Genetic Polymorphisms in *COL1A2* gene and the Risk of Tendinopathy: Case-Control Study

Polimorfismos genéticos no gene COL1A2 e o risco de tendinopatia: Estudo de caso-controle

Lucas Rafael Lopes[1,2] João Antônio Matheus Guimarães[3] Marcus Vinicius Galvão Amaral[3] Camili Gomes Pereira[1] Victor Soares Wainchtock[1] Rodrigo Araujo Goes[3] Vitor Almeida Ribeiro de Miranda[3] Jamila Alessandra Perini[1,2,3]

[1] Pharmaceutical Science Research Laboratory, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, RJ, Brazil
[2] Graduate Program in Public Health and Environment, Escola Nacional de Saúde Pública, Fundação Oswaldo Cruz, Rio de Janeiro, RJ, Brazil
[3] Research Division, Instituto Nacional de Traumatologia e Ortopedia Jamil Haddad, Rio de Janeiro, RJ, Brazil

Address for correspondence Jamila Alessandra Perini, PhD, Instituto Nacional de Traumatologia e Ortopedia (INTO), Research Division, Av. Brasil, 500, Caju, Rio de Janeiro, RJ, 20940-070, Brazil (e-mail: jamilaperini@yahoo.com.br).

Rev Bras Ortop 2023;58(3):478–486.

Abstract

Objective To evaluate the influence of polymorphisms on genes encoding type I collagen and the genetic susceptibility of tendinopathy.

Methodology Case-control study involving 242 Brazilian athletes from different sports modalities (55 cases of tendinopathy and 187 controls). The polymorphisms *COL1A1* (rs1107946) and *COL1A2* (rs412777, rs42524, and rs2621215) were analyzed by the TaqMan system. Odds ratio (OR) with their 95% confidence intervals (CIs) were calculated using a nonconditional logistic regression model.

Results The mean age was 24.0 ± 5.6 years old and 65.3% were men. Of the 55 cases of tendinopathy, 25.4% had > 1 affected tendon, the most frequent being patellar (56.3%), rotator cuff (30.9%) and elbow or hand flexors (30.9%). Age and amount of time of sports practice were associated with a higher chance of presenting tendinopathy (5 and 8 times, respectively). The frequency of variant alleles in control and case patients, respectively, was: *COL1A1* rs1107946 24.0 and 29.6%; *COL1A2* rs412777 36.1 and 27.8%; rs42524 17.5 and 25.9%; and rs2621215 21.3 and 27.8%. After adjusting for

Keywords
- athletes
- tendinopathy
- collagen type 1
- polymorphism, genetic

The multicenter work was developed at the Universidade do Estado do Rio de Janeiro (UERJ-ZO), Fundação Oswaldo Cruz (Fiocruz), and Instituto Nacional de Traumatologia e Ortopedia Jamil Haddad (INTO), Rio de Janeiro, RJ, Brazil.

received
January 19, 2022
accepted
September 12, 2022

DOI https://doi.org/10.1055/s-0042-1757959.
ISSN 0102-3616.

Thieme Revinter Publicações Ltda., Rua do Matoso 170, Rio de Janeiro, RJ, CEP 20270-135, Brazil

confounding factors (age and years of sports practice), *COL1A2* rs42524 and rs2621215 polymorphisms were associated with increased risk of tendinopathy (OR = 5.5; 95% CI = 1.2–24.6 and OR = 3.9; IC95% = 1.1–13.5, respectively). The haplotype *COL1A2 CGT* was associated with low risk for disease development (OR = 0.5; 95%CI = 0.3–0.9).
Conclusion Age (≥ 25 years old), time of sports practice (≥ 6 years) and polymorphisms in the *COL1A2* gene increased the risk of developing tendinopathy.

Resumo

Objetivo Avaliar a influência de polimorfismos nos genes que codificam o colágeno tipo I e a suscetibilidade genética da tendinopatia.
Metodologia Estudo caso-controle envolvendo 242 atletas brasileiros de diferentes modalidades esportivas (55 casos de tendinopatia e 187 controles). Os polimorfismos *COL1A1* (rs1107946) e *COL1A2* (rs412777, rs42524 e rs2621215) foram analisados pelo sistema TaqMan. As razões de chance (OR) com seus intervalos de confiança (IC) de 95% foram calculadas usando um modelo de regressão logística não-condicional.
Resultados A média de idade foi de 24,0 ± 5,6 anos e 65,3% eram homens. Dos 55 casos de tendinopatia, 25,4% apresentaram mais de um tendão acometido, sendo os mais frequentes o patelar (56,3%), o manguito rotador (30,9%) e o do cotovelo ou flexores das mãos (30,9%). A idade e o tempo de prática esportiva foram associados a uma maior chance de apresentar tendinopatia (5 e 8 vezes, respectivamente). A frequência dos alelos variantes nos controles e casos, respectivamente, foi: *COL1A1* rs1107946 24,0 e 29,6%; *COL1A2* rs412777 36,1 e 27,8%; rs42524 17,5 e 25,9%; e rs2621215 21,3 e 27,8%. Após ajuste pelos fatores de confundimento (idade e anos de práticas esportiva), os polimorfismos *COL1A2* rs42524 e rs2621215 foram associados a um risco aumentado de tendinopatia (OR = 5,5; IC95% = 1,2–24,6 e OR = 3,9; IC95% = 1,1–13,5, respectivamente). O haplótipo *COL1A2 CGT* foi associado a um baixo risco para desenvolvimento da doença (OR = 0,5; IC95% = 0,3–0,9).
Conclusão A idade (≥ 25 anos), o tempo de prática esportiva (≥ 6 anos) e polimorfismos no gene *COL1A2* aumentaram o risco de desenvolvimento da tendinopatia.

