Association Between a CTGF Gene Polymorphism and Systemic Sclerosis in a French Population

BRIGITTE GRANEL, LAURENT ARGIRO, ERIC HACHULLA, ISABELLE FAJARDY, PIERRE-JEAN WEILLER, JEAN-MARC DURAND, YVES FRANCES, ANNE-MARIE DOMBEY, SANDRINE MARQUET, NATHALIE LESAVRE, PATRICK DISDIER, FANNY BERNARD, PIERRE-YVES HATRON, and CHRISTOPHE CHEVILLARD

ABSTRACT. Objective. Systemic sclerosis (SSc) is a life-threatening autoimmune disease characterized by chronic fibrosis of the skin and internal organs. Connective tissue growth factor (CTGF) is believed to be a primary mediator of chronic fibrosis. We assessed the possible association between 7 single-nucleotide polymorphisms (SNP) in the CTGF gene and scleroderma in a French population (registration number 2006/0182).

Methods. We conducted a case-control study with 241 scleroderma patients and 269 controls. Seven SNP were genotyped using the TaqMan system. Univariate and multivariate analyses were performed. In silico electrophoretic mobility shift assay (EMSA), and reverse transcriptase polymerase chain reaction analyses were done to assess the effect of the SNP on CTGF gene expression.

Results. The frequency of the rs9399005TT genotype was significantly lower in SSc patients than in controls. This association remained significant after adjustment for gender. An association was detected between the rs9399005 and the diffuse and limited cutaneous forms. Multivariate analysis between SSc patients and controls taking into account all 7 SNP and sex revealed that only sex and rs9399005 SNP were associated with disease. DNA analysis by EMSA indicated that the T allele bound nuclear factors that were also bound by the C allele. The binding affinity was higher for the T allele. Analysis of the human database and experiments with human hepatocyte cell line indicated the existence of an alternative transcript containing the rs9399005 polymorphism in its 3'UTR region. In silico analysis indicated that this polymorphism may alter the structure of CTGF messenger RNA.

Conclusion. These findings suggest that CTGF gene polymorphisms may contribute to susceptibility to scleroderma. (First Release Dec 23 2009; J Rheumatol 2010;37:351–8; doi:10.3899/jrheum.090290)

Key Indexing Terms: SCLERODERMA CONNECTIVE TISSUE GROWTH FACTOR POLYMORPHISM DISEASE SUSCEPTIBILITY ELECTROPHORETIC MOBILITY SHIFT ASSAY

Systemic sclerosis (SSc) is a connective tissue disorder characterized by microvascular abnormalities, autoimmunity and fibrosis, affecting skin and internal organs including lungs and digestive tract. SSc is characterized by deposition of collagen and other extracellular matrix substances in connective tissue, leading to high morbidity and mortality. SSc is a complex, polymorphic, and heterogenous disorder. It is believed to be the consequence of external triggers acting...
ing upon a genetically susceptible host. Transforming growth factor beta (TGF-ß) may be central to the initiation of the chronic fibrosing disorder \(^1\), although its role in the maintenance of the fibrotic phenotype in SSc remains unclear. Unlike TGF-ß, expression of the connective tissue growth factor (CTGF, also known as CCN2), an autocrine stimulator released in response to TGF-ß, correlates well with the extent of skin sclerosis and the severity of pulmonary fibrosis in this disease \(^3\). These observations suggest that CTGF is involved in SSc pathogenesis. Moreover, CTGF messenger RNA (mRNA) is constitutively expressed and strongly induced after TGF-ß stimulation in skin specimens from sclerotic SSc lesions, compared to healthy skin control\(^4\). CTGF is a profibrotic protein that induces proliferation, collagen synthesis, and chemotaxis of fibroblasts. Its presence is associated with excessive matrix protein deposits\(^6\), due at least in part to its action on a CTGF response element in the type I collagen (COL1A2) gene promoter\(^7\). Subcutaneous injection of TGF-ß into newborn mice results in only a transient fibrotic response, while co-injection of TGF-ß and CTGF results in sustained fibrosis\(^8\). This work in animal models suggests a 2-step process of fibrosis in SSc: first, TGF-ß initiates the wound healing response and then CTGF promotes sustained and chronic deleterious fibrosis in a TGF-ß-independent manner\(^5\). Thus, constitutive overexpression of CTGF by fibroblasts is expected to contribute directly to chronic persistent fibrosis, the characteristic feature of the scleroderma phenotype. A genome-wide association study for regions of SSc susceptibility, performed with a Choctaw Indian population of North America with high incidence of SSc, revealed potential SSc-associated regions. One of them was located at 6q23-27, where the CTGF gene maps\(^10\). Association studies between CTGF and SSc have been recently performed that lead to conflicting results. Associations were found with the polymorphism rs6918698 (–945C/G) located in the promoter region in 2 studies\(^11,12\). However, these associations were not replicated in other studies\(^1,3,14\). We therefore investigated whether genetic variation of CTGF could confer susceptibility to SSc. We genotyped polymorphisms in this gene and tested for association with SSc in a French European Caucasian population. Associated polymorphism was functionally characterized.

