Genetic Variants Involved in Extracellular Matrix Homeostasis Play a Role in the Susceptibility to Frozen Shoulder: A Case-Control Study.

Carina Cohen,[1] Mariana F. Leal,[1,2] Leonor C. Loyola,[1,2] Sidney E. B. Santos,[3] Ândrea K. C. Ribeiro-dos-Santos,[3] Paulo S. Belangero,[1] Eduardo A. Figueiredo,[1] Andre Wajnsztejn,[1] Adrielle M. de Oliveira,[1,2] Marlia C. Smith,[2] Carlos V. Andreoli,[1] Alberto de Castro Pochini,[1] Moises Cohen,[1] Benno Ejnisman,[1] Flávio Faloppa[1]

[1]Departamento de Ortopedia e Traumatologia, Universidade Federal de São Paulo, São Paulo, SP, Brazil, [2]Disciplina de Genética, Departamento de Morfologia e Genética, Universidade Federal de São Paulo, São Paulo, SP, Brazil, [3]Laboratório de Genética Humana e Médica, Instituto de Ciências Biológicas, Universidade Federal do Pará, Belém, PA, Brazil

Received 28 April 2018; accepted 26 December 2018
Published online in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/jor.24228

ABSTRACT: Frozen shoulder is a condition of loss of active and passive motion as result of inflammatory contracture and fibrosis of the joint capsule. We hypothesize that genetic variants in genes involved in these processes such as genes that play a role in extracellular matrix homeostasis (collagens, glycoproteins, genes involved in TGFβ signaling, and metalloproteinases and its inhibitors) may contribute to the susceptibility to frozen shoulder. We evaluated eighteen SNPs of genes involved in extracellular matrix homeostasis in 186 cases ($N_{females} = 114$; $N_{males} = 72$) of frozen shoulder and 600 age-matched controls ($N_{females} = 308$; $N_{males} = 292$). Multivariate logistic regressions were carried out with age, gender, genetic ancestry, and common comorbidities as covariates. Carriers of the C allele of *MMP13* rs2252070 and G/G *MMP9* (rs17576 A>G/rs17577 G>A) haplotype may have an increased risk of frozen shoulder ($p = 0.002$, OR = 1.64, 95%CI = 1.20–2.26, and $p = 0.046$, OR = 1.40, 95%CI = 1.01–1.95, respectively), especially in females ($p = 0.005$, OR = 1.91, 95%CI = 1.22–2.99, and $p = 0.046$, OR = 1.59, 95%CI = 1.01–2.51, respectively). In females, the G allele of *MMP9* rs17576 tended to contribute to the susceptibility to the studied disease ($p = 0.05$, OR = 1.51, 95%CI = 0.97–2.33). In contrast, the presence of the C allele of *TGFB1* rs1800470 seems to be associated with a reduced risk ($p = 0.04$, OR = 0.47, 95%CI = 0.23–0.96) while the GG-genotype of *TGFBR1* rs1590 was associated with increased risk ($p = 0.027$, OR = 4.11, 95%CI = 1.17–14.38) to frozen shoulder development in males. Thus, we identified genetic variants that were independent risk factors that can aid in the risk assessment of frozen shoulder reinforcing the involvement of MMP and TGFβ signaling in disease development. © 2019 Orthopaedic Research Society. Published by Wiley Periodicals, Inc. J Orthop Res 37:948–956, 2019.

Keywords: frozen shoulder; susceptibility; polymorphism; extracellular matrix; metalloproteinases

Frozen shoulder (FS), also known as adhesive capsulitis, is a debilitating condition in which patients present with limited active and passive shoulder motion. The loss of motion is a result of an inflammatory contracture of the joint capsule, which seems to lead to capsular fibrosis.[1]

FS occurs in 3–5% of the general population.[2] It is found predominantly among women aged 40–60.[3] In diabetes patients, the FS incidence increases to 10.8–29%.[4–6] Dyslipidemia[7,8] and high blood pressure[7] seem to increase the risk of this shoulder injury. Hypothyroidism is also a FS risk factor, especially in females.[9,10] Possible racial differences have been described.[4]

The etiology of FS remains unclear. Genetic factors may be involved in susceptibility, and its heritability is estimated to be 42%.[11] Single nucleotide polymorphisms (SNPs) are the most studied type of genetic variations, and knowing which SNPs are associated with complex diseases provides insights into underlying etiologic mechanisms.[12] Xu et al. originally described that the rare variant of the metalloproteinase *MMP3* rs650108 was associated with increased susceptibility to FS in a Chinese Han population.[13]

The shoulder capsule is composed of cellular and fibrous elements. Types I (encoded by *COL1A1* and *COL1A2*s), III (encoded by *COL3A1*), and V fibrillar collagens (mainly encoded by *COL5A1* and *COL5A2* in joint tissues) are the most common types present in the extracellular matrix (ECM) of this tissue.[14] Matrix metalloproteinases (MMPs) are enzymes responsible for tissue remodeling and ECM degradation.[15] These enzymes are classified based on their substrate preference, including collagenases (e.g., MMP1 and MMP13), stromelysins (e.g., MMP3), and gelatinases (e.g., MMP2 and MMP9).[15] MMPs are inhibited by tissue inhibitor of metalloproteinases (TIMP), including TIMP2.[16] A balance between MMPs and TIMPs is necessary for tissue maintenance and remodeling. As described above, a SNP of *MMP3* (rs650108) may play a role in the FS etiology.[13]

Rodeo et al. previously suggested that cytokines, such as transforming growth factor beta (TGFβ), may play a key role in the inflammatory and fibrotic processes that occur in FS.[17] TGFβ activity is mediated by signaling receptors such as TGFβ receptor I (encoded by the *TGFBR1* gene).[18] TGFβ regulates and is regulated by several ECM proteins, including collagens, fibronectin 1 (FN1) and tenascin C (TNC).[19,20] However, both FN1 and TNC glycoproteins act in TGFβ modulation.[21,22] Increased *TNC* and *FN1* mRNA expression is a possible a marker of capsule injury.[23] Upregulation of *TGFBR1* seems also to be

Carina Cohen and Mariana F. Leal contributed equally to this work.