Palavras-chave
- atletas
- tendinopatia
- colágeno tipo 1
- polimorfismo genético

Introduction

Tendinopathy is a multifactorial tendon disease, representing 1 to 6% of musculoskeletal injuries in the general population and 20 to 50% in athletes.[1,2] Initially, there is a physiological change in the tendon tissue, which may worsen with continuous stress or failures in the healing process, causing inflammation, pain, and even tissue degeneration.[2–4] The prevalence of the disease varies between the different parts of the body and according to age, sex, and occupation.[2] In athletes, the affected tendons vary according to sports modality, but mainly affect the Achilles tendon, the patellar tendon, rotator cuff tendons and tendons that originate from the lateral epicondyle.[1,2,5] The interaction between mechanical and biological factors is associated with the development of tendinopathies.[2–4,6]

Tendons are composed of a dense extracellular matrix (ECM) with low cellularity, but with high content of structural proteins, mainly type I collagen, a protein with heterotrimer morphology, composed of two α1 chains and one α2.[2,4] This structure offers the tendon tissue unique mechanical properties in terms of elasticity and tension, allowing the transmission of load between muscle and bone, promoting joint movement. The musculotendinous and osteotendinous junctions have specific histological characteristics.[7,8] The α1 chains are encoded by the *COL1A1* gene, located on chromosome 17, and the α2 chain is encoded by the *COL1A2* gene, and is located on chromosome 7.[4,9] It has been observed that the reduction in gene transcription of *COL1A1* and *COL1A2* in mice altered the properties of tendon strength, elasticity, tension, and healing.[10]

Recent studies have shown the influence of single nucleotide polymorphism (SNP) on the susceptibility of tendinopathy.[2,3,6,11] Genetic polymorphism is characterized by the substitution, deletion, or insertion of nucleotides in DNA tape, occurring in > 1% of the population, which may result in changes in the expression or functionality of gene products, determining individual characteristics, including susceptibility to certain diseases and response to some medications. The possible combinations of two alleles that characterize an SNP can form three possibilities of genotypes (heterozygous, wild homozygous or variant), which may or may not differ in phenotypes.[12] The combination of strongly correlated alleles

on the same chromosome is generally inherited as a unit, called haplotypes, because there is no genetic recombination between sites.[13]

COL1A1 and *COL1A2* genes are polymorphic and may alter the expression or biological function of type 1 collagen.[9,11,14] The SNPs *COL1A1* rs1800012 ($C > A$) and rs1107946 ($G > T$), present in the promoter region, were associated with soft tissue injuries of the musculoskeletal system in South African athletes of various sports modalities and in a cohort of nonathletes from the United Kingdom.[9,15] To date, no study has evaluated the influence of SNPs of the *COL1A2* gene on tendinopathy. Thus, the aim of the present study was to evaluate the contribution of SNPs *COL1A1* rs1107946 and *COL1A2* rs412777, rs42524 and rs2621215 as factors associated with genetic susceptibility of tendinopathy in Brazilian athletes.

Methodology

A retrospective observational case-control study was conducted with 242 Brazilian athletes, approved by the Human Research Ethics Committee of a tertiary hospital of the Brazilian Unified Public Health System (number 81225817.0.0000.5273). All participants signed the Informed Consent Form (TCLE, in the Portuguese acronym) and self-completed a questionnaire previously validated by experts in the area, available online in a previous study.[1] At the end of data collection, a trained observer verified and confirmed with each athlete the information. The database was filled out by a trained researcher, with double verification by two other different researchers, to ensure the veracity of the information entered in the database.

The inclusion criteria were federated athletes aged between 18 and 45 years old and of different sports modalities, who were recruited from February 2018 to November 2019. Athletes with no data on musculoskeletal injury and/or who had no biological material collected were excluded.

►**Fig. 1** shows the flowchart of the 242 athletes included in the study; 187 athletes with no history of injury in the musculoskeletal system (control group) and 55 athletes reported clinically diagnosed tendinopathy (case group), confirmed with magnetic resonance imaging (MRI). All diagnoses of tendinopathy were evaluated by two specialized orthopedic surgeons, as described in previous studies, referring to the selection criterion of cases of tendinopathy.[3,6]

Genomic DNA was obtained from an oral mucosa sample collected by sterile and isolated swab with the QIAmp DNA Mini Kit extraction kit (Qiagen, Hilden, Germany), following the recommendations of the manufacturer. The analyses of the SNPs of interest were performed by the real-time polymerase chain reaction (RT-PCR) technique, using validated allelic discrimination assays using TaqMan system (Thermo Fisher Scientific, Waltham, MA, USA) as described in the literature.[16] The TaqMan system for allelic discrimination consists of a set of primers and oligonucleotide probes designed specifically for each target SNP. The 2 probes are marked with different fluorescence, allowing the identification of the two possible alleles (*COL1A1* rs1107946 $G > T$ and *COL1A2* rs412777 $A > C$, rs42524 $G > C$ and rs2621215 $T > G$) present in the sample of the individual. The fluorescence intensity is captured by the equipment, discriminating the genotype of each individual (►**Fig. 2**). To ensure the quality control of the analysis, for each SNP, two standardized positive controls of each genotype were used, in addition to two negative controls, as described in the literature.[6]

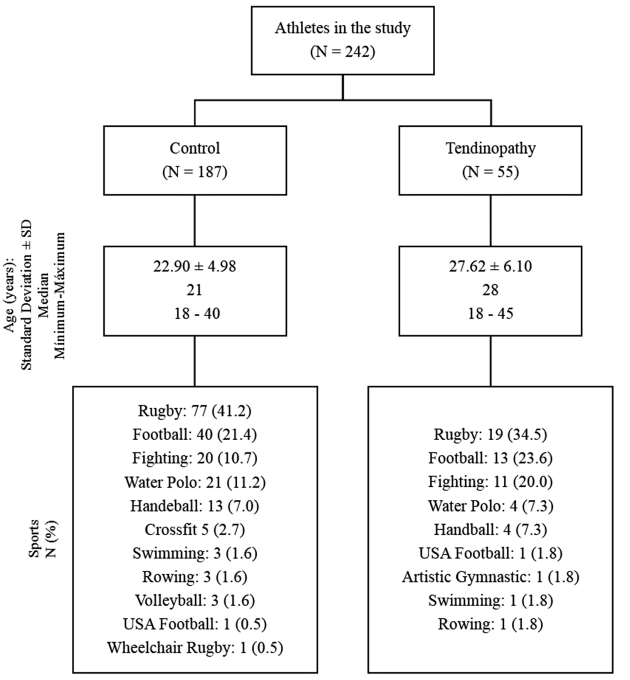

Fig. 1 Flowchart of the participants included in the study.