**MATERIALS AND METHODS**

**Patients and controls.** We included 510 consecutive unrelated subjects, comprising 241 SSc patients classified according to the LeRoy’s criteria\(^15\) and 269 healthy controls, all of French Caucasian origin. Scleroderma patients were recruited by the Internal Medicine Departments in Marseilles (111 patients) and Lille (130 patients). Healthy controls were recruited at the Blood Bank Centre (volunteer blood donors), or included by the Clinical Investigation Centre of the Mediterranean University. They did not suffer from autoimmune disease. Controls were matched by age (± 10 years) and ethnicity to patients. The Ethics Committee of the Mediterranean University approved the study and all patients and controls gave written informed consent for all procedures (Direction Générale de la Santé, registration number 2006/0182).

**Blood samples and DNA preparation.** We collected 5 to 15 ml of blood on lithium heparin and stored it at –20°C. Genomic DNA was extracted from total blood using a phenol-chloroform method and an Autogen NA2000 (Genework; Pont-à-Mousson, France)\(^16\).

**Genotyping.** Genotype studies were performed using the TaqMan system (allelic discrimination using the 5’nuclease assay, Applied Biosystems, Courtabeuf, France) with an ABI Prism 7900 Sequence Detector (Applied Biosystems), according to the manufacturer’s instructions. Replication of 50 samples in a double-blind protocol confirmed 100% reproducibility. Seven single nucleotide polymorphisms (SNP) were genotyped: rs6918698, rs1931002, rs9493150, rs12526196, rs12527705, rs9399005, and rs12527379 (Figure 1A). These SNP were chosen for several reasons: (1) SNP rs6918698 (G-945C) mapping in the promoter of the gene was previously studied by Fonseca, et al\(^11\) and found to be a functional SNP associated with susceptibility to SSc; (2) rs1931002, rs9493150, rs9399005, and rs12527379 were selected as Tag SNP according to HapMap Data on the Caucasian reference population. The tag SNP were selected on a region covering the CTGF gene ± 10 kb, according to HapMap. For this selection, correlation groups were defined according to the \(r^2\) values. The cutoff value was 0.8; and (3) the rs12526196 and rs12527705 were added to cover the distal 3’ region of the gene (Figure 1A).

**Statistical analyses.** SPSS software was used for statistical analyses. Chi-squared tests were performed to compare genotype frequencies between cases and controls. Fisher’s exact tests were used as appropriate. Because SSc is an autoimmune disorder preferentially affecting women, and because the control group in our study could not be matched for gender, we performed a stepwise binary logistic regression analysis on the whole population including gender and polymorphisms as covariates. Because of the number of statistical tests performed, and to keep the type I error < 5%, p values were corrected as appropriate according to Nyholt’s procedure for multiple testing correction\(^17,18\). Associations remained significant after this correction (Table 3). The effective number (Meff) of independent tests was estimated by using the single nucleotide polymorphism spectral decomposition (SNPSpD) Web site (http://genepi.qimr.edu.au/general/daleN/SNPSpD/; accessed October 23, 2009): the Meff value was estimated to be 5.4 in our study. Main associations detected remained significant after correction. Linkage disequilibrium analysis was done with the haplview software.

**Nuclear extract preparation.** Nuclear extracts were prepared from a human hepatocyte cell line (HEPG2) stimulated for 1 hour with dexamethasone (1 \(\mu\)M) and from human normal dermal fibroblasts cell line (GM0038). The nuclear and cytoplasmatic extraction reagents from Pierce were used. Nuclear protein concentrations were determined with the bicinchoninic acid protein assay reagent (Pierce).