Grant sponsor: Conselho Nacional de Desenvolvimento Cientfico e Tecnológico; Grant sponsor: Fundação de Amparo à Pesquisa do Estado de São Paulo; Grant numbers: 2011/22548-2, 2012/14768-5, 2013/23332-9, 2015/02285-8, 2016/01392-8, 2016/15785-1.

Correspondence to: Mariana F. Leal (T: +55-11-55716621; F: +55-11-55716621; E-mail: Mariana.morf@epm.br)

dependent on symptom duration, and thus TGFβ signaling may be involved in FS.[23] Because *TNC*, *FN1*, and *TGFBR1* may play a role in FS by contributing to capsule inflammation and fibrosis, we hypothesize that genetic variants in these genes, as well as in *COL1A1* and *COL5A1*, *TGFB1*, *TIMP2*, and *MMP*s, may contribute to the susceptibility to this shoulder condition.

Here, we investigated whether 18 SNPs in genes involved in ECM structure and homeostasis may contribute to the risk of idiopathic FS. To elucidate whether genetic variants may be independent risk factors, we included in our models genetic ancestry and epidemiological risk factors for the disease to avoid population stratification or an intrinsic association between the studied SNPs and common comorbidities previously associated with FS incidence. This approach was chosen to identify SNPs that may add predictive information beyond the clinical variables that are currently easily accessed in clinical practice. We originally demonstrated the involvement of SNPs in *MMP2*, *MMP9*, *MMP13*, *TGFB1*, and *TGFBR1* in susceptibility to FS.

MATERIALS AND METHODS

Subject Selection

This is a case–control study involving 786 individuals. Patients and controls were recruited at the Hospital São Paulo of the Universidade Federal de São Paulo, Brazil. Each subject agreed to participate by signing a written consent form before data and sample collection approved by the ethics committee of the Universidade Federal de São Paulo (approval number:CEP 1918/11).

Between 2012 and 2016, 186 patients with idiopathic FS or history of this disease were recruited ($N_{\text{females}} = 114$; $N_{\text{males}} = 72$). During the diagnostic clinical evaluations, the patients presented pain, loss of motion and severe limitations during daily activities; functional restriction of active and passive shoulder motion and no history of trauma or previous shoulder pathologies. Magnetic resonance imaging (MRI) was used to exclude secondary stiff shoulder. Exclusion criteria were: Generalized arthritis; previous shoulder compromise, such as major trauma, fracture, rotator cuff tear, calcifying tendonitis, or shoulder instability; and superior labral anterior and posterior (SLAP) lesions.

The control group consisted of 600 age-matched subjects (±1 years) and given an a priori estimated ratio of 1:3 cases–controls ($N_{\text{females}} = 308$; $N_{\text{males}} = 292$). All controls were physically active in their daily lifestyle; however, sports participation or occupational activities were not used as inclusion or exclusion criteria in both case and control groups. Exclusion criteria were: History of FS; history of other shoulder disease; family history of shoulder injuries; genetic syndromes.

All patients answered a questionnaire concerning gender, age, smoking habits and other medical conditions (Table 1 and Supplementary Table S1). High blood pressure, dyslipidemia, diabetes and thyropathy (mainly hypothyroidism) were the main comorbidities observed in cases and controls.

DNA Extraction

Genomic DNA was extracted from fresh white blood cells using Gentra Puregene Blood Kit (Qiagen, Valencia, CA). DNA concentration and quality were determined using a Nanodrop ND-1000 (Thermo Scientific, Wilmington, DE) and stored at −20°C.

Molecular Ancestry

To identify SNPs that may be associated with the risk of FS, we initially evaluated their frequency in a subpopulation in which we determined the genetic ancestry to avoid population stratification bias. Genetic ancestry was determined for 164 randomly selected cases and 460 gender-matched controls whose age was the same or up to 2 years older than selected cases. This older age was used because they were more exposed to possible environmental factors that can contribute to FS risk and did not develop the condition (Table 1). Subjects were genotyped for a set of 61 biallelic

Table 1. Distribution of Clinical Variables and Genetic Ancestry in Subjects With and FS

	All Subjects ($n = 786$)			All Subjects With Genetic Ancestry Data ($n = 624$)		
Variables	Controls ($n = 600$)	Cases ($n = 186$)	p-value	Controls ($n = 460$)	Cases ($n = 164$)	p-value
Age (years; mean ± SD)	53.49 ± 11.73	52.66 ± 9.04	0.373	54.49 ± 11.01	52.50 ± 9.27	0.039*
Gender [n(%)]						
Female	308 (51.3)	114 (61.2)	0.02*	263 (57.2)	97 (59.1)	0.729
Male	292 (48.7)	72 (38.8)		197 (42.8)	67 (40.9)	
Genetic ancestry (%; mean ± SD)						
African component	–	–	–	23 ± 21	10 ± 13	<0.001*
Ameridian component	–	–	–	11 ± 10	8 ± 6	<0.001*
European component	–	–	–	62 ± 24	75 ± 25	<0.001*
Smoker [n(%)]	62 (10.3)	17 (9.1)	0.392	52 (11.3)	15 (9.1)	0.282
High blood pressure [n(%)]	230 (38.4)	43 (23.1)	<0.001*	176 (38.3)	40 (24.4)	0.001*
Dyslipidemia [n(%)]	52 (8.7)	51 (27.4)	<0.001*	39 (8.5)	45 (27.4)	<0.001*
Diabetes [n(%)]	124 (20.7)	24 (12.9)	0.012*	101 (22.0)	21 (12.8)	0.008*
Thyropathy [n(%)]	62 (10.4)	57 (30.6)	<0.001*	49 (10.6)	48 (29.3)	<0.001*
Hypothyroidism [n(%)]	35 (5.8)	39 (21)	<0.001*	30 (6.5)	31 (18.9)	<0.001*

n: Number of individuals; SD: standard deviation. *Significant difference between groups ($p < 0.05$) by t-test for independent samples (quantitative variables) or X^2 test (categorical variables).

validated short insertion/deletion polymorphisms (INDELs) using the ABI PRISM 3130 Genetic Analyzer (Applied Biosystems, Foster Ciyt, CA) and analyzed with GeneMapper v3.2 software (Applied Biosystems) in a single center, as previously described.[24] The ABIGS LIZ-500 ladder (Applied Biosystems) was used as a reference. Standards of known size were included in each assay for quality control.[24]

The admixture model assumes that each subject inherits part of his or her ancestral markers from ancestral populations. Based on this assumption, the results were plotted against the three parental populations (African, Amerindian, and European) that constitute the Brazilian population to perform ancestry stratification. Structure v2.3.4 software was used to estimate admixture.