The sample size was used in the Epi Info 7 program, version 7.1.3. (http://wwwn.cdc.gov/epiinfo/html/downloads.htm) to detect differences between groups (case and control), assuming a ratio between cases and controls of 1:3 and odds ratio (OR) of 2.5 with a power of 0.8 and 5% of type I error.

The normal distribution of continuous variables in the studied population was verified by the Shapiro-Wilk test. Comparisons of continuous variables between tendinopathy cases and control groups were performed using the Student t-test, and they were presented as ± standard deviation (SD). Categorical variables, as well as the distribution frequencies of genotypes and alleles between the two groups were expressed in percentage and determined by the chi-squared test ($\chi 2$) or by the Fisher exact test, when applicable.

The frequencies of alleles and genotypes of *COL1A1* and *COL1A2* SNPs were determined by direct counting of alleles, and then the Hardy-Weinberg balance (HWE) was calculated. Haplotype patterns and binding imbalance coefficients (D' is the degree of imbalance in the module and R2 is the degree of correlation) were inferred using the Haploview program, as described in the literature.[16,17] The binding imbalance describes the combinations of alleles that occur in a sample due to the formation of haplotypes, calculated from the frequency of the individual alleles.[13] The combined analysis

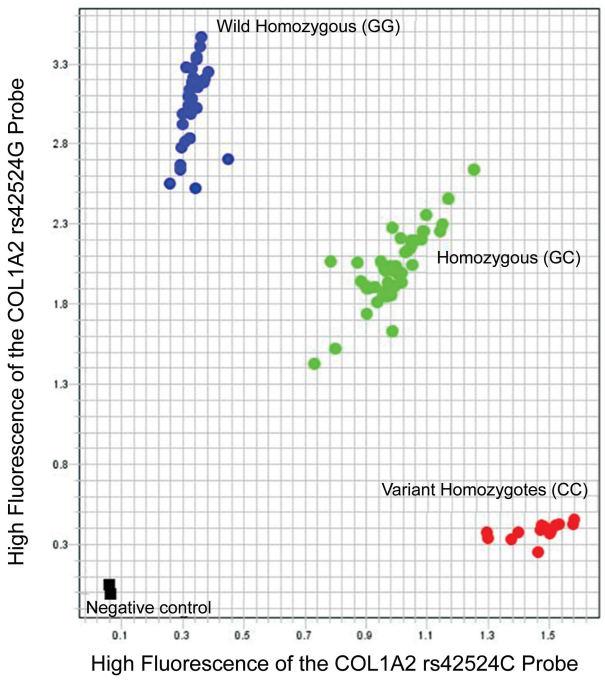

Fig. 2 Analysis of polymorphisms by real-time PCR using the Taqman system. **Label:** Example of discrimination of genotypes of *COL1A2* rs42524*G* > *C* polymorphism. The blue circles, which present high fluorescence of the *COL1A2* rs42524*G* probe, are the patients with wild homozygous genotype (*GG*). The red circles, which present high fluorescence of the *COL1A2* rs42524*C* probe are the variant homozygotes (*CC*) and the green circles are heterozygotes (*CG*), because they present fluorescence of both probes. Black squares are negative controls (water), which should not present PCR amplification and, consequently, fluorescence.

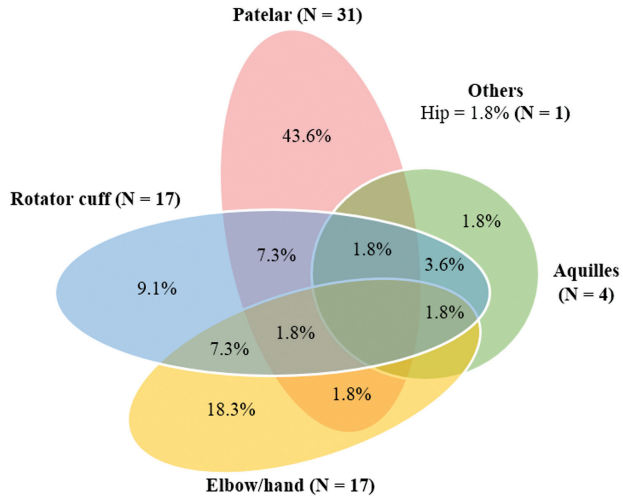

Fig. 3 Distribution of sites affected by the disease. **Label:** Of the athletes who reported more than one site of tendinopathy, 7.3% ($n = 4$) had rotator cuff and elbow/hand tendons, 7.3% ($n = 4$) were in the patellar and rotator cuff, 3.6% ($n = 2$) in the Achilles cuff and rotator cuff, 1.8% ($n = 1$) in the Achilles and elbow/hand, 1.8% ($n = 1$) in the patellar, rotator cuff, and elbow/hand, and 1.8% ($n = 1$) in the patellar, rotator, and Achilles cuff.

of alleles (haplotypes) is possibly clinically more relevant than the individual analysis of each isolated SNP.[17]

The magnitude of the association between the presence of polymorphisms and tendinopathy was estimated by Odds Ratio (OR), with their respective 95% confidence intervals (95%CIs), using a binary logistic regression model. The construction of the final model was based on the degree of statistical significance in the univariate analysis and on the biological importance of the variables studied, attributing an input significance level < 0.25 ($p \leq 0.25$) and remaining with an output level of 0.05 ($p \leq 0.05$). All statistical analyses were performed using the IBM SPSS Statistics for Windows version 20.0 (IBM Corp., Armonk, NY, USA).