**Electrophoretic mobility shift assay (EMSA).** Complementary single-stranded oligonucleotides were synthesized commercially; they were designed to span a region extending approximately 10 base pairs (bp) on either side of the variant nucleotide, as follows: Rs9399005 oligonucleotide forward: CTA ACC TAT AA/C(T) GGC CAG AGA GGT ACA AA, Rs9399005 oligonucleotide reverse: TTT GTA CCT TCG CC/G(A) TTA TAG GTT AG.

Complementary strands were annealed by placing reactions (sense and antisense oligonucleotides) in a boiling water bath for 10 min, then cooling to room temperature. Binding reactions were set up with LightShift Chemiluminescent EMSA Kit (Pierce). Aliquots of 20 fmol of complementary DNA were incubated at room temperature for 20 min with 4 \(\mu\)g of nuclear extract in 10 mM Tris, 50 mM KCl, 1 mM DTT, 2.5% glycerol, 5 mM MgCl\(_2\), 50 ng/\(\mu\)l poly d(I-C), 0.05% NP-40, pH 7.5. Reactions were loaded onto an 8% non-denaturing polyacrylamide gel and run for 150 min at 110 V. Free DNA and DNA-protein complexes were transferred to nylon N+ membranes by capillary action. Binding was detected according to the manufacturer’s instructions (Pierce).

**RNA preparation and cDNA synthesis.** Total RNA was extracted using the
TRIzol reagent (Life Technologies) and following the manufacturer’s instructions. First strand cDNA was synthesized from 2 µg of total RNA using the High-Capacity cDNA Archive Kit (Applied Biosystems) according to the manufacturer’s protocol. To minimize variations in reverse transcriptase efficiency, all samples from a single experiment were reverse transcribed simultaneously.

Polymerase chain reaction (PCR). PCR amplifications were carried out in 30 µl reactions containing 100 ng DNA or cDNA, 10 mM Tris-HCl pH = 9.0, 0.1% TritonX-100, 50 mM KCl, 0.2 mg/ml BSA, 1.5 mM MgCl2, 1 µM of each primer, 1 mM of dNTP and 1.5 U Taq polymerase (Invitrogen). Following the initial denaturation step (94˚C 5 min), samples were subjected to 35 cycles of PCR consisting of 94˚C for 1 min, annealing temperature for 45 s, and 72˚C for 45 s. The primers used for exon3-exon4 amplification were exon3 forward 5’AGCAGCTGCAAGTACCAGTG3’, exon4 reverse 5’CCAGGCAGTTGGCCTAATC3’. The primers used to detect the rs9399005 marker in cDNA were forward 5’AAGTTCAGAACCTAGAGCA3’ and reverse 5’TCA GTC TCC ATT AAC CCT GTT G3’.

RESULTS
The clinical and immunological characteristics of the patients are presented in Table 1. Patients and controls were

Table 1. Demographic data and disease characteristics of patients with SSc (n = 241).

<table>
<thead>
<tr>
<th>Data</th>
<th>Patients, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age (yrs ± SD)</td>
<td>57.8 ± 12</td>
</tr>
<tr>
<td>Female</td>
<td>208 (86.3)</td>
</tr>
<tr>
<td>Limited cutaneous form</td>
<td>156 (64.7)</td>
</tr>
<tr>
<td>Diffuse cutaneous form*</td>
<td>77 (30.3)</td>
</tr>
<tr>
<td>Pulmonary fibrosis**</td>
<td>73 (28.9)</td>
</tr>
<tr>
<td>Pulmonary arterial hypertension**</td>
<td>26 (10.8)</td>
</tr>
<tr>
<td>Positive for anti-nuclear antibodies***</td>
<td>202 (83.8)</td>
</tr>
<tr>
<td>Positive for anticientromere antibodies</td>
<td>95 (39.4)</td>
</tr>
<tr>
<td>Positive for anti-topoisomerase I</td>
<td>68 (28.2)</td>
</tr>
</tbody>
</table>

SSc: systemic sclerosis. * According to the classification of LeRoy EC, et al15. ** Pulmonary fibrosis noted on computed tomography scans and/or abnormal pulmonary function tests (forced vital capacity less than 75%). Pulmonary arterial hypertension was confirmed in all cases by results of a right heart catheterization. *** Antinuclear antibodies were detected by indirect immunofluorescence (IIF on HEp2) with a screening dilution of 1/100. Anticientromere and antitopoisomerase I antibodies were detected by commercially available kit (Elia from Phadia).
from a French European Caucasian population; they were all genotyped for 7 CTGF polymorphisms and the association with SSc was tested for the rs6918698, rs1931002, rs9493150, rs12526196, rs12527705, rs9399005, and rs12527379 SNP, in comparison with controls (Tables 2 and 3). All the SNP were in Hardy-Weinberg equilibrium (Tables 2 and 3).