Genotyping

Details of the studied genetic variants are shown in Supplementary Table S2. All SNPs were in genes associated with ECM structure and homeostasis and/or fibrosis process. SNPs were selected based on at least one of the following criteria: Be functional; previously associated with the risk of orthopedic disease/injuries; previously associated with the risk of comorbidities that are related with FS development (see references in Supplementary Table S2). The minor allele frequency (MAF) based on the 1000 genomes project was used to selected SNPs for each we could detect an estimated 1.4–2.1genotype relative risk with 80% of expected power at the 5% significance level in our sample (http://csg.sph.umich.edu//abecasis/cats/gas_power_calculator/index.html).

Genotyping was performed by 5′ nuclease PCR using the TaqMan allelic discrimination system (Supplementary Table S2; Life Technologies, Grand Island, NY) in a ViiA 7 Real-Time PCR System (Life Technologies) in a single center. Blank and known samples representative of each genotype were included in all runs. Genotype calling was performed using TaqMan Genotyper Software (Life Technologies). Analyses were performed in single-replicates; however, replicates were performed if possible low-quality data was detected. Genotype calling, as well as genetic ancestry analysis, were performed by investigators blind for clinical data during the analyses. This study followed the The STrengthening the REporting of Genetic Association studies (STREGA) recommendations.[25]

Models

Designating the minor (rare) allele as m and the major allele as M, the association between the SNPs and FS risk was assessed under the following genetic models: A) dominant model (Mm + mm vs. MM), where the presence of the rare allele is involved in an increased or reduced risk of FS; B) recessive model (mm vs. Mm + MM), where the presence of two copies of the rare allele is necessary to modify the phenotype; and C) additive model in which samples were coded as 0 (MM), 1 (Mm) or 2 (mm). In the additive model, individuals with two copies of the rare allele would be more susceptible to the disease than individuals with one copy; however, individuals with one copy of the rare allele would have a greater risk of disease than individuals without the allele rare. In all analyses, the common homozygote genotype in the control population was defined as the reference category.

Statistical Analyses

The chi-square test was used to compare gender, smoking habits and the presence of medical conditions between patients with FS and controls. T-test was used to compare age and genetic ancestry between cases and controls. These analyses were performed using SPSS v.18 software (IBM, Armonk, NY).

Samples with missing data (genotype) were excluded from statistical analysis. Allele and genotype frequencies were calculated for polymorphisms, and the chi-square test was used to investigate deviation from Hardy-Weinberg Equilibrium (HWE).

The association between the case–control status and each individual SNP, as measured by the odds ratio (OR) and its corresponding 95% confidence interval (CI), was estimated using unconditional logistic regression. First, the logistic regression was performed using sex and age as covariates. p values are usually adjusted in studies aiming to identify SNPs associated with the risk of joint injuries, including in the previous study with FS.[13]

Second, European and African ancestry were included as covariates, as this correction is essential to avoid bias due to an association between a SNP and genetic ancestry in an admixed population. To avoid multicollinearity, Amerindian ancestry was not included in the model because it could be inferred based on the percentages of the other ancestries.

Third, logistic regression was performed using sex, age, genetic ancestry, high blood pressure, dyslipidemia, diabetes and hypothyroidism (the main thyropathy in our population) as covariates. These covariates were previously described as epidemiological risk factors for FS, and their distributions differed between cases and controls in our population (Table 1). As a validation step, logistic regression was performed for the SNPs identified as independent risk factors in the previous analyses (adjusted by genetic ancestry) with all recruited individuals ($n = 786$). In this step, sex, age, and the main comorbidities were included as covariates. Genetic ancestry was not included. A final model considered more than one SNP as a possible risk factor.

The models described above were also evaluated separately among females and males. All model were evaluated using SNPstat software.[26]

Linkage disequilibrium (LD) and the associations between haplotypes of selected SNPs of a gene and the risk of FS were estimated using SNPstat software.[26] First, p values were adjusted for gender, age, and genetic ancestry. Then, p values were adjusted for gender, age, and comorbidities in the analyses involving all recruited subjects. To allow the investigation of a larger number of samples necessary for haplotype analyses, genetic ancestry was not included as covariable in the validation set.

A p-value of 0.05 was considered statistically significant.

RESULTS

Medical Conditions and Genetic Ancestry

The most frequent medical conditions in cases and controls were high blood pressure, dyslipidemia, diabetes and thyropathy, mainly hypothyroidism (Table 1 and full data at Supplementary Table S1). Dyslipidemia and hypothyroidism were more frequent in patients with FS than in controls ($p < 0.001$). Conversely, high blood pressure and diabetes were more common in controls ($p < 0.05$). Females with FS also presented increased frequency of dyslipidemia and hypothyroidism and reduced frequency of high blood pressure and diabetes in relation to the recruited controls (all $p < 0.001$; Supplementary Table S1). The frequency of dyslipidemia was significantly higher

among males with FS in relation to controls ($p < 0.001$; Supplementary Table S1).

Males with FS had a higher frequency of high blood pressure ($p = 0.002$) and diabetes ($p = 0.007$) than females. In contrast, hypothyroidism was more frequent in females than in males ($p = 0.004$; Table 1).

All subjects had mixed genetic ancestry. European ancestry was predominant in both groups, followed by African and Amerindian components (Table 1 and Supplementary Table S1). Only 17 individuals had less than 80% of European, African and Amerindian ancestry, which could be result of other population ancestry, such as Asian.