Results

Of the 55 tendinopathy cases, 31 (56.3%) reported patellar tendon disease, 17 (30.9%) in rotator cuff tendons, 17 (30.9%) in the elbow or hand flexors, 4 (7.3%) in the Achilles tendon, and 1 (1.8%) in the middle gluteus tendon (►**Fig. 3**). In addition, 14 (25.4%) athletes reported > 1 tendon affected by the disease.

The mean age of the 242 athletes was 24.0 ± 5.6 years old (18 to 45 years old), the mean body weight was 74.7 ± 16.3 Kg (48 to 128 Kg), the mean height was 1.7 ± 0.1 m (1.51–2.05), the mean body mass index (BMI) was 24.7 ± 3.5 Kg/m^2 (17.5 to 41.8 Kg/m^2), the mean years of sports practice were 9.5 ± 6.1 (1 to 34), and the mean hours per training week was 13.0 ± 7.7 (1 to 42). The demographic, clinical, sports and training tendinopathy variables and controls are found in ►**Table 1**. The variables were evaluated to identify the possible confounding factors in the analysis of association between SNPs and tendinopathy. After applying the nonconditioned logistic regression model, only age and years of practice remained associated with the risk of tendinopathy in the studied population (Table 1).

The success rate of the genetic analysis was 98.8% for SNP *COL1A1* (rs1107946), 99.2% for *COL1A2* rs42524, 97.9% for rs412777 and rs2621215. The genotypic distributions of all SNPs were in HWE, indicating that the frequencies of alleles remained unchanged throughout the generations. The frequencies of the smallest variant allele of the studied SNPs are shown in ►**Fig. 4**, in which there was no significant difference in the distribution between the tendinopathy and control groups. After adjusting for confounding factors (age, sports modality, and years of sports practice), variant genotypes *COL1A2* rs42524 *CC* and *COL1A2* rs2621215 *GG* were associated with increased risk (approximately 5.5 and 4 times, respectively) of tendinopathy (►**Table 2**).

Eight *COL1A2* (rs412777, rs42524, rs2621215) SNPs haplotypes were inferred and the frequency of the *COL1A2 AGT* wild haplotype was 48.4 and 50.9% in controls and tendinopathy cases, respectively (►**Fig. 5**). After adjustment for confounding variables, variant *COL1A2 CGT* haplotype was negatively associated with tendinopathy (OR = 0.48; 95%CI = 0.25–0.93).

Discussion

Tendinopathy is a musculoskeletal system disorder very common in sports, with a prevalence of 10 to 50%, with

Table 1 Epidemiological and clinical analysis of the individuals included in the study ($n = 242$)

Variables	Control ($n = 187$)	Tendinopathy ($n = 55$)	*p-value*[a]	Adjusted OR (95%CI)[b]
Age (years old)	***n* (%)**			
≤ 20	84 (44.9)	7 (12.7)	< **0.001**	1[c]
21 to 24	46 (24.6)	10 (18.2)		2.63 (0.92–7.54)
25 to 29	33 (17.7)	18 (32.7)		5.53 (2.06–14.86)
≥ 30	24 (12.8)	20 (36.4)		5.16 (1.79–14.83)
Gender				
Female	63 (33.7)	21 (38.2)	0.54	1[c]
Male	124 (66.3)	34 (61.8)		0.64 (0.31–1.32)
Weight (Kg)				
≤ 64.00	49 (26.2)	12 (21.8)	0.53	1[c]
64.01 to 73.50	48 (25.7)	12 (21.8)		0.93 (0.34–2.56)
73.51 to 83.50	43 (23.0)	18 (32.7)		1.65 (0.65–4.21)
> 83.50	47 (25.1)	13 (23.7)		0.85 (0.32–2.26)
Height (m)				
≤ 1.67	50 (26.7)	17 (30.9)	0.67	1[c]
1.67 to 1.74	46 (24.6)	10 (18.2)		0.47 (0.17–1.27)
1.75 to 1.80	42 (22.5)	15 (27.3)		1.11 (0.45–2.77)
> 1.80	49 (26.2)	13 (23.6)		0.69 (0.27–1.77)
Alcohol consumption				
No	82 (43.9)	20 (36.4)	0.32	1[c]
Yes	105 (56.1)	35 (63.6)		1.19 (0.59–2.38)
Smoking				
No	177 (94.7)	53 (96.4)	0.61	1[c]
Yes	10 (5.3)	2 (3.6)		0.43 (0.08–2.30)
Years of sports				
≤ 5	69 (36.9)	3 (5.5)	**0.001**	1[c]
6 to 10	60 (32.1)	17 (30.9)		8.62 (2.30–32.25)
11 to 15	40 (21.4)	16 (29.1)		10.72 (2.82–40.77)
> 15	18 (9.6)	19 (34.5)		11.00 (2.78–43.53)
Weekly training hours				
≤ 8	62 (32.8)	14 (25.5)	0.27	1[c]
9 to 12	56 (30.2)	13 (23.6)		0.83 (0.32–2.13)
13 to 18	37 (19.8)	13 (23.6)		1.30 (0.49–3.43)
> 18	32 (17.2)	15 (27.3)		1.42 (0.53–3.83)

[a]p-value calculated by Fisher's χ^2 or exact test, when necessary; [b]OR = *Odds ratio,* odds ratio adjusted for age, years of training; CI = 95% confidence interval; [c]reference group.

the affected tendon varying according to the sports modality of the athlete.[1,2] The patellar tendon is among the most vulnerable and frequent to suffer excessive use injuries in the lower limbs.[1,18] In the present study, ~ 60% of the athletes reported patellar tendon disease, being more frequent in rugby, soccer, and football. The increase in tension and mechanical load during knee flexion generates cellular changes, with degeneration of collagen fibril that decreases the structural properties of the tissue.[19] Rotator cuff and elbow or hand flexors tendinopathies were the most frequent (~ 62%) in the upper limbs, especially in rowing, artistic gymnastics, swimming, wrestling sports and, water polo. The shoulder is widely used in sports that require greater demand and strength of the upper limbs, such as throwing movements in water polo and in attack and defense exercises required in combat sports.[1,20]