Linkage disequilibrium analysis performed on controls indicated that rs12526196 and rs12527705 markers were strongly correlated (r² value = 1). No other correlation was detected (0.01 < r² value < 0.44) between the other polymorphisms, indicating that the genotyped markers were good tag SNP.

A difference in CTGF genotype distribution between SSc patients and controls was observed for only the rs9399005 T/C polymorphism (Table 3). No association was detected for the other 6 tested polymorphisms (Table 2). In detail, the rs9399005 TT genotype was significantly less frequent in patients with SSc as a whole (2.5%) than in controls [10.0%; p = 0.001, OR = 0.23, 95% CI (0.09-0.56)]. SSc is a heterogeneous disease and cases are usually classified into 2 cutaneous forms based on the extent of skin fibrosis: diffuse cutaneous and limited cutaneous forms. The prevalence of the TT genotype also differed between both these subgroups of patients and controls: 2.6% of the diffuse cutaneous subgroup vs 10.0% of the control group had the TT genotype [p = 0.001, OR = 0.23, 95% CI (0.09-0.56)]; and 2.6% of the limited cutaneous subgroup and 10.0% of the control group had the TT genotype [p = 0.008, OR = 0.24, 95% CI (0.08-0.69); Table 3]. No significant association was observed between the rs9399005 SNP marker and the serum specific autoantibody profile: uncorrected p value of

---

Table 2. Genotype and allele distributions of the CTGF SNP in patients with SSc and controls.

<table>
<thead>
<tr>
<th>SNP/Subgroup</th>
<th>No. Subjects</th>
<th>GG (n (%)</th>
<th>GC (n (%))</th>
<th>CC (n (%))</th>
<th>G/C Frequency (%)</th>
<th>Allele AS* and HW †</th>
<th>AS* and HW †</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs6918698</td>
<td>Controls</td>
<td>269</td>
<td>83 (30.9)</td>
<td>115 (42.8)</td>
<td>71 (26.4)</td>
<td>52.2/47.8</td>
<td>AS: p = 0.61</td>
</tr>
<tr>
<td></td>
<td>Patients with SSc</td>
<td>241</td>
<td>61 (28.2)</td>
<td>111 (51.4)</td>
<td>44 (20.4)</td>
<td>48.3/51.7</td>
<td>HW: p = 0.06</td>
</tr>
<tr>
<td>rs12527705</td>
<td>Controls</td>
<td>268</td>
<td>1 (0.4)</td>
<td>45 (16.8)</td>
<td>222 (82.8)</td>
<td>91.2/8.8</td>
<td>AS: p = 0.16</td>
</tr>
<tr>
<td></td>
<td>Patients with SSc</td>
<td>223</td>
<td>3 (1.3)</td>
<td>45 (20.2)</td>
<td>175 (78.5)</td>
<td>88.5/1.5</td>
<td>HW: p = 0.72</td>
</tr>
<tr>
<td>rs12526196</td>
<td>Controls</td>
<td>267</td>
<td>TT (83.1)</td>
<td>TA (16.5)</td>
<td>AA (0.4)</td>
<td>91.4/8.6</td>
<td>AS: p = 0.10</td>
</tr>
<tr>
<td></td>
<td>Patients with SSc</td>
<td>216</td>
<td>167 (77.3)</td>
<td>47 (21.8)</td>
<td>2 (0.9)</td>
<td>88.2/11.8</td>
<td>HW: p = 0.75</td>
</tr>
<tr>
<td>rs9493150</td>
<td>Controls</td>
<td>268</td>
<td>135 (50.4)</td>
<td>47 (17.2)</td>
<td>36 (13.4)</td>
<td>68.5/31.5</td>
<td>AS: p = 0.68</td>
</tr>
<tr>
<td></td>
<td>Patients with SSc</td>
<td>212</td>
<td>109 (51.4)</td>
<td>78 (36.8)</td>
<td>25 (11.8)</td>
<td>69.8/30.2</td>
<td>HW: p = 0.03</td>
</tr>
<tr>
<td>rs1931002</td>
<td>Controls</td>
<td>269</td>
<td>TT (9.7)</td>
<td>TC (57.2)</td>
<td>210 (71.2)</td>
<td>88.7/11.3</td>
<td>AS: p = 0.14</td>
</tr>
<tr>
<td></td>
<td>Patients with SSc</td>
<td>202</td>
<td>2 (1.0)</td>
<td>30 (14.9)</td>
<td>170 (84.2)</td>
<td>91.6/8.4</td>
<td>HW: p = 0.68</td>
</tr>
<tr>
<td>rs12527379</td>
<td>Controls</td>
<td>226</td>
<td>GG (43.1)</td>
<td>GA (38.6)</td>
<td>49 (18.4)</td>
<td>62.4/37.6</td>
<td>AS: p = 0.55</td>
</tr>
<tr>
<td></td>
<td>Patients with SSc</td>
<td>226</td>
<td>81 (35.8)</td>
<td>111 (49.1)</td>
<td>34 (15.0)</td>
<td>60.4/39.6</td>
<td>HW: p = 0.01</td>
</tr>
</tbody>
</table>