European component was more frequent among cases than controls ($p < 0.001$; Table 1 and Supplementary Table S1). In contrast, both Amerindian and African components were more frequent among the controls ($p < 0.001$; Table 1 and Supplementary Table S1).

SNPs and FS Susceptibility

Alleles and genotypes were in HWE, with the exceptions of *COL1A1* rs1800012 and *COL5A1* rs12722, which deviated significantly from HWE in the studied population ($p = 0.027$ and $p = 0.015$; respectively) and mainly in the controls ($p = 0.006$ and $p = 0.037$, respectively). Although these SNPs could be considered in HWE if a Bonferroni correction was applied after multiple comparisons, the results of *COL1A1* rs1800012 and *COL5A1* rs12722 should be considered with care.

The genotype distributions of the studied SNPs for the cases and controls are shown in Supplementary Tables S3–S5. SNP *MMP2* rs243865 ($p = 0.049$, OR = 1.39, 95% CI = 1.00–1.93; additive model) and *MMP3* rs522616 ($p = 0.04$, OR = 0.45, 95%CI = 0.21–0.98; recessive model) were associated with FS development when adjusted for age and gender (Supplementary Table S3). However, no significant association between these SNPs and disease susceptibility was detected after adjusting for genetic ancestry.

The rare allele of *MMP13* rs2252070 was an independent risk factor for FS, including in an additive model that adjusted for gender, age, genetic ancestry and the main comorbidities ($p = 0.002$, OR = 1.64, 95%CI = 1.20–2.26; Supplementary Table S3). Therefore, both homozygous and heterozygous individuals for the C-allele of *MMP13* rs2252070 may be at increased risk of developing FS; however, CC-genotype carriers are even more susceptible. This finding was confirmed when a multivariate logistic regression was carried out using the data of all recruited patients (Fig. 1A; Table 2).

When only females were evaluated, the C-allele of *MMP13* rs2252070 was also an independent risk factor for FS ($p = 0.005$, OR = 1.91, 95%CI = 1.22–2.99; additive model adjusting for gender, age, genetic ancestry and comorbidities; Supplementary Table S4; Fig. 1A; Table 2). Despite the higher frequency of the C allele among cases, this SNP was not associated with disease susceptibility in males ($p > 0.05$ for all models; Fig. 1A; Supplementary Table S5; Table 2).

Comparing the genotype frequency of *MMP9* rs17576 in females, initially we observed that the G-allele (minor allele) tended to contribute to risk of the disease ($p = 0.05$, OR = 1.51, 95%CI = 0.97–2.33; additive model adjusting for age, genetic ancestry and comorbidities; Supplementary Table S4). When all recruited females were evaluated (including samples without genet ancestry data), providing a larger sample size, the rare variant of *MMP9* rs17576 was significant associated with disease susceptibility

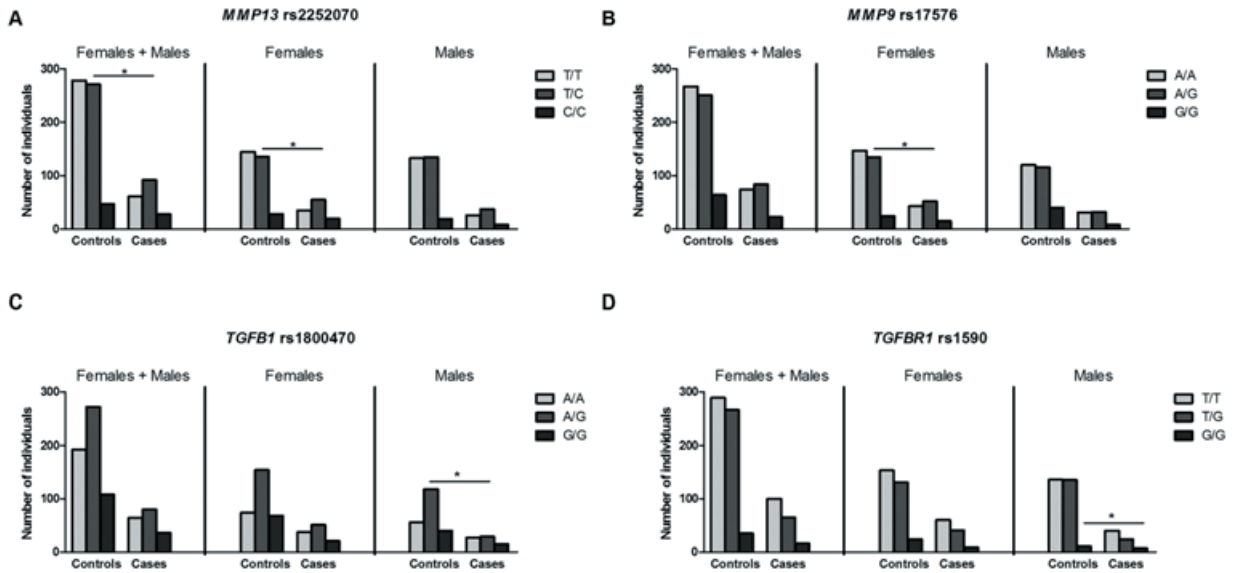

Figure 1. Genotype frequencies in the recruited individuals with or without FS. (A) *MMP13* rs2252070; (B) *MMP9* rs17576; (C) *TGFB1* rs1800470; (D) *TGFBR1* rs1590. *Significant difference between groups after adjustment by age, gender and comorbidities ($p < 0.05$).