Age and time of sports practice were associated with the risk of the presence of tendinopathy, according to other

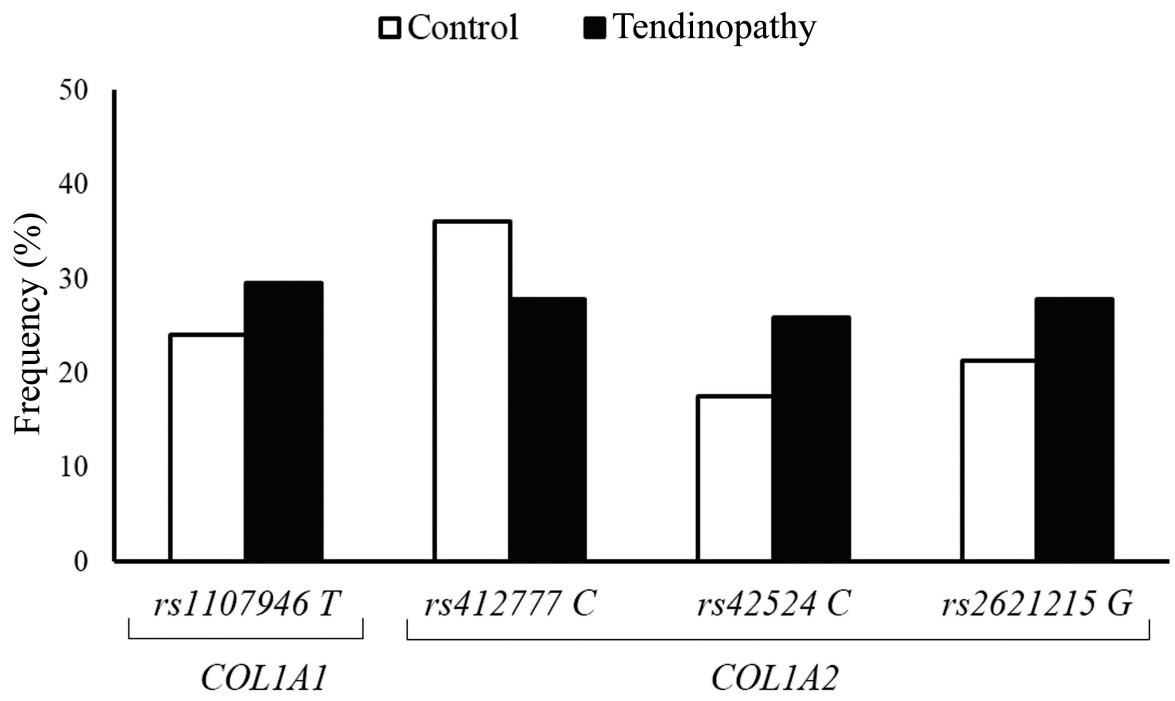

Fig. 4 Distribution of the allelic frequency of *COL1A1* and *COL1A2* polymorphisms in the studied population ($n = 242$). **Label:** There was no significant difference between the groups.

Table 2 Analysis of the association of *COL1A1* and *COL1A2* polymorphisms with the development of tendinopathy

Variables	Control ($n = 187$)	Tendinopathy ($n = 55$)	*p-value*[a]	Adjusted OR (95%CI)[b]
COL1A1 rs1107946[d]	*n* (%)			
GG	110 (59.5)	29 (53.7)	0.13	1[c]
GT	61 (33.0)	18 (33.3)		1.22 (0.57–2.61)
TT	14 (7.5)	7 (13.0)		2.60 (0.83–8.13)
COL1A2 rs412777[e]				
AA	75 (41.0)	28 (51.9)	0.27	1[c]
AC	84 (45.9)	22 (40.7)		0.70 (0.34–1.44)
CC	24 (13.1)	4 (7.4)		0.33 (0.09–1.16)
rs42524[f]				
GG	125 (67.2)	32 (59.3)	**0.03**	1[c]
GC	57 (30.6)	16 (29.6)		1.21 (0.57–2.55)
CC	4 (2.2)	6 (11.1)		**5.47 (1.22–24.58)**
rs2621215[e]				
TT	113 (61.7)	30 (55.6)	0.16	1[c]
TG	62 (33.9)	18 (33.3)		1.24 (0.59–2.62)
GG	8 (4.4)	6 (11.1)		**3.91 (1.13–13.48)**

[a]p-value calculated by the Fisher χ^2 or exact test, when necessary; [b]OR = *Odds ratio*, odds ratio adjusted for age, years of training; CI = 95% confidence interval; [c]reference group; [d]Information obtained from 239 athletes (185 controls and 54 cases of tendinopathy); [e]Information obtained from 237 athletes (183 controls and 54 cases of tendinopathy); [f]Information obtained from 240 athletes (186 controls and 54 cases of tendinopathy).

studies in the literature.[1,2] Advancing age can cause tendon aging, causing a change in tendon vascularization and, consequently, forming a less rigid structure vulnerable to undergoing tissue stress.[21] The present study found that older athletes, > 25 years old, and with > 10 years of sports practice were strongly associated with tendinopathy in Brazilian athletes (about 5 and 11 times, respectively), which corroborates a previous study, which observed an 8 times