* AS: Association study: univariate analyses were performed to compare genotype frequencies between cases and controls. † HW: Hardy-Weinberg equilibrium: Tests were done on control group. The selected cut off p value was 0.001. GTGF: connective tissue growth factor; SSc: systemic sclerosis; SNP: single-nucleotide polymorphisms.

Table 3. Frequencies of rs9399005 SNP genotypes and alleles in patients with SSc and controls.

<table>
<thead>
<tr>
<th>Population</th>
<th>No. Subjects</th>
<th>TT (n (%))</th>
<th>TC (n (%))</th>
<th>CC (n (%))</th>
<th>TC + CC (n (%))</th>
<th>T allele (p corr)</th>
<th>p value* (p corr)</th>
<th>OR CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Controls</td>
<td>269</td>
<td>27 (10.0)</td>
<td>94 (34.9)</td>
<td>148 (59.8)</td>
<td>242 (90.0)</td>
<td>27.50</td>
<td>0.001 (0.005)</td>
<td>0.23 (0.09-0.56)</td>
</tr>
<tr>
<td>SSc patients</td>
<td>241</td>
<td>6 (2.5)</td>
<td>91 (37.3)</td>
<td>144 (59.8)</td>
<td>235 (95.7)</td>
<td>21.36</td>
<td>0.001 (0.005)</td>
<td>0.23 (0.09-0.56)</td>
</tr>
<tr>
<td>dcSSc</td>
<td>77</td>
<td>2 (2.6)</td>
<td>28 (36.4)</td>
<td>47 (61.0)</td>
<td>75 (97.4)</td>
<td>20.78</td>
<td>0.001 (0.005)</td>
<td>0.23 (0.09-0.56)</td>
</tr>
<tr>
<td>lcSSc</td>
<td>156</td>
<td>4 (2.6)</td>
<td>58 (37.2)</td>
<td>94 (60.2)</td>
<td>152 (97.4)</td>
<td>21.15</td>
<td>0.008 (0.043)</td>
<td>0.24 (0.08-0.69)</td>
</tr>
</tbody>
</table>

SSc: systemic sclerosis; dcSSc: diffuse cutaneous SSc; lcSSc: limited cutaneous SSc. CI: confidence interval. P value: univariate analyses were performed to compare genotype frequencies (TT vs TC + CC) between cases and controls. p corr: p values were corrected according to the method for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other17,18. Hardy-Weinburg equilibrium test was done on control group. The obtained value (0.13) was not significant. The selected cutoff p value was 0.001.
0.022 for comparison between patients positive for antici-
tromere antibodies (ACA) and controls and uncorrected p
value of 0.998 for comparison between patients positive for
antitopoisomerase I (anti-topo I) and controls.

There was a difference in sex ratio between SSc patients
(86.3% women) and control subjects (44% women), thus we
included sex as a covariate in analyses. A multivariate an-
alysis including sex and the polymorphism rs9399005T/C as
covariates confirmed the association between this SNP and
SSc disease [sex: p < 0.001, OR = 8.26, 95% CI
(5.26–12.82), and rs9399005T/C polymorphism p = 0.001,
OR = 0.19, 95% CI (0.07–0.50)].