Table 2. Risk Factors for FS Susceptibility in a Multivariate Logistic Regression

	All Subjects		Females		Males	
Variable	*p*-Value	OR (95%CI)	*p*-Value	OR (95%CI)	*p*-Value	OR (95%CI)
Age	0.188	0.98 (0.97–1.01)	0.355	0.99 (0.96–1.01)	0.001*	0.94 (0.91–0.97)
Gender (Reference: Male)	0.112	1.36 (0.93–1.98)	–	–	–	–
rs2252070 (Reference: TT; additive model)	<0.001*	1.68 (1.27–2.21)	0.001*	1.84 (1.27–2.63)	–	–
rs17576 (Reference: AA; additive model)			0.05#	1.47 (1.00–2.16)	–	–
rs1800470 (Reference: AA; dominant model)	–	–			0.008*	0.41 (0.21–0.79)
rs1590 (Refence: TT; recessive model)	–	–	–	–	0.010*	3.08 (1.31–7.23)
High blood pressure (Reference: Absence)	<0.001*	0.31 (0.19–0.50)	<0.001*	0.18 (0.09–0.38)	–	–
Dyslipidemia (Reference: Absence)	<0.001*	7.04 (4.13–11.98)	<0.001*	8.03 (3.57–18.02)	<0.001*	8.19 (3.84–17.51)
Diabetes (Reference: Absence)	–	–	0.045*	0.40 (0.16–0.98)	–	–
Thyropathy (Reference: Absence)	<0.001*	4.08 (2.38–6.99)	<0.001*	6.03 (3.09–11.77)	–	–

OR: odds ratio; CI: confidence interval. *risk factor for FS susceptibility ($p < 0.05$). #Tendency to be a risk factor ($p = 0.05$).

($p = 0.044$, additive model adjusting for age and comorbidities; Fig. 1B). However, in a model considering both *MMP9* rs2252070 and rs17576, the rs17576 variant only tended to contribute to the risk of FS ($p = 0.05$; Table 2).

In males, the presence of the T-allele of *TIMP2* rs2277698 was associated with the risk of FS in a dominant model adjusted for age and ancestry ($p = 0.035$, OR = 2.15, 95%CI = 1.06–4.36; dominant model; Supplementary Table S5). However, it was not an independent risk factor when comorbidities were included in the model.

The SNPs *TGFB1* rs1800470 and *TGFBR1* rs1590 were also involved in susceptibility to FS in males (Supplementary Table S5). The C-allele of rs1800470 was associated with a reduced risk of FS ($p = 0.04$, OR = 0.47, 95%CI = 0.23–0.96; dominant model adjusting for age, genetic ancestry and comorbidities). GG-genotype carriers of *TGFBR1* rs1590 had an increased risk of this condition ($p = 0.027$, OR = 4.11, 95%CI = 1.17–14.38; recessive model adjusting for age, genetic ancestry and comorbidities). These genetic variants were confirmed as independent risk factors for FS susceptibility when all recruited males were evaluated (Table 2; Fig. 1C and D).

Table 2 shows p values, OR and 95%CI of the final model including the significant SNPs and clinical variables as possible risk factors adjusted for age and gender (when necessary). It is important to highlight that diabetes was an independent risk factor only among females, and dyslipidemia was the unique comorbidity associated with FS risk in males ($p < 0.05$).

Haplotypes and FS Susceptibility

For the haplotype analyses, only polymorphisms in HWE were included. There was strong linkage disequilibrium in *TGFB1*, *TGFBR1*, *MMP2*, *MMP3*, and *MMP9* loci.

For between-group comparisons, p values were first adjusted for gender, age and genetic ancestry. Comparing the *MMP9* haplotypes between cases and controls, a significant association was detected between the G/G (rs17576 A>G/rs17577 G>A) haplotype and FS ($p = 0.046$, OR = 1.40, 95%CI = 1.01–1.95; Table 3) and when only females were evaluated ($p = 0.046$, OR = 1.59, 95%CI = 1.01–2.51; Table 3). In females, the T/C/G *MMP2* (rs243865 C>T/rs2285053 C>T/rs243866 G>A) haplotype was also associated with the susceptibility of FS ($p < 0.0001$; Table 3). However, the frequency of this haplotype was low in the studied population. These findings were confirmed in the analysis involving all recruited cases and controls with adjustment for age, gender and main comorbidities (Supplementary Table S6).

DISCUSSION

The present study provides additional evidence that FS is a complex disease that involves both environmental and genetic risk factors. We identified SNPs or haplotypes that were independent risk factors for this injury even in models that included common comorbidities as covariates. Therefore, analysis of these genetic variations can contribute to FS risk assessment. Currently, the diagnosis of FS is often one of exclusion, which may be difficult because the disease can share symptoms with many other shoulder conditions.[2] In that way, knowledge of the molecular basis of a disease can help in the development of improved diagnostic tools.

We observed that carriers of the C-allele of *MMP13* rs2252070 and, especially female, may present increased FS susceptibility. Moreover, CC-genotype carriers are even more susceptible. The rs2252070 is a functional variant in the Sp1 (a transcription factor) binding site of *MMP13* promoter, and this polymorphism has been found to influence gene expression.[27] The C-allele of rs2252070, but not the T-allele, can bind Sp1 and promote *MMP13* expression.[28] Similarities between the fibrosis process of FS and Dupuytren disease have been reported[29] and increased *MMP13* expression has been described in this disease.[30] Thus,