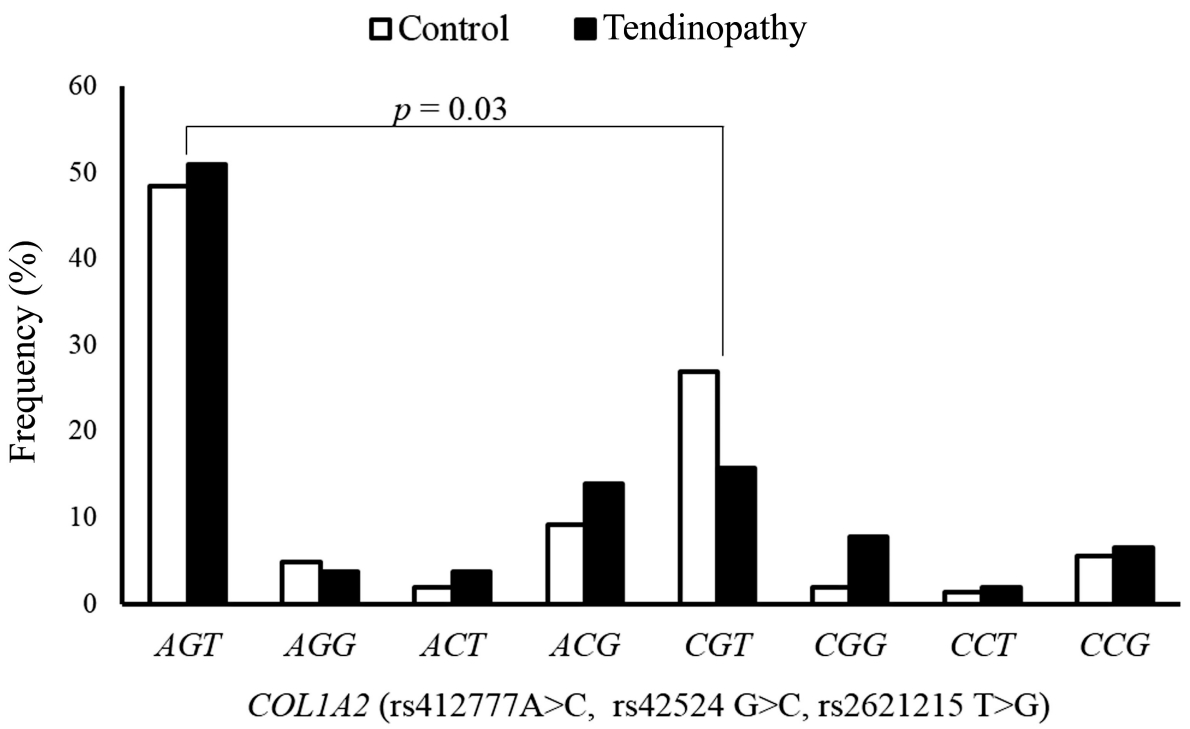

Fig. 5 Distribution of haplotypic frequency of *COL1A2* polymorphisms in the studied population ($n = 242$). **Label:** Eight *COL1A2* haplotypes, involving polymorphisms rs412777, rs42524, rs2621215, were found in the studied population. There was a significant frequency distribution difference of the AGT and *CGT haplotypes*.

risk associated with tendinopathy in the patellar tendon, rotator cuff and Achilles in Brazilian volleyball athletes > 24 years old and > 15 years of sports practice.[3] Also, the prevalence of the disease was observed in 17% of adult athletes compared with 5.6% in adolescents.[22] In addition, intense exposure to volume and frequency of training can cause recurrent tendon injuries, with an incidence of 0.12 injuries/1,000 hours of training or competition time.[23] Different training models, such as jump load and repetitive movements, may increase the risk of developing tendon injury.[2]

Genetic variations have been presented as intrinsic risk factors associated with the development of tendinopathy.[2,4,11,24] The gene-environment interaction of some modifiable risk factors, especially in athletes, related to the intensity and volume of training or competitions, and not modifiable, such as the genetic profile of the athlete, may result in a synergistic effect on the manifestation of tendon injury.[25] A systematic review study involving 17 studies on the influence of genes linked to collagen structure and tendon homeostasis observed the relevance of the contribution of polymorphisms in genes encoding collagen with susceptibility to Achilles tendinopathy.[26] Patients with shoulder instability, for example, have lower expression of *COL5A1*.[27]

The injured tendons have higher collagen expression of types I and III.[9,15] Considering type I collagen, *COL1A1* rs1107946 $G > T$ polymorphism is related to differences in transcriptional activity of the gene, in which the G allele presents higher transcription efficiency when compared to the T-allele.[28] Ficek et al.[15] observed that the *COL1A1 GT* (rs1107946-rs1800012) haplotype is associated with a lower risk of rupture of the anterior cruciate ligament in a group of professional football players, but did not find an association with tendinopathy, as observed in the present study, in which *COL1A1* rs1107946 SNP was not associated with the risk of tendinopathy. To date, there is no report of studies evaluating the influence of the *COL1A2* gene on the etiology of tendinopathy, making the present study a pioneer in describing the risk association of variant genotypes *COL1A2* rs42524 *CC* and rs2621215 *GG* with susceptibility to the disease. Recently, Perini et al. observed that variant genotypes *COL1A2* rs42524 *CC* and *COL1A2* rs2621215 *GG* presented a chance of ~ 5 times for noncontact rupture of the anterior cruciate ligament.[16] The substitution of the amino-acid alanine by proline (Ala > Pro) of *COL1A2* rs42524 $G > C$ SNP promotes a change in the stability of the triple collagen helix.[14] *COL1A2* rs2621215 $T > G$ SNP may interact with other functional polymorphisms and affect the removal of introns from this gene causing damage to collagen structure.[29] Thus, the presence of variant alleles *COL1A2* rs42524 *C* and rs2621215 *G* may produce a less flexible collagen, making it more vulnerable to tissue stress.