We also performed an analysis restricted to the female
population of patients and controls. The frequency of the
rs9399005 TT genotype was significantly lower among the
female patients with SSc (2.9%) than among the female
control group (11.5%; p = 0.002; Table 4). The frequency
of the rs9399005 TT genotype was 2.8% in the female limited
cutaneous subgroup vs 11.5% in the female control
group (p = 0.005); there was also a trend for association
between this SNP and the female diffuse cutaneous sub-
group (p = 0.069).

Multivariate regression analysis between SSc patients
and controls, taking into account all 7 SNPs and sex,
revealed that only sex and the rs9399005 SNP were associ-
ated with disease. Being a woman was a risk factor [p <
0.001, OR = 8.26, 95% CI (5.03–13.51)] for SSc, while the
rs9399005T/T genotype was protective [p = 0.002, OR =
0.21, 95% CI (0.07–0.57)]. The same multivariate analysis
done with a limited cutaneous subgroup gave similar results:
again, the rs9399005T/T genotype was protective [p =
0.017, OR = 0.25, 95% CI (0.08–0.78)]. Multivariate analy-
sis on the diffuse cutaneous subgroup gave a trend of asso-
ciation only for the rs9399005 SNP (p = 0.054).

We then analyzed the effect of the rs9399005 polymor-
phism, in vitro by EMSA with nuclear extracts from dexa-
methasone-stimulated hepatocytes. A DNA/protein complex
was formed with both the rs9399005T allele and the
rs9399005C allele. However, the binding efficiency was
higher with the rs9399005T allele than with the rs9399005C
allele (Figure 1B). Some competition reactions were carried
out with 100 (5×), 200 (10×), and 400 (20×) fmol of unbi-
otinylated rs9399005A/T and G/C probes. The unbiotinylat-
ed rs9399005G/C probe did not compete with the biotiny-
lated rs9399005A/T probe for binding. The unbiotinylated
rs9399005A/T probe partially competed with the biotinylat-
ed rs9399005A/T probe (at 20× concentration, no binding
was observed confirming the specificity of the binding/
bond; Figure 1). We performed the same analysis with the
nuclear extract from unstimulated human normal dermal
fibroblasts (GM00038). Here again, the binding efficiency
was higher with the rs9399005T allele than with the
rs9399005C allele (data not shown).

Only one CTGF transcript (NM_001901) was found in
the genome database: it is 2358 bp long and has a 1083 bp
3’UTR region. Blast analysis of this reference sequence
identified another entry describing a longer CTGF transcript
with a longer 3’UTR region (CA413027); the untranslated
region is 1604 bp long and contains the rs9399005 poly-
morphism. However, this sequence has been reported only
once. To confirm the existence of this transcript, we extract-
ed mRNA from stimulated HEPG2, treated it with DNase
and used reverse transcriptase-polymerase chain reaction
to test for CTGF transcripts. The cDNA products were tested
with 2 pairs of primers (Figure 2A). First, we tested a primer
pair covering exon 3 and exon 4. The sizes of the amplified
products from genomic DNA and cDNA were different (381
bp and 250 bp, respectively), indicating the absence of DNA
contamination. The second pair of primers flanked a short
sequence containing the marker rs9399005. An amplifica-
tion product of 173 bp was obtained from both genomic
DNA and stimulated hepatocytes cDNA (Figure 2A). These

---

**Table 4. Frequencies of rs9399005 SNP genotypes in cases vs controls according to gender.**

<table>
<thead>
<tr>
<th>Population</th>
<th>No. Subjects</th>
<th>TT n (%)</th>
<th>TC + CC n (%)</th>
<th>p*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Women</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Controls</td>
<td>113</td>
<td>13 (11.5)</td>
<td>100 (88.5)</td>
<td>0.002</td>
</tr>
<tr>
<td>Patients with SSc</td>
<td>208</td>
<td>6 (2.9)</td>
<td>208 (97.1)</td>
<td>0.002</td>
</tr>
<tr>
<td>Diffuse cutaneous form</td>
<td>60</td>
<td>2 (3.3)</td>
<td>58 (96.7)</td>
<td>0.069</td>
</tr>
<tr>
<td>Limited cutaneous form</td>
<td>143</td>
<td>4 (2.8)</td>
<td>139 (97.2)</td>
<td>0.005</td>
</tr>
<tr>
<td>Positive for anticentromere antibodies</td>
<td>91</td>
<td>1 (1.1)</td>
<td>90 (98.9)</td>
<td>0.003</td>
</tr>
<tr>
<td>Positive for antitopoisomerase</td>
<td>57</td>
<td>0 (0)</td>
<td>57 (100)</td>
<td>0.005</td>
</tr>
<tr>
<td><strong>Men</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Controls</td>
<td>144</td>
<td>14 (9.7)</td>
<td>130 (90.3)</td>
<td>Not done</td>
</tr>
<tr>
<td>Patients with SSc</td>
<td>33</td>
<td>0 (0)</td>
<td>33 (100)</td>
<td>Not done</td>
</tr>
</tbody>
</table>