Table 3. Haplotype Frequencies in Cases and Controls

			All Subjects			Females			Males		
Gene	Haplotype		Controls	Cases	*p*-Value[a]	Controls	Cases	*p*-Value[a]	Controls	Cases	*p*-Value[a]
TGFB1	rs1800470/ rs1800469	A/G[b]	0.5063	0.5488		0.4963	0.5361		0.5201	0.5672	
		G/A	0.3887	0.3811	0.5	0.4034	0.3814	0.54	0.3677	0.3806	0.72
		G/G	0.1008	0.0701	0.44	0.0958	0.0825	0.82	0.1084	0.0522	0.17
		A/A	0.0042	NA	1	0.0045	NA	1	0.0038	NA	1
TGFBR1	rs11568785/ rs1590	A/T[b]	0.6492	0.6616		0.6452	0.6725		0.6546	0.6493	
		A/G	0.293	0.2713	0.68	0.2978	0.2657	0.71	0.2865	0.2761	0.86
		G/T	0.0578	0.0671	0.8	0.057	0.0595	0.86	0.0589	0.0746	0.54
		G/G	0	0	–	0	0.0024		0	0	–
MMP2	rs243865/ rs2285053/ rs243866	C/C/G[b]	0.7026	0.6646		0.6844	0.6701		0.7359	0.6567	
		T/C/A	0.1427	0.189	0.52	0.1597	0.1907	0.66	0.111	0.1866	0.11
		C/T/G	0.1547	0.1433	0.46	0.1559	0.134	0.2	0.1442	0.1567	0.7
		T/C/G	NA	0.003	–	NA	0.0052	<0.0001*	0.0089	0	–
		T/T/A	0	0	–	0	0	–	0	0	1
MMP3	rs679620/ rs522616	T/T[b]	0.3971	0.4207		0.3821	0.3969		0.4174	0.4552	
		C/T	0.3044	0.3232	0.59	0.3213	0.3711	0.62	0.2816	0.2537	0.99
		C/C	0.2969	0.2561	0.42	0.2966	0.232	0.34	0.2975	0.291	0.88
		T/C	0.0016	0	1	0	0	–	0.0035	0	1
MMP9	rs17576/ rs17577	A/G[b]	0.6899	0.6555		0.7064	0.6443		0.6678	0.6716	
		G/G	0.1695	0.2409	0.046*	0.1643	0.2423	0.046*	0.1766	0.2388	0.44
		G/A	0.1366	0.1037	0.41	0.1247	0.1134	0.9	0.1525	0.0896	0.23
		A/A	0.004	0	1	0.0046	0	1	0.0031	0	1

[a]*p*-value adjusted by age, gender (when applied) and genetic ancestry. [b]Reference haplotype. *Significant difference between groups; $p < 0.05$.

our study highlights the involvement of MMP collagenase in the etiology of FS.

We also identified the G (minor allele)/G (major allele) *MMP9* (rs17576/rs17577) haplotype as an independent risk factor for the disease, especially in females. This *MMP9* haplotype conferred 1.4 times the risk of FS in general and 1.7 times the risk among females. Although less predictive than *MMP13* rs2252070, the G-allele of rs17576 was also associated with disease risk among females independently from the rs17577 genotype. A haplotype represents a set of alleles which are inherited together and represent a specific interval on a chromosome and may be associated with a phenotype. Both rs17576 and rs17577 are missense polymorphisms and therefore involve the substitution of one amino acid for another, which may result in altered protein activity. rs17576 is located in the MMP9-specific fibronectin type II domains, which presumably enhance substrate binding,[31,32] and rs17577 is located in the hemopexin domain, which seems to negatively affect the substrate and inhibitor binding.[33,34] FN1 is among the proteins with a fibronectin type II domain that can be degraded by MMP9. However, FN1, which is increased in the capsule of patients with FS,[23] is also able to induce an increase of MMP9 expression.[35] The feedback between these proteins is important for tissue remodeling because, in this context, MMP9 has been suggested to release or activate a number of cytokines and growth factors, including TGFB1.[36]

The gelatinase *MMP2* T (minor-allele)/C (major-allele)/G (major-allele) haplotype (rs243865/rs2285053/rs243866) was also associated with FS susceptibility, though only among females. MMP2 presents proteolytic activity against matrix and nonmatrix proteins. Thus, it is expected that genetic variants that directly affect gene expression and/or protein function may impact the progression of pathological processes involving tissue remodeling,[37] such as FS. The rs243865 is a known functional SNP that interrupts a *MMP2* Sp1 site. It was previously demonstrated that the rs243865 C-allele presents increased activity and expression in relation to the T-allele.[37] It is worth noting that the T/C/G haplotype was infrequent in our population. Thus, investigation in a larger population is necessary to elucidate how this haplotype impacts the female FS risk susceptibility.

Both SNPs of *MMP13* and *MMP2* associated with FS risk are located in Sp1 sites. Regions rich in Sp1 sites are influenced by DNA methylation,[38] a dynamic mechanism of gene expression regulation. DNA methylation is a mechanism of control of *MMPs* expression in shoulder injuries.[39] Moreover, the impact of methylation in SNPs was shown in osteoarthritis susceptibility.[40] Thus, the dynamic nature of FS with the shoulder function improving overtime suggests that SNPs that

have an effect in the DNA methylation pattern may play a key role in the disease susceptibility.

A previous study in a Chinese population also suggested the involvement of *MMP3* rs650108 in FS risk. We did not evaluate this SNP; however, like Xu et al., we evaluated the rs679620 and no association between this genetic variant and the risk of FS when adjusting for gender and age was detected. Conversely, the CC-genotype of *MMP3* rs522616 was significantly more common among controls than among FS patients after adjusting for the same variables. However, the results were not significant once genetic ancestry was added to the model. This finding highlights the necessity of ancestry adjustment in a genetic study, especially in admixed populations. Large-scale genome-wide studies such as HapMap and 1000 genomes projects demonstrate that there is a broad range in SNPs frequencies among populations from different continents. Supplementary Table S2 exemplifies the range in SNPs frequencies by showing the MAF of the studied genes in African, Amerindian, and European populations according the 1000 genomes project. In populations with massive interethnic admixture, the establishment of one's ethnicity for research purposes is an imprecise task, and physical appearance may not be an efficient indicator of an individual's ancestry.[24,41] Therefore, genetic ancestry determination is essential to avoid the description of SNPs associated with disease risk when the SNPs are truly associated with a subpopulation in the studied samples.

In males, *TIMP2* rs2277698, *TGFB1* rs1800470, and *TGFBR1* rs1590 seem to be involved in the susceptibility to FS. However, only *TGFB1* rs1800470 and *TGFBR1* rs1590 were independent risk factors of this shoulder injury. The rare allele *TGFB1* rs1800470 (missense SNP) seems to be reduce the risk of FS. Conversely, *TGFBR1* rs1590 is located in the 3′-UTR region, which is involved in the regulation of mRNA stability, and the rare homozygous variant seems to be associated with increased disease risk. These findings reinforce the involvement of TGFβ signaling in FS, as previously demonstrated by gene expression analysis.[23]

It is worth noting that none of the SNPs associated with FS among females were significantly associated with the risk of this disease in males. In females, the significant SNPs were located in *MMP* genes, while SNPs of *TGFB1* and *TGFBR1* were associated with FS susceptibility in males. We cannot exclude that distinct genetic factors may play different roles in the disease development due to the interaction and function of diverse proteins in hormone-dependent pathways. Moreover, we believe that this gender difference may have interfered in the observation of only *MMP13* rs2252070 as a genetic factor in the analysis including both genders.