COL1A2 CGT haplotype, formed only by the variant allele of SNP rs412777*C*, and the other wild two (rs42524*G* and rs2621215*T*), presented a protective factor for the development of tendinopathy. The exchange of the nucleotide adenine by cytosine in exon 25 does not promote the exchange of the aminoacid proline at position 392 of the α2 chain of type

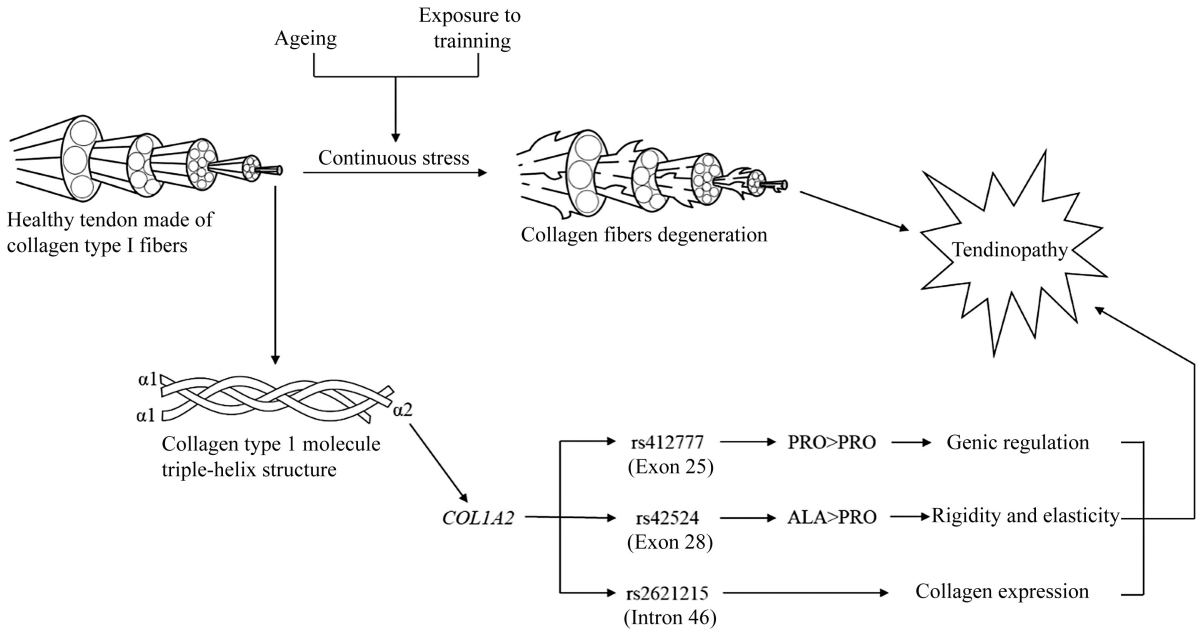

Fig. 6 Hypothesis of mechanism of the development of tendinopathy in the presence of intrinsic factors (age and polymorphisms of *COL1A1* and *COL1A2*) and extrinsic (years of sports training).

1 collagen. The mechanisms by which the different *COL1A2* gene polymorphisms affect its expression and/or function are not yet fully known. However, SNP rs412777 may be in imbalance in connection with a functional mutation and may influence gene expression and interfere with the mechanism of removal of introns and union of the exons, which are fundamental for protein formation.[30]

Finally, based on the results found, a mechanism of hypothesis of the gene-environment interaction between the age of the athlete, the time of practice in sport and the contribution of *COL1A2* SNPs (rs412777, rs42524, rs2621215) in the development of tendinopathy was suggested. Factors such as aging and mechanical stress can contribute to increased demand or irregular distribution of the tensile load, causing tissue changes.[2] Type 1 collagen fibers present differences between individuals due to the presence of polymorphisms in the genes (*COL1A1* and *COL1A2*), which encode the protein.[9,14,15] *COL1A2* SNPs can influence gene expression, increasing stiffness and decreasing tissue elasticity.[14,30] Thus, we suggest that gene-environment interaction (*COL1A2* SNPs added to aging and high exposure to training) can contribute synergistically to the development of tendinopathy in athletes (►**Fig. 6**).

This is a pioneering and innovative study, which can contribute to the biotechnological and clinical advancement in personalized medicine; however, the sample size of the tendinopathy group of cases was the main limitation of the study, not allowing stratified analysis of the disease in tendons with different biomechanical activities. Moreover, the present study did not differentiate insubstantial tendinopathies from those of insertional ones or that occur at the myotendinous junction, which should be a reason for investigation in future analytical studies because they are diseases of a diverse clinical spectrum. However, we have the science that even with differences in anatomical sites, all tendons are composed of type 1 collagen. We also have a variety of sports modalities, which differ in the type of training, biomechanical demands and, consequently, in the tendons affected by the disease. However, the possible confounding variables were inserted in the logistic regression model to evaluate the real influence of SNPs on the development of tendinopathy. Thus, the results of the present study may contribute as a database for future studies in order to build a database with different populations so that it is possible to identify modifiable and nonmodifiable risk factors associated with the development of tendinopathy. Individualized programs of injury prevention using genetic information can contribute to the promotion of the health and well-being of individuals, besides being useful diagnostic tools in the clinical practice of the orthopedist.

Conclusion

Athletes > 25 years old, with > 6 years of sports practice, with genotypes *COL1A2* rs42524 *CC* and rs2621215 *GG* have a higher risk of developing tendinopathy, while the *COL1A2 CGT* variant haplotype presented a protective effect on the development of the disease.

Contribution of the Authors

The authors contributed individually and significantly to the development of this article: Perini J. A., Lopes R. L., and Goes R. A. participated in the conception and design of the study. Perini J. A., Lopes R. L., and Goes R. A. gathered the data and developed the database. Perini J. A., Lopes R. L., Pereira C. G., and Wainchtock V. S. carried out the

experiments and statistical analyses. Perini J. A., Lopes R. L., Amaral M. V. G., and Miranda V. A. R. performed data analysis and interpretation. Perini J. A., Lopes R. L., Pereira C. G., and Wainchtock V. S. wrote the manuscript. Guimarães J. A. M. and Amaral M. V. G. performed the critical revision of the manuscript for important intellectual content. All authors read and approved the final manuscript.