* P value: univariate analyses were performed to compare genotype frequencies between cases and controls
(analysis was limited to women). P values remained significant after correction according to the method for mul-
tiple testing for SNP in linkage disequilibrium with each other17,18, except for the association between the
rs9399005 SNP and the diffuse cutaneous form of the disease.
results confirmed the existence of an alternative CTGF transcript with a longer 3'UTR.

We used the RNAfold program to predict the minimum free energy secondary structure of the long CTGF RNA containing the rs9399005 polymorphism. Two different structures were predicted (Figure 2B) with similar energies: allelele T: $\Delta G = -1016.49$ kcal/mol; allele G: $\Delta G = -1018.00$ kcal/mol. The 2 predicted secondary structures differed in the central part of the molecules (Figure 2B).

**DISCUSSION**

Our study shows that the CTGF rs9399005 T/C marker is associated with SSc, and indeed with both diffuse and limited cutaneous forms of the disease. These associations were significant after correction for multiple statistical testing. The sex ratio was different between patients and controls, with women being much more affected by SSc than men, so we tested whether this difference affected the associations identified. Inclusion of sex in the statistical analysis con-
firmed the association between the rs9399005 marker and SSc.

For women, the CTGF rs9399005 T/C marker was significantly associated with limited forms of the disease (p = 0.005), and there was a trend of association with diffuse forms (p = 0.069). The finding of a trend was probably due to the small number of women presenting the diffuse cutaneous form (n = 60), which is typically the rarer form of SSc. Moreover, the rs9399005 TT genotype was less frequent in women positive for the 2 SSc classical autoantibodies (ACA and anti-topo I) than in control women. Thus, CTGF rs9399005 T/C seems to be a marker of the disease, and is not restricted to a particular subgroup of patients characterized by the extent of skin involvement (cutaneous limited or cutaneous diffuse) or the presence of specific serum autoantibodies (ACA or anti-topo I). Because of the small numbers of patients and to avoid multiple statistical tests, we decided not to test association with other subgroups of patients [for example, patients with pulmonary fibrosis (n = 73) or pulmonary arterial hypertension (n = 26)].

No association was detected between the disease and the other SNP tested (rs6918698, rs1931002, rs9493150, rs12526196, rs12527705, and rs12527379). In particular, no association was detected with the rs6918698 polymorphism (p = 0.62). The same absence of association was found when sex was included as a covariate (p = 0.99). Previous genetic association studies of CTGF in SSc reported discordant results for this marker (also called G-945C11-14). Fonseca, et al included 2 study groups, all of United Kingdom origin. In the first group, the authors found an excess of the GG genotype in the SSc group (27.5% of patients) relative to controls (18%, p = 0.03). This was replicated in the independent set (group 2): 32.3% of the SSc patients were GG homozygotes vs 19.9% of the controls (p = 0.001). When the 2 groups were merged, the difference in the frequency of GG homozygotes was statistically significant [30.4% of SSc patients were GG homozygotes vs 19.2% of the controls; p < 0.001, OR = 2.2, 95% CI (1.5-3.2)]. However, Gourh, et al failed to replicate the association between this SNP and SSc in a large cohort of subjects (994 SSc patients and 668 controls), all from North America: 29% of SSc patients were GG homozygotes vs 29.8% of the controls (p = 0.83). This lack of association was also found by Rueda, et al. In their study, the GG genotype frequency was not significantly different between cases and controls in 7 independent study populations from Spanish, French, Dutch, German, British, Swedish, and North American origin: the GG genotype frequency was observed between 24.4% and 29.9% in cases, and between 26.6% and 33.7% in controls. Moreover, in our study, the GG frequency in both subgroups (cases and controls) was similar to that observed by Gourh, et al and Rueda, et al.