The proportion of males and females in our study is in agreement with the predominance of the disease among females.[3] Although the male population was small, the FS risk in men seems to be less influenced by common clinical comorbidities. The selection of individuals with clinically relevant phenotypes it is an intuitive approach by which to study the underlying causes of a common disease. This strategy assumes that these patients are the most informative and thus should be studied separately, rather than being included in larger series of patients who might dilute the information that they can provide.[42]

While it is original, this study presents some limitations. First, the sample size is small for a genetic study. Although we demonstrated that the evaluation of genetic variants may help in the understanding of FS etiology; validation in larger populations is still necessary, especially taking in account the necessity of genetic ancestry adjustment. Second, *p*-values were not adjusted for multiple comparisons; however, variants/genes were rationally selected based on their function in joint tissues and thus it is not a "hypothesis-free" investigation. Moreover, considering the demographic heterogeneity of patients and controls, we prioritized the correction for clinical and genetic ancestry that could lead to false positive discoveries due to sample stratification. Third, we analyzed 18 SNPs, and the analysis of other genetic variants in the studied genes or in other genes involved the tissue remodeling, inflammation or fibrosis is needed. Indeed, FS is a multifactorial disease, and therefore several genetic variants may contribute small effects to its susceptibility. Even so, the analysis of genetic variants may be used as a less subjective tool to determine FS risk after further investigations.

In conclusion, carriers of the C-allele of *MMP13* rs2252070 and of the G/G *MMP9* (rs17576/rs17577) haplotype may present increased risk of FS, especially in females. In females, the T/C/G *MMP2* (rs243865/rs2285053/rs243866) haplotype was also associated with FS risk, and the G-allele of rs17576 tended to contribute to susceptibility to the studied disease. In contrast, the C-allele of *TGFB1* rs1800470 reduces the risk of FS development, and the GG-genotype of *TGFBR1* rs1590 increases the risk of the disease in males.

AUTHOR CONTRIBUTIONS

MFL, CC, MC conceived and designed the experiments. CC and MFL were involved in data collection. CC, MFL, AW, PSB, EAF, CVA, ACP, BE, and FF were responsible by samples collection. MFL, LCL, SEBS, AKCRS, and AMO was involved in the genetic analysis. MFL were responsible literature search. MFL and MCS were involved in statistical analysis. MFL and CC wrote the first draft of the manuscript. All authors listed have contributed to all subsequent drafts, and have approved the final manuscript.

ACKNOWLEDGMENTS

This study was supported by the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) under grant #2011/22548-2 (PI: MFL) and fellowships #2012/14768-5

(MFL), #2016/01392-8 (AMO), #2016/15785-1 (AMO); #2013/23332-9 (LCL), #2015/02285-8 (LCL), and Conselho Nacional de Desenvolvimento Cientfico e Tecnológico (CNPq) under fellowship to MCS, SEBS, AKCRB, and MC. We are grateful to scientific support of Sintia Iole Belangero, Ph.D., and Fernanda Wisnieski, Ph. D., from the Universidade Federal de São Paulo (UNIFESP) and Rommel Rodŕguez Burbano, Ph.D, from the Universidade Federal do Pará (UFPA).

