>+1245G/T</td>
<td>0.997†</td>
<td>0.036</td>
<td>G-G</td>
<td>67.9</td>
</tr>
<tr>
<td></td>
<td>(rs1107946)</td>
<td>(rs180012)</td>
<td></td>
<td></td>
<td>G-T</td>
<td>16.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T-G</td>
<td>15.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T-T</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Total</td>
<td>100.0</td>
</tr>
</tbody>
</table>

a) D’=the deviation of the observed frequency of a haplotype from the expected. D’=0.977 suggest on very high LD between *COLIA1* -1997G/T and *COLIA1*+1245G/T

b) Haplotypes are in decreasing frequency

† p<0.001
Table 3. Haplotypes frequencies in ACLR and controls (under a recessive mode of haplotype inheritance)

<table>
<thead>
<tr>
<th>Haplotypes</th>
<th>Frequency (%)</th>
<th>Hap.score</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>combined ACLR and Controls</td>
<td>ACLR</td>
<td>Controls</td>
</tr>
<tr>
<td>G</td>
<td>G</td>
<td>67.9</td>
<td>68.1</td>
</tr>
<tr>
<td>G</td>
<td>T</td>
<td>16.9</td>
<td>14.3</td>
</tr>
</tbody>
</table>

ACLR=anterior cruciate ligament rupture group; Hap.score= statistics score for haplotypes