Financial Support
Carlos Chagas Filho Foundation for Research Support of the State of Rio de Janeiro (FAPERJ) and the National Council for Scientific and Technological Development (CNPq).

Conflict of Interests
The authors have is no conflict of interests to declare.

References

1 Goes RA, Lopes LR, Cossich VRA, et al. Musculoskeletal injuries in athletes from five modalities: a cross-sectional study. BMC Musculoskelet Disord 2020;21(01):122
2 Millar NL, Silbernagel KG, Thorborg K, et al. Tendinopathy. Nat Rev Dis Primers 2021;7(01):1
3 Salles JI, Lopes LR, Duarte MEL, et al. Fc receptor-like 3 (-169T>C) polymorphism increases the risk of tendinopathy in volleyball athletes: a case control study. BMC Med Genet 2018;19(01):119
4 Collins M, Raleigh SM. Genetic risk factors for musculoskeletal soft tissue injuries. Med Sport Sci 2009;54:136–149
5 Ackermann PW, Renström P. Tendinopathy in sport. Sports Health 2012;4(03):193–201
6 Lopes LR, de Miranda VAR, Guimarães JAM, et al. Association of TNF-α -308G > A polymorphism with susceptibility to tendinopathy in athletes: a case-control study. BMC Sports Sci Med Rehabil 2021;13(01):51
7 Buchanan CI, Marsh RL. Effects of exercise on the biomechanical, biochemical and structural properties of tendons. Comp Biochem Physiol A Mol Integr Physiol 2002;133(04):1101–1107
8 Nourissat G, Berenbaum F, Duprez D. Tendon injury: from biology to tendon repair. Nat Rev Rheumatol 2015;11(04):223–233
9 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
10 Guerquin MJ, Charvet B, Nourissat G, et al. Transcription factor EGR1 directs tendon differentiation and promotes tendon repair. J Clin Invest 2013;123(08):3564–3576
11 Cohen C, Figueiredo EA, Belangero PS, Andreoli CV, Leal MF, Ejnisman B. Genetic Aspects in Shoulder Disorders. Rev Bras Ortop 2020;55(05):537–542
12 Karki R, Pandya D, Elston RC, Ferlini C. Defining "mutation" and "polymorphism" in the era of personal genomics. BMC Med Genomics 2015;8:37
13 Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. Science 2008;322(5903):881–888
14 Yoneyama T, Kasuya H, Onda H, et al. Collagen type I alpha2 (COL1A2) is the susceptible gene for intracranial aneurysms. Stroke 2004;35(02):443–448
15 Ficek K, Cieszczyk P, Kaczmarczyk M, et al. Gene variants within the COL1A1 gene are associated with reduced anterior cruciate ligament injury in professional soccer players. J Sci Med Sport 2013;16(05):396–400
16 Perini JA, Lopes LR, Guimarães JAM, 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(01):154
17 Perini JA, Cardoso JV, Berardo PT, et al. Role of vascular endothelial growth factor polymorphisms (-2578C>A, -460 T>C, -1154G>A, +405G>C and +936C>T) in endometriosis: a case-control study with Brazilians. BMC Womens Health 2014;14:117
18 Florit D, Pedret C, Casals M, Malliaras P, Sugimoto D, Rodas G. Incidence of Tendinopathy in Team Sports in a Multidisciplinary Sports Club Over 8 Seasons. J Sports Sci Med 2019;18(04):780–788
19 Aicale R, Oliviero A, Maffulli N. Management of Achilles and patellar tendinopathy: what we know, what we can do. J Foot Ankle Res 2020;13(01):59
20 Hams AH, Evans K, Adams R, Waddington G, Witchalls J. Shoulder internal and external rotation strength and prediction of subsequent injury in water-polo players. Scand J Med Sci Sports 2019;29(09):1414–1420
21 Li HY, Hua YH. Achilles Tendinopathy: Current Concepts about the Basic Science and Clinical Treatments. BioMed Res Int 2016;2016:6492597
22 Cassel M, Baur H, Hirschmüller A, Carlsohn A, Fröhlich K, Mayer F. Prevalence of Achilles and patellar tendinopathy and their association to intratendinous changes in adolescent athletes. Scand J Med Sci Sports 2015;25(03):e310–e318
23 Hägglund M, Zwerver J, Ekstrand J. Epidemiology of patellar tendinopathy in elite male soccer players. Am J Sports Med 2011;39(09):1906–1911
24 Assunção JH, Tenrreiro BF, Gracitelli MEC, Malavolta EA, Ferreira Neto AA. Family predisposition for rotator cuff tear and other tendinopathies - a case-control study. Rev Bras Ortop 2020;55(04):470–475
25 Andarawis-Puri N, Flatow EL, Soslowsky LJ. Tendon basic science: Development, repair, regeneration, and healing. J Orthop Res 2015;33(06):780–784
26 Kozlovskaia M, Vlahovich N, Ashton KJ, Hughes DC. Biomedical risk factors of achilles tendinopathy in physically active people: a systematic review. Sports Med Open 2017;3(01):20
27 Belangero PS, Leal MF, de Castro Pochini A, Andreoli CV, Ejnisman B, Cohen M. Profile of collagen gene expression in the glenohumeral capsule of patients with traumatic anterior instability of the shoulder. Rev Bras Ortop 2014;49(06):642–646
28 García-Giralt N, Enjuanes A, Bustamante M, et al. In vitro functional assay of alleles and haplotypes of two COL1A1-promoter SNPs. Bone 2005;36(05):902–908
29 Meng Q, Hao Q, Zhao C. The association between collagen gene polymorphisms and intracranial aneurysms: a meta-analysis. Neurosurg Rev 2019;42(02):243–253
30 Suuriniemi M, Mahonen A, Kovanen V, Alén M, Cheng S. Relation of PvuII site polymorphism in the COL1A2 gene to the risk of fractures in prepubertal Finnish girls. Physiol Genomics 2003;14(03):217–224