REFERENCES

1. Tamai K, Akutsu M, Yano Y. 2014. Primary frozen shoulder: brief review of pathology and imaging abnormalities. J Orthop Sci 19:1–5.
2. Manske RC, Prohaska D. 2008. Diagnosis and management of adhesive capsulitis. Curr Rev Musculoskelet Med 1:180–189.
3. Neviaser AS, Hannafin JA. 2010. Adhesive capsulitis: a review of current treatment. Am J Sports Med 38: 2346–2356.
4. Wang K, Ho V, Hunter-Smith DJ, et al. 2013. Risk factors in idiopathic adhesive capsulitis: a case control study. J Shoulder Elbow Surg 22:e24–e29.
5. Kordella T. 2002. Frozen shoulder & diabetes. Frozen shoulder affects 20 percent of people with diabetes. Proper treatment can help you work through it. Diabetes Forecast 55:60–64.
6. Balci N, Balci MK, Tuzuner S. 1999. Shoulder adhesive capsulitis and shoulder range of motion in type II diabetes mellitus: association with diabetic complications. J Diabetes Complications 13:135–140.
7. Austin DC, Gans I, Park MJ, et al. 2014. The association of metabolic syndrome markers with adhesive capsulitis. J Shoulder Elbow Surg 23:1043–1051.
8. Lo SF, Chu SW, Muo CH, et al. 2014. Diabetes mellitus and accompanying hyperlipidemia are independent risk factors for adhesive capsulitis: a nationwide population-based cohort study (version 2). Rheumatol Int 34:67–74.
9. Schiefer M, Teixeira PF, Fontenelle C, et al. 2017. Prevalence of hypothyroidism in patients with frozen shoulder. J Shoulder Elbow Surg 26:49–55.
10. Milgrom C, Novack V, Weil Y, et al. 2008. Risk factors for idiopathic frozen shoulder. Isr Med Assoc J 10:361–364.
11. Hakim AJ, Cherkas LF, Spector TD, et al. 2003. Genetic associations between frozen shoulder and tennis elbow: a female twin study. Rheumatology (Oxford) 42:739–742.
12. Elands RJ, Simons CC, Riemenschneider M, et al. 2017. A systematic SNP selection approach to identify mechanisms underlying disease aetiology: linking height to post-menopausal breast and colorectal cancer risk. Sci Rep 7:41034.
13. Xu Q, Gai PY, Lv HL, et al. 2016. Association of MMP3 genotype with susceptibility to frozen shoulder: a case-control study in a Chinese Han population. Genet Mol Res 15:1–7.
14. Kaltsas DS. 1983. Comparative study of the properties of the shoulder joint capsule with those of other joint capsules. Clin Orthop Relat Res 20–26.
15. Woessner JF, Jr. 1994. The family of matrix metalloproteinases. Ann N Y Acad Sci 732:11–21.
16. Mittal R, Patel AP, Debs LH, et al. 2016. Intricate functions of matrix metalloproteinases in physiological and pathological conditions. J Cell Physiol 231:2599–2621.
17. Rodeo SA, Hannafin JA, Tom J, et al. 1997. Immunolocalization of cytokines and their receptors in adhesive capsulitis of the shoulder. J Orthop Res 15:427–436.
18. Moore-Smith L, Pasche B. 2011. TGFBR1 signaling and breast cancer. J Mammary Gland Biol Neoplasia 16:89–95.
19. Cordova A, Tripoli M, Corradino B, et al. 2005. Dupuytren's contracture: an update of biomolecular aspects and therapeutic perspectives. J Hand Surg Br 30:557–562.
20. Chiquet-Ehrismann R, Tucker RP. 2011. Tenascins and the importance of adhesion modulation. Cold Spring Harb Perspect Biol 3:1–19.
21. Dallas SL, Sivakumar P, Jones CJ, et al. 2005. Fibronectin regulates latent transforming growth factor-beta (TGF beta) by controlling matrix assembly of latent TGF beta-binding protein-1. J Biol Chem 280:18871–18880.
22. Carey WA, Taylor GD, Dean WB, et al. 2010. Tenascin-C deficiency attenuates TGF-ss-mediated fibrosis following murine lung injury. Am J Physiol Lung Cell Mol Physiol 299:L785–L793.
23. Cohen C, Leal MF, Belangero PS, et al. 2016. The roles of Tenascin C and Fibronectin 1 in adhesive capsulitis: a pilot gene expression study. Clinics (Sao Paulo) 71:325–331.
24. Ramos BR, D'Elia MP, Amador MA, et al. 2016. Neither self-reported ethnicity nor declared family origin are reliable indicators of genomic ancestry. Genetica 144:259–265.
25. Little J, Higgins JP, Ioannidis JP, et al. 2009. STrengthening the REporting of Genetic Association studies (STREGA)-an extension of the STROBE statement. Eur J Clin Invest 39:247–266.
26. Sole X, Guino E, Valls J, et al. 2006. SNPStats: a web tool for the analysis of association studies. Bioinformatics 22:1928–1929.
27. Yoon S, Kuivaniemi H, Gatalica Z, et al. 2002. MMP13 promoter polymorphism is associated with atherosclerosis in the abdominal aorta of young black males. Matrix Biol 21:487–498.
28. Shi M, Xia J, Xing H, et al. 2016. The Sp1-mediaded allelic regulation of MMP13 expression by an ESCC susceptibility SNP rs2252070. Sci Rep 6:27013.
29. Hand GC, Athanasou NA, Matthews T, et al. 2007. The pathology of frozen shoulder. J Bone Joint Surg Br 89:928–932.
30. Johnston P, Chojnowski AJ, Davidson RK, et al. 2007. A complete expression profile of matrix-degrading metalloproteinases in Dupuytren's disease. J Hand Surg Am 32:343–351.
31. Allan JA, Docherty AJ, Barker PJ, et al. 1995. Binding of gelatinases A and B to type-I collagen and other matrix components. Biochem J 309:299–306.
32. O'Farrell TJ, Pourmotabbed T. 2000. Identification of structural elements important for matrix metalloproteinase type V collagenolytic activity as revealed by chimeric enzymes. Role of fibronectin-like domain and active site of gelatinase B. J Biol Chem 275:27964–27972.
33. Li J, Brick P, O'Hare MC, et al. 1995. Structure of full-length porcine synovial collagenase reveals a C-terminal domain containing a calcium-linked, four-bladed beta-propeller. Structure 3:541–549.
34. Murphy G, Knauper V. 1997. Relating matrix metalloproteinase structure to function: why the "hemopexin" domain?. Matrix Biol 15:511–518.
35. Sen T, Dutta A, Maity G, et al. 2010. Fibronectin induces matrix metalloproteinase-9 (MMP-9) in human laryngeal carcinoma cells by involving multiple signaling pathways. Biochimie 92:1422–1434.
36. Kobayashi T, Kim H, Liu X, et al. 2014. Matrix metalloproteinase-9 activates TGF-beta and stimulates fibroblast contraction of collagen gels. Am J Physiol Lung Cell Mol Physiol 306:L1006–L1015.
37. Price SJ, Greaves DR, Watkins H. 2001. Identification of novel, functional genetic variants in the human matrix metalloproteinase-2 gene: role of Sp1 in allele-specific transcriptional regulation. J Biol Chem 276:7549–7558.
38. Graff JR, Herman JG, Myohanen S, et al. 1997. Mapping patterns of CpG island methylation in normal and neoplastic cells implicates both upstream and downstream regions in de novo methylation. J Biol Chem 272:22322–22329.

39. Leal MF, Caires Dos Santos L, Martins de Oliveira A, et al. 2017. Epigenetic regulation of metalloproteinases and their inhibitors in rotator cuff tears. PLoS ONE 12:e0184141.
40. Reynard LN, Bui C, Syddall CM, et al. 2014. CpG methylation regulates allelic expression of GDF5 by modulating binding of SP1 and SP3 repressor proteins to the osteoarthritis susceptibility SNP rs143383. Hum Genet 133:1059–1073.
41. Parra FC, Amado RC, Lambertucci JR, et al. 2003. Color and genomic ancestry in Brazilians. Proc Natl Acad Sci USA 100:177–182.
42. Perez-Gracia JL, Gurpide A, Ruiz-Ilundain MG, et al. 2010. Selection of extreme phenotypes: the role of clinical observation in translational research. Clin Transl Oncol 12:174–180.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article.