Insufficient statistical power is one of the possible explanations for these diverse findings, even if this polymorphism is highly informative. However, this is unlikely to be the cause of the discrepancy between the Fonseca and Gourh studies, as they both involved large cohorts of subjects. Fonseca, et al suggest that the disagreement between these studies may be due to different gene-environment interactions in the 2 populations or to differences in the genetic backgrounds (between American and English populations).

According to this second hypothesis, in some populations, the previous associated SNP (rs6918698) may also be in linkage disequilibrium or correlated (r²) with another marker (the functional one) located within or outside the same gene. It is possible that in a different population the associated SNP (rs6918698) is not in linkage disequilibrium or correlated with the causal SNP. In that population, rs6918698 would not be associated with disease but the true disease-associated SNP would be.

The marker highly correlated to the SNP rs6918698, found associated with SSc in the Fonseca study, may be the rs9399005 polymorphism. In our study population these 2 markers are not correlated (r² = 0.36). Moreover, on the HapMap reference populations, the markers rs9399005 and rs6918698 were genotyped and r² values are highly variable from one population to the other (0.03 < r² < 0.74; populations analyzed in HapMap: ASW: African ancestry in Southwest USA, CEU: Utah residents with Northern and Western European ancestry from the CEPH (Centre d’Etude du Polymorphisme Humain; Evry, France) collection, CHB: Han Chinese in Beijing, China, CHD: Chinese in Metropolitan Denver, Colorado, GIH: Gujarati Indians in Houston, Texas, JPT: Japanese in Tokyo, Japan, LWK: Luhya in Webuye, Kenya, MEX: Mexican ancestry in Los Angeles, California, MKK: Maasai in Kinyawa, Kenya, TSI: Tuscans in Italy, and YRI: Yoruban in Ibadan, Nigeria). Thus, in some populations these 2 markers are correlated (r² > 0.7), while in other populations, the 2 polymorphisms are not correlated (low r² values), which could explain the difficulty of confirming the association in replicate analysis.

The findings of our genetic association studies involving the CTGF gene suggest an association between the rs9399005 SNP marker of this gene and the disease. EMSA showed that this DNA sequence is involved in a DNA/protein complex, and that the protein binding is more efficient for the T allele. EMSA findings clearly depend on the source of the nuclear extract. Here, we used nuclear extracts from a stimulated HEPG2 that produces CTGF mRNA and confirmed our results on nuclear extracts from human normal skin fibroblasts.

Analysis of the RNA indicated that there are at least 2 alternative transcripts for the CCN2 gene differing by the 3’UTR region. Moreover, the polymorphism rs9399005 may modify the secondary structure of the whole mRNA molecule and may therefore affect stability of the transcript (Figure 2B).

The 3’UTR region can be viewed as a regulatory region required for appropriate expression of many genes. This
region can regulate translation, control the deadenylation and readenylation processes that affect the length of the poly(A) tail and the nuclear export, and can determine subcellular targeting and rates of degradation of mRNA.

A cis-acting determinant, common to both transcripts, acts as a post-transcriptional repressive element. It binds nuclear factors and inhibits gene expression. However, this element alters neither the nuclear export status of cis-linked mRNA nor its intracellular stability, suggesting that there are other regulatory sequences for these functions.

The rs9399005T polymorphism may affect the transcription or stabilization of CTGF mRNA. We are currently studying patients and controls to determine whether the rs9399005T allele alters the relative abundance of the 2 transcripts.

CTGF is a potential specific target for therapeutic approaches to fibrotic diseases such as SSc. Work with animal models of fibrosis induced by TGF-β injection have recently shown that it may be possible to inhibit skin fibrosis in vivo by treatment with neutralizing monoclonal antibodies to human CTGF (anti-CTGF antibodies). Our findings implicate the CTGF gene in the pathogenesis of SSc, and particularly in the fibrotic process of the disease. Although we have not identified novel diagnostic tools with direct consequences for clinical practice, our work contributes to the understanding of the CTGF pathway in SSc. It consequently provides a basis for further analysis of this SNP in independent cohorts of SSc patients and controls, as replication is needed to validate genetic associations. Modulation of the CTGF pathway can offer new targeted therapeutic strategies to control the progression of fibrosis in SSc.

ACKNOWLEDGMENT

We thank patients and controls for their participation in this study. We thank A. Dessein for helpful discussions and comments on the manuscript.

REFERENCES